Credit Card Segmentation

Akshay Rahate

December 19, 2020

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.

Data Attributes

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

Data Attributes

PURCHASES_FREQUENCY - Frequency of purchases (percentage of months with at least on 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

Data Attributes

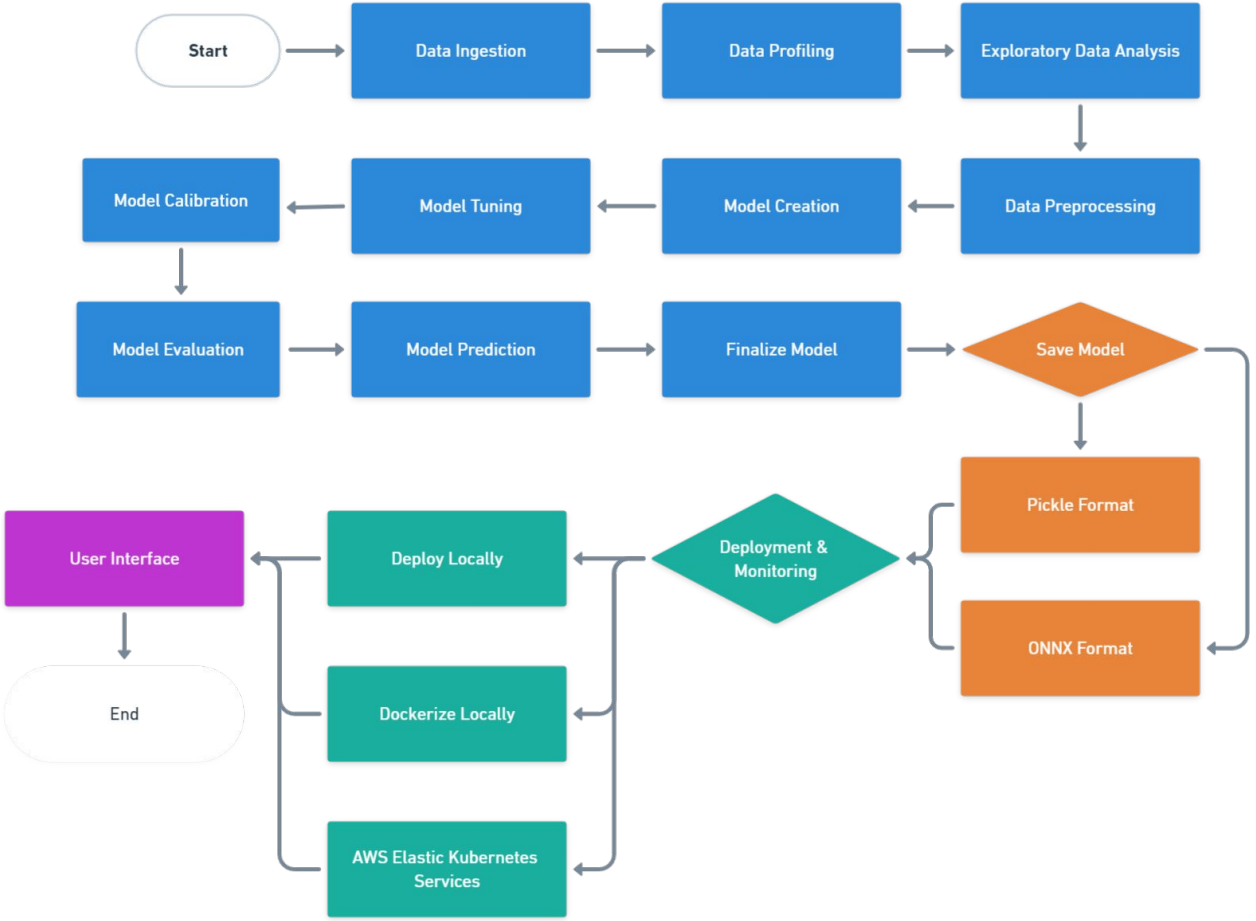
PAYMENTS - Total payments (due amount paid by the customer to decrease their statement balance) in the period

MINIMUM_PAYMENTS Total minimum payments due in the period.

