

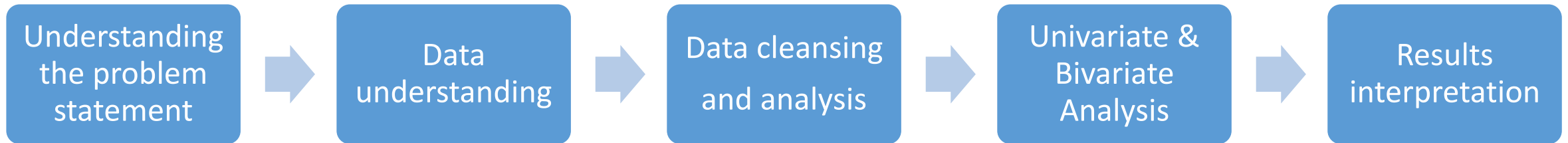
# Lending Club Case Study Submission

Name: Subham Kumar, Karthik R A

## Objective:

1. To understand the driving factors behind loan default.
2. Also identify the risky loan applicants who are most likely to default the loan.

Approach used for given problem statement:

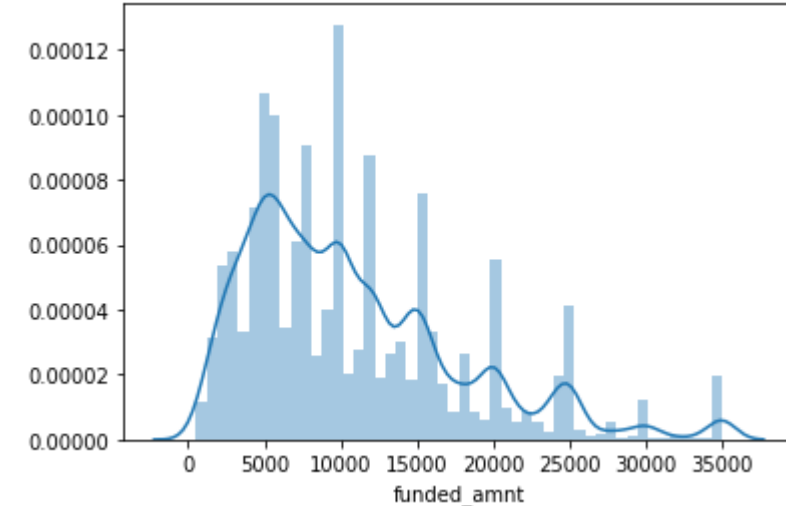
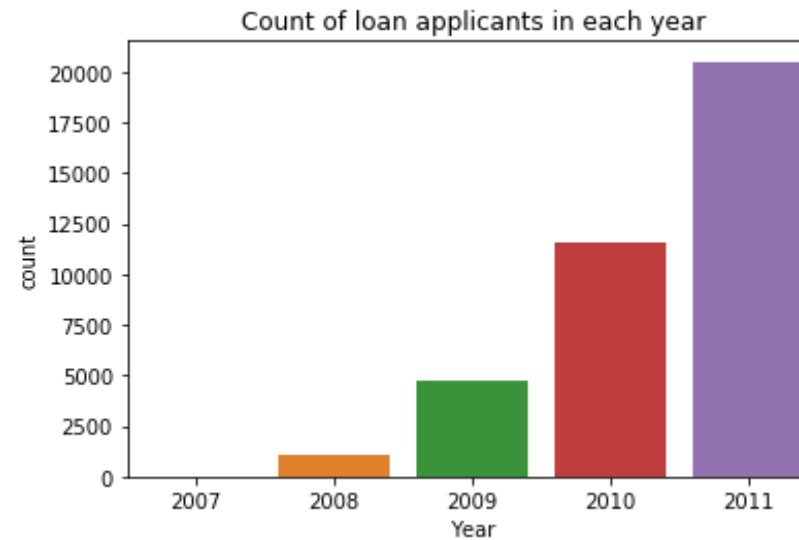
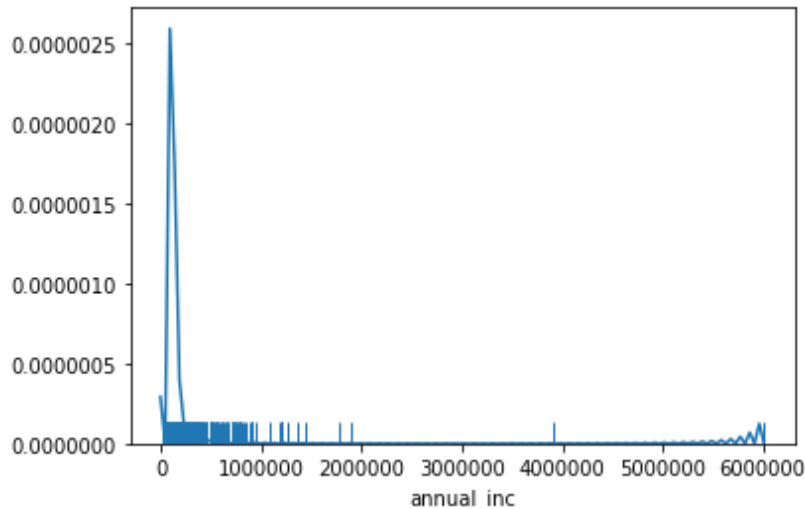
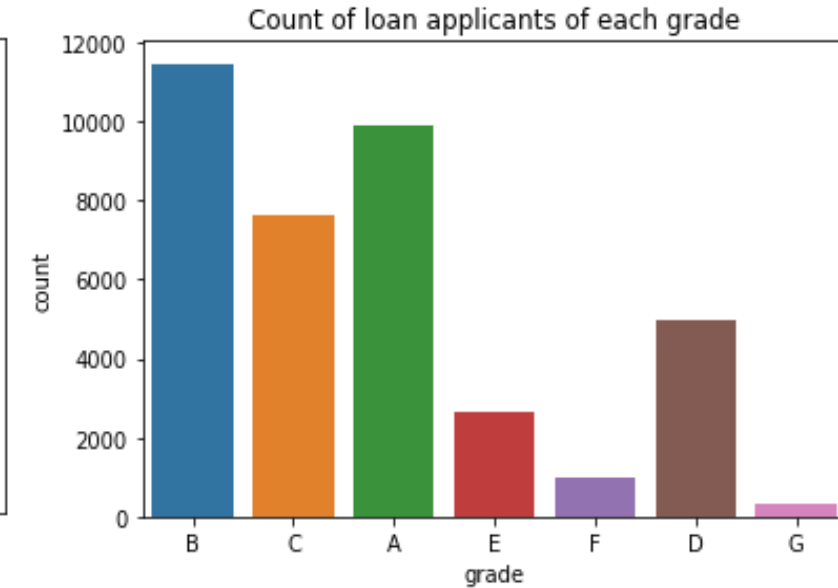
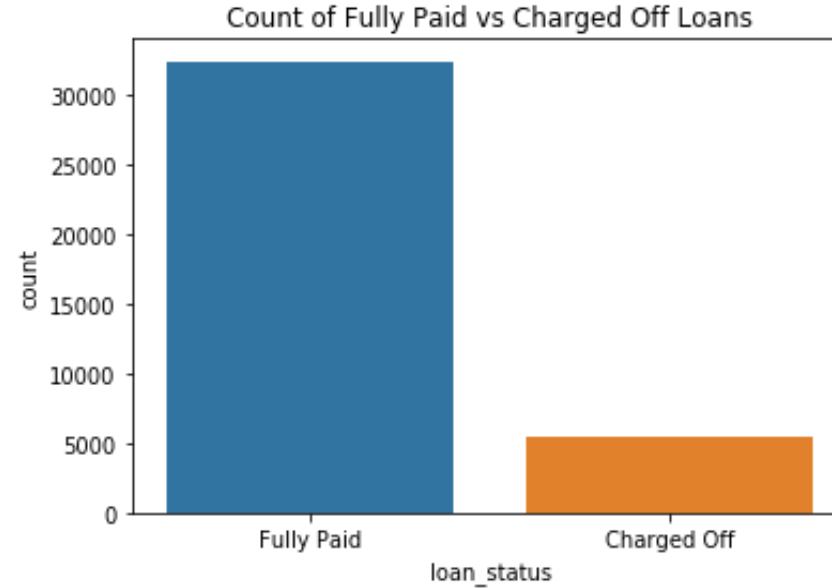


