



Département : Génie Industriel

Entreprise : Société Générale

## **Mémoire de projet de fin d'études**

Pour l'obtention du diplôme d'Ingénieur d'État en Génie Industriel  
Option Data Science et Intelligence Artificielle

---

# **Mesure de la Performance et Prédiction de la Rentabilité des Agences Bancaires par l'Intelligence Artificielle et la Business Intelligence : Application à la Société Générale Algérie**

---

Réalisé par :

**Mr. Hamza BERRABAH**  
**Mr. Billel NOURINE**

Sous la direction de :

**Mr. Youcef HAMANA**  
**Mr. Oussama ARKI**

Présenté et soutenu publiquement le 6 juillet 2024

### **Composition du jury :**

Président	Mr. Iskander ZOUAGHI	MCA	ENP
Examineur	Mr. Hakim FOURAR-LAIDI	MCA	ENP
Promoteur	Mr. Oussama ARKI	MCA	ENP





Département : Génie Industriel

Entreprise : Société Générale

## **Mémoire de projet de fin d'études**

Pour l'obtention du diplôme d'Ingénieur d'État en Génie Industriel  
Option Data Science et Intelligence Artificielle

---

# **Mesure de la Performance et Prédiction de la Rentabilité des Agences Bancaires par l'Intelligence Artificielle et la Business Intelligence : Application à la Société Générale Algérie**

---

Réalisé par :

**Mr. Hamza BERRABAH**  
**Mr. Billel NOURINE**

Sous la direction de :

**Mr. Youcef HAMANA**  
**Mr. Oussama ARKI**

Présenté et soutenu publiquement le 6 juillet 2024

### **Composition du jury :**

Président	Mr. Iskander ZOUAGHI	MCA	ENP
Examineur	Mr. Hakim FOURAR-LAIDI	MCA	ENP
Promoteur	Mr. Oussama ARKI	MCA	ENP

# *Dédicaces*

*Je souhaite sincèrement dédier ce travail :*

*À la personne la plus chère à mon cœur, ma mère, cette sacrée femme qui fait constamment des sacrifices pour me voir réaliser mes ambitions et atteindre mes buts.*

*À mon père, ce brave héros que j'admire beaucoup, occupant une place précieuse dans mon cœur et qui a été le socle solide de mon éducation, instaurant les valeurs qui m'ont guidé tout au long de mon parcours.*

*À mes adorables frère et sœur, Oussama et Asma, que j'aime plus que tout et qui ont attentivement suivi chaque étape de mon évolution avec une admiration qui me touche profondément. Mon rêve le plus cher est de vous voir réussir en vous soutenant de tout mon cœur.*

*À ma nièce et mon neveu, Ryma et Anes, que j'aime énormément. Que Dieu les garde.*

*À mes cousins et cousines, tantes et oncles.*

*À la mémoire de mes grands-pères et de ma grand-mère paternelle, avec qui j'ai passé des étés inoubliables, et en honneur de ma grand-mère maternelle encore avec nous, que Dieu la garde.*

*À mes chers amis qui ont été présents à chaque étape de ma vie.*

*À l'ensemble de mes camarades de la promotion Data Science & IA.*

*À mon binôme du primaire, collège, lycée et avec qui j'ai partagé ce travail, Hamza. Une très belle manière de conclure ce chapitre de notre parcours.*

*Et finalement, à tous ceux qui ont vu en moi un potentiel, qui ont cru en mes compétences et en ma réussite.*

***Billel***

# *Dédicaces*

*Je dédie ce travail :*

*À ma mère,*

*Ton amour inconditionnel, ton soutien sans faille et tes innombrables sacrifices ont été les fondations de mon succès. Ta présence lumineuse a guidé chacun de mes pas tout au long de ce voyage. Je te dois tout ce que j'ai accompli.*

*À mon père,*

*Tes conseils avisés, ta sagesse et ta confiance en moi m'ont continuellement inspiré. Tu m'as encouragé à viser toujours plus haut et à croire en mes capacités.*

*À mon frère Samy et ma sœur Khadîdja,*

*Votre soutien indéfectible et vos encouragements constants ont été des piliers essentiels dans toutes mes aventures. Votre présence m'a apporté une motivation inépuisable.*

*À mon neveu Malik,*

*À toi, petit rayon de soleil de notre famille, ta joie de vivre et ton sourire illuminent nos journées. Bien que très jeune, tu apportes une immense joie et un réconfort précieux à tous ceux qui t'entourent.*

*À ma famille élargie,*

*Votre amour inconditionnel et vos prières m'ont accompagné à chaque étape. Votre soutien a été inestimable et je vous en suis profondément reconnaissant.*

*À mes chers amis,*

*Votre amitié sincère, votre soutien moral et les moments de détente partagés ont rendu ce parcours bien plus agréable. Vous avez apporté une touche de légèreté et de joie à cette expérience exigeante.*

*À mon binôme du primaire Billel,*

*Ton partenariat exceptionnel et ta détermination ont été des atouts précieux. Ensemble, nous avons relevé de nombreux défis et accompli de grandes choses.*

***Hamza***

# Remerciements

Nous adressons nos vifs remerciements à notre encadrant de la SGA, M. Youcef HAMANA, superviseur du département Contrôle de Gestion, pour son encadrement rigoureux, ses conseils avisés, son engagement inébranlable et sa disponibilité inestimable tout au long de notre projet. Il a été bien plus qu'un simple encadrant, il a été un mentor pour nous, et nous avons énormément appris de son expertise et de ses connaissances. Malgré son calendrier chargé, il a su trouver du temps pour nous guider et nous soutenir, ce qui a grandement contribué à la réussite de notre projet. Nous lui en sommes profondément reconnaissants.

Nous remercions M. Oussama ARKI, notre promoteur, pour son encadrement et ses conseils.

Nous exprimons notre gratitude à l'ensemble de l'équipe CDG pour leur soutien constant, qui a été une source d'encouragement et de motivation tout au long de ce projet, ainsi qu'à M. Jugurtha HAMMOU, directeur adjoint financier, pour son accueil chaleureux et son intégration bienveillante dès le début de notre parcours.

Nous souhaitons exprimer notre sincère gratitude et reconnaissance envers l'ensemble du personnel de la SGA, et plus particulièrement envers ceux de la direction financière. Ils nous ont permis de travailler dans un environnement agréable. Nous souhaitons également exprimer notre gratitude particulière à M. Mohamed SELMANI, ancien responsable CDG, pour nous avoir permis d'intégrer la SGA. Son accueil et sa confiance ont été déterminants pour nous offrir cette opportunité d'apprentissage et réalisation de notre projet.

Nous remercions par avance les membres du jury, qui nous ont fait l'honneur d'évaluer notre travail.

Nous remercions également l'ensemble des enseignants du département Génie Industriel qui nous ont accompagnés tout au long de notre formation.

# ملخص

قطاع البنوك الجزائري، الذي يشهد تطوراً مستمراً، يتطلب أدوات فعالة لقياس وتحسين أداء وكالاته. تواجه (SGA)، وهي فرع من المجموعة العالمية SG، هذه التحديات للحفاظ على تنافسيتها.

يهدف هذا المشروع الختامي إلى تقديم نموذج لتحليل وتوقع ربحية وكالات SGA، من خلال دمج ذكاء الأعمال والتحليل الإحصائي والنمذجة التنبؤية. الهدف هو إنشاء نظام قرارات استراتيجي يتيح تقييم الأداء الحالي بدقة باستخدام مؤشرات الأداء الرئيسية (KPIs)، وتطوير نماذج تنبؤية قوية لتقدير الربحية المستقبلية للوكالات، مما يعزز تنافسية SGA في سوق يتسم بالرقمنة والمتطلبات المتزايدة.

---

**الكلمات المفتاحية:** الأداء، التنبؤ، الربحية، البنك، مؤشرات الأداء الرئيسية، ذكاء الأعمال، التحليل الإحصائي، نماذج تنبؤية

---

# Abstract

The Algerian banking sector, in constant evolution, requires effective tools to measure and improve the performance of its branches. Société Générale Algérie (SGA), a subsidiary of the global Société Générale group, must address these challenges to remain competitive.

This final year project proposes a model for analyzing and predicting the profitability of SGA's bank branches, integrating Business Intelligence, statistical analysis, and predictive modeling. The objective is to propose a strategic decision-making system that accurately evaluates current performance through key performance indicators (KPIs) and develops robust predictive models to estimate future profitability of the branches, thereby enhancing SGA's competitiveness in an increasingly digital and demanding market.

---

**Mots clés :** Performance, Prediction, Profitability, Bank, KPIs, Business Intelligence, Predictive Modeling.

---

# Résumé

Le secteur bancaire algérien, en constante évolution, nécessite des outils efficaces pour mesurer et améliorer la performance de ses agences. Société Générale Algérie (SGA), filiale du groupe mondial Société Générale, doit relever ces défis pour rester compétitive.

Ce projet de fin d'études propose un modèle d'analyse de la performance et de prédiction de la rentabilité des agences bancaires de SGA, intégrant la Business Intelligence, l'analyse statistique et la modélisation prédictive. L'objectif est de proposer un système d'aide à la décision stratégique, permettant d'évaluer avec précision la performance actuelle à travers des indicateurs clés de performance (KPIs) et de développer des modèles prédictifs robustes pour estimer la rentabilité future des agences, renforçant ainsi la compétitivité de SGA dans un marché de plus en plus digitalisé et exigeant.

---

**Mots clés :** Performance, Prédiction, Rentabilité, Banque, KPIs, Business Intelligence, Modélisation.

---

# Tables des matières

Table des figures

Liste des tableaux

Liste des signes et acronymes

<b>I.</b>	<b>Introduction Générale.....</b>	<b>15</b>
<b>II.</b>	<b>Premier Chapitre : État des lieux .....</b>	<b>16</b>
	Introduction .....	17
	<b>1 Organisme D'accueil :.....</b>	<b>17</b>
	1.1 Présentation du Groupe Société Générale :.....	17
	1.2 Société Générale Algérie :.....	18
	<b>2 Etude de L'Existant.....</b>	<b>27</b>
	2.1 Mesure de Performance et prévisions au sein de la SGA :.....	27
	2.2 Analyse Interne et Externe :.....	28
	<b>3 Énoncé de la Problématique .....</b>	<b>31</b>
	Conclusion.....	31
<b>III.</b>	<b>Deuxième Chapitre : État de l'art.....</b>	<b>32</b>
	Introduction .....	33
	<b>1 Modernisation Bancaire en Algérie : Vers un Futur Numérique Inclusif .....</b>	<b>33</b>
	<b>2 Fondements du Secteur Bancaire.....</b>	<b>34</b>
	2.1 La Finance Bancaire .....	34
	2.2 Les Marchés Bancaires : .....	34
	2.3 Les Risques Bancaires : .....	35
	<b>3 KPIs dans le Secteur Bancaire (Indicateurs de Performance Clés) :.....</b>	<b>36</b>
	3.1 Produit Net Bancaire (PNB) : .....	36
	3.2 Les Frais Généraux (FG) : .....	39
	3.3 Le Résultat d'Exploitation (RE) : .....	41
	3.4 Les Crédits : .....	43
	3.5 Les Dépôts : .....	44
	<b>4 Technologies Financières (FinTech) .....</b>	<b>45</b>
	4.1 La Business Intelligence (BI) : Définition et Impact sur la Performance Bancaire .....	45
	4.2 Justification du Choix de la BI par rapport à d'Autres Outils .....	46
	4.3 Data Science et Intelligence Artificielle (DSIA) : Application à la Prédiction de la Rentabilité Bancaire .....	47
	4.4 Justification du Choix de l'IA par rapport à d'Autres Outils.....	49

4.5	Intersection entre la BI et l'IA :.....	50
4.6	Machine Learning et Séries Temporelles :.....	51
<b>5</b>	<b>Méthodologie de Travail : .....</b>	<b>62</b>
5.1	Processus ETL : Mesure de Performance des agences bancaires.....	62
5.2	Approche CRISP-DM : Prédiction de la Rentabilité des Agences Bancaires .....	63
	<b>Conclusion : .....</b>	<b>64</b>

#### **IV. Troisième Chapitre : Mesure de la Performance des agences bancaires par la Business Intelligence ..... 66**

<b>1</b>	<b>Analyse des Données et Création des Tableaux de Bord :.....</b>	<b>67</b>
1.1	Identification des Indicateurs Clés de Performance (KPI).....	67
1.2	Collecte de Données .....	69
1.3	Prétraitement des Données :.....	70
1.4	Modélisation et Structuration des Données :.....	72
1.5	Chargement de données et Réalisation des Tableaux de Bord :.....	76
<b>2</b>	<b>Étude de Cas : Réalisation du Tableau de Bord du Produit Net Bancaire :.....</b>	<b>77</b>
	<b>Conclusion : .....</b>	<b>84</b>

#### **V. Quatrième Chapitre : Prédiction de la Rentabilité des Agences Bancaires par l'Intelligence Artificielle ..... 85**

<b>1</b>	<b>Analyse, Modélisation et Évaluation des Prévisions Financières Approche Complète .....</b>	<b>86</b>
1.1	Identification des variables à prédire : .....	86
1.2	Importation des données : .....	87
1.3	Retraitement et Préparation de données :.....	87
1.4	Modélisation : .....	87
1.5	Évaluation : .....	88
1.6	Déploiement :.....	88
<b>2</b>	<b>Etude de cas : Prédiction des Frais généraux .....</b>	<b>89</b>
2.1	Importation des données .....	89
2.2	Retraitement et Préparation de la Base de Données :.....	89
2.3	Séparation des données .....	90
2.4	Vérification de la Stationnarité des Séries Temporelles.....	91
2.5	Modélisation : .....	93
2.6	Résultats et Evaluation.....	96
2.7	Interprétation des résultats .....	98
<b>3</b>	<b>Etude Spécifique : Analyse Prédicative des Frais Personnel pour l'Agence 1 .....</b>	<b>99</b>
3.1	Analyse des Données .....	99
3.2	Analyse et évaluation des résultats des modèles pré-entraînés : .....	109

<b>4</b>	<b>Application Utilisateur :.....</b>	<b>111</b>
4.1	Présentation de l'outil : .....	111
4.2	Fonctionnalités .....	112
<b>VI.</b>	<b>Conclusion Générale.....</b>	<b>116</b>
<b>VII.</b>	<b>Bibliographie .....</b>	<b>117</b>
<b>VIII.</b>	<b>Annexe A : Présentation des tableaux de bord.....</b>	<b>118</b>
<b>IX.</b>	<b>Annexe B : Présentation des outils utilisés .....</b>	<b>121</b>

# Table des figures

Figure 1.1 - Répartition des agences de la SGA.....	18
Figure 1.2 - Organigramme de la SGA.....	20
Figure 1.3 - Organigramme de la Direction Financière.....	21
Figure 1.4 - Organigramme du Contrôle de Gestion .....	22
Figure 1.5 - Schéma récapitulatif de la procédure budgétaire .....	25
Figure 2.1 - Les composantes du PNB .....	37
Figure 2.2 - Les composantes de la Marge d'intérêt .....	37
Figure 2.3 - Les composantes des commissions .....	38
Figure 2.4 - Les composantes des Frais généraux .....	39
Figure 2.5 - Les composantes des Frais Personnels.....	40
Figure 2.6 - Les composantes des Autres charges d'exploitation .....	40
Figure 2.7 - Les composantes du résultat d'exploitation .....	42
Figure 2.8 - Schéma récapitulatif des différents KPI de la banque .....	43
Figure 2.9 - Data Science vs Business Intelligence .....	50
Figure 2.10 - Architecture du RNN.....	59
Figure 2.11 - Architecture du LSTM .....	60
Figure 2.12 - Processus ETL.....	63
Figure 2.13 - Cycle de vie de la méthodologie CRISP-DM.....	63
Figure 3.1 - Étapes de la mesure de performances.....	67
Figure 3.2 - Graphe des dépendances Fonctionnelles .....	74
Figure 3.3 - Modèle Conceptuel (Entité-Association) .....	75
Figure 3.4 - Collecte et importation des données.....	77
Figure 3.5 - Split des Chapitres .....	78
Figure 3.6 - Prétraitement des valeurs manquante .....	78
Figure 3.7 - Sélection des colonnes.....	79
Figure 3.8 - Groupement des données et Calcul des Variables Manquantes .....	79
Figure 3.9 - Agrégation des Données en DataFrames Mensuels .....	80
Figure 3.10 - Dashboard du Produit Net Bancaire.....	81
Figure 3.11 - Dashboard des Commissions .....	82
Figure 3.12 - Dashboard de la Marge d'intermédiation .....	83
Figure 3.13 - Dashboard des Autres Produits d'Exploitation.....	83

<b>Figure 4.1 - Les étapes de la prévision de la rentabilité .....</b>	<b>86</b>
<b>Figure 4.2 - Importation des données .....</b>	<b>90</b>
<b>Figure 4.3 - Séparation des données .....</b>	<b>91</b>
<b>Figure 4.4 - Data-Frame des Frais Généraux de l'agence 2 .....</b>	<b>91</b>
<b>Figure 4.5 - Vérification de la Stationnarité .....</b>	<b>92</b>
<b>Figure 4.6 - Transformation des séries temporelles.....</b>	<b>93</b>
<b>Figure 4.7 - Histogramme des Soldes annuels des 'Frais Personnel' de l'agence 1.....</b>	<b>101</b>
<b>Figure 4.8 - Boite à Moustaches des mouvements des FP de l'agence 1 .....</b>	<b>101</b>
<b>Figure 4.9 - Variations des FP de l'agence 1 (Avant différenciation).....</b>	<b>102</b>
<b>Figure 4.10 - Variations des FP de l'agence 1 (Après différenciation).....</b>	<b>102</b>
<b>Figure 4.11 - Décomposition de la série temporelle des FP de l'agence 1.....</b>	<b>104</b>
<b>Figure 4.12 - Test de stationnarité (Avant différenciation).....</b>	<b>105</b>
<b>Figure 4.13 - Test de stationnarité (Après différenciation).....</b>	<b>106</b>
<b>Figure 4.14 - Fonction d'Autocorrélation des mouvements des FP de l'agence 1 .....</b>	<b>107</b>
<b>Figure 4.15 - Fonction de la partielle Autocorrélation (PACF) des mouvements des FP de l'agence 1 .....</b>	<b>108</b>
<b>Figure 4.16 - Prévisions du modèle RNN pour les frais personnels de l'agence 1 .....</b>	<b>109</b>
<b>Figure 4.17 - Prévisions du modèle LSTM pour les frais personnels de l'agence 1 .....</b>	<b>109</b>
<b>Figure 4.18 - Prévisions du modèle Auto-Arima pour les FP de l'agence 1 .....</b>	<b>110</b>
<b>Figure 4.19 - Structure de l'outil de prévision .....</b>	<b>111</b>
<b>Figure 4.20 - Choix de la date de fin de prévision .....</b>	<b>112</b>
<b>Figure 4.21 - Choix du KPI à prédire .....</b>	<b>112</b>
<b>Figure 4.22 - Choix de la sous-famille des KPIs à prédire.....</b>	<b>113</b>
<b>Figure 4.23 - Choix du type d'agence pour la prédiction.....</b>	<b>113</b>
<b>Figure 4.24 - Choix de l'agence pour la prédiction.....</b>	<b>114</b>
<b>Figure 4.25 - Prévisions générés et graphique des prédictions.....</b>	<b>115</b>
<b>Figure A-1 - Dashboard 'Accueil' .....</b>	<b>119</b>
<b>Figure A-2 - Dashboard 'PNB par marché' .....</b>	<b>119</b>
<b>Figure A-3 - Dashboard 'PNB par client corpo' .....</b>	<b>119</b>
<b>Figure A-4 - Dashboard 'PNB par client Retail' .....</b>	<b>119</b>
<b>Figure A-5 - Dashboard 'Frais Généraux' .....</b>	<b>119</b>
<b>Figure A-6 - Dashboard 'Frais Personnel' .....</b>	<b>119</b>
<b>Figure A-7 - Dashboard 'Autres Charges d'Exploitation' .....</b>	<b>120</b>
<b>Figure A-8 - Dashboard 'Résultat d'Exploitation' .....</b>	<b>120</b>
<b>Figure A-9 - Dashboard 'Encours Moyen de Crédits' .....</b>	<b>120</b>
<b>Figure A-10 - Dashboard 'Encours Crédits fin de période'.....</b>	<b>120</b>
<b>Figure A-11 - Dashboard 'Encours Moyen de Dépôts' .....</b>	<b>120</b>
<b>Figure A-12 - Dashboard 'Encours Dépôts fin de période' .....</b>	<b>120</b>
<b>Figure A-13 - Dashboard 'Comparatif Agences' .....</b>	<b>120</b>

# Liste des tableaux

<b>Tableau 1.1 - Modèle type de document budgétaire.....</b>	<b>24</b>
<b>Tableau 3.1 - Les KPI sélectionnés pour la mesure de performances .....</b>	<b>69</b>
<b>Tableau 3.2 - La table ‘Agence’ .....</b>	<b>72</b>
<b>Tableau 3.3 - La table ‘Marché’ .....</b>	<b>72</b>
<b>Tableau 3.4 - La table ‘Client’ .....</b>	<b>73</b>
<b>Tableau 3.5 - La table ‘Date’ .....</b>	<b>73</b>
<b>Tableau 3.6 - La table ‘PNB’ .....</b>	<b>73</b>
<b>Tableau 3.7 - La table ‘Frais Généraux’ .....</b>	<b>73</b>
<b>Tableau 3.8 - La table 'Encours de Crédits' .....</b>	<b>73</b>
<b>Tableau 3.9 - La table 'Encours de Dépôts’ .....</b>	<b>74</b>
<b>Tableau 3.10 - La table ‘Résultat Net Comptable’ .....</b>	<b>74</b>
<b>Tableau 4.1 - Modélisation - Compilation et Adaptation .....</b>	<b>96</b>
<b>Tableau 4.2 - Modélisation - Optimisation.....</b>	<b>96</b>
<b>Tableau 4.3 - Tableau des résultats de l'évaluation des prévisions des FPs .....</b>	<b>98</b>
<b>Tableau 4.4 - Statistiques descriptives pour la série temporelle FP de l'agence 1 .....</b>	<b>100</b>
<b>Tableau 4.5 - Evaluation des modèles appliqués aux Frais personnels de l'agence 1 ....</b>	<b>110</b>

# Liste des signes et acronymes

SG	Société Générale
SGA	Société Générale Algérie
CDG	Contrôle de gestion
KPI	Indicateurs clés de performance
CRM	Compte de résultats mensuels
PIB	Produit intérieur brut
FMI	Fonds monétaire international
PNB	Produit Net Bancaire
ROE	Return on Equity
FG	Frais Généraux
FP	Frais Personnels
PME	Petite et Moyenne Entreprise
GE	Grande Entreprise
BI	Business Intelligence
IA	Intelligence Artificielle
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
RMSE	Root Mean Squared Error
MSE	Mean Squared Error
MAE	Mean Absolute Error

# Introduction Générale

Le secteur bancaire, pilier essentiel de l'économie, joue un rôle primordial par la mobilisation des ressources financières, l'intermédiation entre les acteurs et la stimulation de l'activité économique. Pour rester compétitives, les banques doivent continuellement suivre et améliorer leurs performances en optimisant leurs méthodes de mesure et de prévision. Société Générale Algérie, filiale du groupe mondial Société Générale, se trouve à la croisée des chemins dans ce contexte dynamique et en constante évolution.

Dans un environnement marqué par des fluctuations constantes, la performance et la rentabilité des agences bancaires sont devenues des enjeux majeurs. Actuellement, la gestion des agences de Société Générale Algérie repose sur des processus manuels et traditionnels, limitant leur capacité d'adaptation rapide aux changements du marché et aux besoins des clients. Ce défi soulève une question centrale : *Comment mesurer efficacement la performance actuelle et prédire la rentabilité future des agences dans un tel contexte ?*

Notre projet de fin d'études vise à répondre à cette question en proposant une analyse approfondie des performances des agences de Société Générale Algérie, basée sur des données historiques et actuelles. L'objectif est double : d'une part, évaluer avec précision la situation actuelle à travers des indicateurs clés de performance (KPIs) ; d'autre part, développer des modèles prédictifs robustes pour estimer la rentabilité future des agences à l'échelle nationale. Ce travail fournira ainsi un outil décisionnel stratégique permettant d'améliorer et d'optimiser la gestion du réseau d'agences.

L'intérêt de notre travail réside dans sa capacité à offrir une vision globale et nuancée de la performance et la rentabilité des agences, tout en identifiant les leviers d'action et les perspectives d'amélioration. En se basant sur une méthodologie rigoureuse qui combine la Business Intelligence, l'analyse statistique et la modélisation prédictive, ce projet détermine les bases nécessaires pour une prise de décision éclairée, contribuant à l'efficacité opérationnelle et à la compétitivité à long terme de Société Générale Algérie.

Ce présent document s'articule autour de quatre chapitres principaux. Le premier chapitre, intitulée "État des lieux", comprend une présentation détaillée de Société Générale Algérie, y compris sa répartition, ses produits et services, ses performances clés, et son département de contrôle de gestion, ainsi que notre diagnostic et analyse interne et externe identifiant les facteurs influents et les limitations de l'approche actuelle. Le deuxième chapitre, "État de l'art", offre une vue d'ensemble du secteur bancaire algérien, détaille les réformes et modernisations, et examine l'impact de l'analyse financière et de l'étude de la rentabilité en contexte bancaire, il explore les risques associés, et les principaux KPIs. Il aborde également les avancées en technologie financière, telles que la Business Intelligence et l'application de la data science et de l'intelligence artificielle en finance. Le troisième chapitre, "Mesure de la Performance des agences bancaires par la Business Intelligence", décrit notre solution consacrée à l'identification des KPIs, la collecte et le prétraitement des données, la conception des tableaux de bord pour l'analyse. Enfin le quatrième chapitre "Prédiction de la Rentabilité des Agences Bancaires par l'Intelligence Artificielle", comprend la modélisation prédictive, l'évaluation des résultats et le déploiement de notre solution proposée, pour simplifier l'accès aux informations clés et favoriser une prise de décision éclairée et rapide.

---

**Premier Chapitre**  
**État des lieux**

---

## Introduction

Ce chapitre présente un état des lieux détaillé de la Société Générale Algérie (SGA), filiale du groupe Société Générale. En mettant en lumière l'historique, la structure organisationnelle, les produits et services offerts, ainsi que la répartition géographique de ses agences, nous visons à fournir une compréhension complète du contexte opérationnel de la SGA. Nous analysons également les missions clés de la direction financière et du département de contrôle de gestion, tout en décrivant les procédures budgétaires et les documents de suivi utilisés. Cette vue d'ensemble est essentielle pour identifier les forces et les faiblesses des méthodes actuelles, et poser les bases d'améliorations pour mesurer en mieux la performance et prédire la rentabilité des agences à l'échelle national.

## 1 Organisme D'accueil :

### 1.1 Présentation du Groupe Société Générale :

Fondée en 1864, la Société Générale est une banque française de renommée internationale, dont le siège est établi à Paris. Reconnue pour sa longue histoire et son envergure mondiale, elle se positionne comme une des banques les plus importantes de France et du monde, avec une présence significative dans de nombreux pays.

Le groupe opère à travers trois segments d'activité principaux : la banque de détail et les services financiers spécialisés, la banque d'investissement et de financement, ainsi que la gestion d'actifs et les marchés mondiaux. Pour les particuliers et les PME, elle propose une gamme étendue de services tels que les comptes d'épargne, les prêts, les cartes de crédit, et les solutions d'assurance, tandis que sa branche de services financiers spécialisés inclut la gestion de patrimoine et la banque privée. En 2021, la Société Générale servait environ 31 millions de clients à travers le monde, employant 133 000 personnes dans divers domaines allant de la banque de détail à la technologie de l'information.

Société Générale offre une large palette de conseils, solutions financières sur mesure aux particuliers, entreprises et investisseurs institutionnels, autour de 3 pôles métiers complémentaires :

- **La banque de détail** en France repose sur la complémentarité de trois enseignes : Société Générale, Crédit du Nord et Boursorama
- **La banque de détail à l'international services financiers et assurances** est présente dans 36 pays et s'appuie sur la mise en œuvre d'un modèle de banque universelle adapté aux spécificités locales.
- **La banque de financement et d'investissement** (Société Générale Corporate & Investment Banking) est présente sur les principales places financières et représente 28% du chiffre d'affaires

Les activités de gestion d'actif et services aux investisseurs sont composées de la Banque Privée, de Gestion d'Actifs, de Gestion de Titres et de courtage sur produits dérivés et représente 11% du chiffre d'affaires

## 1.2 Société Générale Algérie :

La Société Générale Algérie, filiale du groupe Société Générale, marque sa présence en Algérie depuis 24 ans, se distinguant comme le leader bancaire à capitaux privés du pays, tant en termes de réseau d'agences que de clientèle. Elle offre un éventail complet de produits et services financiers destinés aux particuliers, ainsi qu'aux petites, moyennes et grandes entreprises. Son portefeuille de services en banque de détail comprend, entre autres, les comptes courants, les offres de crédit et les solutions bancaires digitales. Par ailleurs, elle fournit des services de banque d'investissement de haut niveau, incluant le conseil en fusion-acquisition, le financement de projets et la gestion d'actifs. Société Générale Algérie facilite l'accès à ses services grâce à son site web et à son application mobile, permettant à ses 450 000 clients de gérer aisément leurs opérations bancaires. La filiale emploie environ 1 500 collaborateurs, renforçant ainsi sa présence locale et son engagement envers le marché algérien.

### 1.2.1 Répartition et Spécialisation des Agences de la Société Générale Algérie :

La Société Générale Algérie (SGA) a établi une présence robuste à travers le pays avec des agences bancaires stratégiquement réparties sur 31 wilayas, couvrant des régions diverses et dynamiques, allant des grands centres urbains aux localités plus rurales dont 13 Centres d'Affaires dédiés à l'activité de la clientèle des Entreprises.

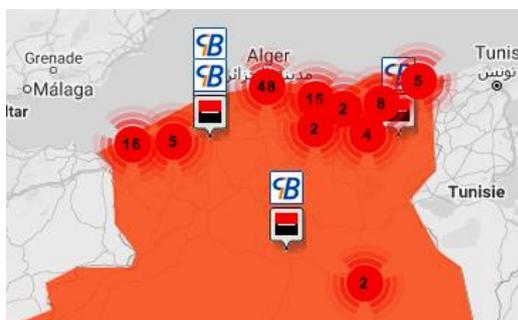


Figure 1.1 - Répartition des agences de la SGA

Ain Temouchent ; Alger ; Annaba ; Batna ;  
Bejaia ; Biskra ; Blida ; Bordj Bou Arreridj ;  
Bouira ; Boumerdes ; Chlef ; Constantine ;  
Ghardaia ; Jijel ; Khenchla ; M'sila ; Mascara ;  
Médéa ; Mostaganem ; Oran ; Ouargla ;  
Relizane ; Setif ; Sidi bel Abbes ; Skikda ; Souk  
Ahras ; Tiaret ; Tipaza (Koléa) ; Tizi-ouzou ;  
Tlemcen

Cette vaste répartition géographique permet à la SGA de toucher une clientèle variée et de répondre efficacement aux besoins spécifiques de chaque région. Les agences de la SGA se classent en trois catégories principales, chacune ciblant des segments de marché distincts :

**Agences Retail :** Ces agences sont principalement orientées vers les particuliers et offrent une gamme complète de services bancaires, incluant les comptes de dépôt, les prêts personnels, les cartes de crédit, et les solutions d'épargne. Elles sont conçues pour servir la grande majorité des clients individuels, proposant des services accessibles et adaptés aux besoins des populations locales.

**Agences Corporate :** Spécifiquement dédiées aux grandes entreprises, ces agences offrent des services bancaires spécialisés tels que le financement des entreprises, la gestion des flux de trésorerie, le commerce international, et les solutions de financement structuré. Situées

principalement dans les zones économiques clés telles qu'Alger, Oran, et Annaba, elles sont stratégiquement placées pour servir les grandes entreprises et les acteurs économiques majeurs.

**Agences Mutualisées** : Opérant comme des centres de services partagés, ces agences fournissent des services à la fois aux clients particuliers et aux entreprises. Elles jouent un rôle crucial dans la facilitation des interactions entre différents types de clients et permettent une synergie entre les services offerts aux particuliers et aux petites entreprises, optimisant ainsi l'efficacité opérationnelle et la satisfaction client.

La répartition des agences de la SGA à travers différentes wilayas, comme Alger, Oran, Constantine, et Tizi-Ouzou, assure une couverture étendue et une proximité avec les clients, ce qui est essentiel pour maintenir la confiance et favoriser une relation durable. Cette stratégie géographique combinée à la diversification des types d'agences permet à la SGA de maximiser son impact sur le marché algérien, en adaptant ses services aux caractéristiques économiques et démographiques de chaque région. En somme, la SGA, par cette structuration en réseau, renforce sa capacité à répondre de manière flexible et ciblée aux besoins diversifiés de sa clientèle, contribuant ainsi à sa croissance et à sa rentabilité dans le secteur bancaire concurrentiel algérien.

### 1.2.2 Les produits, services et marchés de la SGA

La Société Générale Algérie offre une gamme diversifiée et innovante de services bancaires à plus de 450 000 clients Particuliers, Professionnels et Entreprises :

- Comptes bancaires.
- Prêt de financement de l'habitat (Achats et constructions à moyen terme de 5 à 7 ans ou à long terme plus de 7 ans et des crédits d'amélioration de l'habitat).
- Crédits à la consommation (crédits classiques, motos, autos).
- Financement de créances aux professionnels (crédits spot, escompte...).
- Placements, épargnes, cartes bancaires, ainsi que des services de leasing.

Le marché bancaire est le lieu où se rencontrent des particuliers et des entreprises en quête de financement, le secteur d'intervention de la SGA est composé de :

- 44 % de PME
- 39 % de grandes entreprises
- 17 % de clientèle particuliers et professionnels

La SGA dispose d'un portefeuille clients comprenant **5 232 Entreprises** réparties comme suit : 88% PME, 12% Grandes Entreprises.

Le portefeuille clients entreprises représente 83% du total des engagements de la banque, 17% des engagements étant des clients Particuliers et Professionnels.

Voici l'organigramme de la SGA

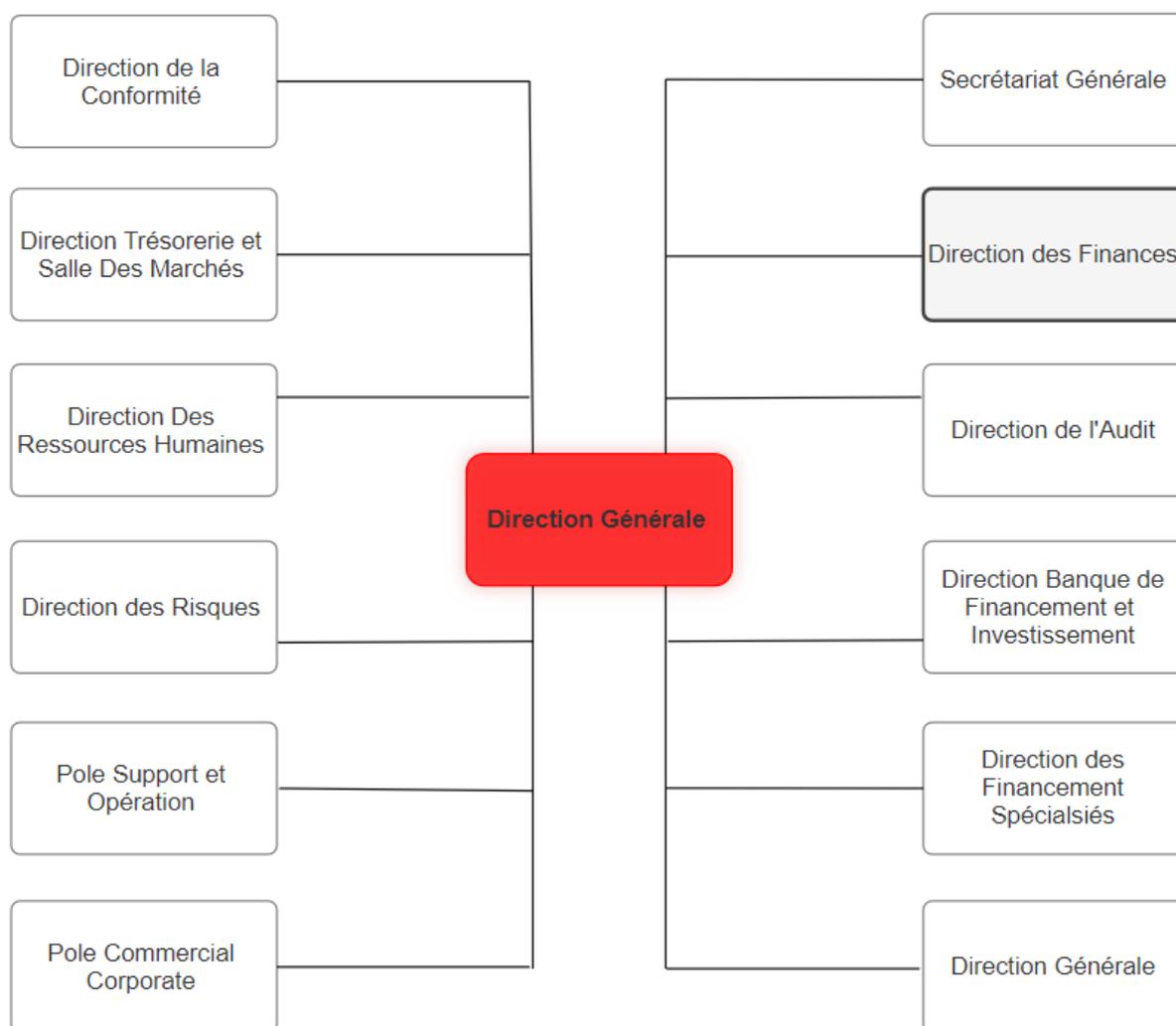


Figure 1.2 - Organigramme de la SGA

### 1.2.3 Direction des Finances :

La Direction Financière de Société Générale Algérie joue un rôle pivot dans la stratégie de gestion et de développement de la banque, incarnant le nerf central de ses activités économiques et financières.

La Direction Financière de Société Générale Algérie est chargée de plusieurs missions cruciales qui sous-tendent la performance et la stabilité de l'institution. Parmi ces missions, on compte la planification financière, qui implique l'élaboration de prévisions budgétaires et la mise en place d'objectifs financiers à court, moyen, et long terme pour la banque ainsi que ses agences. Cette direction assure également la gestion des risques financiers, en identifiant, analysant, et mettant en œuvre des stratégies pour mitiger les risques de marché, de crédit, et opérationnels. L'optimisation de la structure du capital, la gestion des performances des agences bancaires, de liquidité et des flux de trésorerie, ainsi que le contrôle et le reporting financier qui figurent également parmi ses responsabilités. En outre, la Direction Financière est impliquée dans les opérations de financement et d'investissement, veillant à ce que la banque dispose des ressources nécessaires pour ses activités et à l'identification d'opportunités d'investissement rentables.

