Customer Recommendation With Minimized Bank Transaction

Preshika Basnet| Professors: Dr. Yangyang Tao School of Computing and Analytics | Northern Kentucky University

Abstract

Personalized customer recommendation has garnered significant interest in recent years due to its potential to revolutionize business strategies. By leveraging recommendation systems, businesses can effectively target potential customers, enhance user experience, and ultimately boost revenue. In this study, we explore the application of traditional collaborative filtering algorithms to a streamlined bank transaction dataset. The primary objective is to assess the performance of these algorithms when detailed customer information is unavailable, a common scenario in real-world datasets with privacy constraints. Our experimental evaluation focuses on three prominent collaborative filtering techniques: User-Based Collaborative Filtering, Item-Based Collaborative Filtering, and Model-Based Collaborative Filtering. By conducting a comparative analysis, we demonstrate how each approach performs under the given dataset limitations. The findings of this study provide valuable insights into the adaptability and effectiveness of traditional collaborative filtering methods, offering practical implications for designing recommendation systems in similar constrained environments.

Dataset Overview

The dataset we are using is a minimized bank transaction data. It contains 25 columns and more than 15 million records. The whole dataset is in Figure 1 bubble plot according to the geographic locations..

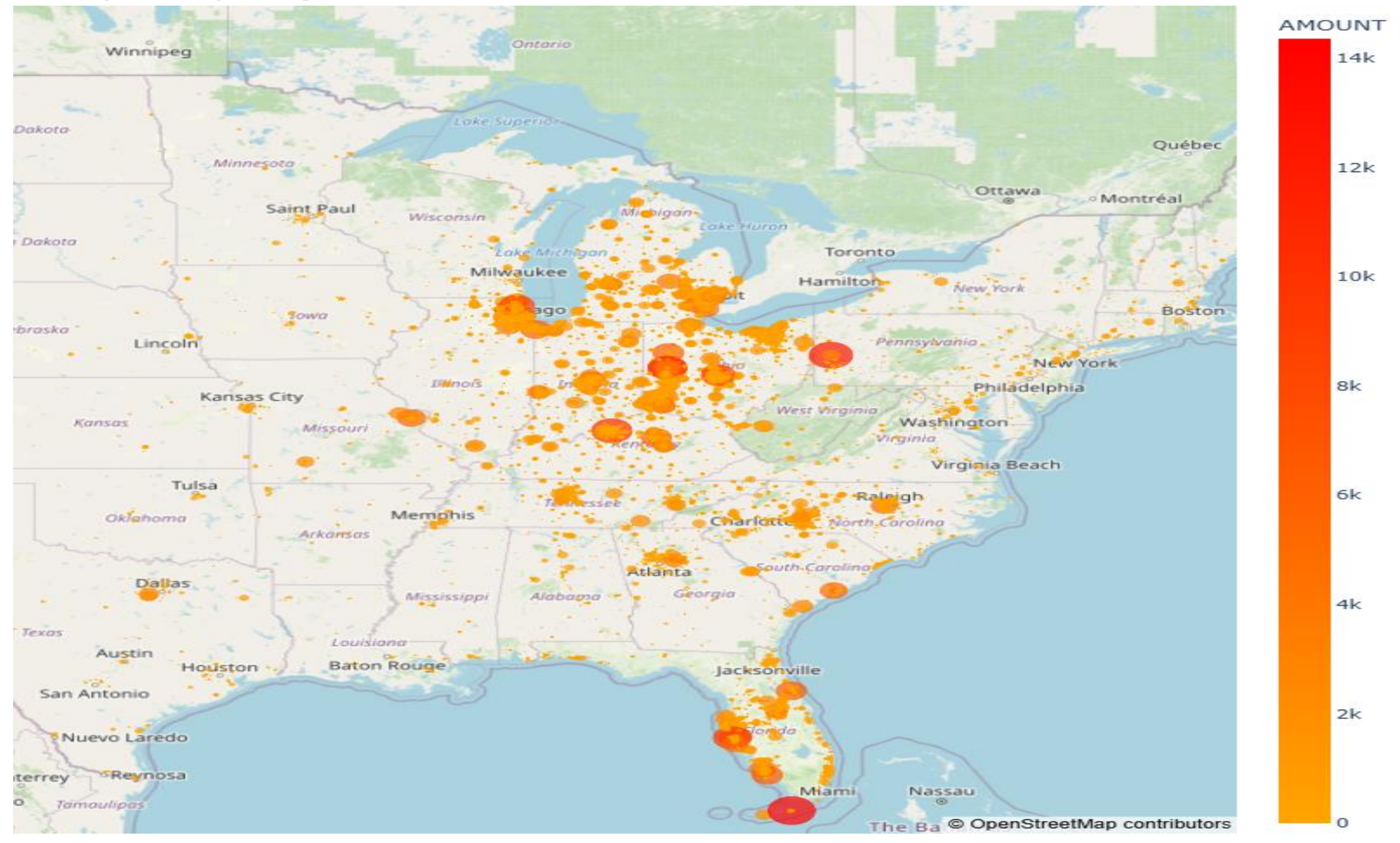


Figure 1. minimized bank transaction data.

