Lending Club Case Study

A data driven analysis on loan approval process and decisioning.

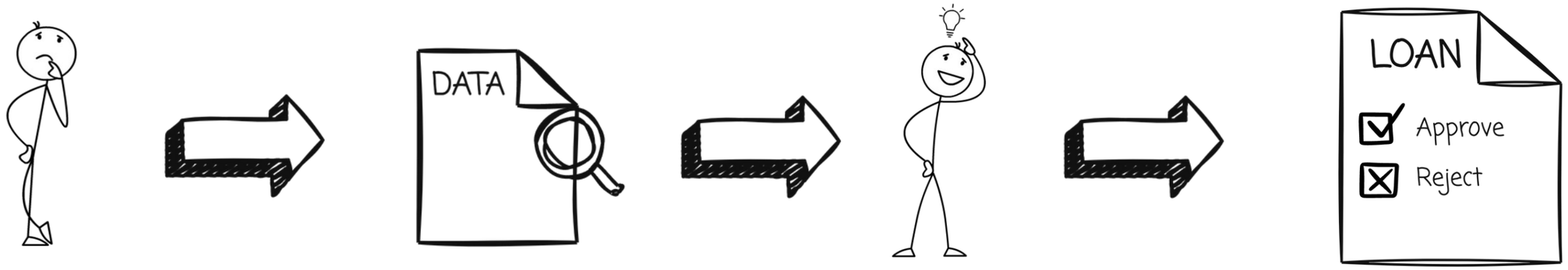
By
Nikhil Chakravarthy
Prachi Sharma

Problem Statement & Challenges

This is a case study of a leading consumer finance company that specializes in lending various types of loans to urban customers. The purpose of this case study is for us to perform Exploratory Data Analysis (EDA) on the given dataset and identify patterns that indicate if a person is likely/unlikely to default, which can be used for taking suitable actions by the company.

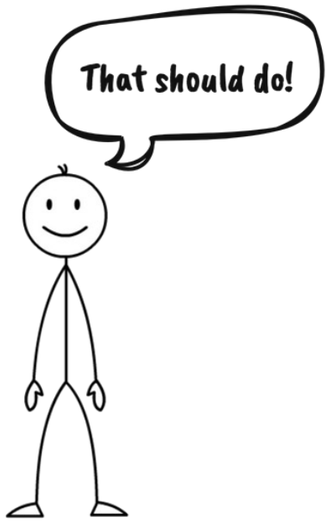
Based on the available data, there are two types of risks/challenges that might affect the company's decision on loan approval process which are as follows:

1. If the applicant is **likely to repay the loan**, then not approving the loan results in a loss of business to the company.
2. If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company.



