

# Exploratory Data Analysis on Lending Club case study

By:

Pradeep Roy

Manoj Akella

## What is Lending Club?

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

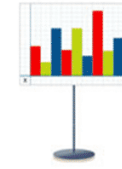
## How Lending Club Works



**Borrowers** apply for loans.  
**Investors** open an account.



**Borrowers** get funded.  
**Investors** build a portfolio.



**Borrowers** repay automatically.  
**Investors** earn & reinvest.

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

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.

Following methods used are:

1. Understanding Dataset
2. Data Cleaning
3. Data Preparation
4. Exploratory Data Analysis

Data Cleaning steps used:

- Delete columns: delete unnecessary columns
- Remove outliers: remove high and low values that would disproportionately affect the results of the analysis
- Missing values: treat missing values with appropriate approach

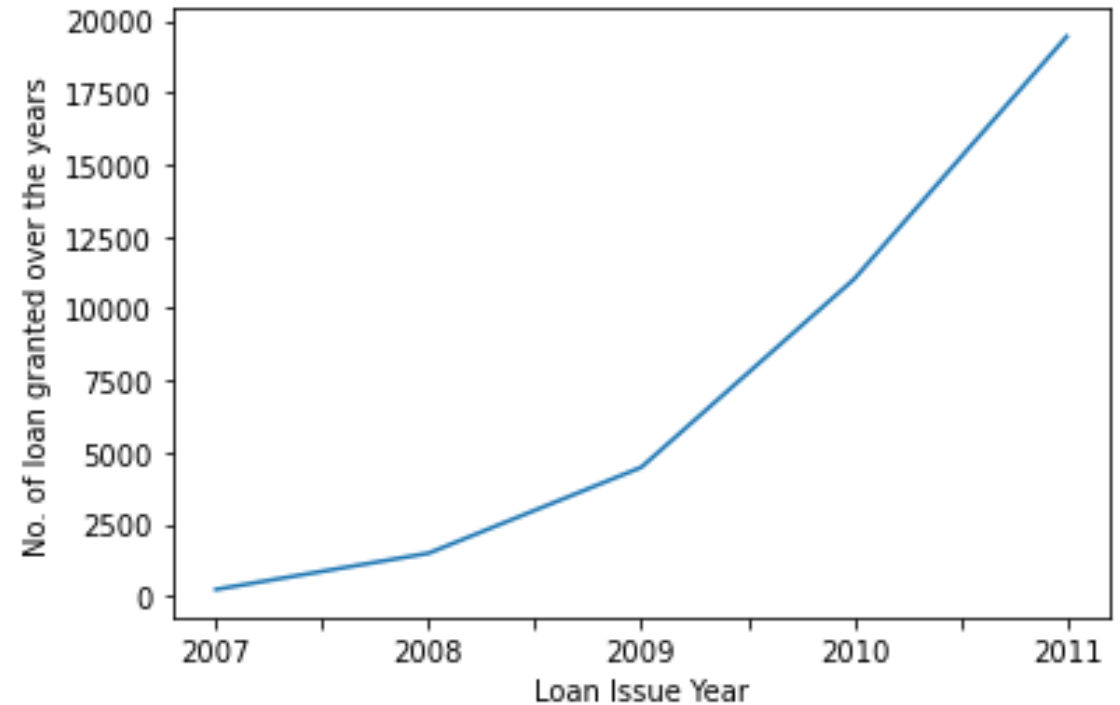
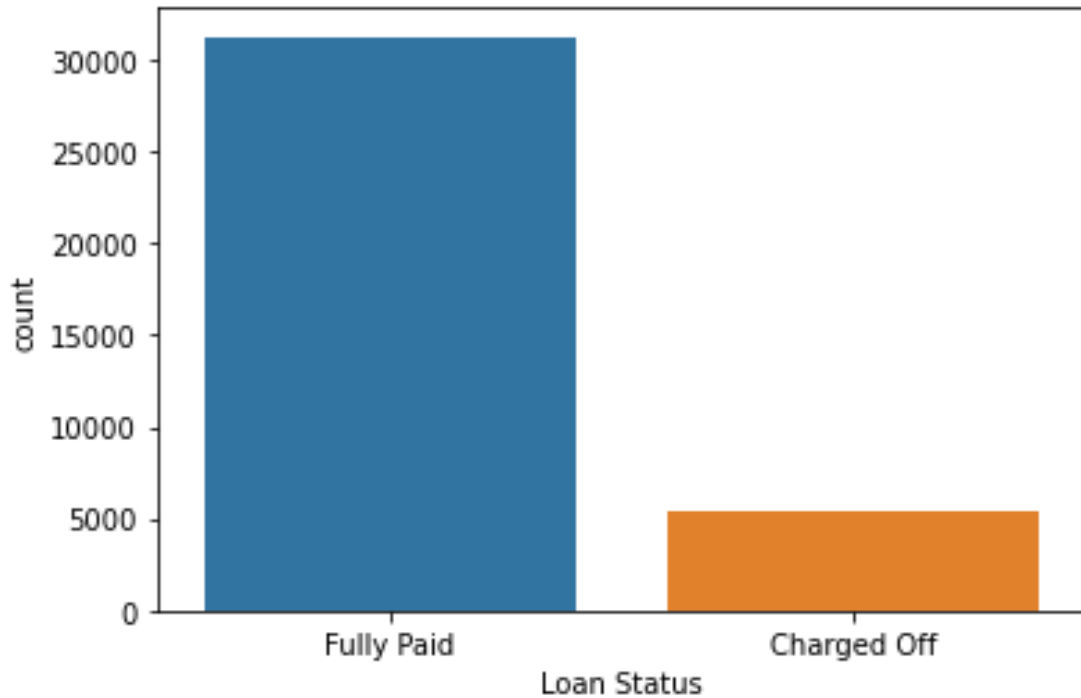
# Analysis

- The objective of the case study is to analyze and understand how the loan applicant and the 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
  - Extracted derived variables e.g. issue\_month, issue\_year to facilitate further analysis
- For **Univariate analysis/Segmented univariate analysis**, we have used
  - Bar chart to check the distribution of all the driver variables
  - Box plots to detect and remove outliers
- We have performed **Bi-Variate and Multi-Variate analysis** to understand how different variables influence each other.

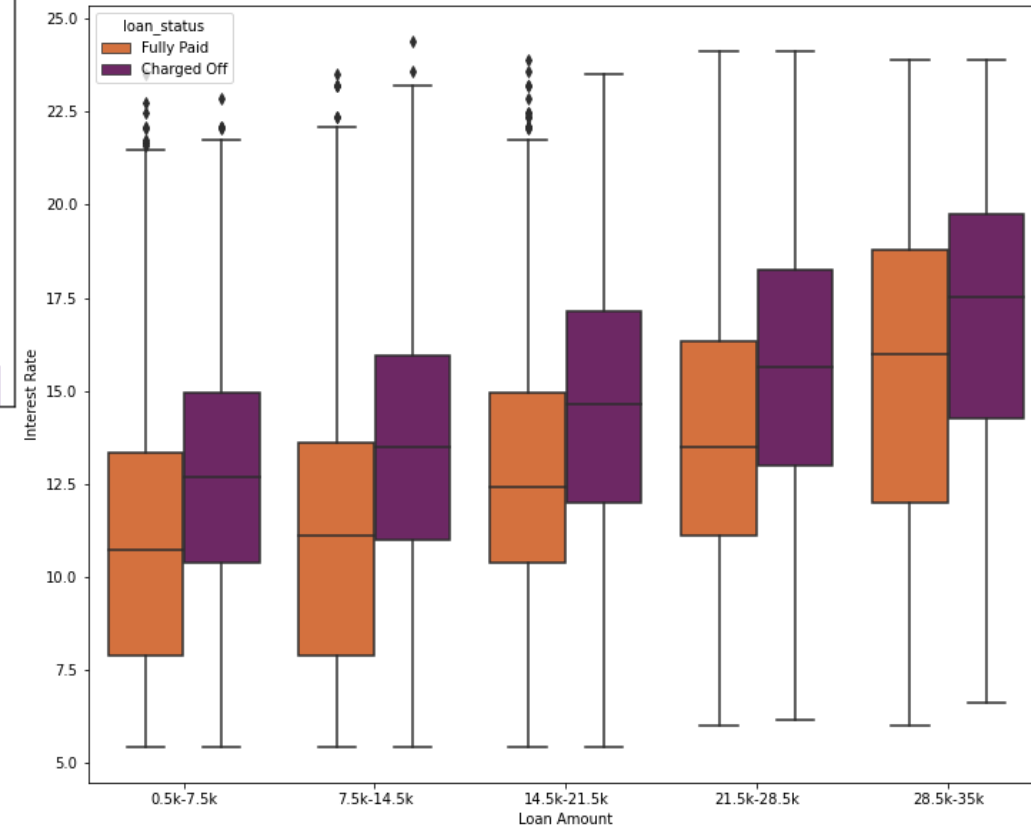
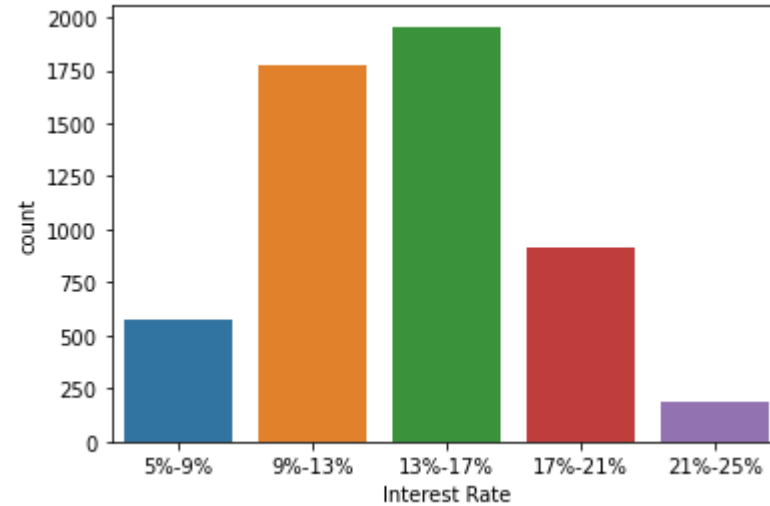
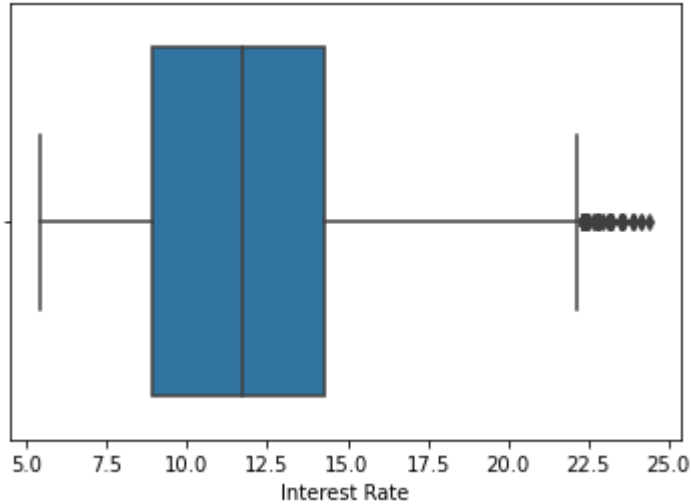
# Loan Status Analysis

## OBSERVATIONS:

- Majority of the loans are fully paid.
- About 14% loans are defaulting.
- The loan has increased exponentially over the years.



