



**UNIVERSITÉ
DE LORRAINE**

Rapport de Projet

**Filière : Ingénierie Mathématique pour la science
des Données**

Niveau : M2

Sujet : Visualisation de la base de données Sales

Encadré par : Mme Céline TREUILLIER

Élaboré par : Yassine MARNISSI, Zineb ZAHOUANI

Année Universitaire : 2023-2024

Chapitre 1 : Introduction du projet

1.1 Contexte du projet

Notre projet vise à identifier les tendances de vente et les préférences des clients en analysant des données sur les achats par produit, région, et période. À travers la visualisation, nous cherchons à fournir des insights clairs pour guider les décisions stratégiques. L'idée, c'est d'identifier des tendances, de voir quels produits se vendent bien, et de comprendre les préférences des clients, pour aider à prendre de bonnes décisions.

1.2 Objectives

L'objectif général de notre projet est de déployer des visualisations de données pour analyser et comprendre les interactions complexes entre clients, produits, régions et temporalités dans notre jeu de données. En explorant ces dimensions, nous aspirons à :

- **Identifier les tendances et les motifs** : Révéler les dynamiques de vente au fil du temps et à travers différentes régions.
- **Comprendre le comportement des clients** : Analyser les préférences et les habitudes d'achat des clients pour mieux répondre à leurs besoins.
- **Évaluer la performance des produits** : Distinguer les produits les plus performants des moins performants.
- **Optimiser la stratégie de distribution régionale** : Identifier les régions clés qui contribuent le plus aux ventes

1.3 Présentation de la base de données

Notre base de données est un dataset de l'entreprise contenant des informations complètes sur les transactions de vente, réparties en trois feuilles Excel distinctes. La feuille "Orders" détaille plus de 10 000 commandes, incluant les dates de commande et d'expédition, les modes d'expédition, les informations clients, les segments de marché, ainsi que les détails géographiques des transactions. De plus, chaque commande est caractérisée par des informations sur les produits, la catégorie, la sous-catégorie, le nom du produit, les quantités vendues, les remises appliquées et les profits.

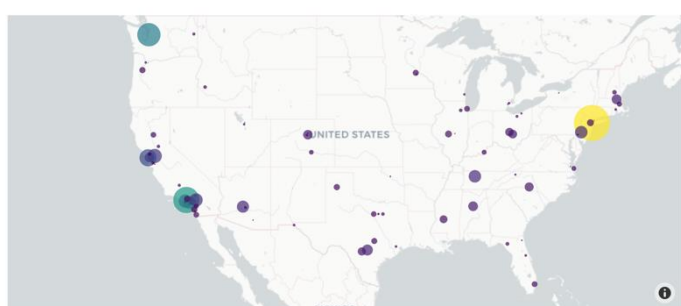
Chapitre 2 : Visualisations Simples et Analyses Préliminaires

Dans cette partie nous allons aborder la visualisation simple de la base de données suivant plusieurs analyses, dans chacune des analyses viendront pour répondre à des questions et des objectifs précis.

Nous allons détailler une question par type d'analyse et veuillez trouver en annexe la suite des questions et des graphes.

3.1 Analyse descriptive : Quelle est la distribution géographique des ventes ?

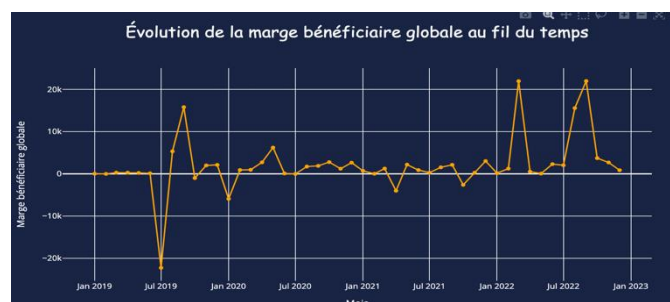
Distribution géographique des ventes par ville



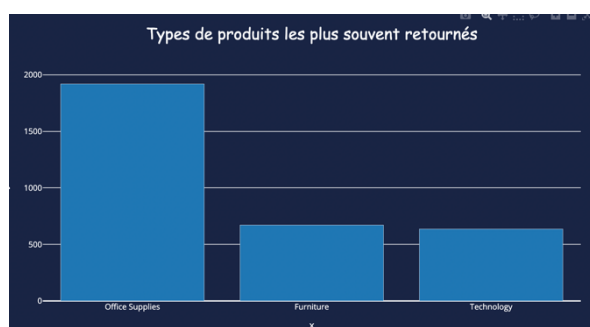
La carte montre la distribution des ventes aux États-Unis, avec des cercles colorés représentant le volume des ventes par ville, plus le cercle est gros et foncé, plus les ventes sont élevées. Cette visualisation nous a été utile pour identifier les marchés clés et informer les décisions stratégiques.

3.2 Analyse temporelle : Comment évolue la marge bénéficiaire globale au fil du temps ?

Le graphique représente l'évolution de la marge bénéficiaire globale de janvier 2019 à janvier 2023. La ligne orange indique les fluctuations mensuelles de la marge, avec des pics notables et quelques chutes en-dessous de zéro, ce qui suggère des pertes pendant certains mois.

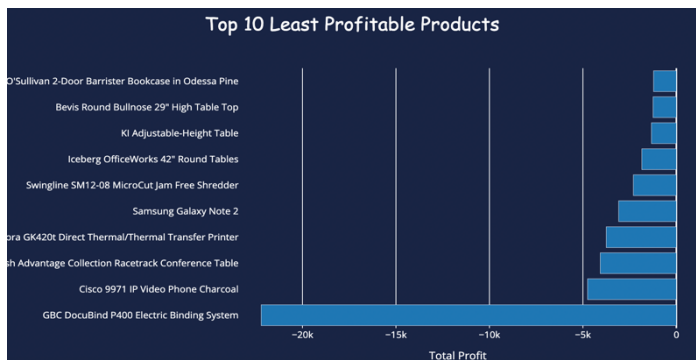


3.3 Analyse des retours : Quels types de produits sont les plus souvent retournés ?



Notre graphique en barres représente les types de produits les plus souvent retournés, classés en trois catégories : fournitures de bureau, meubles et technologie. Les fournitures de bureau sont de loin la catégorie la plus retournée, suivie des meubles et de la technologie, bien moins retournés en comparaison.

3.4 Analyse de la rentabilité : Quels sont les produits les moins rentables ?



Le graphique en barres horizontales présente les 10 produits les moins rentables pour une entreprise, avec des pertes allant jusqu'à 20 000. Les produits sont variés, incluant des meubles, des appareils électroniques et des fournitures de bureau. Chaque barre indique la perte totale associée à chaque produit, mettant en évidence ceux qui impactent négativement la rentabilité.

