Lending Club Case Study

ABHISHEK SAXENA

PRANAV KUMAR PEDDI RAVI

Table of Contents

Problem Statement

When the company (Lending Club in this case) receives a loan application, the company must decide on loan approval/ rejection based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

1. If the customer is more likely to repay the loan, then rejecting it would result in loss of business for the company.
2. If the customer is going to be a defaulter, then approving such application may result in financial loss to the lender.

Given the dataset from Lending Club about past loan applications with status as 'Defaulted' or 'Not Defaulted' the aim is to find out the patterns which indicate if a person is likely to default, which may be used to take further actions approving/ denying loan applications; some suggested actions may include:

- Lending at high interest rates to risky applications.
- Reducing the amount to be allocated.
- Rejecting the complete application.
- Reducing the term of the application.

As part of this exercise, we will perform Exploratory Data Analysis on the given data set to understand the relationship below different set of attributes e.g., Consumer attributes, loan attributes. Portfolio Risk analysis will help the lender on taking various decisions and avoid losses.

Technologies Used

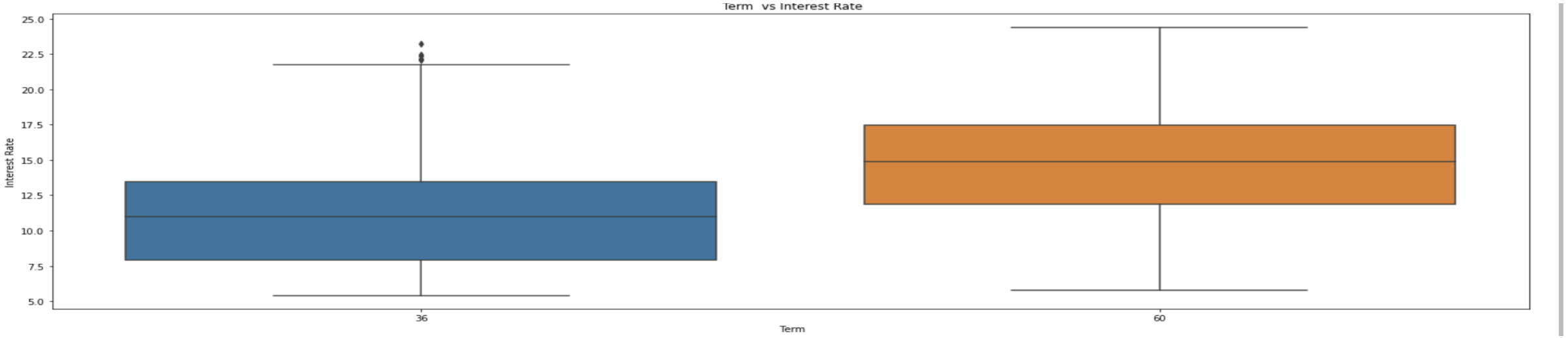
- Python 3.7
- Pandas 1.1.3
- NumPy 1.19.2
- Matplotlib 3.2.2
- Seaborn 0.11.0

Exploratory Data Analysis (EDA)

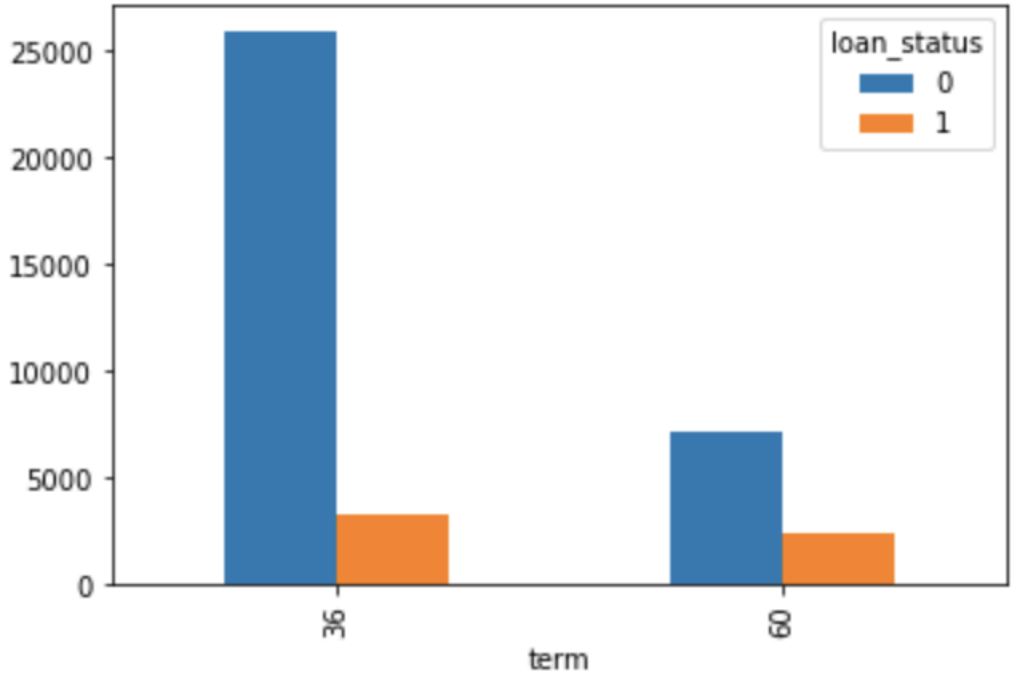
Dataset and corresponding data dictionary has been provided by Upgrad which contains the complete loan data for all loans issued through the time period 2007 t0 2011. It contains a total of 39718 Rows with 111 columns. After loading the dataset, following steps were performed:

1. Check for missing values across all columns and dropped the columns which have values not available, or the presence is less then 30%
2. Dropped all columns which have common values across all rows.
3. For data cleansing, removed unwanted special characters, text from columns like **term** and **interest rate**.
4. Converted the Date Time columns in the uniform format.
5. Removed all the columns with Customer Behavioral Data.
6. Deleted all the rows with Loan Status as current as the analysis should only include **Fully Paid** or **Charged Off** applications.
7. Assigned the values 0/1 for loan status as Fully Paid/ Charged Off to convert the loan status in Categorical values.
8. To convert continuous data (loan amount, DTI, Interest Rate, Annual Income) into categorical data used binning with separate labels so they can be part of analysis.

