

MÉMOIRE DE FIN D’ÉTUDE

En vue de l’obtention du Diplôme de Master II

Option : Exepert en Big Data

SYSTÈME INTELLIGENT DE

PRÉDICTION DU VOLUME DES VENTES

|  |  |
| --- | --- |
| *Réaliser par* | : FATIMA ZEHRAOUI |
| *N° de candidat* | : 140661 |
| *Encadrant* | : Anis BENAMMAR |
| *Tuteur* | : Mouley Driss AIT EL MOUDEN |



Année universitaire : 2023 / 2024

*Avant-propos*

Tout d’abord, je tiens à remercier M. Moulay Dris AIT ELMOUDEN, mon tuteur, pour sa disponibilité, son attention et sa confiance tout au long de mon apprentissage au sein de son entreprise Rapide Service Car 2.

Je remercie également toute l’équipe pédagogique de l’école ESPIC et les intervenants professionnels responsables de ma formation, pour avoir assuré la partie théorique de celle-ci.

Je tiens à témoigner toute ma reconnaissance aux personnes suivantes, pour leur aide dans la réalisation de ce rapport : M. Anis BENAMARA, ses conseils avisés m’ont permis de naviguer avec assurance à travers les complexités du modèle ARIMA, de l’adapter avec précision aux spécificités des données de vente du garage Rapide Service Car 2, et d’obtenir des résultats significatifs. Grâce à son expertise et sa disponibilité, j’ai pu développer une meilleure compréhension des enjeux cruciaux liés à la prévision des ventes. Ce travail m’a également permis de renforcer mes compétences en analyse de données, une compétence qui me sera précieuse tout au long de ma carrière.

Ce projet n’aurait pu être mené à bien sans ses encouragements, sa rigueur et sa passion pour le domaine. Pour tout cela, je lui adresse mes plus sincères remerciements. Je remercie aussi le Président du jury de bien vouloir lire ce rapport et m’évaluer.

Et enfin, je remercie mes parents, mes frères et sœurs, ainsi que mes amis pour leur soutien et leur confiance sans qui je n’aurais jamais pu connaître et vivre une telle expérience.

*Avant-propos*

Le présent rapport a pour objectif de fournir un aperçu détaillé de mon année d’apprentissage au sein de l’entreprise Rapide Service Car 2, dans le cadre de mon mastère en Big Data à l’École Supérieure de Pratiques Informatiques et Commerciales (ESPIC). Cette expérience s’inscrit dans un parcours académique et professionnel visant à approfondir mes connaissances et compétences dans le domaine de l’analyse de données massives.

Au cours de cette année, j’ai eu l’opportunité de m’immerger dans un environnement de travail dynamique et stimulant, propice à l’innovation et à l’expérimentation. Je m’efforcerai, dans ce rapport, de mettre en avant non seulement les missions qui m’ont été confiées, mais également les technologies et méthodologies que j’ai eu l’occasion d’appliquer. J’aborderai également les succès que j’ai pu rencontrer, ainsi que les contraintes et difficultés qui ont jalonné mon parcours, et qui ont contribué à mon apprentissage.

Mon intérêt prononcé pour l’exploitation et le traitement des données massives, ainsi que pour l’application d’études statistiques et la prédiction de données, m’a conduit à rechercher une entreprise dont le cœur de métier repose sur l’utilisation et le développement de techniques d’apprentissage automatique. L’objectif de mon alternance était de contribuer activement au développement de solutions qui permettent de résoudre des problématiques complexes et de donner un sens tangible aux données.

C’est dans cette optique que j’ai choisi d’effectuer ma seconde année d’alternance chez Rapide Service Car 2, une entreprise reconnue pour son rôle de créateur et fournisseur de données originales. La mission de cette entreprise, axée sur la production de données de haute qualité, s’aligne parfaitement avec mes aspirations professionnelles et mes objectifs d’apprentissage.

Dans les sections qui suivent, je présenterai en détail les différentes missions auxquelles j’ai participé, ainsi que les méthodes et outils utilisés pour mener à bien ces projets. Je mettrai également en lumière les défis techniques auxquels j’ai été confrontée tout au long de mon parcours, notamment en ce qui concerne l’analyse et la prédiction des indicateurs clés de performance (KPI) de l’entreprise, en m’appuyant sur ses données internes. Ce rapport vise ainsi à démontrer l’impact de mon travail au sein de Rapide Service Car 2, tout en offrant une réflexion sur les enseignements tirés de cette expérience enrichissante.

**Table des matiéres**

[Introduction générale 10](#_Toc177386204)

[Etat de l’art 12](#_Toc177386205)

[1.1 Introduction 13](#_Toc177386206)

[1.2 Présentation de l’entreprise Rapide Services Car II 13](#_Toc177386207)

[1.2.1 Historique et Contexte de l’Entreprise 13](#_Toc177386208)

[1.2.2 Les Services Proposés 13](#_Toc177386209)

[1.2.3 Les Ressources Humaines et Matérielles 14](#_Toc177386210)

[1.2.4 L’importance du CRM GAD Garage dans la gestion quotidienne 14](#_Toc177386211)

[1.2.5 Impact de la Digitalisation sur l’Entreprise 18](#_Toc177386212)

[Revue de la littérature 19](#_Toc177386213)

[2.1 Introduction 20](#_Toc177386214)

[2.2 Machine Learning pour des données dépendantes 20](#_Toc177386215)

[2.3 Généralités sur les Séries Temporelles 22](#_Toc177386216)

[2.4 Machine Learning pour les Séries Temporelles 23](#_Toc177386217)

[2.4.1 Modèles Linéaires 24](#_Toc177386218)

[2.4.2 Modèles non Linéaires 25](#_Toc177386219)

[2.5 Modèle ARIMA 25](#_Toc177386220)

[2.5.1 Partie autorégressive (AR) 25](#_Toc177386221)

[2.5.2 Partie d’intégration (I) 25](#_Toc177386222)

[2.5.3 Partie moyenne mobile (MA) 26](#_Toc177386223)

[2.6 Conclusion 27](#_Toc177386224)

[Modélisation et prédiction avec ARIMA 28](#_Toc177386225)

[3.1 Introduction 29](#_Toc177386226)

[3.2 Analyse des données 29](#_Toc177386227)

[3.3 Analyse exploratoire des données 29](#_Toc177386228)

[3.3.1 Description des données 29](#_Toc177386229)

[3.3.2 Statistiques descriptives 30](#_Toc177386230)

[3.3.3 Visualisation des données 31](#_Toc177386231)

[3.4 Construction du modèle ARIMA 32](#_Toc177386232)

[3.5 Analyse de la stationnarité 33](#_Toc177386233)

[3.5.1 Identification des paramètres 34](#_Toc177386234)

[3.5.2 Estimation des paramètres du modèle 36](#_Toc177386235)

[3.5.3 Validation du modèle 37](#_Toc177386236)

[3.6 Prédiction des ventes 38](#_Toc177386237)

[3.6.1 Évaluation de la performance des prédictions 40](#_Toc177386238)

[Présentation de l’application et ces fonctionnalités 42](#_Toc177386239)

[4.1 Introduction 43](#_Toc177386240)

[4.2 Environnement matériel et logiciel 43](#_Toc177386241)

[4.2.1 Python 43](#_Toc177386242)

[4.2.2 statsmodels 44](#_Toc177386243)

[4.2.3 NumPy 44](#_Toc177386244)

[4.2.4 Sklearn (Scikit-learn) 45](#_Toc177386245)

[4.2.5 Pandas 45](#_Toc177386246)

[4.2.6 Ploty 46](#_Toc177386247)

[4.2.7 Streamlit 46](#_Toc177386248)

[4.3 Environnement de développement 47](#_Toc177386249)

[4.3.1 Visual Studio Code 47](#_Toc177386250)

[4.3.2 Google Colab 47](#_Toc177386251)

[4.4 Présentation de l’interface et fonctionnalités 48](#_Toc177386252)

[4.4.1 Page de présentation 48](#_Toc177386253)

[4.4.2 Section analyse revenue 50](#_Toc177386254)

[4.4.3 Section anlayse de profit 51](#_Toc177386255)

[4.4.4 Section Prédiction 52](#_Toc177386256)

[4.4.5 Quelques problèmes rencontrés 53](#_Toc177386257)

[4.5 Conclusion 53](#_Toc177386258)

[Conclusion Générale 54](#_Toc177386259)

[Bibliographie 56](#_Toc177386260)

Table des figures

* 1. Fenétre principal du CRM GAD . . . . . . . . . . . . . . . . . . . . . 13
  2. Table de visualisation d’informations . . . . . . . . . . . . . . . . . . 14
  3. Page d’enregistrement d’un nouveau client . . . . . . . . . . . . . . . . 15
  4. Machine learning supervisé . . . . . . . . . . . . . . . . . . . . . . . . 18
  5. Série temporelle aléatoire avec tendance . . . . . . . . . . . . . . . . . 19
  6. Série temporelle aléatoire avec composante saisonnière . . . . . . . . . 20
  7. Série temporelle aléatoire avec valeurs aberrantes . . . . . . . . . . . . 20
  8. Prédiction sur une série temporelle avec le modéle ARIMA . . . . . . . 24
  9. Décomposition de la série temporelle des ventes . . . . . . . . . . . . . 28
  10. Fonction d’auto-corrélation (ACF) des ventes différenciées. . . . . . . . 31
  11. Fonction d’auto-corrélation partielle (PACF) des ventes différenciées. . 32
  12. Prédiction des ventes des pièces automobiles . . . . . . . . . . . . . . 36
  13. Python logo . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
  14. statsmodels logo . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
  15. NumPy logo . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
  16. Scikit-learn logo . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
  17. Pandas logo . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
  18. Ploty logo . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
  19. Streamlit logo . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44
  20. Visual Studio Code logo . . . . . . . . . . . . . . . . . . . . . . . . . 44
  21. Google Colab logo . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
  22. Écran d’accueil . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
  23. Menu principal et filtres . . . . . . . . . . . . . . . . . . . . . . . . . . 47
  24. Dashboard sur le revenue de l’entreprise . . . . . . . . . . . . . . . . . 48
  25. Dashboard sur le profit de l’entreprise . . . . . . . . . . . . . . . . . . 49
  26. Résumé sur le modèle . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
  27. ARIMA forecasting . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50

# Introduction générale

Dans le secteur de la réparation et de l’entretien automobile, la capacité à anticiper le volume de ventes sur une période donnée revêt une importance capitale. Une prévision efficace permet non seulement de gérer les stocks de pièces détachées de manière optimale, mais aussi d’améliorer la planification des interventions et de répondre de manière proactive à la demande des clients. Au sein de Rapide Services Car II, où j’ai eu l’opportunité d’effectuer mon alternance, la prévision des ventes constitue un enjeu fondamental pour accroître la performance et l’efficacité des opérations.

L’avènement des technologies de Big Data et de l’intelligence artificielle (IA) ouvre de nouvelles perspectives pour la mise en place de systèmes intelligents de prédiction, capables de fournir des estimations précises du volume d’activité futur. Ces systèmes tirent parti des données historiques internes, telles que les ventes passées, les types de réparations effectuées et les pièces utilisées, ainsi que des données externes comme les tendances du marché automobile, les périodes de vacances et les conditions climatiques, afin d’élaborer des modèles de prévision.

Pour Rapide Services Car II, l’implémentation d’un système intelligent de prédiction permettrait d’anticiper les pics d’activité, d’optimiser l’approvisionnement en pièces détachées et de rationaliser la gestion des ressources humaines. Ce projet a pour objectif de développer un tel système, offrant ainsi une visibilité claire sur les ventes futures et maximisant la rentabilité du garage, tout en garantissant la satisfaction des clients.

