

VISUALISATION DES DONNEES DE VENTES SUR UNE PERIODE SPECIFIQUE AVEC DES GRAPHIQUES ADAPTES

Par : NGELEKA MUSANGU Steve, DEA UNIKIN, 2022-2023

Introduction

Dans le monde des affaires, la capacité à interpréter et analyser les données de ventes est cruciale pour la prise de décisions stratégiques. Les entreprises génèrent une quantité massive de données au quotidien, notamment des informations sur les ventes, les revenus, et les tendances du marché. La visualisation des données est une méthode puissante pour transformer ces données brutes en informations claires et exploitables. En présentant les données de manière visuelle, les entreprises peuvent identifier rapidement les tendances, les anomalies et les opportunités cachées.

La visualisation des données de ventes sur une période spécifique permet de suivre la performance commerciale, d'évaluer l'efficacité des stratégies de vente, et de prévoir les tendances futures. Les graphiques adaptés jouent un rôle essentiel dans cette analyse, en facilitant la compréhension des données complexes et en rendant les informations accessibles à tous les niveaux de l'organisation.

Problématique

Comment les entreprises peuvent-elles utiliser la visualisation des données pour optimiser l'analyse de leurs ventes sur une période spécifique, en choisissant les types de graphiques les plus adaptés pour chaque type de données et objectif d'analyse ?

Revue de littérature

La visualisation des données de ventes sur une période spécifique est cruciale pour comprendre les tendances et prendre des décisions éclairées. Les bibliothèques Python telles que Matplotlib et Seaborn permettent de créer des graphiques simples et statistiques, tandis que Plotly offre des visualisations interactives et dynamiques. Les techniques avancées de visualisation, telles que les cartes thermiques et les diagrammes de Sankey, apportent une compréhension plus détaillée des relations complexes et des flux de données. L'utilisation combinée de ces techniques améliore significativement l'analyse des performances de vente et la prise de décision basée sur les données.

Quelques questions de recherche

Partant de notre problématique, nous avons relevés six questions qui font partie de notre recherche:

- Quelle est la répartition des ventes par catégorie de produits ?
- Comment varient les ventes au fil du temps sur la période spécifique ?
- Quelle est la performance des ventes dans différentes régions ou villes ?

- Y a-t-il une corrélation entre les remises accordées et le profit réalisé ?
- Quelle est la contribution des différents segments de clients aux ventes totales ?
- Comment varie le coût d'expédition par rapport aux ventes ?

Nous répondrons à ces questions en utilisant les codes python et en justifiant le choix porté sur chaque graphique par rapport à chacune des questions.

- a. Quelle est la répartition des ventes par catégorie de produits ?

Pour répondre à cette question, nous avons opté pour trois graphiques :

