

**PRÉVISION DES VENTES DES EPICERIES FAVORITA EN EQUATEUR PAR LA MÉTHODE HOLT-WINTERS**

**SUJET** :

***Pour le diplôme d’Ingénieur Statisticien et Data***

**MINI-PROJET**

**I**nstitut **S**upérieur de **S**tatistique, d’**E**conométrie et de **D**ata **S**cience

*Ministère de l’Enseignement Supérieur et de la Recherche Scientifique*

**---------------------------------------------------------------------**

*Union – Discipline - Travail*

**REPUBLIQUE DE CÔTE D’IVOIRE**

**ENCADREUR PEDAGOGIQUE :**

**M. Didier Martial Akposso**

Directeur Fondateur de l’INSSEDS

Economiste-Statisticien

Data-Scientist

Présenté par :

**SAVANE MORY**

Promotion 2023-2024

Ingénierie Statistique et Data Science

**Période de rédaction : Décembre 2023**

© Année académique : 2023-2024

# Avant-Propos

Ce travail s'inscrit dans le cadre de la formation au cycle d’ingénieur Statisticien et Data à l’Institut Supérieur de Statistique, d’Économétrie et de Data Science (INSSEDS). L'objectif de ce mini-projet est d'évaluer la compréhension du cours portant sur l'Analyse des séries temporelles et la prédiction à l’aide de la méthode Holt-Winters. Il est important de souligner que ce projet est strictement réalisé à des fins académiques et n'engage en aucun cas la responsabilité de l’école.

En tant qu'auteur, nous reconnaissons humblement que notre travail ne prétend pas avoir exploré tous les aspects nécessaires pour conférer à cette étude le caractère d'une recherche véritablement approfondie. Il s'agit plutôt d'une exploration limitée dans le cadre des enseignements dispensés au sein de notre formation.

Nous souhaitons ainsi offrir une contribution modeste à la compréhension des concepts étudiés tout en restant conscient de la complexité et de l'étendue du domaine de l'analyse des séries temporelles et de la prédiction. Ce travail ne prétend pas être exhaustif, mais plutôt constituer une étape préliminaire dans notre parcours académique.

Nous exprimons notre gratitude envers nos enseignants et toute l'équipe pédagogique de l'INSSEDS pour leur encadrement et leurs enseignements enrichissants qui ont contribué à la réalisation de ce projet.

Enfin, nous espérons que ce travail pourra susciter des réflexions et des discussions fructueuses au sein de la communauté académique et être une source d'apprentissage continu pour tous ceux qui s'intéressent à ce passionnant domaine.

# Sigles

ARMA : Autorégressif et Moyenne Mobile

ACF : Autocorrelation Function

PACF : Partial Autocorrelation Function

# Liste des illustrations

# Listes des figures

[Figure 1: Illustration de la conversion 9](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491587)

[Figure 2: Visualisation des valeurs aberrantes 9](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491588)

[Figure 3: Evolution et Répartition des ventes par Mois 10](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491589)

[Figure 4: Histogramme et densité de la distribution des ventes 11](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491590)

[Figure 5: Visualisation des différents types de modèle 14](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491591)

[Figure 6: Décomposition de la série en Tendance, Saisonnalité et Résidu 15](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491592)

[Figure 7: Autocorrélogrammes 15](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491593)

[Figure 8: Résultat du modèle Holt-Winters 16](#_Toc153491594)

[Figure 9: Entrainement et test du modèle 17](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491595)

[Figure 10: Graphiques des tests 18](#_Toc153491596)

[Figure 11: Prévision des ventes sur 12 mois 19](file:///C:\Users\THINKPAD\OneDrive\Bureau\Miniprojets\MiniProjet2\MiniProjet2_SavanéMory_2023.docx#_Toc153491597)

# Liste des tableaux

[Tableau 1: Dictionnaire des données 8](#_Toc153491686)

[Tableau 2: Résumé numérique des quantités vendues 11](#_Toc153491687)

[Tableau 3: Test de Stationnarité et normalité 16](#_Toc153491688)

[Tableau 4: Coefficients du modèle Holt-Winters 17](#_Toc153491689)

[Tableau 5: Résultat des tests 18](#_Toc153491690)

[Tableau 6: Prédiction des valeurs 19](#_Toc153491691)

# Sommaire

[Avant-Propos 1](#_Toc153576832)

[Sigles 2](#_Toc153576833)

[Liste des illustrations 3](#_Toc153576834)

[Listes des figures 3](#_Toc153576835)

[Liste des tableaux 3](#_Toc153576836)

[Sommaire 4](#_Toc153576837)

[Résumé 6](#_Toc153576838)

[Abstract 7](#_Toc153576839)

[Introduction Générale 9](#_Toc153576840)

[I. STATISTIQUE DESCRIPTIVE DES VENTES 11](#_Toc153576841)

[Introduction 11](#_Toc153576842)

[1. Prétraitement des données 11](#_Toc153576843)

[a) Présentation des données 11](#_Toc153576844)

[b) Nettoyage des données 12](#_Toc153576845)

[2. Analyse descriptive de l’évolution des ventes entre 2013 et 2017 13](#_Toc153576846)

[a) Analyse des graphiques et des indicateurs 13](#_Toc153576847)

[b) Interprétation 15](#_Toc153576848)

[Conclusion 15](#_Toc153576849)

[II. ANALYSE TEMPORELLE ET PREVISION DES VENTES 16](#_Toc153576850)

[Introduction 16](#_Toc153576851)

[1. Série temporelle 16](#_Toc153576852)

[a) Définition des concepts 16](#_Toc153576853)

[b) Le modele de prévision de Holt-Winters 17](#_Toc153576854)

[2. Conditions préliminaires 18](#_Toc153576855)

[a) Décomposition de la série 18](#_Toc153576856)

[b) Vérification de la stationnarité 18](#_Toc153576857)

[3. Prévision des ventes d’articles 20](#_Toc153576858)

[a) Construction du modèle 20](#_Toc153576859)

[b) Validation du modèle 21](#_Toc153576860)

[c) Prédiction des ventes pour les 12 prochains mois. 23](#_Toc153576861)

[Conclusion 24](#_Toc153576862)

[Conclusion générale 25](#_Toc153576863)

[Annexe 27](#_Toc153576864)

[Bibliographie et Webographie 30](#_Toc153576865)

# Résumé

L'étude vise à élaborer une technique robuste de prévision des ventes pour les épiceries Favorita en Équateur en utilisant la méthode de Holt-Winters. Ancrée dans le contexte dynamique du commerce de détail, l'objectif principal est de déterminer si ce modèle peut efficacement saisir la tendance à la hausse des ventes et la saisonnalité, fournissant ainsi des prévisions fiables pour orienter la gestion des ventes et les décisions stratégiques de l'entreprise.

