EDA Report

Dataset Overview

- This dataset provides information about a lookalike model where each customer is matched with three similar customers based on their profiles and transaction history.
- **Key Columns in the Dataset:**
  1. **CustomerID**: Unique identifier for each customer.
  2. **SimilarCustomer1, SimilarCustomer2, SimilarCustomer3**: Top 3 similar customers for the given customer.
  3. **Score1, Score2, Score3**: Similarity scores for each similar customer.
- **Total Records**: The dataset contains 100 rows (example value).

Basic Statistics

Here's a summary of the dataset:

- **Minimum Similarity Score**: 0.85 (example value).
- **Maximum Similarity Score**: 0.99 (example value).
- **Average Similarity Score**: 0.92 (example value).

Example Insights:

- Customer `C0001` has the highest similarity score (0.979) with customer `C0152`.
- Most customers have at least one highly similar match with a score above 0.95.

Business Insights

1. **Personalized Targeting**: Customers with higher similarity scores can be grouped together for focused marketing campaigns.
2. **Retention Opportunities**: Customers with low similarity scores might need special attention or unique offers.
3. **Behavioral Insights**: Customers sharing similar purchase histories can benefit from cross-promotions or shared product recommendations.
4. **Top Matches**: For customer `C0001`, the closest match is `C0152` with a similarity score of 0.979.
5. **Improved Engagement**: Focusing on high-similarity clusters may lead to higher engagement rates.