LENDING CLUB CASE STUDY

ATHIRA S MENON

POOJA DALVI

PROBLEM STATEMENT

Background :

- Lending Club is a consumer finance company which specializes in lending various types of loans to urban customers.
- 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:
    1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
    2. 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.

Objective :

- Able to identify risky loan applicants so that 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.
- Perform an analysis to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- So that company can utilize this knowledge for its portfolio and risk assessment.

ANALYSIS APPROACH

- Data Exploration :
  - Understanding the metadata
  - Checking for presence of null value or duplicate rows
- Data Cleaning
  - Perform imputation techniques for missing value.
  - Dropping irrelevant columns
  - Standardizing values
  - Removing columns with single value

ANALYSIS APPROACH

- Univariant analysis
  - Understand the statistics of each column with describe() function in pandas
  - Plot box plot and distribution chart to understand data distribution.
  - Identify outliers: Outliers values are removed from data.
  - Plot count plot for each column and observations are noted for both categorical and continuous variables
- Bivariant Analysis
  - Correlation matrix is used to understand the correlation among data variables.
  - Analyze annual income with other column using barplot for more insights
  - Analyze loan amount with other column using barplot for more insights

OBSERVATIONS – DATA EXPLORATION & DATA CLEANING

Total no. of attributes : 111

Total no. of rows : 39717

No. of Duplicate rows : 0

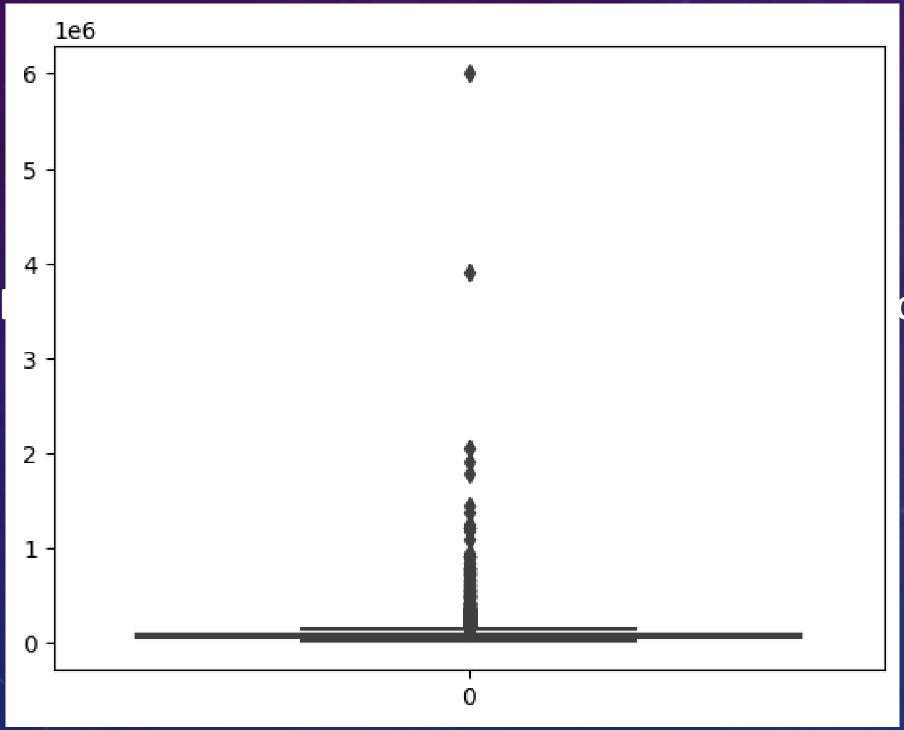
No. of Columns having all null values : 54

No. of Columns having single value : 9

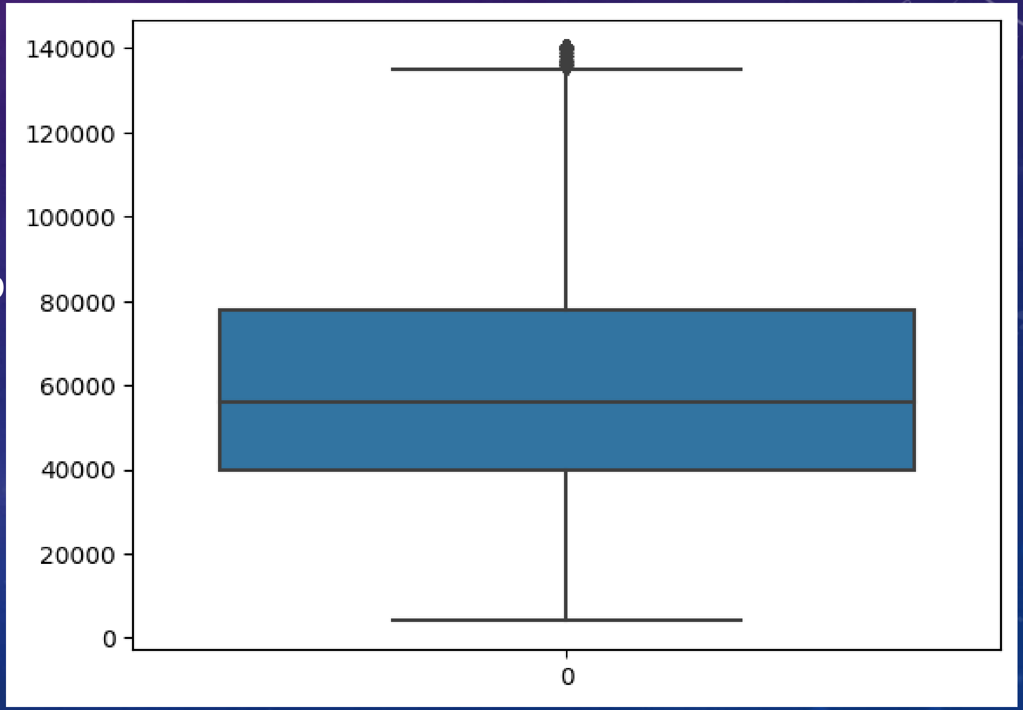
Final shape of data after removing irrelevant data : (39717, 21)

