Lending Club Case Study

TEAM MEMBERS:

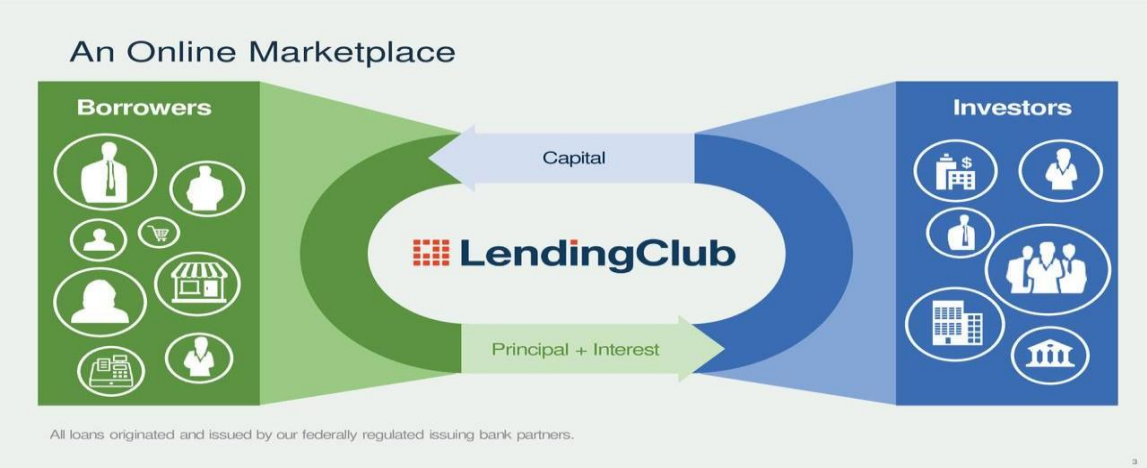
SAGNIK BANERJEE

RAKESH KUMAR SHAH

Introduction

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.



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

Problem Statement

The company basically 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.

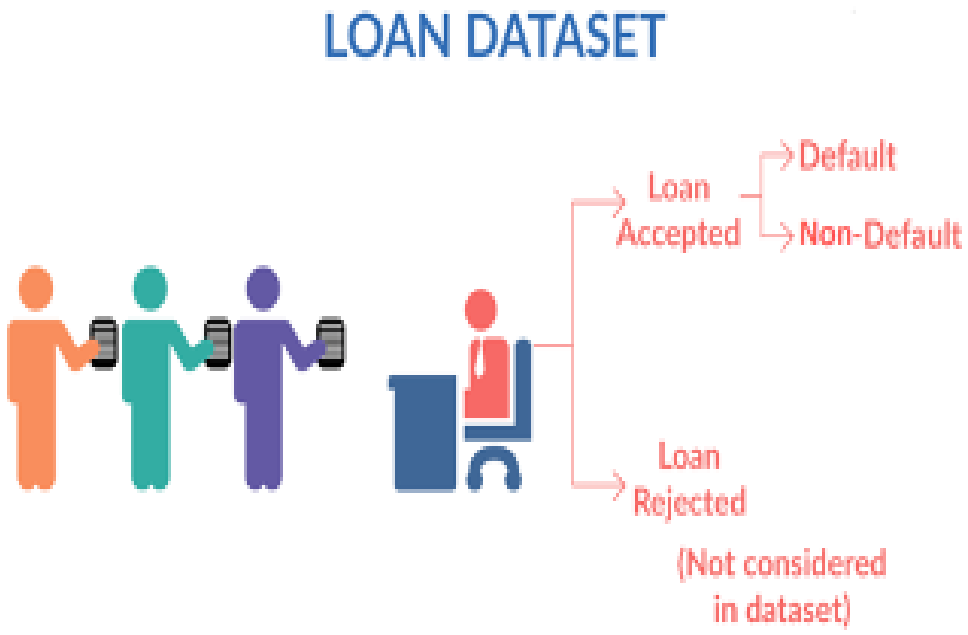
The main aim of this case study is to identify Risky Loan Applicants, thereby cutting down the amount of credit loss of the company.

Solution Approach

As part of this case study we have taken the below mentioned steps as part of the solution:

1. Data understanding
2. Data cleaning (cleaning missing values, removing redundant columns etc.)
3. Data Analysis
4. Recommendations

Data Understanding



Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan.

As part of this case study we have just considered Fully paid and Charged-off Loan status, from the dataset.

Data Cleaning

In Data Cleaning we performed the various steps.

1. We removed the columns which were having more then 60% of the records NULL and dropped duplicates.
2. We removed the columns where it was having unique records throughout and also the unique columns.
3. We identified the behavioral columns which was having data related to the applicants post loan approval were removed.
4. We filled up all the missing value by taking appropriate actions such as mode, mean and median.

Data Analysis

Post Data Cleaning and making the data proper by filling missing value we performed the various Analysis such as:-

1. Univariate Analysis.

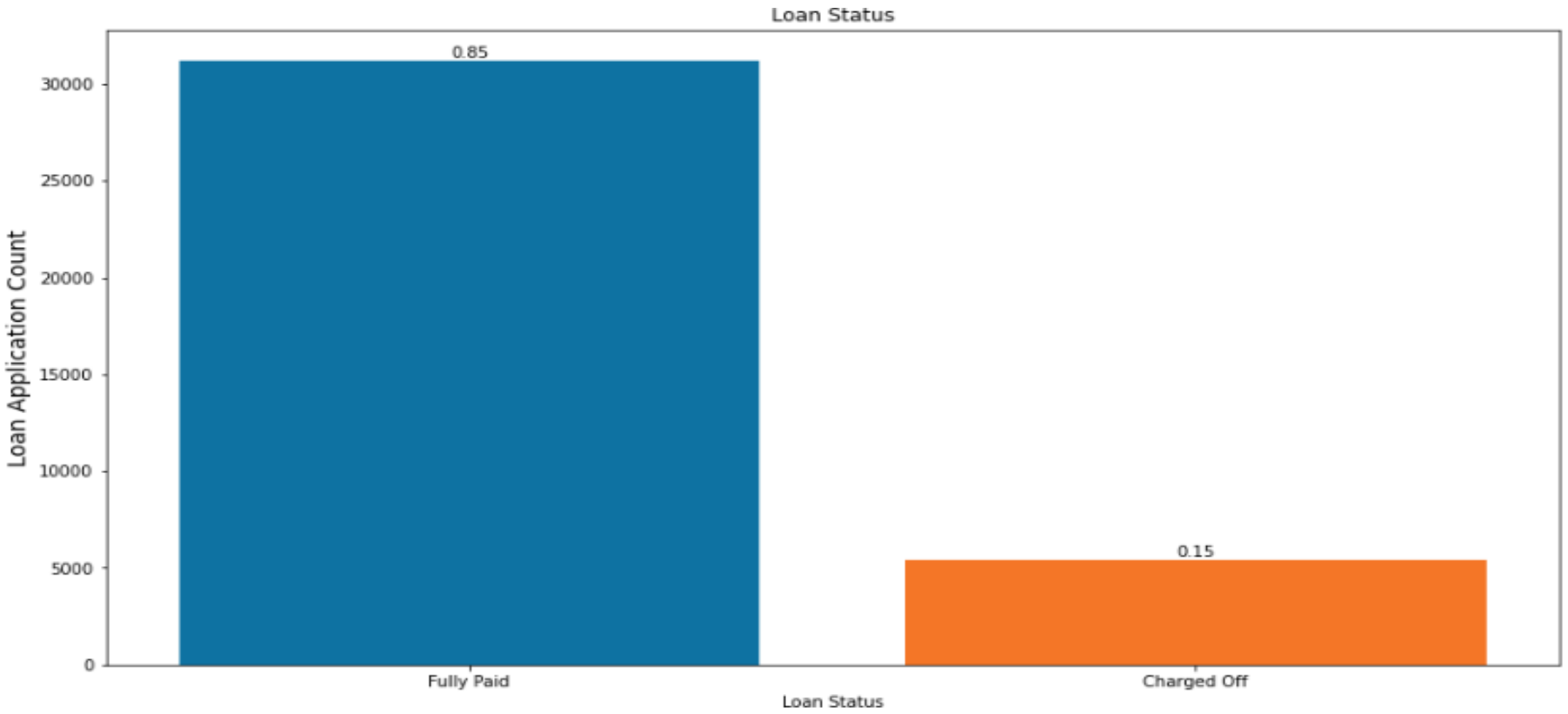
   Various Analysis was performed on

   1.Categorical Variables

   2.Continuous Variables

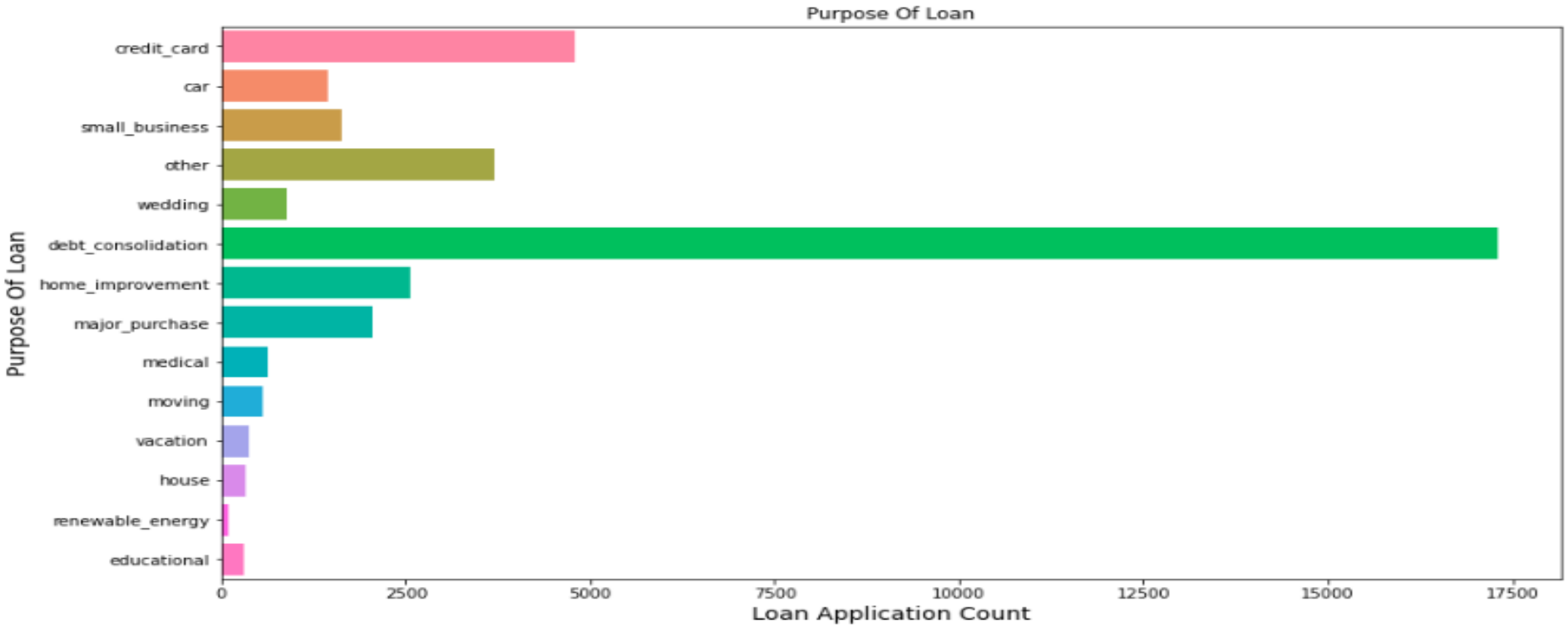
2. Bivariate Analysis.

