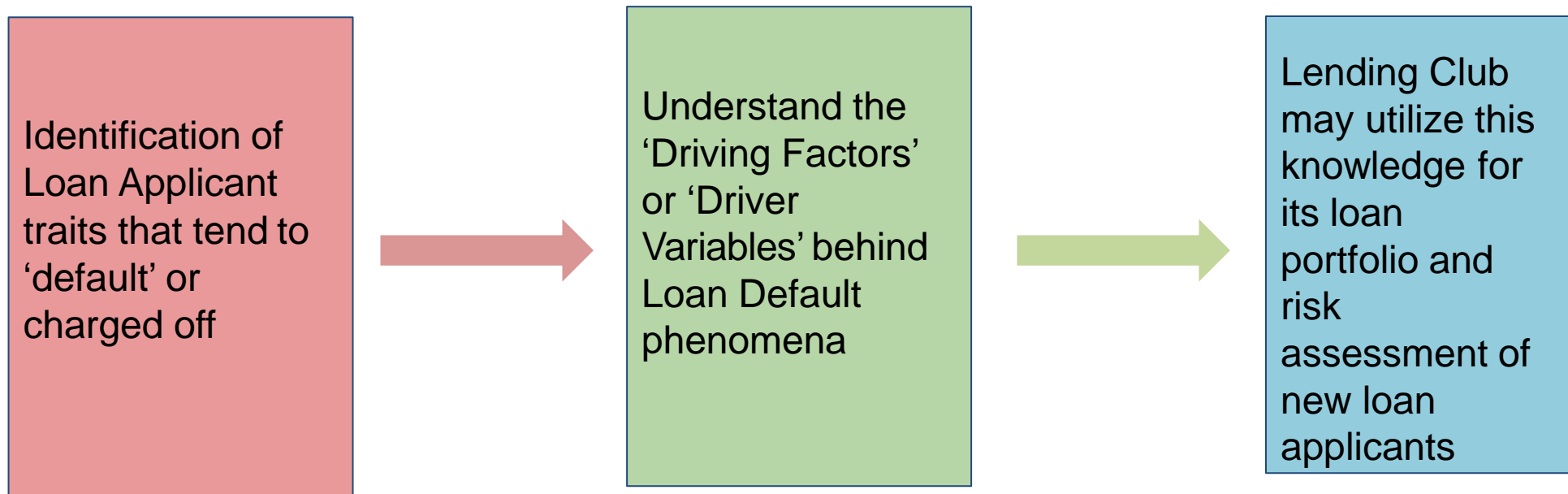# LENDING CLUB CASE STUDY SUBMISSION
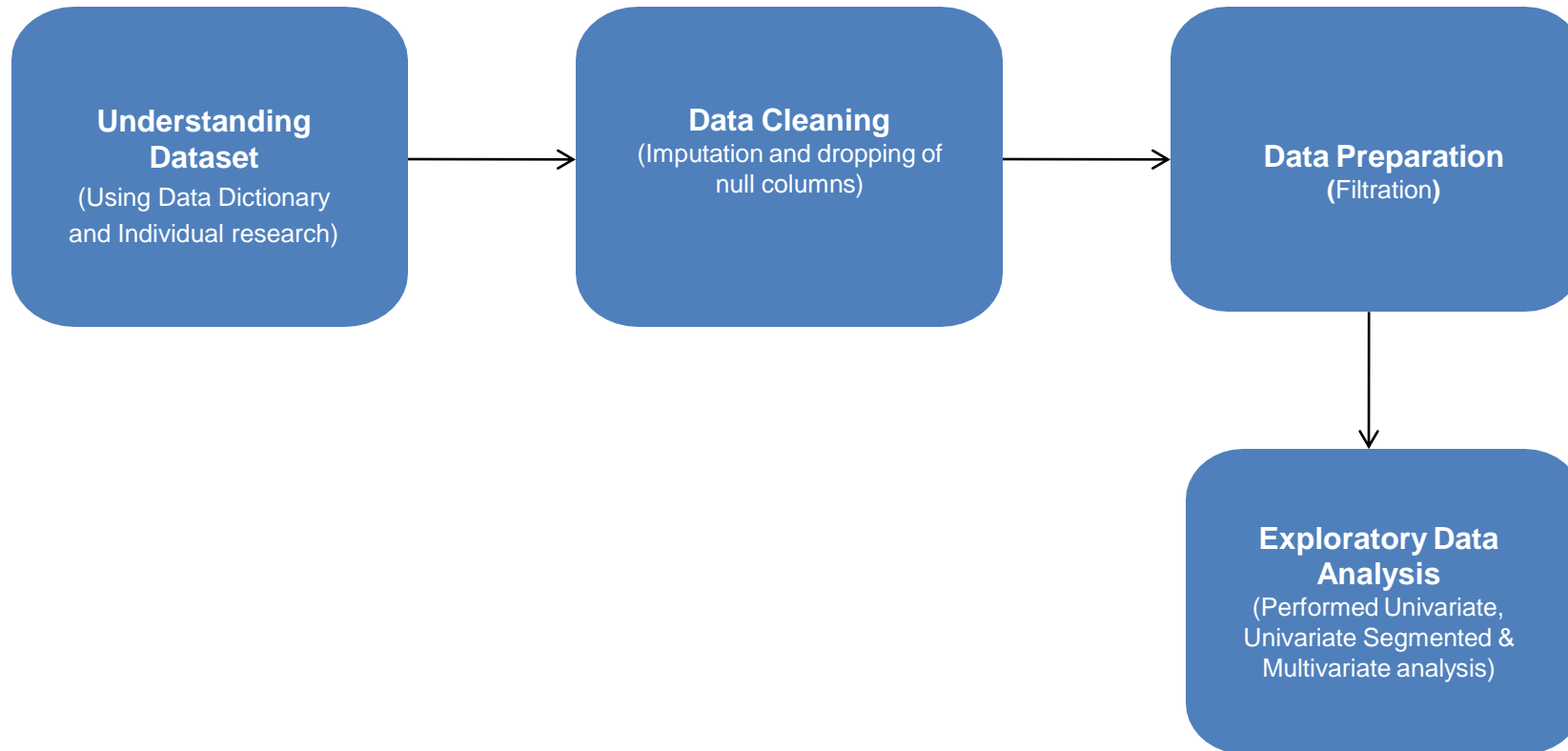
Group Members:

1. **Priya Aggarwal** (Priya27aggarwal@gmail.com)
2. **Shubham Gupta** (shubh.mech15@gmail.com)

Case Study Objectives

Identification of Loan Applicant traits that tend to 'default' or charged off

→

Understand the 'Driving Factors' or 'Driver Variables' behind Loan Default phenomena

→

Lending Club may utilize this knowledge for its loan portfolio and risk assessment of new loan applicants

Problem solving methodology

Data Cleaning Steps

- **Delete columns**: Drop columns with null values

- **Remove outliers**: Remove high and low values that would affect the results of analysis.

- **Missing values:** Treat missing values with appropriate approach i.e. mean for numeric values or mode for categorical values.

- **Duplicate data:** Remove identical rows, remove rows where some columns are identical.

- **Filter rows:** Filter by segment, filter by date period to get only the rows relevant to the analysis.

Analysis

- The essence of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.

