

EXTRACTION, ANALYSE ET VISUALISATION DES DONNÉES D'UN SITE DE E-COMMERCE

29 Avril 2024



L'Équipe



Nidèle Sonna Jeufu

Data Analyst, Cybersecurity
Analyst



Léonce Thérèse Pidy pidy

Data Analyst, IT specialist,
Phd student



Ibrahim Halil Njoya

Data Analyst, Mechatronics
Designer



1

Contexte de l'étude

2

Problématique de l'étude

3

Objectif de l'étude

4

Methodologie

5

Résultats

6

Recommandations



1

Contexte de l'étude

Contexte du projet

Gold Tech Analysis, une startup spécialisée dans l'analyse de données, s'est engagée dans un projet d'importance cruciale pour une entreprise de E-commerce.

Le projet vise à résoudre les problématiques métier rencontrées quotidiennement par cette entreprise en extrayant, analysant et visualisant ses données.

En tant que **Data Analyst**, notre mission consiste à **extraire les données du site** de E-commerce, **reconstituer la base de données**, et **analyser ces données** afin de proposer des graphes décisionnels essentiels au client pour prendre des décisions stratégiques éclairées.

The background features a collage of financial and business-related images. On the left, a circular inset shows a line graph with multiple data series plotted on a grid. Above it, another circular inset shows a calculator and a document with a table. On the right, a circular inset shows a calculator and some papers. The entire background is decorated with large, light gray diagonal stripes.

2

Problématique de l'étude

Problématique

La grande majorité des clients de Gold Tech Analysis font face à une multitude de défis, notamment:

- La gestion efficace des stocks,
- L'optimisation des prix,
- L'amélioration de l'expérience client.

Notre client possède un site de e-commerce et rencontre également ces défis.



L'exploitation des données issues du site du client peut-elle apporter des solutions aux défis rencontrés?

Quelles décisions stratégiques ces données permettront-elles de prendre?

The background features a collage of financial documents, including line graphs and spreadsheets, with a large, light gray stylized number '3' overlaid. Two circular cutouts show close-ups of calculators and papers.

3

Objectif de l'étude

Objectif du projet

Exploiter les données du site de E-commerce pour aider le client à prendre des décisions stratégiques.



The background features a collage of financial and business-related images. On the left, a circular inset shows a line graph with multiple data series plotted on a grid. Above it, another circular inset shows a calculator and a spreadsheet. On the right, a larger circular inset shows a calculator and some papers. The central text is framed by a teal border. The overall design is modern and professional, using a teal and grey color scheme.

4

Méthodologie

Méthodologie



Ciblage précis

Planification et organisation du travail



Extraction des données

Utilisation de la bibliothèque Python BeautifulSoup pour extraire les données du site de E-commerce de manière automatisée.



Reconstitution de la BD

Reconstitution et mise en place de la base de données à l'aide du langage SQL pour assurer une gestion efficace des données.



Analyse et visualisation des données

Analyse descriptive des données pour comprendre les tendances et les comportements des clients.

Visualisation des données à l'aide de graphiques décisionnels pour faciliter la prise de décision du client.

Extraction des données

Beautiful Soup

Cette bibliothèque nous a permis d'analyser le HTML des pages du site et d'extraire les informations pertinentes telles que le titre du livre, sa catégorie, sa note, sa quantité en stock et ses prix...

Requests

Cette bibliothèque nous a permis de récupérer le contenu HTML des différentes pages web.



Extraction des données

CSV

Cette bibliothèque nous a permis d'effectuer la manipulation de fichiers CSV, ce qui peut être utile pour stocker les données extraites dans un format tabulaire.

re

Cette bibliothèque nous a permis d'effectuer des opérations de recherche et de traitement de chaînes, ce qui peut être utile lors de l'extraction de données spécifiques à partir du contenu HTML.

Reconstitution et mise en place de la base de données

SQL

Une fois les données extraites, nous les avons reconstituées et intégrées dans une base de données SQL.

MySQL

Nous avons utilisé la bibliothèque `mysql.connector` pour établir une connexion à la base de données MySQL et insérer les données extraites dans des tables appropriées.