Ainsi, ce mémoire vise à concevoir et à mettre en place un système intelligent de prédiction du volume des ventes pour Rapide Services Car II. En s’appuyant sur les données collectées durant mon alternance, ce projet intégrera les technologies de Big Data et de machine learning afin de créer un outil prédictif performant, adapté aux spécificités du garage et à son environnement.

Dans le chapitre 1 on présentera les défis et les enjeux liés à la prévision des ventes dans le secteur automobile, ainsi que l’importance d’une telle démarche pour l’efficacité opérationnelle.

Dans le chapitre 2 on décrira les technologies de Big Data et les outils de machine learning employés pour développer le système de prédiction, notamment Python, ainsi que des bibliothèques telles que Pandas, NumPy, et Scikit-learn.

Introduction générale

Dans chapitre 3 on détaillera la méthodologie adoptée pour concevoir et implémenter le système prédictif, y compris la collecte des données, le traitement des données et la création des modèles de prévision.

Dans chapitre 4 on présentera l’interface utilisateur du système, en mettant en avant les différentes fonctionnalités, telles que l’analyse des revenus, la prédiction des ventes et la visualisation des données et on analysera les résultats obtenus grâce au système de prédiction, en discutant de l’impact sur la gestion des stocks et la planification des interventions.

Chapitre 1

# Etat de l’art

## 1.1 Introduction

Avant d’intégrer le mastère en Big Data à l’ÉSPIC, j’ai étudié à l’université Sorbonne Paris Nord, où j’ai développé une passion pour la programmation et l’algorithmique. Lors de ma première année d’alternance à la mutuelle Intégrance, j’ai travaillé comme data manager puis développeur Python sur un projet de migration de données.

J’ai ensuite rejoint Rapide Services Car 2 en tant qu’alternante Data Scientist. J’y ai développé des modèles de machine learning pour améliorer la qualité des données et participé à l’analyse géographique des données. Ce chapitre explorera les technologies et méthodes pertinentes dans le domaine du Big Data qui ont guidé mon apprentissage et mes projets.

## 1.2 Présentation de l’entreprise Rapide Services Car II

### 1.2.1 Historique et Contexte de l’Entreprise

Fondée en 2017, Rapide Services Car II est un acteur local majeur dans le domaine de l’automobile, situé au 105 rue Etienne Dolet, à Pierrefitte-sur-Seine. L’entreprise est née de la volonté de répondre aux besoins croissants de maintenance et de réparation des véhicules dans la région. Elle s’est rapidement imposée comme un partenaire de confiance pour les automobilistes, grâce à son service client axé sur la qualité et son expertise technique dans des domaines variés tels que la mécanique, l’électronique embarquée, et la carrosserie.

Au fil des années, Rapide Services Car II a su évoluer avec l’industrie automobile, notamment face à l’introduction de nouvelles technologies dans les véhicules (voitures hybrides, électriques) et les outils de gestion numérique. L’entreprise met un point d’honneur à proposer des services personnalisés tout en garantissant la satisfaction des clients à chaque étape, que ce soit pour un simple entretien ou l’achat d’un nouveau véhicule.

### 1.2.2 Les Services Proposés

Rapide Services Car II se distingue par la diversité des services qu’elle propose à ses clients :

— Entretien courant : Vidange, remplacement de filtres, contrôles périodiques, vérification des systèmes de freinage, de suspension, etc. Ces services permettent aux clients de maintenir leurs véhicules en bon état, tout en prolongeant leur durée de vie.

— Réparation mécanique : Le garage dispose d’équipements de pointe pour diagnostiquer et réparer les pannes mécaniques ou électroniques sur tous types de véhicules, qu’il s’agisse de voitures thermiques, hybrides ou électriques.

— Carrosserie et peinture : Un atelier spécialisé en carrosserie prend en charge la réparation des dégâts esthétiques ou structurels. Les services vont de la réparation de petites bosses à la remise en état complète après un accident, incluant la peinture.

— Vente de véhicules : Rapide Services Car II propose une large gamme de véhicules neufs et d’occasion. Les clients bénéficient de conseils personnalisés pour l’achat de voitures adaptées à leurs besoins, avec des options de financement.

— Location de véhicules : Afin de répondre aux besoins temporaires de mobilité de sa clientèle, l’entreprise propose un service de location de courte et longue durée.

— Diagnostic électronique : L’évolution des technologies embarquées dans les véhicules rend nécessaire l’utilisation d’outils de diagnostic avancés pour détecter les problèmes électroniques. L’atelier dispose de ces équipements pour intervenir sur des systèmes complexes tels que l’ABS, l’ESP, les capteurs ou les calculateurs.

### 1.2.3 Les Ressources Humaines et Matérielles

Rapide Services Car II emploie une équipe de trentaines collaborateurs qualifiés et expérimentés dans divers domaines : mécaniciens, techniciens en diagnostic, carrossiers, et conseillers clientèle. Tous ces employés suivent régulièrement des formations pour rester à jour avec les dernières avancées technologiques dans le secteur automobile.

L’entreprise dispose également d’un atelier moderne, équipé d’outils de diagnostic électronique, de ponts élévateurs, de machines de montage et d’équilibrage de pneus, ainsi que d’une cabine de peinture. Ces installations garantissent la qualité des réparations effectuées et permettent d’accueillir plusieurs véhicules simultanément, optimisant ainsi les délais d’intervention.

### 1.2.4 L’importance du CRM GAD Garage dans la gestion quotidienne

Le CRM interne GAD Garage est l’un des éléments centraux dans la gestion de l’entreprise. Cet outil de gestion de la relation client est spécifiquement conçu pour répondre aux besoins d’un garage automobile. GAD Garage permet de suivre chaque client, véhicule, opération, ou vente de manière centralisée et automatisée.

1. Gestion des Rendez-vous et du Suivi des Véhicules

Grâce à GAD Garage, la planification des rendez-vous est optimisée. Chaque intervention (révision, réparation, contrôle technique) est enregistrée dans le système, avec des rappels automatiques envoyés aux clients pour leurs prochaines visites. Ce suivi proactif permet de fidéliser la clientèle en assurant une gestion rigoureuse de l’entretien des véhicules.

Le CRM permet également d’attribuer des tâches spécifiques aux techniciens et de suivre l’état d’avancement de chaque réparation en temps réel, garantissant ainsi une gestion fluide des opérations.

1. Base de Données Clients et Historique Véhicule

GAD Garage centralise l’ensemble des informations clients dans une base de données sécurisée. Chaque client possède une fiche détaillée comprenant :

— Les informations de contact et les préférences.

— L’historique des interventions sur ses véhicules.

— Les devis, factures et contrats de vente ou de location.

Cette gestion fine de la relation client permet de mieux connaître les besoins de chaque utilisateur, et ainsi de personnaliser les offres et promotions. De plus, la traçabilité de l’historique des véhicules permet d’offrir un service plus précis, notamment lors de la revente de véhicules d’occasion.

1. Gestion des Ventes et du Stock de Véhicules

Pour le département des ventes, GAD Garage offre un suivi complet des véhicules en stock (neufs et d’occasion). Chaque véhicule est enregistré avec son état, ses caractéristiques, et son prix. Lorsqu’un client achète un véhicule, le système génère automatiquement les documents nécessaires (contrats, factures) et met à jour les informations de stock.

En parallèle, les conseillers commerciaux peuvent suivre leurs objectifs de vente, et générer des rapports pour analyser les performances du mois ou de l’année.

1. Statistiques et Analyse des Performances

Un autre atout majeur de GAD Garage est sa capacité à produire des rapports statistiques détaillés. Ces rapports permettent à la direction de suivre en temps réel les indicateurs de performance clés (KPI) tels que :

— Le nombre de réparations effectuées par mois.

— Le taux de satisfaction des clients.

— Le chiffre d’affaires généré par la vente de véhicules ou par les services de réparation.

Grâce à ces données, l’entreprise peut ajuster ses stratégies commerciales et opérationnelles, en se basant sur des analyses précises des tendances et des besoins du marché.

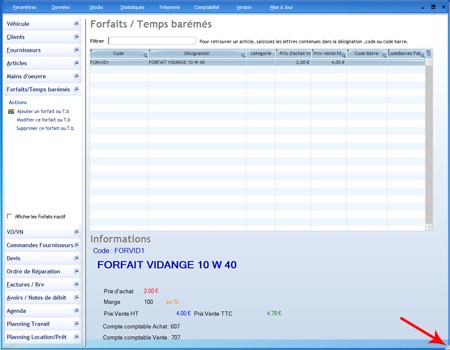


FIGURE 1.1 – Fenétre principal du CRM GAD

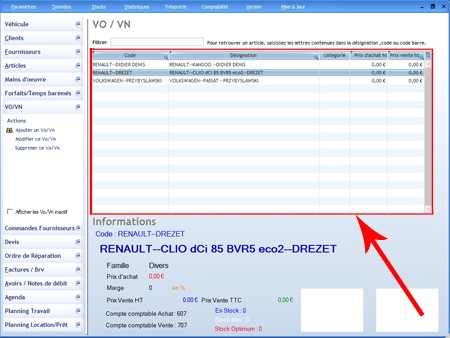


FIGURE 1.2 – Table de visualisation d’informations

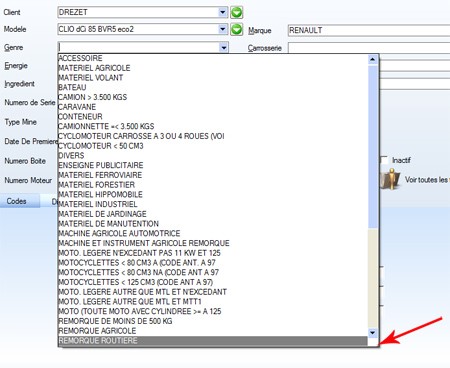


FIGURE 1.3 – Page d’enregistrement d’un nouveau client

### 1.2.5 Impact de la Digitalisation sur l’Entreprise

L’intégration du CRM GAD Garage représente une étape clé dans la transformation numérique de Rapide Services Car II. En automatisant et en centralisant une grande partie des processus de gestion, l’entreprise a gagné en efficacité et en réactivité. De plus, la digitalisation permet d’offrir une meilleure expérience aux clients, qui reçoivent des informations plus rapidement et bénéficient d’un service plus personnalisé.

Chapitre 2

# Revue de la littérature

## 2.1 Introduction

La prédiction des ventes est un domaine crucial pour les entreprises, car elle permet de mieux planifier les ressources, d’optimiser la gestion des stocks et de maximiser les profits. Diverses méthodes et modèles ont été développés au fil des années pour améliorer l’exactitude des prévisions. Dans ce contexte, les modèles de séries temporelles, et en particulier le modèle *ARIMA* (*Auto-Regressive Integrated Moving Average*), ont montré leur efficacité dans la prévision des ventes dans différents secteurs.

Ce chapitre propose une revue de la littérature sur les méthodes de prédiction des ventes, en mettant l’accent sur les séries temporelles et le modèle *ARIMA*. Nous commencerons par examiner les études antérieures qui ont exploré la prédiction des ventes, suivies d’une discussion sur les techniques couramment utilisées pour analyser les séries temporelles. Ensuite, nous aborderons l’application spécifique des modèles *ARIMA* dans divers domaines, en soulignant leurs avantages et limitations. Cette revue de la littérature fournira un cadre théorique solide pour justifier l’utilisation du modèle

*ARIMA* dans notre étude sur les ventes de pièces automobiles au garage Service Car Rapide 1.