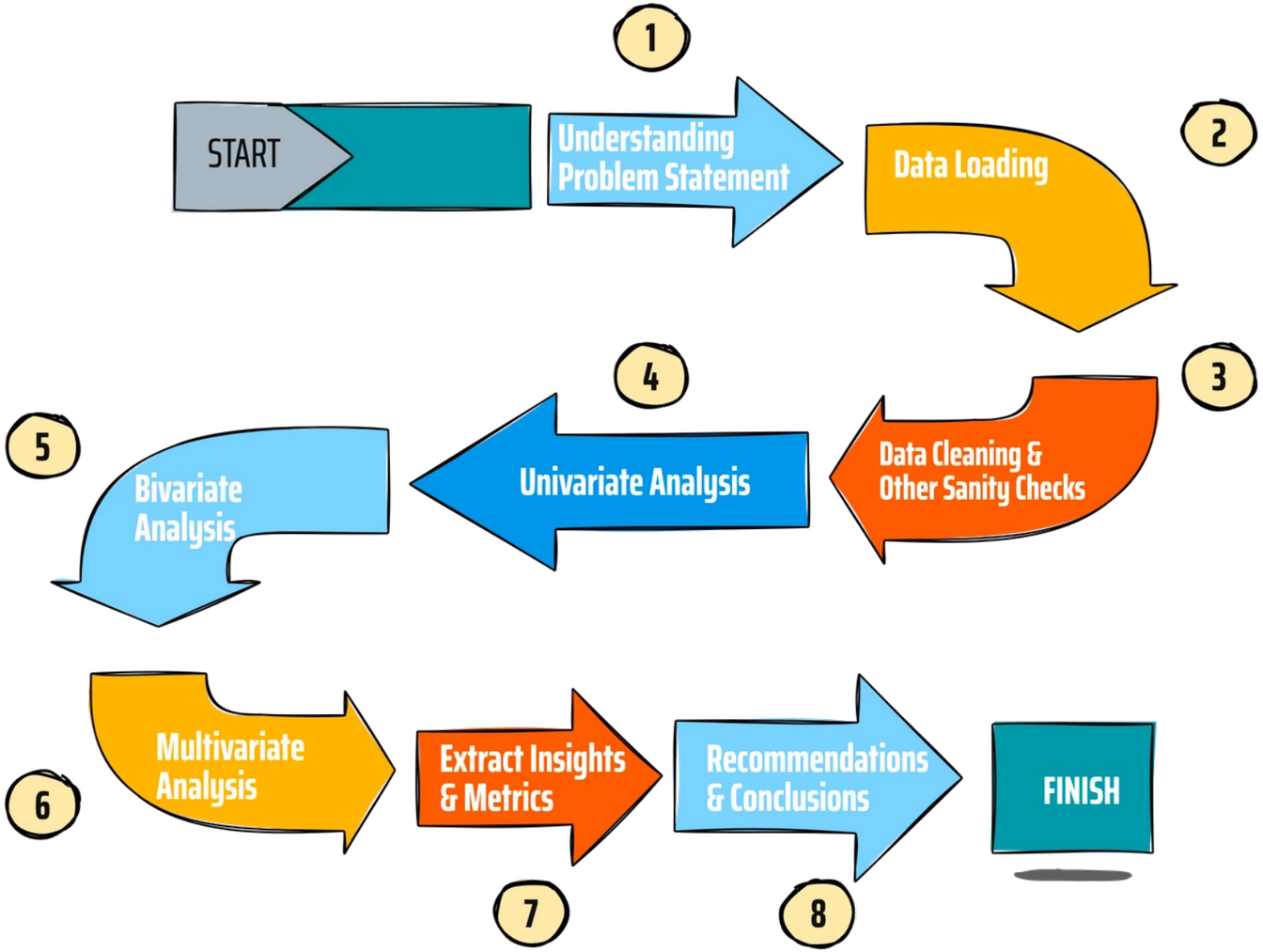
Assumptions

1. We can’t extract any relevant insights from current/active borrowers’ data as they are already in the process of paying the loan back. Hence, the data pertaining to current/active borrowers can be dropped from the original dataset.
2. The columns which have missing value percentage greater than 50% were dropped from the original dataset as they will be of no use in EDA.
3. For columns, whose missing value percentage under the acceptable range, they were imputed for accurate analysis. For numerical columns, median was used for imputation. Categorical columns were imputed with “NA”.
4. Few columns like “tax_liens”, “chargeoff_within_12_mths” contain either 0 or NA indicating no relevance. Hence, they were dropped from the original dataset before EDA.
5. While segmentation, it was observed that few numerical columns contained very few unique values (~10). So, they were considered “categorical columns”.
6. Even though few outliers were observed, they were not dropped from the original dataset so that they can be represented in boxplots.



Analysis & Approach

To uncover hidden insights of the borrowers’ data set, Exploratory Data Analysis (EDA) was performed in a stepbystep process as shown below.



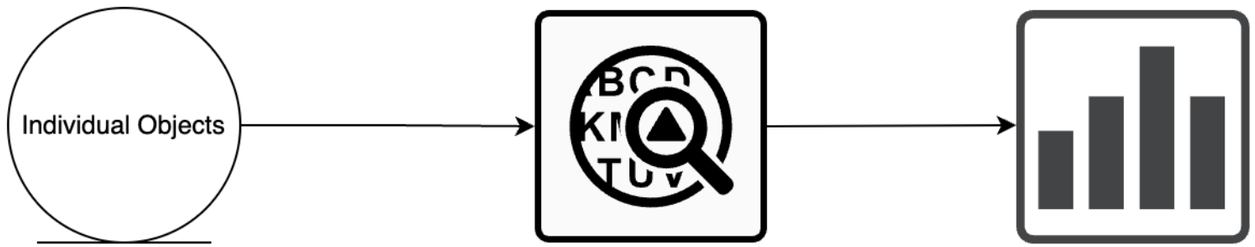
Univariate Analysis

1. Performed extensive univariate analysis on all the available numerical and categorical columns.
2. The graphs were plotted as follows and type of columns.

