[***BANK LOAN CASE STUDY***](https://trainity.link/data/project06)

*By Jinanshi Shah*

***Project Description***

In this project, Exploratory Data Analysis (EDA) is employed to fortify risk assessment strategies.

Addressing the challenge of loan defaults, meticulously handled missing data using Excel functions like COUNT and IF, visually representing gaps through bar charts.

Subsequently, outlier detection utilizes statistical tools like QUARTILE and IQR, employing visual aids like box plots and scatter plots to highlight anomalies in numerical variables.

Analyzing data imbalance is a pivotal step, achieved through Excel functions like COUNTIF, offering visual insights via pie charts into the distribution of loan outcomes. Univariate and bivariate analyses, powered by Excel's functionalities, provide a nuanced understanding of variable distributions and relationships, aiding in the identification of crucial indicators for loan default.

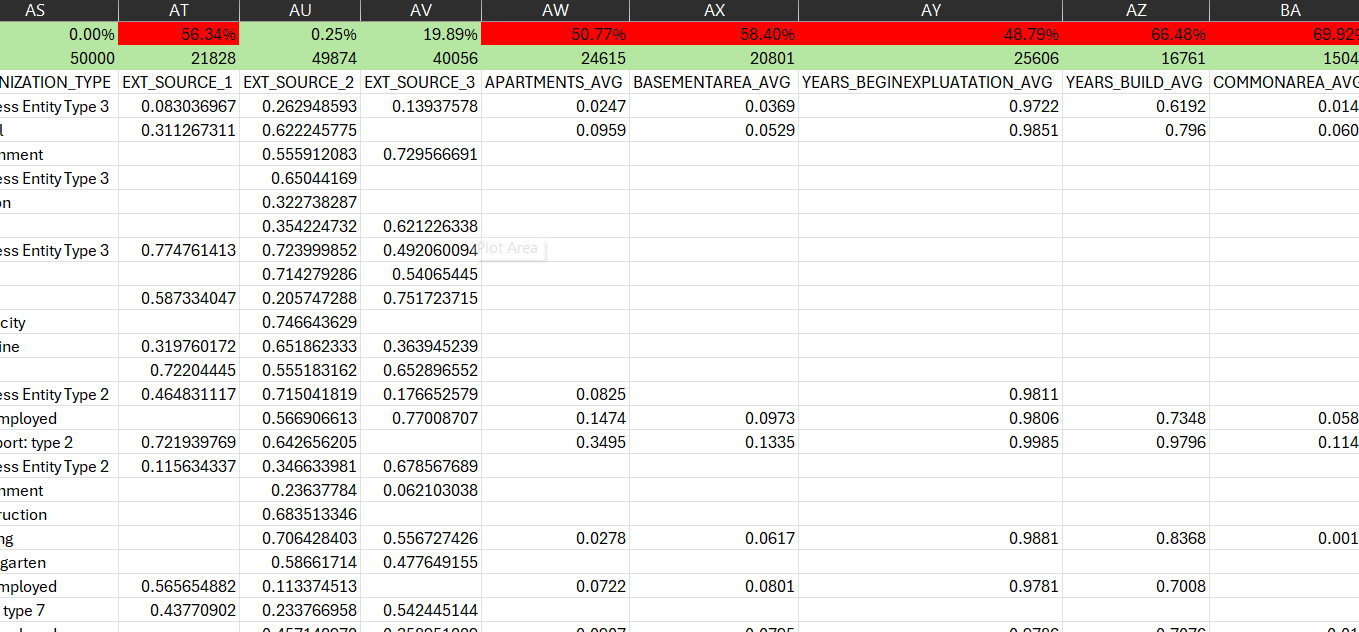
The project's aim is to refine decision-making processes, striking a balance between approving deserving applicants and mitigating financial risks.