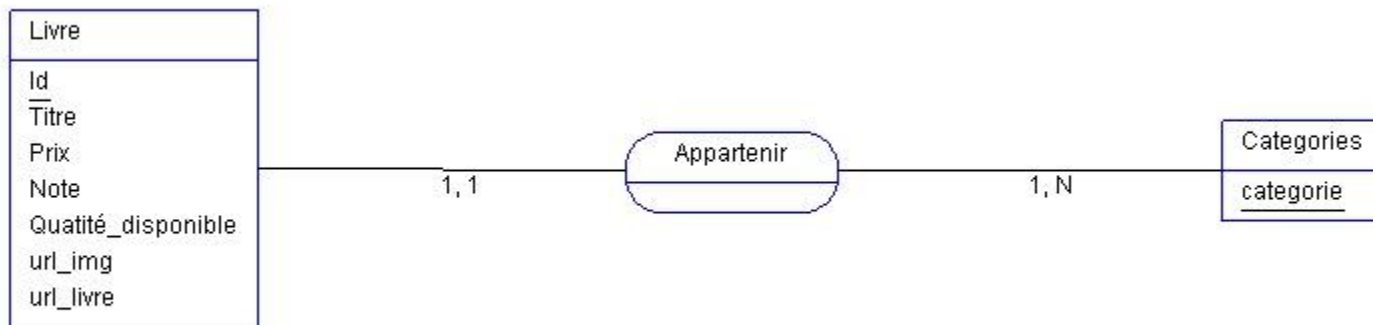
Reconstitution et mise en place de la base de données

`urllib.parse.urljoin, urllib.parse.urlparse`

Ces imports ont principalement été utilisés lors de la construction d'URL absolues à partir d'URL relatives et pour l'analyse d'URL, ce qui peut être utile lors de la reconstitution de la base de données en construisant des liens vers les pages individuelles des livres.

Représentations schématiques de votre base de données

Modèle conceptuel des données (MCD)



Représentations schématiques de votre base de données

Modèle logique des données (MCD)

```
Livre (Id_Livre, Titre_Livre, Prix_Livre, Note_Livre, Quatité_disponible_Livre, url_img_Livre, url_livre_Livre, #categorie_Categories)  
Categories (categorie_Categories)
```

Analyse descriptive des données

Après avoir reconstitué la base de données, nous avons procédé à une analyse descriptive des données pour comprendre les tendances et les caractéristiques des livres disponibles sur le site de E-commerce.

Pandas

Nous avons utilisé des fonctionnalités de Pandas pour explorer les données, calculer des statistiques descriptives et identifier des insights clés tels que les catégories les plus populaires, les prix moyens et la distribution des notes attribuées aux livres.

Analyse descriptive des données

Numpy

Cette la bibliothèque a été utilisé pour effectuer des opérations mathématiques et des manipulations de tableau sur les données.

Matplotlib

Cette la bibliothèque a été utilisé pour la visualisation des données car le module matplotlib.pyplot est utilisé pour créer des graphiques et des visualisations à partir des données extraites et analysées.

