Name: Asritha Veeramaneni

Assignment Tasks:

Task 2: Lookalike Model

Objective:

To build a Lookalike Model that recommends the top 3 similar customers based on their profile and transaction history. The recommendations include similarity scores and use both customer and product information.

Dataset Overview:

1. **Customers.csv**: Contains customer details like CustomerID, Region, and SignupDate.
2. **Products.csv**: Includes ProductID, Category, and Price.
3. **Transactions.csv**: Includes transactional data like TransactionID, CustomerID, ProductID, Quantity, TotalValue, and TransactionDate.

Methodology:

1. **Data Merging**:
    - Combined all three datasets using CustomerID and ProductID to form a consolidated dataset.
2. **Feature Engineering**:
    - Aggregated key features for each customer:
        - TotalValue: Total spending.
        - Quantity: Total items purchased.
        - Category: Most purchased product category.
        - Region: Customer's geographic region.
    - Applied one-hot encoding for categorical features (Region and Category).
    - Normalized numerical features (TotalValue and Quantity) using Min-Max scaling.
3. **Similarity Calculation**:
    - Used **Cosine Similarity** to calculate pairwise similarity between customer profiles.
    - Identified the top 3 most similar customers for each customer based on similarity scores.
4. **Output Format**:
    - Stored results in a CSV file named Lookalike.csv with the format:

Map<cust_id, List<cust_id, score>>

Results:

Here is the document to check the results

https://docs.google.com/spreadsheets/d/1gSUGBokbKPaHu7P-ArN96XHDll1Xk-Cj6OB1lIKuoXk/edit?gid=0#gid=0

Insights:

1. **High Similarity Scores Indicate Strong Profile Matches**:
    - Customers like C0007 and C0146 achieve a perfect similarity score (1.0000), showing identical behaviour or purchase patterns.
    - High scores (>0.99) across other pairs suggest strong alignment in their preferences.
2. **Customers with Similar Purchase Histories**:
    - C0001 is closely aligned with C0184 and C0048, likely due to shared product preferences and similar total spending.
    - C0005 and C0013 share a nearly identical profile with a score of 0.9999, indicating significant overlaps in behaviour.
3. **Geographic and Category Alignment**:
    - C0014 aligns strongly with C0060 (0.9999), possibly due to shared region and dominant category purchases.
    - C0018 and C0122 show strong similarities (0.9996), suggesting similar purchasing behaviour across categories.
4. **Utility for Targeted Marketing**:
    - Recommendations such as those for C0019 (aligned with C0073 and C0070) can be used to cross-sell or upsell products effectively.
    - Identifying highly similar customers helps tailor marketing campaigns based on shared preferences.
5. **Opportunities for Cross-Selling**:
    - Customers like C0020 with similarities to C0157 and C0050 can be targeted with offers related to shared purchase trends, boosting retention and sales.

Conclusion

- The Lookalike Model successfully identifies the top 3 similar customers for the first 20 customers based on transaction history and profile data.
- The results can be leveraged for personalized marketing, cross-selling, and customer engagement strategies.