

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



Société Générale Algérie

Département : Génie Industriel
Entreprise : Société Générale
Algérie

Mémoire de Projet de Fin d'Études

En vue de l'obtention du diplôme d'Ingénieur d'État en Génie Industriel
Option : Data science et intelligence artificielle

Étude de l'impact des encours bancaires sur le produit net bancaire :
Analyse, prévision temporelle et déploiement via une interface utilisateur.

Application : Société Générale Algérie

MEDJAOURI Insaf

ZERGOUG Lina

Sous la direction de Mme. BOUCHAFAA Bahia

Présenté et soutenu publiquement le

(25/06/2023)

Composition du jury :

Président	M. FOURAR-LAIDI Hakim	MCA	ENP
Examineur	M. ARKI Oussama	MCB	ENP
Promoteur	Mme. BOUCHAFAA Bahia	MCA	ENP

ENP 2023

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



Société Générale Algérie

Département : Génie Industriel
Entreprise : Société Générale
Algérie

Mémoire de Projet de Fin d'Études

En vue de l'obtention du diplôme d'Ingénieur d'État en Génie Industriel
Option : Data science et intelligence artificielle

Étude de l'impact des encours bancaires sur le produit net bancaire :
Analyse, prévision temporelle et déploiement via une interface utilisateur.

Application : Société Générale Algérie

MEDJAOURI Insaf

ZERGOUG Lina

Sous la direction de Mme. BOUCHAFEEA Bahia

Présenté et soutenu publiquement le

(25/06/2023)

Composition du jury :

Président	M. FOURAR-LAIDI Hakim	MCA	ENP
Examineur	M. ARKI Oussama	MCB	ENP
Promoteur	Mme. BOUCHAFEEA Bahia	MCA	ENP

ENP 2023

ملخص

هدف هذا العمل هو توقع تطور الأرصدة المصرفية للشركات بهدف الحصول على رؤية شاملة لتطور الدخل المصرفي الصافي في بنك سوسيتيه جنرال الجزائر. لتحقيق هذا الهدف، تم استخدام مختلف تقنيات تحليل البيانات ونماذج التنبؤ.

يبدأ الأسلوب المتبع بتحديد المشكلة من منظور العمل والتقني، تليها جمع البيانات وإعدادها لاستخدامها في نماذج التنبؤ. في النهاية، يتم نشر الحل على شكل واجهة مستخدم.

تسمح حلولنا بقياس فعالية استراتيجيات الأعمال على المدى الطويل وتمكين اتخاذ قرارات مدروسة ومواجهة تحديات السوق المالية.

كلمات مفتاحية: أرصدة الشركات، الدخل المصرفي الصافي، نماذج التنبؤ، تحليل البيانات، واجهة المستخدم.

Abstract

This work aims to predict the evolution of corporate banking balances in order to obtain a comprehensive view of the evolution of the Net Banking Income (NBI) within Société Générale Algérie. To achieve this objective, various data analysis techniques and prediction models have been used.

The implemented approach starts with defining the problem from both a business and technical perspective, followed by the collection and preparation of data for use in prediction models. Finally, the solution is deployed in the form of a user interface.

Our solution enables the measurement of long-term effectiveness of business strategies, facilitates informed decision-making and addresses the challenges of the financial market.

Keywords : Corporate balances, NBI, Prediction models, Data analysis, User interface

Résumé

Le présent travail vise à prédire l'évolution des encours bancaires corporate dans le but d'obtenir une vision globale de l'évolution du Produit Net Bancaire (PNB) au sein de la Société Générale Algérie. Pour atteindre cet objectif, différentes techniques d'analyse de données et modèles de prédiction ont été utilisés.

La démarche mise en place commence par la définition du problème à la fois d'un point de vue métier et technique, suivie de la collecte et de la préparation des données pour les utiliser dans les modèles de prédiction. Enfin, la solution est déployée sous la forme d'une interface utilisateur.

Notre solution permet de mesurer l'efficacité des stratégies commerciales à long terme, de planifier en prenant des décisions éclairées et de faire face aux enjeux du marché financier.

Mot clés : Encours corporate, PNB, Modèles de prédiction, Analyse des données, Interface utilisateur.

Dédicace

Louange à Dieu tout-puissant qui m'a permis de vivre ce jour tant attendu.

Je souhaite dédier ce travail à

Ma mère bien-aimée, qui est une source inépuisable de tendresse et de patience, et qui a fait d'innombrables sacrifices. Je tiens à te remercier pour ton amour, ta générosité et ton soutien. Que Dieu tout-puissant te protège et t'accorde une longue vie remplie de bonheur, de santé et de prospérité.

À mon cher père, qui m'a transmis l'amour pour ce métier d'ingénieur et m'a enseigné le sens du travail et de la responsabilité. Je te suis reconnaissante pour tes sacrifices, ta compréhension et ton soutien. Que Dieu te préserve pour nous et t'accorde santé et bonheur.

À mes deux frères, ma plus grande fierté ! Je leur souhaite une longue vie pleine de succès et de santé. Que Dieu les protège, inch'Allah.

Je souhaite également rendre hommage à ma grand-mère paternelle, ainsi qu'à la mémoire de mes grands-parents maternels et de mon grand-père paternel. J'imagine quelle serait votre joie aujourd'hui. Vous pouvez être fière de votre petite-fille. Que Dieu vous accueille dans son vaste paradis.

À toutes mes cousines, mes cousins, mes oncles et mes tantes, ainsi qu'à mes camarades et amis que j'estime, en particulier à mon binôme Lyna. J'ai eu la chance de réaliser ce projet de fin d'études dans les meilleures conditions à tes côtés, je t'aime ! À mes meilleurs amis Yasso, Safa, Imad, Manou " Le parisien ", Mima " meilleure pâtissière ", Abdou, Omar à tous les moments que nous avons partagés ensemble, ainsi qu'à Yousra " ma youss " et Sarah " sarssoura " , avec qui j'ai passé d'agréables moments. À mon groupe de " khbiichh " avec mon cher ami Anes " metalii " et Lyna " ma Lynouch d'amour " Je vous souhaite une grande réussite dans vos vies et que nos liens d'amitié se perpétuent éternellement.

Je tiens à exprimer ma profonde gratitude à Madame Boucahafa, mon encadrante de projet. Votre soutien, vos conseils avisés et votre expertise ont été d'une valeur inestimable tout au long de ce projet. Je suis profondément reconnaissante de vous avoir comme encadrante de projet,

Insaf

Dédicace

Je dédie ce travail,

À mes parents, mes plus grands soutiens et sources d'inspiration. Votre amour inconditionnel, vos encouragements constants et vos sacrifices ont été les piliers de ma réussite. Je vous suis profondément reconnaissant pour tout ce que vous avez fait pour moi, je vous aime.

À mes deux merveilleux petits frères, mes complices d'aventures et mes sources de bonheur. Même à leur jeune âge, leur présence joyeuse et leurs mots d'encouragement m'ont donné la force.

à ma chère grand-mère "Amamus", une âme lumineuse qui était une véritable source de joie de vivre et de rires, illuminant chaque instant de nos vies. Ce travail est un témoignage de l'amour et de l'inspiration qu'elle a apportés à ma vie.

À ma grand-mère bien-aimée et à mon grand-père adoré, vous êtes les piliers de notre famille et des exemples extraordinaires à suivre, que Dieu vous protège.

À ma meilleure amie Fifi, qui occupe une place spéciale dans mon cœur. Sa présence dans ma vie est d'une valeur inestimable et je l'aime énormément. Notre amitié est une source de bonheur et de réconfort.

À mon binôme, avec qui j'ai eu le privilège de travailler. Je ne regrette pas une seule seconde d'avoir partagé cette expérience avec elle. Travailler à ses côtés était comme un moment de repos, car sa présence et son soutien ont rendu le travail si agréable. Je l'aime tellement.

À mes cousins et cousines, tantes et oncles, et à mes chers amis : Maria, Kahina , Yousra "ra3ed" , sarsouraa et sans oublier aness metalii wa layaliiiii ra3diya avec rokho.

À mon encadrante exceptionnelle, Mme Bouchafaa, je dédie ce travail avec une profonde gratitude. Votre présence et votre expertise ont été une véritable bénédiction dans la réalisation de ce projet.

Lina

Remerciements

Nous exprimons nos sincères remerciements à Mme. Bouchafaa pour ses précieux conseils, sa disponibilité et son accompagnement rigoureux et constant, qui ont été déterminants pour la réalisation de notre travail de fin d'études.

Nous tenons également à adresser nos chaleureux remerciements au président du jury, M. Fourar, ainsi qu'à l'examineur, M. Arki, pour avoir consacré leur temps et apporté leur expertise lors de l'évaluation de ce projet.

Nous sommes profondément reconnaissants envers nos promotrices au sein de SGA, Mme. BOUKELLAL Cylia du département stratégie performance corporate et Shakra du département SIOP. Leur aide précieuse dans le développement de la problématique, la recherche de solutions et leur accompagnement tout au long de notre stage à la Société Générale ont été d'une valeur inestimable.

Nous souhaitons également exprimer notre gratitude envers l'ensemble de nos enseignants du département du Génie Industriel, qui ont partagé avec générosité leur savoir-faire et leurs connaissances tout au long de notre formation. Nous tenons notamment à remercier le trio formé par M. Fourar, M. Arki et M. Nahili pour leur soutien indéfectible ainsi M.Zouaghi qui a la capacité d'encourager et inspirer ses étudiants.

Enfin, nous tenons à remercier toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce projet. Votre appui et votre implication ont été essentiels pour notre réussite. Nous vous sommes infiniment reconnaissants.

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

Introduction générale :.....	14
État des lieux	17
1.1 Le marché financier	17
1.2 La banque	18
1.3 L'histoire d'intégration des banques multinationales en Algérie	18
1.4 Groupe Société Générale	19
1.4.1 Les missions de SG	19
1.4.2 Les pôles de métiers de SG	19
1.4.3 Les valeurs de SG.....	19
1.5 Présentation de société générale Algérie	20
1.5.1 Segmentation des clients.....	21
1.5.2 Les projets de la SGA	21
1.5.3 L'organigramme de la SGA	22
1.6 Les départements clés concernés par la mission à la Société Générale Algérie.....	23
1.6.1 Département stratégie et performance corporate.....	23
1.6.2 La direction des Systèmes d'informations, Organisations et Projets	23
1.6.3 Direction des finances	24
1.6.4 Le top management.....	24
1.6.5 Analyse des interactions entre les départements concernés.....	24
La performance bancaire et la décortication de l'indicateur du Produit Net Bancaire (PNB)	27
2.1 La performance bancaire.....	27
2.2 Les facteurs influant sur la performance bancaire	28
2.2.1 Les facteurs internes	28
2.2.2 Les facteurs externes	30
2.3 Exploration des mesures de rentabilité dans l'évaluation de la performance bancaire	31
2.3.1 De la rentabilité à la rentabilité bancaire.....	31
2.3.2 Les instruments de mesure de la rentabilité.....	32
2.3.3 Les ratios déterminants de la rentabilité : liquidité, solvabilité et risque	36

2.4	Décryptage du PNB : une approche essentielle pour évaluer la performance bancaire	38
2.4.1	Les sources clés de revenus pour les banques : Les moteurs de croissance du PNB.....	38
2.3.2	Au-delà des indicateurs financiers : le PNB comme mesure holistique de la performance bancaire	39
2.4	Problématique.....	40
	Outils et méthodes pour l'analyse et la prévision des séries temporelles	43
3.1	Investigation statistique de la relation significative entre les variables	43
3.1.1	La régression multiple.....	43
3.1.2	Le coefficient standardisé	44
3.1.3	La corrélation.....	45
3.1.4	Détection de la causalité temporelle : Évaluation par le test de Granger	45
3.2	Les séries temporelles	46
3.2.1	Définition.....	46
3.2.2	La notion de stationnarité	46
3.2.3	La décomposition d'une série temporelle	47
3.2.4	Les transformations de séries temporelles.....	49
3.2.5	L'étude des séries temporelles.....	49
3.3	La prévision des données.....	50
3.3.1	Définition.....	50
3.3.2	Les méthodes de prévision.....	50
3.4	Exploration, Analyse et Prévision des Séries Temporelles	51
3.4.1	Exploration et analyse des données : Synergie entre SQL et Power BI.....	51
3.4.2	Pré-étude de la nature de la série temporelle pour le choix des modèles	52
3.4.3	Prévision des séries temporelles : Modèles statistiques, apprentissage automatique et l'approche Prophet.....	52
3.4.4	Evaluation des modèles.....	55
3.5	Interface utilisateur	56
3.5.1	Définition.....	56
3.5.2	Le backend et le frontend.....	56
3.5.3	La liaison entre le backend et le frontend.....	56
	Conception de la solution.....	59
4.1	La collecte de données.....	59

4.2	Exploration de données.....	59
4.3	Étude approfondie de la relation entre les encours et le PNB.....	61
4.3.1	Régression multiple et coefficient standardisé.....	62
4.3.2	La causalité.....	64
4.4	La collecte et la consolidation des données.....	65
4.5	Exploration et compréhension de données.....	67
4.5.1	Décomposition de la série temporelle.....	69
4.5.2	Evaluation de la stationnarité.....	70
4.6	Application des modèles prédictifs.....	71
4.6.1	ARIMA.....	71
4.6.2	SARIMA.....	74
4.6.3	XGBOOST.....	75
4.6.4	Fb prophet.....	76
4.7	Évaluation des erreurs.....	78
4.8	Étude complémentaire.....	78
4.9	Déploiement de la solution.....	80
4.9.1	Implémentation de l'interface de prédiction.....	80
	Conclusion Générale.....	83
	Références.....	86
	Annexe.....	90

Liste des tableaux

4.1 – CARACTERISTIQUES DES VARIABLES : NOMS DE COLONNES ET TYPES DE VARIABLES....	62
4.2 – RESULTATS DE L'ANALYSE STATISTIQUE POUR LES ENCOURS ET L'IMPORT 2022	64
4.3 – RESULTATS DU TEST DE STATIONNARITE (ADF) POUR CHAQUE TRANSFORMATION.....	71
4.4 – ÉVALUATION DE LA QUALITE DE PREDICTION DE CHAQUE MODELE	78

Liste des figures

1.1 Croissance du marché financier entre 2022-2023 par région	17
1.2 Création monétaire	18
1.3 Un profil géographique diversifié du groupe SG	20
1.4 Réseau de la Société Générale Algérie	21
1.5 L'organigramme de la banque SGA	22
1.6 Illustration de la coordination entre les départements impliqués	25
2.1 Les composantes de la performance	28
2.2 Schéma récapitulatif du calcul du Produit Net Bancaire (PNB)	38
2.3 Tableaux de bord de PNB provenant de différents rapports financiers de banques	40
3.1 Les composantes d'une série chronologique	48
3.2 Exemple de schéma additif	48
3.3 – Exemple de schéma multiplicatif	48
3.4 Diagramme de flux du processus SQL	51
3.5 Diagramme d'architecture de Power BI	52
3.6 Un schéma qui illustre les avantages du modèle XGboost	54
3.7 Un schéma qui illustre la relation entre le backend et le frontend	57
4.1 Requête SQL pour l'extraction des données de PNB, encours et secteurs d'activité des clients Corporate pour l'année 2022	59
4.2 Page d'analyse du PNB - rapport Power BI	60
4.3 Page d'analyse des différentes variables par secteur d'activité - rapport Power BI	61
4.4 la matrice de corrélation	61
4.5 Résultat de la régression ols	63
4.6 Capture des résultats : Impact relatif des variables indépendantes sur le PNB (coefficients standardisés)	64
4.7 Requête SQL pour l'extraction des encours des clients corporate pour la période janvier 2012 - mars 2023	66
4.8 Requête SQL pour l'extraction et l'agrégation mensuelle des encours clients corporate ...	66
4.9 Requête SQL pour la création d'une base de données consolidée des encours mensuels clients corporate (janvier 2012 - mars 2023)	66
4.10 Évolution de l'engagement à partir de 2012 jusqu'à 2023	67
4.11 Évolution de l'engagement total par année	67
4.12 Tableau de bord interactif Power BI : Évolution de la série temporelle par année	68
4.13 Page du tableau de bord : Récapitulatif des années de baisse des encours	68
4.14 la décomposition de la série temporelle	69
4.15 Graphique de PACF	72
4.16 Graphique de la fonction d'autocorrélation (ACF)	72
4.17 Comparaison des données différencié, de l'ensemble d'entraînement et de l'ensemble de test	72
4.18 Visualisation des données initiales, de l'ensemble de test et des prédictions futures par le modèle ARIMA	73
4.19 Motifs de saisonnalité dans les données.....	74

4.20 Comparaison des données initiale, de l'ensemble d'entraînement et de l'ensemble de test à partir du janvier 2020	75
4.21 Visualisation des données initiales, de l'ensemble de test et des prédictions futures avec le modèle SARIMA	75
4.22 Comparaison des données initiaux et de l'ensemble de test	76
4.23 Visualisation des données initiales, de l'ensemble de test et des prédictions futures avec le modèles XGboost	76
4.24 Graphique interactif qui permet de visualiser les prédictions et les données historiques.	77
4.25 la décomposition des données en tendance	77
4.26 Visualisation des données initiales et des prédictions futures avec le modèles Fb prophet	79
4.27 la décomposition des données en tendance	79
4.28 Interface de visualisation des valeurs prédites en fonction de la date sélectionnée.....	81

Liste des abréviations

ACF Autocorrelation Function
ADF Augmented Dickey Fuller
API Application Programming Interface
ARIMA Autoregressive Integrated Moving Average
CMR Le coût moyen des ressources
CRCM Centre de Relation Clients Multimédia
CSS Cascading Style Sheets
DDL degrés de liberté
ESS Explained Sum of Squares
GE Grande entreprise
HTML Hyper Text Markup Language.
JSON JavaScript Object Notation
KPI Key Indicator Performance
MAPE mean absolute percentage error
NRMSE normalized root mean squared error
PACF Partial autocorrelation function
PIB Produit intérieur brut
PME petite moyenne entreprise
PNB produit net bancaire
RBE Résultat Brut d'Exploitation
RE Résultat d'Exploitation
RME Rendement Moyen des Emplois
RN Résultat net
ROA ou Return on assets
ROE Return on equity
RSS Residual Sum of Squares
SQL Structured Query Language
SG Société Générale
SGA Société Générale Algérie
SIOP Systèmes d'informations, Organisations et Projets
TSS Total Sum of Squares
WMAPE weighted mean absolute percentage error

Introduction Générale

Introduction générale :

Dans le paysage économique d'un pays, le secteur financier et l'environnement bancaire occupent une place prépondérante en jouant un rôle essentiel dans le développement économique et la stabilité financière. Les institutions financières, telles que la Société Générale Algérie, jouent un rôle clé en facilitant les échanges économiques, en offrant des services financiers indispensables et en soutenant le développement des entreprises et des particuliers.

La Société Générale Algérie, en tant que filiale d'un groupe bancaire international de premier plan, bénéficie d'une expertise et d'une expérience solides dans la prestation de services bancaires et financiers de qualité. Elle s'engage à répondre aux besoins financiers variés de sa clientèle en proposant des produits et des solutions adaptés.

Au sein de la Société Générale Algérie, le département de stratégie et performance corporate joue un rôle crucial dans l'analyse des données et l'élaboration de stratégies visant à optimiser les performances financières de la banque.

Durant notre stage au sein de ce département, nous avons été confrontés à une problématique majeure : celui d'obtenir une perspective globale du Produit Net Bancaire (PNB) corporate, qui constitue le principal indicateur des revenus générés par les activités de SGA, dans les prochaines années. Cependant, nous avons été confrontés à une contrainte significative, à savoir le manque de données disponibles. Parmi les différentes variables pouvant influencer le PNB, les encours bancaires ont émergé comme étant l'élément prédominant. Il est donc essentiel de mener une évaluation approfondie de l'impact des variations des encours bancaires sur le PNB. Ces encours, qui reflètent les montants prêtés par la banque à sa clientèle, revêtent une importance cruciale en tant qu'indicateur clé de l'activité financière de SGA. De plus, la capacité à prédire avec précision les encours bancaires revêt une importance primordiale pour garantir une planification financière rigoureuse, une gestion des risques appropriée et une prise de décisions stratégiques éclairées.

Ainsi, la question qui se pose est la suivante : Dans quelle mesure les encours bancaires ont-ils un impact significatif sur le Produit Net Bancaire (PNB) de la Société Générale Algérie, et comment cette relation doit-elle être prise en compte pour assurer une gestion financière optimale ? De plus, comment pouvons-nous effectuer une prédiction précise des encours bancaires et anticiper les tendances futures en utilisant des modèles de prévision avancés et des techniques d'analyse de données ?

Afin de répondre à cette problématique, nous avons exploré différentes solutions potentielles, en mettant l'accent sur l'utilisation de modèles de prévision basés sur les séries temporelles telles que l'ARIMA, le SARIMA, le XGBoost et le FB Prophet, nous avons cherché à déterminer quel modèle offre les meilleures performances de prédiction des encours bancaires, après avoir effectué une analyse approfondie et rigoureuse de la corrélation et de la causalité entre les encours bancaires et PNB.

Ce travail est composé de quatre chapitres :

1- Le premier chapitre de cette étude se focalise sur le marché financier et domaine bancaire en Algérie, en mettant l'accent sur l'histoire de l'intégration des banques multinationales dans le pays. Plus précisément, il examine le cas de Société Générale Algérie, une institution bancaire de renom et explore les départements clés qui y sont impliqués. L'objectif de ce chapitre est de comprendre la relation entre ces départements et leur contribution à la mission globale de Société Générale Algérie.

2- Le deuxième chapitre s'intéresse à la performance bancaire, en se concentrant sur les facteurs qui influent sur cette performance. L'objectif principal de ce chapitre est de fournir une compréhension approfondie de la performance bancaire et d'explorer les différentes mesures et méthodes utilisées pour évaluer cette performance en se focalisant spécifiquement sur le Produit Net Bancaire (PNB).

3-Le troisième chapitre de ce rapport mettra en évidence les concepts clés nécessaires pour évaluer l'impact significatif des encours bancaires sur le PNB. Il présentera également divers modèles de prédiction et métriques afin de vérifier la validité des résultats obtenus, en commençant par une extraction et visualisation des séries temporelles. De plus, ce chapitre sera consacré à la présentation des principes fondamentaux liés à la réalisation d'une interface utilisateur, tant du côté backend que frontend.

4- Enfin, dans le quatrième et dernier chapitre, nous mettrons en œuvre la solution en détaillant les différentes étapes nécessaires. Cela comprendra la collecte et l'exploration des données, la visualisation en temps réel de ces données, la présentation des résultats de l'étude approfondie de la relation entre les encours et le PNB corporate, ainsi que la création et l'évaluation des modèles de prédiction. Enfin, nous aborderons le déploiement de la solution.

Le fruit de notre travail se matérialise par la création d'un tableau de bord sur Power BI, permettant l'analyse du PNB corporate en fonction de diverses variables. Ce tableau de bord offre une visualisation claire et interactive des données, facilitant ainsi leur analyse et la prise de décisions éclairées. Parallèlement, nous avons développé une interface de prédiction des encours bancaires corporate qui permet aux utilisateurs d'obtenir des estimations précises des valeurs des encours à des dates futures. Cette interface offre une perspective anticipative précieuse pour la planification et la gestion des activités bancaires.

Chapitre 1 : État des lieux

Chapitre 1

État des lieux

Afin de cerner de manière approfondie la problématique, nous procéderons tout d'abord à une présentation de l'état des lieux en mettant en exergue le positionnement de la banque dans le marché financier. Ensuite, nous aborderons la Société Générale et son rôle dans l'économie mondiale. Nous présenterons également la Société Générale Algérie, en mettant en avant ses valeurs, ses services et ses projets. Enfin, nous mettrons en lumière les départements clés qui ont contribué à la réalisation de ce projet.

1.1 Le marché financier

L'histoire des marchés financiers remonte à plusieurs siècles, remontant à la création des premières bourses de commerce dans les villes italiennes de Florence et de Venise au 14ème siècle. Néanmoins, le marché financier est devenu de plus en plus sophistiqué qu'au 20ème siècle avec l'essor de la technologie et de l'informatique. Les ordinateurs ont permis de traiter les transactions beaucoup plus rapidement et les investisseurs ont pu accéder aux marchés financiers du monde entier grâce à Internet.

De nos jours, le marché financier est un marché mondial qui est en constante évolution en réponse aux évolutions économiques, technologiques et politiques où les investisseurs peuvent acheter et vendre des titres financiers tels que des actions, des obligations, des devises, des options et des contrats à terme. Le marché financier est souvent utilisé pour collecter des capitaux pour les entreprises, les gouvernements et d'autres organisations. Il est souvent considéré comme un indicateur de la santé économique d'un pays.

Le marché financier est composé de différents acteurs tels que les investisseurs, les courtiers, les banques d'investissement et les bourses de valeurs. Les investisseurs achètent des titres dans l'espoir de réaliser un profit en vendant ces titres à un prix supérieur à leur prix d'achat. Les courtiers facilitent ces transactions en mettant en relation les acheteurs et les vendeurs de titres. Les banques d'investissement, quant à elles, aident les entreprises à émettre des titres pour lever des capitaux et elles sont considérées comme des acteurs majeurs.

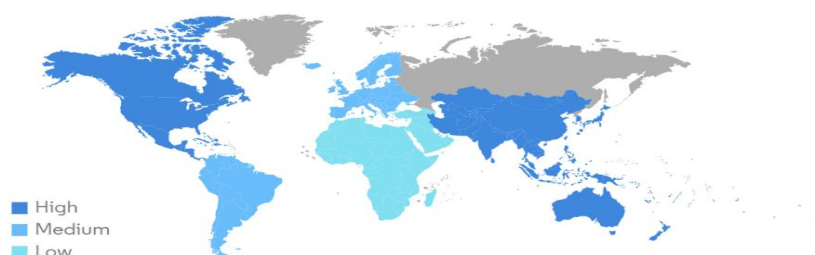


Figure 1.1 – Croissance du marché financier entre 2022-2023 par région, source : mordorintelligence.com récupérée le 03 mars 2023

Les banques et le marché financier sont interdépendants et travaillent ensemble pour faciliter l'accès des investisseurs aux marchés financiers et pour fournir des services financiers importants aux entreprises et aux particuliers.

1.2 La banque

«Établissement financier qui, recevant des fonds du public, les emploie pour effectuer des opérations de crédit et des opérations financières, et est chargé de l'offre et de la gestion des moyens de paiement», source : [larousse](#) .

Une banque est une institution financière qui opère en collectant des fonds auprès de clients déposants et en utilisant ces fonds pour accorder des prêts à d'autres clients emprunteurs. Les banques offrent également une gamme de services financiers supplémentaires, tels que des comptes courants, des comptes d'épargne, des cartes de crédit et des services de change. Les banques sont réglementées par les autorités gouvernementales pour garantir leur stabilité financière et protéger les intérêts des clients déposants. Les banques ont également la capacité de créer de la monnaie en prêtant davantage que les fonds qu'elles ont reçus, ces fonds ne sont pas conservés sous forme de liquidités physiques, mais plutôt sous forme d'écritures comptables ou de chiffres dans les comptes des clients. Ce qui leur confère un rôle important dans la régulation de la masse monétaire et de l'économie en général.

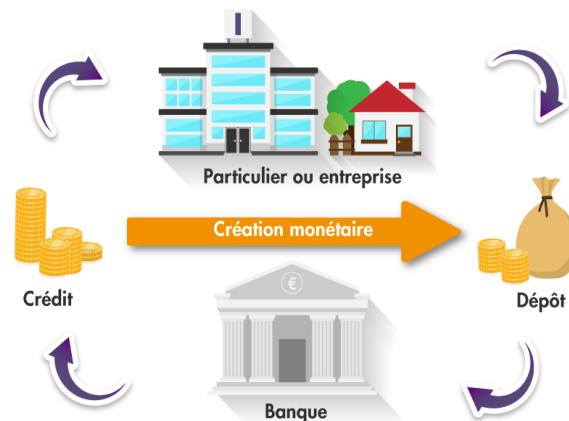


Figure 1.2 – Création monétaire, source : [lafinancepourtous.com](#) récupéré le 20 mars 2023

1.3 L'histoire d'intégration des banques multinationales en Algérie

L'intégration des multinationales et des banques en Algérie a une longue histoire qui remonte à l'époque coloniale française. En effet, dès le début du 20ème siècle. Les entreprises françaises ont commencé à s'installer en Algérie, ces entreprises opéraient principalement dans les secteurs de l'agriculture, de l'industrie et du commerce.

Dans les années 1920, les banques françaises ont créé des filiales en Algérie pour répondre aux besoins financiers des colons et des entreprises françaises présentes dans le pays. Cependant, après l'indépendance de l'Algérie en 1962, le pays a nationalisé plusieurs entreprises françaises,

y compris des banques. Les banques françaises qui avaient des filiales en Algérie ont été obligées de les céder au gouvernement algérien. Cependant, dans les années 1990, l'Algérie a commencé à ouvrir son économie aux investissements étrangers, dans le cadre d'un processus de libéralisation économique. Les multinationales et les banques étrangères ont donc commencé à revenir en Algérie, attirées par les opportunités d'investissement dans les secteurs de l'énergie, des infrastructures et de la finance.

Le groupe Société Générale a été l'une des premières banques étrangères à revenir en Algérie après la libéralisation économique.

1.4 Groupe Société Générale

1.4.1 Les missions de SG

Le groupe s'engage en tant que partenaire de confiance dans les projets qui se base sur la transformation positive du monde, en relevant les défis de l'époque, et en voyant des opportunités. Ainsi, ses principales missions sont :

- Offrir des solutions financières sur mesure pour protéger et gérer les épargnes.
- Financer les projets.
- Offrir des services innovants.
- Sécuriser les échanges.
- Protéger les clients dans leur vie et dans leurs activités.

1.4.2 Les pôles de métiers de SG

Il propose une palette de conseils et de solutions répondant aux besoins de financement de leurs clients particuliers, entreprises et investisseurs institutionnels. Les trois pôles de métiers complémentaires de ce Groupe sont :

- Banque de détail en France qui offre des services financiers aux particuliers, tels que des comptes courants, des prêts immobiliers, des cartes de crédit et des prêts personnels.
- Banque de détail à l'international spécialisé d'équipement finance et de location longue durée et gestion de flottes automobiles.
- Banque de Grande clientèle et Solutions investisseurs rassemble les équipes dédiées à la gestion d'actifs, à la banque de financement et d'investissement, au métier titres, et aux activités de banque privée.

1.4.3 Les valeurs de SG

En janvier 2020, Société Générale a défini sa raison d'être pour construire un avenir meilleur et durable avec ses clients grâce à des solutions financières responsables et innovantes. « Notre raison d'être se veut engageante et être l'expression d'une ambition collective. Elle agit comme la clé de voûte de nos choix stratégiques et la boussole pour guider nos actions au quotidien » (Frédéric Oudéa, Directeur général).

Cet extrait écrit par Frédéric Oudéa résume bien la raison d'être de ce groupe mondial. Esprit d'équipe, Innovation, Responsabilité et Engagement sont les 04 valeurs de ce groupe. Son slogan est : C'EST VOUS L'AVENIR.

