LENDING CLUB CASE STUDY

CONTRIBUTORS:

ANKIT KUMAR

VISHNU PRASAD V S

CONTENTS

1. PROBLEM STATEMENT

- 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
- The data provided contains the information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

2. DATA UNDERSTANDING

- Getting the initial understanding of the provided loan data such as-
  - A brief summary of Dataframe
    - Total records: 39717
    - Total columns: 111
  - Column types
    - 74 columns are float, 13 int and 24 objects
  - Checking null value counts for the columns.
    - Total 55 columns have all the null values and 3 columns have more than 90% null values

3. DATA CLEANING AND MANIPULATION

3.1) Dropping columns and rows

- Dropping null columns
  - Dropping columns with more than 90% null values.
  - Total 56 columns dropped
- mths_since_last_delinq is having 64% null values. This column is part of customer's loan behavior and can be dropped.
- Dropped columns with unique value count.
  - Total 9 columns are dropped
- Null Rows.
  - There is no rows with all null values.

3.2) Standardising values

- Few columns have some prefix and suffix appended which are not required to trimming those values
  - term, Int_rate, emp_length: trimming these values to use as numeric
- Changed issue_dt to date and added derived columns issue_month and issue_year

3.3) Filter Data

- Excluded records with loan_status "Current"- Because, with this status we are not yet sure whether the borrower is going to default or not

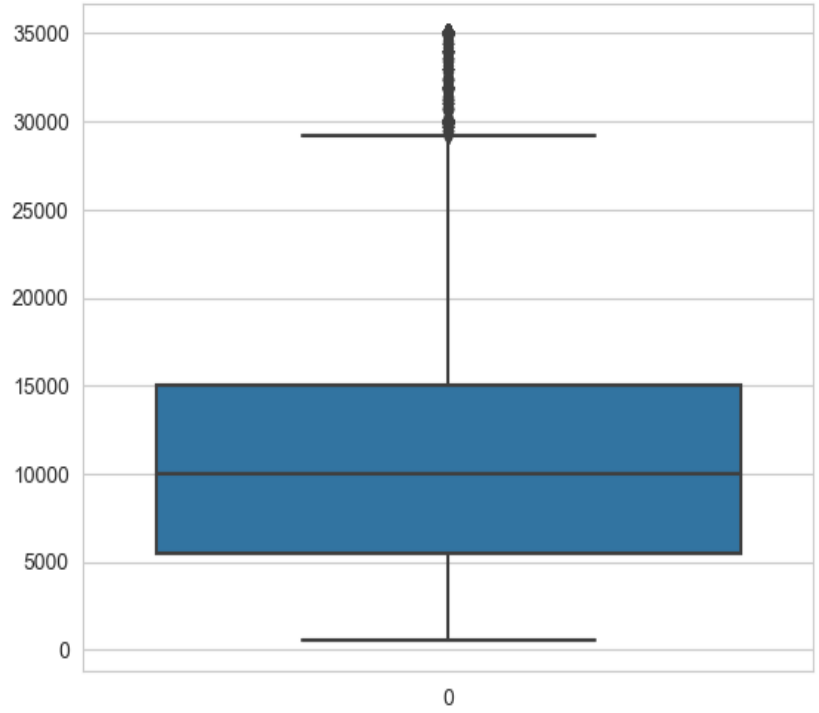
3.4) Exclude irrelevant columns

- The columns that can be dropped because they don’t contribute in deciding the loan default or not:
  - Id, member_id, funded_amnt, funded_amnt_inv, emp_title, title, url, desc, zip_code, sub_grade, verification_status, open_acc
- Below attributes are affected by whether the loan is approved or not. Hence these cannot be relied on for analysis:
  - revol_bal, revol_util, out_prncp, out_prncp_inv, total_acc, recoveries, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, total_rec_late_fee, collection_recovery_fee, last_pymnt_d, last_pymnt_amnt, last_credit_pull_d, earliest_credit_line, delinq_2yrs, delinq_2yrs, inq_last_6mths, open_acc, pub_rec_bankruptcies, mths_since_last_delinq

3.5) Handling Outliers

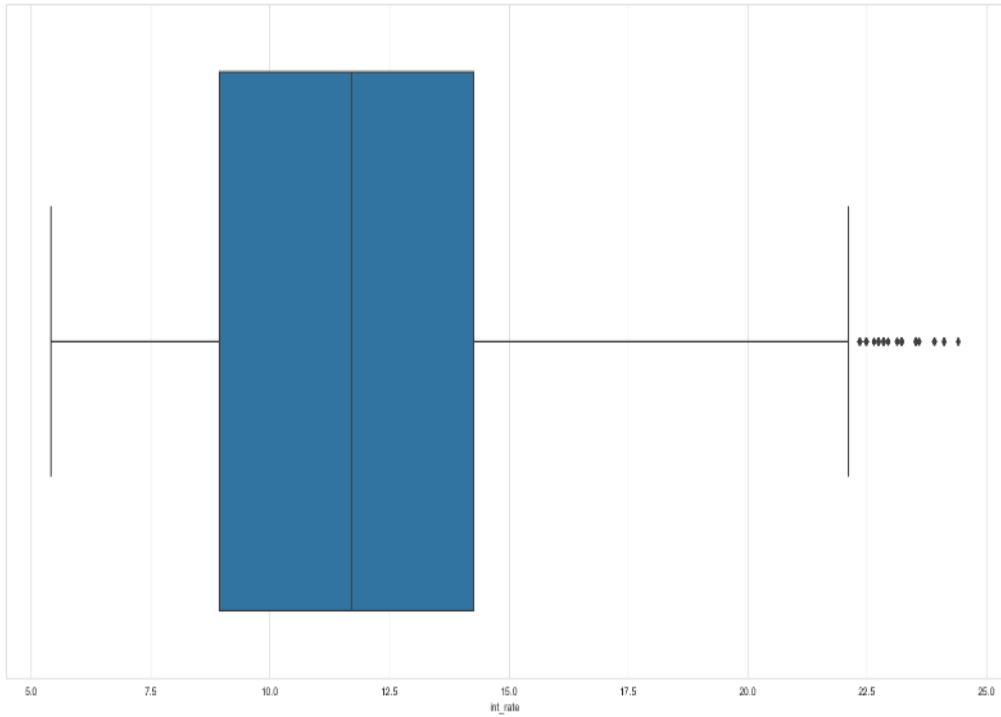
We found some outliers in loan_amount, int_rate, annual_inc and installment and removed those

```
In [32]: sns.boxplot(loan_data['loan_amnt'])
Out[32]: <Axes: >
```



Loan_amount

```
In [35]: plt.figure(figsize=(20, 10 ))
         sns.boxplot(data=loan_data, x='int_rate')
         plt.show()
```