Voici l'organigramme de la direction des Finances :

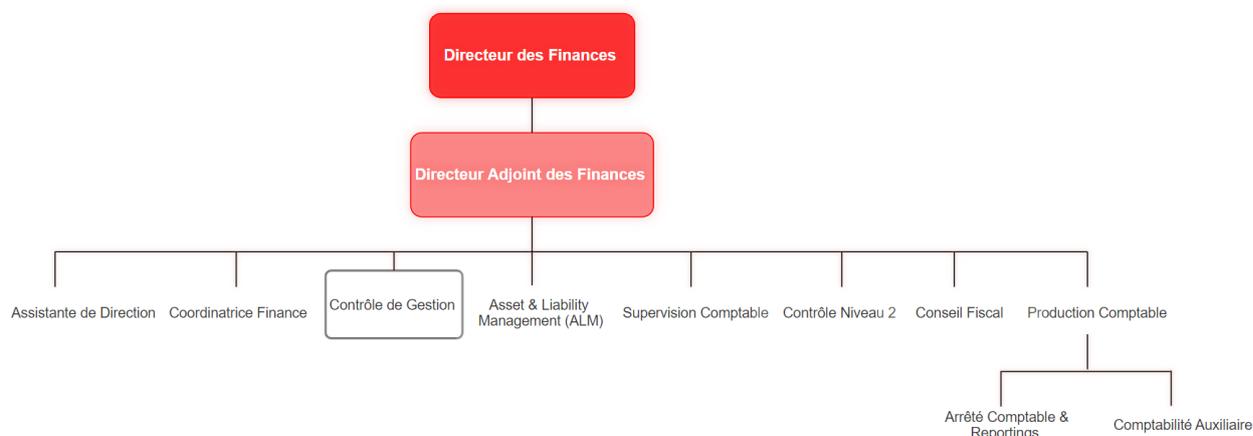


Figure 1.3 - Organigramme de la Direction Financière

#### 1.2.4 Département Contrôle de Gestion :

Le Département de Contrôle de Gestion, essentiel au sein de la Direction Financière de Société Générale Algérie (SGA), assume un rôle primordial dans le pilotage économique de la banque. Sa mission centrale est d'assurer le suivi de la performance financière et opérationnelle de l'ensemble des activités de l'institution. En véritable partenaire stratégique, ce département établit les budgets, réalise des analyses financières approfondies et développe des prévisions économiques, permettant ainsi une gestion optimisée des ressources. Il joue également un rôle crucial dans l'évaluation de la rentabilité des divers segments d'activité et projets de la banque, fournissant une base solide pour la prise de décisions stratégiques éclairées. Le Contrôle de Gestion est responsable de l'élaboration des tableaux de bord et des reportings réguliers, offrant ainsi une visibilité claire sur les performances réelles par rapport aux objectifs fixés. Par l'analyse des écarts et la proposition de mesures correctives, ce département assure une gestion proactive des coûts et des revenus, contribuant de manière significative à l'amélioration continue de l'efficacité et de l'efficacité opérationnelle de la SGA.

Voici l'organigramme du département Contrôle de Gestion :

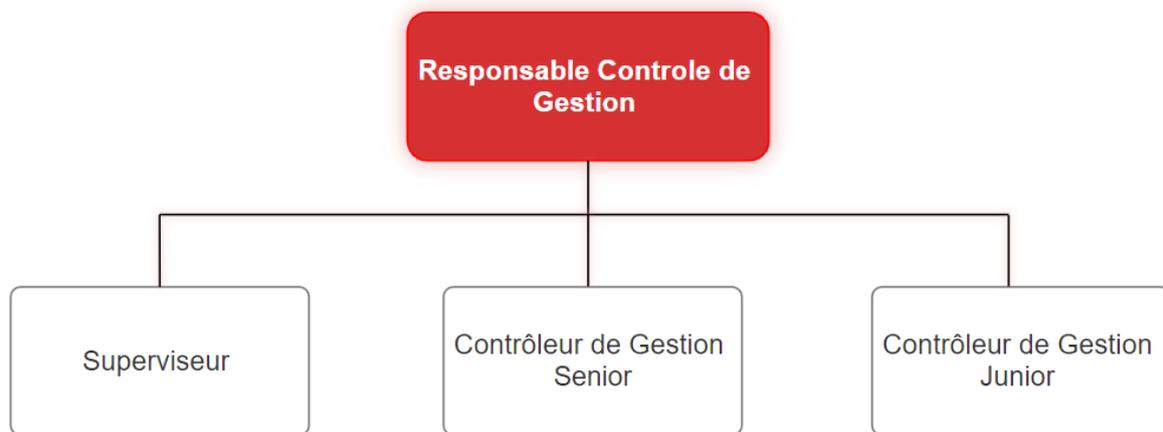


Figure 1.4 - Organigramme du Contrôle de Gestion

### Domaine et Activités du Contrôle de Gestion :

Nous allons maintenant parler de nombreuses activités de CDG, Au cœur du domaine d'activité du Contrôle de Gestion chez Société Générale Algérie, les aspects clés comprennent :

- **Surveillance et Analyse des Performances** : Examen continu des indicateurs clés de performance, analyse des coûts, des revenus, et des marges par branche et unité opérationnelle.
- **Détection des écarts** : Identification proactive des divergences entre les résultats réels et les prévisions, permettant de repérer les opportunités d'amélioration et les zones à risque.
- **Modélisation Financière et Analyse de Scénarios** : Élaboration de modèles prédictifs et analyses de différents scénarios pour évaluer l'impact des décisions stratégiques et des fluctuations du marché sur la santé financière de l'entreprise.
- **Évaluation des Investissements et Analyse de Rentabilité** : Analyse détaillée des nouveaux produits et services, ainsi que des projets d'investissement pour évaluer leur potentiel de rentabilité.
- **Suivi des Coûts Opérationnels** : Contrôle rigoureux des dépenses opérationnelles pour identifier les possibilités de réduction des coûts et d'optimisation des ressources.
- **Support à la Prise de Décision** : Fourniture d'insights financiers et d'analyses approfondies pour guider la prise de décisions stratégiques et opérationnelles.
- **Optimisation de la Performance Financière** : Contribution active à l'amélioration de l'efficacité opérationnelle et à l'augmentation de la performance financière globale de la banque.

Ces points soulignent l'importance du Contrôle de Gestion dans la stratégie globale et les opérations quotidiennes de Société Générale Algérie, en assurant une gestion financière efficace et en soutenant la prise de décisions éclairées.

### 1.2.5 Les procédures et reportings liés à la performance :

Dans cette section, nous décrivons les différents outils et documents utilisés pour la gestion et le suivi des activités au sein de la SGA. Nous mettrons en lumière deux éléments clés : les tableaux de bord et les reportings, ainsi que la procédure budgétaire.

#### Tableaux de bord et reportings :

Un tableau de bord de gestion (Dashboard) est un document clair et synthétique qui regroupe un ensemble d'indicateurs dont le responsable d'une entité a besoin au fur et à mesure pour la piloter, il permet d'analyser rapidement l'évolution des indicateurs et de déterminer rapidement des actions correctives. Il s'agit, par conséquent, d'un outil de prise de décision à court terme permettant à la fois d'orienter et de piloter.

Le reporting est un dispositif qui permet de remonter l'information selon une périodicité préalablement définie (souvent mensuelle) à l'usage de la ligne hiérarchique. Il permet à celle-ci d'exercer un contrôle a posteriori des responsabilités déléguées.

Contrairement au tableau de bord opérationnel qui est tourné vers l'action, le reporting est plutôt un outil de contrôle a posteriori des responsabilités déléguées aux managers, dont la finalité est de vérifier que les objectifs délégués ont bien été atteints, il s'agit donc plus d'évaluer la performance. Pour autant, en dehors du contrôle des subordonnés, le reporting offre, au travers d'une information agrégée (principaux objectifs, faits marquants de la période, compte d'exploitation et bilan, avancement des plans d'action, etc.), une vision globale des réalisations,

#### La procédure budgétaire :

Le budget permet la projection des besoins en ressources financières nécessaires à la réalisation des objectifs stratégiques et plans d'action de chaque entité, il doit couvrir la totalité des activités de l'entreprise, doit être exhaustif, détaillé et valorisé.

Le contrôleur de gestion joue un rôle central dans ce processus de construction budgétaire, il est le garant de la cohérence globale de celle-ci, sa position lui permet d'accompagner l'ensemble des décideurs opérationnels dans la définition de leurs objectifs et le choix de leurs ressources, pour que le processus budgétaire soit efficace, il est donc nécessaire que le contrôle de gestion ait un rôle de médiateur.

A la fin du premier semestre le service contrôle de gestion entame les étapes constituant la procédure budgétaire de l'année à suivre (N+1)

Cinq phases clés peuvent être mises en évidence dans son organisation :

- **La première phase** : consiste en l'élaboration du planning par le groupe retraçant toutes les dates limites ainsi que les différentes missions nécessaires au bon déroulement de la procédure budgétaire.
- **La seconde phase** : concerne la création et la mise en place des Templates. Ces derniers regroupent l'évolution du budget, le réalisé mensuel ainsi que des champs à remplir par les différents métiers de la banque sur leurs objectifs et dépenses.

Toutes les structures de la banque sont concernées par la procédure budgétaire de la manière suivante :

- ✓ Les directions corporate, Retail, Trésorerie et la salle des marchés sont interrogées sur les prévisions liées aux indicateurs d'activité commerciale.
- ✓ Les structures restantes qui sont des structures gestionnaires sont sollicitées pour la collecte des prévisions liées aux charges d'exploitation et d'investissement.

## Chapitre I : État des lieux

- ✓ Le Secrétariat Général
- ✓ La Direction financière
- ✓ La Direction des ressources humaines
- ✓ La Direction des systèmes d'information
- ✓ La Direction de l'immobilier
- ✓ La Direction des moyens de paiements
- ✓ La Direction des Risques
- ✓ La Directions Marketing

Une fois la procédure budgétaire lancée, les Templates sont diffusé par la direction financière aux métiers avec un délai qui doit être respecté, suite à ça plusieurs entretiens entre CDG et métiers sont effectués pour revoir et éclaircir les chiffres avancés et assurer la cohérence globale de la trame budgétaire

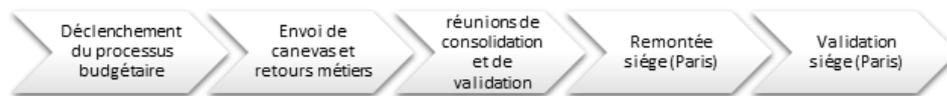
Il est important que les documents, mais aussi les méthodes de chiffrage et de calcul, soient standardisés d'une direction à l'autre afin de permettre les validations et les arbitrages par la direction générale ainsi que la consolidation par le contrôle de gestion

Voici un modèle type de document budgétaire (tableau de bord) :

Revenus ou indicateurs d'activité	Rappel réalisé N-1	Objectif N	Estimé N	Proposition de budget N+1	Variation N+1/estimé N
Commentaires					

*Tableau 1.1 - Modèle type de document budgétaire*

- **La troisième phase** : une fois que le travail avec les métiers est finalisé, que les chiffres sont justifiés et que les canevas sont établis. Le contrôle de gestion met en place une présentation du budget pour la direction générale, qui peut demander certaines modifications, suite à ça une deuxième réunion est organisée pour la validation final des chiffres
- **La quatrième phase** : est caractérisée par la remontée des chiffres via Finlab (logiciel intra groupe) et l'organisation d'une réunion entre la direction générale SGA et le siège afin de justifier les hypothèses de construction du budget et les différentes variations par rapport au réalisé N et N+1, des modifications peuvent être exigées par le siège avant leurs approbations finales
- **La cinquième phase** : la dernière étape est la diffusion des chiffres aux services concernés, des notifications d'objectifs à atteindre sont communiquées aux métiers concernant les encours, les indicateurs d'activité etc.



*Figure 1.5 - Schéma récapitulatif de la procédure budgétaire*

Une fois le budget élaboré et validé et selon la périodicité souhaitée par la banque, le contrôle de gestion doit analyser l'exécution du budget et déterminer les écarts par rapport aux prévisions et effectuer des reportings périodiques dédiés au suivi de l'activité de la banque et indiquant les principales constatations et recommandations

Les reportings relatifs au suivi budgétaire sont de deux catégories :

- ✓ Reportings destinés au Groupe
- ✓ Reportings locaux

Ci-dessous nous allons présenter certains de ces reportings :

### **Reportings destinés au groupe :**

- Liasse magnitude : reporting mensuel qui reprend le suivi des indicateurs d'activité en termes d'encours et de production, de compte de résultat ainsi qu'un suivi détaillé des frais généraux
- Présentation des résultats : reprend une analyse approfondie des indicateurs financiers tels que le compte de résultat, les encours et la production afin de détecter l'ensemble des variations par rapport au budget, à l'estimé et au réalisé
- Reporting coût IT : reporting trimestriel, il met en évidence les dépenses informatiques
- Reporting Trade : présentation annuelle des données du commerce extérieur : crédits documentaires, remises documentaires et garanties internationales

### **Reportings locaux :**

Ici nous pouvons identifier les reportings d'exploitation des reportings du commerce extérieur

- ✓ Reportings d'exploitation globale :

Compte de Résultat Mensuel (CRM) : rapport mensuel qui reprend la composition du résultat de la banque ainsi que la justification des variations entre réalisation, budget et estimé à savoir :

- Marge d'intermédiation (= intérêts reçus – intérêts payés)
- Commissions (dont gains de change)
- Autres produits et charges d'exploitation
- Produit net bancaire (= produits d'exploitation bancaire – charges d'exploitation bancaire)
- Frais de personnel

## Chapitre I : État des lieux

- Amortissement
- Autres charges d'exploitation
- Total des frais généraux
- Résultat brut d'exploitation ( $R.B.E = P.N.B - \text{charges de fonctionnement} - \text{dotations aux amortissements}$ )
- Coût net du risque
- Résultat net comptable

CRM par agence : reporting mensuel qui reprend le compte de résultat en décomposant le PNB et les FG par agences de la manière suivante :

- Marge d'intérêt
- Intérêts de refinancement
- Commissions nettes
- Frais personnel
- Amortissement
- Pourcentage du PNB par effectif
- Coefficient d'exploitation par agence ( $= \text{charges d'exploitation par agence} / \text{P.N.B par agence}$ )
- Pourcentage des FG par agence
- Pourcentage du RBE par agence

✓ Tableau de bord banque :

- Reporting mensuel qui regroupe les indicateurs suivants :
  - Activité et résultats : reprend la composition détaillée du résultat net comptable en mettant en exergue le pourcentage de réalisation par rapport au budget et aux réalisations du mois de l'exercice précédent
  - Statistiques commerciales : ça concerne les ouvertures et fermetures de comptes ainsi que les souscriptions aux différents produits
- Dépôts : rend compte des encours des différents dépôts par marché sur le mois encours ainsi que la variation et l'écart par rapport au mois précédent
- Effectifs : reprend le nombre d'employés de la banque
- Nombre de clients : reprend le total de la clientèle du mois en cours
- Tableau de bord retail : tableau mensuel qui met en évidence le nombre de clients, de comptes, de dépôts, d'emplois et d'engagement par agence
- Tableau de bord corporate : il reprend les mêmes indicateurs que pour le retail, mais pour les business center et la direction des grandes entreprises
- PNB par client : reporting trimestriel, il reprend les informations concernant la répartition du PNB par client
- PNB par marché : ce fait trimestriellement, il reprend l'ensemble du PNB par marché (CLIPRO, CLIPRI, PME, GE)
- FG par structure : document mensuel met en évidence l'ensemble des dépenses liées à une structure, il est envoyé à chaque structure concernée

## Reporting du commerce international

- Reporting Trade : document trimestriel qui regroupe les indicateurs relatifs aux crédits documentaires, remises documentaires, et garanties internationales
- Synthèse production comex : reporting qui reprend l'évolution de la production de crédoc et remdoc en nombre et en montant
- PNB Trade : élaborer mensuellement, il reprend l'activité du commerce extérieur de la banque
- Production Crédoc : reporting mensuel qui regroupe l'ensemble des opérations crédoc par agence (nombre, volume, détail clients ...)
- Production Remdoc : regroupe l'ensemble des opérations remdoc par agence (nombre, montant et client)

## 2 Etude de L'Existant

Dans cette partie, nous analysons les méthodes actuelles de la Société Générale Algérie pour mesurer la performance et prédire la rentabilité de ses agences bancaires, en identifiant les facteurs influents et les limitations de l'approche actuelle. Nous formulons la problématique de notre projet et présentons une méthodologie rigoureuse, pour améliorer la gestion de la performance et la prévision de la rentabilité, renforçant ainsi la compétitivité de la banque.

### 2.1 Mesure de Performance et prévisions au sein de la SGA :

La Société Générale Algérie emploie plusieurs indicateurs clés de performance (KPIs) pour évaluer l'efficacité de ses agences, incluant le Produit Net Bancaire (PNB), les Frais Généraux (FG), les Encours de Crédits, les Encours de Dépôts, le Résultat Net d'Exploitation, le LTD (Loan To Deposit), le ratio Cost-Income, qu'on développera plus tard dans le chapitre suivant. Ces indicateurs sont présentés dans une section de tableaux et des fichiers Excel comportant les soldes ainsi que le budget.

Toutes ces variables financières sont obtenues à partir des extractions et calculées en global banque et aussi par agence pour la majorité d'entre eux.

L'objectif principal de la première partie qui est la mesure de la performance est d'aider à la prise de décision, avoir l'impact de chaque agence sur les revenus bancaires et savoir quelles sont les moins performantes afin que chaque responsable et gestionnaire d'agence puisse atteindre les objectifs prédéfinis à partir de la meilleure agence.

Le second objectif celui de la deuxième partie est de faire une prévision et prédire le comportement financier futur de ces agences dont nous avons mesurés leurs performances.

**Diagnostic 1 :** *L'approche actuelle est entravée par le manque de données fréquentes et la sous-prise en compte des variables pertinentes, de plus, les analyses et les prévisions effectuées sont principalement faits sur Excel, ce qui réduit la capacité à capturer les tendances en temps réel et à anticiper les impacts potentiels de la non linéarité sur la performance des agences.*

**Diagnostic 2 :** *Les prévisions actuelles reposent sur une approche globale plutôt que granulaire, ce qui peut conduire à une perte de précision. Le split se fait par des clés de répartition déterminées en fonction de poids, ce qui ne tient pas suffisamment compte des*

*spécificités de chaque agence. Par conséquent, les décisions stratégiques sont prises sur la base de données potentiellement biaisées, ce qui pourrait entraîner des conséquences néfastes pour la rentabilité globale de la banque.*

Pour remédier à ces lacunes et gérer en mieux le réseau d'agences, nous proposons une solution intégrée reposant sur l'utilisation des techniques de data science et d'intelligence artificielle qui aidera la SGA, non seulement à améliorer la mesure de la performance et prévoir la rentabilité, mais aussi mieux s'adapter aux défis et aux opportunités du marché, renforçant ainsi sa position concurrentielle et sa viabilité à long terme.

## **2.2 Analyse Interne et Externe :**

Dans cette section, nous réalisons une analyse approfondie de l'environnement interne et externe de la Société Générale Algérie (SGA). Cette analyse est essentielle pour identifier les facteurs influençant la performance et la rentabilité des agences bancaires de la SGA. Elle comprend deux volets principaux : l'analyse de l'environnement macroéconomique à travers le modèle PESTEL et l'évaluation concurrentielle et de marché via le modèle des Cinq Forces de Porter. De plus, une analyse SWOT permettra d'identifier les forces, faiblesses, opportunités et menaces auxquelles la SGA est confrontée.

### **2.2.1 Analyse de l'Environnement Macroéconomique : PESTEL**

À travers une analyse PESTEL, nous évaluons l'influence des facteurs politiques, économiques, technologiques, écologiques et légaux sur la SGA. Cette perspective macroéconomique révèle les dynamiques du marché et les défis spécifiques au secteur bancaire algérien.

#### **Politique**

Le cadre réglementaire algérien, en constante évolution, influence directement la manière dont les agences bancaires opèrent et se conforment aux normes.

La réglementation peut affecter les pratiques en matière de reporting financier et de transparence, ce qui est essentiel pour une mesure précise des performances. De plus, les politiques de privatisation et les initiatives gouvernementales en faveur de l'inclusion financière peuvent ouvrir de nouvelles opportunités de marché et nécessiter des ajustements dans la stratégie des agences.

#### **Économique**

Les conditions économiques, telles que la croissance économique, les taux d'intérêt, et l'inflation, impactent directement la rentabilité des agences.

Les périodes de récession ou de faible croissance économique peuvent réduire la demande de crédit et augmenter les taux de défaut, affectant ainsi les prévisions de rentabilité.

La diversification de l'économie algérienne influence également le développement de nouveaux produits et services bancaires adaptés à divers secteurs économiques.

#### **Technologique**

La digitalisation et l'innovation technologique transforment les opérations bancaires. Pour les agences de la Société Générale Algérie, l'intégration de technologies avancées telles que l'analyse de données et l'intelligence artificielle peut améliorer la précision des prévisions de

rentabilité. Les agences doivent être équipées pour exploiter ces technologies afin d'optimiser la gestion des risques et d'améliorer l'efficacité opérationnelle.

### **Écologique**

La prise de conscience croissante des enjeux environnementaux pousse les banques à investir dans des projets de finance durable et à proposer des produits éco-responsables. Les agences qui intègrent des critères environnementaux dans leurs opérations et leurs évaluations de crédit peuvent non seulement attirer des clients soucieux de l'environnement, mais aussi minimiser les risques associés aux changements climatiques et réglementaires.

### **Légal**

Les normes légales, en particulier celles liées à la protection des données et à la cybersécurité, sont cruciales pour les opérations numériques des agences. Le respect strict de ces lois est nécessaire pour préserver la confiance des clients et éviter les sanctions réglementaires. Les implications légales doivent donc être intégrées dans la mesure des performances et les prévisions de rentabilité pour garantir la conformité et la sécurité des opérations.

Cette analyse PESTEL fournit un cadre pour comprendre les facteurs macroéconomiques qui influencent la mesure des performances et la prédiction de la rentabilité des agences bancaires. En intégrant ces facteurs dans l'analyse, la banque peut élaborer des stratégies plus robustes et résilientes face à un environnement en mutation.

## **2.2.2 Évaluation Concurrentielle et de Marché : Modèle des Cinq Forces de Porter**

L'application du modèle des cinq forces de Porter offre une compréhension de l'intensité concurrentielle au sein du secteur. Cette analyse met en évidence les principaux leviers compétitifs et les stratégies potentielles pour la SGA.

**Rivalité entre Concurrents Existantes :** Intensité élevée due à la présence de nombreuses banques locales et internationales ; nécessité pour la SGA de se distinguer par la qualité de service et l'innovation.

**Menace des Nouveaux Entrants :** Modérée, limitée par des barrières réglementaires importantes ; risque accru avec l'émergence des néo-banques et des fintechs.

**Pouvoir de Négociation des Clients :** Élevé, les clients sont mieux informés et demandent des services plus compétitifs et de meilleure qualité, poussant la SGA à innover et personnaliser ses offres.

**Menace des Produits de Substitution :** Croissante avec l'émergence des fintechs et des solutions de paiement mobile, offrant des alternatives pratiques et économiques aux services traditionnels.

**Pouvoir de Négociation des Fournisseurs :** Modéré, principalement avec des partenaires technologiques et financiers, la SGA a la flexibilité de choisir parmi plusieurs options du marché.

## **2.2.3 Analyse SWOT de la Société Générale Algérie**

L'analyse SWOT permet d'évaluer les forces, les faiblesses, les opportunités et les menaces d'une organisation. Pour la Société Générale Algérie (SGA), cette analyse aide à

développer des stratégies adaptées pour renforcer ses avantages compétitifs et se préparer aux défis futurs.

### **Forces**

**Réseau d'Agences Étendu :** La SGA bénéficie d'un vaste réseau d'agences bien établi qui lui permet de toucher une large clientèle et de capter significativement le marché algérien.

**Expertise Internationale :** Faisant partie du groupe Société Générale, la SGA tire parti de l'expertise internationale et des meilleures pratiques globales, ce qui renforce sa capacité à innover et à maintenir des standards élevés.

**Adoption Précoce de Solutions Numériques :** La SGA a été parmi les premières banques en Algérie à adopter des solutions numériques, ce qui lui a permis de se positionner favorablement face à la digitalisation croissante des services financiers.

### **Faiblesses**

**Dépendance sur le Marché Local Volatile :** La forte dépendance de la SGA sur le marché algérien, caractérisé par sa volatilité économique et politique, rend ses performances susceptibles aux fluctuations locales.

**Perception de l'Expérience Client à Améliorer :** Malgré des avancées numériques, la perception de l'expérience client nécessite des améliorations pour répondre aux attentes modernes des consommateurs, en particulier dans un contexte où la concurrence s'intensifie.

### **Opportunités**

**Croissance du Secteur des Fintechs :** La montée des fintechs offre à la SGA l'opportunité de former des partenariats stratégiques ou d'innover en interne pour développer de nouveaux produits et services financiers.

**Potentiel d'Expansion Régionale :** Avec sa base solide en Algérie, la SGA a l'opportunité d'élargir sa présence dans d'autres marchés régionaux, exploitant ainsi de nouveaux segments de clientèle.

**Demande pour des Services Bancaires Innovants :** L'augmentation de la demande pour des services bancaires personnalisés et technologiquement avancés représente une opportunité significative pour la SGA de se différencier et d'attirer une clientèle plus large et diversifiée.

### **Menaces**

**Concurrence Accrue :** L'intensification de la concurrence, tant de la part des banques traditionnelles que des nouveaux acteurs, met la SGA face à la nécessité d'innover constamment. Comment la SGA peut-elle maintenir son avantage concurrentiel et innover suffisamment pour répondre à cette pression croissante ?

**Réglementations Changeantes :** Les réglementations financières en constante évolution peuvent limiter certaines opérations ou exiger des ajustements coûteux. Comment la SGA peut-elle s'adapter rapidement et efficacement aux changements réglementaires tout en minimisant les coûts et les disruptions opérationnelles ?

**Risques Liés à la Cybersécurité :** Dans un environnement de plus en plus numérisé, les risques de cybersécurité représentent une menace croissante pour la sécurité des données clients et la stabilité opérationnelle de la banque. Quelles mesures la SGA doit-elle mettre en place pour renforcer sa cybersécurité et assurer la protection des données sensibles de ses clients face à des menaces évolutives ?

### 3 Énoncé de la Problématique

Après avoir réalisé notre diagnostic, nous avons clairement identifié l'importance de la mesure des performances et de la prévision pour les activités de gestion des agences bancaires, ainsi que le besoin de concevoir un outil de performance et de rentabilité. Cet exercice, essentiel pour Société Générale Algérie comme pour toute banque souhaitant maintenir sa compétitivité et assurer sa croissance.

Notre projet vise, dans une première phase, à déterminer les facteurs influençant la performance et à moderniser les méthodes d'évaluation des agences, en partenariat avec le département du Contrôle de Gestion (CDG). Nous envisageons de proposer une solution Business Intelligence en réalisant des tableaux de bord aux KPIs prédéfinis pour mieux visualiser et cerner la performance, *Quels seront les facteurs clés de performance (KPIs) à prendre en compte ? Comment la Business Intelligence peut-elle aider à mesurer la performance des agences bancaires ?*

La deuxième phase consiste à intégrer ces KPIs dans des modèles prédictifs statistiques et d'apprentissage automatique avancés pour faire la prévision de ces indicateurs à partir de Janvier 2024. *Comment ces modèles peuvent-ils aider à moderniser les prévisions de la rentabilité des agences ? Quelle sera la forme finale de notre outil de mesure de la performance et prédiction de la rentabilité ?*

### Conclusion

Ce chapitre a mis en lumière le rôle essentiel de la Société Générale Algérie (SGA) dans le secteur bancaire national, en détaillant ses services variés, sa présence géographique, et son impact économique. Nous avons exploré l'importance du département de Contrôle de Gestion dans la performance financière de la SGA, ainsi que les procédures budgétaires et les outils de reporting qui optimisent ses opérations.

En fournissant un diagnostic stratégique approfondi, ce chapitre a également exploré les méthodes actuelles de mesure de la performance et de prévision de la rentabilité des agences bancaires de la SGA, ainsi que les facteurs internes et externes qui influencent ces processus. En effet, nous avons identifié les principaux leviers et contraintes qui façonnent la dynamique opérationnelle et stratégique de la banque.

Cette analyse a permis d'expliquer notre problématique et son origine en illustrant comment notre solution pourrait améliorer la gestion de la performance et la prévision de la rentabilité des agences bancaires de la SGA. Ainsi, ce chapitre pose les bases nécessaires pour les recommandations et les solutions qui seront développées dans les chapitres suivants, visant à renforcer l'efficacité et la compétitivité de la Société Générale Algérie.

---

## **Deuxième Chapitre**

### **État de l'art**

---

## Introduction

Ce chapitre présente un état de l'art sur la modernisation du secteur bancaire en Algérie et les technologies financières. Nous débutons par une vue d'ensemble du secteur bancaire algérien et son évolution, suivie de la définition des concepts essentiels du secteur, notamment les indicateurs clés de performance (KPIs). Nous explorons ensuite les technologies financières, les méthodologies de travail pertinentes et leur impact sur la performance financière et la rentabilité des agences bancaires. Cette analyse vise à fournir une compréhension approfondie du paysage bancaire algérien et des outils technologiques modernes utilisés pour optimiser la gestion des performances et assurer une rentabilité accrue.

### 1 Modernisation Bancaire en Algérie : Vers un Futur Numérique Inclusif

Le secteur bancaire algérien, historiquement caractérisé par sa rigidité et son opacité, est en pleine mutation face aux impératifs de modernisation et de digitalisation. Ces transformations sont motivées par les attentes des consommateurs, la concurrence et les directives réglementaires nationales et internationales. En réponse, les banques, y compris la SGA, ajustent leurs stratégies pour améliorer la performance des agences et la rentabilité globale.

En 2023, l'Algérie a démontré une résilience économique notable avec une croissance du PIB de 5,3% malgré une inflation ascendante. La Banque mondiale et le FMI prévoient des perspectives économiques positives pour 2024, soutenues par une hausse des exportations hors hydrocarbures. Ces développements économiques créent un environnement favorable pour l'innovation et la modernisation bancaire.

Le secteur financier et bancaire algérien a connu des réformes profondes visant à sa modernisation et à l'amélioration de sa compétitivité en 2023. L'adoption de la nouvelle loi bancaire et monétaire a constitué un pivot dans cette évolution, introduisant des mécanismes pour la finance islamique, renforçant la gouvernance de la Banque d'Algérie, et jetant les bases pour le développement d'une monnaie numérique nationale, le Dinar Algérien Digital. Ces mesures reflètent un engagement vers l'inclusion financière et la digitalisation du secteur, visant à le préparer pour les défis futur.

La facilitation de l'accès au crédit et l'augmentation des fonds disponibles pour l'économie témoignent également de ces efforts de modernisation. Plus de 10.000 milliards de dinars de crédits ont été alloués à l'économie par les banques et les établissements financiers, soulignant l'engagement de l'État et des institutions financières à soutenir la croissance économique et l'internationalisation du secteur bancaire algérien.

Avec l'adoption croissante de la technologie dans tous les secteurs d'activité, le secteur bancaire n'est pas en reste. La digitalisation des services bancaires est devenue un levier crucial pour améliorer l'efficacité, la sécurité et la personnalisation des services. Société Générale Algérie, en particulier, a commencé à intégrer des solutions numériques pour automatiser les opérations et mieux répondre aux besoins des clients. Cette section illustre comment la

transformation numérique influence les performances des agences bancaires et les préparatifs nécessaires pour migrer vers un environnement plus numérisé.

## 2 Fondements du Secteur Bancaire

### 2.1 La Finance Bancaire

La finance bancaire se distingue de la finance d'entreprise par plusieurs aspects spécifiques liés aux opérations et aux risques auxquels les banques sont exposées, elle englobe diverses activités financières, incluant la gestion des dépôts et des prêts, la gestion des risques, et la gestion des investissements. Ces opérations sont cruciales pour assurer la liquidité et la rentabilité des banques.

- **Gestion des Dépôts et des Prêts :** La gestion efficace des dépôts et des prêts est essentielle pour maintenir la liquidité et la rentabilité de la banque.
- **Gestion des Risques :** Les banques doivent gérer divers risques (crédit, marché, opérationnel) pour assurer leur stabilité financière.
- **Gestion des Investissements :** Les banques investissent dans divers instruments financiers pour maximiser les rendements tout en minimisant les risques.
- **Services Financiers aux Clients :** Les services comme les comptes chèques, les comptes d'épargne, et les prêts hypothécaires sont essentiels pour attirer et fidéliser les clients.
- **Conformité Réglementaire :** Les banques doivent se conformer à des réglementations strictes pour garantir la stabilité du système financier et protéger les intérêts des parties prenantes.

### 2.2 Les Marchés Bancaires :

Les marchés bancaires jouent un rôle essentiel dans l'économie en répondant aux besoins financiers diversifiés des différents segments de la société.

#### **Le marché Retail :**

Le marché Retail se compose de deux segments principaux : le marché CLIPRI et le marché CLIPRO.

- Le marché **CLIPRI**, dédié à la clientèle particulière, constitue un segment essentiel du secteur bancaire, offrant une gamme diversifiée de produits et services adaptés aux besoins individuels des clients. Des comptes chèques aux cartes de retrait en passant par les prêts aux particuliers (crédit conso, crédit immo), les banques rivalisent pour attirer et fidéliser les clients particuliers en proposant des offres compétitives et des services de qualité, tout en mettant l'accent sur une expérience client satisfaisante.
- Le marché **CLIPRO**, quant à lui, se concentre sur la clientèle professionnelle, englobant les petites entreprises ainsi que les professions libérales. Les banques y proposent une variété de solutions financières conçues spécifiquement pour répondre aux besoins des entreprises en matière de gestion financière, de financement des investissements et de

services de trésorerie. L'objectif est d'accompagner ces entreprises dans leur croissance et leur développement en leur offrant un soutien financier et des solutions adaptées à leurs activités.

### **Le marché Corporate :**

Le marché Corporate se compose de deux segments principaux : le marché PME et le marché GE.

- Le marché **PME**, spécifiquement axé sur le soutien des petites et moyennes entreprises, joue un rôle crucial dans le paysage financier en offrant des services financiers variés, tels que le crédit-bail, les lignes de crédit et les services de change. Les banques fournissent aux PME l'accès aux ressources financières nécessaires à leur croissance et à leur succès, contribuant ainsi à stimuler l'activité économique et à créer des emplois.
- Le marché **GE**, dédié aux grandes entreprises et aux multinationales, offre des solutions financières sophistiquées et sur mesure pour répondre aux besoins complexes de ces acteurs économiques majeurs. Des opérations sur le marché des capitaux au financement structuré en passant par les services de trésorerie internationale, les banques agissent en tant que partenaires stratégiques, accompagnant les grandes entreprises dans leurs projets d'expansion, de gestion des risques et de création de valeur.

### **2.3 Les Risques Bancaires :**

Toujours dans le contexte de la mesure des performances et la prédiction de la rentabilité des agences bancaires de la Société Générale Algérie (SGA), il est crucial de comprendre et de gérer efficacement les différents risques bancaires. Ces risques peuvent influencer directement la stabilité financière de chaque agence et, par conséquent, affecter les prévisions de rentabilité à court et à long terme.

- **Risques Structurels :**

Les risques structurels dans le secteur bancaire se rapportent à la "structure" de la banque, observable à travers l'analyse de son bilan comptable. Ces risques sont essentiels à évaluer puisqu'ils peuvent impacter directement la capacité d'une agence à maintenir sa solvabilité et sa liquidité face aux fluctuations économiques

- **Risque de taux :**

Ce risque représente la vulnérabilité de la banque aux variations des taux d'intérêt. Une hausse des taux peut diminuer la valeur des obligations en portefeuille et affecter les marges sur les produits basés sur les intérêts, influençant ainsi directement la rentabilité des agences.

- **Risque de change :**

Pour les agences opérant avec des transactions en devises étrangères, les fluctuations des taux de change peuvent entraîner des pertes non anticipées, affectant la stabilité financière de l'agence.

- **Risque de liquidité :**

Ce risque est particulièrement pertinent pour les agences ayant des obligations de retrait immédiat. Une mauvaise gestion de la liquidité peut forcer l'agence à vendre des actifs précipitamment, souvent à perte, compromettant ainsi sa performance financière.

- **Risque de marché :**

Ce risque concerne la perte potentielle due à des changements dans les conditions de marché, affectant les actifs et passifs de l'agence. La volatilité du marché peut influencer sur les prévisions de rentabilité et exige une gestion prudente et proactive.

- **Risque de contrepartie :**

La défaillance d'une contrepartie à honorer ses engagements peut entraîner des pertes significatives pour l'agence, impactant directement sa performance.

- **Risque de réputation :**

Des incidents affectant l'image de la banque peuvent diminuer la confiance des clients et des investisseurs, réduisant potentiellement les dépôts et autres formes de revenus.

- **Risque opérationnel :**

Les erreurs de processus, les défaillances systémiques, ou les événements externes peuvent causer des pertes substantielles, nécessitant une surveillance rigoureuse et des systèmes de contrôle interne robustes.

- **Autres Risques :**

Les risques non structurels comprennent ceux qui ne sont pas directement liés à la structure financière de la banque mais qui sont tout aussi critiques pour la gestion des performances des agences

### **3 KPIs dans le Secteur Bancaire (Indicateurs de Performance Clés) :**

Les KPIs sont des mesures quantitatives utilisées pour évaluer l'efficacité et l'efficacité des opérations d'une organisation. Dans le contexte bancaire, ces indicateurs aident à mesurer la performance financière, la satisfaction client et l'efficacité opérationnelle.

#### **3.1 Produit Net Bancaire (PNB) :**

Le PNB est un indicateur clé de la performance financière d'une banque, équivalent à la valeur ajoutée pour les entreprises. Il est constitué par la somme de la marge d'intermédiation, les commissions et les autres produits d'exploitation.

