**Project 3) - End to End Case Study-BankingCredit Card segmentation**

**Objective** –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.

**Data Availability & Business Problem:**

**CUST\_ID**: Credit card holder ID

**BALANCE**: Monthly average balance (based on daily balance averages)

**BALANCE\_FREQUENCY:** Ratio of last 12 months with balance

**PURCHASES:** Total purchase amount spent during last 12 months

**ONEOFF\_PURCHASES:** Total amount of one-off purchases

**INSTALLMENTS\_PURCHASES**: Total amount of installment purchases

**CASH\_ADVANCE**: Total cash-advance amount

**PURCHASES\_ FREQUENCY**: Frequency of purchases (Percent of months with at least one

purchase)

**ONEOFF\_PURCHASES\_FREQUENCY:** Frequency of one-off-purchases

**PURCHASES\_INSTALLMENTS\_FREQUENCY**: Frequency of installment purchases

**CASH\_ADVANCE\_ FREQUENCY**: Cash-Advance frequency

**AVERAGE\_PURCHASE\_TRX**: Average amount per purchase transaction

**CASH\_ADVANCE\_TRX**: Average amount per cash-advance transaction

**PURCHASES\_TRX**: Average amount per purchase transaction

**CREDIT\_LIMIT**: Credit limit

**PAYMENTS:** Total payments (due amount paid by the customer to decrease their statement

balance) in the period

**MINIMUM\_PAYMENTS:** Total minimum payments due in the period.

**PRC\_FULL\_PAYMEN**: Percentage of months with full payment of the due statement balance

**TENURE**: Number of months as a customer

Data =CC\_GENERAL.csv

**Identify the characteristics of each segment and define the strategy of each segment based on the business problem.**

**Historical Data** –

* Data of 18 attributes and 3 other attribute created based on problem of Customer
* Data of 8950 customers

**Tools Used –** Python (numpy, pandas, matplotlib, , sklearn,PCA,Elbowanalysis,silhoute coefficient,kmeans,, etc.), Excel

**Techniques –**

### Clustering evaluation (Finding optimal number of clusters)

This explains how to find optimal number of clusters in a given dataset by using various techniques. Different techniques discussed here are

* Dendogram
* Elbow method
* Silhoutte score Analysis We will first load the data into dataframe and scale the features and create clusters. And then various metrics are calculated to validate the number of cluster creations and what will be the optimal number of clusters. The Silhouette Coefficient is a common metric for evaluating clustering "performance" in situations when the "true" cluster assignments are not known.

**Challenges** –

* Data Collection
* Feature Selection

------------------------------------------------------------------------------------------------------------------------------------------