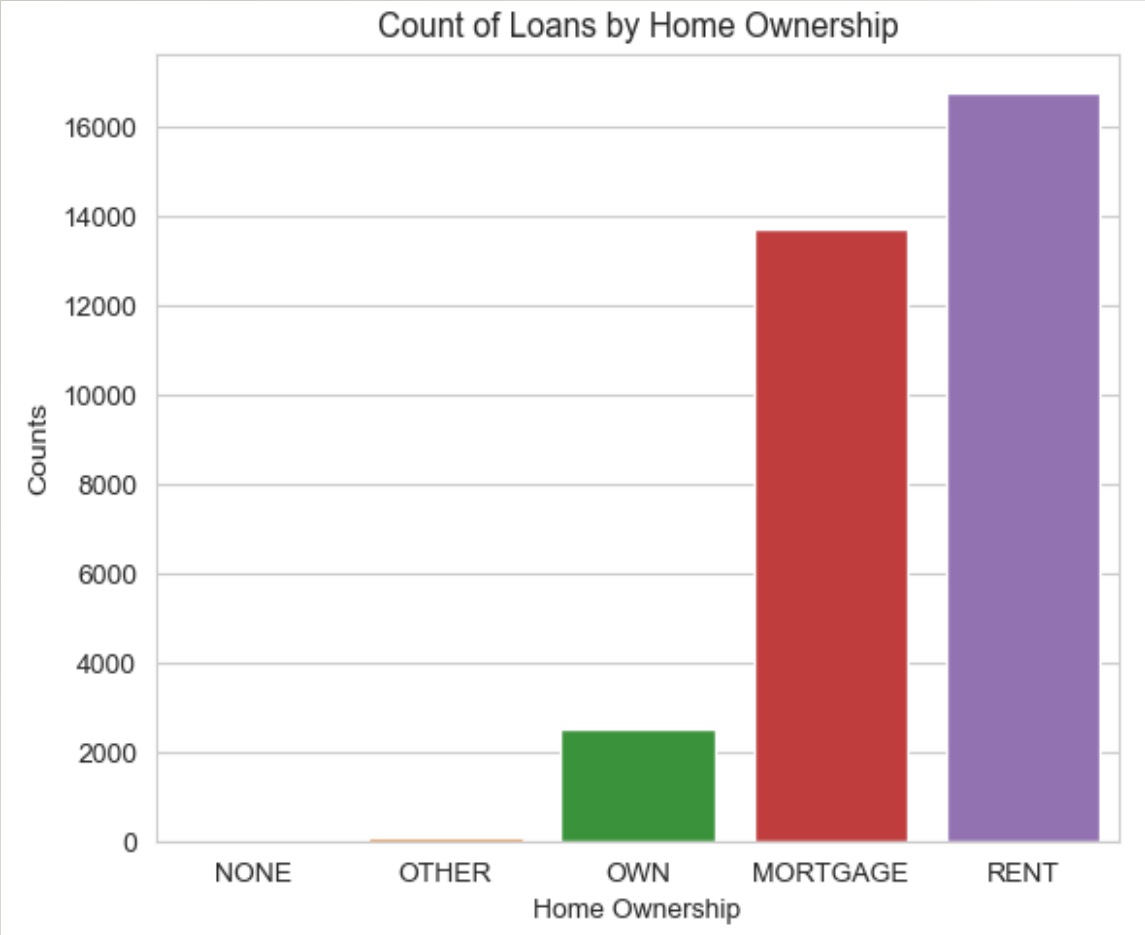


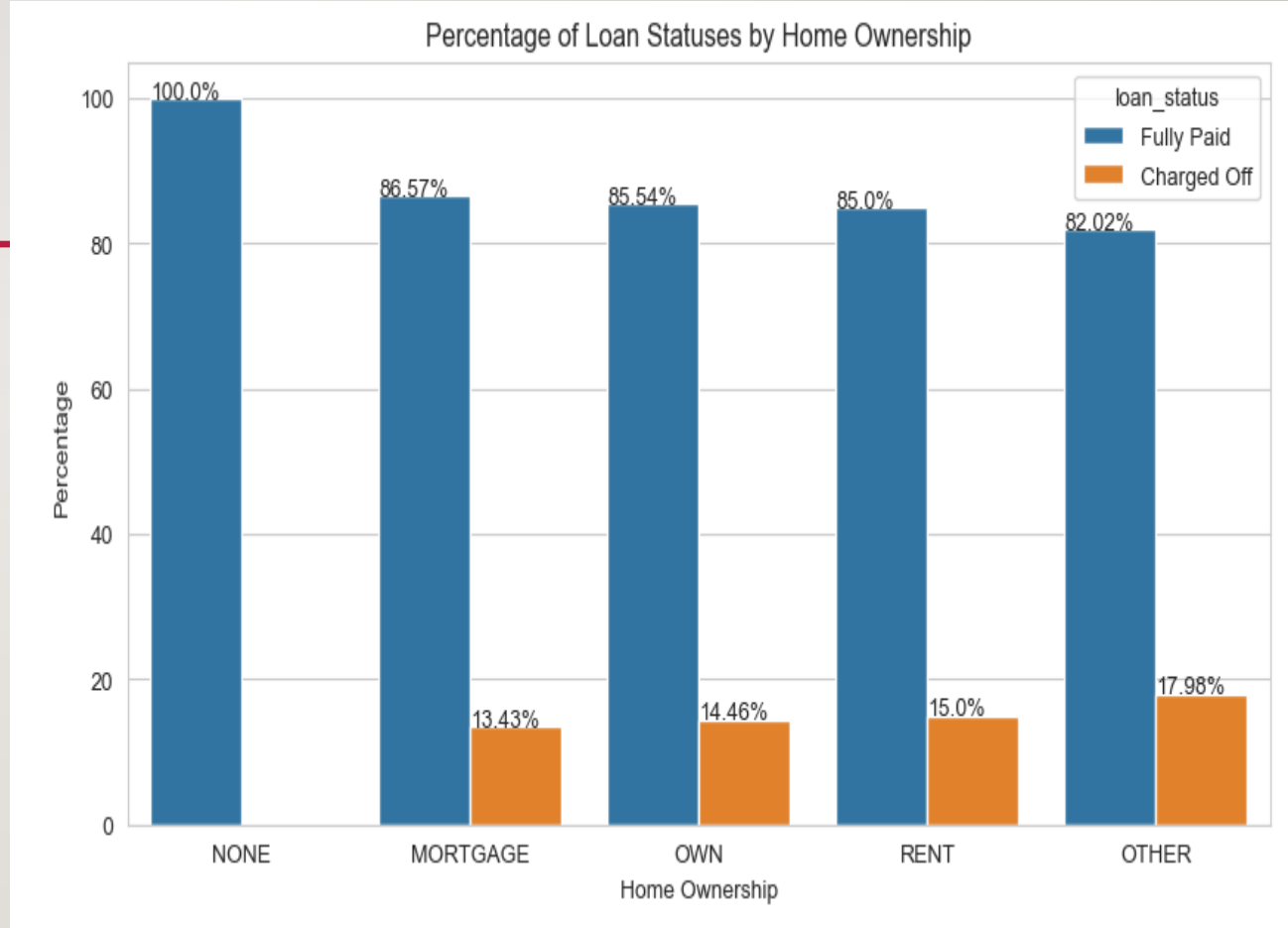
int_rate (Interest Rate)

4. UNIVARIATE ANALYSIS

- We have plotted some countplots and extracted useful information from the below categorical variables.
    - home_ownership, purpose, addr_state, term_in_months, grade, emp_length, issue_year, loan_amnt, int_rate, annual_inc, pub_rec, Installment and installment_bkt