# Data Cleaning and Analysis

1. Read the dataset and remove the columns which are not required for further analysis (like customer behaviour variables).
2. Calculate the percentage of null values in different columns.
3. Convert the emp\_length column to numeric datatype, by extracting only numbers from this column.
4. Missing values in emp\_length column are assigned with median value of employee length.
5. Add a new column 'Year' to extract the year in which the loan was taken. The value in this column is derived from 'issue\_d' column.
6. Remove rows where loan status is current, as these rows won't provide us any insight of loan defaulting factors.
7. Once the dataset is cleansed, identify the unique values in each of the columns to see its variations. This will help to understand whether the values in a column is categorical / continuous.

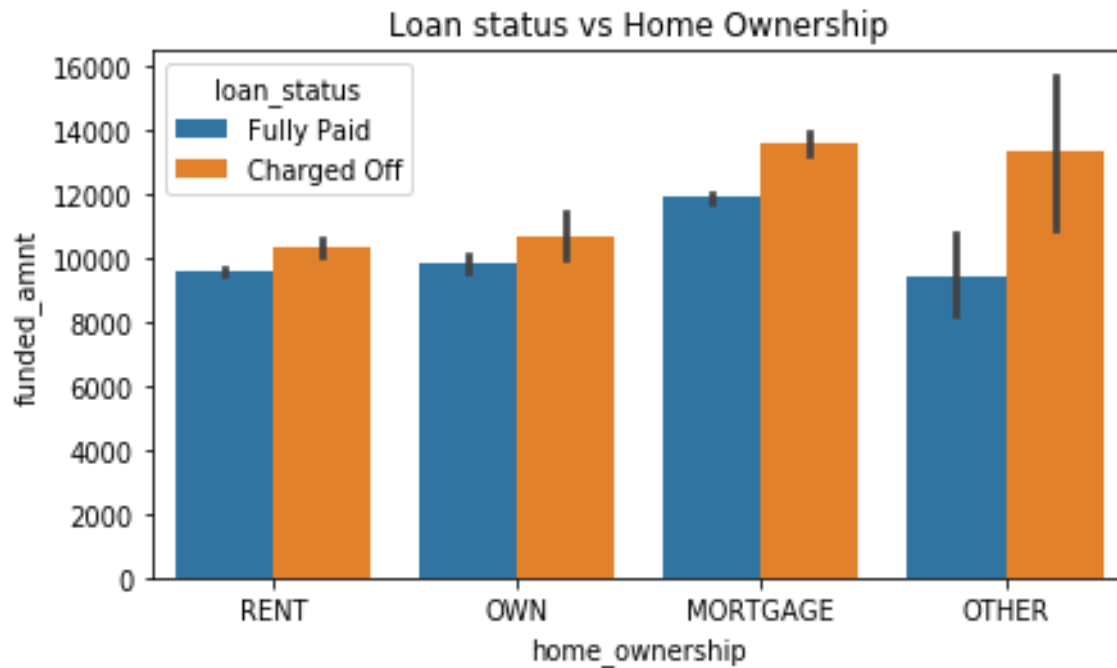
## Univariate Analysis Results:

1. Around 14.5% of people defaulted out of total people who applied for loan.
2. Majority people have Grades A & B who are taking loans.
3. Loan granted has increasing trend Year on Year basis.
4. Mostly loan applicants lie between 40-50k income bracket.
5. Applied loan amount for peaks around 5k and sees a downward trend henceforth.