Univariate Categorical Variable Analysis – Loan Status



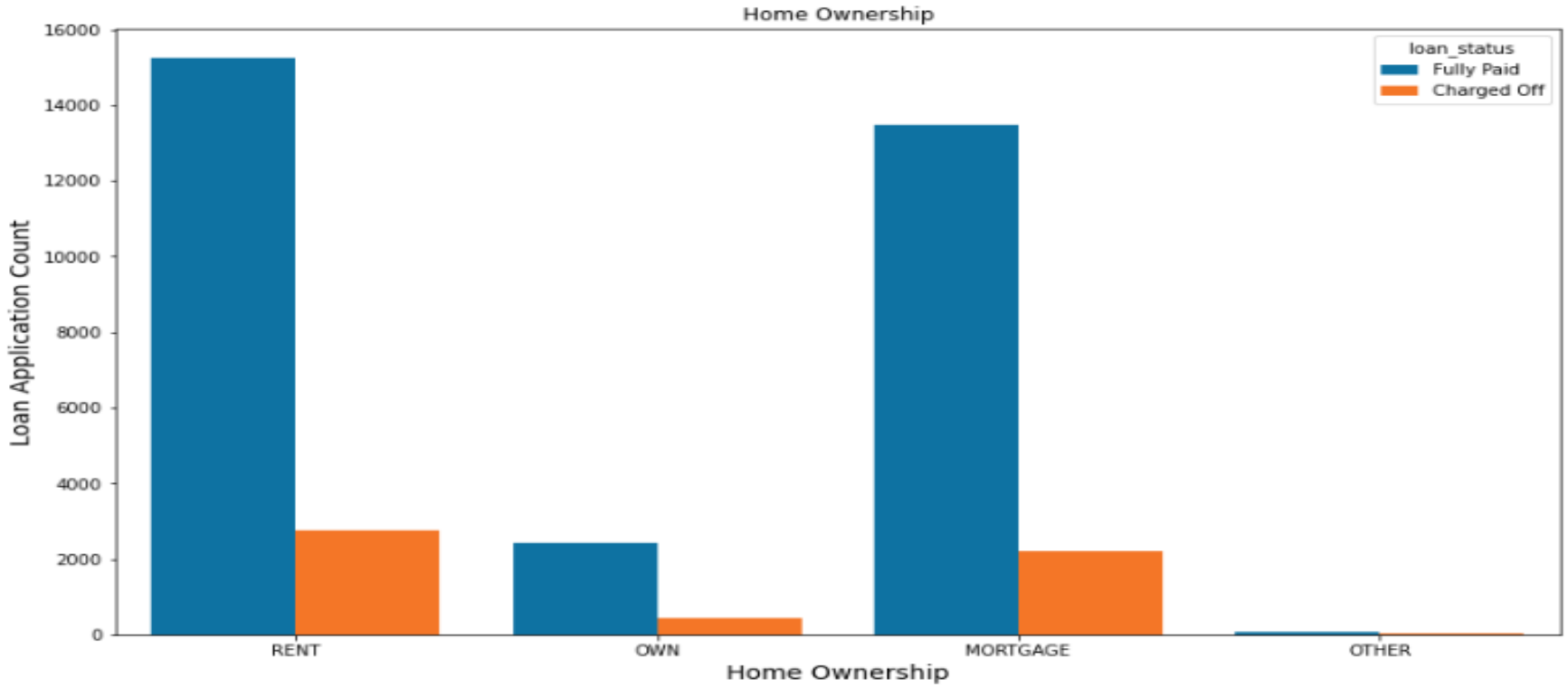
The above plot shows that around 15% loans were charged off out of the total loans issued

Univariate Categorical Variable Analysis – Purpose Of Loan



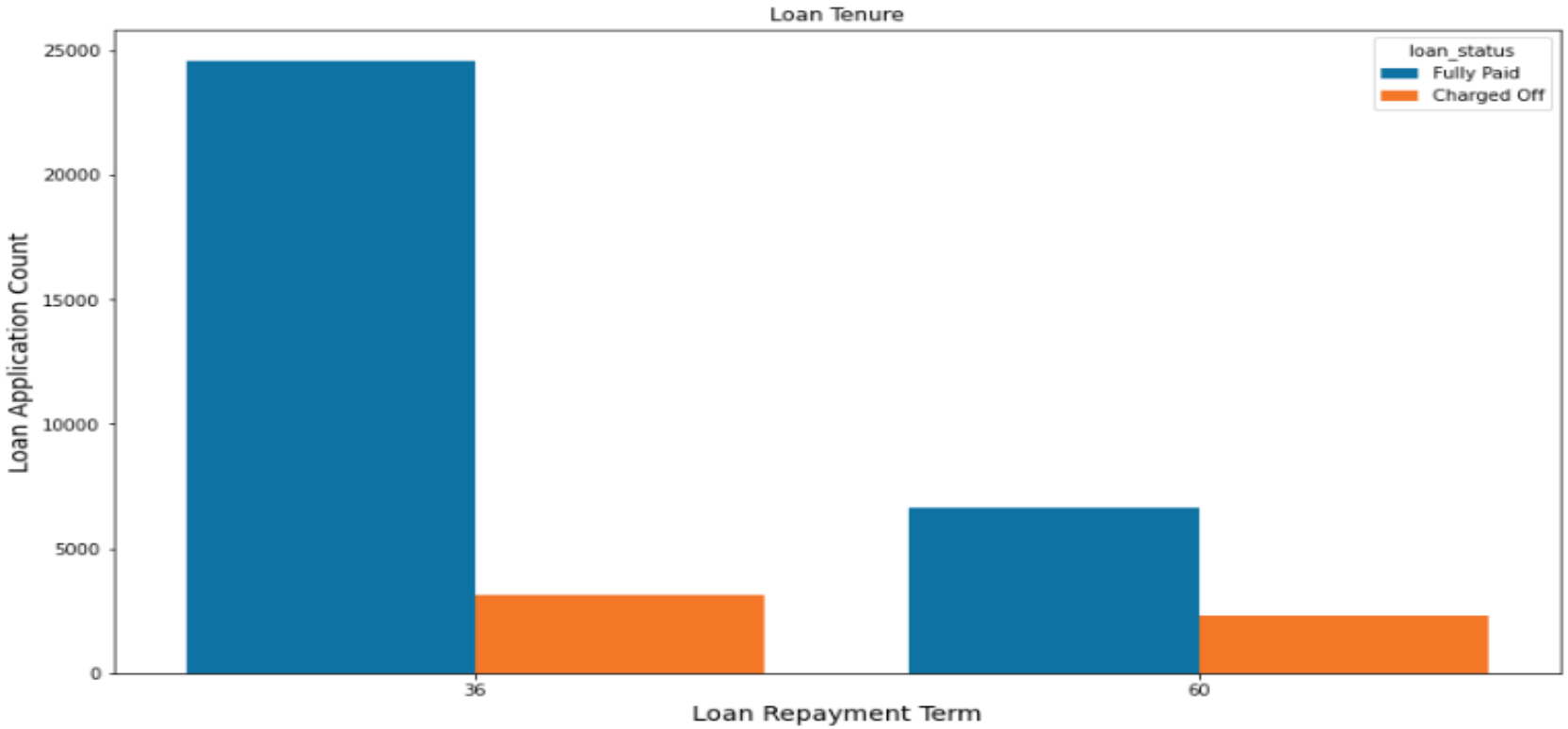
The above figure shows that most of the loans were taken for the purpose of debt consolidation.

Univariate Categorical Variable Analysis – Home Ownership



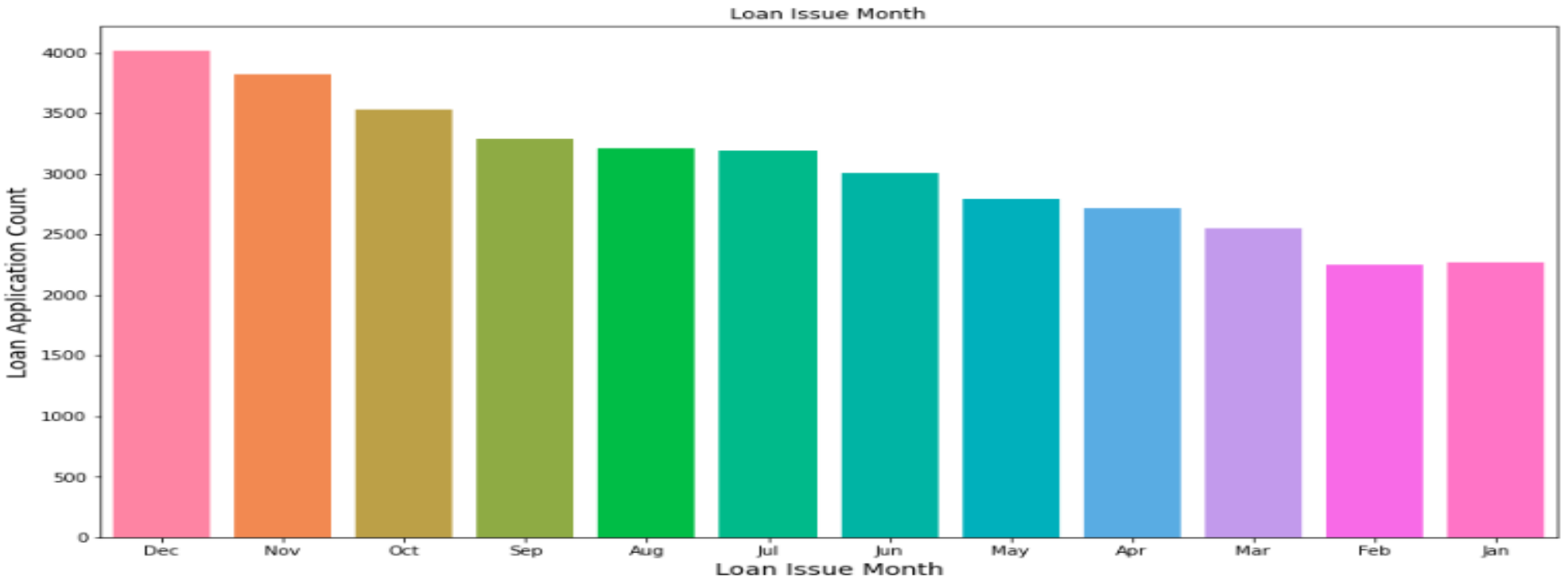
From the above figure below analysis we can conlcude that most of loan applicants living in rented home or mortgazed their home. They also have highest number of deafulters

Univariate Categorical Variable Analysis – Loan Tenure



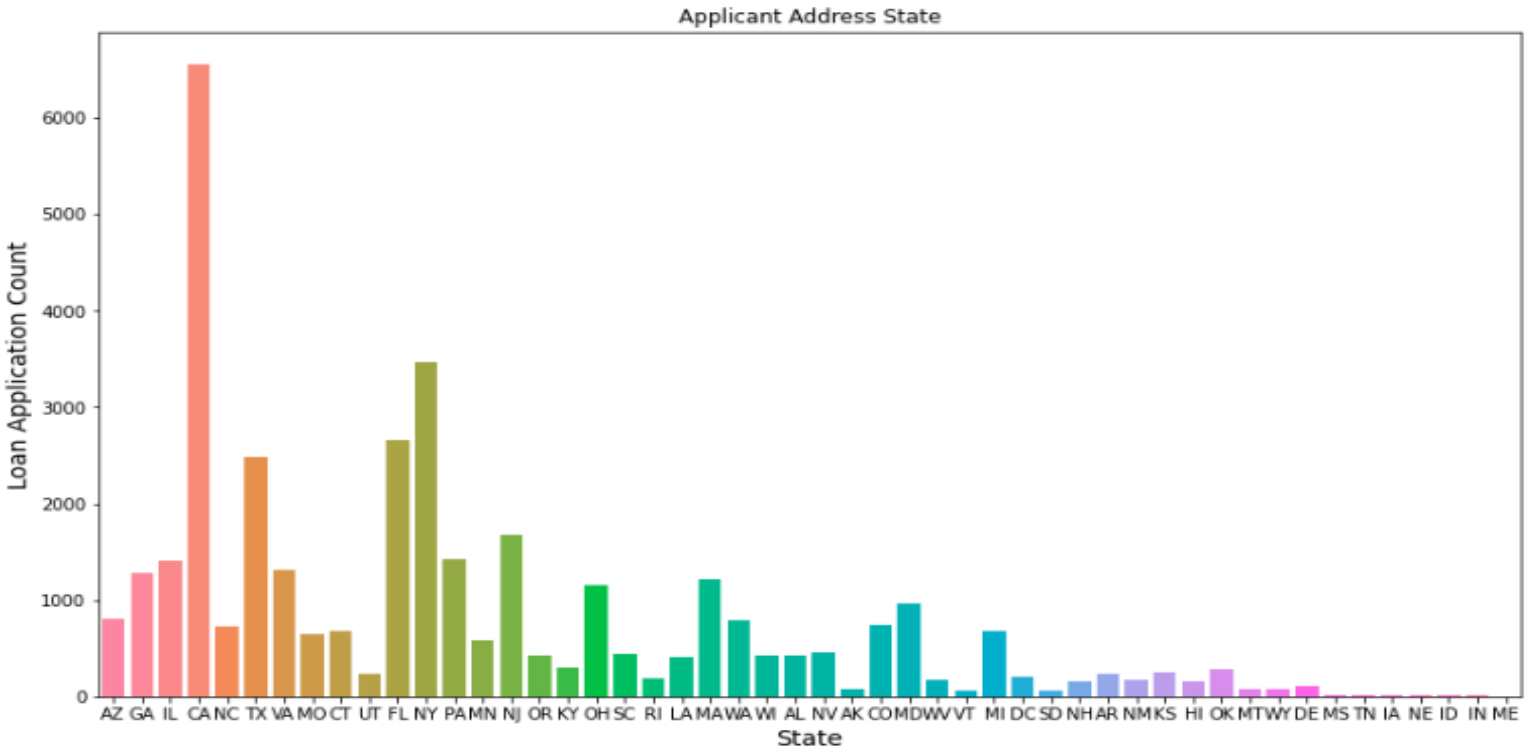
For the above figure, we can say that therer are more applicants taking loan for 36 months and hence more percentage of applicants are getting charged off as compared to 60 months tenure

