

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
A. Contexte.....	2
B. Objectif	2
C. Présentation des données	2
Identifiants	2
Informations sur les commandes.....	2
Informations sur le client	2
Informations sur les produits.....	3
II. Méthodologie.....	3
A. Préparation des données.....	3
B. Démarche analytique.....	3
III. Analyse	4
A. Exploration de la Structure Globale des Ventes	4
B. Analyse des Sous-Catégories et Écarts Régionaux.....	6
C. Analyse des Segments Clients.....	8
D. Étude des Facteurs Opérationnels et Temporels	8
IV. Discussion.....	10
A. Limites.....	10
B. Recommandations	10
V. Conclusion	10
A. Résumé :	10
B. Pistes futures	10

I. Introduction

A. 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.

B. Objectif

L'objectif principal de cette analyse est d'identifier les facteurs clés influençant les performances de l'entreprise dans un marché dynamique. 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.

C. 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

II. Méthodologie

A. Préparation des données

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.

B. Démarche analytique

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

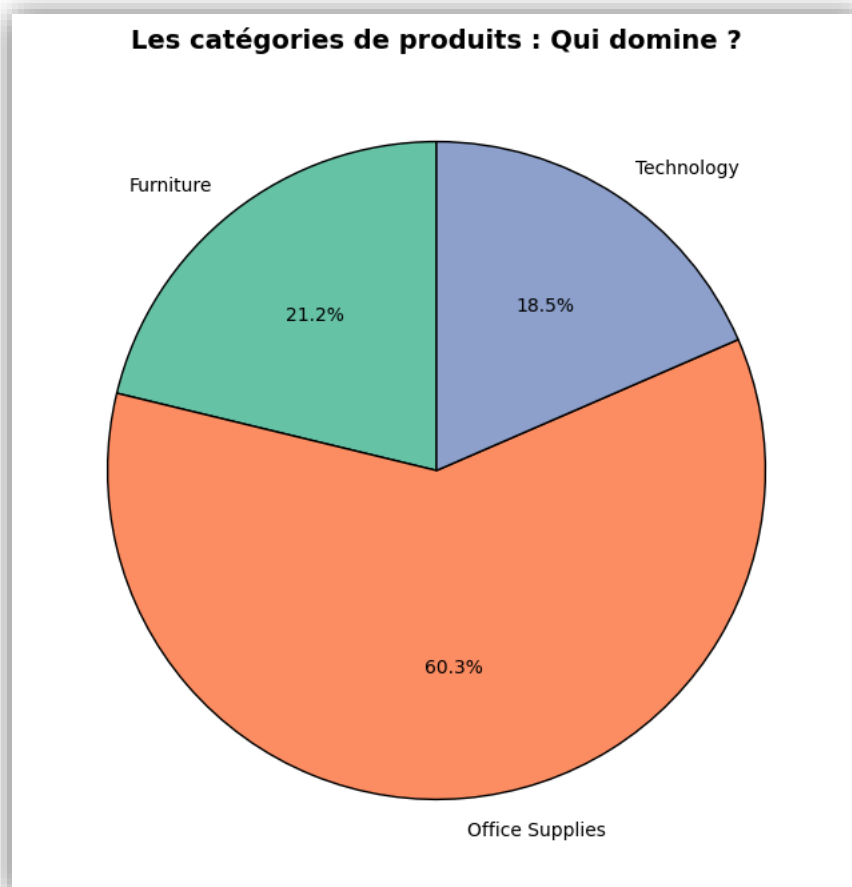
dans l'analyse, elle peut offrir des pistes intéressantes, notamment si certains segments freinent la rentabilité de produits a priori performants. Cette analyse est donc intégrée de manière secondaire, uniquement si elle apporte une information exploitable.

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.