L'analyse des données historiques de 2013 à 2017 révèle une croissance constante des ventes, avec des performances marquées en décembre, atteignant environ 25 millions d'unités. La distribution des ventes, caractérisée par deux sommets successifs, indique une propension à vendre en grande quantité. La méthode de Holt-Winters est validée par des tests de résidus démontrant sa robustesse avec des résidus stationnaires, un bruit blanc, et une distribution gaussienne.

En guise de recommandation, nous conseillons l'adoption de la méthode de Holt-Winters pour la prévision des ventes. Les prévisions générées doivent être mises en œuvre pour optimiser la gestion des stocks, planifier les opérations, surtout pendant les périodes de pointe. Toutefois, il est crucial de reconnaître les limites de l'analyse, notamment la dépendance aux tendances passées. Des améliorations potentielles incluent l'intégration de variables externes et une évaluation continue du modèle pour une adaptation proactive aux changements. Cette étude contribue ainsi à la précision des prévisions dans le secteur du commerce de détail, offrant des perspectives stratégiques pour les épiceries Favorita en Équateur.

Mots clés : Séries temporelles ; Prédiction ; méthode de Holt-Winters

# Abstract

This study aims to develop a robust sales forecasting technique for Favorita supermarkets in Ecuador using the Holt-Winters method. Rooted in the dynamic context of the retail industry, the primary objective is to determine whether this model can effectively capture the upward trend in sales and seasonality, thus providing reliable forecasts to guide inventory management and strategic decision-making for the company.

Analysis of historical data from 2013 to 2017 reveals a consistent growth in sales, with notable performances in December, reaching around 25 million units. The sales distribution, characterized by two successive peaks, indicates a tendency to sell in large quantities. The Holt-Winters method is validated through residual tests, demonstrating its robustness with stationary residuals, white noise, and a Gaussian distribution.

In conclusion, we recommend the continued adoption of the Holt-Winters method for sales forecasting. The generated forecasts should be implemented to optimize inventory management and operational planning, especially during peak periods. However, it is crucial to acknowledge the limitations of the analysis, including reliance on past trends. Potential improvements involve incorporating external variables and ongoing model evaluation for proactive adaptation to changes. This study contributes to the accuracy of forecasts in the retail sector, offering strategic insights for Favorita supermarkets in Ecuador.

Key words: Temporal series; prediction; Holt-Winters method

# Introduction Générale

Dans le contexte dynamique du commerce de détail, où la compétitivité est cruciale, la capacité à anticiper les tendances de vente demeure un impératif pour les entreprises. Notre étude se positionne au cœur de cette nécessité en cherchant à estimer correctement les ventes et à gérer les stocks. Le problème clé est de prévoir les besoins en ventes et en stocks pour chaque emplacement afin d’éviter les surstocks et les sous-stocks, permettant ainsi à l’entreprise d’offrir le meilleur service client tout en réduisant les pertes et en garantissant la durabilité du magasin. Face à l'évolution constante du marché, l'importance stratégique d'outils de prévision précis et adaptables devient indéniable pour orienter les décisions opérationnelles et optimiser la gestion des stocks.

Au cœur de notre réflexion réside la question centrale de savoir si ***le modèle de Holt-Winters peut véritablement saisir la dynamique de la tendance à la hausse des ventes et la saisonnalité propre aux épiceries Favorita en Équateur***. Cette interrogation constitue le fil conducteur de notre étude, visant à évaluer la pertinence de ce modèle dans la fourniture de prévisions fiables, cruciales pour une gestion proactive des stocks et des décisions stratégiques éclairées.

Notre analyse a donc pour objectif principal d’élaborer une technique de prédiction robuste pour les épiceries Favorita en Équateur.

L’atteinte de cet objectif passera par la résolution des objectifs spécifiques ci-dessous :

* Déterminer la présence de tendance
* Déterminer la présence de saisonnalité
* Eliminer les effets de tendance et de saisonnalité
* Vérifier les résidus

Pour mener à bien ce travail, nous posons les hypothèses suivantes :

1. L’évolution des ventes a une tendance haussière et une saisonnalité
2. Le modèle de Holt-Winters est adapté pour la prévision des ventes

Pour atteindre ces objectifs, nous avons adopté une méthodologie rigoureuse. L'analyse repose sur les données historiques des ventes de 2013 à 2017 des épiceries Favorita en Équateur. Le choix délibéré de la méthode de Holt-Winters, en raison de sa capacité à modéliser conjointement la tendance et la saisonnalité, structure notre approche. Les tests de résidus ont été appliqués comme un indicateur critique de la qualité du modèle. Notre démarche intègre également une réflexion approfondie sur les limites inhérentes à l'analyse, offrant ainsi des perspectives pour une amélioration continue des résultats.

Ainsi, afin de ressourdre efficacement cette étude, nous l’avons érigé essentiellement en deux parties, d’une part, l’analyse descriptive des ventes et d’autres parts, la prédiction des ventes sur 12 mois.

# STATISTIQUE DESCRIPTIVE DES VENTES

## Introduction

Le nettoyage des données ou data cleansing est une étape essentielle de l’analyse statistique et la Data Science. Il s’agit de corriger ou supprimer des enregistrements inexacts dans des jeux de données afin de pouvoir les exploiter par la suite.

C’est un processus qui vise, de manière générale, à améliorer **la qualité des données**. En pratique, il consiste à **importer les données** dans un logiciel statistique, dans notre Jupyter Notebook du logiciel Python. Ensuite, faire une **première visualisation** pour avoir une idée succincte de la structure de nos données et enfin, identifier et corriger les valeurs manquantes, valeurs aberrantes, valeurs extrêmes afin de qu’elles puissent être plus cohérentes et sans erreurs.

## Prétraitement des données

### Présentation des données

Les données utilisées dans ce rapport sont issues de la plateforme dédiée à la Data Science Kaggle.

Elles portent sur les quantités d’articles vendues sur quatre années de 2013 à 2017.

Le dictionnaire des données se présente comme suit :

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Nature/Type | Description | Modalités |
| date | Object (Qualitative) | Les moments d’enregistrement des opérations | Format : YYYY-MM-DD  Débute du 2013-01-01 au 2017-08-15 |
| id | int64 (Quantitative) | Identifiant des familles de produits | Nombres entiers |
| store\_nbr | int64 (Quantitative) | Nombre de magasin | 9 |
| family | object (Qualitative) | Les différents types de produits vendus | 33 différentes familles de produits |
| sales | float64 (Quantitative) | Les quantités d’articles vendues | Numérique |
| onpromotion | int64 (Quantitative) | Nombre de produits en promotion | Nombres entiers |

Tableau 1: Dictionnaire des données

*Source : Réalisé par l’auteur*

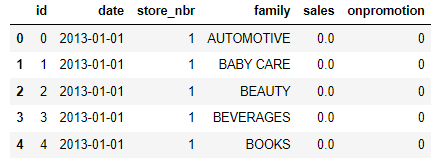
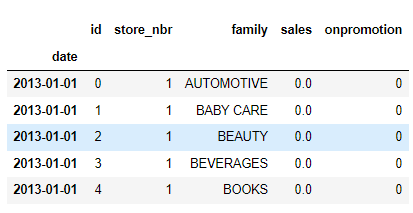
La variable d’intérêt dans cette analyse est la variable « sales ».

