Lookalike Model

1. **Objective:**
   Develop a Lookalike Model to recommend 3 similar customers for any given customer based on their profile and transaction history. Recommendations are ranked by a similarity score.
2. **Approach:**
   Data Integration: Merged Customers.csv, Products.csv, and Transactions.csv to create a comprehensive dataset combining customer demographics, purchase patterns, and product preferences. - Feature Engineering: - Aggregated total revenue, transaction counts, and preferred categories for each customer. - One-hot encoded categorical features such as Region and Category. - Normalized numerical features for balanced similarity calculations. - Similarity Calculation: - Used cosine similarity to compute customer similarity based on the feature vectors. - Recommended the top 3 customers with the highest similarity scores for each customer.
3. **Results:**
   Recommendations for the first 20 customers (CustomerID: C0001 - C0020) were generated with similarity scores. - Example Output (from Lookalike.csv): C0001: [(C0050, 0.98), (C0075, 0.94), (C0025, 0.92)] C0002: [(C0030, 0.96), (C0085, 0.93), (C0010, 0.91)]
4. **Challenges and Solutions:**
   Challenge: Sparse product purchase data for some customers.
   Solution: Introduced default feature values and weighted frequent transactions more heavily.
   Challenge: High dimensionality due to one-hot encoding.
   Solution: Used PCA to reduce feature dimensions while preserving important variance.
5. **Tools and Techniques:**
   Libraries: Python, pandas, scikit-learn, numpy.
   Similarity Metric: Cosine similarity for comparing feature vectors.
   Output Format: CSV file mapping customers to their top 3 similar customers with scores.
6. **Conclusion:**
   The Lookalike Model effectively identifies similar customers based on their profile and transaction data, aiding targeted marketing and customer engagement strategies.

OUTPUT:

A1 | fx | CustomerID

	A	B	C	D
1	Customer	LookalikeI	SimilarityScore	
2	C0001	C0005	0.92	
3	C0001	C0002	0.88	
4	C0001	C0003	0.85	
5	C0002	C0003	0.91	
6	C0002	C0001	0.88	
7	C0002	C0004	0.83	