

# Credit Card Segmentation

Submitted by:

Ganesh Karthik Kommineni

# Introduction

## 1.1 Problem Statement

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.

## 1.2 DATA

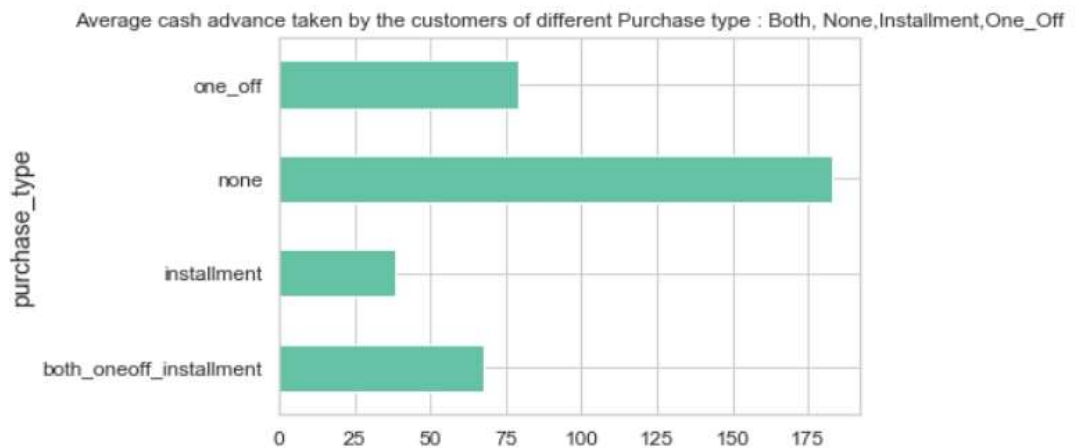
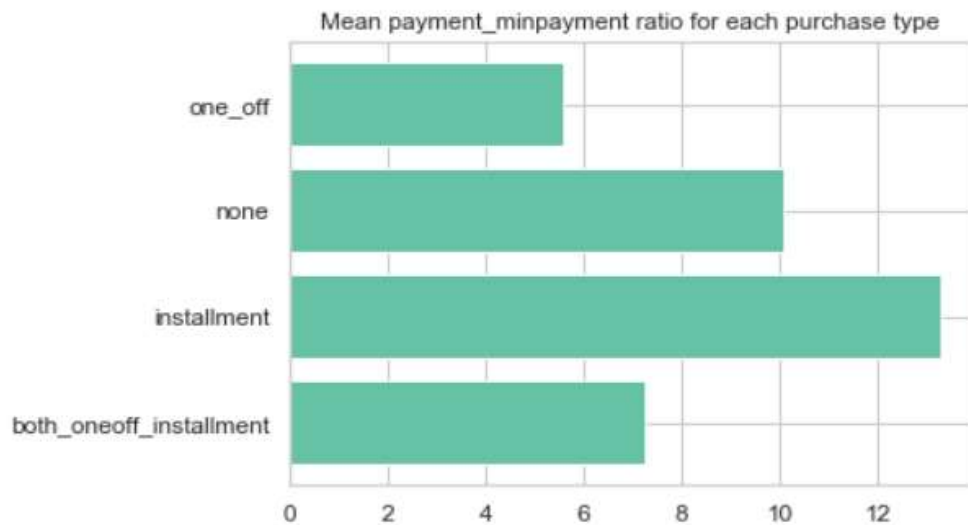
### Attributes:-

