

LENDING CLUB CASE STUDY

Group Members:

Yogesh Ghate

Arti Saraswat

CONTENT

- About Lending Club
- Problem Statement
- Find columns having more than 60% null values
- Fix missing Values
- Standardize Values
- Removing Outliers
- Univariate Analysis on Continuous Variables
- Univariate Analysis on Categorical Variables
- Segmented Univariate on Categorical Variables
- Segmented Univariate on Continuous Variables
- Bivariate on Continuous Variables
- Bivariate on Categorical Variables
- Derived Metric
- Recommendations based on Univariate Analysis
- Recommendations based on Bivariate Analysis

ABOUT LENDING CLUB :

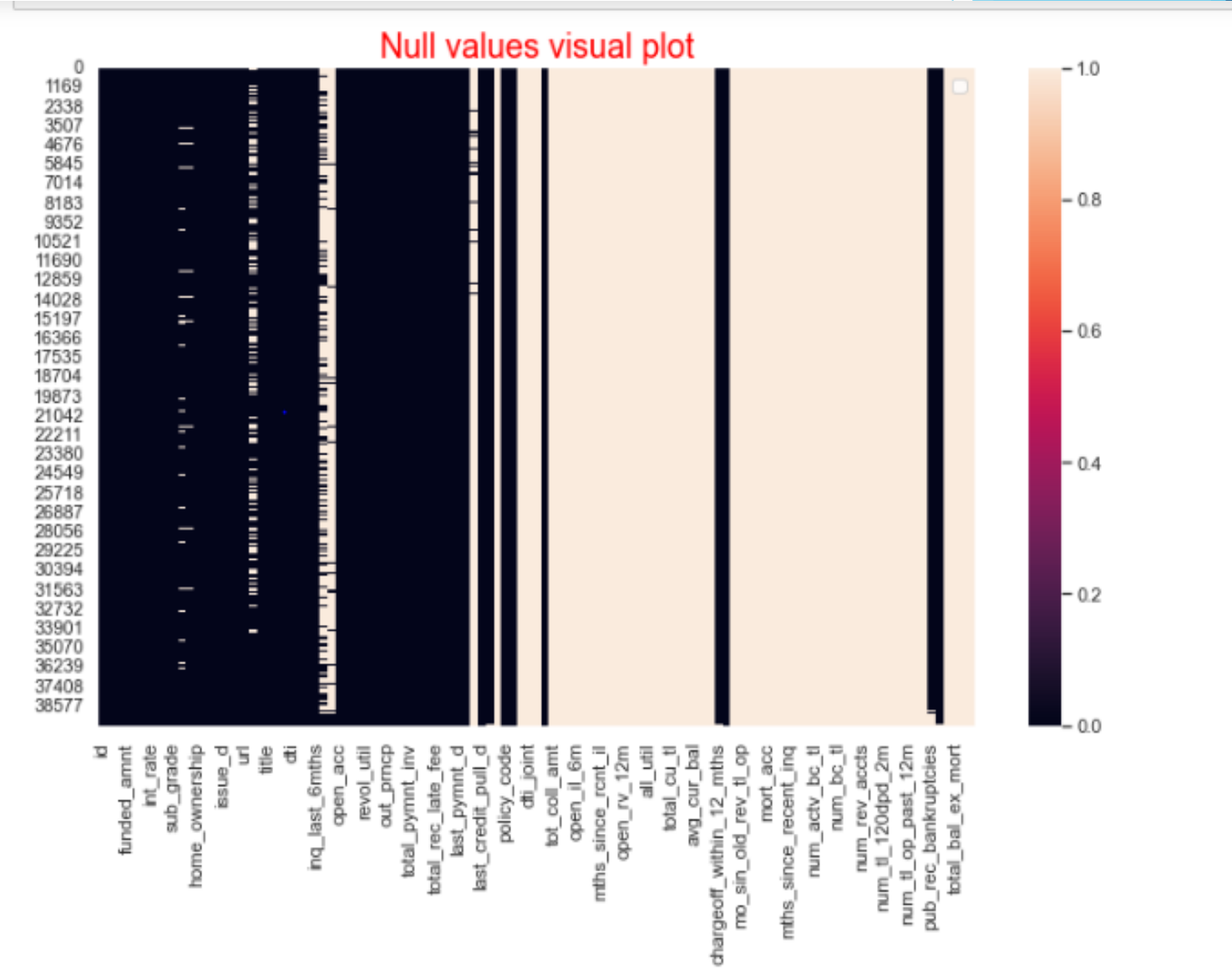
- Lending Club is a peer-to-peer lending company headquartered in San Francisco, California.
- It was the first peer-to-peer lender to register its offerings as securities with the Securities and Exchange Commission (SEC), and to offer loan trading on a secondary market.
- The company claims that \$15.98 billion in loans had been originated through its platform up to December 31, 2015

PROBLEM STATEMENT

- The data given 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 by understanding how consumer attributes and loan attributes influence the tendency of default.
- When a person applies for a loan, there are two types of decisions that could be taken by the company:
 - Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:
 1. Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
 2. 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'.
 3. 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
 - Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements)

Finding columns with more than 60% null values

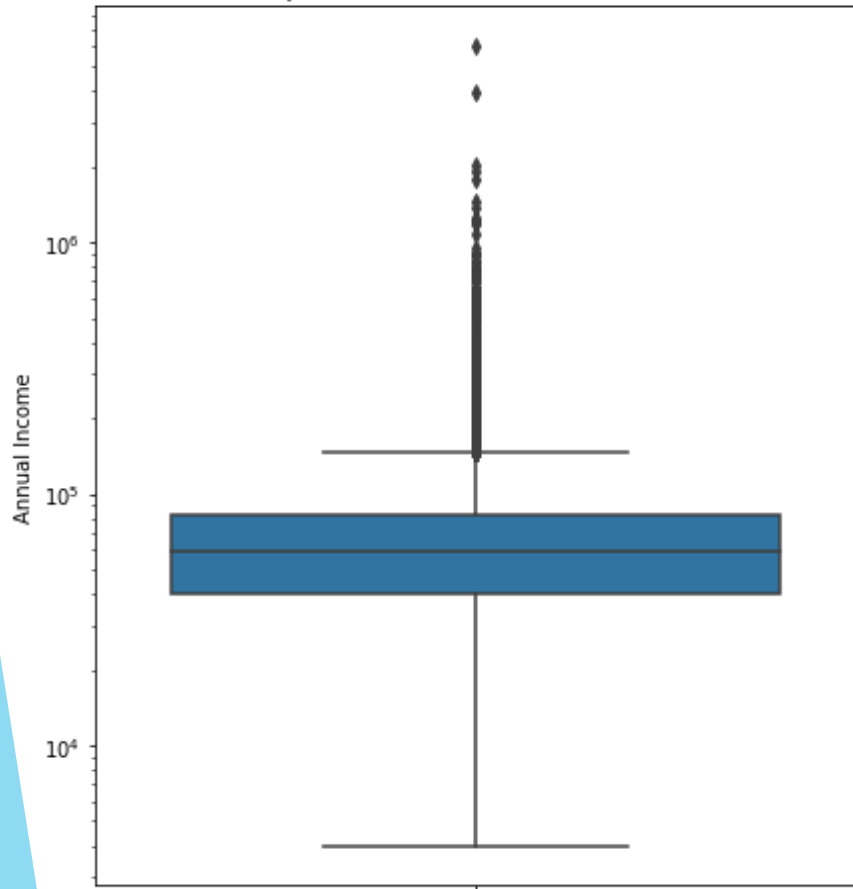
Percentage of null values		No.of columns
0	10% or less	53
1	20% or less	53
2	30% or less	53
3	40% or less	54
4	50% or less	54
5	60% or less	54
6	70% or less	55



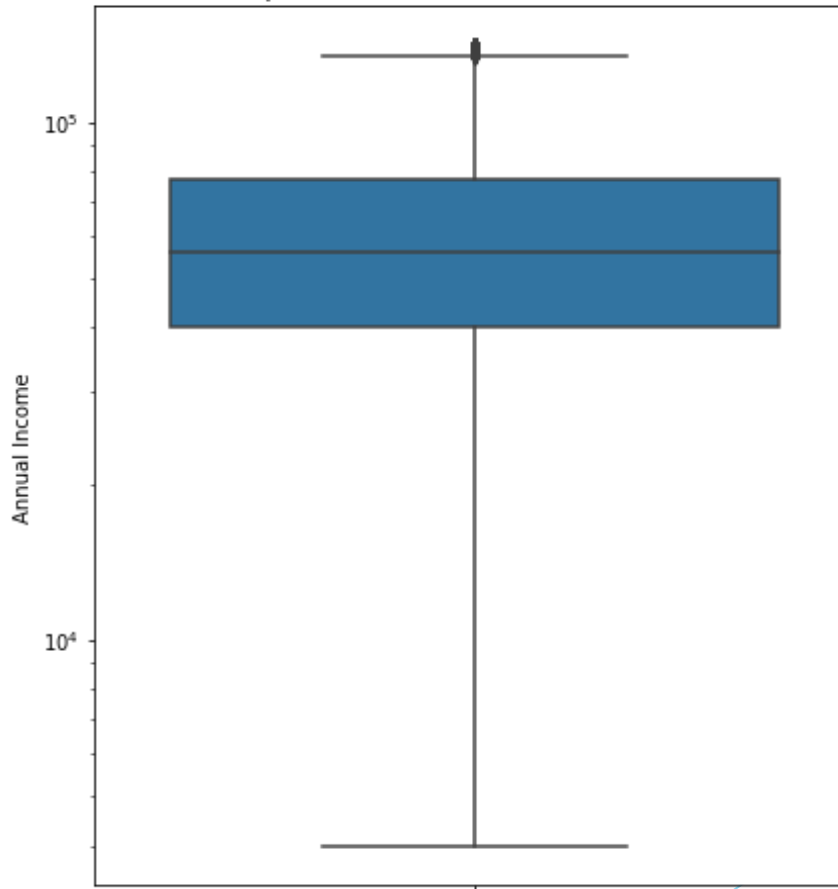
Removing Outliers

- Annual income contains a large outliers as shown in the 1st plot
- Outliers are those whose values which are falling after .95 percentile
- Data shape after dropping outliers of annual income: 36642 rows and 14 columns