## 2.2 Machine Learning pour des données dépendantes

L’apprentissage automatique, ou Machine Learning (ML) est un ensemble de techniques et d’algorithmes qui permettent de faire apprendre à un ordinateur à réaliser une certaine tâche. Ces algorithmes sont parvenus à de très bons résultats pour de nombreux domaines d’applications. [1]

Théoriquement, on considère deux ensembles X et Y. On suppose que l’on observe un jeu de *n* points de données D*n* = (*x*1*,y*1)*,...,*(*xn,yn*). Pour chaque point de donnée *i*, *xi* ∈ X est un vecteur de caractéristique, et *yi* ∈ Y est une étiquette. On en parle plus souvent en utilisant les termes anglais de "feature" et "label". On suppose que ces données ont été engendrées indépendamment à partir d’un certain couple de variables aléatoires (*X,Y* ) de loi jointe *P*, autrement dit, on suppose qu’il existe une certaine relation constante entre features et labels.

L’objectif central du Machine Learning est d’entraîner une certaine fonction *g* : X → Y tel que *g*(*X*) soit un bon estimateur de *Y* . La fonction *g* est appelée fonction d’apprentissage (ou modèle). Pour estimer la qualité d’une fonction d’apprentissage, on introduit une certaine perte *L* : Y × Y → R+ qui évalue la qualité d’un tel estimateur *g* en définissant une erreur théorique associée à cet estimateur :

*L*(*g*) = E[*L*(*g*(*X*)*,Y* )] (2.1)

Lorsque l’on s’intéresse au Machine Learning d’un point de vue théorique, on introduit en général deux notions. La première est la notion d’estimateur de Bayes *g*, qui est le meilleur estimateur possible parmi toutes les fonctions mesurables de X vers Y, c’est à dire que *gg*=*L*(*g*).

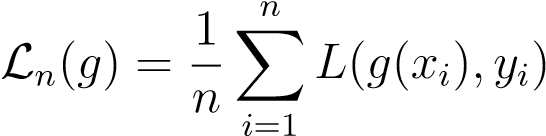
La seconde notion est celle de l’ensemble de modèle S parmi lequel on cherche notre estimateur. On définit donc également le meilleur modèle théorique *g*∗S = *g* ∈ S*L*(*g*) dans cet ensemble S.

Un algorithme de Machine Learning est un choix d’un ensemble S, et d’une méthode pour choisir un certain estimateur *g*ˆ ∈ S. Pour évaluer la qualité de l’algorithme, il faut contrôler l’excès de risque : *L*(*g*ˆ) − *L*(*g*∗). Contrairement à ce qui se passe en statistique classique, on ne cherche donc pas à savoir si l’on parvient à trouver le "bon" modèle, ou un modèle "proche" du bon modèle, mais si l’on trouve un modèle dont l’erreur n’est pas trop loin de celle du meilleur modèle. On décompose généralement cet excès de risque en introduisant :

D’une part l’erreur d’approximation *L*(*g*S)−*L*(*g*), qui diminue lorsque l’on augmente l’ensemble de modèle S, D’autre part l’erreur d’estimation, qui augmente en général avec la taille l’ensemble de modèle S.

Il existe de nombreux résultats pour borner l’excès de risque, mais ils dépendent de la taille de l’espace des hypothèses *S* et de la manière de choisir la fonction *g*ˆ qui minimise le risque empirique sur les données disponibles.

La manière classique de choisir *g*ˆ est de le considérer comme le minimiseur du risque empirique L*n* défini par la formule :

 (2.2)

Dans ce cadre, *g*ˆ =*g*∈*S* L*n*(*g*). On dit alors que *g*ˆ est un minimiseur du risque empirique et on dispose d’un grand nombre de résultats permettant de contrôler l’excès de risque L(*g*ˆ) − L(*g*∗).

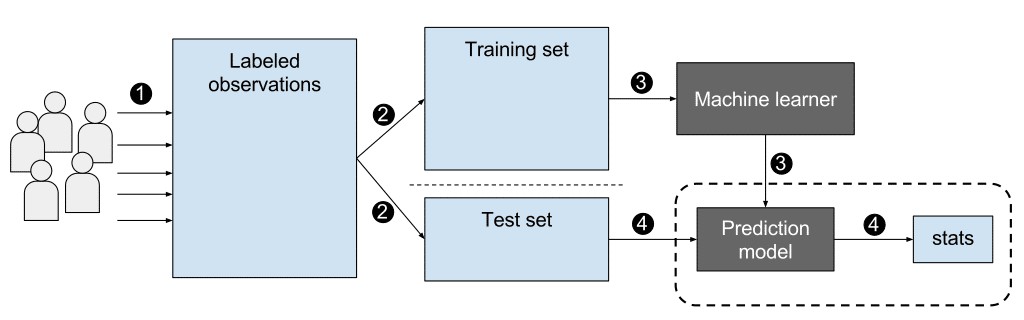


FIGURE 2.1 – Machine learning supervisé

## 2.3 Généralités sur les Séries Temporelles

Une série temporelle est une suite de données ordonnées dans le temps. Formellement, une série temporelle  est une suite de *T* observations d’une variable aléatoire indexée par le temps *t*. [2]

Les séries temporelles présentent des propriétés statistiques particulières, comme :

— La dépendance temporelle : les observations successives sont généralement corrélées.

— La non-stationnarité : les propriétés statistiques (moyenne, variance, autocorrélation) peuvent varier dans le temps.

— La saisonnalité : des motifs périodiques peuvent apparaître dans la série.

Ces caractéristiques spécifiques aux séries temporelles doivent être prises en compte lors de leur analyse et de leur modélisation.

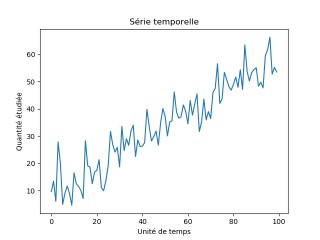


FIGURE 2.2 – Série temporelle aléatoire avec tendance

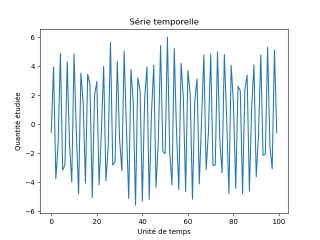


FIGURE 2.3 – Série temporelle aléatoire avec composante saisonnière

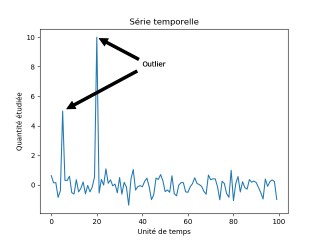


FIGURE 2.4 – Série temporelle aléatoire avec valeurs aberrantes

## 2.4 Machine Learning pour les Séries Temporelles

Mathématiquement, une série temporelle est une suite de variables aléatoires (*Xt*) à valeurs dans un certain espace X . On dispose d’une certaine réalisation de ces variables aléatoires (*x*1*,...,xn*), et on cherche à prédire les valeurs les plus probables pour les suivantes.

Plus formellement, le problème de prédiction de séries temporelles consiste à trouver des valeurs (*xn*+1*,...,xn*+*h*) qui minimisent l’espérance de la fonction de perte *L* :

*h*

min*h* XE[*L*(*xn*+*i,Xn*+*i*)] (2.3)

{*xn*+*i*}

*i*=1 *i*=1

Où :

— *h* est l’horizon de prédiction, c’est-à-dire le nombre de valeurs "en avance" que l’on cherche à prédire.

— *L* est une fonction de perte à valeurs réelles positives telle que *L*(*x,y*) mesure l’erreur commise lorsque l’on prédit *x* à la place de *y*.

Pour pouvoir prédire le futur d’une série temporelle, on s’appuie généralement sur son passé. Cela signifie que *Xt*+*h* s’exprime à l’aide d’une certaine fonction *fh*, dite fonction d’apprentissage (ou modèle), appliquée aux *a* dernières valeurs connues :

*Xt*+*h* = *fh*(*Xt,...,Xt*−*a*) (2.4)

Le choix de la fonction *fh* et de l’horizon *h* dépend de la nature de la série temporelle et de l’application visée.

Les méthodes d’apprentissage automatique peuvent être particulièrement utiles pour traiter les séries temporelles, notamment pour des tâches de prévision, de classification ou d’extraction de caractéristiques.

### 2.4.1 Modèles Linéaires

Une approche classique consiste à utiliser des modèles linéaires tels que les modèles ARIMA (Autoregressive Integrated Moving Average). Ces modèles capturent la structure de dépendance temporelle de la série sous forme d’un processus autorégressif et de moyennes mobiles.

Le modèle ARIMA(*p,d,q*) s’écrit sous la forme :

*p q*

(1 − X*ϕiLi*)(1 − *L*)*dyt* = (1 + X*θjLj*)*ϵt* (2.5)

*i*=1 *j*=1

Où *L* est l’opérateur retard, *ϕi* les paramètres autorégressifs, *θj* les paramètres de moyennes mobiles et *ϵt* un bruit blanc.

### 2.4.2 Modèles non Linéaires

Lorsque les séries temporelles présentent des comportements non linéaires, les modèles linéaires peuvent s’avérer insuffisants. On peut alors se tourner vers des modèles non linéaires tels que :

— Les réseaux de neurones récurrents (RNN), comme les LSTM, qui peuvent capturer des dépendances temporelles complexes.

— Les modèles à noyaux, comme les processus gaussiens, qui permettent de modéliser des non-linéarités.

— Les modèles à régimes switching, comme les modèles de Markov switching, qui peuvent s’adapter à des changements de régimes dans la série.

## 2.5 Modèle ARIMA

Le modèle ARIMA (Autoregressive Integrated Moving Average) est une classe de modèles de séries temporelles linéaires très utilisée pour analyser et prédire des données temporelles. Ce modèle combine trois composantes principales : la partie autorégressive (AR), la partie d’intégration (I) et la partie moyenne mobile (MA).[3]

### 2.5.1 Partie autorégressive (AR)

La partie autorégressive du modèle ARIMA capture la dépendance linéaire entre les valeurs actuelles et les valeurs passées de la série temporelle. Le modèle AR(p) s’écrit sous la forme :

*Xt* = *c* + *ϕ*1*Xt*−1 + *ϕ*2*Xt*−2 + ··· + *ϕpXt*−*p* + *εt* (2.6)

Où

*ϕ*1*,ϕ*2*,...,ϕp* sont les paramètres autorégressifs et *εt* est un bruit blanc.

### 2.5.2 Partie d’intégration (I)

La partie d’intégration du modèle ARIMA capture les tendances non stationnaires dans la série temporelle. Le modèle I(d) prend en compte les différences d’ordre *d* nécessaires pour rendre la série stationnaire, selon l’équation :

(1 − *B*)*dXt* = *Yt* (2.7)

Où : *B* est l’opérateur de retard et *Yt* est une série stationnaire.

### 2.5.3 Partie moyenne mobile (MA)

La partie moyenne mobile du modèle ARIMA capture la dépendance linéaire entre les valeurs actuelles et les erreurs passées de la série temporelle. Le modèle MA(q) s’écrit sous la forme :

*Xt* = *µ* + *εt* + *θ*1*εt*−1 + *θ*2*εt*−2 + ··· + *θqεt*−*q* (2.8)

Où

*θ*1*,θ*2*,...,θq* sont les paramètres de la moyenne mobile et *εt* est un bruit blanc.

Le modèle ARIMA général, noté ARIMA(p,d,q), combine les trois composantes précédentes et s’écrit sous la forme :

*ϕ*(*B*)(1 − *B*)*dXt* = *θ*(*B*)*εt* (2.9)

Où

*ϕ*(*B*) = 1−*ϕ*1*B* −*ϕ*2*B*2 −···−*ϕpBp* et *θ*(*B*) = 1+*θ*1*B* +*θ*2*B*2 +···+*θqBq*.

Le modèle ARIMA est très flexible et peut s’adapter à une large gamme de comportements des séries temporelles. Son utilisation nécessite cependant une analyse approfondie des données pour identifier les paramètres appropriés (p, d, q). [4]

Les paramètres p, d et q du modèle ARIMA(p,d,q) ont les significations suivantes :

— p : ordre de la partie autorégressive (AR). Ce paramètre indique le nombre de valeurs passées à prendre en compte pour prédire la valeur actuelle.

