

Lookalike Model Explanation

The goal of the Lookalike Model is to identify and recommend the **top 3 similar customers** based on their **profile and transaction history**. The model takes the following approach:

Step 1: Data Preparation

We start by preparing the data for analysis. The Transactions.csv dataset is used to calculate the total spend (Total Value) and the total quantity of products purchased (Quantity) for each customer. These two features—Total Value and Quantity—are crucial for identifying patterns of customer behaviour and determining similarity.

Step 2: Feature Engineering

We aggregate the transaction data by CustomerID to compute the following features:

- **Total Spend (Total Value):** The sum of all transactions made by the customer.
- **Total Quantity (Quantity):** The total quantity of products purchased by the customer.

These features represent the customer's overall purchasing behaviour and will be used to compare customers against each other.

Step 3: Data Normalization

Cosine similarity requires normalized data to ensure that the comparison is based on relative patterns of purchase rather than the absolute values, which could be skewed by customers with high overall spending or large quantities. We use **Standard Scaler** from scikit-learn to normalize the data, so each feature (i.e., Total Value and Quantity) has a mean of 0 and a standard deviation of 1.

Step 4: Similarity Calculation using Cosine Similarity

Next, we calculate the **cosine similarity** between customers. Cosine similarity is a metric used to measure how similar two vectors are, irrespective of their magnitude. It calculates the cosine of the angle between two vectors in a multi-dimensional space, and the value ranges from -1 (completely dissimilar) to 1 (completely similar). Here, the vectors represent customers' normalized Total Value and Quantity features.

We use the cosine similarity function from scikit-learn to compute a similarity matrix, which gives us the similarity scores between each pair of customers.

Step 5: Identifying Lookalikes

Once the similarity matrix is generated, we identify the **top 3 most similar customers** for each customer. The logic is as follows:

1. For each customer, we extract their similarity scores with all other customers.
 2. We sort these scores in descending order to identify the most similar customers (i.e., those with the highest similarity scores).
 3. The top 3 customers (excluding the customer themselves) are selected as their lookalikes.
-

Step 6: Storing and Exporting Results

The results of the lookalike analysis are stored in a dictionary, where each key is a CustomerID, and the value is a list of tuples containing the **CustomerID** and the **similarity score** of their top 3 lookalikes. This dictionary is then transformed into a pandas DataFrame for easy manipulation and export.

The DataFrame contains the following columns:

- CustomerID: The ID of the customer for whom lookalikes are being identified.
- Lookalike1, Lookalike2, Lookalike3: The IDs of the top 3 lookalike customers.
- Score1, Score2, Score3: The corresponding similarity scores for the lookalike customers.

Finally, the DataFrame is exported to a CSV file (Lookalike.csv) for further analysis and reporting.

Step 7: Output Verification

A sample of the output is displayed to ensure that the data is being correctly processed and saved. The CSV file provides an organized view of the top 3 lookalikes for each customer, along with their similarity scores.

Key Insights from the Model:

- The **lookalike customers** are identified based on their purchasing behaviours, ensuring that the recommendations are closely aligned with the original customer's spending patterns and product preferences.
 - **Cosine similarity** ensures that the model captures the relative similarities in behaviour rather than just focusing on the size of purchases, making the recommendations more meaningful.
-

Business Use Case:

This model can be used in various ways:

- **Targeted marketing:** By identifying similar customers, businesses can craft marketing campaigns tailored to those lookalikes, promoting products that are likely to appeal to them.
- **Customer segmentation:** This model can help in segmenting customers into groups of similar profiles, enabling personalized customer experiences and improving customer retention strategies.

Conclusion

This model not only provides businesses with a way to identify lookalikes but also helps in delivering more relevant experiences and offers to their customers, ultimately leading to increased customer satisfaction and retention. The insights generated through this model can significantly impact the effectiveness of business decisions related to product recommendations, marketing campaigns, and customer loyalty programs.