- **Graphique à barres** : Le graphique à barres montre clairement quelles catégories de produits génèrent le plus de ventes. Par exemple, si la catégorie "Technologie" a la plus grande barre, cela indique qu'elle est la plus performante en termes de ventes.
- **Graphique en secteurs** : Le graphique en secteurs donne une vue d'ensemble de la part de marché de chaque catégorie. Cela aide à visualiser les proportions relatives.
- **Histogramme** : L'histogramme montre la distribution des ventes dans différentes plages. Cela peut aider à identifier des anomalies ou des catégories avec des ventes particulièrement élevées ou faibles.

```
[2]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Chargement des données
data = pd.read_csv('superstore.csv')

# Répartition des ventes par catégorie de produits
sales_by_category = data.groupby('Category')['Sales'].sum().reset_index()

# Graphique à barres
plt.figure(figsize=(12, 6))
sns.barplot(x='Category', y='Sales', data=sales_by_category)
plt.title('Répartition des ventes par catégorie de produits')
plt.xlabel('Catégorie')
plt.ylabel('Ventes')
plt.show()

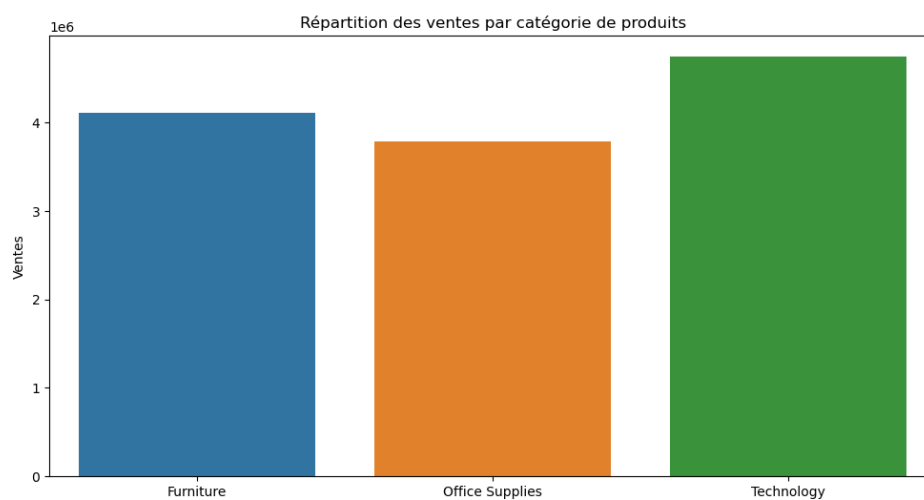
# Graphique en secteurs
plt.figure(figsize=(8, 8))
plt.pie(sales_by_category['Sales'], labels=sales_by_category['Category'], autopct='%1.1f%%', startangle=140)
plt.title('Répartition des ventes par catégorie de produits')
plt.show()

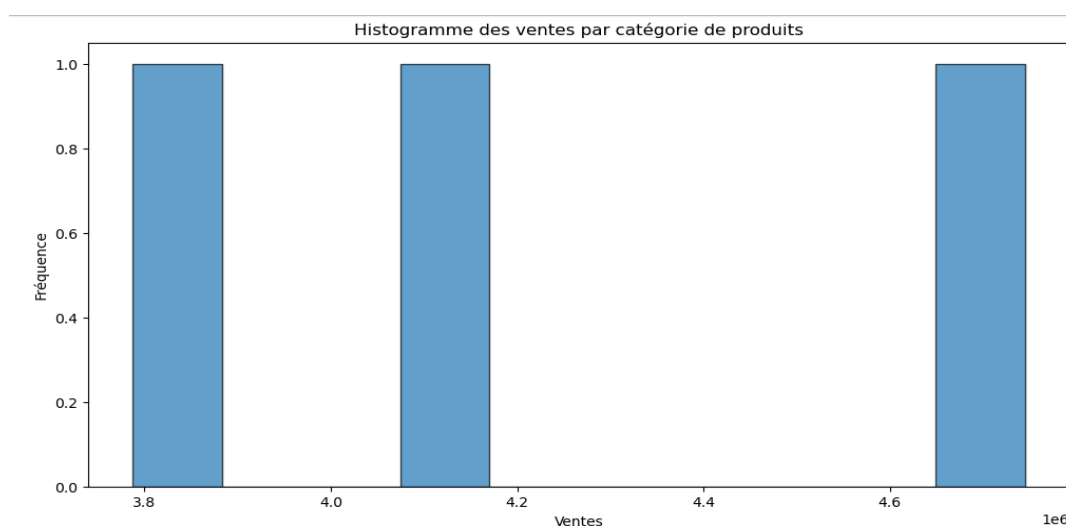
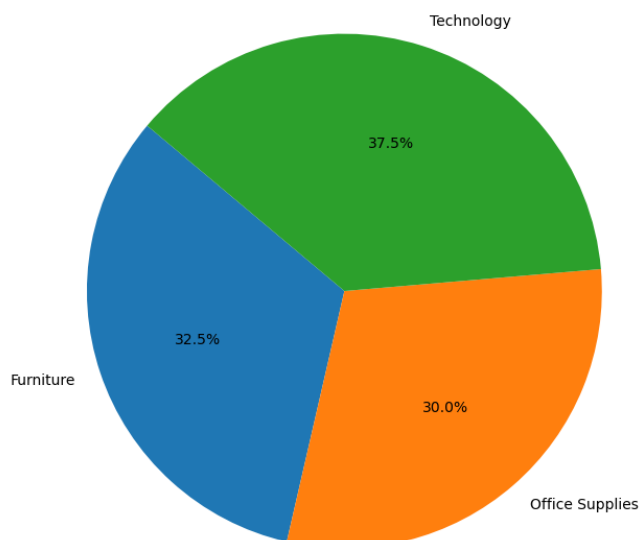
# Graphique en histogramme (en utilisant un binning des ventes)
plt.figure(figsize=(12, 6))
plt.hist(sales_by_category['Sales'], bins=10, edgecolor='k', alpha=0.7)
plt.title('Histogramme des ventes par catégorie de produits')
plt.xlabel('Ventes')
plt.ylabel('Fréquence')

# Paramètres de la solution
solution_1 = sales_by_category.to_dict()
```

Fig1 : Codes pour la première question

Résultats graphiques





b. Comment varient les ventes au fil du temps sur la période spécifique ?

En ce qui concerne cette question, nous avons opté pour les graphiques suivants :