— d : ordre de différenciation (I). Ce paramètre indique le nombre de différences nécessaires pour rendre la série temporelle stationnaire.

— q : ordre de la partie moyenne mobile (MA). Ce paramètre indique le nombre d’erreurs passées à prendre en compte pour prédire la valeur actuelle.

Le choix des valeurs appropriées pour p, d et q est une étape cruciale dans la modélisation ARIMA. Ces paramètres doivent être déterminés en fonction des caractéristiques de la série temporelle étudiée.

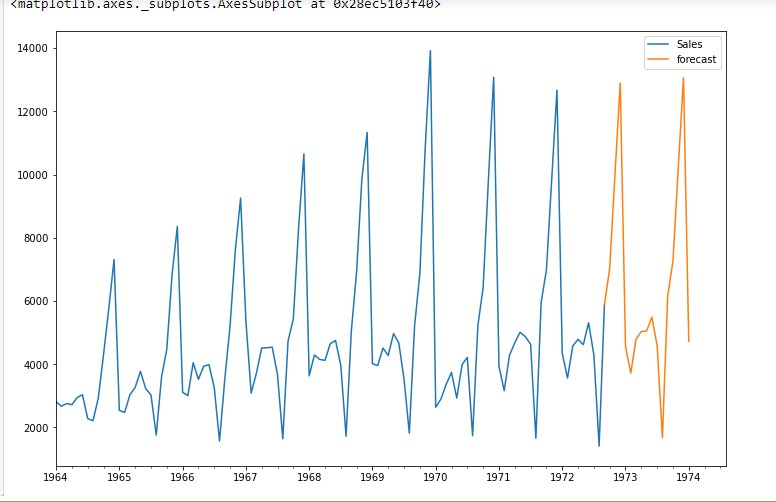


FIGURE 2.5 – Prédiction sur une série temporelle avec le modèle ARIMA

## 2.6 Conclusion

En conclusion pour ce chapitre, l’intégration du modèle ARIMA dans les approches de machine learning apporte une valeur ajoutée substantielle pour l’analyse et la prévision des séries temporelles. La combinaison de la puissance de modélisation du modèle ARIMA et des capacités d’apprentissage automatique offre de nouvelles perspectives passionnantes pour relever les défis liés aux données temporelles dans de nombreux domaines d’application.

Chapitre 3

# Modélisation et prédiction avec ARIMA

## 3.1 Introduction

La modélisation des séries temporelles est une technique essentielle pour analyser et prévoir l’évolution des données séquentielles. Parmi les différents modèles disponibles, le modèle *ARIMA* (*Auto-Regressive Integrated Moving Average*) est particulièrement prisé pour sa flexibilité et sa capacité à capturer divers patterns dans les données. Dans ce chapitre, nous nous concentrerons sur l’application du modèle *ARIMA* pour prédire les ventes de pièces automobiles au garage Service Car Rapide 1.

Nous commencerons par une analyse exploratoire des données historiques de ventes pour identifier les tendances, les saisonnalités et les anomalies éventuelles. Ensuite, nous procéderons à la construction du modèle *ARIMA* en déterminant les paramètres adéquats (p, d, q), suivie de l’estimation et de la validation du modèle. Enfin, nous utiliserons le modèle développé pour prédire les ventes futures et évaluerons la performance des prédictions obtenues.

Ce chapitre vise à démontrer l’efficacité du modèle *ARIMA* dans la prévision des ventes et à fournir des insights pratiques pour améliorer la gestion des stocks au garage Service Car Rapide 1.

## 3.2 Analyse des données

## 3.3 Analyse exploratoire des données

L’analyse exploratoire des données (EDA) est une étape cruciale dans la compréhension des caractéristiques et des structures sous-jacentes des séries temporelles. Elle permet d’identifier les tendances, les saisons, et les anomalies qui pourraient influencer la modélisation et la prévision des ventes. Dans cette section, nous effectuons une EDA sur les données historiques des ventes de pièces automobiles du garage Service Car Rapide 1.

### 3.3.1 Description des données

Les données utilisées pour cette étude proviennent des archives de ventes du garage Service Car Rapide 1. Elles couvrent une période entre mars 2018 et juin 2024 et incluent les variables suivantes :

— Date : La date de la transaction.

— Produit : Le type de pièce automobile vendue.

— Quantité vendue : Le nombre de pièces vendues par transaction.

— Prix unitaire : Le prix de vente unitaire de chaque pièce.

— Region : La région d’où le client est venu.

— Feedback : La note du client par rapport au service.

— Montant de gain : Le montant gagner par rapport au prix d’achat de la piéce.

### 3.3.2 Statistiques descriptives

Le tableau 3.1 présente les principales statistiques descriptives pour deux variables clés de nos données sur les ventes de pièces automobiles de mars 2018 à juin 2024.

|  |  |  |
| --- | --- | --- |
| Statistiques | Total TTC (C) | Quantité vendue (piéce) |
| Moyenne | 20665.390526 | 135.118421 |
| Médiane | 18430.045000 | 102.000000 |
| Écart-type | 8440.860071 | 147.581658 |
| Minimum | 3932.910000 | 33.000000 |
| Maximum | 45980.540000 | 1151.000000 |

TABLE 3.1 – Statistiques descriptives des ventes de pièces automobiles de mars 2018 à juin 2024.

— Moyenne :

— *Total TTC* : La moyenne des ventes s’élève à 20,665.39 C, indiquant un chiffre d’affaires relativement élevé.

— *Quantité vendue* : La moyenne de 135.12 pièces vendues montre une performance stable dans les ventes.

— Médiane :

— La médiane des ventes (18,430.05 C) est inférieure à la moyenne, ce qui peut indiquer une distribution asymétrique, probablement influencée par quelques ventes exceptionnellement élevées.

— Pour la quantité, la médiane de 102 pièces suggère que la majorité des transactions se situent autour de ce nombre.

— Écart-type :

— Un écart-type de 8,440.86 C pour le Total TTC indique une variabilité significative dans les ventes, signifiant que certaines ventes sont bien au-dessus ou en dessous de la moyenne.

— Pour la quantité, un écart-type de 147.58 pièces indique également une grande variation dans le nombre de pièces vendues par transaction.

— Minimum et Maximum :

— Le minimum pour le Total TTC (3,932.91 C) et la quantité (33 pièces) suggère que certaines ventes sont relativement faibles.

— Le maximum, avec 45,980.54 C et 1,151 pièces, montre qu’il y a des transactions très importantes, ce qui pourrait être dû à des ventes en gros ou des contrats importants.

### 3.3.3 Visualisation des données

La visualisation des données est une étape essentielle pour identifier les patterns temporels. Les graphiques suivants illustrent l’évolution des ventes au fil du temps.

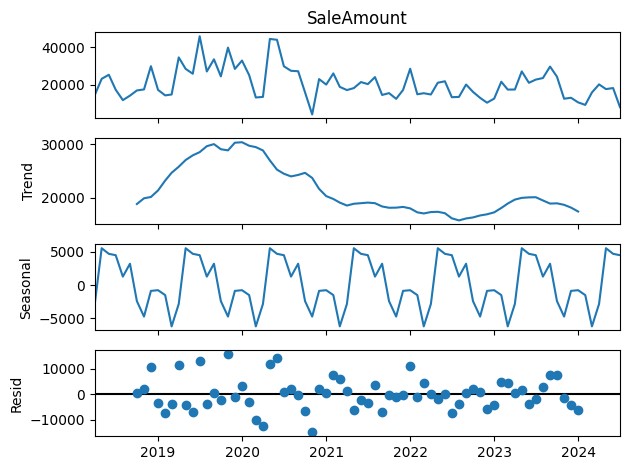


FIGURE 3.1 – Décomposition de la série temporelle des ventes

#### 1. Composante de Tendance

La composante de tendance montre une évolution générale des ventes au fil du temps. On observe que :

— Les ventes ont tendance à augmenter jusqu’à un certain point, puis montrent une légère diminution.

— Cette tendance pourrait être influencée par des facteurs externes tels que les changements économiques, les promotions ou les fluctuations du marché.

#### 2. Composante Saisonnière

La composante saisonnière indique les variations récurrentes des ventes qui se produisent à intervalles réguliers :

— On remarque des pics réguliers dans les ventes qui semblent correspondre à des périodes spécifiques chaque année.

— Ces variations peuvent être dues à des saisons de vente, comme les fêtes ou les promotions annuelles, qui influencent le comportement d’achat des consommateurs.

#### 3. Composante des Résidus

La composante des résidus représente les variations aléatoires des ventes qui ne peuvent pas être expliquées par la tendance ou la saisonnalité :

— Les résidus semblent être distribués de manière aléatoire autour de zéro, ce qui est un indicateur positif pour un modèle de prévision.

— Une absence de structure dans les résidus suggère que le modèle de décomposition est adéquat pour capturer les dynamiques de la série temporelle.

## 3.4 Construction du modèle ARIMA

La modélisation ARIMA (*Auto-Regressive Integrated Moving Average*) est une méthode statistique puissante utilisée pour analyser et prévoir les séries temporelles. Elle combine trois composants : l’auto-régression (AR), la différenciation intégrée (I), et la moyenne mobile (MA). Dans cette section, nous détaillons les étapes suivies pour construire le modèle ARIMA adapté aux ventes de pièces automobiles du garage Service Car Rapide 1.

## 3.5 Analyse de la stationnarité

L’analyse de la stationnarité est une étape cruciale dans l’étude des séries temporelles. Une série stationnaire se caractérise par une moyenne et une variance constantes dans le temps, ainsi qu’une autocorrélation qui décroît rapidement lorsque le décalage augmente. À l’inverse, une série non stationnaire présente des variations de sa moyenne et/ou de sa variance au cours du temps.

Pour tester la stationnarité de la série, nous avons utilisé le test de Dickey-Fuller augmenté (ADF). Ce test permet de vérifier la présence d’une racine unitaire dans la série, ce qui indiquerait une non-stationnarité. [5]

Les résultats du test ADF sont les suivants :

Résultats du test de Dickey-Fuller augmenté :

|  |  |
| --- | --- |
| Statistique de test | -5.079770601791765 |
| p-value | 1.5393235766400614e-05 |
| Valeurs critiques :  1% | -3.520713130074074 |
| 5% | -2.9009249540740742 |
| 10% | -2.5877813777777776 |
| Interprétation | La série temporelle est stationnaire |

TABLE 3.2 – Résultats du test de stationnarité

— Statistique de Test :

— La statistique de test est de -5.079770601791765, ce qui est comparé aux valeurs critiques pour évaluer la stationnarité.

— p-value :

— La p-value est très faible (1.5393235766400614e-05), bien inférieure à 0.05, indiquant une forte preuve contre l’hypothèse nulle.

— Valeurs Critiques :

— À 1% : -3.520713130074074

— À 5% : -2.900924954074742

— À 10% : -2.587871377777777

— La statistique de test (-5.079770601791765) est inférieure à toutes ces valeurs critiques.

La série temporelle est stationnaire, ce qui signifie que ses propriétés statistiques ne changent pas au fil du temps.

### 3.5.1 Identification des paramètres

La première étape dans la construction d’un modèle ARIMA consiste à identifier les paramètres appropriés, notés *p*, *d* et *q*. Ces paramètres représentent respectivement :

— *p* : Le nombre de termes auto-régressifs.

— *d* : Le nombre de différenciations nécessaires pour rendre la série stationnaire.

— *q* : Le nombre de termes de moyenne mobile.

Pour déterminer ces paramètres, nous utilisons l’auto-corrélation partielle (PACF) et les fonctions d’auto-corrélation (ACF).

Auto-corrélation et auto-corrélation partielle

Les graphes ACF et PACF sont utilisés pour identifier les valeurs potentielles de *p* et *q*. Les figures 3.2 et 3.3 montrent respectivement les graphes ACF et PACF des ventes différenciées.