OBSERVATION– UNIVARIANT ANALYSIS

- ce of o



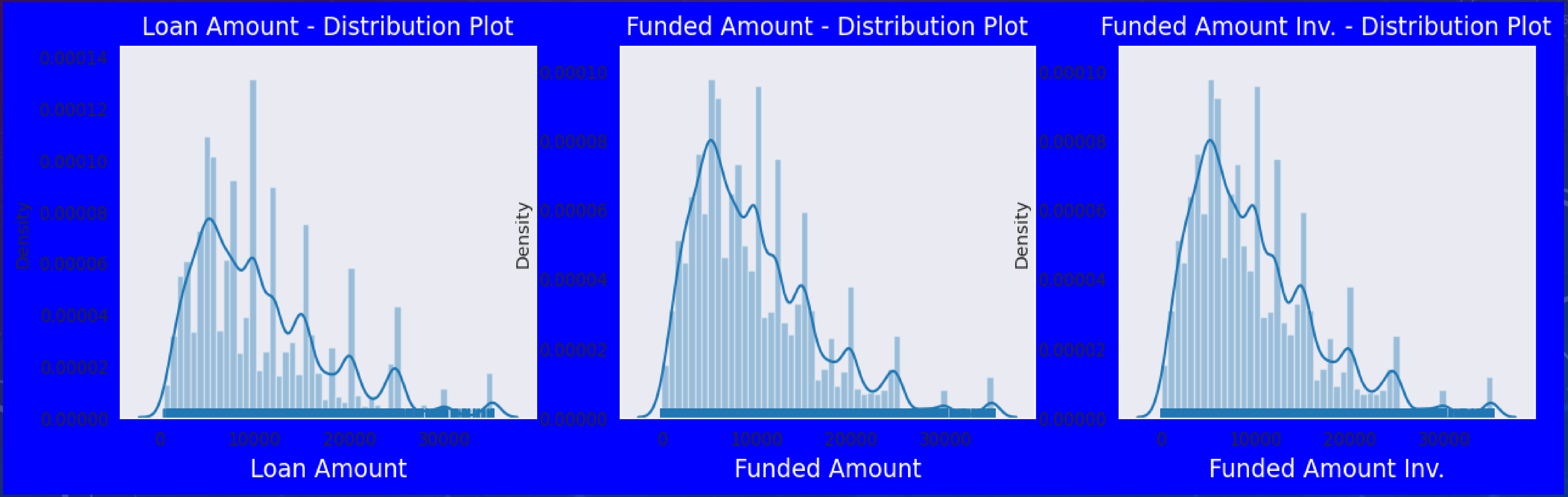
Box plot with outlier



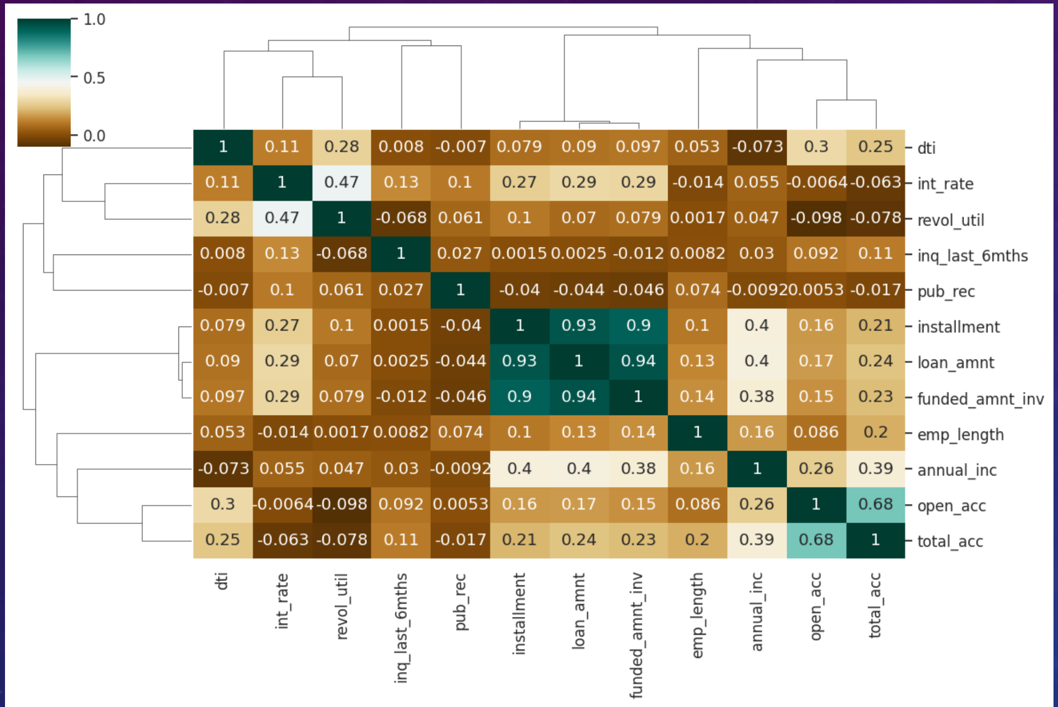
Box plot after removing outlier

OBSERVATION – UNIVARIANT ANALYSIS

Distribution plot of loan amount, funded amount and funded amount Inv. Reveals that data distribution is same for all three variables hence analysis is done only on loan amount.



