Lending Club Case Study

Gaurav Ahuja

Objective

The Objective of this case study is to implement EDA technique on a real world problem and understand the insights and present in a business first manner via presentation.

Benefits of the case study:

- ➢ Gives a idea about how EDA is used in real life business problems.
- ➢ It also develops a basic understanding of risk analytics in banking and financial services.
- ➢ How the data is used to minimize loss of money while lending it to clients.
- ➢ It improves our understating of visualization and what charts to use for real life data.

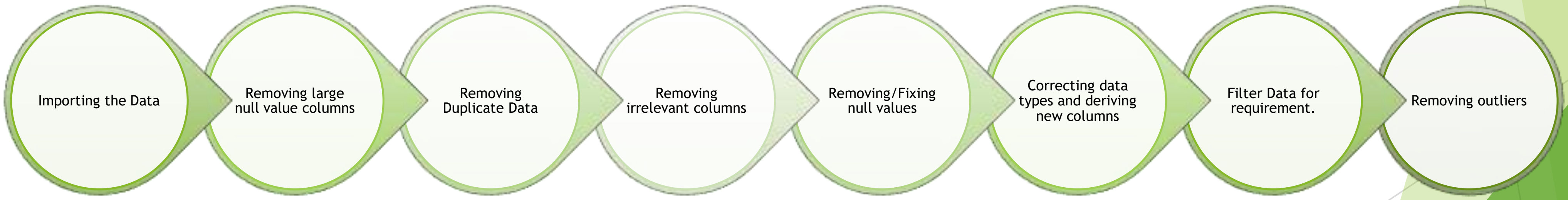
Business Understanding

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

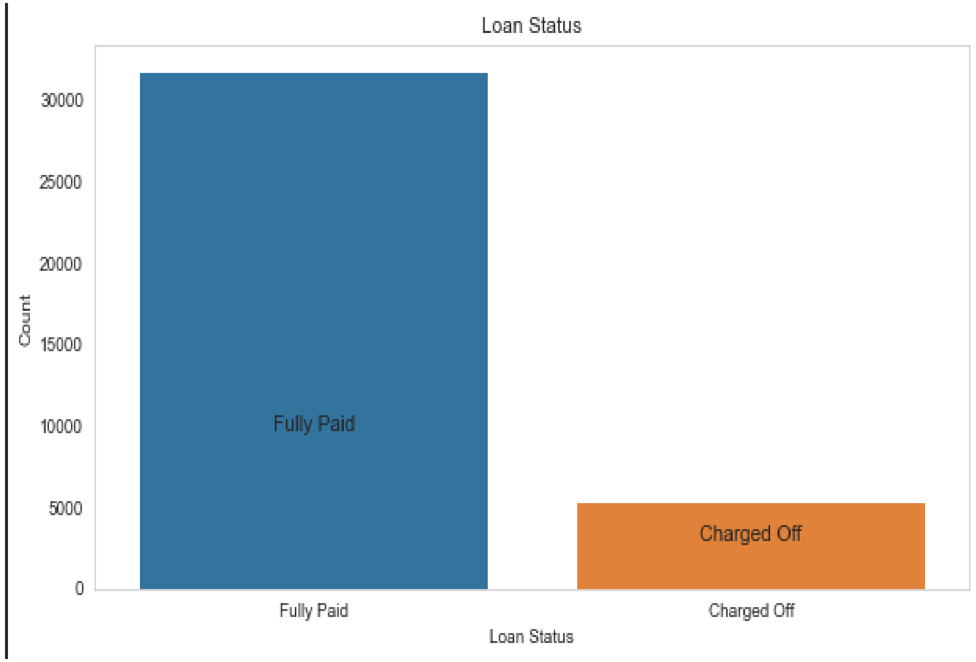
Dataset Details:

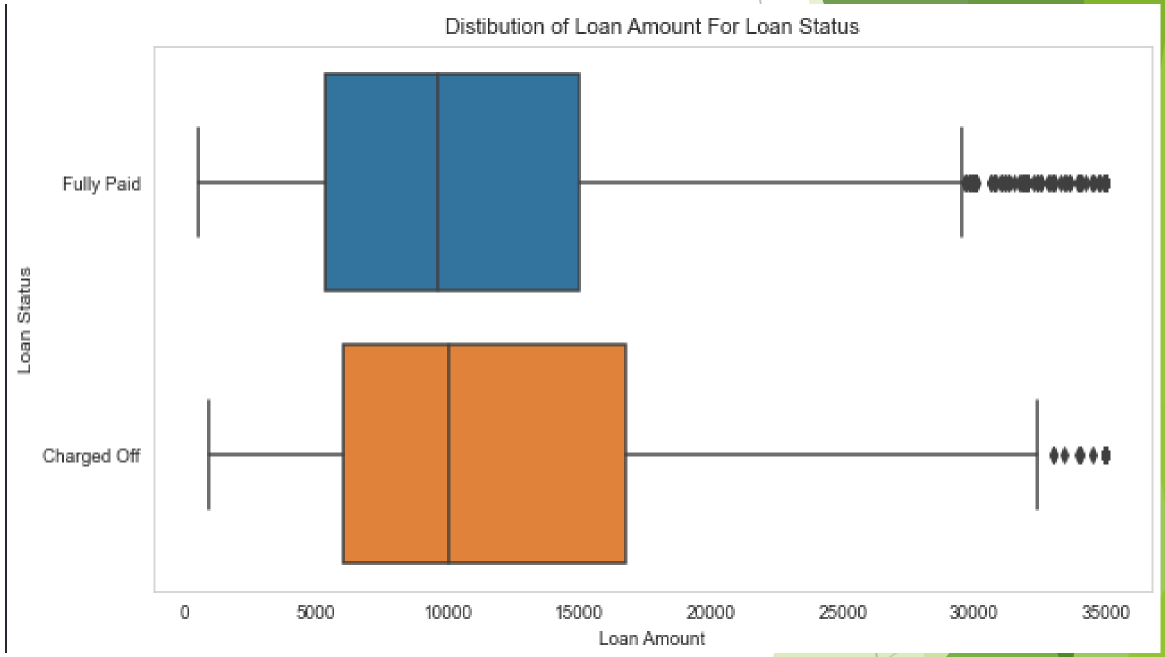
The data given below contains information about past loan applicants and whether they 'defaulted' or not. Data has details regarding approved loan not the rejected ones. It has 3 status of loan which is Fully Paid, Current and Charged-Off.

