

Lending Club Case Study

Submission

An illustration showing two hands in business suits. The left hand holds a white document with a signature and horizontal lines. The right hand holds a stack of green banknotes with a dollar sign. The background is a light beige color.

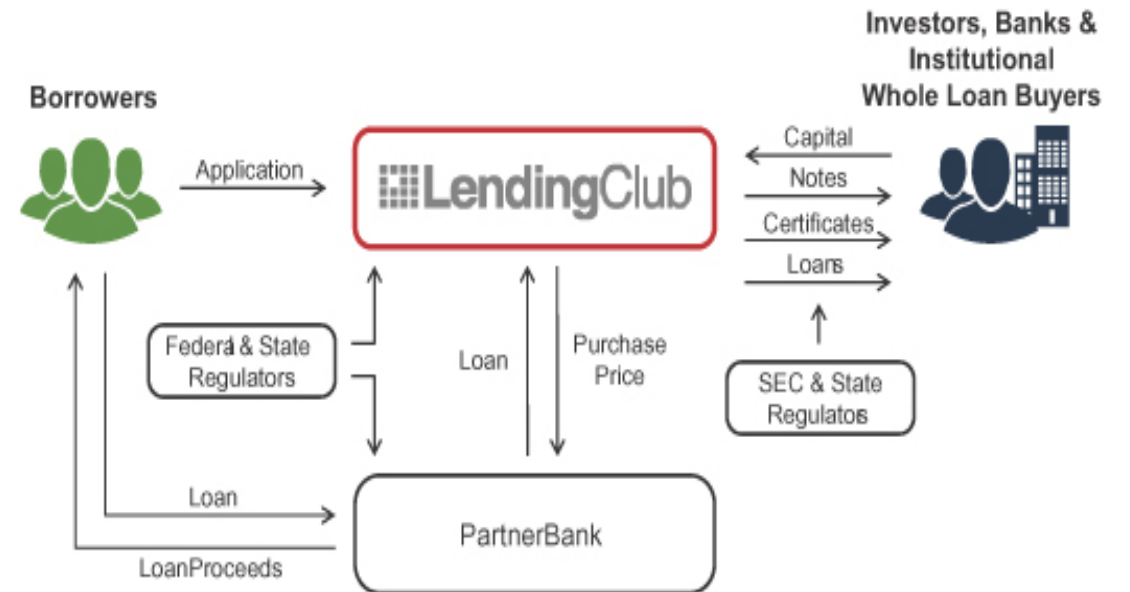
Group Members:

Vaibhav Shukla

Pankaj Kumar Sharma

Abstract

- **Lending Club** is a market place which lending various types of Personal loans , Business loans and Financing of medical procedures. When company receives a loan application, the company have to take decision of loan approval based on the applicant's profile.
- As an employee of Lending Company our aim is to analyse the past data . The data contains all the information about Applicant like annual income home ownership and past loan whether they '**defaulted**' or not.
- The aim is to identify patterns which indicate if a Applicant is likely to be default, which may be used for taking decision such as denying loan, reducing the amount of loan, lending (risky applicants) at a higher interest rate etc.



Problem Solving Approach – Flow chart

Data Understanding

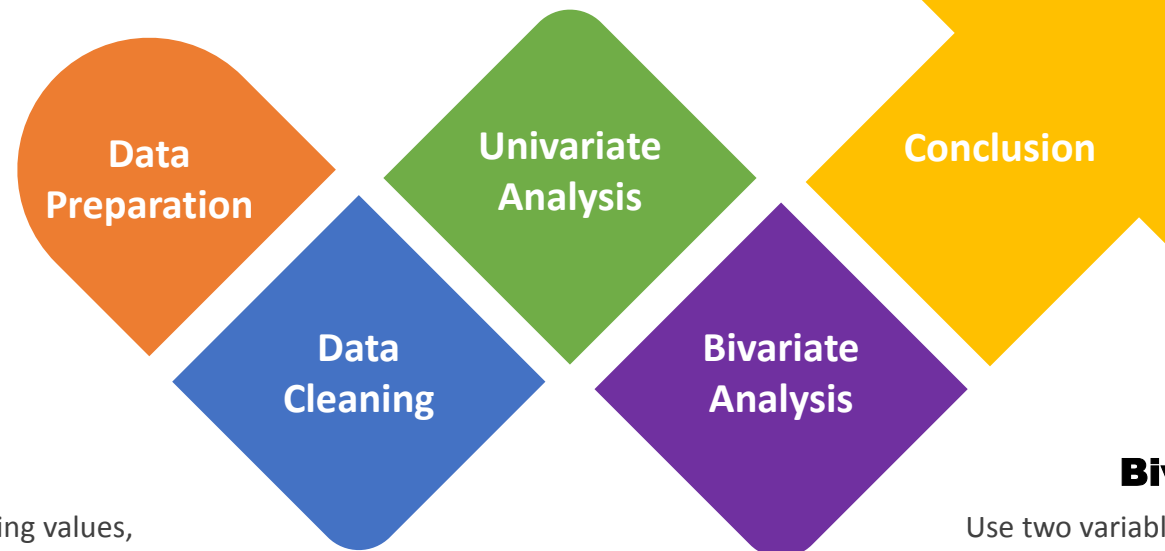
Read loan dataset, check sample data, size and datatype, then find out the Target variable

Univariate Analysis

Deal with single variable to analyse the important factors. Find the driver variables. Plot variables with loan status and find out the inference.

Conclusion

Using analysis conclude the highly impact variables which is results in loan Status. Similarly conclusion is drawn by analyzing combination of variables impacting the loan default.



Data Cleaning

Cleaning the dataset by dealing with missing values, removing redundant columns, correcting data type, depending on requirements.

Bivariate Analysis

Use two variables to plot the charts, heat map, scatter plots. Find out the impact of combination of variables on loan status.

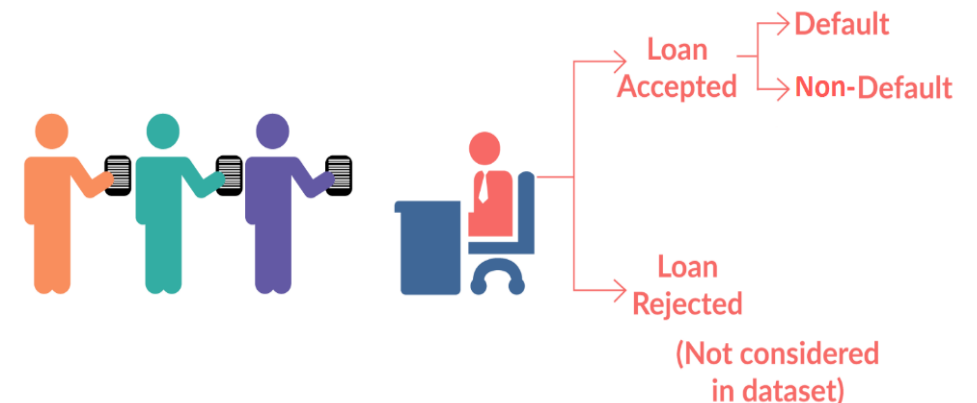
Data Understanding

- Loan data is imported from “loan.csv” file. The raw data contains 39717 rows and 111 columns.
- Checked the datatype of various columns.
- There are different types of variables present in the dataset like:
 - **Applicant Demographics** (Employment length, Employment title, Annual income, etc.)
 - **Loan information & its characteristics** (Loan amount, Funded amount, Loan grade, Loan status, Interest rate, etc.)
 - **Customer behavior** (application type, loan purpose, delinquency 2 year, revolving balance, etc.)
- Now, applicant behavior variables are not available at the time of loan application , thus it can not be used for prediction of loan approval.
- We also checked the columns which contains numerical data and extract it for further data insights.
- Understanding of date attributes and convert them to the correct date format.
- We identified target variable columns , in this case “loan_status” is target column.