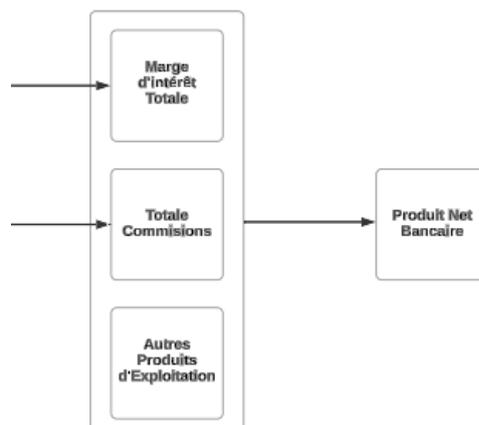


Figure 2.1 - Les composantes du PNB

### 3.1.1 Marge d'intérêt Totale :

Elle est le résultat de la différence entre les intérêts et autres revenus perçus sur les actifs financiers et les coûts de financement des ressources mobilisées. Elle comprend des éléments tels que la rémunération interne des ressources, la marge leasing, la facturation interne des emplois.



Figure 2.2 - Les composantes de la Marge d'intérêt

- **Marge d'intérêt** : C'est probablement la plus importante et la plus directe des composantes de la marge d'intermédiation. Elle représente la différence entre les intérêts perçus sur les prêts, crédits et autres actifs financiers, et les intérêts payés sur les dépôts, dettes et autres passifs financiers. Cette marge reflète l'efficacité avec laquelle une banque gère la répartition de ses ressources et passifs pour maximiser les revenus tout en minimisant les coûts.
- **Marge leasing** : Relatif aux activités de leasing, ce terme désigne la différence entre les paiements reçus des clients pour l'utilisation des équipements ou véhicules loués et les coûts associés au financement de ces biens. La marge de leasing est donc le profit réalisé sur les contrats de leasing après déduction des intérêts payés sur les fonds empruntés pour acheter les biens loués.
- **Rémunération interne des ressources** : Cette composante représente les intérêts et autres revenus générés par l'utilisation des fonds déposés par les clients ou d'autres institutions financières. Cette rémunération est calculée à travers le taux de cession interne (TCI)

- **Facturation interne des emplois** : Cette composante fait référence aux coûts internes générés par l'utilisation des ressources financières au sein de différentes divisions de la banque. Elle peut inclure, par exemple, les coûts associés à l'allocation de capital entre différents départements ou filiales selon leurs besoins en financement. Cette facturation est aussi calculée à travers le taux de cession interne (TCI)

### 3.1.2 Commissions :

Les revenus de commissions sont générés par des services spécifiques fournis aux clients, incluant les opérations de commerce extérieur, les services relatifs aux opérations bancaires quotidiennes, les changes manuels, et divers types de financement. Les commissions peuvent aussi inclure les revenus pour caution domestique, réémissions de garantie internationale, services monétiques et bancaires, et d'autres services spécifiques.



Figure 2.3 - Les composantes des commissions

- **Opérations de commerce extérieur** : Ces commissions sont générées par les services liés au commerce international, incluant le financement du commerce, les lettres de crédit, les garanties bancaires internationales, et la gestion des transactions en devises étrangères. Ces services aident les entreprises à sécuriser et à financer leurs opérations internationales.
- **Services relatifs aux opérations bancaires quotidiennes** : Il s'agit des frais perçus pour la gestion des comptes courants, les transactions bancaires comme les virements, les prélèvements automatiques et la gestion des chèques. Ces commissions couvrent les coûts opérationnels de ces services et représentent une source de revenus régulière pour la banque.
- **Changes manuels** : Ces commissions sont issues des opérations de change manuel, où la banque agit comme intermédiaire pour le client souhaitant convertir des devises. La banque peut percevoir une commission basée sur le montant de la transaction ou offrir un taux de change moins favorable, générant ainsi un profit.
- **Divers types de financement** : Inclut les commissions provenant de services de financement spéciaux tels que les prêts hypothécaires, les crédits à la consommation, et les financements de projets. La banque peut percevoir des frais de dossier, des frais de gestion de prêt ou des commissions basées sur le montant financé.
- **Revenus pour caution domestique** : Ces commissions sont perçus lorsque la banque fournit une garantie au profit d'un client, assurant à un tiers que certaines obligations

contractuelles seront remplies. La commission est souvent un pourcentage de la somme garantie.

- **Réémissions de garantie internationale** : Similaires aux cautions domestiques, mais dans un contexte international. La banque garantit les engagements d'un client envers des partenaires étrangers, sécurisant des transactions internationales complexes.
- **Services monétiques et bancaires** : Les commissions de ce segment incluent les frais pour l'utilisation de cartes de débit, les paiements électroniques, les distributeurs automatiques de billets (DAB), et autres technologies de paiement électronique. La banque peut percevoir des frais par transaction, des frais mensuels, ou des frais annuels pour l'utilisation de ces services.
- **Autres services spécifiques** : Peut inclure des services tels que la gestion de fortune, la consultation financière, les assurances, et d'autres services consultatifs pour lesquels la banque perçoit des commissions.

### 3.1.3 Autres produits et charges d'exploitation :

Cette catégorie englobe les revenus et les coûts qui ne sont pas directement liés aux activités d'intermédiation ou de commission. Cette catégorie est assez variée et peut inclure des éléments moins fréquents mais importants pour la gestion financière de l'entreprise.

## 3.2 Les Frais Généraux (FG) :

Les frais généraux, souvent désignés sous l'acronyme FG, englobent l'ensemble des coûts nécessaires au fonctionnement d'une entreprise mais qui ne sont pas directement attribuables à la production de biens ou à la prestation de services spécifiques. Dans le contexte bancaire, ces frais ne sont pas directement liés aux activités d'intermédiation financière (comme les prêts ou les dépôts) ni aux commissions générées par des services spécifiques (comme les transactions de change ou la gestion de comptes). Ils représentent plutôt les dépenses nécessaires pour maintenir l'infrastructure et les opérations de la banque en état de marche.

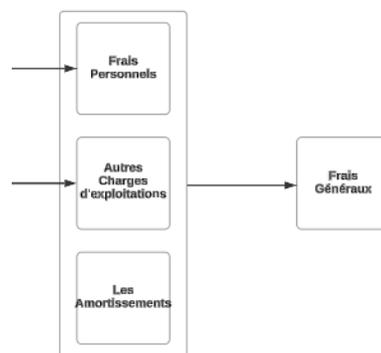


Figure 2.4 - Les composantes des Frais généraux

### 3.2.1 Les Frais de Personnel (FP) :

Les frais de personnel représentent une part importante des frais généraux et comprennent plusieurs éléments liés à la rémunération et aux avantages offerts aux employés

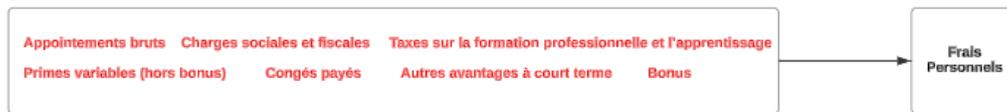


Figure 2.5 - Les composantes des Frais Personnels

- **Appointements bruts** : Salaire de base payé aux employés avant déductions.
- **Charges sociales et fiscales** : Contributions de l'employeur aux assurances sociales et taxes sur les salaires.
- **Congés payés** : Rémunération versée aux employés durant leurs périodes de congé légal.
- **Bonus** : Rémunérations supplémentaires basées sur la performance individuelle ou collective.
- **Primes variables (hors bonus)** : Paiements incitatifs basés sur la réalisation d'objectifs spécifiques, autres que les bonus traditionnels.
- **Autres avantages à court terme** : Avantages divers offerts aux employés, comme les primes de paniers, les contributions aux transports, etc.
- **Taxes sur la formation professionnelle et l'apprentissage (TFPA)** : Taxes spécifiques liées aux coûts de personnel, selon la législation locale.

### 3.2.2 Autres Charges d'Exploitation :

Les Autres Charges d'Exploitation incluent tous les coûts non liés directement à la rémunération des employés mais nécessaires à la gestion quotidienne et opérationnelle de la banque.



Figure 2.6 - Les composantes des Autres charges d'exploitation

- **Provisions pour risques et charges** : Montants servant à couvrir des risques identifiés ou des charges anticipées.

- **Refacturation interne entre entités du groupe** : Charges transférées entre différentes entités de la banque pour des services internes.
- **Taxes et impositions** : Taxes gouvernementales ou locales applicables aux activités de la banque.
- **Formation** : Coûts liés à la formation professionnelle et au développement des compétences des employés.
- **Affranchissement** : Frais de gestion du courrier et des envois postaux.
- **Fourniture et consommables** : Achat de fourniture de bureau et autres consommables nécessaires au fonctionnement quotidien.
- **Frais de communication et télécommunication** : Coûts des services de communication, y compris téléphonie, internet, et autres médias.
- **Frais de missions** : Dépenses liées aux déplacements professionnels, y compris les voyages, l'hébergement, et les indemnités journalières.
- **Frais immobiliers** : Coûts liés à l'entretien, le gardiennage et à la gestion des propriétés de la banque.
- **Frais informatiques** : Dépenses en matériel et logiciels informatiques, maintenance et support technique.
- **Recours à l'extérieur** : Coûts pour services externes, consultants, ou sous-traitance.
- **Autres frais divers** : Dépenses diverses non classées ailleurs.

### 3.2.3 Les Amortissements :

Les amortissements, dans le contexte des frais généraux (FG) d'une entreprise telle qu'une banque, représentent la méthode comptable utilisée pour allouer le coût d'un actif tangible sur sa durée de vie utile. Cette pratique permet de répartir le coût initial de l'actif sur plusieurs périodes comptables, ce qui aide à refléter la diminution de la valeur de l'actif due à l'usage, à l'usure, ou à l'obsolescence.

### 3.3 Le Résultat d'Exploitation (RE) :

Le résultat d'exploitation est un indicateur financier essentiel qui mesure la performance opérationnelle d'une entreprise, comme une banque, indépendamment de ses activités non opérationnelles et des éléments financiers tels que les intérêts et les impôts. Il est calculé en prenant en compte plusieurs éléments clés qui reflètent l'efficacité avec laquelle l'entreprise gère ses opérations principales.

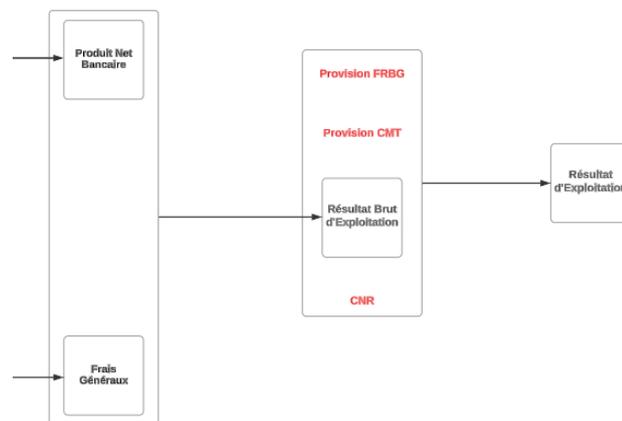


Figure 2.7 - Les composantes du résultat d'exploitation

- **Prévisions FRBG & CMT** (Fonds pour risque bancaires généraux & Crédit à moyen terme) : Il s'agit d'une provision calculée sur la base de la maturité résiduelle des crédits de la banque, la constitution de cette provision est exigée par le régulateur (banque d'Algérie), au même titre que la provision pour CMT qui représente 5% des encours CMT
- **Coût Net du Risque (CNR)** : Le CNR représente le coût des risques associés au crédit et d'autres types de risques financiers. Il est calculé en soustrayant les recouvrements de créances précédemment amorties ou provisionnées des nouvelles charges pour dépréciation et provisions pour pertes sur prêts. Le CNR est un indicateur crucial pour les banques car il reflète directement l'efficacité de leur gestion du risque de crédit.
- **Résultat Brut d'Exploitation (RBE)** : Le RBE (Excédent brut d'exploitation) dans le jargon financier international, mesure la rentabilité opérationnelle. Il est calculé comme étant le revenu généré par les activités principales de la banque moins les coûts opérationnels directs, y compris les salaires, les frais généraux, et autres charges d'exploitation directes. Le RBE est souvent utilisé pour évaluer l'efficacité avec laquelle une entreprise génère du profit à partir de ses opérations principales, sans être influencé par ses décisions de financement, sa structure fiscale, ou ses investissements en capital.

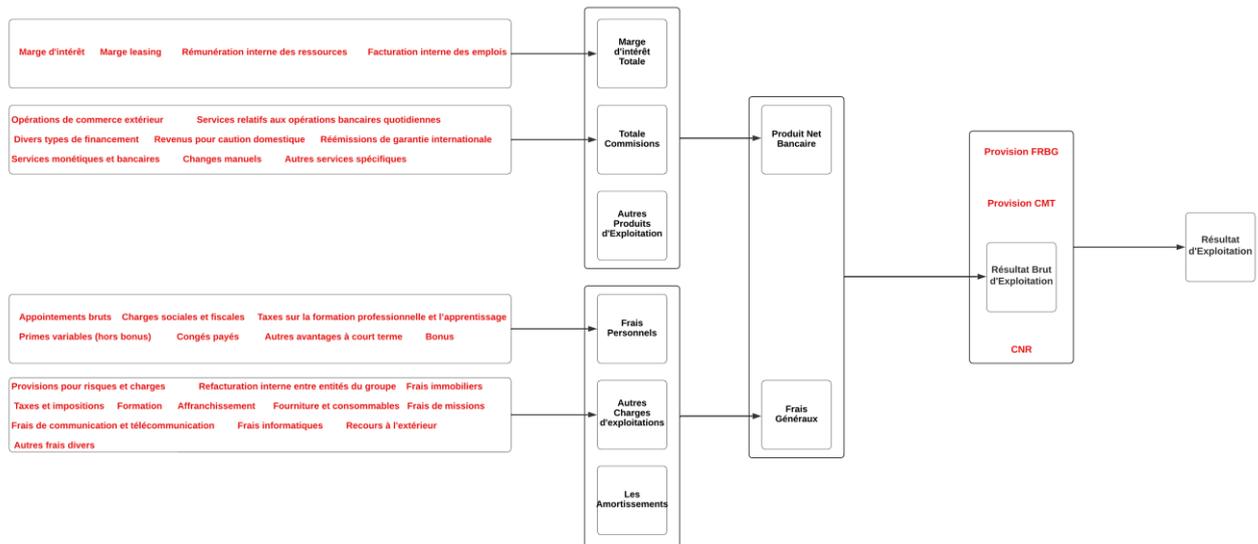


Figure 2.8 - Schéma récapitulatif des différents KPI de la banque

### 3.4 Les Crédits :

Les crédits, dans le contexte bancaire, se réfèrent aux sommes d'argent prêtées par une banque ou une institution financière à ses clients, qui s'engagent à rembourser le montant principal avec des intérêts et éventuellement d'autres frais selon les termes du contrat de prêt. Les crédits sont une composante essentielle de l'activité bancaire, car ils facilitent la circulation de l'argent dans l'économie et permettent aux individus, entreprises et autres entités d'investir, de consommer ou de gérer leurs finances plus efficacement. Les types de crédits sont :

- **Crédit CMT** : représente les crédits moyen termes, en d'autres termes, ce sont des crédits qui ont durée de deux à sept ans. En contrepartie, le bénéficiaire du crédit doit fournir des garanties à la banque, qui est variable en fonction de du montant demandé.
- **Crédit SPOT** : Ce type de crédit est utilisé par les entreprises qui ont besoin d'un remède pour les problèmes de trésorerie. Ce dernier est un crédit court terme, de durée qui peut aller de quinze jours à une année.
- **Crédit syndiqué** : est un crédit qui est partagé entre plusieurs banques pour le compte d'un seul client. Ce cas de figure arrive lorsque le montant du crédit est trop important pour une seule banque, le bénéficiaire du crédit doit fournir des garanties à la banque, qui est variable en fonction de du montant demandé.
- **Crédit Leasing** : Le leasing offre aux clients la possibilité de louer un équipement professionnel pour une durée qui peut aller de trois à cinq ans. À la fin de la période convenue, la banque propose la possibilité de devenir propriétaire du bien (option d'achat)
- **Crédit Cash Pharm** : Ce type de crédit concerne le marché pharmaceutique. Il est important car généralement les remboursements CNAS & CAMSSP peuvent être longs,

ce qui peut mettre les différents pharmaciens dans des situations financières contraignantes. Le CashPharm est la solution à ce problème de trésorerie car il permet de subvenir à ce besoin de courte durée avec des délais de remboursements adaptés.

- **Crédit à la consommation** : est un type de prêt qui concerne les opérations autres que celles relatives à l'immobilier comme pour financer les achats de biens et services, comme les grosses dépenses en biens d'équipement, destinés aux particuliers tels que les électroménagers
- **Crédit Automobile** : comme son nom l'indique, est un crédit qui permet de financer l'achat de véhicules.
- **Crédit Immo** : Cette abréviation utilisée dans le secteur bancaire veut dire crédit immobilier. Ce dernier permet de financer l'ensemble ou une partie de l'acquisition d'un bien immobilier, de l'opération de construction, ou des travaux sur ce dernier.

### 3.5 Les Dépôts :

Les dépôts sont des fonds que les clients confient à une banque pour les conserver en toute sécurité. Ces fonds peuvent être placés dans divers types de comptes, tels que les comptes courants, les comptes d'épargne, ou les dépôts à terme. Les dépôts jouent un rôle crucial dans le fonctionnement des banques, car ils constituent une source principale de financement pour leurs activités de prêt. Les types de Dépôts sont :

- **Dépôt à vue** : Pour ce type de dépôt, la banque ne paye aucun intérêt au client, il peut procéder au retrait ou au versement sans aucune condition.
- **Dépôt épargne** : Un compte d'épargne est un dépôt d'argent à vue dans une banque ou un organisme assimilé, rapportant un intérêt et ne permettant généralement pas d'être utilisé pour faire directement des paiements.
- **Dépôt à terme** : dépôt à terme a une durée fixée à l'avance et rapporte un intérêt à condition que le client titulaire du compte ne retire pas l'argent avant l'échéance (fixée entre un mois et dix ans selon les établissements bancaires). Plus la durée d'immobilisation du dépôt est longue et plus le taux de rémunération est élevé.

## 4 Technologies Financières (FinTech)

La FinTech, abréviation de "technologie financière," désigne l'intégration de technologies innovantes dans les services financiers pour améliorer leur efficacité, accessibilité et personnalisation. Les solutions FinTech ont transformé le paysage financier en rendant les services plus rapides et mieux adaptés aux besoins, tout en facilitant une prise de décision éclairée. Parmi les technologies clés de la FinTech, on trouve la blockchain, l'intelligence artificielle (IA), les big data, les paiements mobiles et, bien sûr, la business intelligence (BI).

Dans cette partie, nous allons explorer les technologies que nous avons choisies pour notre contribution : la Business Intelligence (BI) et l'intelligence artificielle (IA), ainsi que leurs impacts sur la performance bancaire et les méthodologies de travail utilisées dans ce contexte. Nous débutons par une définition de la BI et son impact dans les banques pour améliorer la performance. Ensuite, nous examinons le processus l'ETL (Extraction, Transformation, Chargement), qui est essentiel pour une gestion efficace des données.

Puis, nous abordons le rôle principal de l'IA et les méthodologies existantes, en nous concentrant sur celle que nous avons choisie, le CRISP-DM (Cross Industry Standard Process for Data Mining), et sa contribution à la prévision des indicateurs clés de performance (KPIs). Enfin, nous discutons de l'intersection entre la BI et l'IA, qui partagent une synergie substantielle pour améliorer la prise de décision et optimiser la rentabilité des agences bancaires.

### 4.1 La Business Intelligence (BI) : Définition et Impact sur la Performance Bancaire

La Business Intelligence (BI) est une combinaison de processus, de politiques, de culture et de technologies permettant la collecte, la manipulation, le stockage et l'analyse de données provenant de sources internes et externes. Son objectif est de communiquer des informations, de créer des connaissances et d'éclairer la prise de décision. La BI permet aux entreprises de rendre compte de leurs performances, de découvrir de nouvelles opportunités commerciales et de prendre de meilleures décisions concernant les concurrents, les fournisseurs, les clients, les questions financières, les questions stratégiques, les produits et les services.

Dans le secteur bancaire, où la gestion efficace des données est primordiale, la Business Intelligence joue un rôle central. Les banques collectent une quantité massive de données provenant de diverses sources telles que les transactions, les comptes clients, les prêts et les investissements. La BI leur permet d'analyser ces données pour comprendre le comportement des clients, évaluer les risques, optimiser les opérations et améliorer la rentabilité. Spécifiquement pour Société Générale Algérie (SGA), nous allons utiliser la BI pour mesurer la performance des agences à l'échelle nationale.

Dans la BI, il existe plusieurs modèles de données qui peuvent être utilisés pour organiser et structurer les données de manière efficace. Parmi ces modèles, on retrouve notamment :

- **Modèle en étoile (Star Schema) :** Le modèle en étoile est l'un des modèles de données les plus couramment utilisés. Il se compose d'une table centrale de faits entourée de plusieurs tables de dimensions. La table de faits contient les mesures numériques, tandis que les tables de dimensions contiennent des attributs des données, tels que les

informations sur les produits, les clients, les dates, etc. Ce modèle facilite l'agrégation des données et la création de rapports analytiques.

- **Modèle en flocon (Snowflake Schema) :** Le modèle en flocon est une extension du modèle en étoile où les tables de dimensions sont normalisées en plusieurs tables pour réduire la redondance des données. Cela peut être utile lorsque les dimensions ont des hiérarchies complexes ou des attributs répétitifs. Bien que cela puisse rendre la modélisation plus complexe, cela peut également réduire l'espace de stockage et améliorer les performances des requêtes.
- **Modèle en constellation (Constellation Schema) :** Le modèle en constellation, également connu sous le nom de modèle en galaxie, est une combinaison de plusieurs modèles en étoile. Il permet de gérer des cas où plusieurs tables de faits partagent des tables de dimensions communes. Dans ce modèle, plusieurs tables de faits peuvent être utilisées pour représenter différentes mesures. Les tables de dimensions peuvent être partagées entre les différentes tables de faits, offrant une analyse cohérente et intégrée des données. Ce modèle est particulièrement utile dans les environnements complexes où des analyses multidimensionnelles sont nécessaires, permettant une vue globale et détaillée de différentes facettes de l'activité d'une organisation.

L'implémentation de la BI au sein d'une banque a un impact significatif sur la mesure des performances de ses agences bancaires. Voici quelques-uns des principaux bénéfices :

- **Optimisation des Ressources :** La BI permet d'identifier les agences performantes et celles qui le sont moins, ce qui aide à mieux allouer les ressources comme le capital, le personnel et la technologie.
- **Amélioration de la Prise de Décision :** Avec des données précises et des analyses fiables, les décideurs peuvent planifier efficacement des actions futures, comme l'ouverture de nouvelles agences ou l'amélioration des infrastructures existantes.
- **Stratégie d'Expansion et de Consolidation :** La compréhension des tendances de performance permet à la banque de se concentrer sur les marchés les plus lucratifs pour l'expansion ou la consolidation.
- **Amélioration du Service Client :** L'analyse des performances peut révéler des insights sur la satisfaction des clients et l'efficacité des services, permettant ainsi d'améliorer l'offre de services.
- **Gestion des Risques :** Les analyses rigoureuses aident à mieux gérer les risques financiers, tels que les risques de crédit et de taux d'intérêt, ainsi que les risques opérationnels.

## 4.2 Justification du Choix de la BI par rapport à d'Autres Outils

Bien que plusieurs outils et techniques puissent être utilisés pour mesurer la performance bancaire, la Business Intelligence (BI) offre des avantages distincts qui justifient son choix :

- **Intégration et Cohérence des Données** : La BI permet d'intégrer des données provenant de multiples sources et de les transformer en une vue cohérente, facilitant ainsi une analyse plus précise et complète.
- **Flexibilité et Adaptabilité** : La BI offre une flexibilité inégalée pour adapter les analyses aux besoins spécifiques de la banque, y compris l'intégration de nouvelles sources de données et l'ajustement des transformations pour s'adapter aux changements des exigences commerciales.
- **Support pour l'Analyse Avancée** : En préparant les données de manière structurée, la BI facilite l'application de techniques d'analyse avancées, telles que l'analytique prédictive et le data Mining, permettant de générer des insights approfondis et de prévoir la rentabilité future des agences bancaires.
- **Efficacité de la Prise de Décision** : Avec des données propres et bien organisées, les décideurs peuvent accéder rapidement aux informations pertinentes, ce qui améliore la qualité et la rapidité de la prise de décision.

### 4.3 Data Science et Intelligence Artificielle (DSIA) : Application à la Prédiction de la Rentabilité Bancaire

La data science est une approche multidisciplinaire visant à extraire des informations exploitables à partir des données brutes. Elle englobe des étapes critiques telles que la préparation des données, l'analyse exploratoire, le développement de modèles statistiques et de machine Learning, ainsi que le déploiement de ces modèles pour automatiser et améliorer les prises de décision. La data science est essentiellement tournée vers la prédiction et l'analyse prédictive, jouant un rôle primaire dans l'anticipation des tendances et des comportements futurs.

L'intelligence artificielle (IA) représente un domaine de l'informatique focalisé sur la création de systèmes capables d'exécuter des tâches qui nécessitent l'intelligence humaine. Ces systèmes apprennent à partir de grandes quantités de données, identifiant des patterns et prenant des décisions basées sur ces informations, tout en s'améliorant continuellement grâce à l'expérience. L'IA inclut un vaste éventail de techniques et approches, parmi lesquelles le machine Learning (apprentissage automatique), les réseaux de neurones, et l'apprentissage profond sont particulièrement prédominants.

L'IA est devenue un pilier central dans le secteur financier, transformant radicalement la manière dont les institutions financières opèrent et prennent des décisions.

L'intégration de l'intelligence artificielle (IA) dans la prévision des indicateurs clés de performance (KPIs) des agences bancaires offre un potentiel considérable pour améliorer la modélisation des scénarios de performance et évaluer l'impact sur la rentabilité des institutions financières. Elle permet de combiner des techniques avancées de machine Learning avec l'analyse des séries temporelles, ouvrant ainsi de nouvelles perspectives pour capturer des modèles de comportement complexes et anticiper la réaction des agences face à des conditions économiques et financières variables.

Les techniques de machine Learning appliquées aux séries temporelles permettent de détecter des modèles, des saisons et des comportements non linéaires, ce qui est essentiel pour prédire les variations des KPIs tels que les frais généraux, le produit net bancaire, les encours de crédits et les encours de dépôts. Grâce à l'apprentissage automatique, les modèles peuvent apprendre à partir des données historiques, en prenant en compte les tendances passées, les cycles saisonniers, les effets de long terme et les relations complexes non linéaires entre les différentes variables, il est possible d'obtenir une meilleure estimation des performances futures, une évaluation plus précise de la rentabilité et une meilleure prise de décision pour les agences bancaires.

### **4.3.1 Les domaines de l'intelligence Artificielle :**

L'intelligence artificielle (IA) englobe un vaste éventail de sous-domaines, chacun offrant des techniques et des approches uniques pour résoudre des problèmes complexes et automatiser des processus. En exploitant des modèles sophistiqués et des algorithmes avancés, ces domaines permettent aux systèmes de traiter, d'analyser et d'apprendre à partir de grandes quantités de données. La section suivante explore les principaux domaines de l'IA, en détaillant leurs fonctions, applications et impacts dans divers secteurs. Chacun de ces domaines contribue de manière significative à l'évolution et à l'innovation technologique, offrant des solutions avant-gardistes qui transforment les industries et améliorent notre quotidien.

#### **Apprentissage machine (Machine Learning) :**

L'apprentissage machine est une branche de l'IA qui se concentre sur le développement de systèmes capables d'apprendre à partir de données sans être explicitement programmés. Les algorithmes d'apprentissage machine sont entraînés sur des ensembles de données pour reconnaître des schémas et des tendances, et sont utilisés dans une variété de domaines, y compris la classification, la régression, le clustering et la recommandation et la prévision.

#### **Réseaux neuronaux artificiels (Neural Networks) :**

Les réseaux neuronaux sont des modèles informatiques inspirés du fonctionnement du cerveau humain. Ils sont composés de couches de neurones interconnectés qui transforment l'entrée en sortie en passant par des processus d'apprentissage itératif. Les réseaux neuronaux sont largement utilisés dans des domaines tels que la vision par ordinateur, le traitement du langage naturel, et la reconnaissance de la voix.

#### **Traitement du langage naturel (Natural Language Processing, NLP) :**

Le traitement du langage naturel est une branche de l'IA qui se concentre sur la compréhension et la génération du langage humain par des ordinateurs. Les algorithmes de NLP sont utilisés pour analyser et comprendre le texte écrit ou parlé, et sont utilisés dans des applications telles que la traduction automatique, l'analyse de sentiment, et les chatbots.

#### **Vision par ordinateur (Computer Vision) :**

La vision par ordinateur est un domaine de l'IA qui se concentre sur l'analyse et l'interprétation des images et des vidéos par des ordinateurs. Les algorithmes de vision par ordinateur sont utilisés pour détecter des objets, reconnaître des visages, estimer des poses, et bien d'autres applications, dans des domaines aussi variés que la surveillance, la médecine et l'automobile.

### **Traitement automatique des données (Automated Data Processing) :**

Le traitement automatique des données est une branche de l'IA qui se concentre sur l'automatisation des tâches de collecte, de nettoyage, de transformation et d'analyse des données. Les techniques d'automatisation des données utilisent des algorithmes d'apprentissage machine pour accélérer et optimiser ces processus, permettant aux entreprises de gagner du temps et de l'efficacité dans leurs opérations.

## **4.4 Justification du Choix de l'IA par rapport à d'Autres Outils**

L'adoption de l'intelligence artificielle (IA) dans le domaine de la finance bancaire présente de nombreux avantages par rapport aux autres outils et techniques traditionnels. Voici les principales raisons qui justifient notre choix de l'IA :

- **Précision et efficacité accrues**

Les modèles d'IA, grâce à leur capacité à apprendre et à s'adapter continuellement, offrent une précision accrue dans les prédictions et les analyses. Les techniques d'IA permettent de réduire les erreurs humaines et d'améliorer la fiabilité des résultats. De plus, l'IA peut automatiser de nombreuses tâches analytiques, augmentant ainsi l'efficacité opérationnelle et permettant aux analystes de se concentrer sur des activités à plus forte valeur ajoutée.

- **Détection de patterns complexes et non linéaires**

L'IA excelle dans la détection de motifs complexes et non linéaires au sein des données, des patterns souvent indétectables par les méthodes statistiques traditionnelles. Les modèles d'IA, tels que les réseaux de neurones, peuvent capturer les relations subtiles et les interactions entre les variables, offrant une compréhension plus approfondie et des insights plus précis sur les tendances et les comportements futurs.

- **Adaptabilité et évolutivité**

Les solutions basées sur l'IA sont hautement adaptables et évolutives. Elles peuvent être facilement ajustées pour répondre à de nouveaux défis ou intégrer des données supplémentaires sans nécessiter de restructuration majeure. Cette flexibilité permet aux institutions financières de s'adapter rapidement aux changements du marché et aux nouvelles réglementations, tout en maintenant des performances optimales.

- **Capacité d'apprentissage continu**

Les systèmes d'IA sont conçus pour apprendre en continu à partir de nouvelles données, ce qui leur permet de s'améliorer avec le temps. Cette capacité d'apprentissage continu garantit que les modèles restent pertinents et précis, même dans des environnements dynamiques et changeants. Contrairement aux outils traditionnels qui peuvent nécessiter des mises à jour manuelles fréquentes, les modèles d'IA peuvent s'adapter automatiquement aux nouvelles conditions et aux évolutions du marché.

- **Intégration de multiples sources de données**

L'IA permet d'intégrer et d'analyser des données provenant de diverses sources, y compris des données structurées et non structurées. Cette capacité à unifier des informations disparates en une vue cohérente et exploitable est essentielle pour obtenir des insights holistiques et prendre

des décisions éclairées. Les outils traditionnels, en revanche, peuvent avoir des difficultés à gérer et à intégrer des sources de données hétérogènes de manière efficace.

#### 4.5 Intersection entre la BI et l'IA :

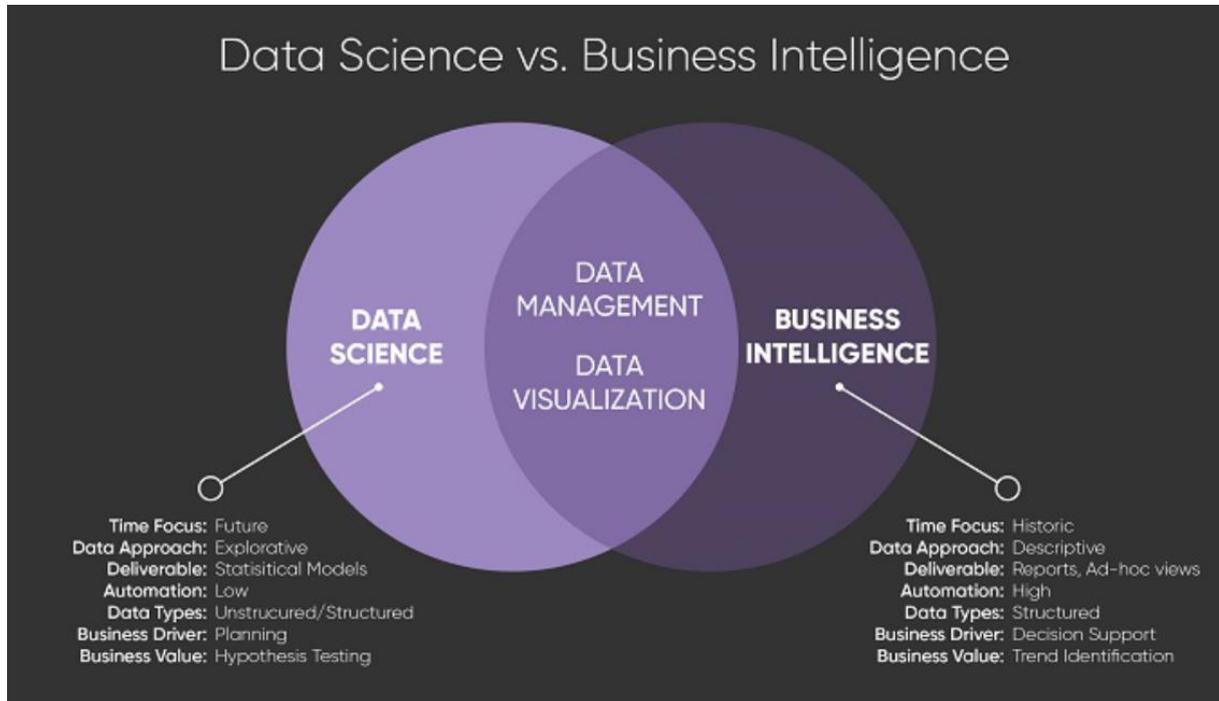


Figure 2.9 - Data Science vs Business Intelligence

Comme le montre la figure ci-jointe, la data science et l'IA partagent une intersection substantielle avec le business intelligence (BI). Cet espace commun inclut la gestion des données et la visualisation des données, deux piliers essentiels pour transformer des données brutes en insights actionnables. Alors que le BI se concentre principalement sur l'analyse descriptive des données historiques pour informer les décisions actuelles, la data science et l'IA adoptent une approche plus exploratoire, visant à modéliser et prédire les futurs possibles. Cette synergie permet aux organisations de non seulement comprendre où elles se trouvent, mais aussi d'où elles viennent et où elles pourraient aller.

En combinant la BI avec la data science et l'IA, Société Générale Algérie peut équilibrer la compréhension des tendances passées avec la prédiction de futures opportunités ou défis de toutes les variables que nous avons déterminés, assurant ainsi une prise de décision basée sur une connaissance approfondie et une anticipation stratégique.

Dans le contexte actuel de la transformation numérique, l'intelligence artificielle (IA) et les modèles statistiques jouent un rôle crucial dans l'optimisation des performances et de la rentabilité des agences bancaires. Ces techniques permettent non seulement d'améliorer la prise de décision stratégique mais aussi d'optimiser les ressources et de maximiser la rentabilité.

## 4.6 Machine Learning et Séries Temporelles :

L'intégration du machine learning et de l'analyse des séries temporelles ouvre de nouvelles perspectives pour la prédiction et l'optimisation des performances financières. Ces deux domaines, lorsqu'ils sont combinés, permettent de tirer parti des vastes quantités de données temporelles générées par les activités bancaires pour améliorer les prévisions, identifier les tendances et prendre des décisions éclairées. Dans cette section, nous explorons d'abord les concepts clés et les principes fondamentaux du machine learning, avant de nous plonger dans les spécificités de l'analyse des séries temporelles, mettant en lumière leurs applications et techniques pour la prédiction des indicateurs financiers critiques des agences bancaires.

### 4.6.1 Définition et Principes Fondamentaux :

Le machine learning, ou apprentissage automatique, est une sous-discipline de l'IA qui se concentre sur le développement d'algorithmes permettant aux ordinateurs d'apprendre à partir des données. Les algorithmes de machine learning identifient des patterns et utilisent ces connaissances pour effectuer des prédictions ou des décisions sans être explicitement programmés pour chaque tâche.

Les principes fondamentaux du machine learning incluent :

**Supervised Learning (Apprentissage supervisé) :** Les algorithmes apprennent à partir de données étiquetées, c'est-à-dire que chaque exemple d'apprentissage est accompagné de la réponse correcte. Par exemple, pour prédire les produits nets bancaires d'une agence, les données historiques avec les valeurs de ces produits sont utilisées pour entraîner le modèle.

**Unsupervised Learning (Apprentissage non supervisé) :** Les algorithmes cherchent à identifier des structures ou des motifs dans des données non étiquetées. Cela peut être utilisé pour segmenter les clients en fonction de leurs comportements bancaires afin d'optimiser les services proposés.

**Reinforcement Learning (Apprentissage par renforcement) :** Les algorithmes apprennent à prendre des décisions en interagissant avec un environnement et en recevant des récompenses ou des punitions en fonction de leurs actions. Ce type d'apprentissage peut être appliqué pour optimiser les opérations bancaires en temps réel.

### Capacité et Surapprentissage

Il y a un équilibre délicat entre la capacité d'un modèle à s'adapter aux données d'entraînement et à généraliser aux données de test. Un modèle trop complexe peut s'adapter aux données d'entraînement mais échouer à généraliser (surapprentissage), tandis qu'un modèle trop simple peut échouer à capturer des tendances significatives dans les données d'entraînement (sous-apprentissage). Par exemple, un modèle trop complexe pourrait capturer du bruit plutôt que de véritables tendances, rendant les prévisions peu fiables.

### **Compromis Biais -Variance**

Le biais se réfère à l'erreur systématique introduite par le choix d'un modèle spécifique, tandis que la variance se réfère à l'erreur due à la sensibilité du modèle à la fluctuation dans les données d'entraînement. Un équilibre approprié entre le biais et la variance est essentiel pour obtenir un modèle performant. Dans le contexte bancaire, cela signifie choisir des modèles qui capturent les tendances générales sans être trop influencés par les fluctuations saisonnières des données financières.