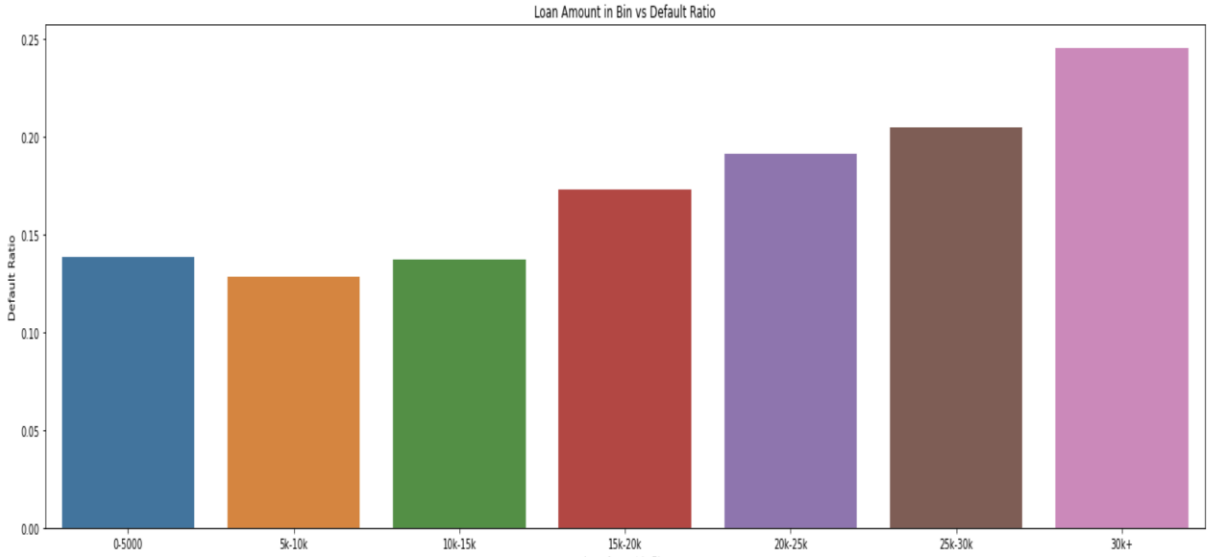
Term Analysis

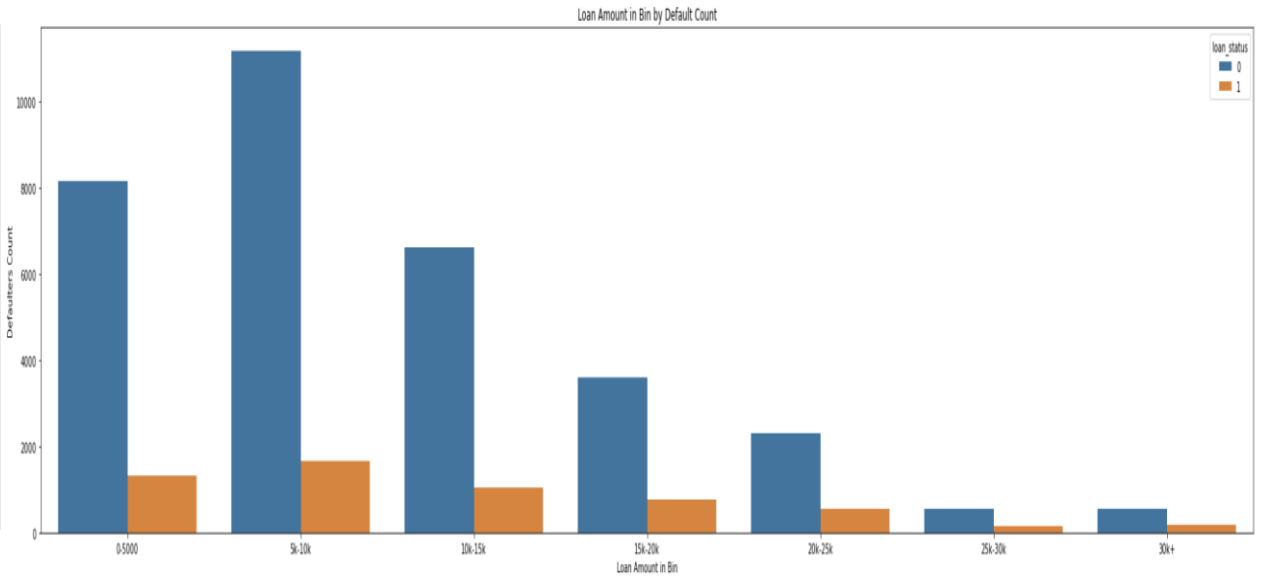


- On graph Term vs Defaulters, we could observe an increase in defaulter as the term increases
- On graph Interest rate vs Term, we could observe that term and the interest rate are positively co-related
- We could see that loners who are talking higher loan term are most likely to default

