



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



Département de génie industriel

Entreprise Djezzy

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

*Projet BI : de l'extraction à l'analyse de données
pour le ciblage des campagnes Marketing*

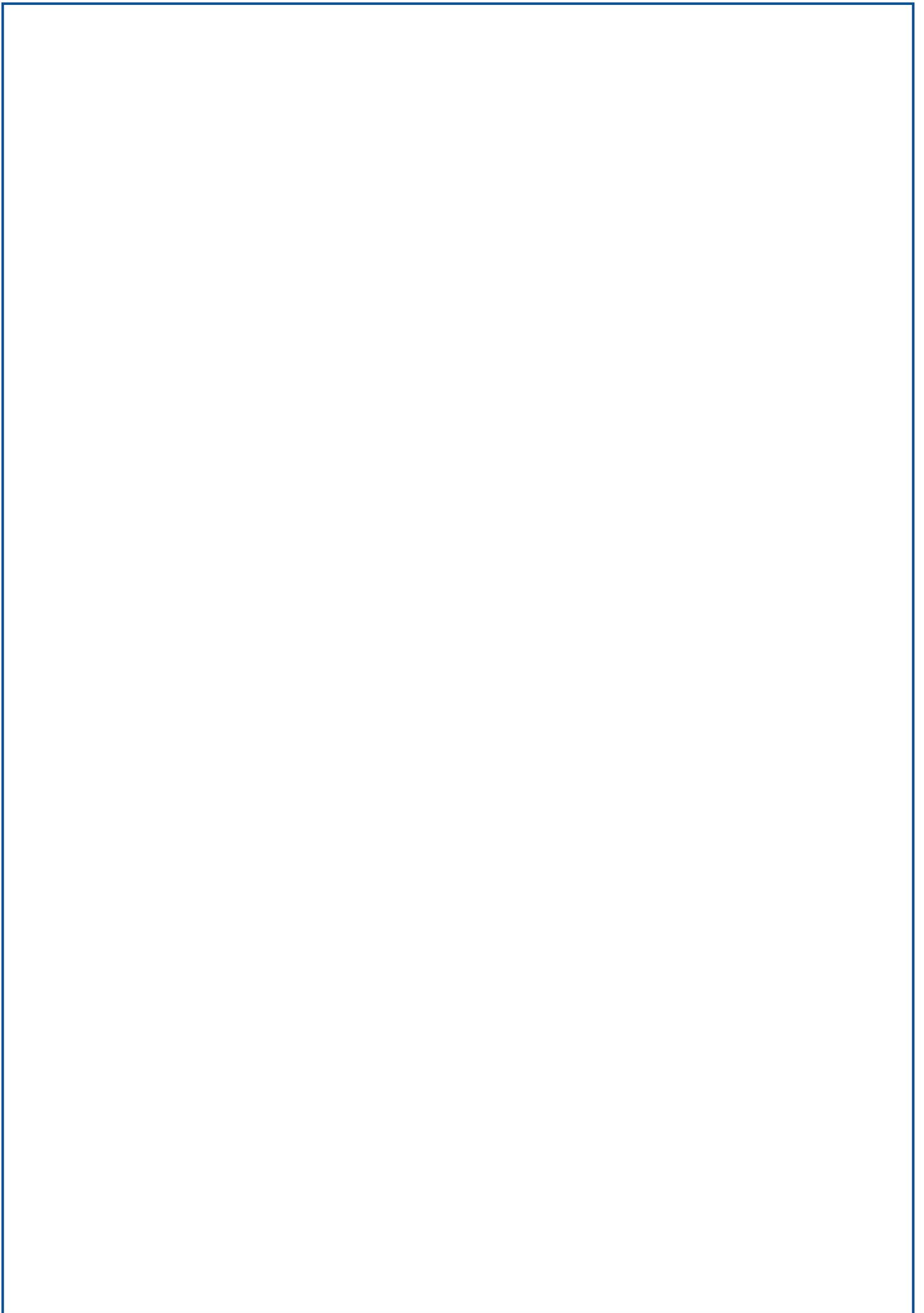
Elaboré par : Moussous Hanane
Rarrbo Amine

Sous la direction de : Mr Reda Gourine et Souhila Abriche

Présenté(e) et soutenue publiquement le 02/07/2019

Composition du Jury :

Président	Mme.Fatima Nibouche	MCA	ENP
Promoteur	Mr.Redha Gourine	Docteur	ENP
Examineur	Mme. Nadjwa Boukadoum	MAA	ENP





REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



Département de génie industriel
Entreprise Djazzy

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

*Projet BI : de l'extraction à l'analyse de données
pour le ciblage des campagnes Marketing*

Elaboré par : Moussous Hanane
Rarrbo Amine

Sous la direction de : Mr Reda Gourine et Souhila Abriche

Présenté(e) et soutenue publiquement le 02 /07 /2019

Composition du Jury :

Président	Mme.Fatima Nibouche	MCA	ENP
Promoteur	Mr.Redha Gourine	Docteur	ENP
Examineur	Mme. Nadjwa Boukadoum	MAA	ENP

ENP 2019

Dédicaces

Nous dédions ce travail à nos chers parents. Aucun hommage ne pourrait être à la hauteur du soutien Dont ils ne cessent de nous apporter.

Que dieu leur procure Bonne santé et longue vie, à nos frère et sœurs. Ainsi qu'à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible.

Remerciements

Nous remercions dieu le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire. Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu voir le jour sans l'aide et l'encadrement de Mr Selama Mohamed, un data scientist aux compétences multiples qui a partagé avec nous son savoir et son temps et pour cela nous lui serons à vie reconnaissants.

A Mr Gourine Reda pour sa rigueur de travail, sa disponibilité et ses précieux conseils durant la préparation de ce mémoire.

Nos remerciements s'adressent aussi à tous nos professeurs pour leur générosité et la grande patience dont ils ont su faire preuve tout au long de notre formation d'ingénieurs.

Nous remercions également les employés du service EIM et CVM de Djezzy qui ont toujours été disponibles pour répondre à nos interrogations, ainsi qu'aux personnes qui nous ont aidé et soutenue de près ou de loin à la réalisation de notre travail.

Résumé

ملخص

العديد من الشركات، خاصة شركات الاتصالات السلكية واللاسلكية، تجد نفسها الآن في حوزة كميات كبيرة من البيانات. هذه البيانات، بعد المعالجة والتحليل، تولد معلومات تسهل في اتخاذ القرارات على عدة مستويات. يهدف هذا المشروع، الذي يغطي جميع مراحل مشروع في ذكاء الأعمال، إلى أتمتة حساب مؤشرات الأداء الرئيسية التي تمكن إدارة قيمة في استهداف، من خلال حملات التسويق، مشتركين جديدي المدفوعة مسبقاً، وهذا من خلال تقديم أفضل ما يناسب سلوكياتهم. الإشكالية هي: ما هي البيانات، وكيفية استغلالها، وتنظيمها وتقديمها لأتمتة حساب مؤشرات الأداء الرئيسية هذه من أجل تحسين استهداف هذه الحملات التسويقية؟

لمعالجة هذه الإشكالية، تم اقتراح حل يتم البدء به من خلال تصميم مخزن بيانات من مستودع بيانات موجود والذي سيتم استغلال بياناته بسهولة. لحساب هذه المؤشرات تلقائياً، سيتم تجميعها وترتيبها حسب كل مشترك في جدول يسمى مجموعة البيانات التحليلية، وكل هذا يشكل الجزء الثاني من هذا الحل. أما الجزء الثالث، فهو مخصص لتنفيذ خوارزمية تصنيف (ك-معدلات) تجزأ المشتركين لتيسير استهدافهم. وأخيراً، كنقطة أخيرة يأتي تصميم لوحة القيادة التي تشمل هذه المؤشرات، مما يسمح تصور النتائج التي تم الحصول عليها في الأجزاء السابقة.

الكلمات الرئيسية

ذكاء الأعمال، مؤشرات الأداء الرئيسية، مستودع البيانات، مخزن البيانات، برمجة حساب المؤشرات آلياً، استخراج البيانات، تصنيف-معدلات، لوحة القيادة

ABSTRACT:

Numerous companies, especially telecommunications ones, nowadays find themselves in possession of large amounts of data, this data, after processing and analysis, generates relevant information on which to take decision at different levels. This project, covering all stages of a Business intelligence (BI) project, aims to automate the calculation of key performance indicators (KPIs) that enable Customer Value Management (CVM) of Djezzy to target, through marketing campaigns, its prepaid subscribers, and this according to their behaviors. The resulting problem is therefore: What data, how to exploit it, organize it and present it to automate the calculation of these KPIs in order to optimize the targeting of these marketing campaigns?

To address this problem, a solution has been proposed, it is initiated by the design of a Datamart (DM) from the existing data warehouse (DWH), to which data will be easily exploited, for the automated calculation of these indicators, which will be grouped and ordered per subscriber in a table called Analytical Dataset (ADS) monthly updated, constituting the second part of this solution. As a third part, the implementation of a

classification algorithm (K-means) segmenting subscribers to facilitate their targeting is involved. Finally, as a final point comes the design of a dashboard that includes these indicators, thus allowing the visualization of the results obtained in the previous parts.

Keywords: Business intelligence, key performance indicators, data warehouse, DM, indicator calculation automation, ADS, data mining, classification, K-means, dashboard.

Résumé :

Beaucoup d'entreprises notamment celles de télécommunication se retrouvent aujourd'hui en possession d'importantes quantités de données, ces données, après traitement et analyse génèrent des informations pertinentes sur lesquelles s'appuie la prise de décision à différents niveaux. Ce projet, abordant toutes les étapes d'un projet en Business intelligence (BI) a pour objectif d'automatiser le calcul des indicateurs clé de performance (KPI) du service « Customer Value Management » (CVM) de l'entreprise Djezzy, lui permettant de cibler, par des campagnes marketing les abonnés prépayés de celle-ci, et via l'offre la mieux adaptée à leurs comportements. La problématique qui en découle est la suivante : **Quelles sont les données, comment les exploiter, les organiser et les présenter pour automatiser le calcul de ces KPI afin d'optimiser le ciblage de ces campagnes marketing ?**

Pour répondre à cette problématique, une solution a été proposée, celle-ci est initiée par la conception d'un Datamart à partir d'entrepôt de données (EDD) existant de l'entreprise, dont les données seront facilement exploitées, pour le calcul automatisé de ces KPI, qui seront regroupés et ordonnés par abonné dans une table nommée Analytical Dataset (ADS) mise à jour mensuellement, constituant ainsi la seconde partie de cette solution. Comme troisième partie intervient l'implémentation d'un algorithme de classification (K-means) segmentant les abonnés afin de faciliter leur ciblage. Enfin, comme dernier point vient la conception d'un tableau de bord reprenant ces indicateurs, permettant ainsi la visualisation des résultats obtenus dans les précédentes parties.

Mots clés : Business intelligence, indicateurs clés de performance, entrepôt de données, magasin de données, automatisation du calcul des indicateurs, ADS, exploration de données, classification, K-moyennes, tableau de bord.

Tables des matières

Liste des tableaux

Liste des figures

Liste des sigles et abréviations

Introduction.....13

Première partie : Aspects théoriques et cadre de stage

Chapitre 01 : Généralités en BI et en data mining

1. BI et prise de décision.....	18
2. Les phases d'un projet BI.....	19
2.1. Phase d'alimentation.....	19
2.1.1. Extraction.....	20
2.1.2. Transformation	20
2.1.3. Chargement.....	20
2.2. Phase de modélisation (data warehousing).....	20
2.2.1. Définition d'un entrepôt de données	21
2.2.2. Définition d'un DM	22
2.2.3. Approches académiques de modélisation.....	23
2.2.3.1. L'approche d'Inmon (ou approche Top-Down)	23
2.2.3.2. L'approche de Kimball (ou approche Bottom-Up)	23
2.2.4. Modélisation dimensionnelle.....	24
2.2.4.1. Définition de la modélisation dimensionnelle	24
2.2.4.2. Table des faits	25
2.2.4.3. Table des dimensions.....	25
2.2.4.4. Étapes de la modélisation dimensionnelle.....	28
2.2.4.5. Modèles d'un entrepôt de données	28
2.3. Phase d'analyse.....	32
2.3.1. Data mining	32
2.3.1.1. Les techniques d'exploration de données	33
2.3.1.2. Data Mining dans la gestion de la relation client	34
2.3.1.3. Segmentation des clients et exploration des données	35

2.3.1.4.	Types de segmentation	35
2.3.1.5.	Méthodologie de la segmentation comportementale	36
2.3.1.6.	Définition du clustering	37
2.3.1.7.	Algorithme k-means	38
2.3.2.	Online Analytical Processing	41
2.3.3.	Reporting et tableau de bord.....	41
2.3.3.1.	Le Reporting	41
2.3.3.2.	Le tableau de bord	43

Chapitre 02 : Cadre de stage

1.	Présentation de l'entreprise Djazzy	44
1.	Présentation du service Enterprise Information Management (EIM)	45
2.	Présentation du service Customer Value Management (CVM).....	46
3.1.	Key Performance Indicators et campagnes marketing	46
3.1.1.	Campagne marketing au niveau du CVM	46
3.1.2.	Key Performance Indicator.....	47
3.1.3.	Analyse de l'existant	47
4.	Présentation synthétique du Projet	48
4.1.	Problématique et hypothèse.....	48
4.2.	Méthodologie de travail.....	50
4.3.	Contraintes surmontées.....	52

Deuxième partie : Solution Bi mise en place

Chapitre 01 : Organisation des données et calcul des KPI

1.	Conception du DM	55
1.1.	Réalisation technique du DM	58
1.1.1.	Création des dimensions	59
1.1.1.1.	Création de vues sur les tables.....	59
1.1.1.2.	Application d'outils ELT sur les tables	59
1.1.2.	Création des tables de faits	69
1.2.	Présentation du DM sur Business Object :	70
1.3.	Apports du DM :	71
2.	Conception de l'ADS :	72
2.1.	Du DM à l'ADS :	73

2.1.1.	KPI retenus	73
2.2.	Réalisation technique de l'ADS	76
2.2.1.	Calcul des KPI et chargement des tables	77
2.2.2.	Création et chargement des tables temporaires	77
2.2.3.	Création et chargement d'une table cible :	78
2.2.4.	Création et chargement de l'ADS.....	78
2.2.4.1.	Création	78
2.2.4.2.	Chargement.....	79
2.2.5.	Automatisation du calcul des KPI	79

Chapitre 02 : Segmentation et tableau de bord

1.	Segmentation comportementale des abonnés.....	82
1.1.	Langage et algorithme utilisés.....	82
1.2.	Données utilisées	83
1.3.	Implémentation de l'algorithme sur R.....	85
1.3.1.	Intégration des données	85
1.3.2.	Traitement des données	86
1.3.2.1.	Suppression de l'attribut « Subs_id ».....	87
1.3.2.2.	Détection et traitement des données Aberrantes (Outliers).....	87
1.3.2.3.	Suppression des variables qualitatives	87
1.3.2.4.	Normalisation des données.....	88
1.3.3.	Application de l'algorithme.....	88
1.3.3.1.	Détermination du nombre optimal de classes.....	88
1.3.3.2.	Lancement de l'algorithme.....	89
1.4.	Fouille et analyse des résultats	89
1.4.1.	Exploration des résultats	89
1.4.2.	Analyse des résultats	90
1.4.3.	Impact financier de la segmentation.....	100
2.	Création d'un tableau de bord	102
2.1.	Tableau desktop.....	102
2.2.	Réalisation du tableau de bord.....	102
	Conclusion.....	105
	Bibliographie.....	107
	Annexes.....	109

Liste des tableaux

Tableau 1: Comparaison entre les systèmes transactionnel et décisionnel	22
Tableau 2: Comparaison des approches Top Down et Bottom Up	24
Tableau 3: Comparaison entre les mesures d'une table de faits et les attributs d'une dimension (C.Desrosiers, 2011).....	27
Tableau 4 : Comparaison entre les modèles en étoile et en flocons.....	31
Tableau 5: Modèles d'exploration des données dans le CRM.....	34
Tableau 6 : Avantages et inconvénients de l'algorithme k-means	40
Tableau 7: Opérations de sélection, de rotation et de forage d'un hypercube	42
Tableau 8: Catégories et description des Familles des KPI retenus.....	73
Tableau 9: Liste et description des KPI Outgoing calculés à partir de la table « Fact_Traffic »	74
Tableau 10: Liste et description des KPI Incoming calculés à partir de la table « Fact_Traffic »	75
Tableau 11: Liste et description des KPI calculés à partir de la table « Fact_Refill ».....	76
Tableau 12: Liste des tables et des macros pour chaque catégorie de KPI.....	77
Tableau 13: Liste des attributs de l'ADS, leur type et leur description.....	79
Tableau 14: Liste et description des variables qualitatives retenues pour de la segmentation	85
Tableau 15: % de l'effectif et inertie intra-classe de chaque classe.....	90
Tableau 16: Coordonnées des centres de gravité des classes.....	91
Tableau 17: Les valeurs en % de la variable « Segment_value» pour chaque classe	94
Tableau 18: Les valeurs en % de la variable « SubscriberProfile » pour chaque classe.....	95
Tableau 19: Les valeurs en % de la variable « Tenure » pour chaque classe	95
Tableau 20: Les valeurs en % de la variable « Device_type » pour chaque classe.....	95
Tableau 21: Taux de réponse sans et avec segmentation par valeur d'abonné.....	100
Tableau 22: Gain en DZA lors d'un ciblage par segments.....	101

Liste des figures

Figure 1: Phase d'intégration des données dans un projet BI.....	19
Figure 2: Schéma de l'approche Top-Down.....	23
Figure 3: Schéma de l'approche Bottom-Up	24
Figure 4: Présentation d'une table des faits	25
Figure 5: Présentation d'une dimension	26
Figure 6 : Exemple illustrant la hiérarchie et le grain d'une dimension.....	27
Figure 7: Etapes de la modélisation dimensionnelle.....	28
Figure 8: Exemple d'un modèle en étoile (Inmon, 2005).....	29
Figure 9: Exemple d'un modèle en flocons (R & Ross M, 2008).....	30
Figure 10 : Exemple d'un modèle en Constellation.....	32
Figure 11: Processus du Knowledge Discovery in Databases	33
Figure 12: Arbre Hiérarchique	37
Figure 13: Représentation des Centres d'inerties intra et inter-classe dans un nuage de points	39
Figure 14: Elbow plot pour déterminer le nombre optimal de classes.....	40
Figure 15 : Opérations de sélection, de rotation et de forage d'un hypercube.....	41
Figure 16 : Organigramme de Djezzy	45
Figure 17: Organigramme du service Enterprise Information Management	45
Figure 18: Schéma de l'approche Top Down	56
Figure 19: Modèle physique et script de création de la dimension Dim_SubS	59
Figure 20:Modèle physique et script de création de la dimension « Dim_charge_type »	63
Figure 21: Script FastExport pour la dimension « Dim_Charge_Type »	64
Figure 22: Script FastLoad pour la dimension « Dim_Charge_Type »	64
Figure 23: Script du BTEQ 1 pour la dimension « Dim_Charge_Type ».....	65
Figure 24: Script du BTEQ 2 pour la dimension Dim_Charge_Type.....	65
Figure 25: Script du BTEQ 3 pour la dimension « Dim_Charge_Type ».....	66
Figure 26: Script d'exécution du fichier. Bat exécutant les scripts Fastexport, Fastload, et les scripts BTEQ pour la dimension « Dim_Charge_Type »	66
Figure 27: Script BTEQ 1 pour la dimension « Dim_Cell_Loc »	68
Figure 28: Script du BTEQ 2 pour la dimension « Dim_Cell_Loc »	68
Figure 29: Script d'exécution du fichier. Bat exécutant les scripts Fastexport, Fastload, et les scripts BTEQ pour la dimension « Dim_Cell_Loc »	69
Figure 30: Script de création de la table de faits « Fact_traffic »	69
Figure 31: Script de création de la table de faits « Fact_Refill ».....	70
Figure 32: Interface de BO accompagné du modèle du DM.....	71
Figure 33: Modèle physique du DM sur BO.....	71
Figure 34: Exemple de création d'un rapport sur BO (revenu journalier).....	72
Figure 35: Données contenu dans le DM	73
Figure 36: Script de calcul de KPI sur le revenu appels sortant pour le mois M1 (Mars 2019)	77
Figure 37: Script de Création de la table « Target_Base».....	78
Figure 38: Script de création de la Macro de chargement de la table « Target Base ».....	78

Figure 39: Script BTEQ exécutant les macros de chargement des tables des KPI et de l'ADS	80
Figure 40: Script du fichier. Bat exécutant le BTEQ de chargement des tables de KPI et de l'ADS	80
Figure 41: Script et temps d'exécution des macros de chargement.....	80
Figure 42: Chargement des données dans R dans une table nommée « dataset »	86
Figure 43: Suppression de l'attribut Subs_id.....	86
Figure 44: Visualisation des valeurs de la variable NBER_CALLS_Offnet_IN_M1 via la Boite à moustaches.....	87
Figure 45: Script de création de la fonction « remove_outliers » sous R	87
Figure 46: Script de suppression des variables qualitatives et mise à la même échelle des variables quantitative.....	88
Figure 47: Elbow plot pour déterminer le nombre optimal de classes.....	88
Figure 48: Script d'affichage du graphe « Elbow plot ».....	89
Figure 49: Script d'exécution de l'algorithme K-means	89
Figure 50: Script d'affectation des abonnés à une classe	92
Figure 51: Script de création d'un fichier CSV contenant les valeurs de la table dataset_final.	93
Figure 52: Script de création de la table « Clustering_Kmeans »sur Teradata	93
Figure 53: Script BTEQ pour la suppression des tables d'erreurs.....	93
Figure 54: Script Fastload pour la table « clustering_kmeans »	93
Figure 55: Script du fichier.Bat exécutant le BTEQ et le Fastload.....	94
Figure 56: Script de rajout de l'attribut Cluster_num à la table ADS_Kmeans	94
Figure 57: Histogrammes en % du trafic appels et données mobiles par classe.....	96
Figure 58: Histogrammes en % du trafic sms on net, off net et incoming par classe	96
Figure 59: Script de regroupement des classes 5 et 6	100
Figure 60: Tableau de bord créé sur tableau software.....	104

Liste des sigles et abréviations

ADS	<i>Analytical Dataset</i>
Bat	<i>Batch</i>
BI	<i>Business Intelligence</i>
BO	<i>Business Object</i>
BTEQ	<i>Batch Teradata Query</i>
CRM	<i>Customer Relationship Management</i>
CSV	<i>Comma-Separated-Values</i>
DM	<i>Data Mart</i>
DWH	<i>Data warehouse</i>
EDD	<i>Entrepôt de données</i>
EIM	<i>Enterprise Information Management</i>
ET	<i>Error Table</i>
ETL	<i>Extract Transform Load</i>
Gb	<i>Gigabit</i>
HV	<i>High Value</i>
IT	<i>Information Technology</i>
KDD	<i>Knowledge Discovery in Databases</i>
KPI	<i>Key Performance Indicator</i>
LV	<i>Low Value</i>
MV	<i>Medium Value</i>
ODBC	<i>Open Database Connectivity</i>
OLAP	<i>Online Analytical Processing</i>
SGBD	<i>Système de gestion de bases de données</i>
SQL	<i>Structured Query Language</i>
TPT	<i>Teradata Parallel Transporter</i>
UV	<i>Uniqueness violations</i>
VHV	<i>Very High Value</i>
VLV	<i>Very Low Value</i>
WT	<i>Work table</i>

Introduction

Notre stage, d'une durée de 4 mois s'est déroulé au sein de Djezzy Télécom, plus précisément au niveau des services « Customer Value Management » (CVM) en tant que client du projet, et « Enterprise Information Management » (EIM) pour la réalisation de celui-ci. Il couvre la Business intelligence (BI) et la data science, deux domaines en plein essor complémentaires l'un de l'autre.

Nombreuses entreprises, et particulièrement de télécommunication se retrouvent aujourd'hui submergées de données, dont l'analyse et la gestion tendent à leurs donner un sens, en une information sur laquelle, les managers se basent pour la prise de décision effective. C'est là qu'intervient-la BI, elle constitue un outil d'aide à la décision qui permet d'avoir une vue d'ensemble des activités de l'entreprise et de son environnement, elle permet la collecte, la consolidation, la modélisation et la restitution des données à l'entreprise pour fournir des informations et des analyses significatives aux employés, clients, fournisseurs, et partenaires de celle-ci.

Djezzy a à sa disposition un entrepôt de données (EDD) ou Data Warehouse en anglais (DWH) constituant la seule source de données exploitable, consultable et commune à tous les services de l'entreprise, il est implémenté dans leur système de gestion de base de données « Teradata », il contient un nombre important de bases de données, chacune contenant un ensemble conséquent de tables.

Le service « Customer Value Management » de Djezzy a pour mission de cibler, via des campagnes marketing les abonnés prépayés de l'entreprise avec l'offre la mieux adaptée à leur trafic réseau en termes d'appels, de sms et de consommation en données mobiles. Ce ciblage prend en compte les valeurs d'indicateurs clés ou « Key Performance Indicators » (KPI) en anglais de la période en cours. Cette opération de ciblage débute par la planification de la campagne relative à l'offre prépayée que l'on veut vendre, une fois celle-ci connue, on en déduit la liste des KPI qui y sont relatifs, comme par exemple le nombre d'abonnés déjà souscrits à cette offre ou encore le revenu généré par celle-ci. Comme ces KPI ne sont pas connus à l'avance mais plutôt après avoir pris connaissance de la campagne en question, ils ne pourront être calculés qu'une fois celle-ci connue. Pour effectuer ce calcul, le service CVM dispose d'analystes ayant des compétences en informatique et en programmation mais aussi une bonne compréhension des aspects marketing dans le secteur des télécoms, ils sont chargés de prendre connaissances de la liste des KPI de la campagne planifiée et de les calculer pour le lancement de celle-ci. Ce calcul se fait en exploitant les données d'un ensemble de table contenant des informations relatives aux abonnés, aux offres et au trafic réseau. Ceci est

réalisé sur le système de gestion de base de données de Djezzy « Teradata » via des requêtes SQL sur ces tables. Une moyenne de 30 KPI est calculée pour une campagne, dont parmi eux des indicateurs clé et d'autres qui ne le sont pas, cependant ils sont tous confondus et communément appelés KPI par abus de langage par l'équipe business CVM. Le temps de calcul de ces KPI varie selon leur nombre, leur complexité de calcul et la facilité d'accès et d'exploitation des tables de données qui lui sont nécessaires, en moyenne les analystes CVM l'effectuent en 4 jours.

Le calcul des KPI est retardé par deux facteurs, le premier est le retard pris dans l'accès et l'exploitation des données des tables du DWH, en effet comme l'entrepôt de données est la seule source de données et est donc commune à tous les services de l'entreprise, elle contient un nombre très important de tables, chacune se trouvant dans une des bases de données de celui-ci. Par conséquent lorsqu'on souhaite avoir accès à des données spécifique à un service tel que le trafic réseau des abonnés pour le CVM ou encore les données sur les ventes pour le service achat, on se doit de connaître les noms des tables, ceux des bases de données où elles s'y trouvent et veiller à ce que les tables en question ne sont pas en cours d'exploitation car dans le cas échéant elles seront impossible d'accès par l'utilisateur jusqu'à ce l'exploitation en cours ne se termine.

Le deuxième facteur réside dans le fait que les analystes CVM se retrouvent à calculer à chaque planification d'une campagne les KPI qui lui sont relatifs, ils se retrouvent à calculer parfois les mêmes KPI qui sont communs à plusieurs campagnes, ceci est aussi un facteur de retard et une sous-optimisation dans le processus de ciblage des campagnes marketing par le CVM.

Un retard engendré par les deux facteurs précédemment cités impacte le processus de ciblage des campagnes, en prenant du retard pour le lancement de celles-ci qui peuvent devenir parfois obsolètes et non d'actualité pour la période en cours, ce retard peut aussi provoquer le départ d'abonnés qui étaient susceptible de partir et qu'on pouvait fidéliser avec la bonne campagne au bon moment.

Après avoir pris connaissance des missions du service CVM, des sous-optimalité dans l'accès et la manipulation des données, des éventuels retards pris lors de l'exécution des campagnes dû aux deux facteurs précédemment cités, nous avons construit notre problématique soulevé par les questions suivantes : **Quelles sont les données, comment les exploiter, les organiser et les présenter pour optimiser le ciblage des campagnes marketing sur service CVM ?**

Pour répondre à cette problématique, nous avons proposé une solution constituée de 4 parties, la première consiste à concevoir un magasin de données ou datamart en anglais pour le service CVM pour tout d'abord, lui faciliter l'accès aux données, puis pour calculer les KPI sur lesquels il se base lors du ciblage des campagnes. La seconde partie consiste à automatiser

le calcul de ces KPI qui se fait actuellement manuellement, puis de les regrouper dans une unique table appelé « Analytical Dataset ». Ces deux parties précédemment citées répondent à elles seules à la problématique abordée plus haut cependant toujours pour optimiser le ciblage de ces campagnes, nous sommes allés plus loin, en exploitant ces KPI par des outils d'analyse en BI pour effectuer une segmentation comportementale des abonnés prépayés de Djezzy, pour en prendre connaissance des profils d'abonnés qui ressortent, et pour mettre en avant l'importance d'un ciblage par une campagne constituant ainsi la troisième partie de notre solution. Ensuite présenter ces indicateurs dans un tableau de bord dynamique consultable, pour le CVM comme dernière partie.

Cet écrit est rédigé comme suit : dans un premier temps, nous présenterons brièvement l'entreprise Djezzy, et les deux services CVM et EIM. Nous présenterons ensuite la méthodologie de travail suivie dans la formulation de notre problématique et de la solution proposée qui seront aussi abordées dans cette première partie du mémoire.

Dans un second temps, s'inscrivant dans le cadre théorique sur lequel nous nous sommes appuyés lors de la réalisation de notre projet, nous aborderons des généralités sur la Business intelligence, les méthodes et outils qu'elle utilise et qui ont été implémentés dans notre travail, tels que ceux de data mining, ou encore le tableau de bord et le reporting.

Enfin, nous détaillerons la solution proposée, initiée par la conception d'un Data Mart (DM) dont les données seront facilement exploitées pour le calcul automatisé d'indicateurs clés de performance (KPI), qui seront regroupés et ordonnés par abonné dans une table nommé Analytical Dataset (ADS) mensuellement mise à jour, constituant ainsi la seconde partie de notre projet.

Comme troisième partie intervient l'implémentation d'un algorithme de classification (K-means) segmentant les abonnés afin de faciliter leur ciblage. Enfin, comme dernier point de notre solution vient la conception d'un tableau de bord reprenant les principaux indicateurs clés, permettant ainsi la visualisation des résultats obtenus dans les précédentes parties.

Première partie :
Aspects théoriques et cadre
de stage

Chapitre 01 :

Généralités en BI et en data mining

Dans ce chapitre, nous présenterons les bases théoriques qui nous ont servi lors de la réalisation de notre projet BI, de l'étape de collecte, d'organisation et de stockage à celle de l'analyse des données. Les différentes méthodes et outils en BI et en data mining seront également abordés.

1. BI et prise de décision

Pour de nombreuses entreprises, les nouvelles technologies causent une surcharge d'informations, laissant les décideurs débordés d'informations inadéquates, incorrectes, incohérentes et trompeuses. En effet, les différentes méthodes de récupération et de traitement de ces informations souvent inutile peuvent mobiliser de nombreuses ressources. Pourtant, certaines entreprises parviennent à fournir à leurs décideurs des données brutes traitées et automatiquement consolidées, présentées dans un format compréhensible et facile à comprendre. Ces sociétés fournissent des informations pertinentes pour une prise de décision rapide et profonde. Qu'est-ce que ces entreprises ont que les autres n'ont pas? Des Fonctionnalités et processus de BI.

▪ BI

La BI ou informatique décisionnelle en français est un domaine de recherche qui se concentre sur les aspects théoriques et pratiques pour obtenir une base d'informations solide pour la prise de décision. L'objectif principal étant de fournir par le biais de diverses solutions, les bonnes connaissances aux bonnes personnes au moment opportun. Cela nécessite la bonne combinaison de systèmes Information Technology (IT), d'architectures, de structures de données, de processus de collecte de données.

Lors de l'évaluation des capacités de la BI, il y a quatre niveaux à prendre en compte:

- **Reporting** :Le Reporting est une fonctionnalité essentielle des outils de BI, l'objectif étant de créer des rapports récurrents et standard de manière efficace et conviviale. Les rapports sont prédéfinis et statiques par nature, générés soit à la demande d'un utilisateur final, soit actualisés périodiquement via un planificateur automatique.L'objectif étant la consolidation et à l'agrégation de données provenant de diverses sources selon une approche répétitive (automatisée, idéale) à partir de sources de données fiables (voir figure 35, page 58).
- **Les tableaux de bord** :Les tableaux de bord contiennent les données stratégiques de l'entreprise agrégées à un haut niveau, des présentations complètes comparables et des indicateurs de performance consolidés. Ils comprennent des rapports statiques et interactifs avec des données traduites en graphiques, diagrammes, jauges et illustrations pour simplifier la communication de sujets complexes (voir figure 61, page 88).

- **L'Analyse :** Au niveau de l'analyse, les systèmes BI fournissent non seulement des informations consolidées que les utilisateurs peuvent détailler et filtrer, mais également des prévisions et des analyses de tendance permettant de développer de nouvelles informations (basées sur les données brutes).
- **L'Analytique :** Au niveau supérieur d'un système de BI, des analyses de données intelligentes automatisées reposent sur une logique « floue » sophistiquée, basée sur des fonctions conviviales mais puissantes, le système de BI peut récupérer des informations significatives tout en masquant la complexité sous-jacente de l'interprétation des données. Les scénarios de simulation et les fonctionnalités de simulation fournissent une aide à la décision avancée et personnalisée. (Martin, 2011)

2. Les phases d'un projet BI

La BI est une solution informatique permettant de collecter, de traiter puis d'analyser les données de l'entreprise selon des critères spécifiques. Les résultats de cette analyse permettent aux dirigeants d'obtenir une vue globale sur leur activité, une meilleure compréhension du comportement client et une meilleure réactivité face au marché. La solution BI est constituée de 3 phases (Gomes, 2015) :

- La phase d'alimentation.
- La phase de modélisation.
- La phase d'analyse.

2.1. Phase d'alimentation

La phase d'alimentation ou intégration des données, utilise l'outil Extract Transform Load (ETL). Un ETL permet l'Extraction, la Transformation et le Chargement de données depuis des sources diverses (bases de données, fichiers) vers des cibles préalablement définies comme le montre le schéma ci-dessous.

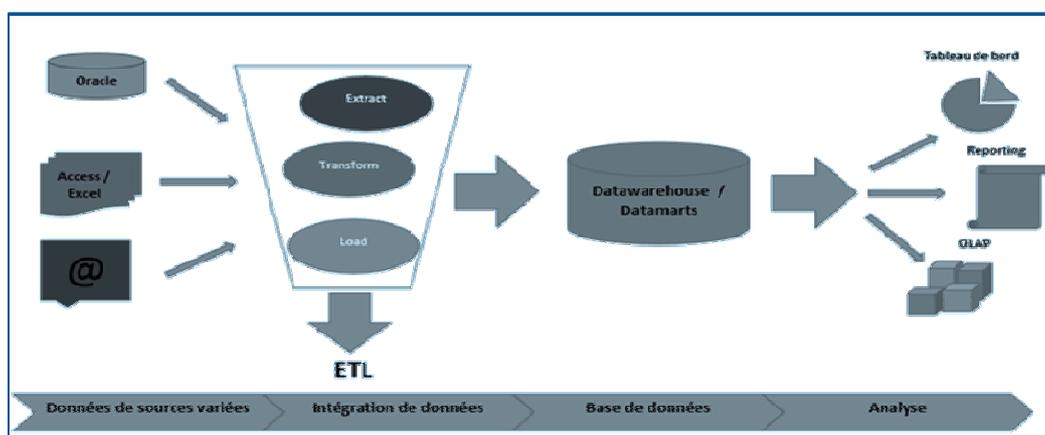


Figure 1: Phase d'intégration des données dans un projet BI

2.1.1. Extraction

La première partie du processus ETL consiste à extraire les données des systèmes sources. Dans de nombreux cas, c'est l'aspect le plus difficile de l'ETL, par ce qu'il faut établir des connexions vers des systèmes tiers ou exporter des volumes parfois importants de données sous forme de fichiers textes par exemple. Les formats de données communes sont la source de données relationnelles et des fichiers plats.

3 méthodes d'extraction sont utilisées :

Push : « Pousse » les données vers la staging area, si le système est occupé il ne poussera jamais les données.

Pull : « Tire » les données de la source vers la staging area, peut surcharger le système s'il est en cours d'utilisation.

Push-Pull : C'est le mélange des deux méthodes. La source prépare les données à envoyer et prévient la staging area qu'elle est prête. La staging va récupérer les données. Si la source est occupée, la staging fera une demande plus tard.

2.1.2. Transformation

L'étape de transformation applique une série de règles appelées règles de gestion. Pour certaines sources de données, il faudra une manipulation très légère – voire aucune – sur les données. Dans d'autres cas, un ou plusieurs des types de transformation suivants peuvent être nécessaires pour répondre aux besoins technico-commerciaux et de la base de données cible :

- Sélectionner uniquement certaines colonnes à charger.
- L'agrégation de certaines données.
- La suppression des données redondantes.
- Fractionnement d'une colonne en plusieurs colonnes...

2.1.3. Chargement

La phase de chargement qui est la dernière charge les données dans l'entrepôt de données (EDD).

2.2. Phase de modélisation (data warehousing)

Comme la modélisation relationnelle, la modélisation dimensionnelle est une technique de conception logique mieux adaptée aux applications de BI et à l'entreposage de données.

Le but de la modélisation dimensionnelle est de faciliter l'élaboration des rapports BI, de la requête et de l'analyse.

Les concepts clés de la modélisation dimensionnelle sont les faits, les dimensions et les attributs. Les dimensions peuvent avoir des hiérarchies différentes et avoir des attributs qui définissent le qui, quoi, où et pourquoi du modèle dimensionnel. Le grain, ou le niveau de granularité, est un autre concept clé avec la modélisation dimensionnelle, car il détermine le niveau de détail. Les faits, les dimensions et les attributs peuvent être organisés suivant deux schémas en étoile ou en flocon. Le choix du schéma dépend de certaines variables telles que le type de rapport que le modèle doit faciliter et le type d'outil de BI utilisé.

2.2.1. Définition d'un entrepôt de données

Collection de données thématiques, intégrées, non volatiles et historiées, organisées pour le support d'un processus d'aide à la décision (Inmon, 2005) :

- **Thématiques (subject-oriented)** :les données du DWH sont organisées, de telle sorte que tous les éléments se rapportant à un même événement ou objet réel sont liés entre eux.
 - **Intégrées (integrated)**:la base de données contient des données de la plupart ou de la totalité des applications de l'organisation.
 - **Non volatiles (non-volatile)** :les données du DWH ne sont jamais réécrites ou supprimées une fois engagées, les données sont statiques, en lecture seule et retenues pour les futurs rapports de synthèse.
 - **Historisées (time variant)** :les différentes modifications apportées aux données de la base de données sont suivies et enregistrées, des rapports synthétiques peuvent ainsi être élaborés.
- **Différence entre un DWH et un système transactionnel :**

Système transactionnel Données primitives/ données opérationnelles	Système décisionnel (DWH) Données dérivées
Orienté vers l'application	Orienté sujet
Détaillé	Résumé, autrement raffiné
Précis, au moment de l'accès	Représente les valeurs au fil du temps. Instantanés
Sert la communauté cléricale	Sert la communauté managériale
Peut être mis à jour	N'est pas mis à jour
Exécuter répétitivement	Exécution heuristique
Exigences de traitement comprises a priori	Exigences de traitement non comprises a priori
Sensible aux performances	Performance détendue
Accès à une unité à la fois	Accès à un ensemble à la fois
Contrôle de la mise à jour une préoccupation majeure en termes de propriété	Contrôle de mise à jour sans problème
Haute disponibilité	Disponibilité détendue

Non-redondance	La redondance est un fait de vie
Structure statique, contenu variable	Structure flexible
Transaction pilotée	Analyse pilotée
Petite quantité de données utilisées dans un processus	Grande quantité de données utilisées dans un processus
Prend en charge les opérations quotidiennes	Prend en charge les besoins managériaux
Forte probabilité d'accès	Faible, moyenne probabilité d'accès
Gérer dans son intégralité	Gérée par des sous-ensembles

Tableau 1: Comparaison entre les systèmes transactionnel et décisionnel

▪ Implémentation du DWH dans un SGBD :

Les deux problématiques fondamentales des DWH sont l'optimisation et la simplification :

Comment rester performant et lisible avec de très gros volumes de données et des requêtes portant sur de nombreuses tables (impliquant beaucoup de jointures) ?

On utilise massivement :

- Les vues concrètes ¹: Un DWH procède par copie depuis le ou les systèmes transactionnels.
- La dénormalisation ² : Un DWH est hautement redondant.

A noter que le caractère statique du DWH efface les inconvénients de ces techniques lorsqu'elles sont mobilisées dans des systèmes transactionnels(CROZAT, 2017).

2.2.2. Définition d'un DM

Un DM est un sous-ensemble du DWH de l'entreprise, obtenu par extraction et agrégation des données de celui-ci.

Les DWH et DM se distinguent par le spectre qu'ils couvrent :

- Le DWH recouvre l'ensemble des données et problématiques d'analyse visées par l'entreprise.
- Le DM recouvre une partie des données et problématiques liées à un métier ou un sujet d'analyse en particulier(CROZAT, 2017).

¹Une vue concrète est un stockage statique (dans une table) d'un résultat de requête. Elle permet de traiter le problème des requêtes dont les temps de calcul sont très longs et les fréquences de mise à jour faible.

²Processus consistant à regrouper plusieurs tables liées par des références, en une seule table, en réalisant statiquement les opérations de jointure adéquates. L'objectif de la dénormalisation est d'améliorer les performances de la BD en recherchant sur les tables considérées, en implémentant les jointures plutôt qu'en les calculant.

