

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Ecole Nationale Polytechnique



Département de Génie Industriel

Mémoire de projet de fin d'études
pour l'obtention du diplôme d'ingénieur d'état en Management Industriel

Thème

Application du Data Mining dans la segmentation des clients dans le
secteur automobile.

Cas d'application : Renault Algérie.

Présenté par : M^{lle} BENAICHE Hadjer
M. HADDAM Mohammed Yacine

Sous la direction de : Dr. ZOUAGHI Iskander (ENP)
Mme. BENARAB Sarah (Renault Algérie)

Présenté et soutenu publiquement le (20/06/2016)

Composition du jury :

Présidente : Mme. ABOUN Nacera, MAA, ENP
Examineur : M. BOUKABOUS Ali, MAA, ENP
Promoteur : Dr. ZOUAGHI Iskander, MCB, ENP

DEDICACES

Je dédie ce travail aux êtres les plus chers à mon cœur :

A ma précieuse maman, qui par son amour, son dévouement, ses nobles valeurs et son courage, m'a toujours inspirée et soutenue.

A mon cher papa, pour ses riches enseignements, sa gentillesse et ses sacrifices consentis pour mon éducation.

A Faten et Salim pour leur bonté et éternels encouragements. Merci de remplir ma vie de joie.

A ma première école dans la vie, mon grand-père Rabi que j'aime tant. Ces lignes sont pour moi, la traduction de mon incommensurable redevance et remerciement.

A mon binome et ami Yacine, pour sa persévérance et sa bonne humeur.

Hadjer

A ma très chère mère, symbole de bonté et source de tendresse.

Pour son dévouement, ses encouragements et ses prières, sans lesquels rien n'aurait été possible.

A mon cher père, source d'inspiration. Pour son engagement sans faille, et ses conseils avisés.

A mon adorable sœur Rym et mon frère Lotfi, pour leur soutien fraternel et leur complicité.

A ma binome Hadjer, pour sa patience, son esprit aiguisé, sa rigueur et son professionnalisme.

Yacine

REMERCIEMENTS

Nos sincères remerciements et notre gratitude vont tout d'abord à notre encadreur, Dr. Iskander ZOUAGHI, pour son engagement à nos côtés tout au long de notre travail, pour ses conseils avisés et sa disponibilité permanente ainsi que pour ses hautes qualités professionnelles et humaines qui nous ont beaucoup touchés.

Veillez trouver ici, l'expression de notre profond respect et de notre grande estime.

Nous tenons également à remercier chaleureusement notre encadreur au sein de l'entreprise Renault Algérie, Mme. Sarah BENARAB, pour son aide inestimable, ses encouragements et son soutien inconditionnel dans les moments difficiles.

Nos vifs remerciements s'adressent aussi aux équipes des directions commerciale et marketing de Renault Algérie pour leur intérêt et leur précieuse contribution à l'élaboration de ce travail, ainsi qu'à l'ensemble du personnel de l'entreprise pour son chaleureux accueil.

Nous profitons de cette occasion pour exprimer notre reconnaissance envers l'ensemble du corps académique du département Génie Industriel de l'Ecole Nationale Polytechnique, pour les valeurs transmises et les enseignements prodigués durant les trois années de spécialité.

Nous souhaitons remercier tout particulièrement Mme. Nacéra ABOUN de nous avoir fait l'honneur de présider notre jury ainsi que M. Ali BOUKABOUS de nous avoir accordé de son temps pour examiner notre travail.

Hadjer BENAICHE

Yacine HADDAM

ملخص

الهدف من هذا المشروع هو المساهمة في تحسين استراتيجيات التسويق لرونو الجزائر، من خلال استخدام تقنيات التنقيب عن المعطيات في تقسيم الزبائن لتحقيق هذا الهدف، دراسة اولية تسمح لنا بتجزئة زبائن الشركة من اجل تحديد توقعات كل جزء وتحسين أداء حملات إطلاق منتج جديد. بعد ذلك، ننجز دراسة ثانية لتجزئة السوق، وذلك لتمكين الشركة من مواجهة التحديات الجديدة التي يواجهها قطاع السيارات في الجزائر.

كلمات البحث : إدارة العلاقات مع الزبائن، تجزئة الزبائن، تنقيب عن المعطيات، تجميع، تصنيف، صناعة السيارات.

Abstract

The objective of this project is to contribute to the improvement of marketing strategies of Renault Algeria through the use of Data Mining techniques in customer segmentation. To achieve this objective, a first study allow us to establish a segmentation of the corporate customers in order to identify the expectations of each segment and optimize the performance of product launch campaigns. Then, we conduct a second study to perform a market segmentation, in order to enable the company to meet the new challenges that face the automotive sector in Algeria.

Keywords : Customer Relationship Management, Customer segmentation, Data mining, Clustering, Classification, Automotive Industry.

Résumé

L'objectif de ce projet est de contribuer à l'amélioration des stratégies Marketing de l'entreprise Renault Algérie, à travers l'utilisation des techniques de Data Mining dans la segmentation des clients.

Pour réaliser cet objectif, une première étude nous permet d'établir une segmentation des clients de l'entreprise afin d'identifier les attentes des différents segments et d'optimiser la performance des campagnes marketing relatives au lancement d'un nouveau produit. Nous réalisons ensuite une seconde étude visant la segmentation du marché, dans le but de permettre à l'entreprise de répondre aux nouveaux enjeux auxquels fait face le secteur automobile en Algérie.

Mots clés : Gestion de la relation client, Segmentation des clients, Forage des données, Classification, Classement, Industrie automobile.

Table des matières

LISTE DES TABLEAUX.....	7
LISTE DES FIGURES.....	8
LISTE DES ABREVIATIONS	10
INTRODUCTION GENERALE	11
CHAPITRE I : GESTION DE LA RELATION CLIENT ET APPLICATION DU DATA MINING DANS LA SEGMENTATION DES CLIENTS.	13
Introduction.....	14
Section 1 : Gestion de la relation client	15
1. Développement de la relation client.....	15
2. Définition et approches CRM	16
3. Processus CRM	18
4. Implantation d'une stratégie CRM en entreprise	20
Section 2 : Application du Data Mining dans la segmentation des clients.....	22
1. Concept de Data Mining	22
2. Techniques de Data Mining	23
3. Logiciels de Data Mining.....	38
4. Segmentation des clients	39
5. Utilisation du Data Mining dans la segmentation des clients	46
Conclusion	51
CHAPITRE II : PRESENTATION DE L'ENTREPRISE ET EXPLOITATION DE LA BASE DE DONNEES CLIENTS.	52
Introduction.....	53
Section I : Présentation de l'entreprise	54
1. Le Groupe Renault dans le monde	54
2. Renault en Algérie.....	58
3. Définition de la problématique.....	66

Section II : Diagnostic et exploitation de la base de données clients	70
1. Cycle de vie du client et points de collecte des données.....	70
2. Système d'information client (SIC)	74
3. Diagnostic de la base de données clients : BCS.....	76
4. Exploitation de la base de données clients	80
Conclusion	100
CHAPITRE III : METHODOLOGIE, CONCEPTION DES MODELES ET INTERPRETATION DES RESULTATS.	101
Introduction	102
Section I : Segmentation des clients de Renault Algérie	103
1. Collecte des données	103
2. Sélection des données.....	107
3. Préparation des données	108
4. Transformation des données.....	108
5. Application du Data Mining.....	111
6. Interprétation des résultats	117
Section II : Segmentation du marché automobile algérien	126
1. Collecte des données	126
2. Sélection des données.....	127
3. Préparation des données	127
4. Transformation des données.....	128
5. Application du Data Mining.....	132
6. Interprétation des résultats.....	140
Conclusion	148
CONCLUSION GENERALE	149
BIBLIOGRAPHIE	151
ANNEXES	155

Liste des tableaux

Tableau I. 1: Types de segmentations des clients en fonction des besoins marketing	45
Tableau II. 1 : Répartition des clients BCS par marque	77
Tableau II. 2: Source et pourcentage des décalage des champs de la base BCS.	78
Tableau II. 3: Récapitulatif de la fusion des bases de données BCS et INCADEA.....	82
Tableau II. 4: Seuils des casiers de regroupements pour la discrétisation du score RFM.	85
Tableau II. 5: Résultats de la classification non supervisée pour les trois méthodes.....	87
Tableau II. 6: Contribution de chaque variable dans la constitution des clusters	90
Tableau II. 7: Résultats de la régression logistique au 3 ^{ème} pas.	98
Tableau III. 1: Comparaison des méthodes de sondage	105
Tableau III. 2: Statistiques de l'enquête réalisée par le Call Center.	107
Tableau III. 3: Caractéristiques des 7 clusters de la première segmentation des clients.....	113
Tableau III. 4: Caractéristiques des 7 clusters de la deuxième segmentation des clients.	117
Tableau III. 5: Clusters obtenus après la première classification_ Algorithme K-means.....	135
Tableau III. 6: Clusters obtenus après la seconde classification_ Algorithme TwoStep.	138
Tableau III. 7: Catégories démographiques et socioéconomiques des 7 segments.....	141

Liste des figures

Figure I. 1 : Représentation de l'approche tridimensionnelle du CRM.	18
Figure I. 2: Les processus CRM.....	20
Figure I. 3: Etapes du Processus de Data Mining selon CRISP-DM.	23
Figure I. 4: Approches de classification.....	34
Figure I. 5: Déroulement de la classification ascendante hiérarchique.....	35
Figure I. 6: Coupure du dendrogramme	36
Figure I. 7: Classification des méthodes de Data Mining dans le CRM	46
Figure II. 1: Actionnariat du Groupe Renault au 31 Décembre 2015	54
Figure II. 2: Modules communs à Renault et Nissan.	56
Figure II. 3: Pourcentages de participation de Renault et Nissan dans l'Alliance.....	56
Figure II. 4: Structure du Groupe Renault (en % d'actions émises)	58
Figure II. 5: Organigramme de Renault Algérie.	60
Figure II. 6: Ventes par marque en 2015.....	64
Figure II. 7: Ventes des marques Renault, Dacia, Peugeot et Hyundai, de 2011 à 2015.....	65
Figure II. 8: Evolution des ventes de véhicules neufs entre 2008 et 2015 en Algérie.	66
Figure II. 9: Cartographie niveau 2 des processus intervenant dans le cycle de vie du client.	71
Figure II. 10: Cartographie des points d'interaction_Phase achat d'un véhicule neuf.	73
Figure II. 11: Cartographie des points d'interaction_Phase achat des services après-vente. ..	74
Figure II. 12: Architecture du système d'information client de Renault Algérie	75
Figure II. 13: Interactions entre l'entreprise Renault Algérie et ses clients/prospects.....	76
Figure II. 14: Exemple de décalage des lignes dans la base BCS.....	78
Figure II. 15: Flux du modèle de transformation de données sur SPSS Modeler.....	83
Figure II. 16: Histogramme des effectifs, issu de la discrétisation du score RFM.	86
Figure II. 17: Flux du modèle Profil_client, sur SPSS Modèle.....	86
Figure II. 18: Flux du modèle prédictif, sur IBM SPSS Modeler.	88
Figure II. 19: Résultat de mesure de la silhouette du modèle K-means_ IBM SPSS Modeler	89
Figure II. 20: Comparatif des clusters de la classification k-means.....	91
Figure II. 21: Arbre de décision généré par la méthode C&RT 1 sur IBM SPSS Modeler.	93
Figure II. 22: Résultat de l'analyse de prédiction du modèle de régression logistique.....	98
Figure III. 1: Etapes du Knowledge Discovery	102
Figure III. 2: Méthodes de sondage	104
Figure III. 3: Codage des données sur IBM SPSS Statistics (Variable : Motorisation).....	108
Figure III. 4: Rotation de la première matrice des composantes avec IBM SPSS Statistics.	109
Figure III. 5: Rotation de la deuxième matrice des composantes avec IBM SPSS Statistics.	110
Figure III. 6: Récapitulatif du modèle Two-Step_ Première segmentation des clients.....	112
Figure III. 7: Modèle de la première classification des clients sur IBM SPSS Modeler.....	112
Figure III. 8: Répartition des segments de véhicules dans les clusters 2 et 5.	113
Figure III. 9: Classement des clients en fonction de : Score, IBM SPSS Modeler.	115
Figure III. 10: Résultat du modèle de classement : Score.....	115
Figure III. 11: Modélisation de la deuxième classification des clients, IBM SPSS Modeler.	116
Figure III. 12: Comparaison des clusters en fonction de la variable APV.....	118
Figure III. 13: Réseau bayésien obtenu.	119
Figure III. 14: Table des probabilités conditionnelles du noeud \$T-TwoStep.....	119

Figure III. 15: Profiling sociodémographique des clusters cibles.....	121
Figure III. 16: Comparaison des clusters en fonction de la variable : Caractéristiques.....	122
Figure III. 17: Comparaison des clusters en fonction de la variable : Finance.....	123
Figure III. 18: Comparaison des clusters en fonction de la variable : Prestige.....	124
Figure III. 19 : Comparaison des segments clients en fonction des sources d'informations..	125
Figure III. 20: Codage des variables sur IBM SPSS Statistics.....	128
Figure III. 21: Calcul d'une variable sur IBM SPSS Statistics.....	129
Figure III. 22: Rotation de la matrice des composantes principales, IBM SPSS Statistics. ..	130
Figure III. 23: Résultat de la rotation de la matrice des composantes, IBM SPSS Statistics.	131
Figure III. 24: Première classification non supervisée automatique_IBM SPSS Modeler. ...	134
Figure III. 25: Résultats du modèle K-means à 5 clusters.	134
Figure III. 26: Deuxième classification non supervisée automatique_IBM SPSS Modeler. .	136
Figure III. 27: Résultats du modèle TwoStep à 7 clusters.	137
Figure III. 28: Modélisation de la régression logistique sur IBM SPSS Modeler.	139
Figure III. 29: Résultat de l'analyse du modèle de régression logistique.....	140
Figure III. 30: Comparaison des clusters en fonction de la variable FINANCE.	142
Figure III. 31: Comparaison des clusters en fonction de : CARACTERISTIQUES.	143
Figure III. 32: Comparaison des clusters en fonction de la variable UTILITAIRE.....	144
Figure III. 33: Comparaison des clusters en fonction de la variable PRESTIGE.	145
Figure III. 34: Comparaison des clusters en fonction des sources d'information utilisées....	146
Figure III. 35: Utilisation des sources d'information traditionnelles par les clusters.....	147

Liste des abréviations

AADL : Agence Nationale de l'Amélioration et du Développement du Logement.

ANSEJ : Agence Nationale de Soutien à l'Emploi des Jeunes.

ABS : Anti Blocking System.

BCS : Base Clients Simplifiée.

CAH : Classification Ascendante Hiérarchique.

CLV : Customer Lifetime Value.

DMS : Dealer Management System.

F1 : Formule 1.

INSEE : Institut National de la Statistique et des Etudes Economiques.

JV : Joint-Venture.

KDD : Knowledge Discovery in Databases.

LMT : Lead Management Tool.

ONS : Office National des Statistiques.

PR : Pièces de Rechange.

RSM : Renault Samsung Motors.

SOM : Self Organized Map.

SPA : Société Par Actions.

SIC : Système d'Information des Clients.

VUAC : Véhicules Utilitaires aux Adaptations Complémentaires.

Introduction générale

Dans un environnement économique marqué par une concurrence renforcée et exacerbée, le consommateur devient exigeant, mieux renseigné et de plus en plus regardant sur les différentes caractéristiques proposées par chaque produit. Dans une telle situation, il devient primordial de se renouveler, de se tenir au courant des tendances et surtout, de comprendre le client afin de s'adapter au mieux à ses multiples exigences et de répondre à ses aspirations, de plus en plus diversifiées. Ainsi, la constance de la relation client est parvenue à se placer au cœur des préoccupations des entreprises qui se sont inscrites dans une démarche dynamique de connaissance du client, condition indispensable pour faire face aux nouveaux enjeux du marché.

Dans ce contexte, la segmentation des clients en des groupes homogènes d'individus en fonction de leurs similarités, apparaît sans doute comme un concept clé du marketing en ce qu'il reconnaît que les clients sont différents et leurs besoins variés. Elle permet aux entreprises d'identifier les caractéristiques communes des différents segments et d'optimiser ainsi l'allocation des ressources relatives aux campagnes marketing, tout en personnalisant de façon adaptée la relation avec les clients. Plusieurs critères peuvent être utilisés pour la segmentation : géographiques, sociodémographiques, psychologiques ou comportementaux. Une segmentation peut également être multicritères.

Segmenter efficacement les clients requiert l'acquisition d'informations pertinentes, permettant de décrire et de caractériser les clients afin de pouvoir les regrouper dans des segments en fonction de leurs similarités. C'est donc dans le but de munir l'entreprise avec diverses données sur sa clientèle que les systèmes d'information intègrent les processus de gestion de la relation client, au sein des entreprises.

Toutefois, à l'air du digital, de nouveaux usages émergent et des capacités d'information, de partage et de consommation se transforment et se développent. Le concept des clients hyper-connectés envahit le domaine du marketing et leurs points de contact avec les entreprises se multiplient. Ces dernières, autrefois préoccupées par la collecte de données sur les clients, font face aujourd'hui à un nouveau challenge, celui d'exploiter le plus efficacement possible, les grandes quantités de données clients, stockées dans des systèmes d'information complexes.

A cet effet, un ensemble de techniques et de méthodes statistiques, mathématiques et informatiques, dites de Data Mining, intègrent le domaine du Marketing, permettant ainsi d'extraire des connaissances précises sur les clients, et ce, à partir de plusieurs bases de données volumineuses. Aujourd'hui, l'application des techniques de Data Mining joue un rôle crucial dans l'obtention d'une segmentation efficace de la clientèle et contribue de ce fait, au succès et à la pérennité de la relation de l'entreprise avec ses clients (Zeng et al., 2012).

Dans le secteur automobile mondial, déjà très concurrentiel et fortement exposé à divers facteurs d'influence, économiques et sociaux, l'exigence des clients a provoqué l'apparition de cycles de vie produit de plus en plus courts. En Algérie, les caractéristiques relatives à ce secteur rajoutent une difficulté supplémentaire aux entreprises implantées dans le pays. En effet, l'influence croissante des législations sur le marché automobile national a conduit à une véritable crise, obligeant ainsi les constructeurs automobiles à maîtriser impérativement l'ensemble des coûts, notamment marketing, à travers un ciblage optimal des clients. Renault Algérie, bien que leader dans le pays depuis plusieurs années, n'échappe pas moins à la règle et voit son chiffre d'affaire fortement diminué à cause de l'importante réduction des importations.

La problématique qui se pose donc est la suivante : Comment permettre à l'entreprise d'optimiser ses coûts marketing de façon à minimiser la perte du chiffre d'affaire causée par la nouvelle réglementation et conserver sa position de leader dans le marché ?

C'est dans le but d'apporter les réponses nécessaires à cette préoccupation de l'entreprise Renault Algérie, que le présent projet est implémenté, avec comme objectif, l'application des outils de Data Mining dans la segmentation efficace des clients, et est organisé comme suit :

Une revue de littérature sur le développement de la gestion de la relation client ainsi que sur les différentes techniques de Data Mining et leur application dans la segmentation de la clientèle est d'abord proposée dans le premier chapitre du rapport, afin de cerner les aspects théoriques, nécessaires pour le bon déroulement de notre projet.

Ensuite, nous présentons, dans la première partie du deuxième chapitre, l'entreprise Renault Algérie ainsi que l'environnement dans lequel elle évolue, à savoir, le marché automobile algérien, dans le but d'identifier les principaux objectifs d'une segmentation de la clientèle. La seconde partie du deuxième chapitre est consacrée au diagnostic et à l'exploitation de la base clients en vue de proposer un aperçu global sur la structure des clients de l'entreprise.

Afin d'affiner notre analyse de la typologie des clients de Renault Algérie, nous proposons, dans la première section du troisième chapitre, une segmentation approfondie de la clientèle, à partir des résultats d'une enquête que nous avons menée auprès des clients de l'entreprise. Par ailleurs, une seconde étude est réalisée auprès des acheteurs de véhicules de toutes marques confondues, et présentée dans la seconde partie du troisième chapitre, afin de segmenter le marché automobile algérien, identifier les segments cibles pour la stratégie marketing de l'entreprise et lui permettre ainsi de faire face aux nouveaux enjeux auxquels elle est confrontée. Enfin, nous restituons, dans la conclusion, les principaux résultats obtenus grâce à l'application du Data Mining dans la segmentation des clients.

CHAPITRE I :

**Gestion de la relation client et Application
du Data Mining dans la segmentation des clients.**

Introduction

Le présent chapitre consiste à définir le cadre théorique sur lequel se base notre travail. Il est organisé à cet effet en deux principales sections.

La première section décrit l'évolution de la relation de l'entreprise avec ses clients provoquant l'apparition et le développement du CRM, avant d'aborder, plus en détail, ses différents processus et approches.

Le but de la seconde section est d'introduire le concept de Data Mining ainsi que ses principales techniques, tout en se focalisant sur les méthodes utilisées dans le cadre de ce projet, pour ensuite présenter une revue de littérature sur les différentes études utilisant les techniques de Data Mining dans la segmentation des clients.

Section 1 : Gestion de la relation client

L'entretien d'une relation mutuellement bénéfique avec les clients est devenu l'une des préoccupations majeures des entreprises d'aujourd'hui. C'est d'ailleurs le principal objectif d'une gestion de la relation client.

1. Développement de la relation client

Sur un marché de plus en plus concurrentiel, l'effort essentiel pour toute entreprise concerne l'attraction et la rétention des clients, et ce, en leur créant plus de valeur.

En 50 ans, nous avons assisté à une véritable mutation du marketing. En effet, dans les années 70, la combinaison de la maîtrise des coûts et l'amélioration des processus métiers ont permis aux entreprises de proposer de larges gammes de produits adéquates aux segments ciblés. Par la suite, durant les années 80, le consommateur est devenu particulièrement sensible à la qualité, poussant les firmes à se concentrer davantage sur le développement des produits et des services.

Dès le début des années 90, profitant d'un bouleversement des canaux de communication, le marché connaît une inversion du paradigme marketing se traduisant par un passage d'une orientation « produit » à une orientation « client », et marque ainsi le début du recentrage sur le client.

L'expansion de l'utilisation de l'internet dans les années 2000 pousse le recentrage client encore plus loin, permettant même aux entreprises, convaincues des bénéfices engendrés de la satisfaction de leur clientèle, de proposer une offre personnalisée pour chaque client, incarnant la notion de personnalisation de mass (*Mass Customisation*). En effet, augmenter la satisfaction des clients en leur offrant une valeur supérieure et qui répond davantage à leurs attentes permet aux entreprises de réaliser de meilleurs bénéfices financiers (Kotler, 2014).

L'intérêt d'une relation bénéfique avec les clients n'est pas un nouveau concept dans le domaine des affaires. La concurrence accrue et l'infidélité croissante des clients ont contraint les entreprises à mettre en œuvre de nouveaux outils pour mieux se positionner dans le marché et de développer la fidélité des consommateurs en proposant des produits et des services plus personnalisés. Ce paradigme, développé sur le principe que les clients varient selon leurs besoins, préférences, comportements d'achat et sensibilité aux prix, constitue le socle du marketing relationnel.

D'ailleurs, la fidélité des clients apparait comme étant le premier objectif du marketing relationnel (Asunción et al, 2004). Le défi, pour les entreprises, serait alors de développer cette loyauté à travers l'établissement de relations fructueuses avec leurs clients (Newell et al, 2000).

De plus, l'aspect immatériel de la relation avec le client n'est pas facilement duplicable par la concurrence. En offrant une valeur différente, l'entreprise se démarque par un avantage concurrentiel unique. (Roberts et al. 2003).

Enfin, la croissance d'utilisation de l'internet et l'émergence des objets connectés, augmentent considérablement les opportunités marketing en transformant la manière dont les relations entre les entreprises et leurs clients sont gérées. (Ngai, 2005).

Le développement de la relation de l'entreprise avec sa clientèle ces dernières années, nous mène à considérer cette relation comme un paradigme d'organisation qui s'appuie sur la fidélisation (Ryals et Knox, 2001). La mise en œuvre d'un tel paradigme est caractérisée par la gestion de la relation client, littéralement connue sous le nom de Customer Relationship Management (CRM) (Ryals et Payne, 2001).

2. Définition et approches CRM

Le Customer Relationship Management (CRM) est fondé sur la philosophie de la création, le développement et l'amélioration de la relation d'une entreprise avec ses clients. Cette approche vise à améliorer et maximiser la valeur de la clientèle, et donc la rentabilité de l'entreprise (Frow and Payne, 2005).

Dans la pratique, le but est d'améliorer l'expérience relationnelle du client avec l'entreprise, ce qui conduit à la fidélité des acheteurs et l'augmentation des profits (Chou et al, 2002).

Bien que le CRM soit largement reconnu comme une importante approche stratégique de création de valeur, il n'existe pas actuellement de définition généralement acceptée et qui fasse l'unanimité dans la littérature. Parmi les définitions existantes, nous pouvons citer les suivantes.

Le management de la relation client (CRM) est un ensemble de processus supports qui permettent aux processus stratégiques de l'entreprise d'établir une longue, rentable et profitable relation avec des clients spécifiques (Ling et Yen, 2001). Les données clients et les systèmes d'informations forment ainsi le socle sur lequel se base la construction d'une stratégie CRM réussite.

Swift (2001) définit le CRM comme une approche de l'entreprise qui permet de comprendre et d'influencer le comportement des clients à travers une communication adéquate, afin de fidéliser et d'augmenter la rentabilité de chaque client.

De plus, selon Kotler et Armstrong (2014), le CRM concerne la gestion des informations détaillées sur les clients individuels, à chaque point de contact et d'interaction avec le client, afin de maximiser sa fidélité. Ils le définissent également comme un alignement de la stratégie, des processus et de la technologie pour la gestion des clients, et de tous les processus orientés client de l'entreprise et ses partenaires.

En outre, Parvatiyar et Sheth (2001) assimilent le CRM à une stratégie complète de processus d'acquisition, de maintien et de partenariat accompagnée d'une sélection du consommateur dans le but de créer plus de valeur pour l'entreprise et pour le client. Cela implique l'intégration de tous les processus de l'organisation : marketing, ventes, service client ainsi que la chaîne logistique afin d'atteindre une plus grande efficacité dans la valeur délivrée au client.

Greenberg (2002) présente, dans ses travaux, le CRM comme étant un ensemble complet de processus et de technologies pour la gestion des relations avec les clients potentiels et actuels et des partenaires commerciaux à travers le marketing, les ventes et le service quel que soit le canal de communication.

Ainsi, nous pouvons constater que certains auteurs limitent l'intervention du CRM à la relation de l'entreprise avec ses clients actuels, tandis que d'autres tendent à élargir son champ d'application à l'acquisition de nouveaux clients. Par ailleurs, ces définitions mettent l'accent sur l'importance de considérer le CRM comme un processus global, aligné sur la stratégie de l'organisation, d'acquisition et la fidélisation des clients à travers un mix marketing personnalisé se basant sur les informations récoltées à chaque point d'interaction avec le consommateur.

En nous appuyant sur les précédentes définitions et en prenant en considération le fait que la relation client de l'entreprise commence au début du cycle de vie du client, c'est-à-dire, à la phase prospect, nous définissons le CRM comme une stratégie globale optimisant l'utilisation des différentes ressources de l'entreprise, notamment technologiques, dans la poursuite de la compréhension et fidélisation des clients actuels et potentiels afin de maximiser la rentabilité de l'entreprise.

Au-delà de la simple définition du CRM, une analyse détaillée des différentes définitions et des descriptions du CRM proposées par des académiciens et les professionnels a été réalisée par Zablah et al. (2004). Il en résulte que le CRM a, implicitement ou explicitement, été considéré comme une approche tridimensionnelle : marketing, technologique et organisationnelle.

L'approche marketing concerne les traitements liés à l'identification du client par l'analyse des bases de données sur la clientèle. Ce qui permet de mettre en place des stratégies de segmentation marketing efficaces, d'élaborer une relation personnalisée grâce à la multiplication des points d'interactions et d'améliorer l'image de marque auprès des consommateurs. Le CRM permet ainsi aux entreprises, d'une part, d'investir dans les clients les plus rentables, en minimisant par la même occasion les investissements dans les clients considérés moins « précieux » (Verhoef et Donkers 2001). Et d'autre part, d'améliorer la réactivité de l'organisation et répondre au mieux aux besoins des consommateurs (Xu and Walton, 2005).

L'approche technologique comprend l'ensemble des outils mis en place pour collecter, traiter et restituer les informations relatives aux clients ainsi que les différents moyens de communication, à savoir, les sites web, réseaux sociaux, forums, centres de réclamation... En bref, cela concerne l'ensemble des systèmes d'informations qui soutiennent les processus métiers du marketing, comprenant plusieurs fonctionnalités dont le Data Warehouse, le Data Mining, la business intelligence et les outils de Reporting. Dans cette approche, le CRM est considéré comme la technologie utilisée pour fusionner les ventes, le marketing et le système d'information pour établir des partenariats rentables avec les clients.

L'approche organisationnelle nécessite une réflexion globale sur l'organisation de l'entreprise. Dans ce contexte, l'intégration d'une approche processus orientée client au sein de l'organisation est indispensable pour augmenter la satisfaction des clients et optimiser ainsi la

rentabilité de l'entreprise. Le CRM se préoccupe aussi de la création et l'exploitation des relations avec les entités extérieures à l'entreprise, notamment les canaux de distribution. (Srivastava et al.,1999).

La figure I.1 illustre l'approche tridimensionnelle précédemment détaillée.

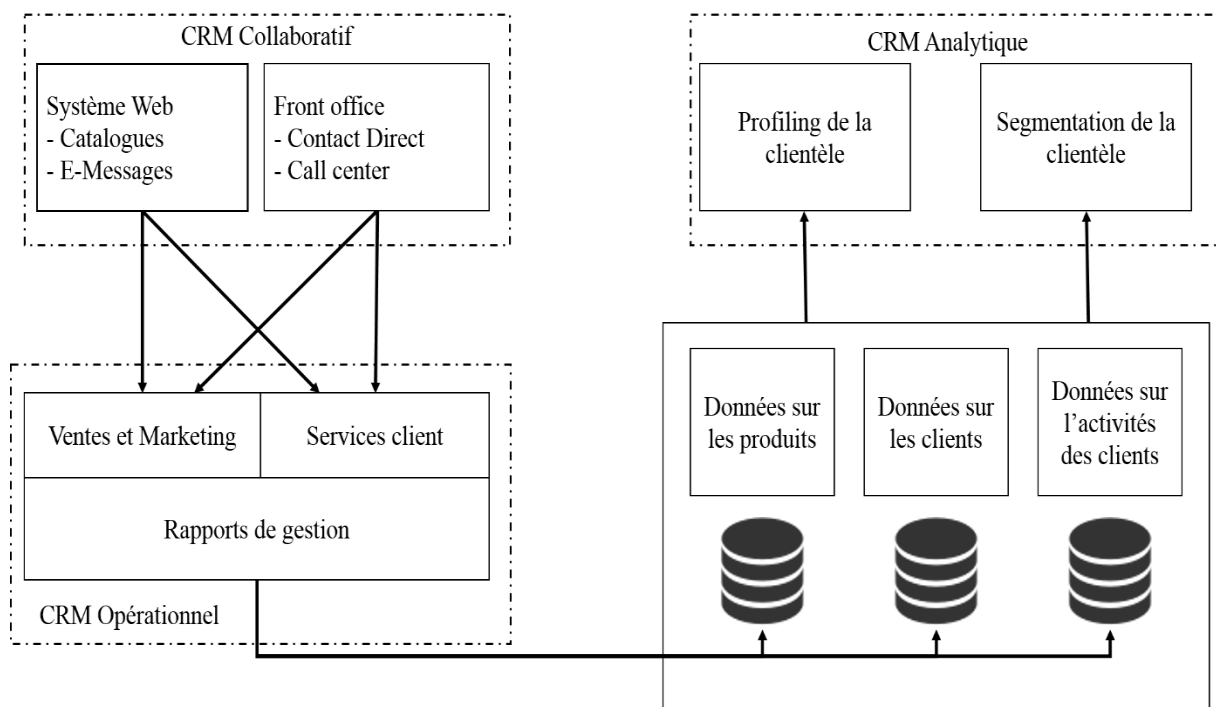


Figure I. 1 : Représentation de l'approche tridimensionnelle du CRM (Source : Zablah et al. (2004).

3. Processus CRM

La littérature présente principalement deux approches de définition des processus CRM.

La première approche, initiée par Greenberg (2004) se focalise sur l'intégration de l'approche centrée client dans l'entreprise. Elle classe les processus du CRM en trois macro processus : analytique, opérationnel et collaboratif.

Le CRM analytique gère les bases de données des processus opérationnels et analyse les informations récoltées au sein des DataWarehouse ou DataMart. Le but est de caractériser le comportement d'achat des clients et soutenir ainsi les processus décisionnels au niveau stratégique, en fournissant aux managers concernés des rapports et des analyses qui leur permettent d'allouer le plus efficacement possibles les ressources de l'entreprise.

Le CRM opérationnel appuie les processus décisionnels par l'intégration et l'automatisation des processus métiers horizontaux en interaction avec le client. Il constitue un segment important du CRM puisqu'il utilise les résultats d'activités relatives au CRM analytique et les transforme en activités à valeur ajoutée pour les clients. Le CRM opérationnel permet ainsi l'automatisation des campagnes marketing et de la force de vente.

Le CRM collaboratif repose essentiellement sur un échange efficace d'informations avec le client ainsi qu'avec les différentes fonctions de l'entreprise. Il représente ainsi la capacité du logiciel à regrouper les informations venant de différents canaux de communication.

La deuxième approche de définition des processus CRM s'appuie sur les différentes étapes de la relation client de l'entreprise.

Dans ce sens, Kracklauer et al. (2004) définissent quatre processus fondamentaux au CRM, à savoir l'identification, l'attraction, la rétention et le développement de la clientèle.

L'identification de la clientèle, première étape du CRM, différencie les principaux groupes de clients qui sont susceptibles d'être les plus rentables pour l'entreprise. Elle permet aussi d'identifier, d'analyser et de cibler le comportement, le besoin et le potentiel d'un segment spécifique.

Le principal élément de *l'attraction de la clientèle*, deuxième étape du processus, est le marketing direct. Ce processus dirige les efforts et les ressources de l'entreprise pour motiver les clients cibles à l'achat, et ce, à travers des canaux de communications adéquats.

La rétention de la clientèle est définie comme étant la mission principale du CRM. Elle veille à la satisfaction du client par la comparaison de ses attentes avec sa perception d'être satisfait. De nombreuses méthodes du marketing personnalisé sont utilisées dans ce sens, tels que les systèmes de recommandation qui se basent sur l'analyse des comportements d'achat, le traitement des requêtes et les programmes de fidélisation.

Le développement de la clientèle a pour objectif l'expansion de l'intensité de la valeur du client pour l'entreprise. De nombreuses approches sont mises en œuvre pour augmenter la profitabilité du client telles que l'étude du panier d'achat ou de la valeur à vie du client. D'autres approches, telles que le Up / Cross Selling, visent à augmenter la valeur de l'achat par des offres promotionnelles en montée de gamme ou l'association de services complémentaires qui sont étroitement liés aux besoins de l'acheteur.

La figure I.2 illustre les différents processus CRM proposés par Kracklauer et al. (2004).

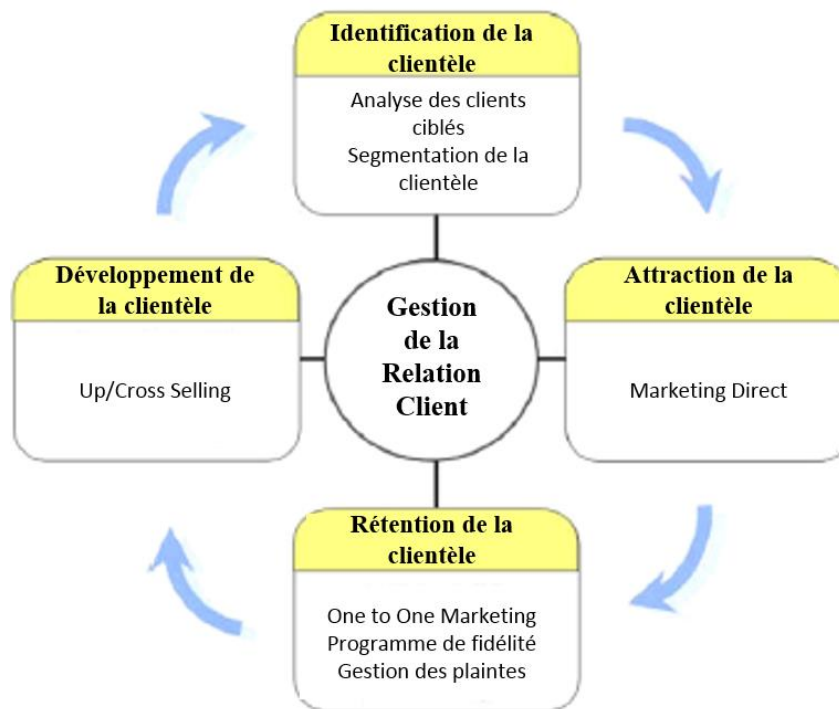


Figure I. 2: Les processus CRM (Source : Kracklauer et al. ,2004).

4. Implantation d'une stratégie CRM en entreprise

Les approches CRM ont émergé comme les perspectives stratégiques de haute priorité dans un bon nombre d'organisations (Coltman 2007). Les principaux objectifs visés à travers l'implantation d'une stratégie CRM dans les entreprises sont l'accroissement de la valeur à vie de la relation client et l'optimisation de l'efficacité des actions marketing et processus de ventes.

En effet, les entreprises emploient une approche CRM pour son potentiel de création de valeur commerciale puisqu'une approche CRM réussie est censée améliorer la fidélité des clients. De plus, le CRM améliore les performances des différentes fonctions de l'entreprise telles que le marketing, les ventes et le service après-vente, et ce, grâce à l'information fournie par le CRM et qui permet de voir toutes les interactions avec les clients et donc fournir une haute qualité de service.

Dans ce contexte, la littérature existante suggère que la mise en œuvre du CRM impacte un certain nombre de fonctions au sein d'une organisation, y compris les ventes, le marketing, les services après-vente et le système d'information. Campbell (2003) suggère que les ressources humaines, les processus et les technologies représentent l'élément clé à prendre en considération pour le succès d'une initiative CRM. De plus, l'économie est aujourd'hui axée sur la capitalisation de la connaissance. Les entreprises doivent mettre en œuvre des solutions CRM globales et intégrées, impliquant toutes les fonctions de l'organisation et travaillant en groupe pour partager l'information et fournir une seule vue sur le client (Yusuf et al., 2004).

Selon Bose (2002) les entreprises de services sont considérées comme les plus susceptibles d'une implémentation d'une démarche CRM, en raison du fait qu'ils disposent de processus déjà existant permettant de collecter et accumuler d'énormes quantités de données, sur chaque client. La gestion de la relation client est sans doute d'une importance vitale pour de nombreuses entreprises. Cependant, un aspect important présent dans la littérature et qui ne doit pas être négligé est la difficulté de l'implantation même d'une approche CRM, qui peut parfois se soldée par un échec. (Chen et Ching, 2004)

En plus de la difficulté d'implantation, une entreprise, même étant orientée consommateur, pourrait ne pas dégager le bénéfice nécessaire pour rentabiliser ses relations clients. Dans ce sens, plusieurs chercheurs, groupes de conseils et consultants ont publié des critiques importantes concernant l'adoption d'une approche CRM, mettant en avant le faible retour sur investissements du CRM.

A ce titre, le groupe METZ présente, en conclusion ses études, que 55 % des projets CRM ne produisent pas les résultats escomptés, tandis que Preslan et Keltz (2003) montrent que seulement 16 % des projets CRM ont réellement un retour sur investissements intéressant pour les entreprises. Le nombre considérable d'échecs d'adoption d'une approche CRM peut être expliqué par le fait que l'intégration du CRM au sein d'une entreprise doit nécessairement s'accompagner par un changement dans la façon qu'a cette dernière d'interagir et de communiquer avec ses clients.

Enfin, il est important de mentionner que, par sa définition, une démarche CRM possède une certaine dimension stratégique non négligeable. Chaque entreprise devrait être convaincue de son utilité et prête à mettre en œuvre tous les moyens possibles pour réussir son intégration.

Aujourd'hui, grâce aux nouvelles technologies informatiques, le CRM se développe à travers une plus grande collecte de données sur les clients. Toutefois, le traitement de ces données se complique avec au fur et à mesure de l'accroissement des quantités collectées. C'est ainsi que le Data Mining s'est introduit dans les processus CRM des entreprises.

Section 2 : Application du Data Mining dans la segmentation des clients

Avec la taille de plus en plus croissante des bases de données des entreprises, il devient difficile aujourd'hui d'en extraire des informations utiles pour la prise de décision. Un traitement préalable de ces données devient donc nécessaire. C'est dans ce contexte qu'est apparu le Data Mining.

1. Concept de Data Mining

Historiquement, le concept apparaît en 1989 sous le nom de *Knowledge Discovery in Databases* (KDD) avant que le terme « Data Mining » n'apparaisse pour la première fois en 1991.

Aujourd'hui, de nombreux auteurs associent le terme Data Mining à celui de KDD, d'autres s'accordent à dire que le Data Mining représente le processus principal du KDD.

Dans ce contexte, et de manière générale, le KDD est représenté par trois processus. Le premier consiste en la préparation des données avant l'application des techniques de Data Mining, c'est-à-dire, le nettoyage, l'intégration, la sélection et la transformation des données. Le deuxième processus est le processus clé du KDD, il s'agit du Data Mining. Il consiste en l'application de différents algorithmes dans le but d'extraire des informations utiles à partir des données. Enfin, le troisième processus évalue les résultats fournis précédemment, en fonction de considérations pratiques et des besoins des utilisateurs. (Han et Kamber. 2006).

L'objectif principal du Data Mining est la découverte d'informations intéressantes et non soupçonnées dans de grands volumes de données (Hand, 2001), et ce, en identifiant des modèles valides, nouveaux, potentiellement utiles et ultimement compréhensibles, à partir des données. (Fayyad et Piatetski-Shapiro, 2001).

Ainsi, nous pouvons définir le Data Mining comme étant un processus faisant intervenir différents outils et techniques d'analyse et de modélisation dans le but d'extraire des informations utiles à partir de larges quantités de données.

Le *Cross Industry Standard Process for Data Mining* (CRISP-DM) présente un modèle décrivant les différents processus de Data Mining.

Ce modèle, résumé dans la figure I.3, est très utilisé par les experts en Data Mining dans différentes industries. Il est constitué de 6 phases regroupées en un seul processus cyclique.

La première phase consiste en la compréhension du métier à travers l'identification des objectifs et de la situation actuelle de l'entreprise, l'établissement l'identification des objectifs du projet de Data Mining ainsi que le développement d'un plan pour le projet.

La deuxième phase permet de comprendre les données, et ce, par la collecte, la description, l'exploration et la vérification de la qualité des données.

La troisième phase du processus vise le nettoyage et la préparation des données selon les besoins du projet.

Cependant, dans le domaine du Marketing, des ventes et du CRM, une autre classification, reposant sur la nature des tâches et fonctions que peut réaliser le Data Mining, est proposée par Berry et Linoff (2004).

Ces fonctions sont présentées comme suit :

Le classement, aussi appelé « classification » par les auteurs anglo-saxons, est l'une des fonctions les plus répandues du Data Mining. Comme son nom l'indique, cette méthode consiste à classer des objets (ou des individus d'une population) dans des classes bien définies préalablement et dont le nombre est limité. Parmi les techniques de Data Mining utilisées dans le classement, on trouve les arbres de décision, la méthode des plus proches voisins et les réseaux de neurones.

L'objectif d'une méthode d'*estimation* est de trouver les valeurs de variables continues et inconnues telles que le revenu, la taille, etc. En pratique, les méthodes d'estimation sont souvent utilisées pour remplir les fonctions des méthodes de classement en créant une nouvelle variable (par exemple : un score) qui permettra de classer plus facilement les individus à l'aide d'une méthode de classement. Parmi les méthodes d'estimation, nous pouvons citer les modèles de régression et les réseaux de neurones.

Les méthodes de *prédiction* remplissent pratiquement les mêmes fonctions que les méthodes de classement ou d'estimation. La différence réside dans le fait qu'il existe une relation temporelle entre les variables d'entrée et la variable dite cible, à prédire. De ce fait, n'importe quelle technique de classement ou d'estimation peut être utilisée pour la prédiction tant qu'on dispose de données (ou d'échantillon) d'apprentissage dont les valeurs de la variable à prédire sont connues. Ces données historiques sont utilisées pour construire un modèle qui explique le comportement observé de la variable.

La fonction d'un *groupement par affinité* est de déterminer quels objets (ou individus) vont ensemble. La nature de cette tâche a rendu populaire l'utilisation des règles d'association, notamment dans le secteur de la grande distribution où elles sont surnommées : Analyse du panier de la ménagère, car elles permettent d'analyser les produits simultanément achetés par un client. Le groupement par affinité peut ainsi servir de base, en Marketing, pour des opérations de cross-selling¹ ou pour grouper des produits ou des services.

La *classification*, aussi appelée *Clustering* par les auteurs anglo-saxons, est la fonction qui consiste en la segmentation d'une population hétérogène en un nombre de classes ou de *clusters* plus homogènes. Contrairement aux méthodes de classement, la classification ne s'appuie pas sur des classes préalablement définies puisque les données seront regroupées selon leurs similarités. C'est donc à l'utilisateur de déterminer la signification de chaque classe. En Marketing, la classification est souvent utilisée préalablement à d'autres méthodes de Data Mining ou de modélisation. Par exemple, la classification peut être utilisée comme première étape d'une analyse de marché. Ainsi, au lieu de définir une stratégie de promotion à tout le marché, il convient de le diviser en classes afin d'adapter les stratégies marketing aux caractéristiques de chaque classe. Parmi les techniques de classification les plus utilisées, nous

¹ Une action de cross-selling consiste à « Proposer, au moment de la commande ou plus tard, la vente d'un produit complémentaire à celui acheté ». Source : www.definitions-marketing.com

pouvons citer la classification hiérarchique, le partitionnement ou encore un type de réseaux de neurones appelé *Self Organizing Map* (SOM).

Le *Profiling* est une fonction très pertinente dans certains domaines tels que la sociologie, l'économie, les sciences politiques ou encore, le Marketing. Elle consiste à décrire, pour mieux expliquer, le comportement des individus, des produits ou des processus. De nombreuses méthodes peuvent être utilisées dans ce sens. Par exemple, les arbres de décision peuvent être employés afin de réaliser des profils de clients en respectant un certain niveau de revenu. Par ailleurs, les règles d'associations et les techniques de classification constituent également de solides outils pour le Profiling.

Dans le cadre de notre projet de Data Mining, nous utilisons différentes techniques de classement et de classification en raison de leur adéquation avec notre objectif d'étude, à savoir, la segmentation des clients. De plus, dans le but de préparer les données avant l'étape d'analyse, nous ferons appel à l'analyse factorielle afin de réduire les dimensions organisant les relations entre les différentes variables. Il convient alors de la présenter brièvement avant d'introduire les différentes méthodes utilisées.

Cependant, étant donné qu'une confusion existe parfois autour des termes employés pour définir quelques techniques de Data Mining, nous jugeons utile de clarifier la terminologie utilisée dans le présent rapport.

Ainsi, nous utiliserons le terme « Classification » pour désigner le « Clustering », utilisé par les auteurs anglo-saxons, ou encore la « Segmentation », terme utilisé par certains auteurs francophones. Par ailleurs, nous emploierons les termes « Classement » et « Arbres de décision » au sujet des méthodes dites de « Classification » et « Decision Trees », respectivement, par les auteurs anglo-saxons.

2.1. Analyse factorielle

L'analyse factorielle est une technique statistique multivariée d'analyse des données, qui permet de représenter géométriquement et le plus fidèlement possible les données (individus d'une population ou des variables) dans un espace factoriel réduit ainsi que de détecter les éventuelles relations existantes entre les variables, ce qui rend possible la synthèse d'un grand nombre de variables pour pouvoir ensuite en tirer des informations pertinentes. C'est d'ailleurs l'une des caractéristiques les plus appréciées des méthodes d'analyse factorielle.

Concernant le modèle mathématique, dans l'analyse factorielle, chaque variable est représentée comme étant une combinaison linéaire des facteurs fondamentaux. Pour des variables standardisées, le modèle factoriel s'inscrit comme suit :

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \dots + A_{im}F_m + V_iU_i$$

Avec :

X_i : $i^{\text{ème}}$ variable centrée réduite.

A_{ij} : Coefficient de régression multiple centré réduit de la variable i sur le facteur commun j .

F : Facteur commun.

V_i : Coefficient de régression de la variable i sur la composante spécifique i

U_i : Composante spécifique pour la variable i (aussi appelée : terme d'erreur).

m : Nombre de facteurs communs.

Les facteurs communs, à leur tour, sont représentés par des combinaisons linéaires des variables :

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \dots + W_{im}X_m$$

Avec :

F_i : Estimation du $i^{\text{ème}}$ facteur.

W_{ij} : Score factoriel (aussi appelé, coefficient de pondération).

k : Nombre de variables.

Généralement, les coefficients de pondération sont choisis de sorte à ce que le premier facteur explique la plus grande partie de la variance totale, le deuxième facteur explique la deuxième plus grande partie de la variance totale, et ainsi de suite.

Avant d'utiliser une analyse factorielle, il convient d'analyser l'indice de Kaiser-Meyer-Olkin (KMO). Ce dernier indique s'il est possible d'expliquer les corrélations entre les paires de variables par d'autres variables incluses, ce qui justifierait l'utilisation d'une analyse factorielle. Afin d'apprécier cet indice, dans le domaine du marketing, l'échelle Kaiser propose l'échelle suivante :

0.90 ou + : Merveilleux.

0.80 à 0.90 : Méritoire

0.70 à 0.80 : Moyen.

0.60 à 0.70 : Médiocre.

0.50 : Misérable.

Sous 0.50 : L'analyse factorielle n'est pas recommandée.

Il est également fréquent d'avoir recours à une rotation orthogonale dite Varimax qui consiste à associer chacune des variables à un nombre réduits de facteurs en maximisant la variance intra-facteurs. Chaque facteur sera donc représenté par un nombre limité de variables, facilitant ainsi l'interprétation de l'analyse.

Différentes méthodes existent pour réaliser une analyse factorielle et chacune est adaptée à un type de données particulier. Parmi les plus utilisées, nous pouvons citer :

a) L'Analyse en Composantes Principales (ACP) :

L'Analyse en Composantes Principales (ACP) est utilisée lorsque chaque individu i de la population est décrit par p variables quantitatives notées $X^1, \dots, X^j, \dots, X^p$ (les données sont représentées dans un tableau individu \times variables quantitatives). Elle permet de transformer plusieurs variables corrélées entre elles en variables indépendantes (décorrélées entre elles) appelées composantes principales ou axes principaux, permettant ainsi de mieux synthétiser l'information.

Cette synthèse est réalisée projetant le nuage d'individus, initialement représenté dans un espace de dimension p , en dimension q (équivalent à q facteurs avec $q < p$), tout en respectant au maximum les distances entre les individus.

Pour représenter le nuage des individus dans un espace réduit et le plus fidèlement possible, les q facteurs choisis pour engendrer le sous espace de dimension q doivent maximiser la dispersion du nuage, qui se mesure par l'inertie de ce dernier (somme des variances des variables). Il s'agit donc de choisir les q facteur de façon à minimiser la perte d'inertie causée par la réduction de l'espace de représentation.

Pour choisir les axes factoriels, on commence par déterminer l'axe maximisant l'inertie projetée sur l'axe, ensuite, parmi les axes indépendants au premier (non corrélés à l'axe), on détermine, une fois de plus, l'axe qui maximise l'inertie projetée (qui sera forcément inférieure à celle du premier axe). Enfin, l'opération est itérée avec des inerties décroissantes.

Pour simplifier les calculs, les variables sont généralement centrées (en leur soustrayant leur moyenne), ce qui ne modifie en rien les résultats de l'ACP (qui devient une ACP centrée). Néanmoins, il est aussi fréquent de réduire les variables (en divisant chaque donnée par l'écart-type de la variable concernée), on parlera alors d'ACP centrée réduite, fortement recommandée pour des variables hétérogènes.

b) L'analyse factorielle des correspondances AFC :

L'AFC est une méthode qui permet d'étudier la relation entre deux variables qualitatives. Cette étude se fait sur la base d'un tableau de contingence (x_{ij}) , pour deux variables A et B de modalités $(a_k)_k$ et $(b_l)_l$, définit comme suit :

$$x_{ij} = \text{nombre d'individus } x, \text{ tels que } A(x) = a_i \text{ et } B(x) = b_j.$$

On définit également, en AFC, la notion de profils-lignes et de profils-colonnes. Ainsi, le $i^{\text{ème}}$ profil-ligne s'écrit : $\left\{ \frac{x_{i1}}{x_{i+}}, \dots, \frac{x_{ih}}{x_{i+}}, \dots, \frac{x_{ik}}{x_{i+}} \right\}$ avec k le nombre de modalités de la variable A et x_{i+} le total des effectifs de la modalité i de la variable B.

Le $h^{\text{ème}}$ profil-colonne s'écrit : $\left\{ \frac{x_{1h}}{x_{h+}}, \dots, \frac{x_{ih}}{x_{h+}}, \dots, \frac{x_{lh}}{x_{h+}} \right\}$ avec l , le nombre de modalités de la variable B et x_{h+} le total des effectifs de la modalité h de la variable A.

L'analyse se fait en réalisant une ACP au tableau des profils-lignes (les individus de l'ACP deviennent les modalités de la variable B) et une autre ACP au tableau des profils-colonnes.

En réalisant les deux représentations des ACP, nous montrons qu'elles se correspondent, ce qui permet de les superposer afin d'obtenir un nuage de points représentant à la fois les modalités de la variable A et celles de la variable B. L'interprétation de ce graphique permettra d'expliciter la relation entre les deux variables (ainsi que la relation entre les modalités des deux variables).

c) L'Analyse des Correspondances Multiples ACM :

L'ACM est une extension de l'AFC à p variables qualitatives ($p > 2$) avec une généralisation particulière du tableau de contingence. Cette généralisation se traduit par un tableau de contingence multiple appelé tableau de Burt.

Le tableau de Burt est un tableau symétrique qui se présente comme une matrice carrée $c \times c$ constituée de p^2 sous-matrices où chacune de ces sous-matrices diagonale est associée à une des p variables.

Chaque ligne de la matrice correspond donc à une modalité de l'une des variables (de même pour les colonnes). Au croisement de la $k^{ième}$ modalité x_{ik} de X_i et de la $l^{ième}$ modalité x_{jl} de X_j on trouve le nombre α d'individus x tels que $X_i(x) = x_{ik}$ et $X_j(x) = x_{jl}$. Il en découle qu'au croisement des deux mêmes modalités on retrouve les effectifs de cette modalité et au croisement de deux modalités différentes d'une même variable, on trouve des 0. Enfin, l'ACM consiste à réaliser une AFC au tableau de Burt obtenu.

2.2. Méthodes de classement

Les méthodes de classement font partie des techniques prédictives du Data Mining dont l'objectif est d'estimer la valeur d'une variable cible d'un individu, en fonction d'autres variables dites explicatives.

Ce qui distingue les méthodes de classement des méthodes d'estimation et de prédiction est le type de la variable à prédire : Qualitatives dans le cas du classement et continues dans le cas de l'estimation et de la prédiction.

Le classement consiste à placer les individus d'une population dans des classes précédemment définies. Cette affectation se fait en fonction d'un modèle construit à l'aide d'un algorithme ou d'un ensemble de règles.

Parmi les différentes techniques de classement existantes, nous utiliserons les arbres de décision, le classificateur bayésien ainsi que la régression logistique, pour leur efficacité et la facilité d'interprétation qu'ils proposent.

a) Les arbres de décision :

Les arbres de décision permettent d'expliquer une valeur à partir d'une série de variables discrètes ou continues, présentées dans une matrice X à m observations et n variables, associée à un vecteur Y à expliquer.

Ce sont des méthodes inductives qui présentent de nombreux avantages : elles sont non paramétriques, non linéaires et assez performantes.

Il existe deux types d'arbres de décision selon les valeurs de la variables Y à expliquer : Les arbres de régression, utilisées dans le cas où la variable Y présente des valeurs continues, et les arbres de classement, utilisées lorsque les valeurs de la variable Y sont qualitatives.

Les arbres de décision partitionnent les individus en produisant les groupes les plus homogènes possibles du point de vue de la variable à prédire, tout en tenant compte d'une hiérarchie de la capacité prédictive des variables considérées.

Cette hiérarchie permet de visualiser les résultats dans un arbre et de constituer des règles explicatives.

Les règles de définition des arbres de décision se basent nécessairement sur plusieurs itérations. À chacune d'elles, on divise les individus en k classes, afin d'expliquer la variable de sortie. (Tufféry, 2011),

La première division est obtenue en choisissant la variable explicative fournissant la meilleure séparation des individus. Elle définit de ce fait des sous-populations, représentées par les « nœuds » de l'arbre et à chaque nœud est associée une mesure de proportion qui permet d'expliquer l'appartenance à une classe ou la signification d'une variable de sortie.

L'opération est ensuite répétée pour chaque sous-population, jusqu'à ce qu'aucune séparation ne soit possible.

On obtient alors des nœuds terminaux, appelés « feuilles » de l'arbre. Chaque feuille est caractérisée par un chemin spécifique à travers l'arbre qu'on appelle « une règle » et l'ensemble des règles pour toutes feuilles constitue le modèle final.

Parmi les arbres de décision, traitant des variables continues, les plus répandues, nous pouvons citer l'algorithme CHAID. Il s'agit d'un algorithme connu pour son efficacité, qui utilise la mesure du χ_2 d'écart à l'indépendance et le t de Tschuprow.

Par ailleurs, les algorithmes CART et C4.5, se basent respectivement sur l'index de Gini (on parle aussi d'indice de concentration) et sur la notion d'entropie. Dans ce cas, l'ajustement de la taille de l'arbre est procédé par post-élagage, c'est-à-dire par l'arbre le plus pur avec toute la segmentation, puis un critère est choisi pour comparer des arbres de tailles différentes.

b) Classificateur bayésien :

L'algorithme de classification de Bayes est une méthode probabiliste basé sur l'application du théorème de Bayes avec l'hypothèse naïve, c'est-à-dire que les variables explicatives X_i sont supposées indépendantes conditionnellement à la variables cible Y .

Toutefois, il est nécessaire d'estimer en données d'entrée les probabilités conditionnelles de chaque classe ou variable $P(X = x_i|Y = c)$. Même si l'hypothèse est forte, ce modèle s'avéré très efficace sur des jeux de données volumineux.

L'algorithme de classification de bayes se présente comme suit :

Etant donné une classe de variable Y et des vecteurs de variables dépendantes de x_1 à x_n , selon le théorème de Bayes, on obtient la relation suivante :

$$P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n|y)}{P(x_1, \dots, x_n)}$$

En utilisant l'hypothèse naïve d'indépendance, on obtient :

$$P(x_i|y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i|y)$$

Donc pour chaque i , cette relation est simplifiée comme suit :

$$P(y|x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i|y)}{P(x_1, \dots, x_n)}$$

Comme $P(x_1, \dots, x_n)$ est une constante donnée par les variables d'entrée, on peut utiliser la règle de classification suivante :

$$P(y|x_1, \dots, x_n) \propto P(y) \prod_{i=1}^n P(x_i|y)$$

Nous pouvons ensuite utiliser l'estimateur de maximum à posteriori, une méthode dérivée du maximum de vraisemblance, pour estimer les paramètres d'une densité de probabilité reliés à un échantillon donné.

La méthode du maximum à posteriori consiste à trouver la valeur $\hat{\theta}_{MAP}$ qui maximise la grandeur $L(\theta)p(\theta)$ où $L(\theta)$ est la vraisemblance et $p(\theta)$ est la distribution à priori des paramètres θ .

En utilisant l'estimateur de maximum a posteriori (MAP) afin d'estimer $P(y)$ et $P(x_i|y)$, la formule obtenue de la fréquence relative de la classe y est la suivante :

$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^n P(x_i|y)$$

Les différents classificateurs bayésiens naïfs se différencient principalement par leurs hypothèses posées sur la distribution de $P(x_i | y)$.

En dépit de l'apparence trop simplifiée de leurs hypothèses, les classificateurs de Bayes naïfs ont fait leurs preuves dans de nombreuses applications réelles, comme dans le cas des filtres de messagerie (spam).

Et pour cause, les naïfs de Bayes peuvent être extrêmement rapides par rapport à d'autres méthodes sophistiquées, et ce, du fait, du découplage des distributions de la classe conditionnelle, qui signifie que chaque distribution peut être estimée de manière indépendante comme une distribution unidimensionnelle (Zhang, 2004).

c) Régression logistique :

Dans la pratique, il arrive souvent que les méthodes d'estimation soient utilisées dans un objectif de classement des individus.

