



ROYAUME DU MAROC  
HAUT COMMISSARIAT AU PLAN  
INSTITUT NATIONAL DE STATISTIQUE  
ET D'ÉCONOMIE APPLIQUÉE  
INSEA



*Sujet:*

# ANALYSE DES TENDANCES DE VENTE ET COMPORTEMENTS D'ACHAT

Réalisé par : EL MOTCHOU

Abdellah

# Table des matières

1	Introduction . . . . .	3
1.1	Contexte et Objectifs . . . . .	3
1.2	Présentation des Données Utilisées . . . . .	3
2	Préparation des Données . . . . .	5
2.1	Chargement et Nettoyage des Données . . . . .	5
2.2	Gestion des Valeurs Manquantes et des Erreurs . . . . .	5
3	Analyse Exploratoire (EDA) . . . . .	6
3.1	Statistiques Générales . . . . .	6
3.2	Visualisations Clés . . . . .	7
3.3	Corrélation entre les Variables . . . . .	7
3.4	Interprétation des Résultats . . . . .	7
4	Tableau de Bord Interactif . . . . .	8
4.1	Fonctionnalités du Dashboard Streamlit . . . . .	8
4.2	Explication des Visualisations . . . . .	9
5	Insights et Recommandations . . . . .	9
5.1	Principales Observations . . . . .	9
5.2	Principales Observations . . . . .	10
6	Conclusion . . . . .	10

# 1 Introduction

## 1.1 Contexte et Objectifs

L'objectif de ce projet est d'analyser les tendances de vente et les comportements d'achat à travers un tableau de bord interactif. Grâce à une exploration approfondie des données, il vise à :

- Comprendre l'évolution des ventes sur différentes périodes.
- Identifier les produits les plus et les moins vendus.
- Analyser le comportement des clients (clients fidèles, clients les plus rentables).
- Étudier l'impact du délai de livraison sur les ventes.
- Visualiser les performances des ventes par région et par catégorie de produit.
- Permettre une exploration dynamique des données avec des filtres interactifs.

Ce projet aide à prendre des décisions stratégiques en optimisant les ventes et en améliorant l'expérience client.

## 1.2 Présentation des Données Utilisées

La base de données utilisée dans ce projet contient des informations détaillées sur les ventes de produits dans différentes régions. Chaque ligne représente une transaction individuelle et comprend plusieurs attributs permettant d'analyser les tendances des ventes, le comportement des clients et l'impact de divers facteurs sur les performances commerciales.

Les principales colonnes de la base de données sont :

- **Identifiants des Transactions :**
  - **Row ID** : Identifiant unique de chaque ligne.
  - **Order ID** : Identifiant unique de la commande, permettant de regrouper plusieurs produits dans une même transaction.
- **Informations Temporelles :**
  - **Order Date** : Date à laquelle la commande a été passée.
  - **Ship Date** : Date d'expédition de la commande, utilisée pour calculer le délai de livraison.

- **Détails sur la Livraison :**
  - **Ship Mode** : Mode d'expédition choisi (Standard, Express, etc.), permettant d'étudier son impact sur le délai de livraison et la satisfaction client.
- **Informations Clients :**
  - **Customer ID** : Identifiant unique du client.
  - **Customer Name** : Nom du client.
  - **Segment** : Type de client (Consommateur, Entreprise, Marché domestique), utile pour segmenter les analyses.
- **Localisation Géographique :**
  - **Country** : Pays (ici, uniquement "United States").
  - **City** : Ville où la commande a été effectuée.
  - **State** : État correspondant à la ville.
  - **Postal Code** : Code postal de l'adresse de livraison.
  - **Region** : Région du client (Ouest, Est, Sud, Centre), permettant une analyse géographique des ventes.
- **Détails des Produits :**
  - **Product ID** : Identifiant unique du produit.
  - **Product Name** : Nom du produit vendu.
  - **Category** : Catégorie du produit (Mobilier, Technologie, Fournitures de bureau).
  - **Sub-Category** : Sous-catégorie du produit (ex : Chaises, Téléphones, Papier, etc.).
- **Informations Financières :**
  - **Sales** : Montant total de la vente pour le produit correspondant.

