

# Lookalike Model Report

## 1. Introduction

The Lookalike Model aims to identify and recommend similar customers based on their profile and transaction history. This model utilizes customer attributes and spending behavior to generate similarity scores and recommend three closest matches for each customer.

## 2. Data Preparation

The dataset consists of three CSV files:

- **Customers.csv**: Contains CustomerID, Name, Region, and SignupDate.
- **Products.csv**: Contains ProductID, ProductName, Category, and Price.
- **Transactions.csv**: Contains TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, and Price.

Steps taken:

- Converted **SignupDate** and **TransactionDate** to datetime format.
- Aggregated customer transaction data to calculate **Total Spending (TotalValue)** and **Number of Transactions (TotalTransactions)**.
- Merged aggregated transaction data with customer profiles.
- Encoded categorical features (Region) using one-hot encoding.
- Standardized numerical features to ensure uniform scaling.

## 3. Similarity Computation

- Used **Cosine Similarity** to measure the similarity between customers based on numerical attributes.
- Computed a similarity matrix for all customers.
- Extracted the top 3 most similar customers for each individual, excluding themselves.

## 4. Results

- The top 3 most similar customers were identified for the first 20 customers (C0001 - C0020).
- The results were saved in a CSV file **Lookalike.csv**, structured as follows:
  - **CustomerID | Lookalike1 | Score1 | Lookalike2 | Score2 | Lookalike3 | Score3**
- Similarity scores range between 0 (no similarity) and 1 (high similarity).

## 5. Business Insights

- This model helps in **targeted marketing campaigns** by identifying customers with similar buying behavior.
- It assists in **personalized product recommendations** based on similar customer preferences.
- Customer segmentation can be enhanced by analyzing clusters of similar customers.
- The model can be leveraged for **customer retention strategies**, identifying those likely to churn.

## 6. Limitations & Future Improvements

- The model currently considers **transactional and profile-based features**, but could be enhanced with product preferences.
- **Incorporating product category trends** in similarity computation may improve recommendations.
- **Including more advanced techniques like clustering or deep learning embeddings** could refine lookalike accuracy.

## 7. Conclusion

The Lookalike Model successfully identifies similar customers using a data-driven approach. This method can be instrumental in refining business strategies, customer engagement, and marketing initiatives.