L'une des méthodes d'estimation les plus utilisées pour le classement et les plus fiables est la régression logistique. Elle est également très appréciée pour sa généralité, son interprétabilité et sa robustesse, ce qui lui vaut de nombreux domaines d'application tels que le Marketing. (S. TUFFERY. 2012). C'est d'ailleurs pour cela que nous utilisons la régression logistique dans notre projet.

La régression logistique est une méthode prédictive du Data Mining qui sert à construire un modèle servant à prédire une variable cible qualitative à partir d'autres variables explicatives quantitatives ou qualitatives.

L'objectif de la régression logistique, comme celui de toute régression, est de modéliser l'espérance conditionnelle $E(Y/X=x)$. Dans le cas où la variable à prédire Y est continue (régression linéaire), il s'agit de déterminer la moyenne de Y selon les valeurs des variables explicatives. Le modèle de la régression linéaire s'applique dans le cas des variables qualitatives, le plus souvent binaires. Pour cela, la régression logistique consiste, non pas à modéliser la valeur de Y mais la probabilité qu'elle se réalise. C'est-à-dire : $E(Y/X=x) = P(Y=1 / X)$.

La facilité d'interprétation d'une régression logistique provient du fait que les coefficients estimés par le modèle soient mathématiquement liés à l'odds-ratio (OR), aussi appelé rapport des côtes, qui représente la force de l'association entre une variable explicative et la variable cible.

Le modèle mathématique le plus utilisé pour une régression logistique est le modèle Logit. Dans le cas d'une régression logistique binaire simple, c'est-à-dire avec une seule variable explicative, il se présente comme suit :

$$P(Y=1 / x) = f(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}, \text{ } f \text{ étant la fonction logistique.}$$

D'autres modèles, moins utilisés, existent pour la régression logistique. Il s'agit du Probit et du Log-log.

L'extension vers un modèle à plusieurs variables, se fait comme suit :

$$P(Y=1 / X) = f(X) = \frac{\exp(\alpha + \sum \beta_i x_i)}{1 + \exp(\alpha + \sum \beta_i x_i)}$$

A chaque variable explicative x_i on associe un coefficient β_i et OR_i , égal à $\exp(\beta_i)$.

Lors du choix des variables explicatives à inclure dans le modèle, il est fortement déconseillé d'utiliser deux variables fortement liées entre elle, on parle alors de colinéarité entre les variables. Ce phénomène a pour conséquence un modèle de régression de qualité médiocre.

Lorsque la variable cible est à k modalités ($k > 2$), on parle de modèle multinomial. Dans ce cas, il s'agit de modéliser $(k-1)$ rapports de probabilités, c'est-à-dire, considérer l'une des modalités comme référence et exprimer $(k-1)$ Logit par rapport à cette référence.

Le plus souvent, l'évaluation globale du modèle de régression logistique se fait à travers le pseudo- R^2 et la matrice de confusion.

2.3. Méthodes de classification

La classification est aussi appelée *Clustering* par les auteurs anglo-saxons, est une technique descriptive très répandue parmi les techniques d'analyse de données et de Data Mining.

Elle est utilisée lorsqu'on veut déterminer des sous-ensembles distincts et homogènes susceptibles de subir des traitements différenciés. De ce fait, elle est souvent employée en médecine, économie, agronomie ainsi que dans différentes sciences humaines telles que le Marketing où elle est particulièrement adaptée à la constitution de différents profils de client d'une entreprise. D'ailleurs, en Marketing, les termes « *segmentation* » ou « *typologie* » sont souvent utilisés pour désigner la classification.

La classification peut être définie comme étant une méthode statistique qui permet de regrouper des objets (des individus ou des variables) en un nombre limité de classes.

Ces classes étant exclusives et respectant l'homogénéité interne et l'hétérogénéité externe, c'est-à-dire que les objets appartenant à une même classe sont les plus similaires possibles tandis que deux objets appartenant à deux classes distinctes sont les plus différents possibles. Cette caractéristique de la classification est mesurée à l'aide des inerties intraclasse et interclasse.

L'inertie totale I d'une population s'inscrit :

$$I = \sum_{i \in I} p_i (x_i - \bar{x})^2 \text{ Avec } p_i \text{ le poids de l'objet } x_i \text{ et } \bar{x} \text{ la moyenne des } x_i.$$

L'inertie d'une classe mesure la concentration des points de cette classe autour de son centre de gravité, elle est donc définie de la même manière que pour l'analyse factorielle en remplaçant la moyenne des x_i par le barycentre de la classe. L'inertie de la classe j s'inscrit alors comme suit : $\sum_{i \in I_j} p_i (x_i - \bar{x}_j)^2$

L'inertie intraclasse est par définition la somme des inerties des k classes :

$$I_A = \sum_{j=1}^k I_j$$

Ceci étant, plus la classe est homogène, plus son inertie sera faible, ce qui implique que la segmentation sera meilleure si I_A est plus petite.

L'inertie interclasse I_R s'inscrit comme suit :

$$I_R = \sum_{j \in k} \left(\sum_{i \in I_j} p_i \right) (\bar{x}_j - \bar{x})^2$$

Cela signifie que plus I_R est grand, plus les classes seront éloignées les unes des autres (signe d'une bonne classification).

Nous verrons dans les sections suivantes que, pour des raisons de complexité, il est très difficile d'obtenir une classification optimale des individus. Cependant, l'objectif des précédentes techniques est d'obtenir une classification qui soit de « bonne qualité », c'est-à-dire qui respecte le principe de cohésion interne et externe.

Plusieurs indicateurs ont été définis afin de mesurer la qualité d'une classification. Parmi les indicateurs les plus fiables, nous pouvons citer l'indice Silhouette, dont l'expression mathématique se présente comme suit :

$$-1 \leq S(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \leq 1$$

Le but de l'indice Silhouette (Rousseeuw, 1987) est de vérifier si chaque individu a été bien classé. Pour cela, et pour chaque individu i de la partition, on calcule la valeur $a(i)$ qui représente la distance moyenne qui sépare l'individu des autres individus de la classe à laquelle il appartient et $b(i)$ représente la distance moyenne qui le sépare des individus appartenant à la classe la plus proche.

Quand $S(i)$, l'indice Silhouette, est proche de 1, la distance qui le sépare de la classe la plus proche est très supérieure à celle qui le sépare de sa classe, l'individu est donc bien classé. Par contre, si $S(i)$ est proche de -1, cela veut dire que l'individu est mal classé.

Si l'indice $S(i)$ est proche de 0 alors il pourrait également être classé dans la classe la plus proche.

Il existe principalement deux types d'approches dans la classification, qui sont parfois combinées pour améliorer le résultat de la segmentation, donnant ainsi une classification appelée mixte. Il s'agit de la classification hiérarchique, ascendante (dite agglomérative) ou descendante (dite divisive, elle est moins utilisée) et de la classification non hiérarchique, utilisant le principe de partitionnement.

Ces deux approches sont représentées dans la figure I.4.

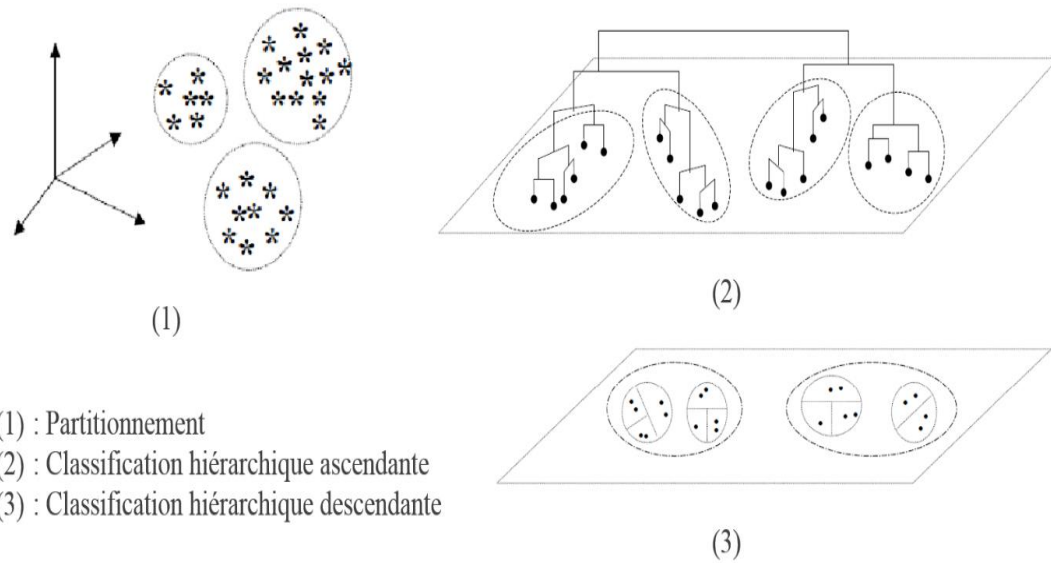


Figure I. 4: Approches de classification (Decaestecker et Saerens, 2007).

a) Classification hiérarchique :

La classification ascendante hiérarchique CAH étant plus souvent utilisée que la classification descendante hiérarchique, notamment en Marketing, nous choisissons de ne présenter que la première méthode dans notre rapport.

L'objectif de la CAH est de regrouper, par agglomération progressive, les objets les plus proches deux à deux en partant de N classes représentant les N objets isolés jusqu'à l'obtention d'une seule classe qui contient tous les éléments. Durant la classification, les classes sont fusionnées successivement en classes plus grandes selon leur degré de similarité. Enfin, les regroupements successifs effectués sont représentés sous forme d'un arbre binaire ou *dendrogramme*.

L'algorithme de la CAH peut se résumer dans la figure I.5.

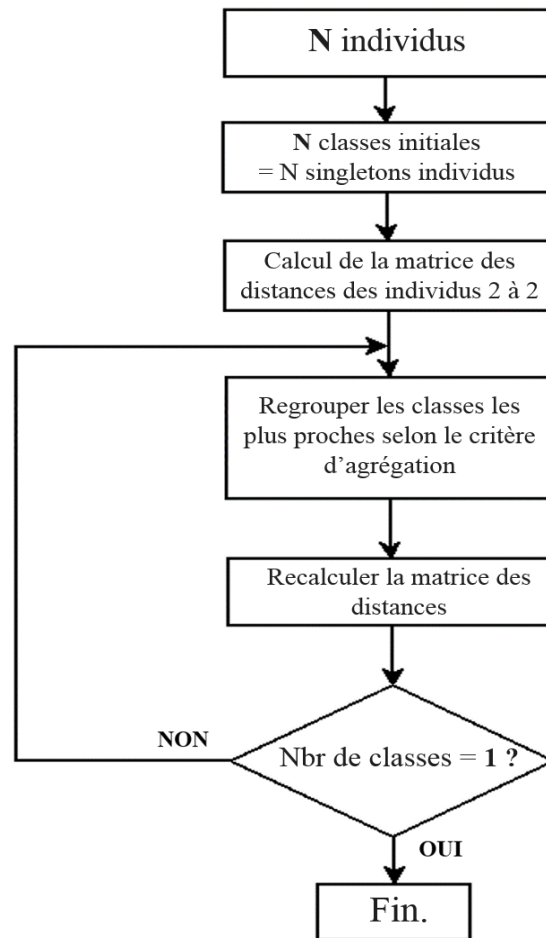


Figure I. 5: Déroulement de la classification ascendante hiérarchique.

Avant d'entamer la classification, il est nécessaire de définir une distance entre les individus (qu'on appelle aussi dissemblance). Pour le calcul de cette distance, on fait souvent appel à la distance euclidienne :

$$L_2 = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$

Mais la distance de City-block (ou de Manhattan) est également utilisée, notamment lorsqu'on cherche à réduire l'impact des individus hors norme :

$$L_1 = \sum_{i=1}^N |x_i - y_i|$$

Un autre critère doit lui aussi être défini avant la classification, il s'agit de l'indice d'agrégation puisqu'il détermine le regroupement des classes. L'indice d'agrégation dépend de la distance choisie pour représenter la distance entre les classes, et là encore, plusieurs méthodes existent pour le choix de la distance. Les méthodes les plus connues sont :

La méthode « du saut maximum » ou « *complete linkage* » consiste à produire des classes de diamètres égaux, elle est donc très sensible aux individus hors norme. Elle se calcule comme suit : $D_{max}(C_1, C_2) = \max\{d(x_i, x_j)\}, x_j \in C_i$ avec C_i la classe i .

La méthode « du saut minimum » ou « *single linkage* » consiste à regrouper des classes en considérant le minimum des distances entre toutes les paires d'individus entre deux classes. Elle n'est pas adaptée dans le cas où deux classes éloignées sont reliées par une chaîne d'individus car elles risquent de se retrouver groupées. Sa formule mathématique est la suivante : $D_{min}(C_1, C_2) = \min\{d(x_i, x_j)\}, x_j \in C_i$ avec C_i la classe i .

La distance dite de « *Ward* » tend à minimiser la baisse de l'inertie interclasse lors du passage d'une classification à $k+1$ classes à une classification à k classes. Etant donné qu'une bonne classification est une classification dont l'inertie interclasse est élevée, la distance de *Ward* correspond le mieux à l'objectif de la classification. Elle est cependant sensible aux individus hors norme et moins adaptée aux classes allongées. Son calcul mathématique se fait comme suit :

$D_{ward}(C_1, C_2) = \frac{n_1 * n_2}{n_1 + n_2} d(\mu_1, \mu_2)^2$; Avec n_1 et n_2, μ_1 et μ_2 les effectives et les barycentres de C_1 et C_2 respectivement.

La détermination de la partition des individus en classes se fait par la coupure de l'arbre obtenu avant une perte jugée trop importante de l'inertie. La figure I.6 illustre le choix de la partition par coupure de l'arbre.

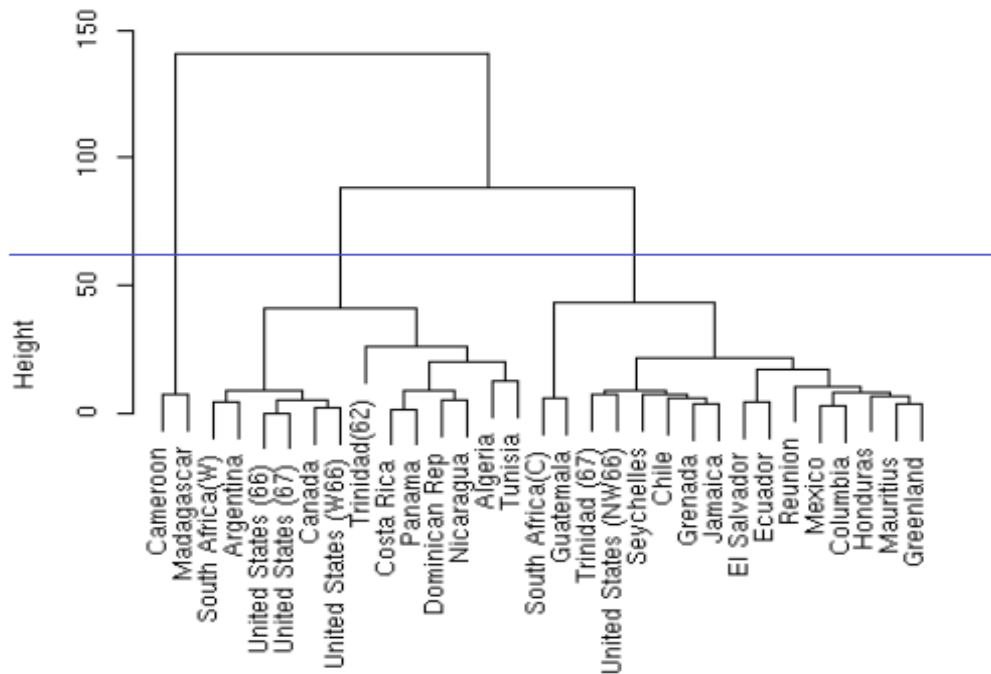


Figure I. 6: Coupure du dendrogramme (Source : www.aliquote.org).

Parmi les avantages de la classification hiérarchique nous pouvons citer l'absence de problème d'initialisation des classes et la possibilité de détecter les classes de formes variables.

En revanche, l'un des inconvénients les plus importants qu'elle présente est son critère de partitionnement local, c'est-à-dire qu'il n'y pas de réaffectation possible des objets durant le déroulement de l'algorithme.

b) Le partitionnement :

Les méthodes de partitionnement, contrairement aux méthodes hiérarchiques, produisent une partition de l'espace des objets plutôt qu'une structure hiérarchique organisationnelle. Pour des raisons évidentes de complexité, il ne s'agit pas de déterminer toutes les partitions possibles avant d'en choisir la meilleure, mais d'utiliser une approche heuristique offrant une « bonne partition », celle qui minimise l'inertie intraclasse, mais qui ne sera pas nécessairement la partition optimale.

L'une des méthodes de partitionnement les plus connues est celle des k-means dont nous présentons l'algorithme ci-dessous.

Entrée : k le nombre choisi de classes et d la mesure de dissimilarité sur l'ensemble X d'objets.

Sortie : Une partition $C = \{C_1, C_2, \dots, C_k\}$.

- 1) Choisir aléatoirement k individus comme étant les centres des k classes : $x_{10}^*, x_{20}^*, \dots, x_{k0}^*$
- 2) Constitution d'une partition initiale en k classes en affectant chaque objet $x_i \in X$ au centre le plus proche : $C_l = \left\{ x_i \in X \mid d(x_i, x_{l,0}^*) = \min_{h=1,2,\dots,k} d(x_i, x_{h,0}^*) \right\}$
- 3) Redéfinition des centres des classes.
- 4) Constitution d'une nouvelle partition $C_t = \{C_1, C_2, \dots, C_k\}$ en réaffectant chaque objet au centre le plus proche : $C_l = \left\{ x_i \in X \mid d(x_i, x_{l,t}^*) = \min_{h=1,2,\dots,k} d(x_i, x_{h,t}^*) \right\}$
- 5) Redéfinition des centres des classes obtenues : $x_{l,t+1}^*, \dots, x_{k,t+1}^*$
- 6) Répétition des étapes 3) et 4) jusqu'à ce que les centres des classes se stabilisent.

Les méthodes de partitionnement présentent de nombreux avantages, tels que leur complexité linéaire qui permet le traitement d'un grand volume de données ainsi que la possibilité de l'amélioration incrémentale des nouvelles données.

Cependant, leur inconvénient majeur concerne le fait que le nombre de classes soit déterminé à l'avance. De ce fait, la partition résultante dépend forcément du choix initial des centres des classes.

c) Classification mixte :

En plus des techniques de classification précédemment citées, il existe des méthodes de classification dites mixtes.

L'algorithme de ces méthodes, appelé Two-Step, se déroule en deux étapes : une classification initiale est réalisée par une méthode de partitionnement avec un nombre de classes généralement fixé entre 10 et 100, qui sera ensuite suivie d'une classification ascendante hiérarchique sur les centres des classes obtenues par partitionnement.

Ainsi, l'algorithme Two-Step permet de combiner le principal avantage d'une méthode hiérarchique, c'est-à-dire l'obtention d'un nombre optimal de classes, avec celui des méthodes de partitionnement, à savoir, la rapidité des calculs.

3. Logiciels de Data Mining

3.1. Choix du logiciel

Le développement et l'application d'algorithmes de Data Mining nécessite l'utilisation d'outils informatiques adaptés et performants. Dans ce sens, le nombre de logiciels ne cesse de croître et le choix de l'outil le plus approprié devient de plus en plus difficile.

Tout d'abord, les logiciels de modélisation en Data Mining les plus connus sont les logiciels commerciaux, parmi lesquels nous pouvons citer SPSS Modeler de l'éditeur IBM, Entreprise Miner de SAS et XLMiner. Ces logiciels fonctionnent généralement en architecture client/serveur, c'est-à-dire qu'ils sont en interface entre l'utilisateur et la base de données. Très complets en termes d'outils de traitement et d'analyse des données, ils proposent aussi de puissantes fonctions de visualisation des données pour la création d'un rendu visuel facilitant l'interprétation des résultats.

Ensuite, les logiciels Open source comme Tanagra, Orange ou Weka, contrairement aux logiciels commerciaux, sont de libre droit d'utilisation et permettent une certaine liberté de modification du programme. Néanmoins, l'établissement d'une architecture avec un serveur n'est possible qu'avec une modification du programme personnalisée, de plus ils ne permettent pas de traiter des bases de données volumineuses.

Enfin, parmi les logiciels spécialisés dans certaines fonctionnalités, nous pouvons citer Startminer pour les réseaux de neurones, Microsoft Analysis Service conçu par Microsoft pour les arbres de décision et Wizwhy pour les règles d'associations.

Dans notre cas, nous optons pour la solution SPSS Modeler 18 éditée par IBM. Cette application de Data Mining et d'analyse textuelle à interface visuelle permet à l'utilisateur de tirer profit des algorithmes statistiques et de Data Mining sans compétence de programmation particulière.

3.2. Connaissance du logiciel

Le logiciel IBM SPSS Modeler a été initialement nommé Clementine par ses créateurs, Integral Solution Limited. Ce nom a été utilisé jusqu'à l'achat de la solution SPSS par le groupe IBM en 2009. Le produit a ensuite été rebaptisé IBM SPSS Modeler, son nom actuel.

IBM SPSS Modeler est une solution qui permet d'analyser de nombreux types de données, peu importe où ces dernières sont stockées (par exemple dans un entrepôt de données, une base de données, un fichier plat) et indépendamment du fait que les données soit structurées (Base de données relationnelles) ou non structurées (Texte, image). De plus, avec son architecture client-serveur, SPSS Modeler peut analyser les données à la source pour une exécution en temps réel, ce qui permet de minimiser le mouvement des données et augmenter ainsi les performances. Il peut également fonctionner indépendamment ou en systèmes opérationnels sur plusieurs plateformes dans le système d'information existant, ce qui permet un déploiement très flexible et l'optimisation des investissements informatiques déjà existants.

Modeler résout une grande variété de problématiques grâce un large éventail d'analyses allant de l'analyse descriptive à des algorithmes beaucoup plus avancés tels que la modélisation automatique, l'analyse de texte, l'analyse des entités, analyse des réseaux sociaux, la prise de décision et l'optimisation.

SPSS Modeler est un logiciel très intuitif avec une interface graphique lui permettant d'être manié par une large catégorie d'utilisateurs : professionnels issus des domaines non techniques jusqu'aux statisticiens et experts des sciences de données. L'interface et le langage graphique que propose SPSS Modeler permettent un apprentissage rapide par les utilisateurs et commence facilement à découvrir les résultats.

4. Segmentation des clients

Le concept de segmentation de la clientèle a été introduit par l'expert en marketing, l'américain Wendell R. Smith, au milieu des années 1950.

Il la définit comme suit : « La segmentation de la clientèle se réfère à la classification des clients par leurs valeurs, exigences, préférences et d'autres facteurs dans le cadre de la stratégie de l'organisation, modèle d'affaires et du marché ciblé » (Smith, 1956).

Aujourd'hui, la segmentation du marché, qui représente en marketing un ensemble de clients, est encore très utilisée, notamment, lorsqu'on souhaite regrouper les différents types de clients selon des caractéristiques spécifiques (Albert, 2003). Elle consiste de ce fait en la création de sous-groupes significatifs d'individus ou d'objets (Haaïjer et al. 1998).

L'objectif de la segmentation est de réduire le nombre d'entités traitées en un nombre gérable de groupes qui sont mutuellement exclusifs et qui partagent des caractéristiques bien définies (Yankelovich et Meer, 2006).

Une segmentation efficace est une segmentation qui produit une information complète sur les clients ainsi que sur leurs exigences et besoins et donc sur la position concurrentielle de chaque entreprise dans le marché (Cooil et al. 2008).

Il arrive parfois que certains auteurs distinguent la segmentation du marché de celle des clients.

Dans ce sens, Cooil (2008) affirme qu'une segmentation du marché est différente de la segmentation des clients. La première examine l'existence des marchés intéressants pour les produits actuels (ou nouvellement développés) d'une entreprise tandis que la seconde concerne l'analyse structurelle des segments de clients existants, dans le but de de maintenir et d'améliorer le service proposé par l'entreprise.

Toutefois, l'importance d'analyser les clients actuels d'une entreprise, tout comme le positionnement concurrentiel dans le marché, fait partie intégrante du processus de développement de la stratégie d'une entreprise (Campbell, 2001).

Les deux segmentations consistent en la définition de groupes homogènes de clients, la principale différence relevée entre les segmentations concerne donc les objectifs visés.

Par ailleurs, deux approches méthodologiques peuvent être utilisées pour la segmentation des clients : la segmentation a priori et la segmentation a posteriori.

La segmentation a priori se base sur des caractéristiques connues. Les groupes sont choisis parmi une population et déclarés, à l'avance, comme « segments » (Cooil et al. 2008).

La segmentation a posteriori, en revanche, ne définit les segments qu'après la collecte de données. Ainsi, la segmentation est réalisée grâce à l'analyse des données collectées (Cooil et al. 2008). Cette approche permet aux analystes de se concentrer sur les segments et leurs caractéristiques pour développer des modèles prédictifs afin d'améliorer la performance du service au client (Cooil et al. 2008).

Par ailleurs, une segmentation des clients, réalisée avant ou après la collecte des données, utilise différentes méthodes d'analyse. Ces méthodes sont regroupées en deux principales catégories : l'analyse qualitative et l'analyse quantitative.

a) Segmentation des clients par l'analyse qualitative :

L'analyse qualitative de la clientèle est un type de segmentation facilement réalisable. Dans ce cas, l'appartenance des clients à des segments ne se base pas sur des chiffres, mais plutôt sur des facteurs qualitatifs tels que les facteurs démographiques, géographiques, psychographiques et comportementaux (Cooil et al. 2008).

Malgré sa simplicité, ce modèle de segmentation fournit des informations pratiques et rapides sur la composition de la clientèle.

La méthode descriptive et l'analyse du portefeuille client sont deux méthodes très utilisées dans le cadre de la segmentation des clients par l'analyse qualitative.

- Méthode descriptive :

La segmentation descriptive est unidimensionnelle, c'est-à-dire qu'elle est obtenue en appliquant une seule dimension d'attributs. Par conséquent, le nombre de segments est fonction du niveau de précision du critère choisi.

L'approche est souvent utilisée dans la segmentation des clients des produits d'une large gamme. Elle révèle les préférences des caractéristiques des produits et les intentions d'achat.

Bien que cette segmentation soit élémentaire, elle peut être considérée comme un support de prise de décision managérial, grâce à l'aperçu utile qu'elle fournit sur la structure de la clientèle et leurs caractéristiques.

- Analyse du portefeuille client :

Les modèles de portefeuille client, également appelés Compétitive Profile Matrix (CPM), sont un outil très connu dans la segmentation des clients.

Souvent, les paramètres pris en compte dans cette approche de segmentation sont purement qualitatifs, ce qui justifie le classement de cette méthode avec les approches qualitatives.

Cependant, ces paramètres peuvent également être constitués de combinaisons mixtes (qualitatifs et quantitatifs) ou purement quantitatifs.

Selon Zolkiewski et Turnbull (2002) ainsi que Terho et Halinen (2007), les modèles de portefeuille client ou matrices (CPM) permettent d'évaluer la répartition des ressources selon les groupes de clients identifiés par des dimensions clés, telles que la part de marché, la croissance du marché, l'attractivité du marché et la position concurrentielle.

La simplicité de la méthode est cependant vivement critiquée. Vue comme un modèle servant uniquement à la visualisation des données, elle ne permettrait pas une analyse prescriptive de ces données.

b) Segmentation des clients par l'analyse quantitative :

Contrairement aux modèles précédents, l'analyse quantitative repose sur les chiffres et les mesures qui sont extraites des entreprises au cours de leurs activités.

Encore une fois, à l'opposé des concepts de segmentation qualitative, plusieurs niveaux de complexité existent pour les méthodes quantitatives, et ce, en fonction du but de la segmentation et des données disponibles.

Dans le cadre de la segmentation quantitative, nous pouvons citer les méthodes suivantes :

- Analyse ABC :

L'analyse ABC est un outil de la segmentation aidant à la prise de décision qui peut être utilisé à plusieurs fins, et ce, en raison du fait qu'il se base sur des principes simples et non arithmétiques.

Cette méthode est réalisée en deux étapes. Dans la première étape, une segmentation est obtenue en classant successivement les clients selon de leurs volumes d'achat, ensuite, selon la valeur des produits achetés. Dans l'étape suivante, les deux analyses sont simplement combinées.

Il en résulte une matrice, à partir de laquelle les nouveaux et les plus fiables segments de clients peuvent être tirés.

- Analyse RFM :

Le modèle de segmentation RFM est l'un des modèles les plus utilisés dans la segmentation des clients.

Il distingue l'importance des clients par trois paramètres (au nombre des lettres qui constituent le nom du modèle) : l'intervalle de consommation, la fréquence d'achat et l'argent dépensé par le client.

Ainsi, le R du modèle RFM symbolise la « Récence », il désigne donc le moment du dernier achat enregistré du client. Plus l'intervalle est court, plus R est grand. F représente la « Fréquence » d'achat du client en une période déterminée. Et enfin, le M symbolise le terme « Monétaire ». Il désigne la valeur monétaire dépensée par le client durant une période donnée.

Les recherches montrent, d'une part, que plus le R et le F présentent de grandes valeurs, plus les clients sont susceptibles de réitérer l'achat. D'autre part, plus le M est grand, plus les clients sont susceptibles de réagir aux nouvelles offres de produits et services. (Wu et al. 2005).

Une fois que les clients possèdent un « score RFM », ils peuvent être regroupés en segments en fonction de leur potentiel et l'analyse de ce potentiel constitue une base pour la prise de décision CRM qui permet de fixer les fréquences de contact avec le client (Miglautsch, 2001).

L'objectif de la notation RFM est de prédire le futur comportement des clients et permettre ainsi une meilleure adaptation de la stratégie marketing aux différents segments.

- Matrice de la valeur client (Customer Value Matrix) :

La matrice de la valeur client est une méthode assez similaire à l'analyse ABC en termes d'applicabilité et de simplicité.

Pour obtenir les quadrants de la matrice, les valeurs moyennes des deux facteurs, la fréquence (nombre d'achat) et la valeur monétaire (montant de l'achat), sont calculées. En conséquent, chaque facteur est divisé en deux parties. De cette manière, quatre segments de tailles égales sont formés.

D'une manière générale, le principe de la matrice de la valeur client peut également être utilisé des analyses se basant sur d'autres facteurs tels que les facteurs géographiques, démographiques ou des facteurs liés à l'acquisition de nouveaux clients.

- Customer Lifetime Value (CLV)

Customer Lifetime Value (CLV) est une méthode qui permet de segmenter les clients en fonction de leur contribution au profit de l'entreprise (Kim et al, 2006).

Elle peut être définie comme étant le bénéfice ou la perte nette de l'entreprise d'un client pendant toute la durée des transactions de ce client avec la société. (Jain et Sing ,2002).

En d'autres termes, la valeur qu'un client apporte à l'entreprise est composée du revenu total apporté par ce dernier duquel on déduit les coûts des ventes et de promotion (coûts de rétention), tout en tenant compte du taux d'actualisation. De plus, pour atteindre une perspective à long terme, les valeurs annuelles sont additionnées.

Nous illustrons, à titre d'exemple, le modèle CLV proposé par Jain et Singh (2002) comme suit :

$$CLV = \sum_{i=1}^n \frac{(R_i - C_i)}{(1 + d)^{i-0.5}}$$

Où :

i : La période relative à une transaction client générant un flux de trésorerie.

n : Le nombre total de périodes pendant la durée de vie représentée du client.

R_i : Le revenu généré par le client dans la période (i).

C_i : Le total des coûts permettant la génération des revenus (R_i).

d : Le taux de remise annuel.

Cependant, le modèle CLV ne permet pas de capturer tous les différents aspects ayant une incidence sur la valeur du client durant toute sa durée de vie au sein de l'entreprise. Par exemple, les coûts d'acquisition (souvent importants) ne sont pas inclus dans le calcul. Il offre néanmoins une segmentation efficace.

Une autre typologie de la segmentation, réalisée en fonction des objectifs qui ont conduit à son utilisation, est proposée par Chorianopoulos et Tsiptsis (2009) comme suit :

Segmentation basée sur la valeur client : Dans ce cas, les consommateurs sont regroupés selon la valeurs qu'ils apportent à l'entreprise. Ce type de segmentation est considéré comme le plus important, car il permet d'identifier les clients les plus précieux et de suivre l'évolution de leur valeur potentielle au fil du temps. Elle est également utilisée pour diversifier les stratégies d'offre et d'optimiser l'allocation des ressources marketing en vue de l'acquisition de nouveaux clients.

Segmentation basée sur le comportement d'achat : Cette segmentation, appréciée pour son efficacité et son utilité, est largement utilisée du fait de la disponibilité des données. En effet, les données requises comprennent les informations sur les produits et leur utilisation. Les clients sont ainsi répartis à partir de leurs identifiants en fonction de leurs habitudes d'achat. La segmentation comportementale est généralement utilisée pour le développement de nouveaux produits et la conception de programmes de fidélisation.

Segmentation basée sur la propension : Cette segmentation permet de grouper les consommateurs en fonction d'un score attribué à chacun d'entre eux. Ce score de propension représente le taux de l'attitude, la tendance ou bien le penchant qu'a le client pour une certaine offre particulière. Plusieurs exemples de score de propension sont utilisés. Nous pouvons en citer le score de désabonnement, le score de prise d'un crédit ou bien le score d'un intérêt pour des produits ou services supplémentaires (Cross-selling). Les scores de propension peuvent être combinés avec d'autres systèmes de segmentation afin de mieux cibler les campagnes marketing. Par exemple, le schéma de la segmentation des valeurs à risques est développé en combinant les propensions de la valeur du segment avec le risque d'achat.

Segmentation basée sur la fidélité : Elle permet d'identifier les différents groupes de clients en fonction de leur statut de fidélité et de séparer les clients loyaux des consommateurs occasionnels. Les segments sont créés par l'application de simples règles ou de ratio d'expert du métier, ou encore, à partir des modèles de clusters issus d'enquêtes ou d'informations sur les clients. La segmentation basée sur la fidélité peut être associée à des comportements d'achats spécifiques et des attributs de la base de données clients. Pour cela, l'entreprise peut commencer par une étude de marché pour révéler les segments de fidélité des clients avant de construire un modèle de classification des clients avec comme champ cible : le segment de fidélité du client. De cette façon, elle sera en mesure d'identifier les comportements associés à chaque segment de fidélité et de les utiliser pour extrapoler les résultats de la segmentation de fidélité à l'ensemble de la clientèle.

Segmentation démographique : Cette segmentation, regroupant les clients en fonction de leurs caractéristiques démographiques, est largement utilisée car considérée comme ayant une forte influence sur les besoins, préférences, et comportements de consommation des clients. Les segments sont généralement créés avec des règles simples sur les attributs de la base de données clients tels que le genre, l'âge, le statut social, l'éducation, l'état matrimonial, la propriété, le revenu familial, le nombre d'enfants, l'âge des enfants, et la profession. Il existe plusieurs variantes de la segmentation démographique, comme la segmentation du cycle de vie familial, principalement déterminée par l'âge du client, le statut matrimonial le nombre et l'âge des enfants. Nous pouvons retrouver des segments du cycle de vie familiale tel que : Jeunes / célibataires, Jeunes familles / pas d'enfants, Jeunes familles / moins de 5 enfants, Personnes âgées à la retraite. La segmentation démographique est principalement utilisée pour améliorer les regroupements de client révélés par les segmentations basées sur le comportement, la valeur ou la propension.

Segmentation basée sur les besoins : Cette dernière est utilisée pour étudier les besoins, perceptions et préférences des clients. Les informations pertinentes ne peuvent être recueillies que par le biais d'études de marché lorsque les segments des clients sont clairement identifiés. La segmentation basée sur les besoins est principalement combinée avec des clusters issus des segmentations basées sur le comportement et sur la valeur du client pour enrichir le profil des consommateurs à travers des données qualitatives. Ce type de segmentation permet de

concevoir et développer de nouveaux produits ou services, et d'adapter la communication pour chaque segment avec un message sur mesure pour développer l'image de marque souhaitée.

Dans le tableau 1 suivant, Chorianopoulos et Tsipsis (2009) résument les principaux besoins des entreprises en termes de stratégies marketing en mettant en avant le type de segmentation adéquat pour chaque besoin ainsi que les outils et techniques analytiques appropriés.

Tableau I. 1: Types de segmentations des clients en fonction des besoins marketing (Source : Chorianopoulos et Tsipsis, 2009, p192).

Besoin de l'entreprise	Critères de segmentation appropriés	Outils et techniques analytiques
Conception et développement de nouveaux produits.	Comportements et Besoins.	Combinaison entre Data Mining et étude de marché/ Analyse factorielle et de classement.
Conception de stratégies d'offre de produits personnalisés.	Comportements.	Data Mining / Analyse factorielle et de classement.
Image de marque et stratégie de communication.	Besoins.	Étude de marché/ Analyse factorielle et de classement.
Différenciation des clients.	La valeur du client combinée avec d'autres attributs tels que la durée du cycle de vie du client.	Analyse des données / Business intelligence.
Allocation des ressources et priorisation des campagnes marketing visant à développer et retenir la clientèle.	La valeur du client avec une étude approfondie sur les motivations d'achat du client.	Étude de la valeur du marché et identification des décideurs d'achat.
Identification de groupes cibles pour campagnes marketing.	Les scores de propension dérivés de modèles de classification pertinents.	Data Mining en utilisant les modèles de classification en fonction de leurs scores de propension et de leur probabilité d'achat.

5. Utilisation du Data Mining dans la segmentation des clients

La segmentation est une simplification de la complexité désordonnée qui permet de faire face à de nombreux clients individuels, chacun ayant des besoins distincts et une valeur potentielle (Wilson et al. 2002). Les méthodes de segmentation traditionnelles sont basées sur des classifications expérientielles, une visualisation des données ou des méthodes de statistique simples.

Avec le développement des TIC et du CRM, de plus en plus de données sont collectées sur les clients, nécessitant des analyses beaucoup plus complexes, auxquelles les méthodes de segmentation traditionnelles ne peuvent faire face. Le Data Mining apparait alors comme une solution, aux spécialistes du Marketing, leur permettant d'extraire de précieuses informations stratégiques à partir d'une grande quantité de données extraites des systèmes transactionnels de l'entreprise. Dans ce contexte, Naim et Bottomely (2003), affirment que le Data Mining permet à une société de mesurer le comportement de ses clients à partir de bases de données de plus de 100 attributs, contre de trois ou quatre attributs pour les modèles de statistiques traditionnels.

Dans ce qui suit, nous détaillons les approches utilisées par différents auteurs dans la segmentation des clients des entreprises en général, et celles du secteur automobile en particulier.

5.1. Segmentation des clients des entreprises de différents secteurs d'activités :

La littérature indique que différents modèles et algorithmes sont utilisés pour la segmentation des clients d'une entreprise. Dans ce sens, Chan (2006) a réalisé une classification des différentes techniques de Data Mining utilisées dans le CRM des entreprises, à partir d'une revue de littérature. Cette classification est illustrée dans la figure I.7 suivante :

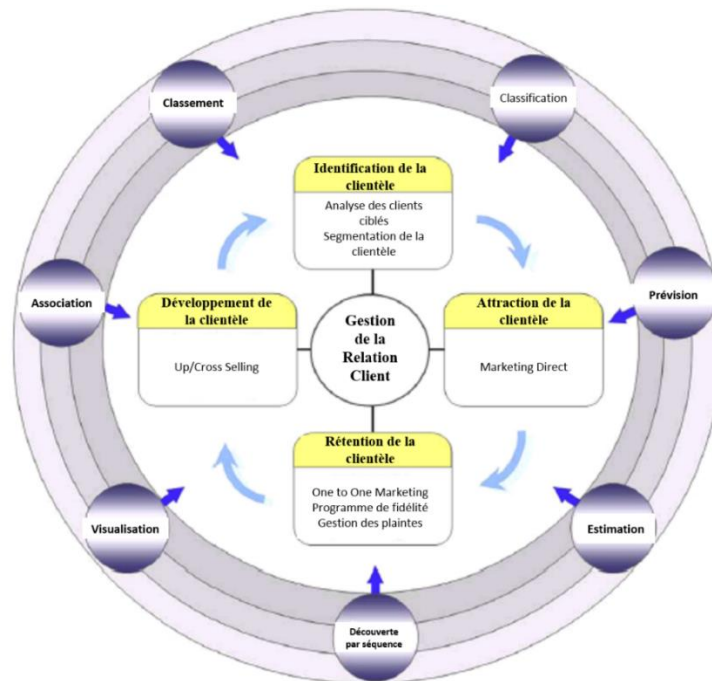


Figure I. 7: Classification des méthodes de Data Mining dans le CRM (Chan, 2006).

Cette classification présente, d'une part, les principales dimensions du CRM : L'identification, l'attraction, la rétention et le développement de la clientèle. Et d'autre part, une des approches utilisées pour la classification des méthodes de Data Mining et qui les regroupe en sept principales méthodes, à savoir : L'association, le classement, la classification, la prévision, l'estimation, la découverte par séquence et la visualisation.

Le modèle de classement est le plus couramment utilisé dans les analyses CRM pour prédire les comportements de futurs clients (Chan, 2008). Quant aux réseaux de neurones et les méthodes de Clustering, elles sont très souvent appliquées dans la classification et le regroupement des clients ainsi que dans la prédiction.

Les arbres de décision et les règles d'association sont aussi populaires que les réseaux de neurones, et pour cause, la logique de ces deux méthodes est plus facilement utilisée par les personnes du métier que les réseaux de neurones.

Chan (2008) classe les méthodes de segmentation de la clientèle existants, selon leur orientation, en deux principales catégories : les techniques de segmentation orientées méthodologie et les techniques orientées application.

La plupart des études de l'orientation méthodologie poussent à la modification de certaines techniques de classification de données, telles que les SOM, ou leur combinaison avec d'autres méthodes de Data Mining pour obtenir des segments et des sous-segments.

Parmi les auteurs utilisant les méthodes de segmentation orientée méthodologie, nous pouvons citer Lee et al. (2004) qui ont développé une méthode de segmentation du marché international en deux phases. La première utilise l'inférence statistique pour tester la différence entre les facteurs de regroupement. La seconde phase est réalisée par la méthode de classification SOM à deux niveaux afin de développer les sous-segments.

Par ailleurs, Hung et al. (2007) utilisent la nouvelle approche du Learning Machine en automatique, le « Support Vector Clustering » pour une segmentation marketing.

Enfin, dans le marché des achats en ligne, Kim et Ahn (2008), proposent une autre approche de segmentation basée sur les algorithmes génétiques.

« Les recherches orientées application doivent trouver la meilleure méthode pour résoudre les problèmes de segmentation dans des applications spécifiques » (Chan, 2008). A cet effet, ils définissent généralement une nouvelle variable pour la procédure de classification ou utilisent plusieurs variables dans des étapes séquentielles de classification.

Dans la littérature relative à l'approche orientée application, le CLV est très souvent présent et de nombreux développements y sont apportés.

Par exemple, Hwang et al. (2004) suggèrent un nouveau modèle CLV dans le secteur de la télécommunication, qui prend en compte les bénéfices actuels générés par le client, son potentiel ainsi que la valeur probable de ce dernier. Kim et al. (2006), en revanche, proposent un modèle probabiliste général pour l'analyse de la valeur client. Ce modèle, illustré par un cas d'étude dans une entreprise de télécommunication, a permis de détecter deux fois plus de

segments que les méthodes utilisées précédemment, et ce, grâce à la combinaison des méthodes de classification et de classement.

Une autre variable d'entrée, très utilisée pour la segmentation de la clientèle orientée application, est le score RFM (précédemment expliqué).

Dans ce contexte, Hsieh (2004) utilise une carte SOM pour identifier les groupes de clients en fonctions de leurs comportements quant au remboursement d'achat ainsi que la récence, fréquence, et valeur monétaire des achats dans le but de prédire un score comportemental. Ce score lui a ensuite permis de classer les clients d'une banque en trois principaux groupes de clients. Enfin, les groupes résultants ont été classés par rapport aux attributs des clients, et ce, en utilisant les règles d'association.

Cheng and Chen (2009) ont fait appel, à leur tour, à une nouvelle procédure de segmentation en combinant les valeurs quantitatives des attributs RFM avec K-means dans la théorie des ensembles rugueux afin d'extraire des règles d'association non supervisées.

Dernièrement, Namvar et al (2011), ont proposé une autre approche pour optimiser la segmentation de la clientèle. Elle consiste en l'utilisation de classifications SOM en deux étapes en considérant les données démographiques ainsi que les attributs RFM pour déterminer les premiers segments. Ensuite, un modèle CLV proposé par Hwang (2004) est utilisé pour réaliser une sous-segmentation des premiers segments obtenus.

En outre, dans l'approche proposée par Chan (2008), une première segmentation, basée sur le Scoring des attributs RFM par les algorithmes génétiques, est utilisée. Cette segmentation est ensuite complétée par une classification CLV des clients.

D'autres variables peuvent également être utilisées. Par exemple, Lee et Park (2005) ont utilisé les données d'une enquête de satisfaction du consommateur. Pour cela, ils proposent un système multi-agent pour la réalisation d'une première segmentation SOM basée sur des données sociodémographiques suivie d'un classement grâce à un algorithme dérivé des règles d'association.

Chang et al. (2007) utilisent une combinaison d'une analyse K-means à des règles d'association afin de proposer un modèle d'anticipation du comportement d'achat d'un potentiel client. Ce modèle est déduit du comportement d'achat des clients fidèles.

Stone et al. (2007) proposent une nouvelle approche de segmentation de la clientèle en deux étapes : La première utilise les K-means arithmétiques afin de regrouper les clients en différents segments en fonction de caractères dits de « survie similaire ». La deuxième se base sur une analyse de survie pour prédire la fonction de survie de chaque segment avant de tester la validité du groupement et définir ainsi le comportement des clients.

Un autre modèle d'identification du comportement d'achat des consommateurs de lait, utilisant la classification SOM, est développé par Rajeshri et al. (2013).

Dans le secteur de l'habillement, Brito et al. (2014) mettent en place un modèle prédictif du comportement d'achat pour une entreprise de personnalisation de vêtements. Ce modèle se

base sur une segmentation de leur base de données CRM par une classification SOM suivie par une caractérisation des regroupements par la méthode des règles d'association.

La même année, Afjeha et Darvishib utilisent un questionnaire en ligne pour proposer une segmentation du marché de l'achat par internet. Ils utilisent, à cet effet, une segmentation du comportement d'achat basée sur une régression logistique des données psychographiques² suivie d'une sous-segmentation des attributs personnels, par une classification SOM.

Enfin, Haiying (2015), suggère un modèle de segmentation de la clientèle pour les sites de e-commerce. Ce modèle utilise deux classifications successives. La première, non supervisée, utilise les règles d'associations pour segmenter les clients potentiels. La deuxième, supervisée, permet de mieux cibler les clients en faisant appel aux arbres de décision.

5.2. Segmentation des clients dans le secteur automobile :

Entre des pressions permanentes pour l'optimisation des coûts, une concurrence acharnée et des clients de plus en plus informés et exigeants, l'industrie automobile continue de faire face à de nombreux challenges ces dernières décennies. En parallèle, l'évolution du Big data et des technologies analytiques offre aux spécialistes du secteur des possibilités insoupçonnées pour la résolution de ces problèmes et la prise de décisions.

Du domaine de la production aux études de satisfaction, le Data Mining s'incruste ainsi de plus en plus dans les différents métiers du secteur automobile.

En marketing, le Data Mining est utilisé, entre autres, pour la segmentation des clients en vue de l'optimisation des campagnes marketing, le développement de nouveaux produits et services, ou encore, afin d'identifier le comportement des nouveaux acheteurs.

Dans la littérature, deux études de cas, portant sur l'utilisation du Data Mining dans la segmentation des clients dans l'industrie automobile, sont présentées.

La première étude de cas, réalisée par Chu-Chai Henry Chan (2008), est orientée « application ». L'étude y propose une nouvelle approche de segmentation combinée à un ciblage de la clientèle dans le but de construire des stratégies efficaces de campagnes marketing. Pour ce faire, l'auteur utilise une base de données de 4000 clients du constructeur japonais Nissan. L'approche consiste à identifier le comportement du consommateur en utilisant les données RFM avant d'évaluer les segments obtenus à l'aide du modèle CLV. Une fois la segmentation réalisée, l'auteur propose l'utilisation des algorithmes génétiques afin de sélectionner les clients appropriés pour chaque campagne. Les résultats de l'étude démontrent que cette méthode permet de cibler plus efficacement les clients qu'une sélection aléatoire.

La seconde étude, plus récente, est plutôt orientée « méthodologie ». Chih-Fong Tsai, Ya-Han Hu et Yu-Hsin Lu (2013) y étudient le problème de segmentation des clients d'un concessionnaire à Taiwan.

² Un critère psychographique est un critère de segmentation ou de ciblage basé sur les styles de vies, les croyances, les valeurs, les personnalités des consommateurs. Source : www.definitions-marketing.com

La segmentation prend en considération les données relatives au comportement des acheteurs mais aussi à leur satisfaction. Les auteurs mettent en avant le fait que la plupart des études précédentes utilisaient une seule méthode de classification des clients, ils décident alors d'utiliser deux algorithmes de segmentation afin d'augmenter la fiabilité de leur étude.

La première segmentation utilise l'algorithme K-means tandis que la deuxième se base sur le modèle de mélange gaussien, aussi appelé l'algorithme EM (pour Expectation Maximization).

Après une comparaison des résultats obtenus à l'aide des deux méthodes, les résultats indiquent que 4 segments de clients (loyaux, potentiels, importants et désabonnés) peuvent être définis et donc traités séparément.

Conclusion

Ce chapitre a présenté une base théorique pour le déroulement de notre projet, en abordant les différentes approches de la gestion de la relation client ainsi que les divers outils statistiques, mathématiques et logiciels, couramment utilisés dans le cadre d'une analyse Data Mining.

Il a également mis en exergue l'important rôle qu'occupe aujourd'hui le Data Mining dans l'optimisation des relations de l'entreprise avec ses clients, et ce, en intégrant les différents processus du CRM et en servant différents objectifs marketing.

Enfin, une revue bibliographique a été présentée afin d'illustrer les différentes applications du Data Mining dans la segmentation des clients de façon générale ainsi que son utilisation pour la segmentation des clients dans le secteur automobile, en particulier.

Une fois les concepts théoriques assimilés, nous proposons, dans le chapitre suivant, une présentation de la structure d'accueil ainsi qu'un diagnostic des données existantes en entreprise.

CHAPITRE II :

**Présentation de l'entreprise
et Exploitation de la base de données clients.**

Introduction

Avant d'aborder les détails de l'étude réalisée dans le cadre de notre projet de fin d'études, il est nécessaire d'introduire l'entreprise ainsi que le contexte général dans lequel s'est déroulé notre travail.

A cet effet, nous présentons, dans la première section de ce chapitre, une description de l'entreprise Renault Algérie, filiale du Groupe Renault, ainsi qu'une analyse du marché automobile algérien dans lequel évolue l'entreprise avant d'aborder l'identification de la problématique suivie d'un diagnostic de la base de données clients, nécessaire à une définition plus précise de la problématique.

Enfin, une exploitation de la base de données clients en vue de la réalisation d'une première segmentation est réalisée et présentée dans la seconde section du présent chapitre.

Section I : Présentation de l'entreprise

L'objectif de la première section du chapitre est d'introduire le contexte général dans lequel se déroule notre projet, et ce, à travers une présentation de l'entreprise ainsi que de l'environnement influençant et influencé par les activités de cette dernière.

1. Le Groupe Renault dans le monde

1.1. Présentation générale et historique

Le Groupe Renault est un constructeur automobile international dont les ventes de véhicules ont dépassé les 2,8³ millions en 2015, et ce, dans 125 pays dans le monde.

Le Groupe a réalisé, la même année, un chiffre d'affaires de 45 327 M € et a détenu un capital de 1 127 M € réparti entre différents actionnaires, comme le montre la figure II.1.

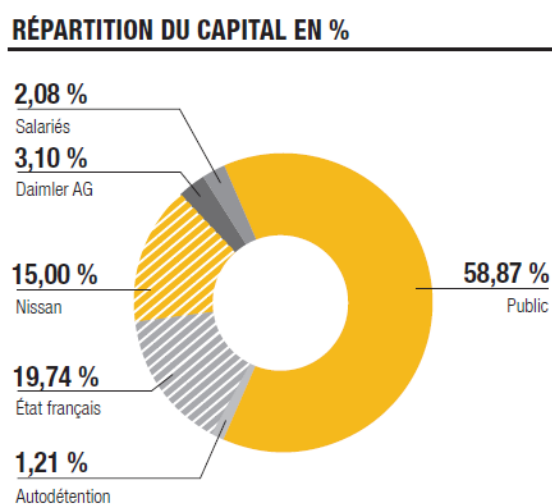


Figure II. 1: Actionnariat du Groupe Renault au 31 Décembre 2015 (Document de référence 2016, du Groupe Renault)

Tout commence en 1898 lorsque Louis Renault a créé la boîte de vitesse à prise directe permettant à sa marque de se spécialiser dans la construction de véhicules particuliers et de taxis.

L'entreprise se développe peu à peu, notamment durant la première guerre mondiale, en produisant de nombreux camions, chars et moteurs d'avions.

Renault est ensuite nationalisée par le gouvernement français en Janvier 1945. Elle devient « Régie Nationale des Usines Renault » et produit, dans les années 70, l'un des plus grands succès commerciaux du Groupe, la Renault 5. Ce n'est qu'en 1990 que Renault devient une Société anonyme avant sa privatisation, 6 années plus tard.

³ Dans la partie qui va suivre, les chiffres et les statistiques ont été tirés à partir de rapports et documents de référence du Groupe Renault, accessibles à partir du site : www.group.renault.com

L'année 1999 est une autre année historique pour la marque puisqu'elle représente la signature de l'alliance avec Nissan ainsi que l'acquisition de la marque roumaine Dacia, suivie, en 2000, par l'acquisition de la nouvelle marque Samsung, en Corée du Sud.

L'engagement de l'entreprise dans le sport automobile a fini par payer puisque pour la deuxième année consécutive, en 2006, Renault F1 Team est sacrée double Champion du monde constructeurs et Pilotes.

En 2013, Renault lance la ZOE, sa première voiture 100% électrique et devient vite leader des ventes de véhicules électriques en Europe.

Aujourd'hui, Renault est la première marque française dans le monde et emploie plus de 120 000 collaborateurs (dont 18,8% sont des femmes) répartis dans les 5 régions du groupe : Amériques, Asie-Pacifique, Région Afrique, Moyen-Orient, Inde (AMI), Eurasie et l'Europe.

1.2. Alliances et partenariats

Le Groupe Renault a, depuis longtemps, multiplié les partenariats et les alliances stratégiques avec les professionnels dans le but de créer des synergies et de générer toujours plus de valeur.

Au cœur de ces partenariats, nous retrouvons l'emblématique Alliance entre Renault et Nissan, à travers laquelle les deux marques ont réussi à nouer des collaborations stratégiques solides avec plusieurs autres constructeurs automobiles, notamment avec Daimler, AVTOVAZ et Mitsubishi Motors.

a) Alliance Renault-Nissan :

L'Alliance Renault-Nissan est un partenariat stratégique qui a été signé en 1999 entre la marque Française Renault et la marque Japonaise Nissan.

Quelques 17 années plus tard, l'Alliance transnationale devient le partenariat le plus durable et le plus productif de l'histoire de l'industrie automobile.

En 2015, l'Alliance Renault-Nissan a réalisé des ventes record de plus de 8,5 millions de véhicules, sous 8 marques différentes (Renault, Dacia, Renault Samsung Motors, Nissan, Infiniti, Datsun, Venucia et Lada), et ce, dans plus de 200 pays à travers le monde. Un chiffre qui représente près de 10% du marché mondial la même année, permettant ainsi à l'Alliance de se hisser à la 4^{ième} place du classement des constructeurs automobiles à l'échelle internationale, derrière Toyota, Volkswagen et General Motors.

L'Alliance a fortement contribué à l'amélioration des performances de Renault et Nissan à travers la mise en commun des moyens de production et logistiques, et ce, sans altérer l'identité de chacune. En effet, l'Alliance a mis en place une approche appelée CMF (Common Module Family) qui permet aux deux marques de réaliser d'importantes économies d'échelle en standardisant des pièces et des modules non visibles par les clients tout en conservant la diversité au niveau du design. Les modules communs à Renault et Nissan sont représentés dans la figure II.2.

COMMON MODULE FAMILY (CMF) : 4+1 BIG MODULES

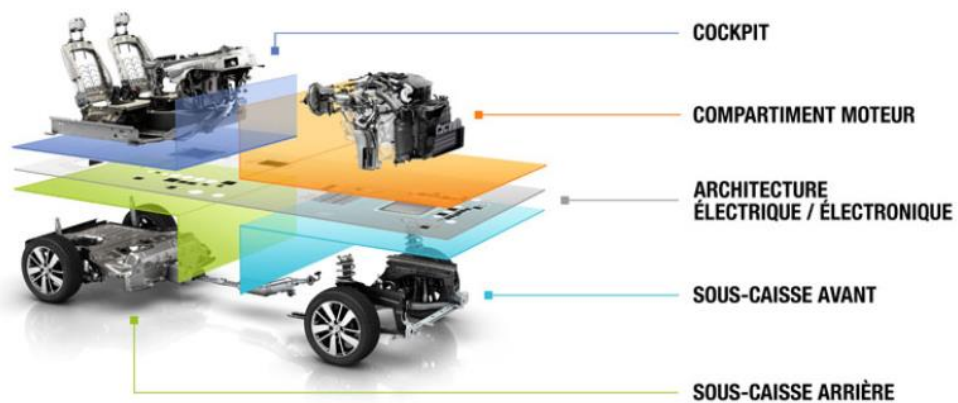


Figure II. 2: Modules communs à Renault et Nissan. (Source : www.group.renault.com).

En plus de la fabrication et de la logistique, et toujours dans le cadre de l'Alliance Renault-Nissan, les deux marques n'hésitent pas à mettre en commun d'autres fonctions opérationnelles clés dans l'industrie automobile, à savoir, l'Ingénierie, les Achats et la Ressources Humaines.

Les participations des marques Renault et Nissan dans l'Alliance sont représentées dans la figure II.3.

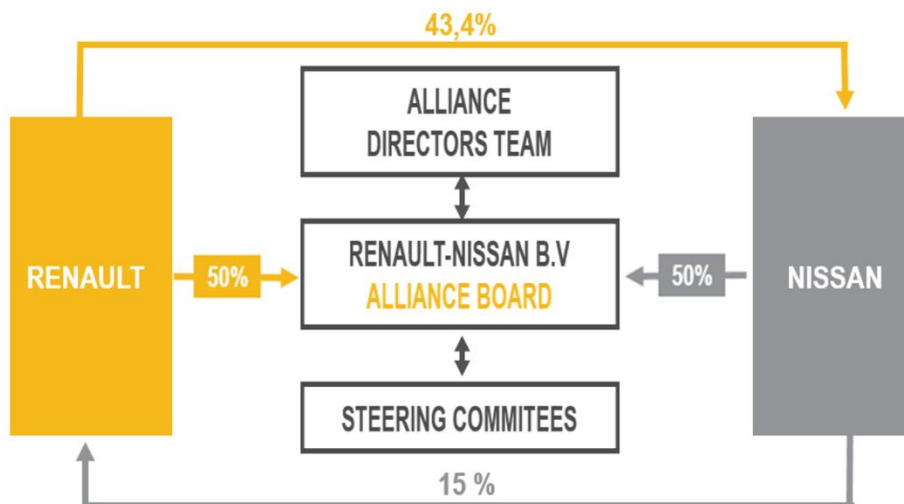


Figure II. 3: Pourcentages de participation de Renault et Nissan dans l'Alliance.

b) AVTOVAZ :

En 2008, l'Alliance Renault-Nissan a signé un partenariat avec le premier constructeur automobile russe : AVTOVAZ. Une collaboration stratégique qui a permis à l'Alliance de disposer, avec AVTOVAZ, de 4 usines en Russie : Togliatti, Moscou, Saint Pétersbourg et Izhevsk, qui produisent différents modèles tels que : Renault Logan et Sandero, Lada XRAY, Nissan Almera et Lada Largus.

En raison de l'importance de cette collaboration, les deux partenaires ont mis en place, en 2013, une structure commune d'achats en Russie réalisant ainsi des économies non négligeables.

c) DAIMLER :

Le 7 avril 2010 marque le partenariat entre l'Alliance Renault-Nissan avec l'un des premiers constructeurs automobiles haut de gamme dans le monde sous la marque Mercedes-Benz, le groupe allemand Daimler AG. Dans le cadre de cet accord, l'Alliance participe au capital de Daimler à hauteur de 3,1%. En échange, Daimler participe à son tour, avec le même pourcentage, dans chacun des capitaux de Renault et de Nissan.

