Lending Club Case Study

Group Members:-

1)Mihir Pawar (Group Facilitator)

2)Bharadwaj Mokshagundam

UpGrad

iiit-b International Institute of Information Technology Bangalore

Steps Involved

1)Understanding of the problem statement
2)Understanding of the dataset using data dictionary
3)Data Cleaning
4)Data Preparation
5)Exploratory Data Analysis

Data Cleaning and Data Preparation

1)Deleted unnecessary columns from the data.

2)Treat Missing values as per domain knowledge and statistical approach.

3)Removed columns with same values throughout the column.

4)Removed columns with too many unique values.

5)Created new columns(Derived Metrics) that give us new information about data.
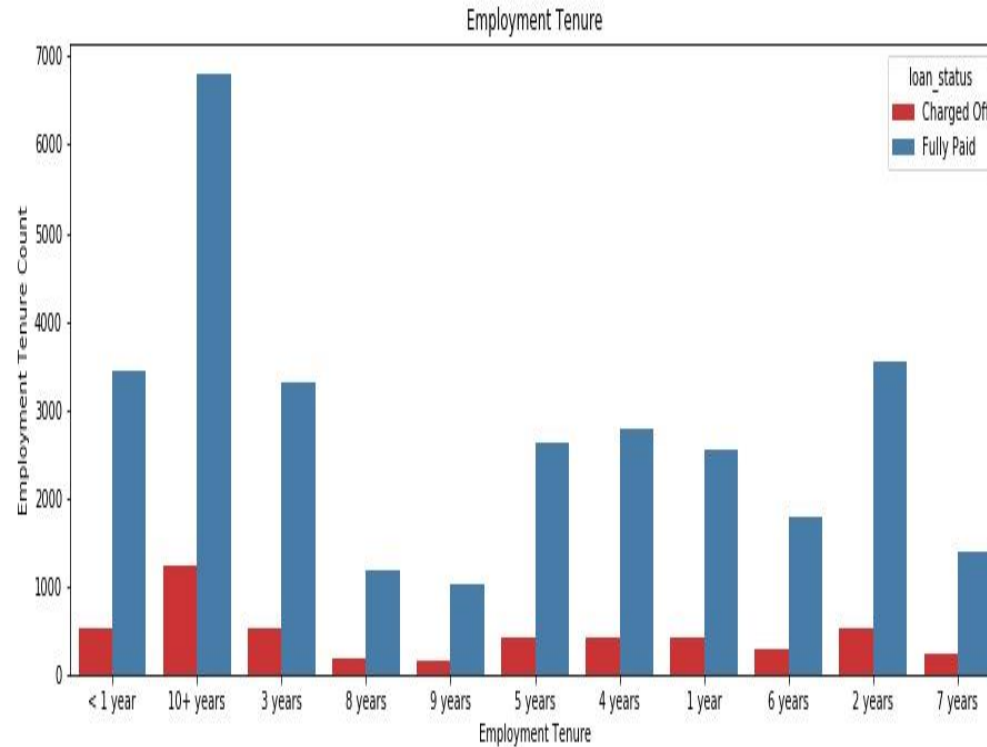
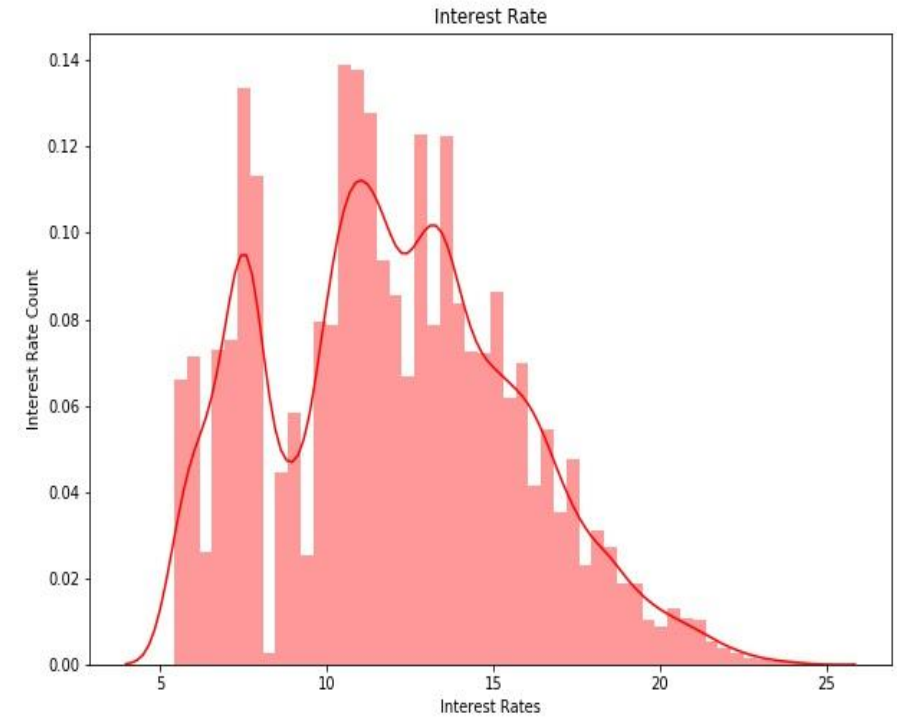# Exploratory Data Analysis