Univariate Categorical Variable Analysis (Derived)– Loan Issue Month



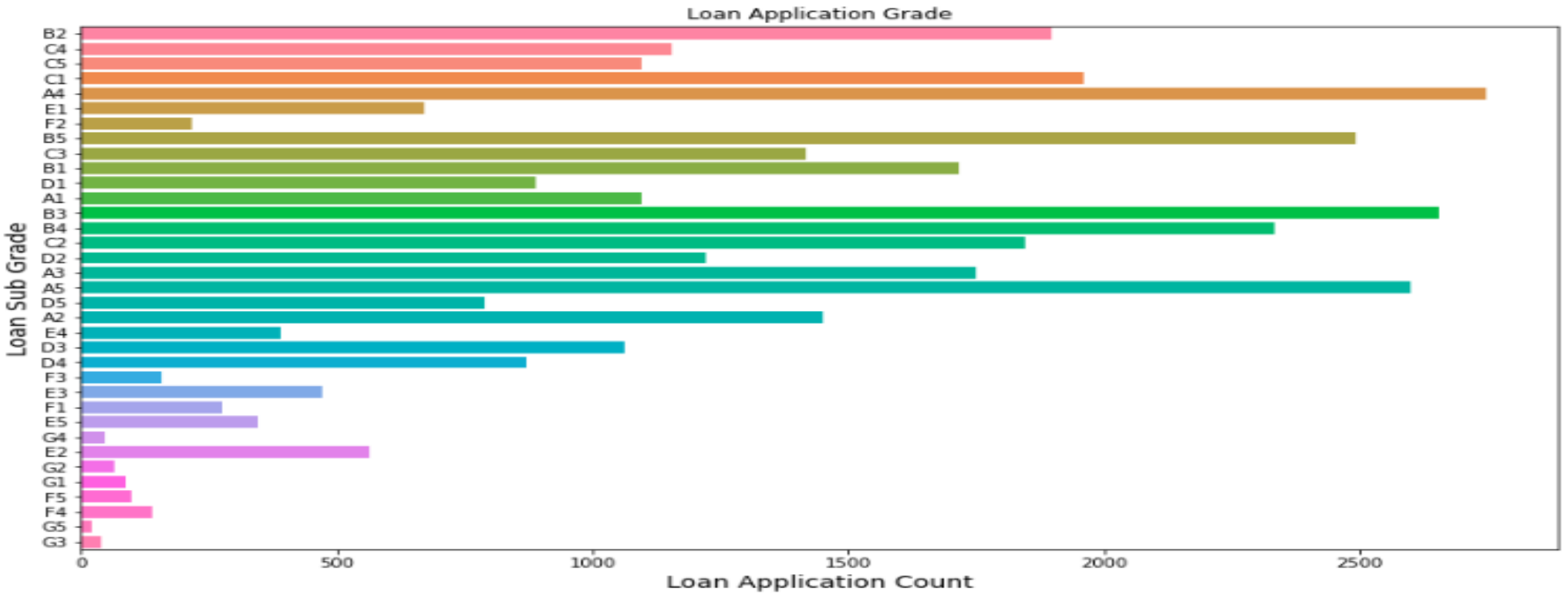
From the above figure we can conclude that People are taking more loans during Holiday seasons of November and Decemeber

Univariate Categorical Variable Analysis – application address Sate



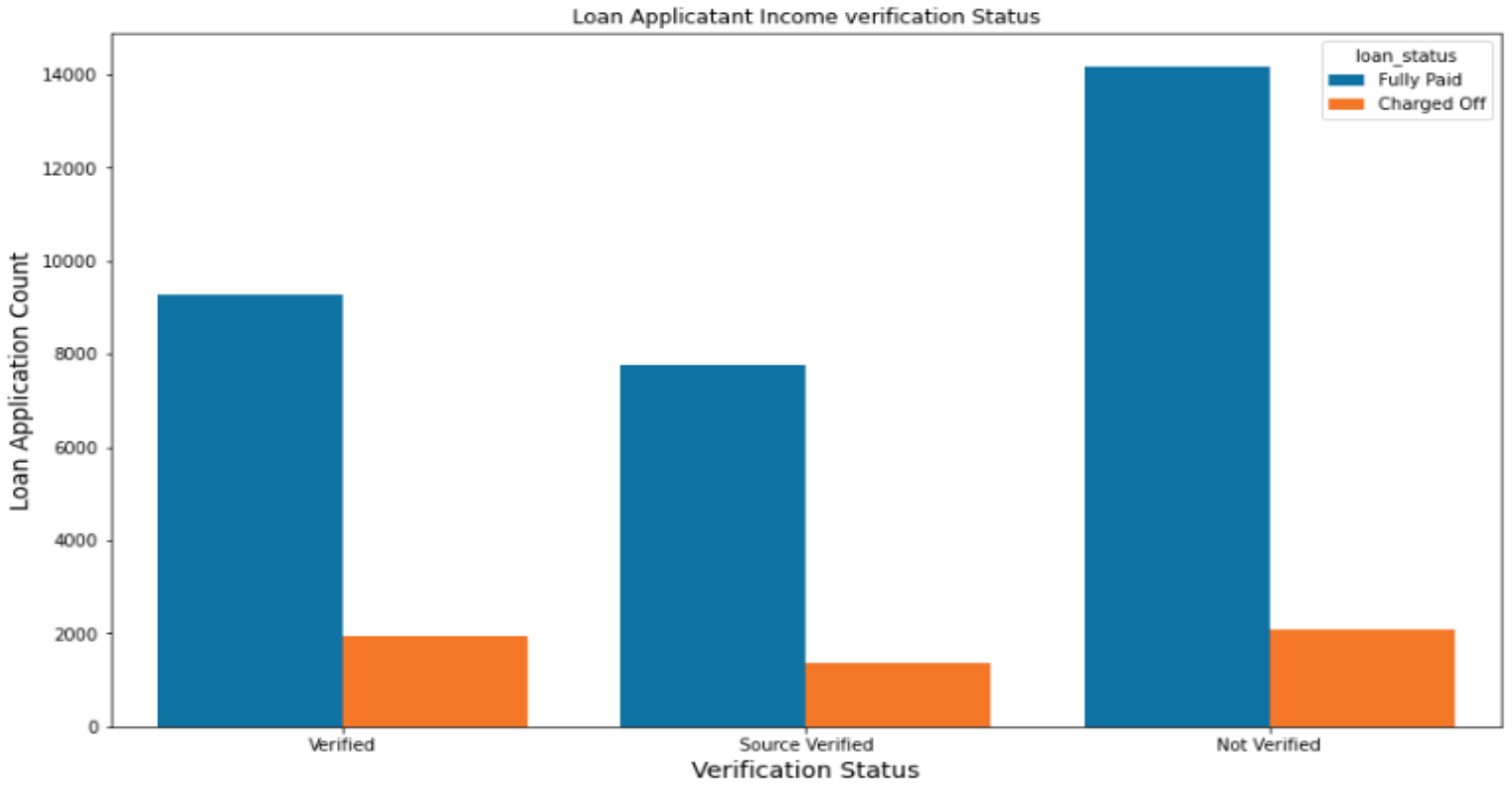
Top Loan Applications originate from well known progressive states such as CA, NY, FL, TX with larger degree of urbanization, tax rates and higher than average income levels.

Univariate Categorical Variable Analysis – Loan Application Grade



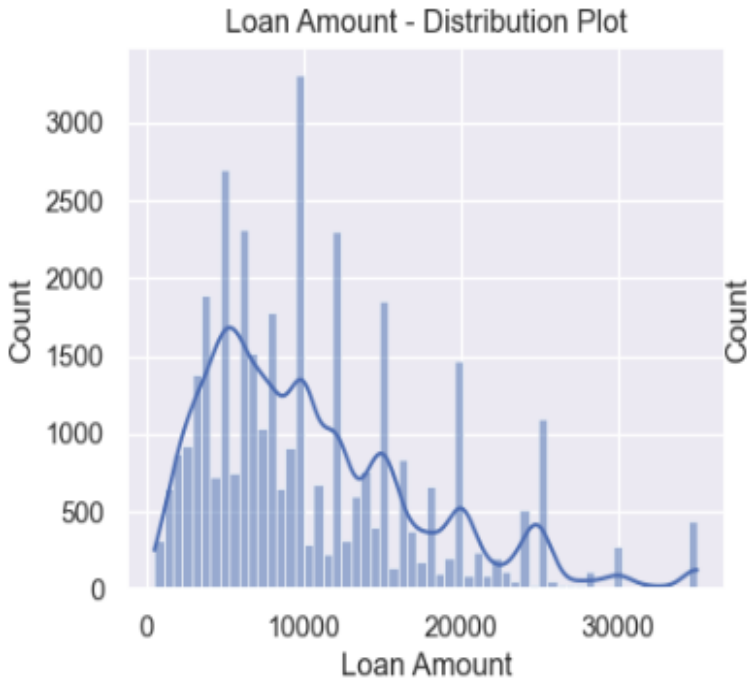
From the above figures,it is evident that, most of the loan accounts comprise of the top 3 grades (A, B and C). Most of the applicants are from B grade.

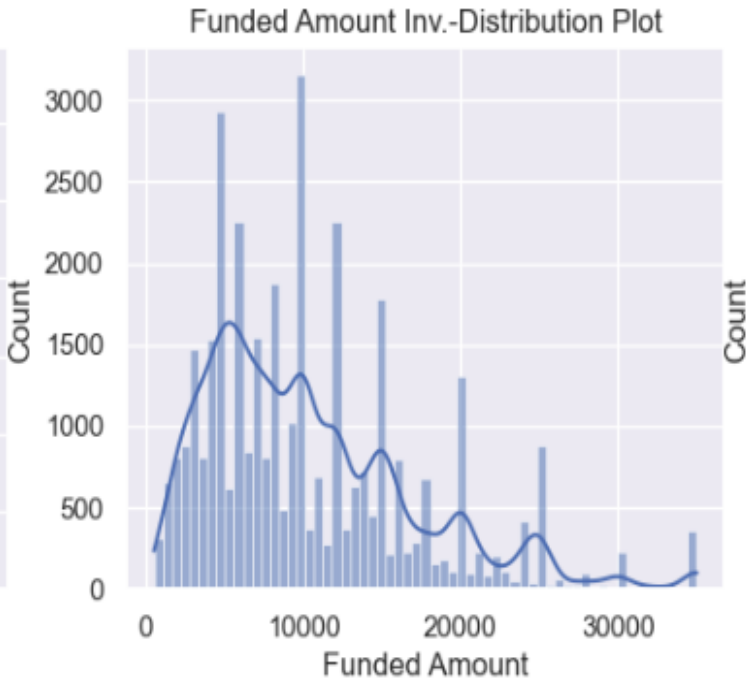
Univariate Categorical Variable Analysis – Verification Status

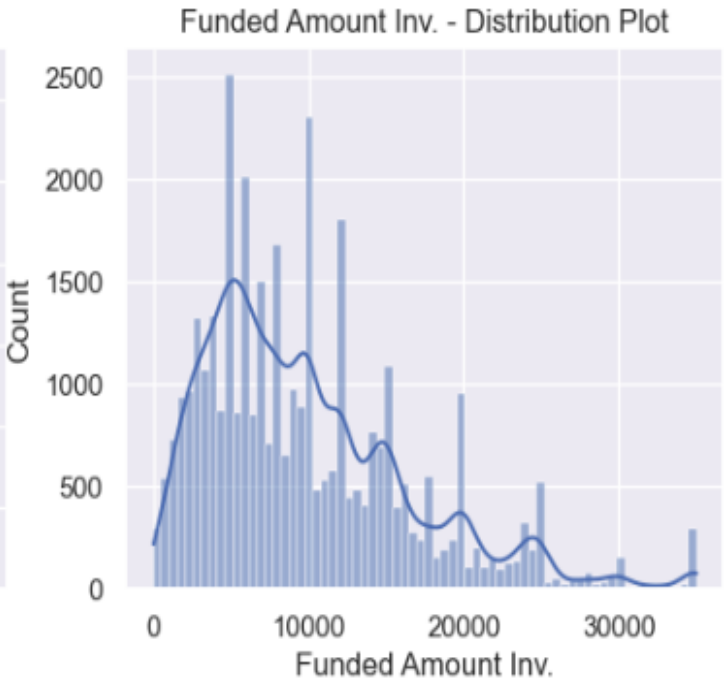


For the above figure it is clearn that most of the Loan applicants income source is not verified and has highest loan charged-off counts. Hence the recomendationshould be verify all the applicant income source.

