PROJECT REPORT

Title: Credit Data Discovery & Analysis

Created by:

- Shubham Gaur
- Harinandan Praveen
- Attah Hussain Muhammed
- Rabeya Khatun

Introduction:

This report presents the progress of the Credit Exploratory Data Analysis (EDA) project. The primary objective of this project is to analyze credit data to identify patterns, trends, and anomalies that can inform credit risk management strategies. This EDA is a crucial step in developing more robust credit scoring models and improving decision-making processes in credit allocation. The project involves a comprehensive analysis of a large dataset containing various credit related attributes. Key areas of focus include understanding the distribution of credit scores, analyzing the impact of different variables on credit risk, and identifying potential indicators of default risk.

Objective and Focus:

The project aims to provide practical experience in EDA with a focus on understanding risk analytics in banking and financial services, revolving around loan providing companies. The main objective is to analyze credit data to identify reliable loan applicants and minimize the risk of lending to potential defaulters, thereby enhancing credit risk management strategies and decision-making in credit allocation.

Problem Statement and Dataset Analysis:

- Loan companies face challenges in assessing the risk of lending to individuals with poor or no credit history. The project involves a comprehensive analysis of a large dataset containing various credit-related attributes, including loan applications and applicants' credit history. It focuses on understanding the distribution of credit scores, analyzing the impact of different variables on credit risk, and identifying potential indicators of default risk.
- The dataset includes information like credit scores, loan amounts, repayment histories, and demographic information of borrowers. It is composed of three files: 'application_data.csv', 'previous_application.csv' .

Data Collection, Preprocessing, and Analysis:

- The data collection phase involved aggregating credit information from multiple sources. Rigorous data cleaning and preprocessing were conducted to ensure data quality, involving handling missing values, correcting inconsistencies, and normalizing formats. This also included categorizing continuous variables and encoding categorical variables.
- The exploratory data analysis phase employed statistical and visualization techniques to understand key variables, identify correlations, and detect outliers. Key findings include distinct patterns in borrowing behavior across different credit score ranges, significant relationships between loan amount, repayment history, and credit scores, and identification of unusual cases indicating potential fraud or data entry errors.

Methodology Overview:

The methodology for the Credit Exploratory Data Analysis project encompasses several steps, each critical for deriving meaningful insights from the credit dataset.

- **Data Collection:** The initial phase involved gathering a comprehensive dataset from a reliable financial database. This dataset includes variables

such as credit scores, loan amounts, interest rates, repayment terms, borrower demographics, and credit history.

- **Data Cleaning:** The next step was to clean the data, which involved handling missing values, removing duplicates, and correcting inconsistencies. This process ensures the accuracy and reliability of the analysis.
- **Data Exploration:** Using statistical and visualization tools, the data was explored to understand distributions, relationships between variables, and potential outliers. Techniques like histograms, scatter plots, and correlation matrices were employed.
- **Statistical Analysis:** Various statistical methods were used to analyze the data. This included descriptive statistics to summarize the data and inferential statistics to draw conclusions about the population based on the sample data.
- **Data Visualization:** To effectively communicate the findings, data visualization techniques were extensively used. Graphs, charts, and heat maps were created to illustrate the relationships and patterns within the data.

This methodology provided a structured approach to analyze the credit data comprehensively, ensuring that the findings are robust and insightful.

Significance:

This analysis not only aids in immediate credit risk management but also paves the way for developing more sophisticated models for future financial stability.

Expected Results and Insights:

The project's goals included identifying and addressing data outliers and imbalances, conducting various analyses to understand factors influencing loan default, and making informed lending decisions. The assignment aimed to enhance understanding of EDA in a real-world financial context, with a focus on identifying patterns that can inform better lending decisions. The emphasis was on comprehensive data analysis and clear presentation of findings.

These insights are instrumental in understanding the factors that influence credit risk and can guide the development of more sophisticated credit risk models.