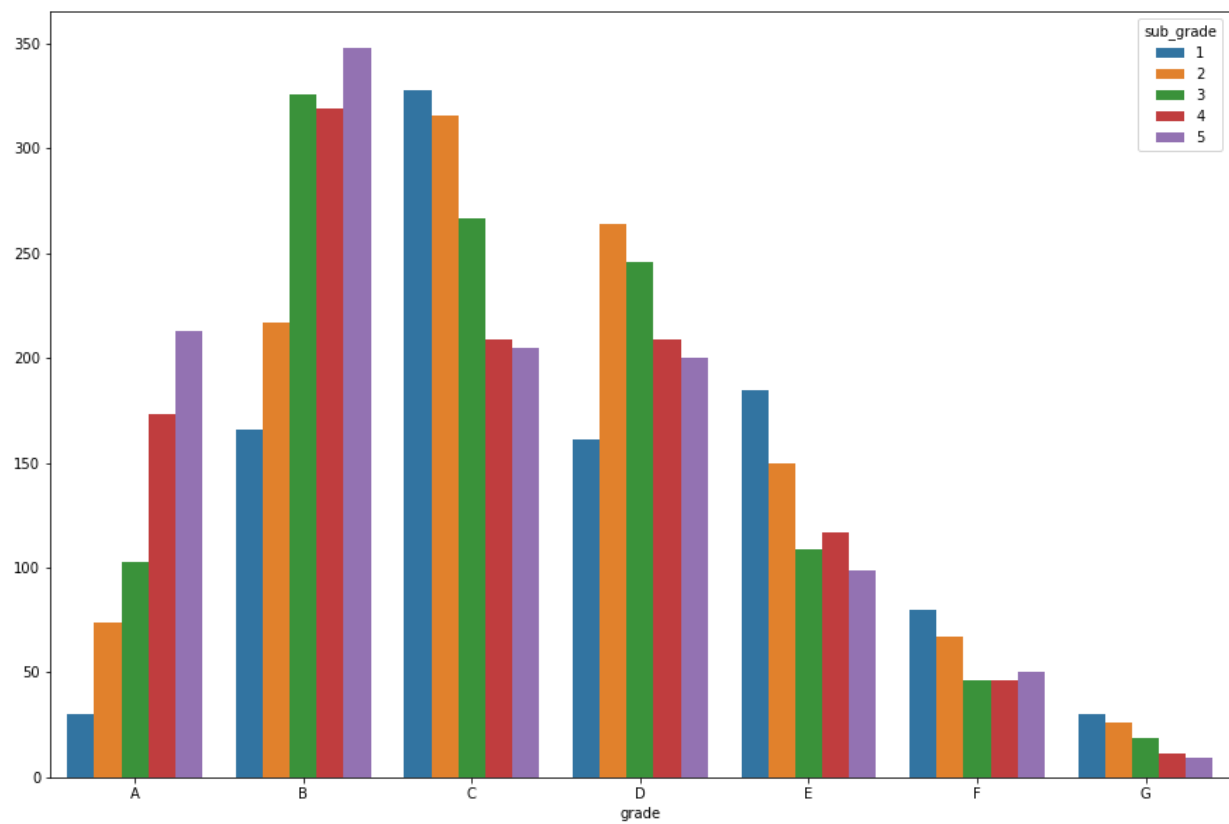
# Interest Rate Analysis



## OBSERVATIONS:

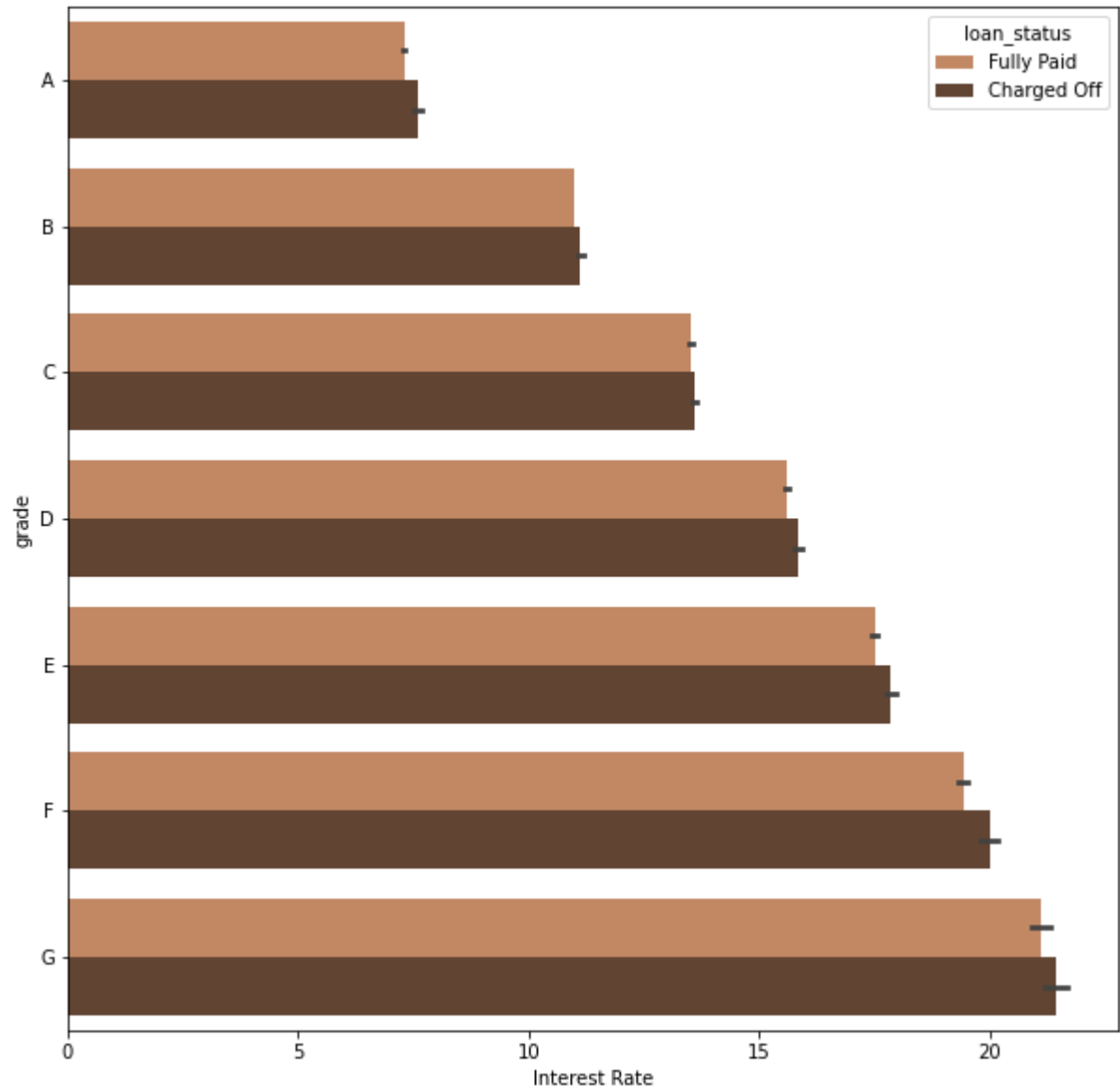
- Average interest rate is 12 %. After 75%, interest rate rose up to 25% from 15%.
- Probability of defaulting is more if loan applicant's interest rate is between 13% to 17%.
- [VERY IMP] The interest rate of charged off loans is very high than that of fully paid loans across all the loan amount groups. This can be really strong driving factor for loan defaulting.

# Grade Analysis



## OBSERVATIONS:

- Probability of defaulting is more if loan applicant grade is 'B'.
- There is a high probability of defaulting when applicants taking loan belong to grade G and interest rate above 20%.

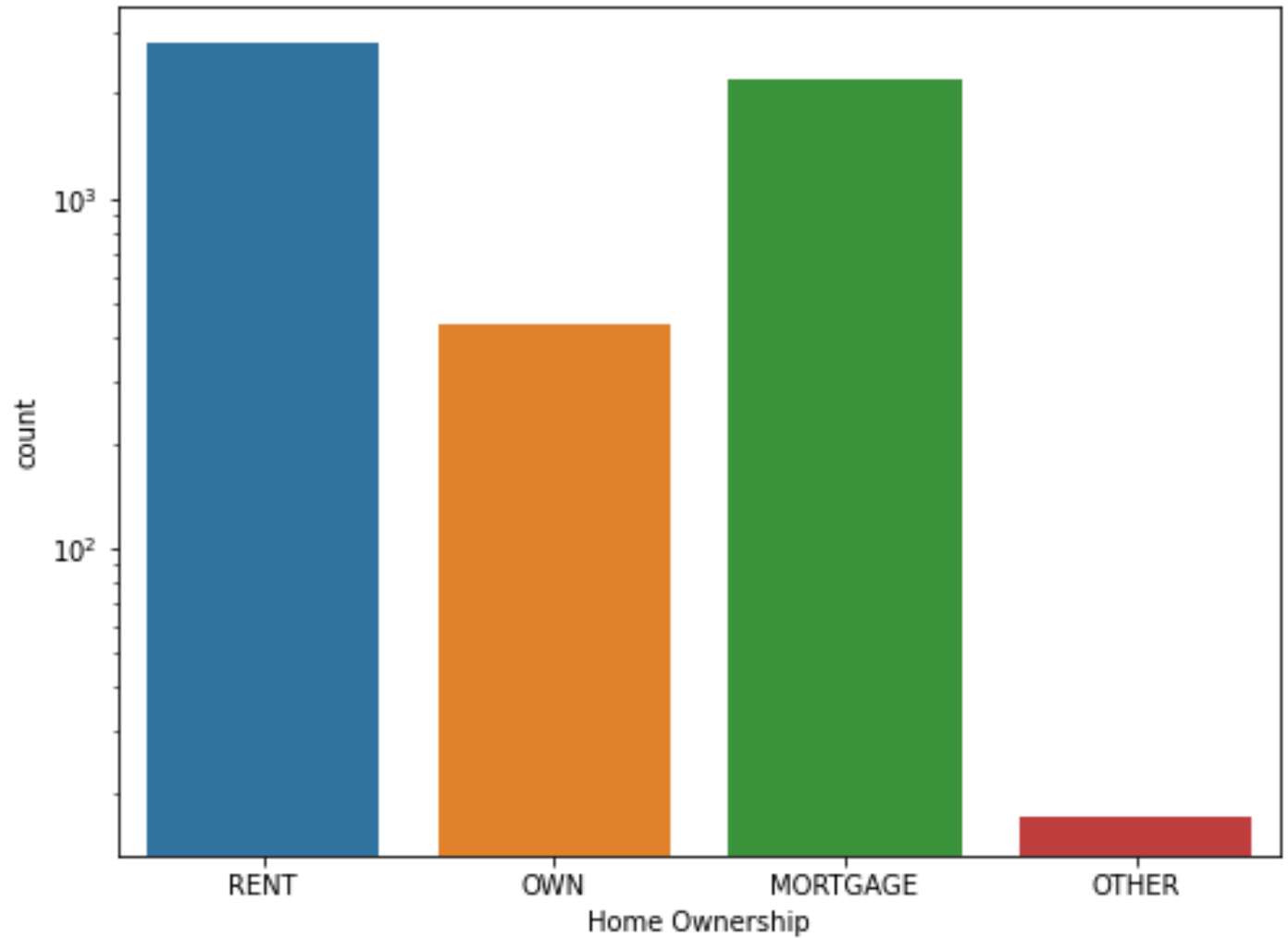


# Home Ownership Analysis

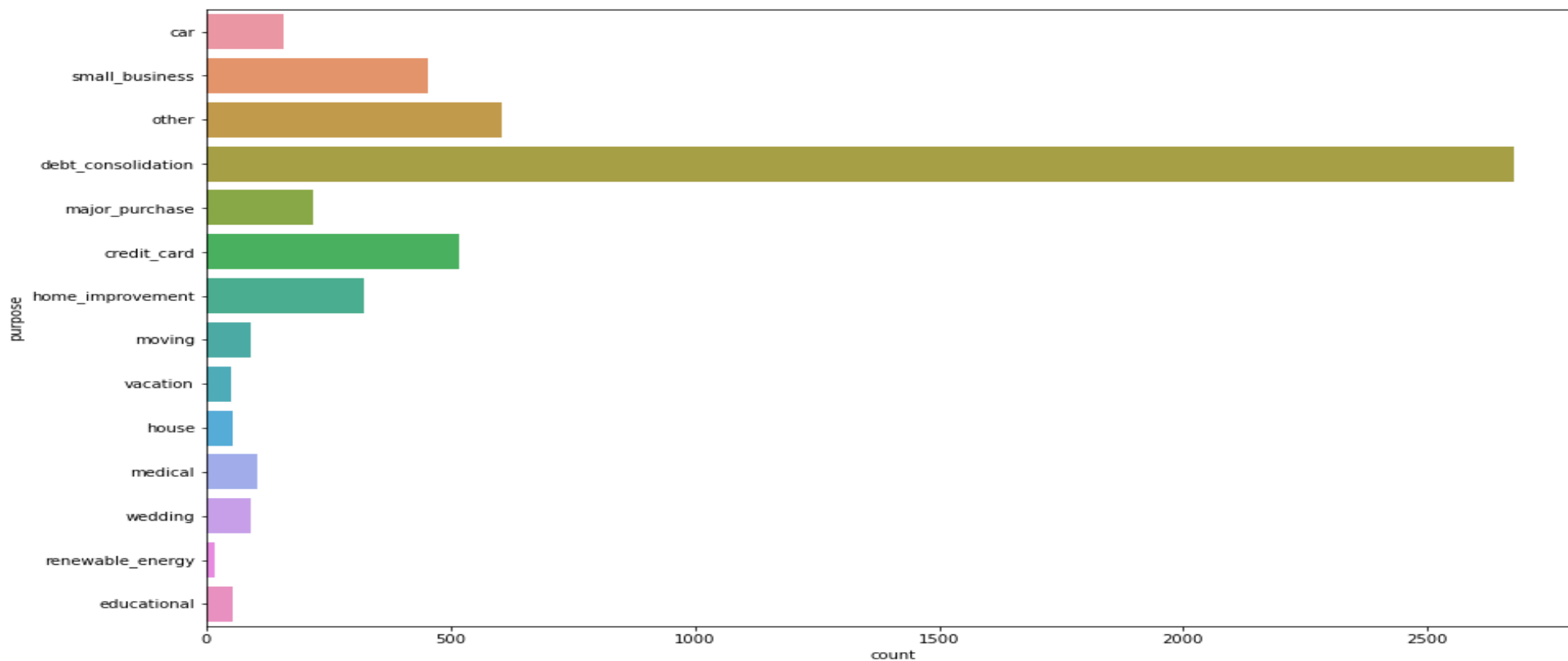
---

## OBSERVATIONS:

- Probability of defaulting is more if loan applicant's home ownership is 'RENT'



# Loan Purpose Analysis



## OBSERVATIONS:

- Probability of defaulting is more if loan applicant's purpose is debt consolidation.

