Lending Club Case Study

Group Members:
Akash Sharma
Mayank Bajaj

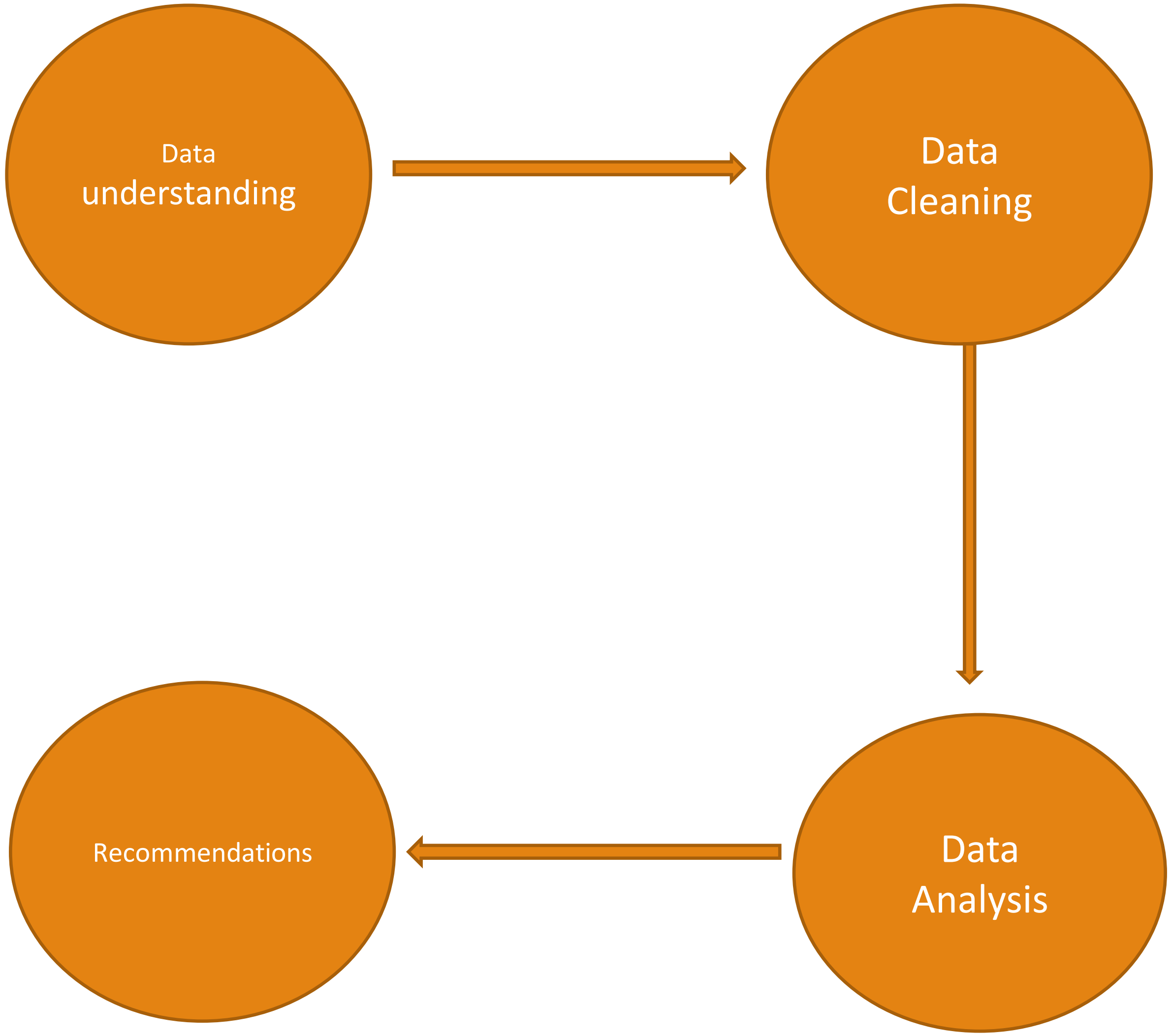
Business Objective

Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters’.

If we are able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Steps for Analysis



Data Understanding

The business objective is to take a decision whenever they receive a loan application whether to reject or approve based on certain variables and recommendations.

Dataset Details:

The data given below contains information about past loan applicants and whether they have been ‘charged off’ or not. Data has details regarding approved loans only, not the rejected ones. It has 3 status of loan which is Fully Paid, Current and Charged-Off.

Current loans wouldn’t help us in identifying any behavior, so only Charged-Off and Fully-Paid data can contribute to our analysis.

Data also contains some behavioral information that one only can possess after loan approval, so this data also wouldn’t help us in identifying any pattern for loan defaulters.

Data Cleaning

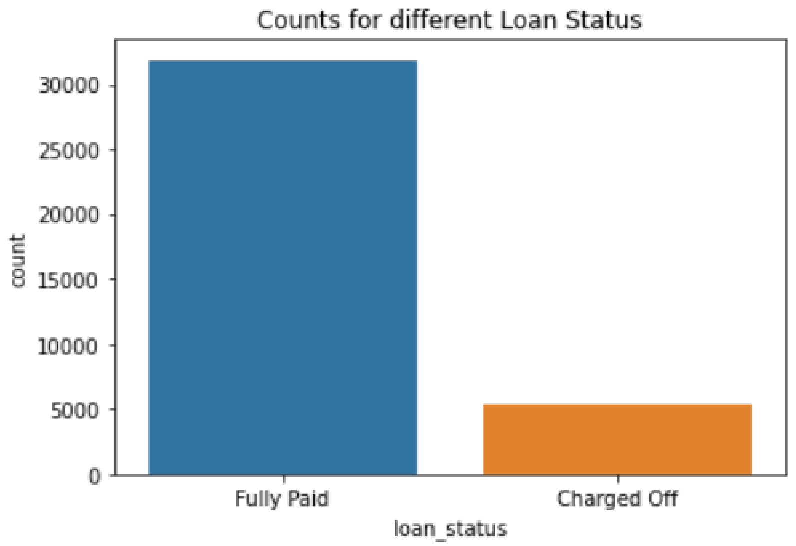
- Ø Checking Columns with only Null Values and removing them
- Ø Checking columns with more than 50% Null Values and Removing them
- Ø Dropping Columns that only contain a single value for the entire column
- Ø Dropping behavioral columns that can only be calculated once loan is approved/irrelevant such as delinq_2yrs,earliest_cr_line,inq_last_6mths, etc.
- Ø Dropping columns which has masked information and those which can’t contribute to analysis based on domain knowledge.
- Ø Fixing Null values in remaining columns and creating new columns for Analysis.
- Ø Removing outliers

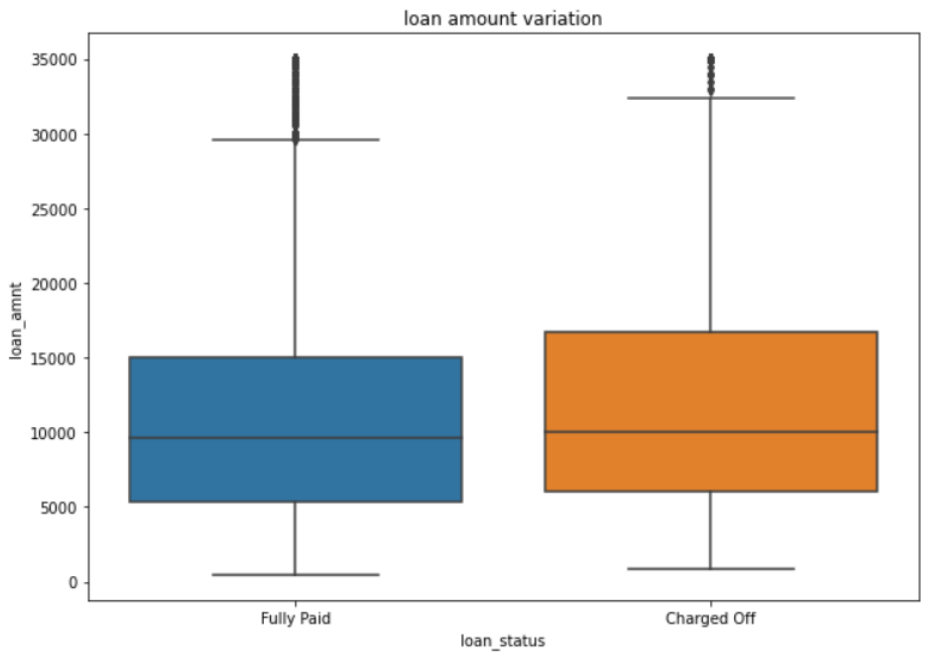
Data Analysis with Visualization

Loan Status and Amount

Loan Status: The number of charged off loan is much smaller(14.5%) compared to total count.