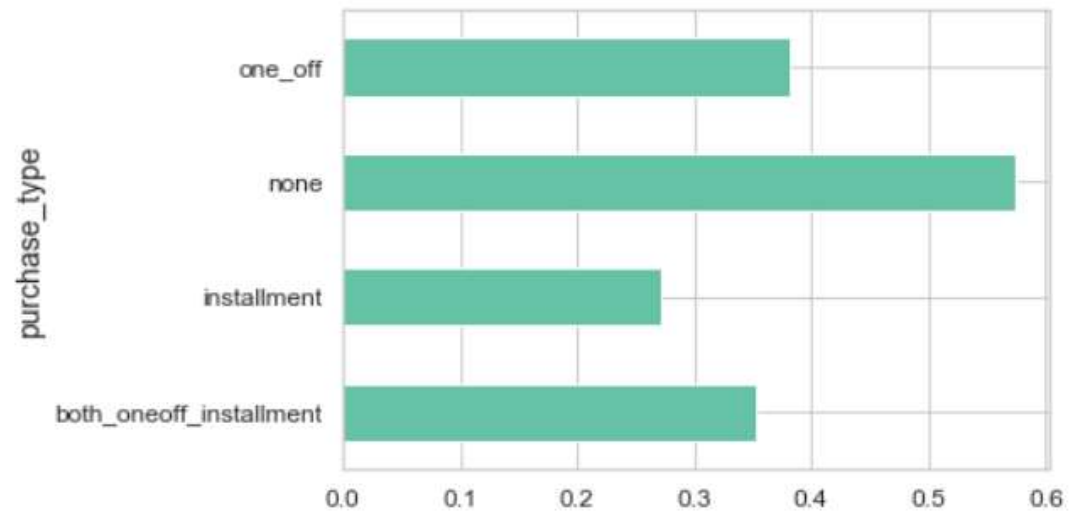
1. CUST\_ID Credit card holder ID
2. BALANCE Monthly average balance (based on daily balance averages)
3. BALANCE\_FREQUENCY Ratio of last 12 months with balance
4. PURCHASES Total purchase amount spent during last 12 months
5. ONEOFF\_PURCHASES Total amount of one-off purchases
6. INSTALLMENTS\_PURCHASES Total amount of installment purchases
7. CASH\_ADVANCE Total cash-advance amount
8. PURCHASES\_FREQUENCY-Frequency of purchases (percentage of months with at least on purchase)
9. ONEOFF\_PURCHASES\_FREQUENCY Frequency of one-off-purchases
10. PURCHASES\_INSTALLMENTS\_FREQUENCY Frequency of installment purchases
11. CASH\_ADVANCE\_FREQUENCY Cash-Advance frequency
12. AVERAGE\_PURCHASE\_TRX Average amount per purchase transaction
13. CASH\_ADVANCE\_TRX Average amount per cash-advance transaction
14. PURCHASES\_TRX Average amount per purchase transaction
15. CREDIT\_LIMIT Credit limit
16. PAYMENTS-Total payments (due amount paid by the customer to decrease their statement balance) in the period
17. MINIMUM\_PAYMENTS Total minimum payments due in the period.
18. PRC\_FULL\_PAYMENT- Percentage of months with full payment of the due statement balance
19. TENURE Number of months as a customer

## IMPLEMENTATION

As given in problem statement we have to define a marketing strategy by dividing customers into groups based on their purchases, cash payments, credit limits, advance payments etc. by using Clustering Algorithm.