- **Graphique de ligne** : Le graphique de ligne montre les tendances des ventes mois par mois. Des pics ou des creux peuvent être liés à des événements spécifiques, des promotions ou des saisons.
- **Graphique à barres** : Ce graphique permet de voir les ventes mensuelles de manière distincte, facilitant la comparaison directe entre différents mois.
- **Graphique en aires** : Le graphique en aires met en évidence la croissance cumulée des ventes au fil du temps, ce qui est utile pour comprendre la tendance générale.

```
[3]: # Conversion de la colonne de date
data['Order.Date'] = pd.to_datetime(data['Order.Date'])

# Ventes au fil du temps
sales_over_time = data.groupby(data['Order.Date'].dt.to_period('M'))['Sales'].sum().reset_index()
sales_over_time['Order.Date'] = sales_over_time['Order.Date'].dt.to_timestamp()

# Graphique de ligne
plt.figure(figsize=(14, 7))
sns.lineplot(x='Order.Date', y='Sales', data=sales_over_time)
plt.title('Évolution des ventes au fil du temps')
plt.xlabel('Date')
plt.ylabel('Ventes')
plt.show()

# Graphique à barres
plt.figure(figsize=(14, 7))
sns.barplot(x='Order.Date', y='Sales', data=sales_over_time)
plt.title('Ventes mensuelles')
plt.xlabel('Date')
plt.ylabel('Ventes')
plt.xticks(rotation=45)
plt.show()

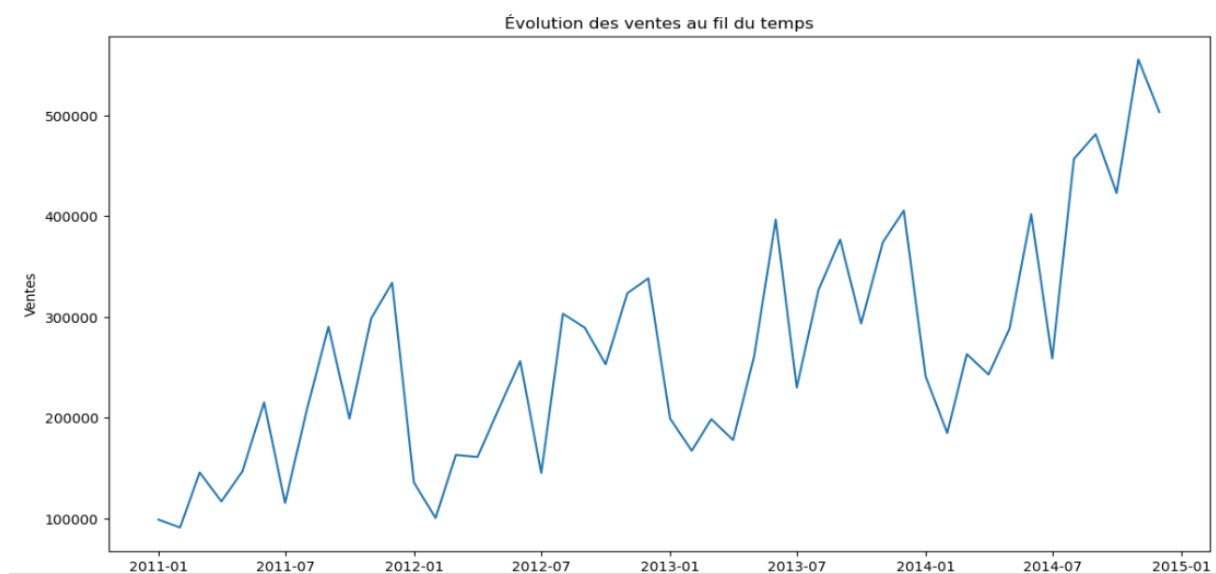
# Graphique en aires
plt.figure(figsize=(14, 7))
plt.fill_between(sales_over_time['Order.Date'], sales_over_time['Sales'], alpha=0.3)
plt.title('Ventes cumulées au fil du temps')
plt.xlabel('Date')
plt.ylabel('Ventes')

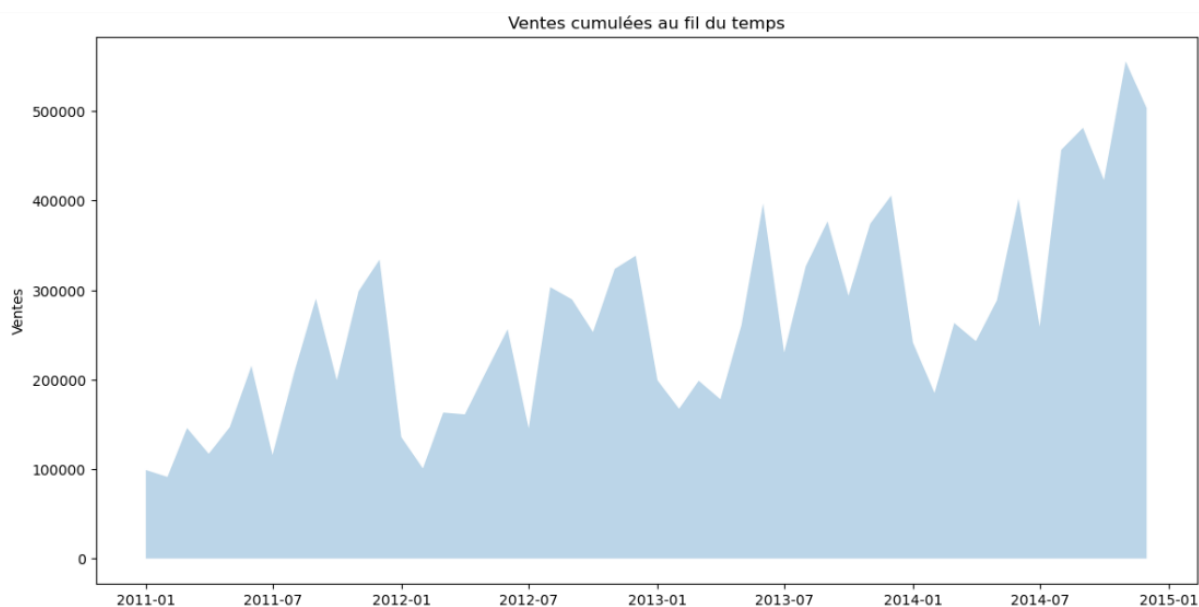
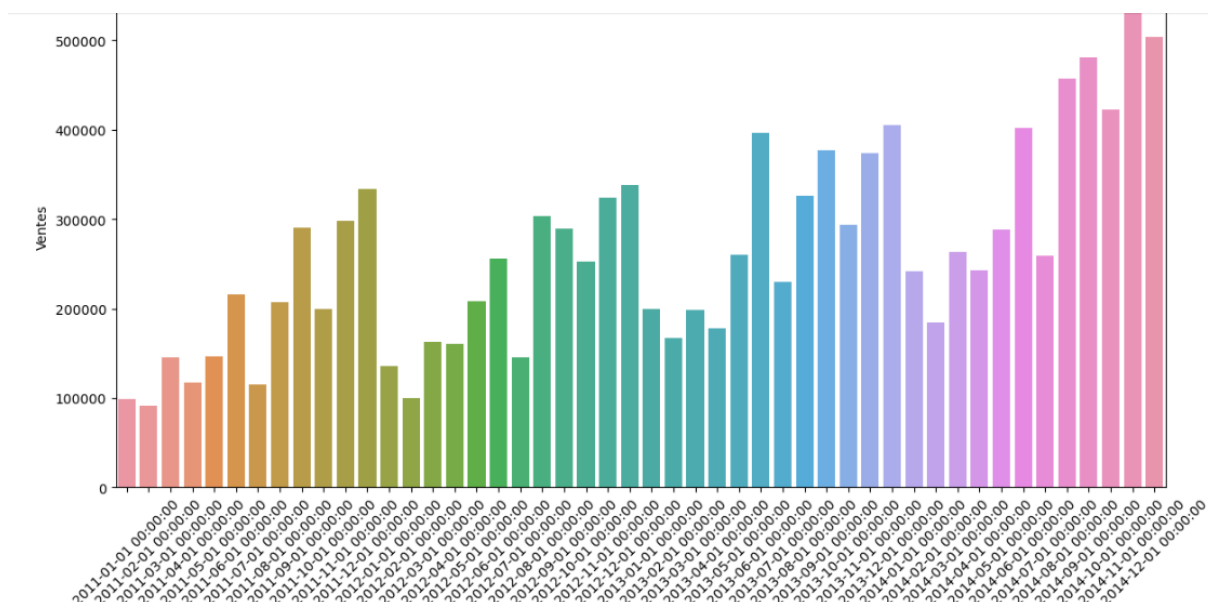
plt.show()

# Paramètres de la solution
solution_2 = sales_over_time.to_dict()
```

Fig2 : Codes python de la deuxième question

Résultats graphiques





c. Quelle est la performance des ventes dans différentes régions ou villes ?