Ce partenariat avec Daimler a permis à Renault de créer un nouveau véhicule utilitaire, Citan, basé sur la technologie Renault et produit par la marque Mercedes-Benz, de développer une nouvelle architecture commune à Renault Twingo et la nouvelle génération smart de Daimler et d'accroître la vente de moteurs diesel et essence à faible consommation. En plus de Daimler et d'AVTOVAZ, l'Alliance a établi, en 2013, des relations industrielles étroites avec Mitsubishi et DONGFENG, déjà partenaires du constructeur japonais Nissan.

1.3. Activités du Groupe

En plus des participations dans Nissan et AVTOVAZ, les activités du Groupe Renault s'insèrent dans deux principaux secteurs :

L'Automobile : Renault s'occupe de la conception, la fabrication et la commercialisation de :

Véhicules particuliers, sous les trois marques Renault, Dacia et Renault Samsung Motors (RSM), sachant que, dans certains pays, les véhicules fabriqués par Dacia et RSM peuvent être commercialisés sous la marque Renault. Ainsi, le Groupe Renault est présent dans le monde dans le segment des voitures compactes représentées par les segments A, B et Assimilés VP, avec une large gamme de modèles tels que : KWID, LOGAN, SANDERO, SYMBOL, CLIO, CAPTUR, SCALA et KANGOO, dans les segments C et D (y compris les SUV et monospace) avec des modèles comme MEGANE, FLUENCE, DUSTER, KADJAR, SCENIC et ESPACE ainsi que dans le segment S avec LATITUDE et TALISMAN. Des modèles complétés par des véhicules sportifs issus de la gamme Renault, tels que Clio Renault Sport et Megane Renault Sport.

Véhicules utilitaires avec le service Renault Pro+ permettant au Groupe de répondre efficacement aux exigences spécifiques des clients professionnels, notamment grâce à l'offre sur mesure et les transformations à la carte proposées par le service.

Véhicules électriques vendus exclusivement sous la marque Renault. Actuellement, 3 voitures sont déjà commercialisées (TWIZY, KANGOO Z.E. et ZOE), permettant ainsi au Groupe de conserver son leadership dans le segment en réalisant un total de 83 300 ventes de véhicules électriques dans le monde, depuis le lancement du premier modèle électrique en 2011.

Le Groupe assure également la distribution de ses produits à travers son réseau commercial, notamment, via sa filiale Renault Retail Groupe. En plus de la vente de véhicules neufs, le Groupe Renault s'occupe de la commercialisation de pièces de rechange et des

véhicules d'occasion ainsi que de l'aménagement des véhicules Renault à travers la Business Unit : Renault TECH.

Des activités auxquelles s'ajoute la gamme mécanique Renault issue de l'activité Business to Business qui est prise en charge par une direction dédiée (Au-delà de l'alliance Renault-Nissan). Cette activité permet au Groupe de proposer des organes mécaniques résultants des coopérations automobiles comme, par exemple, avec Daimler, AVTOVAZ et General Motors-Opel.

Les différents services : Même si les activités principales concernent la branche automobile, qui a contribué à hauteur de 95,1 % au chiffre d'affaires du Groupe en 2015, Renault continue de développer ses activités dans le secteur du service. Ces dernières regroupent le financement des ventes, la location, l'entretien et les contrats de service. A cet effet, la RCI Banque, une filiale à 100% de Renault s.a.s, a été créée afin d'assurer le financement des ventes des différentes marques du Groupe Renault ainsi que celles du Groupe Nissan (Nissan, Infiniti, Datsun), et ce, en proposant aux particuliers des offres de crédits et des solutions locatives pour les particuliers, en offrant des solutions performantes adaptées aux entreprises et en assurant le financement des stocks des réseaux de l'Alliance.

La figure II.4 présente, synthétiquement, la structure du Groupe Renault.

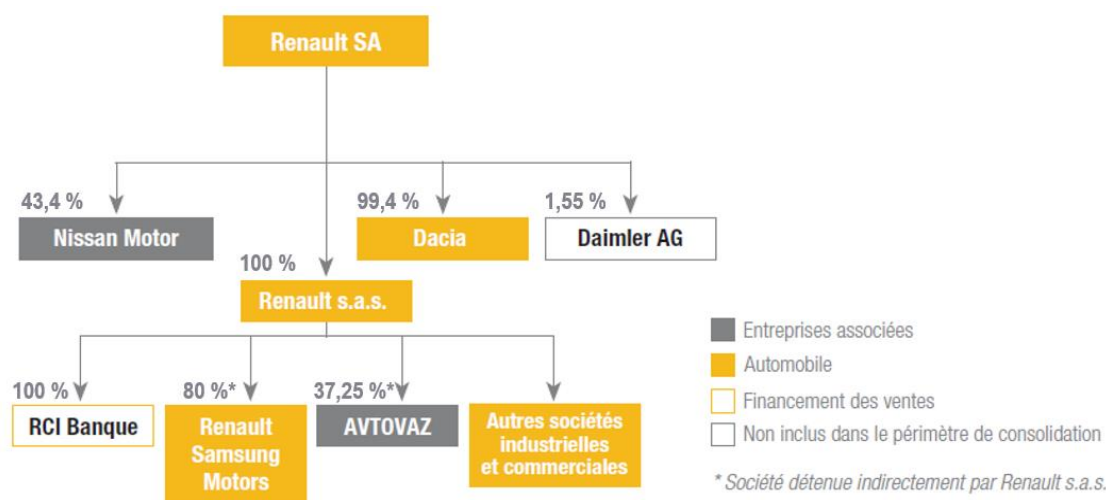


Figure II. 4: Structure du Groupe Renault (en % d'actions émises). (Source : Document de référence Renault 2016)

2. Renault en Algérie

La présence de Renault en Algérie remonte au 13 Novembre 1922, avec la fondation de la « Société Algérienne des Automobiles Renault (SADAR) », chargée de la commercialisation des véhicules de la marque dans le pays. Une présence confortée par l'ouverture, en 1959, de l'usine d'assemblage : « Construction des Automobiles Renault en Algérie (CARAL) » qui produira la R4, la R8 et la R16.

Toutefois, après leur fusion, les deux entreprises cessent leurs activités suite à leur nationalisation en 1969 et c'est donc l'entreprise publique SONACOME, puis son avatar ENDVP, qui assurent les fonctions d'importation, de distribution et de maintenance des véhicules particuliers et petits utilitaires.

Après la fin des contrats d'Autorisations d'importations de Véhicules (AIV), Renault revient en Algérie en 1987 avec l'ouverture d'un bureau de liaison à Alger.

Ce n'est qu'en 1998 que Renault Algérie S.P.A est créée, permettant à la marque du losange de réaliser des chiffres de croissance élevés dans le pays et d'ouvrir, en 2014, une usine d'assemblage à Oran : « Renault Algérie Production (RAP) », une JV entre le Groupe Renault (49%), le Fond National d'Investissement FNI (17%) et la SNVI (34%), qui donnera naissance à la première voiture « Made in Bladi ».

2.1. Entreprise Renault Algérie

L'entreprise Renault Algérie, sise à la zone industrielle d'Oued Smar, Alger, est une S.P.A au capital social de plus de 1 Milliard de Dinars en 2015.

Présente dans le pays depuis 1998, la filiale du Groupe Renault maintient sa position de leader dans le marché algérien de l'automobile pour la 10^e année consécutive, avec 90 182 véhicules vendus en 2015 et une part de marché record de 35,6% la même année.

En plus de la vente de véhicules neufs (Véhicules aux Particuliers, Véhicules Utilitaires et Véhicules Utilitaires à Adaptation Complémentaire), la prestation de services après-vente (Entretien, Réparation Mécanique et Carrosserie, Installation d'accessoires, Assistance et vente de pièces de rechange), la production de Véhicules aux Particuliers (à travers Renault Algérie Production) et la formation des stagiaires du réseau d'agents agréés ainsi que toute personne désireuse d'acquérir l'un des nombreux savoirs automobiles (Renault Algérie Académie), le groupe Renault est le seul constructeur en Algérie qui propose à sa clientèle la vente de véhicules d'occasion.

Aujourd'hui, l'entreprise est présente en Algérie avec 8 principales installations : Le siège Renault Algérie, 3 succursales (Oued Smar, les Grands Vents, Oran), le centre livreur de Chéraga, CI Jijel, Renault Algérie Académie, le magasin central PR, et pas moins de 67 agents agréés, répartis dans l'ensemble du territoire national.

L'organigramme présenté dans la figure II.5 résume les principales fonctions prises en charge par plus de 400 collaborateurs de l'entreprise.

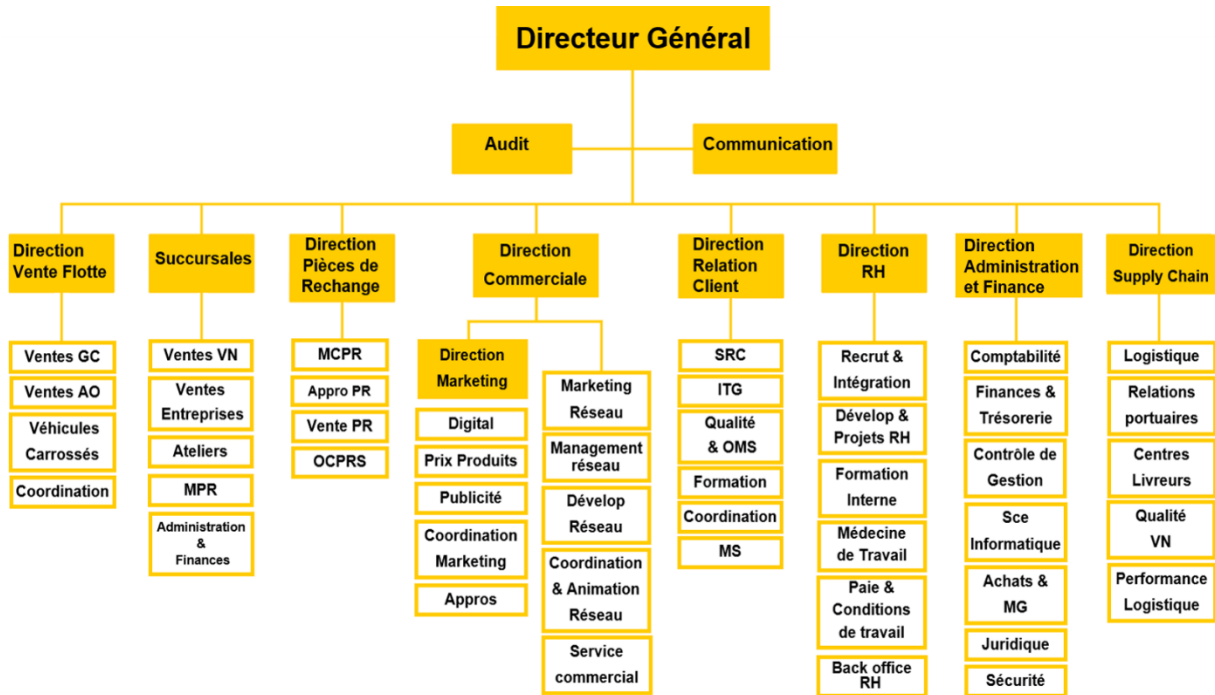


Figure II. 5: Organigramme de Renault Algérie.

A partir de l'organigramme, nous pouvons constater que la diversification des activités de l'entreprise Renault Algérie s'est traduite par une fonction Marketing transversale.

D'une part, la direction commerciale qui s'occupe de l'activité principale de l'entreprise, à savoir, la vente de véhicules neufs, possède sa propre direction Marketing, à laquelle sont attachés les départements Prix produit, Approvisionnement, Digital, Coordination et Publicité.

D'autre part, la direction des ventes flottes, la direction des Pièces de Rechanges et les succursales ont leurs propres responsables Marketing.

Par ailleurs, l'ensemble du réseau commercial de Renault Algérie, qui n'emploie pas forcément des spécialistes dans le Marketing (y compris les succursales), peut compter sur le soutien du département Marketing Réseau, de la direction commerciale, qui s'occupe de la coordination des actions entreprises par les agents ainsi que leur cohérence avec l'image de marque de l'entreprise.

Ainsi, chaque unité commerciale possède sa propre structure Marketing, plus ou moins indépendante (hors campagnes transversales telles que les lancements de produits). De plus, mis à part la direction Pièces de Rechange, toutes les autres structures transitent par le service Publicité de la direction Marketing pour leurs campagnes de communication publicitaires au lieu de contacter directement l'agence attitrée (Agency Of Record AOR), pour des raisons logistiques mais aussi et surtout pour l'expertise de ce service dans le domaine de la communication.

2.2. Analyse du marché de Renault Algérie

Afin de préparer l'identification de la problématique sur laquelle portera notre projet, nous décidons de procéder à une analyse exhaustive de l'environnement qui influence et est influencé par l'ensemble des activités de l'entreprise.

Pour ce faire, nous réalisons une analyse du macro-environnement de l'entreprise, basée sur la méthode PESTEL, suivie d'une évaluation du microenvironnement à l'aide des 5+1 Forces de PORTER.

a) Analyse du macro-environnement avec PESTEL :

La compréhension et l'anticipation du macro-environnement dans lequel évolue l'entreprise est un facteur déterminant pour la pérennité de cette dernière. A cet effet, nous utiliserons l'analyse PESTEL afin d'identifier l'influence que peuvent exercer les facteurs Politiques, Economiques, Sociologiques, Technologiques, Ecologiques et Légaux sur l'entreprise Renault Algérie.

Du point de vue Politique, une grande importance est accordée au secteur de l'industrie mécanique. Et pour cause, la facture d'importation de véhicules a atteint les 6,34 milliards de dollars en 2014, soit 9% de la facture globale d'importation, et ce, dans une situation économique lourdement marquée par la crise pétrolière. Pour y remédier, l'Etat décide de limiter les importations de véhicules (un peu moins de 400 000 en 2014) à près de 300 000 unités en 2015 pour enfin les réduire à 83 000 véhicules en 2016. Une situation qui a conduit à une réelle dynamique d'investissement dans la production automobile en Algérie, notamment avec la création de l'usine Renault Algérie Production en Novembre 2014 ainsi l'ouverture prochaine des usines de PSA Peugeot Citroën, Volkswagen et Fiat⁴, en 2017.

L'économie du pays étant principalement alimentée par la rente des hydrocarbures, l'évolution des cours détermine fortement le pouvoir d'achat de la population algérienne. Un pouvoir d'achat qui s'est détérioré, notamment avec la récente crise pétrolière et le retour de l'inflation en 2015 (près de 4,8 % selon le ministère des finances). Cependant, afin de déterminer le potentiel du marché automobile algérien, deux indicateurs peuvent être utilisés : Le taux d'équipement automobile de la population⁵ et l'âge moyen du parc roulant. Le premier avoisinait, en 2014, les 130 véhicules par 1000 habitants, en Algérie, contre 600 en moyenne, la même année en France. Quant au deuxième, il indique un parc automobile vieillissant, avec un âge moyen de 16 ans ⁶en 2014 en Algérie contre environs 8 ans ⁷la même année en France. Des résultats qui permettent à Renault Algérie d'affirmer que le marché algérien de l'automobile est « à fort potentiel sur le long terme ». Un marché dans lequel l'entreprise est leader depuis une dizaine d'années, et ce, en dépit d'une concurrence très agressive sur le marché.

⁴ Source : Documents internes Renault Algérie.

⁵ « Sur un marché donné, le taux d'équipement correspond au nombre de ménages ou d'individus équipés du bien divisé par le nombre total de ménages ou d'individus. Le taux d'équipement est un indicateur de potentiel / saturation d'un marché et permet notamment de voir s'il s'agit d'un [marché de renouvellement](#) ou d'un [marché de premier équipement](#) ». B.Bathelot, le 2 avril 2015. www.definitions-marketing.com.

⁶ Selon Gilles Guillaume, édité par Yves Clarisse. www.zonebourse.com

⁷ Selon le rapport du Comité des Constructeurs Français d'Automobiles

L'environnement macrosociologique dans lequel évolue l'entreprise Renault Algérie est principalement caractérisé par une population jeune (la deuxième plus jeune au Maghreb) avec une moyenne des âges de 27,3 ans⁸ en 2014 et un taux de chômage qui avoisine les 11,2%⁹ en 2015. L'automobile représente néanmoins le 3^{ème} poste de dépense des foyers Algériens dont seulement 46% sont dotés d'une automobile¹⁰. Il est également important à signaler que les femmes prennent de plus en plus le volant en Algérie, à titre illustratif, les concessionnaires estiment que plus de 50% des clients ayant acheté des véhicules durant le salon de l'automobile 2015 sont des femmes¹¹. Par ailleurs, l'opinion de la population concernant les marques françaises reste mitigée. D'une part, les repères culturels francophones présents dans le pays constituent des facteurs d'attachement aux marques françaises. D'autre part, en plus du passif colonial qui représente une source de rejet pour une certaine tranche sociale, les jeunes sont de plus en plus orientés vers des repères culturels orientaux.

Les importantes avancées technologiques de ces dernières années ont poussé les constructeurs automobiles à accorder une importance considérable à la recherche et au développement. En plus des progrès réalisés sur la motorisation écologique qui n'est pas encore destinée aux pays sous-développés ou en voie de développement comme l'Algérie, nous assistons à l'apparitions de nouvelles fonctionnalités dans les véhicules commercialisés en Algérie tels que : l'*Easy Park Assist*, le détecteur de fatigue, régulateur de vitesse adaptatif, etc.

L'écologie influence l'industrie automobile mais la subit également. D'une part, les véhicules causent d'importantes nuisances environnementales qui se traduisent en pollution atmosphérique et émission de gaz à effet de serre. D'autre part, l'intervention de l'Etat dans la protection de l'environnement conditionne les activités du secteur automobile. Cependant, les législations anti-pollution en vigueur en Algérie sont nettement moins strictes qu'en Europe, aux USA ou au Japon. En effet, nous parlons toujours de motorisations aux normes Euro 3 pour le Diesel et Euro 4 pour l'essence en Algérie, alors qu'en Europe et dans plusieurs autres pays du monde, une motorisation aux normes Euro 6 est en vigueur. Ceci s'explique par la qualité du carburant dans le pays, considéré parmi les plus pollués en Afrique.

Les facteurs légaux et réglementaires affectent d'une façon significative le marché automobile Algérien. En effet, les entreprises automobiles font constamment face à des contraintes légales relatives à des normes en matière de sécurité, de pollution, etc. Par exemple, l'entrée en vigueur du nouveau cahier des charges, la nouvelle taxe véhicules neufs ou encore le nouveau processus d'homologation des véhicules utilitaires en 2015 ont fortement conditionné le marché automobile en Algérie.

L'un des principaux points que nous pouvons retenir de cette analyse PESTEL est la nécessité d'investir dans la production nationale. Renault Algérie a déjà implanté une usine à Oran qui produit actuellement la Symbol et bientôt la Sandero de la marque Dacia.

Cependant, comme nous l'avons précisé plus haut, d'autres constructeurs comptent implanter leurs usines en Algérie très prochainement, et ce, avec des capacités de production

⁸ Source : Site d'information américain GlobalPost.com

⁹ Source : Office National des Statistiques

¹⁰ Source : Communiqué de Quentin Le Helet, directeur de GIPA Maghreb (société spécialisée dans l'analyse des données automobiles)

¹¹ Source : www.algerie-focus.com

égales ou supérieures à la capacité actuelle de l'usine de Renault. Une situation qui risque de redessiner le marché automobile Algérien.

De plus, l'importante baisse dans les quotas d'importation accordés par le gouvernement algérien aux constructeurs et concessionnaires automobiles, réduit les ventes sous importation de Renault Algérie à 15 000 véhicules pour l'année 2016 (le quota le plus important des 40 accordés) pour un montant total ne dépassant pas les 138 M \$ (sachant que ce montant avoisinait les 1 000 M\$ en 2015). En gardant la même stratégie, l'entreprise verrait son chiffre d'affaires, relatif aux importations, chuter considérablement. D'où la nécessité de repenser le mix produit de façon à optimiser la rentabilité, de développer d'autres sources de revenu tels que le service après-vente et de réduire au maximum les coûts, notamment ceux du Marketing.

b) Evaluation de la position concurrentielle de l'entreprise :

L'environnement concurrentiel d'une entreprise, qui se limitait aux concurrents directs de cette dernière, se compose aujourd'hui de l'ensemble des acteurs susceptibles d'affecter les bénéfices de l'entreprise. C'est dans ce contexte que nous utiliserons les 5 +1 Forces de Porter afin de déterminer ces acteurs ainsi que leur impact sur les activités de Renault Algérie.

Renault Algérie, étant une filiale du Groupe Renault, se fait livrer par les usines du Groupe à travers le monde. Cette négociation se fait principalement sur la base des volumes demandés puisque de gros volumes engendrent plus d'économies d'échelles pour les usines. Ainsi, dans un marché automobile algérien réduit de moitié, les ventes de l'entreprise réalisées sur la base des importations en seront forcément affectées. Le pouvoir de négociation des fournisseurs est donc : Fort.

La pouvoir de négociation des clients dans le secteur automobile est souvent jugé important en raison de la faible différenciation entre les produits et du fait qu'il est de plus en plus difficile, aujourd'hui, de fidéliser les clients à la marque. Toutefois, dans le contexte actuel du marché algérien, caractérisé par une demande largement supérieure à l'offre (en raison d'une limitation importante des importations), les experts du secteur au sein de l'entreprise Renault Algérie s'accordent à considérer le pouvoir de négociation des clients comme étant : Très faible.

Les produits de substitution de la voiture sont des produits remplissant les mêmes fonctions mais avec des technologies différentes. Ainsi, nous avons, d'une part, les voitures utilisant des technologies plus propres telles que les voitures électriques, et d'autre part, les deux-roues et les transports collectifs. Aujourd'hui, les voitures « écologiques » ne se sont pas encore implantées dans le marché automobile Algérien, ne permettant pas ainsi à ces dernières de remplacer les voitures existantes. Concernant la seconde catégorie de substituts, outre la concurrence importante des trains et des avions sur les longues distances, aucun des moyens de transport en commun ou des deux-roues ne semble, à ce jour, remplacer le confort d'une voiture en Algérie. La menace des produits de substitution est donc relativement : Faible.

La menace des nouveaux entrants dans le marché automobile Algérien est considérée, elle aussi, comme étant : faible, notamment dans un contexte de limitation des importations. En effet, les autorités algériennes fournissent de plus en plus difficilement les licences d'importations aux nouveaux concessionnaires. Le seul moyen pour un nouveau constructeur automobile de pénétrer le marché algérien serait alors d'implanter une usine de production dans

le pays, ce qui représente un investissement très important, confirmant ainsi la faible menace que représentent les nouveaux entrants, au moins, sur le court terme.

Bien que Renault et Dacia aient occupé les 1^{ière} et 2^{ième} places du classement des marques selon les ventes réalisées en 2015, la concurrence n'en demeure pas moins intense. En effet, nous pouvons remarquer, dans la figure II.6, que les quantités des ventes réalisées par chacune des deux marques sont proches de celles de leurs principaux concurrents : Peugeot et Hyundai.

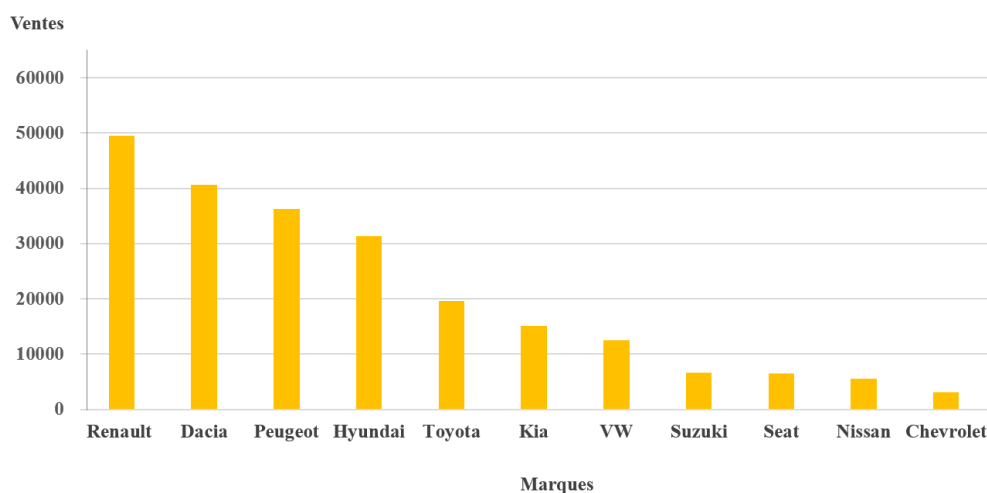


Figure II. 6: Ventes par marque en 2015. (Source : Documents internes de Renault Algérie).

De plus, en analysant les ventes des premières marques de 2015 durant ces 5 dernières années, présentées dans la figure II.7, nous pouvons constater l'intensité de la concurrence entre ces marques qui se partagent le podium.

Par ailleurs, en s'appuyant sur l'annexe 1, nous remarquons que le groupe Renault est incontestablement leader dans le segment M0, son principal segment (qui représente 37% des ventes du groupe) avec la double offre Dacia Logan et Renault Symbol ainsi que dans le SUV, avec Dacia Duster. Par contre, le groupe est fortement challengé par son concurrent Peugeot dans le segment I représenté par Renault Clio et Dacia Sandero et est largement dépassé par ses concurrents dans le segment M1, en raison du cycle de vie très courts de la Renault Megane.

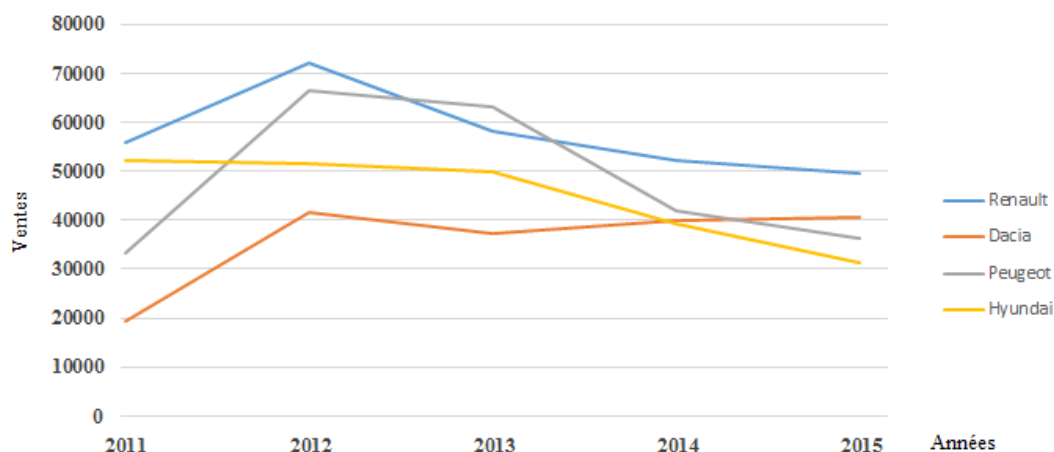


Figure II. 7: Evolution des ventes des marques Renault, Dacia, Peugeot et Hyundai, de 2011 à 2015. (Source : Documents internes de Renault Algérie).

Les décisions gouvernementales affectent fortement le marché automobile algérien provoquant ainsi de grands bouleversements dans les ventes de ces dernières années, comme le montre la figure II.8.

En effet, selon les experts du secteur automobile, les importantes revalorisations salariales accordées par le gouvernement algérien en 2011, à défaut de permettre aux fonctionnaires concernés d'acquérir des logements, ont provoqué un record des ventes en 2012, dépassant les 433 000 véhicules soit presque le double des ventes enregistrées en 2010. Cette hausse est également alimentée par l'augmentation des ventes de véhicules utilitaires suite au programme d'incitation de l'ANSEJ.

Cette situation ne persiste cependant pas, puisque le nouveau décret automobile obligeant les constructeurs d'équiper tous les véhicules neufs d'airbags et d'ABS ainsi que la relance des projets sociaux AADL en 2013 et 2014, provoquent la baisse des volumes des ventes en 2014.

Une baisse qui persiste en 2015, avec l'entrée en vigueur du nouveau cahier des charges, la nouvelle taxe pour les voitures neuves, la nouvelle procédure d'homologation des VUAC ainsi que la limitation des importations.

Ces réglementations continuent, en 2016, d'influencer le marché automobile en Algérie, notamment avec les nouveaux quotas d'importation.

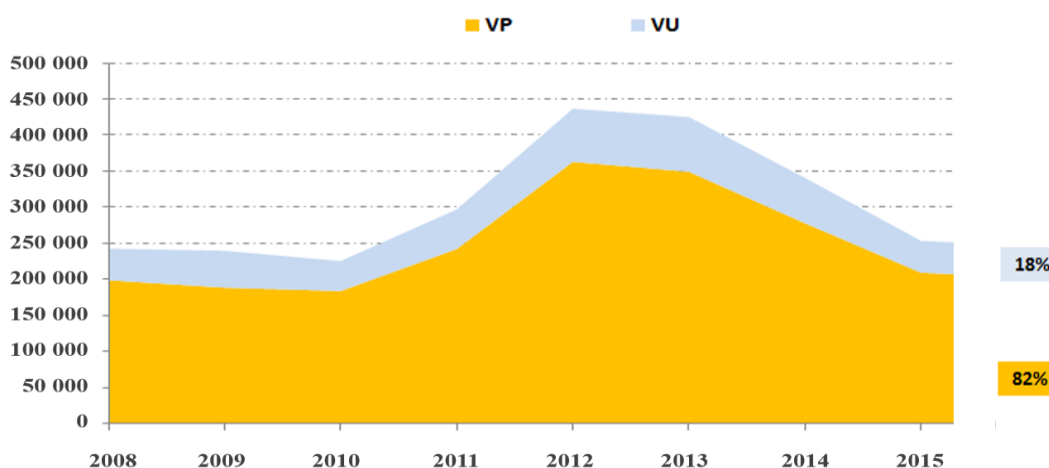


Figure II. 8: Evolution des ventes de véhicules neufs entre 2008 et 2015 en Algérie. (Source : documents de Renault Algérie).

Les résultats de l'analyse concurrentielle basée sur les 5 + 1 Forces de PORTER nous permet de mettre en évidence le contexte réglementaire fluctuant dans lequel évolue le marché automobile algérien, animé par une forte concurrence.

3. Définition de la problématique

3.1. Analyse du contexte

Le Groupe Renault est devenu, depuis quelques années, spécialiste dans les voitures populaires, notamment avec la révolution *Low Cost*¹² qu'il a entamé avec la marque Dacia. Cependant, l'ancien constructeur de voitures de luxe avant-guerre vise désormais la conquête de nouveaux segments, notamment celui du haut de gamme.

En effet, les multiples alliances et partenariats permettent à Renault d'envisager le haut de gamme à moindre coût. C'est d'ailleurs consciente de cet avantage concurrentiel que la marque au losange décide d'opter pour une stratégie « Back to Premium » soutenue par le lancement de produits tels que la luxueuse berline Talisman, la nouvelle Megane ou encore, le nouvel Espace.

Renault Algérie est une filiale du Groupe Renault. Elle partage de ce fait la même vision et les mêmes objectifs stratégiques. Cette dépendance s'est traduite, entre autres, par le lancement de voitures plus prestigieuses que celles proposées habituellement par la marque au client algérien. Les modèles concernés sont : La Kadjar (sortie en Mars 2016), la nouvelle Megane (prévue pour le deuxième semestre 2016) ainsi que la Talisman (prévue pour 2017).

Toutefois, la marque ne souhaite pas se défaire de sa stratégie du *Low Cost*, qui lui a d'ailleurs valu le rang de leader en Algérie, bien au contraire, puisque l'usine d'Oran devra

¹² Le Low Cost est « une stratégie commerciale consistant à proposer un bien ou un service à un prix inférieur à ceux que pratiquent habituellement les entreprises concurrentes ». (Source : www.larousse.fr)

entamer la production de la Dacia Sandero cette année, en plus de la Symbol Made in Bladi produite par l'usine depuis le 10 Novembre 2014.

Par ailleurs, le Groupe Renault souhaite également s'adapter aux attentes, de plus en plus exigeantes, des clients d'aujourd'hui. Pour ce faire, il lance le programme C@RE (pour *Customer Approved Renault Experience*) en 2012, dans 7 pays pilotes, dont l'Algérie, avant d'être déployé dans 30 pays d'ici la fin 2016.

Un programme encore rare dans le secteur automobile dont le but ultime est l'optimisation de la relation client à travers une démarche multicanal (en unissant le Web au réseau physique) afin de prendre en considération l'ensemble du processus commercial d'un client, c'est-à-dire la conquête, la conversion et la fidélisation.

En outre, ce programme comprend plus de 40 projets dont : le portail My Renault qui permet aux clients d'accéder à un espace personnalisé, la création des Renault Stores, la formation comportementale des équipes technico-commerciales au sein de la Renault Académie ainsi que la fusion des bases de données afin d'obtenir un seul historique client, en plus de celui des produits.

L'un des principaux objectifs à travers le déploiement d'un tel programme consiste à permettre au client de : « bénéficier d'offres de fidélisation adaptées à son profil et à son véhicule ». (www.media.renault.com. Communiqués ; 17 avril 2013).

Cependant, une enquête appelée NCBS (pour New Car Buyers Survey) ¹³a été réalisée par un cabinet indépendant auprès de 1 993 acheteurs algériens de véhicules neufs.

En analysant les données de cette enquête¹⁴, nous avons constaté, qu'en dépit des efforts de l'entreprise dans l'amélioration de la relation client :

- La satisfaction globale des acheteurs vis-à-vis des deux marques Renault et Dacia figure parmi les plus faibles.
- Malgré le fait que les prix moyens proposés par Renault et Dacia soient les plus bas dans les différents segments, la satisfaction des clients vis-à-vis du critère « prix » est la moins élevée.
- En général, peu de clients des marques Renault et Dacia achètent des services complémentaires chez leurs concessionnaires.
- Parmi les différentes raisons d'achat des clients, Renault est la marque la moins choisie pour son « Image et Statut ».
- Le taux des acheteurs ayant envisagé l'achat d'une voiture d'occasion est le plus élevé pour la marque Renault que pour les autres marques considérées.

¹³ NCBS (New Car Buyers Survey) est une enquête qui a été réalisée entre Mars 2015 et Aout 2015 sur un échantillon de 1 993 acheteurs de voitures neuves (VP) en Algérie de 13 marques différentes (dont Renault et Dacia). Les ventes couvertes sont de 136 623 unités sur une période allant de Septembre 2014 à Mars 2015.

¹⁴ Un aperçu du rapport que nous avons réalisé sur l'enquête NCBS 2015, est présenté en annexe 2.

3.2. Identification de la problématique

La connaissance des objectifs stratégiques de l'entreprise ainsi que les différentes analyses que nous avons réalisées, à savoir, l'analyse des facteurs macro-environnementaux par la méthode PESTEL, l'évaluation de la position concurrentielle par les 5 + 1 forces de PORTER et l'analyse de l'enquête NCBS 2015, nous ont permis d'apprécier la situation actuelle dans laquelle évolue l'entreprise et de résumer les différents challenges auxquels elle est confrontée comme suit :

- Une forte intensité concurrentielle.
- La volonté de l'entreprise de diversifier ses offres sur différents aspects (mix produit, versions et services après-vente) dans le but de rehausser l'image de marque.
- L'engagement de l'entreprise dans le développement de la relation avec les clients et leur fidélisation.
- Construire une stratégie efficace pour la commercialisation des véhicules produits en Algérie.
- La nécessité de rentabiliser chaque cout investi, notamment en termes de campagnes marketing.

Dans ce contexte, lors d'entretiens avec les services Marketing et Commercial, nous avons convenu de l'importance de mieux cerner les caractéristiques des différents segments de clients afin de répondre au mieux à leurs attentes et exigences.

En effet, en marketing, la segmentation est un outil de différenciation qui permet aux entreprises de gagner en efficacité à travers la construction de stratégies différenciées, adaptées aux différents groupes ciblés. Elle constitue de ce fait un avantage concurrentiel non négligeable.

Comme mentionné dans le premier chapitre, il existe plusieurs types de segmentations, servant différents objectifs.

Les segmentations les plus connues et les plus utilisées sont celles réalisées à partir de la valeur apportée par un client ou encore des données démographiques de ce dernier.

Toutefois, l'accroissement de l'intensité concurrentielle pousse de plus en plus d'entreprises à s'orienter vers des segmentations offrant une vue plus exhaustive sur les clients, et ce, en ajoutant d'autres caractéristiques telles que les besoins, préférences, etc.

Dans la pratique, une segmentation des clients d'une entreprise peut être réalisée soit, à partir des bases de données existantes au sein de cette dernière, soit en utilisant des données récoltées à partir d'études de marché sur les clients. Il arrive également qu'une étude de marché soit réalisée lorsqu'une entreprise souhaite analyser l'ensemble du marché.

Pour notre projet, en prenant en considération le fait qu'une segmentation traditionnelle (basée uniquement sur la valeur apportée par le client ou les données démographiques de celui-ci) ne soit plus suffisante pour permettre à l'entreprise de prendre des décisions pertinentes, nous avons décidé de compléter notre analyse en incluant d'autres segmentations prenant en compte les critères comportementaux des acheteurs.

CHAPITRE II: Présentation de l'entreprise et Exploitation de la base de données clients.

Toutefois, de telles informations n'étant souvent pas (ou partiellement) disponibles au sein des bases de données de l'entreprise, nous jugeons utile de commencer notre étude par un diagnostic de la base de données de l'entreprise afin d'identifier le type de segmentation qu'il est possible de réaliser à partir de ces données avant d'envisager de réaliser une étude de marché.

Section II : Diagnostic et exploitation de la base de données clients

L'objet de cette section est de diagnostiquer la base de données clients de l'entreprise Renault Algérie, afin d'évaluer la possibilité de réaliser une segmentation efficace et exhaustive des clients de l'entreprise.

Avant cela, il est nécessaire de comprendre l'origine des flux d'informations relatives aux clients ainsi que le processus de collecte et stockage des données.

Pour ce faire, nous commençons par identifier les différents points de collecte de données clients tout au long de leur cycle de vie avant de procéder à la description des systèmes d'information clients (SIC) et au diagnostic effectif des bases de données.

Ce diagnostic comprend le nettoyage et de la préparation des données dans le but de permettre une éventuelle segmentation des clients à partir des données collectées par l'entreprise.

Enfin, une segmentation des clients à partir des données existantes est présentée et discutée à la fin de la présente section.

1. Cycle de vie du client et points de collecte des données

Le cycle de vie client est désigné comme l'enchaînement des événements influençant son comportement.

Durant les différents stades de la relation de l'entreprise avec ses clients, ces derniers sont en interaction avec différents canaux de communication, où de nombreuses informations les concernant sont collectées.

Le cycle de vie des clients du Groupe Renault Algérie se caractérise par deux phases distinctes : l'achat d'un véhicule neuf et l'achat des services après-vente.

Nous résumons l'ensemble des processus client intervenant tout au long du cycle de vie client dans une cartographie de niveau deux, présentée dans la figure II.9 suivante :

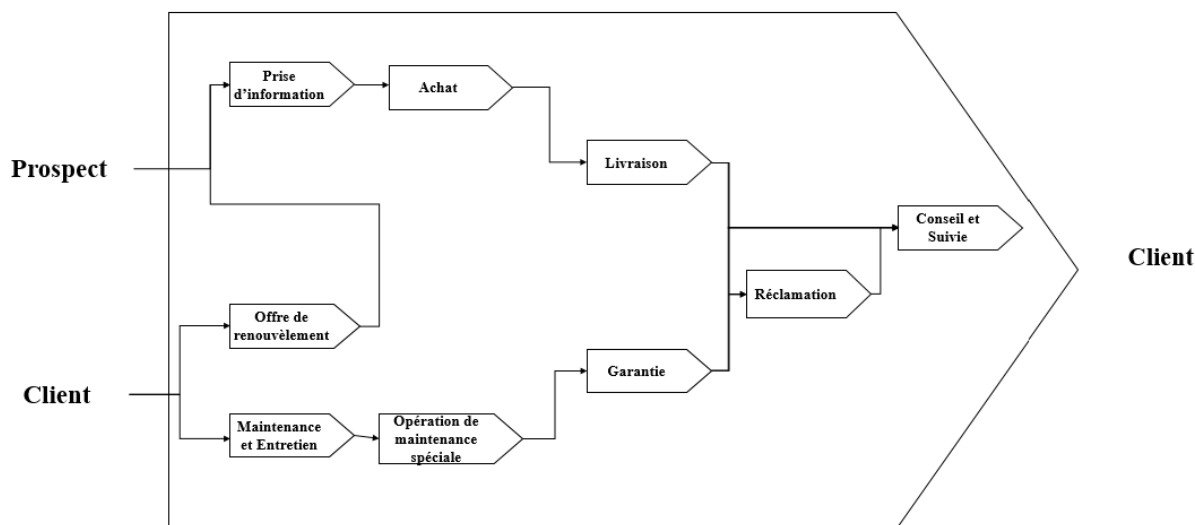


Figure II. 9: Cartographie de niveau deux des processus intervenant dans le cycle de vie du client de l'entreprise Renault Algérie.

Phase 1 : Achat du véhicule neuf

L'acquisition d'un véhicule neuf par un client passe par les étapes suivantes : la prise d'information (prospect), l'achat et enfin, la livraison et le conseil.

a) Processus de prise d'information :

L'entreprise met à disposition des prospects un large panorama de canaux de communication qui lui permettent de transmettre ses offres commerciales mais aussi de collecter des données sur les prospects. Ainsi, le processus de prise d'information constitue la première interaction entre les futurs clients et l'entreprise.

Cette interaction peut se faire lorsque le prospect se rend au show-room de la marque afin de s'informer sur les offres commerciales proposées. Dans ce cas, un conseiller client se charge de prendre les coordonnées personnelles du prospect. Malheureusement, aucun système d'information n'est mis en place pour la collecte et la transmission de ces informations à l'entreprise, sauf dans le cas où le prospect décide de commander un véhicule.

Il existe toutefois un autre point d'interaction entre les prospects et l'entreprise. Il s'agit du canal digital. En effet, lorsqu'un prospect consulte le site internet de la marque afin de demander des informations sur une offre en particulier une brochure technique et/ou commerciale est envoyée par e-mail au demandeur qui renseigne un formulaire du site internet de la marque avec ses coordonnées et informations personnelles.

Concernant la communauté interactive sur les réseaux sociaux, un prestataire de service se charge de la diriger vers le site web de la marque, ou bien vers le centre d'appel, qui constitue à son tour un autre point d'interaction avec les prospects mais aussi avec les clients de l'entreprise.

Toutes les requêtes et demandes d'information sont groupées dans un même système d'information de traitement des prospects appelé « LMT », littéralement en anglais « Lead Management Tool ».

Les prospects regroupés sur le LTM sont contactés par l'un des conseillers client du centre d'appel afin de fournir les informations demandées sur une offre en particulier et d'évaluer leur intérêt d'achat ou encore de leur proposer un rendez-vous chez le concessionnaire le plus proche de chez eux pour passer commande ou pour demander un test-drive.

Cependant, après la prise de rendez-vous avec le concessionnaire, aucun retour d'information de la part des concessionnaires ne se fait, rendant ainsi impossible le calcul du taux de conversion des prospects en clients et donc, le retour sur investissement des campagnes marketing réalisées par le canal digital.

En résumé, la collecte de données des prospects dans la phase « prise d'information » n'est réalisée qu'à travers le canal digital.

b) Processus d'achat :

Lorsqu'un prospect décide d'acheter un véhicule, il passe une commande au niveau des show-rooms. Le conseiller client transmet alors la commande sur le système d'information SEDRE en renseignant obligatoirement le nom, prénom, numéro de téléphone et e-mail du client. Un dossier physique comprenant plusieurs documents tels que la photocopie de la carte d'identité, un formulaire d'informations et le chèque de paiement, est alors constitué et remonté au service commercial du Groupe Renault Algérie pour la confirmation de la commande et l'établissement de la facture et la carte jaune.

Ainsi, les données des clients sont collectées d'une part, à travers le système d'information SEDRE mis à disposition des différents agents agréés et d'autre part, à travers le dossier physique au niveau du service commercial de l'entreprise.

c) Processus de livraison et conseil :

Le client est demandeur de conseils et astuces sur l'utilisation et l'entretien de son véhicule, tout au long de sa relation avec le constructeur. Pour ce faire, l'entreprise a récemment mis à la disposition de ses clients un canal privilégié qui est l'espace personnel My Renault qui permet de suivre la commande de la voiture ainsi que son entretien, d'établir les devis d'entretien et de profiter de diverses offres et avantages exclusifs aux clients inscrits sur l'espace My Renault.

La souscription à l'espace My Renault constitue le dernier point d'interaction avec le client dans la phase d'achat. Elle est prise en charge par le département CRM de l'entreprise qui a prévu à cet effet, plusieurs campagnes d'appel des clients qui sont effectuées par le centre d'appel afin de renseigner les informations personnelles des clients sur l'espace My Renault et de rappeler les nombreux avantages de leur souscription à ce service.

La cartographie présentée dans la figure II.10 montre les potentiels points d'interaction entre l'entreprise et le client/prospect avec l'entreprise tout au long de la phase d'achat d'un véhicule neuf.

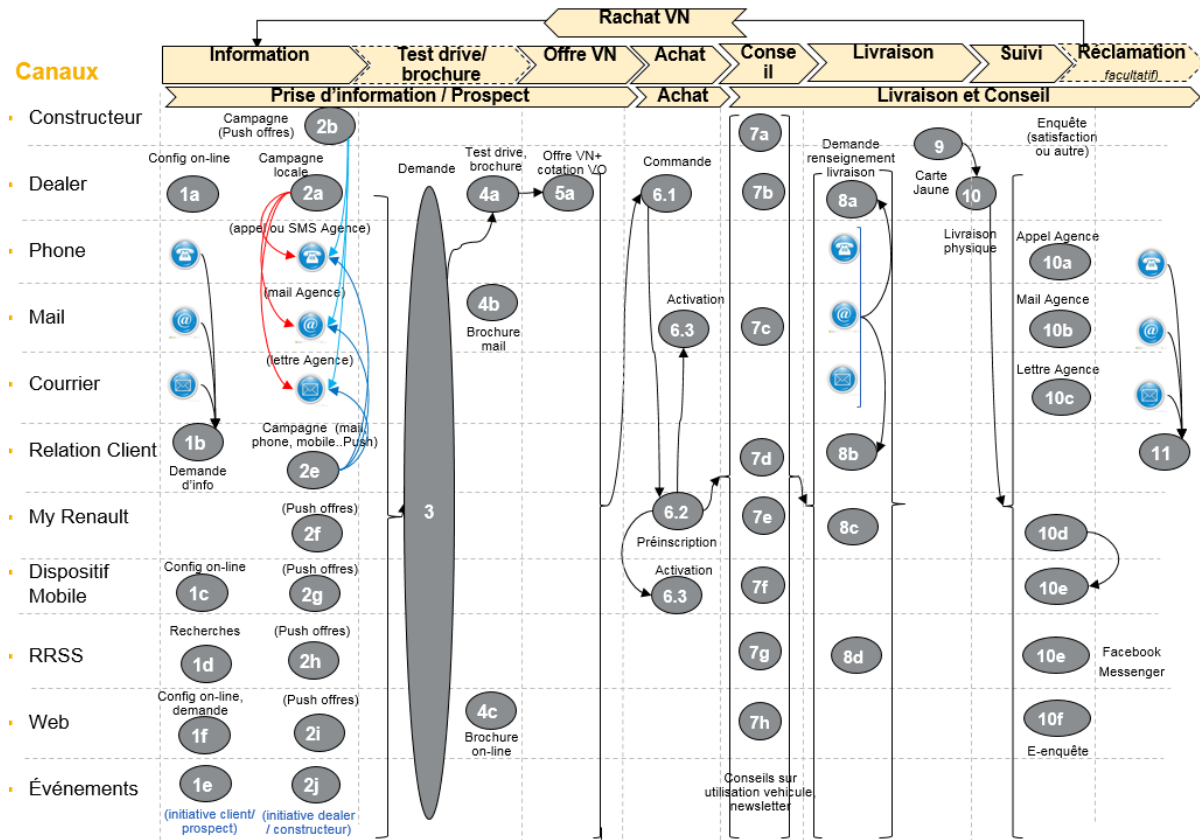


Figure II. 10: Cartographie des points d'interaction entre les clients et l'entreprise, durant la phase d'achat d'un véhicule neuf.

Phase 2 : Services après-vente :

La phase de services après-vente commence après la livraison du véhicule et l'interaction entre les deux parties est initiée soit par l'entreprise soit par le client lui-même.

Dans le premier cas, les conseillers client du centre d'appel contactent les clients après 3 mois de l'achat du véhicule afin de lui proposer la souscription à l'espace My Renault et lui proposer des accessoires, différents contrats de service, etc. Les clients inscrits sur My Renault reçoivent également des notifications relatives aux opérations de maintenance du véhicule ainsi qu'aux éventuels problèmes techniques qu'il pourrait rencontrer.

Dans le deuxième cas, le client contacte l'entreprise en passant par les réseaux sociaux, le site web de la marque ou directement, par le centre d'appel afin de se renseigner sur les différents services après-vente proposés par l'entreprise.

Lors des passages à l'un des ateliers des agents agréés, des informations sur les clients et leurs véhicules sont collectées dans les systèmes de gestion des vendeurs. Ainsi, en plus des données relatives à l'opération de maintenance ou de réparation réalisée, l'agent renseigne le kilométrage du véhicule et actualise les informations personnelles du propriétaire du véhicule.

Toutefois, ces données ne sont remontées à l'entreprise que depuis la mise en place du système d'information DMS, littéralement Dealer Management System, qui est actuellement en cours de déploiement.

Par ailleurs, les enquêtes de satisfaction ainsi que les différentes campagnes marketing réalisées constituent un autre moyen de collecte de données qui permet d'actualiser les informations enregistrées sur les clients.

La cartographie présentée dans la figure II.11 montre les potentiels points d'interaction entre les clients et l'entreprise tout au long de la phase du service après-vente.

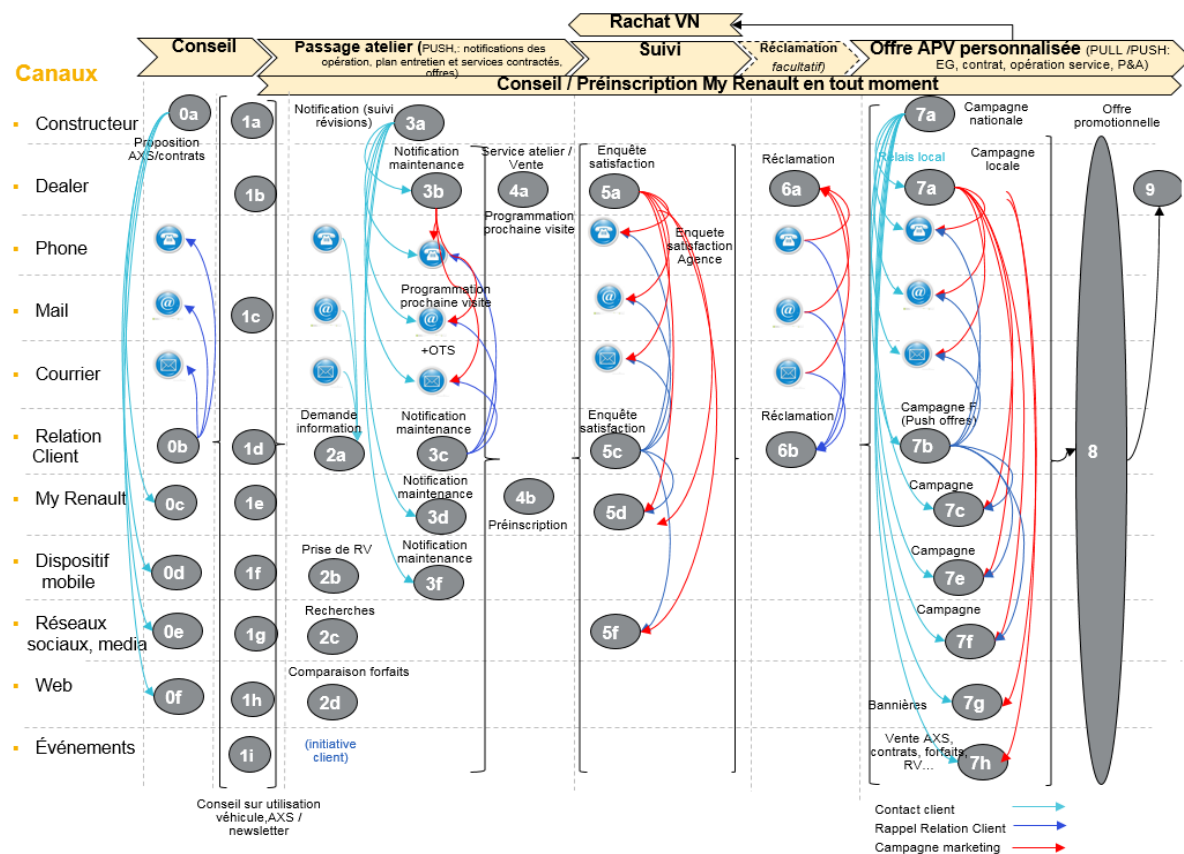


Figure II. 11: Cartographie des points d'interaction entre les clients et l'entreprise, durant la phase d'achat des services après-vente.

2. Système d'information client (SIC)

Renault Algérie, étant une filiale du groupe Renault, dispose de la même architecture des systèmes d'information, centralisée, que celle des autres filiales du Groupe.

Le système d'information client (SIC) actuellement mis en place au sein de Renault Algérie, est résumé dans la figure II.12.

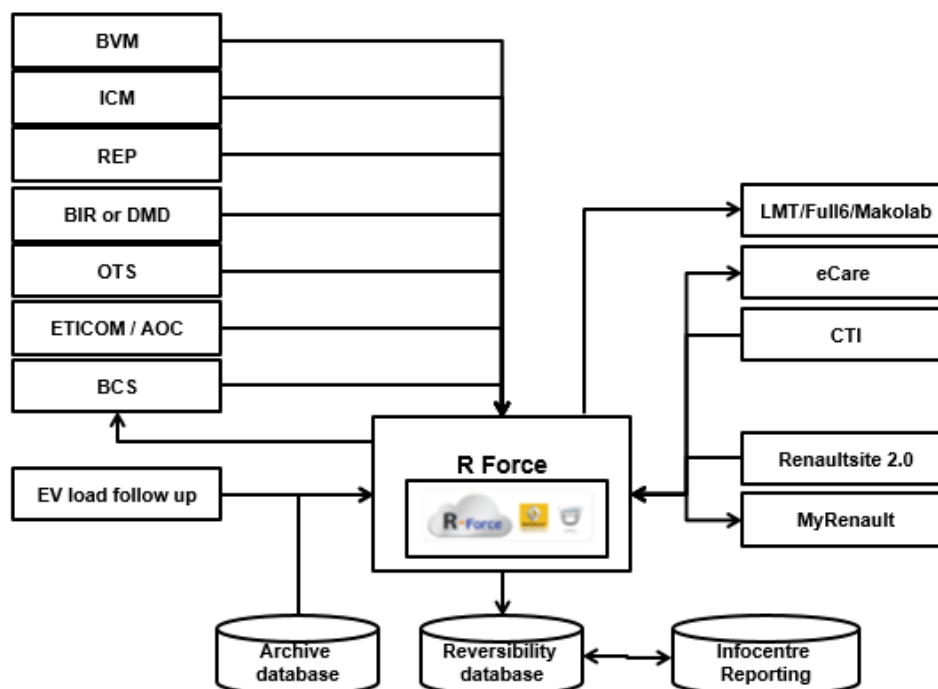


Figure II. 12: Architecture du système d'information client de Renault Algérie (Source : Document interne).

Le SIC est composé de systèmes d'information centralisés au niveau du Groupe Renault. Chaque système est dédié à un processus métier et relié au système d'information R-Force¹⁵.

2.1. R-Force

R-Force est un système d'information de gestion de la relation client qui a été spécialement conçu pour Renault par le leader des systèmes d'information CRM Sales Force.

R-Force représente le cœur du SIC car il est en interface avec tous les autres systèmes d'information relatifs aux véhicules, à la base de données clients et prospects ainsi qu'au web-service. Il permet surtout d'actualiser les données de la base client (BCS).

2.2. Base Client Simplifiée (BCS)

BCS est la principale base de données des clients de l'entreprise. Elle regroupe l'ensemble des données relatives aux clients allant des coordonnées, dates de naissance, informations sur l'achat, caractéristiques du véhicule acheté, etc. Il s'agit également de la seule base de données en interface en mode actualisation avec le système d'information R-Force.

Il est à noter le système d'information de la gestion des vendeurs DMS est en cours d'implémentation. Son interconnexion avec le SIC qui permettra d'actualiser les données BCS à chaque entrée atelier est prévue pour la fin de l'année 2016.

¹⁵ L'annexe 3 présente la description des différents systèmes d'information, leur rôle ainsi que leur connexion avec les autres systèmes.

La figure II.13 résume les interactions de Renault Algérie avec ses clients tout le long de leur cycle de vie au sein de l'entreprise.

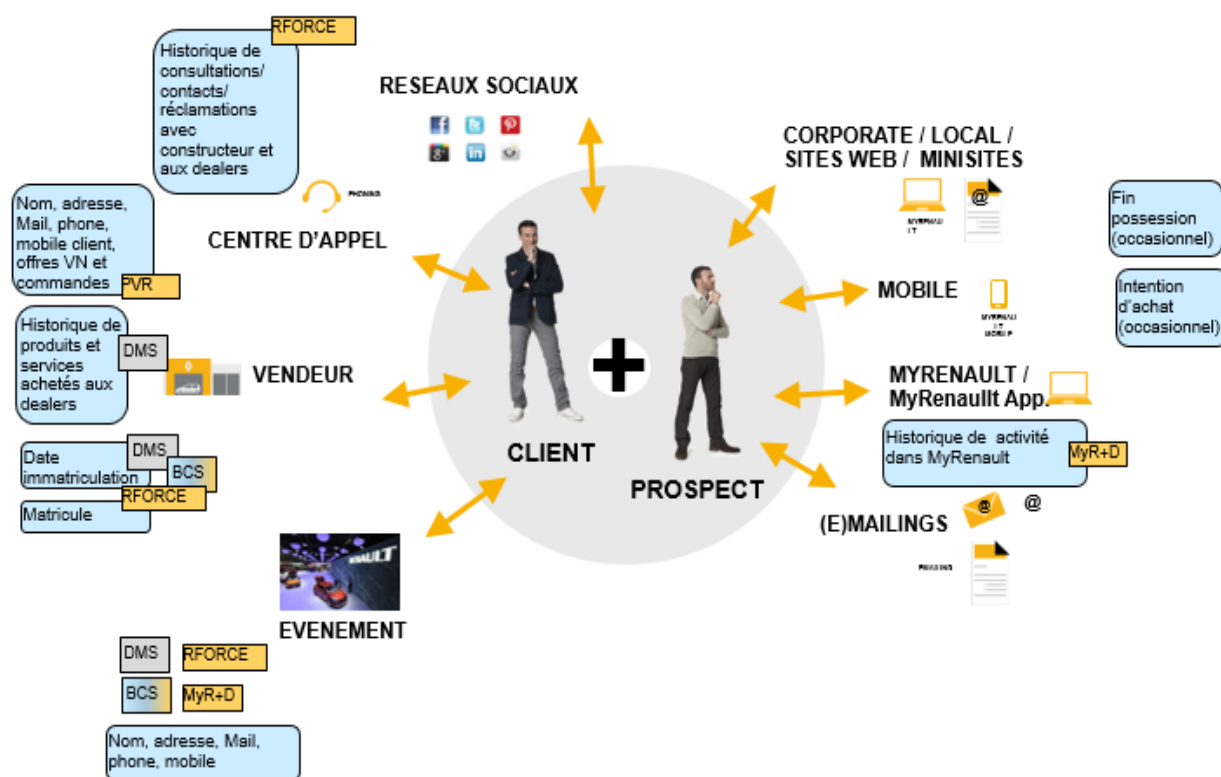


Figure II. 13: Modélisation des interactions entre l'entreprise Renault Algérie et ses clients/prospects.

3. Diagnostic de la base de données clients : BCS

3.1. Connaissance de la base de données :

La base BCS constitue le Data Mart du SIC. Elle contient exactement 104 champs d'information sur les clients, qui peuvent être regroupés en 3 dimensions, comme suit :

- La dimension « Sociodémographique »

Cette dimension est constituée de 74 attributs concernant les informations générales sur le client telles que son nom, prénom, date de naissance, profession, adresse, coordonnées, accord pour le contacter, langue de communication, etc.

Cependant, seules les informations nécessaires à l'établissement de la carte jaune sont renseignées, c'est-à-dire le nom, prénom, sexe et adresse du client.

Par ailleurs, les différentes campagnes marketing réalisées auprès des clients permettent d'actualiser et de renseigner davantage la base de données BCS.

