

LENDING CLUB CASE STUDY SUBMISSION

Group Members:

- 1. Bhavana Mayreddy**
- 2. Manjunath Sirur**

Problem Statement :

Analyse the past loans data to determine the categories / parameters which can predict the risk of the loans to be defaulted.

Case Study Objectives :

- Track and identify the attributes of the loan applicants that are mostly tending towards 'default' the payback
- Analyse and understand the attributes / driving factors that are behind the default loans
- These knowledge may be utilized by the Lending Company for assessing the risk of new loan applicants

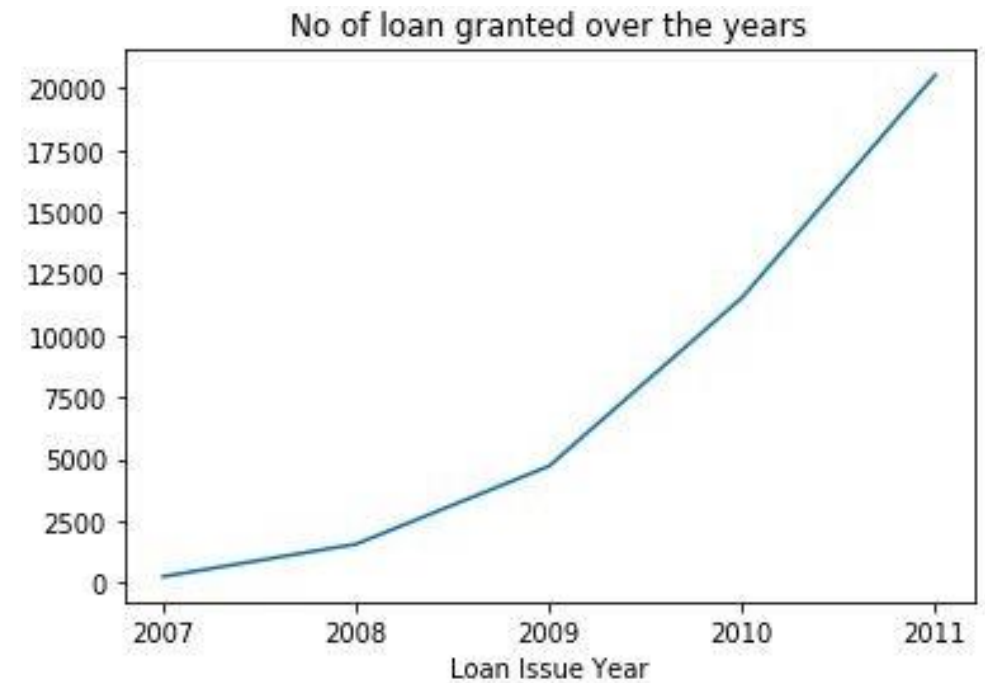
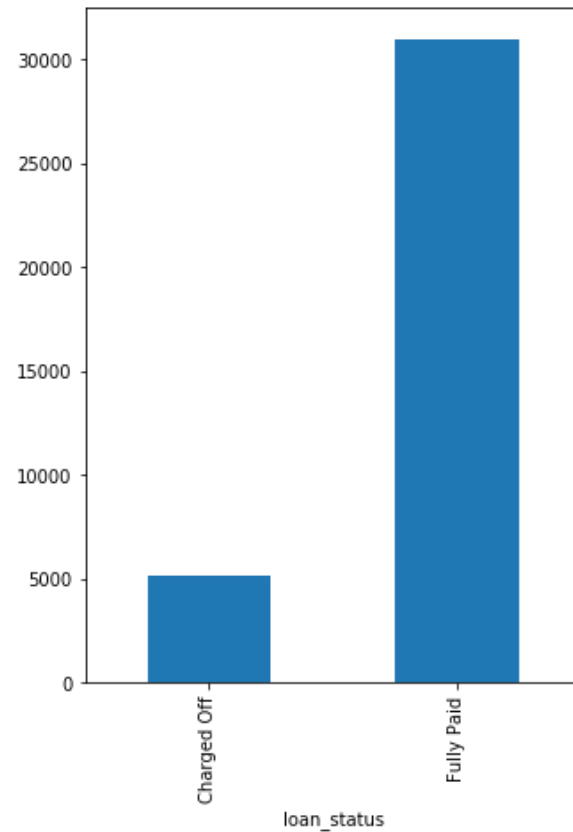
Data Cleaning Steps

- **Delete columns:** Delete unnecessary columns.
 - Columns with null values >60%, with single category values and not available at the time of loan application
- **Remove outliers:** Remove high and low values that would affect the results of your analysis.
- **Missing values:** Treat missing values with appropriate approach.
 - Missing values <5% of the data are removed else treated
- **Duplicate data:** Remove identical rows, columns which same information can be ignored.
- **Modify data :** Modify the columns to understand it better
 - Recreate/create the columns by modifying the data
 - Change the columns to appropriate datatypes
- **Filter rows:** Filter by segment, filter by date period to get only the rows relevant to the analysis.
- **Relevant data :** Data is trimmed down by ignoring current status loans.

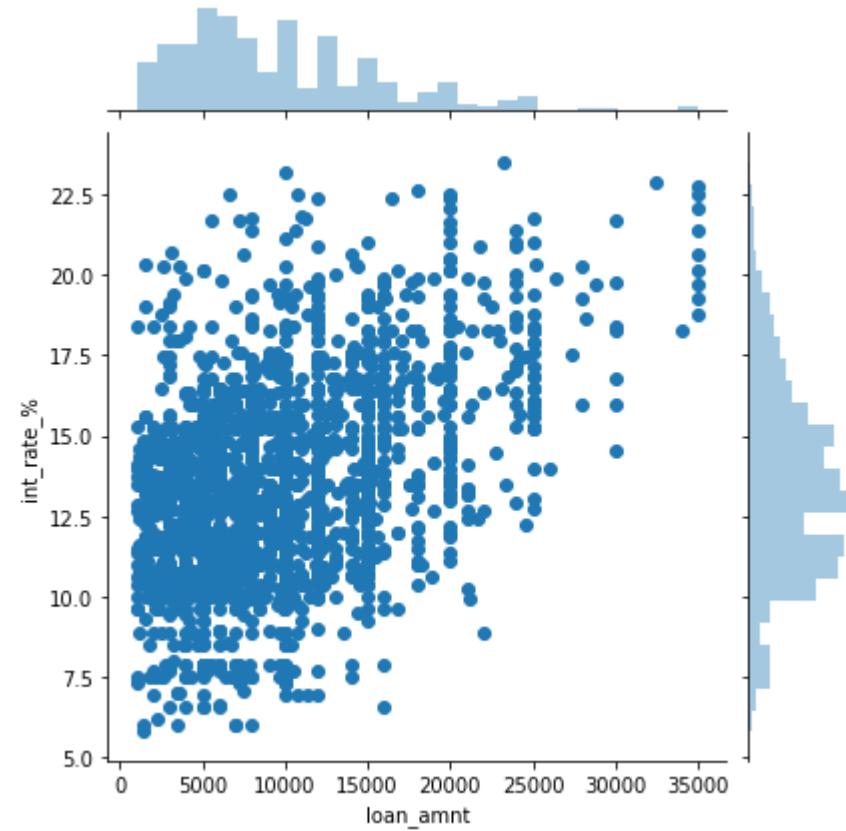
Analysis

- The project is about 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
 - Modify /Create columns necessary to analyse the data
- During **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



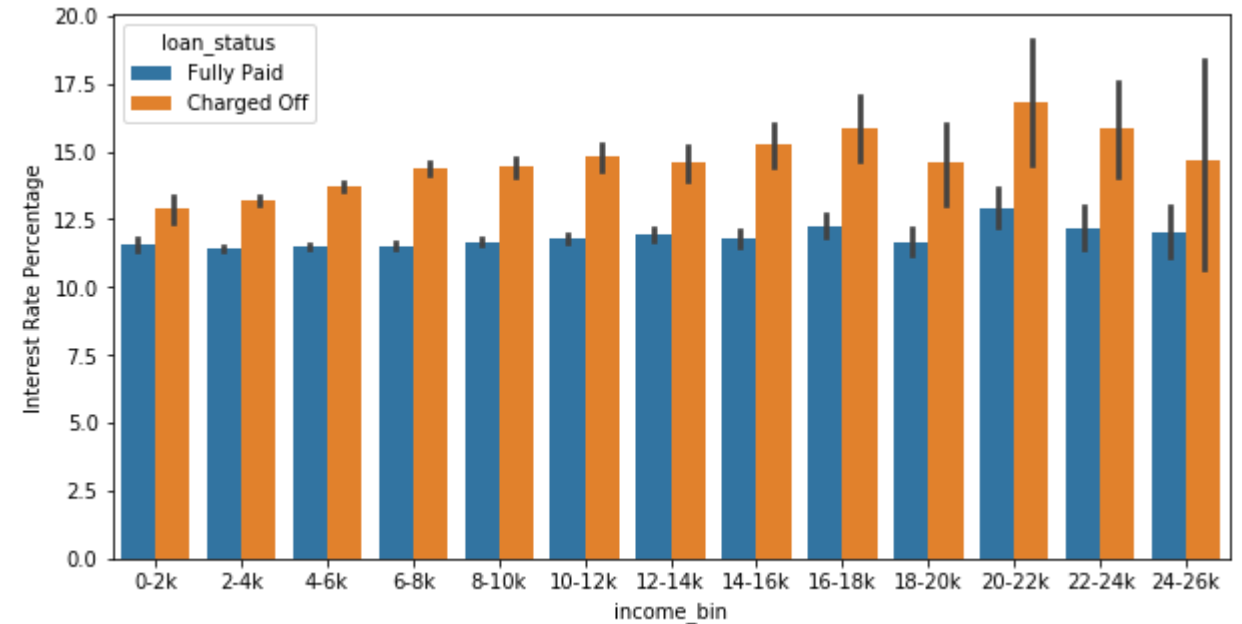
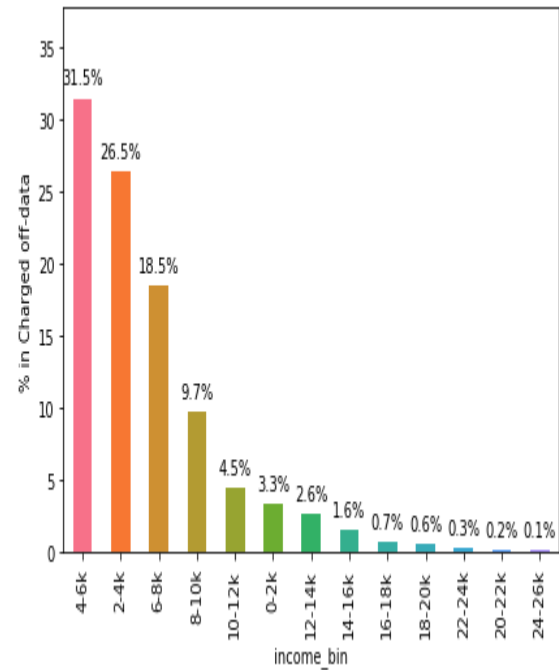
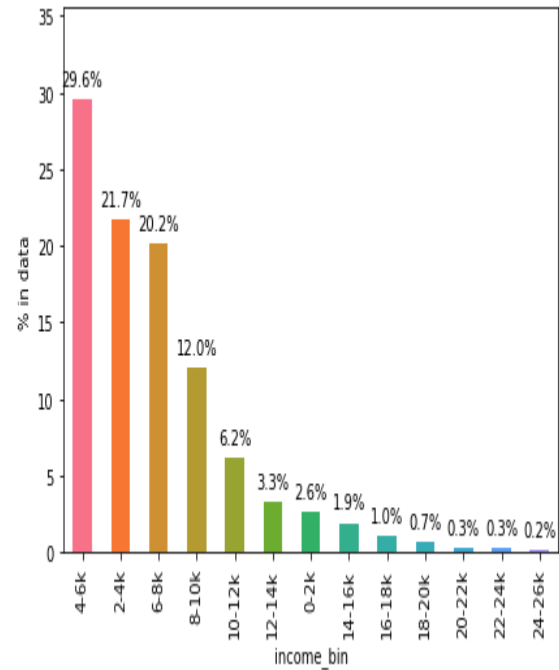
About 14% of the loan are defaulting.