Dataset (Cont.)

- The dataset contains the merchandize information and the location information. But there is no user information in this dataset. There are large amount of missing value in the email and website columns. We are working on the cleaned dataset which contains around 170,000 records. Most of the transaction are small transactions between [0,2000] in Figure 2. 98% of small transactions are below 250 in Figure 3.

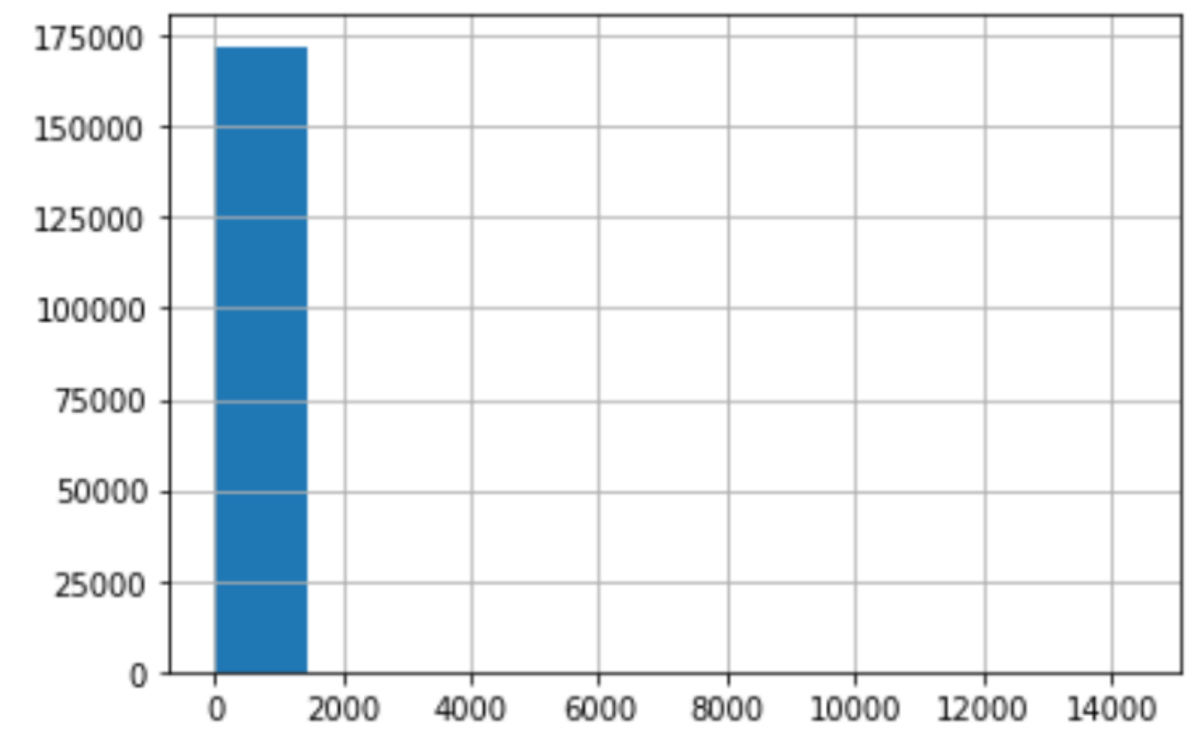


Figure 2. small transactions between [$0,$2000].