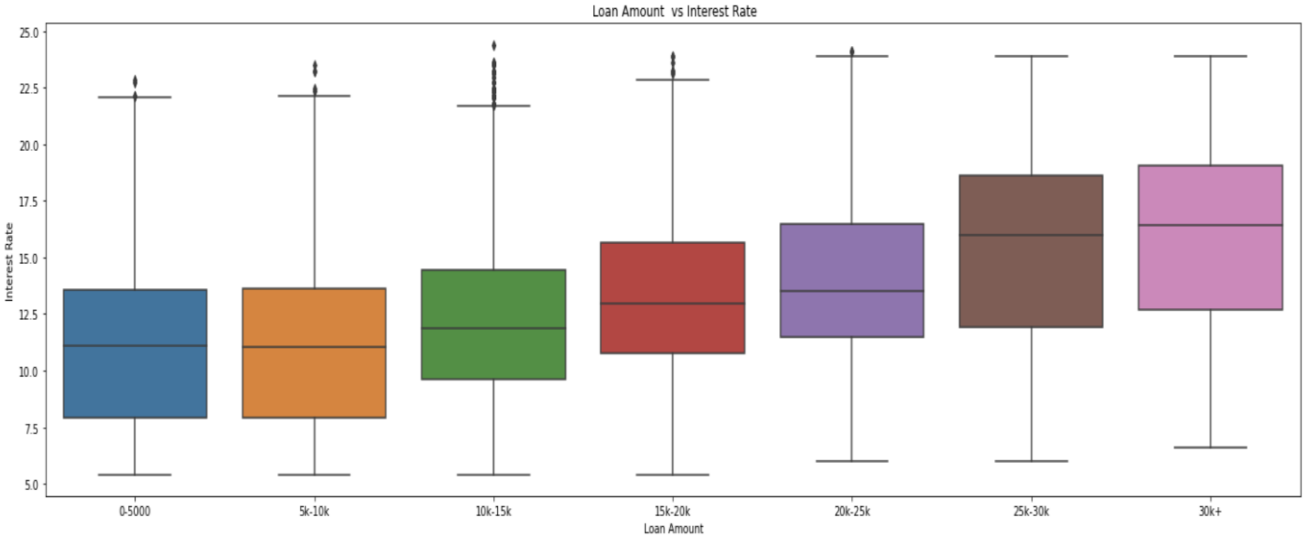


Loan Amount Analysis

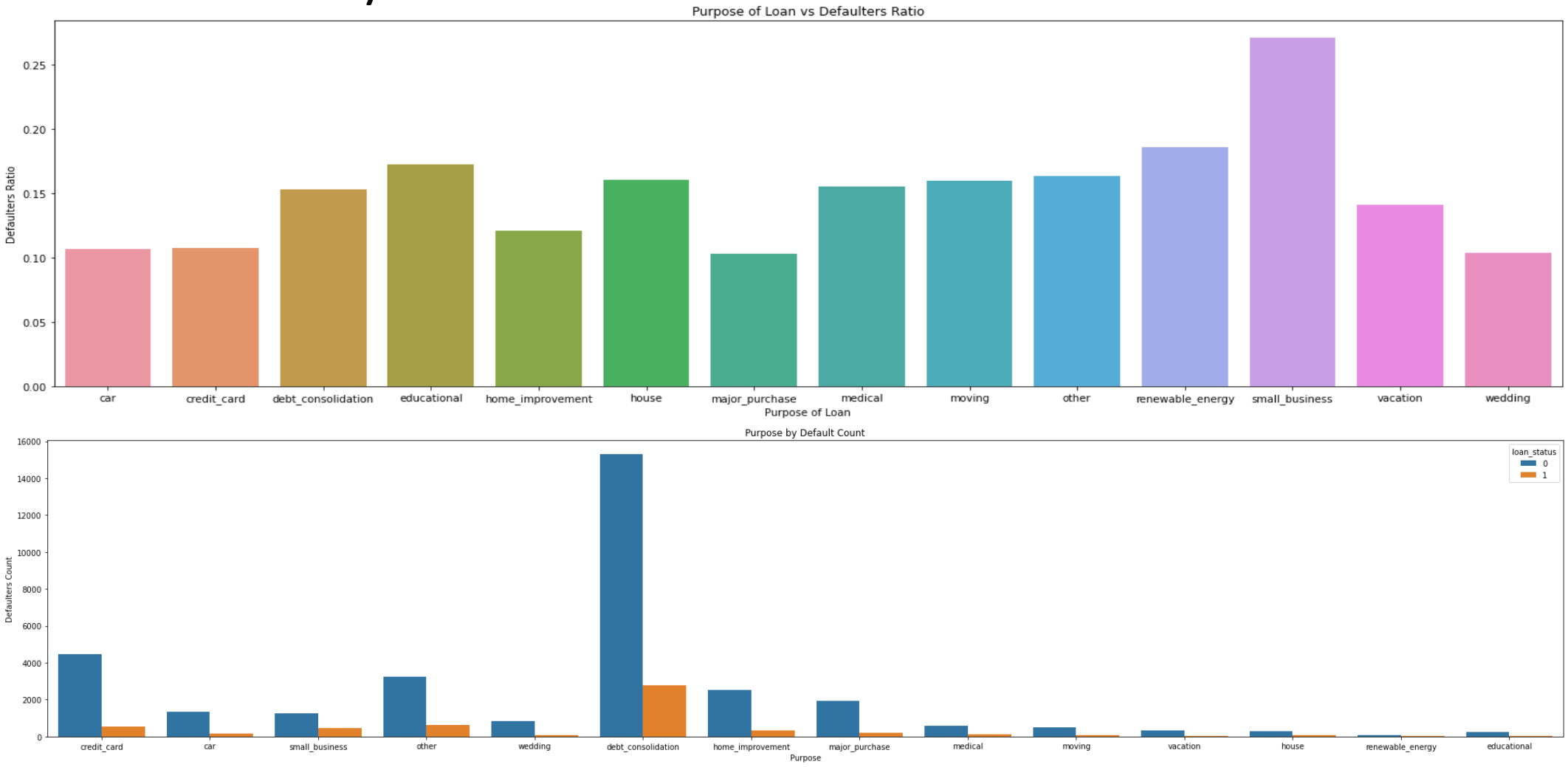




- On graph Loan Amount vs Default Ratio, we could observe an increase in defaulter as the loan amount increases
- On graph Loan Amount vs Default Count, we could observe that the max number of defaulters are for the range of 5-10k. Though the percentage of defaults is less as observed in Loan Amount vs Default Ratio
- On graph, Loan Amount vs Interest Rate, we could observer that these two factors are correlated positively.
- We could possibly deduce that for amount(s) greater than 15k, Lending Club should be careful in giving out the loans