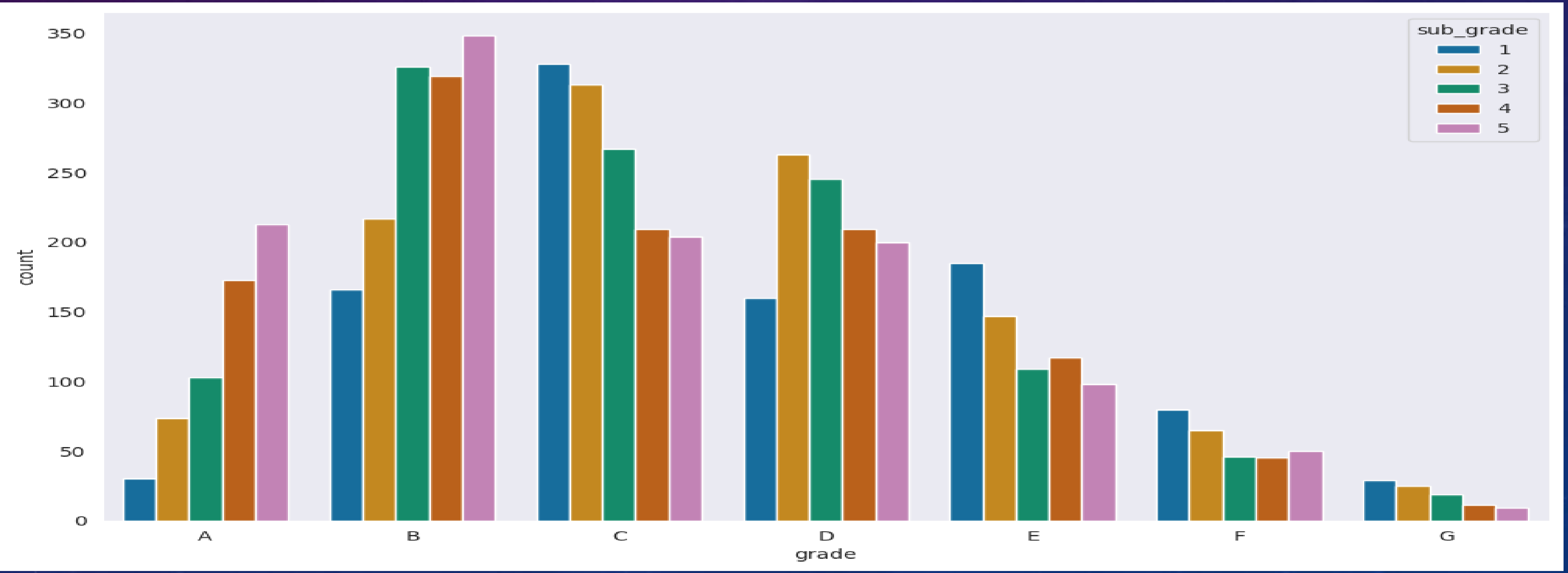
OBSERVATION – CORRELATION MATRIX



- Loan amount, investor amount, funding amount are strongly correlated.
- Annual income with DTI(Debt-to-income ratio) is negatively correlated.
- Debt income ratio is the percentage of a consumer's monthly gross income that goes toward paying debts. That means when annual income is low DTI is high & vice versa.
- Positive correlation between annual income and employment years.That means income increases with work experience

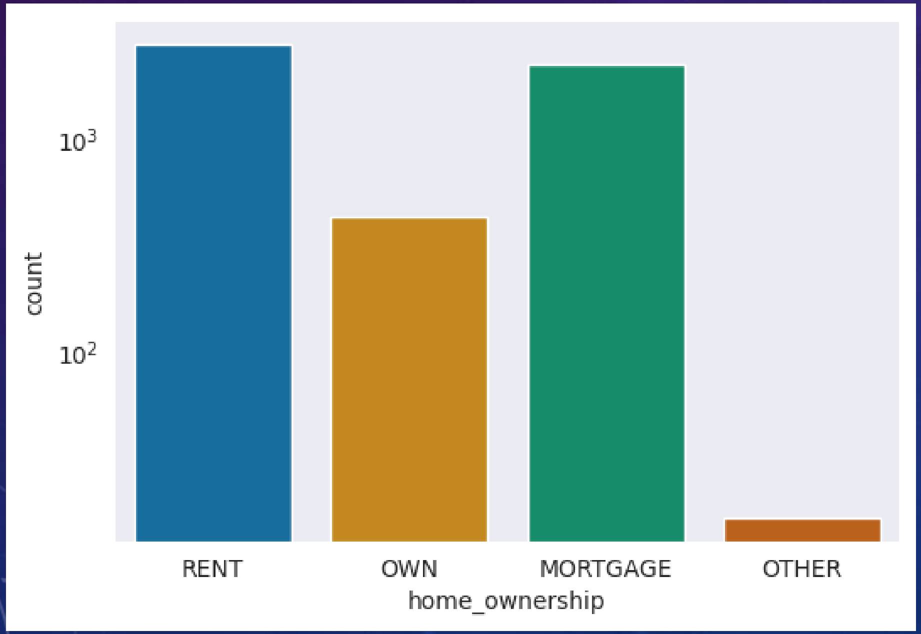
OBSERVATION – UNIVARIANT ANALYSIS

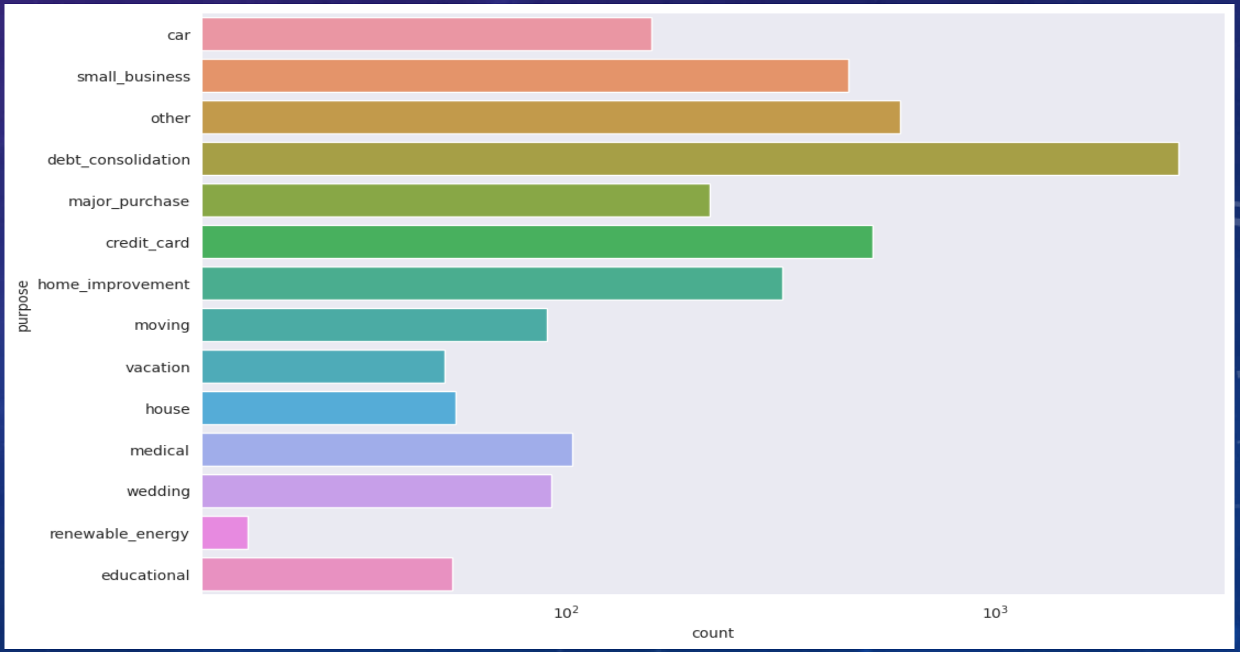
- Most borrowers fall under B grades and sub grade as 5 are likely to be defaulters