La mission de la Société Générale est de construire ensemble, avec ses clients, un avenir meilleur et durable en apportant des solutions financières responsables et innovantes.



Figure 1.1– Un profil géographique diversifié du groupe SG, source : Rapport intégré 2021-2022 Récupéré le 15 mars 2023

1.5 Présentation de société générale Algérie

Société Générale Algérie, qui appartient à 100% au Groupe Société Générale, est l'une des premières banques privées à avoir établi sa présence en Algérie en 2000. Son réseau ne cesse de se développer et compte actuellement 91 agences réparties dans 35 wilayas, incluant 13 centres d'affaires dédiés aux clients entreprises.

La banque emploie également une équipe de professionnels expérimentés et qualifiés de 1500 collaborateurs pour répondre aux besoins de ses clients en proposant des services dans les deux segments du marché, le retail et le corporate, en offrant une gamme variée et novatrice de services bancaires à plus de 230 000 clients, y compris les particuliers, les professionnels et les entreprises.

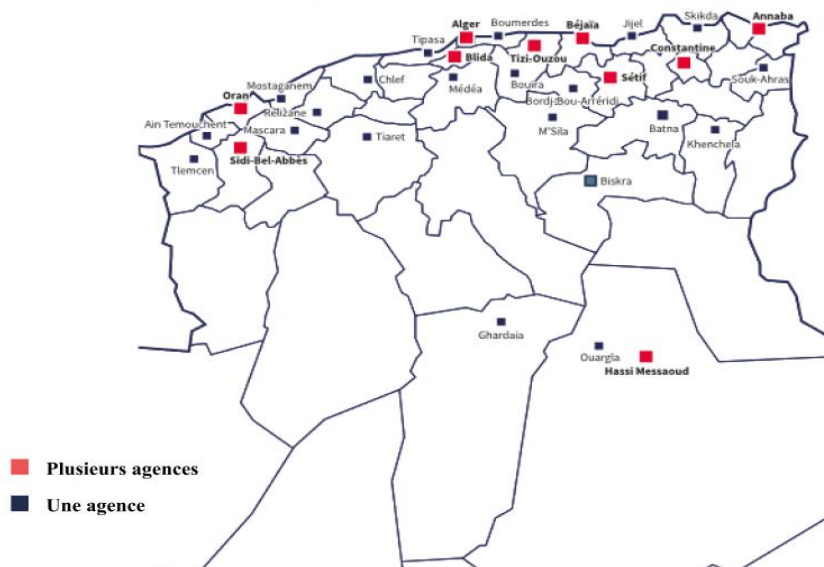


Figure 1.4 – Réseau de la Société Générale Algérie, source : documents internes à SGA

Récupéré le 14 mars 2023

1.5.1 Segmentation des clients

- **Retail :**

La segmentation des clients dans le secteur du retail se divise en deux types :

Particulier est une personne physique qui n'agit pas à des fins professionnelles ou commerciales. Les particuliers peuvent utiliser les services des banques pour gérer leur argent, épargner, emprunter, investir, etc.

Professionnel est une personne exerçant une activité professionnelle réglementée, comme un avocat, un médecin, un expert-comptable, etc. Les professionnels peuvent avoir des besoins financiers spécifiques liés à leur profession, tels que la gestion de leur trésorerie, la gestion de leur patrimoine, la fiscalité, etc. Les professionnels peuvent également avoir des besoins d'emprunt spécifiques pour financer leur activité professionnelle.

- **Corporate :**

Un client corporate est une entreprise ou une organisation qui produit des biens ou des services en vue de réaliser des profits. Les entreprises peuvent être de différentes tailles : petite, moyenne et grande entreprise.

1.5.2 Les projets de la SGA

SGA propose une gamme de services et produits à ses clients pour offrir plus de commodité et renforcer la relation avec eux.

Projets de développement d'applications en ligne

Ils sont destinés aux clients corporate et professionnels, tels que l'APPLI BtoB, l'APPLI SGA PRO et SOGECASH NET. Toutes les trois offrent des solutions de gestion de comptes bancaires à distance plus rapide et plus efficace, de réalisation d'opérations courantes et de virements en temps réel, à tout moment et en toute sécurité. La communication a été lancée à cet effet, déclinée en affiches au niveau des agences concernées, ainsi que sur les plateformes digitales (Facebook, LinkedIn & Web site Société Générale Algérie).

Centre de Relation Clients Multimédia

SGA dispose d'un CRCM en support au réseau d'agences, au service des clients de 08h30 à 20H30 7jours/7, jours fériés inclus. Le CRCM assure le service client, l'assistance aux produits de banque à distance et la pré-vente avec prises de RDV en agence via différents canaux : téléphone, Mail, interactions Web.

1.5.3 L'organigramme de la SGA

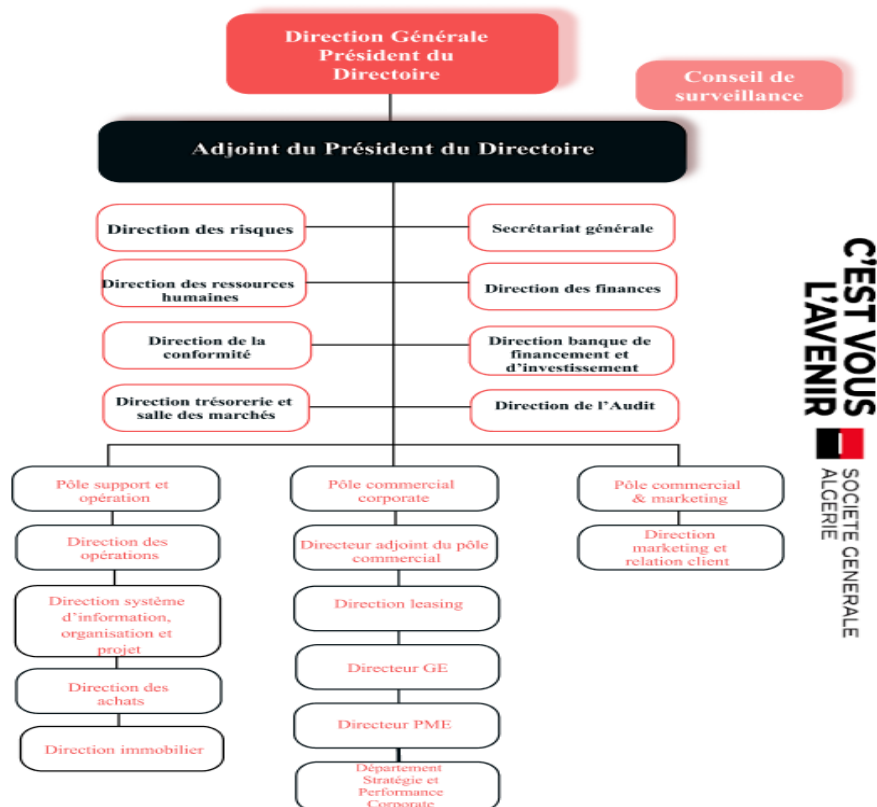


Figure 1.1– L'organigramme de la banque SGA, organigramme réalisé par les auteurs de ce rapport le 16 mars 2023

1.6 Les départements clés concernés par la mission à la Société Générale Algérie

1.6.1 Département stratégie et performance corporate

Le département Performance Stratégie Corporate est l'un des départements les plus importants de la Société Générale Algérie. Il est responsable du suivi de la performance de la banque pour améliorer ses résultats financiers. Le département comprend trois services : le service de pilotage procédure, le service de forçage et le service pilotage corporate :

- Le service de procédure est responsable de l'élaboration et de la mise en œuvre de procédures pour améliorer l'efficacité opérationnelle de la banque.
- Le service de forçage a pour mission de contraindre les débiteurs à payer leurs dettes en saisissant leurs fonds détenus sur des comptes bancaires.
- Le service pilotage corporate joue un rôle crucial dans l'activité commerciale des business center, des grandes entreprises et des petites et moyennes entreprises.

Le dernier service est le plus décisif des trois services, il est chargé de :

- Piloter et surveiller l'activité commerciale du marché corporate (GE, PME).
- Extraire les données à partir de différentes sources telles que les systèmes de gestion de la banque, les systèmes comptables et les enquêtes de satisfaction des clients.
- Réaliser des tableaux de bord dynamiques avec Excel et Power BI afin de fournir une vue globale de la performance du marché corporate en temps réel.
- Utiliser divers outils pour aider le top management à prendre des décisions éclairées et élaborer des stratégies visant à améliorer les résultats de la banque, en se basant sur plusieurs indicateurs de performance clés (KPI).
- Proposer des stratégies pour améliorer les performances financières.

En résumé, ce département est crucial pour la Société Générale Algérie, ses trois services jouent des rôles clés dans l'accomplissement des objectifs de cette banque.

1.6.2 La direction des Systèmes d'informations, Organisations et Projets

La SIOP est placée sous la supervision directe du Pôle Support et Opération de Société Générale Algérie. Elle a pour objectif d'établir et de vérifier la mise en œuvre de la politique informatique ainsi que de gérer les processus métiers, en s'assurant du respect des normes et des standards en matière de technologies de l'information et de systèmes d'informations. Elle est également responsable de la direction et du suivi des projets, qu'ils soient internes ou externes.

- Missions et activités de la SIOP

La Direction des Systèmes d'informations, Organisations et Projets a plusieurs missions cruciales :

- Maintenir le bon fonctionnement des systèmes techniques et applicatifs de la banque

- Piloter et réaliser les projets en conformité avec la stratégie de l'entreprise
- Rester à l'affût des nouvelles opportunités technologiques pour répondre aux futurs besoins de l'entreprise
- Garantir la protection des actifs informationnels de la banque en termes de confidentialité, intégrité, disponibilité et traçabilité,
- Maintenir la cohérence de l'infrastructure technologique en accord avec les besoins et la stratégie de la banque.

1.6.3 Direction des finances

Au sein de Société Générale Algérie, la direction des finances est responsable de la gestion financière de la banque. Elle a pour mission de superviser l'ensemble des activités financières et la comptabilité. Elle est également responsable de l'élaboration et du suivi du budget de la banque, ainsi que du calcul des indicateurs de performance financière tels que le produit net bancaire (PNB) et le résultat net. La direction des finances travaille en étroite collaboration avec les autres départements de la banque, notamment la direction générale, le pôle commercial corporate, afin de garantir la viabilité financière de la banque et de contribuer à la réalisation des objectifs stratégiques de l'entreprise. Elle est également chargée de veiller au respect des normes et des réglementations en matière financière, tant au niveau national qu'international.

1.6.4 Le top management

Le Top management au sein de SGA son rôle consiste à définir la stratégie globale de la banque en se basant sur les résultats des différents départements. Ils prennent des décisions importantes telles que l'expansion dans de nouveaux marchés, l'octroi de prêts, l'investissement dans de nouvelles technologies, etc. Le Top management travaille en étroite collaboration avec les différents départements de la banque pour s'assurer que la stratégie est mise en œuvre de manière efficace.

1.6.5 Analyse des interactions entre les départements concernés

Le département performance stratégie corporate de la Société Générale Algérie travaille en étroite collaboration avec la direction de la finance qui est responsable de la gestion des budgets et de tout ce qui concerne la comptabilité de la banque. Des tableaux de bord seront réalisés à partir de ces données financières au niveau du département corporate puis ils seront présentés au top management.

De plus, il collabore avec la SIOP (Système d'Information, Organisation et Projet), cette collaboration est essentielle pour le département performance stratégie Corporate, car le système d'information de la banque assure le bon fonctionnement des systèmes techniques, la collecte, le traitement, le stockage et la diffusion des données nécessaires à la prise de décision. Les indicateurs de performance sont calculés par la direction de la finance, mais leur collecte et leur analyse nécessitent une étroite collaboration avec la SIOP pour garantir la fiabilité des données et leur mise à disposition en temps voulu. De plus, la SIOP est responsable de l'élaboration des projets techniques qui permettent de mettre en œuvre les stratégies et les plans d'actions élaborés par le top management.

En somme, le département Performance Stratégie Corporate est un acteur clé de la Société Générale Algérie qui s'appuie sur une collaboration étroite entre la finance et la SIOP pour atteindre ses objectifs.



Figure 1.6– Illustration de la coordination entre les départements impliqués, schéma réalisé par les auteurs de ce rapport le 16 mars 2023

Après avoir présenté l'état des lieux, nous allons nous pencher par la suite sur le concept de performance au sein de Société Générale Algérie, permettant ainsi de mieux appréhender les critères et les mesures utilisés pour évaluer l'efficacité et le succès de la banque dans son fonctionnement et ses activités.

Chapitre 2 : La performance bancaire et la décortication de l'indicateur du Produit Net Bancaire (PNB)

Chapitre 2

La performance bancaire et la décortication de l'indicateur du Produit Net Bancaire (PNB)

Dans ce chapitre, nous procéderons à une analyse détaillée de la performance bancaire, en mettant l'accent sur les outils de mesure utilisés. Nous explorerons la palette des différents indicateurs de performance, en accordant une attention particulière au Produit Net Bancaire (PNB). Nous expliquerons comment le PNB est calculé, en mettant en évidence les variables qui le génèrent, et nous soulignerons son importance dans l'évaluation de la performance d'une banque. Nous analyserons les obstacles spécifiques rencontrés dans le contexte de la Société Générale Algérie lorsqu'il s'agit de prévoir avec précision le PNB, ce qui souligne l'importance de cette problématique pour la banque.

2.1 La performance bancaire

La performance est définie comme le fait de réaliser quelque chose avec efficacité, avec succès et de manière remarquable. Cette notion de performance est particulièrement pertinente dans le domaine bancaire, car elle implique plusieurs concepts interdépendants qui doivent être évalués ce qui rend la notion de performance bancaire une notion complexe. La capacité de la banque à atteindre ses objectifs est un premier élément à considérer, ce qui implique d'analyser l'efficacité de la banque. Le second concept à prendre en compte est celui de l'efficacité, qui évalue si la banque atteint les meilleurs résultats possibles avec les ressources dont elle dispose dans un temps donné. La pertinence est également un aspect important à considérer, elle se réfère à l'adaptation et à l'adéquation des informations, des outils de gestion, des décisions et des actions aux objectifs spécifiques d'une banque, afin de prendre des décisions éclairées et d'atteindre ces objectifs de manière efficace et efficiente. En fin de compte, la performance bancaire se mesure à la capacité de la banque à atteindre ses objectifs stratégiques et financiers, tout en offrant des produits et services de qualité à ses clients et en respectant les réglementations applicables, tout en prenant en compte l'impact social et environnemental de ses activités.

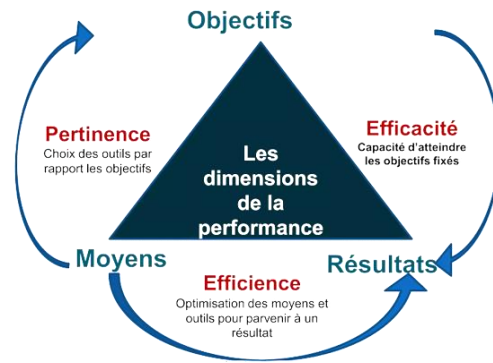


Figure 2.1–Les composantes de la performance, source : fr.slideshare.net récupérée le 11 avril

2023

2.2 Les facteurs influant sur la performance bancaire

Après avoir discuté la définition de la performance bancaire, il est crucial d'examiner les différents facteurs qui peuvent influencer cette dernière. Il est important de noter que la performance bancaire est influencée par un ensemble de facteurs, tant internes qu'externes.

2.2.1 Les facteurs internes

- **La Taille :**

La taille est un critère de performance interprété différemment par plusieurs chercheurs. Certains le considèrent comme un critère de performance, d'autres comme ayant un impact négatif, tandis que d'autres estiment que c'est un critère non significatif.

Néanmoins, une banque plus grande peut bénéficier d'avantages tels que des économies d'échelle, une plus grande diversification du portefeuille et une plus grande capacité à attirer des dépôts importants, ce qui peut se traduire par une augmentation de la rentabilité, une plus grande résilience face aux risques et une meilleure capacité à faire face aux chocs économiques. Cependant, une grande taille peut également entraîner des coûts élevés, une complexité accrue dans la gestion des risques et une perte de flexibilité.

Selon, Genay (1999) à partir de son étude sur les banques japonaises a prouvé que les grandes banques sont les plus performantes comparé aux petites. Ainsi, Halkos et Salamouris (2004) ont également prouvé cette hypothèse à partir des résultats similaires sur un échantillon de 18 banques grecques. Tandis que certaines études montrent le contraire. Cependant, d'après le test de Short (1979) à propos de l'implication de la variable taille dans la rentabilité de la banque, déduit un résultat non significatif. Selon JONGHE, les petites banques sont les plus résistantes aux conditions économiques difficiles, ainsi BARROS affirme que ces dernières ont plus de chance d'être plus performantes. Par ailleurs, selon d'autres études et chercheurs (Demirguc-Kunt et Huizinga, 2000 ; Goddard et al., 2004) ont constaté la corrélation entre la taille et la rentabilité des banques, ainsi que l'impact du capital et d'autres facteurs financiers, réglementaires et externes tels que la corruption sur leur performance.

En somme, il est important de trouver un équilibre entre les avantages et les coûts liés à la taille de la banque. Les résultats de ces études empiriques sur la relation entre la taille et la

performance bancaire sont mitigés, car elles sont appliquées dans différents contextes, pays et périodes.

- **La capitalisation :**

La capitalisation est un facteur important pour la performance bancaire. Elle se réfère à la quantité de fonds propres qu'une banque possède. Les fonds propres sont constitués par les capitaux propres de la banque et les réserves légales. Le ratio des capitaux propres sur les actifs est souvent utilisé comme un indicateur de la solvabilité financière d'une banque sachant que les actifs se réfèrent aux biens détenus tels que les prêts accordés, les investissements et les réserves. Plus le ratio des capitaux propres est élevé, plus la banque est considérée comme solvable et capable de faire face aux risques.

Les nouvelles règles de Bâle II ont introduit des relations basées sur les fonds propres, qui sont établies autour de trois piliers :

- Les exigences minimales de fonds propres des banques.
- Une surveillance prudentielle individualisée.
- La transparence et la discipline de marché.

L'objectif principal est de garantir la sécurité, la solidité et l'efficacité du système bancaire en offrant une protection contre les différents types de risques, notamment le risque de crédit, le risque opérationnel et le risque de marché.

En effet, une banque bien capitalisée est considérée comme plus solide et plus résiliente face aux chocs économiques et financiers. Elle est donc plus susceptible de bénéficier de la confiance des investisseurs et de ses clients, ainsi que d'être en mesure de maintenir ses activités même en période de crise.

- **La liquidité**

La liquidité bancaire se réfère à la capacité de la banque à mobiliser rapidement les fonds nécessaires pour répondre aux retraits de dépôt ou aux demandes de crédit, assurant ainsi la sécurité et la solidité du système bancaire.

Selon Bouwan, la liquidité n'a aucun effet significatif sur la performance, par contre elle a un effet négatif sur la marge d'intérêt net. Une bonne quantité d'actifs liquide est importante pour la banque afin de faire face aux événements imprévus, ainsi la liquidité stockée au bilan permet d'éviter des ventes d'actifs bradées pour satisfaire le besoin de liquidité. Aussi, il précise qu'avoir un taux de liquidité satisfaisant permet de réduire le risque de liquidité.

D'après les différentes recherches obtenues par Molyneux et Thornton (1992) et Pasiouras & Kosmidou (2007) prouve que la liquidité a un impact négatif sur la rentabilité, par ailleurs, Kosmidou et al. (2006) et Bourke (1989) montre qu'elle possède un impact positif. Ainsi, selon (Maghyrech & Shammout, 2004) cette différence est justifiée par l'élasticité de la demande de crédit.

- **Le risque de crédit**

Le risque de crédit est la probabilité que les emprunteurs d'une banque ou d'un établissement financier ne soient pas en mesure de rembourser leur dette conformément aux termes convenus, ce qui peut entraîner un défaut de paiement.

En effet, il ne dépend pas seulement du risque de défaillance mais aussi des risques liés à la rentabilité de la banque. Pour montrer la relation entre le risque de défaillance et la rentabilité plusieurs recherches ont été produites. A travers des résultats obtenus de ces travaux (Athanasoglou, et al, 2006 ; Miller & Noulas, 1997) il a été concrètement prouvé que ce risque a un effet négatif sur la performance de la banque. Car à chaque fois le risque est élevé, plus la valeur des crédits augmentent et le rendement des actifs diminue.

- **L'efficacité :**

L'efficacité se réfère à la capacité d'une banque à utiliser efficacement ses ressources pour générer des revenus et des bénéfices. Elle peut être mesurée par rapport aux coûts de la banque, tels que les coûts opérationnels et les coûts liés au personnel, ainsi qu'aux revenus, tels que les revenus d'intérêts et les revenus de commissions.

Une banque efficace peut augmenter sa rentabilité en maximisant ses revenus et en minimisant ses coûts, ce qui peut conduire à une croissance plus rapide, une plus grande part de marché et une meilleure compétitivité.

Selon les études de (Bourke (1989) et Molyneux & Thornton (1992)) ont montré une relation significative entre la rentabilité et l'efficacité de la banque. Ainsi, les résultats des travaux de (Maghyrech & Shammout, 2004 ; Athanasoglou, et al, 2006...) ont approuvé d'autres variables explicatives comme les frais d'exploitation, total actif et résultat avant impôt, en effet, ils ont conclu que la relation entre la rentabilité et ces variables est une relation inverse.

D'après les auteurs qui ont étudié l'impact de ce critère sur la performance constatent que l'efficacité améliore la performance bancaire, cela en utilisant n'importe qu'elle variable (ROA, ROE et NIM) pour la mesure ainsi que le type de contrôle qui s'exerce (état ou privé)

2.2.2 Les facteurs externes

- **Croissance du PIB**

Elle est considérée comme un indicateur important du développement économique et de la richesse d'un pays sur une période donnée. Elle est calculée en comparant la valeur totale de tous les biens et services finaux produits dans un pays au cours d'une année à la valeur totale de la production de l'année précédente.

Les études empiriques sur la relation entre la croissance du produit intérieur brut PIB et la rentabilité des banques ont donné des résultats divergents. Bien que la croissance du PIB soit considérée comme ayant un impact positif et significatif sur la rentabilité des banques, certaines études ont trouvé une relation positive entre ces deux variables, tandis que d'autres ont trouvé une relation négative ou non significative.

Sufian et al. (2008) ont identifié un impact positif dans leur étude portant sur la relation entre le ROA (Return on Assets) et les déterminants externes ou macro-économiques des banques aux Philippines. Pasiouras et Kosmidou (2007) ont également obtenu des résultats similaires lors de leur étude sur l'impact des facteurs influençant la rentabilité des banques commerciales domestiques et étrangères dans l'Union Européenne.

Bernank et Gertler (1989) ainsi que Demirguç-Kunt et al. (2004) ont même identifié une relation inverse entre la croissance du PIB et la rentabilité des banques. Vong (2009) a constaté, dans son étude sur les banques de Macao, que la croissance du PIB n'a pas d'impact significatif sur le ROA (Return on Assets).

- **L'inflation :**

L'inflation est une augmentation générale et durable des prix des biens et des services dans une économie donnée, ce qui entraîne une diminution de la valeur de la monnaie. Elle peut être causée par une augmentation de la demande pour les biens et les services, une baisse de l'offre de biens et de services, ou une augmentation de la quantité de monnaie en circulation. Des recherches ont été menées par plusieurs chercheurs, notamment Revel en 1979, Debinna Rasiah en 2010, Khrawish en 2010, Sufian en 2008, Perry en 1992 et Pasiouras et Kosmidou en 2007, qui ont confirmé que l'inflation peut avoir un impact négatif sur la rentabilité des banques.

- **Taux d'intérêt :**

Les intérêts se réfèrent généralement aux sommes d'argent payées ou reçues en échange de l'utilisation de l'argent déposé ou emprunté. Le taux d'intérêt, qui est le pourcentage appliqué à ces sommes, est largement considéré comme un déterminant clé de la rentabilité des banques dans de nombreuses études, car il influence directement le revenu net d'intérêt. Le revenu net d'intérêt résulte de la différence entre les intérêts perçus sur les prêts accordés et les intérêts payés sur les emprunts contractés. En conséquence, le taux d'intérêt est souvent présenté comme une variable externe dans les recherches sur les déterminants de la rentabilité des banques, car il est largement influencé par les politiques économiques gouvernementales ainsi que par les conditions du marché de l'offre et de la demande.

Claessens et al. (2016) ainsi que Molyneux et Thornton (1992) ont identifié une relation positive et significative entre le taux d'intérêt et la rentabilité des banques. (Claessens, S, N Coleman and M Donnelly (2016). Ben Naceur (2003) a constaté une relation négative entre le taux d'intérêt et la rentabilité bancaire. (Ben Naceur, S. (2003). Dans son étude portant sur l'impact des variables macro-économiques sur la rentabilité des banques au Nigeria sur la période 1980-2006, Toni Uhomoibhi (2008) a mis en évidence que le taux d'intérêt réel est un déterminant significatif de la rentabilité des banques au Nigeria. Ce résultat est cohérent avec ceux obtenus par Pasiouras et Kosmidou (2007) ainsi que Sufian et al. (2008).

2.3 Exploration des mesures de rentabilité dans l'évaluation de la performance bancaire

2.3.1 De la rentabilité à la rentabilité bancaire

Après avoir étudié les différents facteurs qui influencent la performance bancaire, il est évident que cette dernière doit être mesurée. Il est nécessaire d'utiliser différents indicateurs de mesure de rentabilité. L'analyse de ces indicateurs permet de mesurer la performance d'une banque, de la comparer à ses concurrents et d'identifier les domaines dans lesquels elle peut améliorer sa performance.

« La rentabilité d'un établissement de crédit représente son aptitude à dégager de son exploitation des gains suffisants, après déduction des coûts nécessaires à cette exploitation, pour poursuivre durablement son activité » (NOUY, 1992).

La rentabilité bancaire fait partie des mesures de performance bancaire. Elle permet d'évaluer l'efficacité avec laquelle la banque utilise ses ressources pour générer des bénéfices et atteindre ses objectifs financiers, elle est définie comme sa capacité à générer des profits suffisants à partir de son activité après la déduction des coûts, pour assurer la viabilité à long terme de ses activités. Mesurer la rentabilité est crucial pour évaluer la santé financière et la pérennité de la banque et même pour évaluer la conformité aux exigences réglementaires et prudentielles.

2.3.2 Les instruments de mesure de la rentabilité

Il existe différents instruments pour mesurer la rentabilité des banques, qui sont généralement classés en trois catégories selon NOUY (1992) :

- l'approche par les soldes intermédiaires de gestion
- l'approche par les coûts, les rendements et les marges
- l'approche par les ratios de rentabilité

- L'approche par les soldes intermédiaires de gestion :

L'article de Nouy souligne que la rentabilité des banques ne peut pas être évaluée uniquement par le résultat net, car il peut inclure des produits ou des charges non récurrentes qui peuvent masquer la véritable structure de la rentabilité de la banque. Ainsi, pour une analyse précise de la rentabilité, il est important d'utiliser des soldes intermédiaires pour identifier les éléments qui ont contribué au résultat final. L'idée générale est que l'analyse de la rentabilité d'une banque nécessite une approche plus détaillée que simplement le résultat net.

- **Produit net bancaire (PNB)**

Le produit net bancaire est le résultat de l'activité courante d'une banque, calculé en soustrayant les intérêts payés aux dépôts et les commissions versées pour les services financiers fournis, du montant total des intérêts et commissions perçus par la banque auprès de ses clients. Où les intérêts représentent les revenus générés par les prêts et les investissements de la banque, tandis que les commissions sont des frais facturés pour des services tels que la gestion de compte, la carte bancaire, etc.

$$\text{PNB} = \text{intérêts reçus} + \text{commissions perçues} \\ - \text{intérêts payés} - \text{commissions versées}$$

- **Résultat Brut d'Exploitation (RBE)**

Le RBE correspond à la différence entre le produit net bancaire et la somme des frais généraux et des dotations aux amortissements. Ce ratio permet d'évaluer la capacité de la banque à générer une marge une fois que les coûts liés aux ressources et aux charges de fonctionnement ont été pris en compte. Toutefois, le RBE ne donne pas une indication précise sur la capacité bénéficiaire d'une banque car il ne tient pas compte du coût du risque.

$$\text{RBE} = (\text{PNB} + \text{produits accessoires}) - (\text{frais généraux} + \text{amortissements})$$

- **Résultat d'Exploitation (RE)**

Le résultat d'exploitation correspond à la marge globale générée par une banque à travers toutes ses activités, une fois que les frais de structure et le risque de crédit ont été pris en compte. Il est calculé en soustrayant les dotations aux provisions du résultat brut d'exploitation, ce qui permet de prendre en compte le risque associé aux activités de la banque.

$$\text{RE} = \text{RBE} - \text{dotations nettes aux provisions}$$

- **Résultat net (RN)**

Le résultat net est calculé en prenant en compte plusieurs éléments : le résultat d'exploitation, les produits et charges exceptionnels, la participation des salariés ainsi que les impôts sur les. Il permet de mesurer la différence entre les revenus et les dépenses de la banque.

$$\text{RN} = \text{RE} + \text{autres produits} - \text{autres charges} - \text{participatives. salariés} - \text{impôt sociétés}$$

- L'approche par les coûts, les rendements et les marges

Grâce à cette approche, il est possible d'évaluer la rentabilité d'une banque de manière concise et complète, en prenant en compte l'ensemble de ses activités à l'aide d'indicateurs simples qui permettent de comparer ses performances avec celles des autres banques, que ce soit à l'échelle nationale ou internationale.

- **Le coût moyen des ressources (CMR) :**

Une banque doit collecter des fonds auprès de ses clients et du marché monétaire pour financer ses activités. Cette activité principale de la banque implique des coûts qui varient en fonction de la source et des caractéristiques des fonds collectés. Pour évaluer ces coûts, les banques calculent un coût moyen des ressources afin de déterminer les charges qu'elles supportent. Ce coût est obtenu en divisant la somme des intérêts versés par la somme des encours moyens des ressources collectées.

$$\text{CMR} = \frac{\text{Somme des intérêts versés}}{\text{Somme des encours moyens des ressources}}$$

- **La marge d'intérêt :**

La marge d'intérêt nette est un indicateur important pour mesurer la rentabilité et la croissance d'une banque. Ce ratio permet de calculer la différence entre les intérêts gagnés par la banque sur ses prêts et ceux qu'elle doit payer sur les dépôts des clients.

Plus le ratio est élevé, plus la banque gagne d'argent grâce à cette activité principale de prêt et d'emprunt. La marge d'intérêt nette est donc un indicateur clé pour évaluer la performance financière d'une banque.