### Nettoyage des données

L’objectif principal de cette analyse étant de construire un modèle permettant de prédire les quantités d’articles vendues dans les épiceries Favorita, nous avons alors convertie le type de la variable **date** en type ***datetime*** et l’avons mis comme ***index*** afin de faciliter la manipulation.

Figure 1: Illustration de la conversion

*Source : Réalisé par l’auteur avec Python*



De plus, nos données mensuelles sont exemptes de valeurs manquantes et de valeurs aberrantes/Extrêmes, le graphique ci-dessous nous permet de constater ce fait.

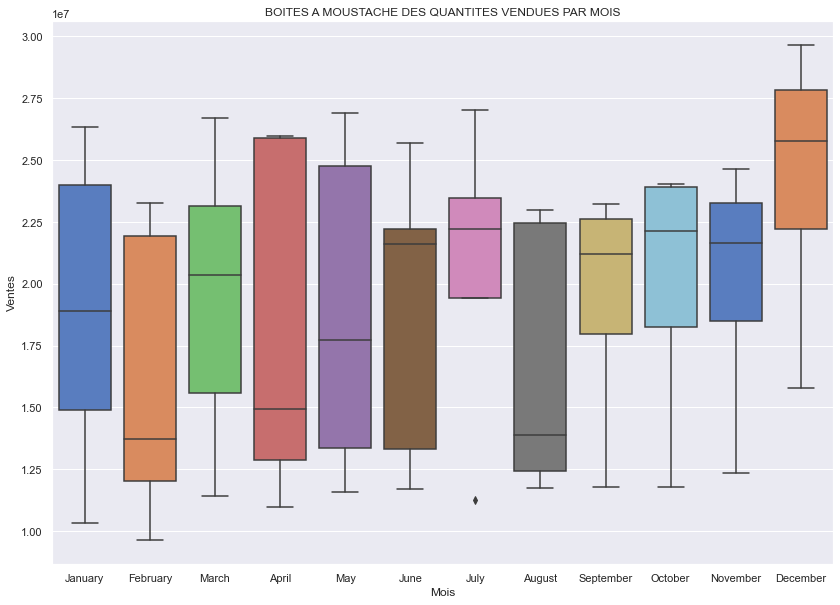


Figure 2: Visualisation des valeurs aberrantes

*Source : Réalisé par l’auteur avec Python*

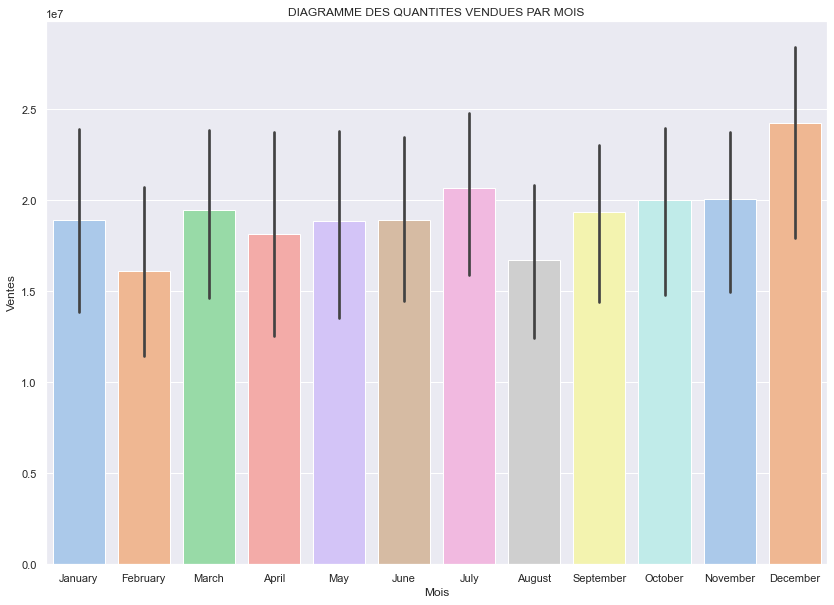
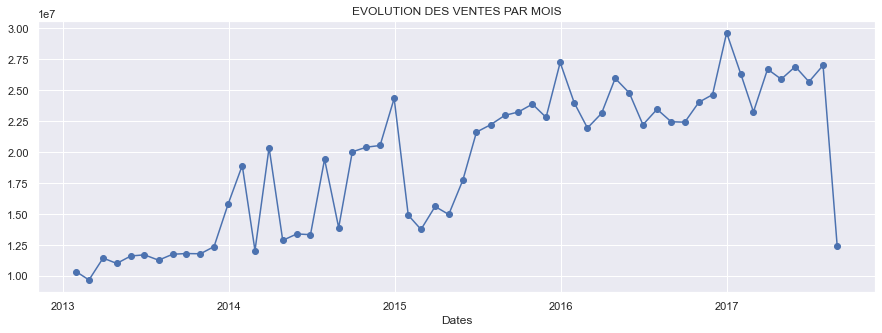
*Source : Réalisé par l’auteur avec Python*

## Analyse descriptive de l’évolution des ventes entre 2013 et 2017

### Analyse des graphiques et des indicateurs

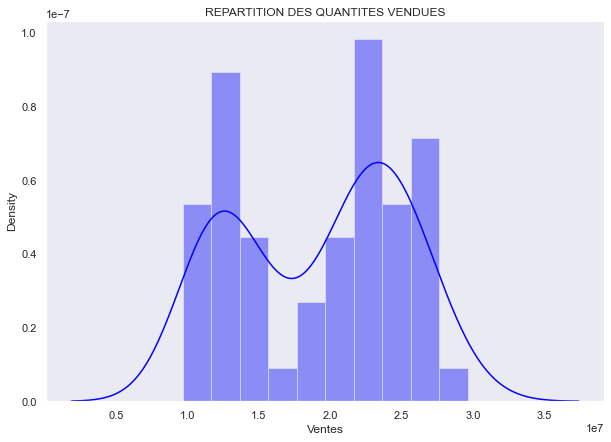
Les graphiques ci-dessous représentent l’évolution et la répartition des quantités d’articles vendus dans les épiceries de la chaine de magasin Favorita par mois sur quatre ans, de 2013 à 2017.