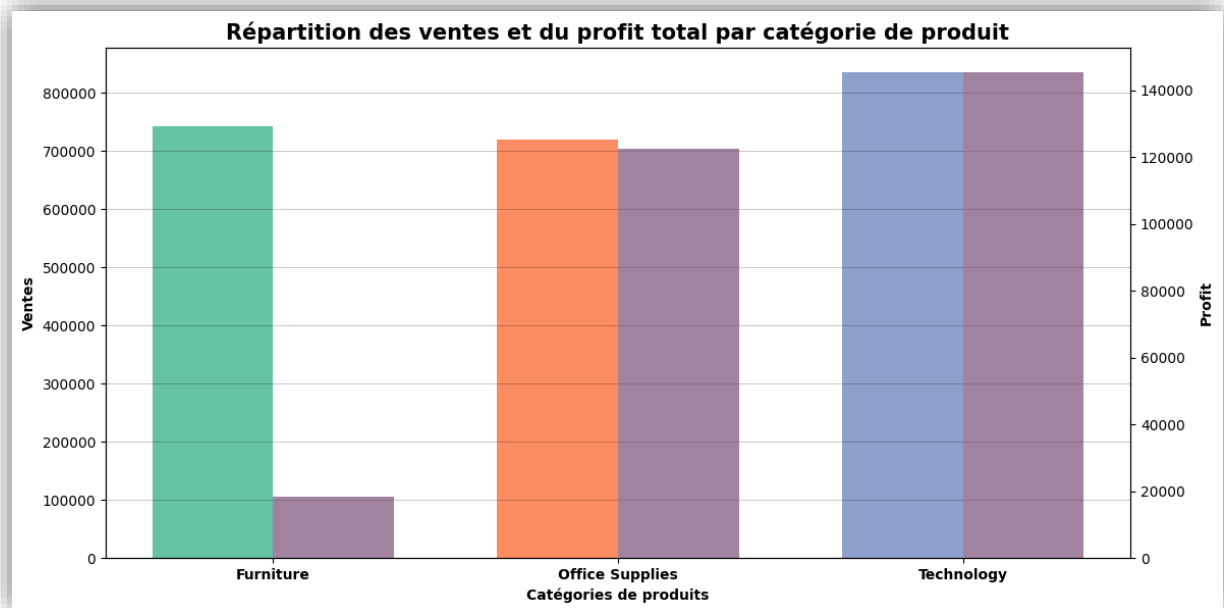
III. Analyse

A. Exploration de la Structure Globale des Ventes



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 %.