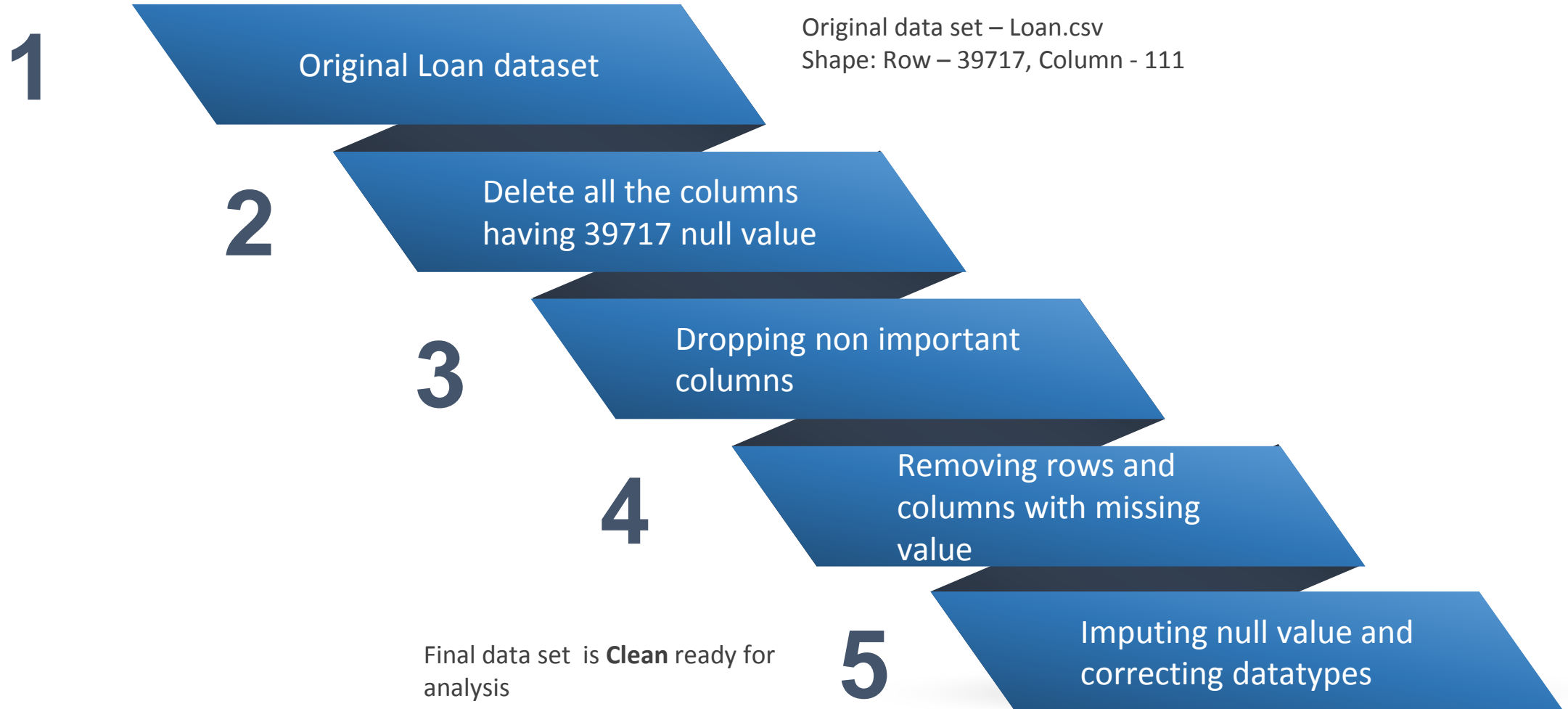
Shape of dataset

Row – 39717, Column - 111

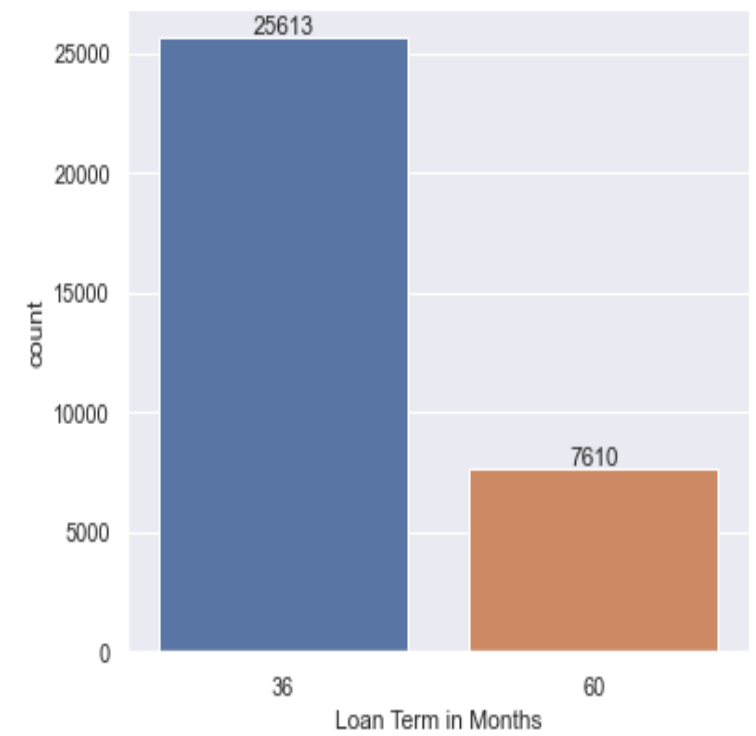
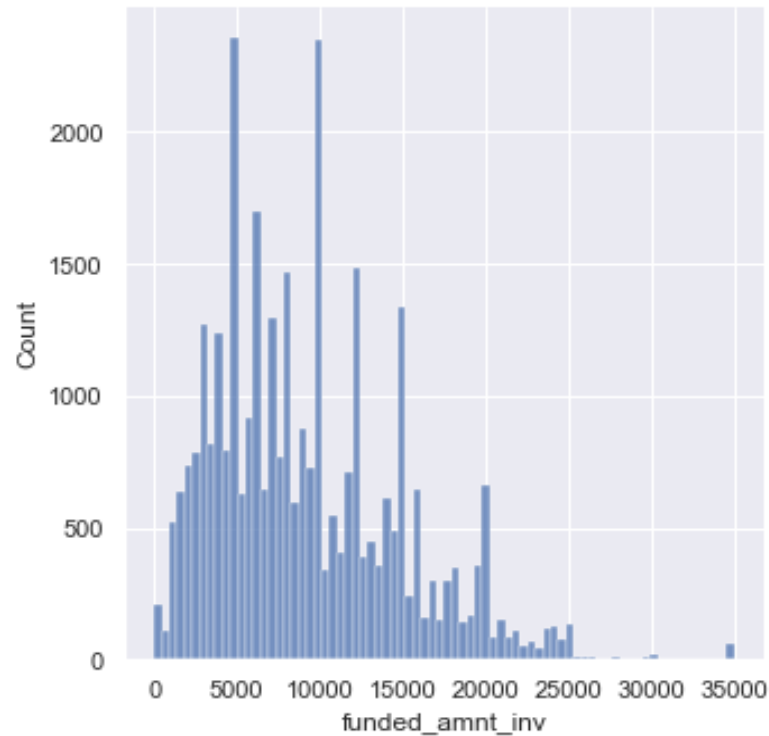
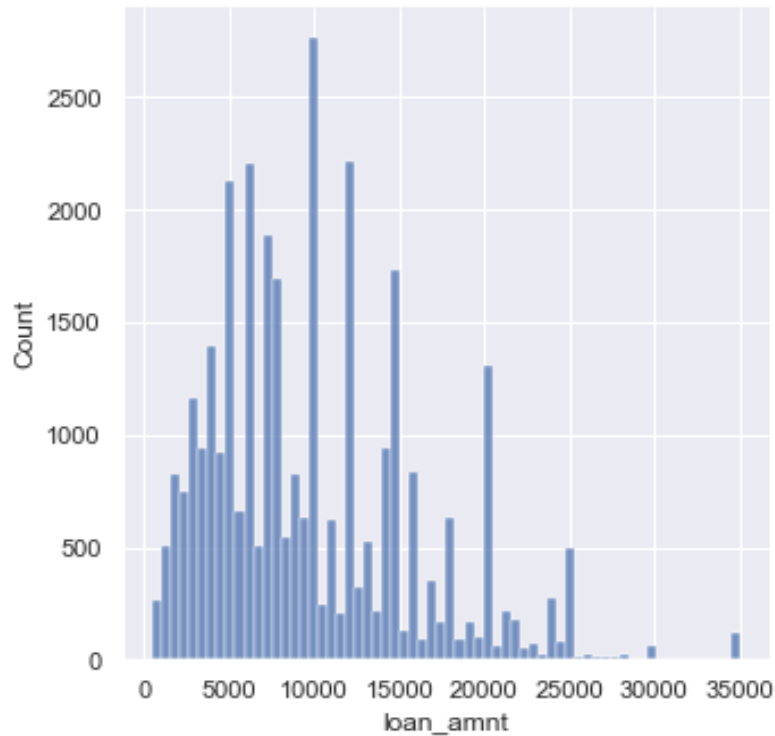
LOAN DATASET