Univariate analysis of home_ownership

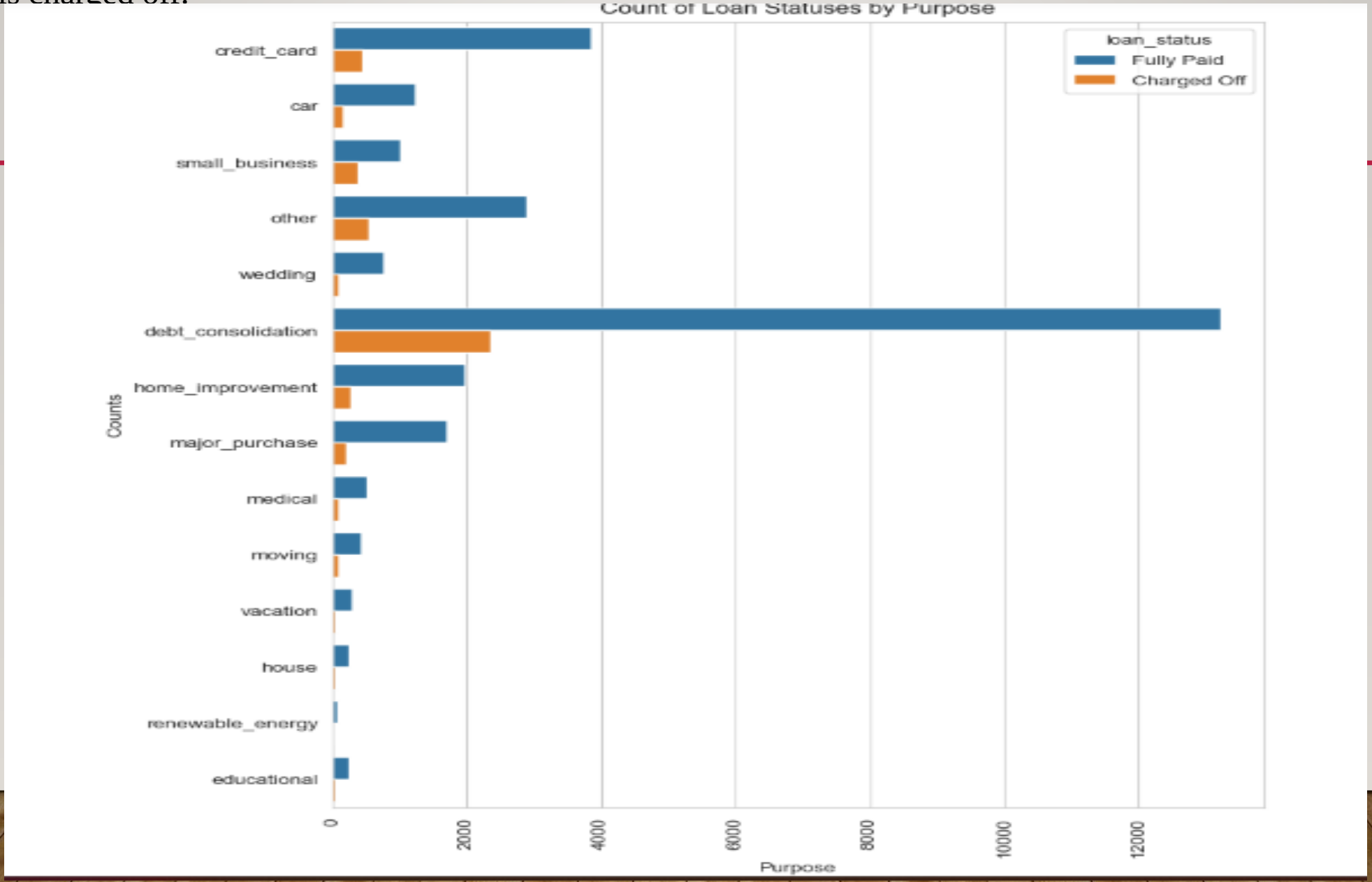




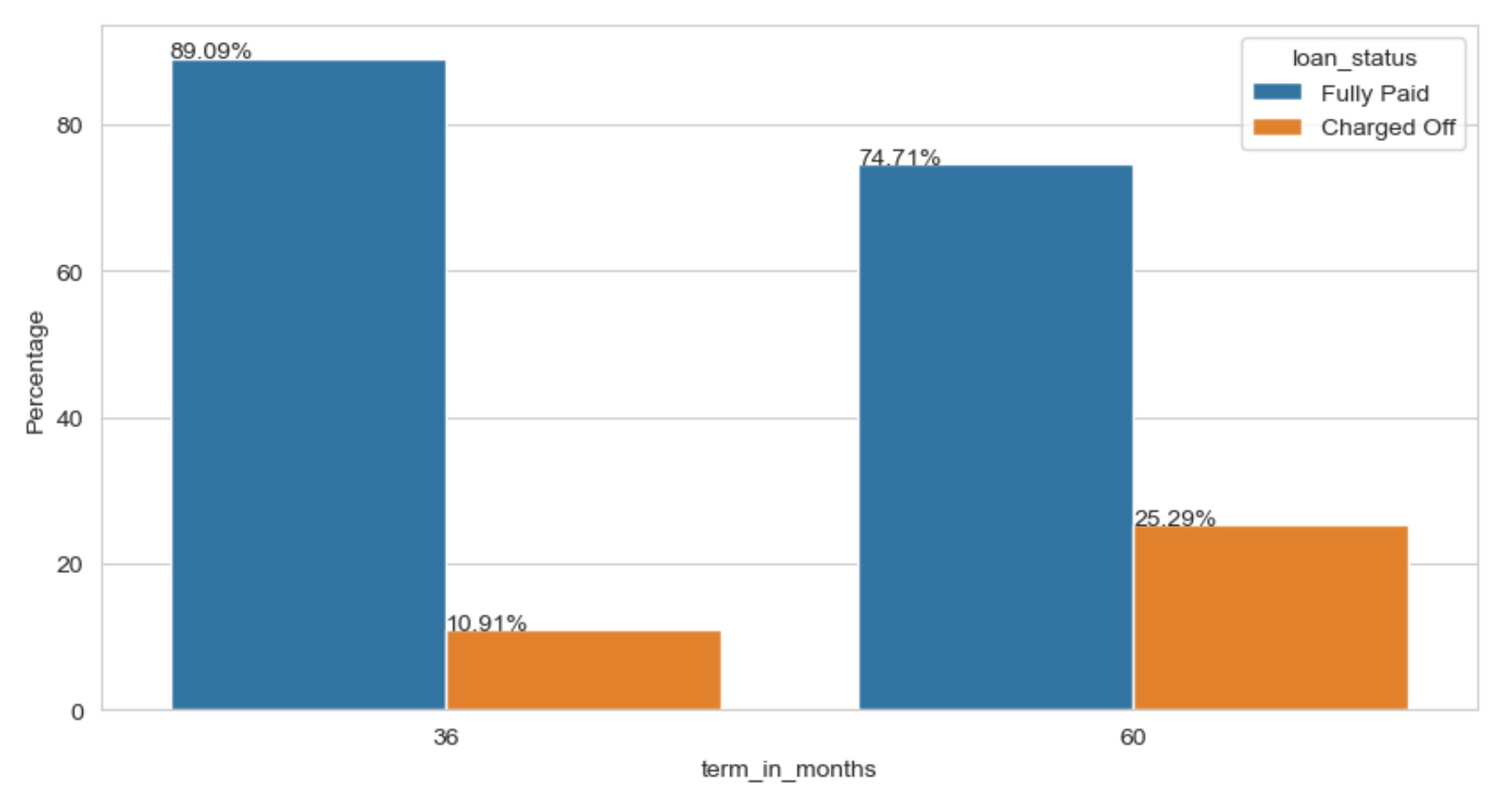
Observation: Majority of borrowers RENT a home as per the above chart, followed by MORTGAGE and OWN. 18% of loans borrowed by the OTHER home_owneres are Charged Off followed by 15% for those who RENT.

Univariate analysis of purpose

observation: majority of the loans are taken for the purpose of debt_consolidation. 27% of loans taken for the purpose of small_business is charged off.

