Lookalike Model - Explanation of Code Steps

1. Merge Datasets:

The transactions, customers, and products datasets were merged to create a comprehensive dataset. This dataset includes information about customer profiles, products, and their respective transactions.

2. Create Pivot Table:

A pivot table was created to capture customer-product purchase behavior. Rows represent customers, columns represent products, and the values are the quantities purchased.

3. Data Standardization:

The data was standardized using StandardScaler to ensure all features contributed equally to the similarity calculations. This step is crucial for models sensitive to varying scales.

4. Cosine Similarity:

Cosine similarity was used to compute similarity scores between customers based on their purchase patterns. This metric measures the cosine of the angle between two vectors, making it ideal for understanding similarity.

5. Extract Top Similar Customers:

A function was written to extract the top 3 most similar customers for a given customer. The results exclude the customer itself to avoid self-comparisons.

6. Save Results:

The results for the first 20 customers (CustomerID: C0001 to C0020) were saved in a CSV file

named 'Lookalike.csv'. The file includes CustomerID, SimilarCustomerID, and the similarity score in the required format.