Amount: It varies from 500 to 35000 with a mean of 9800. Loan amount is majorly small and very few clients have taken large loans and larger it goes we have higher chance of defaulting.

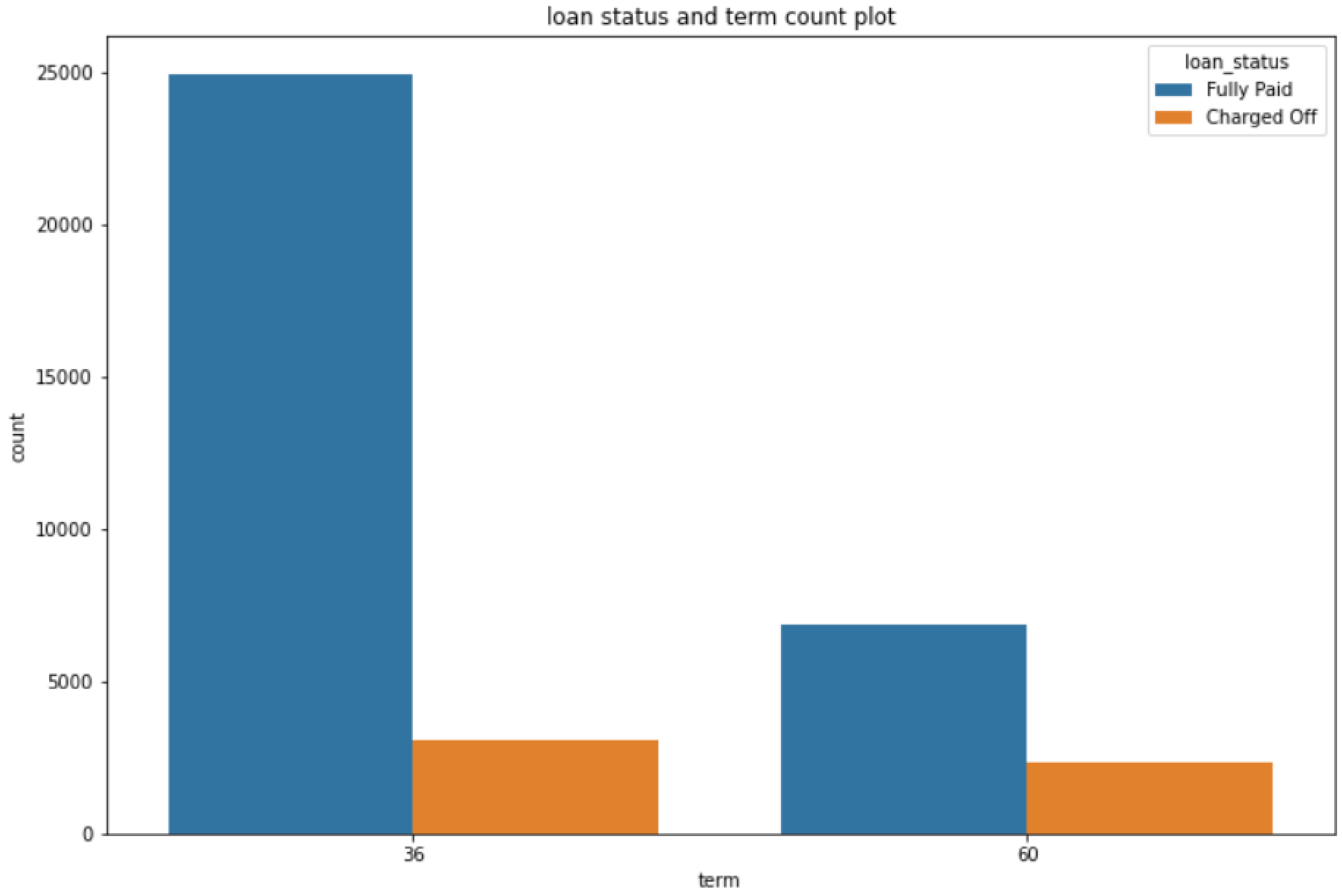





```
count    37173.000000
mean     11034.824335
std       7272.289190
min        500.000000
25%       5400.000000
50%       9800.000000
75%      15000.000000
max      35000.000000
Name: loan_amnt, dtype: float64
Mean loan amt loan_status
Charged Off    12185.797575
Fully Paid     10840.903090
Name: loan_amnt, dtype: float64
median loan amt loan_status
Charged Off    10000.0
Fully Paid      9600.0
Name: loan_amnt, dtype: float64
max loan amt loan_status
Charged Off    35000
Fully Paid     35000
Name: loan_amnt, dtype: int64
min loan amt loan_status
Charged Off    900
Fully Paid     500
Name: loan_amnt, dtype: int64
```

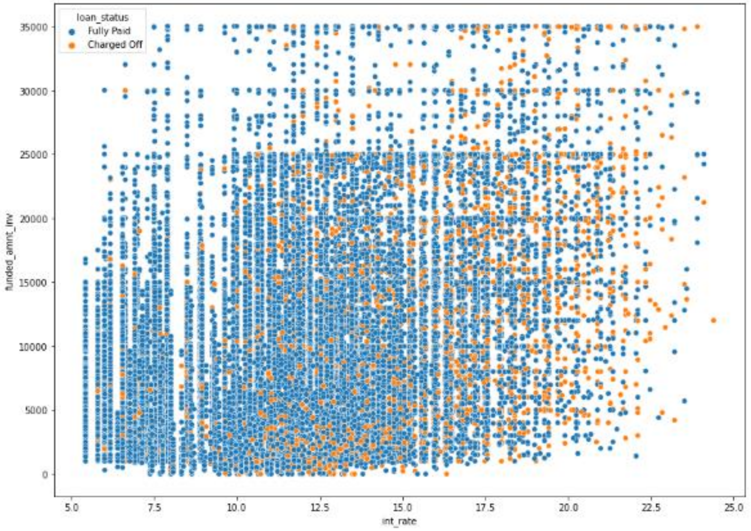
Term

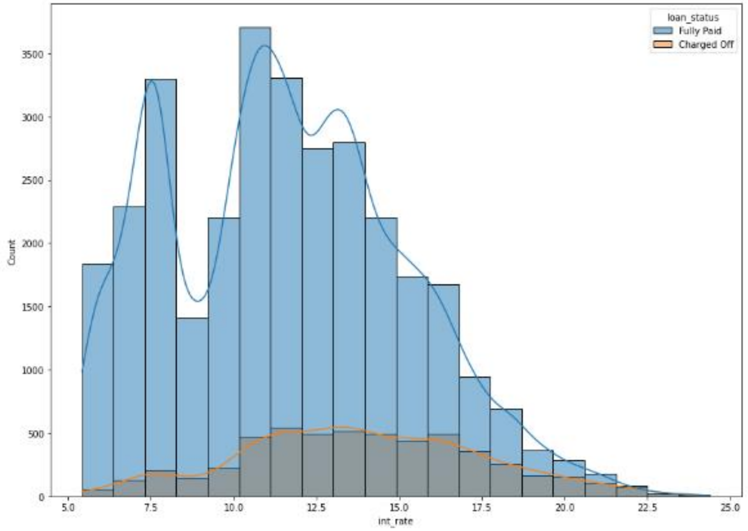
The Loans taken for 36 month term are much more than 60 months and have lower chance of defaulting.