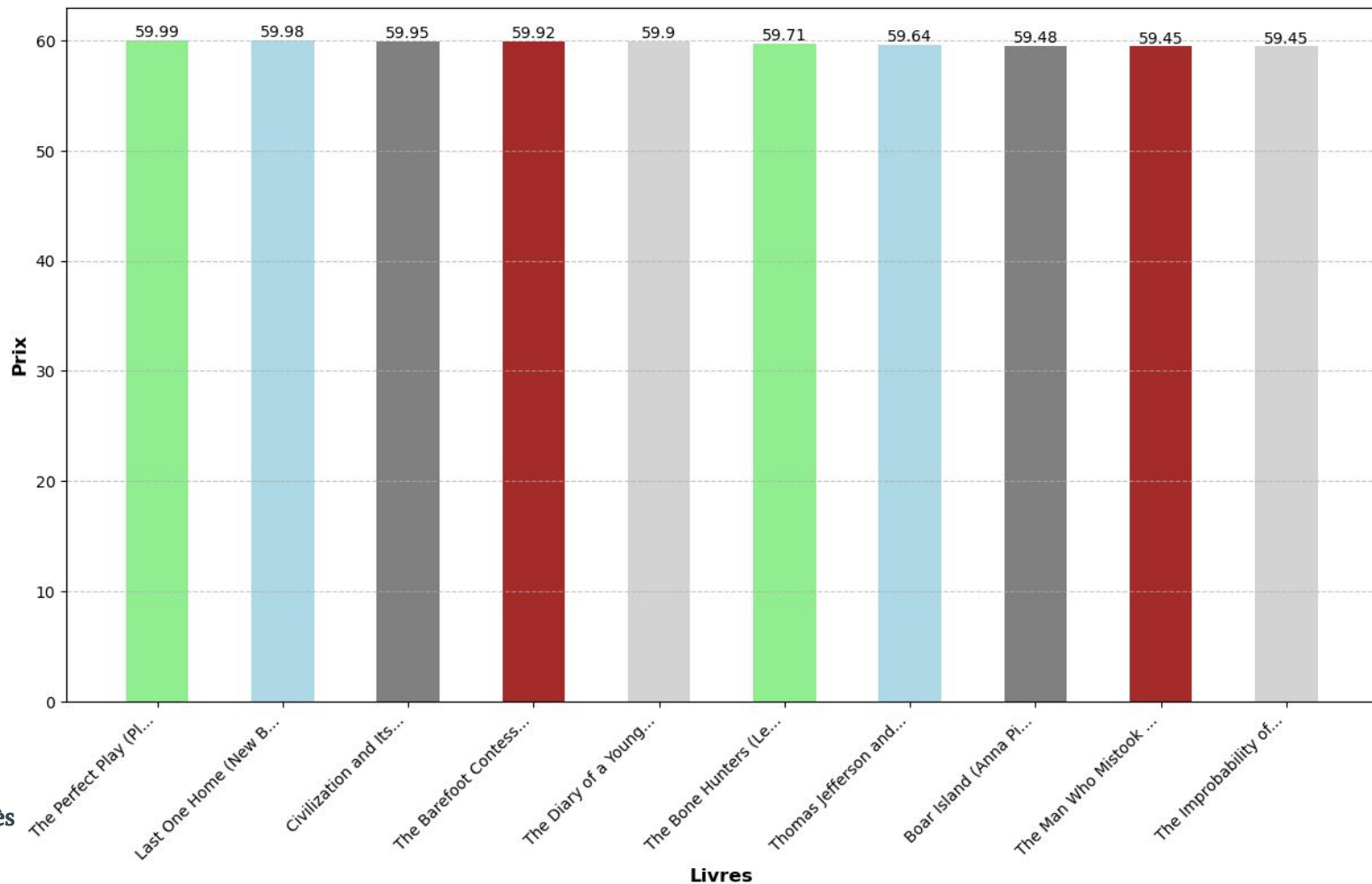


5

Résultats

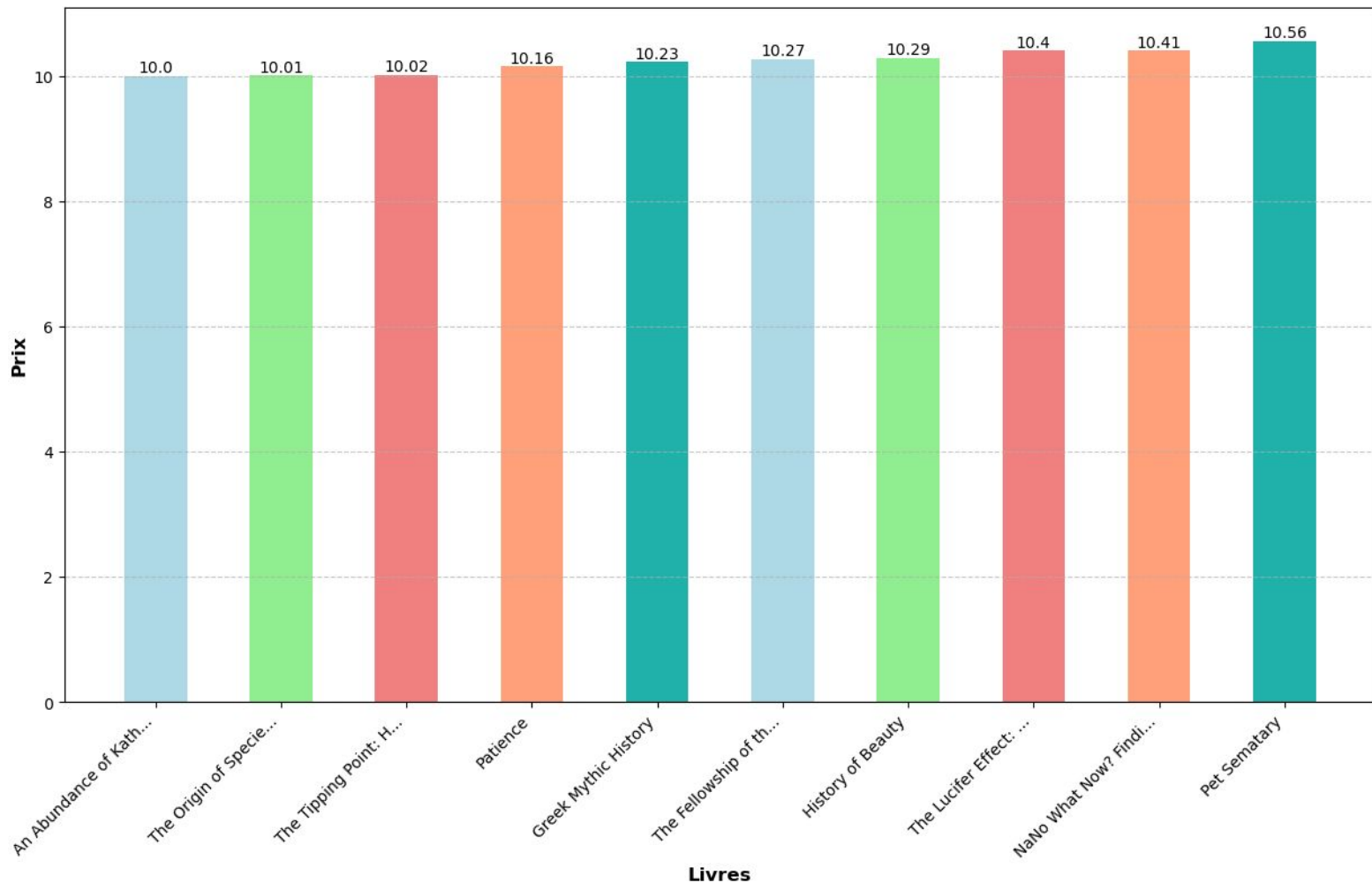


Les 10 produits les plus coûteux