Loan Status
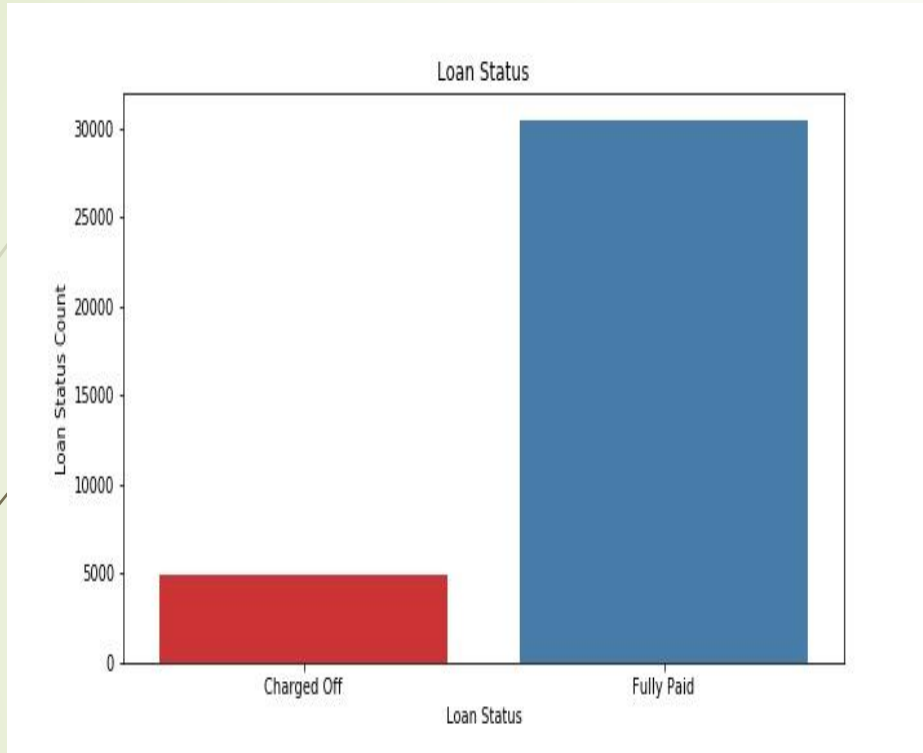
Application Type
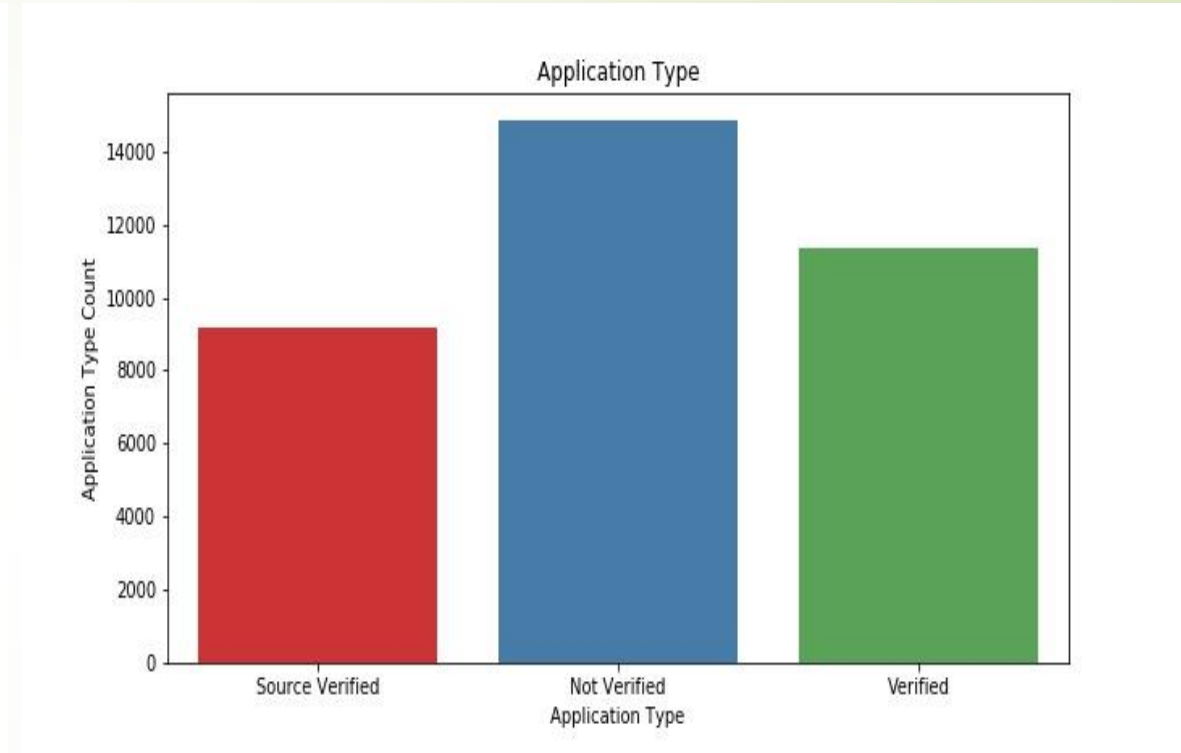
Observations:-
- About 86% applicants loan is fully paid.