- **Rendement Moyen des Emplois (RME) :**

Les banques sont confrontées au défi de calculer avec précision le rendement de l'ensemble de leurs emplois, car chaque emploi possède des caractéristiques propres (type, nature, maturité, etc.) rendant le calcul du coût de ressources difficile. Pour résoudre ce problème, les banques déterminent un rendement moyen de leurs emplois en calculant le rapport entre la somme des intérêts perçus et la somme des encours des emplois.

$$\text{RME} = \frac{\text{Somme des intérêts perçus sur les emplois}}{\text{Somme des encours moyens des emplois}}$$

- **L'écart de taux :**

C'est un indicateur important pour mesurer la rentabilité globale des activités de la banque et évaluer la capacité de la banque à générer des bénéfices. Il est calculé en soustrayant le rendement moyen des emplois du coût moyen des ressources.

$$\text{L'écart de taux} = \text{RME} - \text{CMR}$$

- **L'approche par les ratios de rentabilité**

L'approche par les ratios est une méthode pratique pour évaluer la rentabilité des banques en utilisant des indicateurs fiables et simples à utiliser. Cette approche permet de synthétiser les chiffres clés des soldes intermédiaires de gestion à travers plusieurs ratios, qui reflètent les différents aspects de la rentabilité bancaire.

- **ROE :**

Le ratio de la rentabilité des fonds propres (Return On Equity, ROE en anglais) est un indicateur clé de la performance financière d'une banque. Il mesure sa capacité à générer des bénéfices en rapport avec les fonds propres investis par les actionnaires.

$$\text{ROE} = \frac{\text{bénéfice net}}{\text{moyenne des fonds propres}}$$

Plus le ROE est élevé, plus la banque est rentable par rapport aux fonds propres investis. Le ROE est un indicateur important pour les investisseurs et les actionnaires car il permet d'évaluer la capacité de l'entreprise à générer des rendements sur les fonds qu'ils ont investis.

Cependant, il est important de noter que le ROE peut être influencé par l'effet de levier financier, c'est-à-dire l'utilisation de la dette pour financer les activités de l'entreprise. Une entreprise qui utilise beaucoup de dettes pour financer ses activités peut avoir un ROE plus élevé, mais cela peut également augmenter le risque financier et la volatilité des bénéfices. Par conséquent, il est important d'utiliser le ROE en combinaison avec d'autres indicateurs financiers pour évaluer la performance globale de l'entreprise.

- **ROA :**

Le ROA ou "Return on assets" en anglais, est un indicateur de performance qui permet d'évaluer le taux de rendement de l'ensemble des actifs d'une banque. Il mesure ainsi la capacité de l'entreprise à générer des bénéfices à partir de ses investissements.

$$\text{ROA} = \text{Résultat net} \div \text{actifs nets}$$

Sachant que le résultat net obtenu correspondra à la différence entre les produits et les charges et le total des actifs représentent l'ensemble des ressources détenues par la banque, y compris les prêts accordés aux clients, les investissements, les réserves et les fonds propres.

Un ROA élevé indique que la banque est efficace dans l'utilisation de ses ressources. Dans le cas contraire, la rentabilité de la banque est insuffisante par rapport à ses ressources matérielles et financières

De ce fait, ce ratio ne peut pas être gonflé par effet de levier. En l'absence de dette, ROE et ROA sont égaux. En présence de dette, mieux vaut utiliser le ROA pour évaluer la performance d'une entreprise.

- **Le ratio de la marge nette :**

C'est un indicateur qui permet d'évaluer la rentabilité d'une banque en exprimant le pourcentage du chiffre d'affaire qui reste disponible après déduction de l'ensemble des coûts engagés pour son activité, y compris les charges financières (les intérêts) et fiscales (les impôts).

$$\text{Marge bénéficiaire nette} = (\text{Bénéfice net} / \text{Chiffre d'affaires net}) \times 100$$

S'il est élevé, il indique que l'entreprise a plus de fonds disponibles pour les investissements futurs ou la distribution aux actionnaires. Ainsi, il est considéré comme un indicateur clé de la santé financière d'une banque. Ce ratio est généralement utilisé pour le suivi de la performance d'une banque au fil du temps ou pour comparer la performance de différentes banques.

2.3.3 Les ratios déterminants de la rentabilité : liquidité, solvabilité et risque

Maintenant que la mesure de la performance bancaire, à travers la rentabilité, a été abordée, il est temps d'examiner les ratios clés qui contribuent à cette mesure. Ces ratios sont la liquidité, la solvabilité, et le risque. Il est important de noter que ces trois ratios ne peuvent pas être considérés de manière isolée les uns des autres, ils sont étroitement liés et interdépendants. En ayant une vue d'ensemble de ces ratios et en les gérant de manière intégrée et proactive, les banques peuvent mieux comprendre leur performance financière globale et améliorer leur rentabilité à long terme.

- **Le ratio de liquidité :**

Le ratio de liquidité, ou coefficient de liquidité, est un indicateur financier utilisé pour mesurer la capacité d'une banque à faire face à ses engagements à court terme avec ses actifs liquides mobilisables. En d'autres termes, il mesure la capacité d'une banque à rembourser ses dettes à court terme avec ses disponibilités en espèces et ses investissements à court terme.

Ce ratio est obtenu en évaluant le montant de l'actif à court terme (incluant les stocks) par rapport au passif d'une durée équivalente. Sachant que les passifs représentent les dettes de la banque, y compris les dépôts des clients et les emprunts contractés.

Pour calculer ce ratio, la formule suivante est utilisée :

$$\text{ratio de liquidité} = \frac{\text{Actif court terme}}{\text{passif court terme}}$$

Le ratio de liquidité générale doit être supérieur à 1 pour assurer une certaine marge de sécurité à la banque. En effet, plus ce ratio est élevé, plus l'entreprise dispose de liquidités suffisantes pour faire face à ses échéances à court terme et pour développer ses activités. À l'inverse, un ratio inférieur à 1 signifie que la banque ne dispose pas de suffisamment de liquidités pour honorer ses dettes à court terme, ce qui peut compromettre sa santé financière et son développement. Dans ce cas, on parle de fonds de roulement négatif, sauf en présence de circonstances particulières justifiant cette situation.

Le respect de ce ratio est important pour assurer la solvabilité d'une banque et éviter une crise de liquidité.

- **Les ratios de solvabilité :**

La solvabilité d'une banque fait référence à sa capacité à absorber des pertes financières imprévues sans remettre en question ses engagements envers ses déposants et ses créanciers. Cette capacité est principalement mesurée à travers le ratio de solvabilité, qui a été établi par le comité de Bâle en 1988 sous le nom de ratio Cooke. Selon ce ratio, les banques doivent détenir des fonds propres équivalant à au moins 8% des risques de crédit encourus du fait de leurs opérations, afin de garantir leur stabilité financière. Ce ratio de solvabilité est donc un indicateur clé de la capacité d'une banque à faire face à des chocs économiques et à continuer à fournir des services financiers à ses clients.

Ratio cook = (fonds propre nets /les risques pondérés) \geq 8%

Sachant que :

les risques pondérés = engagements x coefficient de pondération (variable)

Au fil du temps, avec la complexité croissante du monde bancaire et l'innovation financière, le ratio Cooke a révélé ses limites. En effet, il ne prenait pas en compte certains risques tels que le risque opérationnel, le risque de taux d'intérêt et le risque de liquidité. C'est pourquoi il a été révisé pour inclure le risque opérationnel dans son calcul en l'intégrant dans le dénominateur. Ainsi le ratio MC Donough a été mis en place par le comité de Bâle II.

Ratio MC Donough = (Fonds propres /risques (crédit +marché+opérationnel)) \geq 8%

- **Les ratios de division de risque :**

Les ratios de division de risque sont des indicateurs financiers utilisés pour évaluer la répartition des risques de crédit d'une banque sur différents emprunteurs ou secteurs économiques. Ils sont utilisés pour mesurer la diversification du portefeuille de crédits d'une banque et évaluer le risque de défaut de paiement. Les deux ratios de division de risque les plus couramment utilisés sont le ratio de concentration et le ratio d'exposition aux grands risques :

- Le ratio de concentration maximal est de 25%, ce qui signifie que les fonds propres réglementaires d'une banque ne doivent pas représenter plus de 25% des risques de crédit nets pondérés encourus sur un même bénéficiaire.

fonds propres réglementaires / somme des risques de crédit nets pondérés \leq 25%

- Le ratio d'exposition aux grands risques stipule que le montant total des grands risques, représentant entre 10% et 25% des fonds propres réglementaires, ne doit pas dépasser huit fois le montant des fonds propres réglementaires de la banque.

total des grands risques \leq 8 * fonds propres

2.4 Décryptage du PNB : une approche essentielle pour évaluer la performance bancaire

Maintenant que nous avons passé en revue les différents indicateurs utilisés pour mesurer la performance des banques, nous allons nous concentrer sur le produit net bancaire PNB. Comme nous l'avons déjà défini auparavant, le PNB est le résultat de l'activité principale de la banque, à savoir l'intermédiation financière. Nous allons nous pencher plus en détail sur cet indicateur clé de la performance bancaire, examiner de manière profonde les différentes sources qui le génèrent pour mieux comprendre ce chiffre important.

Voici un récapitulatif du calcul du PNB et de ses composantes, comme mentionné précédemment :

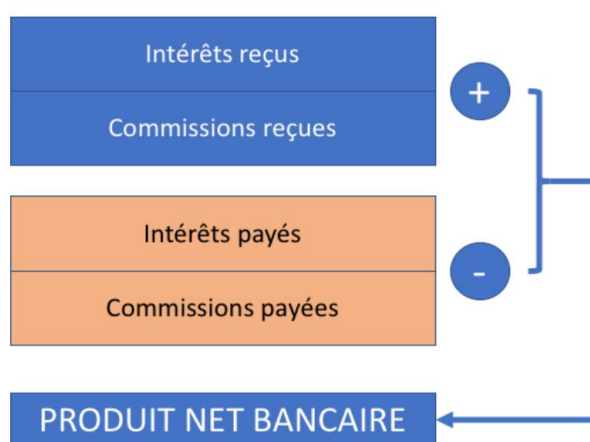


Figure 2.2 – Schéma récapitulatif du calcul du Produit Net Bancaire (PNB), source : openclassrooms.com récupérée le 11 avril 2023

Les intérêts reçus sont les revenus générés par les prêts et les investissements de la banque, tandis que les intérêts payés sont les frais que la banque doit payer sur les dépôts et les emprunts qu'elle a contractés. De même, les commissions perçues sont les frais facturés par la banque pour les services qu'elle fournit à ses clients, tels que les frais de gestion de compte, les frais de transaction, etc. Les commissions versées sont les frais que la banque doit payer pour obtenir des services de tiers, tels que les frais de carte de crédit, les frais de courtage, etc.

Les banques tirent une partie importante de leurs revenus des mouvements créditeurs, encours, export et import. En effet, ces activités sont considérées comme des facteurs clés qui génèrent des revenus pour les banques, sous forme d'intérêts et de commissions.

2.4.1 Les sources clés de revenus pour les banques : Les moteurs de croissance du PNB

- Les mouvements créditeurs :

Les mouvements créditeurs sont des opérations qui ajoutent des fonds au compte bancaire d'un

individu ou d'une entreprise. Il s'agit donc d'une somme qui entre sur votre compte bancaire et qui a un impact positif sur ce dernier. Cela peut inclure des dépôts, des transferts entrants, des remises de chèques et des versements de salaire, entre autres. Les mouvements créditeurs peuvent également générer des commissions ou des intérêts, ou les deux en fonction de la nature du compte, en fonction des termes et conditions établis avec la banque.

- Les encours :

Ils représentent les dettes qu'un débiteur doit à sa banque, ainsi que l'ensemble des crédits qu'un emprunteur doit à sa banque. Elle peut également représenter l'ensemble des avoirs nets incluant les actifs et les passifs à un instant donné, en particulier pour les entreprises. La notion d'encours bancaire est importante lors des calculs de solvabilité d'un emprunteur, afin d'évaluer sa capacité à rembourser ses dettes. Les encours peuvent générer des intérêts pour la banque ou des commissions en fonction de la nature de la dette. Par exemple, si l'encours représente un crédit accordé à un client, la banque peut percevoir des intérêts sur ce crédit. En revanche, si l'encours représente une commission de service liée à un compte bancaire, la banque peut percevoir des commissions en fonction des frais facturés au client.

- Import-export :

Les activités d'import-export sont souvent des transactions à grande échelle qui nécessitent des financements importants. Les banques peuvent offrir des solutions de financement pour simplifier ces transactions. En facilitant le paiement (offrir des services de change de devises) et la documentation liés aux transactions d'import-export la banque peut facturer des commissions. En offrant des services de financement et de gestion de trésorerie pour ces activités, les banques peuvent donc générer des revenus supplémentaires tout en aidant les entreprises à réaliser leurs opérations d'import-export de manière efficace et efficiente.

2.3.2 Au-delà des indicateurs financiers : le PNB comme mesure holistique de la performance bancaire

Il est important de connaître et d'évaluer le produit net bancaire (PNB) car il représente la principale source de revenus d'une banque. Les rapports financiers annuels publiés par les banques locales et mondiales contiennent des tableaux de bord et des statistiques détaillées sur le PNB. Cela est dû à l'importance cruciale du PNB en tant qu'indicateur de la performance financière globale d'une banque. En effet, à partir du PNB, on peut calculer la majorité des ratios et indicateurs qui permettent d'obtenir une vision plus complète de la performance financière de la banque tels que : ROA, ROE, RN, RE, RBE et la marge nette. L'évaluation du PNB permet aux banques de comprendre leur rentabilité, de mesurer l'efficacité de leur stratégie commerciale et de prendre des décisions éclairées pour améliorer leur performance financière en évaluant le montant total des revenus que la banque génère à partir de ses activités principales, telles que l'octroi de prêts, la collecte de dépôts, la gestion de trésorerie et la vente de produits et services financiers.

En comprenant les différents composants du PNB, les investisseurs, les analystes financiers et les régulateurs peuvent évaluer la stabilité financière de la banque, ils peuvent également identifier les sources de revenus les plus rentables et les moins rentables. Par exemple, si une banque constate que ses commissions perçues sont en baisse, elle peut décider de concentrer ses efforts sur la génération de nouvelles sources de revenus de commission ou d'améliorer sa stratégie en mettant en place de nouveaux produits et services ou l'ajustement des tarifs pour améliorer leur rentabilité et stimuler la croissance des commissions.

En évaluant le PNB, les banques peuvent également identifier les tendances et les évolutions du marché financier, ainsi que les opportunités émergentes. Cela peut leur permettre de s'adapter rapidement aux changements du marché et de saisir les opportunités de croissance.

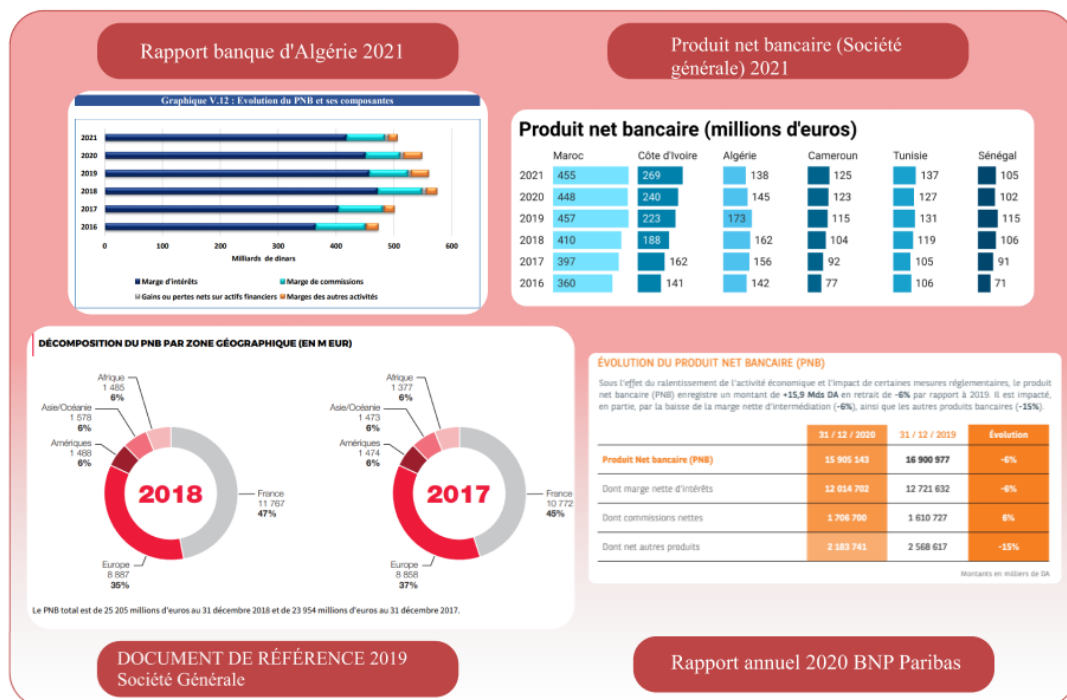


Figure 2.3 –Tableaux de bord de PNB provenant de différents rapports financiers de banques, figure réalisée par les auteurs de ce rapport le 15 avril 2023

Comme on peut le voir sur la figure, ces rapports financiers fournissent une mine d'informations sur la performance de la banque, y compris des statistiques clés telles que le PNB. Ces rapports sont facilement accessibles en ligne et ils sont utiles pour les investisseurs, les analystes financiers et les régulateurs pour évaluer la santé financière des banques.

En somme, le produit net bancaire est une mesure importante qui fournit une indication précieuse sur la performance financière ainsi il permettra aux de prendre des décisions éclairées et de rester compétitives dans un marché financier en constante évolution.

2.4 Problématique

Au sein de la banque, l'une des missions primordiales consiste à piloter et surveiller l'activité commerciale à travers des indicateurs de performance tels que le produit net bancaire (PNB).

Ce dernier permet de mesurer les revenus générés par les activités principales de la banque, telles que l'octroi de prêts, la gestion de trésorerie et la vente de produits et services financiers. Comme mentionné dans le chapitre précédent, il est important de souligner que les clients corporate sont ceux qui génèrent le plus de revenus pour une banque, car ils représentent les grandes entreprises, moyennes et petites, qui ont besoin de services financiers complexes. Par conséquent, le PNB corporate est le PNB sur lequel la banque mise pour augmenter ses revenus et maximiser ses bénéfices.

Avoir une vue globale sur l'évolution du PNB corporate dans les années à venir permet de mesurer l'efficacité des stratégies commerciales à long terme, de planifier et de prendre des décisions éclairées pour faire face aux défis du marché financier qui en constante évolution. En effet, la capacité à prédire le PNB corporate est essentielle pour les banques qui cherchent à maintenir leur compétitivité sur le marché.

Cependant, il y a des défis à relever pour parvenir à cette prédiction, notre objectif a été confronté à des problèmes au sein de la banque concernée. Tout d'abord, le manque d'enregistrement détaillé par segment de client des données du PNB avant 2018 a posé un premier défi. Avant cette date, la banque calculait uniquement le PNB global annuel. Le deuxième défi réside dans le fait qu'une prédiction basée sur seulement cinq ans est incertaine, surtout que 2 ans de cette période ont été impactés par la pandémie COVID-19, qui a eu un impact majeur sur l'économie mondiale.

Pour remédier à cette problématique, nous nous sommes basés sur l'avis des experts qui ont mis en avant l'importance de l'augmentation des encours pour assurer une augmentation du PNB. Notre mission consiste à prouver le lien fort entre les encours et le PNB, en menant une étude et une analyse statistique pertinente. Nous utiliserons ensuite des modèles de séries temporelles adaptés à la problématique pour prédire les encours, cette prédiction des encours permet d'avoir une vue d'ensemble sur l'évolution du PNB corporate dans les années à venir grâce au lien étroit qui existe entre ces deux variables.

Après avoir exposé les défis rencontrés au sein de la Société Générale Algérie, nous aborderons, dans le chapitre suivant, les différentes méthodes utilisées pour résoudre cette problématique et atteindre les objectifs définis précédemment.

Chapitre 3 : Outils et méthodes pour l'analyse et la prévision des séries temporelles

Chapitre 3

Outils et méthodes pour l'analyse et la prévision des séries temporelles

Au cours de ce chapitre, nous explorerons les différents concepts liés à l'étude statistique démontrant l'impact d'une variable X sur une variable Y. Nous aborderons notamment l'analyse exploratoire des variables, qui consiste à examiner et à comprendre les relations entre différentes variables, ainsi que les techniques de prévision des séries temporelles. Enfin nous expliquons les deux paradigmes backend et frontend nécessaires pour concevoir une interface utilisateur.

3.1 Investigation statistique de la relation significative entre les variables

3.1.1 La régression multiple

En général, la régression multiple est une méthode utilisée pour expliquer ou prédire la variance d'une variable dépendante (endogène) en utilisant une combinaison de variables indépendantes (exogènes). La régression est dite multiple lorsque le modèle comprend au moins deux variables indépendantes, alors qu'elle est simple lorsqu'il n'y a qu'une seule variable indépendante. Elle est généralement représentée par une équation de la forme :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Où Y est la variable dépendante, β_0 est l'intercept, β_1 à β_n sont les coefficients de régression correspondant à chaque variable indépendante X_1 à X_n , et ε est l'erreur résiduelle. La régression multiple repose sur des hypothèses importantes. Ces hypothèses sont nécessaires pour garantir la validité et l'interprétation correcte des résultats du modèle :

- **Linéarité** : La relation entre les variables indépendantes et la variable dépendante est linéaire.
- **Homoscédasticité** : Les résidus (erreurs) de la régression ont une variance constante à tous les niveaux des variables indépendantes.
- **Indépendance des résidus** : Les résidus ne sont pas corrélés entre eux.
- **Absence de colinéarité parfaite** : Les variables indépendantes ne sont pas parfaitement corrélées les unes avec les autres.
- **Distribution normale des résidus** : Les résidus suivent une distribution normale.
- **Absence d'observations influentes** : Il n'y a pas d'observations qui ont un effet disproportionné sur les résultats de la régression.

La régression multiple permet d'examiner l'effet de plusieurs variables indépendantes simultanément sur la variable dépendante en utilisant :

Tableau ANOVA : Le tableau ANOVA fournit des informations sur la variance expliquée par le modèle (TSS), la variance expliquée par les variables indépendantes (ESS) et la variance non expliquée (RSS). Il permet de quantifier la contribution de chaque variable indépendante à l'explication de la variance de la variable dépendante. Pour calculer le tableau ANOVA (Analysis of Variance) dans le contexte de la régression multiple, vous aurez besoin des paramètres suivants (Vous trouverez les formules mathématiques de ces paramètres en annexe A) :

DDL (degrés de liberté) : Ce sont les degrés de liberté associés à chaque composant de variance. Le DDL pour TSS est égal au nombre total d'observations moins un, tandis que le DDL pour ESS est égal au nombre de variables indépendantes du modèle. Le DDL pour RSS est égal à la différence entre le nombre total d'observations et le nombre de variables indépendantes du modèle.

TSS (Total Sum of Squares) : Il s'agit de la somme des carrés des différences entre les valeurs observées de la variable dépendante et la moyenne de la variable dépendante.

ESS (Explained Sum of Squares) : C'est la somme des carrés des différences entre les valeurs prédites par le modèle et la moyenne de la variable dépendante. L'ESS représente la part de variance expliquée par le modèle.

RSS (Residual Sum of Squares) : C'est la somme des carrés des résidus, c'est-à-dire les différences entre les valeurs observées de la variable dépendante et les valeurs prédites par le modèle. Le RSS représente la part de variance non expliquée par le modèle.

Coefficient de détermination (R²) : R² mesure la proportion de la variance totale de la variable dépendante expliquée par le modèle de régression. Il fournit une indication de l'adéquation du modèle et de sa capacité à expliquer les variations de la variable dépendante. Une valeur élevée de R² indique un meilleur ajustement du modèle aux données observées.

Le t-test : il est utilisé pour déterminer si chaque coefficient de régression est statistiquement différent de zéro. En d'autres termes, il teste l'hypothèse nulle selon laquelle il n'y a pas de relation linéaire entre la variable indépendante et la variable dépendante. Pour effectuer un t-test, on calcule le t-score pour chaque coefficient de régression en divisant le coefficient de régression par son erreur standard. Le t-score suit une distribution de Student et on utilise une valeur critique de la distribution pour déterminer si le coefficient de régression est significativement différent de zéro.

Statistique F : C'est le rapport entre l'ESS et le RSS, ajusté pour les degrés de liberté correspondants. La statistique F est utilisée pour tester l'ensemble de l'hypothèse nulle selon laquelle tous les coefficients de régression sont nuls, c'est-à-dire qu'il n'y a pas de relation significative entre les variables indépendantes et la variable dépendante. Une valeur élevée de la statistique F, associée à un faible niveau de signification (p-value), indique que le modèle de régression est globalement significatif.

3.1.2 Le coefficient standardisé

Les coefficients standardisés permettent de connaître le sens de la relation entre chaque prédicteur et la variable dépendante. Ce coefficient est calculé en divisant la constante (intercept) et les coefficients de chaque variable explicative par l'écart type de la variable

respective. Cela permet de mesurer l'importance relative des variables, en divisant les coefficients par l'écart type qui mesure la dispersion des données autour de leur moyenne, on les met à l'échelle et on les rend comparables, quelles que soient les unités de mesure des variables explicatives.

Coefficient standardisé = Coefficient / Écart type de la variable explicative

Les coefficients standardisés permettent d'identifier les variables explicatives qui ont une influence plus importante sur la variable cible par rapport à celles ayant des coefficients standardisés plus faibles.

3.1.3 La corrélation

La corrélation est une mesure statistique qui quantifie la relation linéaire entre deux variables. Elle est utilisée couramment pour décrire des relations simples entre variables, indépendamment de la cause et de l'effet. On utilise une mesure sans unité appelée coefficient de corrélation pour décrire les corrélations, qui varie entre -1 et +1 et est généralement représenté par le symbole r . La significativité statistique est indiquée par une valeur p . Par conséquent, les corrélations sont généralement exprimées à l'aide de deux chiffres : r et p .

Lorsque le coefficient de corrélation r est proche de zéro, cela indique une faible relation linéaire entre les variables. Les valeurs positives de r dénotent une corrélation positive, ce qui signifie que lorsque les valeurs d'une variable augmentent, les valeurs de l'autre variable ont tendance à augmenter également. Inversement, les valeurs négatives de r indiquent une corrélation négative, ce qui suggère que lorsque les valeurs d'une variable augmentent, les valeurs de l'autre variable ont tendance à diminuer. Une fois qu'une corrélation significative est établie, il est également important de considérer sa force. Une corrélation positive parfaite est représentée par une valeur de 1, indiquant une relation linéaire étroite et constante entre les variables. De même, une corrélation négative parfaite est représentée par une valeur de -1, ce qui signifie qu'il existe une relation linéaire étroite et inverse entre les variables.

Même si deux variables sont fortement corrélées, cela ne signifie pas nécessairement qu'il existe une relation de cause à effet entre elles.

3.1.4 Détection de la causalité temporelle : Évaluation par le test de Granger

La causalité est le concept qui décrit une relation de cause à effet entre deux variables. Lorsqu'une variable A est considérée comme la cause d'une variable B , cela signifie que des changements ou des variations dans A ont une influence sur les valeurs ou les comportements de B .

L'inférence de Granger est une méthode statistique utilisée pour évaluer la causalité entre deux séries chronologiques, c'est-à-dire des ensembles de données qui sont mesurées dans le temps. Elle est basée sur le principe que si une variable X aide à prédire une autre variable Y , en tenant compte de l'historique des deux variables, alors on peut dire que X a une certaine causalité sur Y .

Le test de causalité de Granger, implémenté dans la fonction `grangercausalitytests` du module `statsmodels.tsa.stattools` en Python, est utilisé pour évaluer la causalité entre deux séries chronologiques. Ce test suit une approche basée sur la régression pour évaluer si l'inclusion des valeurs passées d'une série aide à prédire de manière significative l'autre série.

La fonction `grangercausalitytests` prend comme paramètre le nombre maximum de retards (lags) à inclure dans le modèle qui est nommé `maxlag`, il détermine la portée de l'analyse temporelle pour évaluer la causalité. La fonction retourne un dictionnaire contenant les résultats du test de causalité pour différents retards (lags). Chaque clé du dictionnaire correspond à un retard spécifique. Pour chaque retard, le dictionnaire de résultats contient les statistiques de test, la valeur p et d'autres informations. La statistique de test la plus couramment utilisée est la statistique F, qui est associée à une valeur p. Une valeur inférieure à un seuil prédéfini 0,05 qui suggère une preuve de causalité statistiquement significative entre les séries chronologiques.

3.2 Les séries temporelles

3.2.1 Définition

Les séries temporelles sont des ensembles de données statistiques recueillies à des moments différents et organisées dans l'ordre chronologique. Elle se concentre sur l'évolution d'un phénomène au fil du temps, dans le but de le décrire, de l'expliquer et de le prévoir dans le futur. Elles sont souvent représentées mathématiquement sous forme d'une séquence de valeurs numériques ordonnées dans le temps. Cette séquence peut être représentée par une fonction mathématique telle que :

$$X(t) = \{X_1, X_2, \dots, X_n\}$$

Où $X(t)$ est la séquence de valeurs de la série temporelle à différents moments dans le temps (t) , et X_1, X_2, \dots, X_n sont les valeurs numériques de la série temporelle à chaque instant t . La notation $\{ \}$ indique que les valeurs sont ordonnées dans le temps. La séquence $X(t)$ est donc une fonction qui peut être évaluée à différents instants dans le temps pour obtenir les valeurs correspondantes de la série temporelle (Pour plus de détails, vous pouvez consulter l'annexe B)

3.2.2 La notion de stationnarité

La stationnarité d'une série temporelle fait référence à une propriété importante de cette série, où les propriétés statistiques de la série restent constantes au fil du temps. Plus précisément, une série temporelle est considérée comme stationnaire si elle satisfait les trois critères suivants :

- **Stationnarité en termes de moyenne** : La moyenne de la série temporelle reste constante dans le temps. Cela signifie que la tendance générale de la série ne change pas de manière significative au fil du temps.
- **Stationnarité en termes de variance** : La variance de la série temporelle reste constante dans le temps. Cela implique que l'amplitude des fluctuations de la série reste stable au fil du temps.
- **Stationnarité en termes de covariance** : La covariance entre les observations de la série temporelle à différents moments reste constante. Cela signifie que la relation

entre les observations ne dépend pas du temps, et qu'il n'y a pas de dépendance temporelle significative.

La stationnarité est une condition importante dans l'analyse des séries temporelles car elle permet d'appliquer des modèles et des techniques statistiques spécifiques qui supposent une telle propriété. Voici trois techniques couramment utilisées pour l'évaluer :

- **Analyse visuelle** : Cette technique consiste à examiner graphiquement la série temporelle. Si la série semble présenter des changements systématiques au fil du temps, elle est susceptible de ne pas être stationnaire.
- **Méthode de division du jeu de données** : Cette technique consiste à diviser le jeu de données en deux parties égales. Ensuite, vous pouvez calculer la moyenne et la variance pour chaque partie et les comparer. Si les moyennes et les variances diffèrent considérablement entre les deux parties, cela suggère une non-stationnarité. En revanche, si les moyennes et les variances sont similaires, cela indique une stationnarité probable de la série temporelle.
- **Test de Dickey-Fuller augmenté (ADF)** : Le test de Dickey-Fuller augmenté est une méthode statistique largement utilisée pour évaluer la stationnarité d'une série temporelle. Ce test examine si une série temporelle suit un processus stochastique avec une racine unitaire, ce qui indiquerait une non-stationnarité. Si la valeur p du test est inférieure à un certain seuil généralement 0,05, on rejette l'hypothèse nulle de non-stationnarité et on conclut que la série est stationnaire.