Interest Rate

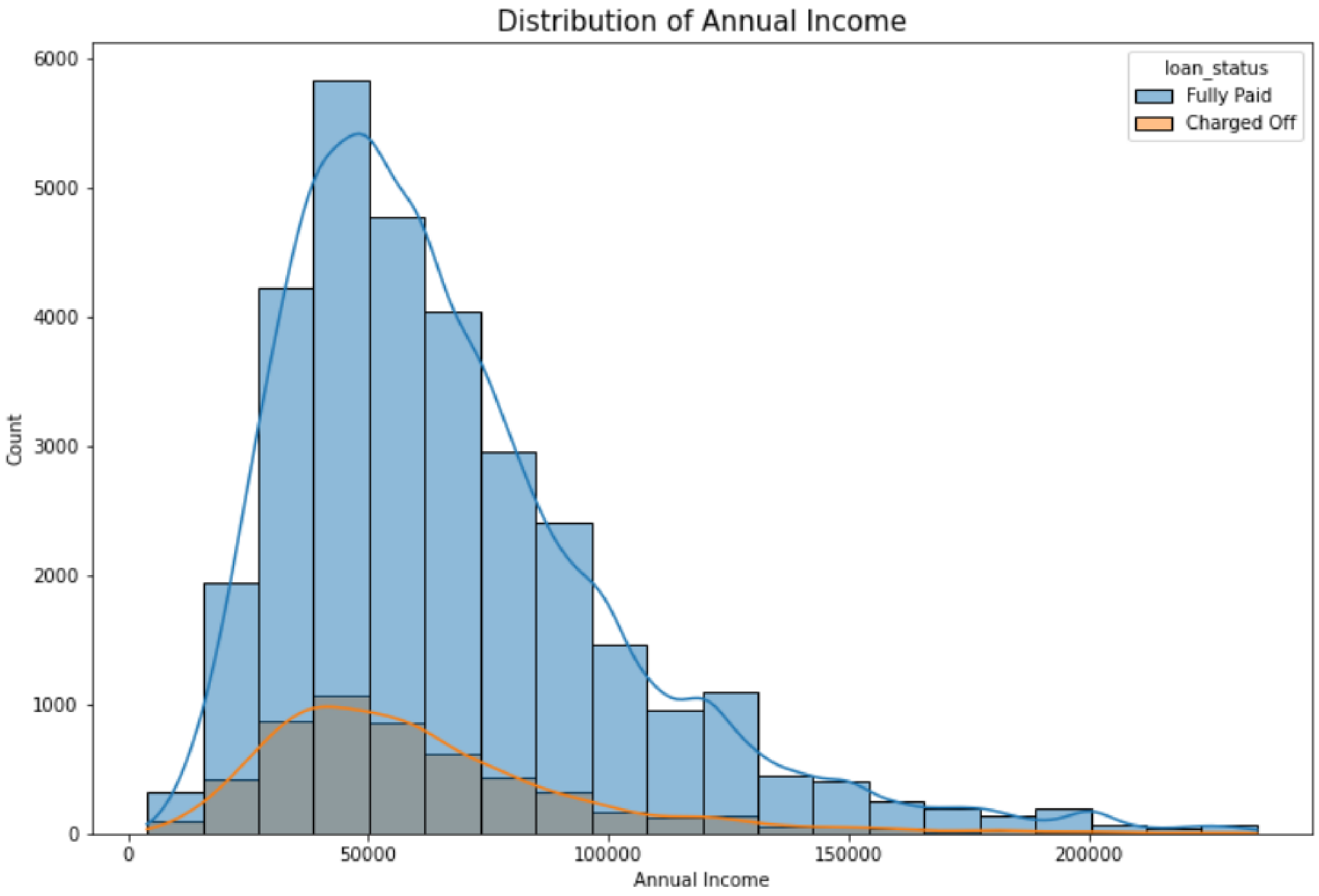
- The Number of loans Taken is higher for the 7.5% - 15% bucket
- The chances of approving a riskier loan increase with higher interest rate
- It shows a huge charged-off chunk after 10% irrespective of loan amount and the percentage of charge-off loans increases with higher interest rates





Annual Income

Applicants with less than 100k annual income are more likely to default and higher annual income are less likely to a defaulter