- Verification status indicates if income was verified by LC, not verified, or if the income source was verified.
- As we can see above, Most of the applicant's income was not verified by the bank.
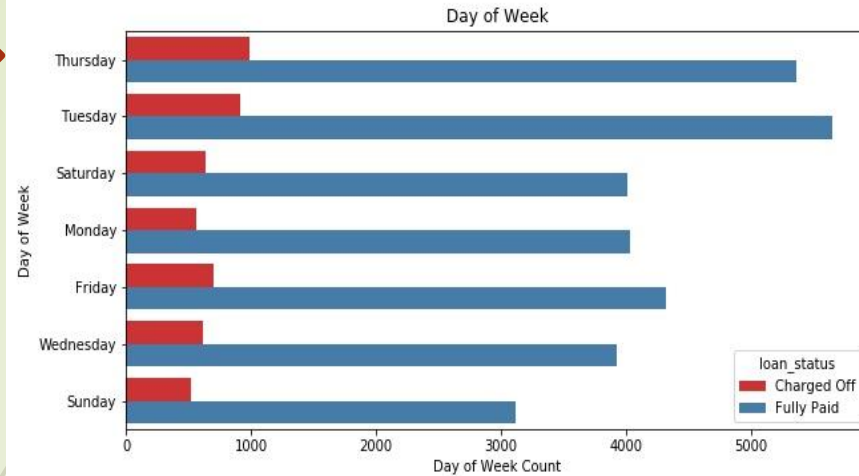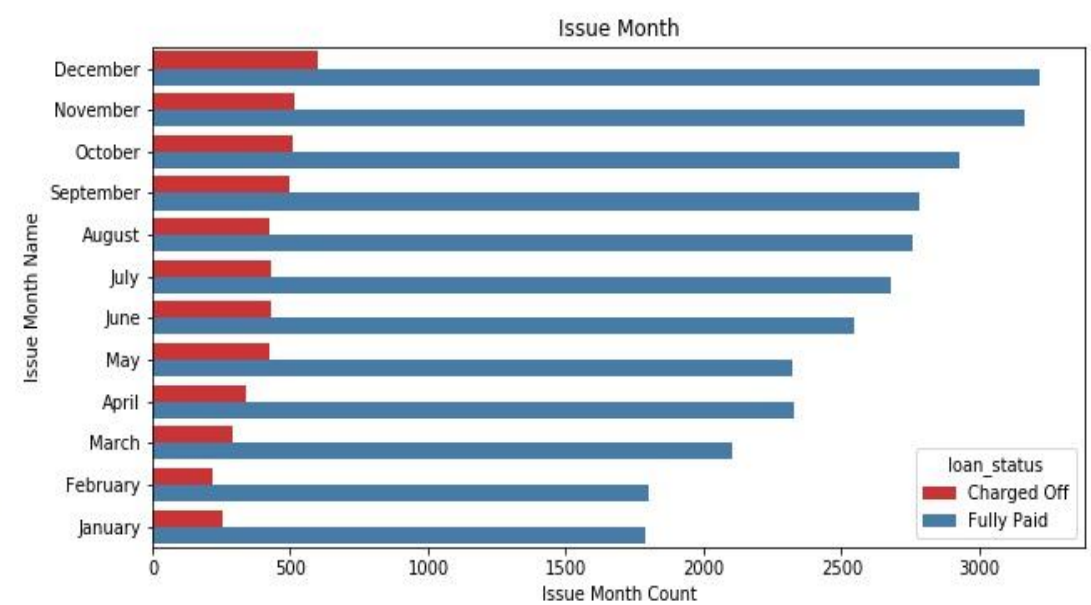
Employment Tenure/Experience