Observation:

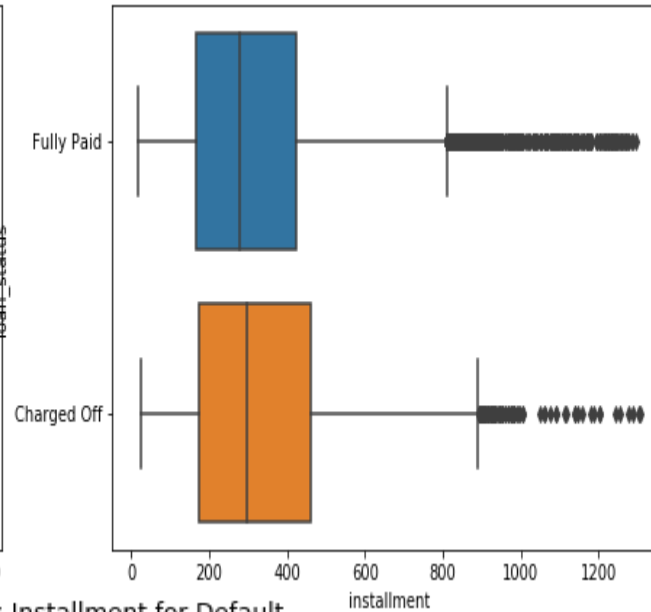
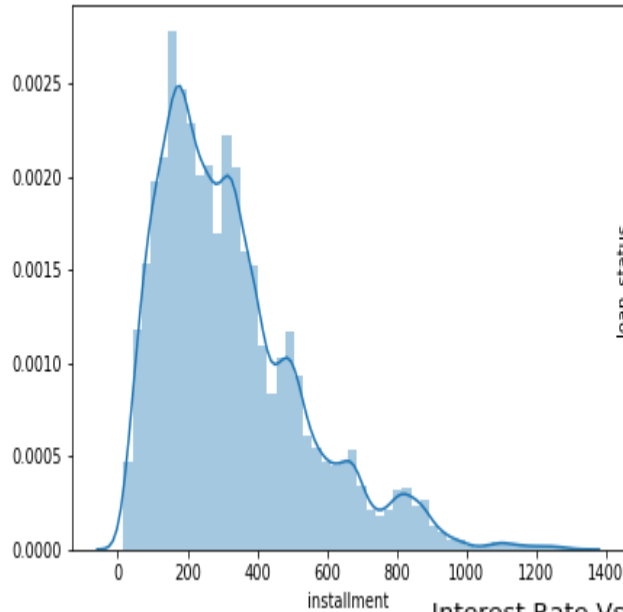
1. As the loan amount increase the interest rate is mostly spread to high interest rate

Loan - Income Vs Interest Rate



Observation:

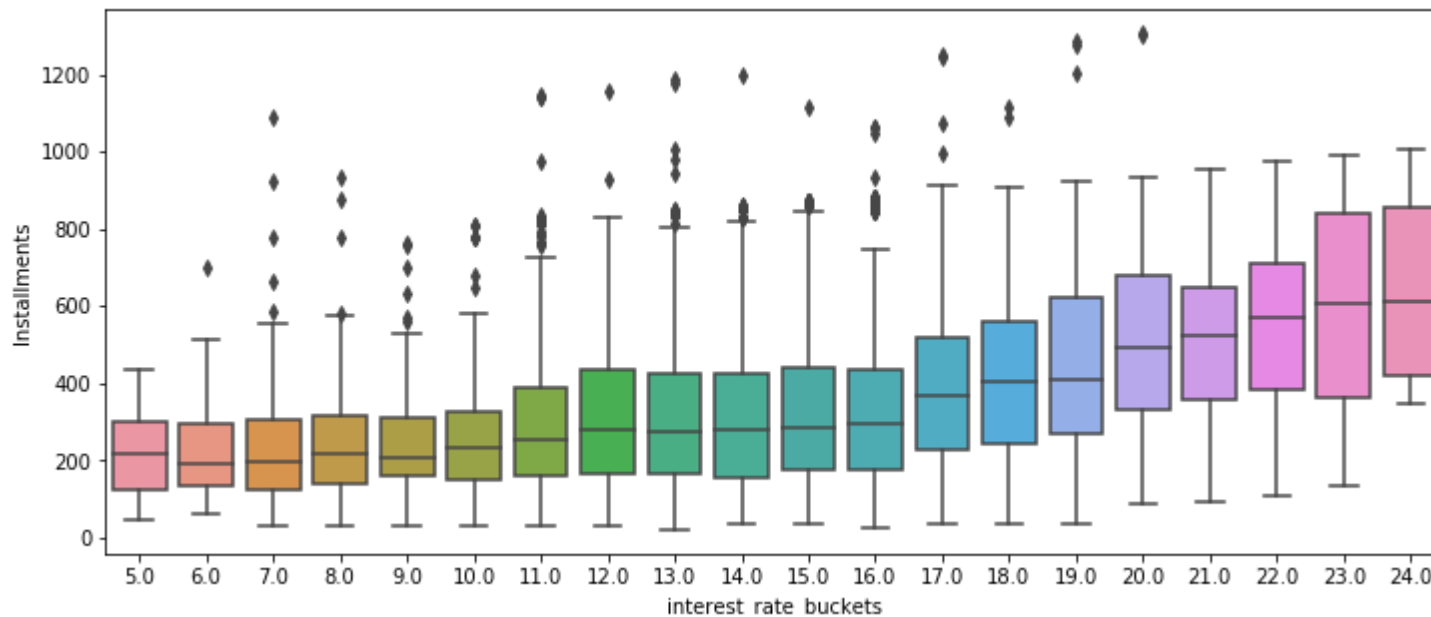
1. Interest rate increases with annual income. Higher interest rate have higher tendency to default the loan.



Observations:

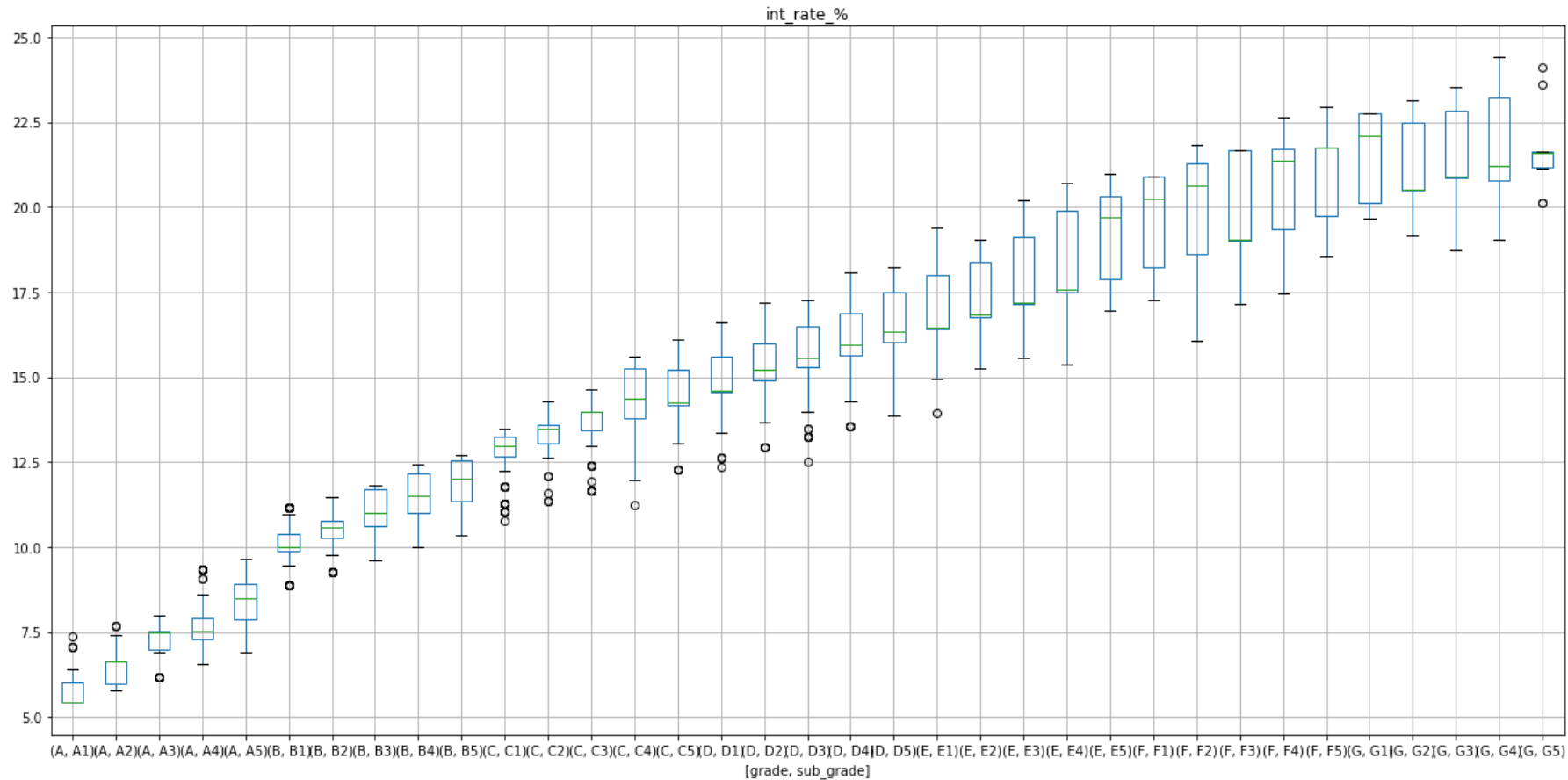
1. Loans are Charged Off have high installment on average
2. Installment increases with interest rate.