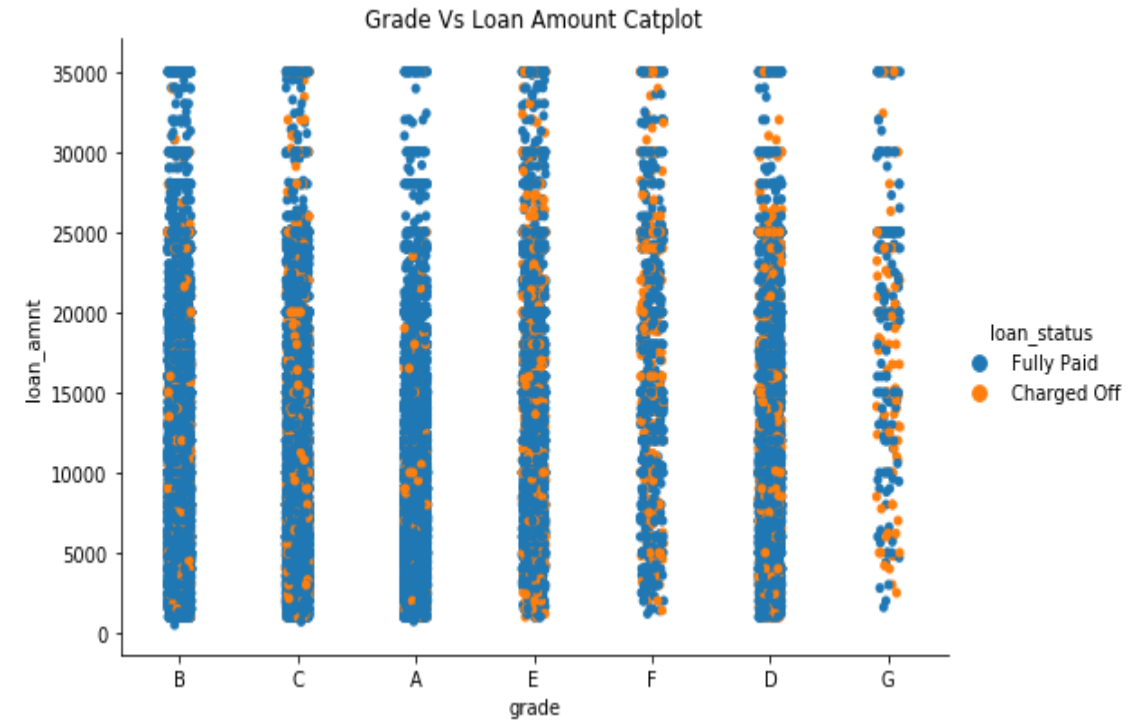


## Bivariate Analysis Results:

1. From the below plot, it is evident that people from OTHER home\_ownership category are most likely to default the loan. The funded amount to these loan applicants must be capped below 8000 – 9000 USD.

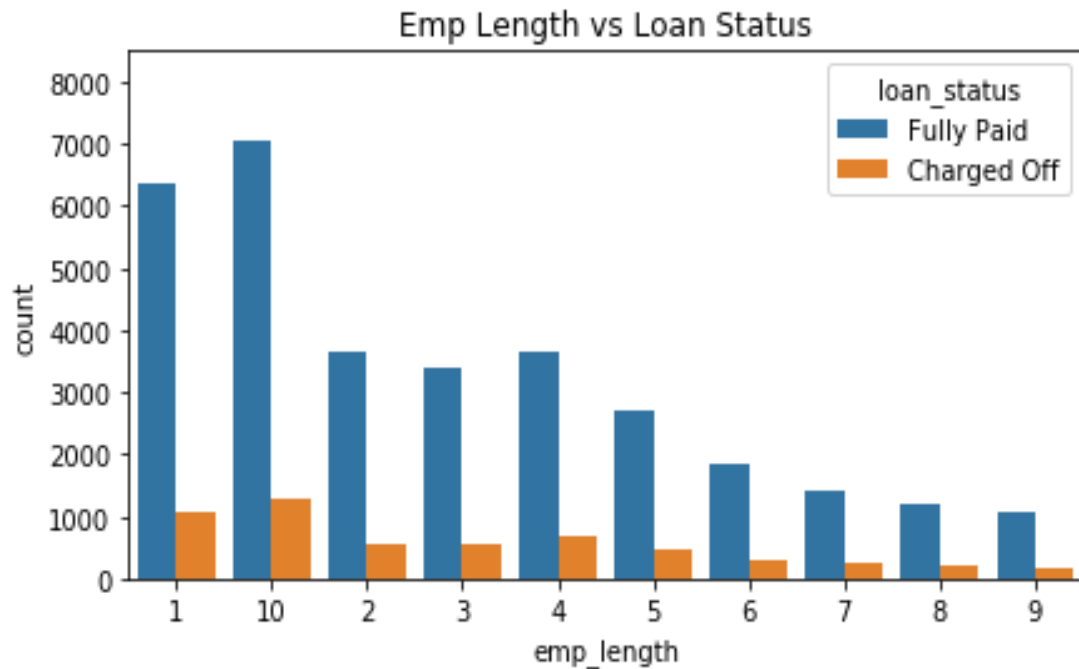


2. From the below catplot, it can be read that the loan applicants of grade G are more likely to default the loan.

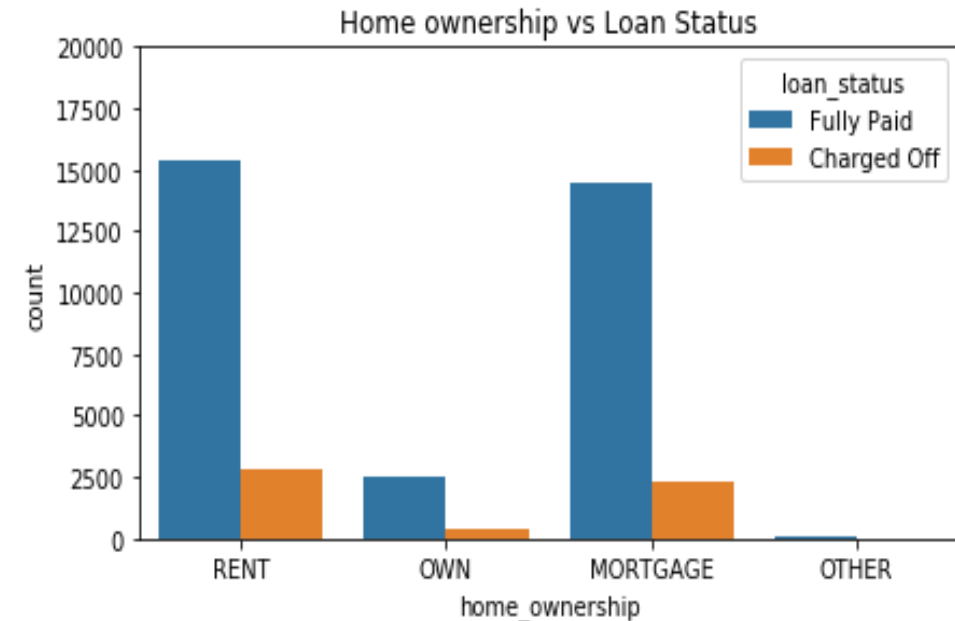


## Bivariate Analysis Results:

1. From the below plot, it can be concluded that loan applicants having experience of 10 years are most likely to repay the loan, as the ratio of fully paid vs defaulters is highest for this category.