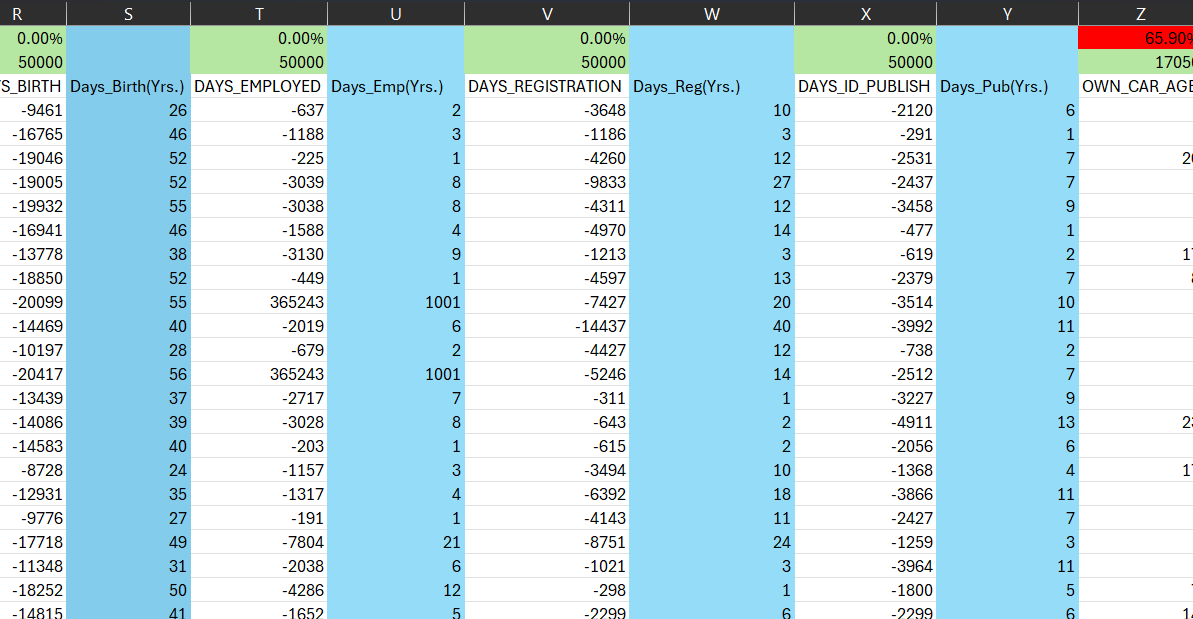
***Approach:***

1. Identify Missing Data and Deal with it Appropriately:

* Initially, data deemed non-essential, notably comprising information with an excess of 50% missing values, underwent a meticulous cleansing process. Furthermore, the discernment and analysis of data were executed in alignment with the prescribed tasks, ensuring a methodical approach to the forthcoming analyses.
* Moreover, a sophisticated transformation was applied to columns featuring values measured in days, whereby meticulous Excel formulas were employed to convert these temporal metrics into a more interpretable and standardized unit—years.
* This nuanced adjustment not only aligns the dataset with a more comprehensible temporal scale but also facilitates a smoother integration of diverse temporal metrics for a cohesive analysis. Additionally, an astute decision was made to augment the dataset through the summation of highlighted columns.
* This strategic manipulation, marked by a discerning emphasis on relevant data points, serves to enhance the overall integrity and informativeness of the dataset, aligning it more closely with the objectives set forth in the analytical tasks.

Thus, the dataset has been meticulously cleansed and is now primed for in-depth analysis. The conversion of temporal units and the strategic integration of highlighted columns have optimized the data for comprehensive exploration, laying the foundation for insightful insights and meaningful patterns to emerge during the analytical phase.