Data Clean-up and preparation process:



Loan Status and Amount

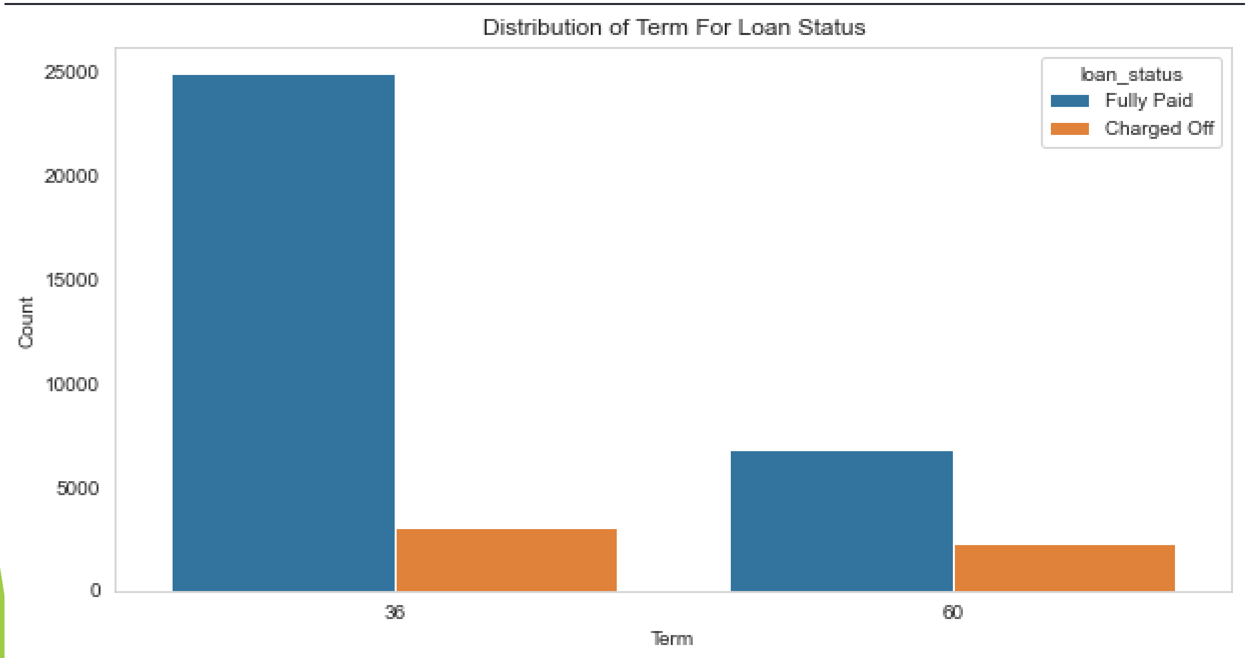


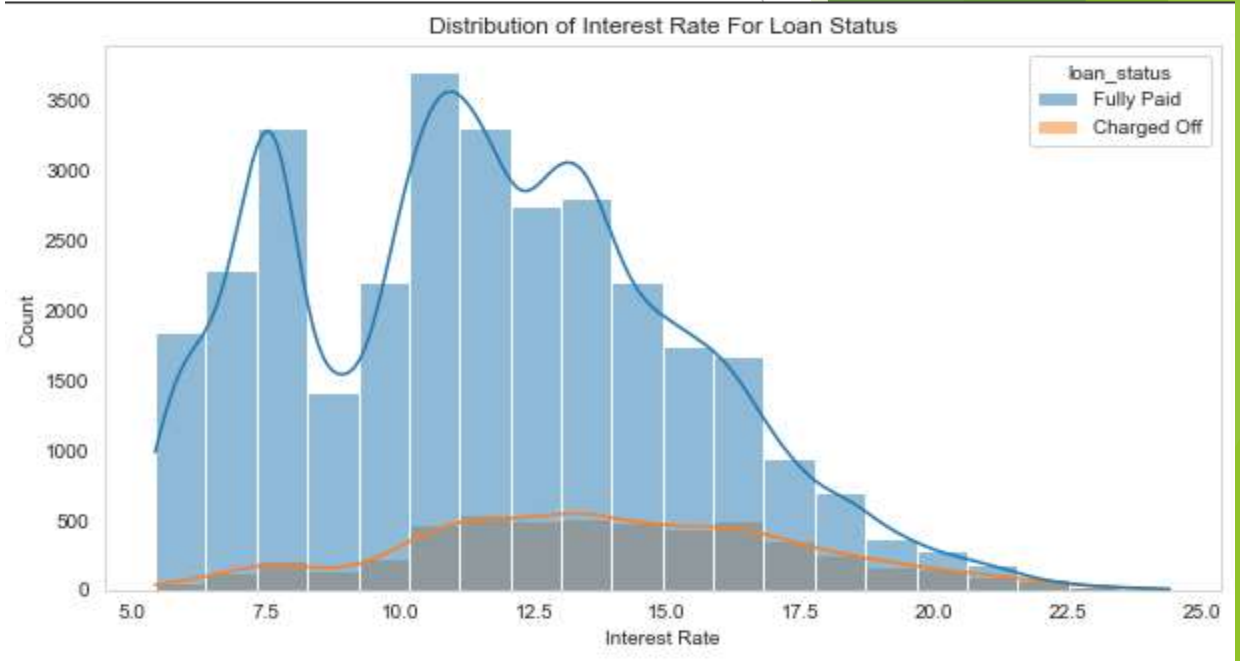


- **Loan Status:** The number of charged off loan is much smaller(14.5%) compared to total count.

- **Loan Amount:** It varies from 500 to 35000 with a median of 10000. Loan amount is majorly small and very few clients have taken large loans and larger it goes we have higher chance of defaulting.