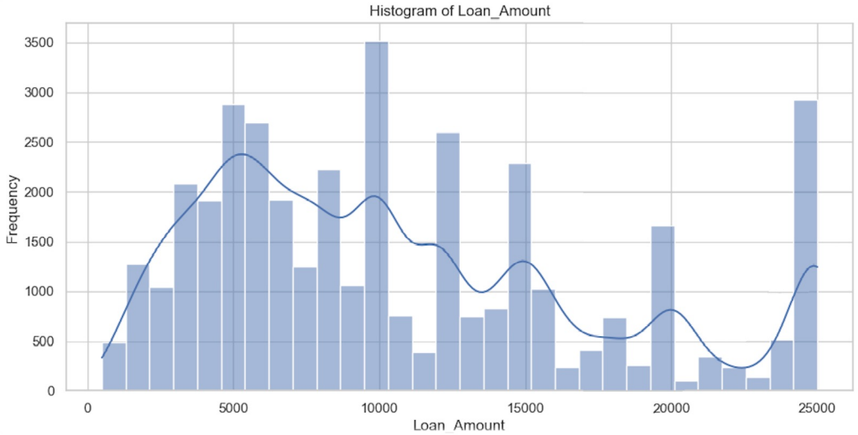
    - Histogram plots for numerical columns to obtain metrics based on frequency.
    - Box plots for categorical columns to highlight the outliers.
    - Count plots for both numerical & categorical columns.

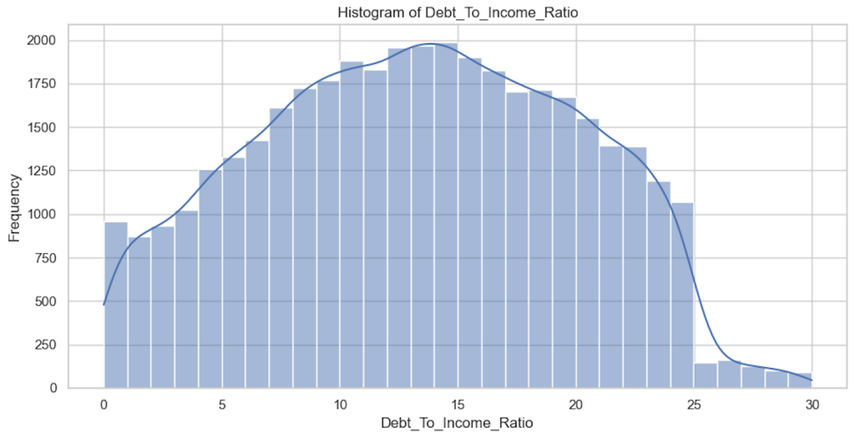
The sample results of the univariate analysis can be seen in the next slide.

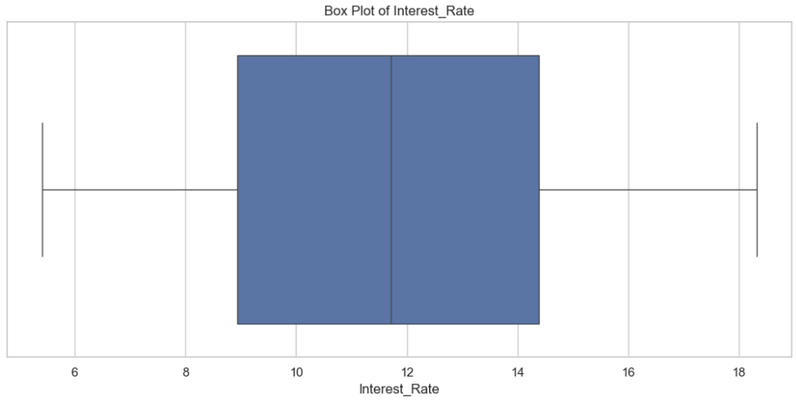
NOTE: Only few results were captured to accommodate the further analyses.