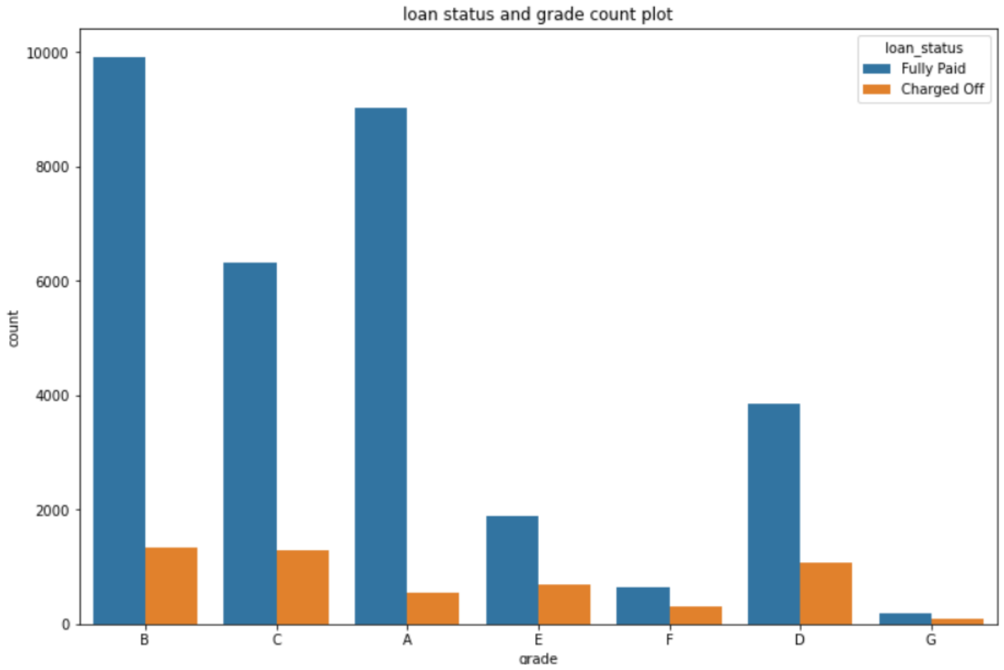


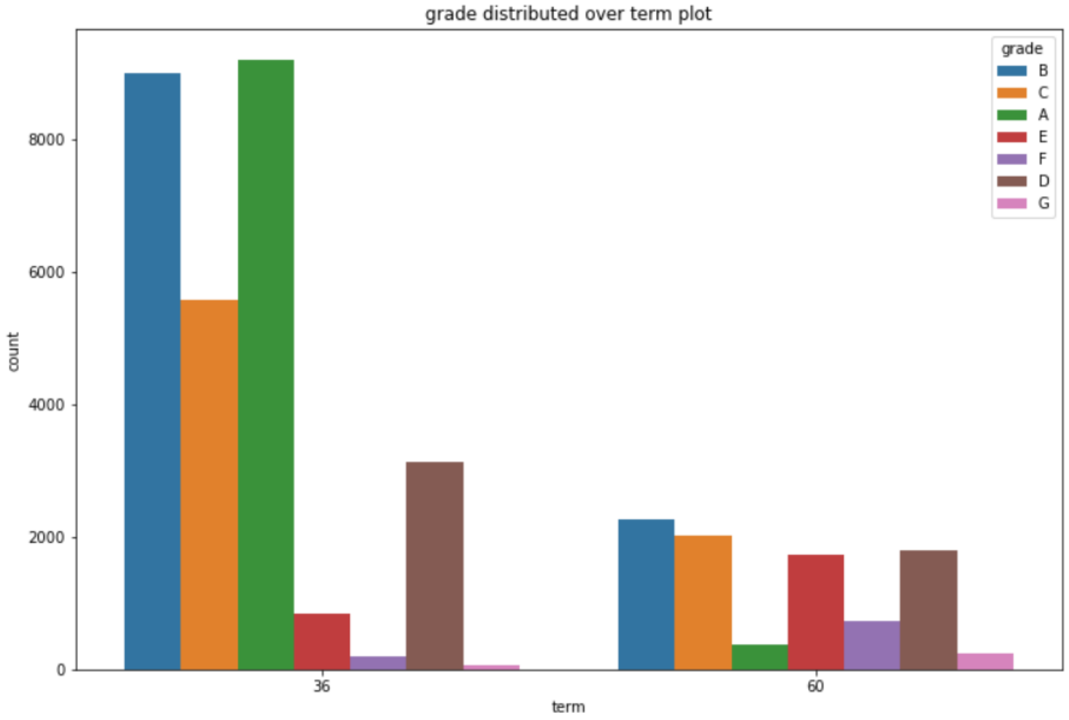
Grade

The loan approved is major of a higher grade as they are of low risk thus low chance of defaulting.

60-month term loans have less number of lower-grade loans with high risk.

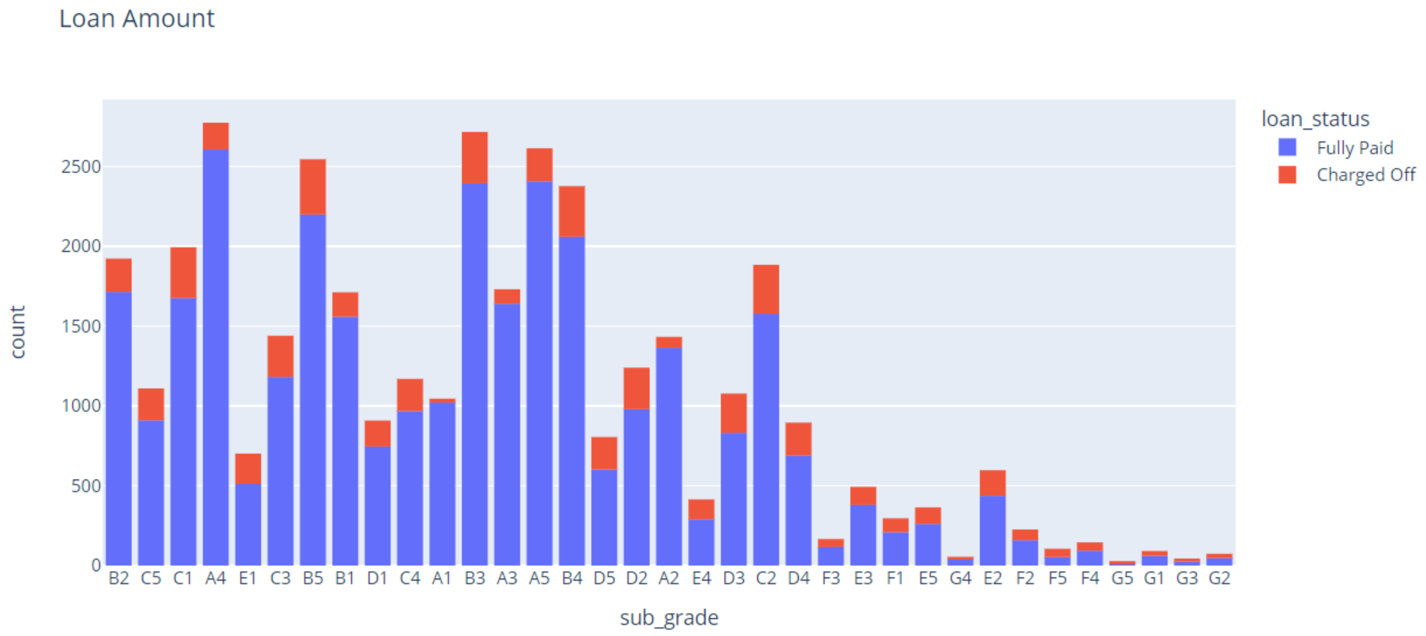
As we move from Grade A to G, the probability that a person will charge off is increasing.

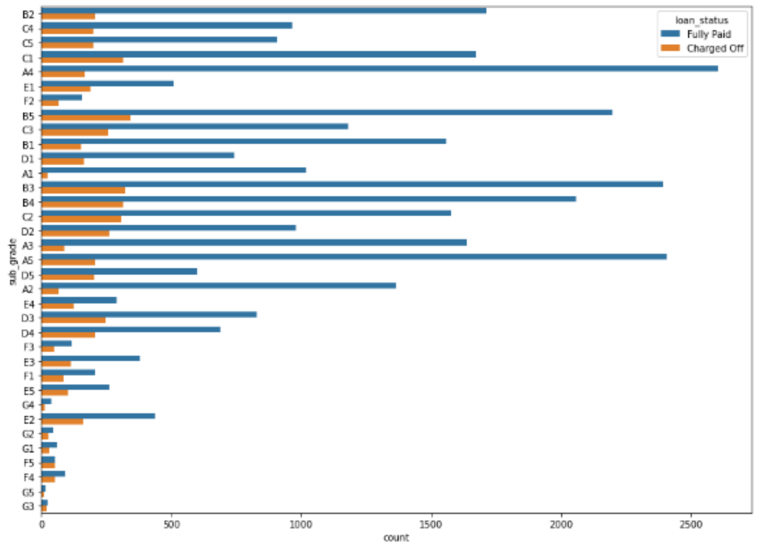


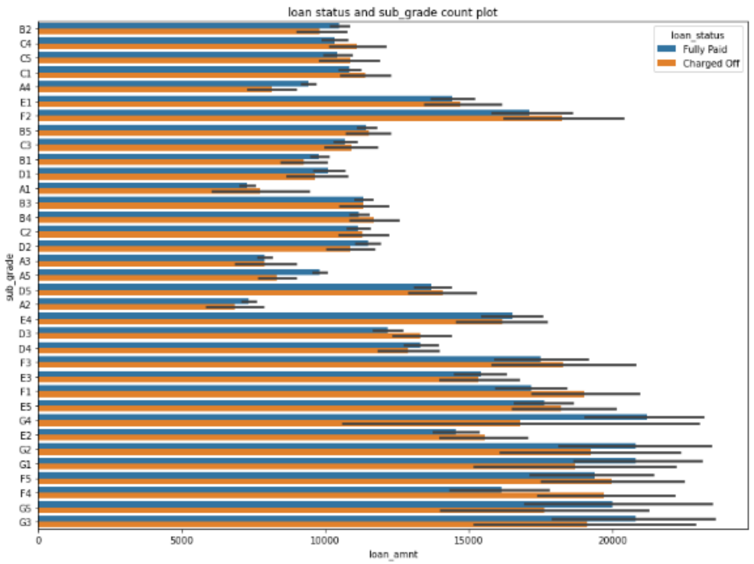


Sub Grade

Mostly Charged off loans belong to D, E,F, G but with mean estimator it also stands out that sub_grade F1, F2, F3 and F4 is more riskier category