3.5 Analyse de la performance régionale : Quels sont les régions ou les villes qui génèrent le plus de ventes ?

Le graphique montre les ventes par ville sous forme de barres verticales, avec une ville particulière se démarquant par un volume de ventes nettement supérieur aux autres, indiquant un marché particulièrement important ou réussi dans cette localité. Les autres villes présentent des ventes plus modérées en comparaison.



3.6 Analyse des variables catégorielles : Comment se répartissent les ventes par catégorie de produit ?



Le diagramme circulaire illustre la répartition des ventes par catégorie de produit. La technologie représente la part la plus importante avec 39,9%, suivie du mobilier (meubles) à 33,3% et enfin des fournitures de bureau à 26,8%. Cela suggère une prédominance des produits technologiques dans les ventes globales.

Chapitre 3 : Visualisations Interactives et Dashboard Opérationnel

4.1 Choix des outils

Plotly : L'outil de visualisation interactif



Lors de notre analyse nous avons fait appel à Plotly qui est une bibliothèque graphique open-source qui permet de réaliser des visualisations interactives et sophistiquées. Elle gère des types de graphiques variés, allant des plus simples aux plus élaborés, comme les graphiques 3D ou les cartes. L'interactivité est au cœur de Plotly, offrant des fonctionnalités de zoom, de déplacement, et d'informations au survol, enrichissant ainsi l'expérience utilisateur.

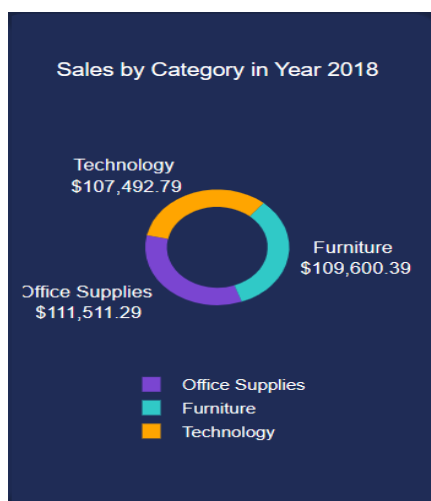
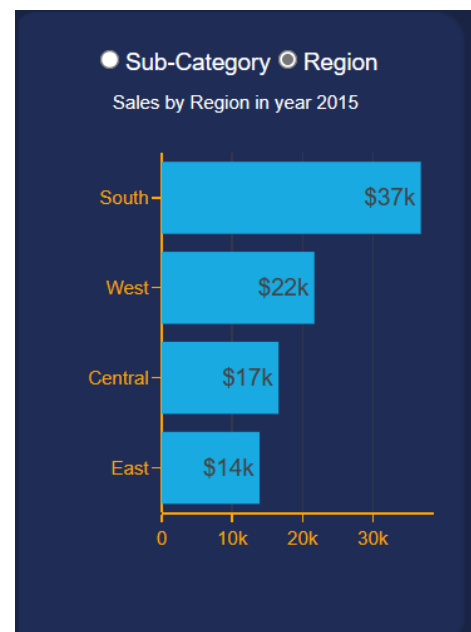
Dash : Le Cadre pour Applications Web Dynamiques



Dash est un framework Python pour la création d'applications web dynamiques et de tableaux de bord interactifs. Basé sur Flask, Plotly.js, et React.js, Dash permet d'intégrer des graphiques interactifs dans des interfaces utilisateur attractives et personnalisées, rendant la visualisation de données plus intuitive et engageante. Dash est composé de deux éléments essentiels à son fonctionnement qui sont les **Layouts** et les **Callbacks**.

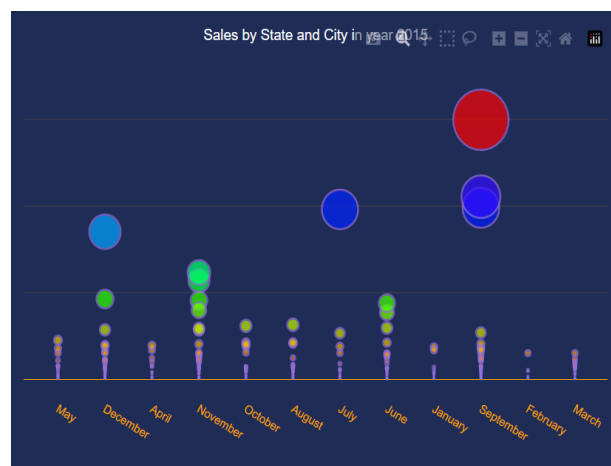
4.2 Conception du dashboard

La fonction mise en place actualise **un graphique à barres** basé sur le choix de l'utilisateur entre 'Sous-catégorie' et 'Région'. Pour 'Sous-catégorie', elle utilise des données de ventes annuelles par sous-catégorie pour créer un graphique. Des infos supplémentaires comme l'année, le segment, et les ventes sont affichées au survol. Le graphique s'intitule "Ventes par sous-catégorie en année" avec l'année en question. Dans le cas de 'Région', la même logique est appliquée pour montrer les ventes par région dans le titre "Ventes par région en année", toujours suivies de l'année.

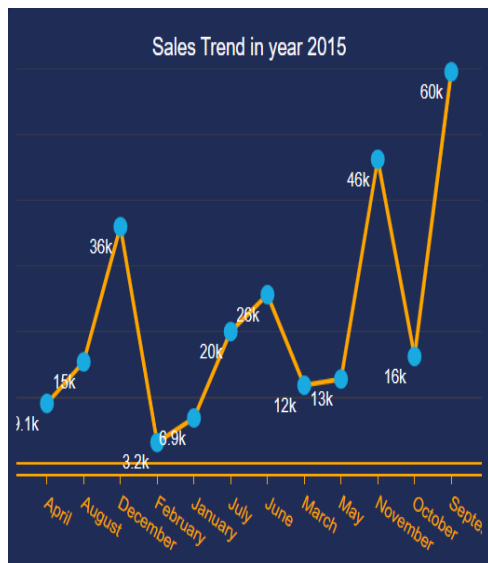


Un graphique en anneau est mis à jour en fonction des interactions de l'utilisateur. Les interactions sont gérées par des "callbacks", qui sont des fonctions qui sont appelées en réponse à des événements, comme le changement de valeur d'un élément d'interface utilisateur.