Boxplot of Annual Income with Outliers



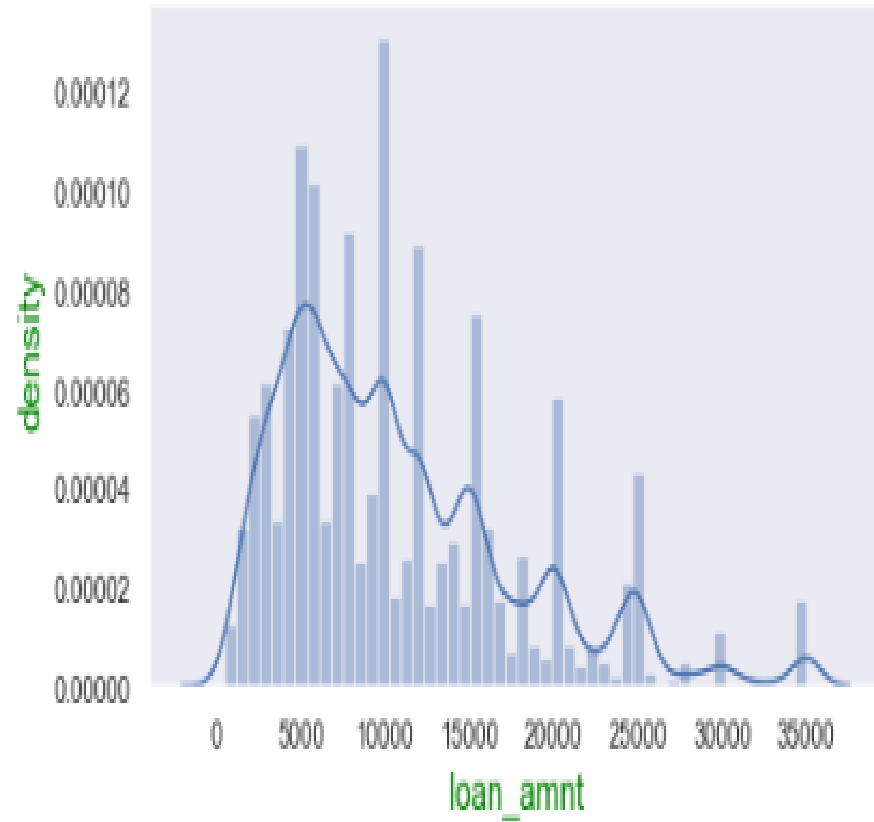
Boxplot of Annual Income without Outliers



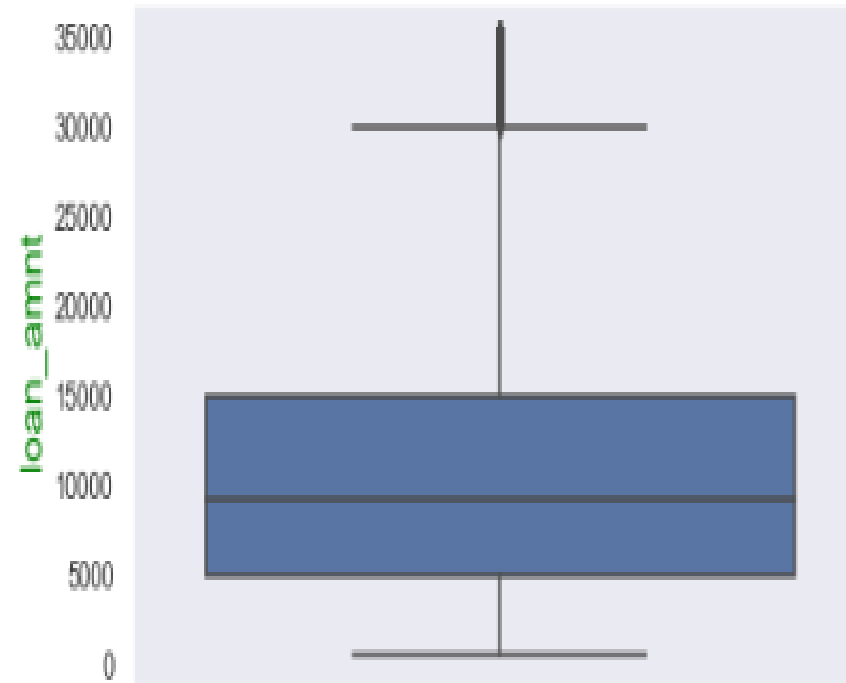
UNIVARIATE ANALYSIS ON LOAN AMOUNT

Majority of the loan amount lies between 5000 to 15000

Distribution Plot

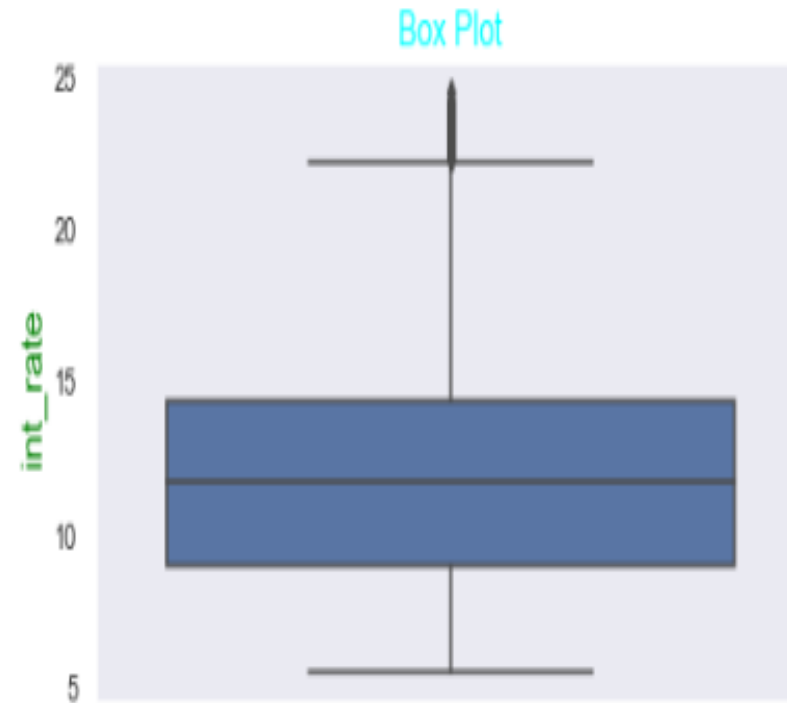
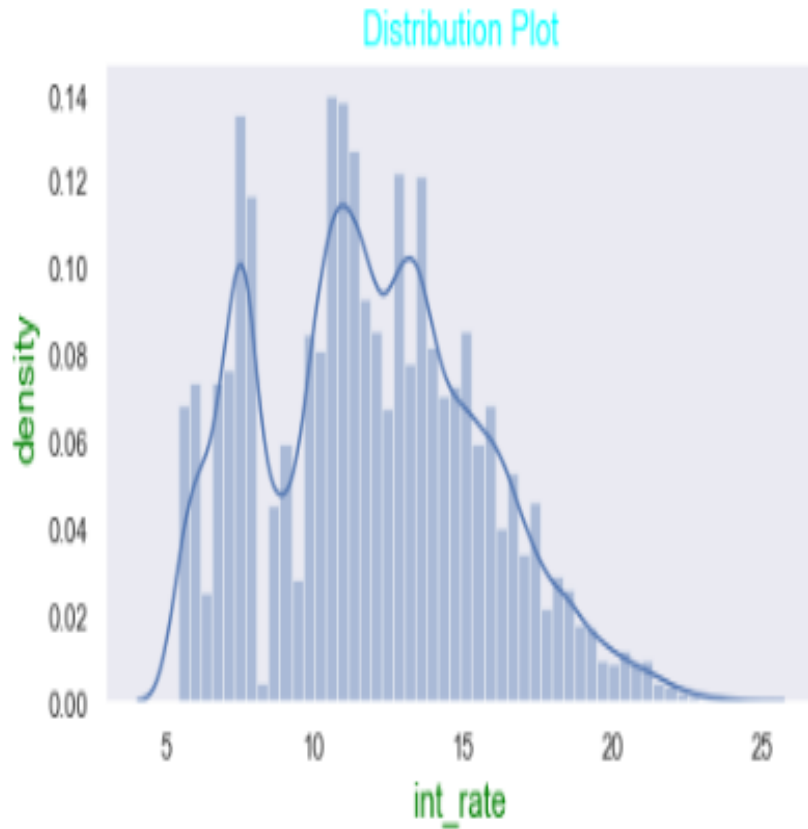


Box Plot



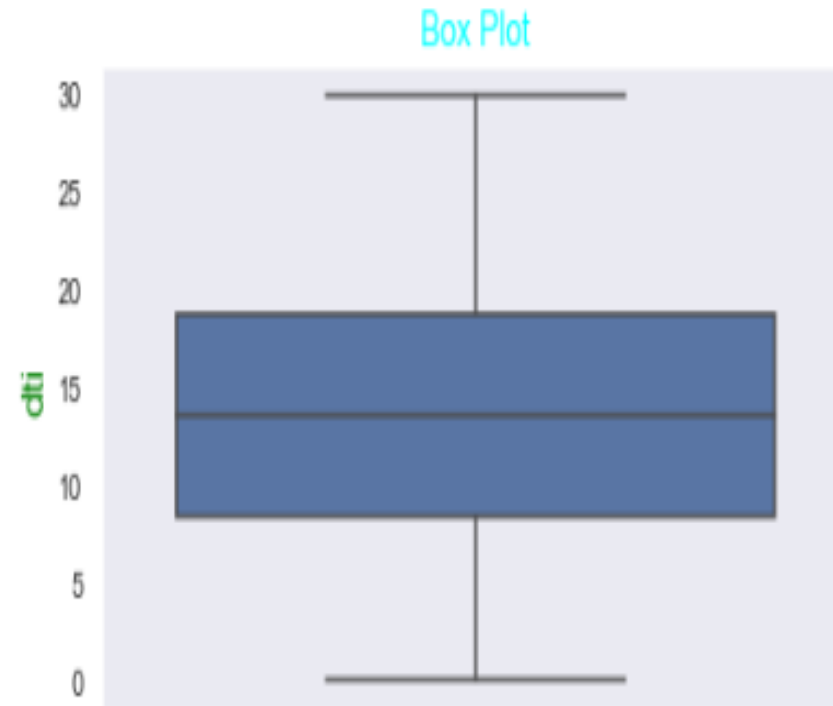
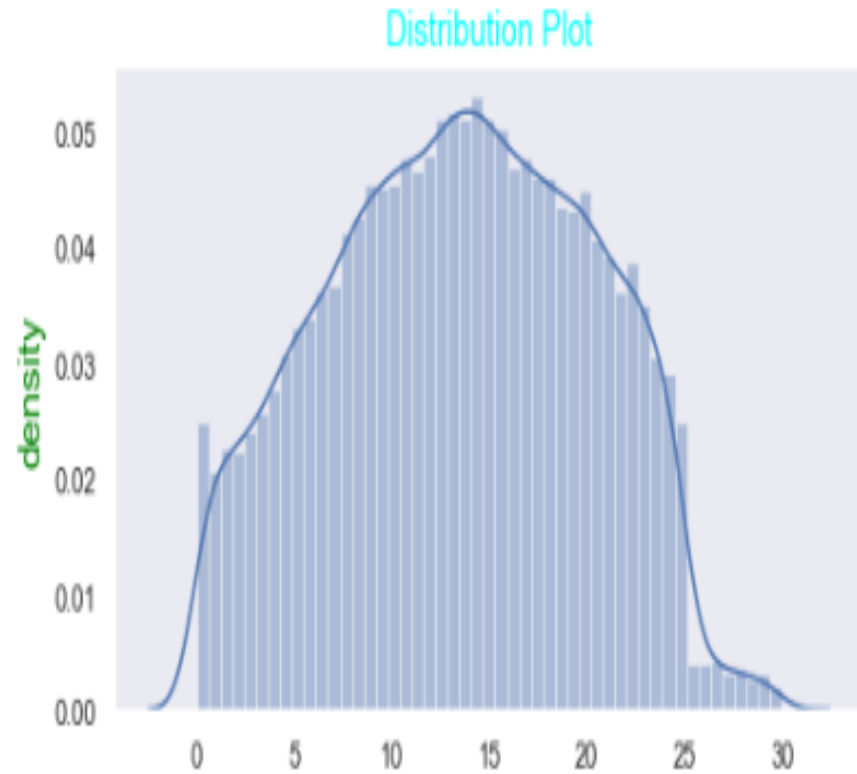
UNIVARIATE ANALYSIS ON INTEREST RATE