Cette base de données offre une **vision complète du processus de vente**, depuis la commande jusqu'à la livraison. Elle permet d'analyser l'évolution des ventes dans le temps, d'identifier les produits les plus rentables, d'évaluer la fidélité des clients et de comprendre l'influence des délais de livraison sur la satisfaction client

## 2 Préparation des Données

### 2.1 Chargement et Nettoyage des Données

L'étape de chargement et de nettoyage des données est essentielle pour garantir la fiabilité et la pertinence des analyses. Avant toute exploration, il est crucial de s'assurer que les données sont correctement formatées, cohérentes et exemptes d'anomalies susceptibles de fausser les résultats.

Les données ont d'abord été extraites d'un fichier CSV contenant les transactions commerciales. Une vérification de l'existence du fichier a été effectuée afin d'éviter toute erreur liée à une absence de source de données. Ensuite, une première exploration a permis d'identifier les types de données présents et de détecter d'éventuelles incohérences.

Les dates de commande et de livraison ont été converties dans un format standardisé afin de faciliter les calculs et l'analyse des tendances temporelles. Cette conversion a permis de corriger des erreurs de formatage et d'assurer une uniformité dans l'ensemble du dataset.

Une attention particulière a été portée à la gestion des doublons. Des vérifications ont été menées pour identifier et supprimer les entrées répétées afin d'éviter les biais statistiques. La suppression des doublons garantit que chaque transaction est comptabilisée une seule fois, ce qui renforce la précision des analyses.

Enfin, une nouvelle variable, appelée *temps de livraison*, a été créée en calculant la différence entre la date de commande et la date d'expédition. Cette variable est cruciale pour comprendre l'impact des délais de livraison sur les performances de vente et la satisfaction client.

Ainsi, toutes ces étapes de nettoyage et de structuration des données permettent de travailler sur un dataset fiable, optimisé et prêt pour l'analyse exploratoire et la construction du tableau de bord interactif.

### 2.2 Gestion des Valeurs Manquantes et des Erreurs

Les données ont été analysées afin de détecter et traiter les valeurs manquantes et les erreurs potentielles. La présence de valeurs manquantes peut poser des problèmes

d'interprétation et affecter la qualité des résultats obtenus.

Dans cette base de données, les valeurs manquantes étaient principalement présentes dans la colonne **Postal Code**. Afin d'éviter la perte d'informations précieuses, ces valeurs ont été remplacées par la moyenne des codes postaux disponibles, garantissant ainsi une cohérence géographique des analyses.

Les erreurs de format ont également été corrigées, notamment celles liées aux dates. Certaines entrées mal enregistrées ont été détectées et supprimées pour éviter toute incohérence dans l'analyse temporelle des ventes.

Enfin, une vérification a été réalisée sur les catégories de données afin d'assurer leur intégrité. Par exemple, les valeurs uniques des colonnes catégorielles, comme **Ship Mode**, **Category** et **Region**, ont été examinées pour détecter d'éventuelles incohérences ou erreurs de saisie.

Grâce à ces corrections et ajustements, la base de données est désormais prête pour une analyse approfondie, garantissant des résultats précis et exploitables.

L'analyse exploratoire des données (EDA) est une étape essentielle permettant de mieux comprendre la structure des données, d'identifier les tendances et de détecter d'éventuels problèmes ou anomalies. Cette phase repose sur des analyses statistiques et des visualisations graphiques.