Un graphique à bulles mis à jour en fonction des interactions de l'utilisateur. Qui sont gérées par un callback `update_graph`. Elle prend en entrée `select_year` pour l'année sélectionnée et `radio_items` pour le segment choisi par l'utilisateur. Le traitement des données consiste d'abord à agréger les ventes selon



l'année, le segment, l'État, la ville et le mois, puis à filtrer les données en fonction des sélections de l'utilisateur. Ensuite, un graphique à bulles est généré à partir des données filtrées, où chaque bulle représente les ventes pour un mois donné, avec la taille des bulles proportionnelle au montant des ventes et la couleur variant en fonction de ce montant. La sortie de la fonction est le graphique à bulles mis à jour, prêt à être affiché dans l'application.



La fonction **update_graph** est un callback utilisé dans une application Dash pour mettre à jour **un graphique en ligne** (line chart) en fonction des sélections de l'utilisateur pour l'année et le segment. Les données de ventes sont traitées en regroupant les ventes par année, segment et mois, puis en filtrant les données selon les sélections de l'utilisateur. Ensuite, un graphique en ligne est généré à partir des données filtrées, où chaque point représente les ventes pour un mois donné.

Chapitre 4 : Conclusion

En conclusion, cette étude nous a permis d'explorer et d'analyser en détail les tendances de vente de l'entreprise. Nos visualisations interactives ont révélé des informations précieuses sur le comportement des clients, la performance des produits et les dynamiques de vente régionales. Bien que nous ayons surmonté plusieurs défis techniques notamment dans les nouvelles bibliothèques comme dash, nous avons tiré des leçons de ces limitations à notre pour développer notre base de données et trouver les bonnes visualisations qui permettent le mieux à répondre aux questions et objectifs définies dans l'introduction.

Chapitre 5 : Bibliographie et Annexes

5.1 Bibliographie :

- [Plotly vs Matplotlib: Lequel est meilleur pour la visualisation de données ? – Kanaries](#)

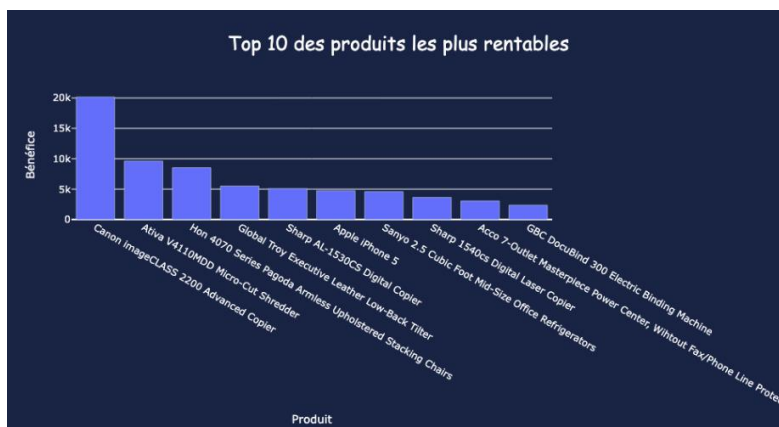
5.2 Annexes :

Analyse descriptive :

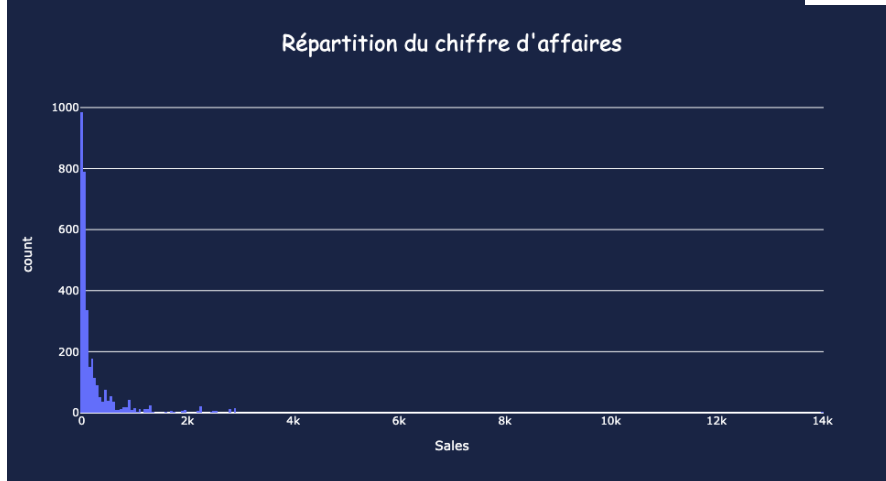
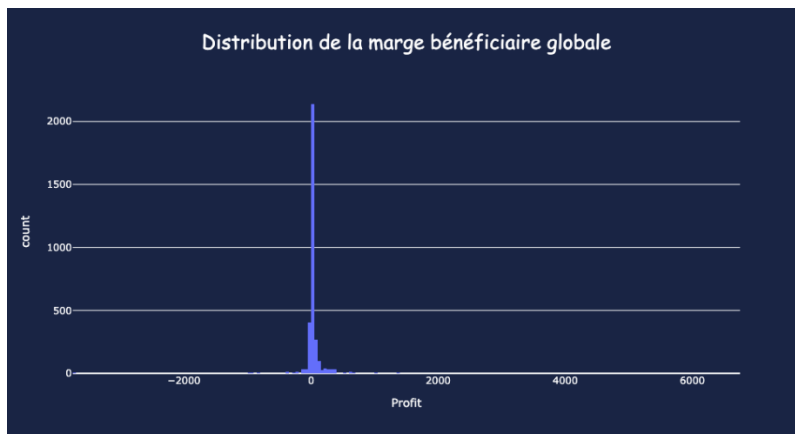
Quelle est la distribution des ventes (Sales) ?



Quels sont les produits les plus vendus en termes de quantité et de bénéfices (Profit) ?

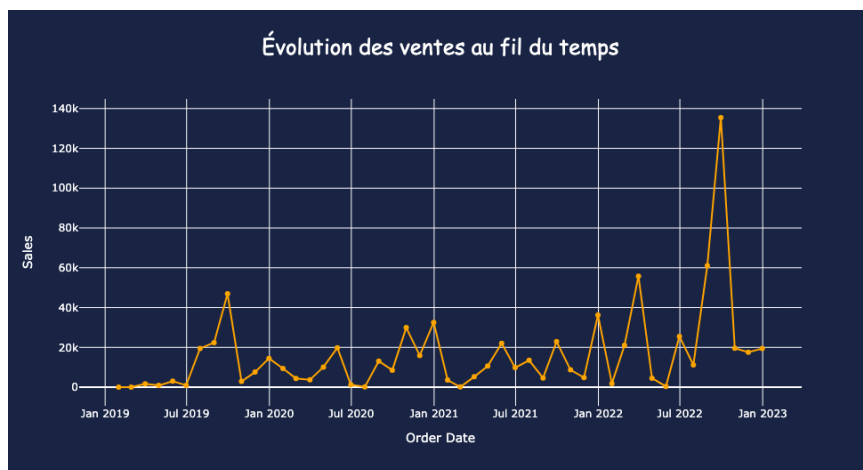


Quelle est la répartition des commandes par mode de livraison (Ship Mode) ?