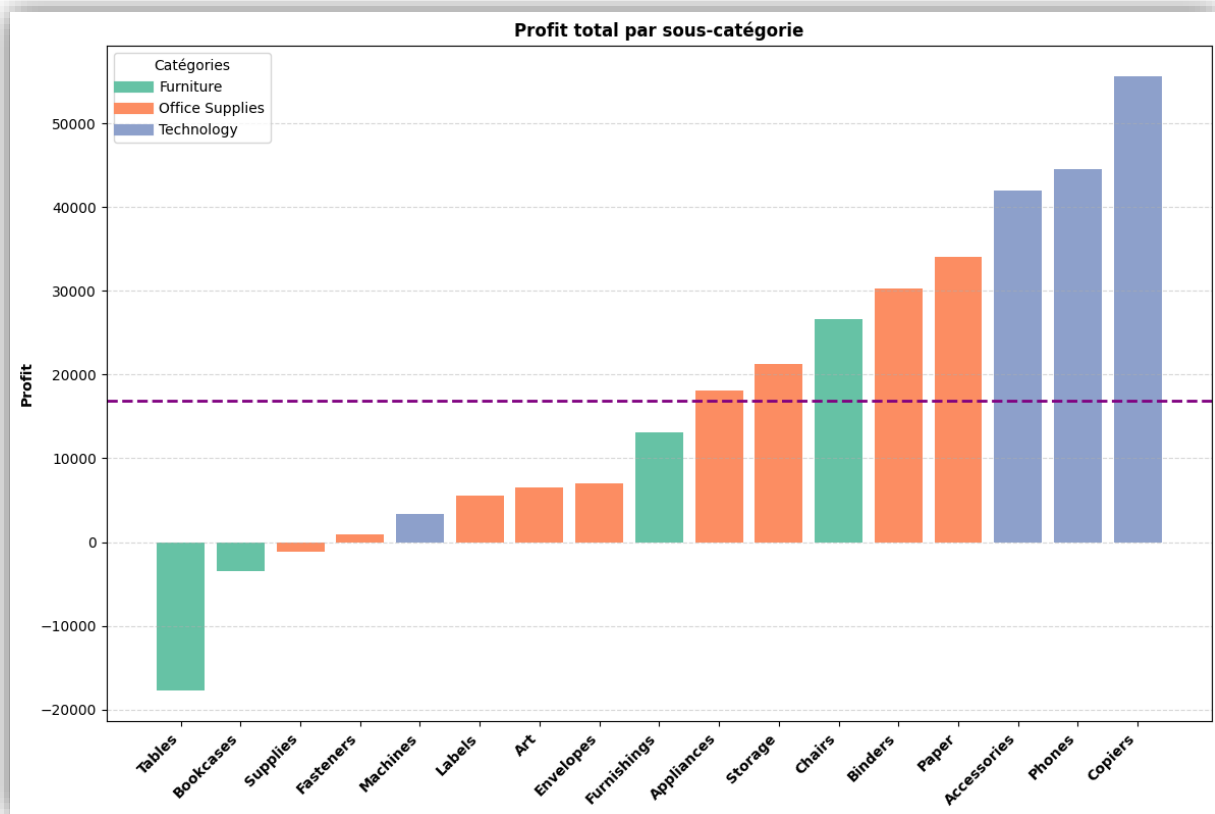


Pour la suite, nous nous intéressons aux détails des commandes, notamment la quantité commandée et le profit généré. On remarque que la quantité de produits commandés est relativement homogène entre les catégories, chacune dépassant les 700 000 unités. La catégorie **Technology** se distingue toutefois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

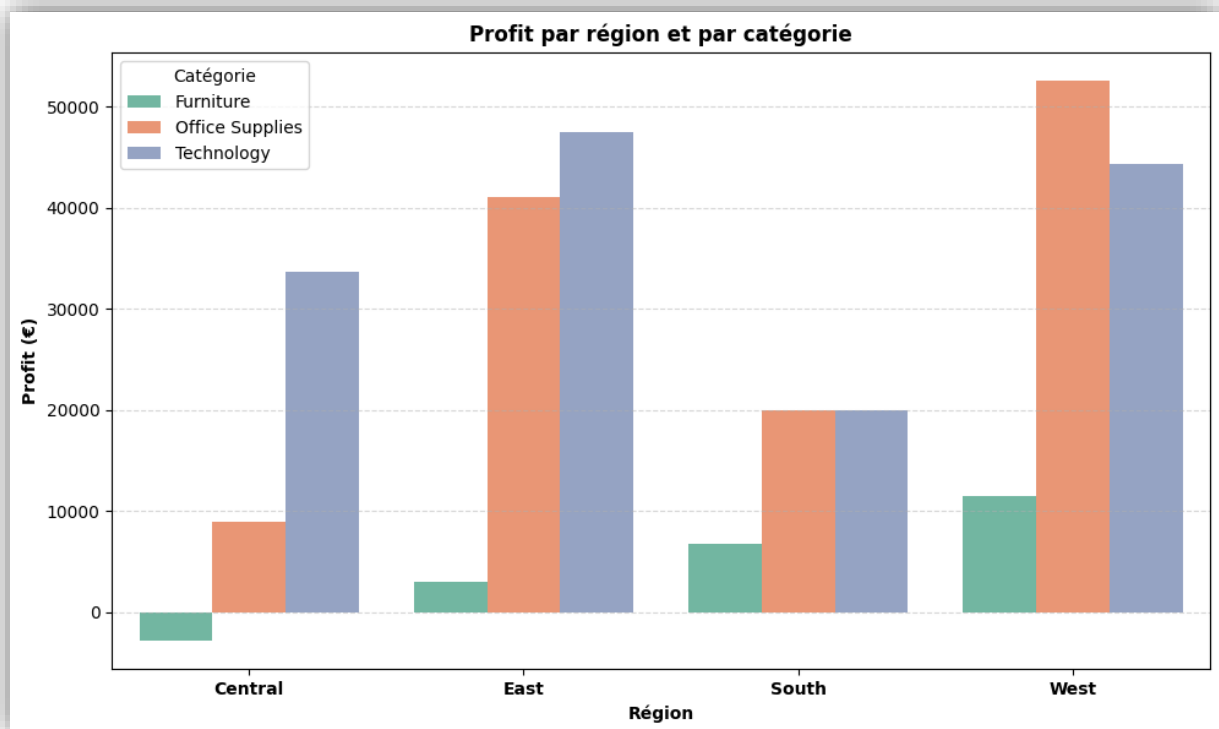
B. Analyse des Sous-Catégories et Écarts Régionaux