Term and Interest Rate

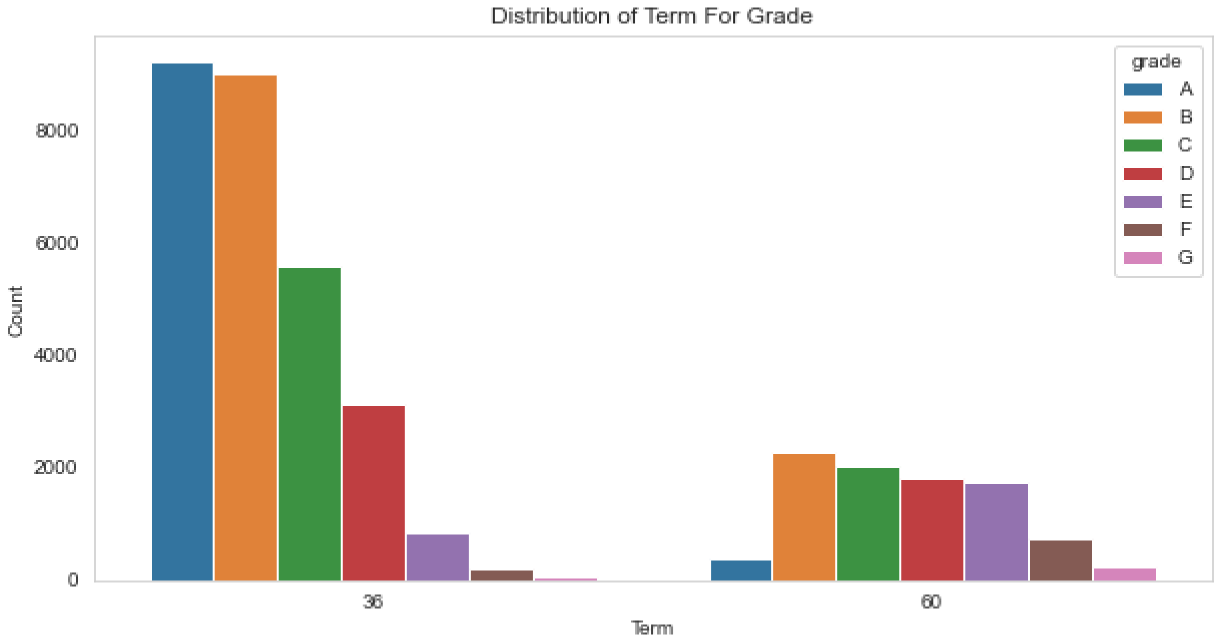


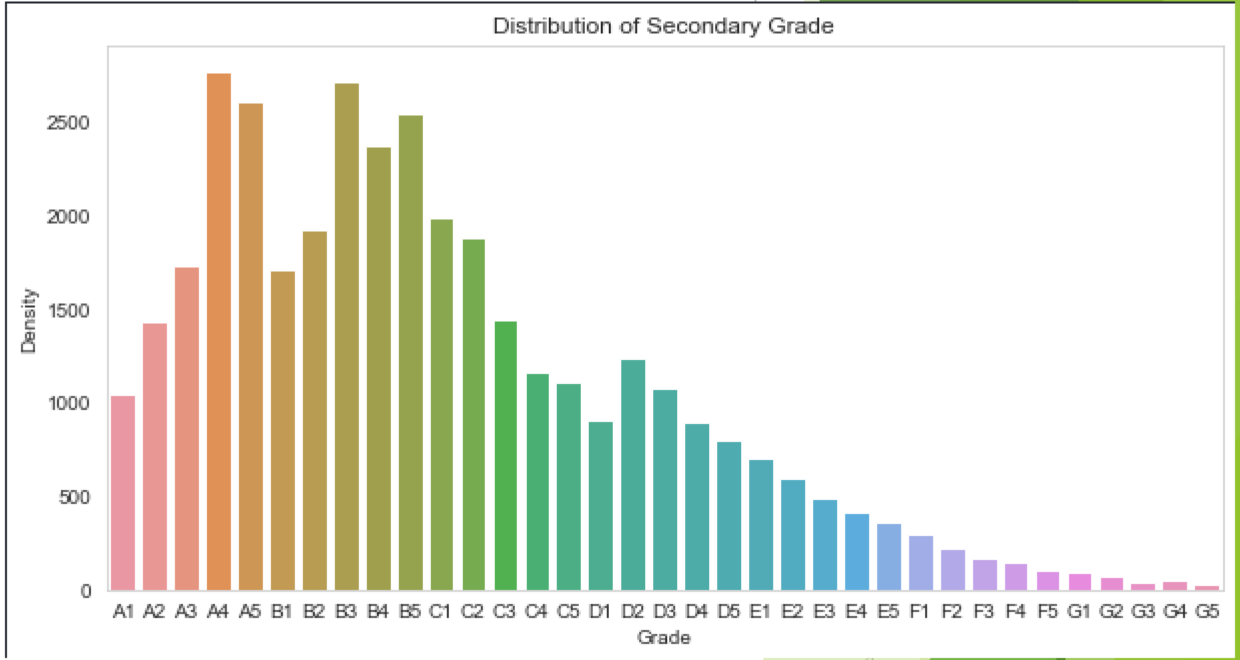


- **Loan Term:** The Loans taken for 36 month term are much more than 60 months and have lower chance of defaulting.

- **Interest Rate:** The count of loan taken varies with interest rate showing peak around in 5-15 bracket and decreasing slowly where as the chance of defaulting increases with interest rate.