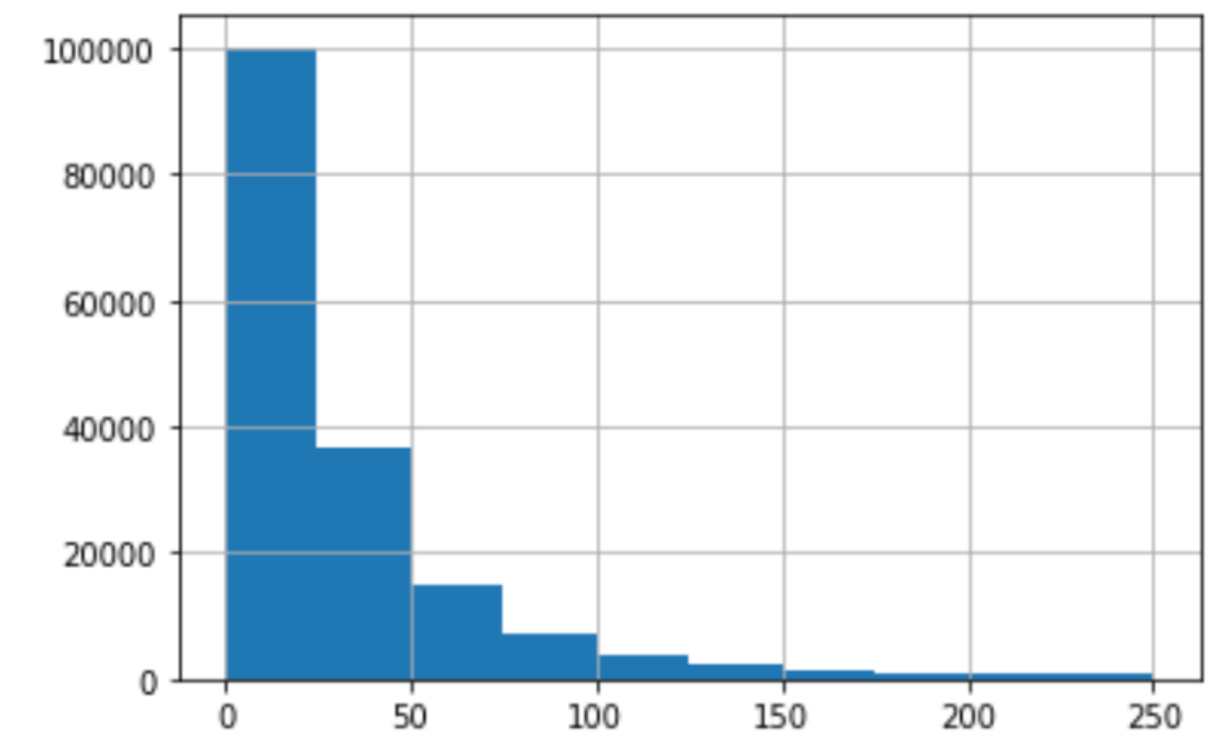


Figure 3. small transactions between below $250.

- We analyze the spending according to the transaction amount in the below table.
- High Spending: Significant expenses include Credit Card Payments ($4,453), followed by Checks ($448), and Business Miscellaneous ($524). Other notable categories are Insurance ($263) and Refunds/Adjustments ($470).
- Moderate Spending: Includes Online Services ($93), Clothing ($108), Shoes ($113), Utilities ($100), and Child/Dependent Expenses (comparable to others in this range).
- Low Spending: Categories such as Restaurants/Dining, Electronics, Retirement Contributions, Dues and Subscriptions, and Deposits all fall under $35, indicating minimal expenditure in these areas.

Transaction Amount Category	Merchant category	Average amount
High Spending	Insurance/Refunds/Adjustments/Credit Card Payments/Checks/Business Miscellaneous	263/470/524/4453/448
Moderate spending	Online Services/Clothing/Shoes/Utilities/Child/Dependent Expenses	93/108/113/100
Low Spending	Restaurants/Dining/Electronics/Retirement Contributions/Dues and Subscriptions/Deposits	35/34/35/21/34

Table 1. small transactions categories (High, Moderate, and Low).

Methods

Three methods are used perform customer recommendation on the cleaned dataset:

- User-Based Collaborative Filtering: This method identifies users who are similar to the active user and makes recommendations based on how much similar users spending. The assumption is: "If users X and Y have similar spending, and X liked an item, Y might also like it." In this study, we use cosine similarity as the main approach for calculating the similarity.
- Item-Based Collaborative Filtering: Instead of focusing on users, item-based CF focuses on finding relationships between items. The assumption is: "If users liked a particular merchant, they may also like similar merchants." Again, cosine similarity is used for calculating the similarity.
- Model-Based Collaborative Filtering: This method builds predictive models using machine learning algorithms to predict user preferences based on past interactions. Unlike user-based and item-based CF, model-based methods focus on building a mathematical representation of users and items. In this study we use alternating least squares (ALS) as the model for recommendation.

Results / Conclusion

We found some interesting finding in the dataset. There are certain categories of spending that requires longer processing time. Such as restaurant and dining. Especially, the Fact Food and Restaurants among the detailed categories. The result is shown in Figure 4 (a) and (b).