- We performed **data cleaning and preparation** on the Loan dataset:
  - Imputed the NA values for all the variables
  - Created two new columns:
    - Profit and Loss column
    - Ratio of funded amount and annual income

- During **univariate analysis** we have created:
  - Histograms and Bar charts to check out the distribution of all the driver variables
  - Box plots to detect the Outliers
  - Performed the Multivariate analysis to understand how different variables interact with each other.
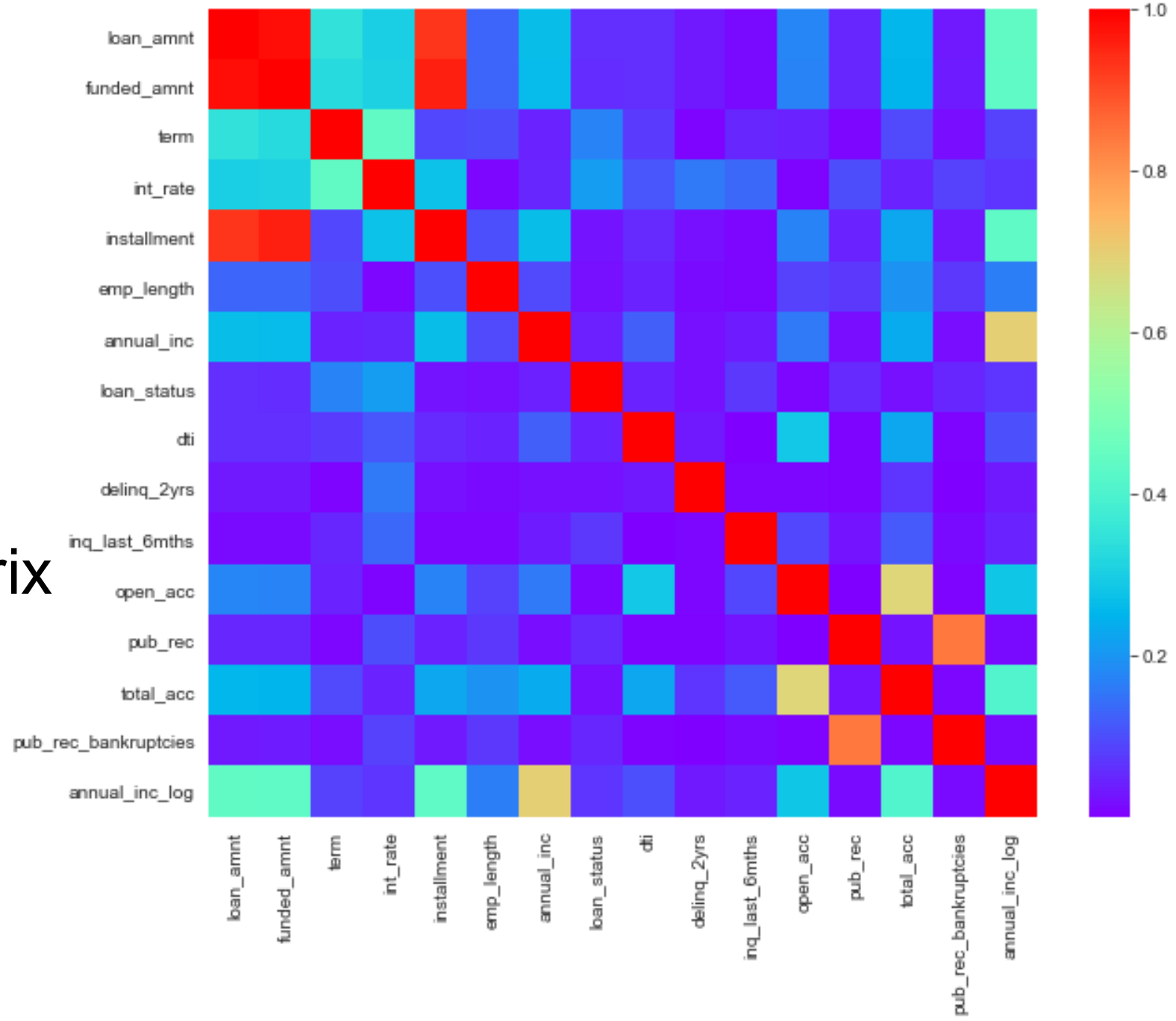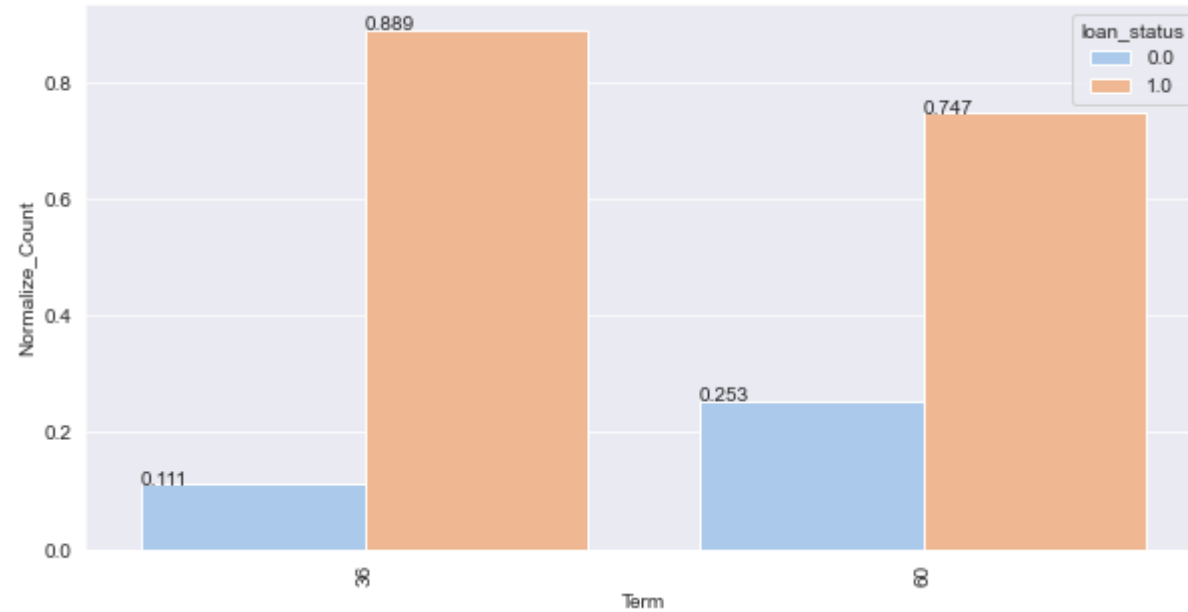
Loan Status Analysis