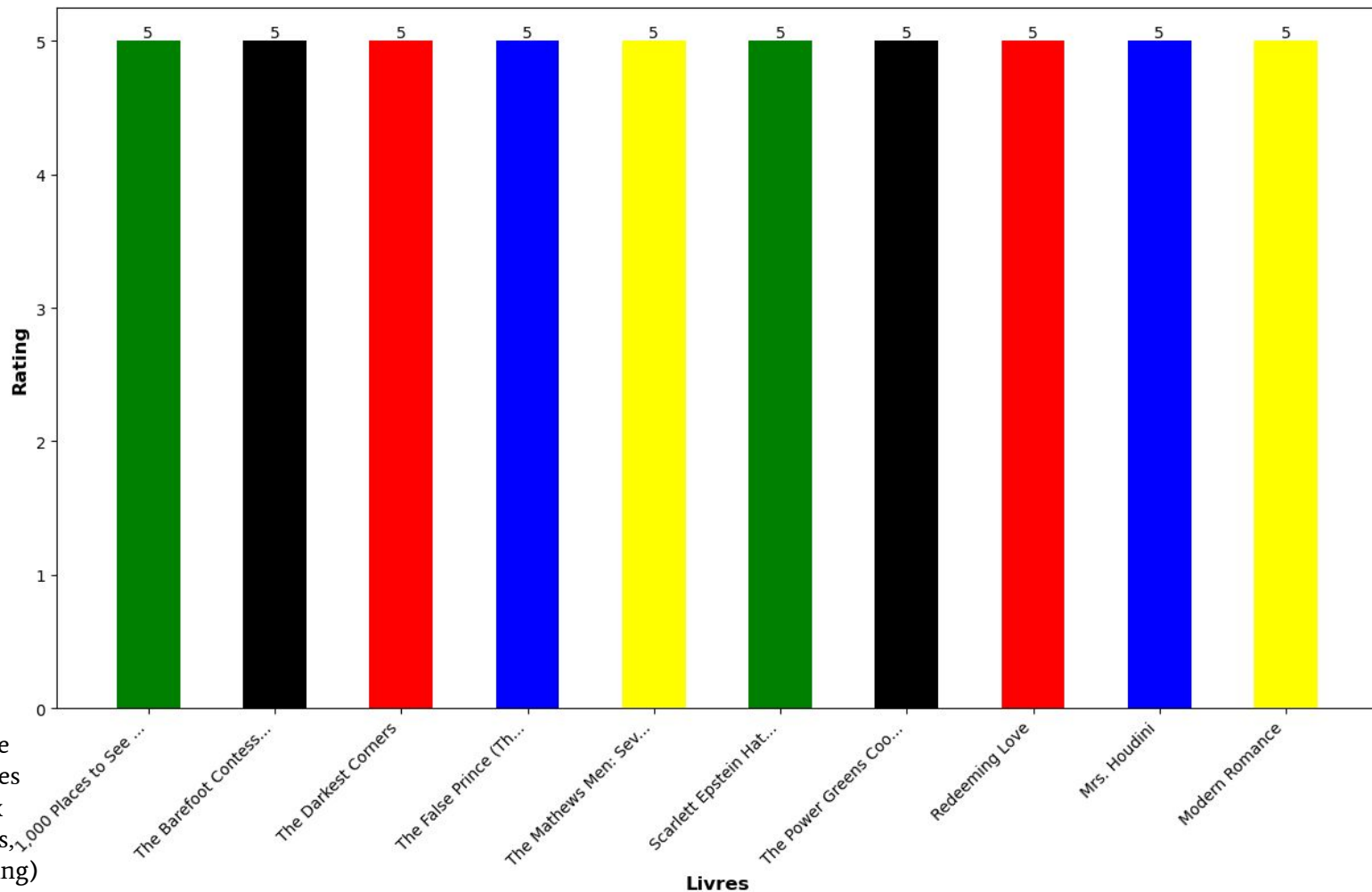
Interprétation 1: Les cinq premiers livres affichent des prix très proches les uns des autres