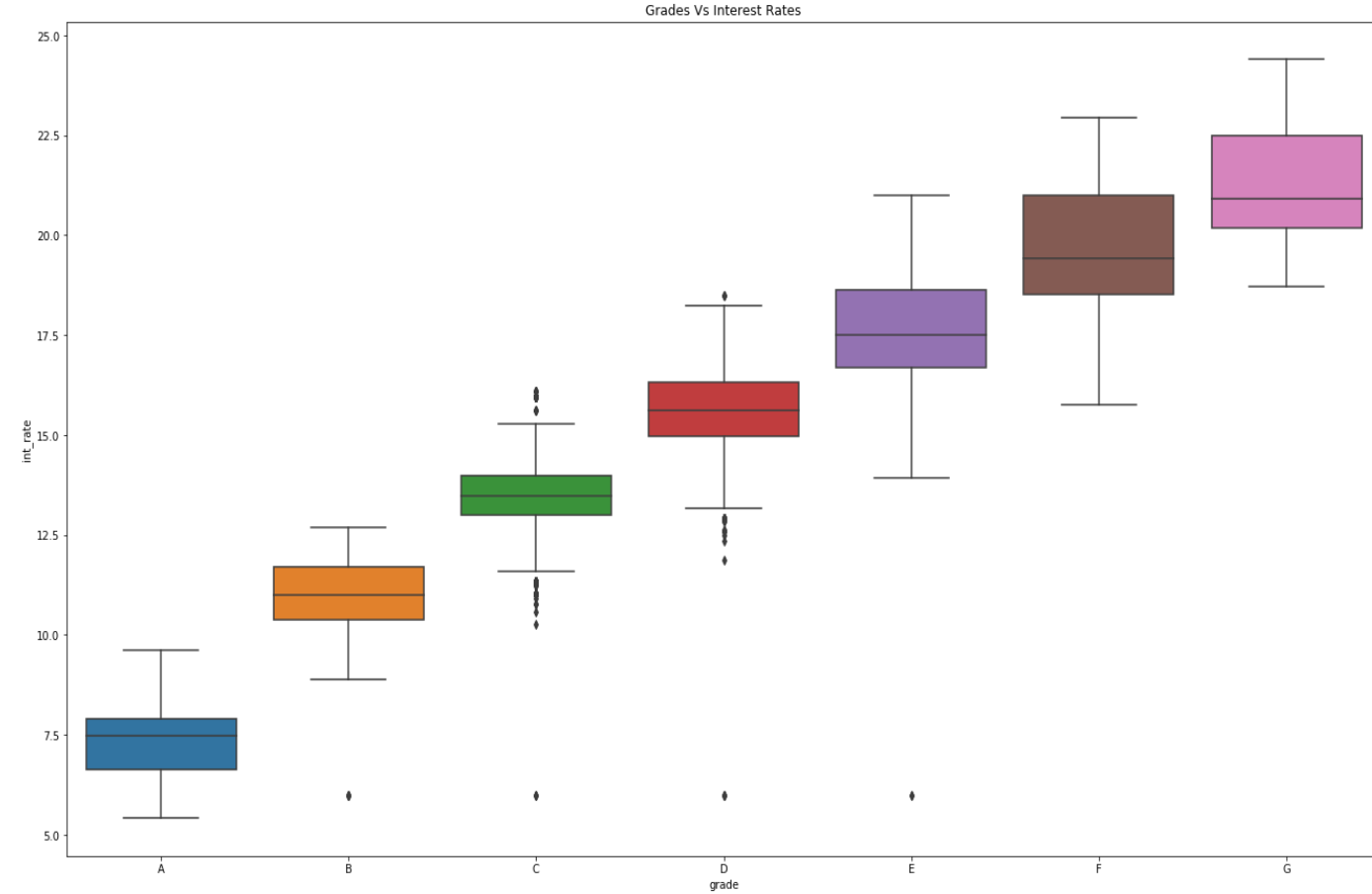
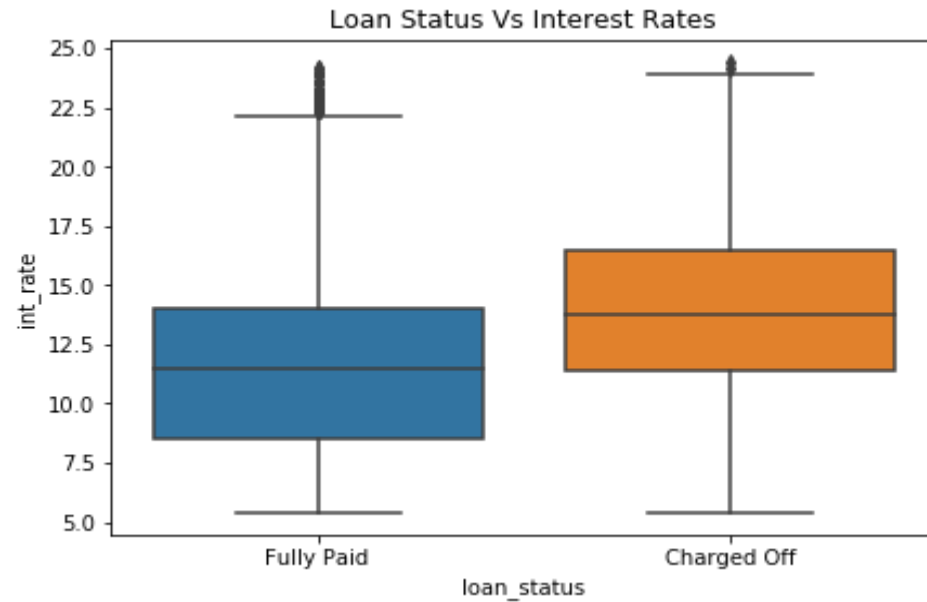


2. Mortgage and rented loan applicants are more likely to repay the loan.



## Bivariate Analysis Results:

From the given two plots, it is evident that loan applicants with grade D to G are provided loans with higher rate of interest and people who are provided with loans with higher rate of interest are more likely to default the loan.

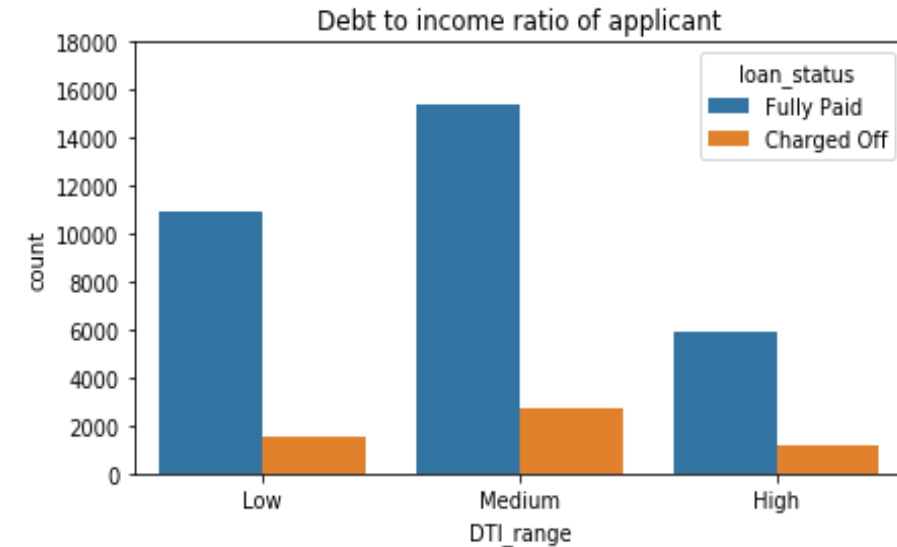
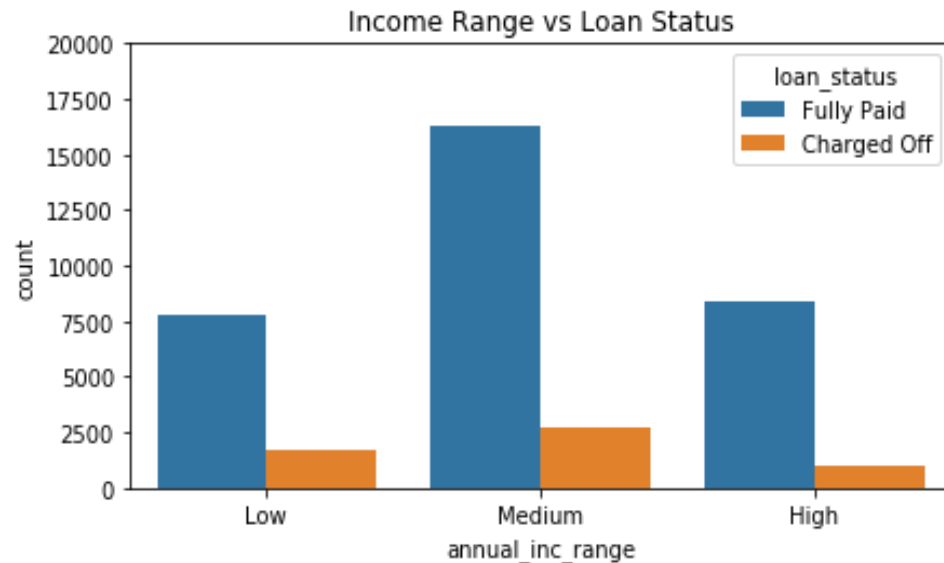