2.2.3. Approches académiques de modélisation

Lors de la conception d'un Data Warehouse, deux approches se confrontent : la méthode de **Bill Inmon** et celle de **Ralph Kimball**, chacune d'entre elles présentant ses avantages et inconvénients (Banguy, 2016)

2.2.3.1. L'approche d'Inmon (ou approche Top-Down)

William H. Inmon est un informaticien américain né en 1945, reconnu comme le père du Data Warehouse. L'approche d'Inmon, souvent opposée à celle de Ralph Kimball est caractérisée comme étant « Top-Down ».

Dans son approche, le Data Warehouse est un référentiel centralisé d'entreprise stockant l'information au niveau le plus détaillé. Des DM modélisés sous forme de schémas en étoile sont ensuite créés à partir de ce Data Warehouse.

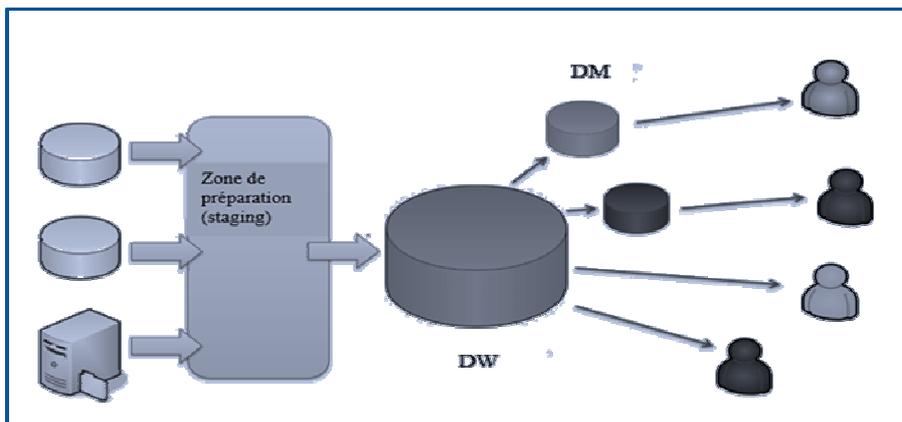


Figure 2: Schéma de l'ap proche Top-Down

2.2.3.2. L'approche de Kimball (ou approche Bottom-Up)

Ralph Kimball est un informaticien et chef d'entreprise américain. Il est connu pour ses ouvrages informatiques sur le thème du décisionnel.

Son approche s'oppose à celle de William H. Inmon, affirmant qu'**un DWH doit être rapide et compréhensible**.

Le DWH peut être vu, selon l'auteur, comme l'union des DMs cohérents entre eux grâce aux dimensions conformes ³(*data Warehouse bus*).

³ Voir infra, page 22

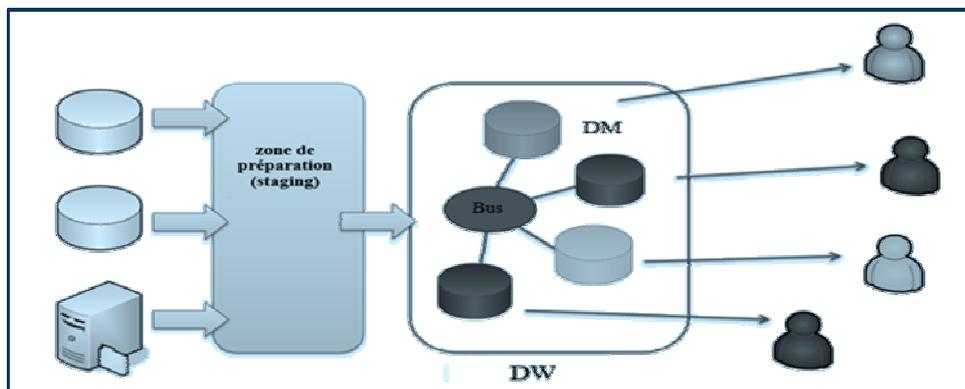


Figure 3: Schéma de l'approche Bottom-Up

▪ **Inmon (Top-down) contre Kimball (Bottom-Up)**

	Ralph Kimball	William H. (Bill) Inmon
Processus	Bottom-Up	Top-Down
Organisation	Conglomérats de DM	Datawarehouse
Schématisation	Etoile	Flocon
Construction	Rapide	Couteux en temps
Maintenance	Difficile redondance à gérer	Facile
Cout	Cout initial modéré	Cout initial élevé
Durée mise en œuvre	Court si mode incrémental	Long
Compétences utiles	Equipe généraliste	Equipe spécialisée
Intégration des données	Par domaine métier	Au niveau de l'entreprise

Tableau 2: Comparaison des approches Top Down et Bottom Up

Le choix de la meilleure approche dépend de l'activité de l'entreprise ainsi que de ses objectifs à long et à court terme sur le plan décisionnel, un mixte des deux est possible.

2.2.4. Modélisation dimensionnelle

2.2.4.1. Définition de la modélisation dimensionnelle

Technique de conception logique permettant de structurer les données de manière à les rendre intuitives aux utilisateurs d'affaires et offrir une bonne performance aux requêtes. Elle Permet de considérer un sujet d'analyse comme des points dans un espace à plusieurs dimensions. (C.Desrosiers, 2011)

Elle organise les données de manière à mettre en évidence le sujet (le fait) et les perspectives d'analyse (les dimensions).

2.2.4.2. Table des faits

Une table qui contient un ou plusieurs faits numériques qui se produisent lors de la combinaison de clés qui définissent chaque enregistrement.

Elle :

- Correspond normalement à un seul processus d'affaires.
- Stocke des mesures générées par les événements du processus.
- Les faits prennent leur valeur au moment où l'évènement d'affaires survient (aspect temporel important).
- Contient typiquement un très grand nombre de lignes.
- Comprend deux types de colonnes :
 - Clés étrangères vers des tables de dimension ;
 - Valeurs numériques qui sont les mesures.
- Modélise des relations de type plusieurs-à-plusieurs (voir illustration, page 44).

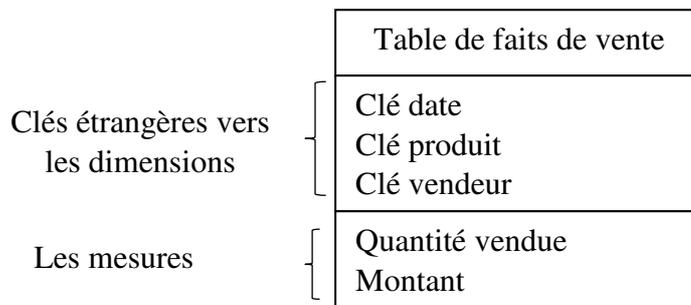


Figure 4: Présentation d'une table des faits

- **Les types de mesures**
 - **Additives** : Valeurs qui peuvent être agrégées par rapport à toutes les dimensions (chiffre d'affaires, quantités vendues, etc.)
 - **Semi-additives** : Valeurs qui peuvent être agrégées par rapport à un certain nombre de dimensions (solde de compte client, niveau de stocks, etc.)
 - **Non-additives** : Valeurs qui ne peuvent pas être agrégées (prix unitaire, ratios, etc.)

2.2.4.3. Table des dimensions

Une dimension est une entité qui établit le contexte commercial des mesures (faits) utilisées par une entreprise. Elle définit le qui, quoi, où et pourquoi du modèle dimensionnel, et groupe des attributs similaires dans une catégorie ou une zone de sujet.

Elle se décline en une table qui contient une clé primaire ⁴unique correspondant à l'un des composants de la clé multiple de la table des faits (voir illustration, page 44).

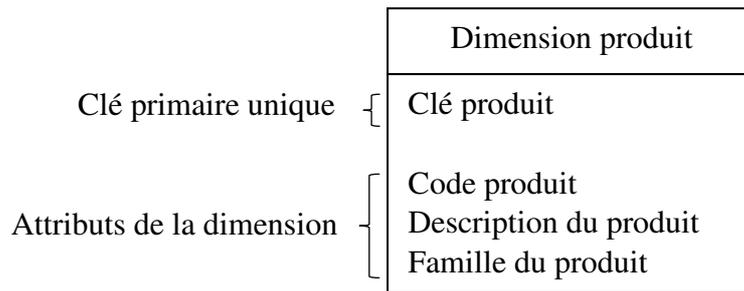


Figure 5: Présentation d'une dimension

▪ Types de dimensions (Zouaghi, 2018-2019)

- **Dimension à évolution lente:** Ses Attributs pourraient subir des changements dans le temps. Cela dépend de l'exigence métier si l'historique des modifications doit être conservé dans l'entrepôt de données.
- **Dimension à évolution rapide :** Ses attributs changent fréquemment et/ou rapidement dans le temps.
- **Dimension conforme:** Cette dimension est commune à plusieurs tables de faits dans une base de données unique ou dans plusieurs entrepôts de données ou magasins de données.
- **Dimension causale:** Cette dimension cause ou provoque des faits.
- **Dimension dégénérée ou factuelle :** Lorsque l'attribut de dimension est stocké dans le cadre d'une table de faits et non dans une table de dimensions distincte. Ce sont essentiellement des clés de dimension pour lesquelles il n'y a pas d'autres attributs.
- **Dimension débarras:** Les dimensions dépotoirs sont souvent créées pour gérer les clés étrangères pour éviter d'avoir un grand nombre de celles-ci dans la table de faits.
- **Dimension rétrécie :** Une dimension rétrécie est un sous-ensemble d'une autre dimension.
- **Dimension statique :** Les dimensions statiques ne sont pas extraites de la source de données d'origine, mais sont créées dans le contexte de l'entrepôt de données telle qu'une dimension date (C.Desrosiers, 2011).

▪ Mesures contre attributs de dimension

⁴ La clé primaire devrait toujours être une clé artificielle composée d'une seule colonne (surrogate key) car elle permet un accès accéléré par index et jointures ; elle ne change jamais contrairement à une clé naturelle ; elle permet d'identifier la même entité dans deux tables même si les tables ont des colonnes différentes.

Mesures	Attributs de dimensions
Dépendent d'un évènement d'affaires	Dépendent de la dimension et non pas des faits
Ont souvent des valeurs continues	Ont souvent des valeurs discrètes
Servent dans le calcul d'indicateurs de performance	Servent à filtrer ou étiqueter des faits

Tableau 3: Comparaison entre les mesures d'une table de faits et les attributs d'une dimension (C.Desrosiers, 2011)

▪ **Hiérarchie et grains des dimensions** (Sherman, 2015)

Un autre aspect du contexte commercial créé par les dimensions est qu'ils sont souvent hiérarchiques. Ces hiérarchies représentent des relations plusieurs-à-un. Parmi les exemples de hiérarchies, citons :

- Structures organisationnelles.
- Catégories de produits ou de services.
- Groupements géographiques.
- Temps. Les années se décompose en quarts, mois, semaines, heures, jours, minutes, en quelques secondes. C'est la hiérarchie classique.

Le « grain » d'une dimension est le niveau de sélection le plus fin possible de cette dimension. Le grain de la dimension Temps est le Mois, Le grain de la dimension Territoire est Région.

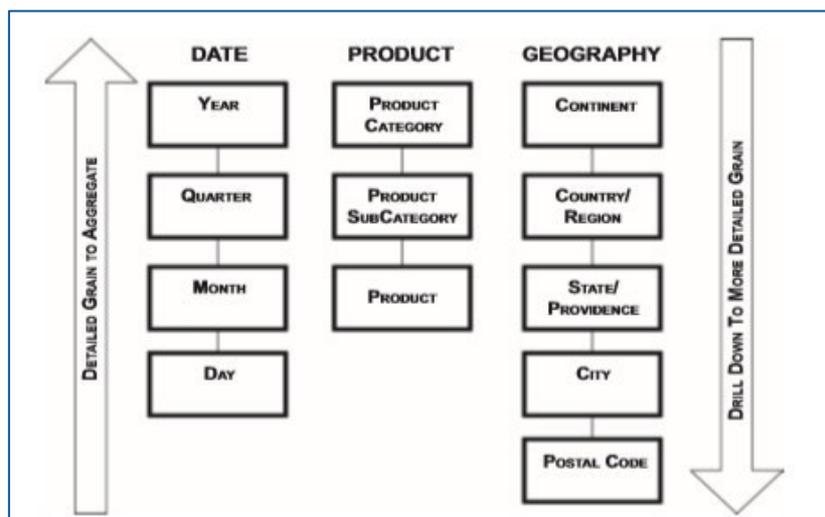


Figure 6 : Exemple illustrant la hiérarchie et le grain d'une dimension

Comme le montre la figure, les données sont agrégées à partir du niveau de données le plus détaillé (c.-à-d. le niveau le plus bas de la hiérarchie jusqu'au niveau le plus élevé). En règle générale, les tableaux de bord et les rapports affichent initialement des données

agrégées, et les gens d'affaires se présentent à des niveaux de détail plus élevés. Le forage vers le haut et vers le bas se déplace le long des hiérarchies pour examiner le niveau de détail nécessaire pour l'analyse de l'entreprise. Du point de vue de la modélisation des données, les hiérarchies sont des séries en cascade de relations plusieurs-à-un à différents niveaux. Chaque niveau correspond à un attribut de dimension.

2.2.4.4. Étapes de la modélisation dimensionnelle

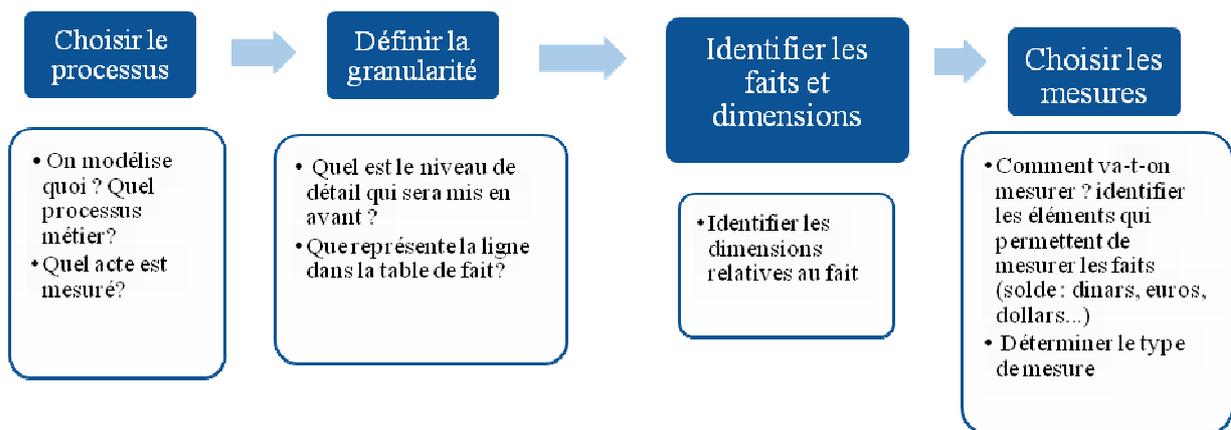


Figure 7: Etapes de la modélisation dimensionnelle

2.2.4.5. Modèles d'un entrepôt de données

Il existe trois types de schémas pour la création d'un modèle dimensionnel : Schéma en étoile, en de flocon et un schéma en constellation(Sfaxi, 2014)

Le choix de schéma à utiliser lors de la création du modèle dimensionnel s'effectue en examinant ces questions :

- Quel type d'analyse essayez-vous d'effectuer sur ces données et à quel point est-il complexe ?
- Quelles sont les exigences et les restrictions analytiques ?
- Dans quelle mesure les données que vous souhaitez interroger et analyser sont-elles cohérentes ?
- Quel outil de BI envisagez-vous d'utiliser ? Bien que différents outils puissent apparaître pour afficher le même type de données, les résultats et les graphiques, ils peuvent être très différents sous les couvertures, et s'appuient sur un schéma spécifique pour les meilleurs résultats.
- **Modèle étoile**

Le schéma en étoile est le schéma le plus courant pour les modèles dimensionnels. Il s'agit d'une table de faits entourée de plusieurs tables de dimension dénormalisées qui n'ont pas de lien entre elles.

Avantages :

- Facilité de navigation.
- Performance des requêtes.
- Gestion des agrégats.

Inconvénients :

- Redondance dans les dimensions.
- Alimentation complexe.

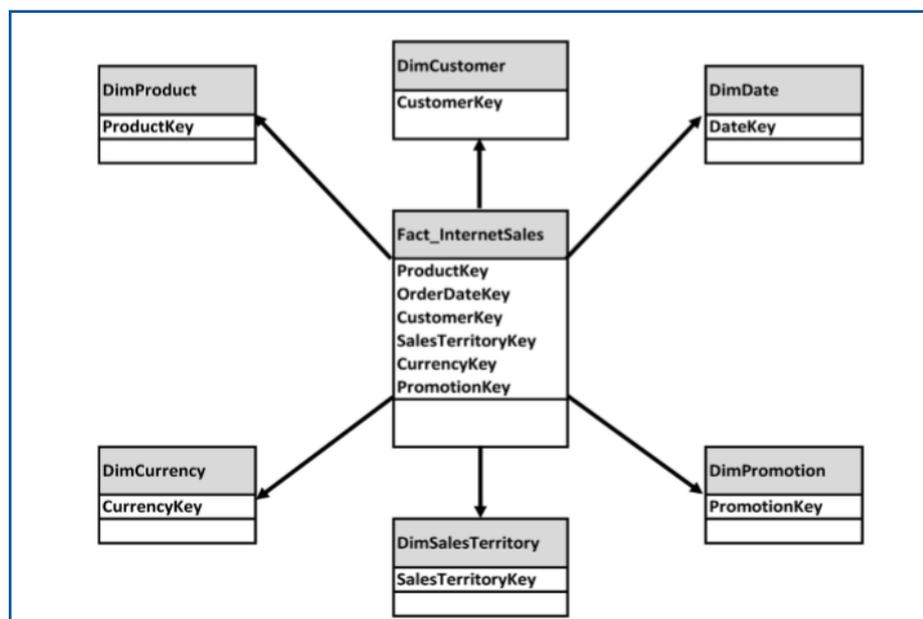


Figure 8: Exemple d'un modèle en étoile (Inmon, 2005)

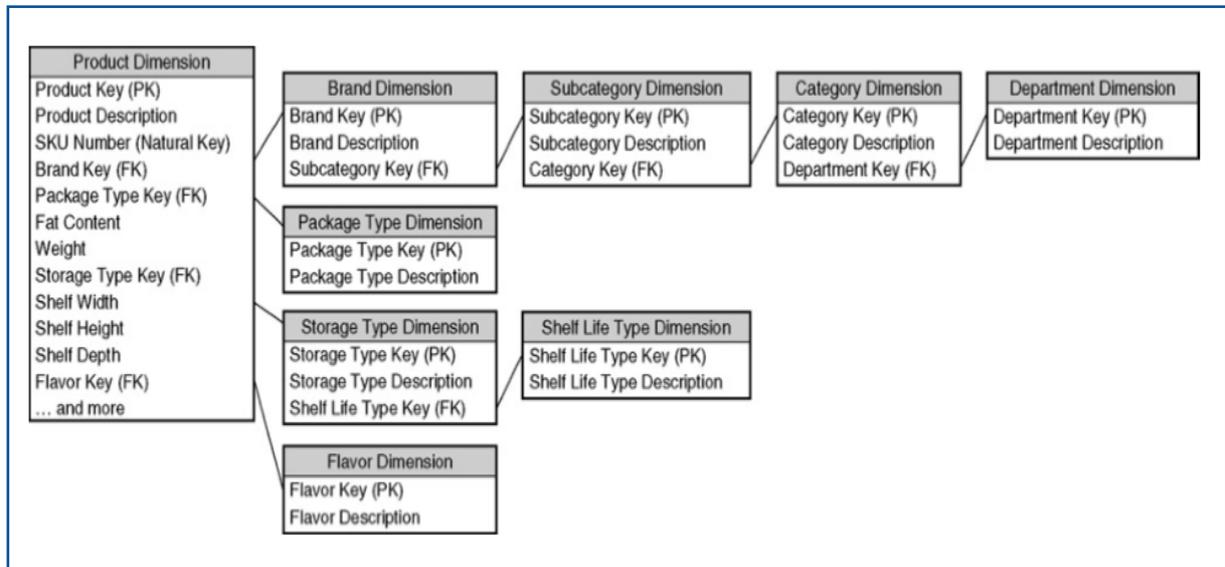
Le regroupement par attributs dans différentes tables de dimension permet à un homme d'affaires d'analyser qui achète quoi, où sont-ils situés, la date d'achat, et s'ils achètent en utilisant quelles promotions.

▪ Modèle en flocons

Le deuxième type de schéma de dimension est le flocon de neige. Il prend le schéma en étoile, avec les faits entourés par des dimensions dénormalisées, une étape plus loin en normalisant⁵ les hiérarchies dans une dimension particulière. Chaque niveau de la hiérarchie

⁵La normalisation est le processus qui permet d'optimiser un modèle logique afin de le rendre non redondant.

dimensionnelle devient sa propre table dimensionnelle avec les clés étrangères créées pour relier la structure hiérarchique. La table de faits stocke la clé étrangère au niveau le plus bas de la hiérarchie dimensionnelle (R & Ross M, 2008).



(Figure 9: Exemple d'un modèle en flocons (R & Ross M, 2008)

Avantages :

- Réduction du volume : gaine en espace de stockage.
- Permettre des analyses par palier sur la dimension hiérarchisée.
- Performances des mises à jour.

Inconvénients :

- Schéma moins intuitif aux utilisateurs d'affaires ;
- Dégradation de la performance à cause des jointures additionnelles.

Ce processus conduit à la fragmentation des données dans plusieurs tables.

Schéma en étoile vs schéma en flocon

	Schéma en étoile	Schéma en flocon
Structure du schéma	Contient des tables de faits et de dimensions.	Contient des tables de sous-dimension comprenant des tables de faits et de dimensions.
Modèle de données	Top-down	Bottom-up
Complexité de requête	Faible	Haute
Jointure utilisée	Moins	Grand nombre de jointure
Utilisation de l'espace	Plus	Moins
Temps consommé lors de l'exécution de la requête	Prend moins de temps	Prend plus en raison de l'utilisation excessive de la jointure.
Utilisation de la normalisation	N'utilise pas la normalisation.	Utilise la normalisation et la dénormalisation.

Tableau 4 : Comparaison entre les modèles en étoile et en flocons

(Différence entre le schéma en étoile et en flocons, 2018)

▪ **Modèle en constellation**

Ce 3^{ème} type de modèle fusionne plusieurs modèles en étoile qui utilisent des dimensions communes. Il comprend donc :

- Plusieurs tables de faits.
- Des tables de dimensions communes ou non à ces tables de faits.

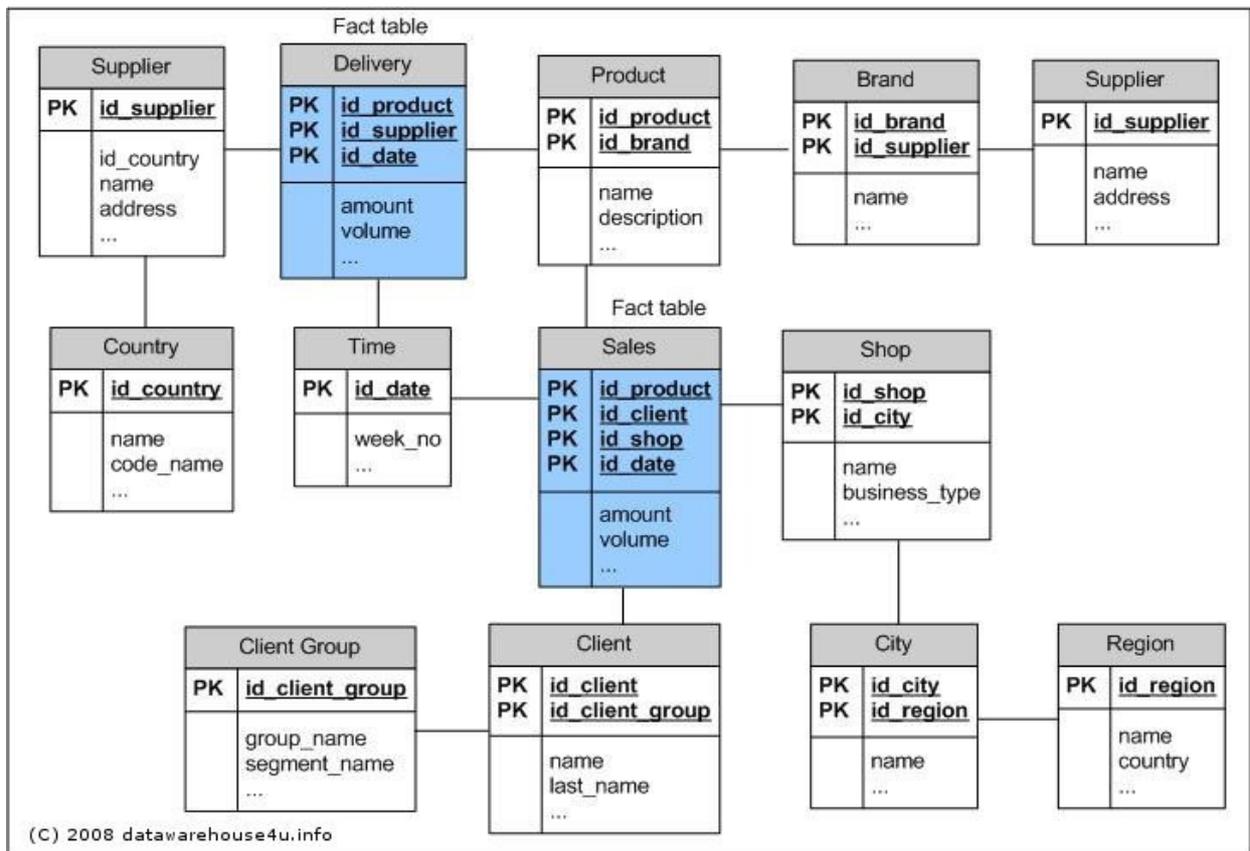


Figure 10 : Exemple d'un modèle en Constellation

(Brisson, 2016)

2.3. Phase d'analyse

La dernière étape de la chaîne décisionnelle est celle de l'exploitation des données. Les données nettoyées, consolidées, accessibles et stockées sont désormais prêtes à être analysées par les utilisateurs finaux ou les spécialistes de l'analyse. Pour ce faire, on utilise différents outils tels que les cubes Online Analytical Processing (OLAP) (pour les analyses multidimensionnelles), le Data Mining (pour chercher des corrélations), ou encore des tableaux de bord présentant les indicateurs clés.

2.3.1. Data mining

Le Data Mining ou fouille de données est un ensemble de méthodes qui consistent à extraire une connaissance, à partir d'une base de données décisionnelle de type DWH ou DM. Le Data Mining est très adapté à des données à grand volume pour la prise de décision. L'exploration de données consiste à résoudre les problèmes en analysant les données déjà présentes dans les bases de données. Il fournit des outils pour l'apprentissage automatisé à partir de données historiques et le développement de modèles pour prédire les tendances et les comportements futurs.

Le Data Mining une étape dans le processus Knowledge Discovery in Databases (KDD)⁶ qui consiste à appliquer l'analyse des données et les algorithmes de découverte qui produisent un recensement particulier de modèles sur les données. (Zouaghi, 2018-2019)

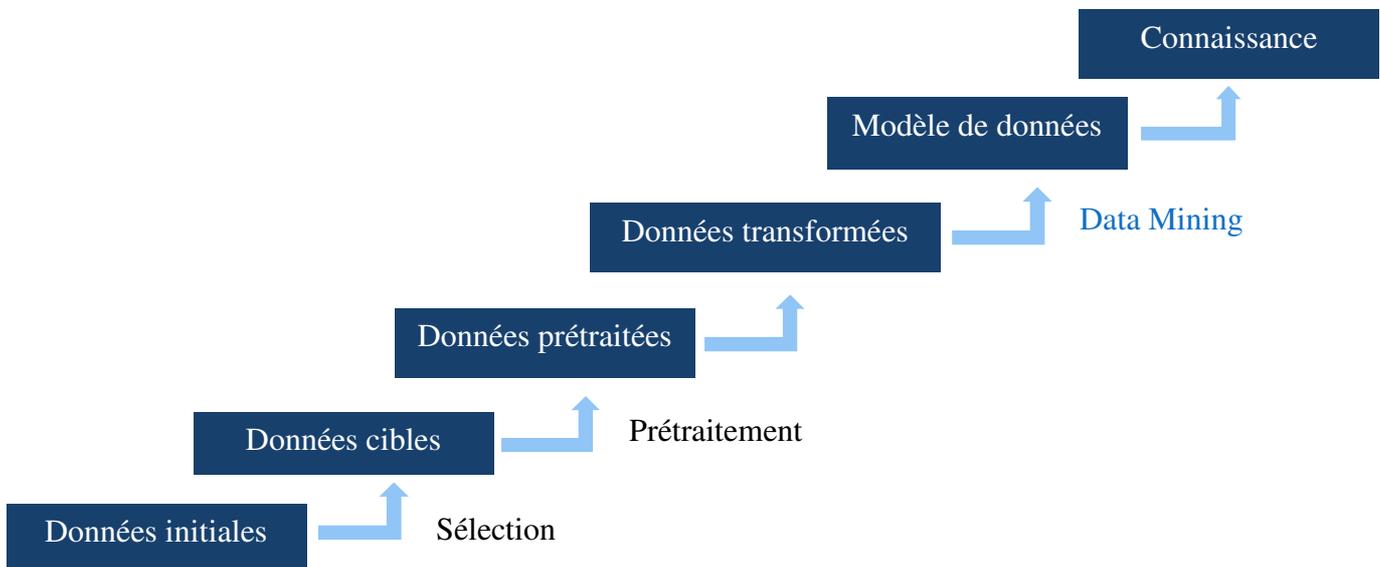


Figure 11: Processus du Knowledge Discovery in Databases

2.3.1.1. Les techniques d'exploration de données

Différentes techniques sont proposées. Elles sont à choisir en fonction de la nature des données et du type d'étude que l'on souhaite entreprendre :

- **Clustering** : L'analyse de clustering est une technique d'exploration de données permettant d'identifier les données qui se ressemblent.
- **Régression** : L'analyse de régression est la méthode d'exploration de données permettant d'identifier et d'analyser la relation entre les variables. Il est utilisé pour identifier la probabilité d'une variable spécifique, compte tenu de la présence d'autres variables.
- **Règles d'association** : Cette technique d'exploration de données permet de trouver une association et une corrélation entre deux éléments ou plus.
- **Détection externe** : Ce type de technique d'exploration de données fait référence à l'observation d'éléments de données dans la base de données qui ne correspondent pas à un modèle ou à un comportement attendu. Cette technique peut être utilisée dans divers domaines, tels que l'intrusion, la détection, la fraude ou la détection de pannes, etc. La détection externe est également appelée analyse des valeurs aberrantes ou extraction de l'outlier.

⁶ Knowledge Discovery in Databases (KDD) est un processus d'identification de modèles valides, nouveaux, potentiellement utiles et compréhensibles.

- **Schémas séquentiels** : Cette technique d'exploration de données permet de découvrir ou d'identifier des modèles ou des tendances similaires dans les données de transaction pour une période donnée.
- **Prédiction** : La prédiction a utilisé une combinaison des autres techniques d'exploration de données comme les tendances, les schémas séquentiels, le clustering, etc. cette méthode analyse les événements ou les instances passées dans une séquence appropriée pour prédire un événement futur(Alton, 2017).

2.3.1.2. Data Mining dans la gestion de la relation client

À des fins de marketing, l'exploration de données est une aide énorme. En utilisant la base de données de la gestion de la relation client CRM, les données démographiques, informations géographiques, psychographiques et comportementales des clients seront utiles pour les segmenter en utilisant des techniques d'exploration de données pour analyser les informations des clients.

L'idée principale de l'exploration de données pour CRM est que les données du passé peuvent contenir des informations qui seront utiles à l'avenir. Cela fonctionne parce que les comportements des clients capturés dans les données d'entreprise ne sont pas aléatoires, mais reflètent les différents besoins, préférences, propensions et traitements des clients.

L'exploration de données peut fournir un aperçu des clients, ce qui est vital pour une stratégie CRM efficace. Il peut aider à des interactions personnalisées avec les clients et donc une satisfaction accrue et des relations clients rentables grâce à l'analyse de données.

Les modèles d'exploration de données peuvent aider dans toutes ces tâches, comme illustré dans le tableau (Ziafat & Shakeri , 2014):

Tâche de modélisation	Techniques de modélisation	Exemple d'application
Association	A priorité, FP-croissance	Analyse du panier de marché
Clustering	K-means, Modèles hiérarchiques	Segmentation
Classification	Arbres de décision Réseaux de neurones Régression...	Prédiction

Tableau 5: Modèles d'exploration des données dans le CRM

2.3.1.3. Segmentation des clients et exploration des données

La segmentation du client est le processus de division des clients en sous-groupes distincts, significatifs et homogènes en fonction de divers attributs et caractéristiques. Il est utilisé comme un outil de marketing de différenciation. Il permet aux organisations de comprendre leurs clients et de créer des stratégies différenciées.

Il existe différents types de segmentation selon les critères de segmentation utilisés. En particulier, les clients peuvent être segmentés en fonction de leur valeur, de leurs informations socio-démographiques, ainsi que de leurs caractéristiques comportementales, de leurs besoins/attitudes.

2.3.1.4. Types de segmentation

Les types de segmentation suivants sont les plus couramment utilisés :

- **Basée sur la valeur** : dans la segmentation basée sur la valeur, les clients sont regroupés en fonction de leur valeur. Il s'agit de l'un des types de segmentation les plus importants puisqu'il peut être utilisé pour identifier les clients les plus précieux et pour suivre les changements de valeur au fil du temps. Il est également utilisé pour différencier les stratégies de prestation de services et pour optimiser l'allocation des ressources dans les initiatives de marketing.
- **Comportementale** : il s'agit d'un type de segmentation très efficace et utile. Il est également largement utilisé car il présente des difficultés minimales en termes de disponibilité des données. Les données requises incluent la propriété du produit et les données d'utilisation qui sont généralement stockées et disponibles dans les bases de données de l'organisation. Les clients sont divisés en fonction de leurs habitudes de comportement et d'utilisation identifiées. Ce type de segmentation est généralement utilisé pour développer des stratégies d'offre de produits personnalisées et pour le développement de nouveaux produits, et la conception de programmes de fidélisation.

Dans la segmentation comportementale, les segments sont identifiés avec l'application de modèles de clustering appropriés sur les données d'utilisation/comportementale qui résident habituellement dans l'entrepôt de données ou les DM de l'organisation. Ainsi, la segmentation comportementale peut être mise en œuvre avec un degré élevé de confiance et un coût relativement faible. Les attributs qui peuvent être utilisés pour la segmentation comportementale incluent la propriété et l'utilisation des produits, le volume/type/fréquence des transactions, l'historique des paiements et des revenus, etc.

- **Socio-démographique et étape de vie** : ce type révèle différents regroupements de clients basés sur des informations sociodémographiques et/ou de phase de vie telles que le sexe, la

race, l'âge, le statut social, l'éducation, l'occupation, le revenu, et l'état matrimonial. Ce type de segmentation est approprié pour promouvoir des produits spécifiques basés sur les cycles de vie.

- **Besoins/attitudes** : ce type de segmentation est généralement basé sur les données d'études de marché et identifie les segments de clients en fonction de leurs besoins, désirs, comportements, préférences et perceptions concernant les services et les produits de l'entreprise. Il peut être utilisé pour soutenir le développement de nouveaux produits et pour déterminer l'image de marque et les caractéristiques principales du produit à communiquer (Ziafat & Shakeri, 2014).

2.3.1.5. Méthodologie de la segmentation comportementale

La méthodologie de segmentation comportementale comprend les étapes principales suivantes:

- **Compréhension des affaires et conception du processus de segmentation** : Cette phase commence par la compréhension des exigences du projet du point de vue de l'entreprise. Il implique des réunions de partage des connaissances et une étroite collaboration entre les mineurs de données et les marketeurs impliqués dans le projet pour évaluer la situation, définir clairement l'objectif commercial spécifique et concevoir l'ensemble de la procédure d'exploration de données.
- **Compréhension, préparation et enrichissement des données** : L'enquête et l'évaluation des sources de données disponibles sont suivies par l'acquisition, l'intégration et le traitement des données pour les besoins de la modélisation de segmentation.
- **Identification des segments avec modélisation de clusters** : Les clients sont divisés en segments distincts à l'aide de l'analyse de clusters.
- **Évaluation des segments révélés** : Dans cette phase, les résultats de modélisation sont évalués et le schéma de segmentation qui répond le mieux aux besoins de l'organisation est sélectionné pour le déploiement.
- **Déploiement de la solution de segmentation, conception et livraison des stratégies différenciées** : Le projet de segmentation se conclut par le déploiement de la solution de segmentation et son utilisation dans le développement de stratégies marketing différenciées et de marketing segmenté (voir segmentation, chapitre 2 de la partie 2).

Une façon de trouver des segments comportementaux est d'utiliser les techniques de clustering tel que la méthode k-means et les méthodes de regroupement hiérarchique (Ziafat & Shakeri, 2014).

2.3.1.6. Définition du clustering

Le clustering est une méthode d'apprentissage non supervisé ⁷. Ainsi, on n'essaie pas d'apprendre une relation de corrélation entre un ensemble de donnée X d'une observation et une valeur Y à prédire, comme c'est le cas pour l'apprentissage supervisé⁸. L'apprentissage non supervisé va plutôt trouver des patterns dans les données. Notamment, en regroupant les choses qui se ressemblent (Benzaki, 2018).

En apprentissage non supervisé, les données sont représentées comme suit :

$$X = \begin{pmatrix} x_{(1,1)} & \dots & x_{(1,n)} \\ \vdots & \ddots & \vdots \\ x_{(m,1)} & \dots & x_{(m,n)} \end{pmatrix}$$

$x_{(m,n)}$: la valeur d'une observation de l'individu m pour la variable n

Chaque ligne représente un individu. Le clustering va regrouper en plusieurs familles (clusters) les individus/objets en fonction de leurs similitudes et différences.

Il existe deux types de clustering :

- **Le clustering hiérarchique** : Suites de partitions emboîtées

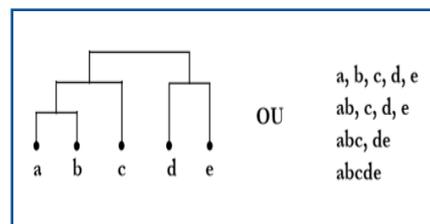


Figure 12: Arbre Hiérarchique

a,b,c,d,e : classes hiérarchiques

Avantage : La lecture de l'arbre permet de déterminer le nombre optimal de classes.

Inconvénient : Coûteux en temps de calcul.

- **Le clustering non-hiérarchique (partitionnement)** : Partition en k classes

Exemple : Centres mobiles et Nuées dynamiques.

Avantage : Permettent la classification d'ensembles volumineux.

⁷ L'apprentissage non supervisé est principalement utilisé en matière de classification, destiné à regrouper un ensemble d'éléments hétérogènes sous forme de sous-groupes homogènes ou liés par des caractéristiques communes. La machine fait alors elle-même les rapprochements en fonction de ces caractéristiques qu'elle est en mesure de repérer sans nécessiter d'intervention externe.

⁸ L'apprentissage supervisé est une technique d'apprentissage automatique où l'on cherche à produire automatiquement des règles à partir d'une base de données d'apprentissage contenant des « exemples » (en général des cas déjà traités et validés).

Inconvénient : On impose au départ le nombre de classes(GONZALEZ, 2016).

2.3.1.7. Algorithme k-means

L'algorithme K-means, tel que développé par Lloyd (1957), Forgy (1965), et MacQueen (1967) est un algorithme non supervisé de clustering non hiérarchique. Il permet de regrouper en clusters distincts les observations du Dataset. Ainsi les données similaires se retrouveront dans un même cluster. Par ailleurs, une observation ne peut se retrouver que dans un cluster à la fois. Une même observation, ne pourra donc, appartenir à deux clusters différents(Rakotomalala, 2014).

▪ Notion de similarité

Pour pouvoir regrouper un jeu de données en cluster distincts, l'algorithme K-means a besoin d'un moyen de comparer le degré de similarité entre les différentes observations. Ainsi, deux données qui se ressemblent, auront une distance de dissimilarité réduite, alors que deux objets différents auront une distance de séparation plus grande.

Les littératures mathématiques et statistiques regorgent de définitions de distance, les plus connues pour les cas de clustering sont :

- **La distance Euclidienne** : C'est la distance géométrique qu'on apprend au collège. Soit une matrice X à n variables quantitatives. Dans l'espace vectoriel E^n La distance euclidienne entre deux observations x_1 et x_2 se calcule comme suit :

$$d(x_1, x_2) = \sqrt{\sum_{j=1}^n (x_{1j} - x_{2j})^2}$$

- **La distance de Manhattan (taxi-distance)** : est la distance entre deux points parcourus par un taxi lorsqu'il se déplace dans une ville où les rues sont agencées selon un réseau ou un quadrillage. Un taxi-chemin est le trajet fait par un taxi lorsqu'il se déplace d'un nœud du réseau à un autre en utilisant les déplacements horizontaux et verticaux du réseau.

▪ Inertie intra-classe W et inter-classes B

Relation fondamentale (Théorème d'Huygens) :

$$\text{Inertie totale} = \text{inertie inter-classe} + \text{inertie intra-classe}$$

$$T = B + W$$

$$\sum_{i=1}^n d^2(i, G) = \sum_{k=1}^K n_k d^2(G_k, G) + \sum_{k=1}^K \sum_{i=1}^{n_k} d^2(i, G_k)$$

G : barycentre global

G_k : barycentre conditionnel de chaque classe k

n : nombre d'observations

n_k : nombre d'individus dans chaque classe k

$d()$: est une mesure de distance caractérisant les proximités entre les individus. Par ex distance euclidienne ou euclidienne pondérée par l'inverse de la variance.