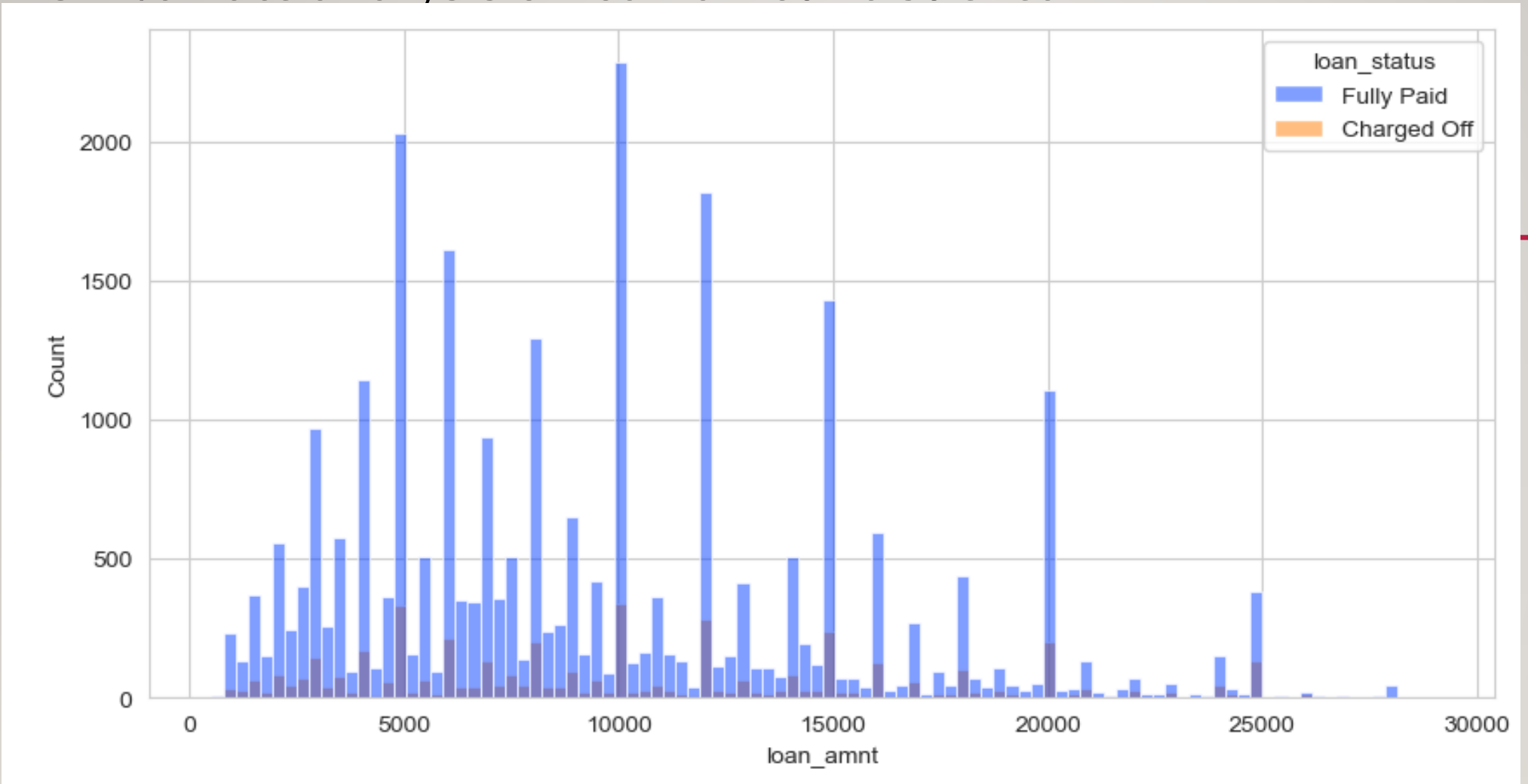


Univariate analysis of terms_in_months



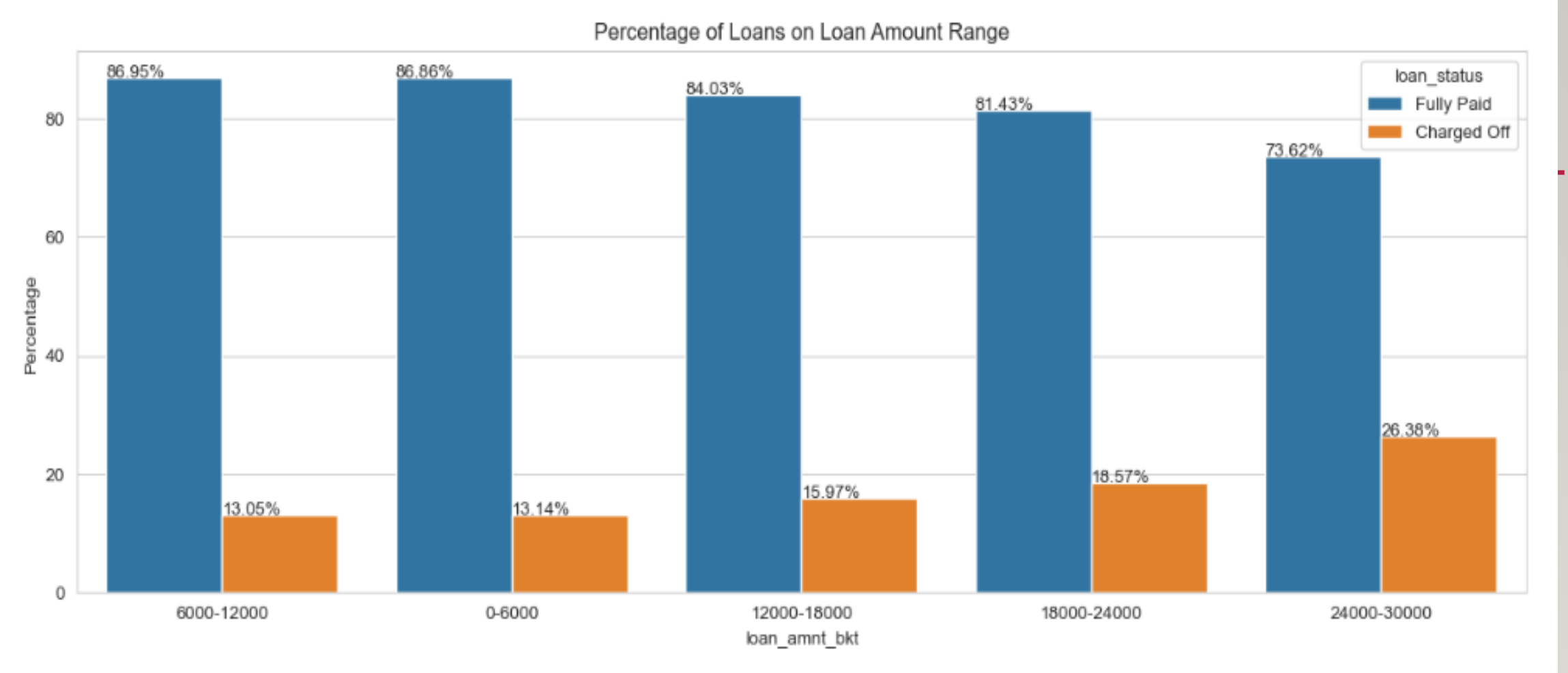
Observation: 25.3% of loans taken for 60 months term is defaulted.