OBSERVATION – UNIVARIANT ANALYSIS

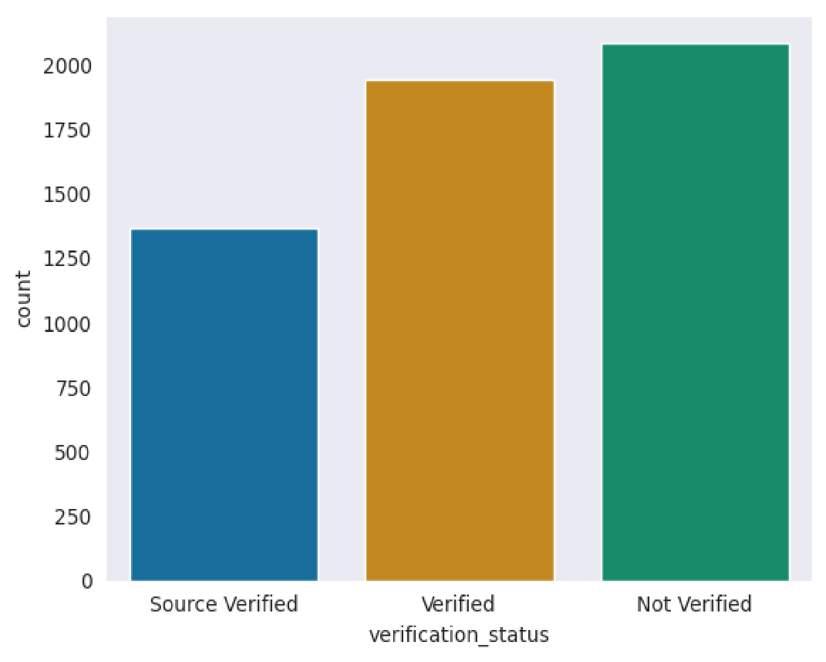
- Most of them living in rented home or mortgaged their home has charged off is high too.
- Most of the loans were taken for the purpose of debt consolidation & paying credit card bill.

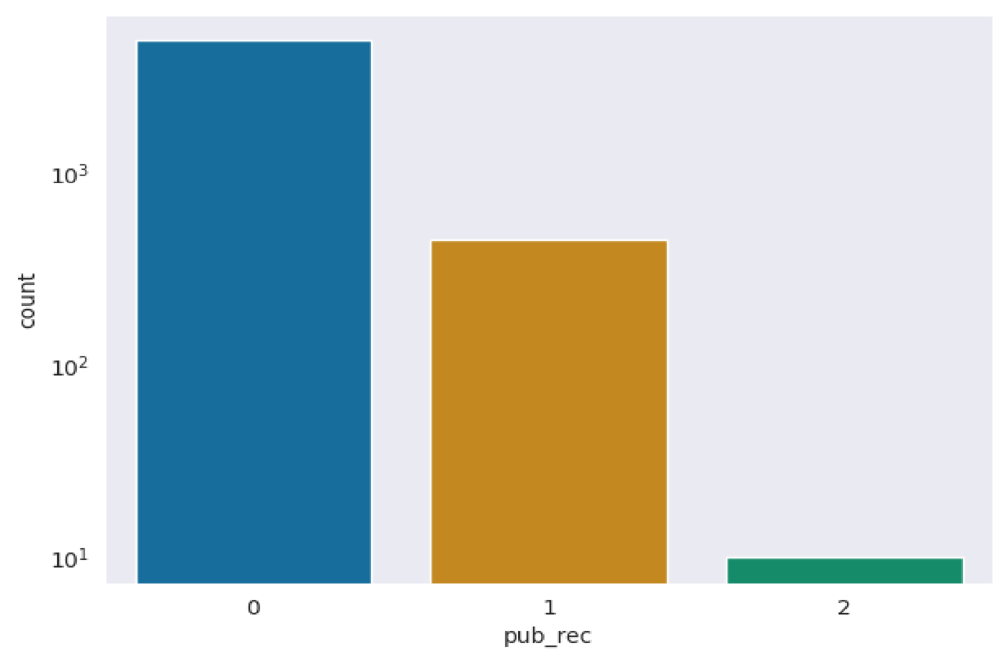




OBSERVATION – UNIVARIANT ANALYSIS

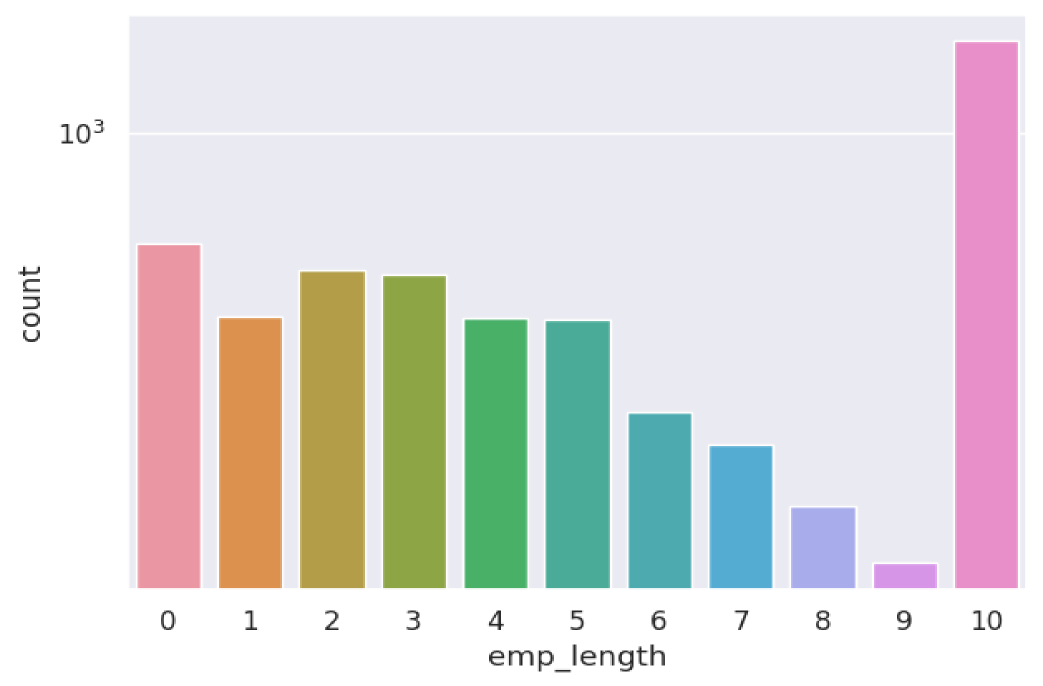
- When loan status is Not verified chance of defaulting loan is more
- Applicants having number of derogatory public records is 0 are more likely to be defaulters