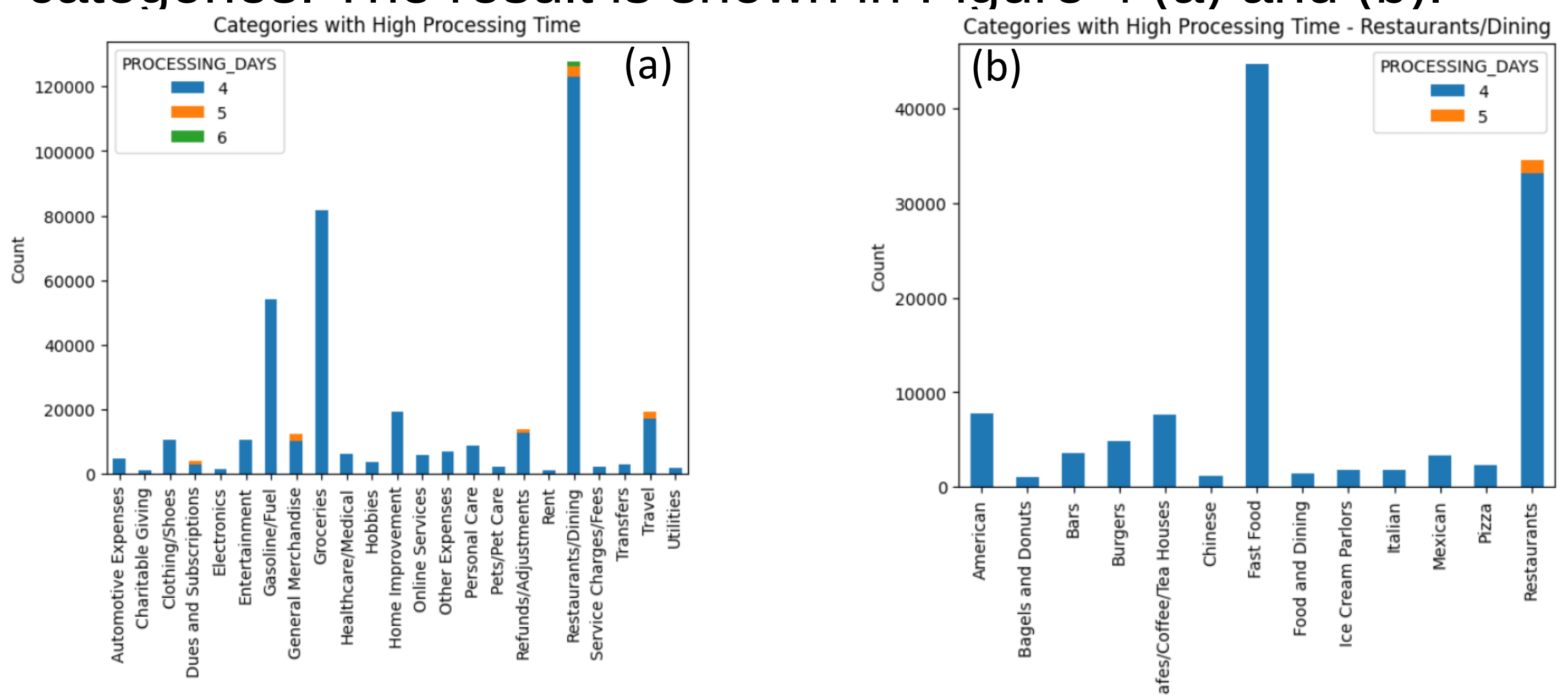


Figure 4 spending categories with high processing time.

The RMSE evaluation of Collaborative Filtering (CF) models highlights that User-Based CF is the most accurate, with the lowest error Values at 0.01. Model-Based CF follows, while Item-Based CF has the highest RMSE, indicating fewer effective predictions.