Les sous-catégories génèrent des profits très variés. Certaines, comme les **Tables**, figurent parmi les plus faibles, tandis que d'autres, telles que les **Accessories**, **Phones**, et **Copiers**, dominent avec des profits nettement supérieurs à 40 000€. La majorité des sous-catégories montrent des résultats modérés, avec des bénéfices allant de 15 000€ à 35 000€, ce qui suggère une distribution assez inégale du profit. Quelques sous-catégories comme **Bookcases** et **Suppliers** présentent même des profits négatifs. Cette dispersion des résultats met en lumière des opportunités d'optimisation dans les sous-catégories les moins performantes.

Malgré la présence de plusieurs sous-catégories peu rentables, voire déficitaires, la moyenne des profits se situe juste en dessous des 20 000 €. Ce niveau moyen témoigne d'un certain équilibre entre les performances.

On peut également analyser les résultats par **catégories**. Le faible profit de **Furniture** s'explique aisément : deux de ses quatre sous-catégories affichent des résultats négatifs. À l'inverse, **Technology** tire clairement son épingle du jeu, en regroupant les trois sous-catégories les plus rentables. Enfin, **Office Supplies** présente un bilan plus contrasté, avec des performances très variables selon les sous-catégories — certaines très profitables, d'autres plus modestes.



Maintenant que nous avons exploré les produits en détail, intéressons-nous à leur répartition géographique.

La région *Ouest* se démarque comme la plus rentable, suivie de près par *l'Est*. En revanche, le *Sud* affiche des performances en retrait, notamment pour les catégories **Office Supplies** et **Technology**, qui y génèrent un profit nettement inférieur à leur niveau habituel. Ce constat s'étend également à la catégorie **Office Supplies** dans la région *Centre*.

Ces écarts de rentabilité pourraient indiquer une inadéquation entre l'offre et la demande dans ces régions, ou un besoin d'adaptation des stratégies commerciales pour mieux répondre aux spécificités locales.

Enfin, bien que les résultats de **Furniture** restent faibles dans l'ensemble des régions, ils ne surprennent plus à ce stade de l'analyse, puisqu'ils ont été largement expliqués précédemment. Ce faible profit semble indiquer une faiblesse structurelle dans cette catégorie, potentiellement due à une combinaison de faible demande et de marges peu compétitives.

C. Analyse des Segments Clients

Conformément à la méthodologie annoncée, certaines variables complémentaires ont pu être explorées. C'est le cas de la variable *Segment*, qui distingue les clients selon leur profil : *Consumer* (Particuliers), *Corporate* (Entreprises) et *Home Office* (Télétravailleurs). L'impact potentiel de cette variable sur le volume de commandes ou la rentabilité a été examiné.

Toutefois, l'analyse ne révèle pas de différences significatives ou de tendance claire entre ces segments. En d'autres termes, les variations de profit ou de commandes semblent davantage liées à d'autres facteurs (comme la région ou la catégorie de produit) qu'au type de client.