- La dimension « Relation avec l'entreprise »

Cette dimension, comme son nom l'indique, concerne les informations relatives à la nature de la relation entre le client et l'entreprise. Elle comprend 14 attributs tels que le type de relation, la date de début et de fin de la relation, la date de commande et la date de livraison, le mode de paiement, etc.

- La dimension « Véhicule »

La dimension véhicule concerne l'ensemble des caractéristiques relatives au véhicule du client. Parmi les 16 attributs qu'elle contient, nous pouvons citer le numéro de châssis du véhicule noté VIN, le modèle, la version ainsi que l'année d'immatriculation.

Cette partie de la base de données BCS ¹⁶est totalement renseignée sauf pour l'année d'immatriculation.

Dans le but d'analyser l'intégralité de la base client BCS, c'est-à-dire, à partir de la date de la mise en place du système en 2010, une extraction de cette dernière nous a été fournie par le département CRM, le 15 Avril 2016.

A cette date, la base contient plus de 478 000 clients des marques Renault et Dacia et chacun des clients est identifié par une clé unique notée : ID Client.

La répartition des clients par marque est présentée dans le tableau suivant :

Tableau II. 1 : Répartition des clients BCS par marque

Marque	Nombre de client	Pourcentage
Renault	298 389	62,3 %
Dacia	180 539	37,7 %

Une analyse détaillée de la base de données nous permet de détecter les anomalies suivantes :

- Il existe de nombreux décalages des lignes par rapport aux champs adéquats, comme le montre la figure II.14 suivante :

¹⁶Nous détaillons dans l'annexe 4 le dictionnaire de données de la base BCS.

CHAPITRE II: Présentation de l'entreprise et Exploitation de la base de données clients.

1	Agreed com	Agreed com	Communicat	Email comm	Action Type	Acro Row_id	Relationship	Relationship	Relationship	New/Used	Annual mile	Annual mile	Current mile	Current mile	Salesman	Payment me	Ord
2	Marque s	Media cor	Date MAJ	Accord Cl	Mode de	Row_id cl	Type de r	Date de d	Date de fi	VN/VO	Km annue	Date de d	Km cumul	Date de r	Nom ven	Identifiari	Date
8004	Y	RENAULT	ALL	20120901	Y	C	1-49QU-511	1	20100326	20120901	VN		0				
8249	Y	RENAULT	ALL	20141207	Y	C	1-TG40-126	1	20141207	20141207	VN		0		37225	20151220	
9872	Y	RENAULT	ALL	20120901	Y	C	1-9EEJ-83	1	20100609		VN		0				
9888	N	RENAULT	ALL	20120116	N	C	1-BJLK-88	1	20120116		VN		0	61007	20140130		
0153	Y	RENAULT	ALL	20150430	Y	C	1-W7B5-634	1	20150430		VN		0				
1465	Y	RENAULT	ALL	20130805	Y	C	1-JFTF-31	1	20130805		VN	18000	20140331	12099	20140331		
1478	Y	RENAULT	ALL	20120901	Y	C	1-BLIP-16	1	20120119		VN		0				
2754	Y	RENAULT	ALL	20120901	Y	C	1-D8U3-517	1	20120607	20121201	VN		0				
2940	Y	RENAULT	ALL	20151223	Y	C	1-10LU6-457	1	20151223		VN		0				1954
3224	Y	RENAULT	ALL	20120901	Y	C	1-9NFO-31	1	20110714		VN		0				
3439	Y	RENAULT	ALL	20120901	Y	C	1-53KR-125	1	20100711	20120901	VN		0				
3633	Y	RENAULT	ALL	20110418	Y	C	1-6G9B-9	1	20101205		VN		0				
3812	Y	RENAULT	ALL	20100829	Y	C	1-5ITA-10	1	20100829		VN		0				
4101	Y	RENAULT	ALL	20120901	Y	C	1-4NXC-456	1	20100520		VN		0	88000	20130804		
4165	N	RENAULT	ALL	20120920	N	C	1-E3KC-35	1	20120920		VN		0				
4508				20	20150622		20	20150622	M	N	MAGHNI		12		TLEMEN	DZ	
4509				20	20110825		30	20130408	M	Y	OUED GOUCINE CENTRE				OUED GOUS	DZ	
4510				20	20130417		20	20140127	M	Y	CITE PALESTINE NÂ°05				RELIZANE	DZ	
4511				0	20100331		20	20120710	M	Y	ALGERIE						ALG
4512				20	20140625		30	20140625	M	N	07 CITE HOUARI BOUMEDIANE MAGHNI		2		TLEMEN	DZ	
4513				20	20131201		30	20141028	M	N	AIN DEFLA				AIN DEFLA	DZ	
4514				20	20100327		20	20100327	M	N	CITE DIAB EL AGIA BTB PAGE A NÂ°16 KOUBA				AIN DEFLA	DZ	

Figure II. 14: Exemple de décalage des lignes dans la base BCS.

Ces décalages sont issus d'une mise à jour des données clients, supposée manuelle. De plus, les sources des décalages sont clairement identifiées et présentées dans le tableau suivant :

Tableau II. 2: Source et pourcentage des décalages des champs de la base BCS.

Code opérateur	Pourcentage de décalages
20	4.96 %
30	12.6 %
31	49.91 %
74	16.28 %
76	16.25 %
Total	100 %

- Les attributs Profession et Nombre d'enfants ne sont renseignés pour aucun client.
- La majorité des numéros de téléphone et les adresses e-mails renseignés lors de la prise de commande sont erronés. Par exemple, des adresses e-mail sont renseignées sous le domaine « @client-renault.com » et des numéros de téléphone sont enregistrés comme suit : 0770000000 ou encore 213023000000, etc.
- Les champs Ville et Code postale sont non normalisés et comprennent de nombreuses erreurs.

- Plusieurs doublons, c'est-à-dire des observations présentant un ID Client et un code VIN identiques (même personne et même châssis du véhicule), sont identifiés.
- Des attributs relatifs au véhicule du client, tels que la motorisation, présentent des taux de remplissage inférieurs à 70%.

3.2. Nettoyage et préparation des données :

Afin de pouvoir utiliser les données de la base de données clients¹⁷, il est nécessaire de corriger les erreurs et anomalies relevées précédemment.

Pour ce faire, nous avons procédé au nettoyage de la base de données, à l'aide de différentes macros sur Excel. Il en résulte les actions suivantes :

- Plus de 2 200 lignes décalées, soit un total de 14 680 champs, sont corrigées.
- 1904 doublons (présentant le même ID Client ET le même numéro de châssis VIN) sont identifiés et supprimés.
- Plus de 280 000 numéros de téléphones présentés sous un format erroné sont identifiés et supprimés.
- 70 % des adresses e-mails du type « @client-renault.com » sont identifiées et supprimées.
- Plus de 1500 dates de naissance incorrectes (format erroné, dates improbables, etc.) sont éliminées.
- Une normalisation de l'ensemble des codes postaux est réalisée.
- Une normalisation et un remplissage des lieux de domiciliation des clients sont réalisés à partir des codes postaux, ce qui a permis de renseigner 6 500 champs vides.
- Un remplissage de l'ensemble des champs vides de l'attribut motorisation est réalisé, soit 35% de la totalité des champs, à partir des codes versions des véhicules.
- Un remplissage de l'ensemble des champs vides de l'attribut type du véhicule (Particulier/Utilitaire) est réalisé, soit 43% de la totalité des champs, à partir de l'attribut modèle du véhicule.

¹⁷ Une analyse descriptive des données nettoyées de la base BCS est réalisée et présentée en annexe 5.

4. Exploitation de la base de données clients

Le diagnostic de la base de données clients de l'entreprise, précédemment présenté, nous permet de constater que les données disponibles sont insuffisantes pour la segmentation efficace des clients de l'entreprise en fonction des objectifs fixés pour notre projet d'étude.

Par exemple, de nombreuses données sociodémographiques telles que la situation familiale, la profession ou le revenu des clients ne figurent pas dans la base BCS. De plus, cette dernière ne nous fournit aucun renseignement sur le comportement d'achat des services après-vente (sachant que le système DMS est en cours de mise en œuvre), ni sur les besoins et préférences des clients par rapport à l'automobile.

Néanmoins, des informations intéressantes sur le comportement d'achat de véhicules neufs des clients telles que le modèle choisi, la fréquence d'achat et la date de la transaction, peuvent être extraites à partir de cette base.

Ainsi, en nous basant sur les résultats du diagnostic BCS ainsi que sur la revue de littérature sur les différents types de segmentation des clients présentée dans le premier chapitre, nous constatons que la segmentation la plus adéquate en fonction des données disponibles est une segmentation basée sur le comportement d'achat des clients.

4.1. La récolte et la sélection des données

Les données nécessaires à la segmentation des clients de l'entreprise en fonction de leur comportement d'achat de véhicules neufs sont celles relatives aux clients. C'est donc pour cela que nous utilisons les données de la base client BCS pour notre étude.

Les attributs que nous retenons pour l'étude et qui sont jugés intéressants pour la segmentation sont les suivants :

- ID Client : l'identifiant du client.
- Sexe : le genre du client.
- VIN : le numéro de châssis du véhicule.
- Marque : la marque du véhicule.
- Modèle : code du modèle du véhicule.
- Energie : motorisation du véhicule.
- Date de commande : la date de commande du véhicule.
- Code postale : code postale de l'adresse de domiciliation de client.

Note : Le reste des données démographiques telles que la date de naissance des clients, ne sont pas retenues car leur taux de remplissage est jugé trop bas.

Cependant, nous constatons qu'une variable importante ne figure pas dans la base de données clients, il s'agit du montant de l'achat.

En raison de l'importance de cette variable dans la segmentation des clients, nous décidons d'enrichir notre base de données, jusque-là exclusivement constituée des données BCS, avec des attributs de la base de données du système local de facturation des véhicules neufs, appelée « INCADEA ».

Cette base de données comporte toutes les transactions monétaires des clients concernant l'achat de véhicules neufs, et ce, depuis Novembre 2012.

Nous retenons, de la base de facturation INCADEA, les attributs suivants :

- VIN : Numéro de châssis du véhiculé (servant de clé).
- Date de comptabilisation : Date de facturation du véhicule.
- Montant : valeur monétaire perçue par l'entreprise.

Avant de fusionner les deux bases de données préalablement sélectionnées, nous devons nous assurer que la clé utilisée est identifiée de manière unique dans les deux bases de données.

Etant donné qu'INCADEA est une base de données représentant les transactions réalisées, le problème de doublons du code VIN ne se pose pas. En revanche, dans la base BCS, nous remarquons que plusieurs ID_Clients peuvent tout à fait avoir le même code VIN, c'est en fait le cas lorsque le véhicule acheté neuf est ensuite vendu à un autre acheteur qui devient inscrit dans la base de données clients.

Pour remédier à ce problème, nous décidons de ne garder que les clients ayant acheté le véhicule neuf à l'entreprise et ce, en sélectionnant uniquement les clients avec la date d'achat (Date de comptabilisation) la moins récente, c'est-à-dire les premiers acquéreurs du véhicule.

4.2. Préparation des données

Etant donné qu'un nettoyage préalable de la base BCS a été réalisé et que les données de la base de données des facturations, INCADEA, sont déjà normalisées, l'étape de préparation des données se résume en la fusion des deux bases déjà sélectionnées, en un seul fichier SPSS.

Le tableau suivant résume le résultat de la fusion des deux bases de données.

Tableau II. 3: Récapitulatif de la fusion des bases de données BCS et INCADEA.

Base de données	BCS	INCADEA	Fusion BCS et INCADEA
Nombre de clients sans doublons	474 765	333 214	272 698
Pourcentage des doublons éliminés	0,82 %	0 %	0,5 %
Pourcentage des clients restants après la fusion	57,43 %	81,84 %	/

4.3. Transformation des données :

La première transformation des données réalisée concerne l'agrégation monovariante de l'attribut Code postal, afin de réduire les modalités de la variable Région qui sera utilisée.

Ces données sont agrégées en fonction de la segmentation régionale des concessionnaires, établie par le Service Réseau Commercial de l'entreprise.

Cette segmentation regroupe les différentes wilayas du territoire national en quatre principales régions : Nord, Est, Ouest et Sud.

Une nouvelle agrégation monovariante est réalisée, cette fois-ci sur l'attribut Modèle. En fait, l'attribut est représenté par un code du modèle de la voiture (par exemple, le code « 46L » désigne la Renault Symbole). En raison de la multitude des véhicules commercialisés par l'entreprise, nous décidons de réduire leur nombre en utilisant la segmentation véhicule utilisée par l'ensemble des constructeurs automobiles en Algérie.

De plus, afin d'analyser le comportement de rachat de véhicule neuf, nous avons créé une nouvelle variable binaire pour chaque modèle, qui prend la valeur 1 si le client a déjà acheté une voiture du modèle et 0 sinon.

Enfin, pour optimiser la segmentation comportementale des clients de l'entreprise, une agrégation RFM est apportée aux données.

Les transformations sur les lignes et les variables de la nouvelle base sont effectuées sur le logiciel SPSS Modeler, comme l'indique le flux de la figure II.15 suivante :

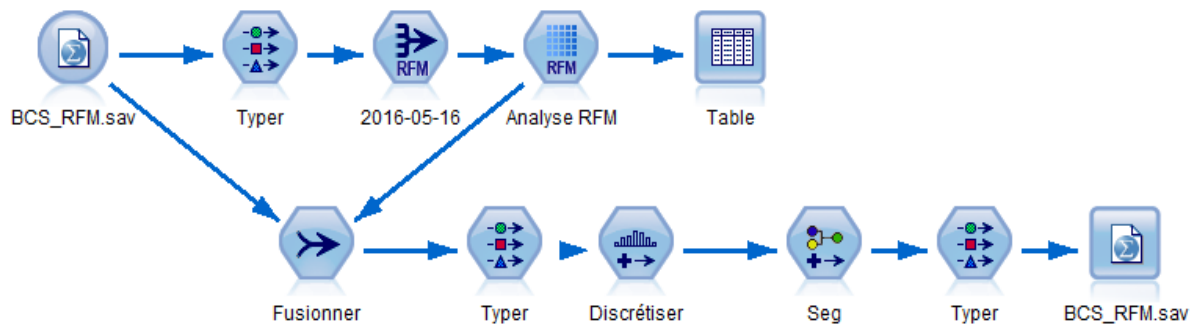


Figure II. 15: Flux du modèle de transformation de données sur SPSS Modeler.

En fait, le nœud Agréger RFM du logiciel IBM SPSS Modeler, permet de réduire les données historiques des transactions des clients, de supprimer les données inutiles et de combiner toutes les données de transactions de chaque client dans une seule ligne.

Avant d'aborder la constitution du score RFM, il convient de détailler l'élaboration de chaque variable nécessaire à l'élaboration de ce score. Pour ce faire, nous présentons la méthodologie adoptée pour l'analyse RFM des clients.

En fait, chaque ligne de la base de données contient les champs relatifs à un client. Les données ont déjà été combinées via des transactions.

Date ou intervalle de la transaction, Nombre de transactions, Total, les clients sont regroupés indépendamment par récurrence, fréquence et valeur monétaire, affecter les ex aequo au hasard dans les classes adjacentes.

Il existe 2 principales méthodes de regroupement :

- Imbriquée : Dans le regroupement par casier imbriqué, un score de récurrence est affecté en premier. Ensuite dans chaque rang de récurrence, les clients reçoivent un score de fréquence et dans chacun, les clients obtiennent leur rang monétaire. L'avantage est de distribuer équitablement les clients sur tous les casiers de regroupement.
- Indépendante : les scores sont affectés aux valeurs de récurrence, de fréquence et monétaires. Ces trois scores sont affectés de manières indépendantes. L'interprétation de chacun des trois composants RFM est donc claire.

Dans notre cas, nous avons décidé d'écarter la méthode imbriquée car elle a pour inconvénient de compliquer l'interprétation des scores de rang de fréquence et monétaire. Par exemple, un client dont le score de fréquence est de 5 dans le groupe de récurrence du score de 5 risque d'avoir une signification autre qu'un client avec un score de fréquence de 5 avec un score de récurrence de 4, puisque le rang de fréquence dépend de celui de récurrence. Par ailleurs, la taille de l'échantillon réduit le risque d'entraîner une distribution moins équitable des scores RFM combinés.

Le nombre de scores RFM combinés possibles est le produit des trois scores.

En cas d'ex aequo c'est-à-dire que deux ou plusieurs valeurs de score RFM combinés sont égales, nous avons choisi de gérer les ex aequo en les affectant aléatoirement aux casiers adjacents, ceci permet de garantir une distribution équitable des casiers en affectant un facteur de variance aléatoire très faible avant tout classement.

Ce processus n'a aucune incidence sur les valeurs d'origine, il n'est utilisé que pour départager entre les ex aequo.

Pour indiquer la dernière date d'achat (Récence), le nombre de transactions effectuées (Fréquence) et la valeur totale des transactions (Monétaire), nous utilisons l'ID Client.

Pour cela, nous définissons la date de calculer la récence par la dernière date d'extraction de la base INCADEA c'est-à-dire le 16 – 05 –2016. Ensuite nous sélectionnerons L'ID client comme ID pour identifier le client et ses transactions, la variable Date de comptabilisation dans le champ date pour calculer la récence et enfin la variable Montant dans le champ valeur pour calculer la valeur monétaire totale des transactions du client.

Après avoir identifié et transformé les données en utilisant le nœud Agréger RFM, nous utiliserons le nœud Analyse RFM pour analyser les nouvelles variables d'agrégation.

Le nœud Analyse RFM permet de déterminer de façon quantitative le potentiel des clients par l'étude de leur dernier achat (Récence), le nombre de leurs achats (Fréquence), et la somme dépensée lors de toutes les transactions (Monétaire).

Les données clients catégorisées sont divisées en un nombre d'intervalle que nous avons fixé comme suit : 5 casiers pour les variables Récence et Monétaire, et 3 casiers pour la variables Fréquence, ceci est justifié par le cycle long de renouvellement d'un véhicule neuf et la récence de la base de données.

Il est important de noter que le nœud Analyse RFM est configuré pour utiliser la création d'intervalles indépendants, de tel sorte qu'il classe et espace les données sur chaque mesure sans tenir compte de la valeur ni des deux autres mesures.

Dans chaque intervalle un score est attribué aux clients, en cas d'ex-aequo le score est conservé dans le même casier pour ne pas biaiser la comparaison entre les clients et le seuil des casiers est toujours recalculé.

Ces scores sont combinés pour calculer un score RFM global. Nous nous sommes entretenus avec les chefs produit de la marque Renault et Dacia, concernant la pondération pour calculer le score global RFM, nous avons conclu à la fin d'une séance de travail d'avoir un score qui peut être lu facilement et qui leur permet d'identifier le casier d'appartenance du client.

La formule de calcul du score global RFM convenue est la suivante :

$$RFM\ Score = Monétaire * 100 + Fréquence * 10 + Récence * 1$$

Bien que l'analyse et le classement des scores RFM représente un outil très pratique, il faut cependant garder à l'esprit que certains facteurs peuvent influencer la répartition des scores RFM sur les 75 casiers possibles et qui peut biaiser l'interprétation de ce score.

CHAPITRE II: Présentation de l'entreprise et Exploitation de la base de données clients.

Afin d'éviter certaines disproportions et des confusions dans l'interprétation du score RFM. Il est impératif de procéder à une normalisation de la distribution, elle sert à obtenir un ordre de grandeur comparable.

Pour ce faire, nous utilisons le nœud Discrétiser qui permet de générer automatiquement des catégories sur un champs de variable continue selon une méthode de regroupement choisie.

Nous optons pour une méthode de regroupement par casier à largeur fixe, elle permet de générer des intervalles de même taille, selon le nombre de casier choisi.

Dans notre étude, nous avons fixé un nombre de 10 casiers en recalculant à chaque fois les seuils des casiers.

La discrétisation de la variable RFM score, nous a permis de transformer les scores en une Nouvelle variable Score RFM_BIN, facilement comparable.

Le tableau suivant donne les seuils inférieurs et supérieurs de chacun des 10 casiers de regroupement.

Tableau II. 4: Seuils des casiers de regroupements pour la discrétisation du score RFM.

Casier	Inférieur	Supérieur
1	111	153.4
2	153.4	195.8
3	195.8	238.2
4	238.2	280.6
5	280.6	323
6	323	365.4
7	365.4	407.8
8	407.8	450.2
9	450.2	492.6
10	492.6	535

Par ailleurs, l'histogramme présenté dans la figure suivante, illustre l'effectif de chaque casier.

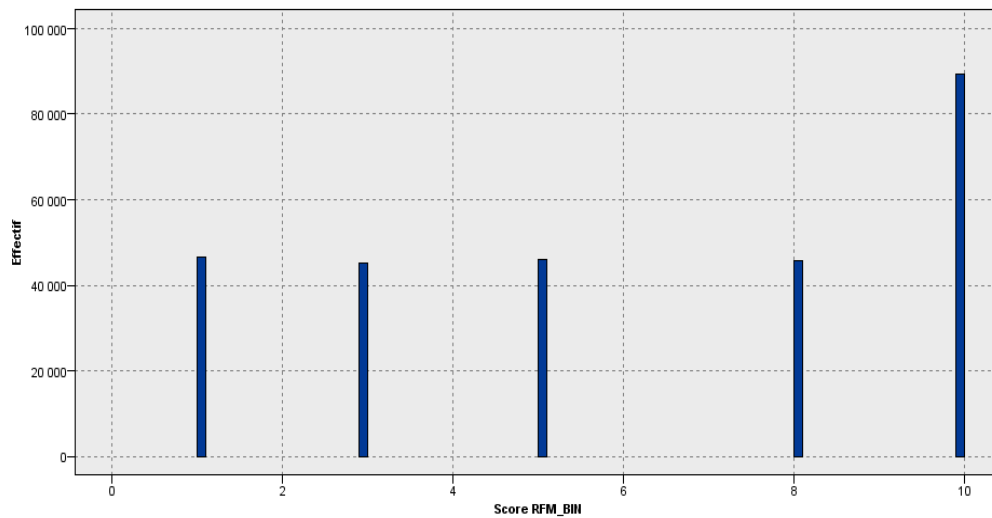


Figure II. 16: Histogramme des effectifs, issu de la discrétisation du score RFM.

Enfin, grâce au nœud Fusionner qui permet de créer à partir de plusieurs enregistrements d'entrée un seul enregistrement de sortie contenant la totalité des variables calculées, nous obtenons une table de données complète pour commencer à les analyser.

4.4. Data Mining :

Dans cette partie, nous utilisons des méthodes de classification et de classement pour construire un modèle qui nous permettra d'identifier les groupes de clients qui constitue le portefeuille de l'entreprise.

Nous détaillons dans la figure II. 17 la construction du modèle établi sur SPSS Modeler pour établir le profil des clients.

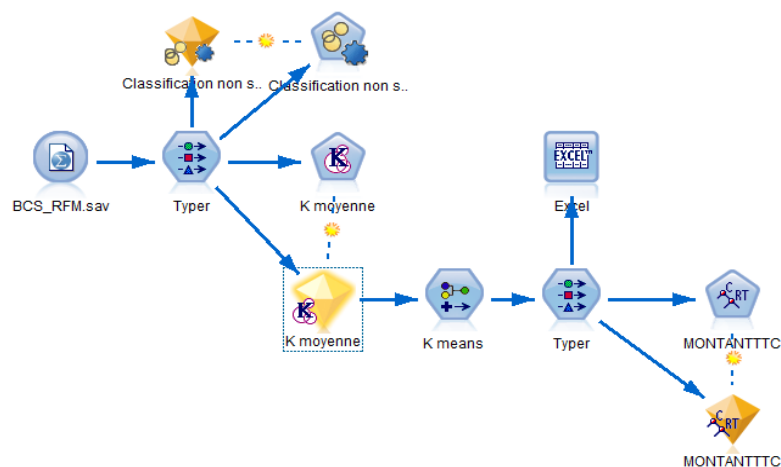


Figure II. 17: Flux du modèle Profil_client, sur SPSS Modeler.

a) Classification des clients :

Dans notre étude, nous avons utilisé le nœud de classification non supervisé qui traite les données sur 3 méthodes différentes qui sont : la méthode Two-step, la méthode des k moyennes et la méthode des réseaux de Kohonen.

Les données d'entrées utilisées dans ce nœud de classement non supervisé sont :

- Score RFM_BIN
- Energie
- Région
- Marque
- Segment

Les résultats obtenus sont résumés dans le tableau suivant :

Tableau II. 5: Résultats de la classification non supervisée pour les trois méthodes.

Modèle	Durée de création	Silhouette	Nombre de clusters
Two-Step	Plus de 5 minutes	0.224	6
K-moyenne	Plus de 5 minutes	0.293	5
Kohonen	Plus de 5 minutes	0.157	12

Nous observons que le modèle K-means donne la meilleure silhouette des trois modèles.

Cependant, le modèle Two-step permet grâce à sa configuration sur SPSS Modeler d'avoir le nombre optimal de cluster significatif par le calcul des critères Bayésien.

K -means est une méthode d'analyse de clusters, elle permet de classer les données en groupes distincts lorsqu'aucun groupe n'est défini au départ. Cette méthode d'apprentissage non supervisé tente de découvrir les tendances au sein des champs d'entrée.

Le modèle K moyenne s'avère souvent la méthode la plus rapide pour classer des jeux de données volumineux. (Guide SPSS Modeler 17,2015)

Par conséquent, nous avons choisi le modèle K-moyenne pour la cohésion intra-classes et pour sa capacité à traiter les données volumineuses, toutefois nous avons fixé le nombre de clusters à 6, le nombre optimal de cluster significatif donné par le modèle Two-Step.

b) Classement : Arbre de décision C&RT

L'objectif de l'utilisation d'un modèle de classement suite à la première classification est de conforter l'analyse et la description des clusters du premier modèle grâce au montant d'achat.

Les algorithmes d'arbre de décision divisent de manière récursive les données en sous-groupes de plus en plus petits. Le modèle de classification C&RT examine en premier lieu les champs d'entrée qui sont : le genre, la marque, la région et le segment afin de définir la meilleure segmentation.

Le découpage définit deux sous-groupes qui sont à leur tour découpez en deux nouveaux. Les résultats de cette classification donneront plus de détails sur la description des clusters de la classification K – means.

c) Prédiction des clients potentiels :

Dans cette partie, nous construisons un modèle prédictif pour estimer le potentiel de renouvellement d'achat d'un véhicule neuf. Les résultats de ce modèle permettront d'optimiser les campagnes marketing de conquête des clients pour l'achat de nouveaux modèles.

La prochaine campagne de conquête des clients effectuée par le centre d'appel concernera le lancement de la nouvelle Dacia Sandero StepWay Made in Bladi. Le modèle classera les clients selon une probabilité de renouvellement d'achat des clients par une Sandero StepWay.

Dans ce qui suit nous détaillons la construction du modèle sur SPSS Modeler.

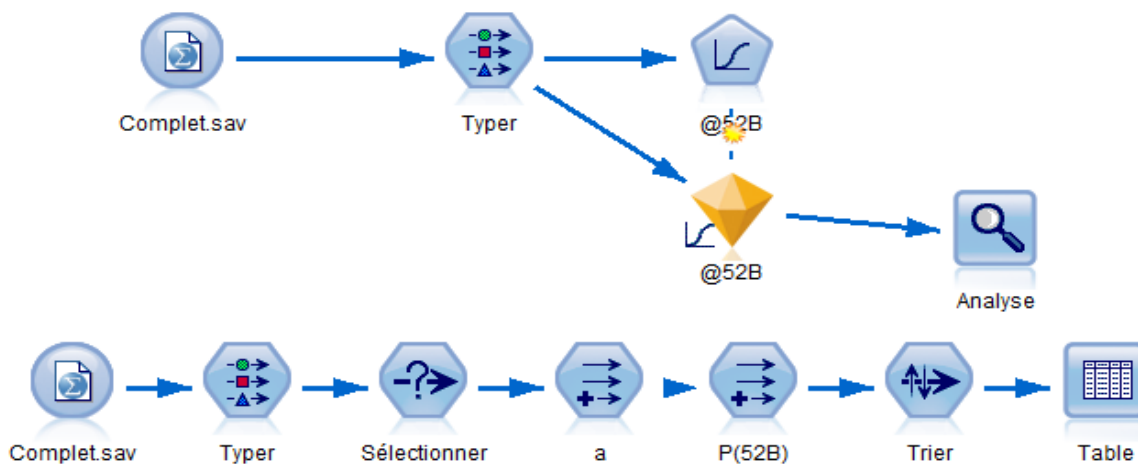


Figure II. 18: Flux du modèle prédictif, sur IBM SPSS Modeler.

Nous optons pour une régression logistique, car elle permet de créer une combinaison de valeurs qui associent les valeurs de champ d'entrée aux probabilités rattachées à chacune des variables de sortie.

Une régression logistique binomiale est appliquée dans ce modèle car la variable cible est une variable binaire (l'achat ou non du modèle 52B), en plus, trois variables d'entrée K-means, Nombre d'achat et le genre, sont pris en compte pour expliquer la variable cible.

Afin d'éviter toutes corrélations entre les variables constituant le modèle de la régression logistique, les variables d'entrées expliquant la segmentation de la variable K-means, ne sont pas pris en compte.

Le modèle de régression logistique reposera sur l'hypothèse fondamentale dite de l'évidence tel que :

$$\text{Ln} \left(\frac{P(52B|1)}{P(52B|0)} \right) = \text{Ln} \left(\frac{P(52B|1)}{1 - P(52B|1)} \right) = \sum_{i=0}^7 a_i x_i + c + \varepsilon(t)$$

Une fois le modèle créé, nous pouvons calculer une probabilité d'appartenance pour chaque client :

$$P(52B) = \frac{e^{a(52B)}}{1 + e^{a(52B)}}$$

$$\text{tel que } a(52B) = a_1 * x_{k\text{-moyenne}} + a_2 * x_{\text{Nombre d'achat}} + a_3 * x_{\text{genre}} + c$$

4.5. Résultats et interprétations

Rappelons que nous avons utilisé un modèle de classification k-means avec 5 variables d'entrées, à savoir, le score RFM discrétisé, l'énergie, la région, la marque et le segment du véhicule.

Un autre modèle de classement d'arbre de décision C&RT, a été généré pour décrire en détails les clusters résultant du K-means.

Les résultats statistiques de la classification K-means sont validés, avec une analyse ANOVA significative pour les 5 variables d'entrées.

La qualité des clusters est donnée par la statique silhouette qui mesure la cohésion et la séparation intra classe. Elle est de 0.3, un niveau correct de qualité des clusters obtenus.

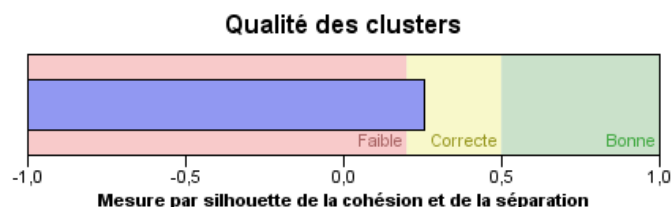


Figure II. 19: Résultat de mesure de la silhouette du modèle K-means sur IBM SPSS Modeler

CHAPITRE II: Présentation de l'entreprise et Exploitation de la base de données clients.

Le tableau suivant résume les résultats de contribution des variables d'entrées pour chaque cluster. Ces derniers ordonnés par importance de taille.

Tableau II. 6: Résultat de la contribution de chaque variable d'entrée dans la constitution des clusters

Cluster	Cluster-4	Cluster-2	Cluster-1	Cluster-3	Cluster-5	Cluster-6
Taille en pourcentage (effectif)	25,2 % (68 659)	19,3 % (52 719)	17,6 % (47 988)	17,1 % (46 677)	13,3 % (36 363)	7,4% (20 292)
Energie-Catégorie la motorisation (Pourcentage)	Essence (98,7 %)	Essence (99,5 %)	Essence (93,2 %)	Diesel (100%)	Essence (88 %)	Essence (90,3 %)
Marque- Marque du véhicule (Pourcentage)	Dacia (57,6 %)	Renault (59,3 %)	Dacia (100%)	Renault (99,3 %)	Renault (87,6 %)	Renault (88,6 %)
Région- Zone d'habitation du client (Pourcentage)	Est (100 %)	Ouest (65,2 %)	Centre (92,6 %)	Centre (56 %)	Centre (98,4 %)	Ouest (99,5 %)
Score RFM - Médiane	5	3	5	10	5	8
Segment – Segment du véhicule (Pourcentage)	M0 (61,4 %)	M0 (100 %)	M0 (59,4 %)	VU (62,2 %)	I (86,3 %)	I (99 %)

La figure suivante représente une comparaison visuelle par ordre de contribution des variables d'entrées pour chacun des clusters.

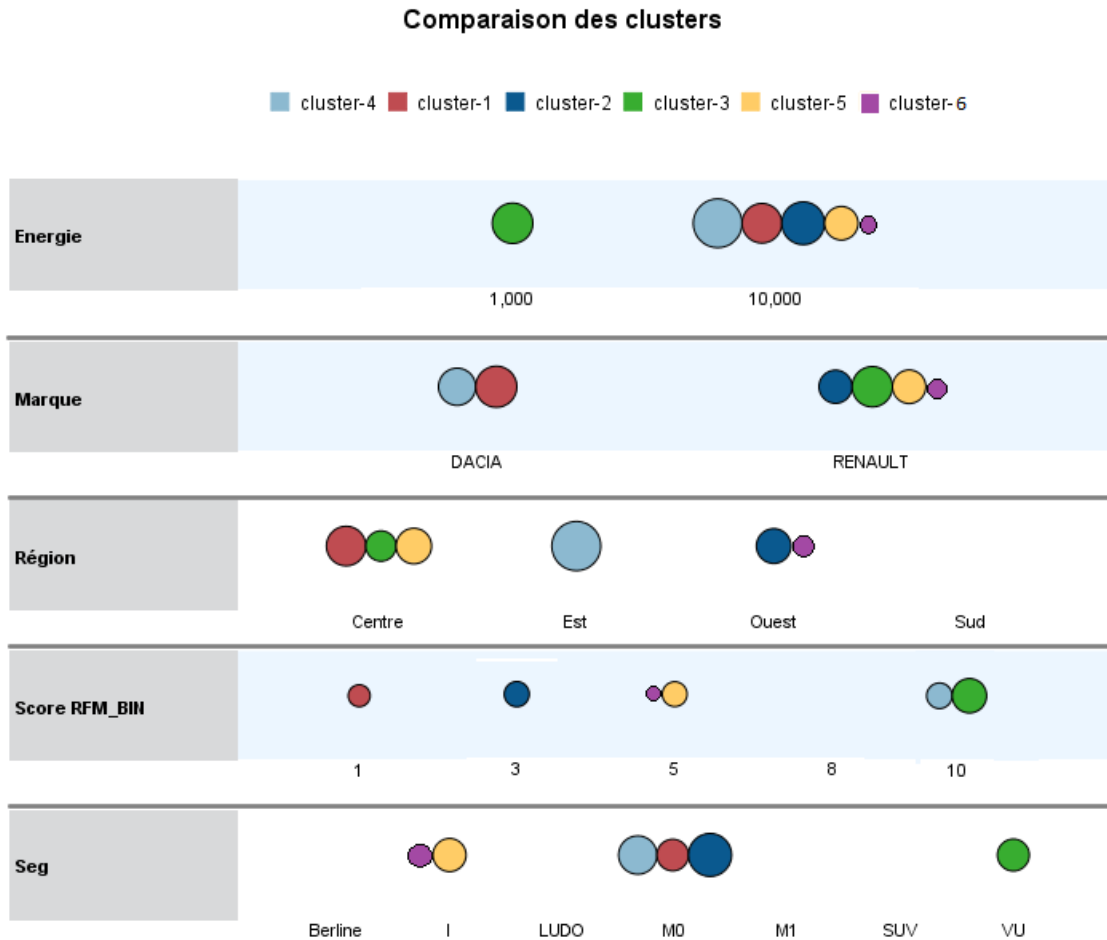


Figure II. 20: Comparatif des clusters de la classification k-means.

Ensuite, nous avons procédé à un classement par la méthode C&RT pour détailler encore plus et décrire au mieux les clusters obtenus par la classification k-means.

Le résultat du classement de la variable cible K-moyenne c'est-à-dire en le cluster auquel appartient chaque ligne en fonction des données d'entrée le genre, la marque, le segment, et le score RFM.

L'arbre de décision d'une profondeur de 7 est présenté dans la figure II.21.

On note que la variable "Genre" n'as aucune contribution dans la classification ni dans le classement, il n'est pas une variable discriminante dans le modèle et ceci peut s'expliquer par le faible taux de clientes, presque 9 % du total des clients.

\$KM-K moyenne

Noeud 0		
Catégorie	%	n
cluster-1	17,516	33430
cluster-2	19,319	36872
cluster-3	17,184	32798
cluster-4	25,210	48115
cluster-5	13,328	25437
cluster-6	7,444	14207
Total	100,000	190859

Région
Amélioration=0,197

Est

Noeud 1		
Catégorie	%	n
cluster-1	0,000	0
cluster-2	0,000	0
cluster-3	11,462	6229
cluster-4	88,538	48115
cluster-5	0,000	0
cluster-6	0,000	0
Total	28,473	54344

Seg
Amélioration=0,043

I; M0; SUV

LUDO; M1; VU

Noeud 3		
Catégorie	%	n
cluster-1	0,000	0
cluster-2	0,000	0
cluster-3	2,713	1338
cluster-4	97,287	47987
cluster-5	0,000	0
cluster-6	0,000	0
Total	25,844	49325

Noeud 4		
Catégorie	%	n
cluster-1	0,000	0
cluster-2	0,000	0
cluster-3	97,450	4891
cluster-4	2,550	128
cluster-5	0,000	0
cluster-6	0,000	0
Total	2,630	5019

Noeud 5		
Catégorie	%	n
cluster-1	41,543	29588
cluster-2	51,760	36865
cluster-3	2,430	1731
cluster-4	0,000	0
cluster-5	4,246	3024
cluster-6	0,021	15
Total	37,317	71223

Région
Amélioration=0,063

Ouest

Centre; Sud

Noeud 11		
Catégorie	%	n
cluster-1	5,945	1547
cluster-2	92,244	24002
cluster-3	1,752	456
cluster-4	0,000	0
cluster-5	0,000	0
cluster-6	0,058	15
Total	13,633	26020

Seg
Amélioration=0,016

Noeud 12		
Catégorie	%	n
cluster-1	62,033	28041
cluster-2	28,456	12863
cluster-3	2,821	1275
cluster-4	0,000	0
cluster-5	6,690	3024
cluster-6	0,000	0
Total	23,884	45203

Marque
Amélioration=0,086

Centre; Ouest; Sud

Noeud 2		
Catégorie	%	n
cluster-1	24,488	33430
cluster-2	27,009	36872
cluster-3	19,462	26569
cluster-4	0,000	0
cluster-5	18,633	25437
cluster-6	10,407	14207
Total	71,527	136515

Seg
Amélioration=0,118

Berline; I; LUDO; M1; VU

Noeud 6		
Catégorie	%	n
cluster-1	5,884	3842
cluster-2	0,011	7
cluster-3	38,041	24838
cluster-4	0,000	0
cluster-5	34,327	22413
cluster-6	21,736	14192
Total	34,210	65292

Seg
Amélioration=0,096

Berline; LUDO; M1; VU

Noeud 13		
Catégorie	%	n
cluster-1	9,133	3842
cluster-2	0,000	0
cluster-3	5,173	2176
cluster-4	0,000	0
cluster-5	52,262	21984
cluster-6	33,432	14063
Total	22,040	42065

Région
Amélioration=0,080

Noeud 14		
Catégorie	%	n
cluster-1	0,000	0
cluster-2	0,030	7
cluster-3	97,567	22662
cluster-4	0,000	0
cluster-5	1,847	429
cluster-6	0,555	129
Total	12,170	23227

CHAPITRE II: Présentation de l'entreprise et Exploitation de la base de données clients.

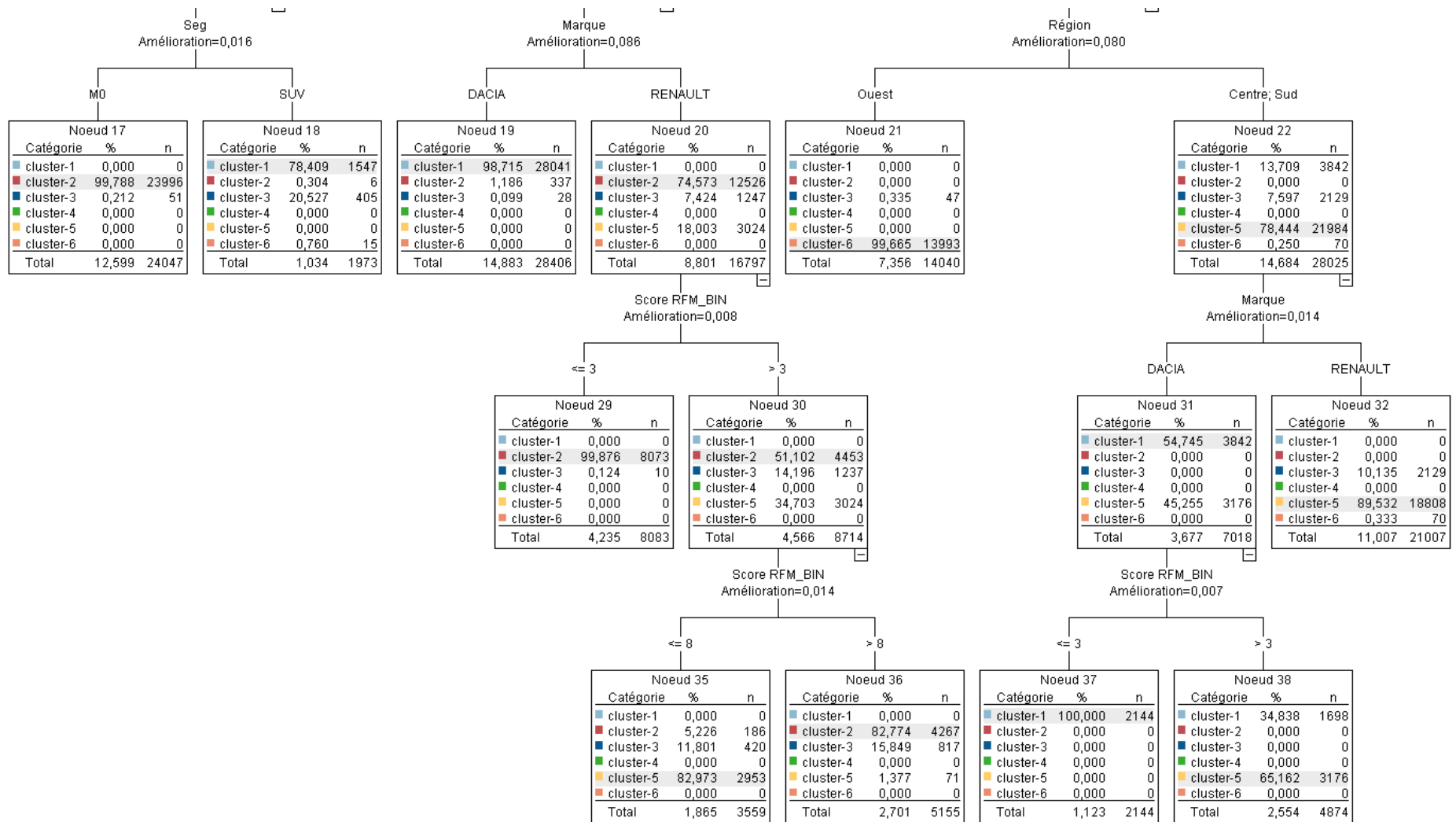


Figure II. 21: Arbre de décision généré par la méthode C&RT 1 sur IBM SPSS Modeler.

Une fois les résultats des différentes méthodes présentés, nous procédons désormais à l'interprétation de la classification obtenue.

Cluster 4 : (Client FAN marque Groupe Renault, Renouvellement +, Segment Mo, I, SUV, Est)

Ce cluster est constitué d'acheteurs de l'est de l'Algérie. Il représente 25 % du portefeuille des clients du Groupe Renault Algérie, avec 28.54 % du chiffre d'affaire total.

Il est composé majoritairement d'homme, environ 94 % des clients et ils sont âgés de 35 -54 ans, habitant majoritairement à l'est du pays avec une médiane du score RFM égale à 5.

Presque tous les véhicules achetés sont des véhicules particuliers de motorisation essence. La marque Dacia représente 57.6 du total des voitures vendues.

Les modèles du segments M0 sont les plus achetés avec 64,1 %, le plus haut pourcentage des 6 clusters avec un penchant pour la Logan de la marque Dacia avec 63 % du total des véhicules M0 de ce cluster.

Concernant le segment I, il représente presque 30 % des véhicules achetés dont deux tiers de la marque Renault avec les deux modèles Clio II et Clio IV. Le reste des véhicules vendus sont des SUV, majoritairement des Dacia Duster.

Pas moins d'un quart des acheteurs du segment M0 ont une récence de 1, c'est-à-dire qu'ils ont acheté un véhicule depuis 4 ans dont 84 % n'ont pas renouvelé d'achat de véhicule neuf chez les deux marques.

Néanmoins plus d'un tiers des ventes du segment M0 sont récentes dont 46 % ont acheté au moins un deuxième véhicule neuf chez le Groupe Renault.

22 % des véhicules du segment I ont été vendus cette dernière année dont 64 % de la marque Dacia dont 55 % ont renouvelé l'achat chez le Groupe Renault. Quant au SUV, 34 % des acheteurs ont une récence égale à 5 dont un tiers sont des anciens clients.

En résumé, ce cluster représente le profil typique des clients de l'est du pays, fans des marques du Groupe Renault sur les trois segments M0, I et SUV avec une préférence pour la marque Dacia. Au total, plus d'un tiers des clients ont acheté un véhicule cette dernière année de la marque Dacia et pour la deuxième fois au moins chez le Groupe Renault.

Cluster 2 : (Nouveau Client VS ancien, M0, Renouvellement -, Fréquence = 1, Majorité de l'Ouest)

Ce cluster très particulier est constitué de clients ayant effectué un achat d'un modèle du segment M0 dont 59.3 % de la marque Renault. Il représente 19.3 % des clients du Groupe Renault avec 22.7 % du chiffre d'affaire depuis 2013.

Les hommes représentent 95% de la population du cluster, majoritairement entre 35 et 54 ans, moins des deux tiers habitent à l'ouest du pays et les autres au centre.

Dans ce cluster il existe deux types de clients :

De nouveaux acheteurs de la Renault Symbol, plus de 35 % ont une récence de 5 dont 60% achètent pour la première fois chez Renault. Même cas pour la Dacia Logan, 30 % des clients ont une récence 5 et parmi eux 76 % pour la première fois. A eux deux ils représentent plus du quart du nombre des clients du cluster.

Le second type de clients sont des acheteurs d'ancienne version de la Renault Symbol et de la Dacia Logan. Il représente 19.75 % des clients du cluster dont 80 % n'ont pas renouvelé d'achat de véhicule neuf chez le Groupe Renault.

Finalement, le profil des clients du cluster deux sont à moitié de nouveaux clients du segment M0 séduit par l'offre d'entrée de gamme que propose Renault, l'autre moitié sont d'anciens clients eux aussi séduits par les anciens modèles d'entrée de gamme de Renault. Les points communs entre les deux types de profils, c'est qu'ils sont séduits par les véhicules du segments M0 mais qu'ils ont une faible fréquence d'achat, la majorité ont acheté un véhicule pour la première fois sans jamais renouveler de voiture neuve chez le Groupe Renault.

Cluster 1 : (Fidèle Dacia, Homme M0, I, SUV, Centre, Renouvellement ++ /Nouveau, Femme, SUV, Centre)

Ce cluster est constitué uniquement de client de la marque Dacia. Il représente 17.6 % du portefeuille client avec une part de 20.5% du chiffre d'affaire, avec une médiane du score RFM égale à 5.

Il est composé de 90 % d'hommes âgés entre 35 et 64 ans, et 10 % de femmes âgées entre 35 et 54 ans, habitant majoritairement au centre du pays.

Il est constitué uniquement de véhicules particuliers de la marque Dacia dont 93.4% de motorisation Essence. 60 % des modèles sont du segment M0, 30 % du segment SUV et le reste du segment I.

On note que plus de 25 % les clients du segments M0 et I sont de nouveaux acheteurs, avec un score de récence de 5, toutefois on remarque que 50 % de ces clients rachètent au moins pour la deuxième fois un véhicule de la marque Dacia, ce sont majoritairement des hommes.

Concernant le segment SUV, avec le modèle Dacia Duster, les hommes représentent 85% des clients. Moins du quart des ventes pour les hommes sont récent, parmi eux 24% sont des fidèles de la marque Dacia, avec des scores de fréquence égale à 2. Les femmes représentent 15%, majoritairement elles sont, pour la première fois, clientes de la marque Dacia.

En résumé, ce cluster représente les clients hommes fidèles de la marque roumaine sur les trois segments confondus où elle est présente. Il comprend aussi de nouvelles clientes de la marque essentiellement dans le segment SUV.

Cluster 3 : M1, VU (Monétaire ++, Récence -, Fréquence --) (Ludospace Fidèle ,Fréquence +)

Le cluster numéro trois comporte tous les modèles de motorisation Diesel vendu par le Groupe Renault, il représente 17.1 % des clients avec 14.6 % du chiffre d'affaire.

Ce groupe de client comprend 91 % d'hommes âgés entre 35 et 44 ans, plus de la moitié sont issus du centre du pays, 24 % de l'ouest et 19 % de l'Est.

Ce cluster comprend de trois segments de modèles à 99.3% de la marque Renault, tous les véhicules utilitaires vendus par le Groupe Renault et qui représente 62.2 % des modèles du cluster. Les 38.8 % restants représentent tous les véhicules particuliers des segments Ludospace et M1 (Compacte), ainsi les modèles des segments I avec une forte note RFM.

Ce cluster se caractérise par des notes RFM très élevées, car 89 % des clients ont un score monétaire de 5. Cependant la fréquence d'achat est la plus faible parmi les six clusters, environs 91% ont acheté un seul véhicule neuf, mais le fait le plus marquant c'est la récence, plus de 78 % des clients ont un score de moins de trois.

Les propriétaires des véhicules utilitaires sont les plus anciens du cluster. Parmi les 62.2 % des clients seulement 5.2 % ont acheté un deuxième véhicule utilitaire chez le Groupe Renault.

Malgré la faible proportion des clients du segment Ludospace, seulement 1.7 % du total des clients, la majorité des clients ont un score monétaire de 5. De plus, 50 % des clients ont une récence de 5, c'est-à-dire qu'ils ont effectué un achat durant cette dernière année, et parmi eux 30 % ont acheté pour la deuxième fois dans le même segment.

Les clients du segments M1 sont d'anciens clients, plus de 70 % ont un score de récence de moins de 3, et moins de 10 % des clients ont acheté un véhicule neuf pour la deuxième fois.

Pour conclure, ce cluster représente les profils des clients à fort potentiel monétaire mais à faible potentiel de renouvellement de véhicules pour les segments des véhicules utilitaires et celui de voitures compacts (M1) de véhicules particuliers.

Cependant les clients du segments Ludospace sont très fidèles à la marque et au segment, avec un fort potentiel monétaire et une fréquence de renouvellement élevée, pas moins du tiers des acheteurs récents ont renouvelé dans le même segment.

Cluster 5 : (Segment I, récence -, Fréquence 0)

Ce groupe de client est composé de 13.3 % de l'ensemble des acheteurs soit une part de 8.34% du chiffre d'affaire. Le cluster a la particularité d'être constitué à 86.3 % du segment I dont 87.6 % de la marque Renault et 88% avec une motorisation essence.

Le cluster est composé à 80% d'hommes âgés de 35 à 54 ans, et de 20 % de femmes âgées de 25 à 44 ans, la majorité des clients habitent au centre du pays. 89% des deux genres de clients ont un score monétaire entre 3 et 4.

Les hommes sont des clients plutôt récents, 29 % ont acheté un véhicule du segment I cette dernière année avec un score de récence de 5, plus d'un cinquième d'entre eux ont acheté un véhicule des deux marques confondues pour la deuxième fois.

Quant aux femmes, généralement c'est leur premier achat avec des scores de récence de 1 à 5 et de faible score de fréquence.

Donc, ce cluster est composé des clients du segments I, majoritairement de la marque Renault. Les profils hommes ont des achats plus récents dont une partie pour la deuxième fois, par contre c'est le premier achat pour les profils femmes.

Cluster 6 : (Homme, Ouest, segment I, Fidèle Renault, Ouest, Renouvellement ++)

Le dernier cluster et le plus petit avec 7.4 % des clients du Groupe Renault Algérie, il représente 5.32 % du chiffre d'affaire. Le cluster est constitué à 99.5 % de modèles du segment I dont 88.4 de la marque Renault et 90.3 % avec une motorisation essence.

Le groupe de clients est composé de 89% d'hommes âgés entre 35 et 54 ans et 11% de femmes âgées entre 25 et 44 ans, la totalité des clients est issue de l'ouest du pays.

Les récents achat de véhicules neufs du segment I effectués par des hommes sont moins récents, seulement 17 % ont un score de récence égale 5, mais avec la particularité d'avoir des scores de fréquence élevés car 50 % des nouveaux clients achètent pour la deuxième un véhicule du segment I.

Toutefois les femmes, ont des scores de récence anciens 68 % ont un score inférieur à 3 avec de faible score de fréquence. Les deux profils ont majoritairement des scores monétaires moyens, 91.4 % des profils ont un score monétaire entre 3 et 4.

En conséquent, les clients de la marque Renault du segment I de l'ouest du pays sont majoritairement moins récents, cependant la moitié des achats récents effectués par les hommes sont pour la deuxième fois dans le même segment I.

Par ailleurs, nous rappelons qu'un deuxième modèle, de prédiction cette fois ci, est utilisé dans le but d'évaluer la probabilité d'achat de la Sandero StepWay Made In Bladi de la marque Dacia, qui va être lancée dans les semaines qui suivent. Il s'agit en fait d'une régression logistique.

La méthode choisie pour la régression logistique binomiale est la méthode pas à pas montante, elle permet de sélectionner les variables explicatives par étapes. A chaque étape, les caractéristiques qui n'ont pas encore été intégrées au modèle sont évaluées et celles qui améliorent de manière significative la prédiction du modèle sont alors ajoutées au modèle.

Variable Cible : la variable binaire 52B

Variable explicative : K-means, genre, nombre d'achat.

Les résultats obtenus au bout du 3ème pas du modèle sont présentés dans le tableau suivant :

Tableau II. 7: Résultats de la régression logistique au 3ème pas.

Variables de l'équation	Colonnes	B	E.S	ddl	Sig.	Exp(B)
Pas 3c	Total					
	Achat	0,365	0,009	1	0	1,441
	Sexe	0,103	0,023	1	0	0,902
	K means			5	0	
	K means (1)	3,215	0,028	1	0	24,903
	K means (2)	-4,08	0,104	1	0	0,017
	K means (3)	0,013	0,501	1	0,63	1,013
	K means (4)	1,213	0,027	1	0,011	3,364
	K means (5)	0,289	0,029	1	0	1,335
	K means (6)	-0,19	0,029	1	0	0,827
	Constante	1,064	0,033	1	0	0,345

On observe que toutes les variables sont significatives, sauf la variable d'appartenance du cluster 3 qui n'est pas explicative et significative dans ce modèle (Sig. 0.63 > 0.05)

Afin de valider notre modèle, nous évaluerons sa capacité à générer des prévisions précises, pour cela nous utilisons le nœud Analyse sur SPSS Modeler qui comparera les valeurs prédites avec les valeurs réelles de notre modèle.

Les résultats obtenus montrent que le modèle prédit à 95.52 % des valeurs correctes pour la prédiction de l'achat d'une Sandero StepWay.

Résultats du champ de sortie @52B

Comparaison de \$L-@52B avec @52B

Correct	260 482	95,52 %
Incorrect	12 216	4,48 %
Total	272 698	

Figure II. 22: Résultat de l'analyse de prédiction du modèle de régression logistique.

Grâce au modèle prédictif établi, nous avons pu évaluer la totalité des clients de la base du Groupe Renault Algérie, pour estimer la probabilité d'achat de la nouvelle version de la Sandero StepWay qui sera produite en Algérie.

CHAPITRE II: Présentation de l'entreprise et Exploitation de la base de données clients.

Ainsi, une base de plus de 4800 clients potentiels avec une probabilité d'achat du nouveau modèle de plus de 63 %, est communiquée au service CRM pour mieux cibler les clients de la prochaine campagne marketing de conquête de client.

Comme le nouvelle Dacia est produite en Algérie, elle est éligible au crédit à la consommation, donc le modèle peut être amélioré par l'ajout d'une nouvelle variable « Score crédit » qui sera basé sur le profil des clients qui ont obtenu un crédit à la consommation pour l'achat de la Renault Symbol.

Conclusion

Ce chapitre nous a tout d'abord permis d'apprécier le contexte dans lequel exerce l'entreprise et d'identifier ainsi les principaux objectifs de la segmentation des clients.

Ensuite, le diagnostic de la base de données clients de l'entreprise, proposé dans la seconde section du présent chapitre, a constitué une base de réflexion sur les différents types de segmentation possibles à partir des données existantes en entreprise.

En nous basant sur les résultats du précédent diagnostic ainsi que sur la base théorique constituée dans le chapitre précédent, nous avons segmenté les clients de l'entreprise en fonction de leur comportement d'achat de véhicules neufs, et ce, en utilisant des méthodes de classement et de classification du Data Mining, avant de présenter un Profiling des segments obtenus.

Enfin, nous avons utilisé les résultats de la segmentation afin de prédire l'achat de la nouvelle Dacia Sandero de la part des clients de l'entreprise, afin de permettre à l'entreprise d'optimiser la prochaine campagne de ciblage marketing.

Ces résultats, bien qu'intéressants, ne nous permettent pas de répondre à l'ensemble des objectifs fixés dans le cadre de notre projet. C'est pour cela que nous détaillons, dans le prochain chapitre, les deux études de marché réalisées auprès des clients de l'entreprise ainsi que des acheteurs de véhicules de toutes marques confondues, en Algérie.

CHAPITRE III :

**Méthodologie, Conception des modèles
et Interprétation des résultats.**

Introduction

Le diagnostic et l'exploitation de la base de données clients de l'entreprise nous permettent d'apprécier la nature des données collectées sur les clients de l'entreprise et donc d'identifier les limites d'une segmentation réalisée à partir de ces données.

Une telle segmentation ne remplissant pas l'ensemble des objectifs fixés pour notre projet, nous présentons, dans la première section de ce chapitre, une segmentation réalisée à partir des résultats d'une enquête sur les clients de l'entreprise.

Par ailleurs, dans le but de soutenir l'entreprise dans sa démarche de diversification des offres, nous dédions la seconde section de ce chapitre à l'étude de marché réalisée afin de segmenter les acheteurs de voitures en Algérie.

Dans la réalisation de chacune des études, nous suivons les étapes du processus *Knowledge Discovery*, présentées dans la figure III. 1.

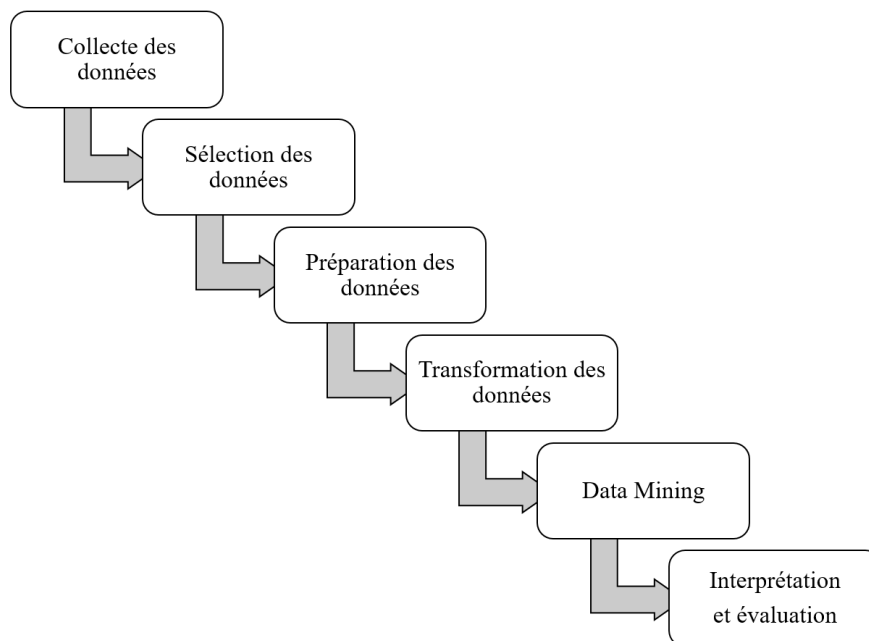


Figure III. 1: Etapes du Knowledge Discovery (Source : Li et Ruan, 2007).

Section I : Segmentation des clients de Renault Algérie

Dans le but de rehausser son image de marque dans le monde, le Groupe Renault met en place la stratégie « Back to Premium » à travers le lancement de nouveaux produits et la montée en gamme de certains modèles existants.

