



Data Storm v5.0

— Semi-Final Round —

Organized By
Rotaract Club of University of Moratuwa
MAY 29, 2024

Powered By
OCTAVE

1 Background To The Business

KJ Marketing, a prominent retail supermarket chain in Sri Lanka, serves a wide range of end consumers. Building on their previous customer classification, the company has identified an opportunity to target high-end customers. They plan to implement personalized marketing strategies to better cater to this segment.



Figure 1: Sample Image

2 Background To The Problem

The company has recently identified that during the past few years using conventional and standard marketing strategies has not been effective to their current customer base. Therefore, in an effort to enhance its marketing strategies, the company is interested in adopting a personalized marketing strategy tailored to individual customer preferences. To effectively implement this strategy, they need to identify their high-end customers, who can be categorized into three distinct groups: Premium, Loyal, and Frequent. To support this initiative, they have provided you with historical customer data. The primary objective is to develop a rule-based approach to identify high-end customers. To ensure the accuracy and effectiveness of the strategy, it is essential to establish a clear framework based on percentile rankings. To facilitate this, a preliminary step involves utilizing clustering techniques to analyze the provided customer data. The purpose of clustering is to identify distinct groups within the customer base.

You are provided with monthly spend, monthly number of visits and monthly basket size for each customer. high end customers consists of premium, loyal and frequent customers, while there are average and low performing customers as well. The company has decided to define target customer segments as defined in the below table 1.

Table 1: Customer Category Data

Customer category	Monthly spend	Monthly visit frequency	Monthly basket size
High end - Premium	very high	high	high
High end - Loyal	high	average	very high
High end - Frequent	high	very high	average
Medium level	average	average	average
Low level	low	low	low

You are tasked with performing clustering on this customer base to understand how the customer segments vary with each other, once this is done, you are required to analyze each segment, and come up with a set of rules defining how each high end customer segments can be identified using a percentile approach on the three given variables.

Eg: Premium customers

- Monthly spend between 95-100th percentile.
- Monthly visit frequency between 80-94th percentile.
- Monthly basket size - between 80-94th percentile.

The company also requires you to come up with interventions to target these customer segments, the business has pre-defined promotion goals that are listed as; **Increase customer retention**, **Increase customer spend**, and **Increase customer visit frequency**.

You are tasked with developing intervention strategies to deliver mass promotions to customers in each segment. Your goal is to create effective strategies that can be shared across entire customer segments. Clearly describe your assumptions, detailed strategies, and the reasoning behind your decisions. Below is a simple idea to create an intervention plan.

Eg:

Customer Segment: Premium

Promotion Goals: The primary goal of this promotion is to increase customer spend.

Customer Preferences: Based on transactional data, identify the items that are frequently purchased by this segment, now find suitable substitutes for these items that have a high margin percentage for the business, and recommend these items to the customers in this segment.

Communication Channels: Promotions can be delivered via multiple channels such as email, SMS, in-app notifications, and direct mail.

3 Datasets and Variable Descriptions

You are provided with customer based dataset for your analysis. It is as follows.

1. Customer data

- **customer_id**: unique identifier of the customers.
- **monthly_spend**: average monetary expenditure by an individual customer during a single month at the store.
- **monthly_visits_frequency**: average number of times a customer visits the store during a single month.
- **monthly_basket_size**: average number of items purchased by a customer during a single month.

4 Deliverables and Evaluation Matrices

4.1 Technical Report - 60% of the total evaluation

Based on your analytical solution, you are required to create a report of your solution (**this needs to be in a proper report format**) with clearly defining the steps, features, feature engineering steps, modelling approaches, evaluation metrics, all the necessary plots/figures and interesting business findings that you can derive from this analysis while answering the following questions.

1. Elaborate on the methodologies implemented to address missing values, duplicates and outliers within the dataset? Please describe any specific techniques used for imputation or exclusion, and the rationale behind these choices.
2. Has feature scaling or standardization been applied to the data? If so, which methods were utilized and explain how these techniques improve the performance of the model?
3. What are the algorithms you considered for this problem, and why did you choose the final algorithm?
4. Were there any challenges faced during model training?
5. Describe the clusters and interpret how it matches with the business customer segment requirements, detail the characteristics that define the different customer segments.
6. What are the finalized rule based conditions you have defined for each segment post clustering? How did you arrive at these conditions, explain your methodology clearly.
7. What are the marketing interventions/strategies that can be applied on the requirements the business has raised, explain your intervention plans as detailed as possible, include any assumptions you make as well.

4.2 Analytical Solution - 40% of the total evaluation

Participants are required to submit their code base as Python files.

- Code base detailing the data loading and cleaning process.
- Code base detailing any descriptive analysis you performed on the data.
- Code base detailing code used for data pre-processing and clustering.
- Code base detailing the clustering evaluations you conducted.
- Code base detailing the steps you used to come up with the percentile rules for each customer segment.

Please ensure that your submission is complete and adheres to these guidelines for a successful evaluation. Include comments for added clarity.