Lending Club Case Study

Introduction

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

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Objective:

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.

Below sequence is followed for this case study

1. Data Understanding
2. Data Cleaning
3. Univariate Data Analysis
4. Bivariate Data Analysis
5. Recommendation

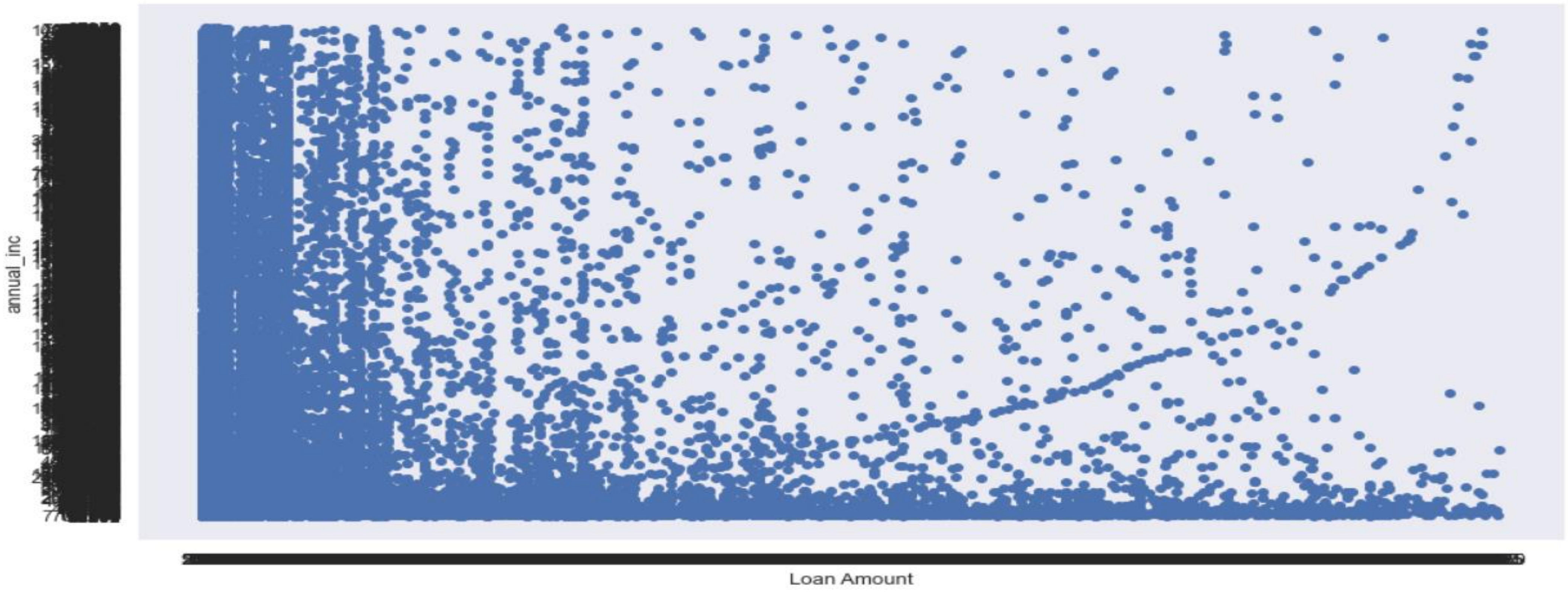
Data Understanding

- We should have important columns with are: Loan_amt, Term, Interest_rate, Grade, Annual Income, Purpose of loan, DTI(debt to income), Emp_Length, Load_Date, Home_Ownership, Verification_Status
- We have customer behavior variables in our data set which are considered before loan in accepted.
- Lot of columns have all null and all zero values. May be these columns were not mandatory columns
- Applicant having greater annual Income and owning a house should be able to pay the loan
- Grade and sub grade should play important role in analysis as they are given on basis of applicants last loans,
- Few columns which are in string format if converted to numeric form will be helpful in our analysis.

Data Cleaning

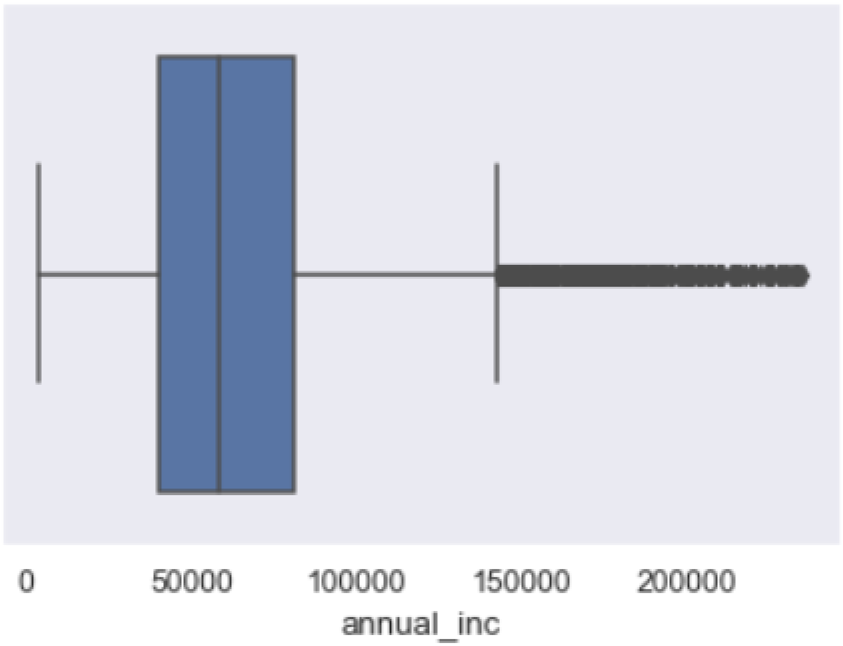
- Remove columns which have high percentage of missing values.
- Remove all the columns with null values
- Remove all the columns with all zero values
- Remove all columns which are customer behavior variables (customer behavior variables should be considered before the loan in accepted)
- Few columns have strings values (eg., employment length 2+ years can changed to 2 and interest rate have values like 2.5% can be changed to 2.5)
- Remove Outliers using annual income

Data Cleaning



In the above scatter plot of Annual Income vs Loan Amount , we can see that density decreases as the annual income increase