Univariate analysis of loan amount bucket



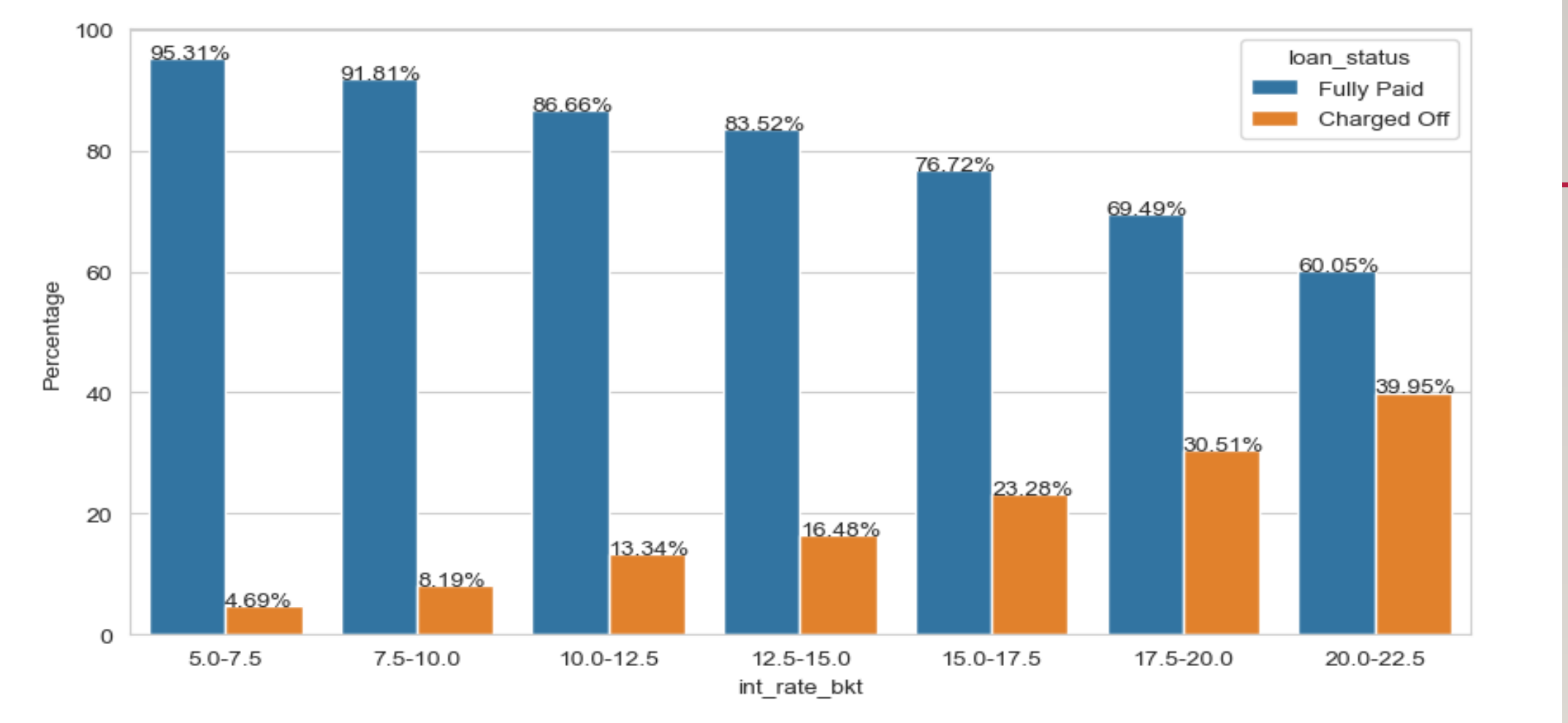
Observation: Most number of loans are granted with loan_amnt of 10000.

Univariate analysis of loan amount bucket



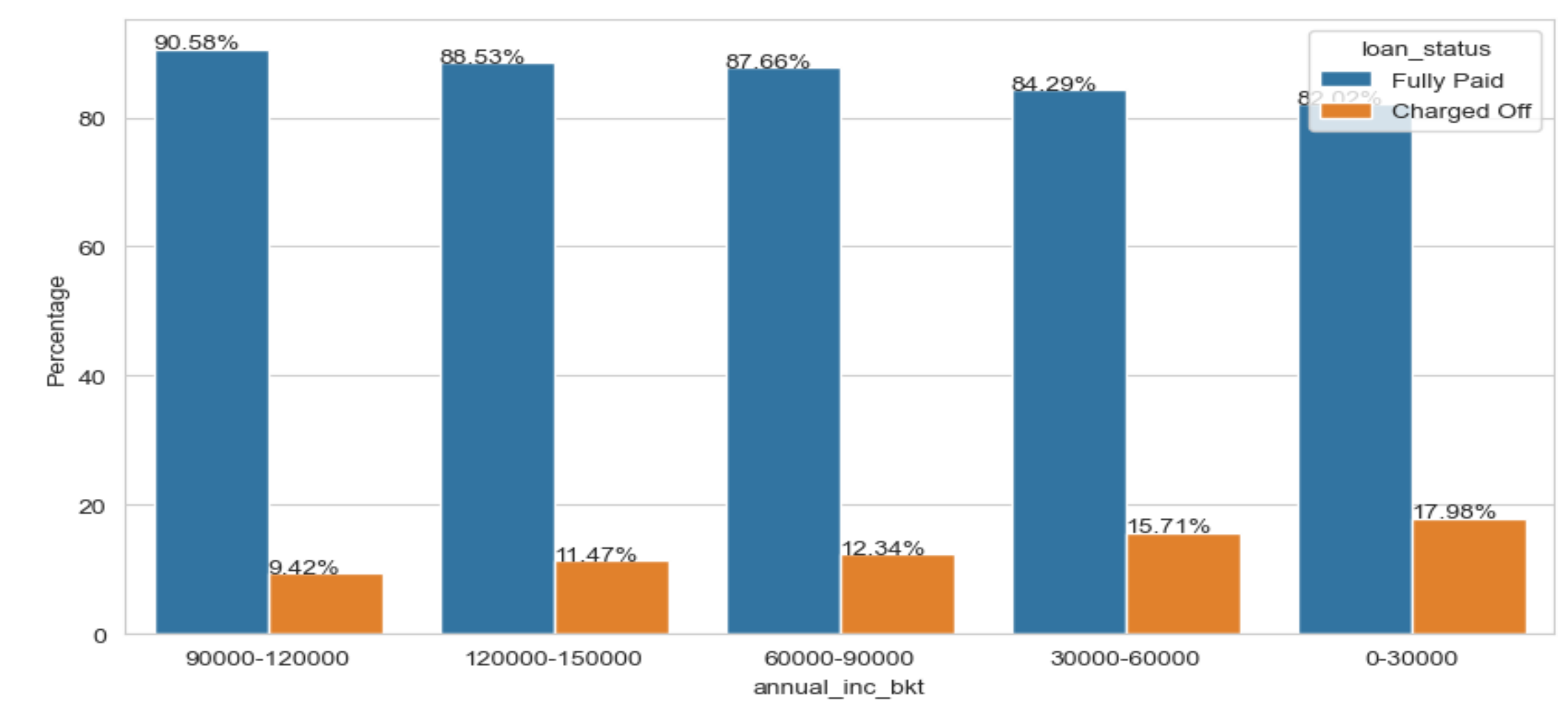
Observation: As the loan amount increases, the chances of defaulting increases. 24-30K have a chance of 26.36% of being Charged Off.

Univariate analysis of interest rate bucket



Observation: As interest rate increases, the chances of defaulting increases. 20-22.5% interest rate have approximately 40% chance of defaulting.

Univariate analysis of annual income bucket



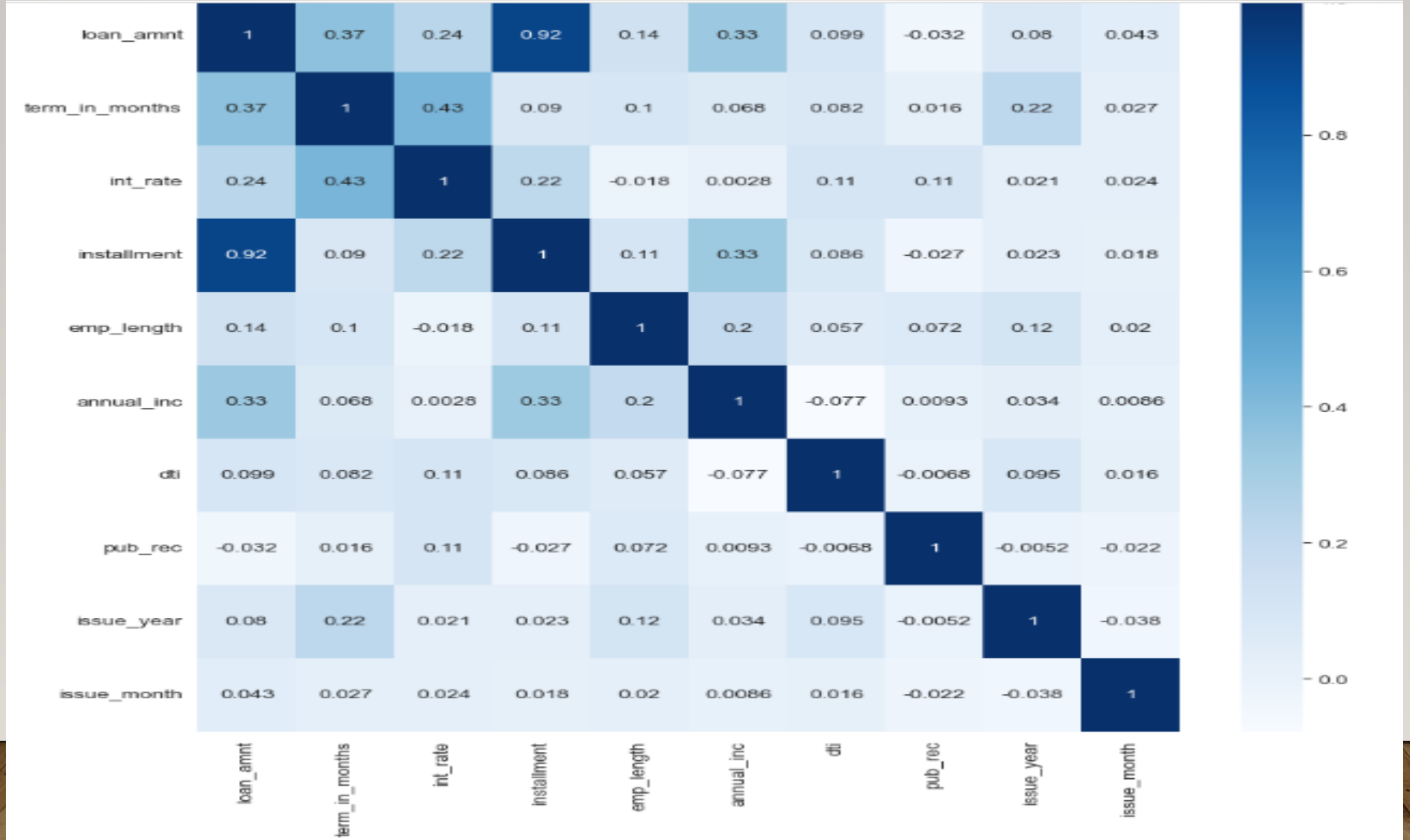
Observation: As annual income decreases the chance of defaulting increases. 0-30K income group have 18% loans Charged Off.