Majority of the loan interest rate lies in the range 8% to 12%



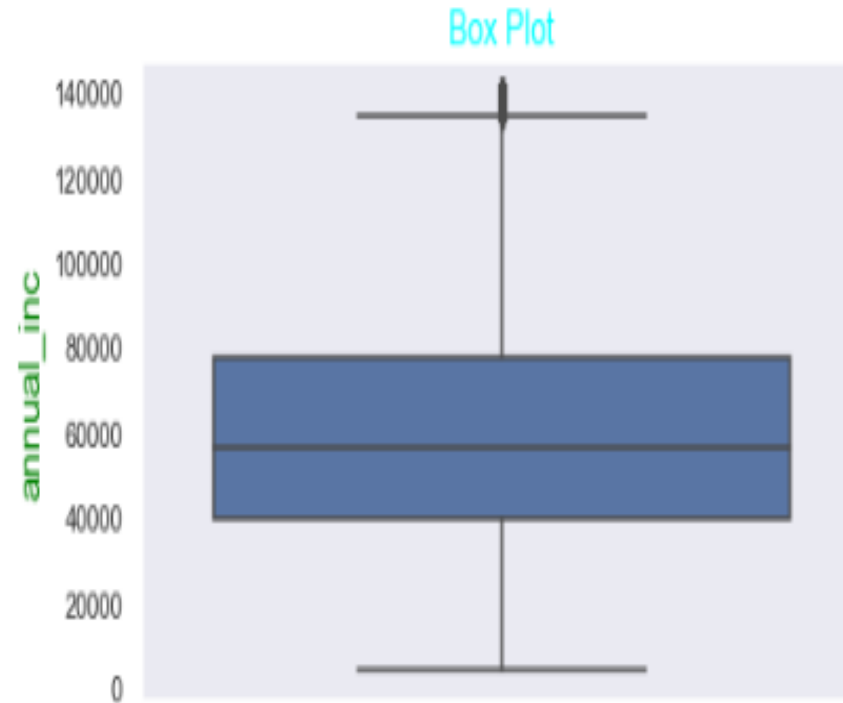
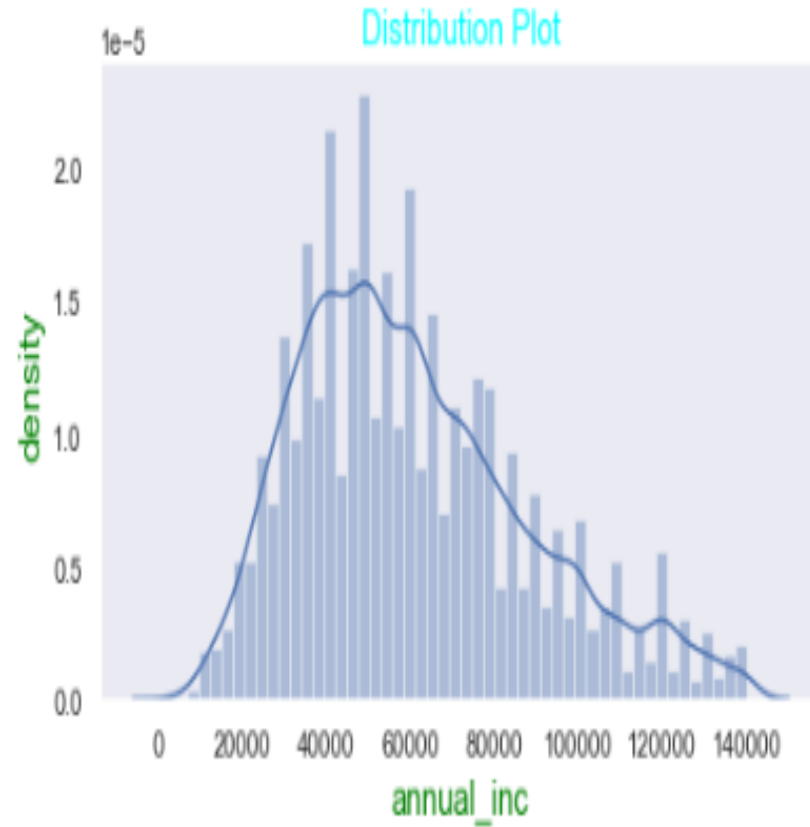
UNIVARIATE ANALYSIS ON DEBT TO INCOME RATIO

Majority of the debt to income ratio lies between 8 to 18

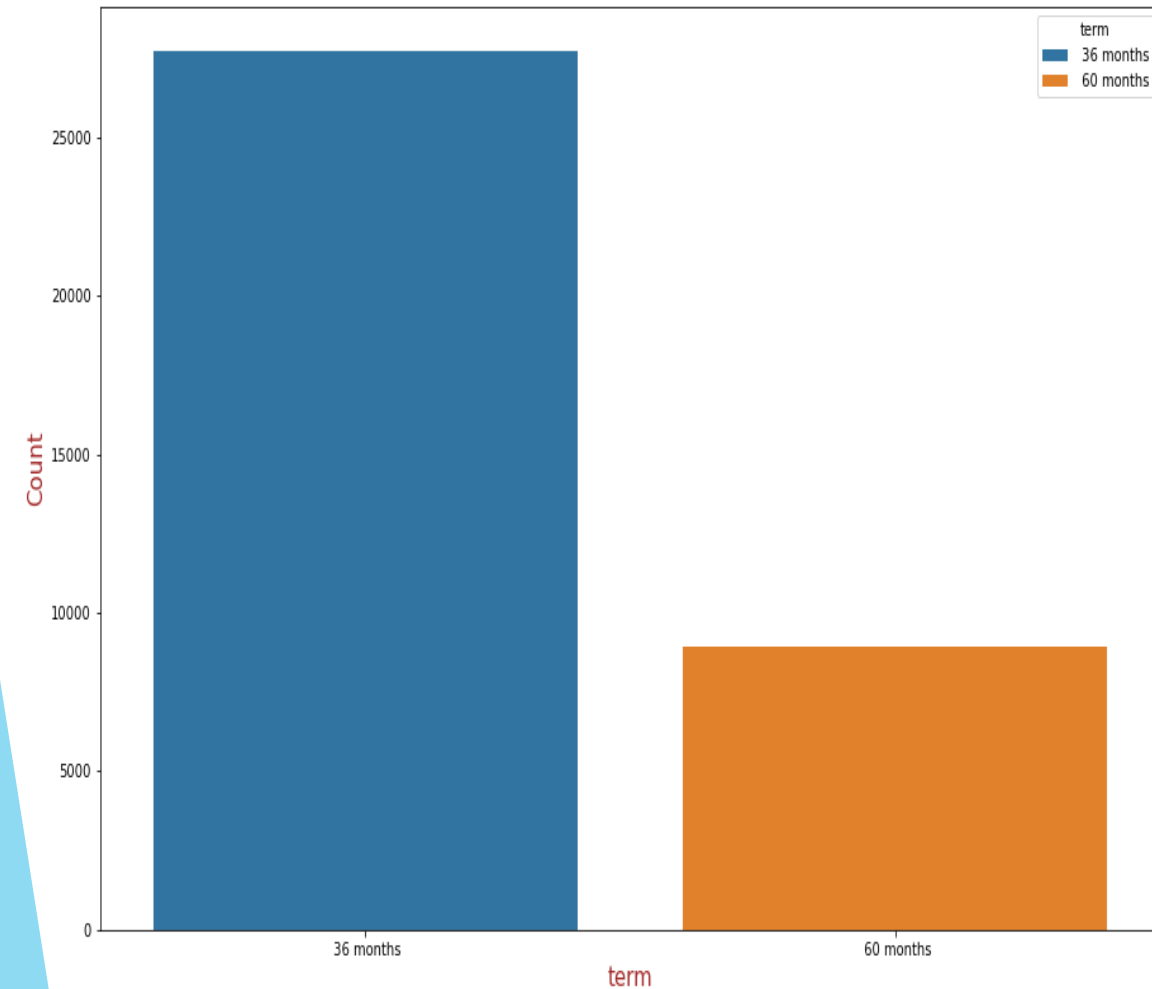


UNIVARIATE ANALYSIS ON ANNUAL INCOME

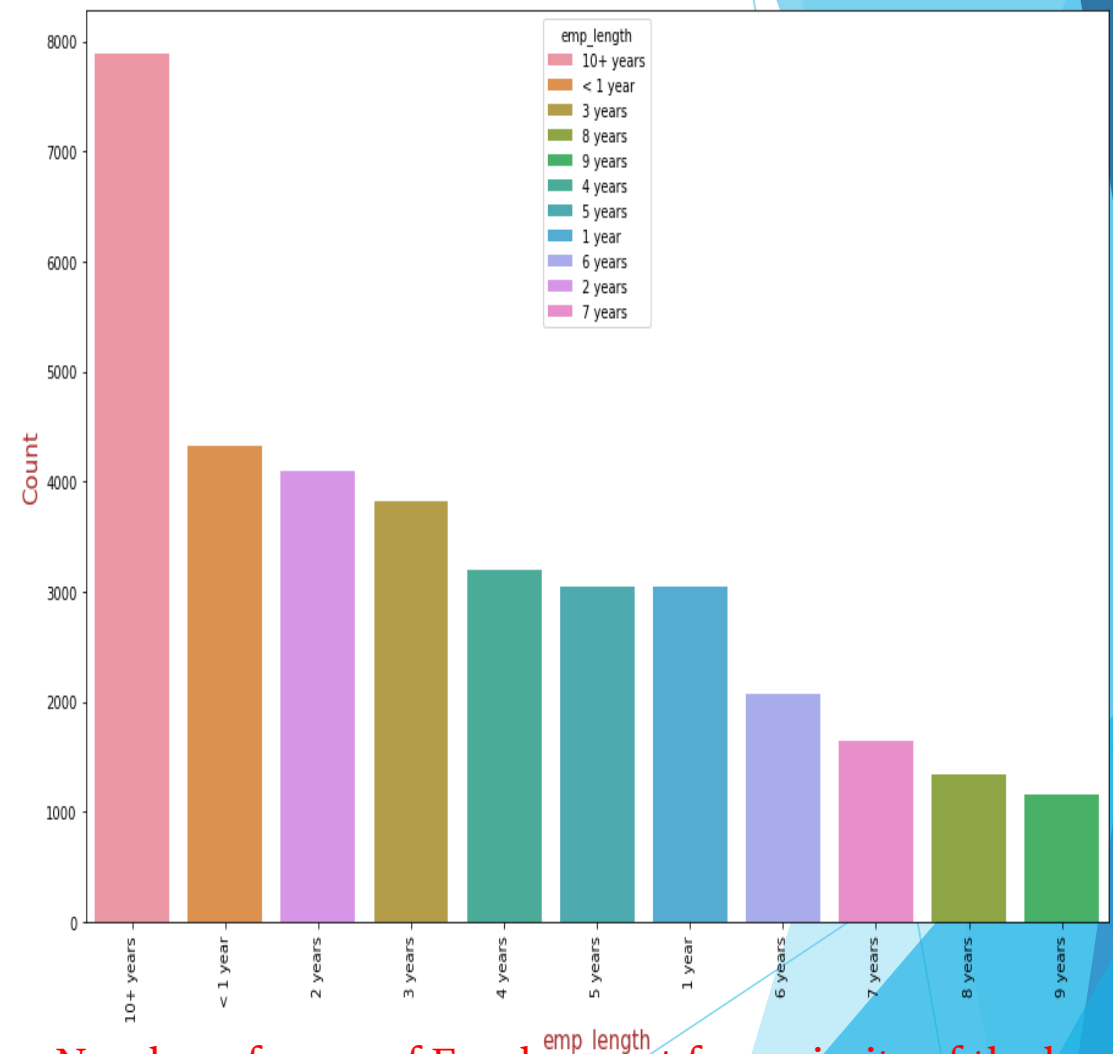
The annual income of majority of the borrowers are in the range 40000 to 80000