Data Cleaning



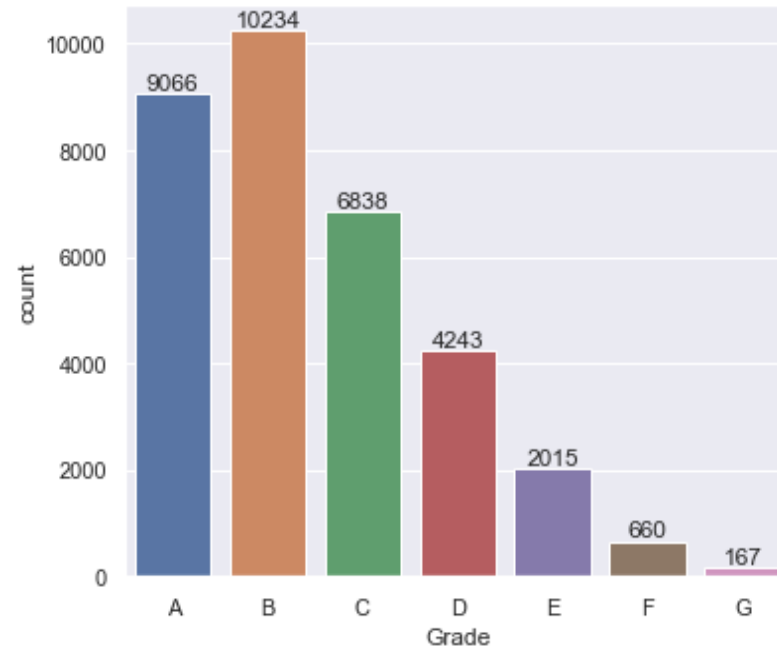
Univariate and Segmented Univariate Analysis



Similar to loan amount more number of loans were sanctioned near 5000 while, the amount 10000 is almost similar to 5000 which were sanctioned

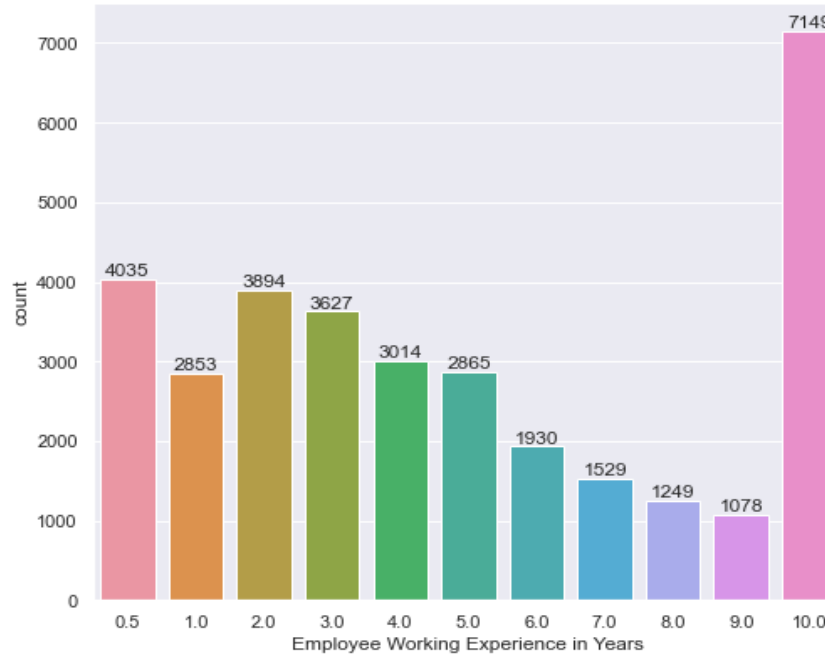
Most of the loans were for 36 months term

Grade vs loan count Bar plot



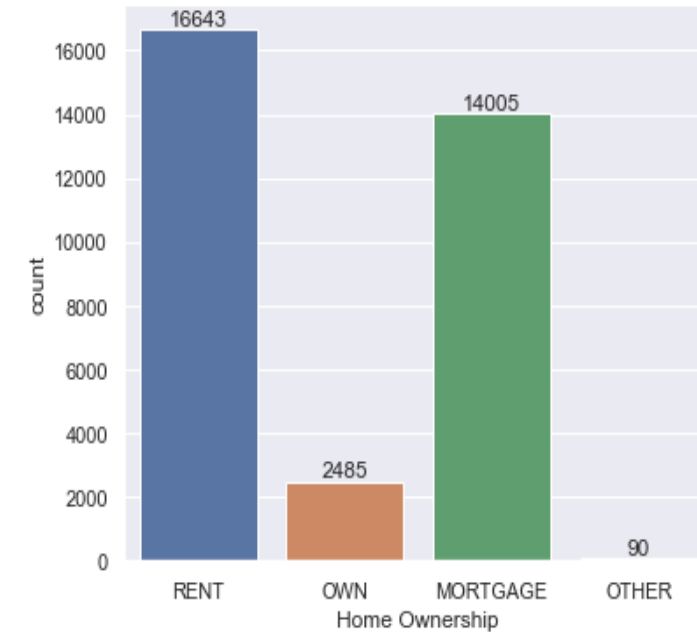
Most of the loans are grade A and B type. Therefore we can say, most of the loans are high graded