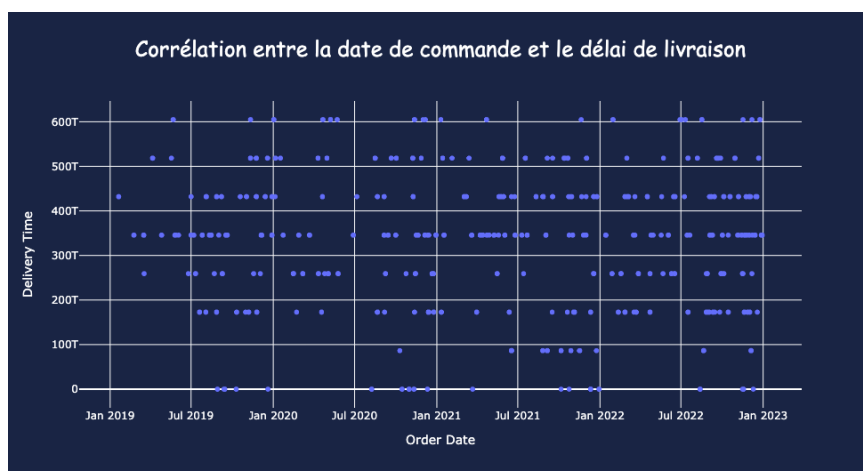


Analyse temporelle :

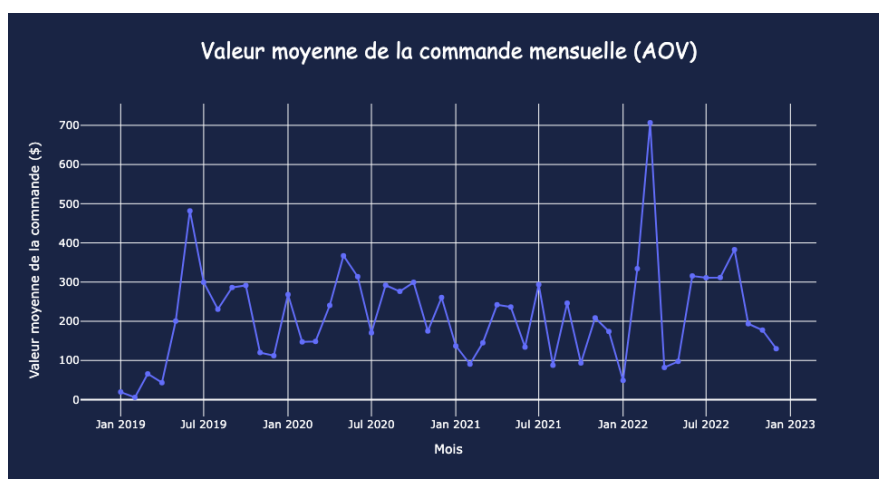
Comment évoluent les ventes au fil du temps ? Y a-t-il des tendances saisonnières ou des pics ?

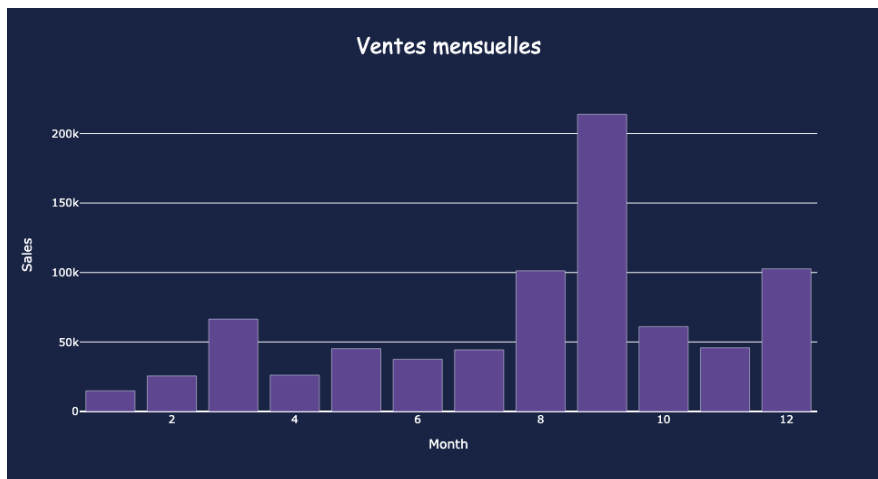


Y a-t-il une corrélation entre la date de commande (Order Date) et le délai de livraison (Ship Date) ?