Interest Rate

Observations:-
- We cab observe that the people with 10+ years of Experience has successfully paid the complete loan.
- Also, the people with 9 years of experience failed to pay the loan(Charged-off)
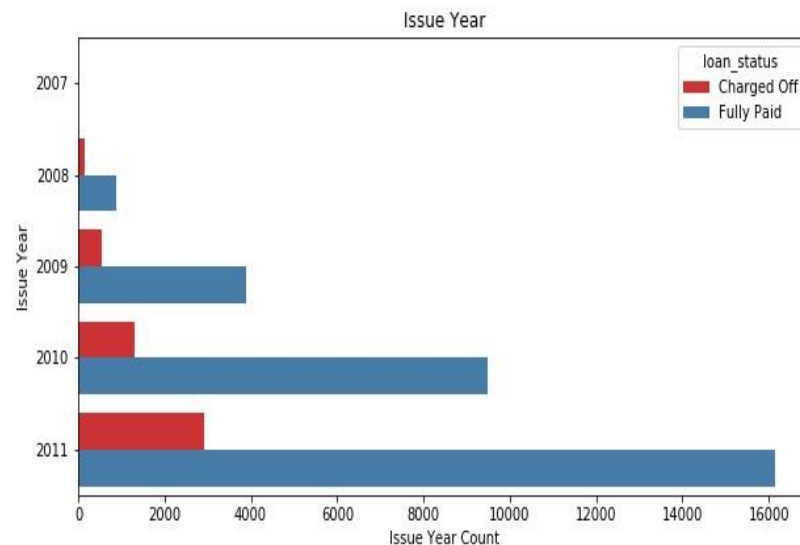
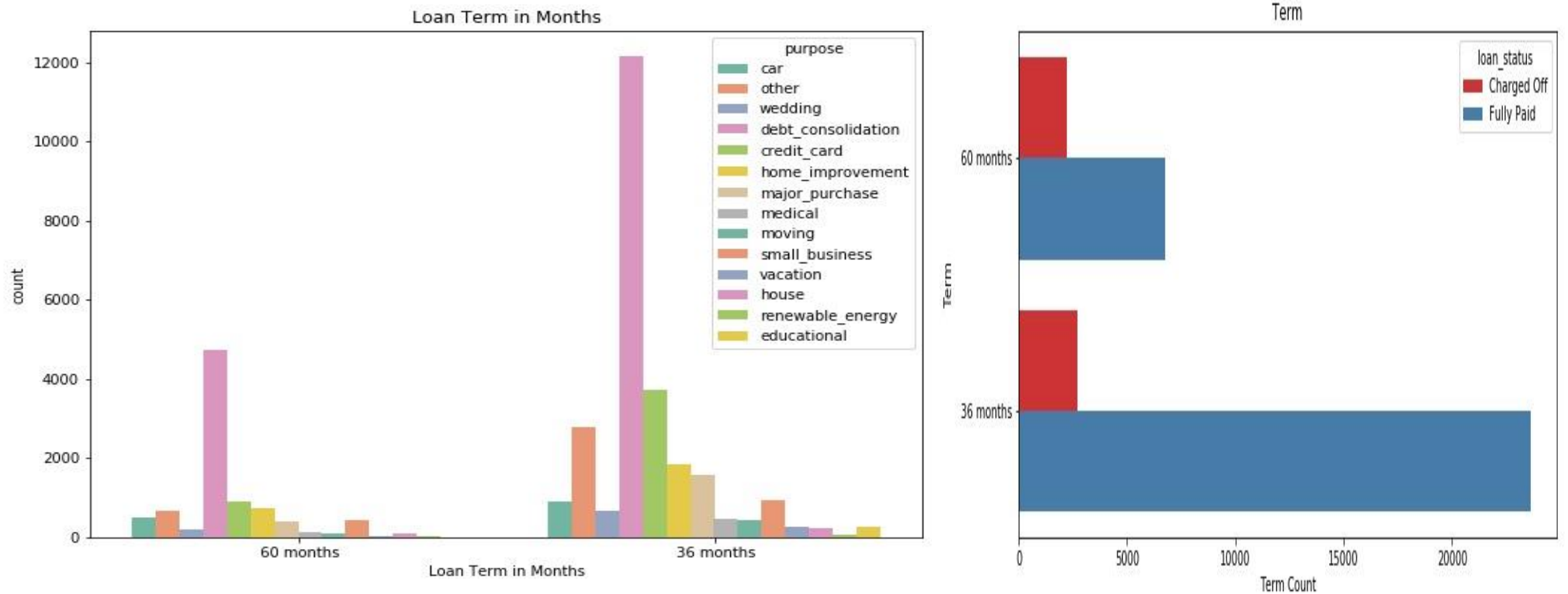- Average interest rate is between 10-15 percent