Interest Rate Vs Installment for Default



Loan Grade, Sub Grade

Boxplot grouped by ['grade', 'sub_grade']

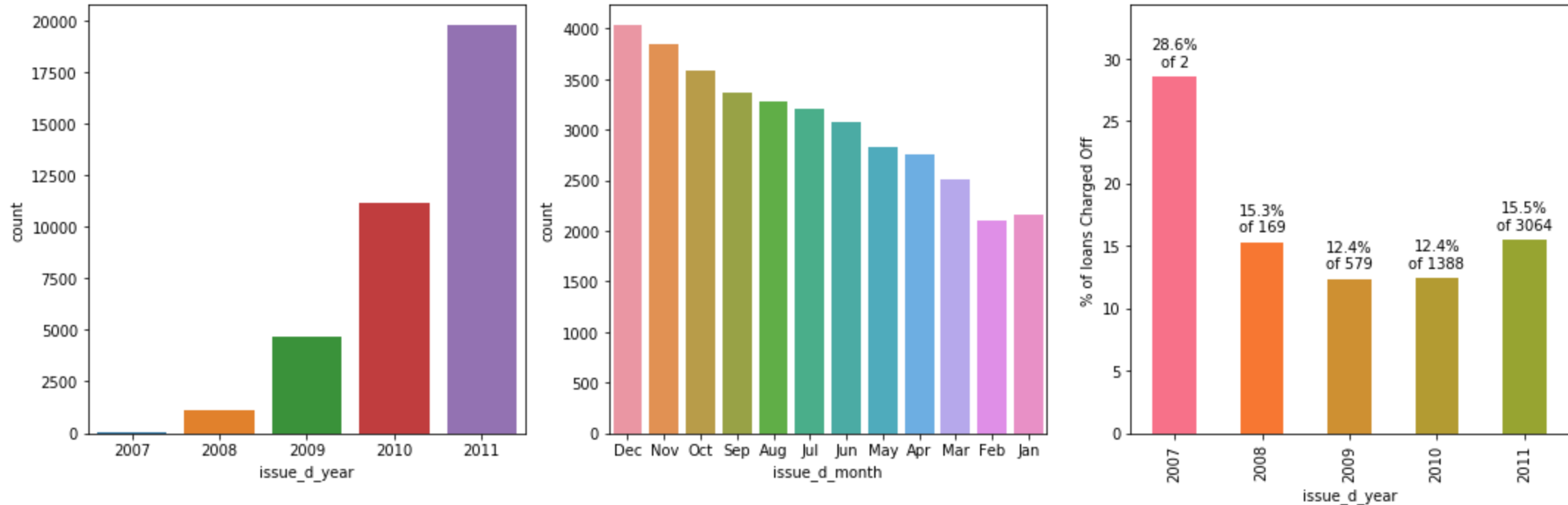


-Note
Grades A is higher
Grade G is lower

Observation:

1. Loans are mostly defaulted for lower grades, subgrades. As the loans at the lower grades have high interest rate and is proportional to installment.

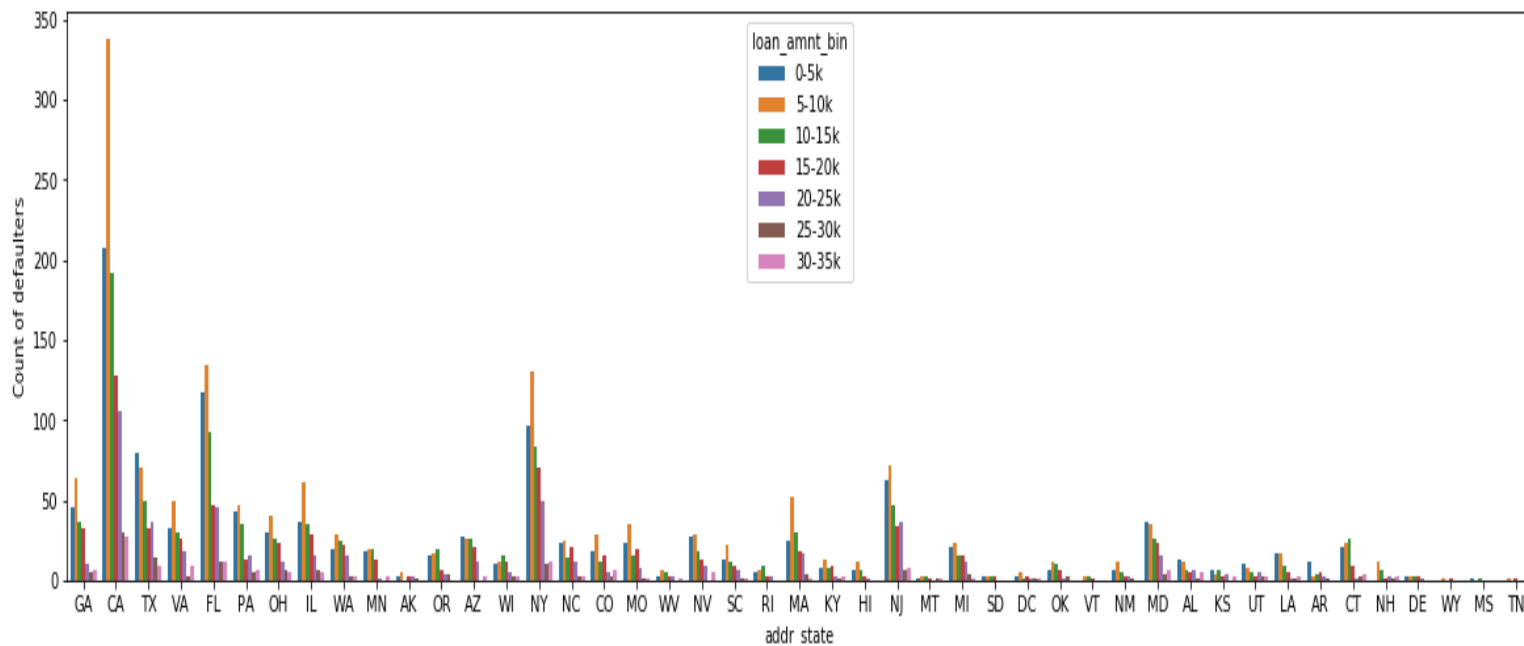
Loan Issuance



Observations:

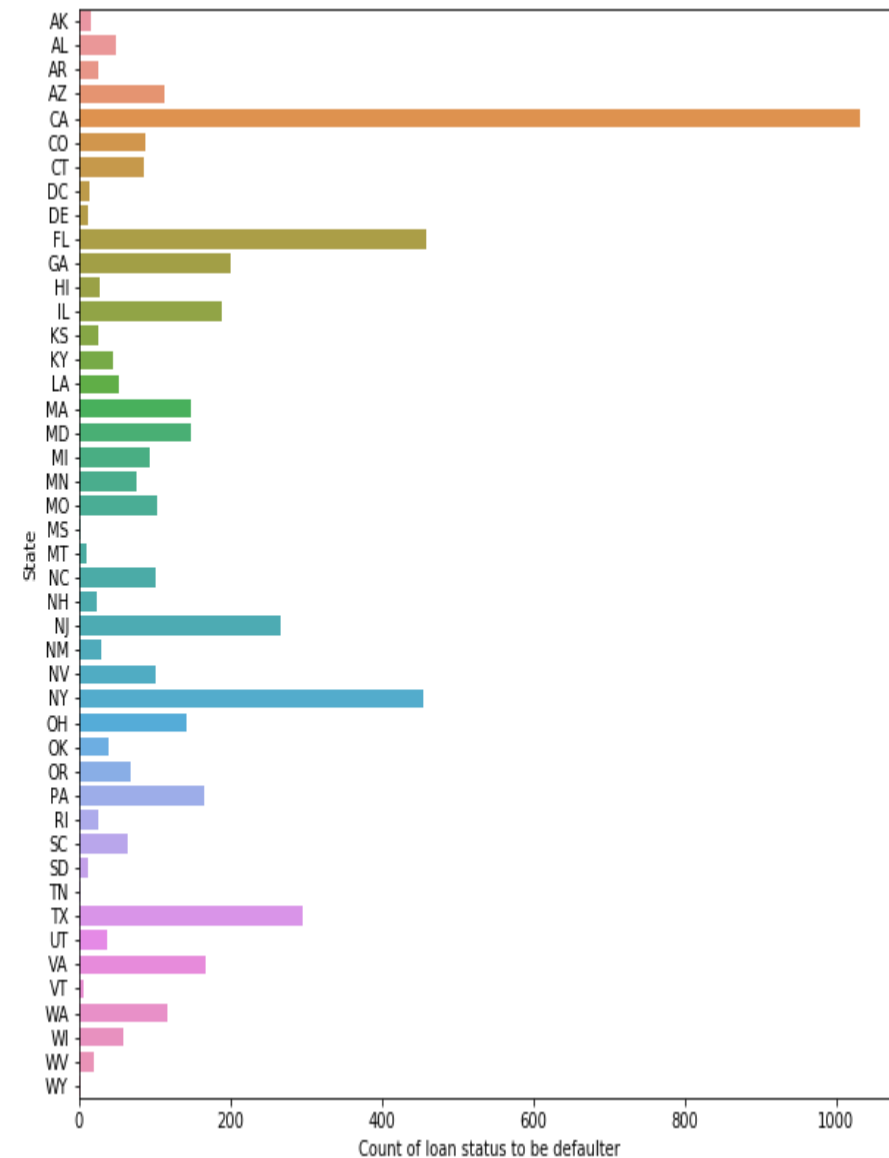
1. LC as an organisation has been giving more loans from 2008 to 2011.
2. The number of loan issued rises over the month from Jan to Dec. December is the month of the year were the maximum number of loans are being issued.

