**Mini Project**

**Domain:**

Financial services data set

**About:**

Finance is a field that is concerned with the allocation (investment) of assets and liabilities over space and time, often under conditions of risk or uncertainty. Finance can also be defined as the art of money management. Participants in the market aim to price assets based on their risk level, fundamental value, and their expected rate of return.

The dataset consists of customer information of L&T financial services. It is a finance dataset, which consists of customers demographics, loan disbursed, asset cost being purchased and the customers previous account and loan history. The dataset also consists of the state and branch id of L&T from where the loan was disbursed, the customer’s account history. It also contains the CNS score and score description provided by the Credit Bureaus of India.

**Challenges:**

It is a challenge for any financial services to target the right people for disbursing the loan. The credit team must analyse various details like CIBIL score, payment history (if available), credit history, geographical location, profession, income, age, education etc. of the customers. This will help in understanding whether the person can pay back the loan amount. Which in turn reduces its NPAs and increases the profitability.

You need to assess what data is available and perform some exploratory and descriptive analytics to identify interesting and useful patterns, trends, and insights.

**What is Expected?**

Being a data analyst, you must come up with a first step document that lists output of your exploratory analysis, any issues or problems you may see with data that need follow up, and some basic descriptive analysis that you think highlights important outcomes/findings from the data. Based on your findings, the next level of analysis will be charted out.

**Recommendations:**

* As a data analyst, what are the approaches you suggest the credit team to identify factors that decide high probability of default? Recommend based on your analysis.

.

**Data Dictionary:**

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| Loan\_Id | Unique Loan id |
| Disbursed\_Amount | Amount of Loan disbursed |
| Asset\_Cost | Cost of the Asset |
| LTV | Loan to Value of the asset |
| Branch\_Id | Branch where the loan was disbursed |
| City\_Code | Code representing City |
| State\_Code | Code representing State |
| Postal\_Code | Postal code of the area |
| Date\_of\_Birth | Date of birth of the customer |
| Employment\_Type | Employment Type of the customer (Salaried/Self Employed) |
| DisbursalDate | Date of disbursement |
| Region\_ID | Code for region of disbursement |
| MobileNo\_Avl\_Flag | If Mobile no. was shared by the customer, then flagged as 1 |
| Aadhar\_flag | If Aaddhar was shared by the customer then flagged as 1 |
| PAN\_flag | If PAN was shared by the customer then flagged as 1 |
| VoterID\_flag | If voter id was shared by the customer then flagged as 1 |
| Driving\_flag | If DL was shared by the customer then flagged as 1 |
| Passport\_flag | If passport was shared by the customer then flagged as 1 |
| PERFORM\_CNS.SCORE | Bureau Score |
| DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS | Number of delinquent accounts in the last six months |
| CREDIT.HISTORY.LENGTH | Credit history in terms of years |
| NO.OF\_INQUIRIES | Number of inquiries made by the customer |
| Loan\_Default (target variable) | Payment default in the first EMI on due date |

**Data Manipulation/Data Visualization/EDA using Python**

Here are some indicative types of analysis you can perform. Please note that this is not an exhaustive list, you may add more

* Come up with appropriate results for the following:
  + Analysis of percentage of default and check whether it is imbalanced data?
  + Replace missing values with appropriate techniques
  + Use the right treatment for outliers in the data
  + Analyse default variable with demographic related data
  + Determine and draw insights on association between default and other variables, both categorical and numerical.

Data Preparation/Analysis tasks like (wherever applicable):

* Univariate, Bi- Variate Analysis and Multi- Variate Analysis
* Missing values identification and treatment
* Outlier analysis and treatment
* Data scaling
* Data transformation
* Feature Engineering

**Statistical Analysis using Python**

* Descriptive statistics for both numerical and categorical and draw few insights from them.
* Perform relevant hypothesis testing

**Model building using ML algorithms**

**Predictive Analysis:**

* Build appropriate predictive model/s on the data.
* Compare various predictive models with appropriate regularization and/or hyper-parameter tuning.
* Evaluate the performance of the model.
* Identify the right metric to evaluate the performance of the model.
* Identify issues and concerns on the given data and suggest the best technique/s to overcome the issues.

**Submission guidelines:**

**Files to be submitted:**

* **Jupyter notebook**
* **Html file**
* **Ppt**
* **Tableau workbook(optional)**

**Start date: 12th April 2022**

**Last date for submission : 22nd April 2022**