## 3 Analyse Exploratoire (EDA)

### 3.1 Statistiques Générales

Une première analyse statistique a été réalisée afin de donner une vue d'ensemble des données. Parmi les principaux indicateurs calculés :

- Nombre total de transactions : **9 800**
- Valeur moyenne des ventes : **230.77 \$**
- Médiane des ventes : **X \$**
- Délai de livraison moyen : **X jours**
- Produits distincts vendus : **X produits**
- Nombre de clients uniques : **X clients**

Ces statistiques de base permettent de mieux cerner la distribution des ventes et d'identifier les premiers axes d'analyse.

## 3.2 Visualisations Clés

Les données ont été visualisées sous différents formats pour mettre en évidence les tendances et les comportements d'achat.

- **Évolution des ventes dans le temps** : L'analyse des ventes sur une période donnée a révélé des variations saisonnières, avec des pics de vente à certaines périodes spécifiques.
- **Répartition des ventes par région** : L'étude des ventes par localisation géographique a mis en évidence les régions les plus performantes en termes de chiffre d'affaires.
- **Produits les plus vendus** : Une analyse des produits a permis d'identifier ceux qui génèrent le plus de revenus ainsi que ceux qui se vendent le moins.
- **Impact du délai de livraison** : L'analyse des délais de livraison a montré qu'un temps d'expédition trop long pouvait influencer négativement les ventes et la satisfaction des clients.

## 3.3 Corrélation entre les Variables

Une matrice de corrélation a été utilisée pour identifier les relations entre les différentes variables du dataset. Cela a permis de comprendre quels facteurs influencent directement les ventes.

## 3.4 Interprétation des Résultats

L'analyse exploratoire a permis de mettre en évidence plusieurs points clés :

- Certains produits et catégories sont plus rentables que d'autres.
- Certaines régions affichent des volumes de ventes nettement supérieurs.
- Un délai de livraison trop long peut avoir un impact négatif sur les achats.
- Les clients récurrents génèrent une part importante du chiffre d'affaires.

Ces premiers résultats ont orienté la suite du projet en identifiant les aspects les plus pertinents à explorer dans le tableau de bord interactif.

## 4 Tableau de Bord Interactif

Le tableau de bord interactif développé avec Streamlit permet une exploration dynamique des données et une meilleure compréhension des tendances des ventes. Il offre une interface utilisateur intuitive, permettant aux utilisateurs de filtrer et d'analyser les ventes selon divers critères.

### 4.1 Fonctionnalités du Dashboard Streamlit

Le tableau de bord inclut plusieurs fonctionnalités essentielles pour l'analyse des ventes :

- **Filtres dynamiques** : L'utilisateur peut filtrer les données selon plusieurs critères, notamment la catégorie de produit, la région, le client et la période d'analyse.
- **Affichage des indicateurs clés** : Des métriques importantes, telles que le chiffre d'affaires total, le nombre de transactions et le délai de livraison moyen, sont mises en avant pour une lecture rapide des performances commerciales.
- **Graphiques interactifs** : Des visualisations dynamiques permettent d'explorer les tendances des ventes et d'identifier les segments les plus performants.
- **Analyse géographique** : Une carte des ventes par région est intégrée pour visualiser la répartition des performances commerciales à travers les différentes zones géographiques.
- **Analyse des clients** : Un classement des meilleurs clients en fonction du chiffre d'affaires généré est affiché pour mieux comprendre leur impact sur les ventes.
- **Étude de l'impact du délai de livraison** : Une analyse des délais de livraison permet d'évaluer leur influence sur les ventes et la satisfaction client.



## 4.2 Explication des Visualisations

Chaque graphique du tableau de bord a été conçu pour répondre à des questions spécifiques et fournir des insights exploitables :

- **Courbe d'évolution des ventes** : Permet d'analyser la tendance des ventes au fil du temps et d'identifier les périodes de forte activité.
- **Histogramme des ventes par catégorie de produit** : Met en évidence les catégories les plus rentables et celles qui génèrent moins de revenus.
- **Carte des ventes par région** : Illustre la répartition géographique des ventes, aidant ainsi à cibler les zones les plus performantes.
- **Graphique des délais de livraison** : Analyse l'impact du délai de livraison sur le volume des ventes et permet d'optimiser la logistique.
- **Classement des clients** : Présente les clients qui réalisent le plus d'achats, ce qui permet d'identifier les clients fidèles et rentables.