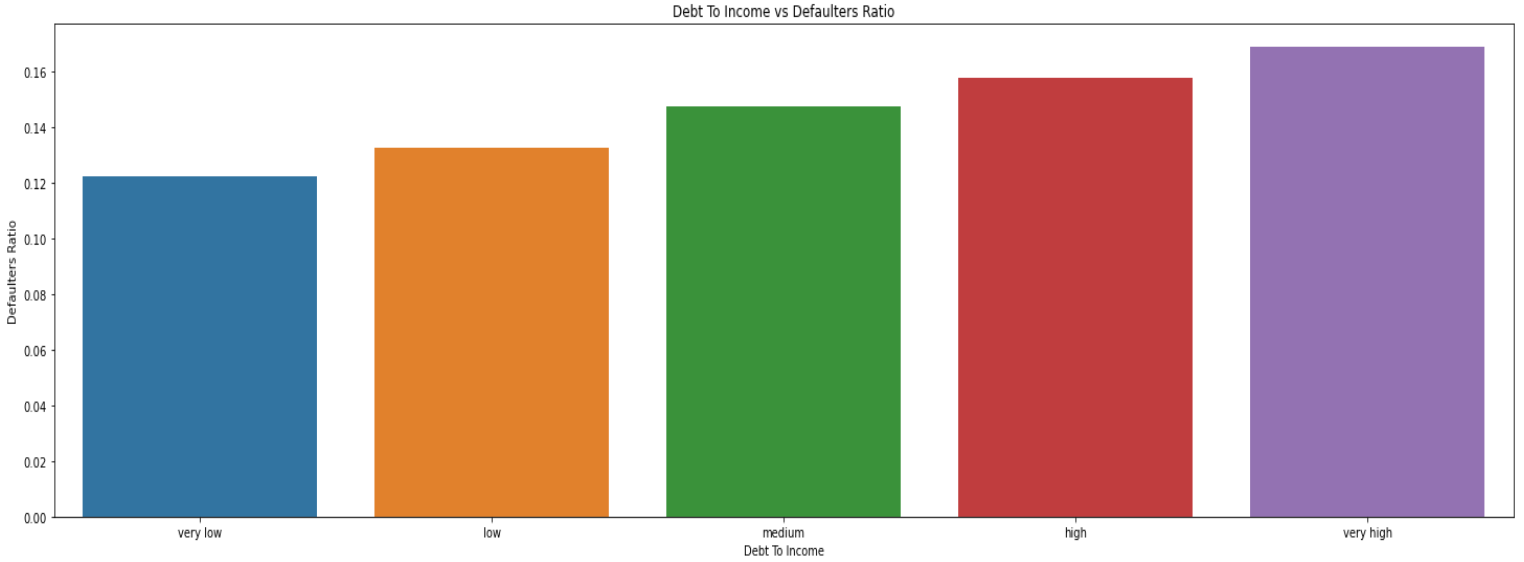
Purpose of Debt Analysis

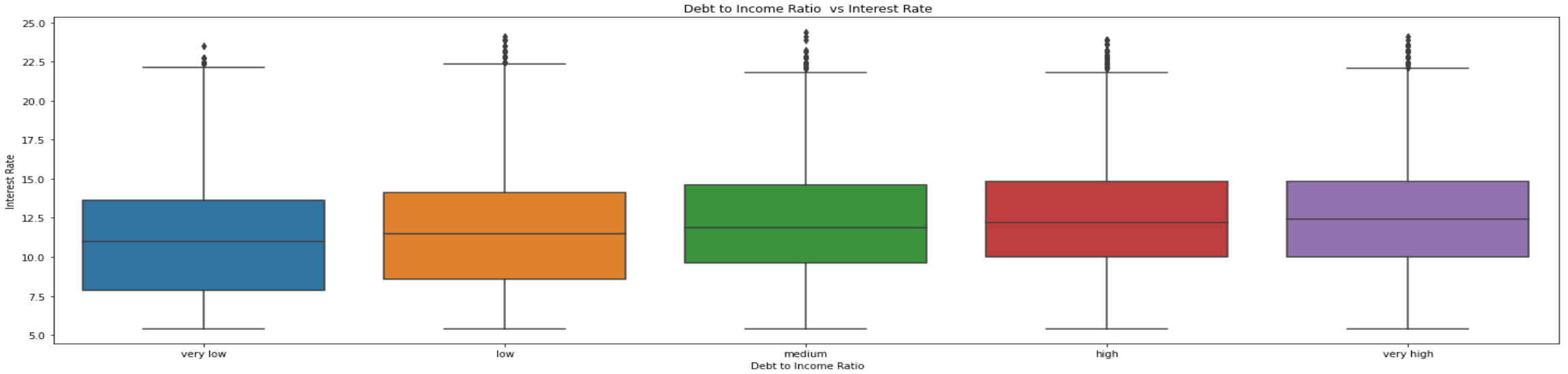


- On graph Purpose vs Default Ratio, we could observe applicants seeking loan for small business are majority of defaulters
- On graph Purpose vs Defaulter's count, we can conclude that credit card and Debt applications are contributing most to the defaulters.