3 PRC_FULL_PAYMENT - Percentage of months with full payment of the due statement balance

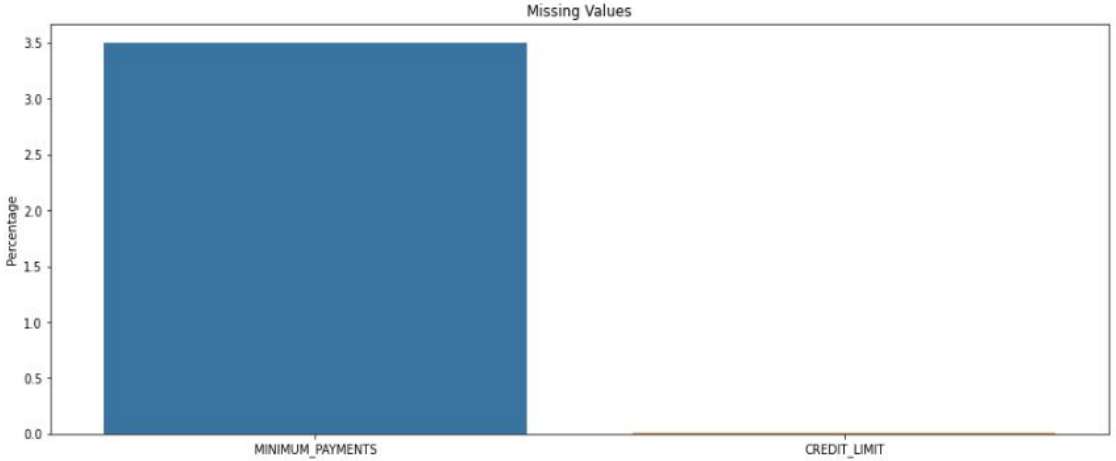
TENURE Number of months as a customer

High Level Architecture



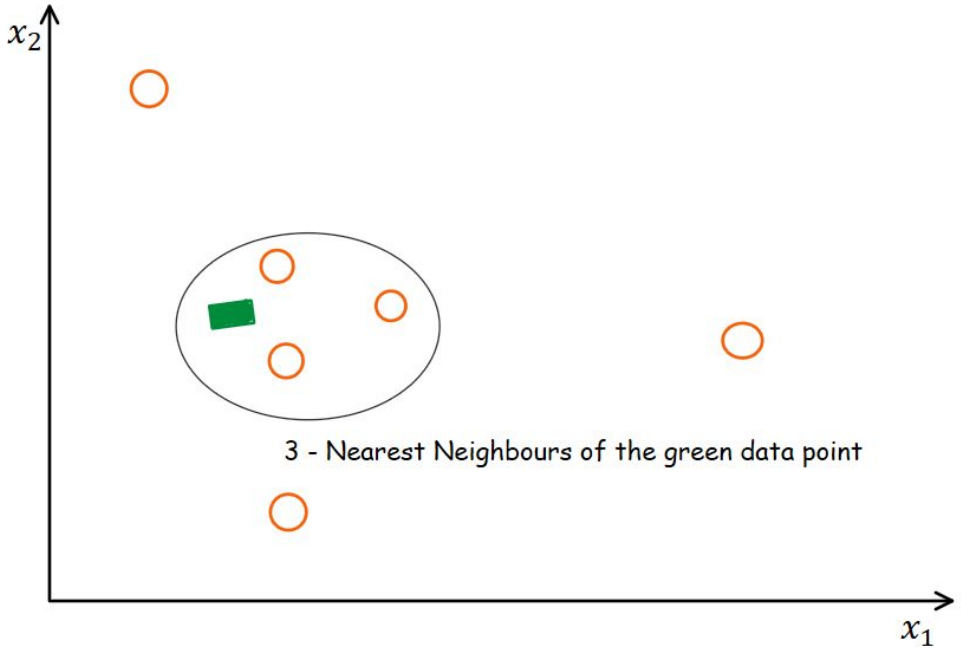
Missing Value Analysis

BALANCE	0
BALANCE_FREQUENCY	0
PURCHASES	0
ONEOFF_PURCHASES	0
INSTALLMENTS_PURCHASES	0
CASH_ADVANCE	0
PURCHASES_FREQUENCY	0
ONEOFF_PURCHASES_FREQUENCY	0
PURCHASES_INSTALLMENTS_FREQUENCY	0
CASH_ADVANCE_FREQUENCY	0
CASH_ADVANCE_TRX	0
PURCHASES_TRX	0
CREDIT_LIMIT	1
PAYMENTS	0
MINIMUM_PAYMENTS	313
PRC_FULL_PAYMENT	0
TENURE	0



kNN Imputation

- Imputes missing data by finding the k closest neighbors to the observation with missing data and then imputing them based on the the non-missing values in the neighbors.



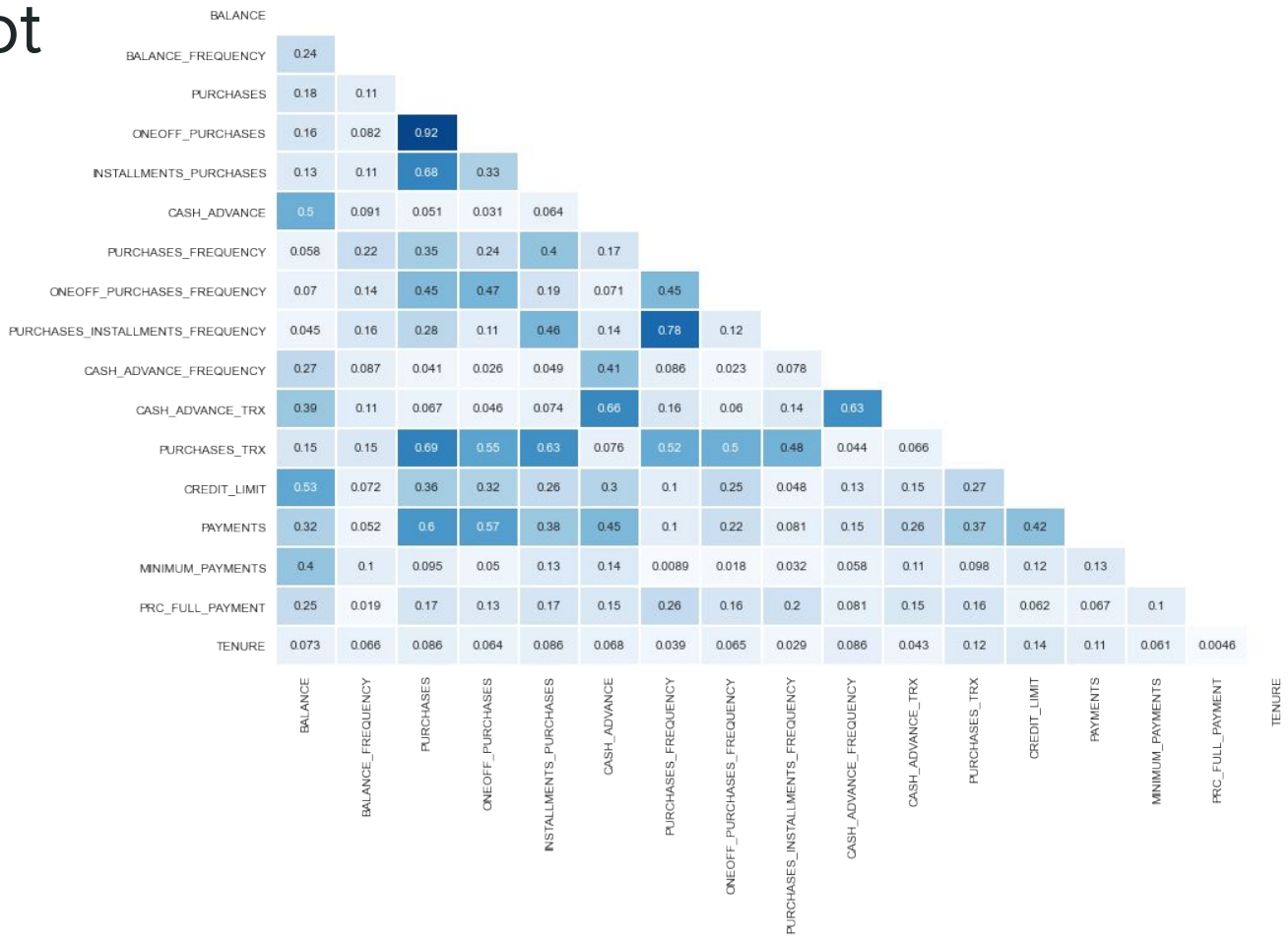
Outlier Analysis

```
BALANCE_FREQUENCY
# values outlier:  0
0.00% of the total data
-------------------------
PURCHASES
# values outlier:  80
0.89% of the total data
-------------------------
ONEOFF_PURCHASES
# values outlier:  74
0.83% of the total data
-------------------------
INSTALLMENTS_PURCHASES
# values outlier:  79
0.88% of the total data
-------------------------
CASH_ADVANCE
# values outlier:  98
1.09% of the total data
-------------------------
PURCHASES_FREQUENCY
# values outlier:  0
0.00% of the total data
```