Employee Experience vs loan count Bar plot



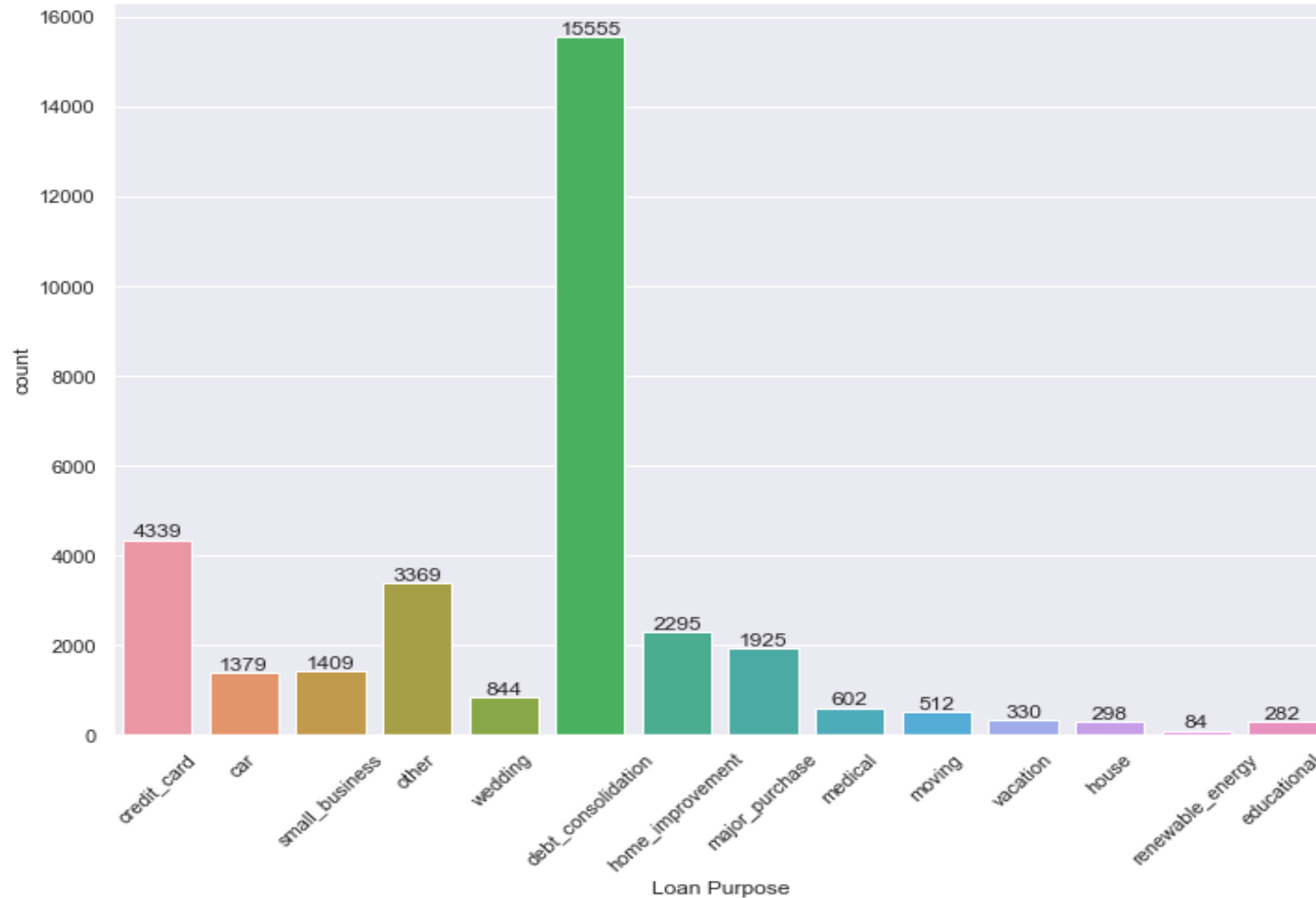
Majority of employees applying for the loans are having more than 10 years of experience

Home ownership vs loan count Bar plot

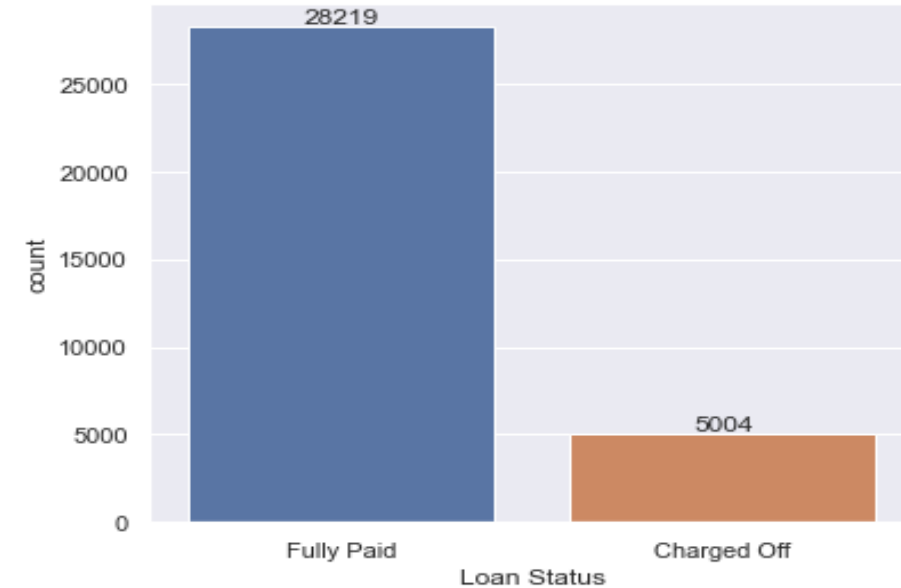


Most of the borrowers are renting the house followed by mortgage while other is the least.