3.2.3 La décomposition d'une série temporelle

Les composantes d'une série chronologique, également appelées facteurs de décomposition, sont des éléments qui aident à mieux comprendre son fonctionnement et son évolution dans le temps. Elles permettent d'analyser les tendances et les variations saisonnières, de modéliser son comportement et de comprendre les différents facteurs qui l'influencent. Les principales composantes d'une série chronologique sont :

- **Tendance** : Il s'agit de l'évolution à long terme de la série temporelle, qui peut être croissante, décroissante ou stable.
- **Saisonnalité** : Il s'agit des fluctuations régulières qui se produisent à intervalles réguliers au cours de l'année.
- **Cycle** : Il s'agit des fluctuations qui se produisent sur une période plus longue que la saisonnalité, généralement de plusieurs années. Ces fluctuations peuvent être dues à des facteurs économiques, sociaux ou politiques.
- **Bruit** : Il s'agit de la variabilité aléatoire de la série temporelle, qui ne peut pas être expliquée par les autres composantes. Le bruit peut être dû à des erreurs de mesure, à des événements imprévus ou à des fluctuations naturelles.

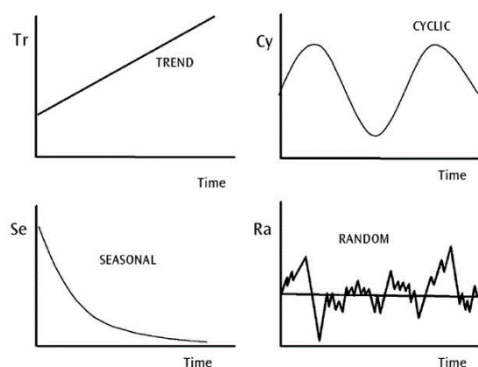


Figure 3.1 – Les composantes d'une série chronologique, source : [researchgate.net](https://www.researchgate.net) récupérée le 18 avril 2023

En fonction des caractéristiques spécifiques de la série chronologique et des objectifs de modélisation, le choix entre un modèle additif ou multiplicatif peut être déterminant pour obtenir une représentation précise et efficace de la série temporelle.

1- Le modèle additif : Dans ce type de modèle, les composantes sont combinées par addition et la tendance est linéaire. Les changements au fil du temps sont toujours effectués d'une même quantité. La saisonnalité est supposée avoir la même fréquence (largeur des cycles) et amplitude (hauteur des cycles) au fil du temps.

$$y(t) = \text{Niveau} + \text{Tendance} + \text{Saisonnalité} + \text{Bruit}$$

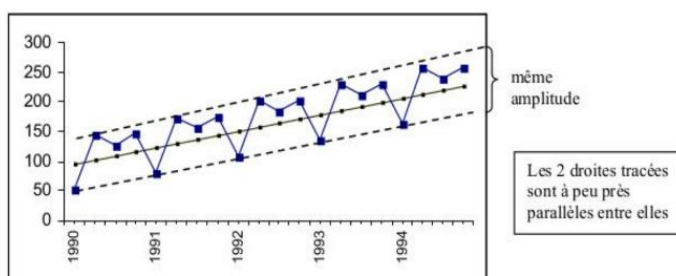


Figure 3.2 – Exemple de schéma additif, source : [memoireonline](https://www.memoireonline.com) récupérée le 30 avril 2023

2- Le modèle multiplicatif : Dans ce type de modèle, les composantes sont combinées par multiplication et sont non linéaires. Les changements ont tendance à augmenter ou diminuer au fil du temps.

$$y(t) = \text{Niveau} * \text{Tendance} * \text{Saisonnalité} * \text{Bruit}$$

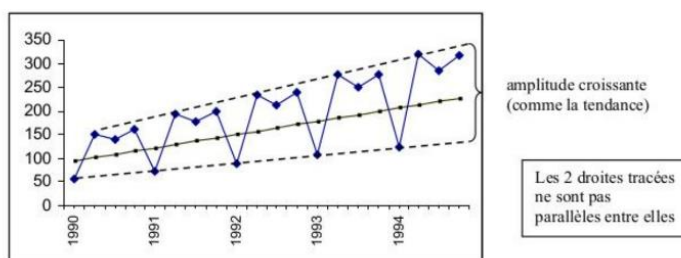


Figure 3.3 – Exemple de schéma multiplicatif, source : [memoireonline](https://www.memoireonline.com) récupérée le 30 avril 2023

3.2.4 Les transformations de séries temporelles

Les transformations de séries temporelles sont des techniques courantes pour modifier la structure de la série temporelle et faciliter son analyse et sa modélisation. Les principales transformations de série chronologique sont les suivantes :

1. **Différenciation** : Il s'agit de la différence entre une observation et l'observation précédente, utilisée pour éliminer la tendance de la série temporelle. Mathématiquement, la différenciation peut être représentée par la formule $\Delta X_t = X_t - X_{t-1}$.
2. **Logarithmique** : Il s'agit de la transformation de la série temporelle en prenant le logarithme de chaque observation, utilisée pour stabiliser la variance et pour modéliser les tendances multiplicatives. Mathématiquement, la transformation logarithmique peut être représentée par la formule $\log(X_t)$.
3. **Normalisation** : Il s'agit de la transformation de la série temporelle en la divisant par une mesure de sa variabilité, telle que l'écart-type ou la plage de variation. Cela permet de comparer les observations entre elles et de faciliter la modélisation. Mathématiquement, la normalisation peut être représentée par la formule $(X_t - \text{moyenne})/\text{écart-type}$.
4. **Agrégation** : Il s'agit de la combinaison de plusieurs observations en une seule période, telle que la moyenne ou la somme des observations, utilisée pour réduire le bruit et pour modéliser les tendances à long terme.

Les transformations de série chronologique permettent de rendre la structure de la série temporelle plus simple et plus compréhensible, tout en préservant les propriétés statistiques importantes. Elles facilitent également la modélisation et la prévision des tendances futures de la série temporelle.

3.2.5 L'étude des séries temporelles

L'étude des séries temporelles a pour objectif principal de comprendre les tendances temporelles et les comportements de données au fil du temps. Les principaux objectifs de l'étude des séries chronologiques comprennent :

1. **Description** : Décrire les caractéristiques principales de la série temporelle, telles que les tendances, les fluctuations saisonnières, les variations aléatoires, les cycles, etc.
2. **Modélisation** : Développer des modèles statistiques qui permettent de quantifier les tendances temporelles et de prévoir les valeurs futures de la série temporelle.
3. **Prévision** : Utiliser les modèles statistiques pour prédire les valeurs futures de la série temporelle et pour évaluer l'incertitude associée à ces prévisions.
4. **Intervention** : Évaluer l'impact des interventions ou des événements sur la série temporelle et déterminer si ces interventions ont eu un effet significatif sur les tendances ou les comportements de données.
5. **Contrôle** : Utiliser les modèles statistiques pour contrôler et surveiller les performances d'un processus au fil du temps, en détectant les anomalies et les déviations par rapport aux valeurs attendues.

Cependant, l'un des objectifs les plus importants de l'étude des séries chronologiques est la prévision.

3.3 La prévision des données

3.3.1 Définition

La prévision représente « l'ensemble des techniques ayant pour but d'évaluer la situation économique à une échéance plus ou moins lointaine », source : [larousse](#) .

La prévision est une méthode qui consiste à prédire des événements futurs ou des tendances en utilisant des données historiques et des méthodes statistiques. Elle permet d'estimer de manière quantitative et qualitative l'occurrence d'un événement futur en se basant sur les données passées et en utilisant des techniques d'analyse et de modélisation prédictive.

3.3.2 Les méthodes de prévision

1- Les méthodes qualitatives :

Les techniques qualitatives se basent principalement sur les avis et les jugements de personnes ayant une connaissance approfondie du contexte de la variable à prédire, ou sur des comparaisons avec des contextes similaires. Ces méthodes sont souvent appliquées lorsque des données historiques ne sont pas disponibles. Ex : changements économiques abruptes. Parmi ces techniques il existe :

- **Le sondage** : Cette méthode est utilisée pour estimer une quantité inconnue en se basant sur les réponses d'un groupe sélectionné de personnes.
- **L'analyse du marché** : Cette analyse permet de comprendre les tendances actuelles et futures du marché, les préférences des consommateurs, la concurrence et les opportunités de croissance.
- **Les jugements d'experts** : Ces méthodes de prévision sont basées sur l'expertise et l'opinion de spécialistes pour établir des prévisions. Parmi ces méthodes, on cite : le panel d'experts qui consiste à recueillir les opinions de plusieurs spécialistes d'un domaine, la méthode Delphi qui permet à des experts différents de formuler leur prévision de manière anonyme et itérative.

Cependant, l'inconvénient majeur des méthodes qualitatives en général est que la subjectivité inhérente peut introduire des biais dans les prévisions, car les opinions des différents experts peuvent s'influencer mutuellement. De manière générale, ces méthodes ont tendance à être plus subjectives et moins fiables que les méthodes quantitatives.

2- Les méthodes quantitatives :

La méthode quantitative repose sur l'analyse de données historiques, en utilisant des techniques basées sur des modèles statistiques ou de machine learning pour améliorer l'exactitude des prévisions.

La qualité de la prévision peut être mesurée en comparant les valeurs prédites avec les valeurs réelles de la série chronologique à des moments futurs. L'erreur de prédiction, notée $ET(h)$, est la différence entre la valeur réelle X_{T+h} et la valeur prédite $\hat{X}_T(h)$. Un intervalle de confiance peut également être calculé pour estimer la précision de la prévision.

3.4 Exploration, Analyse et Prédiction des Séries Temporelles

3.4.1 Exploration et analyse des données : Synergie entre SQL et Power BI

Avec la quantité croissante de données générées chaque jour par la banque, il est crucial d'avoir une gestion efficace des données et des outils adaptés pour tirer pleinement parti de cette richesse d'information. Parmi les outils les plus performants utilisés dans le secteur bancaire pour traiter de grandes quantités de données, Teradata se démarque. Teradata est une plateforme de gestion de données puissante et évolutive qui permet aux banques d'analyser rapidement et efficacement des pétaoctets de données.

Un élément clé de l'écosystème Teradata est le langage SQL. C'est un langage de programmation standardisé conçu spécifiquement pour la gestion et la manipulation des bases de données relationnelles. Il offre une syntaxe claire et facile à comprendre, ce qui en fait un outil essentiel pour interroger, filtrer et agréger les données. Avec SQL, les banques peuvent effectuer des requêtes complexes sur leurs vastes ensembles de données, extraire des informations pertinentes et prendre des décisions basées sur des analyses approfondies. Il permet d'effectuer des opérations telles que la sélection, la jointure, le regroupement et l'agrégation des données, offrant ainsi une flexibilité et une puissance d'analyse essentielles dans le secteur bancaire.

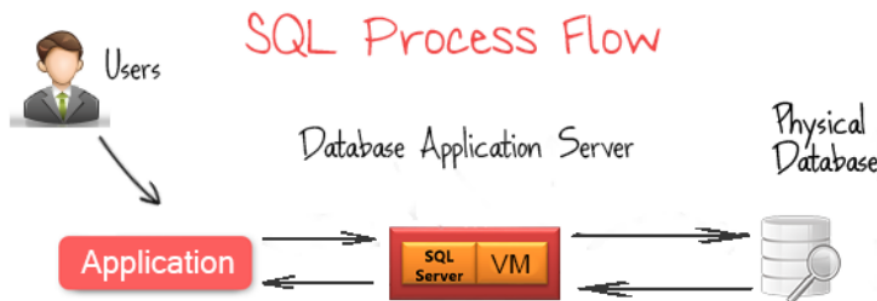


Figure 3.4.1 – Diagramme de flux du processus SQL, source : [way2tutorial](#) récupérée le 2 mai 2023

En complément de l'exploration des données massives, la visualisation interactive joue un rôle essentiel dans la compréhension et l'analyse des informations. Elle permet aux professionnels de la banque de représenter graphiquement les données de manière intuitive et de découvrir des insights significatifs.

Power BI, un outil d'analyse et de visualisation de données qui permet aux utilisateurs de créer des tableaux de bord interactifs, des rapports visuels et des visualisations dynamiques à partir de différentes sources de données, y compris les données massives stockées dans des systèmes tels que Teradata. Power BI offre une interface conviviale et intuitive, permettant aux utilisateurs de manipuler, filtrer et explorer les données en temps réel. Les fonctionnalités de visualisation avancées, telles que les graphiques interactifs, les cartes géographiques, les

diagrammes à secteurs et les tableaux de bord personnalisables, offrent une expérience visuelle riche et immersive.

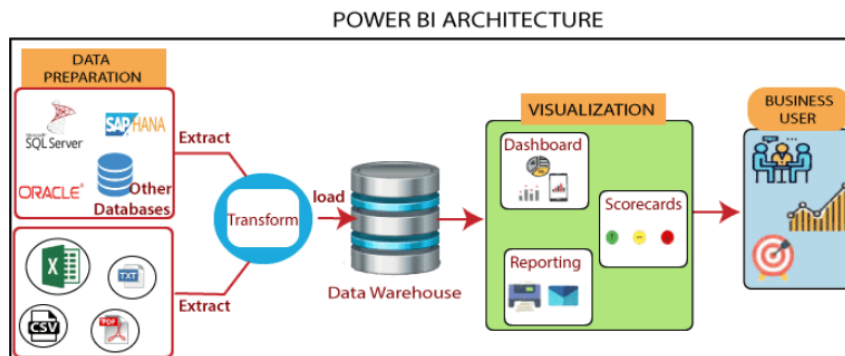


Figure 3.5 – Diagramme d'architecture de Power BI, source : [informationit27.medium](https://www.informationit27.medium.com) récupérée le 2 mai 2023

3.4.2 Pré-étude de la nature de la série temporelle pour le choix des modèles

Dans le cadre de notre projet, nous avons réalisé une pré-étude pour comprendre la nature de notre série temporelle, qui concerne les encours sur une période mensuelle allant de janvier 2012 à mars 2023. Cette pré-étude, dont les résultats sont présentés dans le chapitre 4 de notre solution, nous a permis d'obtenir des informations préliminaires sur la série et de guider notre choix de modèles. Au cours de cette pré-étude, nous avons effectué une analyse de la décomposition temporelle de la série afin d'identifier les différentes composantes telles que les tendances, les saisons et les résidus. Nous avons observé qu'il existe une seule saisonnalité mensuelle dans la série. Nous avons également noté un schéma d'augmentation suivie d'une baisse en 2019, ce qui indique un changement de tendance au cours de cette période spécifique. Ces observations préliminaires nous ont permis de mieux comprendre les caractéristiques de la série et de prendre en compte ces éléments lors du choix des modèles.

À la suite de cette pré-étude, nous avons sélectionné des modèles adaptés à notre série temporelle en tenant compte de sa saisonnalité mensuelle et de son changement de tendance en 2019. Nous avons aussi pris en compte le volume de notre dataset lors du choix des modèles, cette dernière contient un nombre limité de lignes, l'utilisation d'un modèle complexe qui exige une quantité de données plus importante comme LSTM peut être confronté à des problèmes de sur apprentissage où le modèle mémorise les données d'entraînement au lieu de généraliser les motifs sous-jacents. Les modèles ARIMA, SARIMA, XGBoost et Fb Prophet ont été retenus en raison de leurs capacités à modéliser à la fois les tendances et les saisons, ainsi que leurs performances prouvées dans des situations similaires.

3.4.3 Prévision des séries temporelles : Modèles statistiques, apprentissage automatique et l'approche Prophet

1- Les Modèles statistiques

Les modèles statistiques de prédiction sont des méthodes analytiques qui utilisent des techniques statistiques pour modéliser et prévoir les tendances et les modèles dans les données

temporelles. Ces modèles sont généralement basés sur des hypothèses statistiques et sont utilisés pour capturer les relations et les motifs complexes présents dans les séries temporelles. Les modèles ARIMA et SARIMA sont particulièrement adaptés à notre cas avec une seule saisonnalité mensuelle en raison de leurs capacités à modéliser à la fois les tendances et les motifs saisonniers. Cependant, cela ne signifie pas que les autres modèles de prévision ne peuvent pas être utilisés.

D'autres modèles existent, tels que le modèle de régression linéaire ou le modèle de régression polynomiale, cependant, ces modèles peuvent ne pas prendre explicitement en compte la saisonnalité de la série, ce qui peut entraîner une performance moins précise dans la prédiction des valeurs saisonnières. Les modèles VARMA, VAR et VARMAX sont plus adaptés aux séries temporelles multivariées, où plusieurs variables sont prises en compte pour la modélisation et la prédiction. Si on dispose uniquement d'une seule série temporelle, ces modèles ne sont pas appropriés.

Le modèle ARIMA :

Le modèle ARIMA est un modèle statistique utilisé pour modéliser et prévoir les séries temporelles. Il est composé de trois composantes principales : l'auto régression (AR), la moyenne mobile (MA) et l'intégration (I).

La composante autorégressive (AR) du modèle ARIMA utilise les valeurs passées de la série temporelle pour prédire les valeurs futures. Elle suppose que la valeur courante de la série dépend linéairement des valeurs passées et est exprimée en fonction de ces valeurs avec des coefficients spécifiques. La composante de moyenne mobile (MA) du modèle ARIMA utilise les erreurs passées de prédiction pour modéliser les tendances et les motifs résiduels de la série temporelle. Elle suppose que la valeur courante de la série dépend linéairement des erreurs de prédiction passées et est exprimée en fonction de ces erreurs avec des coefficients spécifiques. La composante d'intégration (I) du modèle ARIMA est utilisée pour rendre la série temporelle stationnaire en différenciant la série à des fins de modélisation. La stationnarité est une propriété importante dans les modèles ARIMA, car elle permet de capturer les tendances et les motifs de manière plus précise. ARIMA est un modèle caractérisé par 3 paramètres : p représente l'ordre du terme AR, d représente le nombre de différenciations nécessaires pour rendre la série temporelle stationnaire et q représente l'ordre du terme MA.

En combinant ces trois composantes, le modèle ARIMA est capable de capturer les tendances, les motifs saisonniers et les variations aléatoires dans une série temporelle, ce qui le rend adapté à la modélisation et à la prévision des séries temporelles avec des structures complexes (Pour plus de détails sur ce modèle, voir l'annexe D)

Le modèle SARIMA :

Le modèle SARIMA est une extension du modèle ARIMA. Le terme S dans SARIMA fait référence à la composante saisonnière du modèle. Il inclut des termes supplémentaires d'autorégression, de moyenne mobile et d'intégration pour capturer les motifs saisonniers. Ces termes permettent de modéliser les dépendances saisonnières et d'ajuster le modèle aux variations régulières observées dans la série. Il peut capturer à la fois les variations à court terme et à long terme, offrant ainsi des prévisions plus précises pour les séries temporelles présentant des comportements saisonniers.

2- Les modèles de machine learning

Le machine learning, ou apprentissage automatique en français, est une branche de l'intelligence artificielle qui se concentre sur le développement de techniques permettant aux systèmes informatiques d'apprendre à partir des données et d'améliorer leurs performances sans être explicitement programmés (Pour des informations supplémentaires concernant ce modèle, veuillez consulter l'annexe C). En machine learning, on utilise des modèles, qui sont des algorithmes ou des structures mathématiques, pour apprendre à partir des données et effectuer des tâches spécifiques telles que la classification, la prédiction, le clustering, la reconnaissance d'image, la traduction automatique, etc. Parmi les nombreux modèles de machine learning utilisés pour la prédiction, XGBoost est largement reconnu et très utilisé.

XGBoost :

XGBoost est un algorithme d'apprentissage ensembliste qui combine plusieurs modèles d'arbres de décision pour améliorer les performances de prédiction. Il utilise une technique appelée boosting qui entraîne plusieurs arbres de décision successivement, en accordant une attention particulière aux erreurs commises par les modèles précédents. Il est extrêmement populaire dans les problématiques de data science en raison de sa robustesse et de sa vitesse de calcul. En terme technique, XGBoost utilise un objectif de perte personnalisable et un processus itératif pour entraîner les arbres de décision. Il utilise également des techniques d'élagage et de régularisation pour éviter le surajustement et améliorer la généralisation du modèle. Grâce à sa combinaison de plusieurs arbres de décision, XGBoost est capable de capturer des motifs complexes dans les données et de fournir des prédictions précises.

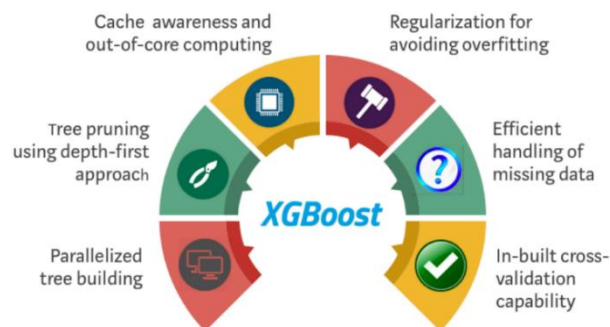


Figure 3.6 - Un schéma qui illustre les avantages du modèle XGboost, source : [larevueia](#) récupérée le 9 mai 2023.

3- Fb prophet :

Prophet, développé par Facebook, est un modèle de prévision des séries temporelles largement utilisé en raison de sa simplicité, de sa flexibilité et de ses performances. Il a gagné en popularité en raison de son approche novatrice pour modéliser les tendances saisonnières et les changements de tendance. Sur le plan technique, Prophet utilise un modèle additif qui combine les composantes saisonnières et les tendances linéaires.

Prophet est capable de détecter automatiquement les changements de tendance, d'ajuster la saisonnalité en fonction des motifs saisonniers observés. Il estime les paramètres du modèle à

l'aide d'une méthode d'optimisation basée sur la minimisation de l'erreur quadratique moyenne entre les valeurs observées et prédites.

L'un des principaux avantages de Prophet est sa facilité d'utilisation. Il propose une interface conviviale avec des paramètres intuitifs, ce qui le rend accessible même aux utilisateurs non experts en statistiques ou en modélisation. De plus, Prophet est capable de gérer efficacement les données manquantes, les valeurs aberrantes et les fluctuations. Prophet est particulièrement adapté aux séries temporelles avec des tendances non linéaires, des changements de tendance et des saisons complexes. Dans notre cas, puisque la série temporelle présente une seule saisonnalité mensuelle et des changements de tendance observés en 2019, Prophet est un choix approprié. Il sera en mesure de modéliser la saisonnalité mensuelle et de détecter les changements de tendance de manière automatique et précise.

3.4.4 Evaluation des modèles

Pour évaluer et comparer les performances des modèles de prédiction, nous avons choisi plusieurs métriques d'erreur afin d'obtenir une vision complète de leur performance, en prenant en compte les caractéristiques spécifiques de nos données. Étant donné que nos données présentent des valeurs élevées et un faible volume, nous avons opté pour les métriques suivantes : le NRMSE, le WMAPE et le coefficient de détermination R-squared (Vous trouverez les formules mathématiques de ces paramètres en annexe A)

Le NRMSE est une mesure de l'erreur quadratique moyenne normalisée par rapport à la plage des valeurs réelles. Il nous permet d'évaluer la précision des prédictions et de comparer les modèles en tenant compte de l'échelle des données. Un NRMSE plus faible indique une meilleure précision du modèle.

Le WMAPE est une version pondérée du MAPE (c'est une mesure de l'erreur absolue moyenne en pourcentage par rapport aux valeurs réelles) qui tient compte de la priorité relative des différentes observations. Cela est particulièrement important lorsque certaines observations sont considérées comme plus importantes que d'autres ou lorsque nous travaillons avec un faible volume de données. Le WMAPE nous permet d'évaluer l'erreur moyenne en pourcentage en prenant en compte la pondération des observations.

Le coefficient de détermination R-squared mesure la proportion de la variance totale des valeurs réelles qui est expliquée par le modèle. Il indique à quel point les prédictions s'ajustent aux données réelles. Un coefficient de détermination R-squared proche de 1 indique un bon ajustement du modèle aux données, tandis qu'une valeur proche de 0 indique un mauvais ajustement.

En utilisant ces métriques, nous sommes en mesure d'évaluer différents aspects de la performance des modèles de prédiction, tels que l'erreur absolue, l'erreur relative et l'ajustement global aux données réelles. Cela nous permet de prendre des décisions éclairées quant au choix du modèle le plus approprié pour notre étude, tout en comprenant l'adéquation de la prédiction aux caractéristiques spécifiques de nos données.

3.5 Interface utilisateur

3.5.1 Définition

Une interface utilisateur interactive connectée à un backend est un système qui permet aux utilisateurs d'interagir avec une application ou un service via une interface visuelle conviviale. Cette interface utilisateur offre des fonctionnalités interactives telles que des formulaires, des boutons et des menus déroulants, permettant aux utilisateurs de fournir des données et de déclencher des actions.

3.5.2 Le backend et le frontend

L'interface utilisateur est étroitement liée aux concepts de backend et de frontend dans le développement logiciel. Le backend désigne la partie d'une application qui gère les opérations côté serveur, telles que la logique métier, la gestion des données et l'exécution des requêtes.

La partie frontend de l'interface utilisateur est responsable de la présentation visuelle et de l'interaction avec les utilisateurs. Elle est généralement développée en utilisant des technologies telles que HTML, CSS et JavaScript pour créer une expérience utilisateur esthétiquement agréable et intuitive.

3.5.3 La liaison entre le backend et le frontend

Pour fournir une expérience utilisateur complète, le frontend et le backend sont deux composants essentiels qui travaillent en tandem. Pour les connecter de manière efficace, on utilise des API et des requêtes HTTP. Flask, un framework web léger et flexible, peut être utilisé dans le backend pour faciliter cette connexion. Flask offre de nombreux avantages lorsqu'il s'agit de relier le frontend au backend. Tout d'abord, il permet de définir facilement des routes pour les différentes fonctionnalités de l'application. Les routes définies dans Flask indiquent les URL et les méthodes HTTP associées à chaque API exposée par le backend.

Voici les étapes de liaison entre le frontend et le backend en utilisant Flask :

- Définition des routes Flask :

Dans Flask, les routes sont les URL vers lesquelles le frontend envoie des requêtes.

- Création d'une API Flask :

Une API est une interface logicielle qui définit les méthodes et les structures de données que le frontend peut utiliser pour interagir avec le backend. En utilisant Flask, vous pouvez créer une API en exposant des routes spécifiques qui représentent les fonctionnalités ou les ressources disponibles dans votre application.

- Traitement des requêtes dans les vues Flask :

Les vues Flask sont des fonctions ou des méthodes associées à chaque route définie. Ces vues sont responsables du traitement des requêtes et de la génération des réponses correspondantes. Dans une vue Flask, vous pouvez accéder aux données de la requête, effectuer des opérations, interagir avec une base de données et retourner des réponses appropriées.

- Gestion des méthodes HTTP :

Les méthodes HTTP, telles que GET, POST, PUT, DELETE, etc., permettent au frontend d'effectuer différentes actions sur les ressources exposées par le backend. Vous pouvez spécifier les méthodes autorisées pour chaque route Flask en utilisant le paramètre (methods) lors de la définition de la route.

- Envoi de requêtes HTTP depuis le frontend :

Le frontend envoie des requêtes HTTP au backend Flask en utilisant des méthodes spécifiques, telles que `fetch()` en JavaScript. Les requêtes peuvent être envoyées à l'URL de la route Flask correspondante, avec les paramètres et les données nécessaires.

- Réception et traitement des requêtes dans les vues Flask :

Lorsque le backend Flask reçoit une requête sur une route spécifique, la vue Flask associée à cette route est appelée. La vue peut accéder aux données de la requête, effectuer des opérations, interagir avec la dataset, et générer une réponse appropriée.

- Génération de réponses dans les vues Flask :

Les vues Flask sont responsables de la génération des réponses à renvoyer au frontend. Selon les besoins, vous pouvez retourner des réponses au format JSON, HTML, ou d'autres formats spécifiques.

- Réception et traitement des réponses côté frontend :

Une fois que le frontend reçoit la réponse du backend Flask, il peut traiter les données reçues et les utiliser pour mettre à jour l'interface utilisateur. Cela peut impliquer l'affichage des données, la modification de l'état de l'application ou la gestion d'erreurs éventuelles.

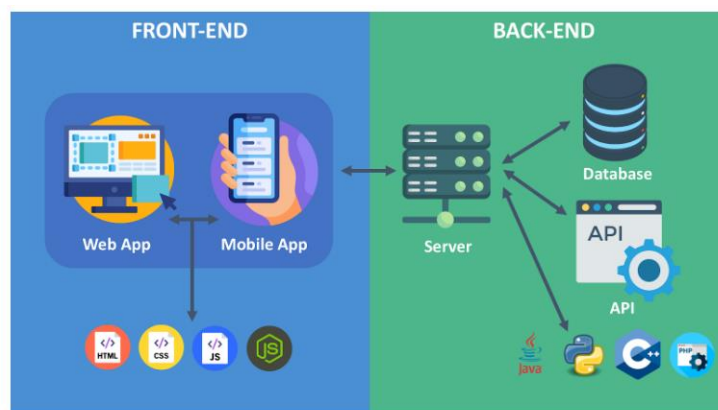


Figure 3.7 – Un schéma qui illustre la relation entre le backend et le frontend, source : techterrotor.com récupérée le 10 mai 2023.

Vous trouverez la présentation des outils utilisés en annexe E.

Après avoir exposé les différentes méthodes utilisées pour résoudre la problématique, nous aborderons, dans le chapitre suivant, les détails de notre solution et présenterons les résultats obtenus grâce à notre approche.

Chapitre 4 : Conception de la solution

Chapitre 4

Conception de la solution

Après avoir défini les outils nécessaires pour la résolution de notre problématique et décrypté le produit net bancaire. Dans ce qui suit, nous allons présenter notre démarche de solution en commençant par la collecte des données.

4.1 La collecte de données

Pour la première partie de notre solution, nous avons extrait les informations relatives au PNB, aux encours, aux mouvements créditeurs, aux importations, aux exportations et aux secteurs d'activité pour les clients de type Corporate en utilisant des requêtes SQL sur le système de gestion de bases de données relationnelles Teradata. Ces données ont été sélectionnées à partir de la table 'Client Table' fournie par la direction de la finance, en filtrant les résultats pour l'année 2022. Nous avons choisi l'année 2022 car cette année est la plus récente pour laquelle nous avons pu obtenir des données complètes sur les clients de type Corporate. Cette opération nous a permis de générer environ 3402 lignes de données, constituant ainsi la base de notre étude visant à prouver la relation forte entre le PNB et les encours.

```
SELECT PNB, Encours, MouvementsCrediteurs, Import, Export, SecteursActivite  
FROM ClientTable  
WHERE Typeclient = 'Corporate' AND Year = 2022;
```

Figure 4.1 – Requête SQL pour l'extraction des données de PNB, encours et secteurs d'activité des clients Corporate pour l'année 2022

4.2 Exploration de données

Dans la partie exploration des données, nous avons utilisé Power BI pour créer un tableau de bord interactif permettant de mieux comprendre les données collectées auparavant. Le tableau de bord contenait plusieurs visualisations et graphiques, tels que des graphiques en surface, un graphique entonnoir et des cartes numériques pour chaque variable. Ces visualisations ont permis d'obtenir une vue d'ensemble des données et de détecter d'éventuelles tendances. Par exemple, nous avons utilisé un graphique entonnoir pour comparer le PNB de différents secteurs d'activité, ce qui a révélé des différences significatives entre les secteurs.