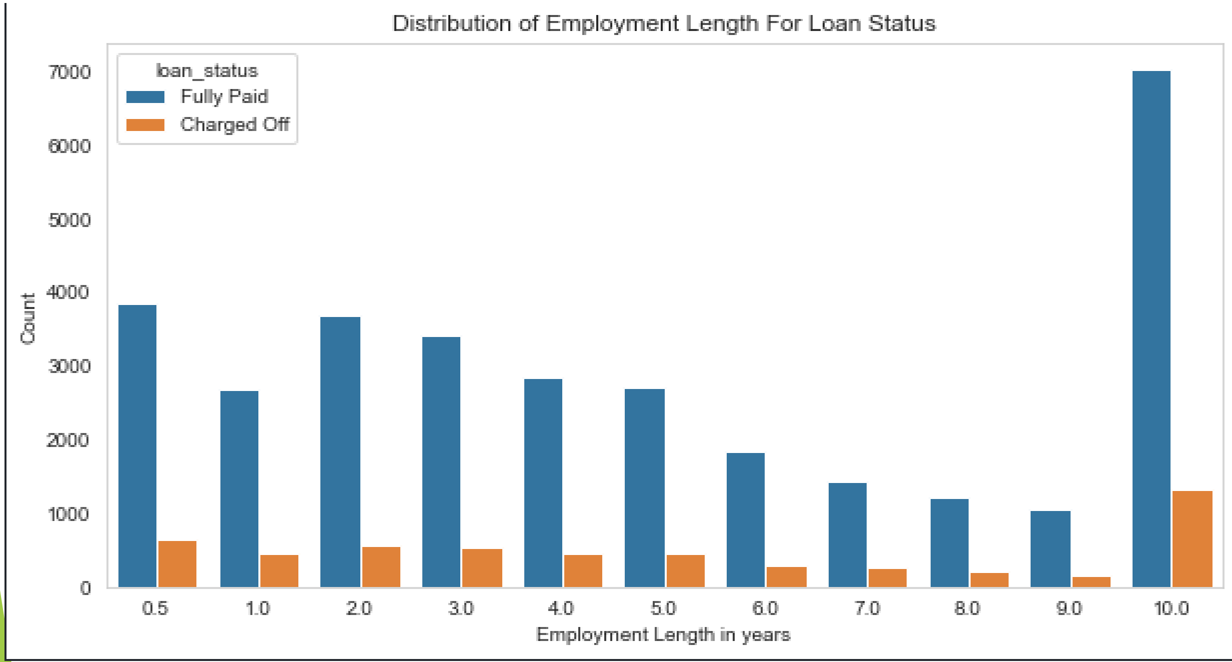
Grade and Sub-Grade

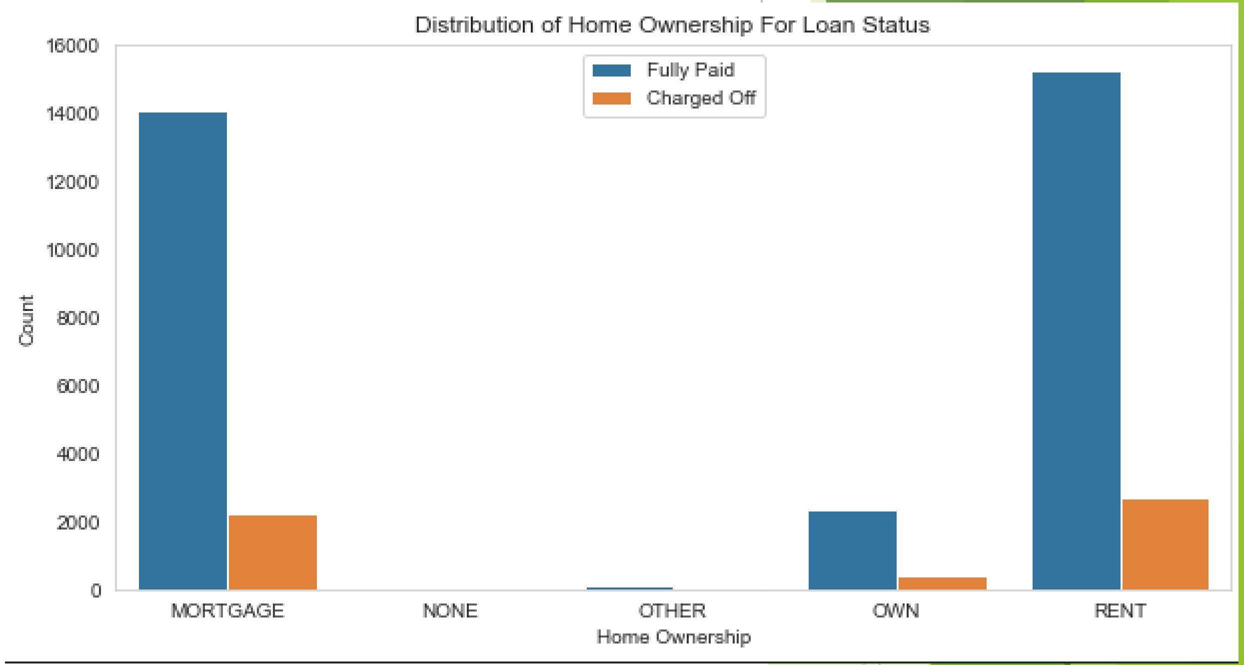




- **Grade:** The loan approved are majorly of higher grade as they are of low risk thus low chance of defaulting. 60 month term loans have larger number of lower grade loans with high risk.
- **Sub Grade:** This provides more insight that the loans within grade are more skewed towards lowered sub grades.