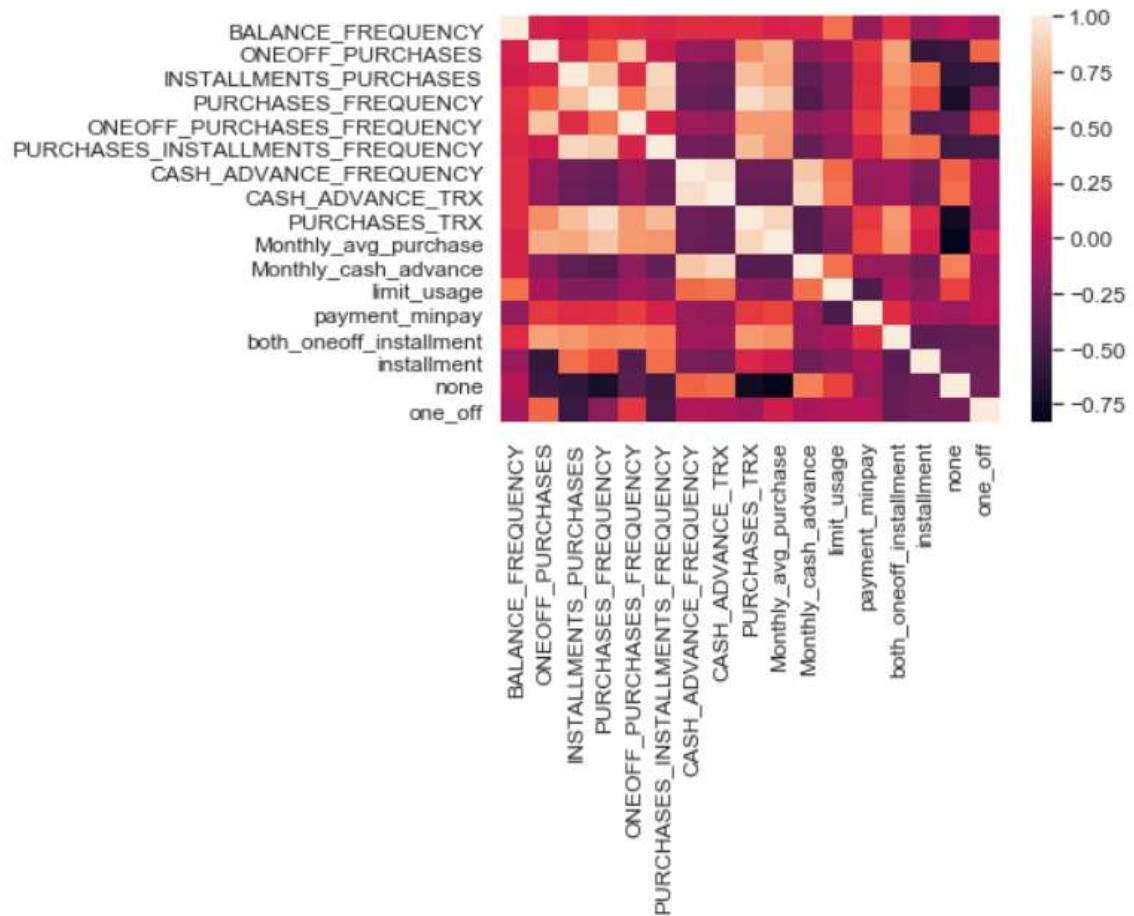
After treating missing values and by deriving new Intelligent KPI's we used plots and got the following insights:-



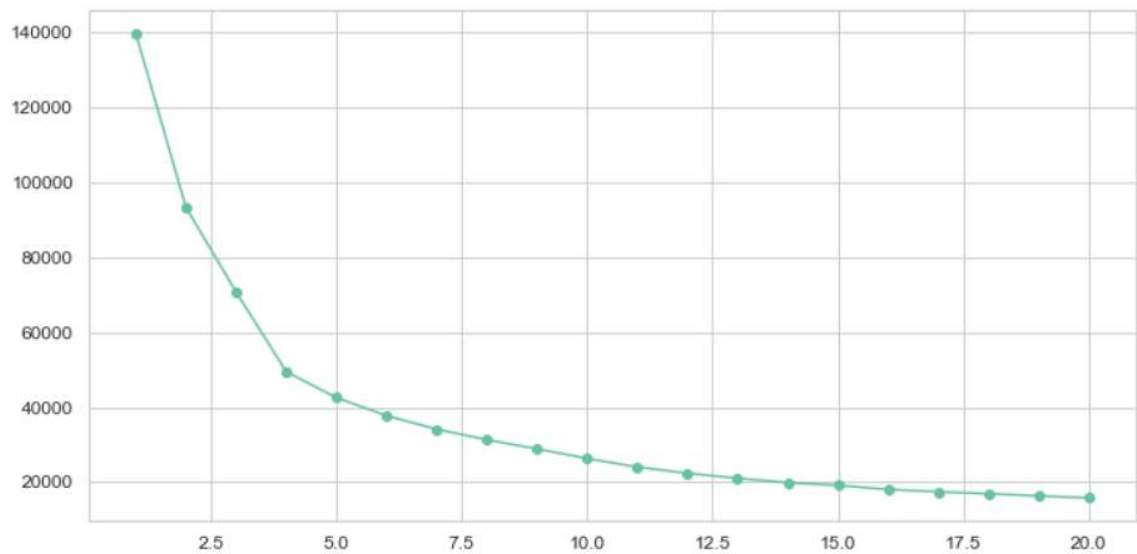


## CORRELATION HEATMAP

Heat map shows that how features are co-related so that applying dimensionality reduction will help negating multi-collinearity in data.