Le secteur d'industrie pharmaceutique et médical (secteur 4 sur la figure), le secteur agroalimentaire(1) sont des secteurs phares pour SGA en raison de leur stabilité, leur importance vitale, leurs besoins en investissements et de leur croissance constante. La demande de produits pharmaceutiques, d'aliments et de services médicaux demeure essentielle pour les consommateurs, ce qui permet aux entreprises de ces secteurs de maintenir une certaine solidité

financière, cette stabilité assure une source régulière d'encours pour les banques. Ces secteurs nécessitent souvent des financements importants et des investissements significatifs en capital pour la production, la recherche et le développement. Le secteur énergétique(6), le transport-logistique(7), l'immobilier(2), la sous-traitance industrielle(3) et les télécommunications(5) sont également des secteurs importants en termes d'encours, bien que leur taux d'encours puisse être légèrement inférieur aux secteurs phares. Ils nécessitent des financements pour leurs activités d'exploration, de développement, de construction et de services, ce qui contribue à la génération d'encours. Ces secteurs sont essentiels pour l'approvisionnement en énergie, le mouvement des marchandises, le développement immobilier, la fabrication et les services de communication. En somme, bien que les taux d'encours puissent varier, ces secteurs restent des acteurs importants dans le paysage bancaire grâce à leurs contributions économiques significatives (Vous trouverez la définition de chaque secteur en annexe F).

Le graphique en entonnoir du tableau de bord suivant illustre la prédominance des secteurs pharmaceutique, agroalimentaire et médical en termes d'encours, suivis par les autres secteurs. Les numéros attribués aux secteurs dans le graphique correspondent aux mêmes secteurs mentionnés et numérotés dans l'analyse précédente.

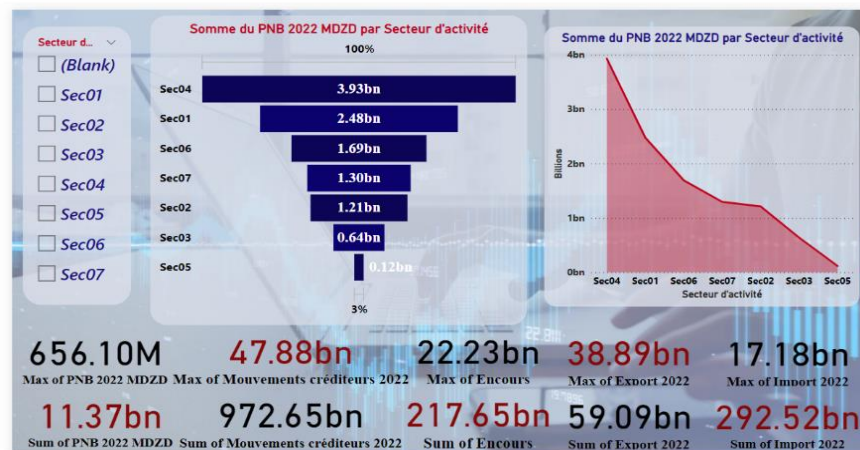


Figure 4.2–Page d’analyse du PNB - rapport Power BI

Nous avons également utilisé des graphiques en surface pour visualiser l'évolution de chaque variable de notre dataset. Ces graphiques nous ont permis de mieux comprendre la problématique et de décider de l'utilisation de la régression multiple pour étudier l'impact de ces variables sur le PNB.

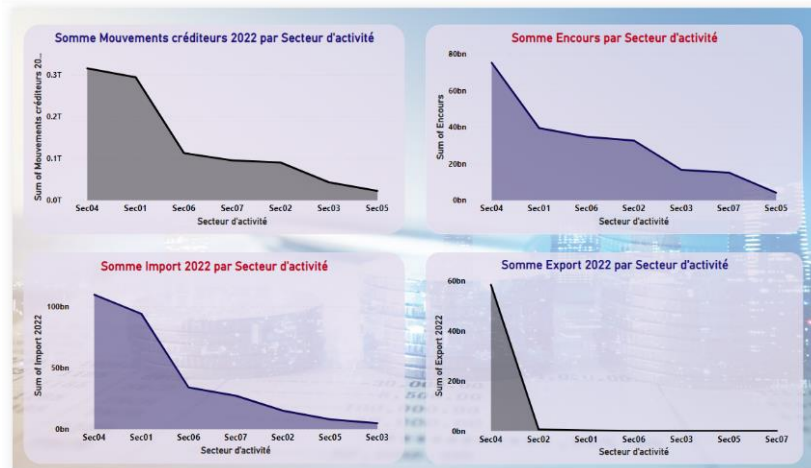


Figure 4.3 –Page d’analyse des différentes variables par secteur d’activité - rapport Power B

En outre, nous avons utilisé une heatmap pour confirmer les résultats de notre analyse de corrélation. Cette heatmap a montré une forte corrélation positive de 98% entre les encours et le PNB, comme indiqué par la couleur la plus foncée. Les autres variables étaient également corrélées, mais dans une moindre mesure : 81% pour les importations, 76% pour les mouvements créditeurs, tandis que les exportations ne présentaient pas de corrélation significative (-3%).

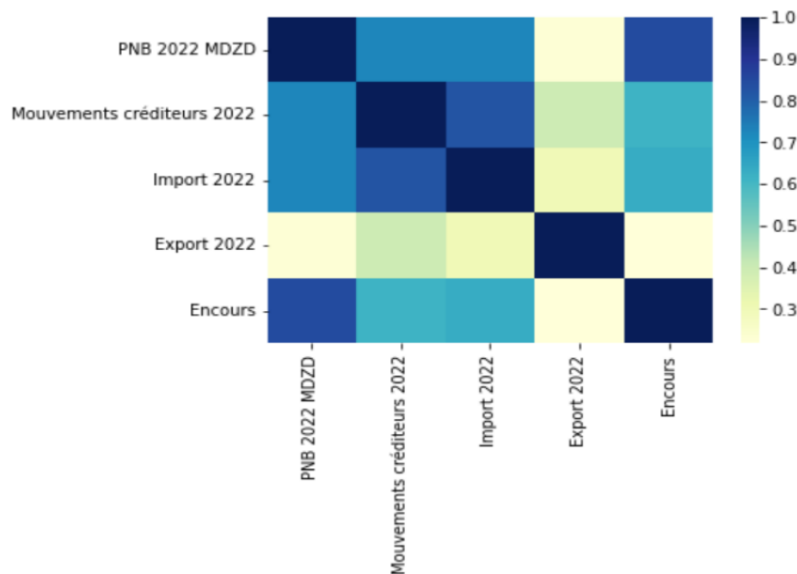


Figure 4.4 –la matrice de corrélation

4.3 Étude approfondie de la relation entre les encours et le PNB

Après avoir collecté et exploré les données nécessaires, il est temps de prouver la relation forte entre les encours et le PNB. Pour cela, une étude et une analyse statistique rigoureuse doivent

être menées. La première étape de cette analyse est l'utilisation de la régression multiple pour observer le lien entre les différentes variables, notamment les encours et le PNB.

4.3.1 Régression multiple et coefficient standardisé

Les étapes clés de l'analyse de régression multiple sur python :

- Importer les données : Après avoir Chargé les bibliothèques essentielles telles que Pandas, Numpy, Matplotlib et statsmodels.api, nous avons importé les données à partir de la source Excel. Notre dataset contient 3402 lignes et 8 colonnes comme suit :

La variable	Le nom de la colonne	Type de la variable
Identifiant du client corporate	ID CLI	Object
Secteur d'activité	Secteur d'activité	Object
PNB	PNB 2022 MDZD	Float64
Mouvement créditeurs	Mouvement créditeurs 2022	Float64
Importation	Import 2022	Float64
Exportation	Export 2022	Float64
Les encours	Encours	Float64

Tableau 4.1 – Caractéristiques des variables : Noms de colonnes et types de variables

- Définir la variable endogène et les variables exogènes : Comme nous l'avons vu précédemment dans le chapitre , le PNB est calculé en ajoutant les intérêts et les commissions générés par les mouvements créditeurs (X1), importations (X2), les exportations (X3) et les encours (X4) qui représentent les variables explicatives ou exogènes qui expliquent le PNB qui est dans ce cas la variable dépendante ou la variable cible (Y) : endogène.
- Créer une matrice X qui contient toutes les variables explicatives. Pour ce faire, nous avons utilisé la fonction concat de la bibliothèque pandas pour concaténer les colonnes correspondant à chaque variable. Ensuite, nous avons ajouté une colonne de constantes à la matrice X en utilisant la fonction add_constant de la bibliothèque statsmodels. Cette constante permet de prendre en compte l'intercept dans la régression.
- spécifier le modèle de régression en utilisant la fonction OLS de la bibliothèque statsmodels. Cette fonction prend en entrée deux arguments : la variable à expliquer (y) et la matrice X des variables explicatives. Elle renvoie un objet model qui contient toutes les informations sur le modèle de régression
- Afficher les résultats de la régression en utilisant la méthode summary de l'objet model.

Voici Les résultats :

OLS Regression Results						
Dep. Variable:	PNB 2022 MDZD	R-squared:	0.962			
Model:	OLS	Adj. R-squared:	0.960			
Method:	Least Squares	F-statistic:	593.8			
Date:	Tue, 09 May 2023	Prob (F-statistic):	2.85e-66			
Time:	21:56:38	Log-Likelihood:	-1764.6			
No. Observations:	100	AIC:	3539.			
Df Residuals:	95	BIC:	3552.			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.036e+06	1.3e+06	0.798	0.427	-1.54e+06	3.61e+06
Mouvements créditeurs 2022	2.876e-05	0.001	0.039	0.969	-0.001	0.001
Import 2022	-0.0071	0.002	-2.931	0.004	-0.012	-0.002
Export 2022	0.0035	0.026	0.135	0.893	-0.048	0.055
Encours	0.0603	0.002	26.710	0.000	0.056	0.065

Figure 4.5 –Résultat de la régression ols

Ces résultats montrent que le coefficient de détermination R^2 est de 0,962, ce qui indique que 96,2% de la variation de la variable dépendante (PNB) est expliquée par les variables indépendantes.

Le coefficient de la constante est de 1,036e+06, mais ce coefficient n'est pas significatif car le p-value associé 0,427 est supérieur à 0,05, le niveau de confiance communément utilisé. Cela signifie que la constante n'a pas un effet significatif sur la variable dépendante.

Le coefficient associé à la variable encours est de 0,0603, ce qui signifie que pour chaque augmentation d'une unité de la variable encours, le PNB augmente de 0,0603 unité. Ce coefficient est significatif car le p-value associé est inférieur à 0,05. Le coefficient associé à la variable import est de -0,0071, ce qui signifie que pour chaque augmentation d'une unité de la variable import, le PNB diminue de 0,0071 unité. Ce coefficient est également significatif. Le coefficient associé à la variable export est de 0,0035. Cependant, ce coefficient n'est pas significatif car le p-value associé est supérieur à 0,05. Pour la variable mouvements créditeurs 2022, le coefficient est de 2.876e-05 avec une p-value est de 0.969, ce qui indique que cette variable n'est pas significative pour expliquer le PNB en 2022.

En somme, les variables encours et importations ont un impact significatif sur le PNB, tandis que les variables exportations et mouvements créditeurs n'ont pas un impact significatif sur le PNB.

Nous avons ensuite calculé le coefficient standardisé (ou bêta) qui mesure l'impact relatif de chaque variable indépendante sur la variable dépendante dans un modèle de régression multiple. Nous avons divisé le coefficient de la variable indépendante par l'écart type de cette variable et en standardisant ainsi toutes les variables dans le modèle. Les résultats suivants permettent de comparer l'impact relatif de chaque variable indépendante, indépendamment de leur unité de mesure ou d'échelle.

Le coefficient standardisé le plus important est celui des encours .Les autres variables indépendantes (Import 2022, Export 2022 et Mouvements créditeurs 2022) ont des coefficients standardisés plus petits et très proches de zéro, ce qui indique que les encours ont Le plus d'impact sur le PNB par rapport aux autres variables.

```

const                1.814342e-02
Mouvements créditeurs 2022  5.033951e-13
Import 2022          -1.243950e-10
Export 2022          6.085525e-11
Encours              1.055292e-09
dtype: float64
1.0552915351609512e-09
Encours
La variable avec l'impact le plus important est "Encours"

```

Figure 4.6 – Capture des résultats : Impact relatif des variables indépendantes sur le PNB (coefficients standardisés)

En se basant sur ces résultats et afin de prouver ce qu'on a constaté en calculant ces coefficients nous avons utilisé une approche de sélection de variables pour éliminer les variables export 2022 et mouvement créditeurs 2022 qui n'ont pas eu d'effet significatif sur le PNB avec le modèle de régression multiple. Nous avons ensuite comparé les performances des deux modèles restants, le modèle Encours et le modèle Import 2022 en termes de R2, F-statistique et RSS.

Les résultats obtenus pour chaque modèle sont les suivants :

La variable X	R2	F test	P value	RSS	ESS	TSS
Encours	95.47%	2068.299	1.083e-67	1.461e+16	3.084e+17	3.23e+17
Import 2022	67.04%	199.349	2.343e-25	1.064e+17	2.165e+17	3.23e+17

Tableau 4.2 – Résultats de l'analyse statistique pour les encours et l'import 2022

Les résultats montrent que le modèle avec les encours est significativement meilleur que le modèle avec les importations en termes de R2 et de F-statistique. Le R2 est beaucoup plus élevé pour le modèle avec les encours (0,95) que pour le modèle avec les importations (0,67). La F-statistique est également beaucoup plus élevée pour le modèle avec les encours (2068) que pour le modèle avec les importations (199). De plus, la p-valeur pour la F-statistique est très faible (1.083e-67), ce qui suggère une forte confiance dans le fait que le modèle avec les encours est significativement meilleur que le modèle avec les importations. De plus, les résultats de RSS confirment que le modèle avec les encours à un meilleur ajustement que le modèle avec les importations. En effet, le RSS est beaucoup plus faible pour le modèle avec les encours (1,46e+16) que pour le modèle avec les importations (1,06e+17), ce qui indique que les données sont mieux ajustées par le modèle avec les encours. Cela confirme les résultats de coefficient standardisé fournis auparavant.

4.3.2 La causalité

Avant d'entamer la seconde étape de notre solution qui est de prédire les encours afin d'avoir une vision globale sur l'évolution du PNB nous avons appliqué la méthode de Granger pour tester si l'inclusion des valeurs passées des encours améliore significativement la prédiction du

PNB. Pour cela, nous avons utilisé la méthode de Granger Causality pour vérifier s'il y avait une causalité entre les variables PNB et Encours. Nous avons divisé notre jeu de données en un ensemble d'entraînement et un ensemble de tests, puis nous avons appliqué la fonction `grangercausalitytests` de `statsmodels` sur notre ensemble d'entraînement. Nous avons fixé le paramètre `maxlag` à 2, ce qui signifie que nous avons testé la causalité jusqu'à deux retards temporels.

La sortie de la fonction `grangercausalitytests` nous a fourni les résultats du test de causalité de Granger pour chaque retard temporel. Les résultats comprenaient la statistique de test F, la valeur p et les degrés de liberté, ainsi que les statistiques de test basées sur le chi carré et le ratio de vraisemblance. Les valeurs p pour chaque retard temporel étaient toutes inférieures à 0,05, ce qui indique une forte preuve de causalité entre les deux variables.

Le test de causalité de Granger est utile pour déterminer si une série temporelle (dans ce cas, les encours) est utile pour prédire une autre série temporelle (dans ce cas, le PNB) en utilisant une approche statistique. Si le test de causalité de Granger révèle que les encours sont significativement utiles pour prédire le PNB, cela justifie l'argument selon lequel il existe une relation causale entre les deux variables et que les encours ont un impact significatif sur le PNB. La statistique F est utilisée pour tester si l'inclusion des valeurs retardées de la variable indépendante aide à prédire la variable dépendante pour différentes valeurs de lag. Les p-values indiquent si la relation est statistiquement significative ou non.

Pour le résultat de `lag = 1`, la statistique F est de 289.2344 et la p-value est de 0.0000, ce qui indique une forte preuve de causalité de Granger. De même, pour le résultat de `lag = 2`, la statistique F est de 143.5422 et la p-value est de 0.0000, indiquant également une forte preuve de causalité de Granger.

Cela suggère que l'inclusion des valeurs passées des encours améliore significativement les prévisions du PNB, cela pourrait indiquer une relation de causalité entre les deux variables, suggérant que les encours ont un impact direct sur le PNB et que les encours sont un prédicteur causalement significatif du PNB, et nous pouvons donc utiliser les encours pour prédire le PNB.

Après avoir prouvé le lien fort entre les encours et le PNB, il était temps de prédire les encours, cependant avant de pouvoir effectuer des prévisions, il était nécessaire de construire une dataset historique fiable. Pour cela, nous avons utilisé TeraData pour effectuer cette opération comme on a fait précédemment. Grâce à cet outil, nous avons pu collecter et stocker les données historiques nécessaires à la prédiction des encours dans un fichier format Excel pour pouvoir l'importer lors de la prédiction.

4.4 La collecte et la consolidation des données

La direction de la finance dispose la dataset complète des encours sous le nom de `table_engagement` contenant des valeurs d'encours pour chaque client, en tenant compte des différents types de clients. Cependant, notre objectif était de sélectionner uniquement les lignes correspondant aux clients de type `corporate`, pour la période allant de janvier 2012 à mars 2023. Pour cela nous avons donc effectué une requête spécifique pour extraire ces lignes. En utilisant les critères de sélection appropriés, nous avons filtré la dataset afin d'obtenir uniquement les encours des clients `corporate` pour la période spécifiée comprise entre janvier 2012 et mars 2023.

```

SELECT *
FROM table_engagement
WHERE type_client = 'corporate'
      AND date_Engagement >= '2012-01-01'
      AND date_Engagement <= '2023-03-31';

```

Figure 4.7–Requête SQL pour l'extraction des encours des clients corporate pour la période janvier 2012 - mars 2023

Ensuite, nous avons procédé à l'agrégation des données mensuelles. Étant donné que les données sont disponibles à un niveau mensuel, nous avons calculé la somme des lignes pour chaque mois afin d'obtenir un engagement total mensuel. Nous avons répété cette opération chaque mois, de janvier 2012 à mars 2023.

```

SELECT DATE_FORMAT(date_engagement, '%Y-%m') AS mois, SUM(engagement) AS total_engagement
FROM table_engagement
WHERE type_client = 'corporate'
      AND date_encours >= '2012-01-01'
      AND date_encours <= '2023-03-31'
GROUP BY mois;

```

Figure 4.8–Requête SQL pour l'extraction et l'agrégation mensuelle des encours clients corporate

Cette requête SELECT utilise la fonction DATE_FORMAT pour extraire le mois au format YYYY-MM à partir de la colonne date_engagement. Ensuite, elle calcule la somme des encours pour chaque mois à l'aide de la fonction d'agrégation SUM. La clause GROUP BY regroupe les résultats par mois. Le résultat de cette requête sera une liste des mois au format YYYY-MM avec les totaux des encours correspondants pour les clients corporate. Nous avons obtenu une table qui contient deux colonnes : mois (au format YYYY-MM) et total_engagement. Les données sont obtenues en utilisant la même requête SELECT que précédemment.

```

SELECT DATE_FORMAT(date_engagement, '%Y-%m') AS mois, SUM(encours) AS total_engagement
FROM table_engagement
WHERE type_client = 'corporate'
      AND date_encours >= '2012-01-01'
      AND date_encours <= '2023-03-31'

GROUP BY mois;

```

Figure 4.9 – Requête SQL pour la création d'une dataset consolidée des encours mensuels clients corporate (janvier 2012 - mars 2023)

Ainsi, à la fin de ce processus, nous avons obtenu une nouvelle dataset couvrant la période de janvier 2012 à mars 2023, contenant les totaux mensuels des encours pour les clients corporate. Cette dataset consolidée nous permettra de procéder à des analyses plus précises et de tirer des conclusions significatives pour notre étude.

4.5 Exploration et compréhension de données

Notre dataset contient deux colonnes : une colonne pour la date et une colonne pour les valeurs des encours. Avant de visualiser notre série temporelle, nous avons commencé par convertir la colonne date en un format de date approprié à l'aide de la fonction `pd.to_datetime()` de la bibliothèque pandas . Cela permettra de manipuler et d'analyser les données temporelles, ensuite nous avons défini cette colonne comme index de notre data frame en utilisant la méthode de `set_index`. Cela signifie que les dates seront utilisées comme référence pour l'organisation des données.

Après avoir préparé nos données convenablement, nous avons utilisé plusieurs méthode de visualisation pour bien comprendre notre série temporelles. Voici les figures qui montrent l'évolution des encours au fil du temps obtenus :

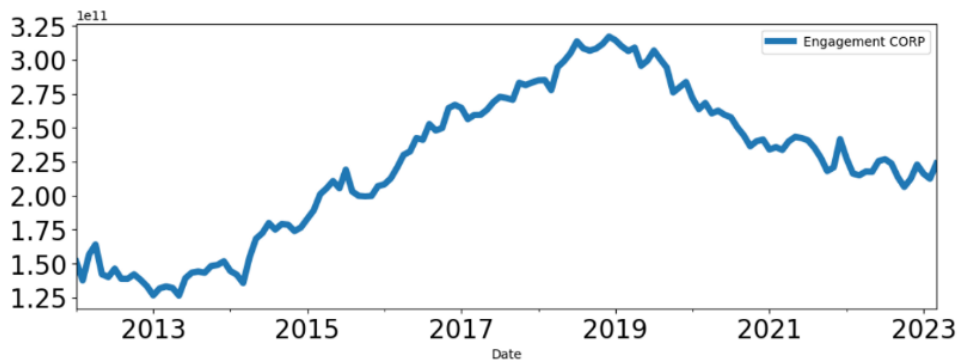


Figure 4.10 – Évolution de l'engagement à partir de 2012 jusqu'à 2023

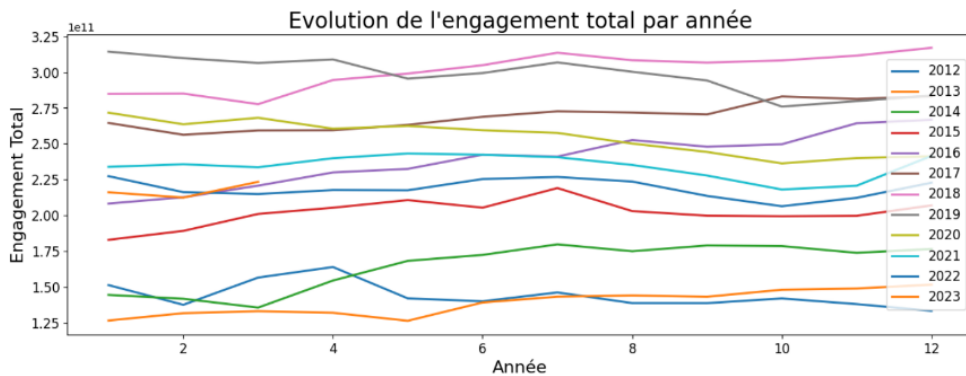


Figure 4.11 – Évolution de l'engagement total par année

D'après l'observation de ces deux graphes, la série temporelle présente une tendance générale à la hausse de 2012 jusqu'à une période donnée. Ensuite, à partir de 2019, on observe une chute significative dans les valeurs de la série temporelle. Cette baisse se poursuit tout au long des années 2020, 2021 et 2022. Cependant, à partir de 2023, la série temporelle semble reprendre progressivement son allure d'augmentation.

Il est important de noter que cette description se base sur une analyse visuelle de la série temporelle et ne fournit pas de valeurs numériques précises pour quantifier les variations

observées. Une analyse plus approfondie, en utilisant power bi qui offre une large gamme de visualisations interactives et qui pourrait fournir des mesures quantitatives et des tendances plus précises sur l'évolution de la série temporelle.

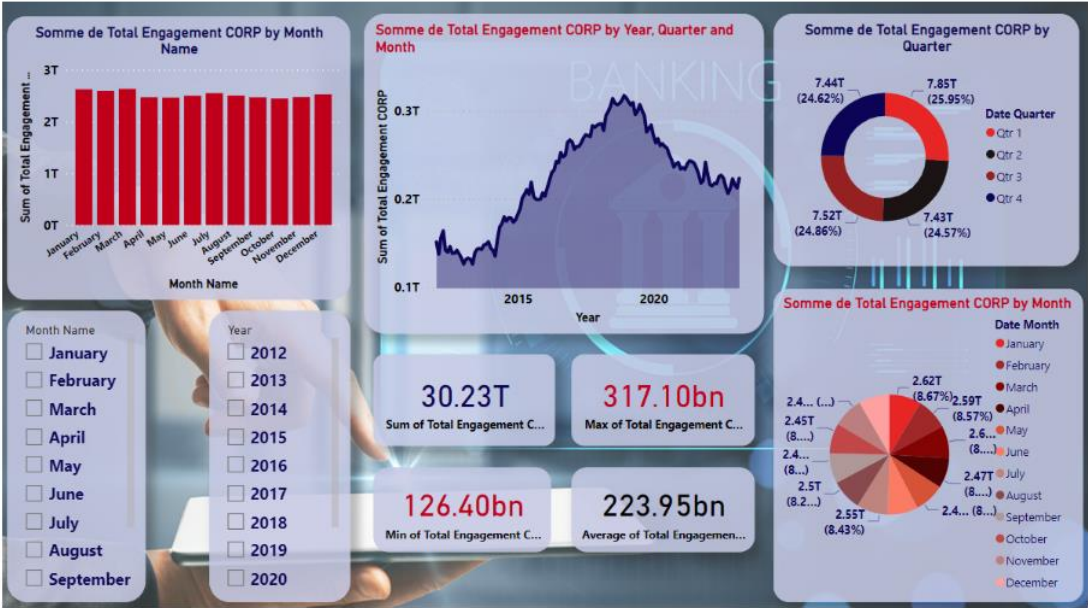


Figure 4.12–Tableau de bord interactif Power BI : Évolution de la série temporelle par année

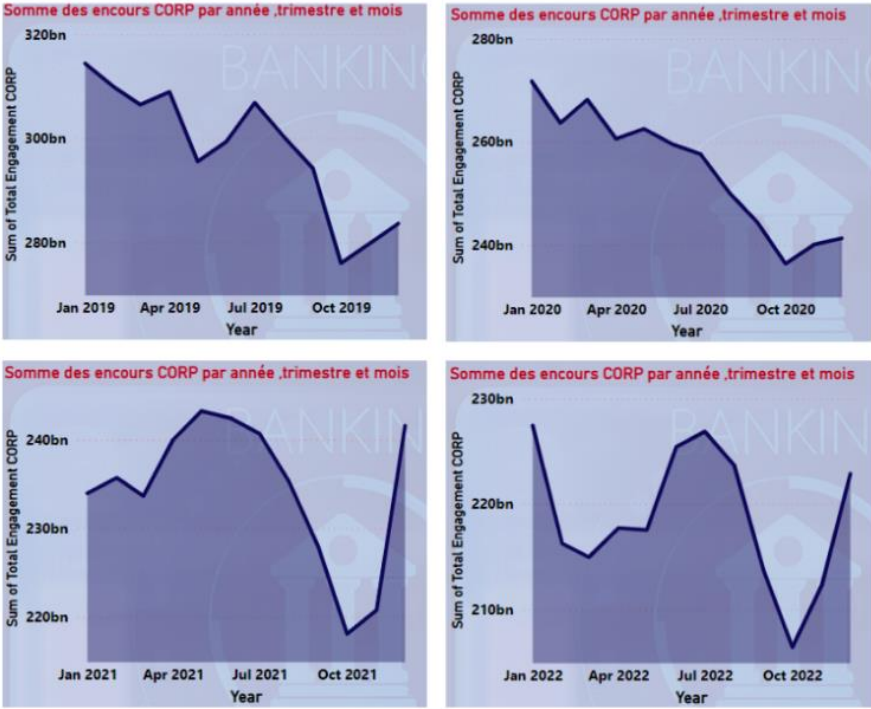


Figure 4.13–Page du tableau de bord : Récapitulatif des années de baisse des encours.

Selon les données fournies, la somme des encours pour les années 2019, 2020, 2021 et 2022 a montré une tendance à la baisse, passant de 3,58 T à 3,06 T, puis à 2,81 T et enfin à 2,63 T respectivement. Cette baisse est attribuée aux conditions politiques de l'année 2019 en Algérie,

ainsi qu'à la période de la pandémie de COVID-19 qui a eu un impact mondial sur l'économie, y compris le secteur financier.

Les conditions politiques survenues en Algérie en février 2019 ont eu un impact significatif sur le secteur financier. Certaines institutions financières ont ressenti les conséquences économiques indirectes de ces conditions politiques qui peuvent avoir des répercussions sur l'économie dans son ensemble, y compris sur les activités bancaires et les niveaux d'encours.

De plus, la pandémie de COVID-19 durant ces trois dernières années a eu un impact majeur sur SGA. Les mesures de confinement, les restrictions de déplacement et les fermetures d'entreprises ont entraîné une baisse de l'activité bancaire. La diminution des importations, en particulier dans le secteur automobile, a contribué à la baisse des encours. Les banques ont été confrontées à des défis tels que la diminution des demandes de prêt, les problèmes de remboursement des emprunts existants et l'augmentation des risques liés à la solvabilité des emprunteurs. En outre, une baisse de la demande de prêts dans le secteur automobile a incité la banque à ajuster les taux d'intérêt à la baisse pour encourager l'emprunt et stimuler l'activité économique dans le but de soutenir l'économie pendant la période de crise.

4.5.1 Décomposition de la série temporelle

La décomposition de la série temporelle en tendance, saisonnalité et résidus est une étape importante dans l'analyse des données temporelles. L'utilisation d'un modèle multiplicatif est justifiée dans ce cas car la série temporelle présente une variation relative plutôt qu'une variation absolue. Étant donné que la série temporelle augmente puis diminue, un modèle multiplicatif est plus adapté pour capturer les variations proportionnelles plutôt qu'un modèle additif qui serait plus approprié pour les variations absolues.