Les 10 produits les moins coûteux



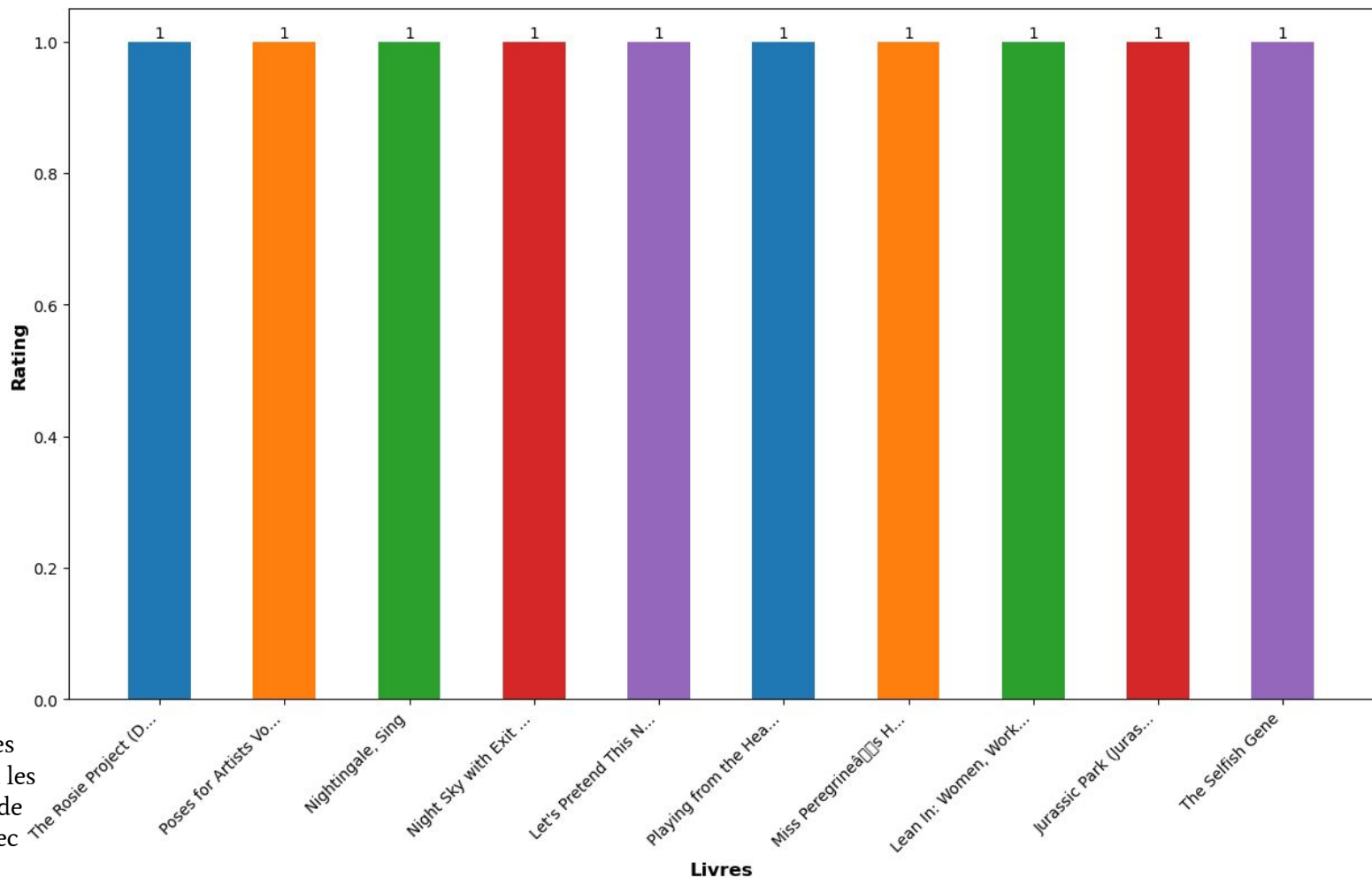
Interprétation 1: Les cinq premiers livres présentent des prix relativement bas.