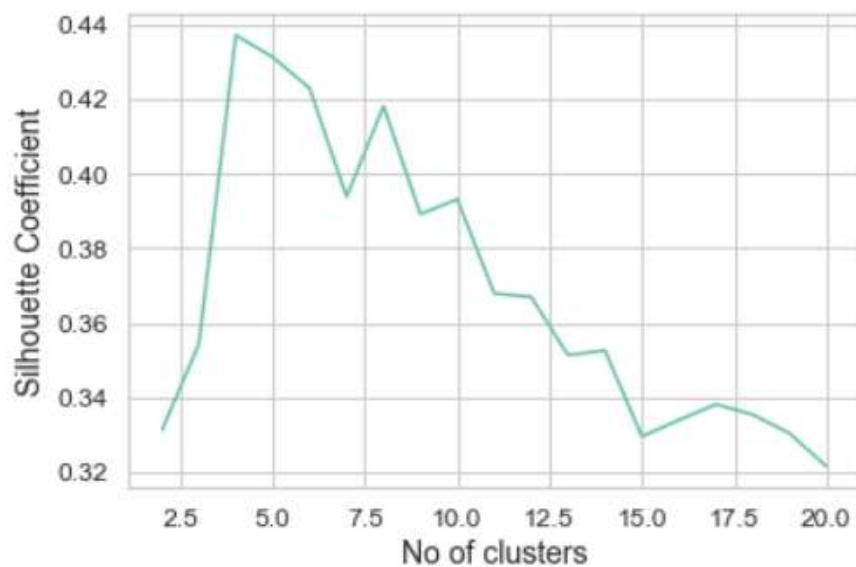


## The Elbow Method

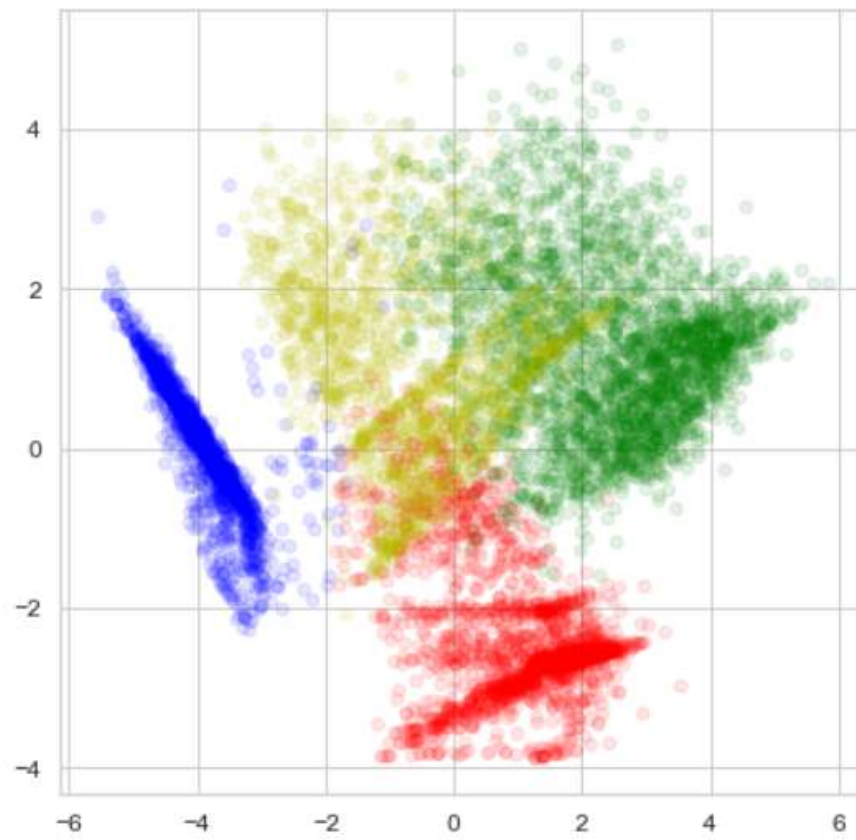


- From the above graph elbow range is 4, 5, and 6. Therefore we can show the behaviour in 4 or 5 or 6 clusters.

