Comprehensive Project Report

Project Name – Credit Card Segmentation

--Kumar Rohit

Edwisor Trainee(Data Science Hiring Program)

Abstract

This case requires trainees to develop a customer segmentation to define marketing strategy. The sample dataset summarizes the usage behaviour of about 9000 active credit card holders during the last 6 months. The file is at a customer level with 18 behavioural variables.

I have used libraries like Python Pandas,Numpy,Matplotlib & Seaborn.Using these I am able to study about each predictor or variable in the data set. Studing of the variable gives an idea as to how each of them gives out a meaning for the data points. I am able to come up with relationships amongst the various variables by also visualizing them.

Table Of Contents

1.Introduction………………………………….…4

2.Data

2.1 Description………………………………..5

2.2 Steps Used in Data Processing …………….6-9

3. Data Visualization……………………..………10-17

4.Project Result-A Business Insight……...………18

Introduction

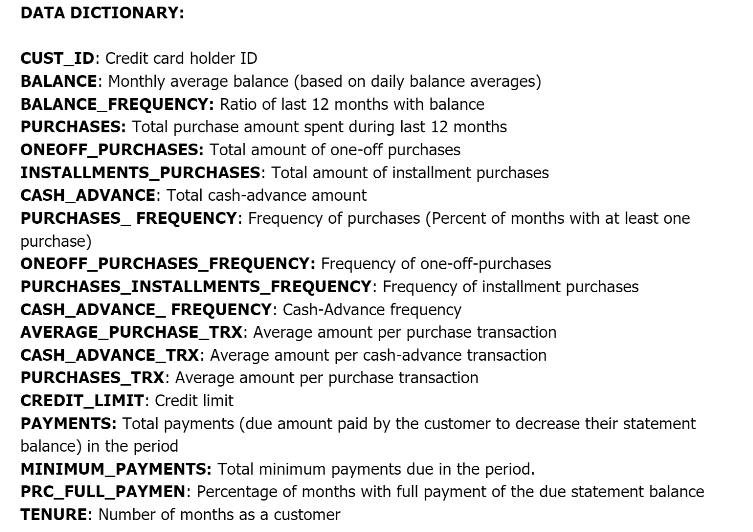
As there is lot of competition amongst banking and other financial institutions for credit card sales in certain time period as deemed fit by them. So, to reach out to maximum customers that may be a probable clients for sales closure they need a strategy for as to which client of theirs are more suitable for them. In my case this case requires to develop a customer segmentation to define marketing strategy.

Data

2.1 Data Description

The sample dataset summarizes the usage behaviour of about 9000 active credit card holders during the last 6 months. The file is at a customer level with 18 behavioural variables.

Data File Name: Credit-Card-Data.xlsx



**2.2 Steps Used in Data Processing**

**\*\*Data Pre-Processing:**

--We have loaded the data into dataframe

--Checked the data shape & its info

--Then we ran Descriptive Analysis on the dataset

--Checked the Null Values summation

##### -\*\*Missing Value Treatment

-Since there are missing values in the data so we are imputing them with median.

**(A). Deriving New KPI**

#### 1a. Monthly\_avg\_purchase and Cash Advance Amount: As the name suggest to find monthly average purchases & Advance cash amount

#### 1b. Purchase\_type : To find what type of purchases customers are making on credit card,lets explore the data.

Notes: I found out that there are 4 types of purchase behaviour in the data set.

1.People who only do One-Off Purchases.

2.People who only do Installments Purchases.

3.People who do both.

4.People who do none.

**So deriving a categorical variable based on the behaviour.**

#### 1c.Limit\_Usage (balance to credit limit ratio):

-Lower value implies customers are maintaining their balance properly. Lower value means good credit score

#### 1d.Payment to minimum payments Ratio

That how many customers clear their dues in full over to the minimum dues that they pay off.

##### Extreme value Treatment

* Since there are variables having extreme values, I am doing log-transformation on the dataset to remove outlier effect

### B.Insights from new KPI's

Average payment\_min payment ratio for each purchase type.

Mean payment\_minpayment ratio for each purchase type

##### *Insight 1: Customers With Installment Purchases are Paying Dues*

##### *Insight 2: Customers who don't do either one-off or installment purchases take more cash on advance*

##### *Insight 3: Customers with installment purchases have good credit score.*

### C.Preparing for Machine learning

Creating Dummies for categorical variable

Before applying PCA we will standardize data to avoid effect of scale on our result. Centring and Scaling will make all features with equal weight

##### Standardrizing data

To put data on the same scale

##### Applying PCA

# Since 5 components are explaining about 87% variance so we select 5 components

Factor Analysis : variance explained by each component-

##### D.Clustering

Pairwise relationship of components on the data

Key performace variable selection. Here i am dropping variables which are used in deriving new KPI

