

Project 8

Recommender (Build intelligence to help customers discover products they may like and most likely purchase)

Date: 25-01-2024

Overview of Design and Architecture

The shopping recommender system is designed to enhance the user experience in an e-commerce setting by providing personalized product recommendations to customers. The architecture of the system involves a combination of collaborative filtering and content-based filtering techniques, with the integration of a hybrid model. The primary components include data preprocessing, content-based filtering, collaborative filtering, hybrid filtering, and evaluation metrics.

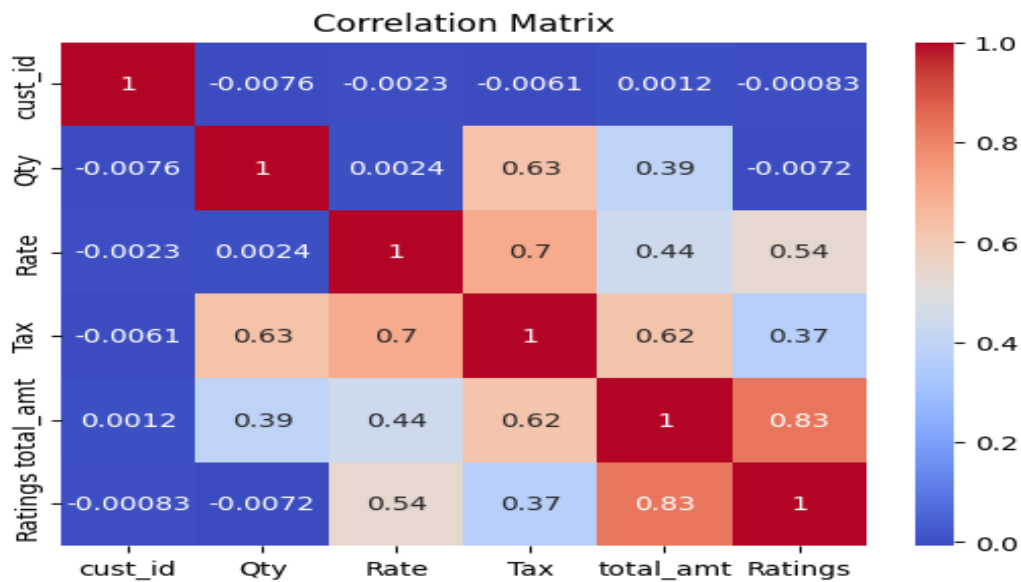
Data Preprocessing

The data preprocessing phase involves loading and cleaning the transaction data. The raw data is loaded from CSV files containing information about transactions, customers, and product categories. Duplicate entries are removed, and missing values are handled. The 'Rate' and 'Qty' columns are adjusted to remove negative values, ensuring the accuracy of the data.