### **Validation et Évaluation**

Les modèles de machine learning doivent être validés sur des ensembles de données indépendants pour évaluer leur performance et éviter le surapprentissage. De plus, différentes métriques d'évaluation peuvent être utilisées selon le problème à résoudre. Pour mesurer la rentabilité d'une agence bancaire, des métriques comme le Mean Absolute Error (MAE) ou le Root Mean Squared Error (RMSE) peuvent être utilisées pour évaluer la précision des prédictions des KPIs.

### **Importance des Données**

Les modèles de machine learning sont aussi bons que les données sur lesquelles ils sont formés. Les données de haute qualité, représentatives et bien préparées sont essentielles pour former des modèles efficaces. Dans le contexte bancaire, cela inclut le traitement des valeurs manquantes, la détection et la gestion des valeurs aberrantes, et la transformation appropriée des variables comme les encours de crédits et dépôts.

### **Interprétabilité et Explicabilité**

L'interprétabilité fait référence à la mesure dans laquelle un humain peut comprendre le processus de prise de décision d'un modèle, tandis que l'explicabilité fait référence à la mesure dans laquelle les prédictions d'un modèle peuvent être expliquées en termes humains. Avec l'essor des modèles de machine learning de plus en plus complexes, tels que les réseaux de neurones profonds, ces deux aspects sont devenus des domaines de recherche actifs. Par exemple, expliquer pourquoi un modèle prédit une baisse des produits nets bancaires est crucial pour les décideurs.

#### **4.6.2 Séries Temporelles**

Les séries temporelles sont des données chronologiques qui enregistrent des observations successives d'un phénomène à des intervalles de temps réguliers. Dans le contexte bancaire, elles sont couramment utilisées pour analyser et prévoir des indicateurs clés de performance (KPIs) comme les frais généraux, le produit net bancaire, les encours de crédits et de dépôts, et d'autres mesures de performance financière des agences.

#### 4.6.2.1 Les Composantes d'une Série Chronologique

Une série chronologique se décompose généralement en plusieurs composantes :

- **Tendance (Trend)** : La composante à long terme qui représente la direction générale de la série. Par exemple, une tendance à la hausse du produit net bancaire peut indiquer une amélioration continue des performances d'une agence.
- **Saisonnalité (Seasonality)** : Les variations périodiques qui se répètent à intervalles réguliers. Par exemple, certaines agences peuvent enregistrer une augmentation des dépôts pendant les périodes de vacances.
- **Cycle (Cycle)** : Les fluctuations qui se produisent à intervalles irréguliers, souvent influencées par des facteurs économiques. Par exemple, les cycles économiques peuvent affecter les taux de crédit et les niveaux de risque associés.
- **Résidu (Residual)** : Les variations irrégulières ou aléatoires qui ne peuvent être attribuées à l'une des composantes ci-dessus, comme des événements imprévus affectant les frais généraux.

#### 4.6.2.2 La Transformation de Série Chronologique

La transformation des séries chronologiques vise à rendre les données stationnaires, c'est-à-dire à éliminer les tendances et les variations saisonnières afin de faciliter la modélisation et la prévision. Les techniques courantes incluent :

- **Transformation Logarithmique** : Une fonction logarithmique est appliquée aux valeurs de la série temporelle pour stabiliser la variance. Par exemple, la transformation  $X_t = \log(Y_t)$  peut être utilisée pour des séries chronologiques à croissance exponentielle, comme les volumes de transactions.
- **Différenciation (Differencing)** : Calculer les différences successives entre les observations pour éliminer les tendances. Par exemple, la première différence est calculée comme suit :

$$Y'_t = Y_t - Y_{t-1}$$

Si cette transformation n'est pas suffisante pour obtenir une série stationnaire, une différenciation de second ordre peut être appliquée :

$$Y''_t = Y'_t - Y'_{t-1}$$

Pour éliminer les effets saisonniers, une différenciation saisonnière est effectuée :

$$Y'_t = Y_t - Y_{t-m}$$

Où  $m$  est le nombre de saisons.

- **Smoothing** : Utiliser des moyennes mobiles ou d'autres techniques pour lisser les données et réduire la variabilité. Cela peut être appliqué pour lisser les fluctuations mensuelles des frais généraux.

- **Décomposition** : Séparer les composantes de tendance, de saisonnalité et de résidu. Par exemple, décomposer les séries de crédits et de dépôts pour analyser séparément les tendances à long terme et les variations saisonnières.

#### 4.6.2.3 Techniques de Prédiction des Séries Temporelles

Les séries temporelles sont une suite d'observations recueillies à des intervalles de temps réguliers. La prédiction des séries temporelles fait référence à l'utilisation de modèles statistiques pour prédire les points de données futurs basés sur les données passées.

Plusieurs techniques sont couramment utilisées :

- **Méthode de lissage exponentiel simple** : Cette technique utilise un facteur de lissage pour pondérer l'impact des observations passées sur les prévisions futures. Elle est utile lorsque les données sont relativement stables sans tendance ou saisonnalité discernable. Le principal inconvénient de cette méthode est qu'elle peut sous-estimer les prévisions lorsque les données présentent une tendance.
- **Méthode de lissage exponentiel double (ou méthode de Holt)** : Cette méthode étend la méthode de lissage exponentiel simple pour inclure une composante de tendance. Elle est plus adaptée lorsque les données présentent une tendance, qu'elle soit croissante ou décroissante.
- **Méthode de lissage exponentiel triple (ou méthode de Holt-Winters)** : Cette méthode ajoute une composante de saisonnalité à la méthode de Holt. Elle est utilisée lorsque les données montrent à la fois une tendance et une saisonnalité.
- **Modèles autorégressifs (AR)** : Ces modèles prédisent les valeurs futures en fonction des valeurs passées. Ils utilisent la dépendance entre une observation et un certain nombre d'observations passées (observations retardées). Ces modèles sont utiles lorsque les données montrent des corrélations à court terme.
- **Modèles de moyenne mobile (MA)** : Ces modèles prédisent les valeurs futures en se basant sur les erreurs des prévisions passées. Ils utilisent la dépendance entre une observation et une erreur résiduelle à partir d'une moyenne mobile appliquée à des observations passées.
- **Modèles ARIMA (AutoRegressive Integrated Moving Average)** : Ces modèles combinent les modèles AR et MA et tentent également de supprimer les tendances dans les données (c'est-à-dire qu'ils sont 'intégrés'). Les modèles ARIMA sont utiles lorsque les données présentent une tendance et/ou une saisonnalité.
- **Modèles de régression** : Ils sont utilisés lorsque la variable que nous voulons prédire est corrélée à une ou plusieurs variables indépendantes. Les modèles de régression peuvent être linéaires ou non linéaires.
- **Modèles de réseaux de neurones** : Ces modèles sont basés sur des systèmes d'apprentissage qui tentent de simuler la manière dont le cerveau humain apprend. Ils sont particulièrement utiles lorsque les relations entre les variables sont non linéaires ou lorsque les données contiennent de nombreuses variables d'entrée.

Chacune de ces techniques à ses propres avantages, limites et hypothèses. L'efficacité de ces méthodes peut varier en fonction des caractéristiques spécifiques de la série temporelle.

### 4.6.3 Algorithmes identifiés :

#### 4.6.3.1 ARIMA :

Le modèle ARIMA (AutoRegressive Integrated Moving Average) est une méthode de modélisation de séries temporelles introduite par George Box et Gwilym Jenkins en 1976. Il est conçu pour gérer les séries temporelles non stationnaires en utilisant des différences pour rendre la série stationnaire avant de l'appliquer à un modèle ARMA (AutoRegressive Moving Average). Les modèles ARIMA sont souvent utilisés dans le domaine de la prévision économique, les prévisions météorologiques, la modélisation de la bourse et bien d'autres domaines où l'analyse de séries temporelles est nécessaire.

#### Paramètres du Modèle ARIMA

Un modèle ARIMA est défini par trois paramètres :  $p$ ,  $d$  et  $q$

- $p$  : l'ordre du composant autoregressif (AR)
- $d$  : l'ordre de la différence nécessaire pour rendre la série temporelle stationnaire (I)
- $q$  : l'ordre du composant de moyenne mobile (MA)

En combinant ces trois composants, le modèle ARIMA est capable de modéliser une variété de séries temporelles complexes.

#### Équations Principales du Modèle ARIMA

##### Composant AutoRegressive (AR) :

L'AR fait référence à la régression linéaire des valeurs actuelles sur les valeurs précédentes de la série temporelle. Il capture les effets de dépendance linéaire à partir des valeurs passées.

$$AR(p) = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

- $Y_t$  : série temporelle à l'instant  $t$
- $C$  : constante
- $\varphi$  : coefficients autoregressifs
- $\varepsilon_t$  : erreur à l'instant  $t$

##### Composant Integrated (I) :

La différenciation est utilisée pour rendre les données stationnaires en supprimant les tendances et les comportements saisonniers. Elle consiste à prendre la différence entre les observations consécutives jusqu'à obtenir une série stationnaire.

$$I(d) : \nabla^d Y_t$$

- $\nabla^d$  : Opérateur de différenciation d'ordre  $d$ .

### Composant Moving Average (MA) :

Le MA utilise la moyenne mobile des erreurs précédentes pour modéliser la relation entre les résidus et les observations actuelles de la série temporelle. Il capture les effets de dépendance entre les résidus.

$$MA(q) : Y_t = C + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

- $\theta$  : coefficients de moyenne mobile

### 4.6.3.2 SARIMA :

Le modèle SARIMA (Seasonal AutoRegressive Integrated Moving Average) est une extension du modèle ARIMA qui prend en compte la saisonnalité des séries temporelles. Il est particulièrement utile lorsque les données présentent des motifs saisonniers réguliers qui doivent être modélisés pour améliorer la précision des prévisions. Le modèle SARIMA intègre des composants saisonniers aux composants non saisonniers d'ARIMA, permettant ainsi de capturer à la fois les tendances régulières et saisonnières dans les données.

#### Paramètres du Modèle SARIMA

Un modèle SARIMA est défini par six paramètres :  $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ , et  $Q$ , ainsi que la périodicité  $s$ .

- $p$  : l'ordre du composant autoregressif (AR) non saisonnier
- $d$  : l'ordre de la différenciation non saisonnière nécessaire pour rendre la série temporelle stationnaire
- $q$  : l'ordre du composant de moyenne mobile (MA) non saisonnier
- $P$  : l'ordre du composant autoregressif (AR) saisonnier
- $D$  : l'ordre de la différenciation saisonnière nécessaire pour rendre la série temporelle stationnaire
- $Q$  : l'ordre du composant de moyenne mobile (MA) saisonnier
- $s$  : la périodicité de la saisonnalité

#### Formulation du Modèle SARIMA

En combinant les composants non saisonniers et saisonniers, un modèle SARIMA peut être noté sous la forme SARIMA ( $p, d, q$ ) ( $P, D, Q$ ) <sub>$s$</sub> , où  $s$  représente la période de saisonnalité. L'équation générale du modèle SARIMA est :

$$\Phi(Bs) \phi(B) \nabla^s D \nabla^d Y_t = C + \Theta(Bs) \theta(B) \varepsilon_t$$

Où :

- $\Phi(Bs)$  et  $\phi(B)$  représentent les polynômes des composants autoregressifs saisonniers et non saisonniers, respectivement.
- $\Theta(Bs)$  et  $\theta(B)$  représentent les polynômes des composants de moyenne mobile saisonniers et non saisonniers, respectivement.
- $\nabla_s D$  et  $\nabla d$  sont les opérateurs de différenciation saisonnière et non saisonnière.

### Application du Modèle SARIMA

Le modèle SARIMA est particulièrement adapté pour les séries temporelles présentant des motifs saisonniers réguliers, comme les cycles économiques trimestriels ou annuels dans les données financières des agences bancaires. En tenant compte à la fois des variations saisonnières et non saisonnières, SARIMA offre une modélisation plus complète et précise des KPIs bancaires.

#### 4.6.3.3 Auto ARIMA :

Le modèle Auto ARIMA est une extension automatisée du modèle ARIMA, conçue pour simplifier le processus de modélisation des séries temporelles en déterminant automatiquement les meilleurs paramètres  $p$ ,  $d$ , et  $q$ .

Cette méthode utilise des critères statistiques pour identifier les valeurs optimales, évitant ainsi la tâche complexe et souvent subjective de sélection des paramètres par l'utilisateur. Auto ARIMA est particulièrement utile pour les séries temporelles complexes où les meilleures valeurs de  $p$ ,  $d$ , et  $q$  ne sont pas immédiatement évidentes.

#### Fonctionnement du Modèle Auto ARIMA

Le modèle Auto ARIMA fonctionne en testant différentes combinaisons de paramètres  $p$ ,  $d$ , et  $q$  et en évaluant chaque modèle basé sur des critères tels que l'Akaike Information Criterion (AIC) ou le Bayesian Information Criterion (BIC). L'objectif est de minimiser ces critères pour identifier le modèle le plus approprié pour les données fournies.

#### AIC (Akaike Information Criterion) :

$$AIC = 2k - 2 \ln(L)$$

Où  $k$  est le nombre de paramètres dans le modèle et  $L$  est la valeur de la fonction de vraisemblance maximale. L'AIC mesure la qualité d'un modèle tout en pénalisant la complexité du modèle. Plus la valeur de l'AIC est faible, meilleur est le modèle.

#### BIC (Bayesian Information Criterion) :

$$BIC = \ln(n) k - 2 \ln(L)$$

Où  $n$  est le nombre d'observations. Le BIC est similaire à l'AIC, mais il pénalise plus fortement la complexité du modèle, favorisant les modèles plus simples.

#### Étapes de Mise en Œuvre du Modèle Auto ARIMA

1. **Identification de la Stationnarité** : Le modèle commence par vérifier si la série temporelle est stationnaire. Si ce n'est pas le cas, il applique des transformations, telles que la différenciation, pour rendre la série stationnaire.
2. **Sélection des Paramètres** : Auto ARIMA effectue une recherche exhaustive ou séquentielle sur les valeurs possibles de  $p$ ,  $d$ , et  $q$ . Il évalue chaque combinaison en utilisant des critères comme l'AIC ou le BIC pour sélectionner le modèle le plus approprié.
3. **Estimation des Paramètres** : Une fois les meilleurs paramètres identifiés, le modèle estime les coefficients autoregressifs et de moyenne mobile ( $\phi$  et  $\theta$ ).
4. **Validation du Modèle** : Le modèle est ensuite validé en utilisant des ensembles de données de test pour vérifier sa capacité de généralisation et éviter le surapprentissage.
5. **Prédiction** : Avec le modèle validé, Auto ARIMA effectue les prédictions sur les valeurs futures de la série temporelle.

### Extension à Auto-SARIMA

Le modèle Auto-ARIMA peut être étendu à Auto-SARIMA (Seasonal Auto-Regressive Integrated Moving Average) en définissant l'hyperparamètre **seasonal = True**. Lorsque cette option est activée, le modèle détermine également les meilleurs hyperparamètres saisonniers  $P$ ,  $D$ ,  $Q$  pour capturer les effets saisonniers dans les données.

#### Auto-SARIMA :

- Le modèle Auto-SARIMA ajoute une composante saisonnière en plus des composants non saisonniers  $p$ ,  $d$ , et  $q$ . Les paramètres saisonniers  $P$ ,  $D$ ,  $Q$  et la période saisonnière  $s$  sont déterminés automatiquement pour mieux capturer les variations périodiques des séries temporelles.
- L'évaluation des modèles avec des critères comme l'AIC ou le BIC prend également en compte les paramètres saisonniers pour sélectionner le modèle optimal.

Ainsi, l'utilisation d'Auto-SARIMA permet d'améliorer la précision des prédictions en tenant compte des variations saisonnières des données, offrant ainsi une modélisation plus complète et précise des séries temporelles.

#### 4.6.3.4 RNN :

Les réseaux de neurones récurrents (RNN) sont des algorithmes d'apprentissage profond spécifiquement conçus pour traiter des séquences de données, telles que les séries temporelles. Contrairement aux réseaux de neurones classiques, les RNN peuvent prendre en compte l'historique des données grâce à une boucle de rétroaction, permettant ainsi de conserver une mémoire des données passées et d'utiliser cette information pour traiter les données futures. Cela les rend particulièrement adaptés à la modélisation des séries temporelles complexes rencontrées dans le secteur bancaire.

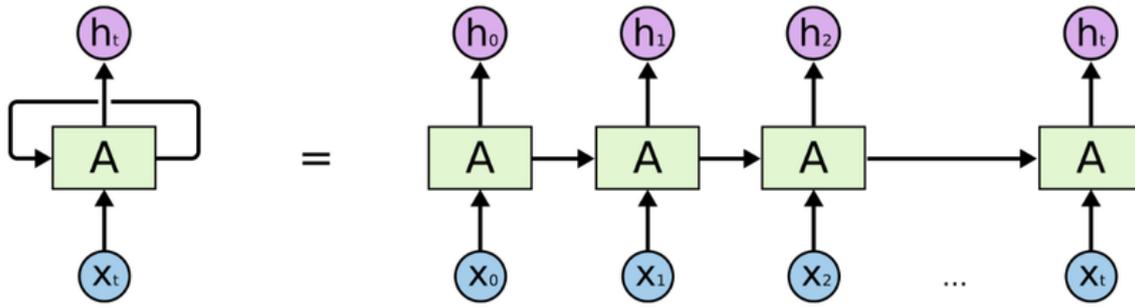


Figure 2.10 - Architecture du RNN

### Fonctionnement du Modèle RNN

La structure de base d'un RNN consiste en une couche cachée de neurones qui est connectée à elle-même. À chaque étape temporelle, le RNN prend en entrée les données actuelles et l'état caché de l'étape temporelle précédente, calcule la sortie et met à jour l'état caché.

- 1- État caché :** Le RNN conserve une mémoire à court terme des données d'entrée précédentes via l'état caché, noté  $h_t$ . Cet état est mis à jour à chaque étape temporelle en utilisant les données d'entrée actuelles  $x_t$  et l'état caché précédent  $h_{t-1}$ .

$$h_t = f(x_t, h_{t-1})$$

- 2- Sortie :** La sortie du RNN à chaque étape temporelle est calculée à partir de l'état caché actuel.

$$y_t = g(h_t)$$

- 3- Fonction de Perte :** La fonction de perte  $J$  mesure l'écart entre la sortie prédite  $y_t$  et la sortie attendue, et est utilisée pour ajuster les paramètres du réseau afin de minimiser cette perte.

$$J = -\frac{1}{N} \sum_{t=1}^N \log(P(y_t | x_1, \dots, x_N))$$

### Défis et Variantes des RNN

Les RNN traditionnels rencontrent des problèmes de vanishing gradients, où la rétropropagation de l'erreur à travers de nombreuses étapes temporelles peut entraîner une diminution rapide du gradient, ralentissant ainsi l'apprentissage. Pour surmonter ce problème, des variantes telles que les LSTM (Long Short-Term Memory) et les GRU (Gated Recurrent Units) ont été développées. Ces variantes utilisent des mécanismes de portes pour réguler le flux d'informations et éviter la disparition des gradients.

Les réseaux de neurones récurrents (RNN) sont des outils puissants pour la modélisation des séries temporelles, en particulier pour les séries présentant des dépendances temporelles complexes. Dans le contexte bancaire, les RNN permettent de prédire avec précision les KPIs critiques. En intégrant des variantes avancées comme les LSTM et les GRU, les RNN peuvent surmonter les problèmes de vanishing gradients et offrir des prévisions robustes et précises,

contribuant ainsi à une meilleure gestion des performances et à l'optimisation de la rentabilité des agences bancaires.

#### 4.6.3.5 LSTM :

Les LSTM (Long Short-Term Memory) sont une variante de réseaux de neurones récurrents (RNN) conçue pour résoudre les problèmes de vanishing gradients souvent rencontrés dans les RNN traditionnels. Ce problème survient lorsque les gradients diminuent exponentiellement à mesure que l'algorithme de rétropropagation se propage dans le temps, rendant l'apprentissage inefficace pour les séquences longues. Les LSTM utilisent des mécanismes de portes pour réguler le flux d'informations et contrôler la quantité d'information stockée dans l'état caché, permettant ainsi de conserver des informations sur de longues périodes.

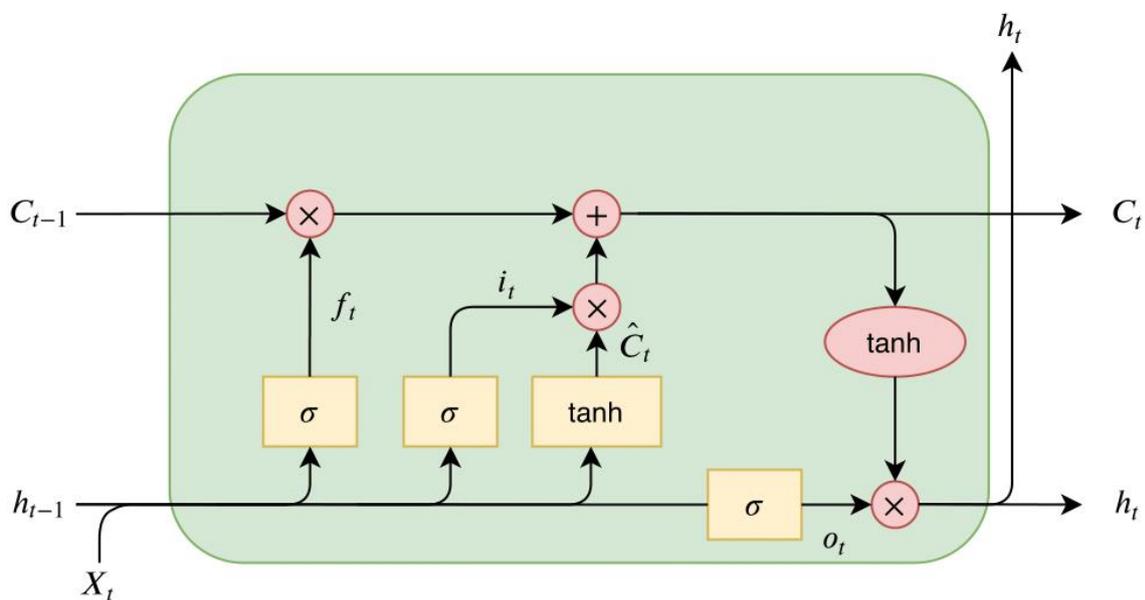


Figure 2.11 - Architecture du LSTM

#### Structure des LSTM

La structure de base d'un LSTM consiste en une cellule LSTM qui inclut trois portes principales : la porte d'entrée, la porte de sortie et la porte d'oubli. Chaque porte contrôle le flux d'informations dans et hors de la cellule LSTM.

- **Porte d'Entrée (Input Gate) :** Contrôle la quantité d'informations nouvelles à ajouter à l'état caché, en fonction des données d'entrée et de l'état caché précédent.
- **Porte d'Oubli (Forget Gate) :** Détermine la quantité d'informations précédentes à oublier, en fonction des données d'entrée et de l'état caché précédent.
- **Porte de Sortie (Output Gate) :** Contrôle la quantité d'informations à sortir de la cellule LSTM, en fonction des données d'entrée et de l'état caché actuel.

### États Cachés

L'état caché dans un LSTM est composé de deux parties :

- **État caché à court terme (Short-Term State)** : Stocke les informations actuelles de la séquence.
- **État caché à long terme (Long-Term State)** : Stocke les informations pertinentes de la séquence jusqu'à présent. Les portes de la cellule LSTM contrôlent la quantité d'informations stockées dans ces états.

### Équations Principales du Modèle LSTM :

#### Équation de la Porte d'Entrée :

$$i(t) = \sigma(W_i [h(t-1), x(t)] + b_i)$$

- Où  $x(t)$  est l'entrée à l'étape temporelle  $t$ ,  $h(t-1)$ , est l'état caché précédent, et  $W_i$  et  $b_i$  sont les poids et biais appris pour la porte d'entrée. La fonction  $\sigma$  représente l'activation sigmoïde.

#### Équation de la Porte d'Oubli :

$$f(t) = \sigma(W_f [h(t-1), x(t)] + b_f)$$

#### Équation de la Porte de Sortie :

$$o(t) = \sigma(W_o [h(t-1), x(t)] + b_o)$$

#### Equation de Mise à Jour de la Cellule :

$$C(t) = f(t) * C(t-1) + i(t) * \tanh(W_c [h(t-1), x(t)] + b_c)$$

- Où  $C(t)$  est l'état de la cellule à l'étape temporelle  $t$  et  $\tanh$  est la fonction d'activation tangente hyperbolique.

#### Equation de l'État Caché :

$$h(t) = o(t) * \tanh(C(t))$$

Les LSTM représentent une avancée significative dans le domaine des réseaux de neurones récurrents, en particulier pour la modélisation des séries temporelles avec des dépendances à long terme. Dans le contexte bancaire, les LSTM permettent de prévoir avec précision des KPIs critiques, en intégrant les informations pertinentes sur de longues périodes. Cela améliore la capacité des banques à planifier, à gérer les risques et à optimiser leur rentabilité, en fournissant des prévisions robustes et précises basées sur les données historiques.

## 5 Méthodologie de Travail :

### 5.1 Processus ETL : Mesure de Performance des agences bancaires

Le processus ETL (Extract, Transform, Load) est une étape clé du traitement des données dans le domaine décisionnel. Il implique l'Extraction des données à partir de différentes sources, la Transformation de ces données selon les besoins spécifiques, puis Load (le chargement) des données transformées dans un entrepôt de données (Data Warehouse) ou un magasin de données (Data Mart) Le processus ETL est une étape primordiale pour garantir la qualité et la pertinence des données utilisées dans les analyses décisionnelles. Il nécessite une planification minutieuse, des outils appropriés et des validations régulières pour s'assurer que les données sont correctement extraites, transformées et chargées dans le système décisionnel.

- **Extraction (Extract) :**

Cette étape consiste à extraire les données brutes à partir de différentes sources de données. Ces sources incluent des bases de données relationnelles, des fichiers plats, des API web, des flux de données en continu, etc.

L'extraction des données peut se faire de manière complète (c'est-à-dire extraire toutes les données à chaque exécution du processus ETL) ou incrémentielle (où seules les données modifiées depuis la dernière exécution sont extraites).

- **Transformation (Transform) :**

Une fois que les données sont extraites, elles sont transformées pour répondre aux besoins spécifiques de l'entrepôt de données ou du système de destination. Les transformations peuvent inclure le nettoyage des données, la normalisation, la fusion de données provenant de différentes sources, la déduplication, la conversion de formats, etc.

Cette étape est souvent la plus complexe du processus ETL, car elle nécessite souvent l'utilisation de langages de programmation ou d'outils ETL spécifiques pour manipuler les données de manière efficace.

- **Chargement (Load) :**

Une fois que les données ont été extraites et transformées, elles sont chargées dans l'entrepôt de données ou le système de destination. Cela peut impliquer l'insertion des données dans une base de données relationnelle, le chargement de fichiers plats dans un système de fichiers, l'envoi de données à un service de cloud computing, etc.

Le chargement des données peut également nécessiter des opérations de validation pour garantir que les données sont correctement chargées et conformes aux règles métier et aux contraintes définies.

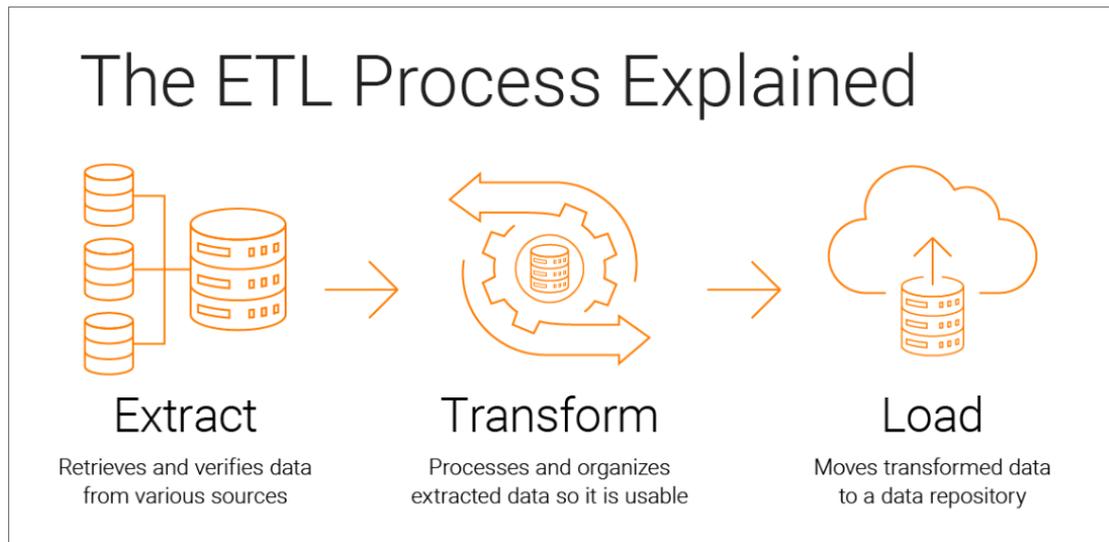


Figure 2.12 - Processus ETL

## 5.2 Approche CRISP-DM : Prédiction de la Rentabilité des Agences Bancaires

Pour réaliser la seconde partie du Projet, nous avons adopté la méthodologie CRISP-DM (Cross Industry Standard Process for Data Mining), reconnue pour sa flexibilité et son efficacité dans les projets de data mining et d'analyse prédictive. Cette méthodologie se compose de six phases, chacune jouant un rôle primaire dans l'atteinte de nos objectifs :

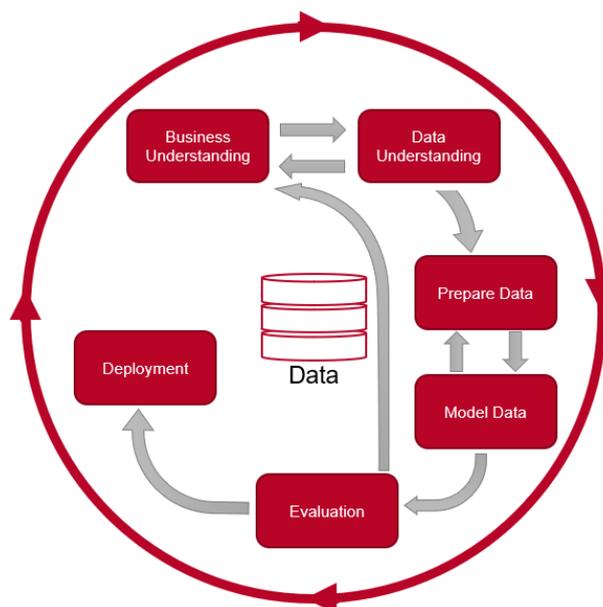


Figure 2.13 - Cycle de vie de la méthodologie CRISP-DM

### **Compréhension du problème et des Besoins Métier :**

- Objectif : Comprendre les objectifs et les exigences de la Société Générale Algérie en matière de prédiction de la rentabilité.
- Actions : Entretiens avec les parties prenantes, analyse des processus existants, identification des KPIs liés à la rentabilité.

### **Collecte et Compréhension des Données :**

- Objectif : Collecter les données nécessaires et en analyser la qualité et la pertinence.
- Actions : Extraction des données historiques (de janvier 2018 à décembre 2023), exploration des données.

### **Préparation des Données :**

- Objectif : Transformer et préparer les données en un format utilisable pour l'analyse et la modélisation.
- Actions : Nettoyage et préparation de données pour être utilisées pour l'analyse et la modélisation. Cette étape peut également inclure l'élimination des données redondantes, la création de nouvelles variables, la normalisation des données, la transformation des dates, l'application de tables pivot pour créer des séries temporelles et la différenciation pour rendre les séries stationnaires.

### **Modélisation :**

- Objectif : Développer des modèles statistiques et d'apprentissage automatique pour la prédiction des KPIs.
- Actions : Sélection des modèles appropriés (modèles statistiques et réseaux de neurones), Entraînement de ces modèles sur les données préparées.

### **Évaluation :**

- Objectif : Valider la pertinence des modèles développés en fonction des critères définis.
- Actions : Utilisation des métriques d'évaluation pour chaque modèle utilisé, ajustement pour améliorer les précisions et validation des modèles avec les parties prenantes.

### **Déploiement :**

- Objectif : Implémenter la solution développée et assurer sa mise en production.
- Actions : Intégration des résultats des modèles dans un outil de performance et prédiction de la rentabilité bancaire, La mise en place d'un système de prise de décision automatisé et présentation des résultats aux parties prenantes.

## **Conclusion :**

Ce chapitre a permis, à travers l'exploration de l'état de l'art, d'identifier les méthodologies de travail pertinentes en se basant sur des concepts théoriques solides. Nous avons mis en lumière l'importance des réformes et des innovations technologiques dans le secteur bancaire algérien, tout en détaillant les indicateurs clés de performance (KPIs) essentiels pour évaluer et optimiser les opérations bancaires. L'intégration des technologies de Business Intelligence (BI) et d'Intelligence Artificielle (IA) a été explorée pour leur capacité à transformer les données en informations exploitables. Enfin, les méthodologies ETL et CRISP-

## **Chapitre II : État de l'art**

DM ont été présentées comme des outils indispensables pour un traitement et une analyse efficace des données, fournissant ainsi une base robuste pour la mise en œuvre de notre solution visant à améliorer la performance et la rentabilité des agences de la Société Générale Algérie.

---

**Troisième Chapitre**

**Mesure de la Performance des  
agences bancaires par la Business  
Intelligence**

---

## Introduction :

Ce chapitre de notre étude se concentre sur la Définition des KPIs clés, la conception de la base de données et la création de tableaux de bord interactifs sur Power BI. Nous suivons un processus structuré pour une analyse détaillée, précise et fiable de la performance des agences bancaires.

## 1 Analyse des Données et Création des Tableaux de Bord :

Afin de mesurer efficacement les performances des agences, nous avons suivi les étapes suivantes :

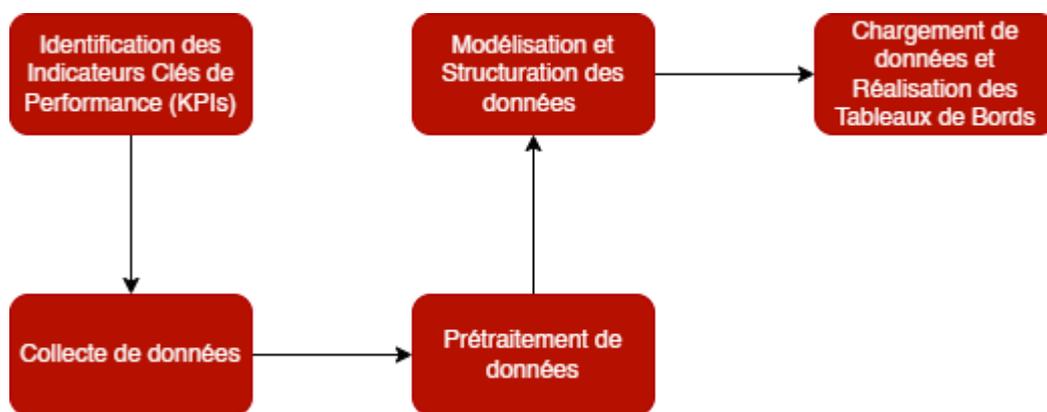


Figure 3.1 - Étapes de la mesure de performances

### 1.1 Identification des Indicateurs Clés de Performance (KPI)

La première étape de notre démarche consiste à identifier les Indicateurs Clés de Performance (KPIs) essentiels pour mesurer la performance des agences bancaires. Cette phase est importante pour garantir une évaluation précise et pertinente.

Pour déterminer les variables clés nécessaires à cette mesure, nous avons entrepris les actions suivantes :

- **Recherche Documentaire** : Nous avons effectué une recherche approfondie sur les bonnes pratiques et les standards de l'industrie bancaire en matière de mesure de performance. Cette recherche a inclus la consultation de publications académiques, de rapports financiers et de benchmarks internationaux.
- **Réunions avec les Experts métiers de la Banque** : Nous avons organisé plusieurs réunions avec les experts internes de la Société Générale Algérie. Ces experts comprennent des contrôleurs de gestion et des responsables financiers. Leurs connaissances approfondies des opérations quotidiennes et des objectifs stratégiques de la banque ont été inestimable pour identifier les KPIs les plus pertinents.

À l'issue de ces recherches et consultations, nous avons identifié 4 grandes familles de KPIs jugés essentiels pour la mesure de la performance des agences bancaires : le Produit Net Bancaire, les frais généraux, l'encours de crédits et l'encours de dépôts.

Le tableau suivant présente en détail les différents KPIs sélectionnés pour chaque famille, illustrant ainsi les aspects spécifiques de la performance des agences bancaires que nous avons jugés cruciaux pour une évaluation complète et précise.

Produit Net Bancaire	Total Marge d'intermédiation	Marge d'intermédiation
		Rémunération interne des ressources
		Facturation interne des emplois
		Marge Leasing
		Marge Coverage
	Total Commissions	Commerce extérieur
		Changes Manuel
		Exploitation Banque
		Autres Financements
		Commissions sur caution domestique
		Réémissions des garanties internationales
		Services, monétiques et banque
	Total APNE	Commissions Coverage
		Autres Produits d'Exploitation
Autres Produits d'Exploitation ventilés		
Frais Généraux	Frais Personnels	APNE Coverage
		Appointements Bruts
		Congés Payés
		Bonus
		Autres Avantages à Court terme
		Taxes sur FPA
		Charges sociales et fiscales taux
	Autres Charges d'Exploitation	Primes variables et commissions
		Provisions pour risques et charges
		Recours à l'extérieur
		Taxes et impositions
		Affranchissement
		Fournitures et consommables
		Frais Informatiques
Frais de Communications		
Frais de missions		
Frais Immobiliers		
Frais Télécommunication		
Autres Frais Divers		
Amortissement	Amortissements	
	Crédit de Consommations	

Encours de Crédits	Encours Moyens	Crédit Moyen Terme (CMT) Crédit Immobiliers
	Encours Fin de Périodes	Crédit Spot Crédit Leasing Crédit Découvert Financement de créances
Encours de Dépôts	Encours Moyens	Dépôts à Vue Dépôt à Terme
	Encours Fin de Périodes	Dépôt de Garantie Epargne Dépôts Devise

Tableau 3.1 - Les KPI sélectionnés pour la mesure de performances

## 1.2 Collecte de Données

Après l'identification des KPIs, la deuxième étape consiste à collecter les données nécessaires pour mesurer ces indicateurs. L'objectif est de réaliser une analyse couvrant la période de janvier 2018 à décembre 2023. Ainsi, nous avons entrepris la collecte de chaque KPI sur cette période afin d'obtenir une vision exhaustive et complète de la performance des agences bancaires sur toute la durée étudiée. Cette collecte de données s'est appuyée sur plusieurs sources internes de la banque incluant :

- **Le fichier Compte de résultat mensuel par Agence**

Compte de résultat mensuel (CRM) par agence contient les résultats globaux cumulés des agences en termes de performances. Ces données, calculées chaque mois par les contrôleurs de gestion, sont stockées dans un fichier Excel pour suivre les résultats globaux de chaque agence bancaire. Les principaux indicateurs suivis incluent le Produit Net Bancaire (PNB), les Frais Généraux (FG), le résultat d'exploitation, et d'autres indicateurs clés de performance.