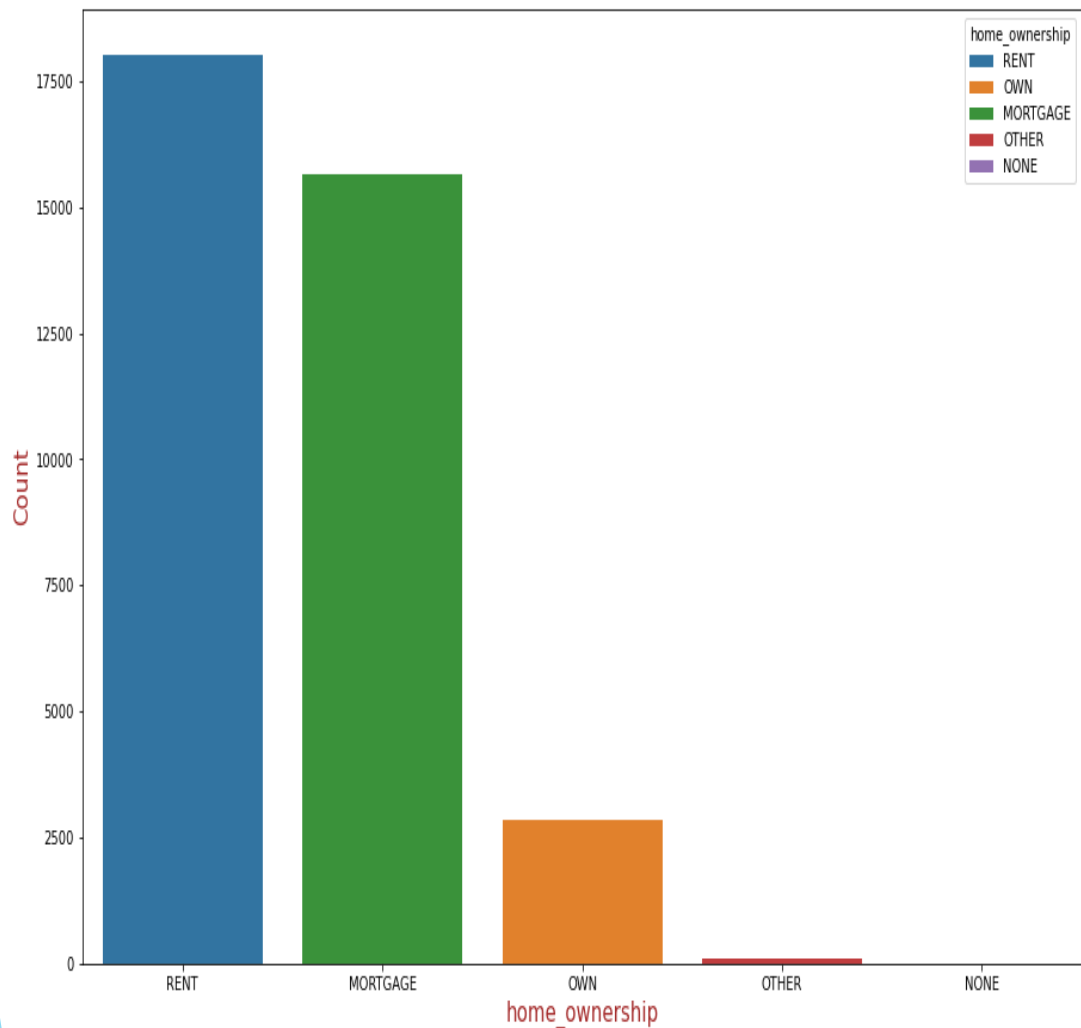
UNIVARIATE ANALYSIS ON CATEGORICAL VARIABLES



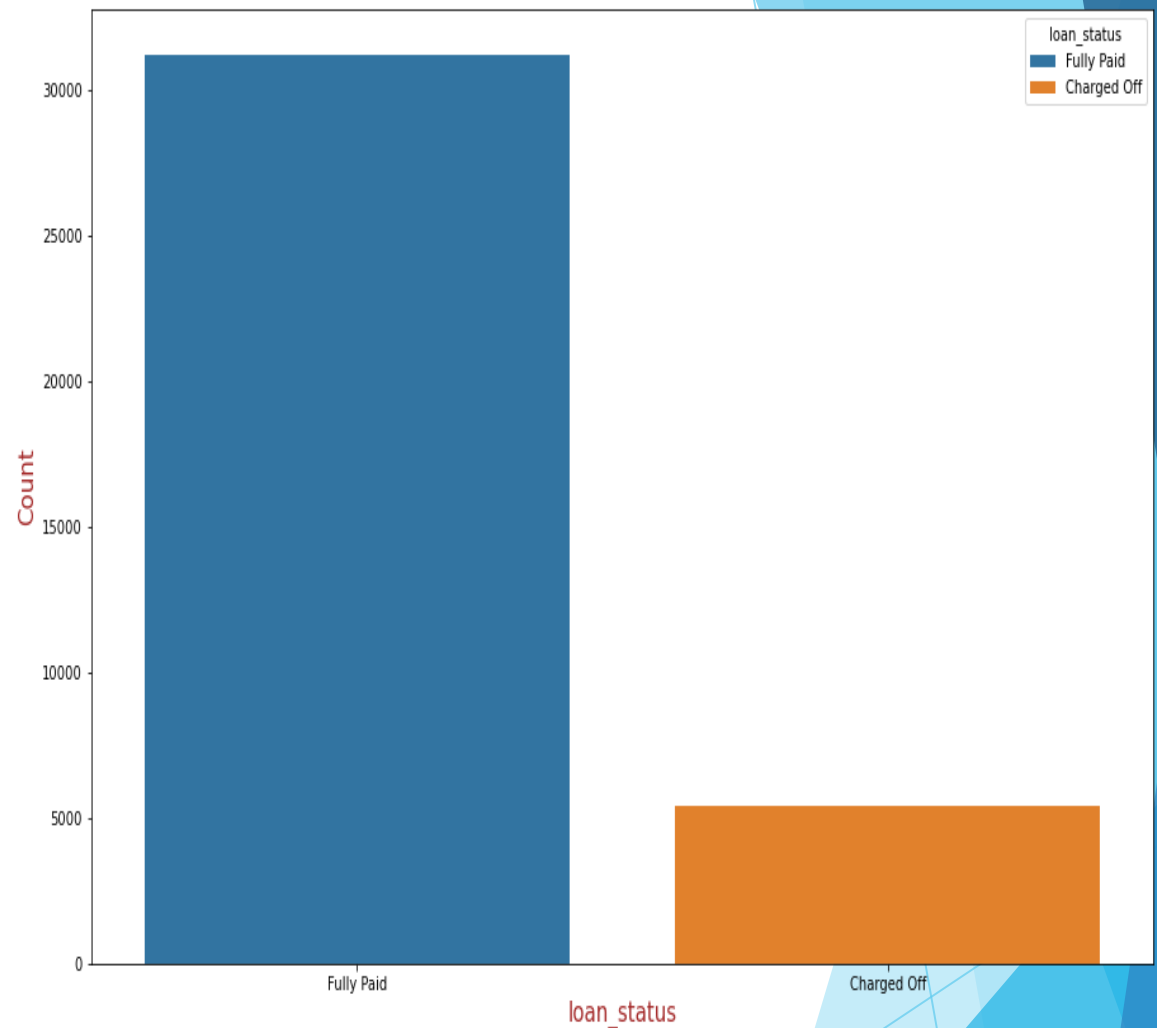
Majority of the borrowers prefer to choose term of 36 months



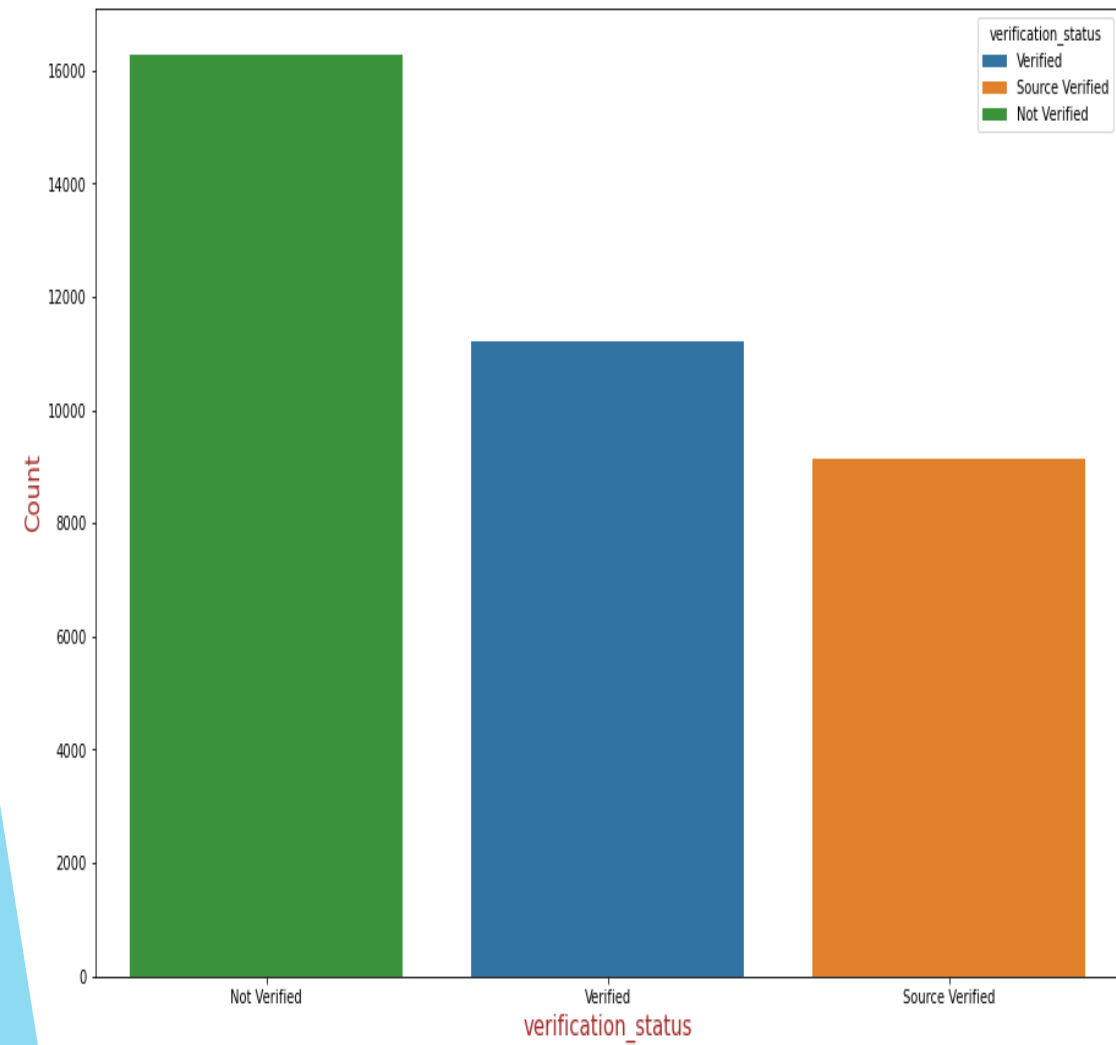
Number of years of Employment for majority of the borrowers are 10+ years



