

Customer Segmentation for E-Commerce

Project Description

Imagine you are part of the Olist (e-commerce in Brazil) Data Scientist Team. You are assigned to help the marketing team create segmentation of Olist customers based on their behavior. The models used in this project are the RFM (Recency, Frequency, & Monetary) Segmentation and K-Means Clustering.

Background

E-commerce is a business model that allows a company or individual to buy or sell their products via the internet. E-commerce users from year to year also increase, especially in Brazil. The following is the data of E-commerce users from 2017 to 2025.

Revenue from the marketplace business model can come from paid features, advertisements, payment gateways, and partnerships. The point is that the more users or traffic on a marketplace, the more revenue you will get. One way to increase users or traffic on a marketplace is to do a **marketing campaign**, such as giving discounts, copywriting, etc.

Companies prepare a budget of 5% — 12% of the total revenue to carry out marketing campaigns. However, **the problem is that the marketing campaign is not right on target, so the company suffers a loss.** A marketing campaign can be measured by how many new users manage to get and how many old users retain to keep using the marketplace. One way to solve the problem is to segment customers or commonly referred to as Customer Segmentation. Customer Segmentation helps marketplace owners to group customers with similar characteristics.

Data Understanding

This data was taken from 2016 to 2018 and publicly distributed on Kaggle in September 2018. The existing data contains customer and seller transactions. To solve this problem, we only need data related to customer transactions because we want to do customer segmentation, which helps us know a customer's behavior in a transaction.

Business Problem

1. How to segment the customers at Olist marketplace so we can divide customers based on their shopping behavior?
2. What kind of treatment for each cluster to increase customer retention rate?

Workflow

1. Data Merging

2. Data Cleaning and Data Pre-Processing

3. Exploratory Data Analysis (EDA) — RFM

[RFM](#) is a technique that can perform Customer Segmentation to determine customers' buying behavior.

The meaning of RFM itself is:

- Frequency: Number of transactions. The bigger the number, the better. Because it means the customers often make transactions.
- Monetary: The spending power of a customer. The bigger the number, the better. Because it means the customer does a lot of spending
- Recency: The last time the customer made a purchase. The smaller the number, the better. Because it means the customers have a short period since they made a transaction since the last time.

The RFM metric is an important indicator of customer behavior segmentation because the **frequency** and **monetary** affect customer lifetime value, and **recency** effect engagement rate. By using RFM, companies can reactivate customers who rarely transact, make customers more loyal, improve marketing strategies to increase revenue from a company.

4. Modeling (K-Means Clustering using RFM Features)

K-Means is an algorithm that can be used for clustering. K-Means algorithm uses the same distance concept as the KNN (K-Nearest Neighbor) algorithm. One of the distance concepts used is [Euclidean distance](#). We have to scale our features because the scale is not the same. The features that we have shows that there are outliers, and we don't want the outlier data to have an effect so that the scaling that can be used is the [Standard Scaler](#).

In cluster analysis, the elbow method is **a heuristic used in determining the number of clusters in a data set**. The method consists of plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use. Based on the Elbow Method and Silhouette Score: We choose 4 clusters.

- **Lost Customers:** Customers who have not made a transaction for a long time, only made one transaction, and have the lowest total transaction.
- **Loyal Customers:** Customers who recently made a transaction, more than one transaction, and the total transaction is high.
- **New Customers:** Customers who recently made a transaction, only made one transaction, and the total transaction is low.
- **Best Customers:** Customers who recently made a transaction, more than one transaction, and the highest total transaction.

Next, we visualize the segmentation results using 2D and 3D scatter plots so that the distribution of each segment can be seen clearly.

Conclusion:

1. **Best Customers:**
 - **Best Customers** were when the customers made transactions recently, made more than one transaction, and had the highest total transactions.
 - Most customers buy the computer accessories category with a credit card payment method.
 - Treatments for these segments are loyalty program/reward points, new product recommendations, and exclusive product offers. (Cross / Up-Selling Strategy)

2. Loyal Customers:

- **Loyal Customers** were when the customers made transactions recently, made more than one transaction, and had the high total transactions.
- Most customers buy the furniture decoration category with a credit card payment method.
- Treatments for this segments are loyalty program/reward points and new product recommendations(Cross / Up-Selling Strategy)

3. New Customers:

- **New Customers** were when the customers made transactions recently, made only one transaction, and had the low total transactions.
- Most customers buy the health beauty category with the credit card payment method.
- Treatments for this segments are giving them a welcome e-mail to build the relationship, offer loyalty program/reward points, and discount vouchers (Cross/Up-Selling Strategy)

4. Lost Customers:

- **Lost Customers** were when the customers not made a transaction for a long time, made only one transaction, and had the lowest total transactions.
- Most customers buy the bed bath table category with the credit card payment method.
- Treatments for these segments are regular limited offers, discount vouchers, campaigns via e-mail, and asking for feedback. (Retention & Reactivate Strategies)