Employment Length & Homeownership

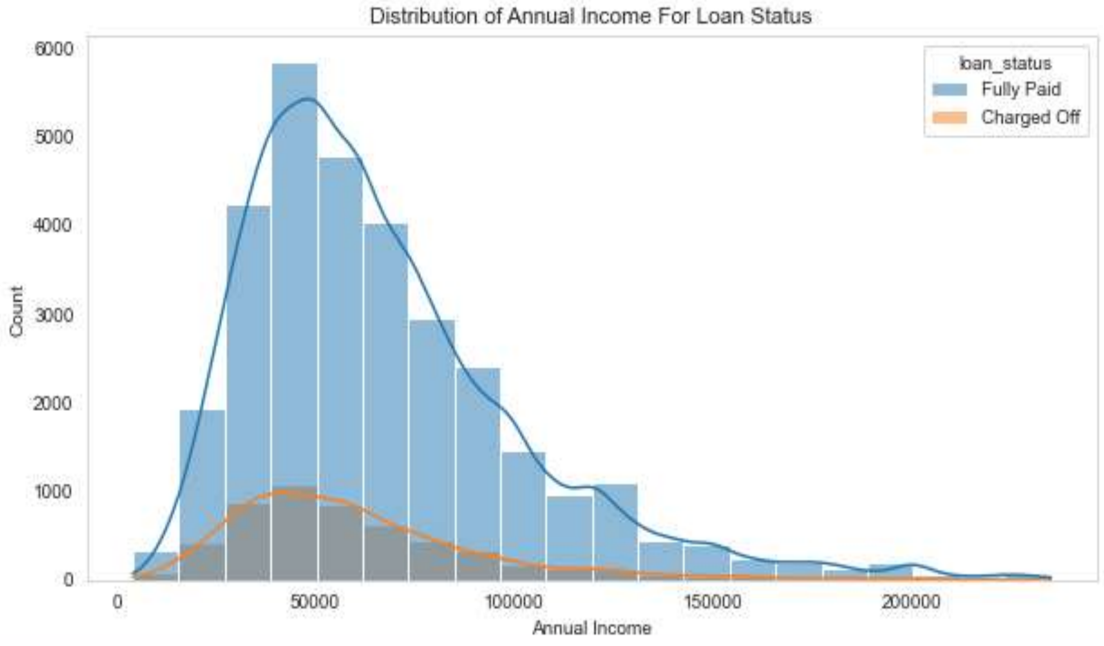


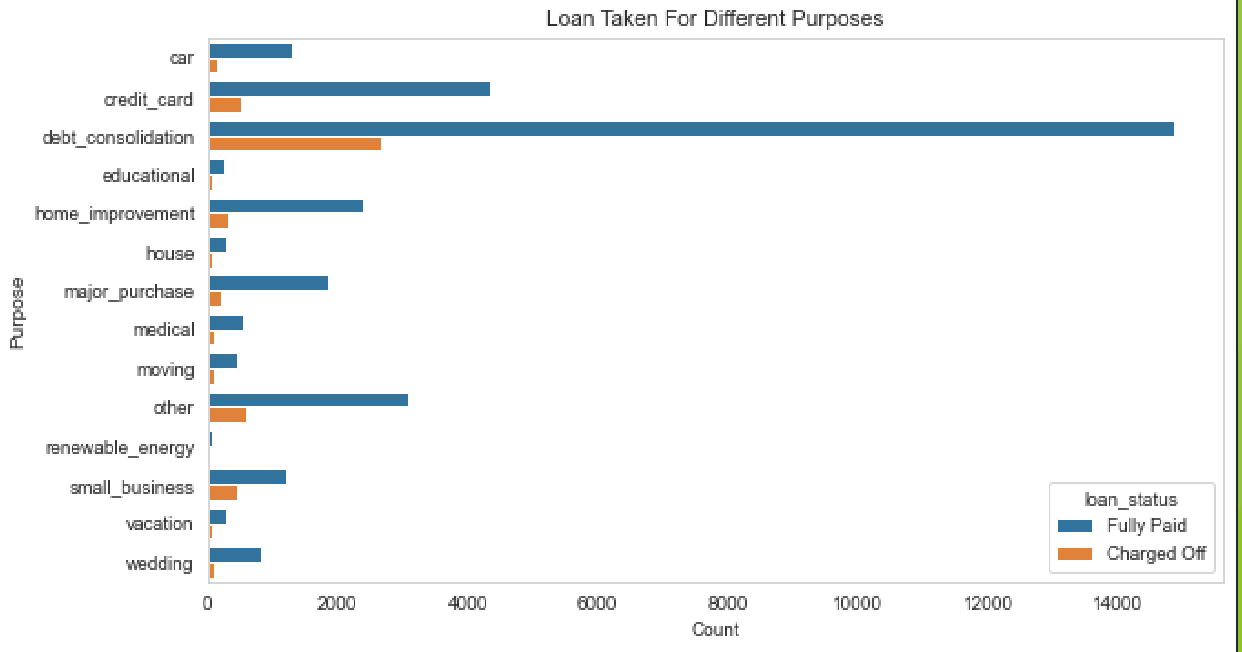


- **Employment Length:** Majority of clients have 10+ years of experience and has highest number of defaulted loan.

- **Home Ownership:** Majority of clients are lacking ownership of any property and are on rent or mortgage and have a higher chance of defaulting.

Annual Income & Purpose

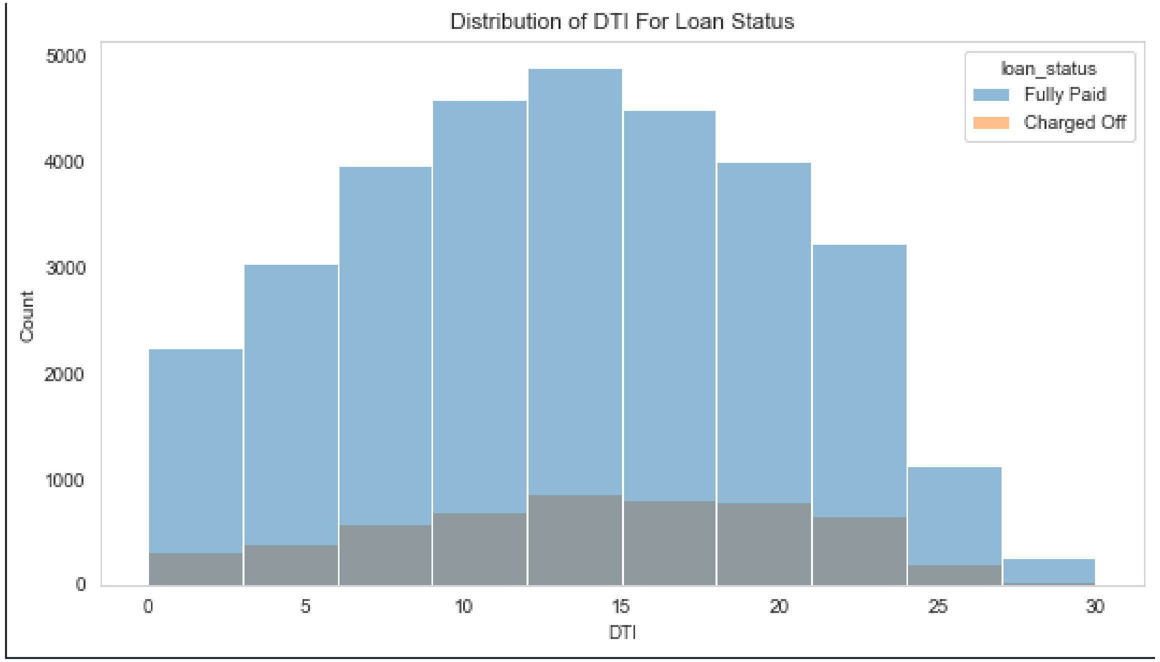


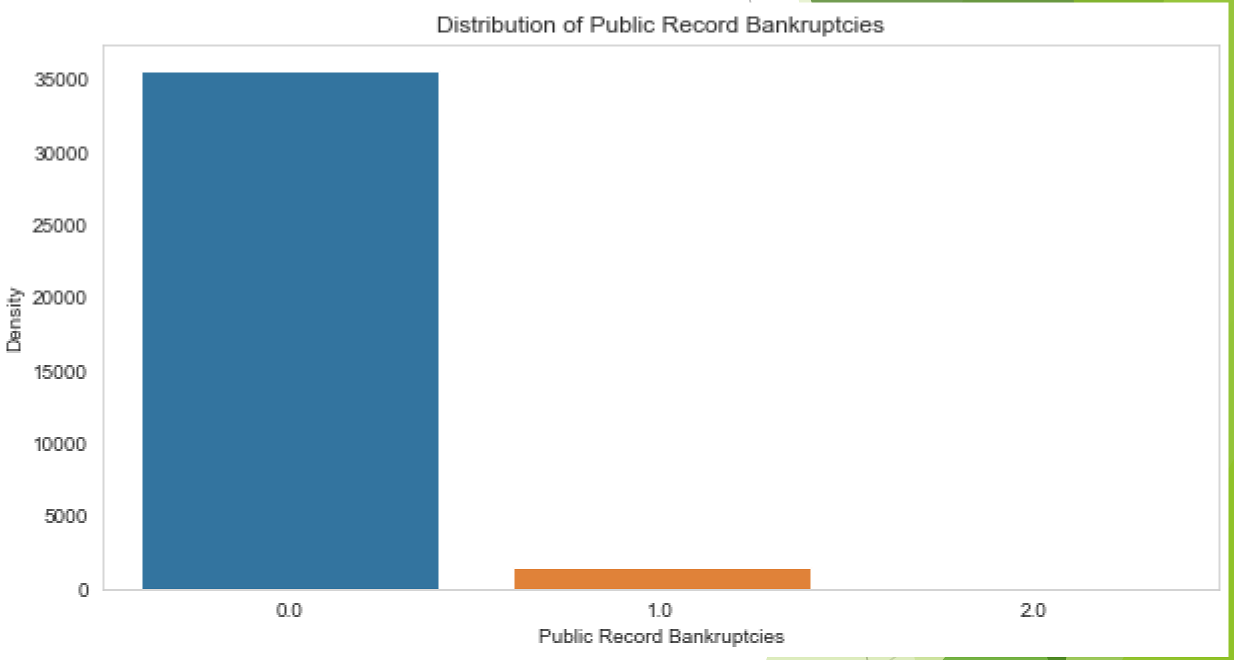


- **Annual Income :** The Majority of clients have low annual income compared to rest and income lower than 50k has higher chance of defaulting.