Par ailleurs, l'entreprise souhaite se rapprocher de ses clients afin de mieux répondre à leurs besoins et attentes et de pouvoir ainsi leur proposer des offres adéquates, notamment en matière de services après-vente.

La première section de ce chapitre est consacrée à l'étude de marché réalisée auprès des clients de Renault Algérie afin d'établir une segmentation efficace des clients.

L'objectif étant de permettre à l'entreprise d'acquérir une meilleure connaissance sur le comportement d'achat des clients en termes de services après-vente ainsi que leurs besoins et attitudes dans le choix d'une voiture, de cibler les clients acheteurs de l'un des nouveaux produits haute gamme de l'entreprise, et enfin, d'adapter sa stratégie de communication en fonction des différents segments.

1. Collecte des données

Les données des clients de l'entreprise Renault Algérie sont collectées à travers un sondage.

a) Choix des méthodes de sondage :

Un sondage est une enquête statistique qui permet de capturer les opinions individuelles à propos d'un sujet spécifique. Il doit être réalisé auprès d'un échantillon représentatif de la population étudiée afin que les résultats obtenus puissent être extrapolés à la population mère.

La détermination de la taille de l'échantillon est une tâche complexe qui implique la considération de facteurs tant quantitatifs que qualitatifs. En règle générale, la plupart des études marketing doivent utiliser des échantillons de taille minimale égale à 200. La fourchette type étant de 300 à 500¹⁸.

Plusieurs méthodes peuvent être utilisées pour réaliser des sondages et chacune présente des avantages et des inconvénients.

Naresh Malhotra (2007) présente différentes méthodes utilisées pour la réalisation de sondages ainsi qu'une liste de 14 critères qu'on prend généralement en compte pour le choix de l'une de ces méthodes.

La figure III.2 illustre les différentes méthodes de sondage proposées par l'auteur.

¹⁸ Source : Etudes Marketing, Naresh Malhotra, 2007.

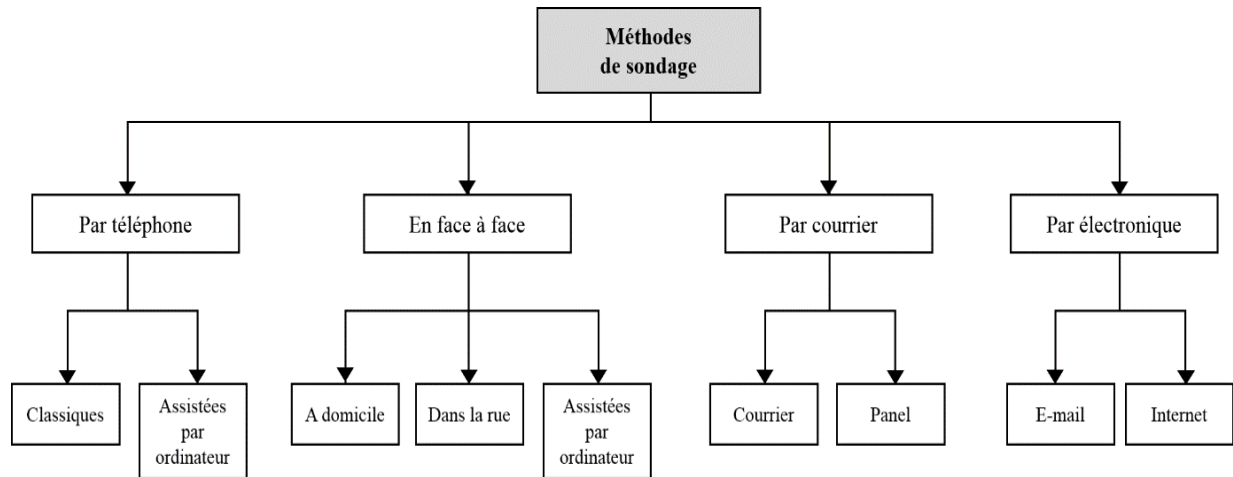


Figure III. 2: Méthodes de sondage (Source : Etudes Marketing, Naresh Malhotra. 2007, p121).

Parmi les critères évoqués dans l'ouvrage, nous nous intéressons plus particulièrement aux critères relatifs à la qualité de l'étude ainsi qu'aux contraintes matérielles de la méthode de sondage. Ainsi, les critères concernés sont les suivants :

- Anonymat du répondant (Il est nécessaire de nous assurer que les répondants sont bien des clients de l'entreprise).
- La diversité des questions.
- Le contrôle de l'échantillon.
- Le contrôle du travail de collecte de données.
- Le taux de réponse.
- La rapidité de la réalisation de l'enquête.
- Le coût.

Dans le Tableau III.1 suivant, nous présentons une partie de l'analyse comparative des méthodes de sondage.

Tableau III. 1: Comparaison des méthodes de sondage. (Source : Etudes Marketing, Naresh Malhotra. 2007, P133).

Critère	Téléphone / CATI ¹⁹	Domicile	Lieux publics	CAPI ²⁰	Courrier	Panel	E-mails	Internet
Diversité des questions	Faible	Elevée	Elevée	Elevée	Moyenne	Moyenne	Moyenne	Moyenne à forte
Contrôle de l'échantillon	Moyen à fort	Potentielle ment élevé	Moyen	Moyen	Faible	Moyen à fort	Faible	Faible à moyen
Contrôle de la collecte	Moyen	Faible	Moyen	Moyen	Elevé	Elevé	Elevé	Elevé
Taux de réponse	Moyen	Elevé	Elevé	Elevé	Faible	Moyen	Faible	Très faible
Rapidité de la réalisation de l'enquête	Elevée	Moyenne	Moyenne à forte	Moyenne à forte	Faible	Faible à moyenne	Elevée	Très élevée
Cout	Modéré	Elevé	Modéré à élevé	Modéré à élevé	Faible	Faible à modéré	Faible	Faible
Anonymat du répondant	Moyen	Faible	Faible	Faible	Elevé	Elevé	Moyen	Elevé

Etant donné que l'entreprise se procure les services d'un Call Center, et ce, pour un montant indépendant du nombre d'enquêtes réalisées, le coût des méthodes de sondage par téléphone ainsi que les campagnes E-mailing sont considérées très faibles.

En prenant en considération ces informations ainsi que la comparaison précédente, nous décidons d'opter pour une méthode de sondage téléphonique assistée par ordinateur, complétée par une campagne d'E-mailing.

b) Conception du questionnaire :

Les étapes suivies lors de la réalisation du questionnaire sont les suivantes :

- Détermination du contenu des questions à poser :

Il existe, dans la littérature, un grand nombre de types de données utilisées dans la segmentation des clients. Le choix des types de données à collecter doit se faire en fonction des objectifs de la segmentation. Dans notre cas d'étude, il s'agit de réaliser une segmentation basée sur le comportement d'achat ainsi que sur les besoins des clients.

¹⁹ Entretien par téléphone assistés par ordinateur.

²⁰ Entretien personnels assistés par ordinateur.

A cet effet, des données sur l'achat de la voiture et des services après-vente effectués par le client ainsi que celles relatives à ses besoins en matière d'automobile sont collectées.

De plus, proposer une stratégie de communication adéquate à chaque segment nécessite l'obtention de données sur l'utilisation des différentes sources d'information sur l'automobile de la part des clients.

Enfin, des données sociodémographiques sont collectées dans le but de réaliser un Profiling des segments obtenus.

- La formulation des questions :

La formulation des questions à poser dans un questionnaire doit se faire de manière à faciliter la compréhension et la formulation des réponses aux répondants mais aussi de façon à surmonter leur réticence à répondre. Nous optons donc pour un vocabulaire simple, des questions structurées (liste déroulante, dichotomiques, case à cocher, etc. Sauf lors des tests du questionnaire), et des intervalles de réponses pour les questions plus délicates (Revenu).

- L'ordre des questions :

Les premières questions sont souvent déterminantes pour assurer la collaboration des répondants. Ainsi, afin de garantir l'instauration d'une relation de confiance avec les participants, nous choisissons de commencer avec des questions générales sur l'expérience d'achat de la voiture et des services après-vente suivies par des questions sur les critères de choix d'un véhicule et l'utilisation des sources d'information avant d'aborder l'éventuel renouvellement de la voiture, pour enfin finir avec les questions relatives aux informations personnelles.

- Forme et disposition du questionnaire :

Comme mentionné précédemment, l'enquête est réalisée auprès des clients de l'entreprise *via* des appels téléphoniques assistés par ordinateur et d'une campagne E-mailing.

A cet effet, concernant les appels téléphoniques, nous avons préparé un questionnaire sous format PDF destiné aux opérateurs du Call Center, accompagné d'un fichier sous format Excel servant à récolter les données.

Pour la campagne E-mailing, un texte incluant un lien de redirection vers un formulaire en ligne réalisé sur Google Forms, est préparé afin d'expliquer l'objectif de l'étude.

c) Tests du questionnaire :

Il est très recommandé de tester tous les aspects d'un questionnaire avant de le diffuser. Ceci inclut le contenu des questions, leur formulation, leur enchaînement ainsi que la forme et l'aspect du questionnaire, dans le but de vérifier la pertinence des questions, leur compréhension ainsi que la durée requise pour y répondre.

Avant le test de notre questionnaire, une journée de formation au sein du Call Center est organisée afin d'apporter les clarifications nécessaires concernant le questionnaire et de

communiquer aux collaborateurs les objectifs de notre étude ainsi que les conditions dans lesquelles elle sera réalisée.

Le test du questionnaire est réalisé, d'une part, à travers l'envoi d'E-mails à nos proches ainsi qu'à des employés de l'entreprise Renault Algérie, tous possédant des voitures des marques Renault ou Dacia, nous permettant ainsi de récolter des feedbacks relatifs au contenu et à la forme du questionnaire, et d'autre part, par l'équipe du Call Center, pour une durée de 5 jours.

Ce test a permis la récolte des réponses de 55 clients, dont 30 ayant répondu à plus de 80% des questions. Selon les estimations du Call center, la durée des appels variait de 30 à 50 minutes. La durée maximale recommandée par ce dernier étant de 30 minutes et prenant en considération le temps limité dont nous disposons pour la réalisation de notre projet, le questionnaire est révisé en vue de la réduction du nombre de questions.

d) Collecte des données :

Les tests réalisés ont permis d'apporter les modifications nécessaires au questionnaire. Une fois la révision de ce dernier achevée, les deux campagnes (téléphoniques et E-mailing) sont lancées, simultanément. Le fichier PDF du questionnaire, le fichier Excel préparé à l'intention du call center ainsi qu'un aperçu de l'Email envoyé aux clients sont présentés dans les annexes 6, 7 et 8 respectivement.

La récolte de données s'est étalée sur une période de 3 semaines, soit, après la constitution d'une base de données de 560 répondants. 210 de ces réponses proviennent du formulaire en ligne. Les statistiques de l'enquête fournies par l'équipe du Call Center sont présentées dans le tableau III.2.

Tableau III. 2: Statistiques de l'enquête réalisée par le Call Center.

Analyse de l'opération enquête clients				
Clients contactés	Clients argumentés	OK	Injoignables	Faux numéros/ Erreur
2151	631	350	990	525

2. Sélection des données

L'étape de sélection des données permet de trier ces dernières en fonction des objectifs de l'étude. Dans notre cas, le questionnaire est spécialement préparé pour collecter les données nécessaires.

Cependant, un premier tri effectué dans la base de données provenant du formulaire en ligne, afin d'écartier 5 doublons ainsi que 7 personnes affirmant n'avoir pas acheté un véhicule de la marque Renault ou Dacia. Ceci réduit le nombre total d'observations à 548.

Un second tri, concernant les questions, est ensuite réalisé afin d'écartier les variables servant à confirmer l'identité des répondants.

3. Préparation des données

Dans notre cas, la préparation des données consiste en la fusion des deux bases de données, créées à partir du formulaire en ligne et de l'opération réalisée par le Call Center, dans un seul fichier SPSS.

4. Transformation des données

Avant de pouvoir recourir aux différentes techniques de Data Mining, il convient d'apporter les transformations nécessaires aux données.

La première modification apportée est le codage des données sur IBM SPSS Statistics, comme le montre la figure III.3 suivante :

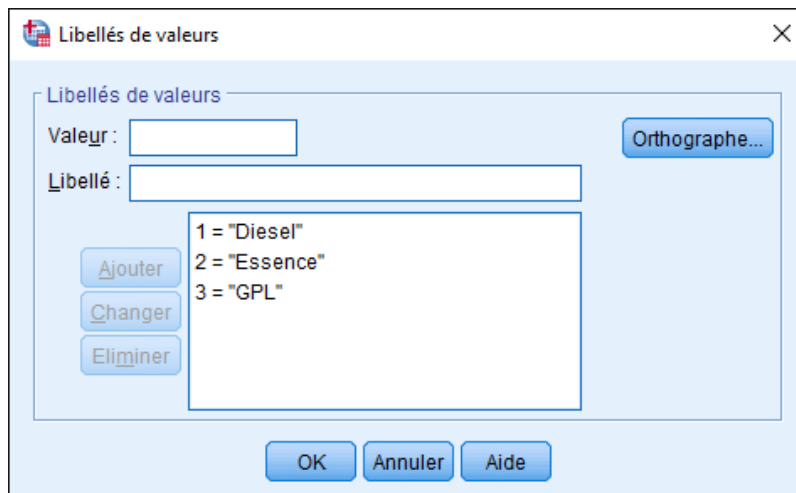


Figure III. 3: Exemple de codage des données sur IBM SPSS Statistics (Variable : Motorisation).

Une discrétisation de la variable Région est ensuite réalisée, regroupant ainsi les 48 wilayas du pays en 4 régions : Nord, Est, Oued et Sud.

Une autre partie de la transformation des données consiste à calculer de nouvelles variables. Ainsi, la variable APV est créée, afin de représenter le comportement d'achat des services après-vente par les clients auprès des ateliers des marques Renault et Dacia, comme suit : $APV = (1 - \text{Aucune}) * ((\text{Maintenance} + \text{Réparation} + \text{ChangementDesPneus} + \text{Carrosserie}) / 4)$.

La variable AgeVoiture est également calculée comme suit : $\text{AgeVoiture} = 2016 - \text{AnneeImmatriculation}$.

Enfin, dans le but de réduire le nombre des critères de choix d'une voiture ainsi que les sources d'information utilisées par les clients pour se renseigner sur les voitures, nous procédons à deux analyses factorielles à l'aide du logiciel IBM SPSS Statistics. Ces variables étant numériques catégorielles, une ACP, accompagnée d'une rotation Varimax, est appropriée.

Dans les deux analyses ACP nous utilisons le critère de Kaiser pour le choix des composantes à retenir, c'est-à-dire celles avec une valeur propre supérieure à 1.

L'indice KMO de la première ACP est équivalent à 0,734, ce qui nous permet de réaliser l'analyse factorielle afin de réduire le nombre des critères de choix.

Les résultats de l'ACP indiquent que 4 composantes principales sont retenues, représentant une valeur propre cumulée de près de 75%. La plus faible qualité de représentation est celle du critère Niveau d'équipement, estimée à 0,7.

La figure III.4 présente les résultats de la rotation de la matrice des composantes obtenue à l'aide du logiciel IBM SPSS Statistics.

Rotation de la matrice des composantes^a

	Composante		
	1	2	3
CriterePrix			,727
CritereConsommation			,784
CritereStatutSocial		,832	
CritereDesign	,445	,606	
CritereConfort	,830		
CritereSecurite	,804		
CritereEquipement	,668	,247	
CritereNouveaute		,819	

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.

Figure III. 4: Résultat de la rotation de la première matrice des composantes principales sur IBM SPSS Statistics.

A partir de la figure précédente, nous constatons que les critères : Confort, Sécurité et Niveau d'équipements, qui représentent les caractéristiques de la voiture, sont les plus corrélés à la première composante principale. Cette dernière nous permet de calculer la variable Caractéristiques, mesurant l'importance accordée par les clients aux différentes caractéristiques de la voiture, comme suit :

$$\text{Caracteristiques} = (\text{Confort} + \text{Equipement} + \text{Securite}) / 3.$$

Les critères de choix les plus corrélés à la deuxième composante principale sont le Statut Social, le Design et la Nouveauté du modèle. Trois critères qui indiquent l'importance accordée au prestige qu'apporte une voiture au client. La variable Prestige est donc calculée comme suit :

$$\text{Prestige} = (\text{StatutSocial} + \text{Design} + \text{Nouveaute}) / 3.$$

Enfin, la forte corrélation existante entre les critères Prix et Consommation de carburant avec la troisième composante principale signifie que cette dernière représente l'importance accordée au critère financier lors de l'achat d'une voiture. La variable Finance est donc créée comme suit : $\text{Finance} = (\text{Prix} + \text{Consommation}) / 2.$

Nous procédons de la même manière pour la réalisation de la deuxième ACP qui concerne les sources d'information utilisées pour se renseigner sur les voitures.

L'indice KMO est estimé cette fois-ci à 0,77, soit un score acceptable pour la réalisation d'une analyse factorielle.

Les résultats de l'ACP, illustrés dans la figure III.5 présentée ci-dessous, indiquent que 3 composantes principales sont retenues, avec une valeur propre cumulée de 82,3% ainsi qu'une qualité de représentation minimale pour la variable AvisConcessionnaires estimée à 0,68.

Rotation de la matrice des composantes^a

	Composante		
	1	2	3
InfoAffiches	,815		
InfosJournaux	,796		
InfosTV	,771		
InfosRadio	,830		
InfosReseauxSociaux	,247	,788	
InfosSiteMarque		,901	
InfosAutresSites		,896	
InfosAvisProches			,803
InfosAvisConcessionnaires			,791

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.

Figure III. 5: Résultat de la rotation de la deuxième matrice des composantes principales avec IBM SPSS Statistics.

La première composante principale est plus corrélée aux variables représentant les sources d'information traditionnelles, à savoir, les Affiches publicitaires, les Journaux, les publicités Télévisées ainsi que la Radio. Il en résulte que la variable Info_Traditionnelle soit calculée comme suit : $Info_Traditionnelle = (Affiches + Journaux + TV + Radio) /4$.

La deuxième composante principale est fortement corrélée aux variables Réseaux sociaux, Site web de la marque et Autres Sites Web. Elle nous permet de créer la variable Info_Connexion, mesurant l'utilisation des sources d'information relatives à internet de la part des clients de l'entreprise, comme suit :

$$Info_Connexion = (ReseauxSociaux + SiteMarque + AutresSites) /3.$$

Enfin, la troisième composante principale étant fortement corrélée aux variables qui mesurent l'utilisation de l'avis des proches et des concessionnaires comme source d'information sur les voitures, nous permet de calculer la variable Info_Avis comme suit :

$$Info_Avis = (AvisProches + AvisConcessionnaires) /2.$$

5. Application du Data Mining

Avant de recourir aux outils de Data Mining, nous établissons un profil général des clients des deux marques Renault et Dacia par une analyse descriptive des caractéristiques de ces individus. Cette analyse est présentée en annexe 9.

a) **Segmentation basée sur le comportement d'achat du véhicule et des services après-vente :**

Le premier objectif de l'utilisation du Data Mining dans la présente étude est de segmenter les clients de l'entreprise en fonction de leur comportement d'achat du véhicule ainsi que des services après-vente.

Comme expliqué précédemment, les méthodes les plus adaptées pour la création de nouveaux segments sont les méthodes de classification. Elles permettent, en effet, de regrouper les individus selon les similarités existantes entre leurs différentes caractéristiques.

Dans notre cas, nous souhaitons étudier l'achat des services en fonction de la marque, des segments de véhicules et de la motorisation choisie. Ces caractéristiques s'ajoutent donc à la variable APV comme entrées dans notre modèle de classification.

Nous utilisons l'indice silhouette, présenté dans le chapitre I, pour l'évaluation de la qualité de la segmentation obtenue ainsi que la comparaison des différentes méthodes de classification. Le nœud Classification non supervisée automatique, dans le logiciel IBM SPSS Modeler, permet de tester simultanément 3 algorithmes de classification : K-means, Two-Step ainsi que l'algorithme de classification neuronale : Kohonen.

L'exécution de ce nœud révèle que l'algorithme Two-Step fournit la meilleure qualité de classification avec 7 clusters créés, suivi de l'algorithme K-means. Toutefois, nous savons que l'utilisation de l'algorithme de K-means nécessite la détermination préalable du nombre de cluster (la valeur par défaut dans le logiciel est de 5) contrairement à Two-Step qui fournit un nombre optimal de clusters. Nous décidons donc de lancer une nouvelle classification K-means, en choisissant le nombre optimal de clusters fournit par l'algorithme Two-Step, soit 7.

La comparaison des indices silhouette des deux méthodes nous permet d'affirmer que la classification par l'algorithme Two-Step est de meilleure qualité, avec un indice silhouette de 0,491, comme le montre la figure III.6 suivante :

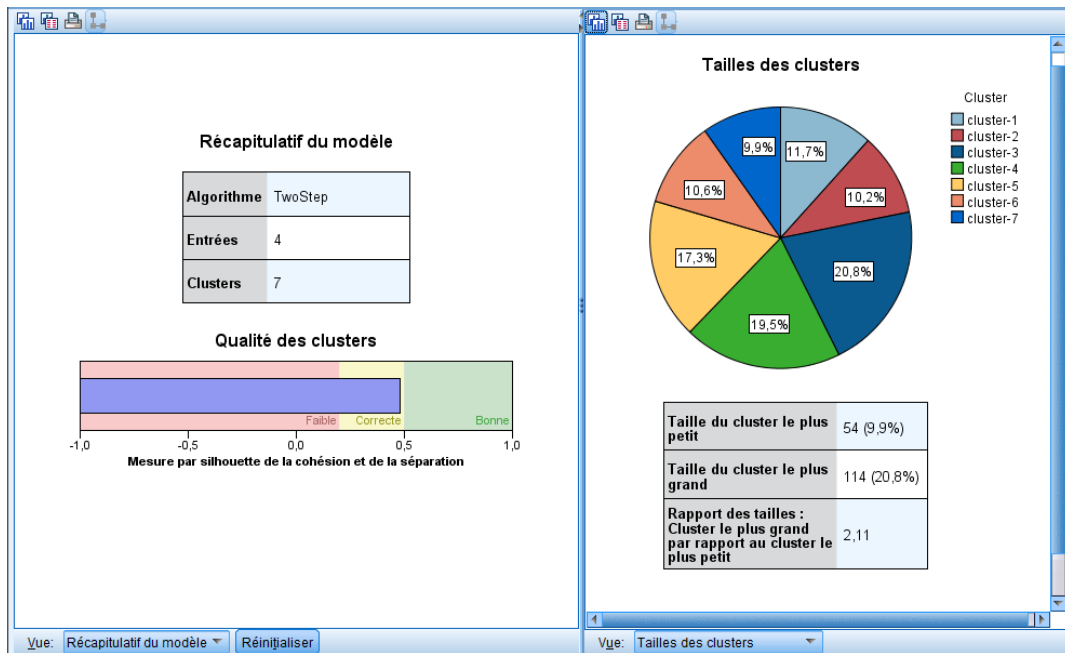


Figure III. 6: Récapitulatif du modèle Two-Step pour la première segmentation des clients.

Le modèle de classification réalisé à l'aide du logiciel IBM SPSS Modeler est présenté dans la figure III.7, comme suit :

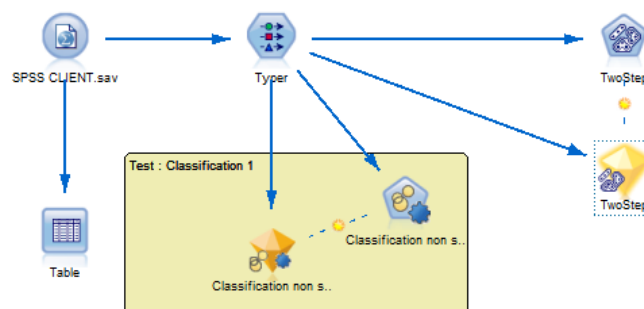


Figure III. 7: Modélisation de la première classification des clients sur IBM SPSS Modeler.

En analysant les résultats du modèle, nous constatons que la variable Segment représente l'indicateur le plus important dans la construction des clusters. Il en résulte que chaque cluster est majoritairement constitué d'une des modalités de la variable Segment, sauf pour les clusters 2 et 5 dont les distributions en fonction des segments véhicules sont représentées dans la figure III.8 suivante :

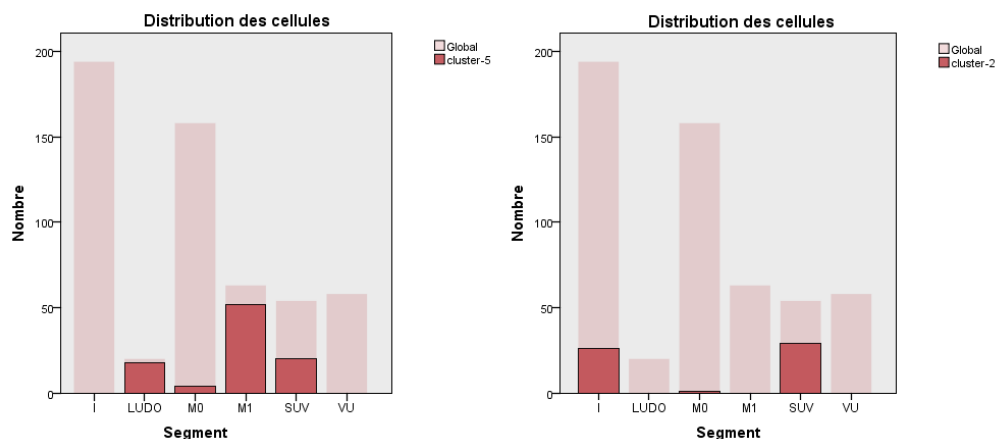


Figure III. 8: Répartition des segments de véhicules dans les clusters 2 et 5.

Afin de faciliter l'interprétation des 7 clusters, présentée dans la prochaine étape de l'étude, nous résumons leurs caractéristiques dans le Tableau III.3 suivant :

Tableau III. 3: Caractéristiques des 7 clusters de la première segmentation des clients.

Cluster	Cluster-1	Cluster-4	Cluster-5	Cluster-3	Cluster-6	Cluster-2	Cluster-7	Global
Taille en pourcentage (effectif)	20,8% (114)	19,0% (104)	17,9% (98)	11,7% (64)	10,6% (58)	10,0% (55)	10,0% (55)	(548)
Segment -Catégorie la plus fréquente (Pourcentage)	I (100%)	M0 (82,7%)	M1 (53,1%)	M0 (100%)	VU (100%)	SUV (52,7%)	I (100%)	/
Marque -Catégorie la plus fréquente (Pourcentage)	Renault (100%)	Renault (100%)	Renault (99,0%)	Dacia (100%)	Renault (100%)	Dacia (100%)	Renault (100%)	Renault (78,1%)
Motorisation -Catégorie la plus fréquente (Pourcentage)	Essence (100%)	Essence (100%)	Diesel (96,9%)	Essence (98,4%)	Diesel (98,3%)	Essence (85,5%)	Diesel (100%)	Essence (59,9%)
APV (Moyenne)	0,34	0,40	0,38	0,20	0,37	0,36	0,35	0,35

b) Prédiction de l'achat d'un nouveau modèle :

L'entreprise Renault Algérie prévoit la commercialisation d'une nouvelle Megane pour les mois à venir. Le lancement de ce produit s'inscrit dans la stratégie globale de montée en gamme du Groupe Renault.

Dans l'objectif de fournir une cible pertinente à l'entreprise Renault Algérie pour le lancement de la Mégane 4, nous réalisons un modèle prédictif des clients appêtent d'acheter le nouveau modèle.

Pour ce faire, nous créons une variable cible qui attribue des scores aux clients, permettant d'identifier ceux avec un fort potentiel d'achat de ceux avec un moyen et un faible potentiel d'achat.

Pour évaluer le potentiel d'achat de la nouvelle Megane, nous nous basons sur les questions relatives au renouvellement d'achat des clients. Ainsi, les variables utilisées pour la création des scores sont les suivantes :

- Renouvellement : Le renouvellement ou non de la voiture dans les 6 prochains mois.
- Renouv_Segment : Le segment souhaité pour le renouvellement.
- Renouv_Prix : Le montant du renouvellement prévu.

L'algorithme utilisé pour l'établissement des scores se présente comme suit :

```
Si (Renouvellement = 0) alors (Score = 0) ;
Sinon
  Si (Renouv_Segment = 3)
    Si (Renouv_Prix >= 5) alors (Score = 2) ;
    Sinon (Score = 1) ;
  Fsi.
  Sinon
    Si (Renouv_Prix >= 5) alors (Score = 1) ;
    Sinon (Score = 0) ;
  Fsi.
Fsi.
```

Nous construisons ensuite un modèle permettant de calculer directement la probabilité d'apparition du phénomène recherché, c'est-à-dire l'achat du nouveau modèle.

L'une des variables d'entrée utilisées dans la construction du modèle est le numéro du cluster issu de la précédente segmentation. Cela nous permettra, d'une part, d'apprécier l'utilité de la segmentation en marketing, et d'autre part, de cibler les segments les plus appétent d'acheter la nouvelle Megane.

Afin d'augmenter la fiabilité du modèle ainsi que sa capacité de généralisation, nous divisons la population étudiée en deux échantillons. Le premier, d'apprentissage, est composé de 60% de l'ensemble des individus, il permet la construction du modèle. Le second, constitué de 40% des individus sert à tester le modèle.

Notre objectif étant de déterminer les segments présentant le meilleur potentiel d'achat, nous choisissons les réseaux bayésiens comme méthode de classement, car les résultats qu'ils proposent permettent de prédire la valeur de n'importe quelle variable du modèle en fonction d'autres variables en mettant en avant les relations de causalité entre les variables, et ce, même si toutes les valeurs des attributs ne sont pas disponibles.

Les variables choisies pour le classement des individus selon le potentiel d'achat du nouveau modèle sont : le segment, le genre, l'âge et la région d'habitation du client, auxquelles

s'ajoutent la fréquence d'achat, l'âge du véhicule possédé ainsi que la distance parcourue, en moyenne par jour. Cette dernière peut être facilement calculée à partir du kilométrage indiqué lors de la première révision, qui se fera nécessairement dans les ateliers de la marque car sous garantie.

Le modèle de classement réalisé à l'aide du logiciel IBM SPSS Modeler est présenté dans la figure III.9 suivante :

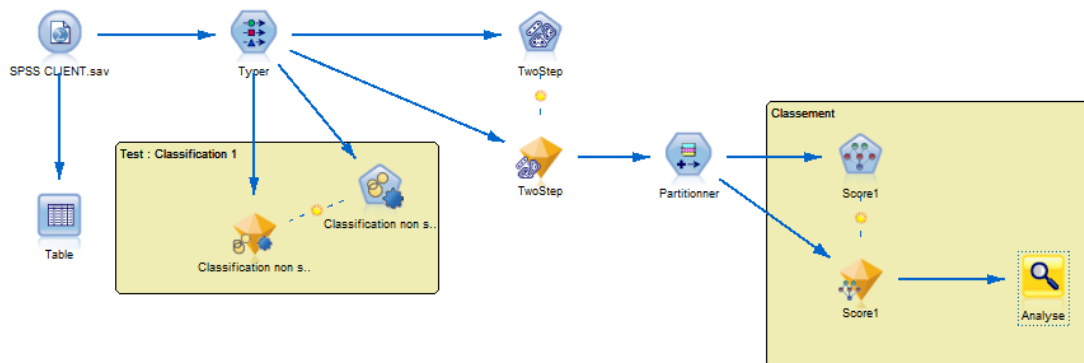


Figure III. 9: Modélisation du classement des clients en fonction de la variable Score sur IBM SPSS Modeler.

L'exécution du modèle sur IBM SPSS Modeler indique une précision générale très acceptable qui avoisine les 80%, comme illustré dans la figure III.10 suivante :

Analyse de [Score] #13

Fichier Edition

Analyse Annotations

Réduire tout Développer tout

Résultats du champ de sortie Score

Comparaison de \$XS-Score avec Score

'Partitionner'	1_Apprentissage		2_Test	
Correct	265	79,82 %	173	80,09 %
Incorrect	67	20,18 %	43	19,91 %
Total	332		216	

OK

Figure III. 10: Résultat du modèle de classement : Score.

c) Segmentation basée sur les besoins des clients en matière d'automobile :

Après avoir segmenté les clients de l'entreprise en fonction de leur comportement d'achat de véhicules et de services après-vente et une fois les segments les plus appétent à acheter le nouveau modèle qui sera bientôt lancé par l'entreprise, nous nous intéressons, dans cette partie, à l'analyse des besoins et exigences des clients, notamment des segments ciblés, vis-à-vis de l'automobile.

Pour ce faire, nous réalisons une nouvelle segmentation qui permet mettre en évidence les critères d'achats d'une voiture des segments précédemment établis. Nous procédons donc de la même façon que pour la classification précédente, c'est-à-dire, en comparant les différentes méthodes de classification à l'aide du nœud classification non supervisée automatique.

Les variables utilisées pour la construction du modèle de classification sont : Le numéro de cluster obtenu par la première classification ainsi que les variables Caractéristiques, Prestige et Finance.

Encore une fois, le modèle présentant la meilleure qualité de classification, estimée à 0,38 est réalisé par l'algorithme Two-Step proposant 7 clusters. Le classement des prédicteurs par importance est le suivant :

- \$T-Two-Step : Le numéro de cluster de la première segmentation.
- Caractéristiques.
- Finance.
- Prestige.

Le modèle obtenu à l'aide du logiciel IBM SPSS Modeler se présente comme suit :

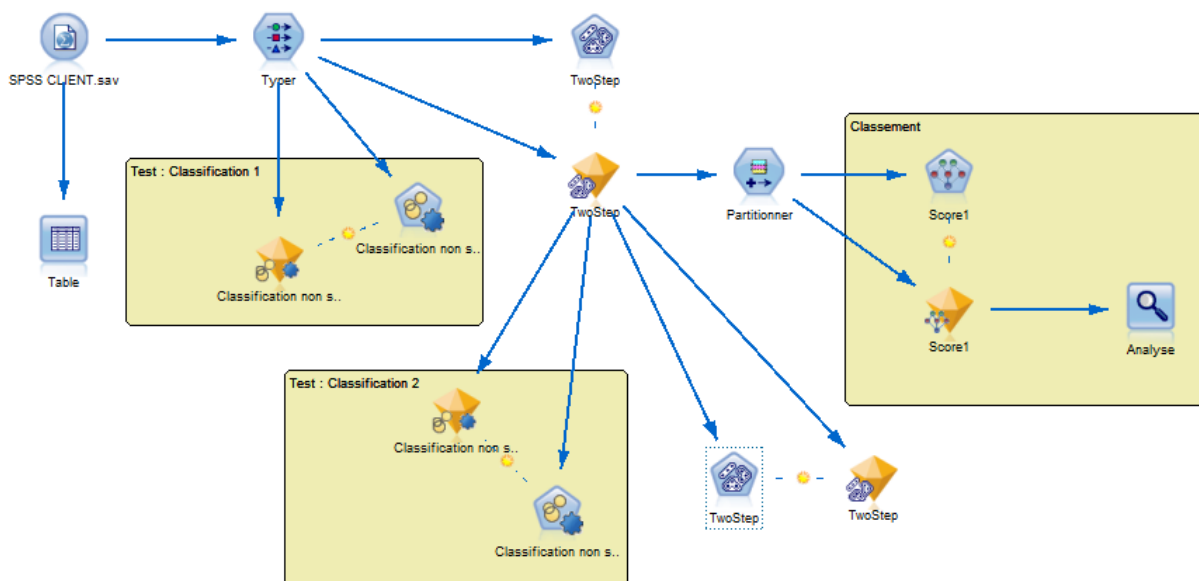


Figure III. 11: Modélisation de la deuxième classification des clients sur IBM SPSS Modeler.

Nous résumons les différentes caractéristiques des nouveaux segments dans le tableau III.4 suivant :

Tableau III. 4: Caractéristiques des 7 clusters de la deuxième segmentation des clients.

Cluster	Cluster-4	Cluster-1	Cluster-3	Cluster-7	Cluster-2	Cluster-5	Cluster-6	Global
Taille en pourcentage (effectif)	20,8% (114)	18,8% (103)	17,9% (98)	11,7% (64)	10,8% (59)	10,0% (55)	10,0% (55)	548
Cluster Two-Step1- Catégorie la plus fréquente (Pourcentage)	cluster-1 (100%)	cluster-4 (100%)	cluster-5 (100%)	cluster-3 (100%)	cluster-6 (98,3%)	cluster-7 (100%)	cluster-2 (100%)	/
CARACTERISTIQUES- Moyenne sur 5	4,20	4,18	4,33	4,11	4,31	4,30	4,34	4,24
FINANCE-Moyenne sur 5	4,29	4,33	4,13	4,38	4,25	4,29	4,35	4,28
PRESTIGE-Moyenne sur 5	3,81	3,64	3,73	3,67	3,71	3,87	3,94	3,75

6. Interprétation des résultats

Dans cette sixième partie de l'étude réalisée auprès des clients de l'entreprise Renault Algérie, nous interprétons les résultats des différents outils de Data Mining précédemment utilisés.

Nous commençons par identifier les segments pertinents pour des actions marketing visant le développement des services après-vente de l'entreprise, grâce au premier modèle de classification utilisé.

Le classificateur bayésien utilisé sur les segments obtenus par la première classification nous permettra ensuite d'identifier et de profiler les segments regroupant les clients appêtent d'acheter le prochain modèle qui sera lancé par l'entreprise, à savoir la Megane 4.

Après cela, une analyse de la dernière classification des clients est réalisée, permettant à l'entreprise de comprendre les besoins et attentes des différents segments, notamment des segments cibles pour le lancement du nouveau modèle.

Une proposition relative aux canaux de communication appropriés pour chaque segment est enfin présentée.

a) Résultats de la première classification :

La situation de crise que connaît le marché automobile algérien, causée par l'importante limitation des importations de véhicules neufs, oblige l'entreprise à développer son activité après-vente afin de compenser l'écart dans le chiffre d'affaires provoqué par la baisse des ventes de véhicules importés.

Le but de cette première classification est de segmenter les clients de l'entreprise en fonction de leur comportement d'achat de véhicules mais aussi des services après-vente.

A cet effet, le modèle de classification basé sur l'algorithme Two-Step utilisé, segmente les clients en 7 principaux clusters, dont les caractéristiques sont présentées dans le tableau III.4 précédemment présenté.

La classification obtenue permet à l'entreprise d'identifier les segments contractant le moins de services auprès des ateliers de la marque et donc de mieux cibler ses actions marketing relatives à la promotion de ces services.

En effet, le tableau III.4, indique que le cluster-3 représentant le segment M0 de la marque Dacia, regroupe les individus les moins acheteurs des services après-vente de la marque, et ce, malgré le fait que les individus du cluster-4, représentant le même segment mais de la marque Renault, soient ceux qui achètent le plus de services après-vente de la marque.

Afin d'obtenir une meilleure visualisation du comportement d'achat des services après-vente des différents clusters, nous réduisons la moyenne de la variable APV de chaque cluster, comme illustré dans la figure III.12 suivante :

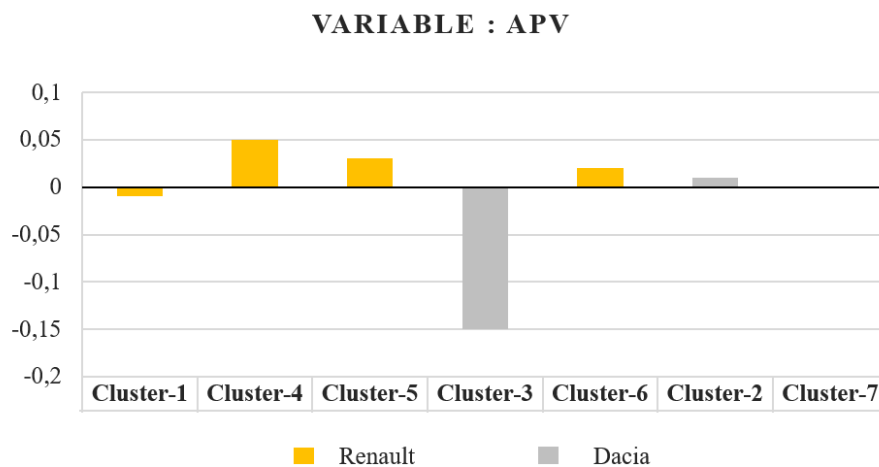


Figure III. 12: Comparaison des clusters en fonction de la variable APV.

En plus du cluster-3, nous constatons que les clusters 1 et 7, représentant tous les deux le segment I de la marque Renault, ont des moyennes APV faibles en comparaison avec les segments M0, M1 et VU de la même marque. Ils constituent de ce fait un manque à gagner pour l'entreprise en matière de vente des services après-vente.

Enfin, il est à noter que les clients Renault en général achètent plus de services auprès des ateliers de la marque, avec une moyenne APV de 0,37, que les clients Dacia, dont la moyenne APV est égale à 0,28.

b) Résultats du modèle de classement :

Afin d'identifier, parmi les segments précédemment établis, ceux présentant le plus de probabilité d'acheter la nouvelle Megane, un score basé sur les intentions de renouvellement de véhicule de la part des clients est créé et utilisé dans un classificateur bayésien.

Le réseau bayésien obtenu qui est présenté dans la figure III.13, met en avant les relations existantes entre les différentes variables utilisées pour la construction du modèle.

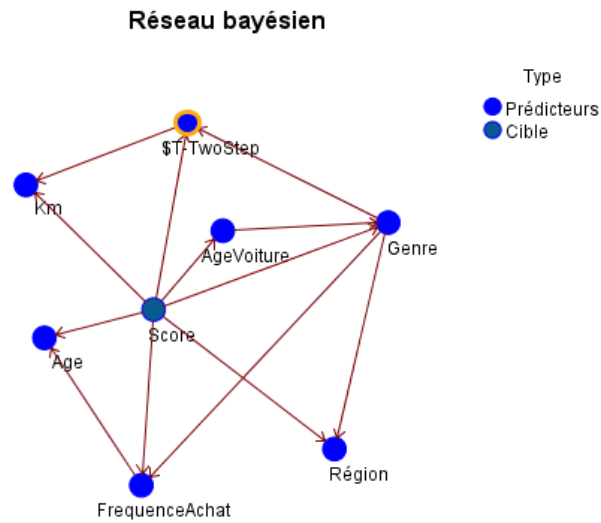


Figure III. 13: Réseau bayésien obtenu.

Nous remarquons, qu'en plus de la variable Score, la probabilité d'appartenance d'un individu à l'un des segments dépend du genre de l'individu.

Ces probabilités conditionnelles de la variable Segment sont détaillées dans la figure III.14 suivante :

**Probabilités conditionnelles de
\$T-TwoStep**

Parents		Probabilité						
Genre	Score	cluster-1	cluster-2	cluster-3	cluster-4	cluster-5	cluster-6	cluster-7
1	0	0,20	0,10	0,12	0,22	0,17	0,10	0,10
1	1	0,11	0,06	0,17	0,17	0,23	0,15	0,11
1	2	0,14	0,05	0,05	0,14	0,43	0,19	0,00
2	0	0,44	0,18	0,07	0,07	0,07	0,09	0,09
2	1	0,33	0,08	0,08	0,08	0,33	0,00	0,08
2	2	0,25	0,25	0,00	0,00	0,00	0,00	0,50

Figure III. 14: Table des probabilités conditionnelles du noeud \$T-TwoStep.

A partir de la figure précédente, nous constatons que les clients masculins appètent d'acheter une Megane 4 ont plus de probabilité d'appartenir au cluster-5, avec une probabilité conditionnelle de 0,43.

En revanche, la gente féminine la plus favorable à l'achat du nouveau modèle a 50% de probabilité d'appartenir au cluster-7.

Rappelons que le cluster-5 est principalement constitué d'individus possédant des voitures du segment I de la marque Renault avec motorisation Diesel et que les individus du cluster-7 possèdent majoritairement des véhicules de segment M1, SUV ou Ludospace de la marque Renault.

Les clusters 5 et 7 représentent donc les segments les plus pertinents pour la campagne marketing visant la promotion de la nouvelle Megane. C'est donc sur ces segments que se basera notre prochaine analyse.

c) Résultats de la seconde classification :

Dans le but d'analyser les besoins et exigences des clients, notamment des segments ciblés, vis-à-vis de l'automobile, une deuxième classification basée sur les notes attribuées aux critères de choix d'une voiture est réalisée.

Les résultats de la classification par la méthode Two-Step indiquent que le prédicteur le plus important de la segmentation est le numéro du cluster obtenu par la première classification.

Les nouveaux clusters sont donc quasiment constitués d'individus de chacun des anciens segments, ce qui nous facilite l'identification des segments cibles.

En effet, les individus appartenant au premier segment cible, c'est-à-dire le cluster-5 précédent, font désormais partie du cluster-3 et les individus du second segment cible, à savoir le cluster-7, appartiennent maintenant au cluster-5. Nous identifions ainsi les nouveaux segments cibles qui sont les clusters 3 et 5.

- Profiling des segments :

Avant de décrire les besoins des clients appartenant aux deux segments, nous procédons à un premier Profiling basé sur les données sociodémographiques des individus.

Pour cela, ne perdons pas de vue que le ciblage des segments est conditionné par le genre des individus et donc que le cluster 3 concerne exclusivement les clients masculins et le cluster 5, la clientèle féminine.

Le Profiling des cluster 3 et 5 sont présentés dans la figure III.15 suivante :

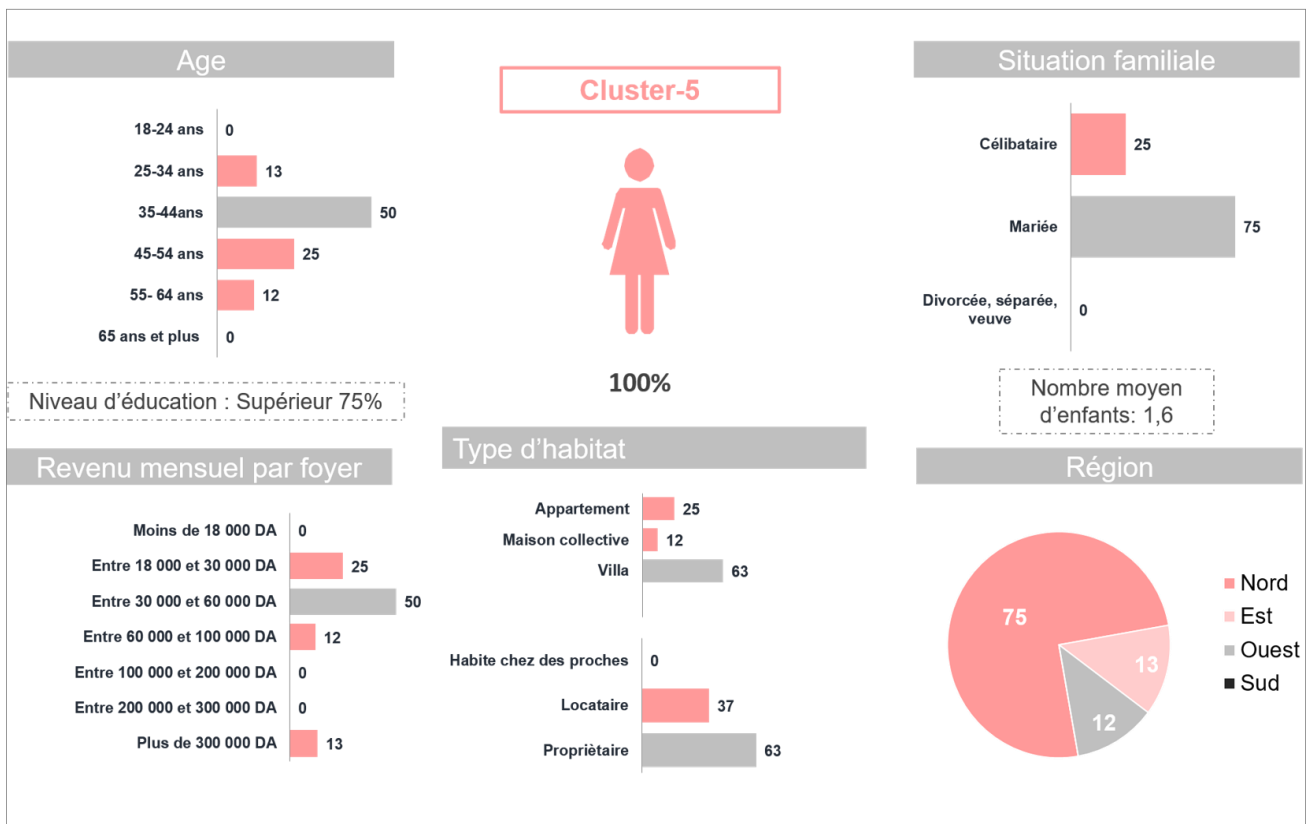
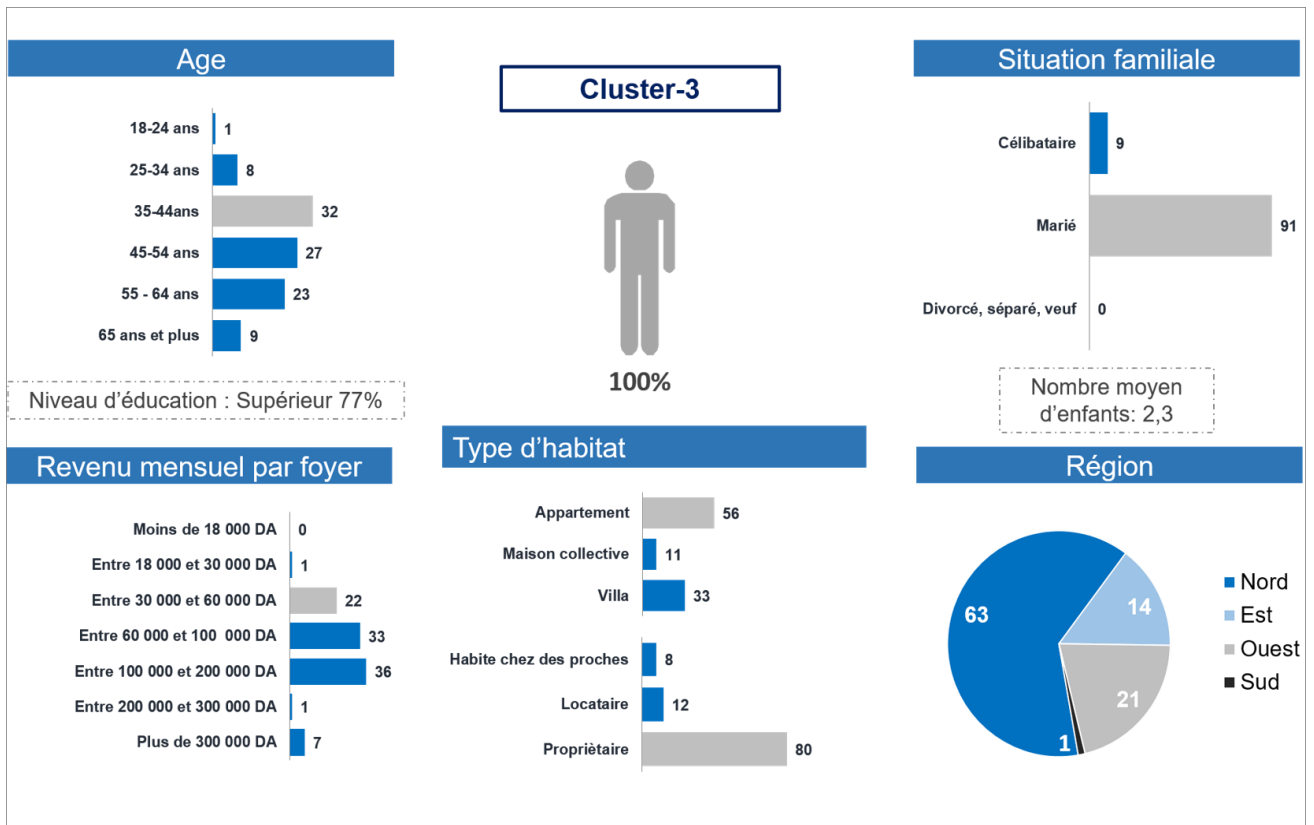


Figure III. 15: Profiling sociodémographique des clusters cibles.

Une fois les profils sociodémographiques établis, nous nous intéressons aux besoins des clients en général, et des segments ciblés en particulier.

A partir du tableau III.4, nous pouvons constater que le critère le plus important dans le choix du véhicule pour les clients Renault Algérie est le critère financier, à savoir, le prix d'achat et la consommation du carburant, suivi de près par les caractéristiques du véhicule et enfin, le prestige qu'apporte la possession d'un véhicule.

Ceci s'applique aussi pour les clients du cluster-3. En revanche, le cluster-5 accorde une importance presque égale au critère financier qu'aux caractéristiques de la voiture, avant de s'intéresser au prestige.

Dans le but de faciliter l'interprétation ainsi que la comparaison des caractéristiques par rapport à chaque segment, nous réduisons les moyennes de chacun des critères d'achat.

La figure III.16 montre les écarts des moyennes des segments de la moyenne générale attribuée à la variable Caractéristiques. Nous remarquons qu'en dépit du fait que le cluster-3 accorde une plus grande importance aux caractéristiques de la voiture que le cluster-5, les deux segments présentent des notes nettement supérieures à la moyenne générale par rapport à ce critère.

Ceci nous informe sur les besoins et attentes des segments cibles en matière de sécurité, confort et niveau d'équipement du véhicule, qui, précisons-le, sont plus prononcés pour les clients de sexe masculin que féminin.

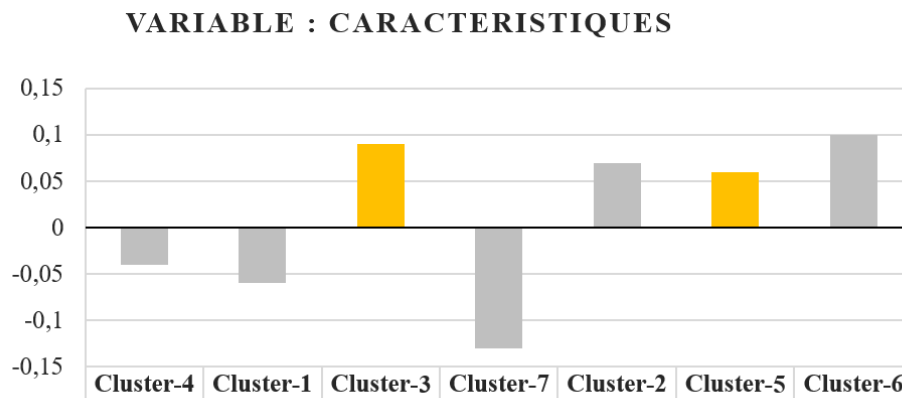


Figure III. 16: Comparaison des clusters en fonction de la variable : Caractéristiques.

La variable Finance est calculée à partir des critères Prix et Consommation de carburant. Elle mesure donc l'importance accordée par le client au critère financier lors du choix du modèle.

La comparaison des clusters en fonction de cette variable, présentée dans la figure III.17, nous montre que le cluster-3 est de loin le moins soucieux des contraintes financières lors de

l'achat d'un véhicule. Par ailleurs, la note attribuée au même critère par le deuxième segment cible, représenté par le cluster-5, est à peine plus importante que la moyenne.

Nous en déduisons que, de manière générale, les clients appètent d'acheter la nouvelle Megane sont moins regardants sur le prix et la consommation de carburant que la plupart des clients de l'entreprise, et que cette différence est beaucoup plus marquée chez hommes.

Une stratégie de commercialisation basée sur la réduction des prix ne serait sans doute pas la plus appropriée.

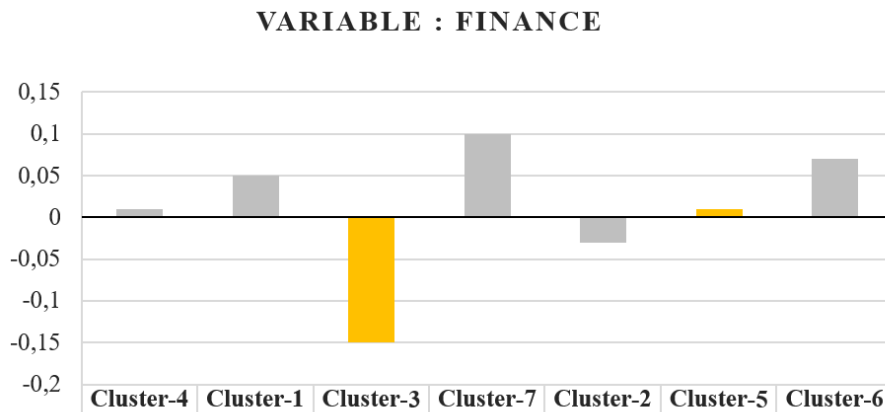


Figure III. 17: Comparaison des clusters en fonction de la variable : Finance.

Enfin, le dernier critère, calculé à partir des notes relatives au Design, la Nouveauté du modèle et le Statut Social, mesure l'importance accordée par les clients au prestige que procure la détention d'une voiture.

La figure III.18 présente une comparaison des segments par rapport à ce critère de choix. Elle nous permet de constater que les individus appartenant au cluster-5 font partie des clients les plus soucieux du prestige relatif à la possession d'un véhicule, tandis que ceux du cluster-3 affichent une note légèrement en dessous de la moyenne pour le même critère.

Ceci nous oriente sur l'intérêt d'inclure le critère du prestige et de la position sociale dans le message de la campagne de communication sur le nouveau modèle, notamment avec la gente féminine.

VARIABLE : PRESTIGE

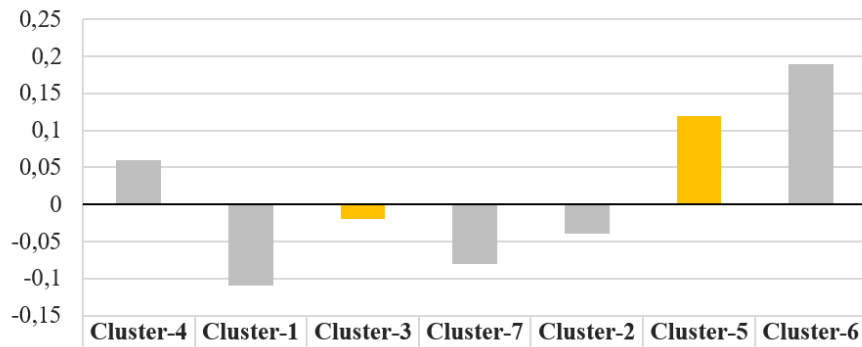


Figure III. 18: Comparaison des clusters en fonction de la variable : Prestige.

En résumé, les caractéristiques de la voiture constituent le principal critère de choix du véhicule pour les clients cibles, masculins, appartenant au cluster-3. Ils accordent en revanche une importance moyenne au prestige que procure l'automobile et très peu pour le critère financier relatif à cette dernière.

Le deuxième segment ciblé pour l'achat de la Megane 4, c'est-à-dire, la clientèle féminine appartenant au cluster-5, accordent pratiquement le même intérêt aux caractéristiques de la voiture que le segment précédent mais est nettement plus intéressée par le prestige que la moyenne. Quant au critère financier, il est presque aussi important pour cette cible que la moyenne des clients de l'entreprise.

- Canaux de communication proposés pour chaque segment cible :

Dans le but de contribuer à la réussite des campagnes marketing de l'entreprise en général et celles relatives au lancement de la nouvelle Megane, en particulier, nous identifions dans la présente partie les canaux de communication adéquats pour chaque segment de clients.

Pour ce faire, nous procédons à une analyse comparative des segments en fonction des variables préalablement calculées, à savoir, Info_Traditionnelle, Info_Connexion et Info_Avis, de la même façon que celle utilisée pour les critères de choix du véhicule.

Ainsi, la figure III.19 suivante présente les moyennes réduites de chaque segment en fonction des différentes sources d'information.