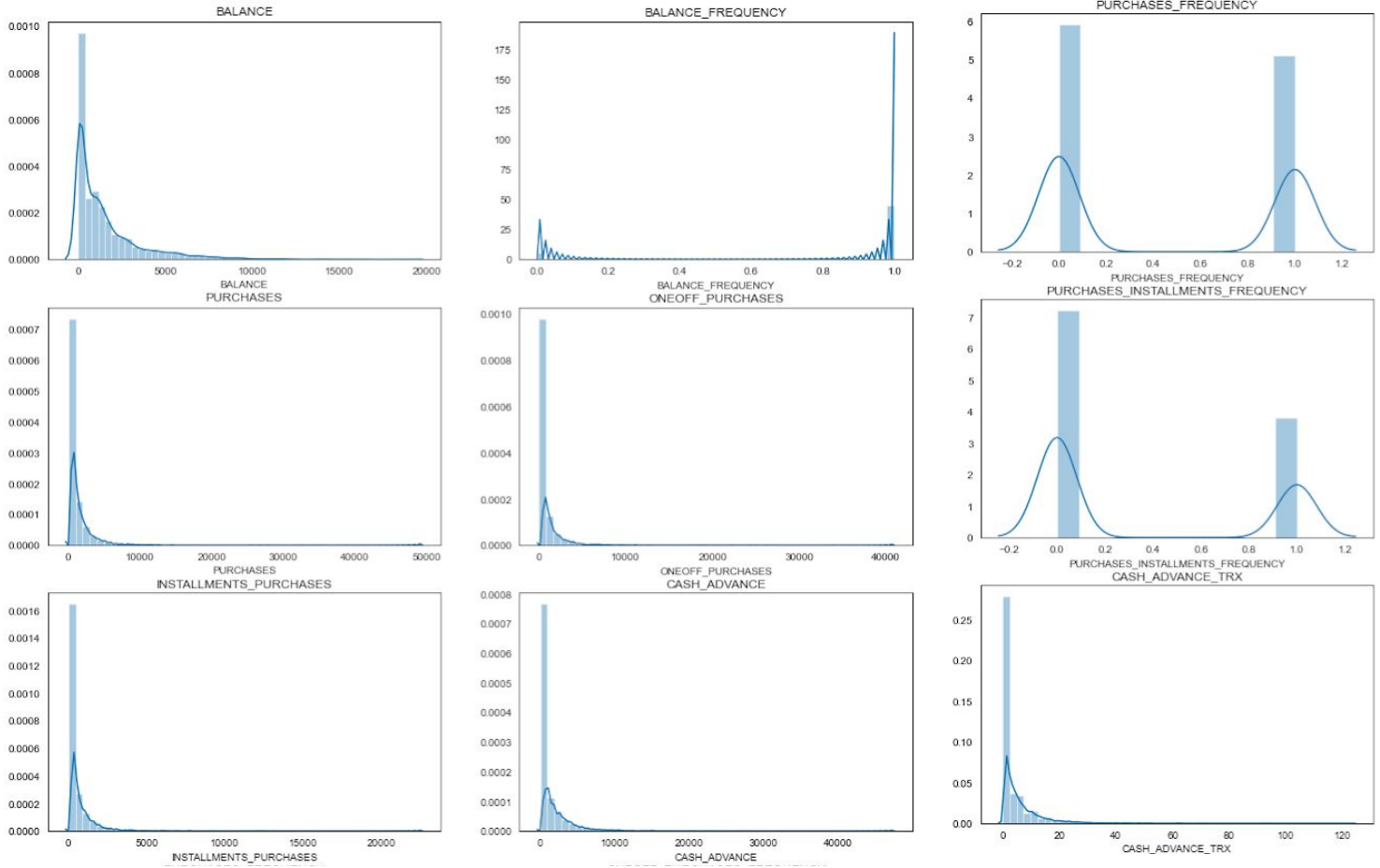
```
ONEOFF_PURCHASES_FREQUENCY
# values outlier:  0
0.00% of the total data
-------------------------------
PURCHASES_INSTALLMENTS_FREQUENC
# values outlier:  0
0.00% of the total data
-------------------------------
CASH_ADVANCE_FREQUENCY
# values outlier:  1
0.01% of the total data
-------------------------------
CASH_ADVANCE_TRX
# values outlier:  80
0.89% of the total data
-------------------------------
PURCHASES_TRX
# values outlier:  95
1.06% of the total data
-------------------------------
CREDIT_LIMIT
# values outlier:  29
0.32% of the total data
```

```
PAYMENTS
# values outlier:  92
1.03% of the total data
-------------------------------
MINIMUM_PAYMENTS
# values outlier:  78
0.87% of the total data
-------------------------------
PRC_FULL_PAYMENT
# values outlier:  0
0.00% of the total data
-------------------------------
TENURE
# values outlier:  204
2.28% of the total data
```