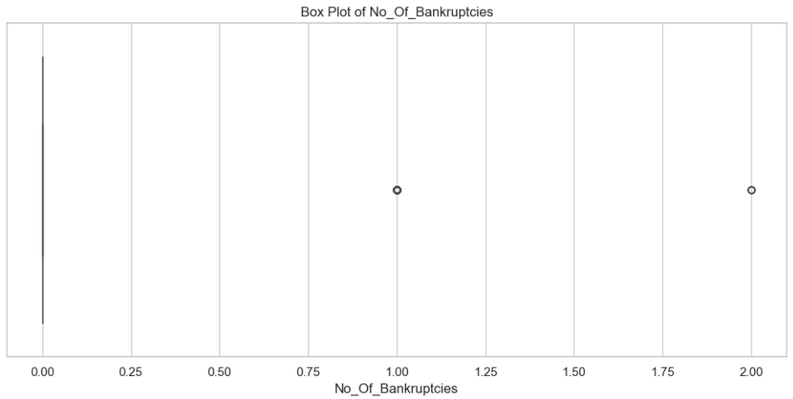


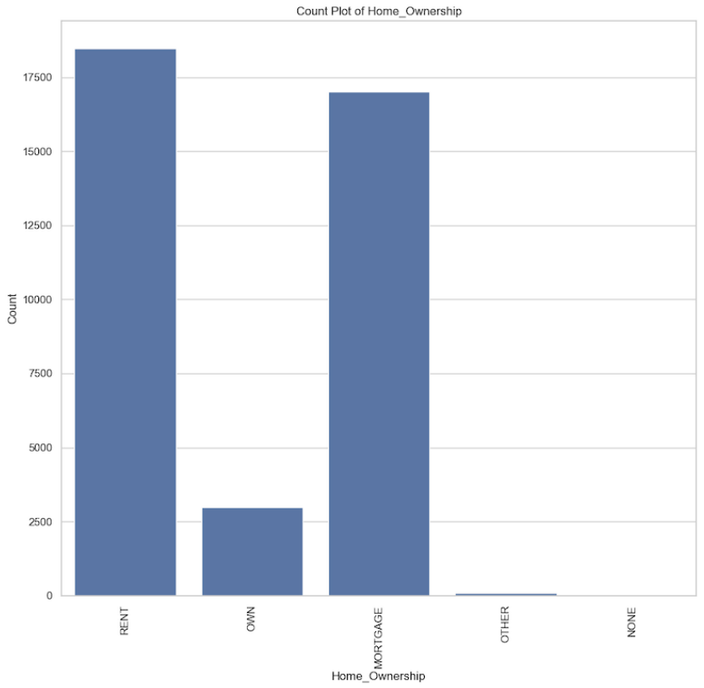
Results Univariate Analysis

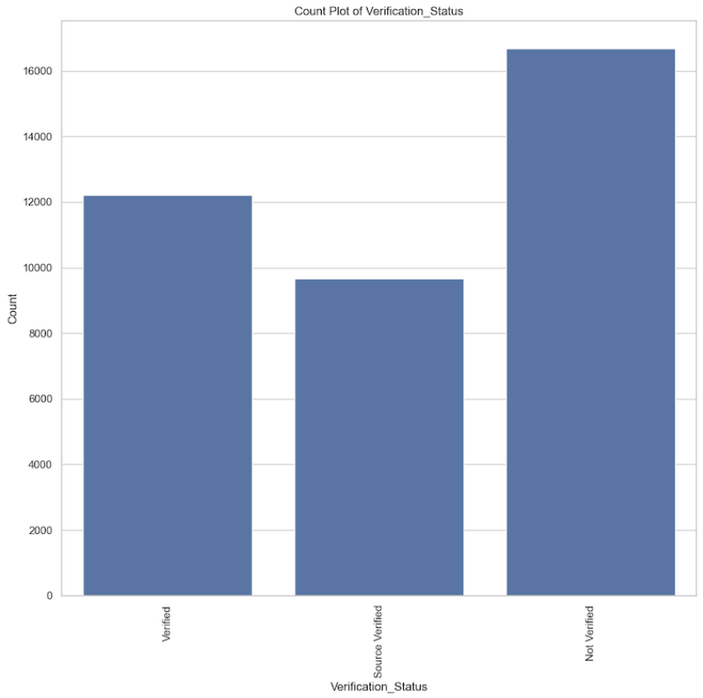












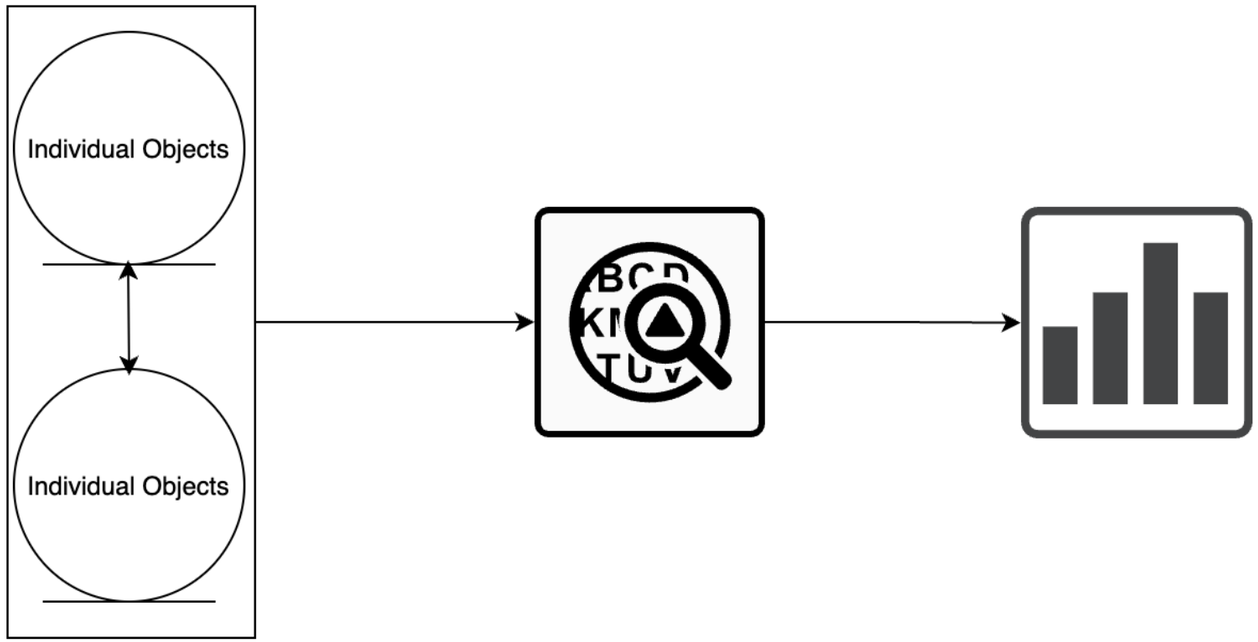
Bivariate Analysis

1. Performed extensive bivariate analysis on all the available numerical and categorical columns to compare their relationships.
