Lookalike Model Report

1. Introduction

- A lookalike model helps identify customers similar to a given customer based on their profile and transaction history. This is valuable for targeting campaigns and enhancing personalization.
- Objectives: The objective is to recommend three similar customers for each of the first 20 customers (C0001–C0020) using a similarity-based approach.

2. Methodology

Data Preparation

- Customer profile data from Customers.csv was combined with transaction summary data derived from Transactions.csv and Products.csv. This resulted in a dataset containing customer-level features, such as average transaction value, purchase frequency, and regional information.
- Numerical features were scaled, and categorical features were encoded where applicable.

Similarity Calculation

- Cosine similarity was used to measure the similarity between customers. This metric considers the angle between feature vectors, making it effective for high-dimensional data.
- The model used average purchase value, purchase frequency, and the number of unique products purchased as features.

Model Development

- The similarity matrix was calculated for all customers, and for each target customer, the top three most similar customers were identified along with their similarity scores.

---

3. Results

	CustomerID	Lookalikes
0	C0001	[('C0164', 0.9684103747672836), ('C0137', 0.96...
1	C0002	[('C0029', 0.9997616343498978), ('C0031', 0.99...
2	C0003	[('C0176', 0.8906401232895584), ('C0027', 0.86...
3	C0004	[('C0075', 0.9976740652389241), ('C0175', 0.99...
4	C0005	[('C0058', 0.9997982043779898), ('C0123', 0.99...

Observations

- Customers with similar purchase behaviour often reside in the same region.
- Customers with high average transaction values frequently have similar lookalikes due to consistent purchasing patterns.

---

4. Conclusion

- The lookalike model effectively identified similar customers for targeted marketing efforts. Key factors influencing similarity include average purchase value and transaction frequency.
- The model can be used for cross-selling, personalized email campaigns, and customer retention strategies.