Pour cette question, nous avons opté pour les graphiques ci-après :

- **Graphique à barres** : Montre quelle région a les ventes les plus élevées. Les différences régionales peuvent être importantes pour les décisions stratégiques.
- **Graphique en secteurs**: Donne une vue d'ensemble de la répartition des ventes entre les régions.
- **Graphique à barres horizontales** : Peut-être plus lisible pour comparer plusieurs régions côte à côte.

```
[4]: # Ventes par région
sales_by_region = data.groupby('Region')['Sales'].sum().reset_index()

# Graphique à barres
plt.figure(figsize=(12, 6))
sns.barplot(x='Region', y='Sales', data=sales_by_region)
plt.title('Performance des ventes par région')
plt.xlabel('Région')
plt.ylabel('Ventes')
plt.show()

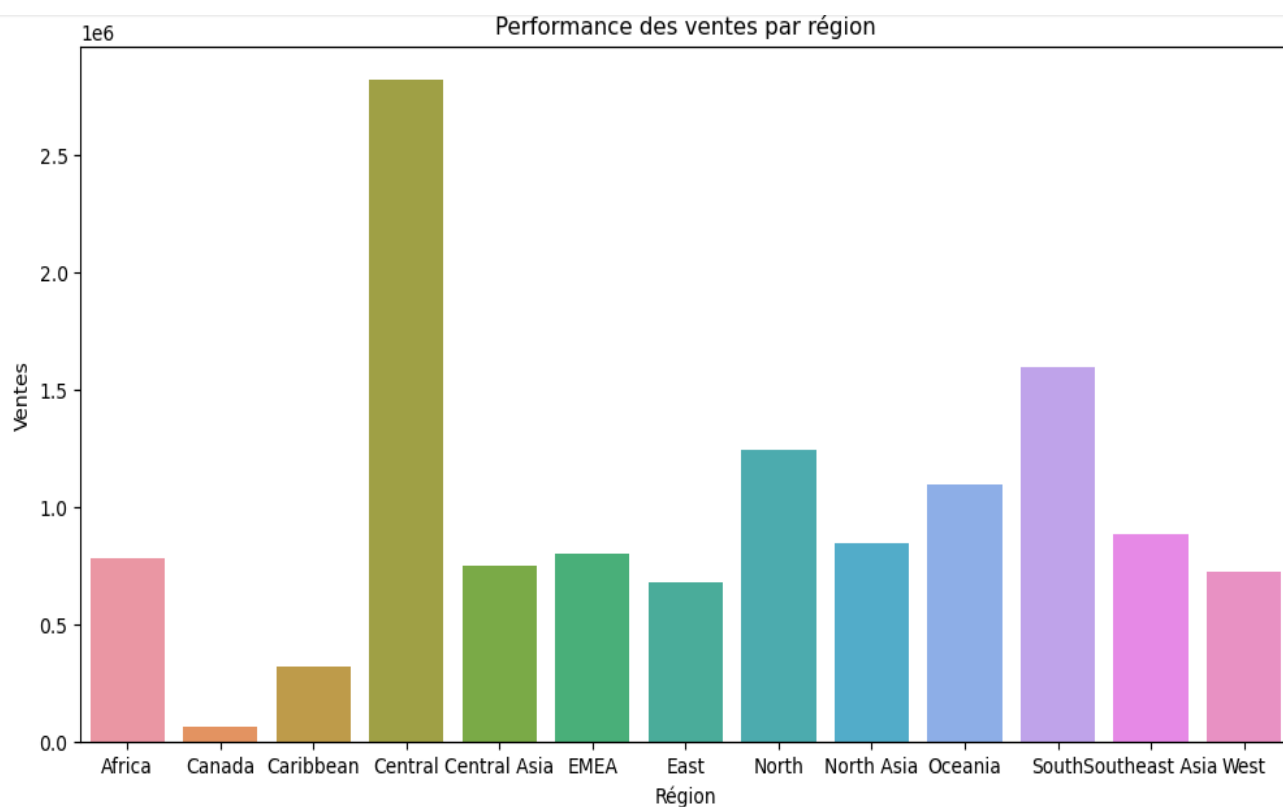
# Graphique en secteurs
plt.figure(figsize=(8, 8))
plt.pie(sales_by_region['Sales'], labels=sales_by_region['Region'], autopct='%1.1f%%', startangle=140)
plt.title('Répartition des ventes par région')
plt.show()

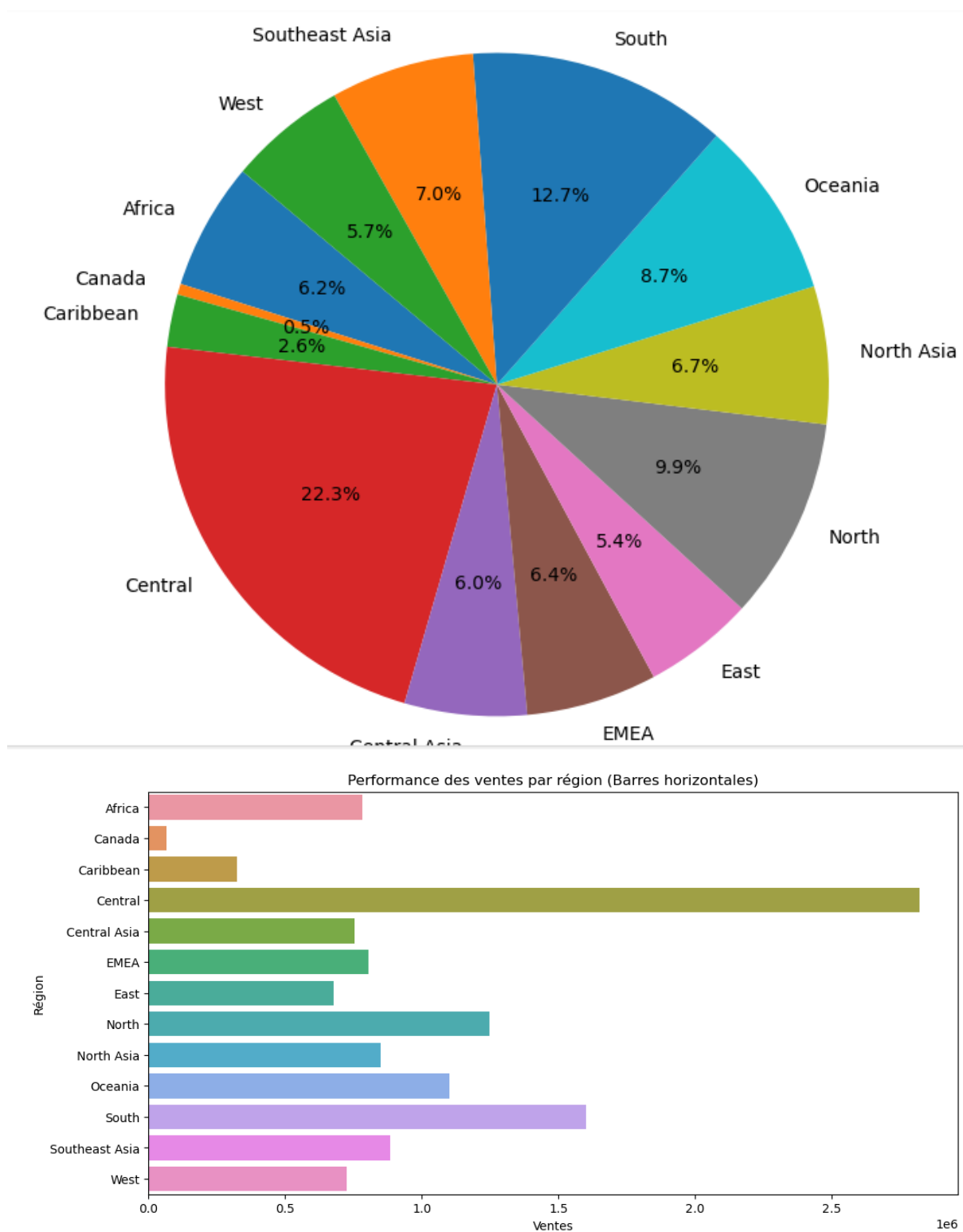
# Graphique en barres horizontales
plt.figure(figsize=(12, 6))
sns.barplot(x='Sales', y='Region', data=sales_by_region)
plt.title('Performance des ventes par région (Barres horizontales)')
plt.xlabel('Ventes')
plt.ylabel('Région')
plt.show()

# Paramètres de la solution
solution_3 = sales_by_region.to_dict()
```