Les 10 produits les plus recommandés



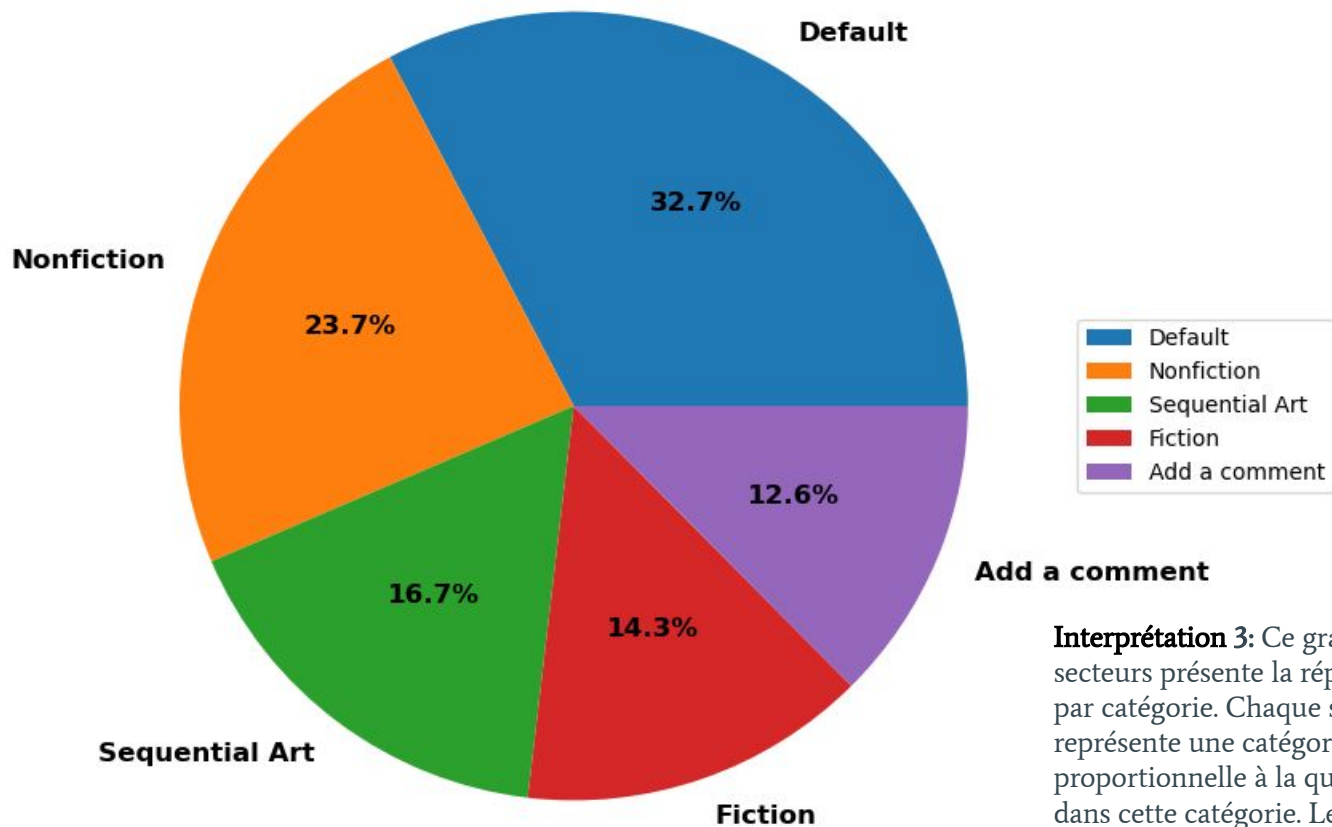
Interprétation 2: Ce graphique affiche les dix livres les mieux notés par les clients, avec leur note (rating) correspondante

