**Name: -** Vatsal Rameshbhai Gohel

**Mail id: -** [**gohel.vatsal11@gmail.com**](mailto:gohel.vatsal11@gmail.com)

**Contact Detail: -** +1-774-253-3960

**LinkedIn Id: -** <https://www.linkedin.com/in/gohelvatsal/>

**GitHub Id: -** <https://github.com/vatsalgohel>

**Project Portfolio**

**Credit Risk Analysis Using Exploratory Data Analysis (EDA)**

**Project Title:**

Credit Risk Analysis for Loan Approval Decision Making

**Problem Statement:**

When a company receives a loan application, it must decide whether to approve the loan based on the applicant’s profile. The decision involves two types of risks:

1. Not approving a loan for an applicant likely to repay, resulting in lost business.
2. Approving a loan for an applicant likely to default, leading to financial loss.

**Project Description:**

**Objective:**

To analyse and understand the factors affecting loan approval decisions to minimize business losses and avoid financial loss due to loan defaults.

**Scope:**

* Analyse applicant profiles to identify significant factors influencing loan repayment capabilities.
* Handle data cleaning, missing values, and outliers in the provided data sets.
* Perform univariate and bivariate analysis to identify trends and patterns in the data.
* Correlate significant variables to understand their impact on loan approval decisions.

**Deliverables:**

* A cleaned and well-prepared data set for analysis.
* Insights and trends based on univariate and bivariate analysis.
* A correlation matrix highlighting significant relationships between variables.
* A comprehensive report detailing findings, recommendations, and actionable insights.

**Methodology:**

**Tools and Technologies Used:**

* Jupyter Notebook
* Python Libraries: pandas, NumPy, matplotlib, seaborn, sklearn

**Process:**

* **Loading Data into Jupyter Notebook:**
  + Load application\_data.csv and previous\_application.csv datasets.
* **Data Inspection and Cleaning:**
  + Understand the data structure using .info(), .shape(), and .describe().
  + Handle missing values by dropping or imputing them.
  + Correct data types and handle outliers.
* **Data Preparation:**
  + Remove unnecessary columns.
  + Group continuous variables for better analysis.
* **Exploratory Data Analysis (EDA):**
  + Perform univariate and bivariate analysis.
  + Identify correlations between significant variables.
  + Merge previous application data with current application data.
* **Conclusion and Recommendations:**
  + Summarize findings and provide actionable recommendations.

**Role and Responsibilities:**

* **Role:** Individual Project
* **Data Collection:** Gather the necessary data sets.
* **Data Cleaning:** Handle missing values, outliers, and data type corrections.
* **Data Analysis:** Perform univariate and bivariate analysis.
* **Correlation Analysis:** Identify significant correlations.
* **Reporting:** Document findings and provide recommendations.

**Results and Impact:**

**Outcome:**

* Identification of key factors affecting loan repayment and default rates.
* Insights into applicant profiles more likely to repay or default on loans.
* Recommendations for loan approval processes to minimize financial risks.

**Impact:**

* Improved decision-making process for loan approvals.
* Reduced financial loss due to loan defaults.
* Enhanced understanding of applicant profiles and risk factors.

**Challenges and Solution:**

**Challenges Faced:**

* Handling a large amount of missing and inconsistent data.
* Identifying and dealing with outliers in numerical data.
* Managing the imbalance in the target variable.

**Solutions Implemented:**

* Dropped columns with more than 40% missing values and imputed other missing values.
* Used Box Plot and Whisker calculations to identify and handle outliers.
* Divided the data set into sub-data sets for more focused analysis of the imbalanced target variable.

**Conclusion and Learnings:**

**Summary:**

The project successfully analysed the factors influencing loan repayment and default rates. Through detailed exploratory data analysis, significant insights were gained into applicant profiles and their likelihood of repaying loans. The recommendations provided can help the company make informed decisions and reduce financial risks.

**Learning Experience:**

* Gained experience in handling large data sets and performing data cleaning.
* Learned advanced techniques for outlier detection and handling.
* Improved skills in univariate and bivariate analysis.
* Enhanced understanding of correlation analysis and its impact on decision-making processes.

**Link for the file:**

<https://drive.google.com/drive/folders/1wM8xXYh7VzlGb3wAwGKjjhow6WQiXApu?usp=sharing>