Fig3 : Codes python de la troisième question

Résultats graphiques





d. Y a-t-il une corrélation entre les remises accordées et le profit réalisé ?

Pour ce qui est de cette question, nous avons opté pour les graphiques suivants :

- **Graphique de dispersion** : Permet de voir la relation directe entre les remises et le profit. Si les points sont dispersés sans tendance claire, cela indique une faible corrélation.

- **Graphique de dispersion avec ligne de régression** : Ajoute une ligne de tendance pour visualiser la corrélation. Une pente positive ou négative indique la direction de la relation.

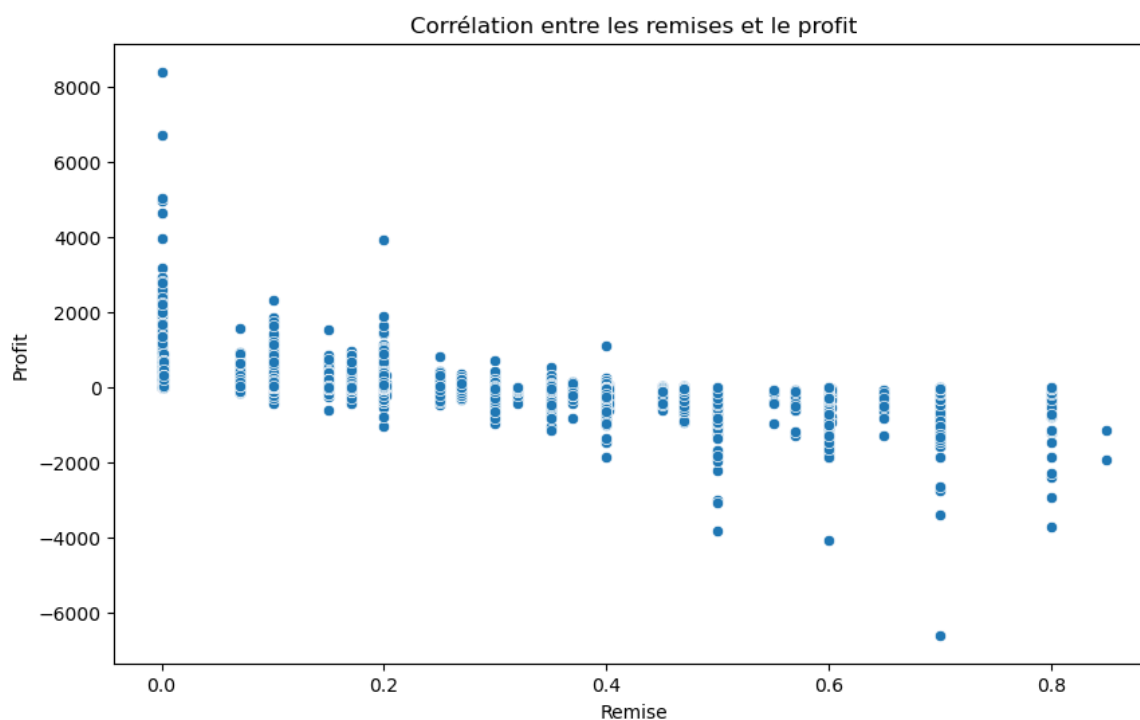
```
[5]: # Graphique de dispersion
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Discount', y='Profit', data=data)
plt.title('Corrélation entre les remises et le profit')
plt.xlabel('Remise')
plt.ylabel('Profit')
plt.show()

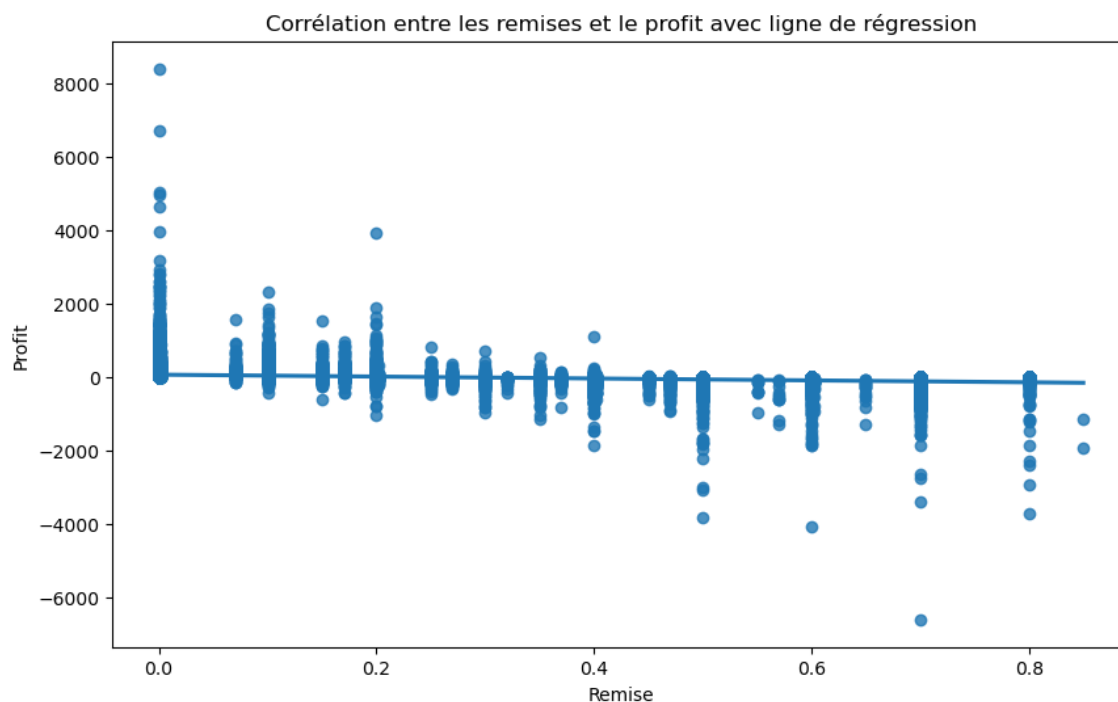
# Graphique de dispersion avec une ligne de régression
plt.figure(figsize=(10, 6))
sns.regplot(x='Discount', y='Profit', data=data)
plt.title('Corrélation entre les remises et le profit avec ligne de régression')
plt.xlabel('Remise')
plt.ylabel('Profit')
plt.show()

# Paramètres de la solution
correlation = data[['Discount', 'Profit']].corr().to_dict()
solution_4 = {'correlation': correlation}
```

Fig4 : Codes python pour la quatrième question

Résultats graphiques





e. Quelle est la contribution des différents segments de clients aux ventes totales ?

Pour donner solution à cette question, nous avons opté pour les graphiques suivants :