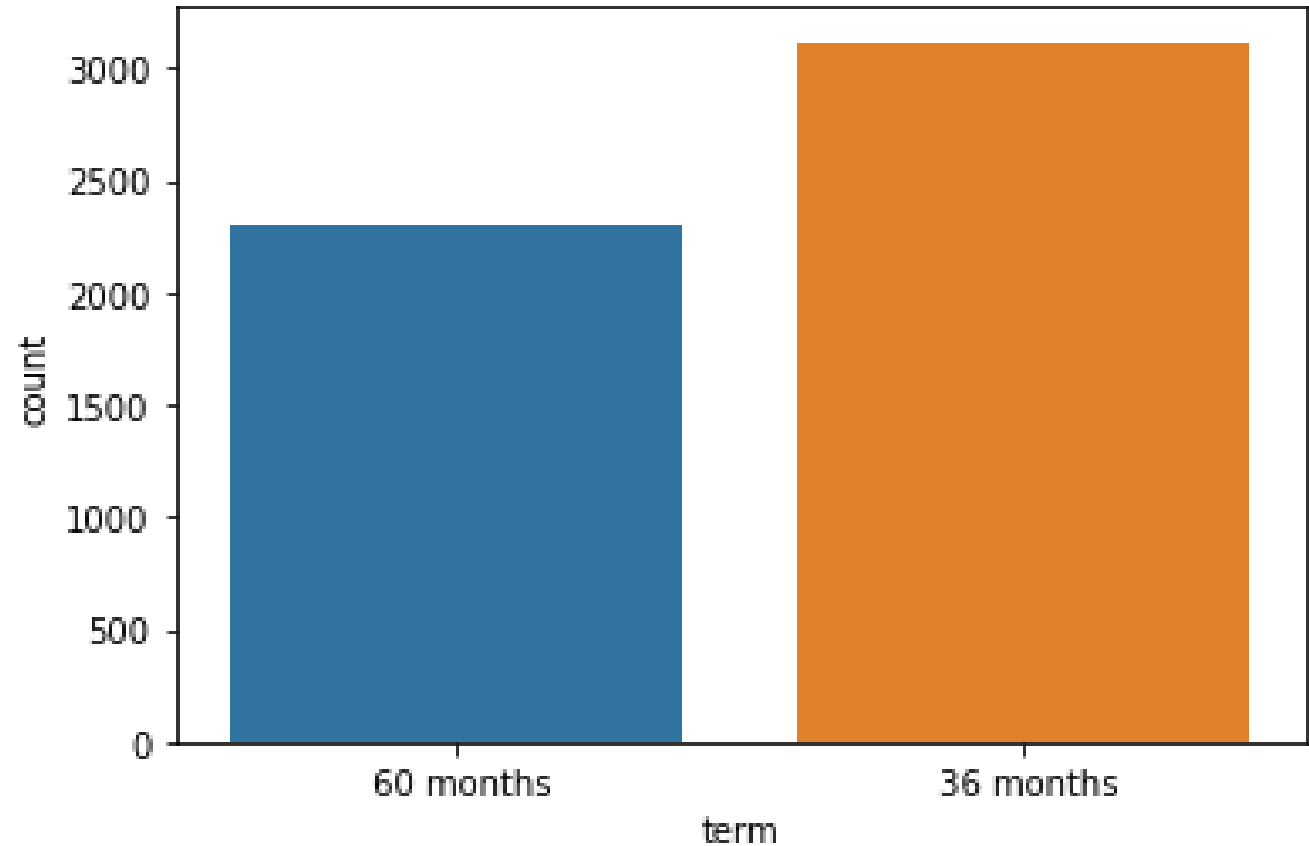


# Loan Term Analysis

---

## OBSERVATIONS:

- Probability of defaulting is more if loan applicant's term is 36 months.

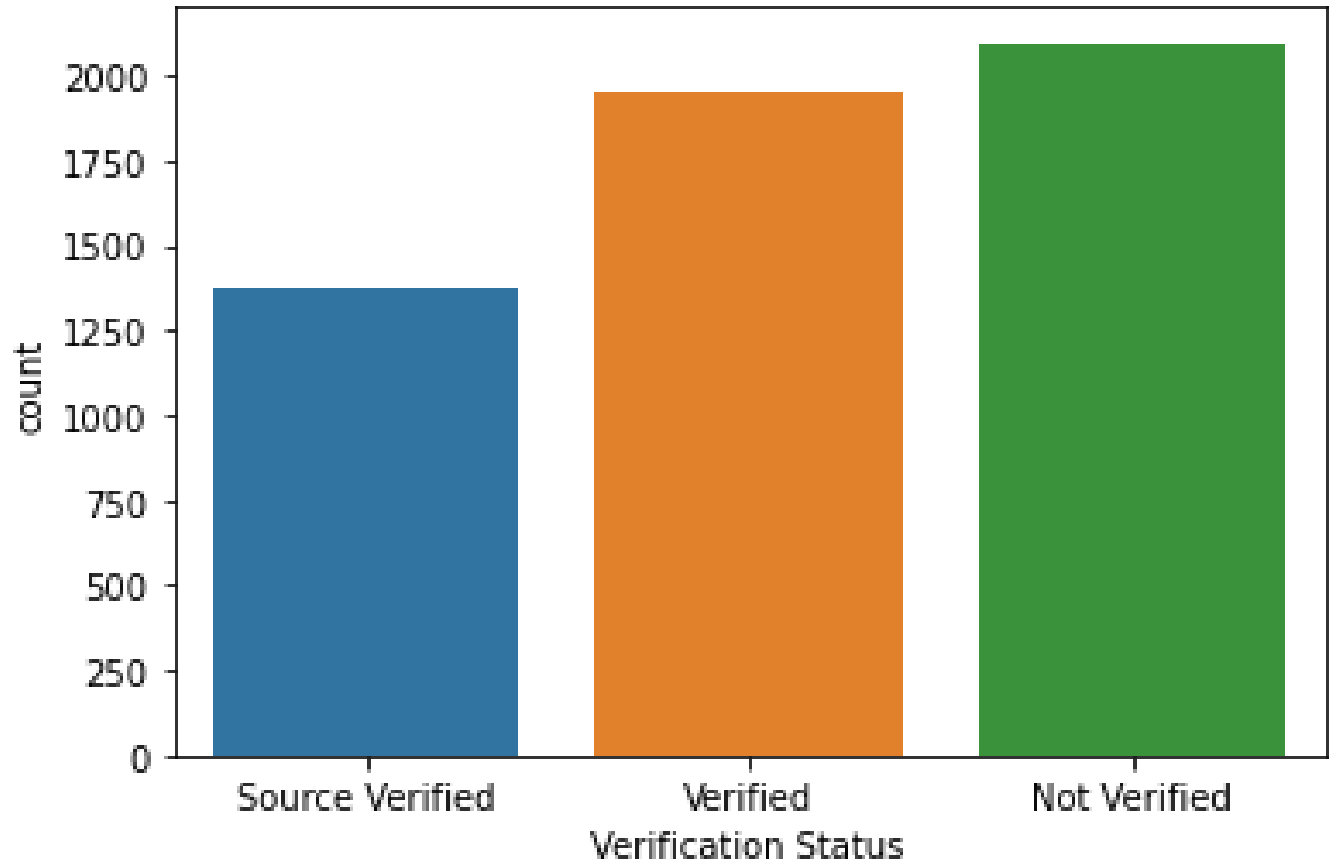


# Verification Status Analysis

---

## OBSERVATIONS:

- Probability of defaulting is more if loan applicant is not verified.