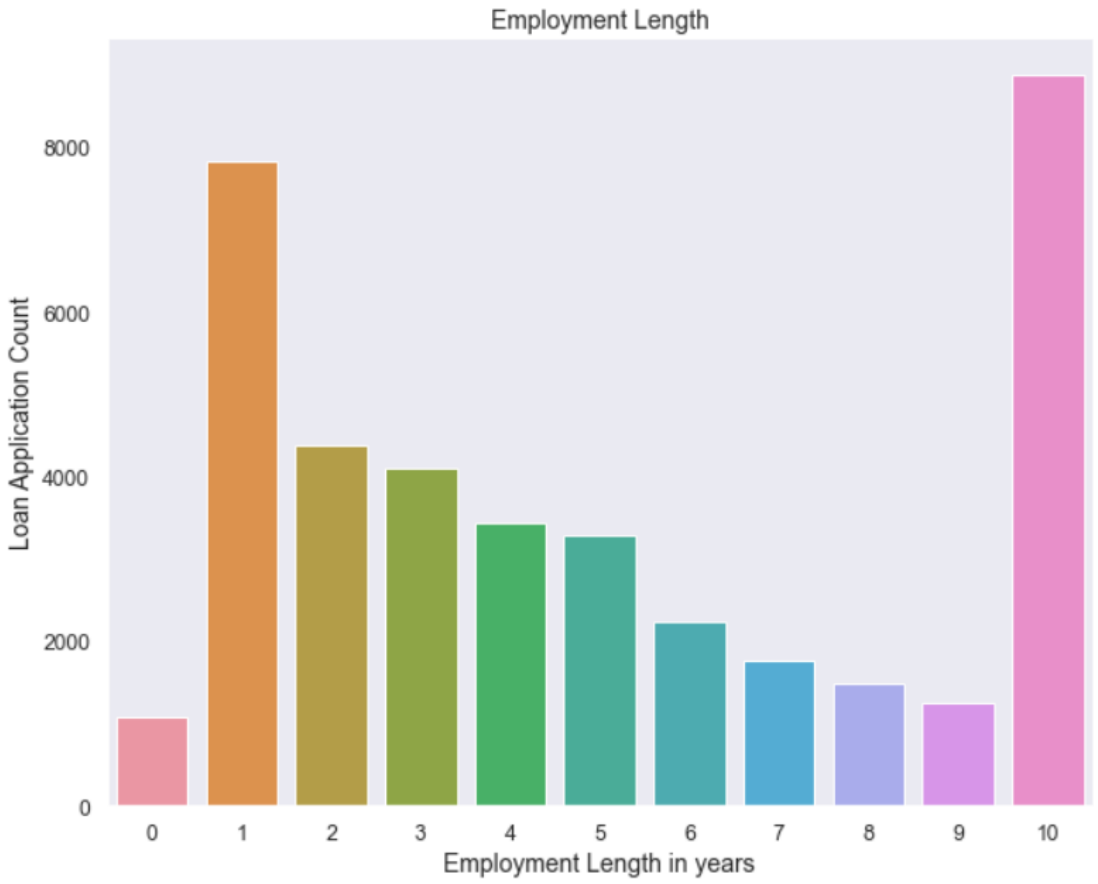
Removing Outliers



Box plot for annual income

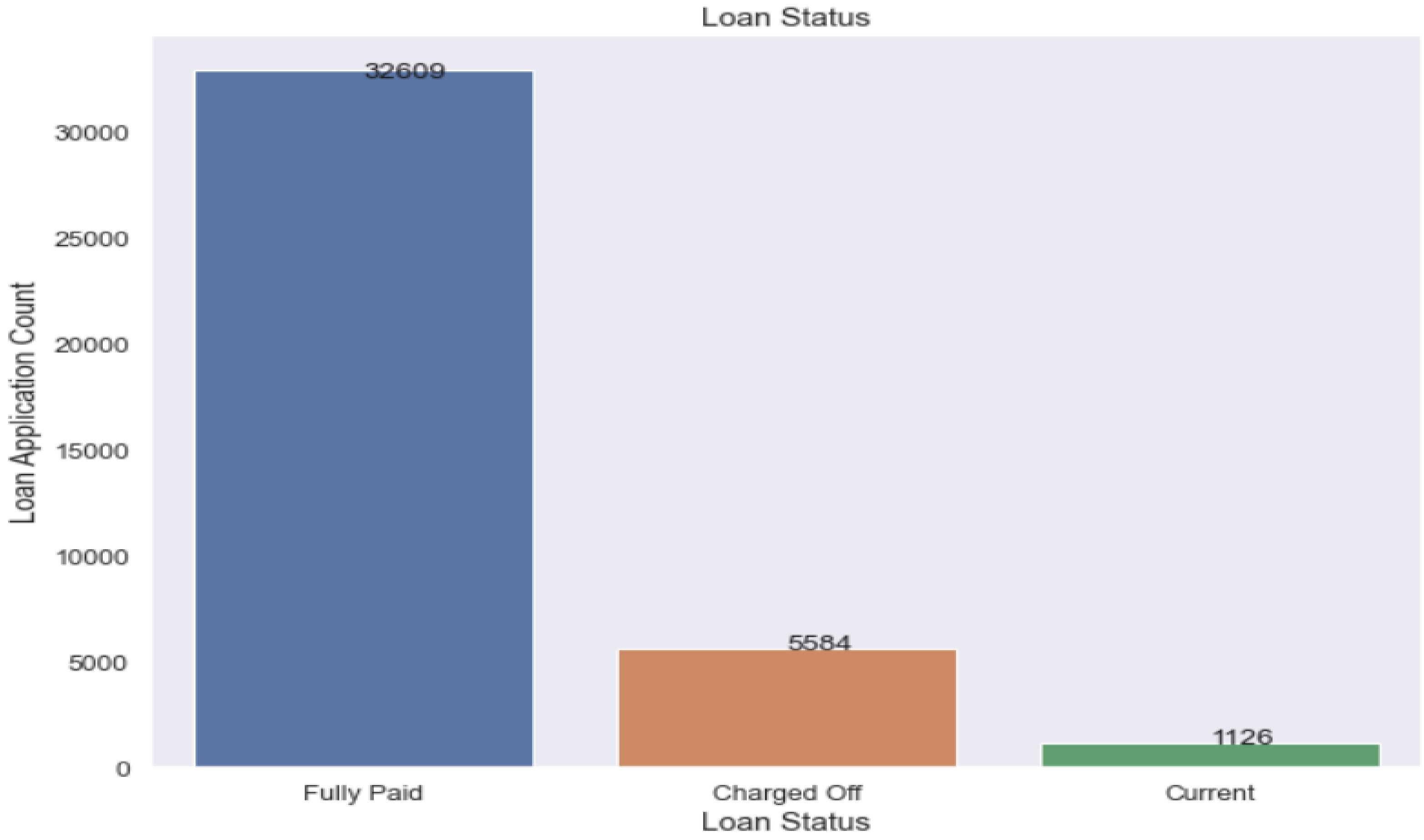
Max annual income value was 6000000 earlier. After removing outliers we get better representation