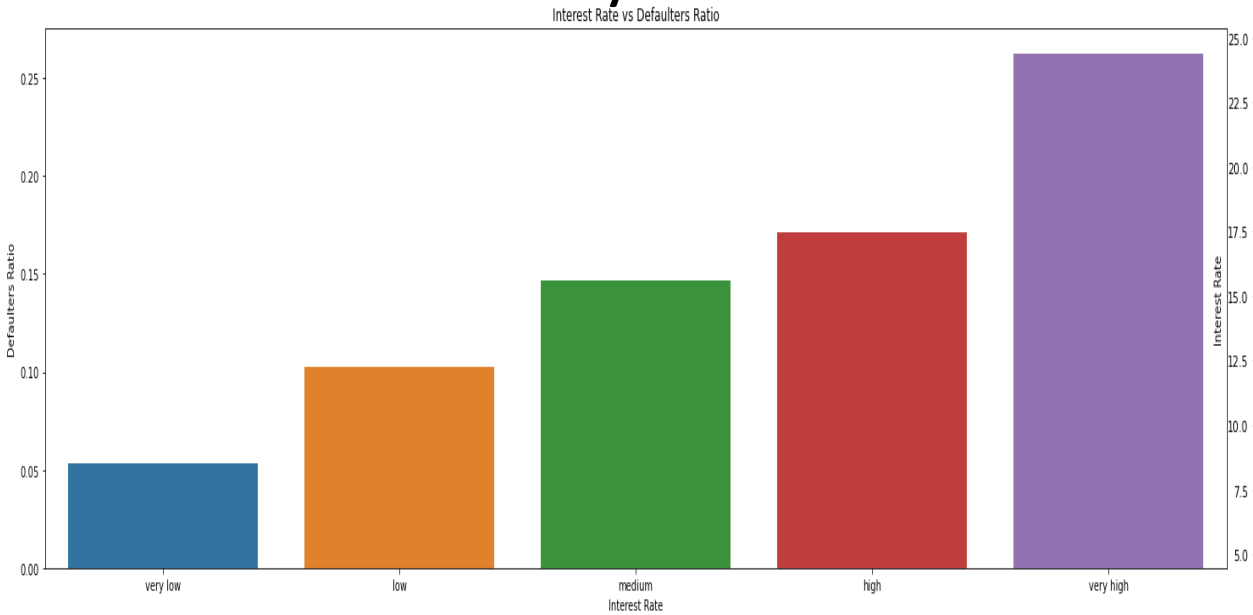
Debt to Income Ratio Analysis

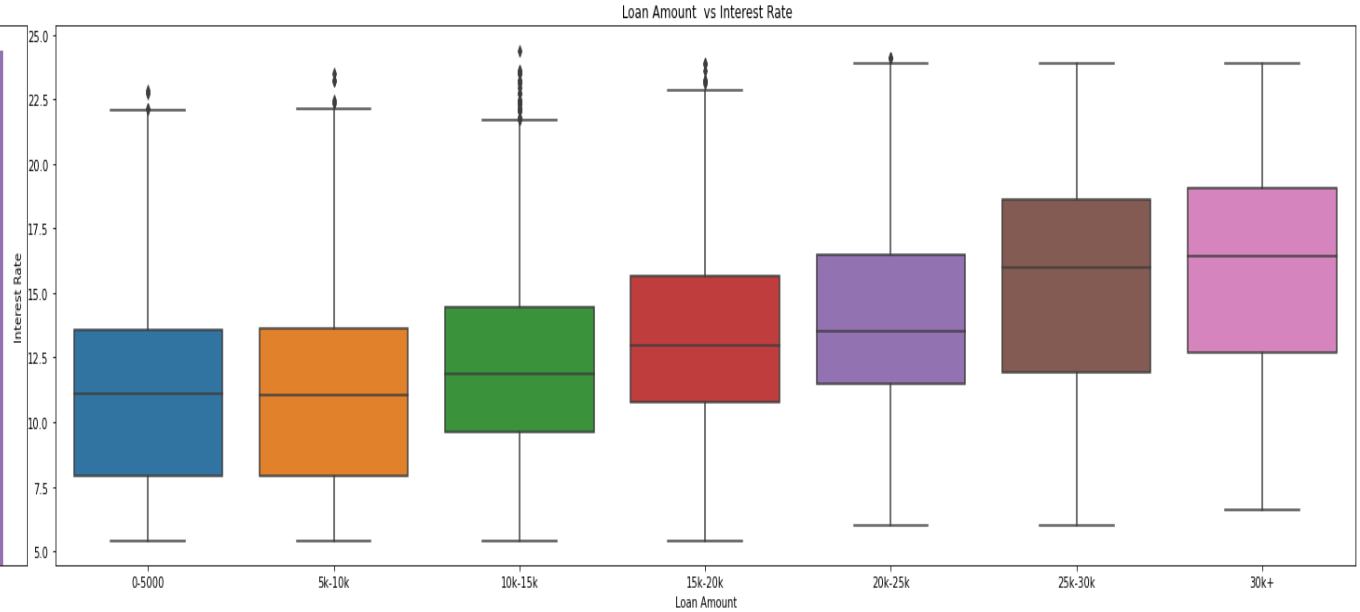
- On graph Defaulters Ratio vs DTI, the number of defaulters increases with the increase in Debt-to-Income ratio.
- Debt to Income Ratio and Interest rates are positively correlated. We could see the median increases with Debt-to-Income Ratio and Interest rates