Figure 3: Evolution et Répartition des ventes par Mois



*Source : Réalisé par l’auteur avec Python*

La courbe d’évolution a une allure croissante et présente des pics généralement à la fin de chaque année qui sont suivis immédiatement par des baisses.

Le diagramme en barre vient éclaircir notre constat concernant les pics en fin d’année. Alors, nous pouvons aisément remarquer que les plus grandes quantités sont vendues dans le mois de décembre avec une quantité avoisinant les 25 million d’articles.

*Source : Réalisé par l’auteur avec Python*

Figure 4: Histogramme et densité de la distribution des ventes

Ce graphique met en exergue un histogramme avec une courbe de densité représentée sur un même plan. La répartition des données est relativement inégale avec une courbe de densité à deux sommets.

|  |  |  |
| --- | --- | --- |
| Indicateur | Valeur | Interprétation |
| INDICATEUR DE TENDANCE CENTRALE ET DE POSITION | | |
| Minimum | 9 658 959 ,78 | La quantité minimale vendue s’élève à plus de 9.600.000 produits |
| Maximum | 29.640.287,58 | La quantité maximale vendue s’élève à plus de 29.600.000 produits |
| Moyenne | 19.172.231,29 | La quantité moyenne vendue est de plus de 19.100.000 produits |
| Ecart-type | 5.806.956,05 | Les quantités vendues oscillent dans l’intervalle  [13 365 275,24 ; 24 979 187,34] |
| INDICATEUR DE FORME | | |
| Skewness | -0.15 | La valeur négative de cet indicateur de forme montre que la distribution des données est étalée à gauche. |
| Kurtosis | -1.43  *Source : Réalisé par l’auteur* | Une valeur du Kurtosis inferieur a 3 met en exergue une distribution platikurtique, c’est à dire quelle est plus plate que la distribution de la loi normale. |

Tableau 2: Résumé numérique des quantités vendues

### Interprétation

L’analyse descriptive des quantités d’articles vendus dans les épiceries Favorita en Equateur nous permet d’affirmer que ces quantités augmentent progressivement au fil des années.

Ces données sont sujettes à des composantes tendancielles et saisonnières. La tendance haussière est aisément identifiable sur la courbe représentant l’évolution des ventes (figure 1). En ce qui concerne la saisonnalité, nous remarquons une augmentation des ventes en fin de chaque année pour atteindre des pics, majoritairement en décembre, dû surement aux fêtes de fin d’année.

La quantité moyenne d’articles vendus par mois s’élève à plus de 19**.** 100.000 articles avec un écart-type de plus ou moins 5**.**806**.**956**,**05. (Tableau 1)

Par ailleurs, suite à la visualisation de la répartition de la distribution (figure 2), nous avons conclu une inégale répartition des données. Le Skewness qui est un indicateur de forme nous autorise à confirmer que la distribution est légèrement étalée à gauche avec une valeur de -0,15. Cette valeur signifie qu’il y a une queue à gauche de la distribution. Cela voudrait dire que le nombre de mois avec de grandes quantités vendues est élevé, en d’autres termes, **l’entreprise vend en grande quantité la majorité des mois**.

# Conclusion

Au terme de cette première partie, il ressort que les quantités d’articles vendues par ces épiceries augmentent au fil des années avec des meilleures performances enregistrées dans les mois de décembre (environ 25 millions).

La distribution présente une forme assez irrégulière avec deux sommets successifs. La valeur du coefficient d’asymétrie nous amène à conclure que cette distribution est relativement étalée à gauche ce qui voudrait dire que l’entreprise vend généralement en grande quantité.

# ANALYSE TEMPORELLE ET PREVISION DES VENTES

## Introduction

Dans cette partie, nous nous attellerons dans un premier temps à la clarification de certains concepts liés aux séries temporelles et présenterons la méthodologie utilisée, ensuite, nous passerons à la construction et au test du modèle et enfin nous ferons la Prévision grâce à la méthode de Holt-Winters.

## Série temporelle

### Définition des concepts

**L'analyse des séries chronologiques[[1]](#footnote-1)** est une méthode utilisée dans diverses industries, telles que la prévision boursière, les prévisions météorologiques, la prévision des cryptomonnaies, etc., pour prévoir les valeurs futures sur la base de données historiques. Dans notre cas, il s’agira de faire la prévision des quantités d’article qui sera vendus. (Köseoğlu 2023)

Toutes les séries chronologiques peuvent être divisées en trois composantes :

**La tendance** ou **trend** : changements lents qui se produisent dans les séries chronologiques.

**La saisonnalité** : modèles de variation qui se répètent à des intervalles de temps spécifiques. Ceux-ci peuvent être hebdomadaires, mensuels, annuels, etc. Ce sont des mouvements périodiques de haut en bas autour du niveau actuel.

**Les résidus** : événements inhabituels qui se produisent dans les données, tels qu'une augmentation soudaine de la fréquence cardiaque d'une personne pendant l'exercice. Ceux-ci provoquent des erreurs aléatoires et sont également appelés « bruit blanc ». (Louis 2021)

L’observation des séries chronologiques permet de distinguer deux grands types de série : celles qui se conforment au modèle multiplicatif et celles qui se conforment au modèle additif. (MAZEROLLE 2006)

**Le modèle additif**: les variations autour du trend demeurent dans une bande de variation à peu près constante.

**Le modèle multiplicatif :** au contraire, les variations autour du trend s’amplifient.

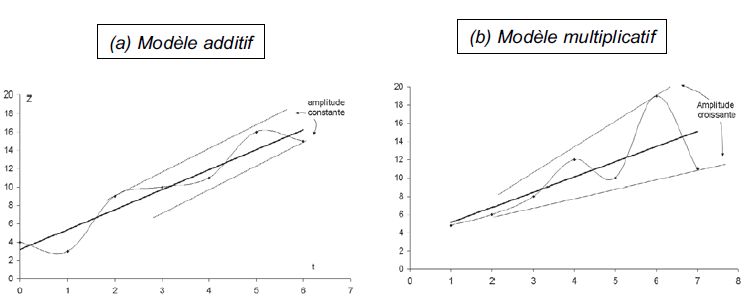


Figure 5: Visualisation des différents types de modèle

*Source :* (MAZEROLLE 2006)

### Le modele de prévision de Holt-Winters

La méthode de prévision utilisée dans ce rapport est la méthode de Holt-Winters ou de lissage exponentiel triple.

C’est est une technique de lissage qui ressemble un peu aux moyennes mobiles. Cette méthode peut être utilisée pour prédire une série temporelle. C'est en fait un cas particulier des modèles Autorégressif et Moyenne Mobile (ARMA). On en distingue différents types (simple, double, triple...). On notera que les lissages exponentiels simple et double ne sont plus vraiment d'actualité donc on présente seulement le lissage exponentiel triple ou méthode de Holt-Winters qui, lui, est parfois utilisé.