- **Inertie inter-classe B** : Dispersion des barycentres conditionnels autour du barycentre global. Indicateur de séparabilité des classes.
- **Inertie intra-classe W** : Dispersion à l'intérieur de chaque groupe. Indicateur de compacité des classes.

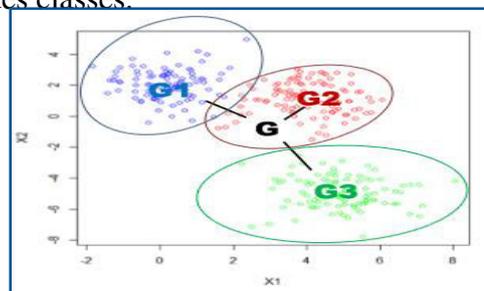


Figure 13: Représentation des Centres d'inerties intra et inter-classe dans un nuage de points

L'objectif de la classification k-means serait de minimiser l'inertie intra-classes W , à nombre de classes K fixé.

Entrée : X (n obs., p variables), K #classes

Initialiser K centres de classes G_k

REPETER

Allocation. Affecter chaque individu à la classe dont le centre est le plus proche ¹

Représentation. Recalculer les centres de classes à partir des individus rattachés ²

JUSQU'À Convergence ³

Sortie : Une partition des individus caractérisée par les K centres de classes G_k

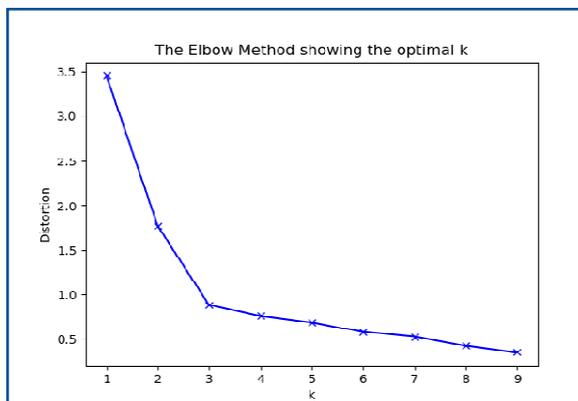


Le processus minimise implicitement l'inertie intra-classe W

²**Propriété fondamentale** : l'inertie intra-classe diminue à chaque étape

- **Détection du nombre de classes optimale**

Une stratégie simple pour identifier le nombre de classes consiste à faire varier K et surveiller l'évolution de l'inertie intra-classe W . L'idée est de visualiser le « coude » où l'adjonction d'une classe ne correspond à rien dans la structuration des données. (Rakotomalala, 2014)



La partition en $K = 3$ classes est la dernière à induire un gain informationnel significatif (une réduction de l'inertie intra implique une augmentation de l'inertie inter)

Figure 14: Elbow plot pour déterminer le nombre optimal de classes

- **Avantages et inconvénients**

Avantages	Inconvénients
Scalabilité : Capacité à traiter les très grandes bases. Seuls les vecteurs des moyennes sont à conserver en mémoire centrale	Lent parce que nécessité de faire passer plusieurs fois les observations.
Complexité linéaire par rapport au nombre d'observations (pas de calcul des distances deux à deux des individus)	L'optimisation aboutit à un minimum local de l'inertie intra-classes W (optimum local ¹²)
	La solution dépend du choix initial des centres de classes. ¹³

Tableau 6 : Avantages et inconvénients de l'algorithme k-means

- **Cas d'utilisation de l'algorithme**

¹¹Nombre d'itérations fixé ou aucun individu ne change de classe ou encore lorsque W ne diminue plus ou lorsque les G_k sont stables

¹²Afin de pallier aux problèmes des optimums locaux, il suffit de lancer le K-means plusieurs fois sur le jeu de données (avec le même nombre de classes mais des initialisations des centres différentes) et voir la composition des classes qui se forment.

¹³Essayer plusieurs configurations de départ et choisir celle qui aboutit à une solution minimisant W

Les champs d'application de K-means sont nombreux, il est notamment utilisé en :

- La segmentation de la clientèle en fonction d'un certain critère (démographique, habitude d'achat, comportement etc.)
- Utilisation du clustering en Data Mining lors de l'exploration de données pour déceler des individus similaires.
- Clustering de documents (regroupement de documents en fonction de leurs contenus).

2.3.2. Online Analytical Processing

Le OLAP, ou Online Analytical Processing, est une **technologie de traitement informatique** (computer processing). Elle permet à un utilisateur de consulter et **d'extraire facilement les données pour les comparer de différentes façons**. C'est un outil inscrit dans analysis services d'aide à la décision bien pratique pour une entreprise. Les données OLAP sont stockées sur une **base multidimensionnelle**, aussi appelées Cubes OLAP, pour faciliter ce type d'analyses Il est ainsi possible d'accéder facilement et rapidement à l'information souhaitée en fonction de plusieurs dimensions(L, 2018). La navigation au sein d'un cube peut être faite en effectuant des opérations de :

- **Sélection** : Slice (tranche de cube), Dice (sous-cube).
- **Rotation** : Pivot (rotation de dimension ou de hiérarchie), Drill-Across (rotation de fait).
- **Forage** : Drill-down (forage vers le bas), Roll-up (forage vers le haut).

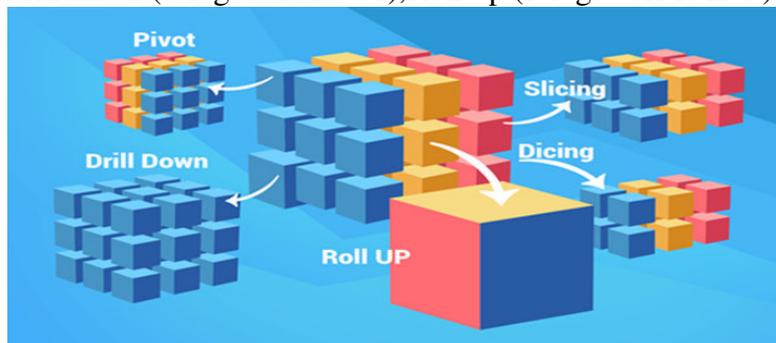


Figure 15 : Opérations de sélection, de rotation et de forage d'un hypercube

2.3.3. Reporting et tableau de bord

2.3.3.1. Le Reporting

Les outils de reporting permettent de restituer les données sous forme de rapport. Il existe deux grands domaines dans le reporting : le reporting ad hoc et le reporting de masse.

- Le **reporting ad hoc** offre la possibilité à l'utilisateur de créer lui-même le rapport qui l'intéresse avec les données qu'il souhaite. L'utilisateur aura accès à des vues métiers spécialement conçus en fonction de ses besoins qui lui permettront de choisir facilement l'information qu'il souhaite. Aucune connaissance en base de données n'est

nécessaire, les vues font la passerelle entre les données stockés et les besoins de l'utilisateur.

- Le **reporting de masse** quant à lui va permettre de créer à l'avance des modèles de rapport qui seront susceptible d'être souvent demandé par les utilisateurs. Le reporting de masse permet de répondre rapidement à un besoin régulier de beaucoup d'utilisateurs. (Gomes, 2015)

2.3.3.2. Le tableau de bord

En BI, le tableau de bord est utilisé dans les entreprises afin de permettre la visualisation de données brutes, les rendant ainsi plus accessibles et compréhensibles. Elle donne du sens à ces données. Pour cela, elle fait appel à différentes représentations visuelles et différents types de hiérarchisation de la donnée.

C'est un moyen efficace d'avoir une vue en temps réel ou différé des enjeux de son activité. L'agrégation de données clés permet de gagner en efficacité et de prendre de meilleures décisions (Tableau de bord (gestion), 2019).

On pourrait penser que le tableau de bord est un outil de Reporting, pourtant les deux présentent des points de différences résumés dans le tableau ci-dessous (Temgoua, 2005) :

Libellé	Tableau de bord	Reporting
Fond		
Destinataire	Service même	Hiérarchie
Indicateurs	- Indicateurs de résultats - Indicateurs de levier	- Indicateurs de résultat
Forme	Graphiques et tableaux	Graphiques et tableaux
Fonction	- Pilotage de la performance - Management de l'entreprise - Gestion de changement	- Evaluation de la performance d'un département

Tableau 7: Opérations de sélection, de rotation et de forage d'un hypercube

Chapitre 02 :

Cadre de stage

Dans ce premier chapitre, nous présenterons brièvement l'entreprise Djazzy, ainsi que les deux services CVM et EIM dans lesquels nous avons réalisé notre projet. Nous aborderons par la suite, le processus de ciblage des campagnes marketing sur lequel s'apparente notre problématique. Cette dernière, ainsi que notre hypothèse seront également abordées.

1. Présentation de l'entreprise Djezzy

Djezzy, un opérateur de télécommunications algérien a été créé en juillet **2001** et a lancé ses activités en février **2002**. Leader des technologies de communications numériques, l'entreprise fournit une vaste gamme de services tels que le prépayé¹ et le postpayé², internet ainsi que les services à valeur ajoutée et le Service Universel des Télécommunications (SUT).

En janvier **2015**, le Fonds National d'Investissement (FNI) prend le contrôle de **51%** du capital de la société alors que le partenaire étranger, le Groupe **Veon**, garde la responsabilité du management de l'entreprise. Djezzy couvre **95 %** de la population à travers le territoire national et ses services **3G** sont déployés dans les **48** wilayas.

Djezzy a déployé ses services **4G** dans **28** wilayas au **31** décembre **2017** avec une couverture de **25%** de la population s'est engagée à couvrir plus de **50%** de la population à l'horizon **2021**.

Djezzy est engagée dans un Processus de Transformation pour devenir l'opérateur numérique de référence en Algérie. L'entreprise est dirigée par Vincenzo Nesci Président Exécutif et Matthieu Galvani, Directeur Général.

Djezzy fait partie du groupe **Veon** (coté à la bourse de Nasdaq (New York) et à Euronext (Amsterdam), une entreprise de communication et de technologie internationale guidée par une vision construite sur des racines entrepreneuriales et dont les valeurs sont basées sur la satisfaction du client, l'innovation, le partenariat et la transparence. (Catalogue d'interconnexion OTA, 2015)

La structure interne de l'entreprise est présentée par l'organigramme suivant :

¹ Un **prépaiement mobile** (aussi appelé pay-as-you-go (PAY) ou prépayé) est une simcard achetée avant l'utilisation du service dont le crédit rechargé sert à payer les services de télécommunication au moment où le service est accessible ou consommé, exemple : DJEZZY HAYLA.

² **Postpaid** est un abonnement mobile pour lequel le service est fourni par accord préalable avec un opérateur de réseau mobile. L'utilisateur dans cette situation est facturé après coup en fonction de son utilisation des services mobiles à la fin de chaque mois, exemple : DJEZZY SMART.

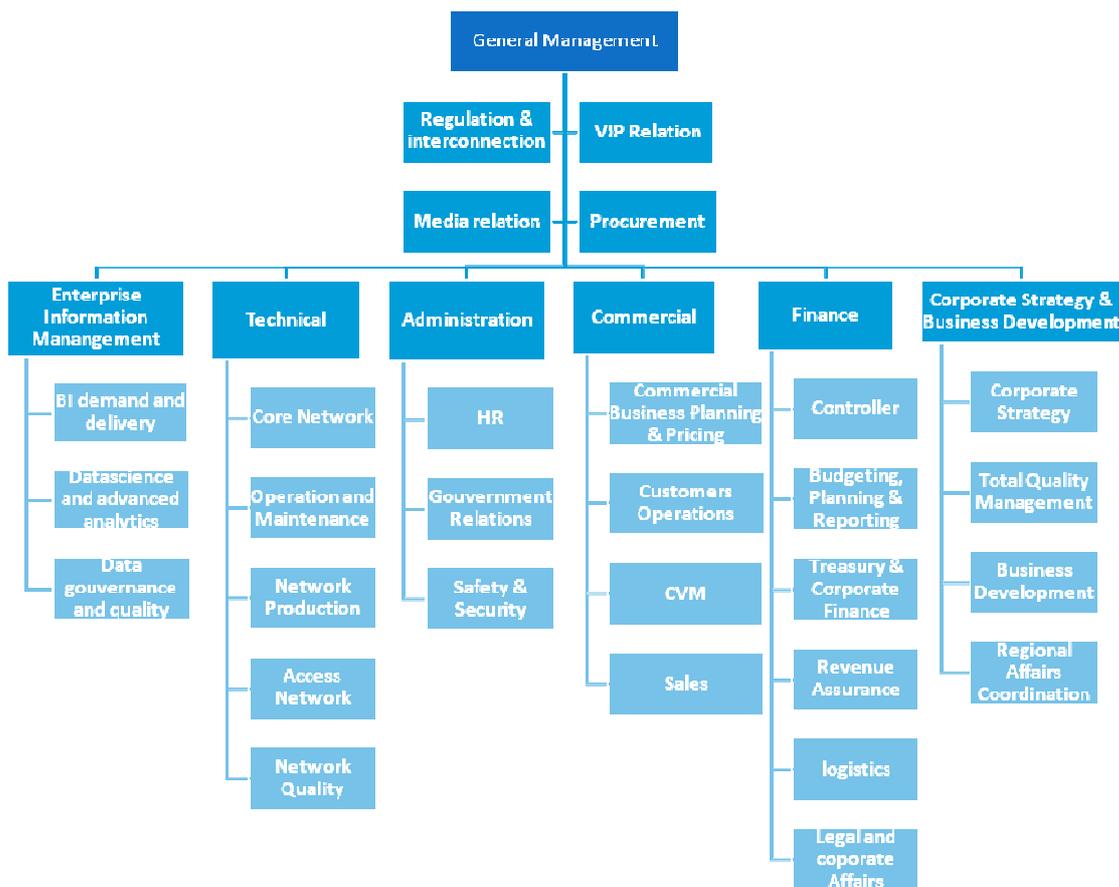


Figure 16 : Organigramme de Djazzy

1. Présentation du service Enterprise Information Management (EIM)

Le service « *Enterprise Information Management* » (EIM) est celui avec lequel nous avons interagis tout au long de notre projet, il regroupe plusieurs sous-services ou l'on effectue la collecte, le traitement, l'exploration, et l'analyse de données, il se subdivise comme suit :

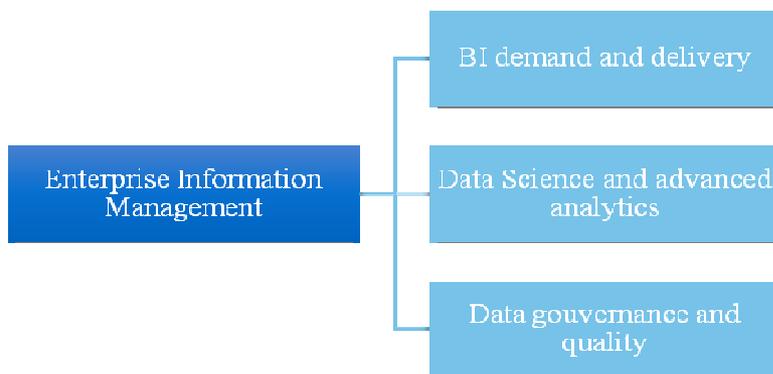


Figure 17: Organigramme du service Enterprise Information Management

Le service dispose de plusieurs ressources logicielles, nous citons :

- **Informatica** : Utilisée pour la collecte et l'intégration des données dans les entrepôts de données existants.
- **Teradata** : Principal Système de gestion de bases de données (SGBD) utilisé dans l'exploitation et le traitement des données, celui-ci a été préféré aux autres SGBD du marché pour sa puissance de traitement et sa facilité d'utilisation.
- **Langages R et Python** : Pour la création de programmes informatiques relatifs à l'exploration et à la visualisation des données.
- **SAP Business Object** : Pour la visualisation, l'analyse et la présentation des données sous formes de rapports et de tableaux de bords.

2. Présentation du service Customer Value Management (CVM)

Le « *Customer Value Management* » (CVM) est un service centré sur le client, qui, comme son l'indique vise à tirer le maximum de profit des abonnés de Djezzy et à en prospecter de nouveaux et cela, via des campagnes marketing.

Il se divise en trois sous-services :

- ➔ **CVM Business** : Ou l'on crée et lance les campagnes.
- ➔ **CVM Operations** : Ou l'on planifie les campagnes.
- ➔ **CVM Data Analytics** : Ou l'on analyse les données collectées sur les campagnes.

3.1. Key Performance Indicators et campagnes marketing

3.1.1. Campagne marketing au niveau du CVM

Les campagnes marketing au niveau du service CVM, sont journalières, hebdomadaires ou mensuelles, elles consistent à promouvoir à un abonné Djezzy l'offre la mieux adaptée aux dernières données récoltées sur son trafic appels, SMS et des données mobiles. Ces campagnes sont de deux types :

- **Indirectes (Publiques)** : La stratégie du CVM se concentre ici sur les canaux où il n'y a pas de communication directe avec les clients par l'entreprise, en les ciblant par exemple, par des panneaux publicitaires et flyers.
- **Directes** : Le service CVM utilise des campagnes de marketing direct pour communiquer un message à leurs clients par courrier, Internet, e-mail, téléphone et autres canaux directs. Les étapes des campagnes de marketing direct sont illustrées dans la figure 3 et expliquées ci-dessous :
 - 1- Recueillir et nettoyer les données nécessaires à partir de différentes sources.

- 2- Analyse et segmentation des clients (clustering) en différents groupes.
- 3- Développement des campagnes marketing ciblées afin de sélectionner les bons clients.
- 4- Exécution de la campagne en choisissant le canal approprié, l'heure appropriée et l'offre appropriée pour chaque campagne.
- 5- Évaluation de la campagne par l'utilisation de groupes de test et de contrôle. L'évaluation implique la division de la population en groupes de test et de contrôle et la comparaison des réponses positives.
- 6- Analyse des résultats de la campagne afin de l'améliorer pour le prochain tour en termes de ciblage, de temps, d'offre, de produit, de communication, etc.(Ziafat & Shakeri , 2014)

3.1.2. Key Performance Indicator

De l'anglais « *Key Performance Indicator* » (KPI) ou indicateur clé de performance, il est utilisé pour déterminer les facteurs pris en compte pour mesurer l'efficacité globale d'entre autres d'un dispositif commercial ou dans le cas échéant, d'un dispositif marketing(Marr, 2012).

3.1.3. Analyse de l'existant

Source de données

Le service CVM, exploite les données disponibles dans le Data warehouse (DWH) que possède l'entreprise. A cet effet, chaque utilisateur a accès aux tables du DWH qu'il manipule directement pour les besoins de son service.

Données et KPI utilisés

Données : comme le service CVM est orienté client, les données qu'il exploite le concerne directement, ce sont des informations :

- Générales : identifiant client, identifiant abonné, dernier appel émis...
- Sur le téléphone : identifiant mobile (TAC), type, système d'exploitation, compatibilité 4G...
- Type de souscription : prépayée ou post-payée.
- Profil global et type et revenu de l'offre.
- Le trafic en appels : appels entrant/sortant, intra/inter opérateur, national ou international.
- Le trafic en SMS : sms envoyé/reçu, intra/inter opérateur, national ou international.
- Le trafic en données mobiles : consommation en données mobiles.

KPI : Sur la base de ces données, les KPI suivants sont calculés :

- **KPI** sur le revenu (DZA) : revenu de chaque transaction, tout type confondu.
- **KPI** sur le nombre : nombre d'sms, nombre de souscription...
- **KPI** sur la durée : exclusifs aux appels.
- **KPI** sur les données mobiles (consommation en gigabit (Gb)).

Calcul des KPI ¹

- Le Calcul des KPI se fait suite à une demande émanant du CVM business.
- Le calcul se fait pour certains KPI seulement pour certains abonnés, visés par la campagne relative à cette demande.
- Le calcul se fait via des requêtes « *Structured Query Language* » (SQL) à partir des tables du DWH.

Utilisation des KPI

Les KPI calculés pour la campagne permettent :

- De connaître quel segment d'abonnés viser.
- De connaître la tendance du marché hebdomadaire et mensuelle.
- De connaître les abonnés partants et donc de ne pas les cibler par une quelconque campagne (KPI : Taux de Churn).

4. Présentation synthétique du Projet

4.1. Problématique et hypothèse

Le service « Customer Value Management » de Djezzy a pour mission de cibler, via des campagnes marketing les abonnés prépayés de l'entreprise avec l'offre la mieux adaptée à leur trafic réseau en termes d'appels, de sms et de consommation en données mobiles. Ce ciblage prend en compte les valeurs d'indicateurs clés ou « Key Performance Indicators » (KPI) en anglais de la période en cours. Cette opération de ciblage débute par la planification de la campagne relative à l'offre prépayée que l'on veut vendre, une fois celle-ci connue, on en déduit la liste des KPI qui y sont relatifs, comme par exemple le nombre d'abonnés déjà souscrits à cette offre ou encore le revenu généré par celle-ci. Comme ces KPI ne sont pas connus à l'avance mais plutôt après avoir pris connaissance de la campagne en question, ils ne pourront être calculés qu'une fois celle-ci connue. Pour effectuer ce calcul, le service CVM dispose d'analystes ayant des compétences en informatique et en programmation mais aussi une bonne compréhension des aspects marketing dans le secteur des télécoms, ils sont chargés de prendre connaissances de la liste des KPI de la campagne planifiée et de les calculer pour le lancement de celle-ci. Ce

¹Tous les KPI cités se déclinent en semaine, en mois, et parfois en jours

calcul se fait en exploitant les données d'un ensemble de table contenant des informations relatives aux abonnés, aux offres et au trafic réseau. Ceci est réalisé sur le système de gestion de base de données de Djezzy « Teradata » via des requêtes SQL sur ces tables. Une moyenne de 30 KPI est calculée pour une campagne, dont parmi eux des indicateurs clé et d'autres qui ne le sont pas, cependant ils sont tous confondus et communément appelés KPI par abus de langage par l'équipe business CVM. Le temps de calcul de ces KPI varie selon leur nombre, leur complexité de calcul et la facilité d'accès et d'exploitation des tables de données qui lui sont nécessaires, en moyenne les analystes CVM l'effectuent en 4 jours.

Le calcul des KPI est retardé par deux facteurs, le premier est le retard pris dans l'accès et l'exploitation des données des tables du DWH, en effet comme l'entrepôt de données est la seule source de données et est donc commune à tous les services de l'entreprise, elle contient un nombre très important de tables, chacune se trouvant dans une des bases de données de celui-ci. Par conséquent lorsqu'on souhaite avoir accès à des données spécifique à un service tel que le trafic réseau des abonnés pour le CVM ou encore les données sur les ventes pour le service achat, on se doit de connaitre les noms des tables, ceux des bases de données ou elles s'y trouvent et veiller à ce que les tables en question ne sont pas en cours d'exploitation car dans le cas échéant elles seront impossible d'accès par l'utilisateur jusqu'à ce l'exploitation en cours ne se termine.

Le deuxième facteur réside dans le fait que les analystes CVM se retrouvent à calculer à chaque planification d'une campagne les KPI qui lui sont relatifs, ils se retrouvent à calculer parfois les mêmes KPI qui sont communs à plusieurs campagnes, ceci est aussi un facteur de retard et une sous-optimisation dans le processus de ciblage des campagnes marketing par le CVM.

Un retard engendré par les deux facteurs précédemment cités impacte le processus de ciblage des campagnes, en prenant du retard pour le lancement de celles-ci qui peuvent devenir parfois obsolètes et non d'actualité pour la période en cours, ce retard peut aussi provoquer le départ d'abonnés qui étaient susceptible de partir et qu'on pouvait fidéliser avec la bonne campagne au bon moment.

Après avoir pris connaissance des missions du service CVM, des sous-optimalité dans l'accès et la manipulation des données, des éventuels retards pris lors de l'exécution des campagnes dû aux deux facteurs précédemment cités, nous avons construit notre problématique soulevé par les questions suivantes : **Quelles sont les données, comment les exploiter, les organiser et les présenter pour optimiser le ciblage des campagnes marketing sur service CVM ?**

Pour répondre à cette problématique, nous avons proposé une solution constituée de 4 parties.Elle consiste à :

1- Concevoir un DM (DM) pour le CVM, qui permet :

- L'accès et la manipulation des données par ce même service sans passer par l'entrepôt de données.
 - Le calcul des KPI retenus.
- 2- Automatiser le calcul de ces KPI qui seront regroupés dans une table de données ordonnée par abonné prépayé et mise à jour, appelé « *Analytical Dataset* » (ADS).
 - 3- Segmenter les abonnés prépayés en classes selon des critères prédéfinis (KPI de l'ADS), afin d'identifier les différences et les similitudes entre ces abonnés en termes de trafic appels, SMS et données, permettant à partir des résultats de cette segmentation de mieux les cibler par des campagnes directes (SMS).
 - 4- Concevoir un tableau de bord analytique pour la visualisation de la dynamique des KPI.

4.2. Méthodologie de travail

Pour mener à bien notre projet, nous avons suivi la méthodologie ci-dessous, celle-ci est inspirée de l'étape « définition du besoin » d'un avant-projet BI de l'auteur Ralph Kimball (Kimball & Ross, 2010).

Etape 1 : Découverte

Avant d'arriver à formuler notre problématique, nous avons tout d'abord demandé à lire le règlement intérieur de l'entreprise ainsi que l'organigramme de l'entreprise au niveau du service des ressources humaines, nous avons demandé qui sera notre responsable, dans quel département serons-nous affectés, quelle est la liberté de circulation entre les départements, quels employés doivent être informés de notre présence, quel sont les matériels, moyens et services mis à notre disposition, qui a accès et qui utilise l'entrepôt de données existant. Nous nous sommes ensuite informés sur les systèmes opérationnels de l'entreprise, et sur les datamart/entrepôts de données actuels, leur utilisation, documentation, contraintes rencontrées, résultats attendus contre ceux obtenus.

Ceci a permis la détection d'un besoin de disponibilité d'un DM pour le service « CVM » permettant de leur optimiser l'accès et le traitement des données en premier lieu, puis l'automatisation du calcul des KPI en second.

Etape 2 : Analyse du besoin

Tout d'abord, nous nous sommes renseignés sur la nécessité d'un DM dédié au service CVM en prenant connaissance du rôle de chaque personne du CVM utilisateur de l'éventuel DM (intitulé du poste, rôle principal, les attentes), nous avons pour ce faire sélectionné les personnes à interviewer, puis nous avons planifié des entretiens et préparé des questionnaires, ces questionnaires couvraient les interrogations sur le type de données que le

CVM manipule, qui sont de types : Identitaires, quantitatives, qualitatives et descriptives. Nous avons ensuite conduit des réunions de lancement à l'intention des utilisateurs du CVM, d'autres avec une équipe du service EIM pour l'identification des données existantes.

Ceci nous a emmené à déduire la conclusion suivante : l'inexistence d'un DM dédié au service CVM contraint celui-ci à s'appuyer sur un calcul manuel (Via des Requêtes SQL) des indicateurs de performances pour le ciblage des campagnes marketing et cela, à la demande (Selon les tendances du marché). Ceci engendre une faible réactivité par rapport aux changements journaliers dans la consommation de chaque abonné et donc les risques suivants sont présents :

- Risque de non fidélisation des clients à forte valeur ajoutée.
- Risque d'investissement (Ciblage par des campagnes) sur des clients à fort taux de Churn (clients bientôt perdus).
- Risque de retard dans le lancement des campagnes, à cause du temps perdu dans le calcul manuel.
- Risque de non prise en compte de tous les KPI dans le ciblage des campagnes, Par oubli ou par inexistence de certains jugés important mais indisponibles.

Suite à la proposition de notre hypothèse à l'équipe CVM, il nous a été demandé de respecter les points suivants :

Données : Ne retenir que les données relatives aux abonnés prépayés, représentant la majorité des clients de Djezzy (97 % du nombre total).

KPI : Retenir les KPI relatifs au revenu, durée d'appel, nombre d'appels et d'sms et consommation en données mobiles.

Offres : Intégrer toutes les offres prépayées actuelles de Djezzy.

Etape 3 : Réalisation technique

Après avoir cerné la problématique, nous avons entamé la réalisation technique de notre solution qui consiste tout d'abord à la conception d'un DM à partir du DWH existant sur le SGBD de l'entreprise, ensuite en exploitant les données de ce DM, nous avons calculé les KPI en les regroupant dans une table nommée ADS, puis nous avons automatisé ce calcul.

Comme seconde partie de notre solution, nous avons réalisé une segmentation comportementale par l'implémentation d'un algorithme de classification via le langage R puis nous avons créé un tableau de bord mettant en avant les principaux indicateurs sur le logiciel tableau desktop (voir partie 2).

4.3. Contraintes surmontées

Comme dans tout projet, nous avons fait face à de nombreuses contraintes :

- Impossibilité au début, d'avoir accès aux données de l'entreprise sans la création d'une session et d'un compte utilisateur.
- Nécessité de travail dans l'une des machines du service EIM à cause de l'absence d'un compte utilisateur sur les notre, et qui n'était pas toujours disponible.
- La nécessité de se former aux nombreuses ressources logicielles dont Teradata, BO, Langage R, tableau desktop.
- Contrainte temporelle lors de l'exécution des différents scripts d'export et de chargement des tables due à la grande quantité de données exportée et chargée.
- Apparition de messages d'erreurs, de bug informatique lors de l'exécution des scripts dont la résolution nécessitait l'intervention d'expert en informatique.
- Nécessité de relancer l'algorithme de classification plusieurs fois avant d'obtenir la bonne classification. Chaque lancement prenait en moyenne une demi-journée due à la grande quantité de donnée manipulée.

Deuxième partie :
Solution BI mise en
place

Chapitre 01 :

Organisation des données et calcul des KPI

Dans ce premier chapitre de la partie solution mise en place, nous présenterons le travail réalisé qui a pour but de répondre au besoin du service CVM, à savoir l'automatisation du calcul d'indicateurs clés pour le service CVM dans le ciblage des campagnes marketing. Ce travail a été découpé en deux parties :

1. La conception d'un Datamart décisionnel (DM) pour le service Customer Value Management CVM.
2. Le traitement puis l'intégration des données du DM dans un Dataset analytique (ADS) constitué des KPI relatifs à ce même service.

1. Conception du DM

La première partie de notre solution consiste à concevoir un DM pour le service CVM facilitant l'accès, l'organisation et la manipulation des données nécessaires pour le calcul des indicateurs utilisés par ce service ayant pour objectif l'aide à la prise de décision et cela sans passer par le DWH existant.

La conception de notre DM s'est effectuée en deux phases, la première concerne l'intégration des données c'est-à-dire la collecte et regroupement des données dans une seule base. La deuxième comprend l'organisation de ces données sous forme d'un DM.

Une fois la conception du DM achevée sur Teradata, nous avons créé un univers pour celui-ci sur le logiciel Business Object afin de permettre aux utilisateurs concernés d'interagir très facilement avec ses données.

Phase 1 : Intégration des données

Cette première phase concerne le processus ETL suivi lors de la conception de notre DM qui consiste en l'extraction, la transformation et le chargement des données (voir phase d'alimentation d'un projet BI, page 15).

Avant d'expliquer le travail réalisé dans cette phase et pour des fins de compréhension, nous présenterons brièvement la gestion des données du DWH représentant notre source de données.

▪ **Gestion des données du SGBD**

La gestion des données de Djezzy se fait sur le SGBD Teradata. Il constitue le système sur lequel nous avons conçu notre DM. Les données y sont stockées sur deux sources Teradata distinctes : **source 1** appelée « Production » et **source 2** appelée « Extension ». Notre DM a été conçu sur la source « **Extension** » qui est donc notre **source cible** car celle-ci n'est jamais interrompue par les jobs de chargement des données qui se font sur la « Production ». Par conséquent, « L'Extension », source de création de notre DM est à tout moment disponible pour consulter et manipuler les données dont parmi elles celles de notre DM.

A noter que chaque source possède un certain nombre de bases de données, les plus exploitées sur l'extension et la production portent respectivement l'appellation **P_EDW_TMP** et **P_EDW_BO_UNIVERSE**.

• **Extraction**

Les données ont été extraites de la source 1 « Production » comme suit :

- En utilisant l'outil d'extraction « FastExport » de Teradata sous forme de fichier plat à destination du même ou d'un autre SGBD.

- En utilisant l'outil Teradata Parallel Transporter (TPT) de Teradata qui combine l'extraction et le chargement de données sous condition que la cible soit Teradata.

- **Chargement**

Les données ont été chargées dans la source 2 « Extension » comme suit :

- En utilisant l'outil de chargement « FastLoad » de Teradata à partir des fichiers plats à destination de Teradata.
- En utilisant l'outil TPT puisqu'il permet également le chargement.

- **Transformation**

Comme les données sont déjà nettoyées et organisées en tables, nous n'avons pas eu besoin de réaliser une quelconque transformation.

Cependant, nous avons eu à effectuer des modifications sur le type et le contenu de certains attributs des tables de données exploitées.

Phase 2 : Organisation des données

L'objectif de cette phase est d'organiser et de préparer les données sous forme d'un DM facilitant par la suite le calcul des KPI.

Mais avant d'aborder cette phase détaillant les étapes de conception de notre DM, il est nécessaire de préciser l'approche suivie lors de la modélisation.

- **Approche de modélisation dimensionnelle**

Comme le DWH de Djezzy représente son référentiel centralisé stockant l'information au niveau le plus détaillé, notre DM modélisé sous forme de schéma en étoile est créé à partir de ce DWH donc l'approche suivie est une approche Top-Down (voir les approches de modélisation, page 19-20).

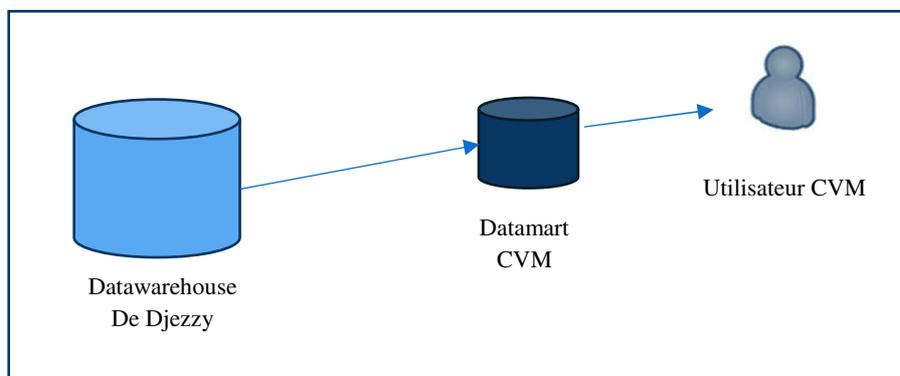


Figure 18: Schéma de l'approche Top Down

Nous allons maintenant détailler les 5 étapes de la modélisation dimensionnelle de notre DM : ((voir les étapes de modélisation d'un DM, page 24

Etape 1 : Définition du processus

Le processus modélisé est celui du « **Ciblage des campagnes marketing** » réalisé par le service CVM.

Ce processus exploite les données sur :

- **Les clients** : les abonnés prépayés.
- **Les produits** : les offres prépayées de Djezzy.
- **Les Ventes** : le trafic sms, appels et données mobiles.

Il a pour :

- **Client utilisateur**: le CVM Business, CVM operations, CVM analytics et EIM (service reporting).
- **Système source**: Teradata (DWH).

Etape 2 : Définition de la granularité

Les grains¹ associés au processus modélisé et qu'on retrouve dans les mesures des tables de faits de notre DM sont les suivants :

Grains :

- Par date : jours, semaines, mois, année
- Par offre
- Par zone géographique : wilaya, commune, région

Mesures :

- Revenu, durée de l'appel, nombre d'sms, consommation en données mobiles

Etape 3 et 4 : Choix des faits et des dimensions

Notre DM permet de répondre aux questions suivantes :

- Quel est le revenu par jour, par wilaya, par offre ?
- Quel est le nombre de clients par Profil ?
- Quelles sont les durées des appels émis et reçus ?
- Quel est le revenu par type de rechargement (flexy, carte) ?

Ces questions nous ont permis de déterminer les faits et les dimensions ci-dessous pour notre modélisation et leurs réponses consistent au calcul d'indicateurs qui y sont relatives.

¹ Voir définition, page 23

- **Fact_Traffic** : Table de faits relative au trafic réseau.
 - **Fact_Refill** : Table de faits relative au rechargement flexy et carte.
 - **Dim_Date** : Dimension relative à la date.
 - **Dim_Subs** : Dimension relative aux informations sur l'abonné prépayé.
 - **Dim_Offer** : Dimension relative aux offres prépayées de Djezzy.
 - **Dim_Cell_Loc** : Dimension relative à l'activité des antennes réseau Djezzy.
 - **Dim_Call_direction** : Dimension relative à la direction de l'appel.
 - **Dim_Charge_Type** : Dimension relative au type de transactions.
 - **Dim_Tariff_Group** : Dimension relative aux transactions par destination.
 - **Dim_Handset** : Dimension relative au téléphone associé à la SIMCARD Djezzy.
 - **Dim_Cell_Area** : Dimension relative à la localisation des antennes réseau Djezzy.
- Pour le détail des tables ci-dessus(Voir annexes 1 à 11)

Etape 5 : Choix des mesures

Les éléments nous permettant de calculer les KPI relatifs aux activités des abonnés prépayés en termes de trafic et de rechargement sont les mesures de nos deux tables de faits correspondant qui sont :

Mesures de la table Fact_Traffic :

- **Billed Amount**: revenu de l'appel, sms et données mobiles.
- **Billed Duration** :durée facturée de l'appel, sms et données mobiles.
- **Actual Duration** :durée totale de l'appel (facturée et non facturée).
- **Number Of Calls** :nombre d'appels et d'sms, et consommation en données mobiles.

Mesures de la table Fact_Refill :

- **Amount** :le revenu du rechargement.
- **Bonus Amount** : le bonus du rechargement.

A noter que toutes les mesures citées ci-dessus sont de type **additif** (voir types de mesure, page 21).

1.1. Réalisation technique du DM

Nous présenterons ici comment nous avons construit les dimensions et les tables de faits sur le SGBD Teradata à partir de requêtes SQL.

1.1.1. Création des dimensions

Nous avons créé les dimensions, soit en effectuant des vues¹ sur des tables existantes sur la source 2 « Extension », soit en exportant les tables de la source 1 « Production » via l'application d'outils Teradata qui y sont dédiés.

1.1.1.1. Création de vues sur les tables

Pour des raisons de sécurité des données du DWH, nous avons, lors de la conception de notre DM, effectué des vues sur les tables dont nous avons eu besoin.

Ci-dessous un exemple de la dimension subscriber créée à partir d'une vue sur la table « DBSS_DIM_SUBS » se trouvant dans la base « P_EDW_BO_UNIVERSE » de la source « Extension ».

DIM_SUBS	dimension subscriber
SUBS_KEY	REPLACE VIEW P_EDW_TMP.DIM_SUBS
CUST_ID	AS
MSISDN	SELECT
SUBS_TYP	SUBS_KEY ,
SUBS_STS	CUST_ID ,
SUBS_CUST_KEY	MSISDN
SUBS_OFFNG_KEY	SUBS_TYP ,
LST_CALL_OUT_DT	SUBS_STS
LST_CALL_IN_DT	SUBS_CUST_KEY
FRST_CALL_DT	SUBS_OFFNG_KEY
SUBS_ACT_DT	LST_CALL_OUT_DT
SUBS_FRST_ACT_DT	LST_CALL_IN_DT
	FRST_CALL_DT
	SUBS_ACT_DT
	SUBS_FRST_ACT_DT
	FROM P_EDW_BO_UNIVERSE.DBSS_DIM_SUBS ;

Figure 19: Modèle physique et script de création de la dimension Dim_Subs

Remarque :

- Les scripts SQL ainsi que le modèle physique des dimensions « Offer », « Date », « Handset » et « Cell_Area » ont été créés de la même façon que la dimension « Subscriber » (Voir annexes 12 à 15)
- La mise à jour journalière de ces dimensions se fait automatiquement à travers des workflows déclenchés par l'équipe ETL du service EIM.

1.1.1.2. Application d'outils ELT sur les tables

Comme il existe deux sources de données, et que leur manipulation n'est autorisée que sur la source « Extension », il a été nécessaire pour la création de certaines dimensions de recourir aux outils ELT de Teradata pour transférer certaines données de la source « Production » vers la source cible « Extension ».

Commençons par définir les outils ELT utilisés (Teradata Tutorial, 2019):

¹**Vue (View)** : Une vue dans une base de données est une synthèse d'une requête d'interrogation de la base. On peut la voir comme une table virtuelle, définie par une requête.

▪ FastExport

L'outil FastExport est utilisé pour exporter les données des tables Teradata dans des fichiers plats. Les données peuvent être extraites d'une ou de plusieurs tables à l'aide de jointure. Comme FastExport exporte les données sous forme de blocs de 64 Ko, il est utile d'extraire de gros volumes de données.

Voici la liste des termes couramment utilisés dans les scripts FastExport :

- **LOGTABLE** : spécifie la table LOG à des fins de redémarrage
- **LOGON** : se connecte à Teradata et lance une ou plusieurs sessions
- **DATABASE** : spécifie la base de données par défaut
- **BEGIN EXPORT** : indique le début de l'export
- **EXPORT** : indique le fichier cible et le format de l'export
- **SELECT** : spécifie la requête « select » pour exporter les données
- **END EXPORT** : spécifie la fin de FastExport
- **LOGOFF** : termine toutes les sessions ainsi que FastExport

Sa commande d'exécution a pour syntaxe : `fexp < nom_du_fichier.fx`

▪ FastLoad

L'outil FastLoad est utilisé pour charger des données dans des tables vides ayant pour type d'attributs une chaîne de caractères. Il ne charge pas les lignes en double, même si la table cible est une table MULTISSET. La table cible ne doit pas avoir d'index secondaire, d'index de jointure et de référence de clé étrangère.