- **Le fichier Résultat Agences**

Le Résultat Agences est un fichier Excel utilisé comme document de travail par les contrôleurs de gestion pour construire le CRM par agences. Ce fichier est créé mensuellement et contient des informations détaillées en cumulé, mais avec une granularité plus élevée en termes de KPIs par rapport au CRM par agence. Ce fichier est une ressource précieuse pour évaluer les performances individuelles de chaque agence bancaire. Il fournit une vue détaillée des résultats financiers, des revenus, des dépenses, des profits et des pertes pour chaque agence.

- **Le fichier Chapitre 6 et 7**

Le fichier Chapitre 6 et 7 est le document le plus détaillé en termes de KPIs par agence. Il contient les détails des sous-familles de chaque grande famille de KPI, avec une colonne "chapitre" définie comme clé qui facilite la recherche de chaque indicateur.

À partir de ce fichier, nous avons pu extraire les détails des commissions pour chaque agence, ainsi que les détails des frais généraux. Ce fichier offre une vue approfondie des performances des agences bancaires, en fournissant des informations détaillées sur les différents aspects de leurs activités. Il permet d'analyser en profondeur les revenus et les dépenses de chaque agence,

en décomposant les KPIs principaux en sous-catégories spécifiques. Par exemple, pour les commissions, le fichier fournit des détails sur les types de commissions générées par agence, comme les commissions sur caution domestiques, les changes manuels, etc. De même, pour les frais généraux, le fichier détaille les différents frais personnels, les amortissements et autres.

- **Fichier PNB par Clients GE PME :**

Ce fichier, construit trimestriellement, contient les résultats en termes de Produit Net Bancaire de chaque client, qu'il soit une Petite et Moyenne Entreprise ou une grande entreprise. Les informations incluent le radical du client, le nom du client et le PNB généré par ce client.

- **Rattachement Client-Agence :**

Ce fichier contient l'appartenance de chaque client à l'agence qui lui correspond. Les informations incluent le radical du client et l'ID de l'agence à laquelle il est rattaché.

- **Fichier PNB par marché :**

Le fichier PNB par marché est construit trimestriellement et détaille le Produit Net Bancaire de chaque marché (CLIPRI, CLIPRO, GE, PME) par agence.

- **Fichier Encours Moyens Crédit :**

Le fichier Encours Moyens Crédit est construit mensuellement et offre une vue détaillée des encours moyens de crédits pour chaque agence. Il comprend les crédits de consommation, les crédits immobiliers, les crédits à moyen terme, et d'autres types de crédits.

- **Fichier Encours Moyens Dépôts :**

Le fichier Encours Moyens Dépôts est construit mensuellement et présente en détail les encours moyens de dépôts pour chaque agence. Il inclut les dépôts à vue, les dépôts à terme, les dépôts de garantie, ainsi que d'autres types de dépôts.

- **Fichier Encours fin de période Crédit :**

Le fichier Encours fin de période des Crédits est construit mensuellement et offre une vue détaillée des encours fin de période de crédits pour chaque agence. Il comprend les crédits de consommation, les crédits immobiliers, les crédits à moyen terme, et d'autres types de crédits.

- **Fichier Encours fin de période Dépôts :**

Le fichier Encours fin de période des Dépôts est construit mensuellement et présente en détail les encours fin de période de dépôts pour chaque agence. Il inclut les dépôts à vue, les dépôts à terme, les dépôts de garantie, ainsi que d'autres types de dépôts.

### 1.3 Prétraitement des Données :

Le prétraitement des données est une étape cruciale dans le processus d'analyse des données, visant à nettoyer, transformer et préparer les données pour une utilisation optimale dans les analyses ultérieures. Cette étape essentielle est fondamentale pour garantir la qualité, la cohérence et la fiabilité des données, et joue un rôle clé dans la création de modèles prédictifs précis et dans la prise de décisions éclairées.

- **Uniformisation des Données : Harmonisation et filtrage des Tableaux Excel**

La première étape du prétraitement des données consiste à rendre uniformes les différentes familles de tableaux Excel. Étant donné que nous avons utilisé des données

historiques de la banque couvrant la période de janvier 2018 à décembre 2023, la construction des tableaux Excel s'est améliorée au fil des années.

Cela nous a conduits à élaborer un modèle fixe de tableaux Excel de 2018 à 2023, dans le but de faciliter la lecture et la manipulation des données. Pour ce faire, nous avons utilisé le langage de programmation Python, en utilisant les bibliothèques Numpy et Pandas.

Cette approche nous a permis de standardiser la structure et le format de nos données, ce qui facilite grandement leur traitement et leur analyse ultérieure.

- **Gestion des Valeurs Manquantes : Préserver l'Intégrité des Données**

La deuxième étape du prétraitement des données implique la gestion des valeurs manquantes. Au cours de notre collecte de données, nous avons remarqué l'absence de certaines tables du CRM par agence, du Résultat Agence, ainsi que des chapitres 6 et 7. Cette absence s'est particulièrement manifestée dans quelques mois, des perturbations opérationnelles ont pu entraîner ces lacunes dans la disponibilité des données.

Face à cette lacune, nous avons pris l'initiative de construire ces tableaux manquants en utilisant la moyenne des valeurs des tables adjacentes. Cette approche nous a permis de préserver l'intégrité de nos données en comblant les lacunes et en garantissant une cohérence dans notre ensemble de données.

- **Standardisation des Unités et Regroupement des KPIs par Agence**

La troisième étape du prétraitement des données est la transformation des résultats pour les uniformiser en une seule unité, à savoir les milliards de DZD (Dinars algériens). Ensuite, chaque KPI est regroupé par agence afin d'obtenir les résultats agrégés pour chaque agence.

Pour réaliser cette tâche, nous avons utilisé la bibliothèque pandas de Python. Cette approche nous a permis de normaliser les données, facilitant ainsi leur comparaison et leur analyse ultérieure. En regroupant les KPIs par agence, nous avons obtenu une vue d'ensemble de la performance de chaque agence, ce qui nous a aidés à identifier les tendances et les modèles significatifs dans les données.

En outre, afin de compléter cette agrégation, nous avons réalisé plusieurs jointures entre les tables. Par exemple, nous avons effectué une jointure entre les tables de PNB par client GE PME et la table de Rattachement client-agence. Cette démarche nous a permis d'obtenir le PNB généré par chaque client dans chaque agence, renforçant ainsi la précision et la pertinence de nos données agrégées.

- **Calcul des Activités Mensuelles Réelles par Agence**

La prochaine étape consiste à effectuer des soustractions entre les tables de chaque KPI de la même année afin d'obtenir l'activité mensuelle réelle pour chaque agence. Étant donné que les données ont été initialement collectées de manière cumulée, une étape essentielle était de décomposer ces données pour obtenir des valeurs uniques pour chaque mois.

Par exemple, le CRM par agence de mars contient les valeurs des KPIs cumulées de janvier à mars. Pour obtenir l'activité unique du mois de mars, nous devons soustraire la table du CRM par agence de mars de la table du CRM par agence de février.

Cette approche a été utilisée pour chaque KPI afin d'obtenir les activités réelles mensuelles. Ce processus nous a permis d'obtenir des données plus granulaires et précises, facilitant ainsi une analyse plus approfondie de la performance mensuelle de chaque agence.

- **Agrégation des Données dans un Tableau Excel Unique**

La dernière étape du processus implique l'agrégation de chaque KPI dans un seul tableau Excel contenant les données de janvier 2018 à décembre 2023 pour chaque agence. Chaque KPI a été représenté dans une colonne distincte, avec une colonne supplémentaire pour la date afin de différencier les résultats par dates.

Cette consolidation des données dans un seul tableau Excel nous a permis de disposer d'un ensemble de données complet et structuré, facilitant ainsi une analyse approfondie de la performance des agences bancaires sur toute la période étudiée. La présence d'une colonne de date nous a également permis d'effectuer des analyses temporelles et de détecter les tendances et les modèles évolutifs au fil du temps.

## 1.4 Modélisation et Structuration des Données :

Afin de tirer le meilleur parti de la base de données que nous avons construite, nous avons envisagé d'utiliser Power BI pour créer des tableaux de bord interactifs contenant des indicateurs de performance pour chaque agence. L'objectif est de permettre un suivi précis et dynamique de l'évolution de chaque agence au fil du temps. Grâce à Power BI, nous pourrions intégrer les données de manière efficace et créer des visualisations dynamiques qui permettront aux utilisateurs de suivre facilement les performances de chaque agence. Afin de réaliser cet objectif nous avons passé par plusieurs étapes :

### 1.4.1 Dictionnaire de données et graphe des dépendances fonctionnelles :

Nous avons élaboré un dictionnaire de données qui répertorie tous les champs et les entités présents dans la base de données. Ce dictionnaire fournit une référence claire pour comprendre les données et leurs significations. De plus, nous avons créé un graphique des dépendances fonctionnelles pour visualiser les relations entre les différentes entités et les dépendances entre les champs. Afin de réaliser notre objectif, nous proposons la création des entités suivantes avec le dictionnaire qui suit pour structurer les données de manière plus appropriée

La table 'Agence' :

Attribut	Signification	Domaine
ID Agence	Identifiant de l'agence	Entier
Nom Agence	Nom de l'agence	Chaîne (30)
Adresse	Adresse de l'agence	Chaîne (50)
Type agence	Indique le type d'agence	Chaîne (15)
Unité commerciale	Indique l'UC de l'agence	Chaîne (30)

Tableau 3.2 - La table 'Agence'

La table 'Marché' :

Attribut	Signification	Domaine
ID Marché	Identifiant du marché	Entier
Marché	Type du marché	Chaîne (30)

Tableau 3.3 - La table 'Marché'

La table 'Client' :

Attribut	Signification	Domaine
Radical Client	Identifiant du client	Entier
Nom	Nom du client	Chaîne (30)
Type	Type du client	Chaîne (30)

Tableau 3.4 - La table 'Client'

La table 'Date' :

Attribut	Signification	Domaine
ID Date	Identifiant de la date	Entier
Date	La date	Date

Tableau 3.5 - La table 'Date'

La table 'PNB' :

Attribut	Signification	Domaine
ID PNB	Identifiant du PNB	Entier
Famille PNB	Famille du PNB	Chaîne (15)
Montant	Montant du PNB	Réel

Tableau 3.6 - La table 'PNB'

La table 'Frais Généraux' :

Attribut	Signification	Domaine
ID FG	Identifiant des FG	Entier
Famille FG	Famille des FG	Chaîne (15)
Montant	Montant des FG	Réel

Tableau 3.7 - La table 'Frais Généraux'

La table 'Encours de Crédits' :

Attribut	Signification	Domaine
ID EC	Identifiant des EC	Entier
Nature EC	Indique la nature des encours	Chaîne (15)
Produit	Type du crédit	Chaîne (15)
Montant	Montant des crédits	Réel

Tableau 3.8 - La table 'Encours de Crédits'

La table 'Encours de Dépôts :

Attribut	Signification	Domaine
ID ED	Identifiant des ED	Entier
Nature ED	Indique la nature des encours	Chaine (15)
Type de dépôts	Type de dépôts	Chaine (15)
Montant	Montant des dépôts	Réel

Tableau 3.9 - La table 'Encours de Dépôts'

La table 'Résultat Net Comptable :

Attribut	Signification	Domaine
ID RNC	Identifiant du RNC	Entier
CNR	Cout Net du Risque	Réel
Prov FRBG	Provisions Fonds pour Risque Bancaire Généraux	Réel
Porv CMT	Provisions Crédit Moyen Terme	Réel
RBE	Résultat Brut d'Exploitation	Réel
RE	Résultat d'Exploitation	Réel

Tableau 3.10 - La table 'Résultat Net Comptable'

Après avoir choisi les entités et défini le dictionnaire des données, voici le **graphe des dépendances fonctionnelles** :

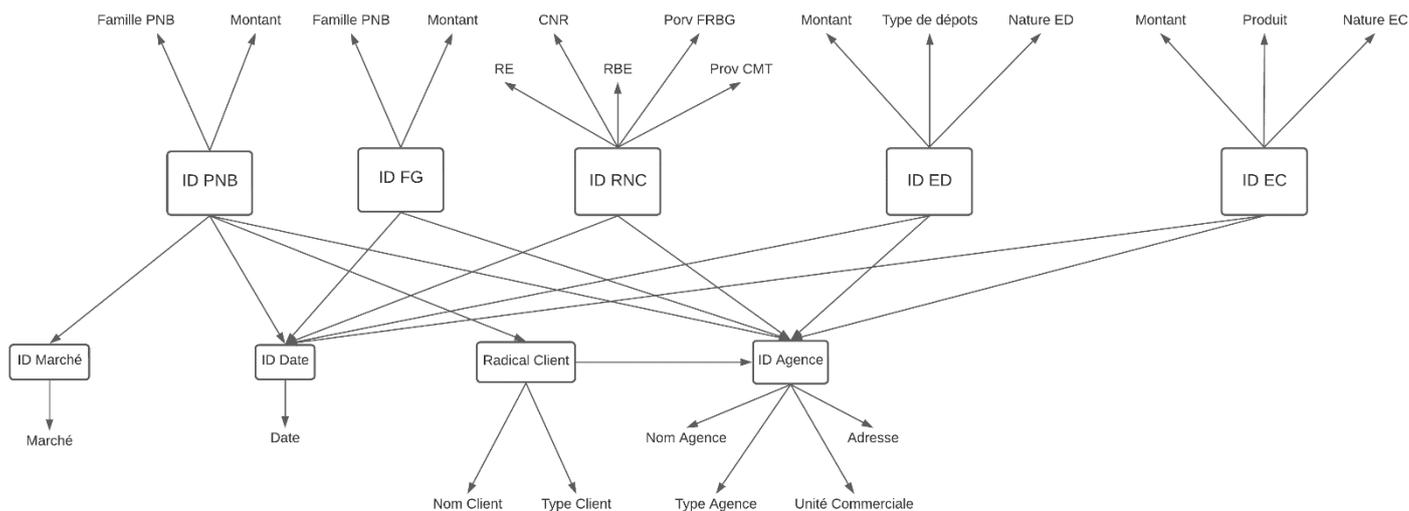


Figure 3.2 - Graphe des dépendances Fonctionnelles

### 1.4.2 Modèle conceptuel (E/A) :

Le modèle conceptuel représente la structure logique de la base de données en utilisant des concepts tels que les classes, les associations et les cardinalités

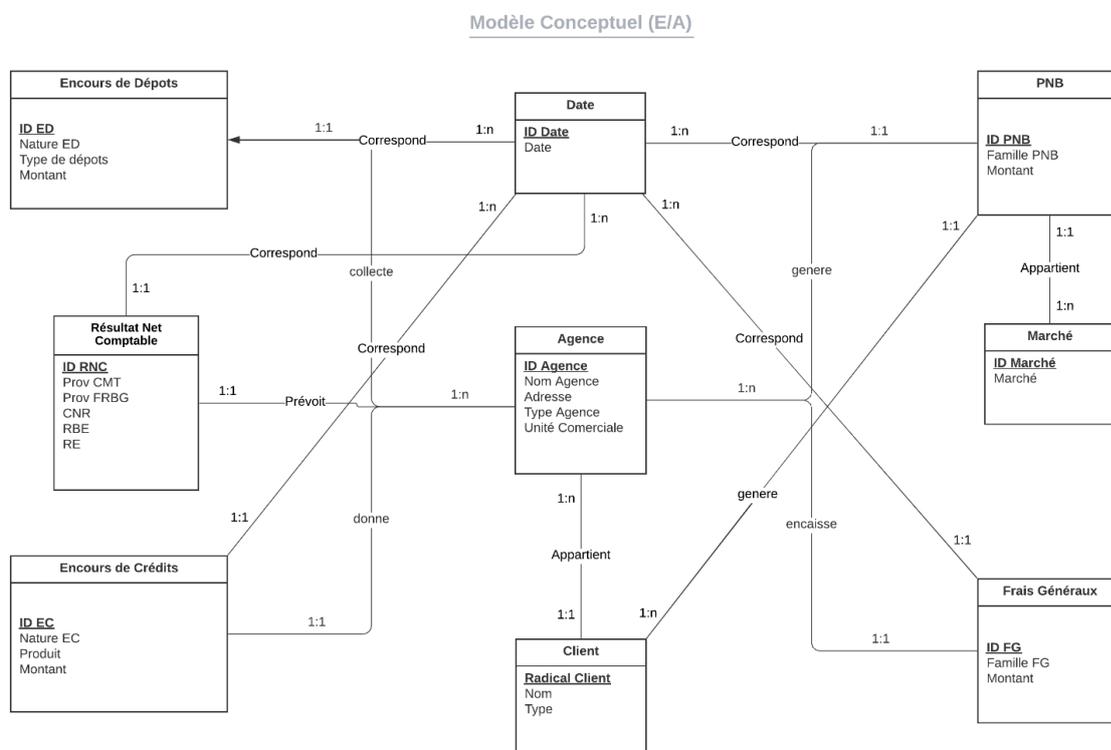


Figure 3.3 - Modèle Conceptuel (Entité-Association)

Le modèle de données adopté est celui de l'architecture en étoile, une approche bien établie dans la modélisation des bases de données pour l'analyse et le reporting. Cette architecture se compose de tables de faits et de tables de dimensions, offrant une structure optimisée pour la manipulation et l'interprétation des données.

Caractéristiques principales :

- Tables de Faits** : Les tables de faits centralisent les mesures quantitatives clés de la banque, le Produit Net Bancaire (PNB), les Frais Généraux, l'Encours de Crédit, l'Encours de Dépôts et le Résultat Net Comptable. Elles sont conçues pour contenir des données détaillées avec des clés étrangères pointant vers les tables de dimensions pertinentes.
- Tables de Dimensions** : Les tables de dimensions fournissent des informations descriptives contextuelles pour enrichir les mesures contenues dans les tables de faits. Les tables de dimensions sont : Agence, Date, Client et Marché

**Avantages du Modèle en Étoile :**

- Simplicité** : L'architecture en étoile offre une structure simple et intuitive, ce qui la rend facile à comprendre et à utiliser pour les analystes et les utilisateurs finaux.

2. **Performance** : En minimisant le nombre de jointures nécessaires entre les tables, ce modèle optimise les performances des requêtes d'analyse, permettant des temps de réponse plus rapides.
3. **Flexibilité** : La conception modulaire de l'architecture en étoile permet l'ajout aisé de nouvelles dimensions ou de nouvelles mesures sans perturber l'ensemble du modèle de données, offrant ainsi une grande flexibilité pour s'adapter aux besoins évolutifs de l'entreprise.

### 1.4.3 Modèle relationnel :

Dans cette section, nous traduisons le modèle conceptuel en un modèle relationnel en utilisant les concepts de tables, d'attributs et de clés primaires et étrangères.

Agence (**ID Agence**, Nom Agence, Adresse, Type Agence, Unité Commerciale)

Client (**Radical Client**, Nom, Type, #ID Agence)

Date (**ID Date**, Date)

Marché (**ID Marché**, Marché)

PNB (**ID PNB**, Famille PNB, Montant, #ID Agence, #ID Client, #ID Date, #ID Marché)

Frais Généraux (**ID FG**, Famille FG, Montant, #ID Agence, #ID Date)

Encours de Crédits (**ID EC**, Nature EC, Produit, Montant, #ID Agence, #ID Date)

Encours de Dépôts (**ID ED**, Nature ED, Type de Dépôts, Montant, #ID Agence, #ID Date)

Résultat Net Comptable (**ID RNC**, CNR, Prov CMT, Prov FRBG, RBE, RE, #ID Agence, #ID Date)

(# : Clé étrangère)

### 1.5 Chargement de données et Réalisation des Tableaux de Bord :

Afin de mieux évaluer les performances et de faciliter la prise de décisions, nous avons chargé les données prétraitées, modélisées, transformées et structurées en tableaux de bord pour chaque variable et sous-variable. Ces tableaux de bord intègrent diverses visualisations, graphiques et filtres répondant aux demandes des décideurs financiers, permettant ainsi de mieux adresser les problématiques identifiées. En tout, nous avons conçu un tableau de bord d'accueil, seize autres pour les différentes variables et sous-variables, ainsi qu'un tableau de bord général pour la comparaison entre les agences. Ces outils fournissent des informations clés et des insights pertinents, soutenant ainsi la prise de décision et la gestion efficace des activités de la banque. Grâce à ces tableaux de bord, les décideurs peuvent accéder rapidement à des données essentielles, analyser les performances de manière détaillée, et prendre des mesures adaptées pour améliorer l'efficacité opérationnelle et la rentabilité des agences bancaires.

De plus, nous avons développé une interface utilisateur permettant d'accéder à la solution. Cette interface offre une expérience utilisateur intuitive et efficace. Les détails concernant cette interface seront abordés en profondeur dans la section "Application Utilisateur".

*Pour mieux comprendre notre méthodologie, nous présenterons une étude de cas portant sur la réalisation et la conception d'un tableau de bord pour la variable 'Produit Net Bancaire' (PNB) et ses sous-variables. Comme indiqué dans le Tableau 3.1.1, le PNB se compose de la marge d'intermédiation, des commissions et des autres produits d'exploitation.*

*Le processus de conception et de réalisation de ce tableau de bord est identique pour toutes les autres variables.*

## 2 Étude de Cas : Réalisation du Tableau de Bord du Produit Net Bancaire :

Dans cette section, nous allons illustrer et détailler la réalisation d'un des tableaux de bord que nous avons créés. Prenons l'exemple du Produit Net Bancaire (PNB). Nous avons suivi plusieurs étapes de traitement de données, d'analyse et de conception. Les sections suivantes détaillent le processus :

### Importation des Bibliothèques

L'importation des bibliothèques nécessaires est une étape primaire pour la manipulation et l'analyse des données. Nous utilisons principalement les bibliothèques pandas pour la gestion des DataFrames, numpy pour les opérations numériques et matplotlib pour la visualisation des données pour toutes les variables.

### Collecte et Transformation des Fichiers Mensuels cumulés

Nous avons collecté les fichiers de résultats des agences pour chaque mois de l'année en cumulés et les transformé en Data-Frames. Cela nous a permis de structurer les données pour une analyse ultérieure.

```
[67]: import pandas as pd
import numpy as np

[68]: chemin_base = r"C:\Users\bn-asuserz\Desktop\Chapitres 6 & 7 2022\Chapitres 6 & 7 retraités "
Chapitre6_7 = {}

for mois in [1,2,3,4,5,6,9,10,12]:

    nom_fichier = f"{chemin_base}{mois:02d}-2022.xlsx"
    sheet_name = 'WF_DetailChap67 '

    try:
        data = pd.read_excel(nom_fichier, sheet_name = sheet_name)
        Chapitre6_7[f"{mois:02d}"] = data

    except FileNotFoundError:
        print(nom_fichier)
```

Figure 3.4 - Collecte et importation des données

### Split des Chapitres pour Obtenir les Composants de chaque sous variable

Nous avons séparé les chapitres pour obtenir les différentes composantes de la marge d'intermédiation, Commissions et Autres produits d'exploitations. Chaque composante est extraite et analysée individuellement pour une meilleure compréhension.

```
[69]: for mois, data in Chapitre6_7.items():

    data.columns = data.iloc[0]
    data = data[1:]
    data = data[data['Nature'] == 'PNB2']

    data = data.iloc[:, :-6]
    data = data[data['Age'] <= 853]

    Chapitre6_7[mois] = data

[71]: def determiner_commissions(chapitres):
    if chapitres in [708200, 702916, 702900, 602700, 708310, 709250, 628300, 608330, 608390, 708300, 608320]:
        return 'Services, monétiques et banque'
    elif chapitres in [702902, 702903, 704361, 702951, 608600, 707000, 707610, 760119]:
        return 'Autres Financement'
    elif chapitres in [702908, 702991, 702990, 760500, 702906, 702920, 702907, 701900, 702909, 709200, 702975, 703900, 601900, 760101, 760200]:
        return 'Exploitation banque'
    elif chapitres in [702930, 702912, 601910]:
        return 'changes Manuel'
    elif chapitres in [760100, 702913, 702910, 702915, 608300, 609100, 702970, 601980, 702911, 702914]:
        return 'Commerce extérieur'
    elif chapitres in [707120, 601920, 608310]:
        return 'Commissions sur caution domest'
    elif chapitres in [707100, 607100]:
        return 'Réémissions de garantie intern'
    else:
        return 'Exploitation banque'
```

Figure 3.5 - Split des Chapitres

### Prétraitement des Valeurs Manquantes

Nous avons géré les valeurs manquantes de différentes manières, selon les directives des responsables, en tenant compte de la signification financière de chaque variable. Les valeurs manquantes ont été soit remplies avec la moyenne des valeurs adjacentes, soit supprimées. Cette approche a permis de maintenir la continuité des données sans introduire de biais.

```
[ ]: for mois, data in resultats_agences.items():
    resultats_agences[mois] = resultats_agences[mois].sort_values(by='Agences', ascending = True)
    resultats_agences[mois] = pd.DataFrame(resultats_agences[mois])
    resultats_agences[mois] = resultats_agences[mois].reset_index(drop = True)
    print(resultats_agences[mois].shape)

[220]: qualitatives = ['Agences']
    colonnes_quantitatives = [col for col in resultats_agences["2022_02"].select_dtypes(include = ['object', 'int64']).columns if col != 'Agences']
    print(colonnes_quantitatives)

["Marge d'Intermediation", "Rémunération interne des ressources", "Facturation interne des emplois", "Marge Leasing", "Marge Coverage", "Commissions PNB 2",
"Commissions leasing ventilées", "Autres commissions Ventilées", "Commission Coverage", "Autres Produits d'Exploitation PNB 3", "Autres Produits d'Exploitation Ventilé", "APNE Coverage"]

[221]: resultat = resultats_agences['2022_09'][colonnes_quantitatives] - resultats_agences['2022_06'][colonnes_quantitatives]
    resultats_agences['2022_08'] = resultat * (2/3) + resultats_agences['2022_06'][colonnes_quantitatives]
    resultats_agences['2022_08'] = pd.concat([resultats_agences['2022_09'][qualitatives], resultats_agences['2022_08']], axis = 1)

[222]: resultats_agences['2022_07'] = resultat / 3 + resultats_agences['2022_06'][colonnes_quantitatives]
    resultats_agences['2022_07'] = pd.concat([resultats_agences['2022_09'][qualitatives], resultats_agences['2022_07']], axis = 1)

[223]: resultat2 = resultats_agences['2022_12'][colonnes_quantitatives] - resultats_agences['2022_10'][colonnes_quantitatives]

[224]: resultats_agences['2022_11'] = resultat2 / 2 + resultats_agences['2022_10'][colonnes_quantitatives]
    resultats_agences['2022_11'] = pd.concat([resultats_agences['2022_12'][qualitatives], resultats_agences['2022_11']], axis = 1)
```

Figure 3.6 - Prétraitement des valeurs manquante

### Standardisation des Unités et Sélection des Colonnes

Nous avons standardisé les montants en les convertissant en milliards de DZD et sélectionnons les colonnes nécessaires. Cela permet une comparaison cohérente entre les différentes agences et périodes.

```
[213]: colonnes_a_conserver = ['Agences', "Marge d'Intermediation", "Rémunération interne des ressources",
                             "Facturation interne des emplois", "Marge Leasing", "Marge Coverage", 'Commissions PNB 2',
                             "Commissions leasing ventilées", "Autres commissions Ventilées", 'Commission Coverage',
                             "Autres Produits d'Exploitation PNB 3", "Autres Produits d'Exploitation Ventilé", 'APNE Coverage']

for mois, data in resultats_agences.items():
    colonnes_existantes = [col for col in colonnes_a_conserver if col in data.columns]
    data_filtré = data[colonnes_a_conserver]

    resultats_agences[mois] = data_filtré

[214]: for mois, data in resultats_agences.items():
        print(resultats_agences[mois].info())
```

Figure 3.7 - Sélection des colonnes

### GroupBy et Calcul des Variables Manquantes

Nous avons regroupé les données par agence et par variable, et calculer les montants. Cela a permis d'obtenir une vue agrégée des données pour chaque agence. Aussi, si une agence n'apparaît qu'en décembre et pas avant, nous l'ajoutons pour tous les mois précédents en lui attribuant un montant de 0 afin d'assurer que toutes les agences soient présentes dans la base de données pour chaque mois.

```
[72]: for mois, data in Chapitre6_7.items():
        data.loc[:, 'Commissions'] = data['Cha'].apply(determiner_commissions)
        Chapitre6_7[mois] = data

[74]: for mois, data in Chapitre6_7.items():
        data = data.groupby(['Age', 'Commissions'])['Sde'].sum().reset_index()
        Chapitre6_7[mois] = data

[75]: Agences_supprimees = [24,26, 100, 150, 200, 250, 260, 300, 350, 400, 450, 500, 505, 550, 600, 650, 700, 750, 800, 850 ]
for mois, data in Chapitre6_7.items():
    data = data[~data['Age'].isin(Agences_supprimees)]
    Chapitre6_7[mois] = data

[78]: agences_uniques = Chapitre6_7['2022_12']['Age'].unique()
commissions_uniques = Chapitre6_7['2022_12']['Commissions'].unique()

for mois, data in Chapitre6_7.items():

    missing_rows = []
    for age in agences_uniques:
        for commission in commissions_uniques:
            if not (( data['Age'] == age) & (data['Commissions'] == commission)).any():
                missing_rows.append({'Age':age, 'Commissions': commission, 'Sde': 0 })

    missing_df = pd.DataFrame(missing_rows)
    data = pd.concat([data, missing_df], ignore_index = True)
    data = data.sort_values(by=['Age', 'Commissions'])
    Chapitre6_7[mois] = data.reset_index(drop = True)
```

Figure 3.8 - Groupement des données et Calcul des Variables Manquantes

### Agrégation des Données en DataFrames Mensuels

Nous avons calculé les données cumulées et prétraitées de chaque variable par mois en faisant la différence des fichiers cumulés puis les regroupant de janvier à décembre et les transformant en fichiers Excel pour la conception de la base de données globale. Cela facilite la manipulation et l'analyse des données.

```
[226]: Résultat_agences = [ resultats_agences["2022_01"], resultats_agences["2022_02"], resultats_agences["2022_03"], resultats_agences["2022_04"],
                        resultats_agences["2022_05"],resultats_agences["2022_06"], resultats_agences["2022_07"], resultats_agences["2022_08"]
                        , resultats_agences["2022_09"], resultats_agences["2022_10"],
                        resultats_agences["2022_11"], resultats_agences["2022_12"]]

noms_des_mois = ['janvier', 'fevrier', 'mars', 'avril', 'mai', 'juin', 'juillet', 'aout', 'septembre', 'octobre', 'novembre', 'decembre']

#dictionnaire pour stocker les CRMs calculés de chaque mois
resultat_agences_par_mois = {}

for i, resultat_agences in enumerate(Résultat_agences) :
    if i == 0 :
        resultat_agences_par_mois [noms_des_mois[i]] = resultat_agences.copy()
    else :
        resultat_agences_par_mois[noms_des_mois[i]] = resultat_agences.copy()
        for column in colonnes_quantitatives :
            resultat_agences_par_mois[noms_des_mois[i]][column] = Résultat_agences[i][column] - Résultat_agences[i-1][column]

[227]: Résultat_agences_par_mois = { 'janvier' : resultat_agences_par_mois['janvier'],
                                   'fevrier': resultat_agences_par_mois['fevrier'],
                                   'mars' :resultat_agences_par_mois['mars'],
                                   'avril' :resultat_agences_par_mois['avril'],
                                   'mai':resultat_agences_par_mois['mai'],
                                   'juin': resultat_agences_par_mois['juin'],
                                   'juillet': resultat_agences_par_mois['juillet'],
                                   'aout': resultat_agences_par_mois['aout'],
                                   'septembre': resultat_agences_par_mois['septembre'],
                                   'octobre': resultat_agences_par_mois['octobre'],
                                   'novembre': resultat_agences_par_mois['novembre'],
                                   'decembre': resultat_agences_par_mois['decembre'] }

mois_noms = {
    'janvier' : '01',
    'fevrier' : '02',
    'mars' : '03',
    'avril' : '04',
    'mai' : '05',
    'juin' : '06',
    'juillet' : '07',
    'aout' : '08',
    'septembre' : '09',
    'octobre' : '10',
    'novembre' : '11',
    'decembre' : '12', }

for mois, dataframe in Résultat_agences_par_mois.items() :
    nom_fichier = f"RA-{mois_noms[mois]}-22.xlsx"
    print(nom_fichier)
    chemin_complet = fr'C:\Users\bn-asuserz\Desktop\Folder\PNB1\PNB1 22\{nom_fichier}'
    dataframe.to_excel(chemin_complet, index = False)
```

Figure 3.9 - Agrégation des Données en DataFrames Mensuels

### Conception des Tableaux de Bord sur Power BI

Les tableaux de bord que nous avons conçus offrent une visualisation interactive et dynamique des données, permettant une analyse approfondie des performances des agences bancaires. Voici une explication détaillée du premier tableau de bord et de son impact sur la mesure des performances.

#### Dashboard Produit Net Bancaire (PNB)

***PNB = Marge d'Intermédiation + Commissions + Autres Produits d'Exploitation***

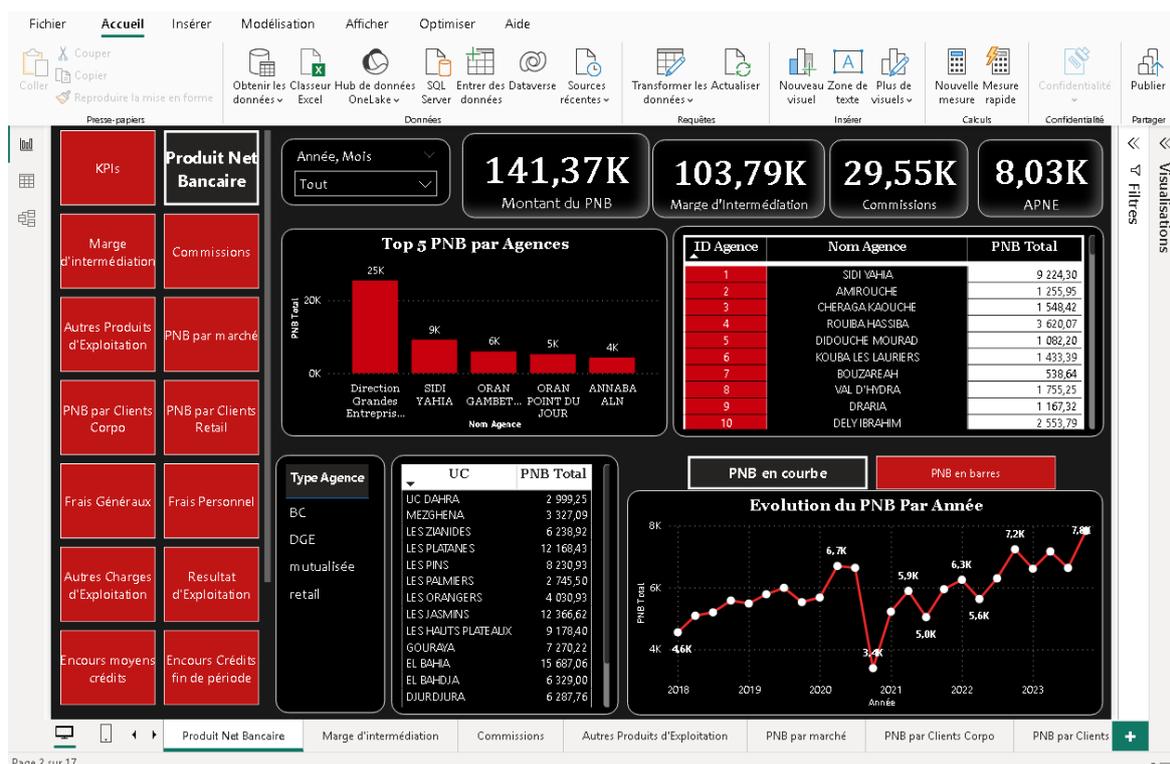


Figure 3.10 - Dashboard du Produit Net Bancaire

Le tableau de bord "Produit Net Bancaire" mesure et visualise les performances globales des agences en termes de PNB. Les éléments clés incluent :

- **Montant du PNB** : Affiche le montant total du Produit Net Bancaire pour la période sélectionnée.
- **Marge d'Intermédiation** : Affiche le montant total de la marge d'intermédiation pour la période sélectionnée.
- **Commissions** : Indique les revenus supplémentaires générés par les services fournis.
- **Autres Produits d'Exploitation** : Fournit une vue complète des autres différentes sources de revenus.

### Fonctionnalités Dynamiques :

- **Filtre Année, Mois** : Sélection des données de janvier 2018 à décembre 2023.
- **Diagramme des TOP 5 PNB par Agence** : Visualise les cinq agences ayant généré le plus de PNB.
- **Filtre Type d'Agence** : Permet une analyse segmentée selon le type d'agence.
- **Tableau des Unités Commerciales par PNB Total** : Récapitule les unités centrales par région.
- **Tableau Contenant ID Agence, Nom Agence, PNB Total** : Vue détaillée de chaque agence.

- **Visuel PNB en Courbe et en Barres** : Montre l'évolution du PNB.

**Impact du Tableau de Bord :**

- **Analyse en Temps Réel** : Filtres interactifs facilitant la prise de décision rapide.
- **Visualisation Intuitive** : Aide à identifier facilement les tendances et écarts de performance.
- **Segmentation des Données** : Analyse plus fine et pertinente par type d'agence et période.
- **Identification des Performances** : Classements et diagrammes pour repérer les meilleures et moins bonnes performances.
- **Historique et Tendances** : Analyse des tendances et performance historiques.