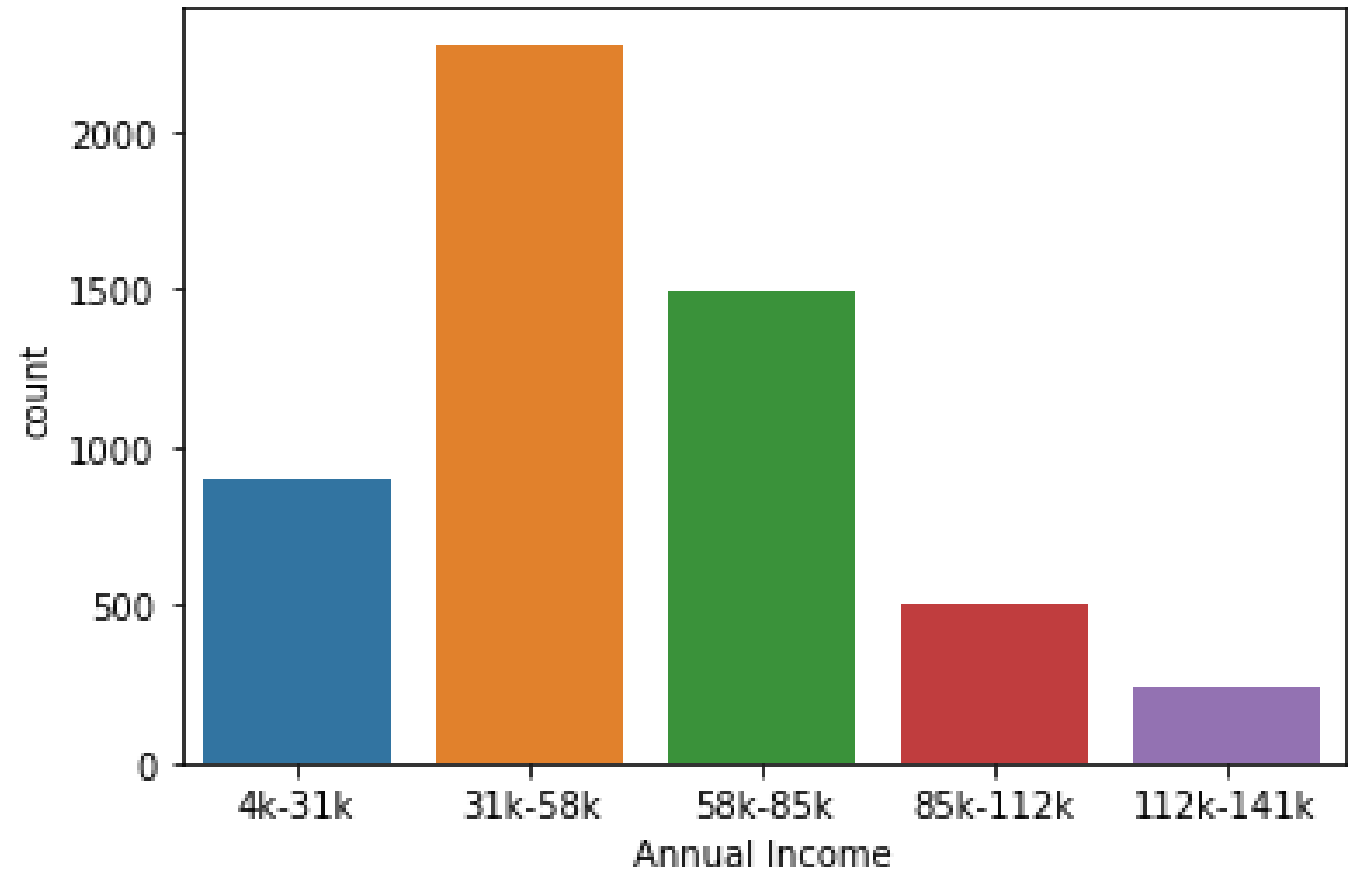


# Annual Income Analysis

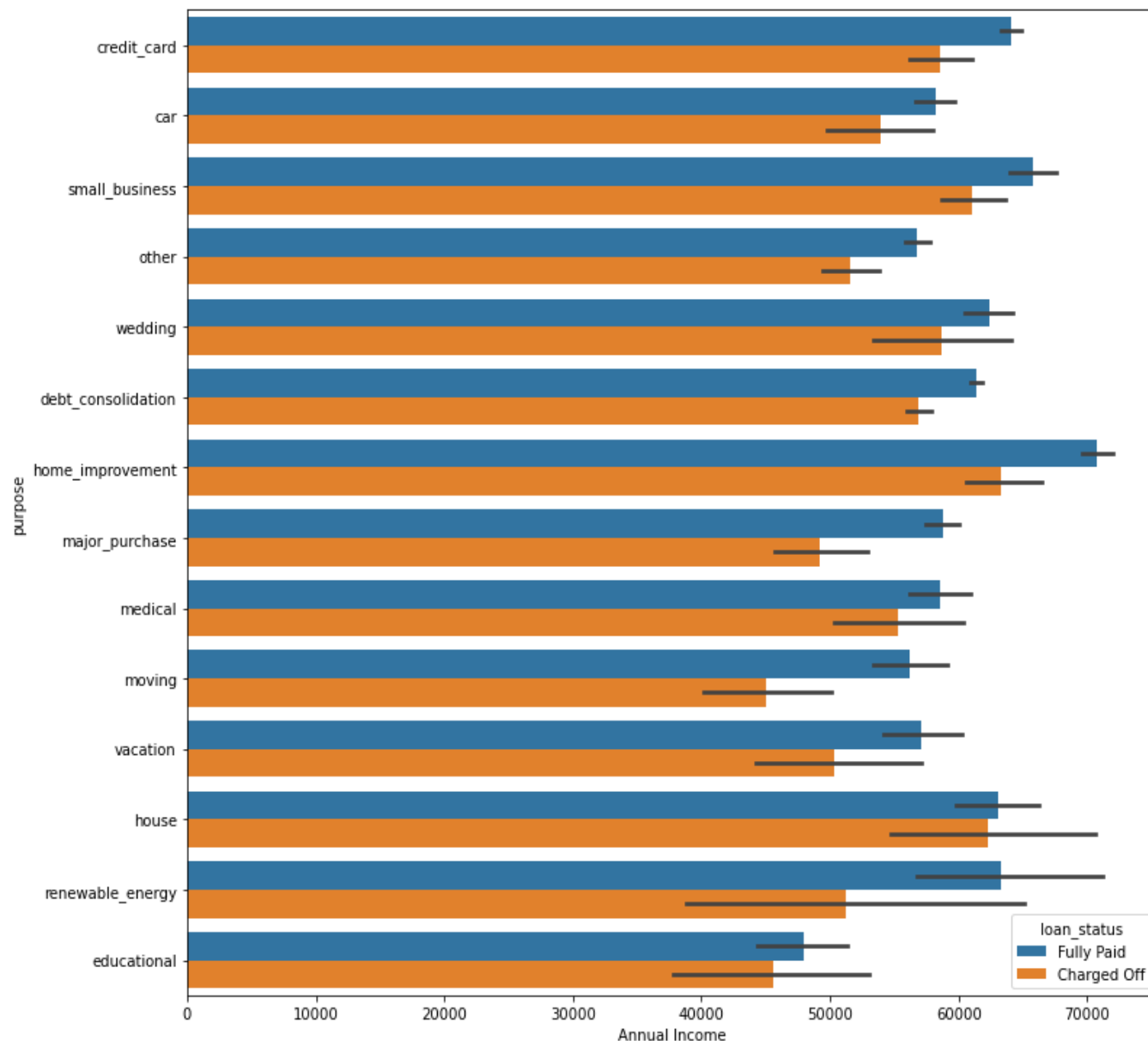
---

## OBSERVATIONS:

- Probability of defaulting is more if loan applicant's annual income is between 31000 to 58000



# Annual Income Analysis-continued



## OBSERVATIONS:

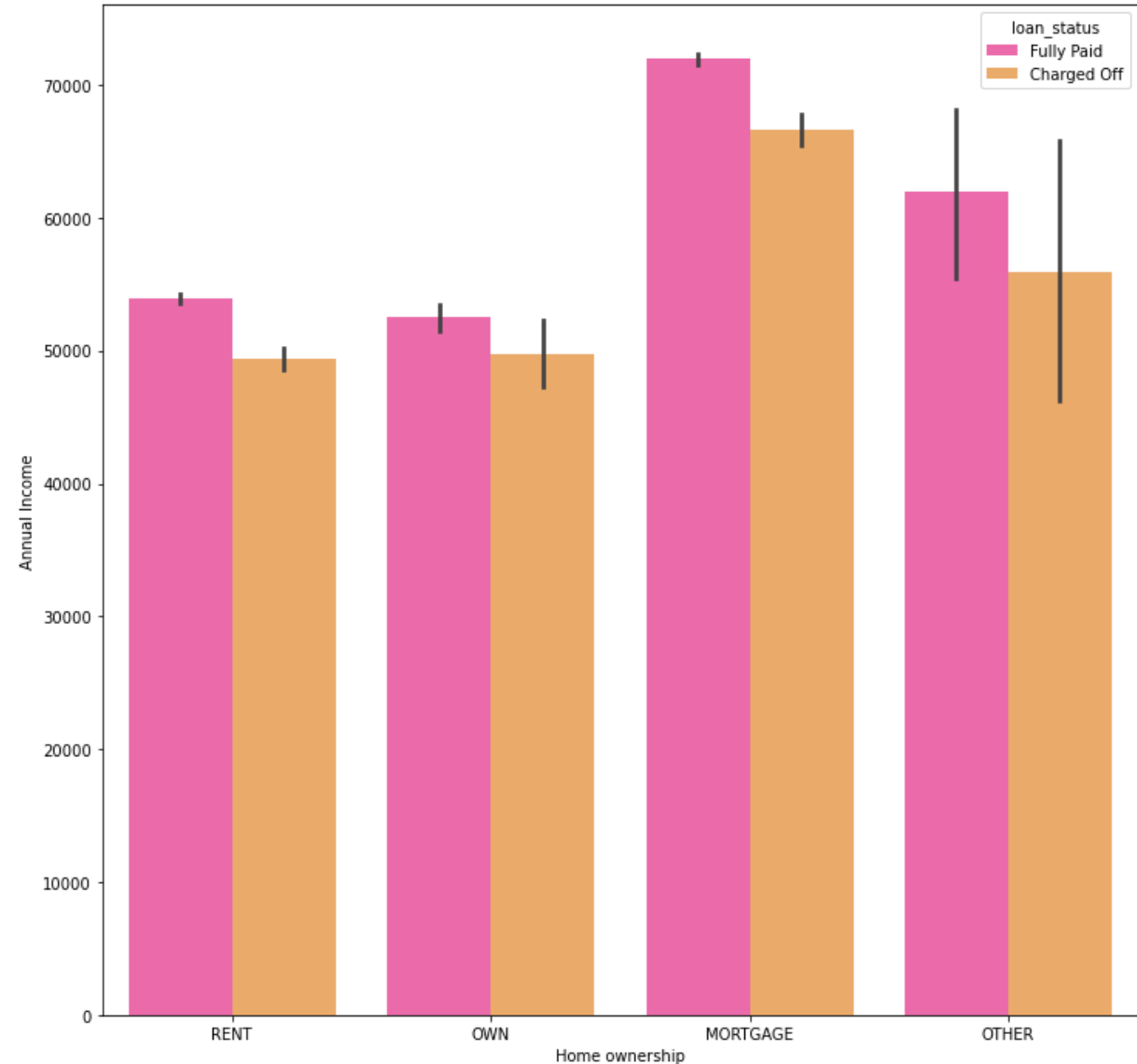
- Though the number of loans applied and defaulted are the highest, as we saw in our analysis earlier, the above analysis shows that the annual income of those who applied is not the highest.
- Applicants with higher salary/annual income has largely applied loans for purposes such as 'home improvement', 'house', 'renewable energy' and 'small business'
- There is a high probability of defaulting when applicants taking loan for 'home improvement' and have income of 60k -70k

# Annual Income Analysis-continued

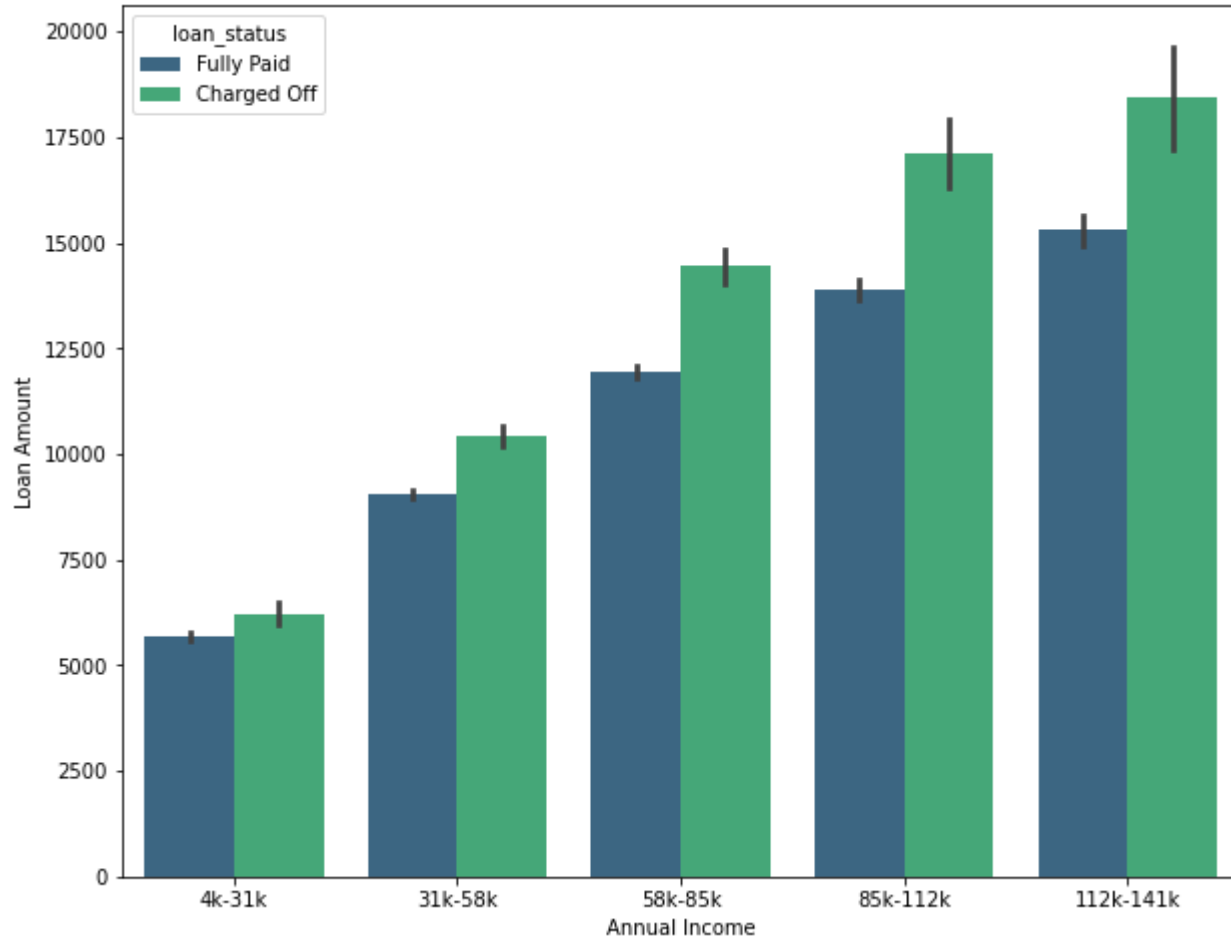
---

## OBSERVATIONS:

- There is a high probability of defaulting when applicants taking loan for whose home ownership is 'MORTGAGE' and have income of 60-70k.

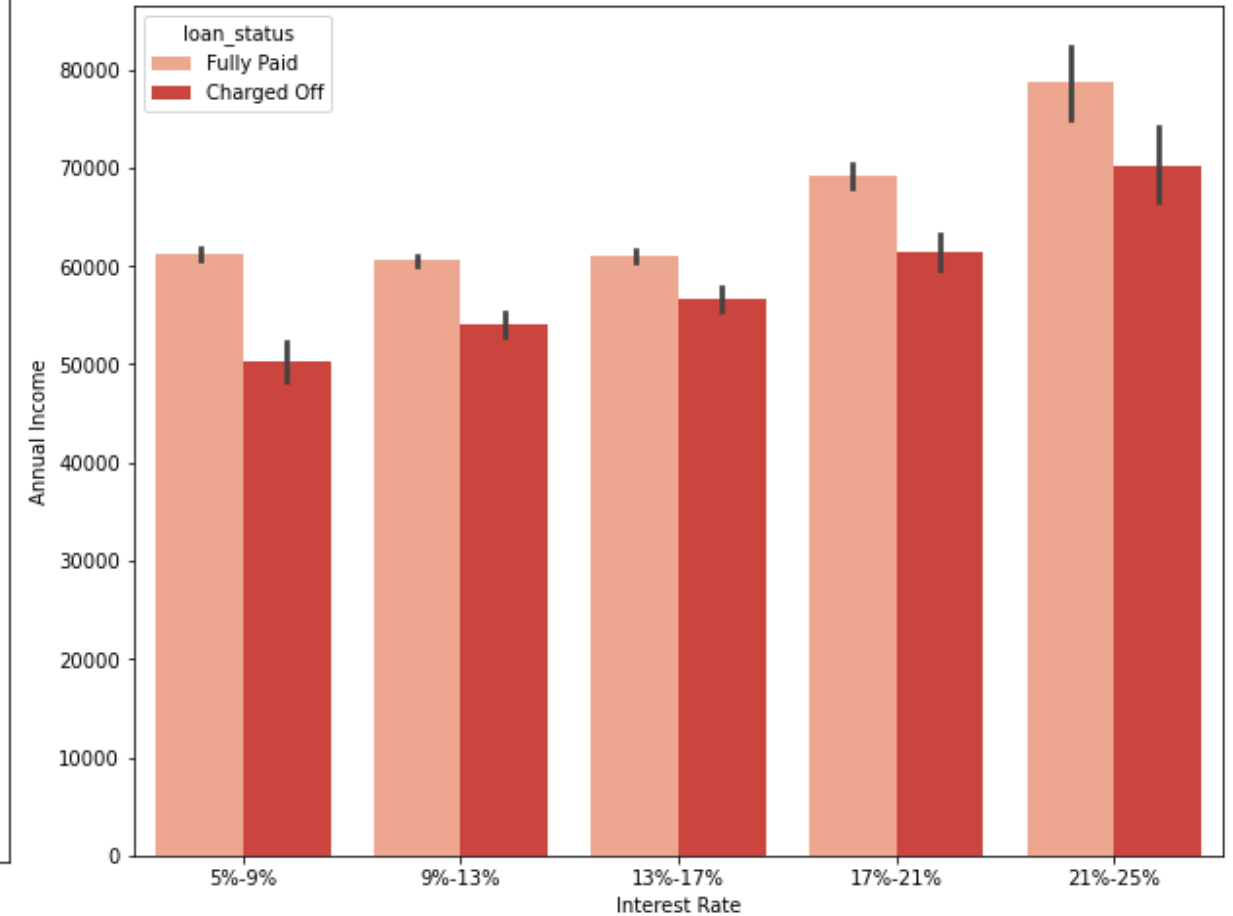


# Annual Income Analysis-continued



## OBSERVATIONS:

- There is a high probability of defaulting when applicants taking high amount of loan.