Loan Distribution Over The Duration
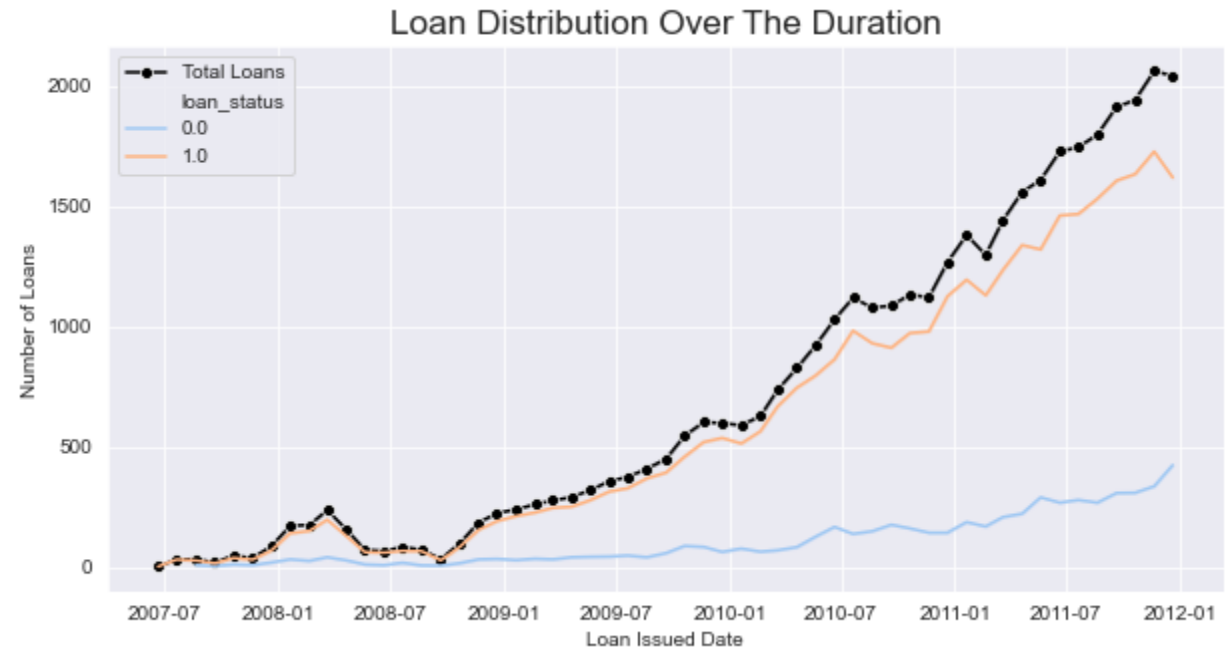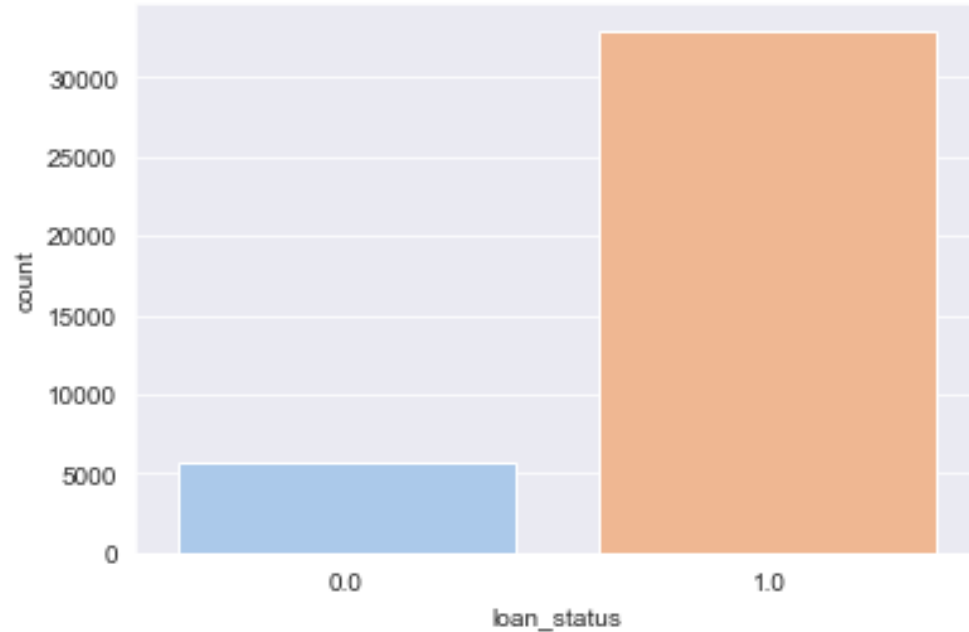
Observations:

1. Most of the loans are Fully Paid.
2. % of loan as charged off or defaulters is very low.
3. The loan has been increasing exponentially over the years as seen in Loan Distribution figure.
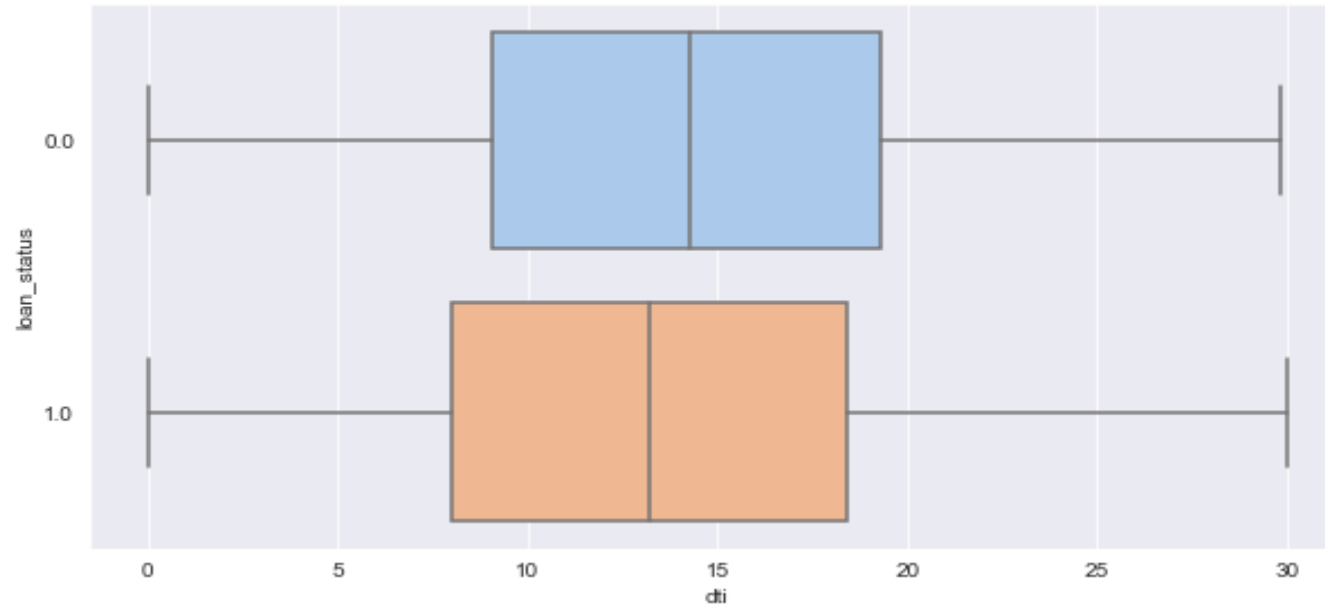
Correlation matrix