Loan purpose vs count Bar plot



Loan status vs count Bar plot

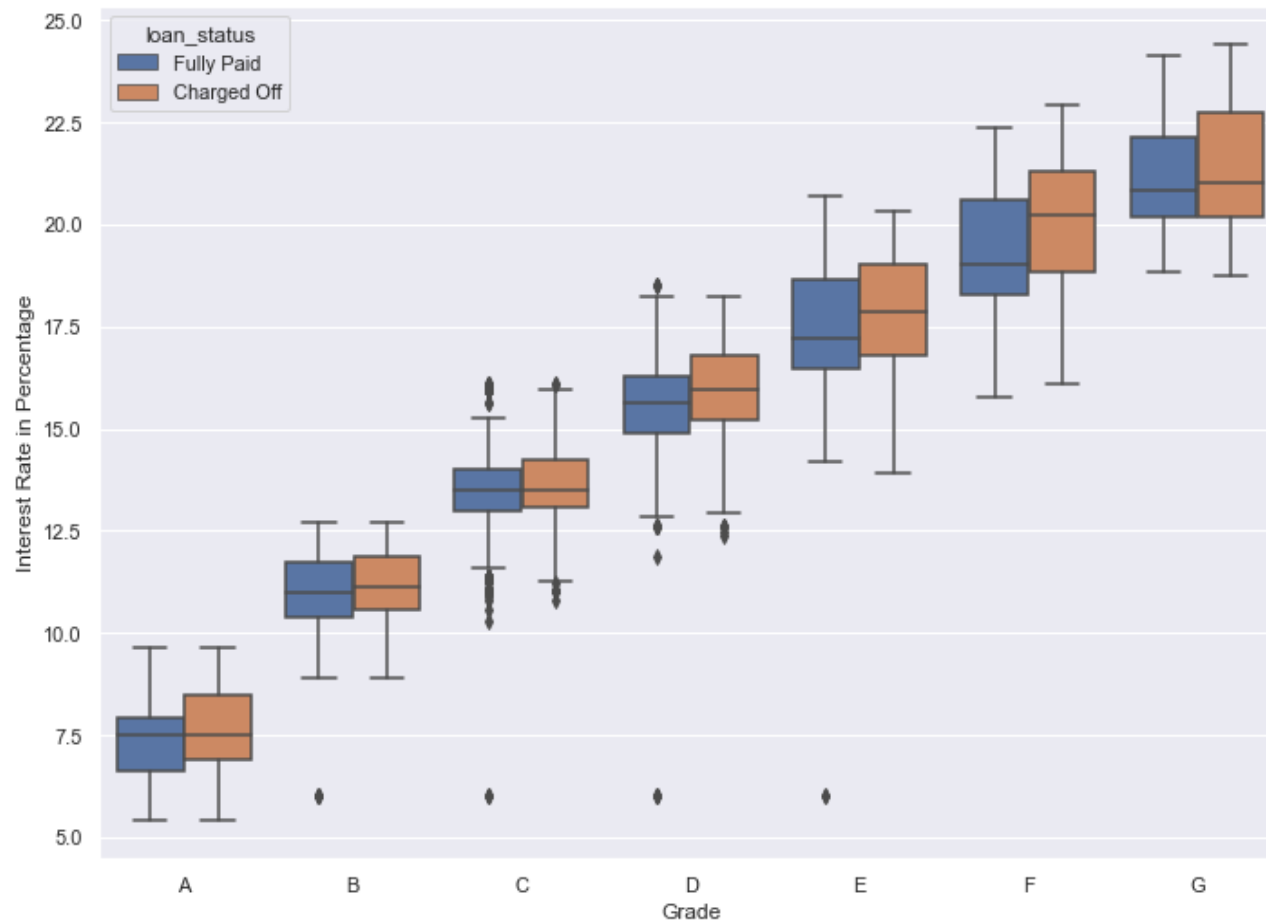


Majority of loans are fully paid

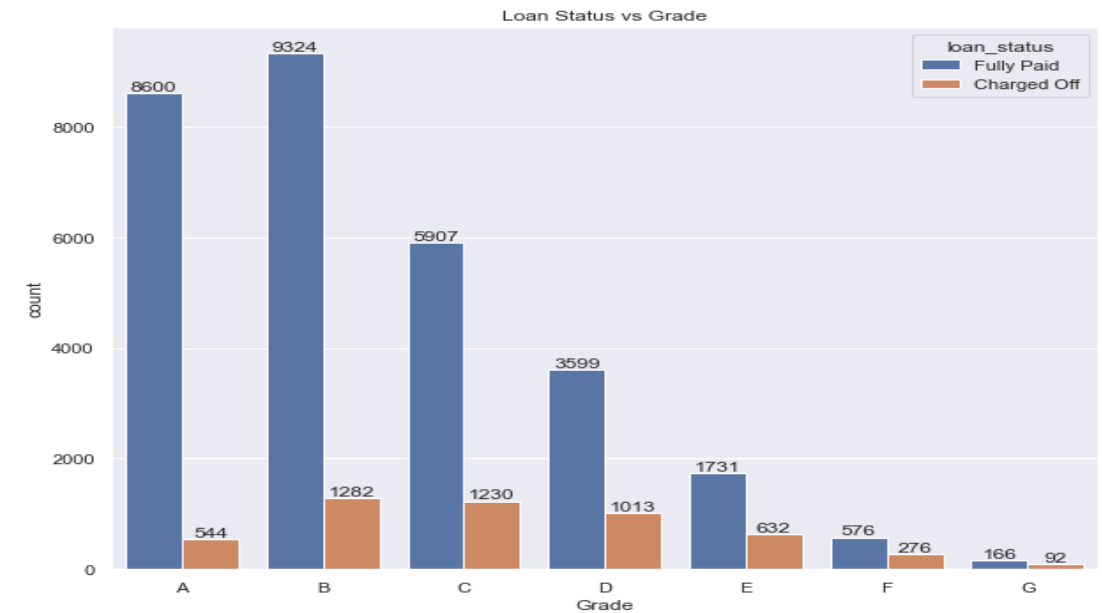
Maximum number of loans are applied for the debt consolidation followed by credit card while renewable energy is the least.

- In Bivariate & Multivariate analysis we plotted bar graph, heat map and subgraph with combination of the important variables identified during univariate analysis .
- From the univariate analysis, we have seen that the chance of default is changing heavily with the purpose. In the bivariate analysis, we can see the patterns of the default against the purpose and the remaining important derived variables one-by-one.

Box plot for Grade vs int. rate percentage wrt loan status

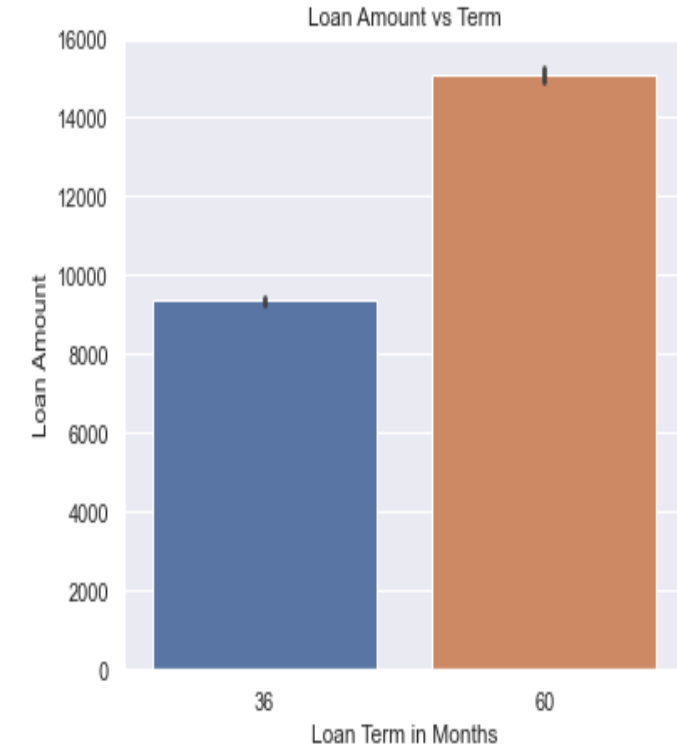
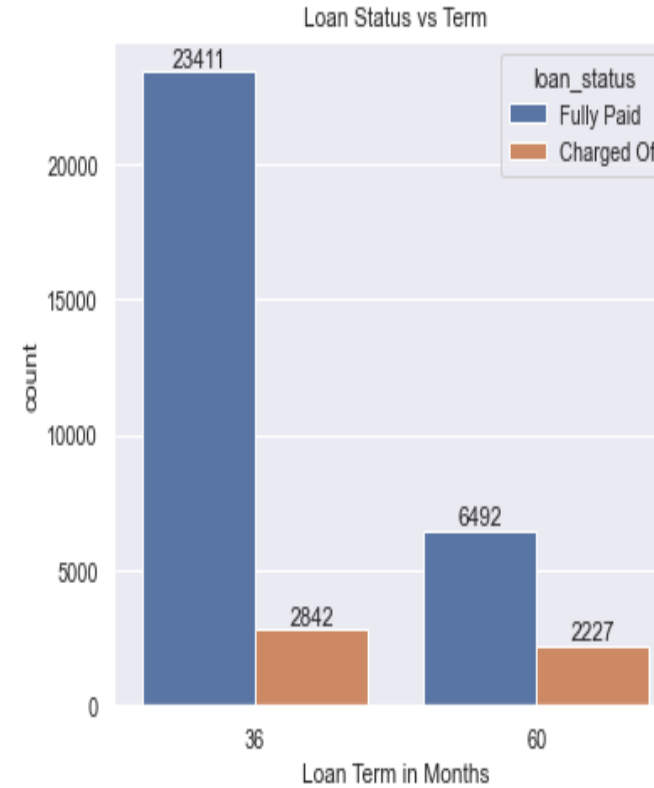
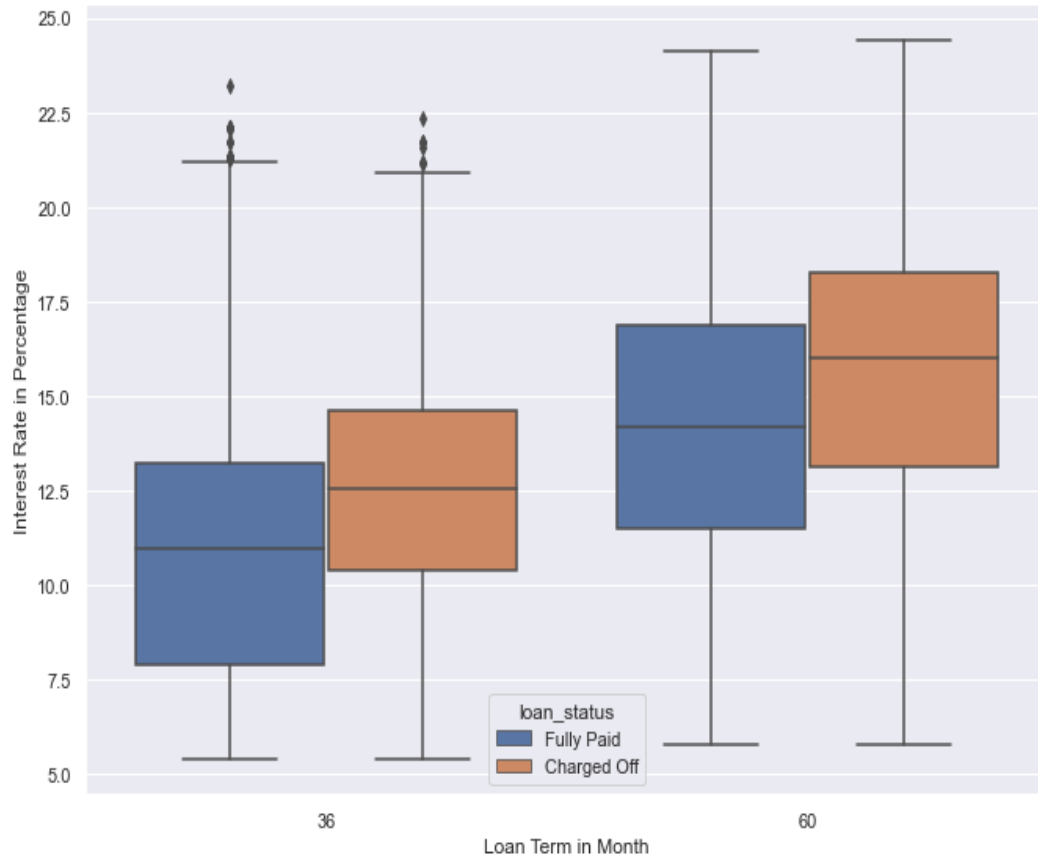


