Business Analytics with R Project BUAN 6356

Credit Card Customer Segmentation

in partial fulfillment for the award of the degree of

Masters in Business Analytics

Under the supervision of

Prof. Sun Shujing

Naveen Jindal School Of Management

Group Members:

1.Mounika Laxmi Kalidindi
2.Sai Sravya Bhupathiraju
3.Sathvik Boda
4.Hanisha Patel
5.Sohail Mohammed Shaik
6.Sai Aditya Jagarlamudi

Business Context/Problem Statement

A credit card is the payment card issued to customers to pay a retailer based on the cardholder's pledge to the card issuer to pay the amount back.The most difficult task for credit card companies is determining who they are selling to. Knowing your customer's personas will allow the company to tailor their targets and services to improve the company's sales.Customer segmentation is the method of breaking down a target market into smaller, more targeted groups.

Data Description and Understanding

Data Source

We extracted the data from the Kaggle.

Source Link: **https://www.kaggle.com/arjunbhasin2013/ccdata**

Data Challenges

- Due to insufficient exposure to the Finance/Banking domain sector, we have faced difficulty in understanding certain variables and link them to correlated variables.

Structure of the Data

The Credit Card Dataset consists of 8950 observations of 18 variables. All the variables are numerical except for Customer_ID.

Descriptive Statistics

- All the frequency columns are with a minimum value "zero" which means there are customers who do not update the balance and purchase frequently.
- Almost 50% of the data in the CASH_ADVANCE_TRX column is with value zero.

Missing/Abnormal Data

Data is missing for two variables. Minimum Payments- 313 missing records and Credit Limit - Only one missing record.

Data Cleaning/Pre-processing

Deletion: Out of 8950 records, the missing record in CREDIT_LIMIT variable accounts for 0.01% which is almost negligible data. Hence, we remove the record.

Imputation: 313 missing records in Minimum_Payments account for nearly 2% of data, so we did not delete these records, instead, we use the imputation methods to fill these missing values.

As the data is right/positively skewed with most of the data falling between 0 to 2000, we imputed the missing records with the **median** value.

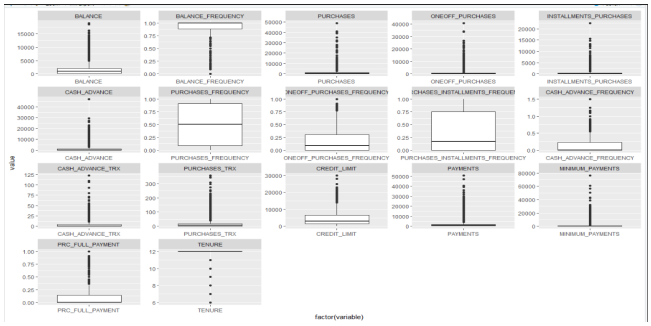
Categorization: Based on the type of the customer purchases, we have categorized into four categories.

Exploratory Data Analysis

Correlation and Multicollinearity

A correlation matrix graph is plotted to understand the correlation between two variables and check if there is any multicollinearity among the variables.

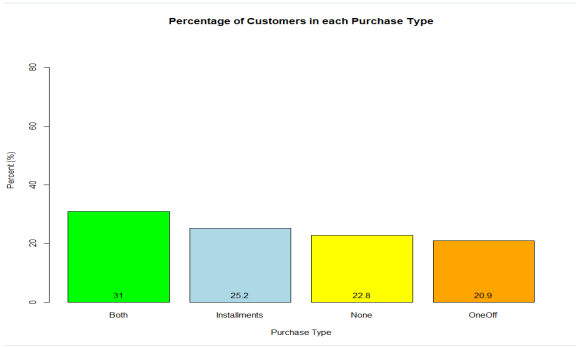
Outliers detection



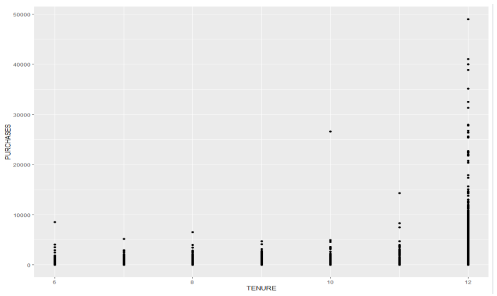
- BALANCE,PURCHASES,ONE-OFF-PURCHASES,CASH_ADVANCE,MINIMUM_PAYMENTS are more **skewed** and with too many outliers. We could normalize them by log transformation, delete the outliers or replace them with max/min in order to deal with the outliers.
- However, considering the variant purchase habits of customers, we did not delete the outliers and include them in the Segmentation.