Univariate Analysis on Employment Length



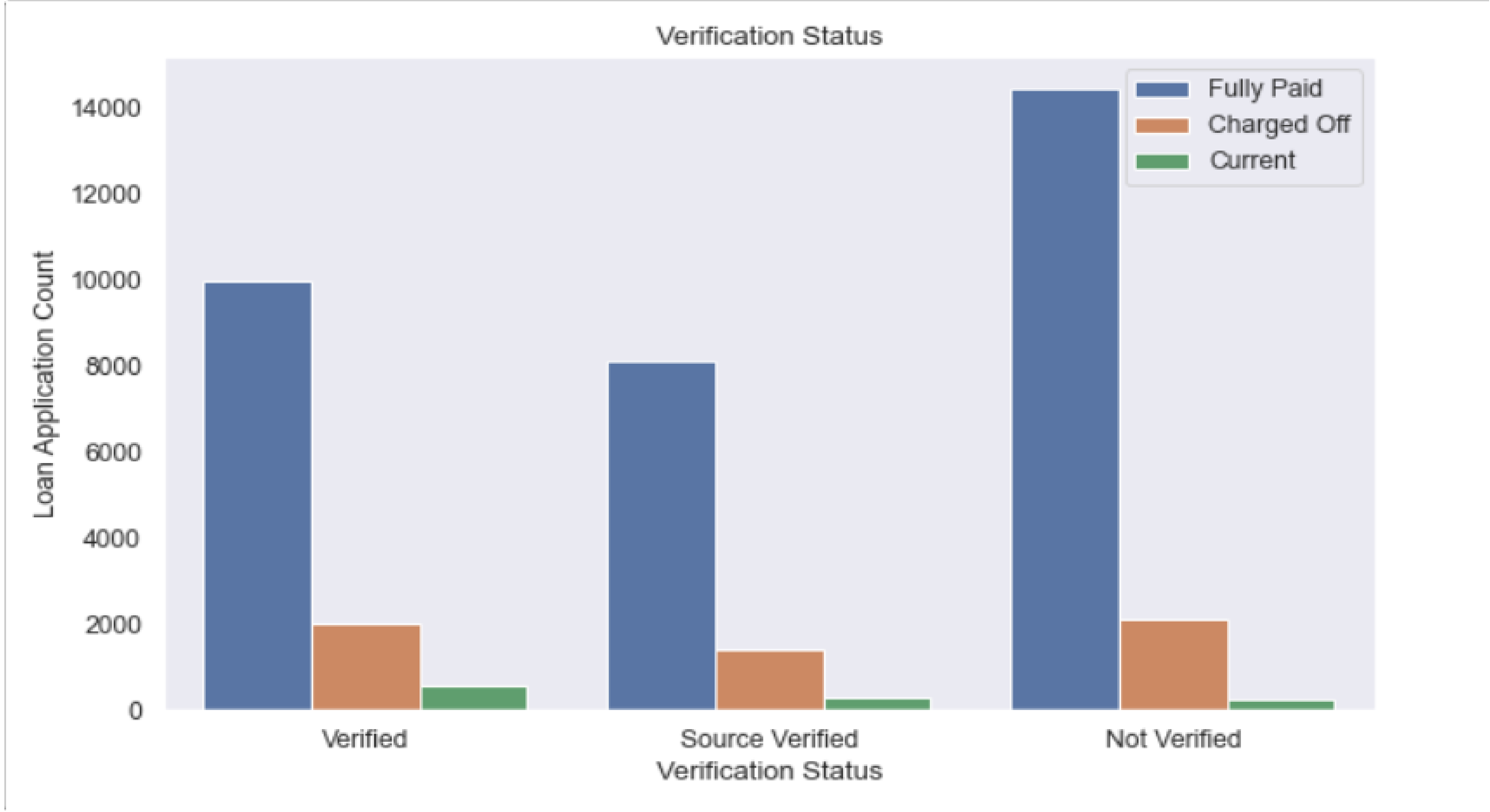
Most of the applicants who apply for loan have employment length of 10 years succeeded by employment length of 1 year and 2 year.

Univariate Analysis on Loan Status



Only 14% loans out of total loan issued were charged off.

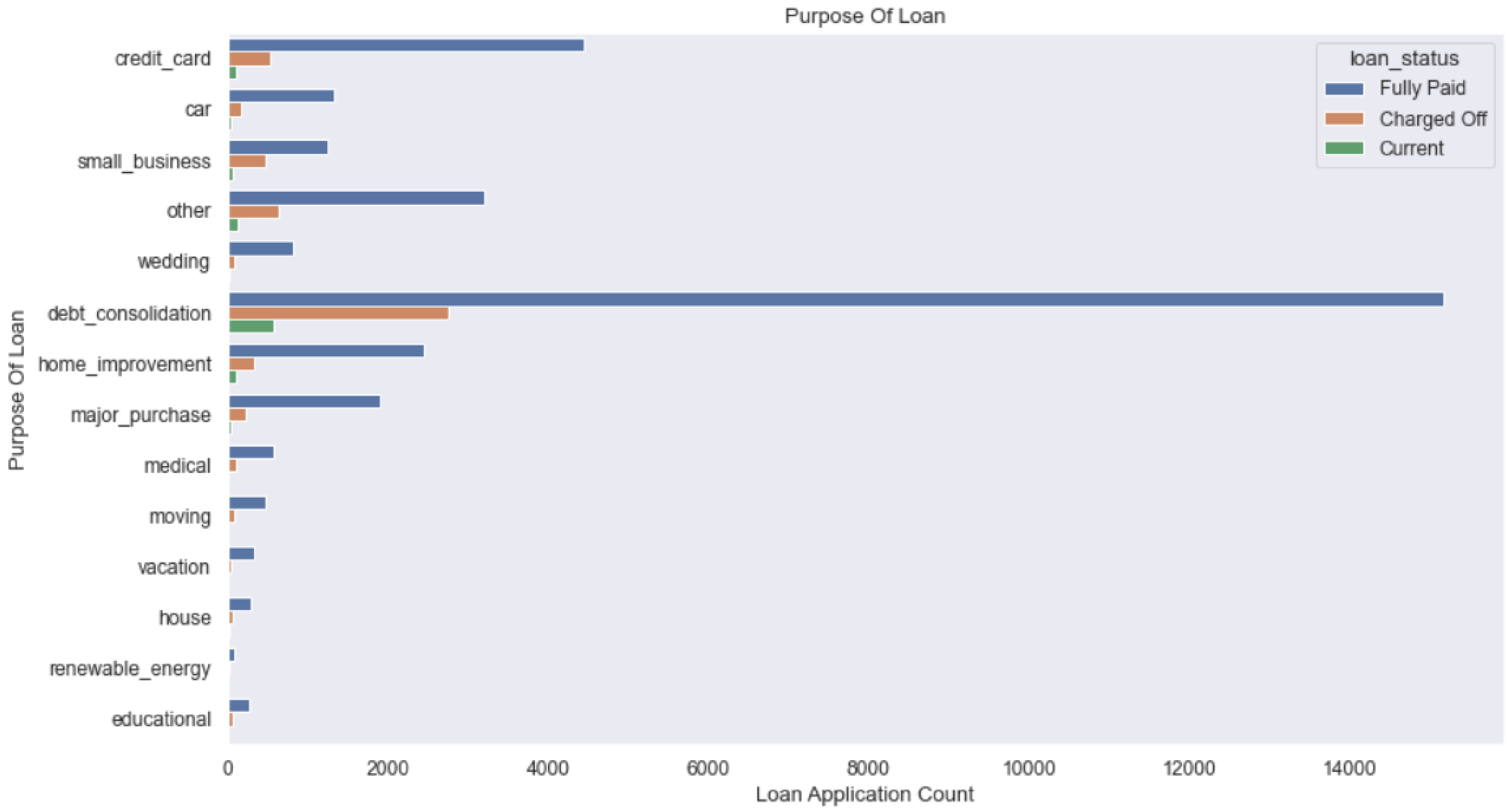
Univariate Analysis on Verification Status