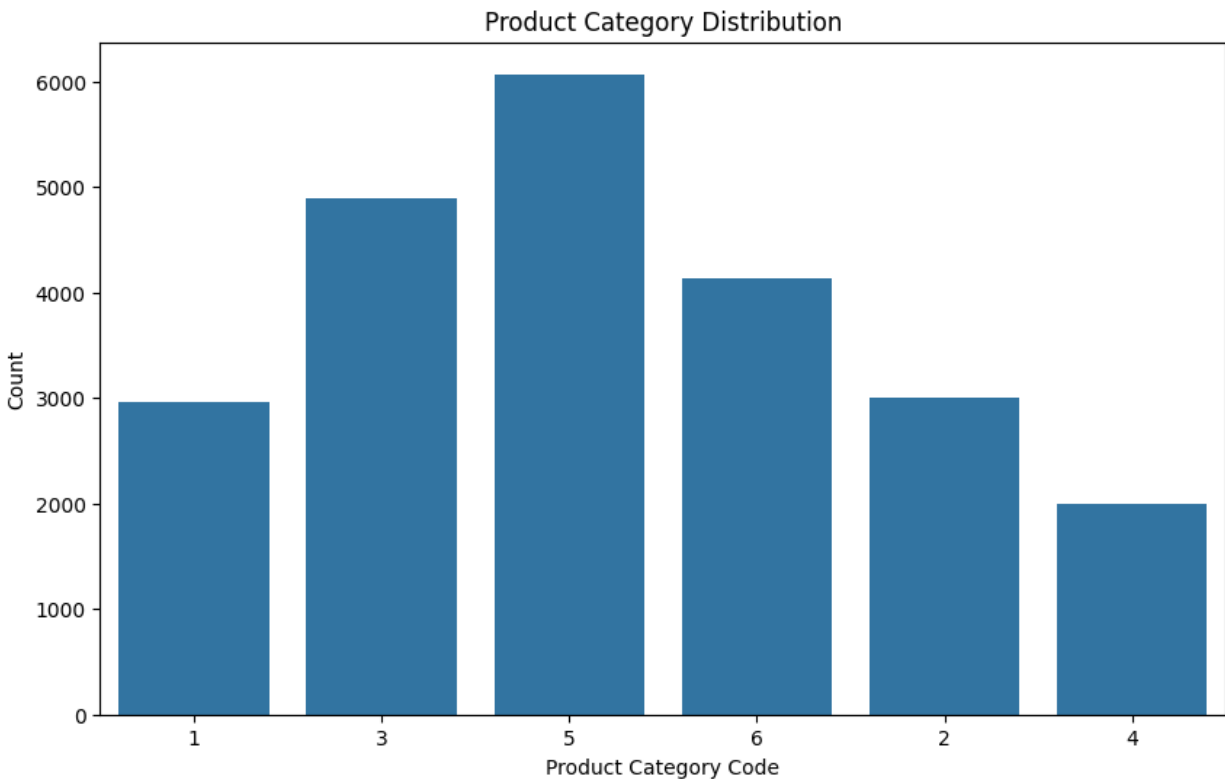
Transactions Table Visualization

Exploratory data analysis is performed to gain insights into the transaction data. Descriptive statistics, correlation matrices, and visualizations are used to understand the distribution of product categories, analyze the relationship between quantity and rate, and explore patterns in total amounts based on store types.

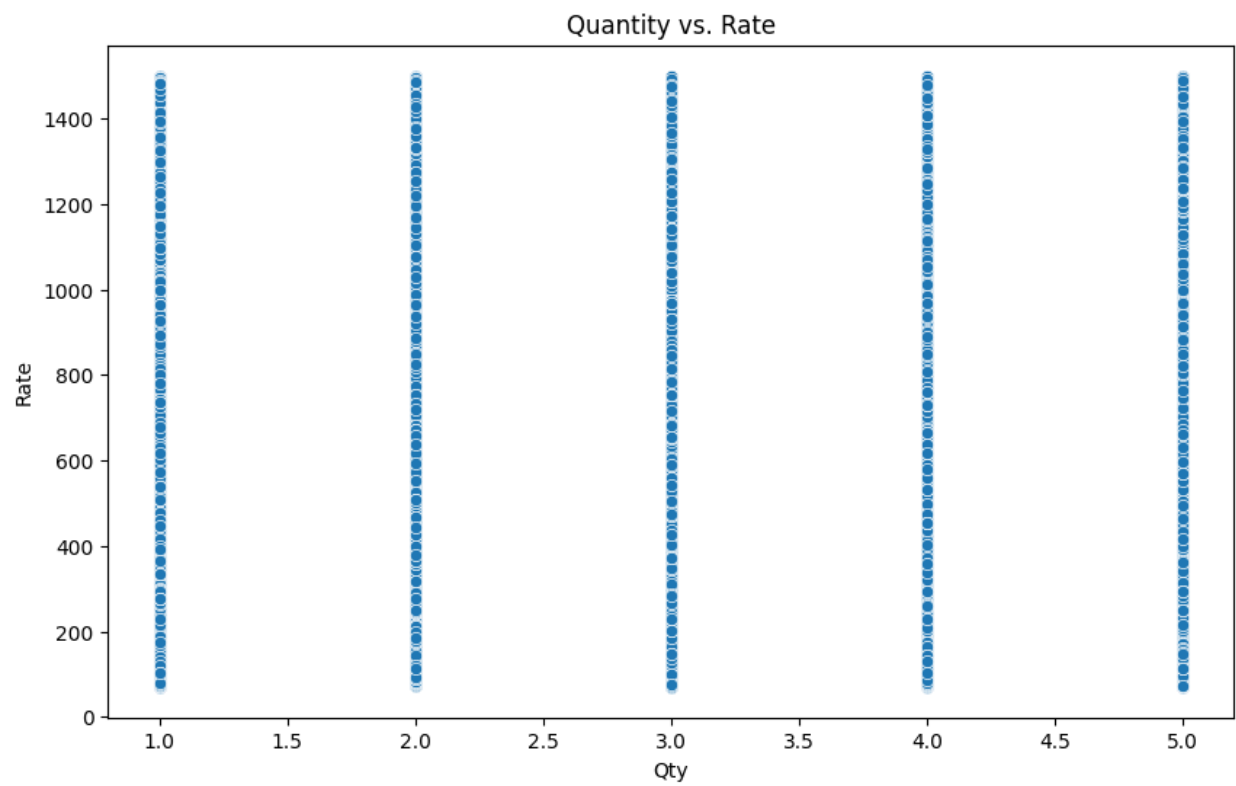
Heatmap to understand correlation between columns