Les 10 produits les moins recommandés



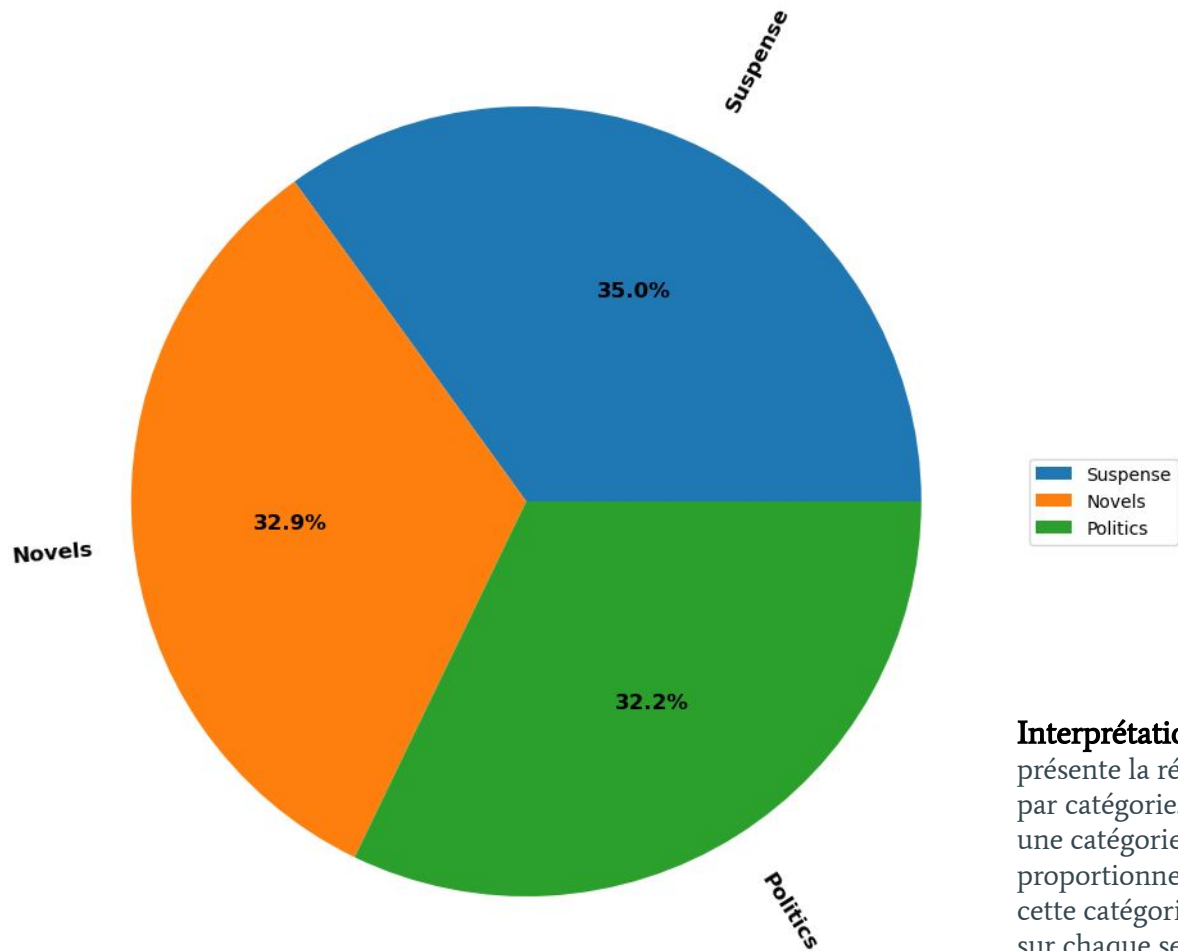
Interprétation 2: Ce graphique présente les cinq livres ayant reçu les notes les plus basses de la part des clients, avec leur note (rating) correspondante.