Seule information potentiellement utile : le nombre total de clients uniques observés dans chaque segment, qui permet d'estimer la taille relative de chaque groupe :

- *Consumer* (Particuliers) : **409** clients uniques
- *Corporate* (Entreprises) : **236** clients uniques
- *Home Office* (Télétravailleurs) : **148** clients uniques

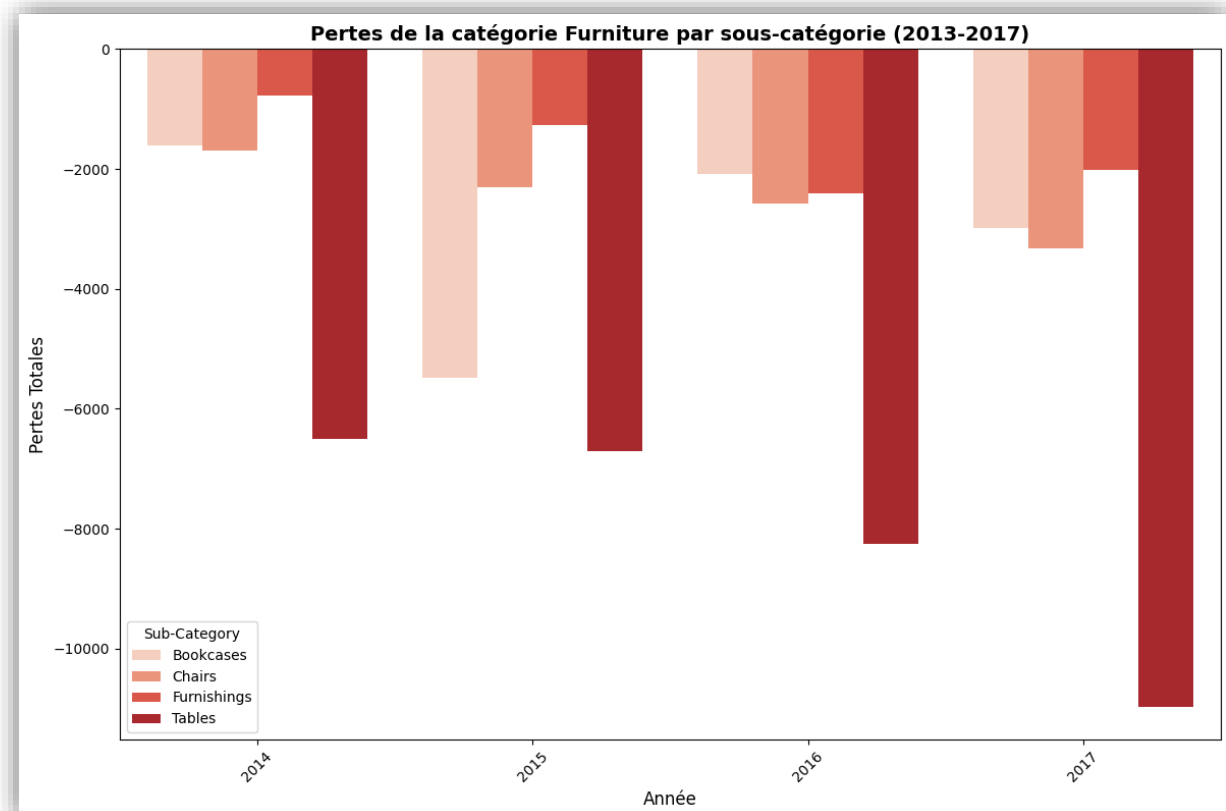
D. Étude des Facteurs Opérationnels et Temporels

Afin de replacer nos données dans un cadre temporel, nous avons commencé par analyser la répartition des dates de commande. La période couverte s'étend de fin 2013 à fin 2017, soit près de quatre années complètes d'activité. Ce contexte est important : une période courte (quelques semaines, par exemple) aurait limité la portée de notre analyse, tandis qu'ici, l'ampleur de l'échantillon permet d'observer des tendances plus durables.

En moyenne, on recense environ 47,82 commandes par semaine. Les extrêmes sont atteints lors de la première semaine (fin 2013) avec seulement 5 commandes, et au début du mois de décembre 2017 avec un pic à 131 commandes hebdomadaires. Ces éléments confirment une certaine dynamique dans la fréquence des commandes au fil du temps.