Applicants who are source verified have slightly less charged off cases but we have less applicants who are source verified

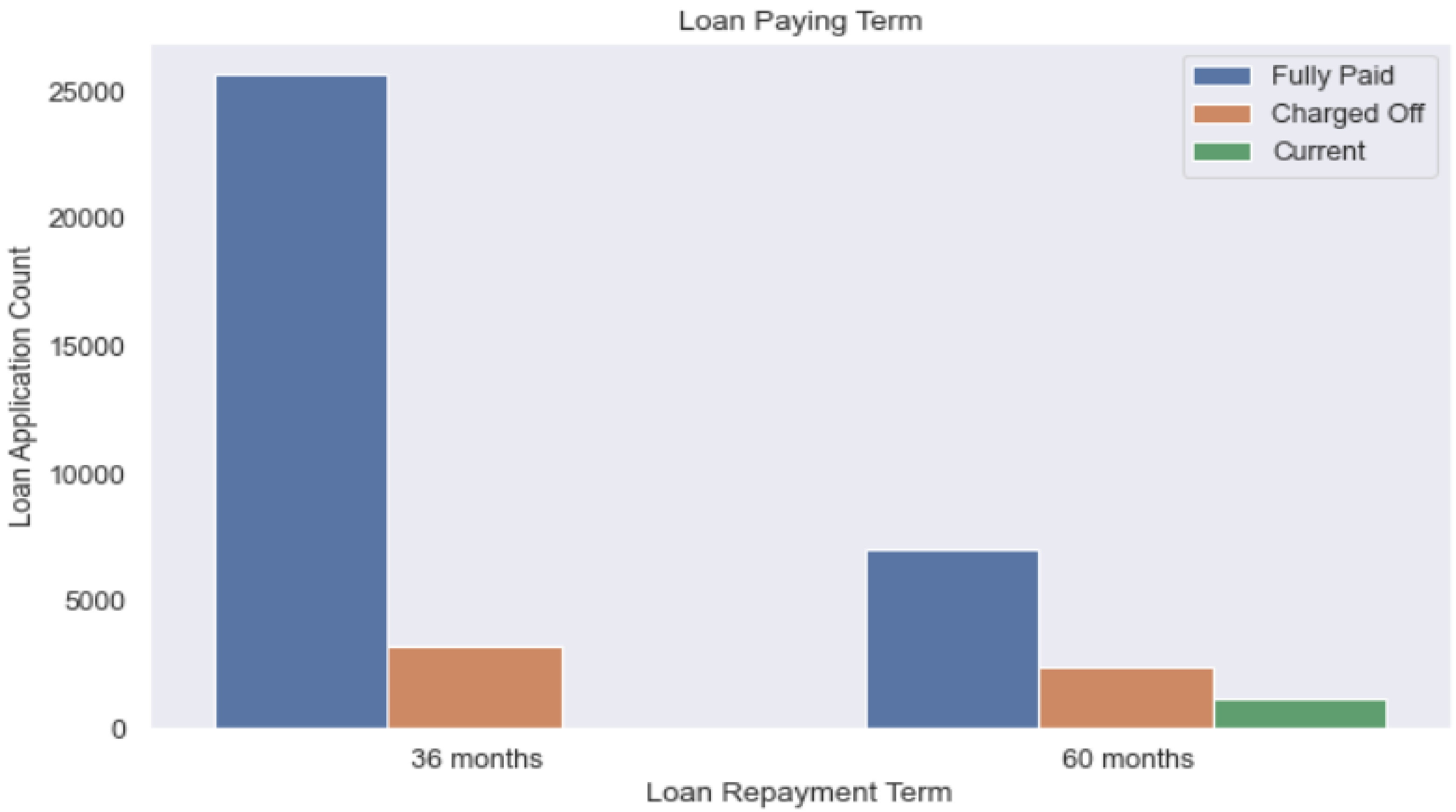
Note: It seems verification status does not show very significant difference for charged off cases

Univariate Analysis on Purpose of Loan



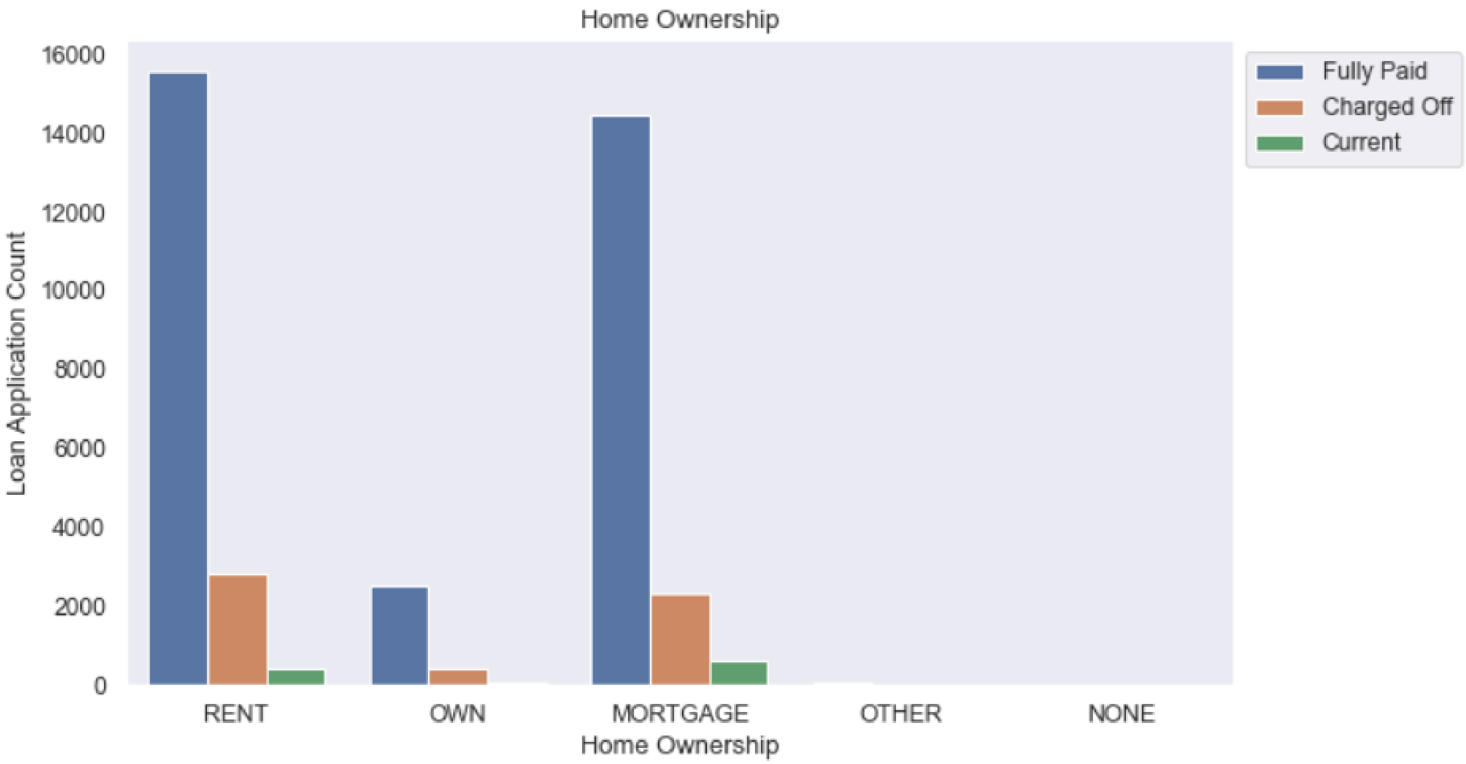
Purpose of most of the loans was debt consolidation and paying credit card bill and their charged off rate is also high