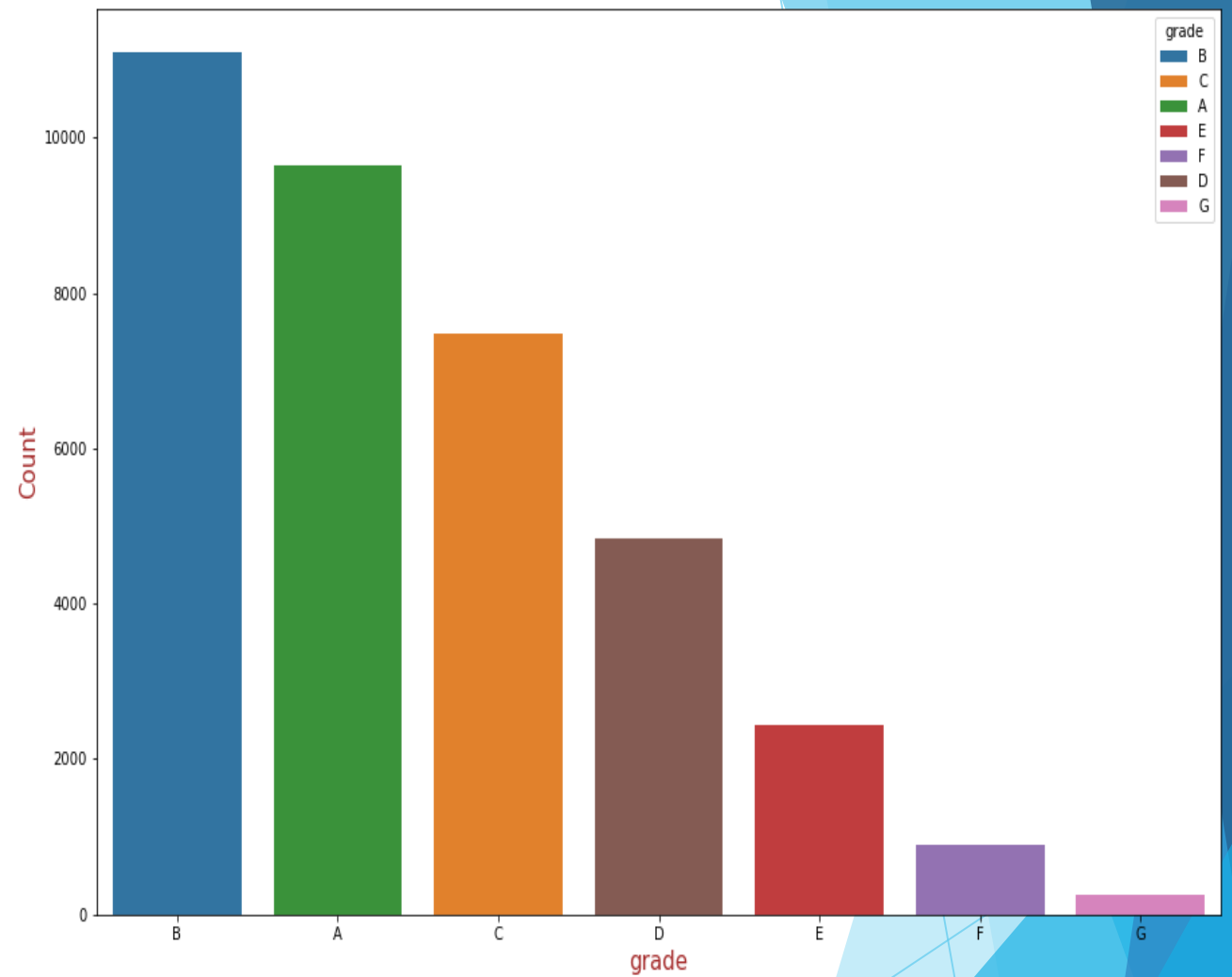
Majority of the borrowers live in a Rented home 18900 whereas 17600 borrowers have a Mortgage on their property



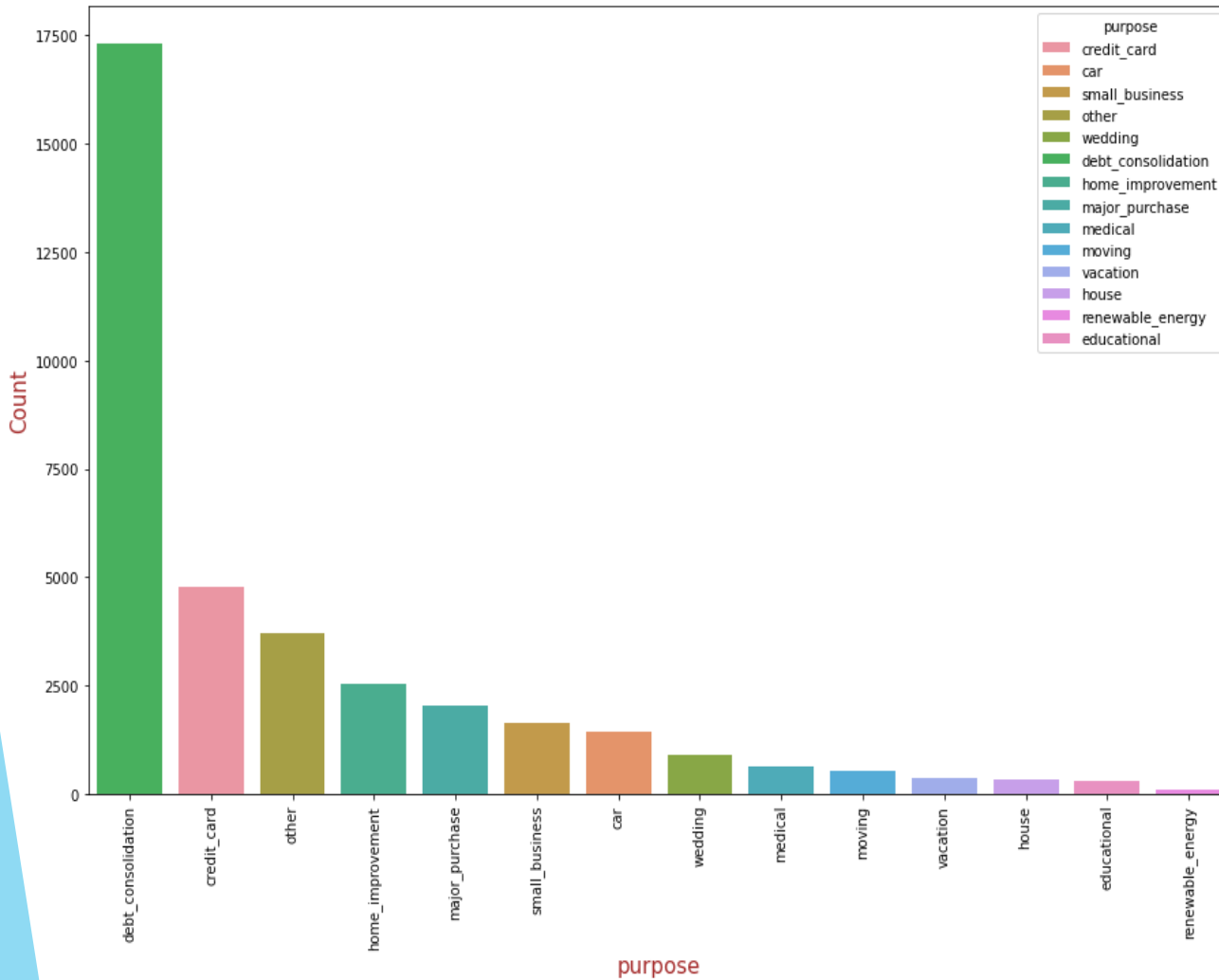
Majority of the loan status is Fully paid (33000) whereas very less borrowers have a status as Charged off (5600)