Comme mentionné précédemment, la catégorie *Furniture* affiche un bénéfice global plus faible malgré un volume de ventes comparable à celui des autres catégories principales. Nous avons donc

choisi de nous concentrer sur cette catégorie afin d'identifier d'éventuels facteurs expliquant cette rentabilité inférieure.



L'évolution du profit par année pour *Furniture* montre une tendance globalement négative, en lien direct avec une augmentation du nombre de commandes. Cette observation renforce notre hypothèse initiale : le problème ne semble pas provenir d'une année exceptionnelle, mais plutôt d'un déséquilibre structurel. La rentabilité de la catégorie *Furniture* souffre donc davantage d'un manque d'équilibre entre les ventes et les marges que d'une conjoncture ponctuelle.

Nous avons également envisagé d'autres pistes d'explication, comme le *Ship Mode*. Toutefois, malgré différentes tentatives d'analyse, aucune corrélation marquante n'a pu être dégagée. Standard Class représente un peu moins de 60 % des livraisons, mais ni cette domination, ni la répartition du profit selon les modes d'expédition ne permettent de formuler une hypothèse solide.

Enfin, nous avons examiné l'impact des remises (*Discounts*). Intuitivement, des remises trop importantes pourraient affecter négativement le bénéfice. Dans le cas de *Furniture*, ce sont les sous-catégories *Tables* et *Bookcases* qui affichent les remises moyennes les plus élevées. Toutefois, cette tendance ne se confirme pas lorsqu'on étend l'analyse à l'ensemble des catégories. Ce facteur ne suffit donc pas à expliquer à lui seul le déficit observé.

IV. Discussion

A. Limites

Dans l'ensemble, l'analyse n'a pas rencontré de freins majeurs. Les données disponibles étaient suffisantes pour explorer les axes jugés pertinents, et aucune contrainte technique ou méthodologique n'est venue entraver significativement le déroulement du travail. Les seules limites rencontrées concernent certains résultats intermédiaires qui, bien qu'explorés, ne se sont pas révélés suffisamment exploitables ou pertinents pour nourrir les conclusions de manière concrète. Ces pistes ont donc été écartées au fil de l'analyse pour conserver un cap clair et utile.

B. Recommandations

L'étude a permis d'identifier une catégorie de produits globalement moins rentable (Furniture), en raison notamment de plusieurs sous-catégories affichant un profit négatif (Tables, Bookcases). Ces constats ont conduit à une exploration approfondie des leviers d'action potentiels au sein de ces sous-catégories déficitaires. Cependant, aucune piste d'optimisation n'a émergé de manière convaincante. En l'état, la recommandation principale serait d'envisager l'arrêt pur et simple de la production ou de la commercialisation des éléments appartenant à ces sous-catégories non rentables, afin de limiter leur impact sur la performance globale.

V. Conclusion

A. Résumé :

L'objectif de cette analyse était d'identifier les leviers susceptibles d'améliorer la rentabilité globale. Pour cela, nous avons étudié les performances par catégorie et sous-catégorie de produits, afin de repérer les zones de fragilité. L'attention s'est portée en priorité sur une catégorie dont le rendement s'est révélé inférieur à celui des autres, avec plusieurs sous-catégories en déficit. À l'inverse, les éléments présentant déjà de bons résultats n'ont pas fait l'objet d'un approfondissement : l'analyse s'est volontairement recentrée sur les segments les plus problématiques, dans une logique de priorisation des actions. Les résultats obtenus permettent d'éclairer les sources de sous-performance et de proposer des pistes concrètes d'optimisation.

B. Pistes futures

Dans le prolongement de ce travail, une piste d'exploration complémentaire consisterait à évaluer le **retour sur investissement** de chaque sous-catégorie ou produit, en intégrant des données supplémentaires liées aux coûts ou aux ressources mobilisées. Une telle approche permettrait de mieux cerner la rentabilité nette de chaque élément, au-delà du seul profit généré. Par ailleurs, des analyses croisées, notamment par corrélation avec d'autres variables (volume de ventes, fréquence des commandes, saisonnalité, etc.), pourraient enrichir la compréhension des dynamiques à l'œuvre et offrir un appui plus fin à la prise de décision stratégique.