Loan Term Analysis

Observation:
1. Loans with 5 year(60 months) period are more likely to default as compare to 3 year(36 months) period loan
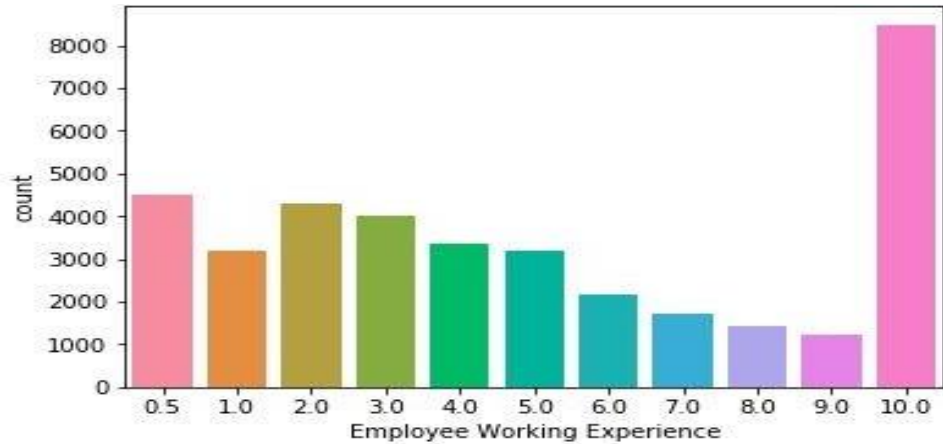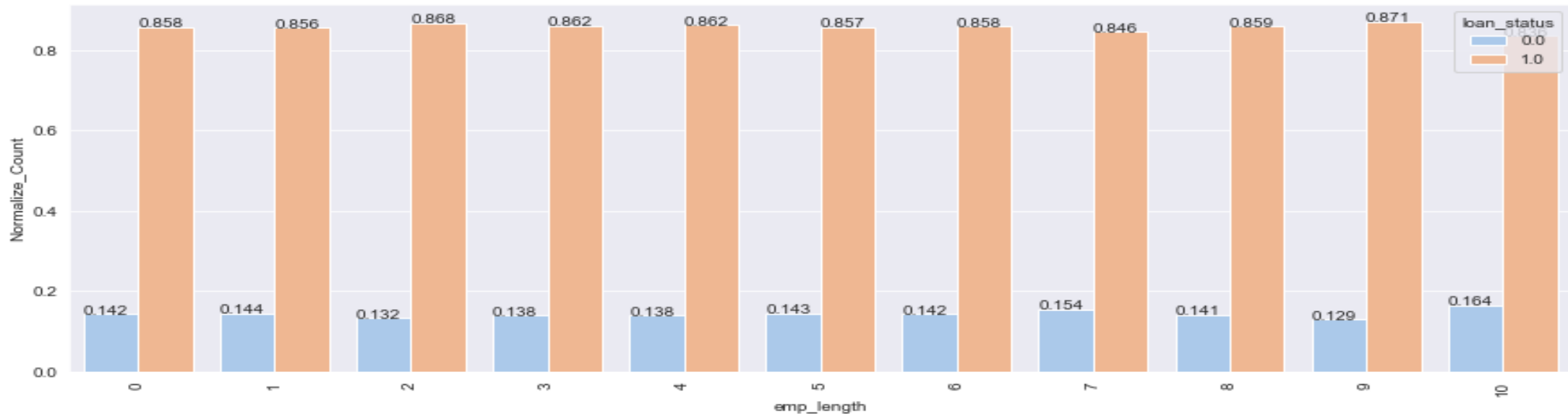
Loan DTI Analysis

Observation:

Higher the dti means higher the risk of default.
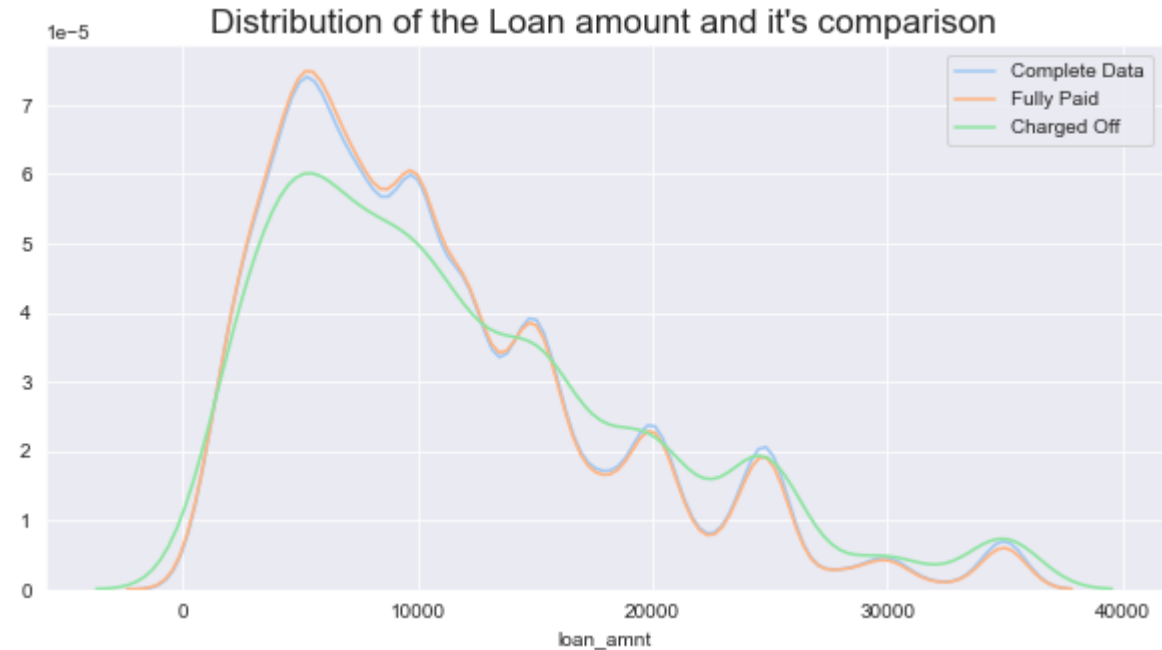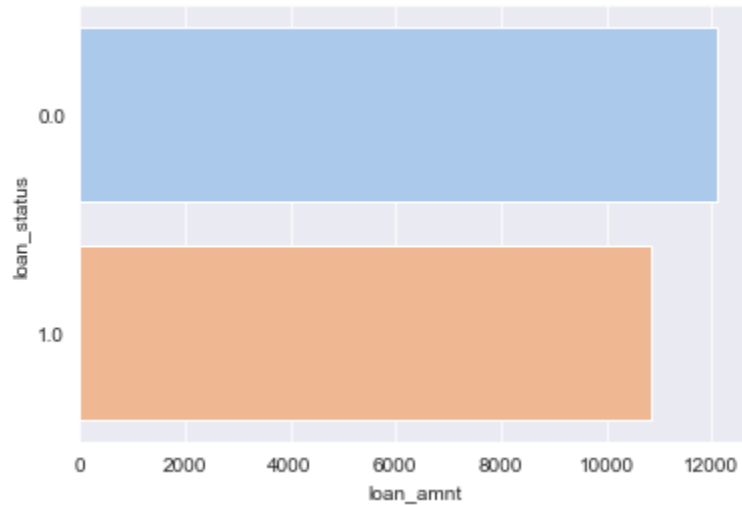
Loan Applicants work Experience Analysis

Observation:

1. Majority of employees applying for the loan have more than 10 years of experience
2. Tendency of person to default the loan with 10 years of experience is also high. So company need to be careful when granting loan.
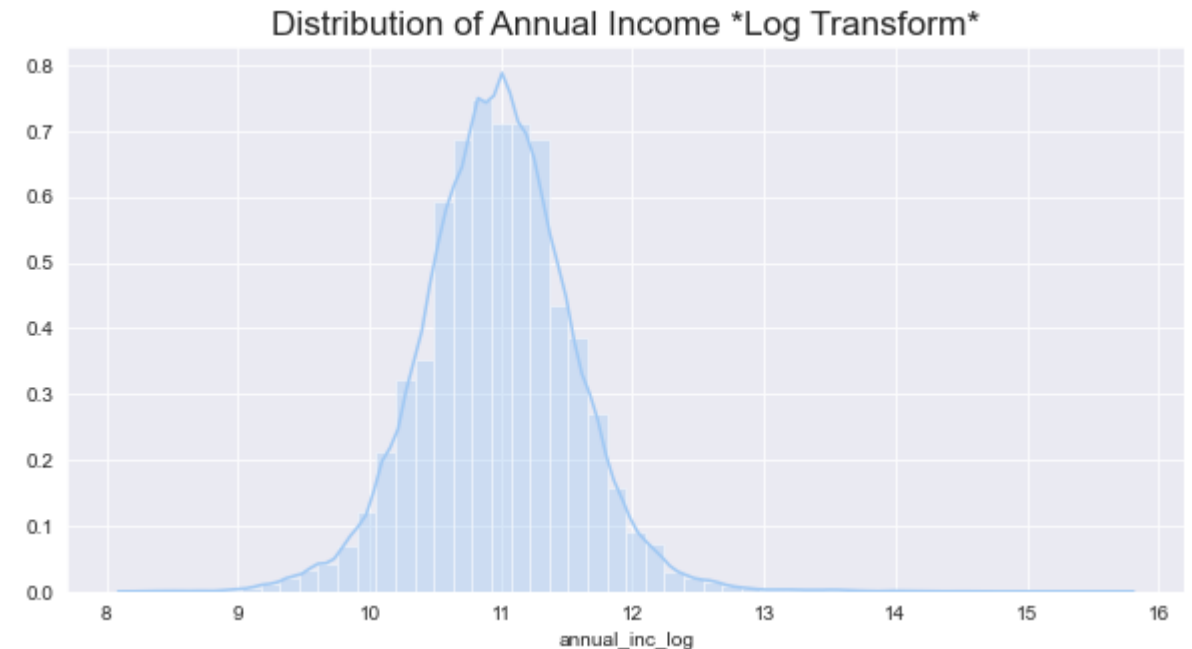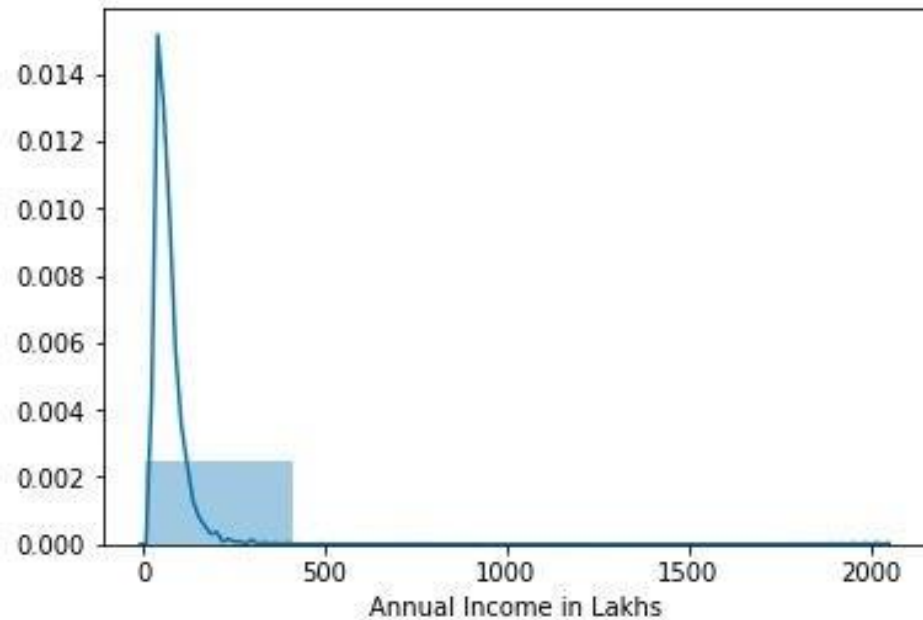
Loan Amount Analysis

Observation:

1. Loan amount is left skewed.
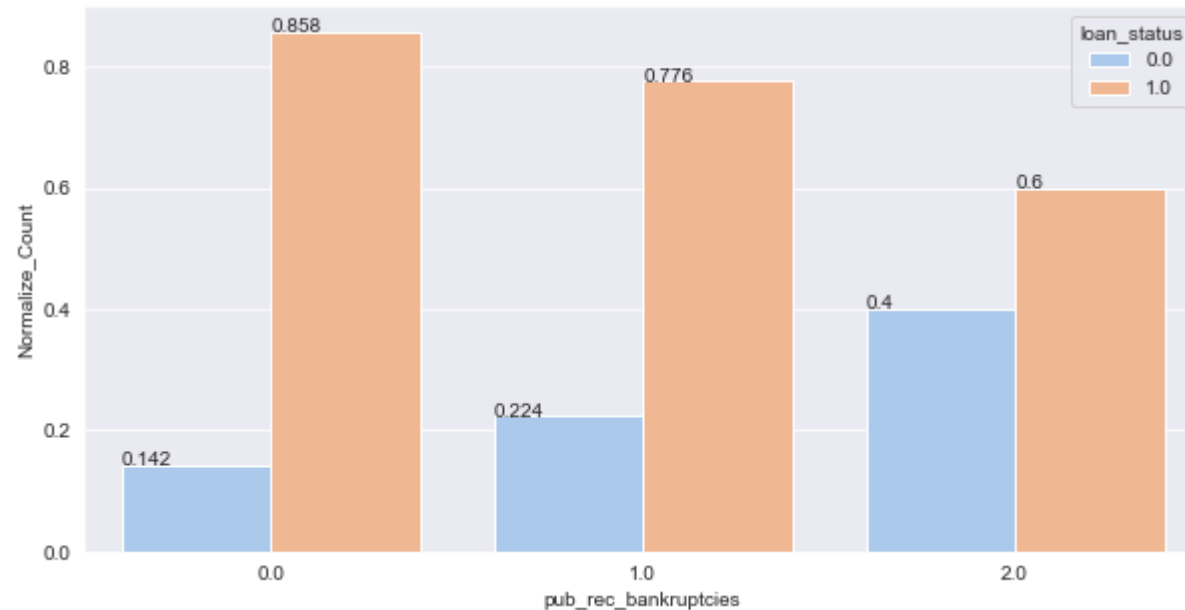2. Both type of loans (Fully Paid & Charged Off) have same distribution

Income Analysis for the loan applicants

Distribution of Annual Income *Log Transform*

Observation:

1. Most of the applicants have income between 0 to 5 lakhs.
2. There are two applicants having income above 30 lakhs
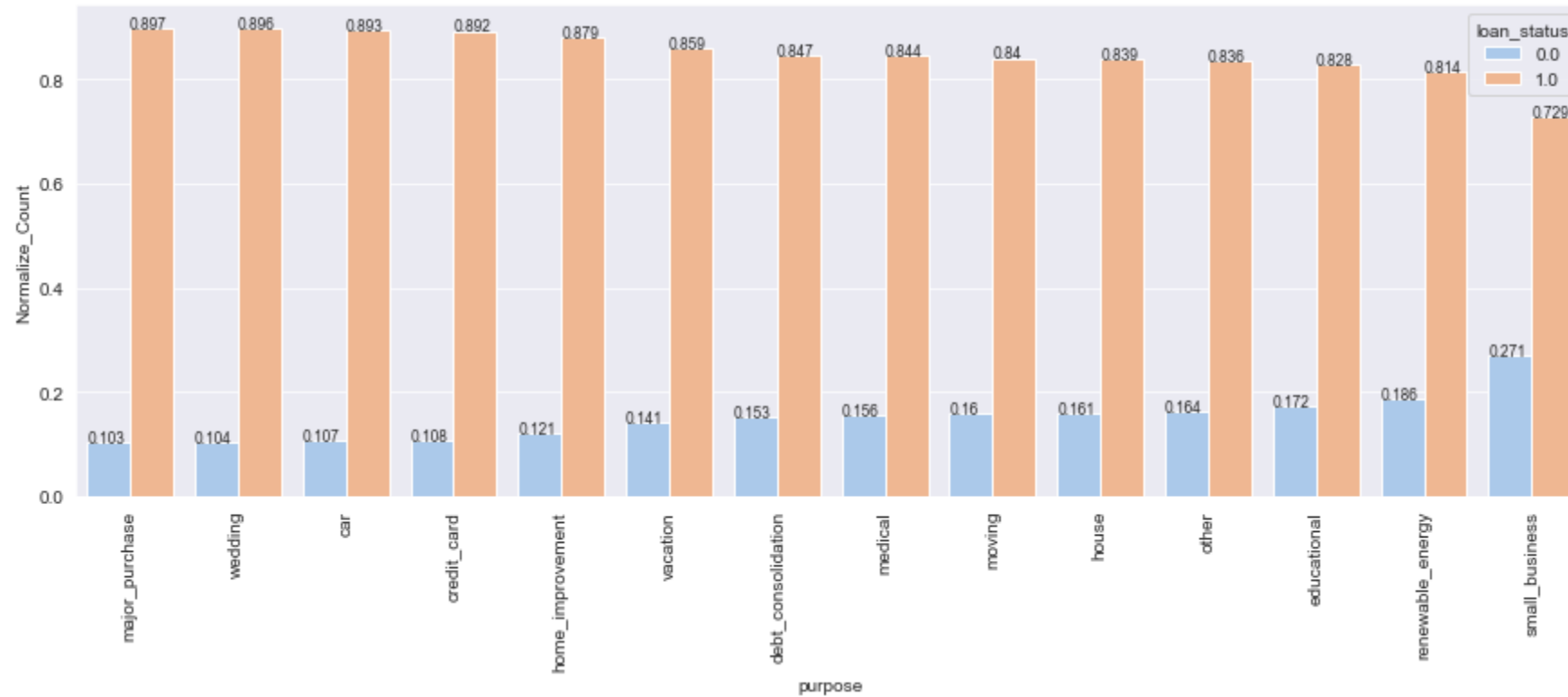3. For ease we used log transform distribution

Analysis for bankruptcies

Observation:
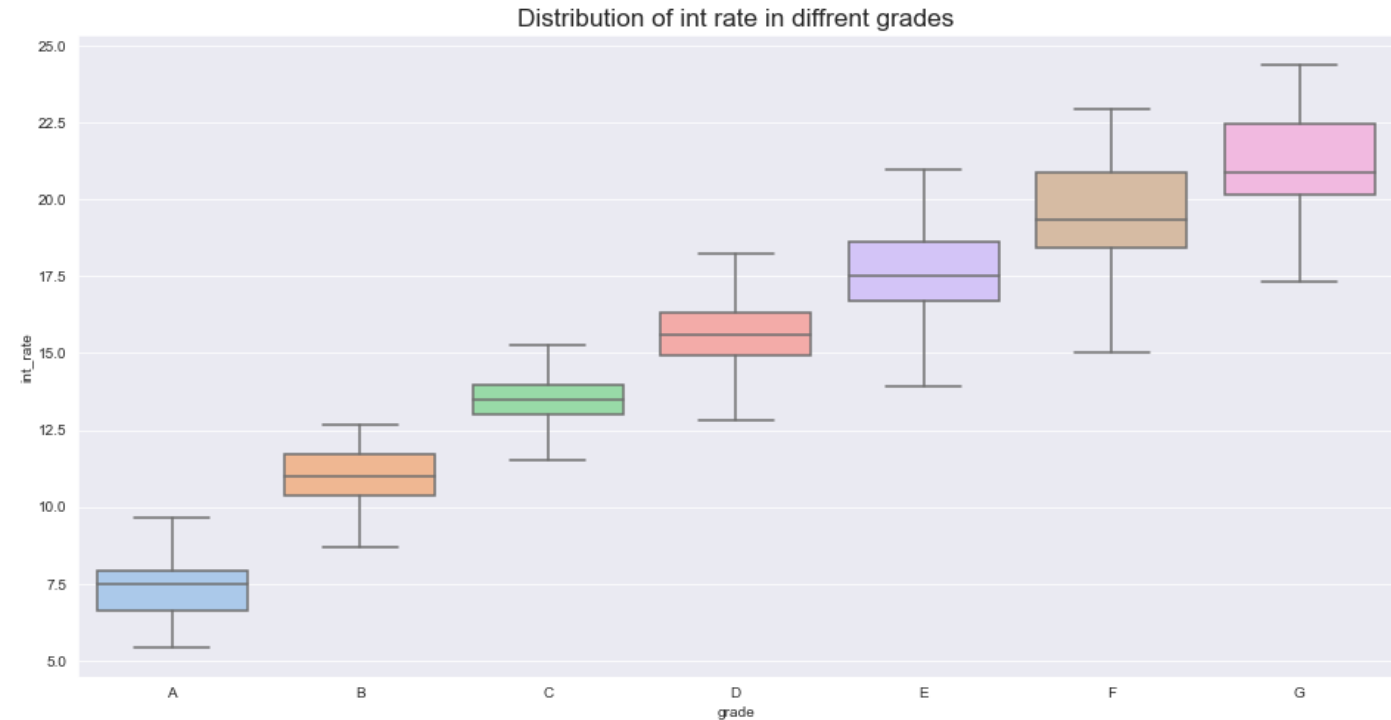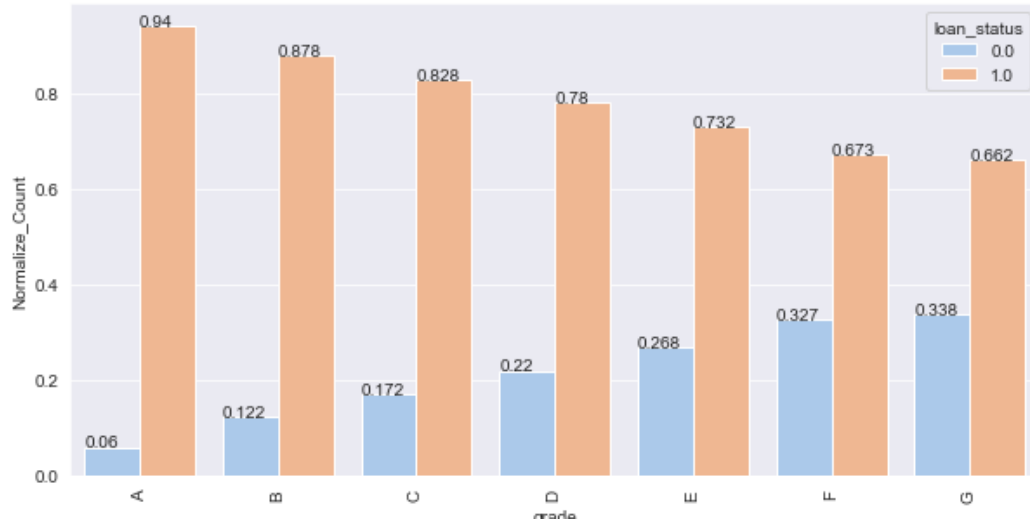1. If a person has record of bankruptcies, He is most likely to default