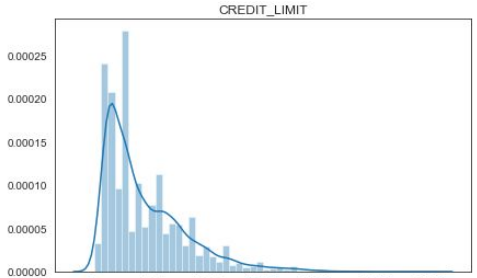
Correlation Plot

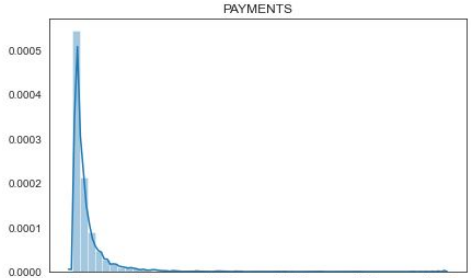


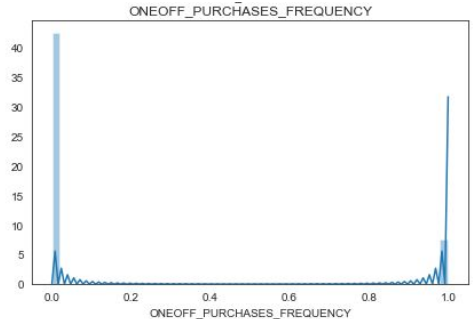
Univariate Analysis

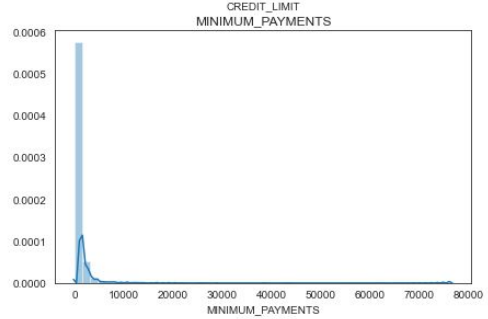


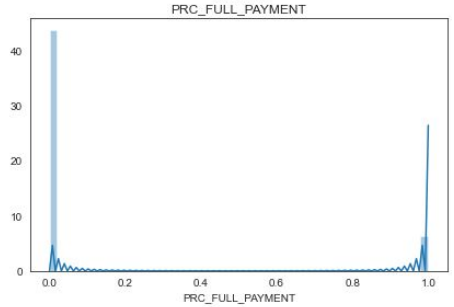
Univariate Analysis

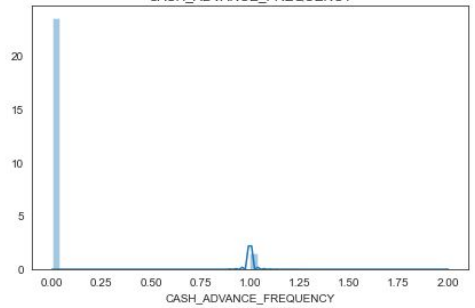


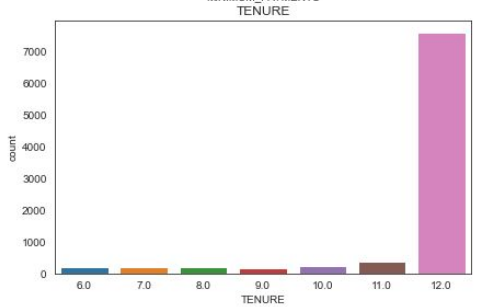


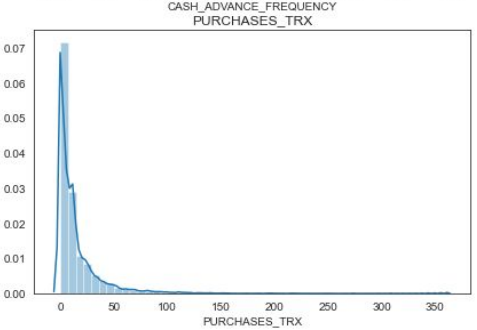












StandardScaler & Normalization

- Standardize features by removing the mean and scaling to unit variance. The standard score of a sample x is calculated as: z = (x - u) / s

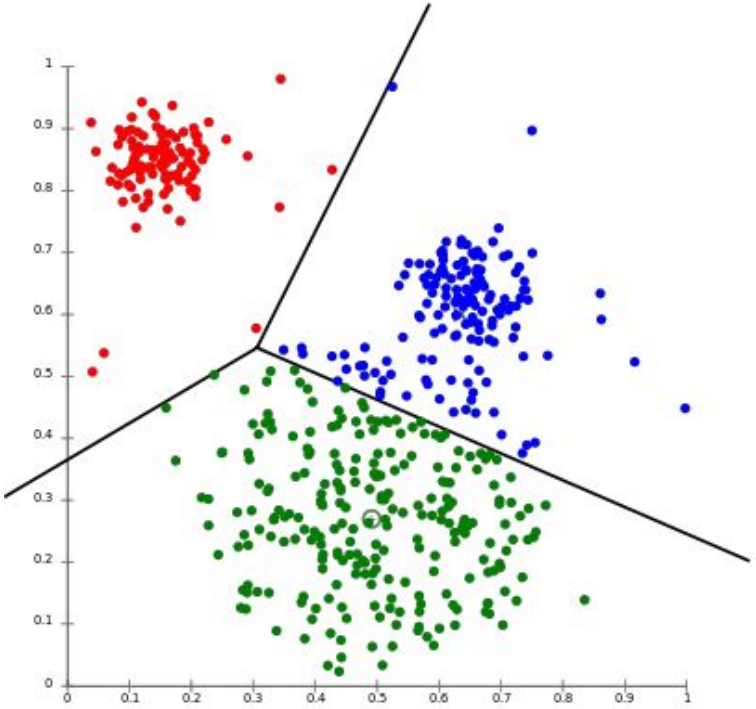
$$X' = \frac{X - \mu}{\sigma}$$

- where u is the mean of the training samples, and s is the standard deviation of the training samples.
- Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