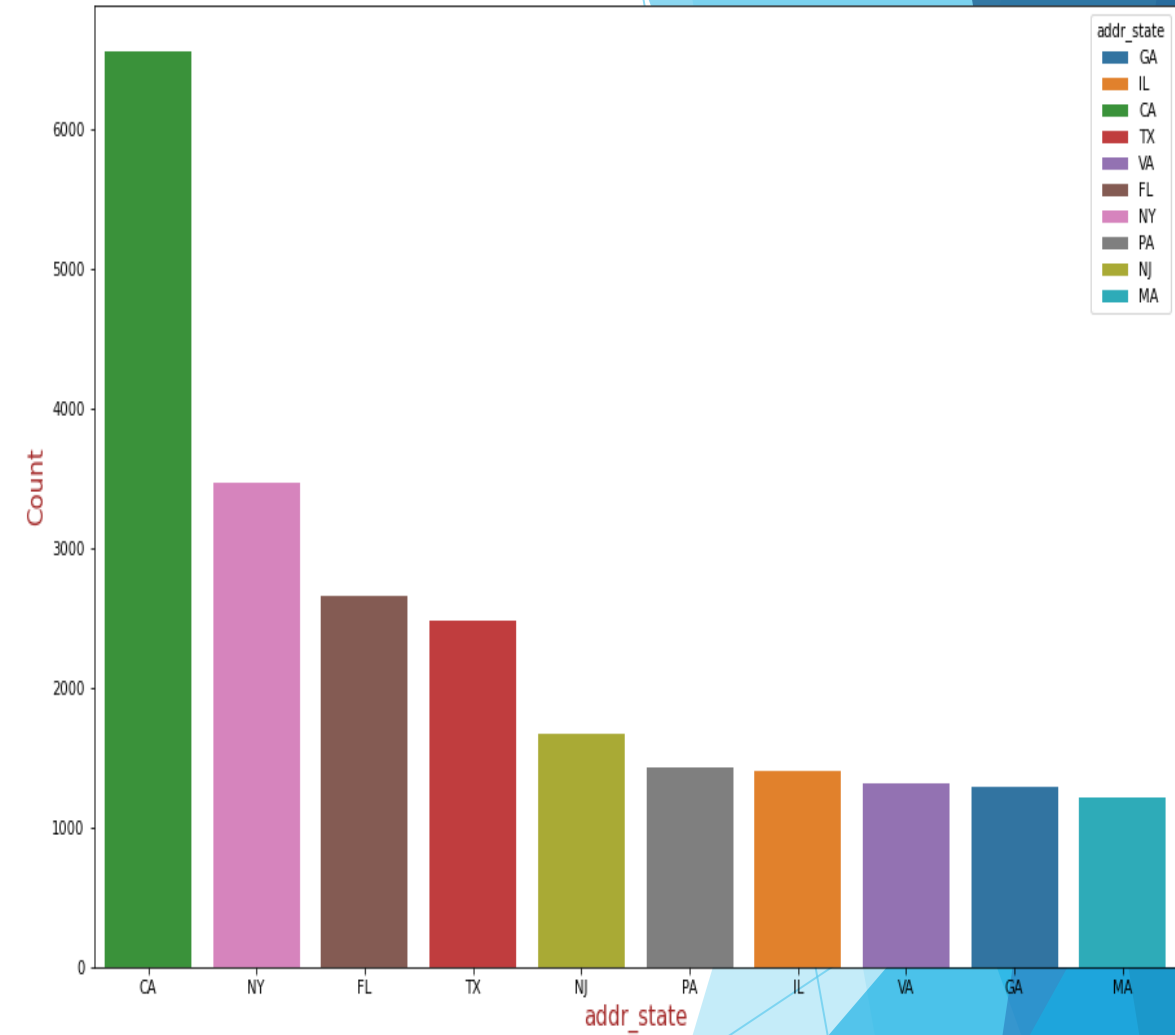
Majority of the borrower's income is Not Verified (~16930)



Most of the borrowers are categorized under Grade B (12020) whereas least are categorized as Grade G (316)

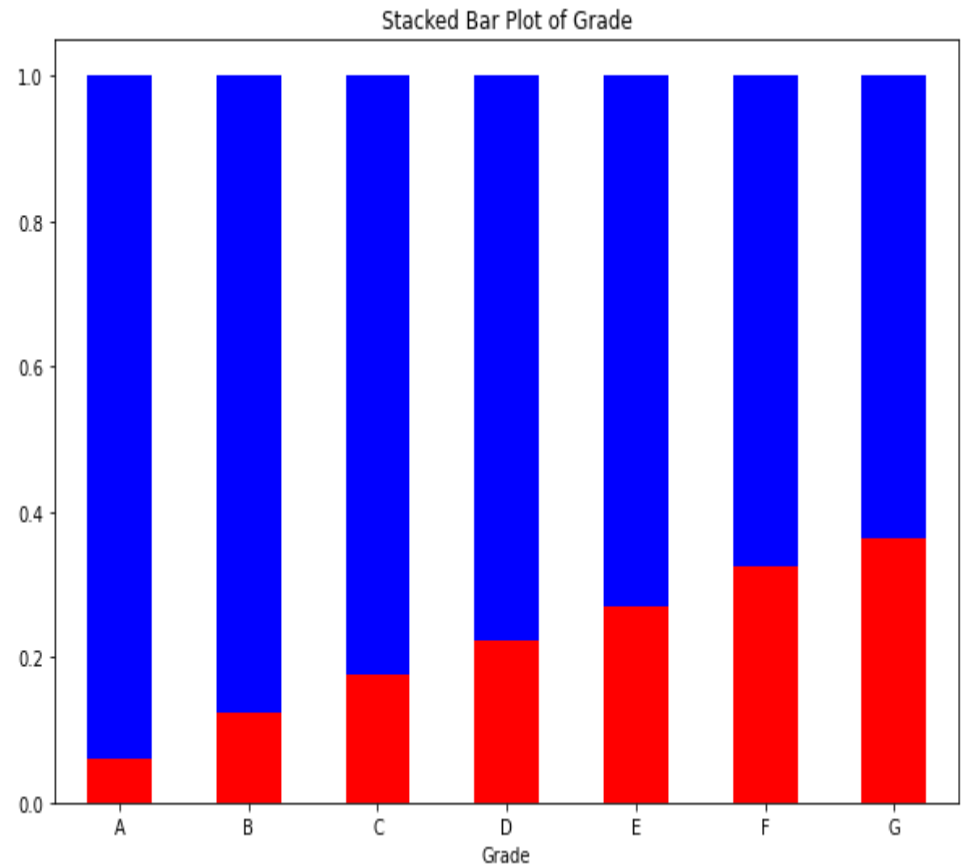


The purpose of applying loan for majority of the borrowers is dept consolidation (18641) followed by credit card (5130)

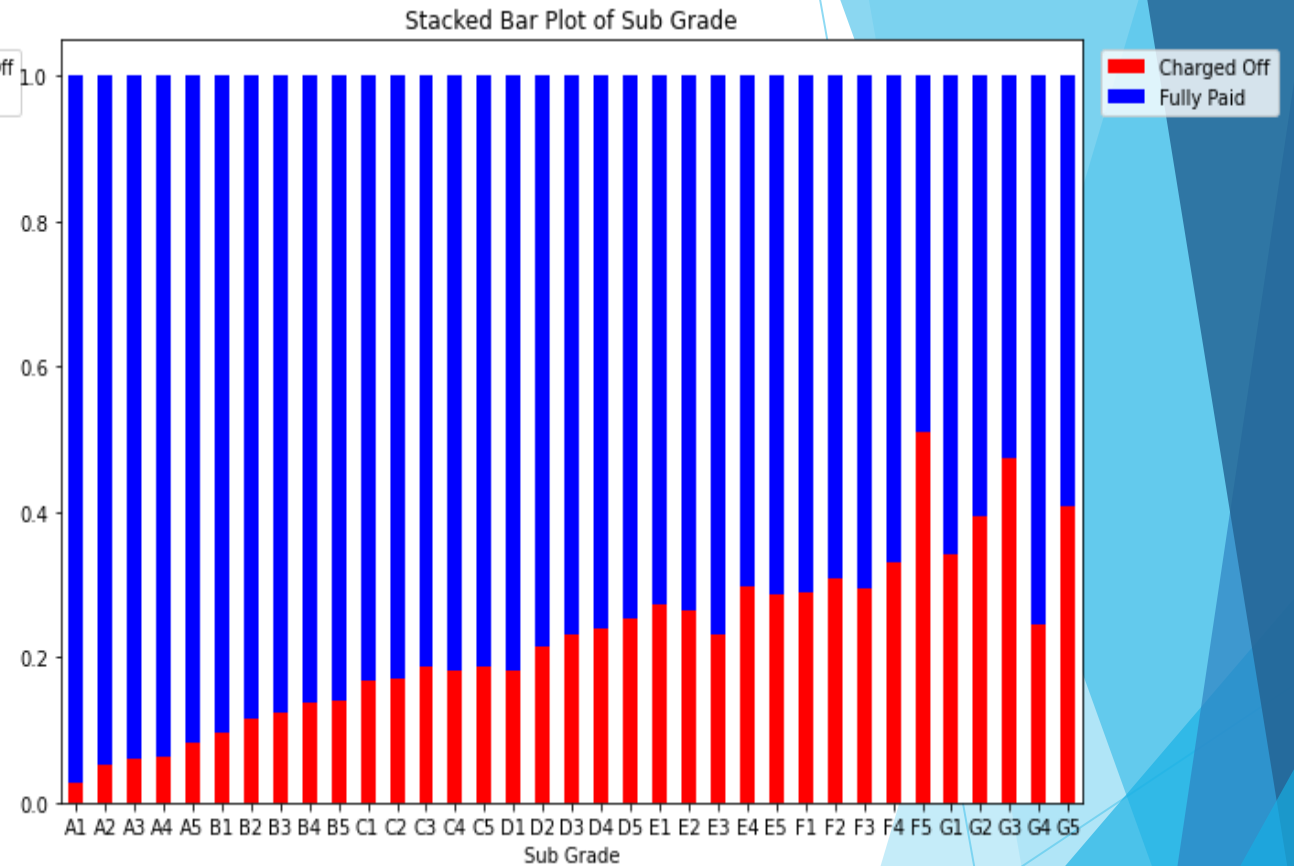


Majority of the borrowers applying for loan are from California state (7099)

