# Introduction

La visualisation des données est une méthode essentielle dans l'analyse de données modernes. Elle permet de transformer des ensembles de données complexes et volumineux en graphiques compréhensibles et visuellement attrayants. Grâce à la visualisation, les analystes peuvent identifier rapidement des tendances, des modèles et des anomalies, facilitant ainsi la prise de décisions informées. Dans le cadre de ce travail pratique pour le cours de visualisation des données en Python, nous avons utilisé des outils de visualisation de données en Python, principalement matplotlib et seaborn, pour analyser un dataset contenant des informations sur l'âge, le genre, le salaire estimé, et les comportements d'achat. Ce rapport présente une analyse approfondie des données en utilisant différentes techniques de visualisation pour extraire et interpréter les informations clés.

# Cadre Théorique

### 1. Visualisation de Données

La visualisation des données consiste à représenter graphiquement des informations afin de rendre leur interprétation plus facile et plus rapide. Elle joue un rôle crucial dans l'analyse des données, permettant de :

* Identifier des tendances et des modèles
* Détecter des anomalies
* Communiquer efficacement des résultats complexes
* Faciliter la prise de décision

### 2. Visualisation des Données avec Python

Python est un langage de programmation populaire pour l'analyse des données grâce à ses nombreuses bibliothèques dédiées à la manipulation et à la visualisation des données. Parmi ces bibliothèques, matplotlib et seaborn se distinguent particulièrement.

### 3. matplotlib et seaborn

* **matplotlib** : C'est une bibliothèque de traçage complète qui permet de créer une grande variété de graphiques, allant des simples diagrammes à lignes aux visualisations plus complexes comme les diagrammes 3D.
* **seaborn** : Construite sur matplotlib, seaborn offre une interface de haut niveau pour dessiner des graphiques statistiques attractifs et informatifs. Elle simplifie la création de visualisations courantes telles que les cartes de chaleur, les boîtes à moustaches et les nuages de points.

# Implémentation Pratique

### Chargement des Données

Les données utilisées dans ce travail pratique proviennent de trois fichiers, dont les dépendances sont spécifiées dans requirements.txt :

* pandas
* numpy
* matplotlib
* seaborn

Les données sont chargées et manipulées à l'aide de pandas, et les visualisations sont créées avec matplotlib et seaborn.

### Préparation des Données

Les données sont nettoyées et préparées pour la visualisation. Cette étape inclut :

* La gestion des valeurs manquantes
* La normalisation des données
* La transformation des variables catégorielles

### Création des Visualisations

Plusieurs types de visualisations ont été réalisés pour explorer les données :

1. **Histogramme de la Distribution des Âges**

python

Copier le code

plt.figure(figsize=(10, 6))

sns.histplot(df['Age'], kde=True)

plt.title('Distribution des Âges')

plt.xlabel('Âge')

plt.ylabel('Fréquence')

plt.show()

1. **Diagramme à Barres pour la Répartition des Genres**

python

Copier le code

plt.figure(figsize=(10, 6))

sns.countplot(x='Gender', data=df)

plt.title('Répartition des Genres')

plt.xlabel('Genre')

plt.ylabel('Nombre')

plt.show()

1. **Boîte à Moustaches pour la Distribution des Salaires par Genre**

python

Copier le code

plt.figure(figsize=(10, 6))

sns.boxplot(x='Gender', y='EstimatedSalary', data=df)

plt.title('Distribution des Salaires Estimés par Genre')

plt.xlabel('Genre')

plt.ylabel('Salaire Estimé')

plt.show()

1. **Nuage de Points pour la Relation entre l'Âge et le Salaire Estimé**

python

Copier le code

plt.figure(figsize=(10, 6))

sns.scatterplot(x='Age', y='EstimatedSalary', hue='Purchased', data=df)

plt.title('Relation entre l\'Âge et le Salaire Estimé')

plt.xlabel('Âge')

plt.ylabel('Salaire Estimé')

plt.show()

1. **Diagramme Circulaire pour la Proportion des Achats**

python

Copier le code

plt.figure(figsize=(10, 6))

df['Purchased'].value\_counts().plot.pie(autopct='%1.1f%%', startangle=90, colors=['lightblue', 'lightgreen'])

plt.title('Proportion des Achats')

plt.ylabel('')

plt.show()

# Discussion

Les visualisations obtenues offrent plusieurs insights intéressants :

* L'histogramme montre que la majorité des individus se situent dans une certaine tranche d'âge, ce qui peut influencer les décisions marketing ciblées.
* Le diagramme à barres révèle une répartition presque égale des genres, ce qui permet de voir si certaines stratégies doivent être adaptées en fonction du genre.
* La boîte à moustaches indique des différences de salaires selon le genre, soulignant des questions potentiellement liées à l'équité salariale.
* Le nuage de points aide à identifier si les jeunes ou les personnes avec des salaires plus élevés sont plus enclins à faire des achats.
* Le diagramme circulaire montre la proportion de clients ayant effectué des achats, fournissant une vue d'ensemble sur l'efficacité des campagnes de marketing.

# Conclusion

La visualisation des données, comme démontré dans ce travail pratique, est un outil puissant et indispensable pour l'analyse des données. L'utilisation de matplotlib et seaborn en Python permet de créer des visualisations variées et informatives, facilitant ainsi la compréhension des données. Les visualisations réalisées ont permis d'identifier des tendances importantes, telles que la répartition des genres, la distribution des âges, et la relation entre l'âge, le salaire et les comportements d'achat. Ces insights peuvent aider à orienter les décisions stratégiques et les actions marketing. En somme, la visualisation des données ne se limite pas à embellir les rapports, elle joue un rôle crucial dans l'exploration et l'interprétation des données, rendant les analyses plus accessibles et exploitables pour un large éventail d'utilisateurs. Ce travail met en évidence l'importance de choisir les bonnes techniques de visualisation pour maximiser l'impact des analyses de données.

Contenu

[Introduction 1](#_Toc170562222)

[Cadre Théorique 2](#_Toc170562223)

[1. Visualisation de Données 2](#_Toc170562224)

[2. Visualisation des Données avec Python 2](#_Toc170562225)

[3. matplotlib et seaborn 2](#_Toc170562226)

[Implémentation Pratique 3](#_Toc170562227)

[Chargement des Données 3](#_Toc170562228)

[Préparation des Données 3](#_Toc170562229)

[Création des Visualisations 3](#_Toc170562230)

[Discussion 6](#_Toc170562231)

[Conclusion 7](#_Toc170562232)