Lookalike Model:

1. Load the Required Libraries

- Import essential libraries like pandas, numpy, StandardScaler (for data normalization), and cosine_similarity (for similarity computation).

2. Load the Datasets

- Read three CSV files:
    - Customers.csv (Customer details)
    - Transactions.csv (Transaction history)
    - Products.csv (Product details)
- Ensure that CustomerID is treated as a **string**, as it contains alphanumeric values (e.g., "C001").

3. Merge the Datasets

- Merge transactions_df with customers_df using the **CustomerID** column.
- Merge the result with products_df using the **ProductID** column.
- This creates a **single dataset** containing customer transactions along with product details.

4. Aggregate Transaction Data for Each Customer

- Group the data by **CustomerID** and compute:
    - TotalSpend: Total value spent by the customer.
    - AvgTransactionValue: Average value of transactions.
    - TotalTransactions: Total number of transactions.
    - LastPurchase: Date of the most recent purchase.

5. Merge Customer Profile Data

- Add customer-related details from customers_df to the aggregated transaction data.

6. Handle Date Columns

- Convert LastPurchase and SignupDate to **datetime format**, handling errors safely.
- Replace missing values with a placeholder date (2000-01-01) to avoid errors in date calculations.

7. Compute Derived Features

- Calculate Recency: Number of days since the last purchase.
- Calculate Tenure: Number of days since customer signup.

- Drop the original date columns as they are no longer needed.

8. Encode Categorical Features

- Convert categorical columns like Region into **one-hot encoded variables** to make them suitable for numerical analysis.

9. Normalize Numerical Features

- Standardize key numerical features (TotalSpend, AvgTransactionValue, TotalTransactions, Recency, Tenure) using StandardScaler.
- This ensures all features have the same scale, improving the accuracy of similarity calculations.

10. Compute Customer Similarity Using Cosine Similarity

- Calculate a **Cosine Similarity Matrix** using numerical features.
- Each row and column in this matrix represents a customer, with values indicating their similarity scores.

11. Define the Lookalike Recommendation Function

- Takes a CustomerID as input and finds the **n most similar customers** based on the similarity matrix.
- Steps involved:
    - **Find the index** of the given customer.
    - **Extract similarity scores** for all customers.
    - **Sort customers** based on similarity, excluding the customer itself.
    - **Return the top n most similar customers** along with their similarity scores.

12. Test the Recommendation System

- Run the function for a given CustomerID (e.g., "C001") and retrieve **3 similar customers** with similarity scores.