- **Purpose:** Loans are taken mostly for debt consolidation followed by credit card payment. Whereas the debt consolidation has highest fully paid loan but also has highest defaulted loans as well.

DTI ratio & Bankruptcy

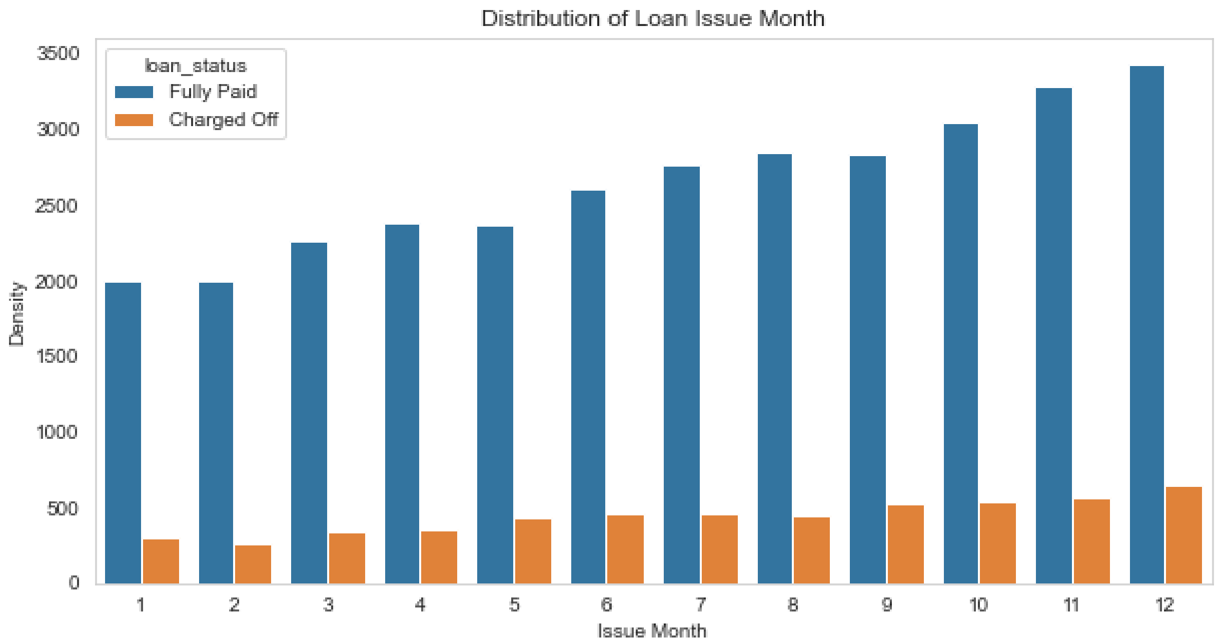


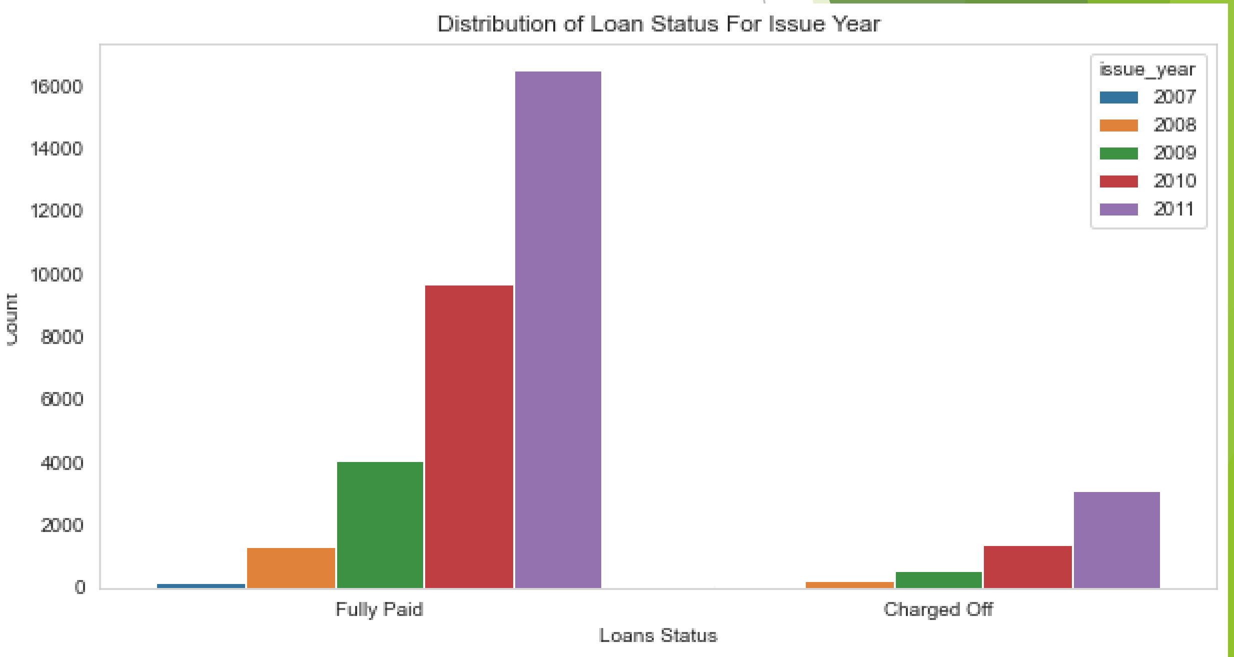


- **DTI:** The large percentage of Clients have a large Debt to Income ratio which shows that lending to such clients can be very risky.

- **Public Recorded Bankruptcy:** Majority of clients have no record of declaring bankruptcy.

