**Data Mining Assignment 5**

**Group 60**

**Problem Statement:** The banks that are receiving customer complaints filed against them will analyse the complaint data to provide results on where the most complaints are being filed, what products/ services are producing the most complaints and other useful data. These datasets fall under the complaints of Credit reporting, Mortgage, Debt Collection, Consumer Loan and Banking Accounting. So need to process that data and do proper mining techniques to extract the important information from that.

Exploratory analysis:

Exploratory analysis is an approach to analysing the data sets to summarize their main characteristics. By using this we can get to know whether the selected features are4 good enough to model and which features are required to for applying mining techniques.

Steps in EDA:

1. Description of data
2. Handling missing data
3. Handling Outliers
4. Understanding relationships and new insights through plots

Description of data:

We can get the full description using describe() function from the numpy arrays.

Handling missing data:

` There are few techniques for handling the missing data.

1. Drop Null or missing values: May cause loss of data
2. Filling the missing values: with mean, median or mode values(in case of numerical values).
3. Predict missing values with Machine learning techniques. Eg: K-Nearest Neighbour, Linear regression

Outlier:

An Outlier was something which is out of range and effecting the whole result.

We can visualize the outlier using different plots.

1. BoxPlot
2. ScatterPlot
3. Z-Score
4. IQR(Inter-Quadrantile range)

Apply clustering techniques:

There are few old and yet effective clustering techniques.

1. K-Means
2. Silhouette coefficients.

By using the above things we can divide the data into different clusters.

From the plot obtained from the cost, the good result was coming at 11 clusters.

1. As part of cohesion and separation for the part of analysis and accuracy we are using
2. DBSCAN
3. K-Mode