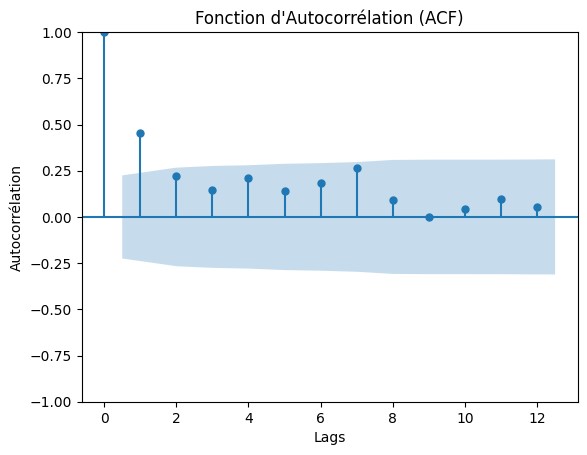


FIGURE 3.2 – Fonction d’auto-corrélation (ACF) des ventes différenciées.

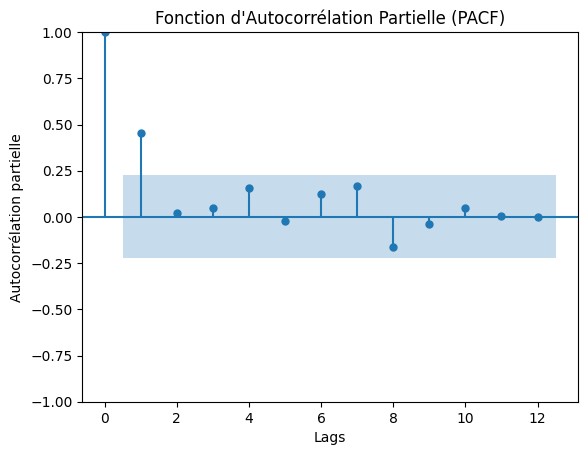


FIGURE 3.3 – Fonction d’auto-corrélation partielle (PACF) des ventes différenciées.

#### Analyse de la Fonction d’Autocorrélation (ACF)

La fonction d’autocorrélation (ACF) mesure la corrélation entre une série temporelle et ses propres valeurs décalées (lags). Voici quelques points clés :[6]

— Au lag 0, l’autocorrélation est toujours égale à 1, ce qui est attendu.

— Les valeurs de l’ACF diminuent progressivement pour les lags suivants, ce qui indique une dépendance décroissante dans le temps.

— Les premiers lags montrent des autocorrélations significatives, suggérant qu’il existe des relations temporelles dans les données.

— À partir de 6 lags, les corrélations deviennent non significatives, ce qui peut indiquer que l’effet des observations passées s’estompe rapidement.

#### Analyse de la Fonction d’Autocorrélation Partielle (PACF)

La fonction d’autocorrélation partielle (PACF) élimine l’effet des lags intermédiaires.

Voici les observations clés :

— La PACF présente une forte autocorrélation au lag 1, suivi d’une diminution rapide.

— Après le lag 1, les valeurs de la PACF deviennent non significatives, ce qui suggère qu’il n’y a pas d’influence significative des lags supérieurs sur les valeurs actuelles.

— Cela peut indiquer que le modèle autorégressif (AR) pourrait être approprié avec un ordre de 1.

L’analyse des fonctions ACF et PACF fournit des informations précieuses pour le choix des modèles de séries temporelles. La présence d’une autocorrélation significative aux premiers lags et l’absence d’influence significative des lags ultérieurs indiquent que les modèles ARIMA, particulièrement ceux avec un ordre d’autorégression de 1, pourraient être efficaces pour modéliser ces données.

### 3.5.2 Estimation des paramètres du modèle

Une fois les paramètres identifiés, nous procédons à l’estimation des coefficients du modèle ARIMA. Cette étape utilise la méthode des moindres carrés pour ajuster le modèle aux données.

Le modèle ARIMA(*p,d,q*) sélectionné pour les ventes de pièces automobiles est le suivant : ARIMA(0,1,2). Les résultats de l’estimation des paramètres sont présentés dans le tableau 3.3.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | coef | std err | z | P>|z| | [0.025 | 0.975] |
| ma.L1 | -0.5098 | 0.119 | -4.284 | 0.000 | -0.743 | -0.277 |
| ma.L2 | -0.2528 | 0.109 | -2.323 | 0.020 | -0.466 | -0.040 |
| sigma2 | 6.142e+07 | 3.49e-10 | 1.76e+17 | 0.000 | 6.14e+07 | 6.14e+07 |

TABLE 3.3 – Estimation des paramètres du modèle ARIMA

#### Interprétation des Résultats

1. \*\*Paramètres de Moyenne Mobile (MA)\*\* :

— ma.L1 : Le coefficient est de -0.5098, avec une erreur standard de 0.119. La statistique *z* est -4.284, et la p-value est 0.000. Cela indique que le terme de moyenne mobile de lag 1 est significatif et négativement associé à la série.

— ma.L2 : Le coefficient est de -0.2528, avec une erreur standard de 0.109. La statistique *z* est -2.323, et la p-value est 0.020. Cela montre également que le terme de moyenne mobile de lag 2 est significatif et négativement associé à la série.

1. \*\*Variance des Résidus (*σ*2)\*\* :

— Le coefficient *σ*2 est de 6.142e+07, avec une erreur standard très élevée (3.49e+17). La statistique *z* est 1.76, et la p-value est 0.000. Cela indique que la variance des résidus est significativement différente de zéro.

1. \*\*Intervalles de Confiance\*\* :

— Les intervalles de confiance à 95% pour ma.L1 (-0.743, -0.277) et ma.L2 (-0.466, -0.040) ne contiennent pas zéro, ce qui confirme la significativité des coefficients respectifs.

— L’intervalle de confiance pour *σ*2 est très large, ce qui pourrait être dû à la grande variabilité des données.

Les résultats montrent que les termes de moyenne mobile ma.L1 et ma.L2 sont significatifs et négativement associés à la série temporelle, ce qui peut indiquer un effet de lissage sur les valeurs précédentes. La variance des résidus est également significativement positive, ce qui est essentiel pour les prévisions. Ces résultats suggèrent que le modèle ARIMA est approprié pour modéliser la série temporelle étudiée.

### 3.5.3 Validation du modèle

La validation du modèle est essentielle pour s’assurer de sa pertinence et de sa robustesse. Nous utilisons plusieurs méthodes de validation, dont l’analyse des résidus et les tests statistiques.

— Test de Ljung-Box (Q) :

— Valeur Q : 0.01

— Probabilité (Prob(Q)) : 0.94

— Interprétation : Une valeur de Prob(Q) élevée (supérieure à 0.05) indique que les résidus ne présentent pas d’autocorrélation significative. Cela suggère que le modèle est adéquat pour capturer la dynamique des données.

— Test de Jarque-Bera (JB) :

— Valeur JB : 8.96

— Probabilité (Prob(JB)) : 0.01

— Interprétation : La faible Prob(JB) (inférieure à 0.05) indique que les résidus ne suivent pas une distribution normale. Cela pourrait suggérer la présence d’asymétrie ou d’excès de kurtosis dans les résidus.

— Hétéroscédasticité (H) :

— Valeur H : 0.26

— Probabilité (Prob(H) (two-sided)) : 0.00

— Interprétation : Une valeur de Prob(H) très faible indique la présence d’hétéroscédasticité dans les résidus, ce qui signifie que la variance des résidus n’est pas constante. Cela peut influencer la précision des estimations des paramètres du modèle.

— Asymétrie et Kurtosis :

— Asymétrie : 0.67

— Kurtosis : 4.04

— Interprétation : Une asymétrie positive (0.67) indique que la distribution des résidus est décalée vers la droite. Une kurtosis supérieure à 3 (4.04) suggère que la distribution a des queues plus lourdes que la normale, ce qui indique un excès de concentration des valeurs extrêmes.

Les résultats indiquent que, bien que le modèle semble adéquat en termes d’autocorrélation des résidus (test de Ljung-Box), il existe des préoccupations quant à la normalité et à l’hétéroscédasticité des résidus. Ces éléments doivent être pris en compte lors de l’interprétation des résultats et pourraient nécessiter des ajustements ou des transformations supplémentaires pour améliorer la spécification du modèle.

## 3.6 Prédiction des ventes

La prévision des ventes futures est une étape cruciale pour la gestion des stocks au sein du garage *Service Car Rapide 1*. L’objectif est d’anticiper la demande en pièces automobiles afin de minimiser les coûts liés aux ruptures de stock ou aux excédents. Pour ce faire, nous avons utilisé le modèle ARIMA, qui a été ajusté aux données historiques des ventes.

Le graphique ci-dessous présente les résultats d’un modèle ARIMA appliqué à des données historiques, avec des prévisions pour les années à venir.

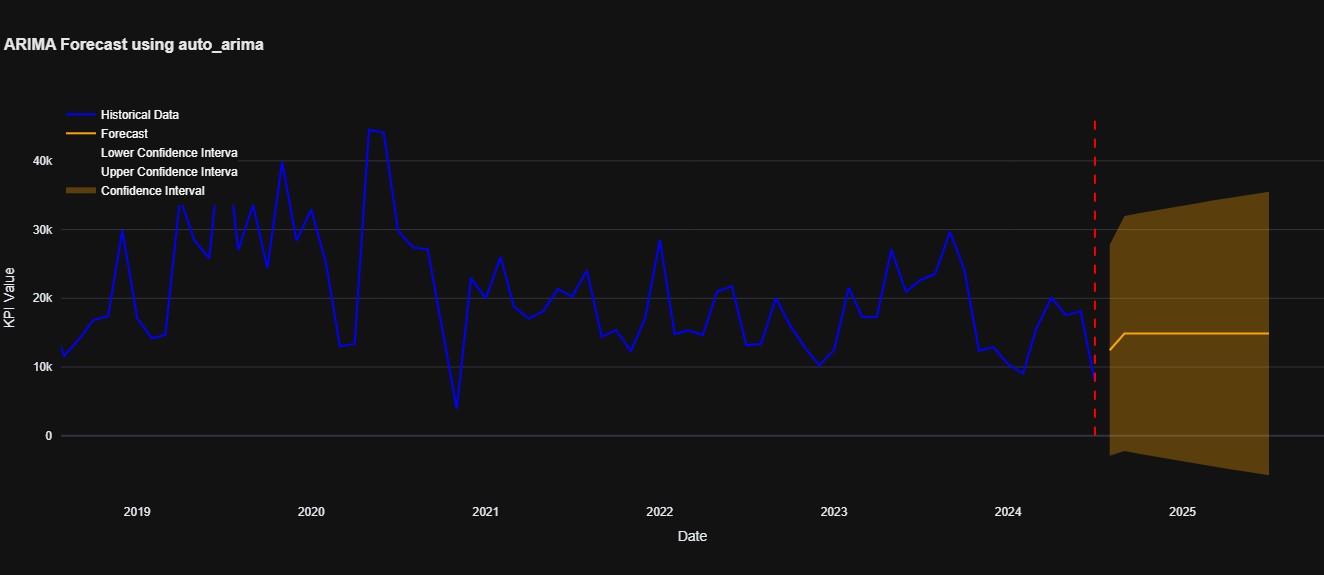


FIGURE 3.4 – Prédiction des ventes des pièces automobiles

#### 1. Données Historiques

- Visualisation : Les données historiques sont représentées par une ligne bleue. Cette ligne montre les fluctuations des valeurs au fil du temps, indiquant la tendance générale et la variabilité. - Tendances : On observe des variations saisonnières et des tendances, ce qui est typique des séries temporelles.