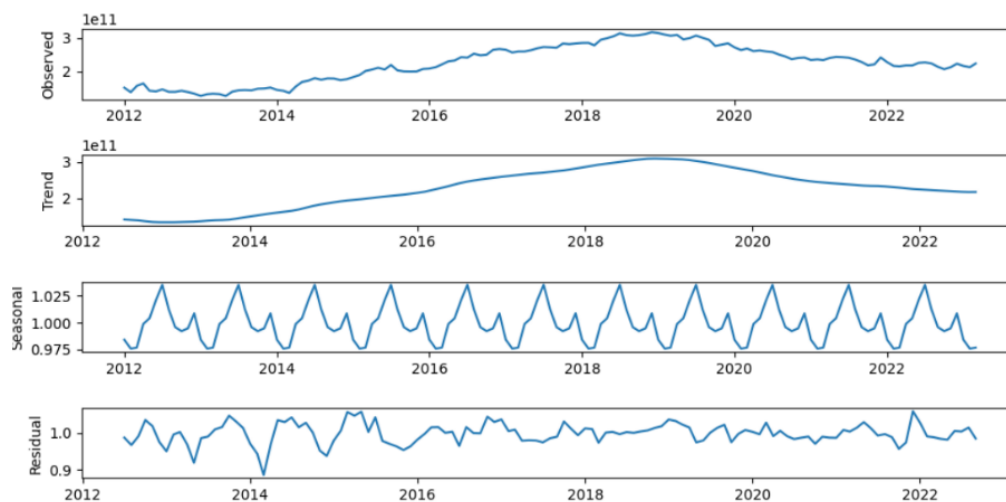


Figure 4.14– la décomposition de la série temporelle

Cette décomposition met en évidence la tendance qu'on a déjà constaté et justifiée auparavant, elle permet également d'identifier la composante saisonnière de la série temporelle, dans notre cas on constate une saisonnalité mensuelle avec un pic majoritairement en juillet et une baisse en octobre. Le pic observé en juillet est attribué aux dépenses accrues des clients corporate durant la période estivale, certains secteurs sont plus actifs et leurs activités nécessitent des

dépenses accrues comme le secteur de transport et services, de plus certaines entreprises ont des projets ou des investissements saisonniers spécifiques à cette période, ce qui se traduit par des dépenses plus élevées.

La baisse observée en octobre dans les encours est expliquée par la fin de la saison estivale, la période de transition vers l'automne ou des fluctuations saisonnières propres à certains secteurs, certaines entreprises peuvent ajuster leurs dépenses en fonction des obligations fiscales et comptables, ce qui peut entraîner une baisse des encours en octobre.

La décomposition nous a permis aussi d'extraire la composante résiduelle de la série temporelle, qui représente les fluctuations aléatoires ou non explicables par la tendance et la saisonnalité.

La vérification si ces résidus présentent les caractéristiques d'un bruit blanc est importante lors de l'analyse de données chronologiques. Un bruit blanc est un processus stochastique caractérisé par des variables aléatoires indépendantes et identiquement distribuées, sans corrélation temporelle. En vérifiant si la série n'est pas un bruit blanc, nous nous assurons qu'elle présente des structures temporelles significatives telles que des tendances, des saisons ou d'autres motifs. Cela permet une modélisation précise et des prédictions pertinentes en prenant en compte ces variations non aléatoires dans les données. Pour évaluer les résidus et vérifier s'ils présentent les caractéristiques d'un bruit blanc, nous avons utilisé deux approches complémentaires, la première approche se trouve dans l'annexe et pour la deuxième approche on a utilisé le test de Ljung-Box, qui est une méthode statistique pour évaluer l'autocorrélation des résidus. Ce test vérifie si les résidus présentent une corrélation temporelle significative, ce qui indiquerait qu'ils ne se comportent pas comme un bruit blanc.

Lorsque nous avons effectué le test de Ljung-Box nous avons obtenu les résultats suivants :

- Statistique de Ljung-Box (lb_stat) : 133.710383
- Valeur p associée (lb_pvalue) : 6.321606e-31

Étant donné que la valeur p associée est extrêmement faible, bien en dessous du seuil communément utilisé de 0,05, nous rejetons l'hypothèse nulle du test de Ljung-Box. Cette conclusion renforce l'idée que les résidus de la série temporelle présentent une corrélation temporelle significative (Nous avons également réalisé une méthode visuelle que vous trouverez en annexe G).

Cela confirme nos observations initiales selon lesquelles les résidus contiennent des structures temporelles non aléatoires. Par conséquent, il est justifié de prendre en compte ces variations non aléatoires lors de la modélisation et des prédictions ultérieures de la série temporelle.

4.5.2 Évaluation de la stationnarité

Avant de commencer toute analyse sur nos données de type série temporelle nous devons vérifier la stationnarité. Nous avons appliqué deux transformations couramment utilisées : la fonction logarithmique et la différenciation. Après avoir effectué ces transformations, nous avons vérifié la présence de stationnarité en utilisant trois méthodes. Les deux premières méthodes se trouvent dans l'annexe H, et voici les résultats du test statistique de Dickey-Fuller augmenté (ADF) pour chaque transformation :

	Les données initiales	Transformation par log	Transformation par différenciation
ADF Statistic	-1.343526	-1.388867	-11.498030
p-value	0.609032	0.587602	4.5633467872743875x 10 ⁻²¹
Stationnarité	Non-stationnaire	Non-stationnaire	Stationnaire

Tableau 4.0.3 – Résultats du test de stationnarité (ADF) pour chaque transformation

Les données initiales ont une p-value supérieure à 0,05, ce qui indique que la série chronologique est non-stationnaire. Après avoir appliqué la transformation logarithmique, nous avons observé une légère amélioration, mais la p-value est toujours supérieure au seuil de 0,05, ce qui indique que la série chronologique est toujours non-stationnaire. Cependant, lorsque nous avons appliqué la différenciation, nous avons obtenu une p-value inférieure à 0,05, ce qui suggère que la série chronologique est devenue stationnaire. Les résultats de ces tests suggèrent que la série chronologique initiale n'était pas stationnaire, mais la différenciation a permis de la rendre stationnaire.

4.6 Application des modèles prédictifs

Les bibliothèques essentielles, telles que Pandas et NumPy pour la manipulation des données, Matplotlib et Seaborn pour la visualisation, statsmodels pour les statistiques, ainsi que Scikit-learn, ont été téléchargées pour chaque modèle. Nous avons également utilisé la fonction `train_test_split` de Scikit-learn pour diviser les données en un ensemble d'entraînement de 90% et un ensemble de test de 10%.

4.6.1 ARIMA

Le modèle ARIMA est caractérisé par 3 paramètres clés : p, d et q. Comme expliqué précédemment dans le chapitre 3 .La sélection des valeurs appropriées pour ces paramètres est cruciale pour obtenir une prédiction précise de la série chronologique. Dans la section suivante, nous allons discuter de la méthode utilisée pour sélectionner les valeurs optimales de p, d et q pour notre modèle ARIMA :

1- Pour déterminer le nombre optimal de termes autoregressifs (p), nous avons utilisé la fonction PACF qui mesure la corrélation partielle entre la série chronologique et ses valeurs passées après avoir tenu compte des relations linéaires de ces valeurs avec d'autres valeurs passées.

Nous avons obtenu le graphe PACF suivant :

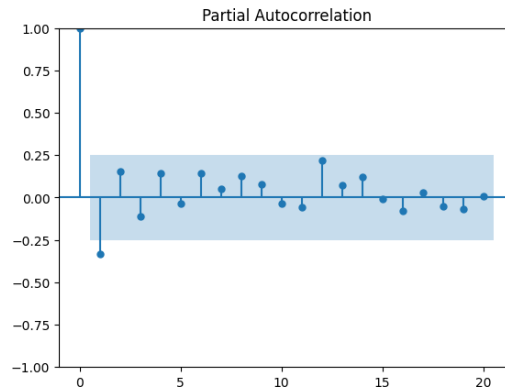


Figure 4.15–Graphique de PACF

Nous pouvons observer que le décalage PACF 1(lag=1) est assez significatif car il se situe nettement au-dessus de la ligne de signification et devient négligeable pour les lags suivants. Cela suggère que nous avons besoin d'un seul terme autoregressif ($p=1$) pour capturer la structure temporelle de notre série chronologique.

2- Le paramètre d correspond au nombre de fois que la série chronologique doit être différenciée pour devenir stationnaire. Après avoir examiné la série chronologique, nous avons constaté que la série n'était pas stationnaire et qu'elle présentait une tendance. Nous avons donc effectué une différenciation simple ($d=1$).

3- Pour déterminer le nombre optimal de termes de $MA(q)$, nous avons utilisé la fonction ACF qui nous a permis d'obtenir le graphe ACF ci-dessous :

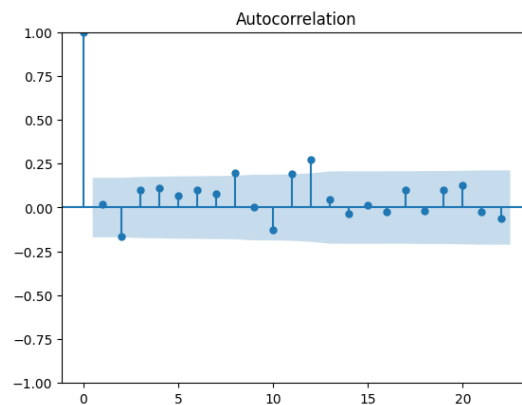


Figure 4.16 – Graphique de la fonction d'autocorrélation (ACF)

Nous pouvons voir que la valeur de la corrélation est significative pour le premier lag (lag=1) et devient négligeable pour les lags suivants. Cela suggère que nous avons besoin d'un seul terme de moyenne mobile ($q=1$) pour capturer la structure temporelle de notre série chronologique.

Après avoir choisi les paramètres optimaux pour le modèle, nous avons divisé notre ensemble de données, après avoir appliqué la différenciation et nous avons créé le modèle avec la méthode `ARIMA()` et l'ajuster avec `fit()`. Ensuite nous avons utilisé le modèle ARIMA que nous venons de créer pour faire des prédictions sur les données de test que nous avons exclues de notre ensemble de données initial avec la méthode `predict`, et nous avons créé un graphe comparatif entre les valeurs réelles des données de test et les prédictions fournies par le modèle:

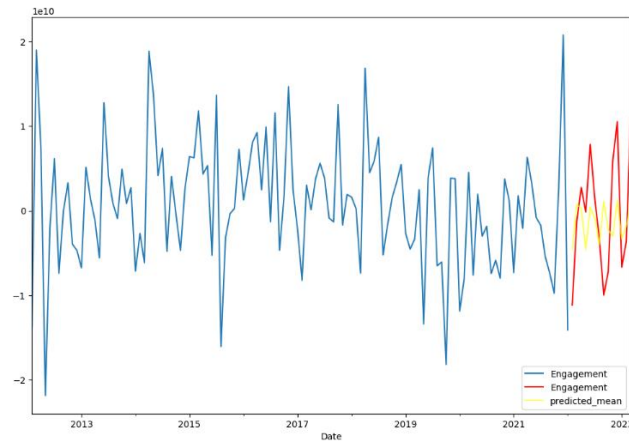


Figure 4.17–Comparaison des données différencié, de l'ensemble d'entraînement et de l'ensemble de test

Par la suite nous avons réalisé des prédictions jusqu'en 2025 en utilisant le modèle que nous avons développé, nous avons appliqué la fonction `cumsum()` à la série temporelle différenciée. Cette étape nous a permis de calculer la somme cumulative des prédictions sur la période considérée. Cependant, il est important de noter que la série utilisée pour les prédictions était différenciée. Par conséquent, pour représenter les résultats finaux dans leur forme originale, nous devons annuler l'effet de la différenciation. Pour ce faire, nous avons appliqué la fonction inverse de la différenciation, qui consiste à effectuer une opération de `cumsum` inversée. Cela nous permet de retrouver les valeurs de la série temporelle dans leur état initial, avant la différenciation.

En utilisant cette technique, nous avons obtenu les valeurs prédites dans la série temporelle originale. Nous avons ensuite tracé le graphe suivant, qui représente l'évolution de la série temporelle prédite jusqu'en 2025, en prenant en compte l'annulation de l'effet de la différenciation :

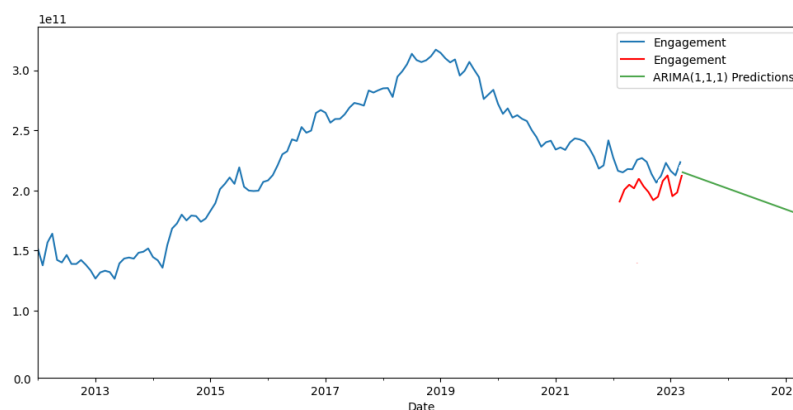


Figure 4.18–Visualisation des données initiales, de l'ensemble de test et des prédictions futures par le modèle ARIMA

Le tracé de la prédiction du test en rouge montre des augmentations et des diminutions similaires à celles des données réelles, ce qui indique que notre modèle est capable de reproduire les motifs et les tendances présents dans la série. En ce qui concerne la prédiction future

représentée en vert, nous constatons une tendance à la baisse de la série temporelle. Cela suggère que, selon notre modèle, la série devrait continuer à diminuer au fil du temps.

4.6.2 SARIMA

Les paramètres du modèle SARIMA sont similaires à ceux du modèle ARIMA. Cependant, il inclut également des paramètres pour modéliser les tendances saisonnières, donc les paramètres seront (P,D,Q) et aussi un autre paramètre : la période saisonnière (s).

Pour la période saisonnière est 12 pour les données avec une saison d'un an comme vous pouvez voir dans le graphe suivant :

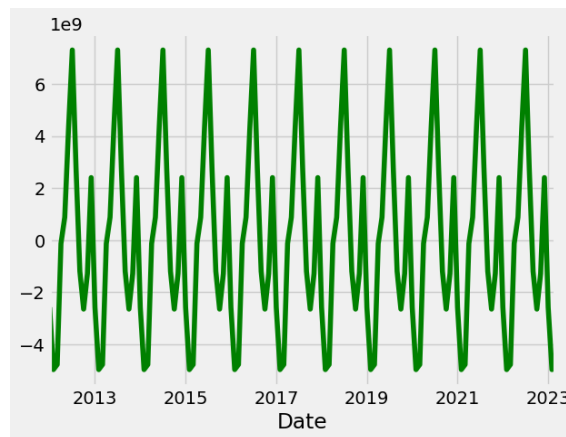


Figure 4.19 – Motifs de saisonnalité dans les données

Pour trouver les paramètres optimaux nous avons utilisé une approche de grille de recherche : Les paramètres p , d et q ont été définis pour prendre des valeurs entre 0 et 2, en se basant sur le modèle ARIMA de base qui avait les paramètres (1,1,1). Nous avons également pris en compte les paramètres saisonniers en générant toutes les combinaisons possibles de triplets p , d , q avec une période saisonnière de 12 pour indiquer une saisonnalité annuelle. En ajustant chaque modèle avec ces paramètres, nous avons utilisé le critère d'information d'Akaike (AIC) pour sélectionner le meilleur modèle. Finalement, nous avons trouvé que le meilleur modèle SARIMA pour notre série chronologique avait des paramètres (1, 1, 1) pour ARIMA et (0, 1, 1, 12) pour les paramètres saisonniers.

Après avoir ajusté le modèle en prenant en compte toute non-stationnarité dans la série chronologique afin de garantir une modélisation précise. Nous avons fait des prédictions sur une période donnée et comparé visuellement ces prédictions avec les données observées ainsi que la marge d'incertitude associée aux prévisions :

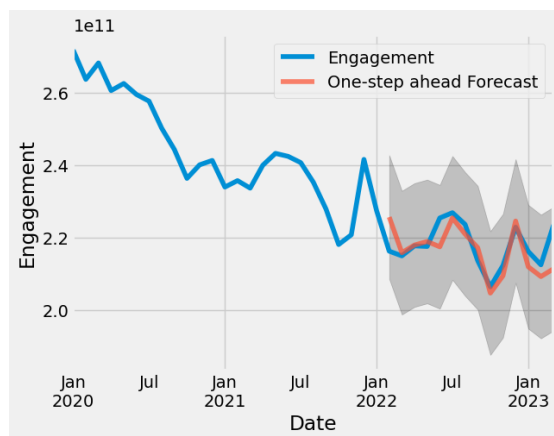


Figure 4.20 – Comparaison des données initiale, de l'ensemble d'entraînement et de l'ensemble de test à partir du janvier 2020

Nous pouvons voir que les prévisions suivent la tendance générale des données observées, avec des hausses et des baisses similaires. Cependant, les valeurs exactes des prévisions peuvent différer légèrement des valeurs observées. Enfin, nous avons effectué des prévisions pour les 10 prochaines années et tracé un graphique illustrant ces prévisions.

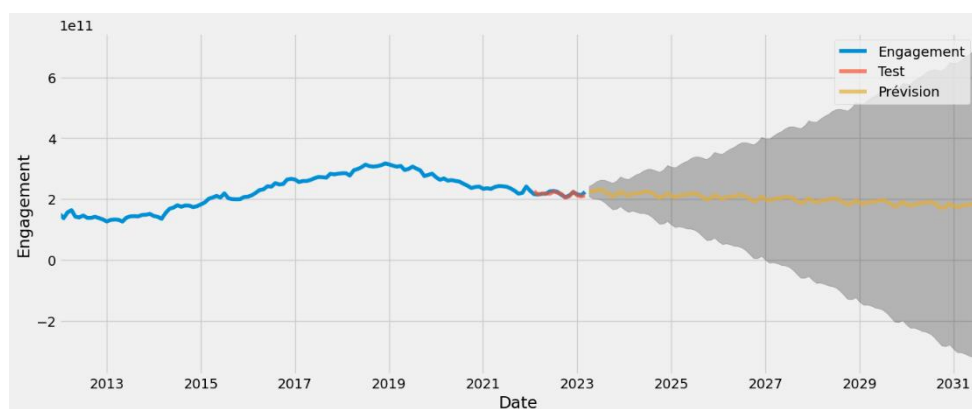


Figure 4.21 – Visualisation des données initiales, de l'ensemble de test et des prédictions futures avec le modèle SARIMA

Dans le graphe, nous pouvons observer que la tendance de la série chronologique continue de diminuer au fil du temps. Cependant, en plus de cette tendance, nous pouvons également voir des variations saisonnières à intervalles réguliers ce qui indique une saisonnalité.

4.6.3 XGBOOST

Nous avons utilisé la fonction "create_features" pour extraire des informations trimestrielles, mensuelles et annuelles à partir de la série chronologique différenciée. Ces informations ont été ajoutées comme nouvelles colonnes dans le dataframe. Ensuite, nous avons divisé la série en deux parties comme nous l'avons mentionné auparavant. La fonction "create_features" a été appliquée à la fois sur l'ensemble d'entraînement et l'ensemble de test pour extraire ces

caractéristiques supplémentaires. Ces caractéristiques ont été utilisées comme entrées pour entraîner le modèle XGBoost, dans le but d'améliorer la précision des prédictions.

Une fois le modèle entraîné avec la fonction `XGBRegressor()` et ajusté avec les paramètres appropriés pour éviter le surapprentissage, nous avons utilisé ce dernier pour effectuer des prédictions sur l'ensemble de test en utilisant les caractéristiques correspondantes. Les prédictions ont ensuite été comparées aux valeurs réelles de l'ensemble de test avec le graphique suivant :

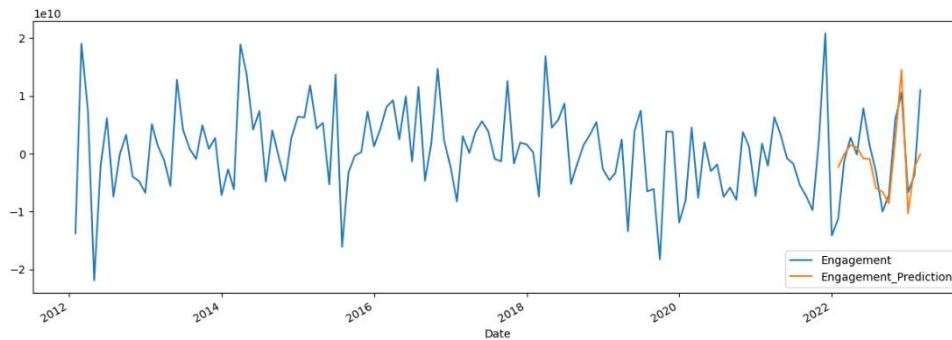


Figure 4.22 – Comparaison des données initiales et de l'ensemble de test

En examinant le graphique obtenu, nous pouvons constater que les variations dans les données d'entraînement ont été également répliquées dans les données de test, bien que les valeurs diffèrent légèrement. Enfin, nous avons tracé deux courbes : la courbe de la série temporelle d'origine et la courbe de la série temporelle prédite après avoir utilisé la fonction `cumsum()` pour obtenir la série temporelle prédite originale.

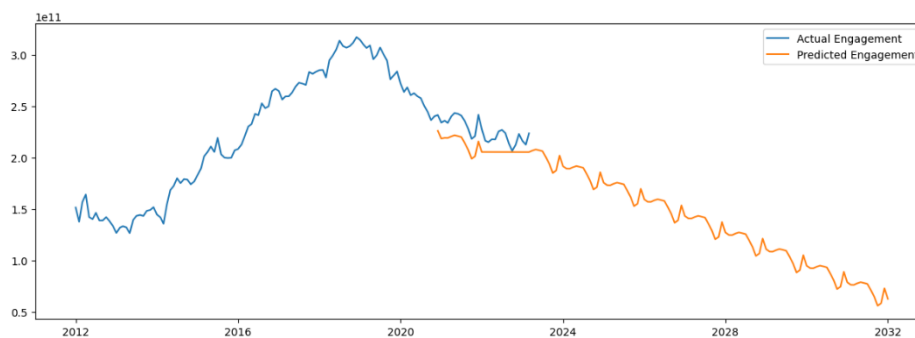


Figure 4.23 – Visualisation des données initiales, de l'ensemble de test et des prédictions futures avec les modèles XGboost

Pour cette courbe, nous pouvons observer que les valeurs d'engagement pour les données de test et d'entraînement suivent la même tendance en termes d'augmentation et de diminution, bien que leurs valeurs soient différentes, mais pas légèrement. De plus, la prédiction de l'engagement jusqu'à 2032 montre une tendance à la baisse. De plus, des fluctuations saisonnières annuelles sont également observables dans les données.

4.6.4 Fb prophet

Nous avons commencé par importer les bibliothèques prophet de fbprophet, et nous avons divisé les données en ensemble d'entraînement ensemble de test. L'étape suivante consiste à réinitialiser l'index et renommer les colonnes de l'ensemble d'entraînement. La méthode

reset_index() réinitialise l'index du dataframe pour que les indices commencent à partir de 0, et columns est utilisé pour renommer les colonnes en ds et y, qui sont les noms de colonnes standard utilisés par prophet. Ensuite nous avons créé un objet prophet et ajusté le modèle aux données d'entraînement. En dernier lieu, nous avons prédit les encours pour les 10 prochaines années en utilisant le modèle ajusté aux données d'entraînement. Cela nous a permis d'avoir un tableau de prédiction, afin d'organiser ce dernier nous avons renommé les colonnes du dataframe prédiction pour une meilleure lisibilité en utilisant la méthode columns. Ensuite, nous avons trié les valeurs par date en utilisant la méthode sort_values.

Afin d'analyser notre nouvelle dataset prédites nous avons fait appel à plusieurs visualisation telles que plot, m.plot_components qui permet de visualiser les différents composants du modèle de manière séparée et de comprendre comment chaque composant contribue à la prédiction globale, plot_plotly(m, FRC) qui trace un graphique interactif qui permet de visualiser les prédictions et les données historiques sur un même graphique avec des intervalles de confiance qui indiquent la plage probable de valeurs futures de plus l'interaction permet une exploration plus flexible des prévisions pour différentes dates, ce qui peut aider à identifier des tendances.

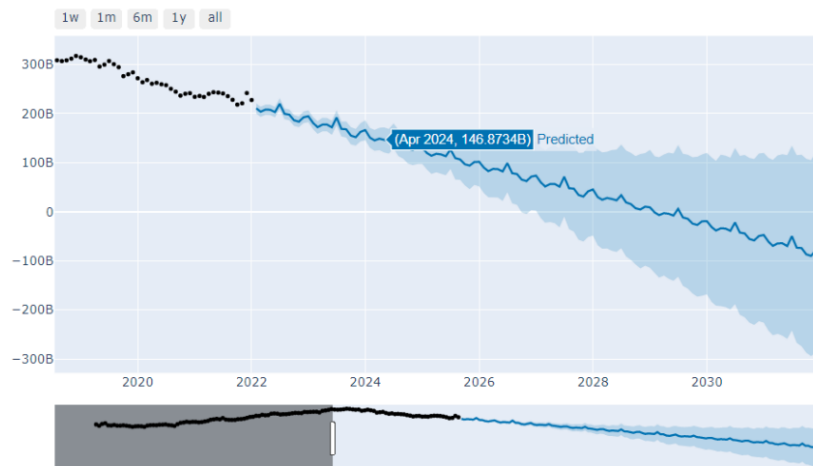


Figure 4.24–Graphique interactif qui permet de visualiser les prédictions et les données historiques

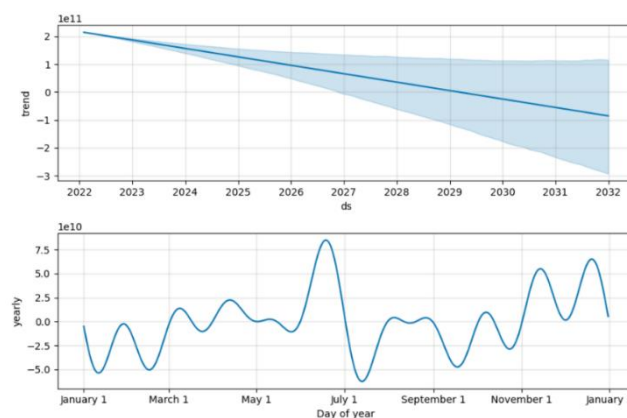


Figure 4.25 – la décomposition des données en tendance

La première figure montre l'historique des données, avec la ligne noire représentant les valeurs réelles et la zone bleue représentant les valeurs prédites avec l'intervalle de confiance. Nous pouvons voir que les encours vont diminuer au fil du temps ce que suggère ainsi la première figure. La deuxième figure montre la décomposition des données en tendance. Nous pouvons voir que la tendance est effectivement à la baisse linéairement, comme les autres figures l'indiquent.

4.7 Évaluation des erreurs

Après avoir appliqué chacun des quatre modèles (ARIMA, SARIMA, XGBoost et FB Prophet) aux données de séries temporelles mensuelles, nous avons évalué la qualité de la prédiction à l'aide de ces métriques : le NRMSE, et le WMAPE et le coefficient de détermination R^2 .

Voici les résultats de chaque modèle :

	ARIMA	SARIMA	XGBoost	Fbprophet
NRMSE	28.829	10.961	24.556	7.750
WMAPE	40.997	9.558	22.914	1.738
R^2	0.16468	0.5106	0.488	0.745

Tableau 4.4 – Évaluation de la qualité de prédiction de chaque modèle

Le modèle SARIMA a montré de meilleures performances par rapport au modèle ARIMA. Cela s'explique par le fait que le modèle SARIMA inclut une composante saisonnière, ce qui lui permet de mieux capturer les variations périodiques présentes dans les données. En intégrant cette composante saisonnière, le modèle SARIMA peut offrir des prédictions plus précises et mieux ajustées aux données réelles. En ce qui concerne le modèle XGBoost, ses performances ont été relativement faibles par rapport aux modèles SARIMA et ARIMA. Enfin, le modèle Fbprophet a présenté les meilleures performances parmi les modèles évalués, avec une erreur moyenne plus faible et une meilleure précision dans les prédictions. Il a également montré une meilleure capacité à expliquer la variance des données par rapport aux autres modèles.

4.8 Étude complémentaire

Il convient de noter que les encours bancaires peuvent être influencés par divers facteurs économiques et sociaux tels que les fluctuations des marchés, les politiques gouvernementales et les événements majeurs tels que la pandémie de COVID-19 et des perturbations économiques. Ces facteurs externes peuvent avoir un impact significatif sur les tendances observées dans les encours bancaires. Par exemple, l'exportation et l'importation ont été temporairement arrêtées. Cette situation a logiquement conduit à une diminution des encours bancaires. Toutefois, avec la relance progressive de ces activités, il est attendu que les encours

bancaires augmentent à nouveau contrairement à ce que nous avons obtenu dans les prédictions précédentes.

Afin de mieux comprendre les tendances sous-jacentes et d'évaluer l'impact réel de ces facteurs externes, nous allons effectuer une nouvelle prédiction en utilisant des données se terminant en décembre 2018. Cette approche nous permettra de démontrer si, en l'absence des effets temporaires du COVID-19 et des événements politiques, les encours bancaires ont tendance à augmenter de manière continue. En appliquant la même méthodologie précédemment pour prédire l'ensemble de la série temporelle, et en évaluant l'erreur nous avons trouvé que le modèle fb prophet est le meilleur modèle (vous trouverez tous les résultats en annexe)

Nous avons obtenu le graphe suivant :

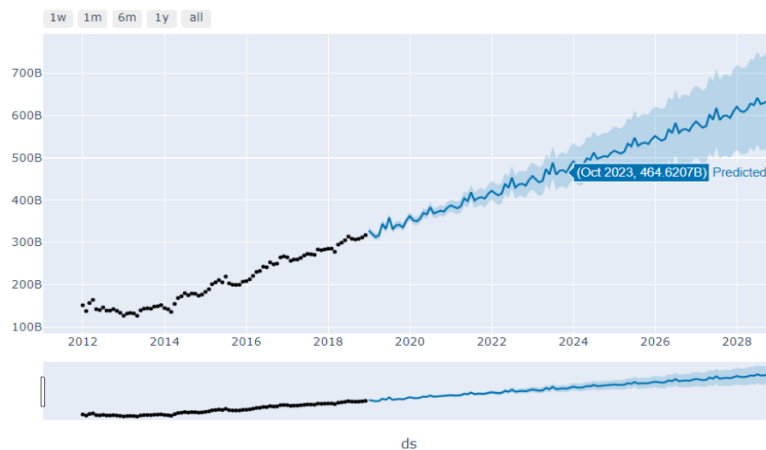


Figure 4.26 – Visualisation des données initiales et des prédictions futures avec le modèle Fb prophet

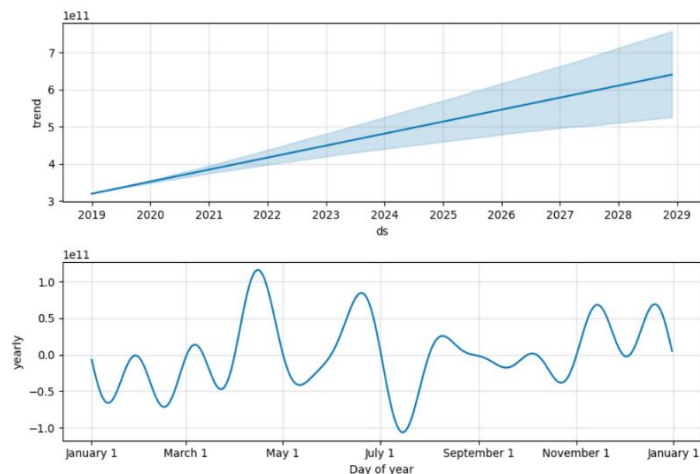


Figure 4.27 – La décomposition des données en tendance

La nouvelle prédiction montre une tendance à la hausse des encours au fil du temps avec une tendance linéaire en augmentation. Cela suggère que la banque peut s'attendre à une amélioration de ses encours dans le futur, malgré les perturbations économiques récentes. Tout d'abord, cela suggère une opportunité d'investissement attrayante, permettant à la banque de

consacrer davantage de ressources pour soutenir cette croissance attendue. De plus, cette tendance offre une perspective favorable pour la rentabilité de la banque, renforçant sa position sur le marché. En identifiant les facteurs de risque, en planifiant stratégiquement cette croissance, en diversifiant le portefeuille et en capitalisant sur les avantages concurrentiels, la banque peut saisir les opportunités qui se présentent et prospérer dans un avenir où les encours augmentent ce qui implique une augmentation du PNB corporate.