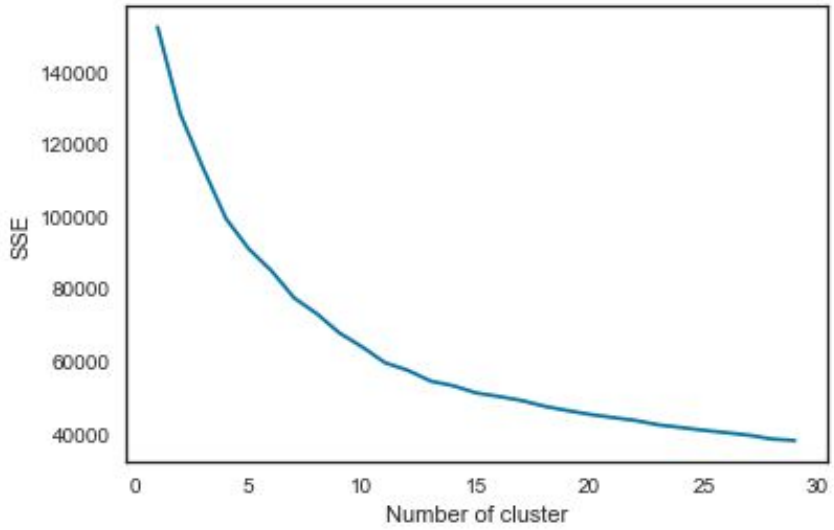
K-Means Clustering

- K-means clustering subdivides a single cluster or a collection of data points into K different clusters or groups.
- The algorithm analyzes the data to find similar data points and assigns each point to a cluster that consists of points with similar characteristics.
- K-Means clustering works by constantly trying to find a centroid with closely held data points.
- This means that each cluster will have a centroid and the data points in each cluster will be closer to its centroid compared to the other centroids.



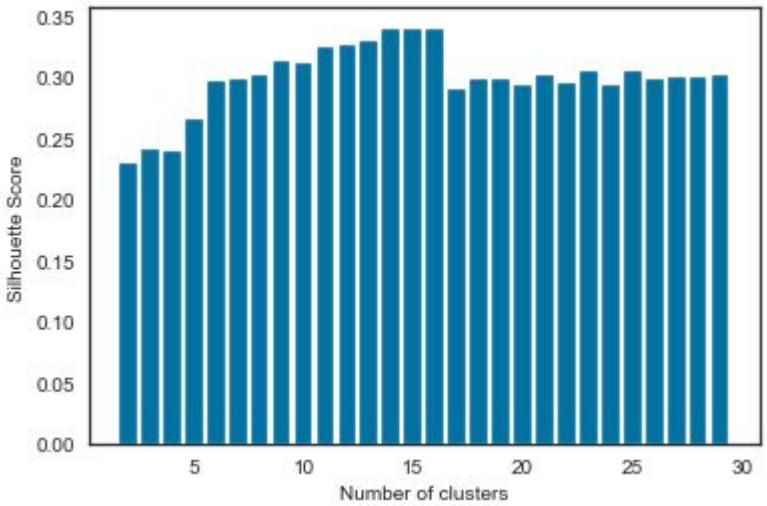
Elbow Plot

- To determine the optimal number of clusters, we have to select the value of k at the “elbow” ie the point after which the distortion/inertia start decreasing in a linear fashion. Thus for the given data, we conclude that the optimal number of clusters for the data is 3.
- If the line chart looks like an arm, then the "elbow" on the arm is the value of *k* that is the best.



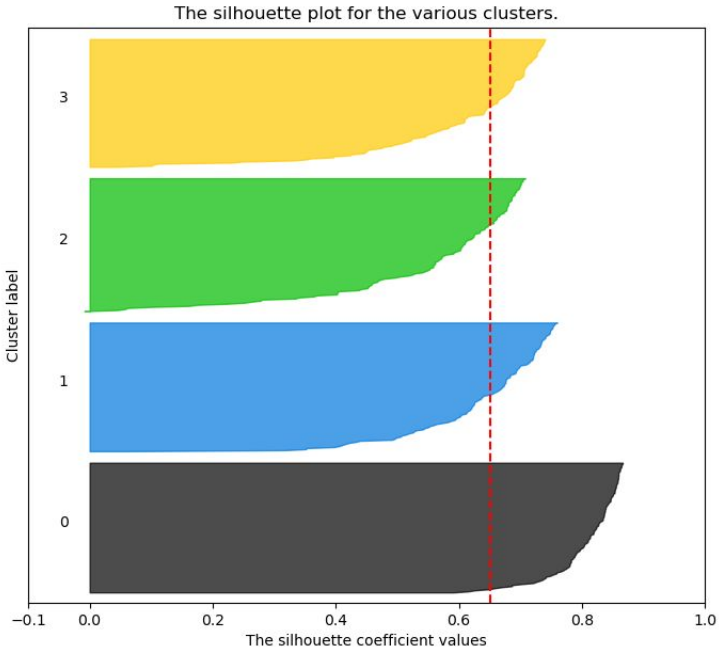
Silhouette Plot

- Silhouette score for a set of sample data points is used to measure how dense and well-separated the clusters are.
- Silhouette score takes into consideration the intra-cluster distance between the sample and other data points within the same cluster (a) and inter-cluster distance between the sample and the next nearest cluster (b).