Bar plot of Grade with respect to loan status



Most of loans are High Grade loans and High Quality loans have low interest rate

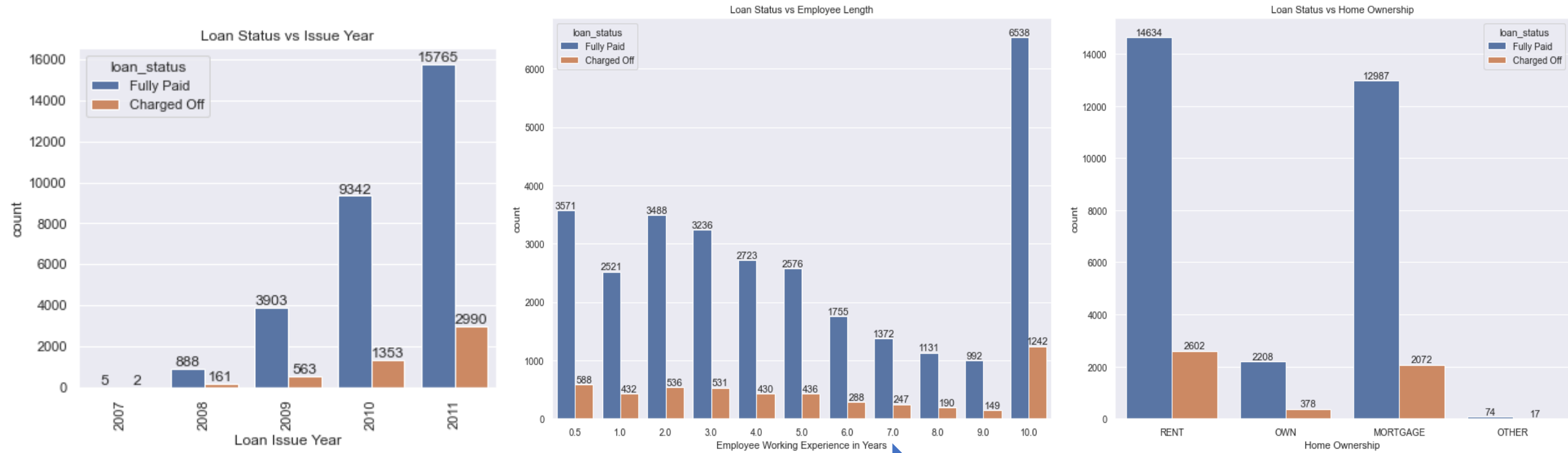
Higher interest rate have higher tendency to default the loan



In both the terms, the chances of charged off are high if the interest rate is more

On an average the default rate is 14%. The default rate is high (34%) when loan term is 60 months as compare to loan term of 36 months (13%).

Higher amount loans are having long term which is 60 months.

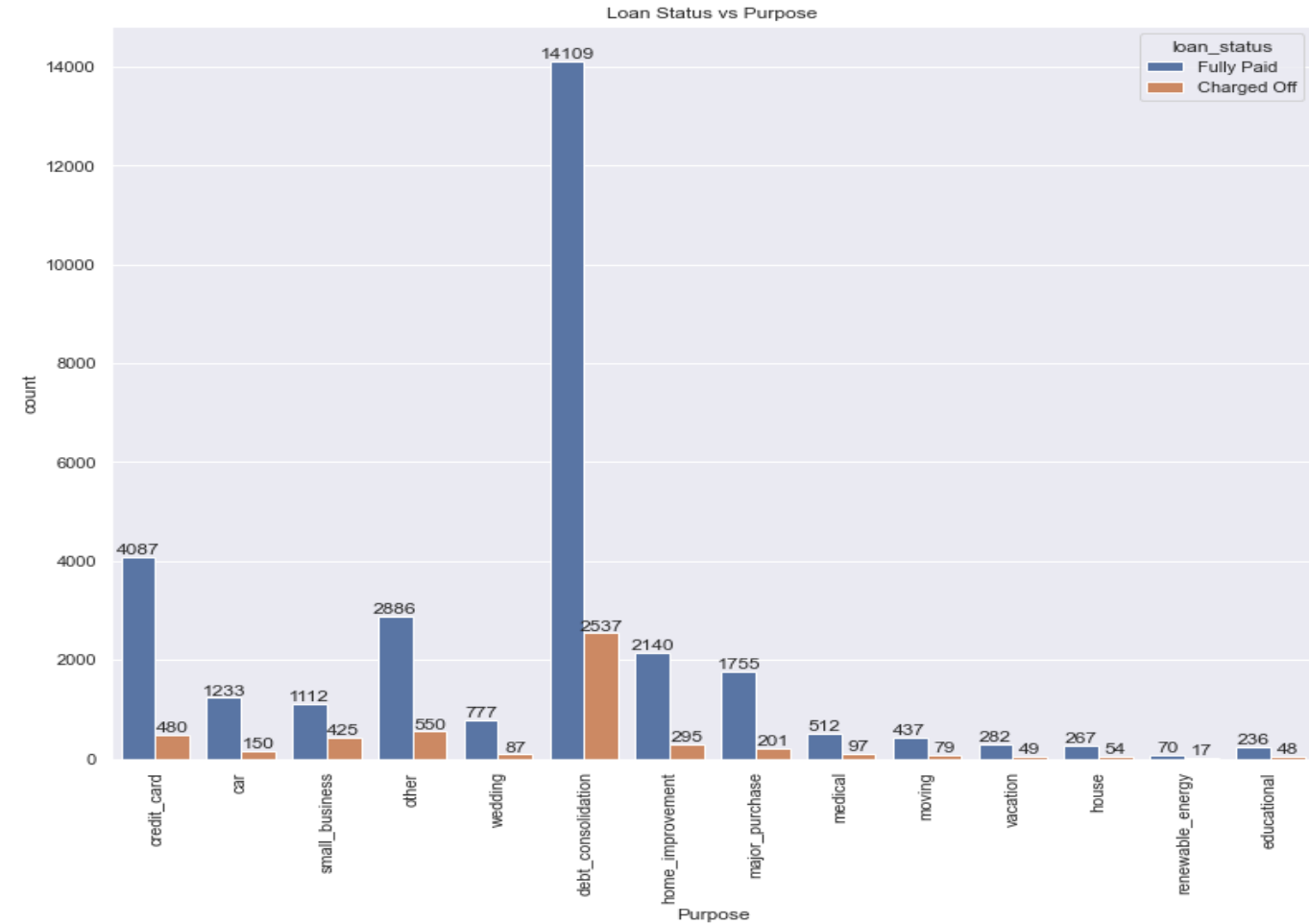
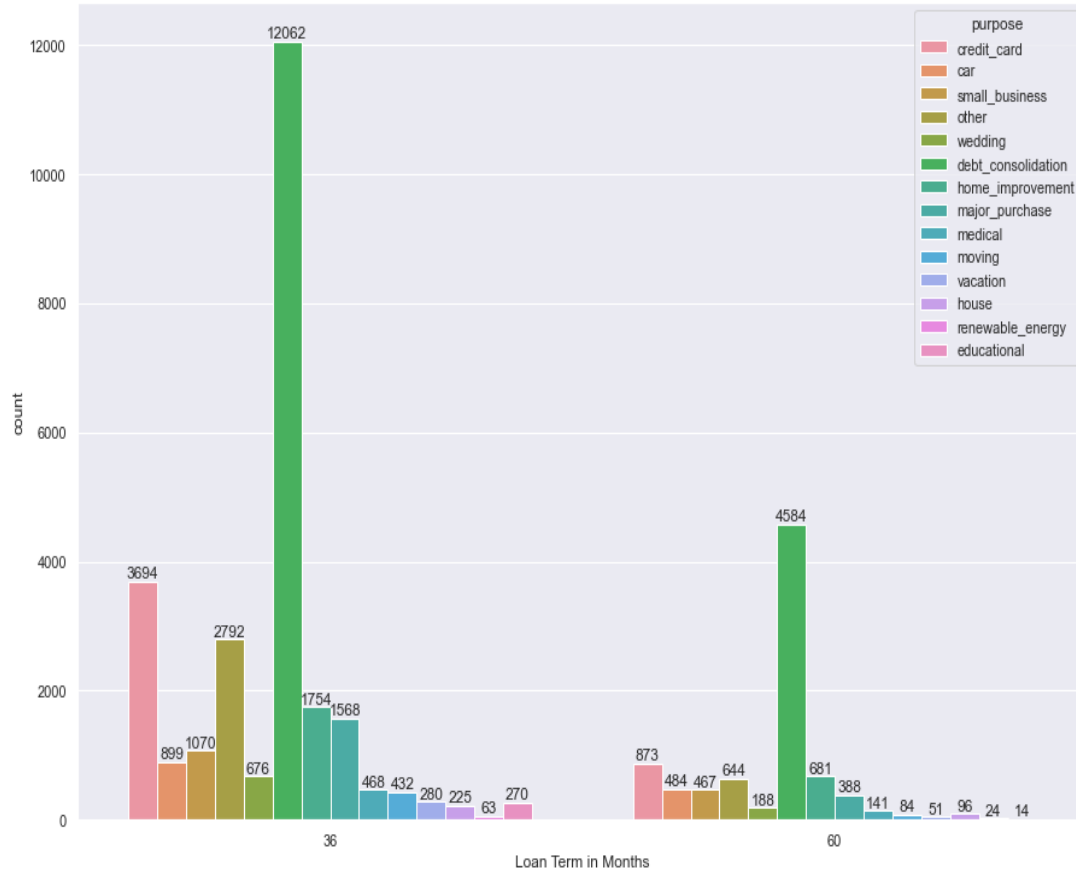


