Lending Club Case Study

Group Members: **Saumya Singh | Venkadesh Kannan**
Batch: **ML C70**

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower-interest-rate loans through a fast online interface.

Lending loans to ‘risky’ applicants is the largest source of financial loss Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In this case, the customers labelled as **'charged-off'** are the **'defaulters'**.

If one can identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants' using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors and variables behind loan default, i.e. the variables which are strong indicators of default. Then company can utilise this knowledge for its portfolio and risk assessment.

To develop our understanding of the domain, we are advised to independently research about risk analytics

(Understanding the types of variables and their significance).

BUSINESS UNDERSTANDING

Of Lending Club Case Study

PROBLEM STATEMENT

In this case study, we need to understand risk of analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

When a person applies for a loan, two types of decisions could be taken by the company:

Approved or Rejected

If the company approves the loan, there are 3 possible scenarios:

First is **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)

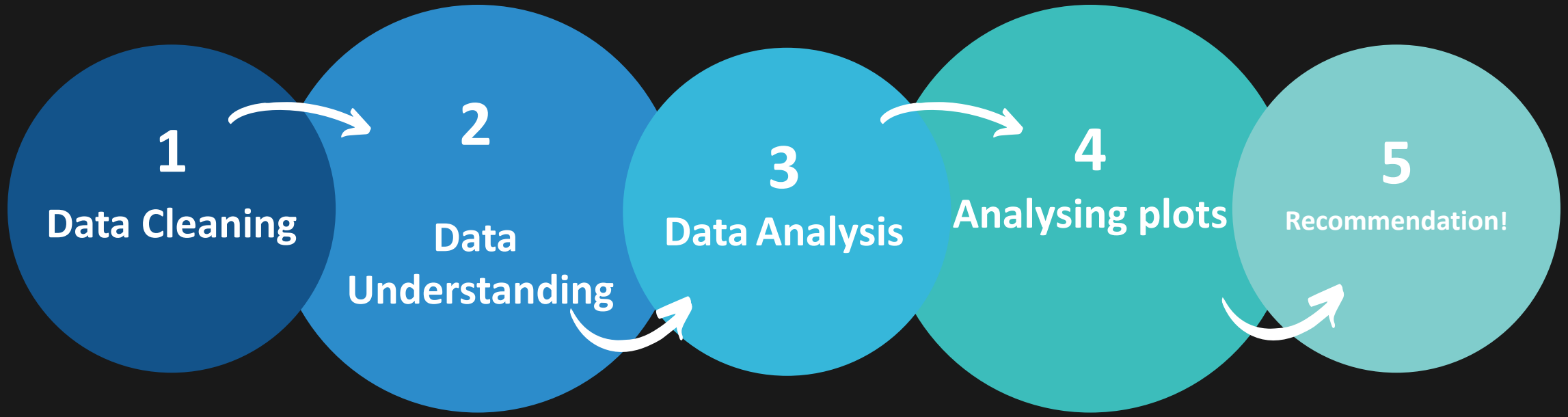
Next is **Current:** Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Lastly is **Charged-off:**

That indicates that the applicant has not paid the instalments in due time for a long period, i.e. he/she has defaulted on the loan

Data Analysis Methodology

5-Step Analysis Process



Data Analysis - Process

5-Steps Explained

DATA CLEANING

Check & Remove

- Duplicates,
- Null columns
- Values, unnecessary variables

DATA UNDERSTANDING

Using **Data Dictionary** understand about all the columns and their domain

DATA ANALYSIS

Univariate Analysis:

Analysing object column with plot distributions

DATA ANALYSIS

Bivariate Analysis:

Analysing the two-variable behaviour like term and loan status with respect to loan amount.