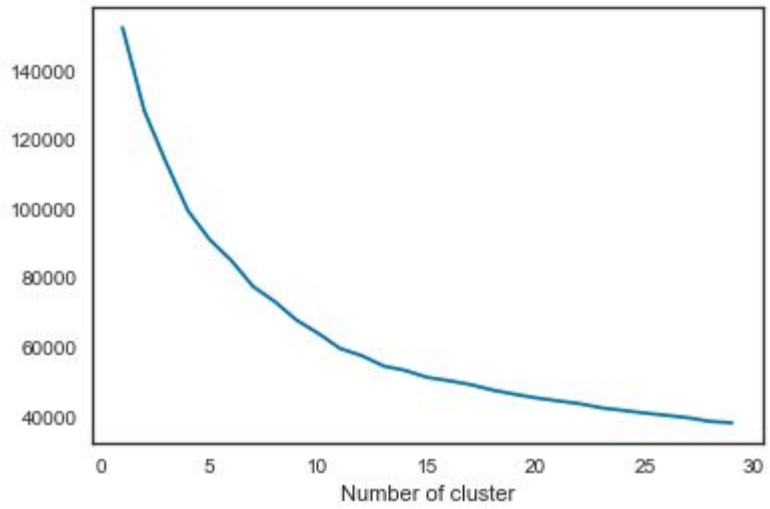
Silhouette Plot

- The silhouette score falls within the range [-1, 1].
- The silhouette score of 1 means that the clusters are very dense and nicely separated. The score of 0 means that clusters are overlapping. The score of less than 0 means that data belonging to clusters may be wrong/incorrect.



Inertia Plot

- It is defined as the mean squared distance between each instance and its closest centroid.
- As the number of clusters increases, closest will be the clusters from the centroids and lower will be the inertia.
- Elbow rule is used in order to find the optimal number of clusters.
- The curve looks like a hand and the number of clusters to be chosen over there should be equal to 8 as after that curve reaches a plateau.



Key Performance Indicator

BALANCE monthly average purchase

CASH_ADVANCE cash advance amount

ONEOFF_PURCHASES Total amount of one-off purchases

INSTALLMENTS_PURCHASES Total amount of installment purchases

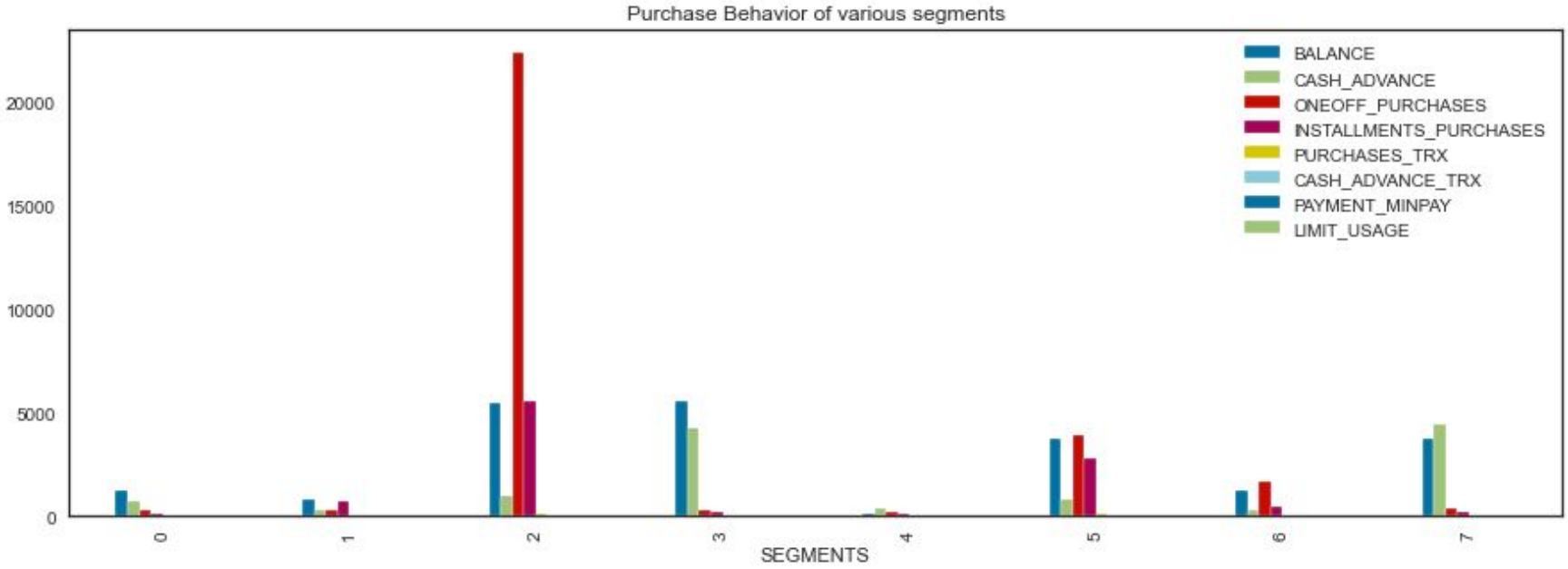
PURCHASES_TRX average amount per purchase

CASH_ADVANCE_TRX cash advance transaction

BALANCE / CREDIT_LIMIT limit usage (balance to credit limit ratio)

