

# Lookalike Model: Recommendation Engine for Similar Customers

## Task 2 Report for eCommerce Transactions Dataset

---

### 1. Overview

The Lookalike Model identifies customers who share similar profiles and behaviors. By analyzing both customer demographic data and their transaction history, the model recommends similar customers for targeted marketing or product strategies.

### 2. Data Used

- **Customers.csv:** Contains demographic information such as CustomerID, Region, and SignupDate.
- **Transactions.csv:** Contains data on customer purchases, including ProductID, Quantity, and TotalValue.

### 3. Approach

The model uses **content-based filtering** to calculate similarity:

- **Customer Profile:** Region, SignupDate
- **Transaction Behavior:** Product preferences, purchase frequency, and spending.

**Cosine similarity** is used to calculate similarity between customers, based on these features. The top 3 most similar customers for each of the first 20 customers (C0001 - C0020) are recommended.

### 4. Lookalike Recommendations

The model outputs a CSV file that shows each customer's top 3 lookalikes and their similarity scores. For example:

#### Python

CustomerID, Lookalikes

C0001, C0003:0.85, C0007:0.78, C0012:0.74

C0002, C0015:0.92, C0020:0.88, C0011:0.82

### 5. Model Evaluation

The model's effectiveness is assessed by the similarity scores and the relevance of the recommendations. Customers with higher similarity scores are considered more alike in terms of behavior and profile. Manual checks confirm that the recommendations align with expected patterns.

### 6. Conclusion

The Lookalike Model offers valuable insights into customer similarities, helping with targeted marketing and personalized strategies. The model can be enhanced further by including additional features or more advanced techniques like collaborative filtering.

## 7. Deliverables

- **Lookalike.csv:** Contains the top 3 similar customers for each target customer.
- **Jupyter Notebook:** Detailed implementation of the model, including code for similarity calculation and generating recommendations.