RECOMMENDATIONS

Analyse all plots and provide recommendations for reducing the loss with Loan status

Data Understanding and Analysis

Overview | Alerts 143 | Reproduction

Dataset statistics

Number of variables	111
Number of observations	1000
Missing cells	56343
Missing cells (%)	50.8%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	867.3 KiB
Average record size in memory	888.1 B

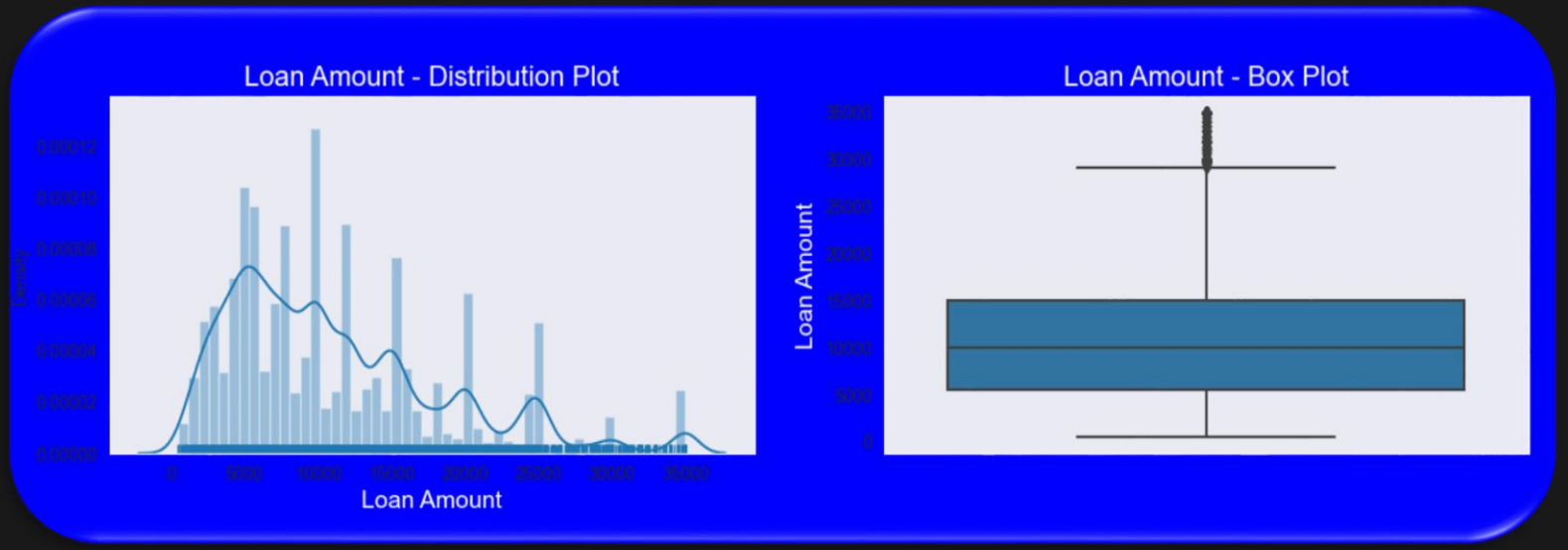
Variable types

Numeric	23
Categorical	21
Text	10
Boolean	2
Unsupported	55

- **We will perform data cleansing by applying:**
  - Remove duplicates & null values
  - Drop unwanted columns/values
- **Exploratory Data Analysis will be performed by applying:**
  - Univariate Analysis
  - Segmented Univariate Analysis
  - Bivariate Analysis
- **Recommendations will be given after analysing all plots and metrics.**