We have extended univaraiate analysis further with other useful attributes and found this observations

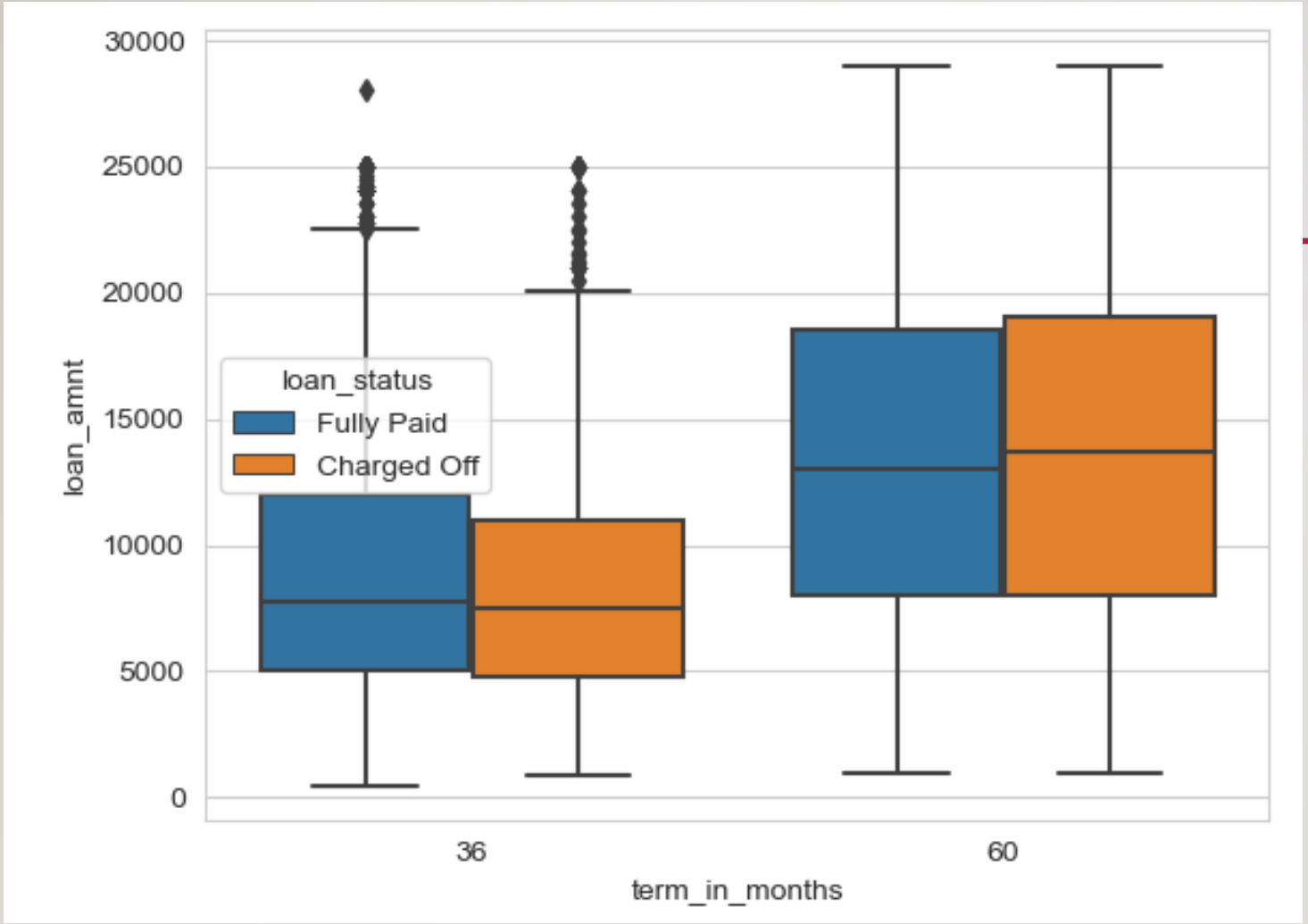
- California takes most number of loans
- 20% of loans taken in Nevada is defaulted.
- As grade goes from A to G, the risk of defaulting increases from 5.99% - 35.33%
- Number of loans granted has increased from 2007 to 2011
- Loans taken in 2007 have 16.26% for defaulting.
- December had the most number of loans taken and also the highest rate of loan being Charged off.
- As number of derogatory public records increases, the risk of defaulting increases. But sufficient data is not available for more than 2. For 2 pub_rec, 22.7% of loans are Charged Off.
- As the installment amount increases, the chances of Charging off increases. Installment amount in the range 700-800 have 17.89% for Charging Off.