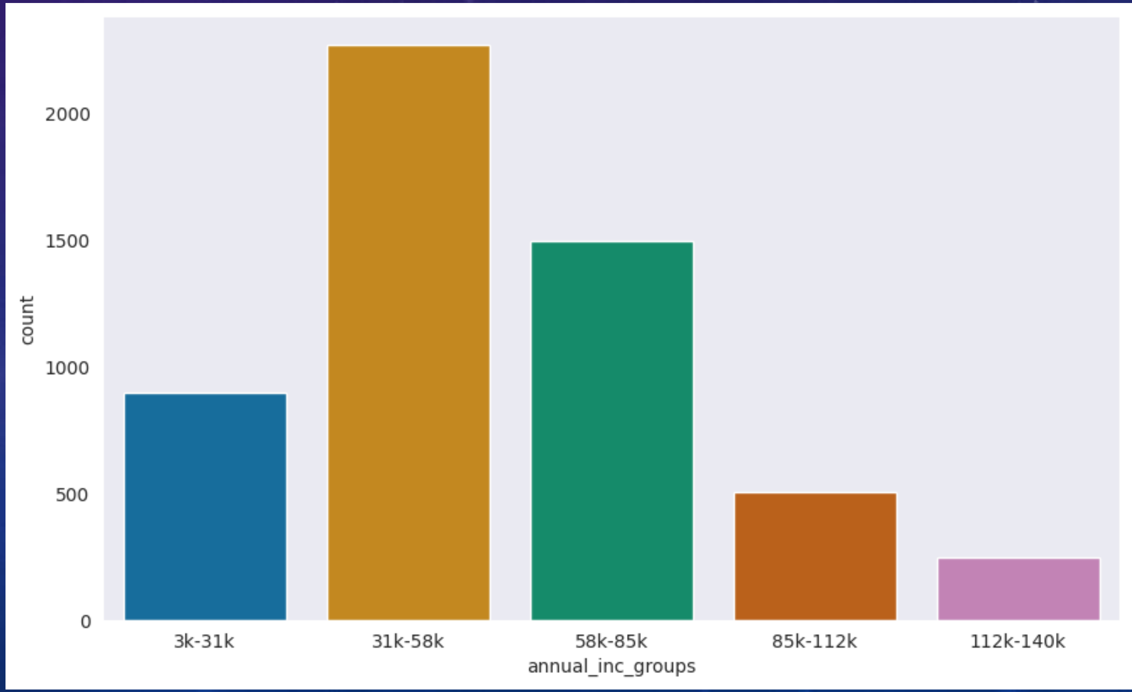




OBSERVATION – UNIVARIANT ANALYSIS

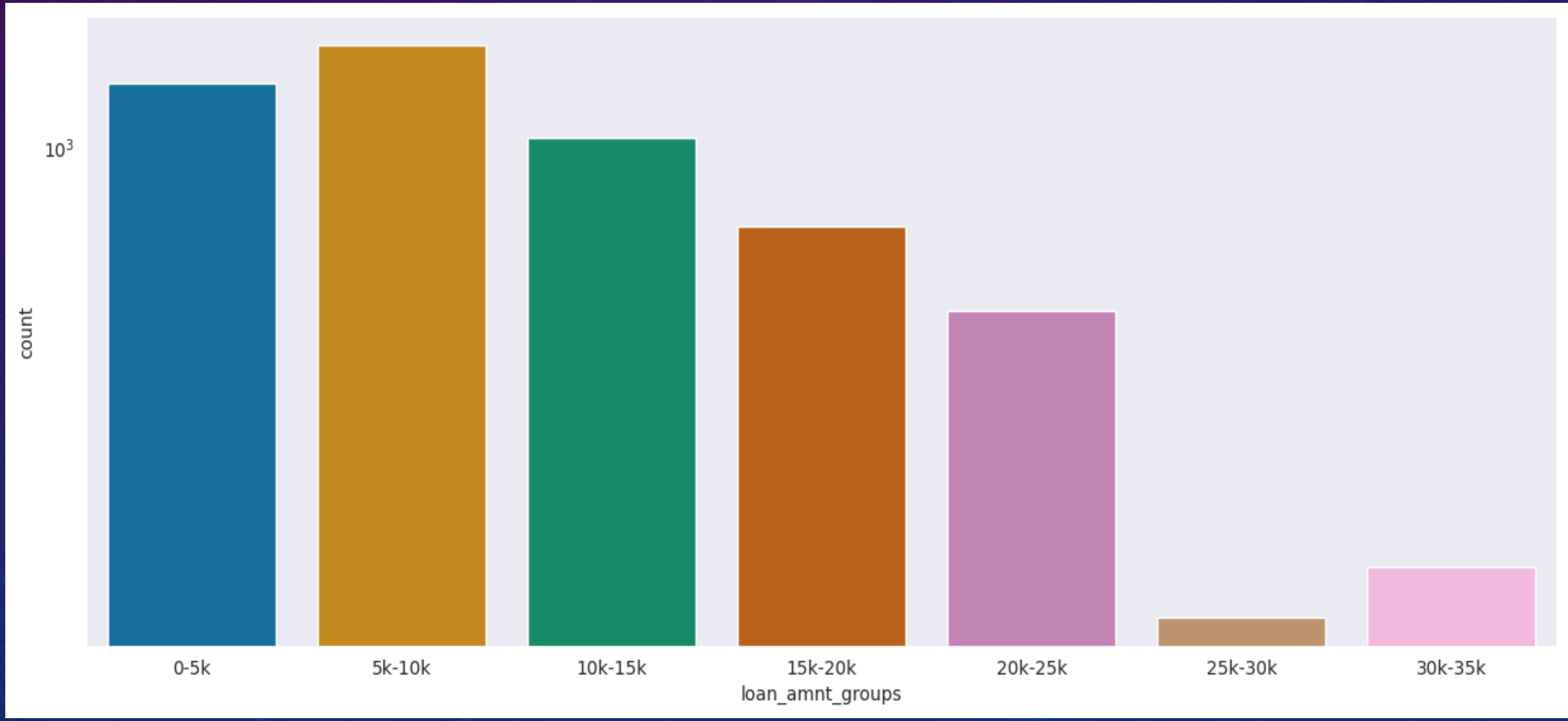
- Employee having more year of experience greater than 10 years are more like to default loan
- Applicants with annual income in range of 31k to 58k are more likely to be defaulters





