|  |
| --- |
| Projet personnel à but formateur et exploratoire |
| Rapport d’Analyse de Données |
| *Exploration, traitement et visualisation de données* |

|  |
| --- |
| Sylvain Vasseur  08/04/2025 |

Table des matières

[**I.** **Introduction** 2](#_Toc195022679)

[A. Contexte 2](#_Toc195022680)

[B. Objectif 2](#_Toc195022681)

[C. Présentation des données 2](#_Toc195022682)

[**Identifiants** 2](#_Toc195022683)

[**Informations sur les commandes** 2](#_Toc195022684)

[**Informations sur le client** 2](#_Toc195022685)

[**Informations sur les produits** 3](#_Toc195022686)

[**II.** **Méthodologie** 3](#_Toc195022687)

[**III.** **Analyse** 3](#_Toc195022688)

[**IV.** **Discussion** 3](#_Toc195022689)

[**V.** **Conclusion** 4](#_Toc195022690)

1. **Introduction**

## Contexte

L’entreprise étudiée évolue dans un secteur très concurrentiel, où la demande des consommateurs ne cesse d'augmenter. Face à un environnement dynamique, cette entreprise cherche à mieux comprendre les facteurs clés influençant ses ventes et ses profits. Plus précisément, elle souhaite analyser quels produits, régions, catégories et segments de clients sont les plus rentables, afin de cibler ceux qui offrent le meilleur potentiel tout en évitant les segments moins performants. L’enjeu de cette analyse est de fournir des recommandations stratégiques permettant à l’entreprise de se différencier et d'optimiser ses résultats dans un marché en perpétuelle évolution.

## Objectif

L'objectif principal de cette analyse est d'identifier les facteurs clés influençant les performances de l’entreprise dans un marchédynamique. Pour ce faire, nous chercherons à analyser les relations entre différentes variables, qu'elles soient **qualitatives** (comme les catégories de produits) ou **quantitatives** (comme les ventes ou les profits). Cette approche complète vise à fournir des recommandations stratégiques pour aider l’entreprise à optimiser ses choix en termes de produits, régions, catégories et segments de clients à cibler, tout en évitant les éléments moins rentables.

## Présentation des données

Le dataset utilisé contient plusieurs variables qui permettent d'analyser les performances de l’entreprise. Voici un aperçu de ces variables :

### **Identifiants**

* **Row ID** (Numérique) : Identifiant unique pour chaque ligne
* **Order ID** (Numérique) : Identifiant unique de chaque commande
* **Customer ID** (Numérique) : Identifiant unique pour chaque client
* **Product ID** (Numérique) : Identifiant unique du produit

### **Informations sur les commandes**

* **Order Date** (Date) : Date de la commande du produit
* **Ship Date** (Date) : Date d'expédition du produit
* **Ship Mode** (Catégoriel) : Mode de livraison spécifié par le client
* **Quantity** (Numérique) : Quantité du produit commandé
* **Discount** (Numérique) : Remise appliquée sur le produit
* **Sales** (Numérique) : Montant des ventes du produit
* **Profit** (Numérique) : Profit ou perte réalisée sur la commande

### **Informations sur le client**

* **Customer Name** (Textuel) : Nom du client
* **Segment** (Catégoriel) : Segment auquel appartient le client
* **Country** (Textuel) : Pays de résidence du client
* **City** (Textuel) : Ville de résidence du client
* **State** (Textuel) : État de résidence du client
* **Postal Code** (Numérique) : Code postal de chaque client
* **Region** (Catégoriel) : Région où le client réside

### **Informations sur les produits**

* **Category** (Catégoriel) : Catégorie du produit commandé
* **Sub-Category** (Catégoriel) : Sous-catégorie du produit commandé
* **Product Name** (Textuel) : Nom du produit

1. **Méthodologie**
   1. **Préparation des données** : Nettoyage, traitement des valeurs manquantes, création de variables.

Avant de procéder à l’analyse, une étape de vérification de la qualité des données a été réalisée. Aucune valeur manquante n’a été détectée dans le jeu de données, ce qui a permis d’éviter les opérations de traitement ou d’imputation classiques. Le seul ajustement nécessaire a concerné le type des variables de dates (Order Date et Ship Date), qui ont été converties au format datetime afin de pouvoir effectuer d’éventuelles analyses temporelles. À ce stade, la création de nouvelles variables n’a pas été jugée nécessaire, les données disponibles étant déjà suffisamment structurées pour répondre aux objectifs de l’analyse.

* 1. **Démarche analytique** : Expliquer l’ordre logique des étapes choisies (ex. : pourquoi commencer par l’exploration, puis corrélation, puis segmentation, etc.).

Afin de répondre à la problématique posée — comprendre les interactions entre les produits, les performances régionales et les marges —, l’analyse a été structurée en plusieurs étapes logiques. L’objectif est de construire une vision d’ensemble cohérente, tout en identifiant les leviers d’optimisation les plus pertinents.