2. The graphs were plotted as follows and type of columns.

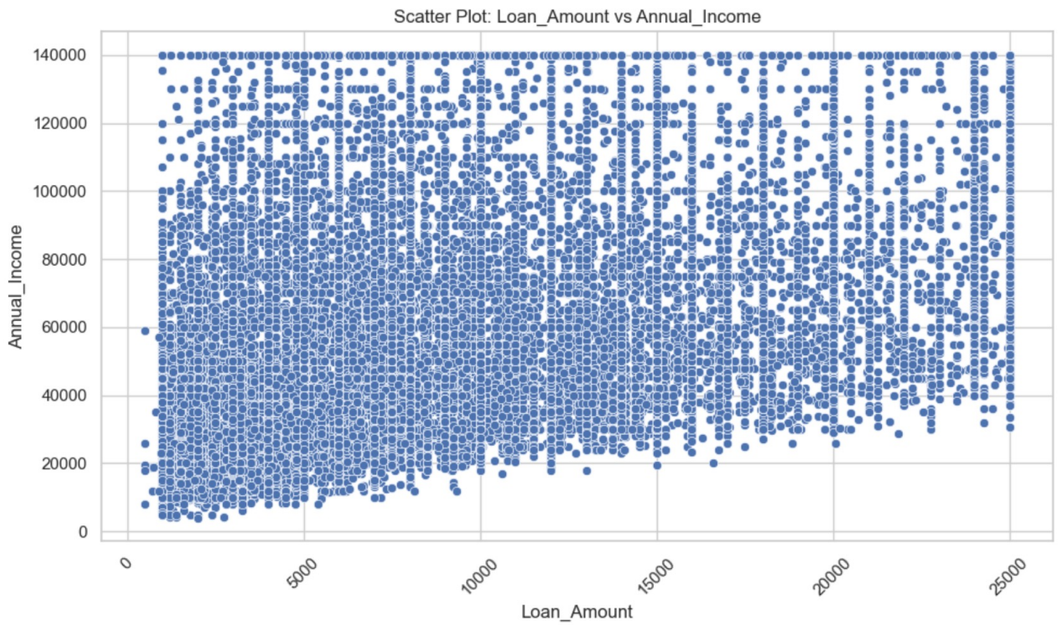
    - Scatter plots for numerical columns vs numerical columns.
    - Box plots for numerical columns vs categorical columns.
    - Bar plots for both categorical columns vs categorical columns.

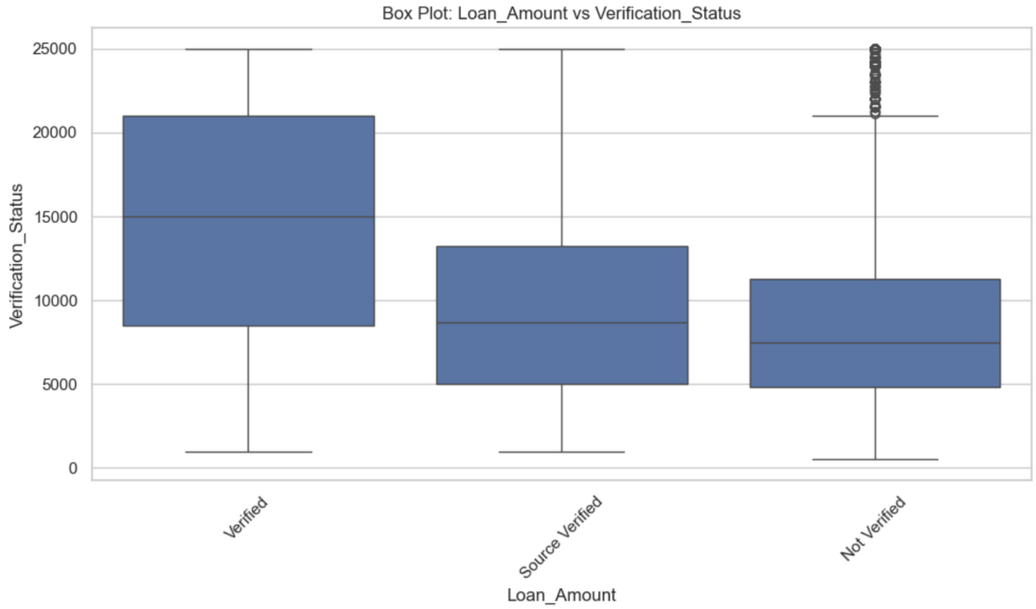
The sample results of the bivariate analysis can be seen in the next slide.