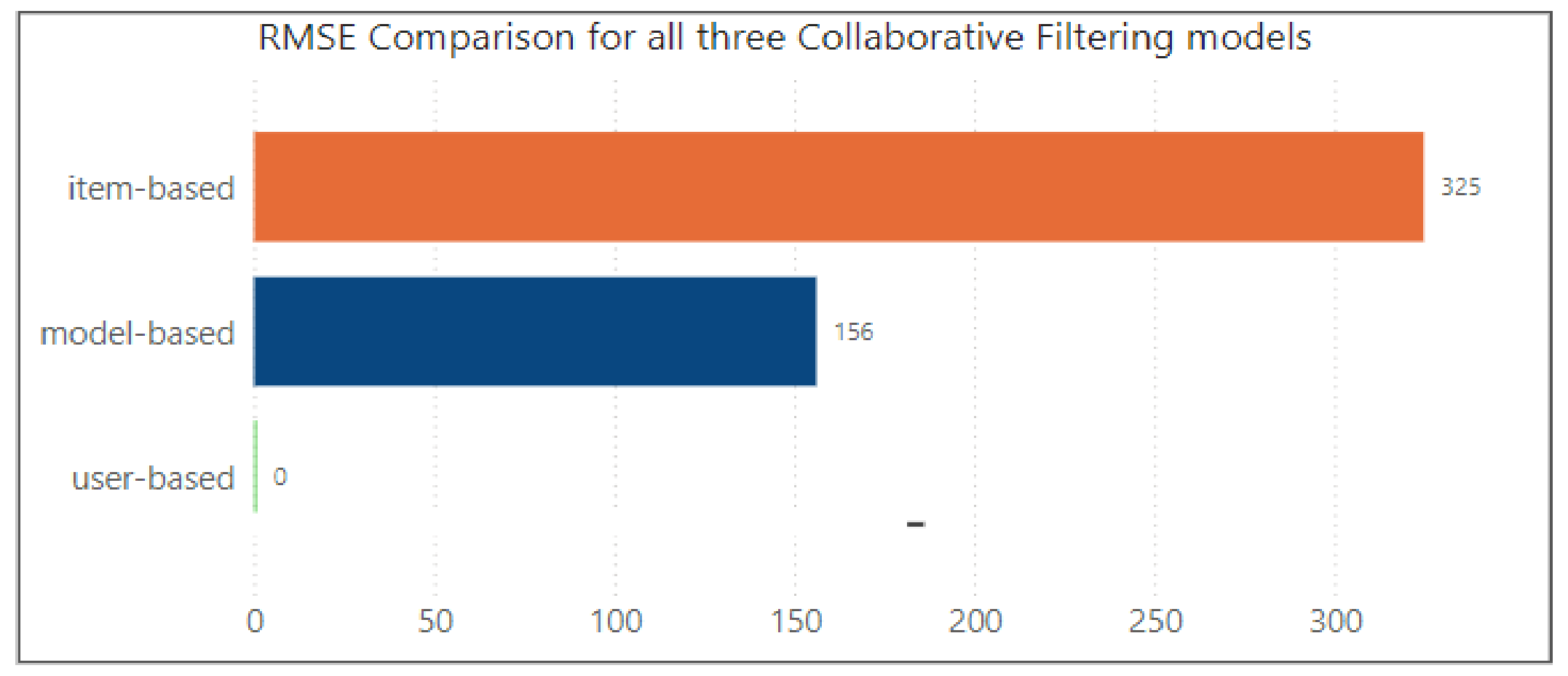
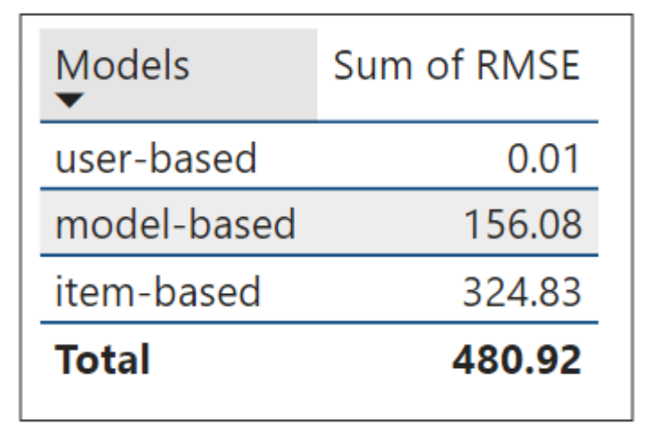


Figure 5. RMSE Values with Collaborative Filtering Models



Models	Sum of RMSE
user-based	0.01
model-based	156.08
item-based	324.83
Total	**480.92**

Table 2. Summary for RMSE Values

Future Directions

We will continue developing recommendation algorithms that leverage privacy-enhancing technologies, such as differential privacy or federated learning, to ensure user confidentiality while maintaining recommendation quality.

Acknowledgements

I would like to thank Dr. Yangyang Tao for your invaluable guidance throughout this project and the Center for Integrative Natural Science and Mathematics (CINSAM) for providing funding for this project.

References

Yujeong, Hwangbo, Kyoung Jun Lee, Baek Jeong, and Kyung Yang Park, Recommendation system with minimized transaction data." Data Science and Management 4 (2021): 40-45