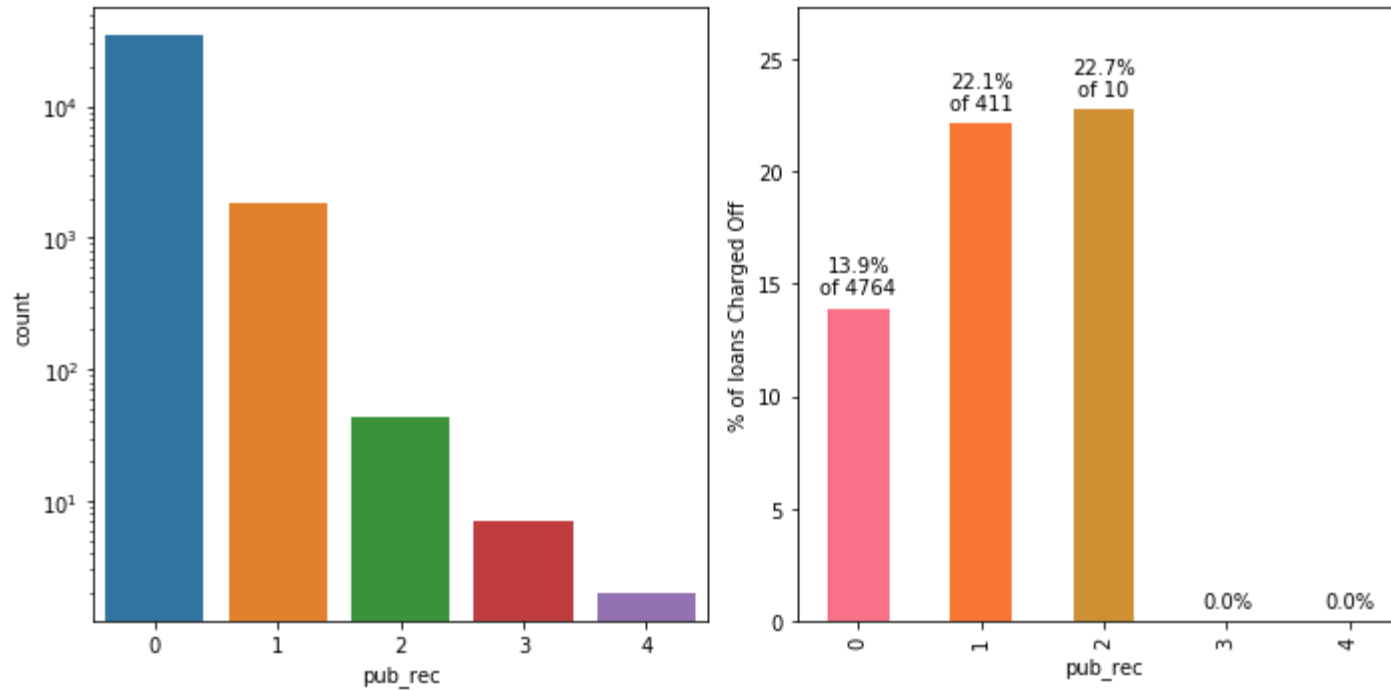
Loan Address-State



Observation:

1. Most of the loans are from California – 18%, Florida – 9%, New York – 7%, as most populated cities.
2. Defaulters are also the in the same place with mostly loan ranges between 5-10k.

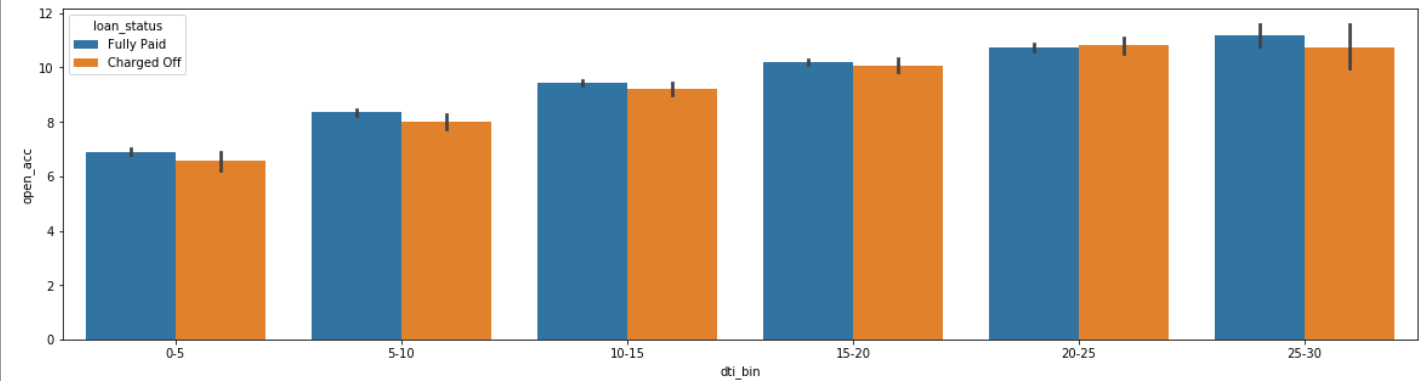
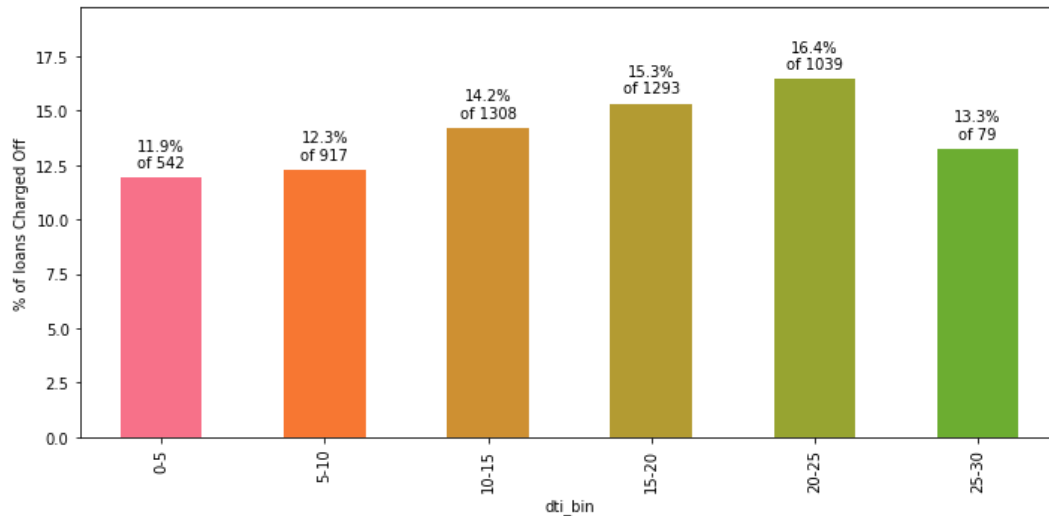




Observation:

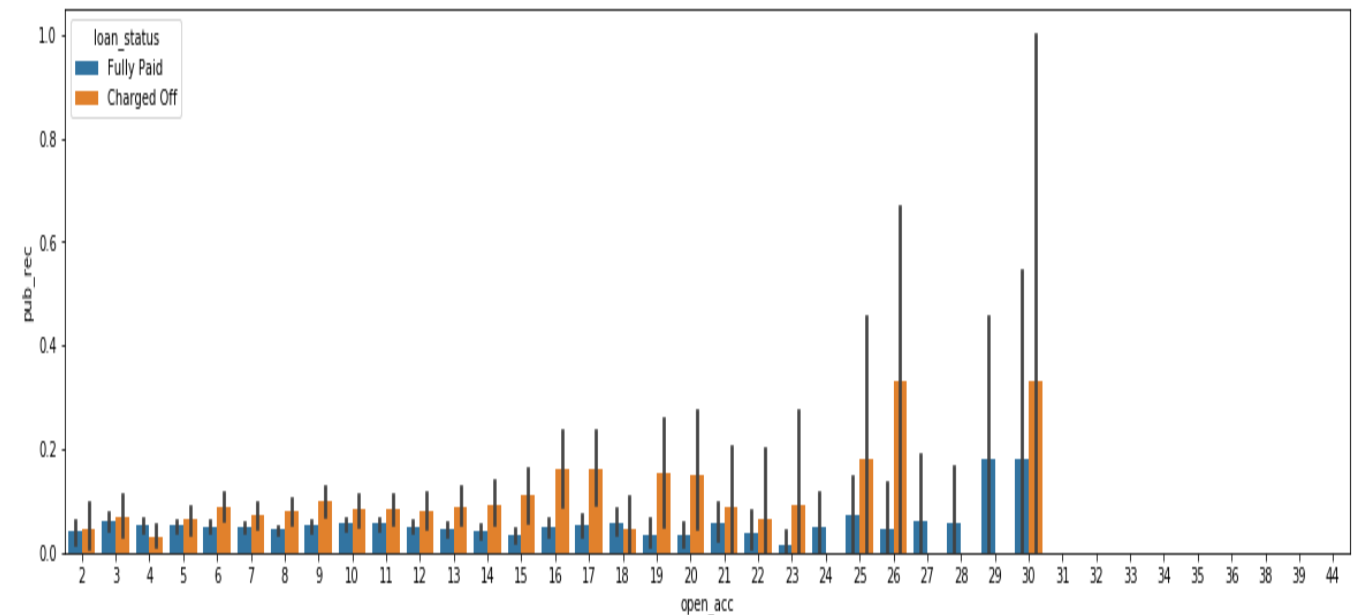
1. 94% have no Public derogatory records. 5% have 1 derogatory record.
2. Having even 1 derogatory record increases the chances of Charge Off significantly.
3. Derogatory public records is highly correlated to pub_rec_bankruptcies

Loan DTI

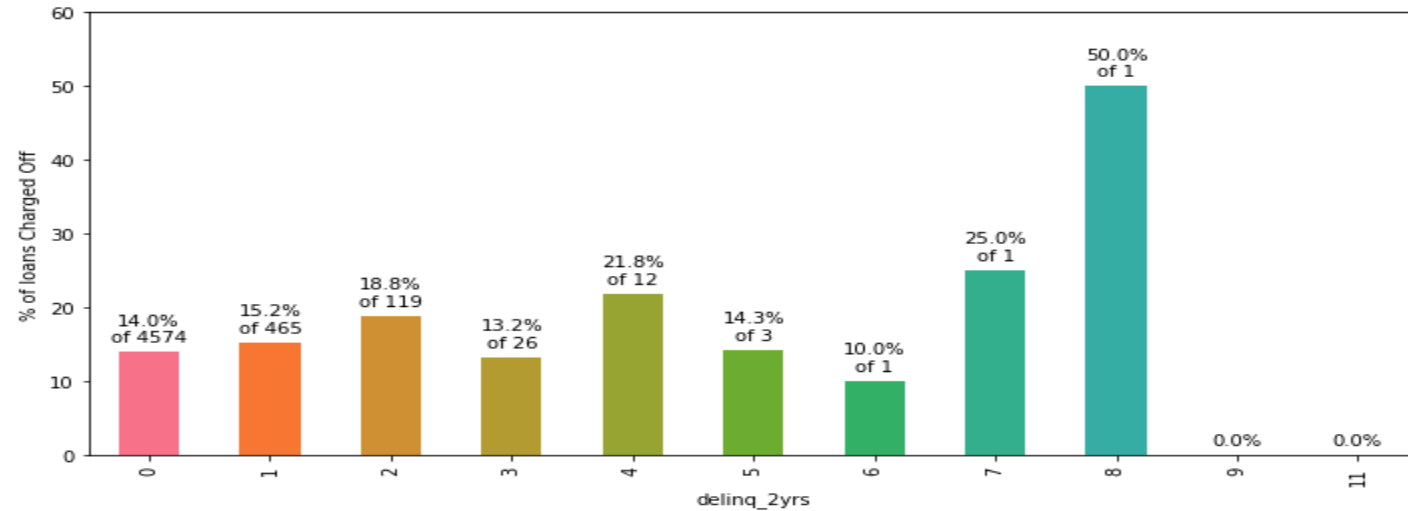


Observation:

1. Higher the dti, more is the chance of defaulting the loan
2. Open_acc is proportional to dti and pub_rec

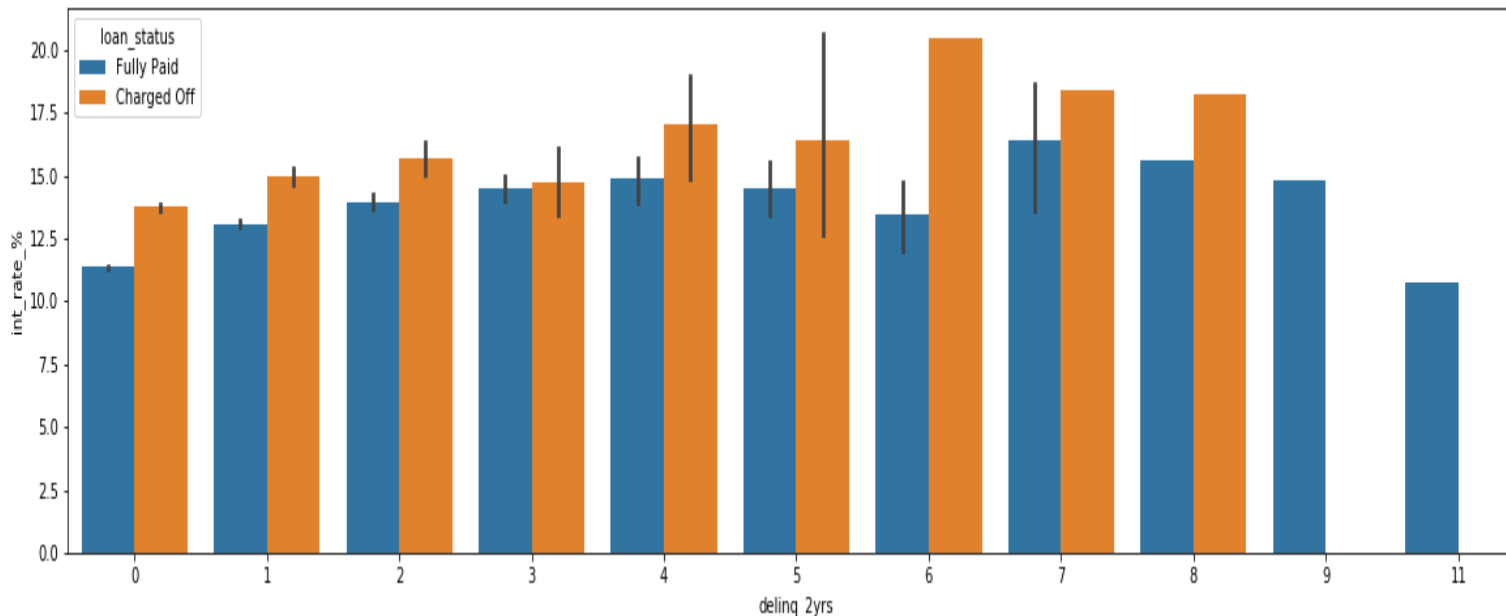


Loan Delinq_2yrs

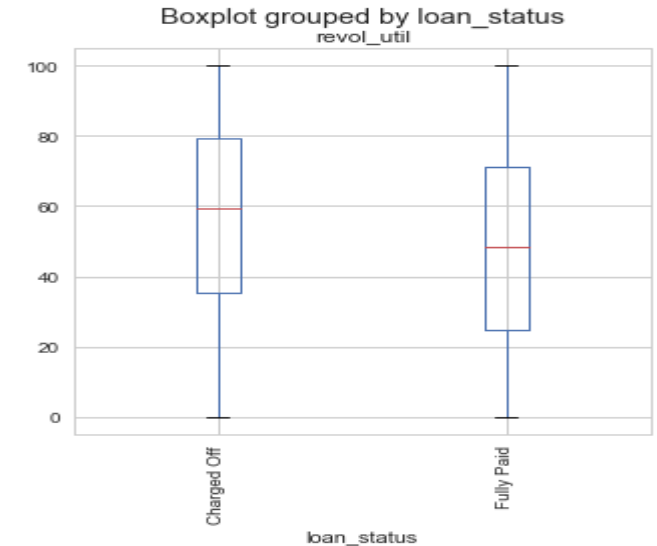
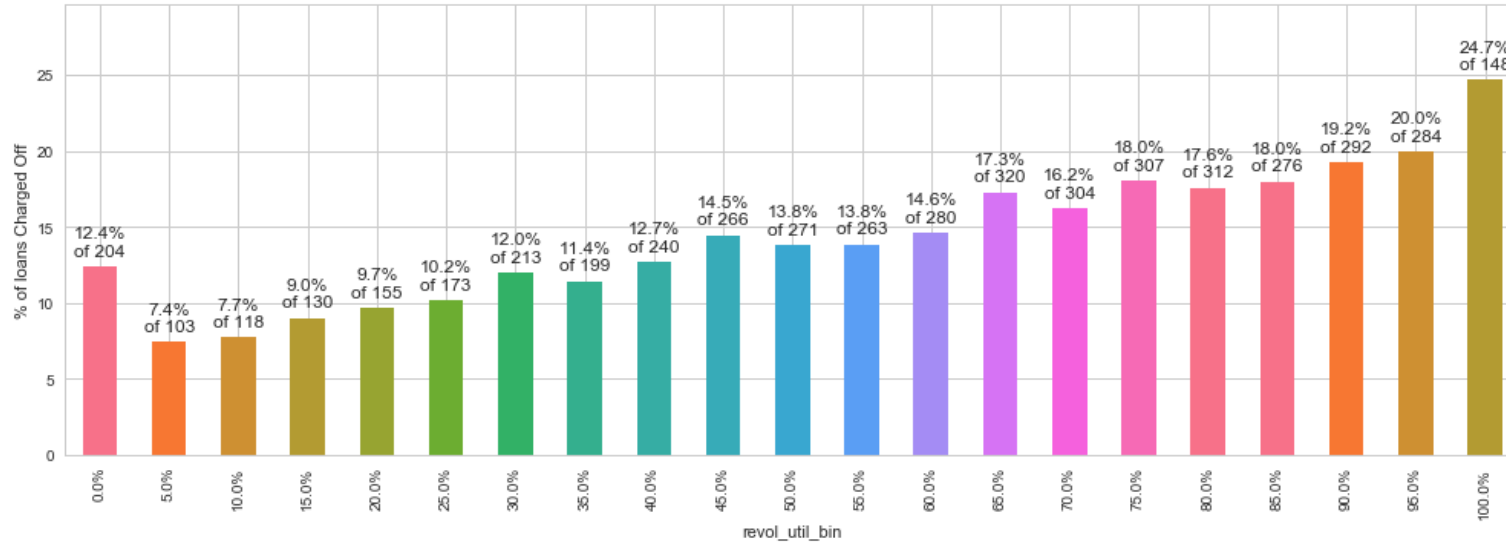


Observation:

1. The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years values ranges between 0 and 11. There are higher chances of default if this variable is greater than 1.

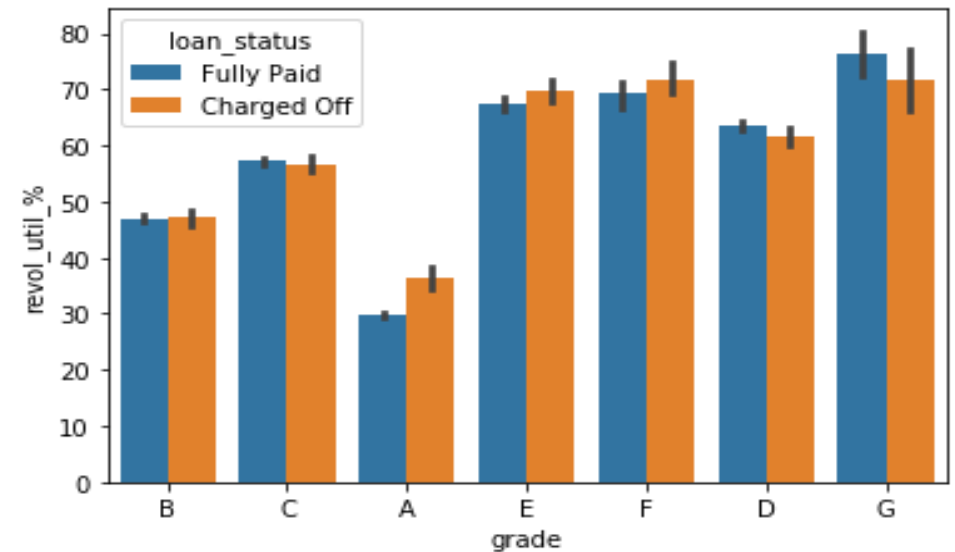


Loan Revol_util Vs Grade

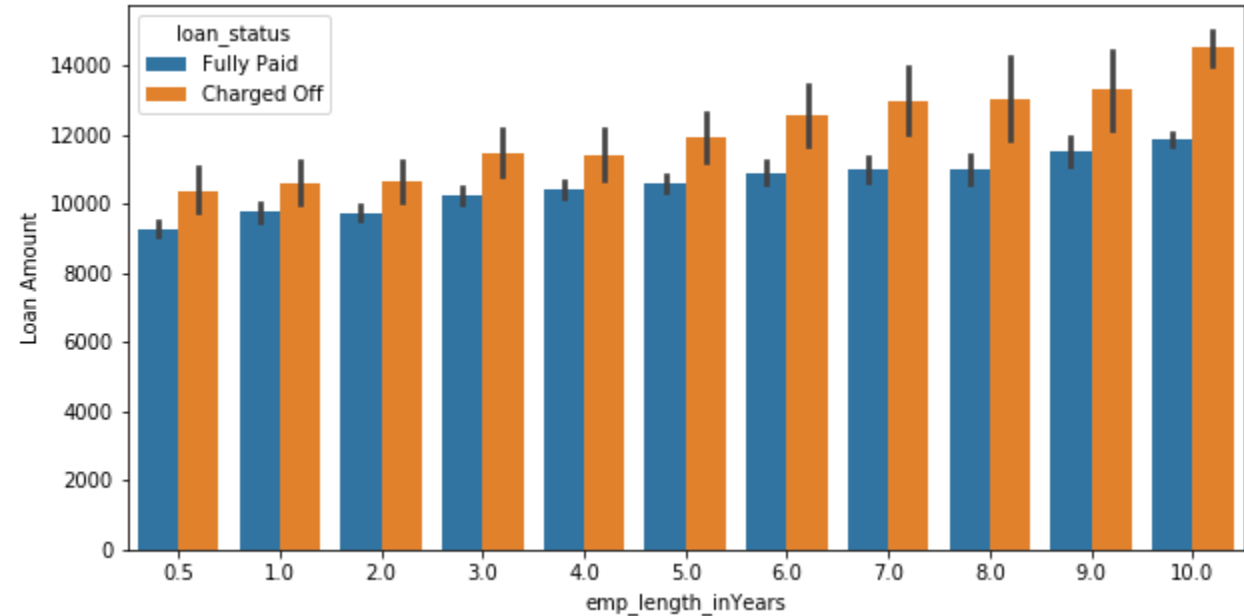
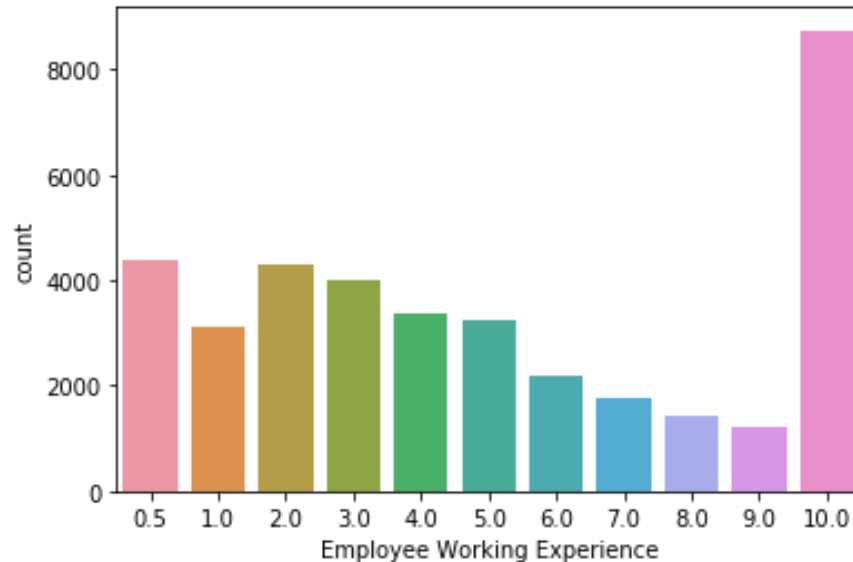