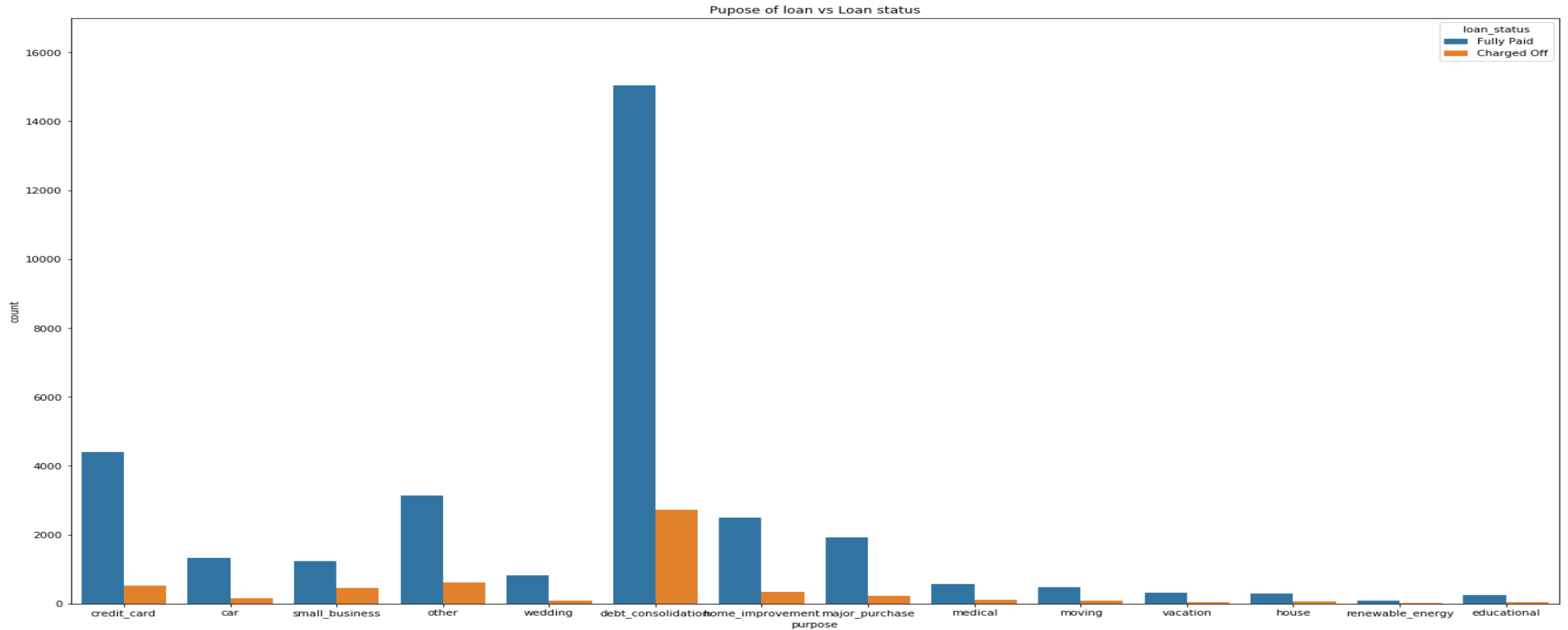
## Bivariate Analysis Results:

1. Ratio of fully paid vs charged off is highest for medium income range loan applicants. Hence this group of people are more likely to repay the loan amount.