Analysis for Loan Purpose

Observation:
1. loan for small business are most likely to get default

Analysis: Interest Rate Vs Grade

Distribution of int rate in diffrent grades

Observation:
1. Lower the Grade Higher the interest rate
2. Lower the Grade (A; Higher & G; Lower) higher the chance of defaulting of loan

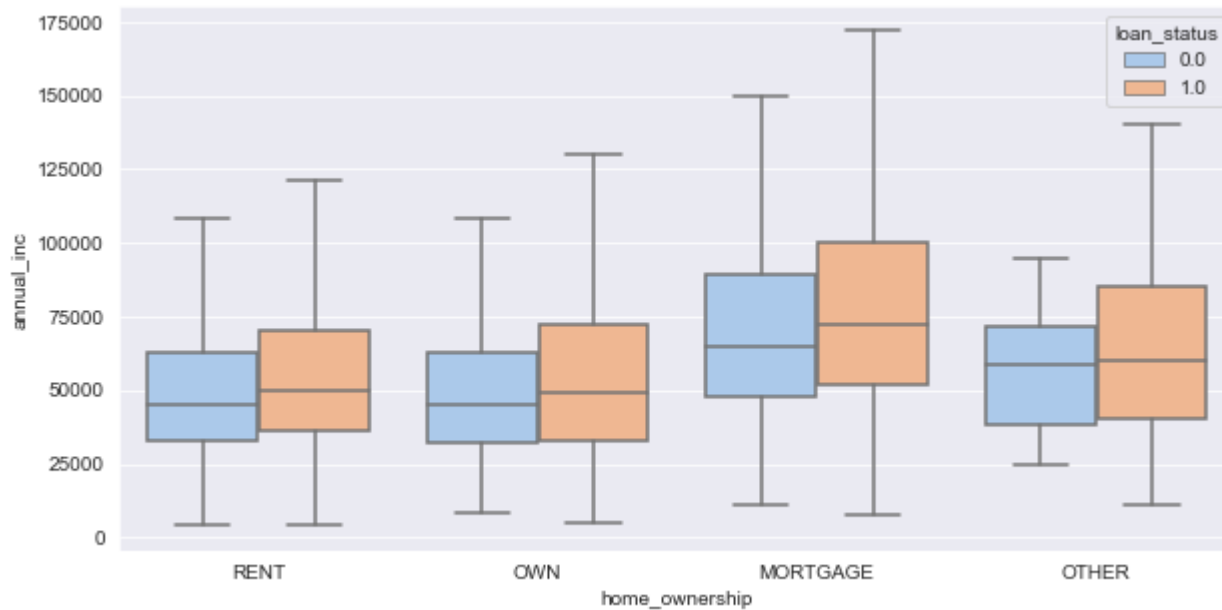# Loan Amount vs Annual Income

Observation:
1. There is a relationship between loan amount and annual income

Annual income Vs home ownership Vs Loan Status

Observation:
In individual group of home_ownership defaulters income range is less.
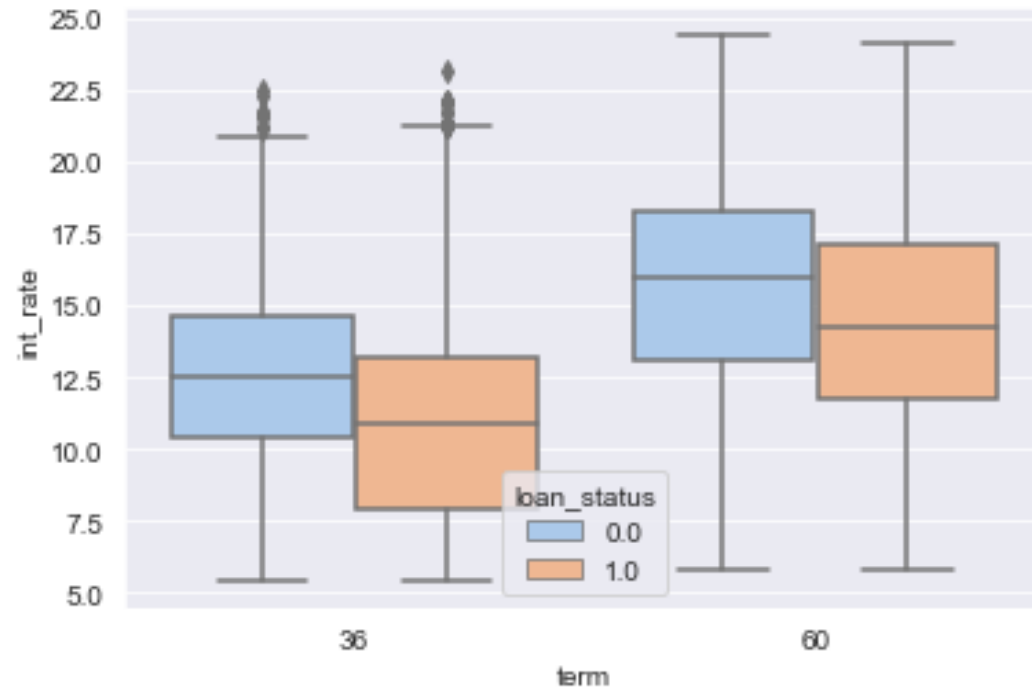
Annual income Vs Verification Vs Loan Status

Observation:
Higher loans even with verified status are defaulter. So, Higher the loan amount has large risk of defaulting the loan.
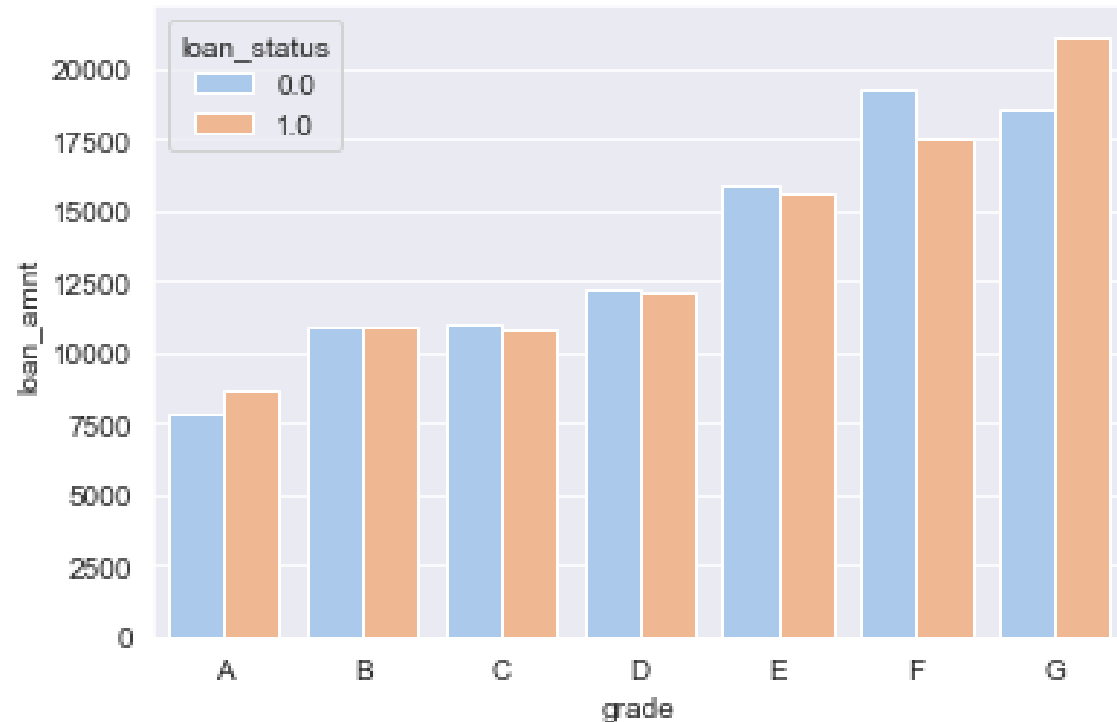
Interest Rate Vs Term Vs Loan Status