- **Graphique circulaire** : Affiche la part des ventes de chaque segment de clients, montrant leur importance relative.
- **Graphique à barres** : Permet de comparer les ventes entre les segments de manière plus granulaire.

```
[6]: # Ventes par segment
sales_by_segment = data.groupby('Segment')['Sales'].sum().reset_index()

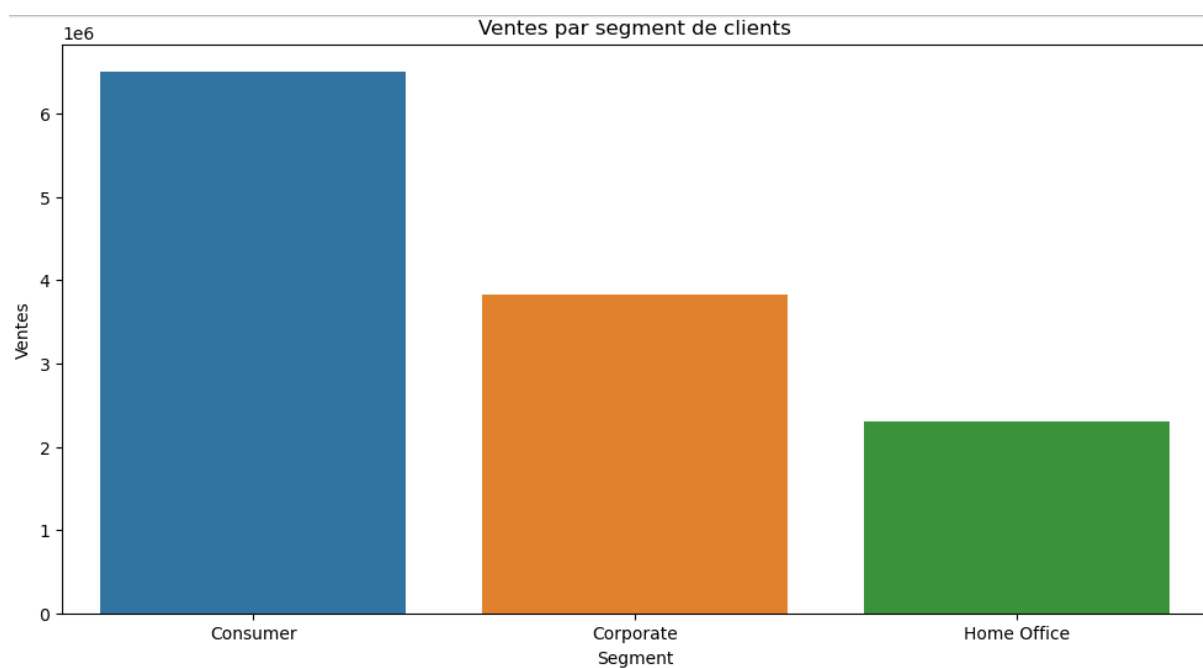
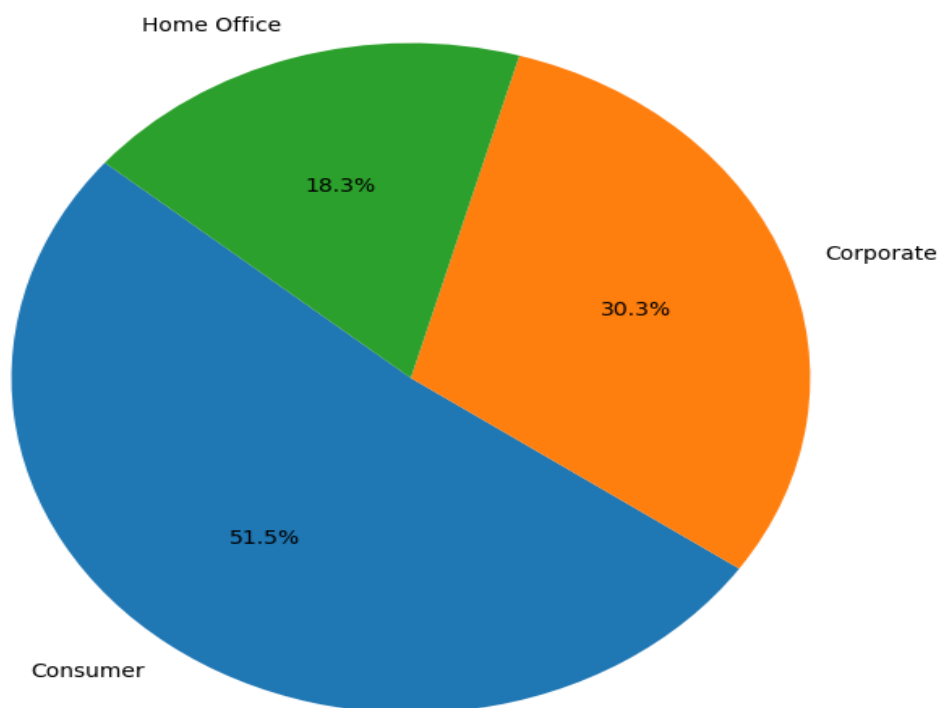
# Graphique circulaire
plt.figure(figsize=(8, 8))
plt.pie(sales_by_segment['Sales'], labels=sales_by_segment['Segment'], autopct='%1.1f%%', startangle=140)
plt.title('Contribution des segments de clients aux ventes')
plt.show()

# Graphique à barres
plt.figure(figsize=(12, 6))
sns.barplot(x='Segment', y='Sales', data=sales_by_segment)
plt.title('Ventes par segment de clients')
plt.xlabel('Segment')
plt.ylabel('Ventes')
plt.show()

# Paramètres de la solution
solution_5 = sales_by_segment.to_dict()
```

Fig5 : Codes python pour la cinquième question

Résultats graphiques



f. Comment varie le coût d'expédition par rapport aux ventes ?

Pour cette question, nous avons opté pour les graphiques suivants :

- **Graphique de dispersion** : Visualise directement la relation entre les coûts d'expédition et les ventes.

- **Graphique de dispersion avec ligne de régression** : La ligne de régression aide à comprendre la tendance générale de cette relation.