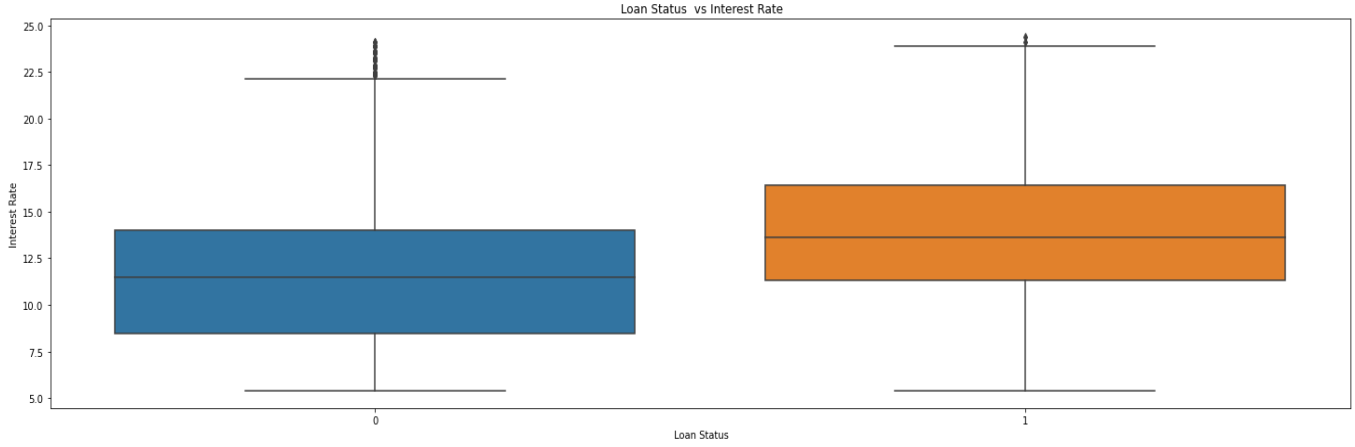


Interest Rate Analysis



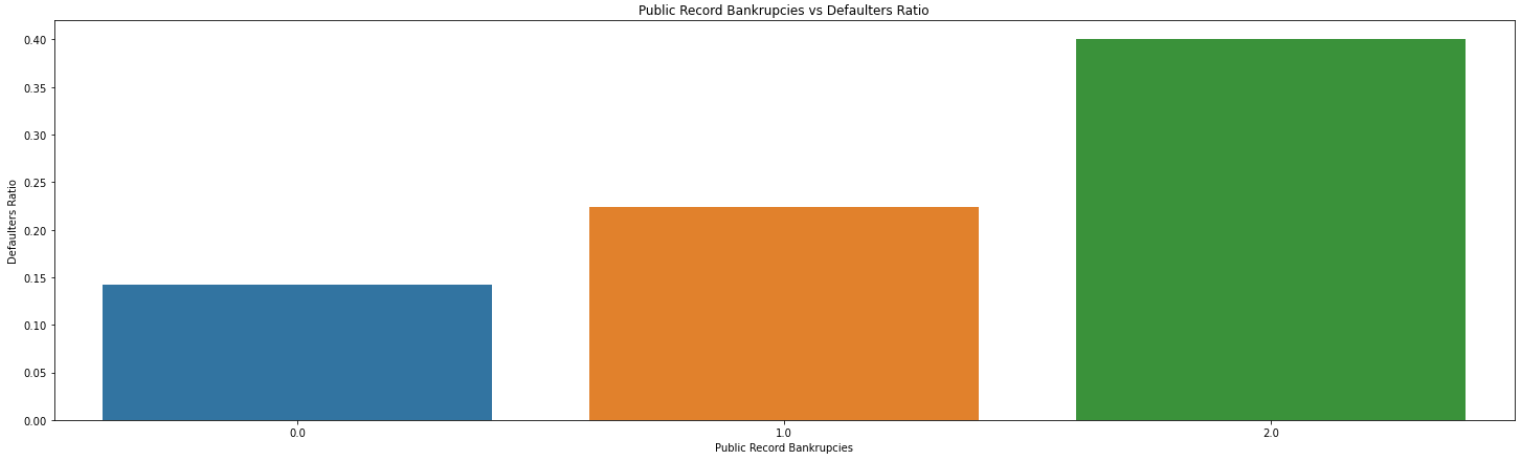


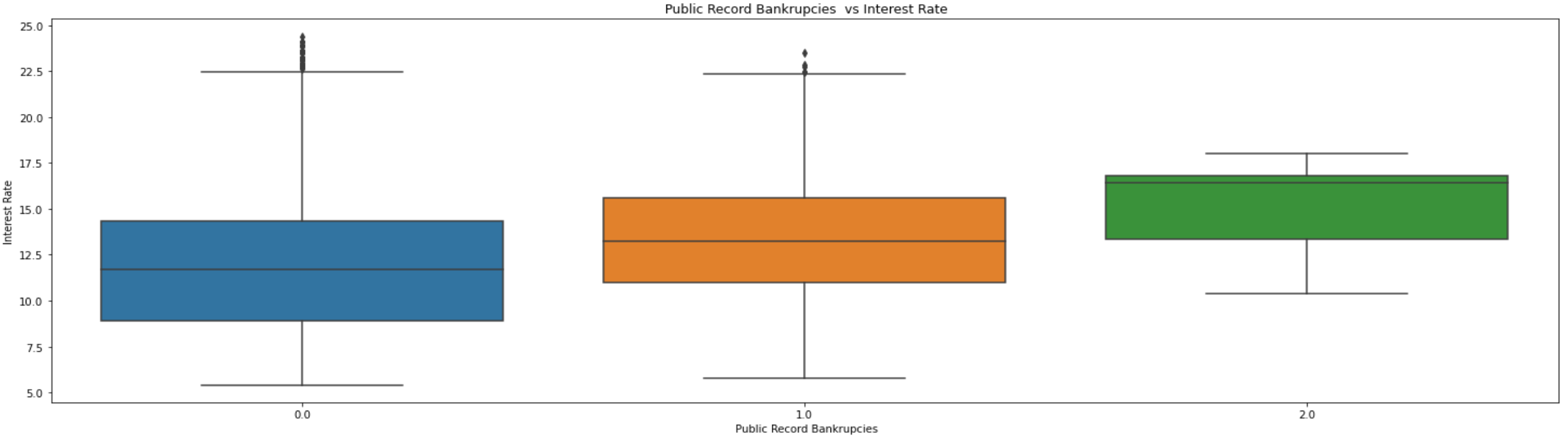
- As Interest Rate increases, the defaulters count also increases in a linear manner.
- On graph Loan Amount vs Interest Rate, we could observe that these 2 variables are positively co-related.
- On graph, Loan Status vs Interest Rate, we could observe that lower interest rate loans are less likely to default.
- We could possibly deduce that interest rate and loan term should be properly calculated to avoid very high interest rates.
- Interest rates over 15% are most likely to default



