Customer Segmentation as the Base for Campaign Planning of a Retail Store (Case Study)

Capstone Project by Akhir Syabani

INTRODUCTION

Customer segmentation is a vital part of retail industry as it helps companies to create and deploy the right strategies. Companies can build a better customer experience or developing attractive sales promotions that will boost their revenues in return.

In this case study, I explored a sample of Customer Segmentation that would help a retail store to plan and budget for their next campaigns. There are two main parts of this project:

- First, I created a customer segmentation based on customers' profile from internal membership data of a retail store, including transaction habits. The goal is to help the management on the decision making, which to produce a valuable input as an additional consideration before they move forward to develop the right strategies according to their available budget.
- Second, I built a model to predict customers' likelihood to response to a campaign based on their previous response toward a campaign or a promotion.

DATA

- For customer segmentation, the data is from **here**. Ideally, the data need to include customers' static data (or demographic) and transaction habits. However, the demographic data is currently unavailable. Within the available data, I used the transaction records of customers which included transaction dates and each of the transaction amount. I counted the data to determine the transaction frequency of each customer, and the average amount of each purchase accordingly. The period is from May 2011 to March 2015. For simplicity, I refer the data as **Customer Transactions** data. The data consist of around 125,000 transactions from 6889 unique customers.

	customer_id	trans_date	tran_amount
0	CS5295	11-Feb-13	35
1	CS4768	15-Mar-15	39
2	CS2122	26-Feb-13	52
3	CS1217	16-Nov-11	99
4	CS1850	20-Nov-13	78
...	...	...	...
124995	CS8433	26-Jun-11	64
124996	CS7232	19-Aug-14	38
124997	CS8731	28-Nov-14	42
124998	CS8133	14-Dec-13	13
124999	CS7996	13-Dec-14	36

125000 rows × 3 columns

	tran_amount
count	125000.000000
mean	64.991912
std	22.860006
min	10.000000
25%	47.000000
50%	65.000000
75%	83.000000
max	105.000000

Figure 1 & 2: Customer Transactions data – summary view

- For the second part, I added the data from [here](), which contained customers' response toward a previous campaign (whether they previously responded or ignored the campaign), we can call it as **Customer Responses**. The available responses were collected from 6884 of 6889 unique customers.

	customer_id	response
0	CS1112	0
1	CS1113	0
2	CS1114	1
3	CS1115	1
4	CS1116	1
...	...	...
6879	CS8996	0
6880	CS8997	0
6881	CS8998	0
6882	CS8999	0
6883	CS9000	0

6884 rows × 2 columns

Figure 3: Customer Responses data – summary view

METHODOLOGY

- I clustered the **Customer Transactions** data using **K-Means clustering**. To determine the optimum K, I used the **elbow method**.
- For the second part, I built the prediction model using **Logistic Regression**, as the target data that is categorical variable (whether customer will likely to respond to a campaign or not, from additional related sample data of **Customer Responses**).

Part 1: CUSTOMER SEGMENTATION ANALYSIS & RESULTS

We have three customer segments based on the clustering:

Cluster 1: Highest Frequency with Big Ticket Sizes or "The Most Loyal Customers"
This cluster has 2215 members, or 32% of total customers. The average frequency of their purchases is around 24 times, with the average amount of around 70 dollars each. *(In the visualization below, represented by the purple area)*

Cluster 2: Lowest Frequency with Small Ticket Sizes or "The Average Customers"
This cluster has 1895 members, or 28% of total customers. The average frequency of their purchases is around 13 times, with the average amount of around 44 dollars each. *(In the visualization below, represented by the blue area)*

Cluster 3: Mid-Range Frequency with Big Ticket Sizes or "The Potential Upgrade"
This cluster has 2779 members, or 40% of total customers, which makes them the biggest group of the three clusters. The average frequency of their purchases is around 17 times, with the average amount of around 70 dollars each. *(In the visualization below, represented by the yellow area)*