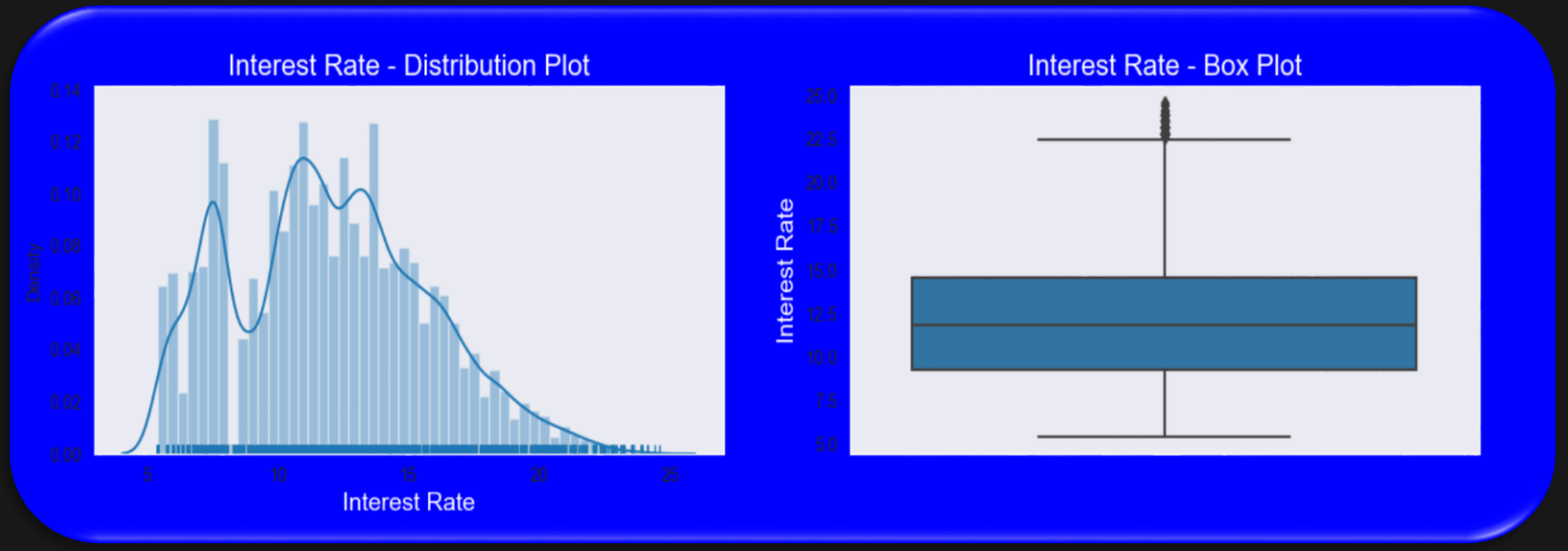
 # Observations :

Below plots show that most of the Loan amounts are in range of 5000 - 15000



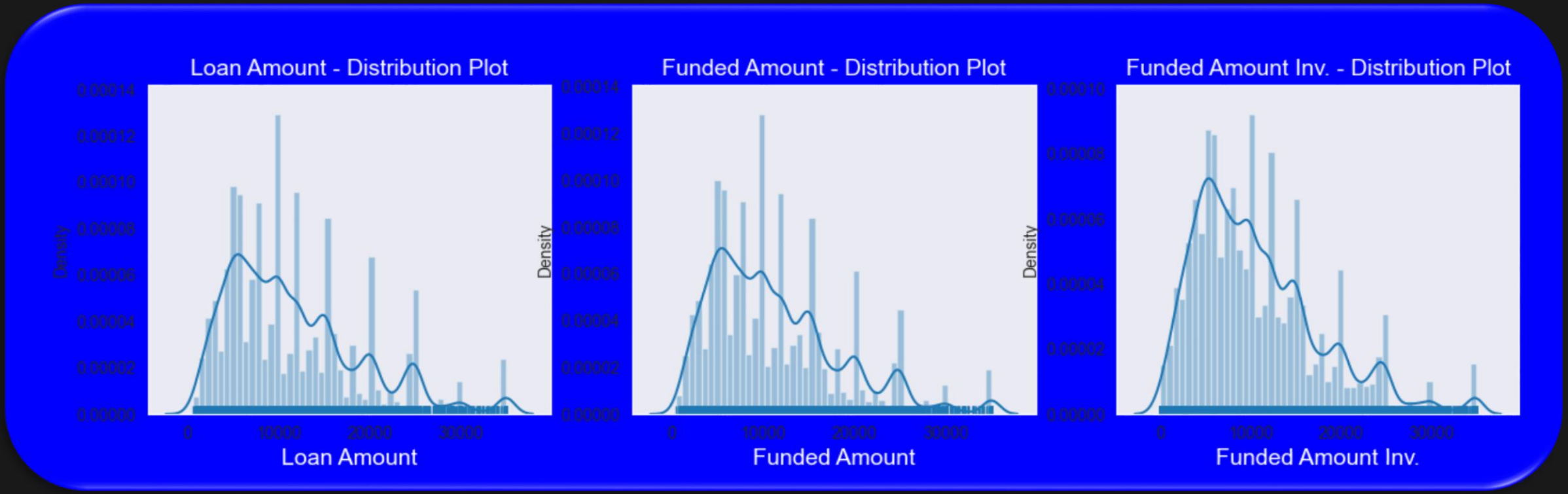
 # Observations :