Issued Day

Issued Month

Issued Year

Observations-
- It can be seen that on Tuesdays, most of the loans are been issued.
- Most of loans are issued in the month of December, followed by November
- Most of the loans were issued in the year 2011, followed by 2010.
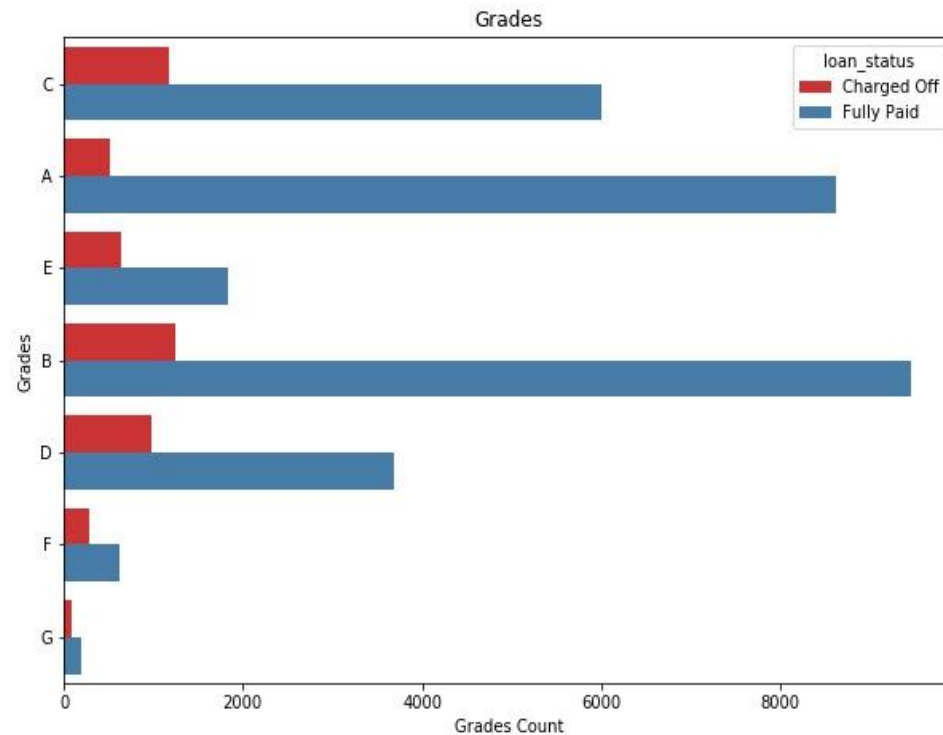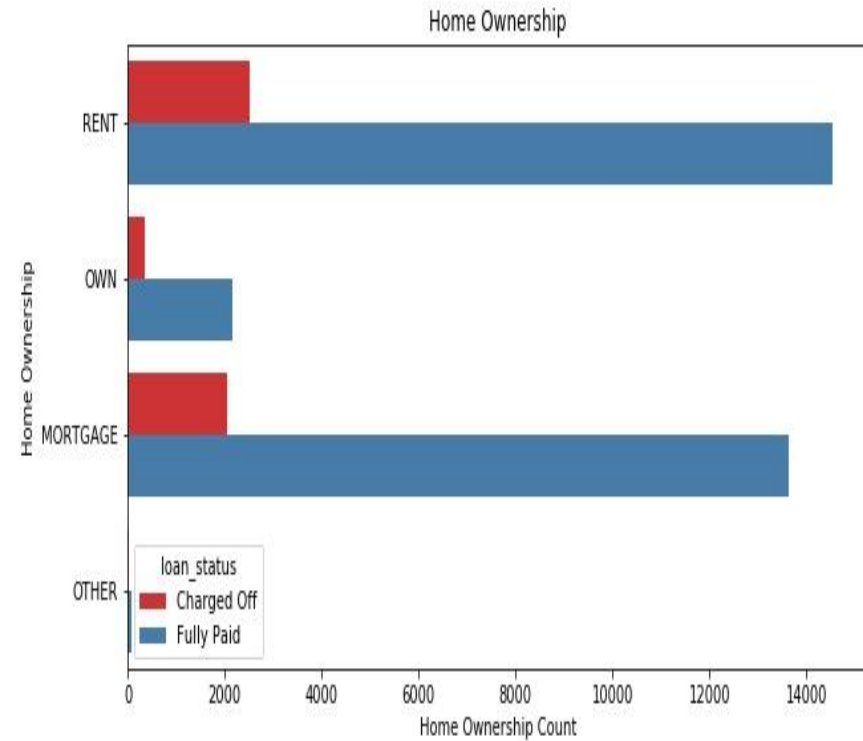
Loan Term Analysis