Observation:
1. Interest rate depends upon term. Higher the term value higher the interest rate.
2. As in previous result of analysis of interest rate and term higher the terms & interest rate, higher the chance of defaulting the loan
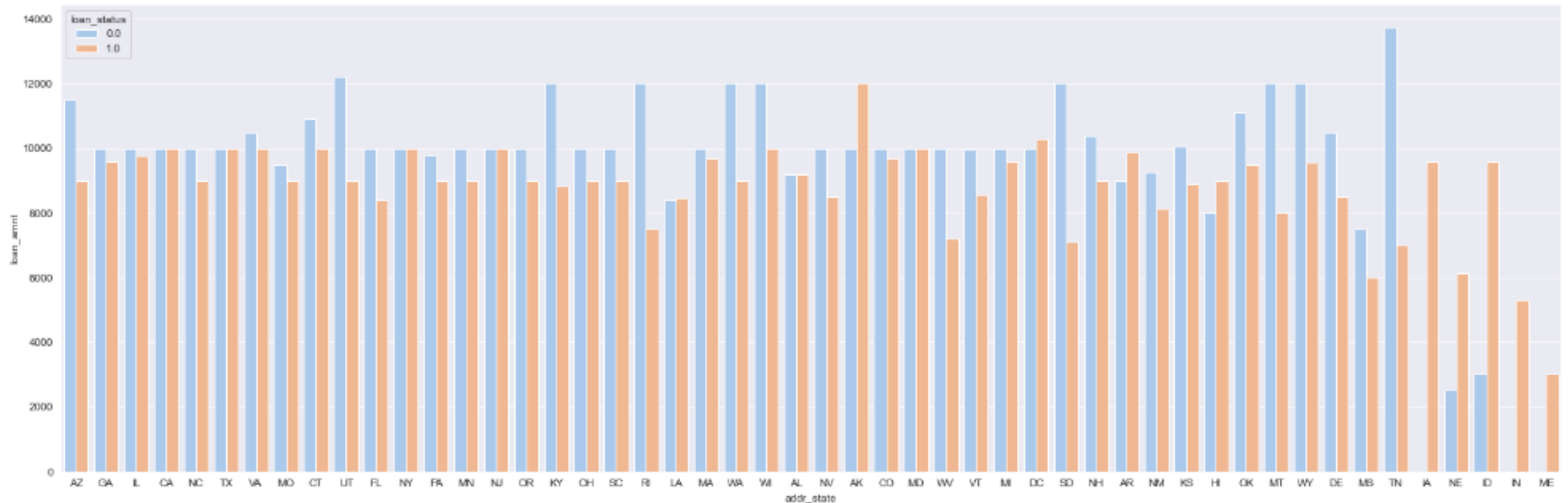
Analysis: Loan Amount Vs Grade Vs Loan Status

Observation:

Higher the loan amount with low grade i.e. F & G having chances for default. So, this combination could defaulting the loan.
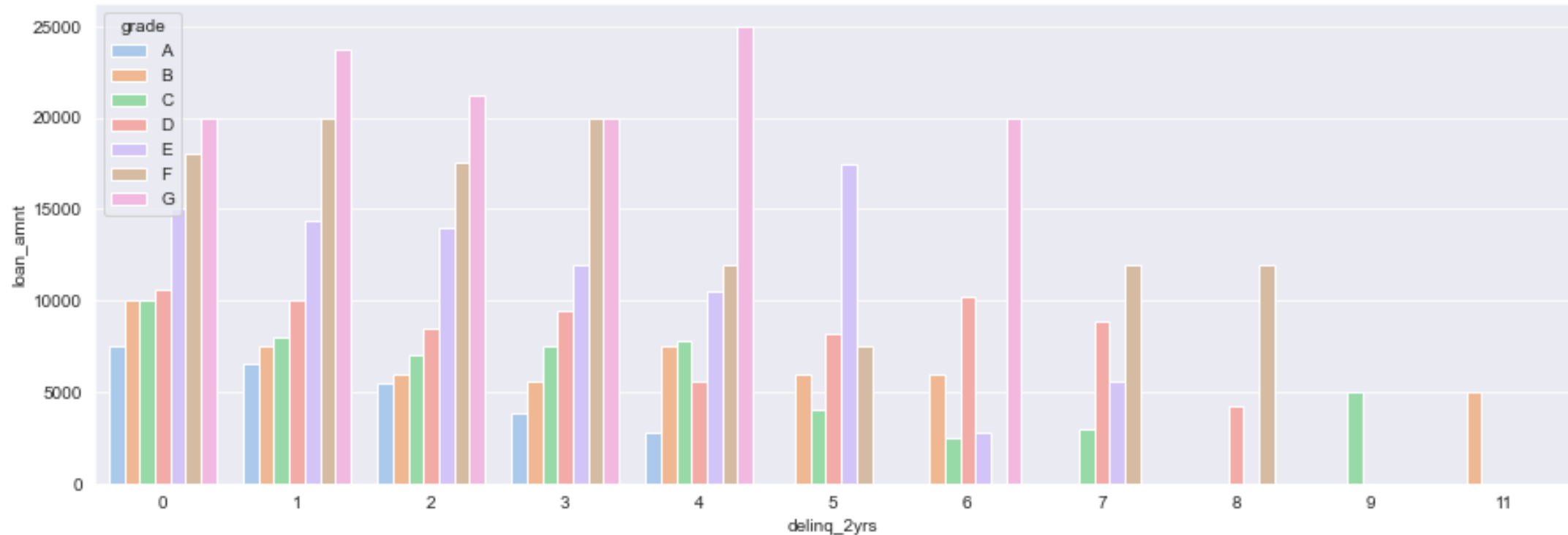
Loan Amount Vs Address State Vs Loan Status

Observation:

WY,MT,TN states have the highest average loan amount that was defaulting the loan.

Loan Amount Vs Delinq 2 yearsVs Loan Status

Observation:

Higher the delinq_2yrs > 1 with higher loan amount has lower the grades. Which shows that this combination could be risky.

Recommendations

Observation:

1. Lower grade loans have high tendency to default.

2. Loans having higher interest rate have more defaulters. Check the background of applicant thoroughly if interest rate is high.

3. Extra check must be done for the applicants having higher dti. There could be higher the risk of default.

4. Loan for small business are most likely to get default.

5. Less annual income with high loan amount is not recommended.

6. Loans with 5 year(60 months) period are more likely to default as compare to 3 year(36 months) period loan

7. Higher the delinq_2yrs > 1 with higher loan amount has lower the grades. Which shows that this combination could be risky.

Recommendation :
We can create score on the base of probability of defaulting based on driving factors,
if there is high probability of defaulting, we can reject the loan.