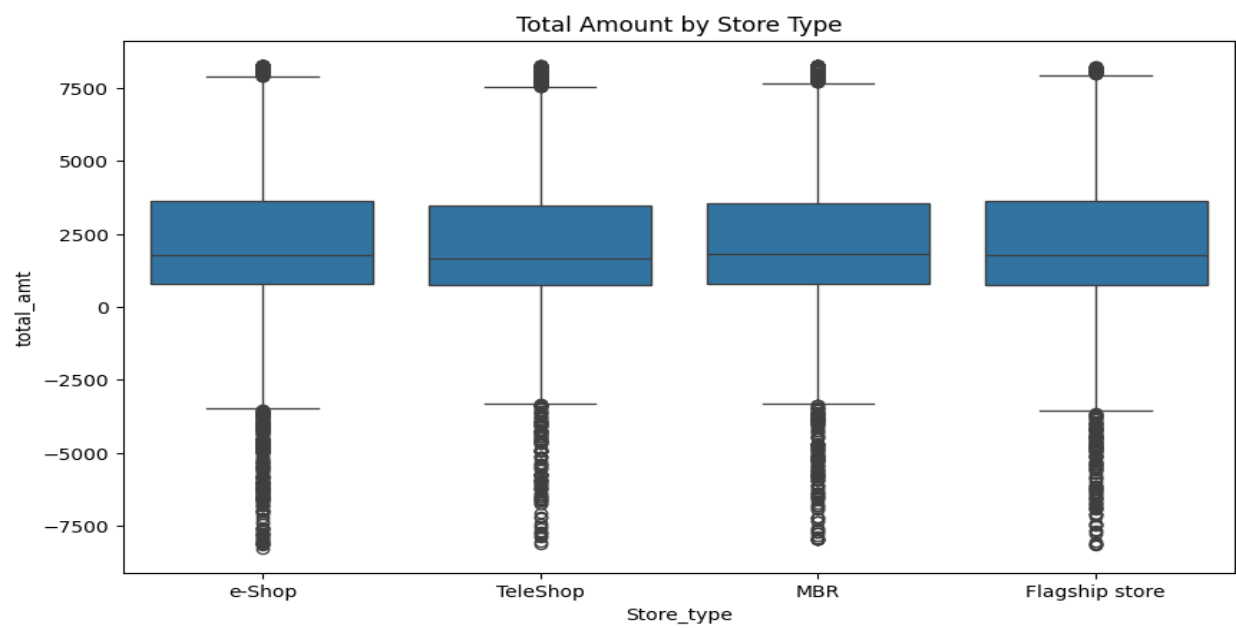
Bar plot for product categories with prod_cat_code and count