Univariate Continuous Variable Analysis – Loan Amount, Funded Amount, Funded Amount INV

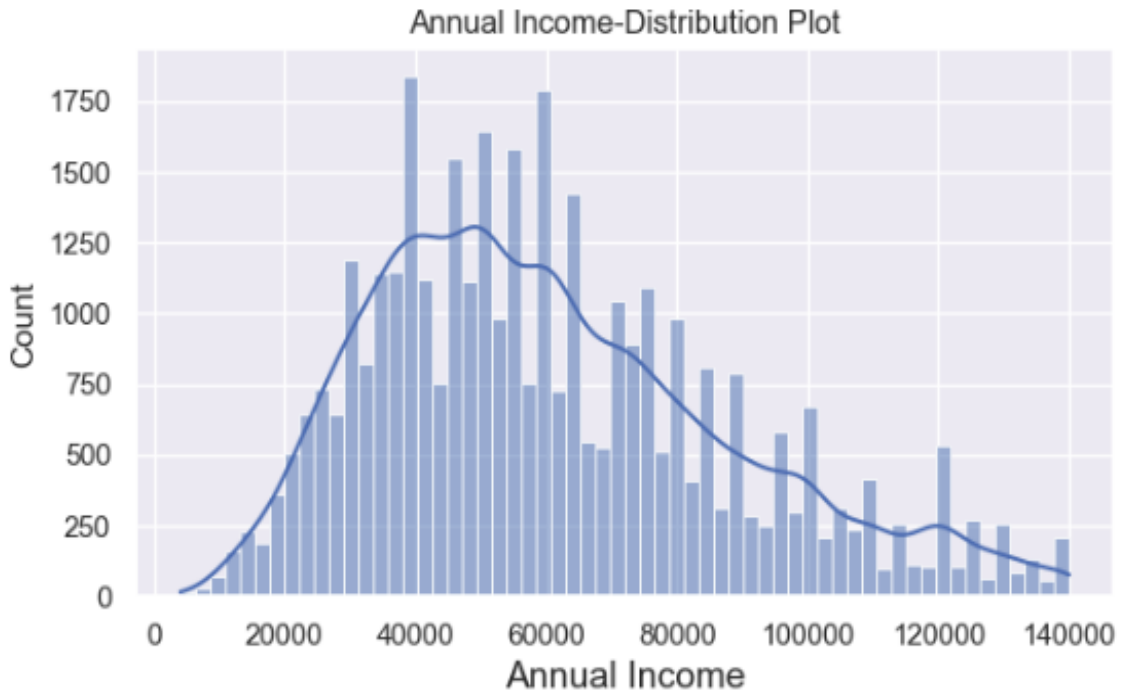


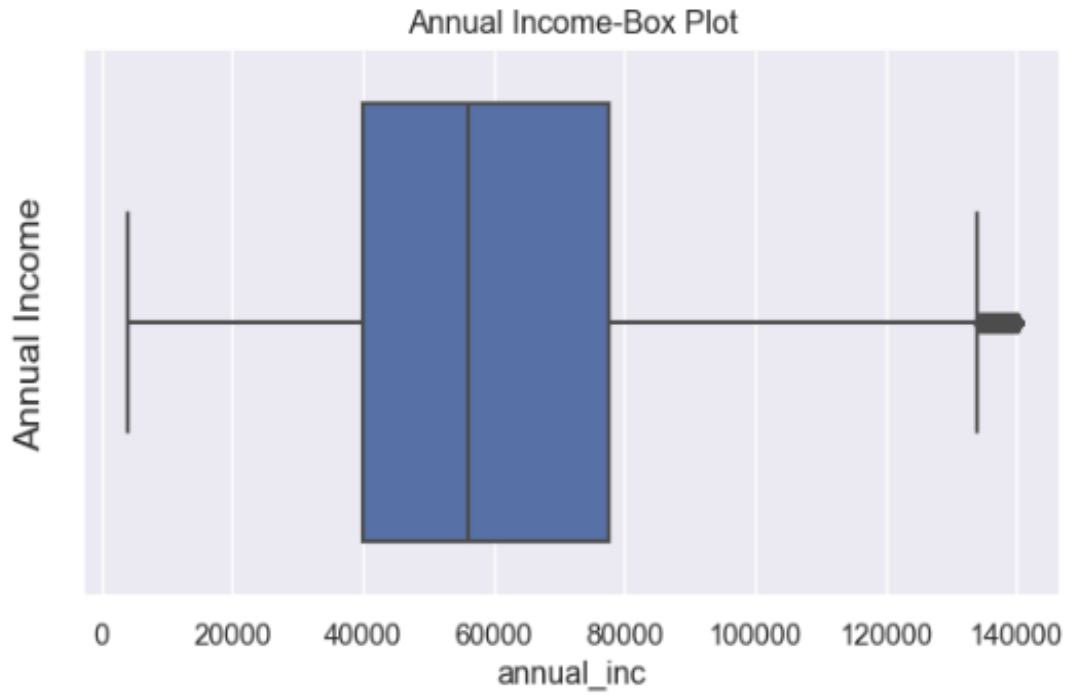




Distribution of amounts for all three plots looks very much similar. No Conclusions can be made from this

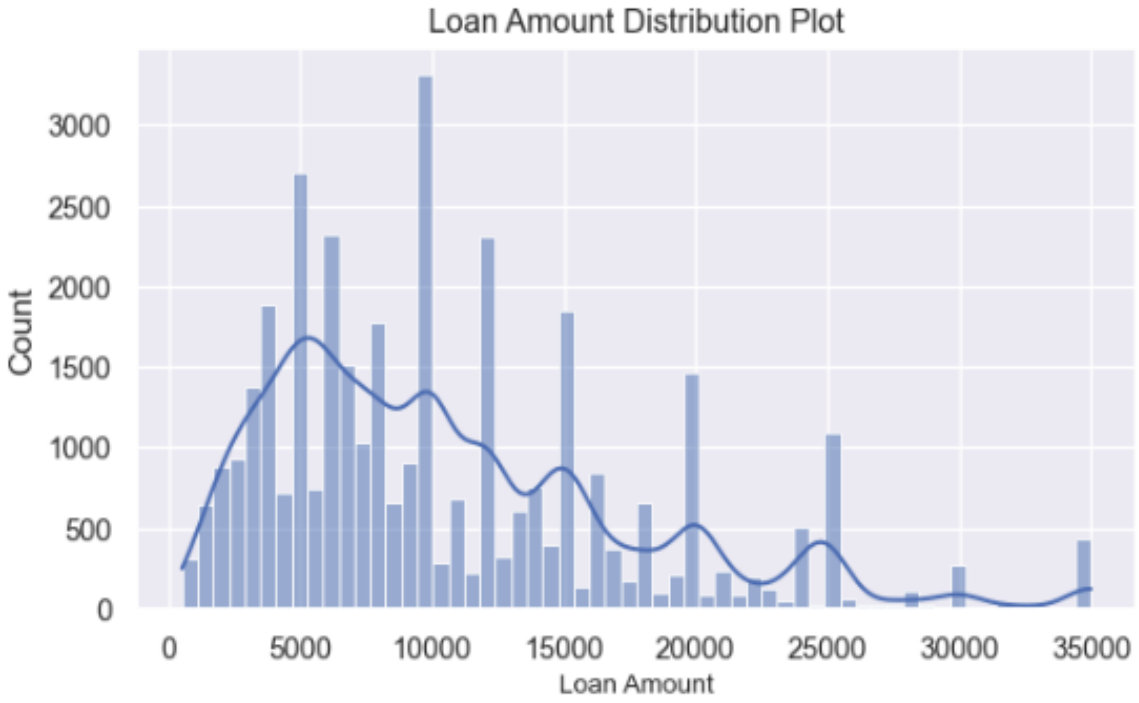
Univariate Continuous Variable Analysis – Annual Income

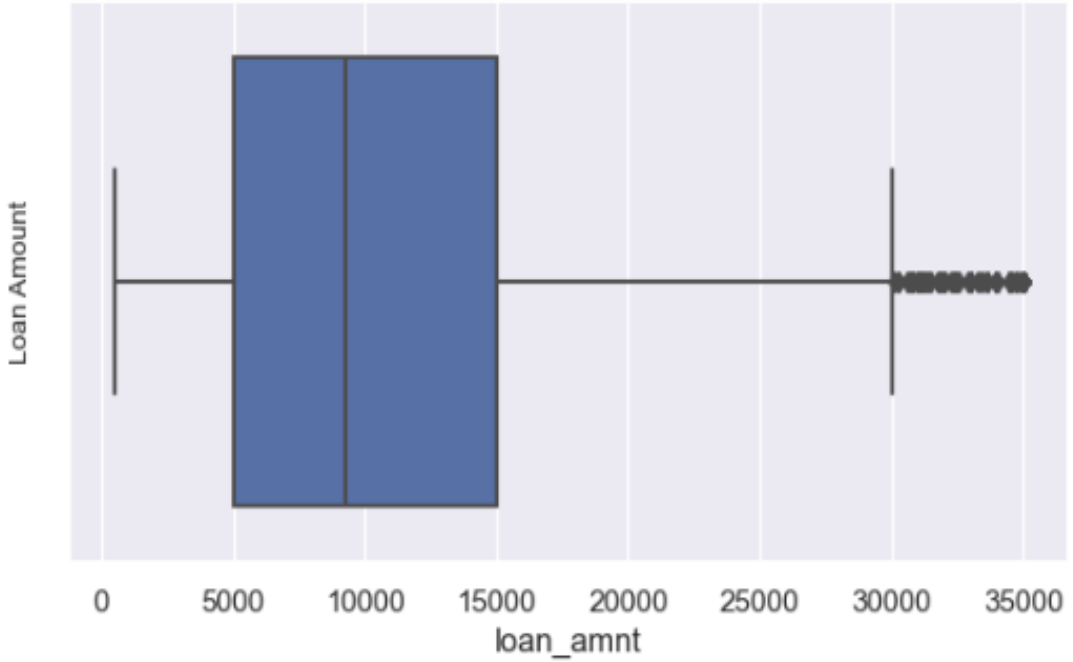




From the above figure we can conclude that most of the Annual Income are in the range 35000 USD to 80000 USD