Observation:

1. revol_util and grade(and therefore int_rate) are correlated in some way.
2. The revol_util is positively correlated to the grade. As the grade goes from A to E the revol_util also increases. This may be because higher loan amounts are associated with higher grades.



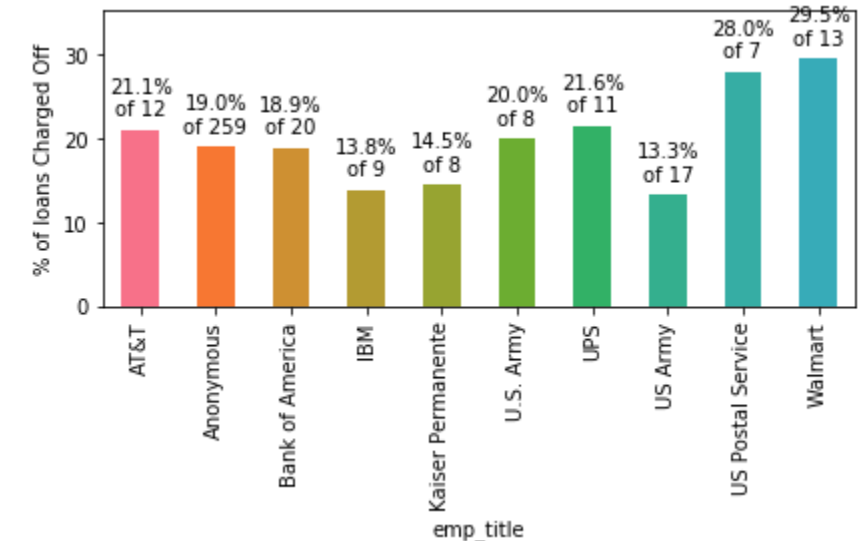
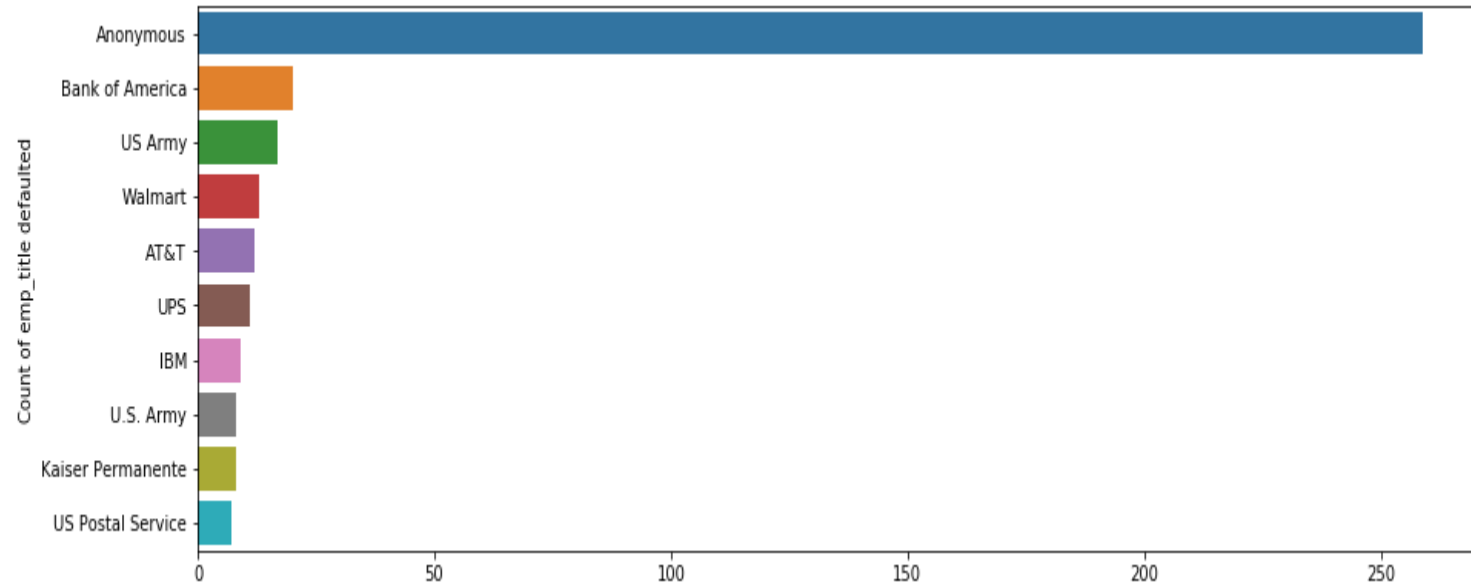
Loan Applicants work Experience Analysis



Observation:

1. Majority of employees applying for the loan have more than 10 years of experience and high loan amount
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 Applicants Company Name Analysis

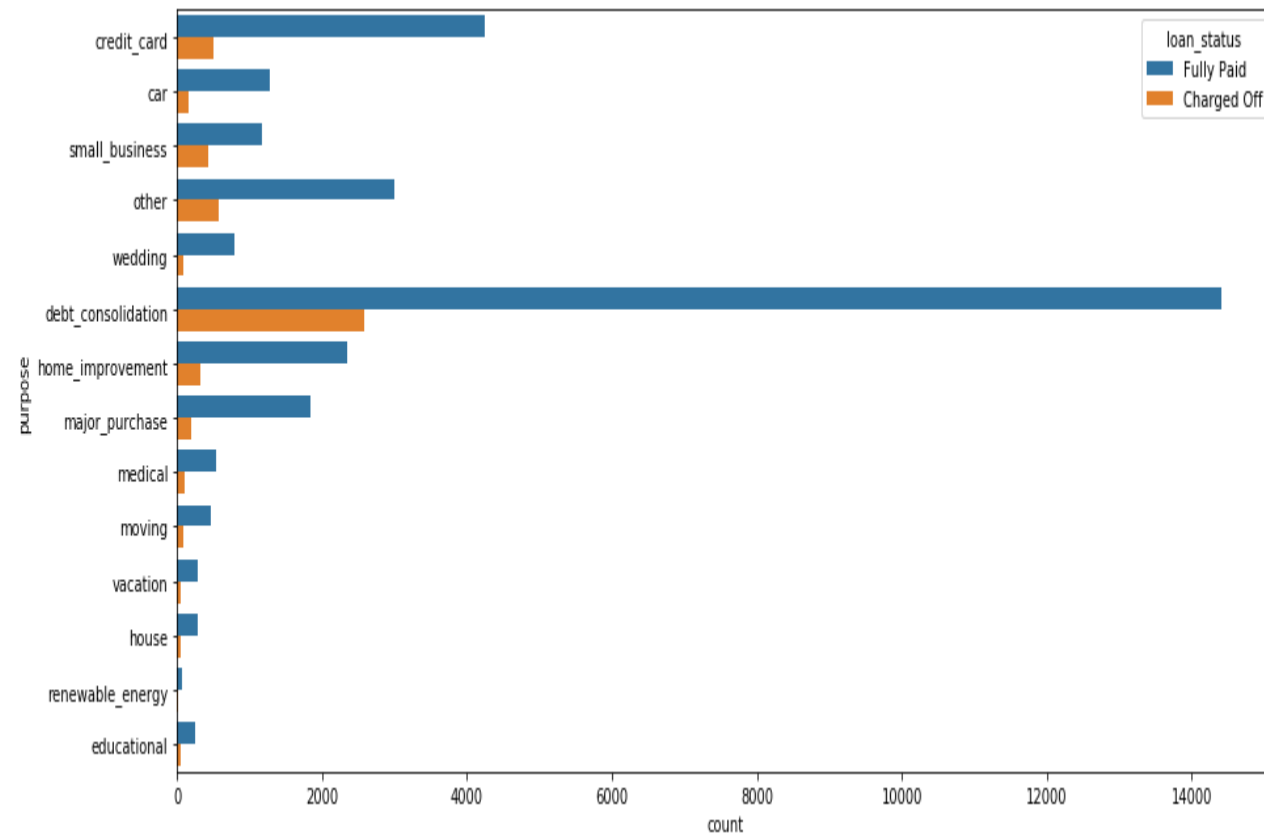


Note: Na values in 'emp_title' are categorised as 'Anonymous'

Observation:

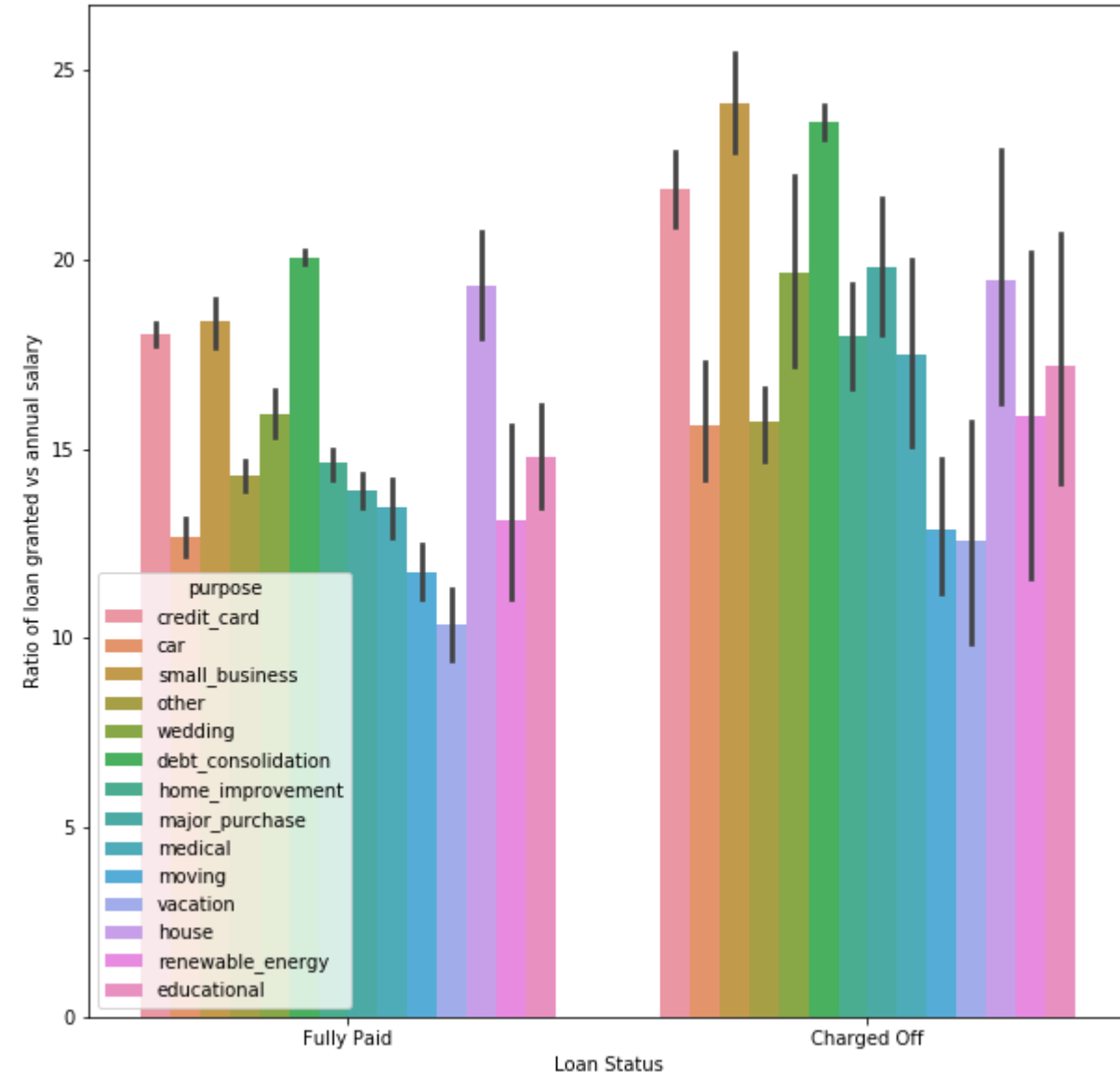
1. Majority of employees applying for the loan are from US Army, Bank of America, AT&T, Walmart.
2. Majority of the loan applicants defaulted does not provide workplace details. So company need to carefully verify.

Purpose for applying Loan Analysis

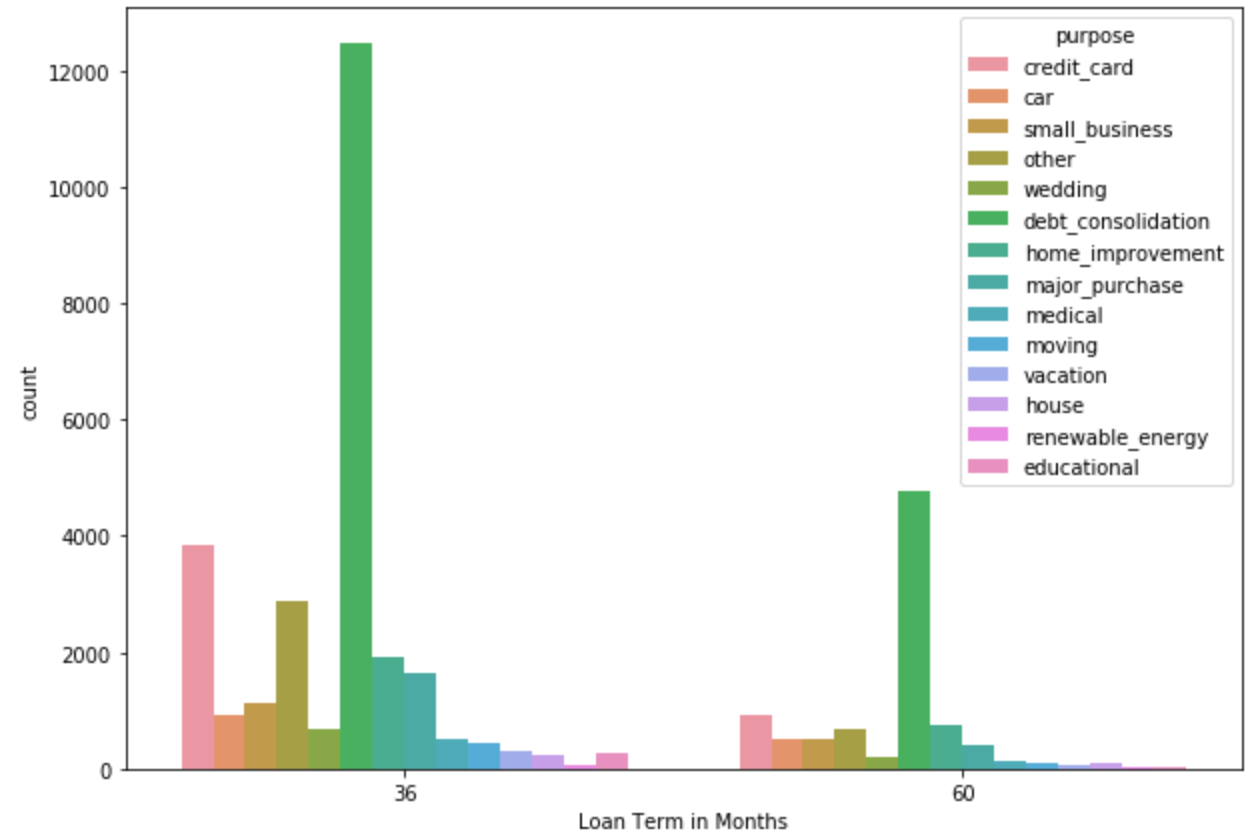
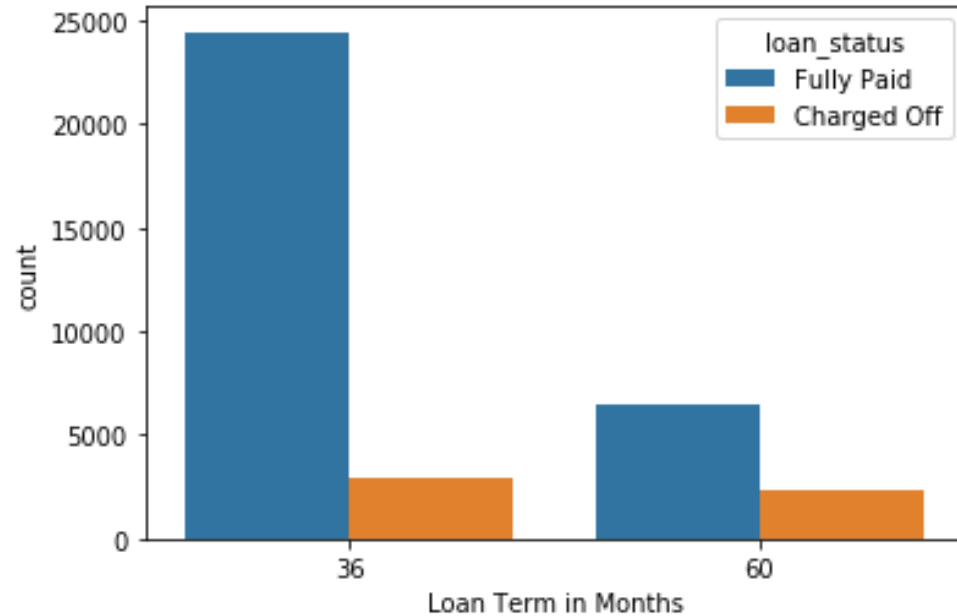


Observation:

1. Most of the loans are granted were for debt consolidation.
2. Applicants having higher income with loan amount tends to default the loan in categories as small business, debt consolidation and credit card.

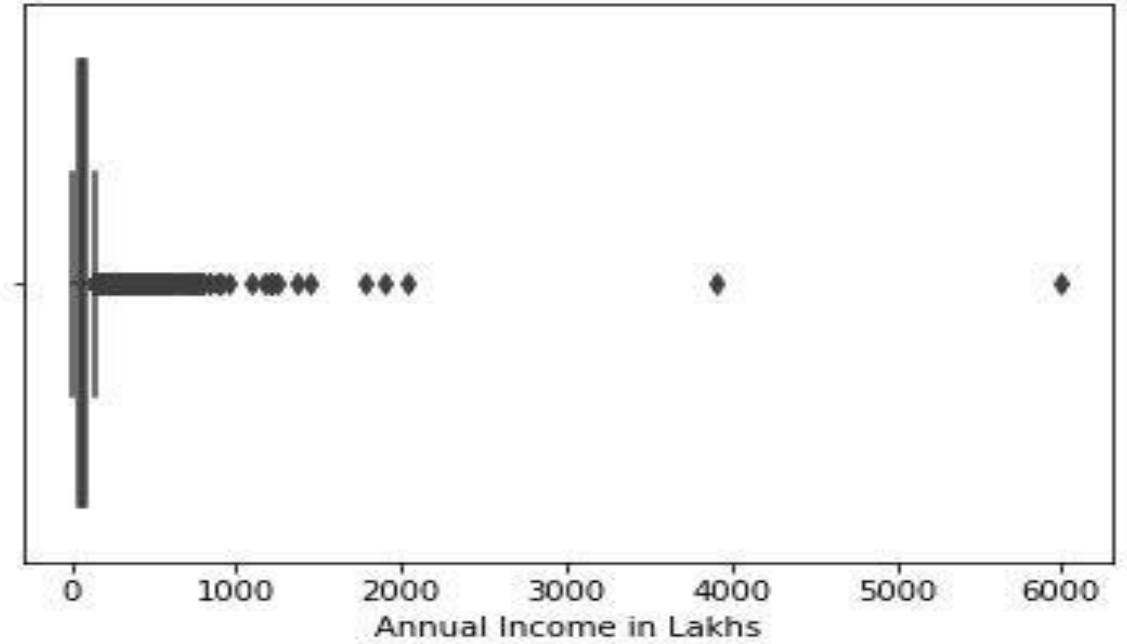
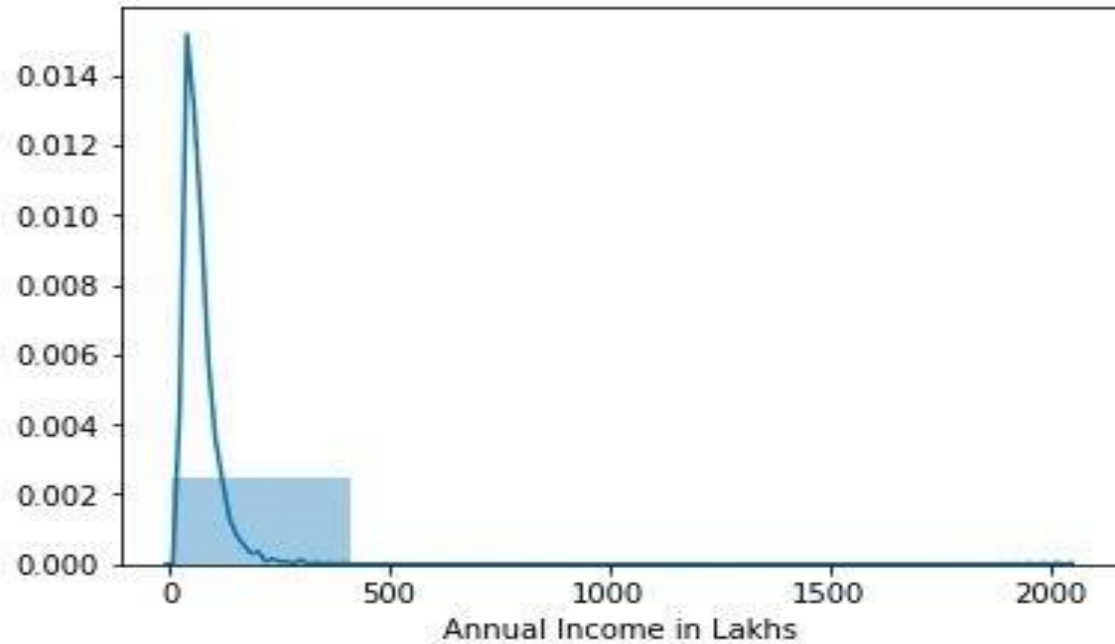