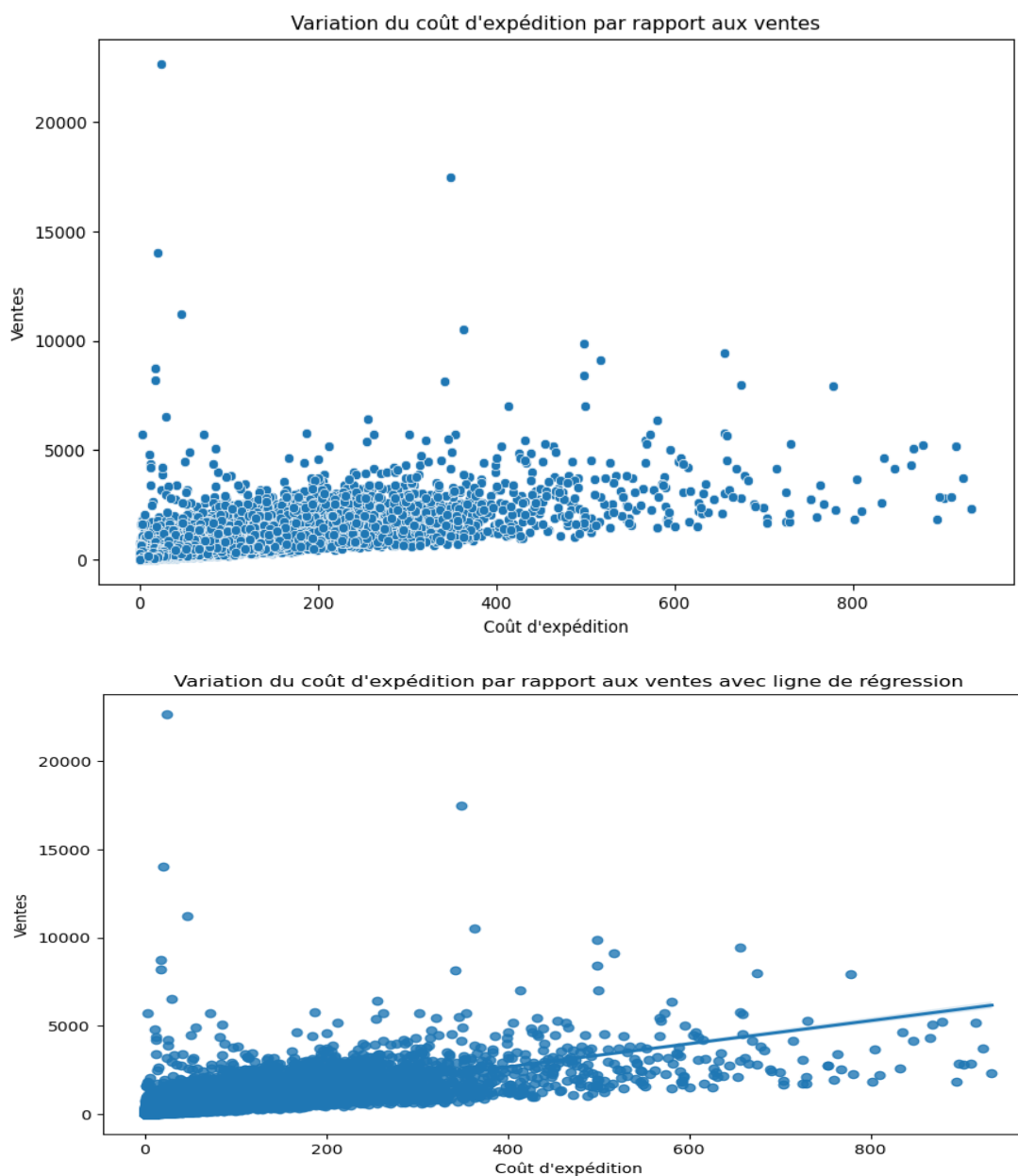
```
[7]: # Graphique de dispersion
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Shipping.Cost', y='Sales', data=data)
plt.title('Variation du coût d\'expédition par rapport aux ventes')
plt.xlabel('Coût d\'expédition')
plt.ylabel('Ventes')
plt.show()

# Graphique de dispersion avec une ligne de régression
plt.figure(figsize=(10, 6))
sns.regplot(x='Shipping.Cost', y='Sales', data=data)
plt.title('Variation du coût d\'expédition par rapport aux ventes avec ligne de régression')
plt.xlabel('Coût d\'expédition')
plt.ylabel('Ventes')
plt.show()

# Paramètres de la solution
shipping_sales_correlation = data[['Shipping.Cost', 'Sales']].corr().to_dict()
solution_6 = {'correlation': shipping_sales_correlation}
```

Fig6 : Codes python pour la sixième question

Résultats graphiques



Commentaires/Discussion sur les résultats

Les analyses visuelles révèlent des informations cruciales sur la répartition des ventes par catégorie de produits, l'évolution des ventes au fil du temps, et la performance des ventes dans différentes régions. Par exemple, les graphiques montrent que certaines catégories de produits dominent les ventes, tandis que d'autres nécessitent une stratégie marketing plus agressive. La tendance temporelle des ventes met en lumière des variations saisonnières et des pics potentiels liés à des événements spécifiques, aidant ainsi à la planification des inventaires et des campagnes promotionnelles. De plus, la performance régionale des ventes permet d'identifier les zones à fort potentiel pour des investissements supplémentaires ou des ajustements stratégiques.

La corrélation entre les remises et les profits, ainsi que la contribution des différents segments de clients aux ventes, sont des indicateurs essentiels pour optimiser la rentabilité. Les graphiques de dispersion et de régression montrent l'impact des remises sur les profits, indiquant qu'une remise trop élevée peut réduire significativement les bénéfices. Comprendre les contributions des segments de clients aide à cibler les efforts de marketing et de fidélisation, en maximisant les ventes des segments les plus rentables tout en développant des stratégies pour engager davantage les segments moins performants.

Enfin, l'analyse des coûts d'expédition par rapport aux ventes révèle des opportunités pour optimiser les opérations logistiques. Une corrélation significative entre ces variables peut indiquer des inefficacités à corriger pour améliorer la rentabilité. En réduisant les coûts d'expédition sans compromettre les ventes, l'entreprise peut accroître sa marge bénéficiaire. Ces insights multidimensionnels, obtenus grâce à une variété de graphiques, permettent de prendre des décisions éclairées pour renforcer les performances globales de l'entreprise.

Conclusion

Nous avons compris que les analyses visuelles fournissent des informations précieuses pour prendre des décisions éclairées en matière de marketing, de gestion des stocks, de logistique et de planification stratégique. En utilisant différents types de graphiques, nous obtenons une vue multidimensionnelle des données, permettant d'identifier les tendances, les anomalies et les opportunités d'amélioration.