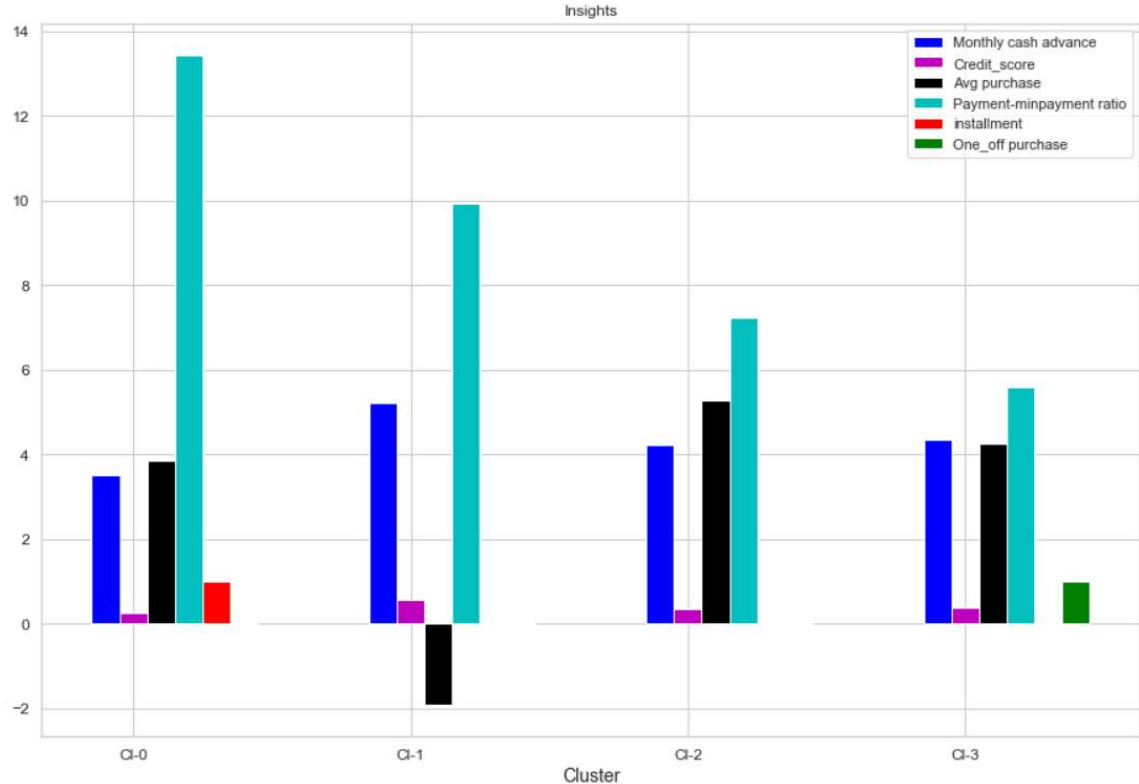
## Silhouette Co-efficient



**Scatter plot of 4 Clusters**



## INSIGHTS OF 4-CLUSTERS THROUGH BARGRAPH



### Clusters are clearly distinguishing behaviour within the customers

**Cluster 0** is the group of customers who have highest Monthly\_avg purchases and doing both instalment as well as one\_off purchases, have comparatively good credit score. This group is about 31% of the total customer base.

**Cluster 1** is taking maximum advance\_cash and is paying comparatively less minimum payment and poor credit\_score & doing no purchase transaction. This group is about 23% of the total customer base.