5. BIVARIATE ANALYSIS

5.1) Correlation Matrix

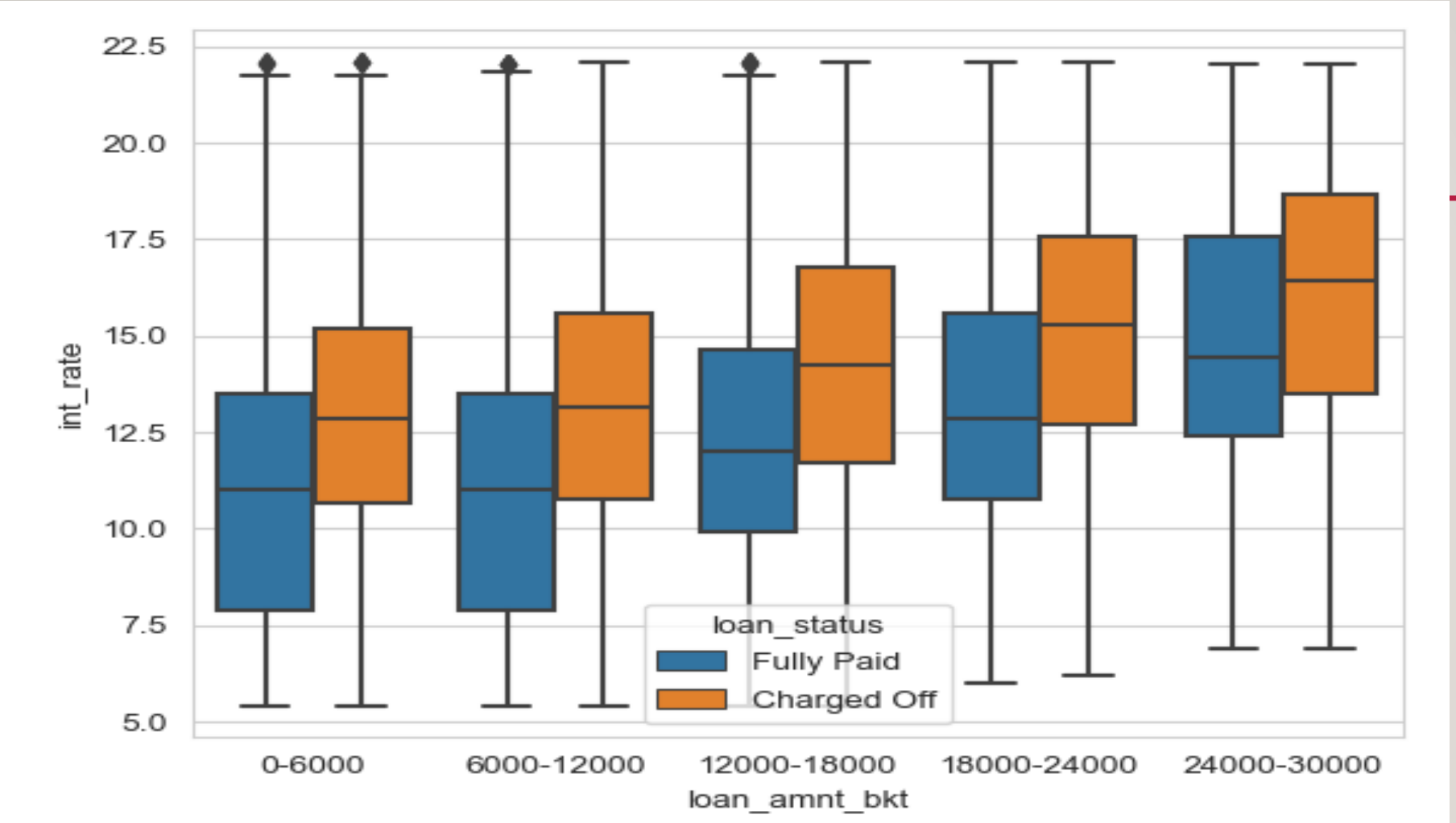


Loan amount vs terms in months



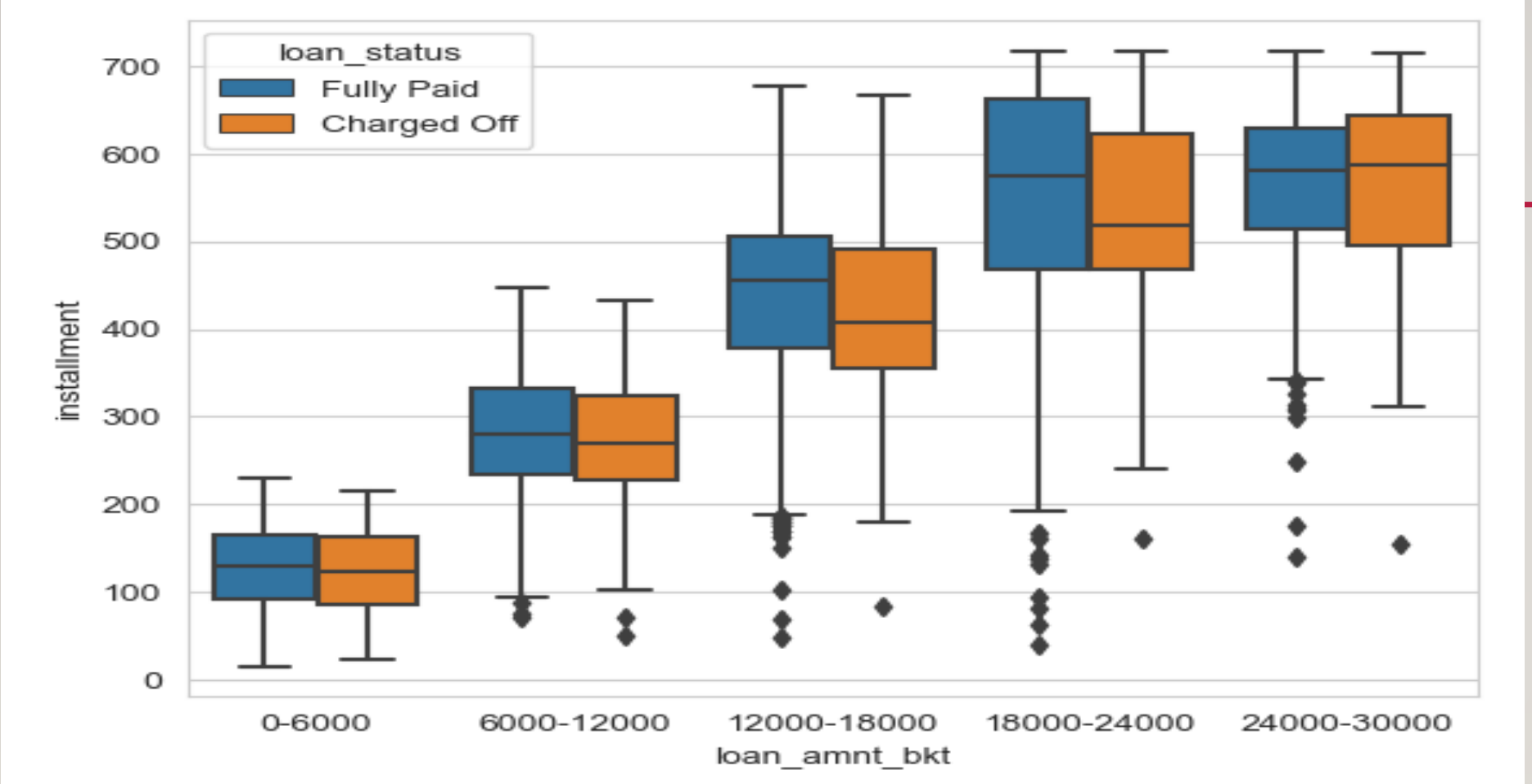
Observation: Higher loan amounts(12K+) for longer duration(60 months) have higher risk of Charging off.