PAYMENTS / MINIMUM_PAYMENTS payments to minimum payments ratio

Cluster Interpretation



Cluster Interpretation

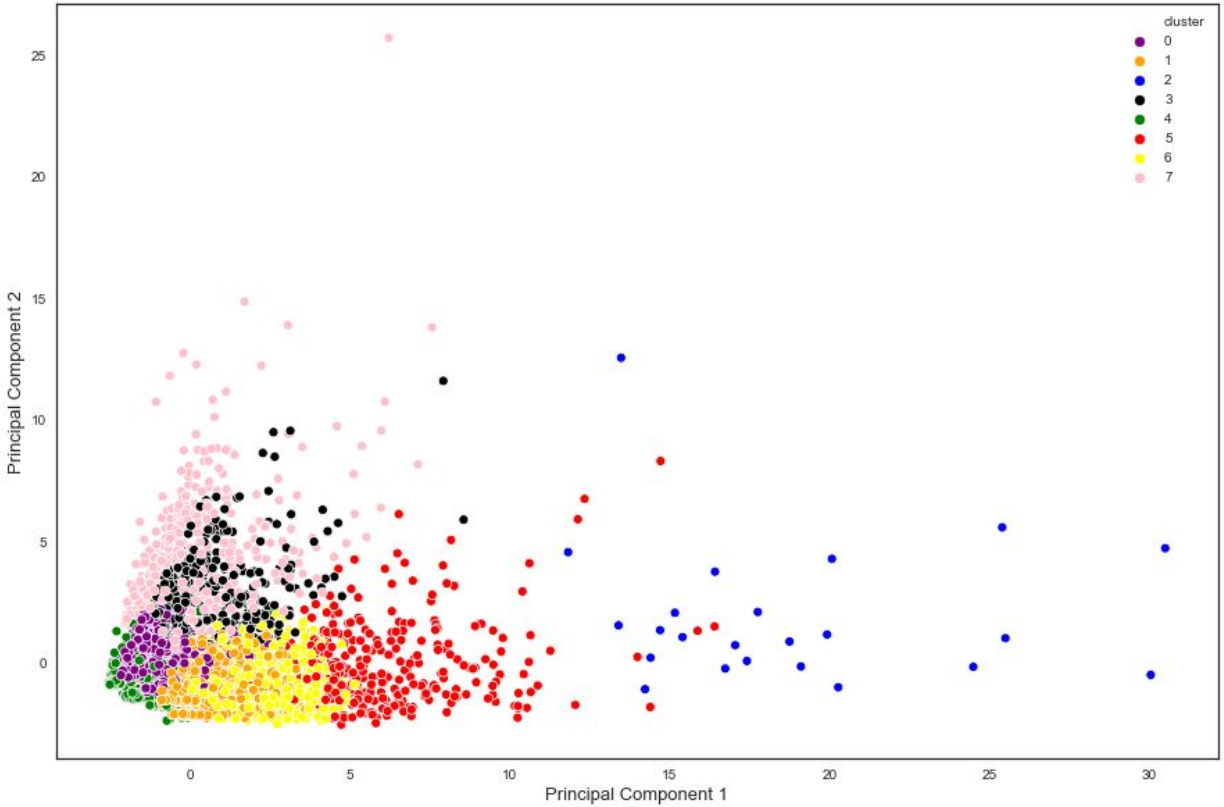
Cluster 3 - This cluster shows significantly high ONEOFF_PURCHASES and INSTALLMENTS_PURCHASES. Also has high BALANCE. Customers in this cluster have very high spending capacity. Targeted marketing campaigns could be term useful.

Cluster 5 & 6 - This cluster shows significantly high BALANCE & CASH_ADVANCE. Potential Customers with the ability to spend more.

Cluster 7 - This cluster shows average spending capability. Marketing campaigns may be fruitful.

Cluster 0, 1, 2 and 4 - This cluster shows below average spending capability. Any marketing campaign will be a futile attempt.

PCA Plot

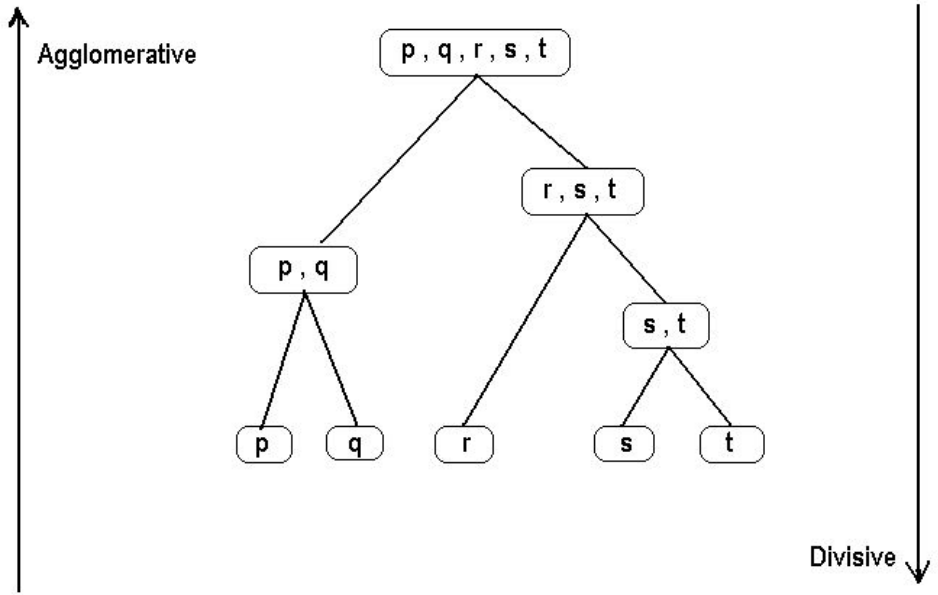

cluster
0
1
2
3
4
5
6
7
Principal Component 2
Principal Component 1
25
20
15
10
5
0
0
5
10
15
20
25
30

DBSCAN

- **Density-based spatial clustering of applications with noise**. It groups together points that are close to each other based on a distance measurement (usually Euclidean distance) and a minimum number of points. It also marks as outliers the points that are in low-density regions.
- **eps**: specifies how close points should be to each other to be considered a part of a cluster. It means that if the distance between two points is lower or equal to this value (eps), these points are considered neighbors.
- **minPoints**: the minimum number of points to form a dense region. For example, if we set the minPoints parameter as 5, then we need at least 5 points to form a dense region.