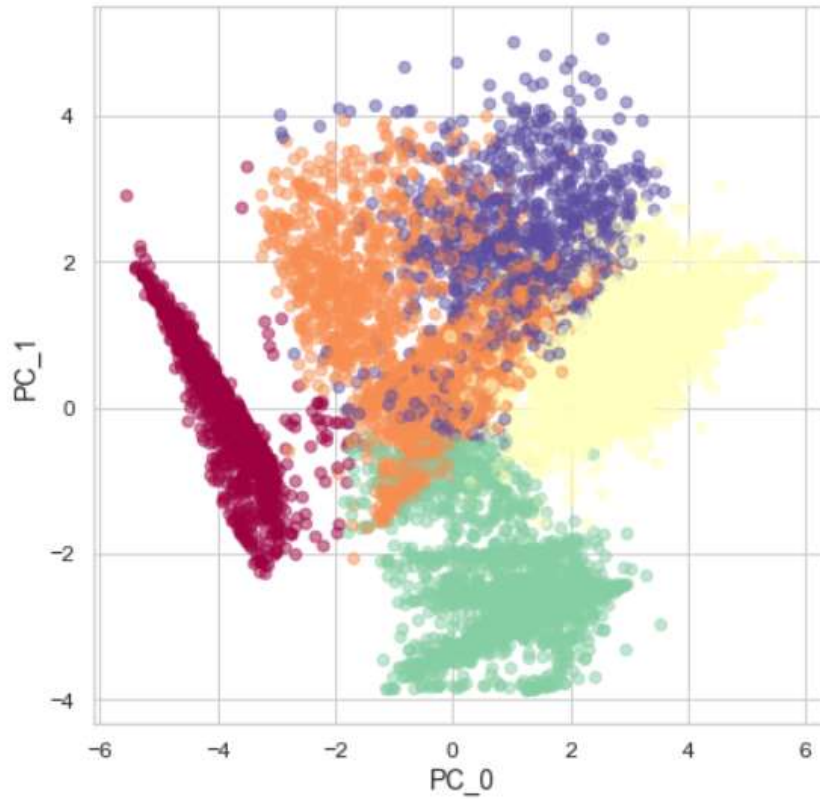
**Cluster 3** customers are doing maximum One\_Off transactions and least payment ratio. This group is about 21% of the total customer base.

**Cluster 2** customers have maximum credit score and are paying dues and are doing maximum instalment purchases. This group is about 25% of the total customer base.