Ce tableau de bord interactif offre ainsi une visualisation efficace des performances de vente et facilite la prise de décision en mettant en évidence les tendances et les axes d'optimisation potentiels.

## 5 Insights et Recommandations

L'analyse des ventes a permis de mettre en évidence plusieurs tendances et points d'amélioration. Cette section résume les principales observations ainsi que les actions recommandées pour optimiser les performances commerciales.

### 5.1 Principales Observations

- **Tendances des ventes** : Les ventes connaissent des fluctuations saisonnières avec des pics à certaines périodes de l'année.
- **Régions les plus performantes** : Certaines régions génèrent un chiffre d'affaires nettement supérieur, tandis que d'autres restent sous-exploitées.
- **Produits les plus rentables** : Certaines catégories de produits contribuent fortement aux revenus, alors que d'autres ont un faible impact.

- **Impact du délai de livraison** : Un temps de livraison trop long peut entraîner une baisse des ventes et une insatisfaction des clients.
- **Clients fidèles et rentables** : Un petit nombre de clients réguliers génère une grande partie des ventes.

## 5.2 Principales Observations

- **Optimiser la gestion des stocks** : S'assurer d'un bon niveau de stock pour les produits les plus demandés afin d'éviter les ruptures.
- **Réduire les délais de livraison** : Améliorer la logistique et collaborer avec des services de livraison plus rapides pour éviter les pertes de ventes.
- **Cibler les régions sous-performantes** : Mettre en place des campagnes marketing adaptées pour stimuler les ventes dans les zones moins actives.
- **Miser sur les produits à forte rentabilité** : Renforcer la promotion des produits les plus vendus et revoir la stratégie pour ceux qui se vendent peu.
- **Fidéliser les clients les plus rentables** : Offrir des réductions ou des avantages aux clients réguliers pour les inciter à acheter davantage.

Ces recommandations permettront d'optimiser la performance commerciale et d'améliorer la satisfaction des clients tout en maximisant les revenus.

# 6 Conclusion

Cette étude sur l'analyse des ventes a permis d'identifier plusieurs tendances majeures influençant la performance commerciale. L'analyse temporelle a révélé des fluctuations saisonnières marquées, avec des périodes de forte activité et des creux nécessitant une stratégie d'optimisation des stocks et des promotions ciblées. L'étude des produits a montré que certaines catégories, notamment la technologie et le mobilier, génèrent un chiffre d'affaires significatif, tandis que d'autres ont une rentabilité plus faible et pourraient bénéficier d'une révision des stratégies de commercialisation. L'impact du délai de livraison a également été mis en évidence : un temps d'expédition trop long peut entraîner une baisse des ventes et une insatisfaction client, ce qui souligne l'importance d'une amélioration logistique. D'un point de vue géographique, certaines

régions se démarquent par un volume de ventes élevé, tandis que d'autres restent sous-exploitées, ce qui suggère l'opportunité de développer des campagnes marketing plus ciblées. Par ailleurs, l'analyse du comportement des clients a révélé qu'un petit groupe de clients fidèles contribue de manière significative au chiffre d'affaires, mettant en avant l'intérêt de renforcer la fidélisation par des offres personnalisées. Afin d'optimiser ces performances, il est recommandé d'améliorer la gestion des stocks, de réduire les délais de livraison, de cibler les régions à faible performance, de promouvoir les produits les plus rentables et de renforcer l'expérience client. Ces actions permettront d'accroître les ventes, d'améliorer la satisfaction des clients et de renforcer la compétitivité de l'entreprise sur le marché.