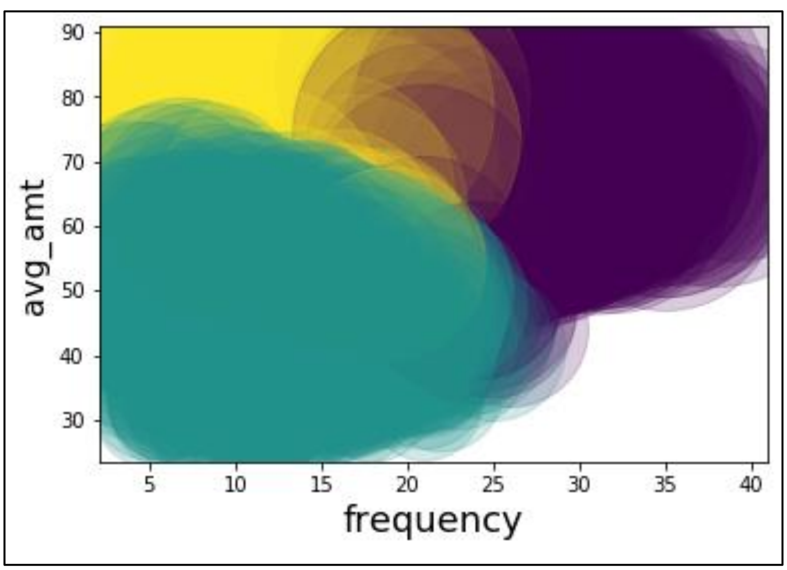

Figure 4: Visualization on the customer segments/clusters based on purchase frequency and average amount spent per customer

Part 2: CUSTOMER RESPONSE PREDICTION MODELING & RESULTS

I have developed the Customer Response model to predict whether customer will respond to a future campaign using Logistic Regression. The model evaluation score is as shown below:

Jaccard Score	f-1 score	log loss
0.90	0.85	0.31

Figure 5: Table of Customer Response Logistic Regression model evaluation scores

Jaccard Score measures the similarity for the two sets of data, which are the independent variables (frequency, average amount/purchase, cluster labels/segment) and target/dependent variable (customer response). f-1 score is to measure accuracy, while Log Loss measures the performance of a classification model where the prediction input is a probability value between 0 and 1. The goal of the machine learning models is to minimize this value. The highest value for Jaccard and f-1 score is 1, while the lowest possible Log Loss is 0 for a perfect model.

The score Jaccard Score and the f-1 score is quite high with 0.90 and 0.85 respectively, while the Log Loss is relatively low at 0.31.

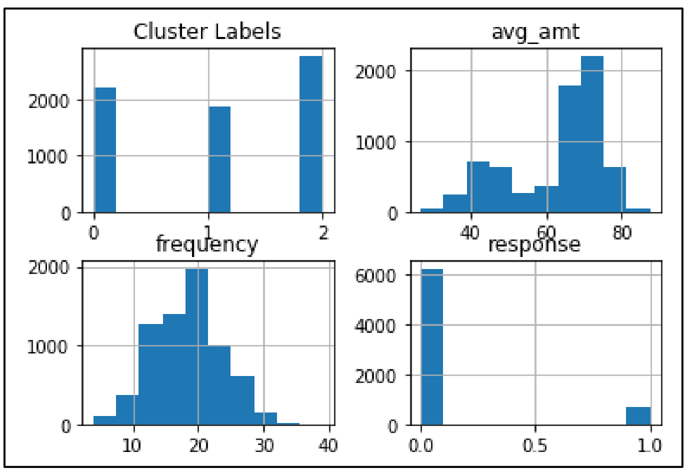

Figure 6: Visualization of all variables used in the Customer Response model

CONCLUSION AND RECOMMENDATIONS

Based on historical transactions, we have three customer segments:

Segment	Proportion	Average Frequency	Average Purchase Amount (Dollars)
The Most Loyal Customers	32%	24	70
The Average Customers	28%	13	44
The Potential Upgrade	40%	17	70

Figure 7: Customer Segmentation/Clustering

Recommendations based on segments for campaign planning:

- The management can focus on creating campaigns for the 'The Potential Upgrade' segment as they make up for 40% (the most) of the customers. The goal of the campaign should mainly focus on increasing their purchase frequency, for example by creating special events every month to attract customers and make them return to the store more regularly.
- The other option is to focus on the 'The Average Customers' to increase their ticket size or purchase amount per visit. The management can create bundling or package promotions to attract bigger purchase amount.
- The decision will largely depend on the store/company's available budget and financial projections based on the customers segment proportions. They may focus on 'The Potential Upgrade' strategies, 'The Average Customers' strategies, or they can probably afford to generally create all the possible campaigns to increase sales.

Other recommendations for future analysis:

1. The management needs to gather additional data on customer demographic information that is currently unavailable, so that the segmentation can be performed more thoroughly. Customer membership data updates mechanism needs to be established.
2. The customer response predicting model has quite favorable evaluation scores. However, since only 10% of the customers did respond to the previous campaign, the management needs to focus on increasing the campaign take-up rate as well.