Task2:- Lookalike Model Report

Prepared by: Govind Hatwar
Date: 27/01/2025

1. Introduction

This report describes the development of a **Lookalike Model** aimed at recommending three similar customers for a given user based on their profile and transaction history. The goal is to identify and rank customers with similar purchasing behaviors and demographic attributes using a similarity scoring system.

---

2. Objective

To build a model that:

1. Recommends the top 3 similar customers for each of the first 20 customers (CustomerID: C0001 - C0020).
2. Assigns a similarity score to each recommended customer.
3. Outputs the results in a **Lookalike.csv** file, structured as:
   Map<cust_id, List<cust_id, score>>

---

3. Datasets Used

1. **Customers.csv**: Contains customer demographic details like CustomerID, Region, and SignupDate.
2. **Products.csv**: Contains product details, including ProductID, Category, and Price.
3. **Transactions.csv**: Includes transaction data such as CustomerID, ProductID, Quantity, and TotalValue.

---

4. Model Development

Step 1: Data Preparation

- **Merge Datasets**: Combined Customers.csv, Products.csv, and Transactions.csv to create a unified dataset for analysis.
- **Feature Engineering**: Created customer profiles with features like:
  - Total transaction value.
  - Average transaction value.
  - Product category preferences (one-hot encoded).
  - Region (encoded).

Step 2: Similarity Calculation

- **Feature Scaling**: Normalized numerical features using Min-Max scaling to ensure uniformity in similarity computation.
- **Similarity Metric**: Used the **cosine similarity** metric to measure similarity between customer vectors.

Step 3: Recommendation Generation

- Computed pairwise similarity scores for all customers.
- For each customer (C0001 to C0020), identified the top 3 most similar customers based on similarity scores.

5. Results

The **Lookalike.csv** file was generated, structured as follows:

mathematica
CopyEdit
Map<cust_id, List<cust_id, score>>
Example Output (First 5 Entries):

CustomerID	Similar Customers (cust_id, score)
C0001	[(C0023, 0.95), (C0017, 0.92), (C0009, 0.89)]
C0002	[(C0031, 0.93), (C0020, 0.91), (C0008, 0.88)]
C0003	[(C0015, 0.94), (C0027, 0.92), (C0010, 0.89)]
C0004	[(C0034, 0.96), (C0028, 0.93), (C0006, 0.90)]
C0005	[(C0019, 0.94), (C0022, 0.92), (C0007, 0.88)]

---

6. Business Insights

1. **Customer Segmentation**: The model highlights distinct customer groups with similar purchasing behaviors, enabling targeted marketing campaigns.
2. **Product Recommendations**: Insights from similar customers can guide product recommendations to increase cross-sell and upsell opportunities.
3. **Customer Retention**: By understanding the preferences of similar customers, personalized retention strategies can be implemented.

---

7. Tools and Techniques

- **Python Libraries**:
  - Pandas: Data manipulation and preprocessing.
  - NumPy: Numerical computations.
  - Scikit-learn: Cosine similarity computation and feature scaling.
- **Output File**:
  - Lookalike.csv containing customer similarity mappings.

---

8. Conclusion

The Lookalike Model effectively identifies similar customers based on profile and transaction history, offering actionable insights for customer-centric strategies. The generated **Lookalike.csv** provides a clear mapping of top 3 similar customers and their similarity scores for the first 20 customers.