2. Loan applicants having high debt to income ratio are more likely to default the loan.



## Bivariate Analysis Results:

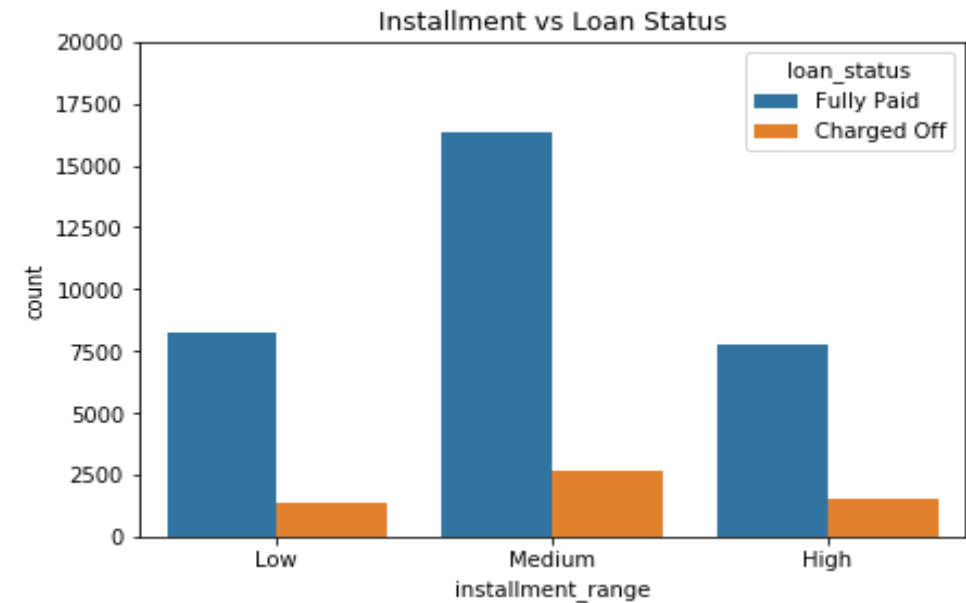
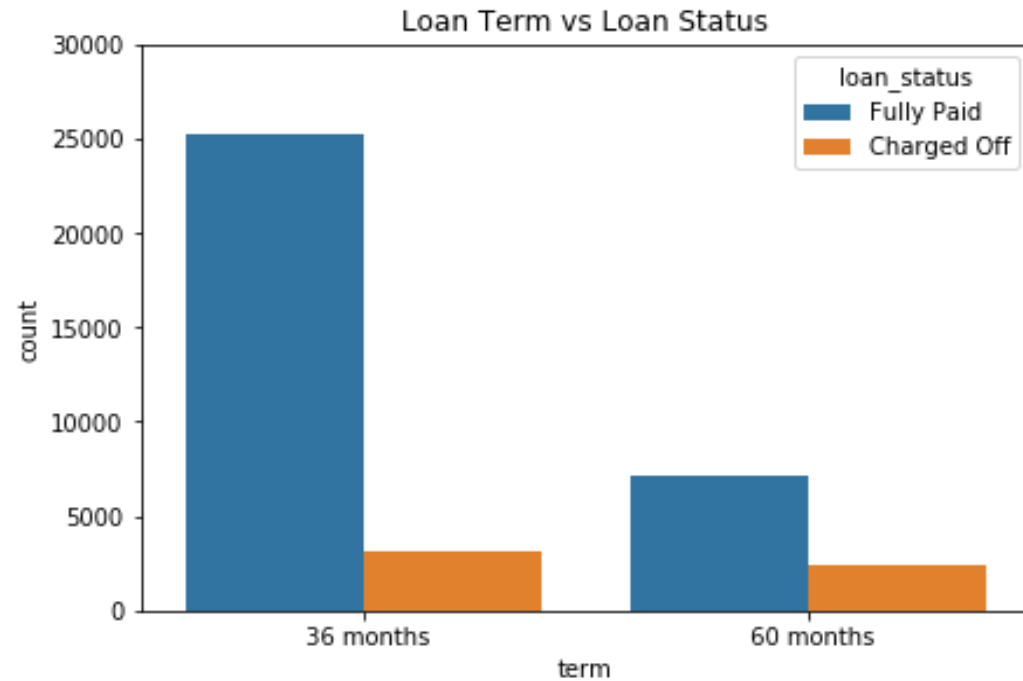


Loan applicants taking loans for debt consolidation are highly likely to repay the loan. Small business loan applicants are most likely to default the loan.

## Bivariate Analysis Results:

1. Loans provided for a tenure of 36 months are most likely to get repaid.

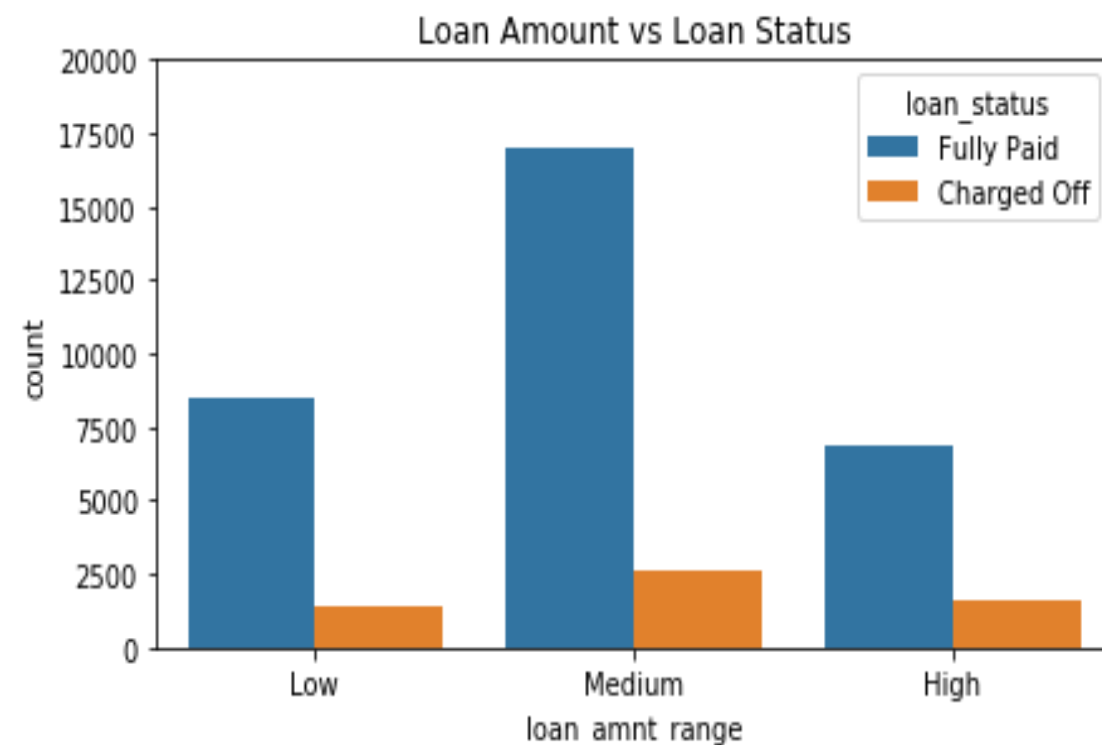
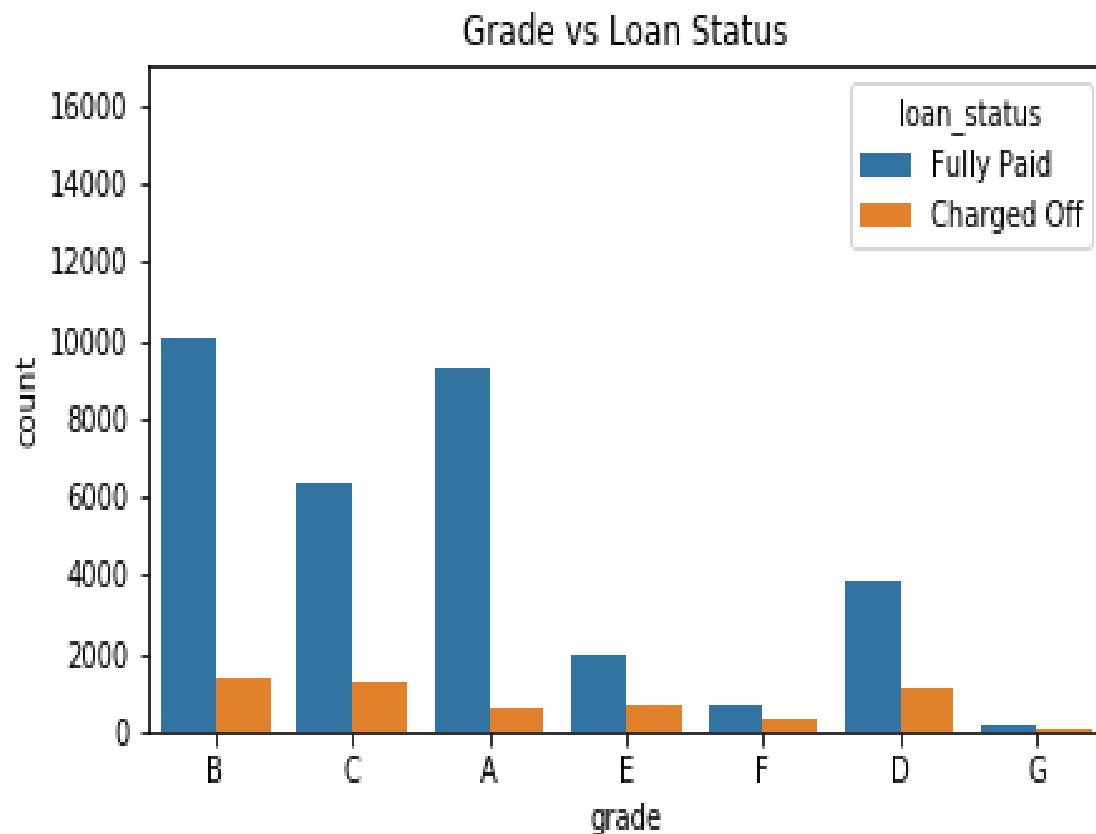
2. High installment range loan applicants are more likely to default.