Below plots show that most of the Interest Rates on loans are in range of 10% - 15%



Univariate Analysis

Distribution of three loan amount fields using a distribution plot



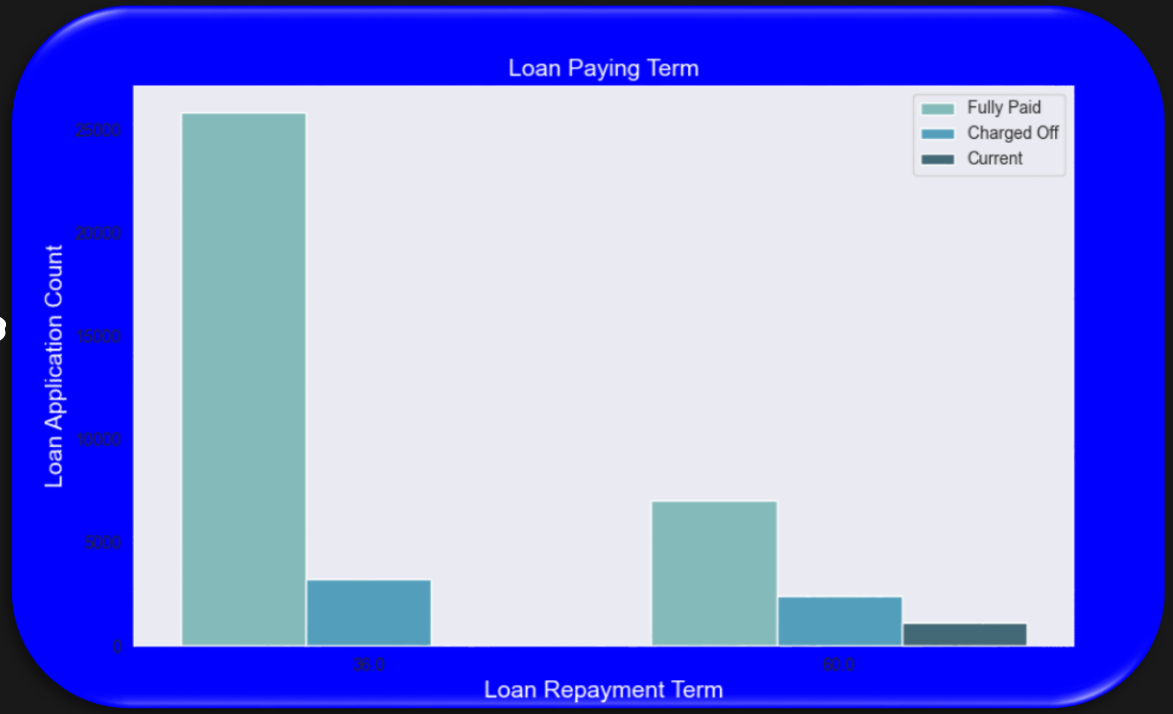
 # Observation:

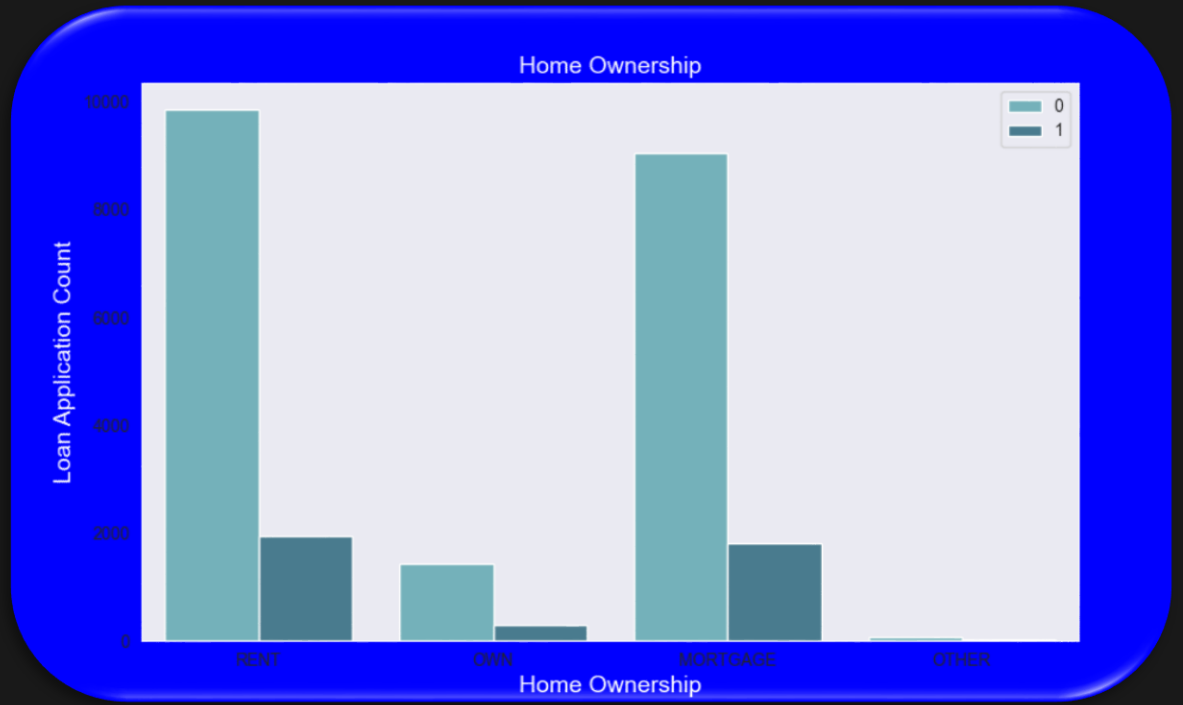
Distribution of amounts for all three looks very much similar.

We will work with only loan amount column for rest of our analysis.

 # Observations :

This plot shows that those who had taken loan to repay in 60 months had more % of number of applicants getting
charged off as compared to applicants who had taken loan for 36 months.