Il est adapté à des séries à tendance localement linéaire avec en plus une composante saisonnière. Ce qui est le cas de notre série.

## Conditions préliminaires

### Décomposition de la série

*Source : Réalisé par l’auteur*

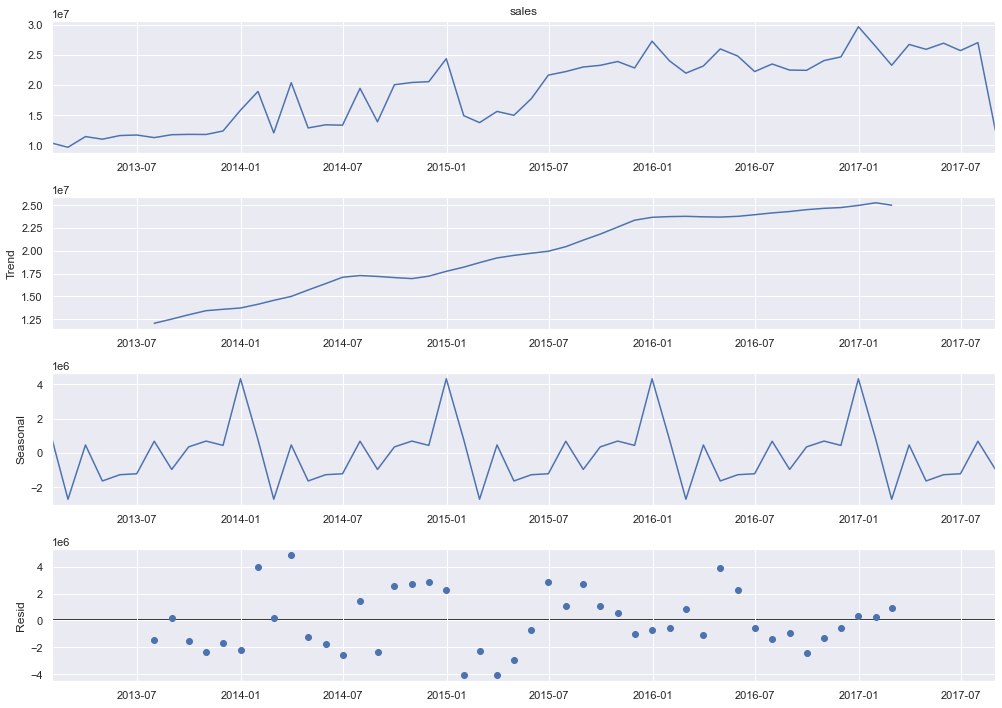
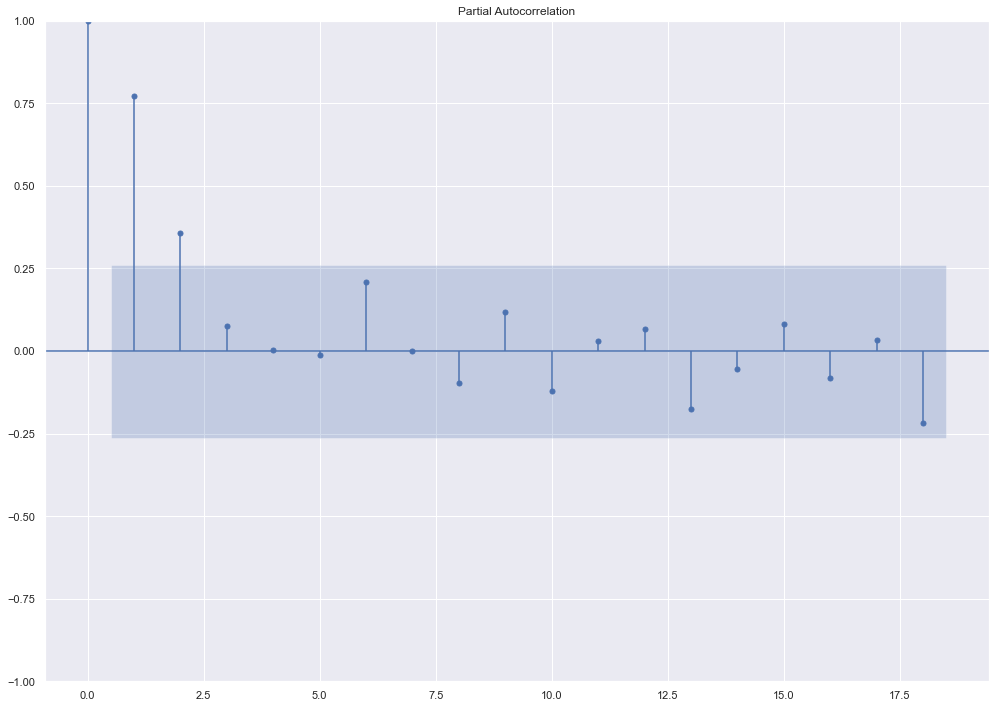
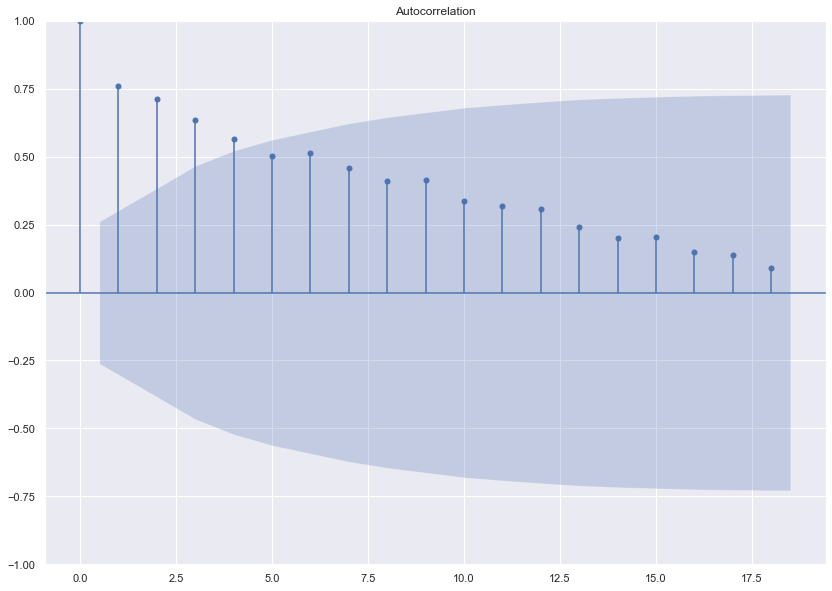


Figure 6: Décomposition de la série en Tendance, Saisonnalité et Résidu

Le graphique ci-dessus nous présente le découpage de la série originale en plusieurs composants indépendants : tendance, saisonnalité, cycle et bruit. Cette décomposition de notre série nous permet d’affirmer que **la tendance est quasi-linéaire et croissante** et aussi que **la saisonnalité est annuelle** avec amplitude constante c’est à dire que les mêmes mouvements se répètent sur les 12 mois de l’année. Ces deux observations majeures nous permettent de confirmer que **notre série est additive**.