## OBSERVATIONS:

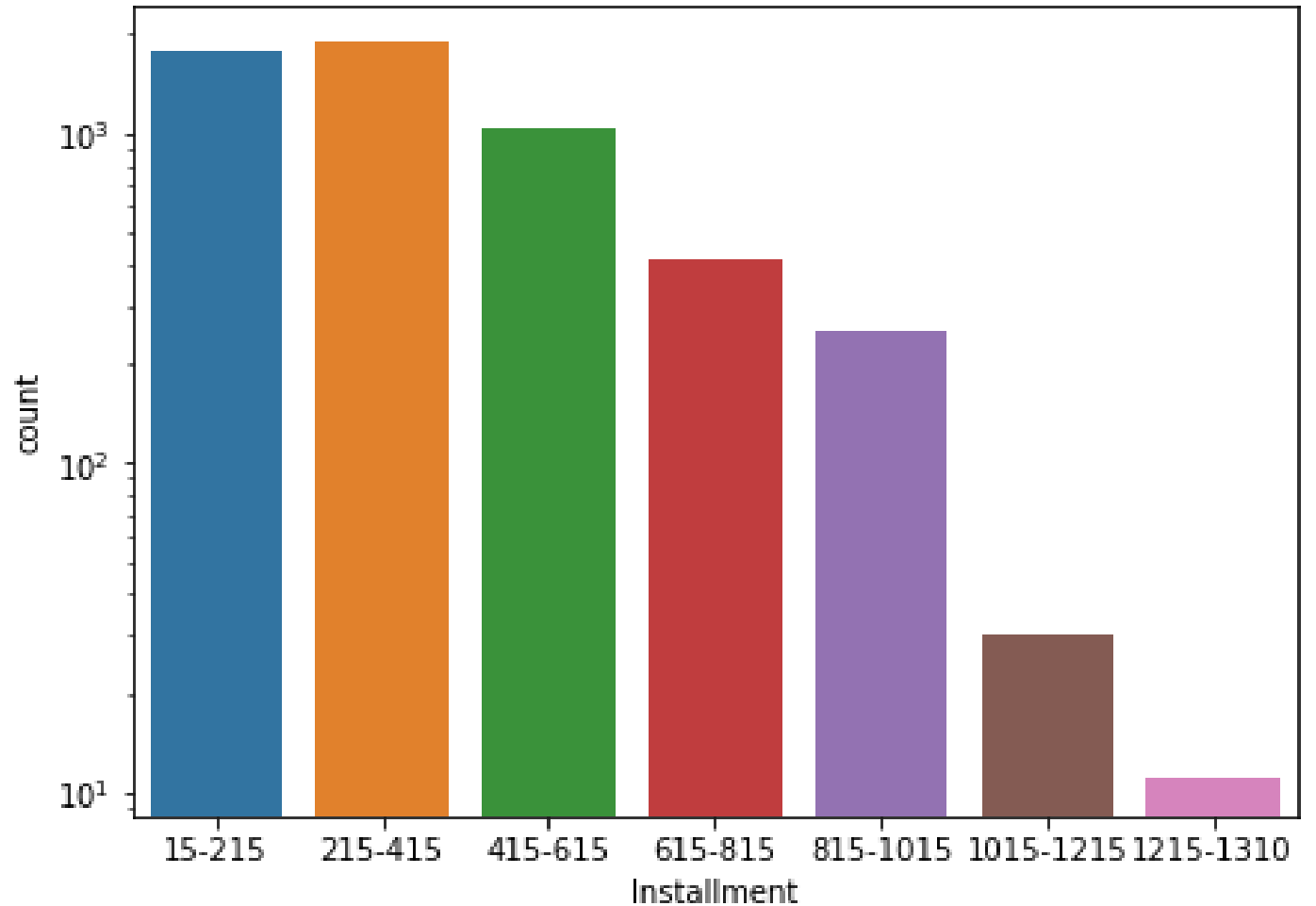
- There is a high probability of defaulting when applicants taking loan receive interest at the rate of 21-25% and have an income of 60k-80k.

# Installment Analysis

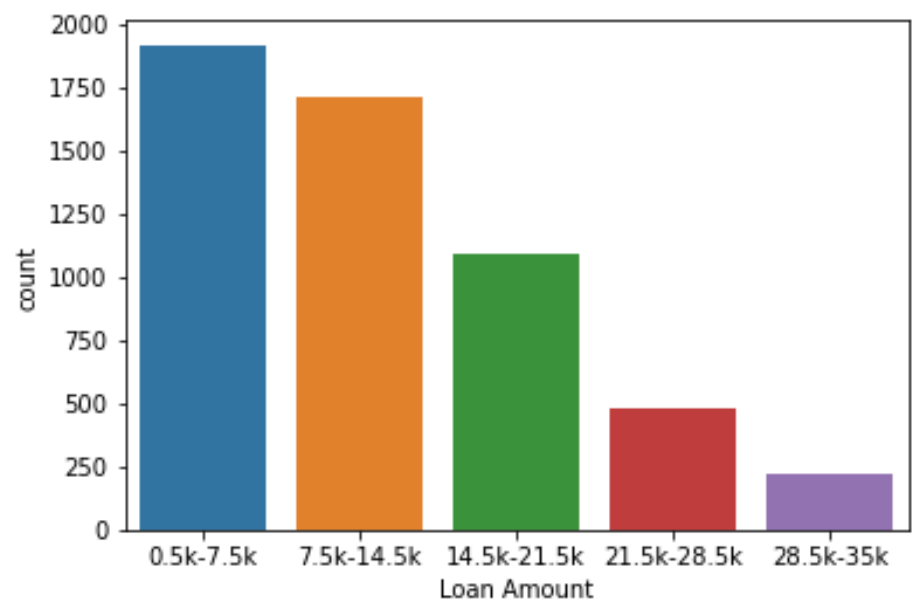
---

## OBSERVATIONS:

- Probability of defaulting is more if loan applicant's installment is between 215 to 415.