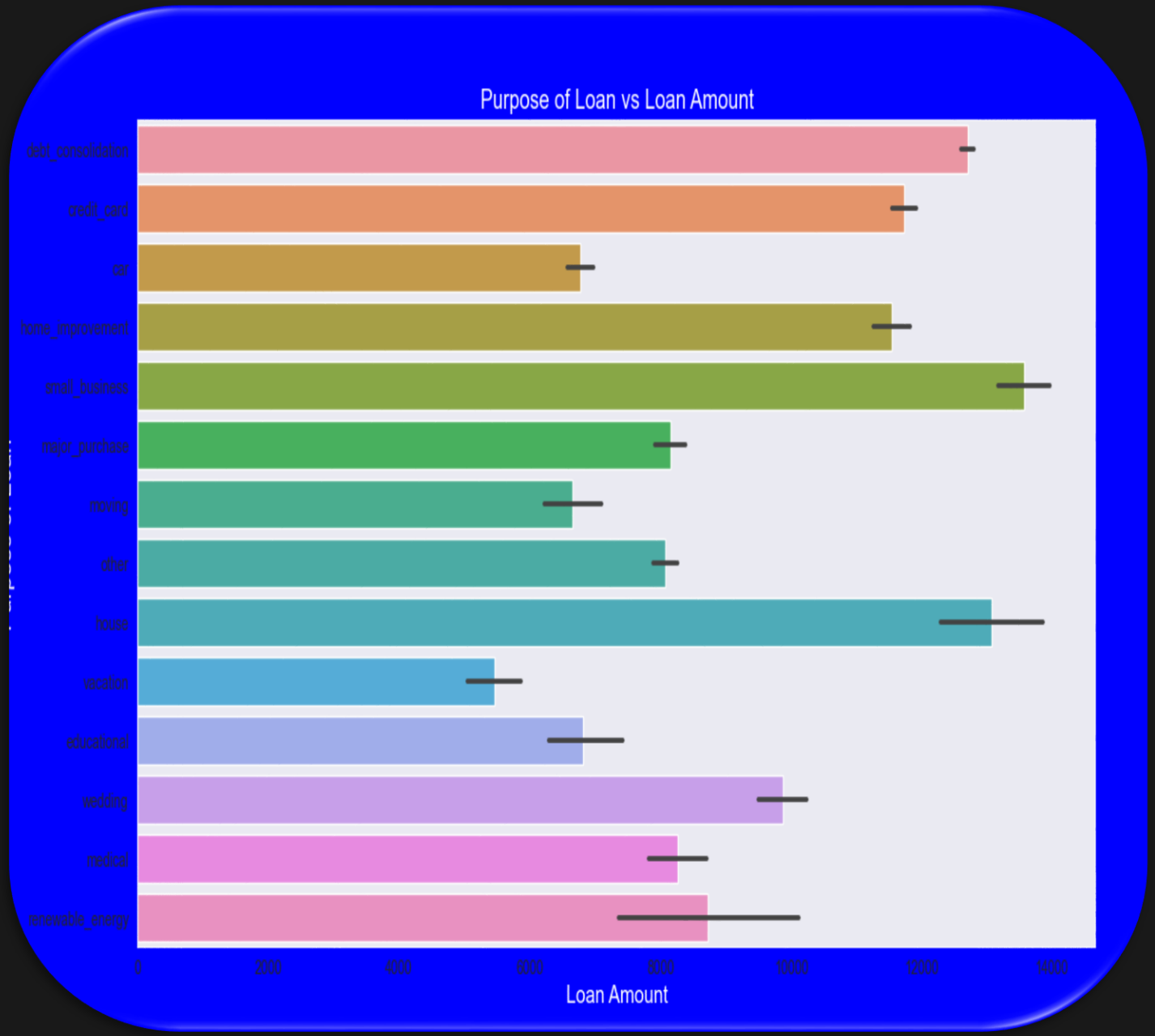




 # Observations :

This plot shows that most of them living in rented home or mortgaged their home.

Applicant numbers are high from these categories so charged off is high too.



Bi-Variate Analysis

Observations:

Loan taken for **Small Business purpose**, **Debt consolidation and Credit cards,** are somewhat evenly distributed as compared to loan taken for other purposes**.**