## Bivariate Analysis Results:

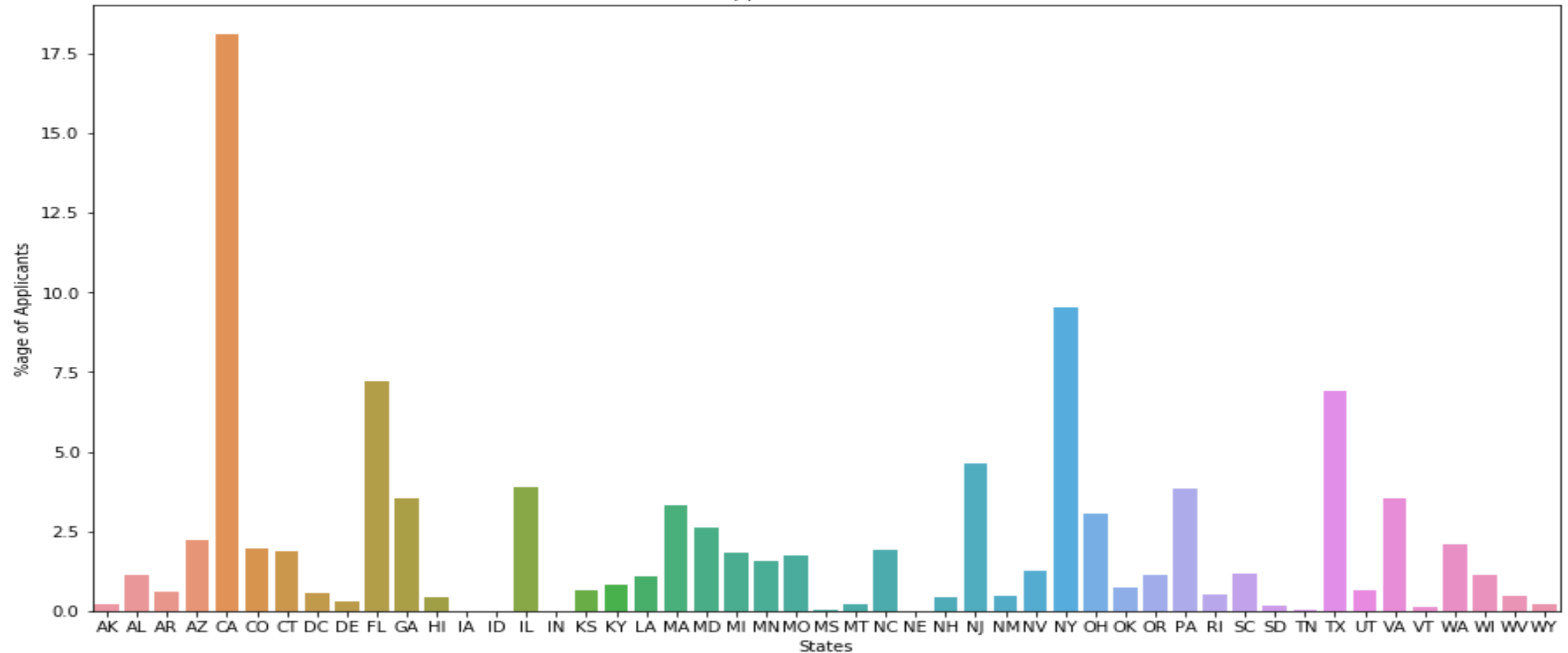
1. Loan applicants of grade A has the least default ratio of loan.