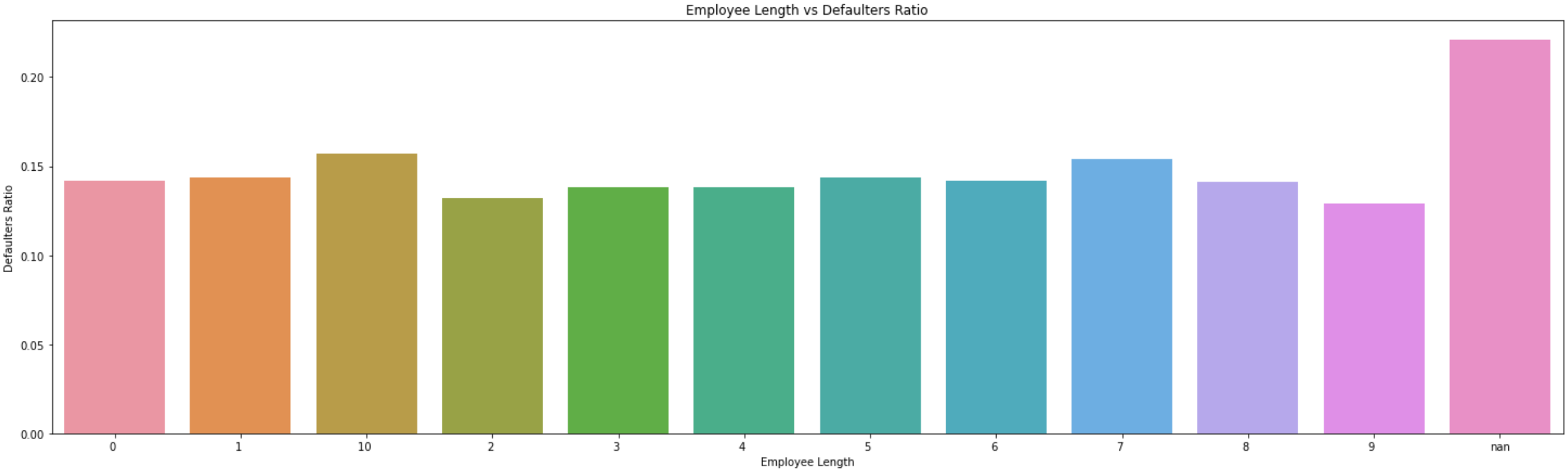
Public Record Bankruptcies Analysis

- On graph in defaulters Ratio vs Public Record Bankruptcies, the number of defaulters increases with the increase in bankruptcies.
- Public Record Bankruptcies and Interest rates are positively correlated. We could see the Median increases with Public Record Bankruptcies and Interest rates
- Loaner with more than 1 public record bankruptcies are likely to default more than 25%