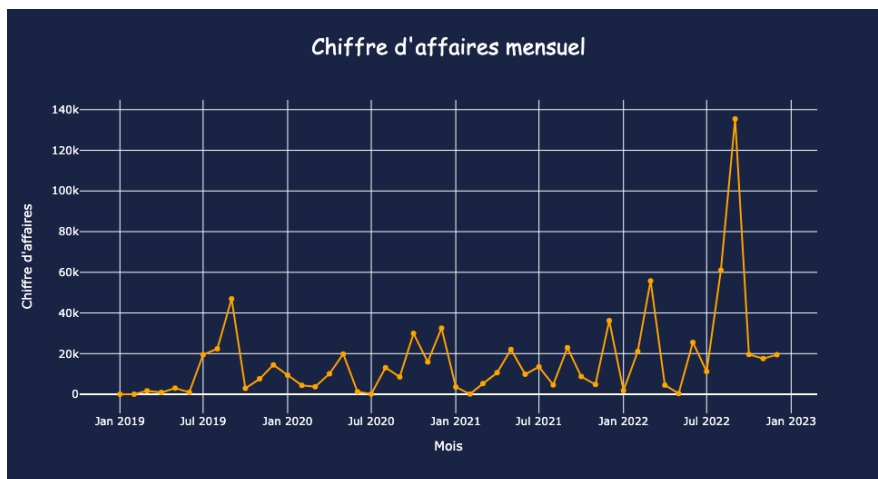


La valeur moyenne de la commande par mois

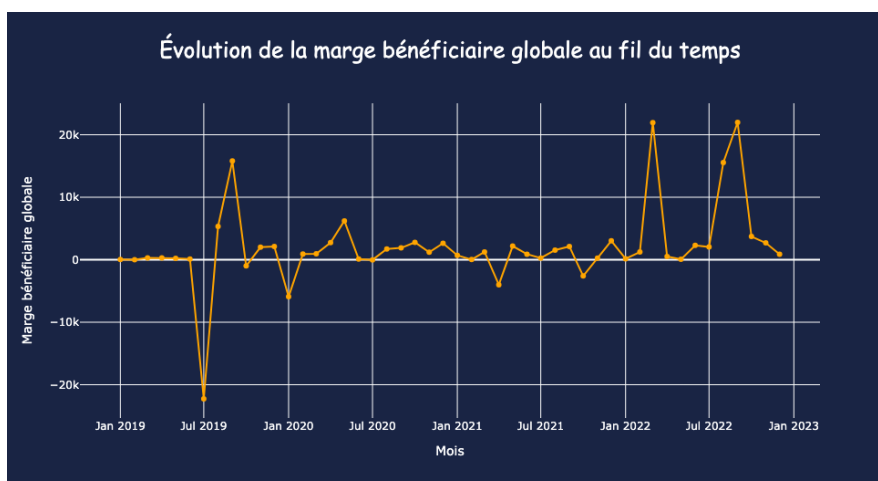




Y a-t-il des tendances saisonnières dans le chiffre d'affaires ?

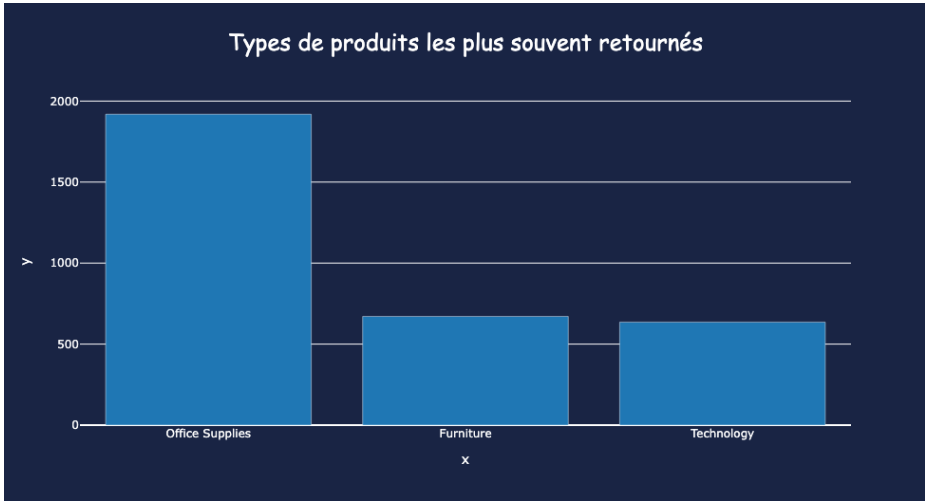


Comment évolue la marge bénéficiaire globale au fil du temps ?

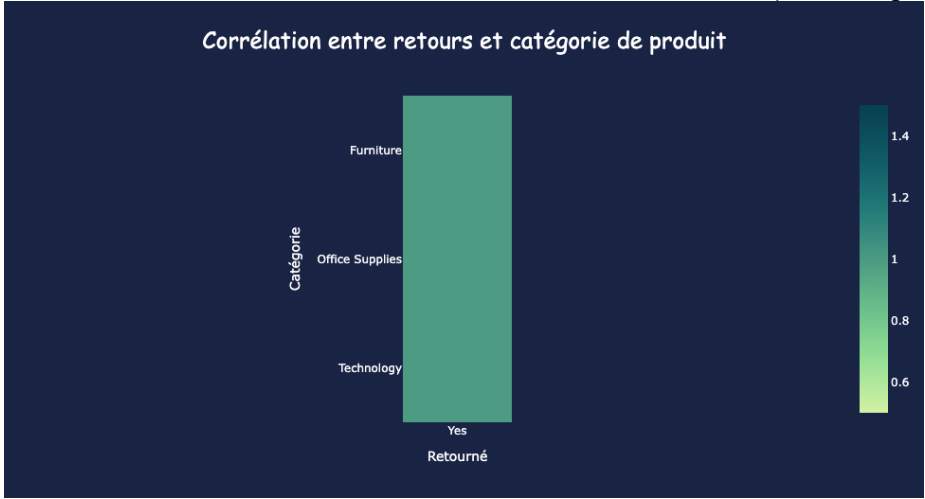


Analyse des retours :

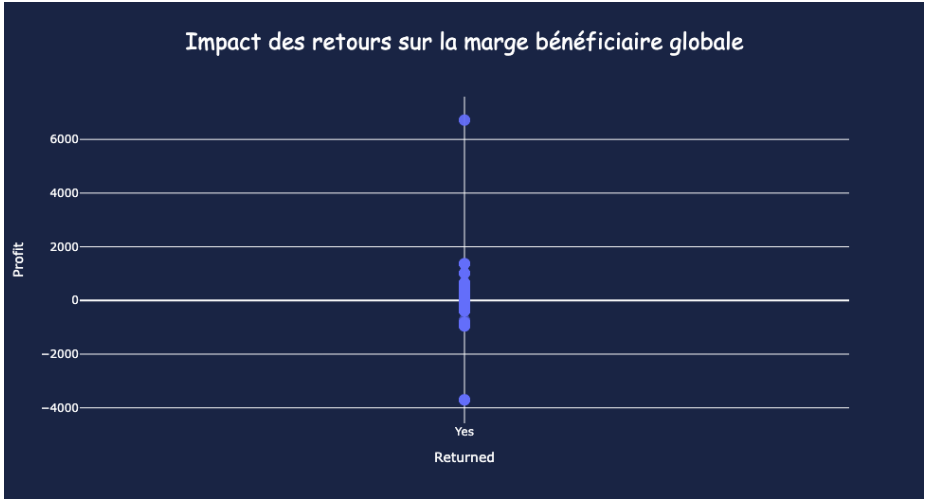
Quels types de produits sont les plus souvent retournés ?