The loans granted has been increasing exponentially over the years, 2011 were having maximum defaulters as compared to other years.

Majority of employees applying for the loan have more than 10 years of experience,

Tendency of Applicant to default the loan with 10 years of experience is also high

Applicants having their home rented or on mortgage have almost equal tendency to default the loan

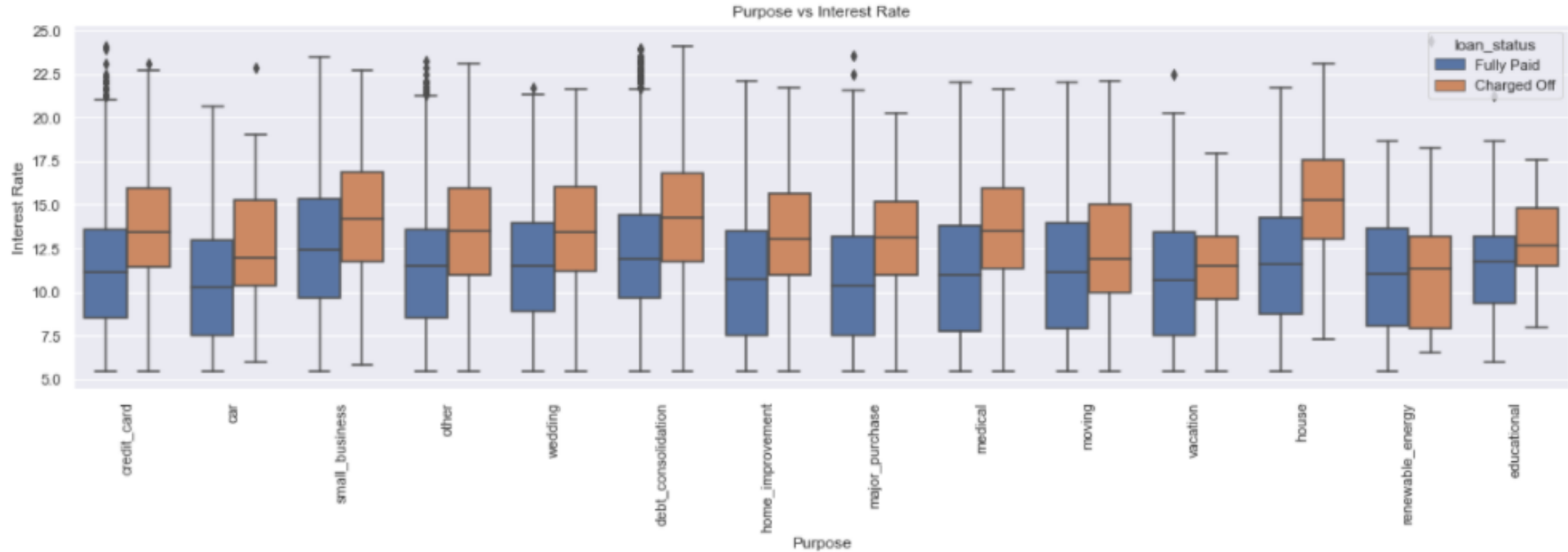


Analysis confirm that chances of default is high when purpose is small business and the term is 36 months.