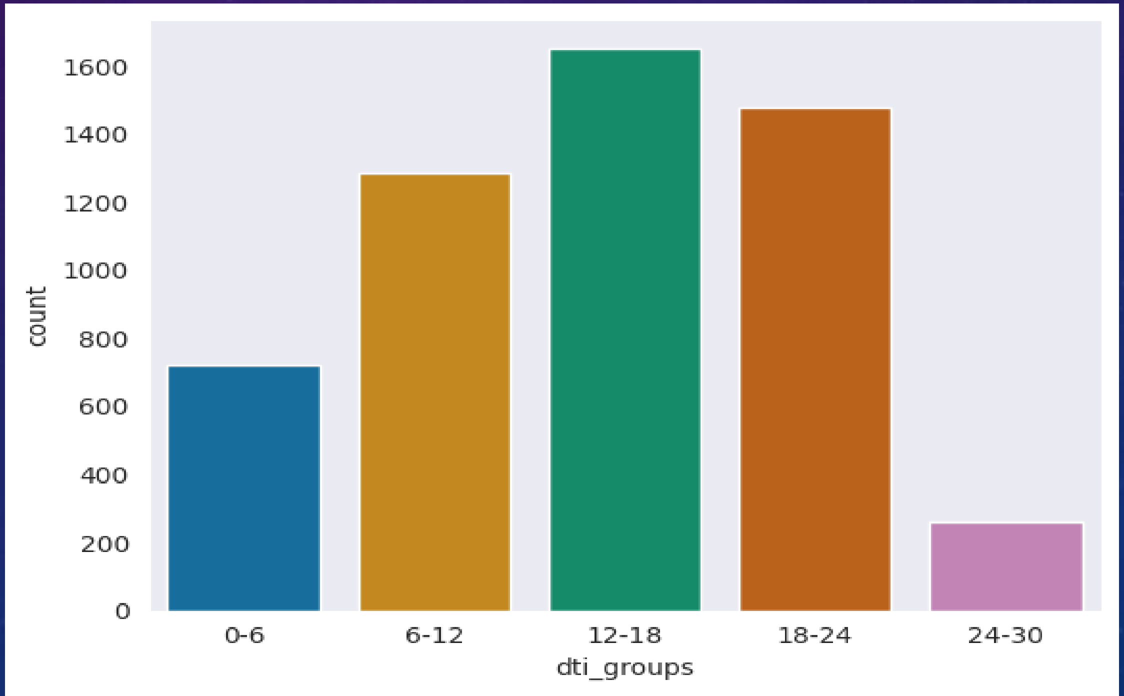
OBSERVATIONS FROM UNIVARIANT ANALYSIS

- Loan amount is between 5429 - 10357 applicants are more likely to be defaulters