### Vérification de la stationnarité

Figure 7: Autocorrélogrammes



**PACF**

**ACF**

*Source : Réalisé par l’auteur*

Les graphiques ci-dessus représentent l’ACF (Autocorrelation Function) et le PACF (Partial Autocorrelation Function).

La représentation de l’ACF montre **une décroissance géométrique lente vers zéro** (0) prouvant la stationnarité de la série, résultat infirmé par le test de Dikey Fuller (ADF) avec une P-value de **0.26** et le PACF présente **deux pics significatifs** signifiant que la valeur à la période **t** est plus liée à la valeur aux périodes **t-1** et **t-2.**

*Source : Réalisé par l’auteur*

|  |  |  |
| --- | --- | --- |
| Test | P-value | Décision |
| Stationnarité (ADF) | 0.26 | P-value > 0,05 : On accepte H0 donc, Non-Stationnarité[[2]](#footnote-2) |
| Normalité (Shapiro) | 0.00051 | P-value < 0,05 : On rejette H0 donc, Non-Normalité |

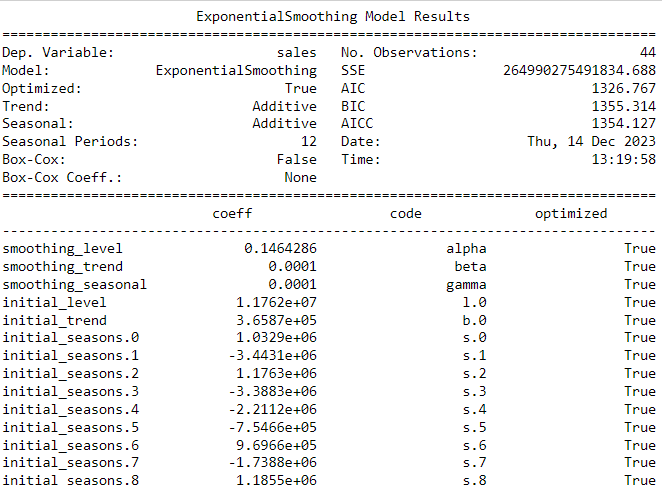
Tableau 3: Test de Stationnarité et normalité

Au total, notre série est **additive avec tendance et saisonnalité**, nos données sont non-stationnaires et non gaussiennes. Nous pouvons maintenant passer à la prévision.

## Prévision des ventes d’articles

### Construction du modèle

Dans le but de construire notre modèle, nous avons séparé le jeu de données en deux séries, une série d’entrainement du modèle allant du 2013-01-31 au 2016-08-31 et une série de test partant du 2016-09-30 au 2017-08-31.

Figure 8: Résultat du modèle Holt-Winters

*Source : Réalisé par l’auteur avec Python*

Tableau 4: Coefficients du modèle Holt-Winters

*Source : Réalisé par l’auteur*

|  |  |
| --- | --- |
| Coefficients | Valeurs |
| Alpha | **0.15** |
| Beta | **0.0001** |
| Gamma | **0.0001** |

Les valeurs de alpha, bêta et gamma sont petit ce qui met à nu **l’importance du passé lointain** pour notre prévision. Le modèle mise donc beaucoup plus sur la stabilité.

*Source : Réalisé par l’auteur avec Python*

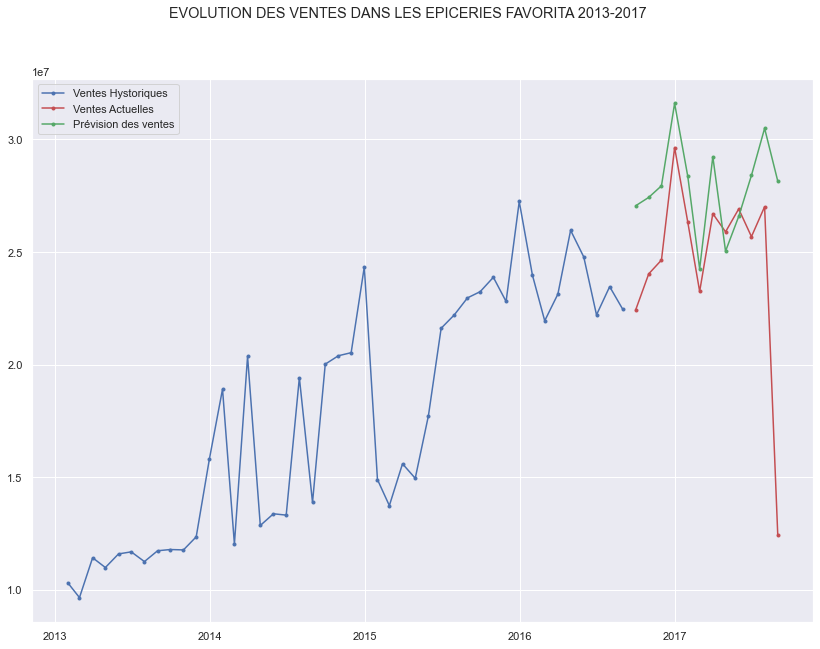


Figure 9: Entrainement et test du modèle

La représentation ci-dessus nous permet de visualiser l’évolution des ventes de 2013 à fin 2016 pour les données d’entrainement (Bleu) et de fin 2016 à fin 2017 pour les données de test (Rouge) et de prévision (Vert). La vérification du modèle se basera sur l’analyse des résidus.

### Validation du modèle

Il s’agit de vérifier si les résidus du modèle choisi suivent un processus bruit blanc. La satisfaction de la seule condition de bruit blanc suffit pour valider le modèle. Mais si les résidus suivent en plus un processus gaussien (loi normal), le modèle serait encore plus intéressant. (Akposso 2022) .

|  |  |  |
| --- | --- | --- |
| Test | P-value | Décision |
| Bruit Blanc (Ljung-Box) | 0.18 | P-value > 0,05, **On accepte H0**: Bruit Blanc |
| Normalité (Shapiro) | 0.75 | P-value > 0,05, **On accepte H0**: Normalité |
| Stationnarité (ADF) | 0.0026 | P-value < 0.0026, **On rejette H0** : Stationnarité |

Tableau 5: Résultat des tests

*Source : Réalisé par l’auteur*

Les résultats ci-dessus nous permettent de conclure la fiabilité de notre modèle de Holt-Winters car nos résidus sont stationnaires et Bruit-blanc Gaussien. Ces résultats sont très satisfaisants et nous rassure concernant les résultats de notre prévision.