Concatenating labels found through K means with data

Mean value gives a good indication of the distribution of data. So we are finding mean value for each variable for each cluster

**Insights**

## Clusters are clearly distinguishing behaviour within customers

## Findings through clustering is validating Insights dervied from KPI. (as shown above in Insights from KPI

**Finding behaviour with 5 Clusters:**

# With 5 clusters :

* we have a group of customers (cluster 2) having highest average purchases but there is Cluster 4 also having highest cash advance & second highest purchase behaviour but their type of purchases are same.
* Cluster 0 and Cluster 4 are behaving similar in terms of Credit\_limit and have cash transactions is on higher side

**So we don't have quite distinguishable characteristics with 5 clusters**

# Finding behavior with 6 clusters

## Insights with 6 clusters

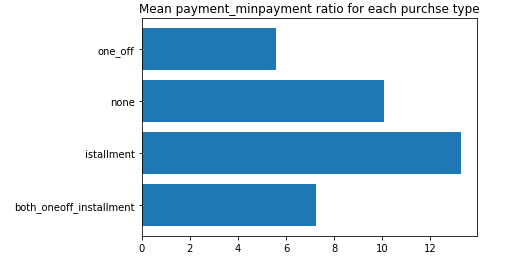
* Here also groups are overlapping .
  + Cl-0 and Cl-2 behaving same

# E.Checking performance metrics for Kmeans

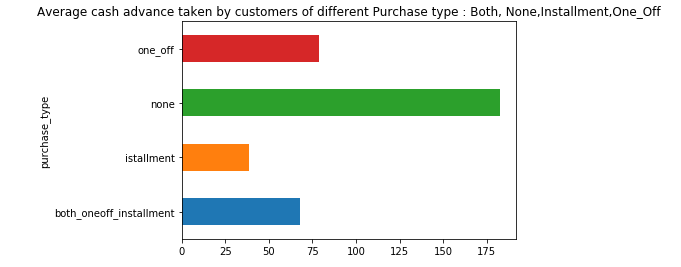
* I am validating performance with 2 metrics Calinski harabasz and Silhouette score

## Performance metrics also suggest that K-means with 4 cluster is able to show distinguished characteristics of each cluster.

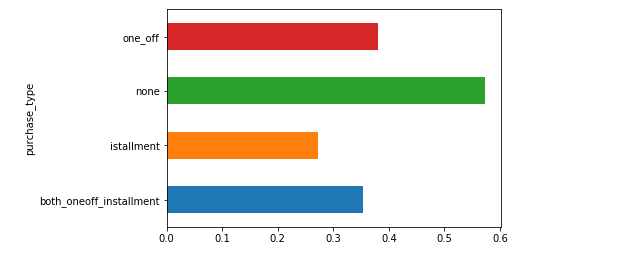
**3.Data Visualization**

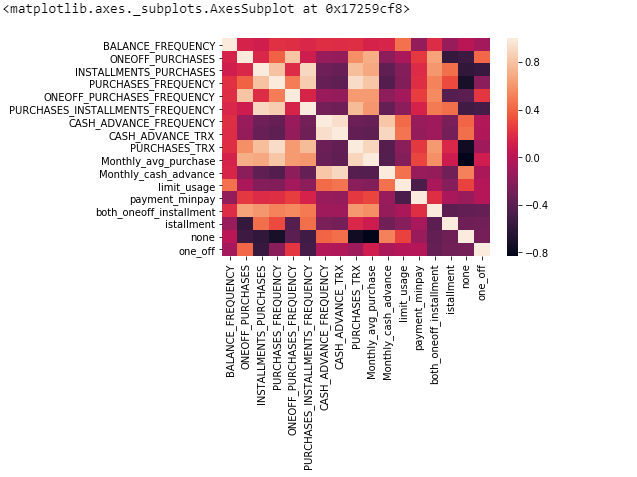


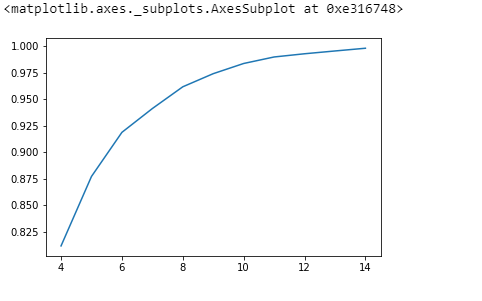
##### *Insight 1: Customers With Installment Purchases are Paying Dues*