**Dashboard Commissions :**

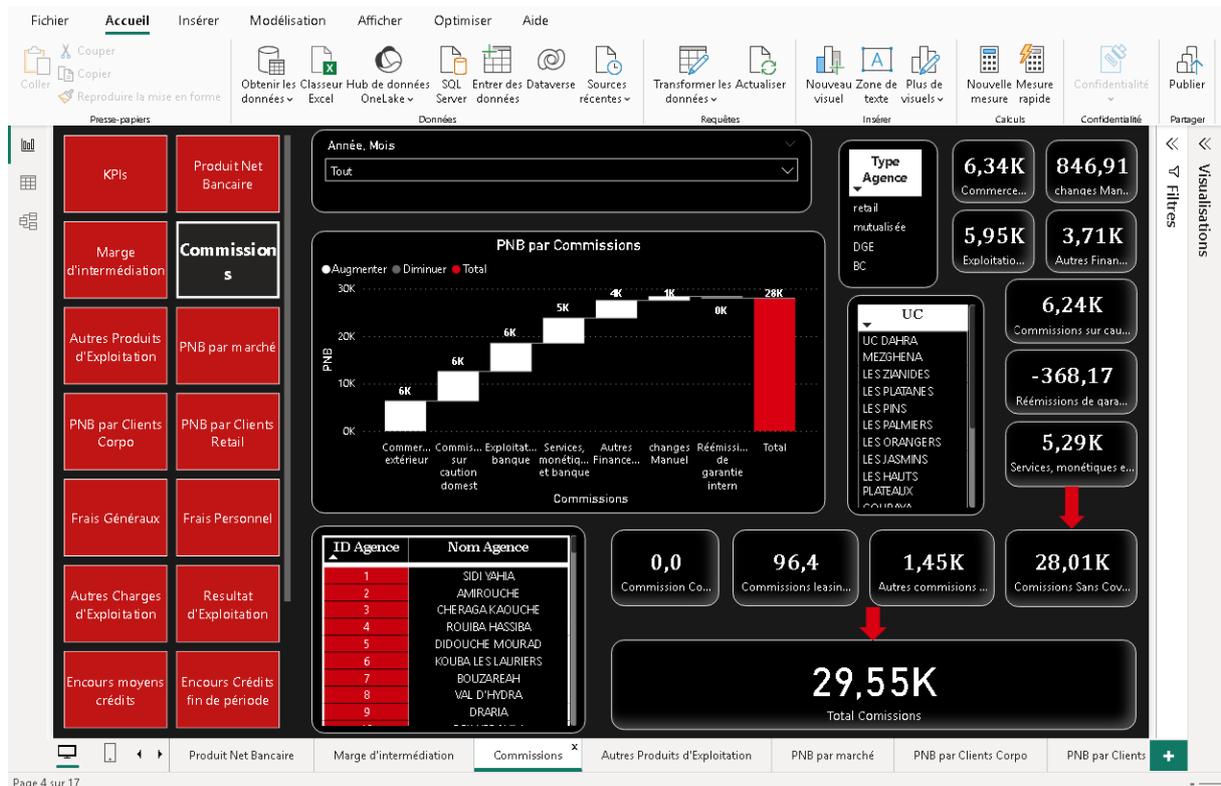


Figure 3.11 - Dashboard des Commissions

**Dashboard Marge d'Intermédiation :**

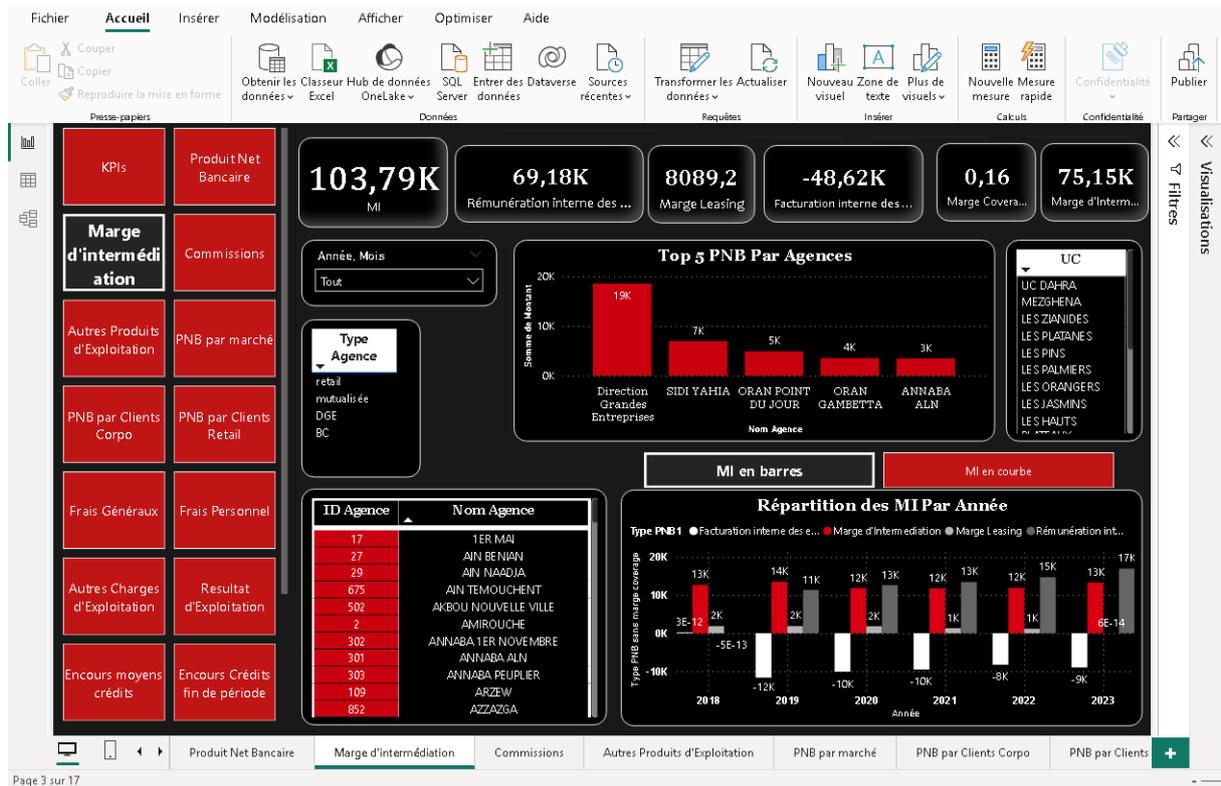


Figure 3.12 - Dashboard de la Marge d'intermédiation

**Dashboard Autres Produits d'Exploitation**

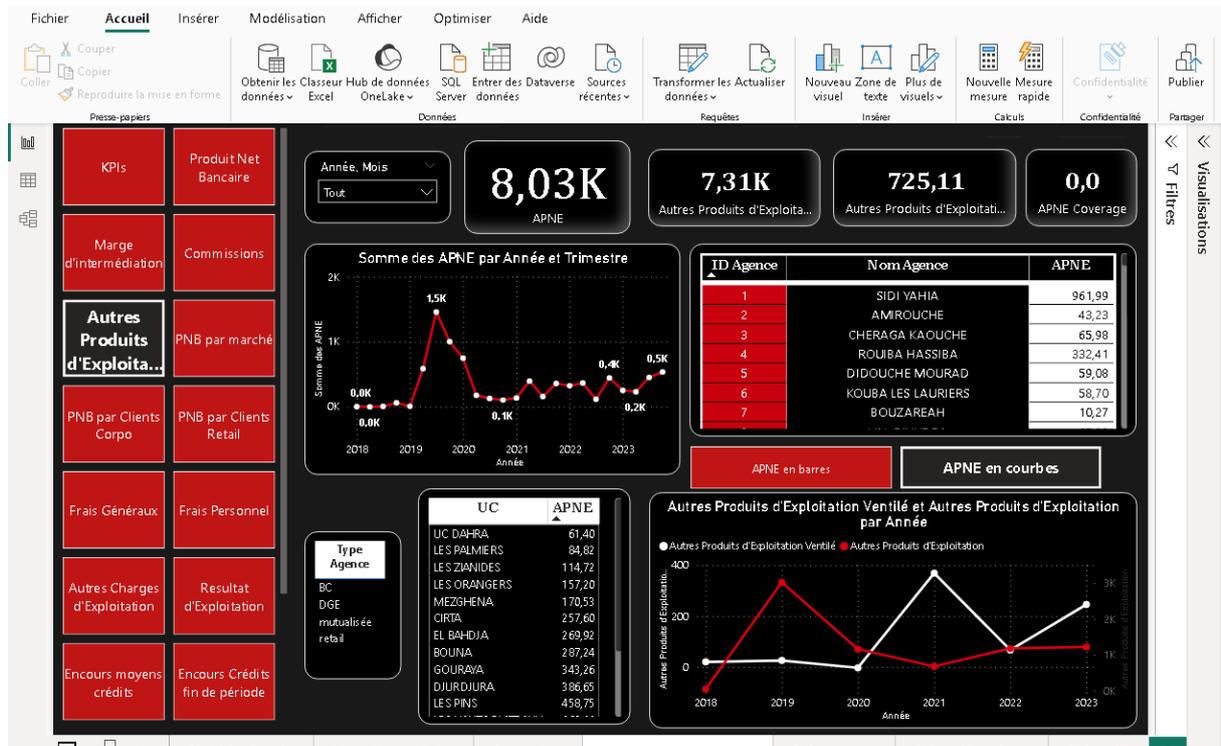


Figure 3.13 - Dashboard des Autres Produits d'Exploitation

## **Conclusion :**

En conclusion, ces tableaux de bord dynamiques offrent une vue complète et des fonctionnalités interactives pour une analyse approfondie et une prise de décision efficace. Les 17 tableaux de bord conçus couvrent une gamme exhaustive de variables financières, chacun étant adapté aux besoins spécifiques des financiers pour une gestion optimale des performances.

---

## **Quatrième Chapitre**

# **Prédiction de la Rentabilité des Agences Bancaires par l'Intelligence Artificielle**

---

## Introduction :

Ce chapitre de notre étude se concentre sur la prédiction et la modélisation des données pour anticiper la rentabilité des agences bancaires. Nous suivons une méthodologie structurée basée sur le processus CRISP-DM pour obtenir des prévisions précises et fiables. Nous utilisons les données historiques, nettoyées et prétraitées dans la première partie pour effectuer les visualisations. Ces données passeront par plusieurs étapes supplémentaires afin d'être correctement préparées pour les modèles de prévision que nous avons identifiés. Afin de prédire efficacement la rentabilité des agences, nous avons suivi les étapes suivantes

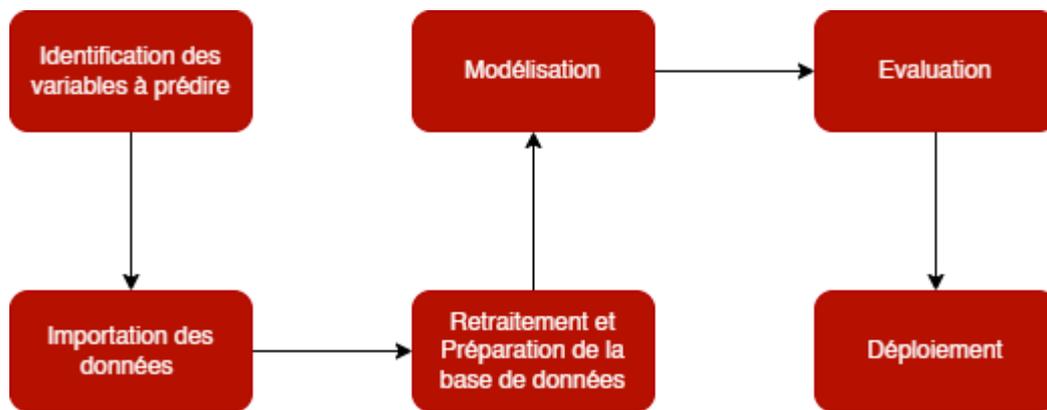


Figure 4.1 - Les étapes de la prévision de la rentabilité

## 1 Analyse, Modélisation et Évaluation des Prévisions Financières Approche Complète

Dans cette section, nous allons explorer en détail notre méthodologie complète pour l'analyse, la modélisation et l'évaluation des prévisions financières des agences bancaires. Notre approche se divise en plusieurs étapes clés, chacune essentielle pour atteindre des prédictions précises et exploitables

### 1.1 Identification des variables à prédire :

Nous avons identifié plusieurs variables à prédire pour chaque agence, qui sont :

- **Le Produit Net Bancaire (PNB)** : les commissions, les autres produits et charges d'exploitation (APNE) et La Marge d'Intermédiation qu'on calculera à partir des prévisions de crédits et de dépôts.
- **Les Frais Généraux (FG)** : les frais personnels, les amortissements, et les autres charges d'exploitation.
- **Les dépôts** : Dépôt à Vue, Dépôt à Terme, Dépôt de Garantie, Epargne et Dépôt Devise

- **Les Crédits** : Crédit Moyen Terme, Crédit Immobilier, Crédit Consommation, Financement de Créance, Crédit Spot, Crédit Découvert, Leasing

Nous avons construit plusieurs fichiers Python (.ipynb), chaque fichier étant dédié à une grande famille de KPI, par exemple (PNB, FG). Cela nous a permis de structurer notre travail de manière appropriée et de réaliser des prédictions spécifiques pour chaque variable.

### 1.2 Importation des données :

Une fois les variables à prédire identifiées, nous avons importé les données correspondantes de la base de données utilisée pour les tableaux de bord précédents. Ces données ont ensuite été retraitées et préparées pour les modèles de prédiction, qu'il s'agisse de modèles statistiques ou de réseaux de neurones.

### 1.3 Retraitement et Préparation de données :

Pour mieux préparer et retraiter les données, nous avons appliqué diverses méthodes statistiques et d'analyse de données afin de comprendre et d'observer le comportement de nos séries temporelles :

- Suppression des nouvelles agences sans historique.
- Regroupement des chapitres de chaque variable et sous-variable.
- Création des tables pour réorganiser nos données par agence et par date, transformant ainsi chaque agence par variable en une série temporelle de 72 instances (de janvier 2018 à décembre 2023).
- Séparation des données des agences dans un dictionnaire de DataFrames, où chaque clé représente l'ID de l'agence et les valeurs correspondent aux différentes variables, chacune dans une colonne.
- Création des fonctions pour vérifier la stationnarité de nos séries temporelles à l'aide du test ADF (Augmented Dickey-Fuller test).
- Pour les séries non stationnaires, nous avons appliqué des méthodes de transformation telles que la différenciation, puis stocké ces séries stationnaires dans un dictionnaire dédié.

### 1.4 Modélisation :

Pour la partie modélisation, Nous avons :

- Identifié dans la section État de l'art, les modèles adéquats pour les prévisions de nos séries temporelles.
- Séparé nos données en ensembles d'entraînement (de 2018 à 2022) et de test (pour l'année 2023) pour chaque variable et pour chaque agence individuellement.
- Entraîné nos modèles et optimisé leur apprentissage automatique.

## 1.5 Évaluation :

Pour la partie évaluation, nous avons mesuré les performances de nos modèles de prédiction à l'aide de métriques spécifiques (MAE, MSE et RMSE).

### Méthodologie d'Évaluation des Erreurs de Prédiction

Pour sélectionner le meilleur modèle de prédiction pour chaque agence, nous avons suivi une méthodologie rigoureuse. Voici les étapes détaillées de notre processus :

#### Stockage des Erreurs des Modèles :

- Nous avons appliqué plusieurs modèles de prédiction aux données de chaque agence pour les différentes sous-variables.
- Pour chaque modèle, nous avons calculé les métriques d'erreur (Mean Squared Error - MSE, Mean Absolute Error - MAE, Root Mean Squared Error - RMSE) afin d'évaluer leurs performances.
- Les erreurs de chaque modèle ont été stockées dans un dictionnaire. Ce dictionnaire a une structure hiérarchique : pour chaque agence, il contient les erreurs pour chaque sous-variable d'une variable.

#### Construction d'une Fonction de Sélection du Meilleur Modèle :

- Nous avons développé une fonction appelée `best_model` dont l'objectif est d'analyser les dictionnaires d'erreurs et de déterminer puis attribuer le modèle le plus performant pour chaque agence et pour chaque sous-variable.
- La fonction `best_model` prend en paramètre les dictionnaires d'erreurs. Elle compare les erreurs de différents modèles et sélectionne celui qui a la valeur minimale pour chaque sous-variable.
- Le résultat est un tableau qui présente les erreurs de prédiction pour les différents modèles (LSTM, SimpleRNN, AutoARIMA) appliqués aux données de toutes les agences. Chaque ligne du tableau correspond à une agence et montre les erreurs obtenues pour chaque modèle, ainsi que le meilleur modèle identifié ayant l'erreur minimale.
- La colonne "Agence" liste les agences avec leurs IDs.
- Les colonnes "Erreur LSTM", "Erreur SimpleRNN" et "Erreur AutoARIMA" indiquent les erreurs de prédiction respectives pour chacun des modèles.
- La colonne "Meilleur Modèle" montre le modèle qui a généré l'erreur la plus faible pour chaque agence.

## 1.6 Déploiement :

Pour la partie déploiement, nous avons développé une interface utilisateur permettant d'accéder aux deux principales composantes de notre solution. Cette interface facilite l'utilisation des outils de prédiction et d'analyse, offrant une expérience utilisateur intuitive et efficace. Les détails concernant cette interface seront abordés en profondeur dans la section "**Application Utilisateur**". Cela inclut la présentation des fonctionnalités et de l'ergonomie de l'interface, ainsi que des exemples d'utilisation.

*Pour mieux comprendre notre méthodologie, nous allons prendre l'exemple d'un fichier Python destiné à faire les prédictions des Frais Généraux. Le processus est le même pour toutes les autres variables.*

## **2 Etude de cas : Prédiction des Frais généraux**

Cette étude de cas illustre notre méthodologie de prédiction de la rentabilité des agences bancaires de la SGA, en se concentrant sur la variable des Frais généraux

### **2.1 Importation des données**

Notre processus de travail commence en important nos données dans un environnement dynamique et interactif, en l'occurrence Jupyter Notebook. Cette plateforme a été choisie pour sa capacité à faciliter l'exploitation du langage de programmation Python, reconnu pour sa puissance et sa flexibilité en matière d'analyse de données. Les données provenaient d'un fichier Excel que nous avons construit lors de la première partie de notre projet. Nous avons choisi la feuille intitulée "Frais généraux" afin d'importer les données historiques de chaque type de frais généraux pour toutes les agences de Janvier 2018 au Décembre 2023

### **2.2 Retraitement et Préparation de la Base de Données :**

Le prétraitement des données est une étape essentielle dans le processus d'analyse, visant à nettoyer, transformer et préparer les données pour une utilisation optimale dans les modèles de prévisions. Cette phase assure la qualité, la cohérence et la fiabilité de nos données, et est fondamentale pour développer des modèles prédictifs précis et prendre des décisions éclairées.

Pour le nettoyage de nos données, nous avons appliqué diverses méthodes d'analyse afin de comprendre et adapter nos séries temporelles. Nous avons commencé par supprimer les nouvelles agences sans historique, puis nous avons regroupé les chapitres de notre variable 'Frais Généraux' et ses sous-variables 'Frais Personnel', 'Amortissements' et 'Autres charges d'exploitation'. Ensuite, nous avons créé des tables pour réorganiser nos données par agence et par date, transformant ainsi chaque agence avec les sous-variables en une série temporelle de 72 instances, couvrant la période de janvier 2018 à décembre 2023.

```
# Importation des bibliothèques
import pandas as pd
import numpy as np
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.statespace.sarimax import SARIMAX
from matplotlib import pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_squared_error
from datetime import datetime, timedelta
from statsmodels.tsa.stattools import adfuller

# Chargement des données depuis un fichier Excel
data = pd.read_excel(r"D:\Downloads\hdd\BaseDeDonnées.xlsx", sheet_name='FG')

# Suppression des agences nouvellement ajoutées qui n'ont pas un historique suffisant
data = data[~data['ID Agence'].isin([24, 26, 150, 260, 400, 505, 271, 610])]

# Groupement des catégories des Frais Généraux (FG) en familles spécifiques
categorie_map = {
    **dict.fromkeys(["Primes variables et commissions (hors bonus)", "Appointements bruts",
                    "Autres avantages à CT", "Charges sociales et fiscales taux", "TAXES SUR FPA", "Bonus", "Congés payés"], "Frais Prsonnel"),
    **dict.fromkeys(["Frais de mission", "Frais télécommunication", "[Refacturation Inter Entités Groupe",
                    "Taxes & impositions", "Recours à l'extérieur", "Frais informatique (-fourniture info)", "Formation",
                    "Affranchissement", "Autres frais divers", "Frais de communication", "Fournitures et consommables",
                    "Provisions pour risques et charges", "Frais immobiliers"], "Autres Charges d'Exploitation"),
    "Amortissements": "Amortissements" }

data['Famille FG'] = data['Famille FG'].map(categorie_map)

# Création d'une table pivot pour réorganiser les données par agence et date
resultat_df = data.pivot_table(
    index=['ID Agence', 'ID Date'],
    columns='Famille FG',
    values='Montant',
    aggfunc='sum',
    fill_value=0 ).reset_index()

resultat_df.columns.name = None

# Fonction pour convertir l'ID de mois en une date réelle
def id_to_date(month_id):
    start_date = pd.Timestamp('2018-01-01')
    target_date = start_date + pd.DateOffset(months=month_id - 1)
    end_of_month = target_date + pd.offsets.MonthEnd(1)
    return end_of_month.date()

# Application de la fonction pour transformer les IDs de date en dates réelles
resultat_df['ID Date'] = resultat_df['ID Date'].apply(id_to_date)

# Création d'une table pivot pour chaque agence et date
pivot_df = resultat_df.pivot_table( index=['ID Agence', 'ID Date'],
    values = ['Amortissements', "Autres Charges d'Exploitation", "Frais Prsonnel"],
    aggfunc = 'sum' )
```

Figure 4.2 - Importation des données

### 2.3 Séparation des données

Nous avons séparé les données des frais généraux (FG) pour chaque agence dans un dictionnaire de Data-Frames appelé 'df\_age'. Chaque Data-Frame à l'intérieur du dictionnaire contient les informations des FG d'une agence spécifique. La clé du dictionnaire représente l'ID de l'agence et les valeurs correspondent aux valeurs de chaque type de FG, chacune dans une

colonne. Pour accéder, par exemple, aux valeurs des frais personnels, des amortissements et des autres charges d'exploitation pour l'agence 2, il suffit d'utiliser 'df\_age [2]'.

```
# Création de dictionnaires pour stocker Les DataFrames individuels pour chaque agence
df_age = {}
for id_agence, groupe in pivot_df.groupby('ID Agence'):
    df_age[id_agence] = groupe
```

Figure 4.3 - Séparation des données

ID Agence	ID Date	Amortissements	Autres Charges d'Exploitation	Frais Prsonnel
2	2018-01-31	12892.840179	3.475896e+05	1.528588e+06
	2018-02-28	11254.573329	2.217082e+05	1.565814e+06
	2018-03-31	11072.422326	4.759107e+05	1.621873e+06
	2018-04-30	11073.062187	4.921786e+05	4.338663e+06
	2018-05-31	11173.432233	3.239805e+05	1.701290e+06
	...	...	...	...
	2023-08-31	75532.210379	2.848787e+05	2.299142e+06
	2023-09-30	62132.340000	1.666861e+05	1.875224e+06
	2023-10-31	70093.821000	2.314573e+05	2.008699e+06
	2023-11-30	74183.210379	1.927789e+06	2.144216e+06
	2023-12-31	76221.210379	-5.526278e+05	4.283947e+06

72 rows × 3 columns

Figure 4.4 - Data-Frame des Frais Généraux de l'agence 2

## 2.4 Vérification de la Stationnarité des Séries Temporelles

### 2.4.1 Test ADF (Augmented Dickey-Fuller)

Le test ADF (Augmented Dickey-Fuller) est un test statistique couramment utilisé en économétrie pour déterminer si une série temporelle est stationnaire ou non. Il est essentiel pour analyser les séries temporelles et détecter la présence de racines unitaires, ce qui indique une non-stationnarité.

La non-stationnarité d'une série temporelle peut poser des problèmes lors de l'analyse et de la modélisation, car elle complique la prévision et la prise de décisions basées sur les tendances et les modèles de données. Le test ADF permet d'évaluer si une série temporelle suit un processus de marche aléatoire ou un processus stationnaire.

Le test ADF repose sur un modèle autorégressif (AR) et utilise la régression pour estimer la présence de racines unitaires dans la série temporelle. La version augmentée du test Dickey-

Fuller prend en compte des retards supplémentaires, ce qui permet de mieux traiter les séries temporelles avec des tendances complexes.

L'hypothèse nulle du test ADF stipule que la série temporelle possède une racine unitaire, indiquant ainsi qu'elle est non stationnaire. Si la valeur du test ADF est inférieure à une certaine valeur critique, l'hypothèse nulle est rejetée, ce qui indique que la série temporelle est stationnaire. En revanche, si la valeur du test ADF est supérieure à la valeur critique, l'hypothèse nulle est acceptée, ce qui signifie que la série temporelle est non stationnaire.

Le test ADF est largement utilisé dans l'analyse des séries temporelles, notamment dans les domaines de la modélisation économique, de la finance et d'autres secteurs où la stationnarité des données est cruciale pour les prévisions et les analyses. Il fournit une approche statistique pour évaluer la présence de racines unitaires et la stationnarité des séries temporelles, ce qui aide à mieux comprendre les modèles et les tendances des données.

Nous avons vérifié la stationnarité de chaque série temporelle dans notre dictionnaire en utilisant une fonction appelée 'check\_stationarity'. Cette fonction prend en paramètre un dictionnaire de Data-Frames et détermine si chaque série temporelle est stationnaire. Elle applique de manière itérative le test de stationnarité ADF (Augmented Dickey-Fuller) à chaque série à l'intérieur du dictionnaire, retournant le résultat pour chaque série.

```
# Fonction pour vérifier si une série temporelle est constante
def is_constant(series):
    return series.nunique() == 1

# Fonction pour vérifier la stationnarité d'une série temporelle avec le test ADF
def check_stationarity(data):
    if is_constant(data):
        return 'Constant series'
    else:
        result = adfuller(data.dropna())
        p_value = result[1]
        return p_value < 0.05

# Vérification de la stationnarité pour chaque colonne de chaque agence
columns = df_age[1].columns
stationarity_results = {}

# Affichage des résultats de la vérification de stationnarité
for agence_id, df in df_age.items():
    stationarity_results[agence_id] = {}
    for col in columns:
        is_stationary = check_stationarity(df[col])
        stationarity_results[agence_id][col] = is_stationary
for agence_id, results in stationarity_results.items():
    print(f'Agence ID : {agence_id} ')
    for columns, is_stationary in results.items():
        print(f"{columns} : {'Stationary' if is_stationary == True else 'Non-Stationary' if is_stationary == False else is_stationary}")
```

Figure 4.5 - Vérification de la Stationnarité

### 2.4.2 Transformation des séries temporelles non stationnaires

Nous avons ensuite appliqué des différenciations à chaque série temporelle non stationnaire afin de les rendre stationnaires et prêtes à être utilisées dans des modèles statistiques de prédiction des séries temporelles, qui requièrent la stationnarité de la série. Notre fonction effectue différents niveaux de différenciation selon la série : si la série est déjà stationnaire, aucune différenciation n'est appliquée. Si une seule différenciation est suffisante pour que la série devienne stationnaire, alors une seule différenciation est appliquée ; sinon, deux sont appliquées. La fonction stocke les séries temporelles stationnaires pour chaque agence dans un

dictionnaire nommé 'df\_diffs', et affiche le nombre de différenciations appliquées pour chaque agence et chaque variable afin de rendre la série stationnaire.

```
agence_stationary_counts = {}

for agence_id, results in stationarity_results.items():
    stationary_count = sum(1 for is_stationary in results.values() if is_stationary == True)
    agence_stationary_counts[agence_id] = stationary_count

print("\nNombre de serie stationnaires par agence:")
for agence_id, count in agence_stationary_counts.items():
    print(f"Agence ID: {agence_id}: {count} series stationnaires")

def difference(data) :
    return data.diff().dropna()

diffs_age = {} # Dictionnaire pour stocker les séries différenciées

for agency_id, df in df_age.items():
    diffs_age[agency_id] = {}
    for column in df.columns:
        if not check_stationarity(df[column]):
            df[column] = difference(df[column])
            diffs_age[agency_id][column] = df[column]
            if not check_stationarity(df[column]):
                print(f"Série {column} de l'agence {agency_id} reste non stationnaire après différenciation")

for agency_id, df in df_age.items():
    for column in df.columns:
        if not check_stationarity(df[column]):
            df[column] = difference(difference(df[column]))
            diffs_age[agency_id][column] = df[column]
            if not check_stationarity(df[column]):
                print(f"Série {column} de l'agence {agency_id} reste non stationnaire après la 2ème différenciation")

stationary_counts = {agency_id: {column: check_stationarity(df[column]) for column in df.columns} for agency_id, df in df_age.items()}

# Affichage des séries différenciées
for agency_id, series in diffs_age.items():
    print(f"Agence ID: {agency_id}")
    for column, data in series.items():
        print(f"Colonne: {column}, Séries différenciées:\n{data.head()}")
```

Figure 4.6 - Transformation des séries temporelles

## 2.5 Modélisation :

La modélisation représente une étape nécessaire dans toute analyse prédictive. Elle implique l'utilisation d'algorithmes statistiques ou de machine Learning pour anticiper une variable d'intérêt. Afin de garantir que nos modèles peuvent généraliser efficacement à partir de nos données et produire des prédictions précises sur des données inédites, il est impératif de diviser notre jeu de données en deux sous-ensembles : l'ensemble d'apprentissage (ou "training set") et l'ensemble de test (ou "test set").

Dans cette section, nous commencerons par examiner une liste de modèles préalablement identifiés à partir de l'état de l'art, dans le but de déterminer les techniques de modélisation les plus pertinentes pour notre analyse. Nous passerons en revue l'architecture des modèles sélectionnés, en fournissant une explication détaillée de leur structure et de leurs principes de fonctionnement.

Par la suite, nous adapterons nos données afin qu'elles correspondent aux exigences structurelles de chaque modèle sélectionné, ce qui nous permettra d'entraîner ces modèles de manière efficace. Bien que cette phase puisse être complexe, elle est indispensable pour générer des prédictions précises et pertinentes à partir de nos données.

### 2.5.1 Identification des modèles

Les modèles identifiés dans l'état de l'art sont les suivants :

- Auto-Arima
- LSTM
- RNN (Réseaux de Neurones Récurrents)
  - **Auto-ARIMA** : Auto-ARIMA est une version automatisée de ARIMA, où les ordres (p, d, q) de ARIMA sont sélectionnés automatiquement en fonction des performances de modélisation sur les données fournies. Cela simplifie le processus de modélisation en éliminant la nécessité de choisir manuellement les ordres.
  - **RNN (Réseaux de Neurones Récurrents)** : Les réseaux de neurones récurrents sont une classe de réseaux neuronaux spécialement conçus pour traiter des données séquentielles en prenant en compte les dépendances temporelles entre les observations successives. Ils sont souvent utilisés pour la modélisation de séries temporelles en raison de leur capacité à gérer la dépendance temporelle.
  - **LSTM (Long Short-Term Memory)** : LSTM est un type de réseau de neurones récurrents (RNN) spécialement conçu pour traiter les séries temporelles et d'autres données séquentielles. Il est capable de capturer les dépendances à long terme dans les séries temporelles grâce à sa structure de mémoire à court et long terme.

### 2.5.2 Préparation des données d'entraînement :

Les modèles sélectionnés peuvent nécessiter des architectures différentes pour la phase d'entraînement :

- **Pour les modèles RNN et LSTM** : Nous avons séparé nos séries temporelles initiales, stockées dans le dictionnaire 'df\_age', en ensembles de données d'entraînement (de 2018 à 2022) et de données de test (pour l'année 2023), pour chaque variable et pour chaque agence individuellement.
- **Pour Auto-ARIMA** : nous avons séparé nos séries temporelles différenciées stockées dans le dictionnaire 'df\_diffs', de la même manière, c'est-à-dire en ensembles de données d'entraînement (de 2018 à 2022) et de données de test (pour l'année 2023), pour chaque variable et pour chaque agence individuellement.

### 2.5.3 Entraînement des modèles et optimisation de l'Apprentissage Automatique

Le tableau ci-dessous est conçu avec plusieurs colonnes pour faciliter la comparaison des différentes approches et pour offrir une meilleure compréhension des caractéristiques de chaque modèle appliqué sur différentes séries temporelles.

- La colonne Modèle indique le nom du modèle utilisé.
- La colonne Série temporelle précise la série temporelle spécifique à laquelle le modèle a été appliqué.

- La colonne Entraînement est subdivisée en deux sous-colonnes : Compilation et Adaptation, qui détaillent les étapes spécifiques du processus d'entraînement du modèle.
- La colonne Architecture/Hyperparamètres est également subdivisée en deux sous-colonnes : Couches et Paramètres, fournissant des détails sur l'architecture du modèle et les hyperparamètres utilisés.

Modèle	Série Temporelle	Entraînement	
		Compilation	Adaptation
<b>Auto-ARIMA</b>	Frais Personnels	/	/
	Amortissement		
	Autres charges		

<b>LSTM</b>	Frais Personnels	Optimiseur = 'Adam' Fonction de perte = MSE Métrique = MAE	Taille des lots de données = 1 Epochs = 150
	Amortissements	Optimiseur = 'Adam' Fonction de perte = MSE Métrique = MAE	Taille des lots de données = 1 Epochs = 150
	Autres Charges	Optimiseur = 'Adam' Fonction de perte = MSE Métrique = MAE	Taille des lots de données = 1 Epochs = 150
<b>RNN</b>	Frais Personnels	Optimiseur = 'Adam' Fonction de perte = MSE	Taille des lots de données = 1 Epochs = 50
	Amortissements	Optimiseur = 'Adam' Fonction de perte = MSE	Taille des lots de données = 1 Epochs = 50
	Autres Charges	Optimiseur = 'Adam' Fonction de perte = MSE	Taille des lots de données = 1 Epochs = 50

Tableau 4.1 - Modélisation - Compilation et Adaptation

Modèle	Architecture/Hyperparamètres	
	Couches	Paramètres
<b>Auto -ARIMA</b>	/	/
		/
		Seasonal = True
		Seasonal = False
<b>LSTM</b>	LSTM (Couche 1) Dropout LSTM (Couche 2) Dense	128 unités, return-sequence = True 0.2 128 unités, return-sequence = False 1 unité
	LSTM (Couche 1) Dropout LSTM (Couche 2) Dense	150 unités, activation = 'relu' 0.2 75 unités, return-sequence = False 1 unité
<b>RNN</b>	SimpleRNN Dense	100 unités, activation = 'tanh' 1 Densité
	SimpleRNN Dense	150 unités, activation = 'tanh' 1 Densité

Tableau 4.2 - Modélisation - Optimisation

## 2.6 Résultats et Evaluation

### 2.6.1 Métriques

**RMSE (Root Mean Square Error)** : Le RMSE est utilisé pour calculer la racine carrée de la moyenne des erreurs quadratiques. C'est une mesure populaire pour évaluer l'exactitude des prédictions d'un modèle. La formule du RMSE est la suivante :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

**MSE (Mean Squared Error)** : Le MSE est la moyenne des carrés des erreurs. Il pénalise fortement les grandes erreurs en raison de la mise au carré, ce qui peut être utile pour mettre en évidence les modèles qui font de grandes erreurs de prédiction. C'est une autre mesure

couramment utilisée pour évaluer l'exactitude des prédictions d'un modèle. La formule du MSE est la suivante :

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

**MAE (Mean Absolute Error) :** Le MAE est la moyenne des valeurs absolues des erreurs. C'est une mesure robuste par rapport aux valeurs aberrantes, car elle ne met pas les erreurs au carré. La formule du MAE est la suivante :

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

### 2.6.2 Évaluation

Pour sélectionner le meilleur modèle de prédiction des Frais Généraux (FG) pour chaque agence, nous avons suivi une méthodologie rigoureuse d'évaluation des erreurs de prédiction. Voici les étapes détaillées de notre processus :

#### Stockage des Erreurs des Modèles :

- Pour chaque modèle, nous avons calculé les métriques d'erreur (Mean Squared Error - MSE, Mean Absolute Error - MAE, Root Mean Squared Error - RMSE) afin d'évaluer leur performance.
- Les erreurs de chaque modèle ont été stockées dans un dictionnaire. Ce dictionnaire a une structure hiérarchique : pour chaque agence, il contient les erreurs pour chaque sous-variable des Frais Généraux.

#### Construction d'une Fonction de Sélection du Meilleur Modèle :

- Nous avons développé une fonction appelée '**best\_model**' dont l'objectif est d'analyser les dictionnaires d'erreurs et de déterminer le modèle le plus performant pour chaque agence et pour chaque sous-variable.
- La fonction **best\_model** prend en paramètre les dictionnaires d'erreurs. Elle compare les erreurs de différents modèles et sélectionne celui qui a la valeur minimale pour chaque sous-variable.
- Le résultat est un nouveau dictionnaire. Pour chaque agence et chaque sous-variable des Frais Généraux, ce dictionnaire indique le modèle ayant généré l'erreur la plus faible.