#### 2. Prévisions

- Prévisions : La ligne orange représente les prévisions futures du modèle ARIMA. Ces prévisions se basent sur les données historiques et les tendances identifiées par le modèle. - Stabilité : La prévision semble se stabiliser autour d’une certaine valeur, indiquant que le modèle prévoit un plateau ou une tendance stable à long terme.

#### 3. Intervalles de Confiance

- Intervalles de Confiance : Les zones ombragées en marron montrent l’intervalle de confiance autour des prévisions. Cela indique la précision des prévisions et l’incertitude associée. - Interprétation : Un intervalle de confiance plus large suggère une plus grande incertitude. À mesure que l’on s’éloigne dans le temps, l’incertitude des prévisions augmente, ce qui est courant dans les prévisions de séries temporelles.

#### 4. Observations et Implications

- Robustesse du Modèle : Le modèle semble capturer la dynamique des données historiques de manière efficace, comme le montre la proximité des prévisions avec les valeurs historiques. - Planification : Les prévisions peuvent être utilisées pour la planification et la prise de décision dans des contextes pratiques, en tenant compte des incertitudes.

#### Conclusion

Le graphique indique que le modèle ARIMA utilisé est efficace pour modéliser les données historiques et fournir des prévisions. Cependant, il est crucial de continuer à surveiller la performance du modèle et d’ajuster si nécessaire, surtout si de nouvelles données montrent des tendances ou des variations significatives.

### 3.6.1 Évaluation de la performance des prédictions

Pour évaluer la performance des prédictions effectuées, plusieurs métriques statistiques ont été utilisées. Parmi les plus courantes, nous avons retenu :

— Mean Absolute Percentage Error (MAPE) : Cette métrique permet de mesurer l’erreur moyenne en pourcentage entre les valeurs réelles et prédites, offrant ainsi une indication claire de l’exactitude des prévisions.

— Root Mean Squared Error (RMSE) : Cette métrique mesure l’erreur quadratique moyenne entre les valeurs réelles et prédites, indiquant ainsi la précision du modèle.

Les performances du modèle ARIMA peuvent être évaluées à l’aide de différentes métriques. Dans ce cas, nous avons calculé le RMSE et le MAE.

#### 1. RMSE (Root Mean Squared Error)

- Valeur : 0.0057 - Interprétation : Le RMSE mesure l’erreur quadratique moyenne entre les valeurs prédites et les valeurs réelles. Une valeur de RMSE de 0.0057 indique que, en moyenne, les prévisions du modèle s’écartent légèrement des données réelles, suggérant que le modèle est relativement précis.

#### 2. MAE (Mean Absolute Error)

- Valeur : 0.0056 - Interprétation : Le MAE mesure l’erreur moyenne absolue entre les valeurs prédites et les valeurs réelles. Une valeur de 0.0056 indique que, en moyenne, les prévisions du modèle s’écartent des valeurs réelles de 0.0056, ce qui témoigne également d’une bonne précision.

#### Comparaison des Métriques

- Les valeurs de RMSE et MAE sont très proches, ce qui est un bon signe. Cela indique que le modèle présente une distribution d’erreurs relativement homogène, sans valeurs aberrantes majeures influençant les prévisions.

Conclusion

Dans ce chapitre, la prédiction des ventes réalisée avec le modèle ARIMA offre un outil précieux pour la gestion prévisionnelle des stocks au sein du garage *Service Car Rapide 1*. Les résultats obtenus montrent une bonne adéquation avec les ventes réelles, ce qui permet de planifier les approvisionnements de manière plus efficace. Cependant, des améliorations pourraient être envisagées en intégrant d’autres variables explicatives ou en utilisant des modèles plus complexes pour affiner les prédictions.

Chapitre 4

# Présentation de l’application et ces fonctionnalités

## 4.1 Introduction

Après présentation et explication du modèle suivi, nous allons voir les outils et environnements qui ont contribué à la codification de ce dernier.

## 4.2 Environnement matériel et logiciel

Processeur : AMD Ryzen 5 5600H with Radeon Graphics 3.30 GHz

Mémoire RAM : 16,0 Go

Type du système : Système d’exploitation 64 bits, processeur x64

Carte graphique : Nvidia Getforce RTX 3060

En ce qui concerne l’environnement logiciel, nous le présentons comme suit :

### 4.2.1 Python

Python est un langage de programmation interprété, multi-paradigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est aussi le langage de programmation open source le plus utilisé par les informaticiens. [11]



FIGURE 4.1 – Python logo

Les principales utilisations de Python par les développeurs sont :

— la programmation d’applications.

— la création de services web.

— la génération de code.

— la méta-programmation.

### 4.2.2 statsmodels

statsmodels est un module Python qui fournit des classes et des fonctions pour l’estimation de nombreux modèles statistiques différents, ainsi que pour la réalisation de tests statistiques, et l’exploration des données. Une liste exhaustive de statistiques de résultats est disponible pour chaque estimateur.

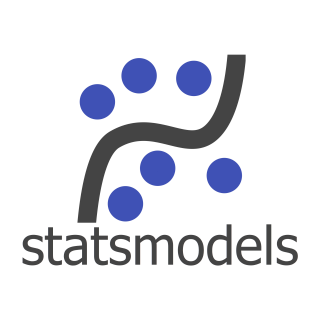


FIGURE 4.2 – statsmodels logo

statsmodels prend en charge la spécification de modèles à l’aide de formules de style R et de DataFrames [6]

### 4.2.3 NumPy

NumPy est une bibliothèque open source pour le langage de programmation Python. Cette bibliothèque est destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux.[12]



FIGURE 4.3 – NumPy logo

### 4.2.4 Sklearn (Scikit-learn)

Sklearn est une bibliothèque d’apprentissage automatique Python libre. Elle fournit dans son Framework une grande bibliothèque d’algorithmes d’implémentation prête à l’emploi dans le cadre. Ces bibliothèques sont particulièrement accessibles aux data

scientists.[13]



FIGURE 4.4 – Scikit-learn logo

En particulier, Sklearn comprend des fonctions d’estimation de forêts aléatoires, de régression logistique, d’algorithmes de classification et de machines à vecteurs de support. Elle est conçue pour fonctionner en harmonie avec d’autres bibliothèques Python libre, comme NumPy et SciPy.

### 4.2.5 Pandas

Pandas est une bibliothèque open source, sous licence BSD, offrant des performances structures de données faciles à utiliser et outils d’analyse de données pour le langage de programmation Python. [14]



FIGURE 4.5 – Pandas logo

La bibliothèque pandas Python qui fournit des données rapides, flexibles et expressives structures conçues pour faire travailler avec des données « relationnelles » ou « étiquetées » à la fois Facile et intuitif.

### 4.2.6 Ploty

La bibliothèque graphique Python de Plotly crée des graphiques interactifs de qualité publication. Exemples de création de diagrammes linéaires, de nuages de points, de graphiques en aires, de graphiques à barres, de barres d’erreur, de boîtes à moustaches, d’histogrammes, de cartes thermiques, de sous-tracés, de plusieurs axes, de graphiques polaires et de graphiques à bulles. [15]



FIGURE 4.6 – Ploty logo

### 4.2.7 Streamlit

Streamlit est une bibliothèque open source Python. Cette bibliothèque est un outil utilisé en machine learning qui vise à représenter graphiquement et à visualiser les données. Dans notre projet, streamlit a été utilisée pour construire l’interface graphique.



FIGURE 4.7 – Streamlit logo

## 4.3 Environnement de développement

Pour concevoir le code, nous avons utilisé deux environnements de développement :

Visual Studio Code et Google Colab.

### 4.3.1 Visual Studio Code

Visual Studio Code est un éditeur de code multiplateforme open source développé par Microsoft. Ce dernier a été développé sous Electron et utilise les fonctionnalités d’édition avancées du projet Monaco Editor.

Conçu principalement pour le développement d’applications utilisant JavaScript, TypeScript et Node.js, le système d’extension bien construit permet à l’éditeur de s’adapter à d’autres types de langage.



FIGURE 4.8 – Visual Studio Code logo

Souvent, les gens confondent entre visual code studio est visual studio. Visual studio est un IDE (Integrated development environment ou environnement de développement en français) de microsoft.

### 4.3.2 Google Colab

Google Colab est un environnement de notebook Jupyter open source qui s’exécute entièrement dans le cloud. De plus, aucune configuration n’est requise et les notebooks crées peuvent être modifiés par les membres de l’équipe en même temps, un peu comme la modification d’un document dans Google Docs.

Colab prend en charge de nombreuses bibliothèques de machine learning populaires qui peuvent être facilement chargées sur votre ordinateur portable comme scikit-learn.



FIGURE 4.9 – Google Colab logo

L’avantage de l’utilisation de Google Colab est qu’il offre un service Cloud gratuit avec un GPU gratuit. Cela veut dire que même si votre machine n’a pas de GPU ( Graphical Processing Units ou unité de traitement graphique en français), le code s’exécutera sans problèmes bien qu’il contient des outils comme ’Tensorflow’ car ’Tensorflow’ nécessite la présence d’un GPU.

L’inconvénient de Colab est la nécessité d’une connexion internet pour l’utiliser et aussi l’obligation d’importer votre projet à chaque fois que vous reconnectez à ce dernier.

## 4.4 Présentation de l’interface et fonctionnalités

Ce chapitre décrit en détail l’application développée pour prédire l’évolution des ventes des pièces automobiles au garage *Service Car Rapide 1*. L’application implémente le modèle ARIMA et offre une interface utilisateur permettant de réaliser des prévisions basées sur les données historiques de ventes. Nous abordons ici l’architecture de l’application, les fonctionnalités principales, ainsi que les technologies utilisées pour sa réalisation.

### 4.4.1 Page de présentation

La page de présentation e tableau de bord de l’application "Rapide Services Car II : Sale’s Analysis" de l’entreprise First Stop. Elle contient diverses informations et graphiques liés à l’analyse des ventes de services automobiles de l’entreprise.