## Behaviour with 5 Clusters

Scatter plot:-

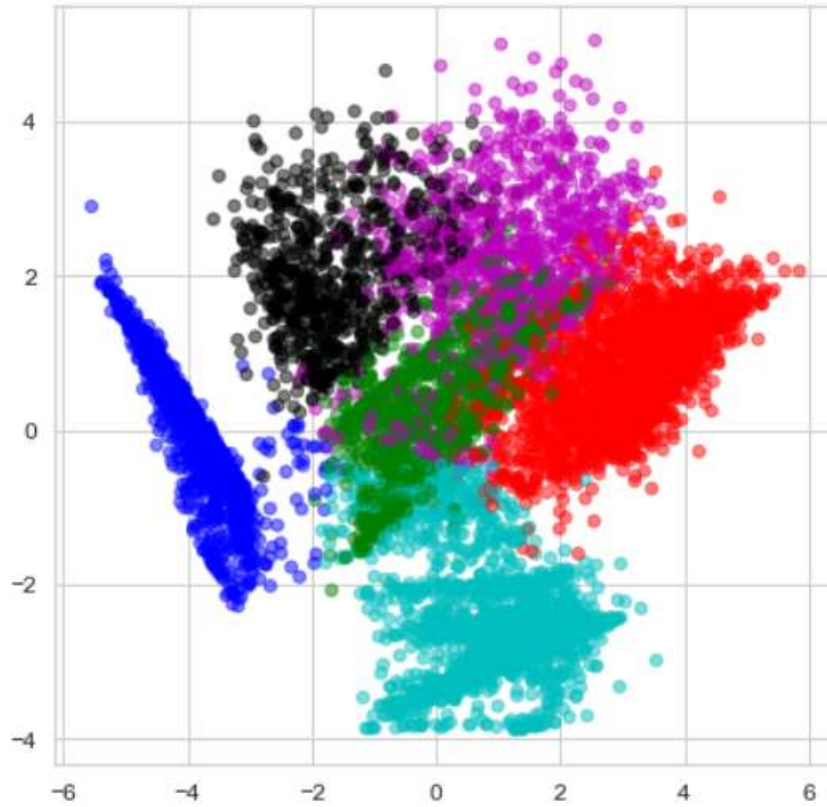


**Conclusion with 5 clusters:**

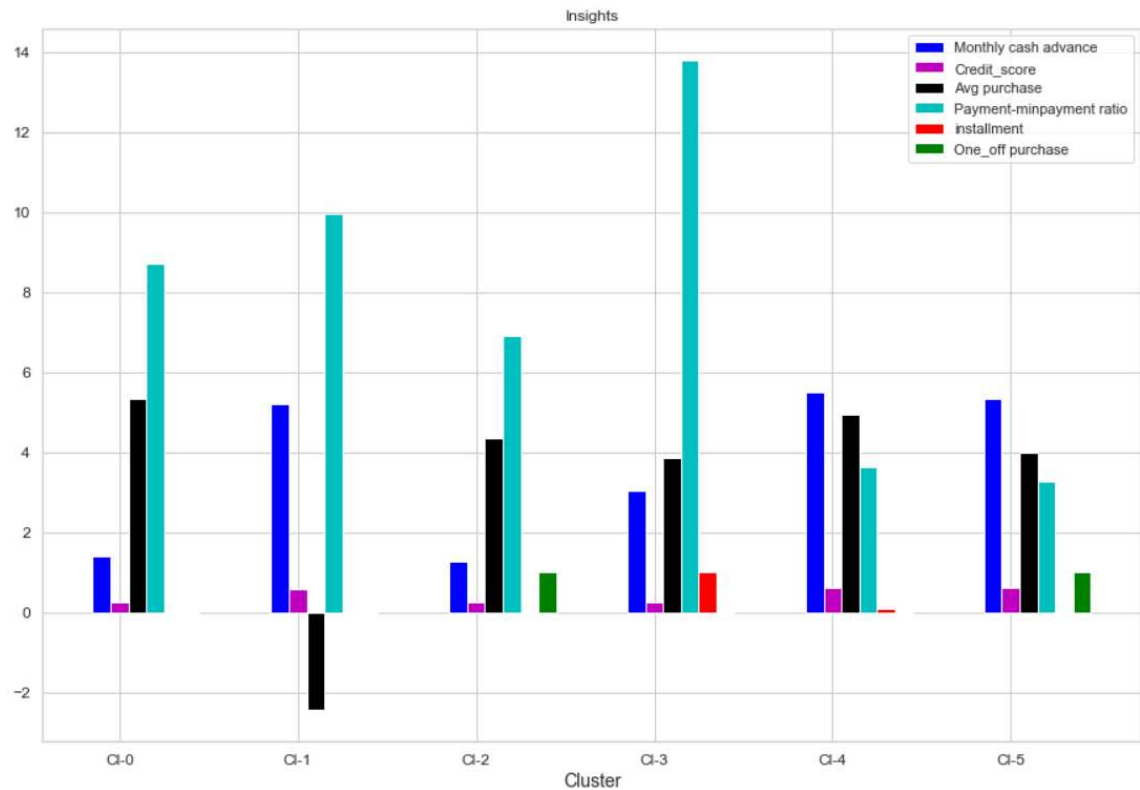
- We have a group of customers (cluster 0) having highest average purchases but there is Cluster 4 also having highest cash advance & second highest purchase behaviour but their type of purchases are same.
- Cluster 0 and Cluster 4 are behaving similar in terms of Credit\_limit and have cash transactions is on higher side.
- So we don't have to quite distinguishable characteristics with 5 clusters