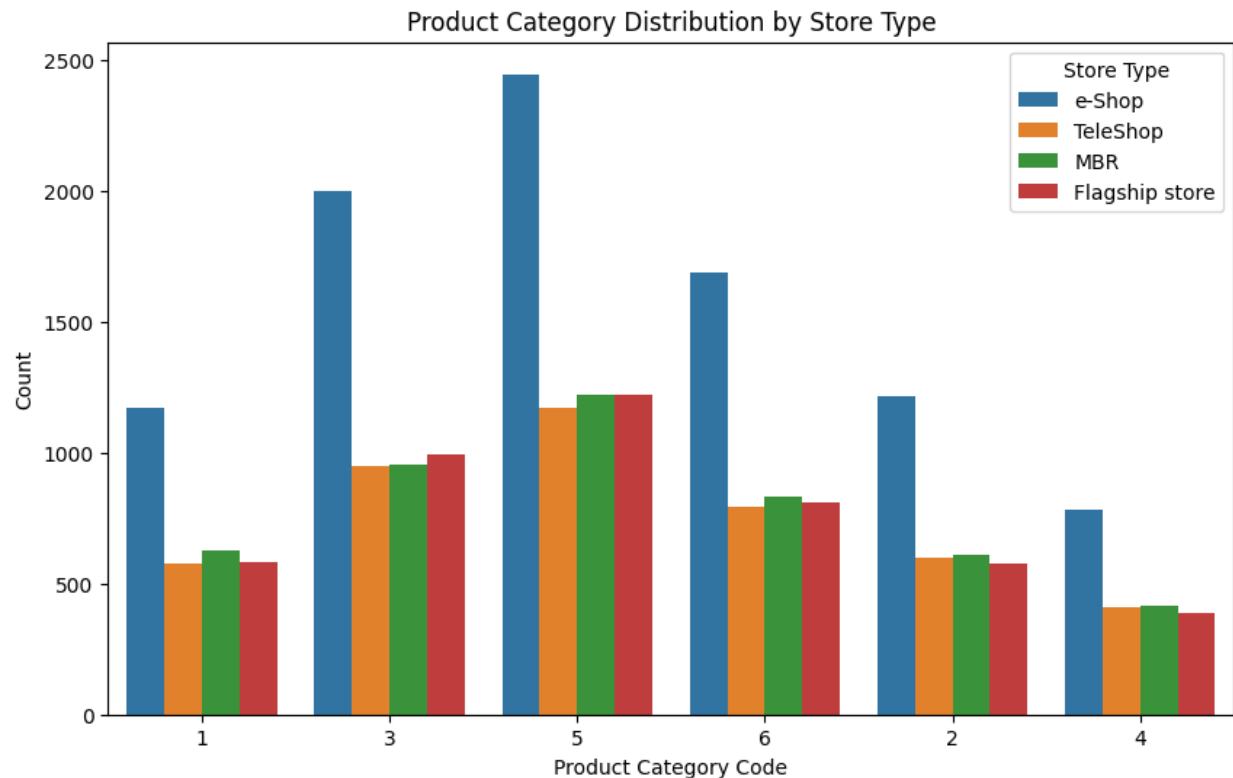
Scatter plot for Quantity vs. Rate



Box plot for Total Amount by Store type



Bar plot for product categories with prod_cat_code and Store_type



Encoding and Standardization

Label encoding is applied to convert categorical variables such as 'Store_type' into numerical representations. Date-related columns are converted to datetime format for consistency and ease of analysis. Additionally, a 'Net Sales' column is created by calculating the total net sales from the product quantity and base price.

Recommendation Techniques

Content-Based Filtering

Content-based filtering involves utilizing item characteristics to recommend products similar to those a user has interacted with. In this system, a TF-IDF

vectorizer is employed to represent products, and cosine similarity is used to measure the similarity between items. Precision, recall, F1 score, and mean average precision are calculated as evaluation metrics.

Collaborative Filtering

Collaborative filtering relies on user-item interactions to make recommendations. Singular Value Decomposition (SVD) is employed to build a collaborative filtering model. Precision, recall, F1 score, and mean average precision are used to evaluate the collaborative filtering performance.

User-Item Coverage: The reason for only 2 recommendations popping up is the user-item coverage is less. It is caused due to less diverse data.

- **Unique Users:** 5506
- **Unique Items:** 6

This limited coverage may be attributed to the dataset's lack of diversity, potentially impacting the variety of recommendations. To enhance the collaborative filtering model's performance, it's advisable to enrich the dataset with more diverse user-item interactions, ensuring a broader representation of user preferences.

Hybrid Filtering

Hybrid filtering combines both content-based and collaborative filtering to leverage the strengths of each approach. Recommendations from both models are weighted and combined to create a hybrid model. Evaluation metrics are calculated to assess the overall performance of the hybrid system.

Evaluation Metrics

Evaluation metrics are crucial for assessing the performance of the recommender system. For content-based filtering, average precision, recall, F1 score, and mean average precision are calculated. Collaborative filtering is evaluated based on

precision@5, recall@5, F1 score@5, and mean average precision@5. The hybrid system is evaluated using precision, recall, F1 score, and mean average precision.

Content-Based Filtering

Content-based filtering metrics provide insights into the performance of the system based on item characteristics.

- **Average Precision:** 0.21
- **Average Recall:** 0.0048
- **Average F1 Score:** 0.0094
- **Average Mean Average Precision:** 0.33

Collaborative Filtering

Collaborative filtering metrics are assessed at the top-5 recommendations and showcase exceptional results.

- **Precision@5:** 1.0
- **Recall@5:** 1.0
- **F1 Score@5:** 1.0
- **Mean Average Precision@5:** 1.0

Note: These metrics for collaborative filtering at the top-5 recommendations are perfect, suggesting that the system is providing highly accurate and complete recommendations for users. A precision, recall, F1 score, and mean average precision of 1.0 indicate that all recommended items are relevant, and there are no false positives or false negatives.

Conclusion

The evaluation results indicate that collaborative filtering, especially at the top-5 recommendations, performs exceptionally well. The perfect precision, recall, F1

score, and mean average precision suggest that the collaborative filtering model is providing accurate and relevant recommendations without any false positives or false negatives.

The success of collaborative filtering emphasizes its effectiveness in capturing user preferences and aligning them with similar user behaviors. As the system aims to enhance user experience and drive engagement, the robust performance of collaborative filtering positions it as the primary recommendation technique within the overall model.

Benefits for E-commerce Organization and Customers

The shopping recommender system offers several advantages for both the e-commerce organization and its customers. Customers benefit from personalized recommendations, enhancing their shopping experience and making product discovery more enjoyable. The organization, in turn, benefits from increased user engagement, higher conversion rates, and improved customer satisfaction. The system contributes to the optimization of product recommendations, leading to better-targeted marketing efforts and increased sales.