La première phase consiste à explorer la structure globale des ventes. Il s'agit d'observer la répartition des catégories de produits, ainsi que leur part dans les ventes et le profit total. Cela permet d’obtenir un premier aperçu de la distribution des produits dans l’offre globale, et de repérer quels types de produits dominent en volume ou en poids financier. À ce stade, la vision reste relativement générale, notamment concernant le profit, car les catégories sont larges. L'objectif est donc avant tout de poser un cadre global avant d'affiner l’analyse au niveau des sous-catégories, où les premières différences marquées de rentabilité pourraient apparaître.

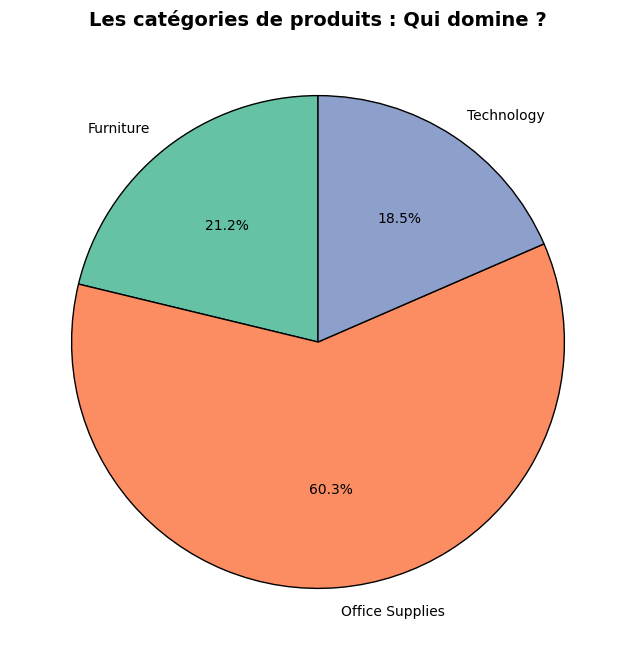
L’étape suivante vise à intégrer la dimension géographique, en croisant les résultats précédents avec les différentes régions. L’objectif ici est de comprendre si certaines zones sont plus productives ou plus rentables que d’autres, à catégorie ou sous-catégorie équivalente. Cette mise en relation offre un premier éclairage sur la productivité régionale, et permet de repérer d’éventuels axes d’optimisation : une région peut, par exemple, vendre beaucoup de produits peu rentables, ou au contraire performer sur des niches rentables. Ces croisements entre production et profit, à la fois par produit et par région, fournissent une base solide pour identifier les déséquilibres à corriger ou les stratégies gagnantes à répliquer.

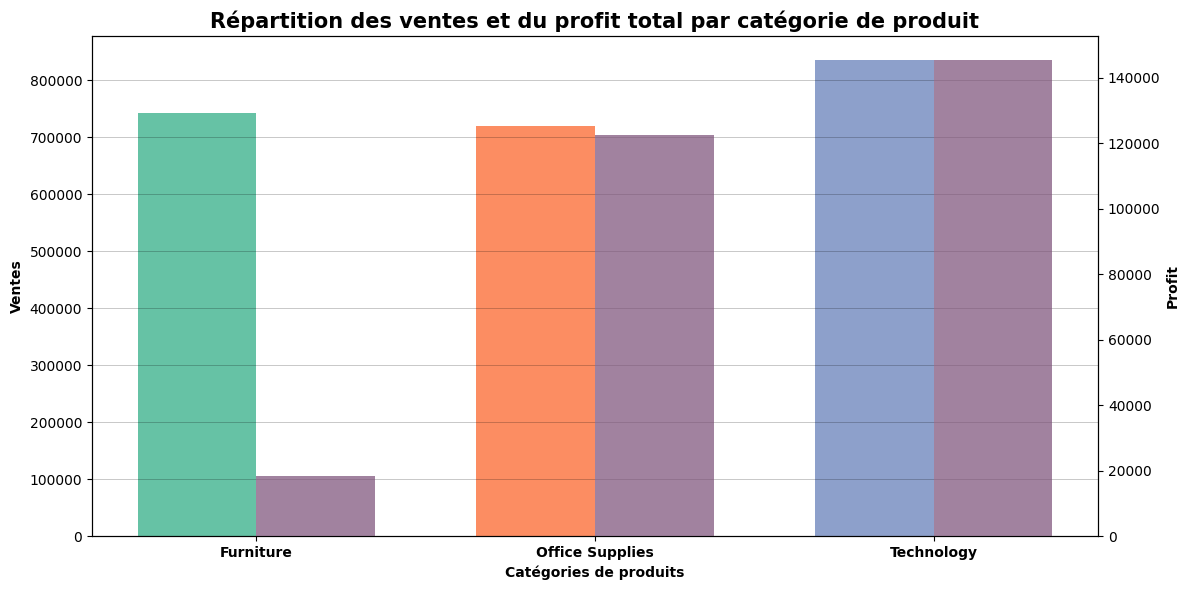
Une fois ce socle construit, certaines variables complémentaires sont analysées dans un but exploratoire. C’est le cas notamment de la variable "Segment", qui permet de répartir les clients selon leur profil : particuliers, entreprises ou télétravailleurs. Bien que cette dimension ne soit pas centrale dans l’analyse, elle peut offrir des pistes intéressantes, notamment si certains segments freinent la rentabilité de produits a priori performants. Cette analyse est donc intégrée de manière secondaire, uniquement si elle apporte une information exploitable.

