Project Report: Bank Loan Case Study

**Project Description**

The project aims to analyze a dataset related to loan applications in a financial company specializing in lending loans. The key objectives include handling missing data, identifying outliers, analyzing data imbalance, performing univariate, segmented univariate, and bivariate analysis, and identifying top correlations for different scenarios. The primary focus is on understanding the factors influencing loan default and gaining insights for informed decision-making.

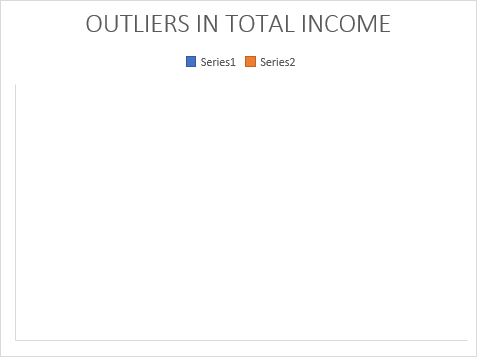
**Approach**

1. **Handling Missing Data**:

* Utilized Excel functions like COUNTBLANK, and IF to identify missing data.
* Eliminated columns with blank percentage more than 30. There were a total of 124 columns after performing EDA the total number of columns left were

1. **Identifying Outliers:**

* Utilized Excel statistical functions, including QUARTILE and IQR, to detect potential outliers.
* Applied conditional formatting and business rules to identify and assess the validity of outliers.
* Visualized outliers using box plots and scatter plots for a clear understanding.



1. **Analyzing Data Imbalance:**

* Used Excel functions like COUNTIF and SUM to calculate the proportions of each class.
* Assessed data imbalance by comparing class frequencies.
* Visualized data distribution using pie charts and bar charts.

1. **Performing Analysis:**

* Conducted univariate analysis using Excel functions for descriptive statistics.
* Employed segmented univariate analysis to compare variable distributions for different scenarios.
* Utilized pivot tables, filters, sorting, and various Excel features for bivariate analysis.

1. **Identifying Top Correlations:**

* Segmented the dataset based on scenarios (e.g., clients with payment difficulties).
* Calculated correlation coefficients using Excel functions like CORREL.
* Visualized correlations using correlation matrices and heatmaps.

**Tech-Stack Used**

Microsoft Excel 2019: Utilized Excel for data analysis, handling missing data, statistical calculations, and visualization.

**Insights**

1. **Missing Data Handling:**

* Identified missing data in various columns and implemented appropriate strategies for imputation.

1. **Outliers and Data Imbalance:**

* Detected outliers in numerical variables, allowing for further investigation or validation.
* Evaluated data imbalance to understand the distribution of the target variable.

1. **Analysis and Correlations:**

* Conducted comprehensive analyses, including univariate, segmented univariate, and bivariate analyses.
* Discovered key insights into variable distributions, relationships, and correlations.

**Result**

* The project has provided a thorough understanding of the loan application dataset.
* Insights gained from handling missing data, identifying outliers, and analyzing data distribution contribute to informed decision-making in loan approval processes.
* The correlation analysis enhances the understanding of factors influencing loan default scenarios.
* This project equips the financial company with valuable insights to improve risk assessments, streamline the loan approval process, and make data-driven decisions for effective credit management.