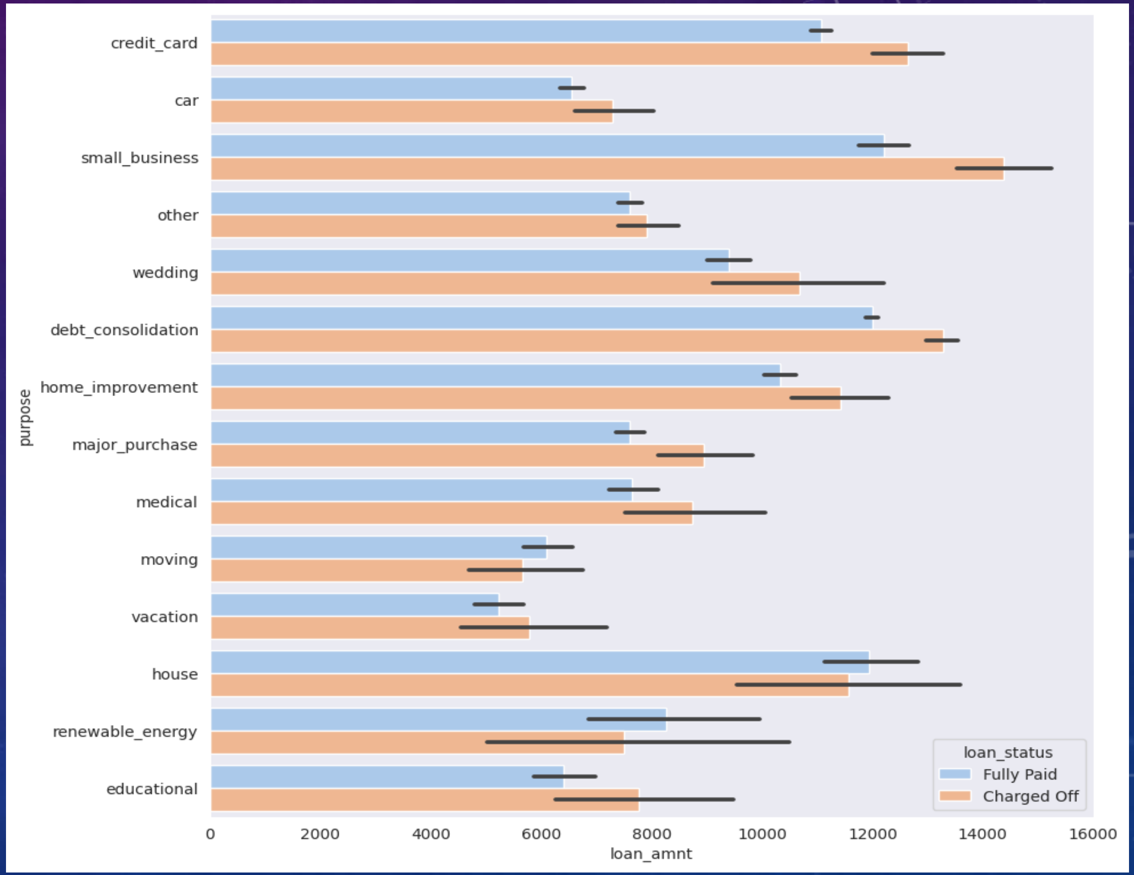
OBSERVATIONS FROM UNIVARIANT ANALYSIS

- DTI is between 12-18 applicants are more likely to be defaulters



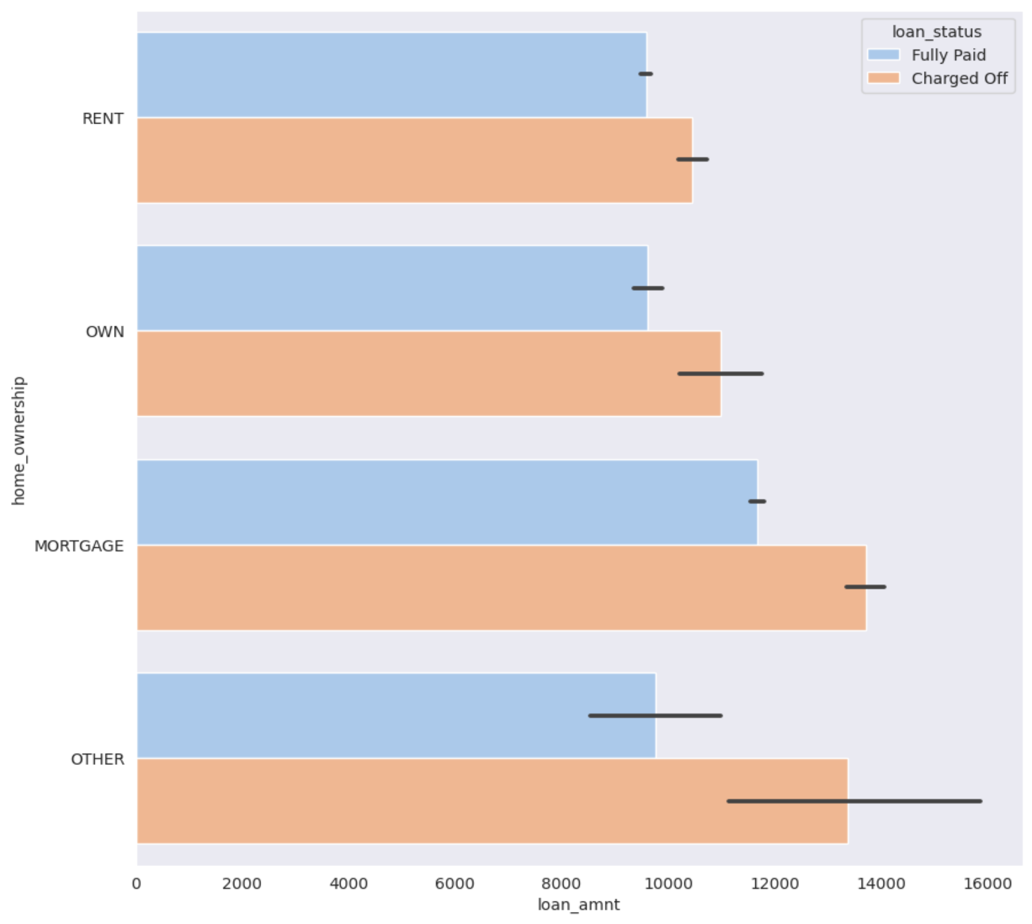
OBSERVATION – BIVARIANT ANALYSIS

- Though the number of loans applied and defaulted are the highest in number for "debt_consolation", the annual income of those who applied isn't the highest.
- Applicants with higher salary mostly applied loans for "home_improvment", "house", "renewable_energy" and "small_businesses"
- Applicants taking loan for 'home improvement' and have income of 60k -70k are more likely to be defaulters



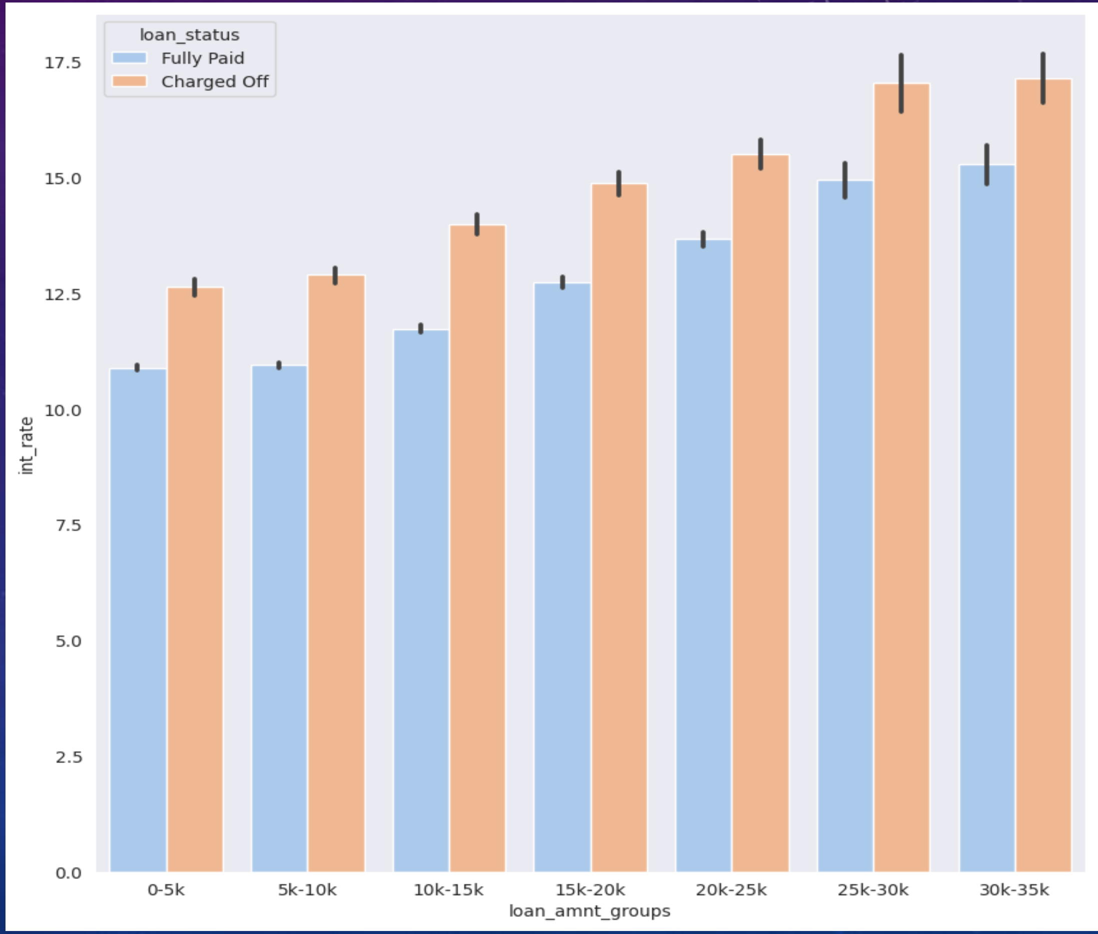
OBSERVATION – BIVARIANT ANALYSIS

- Applicants whose home ownership is 'MORTGAGE and have income of 60-70k are more likely to be defaulters