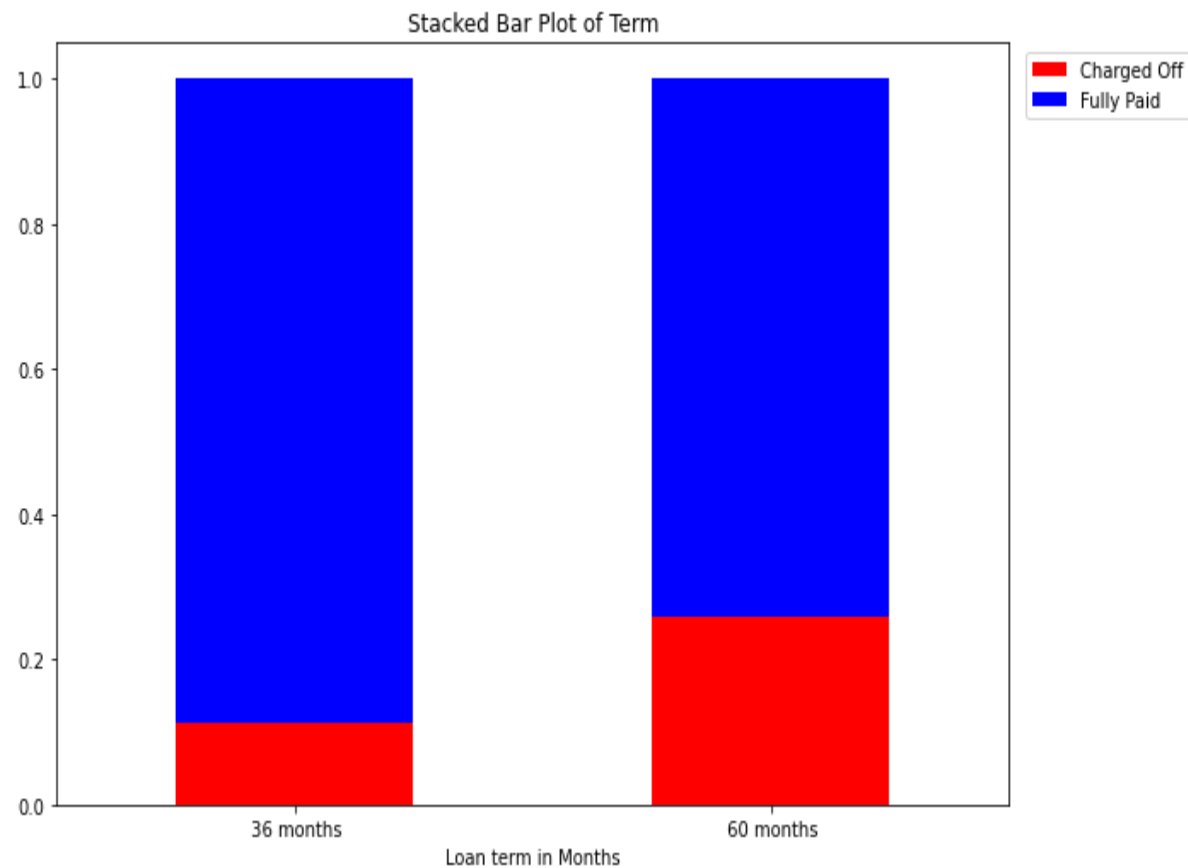
SEGMENTED UNIVARIATE ANALYSIS ON CATEGORICAL VARIABLES



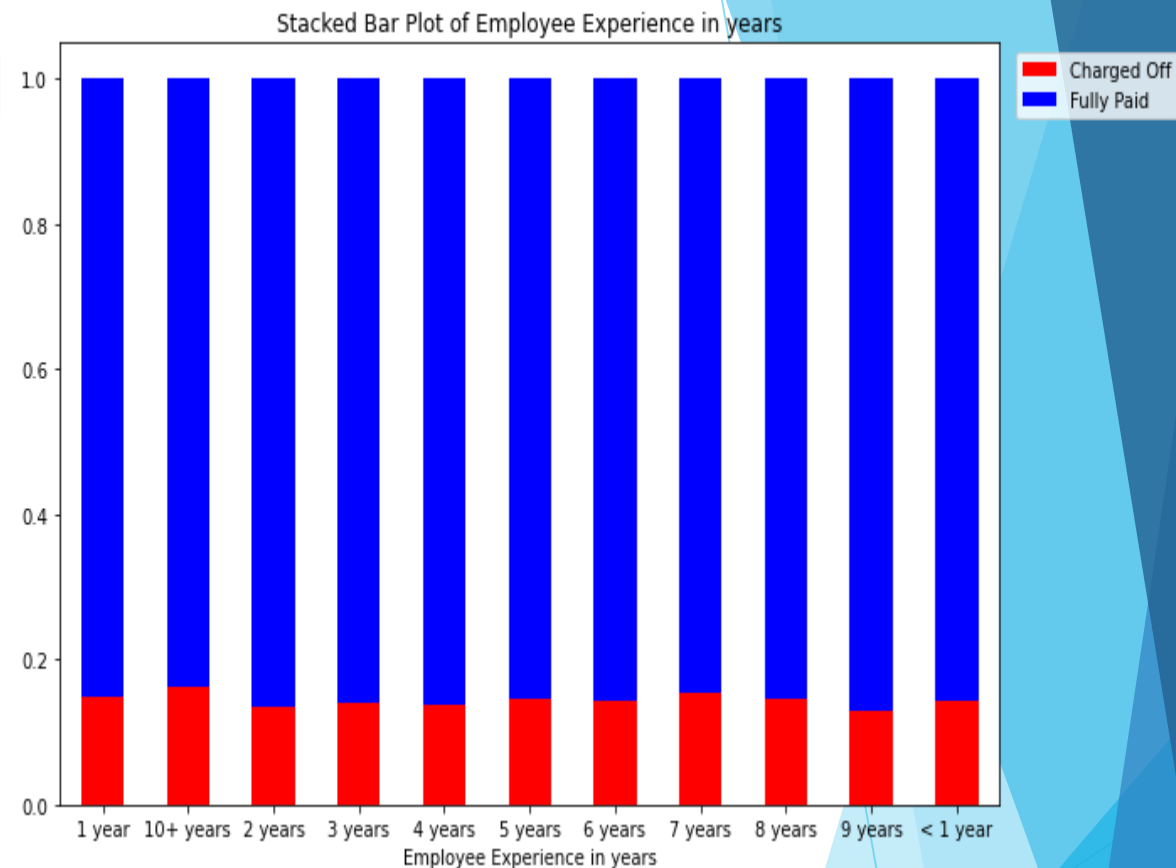
- We can see from the above graph that the Grade G are more prone to Defaulters.
- Going forward from A to G the rate of defaulters are increasing.



Sub Grade G3, G5, F5 is more prone to defaulters

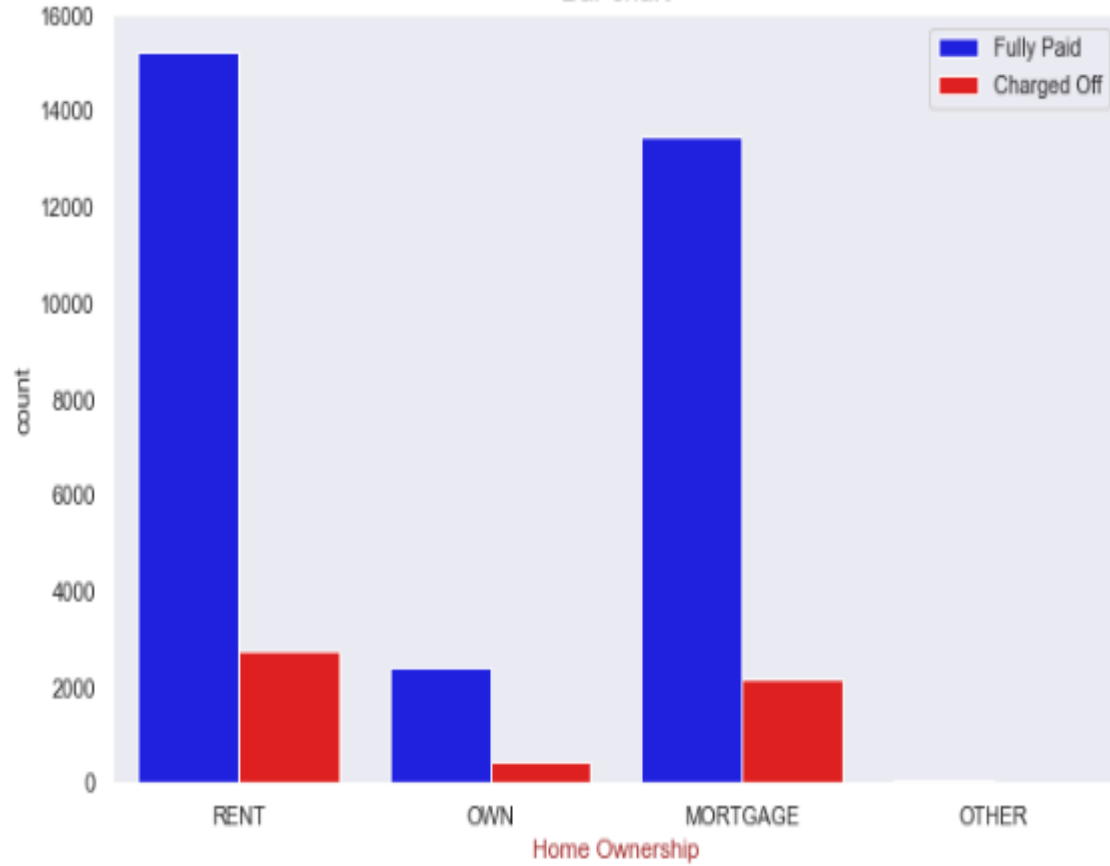


We can see from the above plot that the 60 months loan term is more prone to charge off



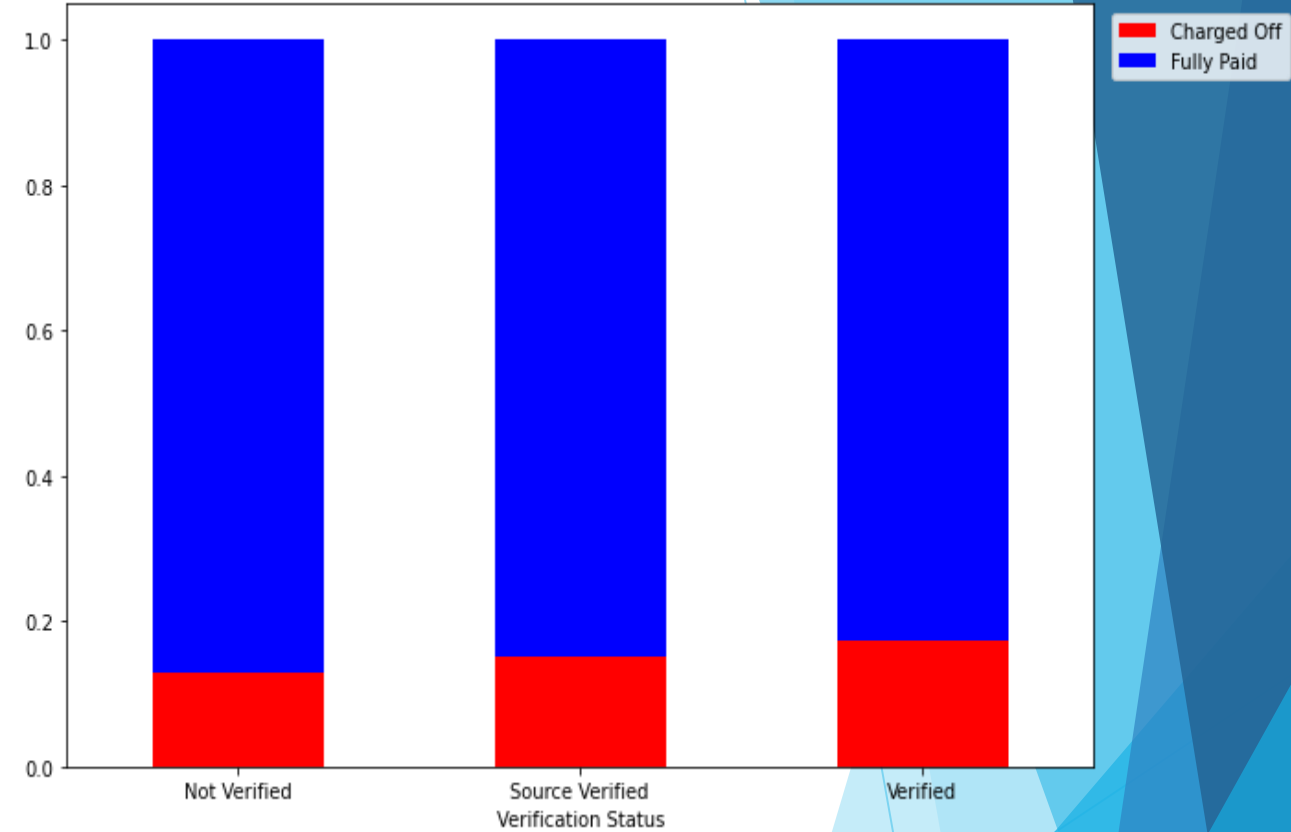
We can see from the above plot that there is not significant affect of no. of years of experience on loan status