Figure 10: Graphiques des tests

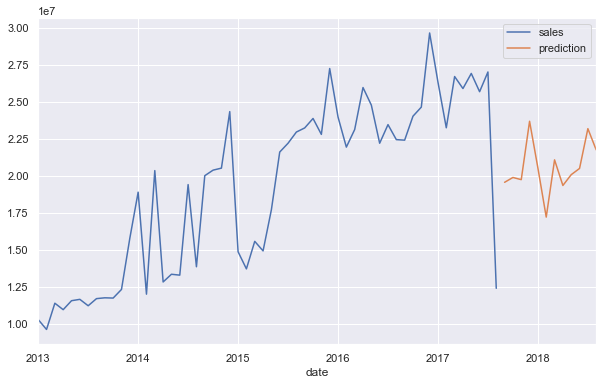
*Source : Réalisé par l’auteur avec Python*

### Prédiction des ventes pour les 12 prochains mois.

Le modèle construit précédemment, nous permet de faire la prévision des quantités d’articles vendus pour les 12 prochains mois (Voir Tableau 6 et la figure 11).

|  |  |
| --- | --- |
| Date | Valeurs Prédites |
| 2017-09-30 | 1.957647e+07 |
| 2017-10-31 | 1.990203e+07 |
| 2017-11-30 | 1.975662e+07 |
| 2017-12-31 | 2.369007e+07 |
| 2018-01-31 | 2.053817e+07 |
| 2018-02-28 | 1.723199e+07 |
| 2018-03-31 | 2.108896e+07 |
| 2018-04-30 | 1.936378e+07 |
| 2018-05-31 | 2.009074e+07 |
| 2018-06-30 | 2.051047e+07 |
| 2018-07-31 | 2.319194e+07 |
| 2018-08-31 | 2.176250e+07 |

Tableau 6: Tableau des valeurs prédites

Figure 11: Prédiction des 12 prochains mois

*Source : Réalisé par l’auteur avec Python*

# Conclusion

Au terme de cette partie, nous retenons que les données soumis à notre analyse sont additives avec tendance et saisonnalité.

La qualité du modèle de Holt Winters a été confirmé par nos tests sur les résidus qui sont **stationnaires, bruit blanc et gaussien**. Le modèle a donc été utilisé pour faire la prévision sur les 12 prochains mois, du 30-09-2017 au 31-08-2018.

# Conclusion générale

Notre étude visait à élaborer une technique robuste de prévision des ventes des épiceries Favorita en Équateur grâce à la méthode de Holt-Winters. La préoccupation majeure était de déterminer si ce modèle pouvait efficacement capturer la tendance à la hausse des ventes et la saisonnalité observée avec pour but ultime de fournir des prévisions fiables afin d'éclairer la gestion des stocks et de soutenir les décisions stratégiques de l'entreprise.

L’analyse des données historiques des ventes de 2013 à 2017 a révélé **une croissance constante des quantités d'articles vendues au fil des années, avec des performances particulièrement fortes en décembre**, où les ventes atteignent environ 25 millions d'unités. **La distribution des ventes présente une forme irrégulière**, indiquant une tendance à vendre en grande quantité. **La méthode de Holt-Winters a été confirmée comme un choix judicieux**, avec des tests de résidus démontrant la robustesse du modèle, ses résidus étant stationnaires, présentant un bruit blanc et suivant une distribution gaussienne.

Sur la base de ces résultats, nous recommandons **l'adoption continue de la méthode de Holt-Winters pour la prévision des ventes**. Il est conseillé de mettre en œuvre les prévisions générées pour optimiser la gestion des stocks, planifier les opérations et améliorer la performance commerciale, en particulier pendant les périodes de pointe comme le mois de décembre. De plus, il serait bénéfique de surveiller de près les facteurs externes susceptibles d'influencer les ventes, tels que les tendances du marché et les changements saisonniers.

Cependant, il est important de noter que toute analyse comporte des limites. Dans notre cas, les résultats sont basés sur des données historiques et supposent une continuité des tendances passées, ce qui peut ne pas toujours être le cas. Les facteurs externes imprévus peuvent également influencer les ventes de manière significative. De plus, la méthode de Holt-Winters suppose une stabilité des paramètres, ce qui peut nécessiter une réévaluation régulière.

Alors, pour améliorer les résultats, des analyses plus approfondies pourraient être envisagées, telles que l'intégration de variables externes (comme les promotions ou les événements spéciaux) dans le modèle. Une évaluation continue de la performance du modèle, avec des ajustements si nécessaire, serait également recommandée. Des analyses de sensibilité pour évaluer l'impact de différentes hypothèses sur les résultats pourraient fournir une compréhension plus approfondie des prévisions générées par le modèle. Enfin, des comparaisons avec d'autres méthodes de prévision pourraient être effectuées pour évaluer la pertinence continue de la méthode de Holt-Winters dans un contexte évolutif.