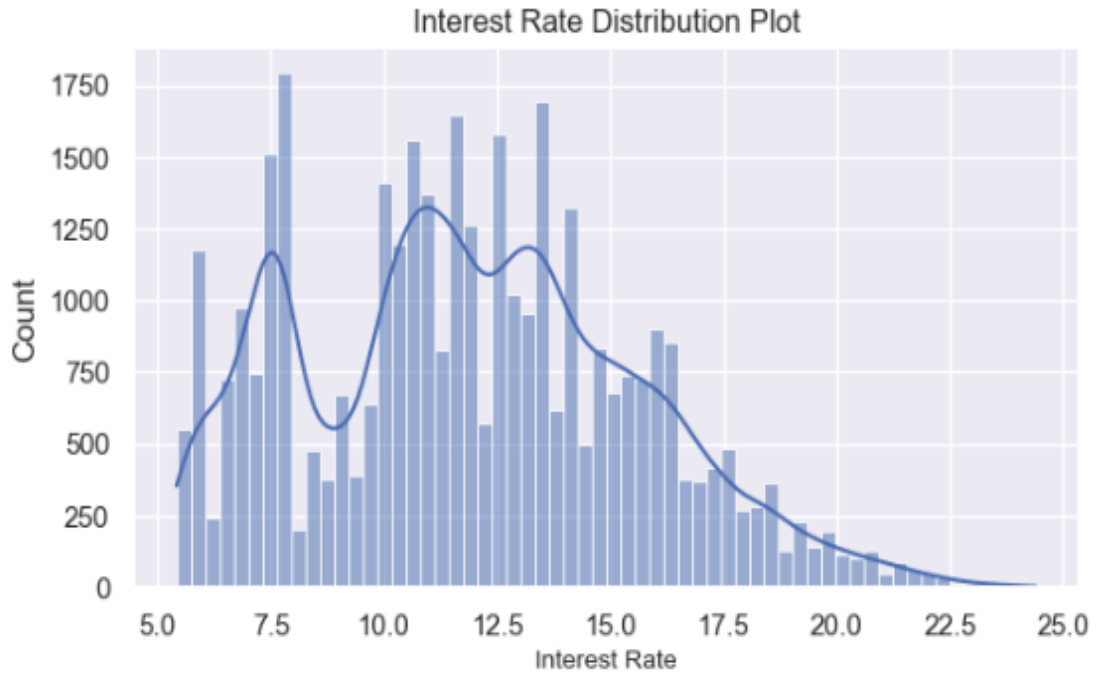
Univariate Continuous Variable Analysis – Loan Amount

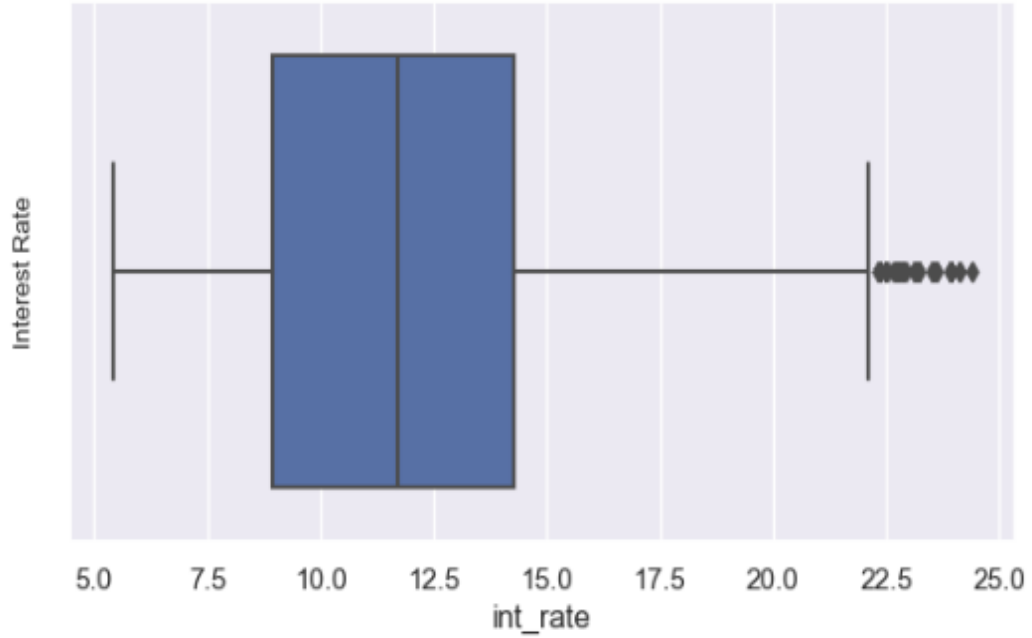




From the above figure we can conclude that most of the Loan Amount are in the range 5000 USD to 15000 USD

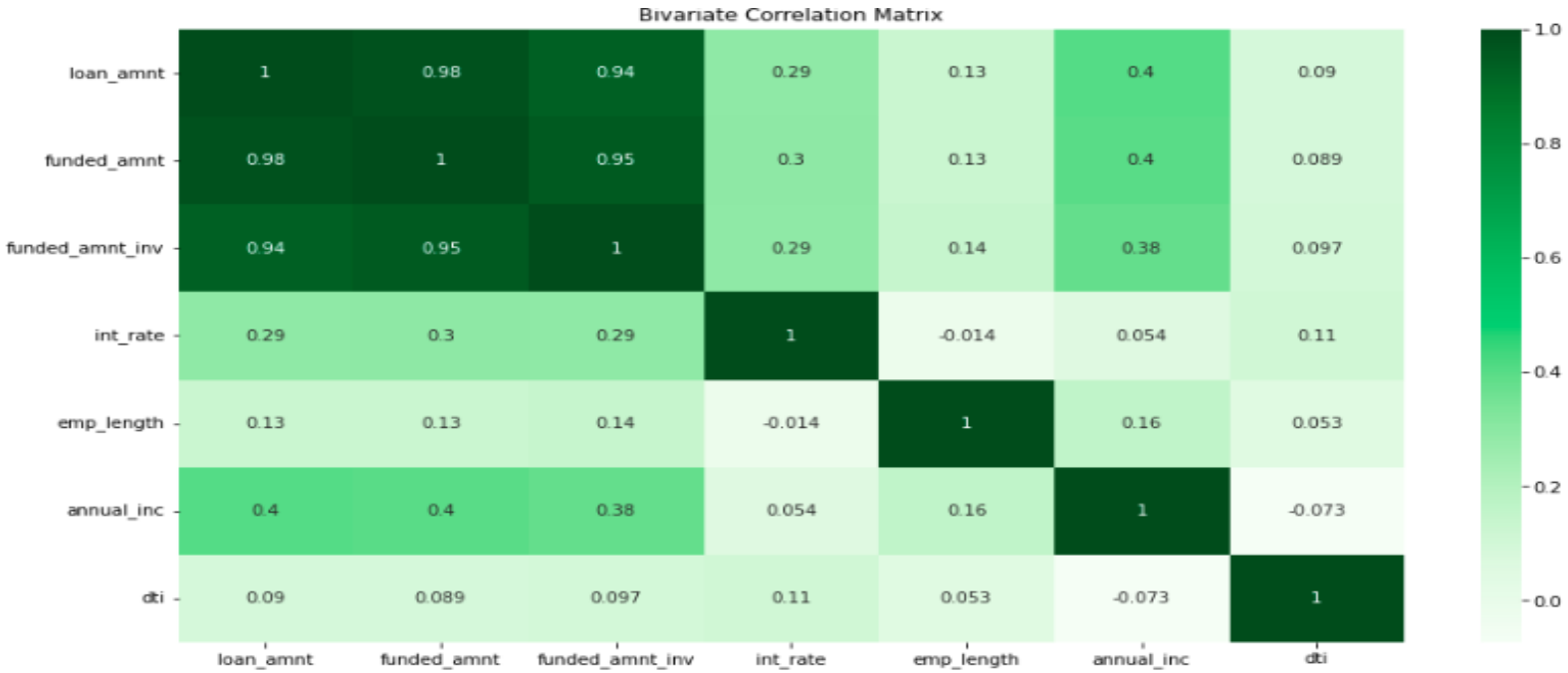
Univariate Continuous Variable Analysis – Interest Rate





From the above figure we can conclude that most of the Interest Rate are in the range 10% to 15%