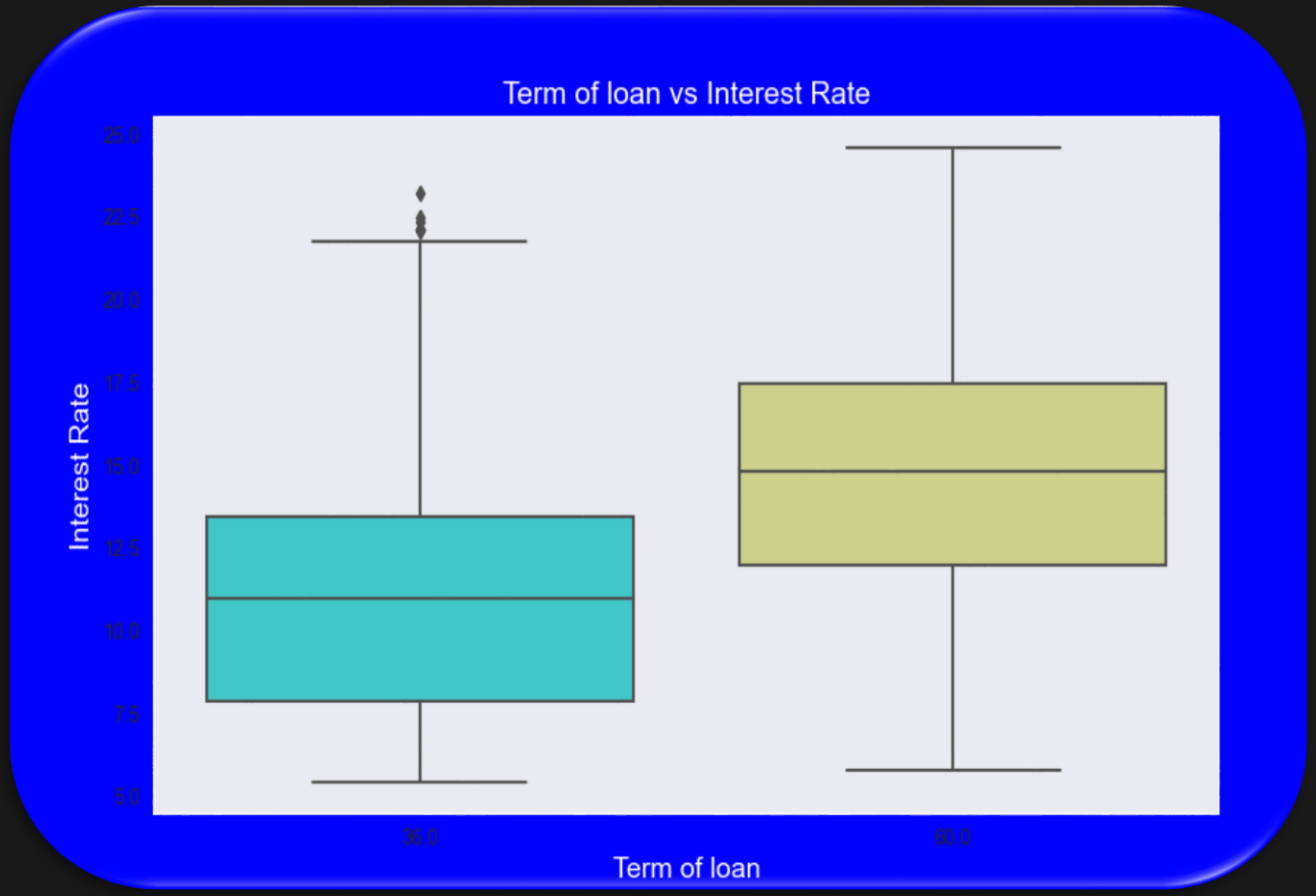
Bivariate Analysis

BarPlot to show the variation of annual income across grade for every loan status

 # Observations:

It is clear that average interest rate is higher for 60 months loan term.

Most of the loans issued for longer term had higher interest rates for repayment.



Bivariate Analysis

Term of Loan vs Interest Rate

 # Observations:

From this we can conclude that the ones getting 'charged off' have lower annual incomes than the ones who paid fully' for every grade (i.e. at same interest range)