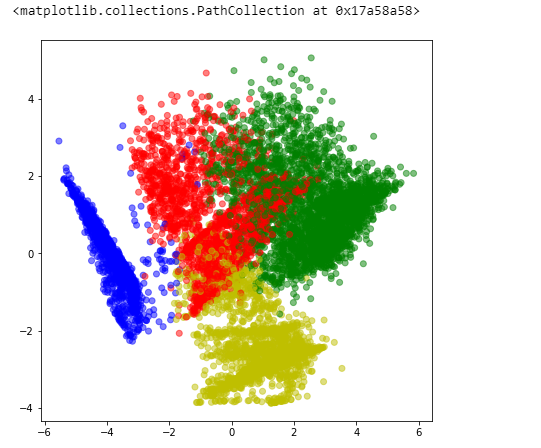


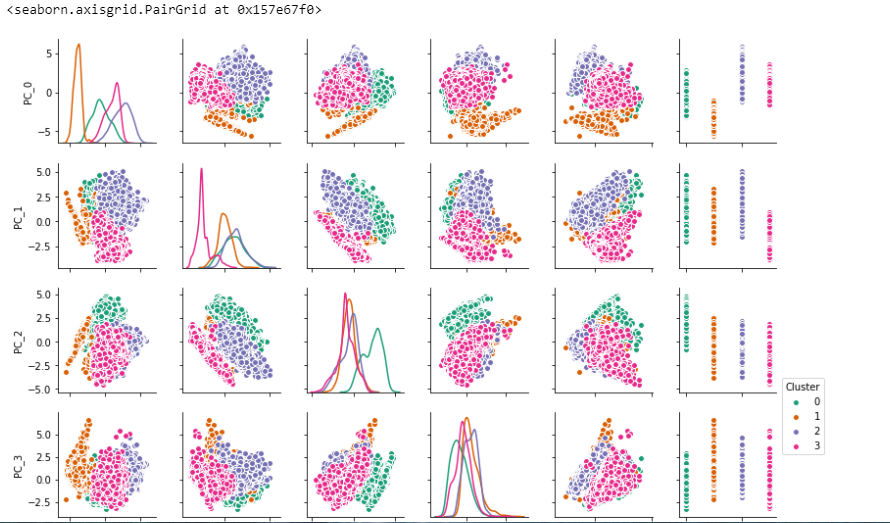
##### *Insight 2: Customers who don't do either one-off or installment purchases take more cash on advance*