4.9 Déploiement de la solution

Dans cette section, nous allons présenter la méthodologie de déploiement de la méthode de prédiction choisie précédemment pour qu'elle soit utilisable par les employés du département stratégie performance corporate de SGA. En effet, nous avons proposé de créer une interface utilisateur intuitive pour rendre la prédiction des indicateurs plus facile et plus rapide. Cette proposition d'interface est utile pour la banque car elle permet aux utilisateurs de simplement sélectionner la variable qu'ils souhaitent prédire, le modèle qu'ils souhaitent utiliser et la date pour laquelle ils souhaitent la prédiction. Cette interface simplifie considérablement le processus de prédiction, car il n'est plus nécessaire de modifier le code Python chaque fois que vous souhaitez effectuer une nouvelle prédiction.

4.9.1 Implémentation de l'interface de prédiction

Cette étape consiste à concevoir le frontend et le backend d'une interface utilisateur qui permettra à l'utilisateur de choisir facilement la variable qu'il souhaite prédire, le modèle de prédiction qu'il veut utiliser, et la date pour laquelle il souhaite effectuer la prédiction. Pour la partie frontend on a utilisé html, css et javascript. Lorsque l'utilisateur soumet ces choix, les valeurs sélectionnées par l'utilisateur sont envoyées au backend pour effectuer la prédiction, par la suite les données sélectionnées sont récupérées et traitées par le backend pour effectuer la prédiction en utilisant le modèle sélectionné. Le backend renvoie la prédiction au frontend et le résultat de la prédiction est affiché à l'utilisateur.

Pour lier notre interface frontend à notre modèle de prédiction en backend nous avons utilisé javascript pour envoyer les données du formulaire au backend Python, puis en utilisant flask pour renvoyer les prédictions à l'interface utilisateur. Nous avons créé une API en utilisant le framework flask. Cette API sera responsable de la communication entre le frontend et le backend. Elle doit être capable de recevoir les données d'entrée (indicateur, modèle, date) envoyées par l'utilisateur à partir de l'interface, de faire appel au modèle pour calculer la prédiction, et de renvoyer le résultat au frontend pour affichage de résultat. Nous devons écrire une fonction python pour récupérer les données nécessaires à partir de cette dataset en fonction des entrées de l'utilisateur. Nous pouvons passer les données récupérées au modèle choisi pour effectuer la prédiction. Ce modèle prend les données récupérées en entrée, les prépare pour être compatibles et fait une prédiction pour les 20 prochaines années avec une fréquence mensuelle. Nous renvoyons uniquement les prévisions pour les 20 prochaines années.

Enfin, nous devons écrire une fonction Python qui génère le résultat HTML à afficher à l'utilisateur. Cette fonction prendra les données de prédiction et les intégrera dans le modèle HTML que nous avons défini.

En résumé nous avons mis en place une interface utilisateur qui permet aux employés de choisir les paramètres d'entrée pour une prédiction et que ces valeurs sont ensuite envoyées au backend via une API Flask. Le backend utilise ensuite une bibliothèque précise pour effectuer une prédiction sur les données en fonction des paramètres sélectionnés par l'utilisateur et renvoie le résultat au frontend pour affichage.



Figure 4.28–Interface de visualisation des valeurs prédites en fonction de la date sélectionnée

Vous trouverez le code de chaque étape en annexe J.

Après avoir présenté les résultats de notre solution et répondu aux besoins des spécialistes, nous aborderons dans la dernière partie la conclusion de ce projet en mettant en évidence l'utilité de la solution développée et en identifiant les possibilités d'amélioration.

Conclusion générale

Conclusion Générale

Dans le secteur bancaire en constante évolution, la capacité à prévoir et à anticiper les tendances est devenue essentielle pour maintenir un avantage concurrentiel. Les banques doivent constamment s'adapter aux changements du marché, à l'évolution des préférences des clients, aux réglementations en évolution et à l'émergence de nouvelles technologies. Dans ce contexte dynamique, la capacité à prédire de manière précise et fiable les évolutions financières devient un élément clé de la réussite.

Lors de notre projet au sein de SGA, nous avons exploré une approche alternative pour contourner les contraintes liées au manque de données, parvenir à estimer et fournir un aperçu de l'allure prévue de l'évolution du PNB corporate dans les années à venir. La méthodologie s'est déroulée selon une séquence d'étapes. Tout d'abord, nous avons entrepris de comprendre le produit net bancaire, son utilité, son processus de calcul et les variables qui le génèrent. Nous avons amorcé notre exploration en décomposant minutieusement son équation pour élaborer une vision approfondie de ses différentes composantes. Ensuite, nous avons extrait le dataset corporate du PNB de l'année 2022 à l'aide de Teradata, cette base contient les mouvements créditeurs, les encours, les importations et exportations qui génèrent les commissions et les intérêts avec des proportions différentes. Pour mieux comprendre nos données et analyser l'importance de chaque secteur sur l'évolution du PNB, nous avons créé un tableau de bord avec Power BI. Cela nous a permis d'obtenir une visualisation claire et dynamique de nos données. Nous avons ensuite procédé à une étude statistique approfondie dans le but de prouver le lien fort entre les encours et le PNB. Pour cela, nous avons réalisé une régression multiple et éliminé les variables qui avaient le moins d'impact sur le PNB, en nous basant sur des statistiques telles que le F statistique, le R2 et l'analyse de variance (ANOVA). Nous avons également utilisé le coefficient standardisé pour mesurer l'impact des encours sur le PNB, calculé le coefficient de corrélation et renforcé notre étude en appliquant le test de Granger, qui a révélé une relation de cause à effet entre le PNB et les encours. La prochaine étape de notre approche a été d'étudier notre série temporelle des encours. Nous avons extrait les données à l'aide de l'outil Teradata et créé un tableau de bord Power BI pour comprendre les événements qui rendent notre série non stationnaire. Nous avons décomposé notre série temporelle pour détecter d'éventuelle saisonnalité et vérifier si notre série présentait un bruit blanc. Ensuite, nous avons choisi quatre modèles de prédiction (ARIMA, SARIMA, XGBoost et FB Prophet) en nous basant sur des critères bien définis. Nous avons ajusté les paramètres de chaque modèle pour optimiser leur performance. Nous avons évalué la performance de ces modèles à l'aide de mesures qui calculent l'erreur entre les valeurs réelles et prédites. Les résultats de nos prédictions ont révélé une baisse remarquable des encours pour les 10 prochaines années. Suite à ce résultat, nous avons réfléchi à l'élimination des années qui ont été impactées par des circonstances externes telles que la pandémie de COVID-19. Nous avons décidé de ne conserver que la période allant de 2012 à fin 2018, qui représente l'année modèle pour SGA, et avons refait la prédiction et l'évaluation comme précédemment. Les résultats de cette dernière prédiction ont révélé une augmentation significative des encours pour les 10 prochaines années. Enfin nous avons déployé notre solution sous forme d'une interface utilisateur monopage qui permettra aux employés du département corporate de choisir le modèle de prédiction ainsi que la variable qu'ils souhaitent prédire, afin d'obtenir une estimation de l'évolution de cette variable.

Les résultats de notre étude ont révélé l'importance d'identifier les fluctuations potentielles, les scénarios défavorables et prendre des mesures préventives pour atténuer les impacts négatifs sur les performances financières de la banque notamment le PNB corporate. Pour cela, il nous semble à la fois nécessaire et utile pour SGA de :

- Diversifier les secteurs cibles : au-delà de ses secteurs phares, la banque peut explorer d'autres secteurs à fort potentiel de croissance en se focalisant sur des industries émergentes telles que les énergies renouvelables et la logistique. Une diversification stratégique permettra de créer de nouvelles opportunités d'affaires et d'élargir la base de clients corporate.

- Proposer des services financiers spécialisés et adaptés aux besoins spécifiques des industries porteuses, en fournissant des lignes de crédit flexibles, des financements structurés ou des produits d'assurance personnalisés. Cela attirera de nouveaux clients corporate.

- Investir dans des solutions technologiques innovantes pour simplifier et accélérer les processus financiers des entreprises. Par exemple, elle peut développer des plateformes de gestion financière intégrées, des solutions de gestion de trésorerie en temps réel ainsi que des applications basées sur l'intelligence artificielle (IA). L'innovation technologique aidera à renforcer la relation avec les clients corporate et à améliorer leur expérience globale.

- Faciliter les transactions internationales, offrir des services de change avantageux, fournir des conseils en matière de commerce international et soutenir les entreprises dans leur expansion à l'étranger. Cela permettra d'attirer de nouveaux clients corporatifs et d'augmenter les flux financiers liés au commerce international.

Notre solution peut être améliorée dans le futur suivant plusieurs axes :

- Étendre la période de collecte de données pour inclure plusieurs années ce qui peut être particulièrement bénéfique pour les modèles de prédiction basés sur les réseaux de neurones.

- Personnaliser davantage l'interface en fonction de leurs besoins spécifiques, cela peut inclure des fonctionnalités telles qu'une assistance contextuelle ou des visualisations interactives pour illustrer les tendances et les projections de manière plus dynamique.

- Exploiter les avancées de l'IA pour améliorer les capacités de prédiction.

- Prise en compte de facteurs externes en les intégrant dans la dataset sous forme d'indicateurs qui peuvent influencer le PNB, tels que les fluctuations du PIB, de l'inflation et du taux de chômage.

Pour finir, notre projet a démontré l'efficacité de la science de données et l'intelligence artificielle dans un contexte spécifique, où l'environnement bancaire est caractérisé par son incertitude et sa nature stochastique. Malgré les défis posés par le manque de données et les facteurs externes notre approche a fourni des résultats en identifiant la variable clés qui influence le PNB et construire des prévisions robustes. Cette réussite démontre la pertinence des méthodes statistiques et d'apprentissage machine dans un environnement complexe et volatile, où les méthodes traditionnelles d'analyse et de prédiction peuvent être insuffisantes. Grâce à notre approche innovante, Société Générale Algérie peut tirer parti des avantages de l'apprentissage machine pour prendre des décisions éclairées, anticiper les tendances du marché et augmenter ses encours corporate, afin d'optimiser ses bénéfices.

Bibliographie

Références

- AKLI Chafiaa et ALIK Leila ; « Performance des agences bancaires : Etude comparative de l'agence 187 et de l'agence 120 du Groupe CPA Tizi-Ouzou Etude Comparative De La Performance des Agences Bancaires Agence 187 et Agence 120 du Groupe CPA Tizi-Ouzou » ; mémoire en science économique ; option : Monnaie, Finance et Banques ; tizi-ouzou ; 2016 ; Consulté sur <https://www.ummo.dz/dspace/handle/ummo/5053>
- AGBODAN M.M. et AMOUSSOUGA F.G. « Les facteurs de performance de l'entreprise », Actualité Scientifique, France, 1995 ; Consulté sur books.google.com
- AMOUR Ounissa , KAB Younes,mémoire de fin d'étude en vue de l'obtention du Diplôme de Master en en Finance d'Entreprise ; «Financement des investissements : Etude comparative entre banque publique (CNEP-Banque) et banque privée (SGA)»; 2014,Consulté sur <http://www.univ-bejaia.dz/xmlui/bitstream/handle/123456789/2875/Financement%20des%20investissements%20%20Etude%20comparative%20entre%20banque%20publique.pdf?sequence=1>
- Alexei Botchkarev ; « Evaluating performance of regression machine learning models using multiple error metrics in Azure Machine Learning Studio » ; 2018 ; Consulté sur https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3177507
- Angèle Renaud ; «Etat de l'art sur la performance» ; 2012 ,Consulté sur hal.science
- BEL HADJ SALAH Maher,mémoire de fin d'étude;«Les déterminants de la performance bancaire : Cas des banques Tunisiennes »;2020 Consulté sur https://ifid-concours.org/biblio/38%C3%A8me_Banque/Bel_Haj_Salah_Maher.pdf
- Bâle II ,«Rappel sur la réglementation Bâle II mise à jour»; le 30 JANVIER 2023; Consulté sur [lafinancepourtous](https://www.bis.org/press/pr1202301.htm)
- Croston, J.D; « Forecasting and stock control for intermittent demands » ;1972 ; Consulté sur [jstor.org/stable](https://www.jstor.org/stable/1600000)
- Danièle Nouy ,«La rentabilité des banques françaises »;2012,Consulté sur [Revue d'économie financière](https://www.revue-efm.fr/)
- Douglas C. Montgomery, Cheryl L. Jennings et Murat Kulahci ; « Introduction to Time Series Analysis and Forecasting » ; 2008; Consulté sur <https://pedro.unifei.edu.br/download/Montgomery.pdf>

-Gregory Heem;« La performance du secteur bancaire : d'une performance financière vers une performance durable»;2021,Consulté sur hal.science

-Guillaume SAUPIN ; « Gradian Boosting , Exploitez les arbres de décision pour le Machine Learning (XGBoost, CatBoost, LightGBM) » ; 2022 ; Consulté sur <https://static.fnac-static.com/multimedia/editorial/pdf/9782409034022.pdf>

- Heba Elsega; «Granger-Causality Inference of the Existence of Unobserved Important Components in Network Analysis » ; Juillet 2021 ; Consulté sur https://www.researchgate.net/publication/353590160_Granger-Causality_Inference_of_the_Existence_of_Unobserved_Important_Components_in_Network_Analysis

-Help center.*Test de racine unitaire (Dickey-Fuller) et de stationnarité d'une série chronologique* [en ligne]. [Consulté le 10/03/2023]. Consulté sur <https://help.xlstat.com/fr/6697-test-de-racine-unitaire-dickey-fuller-et-de>

-JUDIA PEARL ; « Causality, Second Edition : MODELS, REASONING AND INFERENCE » ; 2009 ; Consulté sur <https://yzhu.io/courses/core/reading/04.causality.pdf>

-Keerthana V.*Understanding The Basics of Time Series Forecasting* [en ligne]. [Consulté le 05/04/2023]. Consulté sur <https://www.analyticsvidhya.com/blog/2021/07/abc-of-time-series-forecasting/>

-LORINO PHILIPPE, « Méthodes et Pratiques de la Performance », Les Editions d'Organisation, Paris, 1998 ; Consulté sur <https://www.ummtto.dz/dspace/handle/ummtto/5053>

- Minitab.*Qu'est-ce que la statistique Ljung-Box Q (LBQ) ?* [en ligne]. [Consulté le 13/04/2023]. Consulté sur <https://support.minitab.com/fr-fr/minitab/20/help-and-how-to/statistical-modeling/time-series/supporting-topics/diagnostic-checking/what-is-the-ljung-box-q-statistic/>

- Mordor intelligence.*Financial Analytics Market Size & Share Analysis - Growth Trends & Forecasts (2023 - 2028)* [en ligne]. [Consulté le 23/02/2023]. Consulté sur <https://www.mordorintelligence.com/industry-reports/financial-analytics-market>

-Philippe GARSUAULT et Stéphane PRIAMI ; « La banque : Fonctionnement et stratégies » ; 1997, Consulté sur <http://excerpts.numilog.com/books/9782717834567.pdf>

-Philippe St-Aubin et Bruno Agard ; « Precision and Reliability of Forecasts Performance Metrics » ; 2022, Consulté sur

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwie_a6A_tL_AhWK7LsIHddJDYYQFnoECBcQAQ&url=https%3A%2F%2Fwww.mdpi.com%2F2571-9394%2F4%2F4%2F48%2Fpdf&usg=AOvVaw2-X3PeoWQACQ6Z9m2_vUJl&opi=89978449

-Samir Abderrazek Srairi ; « Analyse des déterminants de la rentabilité des banques commerciales saoudiennes » ; Revue Marocaine de Gestion et d'Économie ; N°2, Janvier-Mai 2010 ; Consulté sur <https://revues.imist.ma/index.php/RMGE/article/view/4665/3299>

-Timothy Z. Keith ; «Multiple Regression and Beyond An Introduction to Multiple Regression and Structural Equation Modeling » ; 2015 ; Consulté sur <http://sadbhavnpublications.org/research-enrichment-material/2-Statistical-Books/Multiple-Regression-and-Beyond-An-Introduction-Multiple-Regression-and-Structural-Equation-Modeling.pdf>

-William W.S. Wei ; « Time series analysis, second edition » ; 2006 ; Consulté sur <https://civil.colorado.edu/~balajir/CVEN6833/lectures/wwts-book.pdf>

-Peter J. Brockwell and Richard A. Davis ; « Introduction to Time Series and Forecasting, Second Edition » ; 2002 ; Consulté sur <http://home.iitj.ac.in/~parmod/document/introduction%20time%20series.pdf>

-Yannig Goude.*Les processus arima* [en ligne]. [Consulté le 15/03/2023]. Consulté sur https://www.imo.universite-paris-saclay.fr/~yannig.goude/Materials/time_series/cours6_ARIMA.pdf

-Zach.*How to Perform a Ljung-Box Test in Python* [en ligne]. [Consulté le 13/04/2023]. Consulté sur <https://www.statology.org/ljung-box-test-python/>

Annexes

Annexe A

Formule mathématique :

1. RSS :

$$RSS = \sum (y_i - \bar{y})^2 \text{ où :}$$

- y_i est la valeur observée de la variable dépendante (ou cible) pour la i-ème observation.
- \bar{y} est la valeur moyenne observée de la variable dépendante.

2. ESS

$$ESS = \sum (\bar{y} - \hat{y}_i)^2 \text{ où :}$$

- \bar{y} est la valeur moyenne observée de la variable dépendante.
- \hat{y}_i est la valeur prédite de la variable dépendante pour la i-ème observation.

3. TSS

$$TSS = \sum (y_i - \bar{y})^2 \text{ où :}$$

- y_i est la valeur observée de la variable dépendante pour la i-ème observation.
- \bar{y} est la valeur moyenne observée de la variable dépendante.

4. NRMSE = (RMSE / (Max - Min)) * 100

Où :

RMSE représente l'erreur quadratique moyenne (Root Mean Square Error)

Max est la valeur maximale de la variable cible

Min est la valeur minimale de la variable cible

5. WMAPE = ($\sum(|A_t - F_t| / \sum(A_t))$) * 100 / n

où :

A_t est la valeur réelle à l'instant t

F_t est la valeur prédite à l'instant t

$\sum(A_t)$ est la somme de toutes les valeurs réelles

n est le nombre total de points de données

Annexe B

Série temporelle

B.1 Domaine d'application :

Les séries temporelles sont omniprésentes dans de nombreux domaines d'application. En voici quelques exemples :

Dans le domaine de l'économie, les séries temporelles sont utilisées pour étudier l'évolution des indicateurs économiques tels que le taux de chômage, les prix des produits de consommation, la croissance économique, etc. Les données économiques sont souvent collectées à des intervalles réguliers, tels que mensuels ou trimestriels.

En finance, les séries temporelles sont largement utilisées pour étudier l'évolution des marchés financiers, tels que les cours des actions, les taux de change, les taux d'intérêt, etc. Les données financières sont souvent collectées à des intervalles de temps très courts, allant de quelques secondes à quelques minutes.

Dans le domaine de l'environnement, les séries temporelles sont utilisées pour étudier l'évolution des phénomènes naturels tels que la météo, la qualité de l'air, les niveaux d'eau, les niveaux de pollution, etc. Les données environnementales sont souvent collectées à des intervalles réguliers, tels que quotidiens ou hebdomadaires.

En médecine, les séries temporelles sont utilisées pour étudier l'évolution de différents aspects de la santé humaine, tels que la tension artérielle, la fréquence cardiaque, le taux de glucose dans le sang, etc. Les données médicales sont souvent collectées à des intervalles de temps très courts, allant de quelques secondes à quelques minutes, en utilisant des dispositifs de surveillance tels que des moniteurs cardiaques ou des lecteurs de glycémie.

En résumé, les séries temporelles sont largement utilisées dans de nombreux domaines d'application, et sont souvent collectées à des intervalles réguliers ou à des intervalles de temps très courts, en fonction de la nature de l'objet d'étude.

Annexe C

Machine learning

C.1 Les types d'apprentissage automatique :

- Apprentissage supervisé : Le modèle est entraîné sur des données étiquetées, où il apprend à prédire des étiquettes ou des résultats spécifiques pour de nouvelles données.
- Apprentissage non supervisé : Le modèle analyse des données non étiquetées afin de découvrir des structures, des modèles ou des regroupements intrinsèques sans étiquettes préalables.
- Apprentissage par renforcement : Le modèle interagit avec un environnement dynamique et apprend à prendre des décisions en maximisant les récompenses à long terme.
- Apprentissage semi-supervisé : Le modèle utilise à la fois des données étiquetées et non étiquetées pour améliorer les performances de prédiction, en tirant parti de l'abondance des données non étiquetées.
- Apprentissage par transfert : Le modèle utilise des connaissances préalablement acquises sur une tâche source pour améliorer ses performances sur une nouvelle tâche similaire

C.2 Le sur apprentissage et le sous apprentissage :

Overfitting, ou surapprentissage en français, se produit lorsque le modèle s'est trop adapté aux données d'entraînement spécifiques au point de ne pas généraliser correctement sur de nouvelles données. En conséquence, le modèle peut mémoriser les exemples d'entraînement au lieu de comprendre les motifs et les relations sous-jacents, ce qui entraîne une mauvaise performance lors de la prédiction de nouvelles données.

L'underfitting, ou sous-apprentissage en français, se produit lorsque le modèle n'est pas suffisamment complexe pour capturer les motifs et les relations des données. Cela se traduit par une performance médiocre, à la fois sur les données d'entraînement et sur les nouvelles données.

C.3 Deep learning :

Le deep learning, également connu sous le nom d'apprentissage profond, est une branche de l'apprentissage automatique (machine learning) qui se concentre sur l'utilisation de réseaux de neurones artificiels profonds pour résoudre des problèmes complexes. Il s'inspire du fonctionnement du cerveau humain et vise à permettre aux ordinateurs d'apprendre et de prendre des décisions de manière autonome.

LSTM (Long Short-Term Memory) est un type spécifique de réseau de neurones récurrents (RNN) largement utilisé dans le domaine du deep learning. Contrairement aux réseaux de neurones classiques, qui peuvent avoir des difficultés à traiter des séquences longues, les LSTM sont conçus pour capturer les dépendances à long terme dans les données séquentielles. L'une des principales caractéristiques des LSTM est leur capacité à conserver et à oublier des

informations au fil du temps grâce à l'utilisation de portes de mémoire. Ces portes permettent au modèle de décider quelles informations doivent être mises à jour, quelles informations doivent être oubliées et quelles informations doivent être transmises aux étapes suivantes du traitement.

Les LSTM sont particulièrement adaptés pour traiter des séquences de données de longue durée, où les dépendances à long terme sont importantes pour la tâche à accomplir. Les LSTM ont généralement besoin d'une grande quantité de données pour fonctionner efficacement. Cela est dû à leur complexité et à leur capacité à capturer des motifs complexes et des dépendances à long terme

Annexe D

Les composantes du ARIMA

D.1 Modèle Auto-régression (AR) :

Il utilise la propriété d'auto-régression de la série, ce qui signifie qu'il est uniquement applicable aux séries qui ont cette propriété. Une série est auto-régressive d'ordre n si sa valeur à un moment donné dépend linéairement des n valeurs précédentes.

Le modèle AR consiste à trouver les coefficients π_i , la constante c et la variance du bruit ϵ_t , en utilisant la formule :

$$x_t = c + \epsilon_t + n \sum_{i=1}^n \pi_i x_{t-i}$$

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

On peut déterminer les coefficients π_i en utilisant une régression linéaire, par exemple. Il est important de noter que ce modèle ne doit être appliqué que s'il y a une corrélation entre la série et une version décalée d'elle-même (autocorrélation).

D.2 Modèle moving average MA :

Plutôt que de se baser sur les valeurs antérieures de la variable à prévoir dans une régression, un modèle de moyenne mobile utilise les erreurs de prévision passées dans un modèle similaire à une régression.

$$x_t = \mu + n \sum_{i=1}^n \theta_i \epsilon_{t-i}$$

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

et Le terme "integrated" (I) dans ARIMA signifie qu'il est nécessaire de différencier la série originale afin de la rendre stationnaire et éliminer toute non-stationnarité potentielle.

D.3 Paramètres :

ARIMA est un modèle caractérisé par 3 paramètres p, d et q
ou :

p représente l'ordre du terme AR

q représente l'ordre du terme MA

d représente le nombre de différenciations nécessaires pour rendre la série temporelle stationnaire.

Et le modèle complet peut s'écrire :

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Annexe E :

Présentation des outils informatiques employés

E.1 Présentation de Python

Python est un langage de programmation moderne et polyvalent, qui offre plusieurs approches pour résoudre des problèmes. Il est interprété, ce qui signifie que le code peut être exécuté sur différents systèmes d'exploitation tels que Windows, Mac et Linux. Python est largement utilisé dans des domaines tels que le calcul scientifique, la science des données, le développement web et même comme langage de script pour étendre les fonctionnalités d'autres logiciels. Ce qui distingue Python, c'est sa vaste communauté et sa multitude de bibliothèques open source disponibles. Cette richesse lui a permis de devenir le langage de programmation le plus populaire et le plus utilisé.

E.2 Présentation des bibliothèques employées

E.2.1 Pandas et NumPy

Pandas est une bibliothèque Python qui simplifie la manipulation de jeux de données tabulaires. Elle permet de charger des données à partir de différentes sources telles que des fichiers CSV, des fichiers Excel et des bases de données. Pandas offre un large éventail de méthodes pour filtrer, extraire, transformer et effectuer diverses opérations sur les données. La bibliothèque est construite sur la base de NumPy, une autre bibliothèque Python pour le calcul et la manipulation de matrices.

E.2.3 Matplotlib et Seaborn

Matplotlib est une bibliothèque Python utilisée pour créer des graphiques et des visualisations de données de manière flexible et personnalisée. Elle offre une large gamme de fonctionnalités pour créer des graphiques en 2D et en 3D, des diagrammes à barres, des histogrammes, des diagrammes en boîte, des nuages de points, des courbes, etc. Seaborn est une bibliothèque Python basée sur Matplotlib, qui offre des fonctionnalités supplémentaires pour la création de graphiques statistiques attrayants et informatifs.

E.2.4 statsmodels

Statsmodels est une bibliothèque qui propose des méthodes statistiques, des modèles économétriques et des tests statistiques pour l'analyse de données. Avec statsmodels.api, vous pouvez effectuer diverses analyses statistiques telles que la régression linéaire, la régression logistique, l'analyse de variance (ANOVA), les séries temporelles, les modèles ARIMA, etc. Il offre également des outils pour effectuer des tests statistiques, des estimations de paramètres, des prédictions et des diagnostics de modèles. Cette bibliothèque fournit également plusieurs

fonctions comme `seasonal_decompos` qui décompose une série temporelle en ses composantes saisonnières, tendanciennes et résiduelles.

E.2.5 Scikit-learn

Scikit-learn est une bibliothèque Python largement utilisée pour l'apprentissage automatique. Elle offre une vaste gamme d'algorithmes et d'outils pour la classification, la régression, le regroupement et d'autres tâches d'exploration de données. Scikit-learn fournit également des métriques couramment utilisées telles que le MSE (Mean Squared Error), le MAE (Mean Absolute Error) et le R2 Score pour évaluer les performances des modèles. De plus, la bibliothèque propose des fonctionnalités pour diviser les données en ensembles d'entraînement et de test, facilitant ainsi l'évaluation et la validation des modèles.

E.2.6 Xgboost :

XGBoost est une bibliothèque d'apprentissage automatique largement utilisée qui se concentre sur la création de modèles de boosting. Elle offre une solution puissante pour combiner efficacement plusieurs modèles simples, ce qui permet de créer un modèle global plus précis. XGBoost est apprécié pour sa capacité à gérer de grandes quantités de données, sa rapidité d'exécution et sa flexibilité. Il est largement utilisé dans divers domaines tels que la régression, la classification et la prédiction.

E.2.7 Prophet

Prophet est une bibliothèque open source développée par Facebook pour la prévision et la modélisation des séries temporelles. Elle est conçue pour simplifier le processus de création de modèles prédictifs basés sur des données temporelles, même pour les utilisateurs non experts en statistiques.

E.3 HTML

HTML, qui signifie HyperText Markup Language (langage de balisage hypertexte), est le langage standard utilisé pour créer et structurer le contenu des pages Web. Il est utilisé pour décrire la structure et la présentation des informations sur une page Web, en utilisant des balises et des éléments spécifiques. HTML repose sur une syntaxe constituée de balises entourant le contenu

E.4 CSS

CSS, qui signifie Cascading Style Sheets (feuilles de style en cascade), est un langage de style utilisé pour définir la présentation et l'apparence des éléments d'une page HTML. Il fonctionne en séparant la structure et le contenu d'une page HTML de son aspect visuel

E.5 JavaScript

JavaScript est un langage de programmation utilisé pour créer des fonctionnalités interactives et dynamiques sur les pages Web. Il est principalement utilisé côté client (dans le navigateur), mais il peut également être utilisé côté serveur (avec des plateformes telles que Node.js). Avec JavaScript, vous pouvez manipuler le contenu HTML, réagir aux événements de l'utilisateur,

effectuer des calculs, envoyer des requêtes HTTP, créer des animations, gérer des formulaires, valider des données et bien plus encore.

E.6 URL et HTTP

URL est l'acronyme de Uniform Resource Locator (Localisateur Uniforme de Ressources en français). Il s'agit d'une chaîne de caractères utilisée pour adresser et localiser des ressources sur Internet. En termes simples, c'est l'adresse web que vous saisissez dans un navigateur pour accéder à un site web ou à une page web spécifique. Une URL typique se compose de plusieurs composants : protocole qui indique le protocole de communication à utiliser, tel que HTTP (Hypertext Transfer Protocol) ou HTTPS (HTTP sécurisé) pour les connexions sécurisées, Nom de domaine et l'emplacement spécifique de la ressource sur le serveur.

E.7 Les tableaux de bord sur power bi

La Business Intelligence (BI) est un ensemble de méthodes, technologies et outils qui permettent aux entreprises de collecter, organiser, analyser et visualiser leurs données afin de prendre des décisions éclairées et de développer leur activité de manière stratégique.