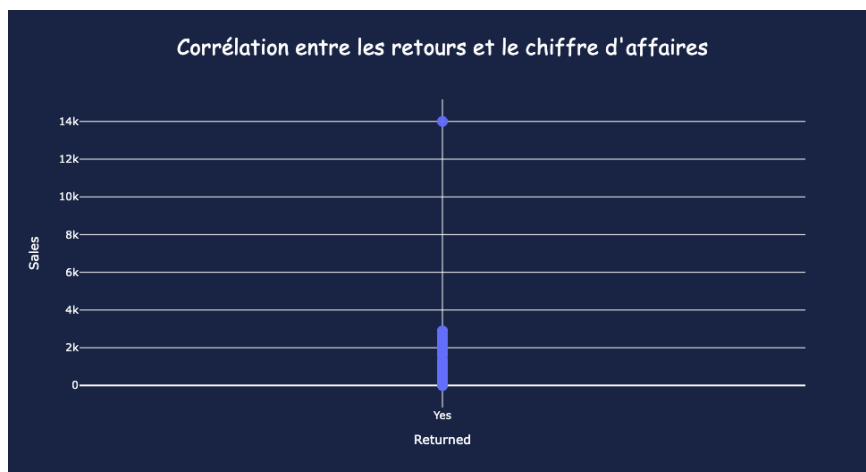
Y a-t-il une corrélation entre les retours et d'autres variables telles que la catégorie de produit, le segment de client ?



Quelle est l'impact des retours sur la marge bénéficiaire globale ?

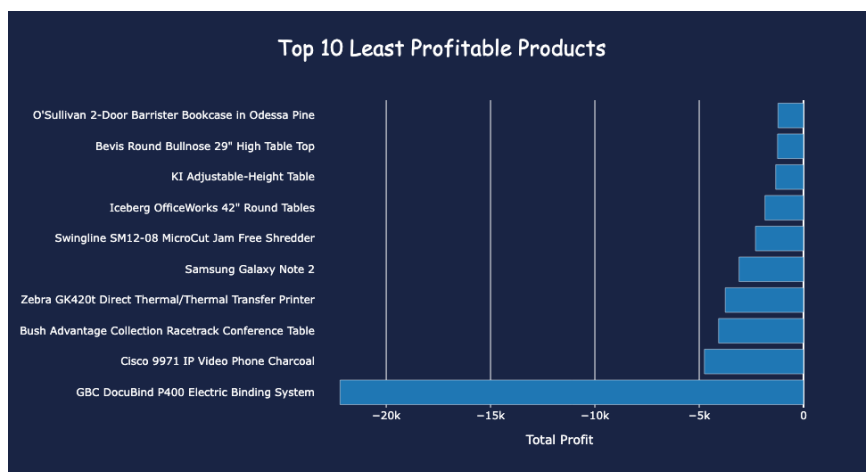


Y a-t-il une corrélation entre les retours et le chiffre d'affaires ?

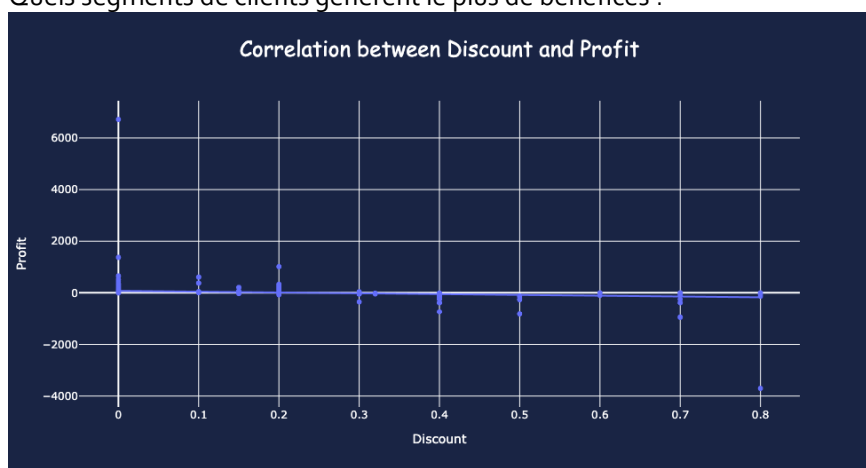


Analyse de la rentabilité :

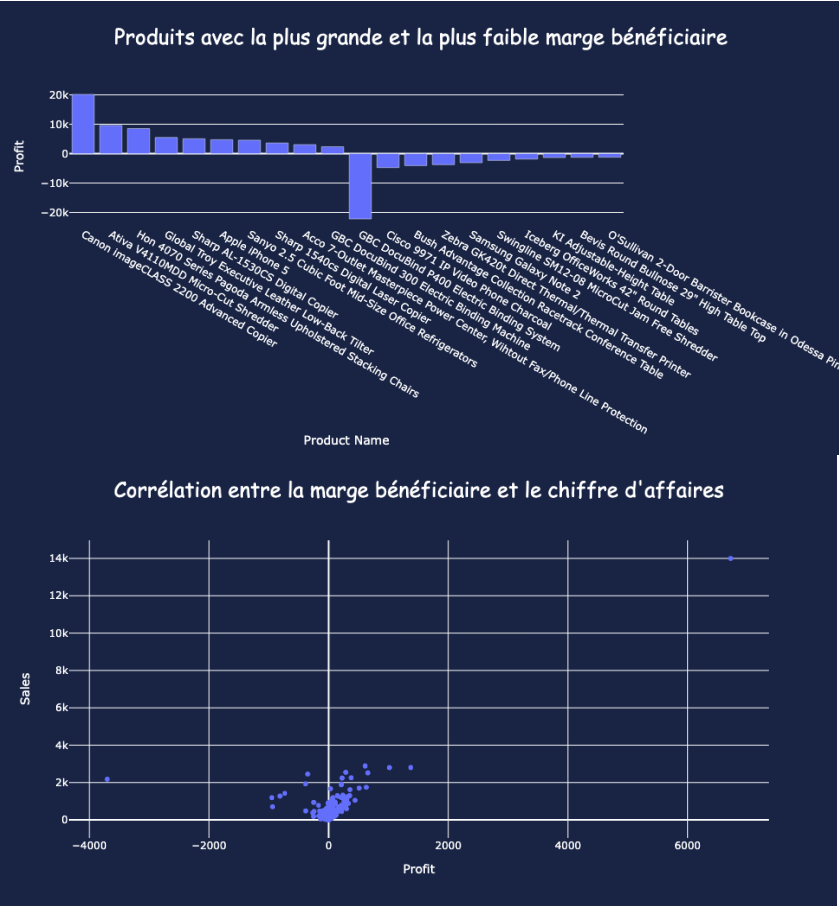
Quels sont les produits les moins rentables ?



Quels segments de clients génèrent le plus de bénéfices ?