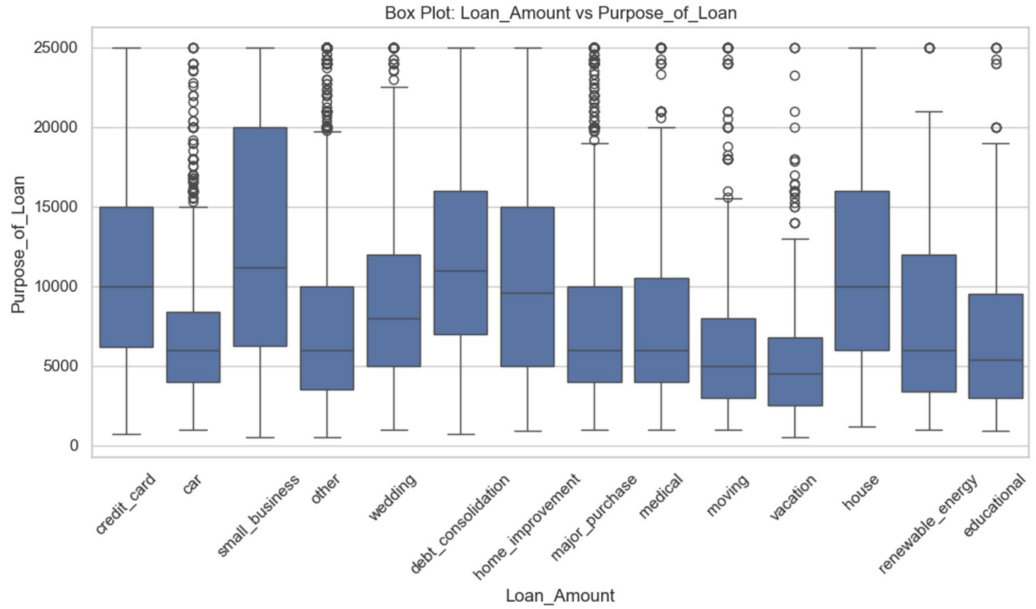
NOTE: Only few results were captured to accommodate the further analyses.