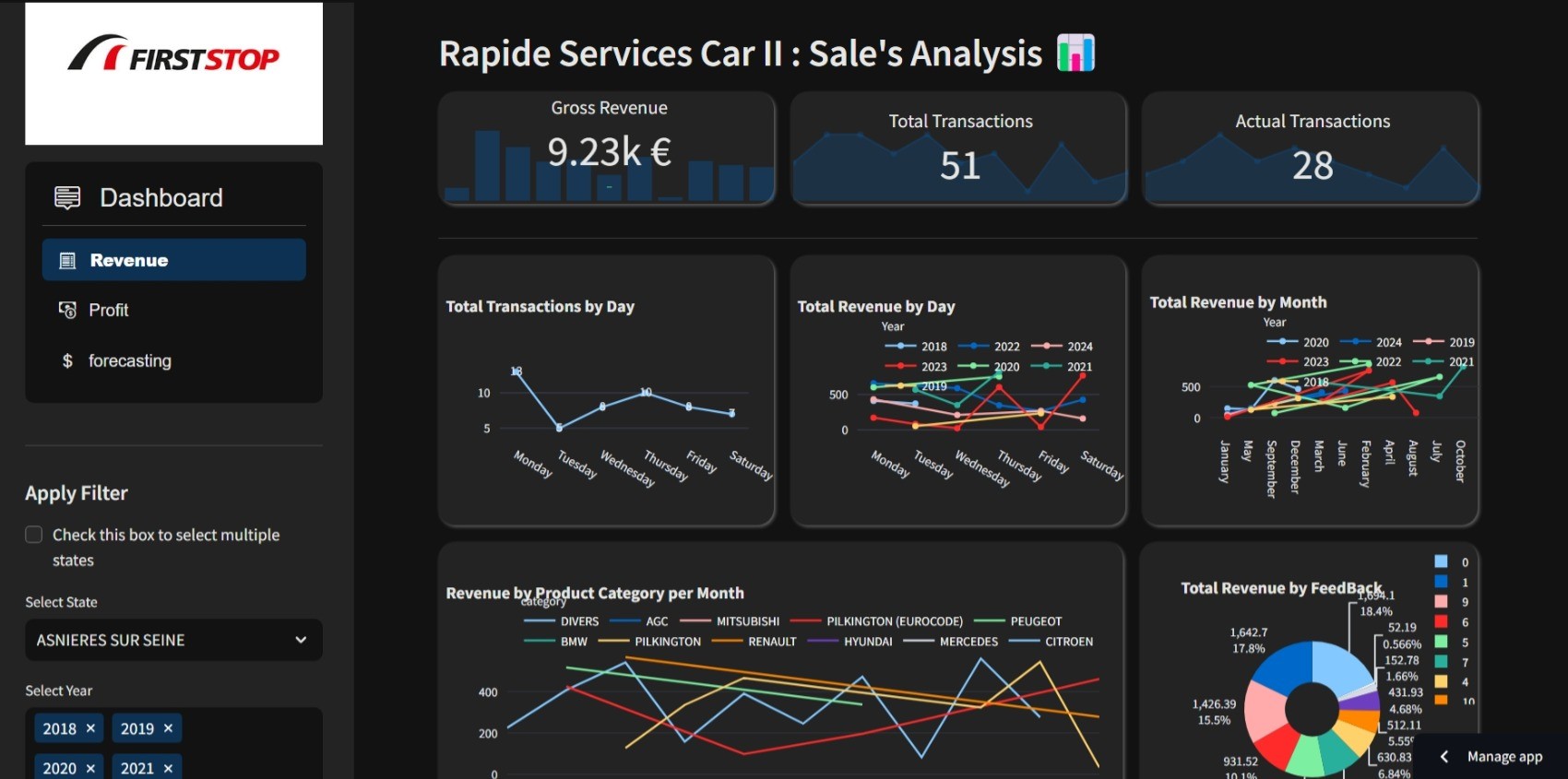


FIGURE 4.10 – Écran d’accueil

— Vue d’ensemble des indicateurs clés

— Accès rapide aux principales fonctionnalités

Fonctionnalités

— Revenue : Affichage des données de revenus

— Profit : Affichage des données de profits

— Forecasting : Affichage des prévisions de performance

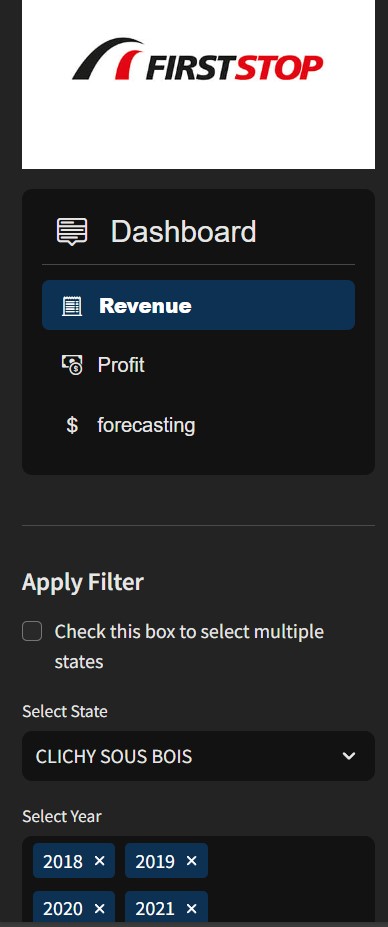


FIGURE 4.11 – Menu principal et filtres

Filtres disponibles

— Sélection d’un ou plusieurs États

— Sélection d’années entre 2018 et 2024

### 4.4.2 Section analyse revenue

cette section présente une analyse des ventes de la société Rapide Services Car II, fournissant diverses informations sur ses performances commerciales. Voici les principales fonctionnalités de cette analyse :

— Chiffre d’affaires brut

— Nombre total de transactions

— Nombre de transactions effectives

— Graphique des transactions totales par jour de la semaine.

— Graphique des recettes totales par jour de la semaine sur plusieurs années.

— Graphique des recettes totales par mois sur plusieurs années.

— Répartition des recettes par catégorie de produit par mois.

— Répartition des recettes totales par type de retour (FeedBack).

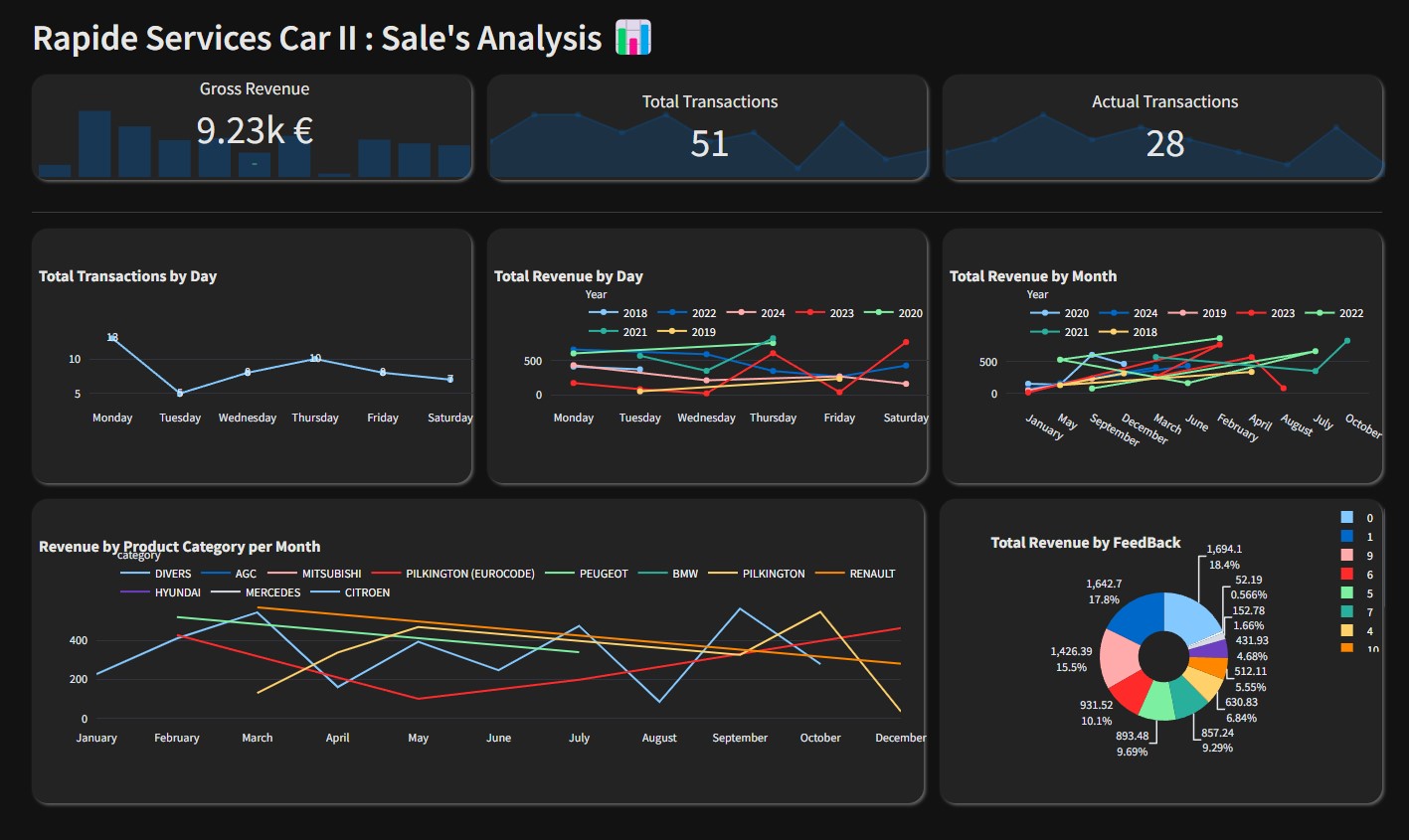


FIGURE 4.12 – Dashboard sur le revenu de l’entreprise

Cette analyse détaillée permet de visualiser les tendances des ventes, les variations hebdomadaires et mensuelles, ainsi que la contribution de chaque catégorie de produit et type de retour au chiffre d’affaires global de l’entreprise. Cela fournit des informations précieuses pour comprendre les performances de l’entreprise et orienter ses stratégies commerciales.

### 4.4.3 Section anlayse de profit

Cette section est pour l’analyse approfondie des performances de vente de Rapide Services Car II, en mettant l’accent sur les données financières et les indicateurs de performance clés. Voici les principales fonctionnalités présentées :

— Bénéfice total

— Marge bénéficiaire

— Graphique de la marge bénéficiaire et du bénéfice par mois.

— Répartition de la marge bénéficiaire par catégorie de produit.

— Classement des 10 principaux produits par bénéfice.

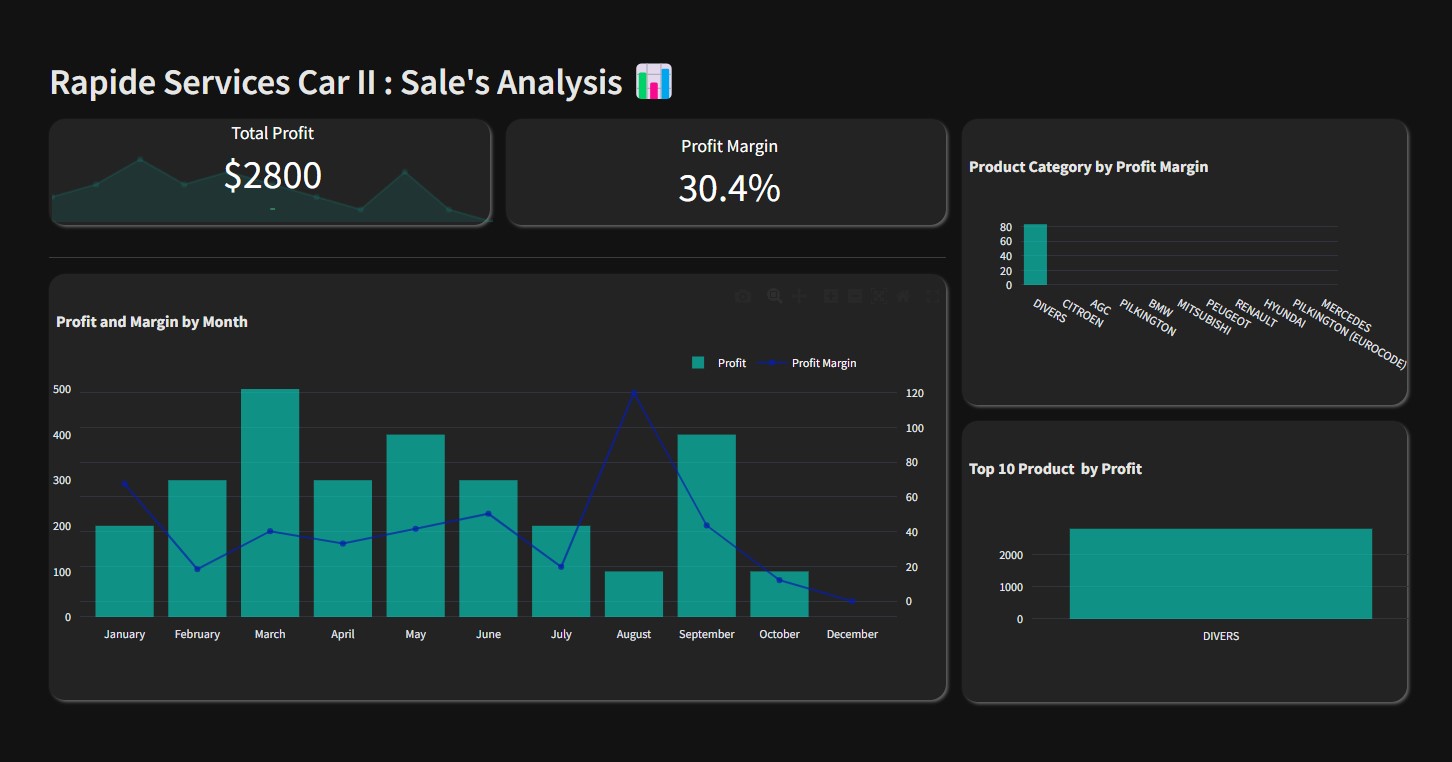


FIGURE 4.13 – Dashboard sur le profit de l’entreprise

Cette analyse financière détaillée permet de comprendre la rentabilité globale de l’entreprise, les tendances mensuelles des bénéfices et des marges, ainsi que les catégories de produits les plus performantes en termes de profitabilité. Ces informations sont essentielles pour prendre des décisions stratégiques éclairées et optimiser les performances de l’entreprise.

