Task 2: Lookalike Model Report

Objective:

To build a Lookalike Model that identifies and recommends the top 3 customers similar to a given user, based on their profile and transaction history. This model uses both customer and product data for recommendations.

Dataset Overview:

1. **Customers.csv**:
    - Contains customer demographic information, including CustomerID, CustomerName, Region, and SignupDate.
2. **Products.csv**:
    - Contains product details, including ProductID, ProductName, Category, and Price.
3. **Transactions.csv**:
    - Contains transaction details, including TransactionID, CustomerID, ProductID, TransactionDate, Quantity, and TotalValue.

Methodology:

1. **Data Preprocessing:**
    - Converted SignupDate and TransactionDate to datetime format and handled invalid dates using pd.to_datetime().
    - Derived key customer metrics, such as:
        - **Account Age:** Days since the customer signed up.
        - **Category Preferences:** Proportion of spending across product categories.
    - Merged transaction data with product details to include product categories in the analysis.
    - Aggregated transaction data for each customer to calculate:
        - **Total Spend:** Total monetary value of transactions.
        - **Average Transaction Value:** Mean value of transactions.
        - **Transaction Count:** Number of transactions.
        - **Purchase Timespan:** Days between first and last purchase.
    - Normalized all numerical features using StandardScaler.
2. **Similarity Calculation:**
    - Constructed a feature matrix containing customer demographic, transaction, and category preference data.
    - Computed a **Cosine Similarity Matrix** to measure pairwise similarity between customers.
3. **Lookalike Recommendations:**
    - For each customer, identified the top 3 most similar customers based on similarity scores.
    - Returned the list of similar customers along with their similarity scores.

Results:

1. **Sample Recommendations:**

CustomerID	Lookalikes
C0001	C0069 (0.5724), C0163 (0.4954), C0048 (0.4790)
C0003	C0163 (0.6663), C0031 (0.5740), C0087 (0.4778)
C0004	C0165 (0.8785), C0104 (0.7493), C0075 (0.6884)
C0006	C0187 (0.6468), C0158 (0.6429), C0048 (0.5591)
C0011	C0174 (0.7040), C0107 (0.6882), C0165 (0.6047)

2. **Output Format:**
    - Generated a CSV file (Lookalike.csv) with two columns:
        - CustomerID: The ID of the customer for whom recommendations were made.
        - Lookalikes: A semicolon-separated list of similar customers and their similarity scores.

Conclusion:

The Lookalike Model successfully identifies similar customers for a given user based on profile and transaction data. This can help in personalized marketing strategies and customer engagement initiatives.