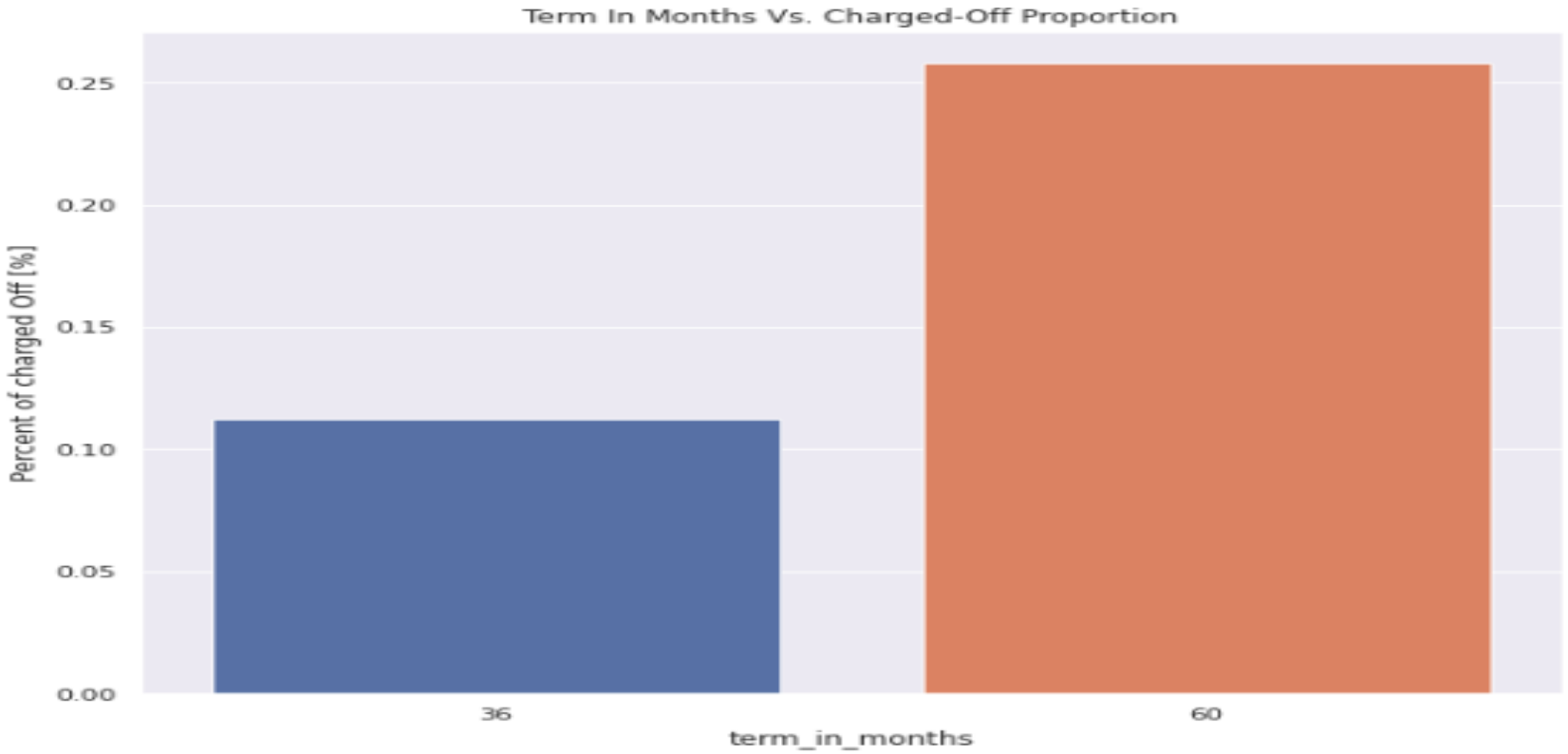
Bivariate Analysis – Correlation Matrix



List of Observations from the above correlation Matrix -

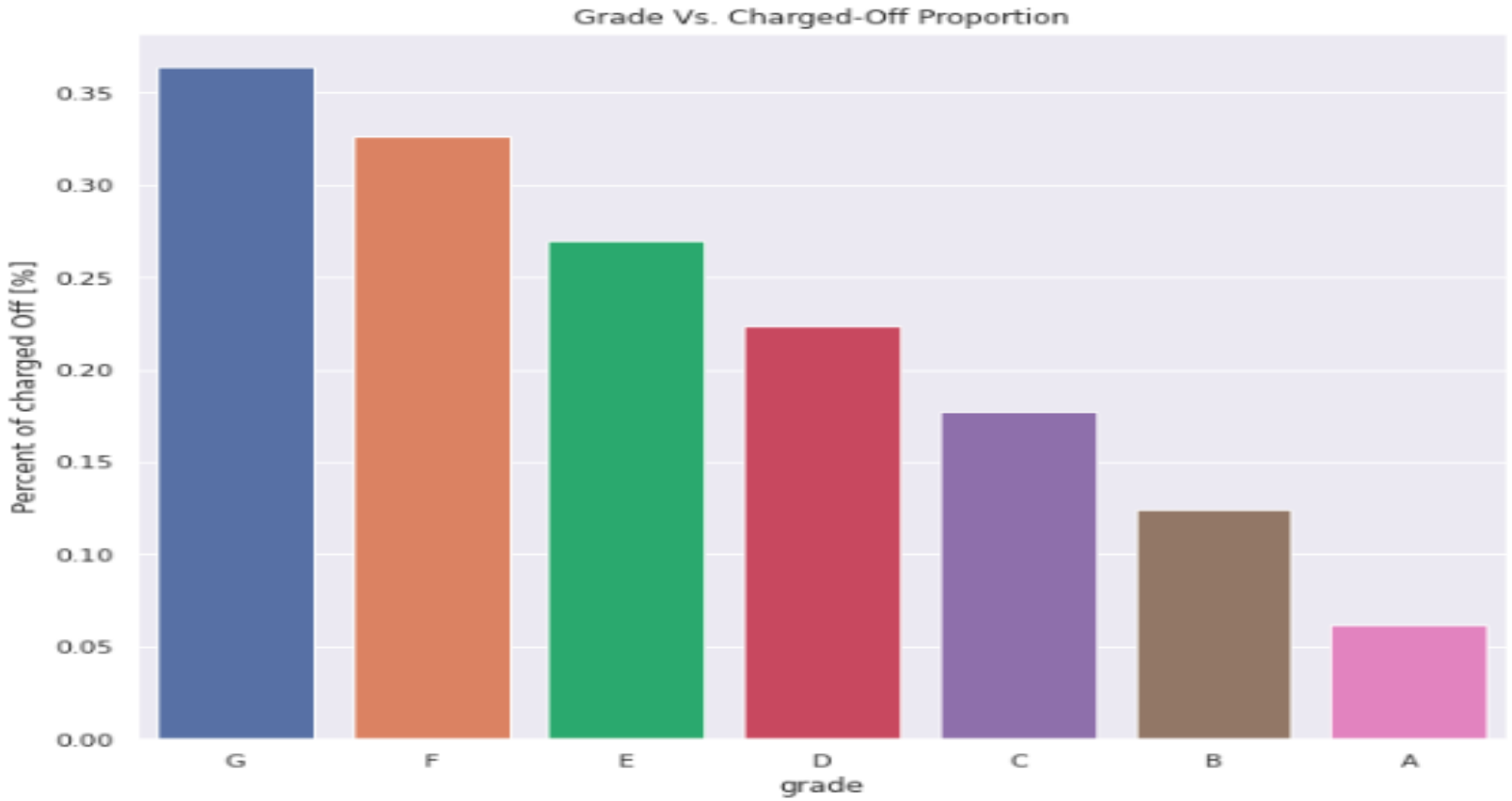
1. Annual Income to Debt To Income Ratio i.e. dti are negatively correlated
2. Loan Amount,Investor Amount and Funding Amount are strongly correlated
3. Positive correlation between Annual Income and employment years
4. Positive correlation between annual income and funded amount that means people with high income gets high funded amount

Bivariate Analysis – Term In Month's Vs Charged Off

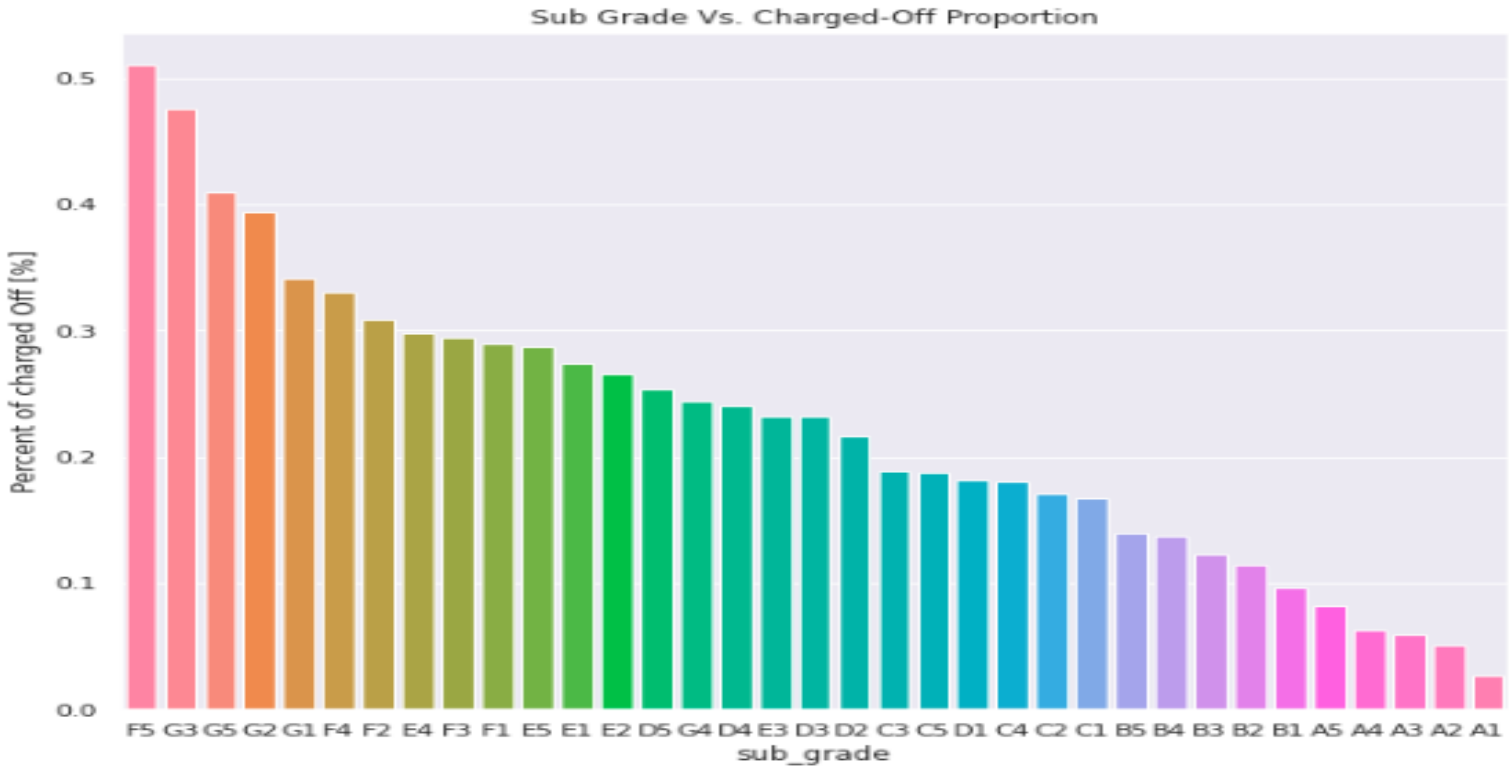


From the above plot it is clear that Loan applicants with 60 months tenure has highest number of Defaults

Bivariate Analysis – Grade Vs Charged Off

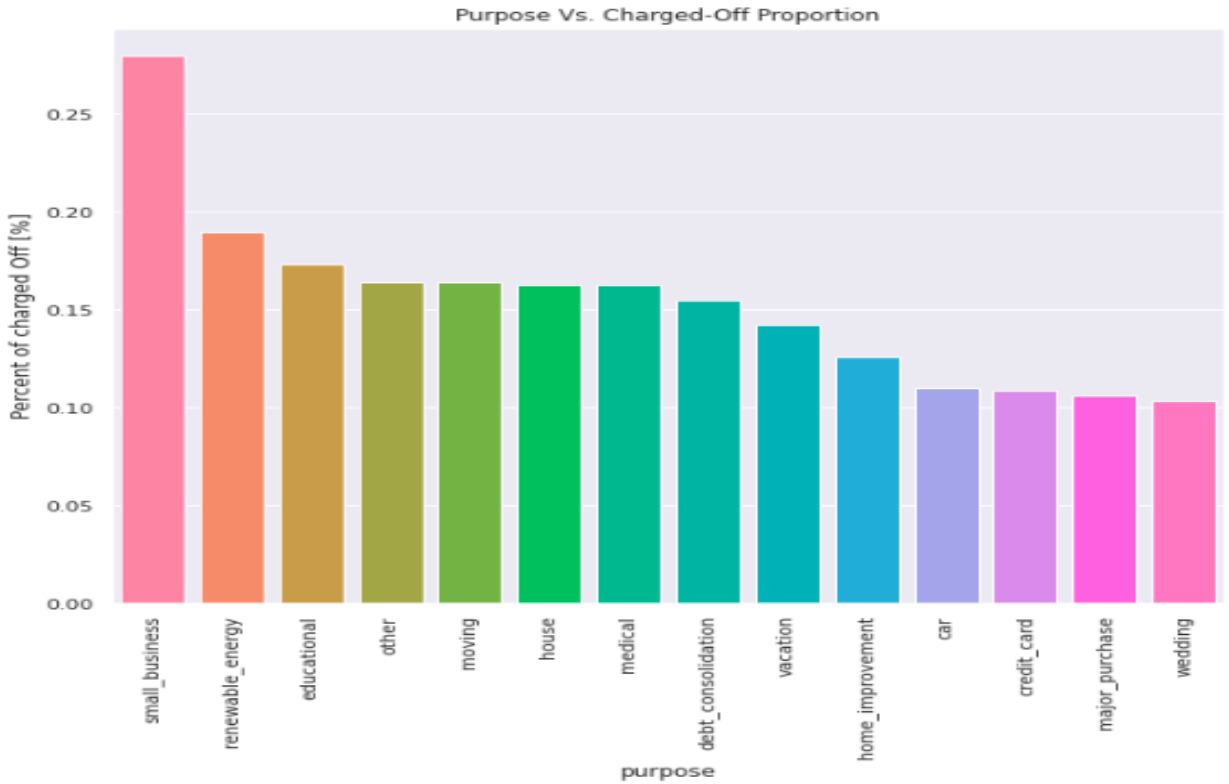


Bivariate Analysis –Sub Grade Vs Charged Off



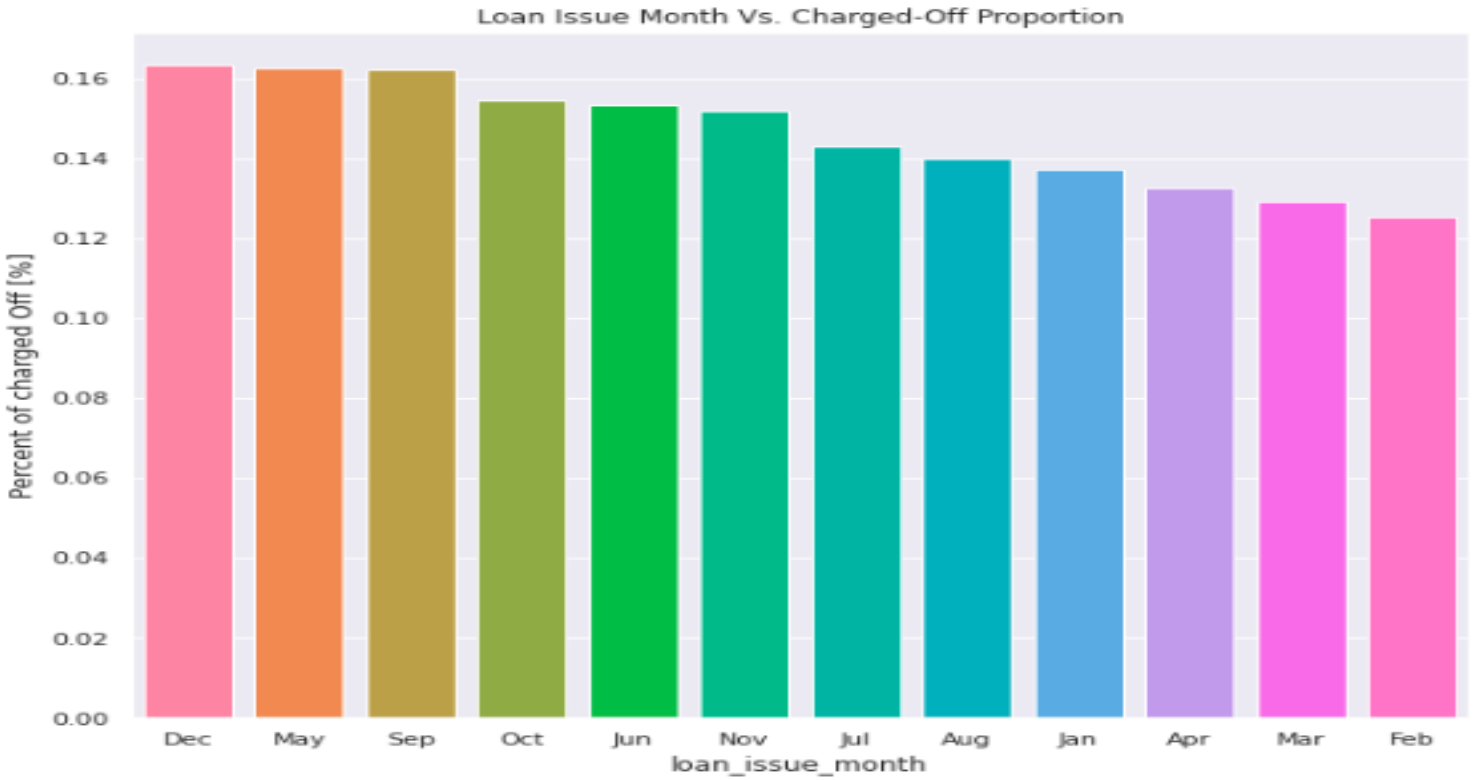
From The above Figures of Grade and Subgrade it is evident that Lower Graded loans Like G, F and E has highest Loan Default percentage

