PROBLEM DESCRIPTION:

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.

OBJECTIVE:

Build a clustering model using Python language or other suitable tools to find how many similar complaints are there in relation to the same bank or service or product .This model will be used to assist banks in identifying the location and types of errors for resolution leading to increased customer satisfaction to drive revenue and profitability.

DATASET SOURCE:

[https://www.kaggle.com/sebastienverpile/consumercomplaintsdata](about:blank)

EXPECTED ACTIVITIES & OUTCOMES

Your activities should include - performing various activities pertaining to the data such as, preparing the dataset for analysis; investigating the relationships in the data set using statistical techniques, visualization; creating a model; evaluating the performance of the clustering model.

Demonstrate the Data Mining process with following activities:

* Problem statement
* Perform exploratory data analysis using the statistical techniques and box plot as applicable
* Preprocess the Data as the data from internet source may be noisy
* Evaluate the model performance using cohesion and separation
* Suggest ways of improving the model
* State all your assumptions clearly and provide clear explanations to explain your stand