Loan Trend over years

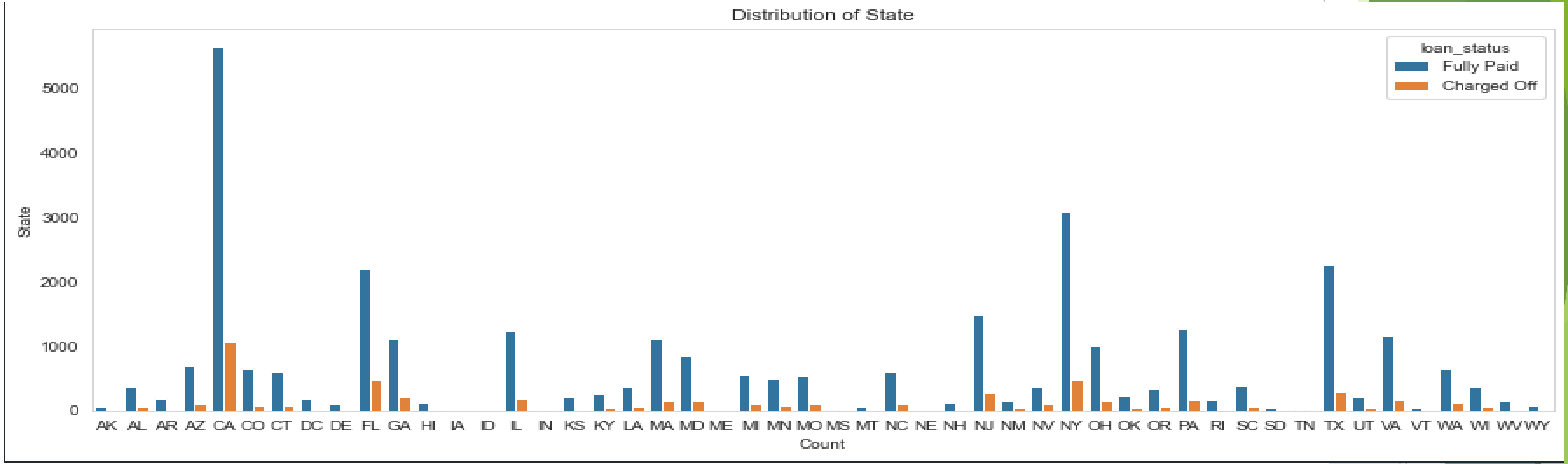




We see a gradual increase in loan taken through the year, with lesser defaulting rate in April ,August, December quarter wise and better more late in year.

With each passing year loan taken are increasing exponentially which indicate we are seeing large increase in DTI ratio and decrease in defaulting rate.

Location Based



For large metropolitan cities we see large number of loans, with higher number of defaulted loans like California, New York, Texas, Florida but have a lower chance of defaulting.

Recommendations

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

1. DTI
2. Grades
3. Verification Status
4. Annual income
5. Pub_rec_bankruptcies

Other considerations for 'defaults' :

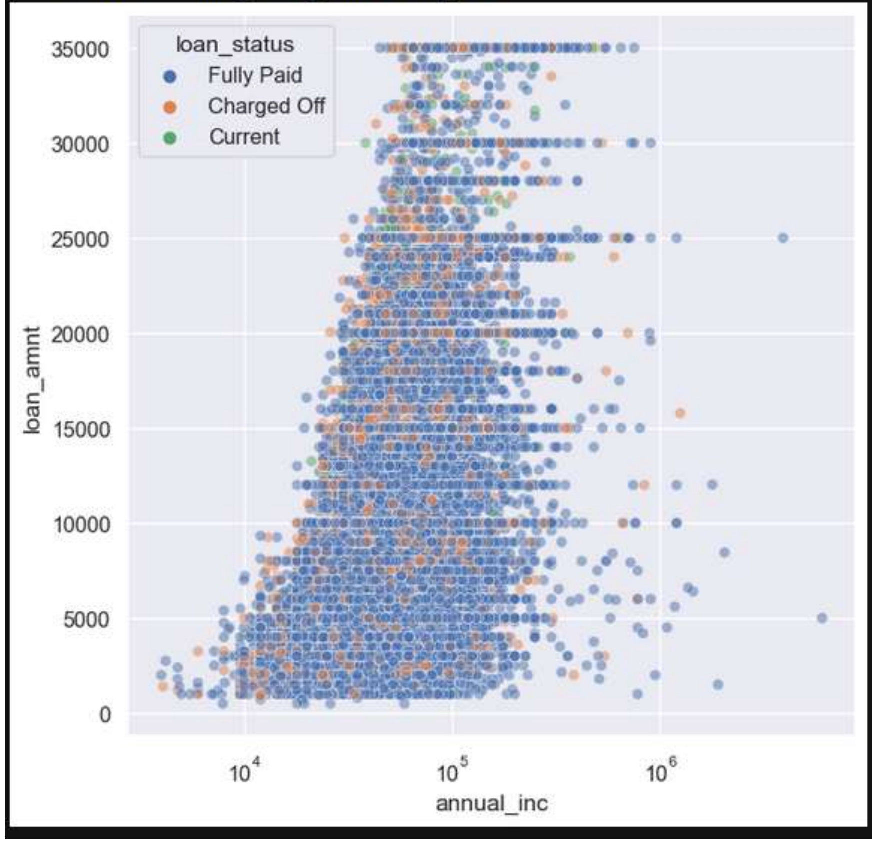
1. Borrowers not from large urban cities like California, new york, texas, florida etc.
2. Borrowers having annual income in the range 50000-100000.
3. Borrowers having Public Recorded Bankruptcy.
4. Borrowers with least grades like E,F,G which indicates high risk.
5. Borrowers with very high Debt to Income value.
6. Borrowers with working experience 10+ years.