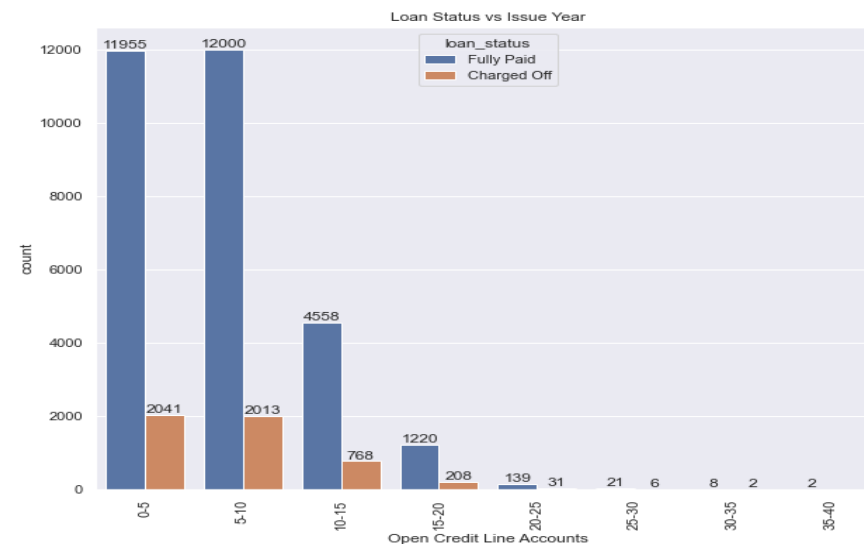
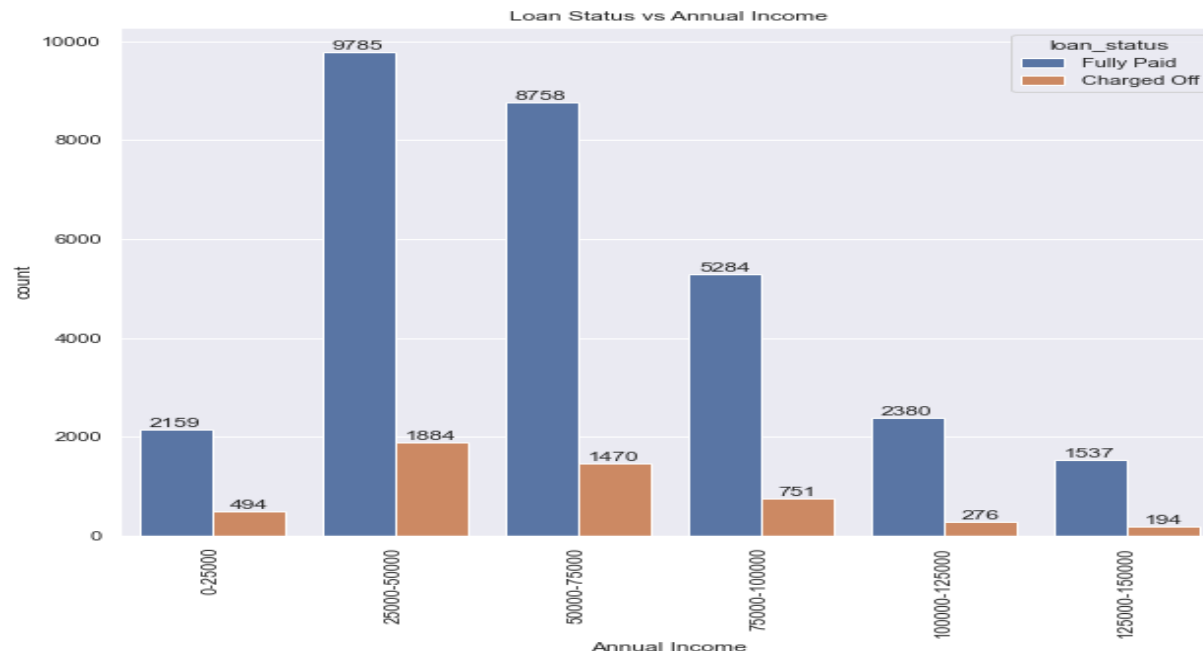
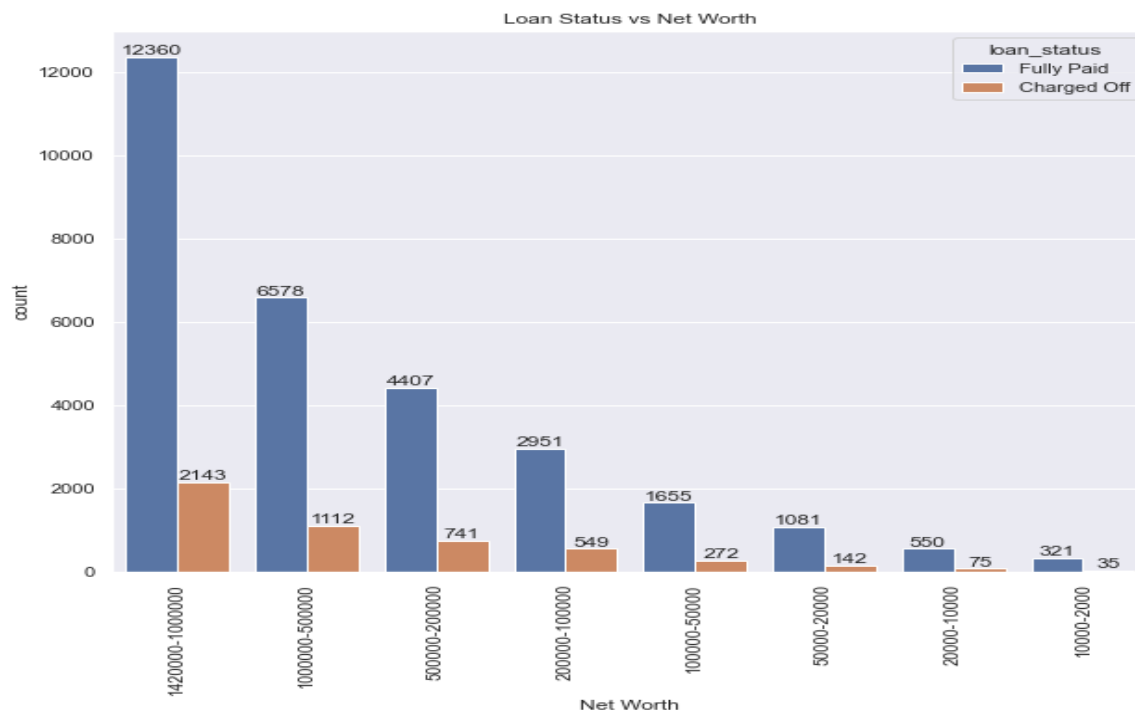


Also for 60 month term the default chance is high when purposes are debt consolidation, educational, house, other, small business and the vacation purposes

The ratio of defaulters is maximum for small scale business borrowers followed by renewable energy.



Normally small businesses, renewable energy are having more defaulters. But if we consider the high interest rate factor then other purposes like house, debt consolidation are also tend to be charged off.



Borrowers whose net worth is between 100000-200000 tend to more defaulters followed by highest net worth borrowers.

Default chances is inversely proportional to the annual income. High the income less chance of default

Borrowers having more credit line open accounts tends to be charged off more

Variables	Inference
Interest Rate	Higher interest rate loans are mostly deterrent. Therefore, denying such risky individuals might be good approach instead of approving them.
Number of Open Credit Lines	Borrowers having more open credit lines tends to be more defaulters. So it is justifiable if such individual's loans are getting handled carefully.
Grade	It is clear from the analysis that high grade loans are having less interest rate and less defaulters. So it might be good opportunity for the investors to invest on high grades also, lower grades could be avoided.
Loan Amount to Annual Income Ratio	Income plays the key role in repaying the loan. If the income of the individual is more then the chances of defaulters are less. Hence, approving their loans will be in the benefit of the industry.
Purpose of Loan	Loans with purpose small businesses, debt consolidation, renewal energy, are more of defaulters. So keeping that in mind, rejecting such loans might be a good idea in the interest of industry.

Conclusions and Recommendations

Sl.no	Insights
1	Though more number of loans are applied for 36 months and has more defaulters, the ratio of charged off loan is more for 60 months.
2	The chance of default is monotonically increasing with the interest rate i.e. the high the interest rate high the default rate chances. Interest rate is high for lower grades. So there is possibility of more charged off for them.
3	The ratio of defaulters is maximum for small scale business borrowers; however, the numbers are high for debt consolidation
4	Defaulter chance is inversely proportional to the annual income. High the income less chance of default.
5	Default chances is increasing with the increase in the debt-to-income ratio.
6	Default chances is directly proportional to the loan amount. Also, chances of default is more when the home_ownership is "other".
7	From analysis it's confirm that chances of default is high when purpose is small_business and the term is 36 months. Also, for 60 month term the default chances is high when purposes are debt_consolidation, educational, house, other, small business and the vacation purposes.
8	Trend shows that default chances are more when the home_ownership status is other and the term is 60 months. Even when the term is 36 months, the default chances are the high if the home_ownership status is other. But there is only one entry (term = 60 months and home_ownership = other), and very less entries in the other home_ownership. So it is better to avoid the loans when the home_ownership is others.
9	In general, educational, house, other and the small business loans are having very high interest rate, debt consolidation with medium high interest rate and renewal energy, house with high interest rate are showing more chances of default .
10	The high income group for all purpose is relatively safe to lend money. The small business is showing high default chances in all income group except very high income group. Loan with purpose renewal energy is showing high chance of default with very low, low and high income bracket.
11.	Loan with purpose car, credit card, home improvement, medical and wedding is showing less chances of default among various category and one of the safest purpose to lend. On the other hand loan with purpose small business, debt consolidation, renewal energy, educational, house and other is showing high chances of default.