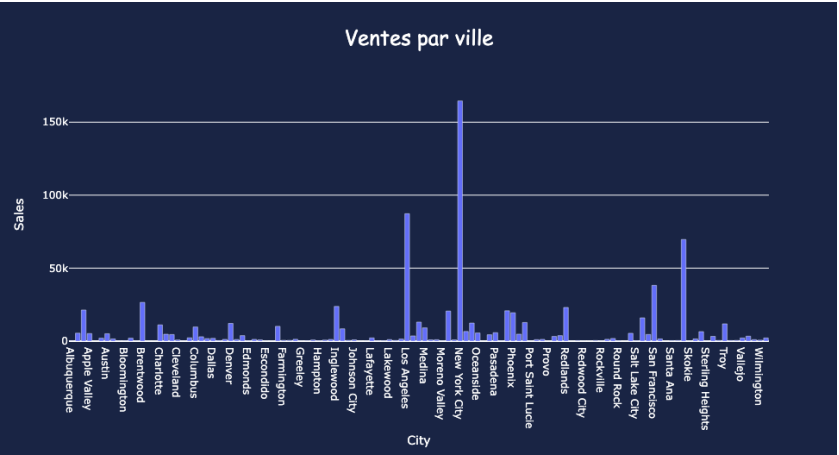


Quels sont les produits avec la plus grande et la plus faible marge bénéficiaire ?

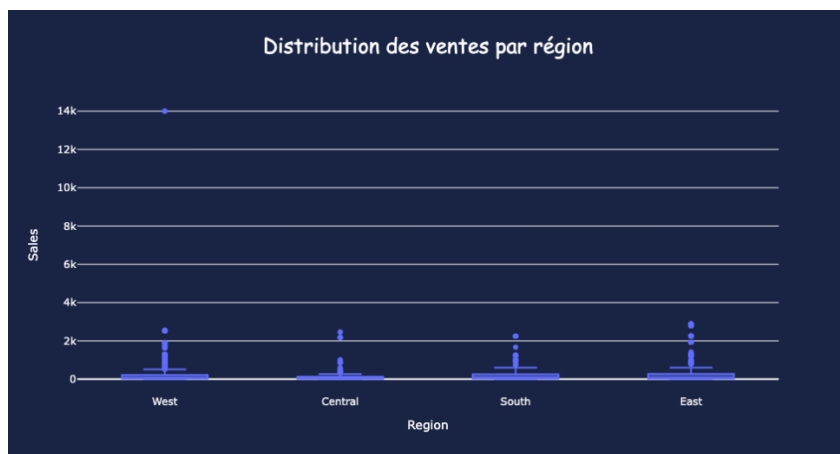


Analyse de la performance régionale :

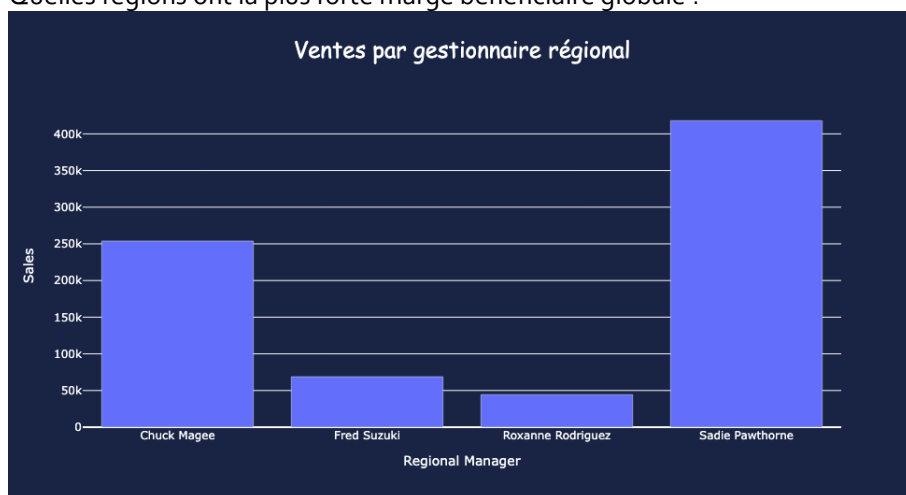
Quels sont les régions ou les villes qui génèrent le plus de ventes ?



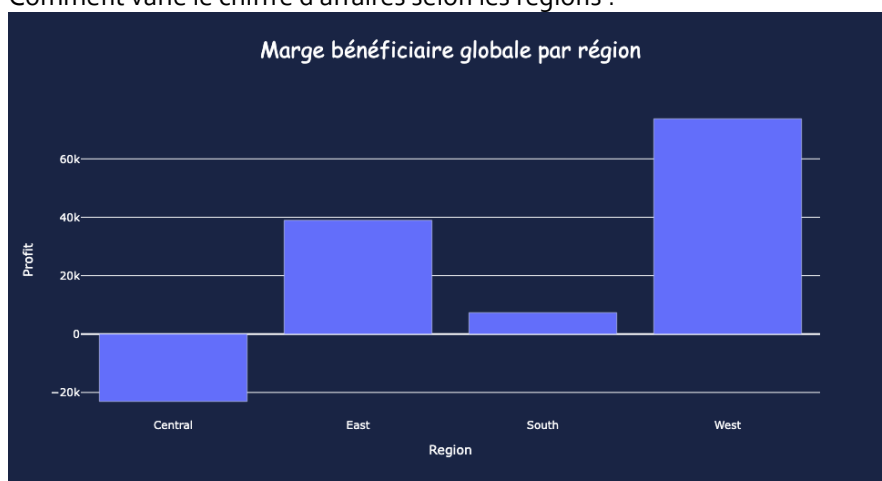
Comment la performance des régions est-elle liée à la gestion régionale (Regional Manager) ?

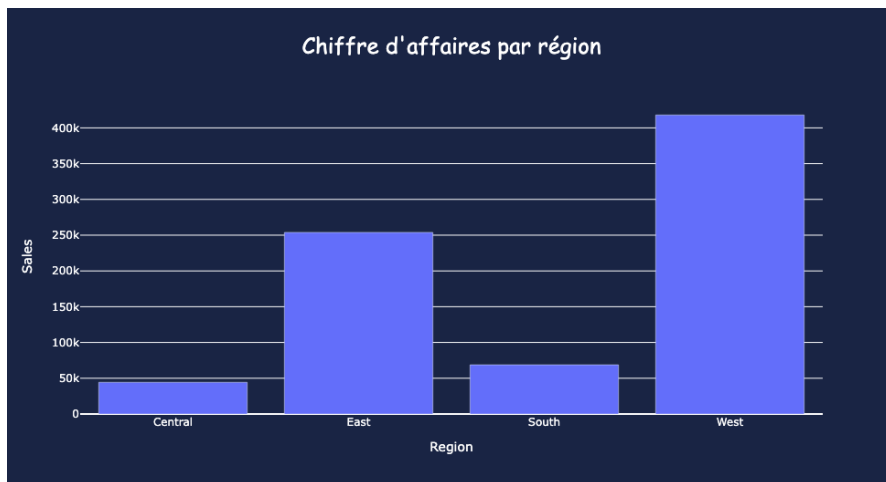


Quelles régions ont la plus forte marge bénéficiaire globale ?