Univariate Analysis on Loan Payment Term



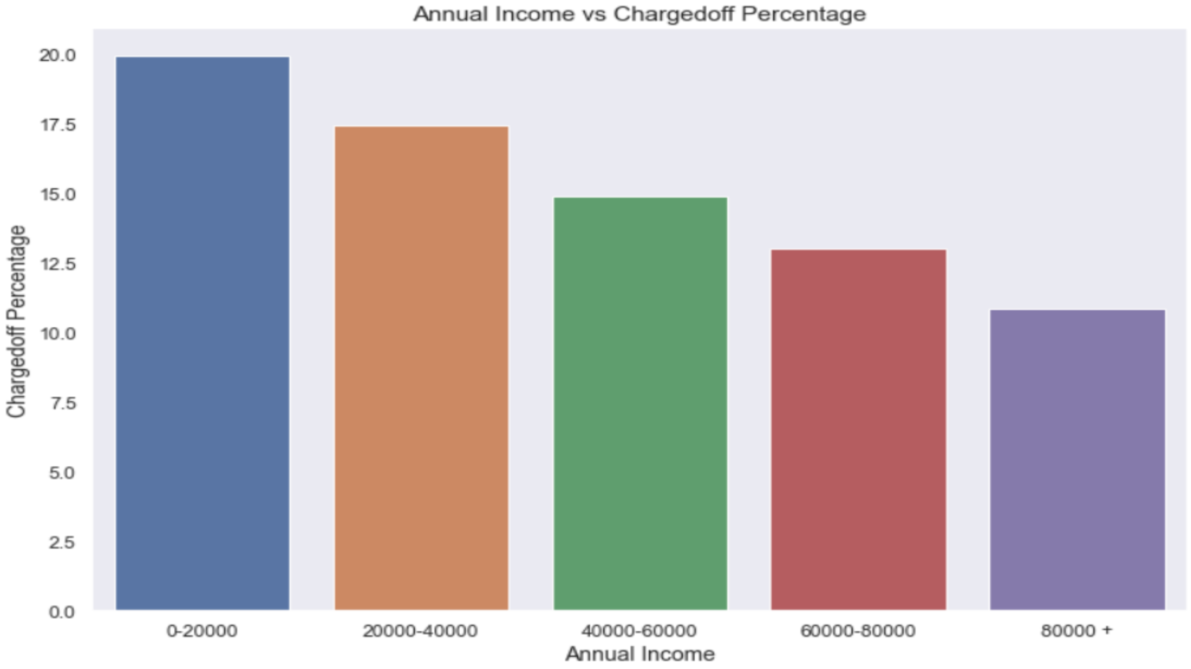
- There are more number applicants for Loan repayment term of 36 months
- Applicants taking loan to repay in 60 months have more charged off as compared to applicants taking loan for 36 months.

Univariate Analysis on Home Ownership



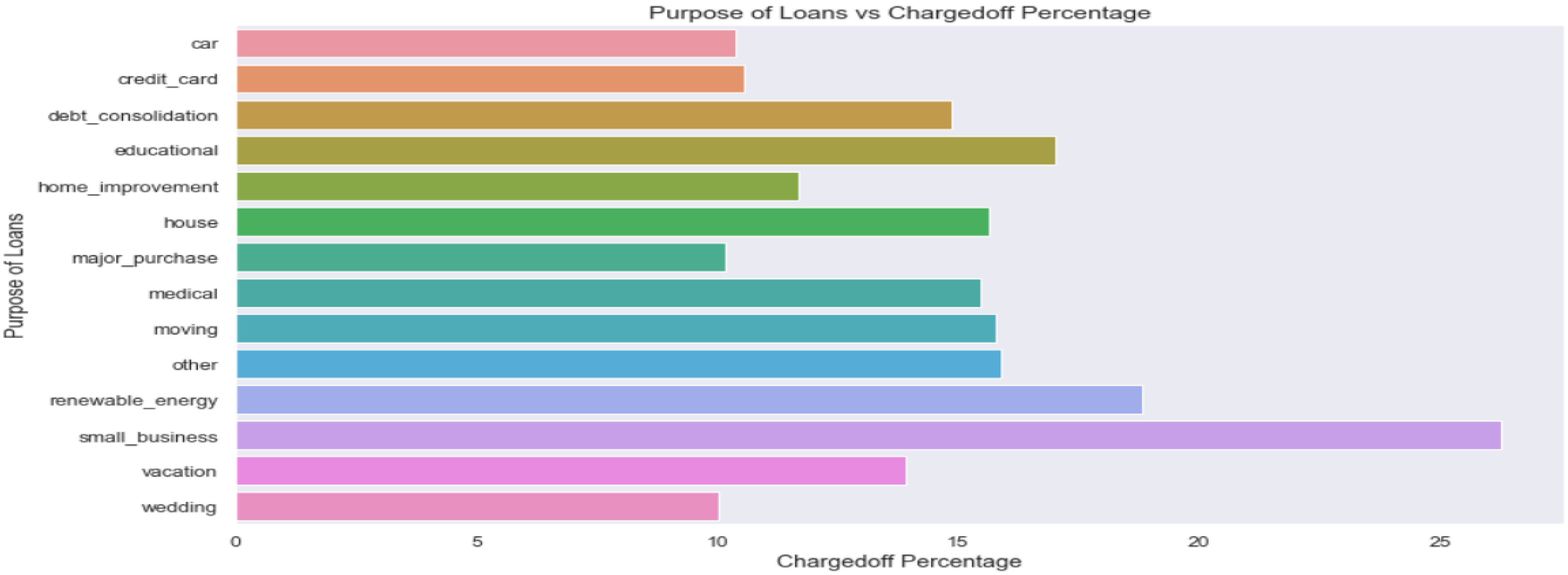
Most applicants live on rent followed by applicants who have mortgaged their house and their charged off rate is respectively high.

