Customer Lookalike Recommendations

The goal of this task is to identify lookalike customers based on their purchasing behavior and characteristics. The primary steps involved merging transaction data with customer and product information, aggregating customer-level features, and calculating similarity scores between customers.

1. Data Preprocessing and Merging

The initial step involved merging three key datasets: **Transactions**, **Customers**, and **Products**. The data was merged based on common identifiers such as **CustomerID** and **ProductID**. This step ensured that all necessary information related to customer transactions, product details, and customer profiles was consolidated for analysis.

2. Customer Features Aggregation

Next, customer-level features were aggregated to summarize the purchasing behavior of each customer:

- **TotalValue**: The total amount spent by the customer.
- **TransactionCount**: The total number of transactions made by the customer.
- **Category**: The most frequently purchased product category by the customer.

These aggregated features provided a comprehensive profile for each customer, representing their purchasing preferences and behaviours.

3. Cosine Similarity Calculation

To find customers with similar purchasing patterns, we used **Cosine Similarity**, a technique for measuring the similarity between two vectors in a multi-dimensional space. The customer features were transformed into numerical values (where needed), and cosine similarity was applied to compute the similarity scores between all customers.

The cosine similarity scores provided a measure of how alike two customers were based on their total spending, transaction frequency, and most purchased categories.

4. Generating Lookalike Recommendations

Using the similarity scores, we identified the top 3 most similar customers for each customer. These recommendations were based on the highest similarity scores, excluding the customer from their own list of recommendations. This process enabled the creation of a lookalike customer list for each individual.

5. Saving the Recommendations

The final list of lookalike customers was saved into a CSV file for further analysis. This structured output will allow for easier interpretation and usage of the recommendations in future marketing or customer engagement strategies.

Key Findings & Observations:

- The system successfully identified customers with similar behaviors, demonstrating the accuracy and usefulness of the **Cosine Similarity** metric.
- The recommendations showed high similarity scores (close to 1), indicating that the model was effective in capturing customer likeness based on purchasing patterns.
- The saved CSV file contains a list of customers and their top 3 lookalikes, which can be used for further analysis or targeted campaigns.

Conclusion:

- This task successfully achieved the goal of identifying lookalike customers by analyzing transaction and customer data. The recommendations were based on behavioral similarities, and the output was structured for practical use in marketing initiatives.