**CREDIT CARD CASE STUDY – SEGMENTATION**

DATA: CC\_GENERAL

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

Data Shape: 9000 x 18

Objective:

To develop a customer segmentation to define the marketing strategy.

Steps to be followed

1. Importing the necessary libraries
2. Importing the data
3. Perform the pandas profiling report and identify the numeric and categorical variables.
4. Prepare the Data Audit Report for the dataset.
5. Perform the basic Data Preparation steps like outliers treatment, missing value treatment. Plot the correlation matrix.
6. Standardise the data to perform the Clustering Analysis.
7. Applying the PCA (Principal Component Analysis) to get the optimal number of components for the variance 70 – 80%.We have observed the number of components to 3 to 5.
8. Export the Loadings matrix and choose the handpicked variables which are seemed to be important from business point of view.
9. Implement the model K means for different K values ie; 3, 4 and 5.
10. To find the optimum K value we need to analyze the data for different K values by exporting the data which is called as profiling output.
11. To find the optimum number of clusters/ K values we use three methods

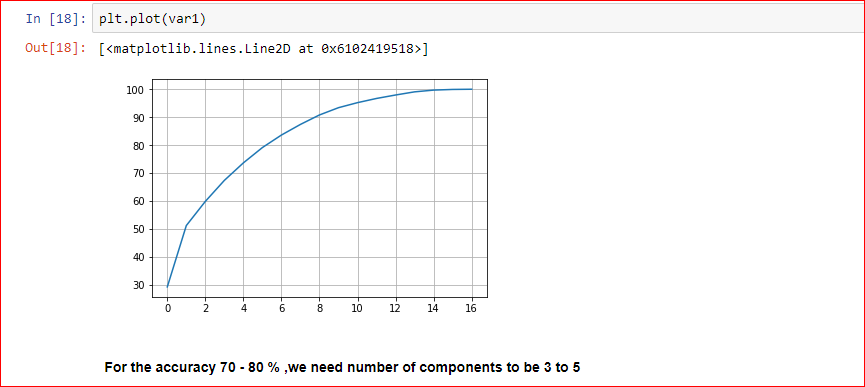
Dendograms

Elbow analysis

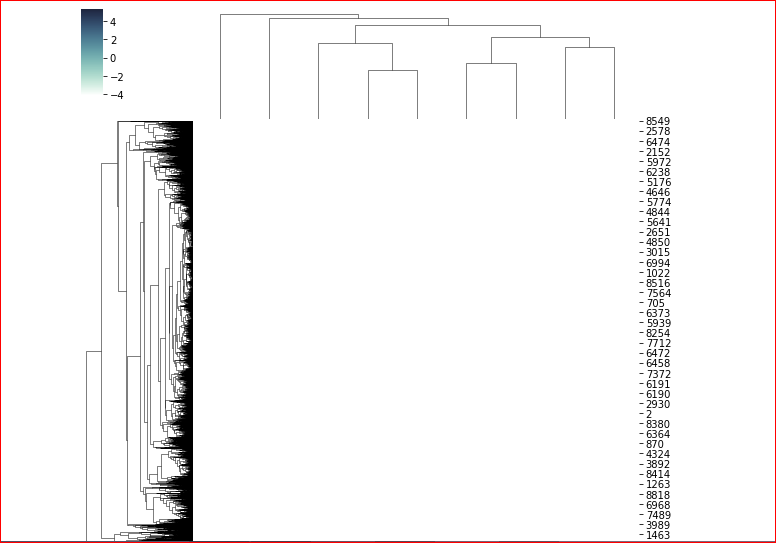
Silhouette coefficient

1. All the above three analysis show that the optimum number of cluster is 5.
2. Perform the DBSCAN Clustering on the data.

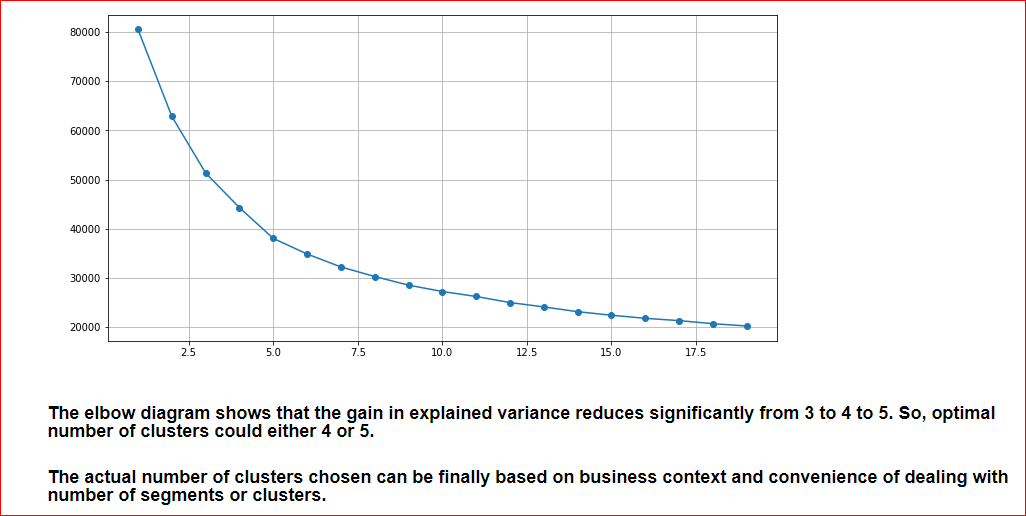
By applying PCA, we get a graph which shows the optimum number of components required with variance of 70 – 80 %.The below snip – it shows the same.



The optimum number of clusters can be found by a method called Dendograms. It shows the number of cluster to be 5.The snip – it shows the same.



The Elbow analysis shows the graph that the optimum numbers of cluster can be 5. The snip - it shows its representation.



The silhouette coefficient is also a method to find optimum number of clusters which shows 5.