Bivariate Analysis of Annual Income on Percentage of charged off applicants



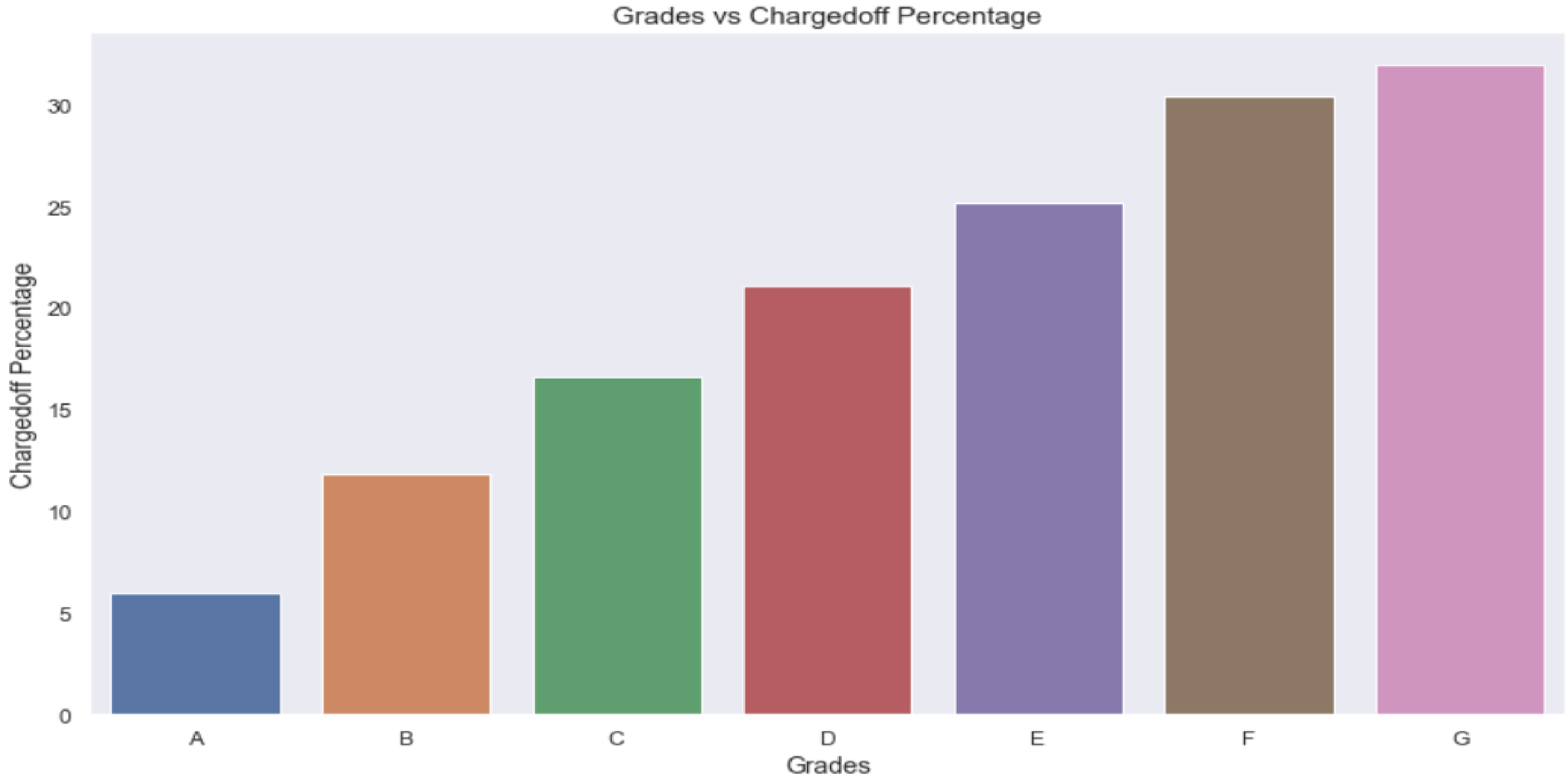
Higher the income range lesser the chances of being charged off

Bivariate Analysis of Purpose of Loan on Percentage of charged off applicants



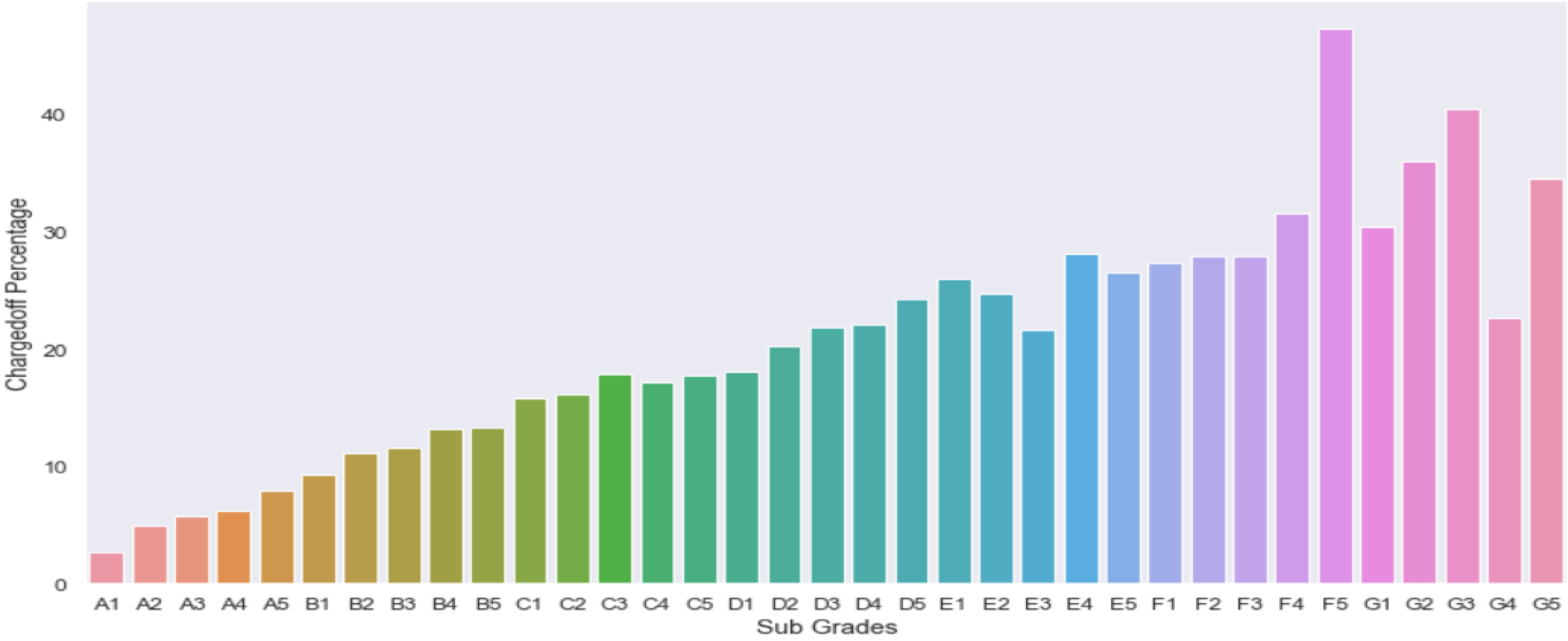
Applicants with purpose as small business have higher chances of getting charged off.