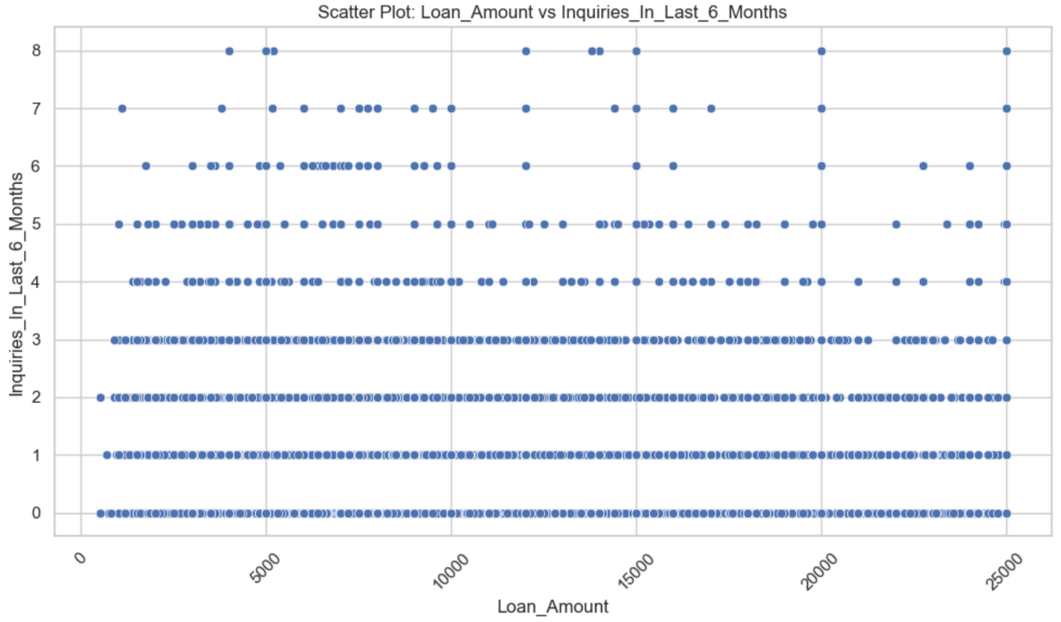


Results Bivariate Analysis





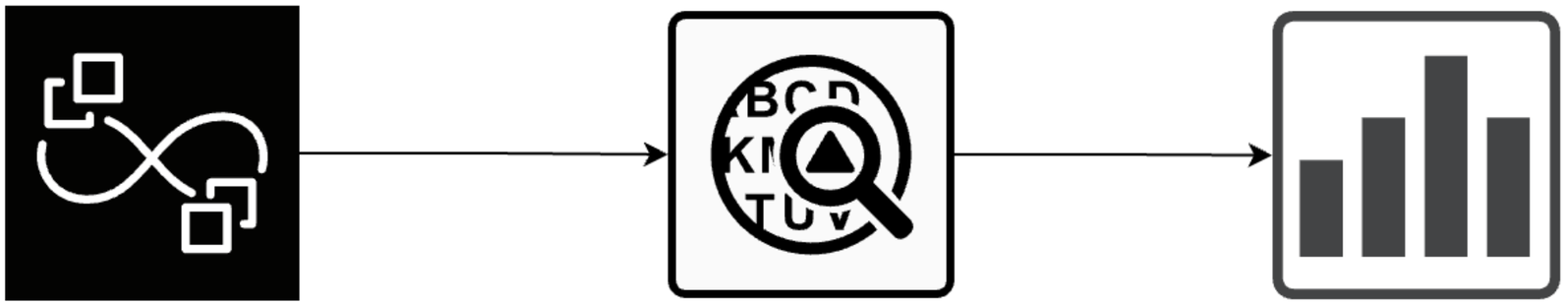




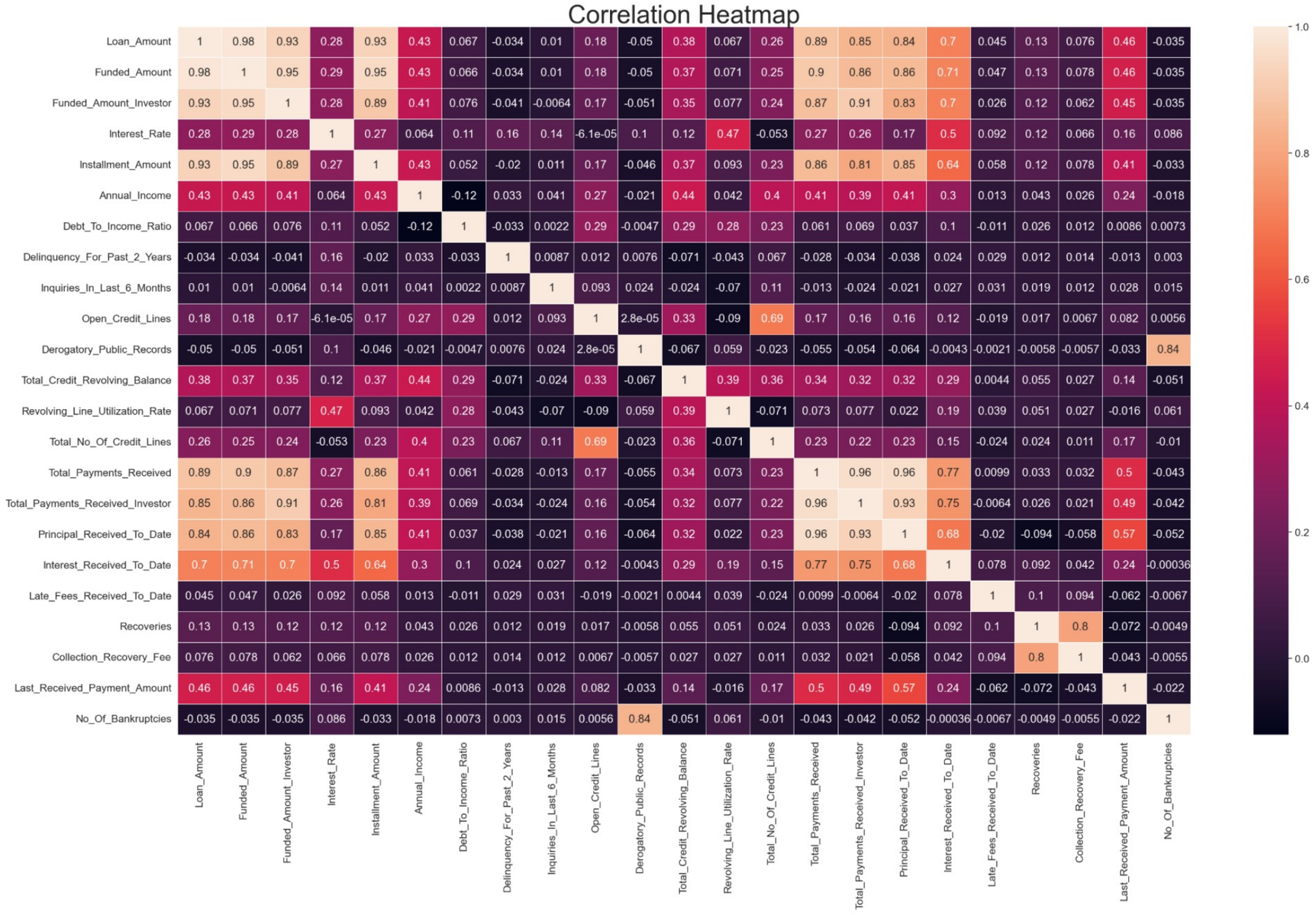
Multivariate Analysis

1. Performed multivariate analysis on all the available numerical to observe their correlation.
2. A heatmap was plotted to watch the correlation matrix.

The results of multivariate analysis can be seen in the next slide.



Results Multivariate Analysis



Derived Metrics & Insights

- Most of the loans are confined to lower amounts, within the range of $2,000 and $10,000.
- Interest rate mostly ranges between 5% and 20%.
- Annual income variance is very high and there are significant high-income outliers.
- Most of the borrowers have Debt-To-Income Ratio (DTI) less than 50% which indicates manageable debt levels.
- Majority of borrowers inquired 0 to 2 times in the last 6 months.
- Most borrowers have a low number of derogatory public records, indicating a generally good credit history.
- Revolving credit balances are generally low, but there are some borrowers with high revolving credit balances.
- The most common loan grades are 'B' and 'C', indicating midlevel credit risk.
- Home ownership status is highly 'RENT' or 'MORTGAGE', with fewer borrowers owning their homes.
- The primary purpose of loans is 'debt consolidation', followed by 'credit card' and 'home improvement’.
- Higher loan amounts tend to be associated with lower interest rates.
- Higher annual incomes are generally associated with higher loan amounts.
- Lower debt-to-income ratios are observed with higher annual incomes.
- Higher grades (A, B) are associated with higher loan amounts.
- Lower grades (C, D) are associated with higher interest rates.
- Borrowers with 'MORTGAGE' and 'OWN' status tend to have higher loan amounts.