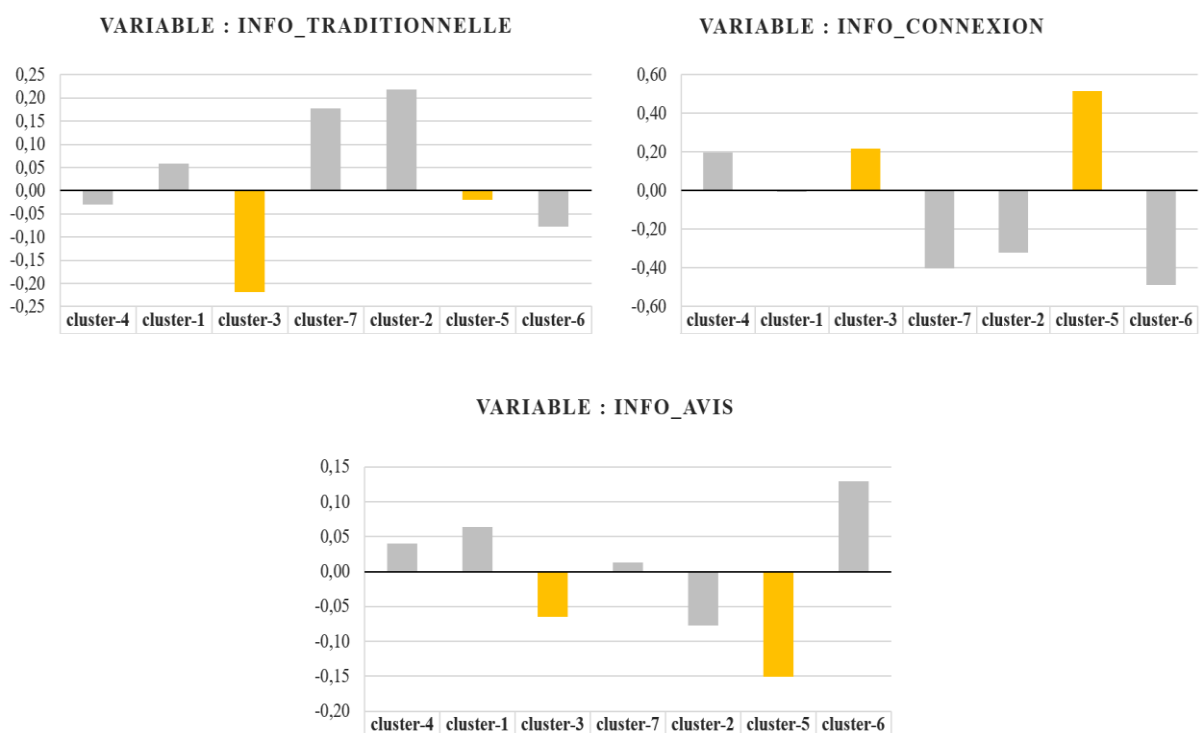


Tableau : Moyenne d'utilisation des sources d'informations sur 5.

Variable	Info_Traditionnelle	Info_Connexion	Info_Avis
Moyenne	2,64	2,87	3,31

Figure III. 19 : Comparaison des segments clients en fonction des sources d'informations.

Nous constatons, d'une part, que les individus appartenant au premier segment cible représenté par le cluster-3, sont les clients qui utilisent le moins les sources d'information traditionnelles pour se renseigner sur les voitures, à savoir, les affiches publicitaires, la TV, les journaux et la Radio. Ces derniers ont également moins recours aux avis des proches et/ou concessionnaires que la moyenne. Par contre, ils utilisent internet à cet effet, plus que la moyenne.

D'autre part, le second segment cible, représenté par le cluster-5, utilise nettement plus les réseaux sociaux, le site web de la marque ainsi que d'autres sites web que la moyenne, pour se renseigner sur les voitures. En revanche, ces clients sont ceux qui se fient le moins aux avis des proches et des concessionnaires sur les voitures. Leur utilisation des sources d'information traditionnelles est cependant proche de la moyenne.

En résumé, le canal de communication le plus pertinent pour les campagnes marketing relatives à la promotion du nouveau produit est le canal digital. Par ailleurs, les affiches publicitaires, les journaux, la TV et la radio peuvent s'avérer efficaces auprès de la clientèle féminine désireuse d'acheter la nouvelle Megane.

Section II : Segmentation du marché automobile algérien

Dans un contexte économique difficile, marqué par la baisse du prix du pétrole, l'Etat algérien se voit contraint de réduire ses factures d'importation, notamment celles relatives à l'automobile. Une situation qui oblige les constructeurs automobiles à investir dans la production nationale des voitures.

Le leader du marché, Renault Algérie, est le premier constructeur automobile à avoir implanté une usine de production dès la fin de l'année 2014 pour la production de la Symbol Made in Bladi.

Plus d'un an plus tard, l'entreprise prévoit déjà le lancement d'autres modèles, à commencer par la Dacia Sandero, prévu pour l'année en cours.

Face à la réduction du chiffre d'affaires, causée par la limitation des importations, il devient nécessaire d'optimiser les lancements de ces produits considérés nouveaux, étant donné que la production nationale de voitures est un phénomène inédit dans le marché automobile algérien.

Dans le but de soutenir l'entreprise Renault Algérie dans sa stratégie de lancement des voitures fabriquées dans le pays, nous procédons à une étude de marché visant à identifier les segments d'acheteurs de voitures les plus favorables à l'achat d'une voiture fabriquée en Algérie en fonction de leurs caractéristiques sociodémographiques mais aussi en prenant en considération leurs comportements, besoins et attentes.

A cet effet, la troisième section de ce chapitre sera organisée selon les 6 étapes du processus *Knowledge Discovery* évoquées précédemment, dans le but d'analyser l'étude du marché réalisée auprès des acheteurs de voitures, de toutes marques confondues.

1. Collecte des données

La collecte des données relatives aux acheteurs de voitures en Algérie s'est faite à travers un sondage.

a) Choix de la méthode de sondage :

En nous appuyant sur le tableau comparatif des différentes méthodes de sondage Tableau III.1 présenté dans la section précédente et en prenant en considération le temps limité dont nous disposons pour la réalisation de notre étude, nous décidons de procéder à un sondage par internet.

b) Conception du questionnaire :

Les étapes de conception du questionnaire adoptées pour la présente étude sont les mêmes que celles suivies dans le cadre de l'étude précédente, à quelques différences près.

En effet, internet étant la seule méthode de sondage utilisée pour cette étude, les contraintes liées au sondage par téléphone telles que le temps moyen d'appel et le manque de support visuel pour le répondant, ne sont plus valides, ce qui nous permet d'enrichir un peu plus notre questionnaire.

Toutefois, parmi les principaux inconvénients d'un sondage par internet est le très faible taux de réponse. Pour remédier à cela, nous décidons, d'une part, de multiplier les sites et les réseaux sociaux dans lesquels nous partageons le questionnaire, et d'autre part, de prévoir une durée de 5 semaines pour la récolte des données.

Le questionnaire est réalisé avec Google Forms.

c) Test du questionnaire :

Avant le lancement de l'enquête en ligne, nous procédons au test du questionnaire auprès de nos proches, mais aussi auprès d'une des communautés d'un réseau social, et ce, pour une durée de 5 jours.

Ce test nous a permis de récolter les réponses de 32 participants et d'apporter ainsi les ajustements nécessaires au questionnaire en fonction de leurs feedbacks.

d) Collecte des données :

Une fois la validation du questionnaire ²¹terminée, il est mis en ligne pour une durée de 5 semaines. Durant cette période 473 réponses sont collectées.

2. Sélection des données

La présente étude consiste à étudier le comportement et les différentes caractéristiques des acheteurs de voitures en Algérie, nous choisissons donc de sélectionner, parmi les 473 répondants, ceux possédant une voiture. Ceci réduit le nombre des réponses à 316, soit 66,8% du total des réponses récoltées.

De plus, 6 doublons sont détectés et supprimés. Ce qui nous amène à un total de 310 réponses, soit 65,5% des réponses récoltées.

3. Préparation des données

Grâce à l'utilisation de questions structurées ainsi que la possibilité de recourir au critère « obligatoire » sur l'ensemble des questions sur Google Forms, nous n'avons pas rencontré de problèmes concernant les données manquantes ou invalides.

En fait, la seule préparation des données effectuée concerne l'intégration des réponses issues des différents sites, forums et réseaux sociaux, dans un fichier SPSS unique.

²¹Le fichier final du questionnaire ainsi qu'un aperçu du formulaire sont présentés en annexe 10 et 11, respectivement.

4. Transformation des données

Avant d'entamer les analyses, il convient de transformer les données de façon à ce qu'elles puissent être utilisées par les différentes méthodes et techniques de Data Mining.

La première transformation apportée concerne le codage des données sur IBM SPSS Statistics, comme présenté dans la figure III.20.

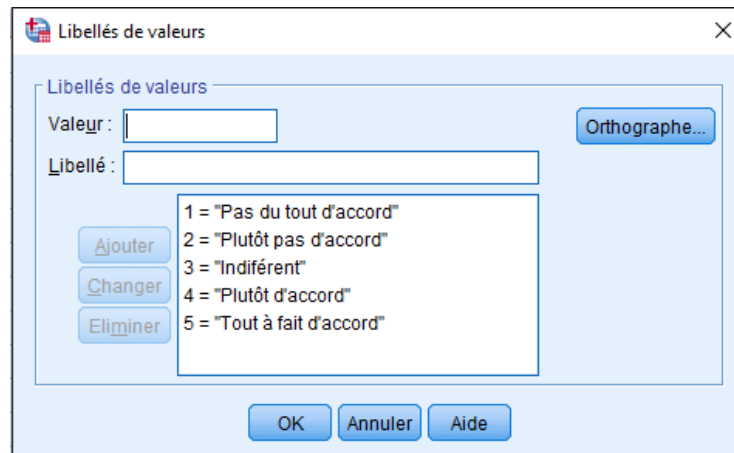


Figure III. 20: Codage des variables sur IBM SPSS Statistics.

Par ailleurs, de nouvelles variables sont créées pour les besoins de l'étude. Ainsi, la variable « AgeVoiture » est créée à partir de la variable « AnnéeImmatriculation » comme suit :
 $AgeVoiture = 2016 - AnnéeImmatriculation$.

La variable « APV », mesurant nombre d'opérations après-ventes réalisées chez le concessionnaire/constructeur, est également créée comme suit :

$$APV = (1 - Aucune) \cdot ((Maintenance + Réparation + ChangementDesPneus + Carrosserie) / 4).$$

La variable ProductionNationale est créée afin d'identifier les automobilistes favorables à l'achat d'une voiture fabriquée en Algérie, et ce, à partir des deux variables mesurant l'accord avec les propositions « Les voitures fabriquées à l'étranger sont de meilleure qualité » et « Ça ne me dérangerait pas d'acheter une voiture fabriquée en Algérie si elle coûte moins chère ».

$$ProductionNationale = (PrixVoitureAlgerie - VoitureEtrangeres) / 2.$$

Une fois la variable continue ProductionNationale calculée, nous discrétisons cette dernière afin de la transformer en une variable binaire indiquant si le consommateur est favorable ou non à l'achat d'une voiture fabriquée en Algérie. La nouvelle variable, ProdBinaire est créée selon le code suivant : `If ProductionNationale > 0 then 1 else 0 endif`. Ainsi, une valeur 1 pour la variable ProdBinaire signifie que le consommateur est favorable à l'achat d'une voiture fabriquée en Algérie.

Une nouvelle variable appelée Segment est introduite afin de regrouper les modèles de voitures des répondants par segments automobiles. A cet effet, chacun des modèles renseignés dans la variable non structurée Modèle est normalisé avant d'être classé dans l'un des segments standards de l'automobile, reposant sur les dimensions des voitures.

De plus, nous décidons de regrouper les 48 wilayas du pays en 4 zones : Centre, Est, Ouest et Sud, dans le but de réduire les modalités de la variable Région.

Un exemple de la procédure de création d'une nouvelle variable est représenté dans La figure III.21.

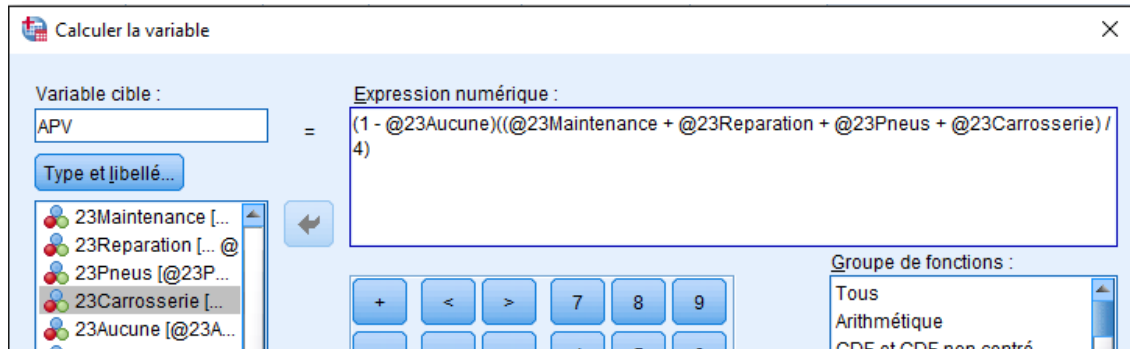


Figure III. 21: Calcul d'une variable sur IBM SPSS Statistics.

Enfin, en raison du nombre important de variables utilisées, des analyses factorielles sont réalisées afin d'identifier les relations entre les différentes variables et pouvoir ainsi réduire les dimensions organisant ces relations.

La première analyse factorielle employée concerne les besoins et attitudes du consommateur capturées à l'aide de 14 variables catégorielles en rapport avec les avis des acheteurs sur l'automobile ainsi que leurs critères de choix d'une voiture.

Les données étant quantitatives, nous utilisons une ACP avec une rotation orthogonale Varimax qui simplifie l'interprétation des facteurs, et ce, à l'aide du logiciel IBM SPSS Statistics.

Dans le volet Options de la fenêtre Analyse Factorielle, nous choisissons de cocher la case « Supprimer les faibles coefficients » (inférieurs à 0,1) de la partie Format d'affichage des coefficients, afin de simplifier la visualisation des résultats.

Le volet Extraction nous permet de sélectionner le critère de choix du nombre de composantes à retenir. Dans notre cas, nous décidons de ne retenir que les axes dont l'inertie est supérieure à 1 (Critère de Kaiser).

Avant d'analyser les résultats de l'ACP, il est nécessaire d'évaluer l'indice de Kaiser-Meyer-Olkin (KMO). Dans notre cas, l'indice KMO est de 0,753, nous pouvons donc procéder à une analyse factorielle de type ACP dans le but de réduire le nombre de variables de notre modèle.

Les résultats de l'ACP indiquent que 4 composantes sont sélectionnées, avec une valeur propre cumulée de 67,51 % de la valeur propre initiale. La plus faible qualité de représentation concerne la variable Prix pour une valeur acceptable de 0,634.

A partir de la rotation de la matrice des composantes, présentée dans la figure III.22, nous pouvons déterminer les variables ayant le plus contribué à la construction de chaque composante dans le but de les regrouper en un nombre réduit de variables.

Rotation de la matrice des composantes^a

	Composante			
	1	2	3	4
Une voiture est plus un moyen de transport qu'une source de plaisir.	-,248	,533	,255	,138
La voiture que je conduis doit refléter ma position sociale.				,829
Prix			,768	
Marque	,600	-,168	,275	
Consommation	-,144	,447	,587	
StatutSocial	,295			,741
Puissance	,677	-,195		
Design	,628		-,193	,186
Confort	,686	,241		,127
Securite	,241	,698		
Revente	,223	,228	,598	
Equipement	,704	,168		,175
Nouveaute	,513	,313		,354
Environnement		,729	,124	-,130

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.

Figure III. 22: Résultat de la rotation de la matrice des composantes principales sur IBM SPSS Statistics.

A partir de la figure précédente, nous pouvons constater que les variables : Marque de la voiture, Puissance du moteur, Design, Confort intérieur, Niveau d'équipements et Nouveauté du modèle sont les plus corrélées au premier axe. Nous en déduisons donc que la première composante principale mesure l'importance accordée par les acheteurs aux caractéristiques de la voiture, ce qui nous permet de créer la variable CARACTERISTIQUES comme suit :

$$\text{CARACTERISTIQUES} = (\text{Marque} + \text{Puissance} + \text{Design} + \text{Confort} + \text{Equipement} + \text{Nouveaute}) / 6.$$

En analysant le deuxième axe, nous déduisons que la deuxième composante principale mesure la perception de la voiture comme un simple moyen de transport de la part des automobilistes ainsi que le degré d'importance qu'ils accordent aux critères : Sécurité et Respect de l'environnement, dans le choix d'une voiture. La relation existante entre ces trois variables signifie que la deuxième composante mesure la vision utilitaire des répondants à l'égard de la voiture et nous permet de créer la variable UTILITAIRE comme suit :

$$\text{UTILITAIRE} = (\text{MoyenDeTransport} + \text{Securite} + \text{Environnement}) / 3.$$

Les variables Prix, Consommation de carburant et Valeur de la voiture à la revente contribuent majoritairement à la constitution du troisième axe factoriel. Nous en déduisons que la troisième composante principale mesure la considération financière des répondants lors du choix d'une voiture. Ainsi, la variable FINANCE est créée comme suit :

$$\text{FINANCE} = (\text{Prix} + \text{Consommation} + \text{Revente}) / 3.$$

Enfin, le degré d'accord avec la proposition « ma voiture doit refléter ma position sociale » ainsi que le critère de choix Statut Social constituent les deux variables les plus corrélées au quatrième axe factoriel. Ceci indique que ce dernier mesure le prestige attendu par les répondants à travers l'acquisition d'une voiture. La variable PRESTIGE est ainsi créée comme suit :

$$\text{PRESTIGE} = (\text{PositionSociale} + \text{StatutSocial}) / 2.$$

Nous procédons de la même manière pour l'analyse factorielle des sources d'informations utilisées lors du choix d'une voiture.

L'indice KMO relatif à cette deuxième ACP est estimé à 0,771, ce qui nous permet de réaliser l'analyse factorielle.

Les résultats de l'ACP montrent la sélection de 4 composantes principales représentant une valeur propre cumulée de 64,93 %. La plus faible qualité est de 0,669 et elle correspond à la variable SalonAutomobile.

La figure III.23 présente les résultats de la rotation de la matrice des composantes principales de la présente ACP.

	Composante			
	1	2	3	4
5Affiches	,730			,295
5Journaux	,801	,127		
5TV	,677	,354		
5Radio	,682		,120	,328
5ReseauxSociaux	,310	,662	,159	-,163
5SiteMarque		,732	-,163	,455
5AutreSite		,803	,132	
5SalonAuto	,208	,128		,720
5MarcheOccasion	,115	,159	,808	
5Proches			,794	,255
5Concessionnaire	,131		,404	,707

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.
a. Convergence de la rotation dans 7 itérations.

Figure III. 23: Résultat de la rotation de la matrice des composantes sur IBM SPSS Statistics.

Les variables les plus corrélées à la première composante principale sont les variables : Affiches, Journaux, TV et Radio, soit, des sources d'informations considérées « traditionnelles ». Cette relation entre les 4 variables précédemment citées, nous permet de créer la variable InfoTraditionnelles comme suit :

$$\text{INFO_TRADITIONNELLES} = (\text{Affiches} + \text{Journaux} + \text{TV} + \text{Radio}) / 4.$$

Par ailleurs, les variables représentant les réseaux sociaux / Forums / Blogs, le site Web de la marque de la voiture ainsi que d'autres sites Web relatifs à l'automobile sont les plus corrélées à la deuxième composante principale. La relation entre ces trois variables mesure le degré d'utilisation d'internet par les acheteurs de voitures dans la recherche d'informations sur l'automobile. La variable INFO_CONNEXION est alors créée comme suit :

$$\text{INFO_CONNEXION} = (\text{ReseauxSociaux} + \text{SiteMarque} + \text{AutreSite}) / 3.$$

Les variables les plus corrélées au troisième axe factoriel sont le marché des voitures d'occasion et l'avis des proches. Les répondants ayant recours à ces deux sources d'informations sont plus sensibles aux avis non professionnels de leur entourage. Les deux variables sont ainsi regroupées en une variable créée à cet effet : AVIS_NONPRO.

$$\text{AVIS_NONPRO} = (\text{MarcheOccasion} + \text{AvisProches}) / 2.$$

Enfin, les variables les plus corrélées à la quatrième composante principale sont le salon automobile et l'avis des concessionnaires. En opposition au troisième axe, les sources d'informations utilisées dans ce cas sont basées sur les avis des professionnels du secteur automobile. La variable AVIS_PRO est créée comme suit :

$$\text{AVIS_PRO} = (\text{SalonAuto} + \text{AvisConcessionnaires}) / 2.$$

5. Application du Data Mining

Rappelons que l'objectif de la présente étude est de segmenter le marché automobile algérien en fonction des comportements et besoins des acheteurs, avant de réaliser un Profiling des segments ciblés, c'est-à-dire, ceux avec une opinion positive sur la production nationale.

Cependant, avant d'entamer l'utilisation des techniques de Data Mining, nous procédons tout d'abord à une analyse descriptive par des méthodes statistiques univariées et bivariées pour une meilleure compréhension des données. Les résultats des statistiques descriptives sont présentés en annexe 12.

a) Discussion sur le choix des méthodes de Data Mining :

La segmentation du marché consiste à diviser les consommateurs en groupes homogènes avec des envies et des besoins différents. Le but étant, pour les entreprises, de proposer des offres adéquates à chaque segment et gagner ainsi en efficacité.

Le Data Mining propose, à cet effet, une multitude de méthodes et d'algorithmes qui permettent de classer des objets ou des individus partageant des caractéristiques similaires dans des groupes distincts.

Ces techniques sont souvent classées en deux catégories, expliquées dans le Chapitre I, qui sont les méthodes de classement et les méthodes de classification.

Etant donné que nous visons à créer des segments homogènes d'individus plutôt que de classer ces derniers dans des segments préalablement définis, les méthodes de classification non supervisées sont les plus appropriées pour notre étude.

Plusieurs algorithmes de classification existent à ce jour et chacun présente une approche différente pour regrouper les individus. Le nœud « Classification non supervisée automatique » d'IBM SPSS Modeler présente les résultats des différents modèles de classification, adéquats au problème traité. Pour les départager, nous utilisons l'indice Silhouette, présenté dans le Chapitre I, qui permet d'évaluer la qualité de la segmentation.

Pour répondre à l'objectif de l'étude de marché, nous commencerons par réaliser une première segmentation en fonction du comportement d'achat de voitures et de services après-vente des consommateurs. Cette classification sera ensuite complétée par une deuxième segmentation basée sur les besoins et les attentes des automobilistes ainsi que leur opinion quant à l'achat d'une voiture produite en Algérie. Ceci nous permettra d'identifier les segments cibles pour le lancement des nouveaux produits issus de la production nationale.

Enfin, nous développerons un modèle prédictif utilisant principalement les données sociodémographiques des consommateurs afin de classer ces derniers dans les segments précédemment déterminés.

b) Classification en fonction du comportement des individus :

Cette première classification vise à segmenter les individus selon le comportement des individus concernant l'achat de voitures et de services après-vente.

Pour cela, en plus des variables liées aux caractéristiques de la voiture, c'est-à-dire, si la voiture est achetée neuve ou d'occasion, l'âge de cette dernière (calculé précédemment à partir de l'année d'immatriculation), le segment auquel elle appartient ainsi que le carburant et le type de boîte à vitesse de la voiture, on ajoute les variables qui décrivent le comportement des individus avant l'achat effectif, c'est-à-dire sa participation ou non au choix du modèle ainsi que le temps de décision pris pour passer commande.

De plus, afin de mesurer l'achat des services après-vente auprès des constructeurs et concessionnaires, nous ajoutons la variable APV, calculée précédemment.

L'ensemble de ces données sont utilisées en variables d'entrée dans le nœud classification non supervisée automatique d'IBM SPSS Modeler. Dans l'onglet Modèle du nœud, nous décidons de classer les modèles par le critère Silhouette afin de déterminer le modèle présentant la meilleure qualité de classification. Nous choisissons également de garder les paramètres par défaut des trois modèles de classification : Two-Step, Réseaux de Kohonen et K-means.

La figure III. 24 illustre le modèle réalisé à l'aide du logiciel IBM SPSS Modeler.

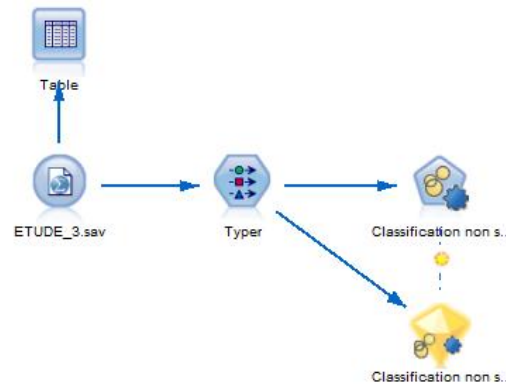


Figure III. 24: Première classification non supervisée automatique IBM SPSS Modeler.

Nous décidons de faire appel au nœud de classification automatique car l'un des avantages majeurs qu'il présente est la possibilité de comparer la qualité mais aussi les résultats de différents algorithmes de classification.

Les résultats obtenus après l'exécution du nœud indiquent que le modèle K-means avec 5 clusters présente le meilleur indice Silhouette, estimé à 0,341. Une qualité jugée correcte, comme le montre la figure III.25 suivante :

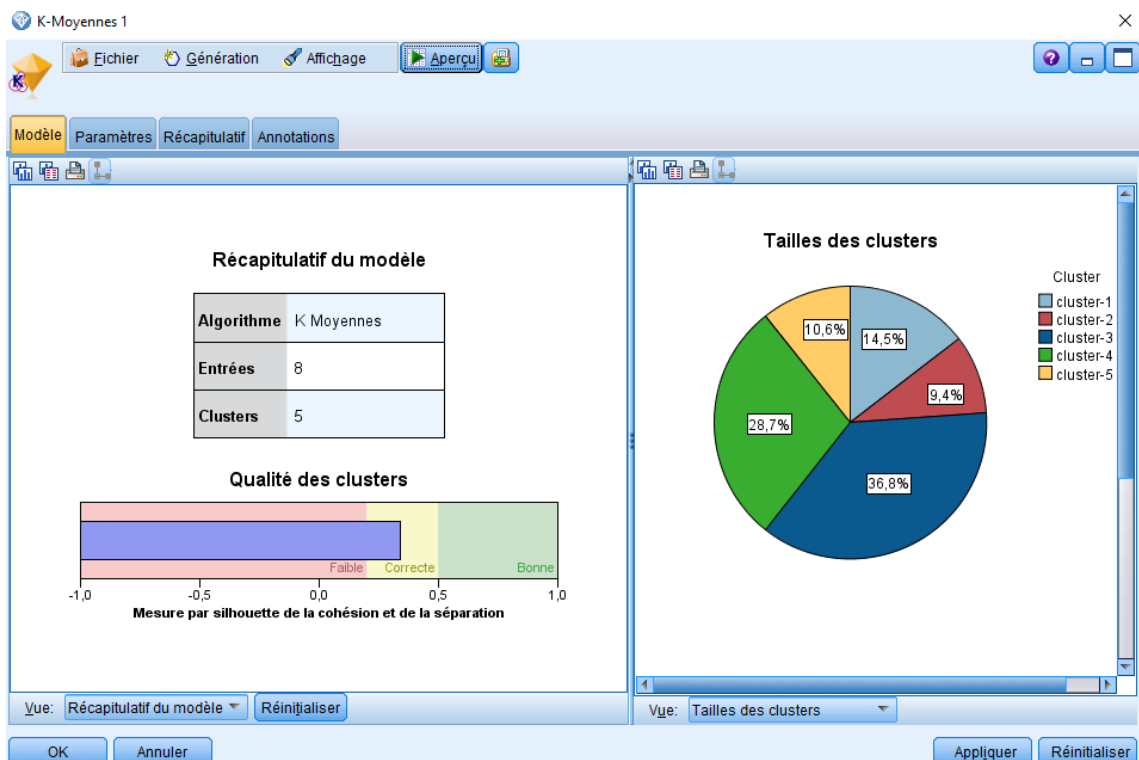


Figure III. 25: Résultats du modèle K-means à 5 clusters.

Une fois le modèle théorique validé par l'indice Silhouette, nous nous intéressons plus en détail à la signification pratique de la classification.

Tout d'abord, nous remarquons que les variables VoitureNeuveOccasion et TypeBoiteVitesse sont les plus importantes dans l'estimation du modèle. Le reste du classement se présente comme suit :

- La participation ou non au choix du modèle (0,53).
- Le type de carburant utilisé (0,47).
- L'âge de la voiture (0,15).
- Le segment auquel appartient le modèle (0,07).
- La variable APV mesurant l'achat des services après-vente auprès des concessionnaires et constructeurs automobiles (0,05).
- Le temps entre la décision d'achat et le moment de la commande (0,01).

Le cluster 3 représente la classe la plus grande avec un effectif de 114, soit 36,8% de l'effectif total, tandis que le cluster 2 ne contient que 29 individus, soit 9,4% du total.

Les clusters 2 et 4 sont constitués, à 100%, d'acheteurs de véhicules d'occasions avec une majorité équipés de boîtes à vitesse manuelles. Cependant, le cluster 2 contient à 100%, des personnes n'ayant pas participé au choix du modèle de leur voiture.

Les clusters 1, 3 et 5 restants se constituent, majoritairement, d'acheteurs de véhicules neufs ayant participé au choix du modèle. Le cluster 5 représente la seule classe formée uniquement d'individus possédant des voitures avec des boîtes à vitesse automatiques. Il détient également la meilleure moyenne APV, égale à 0,34.

Les caractéristiques des 5 clusters issus de cette classification sur IBM SPSS Modeler, sont représentés dans le tableau III.5 suivant :

Tableau III. 5: Clusters obtenus après la première classification_ Algorithme K-means.

Cluster	Cluster-1	Cluster-2	Cluster-3	Cluster-4	Cluster-5
Taille en pourcentage (effectif)	14,5% (45)	9,4% (29)	36,8% (114)	28,7% (89)	10,6% (33)
VoitureNeuveOccasion-Catégorie la plus fréquente (Pourcentage)	Neuve (100%)	Occasion (100%)	Neuve (100%)	Occasion (100%)	Neuve (93,9%)
TypeBoiteVitesse-Catégorie la plus fréquente (Pourcentage)	Manuelle (100%)	Manuelle (75,9%)	Manuelle (100%)	Manuelle (92,1%)	Automatique (100%)
ParticipationChoix-Catégorie la plus fréquente (Pourcentage)	Oui (91,1%)	Non (100%)	Oui (86%)	Oui (100%)	Oui (87,9%)
TypeCarburant-Catégorie la plus fréquente (Pourcentage)	Diesel (100%)	Essence (62,1%)	Essence (98,2%)	Essence (61,8%)	Diesel (57,6%)
AgeVoiture-Médiane	3 ans	6 ans	3 ans	4 ans	1an
Segment-Catégorie la plus fréquente (Pourcentage)	Compacte (51,4%)	Petite Citadine (34,8%)	Citadine (64,8%)	Citadine (51,5%)	Citadine (37%)
APV-Moyenne	0,27	0,22	0,21	0,15	0,34
TempsDecision-Médiane	Moins d'un mois	Moins de deux semaines	Moins d'un mois	Moins d'un mois	Moins d'un mois

En analysant ces caractéristiques, nous constatons que les résultats pratiques présentés par le modèle sont cohérents, ce qui nous permet, en plus des résultats théoriques, de valider le modèle de classification.

c) Classification en fonction des besoins des individus :

Dans le but d'améliorer notre première segmentation, nous décidons de la compléter en introduisant les besoins des consommateurs ainsi que leur avis sur la production nationale de voitures, dans une seconde classification non supervisée.

Les variables d'entrée du nouveau modèle deviennent alors : Le numéro de cluster issu de la première classification K-means, les 4 variables issues de l'analyse factorielle réalisée précédemment, qui sont : CARACTERISTIQUES, UTILITAIRE, FINANCE et PRESTIGE ainsi que la variable binaire ProdBinaire.

Nous procédons de la même manière que pour la première classification, c'est-à-dire que nous faisons appel au nœud de classification non supervisée automatique afin de comparer les qualités de classification des différents algorithmes.

La figure III.26 suivante illustre le modèle de la deuxième classification à l'aide d'IBM SPSS Modeler.

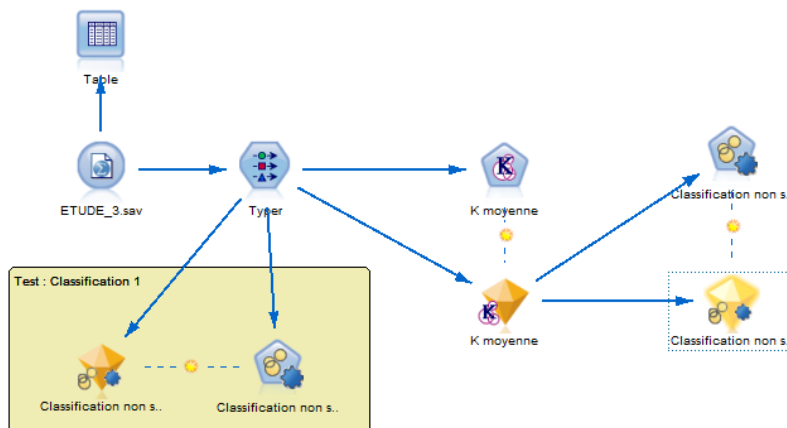


Figure III. 26: Deuxième classification non supervisée automatique_IBM SPSS Modeler.

L'exécution du nœud indique que l'algorithme Two-step présente la meilleure qualité de classification avec un indice Silhouette de 3,26 jugé correct.

Les résultats de la classification sont présentés dans la figure III.27.

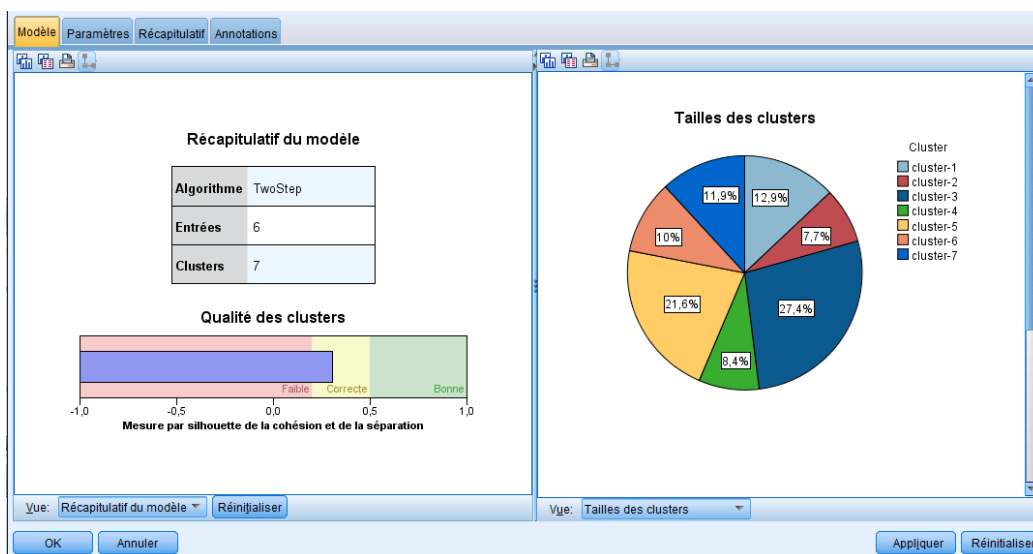


Figure III. 27: Résultats du modèle TwoStep à 7 clusters.

La classification par la méthode Two-Step produit 7 clusters dont le plus grand contient 85 individus, soit 27,4% du total d'observations, tandis que le plus petit regroupe 24 individus, ce qui équivaut à 7,7% de l'ensemble des observations.

La variable la plus importante pour l'estimation du modèle de classification représente le numéro du cluster issu de la première segmentation K-means. Le reste du classement des variables selon leur importance se présente comme suit :

- ProdBinaire.
- FINANCE.
- CARACTERISTIQUES.
- UTILITAIRE.
- PRESTIGE.

Etant donné que la variable représentant le numéro de cluster K-means est la variable la plus importante dans la construction du modèle, chacun des nouveaux clusters obtenus est majoritairement formé à partir des éléments d'un des clusters K-means.

Ainsi, les clusters 4,6 et 7 sont formés, à hauteur d'au moins 90%, des individus des clusters K-means 2, 5 et 1 respectivement, tandis que les clusters 3 et 1 et les clusters 5 et 2 se partagent les individus des segments K-means 3 et 4, respectivement.

Par ailleurs, les caractéristiques des différents segments sont représentées dans le tableau III.6 suivant :

Tableau III. 6: Clusters obtenus après la seconde classification_ Algorithme TwoStep.

Cluster	Cluster-3	Cluster-5	Cluster-1	Cluster-7	Cluster-6	Cluster-4	Cluster-2	Global
Taille en pourcentage (effectif)	27,4% (85)	21,6% (67)	12,9% (40)	11,9% (37)	10,0% (31)	8,4% (26)	7,7% (24)	310
Cluster KM-Catégorie la plus fréquente (Pourcentage)	cluster-3 (100%)	cluster-4 (100%)	cluster-3 (70%)	cluster-1 (100%)	cluster-5 (90,3%)	cluster-2 (100%)	cluster-4 (91,7%)	/
ProdBinaire-Catégorie la plus fréquente (Pourcentage)	Non (100%)	Non (100%)	Oui (100%)	Non (100%)	Non (100%)	Non (100%)	Oui (100%)	Non (79,4%)
FINANCE-Moyenne sur 5	4,46	4,36	4,50	4,49	3,71	4,62	4,62	4,40
CARACTERISTIQUES-Moyenne sur 5	4,04	4,18	4,22	4,32	4,28	4,13	3,83	4,14
UTILITAIRE-Moyenne sur 5	3,78	3,35	3,65	3,73	3,52	3,54	3,62	3,60
PRESTIGE-Moyenne sur 5	3,59	3,53	3,60	3,71	3,82	3,51	3,40	3,60

L'analyse des différents segments obtenus nous permet de constater que les clusters les plus intéressants pour notre objectif d'étude sont les clusters 1 et 2, puisqu'ils sont les seuls segments constitués d'individus ayant une opinion positive sur la production nationale des voitures. Ils représentent, à eux deux, 20,6% de l'ensemble de la population étudiée.

Le Profiling des consommateurs ciblés est un outil important qui permet d'adapter les actions de l'entreprise en fonction des caractéristiques des clients ou prospects, et donc, de bâtir la meilleure offre commerciale possible et d'augmenter la pertinence des stratégies marketing mises en place. C'est d'ailleurs conscients de cet avantage concurrentiel non négligeable que nous présentons, dans l'étape d'interprétation des résultats, le Profiling des deux segments ciblés.

d) Modèle d'estimation :

Les deux modèles de classification précédents nous ont servi à rechercher des classes « naturelles » dans le secteur de l'automobile en Algérie afin de pouvoir cibler les segments les plus intéressants pour notre problématique. Leur établissement a toutefois nécessité la collecte de données à partir de 27 variables relatives aux critères comportementaux et attitudeux précis des répondants.

Exception faite des secteurs tels que la télécommunication, la banque et le commerce en ligne, il est généralement très difficile de se procurer ce type d'informations sur des clients, et encore plus sur l'ensemble du marché.

C'est en fait la raison pour laquelle les experts font souvent appel aux études de marché, qui permettent de généraliser les modèles générés à partir des données d'un échantillon représentatif de la population.

Souvent, les modèles de classification construits à partir des études de marché permettent de préparer un travail de classement des individus dans les classes précédemment définies. Pour cela, on fait souvent appel aux modèles d'estimation.

En fait, comme expliqué dans le Chapitre I, les méthodes d'estimation peuvent réaliser une tâche de classement en utilisant une variable cible qui permet de classer les individus dans les différentes classes.

L'intérêt serait toutefois d'établir un modèle suffisamment efficace et utilisant des données plus faciles à collecter que celles utilisées pour la construction des modèles de classification. Ainsi, pour les prédicteurs, nous utilisons les données socioéconomiques et démographiques des individus ainsi que deux autres variables qui sont la marque de la voiture et son type (achetée neuve ou d'occasion).

Nous utilisons pour cela, le nœud classificateur automatique du logiciel IBM SPSS Modeler, qui permet de comparer les différents algorithmes de classement selon leur précision. Les résultats obtenus à travers l'exécution de ce nœud indiquent que le modèle de régression logistique présente la meilleure performance.

En effet, la régression logistique, contrairement aux autres méthodes de régression, est adaptée dans ce cas étant donné que la variable cible que nous cherchons à prédire est une variable qualitative nominale qui représente le numéro de cluster auquel devrait appartenir l'individu. La variable cible étant nominale à plus de 2 modalités, nous sélectionnons, dans le nœud régression logistique de d'IBM SPSS Modeler, la procédure multinomiale.

Cependant, les données sociodémographiques des individus présentent souvent des corrélations importantes entre les variables, pour y remédier, nous exécutons une analyse factorielle, qui nous permet de présenter les variables dans des facteurs indépendants, que nous utilisons en entrée du modèle. Les résultats de l'analyse factorielle réalisée indiquent que 5 facteurs sont retenus, sachant que le critère de choix du nombre de facteurs sélectionné est celui de Kaiser pour une variance cumulée de 65,31%.

Après l'exécution du modèle de régression, nous ajoutons le nœud de sortie Analyse afin d'évaluer la précision globale du modèle. La figure III. 28 présente le modèle réalisé sur le logiciel IBM SPSS Modeler.

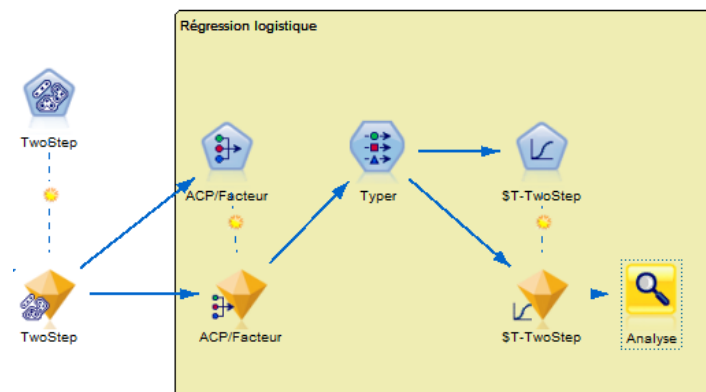


Figure III. 28: Modélisation de la régression logistique sur IBM SPSS Modeler.

L'exécution du nœud d'analyse présente une précision globale de 83,55% pour le modèle, jugée convenable pour le domaine de notre étude.

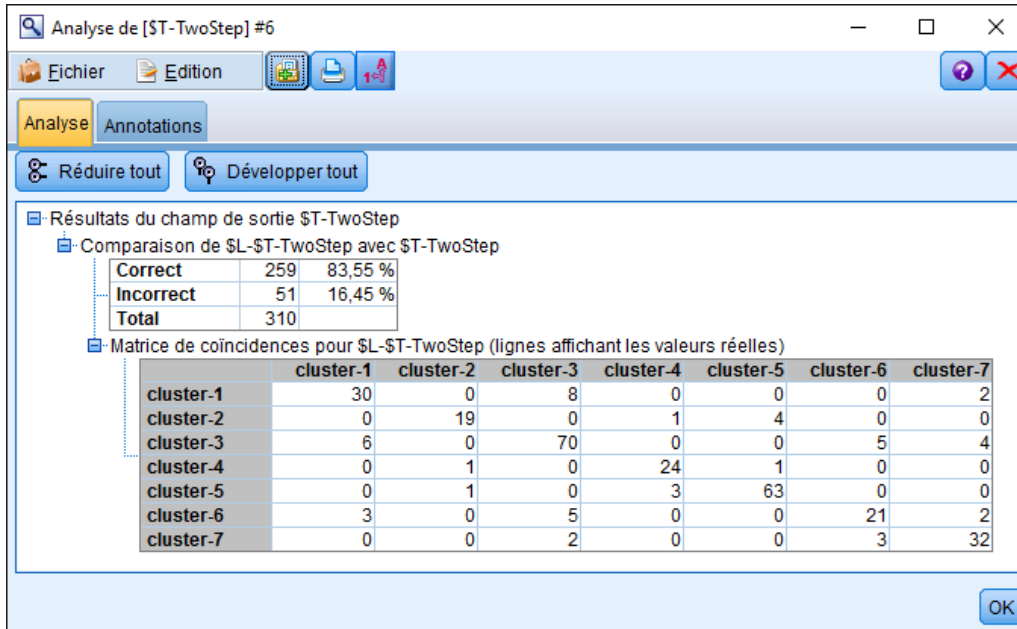


Figure III. 29: Résultat de l'analyse du modèle de régression logistique.

6. Interprétation des résultats

Nous avons réalisé un sondage dans le but de segmenter le marché algérien de l'automobile et de proposer ainsi une stratégie de ciblage à l'entreprise Renault Algérie pour le lancement de ses produits fabriqués en Algérie.

Pour ce faire, nous avons procédé à une double classification basée, en premier lieu, sur les données comportementales des automobilistes et, en second lieu, sur leurs besoins et attitudes vis-à-vis de la production nationale.

A l'issue de la segmentation, nous avons obtenu 7 différents clusters, auxquels nous avons appliqué un modèle de régression logistique utilisant, principalement, des données socioéconomiques et démographiques afin de permettre le classement des individus dans les différents segments, avec une précision de plus de 83%.

Dans cette sixième et dernière étape, nous établissons des profils démographiques détaillés des segments complétés par une analyse comportementale et attitudinale vis-à-vis de l'automobile. Une stratégie de communication pour les deux segments ciblés sera également proposée afin de contribuer à l'optimisation des actions marketing de l'entreprise.

- Profiling des segments :

Le choix de compléter le Profiling démographique des consommateurs par une analyse comportementale et attitudinale se justifie par le fait que des individus au sein d'une même

catégorie démographique peuvent présenter des profils comportements et des besoins très différents.

Dans le tableau III.7 nous résumons les principales catégories démographiques et socioéconomiques des segments obtenus.

Tableau III. 7: Principales catégories démographiques et socioéconomiques des 7 segments.

Cluster	Cluster-1	Cluster-2	Cluster-3	Cluster-4	Cluster-5	Cluster-6	Cluster-7	Global
Taille en pourcentage (effectif)	12,9% (40)	7,7% (24)	27,4% (85)	8,4% (26)	21,6% (67)	10,0% (31)	11,9% (37)	(310)
Genre -Catégorie la plus fréquente (Pourcentage)	Homme (92,5%)	Homme (95,8%)	Homme (61,2%)	Homme (65,4%)	Homme (82,1%)	Homme (74,2%)	Homme (78,4%)	Homme (76,1%)
Age -Catégorie la plus fréquente (Pourcentage)	25-34 (62,5%)	18-24 (41,7%)	25-34 (34,1%)	25-34 (42,3%)	25-34 (44,8%)	25-34 (48,4%)	18-24 (48,6%)	25-34 (41,0%)
Profession -Catégorie la plus fréquente (Pourcentage)	Cadre supérieur (27,5%)	Etudiant (37,5%)	Etudiant (25,9%)	Etudiant (30,8%)	Employé (25,4%)	Cadre supérieur (45,2%)	Etudiant (37,8%)	Etudiant (25,2%)
Niveau d'éducation -Catégorie la plus fréquente (Pourcentage)	Supérieur (97,5%)	Supérieur (91,7%)	Supérieur (96,5%)	Supérieur (80,8%)	Supérieur (82,1%)	Supérieur (96,8%)	Supérieur (86,5%)	Supérieur (90,6%)
Situation familiale -Catégorie la plus fréquente (Pourcentage)	Marié(e) (52,5%)	Célibataire (54,2%)	Célibataire (62,4%)	Marié(e) (53,8%)	Célibataire (58,2%)	Célibataire (51,6%)	Célibataire (59,5%)	Célibataire (55,8%)
Nombre d'enfants -Catégorie la plus fréquente (Pourcentage)	0 (60,0%)	0 (83,3%)	0 (67,1%)	0 (53,8%)	0 (65,7%)	0 (64,5%)	0 (64,9%)	0 (65,5%)
Type d'habitation -Catégorie la plus fréquente (Pourcentage)	Appartement (62,5%)	Villa (33,3%)	Appartement (55,3%)	Appartement (69,2%)	Appartement (46,3%)	Villa (48,4%)	Appartement (48,6%)	Appartement (51,0%)
Propriétaire/Locataire -Catégorie la plus fréquente (Pourcentage)	Propriétaire (57,5%)	Propriétaire (54,2%)	Propriétaire (60,0%)	Propriétaire (65,4%)	Propriétaire (49,3%)	Propriétaire (54,8%)	Propriétaire (51,1%)	Propriétaire (56,1%)
Zone d'habitation -Catégorie la plus fréquente (Pourcentage)	Urbaine (92,5%)	Urbaine (79,2%)	Urbaine (90,6%)	Urbaine (88,5%)	Urbaine (86,6%)	Urbaine (93,5%)	Urbaine (89,2%)	Urbaine (89,0%)
Nbr de voitures dans le foyer -Catégorie la plus fréquente (Pourcentage)	2 (55,0%)	2 / 3 (33,3%)	2 (43,5%)	2 (50,0%)	2 (43,3%)	2 / 3 (32,3%)	2 (48,6%)	2 (44,2%)
Région - Catégorie la plus fréquente (Pourcentage)	Nord (70,0%)	Nord (79,2%)	Nord (71,8%)	Nord (65,4%)	Nord (71,6%)	Nord (67,7%)	Nord (83,8%)	Nord (72,6%)
Revenu mensuel en milliers DA -Catégorie la plus fréquente (Pourcentage)	60-100 (70,0%)	100-200 (79,2%)	100-200 (71,8%)	100-200/30-60 (65,4%)	60-100/100-200 (71,6%)	200-300 (67,7%)	100-200 (83,8%)	100-200 (31,6%)

Rappelons que les segments ciblés pour le lancement et la commercialisation des voitures produites en Algérie sont les clusters 1 et 2. En analysant le tableau précédent, nous pouvons dresser les profils des individus appartenant à ces deux clusters.

L'individu type appartenant au premier segment est un homme marié âgé entre 25 et 34 ans, propriétaire d'un appartement et dont la profession est cadre supérieur. Le revenu mensuel moyen de son foyer se situe entre 60 000 et 100 000 DA.

L'individu type du second cluster est un étudiant célibataire âgé entre 18 et 24 ans. Il habite une villa et le revenu moyen de son foyer se situe entre 100 000 et 200 000 DA.

Au-delà de ces critères sociodémographiques, il est important de s'intéresser aux besoins et attentes des segments afin que l'entreprise puisse adapter son offre commerciale et marketing en conséquence. Pour ce faire, nous réalisons une analyse comparative des différents segments selon les critères de choix d'une voiture. La note totale du segment concernant un critère représente la moyenne de ce segment par rapport à ce critère. Afin de garantir une meilleure visualisation des données comparatives, nous avons réduit les moyennes des segments par la moyenne globale accordée à chaque critère.

La variable FINANCE est calculée à partir des critères Prix, Valeur à la revente et Consommation de carburant. Nous en avons déduit qu'une moyenne élevée attribuée à ce critère signifie une grande importance accordée aux contraintes financières lors du choix d'une voiture.

Ainsi, la figure III.30 présentée ci-dessous, indique que les individus des clusters 1 et 2 font généralement plus attention que la moyenne au critère financier lorsqu'ils choisissent leurs voitures. Cependant, nous remarquons que cette importance est légèrement plus élevée pour les individus du cluster 2 que pour ceux du cluster 1, ce qui pourrait expliquer l'achat des membres du segment 2 de véhicules d'occasion. L'origine de cette attention plus prononcée pour le cluster 2 que le cluster 1 vis-à-vis du critère financier pourrait s'expliquer par le fait que le cluster 1 soit majoritairement constitué de personnes dont la profession est cadre supérieur tandis que la majorité des individus du cluster 2 sont des étudiants âgés de 18 à 24 ans (Rappelons que le revenu mensuel indiqué est celui du foyer de l'individu. Par exemple, le revenu mentionné par un étudiant n'ayant pas d'emploi représente très probablement celui de ses parents).

Ceci étant, nous pouvons raisonnablement émettre l'hypothèse selon laquelle les individus du segment 2 priorise exclusivement le critère financier dans le choix d'une voiture. Une hypothèse qui sera étudiée en analysant les autres critères de choix.

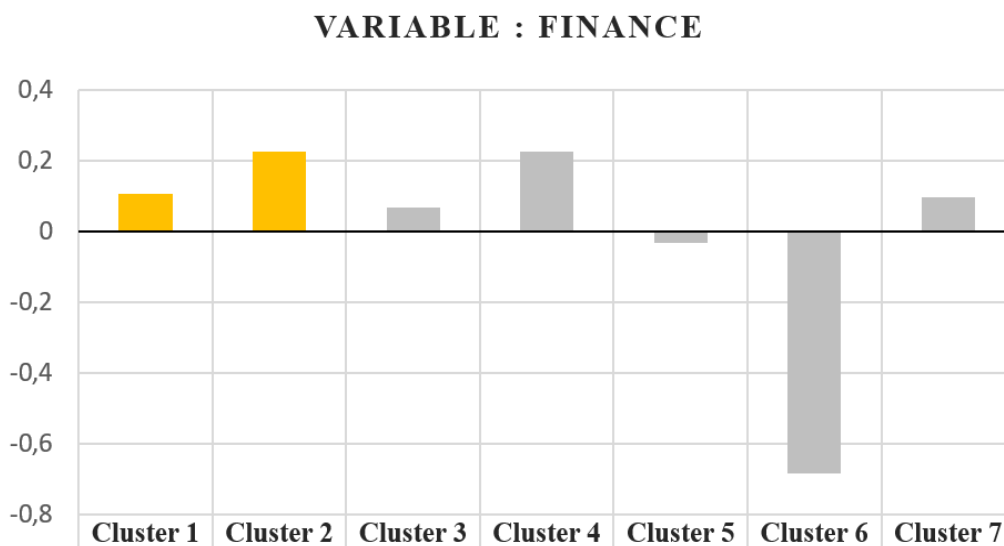


Figure III. 30: Comparaison des clusters en fonction de la variable FINANCE.

La variable CARACTERISTIQUE représente la moyenne des variables Marque, Puissance du moteur, Design extérieur, Confort de la voiture, Niveau d'équipements et Nouveauté du modèle. Nous en déduisons qu'une note élevée attribuée à ce critère signifie que l'individu accorde une grande importance aux différentes caractéristiques de la voiture, aussi bien internes (puissance du moteur, confort et niveau d'équipements) qu'externes (Marque, Nouveauté du modèle et Design).

La figure III.31 montre que le cluster 1 accorde une importance légèrement supérieure à la moyenne aux caractéristiques de la voiture lors du choix. En revanche, le cluster 2 en est nettement moins soucieux. D'ailleurs, la moyenne du cluster 2 par rapport à ce critère est la plus faible parmi les 7 clusters.

Nous en déduisons, d'une part, que pour les individus du segment 1, les caractéristiques de la voiture sont presque aussi importantes dans le choix d'une voiture que le critère financier. Ceci prouve que les deux critères ne sont pas forcément opposés.

D'autre part, les individus du segments 2 privilégient clairement le critère financier lié à l'achat et la possession d'une voiture aux caractéristiques qu'elle peut présenter. Une constatation qui confirme le caractère prioritaire relatif aux coûts chez les automobilistes du segment 2.

Du point de vue démographique et socioéconomique, le cluster le plus semblable au segment 2 est le segment 7. Cependant, ce dernier affiche la note la plus élevée pour le critère CARACTERISTIQUES parmi les 7 segments. Nous remarquons également que l'importance accordée aux critères financiers par ce même segment est moins élevée que celle du cluster 2. Ceci confirme notre hypothèse de travail selon laquelle les individus appartenant au même groupe démographique peuvent parfaitement se comporter de manière différente et présenter des besoins opposés.

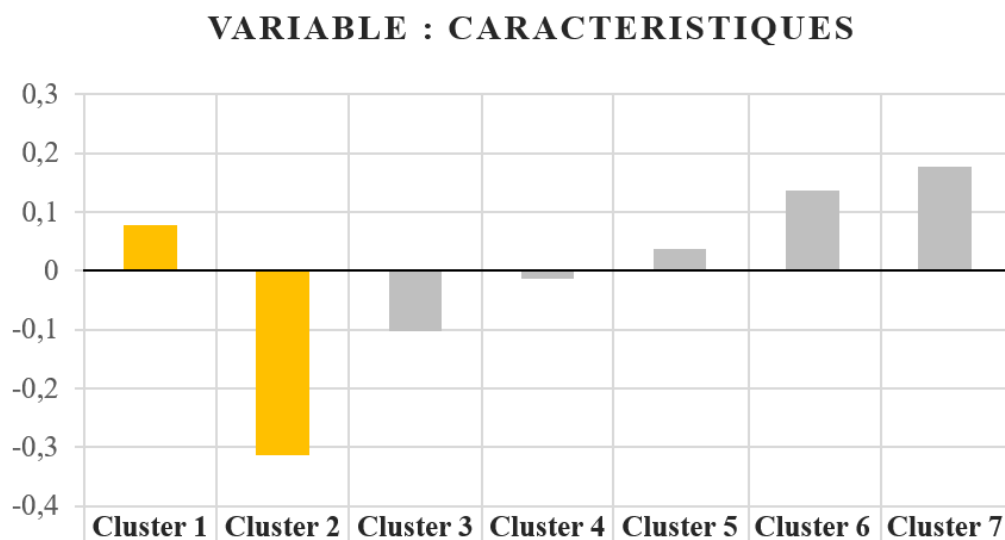


Figure III. 31: Comparaison des clusters en fonction de la variable CARACTERISTIQUES.

La variable UTILITAIRE est calculée à partir du degré d'accord avec la proposition affirmant que la voiture est plus un moyen de transport qu'une source de plaisir ainsi que des variables sécurité et environnement. Trois critères qui révèlent une position neutre par rapport à la voiture chez les répondants et donc une vision utilitaire de l'automobile.

En nous appuyant sur les résultats présentés dans la figure III.32, nous constatons que les critères en rapport avec une vision utilitaire de l'automobile sont légèrement plus importants pour les individus du clusters 1 que la moyenne. Quant à l'importance relative à ces critères pour les individus du clusters 2, elle s'inscrit pratiquement dans la moyenne. Rappelons que cette moyenne, qui est égale à celle de la variable PRESTIGE, est estimée à 3,6 sur 5.

Nous en déduisons que les segments précédemment cités semblent apprécier la voiture pour son aspect pratique sans la réduire pour autant à cela.

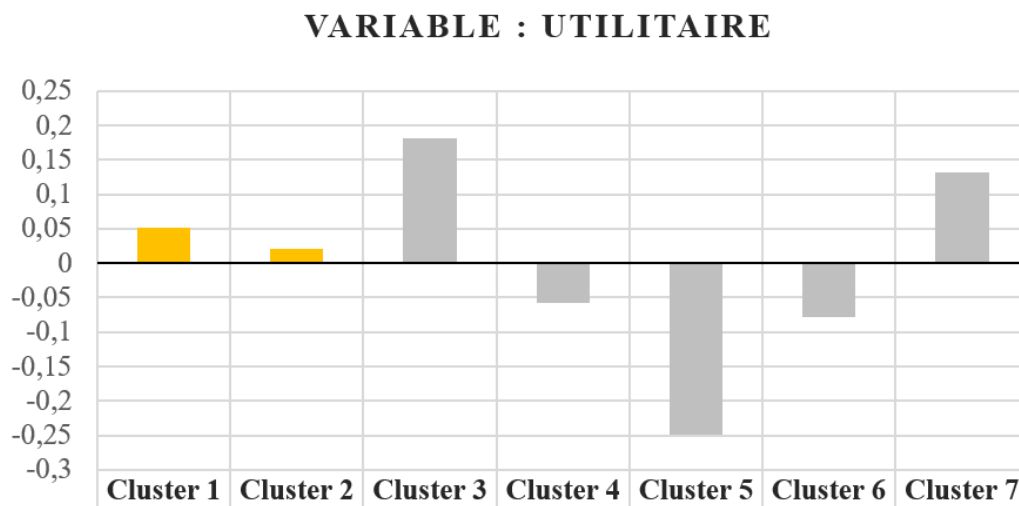


Figure III. 32: Comparaison des clusters en fonction de la variable UTILITAIRE.

Le calcul de la variable PRESTIGE a nécessité l'utilisation la variable représentant le degré d'accord du répondant avec le fait que la voiture qu'il conduit doit représenter sa position sociale ainsi que le critère de choix : Statut Social. Comme son nom l'indique, la variable PRESTIGE mesure donc l'importance accordée au prestige qu'apporte l'achat et la possession d'une voiture à son propriétaire.

La figure III.33 montre que le cluster 2 est le moins soucieux de ce critère, avec une valeur nettement inférieure à la moyenne générale. Ce qui corrobore parfaitement l'hypothèse voulant que la priorité des individus du cluster 2 lors du choix d'une voiture concerne exclusivement le critère financier.

Par ailleurs, l'importance accordée par les individus du cluster 1 au PRESTIGE qu'apporte la voiture s'inscrit parfaitement dans la moyenne générale.

Cependant, le cluster 6, qui présente la meilleure note pour le critère PRESTIGE, possèdent des caractéristiques sociodémographiques similaires à ceux du cluster 1. La seule différence significative notée par rapport à ces caractéristiques est le revenu moyen du foyer.

En effet, le cluster 6 est majoritairement constitué d'individus dont le revenu moyen du foyer se situe entre 200 000 et 300 000 DA, ce qui représente le plus grand revenu moyen parmi les 7 clusters, contre un revenu moyen du foyer s'inscrivant dans l'intervalle 30 000 - 60 000 DA pour le cluster 1. Cette différence de statut social pourrait, éventuellement, être à l'origine de la différence des notes des deux clusters pour la variable PRESTIGE.