Voici la liste des termes couramment utilisés dans les scripts FastLoad :

- **BEGIN LOADING** : identifie la table qui sera chargée
- **ERRORFILES** : identifie deux tables d'erreurs qui seront créées et mises à jour
- **CHECKPOINT** : définit quand faire un point de contrôle
- **SET RECORD** : Spécifie le format du fichier d'entrée
- **DEFINE** : Définit la disposition du fichier d'entrée
- **FILE** : Spécifie le nom et le chemin du fichier d'entrée
- **INSERT** : Insère les enregistrements du fichier d'entrée dans la table cible
- **END LOADING** : Distribue les enregistrements dans la table cible

Sa commande d'exécution a pour syntaxe : `FastLoad <nom_table.fl ;`

▪ BTEQ

L'outil Batch Teradata Query (BTEQ) est un outil puissant dans Teradata qui peut être utilisé pour créer des macros et des procédures stockées. BTEQ peut être utilisé pour :

- Importer des données dans des tables Teradata à partir d'un fichier plat.
- Extraire des données de tables dans des fichiers ou des rapports.

Voici la liste des termes couramment utilisés dans les scripts BTEQ :

- **ACTIVITYCOUNT** : Retourne le nombre de lignes affectées par la requête précédente.
- **ERRORCODE** : Retourne le code d'état de la requête précédente.
- **LABEL** : Attribue une étiquette à un ensemble de commandes SQL.
- **RUN FILE** : Exécute la requête contenue dans un fichier.
- **GOTO** : Transfère le contrôle sur une étiquette.

- **LOGOFF** : Se déconnecte de la base de données et met fin à toutes les sessions.
- **IMPORT** : Spécifie le chemin du fichier d'entrée.
- **EXPORT** : Spécifie le chemin du fichier de sortie et lance l'exportation.

- **Teradata Parallel Transporter (TPT)**

FastLoad, BTEQ et FastExport sont des outils de chargement. Ce sont des outils distincts offrant leur propre niveau de convivialité et de syntaxe. FastLoad est un outil de chargement en vrac. BTEQ est un outil de chargement transactionnel, insérant chaque ligne séparément et offrant une fonctionnalité de restauration.

Teradata propose un outil combinant les fonctionnalités de tous les outils. **Teradata Parallel Transporter (TPT)** fusionne les fonctionnalités de FastLoad, BTEQ et FastExport en un seul outil.

La structure du script TPT se présente comme suit :

- Définir le nom du job pour le script ;
- Fournir une description du job ;
- Définir le schéma de destination ;
- Définir les opérateurs d'extraction et de chargement ;
- Application du job ;
- Il requiert aussi un fichier pour le job.

Pourquoi le TPT

- TPT transporte efficacement de grandes quantités de données en utilisant les principes fondamentaux de parallélisme et d'évolutivité de l'entreposage de données.
- Contrairement aux outils classiques dans lesquels plusieurs sources de données sont généralement traitées en série, TPT peut accéder à plusieurs sources de données en parallèle.

Quand utiliser le TPT

- Lors de l'extraction de grandes quantités de données de bases de données externes non-Teradata.
- TPT peut être utilisé pour charger / décharger des données dans un environnement de base de données Teradata

La commande d'exécution du script TPT est la suivante : **tbuild -f <scriptname> -v <jobvariablename>**

Nous allons maintenant entamer le détail de création des dimensions sur lesquelles nous avons appliqué les outils cités ci-dessus :

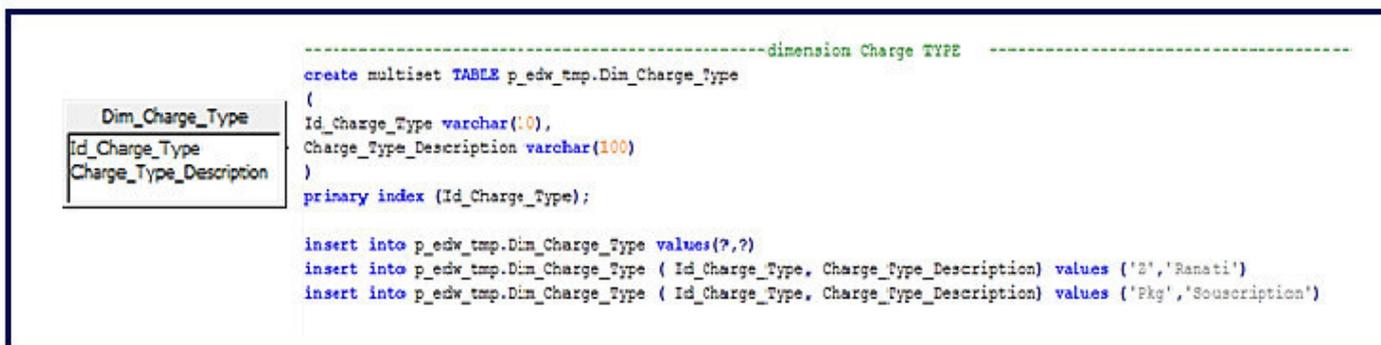
→ Dimension Charge_Type

La démarche suivie pour la création de la dimension est la suivante :

- 1- Pour trouver le nom de la table « Type de rechargement » dans les BDD de la source Production, nous avons utilisé les métadonnées sur les tables via l'instruction :

```
Select*From DBC.Tables where TableName Like '%charge%Type%'
```
- 2- Le nom de la table trouvé est : « VW_Billing_Charge_Type »
- 3- Ensuite, nous avons créé une table multiset vide Dim_Charge_Type contenant deux attributs : « Id_charge_type » et « charge_type_description »
- 4- Nous avons indexé la table avec le « primary index » : Id_chrage_type
- 5- Nous avons ensuite exporté la table de la source production sous forme de **fichier plat (txt)**
- 6- Nous avons importé ce même fichier dans la table vide que nous avons créé sur la BDD « P_EDW_TMP » de la source cible extension.
- 7- Nous avons ensuite chargé la table Dim_Charge_Type via l'import du fichier Txt
- 8- Nous avons au final rajouté les attributs « Z » : Ranati et « Pkg » : souscriptions manquantes dans la table « VW_Billing_Charge_Type » car cette dernière n'est pas à jour.

Pour cette dimension, comme la table source « VW_Billing_Charge_Type » n'était pas volumineuse, nous n'avons eu besoin de recourir aux outils d'export FastExport et TPT, nous avons utilisé une autre fonctionnalité rapide d'export dans les options de Teradata avant d'ajouter les attributs des deux types manquants comme on peut le voir sur le script SQL.



« Figure 20:Modèle physique et script de création de la dimension « Dim_charge_type »

→ Dimension Tariff_Group

Problématique :

- Table source disponible sous le nom « dbss_tariff_group » dans la source « Production ».
- Création de la dimension dans la source 2 « Extension », à partir de la table source « dbss_tariff_group ».
- Mise à jour quotidienne de cette dimension → Automatisation des mises à jour.

Solution :

- Création de deux tables : Dim_Tariff_Group et Dim_tarif_group_stg.
- Utilisation des outils FastExport, FastLoad et BTEQ
- Création de tâches planifiées sur Windows en organisant l'exécution des scripts dans un unique fichier de type « Batch » (Bat).
- **Création des deux tables**

La création des deux tables se fait via l'outil BTEQ.

- **Dim_tariff_group_stg (table intermédiaire) :** c'est une table intermédiaire (staging) ayant les mêmes attributs que la dimension. Elle est créée uniquement pour utiliser le FastLoad. Comme FastLoad ne prend en charge que les chaînes de caractères, et ne charge que si la table est vide, cette table a comme type d'attributs des chaînes de caractères et sera vidée après chaque chargement par un script BTEQ.
- **Dim_Tariff_Group :** Ça sera notre dimension finale.

- Utilisation des outils

- **FastExport :**

```
.LOGON 172.16.53.24/ [redacted] ;
.LOGTABLE P_edw_tmp.tg_log;
DATABASE p_edw bo universe;
]
.BEGIN EXPORT SESSIONS 2;
.EXPORT OUTFILE C:\Users\[redacted]\Desktop\PFE_MOUSSRAR\ETL_tg\tg.txt
MODE RECORD FORMAT TEXT;
select trim ("A") (FORMAT 'X(250)')(CHAR(250))
]
from( SELECT trim(coalesce(Tariff_Group_ID,'')) || '|' ||
trim(coalesce(Tariff_Group_Description,'')) || '|' ||
trim(coalesce(Charge_Type,'')) || '|' ||
trim(coalesce(Destination_Type,'')) || '|' ||
trim(coalesce(DESTINATION_COMPTA,'')) as A FROM p_edw_bo_universe.dbss_tariff_group ) A;
.END EXPORT;
.LOGOFF;
```

« Figure 21: Script FastExport pour la dimension « Dim_Charge_Type »

Remarque :

- Le nom de l'utilisateur ainsi que son mot de passe sont masqués en gris.
- Nous avons créé un fichier txt nommé « Logon_file » permettant l'accès à la base de l'utilisateur pour éviter que le nom et le mot de passe de l'utilisateur soit visible dans le script du FastExport par d'autres utilisateurs.

- **FastLoad :**

```
LOGON 172.16.53.41/[redacted] ;
DATABASE p_edw TMP;
BEGIN LOADING p_edw_TMP.Dim_Tariff_Group_Stg
ERRORFILES tg_ET, tg_UV
CHECKPOINT 10000;
SET RECORD VARTEXT "|";
DEFINE Tariff_Group_ID (VARCHAR(50)),
Tariff_Group_Description (VARCHAR(255)),
Charge_Type (VARCHAR(3)),
Destination_Type (VARCHAR(50)),
DESTINATION_COMPTA (VARCHAR(255)),
FILE = C:\Users\[redacted]\Desktop\PFE_MOUSSRAR\ETL_tg\tg.txt;
INSERT INTO Dim_Tariff_Group_Stg (
Tariff_Group_ID ,
Tariff_Group_Description ,
Charge_Type ,
Destination_Type,
DESTINATION_COMPTA
)
VALUES (
:Tariff_Group_ID ,
:Tariff_Group_Description ,
:Charge_Type ,
:Destination_Type,
:DESTINATION_COMPTA
);
END LOADING;
LOGOFF;
```

« Figure 22: Script FastLoad pour la dimension « Dim_Charge_Type »

Lors de l'exécution de ce script, nous avons rencontré les problèmes suivants :

- Lors de l'export en fichier txt : présence d'espaces entre chaque colonne (attribut), ce sont des tabulations → résolu par la fonction SQL « Trim »
- Présence de caractères spéciaux au début de la première colonne résolue → résolu par la fonction « Coalesce »

- **BTEQ** : Ci-dessous le script BTEQ :

```
.LOGON 172.16.53.24/ ;
DATABASE p_edw_tmp;
drop table p_edw_tmp.tg_log ;
.if ERRORCODE <> 0 THEN .EXIT ERRORCODE;
.LOGOFF;
```

« Figure 23: Script du BTEQ 1 pour la dimension « Dim_Charge_Type »

BTEQ 1 : se charge de supprimer la table Log de la dimension Tarif_Group

```
.LOGON 172.16.53.41/ ;
DATABASE p_edw_tmp;

drop table p_edw_tmp.tg_ET ;
drop table p_edw_tmp.tg_UV ;
drop table p_edw_tmp.Dim_Tariff_Group_stg ;
create MULTISET TABLE p_edw_tmp.Dim_Tariff_Group_stg
(
    Tarif_Group_ID VARCHAR(50) CHARACTER SET LATIN NOT CASESPECIFIC,
    Tarif_Group_Description VARCHAR(255) CHARACTER SET LATIN NOT CASESPECIFIC,
    Charge_Type CHAR(3) CHARACTER SET LATIN NOT CASESPECIFIC,
    Destination_Type VARCHAR(50) CHARACTER SET LATIN NOT CASESPECIFIC,
    DESTINATION_COMPTA VARCHAR(255) CHARACTER SET LATIN NOT CASESPECIFIC)
no PRIMARY INDEX ;
.if ERRORCODE <> 0 THEN .EXIT ERRORCODE;
.LOGOFF;
```

Figure 24: Script du BTEQ 2 pour la dimension Dim_Charge_Type

BTEQ 2 : se charge de :

- Supprimer les deux tables d'erreurs tg_ET et tg_UV créées par défaut.
- Supprimer puis créer la table intermédiaire Dim_Tariff_Group_stg.

```

.LOGON 172.16.53.41/
DATABASE p_edw tmp;
delete from p_edw_tmp.Dim_Tariff_Group ;
insert into p_edw_tmp.Dim_Tariff_Group
select
trim(Tariff_Group_ID)Tariff_Group_ID,
trim(Tariff_Group_Description)Tariff_Group_Description,
trim(Charge_Type)Charge_Type,
trim(Destination_Type)Destination_Type,
trim(DESTINATION_COMPTE) DESTINATION_COMPTE
from      p_edw_tmp.Dim_Tariff_Group_stg ;

delete from p_edw_tmp.Dim_Tariff_Group_stg;

.IF ERRORCODE <> 0 THEN .EXIT ERRORCODE;
.LOGOFF;

```

Figure 25: Script du BTEQ 3 pour la dimension « Dim_Charge_Type »

BTEQ 3 se charge de

- Supprimer le contenu de la dimension si elle est non vide (anciennes données).
- Insérer les données dans la dimension à partir de la table intermédiaire.
- Supprimer le contenu de la table intermédiaire pour un futur remplissage.

• Création de tâches planifiées

Afin d'automatiser le chargement de la dimension Tariff_Group, nous avons créé pour cela un fichier de type « Bat » qui regroupe les syntaxes d'exécution des scripts précédents.

Le fichier Bat en question sera exécuté automatiquement après création d'une tâche planifiée via l'option « Planificateur de tâche » de Windows.

Ci-dessous le script du fichier.Bat :

```

echo "-----preparation logtable -----"
bteq < C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\bteq_tg_1.sql >C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\bteq_tg_log_1.log
echo "-----fastexport-----"
fexp < C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\fx_tg.fx >C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\fx_tg_log.log
echo "-----drop create-----"
bteq < C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\bteq_tg_2.sql >C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\bteq_tg_log_2.log
echo "-----fast load -----"
FastLoad < C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\fl_tg.fl >C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\log_fl_tg_3.log
echo "-----finale table -----"
bteq < C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\bteq_tg_3.sql >C:\Users\...\Desktop\PF_E_MOUSSRAR\ETL_tg\bteq_tg_log_4.log
pause

```

Figure 26: Script d'exécution du fichier .Bat exécutant les scripts Fastexport, Fastload, et les scripts BTEQ pour la dimension « Dim_Charge_Type »

→ **Dimension Cell_Location**

Problématique :

- Table source disponible sous le nom « last_MOST_USED_CELL_IMEI_MONTHLY_DBSS » dans la source production.
- Création de la dimension dans la BDD P_EDW_TMP la source extension, à partir de la table source « last_MOST_USED_CELL_IMEI_MONTHLY_DBSS » de la source Production.
- Mise à jour quotidienne de cette dimension → Automatisation des mises à jour.

Solution :

- Création de deux tables : Dim_cell_loc_stg et Dim_Cell_Loc.
- Utilisation des outils TPT et BTEQ.
- Création de tâches planifiées sur Windows en organisant l'exécution des scripts dans un unique fichier de type « Bat ».

- **Création des deux tables**

- **Dim_cell_loc_stg : (table intermédiaire) :** c'est une table intermédiaire (staging) ayant les mêmes attributs que la dimension. Comme le FastLoad, le TPT ne charge que si la table est vide, cette table sera donc vidée après chaque chargement par un script BTEQ.
- **Dim_Cell_Loc :** c'est la dimension finale.

La création des deux tables se fait via l'outil BTEQ.

- **Utilisation des outils**

- **TPT (Voir annexe 16) :**

- **Nom du job pour le script :** Job_cell_loc
- **Description du job:** 'Job for loading Dim cell_location'
- **Définir le schéma de destination :** W_0_s_Job_cell_loc
- **Opérateur d'export :** W_0_o_Job_cell_loc
- **Opérateur chargement :** W_1_o_Job_cell_loc
- **Application du job.**

- **BTEQ :** ci-dessous les scripts des BTEQ utilisés

```

.LOGON 172.16.53.41/
DATABASE p_edw_tmp;
drop table p_edw_tmp.Dim_cell_log ;

drop table P_EDW_TMP.Dim_Cell_Loc_stg_ET;
drop table p_edw_tmp.Dim_Cell_Loc_stg_UV ;
drop table P_EDW_TMP.Dim_Cell_Loc_stg_WT;

drop table p_edw_tmp.dim_cell_loc_stg;

CREATE MULTISSET TABLE p_edw_tmp.dim_cell_loc_stg ,NO FALLBACK ,
NO BEFORE JOURNAL,
NO AFTER JOURNAL,
CHECKSUM = DEFAULT,
DEFAULT MERGEBLOCKRATIO
]
(
Subs_Id INTEGER,
Last_Most_Used_Cell VARCHAR(20) CHARACTER SET LATIN NOT CASESPECIFIC,
Last_Most_Used_Imei VARCHAR(15) CHARACTER SET LATIN NOT CASESPECIFIC)
NO PRIMARY INDEX ;
]
.IF ERRORCODE <> 0 THEN .EXIT ERRORCODE;
.LOGOFF;

```

« Figure 27: Script BTEQ 1 pour la dimension « Dim_Cell_Loc

BTEQ 1 : se charge de :

- Supprimer les tables : Error Table (ET), Uniqueness violations (UV), Work table (WT) créées par défaut.
- Supprimer puis recréer la table intermédiaire Dim_Cell_Loc_stg.

```

.LOGON 172.16.53.41/
DATABASE p_edw_tmp;

delete from P_EDW_tmp.Dim_Cell_Loc ;
insert into P_EDW_tmp.Dim_Cell_Loc
select * from P_EDW_tmp.Dim_Cell_Loc_stg ;

]
.IF ERRORCODE <> 0 THEN .EXIT ERRORCODE;
.LOGOFF;

```

« Figure 28: Script du BTEQ 2 pour la dimension « Dim_Cell_Loc

BTEQ 2 : Met à jour les données de la dimension Dim_Cell_Loc en la vidant puis en la remplissant de nouveau à partir des données de la table intermédiaire Dim_Cell_Loc_stg.

• Création de tâches planifiées

Nous avons créé un fichier « Bat » qui regroupe les syntaxes d'exécution des scripts précédents dans le but de les automatiser.

Le fichier Bat en question sera exécuté automatiquement après création d'une tâche planifiée via l'option « Planificateur de tâche » de Windows.

Ci-dessous le script du fichier Bat :

```

echo "-----preparation logtable -----"
bteq <C:\Users\...\Desktop\FFE_MOUSSRAR\ETL_loc\bteq_loc_1.sql >C:\Users\...\Desktop\FFE_MOUSSRAR\ETL_loc\bteq_loc_log_1.log
echo "-----tpt-----"
tbuild -L "C:\Users\...\Desktop\FFE_MOUSSRAR\ETL_loc" -f tpt.tpt job_location
echo "-----finale table -----"
bteq < C:\Users\...\Desktop\FFE_MOUSSRAR\ETL_loc\bteq_loc_2.sql >C:\Users\...\Desktop\FFE_MOUSSRAR\ETL_loc\bteq_loc_log_2.log
pause

```

Figure 29: Script d'exécution du fichier. Bat exécutant les scripts Fastexport, Fastload, et les scripts BTEQ pour la dimension « Dim_Cell_Loc »

Le temps d'exécution des scripts des outils utilisés lors de l'export et du chargement était relativement long vu l'importante quantité de données exportées et chargées qui avoisine les **14 millions de lignes**.

1.1.2. Création des tables de faits

Pour créer nos deux tables de faits, nous avons effectué des vues sur deux tables du DWH contenant comme attributs les clés étrangères des dimensions créées ainsi que les mesures nécessaires au calcul des KPI.

→ Table Fact_Traffic

Pour créer cette table, nous avons effectué une vue sur la table « daily_subscriber_traffic_dbss » de la base p_edw_bo_universe dans la source extension en gardant comme attributs que les clés primaires de nos dimensions en tant que clés étrangères. Ci-dessous le script¹ de création de la table :

```

create view p_edw_tmp.Fact_Traffic as
LOCKING P_edw_bo_universe.daily_subscriber_traffic_dbss FOR ACCESS
select
Subscriber_ID as SUBSCRIBER_ID,
TRANSDATE as CALENDAR_DATE ,
Offer_ID as OFFER_KEY,
Operator as OPERATOR,
Call_Direction_Type as ID_CALL_DIRECTION ,
Charge_Type as ID_CHARGE_TYPE,
Tariff_Group_ID as TARIFF_GROUP_ID ,
-----Mesures-----
Number_of_Calls,
Actual_Duration ,
Billed_Duration ,
Billed_Amount
from p_edw_bo_universe.daily_subscriber_traffic_dbss:

```

Figure 30: Script de création de la table de faits Fact_traffic

¹ On voit sur ce script la clé « OPERATOR » qui est la clé primaire d'une dimension dégénérée relative aux données sur les différents opérateurs réseau (voir définition, page 22).

→ **Table Fact_Refill**

Pour créer cette table, nous avons effectué une vue sur l'union des deux tables « VOUCHER_BY_SUBSCRIBER_DETAIL » et « FLEXY_BY_SUBSCRIBER_DETAIL » de la base p_edw_bo_universe dans la source « Extension » en gardant comme attributs que les clés primaires de nos dimensions en tant que clés étrangères.

Ci-dessous le script de création de la table :

```
REPLACE VIEW P_EDW_TMP.Fact_Refill AS

LOCKING P_EDW_BO_UNIVERSE.FLEXY_BY_SUBSCRIBER_DETAIL FOR ACCESS
LOCKING P_EDW_BO_UNIVERSE.VOUCHER_BY_SUBSCRIBER_DETAIL FOR ACCESS

SELECT Subs_id,offer_id, Refill_type, REFILL_DATE , Amount, Bonus_Amount
FROM
] (
SELECT Subs_Id,Offering_Id as offer_id, 'Voucher' Refill_type, REFILL_DATE, CAST (REFILL_VALUE AS DECIMAL (18,2)) Amount, Bonus_Amount
FROM P_EDW_BO_UNIVERSE.VOUCHER_BY_SUBSCRIBER_DETAIL
WHERE REFILL_TYPE <>'PAYMENT'
UNION ALL
SELECT Destination_Subscriber_id Subs_id ,dest_offer , 'Flexy' Refill_type,Transaction_date (DATE) REFILL_DATE, CAST (TRANSFERRED_AMOUNT AS DECIMAL (18,2)) Amount, Bonus_Amount
FROM P_EDW_BO_UNIVERSE.FLEXY_BY_SUBSCRIBER_DETAIL
WHERE R2P_PURPOSE=0
) RECHARGE;
```

Figure 31: Script de création de la table de faits Fact_Refill

1.2. Présentation du DM sur Business Object :

Pour visualiser et exploiter le modèle physique de notre DM sur le logiciel « Business Object » (BO)¹, voici brièvement comment nous avons procédé :

- Nous avons donné un nom à notre Univers² : **DM_Universe**.
- Nous avons donné une description à notre Univers qui apparaîtra sous le module BO.
- Nous avons ensuite créé une **connexion à notre base de données de Teradata (extension)** : en passant par le choix du driver « *Open Database Connectivity* » (ODBC) à utiliser pour se connecter à cette base.

¹Business Objects : Créée en 1990, c'est le premier éditeur mondial des outils d'aide à la décision pleinement intégrés et évolutifs pour l'Entreprise, donnant aux utilisateurs finals un accès intuitif aux informations stockées dans les bases de données et les progiciels applicatifs (ERP).

²Univers : Chez Business Objects, c'est une représentation orientée métier des données qui regroupent les mots de vocabulaire professionnel usuel des utilisateurs. C'est une représentation totale ou partielle de la base de données adaptée à une fonction de l'entreprise, à un groupe d'utilisateurs, ou à une application particulière.

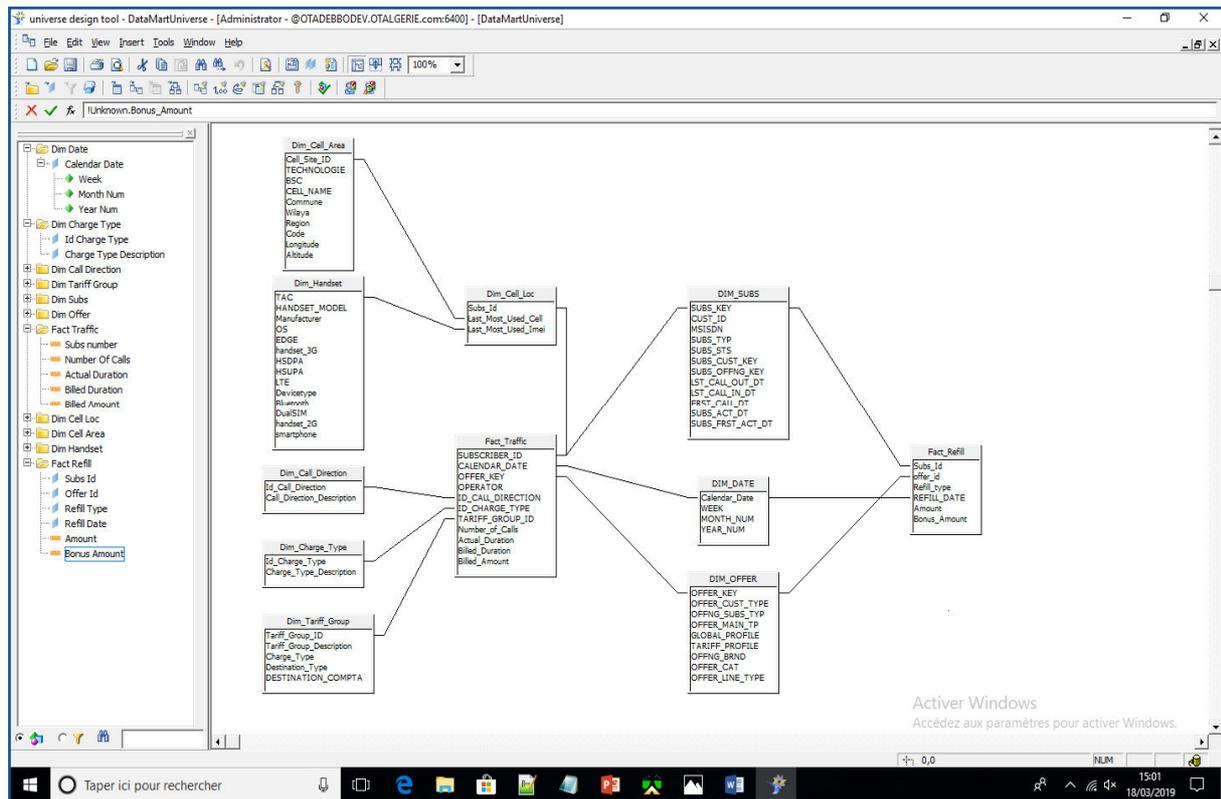


Figure 32: Interface de BO accompagné du modèle du DM

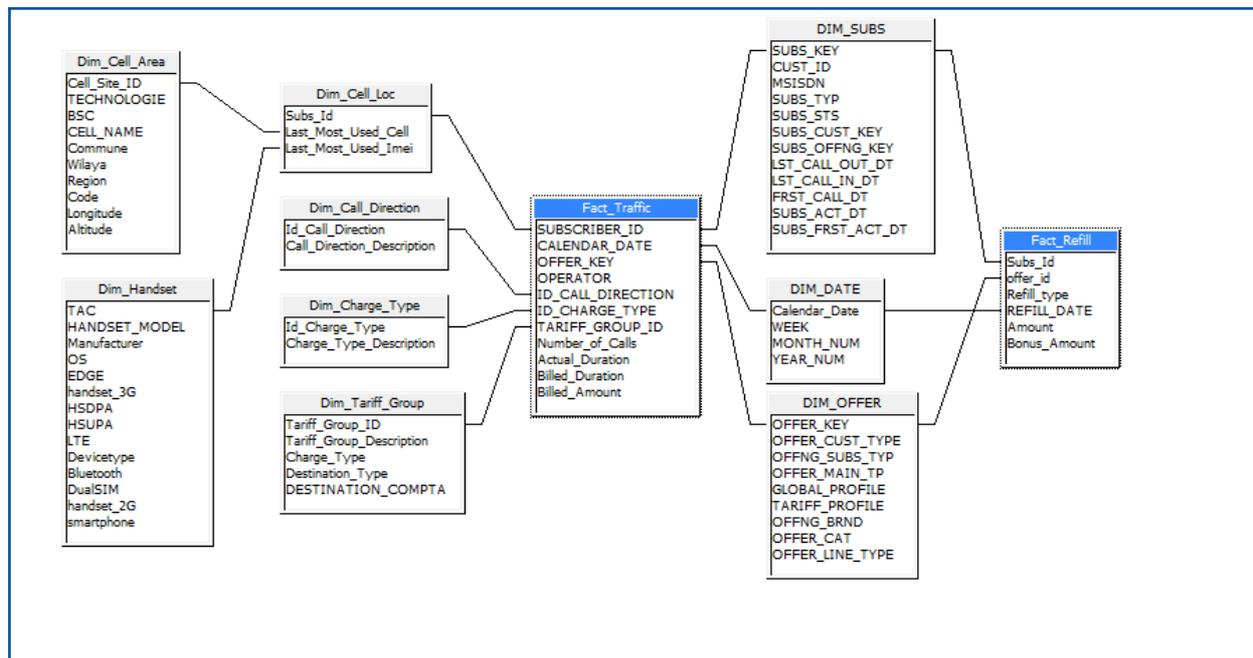


Figure 33: Modèle physique du DM sur BO

1.3. Apports du DM :

Le DM que nous avons créé répond non seulement au besoin exprimé par le service CVM qui est satisfait par les mesures des tables de faits permettant le calcul des KPI, mais aussi son utilité s'étend aux autres services de l'entreprise. En effet, il permet aux utilisateurs finaux

business de créer en quelques clics leurs propres rapports à partir du DM implémenté sur BO et cela sans passer par un intermédiaire expert qui prend un temps nettement supérieur facilitant ainsi l'accès à l'information pour la prise de décision.

Concernant le service CVM, le DM Facilitel'accès aux données qui sont désormais toutes disponible dans la source « Extension » et il Réduitle temps de lancement d'une campagne sachant quele temps mis par les analystes du service CVM pour le calcul des indicateurs clés nécessaires au choix des campagnes et des abonnés visés par celle-ci est approximativement passé de 4 à 1 jour.

Voici un exemple d'utilisation sur WebIntelligence¹ deBO de notre DM pour élaborer un rapport décisionnel :

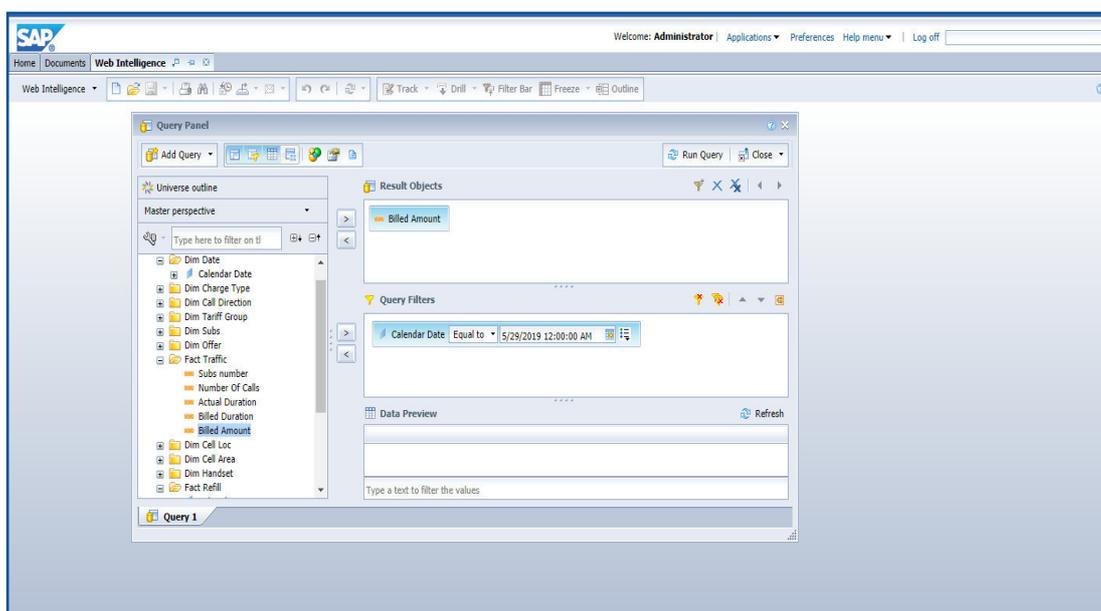


Figure 34: Exemple de création d'un rapport sur BO (revenu journalier)

2. Conception de l'ADS :

Dans cette deuxième partie de notre solution, nous avons entamé la conception d'un outil répondant au besoin exprimé par le service CVM, qui est l'automatisation du calcul des KPI.

Cet outil, appelé « Analytical Dataset » est une table de données regroupant, par abonné prépayé, l'ensemble de ces KPI calculés à partir du DM.

¹**WebIntelligence** : est une solution d'analyse et de reporting de BO, elle répondaux besoins de tous les profils d'utilisateurs et permet d'accéder aux informations stockées dans les bases de données de l'entreprise. Elle permet "en libre-service" de lancer des requêtes, de générer des rapports et d'effectuer des analyses OLAP via le Web au travers d'un browser sur un poste de type client léger.

2.1. Du DM à l'ADS :

Le DM conçu constitue une base de données analytique permettant via ses mesures le calcul des KPI retenus qui sera automatisé et présenté par la suite dans l'ADS.

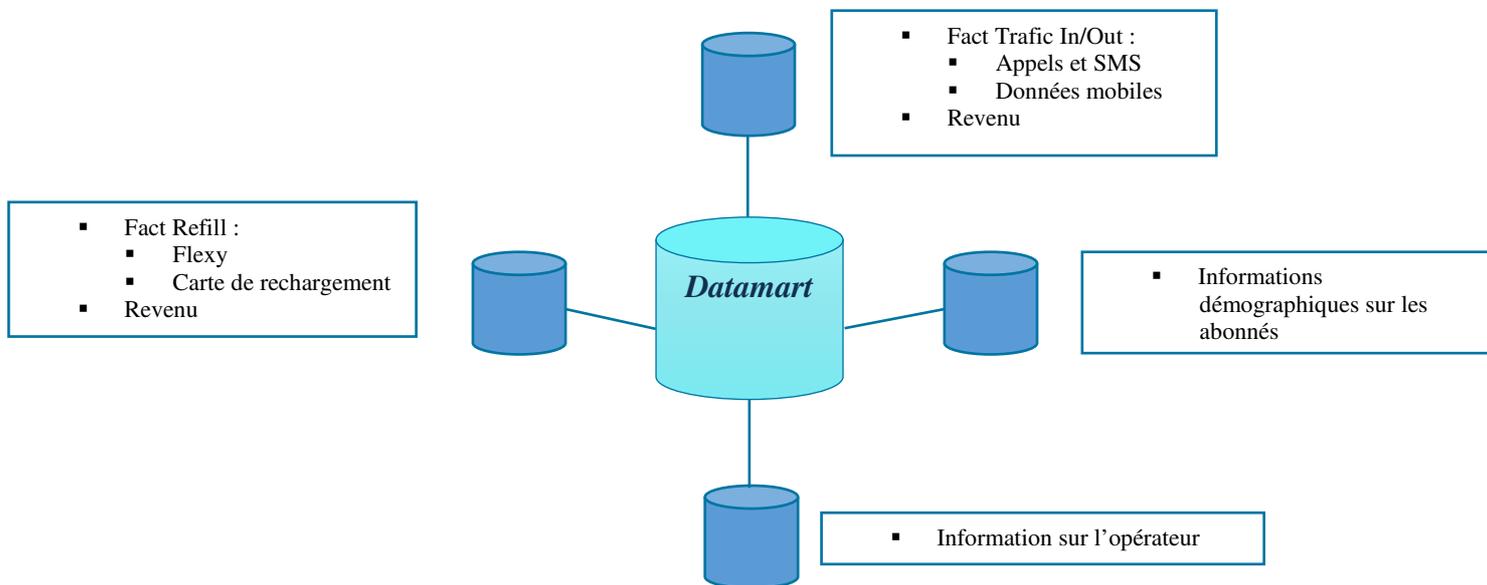


Figure 35: Données contenu dans le DM

2.1.1. KPI retenus

Par abus de langage, tout indicateur est appelé KPI dans l'entreprise, ainsi les indicateurs ci-dessous ne sont pas tous des indicateurs clés permettant la prise de décision.

Avant d'entamer le calcul de ces indicateurs relatifs aux abonnés prépayés, nous les avons catégorisés comme suit :

Famille des KPI	Catégorie	Description
KPI Traffic	Outgoing	Ce sont les KPI relatifs aux abonnés ayant émis des appels et sms (sortant)
	Incoming	Ce sont les KPI relatifs aux abonnés ayant reçus des appels et sms (entrant)
KPI Refill	Cartes (Voucher)	Ce sont les KPI relatifs aux abonnés ayant effectué un rechargement de type Voucher (carte)
	Flexy	Ce sont les KPI relatifs aux abonnés ayant effectué un rechargement de type Flexy

Tableau 9: Catégories et description des Familles des KPI retenus

▪ **KPI Outgoing Traffic**

Catégorie des KPI		Description	KPI
Voice	On-Net	Les appels émis vers Djezzy	NBER_CALLS : Nombre d'appels VOICE_REV : Revenu des appels VOICE_USAGE : Durée des appels
	Off-Net	Les appels émis vers les autres opérateurs et vers le fixe	
	International	Les appels vers l'international	
SMS	On-Net	Les SMS envoyés vers Djezzy	NBER_SMS : Nombre d'SMS REV_SMS : Revenu sur les SMS
	Off-Net	Les SMS envoyés vers les autres opérateurs	
	International	Les SMS envoyés vers l'international	
Data		Consommation en données mobiles	DATA_USAGE_MO : Consommation data en Mo DATA_USAGE_MO_PAYG : Consommation data « Pay as you go ¹ » DATA_REV_MO_PAYG : Revenu sur les data PAYG
Subscription		Souscription à une offre de Djezzy	NBER_SOUSC : Nombre de souscriptions REV_SOUSC : Revenu sur les souscriptions
ARPU		Average revenu par user	ARPU : le revenu moyen par utilisateur (par mois)

Tableau 12: Liste et description des KPI Outgoing calculés à partir de la table « Fact_Traffic »

¹« Pay As You Go » est un service de souscription mobile qui consiste à consommer ce que l'on a payé à l'avance.

- **KPI Incoming Traffic**

		Description	KPI
Voice	Off-Net	Les appels reçus des autres opérateurs et du fixe	NBER_CALLS_IN : nombre d'appels reçus (On-Net, Off-Net et international) NBER_CALLS_IN_OFFNET : nombre d'appels reçus des autres opérateurs
SMS			SMS_IN : nombre d'SMS reçu

Tableau 13: Liste et description des KPI Incoming calculés à partir de la table « Fact_Traffic »

On constate qu'il y a moins de KPI incoming par rapport à l'outgoing, jugés moins important car leur revenu est nul et que seuls les KPI incoming Off-Net (reçus des autres opérateurs) et SMS_IN (reçus) ont été retenus car utiles à l'analyse.

- **KPI_Refill**

KPI	Description
Amount_day_minus_1	Revenu sur le rechargement à j-1
Rev_Refill_Max	Revenu Max rechargé (par mois)
Potentiel	Max revenu rechargé entre les 6 derniers mois - Revenu rechargé M1, permet de déterminer le potentiel d'un abonné durant les 6 derniers mois
variance_refill	La variance sur le revenu des rechargements
ecartype_rev	L'écart type sur le revenu des rechargements
GAP	Moyenne de jours entre 2 recharges (3 mois)
Nber_Refill	Nombre de rechargements
Rev_Refill	Revenu sur les rechargements

Nber_Refill_Flexy	Nombre de rechargements de type Flexy
Rev_Refill_Flexy	Revenu sur les rechargements de type Flexy
Nber_Refill_Voucher	Nombre de rechargements de type Voucher
Rev_Refill_Voucher	Revenu des rechargements de type de Voucher

Tableau 15: Liste et description des KPI calculés à partir de la table « Fact_Refill »

Remarques :

- 1- Chaque KPI sera calculé par semaine et par mois glissant ¹comme suit :
 - M1 : de j-1 à j-30
 - M2 : de j-31 à j-60
 - M3 : de j-61 à j-90
 - M4 : de j-91 à j-120
 - M5 : de j-121 à j-150
 - M6 : de j-151 à j-180
 - W1 : de j-1 à j-7
 - W2 : de j-8 à j-14
 - W3 : de j-15 à j-21
 - W4 : de j-22 à j-28
- 2- Le nombre total de KPI est de **230** calculés pour plus de **14 millions** d'abonnés.

2.2. Réalisation technique de l'ADS

Dans cette partie, nous aborderons la démarche suivie pour aboutir à l'automatisation du calcul des KPI qui se divise en deux parties. La première partie consiste à calculer les KPI par catégorie et ceci via des requêtes SQL, ensuite à les charger dans des tables distinctes (par catégorie) puis dans la table finale (ADS) et ceci via des macros². La deuxième partie est l'automatisation de ces chargements via l'outil BTEQ exécutant l'ensemble des macros dans un fichier Bat mensuellement.

Les opérations de calcul des KPI, de création et de chargement des tables ainsi que les macros se sont faites sur la base P_EDW_TMP de la source extension.

¹Un mois glissant commence à partir du jour ou une campagne marketing est lancée, le principe est identique pour la semaine glissante.

²Une macro est destinée à contenir des instructions SQL exécutées automatiquement une fois que la macro est sélectionnée en tapant son nom et en l'exécutant en tant qu'instruction SQL.