Bivariate Analysis

Plotting Heatmap for any Linear Correlation between Pairs of Variables

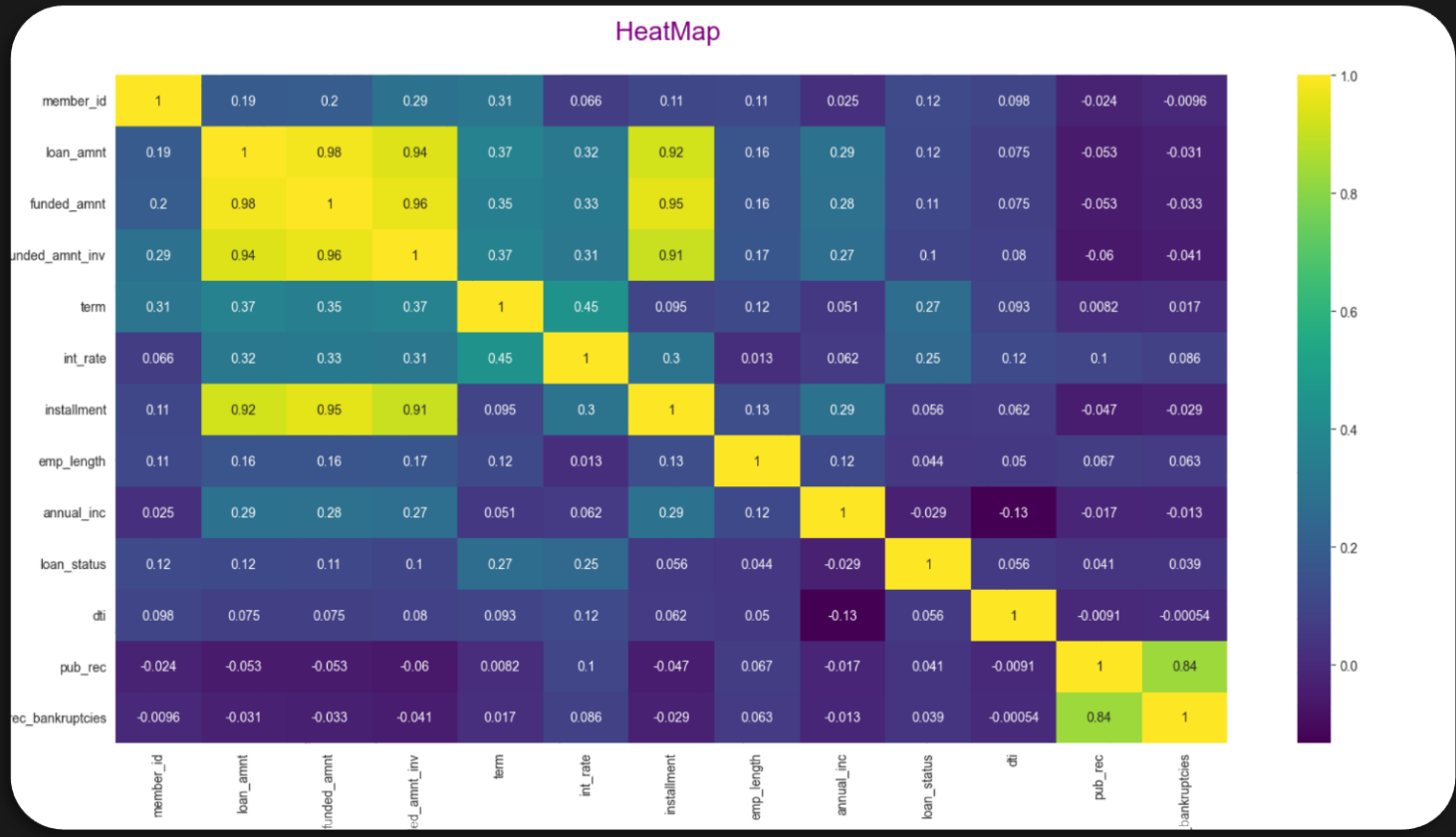
 # Observations:

Loan amount, investor amount, funding amount are strongly correlated.

Annual income with DTI(Debt-to-income ratio) is negatively correlated.

Debt income ratio is the percentage of a consumer's monthly gross income that goes toward paying debts which means when annual income is low DTI is high & vice versa.

Positive correlation between annual income and employment years which means income increases with work experience



Conclusions

Grades are good metric for detecting defaulters. Lending Club should examine more information from borrowers before issuing loans to Low grade (G to A).

Lending Club should control the number of loans issued to borrowers who are from CA, FL and NY to make profits.

Borrowers with mortgage homeownership are taking higher loans and defaulting on the approved loans.

The lending club should stop giving loans to the above category when the loan amount requested is more than 12000.

Conclusions

People with more public derogatory records have more chance of filing a bankruptcy.

The lending club should make sure there are no public derogatory records for borrowers.

.

THANK YOU!

Saumya Singh // Venkadesh Kannan