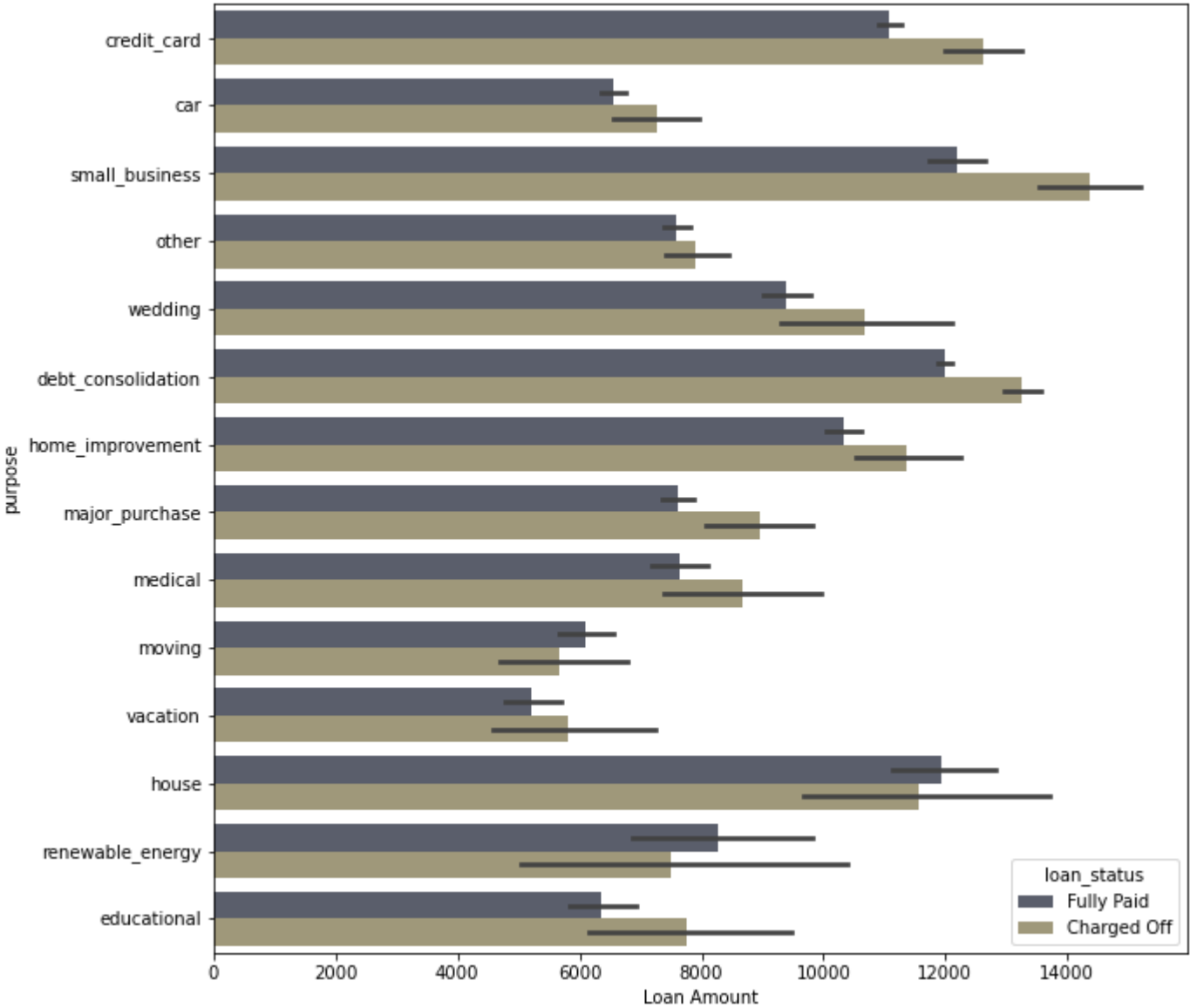


# Loan Amount Analysis



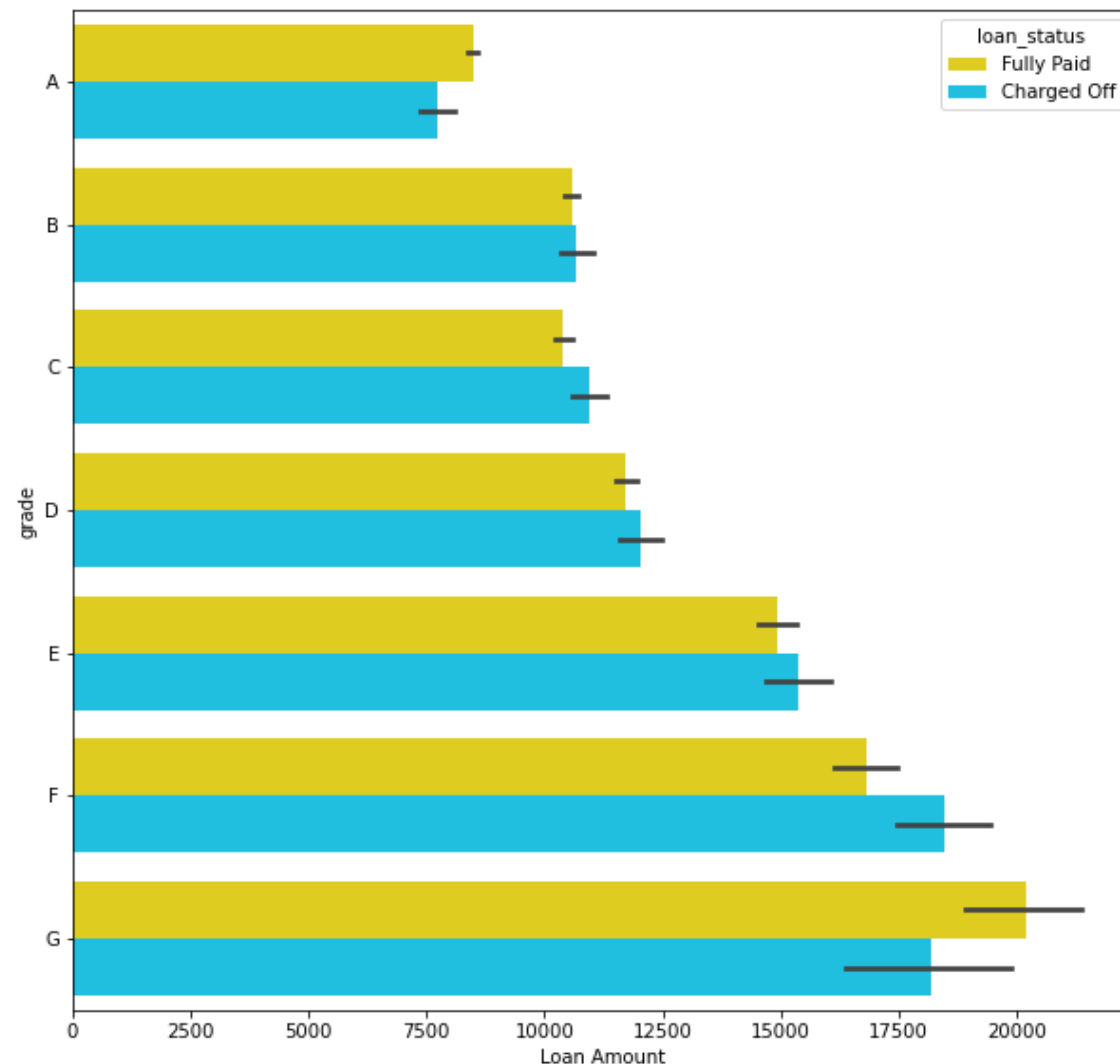
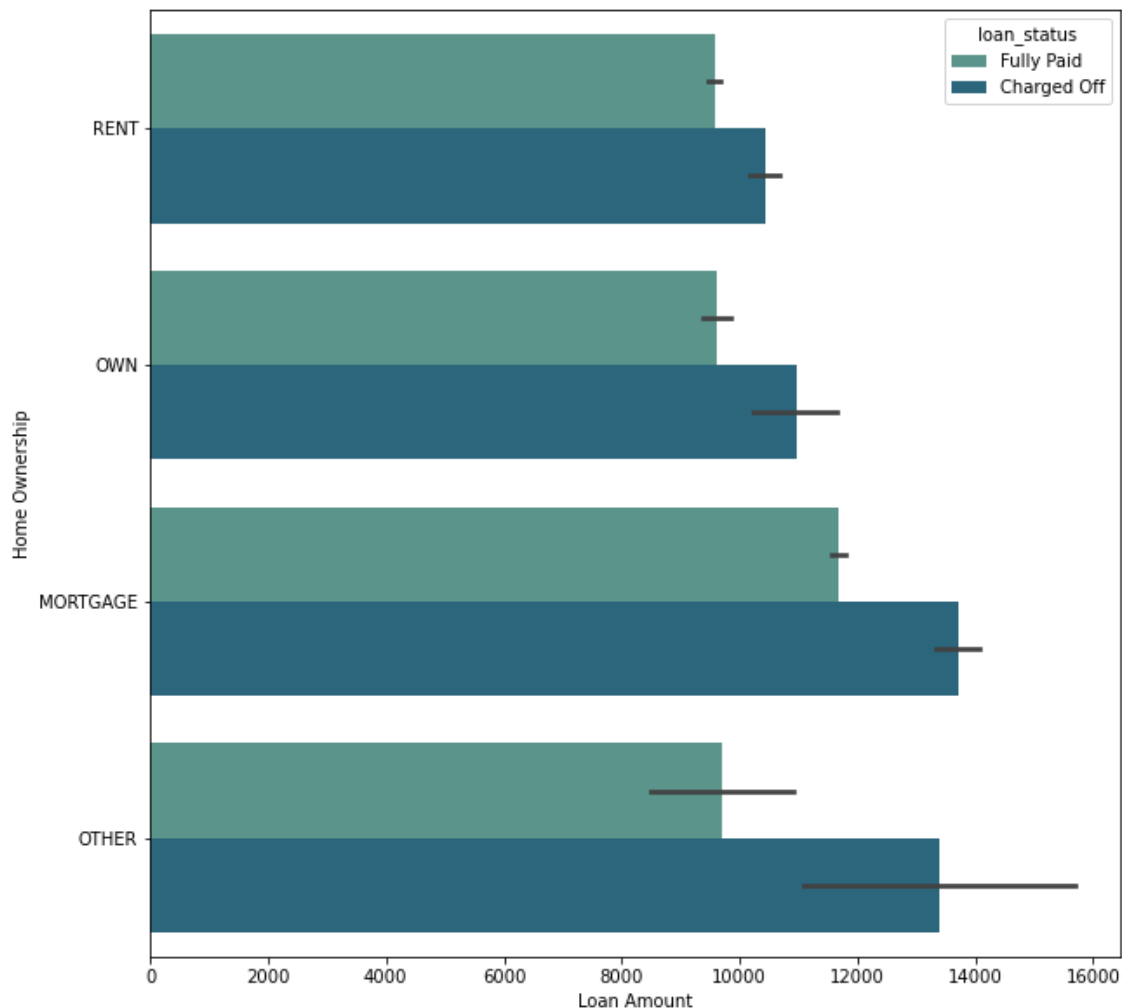
## OBSERVATIONS:

- Probability of defaulting is more if loan amount of app is between 500 to 7500.
- There is a high probability of defaulting when applicant taking loan for small business and the loan amount is greater than 14k.





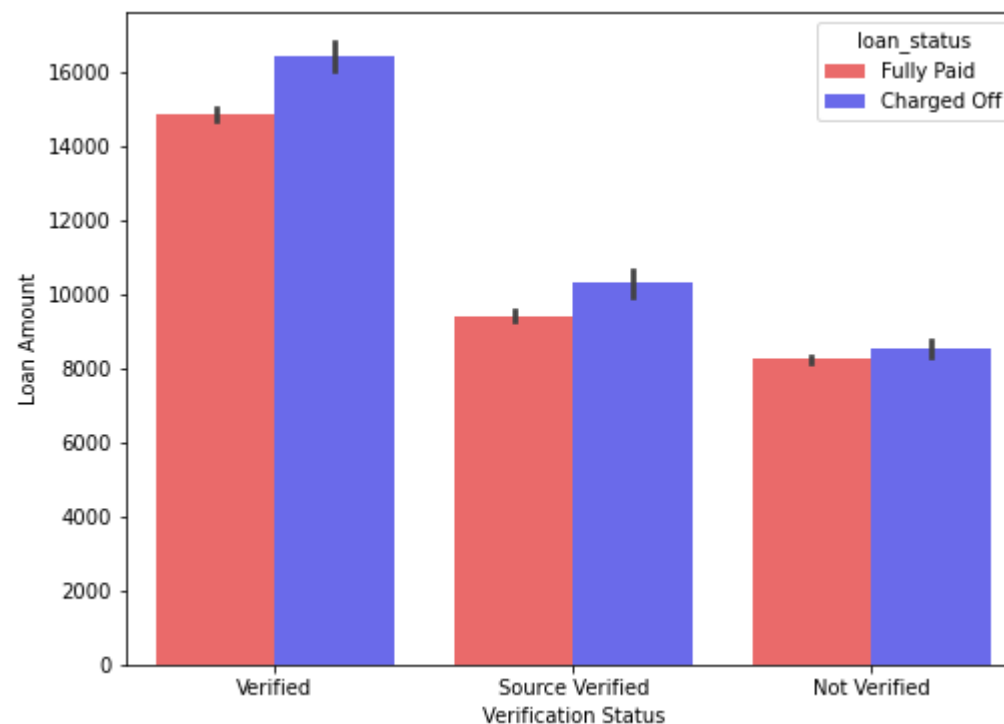
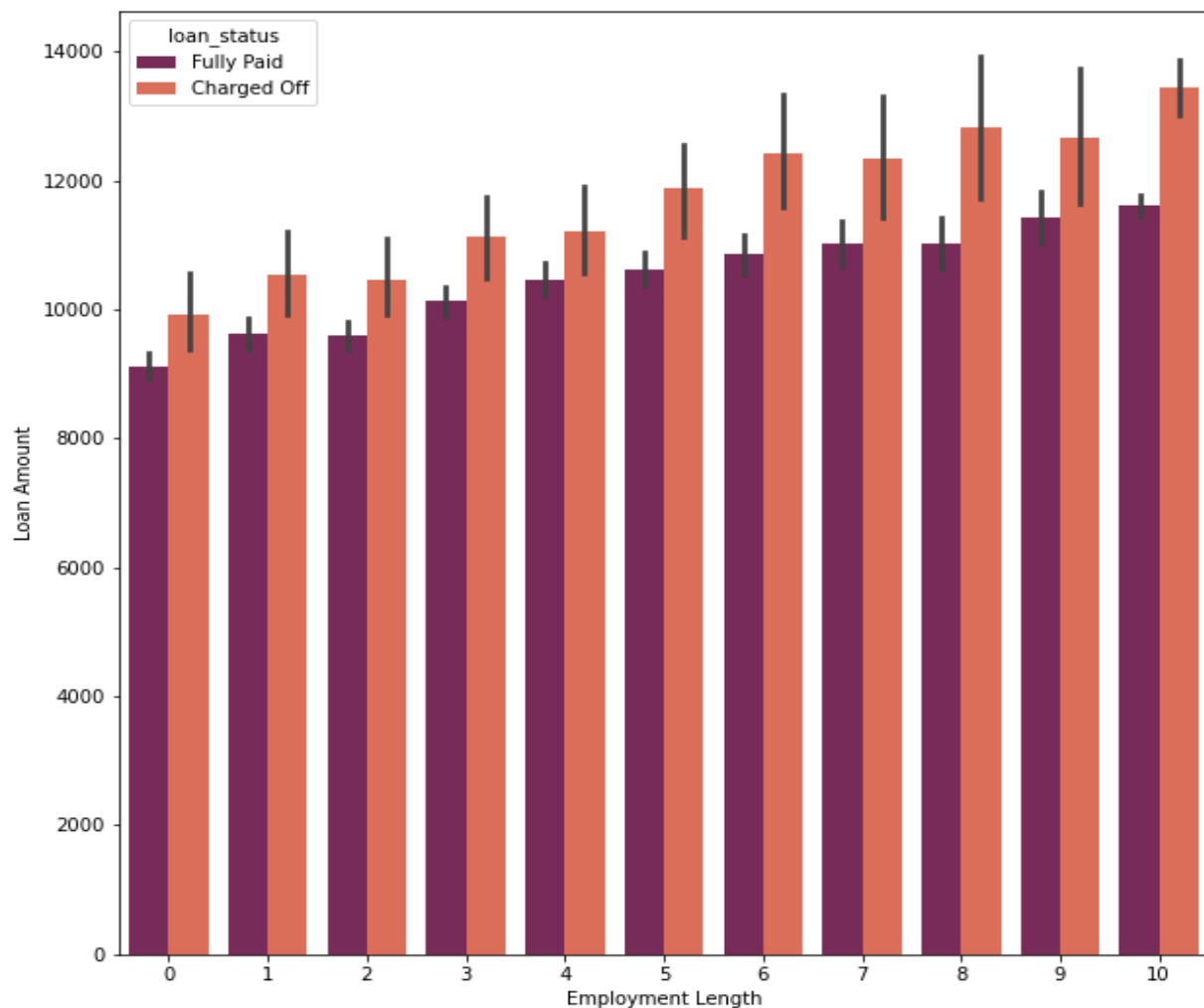
# Loan Amount Analysis-continued



## OBSERVATIONS:

- There is a high probability of defaulting when applicants taking loan whose home ownership is 'MORTGAGE' and have loan of 14-16k.
- There is a high probability of defaulting when applicants taking loan belong to grade F and loan amount is between 15k-20k.