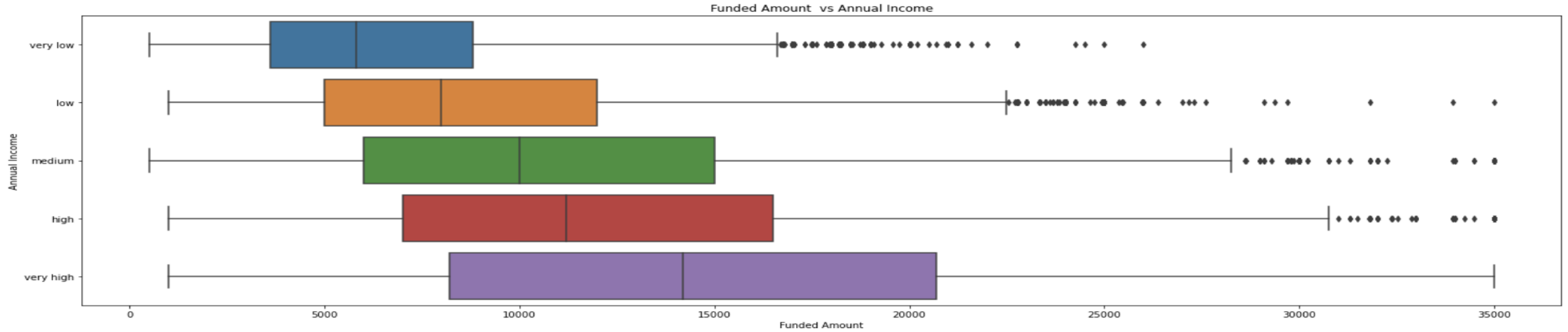


Employment Length Analysis

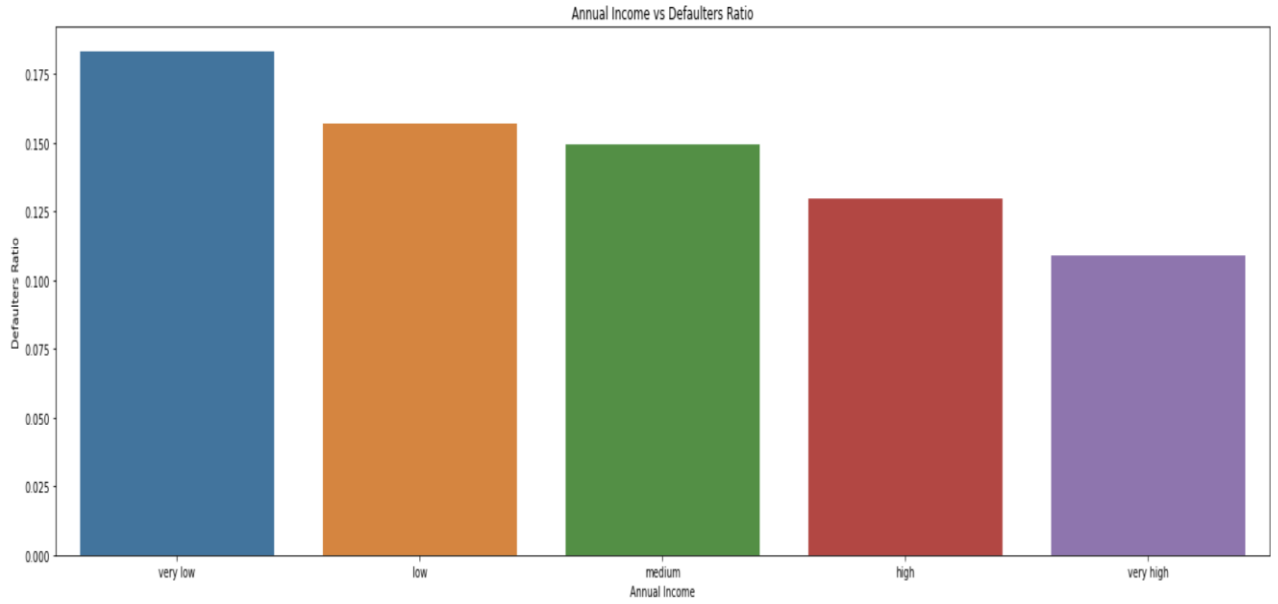


- Applicant’s employment length doesn’t seem to have any significant impact on Defaulter's count.

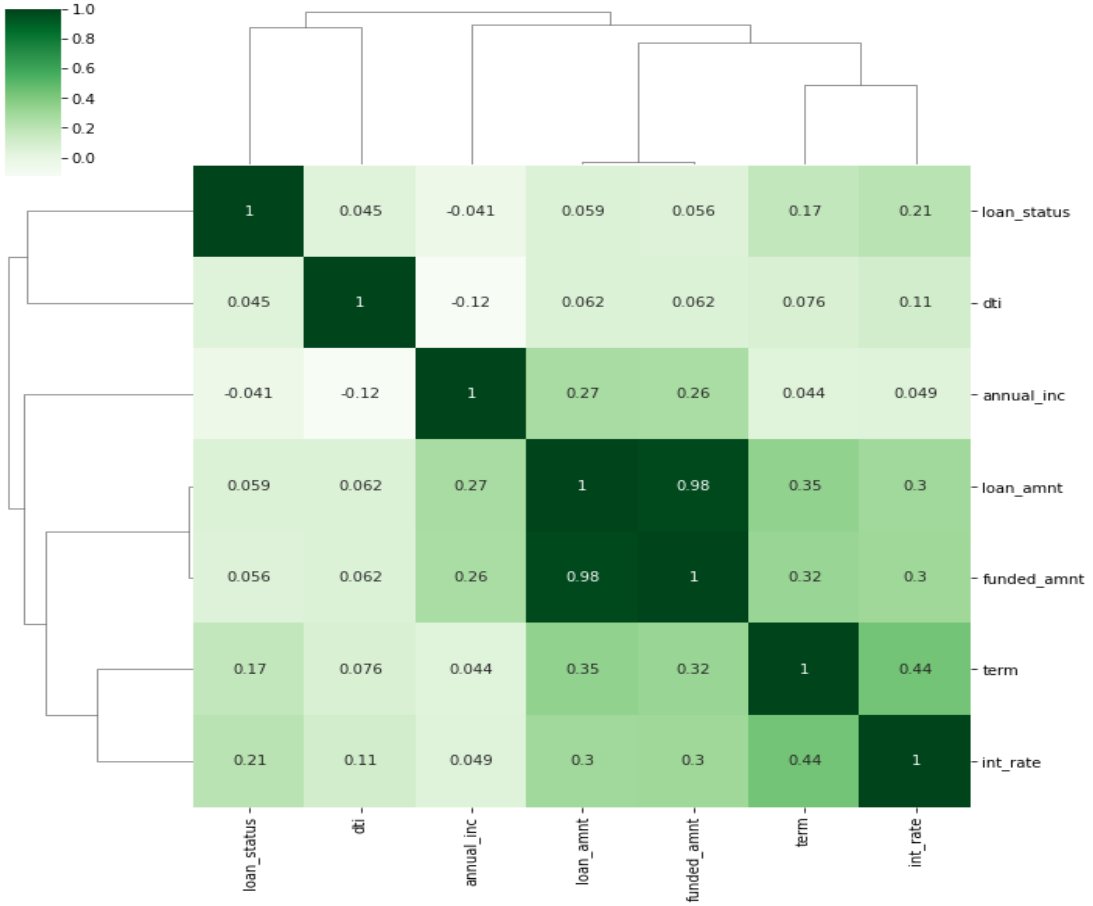
Annual Income Analysis



- Loaners with high income are very unlikely to default
- Loaners with low income has higher percentage to default which is around 17.5%
- Loaners with high annual income are provided with high loan amount
- Annual amount and loan amount are positively correlated



Correlation Matrix (HeatMap)



Pearson coefficient matrix for the 7 most important variables used in the analysis. It shows there is high correlation between Amount requested and Amount funded. Also, all these variables contribute to the decision making as they are mostly correlated.

Observations:

Lending club should consider these factors before giving loan

- Loaners are likely to default more for 60 months period
- For loan amount(s) greater than 15k, loaners are likely to default
- Small businesses are likely to default the most, which is around 25%
- Loaners with higher interest rates are likely to default
- Employee length has no impact on defaulters
- Lower income loaners are likely to default around 17.5%
- Loaners with high Debt to income ratio are likely to default
- Loaner with at least one record in discrepancies are likely to default to 20 % and it increase to more than 25% for loaners over one record in discrepancies
- Debt Consolidation, Credit Card, Home Improvement are the highest defaulting categories.
- Customers with High Loan amount to Annual Income ratio have more chances of turning into Defaulters.

Acknowledgements:

- Original Data Set and Data Dictionary was provided by Upgrad.
- StackOverFlow was used to check the syntax for various parameters for plotting Graphs.