### 2.6.3 Résultats :

Nous avons utilisé les métriques prédéfinies précédemment pour évaluer la performance de nos modèles. Les résultats obtenus pour la prédiction des frais personnels sont présentés ci-dessous :

Agence	RMSE LSTM	RMSE SimpleRNN	RMSE AutoARIMA	MAE LSTM	MAE SimpleRNN	MAE AutoARIMA	MSE LSTM	MSE SimpleRNN	MSE AutoARIMA	Meilleur Modèle
1	0.214380	0.224380	0.352930	0.103898	0.127120	0.127720	0.045959	0.050346	0.124560	LSTM
2	0.122540	0.227270	0.594160	0.051763	0.067415	0.068730	0.015016	0.051652	0.353026	LSTM
3	0.111600	0.217750	0.595130	0.045159	0.064384	0.062232	0.012455	0.047415	0.354180	LSTM
4	0.104930	0.177540	0.094930	0.067307	0.071545	0.055782	0.011010	0.031520	0.009012	AutoARIMA
5	0.082800	0.185790	0.072800	0.085244	0.083418	0.069177	0.006856	0.034518	0.005300	AutoARIMA
6	0.082800	0.178140	0.429760	0.085244	0.060519	0.091358	0.006856	0.031734	0.184694	LSTM
7	0.077900	0.238640	0.067900	0.052666	0.060519	0.046010	0.006068	0.056949	0.004610	AutoARIMA
8	0.118310	0.206440	0.594130	0.075517	0.086650	0.093566	0.013997	0.042617	0.352990	LSTM
9	0.105060	0.095060	0.591130	0.071694	0.053305	0.064551	0.011038	0.009036	0.349435	SimpleRNN
10	0.110400	0.265760	0.558960	0.048839	0.068048	0.067538	0.012188	0.070628	0.312436	LSTM
11	0.076030	0.199930	0.066030	0.079630	0.080354	0.064441	0.005781	0.039972	0.004360	AutoARIMA
12	0.123500	0.216030	0.350240	0.077285	0.091275	0.087833	0.015252	0.046669	0.122668	LSTM
13	0.085617	0.106620	0.373490	0.076352	0.063000	0.081029	0.007330	0.011368	0.139495	SimpleRNN
14	0.085620	0.030620	0.594750	0.008834	0.070787	0.084122	0.007331	0.000938	0.353728	SimpleRNN
15	0.084090	0.074090	0.506930	0.082935	0.068954	0.084328	0.007071	0.005489	0.256978	SimpleRNN
16	0.084170	0.203980	0.327760	0.073589	0.092788	0.087053	0.007085	0.041608	0.107427	LSTM
17	0.090210	0.191120	0.080210	0.074853	0.078759	0.061384	0.008138	0.036527	0.006434	AutoARIMA
18	0.101240	0.267970	0.524050	0.073477	0.087999	0.085723	0.010250	0.071808	0.274628	LSTM
19	0.096600	0.255820	0.326520	0.057692	0.072216	0.069101	0.009332	0.065444	0.106615	LSTM
20	0.089560	0.238340	0.373240	0.054246	0.066010	0.069229	0.008021	0.056806	0.139308	LSTM
21	0.105590	0.095590	0.489620	0.076968	0.062779	0.081927	0.011149	0.009137	0.239728	SimpleRNN
22	0.081970	0.255370	0.071970	0.047279	0.076461	0.060655	0.006719	0.065214	0.005180	AutoARIMA
23	0.089610	0.193660	0.520590	0.069079	0.085401	0.079210	0.008030	0.037504	0.271014	LSTM
24	0.093320	0.264620	0.392280	0.059776	0.076412	0.071557	0.008709	0.070024	0.153884	LSTM
25	0.111300	0.200340	0.473400	0.080650	0.066417	0.078934	0.013043	0.080646	0.223824	LSTM

Tableau 4.3 - Tableau des résultats de l'évaluation des prévisions des FPs

## 2.7 Interprétation des résultats

L'analyse des résultats nous permet de dégager plusieurs points importants sur la performance des modèles de prédiction utilisés. Voici notre analyse détaillée

### Analyse des erreurs :

- **RMSE (Root Mean Squared Error)** : Une RMSE plus basse indique que les prédictions sont en moyenne plus proches des valeurs réelles. LSTM affiche généralement les RMSE les plus basses, ce qui traduit une bonne précision globale.
- **MAE (Mean Absolute Error)** : MAE étant une mesure plus robuste face aux valeurs aberrantes, les résultats confirment également la supériorité de LSTM dans la majorité des cas. Cela souligne que LSTM minimise efficacement les erreurs absolues.
- **MSE (Mean Squared Error)** : Le MSE, qui accentue les grandes erreurs, met en évidence la capacité de chaque modèle à éviter les prédictions fortement erronées. LSTM affiche les MSE les plus bas, démontrant sa fiabilité.

### Performance Globale des Modèles :

- **LSTM (Long Short-Term Memory)** apparaît comme le modèle le plus performant dans la majorité des cas, avec des valeurs de RMSE, MAE et MSE souvent plus basses que les autres modèles. Cela indique une capacité supérieure à capturer les dynamiques temporelles et les relations complexes dans les données des différentes agences.

- **Auto-ARIMA** (AutoRegressive Integrated Moving Average) se distingue particulièrement pour certaines agences spécifiques, notamment les agences 4 et 5. Cela peut suggérer que pour des séries temporelles avec des patterns plus linéaires et saisonniers, Auto-ARIMA est mieux adapté.
- **SimpleRNN** (Simple Recurrent Neural Network) est le meilleur modèle pour quelques agences (comme les agences 9 et 13), ce qui indique que dans certains contextes, des architectures plus simples peuvent suffire à capturer les dynamiques des données. Cependant, il est globalement moins performant comparé à LSTM.

## Conclusion :

L'analyse des résultats des prévisions montre que **LSTM** est globalement le modèle le plus performant, mais **AutoARIMA** et **SimpleRNN** ont aussi leur place pour des cas spécifiques. Cette approche hybride permet d'optimiser les prédictions en fonction des caractéristiques particulières des données de chaque agence, garantissant ainsi des prédictions plus précises et fiables.

## 3 Etude Spécifique : Analyse Prédicative des Frais Personnel pour l'Agence 1

Dans cette section, nous allons passer en revue toutes les étapes entreprises pour prédire les frais personnels de l'agence 1.

### 3.1 Analyse des Données

L'analyse des données est une étape primordiale dans l'étude des séries temporelles. Elle vise à explorer et à comprendre les tendances, les variations et les structures sous-jacentes présentes dans les données collectées sur une période de cinq ans, de 2018 à 2023. Cette analyse permettra d'identifier les modèles saisonniers, les fluctuations significatives, ainsi que les événements exceptionnels ayant impacté les dépenses de l'agence.

#### 3.1.1 Statistiques descriptives

Le résultat de `df.describe()` fournit des statistiques descriptives pour la série temporelle des frais personnels de l'agence 1. Voici une interprétation des différentes mesures statistiques données dans le tableau :

- **Count** : La série temporelle des frais personnels de l'agence 1 comporte un total de 72 observations, indiquant une fréquence de collecte de données mensuelle sur une période de 6 ans.
- **Mean** : La moyenne des frais personnels de l'agence 1 est d'environ 4 897 914, ce qui suggère que, en moyenne, l'agence dépense près de 4,9 millions pour les frais personnels chaque mois.

- **Std** : L'écart-type des frais personnels de l'agence 1 est d'environ 2 427 166, indiquant une variabilité substantielle autour de la moyenne. Cela montre que les frais personnels peuvent fluctuer de manière significative d'un mois à l'autre.
- **Min** : Le montant minimum des frais personnels enregistré est d'environ 2 219 847, ce qui représente la dépense la plus faible observée dans la série. Cela pourrait correspondre à des périodes de réduction des coûts ou de moindre activité.
- **25%** : Le premier quartile (Q1) est d'environ 2 975 104, ce qui signifie que 25% des observations de frais personnels sont inférieures à cette valeur. Cela indique que dans un quart des mois, les frais personnels sont relativement bas.
- **50%** : La médiane (deuxième quartile, Q2) est d'environ 3 920 304, ce qui représente le point central de la série. Cela signifie que 50% des observations sont en dessous de ce montant et 50% sont au-dessus, offrant une indication claire de la tendance centrale des frais personnels.
- **75%** : Le troisième quartile (Q3) est d'environ 6 480 019, ce qui indique que 75% des frais personnels sont inférieurs à cette valeur. Cela montre que dans un quart des mois, les frais personnels atteignent des niveaux relativement élevés.
- **Max** : Le montant maximum des frais personnels enregistré est d'environ 13 436 300, ce qui représente une valeur exceptionnellement élevée par rapport aux autres observations. Ce pic pourrait être dû à des événements extraordinaires ou des dépenses exceptionnelles.

Statistique	Valeur
Count	7.200000e+01
Mean	4.897914e+06
Std	2.427166e+06
Min	2.219847e+06
25%	2.975104e+06
50%	3.920304e+06
75%	6.480019e+06
Max	1.343630e+07

Tableau 4.4 - Statistiques descriptives pour la série temporelle FP de l'agence 1

Ces statistiques descriptives donnent un aperçu global de la distribution et des caractéristiques des frais personnels pour l'agence 1 dans la série temporelle. Ils nous aident à comprendre les tendances, la dispersion et la portée des valeurs des frais personnels, ce qui est utile pour l'analyse et la modélisation ultérieures.

#### Interprétation des statistiques :

**Dispersion et Variabilité** : La grande différence entre le minimum et le maximum, ainsi que l'écart-type élevé, indiquent une grande dispersion dans les frais personnels, reflétant peut-être des variations dans l'activité de l'agence ou des dépenses irrégulières au fil du temps.

**Asymétrie et Distribution** : La différence notable entre la moyenne et la médiane peut suggérer une distribution asymétrique des frais personnels. La présence de valeurs extrêmes, comme le maximum, tire la moyenne vers le haut, indiquant une possible asymétrie positive.

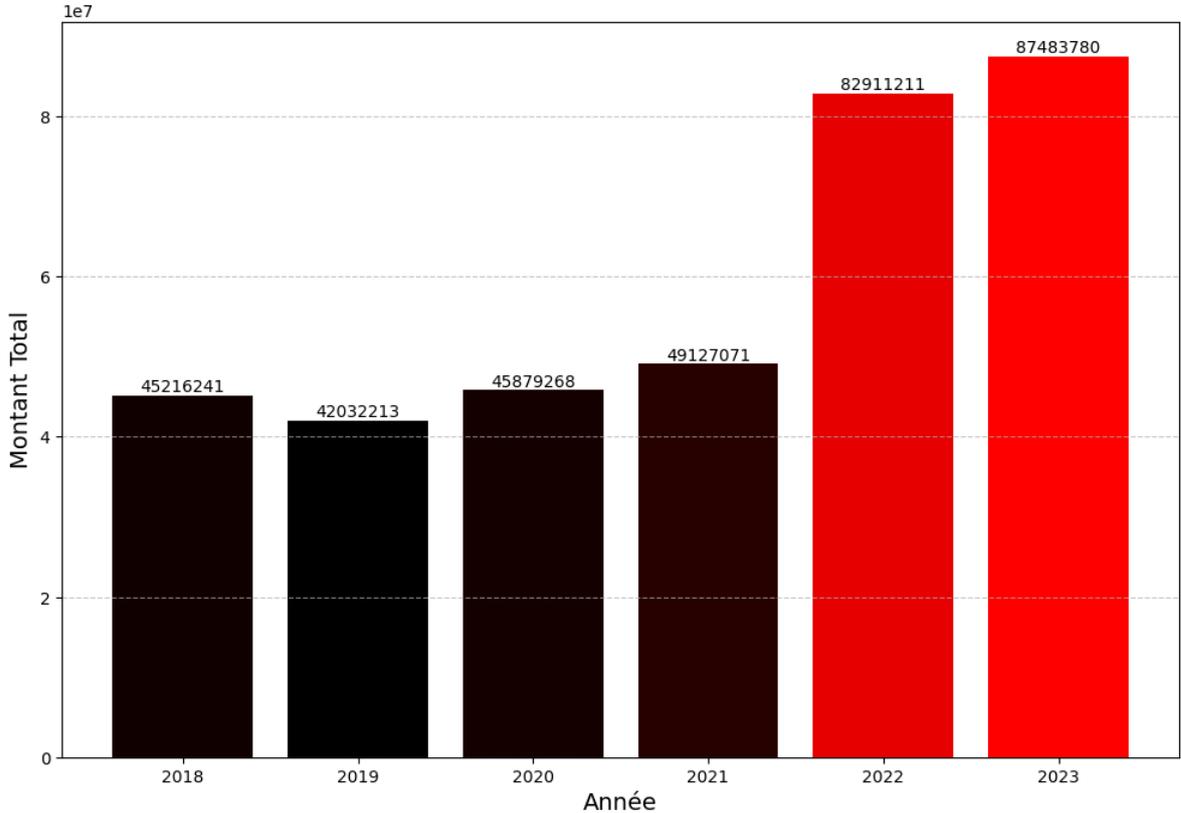


Figure 4.7 - Histogramme des Soldes annuels des 'Frais Personnel' de l'agence 1

3.1.2 Visualisation des données :

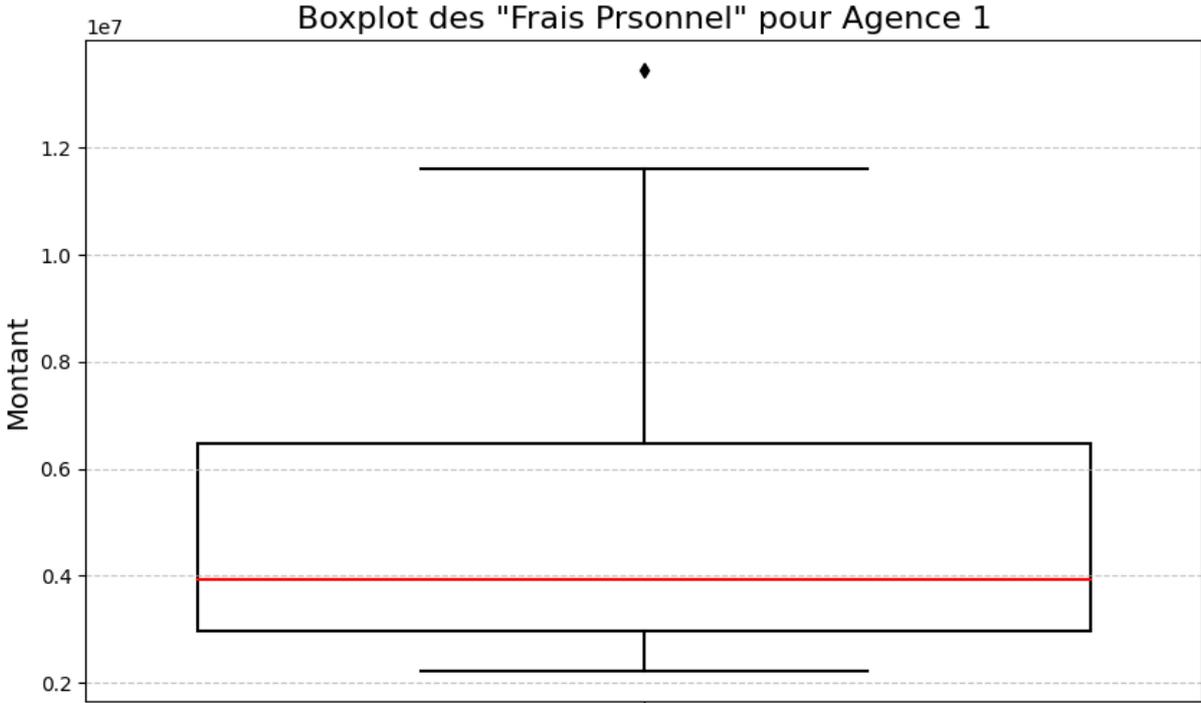


Figure 4.8 - Boite à Moustaches des mouvements des FP de l'agence 1

La série temporelle des frais généraux pour l'agence 1 est devenue stationnaire après l'application d'une seule différenciation.

Avant différenciation :

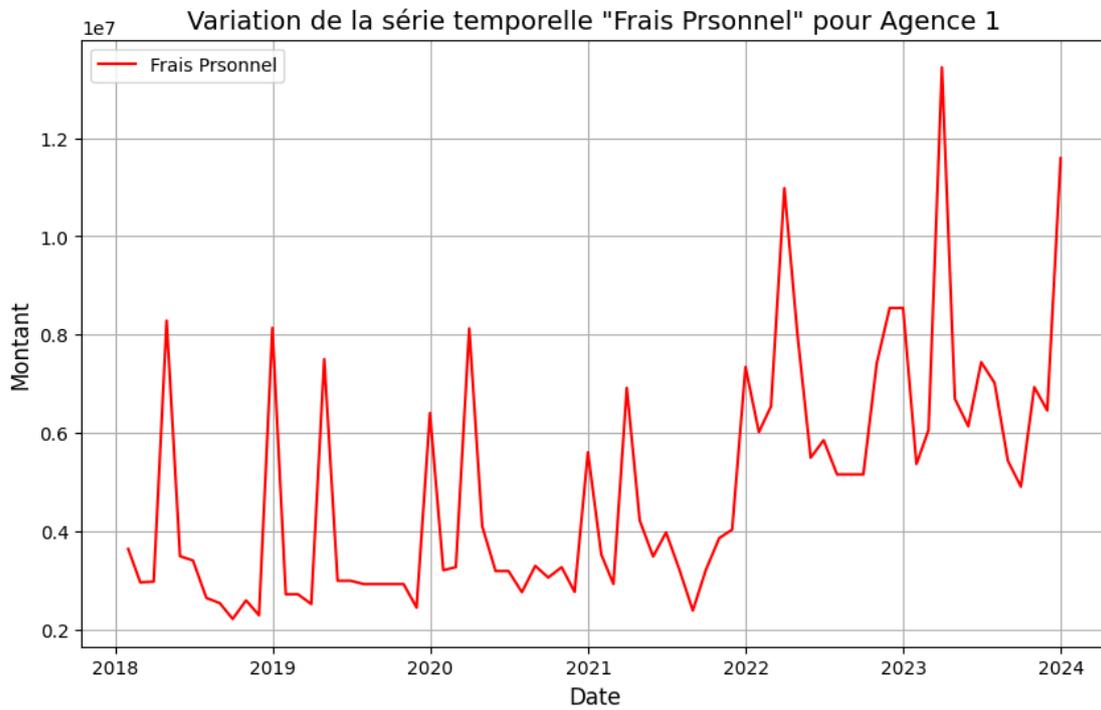


Figure 4.9 - Variations des FP de l'agence 1 (Avant différenciation)

Après différenciation :

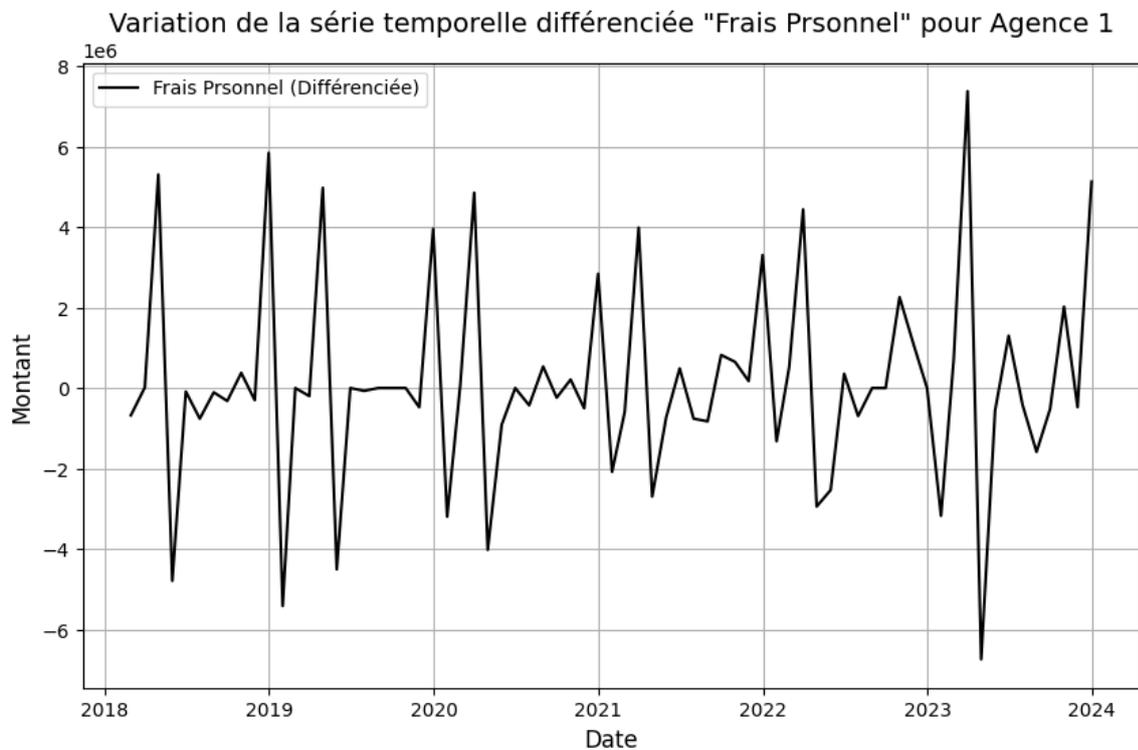


Figure 4.10 - Variations des FP de l'agence 1 (Après différenciation)

## Interprétation des Graphiques des Frais Personnels de l'Agence 1

La Variation de la Série Temporelle :

- **Tendance Générale** : La série temporelle montre une fluctuation significative des frais personnels sur la période de 2018 à 2024, avec des pics et des creux récurrents.
- **Saisonnalité** : Il y a des variations saisonnières apparentes, où des pics élevés de dépenses se produisent à intervalles réguliers, probablement liés à des événements spécifiques comme des périodes de bonus ou des augmentations salariales.
- **Variabilité** : La série présente une grande variabilité, surtout visible avec les pics élevés autour de 2022 et 2023, ce qui peut indiquer des événements ou décisions exceptionnels ayant entraîné des augmentations importantes des frais.

L'Histogramme des Soldes annuels de la série :

- **Distribution Annuelle** : Les dépenses annuelles en frais personnels montrent une tendance à la hausse, particulièrement marquée en 2022 et 2023 probablement liés à des facteurs comme l'augmentation de la masse salariale.
- **Comparaison des Années** : Les années 2018 à 2021 montrent des dépenses relativement stables, mais une augmentation significative est observée en 2022 et 2023, où les montants atteignent des sommets (82,911,211 et 87,483,780 respectivement).
- **Analyse des Données** : Cette augmentation peut être due à des augmentations salariales, des primes exceptionnelles, ou des embauches massives.

La Boite à moustache de la série :

- **Dispersion des Données** : Le boxplot montre une dispersion des frais personnels avec des valeurs allant de  $0,2 \times 10^7$  à environ  $0,6 \times 10^7$  pour la majorité des données.
- **Outliers** : Un outlier est observé au-dessus de  $1,2 \times 10^7$ , ce qui indique une dépense exceptionnelle durant une période.
- **Médiane** : La ligne rouge représentant la médiane se situe juste en dessous de  $0,4 \times 10^7$ , ce qui indique que 50% des frais personnels sont inférieurs à ce montant.
- **Quartiles** : Les valeurs des premier et troisième quartile montrent que 50% des données se situent entre environ  $0,3 \times 10^7$  et  $0,6 \times 10^7$ , indiquant une variabilité modérée dans les dépenses hors des valeurs aberrantes.

En conclusion, la série temporelle des frais personnels présente des tendances et des fluctuations saisonnières marquées, avec des pics réguliers suggérant des cycles annuels dans les dépenses. Les variations significatives, particulièrement visibles en 2022 et 2023, nécessitent une attention particulière pour comprendre les causes sous-jacentes, pouvant être liées à des décisions stratégiques internes. De plus, l'identification d'outliers indique des événements exceptionnels, qu'il est primordial d'analyser pour comprendre les raisons de ces dépenses exceptionnelles et planifier en conséquence.

### 3.1.3 Décomposition des séries temporelles

La décomposition saisonnière est une technique utilisée en analyse des séries temporelles pour diviser une série en plusieurs composantes distinctes : **tendance**, **saisonnalité** et **résidus**. Cette méthode permet de révéler les structures sous-jacentes d'une série temporelle et de mieux comprendre les modèles saisonniers ainsi que les variations à long terme.

L'objectif de la décomposition saisonnière est de séparer les différentes composantes d'une série temporelle. Voici une brève description des composantes clés :

- **Tendance** : Il s'agit de la composante à long terme qui représente l'évolution générale de la série au fil du temps. Elle indique la direction et la tendance globale de la série, qu'elle soit croissante, décroissante ou stable.
- **Saisonnalité** : Cette composante représente les variations périodiques régulières dans la série, souvent dues à des effets saisonniers ou cycliques. La saisonnalité peut être quotidienne, hebdomadaire, mensuelle ou annuelle, selon la fréquence des variations saisonnières présentes dans les données.
- **Résidus** : Ce sont les variations aléatoires ou non expliquées par la tendance et la saisonnalité. Les résidus sont souvent considérés comme du bruit ou des erreurs dans la série temporelle.

La décomposition saisonnière utilise généralement des méthodes mathématiques telles que la méthode de la moyenne mobile ou la décomposition de Fourier pour extraire ces différentes composantes. Une fois la décomposition effectuée, il devient plus facile d'analyser et de modéliser chaque composante séparément, ce qui peut faciliter les prévisions ou l'identification de modèles spécifiques dans les données.

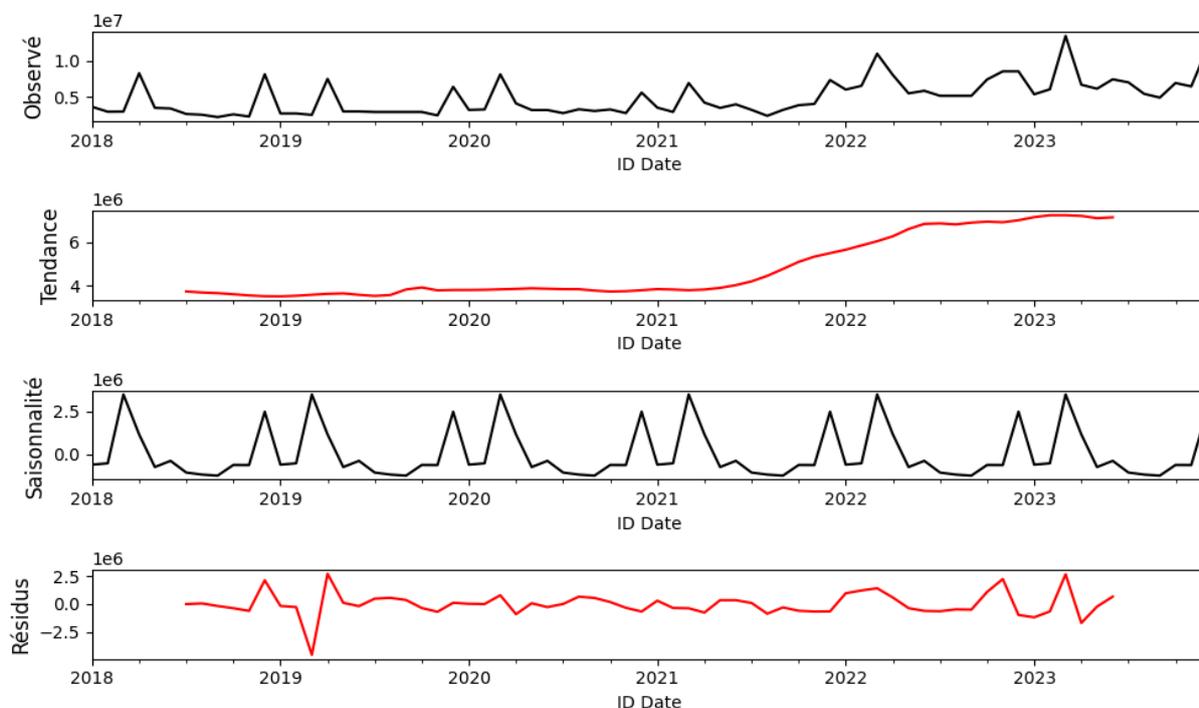


Figure 4.11 - Décomposition de la série temporelle des FP de l'agence 1

### Interprétation :

- **Observé** : La série temporelle originale montre des fluctuations régulières avec des pics et des creux.
- **Tendance** : La tendance, en rouge, indique une augmentation progressive des frais personnels, particulièrement après 2021, suggérant une hausse constante des dépenses.
- **Saisonnalité** : La composante saisonnière, en noir, révèle des cycles annuels réguliers dans les données, ce qui signifie des variations récurrentes chaque année.
- **Résidus** : Les résidus, en rouge, représentent les variations imprévues ou aléatoires dans les données après avoir enlevé la tendance et la saisonnalité. Les résidus doivent idéalement être sans structure pour indiquer que les modèles de tendance et de saisonnalité ont bien capturé les principaux motifs des données.

#### 3.1.4 Test de stationnarité

```
# Test ADF avant différenciation
adf_result_before = adfuller(agence_1_df['Frais Prsonnel'])
print("ADF Test Avant Différenciation:")
print(f"ADF Statistic : {adf_result_before[0]}")
print(f"p-value : {adf_result_before[1]}")
print(f"Num Lags : {adf_result_before[2]}")
print(f"Num Observations Used For ADF Regression and Critical Values Calculation : {adf_result_before[3]}")
print("Critical Values :")
for key, value in adf_result_before[4].items(): # type: ignore
    print(f"    {key} : {value}")
```

✓ 0.0s

```
ADF Test Avant Différenciation:
ADF Statistic : 0.46427760274217966
p-value : 0.9837459334125123
Num Lags : 12
Num Observations Used For ADF Regression and Critical Values Calculation : 59
Critical Values :
1% : -3.5463945337644063
5% : -2.911939409384601
10% : -2.5936515282964665
```

Figure 4.22 - Test de stationnarité (Avant différenciation)

### Interprétation :

- **P-value** : La p-value de 0.9837459334125123 est très élevée, bien au-dessus de 0.05. Cela indique que nous ne pouvons pas rejeter l'hypothèse nulle de Non-Stationnarité.
- **ADF Statistic** : La statistique ADF de 0.46427760274217966 est bien supérieure aux valeurs critiques à 1%, 5% et 10% (-3.5463945337644063, -2.911939409384601, -2.5936515282964665 respectivement). Cela signifie que la série temporelle n'est pas stationnaire.
- **Conclusion** : Avant la différenciation, la série temporelle des frais personnels de l'Agence 1 présente des caractéristiques de non-stationnarité. Cela est confirmé par une p-value élevée et une statistique ADF qui n'est pas inférieure aux valeurs critiques. Une série non stationnaire peut avoir une tendance ou une saisonnalité qui doit être traitée pour des analyses plus précises.

```

# Test ADF après différenciation
adf_result_after = adfuller(diff_agence_1_df['Frais Prsonnel'])
print("\nADF Test Après Différenciation:")
print(f"ADF Statistic : {adf_result_after[0]}")
print(f"p-value : {adf_result_after[1]}")
print(f"Num Lags : {adf_result_after[2]}")
print(f"Num Observations Used For ADF Regression and Critical Values Calculation : {adf_result_after[3]}")
print("Critical Values :")
for key, value in adf_result_after[4].items(): # type: ignore
    print(f"    {key} : {value}")

```

✓ 0.0s

```

ADF Test Après Différenciation:
ADF Statistic : -3.717335920175868
p-value : 0.0038758123878013385
Num Lags : 11
Num Observations Used For ADF Regression and Critical Values Calculation : 59
Critical Values :
1% : -3.5463945337644063
5% : -2.911939409384601
10% : -2.5936515282964665

```

Figure 4.13 - Test de stationnarité (Après différenciation)

### Interprétation :

- **P-value** : La p-value de 0.0038758123878013385 est bien inférieure à 0.05. Cela indique que nous pouvons rejeter l'hypothèse nulle de non-stationnarité.
- **ADF Statistic** : La statistique ADF de -3.717335920175868 est inférieure aux valeurs critiques à 1%, 5% et 10%. Cela signifie que la série temporelle est stationnaire après différenciation.
- **Conclusion** : Après la différenciation, la série temporelle des frais personnels de l'Agence 1 devient stationnaire. La p-value basse et la statistique ADF inférieure aux valeurs critiques montrent que les caractéristiques de tendance ou de saisonnalité ont été éliminées, rendant la série appropriée pour des modèles de prévision.

### 3.1.5 Autocorrélation et Autocorrélation Partielle

L'autocorrélation est une mesure statistique utilisée en analyse des séries temporelles pour quantifier la corrélation entre une observation et les observations précédentes dans la même série temporelle. Elle permet de détecter les motifs de corrélation et les dépendances entre les valeurs successives d'une variable.

Plus précisément, l'autocorrélation mesure la similarité entre les valeurs d'une série temporelle à différents décalages (lags). Elle peut prendre des valeurs allant de -1 à 1 : une valeur de -1 indique une corrélation négative parfaite, 0 indique aucune corrélation, et 1 indique une corrélation positive parfaite.

L'autocorrélation partielle, quant à elle, mesure la corrélation entre une observation et ses décalages (lags), tout en éliminant l'influence des décalages intermédiaires. En d'autres termes, elle permet d'examiner la corrélation directe entre deux observations après avoir pris en compte la corrélation indirecte via les autres observations intermédiaires.

L'autocorrélation et l'autocorrélation partielle sont souvent représentées graphiquement à l'aide de fonctions d'autocorrélation (ACF) et de fonctions d'autocorrélation partielle (PACF), respectivement. Ces graphiques permettent de visualiser les motifs de corrélation et d'identifier les décalages significatifs dans la série temporelle.

Ces mesures d'autocorrélation et d'autocorrélation partielle sont couramment utilisées pour l'analyse des séries temporelles, notamment dans la modélisation et la prévision. Elles fournissent des informations sur la structure de dépendance temporelle des données, ce qui permet de sélectionner des modèles appropriés et d'estimer les paramètres pour une meilleure prédiction des valeurs futures.

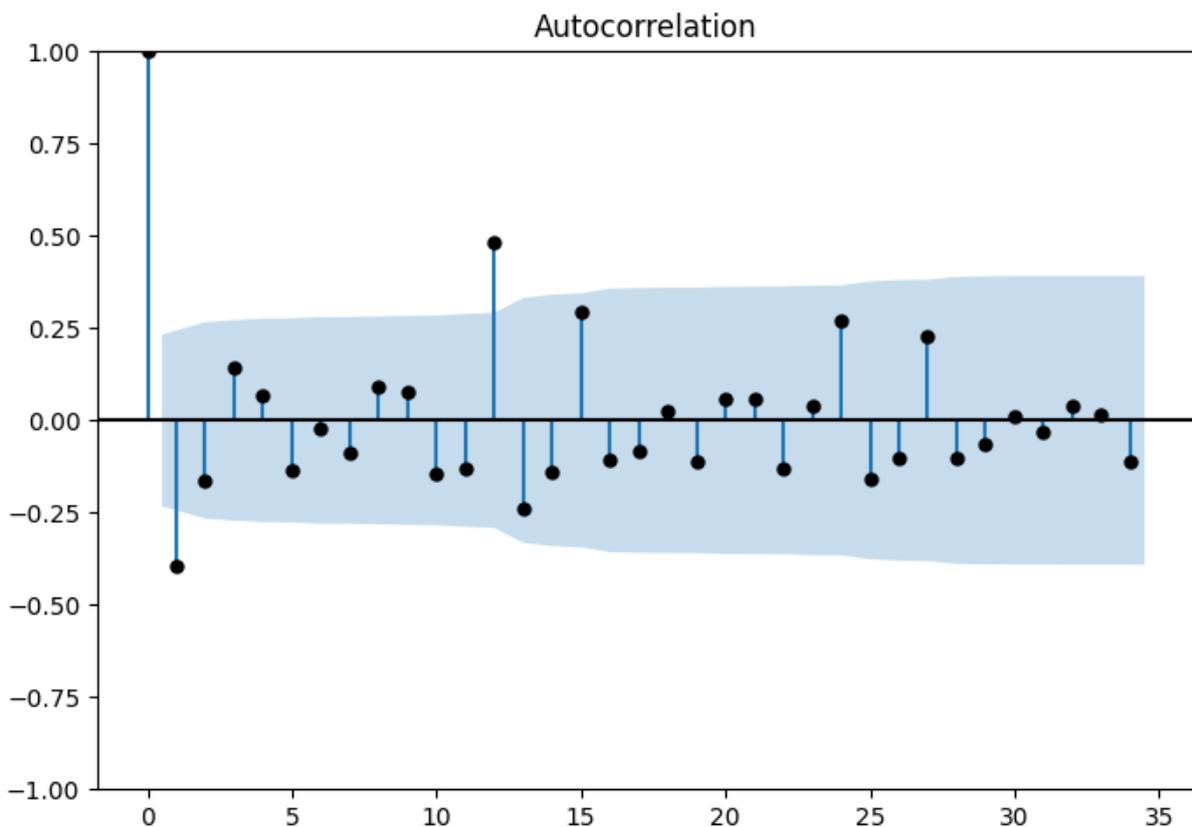


Figure 4.34 - Fonction d'Autocorrélation (ACF) des mouvements des FP de l'agence 1

### Interprétation :

- **Lags Significatifs :** Les premiers lags montrent des autocorrélations significatives. En particulier, le premier lag (lag 1) a une autocorrélation élevée, ce qui est typique dans une série temporelle où les valeurs sont fortement corrélées avec leurs valeurs immédiatement précédentes.
- **Diminution des Corrélations :** La plupart des autocorrélations après les premiers lags ne dépassent pas les limites de confiance (zone bleue), ce qui suggère que les dépendances se dissipent rapidement après quelques lags.

- **Saisonniers** : Il y a des lags significatifs à des intervalles réguliers (par exemple, autour des lags 10, 15, et 25), ce qui peut indiquer des effets saisonniers ou des cycles répétitifs dans les données.

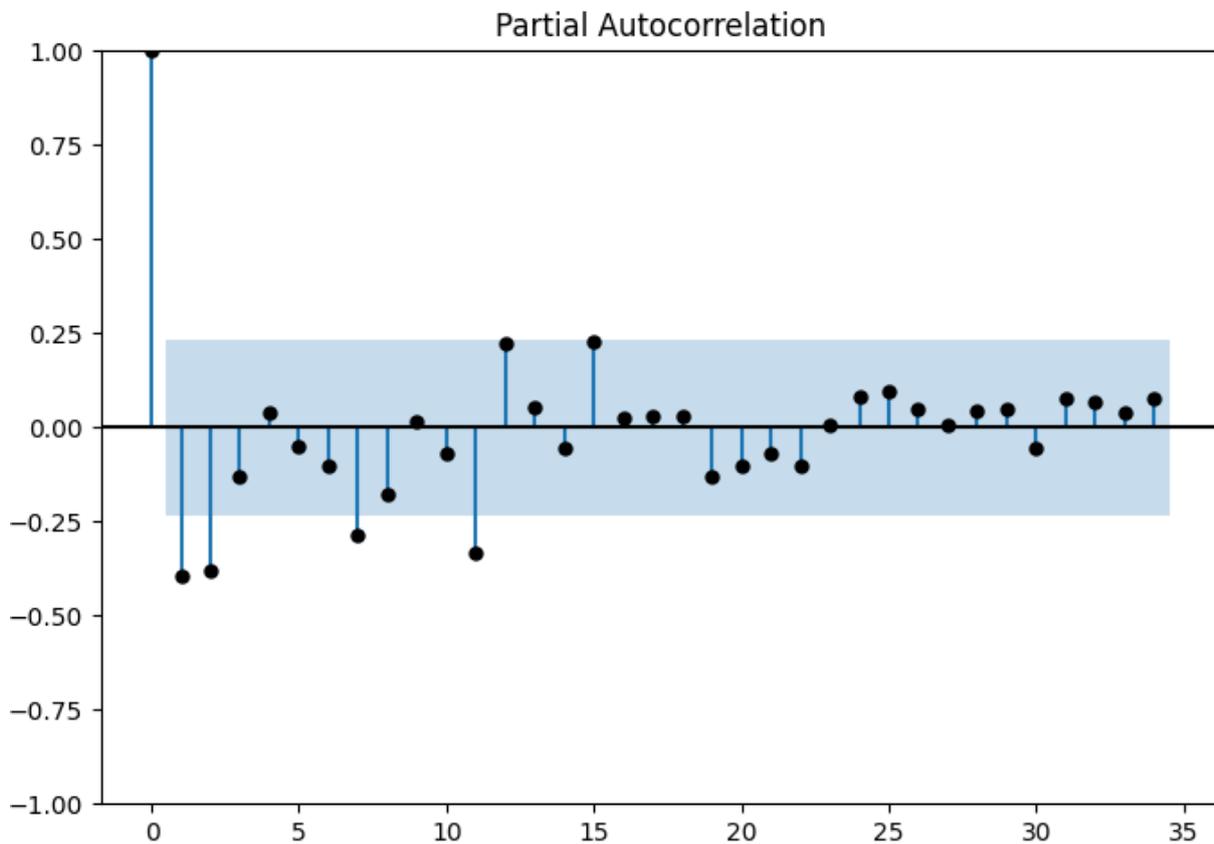


Figure 4.15 - Fonction de la partielle Autocorrélation (PACF) des mouvements des FP de l'agence 1

#### Interprétation :

- **Lags Significatifs** : Le premier lag (lag 1) montre une forte autocorrélation partielle, ce qui est attendu pour une série temporelle où les valeurs récentes influencent fortement les valeurs actuelles.
- **Dépendances à Court Terme** : Les lags suivants montrent des autocorrélations partielles significatives jusqu'au lag 5, puis diminuent. Cela suggère que les dépendances dans les données sont principalement à court terme.
- **Dissipation Rapide** : Après le lag 5, les autocorrélations partielles ne sont pas significatives, ce qui indique que les effets de dépendance se dissipent rapidement et ne persistent pas sur de nombreux lags.