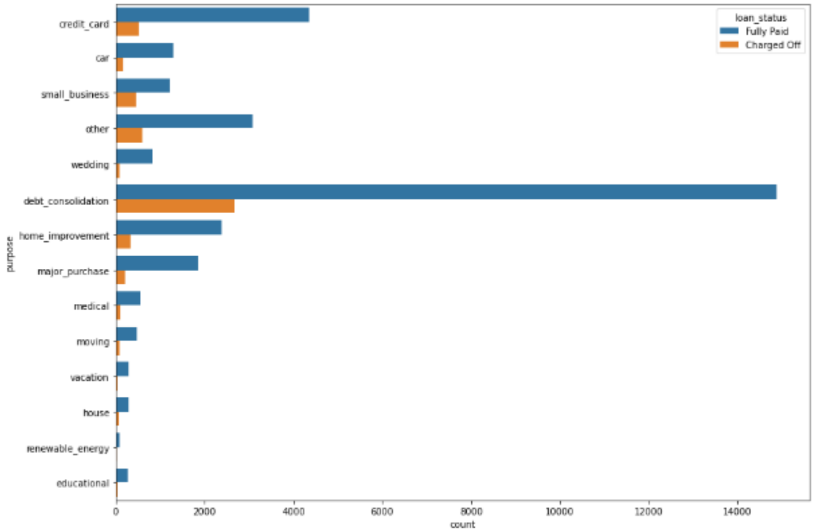


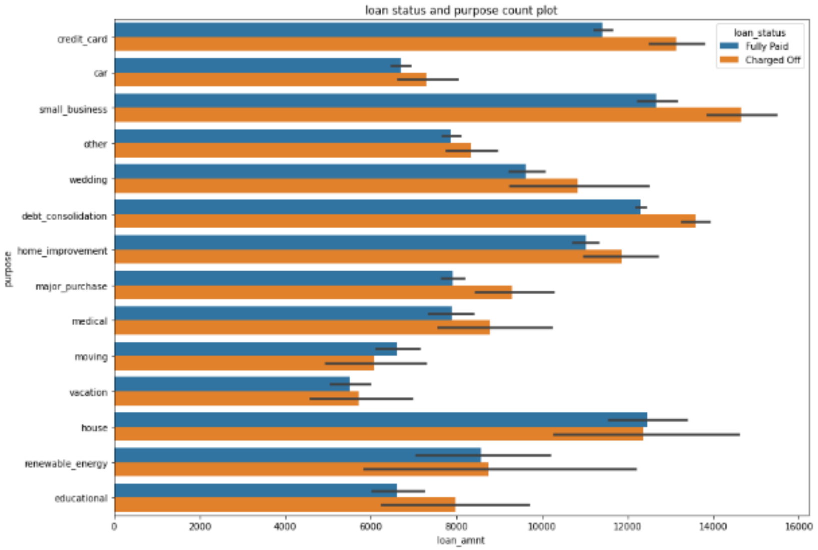


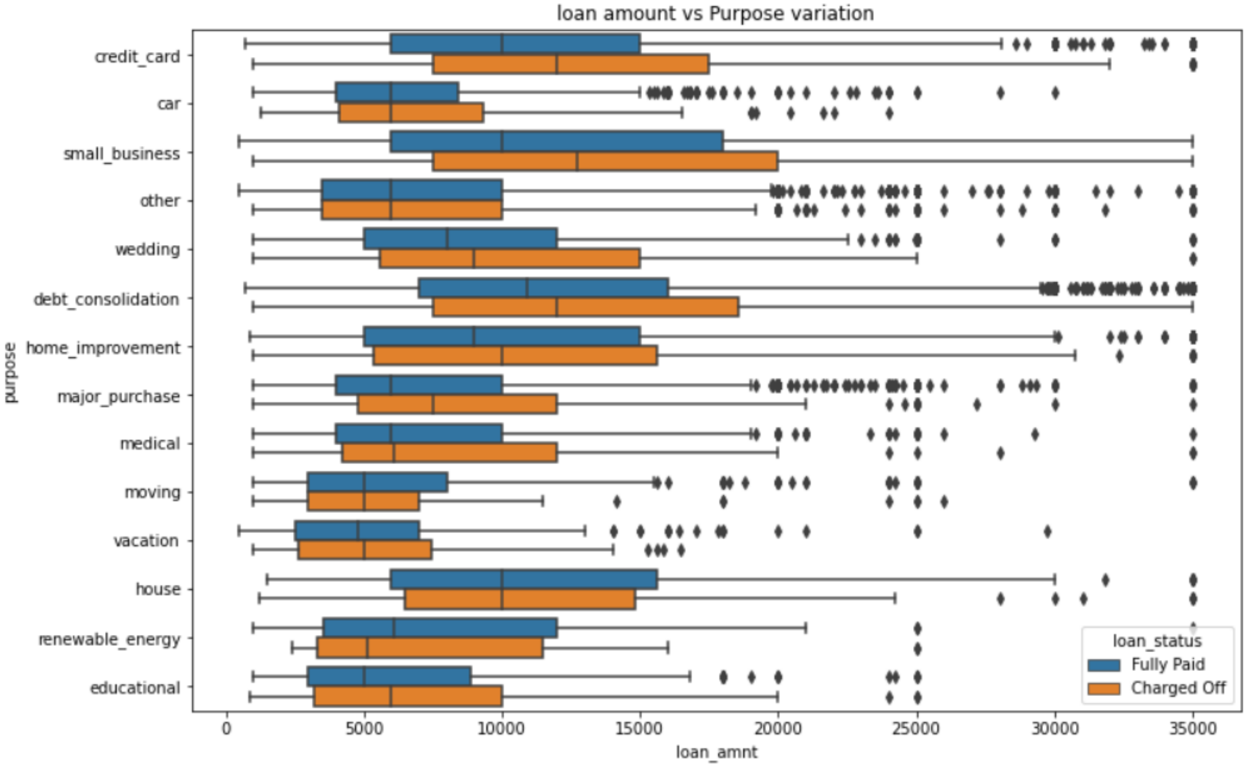
Purpose

Mostly Charged off loans belongs to credit card, debt_consolidation and small business

loan given to category credit_card and debt_consolidation for more than 15-16k falls in risk category and has more chances to get default



