Lookalike Model Report

➢ Objective:

To identify and recommend similar customers based on their profiles and transaction history,

enabling personalized marketing and improved customer retention.

➢ Methodology:

➢ Data Preparation:

- Merged `Transactions`, `Products`, and `Customers` datasets to build a unified view.

- Aggregated customer-level features such as:

    o TotalValue: Total transaction value per customer.

    o Quantity: Total quantity purchased.

    o Price: Average price of products purchased.

    o Category: Most purchased product category.

➢ Feature Engineering:

    o Encoded the `Category` feature using one-hot encoding.

    o Standardized numerical features (`TotalValue`, `Quantity`, `Price`) with `StandardScaler`.

➢ Similarity Calculation:

    o Used 'cosine similarity' to compute similarity scores between customers based on their

    o profiles.

    o Generated recommendations for the first 20 customers, identifying their top 3 lookalikes.

➢ Results:

➢ Sample Recommendations (from Lookalike CSV):

Customer ID	SimilarCustomerID	Similarity Score
C0001	C0069	0.9324
C0001	C0154	0.9235
C0001	C0026	0.8875
C0002	C0029	0.9998
C0002	C0088	0.9848
C0002	C0062	0.9811
C0003	C0038	0.9938
C0003	C0160	0.9440
C0003	C0189	0.9273

➢ Insights:
  o Customers with similar spending patterns, purchase quantities, and product preferences exhibit high similarity scores.
  o Regional and category-specific preferences influence similarity clustering.

➢ Additional Findings:
  o Regional Distribution: South America has the largest customer base (29.5%).
  o Top Revenue Category: `Books` contribute $192,147.47 to total revenue.
  o Monthly Trends: Significant sales growth, with a total of 1,000 transactions and an average value of $689.99.

➢ Recommendations:
  o Use lookalike recommendations for personalized marketing and upselling.
  o Focus campaigns on customers similar to high spenders to drive revenue.
  o Leverage insights into regional and category preferences for targeted promotions.