TASK-2

Lookalike Model

s

Introduction:

The goal of Task 2 is to develop a **Lookalike Model** that recommends three similar customers for each input customer based on their profile and transaction history. This task leverages both customer attributes (e.g., demographics, region) and transactional data (e.g., purchase history, spending patterns) to calculate similarity scores between customers. The recommendations are generated for the first 20 customers (CustomerIDs: C0001 - C0020), and the results are stored in a structured format as Lookalike.csv.

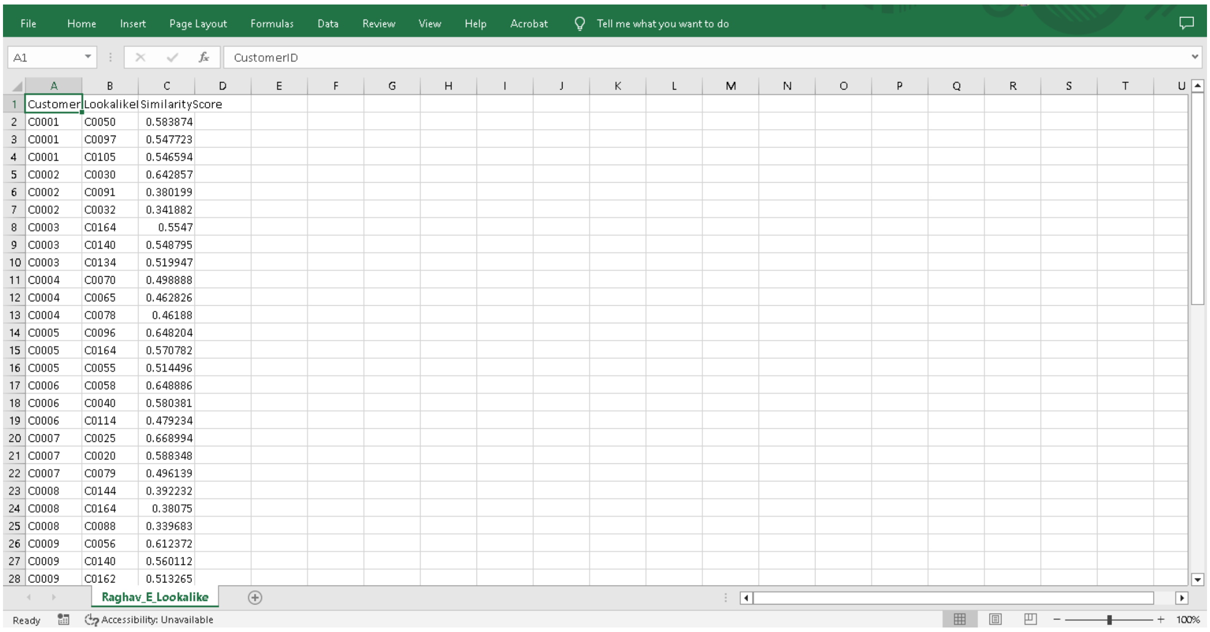
This task evaluates the effectiveness of the model in identifying lookalike customers, which can be applied to personalized marketing strategies or targeted campaigns.

Steps for Model Development :

1. **Data Preprocessing:**
    - o Merge Customers.csv and Transactions.csv on CustomerID to include both customer profiles and transaction details.
    - o Normalize numeric features (e.g., TotalValue, Quantity) to ensure fair comparison.
    - o Encode categorical variables (e.g., Region, ProductCategory) using one-hot encoding or label encoding.
2. **Feature Engineering:**
    - o Create aggregate features for each customer, such as:
        - ▪ Average spending per transaction.
        - ▪ Most purchased product categories.
        - ▪ Total quantity of products purchased.
    - o Generate vectors that represent customer profiles based on these features.
3. **Similarity Calculation:**
    - o Use a distance metric like **cosine similarity**, **Euclidean distance**, or a custom metric to compute similarity between customers based on their profile vectors.
    - o Rank other customers based on their similarity to the input customer.

4. **Recommendation System:**
    - For each input customer, recommend the top 3 most similar customers along with their similarity scores.
    - Store the results in the format: Map<cust_id, List<cust_id, score>>.

OUTPUT :



Customer	Lookalike	SimilarityScore
C0001	C0050	0.583874
C0001	C0097	0.547723
C0001	C0105	0.546594
C0002	C0030	0.642857
C0002	C0091	0.380199
C0002	C0032	0.341882
C0003	C0164	0.5547
C0003	C0140	0.548795
C0003	C0134	0.519947
C0004	C0070	0.498888
C0004	C0065	0.462826
C0004	C0078	0.46188
C0005	C0096	0.648204
C0005	C0164	0.570782
C0005	C0055	0.514496
C0006	C0058	0.648886
C0006	C0040	0.580381
C0006	C0114	0.479234
C0007	C0025	0.668994
C0007	C0020	0.588348
C0007	C0079	0.496139
C0008	C0144	0.392232
C0008	C0164	0.38075
C0008	C0088	0.339683
C0009	C0056	0.612372
C0009	C0140	0.560112
C0009	C0162	0.513265

Conclusion:

This Lookalike Model aims to identify customers with similar profiles, enabling the business to target lookalike groups with personalized marketing efforts. By leveraging both customer and transaction data, the model can effectively group customers with shared behaviours, thus improving campaign precision and increasing revenue potential.