Deliverables:

1. **Lookalike.csv:**

   - A CSV file named "Lookalike.csv" has been generated.
   - It contains the following columns:
     - `CustomerID`: The ID of the target customer.
     - `SimilarCustomerID`: The ID of the lookalike customer.
     - `SimilarityScore`: The cosine similarity score between the target customer and the lookalike customer.

2. **Jupyter Notebook/Python Script:**

   - The provided Python script outlines the model development process.
   - Key steps include:
     - Data loading and preparation:
       - Loading datasets (`Customers`, `Products`, `Transactions`).
       - Merging datasets to create a comprehensive view of customer transactions.
       - Aggregating customer features (total spending, transaction frequency).
       - Creating a pivot table to capture category preferences.
       - Merging features into a single DataFrame.
     - Feature normalization:
       - Standardizing numerical features using `StandardScaler`.
     - Cosine similarity calculation:
       - Computing the cosine similarity between normalized customer feature vectors.
     - Creating a customer similarity DataFrame.
     - Finding lookalike customers:
       - Defining a function to efficiently retrieve the top N most similar customers.
       - Iterating through the first 20 customers and finding their top 3 lookalikes.
     - Creating the "Lookalike.csv" file.
     - Visualizing customer similarity with a heatmap.

Insights:

- **Lookalike Customer Identification:**
  - The model successfully identified the top 3 lookalike customers for each of the first 20 customers based on their purchasing behavior and preferences.
  - For example, for Customer C0001, the top 3 similar customers are:
    - C0069 with a similarity score of 0.933

- ■ C0026 with a similarity score of 0.904
- ■ C0157 with a similarity score of 0.855
- **Model Explanation:**
  - The model utilizes a combination of customer features such as total spending, transaction frequency, and category preferences to identify lookalike customers.
  - Cosine similarity is employed to measure the similarity between customers based on their feature vectors.
  - This approach enables the identification of customers with similar purchasing patterns and preferences, which can be valuable for targeted marketing campaigns, product recommendations, and customer segmentation.

Further Considerations:

- **Feature Engineering:** Exploring additional customer features (e.g., recency, frequency, monetary value - RFM analysis, demographics) could enhance the accuracy of the lookalike model.
- **Model Evaluation:** Evaluating the model's performance using techniques like hold-out validation or cross-validation can provide insights into its predictive accuracy and robustness.
- **Business Application:** The identified lookalike customers can be leveraged for various business applications, such as:
  - Targeted marketing campaigns: Delivering personalized offers and promotions to lookalike customers.
  - Product recommendations: Suggesting products or services that similar customers have purchased.
  - Customer segmentation: Grouping customers into segments based on their lookalike profiles for more effective targeting and personalization.