Observations→

- Borrower with higher annual_inc go for higher loan amount

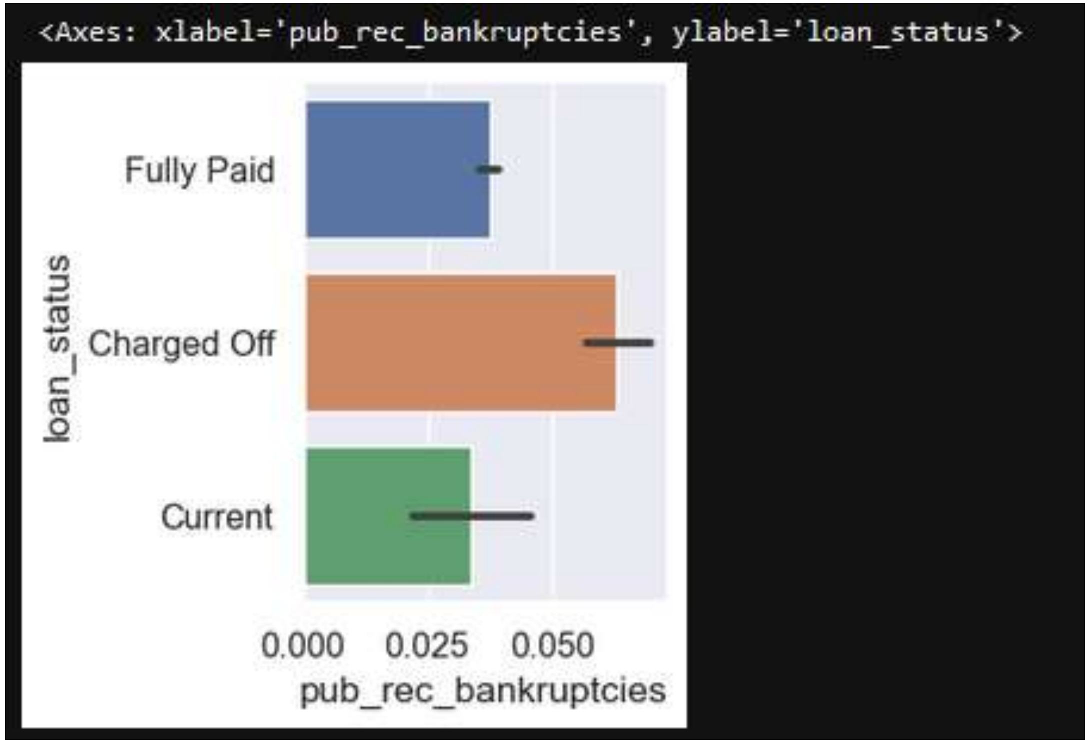
```
f, ax = plt.subplots(figsize=(7, 7))
ax.set(xscale="log")
sns.scatterplot(data=loan_data, hue='loan_status', x='annual_inc', y='loan_amnt', alpha=0.5, ax=ax)
```

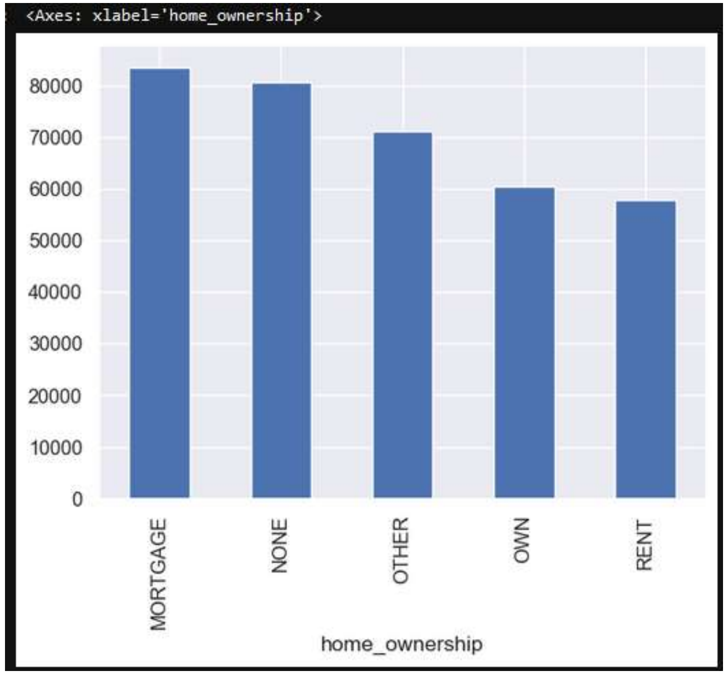
```
<Axes: xlabel='annual_inc', ylabel='loan_amnt'>
```

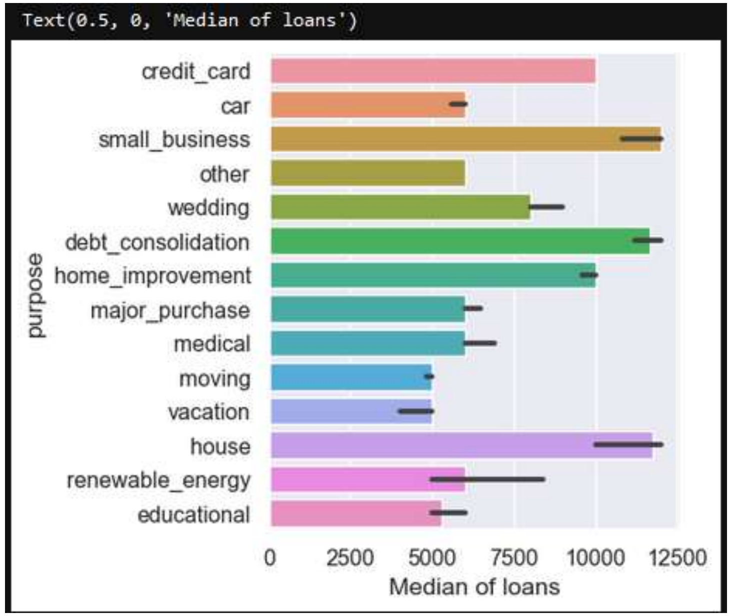


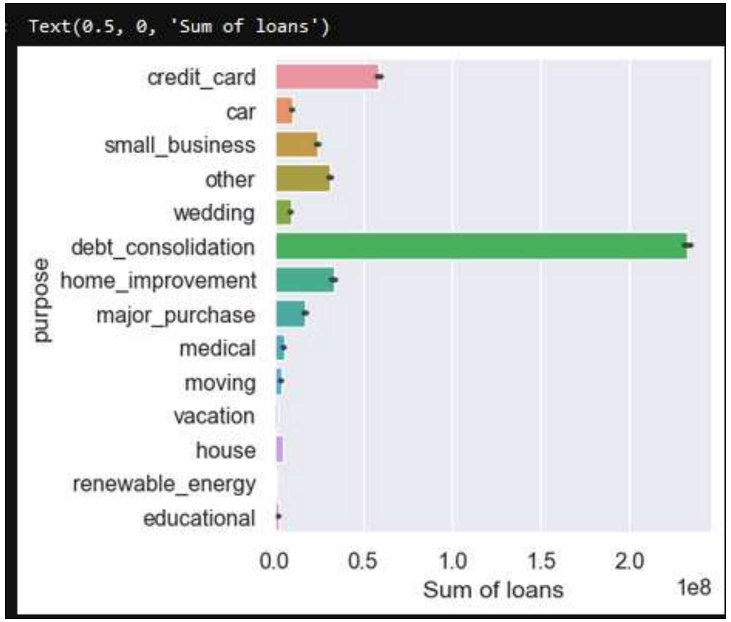
Borrowers who went bankrupt before are most likely to default. People having 2 records of bankruptcies have a higher probability to default than people with no bankruptcy report

- Observation→









Observations->

- credit_card, debt_consolidation, home_improvement, major_purchase are the purposes that have more than 2000 count and at least 10% default rate Median of small business loans are highest and their default rate is also highest. Most of the money is invested in debt_consolidation and their default rate is also pretty high (above 15%).