2.2.1. Calcul des KPI et chargement des tables ¹

Nous avons calculé les KPI par catégorie via des requêtes SQL, ci-dessous un exemple de calcul des KPI sur le revenu appels sortant pour le mois M1 :

```
select
  SUBSCRIBER_ID
  ,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-30 AND :run_date-1
AND ID_CHARGE_TYPE='L' AND TARIFF_GROUP_ID IN ( 'JZYJZY','FWD-JZYJZY')
THEN BILLED_AMOUNT ELSE 0 END) VOICE_REV_ON_NET_M1

FROM P_EDW_TMP.FACT_TRAFFIC AS A
INNER JOIN P_EDW_TMP.DIM_OFFER AS B ON A.OFFER_KEY=B.OFFER_KEY

WHERE ID_CALL_DIRECTION='O'
AND CALENDAR_DATE BETWEEN :run_date-90 AND DATE
GROUP BY 1
;
```

Figure 36: Script de calcul de KPI sur le revenu appels sortant pour le mois M1 (Mars 2019)

2.2.2. Création et chargement des tables temporaires

Après avoir terminé le calcul des 230 KPI, nous avons créé pour chaque catégorie une table vide temporaire qui sera chargée via des macros qui reprennent les mêmes instructions des scripts du calcul. Ci-dessous la liste des tables et des macros pour chaque catégorie de KPI :

Catégorie de KPI	Tables temporaires	Macro de chargement
KPI_Traffic_Outgoing	ADS_OUTGOING_MONTH	LOAD_ADS_OUTGOING_Month
	ADS_OUTGOING_WEEK	LOAD_ADS_OUTGOING_Week
KPI_Traffic_Incoming	ADS_INCOMING	LOAD_ADS_INCOMING
KPI_Refill	ADS_REFILL	LOAD_ADS_REFILL

Tableau 16: Liste des tables et des macros pour chaque catégorie de KPI

- Nous avons créé ces tables temporaires pour des raisons d'optimisation de la durée d'exécution des macros, du chargement des données.
- Le chargement des KPI traffic outgoing s'est fait dans deux tables car il avait mis trop de temps (environ 1 heure) lorsqu'on lui a dédié qu'une seule table.

¹Tous les scripts de calcul des KPI, de la création des tables et des macros sont en annexe (Voir annexes 17 à 24).

- Comme il est visible sur les scripts des macros, celles-ci sont paramétrées par une date nommée « Run_date » permettant à l'utilisateur d'accéder aux KPI calculés correspondant à la date introduite.

2.2.3. Création et chargement d'une table cible :

Nous avons créé une table appelée « Target_Base » dans le but de contenir que les abonnés prépayés avec un statut **actif** c'est-à-dire ceux qui ont fait au moins un trafic (Appels et/ou SMS et/ou data) durant les 90 derniers jours.

Cette table contient en plus de l'attribut « statut de l'abonné », son identifiant, la date du premier appel reçu ou émis ainsi que celle du dernier appel reçu et émis pour calculer par la suite l'âge de l'abonné.

Ci-dessous les scripts de la création de la table ainsi que celui de sa macro de chargement :

```
create multiset table P_EDW_TMP.TARGET_BASE
AS (
SELECT
SUBS_KEY
,SUBS_STS
,LST_CALL_OUT_DT
,LST_CALL_IN_DT
from P_EDW_TMP.DIM_SUBS
where SUBS_STS in ('ACT','SUS','SSD','DRM','RTER')
) with no data
primary index(SUBS_KEY)
```

Figure 37: Script de Création de la table « Target_Base »

```
replace MACRO P_EDW_TMP.LOAD_TARGET_BASE
AS (
delete from P_EDW_TMP.TARGET_BASE;
insert into P_EDW_TMP.TARGET_BASE
SELECT
SUBS_KEY
,SUBS_STS
,FRST_CALL_DT
,LST_CALL_OUT_DT
,LST_CALL_IN_DT
from P_EDW_TMP.DIM_SUBS a
inner join p_edw_tmp.DIM_OFFER b on a.SUBS_OFFNG_KEY =b.OFFER_KEY
where SUBS_STS in ('ACT','SUS','SSD','DRM','RTER') ;
);
```

Figure 38: Script de création de la Macro de chargement de la table « Target Base »

2.2.4. Création et chargement de l'ADS

2.2.4.1. Création

Après avoir créé la table cible « Target_Base » contenant les abonnés actifs, et afin d'en retenir que ceux qui ont fait un trafic sortant, ou entrant, ou bien une opération de rechargement, nous avons effectué une jointure de type « left join » entre la table « Target_Base » et les tables « ADS_Outgoing_Month », « ADS_Outgoing_Week », « ADS_Incoming » et « ADS_Refill », pour obtenir notre table ADS nommée « **ADS_Final** » (Voir annexe 25)

Cette jointure permet de lister :

- Tous les résultats de la table gauche « Target_Base » même s'il n'y a pas de correspondance avec les tables de droites et dans ce cas les colonnes de celles-ci vaudront la valeur nulle.
- Les résultats de l'intersection entre la table de gauche et celles de droite.

Au final, la table ADS contient les attributs suivants :

Nom de l'attribut	Type	Description
RUN_DATE	Date YY/MM/DD	Cet attribut date a été calculé de telle sorte que la date relative aux résultats de l'ADS soit le premier du mois en cours. Elle est calculée comme suit : date-extract (day from date)
SUBS_ID	Integer	La clé primaire de la table ADS, qui est l'identifiant de l'abonné.
FRST_CALL_DT	Timestamp YYYY-MM-DDBHH:MI:SS	La date du premier appel émis ou reçu par l'abonné
LST_CALL_OUT_DT	Timestamp	La date du dernier appel émis par l'abonné
LST_CALL_IN_DT	Timestamp	La date du dernier appel reçu par l'abonné
Tous les KPI Outgoing	Type et description (Voir annexe 26)	
Tous les KPI Incoming		
Tous les KPI Refill		

Tableau 17: Liste des attributs de l'ADS, leur type et leur description

2.2.4.2. Chargement

Le chargement de l'ADS (Voir annexe 27) se fait à la fin du mois alors que les résultats qu'il affiche sont datés du premier de ce même mois pour les raisons suivantes :

- 99% des données du DWH sont prêtes à la fin de chaque mois car le « Call Detail Record », qui est le fichier plat détaillé d'un appel d'une date donnée, ne parvient au DWH de Djazzy que dix à quinze jours après ; ce qui justifie le chargement de l'ADS à la fin du mois.
- L'attribut « Run_Date » renvoie au premier du mois pour avoir un ADS mensuel daté du premier de chaque mois.

2.2.5. Automatisation du calcul des KPI

On entend par automatisation du calcul des KPI, l'exécution automatique des macros de chargement des tables des KPI vu précédemment et de l'ADS, ceci a été réalisé comme suit :

- 1- Utilisation de l'outil BTEQ qui exécute l'ensemble des macros de chargement des tables des KPI et de l'ADS. Ci-dessous le script y correspondant :

```
.LOGON 172.16.53.41/
DATABASE p_edw_tmp;
exec P_EDW_TMP.LOAD_TARGET_BASE;
exec P_EDW_TMP.LOAD_ADS_OUTGOING_MONTH (date-extract(day from date)+1);
exec P_EDW_TMP.LOAD_ADS_OUTGOING_WEEK (date-extract(day from date)+1);
exec P_EDW_TMP.LOAD_ADS_INCOMING (date-extract(day from date)+1);
exec P_EDW_TMP.LOAD_ADS_REFILL (date-extract(day from date)+1);
exec p_edw_tmp.LOAD_ADS_FINAL;
|.IF ERRORCODE <> 0 THEN .EXIT ERRORCODE
.LOGOFF;
```

Figure 39: Script BTEQ exécutant les macros de chargement des tables des KPI et de l'ADS

- 2- Création d'un fichier « Bat » qui regroupe les syntaxes d'exécution du script BTEQ dans le but d'automatiser son exécution. Le fichier Bat en question sera exécuté automatiquement après création d'une tâche planifiée via l'option « Planificateur de tâche » de Windows. Ci-dessous le script correspondant :

```
echo "-----ADS-----"
bteq <"C:\Users\...\.Desktop\PFE_MOUSSRAR\Codes_SQL\BTEQ_ADS.sql" >C:\Users\...\.Desktop\PFE_MOUSSRAR\Execution_ADS.log
pause
```

Figure 40: Script du fichier .Bat exécutant le BTEQ de chargement des tables de KPI et de l'ADS

Note :

L'ADS contient 233 attributs, sur plus de 14 millions d'abonnés.

Le modèle physique de la table ADS se trouve en annexe (Voir annexe 32).

Les temps d'exécution des macros de chargement une par une sont visibles sur le script d'exécution suivant :

```
exec P_EDW_TMP.LOAD_TARGET_BASE; -- 4 sec---
exec P_EDW_TMP.LOAD_ADS_OUTGOING_MONTH (date-extract(day from date)+1);-- 22MN 20S---
exec P_EDW_TMP.LOAD_ADS_OUTGOING_WEEK (date-extract(day from date)+1);----18mn13sec---
exec p_edw_tmp.LOAD_ADS_INCOMING (date-extract(day from date)+1);----12mn10sec-----
exec P_EDW_TMP.LOAD_ADS_REFILL (date-extract(day from date)+1); -- 01min40sec
exec p_edw_tmp. LOAD_ADS_FINAL;-----15sec-----
```

Figure 41: Script et temps d'exécution des macros de chargement

Chapitre 02

Segmentation et tableau de bord

Dans ce deuxième chapitre de la solution mise en place, nous aborderons la phase analyse de notre projet BI pour répondre à la deuxième partie de notre problématique qui est l'exploitation des KPI calculés précédemment, pour le ciblage des campagnes marketing.

Ce travail a été découpé en deux parties :

1. L'implémentation d'un algorithme de classification (K-Means) Pour une segmentation comportementale des clients.
2. La conception d'un tableau de bord analytique pour la visualisation dynamique des KPI.

1. Segmentation comportementale des abonnés

Cette partie de notre projet intervient après avoir achevé l'ADS, c'est la phase d'exploitation et d'analyse des données, elle a pour objectif d'affecter chaque abonné prépayé à une classe d'appartenance comportementale en se basant sur les valeurs des KPI de celui-ci.

Pourquoi une segmentation comportementale ¹?

La concurrence entre les opérateurs mobiles est de plus en plus basée sur le comportement des abonnés. Afin d'améliorer la compétitivité des opérateurs de téléphonie mobile et la valeur client, plusieurs technologies de data mining peuvent être utilisées. L'une des technologies d'extraction de données les plus importantes est le clustering et la segmentation des clients. Cette pratique de ciblage s'est avérée générale et efficace pour le secteur des télécommunications mobiles.

Comme les campagnes marketing effectuées par le service CVM ciblent des segments de clients en se basant sur leur comportement, c'est-à-dire leur trafic en termes d'appels, de sms et de consommation en données mobiles, nous avons jugés, après entretien avec les membres de ce service, que réaliser une segmentation des clients sur la base de leur trafic pourrait se révéler utile lors du lancement d'une campagne.

Bien que le service CVM en collaboration avec l'équipe data science du service EIM a déjà établi une segmentation comportementale, la nôtre diffère du fait qu'elle exploite les données d'une seule table (ADS), et que ses résultats seront intégrés en tant qu'attributs (numéro et description de la classe d'abonnés) dans l'ADS. Ceci met en avant la facilité d'exploitation des données prêtes et organisées de l'ADS pour un lancement futur d'algorithmes d'analyses de données en se basant sur les variables décisionnelles que sont les KPI.

1.1. Langage et algorithme utilisés

▪ R et RStudio

R est un langage de programmation open source pour le calcul statistique et affichage des graphiques.

RStudio est un IDE (Integrated Development Environment) gratuit et open source pour R. Son interface est organisée de manière à ce que l'utilisateur puisse voir clairement le code R, les graphiques et les tables de données. Il offre également une fonctionnalité de type import-Wizard permettant d'importer n'importe quel type de fichier (BENMALEK, 2018).

¹ Voir définition et méthodologie suivie, page 31-32

Nous avons choisi le langage R pour sa facilité de syntaxe et pour les nombreuses bibliothèques de calcul statistique et de visualisation des données qu'il intègre.

▪ Algorithme K-means

L'objectif de notre segmentation est d'aboutir à un certain nombre de classes d'abonnés caractérisées par à un profil client, ce profil est décrit par des variables quantitatives et qualitatives reflétant un comportement client particulier.

Il existe de nombreux algorithmes de classification, chacun présentant des avantages et des inconvénients. Dans notre cas, nous avons opté pour l'algorithme de classification K-means (voir algorithme k-means, page 34) pour les raisons suivantes :

- Il est très facile à comprendre et à implémenter.
- Sa simplicité conceptuelle et sa rapidité.
- Applicable à des données de grande taille et à tout type de données (textuelles aussi).

Le fait que l'algorithme peut traiter efficacement des données de très grande taille, a été la raison principale de notre choix. En effet, nous l'avons appliqué sur les données de l'ADS, une table de plus de **14 millions** de lignes.

Cependant, l'algorithme présente aussi des inconvénients :

- Le nombre de classes doit être fixé au départ.
- Le résultat dépend du tirage initial des centres des classes.

Dans notre cas, le nombre de classes fixé au départ a été un avantage plus qu'un inconvénient, puisque le service CVM concerné a exigé un seuil se situant entre 6 et 10, concordant avec le nombre de profils d'abonnés (SubscriberProfile) déjà établi.

1.2. Données utilisées ¹

Les données exploitées par l'algorithme sont celles contenues dans l'ADS. Plus précisément, nous avons sélectionné un certain nombre d'attributs (KPI) représentant les variables de décision de l'algorithme.

Les attributs retenus de l'ADS sont ceux qui reflètent au mieux le comportement client relatif à leur trafic, et qui feront ressortir, par la suite, des profils d'abonnés aussi distincts que possible.

¹Le mois retenu pour les données utilisées est le mois de Mars.

Les attributs quantitatifs ¹:

- Subs_id ²
- NBER_CALLS_M1
- NBER_CALLS_ON_NET_M1
- NBER_CALLS_OFF_NET_M1
- NBER_CALLS_IN_M1
- NBER_CALLS_OFFNET_IN_M1
- NBER_CALLS_INT_M1
- NBER_SMS_ON_NET_M1
- NBER_SMS_OFF_NET_M1
- SMS_IN_M1
- DATA_USAGE_MO_M1+ DATA_USAGE_MO_PAYG_M1 = VOLUME_DATA

Nous avons rajouté en plus des variables quantitatives ci-dessus, des variables qualitatives pour :

- Ne pas biaiser la segmentation obtenue par l'algorithme avec le profil d'un abonné.
- Renforcer par la suite l'interprétation des résultats donnés par l'algorithme.

Ci-dessous la liste de ces variables :

Nom de la variable	Description	Valeur
Tenure	Elle représente l'âge de l'abonné et se calcule comme suit : Date du dernier appel-date du premier appel	<=3Mois >3Mois et <= 6Mois >6Mois et <= 12Mois >12Mois et <= 24Mois >3 ans et <= 5 ans >5 ans
Value_Segment	Elle représente une segmentation des abonnés selon une valeur attribuée par rapport à des seuils du KPI APRU : Elle est catégorisée comme suit :	Zéro : ARPU =0 DA VLV : 0<ARPU<=500 DA LV : 500<ARPU<=1000 DA

¹ Voir détail des variables qualitatives dans les tableaux, pages 60-61.

²L'attribut « Subs_id » est l'identifiant d'un abonné, il servira à organiser les résultats de l'algorithme par abonné.

	Very Low Value (VLV) Low Value (LV) Medium Value (MV) High Value (HV) Very High Value (VHV)	MV : $1000 < \text{ARPU} \leq 1500$ DA HV : $1500 < \text{ARPU} \leq 2500$ DA VHV : > 2500 DA
DeviceType	Le type de téléphone utilisé par un abonné prépayé	Basic phone Feature phone Smartphone
SubscriberProfile	Elle représente l'offre Djezzy souscrite par un abonné	Mifi Djezzy Go Djezzy Good Djezzy Carte New Djezzy Carte Djezzy Bayna Djezzy Hayla Djezzy PLAY Djezzy Hadra Djezzy HAYLABEZZEF Djezzy Pack 1

Tableau 18: Liste et description des variables qualitatives retenues pour de la segmentation

Les attributs ont été regroupés dans une table nommée « ADS_Kmeans » (Voir annexe 28) dont le chargement s'effectue à partir des données de l'ADS pour toutes les variables quantitatives ainsi que la variable « Tenure », jointe à celles des dimensions **Dim_Cell_loc** et **Dim_Handset** pour la variable « Devicetype », et de la dimension **DIM_OFFER** pour la variable « SubscriberProfile ».

1.3. Implémentation de l'algorithme sur R

1.3.1. Intégration des données

Lors de cette première étape, nous avons tout d'abord établi une connexion entre R et Teradata permettant d'exporter les données de la table « ADS_Kmeans ». Ensuite, nous avons

effectué leur chargement dans R dans une table nommée « dataset », dupliquée sous le nom « dataset_treat » qui sera celle utilisée pour la manipulation :

```

library( DBI)
library(rJava)
library(RJDBC)
library(ggplot2)
#-----Connexion to teradata

drv <- JDBC("com.teradata.jdbc.TeraDriver", Sys.glob("~/jdbcTera/*.jar"))
con <- dbConnect(drv, "jdbc:teradata://172.16.53.42", user = " ", password= " ")
#-----Load the ADS to R

dataset <- dbGetQuery (con, ("SELECT * FROM p_edw_tmp.ads_kmeans"))
dataset_treat<-dataset

```

Figure 42: Chargement des données dans R dans une table nommée « dataset »

1.3.2. Traitement des données

Dans cette partie, nous avons :

- Supprimé l'attribut « Subs_id » ;
- Détecté et traité les valeurs aberrantes des variables quantitatives ;
- Supprimé les variables qualitatives avant de lancer l'algorithme ;
- Standardisé les données à la même échelle (Normalisation des données).

1.3.2.1. Suppression de l'attribut « Subs_id »

Tout d'abord, nous avons supprimé l'attribut « Subs_id » du data frame « dataset_treat » ne représentant pas une variable de décision comme suit :

```
dataset_treat <- subset( dataset_treat, select = -c(Subs_id))
```

Figure 43: Suppression de l'attribut Subs_id

1.3.2.2. Détection et traitement des données Aberrantes (Outliers)

▪ Détection

En visualisant les observations de chaque variable quantitative via la boîte à moustache¹, nous remarquons la présence de valeurs aberrantes², se situant en dehors des limites supérieure et inférieure de la boîte à moustache.

¹Les boîtes à moustaches permettent de visualiser plusieurs paramètres de distribution d'une variable : la médiane, l'intervalle interquartile et la valeur maximale et minimale de la distribution.

²Une valeur aberrante est une observation inhabituellement élevée ou basse. Ce sont les valeurs se trouvant en dehors des limites supérieures correspondant au 95%ile et inférieures correspondant au 5%ile.

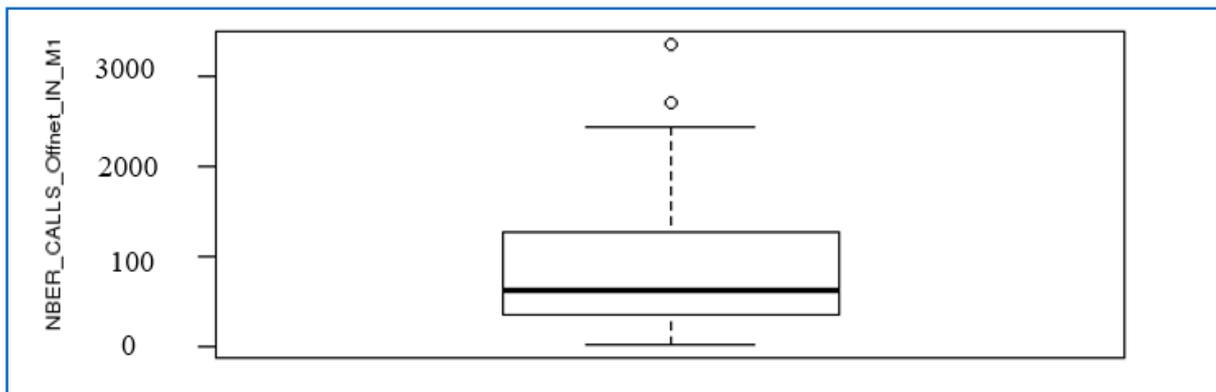


Figure 44: Visualisation des valeurs de la variable NBER_CALLS_Offnet_IN_M1 via la Boite à moustaches

▪ Traitement

Afin d'éliminer les valeurs aberrantes, nous avons utilisé un algorithme de « Capping » en créant une fonction « remove_outliers »¹ dont le principe est le suivant :

Pour les valeurs aberrantes qui se situent en dehors de la $1.5 * IQR (x_{75} - x_{25})$ limites, nous avons plafonné en remplaçant ces observations en dehors de la limite inférieure par la valeur (x_5) correspondant au 5%ile. Celles qui se situent au-dessus de la limite supérieure par la valeur (x_{95}) correspondant au 95% ile (Prabhakaran, 2016).

Le script de la fonction est le suivant :

```
remove_outliers <- function(x) {
  qnt <- quantile(x, probs=c(.25, .75), na.rm = T)
  caps <- quantile(x, probs=c(.05, .95), na.rm = T)
  H <- 1.5 * IQR(x, na.rm = T)
  x[x < ( qnt[1]-H ) ] <- caps[1]
  x[x > ( qnt[2]+H ) ] <- caps[2]
  return(x)
}
```

Figure 45: Script de création de la fonction « remove_outliers » sous R

1.3.2.3. Suppression des variables qualitatives

Enfin, nous avons procédé à la suppression des variables qualitatives de notre data frame afin de n'appliquer l'algorithme que sur les variables quantitatives :

- Car il s'agit de faire une segmentation comportementale en premier lieu.
- Nous aurions pu quantifier ces variables qualitatives mais cela n'était pas nécessaire dans notre cas.

Ces variables seront réajoutées dans la partie analyse des données.

¹Script d'application de la fonction « remove_outliers » sur les variables quantitatives (Voir annexe 29).

1.3.2.4. Normalisation des données

Comme les variables quantitatives possèdent des grandeurs différentes, il était nécessaire de les mettre à la même échelle via la fonction « scale » de R. Ci-dessous le script correspondant :

```
### Suppression des variables qualitatives
datset_scaled <- subset( datset_treat, select = -c(TENURE, Value_Segment, DeviceType, SubscriberProfile))
### Mise à la meme échelle des variables
scaled_data = as.matrix(scale(datset_scaled))
```

Figure 46: Script de suppression des variables qualitatives et mise à la même échelle des variables quantitative

1.3.3. Application de l'algorithme

1.3.3.1. Détermination du nombre optimal de classes

Nous déterminons graphiquement le nombre optimal de classes à retenir, il s'agit de la valeur correspondant à la cassure (au coude) visualisée dans le graphe d'évolution de l'inertie inter-classes en fonction du nombre de classes K :

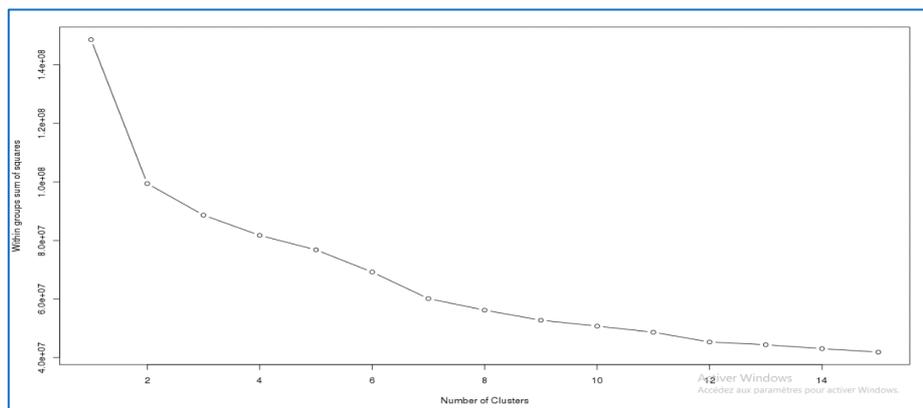


Figure 47: Elbow plot pour déterminer le nombre optimal de classes

On remarque sur le graphe une cassure graduelle qui se stabilise à partir du nombre 8.

Cependant nous avons retenu le nombre de classes égal à 10 pour la raison suivante :

- Après avoir appliqué l'algorithme k-means avec k=8, nous avons obtenu une classe ayant un effectif extrêmement élevé, plus de la moitié de l'effectif total appartenait qu'à une seule classe.
- En retenant le nombre k=10, il est évident d'obtenir une meilleure dispersion (Dispersion inter-classe réduite (voir définitions, page 35)). Ci-dessous le script correspondant :

```

### trouver le nombre optimal de K cluster

###Elbow Methode pour trouver le nombre optimal de classes k

wssplot <- function(scaled_data, nc=15, seed=500){
  wss <- (nrow(scaled_data)-1)*sum(apply(scaled_data,2,var))
  for (i in 2:nc){
    set.seed(seed)
    wss[i] <- sum(kmeans(scaled_data, centers=i)$withinss)
  }
  plot(1:nc, wss, type="b", xlab="Number of Clusters",
       ylab="Within groups sum of squares")
}

### affichage du graphe elbow

wssplot(datset_kmeans)
ggsave("wssplot")

```

« Figure 48: Script d'affichage du graphe « Elbow plot

1.3.3.2. Lancement de l'algorithme

Enfin, nous avons exécuté l'algorithme K-means avec le nombre de classes égal à 10 sur le data frame « datset_kmeans » représentant la table normalisée des données sur les variables quantitatives retenues.

Le temps d'exécution mis par l'algorithme pour afficher les résultats est d'environ **4 heures 30 minutes**.

Ci-dessous le script correspondant :

```

##### Exécution de l'algorithme Kmeans #####
datset_kmeans<-scaled_data
km.res<- kmeans(datset_kmeans,8,nstart=100)

```

Figure 49: Script d'exécution de l'algorithme K-means

1.4. Fouille et analyse des résultats

Dans cette dernière partie, en premier lieu, nous explorerons les résultats donnés par la fonction **K-means** sur **R** en affichant la liste de ses objets qui sont l'effectif et le centre de gravité de chaque classe, l'inertie intra et inter-classes par rapport aux variables quantitatives.

En second lieu, nous analyserons puis nous interprèterons ces résultats sur **Excel**.

1.4.1. Exploration des résultats ¹

L'exploration des résultats de l'algorithme consiste à afficher les mesures suivantes :

- **Les centres de gravité des classes :**Le centre d'une classe est représenté par des coordonnées (en nombre de 8, pour chaque variable), où variable est représentée par un axe spatial. Une coordonnée pour un **centre** = moyenne des valeurs de la variable correspondante dans une classe donnée. L'instruction pour afficher les centres des classes est la suivante : `km.res$centers`

¹ Voir définition inertie intra et inter-classe, page 35

- **L'inertie inter-classe :** Dispersion conditionnelle des barycentres des classes autour du barycentre global. C'est un indicateur de séparabilité des classes. L'instruction pour afficher l'inertie inter-classes est la suivante : `km.res$betweenss`
- **L'inertie intra-classe :** Dispersion à l'intérieur de chaque classe, indicateur de compacité des classes. L'instruction pour afficher l'inertie intra-classe est la suivante : `km.res$withinss`
- **L'inertie totale intra-classe :** C'est la somme des inerties intra-classe. L'instruction pour l'afficher est : `km.res$tot.withinss`
- **L'effectif :** C'est le nombre d'individus par classe. L'instruction pour l'afficher est : `km.res$size`

1.4.2. Analyse des résultats

- 1- Tout d'abord nous avons regroupé puis interprété par classesur Excel les résultats de l'algorithme dans les tableaux suivants :

Classes	% Effectif	% inertie intra-classe
1	1%	14%
2	5%	10%
3	48%	7%
4	18%	16%
5	1%	6%
6	3%	6%
7	10%	4%
8	5%	16%
9	4%	10%
10	6%	12%
Total	14 862 013 (100%)	100%

Tableau 19: % de l'effectif et inertie intra-classe de chaque classe

Classes/Centres	NBER_CALLS_M1	NBER_CALLS_ON_NET_M1	NBER_CALLS_OFF_NET_M1	NBER_CALLS_IN_M1
1	274.189997	212.767957	56.021550	592.71606
2	97.924085	59.077535	40.710382	118.38660
3	8.009095	5.353884	2.567834	16.44413
4	46.720668	29.919666	17.520030	97.16754
5	303.204338	239.755044	65.109508	594.59285
6	337.159853	262.315226	67.166846	258.73300
7	30.938886	18.663426	12.496599	25.74866
8	301.852309	259.585058	47.760536	162.75405
9	56.277317	36.321338	20.788390	263.99999
10	309.942087	260.842855	51.074118	92.71222

NBER_CALLS_Offnet_IN_M1	NBER_CALLS_INT_M1	NBER_SMS_ON_NET_M1	NBER_SMS_OFF_NET_M1	SMS_IN_M1	VOLUME_DATA
158.246687	0.09964234	8.460616	1.4603771	62.328801	0.3521757
37.214967	0.11438936	7.537285	1.6220396	34.066371	6843.0614258
5.089532	0.01837001	0.982165	0.2087194	7.322084	0.1905124
29.619551	0.04984818	3.098787	0.5740750	20.037218	0.3241824
153.713058	0.15686345	12.590800	2.5538612	75.125853	6843.0614258
77.480984	0.17010643	11.238783	2.2856580	59.957680	6843.0614257
8.616167	0.08820007	3.477736	0.8210572	15.377777	6843.0614260
47.395888	0.11059052	6.434302	0.9681215	43.518712	0.3823431
61.074139	0.04347928	5.213346	0.8002600	33.292958	0.3605736
27.753959	0.16431319	9.379671	1.7066934	40.616145	6843.0614258

Tableau 20: Coordonnées des centres de gravité des classes

Nous lisons le tableau **18** comme suit :

Par exemple pour la classe **5**, les abonnés de celle-ci ont émis en moyenne pour le mois de mars, **303** appels dont **230** vers djezzy et **65** vers les autres opérateurs et ont reçus **595** appels dont **153** des autres opérateurs. Ils ont consommé en moyenne **6.8 Gb** d'internet.

Inertie inter-classe : Elle vaut 99,97% de l'inertie totale, par conséquent l'inertie intra classe c'est-à-dire la dispersion entre les individus dans une même classe est très faible et vaut 0.03%, les classes sont donc très distinctes.

- **Classe 3 (48%)** : Cette classe semble contenir les abonnés ayant fait le moins de trafic durant le premier mois (Mars) en termes de : Nombre d'appels, nombre de sms et de consommation en données mobiles.

- **Classe 5 et 6 (1% 3%)** : Ces classes semblent contenir les abonnés ayant fait le plus de trafic sortant durant le premier mois en termes de : Nombre d'appels, nombre de sms et de consommation en données mobiles.
- **Classe 10 (6%)** : Cette classe semble contenir des abonnés utilisateurs des données mobiles ayant un trafic sortant vers Djezzy important.
- **Classe 4 (18%)** : Cette classe semble dominée par des abonnés non utilisateurs des données mobiles ayant un faible trafic appels et sms.
- **Classe 9 (4%)** : Cette classe semble contenir des abonnés ayant un trafic entrant de Djezzy important (in) et une très faible consommation en données mobiles.
- **Classes 8 et 1 (5%) et (1%)** : Cette classe semble être dominée par des abonnés non utilisateurs en données mobiles ayant un fort trafic en termes d'appels émis vers Djezzy.
- **Classe 7 (10%)** : Cette classe semble contenir les abonnés dont le trafic est dominé par l'utilisation des données mobiles.
- **Classe 2 (5%)** : Cette classe semble être dominée par les utilisateurs en données mobiles.

2- Afin de faire ressortir des profils d'abonnés distingués par leur comportement (trafic réseau mesuré par les variables quantitatives) et par leur type en termes d'ancienneté, de modèle de téléphone utilisé, et d'offres souscrites (mesurés par les variables qualitatives), nous avons effectué une analyse croisée entre ces deux types de variables. Pour ce faire nous devons récupérer les données sur les variables qualitatives à partir de la table « ADS_Kmeans » sur Teradata en procédant comme suit :

a- Affectation des abonnés aux classes

En créant un data frame nommé « dataset_final » sur **R** ayant pour colonnes « Subs_id » et « Cluster_num » dont le script est le suivant :

```
dataset_final<-NULL
dataset_final$Subs_id<-dataset$Subs_id
dataset_final$cluster_id=km.res$cluster
```

Figure 50: Script d'affectation des abonnés à une classe

b- Export des données du dataframe vers Teradata

Pour réaliser cet export nous avons :

- Transformé le data frame en fichier Comma-Separated-Values (CSV) nommé « dataset_final » comme suit :

```
write.csv2(datset_final, file = "Kmeans/datset_final.csv", sep = "|", row.names = FALSE, col.names = FALSE)
df_final <- read.csv2(file = "Kmeans/datset_final.csv")
```

Figure 51: Script de création d'un fichier CSV contenant les valeurs de la table `datset_final`

- Créé une table vide sur Teradata nommée « Clustering_Kmeans » ayant pour attributs « Subs_id » et « Cluster_num » dont le script est le suivant :

```
create multiset table p_edw_tmp.Clustering_Kmeans
(
  Subs_id integer
  ,Cluster_num integer
)
no primary index
```

Figure 52: Script de création de la table « Clustering_Kmeans » sur Teradata

- Chargé la table « Clustering_Kmeans » à partir du fichier CSV « datset_final » en utilisant les outils :

BTEQ : Pour supprimer les tables d'erreur engendrées par l'outil FastLoad. Ci-dessous le script correspondant :

```
.LOGON 172.16.53.41/
DATABASE p_edw_tmp;
drop table p_edw_tmp.km_ET;
drop table p_edw_tmp.km_UV ;
.IF ERRORCODE <> 0 THEN .EXIT ERRORCODE;
.LOGOFF;
```

Figure 53: Script BTEQ pour la suppression des tables d'erreurs

FastLoad : Pour charger les données dans la table « Clustering_Kmeans » :

```
LOGON 172.16.53.41/
DATABASE p_edw_TMP;
BEGIN LOADING p_edw_tmp.Clustering_Kmeans
ERRORFILES km_ET, km_UV
CHECKPOINT 10000;
SET RECORD VARTEXT "|";
DEFINE Subs_id (varchar(20)),
cluster_num (varchar(15)),
FILE = C:\Users\ \Desktop\FFE_MOUSSRAR\Clustering\datset_final.csv;
INSERT INTO Clustering_Kmeans (
  Subs_id ,
  Cluster_num
)
VALUES (
  :Subs_id ,
  :Cluster_num
);
END LOADING;
LOGOFF;
```

Figure 54: Script Fastload pour la table « clustering_kmeans

Fichier.Bat : Pour l'exécution des deux précédents scripts

```

echo "-----drop create-----"
bteq < C:\Users\... Desktop\FFE_MOUSSRAR\Clustering\bteq_km.sql >C:\Users\... Desktop\FFE_MOUSSRAR\Clustering\bteq_km.log
echo "-----fast load -----"
FastLoad < C:\Users\... Desktop\FFE_MOUSSRAR\Clustering\dataset_final.fl >C:\Users\... Desktop\FFE_MOUSSRAR\Clustering\dataset_final.log
pause
    
```

Figure 55: Script du fichier.Bat exécutant le BTEQ et le Fastload

- Rajouté l'attribut « Cluster_num » de la table « Clustering_kmeans » à la table ADS_Kmeans via la clé commune « Subs_id » comme suit :

```

alter TABLE p_edw_tmp.ADS_Kmeans add Cluster_num integer;
update a
from p_edw_tmp.ADS_Kmeans a ,p_edw_tmp.Clustering_Kmeans b
set cluster_num = b.cluster_num
where a.subs_id=b.subs_id ;
    
```

Figure 56: Script de rajout de l'attribut Cluster_num à la table ADS_Kmeans

c- Analyse des données sur Excel du nouvel ADS_Kmeans

Nous avons effectué l'analyse des variables qualitatives par classe d'abonnés via les tableaux croisés dynamiques sur Excel. Puis, pour renforcer cette analyse nous avons calculé en % la moyenne des valeurs de chaque variable quantitative pour chaque classe également. Les résultats sont présentés dans les tableaux et graphiques ci-dessous :

Somme de Subs	Étiquettes de colonnes										Total général
Étiquettes de lignes	1	2	3	4	5	6	7	8	9	10	Total général
A - VHV	14,23%	3,39%	0,10%	0,39%	16,41%	8,68%	1,87%	6,58%	0,52%	3,82%	1,47%
B - HV	22,69%	10,80%	0,89%	2,33%	23,55%	19,79%	6,55%	18,03%	3,60%	12,19%	4,59%
C - MV	26,18%	21,81%	2,70%	6,58%	28,35%	33,97%	13,74%	35,86%	8,69%	31,43%	10,08%
D - LV	16,84%	22,99%	5,07%	18,83%	15,33%	19,12%	18,06%	19,80%	21,39%	24,49%	12,78%
E - VLV	17,80%	28,97%	50,54%	62,17%	12,82%	12,07%	35,68%	14,73%	58,93%	16,66%	45,35%
F - ZERO	2,26%	12,04%	40,71%	9,70%	3,53%	6,37%	24,10%	5,00%	6,86%	11,41%	25,72%
Total général	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Tableau 21: Les valeurs en % de la variable « Segment_value » pour chaque classe

Étiquettes de lignes	1	2	3	4	5	6	7	8	9	10	Total général
ALLO	0,41%	0,07%	0,17%	0,49%	0,03%	0,01%	0,04%	0,33%	0,49%	0,01%	0,22%
B2csp1prep	0,00%	0,00%	0,15%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,07%
Distributorprep	0,02%	0,00%	0,37%	0,01%	0,00%	0,00%	0,00%	0,00%	0,01%	0,00%	0,18%
Djezzy BAYNA	0,05%	0,03%	0,09%	0,10%	0,01%	0,00%	0,05%	0,03%	0,11%	0,00%	0,08%
Djezzy Carte	4,84%	0,68%	3,66%	5,13%	0,38%	0,15%	0,87%	2,92%	4,61%	0,05%	3,20%
Djezzy Carte New	60,62%	29,02%	31,05%	39,05%	43,85%	36,10%	20,71%	62,20%	43,64%	25,42%	33,51%
Djezzy GO	5,15%	2,62%	15,03%	16,50%	0,80%	0,28%	3,82%	5,03%	12,99%	0,27%	11,62%
Djezzy GOOD	6,16%	2,22%	8,44%	17,61%	0,54%	0,21%	2,27%	4,77%	14,85%	0,14%	8,54%
Djezzy Hadra	7,25%	5,63%	7,05%	4,62%	7,62%	4,16%	5,33%	6,43%	6,18%	3,32%	6,01%
Djezzy HAYLA	4,69%	20,66%	14,05%	5,66%	13,45%	17,53%	29,78%	6,29%	5,80%	27,72%	14,46%
Djezzy HAYLABEZZEF	7,98%	32,66%	15,43%	7,31%	29,78%	36,91%	29,26%	9,48%	7,77%	37,24%	17,43%
Djezzy PLAY	2,46%	2,88%	2,65%	3,20%	1,99%	1,70%	1,82%	1,95%	3,16%	0,82%	2,53%
Mifi	0,00%	0,00%	0,49%	0,00%	0,00%	0,00%	1,42%	0,00%	0,00%	0,00%	0,37%
Mixteprepaid	0,36%	3,53%	1,30%	0,34%	1,54%	2,94%	4,11%	0,57%	0,39%	5,03%	1,69%
Prepaiddjezzyinternet	0,00%	0,01%	0,07%	0,01%	0,00%	0,00%	0,52%	0,00%	0,01%	0,00%	0,09%
Total général	100,00%										

Tableau 22: Les valeurs en % de la variable « SubscriberProfile » pour chaque classe

Nous remarquons que le profil Djezzy Carte New est dominant pour toutes les classes, car celui-ci représente le profil le plus ancien de Djezzy qui englobe un grand nombre d'offres

Étiquettes de lignes	1	2	3	4	5	6	7	8	9	10	Total général
0	7,77%	6,05%	6,32%	6,76%	7,42%	6,24%	4,81%	8,20%	7,48%	4,88%	6,31%
> 5 Year	51,15%	24,56%	35,39%	57,83%	25,65%	24,96%	16,16%	50,74%	52,05%	15,94%	37,18%
>12M and <= 24 M	11,31%	10,40%	10,05%	9,20%	14,26%	13,04%	8,20%	11,27%	10,38%	10,71%	9,96%
>3 M and <= 6 M	4,64%	9,88%	7,03%	3,60%	10,63%	9,62%	9,58%	3,54%	4,97%	10,37%	6,81%
>3 Year and <= 5 Year	10,86%	6,78%	8,55%	10,63%	8,22%	7,16%	5,18%	11,31%	11,21%	5,40%	8,55%
>6 M and <= 12 M	8,22%	19,79%	10,91%	6,84%	20,60%	25,12%	16,68%	8,95%	7,27%	26,68%	12,27%
New_subs	6,06%	22,54%	21,75%	5,14%	13,21%	13,87%	39,40%	5,98%	6,64%	26,02%	18,92%
Total général	100,00%										

de souscription.