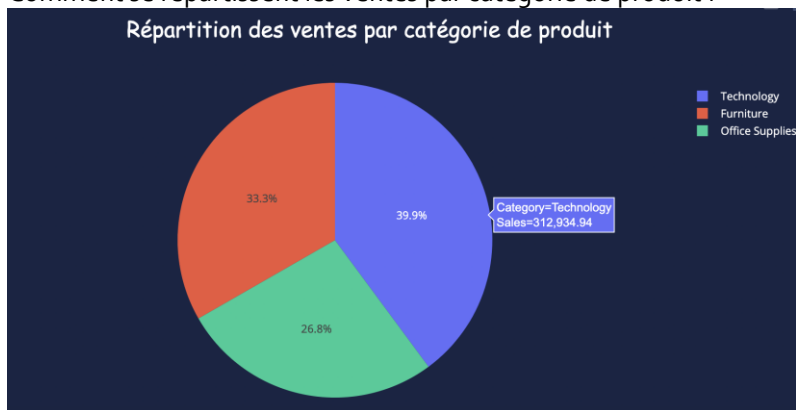
Comment varie le chiffre d'affaires selon les régions ?



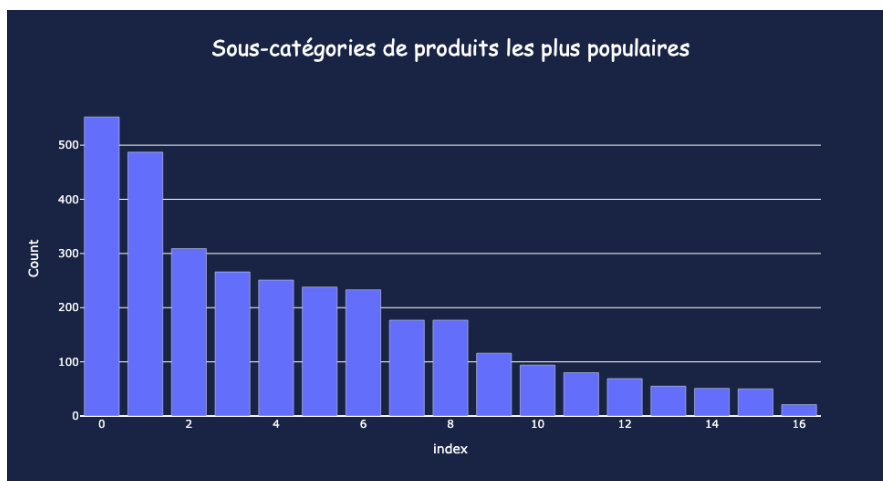


Analyse des variables catégorielles :

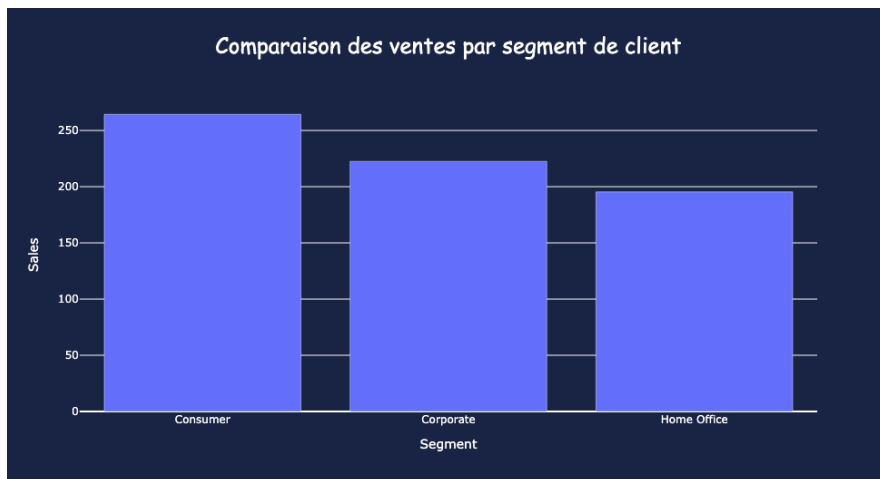
Comment se répartissent les ventes par catégorie de produit ?



Quels sont les sous-catégories de produits les plus populaires ?

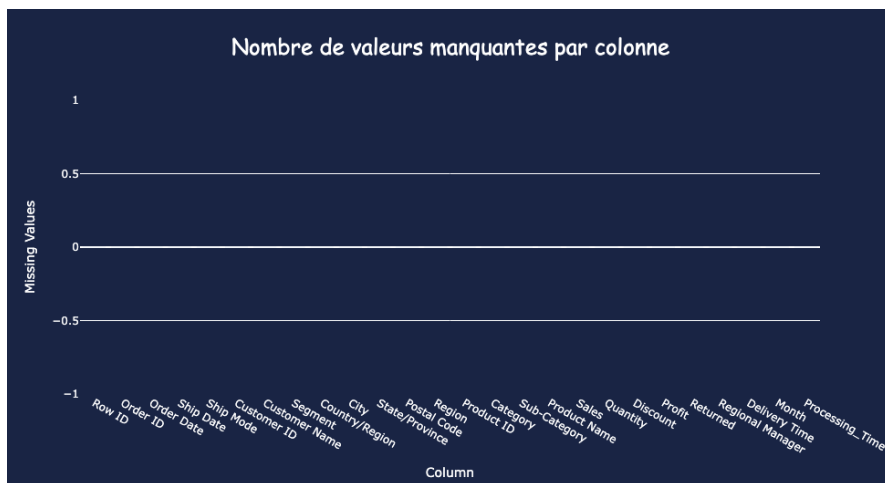


Existe-t-il des différences significatives dans les performances en fonction du segment de client ?



Analyse des données manquantes ou aberrantes :

Y a-t-il des valeurs manquantes dans la dataframe ?



Existe-t-il des valeurs aberrantes ou des erreurs dans les données qui pourraient fausser l'analyse ?