Bivariate Analysis – Purpose Vs Charged Off



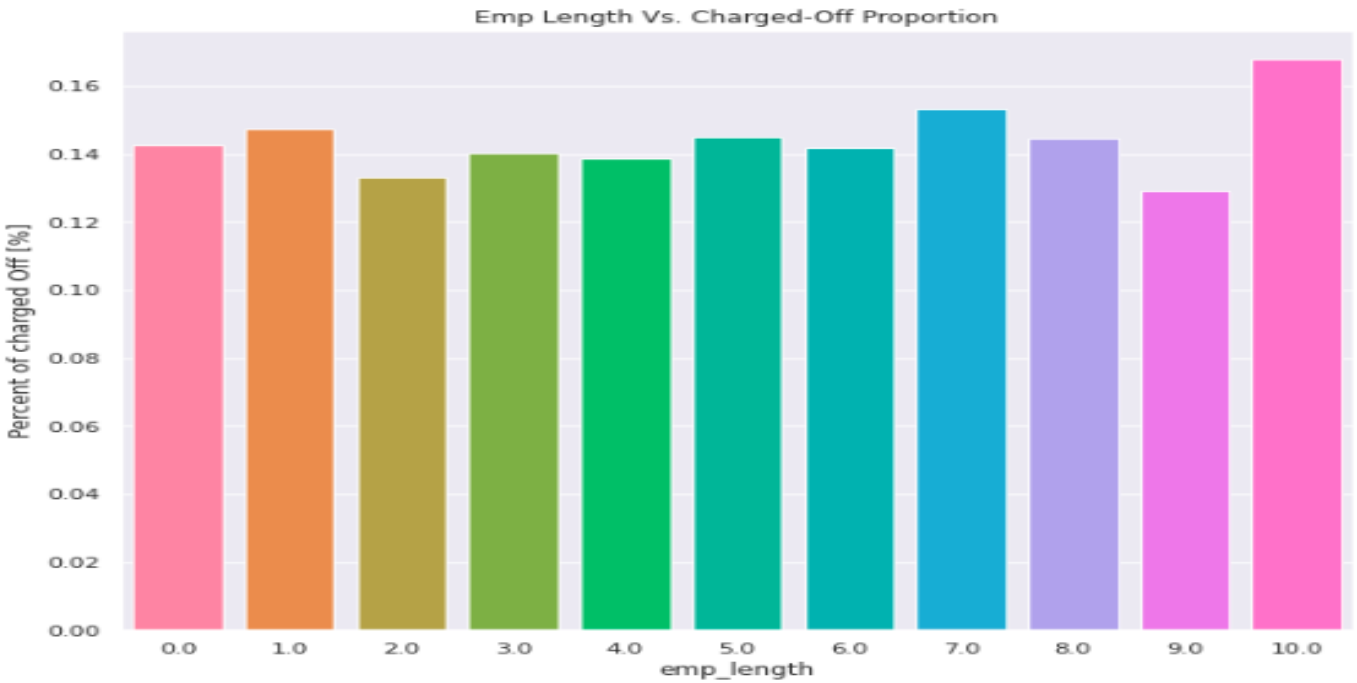
From the above figure it is evedient that Small Business applicants have high chances of getting charged off. Wedding has better chanrged off proportion as compare to other categories.

Bivariate Analysis – Loan Issue Month Vs Charged Off



Most of the Loan applicants defaults during the Holiday season of December

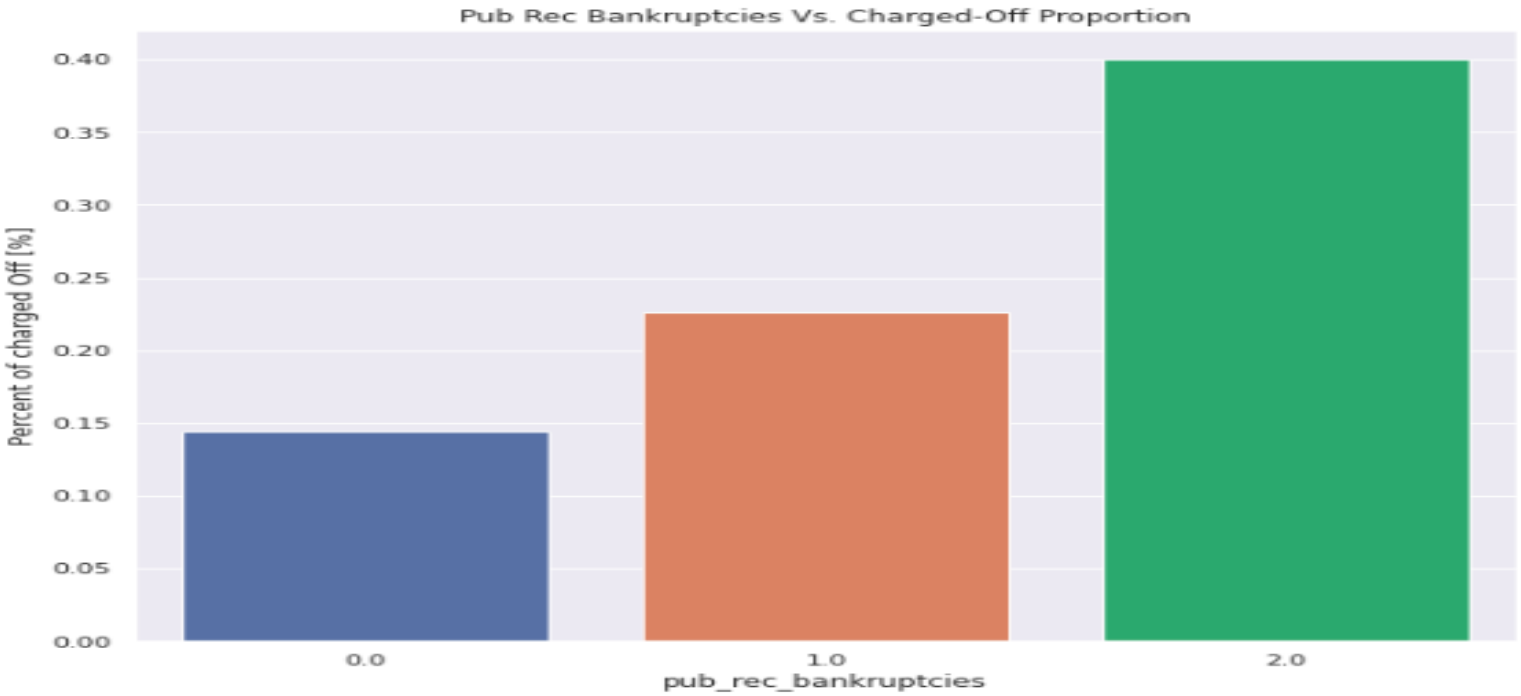
Bivariate Analysis – Emp Length Vs Charged Off



Below Observations can be made from above figire

1. Those who are not working or have less than 1 year of work experience have high chances of getting charged off.
2. It makes sense as with less or no experience they don't have source of income to repay loan.
3. Employees with more than 10 years of experience has also Highest default rate. This might be due to starting of any small business which might not be profitable

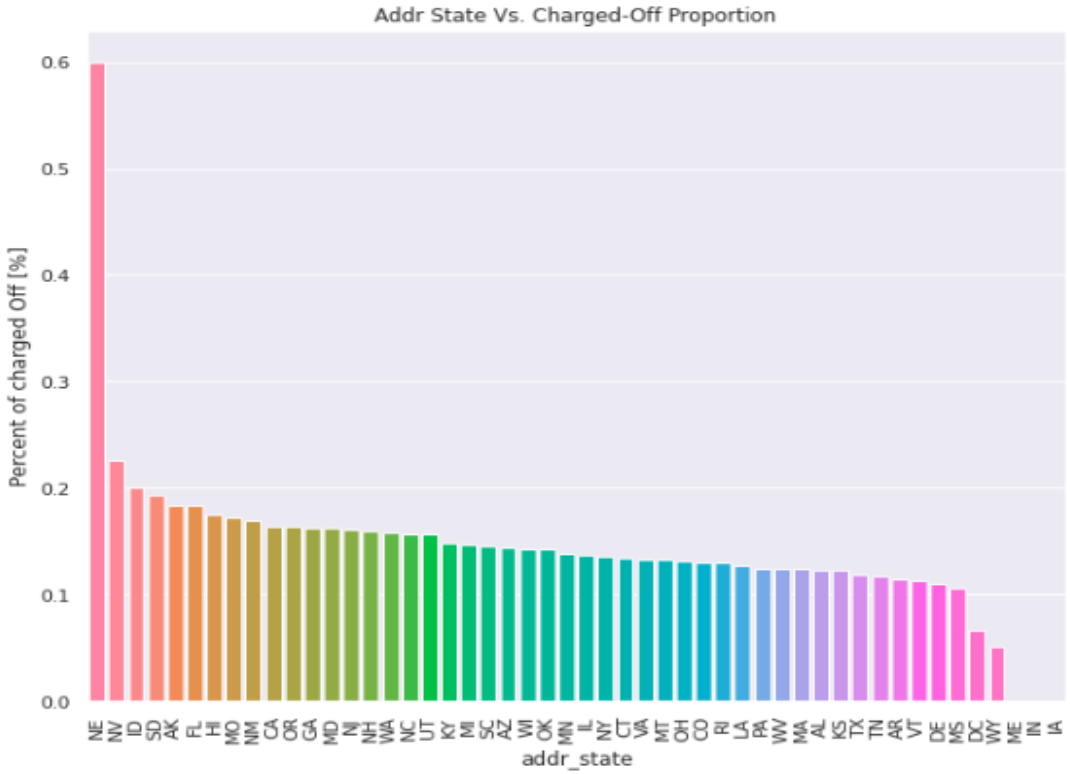
Bivariate Analysis –Public Bankruptcies Vs Charged Off