Tableau 23: Les valeurs en % de la variable « Tenure » pour chaque classe

Étiquettes de lignes	1	2	3	4	5	6	7	8	9	10	Total général
BASIC PHONE	69,05%	6,62%	34,35%	51,64%	13,66%	9,37%	4,88%	66,26%	52,70%	6,83%	33,40%
FEATURE PHONE	11,49%	2,76%	10,80%	13,24%	2,98%	2,34%	2,62%	12,97%	12,28%	2,09%	9,44%
OTHER	1,26%	8,08%	12,07%	2,09%	1,49%	2,37%	23,63%	2,29%	1,81%	8,76%	9,65%
SMARTPHONE	18,20%	82,54%	42,77%	33,02%	81,87%	85,92%	68,87%	18,48%	33,21%	82,32%	47,51%
Total général	100,00%										

Tableau 24: Les valeurs en % de la variable « Device_type » pour chaque classe

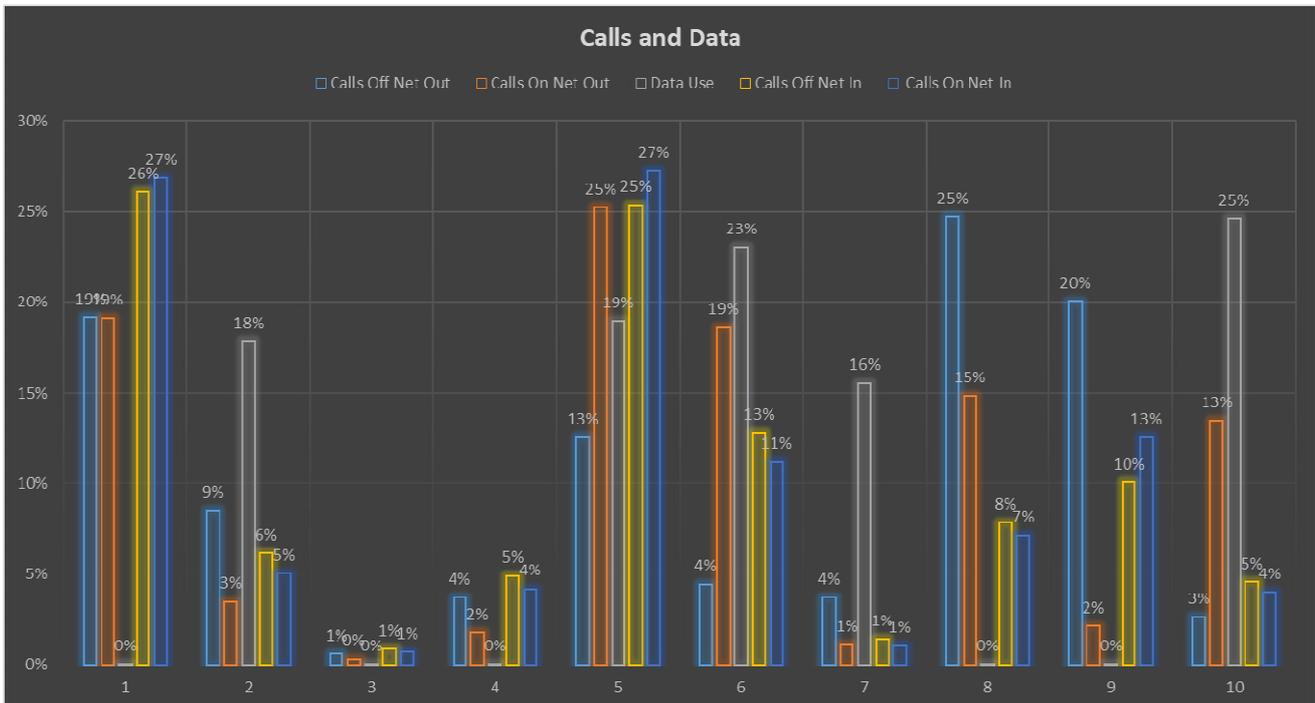


Figure 57: Histogrammes en % du trafic appels et données mobiles par classe

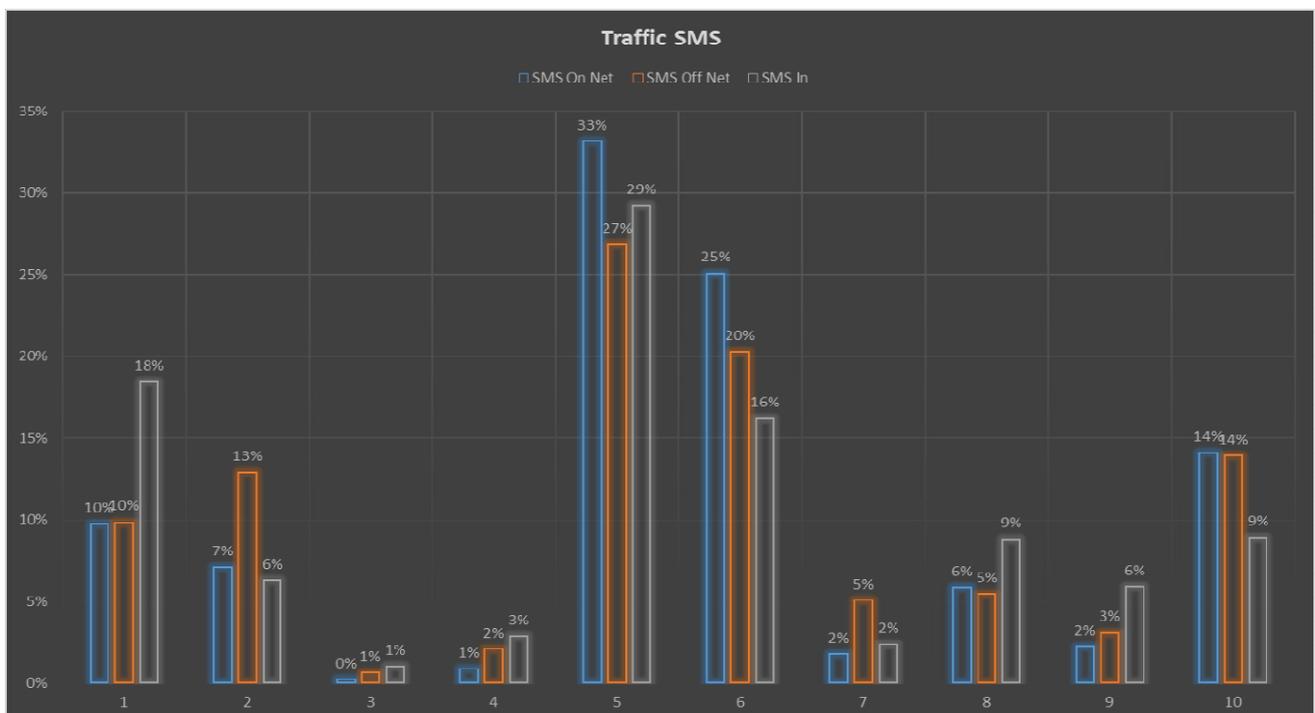


Figure 58: Histogrammes en % du trafic sms on net, off net et incoming par classe

Interprétation des résultats

Classe 3« *Mass Sleeping Customers* »: Cette classe contient les abonnés ayant fait le moins de trafic durant le mois de Mars en termes de nombre d'appels, de sms et de consommation en données mobiles

- Effectif : 48 %
- 1/3 (35%) des abonnés de cette classe ont plus de 5 ans
- 42% des abonnés de cette classe ont moins d'une année
- Ces mêmes %s sont retrouvés pour le type de téléphone 34 % basic phone et 43% smartphone
- 97% des abonnés sont des LV <1000 DA
- Consommation moyenne en données mobiles : 0,19 MB
- Nombre moyen des appels : 8
- Nombre moyen des SMS émis : 0,75

Classes 5 et 6« *Phone Addicts* »: Ces classes contiennent les abonnés ayant fait le plus de trafic sortant durant le mois de Mars en termes de nombre d'appels, d'sms et de consommation en données mobiles.

Bien qu'il y ait des différences dans le nombre d'appels et de sms émis et reçus dans les deux classes, nous avons fait le choix de les regrouper dans une seule classe, car nous jugeons que le comportement des individus de chacune d'elle est le même dans la souscription de l'offre, de plus le fait que les variables qualitatives sont très homogènes a renforcé notre choix.

- Effectif : 4%
- 48% des abonnés ont moins d'un an
- 85% des abonnés ont des Smartphone
- Les offres dominantes sont HAYLA et HAYLA BEZZAF
- 65% des abonnés sont des HV >=1500 DA
- Consommation moyenne en données mobiles : 4,8 GB
- Nombre moyen des appels : 427
- Nombre moyen des SMS émis : 70

Classe 10« *Data lovers and On Net Talkers* »: Cette classe contient les abonnés ayant un trafic élevé d'appels émis vers Djazzy et un trafic sms moyen et data-user.

- Effectif : 6 %
- 82% des abonnés ont un smartphone

- Dominés par HAYLA ET HAYLABEZZEF
- 48% sont des HV
- Consommation moyenne en données mobiles : 5,7 GB
- Nombre moyen des appels : 285
- Nombre moyen des appels émis vers Djezzy : 230
- Nombre moyen des SMS émis : 38

Classe4 « *Traditional Customers* »: Cette classe est dominée par les abonnés non utilisateur des données mobiles, ayant un faible trafic appels et sms.

- Effectif : 18 %
- 60% des abonnés de cette classe ont plus de 5 ans
- 70% des abonnés de cette classe n'ont pas de smartphone
- 90 % des abonnés de cette classe sont des LV (<=1000 DA)
- Les offres sont dominées par DJEZZY GOOD (1000 DA), DJEZZY GO (400 DA)
- Consommation moyenne en données mobiles : 0,32 MB
- Nombre moyen des appels émis : 46
- Nombre moyen des SMS émis : 2

Classe 9 « *Traditional In and Out talkers* » : Cette classe contient les abonnés Non utilisateur des données mobiles data, ayant un fort trafic appels sortant vers les autres opérateurs et un trafic élevé in des appels (point de différence avec la classe 4), et un faible trafic sms.

- Effectif : 4 %
- 53% des abonnés ont plus de 5 ans
- 67% des abonnés n'ont pas de smartphone
- Dominée par les offres DJEZZY GO et DJEZZY GOOD
- 88% des abonnés sont des LV
- Consommation moyenne en données mobiles : 0,36 MB
- Nombre moyen des appels émis : 56
- Nombre moyen des appels reçus : 264
- Nombre moyen des SMS émis : 6

Classe 1« *Calls and SMS loyalists* »:Cette classe contient les abonnés Non user data, fort trafic des appels (In et Out), un moyen trafic sms Out et un trafic sms In élevé

- Effectif : 1%
- 51% des abonnés ont plus de 5 ans
- L'offre dominante est DJEZZY CARTE NEW
- 62% dépensent plus de 1500 DA

- 82% des abonnés n'ont pas de smartphone
- Consommation moyenne en données mobiles : 0,35 MB
- Nombre moyen des appels émis : 416
- Nombre moyen des appels reçus : 592
- Nombre moyen des SMS émis : 26

Classes 8 « *Traditional Out Talkers* »: Cette classe est dominée par les abonnés à fort trafic out, plus ce sont des non-user data avec un faible trafic sms

- Effectif : 5%
- 51% des abonnés ont plus de 5 ans
- L'offre dominante est DJEZZY CARTE NEW
- 62% dépensent plus de 1500 DA
- 82% des abonnés n'ont pas de smartphone
- Consommation moyenne en données mobiles : 0,38 MB
- Nombre moyen des appels émis : 310
- Nombre moyen des appels reçus : 162
- Nombre moyen des SMS émis : 16

Classe 7 « *Data enthusiasts* » : Cette classe contient les abonnés dont le trafic est dominé par l'utilisation des données mobiles.

- Effectif : 10 %
- 65% des abonnés ont moins d'1 an
- 80% des abonnés sont des LV (44% LV)
- L'offre dominante est HAYLA ET HAYLABEZZAF
- 70% des abonnés ont des smartphones
- Consommation moyenne en données mobiles : 3,6 GB
- Nombre moyen des appels émis : 30
- Nombre moyen des appels reçus : 25
- Nombre moyen des SMS émis : 5

Classe 2 « *Data and SMS Likers* »: Cette classe est dominée par les abonnés utilisateur des données mobiles ayant un trafic appels entrant et sortant moyen, et un fort trafic sms sortant vers les autres opérateurs.

- Effectif : 5%
- 83% des abonnés ont des smartphones
- 64% sont des LV
- L'offre dominante est HAYLA BEZZEF

- L'âge des abonnées est diversifié
- Consommation moyenne en données mobiles : 4,1 GB
- Nombre moyen des appels émis : 97
- Nombre moyen des appels reçus : 118
- Nombre moyen des SMS émis : 22

Remarque:

- Nous avons regroupé les classes 5 et 6 car leurs individus présentent les mêmes caractéristiques :

```
----- Regroupement des classes 5 et 6, 9 classes au total-----
update p_edw_tmp.clustering_kmeans
set cluster_num=cluster_num-1
where cluster_num>5
```

Figure 59: Script de regroupement des classes 5 et 6

- Nous avons créé et chargé une table nommée « Traffic_Behavior » contenant le nom, la description et les caractéristiques de chaque classe (Voir annexes 30 et 31).

1.4.3. Impact financier de la segmentation

Après avoir analysé les résultats obtenus, nous allons chiffrer grâce aux données avancées par le service CVM, les éventuels gains lors d'un ciblage par classe d'abonnés via les campagnes directes.

Valeur abonné	Taux de réponse avec ciblage sans segmentation (par individu)	Taux de réponse avec ciblage avec segmentation
Low Value (LV)	2 %	5 %
Medium Value (MV)	4 %	10 %
High Value (HV)	11 %	20 %

Tableau 25: Taux de réponse sans et avec segmentation par valeur d'abonné

Ces données nous ont été communiquées par le service CVM après avoir effectués une segmentation des abonnés. Nous les avons repris pour mettre en avant l'impact financier d'un ciblage par classe d'abonnés.

Classe	Effectif	Revenu mensuel moyen	Cible	Réponse avant la segmentation	Réponse après la segmentation	Gain (DZA)
1	121344	1 476,01	12134,4	1334,784	2426,88	1 611 944,62
2	744877	842,40	74487,7	2979,508	7448,77	3 764 906,31
3	7161999	158,96	716199,9	14323,998	35809,995	3 415 414,08
4	2722627	401,04	272262,7	5445,254	13613,135	3 275 647,00
5	5105883	1 357,74	510588,3	56164,713	102117,66	62 392 154,26
6	1431949	573,14	143194,9	5727,796	14319,49	4 924 243,50
7	678295	1 205,12	67829,5	7461,245	13565,9	7 356 841,83
8	636993	468,09	63699,3	1273,986	3184,965	894 510,16
9	853346	1 000,71	85334,6	9386,806	17066,92	7 685 566,88

Tableau 26: Gain en DZA lors d'un ciblage par segments

Effectif : l'effectif de la classe.

Revenu mensuel moyen : Revenu moyen de la classe d'abonné pour le mois de Mars 2019.

Cible : 10 % de l'effectif ciblé par une campagne (Taux variable selon le cout de la campagne).

Taux de réponse sans segmentation : taux de réponse de l'effectif ciblé avant segmentation.

Taux de réponse avec segmentation : taux de réponse de l'effectif ciblé après segmentation.

Gain : gain généré par le bon ciblage (segmenté).

Nous constatons qu'un ciblage des campagnes par segment d'abonnés génère un bénéfice non négligeable. Celui-ci représente en moyenne 1.18 % du chiffre d'affaire annuel de Djezzy avec comme hypothèse la cible et les taux de réponses retenus.

2. Création d'un tableau de bord

Dans cette dernière partie, nous avons créé un tableau de bord qui met en avant les indicateurs clés qui sont consultés régulièrement par l'utilisateur final lui facilitant la prise de décision. Il regroupe différents types d'indicateurs relatifs au trafic des abonnés prépayés.

Il a été réalisé sur le logiciel tableau, qui était un choix de notre part, car il permet non seulement de concevoir dynamiquement des tableaux de bord, mais aussi pour les avantages suivants:

- Une grande facilité d'utilisation.
- Une analyse 10 à 100 fois plus rapide que la concurrence.
- Des tableaux de bords astucieux, compacts et faciles à comprendre.
- Une utilisation immédiate et sans contraintes techniques.
- Plusieurs sources de données combinées au sein d'un même document.
- Un partage sécurisé des analyses sur le web.

Nous présenterons tout d'abord le logiciel tableau, ensuite la réalisation de notre tableau de bord sur celui-ci.

2.1. Tableau desktop

Tableau desktop est un des produits proposés par l'entreprise tableau permettant la création de tableaux de bord interactifs :

Il permet la connexion à des données sur site ou dans le cloud, que ce soit le Big Data, une base de données SQL, une feuille de calcul ou des applications dans le cloud comme Google Analytics ou Salesforce. Les utilisateurs confirmés peuvent permuter, décomposer et gérer des métadonnées afin d'optimiser les sources de données.

Il gère rapidement des calculs très puissants à partir de données existantes, en glissant et en déposant des lignes de référence.

Il permet aussi de Créer automatiquement des cartes interactives. Les codes postaux intégrés permettent de cartographier instantanément plus de 50 pays du monde entier (Fonctionnalités, s.d.)

2.2. Réalisation du tableau de bord

Afin de réaliser notre tableau de bord dynamique, nous avons :

Tout d'abord, effectué une connexion à notre base de données de Teradata afin d'exporter les données de notre DM pour le calcul et l'affichage des indicateurs clés retenus. Ensuite, pour construire les divers graphiques sur des feuilles de calcul, nous avons utilisé ses tables soit directement, soit en effectuant, pour des fins d'optimisation de calcul, des requêtes SQL

personnalisées implémentées sur tableau. Enfin, nous avons regroupé tous les graphiques des feuilles de calcul dans un tableau de bord.

Le tableau de bord regroupe les KPI suivants :

- **Number of Subscribers per Profile** : nombre d'abonnés par profil de l'offre.
- **Daily Revenue** : le total du revenu journalier.
- **Monthly Refill by Brand** : revenu sur le rechargement par type de celui-ci (Flexy et voucher)
- **% Revenue per profile** : % du revenu par profil de l'offre.
- **Revenue and % of subscribers by wilaya** : revenue et % du nombre d'abonnés par wilaya.
- **% Average revenue per device type** : le revenu moyen par type de téléphone.
- **Evolution of subscription revenue** : évolution du revenu des souscriptions.
- **Traffic Outgoing** : revenu sur le trafic sortant (appels, sms et données mobiles).
- **Data usage evolution** : évolution de l'utilisation des données mobiles.

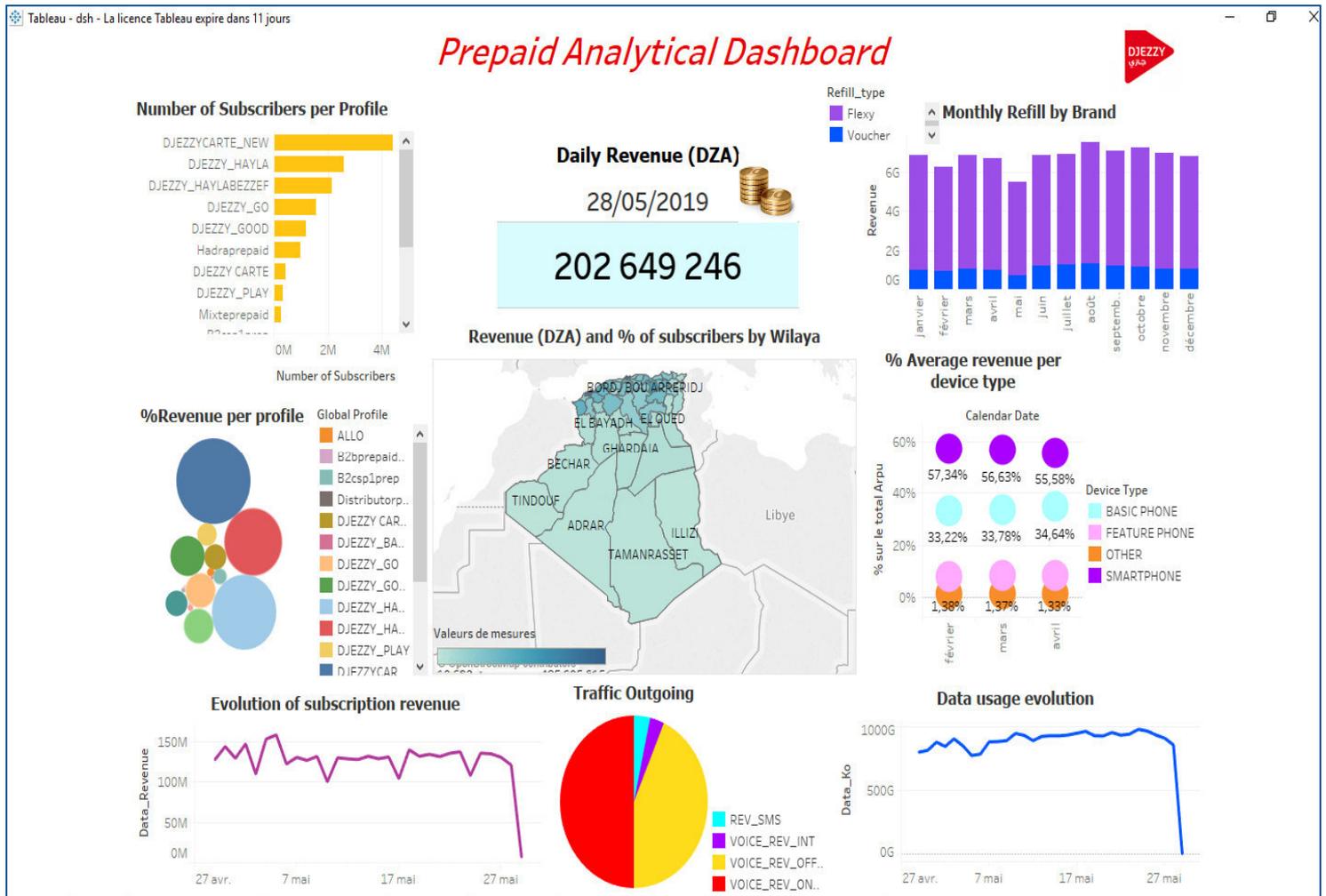


Figure 60: Tableau de bord créé sur tableau software

Nous pouvons remarquer la décroissance subite des deux graphes de « Evolution of subscription revenue » et « Data usage evolution », ceci s’explique par le fait que le chargement des données s’effectue à j-1 et qu’à la date du 28 mai 2019, date de réalisation des graphiques du tableau de bord les données du 27 mai n’avaient pas encore fini de charger.

Conclusion

Ce mémoire de fin d'étude a eu pour objectif de répondre au besoin du service CVM qui consiste à automatiser le calcul des indicateurs clés de performance (KPI) permettant de guider les décideurs vers un ciblage précis et efficace des abonnés prépayés via des campagnes marketing performantes favorisant les retours sur investissements.

Pour ce faire, nous avons proposé en premier lieu de concevoir un DM décisionnel pour ce même service regroupant l'ensemble des données nécessaires pour le calcul des KPI. En second lieu, nous avons regroupé ces KPI dans une table de données ordonnée par abonné prépayé et mise à jour, appelé « Analytical Dataset » (ADS). Enfin, pour automatiser le calcul des KPI nous avons créé un fichier de type « Bat » qui regroupe les syntaxes d'exécution des macros de chargement de l'ADS. Le fichier Bat en question sera exécuté automatiquement après création d'une tâche planifiée via l'option « Planificateur de tâche » de Windows.

Bien que nous ayons répondu au principal besoin du service CVM qui est l'automatisation du calcul des KPI par la conception du DM puis de l'ADS, nous avons voulu par la suite exploiter ces KPI pour des fins d'analyse permettant d'optimiser le ciblage des campagnes de marketing direct. Ceci en réalisant non seulement une segmentation comportementale via l'algorithme k-means dont l'objectif est d'affecter chaque abonné prépayé à une classe d'appartenance, selon le trafic appels, sms et de consommation en données mobiles, mais aussi en visualisant ces indicateurs clés dans un tableau de bord dynamique.

A cet égard, notre projet de fin d'études inclus toutes les phases d'un projet BI, de l'extraction à l'analyse de données. En effet, la première phase qui est l'alimentation en données est intervenue lors du chargement des tables du DM en utilisant des outils d'export et de chargement des données du SGBD Teradata. La deuxième consiste en la modélisation du DM sur ce même SGBD. La dernière phase qui est la phase d'analyse, englobe l'utilisation d'un algorithme de data mining lors de la segmentation des abonnés, et un tableau de bord analytique qui met en avant les indicateurs clés facilitant ainsi la prise de décision.

Pour ce qui est de la valeur ajoutée de notre projet, les utilisateurs business peuvent désormais créer en quelques clics leurs propres rapports à partir du DM implémenté sur BO et cela sans passer par un intermédiaire expert qui prend un temps nettement supérieur, facilitant ainsi l'accès à l'information pour la prise de décision. Concernant le service CVM, le DM facilite l'accès aux données qui sont désormais toutes disponibles dans une même source réduisant le temps de lancement d'une campagne qui est passé approximativement de 4 à 1 jour.

En ce qui concerne la segmentation, nous avons démontré l'impact financier positif d'un ciblage par classe d'abonnés. Celui-ci génère un bénéfice non négligeable qui représente 1.18 % du chiffre d'affaire annuel de Djezzy avec comme hypothèse les valeurs connues d'avance du taux de réponse et du nombre d'abonnés ciblés par classe.

Outre les résultats obtenus, ce stage a été pour nous l'occasion d'améliorer nos connaissances en gestion de base de données, en analyse et en modélisation des données. Autre acquis très précieux est celui de la manière de mener un projet, d'enquêter et de tout mettre en œuvre pour valider une hypothèse. Mais comme tout autre projet nous avons rencontré de nombreuses contraintes, notamment techniques lors de l'exécution des scripts informatiques et lors du chargement de la très grande masse de données.

Bibliographie

Ouvrages :

- BENMALEK, K. (2018). *Segmentation des abonnés de Djezzy selon leurs comportements*. Alger.
- CROZAT, S. (2017). *Introduction aux data warehouses : la modélisation en étoile*. UTC formation.
- Gomes, K. (2015). *State of the art on Business Intelligence*, 12-34. Consulté le juin 4, 2019.
- Inmon, B. (2005). *Building the data warehouse*. Wiley.
- Kimball, R., & Ross, M. (2010). *The Kimball Group Reader*. Indianapolis: Wiley Publishing.
- Marr, B. (2012). *The 75 measures every manager needs to know*. Pearson.
- Martin, A. (2011). *Better Decision Making with*. Chicago: A.T. Kearney Inc.
- R, K., & Ross M. (2008). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* (Vol. Second Edition). Wiley Publishing.
- Sherman, R. (2015). *Business Intelligence Guidebook From Data Integration to Analytics*. (S. Elliot, Éd.) Waltham, USA: ELSEVIER. Consulté le mars 2019
- Varga, M. (2001). *On the Differences of Relational and Dimensional Data Model*. Zagreb.
- Ziafat, H., & Shakeri, M. (2014). Using Data Mining Techniques in Customer Segmentation. *Journal of Engineering Research and Applications*, 70-79.

Sites Internet :

- Brisson, L. (2016). *Business Intelligence Panorama*. Consulté le juin 2019, sur formations.telecom-bretagne: <http://formations.telecom-bretagne.eu/bi/presentations/panorama/#1>
- C.Desrosiers. (2011). *Entrepôts de données et intelligence des affaires, Modélisation Dimensionnelle*. (Hiver, Éd.) Consulté le juin 2019, sur ETS le génie pour l'industrie: https://cours.etsmtl.ca/mti820/public_docs/acetates/MTI820-Acetates-ModelisationDimensionnelle
- Alton, L. (2017, Décembre 22). *The 7 Most Important Data Mining Techniques*. Récupéré sur Data Science Central: www.datasciencecentral.com

- Banguy, Z. (2016, septembre 15). *2 approches pour construire un Data Warehouse (DWH)*.
Récupéré sur Aerow, Experts you need: <https://www.aerow.group>
- Benzaki, Y. (2018, Avril 10). *Tout ce que vous voulez savoir sur l'algorithme K-Means*.
Récupéré sur Mr.Mint: <https://mrmint.fr>
- campagne-marketing*. (2016). Consulté le Avril 2019, sur abileo:
<https://www.abileo.com/campagne-marketing/>
- Catalogue d'interconnexion OTA*. (2015, juin). Consulté le Avril 2019, sur Djezzy:
<http://www.djezzy.dz/>
- Différence entre le schéma en étoile et en flocons*. (2018, Aout 4). Récupéré sur
WayToLearnX: <https://waytolearnx.com>
- EDUCBA (Éd.). (2019). *What's better than a Direct Marketing or Indirect Marketing*.
Consulté le juin 2, 2019, sur EDUCBA: <https://www.educba.com/direct-marketing-vs-indirect-marketing/>
- Fonctionnalités*. (s.d.). Consulté le Mai 2019, sur tableau: <https://www.tableau.com/fr-fr/products/desktop>
- GONZALEZ, P.-L. (2016). *Méthodes de classification*. Consulté le juin 2019, sur
[maths.cnam: http://maths.cnam.fr](http://maths.cnam.fr)
- L, B. (2018, Mars 18). *Définition d'une technologie d'analyse multidimensionnelle*. Récupéré
sur Le big data: <https://www.lebigdata.fr/olap-online-analytical-processing>
- Prabhakaran, S. (2016, Décembre 6). *Outlier detection and treatment with R*. Récupéré sur r-
bloggers: <https://www.r-bloggers.com/outlier-detection-and-treatment-with-r/>
- Rakotomalala, R. (2014). *Méthode des centres mobiles*. Lyon, France.
- Sfaxi, D. L. (2014). *Business Intelligence, modélisation des données décisionnelles*. Consulté
le avril 2019, sur Slidashare: <https://fr.slideshare.net/LiliaSfaxi/chp3-modlisation-multidimensionnelle>
- Tableau de bord (gestion)*. (2019, Avril 16). Récupéré sur Wikipedia:
[https://fr.wikipedia.org/wiki/Tableau_de_bord_\(gestion\)](https://fr.wikipedia.org/wiki/Tableau_de_bord_(gestion))
- Temgoua, S. (2005, Novembre 5). *Différence entre tableau de bord et reporting (ou rapport)*.
Récupéré sur TSS performance: <https://tssperformance.com/difference-entre-tableau-de-bord-et-reporting-ou-rapport/>
- Teradata Tutorial*. (2019). Récupéré sur tutorialspoint: www.tutorialspoint.com

Annexes

Annexe 1: Dictionnaire de donnée de la dimension date.....	i
Annexe 2: dictionnaire de donnée de la dimension offre.....	i
Annexe 3: Dictionnaire de donnée de la dimension type de rechargement.....	ii
Annexe 4: Dictionnaire de donnée de la dimension direction d'appels.....	ii
Annexe 5: Dictionnaire de donnée de la dimension relative à l'activité des antennes réseau Djazzy.....	ii
Annexe 6: Dictionnaire de donnée de la dimension souscripteur.....	iii
Annexe 7: Dictionnaire de donnée de la dimension relative au téléphone associé à la SIMCARD Djazzy.....	iv
Annexe 8: Dimension relative aux transactions par destination.....	iv
Annexe 9: Dictionnaire de donnée de la dimension relative à la localisation des antennes réseau Djazzy.....	v
Annexe 10 : Dictionnaire de donnée de la table de fait rechargement.....	v
Annexe 11: Dictionnaire de donnée de la table de fait trafic.....	vi
Annexe 13: Modèle physique et script de création de la dimension offre.....	vii
Annexe 14: Modèle physique et script de création de la dimension cell_area.....	viii
Annexe 15 : Modèle physique et script de création de la dimension handset.....	viii
Annexe 16: Le script du TPT de la dim_cell_loc.....	x
Annexe 17: Script de la création de la table temporaire « ADS_Outgoing_Month ».....	xii
Annexe 18: Script de création de la macro « LOAD_ADS_OUTGOING_WEEK » de chargement de table temporaire « ADS_Outgoing_Week » pour la catégorie de KPI outgoing.....	xv
Annexe 19: Script de la création de la table temporaire « ADS_Outgoing_Week ».....	xvii
Annexe 20: Script de création de la macro « LOAD_ADS_OUTGOING_MONTH » de chargement de table temporaire « ADS_Outgoing_Month » pour la catégorie de KPI traffic_outgoing.....	xix
Annexe 21: Script de la création de la table temporaire « ADS_INCOMING ».....	xx
Annexe 22: Script de création de la macro « LOAD_ADS_INCOMING » de chargement de table temporaire « ADS_INCOMING » pour la catégorie de KPI traffic_incoming.....	xxi
Annexe 23: Script de la création de la table temporaire « ADS_REFILL ».....	xxii

Annexe 21: Script de la création de la table temporaire « ADS_INCOMING » **Erreur ! Signet non défini.**

Annexe 22: Script de création de la macro « LOAD_ADS_INCOMING » de chargement de table temporaire « ADS_INCOMING » pour la catégorie de KPI traffic_incoming **Erreur ! Signet non défini.**

Annexe 23: Script de la création de la table temporaire « ADS_REFILL » **Erreur ! Signet**

Annexe 24: Script de création de la macro « LOAD_ADS_REFILL » de chargement de table temporaire « ADS_REFILL » pour la catégorie de KPI traffic_refill.....	xxv
Annexe 25: Script de création de l'ADS.....	xxviii
Annexe 26: Le nom et le type des KPI (attributs de la table ADS).....	xxix
Annexe 27: Script de la macro de chargement de la table « ADS_FINAL » nommée « LOAD_ADS_FINAL ».....	xxx
Annexe 28: Script création de la table « ADS_Kmeans ».....	xxxi
Annexe 29: Script de la fonction « remove_outliers » appliquée sur chaque variable quantitative.....	xxxii
Annexe 30: Script de création de la table « Traffic_Behaviour ».....	xxxiii
Annexe 31: Script de chargement de la table « Traffic_Behaviour ».....	xxxiv
Annexe 32: La table ADS_FINAL chargée en Mars 2019.....	xxxv

1) Description des dimensions et des tables de faits (liste, type et description des attributs) :

Dimension		Description
DIM_DATE	La dimension date est en plus de pratique, nécessaire, en effet, est associé pour chaque mesure d'une table de faits au moins une date.	
Liste des Attributs	Type	Description
CALENDAR_DATE	Date format YY/MM/DD	Une date calendaire identifiant la table (PK)
WEEK	Entier BYTEINT	Une semaine calendaire
MONTH_NUM	Entier BYTEINT	Un mois calendaire
YEAR_NUM	Entier SMALLINT	Une année calendaire

Annexe 1: Dictionnaire de donnée de la dimension date

Dimension		Description
DIM_OFFER	Cette dimension contient les informations relatives au type et au contenu des offres Djazzy	
Liste des attributs	Type	Description
OFFER_KEY	Chaîne de caractères	Identifiant de l'offre Djazzy (PK)
OFFER_CUST_TYPE	Chaîne de caractères	Type du client d'une offre Business : Entreprise Individual : abonné individuel non commercial
GLOBAL_PROFILE	Chaîne de caractères	Profil de l'offre <i>Ex</i> : Djazzy_Hayla
OFFER_MAIN_TP	Chaîne de caractères	Même chose que le GLOBAL-PROFILE
TARIFF_PROFILE	Chaîne de caractères	Sous-profil du GLOBAL-PROFILE Par rapport au type du réseau : 2G, 3G, LTE <i>Ex</i> : Djazzy Hayla 3G Djazzy Hayla 3G
OFFER_CAT	Chaîne de caractères	Catégorie de l'offre (c'est un code)

		alphanumérique)
OFFER_LINE_TYPE	Chaîne de caractères	2G ,3G, 4G
OFFNG_SUBS_TYP	Chaîne de caractères (250)	<i>Prepaid</i> : prépayé <i>Postpaid</i> : paiement via facture mensuelle <i>Hybrid</i> : prépayé et postpayé en même temps

Annexe 2: dictionnaire de donnée de la dimension offre

Dimension		Description
Dim_Charge_Type		Cette dimension contient les informations relatives au type de transactions
Liste des attributs	Type	Description
ID_CHARGE_TYPE	Chaîne de caractères(10)	Identifiant du type de transaction (Appels, sms, internet) (<i>PK</i>) <i>Ex</i> : Pkg
CHARGE_TYPE_DESCRIPTION	Chaîne de caractères(100)	Types de transactions : <i>Ex</i> : Internet « Pkg »

Annexe 4507: Dictionnaire de donnée de la dimension type de rechargement

Dimension		Description
Dim_Call_Direction	Cette dimension contient la direction de l'appel	
Liste des attributs	Type	Description
ID_CALL_DIRECTION	Chaîne de caractères (5)	Identifiant sur la direction de l'appel : U, O, I, S (PK)
CALL_DIRECTION_DESCRIPTION	Chaîne de caractères (15)	Description de l'Id_Call_Direction U : Unknown (inconnu) O :Out (émis) I :In (reçu) S: Service Action

Annexe 4: Dictionnaire de donnée de la dimension direction d'appels

Dimension		Description
Dim_Cell_Loc	Cette dimension contient les informations relatives à l'activité des cellules réseaux Djazzy	
Liste des attributs	Type	Description
SUBS_ID	Entier	Identifiant souscripteur en tant que (PK)
LAST_MOST_USED_CELL	Chaîne de caractères (20)	La dernière cellule réseau Djazzy la plus utilisée
LAST_MOST_USED_IMEI	Chaîne de caractères (15)	IMEI : International Mobile Equipment Identity Identifiant du téléphone cellulaire le plus utilisé dernièrement

Annexe 4508: Dictionnaire de donnée de la dimension relative à l'activité des antennes réseau Djazzy

Dimension		Description
DIM_SUBS	Cette dimension contient les informations relatives aux abonnés Djezzy	
Liste des attributs	Type	Description
SUBS_KEY	Entier	Identifiant associé à une souscription (PK)
CUST_ID	Entier	Identifiant de l'utilisateur du numéro associé à la SIMCARD (change après réattribution du numéro de tél)
MSISDN	Chaîne de caractères (25)	Numéro de téléphone associé à la SIMCARD
SUBS_TYP	Chaines de caractères (250)	Type d'un abonné : Prepaid : prépayé Postpaid : paiement via facture mensuelle Hybrid : prépayé et postpayé en même temps
SUBS_STS	Chaîne de caractères (250)	Actif : La SIMCARD de l'abonné est en activité TER : la SIMCARD n'est pas en activité depuis au moins 90 jours
SUBS_CUST_KEY	Entier	
SUBS_OFFNG_KEY	Entier	Clé identifiant une offre Djezzy
LST_CALL_OUT_DT	Timestamp format YYYYYY-MM-DDDBHH:MI:SS	Date du Dernier appel émis
LST_CALL_IN_DT	Timestamp format YYYYYY-MM-DDDBHH:MI:SS	Date du Dernier appel reçu
FRST_CALL_DT	Timestamp format YYYYYY-MM-	Date du premier appel (émis ou reçu)

		DDBHH:MI:SS	
SUBS_ACT_DT		Timestamp format YYYYYY-MM- DDBHH:MI:SS	Date de la dernière activité de la SIMCARD
SUBS_FRST_ACT_DT	Timestamp format YYYYYY-MM- DDBHH:MI:SS	Date d'activation de la SIMCARD	

Annexe 6: Dictionnaire de donnée de la dimension souscripteur

Dimension		Description
Dim_Handset	Cette dimension contient les informations relatives au téléphone associé à la SIMCARD Djezzy	
Liste des attributs	Type	Description
TAC	Entier	Identifiant unique du téléphone (PK)
HANDSET_MODEL	Chaîne	Nom du modèle de téléphone
MANUFACTURER	Chaîne de caractères	Nom du Fabricant du téléphone
OS	Chaîne de caractères	Système d'exploitation du téléphone Ex : Android, ios
HANDSET_3G	Chaîne de caractères	Téléphone 3G ou pas
LTE	Chaîne de caractères	Téléphone 4G ou pas
DEVICETYPE	Chaîne de caractères	Type du téléphone : Smartphone ou pas
BLUETOOTH	Chaîne de caractères	Téléphone possédant l'option Bluetooth ou pas
DUALSIM	Chaîne de caractères	Téléphone Double SIM ou pas
HANDSET_2G	Chaîne de caractères	Téléphone 2G ou pas
SMARTPHONE	Chaîne de caractères	Smartphone ou pas

Annexe 7: Dictionnaire de donnée de la dimension relative au téléphone associé à la SIMCARD Djezzy

Dimension		Description
Dim_tariff_group	Cette dimension contient les informations relatives aux transactions par destination	
Liste des attributs	Type	Description
TARIFF_GROUP_ID	Chaîne de caractères (50)	Identifiant de la transaction par destination (PK)
TARIFF_GROUP_DESCRIPTION	Chaîne de caractères (255)	Description de la transaction par destination <i>Ex:</i> Appel à l'international
CHARGE_TYPE	Chaîne de caractères fixe (3)	SMS, Voice, Data...
DESTINATION_TYPE	Chaîne de caractères(50)	On-net, off-net,
DESTINATION_COMPTA	Chaîne de caractères (30)	International, roaming

Annexe 8: Dimension relative aux transactions par destination

Dimension		Description
Dim_Cell_Area	Cette dimension contient les informations relatives à la localisation des cellules réseaux Djazzy	
Liste des attributs	Type	Description
CELL_SITE_ID	Entier	Identifiant souscripteur (PK)
TECHNOLOGIE	Chaîne de caractères (20)	La dernière cellule réseau Djazzy la plus utilisée
BSC	Chaîne de caractères	IMEI : International Mobile Equipment Identity Identifiant du téléphone cellulaire le plus utilisé dernièrement

CELL_NAME	Chaine de caractères	Nom de la cellule Djazzy
COMMUNE	Chaine de caractères	Commune de la cellule Djazzy
WILAYA	Chaine de caractères	Wilaya de la cellule Djazzy
REGION	Chaine de caractères	Région de la cellule Djazzy : Est, centre, Ouest, Sud
CODE	Chaine de caractères	Code de la cellule Djazzy
LONGITUDE	Chaine de caractères	Longitude de l'emplacement de la cellule Djazzy
ALTITUDE	Chaine de caractères	Altitude de l'emplacement de la cellule Djazzy

Annexe 9: Dictionnaire de donnée de la dimension relative à la Localisation des antennes réseau Djazzy