Enfin, des facteurs plus opérationnels, comme le mode de livraison ou les remises accordées, sont étudiés à titre indicatif. L’idée ici n’est pas de multiplier les visualisations trop détaillées, mais plutôt d’identifier d’éventuelles anomalies ou pistes d’amélioration concrètes. Si un mode de livraison affecte fortement la rentabilité, ou si certaines remises sont mal calibrées, cela peut justifier une révision des pratiques logistiques ou commerciales. De même, un bref regard sur la dimension temporelle peut permettre de détecter des cycles ou des pics d’activité propices à une meilleure allocation des remises.

Cette approche progressive, du général vers le particulier, permet de ne pas se perdre dans les détails dès le départ, tout en gardant la possibilité de creuser des points spécifiques lorsque cela semble pertinent. Elle répond à un double objectif : établir une vision globale pour orienter les décisions stratégiques, tout en identifiant des leviers opérationnels concrets et actionnables.

1. **Analyse**
   1. **Exploration de la Structure Globale des Ventes**

**** Répartition des catégories de produits (barres empilées ou camembert)

Tout d'abord, conformément à la méthodologie établie précédemment, nous commençons par analyser les détails des catégories. On constate une répartition inégale des commandes, avec les **fournitures de bureau (Office Supplies)** représentant plus de 60 % des commandes totales. Les deux autres catégories se partagent les 40 % restants, avec un léger avantage pour les **fournitures (Furniture)** à 21,2 %, contre **la technologie (Technology)** qui enregistre 18,5 %.

Pour la suite, nous nous intéressons aux détails des commandes, notamment la quantité commandée et le profit généré. On remarque que la quantité de produits commandés est relativement homogène entre les catégories, chacune dépassant les 700 000 unités. La catégorie **Technology** se distingue toutefois en enregistrant plus de 800 000 ventes totales.

Cette observation est intéressante, car elle contraste avec le graphique précédent, où l'on a vu que **les fournitures de bureau (Office Supplies)** représentaient plus de 60 % des commandes totales. En dépit de cette forte proportion, la quantité de produits vendus ne se différencie pas de manière significative entre les catégories, ce qui suggère que les commandes sont largement réparties parmi les différentes catégories, même si l'une d'entre elles domine en termes de volume total.

Pour ce qui est du profit, il est intéressant de noter que bien que les **fournitures de bureau** **(Office Supplies)** représentent une grande part des ventes, elles ne génèrent pas proportionnellement autant de profit. Cela suggère peut-être une structure de coûts plus élevée ou des marges bénéficiaires plus faibles dans cette catégorie.

* 1. **Analyse Géographique de la Performance**

 Performance régionale par catégorie de produit (carte géographique ou barres empilées)

 Performance par région, comparée à la production et au profit (barres horizontales)

* 1. **Analyse des Segments Clients**

 Profit par segment de client (barres)

 Part des sous-catégories par segment (barres empilées ou camembert)

* 1. **Étude des Facteurs Opérationnels et Temporels**

 Impact des modes de livraison sur la rentabilité (barres ou boîte à moustaches)

 Impact des remises sur la rentabilité des sous-catégories (lignes ou barres)

 Cycles d’activité et pics de vente ou de profit (lignes ou histogramme)

**Répartition des catégories** de produits (barres)

**Sous-catégories de produits** (barres empilées ou camembert)

**Performance par région** (carte géographique ou barres)

**Segments clients et rentabilité** (barres ou camembert)

**Croisement segments clients / sous-catégories** (barres empilées ou heatmap)

**Impact des modes de livraison** sur la rentabilité (barres)

**Analyse des remises** (barres ou lignes)

**Analyse temporelle** des ventes/remises (lignes)

**Identification d'anomalies** ou opportunités (graphique de tendances)

1. **Discussion**
   1. **Limites** : Ce qui a pu freiner ou restreindre l’analyse (données manquantes, périmètre, temps, etc.).
   2. **Recommandations** : Ce que l’analyse permet de proposer concrètement comme actions ou stratégies.
2. **Conclusion**
   1. **Résumé** : Retour sur les principaux résultats.
   2. **Pistes futures** : Suggestions pour aller plus loin, approfondir ou prolonger le travail.