Loan Tenure (months)for applying Loan Analysis



Observation:

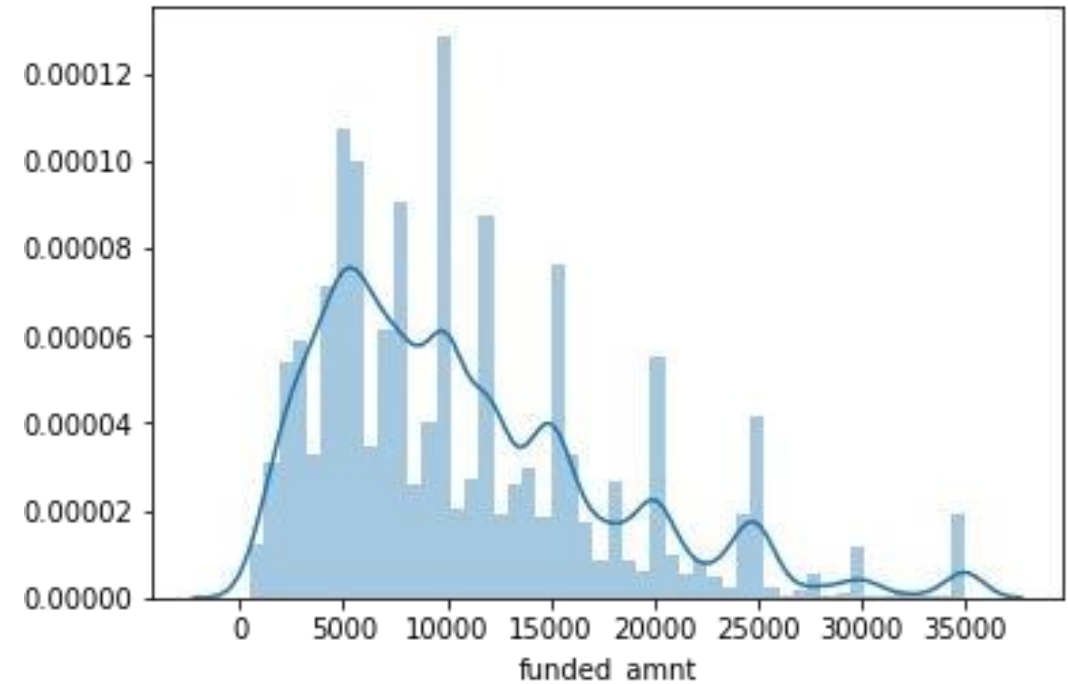
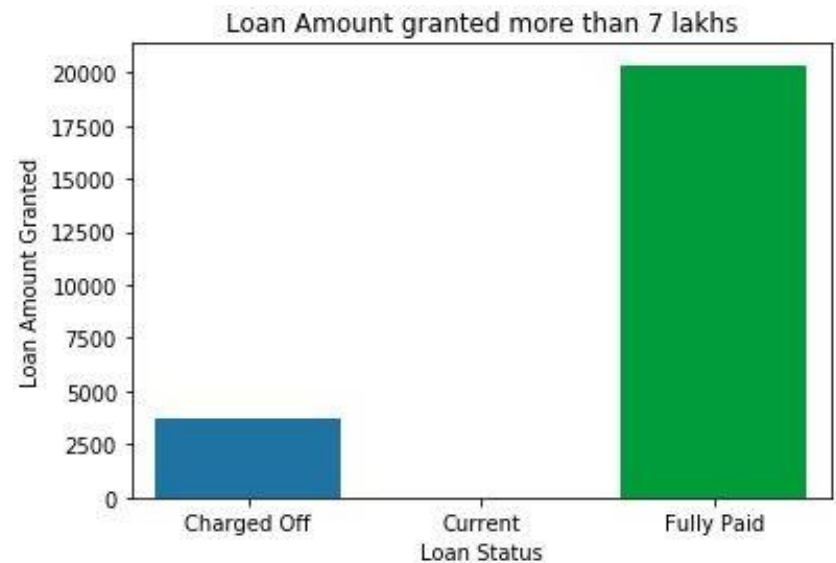
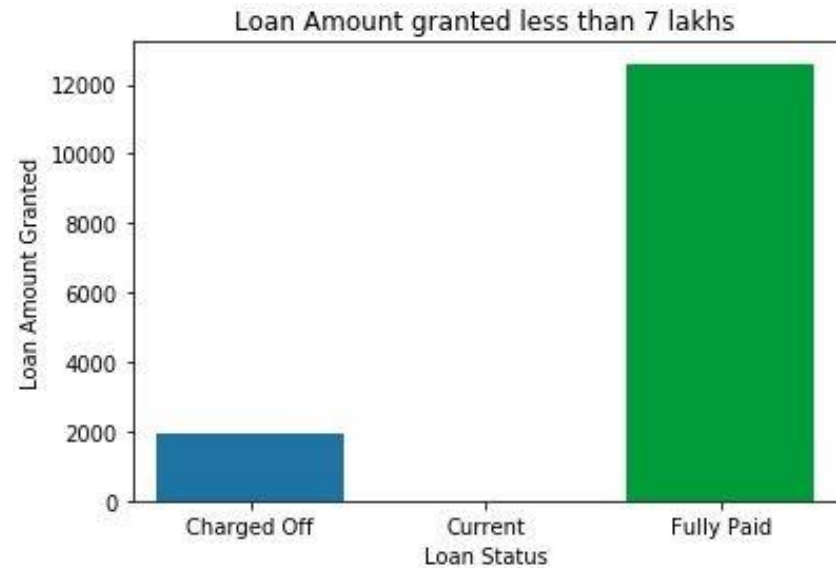
1. Most of the loans granted were of 36 months.
2. Most of the loans having 36 months has high chance to default.
3. Additionally, Debt Consolidation is the most common purpose and loans are granted for 36 months and 60 months.



Observation:

1. Most of the applicants have income between 0 to 5 lakhs.
2. There are two applicants having income above 30 lakhs

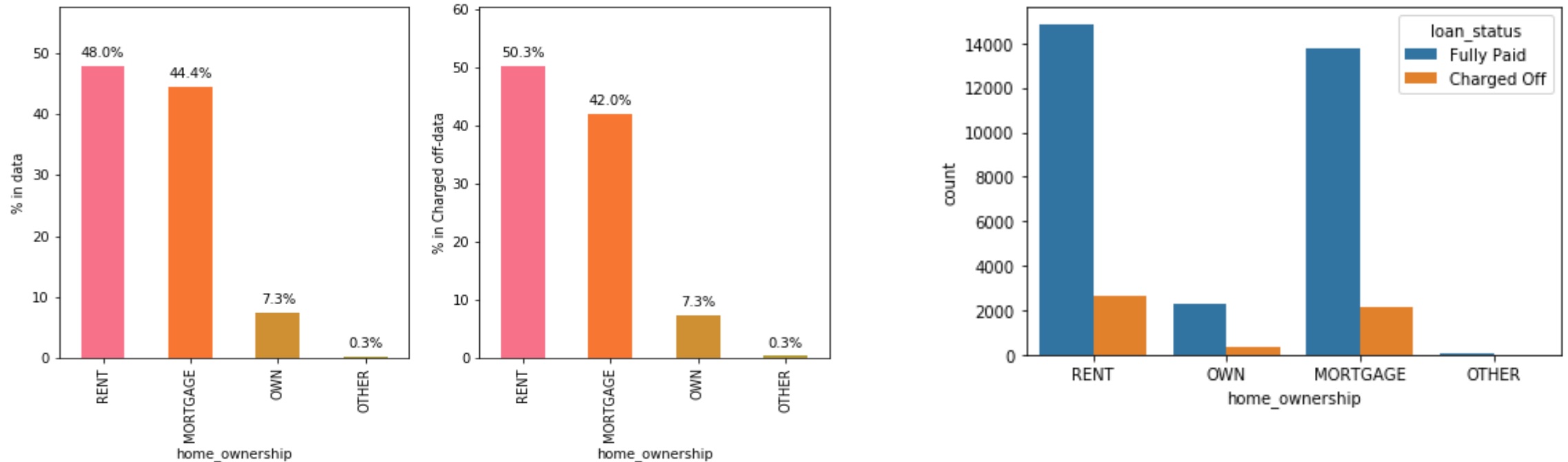
Loan Amount Analysis



Observation:

1. Funded amount is left skewed. Most of the loan amount given is below 7 lakhs.
2. Probability of people with loan amount greater than 7 lakhs tends to default is more than the people with less than 7 lakh

Loan Status by Home Ownership



Observation:

This parameter for the defaulters and the total loan takers resembles same pattern, which points to loans are mostly taken by people on Mortgage & Rent.

It's not a apt to predict that Rent & Mortgage applicants are defaulted more than people owning a place.

Conclusion

Results

1. Low grade loans have high tendency to default. Considering grade A, B, C and so on are higher to lower grade.
2. Loans having higher interest rate have more defaulters
3. Extra checks should be conducted for CA state candidates, as they have high tendency to default.
4. More number of open accounts will lead to a higher probability of increasing DTI and derogatory public records which in turn increases chances of defaulting.
5. Increase in number of delinquent cases in the past will increase the probability of defaulting.