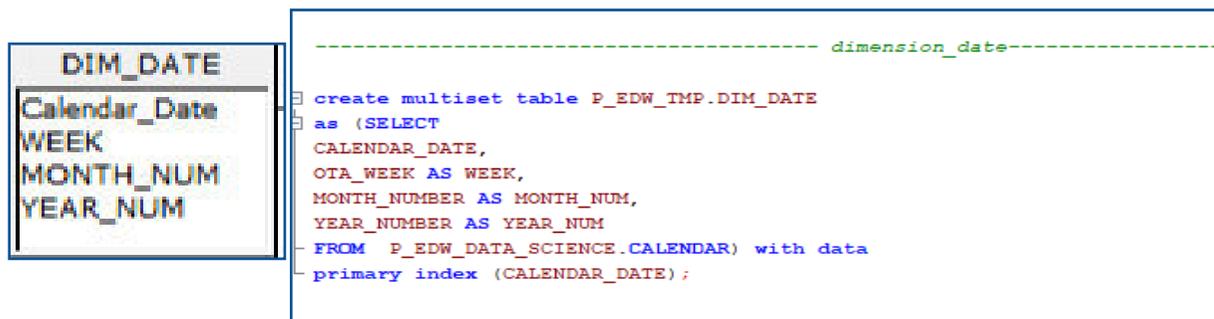
Table de faits		Description
Fact_Refill		Cette table de faits est relative aux deux types de rechargement : « Flexy » et « Voucher »
Liste des clés et attribut	Type	Description
SUBS_ID	Entier	Clé étrangère de la dimension Dim_Subs(FK)
OFFER_ID	Entier	Clé étrangère de la dimension Dim_Offer(FK)
REFILL_TYPE	Chaine de caractères	Attribut sur le type de rechargement : flexy ou voucher
Liste des mesures	Type	Description
AMOUNT	Entier	Mesure sur le revenu (DZA)des rechargements de type flexy et voucher
BONUS_AMOUNT	Entier	Mesure sur le bonus (DZA)suite aux rechargements de type flexy et voucher

Annexe 10 : Dictionnaire de donnée de la table de fait rechargement

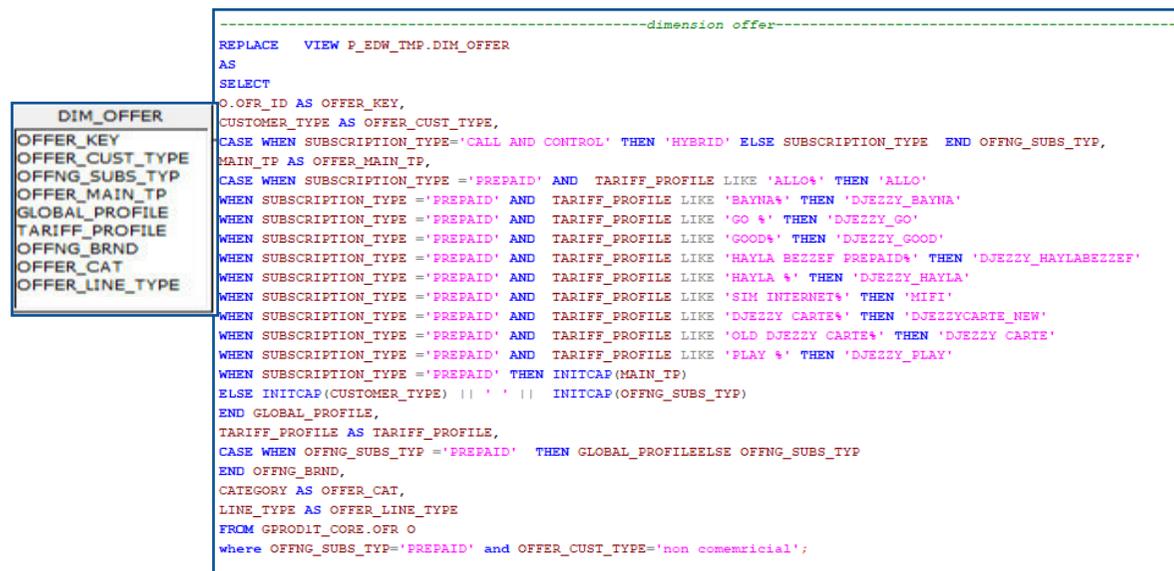
Table de faits		Description
Fact_Traffic	Cette table de faits est relative au trafic des abonnés en termes d'appels, de sms et de consommation en données mobiles.	
Liste des clés	Type	Description
SUBSCRIBER_ID	Entier	Clé étrangère de la dimension Dim_Subs
CALENDAR_DATE	Chaine de caractères	Clé étrangère de la dimension Dim_Date(FK)
OFFER_KEY	Entier	Clé étrangère de la dimension Dim_Offer(FK)
OPERATOR	Entier	Clé étrangère de la dimension Dégénérée « Operator » (FK)
ID_CALL_DIRECTION	Entier	Clé étrangère de la dimension Dim_Call_direction(FK)
ID_CHARGE_TYPE	Entier	Clé étrangère de la dimension Dim_Charge_Type(FK)
TARIFF_GROUP_ID	Entier	Clé étrangère de la dimension Dim_Tariff_Group(FK)
Liste des mesures	Type	Description
Number_of_Calls	Entier	Mesure sur le nombre d'appels et de sms
Actual_Duration	Entier	Mesure sur la durée de l'appel
Billed_Duration	Entier	Mesure la durée facturée de l'appel
Billed_Amount	Entier	Mesure sur le revenu (DZA) de l'appel ou sms ou de l'offre internet

Annexe 11: Dictionnaire de donnée de la table de fait trafic

2) Modèle physique et script de création des dimensions : date, offer, cell_area, hanset,



Annexe 12: Modèle physique et script de création de la dimension date



Annexe 13: Modèle physique et script de création de la dimension offre

Dim_Cell_Area	
Cell_Site_ID	-----dimension_cell_area-----
TECHNOLOGIE	
BSC	
CELL_NAME	create view P_EDW_TMP.Dim_Cell_Area
Commune	as select * from p_edw_bo_universe.cell_site
Wilaya	
Region	
Code	
Longitude	
Altitude	

Annexe 14: Modèle physique et script de création de la dimension cell_area *ea*

Dim_Handset	
TAC	-----dimension_Handset-----
HANDSET_MODEL	
Manufacturer	create view P_EDW_TMP.Dim_Handset
OS	as (
EDGE	select
handset_3G	TAC,---- id unique du modele de tel-- (TAC représente les 8 premiers chiffres du last_most_used_IMEI de la table dim_cell_location)
HSDPA	HANDSET_MODEL,
HSUPA	Manufacturer,
LTE	OS,--- operational system---
Devicetype	EDGE,--2g--
Bluetooth	handset_3G,
DualSIM	HSDPA,
handset_2G	HSUPA,
smartphone	LTE,
	Devicetype,
	Bluetooth,
	DualSIM,
	handset_2G,
	smartphone
	from p_edw_bo_universe.handset_data
)

Annexe 15 : Modèle physique et script de création de la dimension handset

```

DEFINE JOB Job_cell_loc
DESCRIPTION 'Job for loading Dim cell_location'
(
  DEFINE OPERATOR W_1_o_Job_cell_loc
  TYPE LOAD
  SCHEMA *
  ATTRIBUTES
  (
    VARCHAR UserName,
    VARCHAR UserPassword,
    VARCHAR LogTable,
    VARCHAR TargetTable,
    INTEGER BufferSize,
    INTEGER ErrorLimit,
    INTEGER MaxSessions,
    INTEGER MinSessions,
    INTEGER TenacityHours,
    INTEGER TenacitySleep,
    VARCHAR AccountID,
    VARCHAR DateForm,
    VARCHAR ErrorTable1,
    VARCHAR ErrorTable2,
    VARCHAR NotifyExit,
    VARCHAR NotifyExitIsDLL,
    VARCHAR NotifyLevel,
    VARCHAR NotifyMethod,
    VARCHAR NotifyString,
    VARCHAR PauseAcq,
    VARCHAR PrivateLogName,
    VARCHAR TdpId,
  )
  DEFINE SCHEMA W_0_s_Job_cell_loc
  (
    Subs_Id INTEGER,
    Last_Most_Used_Cell VARCHAR(20),
    Last_Most_Used_Imei VARCHAR(15)
  );
  DEFINE OPERATOR W_0_o_Job_cell_loc
  TYPE EXPORT
  SCHEMA W_0_s_Job_cell_loc
  ATTRIBUTES
  (
    VARCHAR UserName,
    VARCHAR UserPassword,
    VARCHAR SelectStmt,
    INTEGER BlockSize,
    INTEGER MaxSessions,
    INTEGER MinSessions,
    INTEGER TenacityHours,
    INTEGER TenacitySleep,
    INTEGER MaxDecimalDigits,
    VARCHAR AccountID,
    VARCHAR DateForm,
    VARCHAR NotifyExit,
    VARCHAR NotifyExitIsDLL,
    VARCHAR NotifyLevel,
    VARCHAR NotifyMethod,
    VARCHAR NotifyString,
    VARCHAR PrivateLogName,
    VARCHAR TdpId,
    VARCHAR TraceLevel,
    VARCHAR WorkingDatabase
  );
);

```

```

APPLY
(
  'INSERT INTO p_edw_tmp.Dim_Cell_loc_stg(Subs_Id ,Last_Most_Used_Cell,Last_Most_Used_Imei)
  VALUES(:Subs_Id ,:Last_Most_Used_Cell,:Last_Most_Used_Imei);'
)
TO OPERATOR
(
  W_l_o_Job_cell_loc[1]

  ATTRIBUTES
  (
    UserName = '██████████',
    UserPassword = '██████████',
    LogTable = 'P_EDW_TMP.Dim_cell_log',
    TargetTable = 'p_edw_tmp.Dim_Cell_loc_stg',
    TdpId = '172.16.53.41'
  )
)
SELECT * FROM OPERATOR
(
  W_0_o_Job_cell_loc[1]

  ATTRIBUTES
  (
    UserName = '██████████',
    UserPassword = '██████████',
    SelectStmt = 'SELECT Subs_Id ,Last_Most_Used_Cell,Last_Most_Used_Imei FROM p_edw_tmp.dim_cell_loc_stg;',
    TdpId = '172.16.53.24'
  )
);
);

```

Annexe 16: Le script du TPT de la dim_cell_loc


```

-----KPI_SMS*/
-----NBER OF SMS
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_M3
-----REVENU
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_M3
/* KPIS_SMS_ON_NET*/
-----NBER OF SMS
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_M3
/* KPIS_SMS_OFF_NET*/
-----NBER OF SMS
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMW') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMW') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMW') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_M3
/* KPIS_DATA*/
-----USAGE
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_M3
-----USAGE PAYAS GO
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_M3
-----REV PAY AS GO
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_M3
/* KPIS_SOUSCRIPTION */
-----NBER OF SOUSC
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='PHG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='PHG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='PHG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_M3
-----REVENU
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='PHG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='PHG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='PHG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_M3

```

```

FROM P_EDW_TMP.FACT_TRAFFIC AS A
INNER JOIN P_EDW_TMP.DIM_OFFER AS B ON A.OFFER_KEY=B.OFFER_KEY
WHERE ID_CALL_DIRECTION='O'
AND CALENDAR_DATE BETWEEN date-90 AND DATE
GROUP BY 1
) with no data
primary index (SUBSCRIBER_ID)

```

Annexe 17: Script de la création de la table temporaire « ADS_Outgoing_Month »


```

SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_W4

-----REVENU-----
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMW','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_W4

/* KPIS_SMS_ON_NET*/
-----NBER OF SMS-----
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_W4

/* KPIS_SMS_OFF_NET*/
-----NBER OF SMS-----
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMW') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMW') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMW') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMW') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_W4

/* KPIS_DATA*/
-----USAGE-----
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_W4

-----USAGE PAYAS GO-----
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_W4

-----REV PAY AS GO-----
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_W4

/* KPIS_SOUSCRIPTION */
-----NBER OF SOUSC-----
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='ENG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='ENG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='ENG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='ENG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_W4

-----REVENU-----
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CHARGE_TYPE='ENG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_W1
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CHARGE_TYPE='ENG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_W2
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CHARGE_TYPE='ENG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_W3
,SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CHARGE_TYPE='ENG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_W4

```

```

FROM P_EDW_IMP.FACT_TRAFFIC AS A

INNER JOIN P_EDW_IMP.DIM_OFFER AS B ON A.OFFER_KEY=B.OFFER_KEY

WHERE ID_CALL_DIRECTION='O'
AND CALENDAR_DATE BETWEEN :run_date-28 AND DATE
GROUP BY 1
);

```

Annexe 18: Script de création de la macro « LOAD_ADS_OUTGOING_WEEK » de chargement de table temporaire « ADS_Outgoing_Week » pour la catégorie de KPI PI outgoing


```

--KPI_SMS*/
--NBER OF SMS
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMG','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMG','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMG','SINT') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_M3
--REVENU
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMG','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMG','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ','SMA','SMG','SINT') THEN BILLED_AMOUNT ELSE 0 END) REV_SMS_M3
/* KPIS_SMS_ON_NET*/
--NBER OF SMS
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMJ') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_ON_NET_M3
/* KPIS_SMS_OFF_NET*/
--NBER OF SMS
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMG') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMG') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='S' AND TARIFF_GROUP_ID IN ('SMA','SMG') THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SMS_OFF_NET_M3
/* KPIS_DATA*/
--USAGE
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='P' THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_M3
--USAGE PAYAS GO
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN CAST (ACTUAL_DURATION AS FLOAT)/ CAST(1024 AS FLOAT) ELSE 0 END) DATA_USAGE_MO_PAYG_M3
--REV PAY AS GO
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='P' AND BILLED_AMOUNT >0 THEN BILLED_AMOUNT ELSE 0 END) DATA_REV_MO_PAYG_M3
/* KPIS_SOUSCRIPTION */
--NBER OF SOUSC
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='FIG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='FIG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='FIG' THEN NUMBER_OF_CALLS ELSE 0 END) NBER_SOUSC_M3
--REVENU
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-30 AND date-1 AND ID_CHARGE_TYPE='FIG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-60 AND date-31 AND ID_CHARGE_TYPE='FIG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN date-90 AND date-61 AND ID_CHARGE_TYPE='FIG' THEN BILLED_AMOUNT ELSE 0 END) REV_SOUSC_M3

```

```

FROM P_EDW_TMP.FACT_TRAFFIC AS A
INNER JOIN P_EDW_TMP.DIM_OFFER AS B ON A.OFFER_KEY=B.OFFER_KEY
WHERE ID_CALL_DIRECTION='O'
AND CALENDAR_DATE BETWEEN :run_date-90 AND DATE
GROUP BY 1
);

```

Annexe 20: Script de création de la macro « LOAD_ADS_OUTGOING_MONTH » de chargement de table temporaire « ADS_Outgoing_Month » pour la catégorie de KPI traffic_outgoing

6) Script de la création de la table temporaire « ADS_INCOMING » :

```

create multiset TABLE p_edw_tmp.ADS_INCOMING as (
select
SUBSCRIBER_ID

-----Voice Incoming-----
sum(case when CALENDAR_DATE between date-7 and date -1 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICFREE','MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_In_W1
sum(case when CALENDAR_DATE between date-14 and date -8 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICFREE','MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_In_W2
sum(case when CALENDAR_DATE between date-21 and date -15 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICFREE','MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_In_W3
sum(case when CALENDAR_DATE between date-28 and date -22 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICFREE','MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_In_W4

sum(case when CALENDAR_DATE between date-30 and date -1 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICFREE','MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_In_M1
sum(case when CALENDAR_DATE between date-60 and date -31 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICFREE','MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_In_M2
sum(case when CALENDAR_DATE between date-90 and date -61 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICFREE','MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_In_M3

-----Voice Incoming off-net-----
sum(case when CALENDAR_DATE between date-7 and date -1 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_Offnet_In_W1
sum(case when CALENDAR_DATE between date-14 and date -8 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_Offnet_In_W2
sum(case when CALENDAR_DATE between date-21 and date -15 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_Offnet_In_W3
sum(case when CALENDAR_DATE between date-28 and date -22 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_Offnet_In_W4

sum(case when CALENDAR_DATE between date-30 and date -1 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_Offnet_In_M1
sum(case when CALENDAR_DATE between date-60 and date -31 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_Offnet_In_M2
sum(case when CALENDAR_DATE between date-90 and date -61 and ID_CALL_DIRECTION='I' and Tariff_Group_ID in ('MICAFREE','MICWFREE','MICCFREE','MICIFREE','MICDELTA','MICICFREE') then Number_of_Calls else 0 end) Voice_Usage_Offnet_In_M3

-----SMS INCOMING-----
sum(case when CALENDAR_DATE BETWEEN DATE-7 AND DATE -1 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMSJET','SMSTAM','SMSTWAT','SMSTINT','SMSTINT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_W1
sum(case when CALENDAR_DATE BETWEEN DATE-14 AND DATE -8 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMSJET','SMSTAM','SMSTWAT','SMSTINT','SMSTINT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_W2
sum(case when CALENDAR_DATE BETWEEN DATE-21 AND DATE -15 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMSJET','SMSTAM','SMSTWAT','SMSTINT','SMSTINT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_W3
sum(case when CALENDAR_DATE BETWEEN DATE-28 AND DATE -22 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMSJET','SMSTAM','SMSTWAT','SMSTINT','SMSTINT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_W4

sum(case when CALENDAR_DATE BETWEEN DATE-30 AND DATE -1 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMSJET','SMSTAM','SMSTWAT','SMSTINT','SMSTINT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_M1
sum(case when CALENDAR_DATE BETWEEN DATE-60 AND DATE -31 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMSJET','SMSTAM','SMSTWAT','SMSTINT','SMSTINT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_M2
sum(case when CALENDAR_DATE BETWEEN DATE-90 AND DATE -61 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMSJET','SMSTAM','SMSTWAT','SMSTINT','SMSTINT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_M3

)
from p_edw_tmp.Fact_Traffic as a
inner join p_edw_tmp.DIM_OFFER as b on a.OFFER_KEY=b.OFFER_KEY

and SUBSCRIBER_ID=129041435
and CALENDAR_DATE>date-90
group by 1 ) with no data
primary index(SUBSCRIBER_ID);

```

Annexe 21: Script de la création de la table temporaire « ADS_INCOMING »

7) Script de création de la macro « **LOAD_ADS_INCOMING** » de chargement de table temporaire « **ADS_INCOMING** » pour la catégorie de KPI **traffic_incoming** :

```

replace MACRO P_EDW_TMP.LOAD_ADS_INCOMING ( run_date date) AS (

DELETE FROM P_EDW_TMP.ADS_INCOMING;
INSERT INTO P_EDW_TMP.ADS_INCOMING
SELECT
SUBSCRIBER_ID

-----VOICE INCOMING-----
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCFREE','MTCAPFREE','MTCVFREE','MTCFREE','MTCIFREE','MTCDELTA','MTCICFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_W1
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCFREE','MTCAPFREE','MTCVFREE','MTCFREE','MTCIFREE','MTCDELTA','MTCICFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_W2
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCFREE','MTCAPFREE','MTCVFREE','MTCFREE','MTCIFREE','MTCDELTA','MTCICFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_W3
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCFREE','MTCAPFREE','MTCVFREE','MTCFREE','MTCIFREE','MTCDELTA','MTCICFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_W4

SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-30 AND :run_date-1 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCFREE','MTCAPFREE','MTCVFREE','MTCFREE','MTCIFREE','MTCDELTA','MTCICFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-60 AND :run_date-31 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCFREE','MTCAPFREE','MTCVFREE','MTCFREE','MTCIFREE','MTCDELTA','MTCICFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-90 AND :run_date-61 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCFREE','MTCAPFREE','MTCVFREE','MTCFREE','MTCIFREE','MTCDELTA','MTCICFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_M3

-----VOICE INCOMING OFF-NET-----
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCAPFREE','MTCVFREE','MTCFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_OFFNET_W1
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCAPFREE','MTCVFREE','MTCFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_OFFNET_W2
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCAPFREE','MTCVFREE','MTCFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_OFFNET_W3
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCAPFREE','MTCVFREE','MTCFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_OFFNET_W4

SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-30 AND :run_date-1 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCAPFREE','MTCVFREE','MTCFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_OFFNET_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-60 AND :run_date-31 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCAPFREE','MTCVFREE','MTCFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_OFFNET_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-90 AND :run_date-61 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('MTCAPFREE','MTCVFREE','MTCFREE') THEN NUMBER_OF_CALLS ELSE 0 END) VOICE_USAGE_IN_OFFNET_M3

-----SMS INCOMING-----
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-7 AND :run_date-1 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMS311','SMS1AM','SMS2WAT','SMS1INT','SMS1INT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_W1
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-14 AND :run_date-8 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMS311','SMS1AM','SMS2WAT','SMS1INT','SMS1INT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_W2
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-21 AND :run_date-15 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMS311','SMS1AM','SMS2WAT','SMS1INT','SMS1INT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_W3
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-28 AND :run_date-22 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMS311','SMS1AM','SMS2WAT','SMS1INT','SMS1INT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_W4

SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-30 AND :run_date-1 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMS311','SMS1AM','SMS2WAT','SMS1INT','SMS1INT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_M1
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-60 AND :run_date-31 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMS311','SMS1AM','SMS2WAT','SMS1INT','SMS1INT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_M2
SUM(CASE WHEN CALENDAR_DATE BETWEEN :run_date-90 AND :run_date-61 AND ID_CALL_DIRECTION='I' AND TARIFF_GROUP_ID IN ('SMS311','SMS1AM','SMS2WAT','SMS1INT','SMS1INT') THEN NUMBER_OF_CALLS ELSE 0 END) SMS_IN_M3

FROM P_EDW_TMP.FACT_TRAFFIC AS A

INNER JOIN P_EDW_TMP.DIM_OFFER AS B ON A.OFFER_KEY=B.OFFER_KEY

WHERE CALENDAR_DATE>:run_date-90
GROUP BY 1 );

```

*Annexe 22: Script de création de la macro « **LOAD_ADS_INCOMING** » de chargement de table temporaire « **ADS_INCOMING** » pour la catégorie de KPI **traffic_incoming***

```
create multiset TABLE p_edw_tmp.ADS_REFILL as (
  select Subs_id
```

```

-----**** REFILL J-1 ***-----
,sum(case when REFILL_DATE = :run_date-1 then Amount else 0 end) Amount_day_minus_1
-----Total REFILL -----

-----Nombre_Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_W1
,sum (case when REFILL_DATE between:run_date-14 and:run_date-8 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_W2
,sum (case when REFILL_DATE between:run_date-21 and:run_date-15 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_W3
,sum (case when REFILL_DATE between:run_date-28 and:run_date-22 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_M1
,sum (case when REFILL_DATE between:run_date-60 and:run_date-31 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_M2
,sum (case when REFILL_DATE between:run_date-90 and:run_date-61 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_M3
-----Revenu Refill -----

,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 then Amount else 0 end) Rev_Refill_W1
,sum(case when REFILL_DATE between:run_date-14 and:run_date-8 then Amount else 0 end ) Rev_Refill_W2
,sum(case when REFILL_DATE between:run_date-21 and:run_date-15 then Amount else 0 end) Rev_Refill_W3
,sum(case when REFILL_DATE between:run_date-28 and:run_date-22 then Amount else 0 end ) Rev_Refill_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 then Amount else 0 end) Rev_Refill_M1
,sum(case when REFILL_DATE between:run_date-60 and:run_date-31 then Amount else 0 end ) Rev_Refill_M2
,sum(case when REFILL_DATE between:run_date-90 and:run_date-61 then Amount else 0 end) Rev_Refill_M3
,sum(case when REFILL_DATE between:run_date-120 and:run_date-91 then Amount else 0 end) Rev_Refill_M4
,sum(case when REFILL_DATE between:run_date-150 and:run_date-121 then Amount else 0 end) Rev_Refill_M5
,sum(case when REFILL_DATE between:run_date-180 and:run_date-151 then Amount else 0 end) Rev_Refill_M6
-----**** REFILL_Flexy ***-----

-----Nombre_Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_W1
,sum (case when REFILL_DATE between:run_date-14 and:run_date-8 and Refill_Type ='flexy'then 1 else 0 end) Nber_Refill_Flexy_W2
,sum (case when REFILL_DATE between:run_date-21 and:run_date-15 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_W3
,sum (case when REFILL_DATE between:run_date-28 and:run_date-22 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_M1
,sum (case when REFILL_DATE between:run_date-60 and:run_date-31 and Refill_Type ='flexy'then 1 else 0 end) Nber_Refill_Flexy_M2
,sum (case when REFILL_DATE between:run_date-90 and:run_date-61 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_M3

-----Revenu Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_W1
,sum(case when REFILL_DATE between:run_date-14 and:run_date-8 and Refill_Type ='flexy' then Amount else 0 end ) Rev_Refill_Flexy_W2
,sum(case when REFILL_DATE between:run_date-21 and:run_date-15 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_W3
,sum(case when REFILL_DATE between:run_date-28 and:run_date-22 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_M1
,sum(case when REFILL_DATE between:run_date-60 and:run_date-31 and Refill_Type ='flexy' then Amount else 0 end ) Rev_Refill_Flexy_M2
,sum(case when REFILL_DATE between:run_date-90 and:run_date-61 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_M3
```

8) Script de création de la macro « LOAD_ADS_REFILL » de chargement de table temporaire « ADS_REFILL » pour la catégorie de KPI traffic_refill :

```

replace MACRO P_EDW_TMP.LOAD_ADS_REFILL (run_date date)
AS (
delete from p_edw_tmp.ADS_REFILL;
insert into p_edw_tmp.ADS_REFILL
select Subs_id

-----**** REFILL J-1 ***-----
,sum(case when REFILL_DATE = :run_date-1 then Amount else 0 end) Amount_day_minus_1
----- Total REFILL -----

-----Nombre_Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_W1
,sum (case when REFILL_DATE between:run_date-14 and:run_date-8 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_W2
,sum (case when REFILL_DATE between:run_date-21 and:run_date-15 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_W3
,sum (case when REFILL_DATE between:run_date-28 and:run_date-22 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_M1
,sum (case when REFILL_DATE between:run_date-60 and:run_date-31 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_M2
,sum (case when REFILL_DATE between:run_date-90 and:run_date-61 and Refill_Type in ('voucher','flexy') then 1 else 0 end) Nber_Refill_M3
----- Revenu Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 then Amount else 0 end) Rev_Refill_W1
,sum(case when REFILL_DATE between:run_date-14 and:run_date-8 then Amount else 0 end) Rev_Refill_W2
,sum(case when REFILL_DATE between:run_date-21 and:run_date-15 then Amount else 0 end) Rev_Refill_W3
,sum(case when REFILL_DATE between:run_date-28 and:run_date-22 then Amount else 0 end) Rev_Refill_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 then Amount else 0 end) Rev_Refill_M1
,sum(case when REFILL_DATE between:run_date-60 and:run_date-31 then Amount else 0 end) Rev_Refill_M2
,sum(case when REFILL_DATE between:run_date-90 and:run_date-61 then Amount else 0 end) Rev_Refill_M3
,sum(case when REFILL_DATE between:run_date-120 and:run_date-91 then Amount else 0 end) Rev_Refill_M4
,sum(case when REFILL_DATE between:run_date-150 and:run_date-121 then Amount else 0 end) Rev_Refill_M5
,sum(case when REFILL_DATE between:run_date-180 and:run_date-151 then Amount else 0 end) Rev_Refill_M6

-----**** REFILL_Flexy ***-----

-----Nombre_Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_W1
,sum (case when REFILL_DATE between:run_date-14 and:run_date-8 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_W2
,sum (case when REFILL_DATE between:run_date-21 and:run_date-15 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_W3
,sum (case when REFILL_DATE between:run_date-28 and:run_date-22 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_M1
,sum (case when REFILL_DATE between:run_date-60 and:run_date-31 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_M2
,sum (case when REFILL_DATE between:run_date-90 and:run_date-61 and Refill_Type ='flexy' then 1 else 0 end) Nber_Refill_Flexy_M3

```

```

----- Revenu Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_W1
,sum(case when REFILL_DATE between:run_date-14 and:run_date-8 and Refill_Type ='flexy' then Amount else 0 end ) Rev_Refill_Flexy_W2
,sum(case when REFILL_DATE between:run_date-21 and:run_date-15 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_W3
,sum(case when REFILL_DATE between:run_date-28 and:run_date-22 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_M1
,sum(case when REFILL_DATE between:run_date-60 and:run_date-31 and Refill_Type ='flexy' then Amount else 0 end ) Rev_Refill_Flexy_M2
,sum(case when REFILL_DATE between:run_date-90 and:run_date-61 and Refill_Type ='flexy' then Amount else 0 end) Rev_Refill_Flexy_M3

-----**** REFILL_Voucher ****-----
-----Nombref Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 and Refill_Type ='Voucher' then 1 else 0 end) Nber_Refill_Voucher_W1
,sum (case when REFILL_DATE between:run_date-14 and:run_date-8 and Refill_Type ='Voucher'then 1 else 0 end) Nber_Refill_Voucher_W2
,sum (case when REFILL_DATE between:run_date-21 and:run_date-15 and Refill_Type ='Voucher' then 1 else 0 end) Nber_Refill_Voucher_W3
,sum (case when REFILL_DATE between:run_date-28 and:run_date-22 and Refill_Type ='Voucher' then 1 else 0 end) Nber_Refill_Voucher_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 and Refill_Type ='Voucher' then 1 else 0 end) Nber_Refill_Voucher_M1
,sum (case when REFILL_DATE between:run_date-60 and:run_date-31 and Refill_Type ='Voucher'then 1 else 0 end) Nber_Refill_Voucher_M2
,sum (case when REFILL_DATE between:run_date-90 and:run_date-61 and Refill_Type ='Voucher' then 1 else 0 end) Nber_Refill_Voucher_M3

----- Revenu Refill -----
,sum(case when REFILL_DATE between:run_date-7 and:run_date-1 and Refill_Type ='Voucher' then Amount else 0 end) Rev_Refill_Voucher_W1
,sum(case when REFILL_DATE between:run_date-14 and:run_date-8 and Refill_Type ='Voucher' then Amount else 0 end ) Rev_Refill_Voucher_W2
,sum(case when REFILL_DATE between:run_date-21 and:run_date-15 and Refill_Type ='Voucher' then Amount else 0 end) Rev_Refill_Voucher_W3
,sum(case when REFILL_DATE between:run_date-28 and:run_date-22 and Refill_Type ='Voucher' then Amount else 0 end) Rev_Refill_Voucher_W4

,sum(case when REFILL_DATE between:run_date-30 and:run_date-1 and Refill_Type ='Voucher' then Amount else 0 end) Rev_Refill_Voucher_M1
,sum(case when REFILL_DATE between:run_date-60 and:run_date-31 and Refill_Type ='Voucher' then Amount else 0 end ) Rev_Refill_Voucher_M2
,sum(case when REFILL_DATE between:run_date-90 and:run_date-61 and Refill_Type ='Voucher' then Amount else 0 end) Rev_Refill_Voucher_M3

-----**** Max_Refill ****-----
,max(case when REFILL_DATE between:run_date-30 and:run_date-1 then Amount else 0 end) Rev_Refill_Max_M1
,max(case when REFILL_DATE between:run_date-60 and:run_date-31 then Amount else 0 end ) Rev_Refill_Max_M2
,max(case when REFILL_DATE between:run_date-90 and:run_date-61 then Amount else 0 end) Rev_Refill_Max_M3
,max(case when REFILL_DATE between:run_date-120 and:run_date-91 then Amount else 0 end) Rev_Refill_Max_M4
,max(case when REFILL_DATE between:run_date-150 and:run_date-121 then Amount else 0 end) Rev_Refill_Max_M5
,max(case when REFILL_DATE between:run_date-180 and:run_date-151 then Amount else 0 end) Rev_Refill_Max_M6

-----**** potentiel ****-----
,(case when
case when Rev_Refill_M1> Rev_Refill_M2 then Rev_Refill_M1 else Rev_Refill_M2 end>
case when Rev_Refill_M3> Rev_Refill_M4 then Rev_Refill_M3 else Rev_Refill_M4 end and
case when Rev_Refill_M1> Rev_Refill_M2 then Rev_Refill_M1 else Rev_Refill_M2 end>
case when Rev_Refill_M5> Rev_Refill_M6 then Rev_Refill_M5 else Rev_Refill_M6 end then
case when Rev_Refill_M1> Rev_Refill_M2 then Rev_Refill_M1 else Rev_Refill_M2 end
when case when Rev_Refill_M3> Rev_Refill_M4 then Rev_Refill_M3 else Rev_Refill_M4 end>
case when Rev_Refill_M5> Rev_Refill_M6 then Rev_Refill_M5 else Rev_Refill_M6 end then
case when Rev_Refill_M3> Rev_Refill_M4 then Rev_Refill_M3 else Rev_Refill_M4 end else
case when Rev_Refill_M5> Rev_Refill_M6 then Rev_Refill_M5 else Rev_Refill_M6 end
end ) - Rev_Refill_M1 as potentiel
-----**** variance ****-----
,var_pop (Amount) variance_refill

-----**** Ecartype ****-----
,sqrt( variance_refill) ecartype_rev
,null gap
from P_EDW_TMP.Fact_Refill as c
--inner join p_edw_tmp.DIM_OFFER as b on c.offer_id =b.OFFER_KEY
where REFILL_DATE BETWEEN:run_date-180 AND DATE
group by 1;

```

```

-----****          GAP  ***-----
update p_edw_tmp.ADS_REFILL
]from (
select
Subs_id
,coalesce(Gap,90) gap
from
] (
select
Subs_id
,gap/nbr as Gap
]from (
select
Subs_id
,count(*) as nbr
,sum(gap) as gap
]from (
SELECT
Refill_Date
,Subs_id
,coalesce(max(Refill_Date) over (partition by Subs_id order by refill_date rows between 1 following and 1 following),date-1) as Next_Recharge
,Next_Recharge-Refill_Date as gap
FROM p_edw_tmp.Fact_Refill
where refill_date>=date-90
-) NextR
group by 1
having nbr>1

-) as OneRech

union all
select
Subs_id
,45 as Gap
]from (
select
Subs_id
,count(*) as nbr
]from (
SELECT
Refill_Date
,Subs_id
,coalesce(max(Refill_Date) over (partition by Subs_id order by refill_date rows between 1 following and 1 following),date-1) as Next_Recharge
,Next_Recharge-Refill_Date as gap
FROM p_edw_tmp.Fact_Refill
where refill_date>=date-90
-) as NextR
group by 1
having nbr=1
-) as OneRech
-) as Total
-) as f
set gap =f.gap
where p_edw_tmp.ADS_REFILL.Subs_id=f.Subs_id ;

):

```

Annexe 24: Script de création de la macro « LOAD_ADS_REFILL » de chargement de table temporaire « ADS_REFILL » pour la catégorie de KPI traffic_refill

0) Script de création de l'ADS :

```

create multiset table p_edw_tmp.ADS_FINAL as (
SELECT
date-extract(day from date) Run_Date
,SUBS_KEY as Subs_id
,FRST_CALL_DT
,LST_CALL_OUT_DT
,LST_CALL_IN_DT
-----Outgoing_KPI-----
,coalesce( ARPU_M1 , 0) as ARPU_M1
,coalesce( ARPU_M2 , 0) as ARPU_M2
,coalesce( ARPU_M3 , 0) as ARPU_M3
,coalesce( VOICE_USAGE_W1 , 0) as VOICE_USAGE_W1
,coalesce ( VOICE_USAGE_W2 ,0) as VOICE_USAGE_W2
,coalesce(VOICE_USAGE_W3 ,0) as VOICE_USAGE_W3
,coalesce(VOICE_USAGE_W4 ,0)as VOICE_USAGE_W4
,coalesce(VOICE_USAGE_M1 ,0) as VOICE_USAGE_M1
,coalesce(VOICE_USAGE_M2 ,0) as VOICE_USAGE_M2
,coalesce(VOICE_USAGE_M3 ,0) as VOICE_USAGE_M3
,coalesce(NBER_CALLS_W1 ,0) as NBER_CALLS_W1
,coalesce(NBER_CALLS_W2 ,0) as NBER_CALLS_W2
,coalesce(NBER_CALLS_W3 ,0) as NBER_CALLS_W3
,coalesce(NBER_CALLS_W4 ,0) as NBER_CALLS_W4
,coalesce(NBER_CALLS_M1 ,0) as NBER_CALLS_M1
,coalesce(NBER_CALLS_M2 ,0) as NBER_CALLS_M2
,coalesce(NBER_CALLS_M3 ,0) as NBER_CALLS_M3
,coalesce( VOICE_REV_W1 ,0) as VOICE_REV_W1
,coalesce( VOICE_REV_W2 ,0) as VOICE_REV_W2
,coalesce( VOICE_REV_W3 ,0) as VOICE_REV_W3
,coalesce( VOICE_REV_W4 ,0) as VOICE_REV_W4
,coalesce( VOICE_REV_M1 ,0) as VOICE_REV_M1
,coalesce( VOICE_REV_M2 ,0) as VOICE_REV_M2
,coalesce( VOICE_REV_M3 ,0) as VOICE_REV_M3

```

(1)

```

,coalesce(VOICE_USAGE_ON_NET_W1 ,0) as VOICE_USAGE_ON_NET_W1
,coalesce(VOICE_USAGE_ON_NET_W2 ,0) as VOICE_USAGE_ON_NET_W2
,coalesce(VOICE_USAGE_ON_NET_W3 ,0) as VOICE_USAGE_ON_NET_W3
,coalesce(VOICE_USAGE_ON_NET_W4 ,0) as VOICE_USAGE_ON_NET_W4
,coalesce(VOICE_USAGE_ON_NET_M1 ,0) as VOICE_USAGE_ON_NET_M1
,coalesce(VOICE_USAGE_ON_NET_M2 ,0) as VOICE_USAGE_ON_NET_M2
,coalesce(VOICE_USAGE_ON_NET_M3 ,0) as VOICE_USAGE_ON_NET_M3
,coalesce(NBER_CALLS_ON_NET_W1 ,0) as NBER_CALLS_ON_NET_W1
,coalesce(NBER_CALLS_ON_NET_W2 ,0) as NBER_CALLS_ON_NET_W2
,coalesce(NBER_CALLS_ON_NET_W3 ,0) as NBER_CALLS_ON_NET_W3
,coalesce(NBER_CALLS_ON_NET_W4 ,0) as NBER_CALLS_ON_NET_W4
,coalesce(NBER_CALLS_ON_NET_M1 ,0) as NBER_CALLS_ON_NET_M1
,coalesce(NBER_CALLS_ON_NET_M2 ,0) as NBER_CALLS_ON_NET_M2
,coalesce(NBER_CALLS_ON_NET_M3 ,0) as NBER_CALLS_ON_NET_M3
,coalesce( VOICE_REV_ON_NET_W1 ,0) as VOICE_REV_ON_NET_W1
,coalesce( VOICE_REV_ON_NET_W2 ,0) as VOICE_REV_ON_NET_W2
,coalesce( VOICE_REV_ON_NET_W3 ,0) as VOICE_REV_ON_NET_W3
,coalesce( VOICE_REV_ON_NET_W4 ,0) as VOICE_REV_ON_NET_W4
,coalesce( VOICE_REV_ON_NET_M1 ,0) as VOICE_REV_ON_NET_M1
,coalesce( VOICE_REV_ON_NET_M2 ,0) as VOICE_REV_ON_NET_M2
,coalesce( VOICE_REV_ON_NET_M3 ,0) as VOICE_REV_ON_NET_M3
,coalesce(VOICE_USAGE_OFF_NET_W1 ,0) as VOICE_USAGE_OFF NET_W1
,coalesce(VOICE_USAGE_OFF_NET_W2 ,0) as VOICE_USAGE_OFF NET_W2
,coalesce(VOICE_USAGE_OFF_NET_W3 ,0) as VOICE_USAGE_OFF NET_W3
,coalesce(VOICE_USAGE_OFF_NET_W4 ,0) as VOICE_USAGE_OFF NET_W4
,coalesce(VOICE_USAGE_OFF_NET_M1 ,0) as VOICE_USAGE_OFF NET_M1
,coalesce(VOICE_USAGE_OFF_NET_M2 ,0) as VOICE_USAGE_OFF NET_M2
,coalesce(VOICE_USAGE_OFF_NET_M3 ,0) as VOICE_USAGE_OFF NET_M3
,coalesce(NBER_CALLS_OFF_NET_W1 ,0) as NBER_CALLS_OFF NET_W1
,coalesce(NBER_CALLS_OFF_NET_W2 ,0) as NBER_CALLS_OFF NET_W2
,coalesce(NBER_CALLS_OFF_NET_W3 ,0) as NBER_CALLS_OFF NET_W3
,coalesce(NBER_CALLS_OFF_NET_W4 ,0) as NBER_CALLS_OFF NET_W4
,coalesce(NBER_CALLS_OFF_NET_M1 ,0) as NBER_CALLS_OFF NET_M1
,coalesce(NBER_CALLS_OFF_NET_M2 ,0) as NBER_CALLS_OFF NET_M2
,coalesce(NBER_CALLS_OFF_NET_M3 ,0) as NBER_CALLS_OFF NET_M3