### 4.4.4 Section Prédiction

Fonctionnalité de prévision :

"Fit ARIMA Model" : Permet d’ajuster un modèle ARIMA (Autorégressive Integrated Moving Average) aux données

Evaluation du modèle :

Cette section affiche les métriques d’évaluation du modèle avec sa précision.

Courbe de prédiction

Cette visualisation permet aux décideurs d’avoir une meilleure compréhension de l’évolution attendue des ventes et des incertitudes associées, ce qui peut les aider dans leurs prises de décision stratégiques.

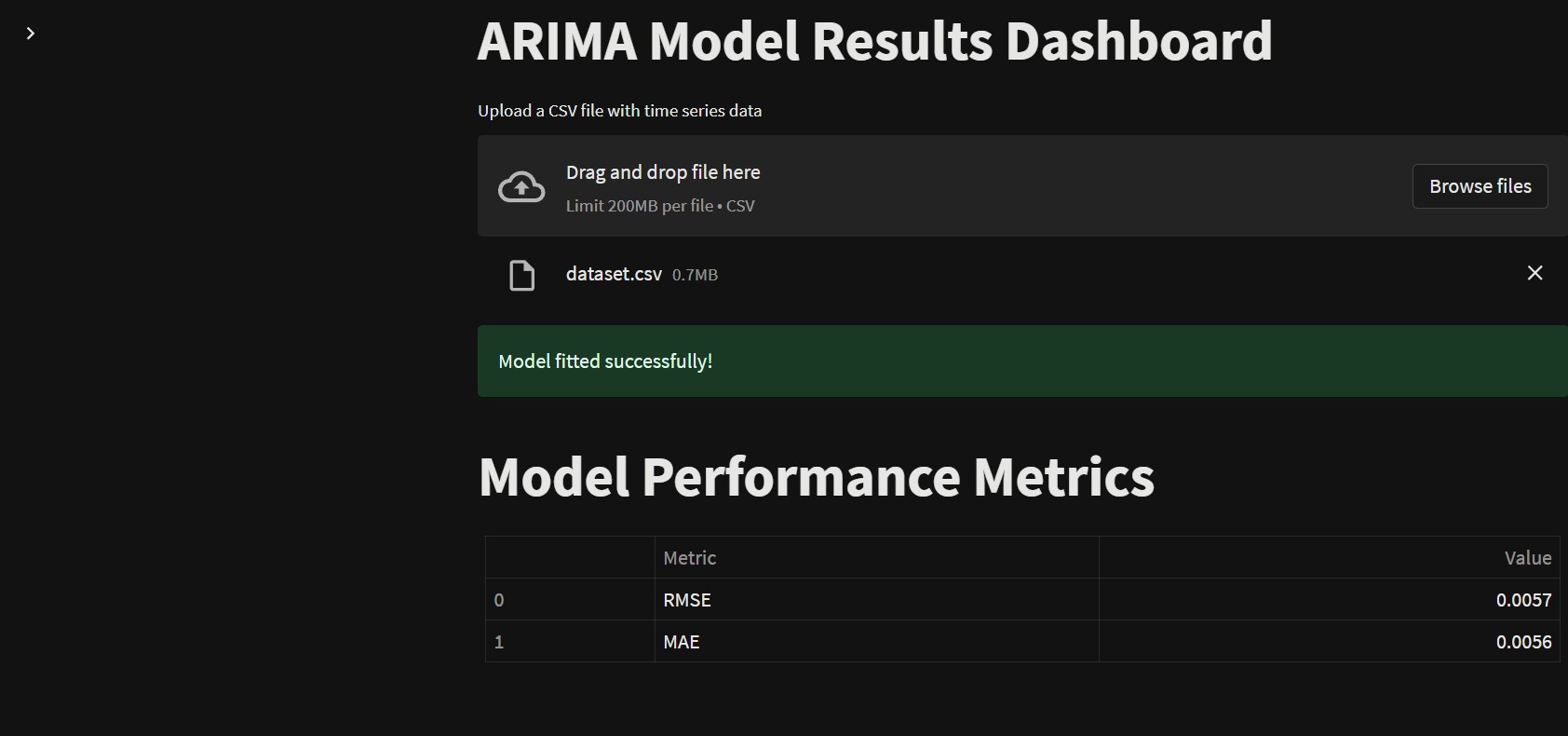


FIGURE 4.14 – Résumé sur le modèle

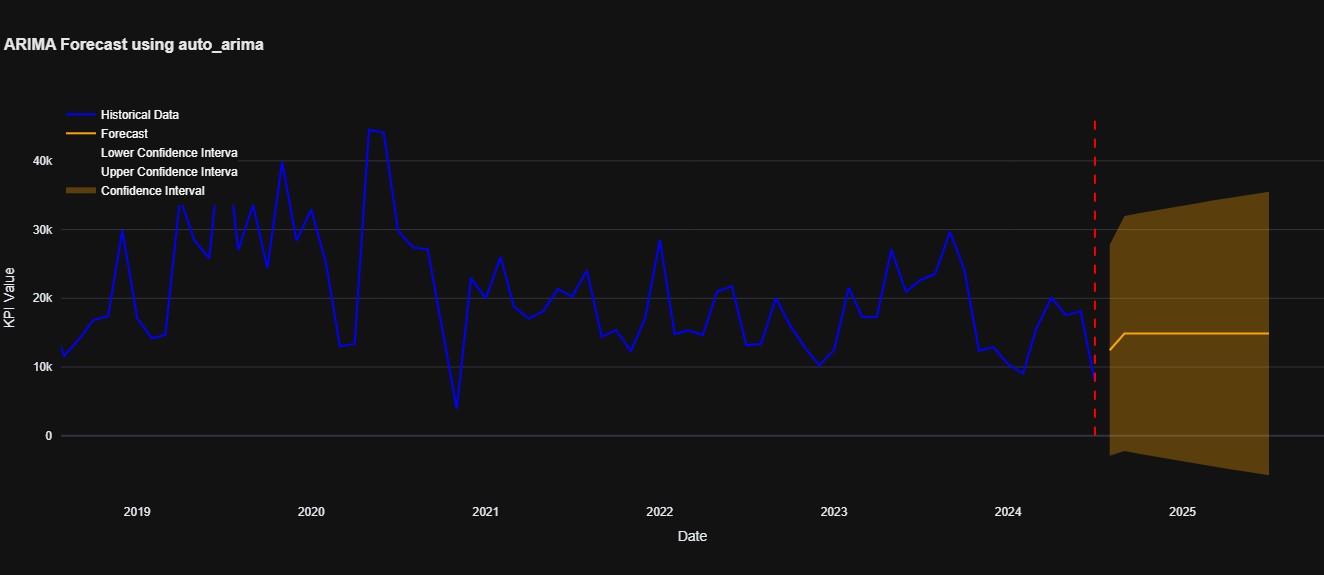


FIGURE 4.15 – ARIMA forecasting

### 4.4.5 Quelques problèmes rencontrés

La méthode donne d’une manière générale de bons résultats, néanmoins nous avons remarqué que par rapport à la taille des données, ça prend plus de temps de calcul et aussi l’erreur d’entraînement stable est du discriminateur 0.69 (69 %).

## 4.5 Conclusion

Parmi les outils et langages qui existent, nous avons présenté ceux qui nous ont été utiles pour réaliser notre projet de manière efficace et qui nous permettent d’avoir les résultats voulus d’une manière rapide. Nous avons aussi vue les résultats obtenus après l’exécution de notre code.

Conclusion Générale

# Conclusion Générale

L’émergence de la Big Data représente une révolution majeure dans la façon dont les entreprises collectent, analysent et exploitent les données. Dans un secteur aussi dynamique que l’automobile, la capacité à tirer parti des données massives s’avère essentielle pour optimiser les opérations, améliorer la satisfaction client et renforcer la compétitivité. Au cours de mon expérience au sein de Rapide Services Car II, j’ai eu l’opportunité d’appliquer des techniques avancées d’analyse des données, en particulier à travers la conception d’un modèle de prédiction basé sur les séries temporelles, utilisant la méthode ARIMA.

Les modèles ARIMA, reconnus pour leur efficacité dans l’analyse des données chronologiques, permettent d’exploiter les tendances et les fluctuations des ventes passées pour anticiper les besoins futurs. L’application de cette méthode sur les données de vente du garage a permis d’identifier des patterns saisonniers et des cycles récurrents, fournissant ainsi des insights précieux pour la planification des stocks de pièces détachées et l’organisation des interventions. Grâce à ces prédictions, Rapide Services Car II est en mesure de mieux gérer ses ressources, d’optimiser ses approvisionnements et de réagir de manière proactive aux variations de la demande.

De plus, l’intégration de modèles prédictifs dans le processus décisionnel stratégique permet de transformer des données brutes en connaissances exploitables. Cela contribue non seulement à une meilleure prise de décision, mais également à une amélioration continue des services offerts aux clients. En anticipant les périodes de forte affluence, le garage peut ainsi ajuster ses opérations et son personnel, garantissant une réactivité optimale et une expérience client de qualité.

En somme, cette expérience souligne l’importance cruciale de la Big Data et de l’analyse prédictive dans le secteur automobile. La mise en œuvre de modèles tels qu’ARIMA non seulement enrichit la compréhension des dynamiques de vente, mais établit également une base solide pour des stratégies d’affaires éclairées. À l’avenir, il sera essentiel pour Rapide Services Car II de poursuivre l’optimisation de ses modèles de prédiction, tout en explorant d’autres techniques avancées d’analyse de données, afin de maintenir son avantage concurrentiel dans un marché en constante évolution.

Bibliographie.

# Bibliographie

1. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis : Forecasting and Control*. 5th Edition. Wiley.
2. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting : Principles and Practice*. 2nd Edition. OTexts. Disponible sur : *https://otexts.com/fpp2/*
3. Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications : With R Examples*. 4th Edition. Springer.
4. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
5. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
6. Chai, T., & Draxler, R. R. (2014). *Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? Geoscientific Model Development*, 7(3), 1247-1250.
7. Kourentzes, N., Petropoulos, F., & Spiliotis, E. (2014). *Forecasting with Exponential Smoothing :*

*The State Space Approach*. *International Journal of Forecasting*, 30(3), 394-401.

1. Liu, W., & Zhang, L. (2018). *A Review of Time Series Prediction Methods for Stock Market*.

*Journal of Forecasting*, 37(7), 785-796.

1. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
3. Python-Wikipédia [en ligne]. Disponible sur : *https://fr.wikipedia.org/wiki/*

*Python\_(langage)*

1. NumPy — Wikipédia [en ligne]. Disponible sur : *https://fr.wikipedia.org/wiki/ NumPy*
2. Scikit-learn — Wikipédia [en ligne]. Disponible sur : *https://fr.wikipedia.org/ wiki/Scikit-learn*
3. SciPy — Wikipédia [en ligne]. Disponible sur : *https://fr.wikipedia.org/wiki/*

*SciPy*

1. Plotly [en ligne]. Disponible sur : *https://plotly.com/python/*