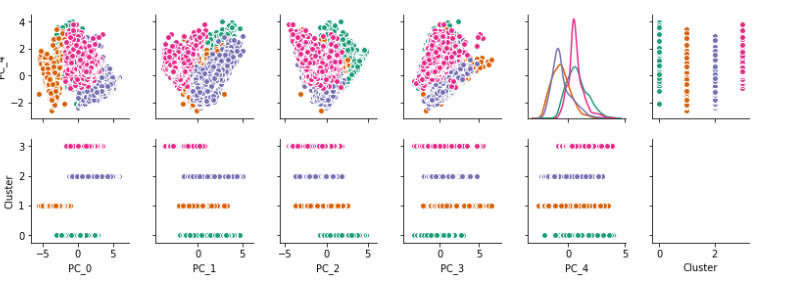


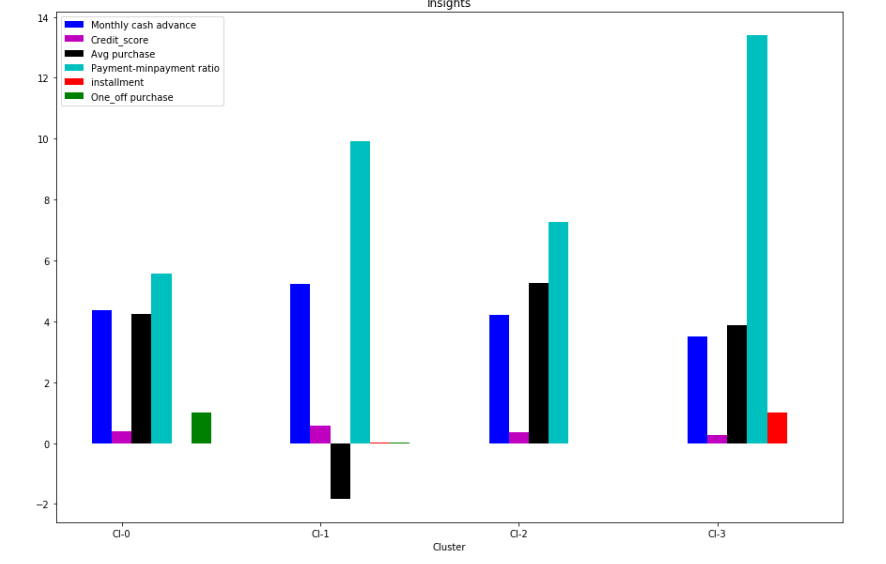


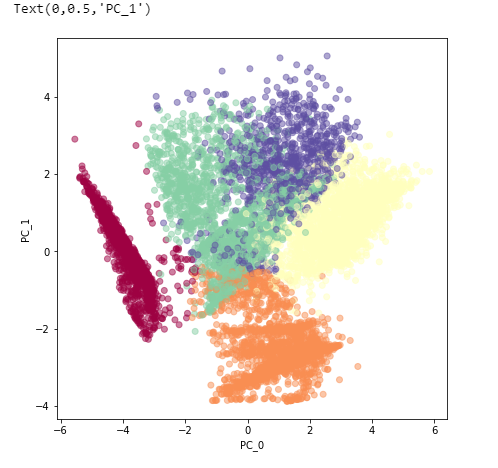


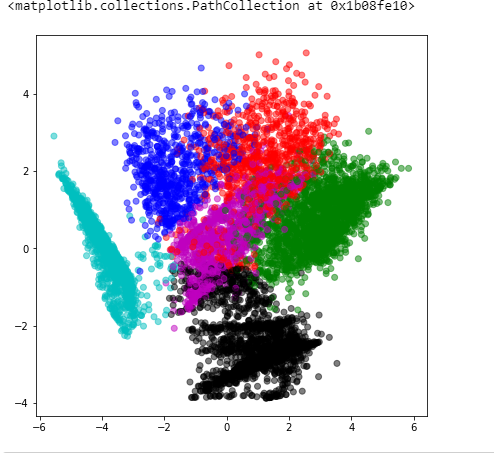


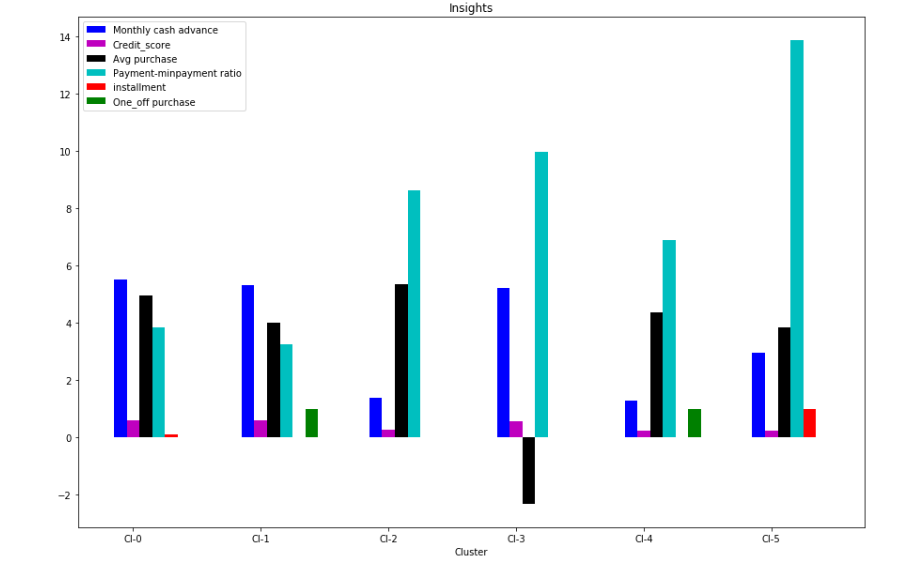


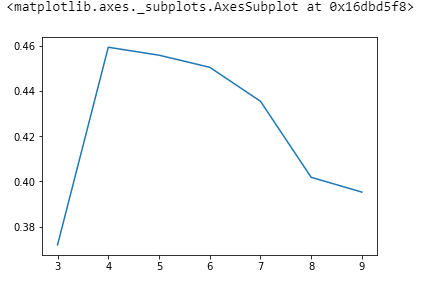


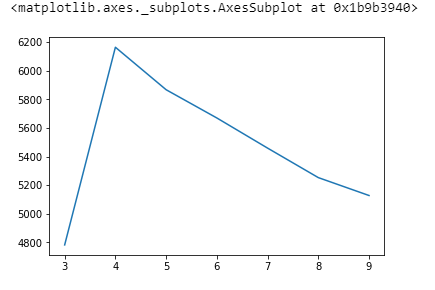












# 4.Project Result-A Business Insight

# Marketing Strategy Suggested:

## a. Group 2

* They are potential target customers who are paying dues and doing purchases and maintaining comparatively good credit score ) -- we can increase credit limit or can lower down interest rate -- Can be given premium card /loyality cards to increase transactions

## b. Group 1

* They have poor credit score and taking only cash on advance. We can target them by providing less interest rate on purchase transaction

## c. Group 0

* This group is has minimum paying ratio and using card for just oneoff transactions (may be for utility bills only). This group seems to be risky group.

## d. Group 3

* This group is performing best among all as cutomers are maintaining good credit score and paying dues on time. -- Giving rewards point will make them perform more purchases.