```

(2)

```

,coalesce( VOICE_REV_OFF_NET_W1 ,0) as VOICE_REV_OFF_NET_W1
,coalesce( VOICE_REV_OFF_NET_W2 ,0) as VOICE_REV_OFF_NET_W2
,coalesce( VOICE_REV_OFF_NET_W3 ,0) as VOICE_REV_OFF_NET_W3
,coalesce( VOICE_REV_OFF_NET_W4 ,0) as VOICE_REV_OFF_NET_W4
,coalesce( VOICE_REV_OFF_NET_M1 ,0) as VOICE_REV_OFF_NET_M1
,coalesce( VOICE_REV_OFF_NET_M2 ,0) as VOICE_REV_OFF_NET_M2
,coalesce( VOICE_REV_OFF_NET_M3 ,0) as VOICE_REV OFF NET_M3
,coalesce(VOICE_USAGE_INT_W1 ,0) as VOICE_USAGE_INT_W1
,coalesce(VOICE_USAGE_INT_W2 ,0) as VOICE_USAGE_INT_W2
,coalesce(VOICE_USAGE_INT_W3 ,0) as VOICE_USAGE_INT_W3
,coalesce(VOICE_USAGE_INT_W4 ,0) as VOICE_USAGE_INT_W4
,coalesce(VOICE_USAGE_INT_M1 ,0) as VOICE_USAGE_INT_M1
,coalesce(VOICE_USAGE_INT_M2 ,0) as VOICE_USAGE_INT_M2
,coalesce(VOICE_USAGE_INT_M3 ,0) as VOICE_USAGE_INT_M3
,coalesce(NBER_CALLS_INT_W1 ,0) as NBER_CALLS_INT_W1
,coalesce(NBER_CALLS_INT_W2 ,0) as NBER_CALLS_INT_W2
,coalesce(NBER_CALLS_INT_W3 ,0) as NBER_CALLS_INT_W3
,coalesce(NBER_CALLS_INT_W4 ,0) as NBER_CALLS_INT_W4
,coalesce(NBER_CALLS_INT_M1 ,0) as NBER_CALLS_INT_M1
,coalesce(NBER_CALLS_INT_M2 ,0) as NBER_CALLS_INT_M2
,coalesce(NBER_CALLS_INT_M3 ,0) as NBER_CALLS_INT_M3
,coalesce( VOICE_REV_INT_W1 ,0) as VOICE_REV_INT_W1
,coalesce( VOICE_REV_INT_W2 ,0) as VOICE_REV_INT_W2
,coalesce( VOICE_REV_INT_W3 ,0) as VOICE_REV_INT_W3
,coalesce( VOICE_REV_INT_W4 ,0) as VOICE_REV_INT_W4
,coalesce( VOICE_REV_INT_M1 ,0) as VOICE_REV_INT_M1
,coalesce( VOICE_REV_INT_M2 ,0) as VOICE_REV_INT_M2
,coalesce( VOICE_REV_INT_M3 ,0) as VOICE_REV_INT_M3
,coalesce(NBER_SMS_W1 ,0) as NBER_SMS_W1
,coalesce(NBER_SMS_W2 ,0) as NBER_SMS_W2
,coalesce(NBER_SMS_W3 ,0) as NBER_SMS_W3
,coalesce(NBER_SMS_W4 ,0) as NBER_SMS_W4
,coalesce(NBER_SMS_M1 ,0) as NBER_SMS_M1
,coalesce(NBER_SMS_M2 ,0) as NBER_SMS_M2
,coalesce(NBER_SMS_M3 ,0) as NBER_SMS_M3
,coalesce(REV_SMS_W1 ,0) as REV_SMS_W1

```

(3)

```

,coalesce(REV_SMS_W2 ,0) as REV_SMS_W2
,coalesce(REV_SMS_W3 ,0) as REV_SMS_W3
,coalesce(REV_SMS_W4 ,0) as REV_SMS_W4
,coalesce(REV_SMS_M1 ,0) as REV_SMS_M1
,coalesce(REV_SMS_M2 ,0) as REV_SMS_M2
,coalesce(REV_SMS_M3 ,0) as REV_SMS_M3
,coalesce(NBER_SMS_ON_NET_W1 ,0) as NBER_SMS_ON_NET_W1
,coalesce(NBER_SMS_ON_NET_W2 ,0) as NBER_SMS_ON_NET_W2
,coalesce(NBER_SMS_ON_NET_W3 ,0) as NBER_SMS_ON_NET_W3
,coalesce(NBER_SMS_ON_NET_W4 ,0) as NBER_SMS_ON_NET_W4
,coalesce(NBER_SMS_ON_NET_M1 ,0) as NBER_SMS_ON_NET_M1
,coalesce(NBER_SMS_ON_NET_M2 ,0) as NBER_SMS_ON_NET_M2
,coalesce(NBER_SMS_ON_NET_M3 ,0) as NBER_SMS_ON_NET_M3
,coalesce(NBER_SMS_OFF_NET_W1 ,0) as NBER_SMS_OFF_NET_W1
,coalesce(NBER_SMS_OFF_NET_W2 ,0) as NBER_SMS_OFF_NET_W2
,coalesce(NBER_SMS_OFF_NET_W3 ,0) as NBER_SMS_OFF_NET_W3
,coalesce(NBER_SMS_OFF_NET_W4 ,0) as NBER_SMS_OFF_NET_W4
,coalesce(NBER_SMS_OFF_NET_M1 ,0) as NBER_SMS_OFF_NET_M1
,coalesce(NBER_SMS_OFF_NET_M2 ,0) as NBER_SMS_OFF_NET_M2
,coalesce(NBER_SMS_OFF_NET_M3 ,0) as NBER_SMS_OFF_NET_M3
,coalesce(DATA_USAGE_MO_W1 ,0) as DATA_USAGE_MO_W1
,coalesce(DATA_USAGE_MO_W2 ,0) as DATA_USAGE_MO_W2
,coalesce(DATA_USAGE_MO_W3 ,0) as DATA_USAGE_MO_W3
,coalesce(DATA_USAGE_MO_W4 ,0) as DATA_USAGE_MO_W4
,coalesce(DATA_USAGE_MO_M1 ,0) as DATA_USAGE_MO_M1
,coalesce(DATA_USAGE_MO_M2 ,0) as DATA_USAGE_MO_M2
,coalesce(DATA_USAGE_MO_M3 ,0) as DATA_USAGE_MO_M3
,coalesce(DATA_USAGE_MO_PAYG_W1 ,0) as DATA_USAGE_MO_PAYG_W1
,coalesce(DATA_USAGE_MO_PAYG_W2 ,0) as DATA_USAGE_MO_PAYG_W2
,coalesce(DATA_USAGE_MO_PAYG_W3 ,0) as DATA_USAGE_MO_PAYG_W3
,coalesce(DATA_USAGE_MO_PAYG_W4 ,0) as DATA_USAGE_MO_PAYG_W4
,coalesce(DATA_USAGE_MO_PAYG_M1 ,0) as DATA_USAGE_MO_PAYG_M1
,coalesce(DATA_USAGE_MO_PAYG_M2 ,0) as DATA_USAGE_MO_PAYG_M2
,coalesce(DATA_USAGE_MO_PAYG_M3 ,0) as DATA_USAGE_MO_PAYG_M3
,coalesce(DATA_REV_MO_PAYG_W1 ,0) as DATA_REV_MO_PAYG_W1
,coalesce(DATA_REV_MO_PAYG_W2 ,0) as DATA_REV_MO_PAYG_W2

```

(4)

```

,coalesce(DATA_REV_MO_PAYG_W3 ,0) as DATA_REV_MO_PAYG_W3
,coalesce(DATA_REV_MO_PAYG_W4 ,0) as DATA_REV_MO_PAYG_W4
,coalesce(DATA_REV_MO_PAYG_M1 ,0) as DATA_REV_MO_PAYG_M1
,coalesce(DATA_REV_MO_PAYG_M2 ,0) as DATA_REV_MO_PAYG_M2
,coalesce(DATA_REV_MO_PAYG_M3 ,0) as DATA_REV_MO_PAYG_M3
,coalesce(NBER_SOUSC_W1 ,0) as NBER_SOUSC_W1
,coalesce(NBER_SOUSC_W2 ,0) as NBER_SOUSC_W2
,coalesce(NBER_SOUSC_W3 ,0) as NBER_SOUSC_W3
,coalesce(NBER_SOUSC_W4 ,0) as NBER_SOUSC_W4
,coalesce(NBER_SOUSC_M1 ,0) as NBER_SOUSC_M1
,coalesce(NBER_SOUSC_M2 ,0) as NBER_SOUSC_M2
,coalesce(NBER_SOUSC_M3 ,0) as NBER_SOUSC_M3
,coalesce(REV_SOUSC_W1 ,0) as REV_SOUSC_W1
,coalesce(REV_SOUSC_W2 ,0) as REV_SOUSC_W2
,coalesce(REV_SOUSC_W3 ,0) as REV_SOUSC_W3
,coalesce(REV_SOUSC_W4 ,0) as REV_SOUSC_W4
,coalesce(REV_SOUSC_M1 ,0) as REV_SOUSC_M1
,coalesce(REV_SOUSC_M2 ,0) as REV_SOUSC_M2
,coalesce(REV_SOUSC_M3 ,0) as REV_SOUSC_M3

```

-----KPI_INcomING-----

```

,coalesce(VOICE_USAGE_IN_W1 ,0) as NBER_CALLS_IN_IW1
,coalesce(VOICE_USAGE_IN_W2 ,0) as NBER_CALLS_IN_W2
,coalesce(VOICE_USAGE_IN_W3 ,0) as NBER_CALLS_IN_W3
,coalesce(VOICE_USAGE_IN_W4 ,0) as NBER_CALLS_IN_W4
,coalesce(VOICE_USAGE_IN_M1 ,0) as NBER_CALLS_IN_M1
,coalesce(VOICE_USAGE_IN_M2 ,0) as NBER_CALLS_IN_M2
,coalesce(VOICE_USAGE_IN_M3 ,0) as NBER_CALLS_IN_M3
,coalesce(VOICE_USAGE_Offnet_IN_W1 ,0) as NBER_CALLS_Offnet_IN_W1
,coalesce(VOICE_USAGE_Offnet_IN_W2 ,0) as NBER_CALLS_Offnet_IN_W2
,coalesce(VOICE_USAGE_Offnet_IN_W3 ,0) as NBER_CALLS_Offnet_IN_W3
,coalesce(VOICE_USAGE_Offnet_IN_W4 ,0) as NBER_CALLS_Offnet_IN_W4
,coalesce(VOICE_USAGE_Offnet_IN_M1 ,0) as NBER_CALLS_Offnet_IN_M1
,coalesce(VOICE_USAGE_Offnet_IN_M2 ,0) as NBER_CALLS_Offnet_IN_M2
,coalesce(VOICE_USAGE_Offnet_IN_M3 ,0) as NBER_CALLS_Offnet_IN_M3

```

(5)

```

,coalesce(SMS_IN_W1 ,0) as SMS_IN_W1
,coalesce(SMS_IN_W2 ,0) as SMS_IN_W2
,coalesce(SMS_IN_W3 ,0) as SMS_IN_W3
,coalesce(SMS_IN_W4 ,0) as SMS_IN_W4
,coalesce(SMS_IN_M1 ,0) as SMS_IN_M1
,coalesce(SMS_IN_M2 ,0) as SMS_IN_M2
,coalesce(SMS_IN_M3 ,0) as SMS_IN_M3

```

-----KPI_refill-----

```

,coalesce(Amount_day_mINus_1 ,0) as Amount_day_mINus_1
,coalesce(NBER_Refill_W1 ,0) as NBER_Refill_W1
,coalesce(NBER_Refill_W2 ,0) as NBER_Refill_W2
,coalesce(NBER_Refill_W3 ,0) as NBER_Refill_W3
,coalesce(NBER_Refill_W4 ,0) as NBER_Refill_W4
,coalesce(NBER_Refill_M1 ,0) as NBER_Refill_M1
,coalesce(NBER_Refill_M2 ,0) as NBER_Refill_M2
,coalesce(NBER_Refill_M3 ,0) as NBER_Refill_M3
,coalesce(REV_Refill_W1 ,0) as REV_Refill_W1
,coalesce(REV_Refill_W2 ,0) as REV_Refill_W2
,coalesce(REV_Refill_W3 ,0) as REV_Refill_W3
,coalesce(REV_Refill_W4 ,0) as REV_Refill_W4
,coalesce(REV_Refill_M1 ,0) as REV_Refill_M1
,coalesce(REV_Refill_M2 ,0) as REV_Refill_M2
,coalesce(REV_Refill_M3 ,0) as REV_Refill_M3
,coalesce(REV_Refill_M4 ,0) as REV_Refill_M4
,coalesce(REV_Refill_M5 ,0) as REV_Refill_M5
,coalesce(REV_Refill_M6 ,0) as REV_Refill_M6
,coalesce(NBER_Refill_Flexy_W1 ,0) as NBER_Refill_Flexy_W1
,coalesce(NBER_Refill_Flexy_W2 ,0) as NBER_Refill_Flexy_W2
,coalesce(NBER_Refill_Flexy_W3 ,0) as NBER_Refill_Flexy_W3
,coalesce(NBER_Refill_Flexy_W4 ,0) as NBER_Refill_Flexy_W4
,coalesce(NBER_Refill_Flexy_M1 ,0) as NBER_Refill_Flexy_M1
,coalesce(NBER_Refill_Flexy_M2 ,0) as NBER_Refill_Flexy_M2
,coalesce(NBER_Refill_Flexy_M3 ,0) as NBER_Refill_Flexy_M3

```

(6)

```

,coalesce(REV_Refill_Flexy_W1 ,0) as REV_Refill_Flexy_W1
,coalesce(REV_Refill_Flexy_W2 ,0) as REV_Refill_Flexy_W2
,coalesce(REV_Refill_Flexy_W3 ,0) as REV_Refill_Flexy_W3
,coalesce(REV_Refill_Flexy_W4 ,0) as REV_Refill_Flexy_W4
,coalesce(REV_Refill_Flexy_M1 ,0) as REV_Refill_Flexy_M1
,coalesce(REV_Refill_Flexy_M2 ,0) as REV_Refill_Flexy_M2
,coalesce(REV_Refill_Flexy_M3 ,0) as REV_Refill_Flexy_M3
,coalesce(NBER_Refill_Voucher_W1 ,0) as NBER_Refill_Voucher_W1
,coalesce(NBER_Refill_Voucher_W2 ,0) as NBER_Refill_Voucher_W2
,coalesce(NBER_Refill_Voucher_W3 ,0) as NBER_Refill_Voucher_W3
,coalesce(NBER_Refill_Voucher_W4 ,0) as NBER_Refill_Voucher_W4
,coalesce(NBER_Refill_Voucher_M1 ,0) as NBER_Refill_Voucher_M1
,coalesce(NBER_Refill_Voucher_M2 ,0) as NBER_Refill_Voucher_M2
,coalesce(NBER_Refill_Voucher_M3 ,0) as NBER_Refill_Voucher_M3
,coalesce(REV_Refill_Voucher_W1 ,0) as REV_Refill_Voucher_W1
,coalesce(REV_Refill_Voucher_W2 ,0) as REV_Refill_Voucher_W2
,coalesce(REV_Refill_Voucher_W3 ,0) as REV_Refill_Voucher_W3
,coalesce(REV_Refill_Voucher_W4 ,0) as REV_Refill_Voucher_W4
,coalesce(REV_Refill_Voucher_M1 ,0) as REV_Refill_Voucher_M1
,coalesce(REV_Refill_Voucher_M2 ,0) as REV_Refill_Voucher_M2
,coalesce(REV_Refill_Voucher_M3 ,0) as REV_Refill_Voucher_M3
,coalesce(REV_Refill_Max_M1 ,0) as REV_Refill_Max_M1
,coalesce(REV_Refill_Max_M2 ,0) as REV_Refill_Max_M2
,coalesce(REV_Refill_Max_M3 ,0) as REV_Refill_Max_M3
,coalesce(REV_Refill_Max_M4 ,0) as REV_Refill_Max_M4
,coalesce(REV_Refill_Max_M5 ,0) as REV_Refill_Max_M5
,coalesce(REV_Refill_Max_M6 ,0) as REV_Refill_Max_M6
,coalesce(potentiel ,0) as POTENTIEL
,coalesce(variance_refill ,0) as VARIANCE_REFILL
,coalesce(ecartype_rev ,0) as ECARTYPE_REV
,coalesce(gap ,0) as GAP

```

```

from P_EDW_TMP.TARGET_BASE TB
left join P_EDW_TMP.ADS_OutgoINg_Month OM on TB.SUBS_KEY=OM.SUBSCRIBER_ID
left join P_EDW_TMP.ADS_OutgoINg_Week OW on TB.SUBS_KEY=OW.SUBSCRIBER_ID
left join P_EDW_TMP.ADS_INCOMING_IG on TB.SUBS_KEY=IG.SUBSCRIBER_ID
left join P_EDW_TMP.ADS_Refill RE on TB.SUBS_KEY=RE.SUBS_ID )
with no data
primary INdex (Subs_id)
PARTITION BY RANGE_N(Run_Date BETWEEN DATE '2006-01-01' AND DATE '2020-12-31' EACH INTERVAL '1' DAY , NO RANGE OR UNKNOWN)

```

Annexe 25: Script de création de l'ADS

10) Le nom et le type des KPI (attributs de la table ADS)

	Column Name	Type	Com ment	Nulla ble	Format
1	Run_Date	DA	?	Y	YY/MM/DD
2	Subs_id	I	?	Y	-(10)9
3	FRST_CALL_DT	TS	?	Y	YYYY-MM-DDBHH:MI:SS
4	LST_CALL_OUT_DT	TS	?	Y	YYYY-MM-DDBHH:MI:SS
5	LST_CALL_IN_DT	TS	?	Y	YYYY-MM-DDBHH:MI:SS
6	ARPU_M1	D	?	Y	-----9999
7	ARPU_M2	D	?	Y	-----9999
8	ARPU_M3	D	?	Y	-----9999
9	VOICE_USAGE_W1	D	?	Y	-----99
10	VOICE_USAGE_W2	D	?	Y	-----99
11	VOICE_USAGE_W3	D	?	Y	-----99
12	VOICE_USAGE_W4	D	?	Y	-----99
13	VOICE_USAGE_M1	D	?	Y	-----99
14	VOICE_USAGE_M2	D	?	Y	-----99
15	VOICE_USAGE_M3	D	?	Y	-----99
16	NBER_CALLS_W1	D	?	Y	-----9
17	NBER_CALLS_W2	D	?	Y	-----9
18	NBER_CALLS_W3	D	?	Y	-----9
19	NBER_CALLS_W4	D	?	Y	-----9
20	NBER_CALLS_M1	D	?	Y	-----9
21	NBER_CALLS_M2	D	?	Y	-----9
22	NBER_CALLS_M3	D	?	Y	-----9
23	VOICE_REV_W1	D	?	Y	-----9999
24	VOICE_REV_W2	D	?	Y	-----9999
25	VOICE_REV_W3	D	?	Y	-----9999
26	VOICE_REV_W4	D	?	Y	-----9999
27	VOICE_REV_M1	D	?	Y	-----9999
28	VOICE_REV_M2	D	?	Y	-----9999
29	VOICE_REV_M3	D	?	Y	-----9999
30	VOICE_USAGE_ON_NET_W1	D	?	Y	-----99
31	VOICE_USAGE_ON_NET_W2	D	?	Y	-----99
32	VOICE_USAGE_ON_NET_W3	D	?	Y	-----99
33	VOICE_USAGE_ON_NET_W4	D	?	Y	-----99
34	VOICE_USAGE_ON_NET_M1	D	?	Y	-----99
35	VOICE_USAGE_ON_NET_M2	D	?	Y	-----99
36	VOICE_USAGE_ON_NET_M3	D	?	Y	-----99
37	NBER_CALLS_ON_NET_W1	D	?	Y	-----9
38	NBER_CALLS_ON_NET_W2	D	?	Y	-----9
39	NBER_CALLS_ON_NET_W3	D	?	Y	-----9
40	NBER_CALLS_ON_NET_W4	D	?	Y	-----9
41	NBER_CALLS_ON_NET_M1	D	?	Y	-----9

	Column Name	Type	Com ment	Nulla ble	Format
193	REV_Refill_M5	D	?	Y	-----99
194	REV_Refill_M6	D	?	Y	-----99
195	NBER_Refill_Flexy_W1	I	?	Y	-(10)9
196	NBER_Refill_Flexy_W2	I	?	Y	-(10)9
197	NBER_Refill_Flexy_W3	I	?	Y	-(10)9
198	NBER_Refill_Flexy_W4	I	?	Y	-(10)9
199	NBER_Refill_Flexy_M1	I	?	Y	-(10)9
200	NBER_Refill_Flexy_M2	I	?	Y	-(10)9
201	NBER_Refill_Flexy_M3	I	?	Y	-(10)9
202	REV_Refill_Flexy_W1	D	?	Y	-----99
203	REV_Refill_Flexy_W2	D	?	Y	-----99
204	REV_Refill_Flexy_W3	D	?	Y	-----99
205	REV_Refill_Flexy_W4	D	?	Y	-----99
206	REV_Refill_Flexy_M1	D	?	Y	-----99
207	REV_Refill_Flexy_M2	D	?	Y	-----99
208	REV_Refill_Flexy_M3	D	?	Y	-----99
209	NBER_Refill_Voucher_W1	I	?	Y	-(10)9
210	NBER_Refill_Voucher_W2	I	?	Y	-(10)9
211	NBER_Refill_Voucher_W3	I	?	Y	-(10)9
212	NBER_Refill_Voucher_W4	I	?	Y	-(10)9
213	NBER_Refill_Voucher_M1	I	?	Y	-(10)9
214	NBER_Refill_Voucher_M2	I	?	Y	-(10)9
215	NBER_Refill_Voucher_M3	I	?	Y	-(10)9
216	REV_Refill_Voucher_W1	D	?	Y	-----99
217	REV_Refill_Voucher_W2	D	?	Y	-----99
218	REV_Refill_Voucher_W3	D	?	Y	-----99
219	REV_Refill_Voucher_W4	D	?	Y	-----99
220	REV_Refill_Voucher_M1	D	?	Y	-----99
221	REV_Refill_Voucher_M2	D	?	Y	-----99
222	REV_Refill_Voucher_M3	D	?	Y	-----99
223	REV_Refill_Max_M1	D	?	Y	-----99
224	REV_Refill_Max_M2	D	?	Y	-----99
225	REV_Refill_Max_M3	D	?	Y	-----99
226	REV_Refill_Max_M4	D	?	Y	-----99
227	REV_Refill_Max_M5	D	?	Y	-----99
228	REV_Refill_Max_M6	D	?	Y	-----99
229	POTENTIEL	D	?	Y	-----99
230	VARIANCE_REFILL	F	?	Y	-9.999999999999999E-999
231	ECARTYPE_REV	F	?	Y	-9.999999999999999E-999
232	GAP	I	?	Y	-(10)9

Annexe 26: Le nom et le type des KPI (attributs de la table ADS)

11) Script de la macro de chargement de la table « ADS_FINAL » nommée « LOAD_ADS_FINAL »

```

replace macro p_edw_tmp. LOAD_ADS_FINAL
AS
(
delete from p_edw_tmp.ADS_Final where Run_date= date-extract(day from date)
insert into p_edw_tmp.ADS_Final
SELECT
date-extract(day from date) Run_Date
, SUBS_KEY as Subs_id
, FRST_CALL_DT
, LST_CALL_OUT_DT
, LST_CALL_IN_DT
-----Outgoing_KPI-----
, coalesce( ARPU_M1 , 0) as ARPU_M1
, coalesce( ARPU_M2 , 0) as ARPU_M2
, coalesce( ARPU_M3 , 0) as ARPU_M3
, coalesce( VOICE_USAGE_W1 , 0) as VOICE_USAGE_W1
, coalesce ( VOICE_USAGE_W2 ,0) as VOICE_USAGE_W2
, coalesce(VOICE_USAGE_W3 ,0) as VOICE_USAGE_W3
, coalesce(VOICE_USAGE_W4 ,0) as VOICE_USAGE_W4
, coalesce(VOICE_USAGE_M1 ,0) as VOICE_USAGE_M1
, coalesce(VOICE_USAGE_M2 ,0) as VOICE_USAGE_M2
, coalesce(VOICE_USAGE_M3 ,0) as VOICE_USAGE_M3
, coalesce(NBER_CALLS_W1 ,0) as NBER_CALLS_W1
, coalesce(NBER_CALLS_W2 ,0) as NBER_CALLS_W2
, coalesce(NBER_CALLS_W3 ,0) as NBER_CALLS_W3
, coalesce(NBER_CALLS_W4 ,0) as NBER_CALLS_W4
, coalesce(NBER_CALLS_M1 ,0) as NBER_CALLS_M1
, coalesce(NBER_CALLS_M2 ,0) as NBER_CALLS_M2
, coalesce(NBER_CALLS_M3 ,0) as NBER_CALLS_M3
, coalesce( VOICE_REV_W1 ,0) as VOICE_REV_W1
, coalesce( VOICE_REV_W2 ,0) as VOICE_REV_W2
, coalesce( VOICE_REV_W3 ,0) as VOICE_REV_W3
, coalesce( VOICE_REV_W4 ,0) as VOICE_REV_W4
, coalesce( VOICE_REV_M1 ,0) as VOICE_REV_M1
, coalesce( VOICE_REV_M2 ,0) as VOICE_REV_M2
, coalesce( VOICE_REV_M3 ,0) as VOICE_REV_M3
, coalesce(VOICE_USAGE_ON_NET_W1 ,0) as VOICE_USAGE_ON_NET_W1
, coalesce(VOICE_USAGE_ON_NET_W2 ,0) as VOICE_USAGE_ON_NET_W2
, coalesce(VOICE_USAGE_ON_NET_W3 ,0) as VOICE_USAGE_ON_NET_W3
, coalesce(VOICE_USAGE_ON_NET_W4 ,0) as VOICE_USAGE_ON_NET_W4
, coalesce(REV_Refill_Flexy_W1 ,0) as REV_Refill_Flexy_W1
, coalesce(REV_Refill_Flexy_W2 ,0) as REV_Refill_Flexy_W2
, coalesce(REV_Refill_Flexy_W3 ,0) as REV_Refill_Flexy_W3
, coalesce(REV_Refill_Flexy_W4 ,0) as REV_Refill_Flexy_W4
, coalesce(REV_Refill_Flexy_M1 ,0) as REV_Refill_Flexy_M1
, coalesce(REV_Refill_Flexy_M2 ,0) as REV_Refill_Flexy_M2
, coalesce(REV_Refill_Flexy_M3 ,0) as REV_Refill_Flexy_M3
, coalesce(NBER_Refill_Voucher_W1 ,0) as NBER_Refill_Voucher_W1
, coalesce(NBER_Refill_Voucher_W2 ,0) as NBER_Refill_Voucher_W2
, coalesce(NBER_Refill_Voucher_W3 ,0) as NBER_Refill_Voucher_W3
, coalesce(NBER_Refill_Voucher_W4 ,0) as NBER_Refill_Voucher_W4
, coalesce(NBER_Refill_Voucher_M1 ,0) as NBER_Refill_Voucher_M1
, coalesce(NBER_Refill_Voucher_M2 ,0) as NBER_Refill_Voucher_M2
, coalesce(NBER_Refill_Voucher_M3 ,0) as NBER_Refill_Voucher_M3
, coalesce(REV_Refill_Voucher_W1 ,0) as REV_Refill_Voucher_W1
, coalesce(REV_Refill_Voucher_W2 ,0) as REV_Refill_Voucher_W2
, coalesce(REV_Refill_Voucher_W3 ,0) as REV_Refill_Voucher_W3
, coalesce(REV_Refill_Voucher_W4 ,0) as REV_Refill_Voucher_W4
, coalesce(REV_Refill_Voucher_M1 ,0) as REV_Refill_Voucher_M1
, coalesce(REV_Refill_Voucher_M2 ,0) as REV_Refill_Voucher_M2
, coalesce(REV_Refill_Voucher_M3 ,0) as REV_Refill_Voucher_M3
, coalesce(REV_Refill_Max_M1 ,0) as REV_Refill_Max_M1
, coalesce(REV_Refill_Max_M2 ,0) as REV_Refill_Max_M2
, coalesce(REV_Refill_Max_M3 ,0) as REV_Refill_Max_M3
, coalesce(REV_Refill_Max_M4 ,0) as REV_Refill_Max_M4
, coalesce(REV_Refill_Max_M5 ,0) as REV_Refill_Max_M5
, coalesce(REV_Refill_Max_M6 ,0) as REV_Refill_Max_M6
, coalesce(potentiel ,0) as POTENTIEL
, coalesce(variance_refill ,0) as VARIANCE_REFILL
, coalesce(ecartype_rev ,0) as ECARTYPE_REV
, coalesce(gap ,0) as GAP
)
from P_EDW_TMP.TARGET_BASE TB
left join P_EDW_TMP.ADS_Outgoing_Month OM on TB.SUBS_KEY=OM.SUBSCRIBER_ID
left join P_EDW_TMP.ADS_Outgoing_Week OW on TB.SUBS_KEY=OW.SUBSCRIBER_ID
left join P_EDW_TMP.ADS_INCOMING IG on TB.SUBS_KEY=IG.SUBSCRIBER_ID
left join P_EDW_TMP.ADS_Refill RE on TB.SUBS_KEY=RE.SUBS_ID;
);

```

Annexe 7560: Script de la macro de chargement de la table « ADS_FINAL » nommée « LOAD_ADS_FINAL »

12)

```

create MULTISET TABLE p_edw_tmp.ADS_Kmeans as
(select
a.Subs_id
,NBER_CALLS_MI
,NBER_CALLS_ON_NET_MI -----ON NET et OFF NET sont des appels Outgoing-----
,NBER_CALLS_OFF_NET_MI
,NBER_CALLS_IN_MI
,NBER_CALLS_OFFNET_IN_MI
,NBER_CALLS_INT_MI
,NBER_SMS_ON_NET_MI -----ON NET et OFF NET sont des outgoing-----
,NBER_SMS_OFF_NET_MI
,SMS_IN_MI
,DATA_USAGE_MO_MI+ DATA_USAGE_MO_PAYG_MI as VOLUME_DATA
,case
.....
when cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) <=3 then 'New_subs'
when cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4)> 3 and cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) <= 6 then '>3 M and <= 6 M'
when cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) > 6 and cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) <=12 then '>6 M and <= 12 M'
when cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) > 12 and cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) <= 24 then '>12M and <= 24 M'
when cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) > 24 and cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) <= 60 then '>3 Year and <= 5 Year'
when cast(run_date as date) - cast (a.FRST_CALL_DT as date) month (4) > 60 then '> 5 Year'
.....
else 0
end TENURE
,CASE
WHEN Coalesce(ARFU_MI,0) <= 0 THEN 'F - ZERO'
WHEN Coalesce(ARFU_MI,0) > 0 AND ARFU_MI <= 500 THEN 'E - VLV'
WHEN Coalesce(ARFU_MI,0) > 500 AND ARFU_MI <= 1000 THEN 'D - LIV'
WHEN Coalesce(ARFU_MI,0) > 1000 AND ARFU_MI <= 1500 THEN 'C - MV'
WHEN Coalesce(ARFU_MI,0) > 1500 AND ARFU_MI <= 2500 THEN 'B - BV'
WHEN Coalesce(ARFU_MI,0) > 2500 THEN 'A - VEV'
ELSE 'Not_detected' END Value_Segment
,CASE
WHEN DEVICETYPE like any ('%Smartphone%', '%Tablet%', '%Tablet%') THEN 'SMARTPHONE'
WHEN DEVICETYPE like any ('%Feature Phone%', '%Feature Phone%') THEN 'FEATURE PHONE'
WHEN DEVICETYPE like any ('%Basic Phone%', '%Basic Phone%') THEN 'BASIC PHONE'
ELSE 'OTHER'
END DeviceType
,case
when GLOBAL_PROFILE like 'Mifi%' then 'Mifi'
when GLOBAL_PROFILE like '%good%' then 'Djezzy GOOD'
when GLOBAL_PROFILE like any ('%Djezzy_Go%', '%DjezzyGo%', '%Djz_Go%') then 'Djezzy GO'
when GLOBAL_PROFILE in ('DjezzyCarte_NEW', 'Isbak') then 'Djezzy Carte New'
when GLOBAL_PROFILE in ('Djezzy Carte', 'Djezzycarte', 'Djezzycarte_Sos', 'Djezzycartedormant') then 'Djezzy Carte'
when GLOBAL_PROFILE like '%BAYNA%' then 'Djezzy BAYNA'
when GLOBAL_PROFILE in ('DJEZZY_HAYLA', 'Djezzyhayla', '7150', 'Lgd') then 'Djezzy HAYLA'
when GLOBAL_PROFILE like '%play%' then 'Djezzy PLAY'
when GLOBAL_PROFILE like '%Hadra%' then 'Djezzy Hadra'
when GLOBAL_PROFILE like '%DJEZZY_HAYLABEZZEF%' then 'Djezzy HAYLABEZZEF'
when GLOBAL_PROFILE like '%Djezzypack%' then 'Djezzy Pack'
else GLOBAL_PROFILE
end SubscriberProfile

from p_edw_tmp.ADS_Final a
left join ( sel * from p_edw_tmp.Dim_Cell_loc) b on a .Subs_id=b.Subs_id
left join p_edw_tmp.Dim_Handset c on c.TAC=substr(b.Last_Most_Used_Imei,1,8)
left join p_edw_tmp.DIM_SUBS z on a.Subs_id=z.SUBS_KEY
left join p_edw_tmp.DIM_OFFER y on y.OFFER_KEY= z.SUBS_OFFNG_KEY
) with data
primary index(Subs_id)

```

Annexe 28: Script création de la table « ADS_Kmeans »

13) Script de la fonction « remove_outliers » appliquée sur chaque variable quantitative

```
##### NBER_CALLS_M1 #####
datset_treat$NBER_CALLS_M1=remove_outliers(datset_treat$NBER_CALLS_M1)
boxplot.default(datset_treat$NBER_CALLS_M1)

##### NBER_CALLS_ON_NET_M1 #####
datset_treat$NBER_CALLS_ON_NET_M1=remove_outliers(datset_treat$NBER_CALLS_ON_NET_M1)
boxplot.default(datset_treat$NBER_CALLS_ON_NET_M1)

##### NBER_CALLS_OFF_NET_M1 #####
datset_treat$NBER_CALLS_OFF_NET_M1=remove_outliers(datset_treat$NBER_CALLS_OFF_NET_M1)
boxplot.default(datset_treat$NBER_CALLS_OFF_NET_M1)

##### NBER_CALLS_OFF_NET_M1 #####
datset_treat$NBER_CALLS_INT_M1=remove_outliers(datset_treat$NBER_CALLS_INT_M1)
boxplot.default(datset_treat$NBER_CALLS_INT_M1)

##### NBER_SMS_M1 #####
datset_treat$NBER_SMS_M1=remove_outliers(datset_treat$NBER_SMS_M1)
boxplot.default(datset_treat$NBER_SMS_M1)

##### NBER_SMS_ON_NET_M1 #####
datset_treat$NBER_SMS_ON_NET_M1=remove_outliers(datset_treat$NBER_SMS_ON_NET_M1)
boxplot.default(datset_treat$NBER_SMS_ON_NET_M1)
datset_treat$NBER_SMS_ON_NET_M1

##### NBER_SMS_OFF_NET_M1 #####
datset_treat$NBER_SMS_OFF_NET_M1=remove_outliers(datset_treat$NBER_SMS_OFF_NET_M1)

##### NBER_SMS_OFF_NET_M1 #####
datset_treat$SMS_IN_M1=remove_outliers(datset_treat$SMS_IN_M1)

##### NBER_SMS_OFF_NET_M1 #####
datset_treat$Volume_data=remove_outliers(datset_treat$Volume_data)
```

Annexe 29: Script de la fonction « remove_outliers » appliquée sur chaque variable quantitative

14) Script de création de la table « Traffic_Behaviour »

```
-----Création d'une table descriptive sur la segmentation comportementale des abonnés -----  
  
create multiset table p_edw_tmp.Traffic_Behaviour  
(Cluster_Num integer  
, Cluster_Name Varchar (50)  
, Cluster_description Varchar (250)  
,cluster_characteristics varchar ( 250)  
)  
primary index (cluster_num)
```

Annexe 7568: Script de création de la table « Traffic_Behaviour »

15) Script de chargement de la table « Traffic_Behaviour »

```
insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)  
|values (1,'Calls and SMS loyalists','cette classe contient les non-user data, à fort traffic des appels (In et Out), moyen traffic sms Out et élevé sms In ',  
'Effectif : 121 344 (1%)  
51% des abonnés ont plus de 5 ans  
-l'offre dominante est DJEZZY CARTE NEW  
-62% dépensent plus de 1500 DA  
-82% des abonnés n'ont pas de smartphone  
-AVG Data : 0,35 MB  
-AVG Calls Out : 416  
-AVG Calls In : 592  
-AVG SMS OUT : 26  
' ) ;  
insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)  
|values (2,'Data and SMS Likers','cette classe est dominée par les souscripteurs user data avec un traffic appels In et Out moyen, fort traffic sms Out off net',  
'Effectif : 744 877 (5%)  
83% des abonnés ont des smartphones  
-64% sont des LV  
-l'offre dominante est HAYLA BEZZEF  
-l'âge des abonnés est diversifié  
-AVG Data : 4,1 GB  
-AVG Calls Out : 97  
-AVG Calls In : 118  
-AVG SMS OUT : 22') ;  
insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)  
|values (3,'Mass Sleeping Customers',  
'cette classe contient les abonnés ayant fait le moins de traffic durant le mois respectivement en termes de : Nombre d appels, nombre SMS et de données ',  
'1/3 (35%) des abonnés de cette classe ont plus de 5 ans  
42% des abonnés de cette classe ont moins d'une année  
, ces memes pourcentages sont retrouvés pour le type de téléphone 34 % basic phone et 43% smartphone  
-97% des abonnés sont des LV <1000 DA  
-AVG Data : 0,19 MB  
-AVG Calls : 8  
-AVG SMS OUT : 0,75  
' ) ;
```

```

insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)
]values (4,'Traditional Customers','cette classe est dominée par les soucripteurs non users , avec un faible trafic appels et sms','Effectif : 2 722 627 (18%)
60% des abonnés de cette classe ont plus de 5 ans
-70% des abonnés de cette classe n ont pas de smartphone
-90 % des abonnés de cette classe sont des LV(low value) ( <=1000 DA)
-les offres sont dominées par DJEZZY GOOD (1000 DA) ,DJEZZY GO (400 DA)
-AVG Data : 0,32 MB
-AVG Calls : 46
-AVG SMS OUT : 2
');
insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)
]values (5,'Phone Addicts',
'cette classe contient les abonnés ayant fait le plus de trafic sortant durant le mois respectivement en termes de : Nombre d appels, nombre d SMS et de données',
'Effectif : 510 483 (4%)
50% des abonnés ont moins dl ans
-85% des abonnés ont des smartphones
-les offres dominantes sont HAYLA et HAYLA BEZZAF
-65% des abonnées sont des HV >=1500 DA
-AVG Data : 4,8 GB
-AVG Calls : 427
-AVG SMS OUT : 70
');
insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)
]values (6,'Data enthusiasts','cette classe contient les abonnés dont le trafic est dominé par l utilisation des données mobiles ','Effectif : 1 431 949 (10%)
-56% des abonnés ont moins d l an
-80% des abonnés sont des LV (44% LV)
-1 offre dominante est HAYLA ET HAYLABEZZAF
-70% des abonnés ont des smartphone
-AVG Data : 3,6 GB
-AVG Calls Out : 30
-AVG Calls In : 25
-AVG SMS OUT : 5
');
insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)
]values (7,'Traditional Out Talkers','cette classe est dominée par les abonnés à fort trafic out, plus ce sont des non-user data, faible trafic sms ',
'Effectif: 678 295 (5%)
51% des abonnés ont plus de 5 ans
-1 offre dominante est DJEZZY CARTE NEW
-62% dépensent plus de 1500 DA
-82% des abonnés n ont pas de smartphone
-AVG Data : 0,38 MB
-AVG Calls Out : 310
-AVG Calls In : 162
-AVG SMS OUT : 16
');
insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)
]values (8,'Traditional In and Out talkers','Non user data, off-netteurs Out avec un trafic in des appels élevé , et faible trafic sms',
'Effectif : 636 993 (4%)
53% des abonnés ont plus de 5 ans
-67% des abonnés n ont pas de smartphone
dominée par les offres DJEZZY GO et DJEZZY GOOD
-88% des abonnés sont des LV
-AVG Data : 0,36 MB
-AVG Calls Out : 56
-AVG Calls In : 264
-AVG SMS OUT : 6
');
insert into p_edw_tmp.Traffic_Behaviour ( Cluster_Num,Cluster_Name,Cluster_description,cluster_characteristics)
]values (9,'Data lovers and On Net Talkers ','On-netteurs(appels out), trafic SMS moyen et data-user','Effectif : 853 346 (6%)
82% des abonnés ont un smartphone
dominés par HAYLA ET HAYLABEZZEF
-48% sont des HV
-AVG Data : 5,7 GB
-AVG Calls : 285
-AVG Calls On Net : 230
-AVG SMS OUT : 38
');

```

Annexe 31: Script de chargement de la table « Traffic_Behaviour »

	Subs_id	VOICE_USAGE_M1	NBER_CALLS_M1	VOICE_USAGE_ON_NET_M1	NBER_CALLS_ON_NET_M1	VOICE_USAGE_OFF_NET_M1	NBER_CALLS_OFF_NET_M1	VOICE_USAGE_INT_M1	NBER_CALLS_INT_M1	NBER_SMS_M1	REV_SMS_M1	NBER_SMS_ON_NET_M1	NBER_SMS_OFF_NET_M1	DATA_USAGE_MO_M1
1	81 133 981													
2	10 850 018													
3	38 395 749													
4	121 032 942													
5	1 834 132													
6	112 017 056													
7	135 054 844													
8	136 723 516													
9	132 480 036													
10	11 612 554													
11	138 656 651													
12	134 148 708													
13	129 640 765													
14	80 725 918													
15	135 817 380													
16	14 949 898													
17	103 763 706													
18	137 486 052													
19	5 934 012													
20	105 432 378													
21	39 656 358													
22	28 971 800													
23	139 154 724													
24	58 186 203													
25	115 244 671													
26	21 922 920													
27	121 421 286													
28	553 804													
29	138 282 459													
30	133 774 516													
31	123 089 958													
32	112 405 400													
33	135 443 188													
34	124 758 630													
35	10 067 763													

S_I	ECARTY PE_REV	GAP
4	0,00	15
6	109,93	15
4	0,00	45
5	0,00	32
7	40,00	22
0	0,00	0
1	0,00	45
1	0,00	45
12	0,00	16
9	502,89	17
14	0,00	45
0	0,00	0
47	406,94	9
2	0,00	11
6	0,00	3
141	352,67	17
6	0,00	24
0	0,00	45
9	0,00	16
20	184,11	6
22	396,32	8
0	0,00	22
0	0,00	45
30	25,09	2
11	25,34	3
5	37,25	7
10	235,70	45
6	0,00	0
16	0,00	45
2	0,00	45
0	4,71	45
0	0,00	0
22	466,96	18
7	47,14	31
14	20,12	2

Annexe 32: La table ADS_FINAL chargée en Mars 2019