Loan amount vs interest rate



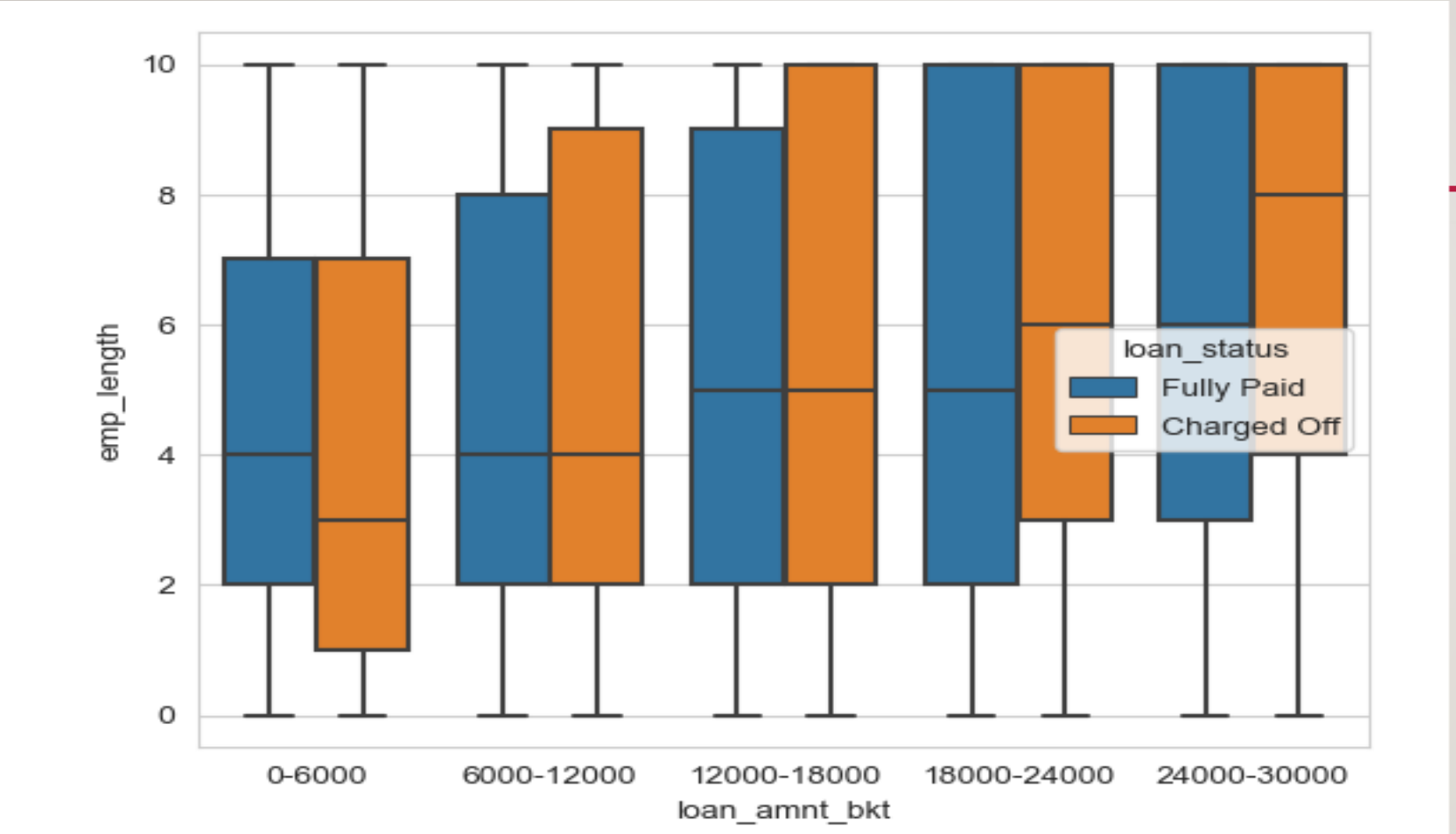
Observation: Interest rate increases as loan amounts increases. Higher interest rate and loan amounts have higher risk of Charging off.

Loan amount vs installments



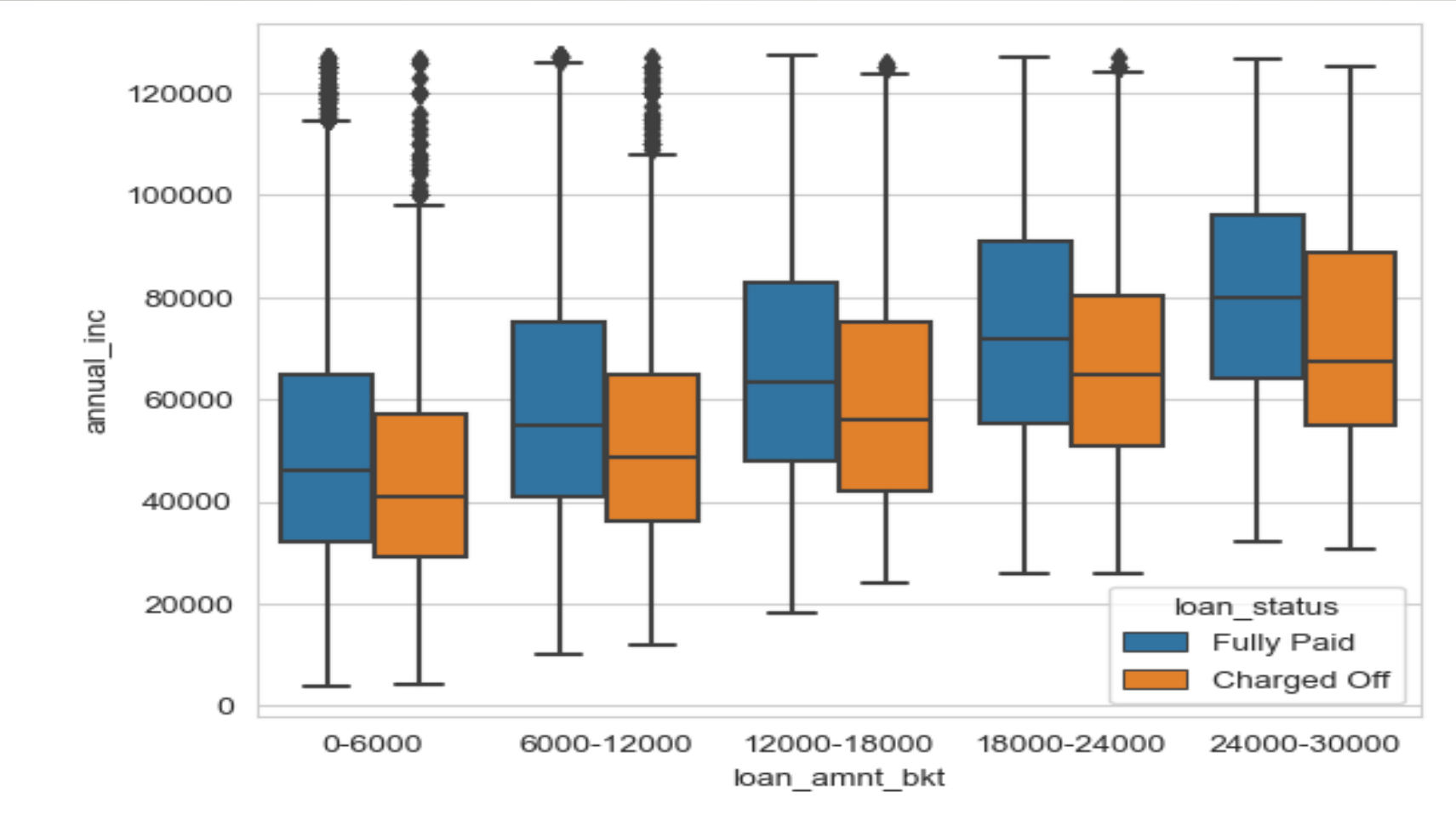
Observation: Installment amount increases as loan amount increases. Higher installment amount and loan amounts have higher risk of Charging Off.

Loan amount vs employment length



Observation: Borrowers with higher employment length opt for higher loan amounts. Higher loan amounts have higher risk of Charging off.

Loan amount vs annual income



Observation: Borrowers who have lower annual income take lower loan amounts. Within each range of loan amounts, those with lower annual income have higher chance of Charging Off.

We have also done bivariate analysis on below fields and found the below observations-

- Loan Amount vs Purpose
- Term In Month vs Interest Rate
- Term In Months vs Annual Income
- Term In Months vs Purpose
- Interest Rate vs Installment
- Interest Rate vs Home Ownership
- Interest Rate vs Purpose
- Interest Rate vs Issue Year
- Employment Length vs Home Ownership

Observations

- OTHER home owners who take high loan amounts have higher risk of Charging Off.
- Borrowers who take high loan amounts for the purpose of credit_card, small_business or debt_consolidation defaults.
- As interest rate increases chances of Charging off increases. Interest rate for longer term loans is higher than that of shorter term thereby making longer term loans to have higher chance of Charging off.
- For each purpose of the loan, those who opted for longer term have higher chance of defaulting.
- Installment increases as interest rate increases. Increase in installment/interest result in higher risk of defaulting.
- House loans provided with high interest rate have higher chance of being Charged Off's
- From 2007 to 2011 there is an increase in interest rate with year. Based on the increase in interest rate, the number of defaults also increased.

6. Conclusion/recommendations

After analysing the data we found below most important points and the most important attributes which must be considered before giving loan.

1. Loan Amount
2. interest rate
3. Annual income
4. No of installments
5. purpose

THANK YOU