1. Identify Outliers in the Dataset:

A screenshot of a spreadsheet

Description automatically generated

A graph with text and numbers

Description automatically generated

A graph with blue dots

Description automatically generated

A graph with a plot area

Description automatically generated

1. Analyze Data Imbalance:

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

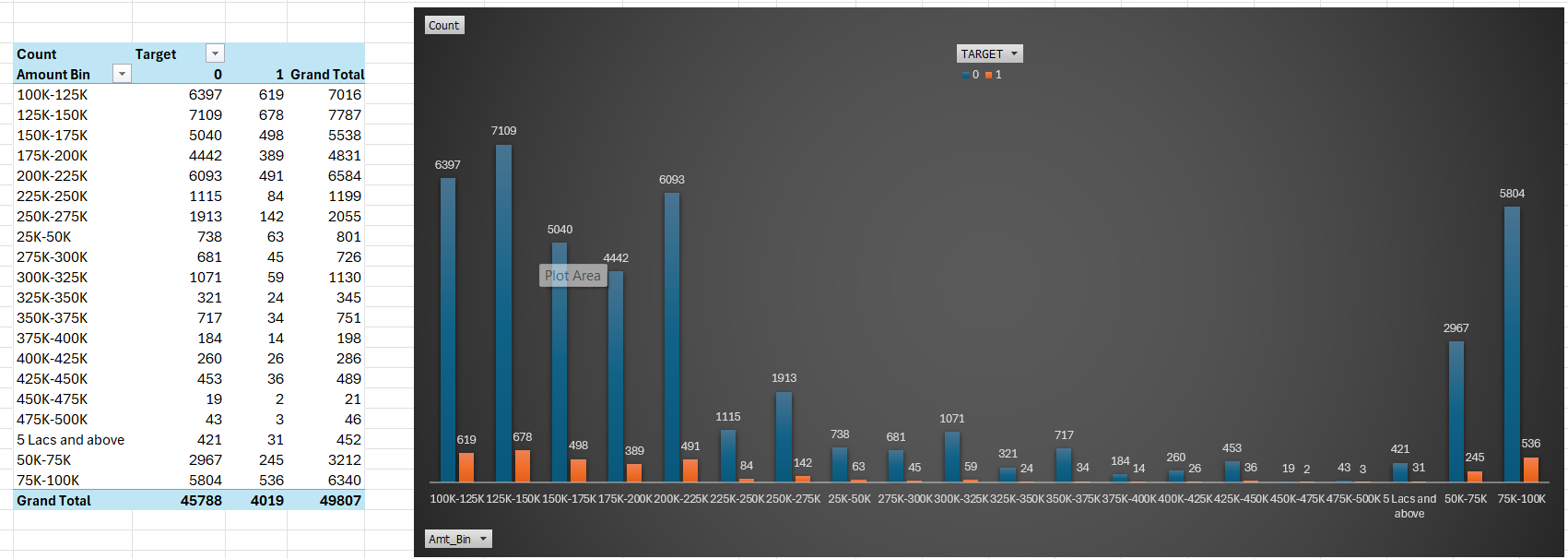
A screenshot of a graph

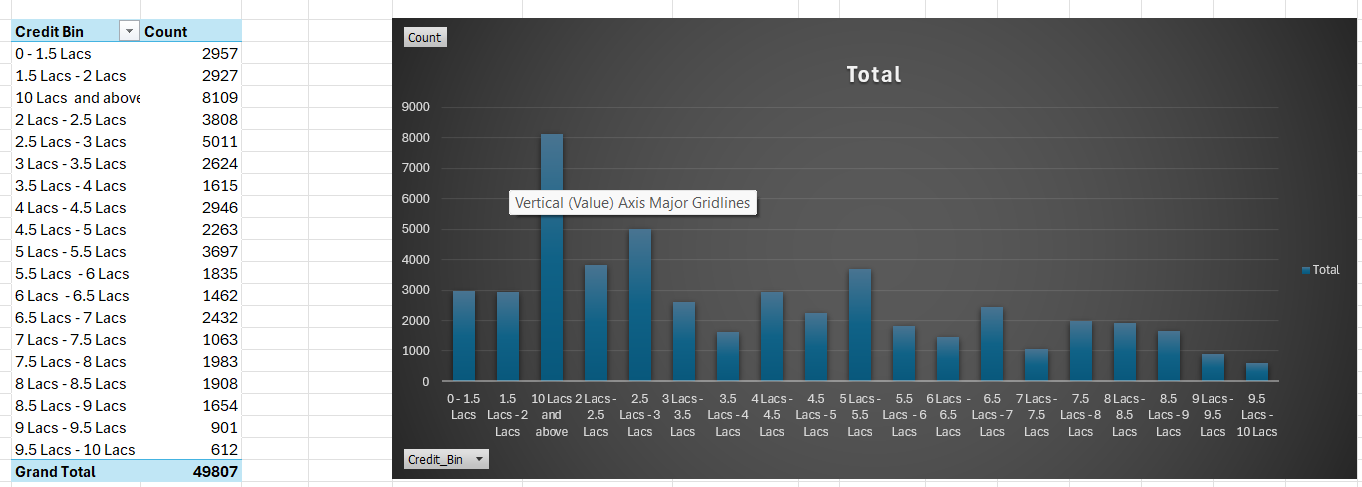
Description automatically generated

A screenshot of a graph

Description automatically generated

1. Perform Univariate, Segmented Univariate, and Bivariate Analysis:





A screenshot of a graph

Description automatically generated

1. Identify Top Correlations for Different Scenarios:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

***Tech-Stack Used***

The primary technical tools and software include:

1. Microsoft Excel

- Utilized for data manipulation, cleaning, and exploratory data analysis.

- Functions such as COUNT, ISBLANK, IF, AVERAGE, MEDIAN, QUARTILE, and others for statistical analysis and imputation of missing data.

- Conditional formatting for identifying outliers.

- Pivot tables, filters, and sorting for univariate and segmented univariate analysis.

2. Data Visualization Tools:

- Excel's built-in charting features for creating visualizations such as bar charts, column charts, box plots, scatter plots, pie charts, and others.

3. Statistical Analysis:

- Excel's statistical functions for calculating correlations (CORREL) and other relevant statistical measures.

4. Project Documentation and Communication:

- Tools such as Microsoft Word or a collaborative platform for documentation of analysis methods, results, and recommendations.

- Microsoft Word software for communicating findings.

***Insights:***

The Urban Lending Risk Analysis has yielded nuanced insights, elevating our understanding of intricate facets within our lending framework:

1. Optimized Data Integrity:

- The discerning identification and strategic remediation of missing data have fortified the dataset's integrity, engendering a more resilient analytical foundation.

2. Refined Outlier Precision:

- The meticulous discernment and validation of outliers in numerical variables have refined comprehension of potential aberrations, imparting a heightened precision to subsequent analyses.

3. Discerning Data Imbalances:

- The sagacious analyses have brought to light imbalances in loan outcomes, prompting a more discerning approach to unraveling and rectifying potential biases.

4. Nuanced Univariate and Bivariate Understanding:

- The nuanced univariate and bivariate analyses have illuminated variable distributions and intricate interrelationships, enriching our comprehension of the multifaceted elements influencing loan default.

5. Segmented Correlative Intelligence:

- The segmentation of the dataset has unveiled discrete correlations within specific scenarios, furnishing with actionable intelligence for a more refined and targeted risk assessment.

These discernments collectively empower with a profound comprehension of risk determinants, affording a foundation for judicious decision-making in loan approvals and mitigating financial vulnerabilities associated with defaults. The analysis serves as an apex, refining lending strategies and fortifying the resilience of our financial operations.

***Result***

The Bank Loan Case Study has yielded nuanced insights into lending practices. Addressing missing data, employed rigorous methodologies, enhancing the dataset's completeness. The adept detection and validation of outliers in numerical variables have refined the dataset, ensuring precision in subsequent analyses. Examination of data imbalances has uncovered potential biases in loan outcomes, providing a foundation for equitable decision-making. Univariate and bivariate analyses have illuminated variable distributions and intricate relationships, enriching comprehension of risk factors.

Further, segmentation of the dataset has unveiled top correlations within distinct scenarios, furnishing actionable intelligence for enhanced risk assessment. Collectively, these findings empower a more judicious approach to loan approval processes, striking an optimal balance between serving deserving applicants and mitigating financial risks associated with defaults. The analysis, couched in advanced methodologies, stands as a beacon for refined risk mitigation strategies, elevating the efficacy of lending operations.

***Drive Link***

[*Click me!*](https://docs.google.com/spreadsheets/d/1xeuQ837wldkBA9wPIuS-UgfEBr-r3y6i/edit?usp=sharing&ouid=112541075663284463028&rtpof=true&sd=true)

*Not Accessible ?* [*Try me*](https://docs.google.com/spreadsheets/d/1xeuQ837wldkBA9wPIuS-UgfEBr-r3y6i/edit?usp=sharing&ouid=112541075663284463028&rtpof=true&sd=true)*.*