### 3.2 Analyse et évaluation des résultats des modèles pré-entraînés :

#### 3.2.1 Résultats :

Les prévisions pour les Frais Personnels, basées sur les données de test allant du Janvier 2023 au Décembre 2023, ont été produites par les modèles RNN, LSTM et Auto-ARIMA de la manière suivante :

RNN:

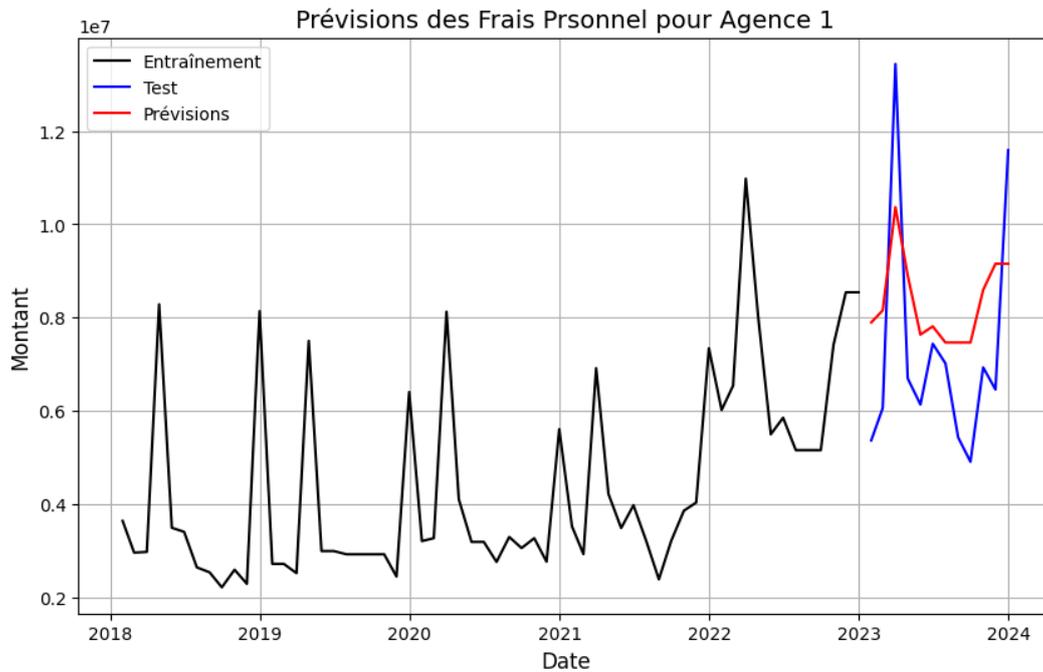


Figure 4.16 - Prévisions du modèle RNN pour les frais personnels de l'agence 1

LSTM:

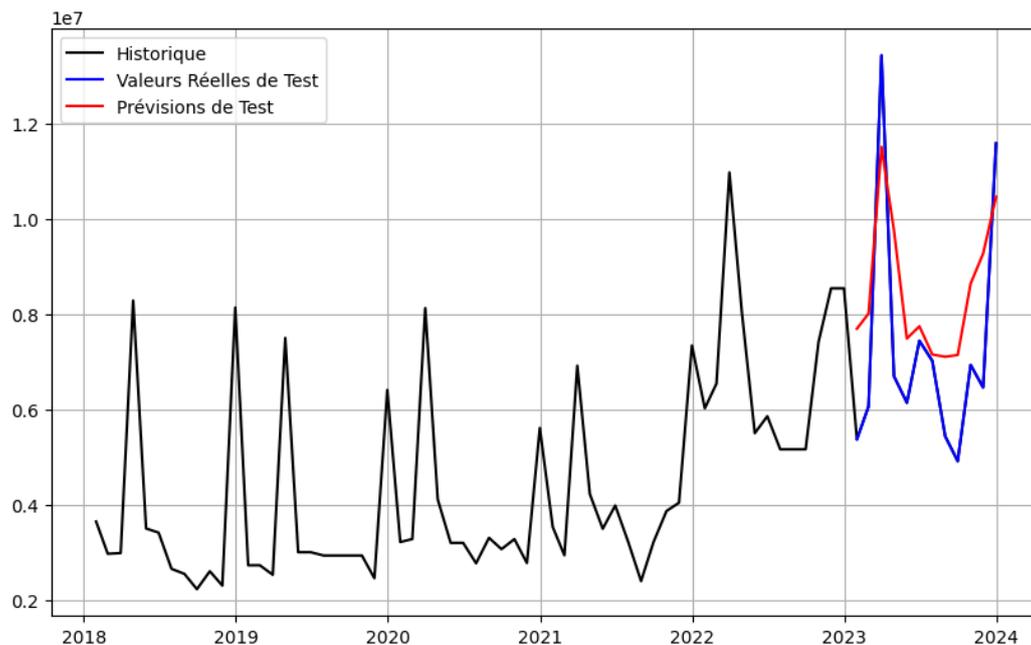


Figure 4.17 - Prévisions du modèle LSTM pour les frais personnels de l'agence 1

Auto-ARIMA:

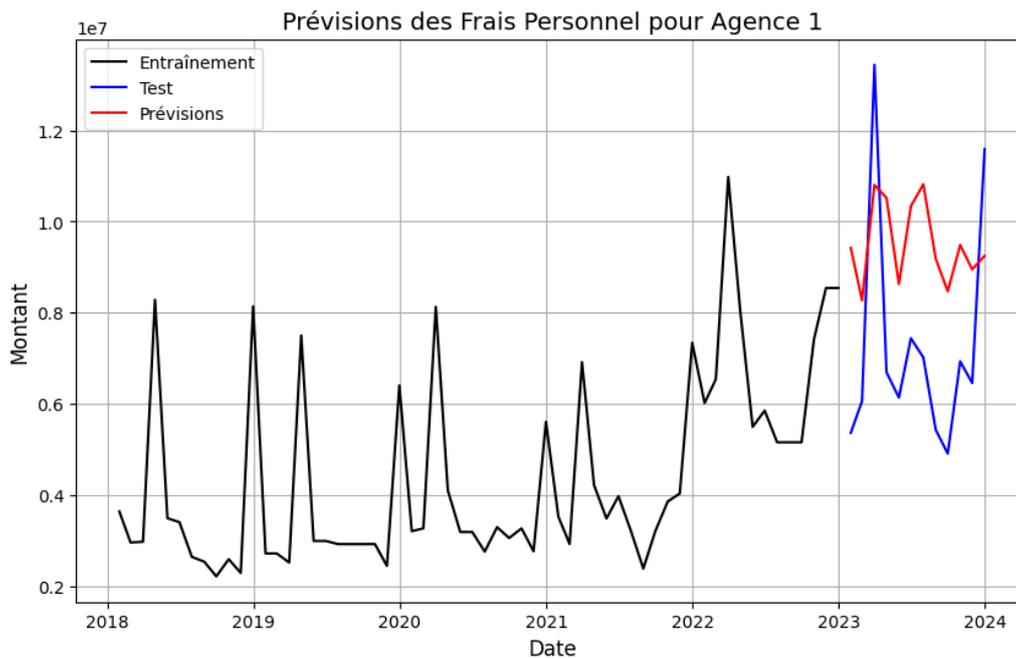


Figure 4.18 - Prévisions du modèle Auto-Arima pour les FP de l'agence 1

### 3.2.2 Évaluation :

Les résultats obtenus sont présentés ci-dessous :

	Modèle	RMSE	MSE	MAE
<b>Frais personnels Pour l'Agence 1</b>	Auto-ARIMA	0.3529	0.1245	0.1277
	LSTM	0.2144	0.0460	0.1084
	RNN	0.2244	0.0504	0.1271

Tableau 4.5 - Evaluation des modèles appliqués aux Frais personnels de l'agence 1

### Conclusion :

D'après les résultats d'évaluation obtenus en utilisant diverses métriques, le modèle **LSTM** s'est révélé être le plus performant pour la prédiction des frais personnels de l'agence 1.

*Le processus d'Analyse et d'Application des modèles est identique pour toutes les autres variables*

## 4 Application Utilisateur :

Pour permettre à tous les utilisateurs, technophiles ou non, de tirer pleinement parti de nos modèles, nous avons développé une application conviviale avec le framework Streamlit. Cette interface simplifie l'utilisation de nos modèles de prévision.

Dans cette section, nous détaillerons les fonctionnalités de notre application et l'outil de prévision des KPIs des agences bancaires de la SGA. L'objectif principal est de fournir une expérience utilisateur optimale, axée sur la convivialité et l'accessibilité.

### 4.1 Présentation de l'outil :

L'outil a été développé en utilisant le framework Streamlit, déployé dans l'environnement de développement intégré VSCode. Streamlit est une bibliothèque open-source en Python qui simplifie la création d'applications web interactives, spécialement conçues pour les projets en science des données et en apprentissage automatique.

Ce framework simplifie considérablement le processus de conception et de partage d'applications centrées sur les données en offrant une interface intuitive et des widgets intégrés pour concevoir des éléments interactifs. Il permet de créer des applications web directement à partir de scripts Python.

L'un des principaux avantages de Streamlit est sa capacité à mettre automatiquement à jour l'application en temps réel lorsque le code est modifié. Cette fonctionnalité est particulièrement utile pour itérer rapidement et explorer efficacement les données.

L'utilisation de VSCode pour déployer Streamlit a offert une expérience de développement fluide et productive. VSCode est reconnu pour ses fonctionnalités robustes telles que l'autocomplétion intelligente, le contrôle qualité du code et la facilité de débogage, ce qui a contribué à accroître notre efficacité tout au long du processus de développement.



Figure 4.19 - Structure de l'outil de prévision

## 4.2 Fonctionnalités

Nous allons maintenant présenter les différentes fonctionnalités de notre application :

**4.2.1 Choix de la date de fin de prévision :** L'utilisateur doit indiquer la date de fin de prévision, c'est-à-dire jusqu'à quel moment les prédictions seront générées. La date de début correspond à la dernière date disponible dans notre jeu de données initial, puisque notre modèle fonctionne avec des données séquentielles. En sélectionnant la date de fin, l'application utilisera notre modèle pour effectuer des prédictions à partir de la dernière date disponible jusqu'à la date de fin spécifiée.



Figure 4.20 - Choix de la date de fin de prévision

**4.2.2 Choix du KPI à prédire :** L'utilisateur peut choisir parmi les quatre grandes familles de KPI disponibles.



Figure 4.21 - Choix du KPI à prédire

**4.2.3 Choix de la sous-famille des KPIs à prédire :** Après avoir choisi la grande famille de KPI, l'utilisateur peut affiner sa sélection en choisissant la variable spécifique à prédire au sein de cette famille.

The screenshot shows the 'Outil de Performance et Rentabilité des Agences Bancaires' interface. At the top, there is a header with the logo 'C'EST VOUS L'AVENIR' and 'SOCIETE GENERALE ALGERIE'. Below the header, the title 'Outil de Performance et Rentabilité des Agences Bancaires' is displayed in a box. The interface includes several dropdown menus: 'Date de fin de prévision' (set to 'Janvier 2024'), 'Variable à prédire' (set to 'Frais Généraux'), and 'Choisissez la Sous-Variante de Frais Généraux' (set to 'Frais Personnel'). A red box highlights the 'Frais Personnel' option in the dropdown menu, which also lists 'Amortissements' and 'Autres Charges des frais'.

Figure 4.22 - Choix de la sous-famille des KPIs à prédire

**4.2.4 Choix du type d'agence pour la prédiction :** Pour faciliter la recherche, l'utilisateur peut d'abord filtrer par type d'agence. Ce premier niveau de filtre aide à affiner la sélection.

The screenshot shows the 'Outil de Performance et Rentabilité des Agences Bancaires' interface. At the top, there is a header with the logo 'C'EST VOUS L'AVENIR' and 'SOCIETE GENERALE ALGERIE'. Below the header, the title 'Outil de Performance et Rentabilité des Agences Bancaires' is displayed in a box. The interface includes several dropdown menus: 'Date de fin de prévision' (set to 'Janvier 2024'), 'Variable à prédire' (set to 'Frais Généraux'), 'Choisissez la Sous-Variante de Frais Généraux' (set to 'Frais Personnel'), and 'Type d'agence' (set to 'Retail'). A red box highlights the 'Retail' option in the dropdown menu, which also lists 'Mutualisée' and 'Corporate'.

Figure 4.23 - Choix du type d'agence pour la prédiction

**4.2.5 Choix de l'agence pour la prédiction :** L'utilisateur peut ensuite choisir l'agence spécifique, soit en sélectionnant le nom de l'agence soit l'ID de l'agence. Lorsque l'utilisateur sélectionne une agence par son nom, l'ID correspondant s'affiche automatiquement, et vice versa.

The screenshot shows a web interface titled "Outil de Performance et Rentabilité des Agences Bancaires". At the top, there is a logo for "C'EST VOUS L'AVENIR" and "SOCIETE GENERALE ALGERIE". Below the title, there are several dropdown menus. The first is "Date de fin de prévision" with "Janvier 2024" selected. Below it is a list of agency names: AMIROUCHE, CHERAGA KAOUICHE, DIDOUCHE MOURAD, KOUBA LES LAURIERS, BOUZAREAH, VAL D'HYDRA, DRARIA, DELLY BRAHIM, and AMIROUCHE (highlighted in red). To the right of this list is a dropdown menu labeled "Choisissez l'ID de l'agence de type Retail" with the value "2" selected.

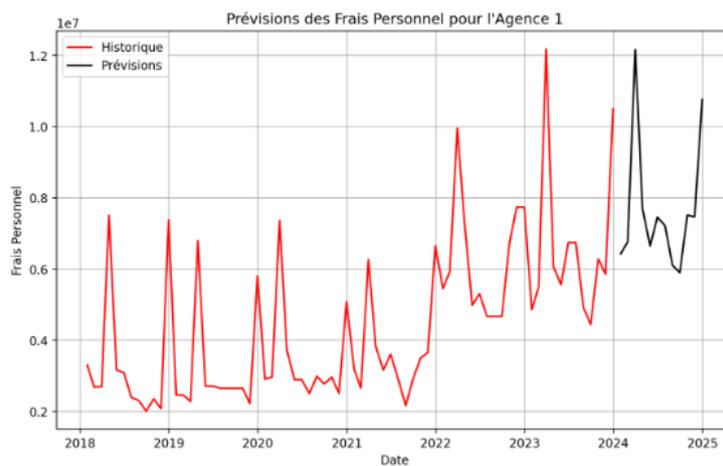
Figure 4.24 - Choix de l'agence pour la prédiction

**4.2.6 Génération des prévisions :** Après avoir sélectionné la date de fin de prévision, la grande famille des KPIs, la sous-famille, le type d'agence et l'agence, l'utilisateur doit cliquer sur le bouton "Prédire" pour générer les prévisions. Les résultats des prévisions sont affichés sur une nouvelle page intitulée "Résultats de la Prédiction". Sur cette page, les éléments suivants seront visibles :

- **Tableau des Prévisions :** Le tableau présente les résultats des prévisions avec deux colonnes :
  - **Date :** Affiche les mois et les années pour lesquels les prévisions sont générées, en commençant de janvier 2024 jusqu'au mois sélectionné par l'utilisateur.
  - **Prévisions de la variable sélectionnée pour l'Agence [ID] :** Affiche les valeurs prédites pour la variable souhaitée pour l'agence sélectionnée.
- **Graphique des Prévisions :** Un graphique linéaire montre les prévisions de la variable sélectionnée par rapport aux données historiques :
  - **Historique :** Les données historiques de la variable pour l'agence sont tracées en rouge.
  - **Prévisions :** Les prévisions générées sont tracées en noir.
- **Bouton Retour au Menu :** Un bouton "Retour au Menu" est situé en bas de la page pour permettre à l'utilisateur de revenir à la page de sélection des paramètres de

prévision. En cliquant sur ce bouton, l'utilisateur est redirigé vers la page principale pour effectuer de nouvelles sélections et générer d'autres prévisions.

Résultats de la Prédiction		
	Date	Prévisions des Frais Personnel pour l'Agence 1
0	Janvier 2024	6428297.04
1	Février 2024	6766532.39
2	Mars 2024	12161973.09
3	Avril 2024	7685482.35
4	Mai 2024	6649237.29
5	Juin 2024	7456780.37
6	Juillet 2024	7224448.42
7	Août 2024	6107604.15
8	Septembre 2024	5902346.49
9	Octobre 2024	7514243.51
10	Novembre 2024	7471143.80
11	Décembre 2024	10762807.32



[Retour au Menu](#)

Figure 4.25 - Prévisions générés et graphique des prédictions

**4.2.7 Accès aux dashboards Power BI :** L'application propose un bouton "Power BI". Lorsque l'utilisateur clique sur ce bouton, il est automatiquement redirigé vers les dashboards des résultats historiques des agences, créés lors de la première partie de notre projet. Cela permet d'accéder facilement à des visualisations détaillées des données historiques.

# Conclusion Générale

Notre projet de fin d'études avait pour ambition de concevoir un outil innovant de mesure de performance et de prédiction de la rentabilité des agences bancaires de la Société Générale Algérie. En adoptant une approche méthodique et en utilisant des techniques avancées de Business Intelligence, de data science et d'apprentissage automatique.

Dans un premier temps, nous avons exploré la direction Financière de la SGA, plus précisément le Contrôle de Gestion (CDG), afin d'acquérir une compréhension approfondie des besoins et des défis spécifiques à notre problématique. Cette phase de Business Understanding a été importante pour établir un diagnostic interne et externe, identifiant ainsi les dysfonctionnements et les opportunités d'amélioration potentielle de la gestion des agences. Ensuite, nous avons mené une étude bibliographique approfondie pour examiner l'utilisation de la BI, de l'IA, et des techniques statistiques dans la mesure de performances et la rentabilité des agences bancaires. Cette étude a permis de créer un pont entre les concepts théoriques et leur application pratique dans le secteur bancaire.

Sur le plan pratique, nous avons développé des tableaux de bord offrant une analyse approfondie des KPIs et des modèles de prévision robustes, s'appuyant sur le processus ETL et la méthodologie CRISP-DM. Ces outils ont servi de fondement à notre système de mesure de performance et de prédiction de la rentabilité, appliqué à diverses variables et agences de la SGA à l'échelle nationale. À l'issue de cette phase, nous avons sélectionné le modèle le plus performant et conçu une interface utilisateur conviviale pour faciliter l'adoption de notre outil par les gestionnaires d'agences et les responsables financiers.

En conclusion, ce projet a mis en évidence l'importance de l'intégration des technologies modernes dans la gestion bancaire. Les résultats obtenus montrent que l'utilisation de la Business Intelligence et du machine Learning peut considérablement améliorer la performance et la rentabilité des agences bancaires. L'adoption et l'extension de cet outil au sein de la Société Générale Algérie renforceront sa compétitivité et son efficacité opérationnelle. Dans une perspective d'amélioration continue de notre solution proposée, il est essentiel pour SGA de se focaliser principalement sur l'amélioration continue des modèles prédictifs, l'intégration de nouvelles sources de données et l'élargissement des fonctionnalités de l'application utilisateur. Cette démarche permettra d'obtenir une vision plus complète et nuancée de la rentabilité et des performances bancaires, renforçant ainsi la capacité de la banque à naviguer dans un marché de plus en plus compétitif et digitalisé.

## Bibliographie :

- [1] Larivière, Michel. *Analyse de la performance des banques*. 2017.
- [2] Rebah, Sid Ahmed. "Le système bancaire algérien : faits et défis". *Revue Algérienne des Sciences Économiques*. 2019.
- [3] Mekidèche, Mouloud. "Modernisation du secteur bancaire algérien : défis et opportunités". *Économie et Finances Algériennes*. 2021.
- [4] Amine, K. "Banking Reforms in Algeria: Progress and Prospects". *African Development Review*. 2020.
- [5] Ahmed, Nassim. *Performance et Services Bancaires en Algérie*. 2019.
- [6] Johnson, Emily. "The Market and Financial Performance of Société Générale Algeria". *Journal of Banking & Finance*. 2020.
- [7] Benamar, Khaled. *Performance et Rentabilité des Banques en Algérie*. 2019.
- [8] Leroy, Philippe. *Diagnostic Stratégique des Banques*. 2020.
- [9] Dubois, Jean-Luc. *CRISP-DM : Méthodologie de Data Mining*. 2018.
- [10] Smith, Robert. "Strategic Diagnosis in Banking : External and Internal Analysis". *Journal of Banking Strategy*. 2021.
- [11] Dupont, Anne. *La Finance Bancaire*. 2020.
- [12] Leclerc, Sophie. *KPIs dans le Secteur Bancaire*. 2018.
- [13] Johnson, Emily. "Banking Performance Metrics : Key Concepts and Measurement". *Journal of Banking & Finance*. 2021.
- [14] Dupuis, Marc. *Business Intelligence : Concepts et Applications*. 2019.
- [15] Lambert, Pierre. *Modèles de Données Bancaires*. 2020.
- [16] Martin, Jean-Luc. *Processus ETL : Techniques et Applications*. 2018.
- [17] Smith, Robert. "Data Science and Artificial Intelligence in Finance: Trends and Applications". *Journal of Financial Technology*. 2021.
- [18] Leclerc, Sophie. "Identification et utilisation des indicateurs clés de performance (KPI) dans le secteur bancaire". *KPI dans le Secteur Bancaire*. 2018.
- [19] Dupuis, Marc. *Collecte et Traitement des Données Bancaires*. 2019.
- [20] Dupont, Jean. *Machine Learning pour les Finances*. 2020.
- [21] Martin, Paul. *Analyse des Séries Temporelles avec Python*. 2019.
- [22] Leroy, Philippe. *Prédiction et Modélisation des Données Financières*. 2021.
- [23] Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019". *Applied Soft Computing*. 2020.
- [24] Pesaran, M. H., Pettenuzzo, D., & Timmermann, A. "Forecasting time series subject to multiple structural breaks". *The Review of Economic Studies*. 2006.
- [25] Lambert, Pierre. *Streamlit pour les Sciences de Données*. 2020.
- [26] Dupuis, Marc. *Développement d'Applications Web avec Python et VSCode*. 2021.
- [27] Smith, Robert. "Building Interactive Data Applications with Streamlit". *Journal of Data Science Applications*. 2021.

---

**Annexe A**

**Présentation des tableaux de bord**

---

Dans cette annexe, vous trouverez les dashboards supplémentaires que nous avons construits pour notre projet. Ces visualisations fournissent des insights précieux sur divers aspects de nos données,

Chaque dashboard a été conçu pour répondre à des questions spécifiques et fournir des visualisations claires et pertinentes pour les utilisateurs finaux.

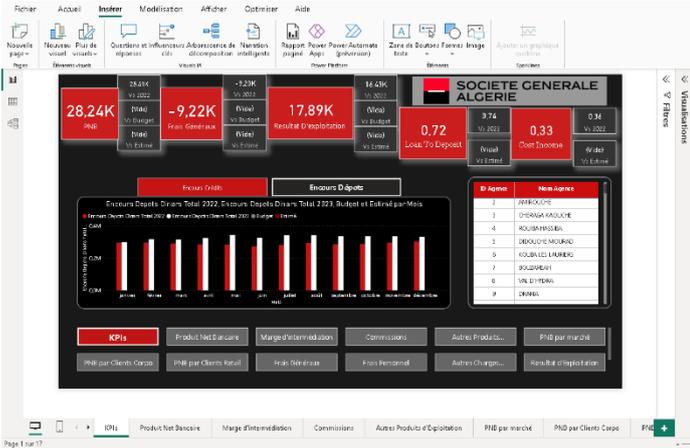


Figure A-1 - Dashboard 'Accueil'

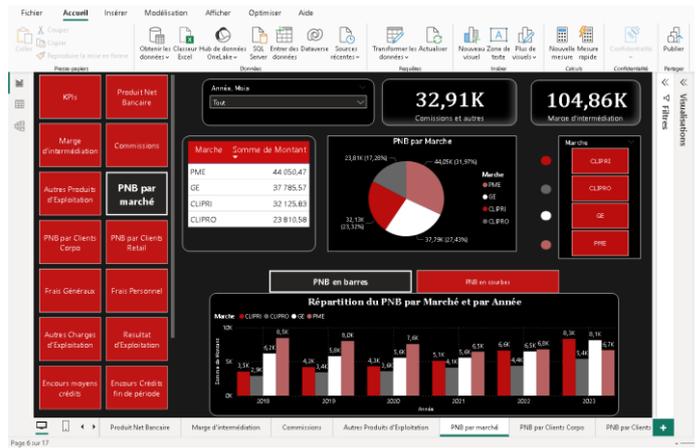


Figure A-2 - Dashboard 'PNB par marché'



Figure A-1 - Dashboard 'PNB par client corpo'

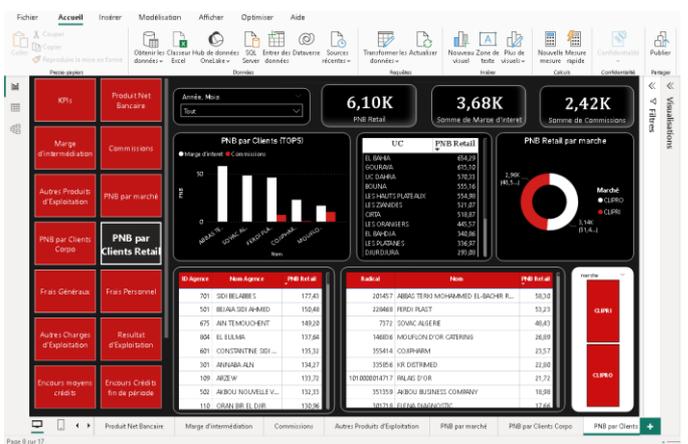


Figure A-2 - Dashboard 'PNB par client Retail'

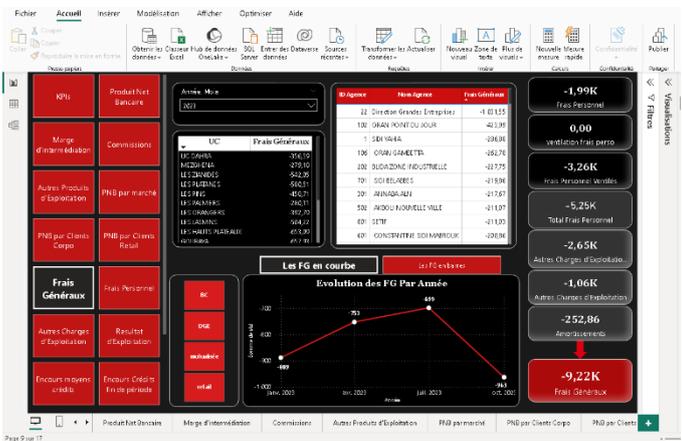


Figure A-5 - Dashboard 'Frais Généraux'

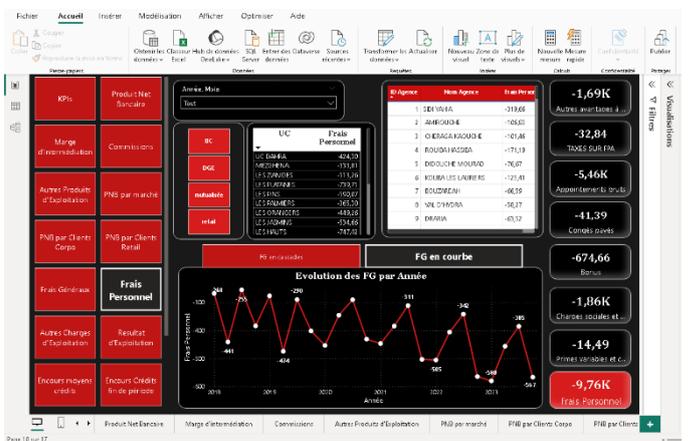


Figure A-6 - Dashboard 'Frais Personnel'

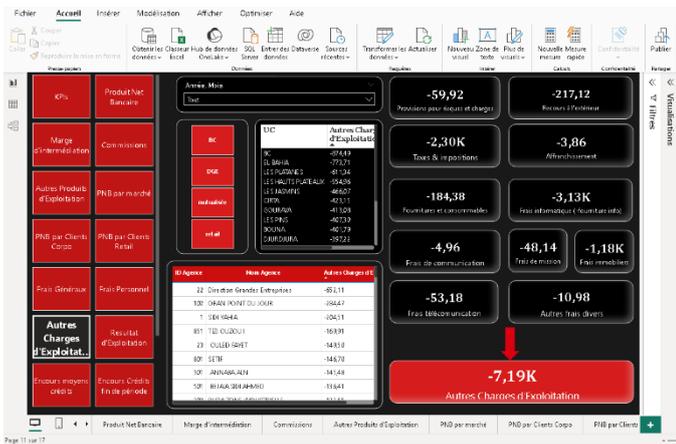


Figure A-7 - Dashboard 'Autres Charges d'Exploitation'

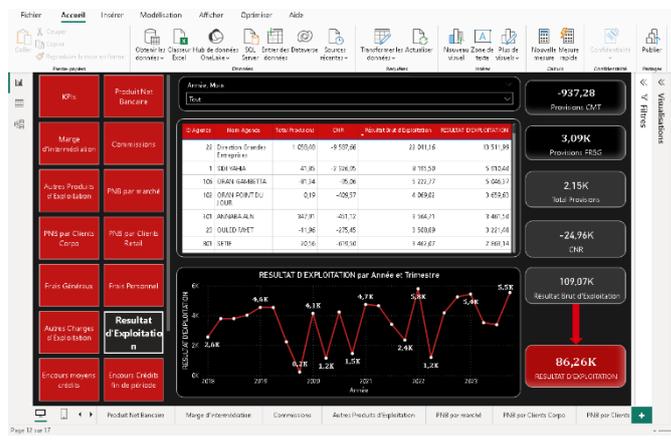


Figure A-8 - Dashboard 'Résultat d'Exploitation'

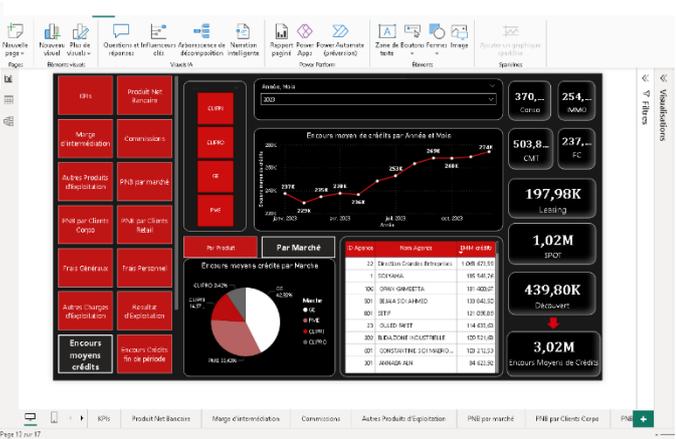


Figure A-3 - Dashboard 'Encours Moyen de Crédits'

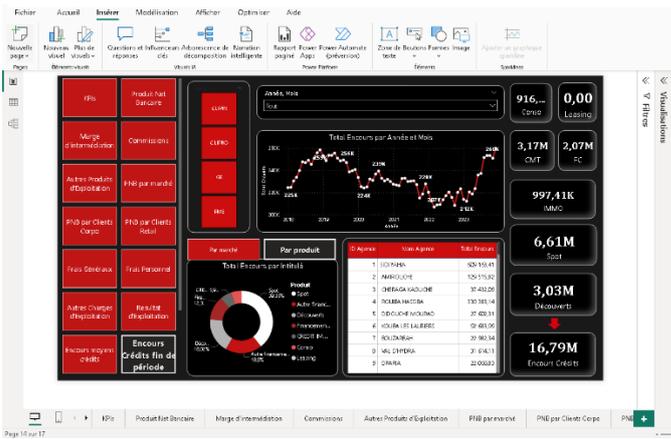


Figure A-4 - Dashboard 'Encours Crédits fin de période'

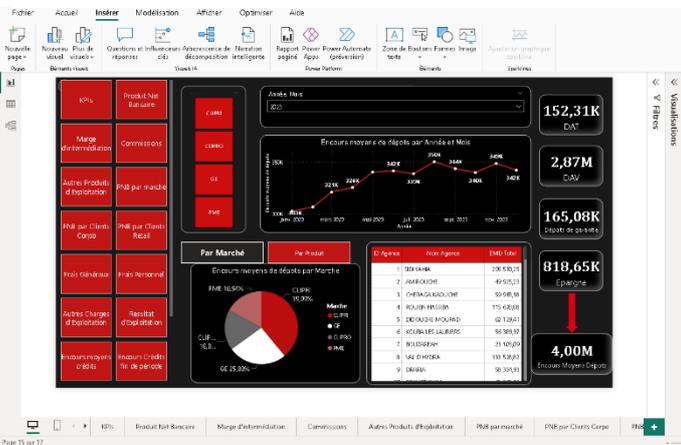


Figure A-11 - Dashboard 'Encours Moyen de Dépôts'

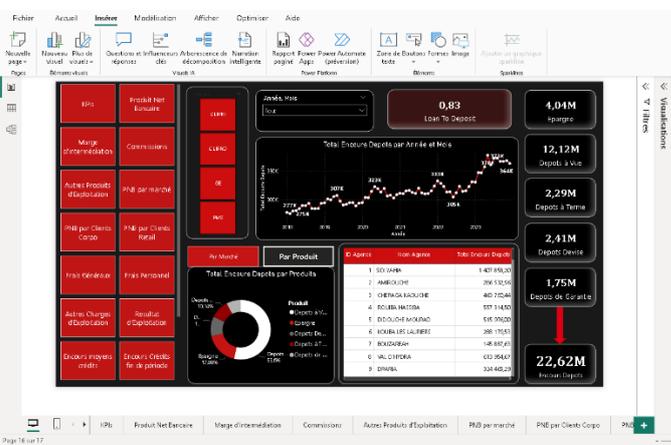


Figure A-12 - Dashboard 'Encours Dépôts fin de période'

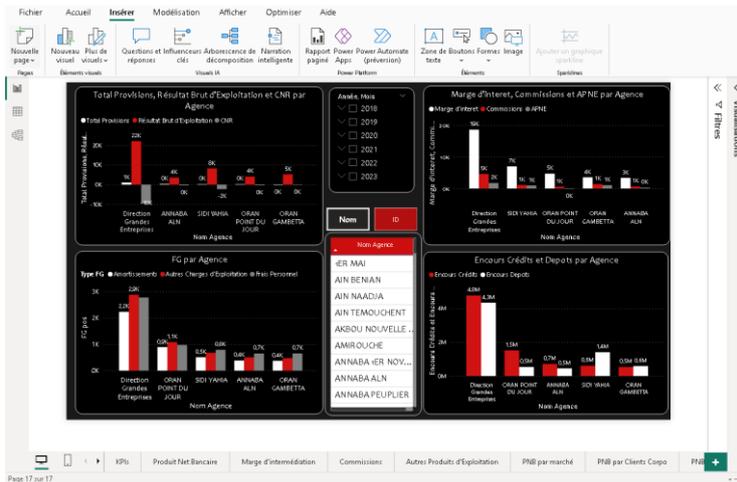


Figure A-5 - Dashboard 'Comparatif Agences'

---

**Annexe B**

**Présentation des outils utilisés**

---

## SAP BusinessObjects

Nous avons utilisé SAP BusinessObjects pour l'extraction de données. SAP BusinessObjects est une suite de logiciels d'intelligence d'affaires (Business Intelligence) proposée par SAP, une société allemande spécialisée dans les solutions logicielles pour les entreprises. SAP BusinessObjects offre une gamme complète d'outils et de fonctionnalités pour la collecte, l'analyse et la visualisation des données, permettant aux organisations de prendre des décisions éclairées basées sur des informations pertinentes et précises.



## DBeaver

Un autre outil utilisé pour l'extraction est DBeaver. DBeaver est un outil de gestion de bases de données universel et gratuit. Il fournit une interface graphique conviviale pour se connecter, naviguer, consulter et gérer différentes bases de données, qu'il s'agisse de bases de données relationnelles (comme MySQL, PostgreSQL, Oracle, SQL Server, etc.) ou de bases de données NoSQL (comme MongoDB, Cassandra, Redis, etc.).



## Python

Étant des data scientists, notre choix s'est porté sur le langage de programmation Python. Python est un langage de programmation interprété, polyvalent et de haut niveau. Il a été créé par Guido Van Rossum et a été publié pour la première fois en 1991. Depuis lors, Python est devenu l'un des langages de programmation les plus populaires au monde en raison de sa simplicité, de sa lisibilité et de sa flexibilité.



## Power BI

Pour la visualisation des données et les rapports interactifs, nous avons utilisé Power BI. Power BI est une suite d'outils d'analyse commerciale de Microsoft qui permet de visualiser vos données et de partager des insights dans toute votre organisation, ou de les intégrer dans une application ou un site web. Il fournit des tableaux de bord en temps réel et permet une collaboration efficace en équipe grâce à ses fonctionnalités de partage.



## **Visual Studio Code (VS Code)**

Pour la partie implémentation de l'outil, nous nous sommes penchés vers l'IDE Visual Studio Code (VS Code). En plus d'une interface utilisateur intuitive et conviviale qui facilite le développement et la navigation dans le code Python, VS Code permet l'analyse de code et contient un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciels de gestion de versions, et supporte le développement web avec des extensions pour divers frameworks, dont Django.