# Loan Amount Analysis-continued



## OBSERVATIONS:

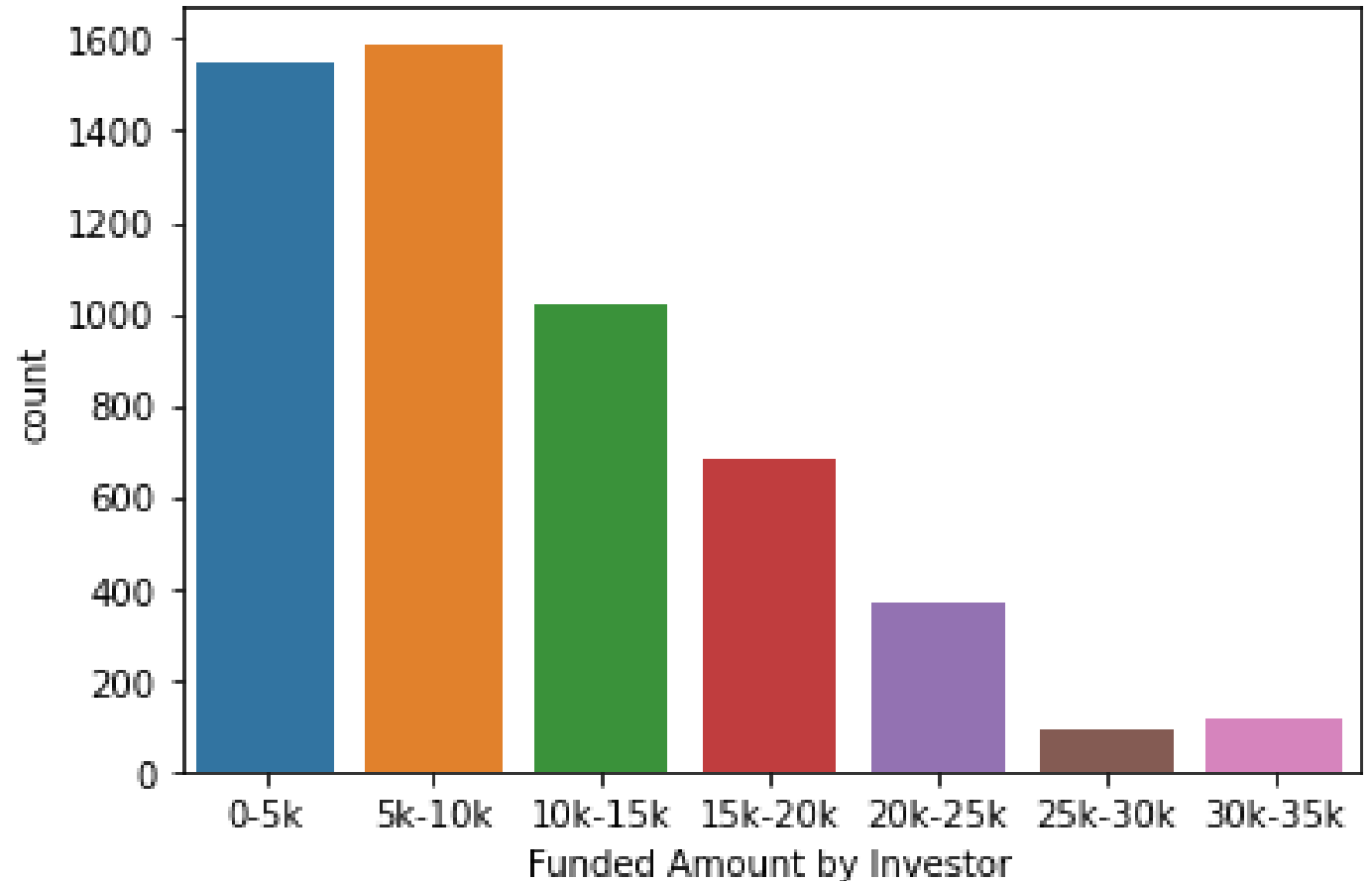
- There is a high probability of defaulting when applicants taking loan have 10 years of employment length and loan amount is 12k-14k.
- There is a high probability of defaulting when applicants taking loan amount above 16k and loan is verified.

# Funded Amount by Investor Analysis

---

## OBSERVATIONS:

- Probability of defaulting is more if funded amount by investor of applicant is between 5000 to 10000.

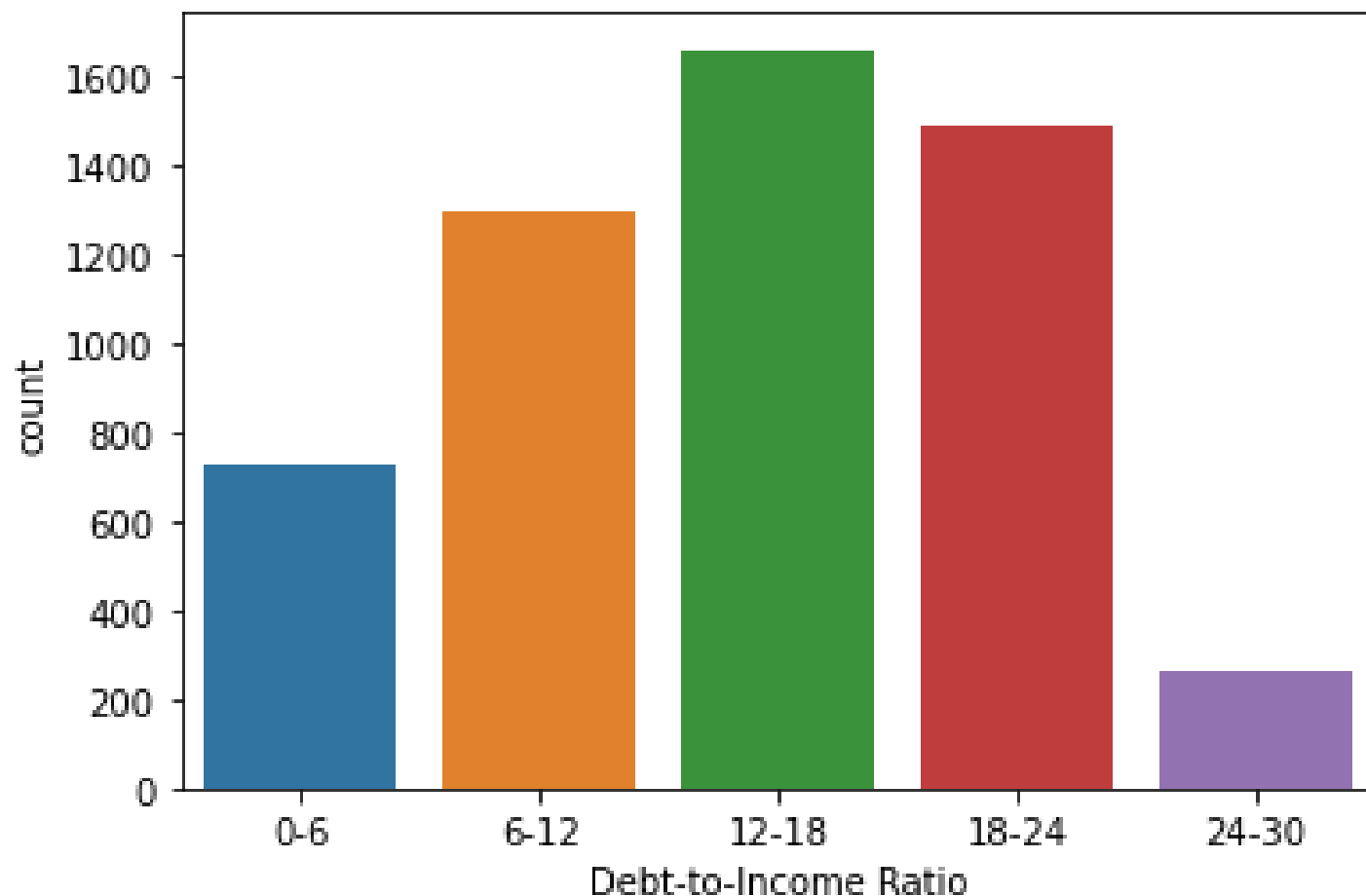


# Debt-to-Income Ratio Analysis

---

## OBSERVATIONS:

- Probability of defaulting is more if debt-to-income ratio of applicant is between 12 to 18.

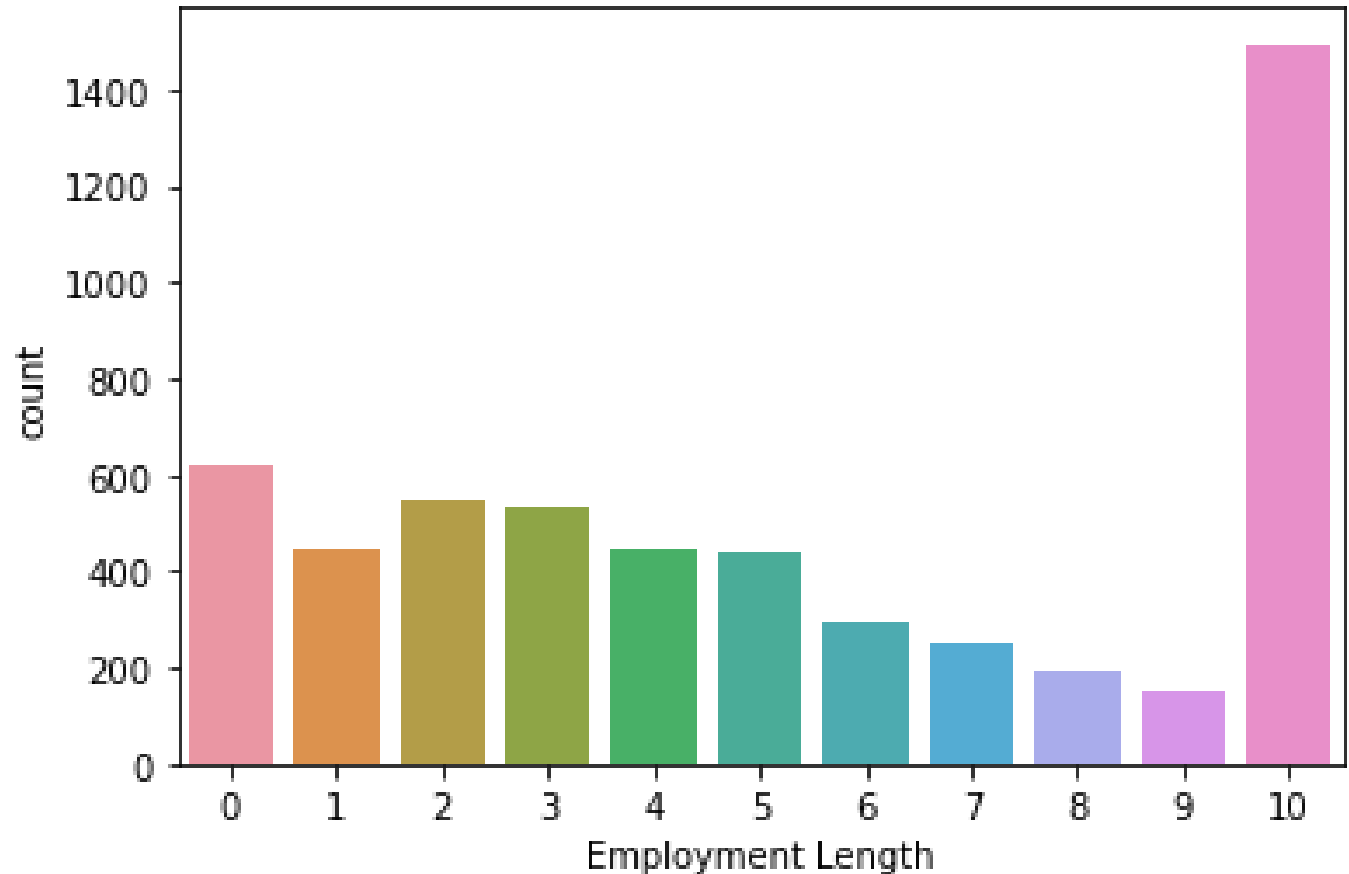


# Employment Length Analysis

---

## OBSERVATIONS:

- Probability of defaulting is more if loan applicant has employment length of 10 or 10+.