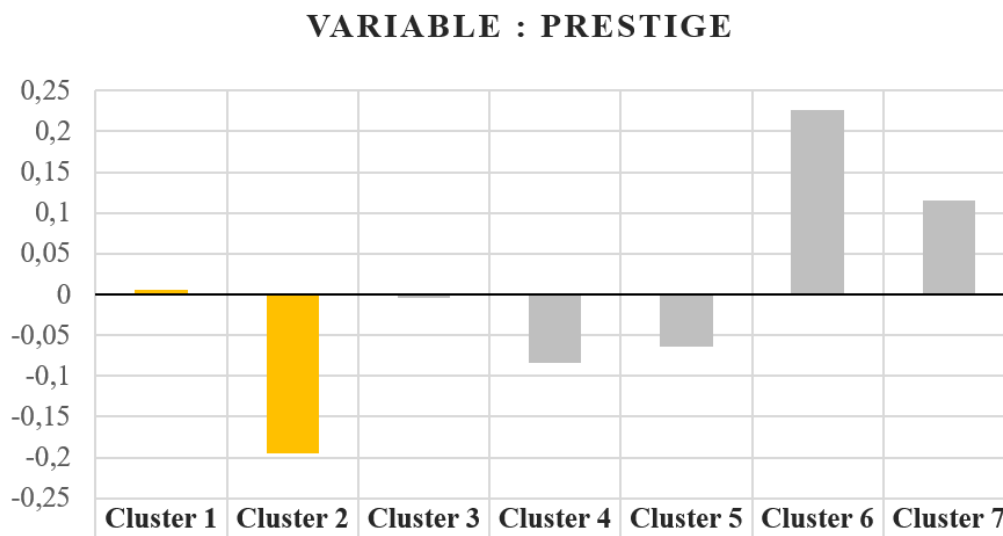


Figure III. 33: Comparaison des clusters en fonction de la variable PRESTIGE.

En résumé, les analyses précédentes nous ont permis de dresser les profils des individus types pour chacun des deux segments ciblés ainsi que de découvrir que les principaux critères de choix du segment 1 s'articulent autour du critère financier et des caractéristiques de la voiture, tandis que le segment 2 s'intéresse exclusivement aux coûts engendrés par l'achat et la possession d'une voiture, délaissant ainsi les caractéristiques et le prestige que pourrait procurer cette dernière.

- Stratégies de communication :

Dans le but de contribuer à l'efficacité des campagnes marketing réalisées par l'entreprise Renault Algérie, dans le cadre des lancements et commercialisation des voitures issues de la production nationale, nous étudions les principales sources d'informations utilisées par les segments ciblés pour se renseigner sur l'automobile. Les résultats de cette étude permettront à l'entreprise d'optimiser le choix et l'utilisation des différents canaux de communication.

Pour ce faire, nous procédons de la même manière que pour le Profiling précédent, c'est-à-dire que nous comparons les notes réduites des 7 clusters concernant les variables relatives à l'utilisation des différentes sources d'information, déterminées à partir d'une analyse factorielle.

Rappelons que les 4 variables calculées après l'ACP sont : INFO_TRADITIONNELLES, INFO_CONNEXION, AVIS_NONPRO et AVIS_PRO.

La figure III.34 illustre cette comparaison.

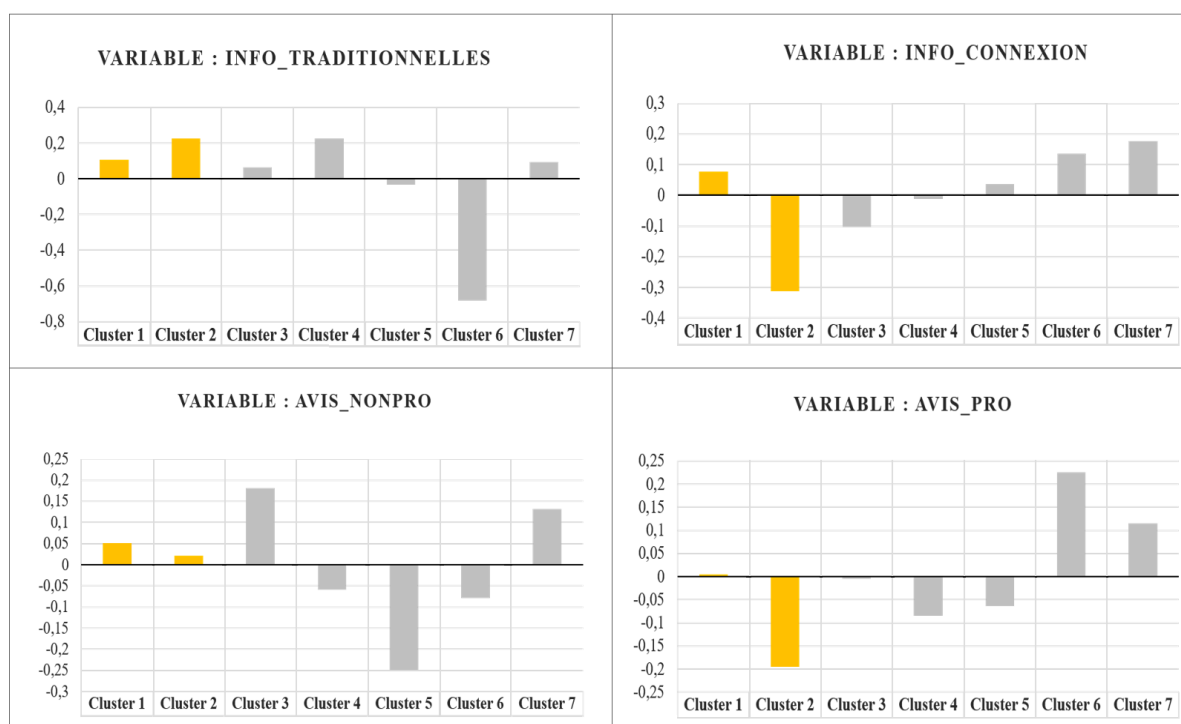


Tableau : Moyennes d'utilisation des sources d'information sur 5.

Variable	INFO_TRADITIONNELLES	INFO_CONNEXION	AVIS_NONPRO	AVIS_PRO
Moyenne	4,39	4,14	3,60	3,59

Figure III. 34: Comparaison des clusters en fonction des sources d'information utilisées.

Nous remarquons que généralement, les sources d'informations les plus utilisées pour se renseigner sur les voitures sont les canaux de communication traditionnelles, suivies de près, par le canal digital.

Par ailleurs, les sources d'information traditionnelles, internet et l'avis des non professionnels sont plus utilisées par les individus du cluster 1 pour la recherche d'informations sur l'automobile que l'avis des professionnels, dont la note relative est légèrement supérieure à la moyenne.

Le cluster 2 utilise beaucoup moins internet et l'avis des professionnels pour se renseigner sur les voitures. Les notes relatives à ces deux variables sont d'ailleurs les plus faibles par rapport aux autres clusters. En revanche, les individus de ce même cluster font plus souvent appel aux sources d'information traditionnelles que les avis des non professionnels sur les voitures, même si ce dernier moyen reste légèrement plus utilisé que la moyenne.

La figure III.35 détaille l'utilisation des sources d'informations traditionnelles par les différents clusters.

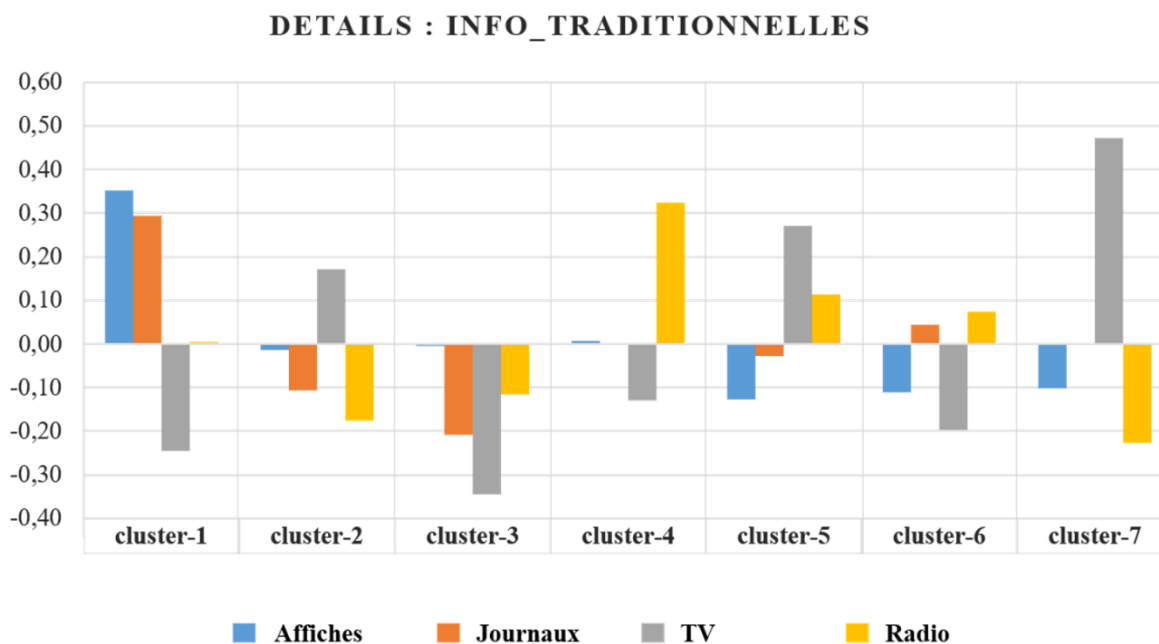


Figure III. 35: Détails d'utilisation des sources d'information traditionnelles par les clusters.

Nous constatons que les individus du premier segment ont beaucoup plus recours aux journaux et affiches publicitaires pour se renseigner sur les voitures que les personnes appartenant au segment 2. Ces dernières sont toutefois plus sensibles aux programmes et publicités télévisées sur l'automobile, contrairement aux individus du segment 1. Enfin, le recours à la radio comme source d'information par le segment 1 est à peine supérieure à la moyenne générale, et beaucoup plus faible pour le segment 2.

En résumé, le canal de communication qui semble le plus approprié pour la stratégie marketing relative à la production nationale des voitures regroupe les sources d'information traditionnelles notamment à travers les affiches publicitaires, les journaux et magazines ainsi que la TV. Par ailleurs, le canal digital est également approprié pour les individus du segment 1. En se basant sur le Profiling précédemment établi, l'entreprise est en mesure de différencier sa stratégie de communication à travers ce canal, en fonction des caractéristiques du segment ciblé. Enfin, la sensibilité des deux segments aux avis des proches, représenté par la variable AVIS_NONPRO, suggère l'intégration de ce paramètre dans le discours commercial de l'entreprise.

Conclusion

Nous avons détaillé, dans le présent chapitre, l'utilisation des outils de Data Mining dans la segmentation des clients en fonction des données récoltées à partir de deux différentes enquêtes.

D'une part, la première enquête que nous avons réalisée nous a permis de collecter différentes données sur les caractéristiques sociodémographiques, comportementales et attitudinales des clients de l'entreprise Renault Algérie. Le but étant de proposer une segmentation efficace des clients afin de contribuer à l'optimisation des actions marketing de l'entreprise.

Pour ce faire, des méthodes de classification nous ont tout d'abord permis d'établir 7 différents segments de clients en fonction de leur comportement d'achat de véhicules et de services après-vente, et donc de fournir à l'entreprise une cible pertinente pour les actions marketing relatives au développement de l'activité après-vente à travers l'identification des segments les moins acheteurs de services.

Par ailleurs, à l'aide d'un modèle de classement, nous avons pu identifier deux segments présentant les meilleures probabilités d'achat du nouveau modèle qui sera bientôt lancé par la marque Renault, à savoir la Megane 4, avant de procéder au Profiling sociodémographique et attitudinal de ces segments.

Enfin, en nous basant sur l'utilisation des différentes sources d'information de la part des clients, nous avons proposé une stratégie de communication adéquate pour chacun des deux segments cibles.

D'autre part, la seconde étude de marché réalisée dans le but de segmenter le marché automobile algérien et de soutenir ainsi l'entreprise dans l'optimisation de ses actions marketing et commerciale, nous a permis de collecter différentes données sociodémographiques, comportementales et attitudinales des acheteurs de véhicules en Algérie. A partir de ces données, nous avons tout d'abord segmenté le marché en 7 principaux cluster avant d'identifier 2 segments cibles servant nos objectifs d'étude.

Ensuite, nous avons mis en place un modèle d'estimation permettant de classer différents individus dans les segments précédemment établis, et ce, à partir de données sociodémographiques, plus faciles à capturer que le comportement ou les besoins des consommateurs.

Enfin, un Profiling des segments cibles nous a permis une meilleure connaissance de leurs caractéristiques et de proposer ainsi à l'entreprise, une stratégie de communication adéquate pour chacun des deux segments cibles.

Conclusion générale

Le travail réalisé consistait en l'application des techniques de Data Mining dans la segmentation efficace des clients pour le compte de l'entreprise Renault Algérie, et ce, dans le but de proposer des solutions pertinentes pour l'optimisation des stratégies marketing de l'entreprise.

La première étape de notre méthodologie de travail concernait la constitution d'une base théorique nous permettant de cerner les différents concepts relatifs à la gestion de la relation client ainsi que les différentes méthodes de Data Mining appliquées à la segmentation des clients, en général, et des clients du secteur automobile, en particulier.

Dans la seconde étape, nous nous sommes intéressés à l'environnement dans lequel évolue l'entreprise, à savoir, le marché automobile algérien, afin d'affiner la définition de la problématique fixée et de cerner davantage les objectifs de la segmentation. De plus, le diagnostic de la base de données clients réalisé, nous a permis de détecter diverses anomalies réduisant la fiabilité de la base.

Une fois les corrections nécessaires apportées, une segmentation des clients réalisée à l'aide de l'algorithme k-means et utilisant un score RFM calculé ainsi que des données existantes en entreprise, a conduit à la constitution de 6 segments clients. Ensuite, un modèle de classement utilisant les arbres de décision est appliqué, dans le but de mettre en avant les règles régissant la première segmentation et rendant ainsi possible le classement des nouveaux clients dans l'un des segments définis. Enfin, un modèle prédictif basé sur la régression logistique est implémenté à partir de ces données afin de permettre à l'entreprise de cibler les clients, avec une fiabilité de plus de 95%, les clients présentant un profil acheteur de la Sandero Made in Bladi, dont la commercialisation est prévue pour les semaines à venir.

La troisième et dernière étape de notre travail est basée sur deux différentes études de marché. La première servant à affiner la segmentation des clients de l'entreprise dans le but d'identifier les meilleures opportunités de développement du service après-vente de l'entreprise, de cibler les segments présentant la meilleure probabilité d'achat de la nouvelle Megane et enfin, d'offrir à l'entreprise une meilleure connaissance du comportement et besoins de ses clients mais aussi de leurs préférences relatives aux canaux de communications. Tout cela, dans le but d'optimiser les coûts des campagnes marketing prévues pour les prochains mois.

La deuxième étude de marché est réalisée dans le but d'optimiser la rentabilité de l'activité de production nationale afin de palier à la réduction du chiffre d'affaires imposée par la réduction des importations. Ainsi, les résultats obtenus nous ont servi pour la segmentation

du marché automobile dans le but d'identifier les différents besoins ainsi que les caractéristiques sociodémographiques et comportementales des acheteurs de véhicules favorables à l'acquisition de véhicule fabriqué en Algérie. Un Profiling des segments cibles permettra à l'entreprise d'adapter sa stratégie marketing et de communication dans la promotion de ses produits fabriqués dans le pays.

Notons toutefois, qu'un certain nombre d'améliorations peut être apporté à ces analyses. Tout d'abord, une meilleure fiabilisation de la base de données clients offrirait certainement une segmentation plus élaborée et plus précise des clients de l'entreprise. En effet, les nombreuses anomalies détectées devraient être traitées de façon permanente, et ce, en limitant les accès en modification à la base et en sensibilisant les agents collaborateurs sur l'importance de la collecte des données sur les clients. De plus, l'intégration des données relatives au montant d'achat du véhicule, des pièces de rechanges ainsi que des services après-vente garantirait une meilleure typologie des clients en fonction de la valeur réelle apportée à l'entreprise. Par ailleurs, une analyse temporelle du comportement des clients constitue un moyen pertinent d'évaluer les performances de l'entreprise dans le temps et d'étudier l'évolution de la clientèle en fonction des précédentes actions marketing entreprises.

Concernant l'étude de marché automobile, rappelons que les contraintes de temps et de coût nous ont conduits à réaliser un sondage par internet afin de récolter les données nécessaires à notre étude. Bien que dans le domaine du marketing, la taille de l'échantillon est considérée parfaitement acceptable, l'hypothèse selon laquelle l'échantillon de l'étude est aléatoire n'est pas forcément vérifiée, puisque les personnes ayant répondu au questionnaire représentent une population plus ou moins « connectée ». Or, une simple connaissance de l'environnement socioéconomique de notre pays nous permet de constater que le taux de connexion demeure assez faible, notamment chez les moins jeunes. Pour remédier à ce critère de biais de la population, nous pouvons proposer une méthode de sondage mixte, moins dépendante sur le canal digital.

De manière générale, le présent travail nous a permis d'apprécier l'avantage concurrentiel non négligeable, que prodigue l'application du Data Mining dans la segmentation des clients, notamment dans le secteur automobile.

BIBLIOGRAPHIE

Abdullateef, Mokhtard et Yusuf, 2004. Effects of Customer Relationship Management Strategy on Call Centre's Employee Intention to Quit: Evidence from Malaysia Call Center, *Procedia - Social and Behavioral Sciences*, Vol.130, pp.305-315, 2004.

Ahn et Kim, 2008. Real-time segmentation of objects from video sequences with non-stationary backgrounds using spatio-temporal coherence. In *Proc.ICIP*, p1554-1547, 2008.

Albert, 2003. Need-based segmentation and customized communication strategies in a complex-commodity industry: A supply chain study. *Industrial Marketing Management*, Vol.32, pp.281-290, 2003.

Amstrong et Kotler, 2014. *The marketing of leadership*, Pearson Education, 2014.

Beerli, Martín et Quintana, 2004. A model of customer loyalty in the retail banking market. *European Journal of Marketing*, Vol.38, pp.253-275, 2004.

Bose, 2002. Customer relationship management: key components for IT success. *Industrial Management & Data Systems*, Vl.102, pp.89-97, 2002.

Campbell, 2003. Creating customer knowledge competence: managing customer relationship management programs strategically. *Industrial Marketing Management*, Vol.34, pp.375-383, 2003.

Chan, 2008. Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert Systems with Applications*, Vol.34, pp.2231-3106, 2008.

Chen et Chen, 2004. Exploring the success factors of eCRM strategies in practice. *The Journal of Database Marketing*, Vol.11, pp.333-343, 2004.

Chen, Kuo, Wu et Tang, 2009. Discovering recency, frequency, and monetary (RFM) sequential patterns from customers purchasing data. *Electronic Commerce Research and Applications*, Vol.8, pp.241–251, 2009.

Chou et al, 2002. Towards a Successful CRM Implementation in Banks: An Integrated Model. *The Service Industries Journal*, Vol.27, pp.1021-1039, 2002.

Coltman, 2007. Can superior CRM capabilities improve performance in banking? *Total Quality Management and Business Excellence*, Vol.20, pp.283–299, 2007.

Cooil, Aksoy et Keiningham, 2008. Approaches to customer segmentation. *Journal of Relationship Marketing*, Vol.6, pp.9-39, 2008.

- Dennis, King et Jayawardhena, 2007.** Consumers online: intentions, orientations and segmentation. *International Journal of Retail & Distribution Management*, Vol. 35, pp.515 - 526, 2007.
- Greenberg et Signori, 2002.** CRM Customer Relationship Management. La tecnologia aiuta a migliorare e rafforzare la relazione con i clienti, Apogeo Editore, 2002.
- Haaijer, Wedel et Vriens, 1998.** Utility covariances and context effects in conjoint MNP model. *Marketing Science*, Vol.17, pp.236-252, 1998.
- Hsu, Chang et Hung, 2007.** Applying SVM to build supplier evaluation model-comparing Likert scale and Fuzzy scale. *IEEE International Conference on Industrial Engineering and Engineering Management*, Conference 2007, pp.6-10, 2007.
- Hung et Kang, 2006.** A Data Mining Case Study in the Underwear Industry for CRM Applications International. *Journal of Management Cases*, Vol.14, pp.12-14, 2006.
- Hwang, Jung et Suh, 2004.** An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *European Journal of Operational Research*, Vol.157, pp.196–217, 2004.
- Jain et Singh, 2002.** Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing*, Vol.16, pp.34-39, 2002.
- Kim, Jung, Suh et Hwang, 2006.** Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, Vol.31, pp.101–107, 2006.
- Kracklauer et Mills et Seifert, 2004.** Collaborative customer relationship management: taking CRM to the next level. Springer, 2004.
- Lee, Lee et Wicks, 2004.** Segmentation of festival motivation by nationality and satisfaction. *Tourism Management*, Vol.25, pp.60-71, 2004.
- Ling, Raymon et Yen, 2001.** Customer relationship management: An analysis framework and implementation strategies. *The Journal of Computer Information Systems*, Vol.41, pp.82-97, 2001.
- Ma, 2015.** A Study on Customer Segmentation for E-Commerce Using the Generalized Association Rules and Decision Tree. *American Journal of Industrial and Business Management*, Vol.5, pp.813, 2015.
- Meuter, Ostrom, Roundtree et Bitner, 2000.** Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters. *Journal of Marketing*: July 2000, Vol. 64, pp. 50-64, 2000.
- Namvar et Khakabimamaghani, 2011.** An approach to optimised customer segmentation and profiling using RFM, LTV, and demographic features. *Expert Systems with Applications*, Vol.37, pp.7120–7131, 2011.

- Naresh Malhotra, 2007.** Etudes Marketing avec SPSS, 5ième édition. PEARSON EDUCATION, France, 2007.
- Natarajan, Pednault, Tipu et Campbell, 2001.** Segmentation-based modeling for advanced targeted marketing. IEEE International Conference on Fuzzy Systems, Volumes 1 à 3, 2001.
- Newell, Goldsmith et Lafferty, 2000.** The Impact of Corporate Credibility and Celebrity Credibility on Consumer Reaction to Advertisements and Brands. Journal of Advertising, Vol.26, pp.23-28, 2000.
- Ngai, 2005.** Customer relationship management research (1992-2002): An academic literature review and classification. Marketing Intelligence & Planning, Vol. 23, pp.582 - 605, 2005.
- Parvatiyar et Sheth, 2001.** Customer relationship management: Emerging practice, process, and discipline. Journal of Economic and Social research, Vol.3, pp.1-34, 2001.
- Payne et Adrian et Frow, 2015.** A Strategic Framework for Customer Relationship Management. Journal of Marketing: October 2005, Vol. 69, pp. 167-176, 2015.
- Preslan et Keltz, 2003.** The customer management applications spending. Report, 2003-2004. Boston, 2003.
- Rajendra, Srivastava, Tasadduq, Shervani et Fahey, 1999.** Marketing, Business Processes, and Shareholder Value: An Organizationally Embedded View of Marketing Activities and the Discipline of Marketing. Journal of marketing, Vol.63, pp.68-79, 1999.
- Ryals et Payne, 2001.** Customer relationship management in financial services: towards information-enabled relationship marketing, Journal of strategic marketing, Vol.9, pp.3-27, 2001.
- Siravenha ,Brito et Pelaes, 2014.** Fuzzy segmentation tool for remote sensing data.Proc. SPIE 9244, Image and Signal Processing for Remote Sensing XX, 924407 (October 23, 2014),2014.
- Smith, 1956.** Product differentiation and market segmentation as alternative marketing strategies. The Journal of Marketing, Vol.21, pp. 3-8, 1956.
- Stéphane TUFFERY,2012.** Data Mining et statistique décisionnelle.4ième édition, EDITIONS TECHNIP, France, 2012.
- Swift, 2001.** Accelerating Customer Relationships: Using CRM and Relationship Technologies. Journal of Business Research, Vol.54, pp. 39-52, 2001.
- Terho et Halinen, 2007.** Customer portfolio analysis practices in different exchange contexts. Journal of Business Research, Vol.60, pp.720-730, 2007.
- Tsai, Hu et Lu, 2015.** Customer segmentation issues and strategies for an automobile dealership with two clustering techniques. Expert Systems: International Journal of Knowledge Engineering and Neural Networks, 2015.

Tsiptsis et Chorianopoulos, 2011. Data mining techniques in CRM: inside customer segmentation, Winsley, 2011.

Verhoef et Donkers, 2001. Predicting customer potential value an application in the insurance industry. Decision support systems, Elsevier,2001.

Wei, Lin et Wu, 2010. A review of the application of RFM model, Ubiquitous Computing Application and Wireless Sensor: UCAWSN-14, 2010.

Wilson, Daniel et McDonald, 2002. Factors for success in customer relationship management (CRM) system. A Review of Research Management Science, Vol.18, pp.193-219, 2002.

Xu et Walton, 2005. Gaining customer knowledge through analytical CRM. Industrial Management & Data Systems, Vol. 105, pp.955 - 971, 2005.

Yankelovich et Meer, 2006. Rediscovering market segmentation, Harvard business review, 84, pp.222, 2006.

Zablah, Bellenger et Johnston, 2004. Customer Relationship Management Implementation Gaps. Journal of Personal Selling & Sales Management, Vol.24, pp.279-295, 2004.

Zolkiewski et Turnbull, 2002. Do relationship portfolios and networks provide the key to successful relationship management? Journal of Service Research, Vol.6, pp.324-335, 2002.

ANNEXES

ANNEXE 1: Classement des ventes par segment, en 2015.

Source : Documents Renault Algérie

Segment	Modèle	Ventes	Classement
M0	LOGAN	22117	1
	SYMBOL	21614	2
	ACCENT	19003	3
	301	4818	4
	C ELYSEE	1782	5
I	208	12775	1
	CLIO 4	9971	2
	SANDERO	9908	3
	CLIO 2	4640	4
	POLO	4119	5
SUV	DUSTER	8343	1
	TIGUAN	1900	2
	M4	1114	3
	SPORTAGE	1094	4
	YETI	841	5
M1	308	4813	1
	LEON	2216	2
	GOLF	1312	3
	MEGANE 3	1208	4
Ludospaces	KANGOO	4203	1
	NEW PARTNER	2530	2
	NEW BERLINGO	2155	3
	CADDY	297	4
Fourgonnettes	KANGOO	5365	1
	NEW PARTNER	4653	2
	NEW BERLINGO	1002	3
	CADDY	1975	4
Grands fourgons	MASTER	1963	1
	BOXER	1487	2
	JUMPER	441	3
	CRAFTER	170	4
	TRANSIT	140	5
Petits fourgons	EXPERT	1444	1
	TRAFIC	287	2
	JUMPY	221	3
	TRANSPORTER	216	4
	MULTIVAN	42	5

ANNEXE 2: Aperçus du rapport de l'enquête NCBS 2015.

(Les pages présentées sont suivantes : 1, 3, 8, 9, 10, 14, 19 et 20)



RAPPORT D'ENQUETE NCBS 2015

Le 16/03/2016

BENAICHE Hadjer & HADDAM Yacine

I. PRESENTATION DE L'ETUDE

ENQUETE NCBS

New Car Buyers Survey (NCBS) est une enquête réalisée sur des récents acheteurs de nouvelles voitures.

La période de vente couverte par l'étude s'étend généralement sur dix mois (de Septembre à Juin).

Les acheteurs interrogés possèdent leurs nouvelles voitures depuis 6 - 45 semaines.

EN ALGERIE

L'enquête NCBS 2015 réalisée en Algérie a couvert une période de vente allant de Septembre 2014 à Mars 2015 *, suivie par une période de récolte des données de Mars 2015 à Aout 2015.

La méthode d'entretien adoptée pour la récolte des données est : Face à face.

ECHANTILLON

Taille de l'échantillon : **1 993** interviews.

Ventes représentées : **136 623** unités

- L'enquête NCBS a été réalisée auprès d'acheteurs de voitures neuves (uniquement des véhicules particuliers) de 13 marques différentes, à savoir : Citroën, Hyundai, Kia, Nissan, Peugeot, Renault, Seat, Skoda, Suzuki, Toyota, Volkswagen, GM Daewoo & Chevrolet et Dacia.
- Au total, 67 modèles de différentes marques sont représentés. Toutefois, les résultats présentés concernent seulement les modèles dont les ventes dépassent les 30 unités. Ainsi, pour les marques Renault et Dacia, les modèles représentés sont :

MODELE	Nb. interviews	Ventes NCBS	% des ventes
Renault Clio (2001) /Symbol/Thalia (2002)	35	2 636	1,9%
Renault Megane III (2009)	49	735	0,5%
Renault Kangoo (2008)	56	2 823	2,1%
Renault Clio IV (2013)	93	5 975	4,4%
Renault Symbol (2013)	63	7 351	5,4%
Dacia Logan (2013)	79	18 090	13,2%

* Suite aux conflits entre GFK et son représentant en Algérie, l'étude a été arrêtée. Par conséquent, seulement les ventes réalisées entre Septembre 2014 et Mars 2015 sont représentées.

Page 3

II. ANALYSE DES DONNEES

RAISONS D'ACHAT

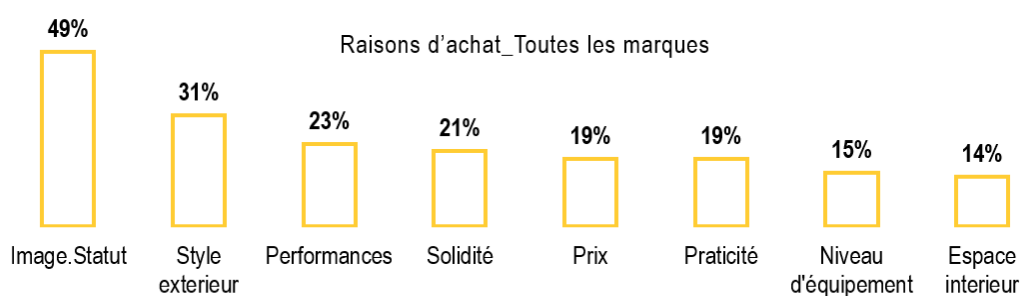
Les raisons d'achat sont traitées dans la **section C** de l'enquête.

La question posée est la suivante :

Pour quelles raisons avez-vous acheté cette voiture en particulier au lieu d'une autre ?

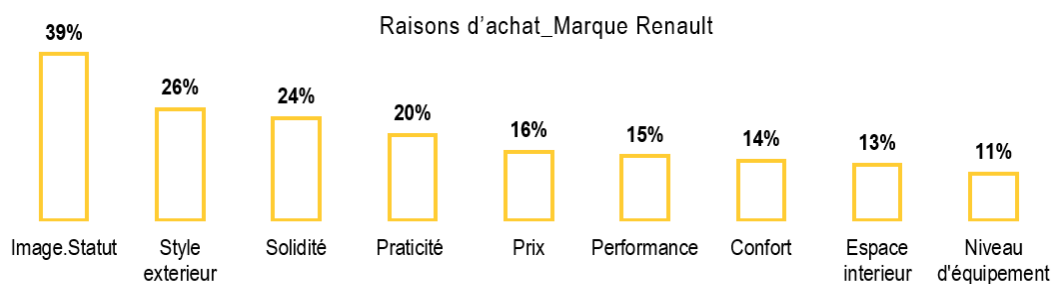
EN GENERAL

Les résultats cumulés pour toutes les marques sont présentées dans la figure suivante* :



MARQUE RENAULT

Renault est la marque la plus choisie pour son « **Confort** ». Elle est toutefois la dernière marque choisie pour son « **Image & Statut** » et sa « **Performance** ».



* Sont représentées les raisons choisies dans plus de 10% des cas.

Page 8

II. ANALYSE DES DONNEES

RAISONS D'ACHAT

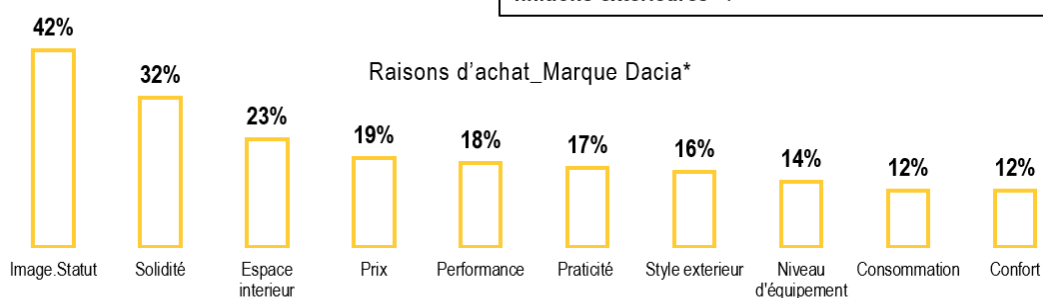
Les raisons d'achat sont traitées dans la **section C** de l'enquête.

La question posée est la suivante :

Pour quelles raisons avez-vous acheté cette voiture en particulier au lieu d'une autre ?

MARQUE DACIA

Dacia est la marque la plus choisie pour sa « **Solidité** » et sa faible « **Consommation** ».
Par contre, elle est la moins choisie pour le « **Style et les finitions extérieures** ».



COTE SEGMENTS

- Dans le segment **M0**, la **Symbol** est la plus choisie pour son « **Prix** » et sa faible « **Consommation** ». Elle est toutefois moins choisie pour son « **Image** », « **Performance** » et « **Tenue de Route** ».
- Dans le segment **M1**, la **Megane III** est la voiture la plus choisie pour sa « **Performance** » et la moins choisie son « **Prix** ».
- Dans le segment **I**, la **Clio IV** est la plus choisie pour son « **Confort** » et son « **Style extérieur** ». Elle est par contre la moins choisie pour son « **Image et Statut** », sa « **Manœuvrabilité** » et sa « **Performance** ».
- Pour la marque Dacia, dans le segment **M0**, la **Logan** est la voiture la plus choisie pour sa « **Solidité** » et sa faible « **Consommation** ». Par contre, elle est la moins choisie pour son « **Style et ses finitions extérieures** ».

* Sont représentées les raisons choisies dans plus de 10% des cas.

Page 9

II. ANALYSE DES DONNEES

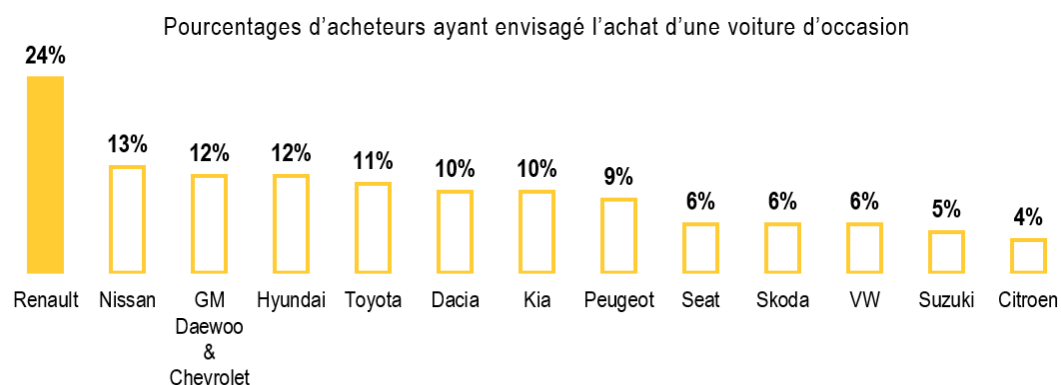
ENVISAGER L'OCCASION

Les résultats de cette partie sont présentés dans la **Section C** de l'enquête.

La question posée est la suivante :

Avant de décider d'acheter une voiture neuve, aviez vous considéré l'achat d'une voiture d'occasion ?

COTE MARQUES



COTE SEGMENTS

1 acheteur sur 3 d'une **Symbol** a envisagé l'achat d'une voiture d'occasion.

Segments	Modèles	Nb.Observations	Oui
M0	Citroen C-Elysee (2013)	77	3,9
	Dacia Logan (2013)	79	11,6
	Peugeot 301 (2012)	111	17,1
	Renault Symbol (2013)	63	33,3
I	Peugeot 208 (2012)	119	5,0
	Renault Clio IV (2013)	93	20,4
M1	Peugeot 308 (2014)	89	5,6
	Renault Megane III (2009)	49*	31,6

II. ANALYSE DES DONNEES

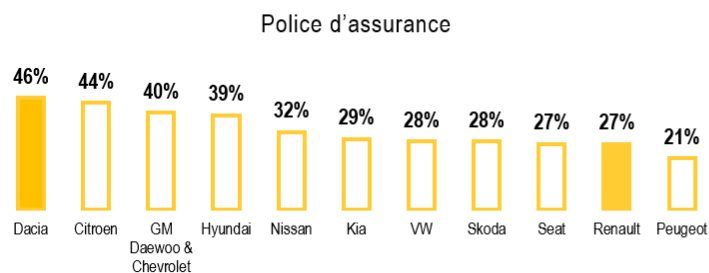
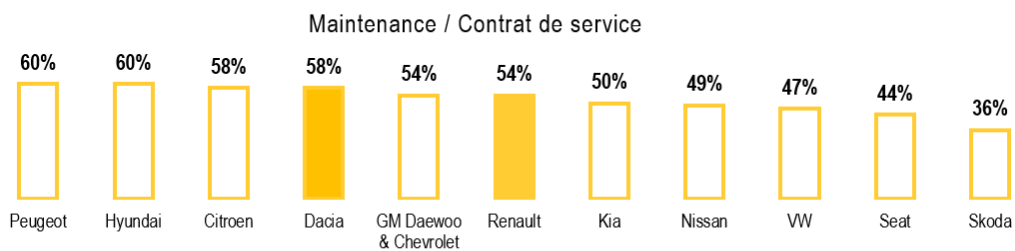
SERVICES COMPLEMENTAIRES

Les résultats de cette partie sont présentés dans la **Section B** de l'enquête. Ils représentent les personnes :
Ayant acheté des services additionnels auprès du concessionnaire/constructeur.

LECTURE

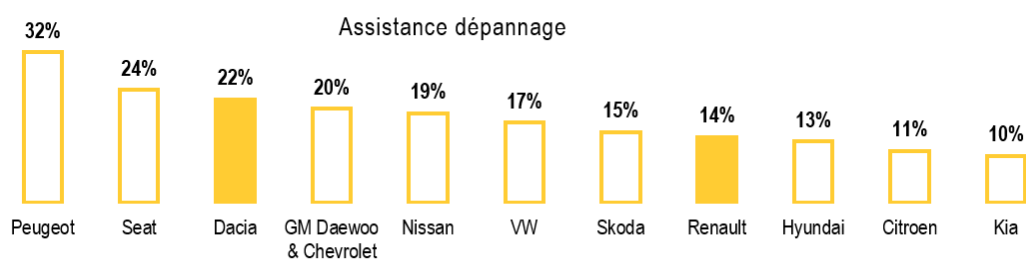
En général, la marque vend peu de services par rapport à ses concurrents.

% des acheteurs ayant pris un service supplémentaire avec l'achat de la voiture.



Très peu de clients de la marque Renault (**27,2%**) contractent leur police d'assurance au près de leur concessionnaire (contre **45,9%** pour Dacia).

Même constat pour l'assistance dépannage, puisque seulement **13,6%** des nouveaux acheteurs Renault se procurent ce service au niveau du concessionnaire.



II. ANALYSE DES DONNEES

SATISFACTION

Les résultats de cette partie sont traités dans la **section E** de l'enquête.

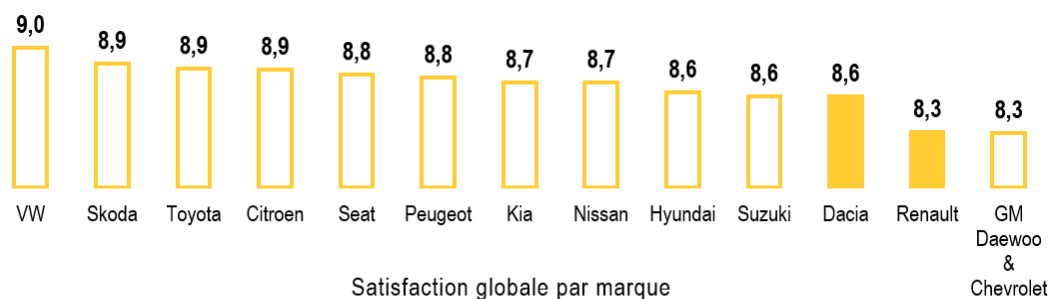
Ils représentent : **La satisfaction globale concernant la nouvelle voiture achetée (note sur une échelle de 1 à 10)**

SATISFACTION GLOBALE

Classement

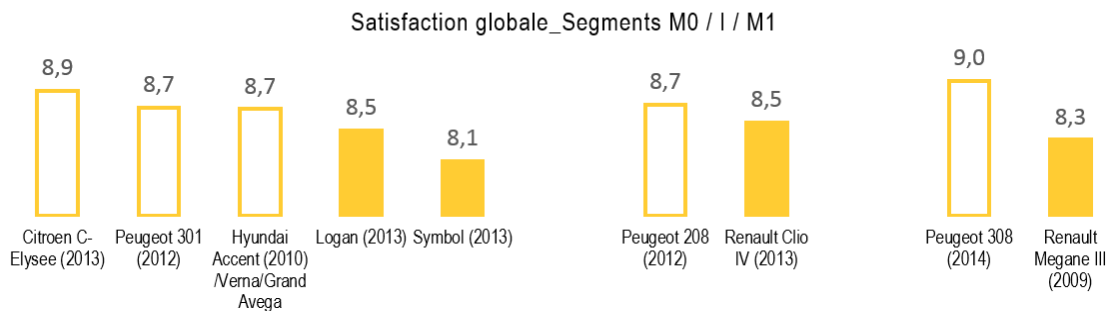
Dacia : **11^{ième}** avec **8,6 sur 10**

Renault : **12^{ième}** avec **8,3 sur 10**



COTE SEGMENTS

Les deux marques **Renault** et **Dacia** sont classées dernières dans les trois segments **M0**, **I** et **M1**



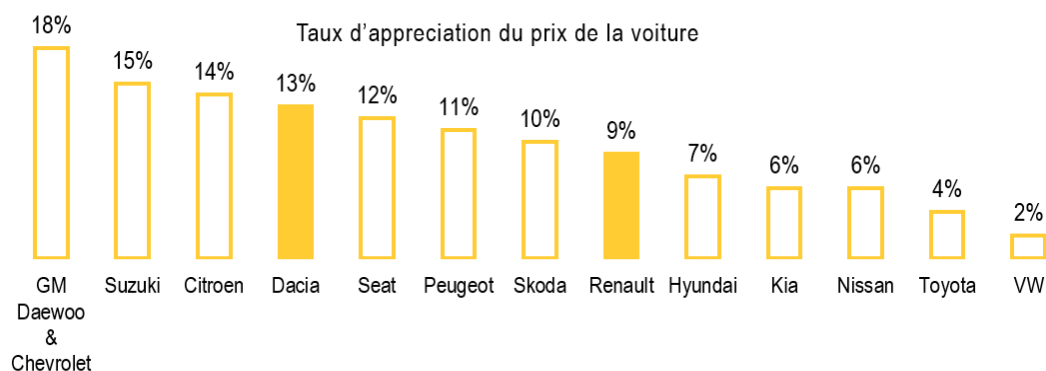
II. ANALYSE DES DONNEES

SATISFACTION

Les résultats de cette partie sont traités dans la **section E** de l'enquête.

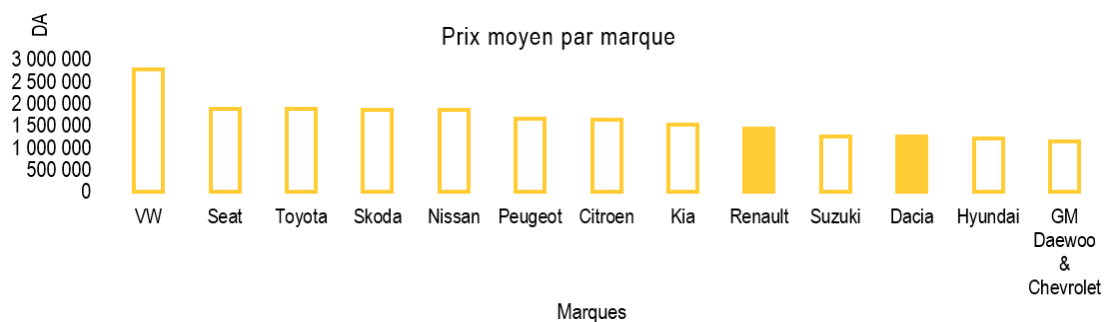
Ils représentent : **La satisfaction globale concernant la nouvelle voiture achetée (note sur une échelle de 1 à 10)**

SATISFACTION : CRITERE PRIX



ON NOTE

Les prix proposés par **Renault** ne sont appréciés que dans **9,8%** des cas, contre **12,2%**, **11,2%** et **10,6%** pour **Seat**, **Peugeot** et **Skoda**, respectivement. Des marques dont les prix moyens par marque sont supérieurs à ceux de Renault.



Page 20

ANNEXE 3 : Définition des SCI de Renault Algérie.

Système source	Système cible	Type d'interface	Données Transférées	Fréquence	Commentaire
BVM	R Force	Batch webservice	Données des véhicules identifiés dans les bases locales relatives aux clients	Rafraîchis par semaine ou par requête d'utilisateur	L'utilisateur final peut retrouver toutes les informations perdues concernant le véhicule grâce au VIN ou le numéro d'enregistrement seulement.
ICM	R Force	Webservice	Conditions de garantie Données historiques de garantie Contrats d'entretien souscrits Contrats d'extension	Requête d'utilisateur	Aucune donnée stockée dans R Force
DMD / BRM	R Force	Batch webservice	Données des revendeurs du réseau. Liste des salariés des revendeurs.	Hebdomadaire	Les données des revendeurs sont localement stockées dans R Force.
OTS	R Force	Webservice	Campagnes de rappels des véhicules	Requête d'utilisateur	
REP	R Force	Webservice	Programme de maintenance des véhicules	Requête d'utilisateur	
ETICOM AOC	R Force	Batch webservice	Description commerciale des véhicules	Hebdomadaire	Reliés à la mise à jour BVM
BCS	R Force	Téléchargement Webservice	Données clients (Client, Véhicule, Propriété)	Requête d'utilisateur	
RForce	Données client	Requête d'utilisateur	Données clients	Requête d'utilisateur	Mise à jour auprès du client pour la création et la modification de données

ANNEXE 4 : Dictionnaire de données de la base BCS.

Attribut	Attribut	Désignation	Type
Customer Id.	ID client	Clé d'identifier client	Texte (10)
Customer Id. MYR	ID MyR	Clé d'identification du membre My Renault	Texte (10)
Customer type	Type client	Type de client	Entier
Customer Status	Etat client	L'état du client	Entier
Staff purchase	Indicateur personnel constructeur	Indicateur personnel constructeur	Booléen
To delete	Client a supprimer	Client à supprimer	Booléen
Special customer	Client spécial	Client spécial	Booléen
Language	Code Langue	Code de la langue	Texte (3)
Id. 1	Identifiant 1		
Id. 2	Identifiant 2		
Decease	Flag indicateur de décès	Indicateur de décès	Booléen
Bankruped	Flag indicateur de cessation d'activité	Indicateur de cessation d'activité	Booléen
Stop comm. global from	Date début stop comm	Date de début de communication	Date/Heure
Stop comm. global to	Date fin stop comm	Date de fin de de communication	Date/Heure
Stop comm. RC from	Date début stop Comm. RC	Date de début de communication avec la relation client	Date/Heure
Stop comm. RC to	Date fin stop Comm. RC	Date de fin de communication avec la relation client	Date/Heure
Address KO	NPAI adresse	N'habite Pas à l'Adresse indiquée	Booléen
Preferred dealer	Affaire de préférence	Le concessionnaire préféré du client	Entier
Selling dealer	Affaire vendeuse	Le concessionnaire vendeur	Entier
After sales dealer	Affaire après-vente	Le concessionnaire après-vente	Entier
Nearest dealer (BIR)	Affaire territoire	Plus proche concessionnaire	Entier
Last name 1 / Company name	Nom 1 / Raison sociale	Nom du client	Texte (15)
Mr/Mrs	Civilité	Civilité	Texte (3)

Title1	Distinction	Distinction	Texte (5)
Title2	Distinction 2	Distinction	Texte (5)
Last name 2	Nom 2	Nom du client 2	Texte (15)
Initials	Initiales	Initiales	Texte (1)
First name 1	Prénom 1	Prénom du client	Texte (15)
First name 2	Prénom 2	Prénom du client 2	Texte (15)
Gender	Sexe	Genre	Booléen
Home phone number	Num tel 1	Numéro de téléphone / Maison	Entier
Home phone KO	Flag Tel 1 KO	Le client accepte d'être contacté par téléphone /Maison	Booléen
Work phone number	Num tel 2	Numéro de téléphone / Travail	Entier
Work phone KO	Flag Tel 2 KO	Le client accepte d'être contacté par téléphone / Travail	Booléen
Mobile phone number	Num tel 3	Numéro de téléphone / Mobile	Entier
Mobile phone KO	Flag Tel 3 KO	Le client accepte d'être contacté par téléphone / Mobile	Booléen
Fax number	Fax	Numéro de téléphone / Fax	Entier
Home E-mail	E-mail - perso	E-mails personale	Texte (25)
Work E-mail	E-mail - pro	E-mails professionnel	Texte (25)
Preferred media	Media de contact	Média préféré du contact	Texte (5)
Birth date	Date de naissance	Date de naissance du client	Date/Heure
Marital status	Etat civil	Etat civil	Texte (10)
Number of adults	Nb de personnes au foyer	Nombre de personne au foyer	Entier
Number of children	Nb d'enfants au foyer	Nombre d'enfants	Entier
Profession	Profession / Catégorie socio professionnelle	Profession	Texte (25)
Creation Origin	Origine de création de l'information	Origine de la création de l'information	Entier
Creation Date	Date de création de l'information	Date de la création de l'information	Date/Heure
Update Origin	Origine de mise à jour de l'information	Origine de la mise à jour de l'information	Entier
Update Date	Date de mise à jour de l'information	Date de la mise à jour de l'information	Date/Heure

Action Type	Mode de création	Mode de création	Booléen
repurchaser vehicle	Re-acheteur véhicule	Re-acheteur de véhicule	Booléen
Street name	Nom de la voie	Nom de la rue	Texte (150)
Address complement 1	Complément adresse 1	Complément adresse 1	Texte (150)
Address complement 2	Complément adresse 2	Complément adresse 2	Texte (150)
Street number	N° de la voie	Numéro de la voie	Entier
Street type	Type de voie	Type de la voie	Texte (15)
City	Ville	Ville	Texte (30)
Country	Pays	Pays	Texte (40)
Postal code	Code postal	Code postal	Entier
Region / district	Province	Wilaya	Texte (20)
PO Box post code	Code postal boîte	Boîte postale / code postal	Entier
PO Box number	N° boîte	Numéro de la boîte postale	Entier
PO Box location	Ville boîte	Boîte postal Ville	Texte (30)
Date of address update	Date MAJ adresse	Date de mise à jour de l'adresse	Date/Heure
Address quality flag	Address quality flag	Qualité de l'adresse	Booléen
Action Type Address	Mode de création	Mode de création de l'adresse	Booléen
Address quality Id	Adresse ID qualité	Id qualité adresse	Entier
Communication agreement	Accord Client	L'accord client	Booléen
Agreed communication brand	Marque sur laquelle porte l'accord	Marque sur laquelle porte l'accord	Texte (7)
Agreed communication medium	Media concerne par l'accord client	Media concerné par l'accord client	Texte (10)
Communication agreement date	Date MAJ de l'accord	Date de mise à jour de l'accord	Date/Heure
Email communication agreement	Accord Client e-mail	Accord communication client	Booléen
Action Type Com Agreement	Mode de création	Mode de création	Booléen

Aco Row_id	Row_id de l'accord client	Id de l'accord client	Entier
Relationship type	Type de relation	Type de relation	Entier
Relationship start date	Date de début de relation	Date début de la relation	Date/Heure
Relationship end date	Date de fin de relation	Date de fin de la relation	Date/Heure
New/Used	VN/VO	Neuf / Occasion	Booléen
Annual mileage	Km annuel parcouru	Km parcouru annuellement	Entier
Annual mileage date	Date de déclaration km annuel	Date de déclaration du Km annuel	Date/Heure
Current mileage	Km cumule	Km Cumulé	Entier
Current mileage date	Date de renseignement du km cumulé	Date de renseignement du KM cumulé	Date/Heure
Salesman	Nom vendeur	Nom du vendeur	Texte (30)
Payment method	Identifiant mode de paiement	Mode de paiement	Texte (15)
Order Date	Date de commande	Date de commande	Date/Heure
Delivery Date	Date de livraison	Date de livraison	Date/Heure
Action Type Cust Veh Relation	Mode de création	Mode de création de l'information	Booléen
Rel Row_id	Row_id de la relation Cli/Veh	Identifiant de la relation client/Véhicule	Entier
Car Id. (VIN)	VIN	VIN	Texte (15)
Brand	Marque	Marque sur laquelle porte l'accord	Texte (7)
Model	Modèle	Modèle	Texte (3)
Version	Version	Version	Texte (15)
Fuel type	Energie	Type d'énergie	Entier
Number of doors	Nb de portes	Nombre de porte	Entier
Gearbox	Type de boite	Boite de vitesse	Booléen
Car type	VP / VU	Type du véhicule	Booléen
Last registration number	N° immatriculation	Numéro d'immatriculation	Texte (10)
Last registration date	Date d'immatriculation	Date d'immatriculation	Date/Heure
Creation Origin Vehicle	Origine de création de l'information	Origine de la création de l'information véhicule	Entier

Creation Date Vehicle	Date de création de l'information	Date de la création de l'information véhicule	Date/Heure
Update Origin Vehicle	Origine de mise à jour de l'information	Origine de mise à jour de l'information véhicule	Entier
Update Date Vehicle	Date de mise à jour de l'information	Date de mise à jour de l'information véhicule	Date/Heure
Action Type Vehicle	Mode de création	Mode de création	Booléen
Veh Row_id	Row_id du véhicule	Id du véhicule	Texte (10)

ANNEXE 5 : Analyse descriptive de la base de données BCS.

Diagnostic BCS Renault / Dacia



DIRECTION
COMMERCIALE

Yacine HADDAM

Hadjer BENAICHE

CONFIDENTIAL
PROPERTY OF GROUPE RENAULT

GROUPE RENAULT

PRESENTATION DE L'ETUDE METHODOLOGIE

Diagnostic de la Base Client Simplifiée, mise en place en 2010.

Périodes d'achat des clients représentées :

- Renault du **21/11/2007** jusqu'au **02/03/2016**.
- Dacia du **09/08/2007** jusqu'au **15/09/2015**.

Données :

Renault BCS		
Client	298 390	100 %
Doublons	1369	0.45%
Clients ayant plus de 4 achats	34 425	11.6%

Dacia BCS		
Client	167 283	100 %
Doublons	535	0.31%
Clients ayant plus de 4 achats	8 787	5.3%



DIRECTION
COMMERCIALE

Yacine HADDAM

Hadjer BENAICHE

CONFIDENTIAL
PROPERTY OF GROUPE RENAULT

2

GROUPE RENAULT

OBSERVATIONS

DECALAGES DES DONNEES

Ajustement de plus **2200** lignes décalées, soit un total de **14 680** champs.