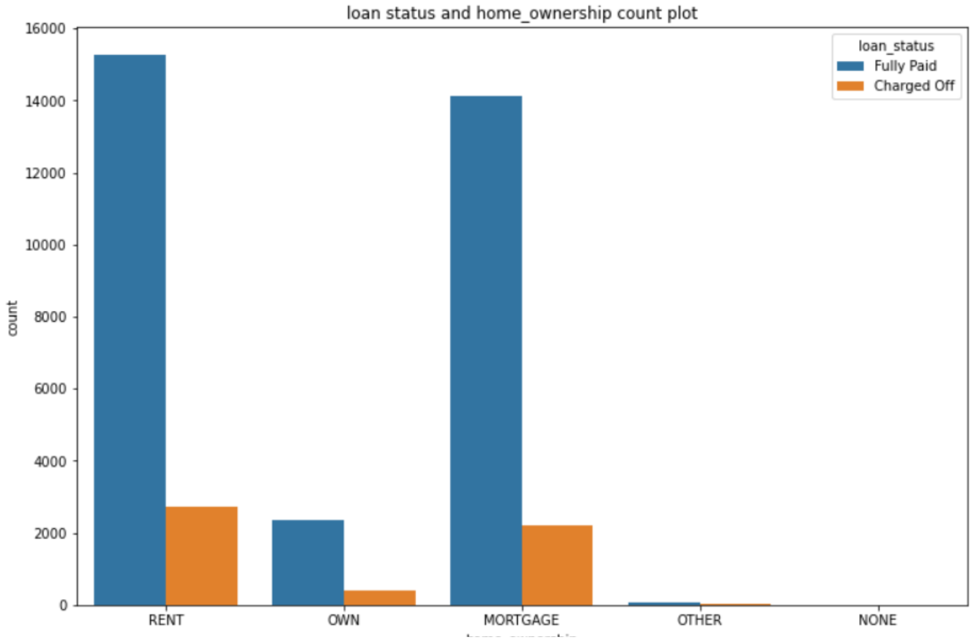


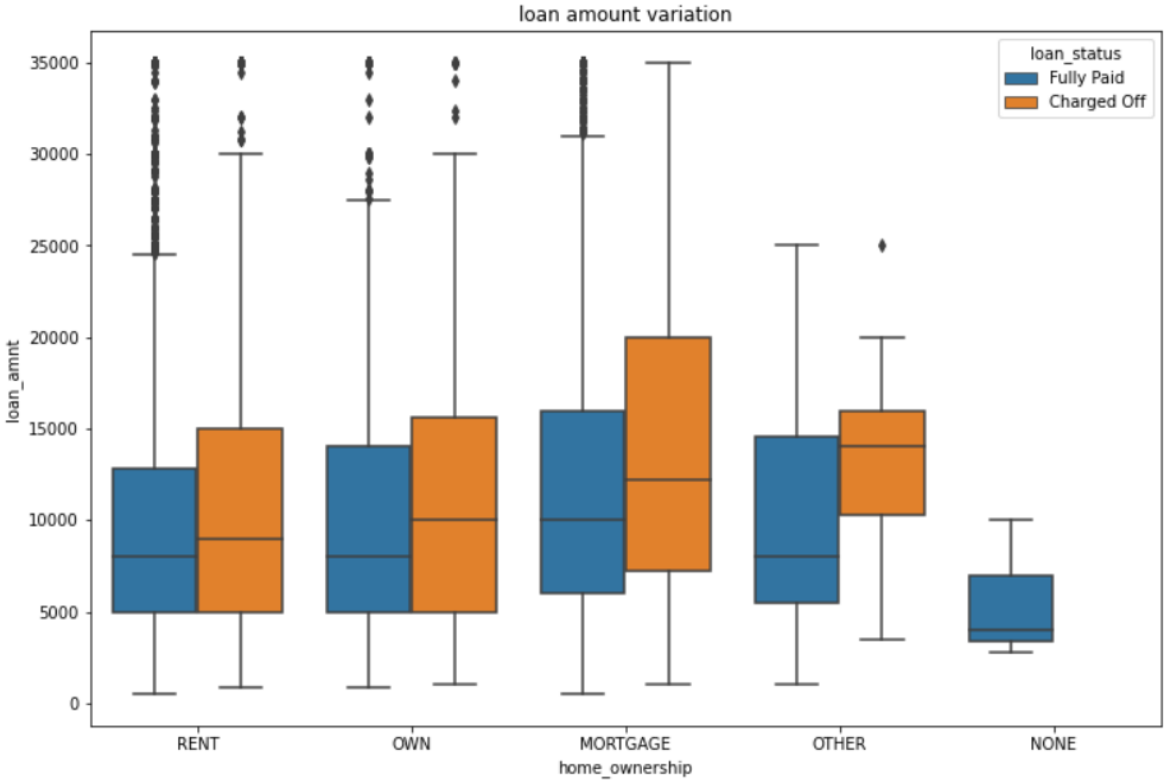


Home Ownership

Defaulters are mostly who don't own a home and are rent/Mortgage

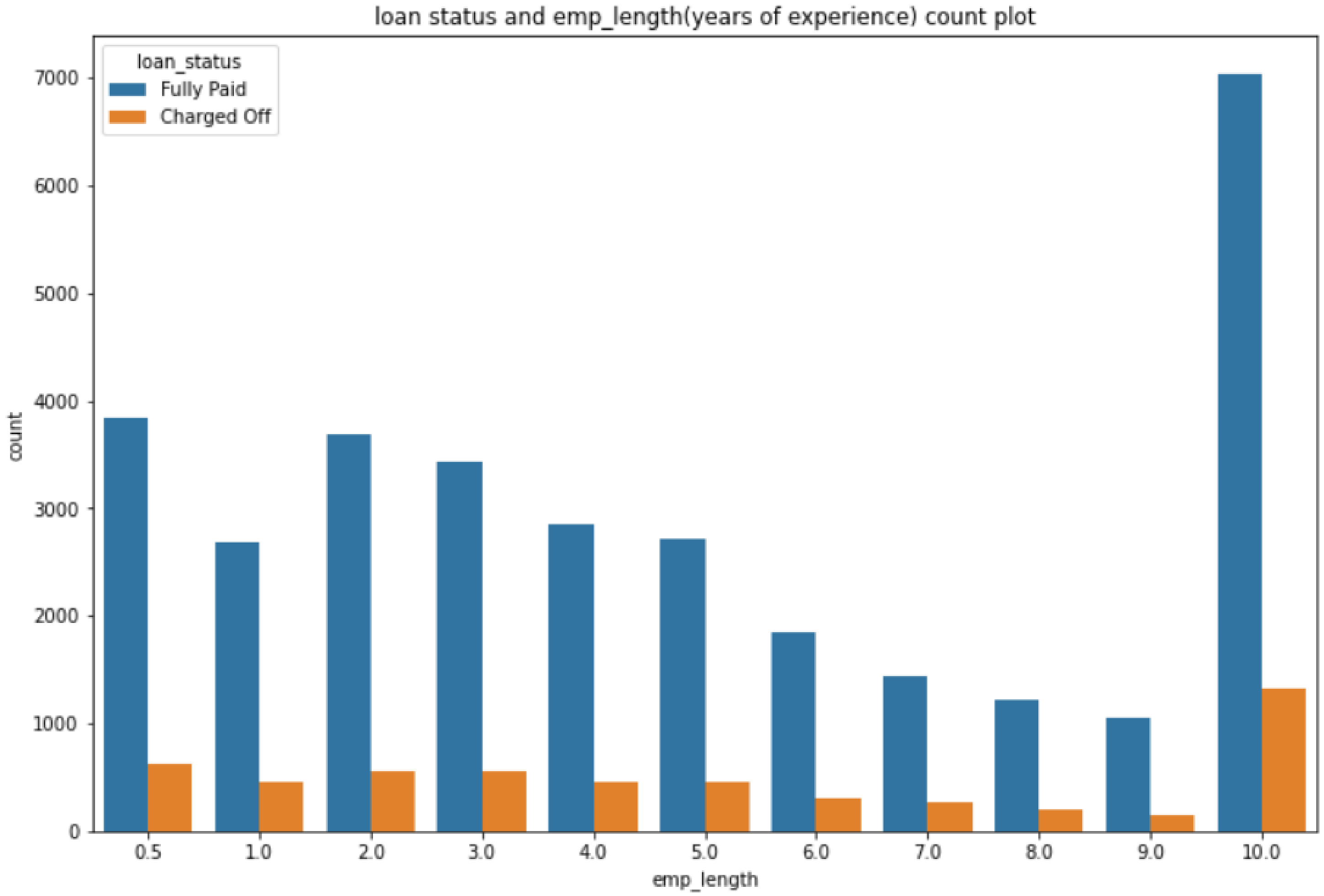
The distribution shows that providing high amount of loans to Mortgage and Rent is riskier





Employment

Applicants with 1 or less than 1 year of experience are more likely to get charged off.