Bivariate Analysis of Applicant Grade on Percentage of charged off applicants



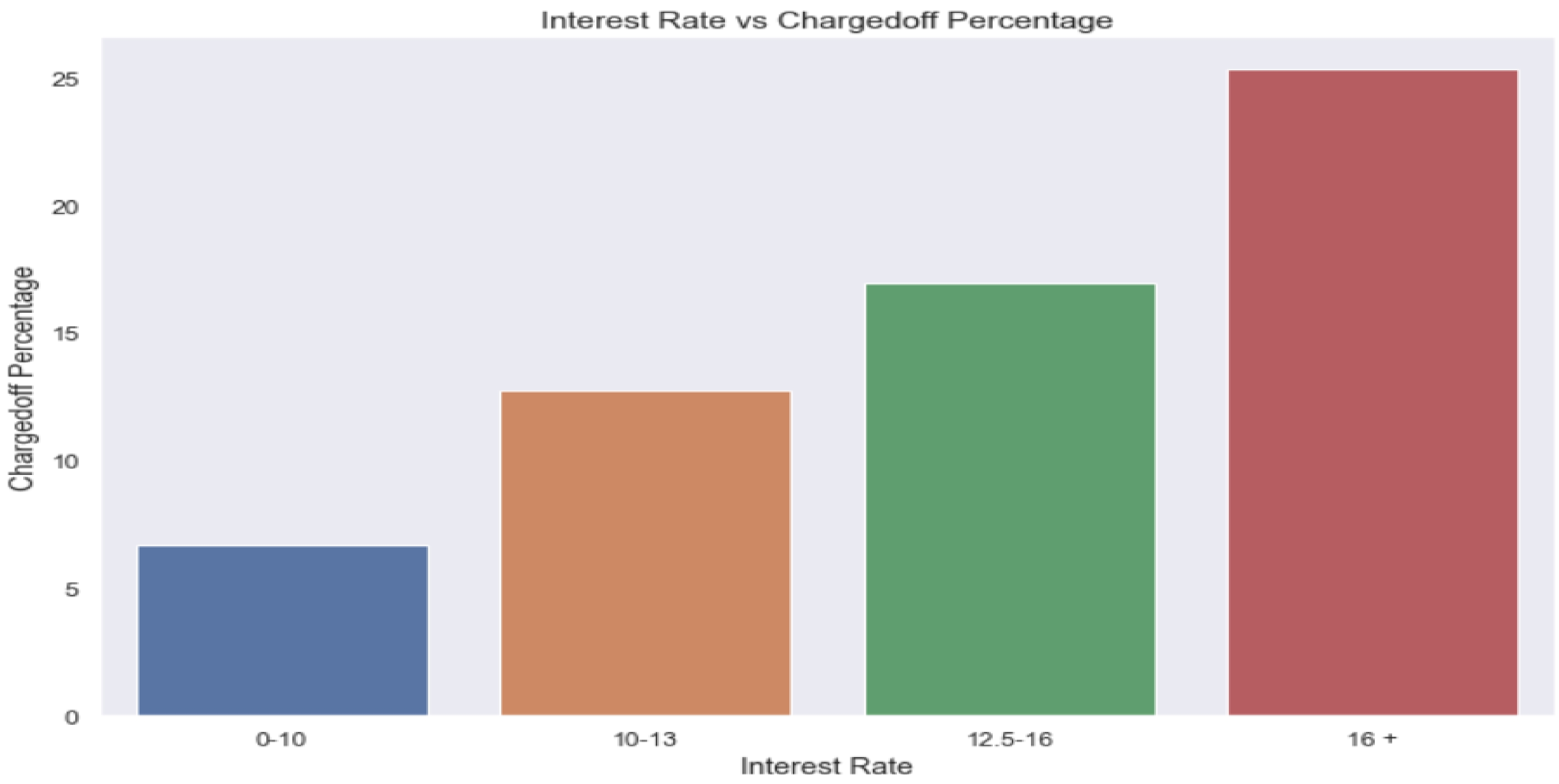
Applicants with better grade have lesser chances of getting charged off

Bivariate Analysis of Applicant Sub-Grade on Percentage of charged off applicants