- 'Source Verified' and 'Verified' statuses are associated with higher loan amounts.
- 'Fully Paid' loans are more common among higher grades (A, B).
- Debt consolidation and home improvement loans generally have higher loan amounts.
- 'Fully Paid' loans are more common among 'MORTGAGE' and 'OWN' statuses.
- Strong positive correlation suggests reliable repayment for larger loans.
- Positive correlation indicates higher repayment capability for borrowers with higher incomes.
- Strong correlation suggests higher grades are associated with a higher likelihood of loan repayment.
- Homeowners and mortgage holders tend to receive larger loan amounts.
- Debt consolidation and home improvement loans are generally larger.

Recommendations & Conclusions

Below are few recommendations and conclusions based on EDA against multiple variables.

Parameters	Recommendation	Defaulter Flag	Conclusion
High Annual Income, Low Debt-To-Income Ratio	Approve higher loan amounts with competitive interest rates.	UNLIKELY	Loan can be approved with confidence.
High Grades (A, B) with Verified Income	Approve loans readily as these applicants show strong repayment records.		
Mid Grades (C) with Moderate Debt-To-Income Ratio	Approve with stricter terms and possibly higher interest rates.	LIKELY	Loan should be approved with caution.
High Revolving Balances but Low Debt-To-Income Ratio	Approve with caution, ensuring the applicant's financial management is stable.		
Low Grades (D, E) with High Debt-To-Income Ratio	Deny the loan or require substantial collateral.	VERY LIKELY	Deny the loan application or request for additional guarantees.
Presence of Derogatory Public Records	Deny the loan or require co-signers or additional guarantees.		
High Interest Rates with Low Annual Income	Deny or approve with significantly reduced loan amounts and higher interest rates.		