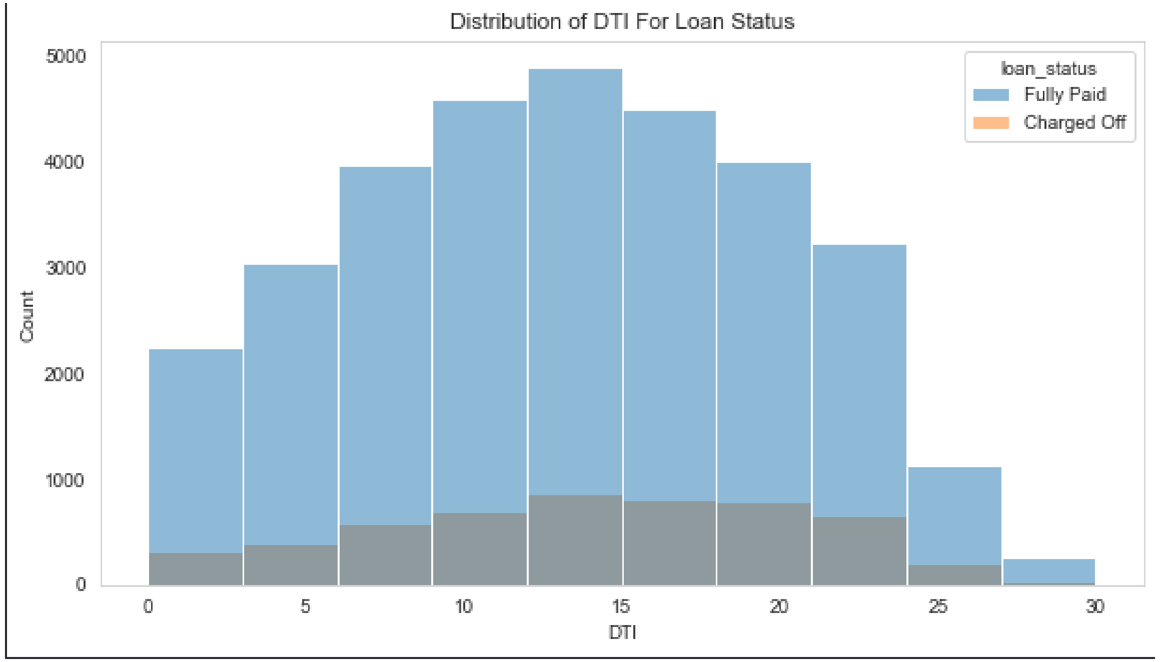


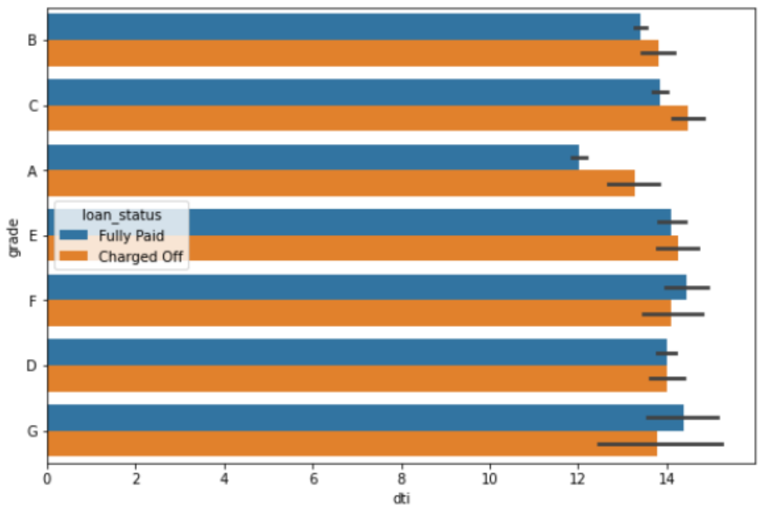
DTI ratio

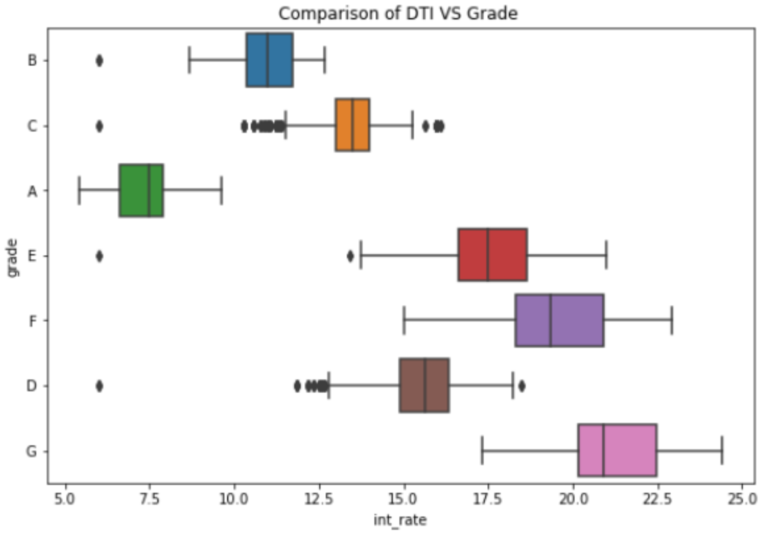
A large percentage of Clients have a large Debt to Income ratio which shows that lending to such clients can be very risky.

Grade A which is the lowest risk also has the lowest DTI ratio which we can say that higher grade like A, B and C has lower rate of default.