Applicants with better sub grade have lesser chances of getting charged off

Bivariate Analysis of Interest Rate on Percentage of charged off applicants

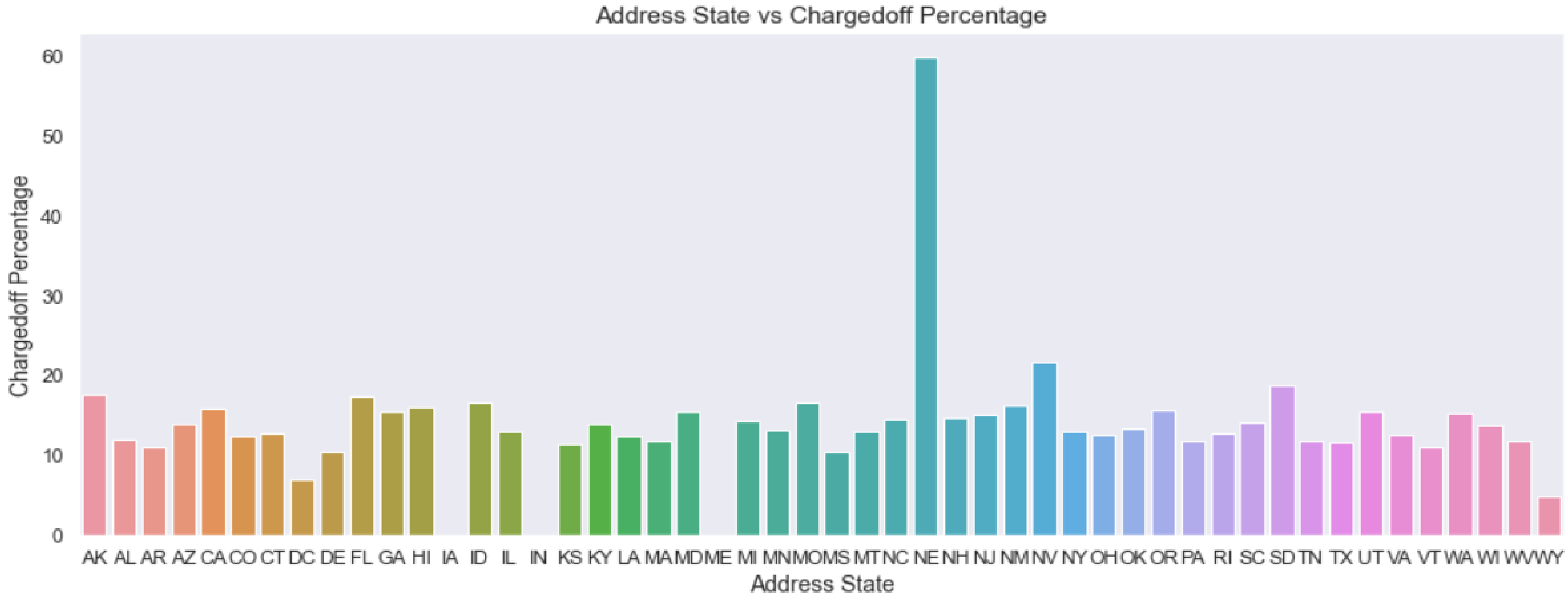


Charged off percentage increases with increase in interest rates.

Bivariate Analysis of Employment Rate on Percentage of charged off applicants

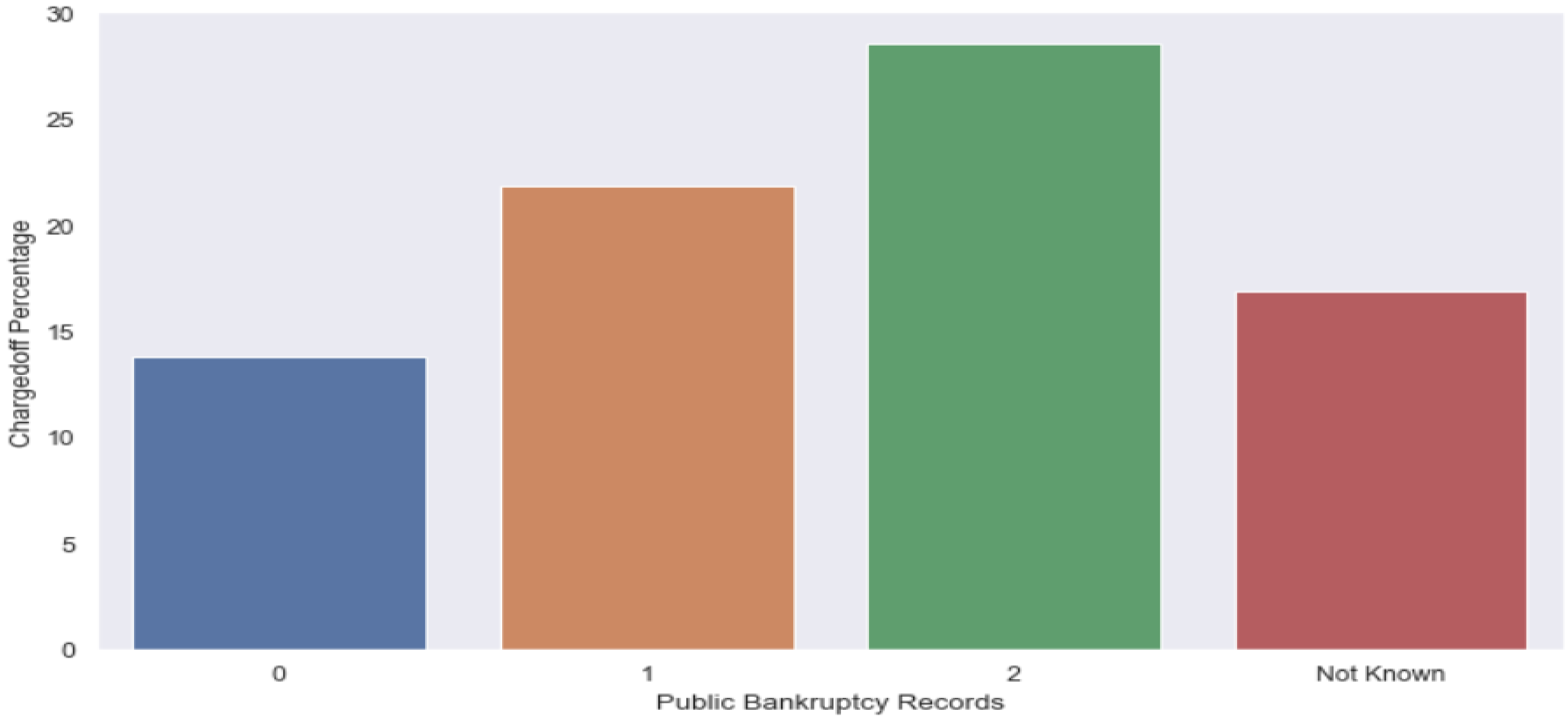
Higher percentage of applicants who are not working or have less than 1 year of work experience get charged off

Bivariate Analysis of Address State on Percentage of charged off applicants



Applicants from states NV,CA and FL have higher number of charged offs.

Bivariate Analysis of Number of public record bankruptcies on Percentage of charged off applicants



Applicants who have defaulted before has more chances of defaulting in future as well.

Recommendations based on Analysis

- Applicants who are under low income bracket have higher chances of being charged off. This would mean higher the income range lesser the chances of being charged off.
- Applicants with purpose as small business have higher chances of getting charged off.
- Applicants getting charged off increases with grade moving from "A" towards "G“
- Applicants who has defaulted before has more chances of defaulting in future as well.
- Applicants who are not working or have less than 1 year of work experience have high chances of getting charged off, reason being no source of income to repay loan.
- Applicant whose purpose of loan is debt consolidation and paying credit card bill have high chances of getting charged off.
- Applicant who have longer loan repayment term have high chances of getting charged off.
- When applicant lives in rented house or their houses are mortgaged chances of getting charged off increases.

Recommendations based on Analysis

- Dti (debt to income ratio) increases with increase in employment length.
- Applicants verification status has not proved to be significant factor for defaulter analysis.
- Applicants from states NV,CA and FL have higher number of charged off
- Applicants with purpose as debt consolidation and paying credit card bill have higher chances of getting charged off.
- Applicants getting charged off increases with sub grade moving from "A1" towards "G5“. It is noted that most cases of applicants being charged of are under F5 subgrade.