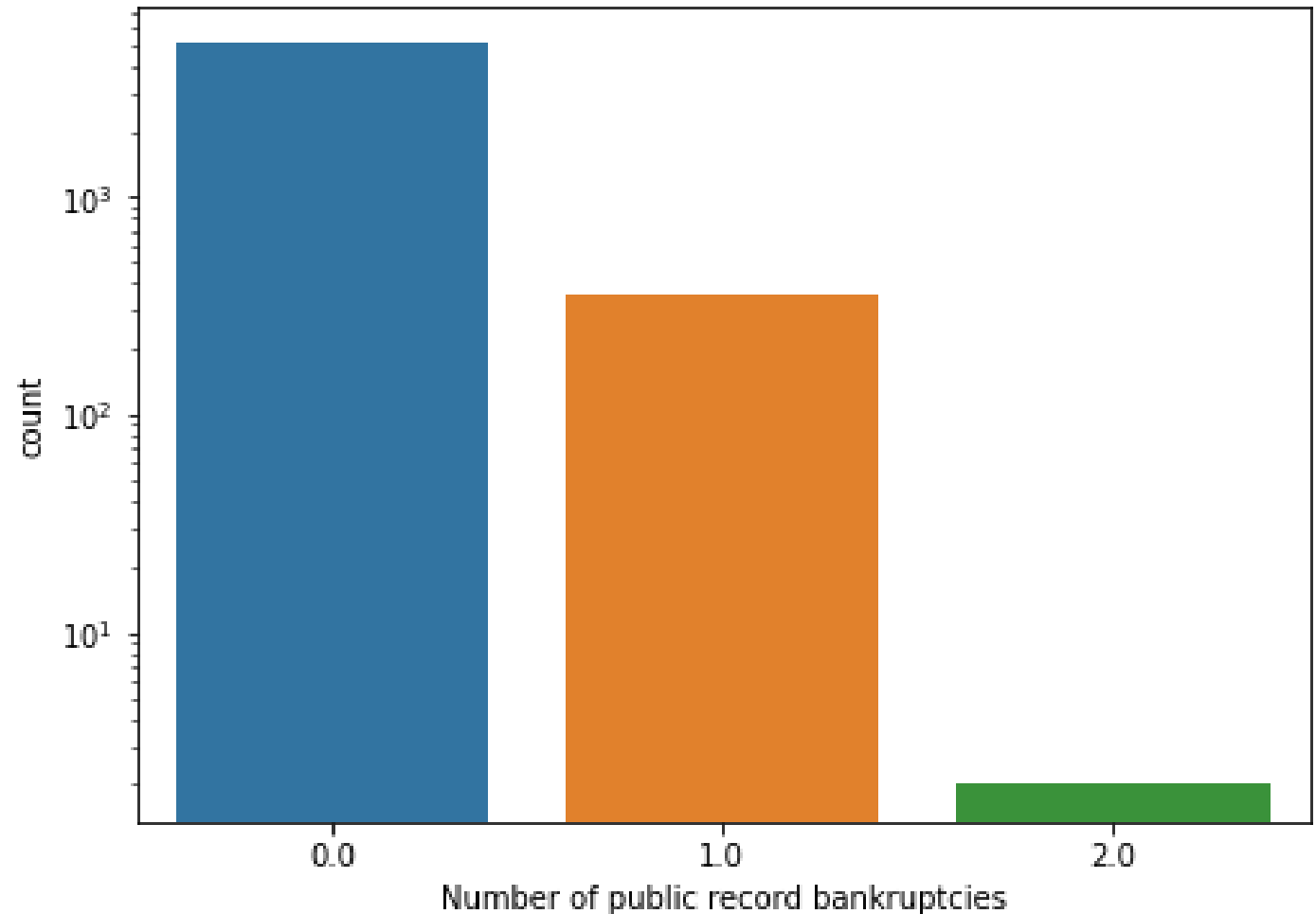


## No. of public record bankruptcies Analysis

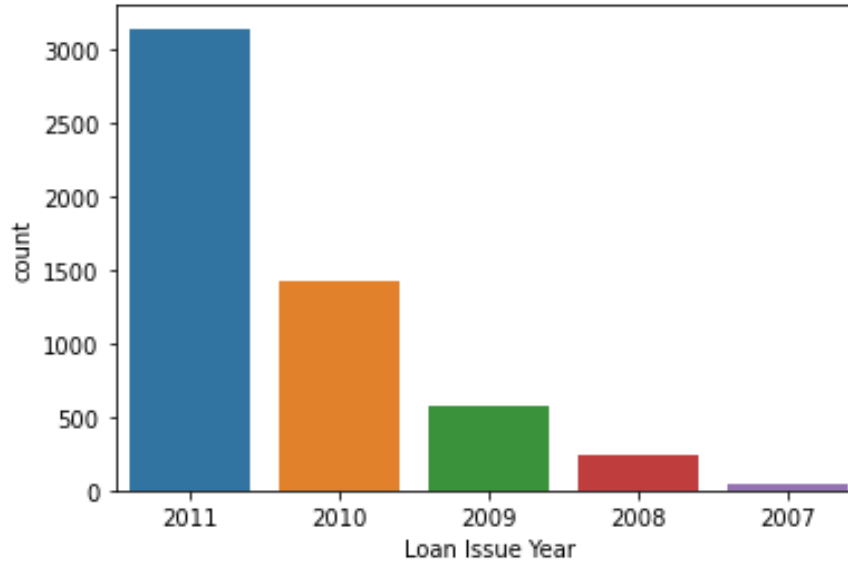
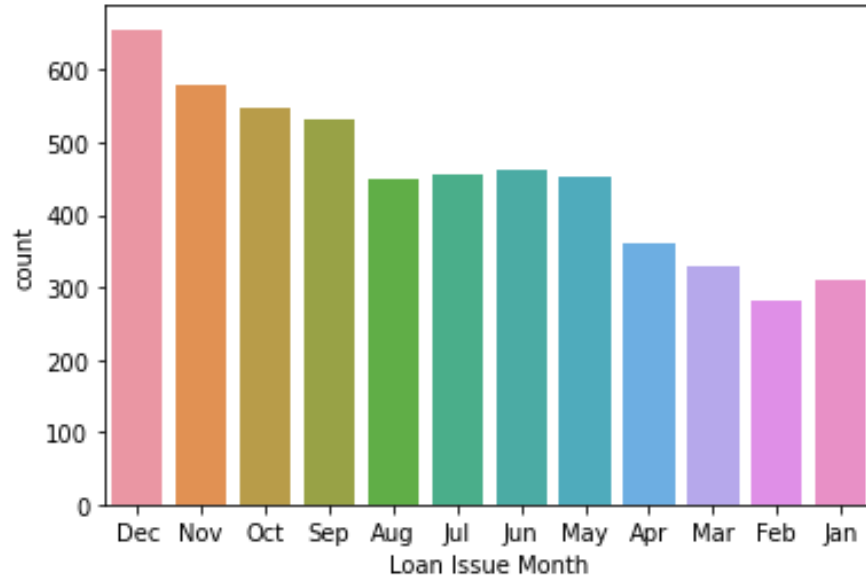
---

### OBSERVATIONS:

- probability of defaulting is more if loan applicant has 0 number of bankruptcies.



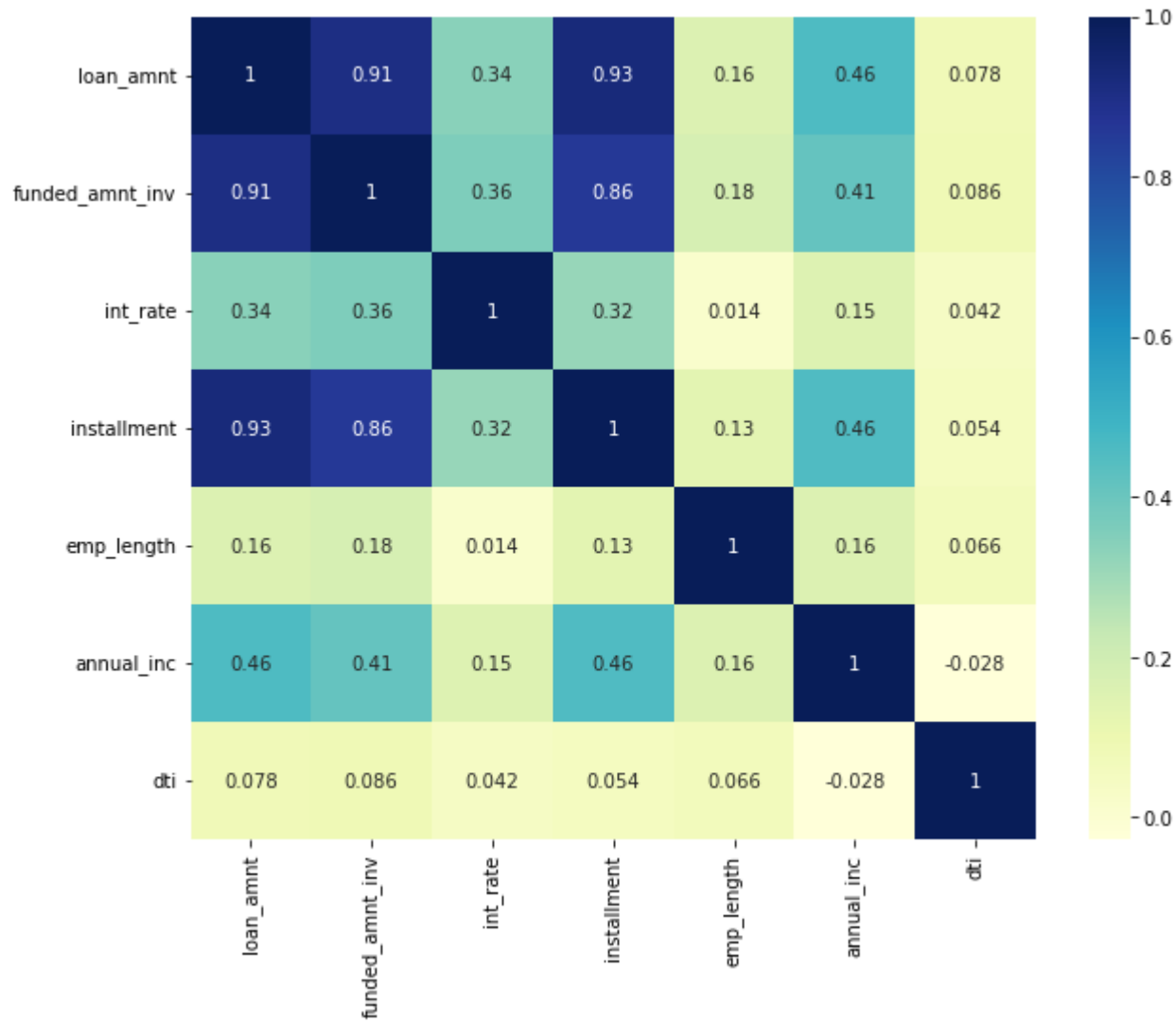
# Loan Issue Analysis



## OBSERVATIONS:

- Probability of defaulting is high for the loan issued in the month of December, possible due to the fact that people take loan to travel and spend money during the Christmas season.
- Probability of defaulting is high for the loan issued in the year of 2011, possibly due to the fact that there was a financial crisis in USA, assuming the data is of US origin.

# Analysis to understand the correlation among variables

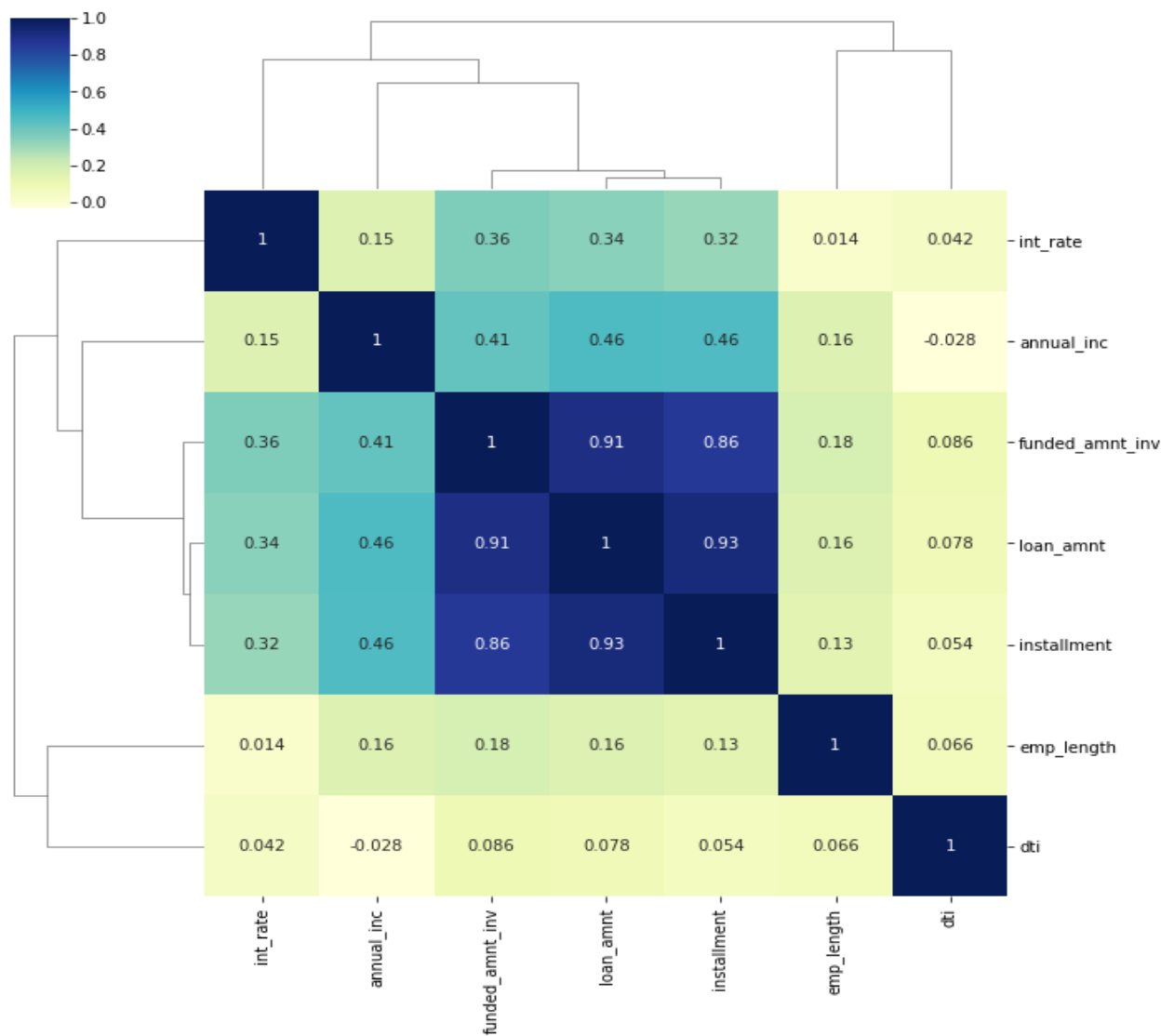


## OBSERVATIONS:

- Probability of defaulting is high for the loan issued in the month of December, possible due to the fact that people take loan to travel and spend money during the Christmas season.
- Probability of defaulting is high for the loan issued in the year of 2011, possibly due to the fact that there was a financial crisis in USA, assuming the data is of US origin.



# Analysis to understand the correlation among variables-continued



## OBSERVATIONS:

The applicants having high chances of defaulting have following correlation behavior:

- Loan Amount, Installment and Funded amount by investors are highly correlated. That means, higher the loan amount, higher the installment, and higher the funded amount by investors.
- Annual income is somewhat correlated but doesn't impact much in terms of funded amount by investors and the loan amount applied.
- Interest rate is not correlated to Annual Income. That means, the annual income of the defaulting loan applicant doesn't influence the interest rate proposed.
- Debt-to-Income ratio is negatively correlated to Annual Income. That means, lesser the annual income, higher the debt.

# Conclusion/Recommendations

---