Lower Grade has higher interest rates and increase chances of default as well



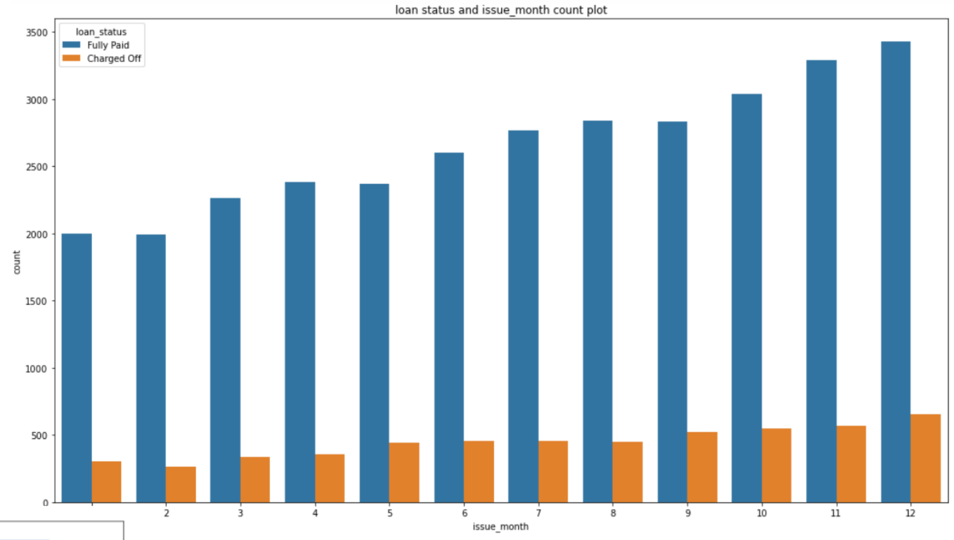


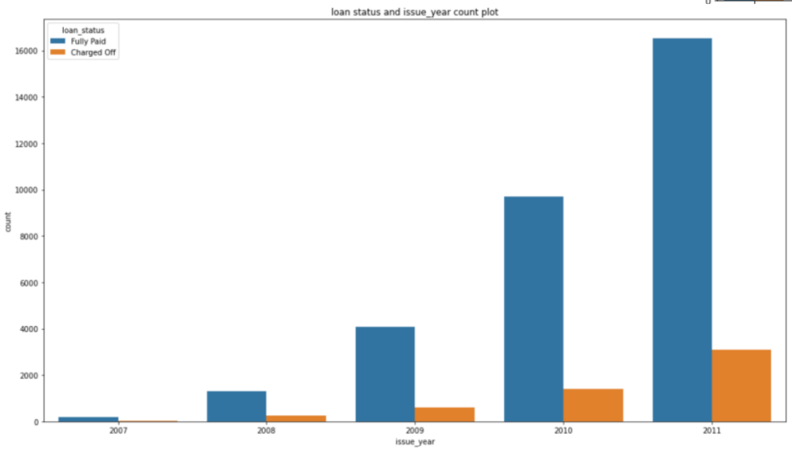


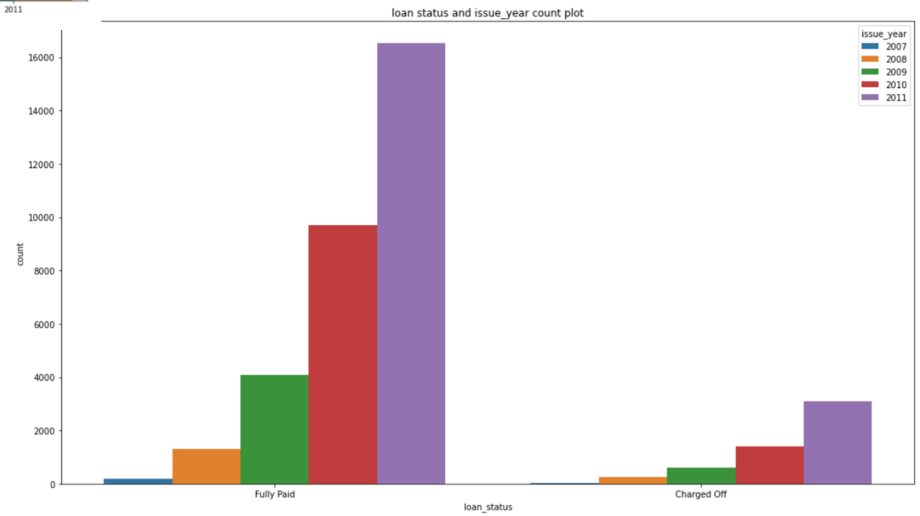
Loans Over the Years

There is an exponential increase in loan applicants over an increasing number of years.

Although there is always an increase at the end of the quarter in loan applications throughout the years.

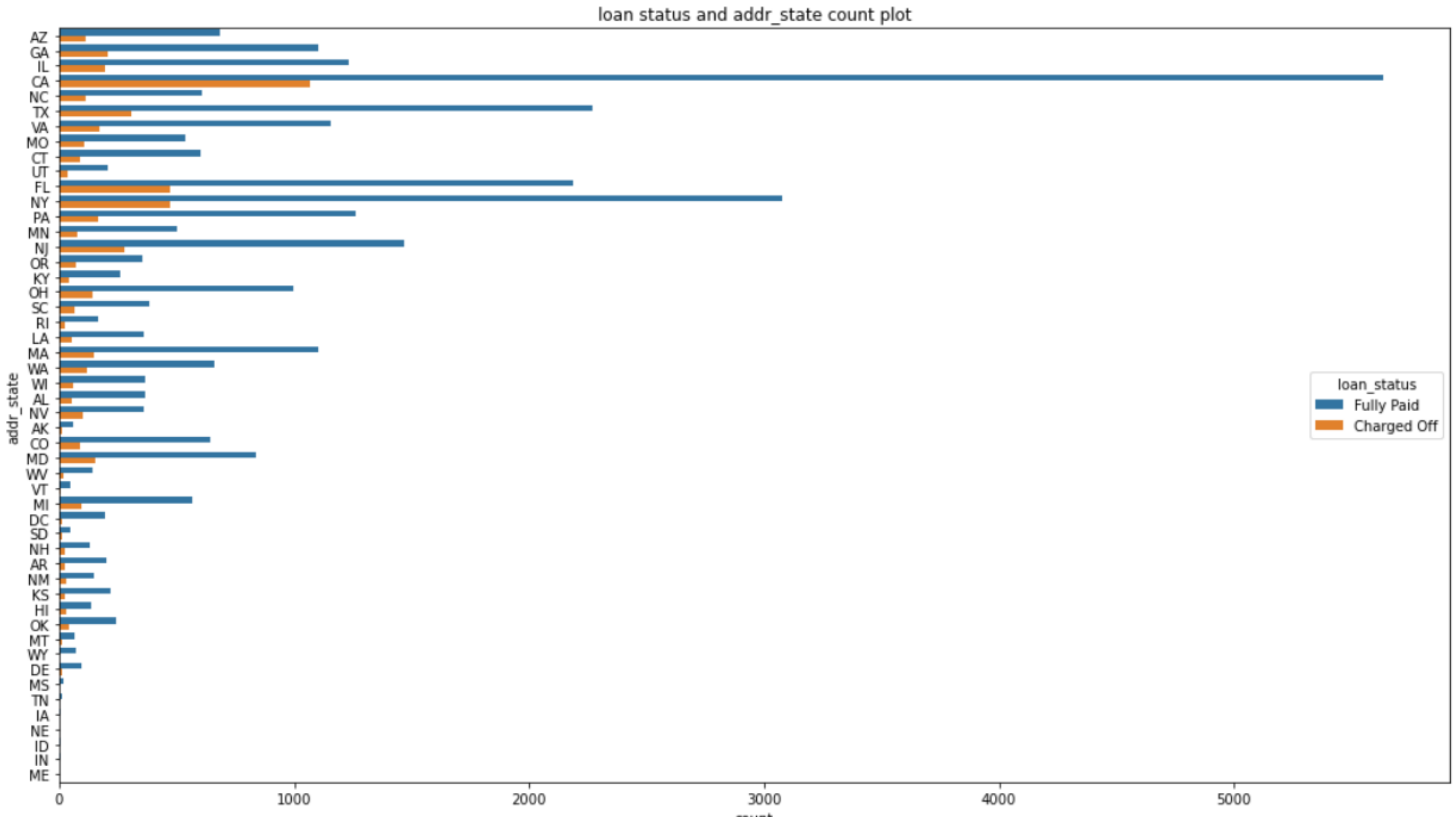






Location

States like NY, CA, FL, NJ, AZ does have a higher chances of approving a riskier loan



Major Driving Factors

Major Driving factors which can be used to predict the chance of defaulting and avoiding Credit Loss:

- Purpose of Loan
- Employment Length
- Grade
- Interest Rate
- Term

Recommendations

There are multiple States with a high probability of charge-off, the highest being NV, NY, CA, FL, NJ, and AZ.

Applicants who take loans for debt consolidation, Credit cards, and small businesses are having high chances of charge-off.

As we move from Grade A to G, the probability that Applicant will charge off is increasing

As Annual income is decreasing, the probability of the Applicant charging off is increasing, Applicants earning less than 100k are a risk, and Applicants earning less than 50k are riskier.

Higher the interest rate, the higher the chances an Applicant can charge off.

Applicants with less than 1 year of experience are more likely to charge off.