Les 5 catégories ayant le plus de produits en stock



Interprétation 3: Ce graphique en secteurs présente la répartition du stock par catégorie. Chaque secteur représente une catégorie, et sa taille est proportionnelle à la quantité de stock dans cette catégorie. Les pourcentages indiqués sur chaque secteur représentent la part de chaque catégorie dans l'ensemble du stock.

Les trois catégories ayant les produits les plus coûteux



Interprétation 4: Ce graphique en secteurs présente la répartition moyenne des prix par catégorie. Chaque secteur représente une catégorie, et sa taille est proportionnelle à la moyenne des prix dans cette catégorie. Les pourcentages indiqués sur chaque secteur représentent la part de chaque catégorie dans l'ensemble des prix moyens.



6

Recommendations



Recommandations

Interpretation 1 (produits les plus et moins coûteux)

- Les cinq premiers livres affichent des prix très proches les uns des autres.
- Les cinq premiers livres présentent des prix relativement bas.

Recommandation 1

- Mettre en avant ces livres comme des options économiques pour attirer les clients soucieux des prix.
- Assurer une diversité de genres pour répondre aux différents intérêts des clients.

Interpretation 2 (produits recommandés)

- D'un côté, nous avons les dix livres les mieux notés par les clients, avec leur note (rating) correspondante et de l'autre les cinq livres ayant reçu les notes les plus basses de la part des clients, avec leur note (rating) correspondante.

Recommandation 2

- Analyser les caractéristiques communes des livres bien notés pour comprendre ce qui plaît aux clients.
- Mettre en avant ces livres dans les promotions ou les recommandations personnalisées pour attirer l'attention des clients sur les produits populaires.
- Examiner les critiques des clients pour comprendre les raisons des notes basses.
- Évaluer si ces livres rencontrent les attentes des clients et s'ils nécessitent des améliorations, telles que des révisions du contenu ou des ajustements de prix.

Recommandations

Interpretation 3

(Répartition du stock par catégorie)

- Ce graphique en secteurs présente la répartition du stock par catégorie. Chaque secteur représente une catégorie, et sa taille est proportionnelle à la quantité de stock dans cette catégorie. Les pourcentages indiqués sur chaque secteur représentent la part de chaque catégorie dans l'ensemble du stock.

Recommandation 3

- Identifier les catégories avec le plus grand nombre de livres en stock pour comprendre les préférences des clients.
- Utiliser ces informations pour planifier les promotions et les campagnes marketing, en mettant en avant les catégories les plus demandées pour attirer les clients.

Interpretation 4

(Répartition moyenne des prix par catégorie)

- Ce graphique en secteurs présente la répartition moyenne des prix par catégorie. Chaque secteur représente une catégorie, et sa taille est proportionnelle à la moyenne des prix dans cette catégorie. Les pourcentages indiqués sur chaque secteur représentent la part de chaque catégorie dans l'ensemble des prix moyens.

Recommandation 4

- Identifier les catégories avec les produits les plus coûteux pour comprendre les tendances de prix et les préférences des clients en matière de produits haut de gamme.
- Adapter les stratégies de tarification et de marketing en fonction de ces données, en mettant en valeur les produits de catégories populaires mais également rentables.
- Explorer des opportunités de segmentation de marché pour cibler les clients intéressés par des produits de luxe ou haut de gamme.

Recommandations

En conclusion, les recommandations présentées offrent des informations précieuses pour orienter les décisions stratégiques dans le domaine de la vente de livres en ligne. Les points clés à retenir sont :

- **Compréhension des tendances de prix**
- **Identification des catégories populaires**
- **Segmentation du marché**
- **Optimisation des prix**

Livrables et Critères de Performance

Les livrables attendus pour ce projet comprennent un dépôt Git contenant le code source de nos scripts d'extraction et d'analyse, un rapport complet décrivant notre méthodologie, nos résultats et nos recommandations, ainsi qu'un fichier SQL de la base de données reconstituée.

Merci pour votre attention