Un tableau de bord sur Power BI est une représentation visuelle et interactive des données et des indicateurs clés de performance (KPI) d'une entreprise ou d'une organisation. Power BI est une plateforme de business intelligence développée par Microsoft. Elle offre des fonctionnalités puissantes pour la visualisation des données, la création de rapports et l'analyse de données, il peut contenir différents types d'éléments visuels, tels que des graphiques, des tableaux, des cartes, des indicateurs de performance, des jauges, des diagrammes et bien d'autres encore. Ces éléments visuels peuvent être configurés pour afficher des données en temps réel, des tendances, des comparaisons, des filtres interactifs, etc.

E.8 Khi 2 :

En statistique, le test du chi² (chi-carré) est une méthode utilisée pour évaluer la relation entre deux variables catégorielles. Il est utilisé pour déterminer si les différences observées entre les fréquences observées et les fréquences attendues dans un tableau croisé sont statistiquement significatives.

E.9 Vraisemblance :

La vraisemblance (likelihood en anglais) est un concept fondamental en statistique et en théorie de l'estimation. Elle est utilisée pour mesurer à quel point un ensemble de données est compatible avec un modèle statistique spécifique.

En termes simples, la vraisemblance est une mesure de la probabilité des données observées, compte tenu des paramètres d'un modèle statistique. Plus précisément, la vraisemblance est définie comme la fonction de densité de probabilité (ou de masse) des données observées, considérée comme une fonction des paramètres du modèle. L'objectif est de trouver les valeurs des paramètres qui maximisent cette fonction de vraisemblance.

Annexe F :

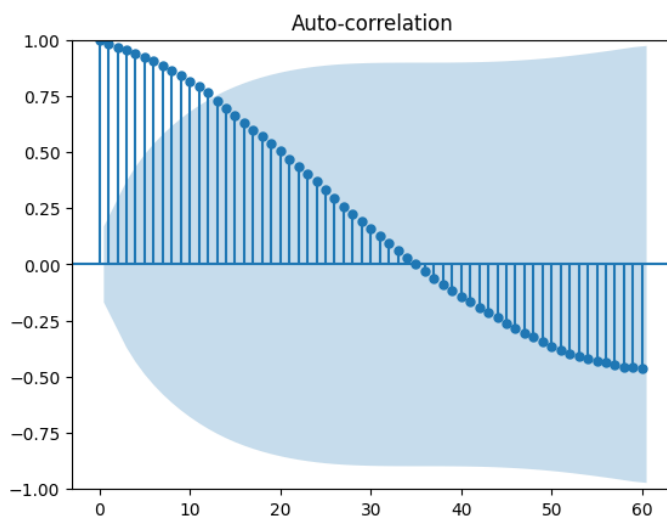
Les secteurs d'activité :

Secteur d'activité	Description
1- Agroalimentaire	Désigne l'ensemble des activités économiques liées à la production, la transformation, la distribution et la commercialisation des produits alimentaires. Cela inclut la production de matières premières agricoles, la transformation de ces matières premières en produits finis , ainsi que la distribution et la vente de ces produits finis
2-Immobilier et BTP	Le secteur immobilier comprend toutes les activités liées à la construction, l'achat, la vente, la location et la gestion de biens immobiliers.Le secteur du BTP (Bâtiment et Travaux Publics) comprend toutes les activités liées à la construction et la rénovation des bâtiments et des infrastructures publiques.
3-Sous traitance industrielle	Le secteur de la sous-traitance industrielle et de services est un secteur qui regroupe l'ensemble des entreprises spécialisées dans la fourniture de services ou de produits spécifiques pour d'autres entreprises, souvent dans le cadre d'un contrat de sous-traitance.
4-Industrie pharmaceutique et médicale	Ce secteur est lié à l'industrie des soins de santé, qui inclut la recherche, la production, la vente et la distribution de médicaments, de dispositifs médicaux et d'autres produits de santé diagnostic.
5-Telecommunication	L'ensemble des activités économiques liées à la transmission de l'information à distance. Cela comprend la téléphonie fixe et mobile, l'Internet, la télévision et la radio diffusées par voie hertzienne, par câble ou par satellite
6-Énergétique	Le secteur énergétique englobe toutes les activités liées à la production, la transformation, le transport, la distribution et la commercialisation des sources d'énergie. Cela inclut notamment l'exploitation des hydrocarbures, l'énergie électrique , ainsi que le transport et la distribution de ces sources d'énergie vers les différents secteurs consommateurs.
7-Transport et logistique	Il regroupe l'ensemble des activités liées à la circulation des marchandises et des personnes, ainsi que leur stockage et leur distribution. Cela inclut le transport routier, ferroviaire, maritime et aérien, ainsi que la logistique et la gestion des flux de marchandises

Annexe G :

ACF

1- le graphe de l'autocorrélation (ACF) pour examiner la corrélation dans les résidus de la série temporelle. L'ACF est un outil graphique qui représente la corrélation entre une observation et ses observations retardées.



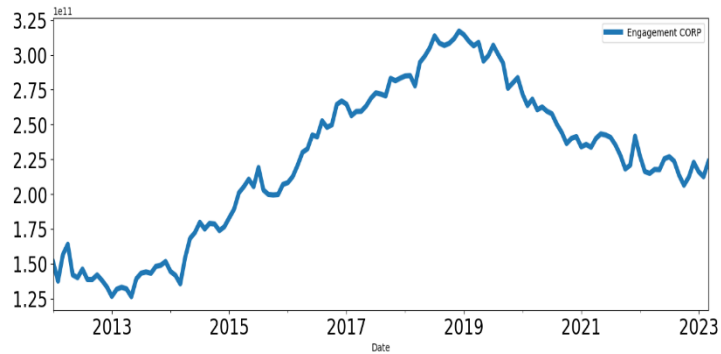
Le fait que les valeurs de l'ACF diminuent lentement suggère qu'il existe une corrélation significative entre les observations à ces retards. Cela signifie que les observations passées ont une influence sur les observations futures à ces intervalles.

Annexe H :

La stationnarité

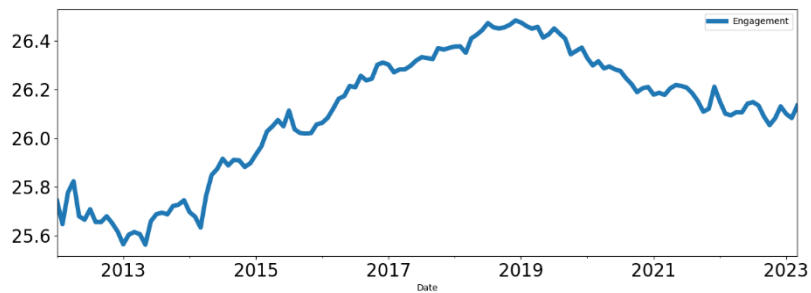
H.1 La visualisation du graphique :

1- Données initiales :



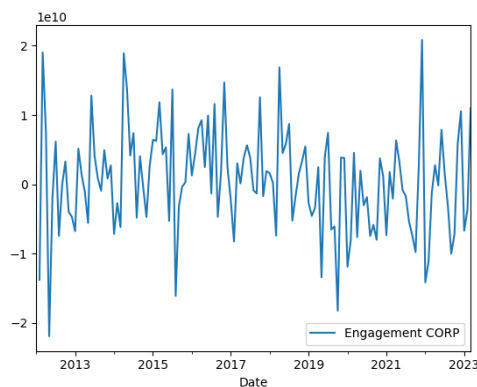
Évolution de l'engagement à partir du janvier 2012 jusqu'à mars 2023

2- Données transformées par Log



Évolution de l'engagement après l'application du logarithme de 2012 à 2023

3- Données transformées par différenciation



Évolution de l'engagement après l'application de la différenciation de 2012 à 2023

Les deux premiers graphiques (données initiales et transformation logarithmique) ne montrent pas de stationnarité claire, tandis que le troisième (transformation par différenciation) présente une forme qui suggère une certaine stationnarité.

H.2 Vérification des caractéristiques statistiques :

Nous avons également examiné la cohérence de la série chronologique en vérifiant si la première partie des données avait des caractéristiques statistiques similaires à la seconde partie. Nous avons divisé la série en deux parties égales et avons comparé les moyennes et les variances de chaque partie

	Les données initiales	Transformation par log	Transformation par différenciation
Moyenne 1	188022116053.950958	25.929469	1798845614.243781
Moyenne 2	260488252837.746277	26.277061	1798845323.462687
Variance 1	2.14454×10^{21}	0.060212	5.6495632×10^{19}
Variance 2	1.1887×10^{21}	0.017547	5.6495633×10^{19}

Comparaison des caractéristiques statistiques entre la première et la seconde partie des données initiales, avec transformation logarithmique et des données différenciées.

Les moyennes des deux parties des données initiales sont très différentes, indiquant une variation significative des niveaux moyens. De plus, les variances des deux parties sont également très élevées et très différentes l'une de l'autre, ce qui suggère une dispersion importante et variable des données. Ainsi, les résultats montrent une certaine stabilité des moyennes après la transformation logarithmique et la différenciation, et les variances sont également similaires et relativement élevées, indiquant une dispersion des données relativement stable. Cela peut suggérer une certaine stationnarité en termes de niveaux moyens, mais des tests statistiques formels sont nécessaires pour confirmer la stationnarité.

Annexe I

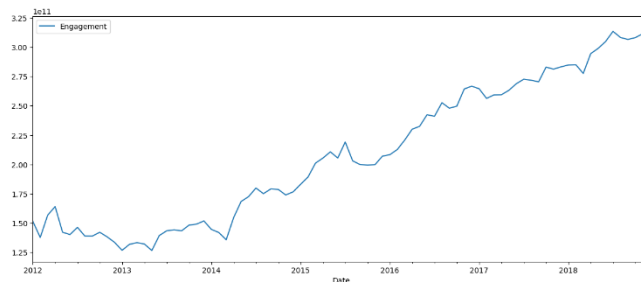
Etude complémentaire

Nous avons suivi les mêmes étapes de la prédiction des données entière, nous allons présenter que les résultats de chaque étapes :

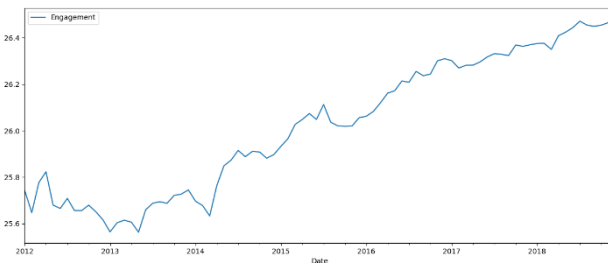
I.1 Evaluation de la stationnarité :

La visualisation du graphique :

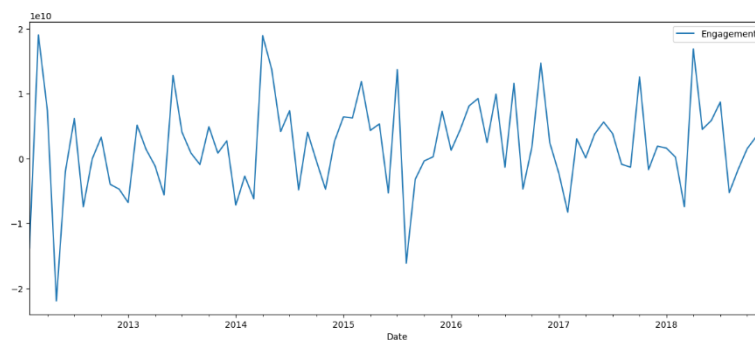
1- données initiales



2- données transformées par log



3-données transformées par différenciation



Les deux premiers graphiques, représentant les données initiales et la transformation logarithmique, ne révèlent pas de tendance claire vers la stationnarité. En revanche, le

troisième graphique, qui présente la transformation par différenciation, montre une forme qui suggère une certaine stationnarité.

I.2 Vérification des caractéristiques statistiques :

	Les données initiales	Transformation par log	Transformation par différenciation
Moyenne 1	156462899780.158722	25.765758	1613571540.817461
Moyenne 2	260583030736.238098	26.276723	2389839175.951220
Variance 1	549067914537921609728	0.020571	66113086897143037952
Variance 2	1263863240737187168256	0.019821	40082827343786131456

I.3 Test de Dickey-Fuller augmenté (ADF):

	Les données initiales	Transformation par log	Transformation par différenciation
ADF Statistic	0.546529	-0.024741	-10.065991
p-value	0.986221	0.956438	1.295887758206393x 10 ¹⁷
Stationnarité	Non-stationnaire	Non-stationnaire	Stationnaire

Les données initiales et les données après avoir appliqué la transformation logarithmique ont une p-value supérieure à 0,05, ce qui indique que la série chronologique est non-stationnaire. Cependant, lorsque nous avons appliqué la différenciation, nous avons obtenu une p-value inférieure à 0,05, ce qui suggère que la série chronologique est devenue stationnaire.

I.4 Évaluation des erreurs

	ARIMA	SARIMA	XGBoost	Fbprophet
NRMSE	30.431	8.5149	33.985	6.0563
WMAPE	28.267	2.438	46.229	1.142
R²	0.194	0.721	0.224	0.813

Annexe J :

Code :

I.1 régression

Définir les variables endogène et exogènes

```
y = data2['PNB 2022 MDZD']  
x1=data2['Mouvements créditeurs 2022']  
x2 = data2['Import 2022']  
x3 = data2['Export 2022']  
x4=data2['Encours']
```

Le coefficient de corrélation

```
# Calculer la corrélation entre les variables "encours" et "PNB"  
correlation = data2["Encours"].corr(data2["PNB 2022 MDZD"])  
  
# Afficher la corrélation  
print("La corrélation entre les variables encours et PNB est de:", correlation)
```

Application du modèle de régression multiple

```
x = sm.add_constant(pd.concat([x1,x2,x3,x4], axis=1))  
model = sm.OLS(y, x).fit()
```

Le coefficient standardisé

```

# Obtenir la liste des coefficients et leurs valeurs
coef = model.params
coef_std = model.params / y.std()
print(coef_std)

# Trouver le coefficient standardisé le plus grand
max_coef_std = max(abs(coef_std[1:]))
print(max_coef_std)

# Trouver la variable correspondante
var_name = coef_std[1:].index[abs(coef_std[1:]) == max_coef_std][0]
print(var_name)

# Afficher le résultat
print(f'La variable avec l\'impact le plus important est "{var_name}"')

```

Comparaison entre les modèles restants : import et encours

```

X = data2[["Import 2022"]]
Y = data2["PNB 2022 MDZD"]

# ajout d'une constante à X pour calculer l'ordonnée à l'origine
X = sm.add_constant(X)

# ajustement du modèle de régression linéaire multiple
model = sm.OLS(Y, X)

# Ajuster le modèle aux données
results = model.fit()
# Effectuer le test F pour la variable "Import"
f_test = results.f_test("Import 2022 = 0")

# Afficher la valeur de F et la p-valeur associée
print(f_test)

```

Granger Causality :

```

from statsmodels.tsa.stattools import grangercausalitytests

```

```

data2=data[['PNB 2022 MDZD', 'Encours']]
train_size = int(len(data2) * 0.7)
train_data, test_data = data2[0:train_size], data2[train_size:len(data2)]

```

```

maxlag = 2
test_results = grangercausalitytests(train_data[['PNB 2022 MDZD', 'Encours']], maxlag, verbose=False)

```

```

for lag in range(1, maxlag+1):
    print(f'lag = {lag}')
    print('Granger causality results:', test_results[lag][0]['ssr_ftest'])
    print('p-value:', test_results[lag][0]['ssr_ftest'][1])

```

I.2 Modèles

Pour la division des données :

```

train,test =train_test_split(df, test_size =0.10 ,shuffle=False)

```

ARIMA :

ACF et PACF :

```

lag_acf = acf(data, nlags =140)
lag_pacf = pacf(data, nlags =60)
fig, ax = plt.subplots(1,2,figsize=(20,5))

plot_acf(lag_acf,ax=ax[0])
plot_pacf(lag_pacf,lags = 20,ax=ax[1])
plt.show()

```

Modèle et son ajustement :

```

from statsmodels.tsa.arima.model import ARIMA
model_arima =ARIMA(train,order=(4,1,3)).fit()

```

Prédire les valeurs de test :

```

pred = model_arima.predict(start=120,end=(len(df)-1))

```

Tracer le graphe du test :

```

train['Engagement'].plot(figsize = (12,8), legend = True)
test['Engagement'].plot( legend = True, color='red')
pred.plot(legend = True, color='yellow', alpha=.7)

```

Appliquer la méthode cumsum et tracer le graphe :

```

# Appliquer la methode cumsum sur les données differenciées test et pred_diff
test_cumsum = test.cumsum()
pred_diff_cumsum = pred_diff.cumsum()

# Ajouter la dernière observation de train aux séries cumulées
last_train_obs = train['Engagement'][-1]
test_pred = pd.Series(last_train_obs + pred_diff_cumsum, index=test.index)
pred_final = pd.Series(last_train_obs + pred_diff_cumsum[len(test):], index=index)

# Afficher les graphiques
data['Engagement'].plot(figsize=(12,8), legend=True)
test['Engagement'].plot(legend=True, color='red')
#test_pred.plot(legend=True, color='blue', alpha=.7)
pred_final.plot(legend=True, color='green', alpha=.7)

```

SARIMA :

Le code pour générer toutes les combinaisons possibles de triplets p, d, q :

```
# Define the p, d and q parameters to take any value between 0 and 2
p = d = q = range(0, 2)

# Generate all different combinations of p, q and q triplets
pdq = list(itertools.product(p, d, q))

# Generate all different combinations of seasonal p, q and q triplets
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]

print('Examples of parameter combinations for Seasonal ARIMA...')
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[1]))
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[2]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[3]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[4]))
```

Nous avons généré toutes les combinaisons possibles de paramètres. En ajustant chaque modèle avec ces paramètres :

```
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = sm.tsa.statespace.SARIMAX(data,
                                             order=param,
                                             seasonal_order=param_seasonal,
                                             enforce_stationarity=False,
                                             enforce_invertibility=False)

            results = mod.fit()

            print('ARIMA{}x{}12 - AIC: {}'.format(param, param_seasonal, results.aic))
        except:
            continue
```

La création et l'ajustement du modèle :

```
mod = sm.tsa.statespace.SARIMAX(data,
                                 order=(1, 1, 1),
                                 seasonal_order=(1, 1, 1, 12),
                                 enforce_stationarity=False,
                                 enforce_invertibility=False)

results = mod.fit()

print(results.summary().tables[1])
```

La prédiction des données test :

```
pred = results.get_prediction(start= 121, dynamic=False)
pred_ci = pred.conf_int()
```

La visualisation du graphe comparatif entre les données de test et les prédictions :

```

ax = data['2020:'].plot(label='observed')
pred.predicted_mean.plot(ax=ax, label='One-step ahead Forecast', alpha=.7)

ax.fill_between(pred_ci.index,
                pred_ci.iloc[:, 0],
                pred_ci.iloc[:, 1], color='k', alpha=.2)

ax.set_xlabel('Date')
ax.set_ylabel('Engagement')
plt.legend()

plt.show()

```

XGBOOST :

La fonction qui permet d'extraire les informations trimestrielles, mensuelles et annuelle :

```

def create_features(df, label=None):
    df['date'] = df.index

    df['quarter'] = df['date'].dt.quarter
    df['month'] = df['date'].dt.month
    df['year'] = df['date'].dt.year

    X = df[['month', 'year', 'quarter']]
    if label:
        y = df[label]
        return X, y
    return X

```

L'application de la fonction sur les données d'entraînement et test :

```

X_train, y_train = create_features(train, label='Engagement')
X_test, y_test = create_features(test, label='Engagement')

```

Le modèle et son ajustement :

```

reg = xgb.XGBRegressor(n_estimators=1000)

reg = reg.fit(X_train, y_train,
              eval_set=[(X_train, y_train), (X_test, y_test)],
              early_stopping_rounds=50,
              verbose=False)

```

Fb prophet

Le modèle avec Fb prophet

```

# Réinitialiser l'index et renommer les colonnes de l'ensemble de test
test_data = test_data.reset_index()
test_data.columns = ['ds', 'y']
# Ajouter une colonne 'ds' à train_data
train_data = train_data.reset_index()
train_data.columns = ['ds', 'y']

# Renommer la colonne 'Engagement CORP' en 'y'
train_data = train_data.rename(columns={'Engagement CORP': 'y'})

# Créer un objet Prophet et ajuster le modèle aux données d'entraînement
m = Prophet()
m.fit(train_data)

# Prévoir pour la période de test
future = m.make_future_dataframe(periods=len(test_data), freq='M', include_history=False)
forecast = m.predict(future)

# Afficher les prévisions pour la période de test
pred = forecast[['ds', 'yhat']]
print(pred)

```

Le graphe interactif de Fb prophet

```

from prophet.plot import plot_plotly, plot_components_plotly

plot_plotly(m, FRC)

```

Métriques :

```

# Calculate RMSE
rmse = np.sqrt(mean_squared_error(y_true, y_pred))
print('RMSE:', rmse)

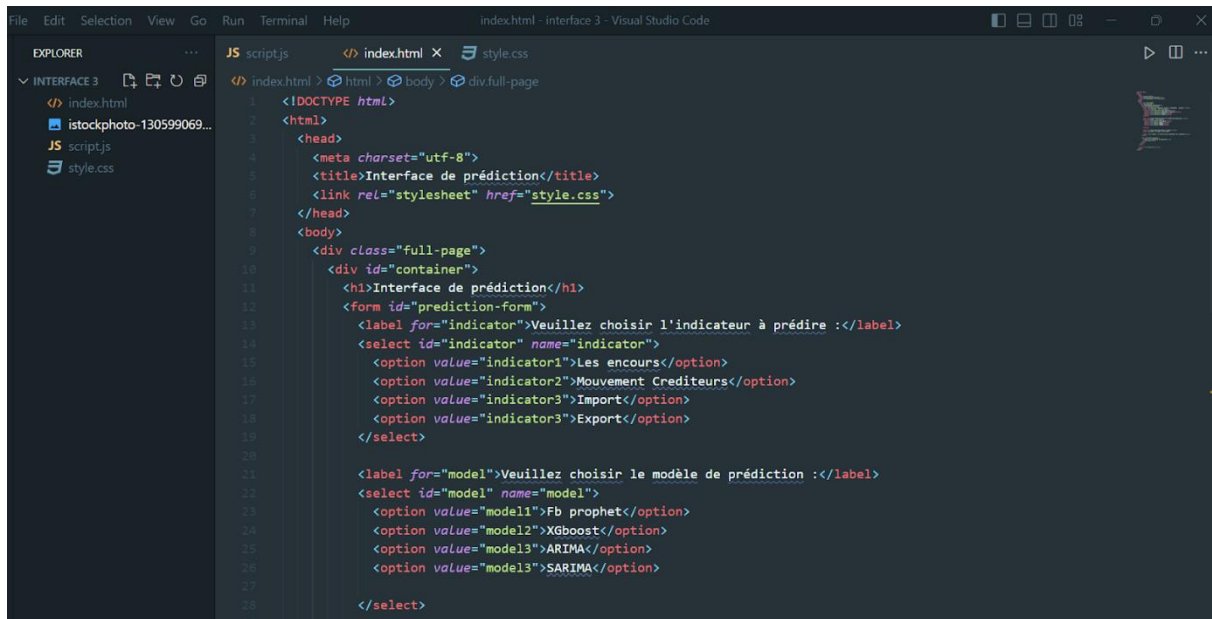
# Calculate WMAPE
wmape = np.sum(np.abs((y_true - y_pred))) / np.sum(np.abs(y_true)) * 100
print('WMAPE:', wmape)

# Calculate NRMSE
nrmse = rmse / (np.max(y_true) - np.min(y_true)) * 100
print('NRMSE:', nrmse)

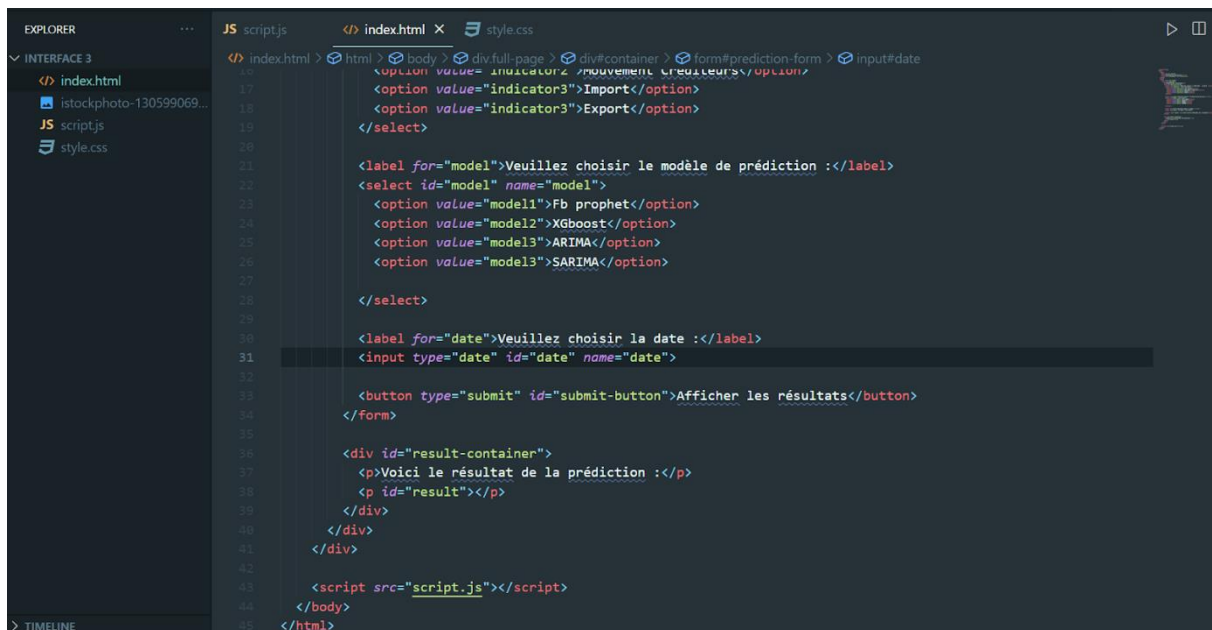
R_square = r2_score(test['Engagement'], test['Engagement_Prediction'])
print('Coefficient of Determination', R_square)

```

I.3 Interface



```
1 <!DOCTYPE html>
2 <html>
3   <head>
4     <meta charset="utf-8">
5     <title>Interface de prédiction</title>
6     <link rel="stylesheet" href="style.css">
7   </head>
8   <body>
9     <div class="full-page">
10      <div id="container">
11        <h1>Interface de prédiction</h1>
12        <form id="prediction-form">
13          <label for="indicator">Veuillez choisir l'indicateur à prédire :</label>
14          <select id="indicator" name="indicator">
15            <option value="indicator1">Les encours</option>
16            <option value="indicator2">Mouvement Credituers</option>
17            <option value="indicator3">Import</option>
18            <option value="indicator3">Export</option>
19          </select>
20
21          <label for="model">Veuillez choisir le modèle de prédiction :</label>
22          <select id="model" name="model">
23            <option value="model1">Fb prophet</option>
24            <option value="model2">XGboost</option>
25            <option value="model3">ARIMA</option>
26            <option value="model3">SARIMA</option>
27          </select>
28        </div>
29      </div>
30    </body>
31 </html>
```



```
16 <option value="indicator2">Mouvement Credituers</option>
17 <option value="indicator3">Import</option>
18 <option value="indicator3">Export</option>
19 </select>
20
21 <label for="model">Veuillez choisir le modèle de prédiction :</label>
22 <select id="model" name="model">
23   <option value="model1">Fb prophet</option>
24   <option value="model2">XGboost</option>
25   <option value="model3">ARIMA</option>
26   <option value="model3">SARIMA</option>
27 </select>
28
29 <label for="date">Veuillez choisir la date :</label>
30 <input type="date" id="date" name="date">
31
32 <button type="submit" id="submit-button">Afficher les résultats</button>
33 </form>
34
35 <div id="result-container">
36 <p>Voici le résultat de la prédiction :</p>
37 <p id="result"></p>
38 </div>
39 </div>
40 </div>
41 </div>
42
43 <script src="script.js"></script>
44 </body>
45 </html>
```



```
File Edit Selection View Go Run Terminal Help style.css - interface 3 - Visual Studio Code
EXPLORER
  INTERFACE 3
    index.html
    istockphoto-130599069...
    JS script.js
    style.css
  style.css > ...
  1 * {
  2   margin: 0;
  3   padding: 0;
  4   box-sizing: border-box;
  5 }
  6 .full-page {
  7   height: 100%;
  8   width: 100%;
  9   background-image: url('istockphoto-1305990690-612x612.jpg');
 10   background-position: center;
 11   background-size: cover;
 12   position: absolute;
 13 }
 14 #container {
 15   width: 500px;
 16   height: 550px;
 17   position: absolute;
 18   top: 50%;
 19   left: 50%;
 20   transform: translate(-50%, -50%);
 21   background-color: rgba(255, 255, 255, 0.32);
 22   border-radius: 20px;
 23   padding: 30px;
 24   box-shadow: 5px 5px 20px 5px rgba(222, 177, 149, 0.792);
 25   background-image: linear-gradient(to bottom, rgba(216, 213, 212, 0.37), rgba(228, 180, 180, 0.498));
```

```
File Edit Selection View Go Run Terminal Help style.css - interface 3 - Visual Studio Code
EXPLORER
  INTERFACE 3
    index.html
    istockphoto-130599069...
    JS script.js
    style.css
  style.css > ...
  29 }
 30
 31 #prediction-form {
 32   background-color: transparent;
 33   border: none;
 34 }
 35
 36 #prediction-form label,
 37 #prediction-form select,
 38 #prediction-form input[type="date"],
 39 #prediction-form button {
 40   display: block;
 41   margin: 10px 0;
 42   width: 100%;
 43   border: none;
 44   border-radius: 5px;
 45   padding: 10px;
 46   font-size: 18px;
 47 }
 48
 49 #container h1 {
 50   color: #C58940;
 51   text-align: center;
 52 }
 53
 54 #submit-button {
 55   background-color: #de9944c2;
 56   color: #fff;
 57   cursor: pointer;
```

```
from flask import Flask, request, jsonify

app = Flask(__name__)

@app.route('/predict', methods=['POST'])
def predict():
    # Récupérer les données du formulaire
    indicateur = request.form['indicateur']
    modele = request.form['modèle']
    date = request.form['date']
```

```

def generate_result_html(predictions):
    # Convertir les prévisions en HTML
    result_html = predictions.to_html(index=False)

    # Intégrer les prévisions dans le modèle HTML
    html = f"""
    <p>Voici le résultat de la prédiction :</p>
    <table>

```

```

import pandas as pd

def get_data(indicator, start_date, end_date):
    # Connexion à la base de données
    conn = sqlite3.connect('data.db')

    # Récupération des données pour l'indicateur choisi et la période spécifiée
    query = f"SELECT * FROM {indicateur} WHERE Date BETWEEN '{start_date}' AND '{end_date}'"
    df = pd.read_sql_query(query, conn, parse_dates=['Date'])

    # Fermeture de la connexion à la base de données
    conn.close()

    # Retourner les données sous forme de DataFrame pandas
    return df

```