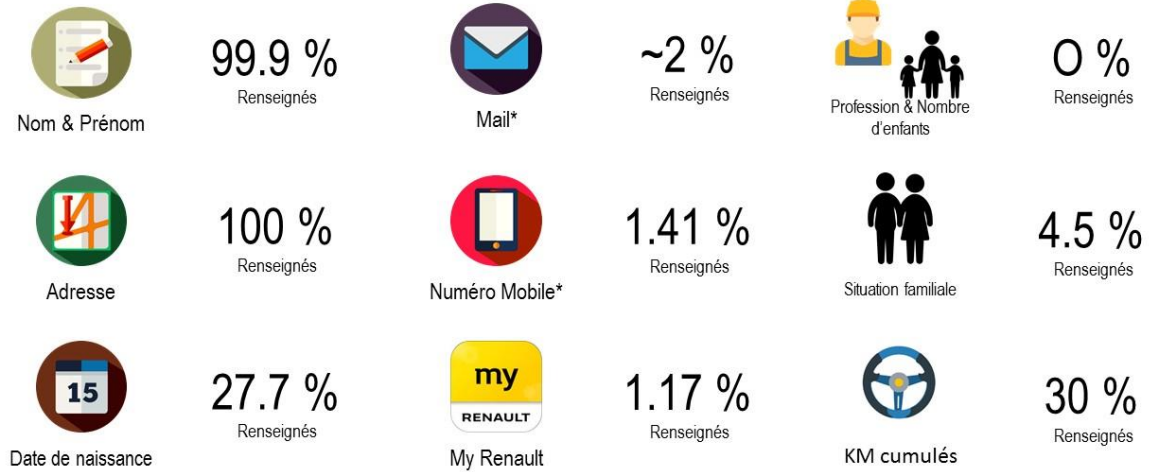
Origine des décalages :

Code opérateur	Pourcentage des décalages
20	4.96 %
30	12.6 %
31	49.91 %
74	16.28 %
76	16.25 %
Total	100 %

J	AK	AL	AM	AN	AO	AP	AQ	AR
phon	Fax number	Home E-mail	Work E-mail	Preferred m	Birth date	Marital statu	Number of a	Number of c
l 3	Fax	E-mail - pi	E-mail - pt	Media de	Date de n.	Stat civil	Nb de per	Nb d'enfa
3E+12	N				FAF			
3E+12	N				FAF	19760126		1
3E+12	N				FAF	19790101		
3E+12	N				FAF	19850922		
3E+12	N				FAF	19830119		
3E+12	N				FAF	19830919		
3E+12	N				FAF	19770922		1
3E+12	N				FAF	19830615		
N					FAF			
3E+12	N				FAF			
3E+12	N				FAF	19631118		
N					FAF			
3E+12	N				FAF	19770922		1
3E+12	N				FAF	19890801		1
3E+12	N				FAF	19790101		1
3E+12	N				FAF	19850922		

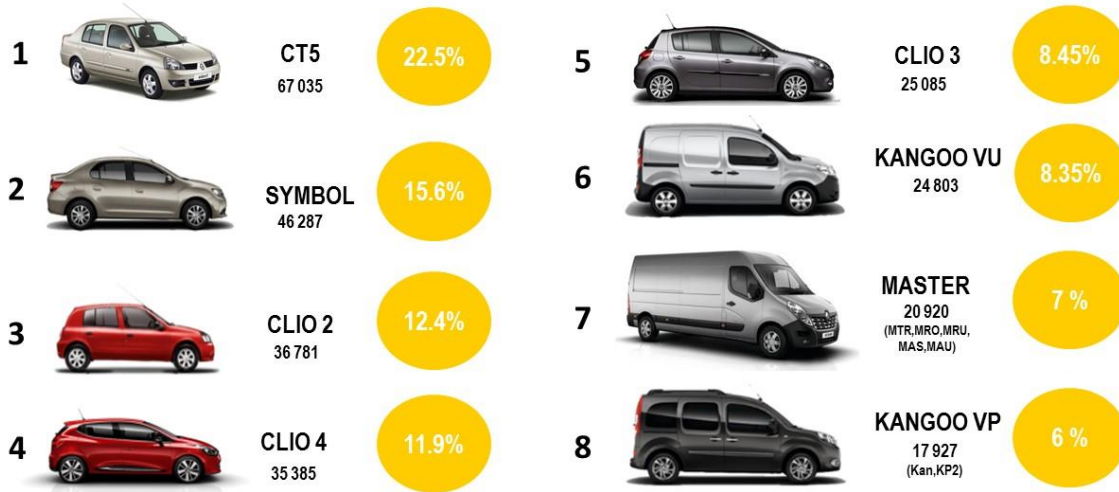
OBSERVATIONS_BCS RENAULT

CHAMPS RENSEIGNES_RENAULT



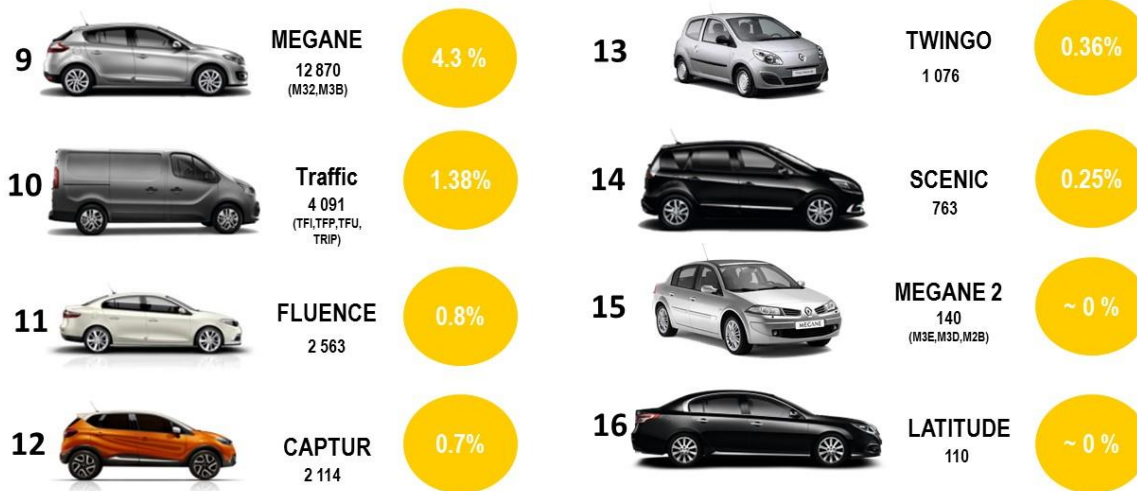
PRESENTATION DE L'ETUDE

CLIENTS PAR MODELE_RENAULT



PRESENTATION DE L'ETUDE

CLIENTS PAR MODELE_RENAULT



OBSERVATIONS
CLIENTS_RENAULT



63.9 %
Du nombre total de clients

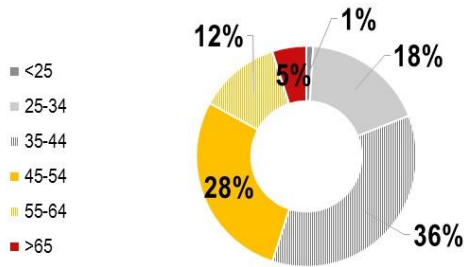


43.4 ans
Age moyen des clients de la marque Renault*

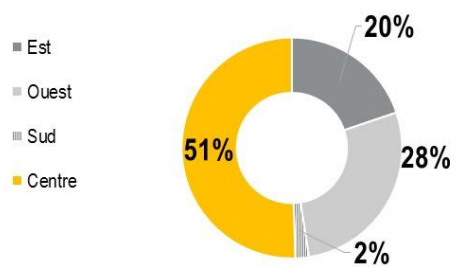


83.24 %
Des véhicules de la marque sont VP

Répartition par tranche d'âge*



Répartition par région**



* Calculé sur une base de 81 656 (champs renseignés)

** Calculé sur une base de 260 662 (champs renseignés)

OBSERVATIONS
CLIENTS RENAULT / HOMME



89 %
Des clients Renault

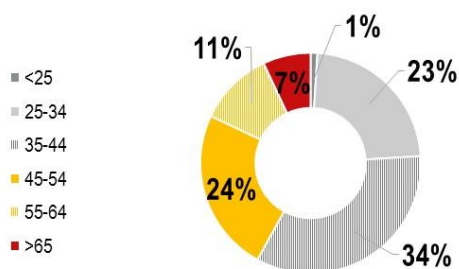


~ 43.2 ans
Age moyen

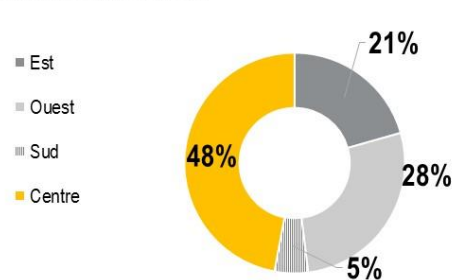


16.57 %
Top modèle 46L

Répartition par tranche d'âge



Répartition par région



OBSERVATIONS

CLIENTS RENAULT / FEMME



11 %
Des clients Renault

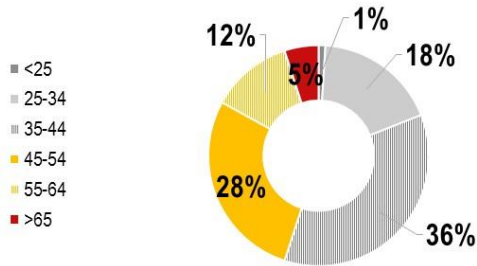


~ 45.5 ans
Age moyen

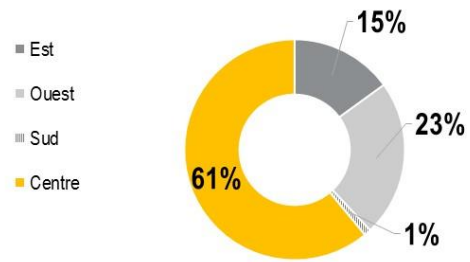


16.57 %
Top modèle CL4

Répartition par tranche d'âge

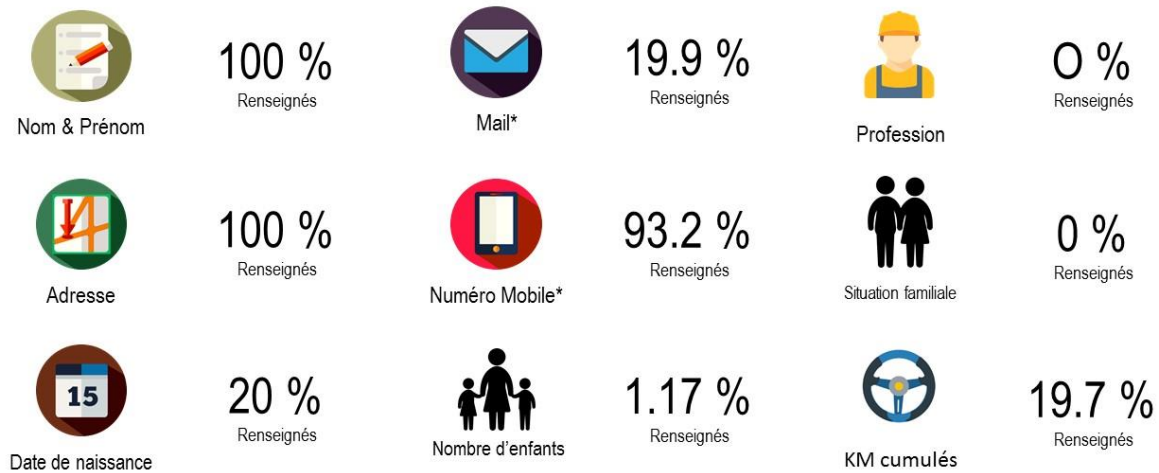


Répartition par région

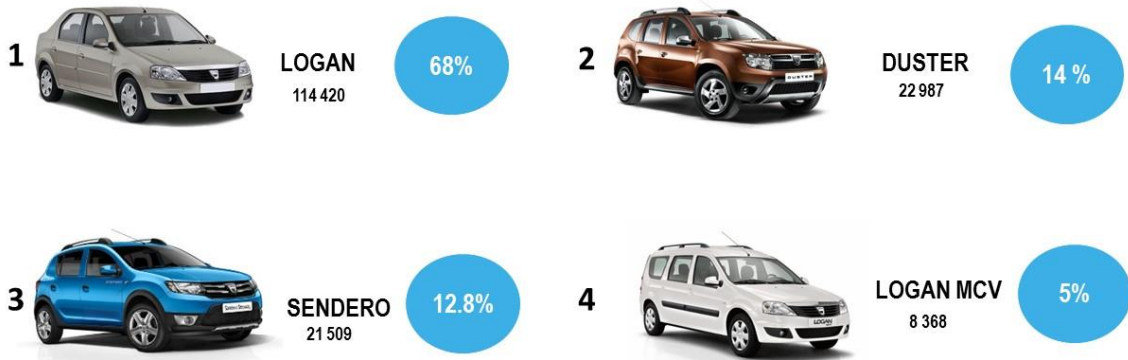


OBSERVATIONS

CHAMPS RENSEIGNES / DACIA



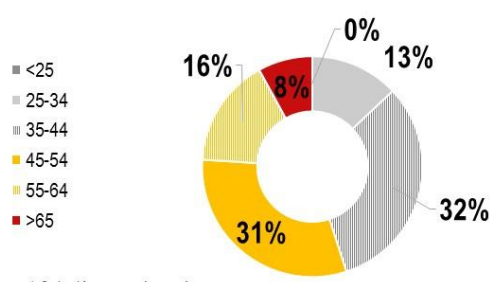
PRESENTATION DE L'ETUDE
CLIENTS PAR MODELE_DACIA



OBSERVATIONS
CLIENTS_DACIA

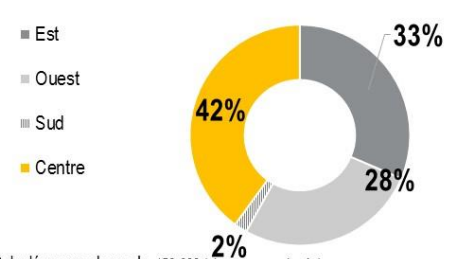


Répartition par tranche d'âge*



* Calculé sur une base de 31 093 (champs renseignés)

Répartition par région**



* Calculé sur une base de 156 688 (champs renseignés)

OBSERVATIONS

CLIENTS DACIA / HOMME



92.4 %
Des clients DACIA

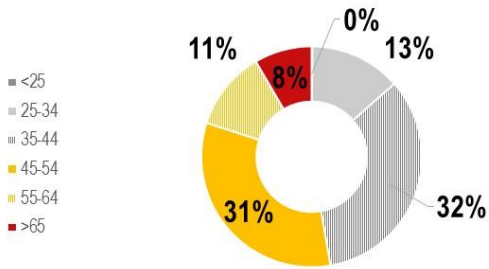


~ 47.2 ans
Age moyen

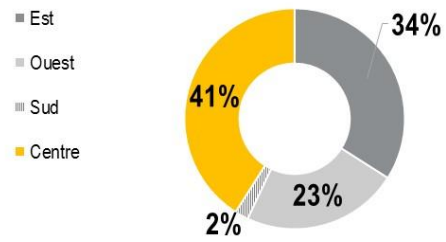


~70 %
Top modèle des achats hommes

Répartition par tranche d'âge



Répartition par région



OBSERVATIONS

MARQUE DACIA / FEMME



7.6 %
Des clients DACIA

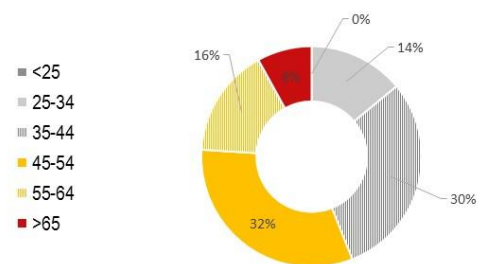


~ 46.7 an
Age moyen

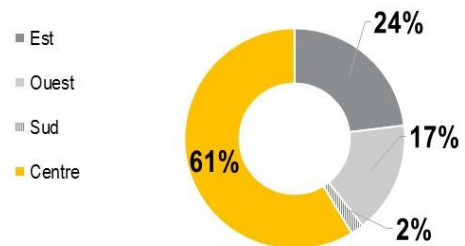


27.8 %
Top modèle des achats des femmes

Répartition par tranche d'âge



Répartition par région



ANNEXE 6 : Questionnaire de l'enquête relative aux clients de l'entreprise.

Il s'agit d'un fichier PDF rédigé à l'intention de l'équipe du centre d'appel pour le sondage téléphonique assisté par ordinateur.

Enquete : Clients de Renault Algérie

Texte de présentation

Bonjour Madame / Monsieur (NOM ET PRENOM DU CLIENT)

Je vous appelle de la part de l'entreprise Renault Algérie.

Nous réalisons actuellement une enquête auprès de nos clients.

Le but de cette enquête est de recueillir vos avis et pour mieux comprendre vos attentes et pouvoir ensuite vous proposer des produits et des services qui correspondent à vos besoins.

Nous souhaiterions donc vous poser quelques questions relatives à votre avis sur l'automobile ainsi que vos démarches avant, pendant et après l'achat de votre voiture.

Renault Algérie vous garantit que vos réponses resteront **strictement confidentielles**.

Notre conversation ne durera qu'une quinzaine de minutes.

Partie A : Questions

Q0. Avez-vous acheté la voiture (**MODELE**) de la marque (**MARQUE**) en (**ANNEE DE FACTURATION**) ? (**NE RIEN SUGGERER _ UNE SEULE REPONSE POSSIBLE**)

- Oui => Continuer
- Non, autre personne disponible => Demander la personne concernée
- Non, autre personne indisponible => Stop inter
- Refus de participer => Stop inter
- Ne sait pas => Stop inter

Q1. Je vais vous citer un certain nombre de critères qu'on prend généralement en compte lors de l'achat d'une voiture, dites-moi, sur une échelle de 1 à 5, quel est le degré d'importance que vous accordez à chacun d'entre eux : (**ENUMERER UN A UN _ UNE SEULE REPONSE POSSIBLE PAR PROPOSITION**)

	1	2	3	4	5
Le prix	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La consommation de carburant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le prestige / le statut social	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le design / la couleur	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le confort / l'espace	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La sécurité	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le niveau d'équipements	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La nouveauté du modèle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q2. Mis à part les opérations prises en charge par la garantie constructeur, quelles sont les opérations que vous réalisez chez les concessionnaires / agents agréés par la marque de votre voiture ? (**ENUMERER PLUSIEURS REPONSES POSSIBLES**)

- Opérations de maintenance (entretien/usure)
- Réparations mécaniques non planifiées, dont panne
- Opérations de remplacement des pneus
- Opérations de carrosserie (Réparation collision, bris de glace)
- Aucune opération, je me rends à des garagistes indépendants

Partie A : Questions

Q3. Je vais vous citer un certain nombre de moyens qu'on utilise pour s'informer sur les voitures. Pouvez-vous me dire votre degré d'utilisation, de 1 à 5, pour chacun d'entre eux ?

(ENUMERER UN A UN _ UNE SEULE REPONSE POSSIBLE PAR PROPOSITION)

	1 (Pas du tout utilisée)	2	3	4	5 (Toujours utilisée)
Affiches publicitaires	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Journaux / magazines	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TV / programmes automobiles	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Radio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Réseaux sociaux / blogs / forums	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Site Web de la marque	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Autres sites Web	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Avis des proches	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Avis des concessionnaires	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4. Combien de kilomètres parcourez-vous, en moyenne, par jour ? (ENUMERER_ UNE SEULE REPONSE POSSIBLE)

- Moins de 20 km
- Entre 20 et 50 km
- Entre 50 et 100 km
- Entre 100 et 200 km
- Plus de 200 km

Partie B : Renouvellement d'achat

Q5. Allez-vous acheter une nouvelle voiture dans les 6 prochains mois ? (NE RIEN SUGGERER_UNE SEULE REPONSE POSSIBLE)

- Oui
- Non
- Je ne sais pas

SI « Non » OU « Je ne sais pas », ALLEZ A LA PARTIE C.

Q6. Vous voulez acheter une voiture neuve ou d'occasion ? (NE RIEN SUGGERER_UNE SEULE REPONSE POSSIBLE)

- Neuve
- D'occasion
- Je ne sais pas

Q7. Dans quel segment (catégorie) souhaitez-vous acheter votre nouvelle voiture ? (ENUMERER ET CITER LES EXEMPLES SI NECESSAIRE_UNE SEULE REPONSE POSSIBLE)

- Petite citadine (Hyundai i10, Kia Picanto, Renault Twingo).
- Citadine (Citroen C3, Peugeot 208, Renault Clio)
- Compacte (Dacia Logan, Peugeot 308, Renault Mégane, VW Golf)
- Ludospace (Citroen Berlingo, Peugeot Partner, Renault Kangoo)
- Berline (Audi A4, Peugeot 508, Volkswagen Passat)
- SUV / 4*4 (BMW X3, Renault Captur, Range Rover Evoque)

Q8. Combien allez-vous dépenser pour cette nouvelle voiture ? (NE RIEN SUGGERER_UNE SEULE PROPOSITION POSSIBLE)

- Moins de 1 000 000 DA
- Entre 1 000 000 et 1 300 000 DA
- Entre 1 300 000 et 1 600 000 DA
- Entre 1 600 000 et 2 000 000 DA
- Entre 2 000 000 et 2 500 000 DA
- Entre 2 500 000 et 3 000 000 DA
- Entre 3 000 000 et 4 000 000 DA
- Plus de 4 000 000 DA

Partie C : Informations générales

(COCHER LE GENRE DU REpondANT)

- Homme
- Femme

Q9. Quelle est votre date de naissance ? (NE RIEN SUGGERER_UNE SEULE REponse POSSIBLE)

Jour	Mois	Année
<input type="text"/>	<input type="text"/>	<input type="text"/>

Q10. Quelle est votre principale profession ? (NE RIEN SUGGERER_UNE SEULE REponse POSSIBLE)

- Agriculteur exploitant
- Cadre supérieur / Profession libérale / patron d'industrie
- Cadre moyen
- Employé
- Ouvrier / Manœuvre (salarié)
- Enseignant
- Médecin / pharmacien
- Commerçant / Artisan à son compte
- Femme au foyer
- Etudiant
- Retraité
- Chômeur / Sans emploi

Q11. Quel est votre niveau d'instruction ? (NE RIEN SUGGERER_UNE SEULE REponse POSSIBLE)

- Niveau primaire
- Niveau moyen
- Niveau secondaire
- Niveau Bac
- Etudes supérieures

Q12. Quelle est votre situation familiale ? (ENUMERER_UNE SEULE REponse POSSIBLE)

- Célibataire
- Marié (e)
- Séparé(e) / Divorcé(e) / Veuf (ve)

Q13. Quel est le nombre d'enfants à votre charge ? (NE RIEN SUGGERER_UNE SEULE REponse POSSIBLE)

- 0
- 1
- 2
- 3
- Plus

Q14. Quel est le type de votre habitation ? (ENUMERER_UNE SEULE REponse POSSIBLE)

- Maison individuelle / villa
- Maison collective
- Appartement

Q15. Etes-vous propriétaire, locataire ou habitez-vous chez des proches ? (UNE SEULE REponse POSSIBLE)

- Propriétaire
- Locataire
- Habite chez des proches

Q16. Quel est le nombre de voitures dans votre foyer ? (NE RIEN SUGGERER_UNE SEULE REponse POSSIBLE)

- 0
- 1
- 2
- 3
- 4 et plus

Partie C : Informations générales

Q17. Quelle est votre wilaya de résidence ?
(NE RIEN SUGGERER_ UNE SEULE
REPONSE POSSIBLE)

Liste déroulante

Q18. Pouvez-vous me donner une
approximation du revenu de votre foyer ?
(ENUMERER SI NECESSAIRE_ UNE
SEULE REPONSE POSSIBLE)

- Moins de 18 000 DA
- Entre 18 000 et 30 000 DA
- Entre 30 000 et 60 000 DA
- Entre 60 000 et 100 000 DA
- Entre 100 000 et 200 000 DA
- Entre 200 000 et 300 000 DA
- Plus de 300 000 DA

FIN

Nous vous remercions vivement pour votre collaboration.

ANNEXE 7 : Fichier Excel de l'enquête relative aux clients de l'entreprise.

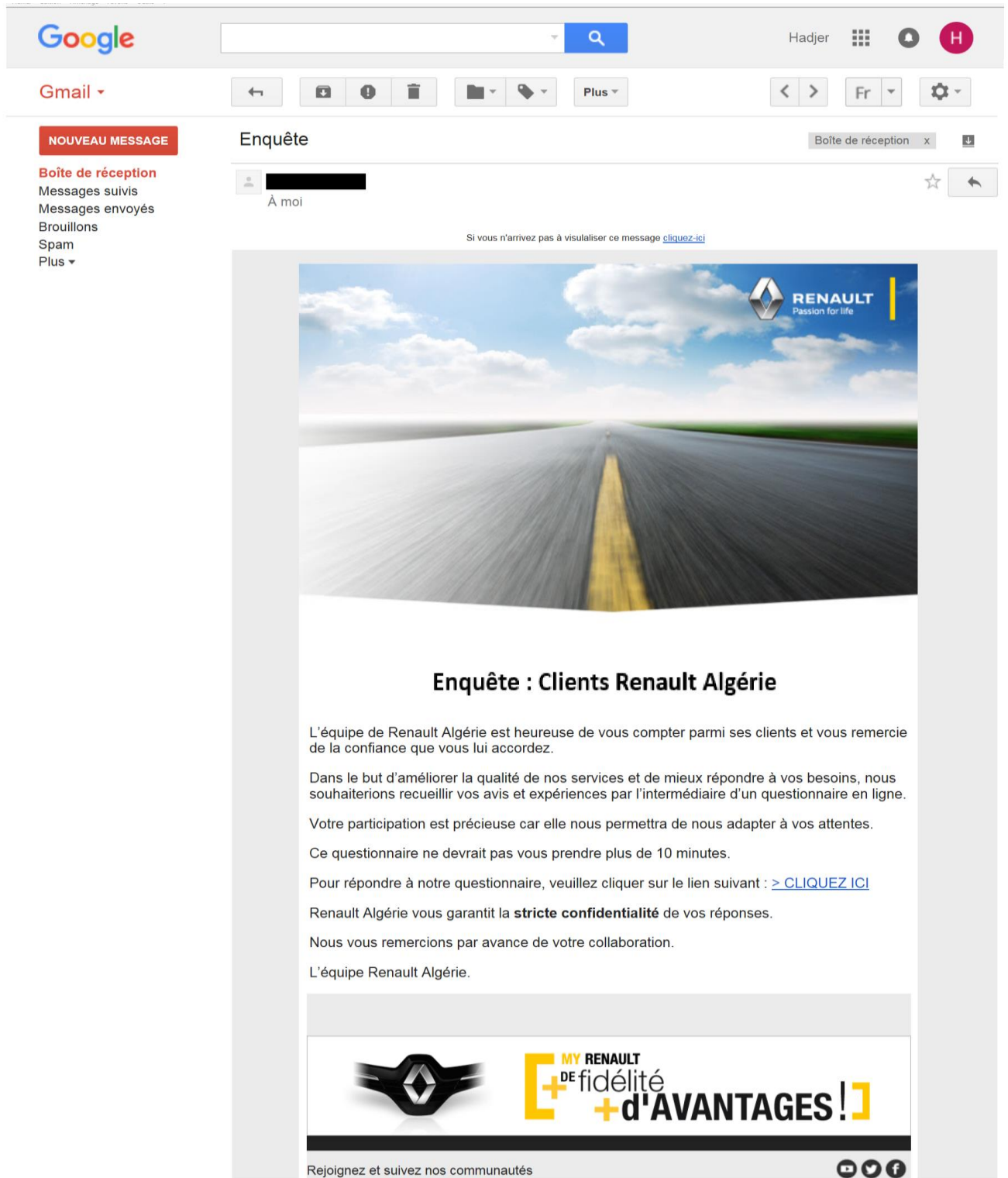
Il s'agit d'un aperçu du fichier Excel créé à l'intention de de l'équipe du centre d'appel dans le cadre de l'enquête réalisée auprès des clients de Renault Algérie.

Note : Pour des raisons évidentes de confidentialité, les noms, prénoms et numéros de téléphones sont masqués dans l'aperçu du fichier.

Identification								Confirmation		Critères de choix			
ID_Client	VIN	Marque	Modèle	Date	Nom	Prénom	Tel 1	Tel 2	Tel 3	Confirmation	Q2. Je vais vous citez un certain nombre de	Le prix	La consommation
1-8BPK-21	VF1BR12SH4	RENAULT	CL3	2 011									
1-462R-672	VF1FW0LB54	RENAULT	KU2	2010									
1-HLDO-1065	VF1SR060H4	RENAULT	CL4	2 012									
1-DNRY-109	VF1LZBK064	RENAULT	FLU	2 012									
1-REWH-234	VF1FW0LB5	RENAULT	KU2	2 014									
1-BR43-92	VF1BR5G0H4	RENAULT	CL3	2 011									
1-VQ80-29	VF1MAF4KC	RENAULT	MAU	2 015									
1-DNRY-508	VF1LZBK064	RENAULT	FLU	2 012									
1-FPVM-107	UU1LSD4CH4	DACIA	90L	2012									
1-RSTX-6	VF14SRAP45	RENAULT	46L	2 014									
1-B871-266	UU1LSDACH4	DACIA	90L	2011									
1-BGCO-295	VF1MAF4KC	RENAULT	MAU	2 011									
1-OTLW-79	VF14SRAP45	RENAULT	46L	2 014									
1-B88Q-82	VF1BR12SH4	RENAULT	CL3	2 011									
1-10BK2-190	VF14SR2P4F	RENAULT	46L	2 015									
1-RQ0U-6	VF1MAF6KJ5	RENAULT	MAU	2 014									
1-XSHU-843	VF1MAF4KE5	RENAULT	MAU	2 015									
1-SUV6-140	VF1SR060H5	RENAULT	CL4	2 014									


ANNEXE 8 : E-mail du sondage sur les clients de l'entreprise.

Il s'agit d'un aperçu de l'e-mail rédigé et envoyé aux clients de Renault Algérie, dans le cadre de notre l'enquête.



The screenshot shows a Gmail inbox with an email titled "Enquête". The email content is as follows:

Si vous n'arrivez pas à visualiser ce message [cliquez-ici](#)



Enquête : Clients Renault Algérie

L'équipe de Renault Algérie est heureuse de vous compter parmi ses clients et vous remercie de la confiance que vous lui accordez.

Dans le but d'améliorer la qualité de nos services et de mieux répondre à vos besoins, nous souhaiterions recueillir vos avis et expériences par l'intermédiaire d'un questionnaire en ligne.

Votre participation est précieuse car elle nous permettra de nous adapter à vos attentes.


Ce questionnaire ne devrait pas vous prendre plus de 10 minutes.

Pour répondre à notre questionnaire, veuillez cliquer sur le lien suivant : [> CLIQUEZ ICI](#)

Renault Algérie vous garantit la **stricte confidentialité** de vos réponses.

Nous vous remercions par avance de votre collaboration.

L'équipe Renault Algérie.



Rejoignez et suivez nos communautés

ANNEXE 9 : Analyse descriptive de l'étude relative aux clients de l'entreprise.

ANALYSE DESCRIPTIVE DE L'ETUDE CLIENTS RENAULT ALGERIE 2016

Hadjer BENAICHE et Yacine HADDAM

1

I. PRESENTATION DE L'ETUDE

METHODOLOGIE

CLIENTS RENAULT ALGERIE 2016 est une enquête réalisée sur un échantillon des clients ayant acheté des véhicules, particuliers ou utilitaires, des marques Renault et Dacia.

La période d'achat couverte par l'étude s'étend du 22 octobre 2010 jusqu'au 16 Mars 2016.

Méthode de sondage : Appels téléphoniques assistés par ordinateur (réalisés par l'équipe du Call Center) accompagnée d'une campagne E-mailing sur toute la base de données des clients.

Durée de l'enquête : 3 semaines, à raison de 5 jours ouvrables par semaine.

Statistiques de l'opération d'appel (résultats transmis par le Call Center) :

Analyse de l'opération enquête clients				
Clients contactés	Clients argumentés	OK	Injoignables	Faux numéros/ Erreur
2151	631	350	990	525

Résultats E-mailing : 210 réponses.

2

ENQUETE CLIENTS RENAULT ALGERIE 2016

II. ANALYSE DESCRIPTIVE

DESCRIPTION DE L'ÉCHANTILLON

Répartition par marque :

Renault : 78,1 %

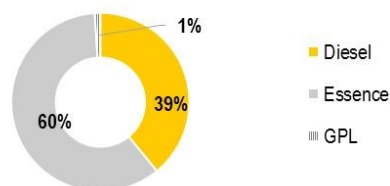
Dacia : 21,9 %

Véhicules représentés (par segment)

Segment	Nb. de véhicules	% de véhicules
M0	158	28,8%
I	195	35,6 %
M1	63	11,5 %
SUV	54	9,9 %
LUDOSPACE	20	3,6 %
VU	58	10,6 %
TOTAL	548	100%

Année d'immatriculation	Nb. de véhicules	% de véhicules
2016	30	5,5 %
2015	114	20,8 %
2014	120	21,9 %
2013	126	23 %
2012	89	16,2 %
2011	43	7,8 %
2010	14	2,6 %
Avant 2010	12	2,7 %

Motorisation



3

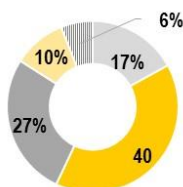
ENQUETE CLIENTS RENAULT ALGERIE 2016

II. ANALYSE DESCRIPTIVE

ANALYSE UNIVARIEE

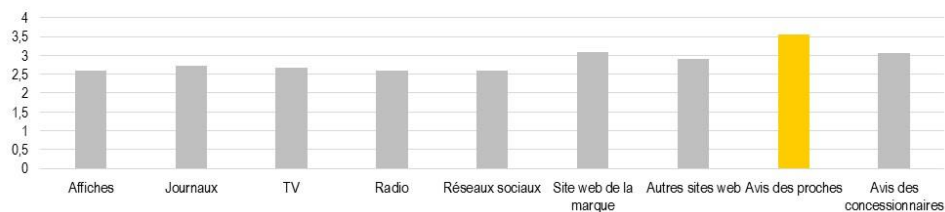
Distance moyenne par jour

- Moins de 20 km
- 20 - 50 km
- 50 - 100 km
- 100 - 200 km
- Plus de 200 km



Opérations	% d'achat
Maintenance	81 %
Réparation	39 %
Changement des pneus	6 %
Carrosserie	13 %
Aucune	16 %

Utilisation des sources d'information

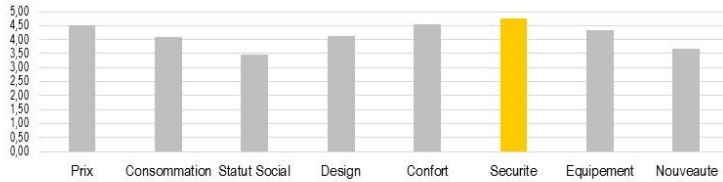


4

ENQUETE CLIENTS RENAULT ALGERIE 2016

II. ANALYSE DESCRIPTIVE
ANALYSE UNIVARIEE

Critères de choix d'une voiture

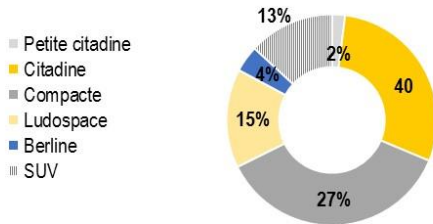


Renouvellement du véhicule

Oui : 78,1 %

Non : 21,9 %

Segment de renouvellement



Montant de renouvellement	% de clients
Moins de 1 000 000 DA	5,5 %
Entre 1 000 000 et 1 300 000 DA	20,8 %
Entre 1 300 000 et 1 600 000 DA	21,9 %
Entre 1 600 000 et 2 000 000 DA	23 %
Entre 2 000 000 et 3 500 000 DA	16,2 %
Entre 3 500 000 et 4 000 000 DA	7,8 %
Entre 1 000 000 et 1 300 000 DA	2,6 %
Plus de 4 000 000 DA	2,7 %

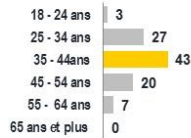
5

ENQUETE CLIENTS RENAULT ALGERIE 2016

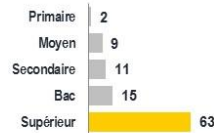
II. ANALYSE DESCRIPTIVE

PROFIL SOCIODEMOGRAPHIQUE_ENTREPRISE

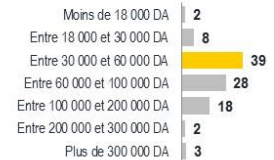
Age



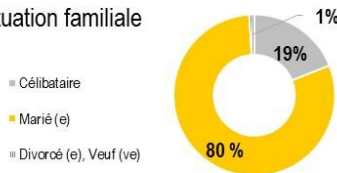
Education



Revenu

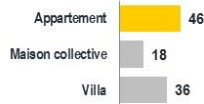


Situation familiale



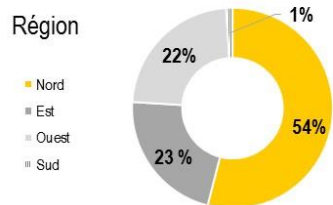
Nombre moyen d'enfants : 1,91

Type d'habitation



Propriétaire : 69,3 %

Région

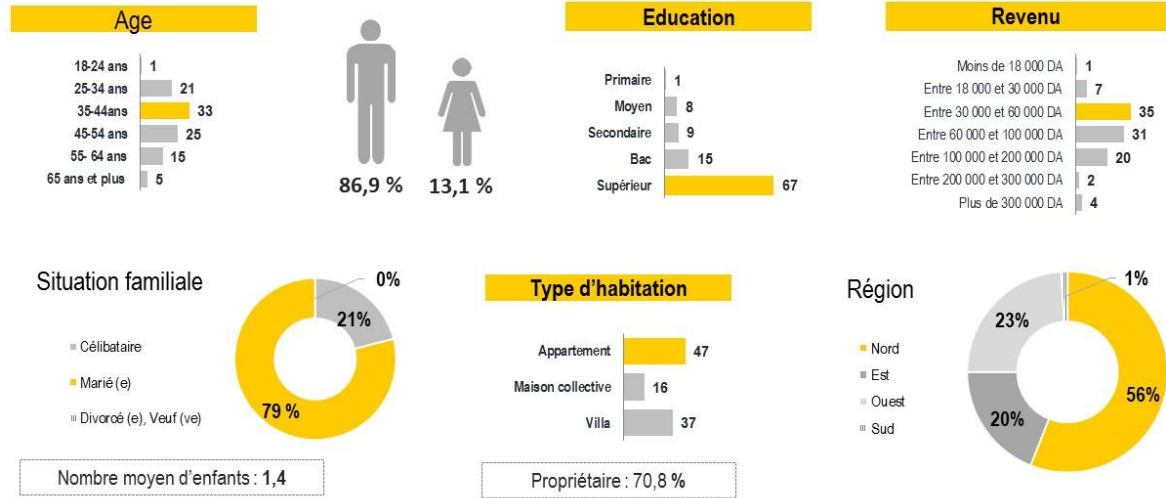


6

ENQUETE CLIENTS RENAULT ALGERIE 2016

II. ANALYSE DESCRIPTIVE

PROFIL SOCIODEMOGRAPHIQUE_RENAULT

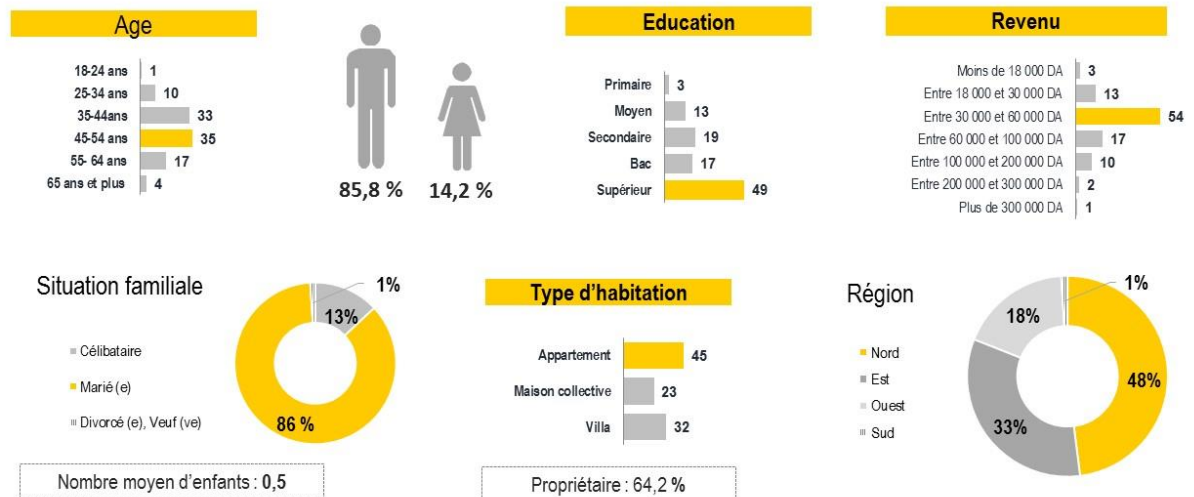


7

ENQUETE CLIENTS RENAULT ALGERIE 2016

II. ANALYSE DESCRIPTIVE

PROFIL SOCIODEMOGRAPHIQUE_DACIA



8

ENQUETE CLIENTS RENAULT ALGERIE 2016

ANNEXE 10 : Fichier texte du questionnaire de l'étude de marché.

Questionnaire : Enquete sur les acheteurs de voitures en Algérie

Bonjour Madame / bonjour monsieur,

Nous sommes deux étudiants en Génie Industriel de l'Ecole Nationale Polytechnique et nous réalisons actuellement une enquête, dans le cadre de notre projet de fin d'études.

Le but de cette enquête est de mesurer la perception et les attentes du consommateur de véhicules, dans le marché de l'automobile en Algérie.

Nous souhaiterions donc vous poser quelques questions afin de connaître votre avis sur l'automobile en général, sur l'expérience d'achat de votre voiture ainsi que sur votre éventuel renouvellement de véhicule.

Il s'agit d'un questionnaire anonyme et toutes vos réponses resteront strictement confidentielles. Cela ne devrait pas vous prendre plus de 15 minutes.

Nous vous remercions par avance de votre collaboration.

Si vous avez des questions concernant le questionnaire, n'hésitez pas à nous écrire à l'adresse suivante :

enquete.automobile.algerie@gmail.com

Votre avis compte pour nous !

Partie A : Avis sur l'automobile

Q1. Quel est votre degré d'accord ou de désaccord avec les propositions suivantes ?

	Pas du tout d'accord	Plutôt pas d'accord	Ni d'accord ni pas d'accord	Plutôt d'accord	Tout à fait d'accord
Les voitures fabriquées à l'étranger sont meilleures en termes de qualité	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La voiture que je conduis doit refléter ma position sociale.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La voiture est plus un moyen de transport qu'une source de plaisir.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ça ne me dérangerait pas d'acheter une voiture fabriquée en Algérie si elle coûte moins chère.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Partie A : Avis sur l'automobile

Q2. Quel est le degré d'importance que vous accordez à chacun des critères suivant dans le choix de votre voiture ?

	Vraiment pas important	Plutôt pas important	Important	Très important	Extrêmement important
Le prix	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La marque	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La consommation de carburant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le prestige / le statut social	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La puissance du moteur	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le design / la couleur	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le confort / l'espace	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La sécurité	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La valeur de la voiture à la revente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le niveau d'équipements	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La nouveauté du modèle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le respect de l'environnement	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Partie A : Avis sur l'automobile

Q3. Lorsque vous vous renseignez sur les voitures, quel est votre degré d'utilisation de chacune des sources d'informations suivantes ?

	1 (Pas du tout utilisée)	2	3	4	5 (Toujours utilisée)
Affiches publicitaires	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Journaux / magazines	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TV / programmes automobiles	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Radio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Réseaux sociaux / blogs / forums	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Site Web de la marque	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Autres sites Web	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Salon de l'automobile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Marché des voitures d'occasion	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Avis des proches	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Avis des concessionnaires	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4. Possédez-vous une voiture ?

- Oui
 Non

SI « Non », ENVOYER VERS PARTIE C.

Partie B : Votre voiture et vous

Q5. Vous avez acheté votre voiture :

- Neuve
- D'occasion
- Je ne l'ai pas achetée, mon entreprise me l'a fournie

SI « Je ne l'ai pas achetée, mon entreprise me l'a fournie », ENVOYER VERS PARTIE C.

Q6. Combien de temps aviez-vous gardé votre ancienne voiture ?

- Moins de 3 mois
- Entre 3 et 6 mois
- 1 an
- 2 ans
- 3 ans
- 4 ans
- Plus de 4 ans
- C'est ma première voiture

Q7. Quelle est la marque de votre voiture ?

Liste déroulante

Q8. Quel est le modèle de votre voiture ?

.....

Q9. Quelle est l'année d'immatriculation de votre voiture ?

Liste déroulante

Partie B : Votre voiture et vous

Q10. Quel est le type de motorisation de votre voiture ?

- Diesel
- Essence
- GPL

Q11. Quel est le type de boîte de vitesse de votre voiture ?

- Automatique
- Manuelle

Q12. Combien de kilomètres parcourez-vous, en moyenne, par jour ?

- Moins de 20 km
- Entre 20 et 50 km
- Entre 50 et 100 km
- Entre 100 et 200 km
- Plus de 200 km

Q13. Aviez-vous participé au choix du modèle de votre voiture ?

- Oui
- Non

Q14. Quelles autres personnes avaient participé au choix de votre voiture ?

- Mon conjoint
- Mes enfants
- Mes parents / Autres membres de ma famille
- Mes amis / collègues
- Mes voisins
- Aucune autre personne

Q15. Mis à part les opérations prises en charge par la garantie constructeur, quelles sont les opérations que vous réalisez chez les concessionnaires / agents agréés par la marque de votre voiture ?

- Opérations de maintenance (entretien/usure)
- Réparations mécaniques non planifiées, dont panne
- Opérations de remplacement des pneus
- Opérations de carrosserie (Réparation collision, bris de glace)
- Aucune opération, je me rends à des garagistes indépendants

Partie C : Votre future voiture et vous

Q16. Allez-vous acheter une nouvelle voiture dans les 6 prochains mois ?

- Oui
- Non
- Je ne sais pas

Q17. Dans quel segment (catégorie) souhaitez-vous acheter votre nouvelle voiture ?

(Les modèles inscrits entre parenthèses sont des exemples pour chaque segment)

- Petite citadine (Hyundai i10, Kia Picanto, Renault Twingo).
- Citadine (Citroen C3, Peugeot 208, Renault Clio)
- Compacte (Dacia Logan, Peugeot 308, Renault Mégane, VW Golf)
- Ludospace (Citroen Berlingo, Peugeot Partner, Renault Kangoo)
- Berline (Audi A4, Peugeot 508, Volkswagen Passat)
- SUV / 4*4 (BMW X3, Renault Captur, Range Rover Evoque)

Q18. Combien allez-vous dépenser pour cette nouvelle voiture ?

- Moins de 1 000 000 DA
- Entre 1 000 000 et 1 300 000 DA
- Entre 1 300 000 et 1 600 000 DA
- Entre 1 600 000 et 2 000 000 DA
- Entre 2 000 000 et 2 500 000 DA
- Entre 2 500 000 et 3 000 000 DA
- Entre 3 000 000 et 4 000 000 DA
- Plus de 4 000 000 DA

Partie D : Informations générales

Q19. Vous êtes :

- Homme
- Femme

Q20. Quel âge avez-vous ?

- Moins de 18 ans
- Entre 18 et 24 ans
- Entre 25 et 34 ans
- Entre 35 et 44 ans
- Entre 45 et 54 ans
- Entre 55 et 64 ans
- Plus de 65 ans

Q21. Quelle est votre principale profession ?

- Agriculteur exploitant
- Cadre supérieur / Profession libérale / patron d'industrie
- Cadre moyen
- Employé
- Ouvrier / Manœuvre (salarié)
- Enseignant
- Médecin / pharmacien
- Commerçant / Artisan à son compte
- Femme au foyer
- Etudiant
- Retraité
- Chômeur / Sans emploi

Q22. Quel est votre niveau d'instruction ?

- Niveau primaire
- Niveau moyen
- Niveau secondaire
- Niveau Bac
- Etudes supérieures

Q23. Quelle est votre situation familiale ?

- Célibataire
- Marié (e)
- Séparé(e) / Divorcé(e) / Veuf (ve)

Q24. Quel est le nombre d'enfants à votre charge ?

- 0
- 1
- 2
- 3
- Plus

Q25. Quel est le type de votre habitation ?

- Maison individuelle / villa
- Maison collective
- Appartement

Q26. Etes-vous propriétaire, locataire ou habitez-vous chez des proches ?

- Propriétaire
- Locataire
- Habite chez des proches

Q27. Vous habitez dans une zone :

- Urbaine (ville)
- Rurale (En dehors de la ville)

Q28. Quel est le nombre de voitures dans votre foyer ?

- 0
- 1
- 2
- 3
- 4 et plus

Partie D : Informations générales

Q29. Quelle est votre wilaya de résidence ?

Liste déroulante

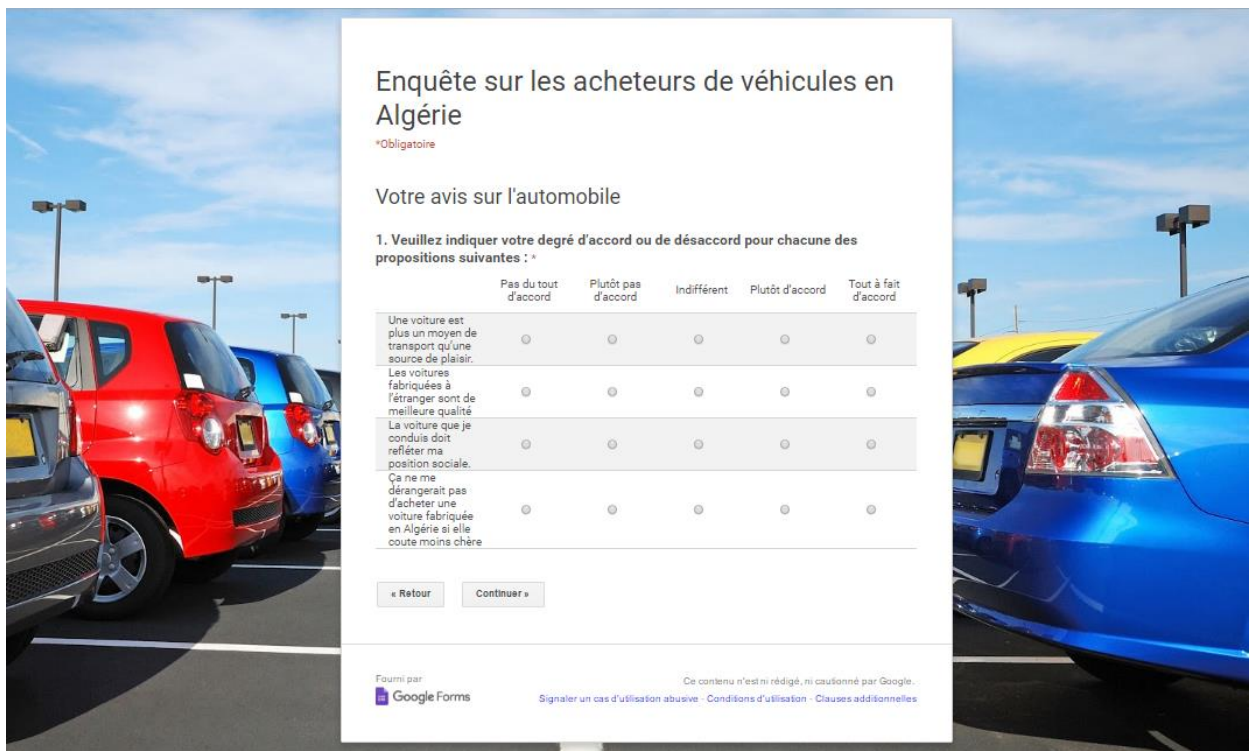
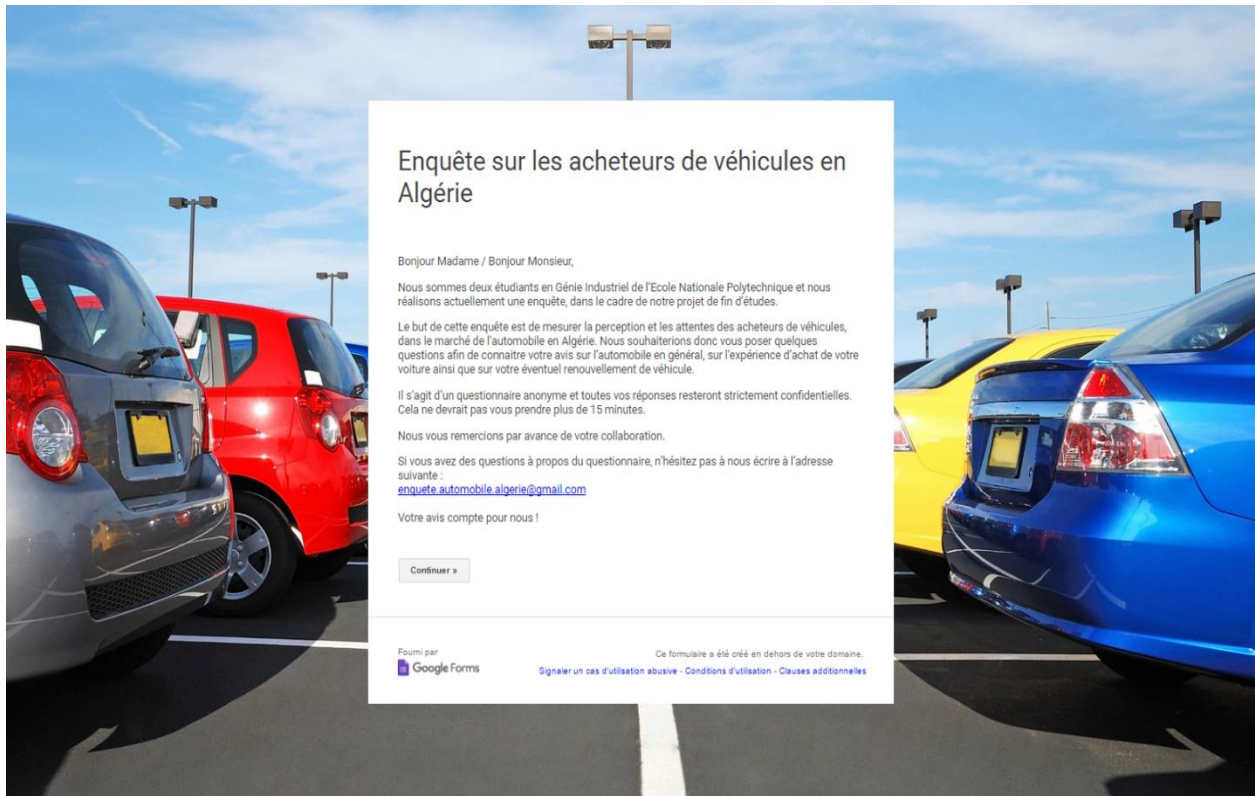
Q30. Quel est le revenu mensuel de votre foyer ?

- Moins de 18 000 DA
- Entre 18 000 et 30 000 DA
- Entre 30 000 et 60 000 DA
- Entre 60 000 et 100 000 DA
- Entre 100 000 et 200 000 DA
- Entre 200 000 et 300 000 DA
- Plus de 300 000 DA

FIN

Nous vous remercions vivement pour votre collaboration.

ANNEXE 11 : Aperçu du questionnaire, en format électronique, de l'étude de marché (pages 1 et 2).



ANNEXE 12 : Analyse descriptive de l'étude de marché.

ANALYSE DESCRIPTIVE DE L'ETUDE MARCHE AUTOMOBILE EN ALGERIE

Hadjer BENAICHE et Yacine HADDAM

1

I. PRESENTATION DE L'ETUDE METHODOLOGIE

L'étude du Marché automobile en Algérie : est une enquête réalisée auprès des acheteurs de véhicules en Algérie, de toutes marques confondues.

Méthode de sondage : Sondage par internet à travers la diffusion d'un formulaire en ligne dans différents réseaux sociaux et les forums spécialisés.

Durée de l'enquête : 5 semaines.

Statistiques :

Analyse de données récoltées			
Réponses récoltées	Acheteurs de véhicules	Doublons	Echantillon final
473	316	6	310

2

ETUDE DU MARCHE AUTOMOBILE EN ALGERIE

II. ANALYSE DESCRIPTIVE

DESCRIPTION DE L'ECHANTILLON

Répartition par Marques (TOP 5) :

Peugeot : **15,8** %
 Volkswagen : **12** %
 Renault : **10** %
 Seat : **6,8** %
 Skoda : **6,1** %

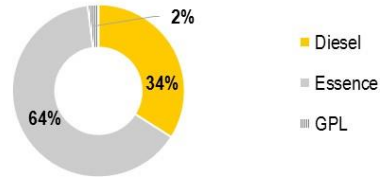
Nbr de véhicules par segment

Segment	Nb. de véhicules	% de véhicules
Petite citadine	40	12,7 %
Citadine	149	48 %
Compacte	85	27,4 %
LUDOSPACE	6	2 %
Berline	8	2,5 %
SUV/4*4	23	7,5 %
TOTAL	310	100%

Répartition par âge du véhicule :

Année d'immatriculation	Nb. de véhicules	% de véhicules
2016	18	5,8 %
2015	44	14,2 %
2014	56	18 %
2013	47	15,2 %
2012	51	16,5 %
2011	13	4,2 %
2010	15	4,9 %
Avant 2010	81	21,2 %

Répartition par motorisation



3

ETUDE DU MARCHE AUTOMOBILE EN ALGERIE

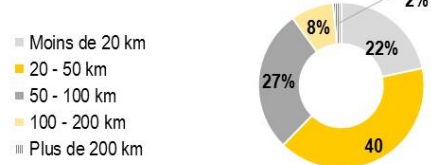
II. ANALYSE DESCRIPTIVE

ANALYSE UNIVARIEE

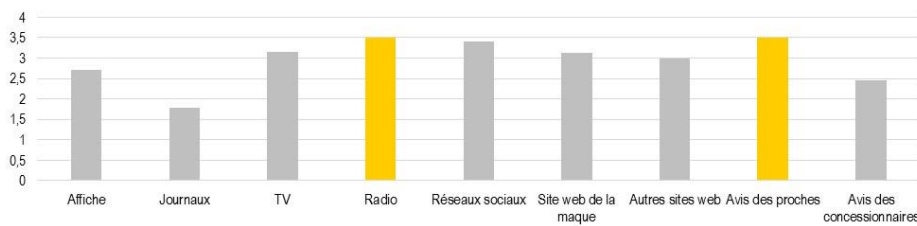
Achat de services après-vente:

Opérations	% d'achat
Maintenance	50,3 %
Réparation	22,3 %
Changement des pneus	6,5 %
Carrosserie	8 %
Aucune	44,5 %

Distance moyenne, parcourue par jour :



Utilisation des sources d'information



4

ETUDE DU MARCHE AUTOMOBILE EN ALGERIE

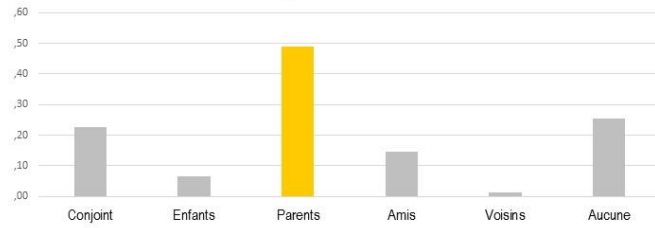
II. ANALYSE DESCRIPTIVE
ANALYSE UNIVARIEE

Participation au choix du véhicule

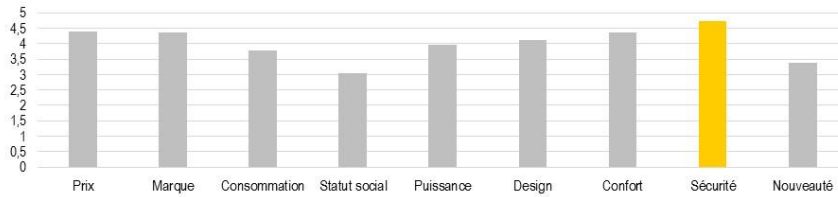
Oui : 82,9 %

Non : 17,1%

Participation au choix du modèle



Critères d'achat d'une voiture



5

ETUDE DU MARCHE AUTOMOBILE EN ALGERIE

II. ANALYSE DESCRIPTIVE
ANALYSE UNIVARIEE

Renouvellement du véhicule

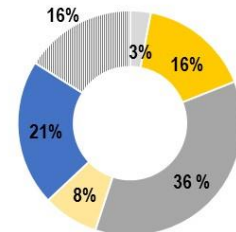
Oui : 21,6 %

Non : 51,6 %

Je ne sais pas : 26,8 %

Segment de renouvellement

- Petite citadine
- Citadine
- Compacte
- Ludospace
- Berline
- SUV/4*4



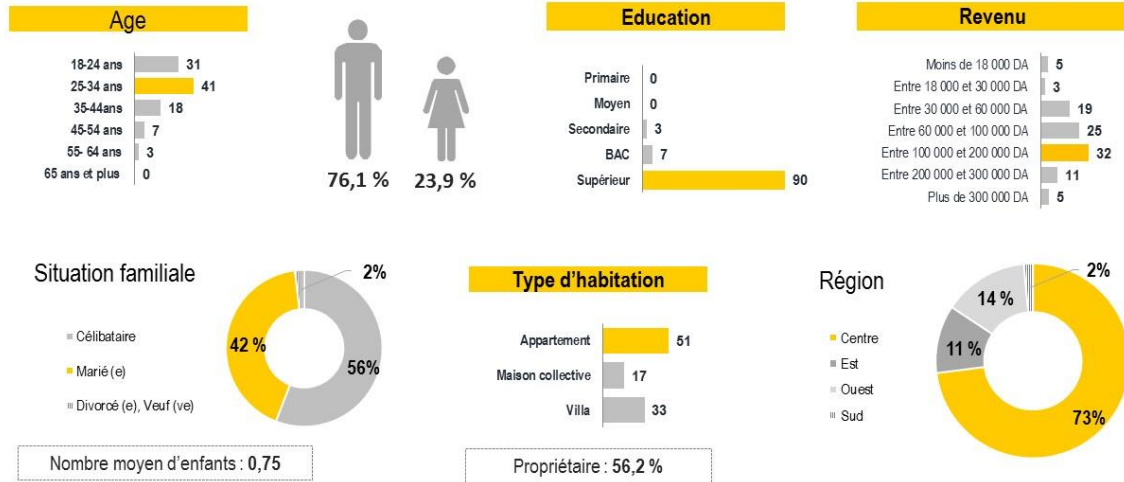
Montant de renouvellement	% de clients
Moins de 1 000 000 DA	1,2 %
Entre 1 000 000 et 1 300 000 DA	14,6 %
Entre 1 300 000 et 1 600 000 DA	20,7 %
Entre 1 600 000 et 2 000 000 DA	22 %
Entre 2 000 000 et 3 500 000 DA	22,2 %
Entre 3 500 000 et 4 000 000 DA	7,3 %
Entre 1 000 000 et 1 300 000 DA	3,7 %
Plus de 4 000 000 DA	8,3 %

5

ETUDE DU MARCHE AUTOMOBILE EN ALGERIE

II. ANALYSE DESCRIPTIVE

PROFIL SOCIODEMOGRAPHIQUE

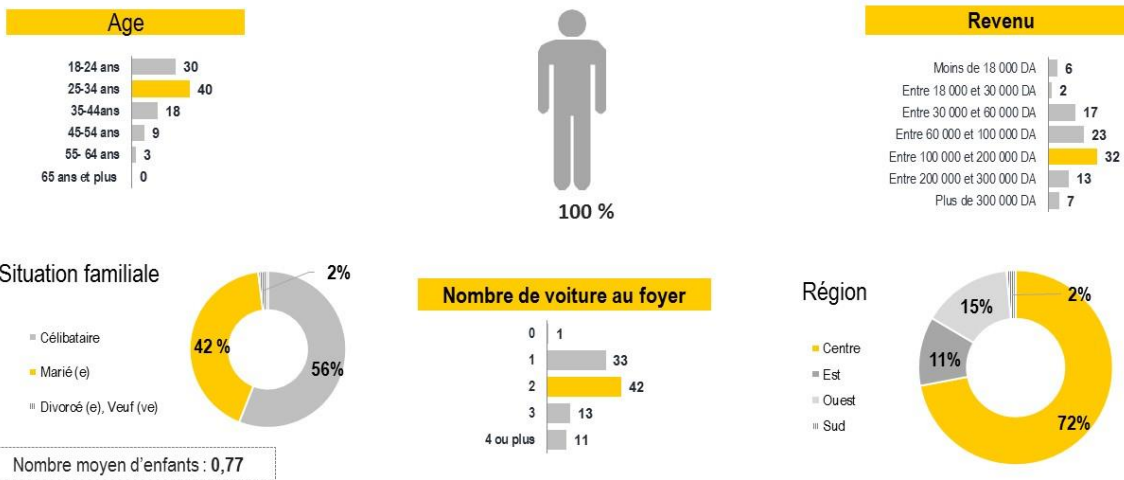


6

ETUDE DU MARCHÉ AUTOMOBILE EN ALGERIE

II. ANALYSE DESCRIPTIVE

PROFIL SOCIODEMOGRAPHIQUE_HOMME



7

ETUDE DU MARCHÉ AUTOMOBILE EN ALGERIE

II. ANALYSE DESCRIPTIVE

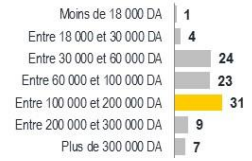
PROFIL SOCIODEMOGRAPHIQUE_FEMME

Age



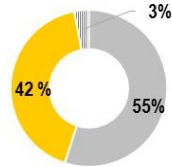
100 %

Revenu



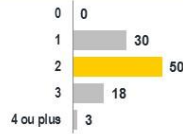
Situation familiale

- Célibataire
- Marié(e)
- Divorcé(e), Veuf(ve)



Nombre moyen d'enfants : 0,71

Nombre de voiture au foyer



Région

- Centre
- Est
- Ouest
- Sud

