Lookalike Model for Customer Segmentation and Product Recommendation

---

Executive Summary

This report details the implementation and results of a lookalike model aimed at identifying customer segments and providing personalized product recommendations for a retail business. The objective is to better understand customer behavior, improve targeted marketing efforts, and drive sales by identifying customers similar to existing high-value or frequent buyers.

Data Overview

The analysis is based on three core datasets:

1. **Customers**: Contains customer details, including Customer ID, Name, Region, and Signup Date.
2. **Products**: Provides information about the products, including Product ID, Name, Category, and Price.
3. **Transactions**: Captures transaction details, such as Transaction ID, Customer ID, Product ID, Quantity, Total Value, and Date of Purchase.

The data was cleaned and merged to provide a unified view, allowing us to perform detailed customer and transaction analysis.

Data Preprocessing and Feature Engineering

1. **Data Merging**: The datasets were merged on the `CustomerID` and `ProductID` fields to create a comprehensive transaction history for each customer, including product details and customer demographics.
2. **Feature Engineering**:
    - **Region Encoding**: The `Region` column was encoded using `LabelEncoder` to create a numeric representation of customer regions.
    - **Customer Features**: Key features were derived, including:
        - `total_spend`: The total amount spent by each customer.
        - `purchase_frequency`: The total number of transactions made by each customer.
        - `region`: The encoded region information for each customer.

Similarity Calculation

A **Cosine Similarity** metric was used to measure how similar customers are based on their spending behavior, purchase frequency, and region. The resulting similarity matrix provides a value between 0 and 1, where 1 indicates identical customers and 0 indicates no similarity.

Recommendation Engine

Using the similarity scores, a recommendation engine was built to suggest customers that are most similar to a given target customer. For instance, if a retailer wants to target customers similar to high-value or frequent shoppers, the model can generate a list of top 5 most similar customers. This allows for efficient and precise targeting in marketing and promotional activities.

Example Output

For example, the top 5 customers most similar to **C0001** were identified as:

1. **C0137** with a similarity score of 1.0
2. **C0152** with a similarity score of 1.0
3. **C0198** with a similarity score of 1.0
4. **C0003** with a similarity score of 1.0
5. **C0002** with a similarity score of 1.0

These customers can now be considered for similar promotional offers or product recommendations based on their shopping behaviors.

Business Insights

1. **Targeted Marketing**: By identifying customers similar to high-value shoppers, the business can run personalized campaigns targeting the most likely buyers, increasing conversion rates.
2. **Product Recommendations**: Understanding customer behavior allows the business to recommend products that similar customers have purchased, potentially increasing sales.
3. **Customer Segmentation**: The model can be used to segment customers into different categories based on their behavior (e.g., high spenders, frequent buyers, or specific region-based preferences). This segmentation can drive more effective marketing strategies, such as tailored offers or region-specific discounts.

Conclusion

The lookalike model has provided valuable insights into customer behavior, and the recommendation engine can be used to optimize marketing efforts. By leveraging similarity scores, businesses can identify high-potential customer segments, create targeted campaigns, and enhance customer satisfaction through personalized product recommendations.