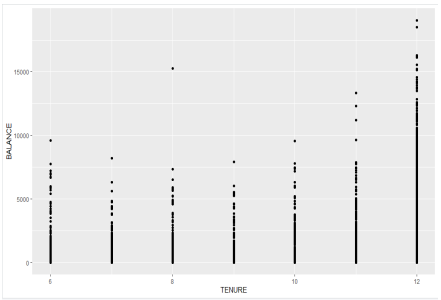
Data Proportion

- We observed that percentage of customers under each category of purchase type are balanced, and there is no particular category where there are fewer/ more customers.

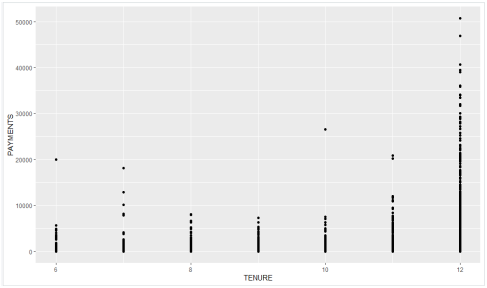


- **Scatter Plot for Purchase and Tenure:** From this, we could see that Purchase value is more for the customers who are associated with the bank for a year.



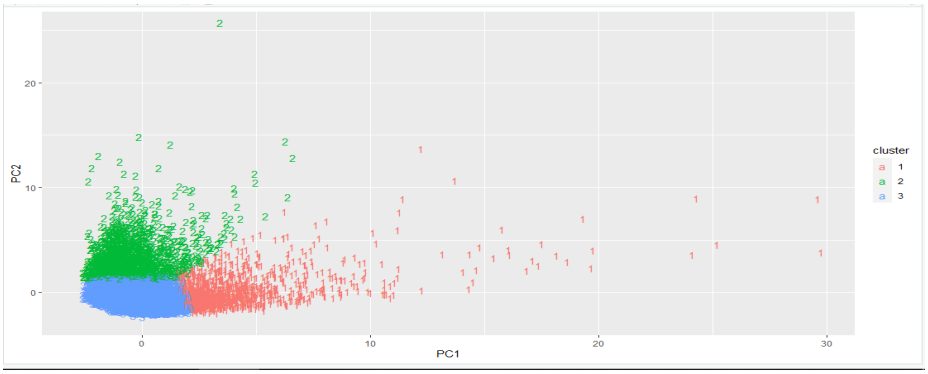


- **Scatter Plot for Tenure and Balance:** Balance value is more for customers associated with the bank for a year. This is because banks would have increased the credit limit for the customers after the completion of one year.
- **Scatter Plot for Tenure and Payments:** Payments increased drastically for the customers with tenure of 12 months.



Clustering:

- Using the elbow plot, we chose to segment the customers into three clusters.
- There are 6105 customers in Cluster 3 which is almost 65% of the data.
- Principal Component Analysis is used to reduce the dimensions without ignoring the meaning of significant variables. We normalized the data and applied PCA and observed change in variance of the data by adding each Principal component.
- At PC10, we observed the appropriate increase in cumulative proportion and decrease in Variance
- The ratio of between-cluster variation to within-cluster variation increased for this PCA+K means clustering method compared to the previous method.So, we used the results from this clustering method for our further analysis.



Interpretations :

Cluster 1 - Transactors

Customers in this cluster have the lowest balance and cash advance and have a high percentage of payments.These customers might have done a fair amount of transactions and paid less in terms of interest.So, they can be classified as Transactors.

Cluster 2 - Revolvers

Customers in this cluster have high balances and cash advances and low purchase frequencies and percentages of payments which indicates that they are one of the most lucrative segments for the providers.These customers can be classified as Revolvers as they use the credit card as loan.

Cluster 3 - Prime Customers

Customers have a high credit limit and the high percentage of payments. These are prime customers who make high purchases and more transactions.

Marketing Strategies:

Cluster-1, we observe that they have more purchases and less payment history in the last six months. We can target this cluster by offering them with lower APRs for a period of time. In this strategy, the cardholders pay reduced APRs. In the promotional period, the customers will show interest in clearing their dues.

Cluster-2, we observe that they have high cash advances. As this behaviour is not a good sign for the company,we can increase APR for cash advances which forces the customers to avoid withdrawing the cash from ATMs and in turn provide Cashbacks such as 3% for gas stations, 2% for groceries and 1% for any other payments that are made via online. This strategy encourages the customers to spend via online or swipe through machines rather than withdrawing cash from ATMs.

Cluster-3, we can infer that this cluster likely consists of affluent people so they might not be interested in the cashbacks. Instead we introduce promotional offers which include upgradation of economy class to business class tickets, premium lounge access and providing a monthly subscription of Netflix or Amazon Prime and issuing them premium titanium cards which makes them feel privileged which leads to loyalty for the company.