Conclusion

1.

Model Overview:

2.

1. A **Lookalike Model** was developed using a **Nearest Neighbors** approach with **Cosine Similarity** to identify customers similar to a given user based on their transaction history and profile information.
2. The model incorporated various customer features, including **total transaction value**, **quantity purchased**, and **average price** paid. In addition, **one-hot encoded regions** were included as features to account for geographic diversity.

3.

Key Insights:

4.

1. By aggregating data for each customer, the model identified the most similar customers based on shared purchasing behavior and demographics, improving the business's ability to target similar customers for personalized marketing.
2. The **Cosine Similarity** metric was used to calculate how closely the profiles of customers match, ensuring that only customers with similar spending patterns and preferences were recommended.

5.

Results:

6.

1. The output was a list of the top 3 most similar customers for each of the first 20 customers (CustomerID: C0001 - C0020) based on their transaction behavior and profile information.
2. The **similarity scores** provided valuable insights into the closeness of customer profiles, which can be used for **targeted marketing** or **personalized recommendations**.

7.

Business Impact:

8.

1. The results from the model can be used to improve customer engagement and **cross-selling opportunities**, as the business can identify customers with similar buying behaviors and preferences.
2. By focusing on customers who resemble existing high-value users, the company can tailor marketing strategies to increase conversions and customer retention.

9.

Future Improvements:

10.

1. The model can be further enhanced by including additional features like **customer lifetime value (CLV)**, **sign-up date**, and **frequency of purchases**.
2. A more sophisticated similarity metric or ensemble models could improve the recommendation quality, and cross-validation can help in evaluating the model's performance more rigorously.

BUSINESS INSIGHTS

1. Top-Selling Product Categories

By analyzing the "Category" column in the Products dataset and linking it with transaction quantities, we can identify the most popular product categories. This insight can guide inventory management and promotional campaigns.

2. Regional Buying Trends

Linking the "Region" column from Customers.csv with Transactions.csv can reveal regional preferences. For instance, certain regions might favor specific product categories, helping to tailor marketing strategies geographically.

3. High-Value Customers

By calculating total spending per customer (using TotalValue in Transactions.csv), we can identify high-value customers. This insight can support personalized loyalty programs to improve customer retention.

4. Seasonal Sales Trends

Analyzing the "TransactionDate" column can uncover sales trends over time. Identifying peak seasons can help optimize inventory and plan seasonal promotions.

5. Customer Retention Patterns

By comparing "SignupDate" with transaction frequencies, we can identify customer retention rates. For instance, customers who remain active after a certain period might be more loyal and profitable.