Below observations can be made from above figure

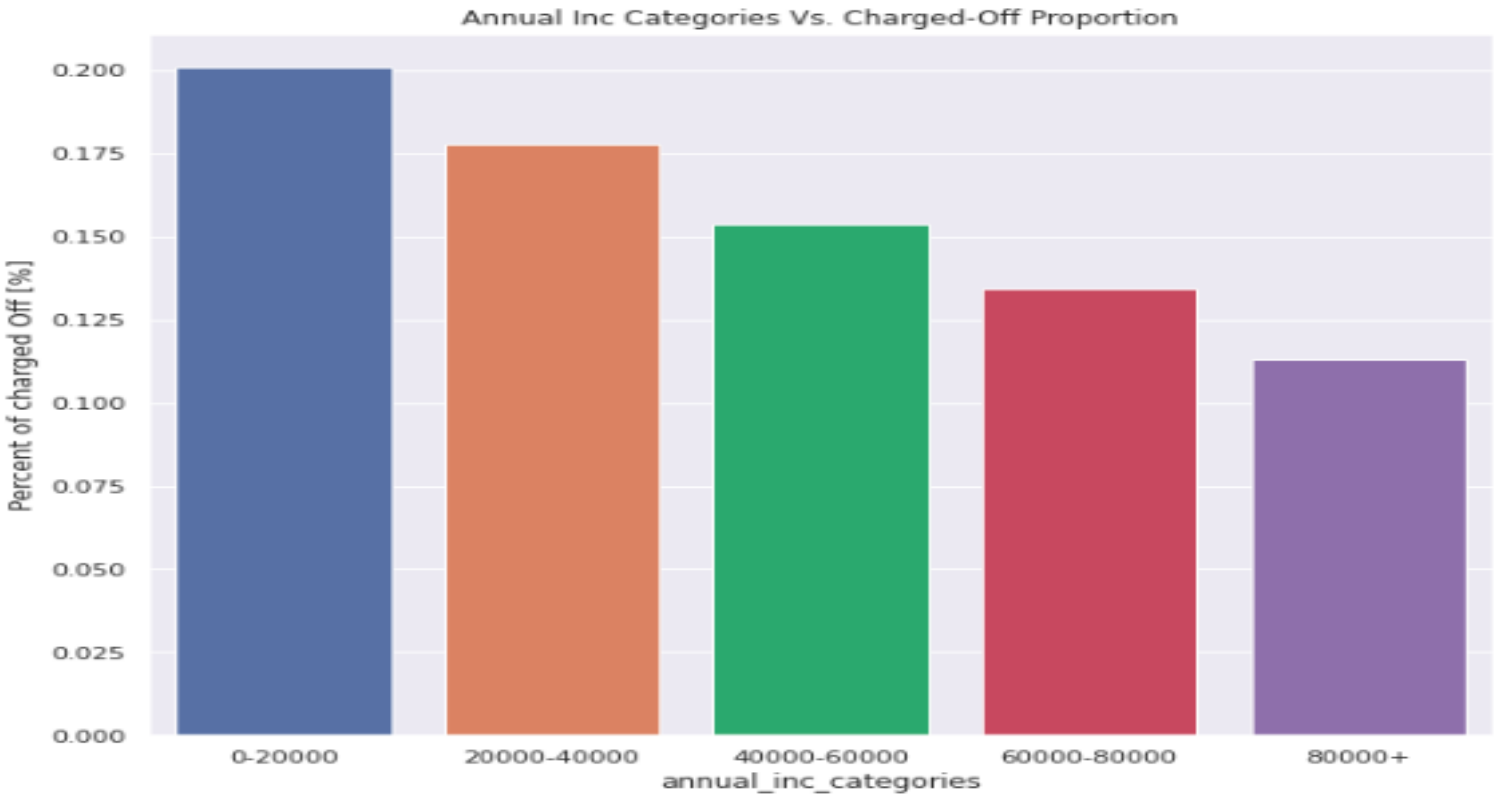
1. Those who already have pub_rec_bankruptcies value 1, have charged off proportion higher than who have no pub_rec_bankruptcies.
2. pub_rec_bankruptcies count 2 has even higher charged off proportion but those numbers are not significant to decide.

Bivariate Analysis –State Vs Charged Off



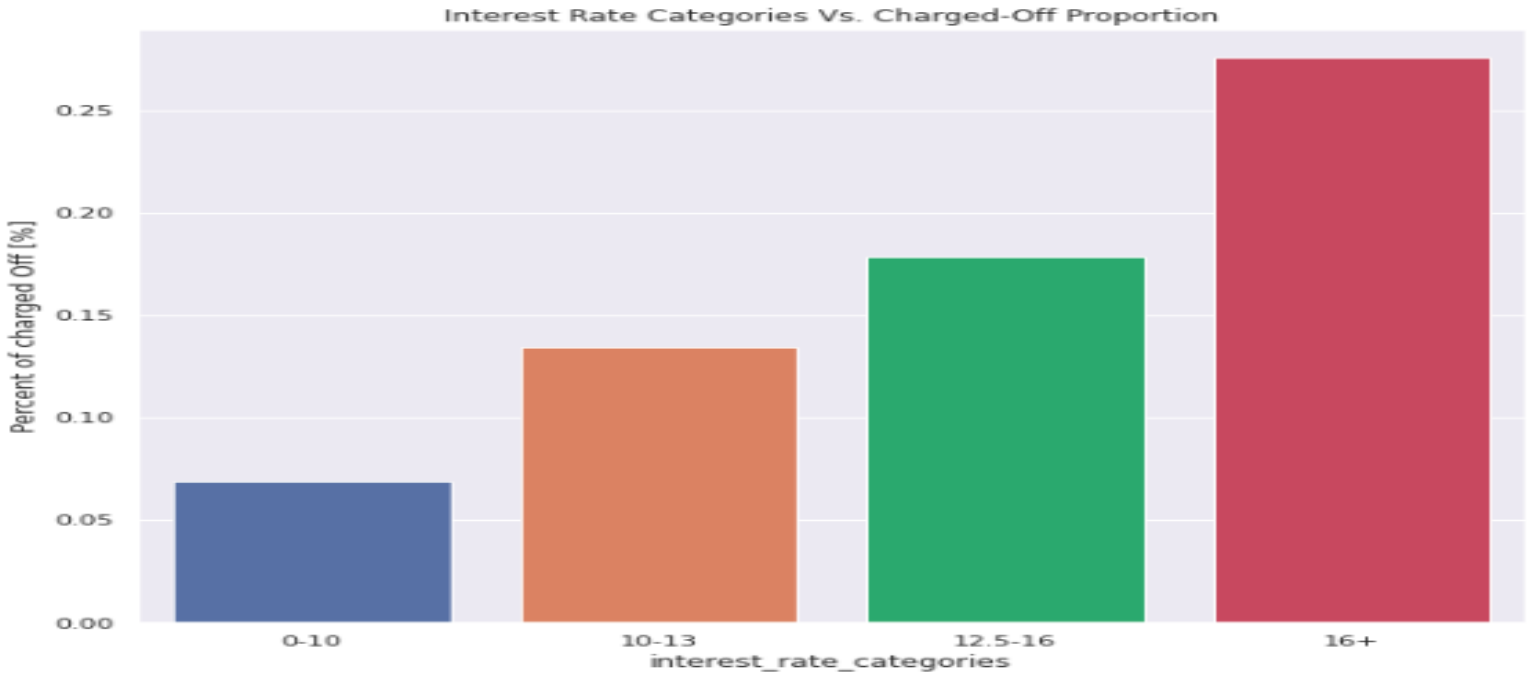
States NE has very high chances of charged off but number of applications are too low to make any decisions. States NV,CA and FL states shows good number of charged offs against good number of applicants

Bivariate Analysis – Annual Inc Vs Charged Off



Income range 80000+ has less chances of charged off.Income range 0-20000 has high chances of charged off.

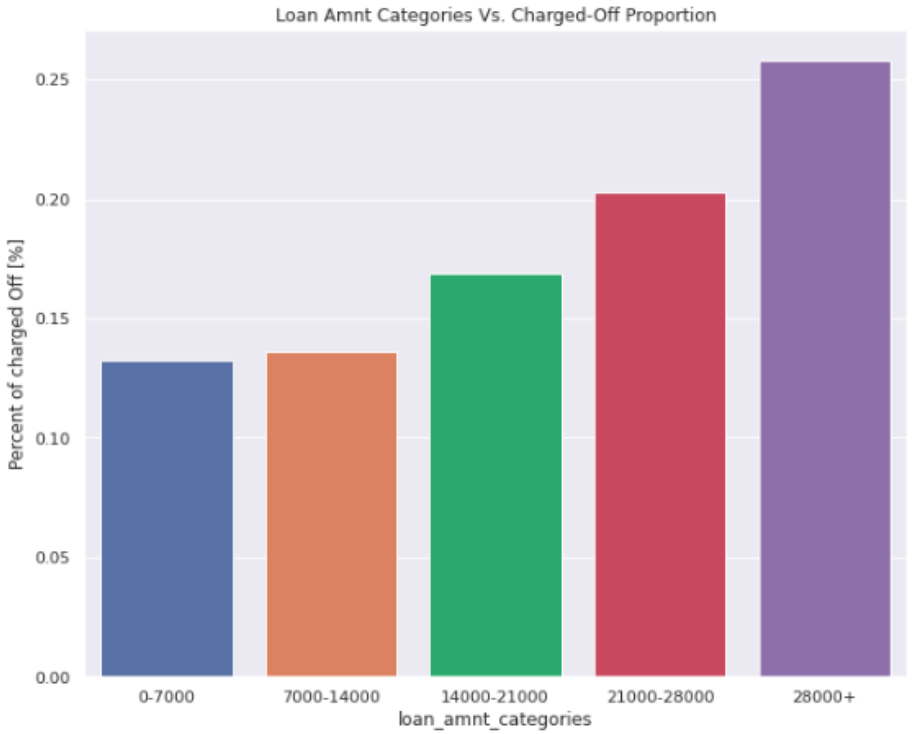
Bivariate Analysis –Interest Rate Vs Charged Off



Below observations are made from above plot

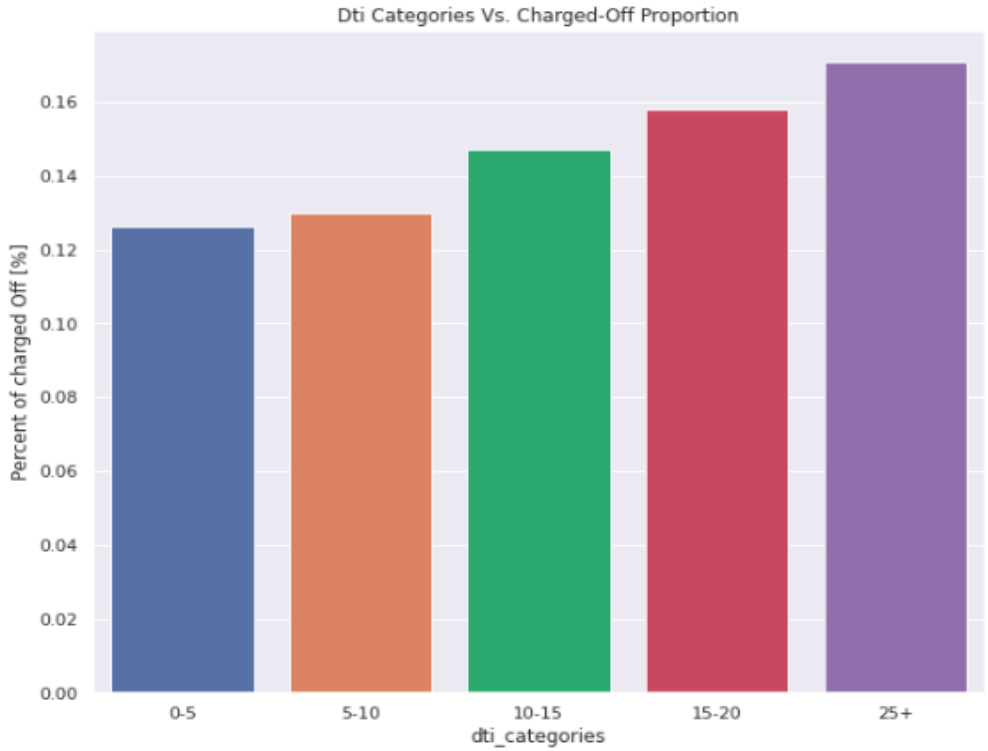
1. Interest Rates which are less than 10% have very less chances of charged off.
2. Interest Rates greater than 16 has the highest chances of charged off.
3. Charged Off Proportion increases with higher interest rates

Bivariate Analysis –Loan Amount Vs Charged Off



From the above figure, we can conclude that when the loan amount is high, the risk of defaulting also increses. Capping the loan amounts < 20K USD can potentially reduce default risk at the surface but could spread the default risk to lower categories of loan amounts

Bivariate Analysis – Diti Vs Charged Off



From the above figure, we can conclude that when the Debt to Income Ratio is high, the risk of defaulting also increses. It is wise to provide loans to candidates who has Debt to Income ratio of less than 15.0

Recommendations

Major Factors Influencing Loan Defaults

1.Higher Interest rate
2.Higher Loan amount
3.Lower Annual income
4.More than 1 Public record of bankruptcies
5.Unverified Verification status
6.Home ownership status for "others", "rent" and "mortgage" categories
7.Lower Grade Loans
8.Higher Loan Tenure
9.Top 5 states of applicant pool
10.Small Business and renewable Energy Business.

Our Recommendations to Lending Club:

1.Prioritize high grade loans.
2.Scrutinize purpose, state and public bankruptcy record.
3.Cap loan amounts beyond >20K where the charge off is higher.
4.Offer reduced repayment tenure less than 60 months.
5.Ensure verification is complete for all loans disbursed.
6.Provide loans to employees with experience between 1 - 10 years.
7.The sub grading system needs improvement to accurately show risk of default
8.Disbursement of Year end or holiday season loans must be avoided.

Thank You