## Behaviour with 6 clusters

Scatter plot:-



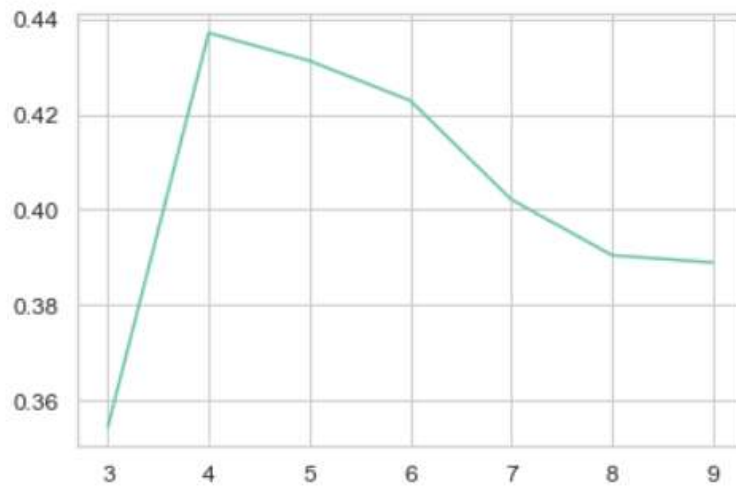
## INSIGHTS OF 6 CLUSTERS USING BARGRAPH



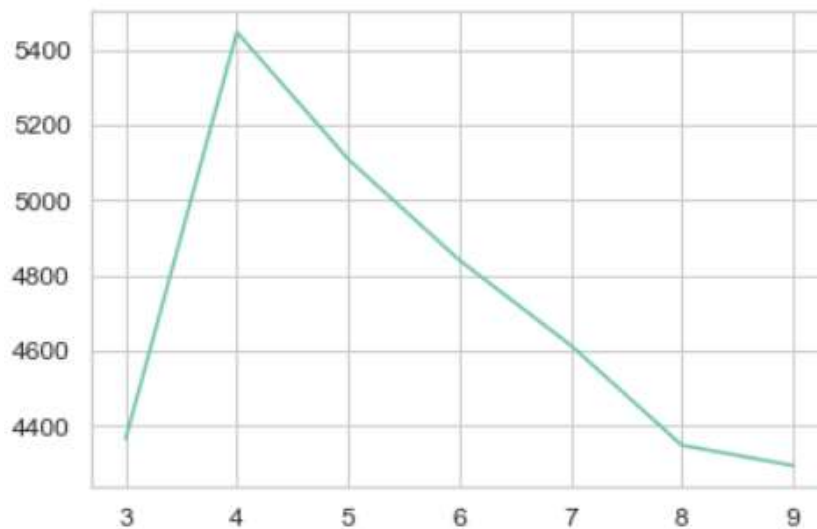
### Conclusion with the 6 clusters

- Here also groups are overlapping.
- Cl-1 and Cl-2 behaving same.

## Performance Metrics for K-means clustering



Calinski harabaz score



Silhouette score

- The performance metrics also says that K-means with 4 clusters is able to show distinguished characteristics of each of the clusters.

## Insights with 4-clusters

**Cluster 0** is the group of customers who have highest Monthly\_avg purchases and doing both instalment as well as one\_off purchases, have comparatively good credit score. This group is about 31% of the total customer base.

**Cluster 1** is taking maximum advance\_cash and is paying comparatively less minimum payment and poor credit\_score & doing no purchase transaction. This group is about 23% of the total customer base.

**Cluster 3** customers are doing maximum One\_Off transactions and least payment ratio. This group is about 21% of the total customer base.

**Cluster 2** customers have maximum credit score and are paying dues and are doing maximum instalment purchases. This group is about 25% of the total customer base.

## **Suggested Target Marketing Strategy:**

### **1. Group 0(cluster-0)**

They are potentially targeted customers who are paying their dues and making purchases and maintaining comparatively very good credit score. They can be offered:

- By increasing credit limit or by lowering down interest rate.
- By giving Premium card / loyalty cards to increase their transactions.

### **2. Group 1(cluster-1)**

This group of customers have poor credit score and taking cash on advance only, so we can target them by providing less interest rates on purchase transactions.

### **3. Group 3(cluster-3)**

This group of customers have minimum paying ratio and using card for just one-off the transactions may be some home bills. They seem to be a risky group.

### **4. Group 2(cluster-2)**

This group of customers are performing best among all as they are maintaining good credit score and paying dues on time.

- We can give them reward points which will make them for more purchases.