Observations:-
- Short term loans(36 months) were fully paid as compared to long term loans.
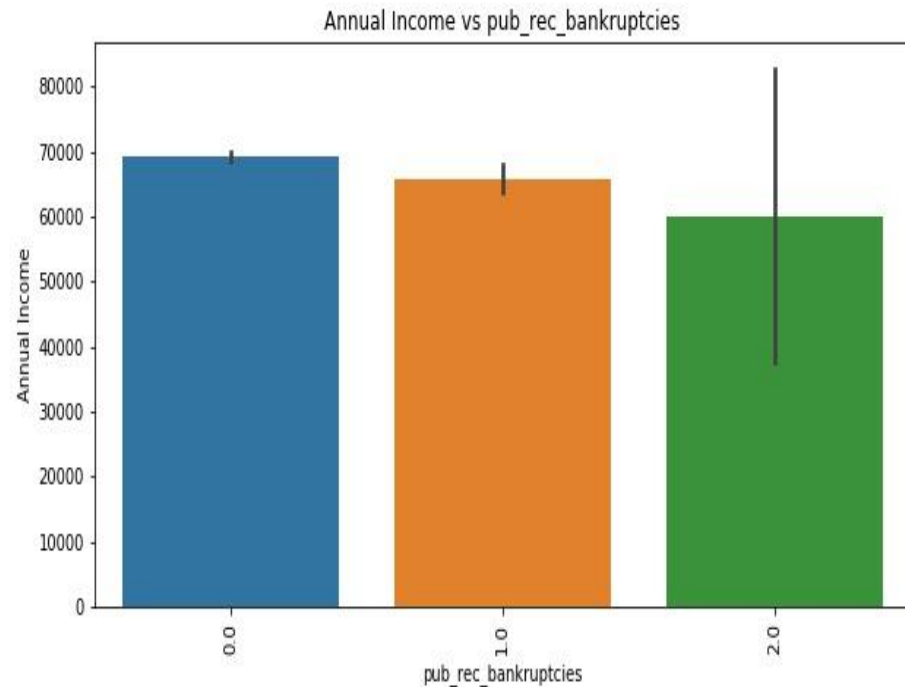- Among all, debt consolidation in both the tenure have high risk of getting default.

