

Credit Card Segmentation

Clustering

Vishal Verma | Data Science

# Problem Statement

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

It is an unsupervised machine learning clustering problem

So I have used multiple cleaning and scaling process and found the best outcome using different hyper parameters

## Expectations & Solutions

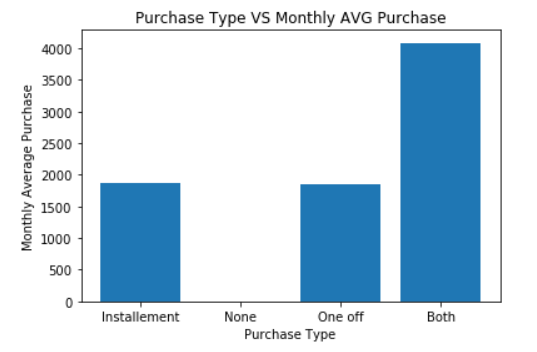
The three things that were expected by the student we’re full filled as mentioned below:

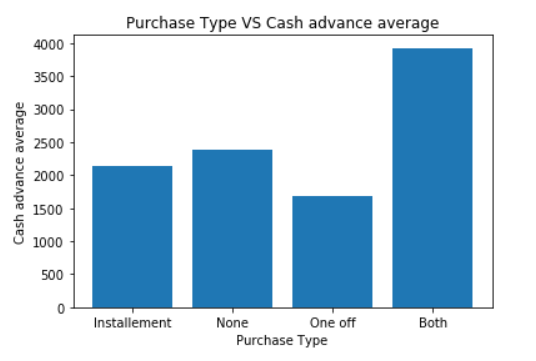
1. Advanced data preparation. Build an ‘enriched’ customer profile by deriving ‘intelligent’ KPI’s

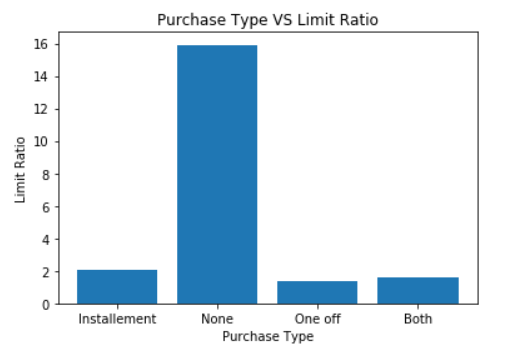
* Created all the KPI’s as mentioned in the problem and used them to further get knowledge about the data of the customers

1. Advanced reporting. Use the derived KPI’s to gain insight on the customer profiles

* Used the derived KPI’s to get insight about the customer.
* There were four categories of customer which had a different outlook on purchases made by the credit cards.
* Some were using it just for cash advance while others used it for one off purchases or installment purchases or both.
* The plots attached below show the insights generated from the KPI’s

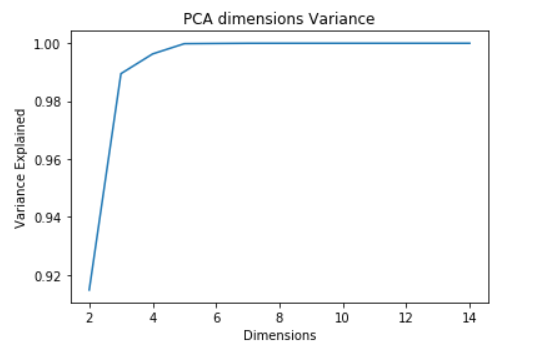




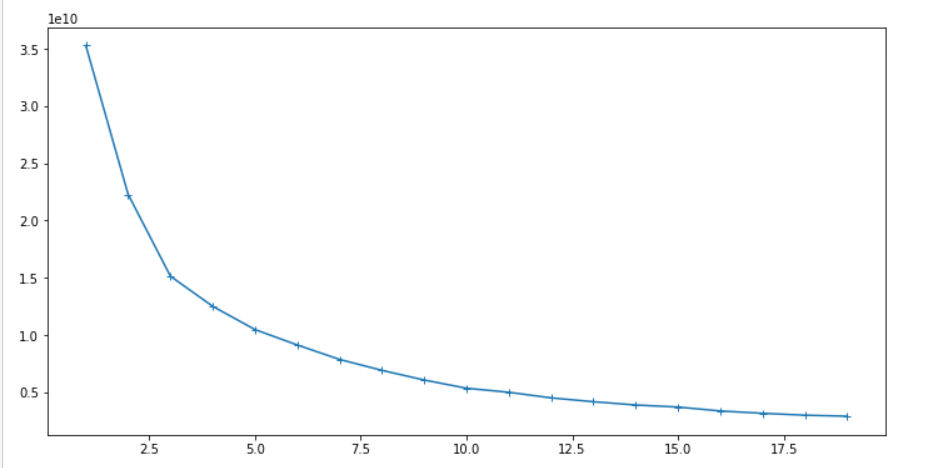


1. Clustering of the credit card customers

* Using the dimension reduction factor analysis using PCA we were able to reduce the dimensionality of the data set.
* To reduce the dimensionality we used PCA with different no. of variables and plotted a graph for that.



* As forecasted in the graph we used 5 variables for the PCA as they were explaining a good amount of variance in the dataset.
* With which we applied a K-means clustering model for which I used elbow method to get the best value for the number of clusters.



* So I used this to get the no. of clusters for the dataset and was able to get the clusters of customers.





* So I used 4 & 5 as the cluster no. for cluster analysis and got two different graph but we’re going with 4 as they don’t overlap much and create a clear distinction between the clusters.

## Summary

This model can help the business in achieving strategic goals by using it as a way to configure or give offers to different set of clusters of customer.

Each cluster represents a certain type of customer with a particular type of purchase habit. So the company can strategies by giving offers to the customer with less spending so that they use the credit card and company can profit on that

They can also offer different type of credit cards based on their limit of usage and can get a better interest rate on the better credit card.

# Instructions to run the code

## Python

* Open Jupyter notebook in the directory where you have stored the solution file
* Change the working directory to get the data from the dataset provided
* Run the cells one by one using shift + enter.

## R

* Open R studio
* Open the R solution file
* Change the working directory to get the data from the dataset provided
* Run the commands in R studio one by one using ctrl+enter