Agglomerative Clustering

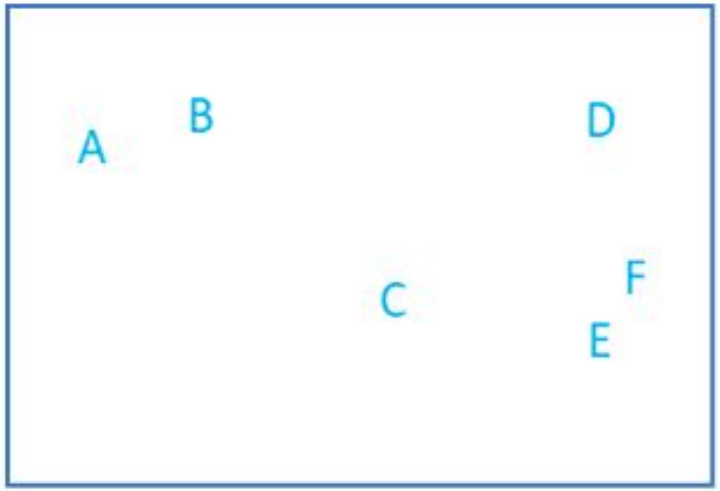
- Bottom up approach. Start with many small clusters and merge them together to create bigger clusters.
- The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects. The result is a tree-based representation of the objects, named dendrogram.

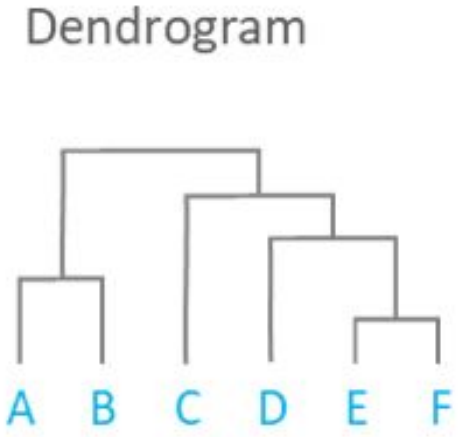


Dendrogram

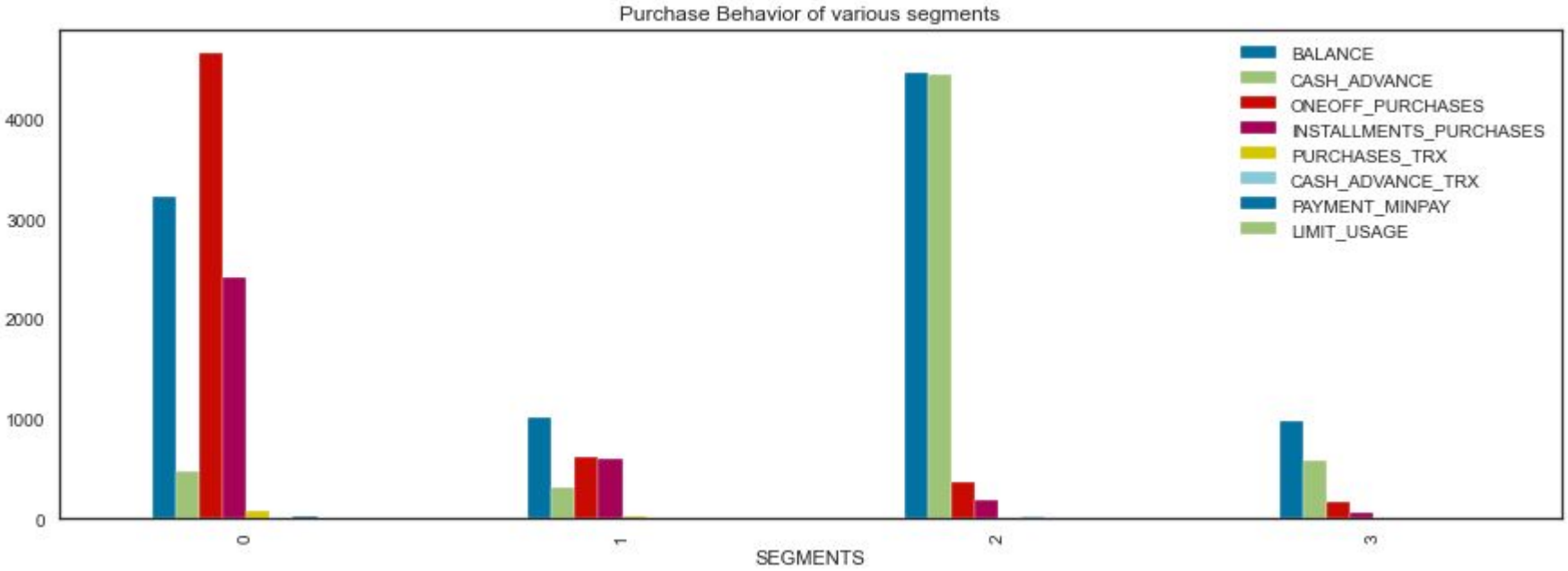
- The Hierarchical clustering Technique can be visualized using a Dendrogram.
- A Dendrogram is a tree-like diagram that records the sequences of merges or splits.
  - Determine the largest vertical distance that doesn’t intersect any of the other clusters.
  - Draw a horizontal line at both extremities.
  - The optimal number of clusters is equal to the number of vertical lines going through the horizontal line.

Dendrogram





Cluster Interpretation



Cluster Interpretation

Cluster 2 - This cluster shows significantly high BALANCE and ONEOFF_PURCHASES. Customers in this cluster have very high spending capacity. Targeted marketing campaigns could be term useful.

Cluster 0 - This cluster shows significantly high BALANCE. Customers willing to make ONEOFF_PURCHASES and INSTALLMENTS_PURCHASES.

Cluster 1 & 3 - This cluster shows average spending capability. Marketing campaigns may be fruitful.

Project objectives

Thank You