CREDIT RISK ANALYSIS

Objective:

The objective of the project is to do a comprehensive analysis of the various factors after which there can be made some indications to the Credit company to identify the patterns of the clients facing issues in repaying the loans and taking decisions to deny the loan, decreasing the loan amount, lending at high interest rates etc., and to make sure not denying the loan to eligible or capable customers.

Approach:

1. **Exploring the dataset & documentation:** This is the first and important step before starting any analysis on the dataset. Looking at all the columns and data present in the dataset gives an overview of the dataset and some idea about the possible values of columns.
   Reading the documentation gives a detailed understanding of the columns/ variables purpose and the usage, their data types, the possible values and description about the columns. This helps in how and what kind of analysis can be done on columns to obtain various insights based on business objective.

2. **Setting up Environment & Identifying, importing necessary libraries:** Downloaded the required CSV file from the given link and imported required libraries to start exploring the dataset in Jupyter notebook by reading CSV file.

3. **Exploring the dataset:** Understand the size, dimensions, attributes of the dataset and information regarding the datatypes of the columns in the dataset.

4. **Handling of missing values:** To ensure the correctness of the data and handling missing values. Delete unnecessary columns from the dataset which may not be useful in the required analysis. If a column contains more than 50 % of the missing values then remove those columns depending on the importance of the same in the analysis.

5. **Classification of categorical and numerical columns:** Identify the datatypes of the columns, analysing what are the categorical columns in the dataset and what are the continuous columns. Change the datatypes of the columns from continuos to categorical where ever required which helps to do the proper analysis.

6. **Univariate analysis:** Conduct univariate analysis on the continuous columns to understand the distribution and characteristics of the individual columns.

7. **Bivariate analysis:** Conduct bivariate analysis on different variables to understand the relationships between variables and impact on the business objective i.e., to understand various patterns which contribute to loan defaults by the customers.

8. **Conclusion:** Analyse the columns individually to identify their significance and patterns by the means of Univariate analysis and conduct the bivariate analysis to analyse and understand the relationships between key parameters of the dataset and conclude the findings for making profitable business decisions.