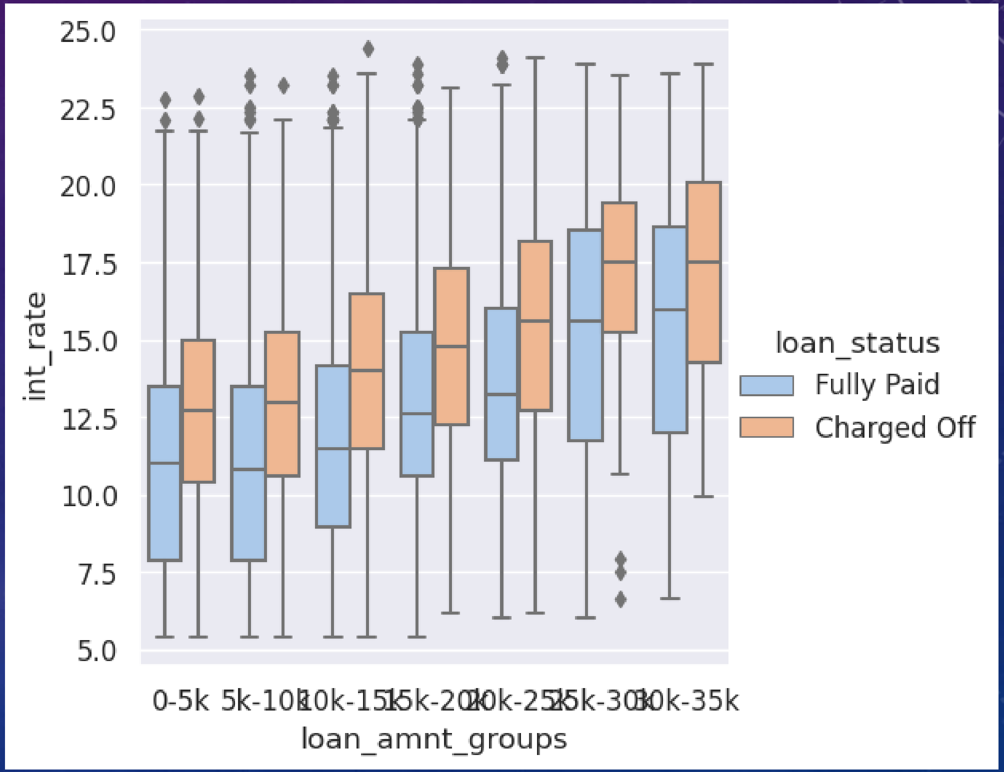
OBSERVATION – BIVARIANT ANALYSIS

- Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k are more likely to be defaulters



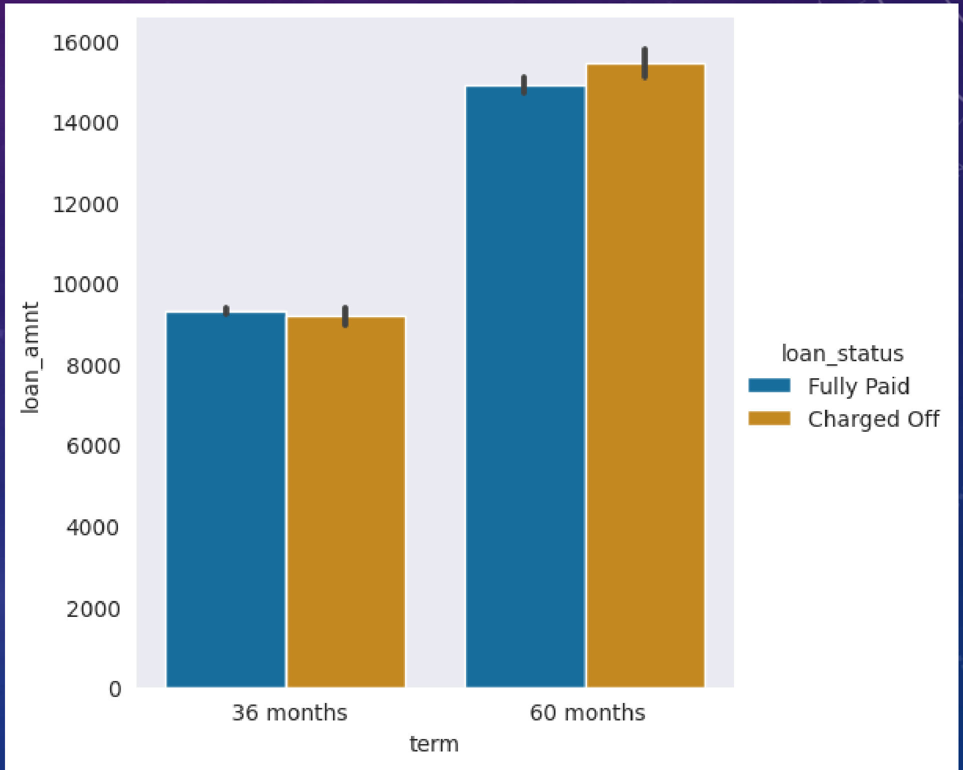
OBSERVATION – BIVARIANT ANALYSIS

- The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan_amount groups.
- This can be a pretty strong driving factor for loan defaulting.



OBSERVATION – BIVARIANT ANALYSIS

- Applicants who applied and defaulted have no significant difference in loan_amounts
- Which means that applicants applying for long term has applied for more loan.



RECOMMENDATIONS

- Major driving factors for to predict chance of defaulters are grades, verification status, annual income, DTI, pub_rec
- Other factors such as borrowers having annual income in the range 50000-100000, Burrowers with working experience 10+ years are more likely to be defaulters.
- Applicants whose home ownership is 'MORTGAGE’ are more likely to be defaulters
- Increased interest rate is a strong factor for deafulting loan amount.

THANKYOU