Grades

Observations:-
- Grades G, F, E and D have higher risk of loan defaulters.
- As the data contains "Other" as a category in Home Ownership, the analysis cant be concluded for this column.
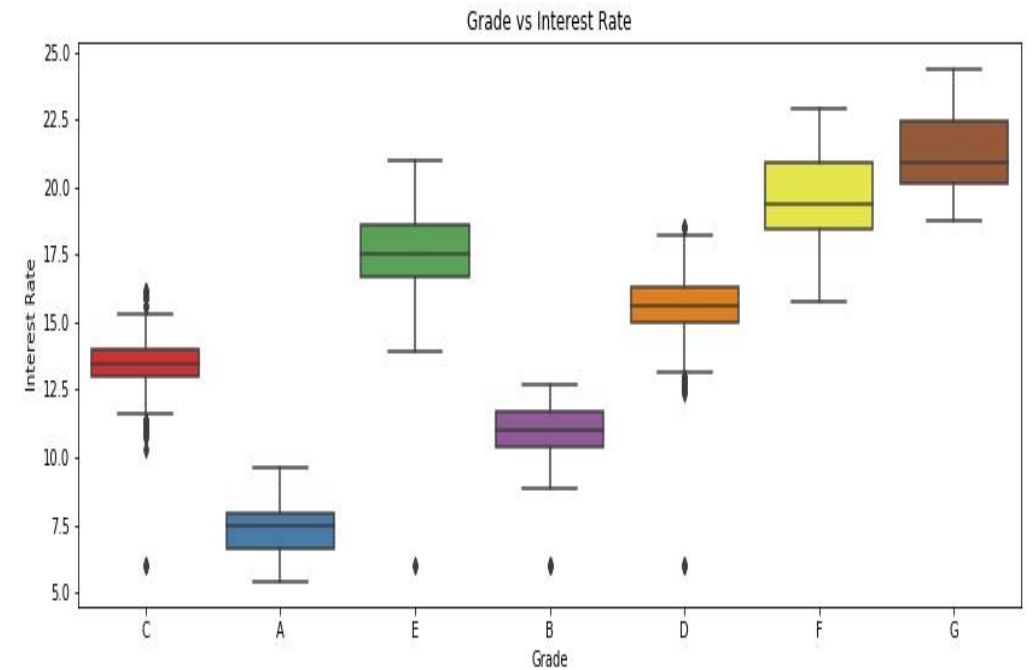
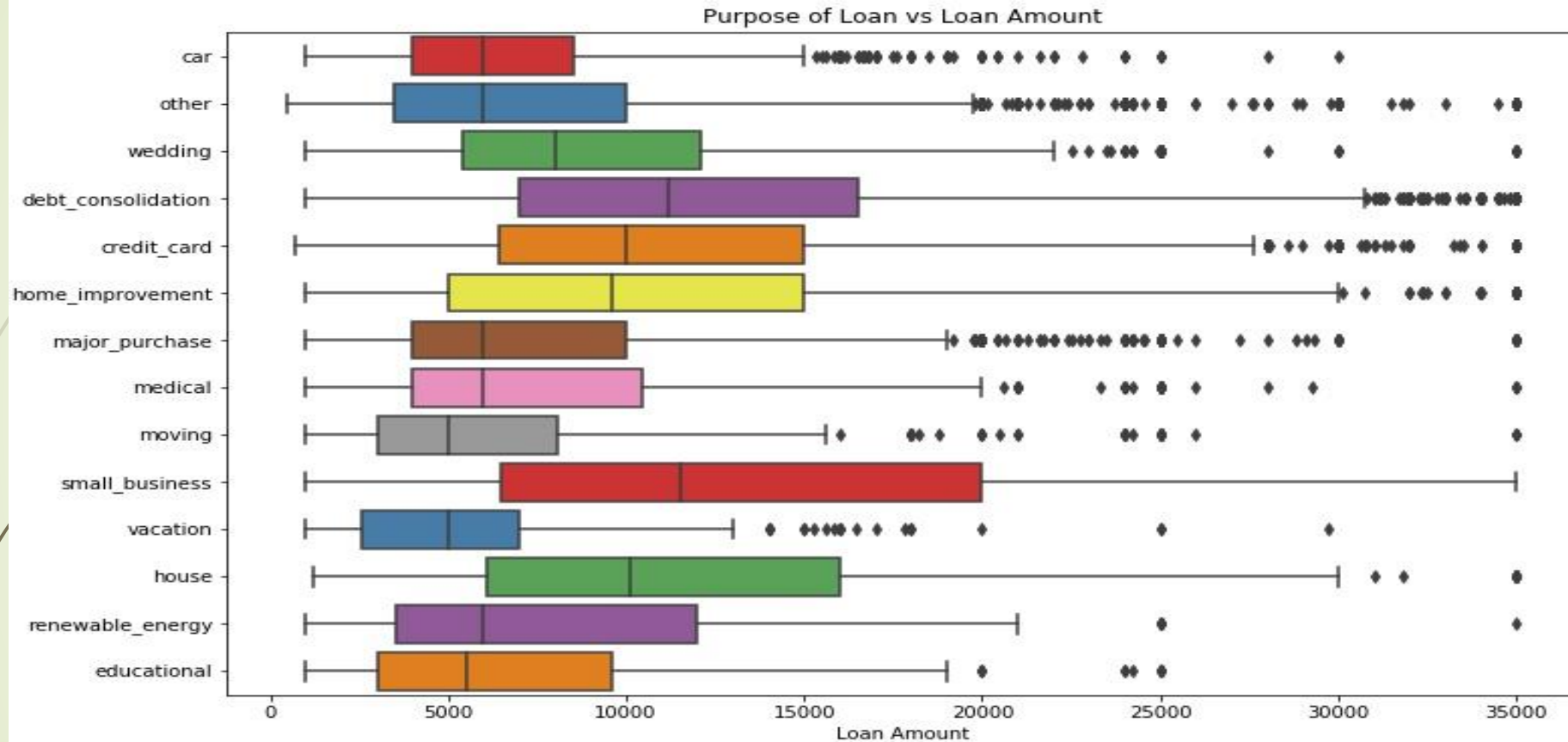Bivariate Analysis

Annual Income vs Public Record Bankrupcies

Grades vs Interest Rate

Observations:-
• The Loans with grades G,F has higher interest rates as compared to Grade A which has lowest Interest Rate

Purpose of Loan vs Loan Amount

- Loans taken for "small_business"," renewable_energy" and "educational" have risk than others.
- Loans taken for "small_business" has a large variance as compared to others, while lowest variance is for "vacation".

Heatmap for Correlation