2. Loans categorized under high loan amount are more likely to default to the loan.



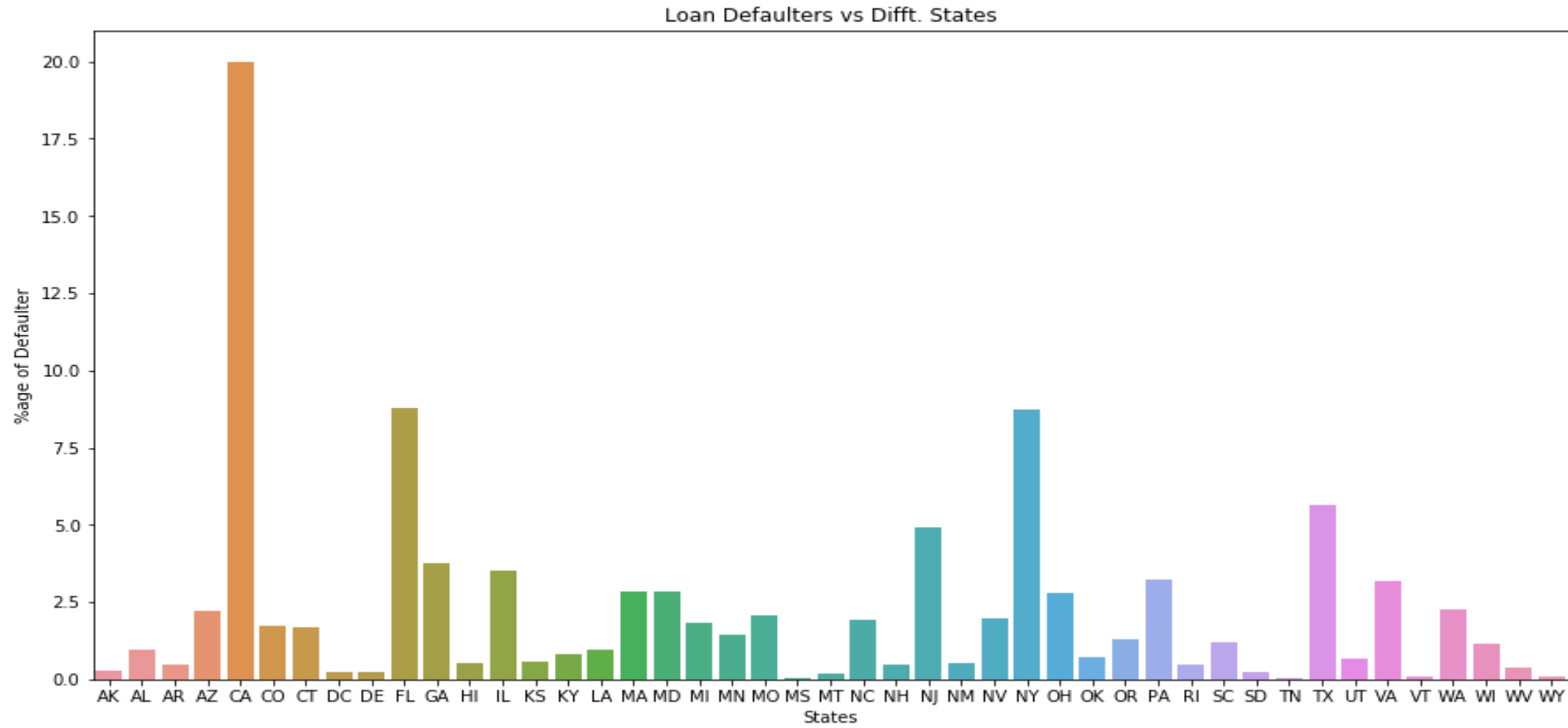
## Bivariate Analysis Results:

Loan Applicants vs Diff. States



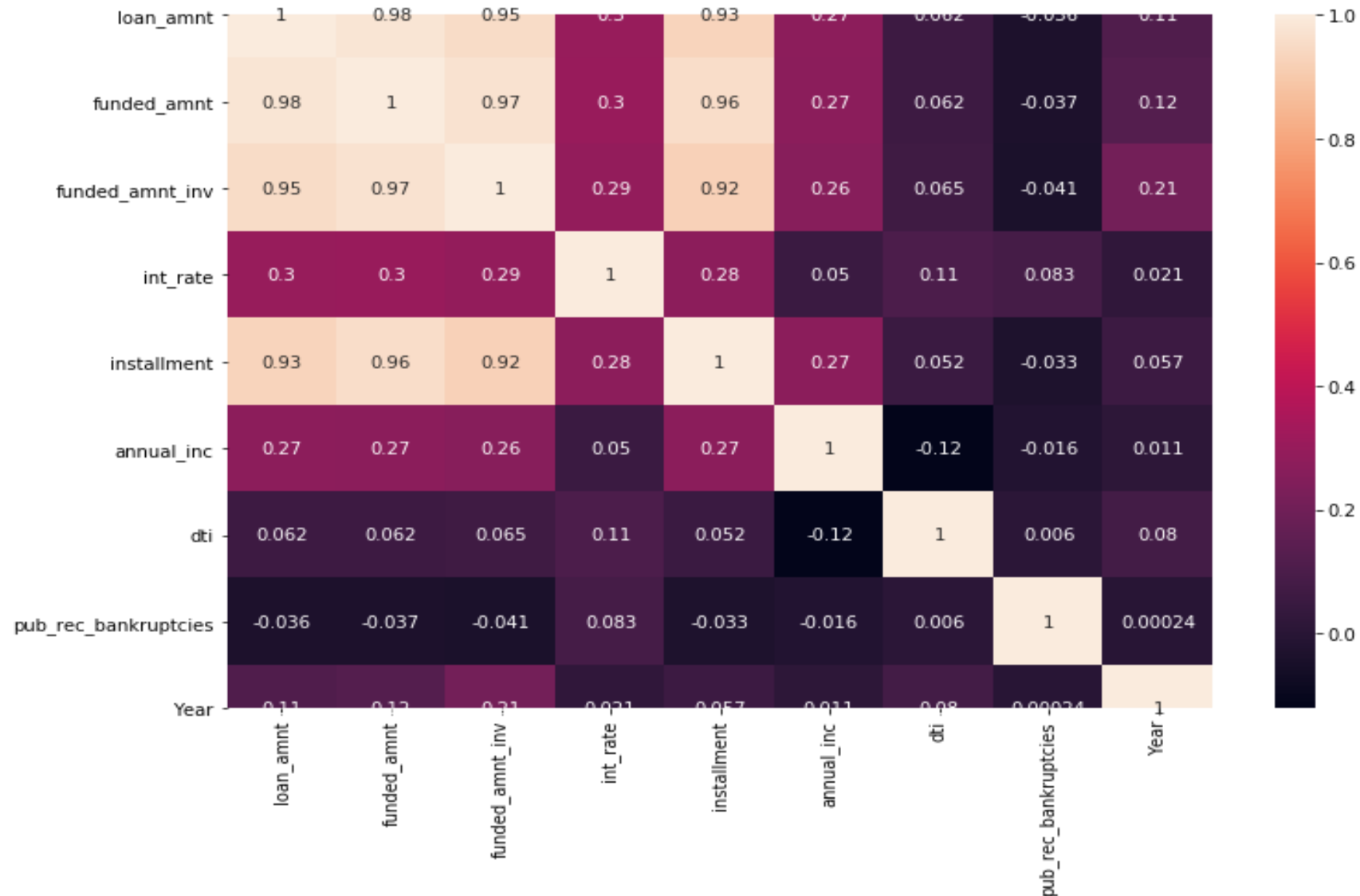
Highest number of loan applicants are from state CA (~17.5 %)

# Bivariate Analysis Results:



Loan applicants from state CA are highest loan defaulters (~20 %)

## Multivariate Analysis Results:



# Conclusions

With the above analysis, it can be concluded that:

1. Loans provided for a tenure of 36 months are most likely to get repaid.
2. Loan applicants belonging to grade A-C are most likely to repay the loan and those from lower grades like G would default on loan.
3. With increase rate of interest, the chances to default the loan increases. From the analysis, it can be concluded that the ROI increase with grades of people moving from A-G.
4. The percentage of loan applicants is maximum from state CA (~17.5%), but the percentage of loan defaulters are also maximum from state CA (~20 %).
5. Loan applicants having experience of 10 years are most likely to repay the loan.
6. Loan applicants having high debt to income ratio are more likely to default the loan.
7. Loan applicants taking loans for debt consolidation are highly likely to repay the loan.