# Annexe

|  |
| --- |
| **Importation des packages** |
| import numpy as np  import pandas as pd  import matplotlib.pyplot as plt  import seaborn as sns  from statsmodels.tsa.holtwinters import ExponentialSmoothing as HWES  import datetime  from sklearn.metrics import mean\_squared\_error  from statsmodels.tsa.seasonal import seasonal\_decompose, STL  from statsmodels.graphics.tsaplots import plot\_acf, plot\_pacf  from statsmodels.tsa.stattools import adfuller  from scipy.stats import shapiro  from scipy.stats.mstats import normaltest  import statsmodels.api as sm  import statsmodels.tsa.api as smt |
| **Importation et Organisation des données par mois** |
| # Importation des donnees  favorita = pd.read\_csv("epicerie.csv", index\_col='date', parse\_dates=True)  df = favorita.copy()  # Organisation par mois  df = df.resample('M').sum()  # Comptage des NaN  df.isna().sum()  # Détecter les doublons  len(df[df.duplicated()])  # Analyse des ventes  sales = df['sales']  sales.describe() |
| **Prétraitement des données** |
| # Valeurs Aberrantes  sns.set(rc={'figure.figsize':(14,10)})  sns.boxplot(x=df.index.month\_name(), y='sales', data=df, palette='muted')  plt.ylabel('Ventes')  plt.xlabel('Mois')  plt.title("BOITES A MOUSTACHE DES QUANTITES VENDUES PAR MOIS")  plt.savefig("BOITES A MOUSTACHE.png", dpi=300)  # Détecter les doublons  len(df[df.duplicated()]) |
| **Visualisation** |
| sns.set\_style('dark')  plt.figure(figsize=(10, 7))  sns.distplot(a=sales, bins=10, color='blue',hist\_kws={"edgecolor": 'white'})  plt.xlabel('Ventes')  plt.title("REPARTITION DES QUANTITES VENDUES")  # Representation des ventes  sns.set\_style("darkgrid")  plt.figure(figsize=(15, 5))  plt.subplot(111)  plt.plot(sales, marker = 'o')  plt.title("EVOLUTION DES VENTES PAR MOIS")  plt.xlabel("Dates")  plt.show()  composants1 = seasonal\_decompose(sales, model='additive')  composants1.plot();  plot\_acf(df.sales);  plot\_pacf(df.sales); |
| **Prevision: Construction du modele** |
| df\_train = sales.iloc[:-12]  df\_test = sales.iloc[-12:]  model = HWES(df\_train, seasonal\_periods=12, trend='add', seasonal='add')  fitted = model.fit()  print(fitted.summary())  sales\_forecast = fitted.forecast(steps=12)  print(df\_test, sales\_forecast)  fig = plt.figure()  fig.suptitle('EVOLUTION DES VENTES DANS LES EPICERIES FAVORITA 2013-2017 ')  past, = plt.plot(df\_train.index, df\_train, 'b.-', label='Ventes Hystoriques')  future, = plt.plot(df\_test.index, df\_test, 'r.-', label='Ventes Actuelles')  predicted\_future, = plt.plot(df\_test.index, sales\_forecast, 'g.-', label='Prévision des ventes')  plt.legend(handles=[past, future, predicted\_future])  plt.show() |
| **Test sur les residus** |
| from statsmodels.stats.diagnostic import acorr\_ljungbox  res=fitted.resid  # Test de Bruit-blanc  acorr\_ljungbox(res, return\_df=True)  # Tes de normalite  shapiro(res)  # stationarite  adf\_testRes = adfuller(res)  print(f'ADF Statistic: {adf\_testRes[0]}')  print(f'p-value: {adf\_testRes[1]}')  def ts\_plot(y, lags=None, title='', fig\_size=(14,12)):  """  Calcul de l'acf, pacf, de l'histogramme et du QQ-plot d'une série temp  """  # on transforme en Series si l'argument y n'en est pas une  if not isinstance(y, pd.Series):  y = pd.Series(y)    # initialisation de la figure et des axes  fig = plt.figure(figsize=fig\_size)  ts\_ax = fig.add\_subplot(311)  acf\_ax = fig.add\_subplot(323)  pacf\_ax = fig.add\_subplot(324)  qq\_ax = fig.add\_subplot(325)  hist\_ax = fig.add\_subplot(326)    # la serie temporelle  y.plot(ax=ts\_ax)  ts\_ax.set\_title(title);    # ACF et PACF  smt.graphics.plot\_acf(y, lags=lags, ax=acf\_ax, alpha=0.05)  smt.graphics.plot\_pacf(y, lags=lags, ax=pacf\_ax, alpha=0.05, method='ywm')    # QQ-plot  sm.qqplot(y, line='s', ax=qq\_ax)  qq\_ax.set\_title('QQ-Plot')    # histogramme  y.plot(ax=hist\_ax, kind='hist', bins=25);  hist\_ax.set\_title('Histogramme');  plt.tight\_layout();  plt.show() |
| **Modèle final** |
| modelFin = HWES(sales, seasonal\_periods=12, trend='add', seasonal='add')  fittedFin = modelFin.fit()  prevision = fittedFin.forecast(12)  sales.plot(legend=True, label='sales', figsize=(10,6))  prevision.plot(legend=True, label='prediction') |

# Bibliographie et Webographie

Akposso, Didier Martial. «PREDICTION DES SERIES TEMPORELLES.» *LISSAGE EXPONENTIEL SIMPLE-DOUBLE-HOLT WINTERS AVEC R.* 2022.

Köseoğlu, Buse. *Medium.* 15 07 2023. https://buse-koseoglu13.medium.com/guide-to-time-series-analysis-with-python-2-moving-average-process-784328325e5f (accès le 12 12, 2023).

Louis. *STATOSCOP.* 04 Mai 2021. https://blog.statoscop.fr/timeseries-1.html (accès le 12 12, 2023).

MAZEROLLE, Fabrice. *Statistique Descriptive.* Marseille: Gualino Editeur, 2006.

**Table des matières**

[Avant-Propos 1](#_Toc153576462)

[Sigles 2](#_Toc153576463)

[Liste des illustrations 3](#_Toc153576464)

[Listes des figures 3](#_Toc153576465)

[Liste des tableaux 3](#_Toc153576466)

[Sommaire 4](#_Toc153576467)

[Résumé 5](#_Toc153576468)

[Abstract 6](#_Toc153576469)

[Introduction Générale 9](#_Toc153576470)

[I. STATISTIQUE DESCRIPTIVE DES VENTES 11](#_Toc153576471)

[Introduction 11](#_Toc153576472)

[1. Prétraitement des données 11](#_Toc153576473)

[a) Présentation des données 11](#_Toc153576474)

[b) Nettoyage des données 12](#_Toc153576475)

[2. Analyse descriptive de l’évolution des ventes entre 2013 et 2017 13](#_Toc153576476)

[a) Analyse des graphiques et des indicateurs 13](#_Toc153576477)

[b) Interprétation 15](#_Toc153576478)

[Conclusion 15](#_Toc153576479)

[II. ANALYSE TEMPORELLE ET PREVISION DES VENTES 16](#_Toc153576480)

[Introduction 16](#_Toc153576481)

[1. Série temporelle 16](#_Toc153576482)

[a) Définition des concepts 16](#_Toc153576483)

[b) Le modele de prévision de Holt-Winters 17](#_Toc153576484)

[2. Conditions préliminaires 18](#_Toc153576485)

[a) Décomposition de la série 18](#_Toc153576486)

[b) Vérification de la stationnarité 18](#_Toc153576487)

[3. Prévision des ventes d’articles 20](#_Toc153576488)

[a) Construction du modèle 20](#_Toc153576489)

[b) Validation du modèle 21](#_Toc153576490)

[c) Prédiction des ventes pour les 12 prochains mois. 23](#_Toc153576491)

[Conclusion 24](#_Toc153576492)

[Conclusion générale 25](#_Toc153576493)

[Annexe 27](#_Toc153576494)

[Bibliographie et Webographie 30](#_Toc153576495)

1. Les séries chronologiques peuvent être enregistrées à intervalles d'une minute, d'une heure, d'un jour, d'un mois, etc. [↑](#footnote-ref-1)
2. Nos données doivent être stationnarisée dans nos prochaines études pour d’autres types modèles, entre autres : ARIMA, … [↑](#footnote-ref-2)