Bar chart

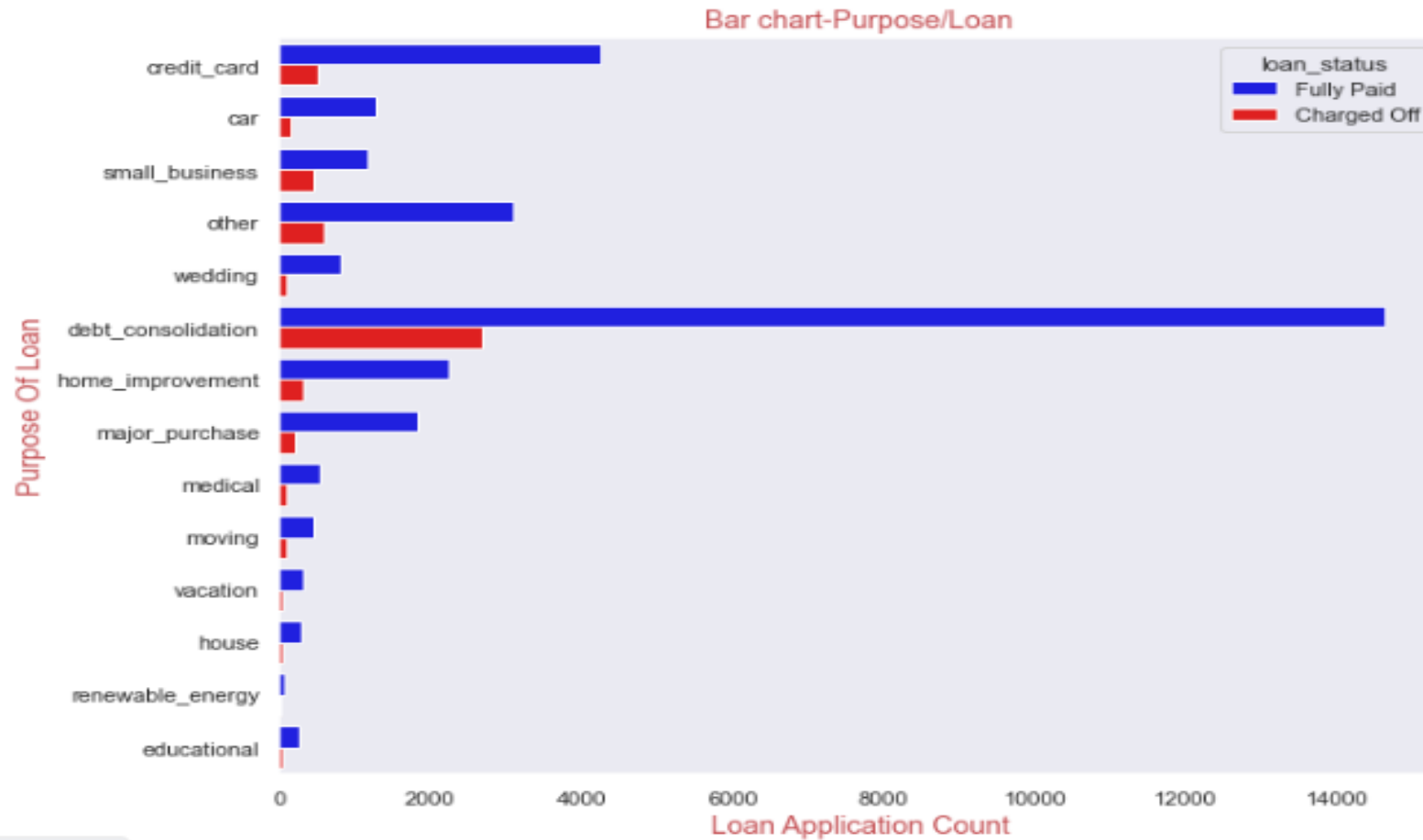


We can see from the above plot that applicants with rented house or with mortgages on their homes are more prone to default

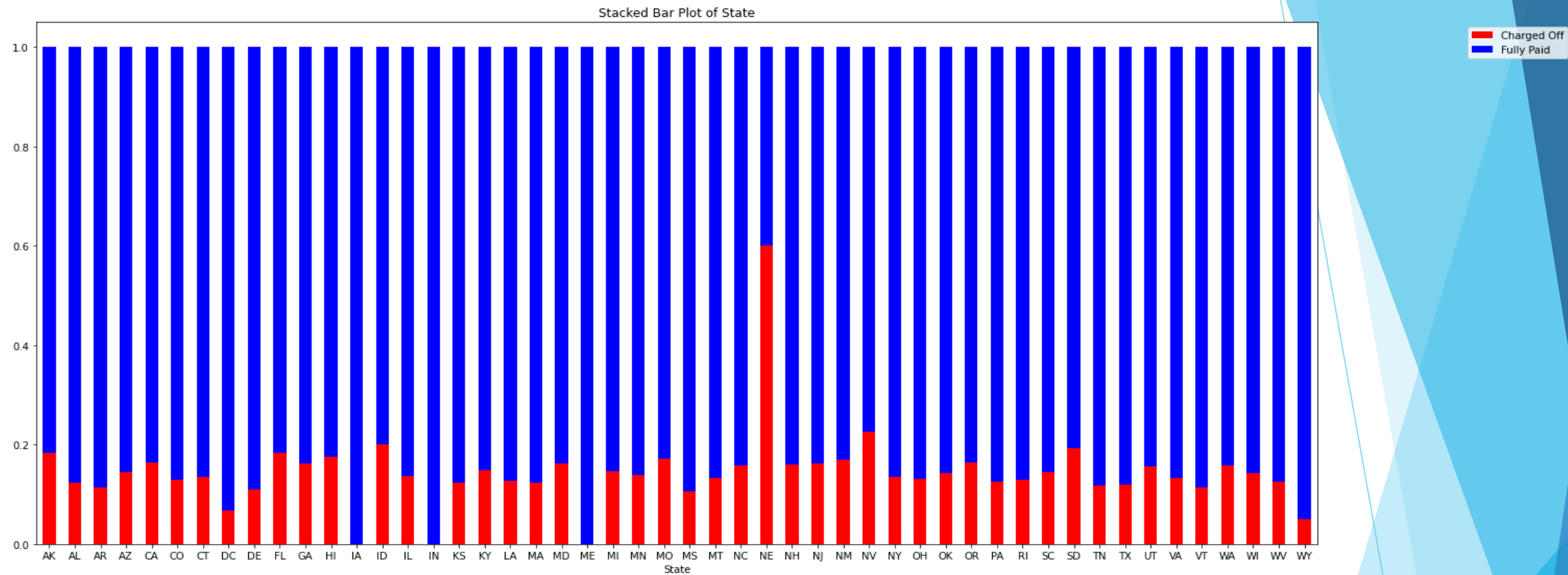
Stacked Bar Plot of Verification Status



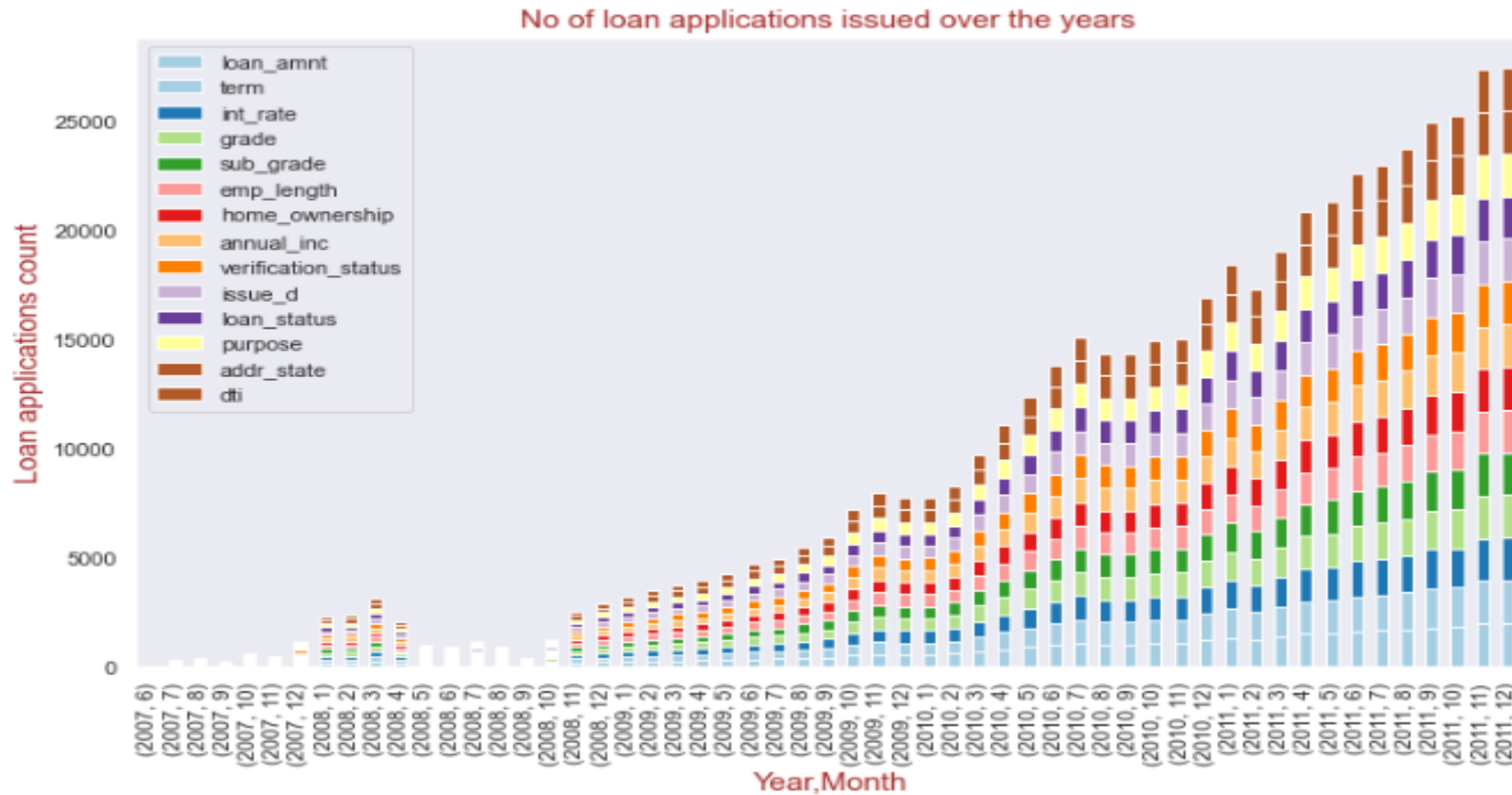
We can see from above plot that the verified loans are more prone to defaulters.



We can see from above plot that the loan taken for the purpose of small business has more prone to defaulters.

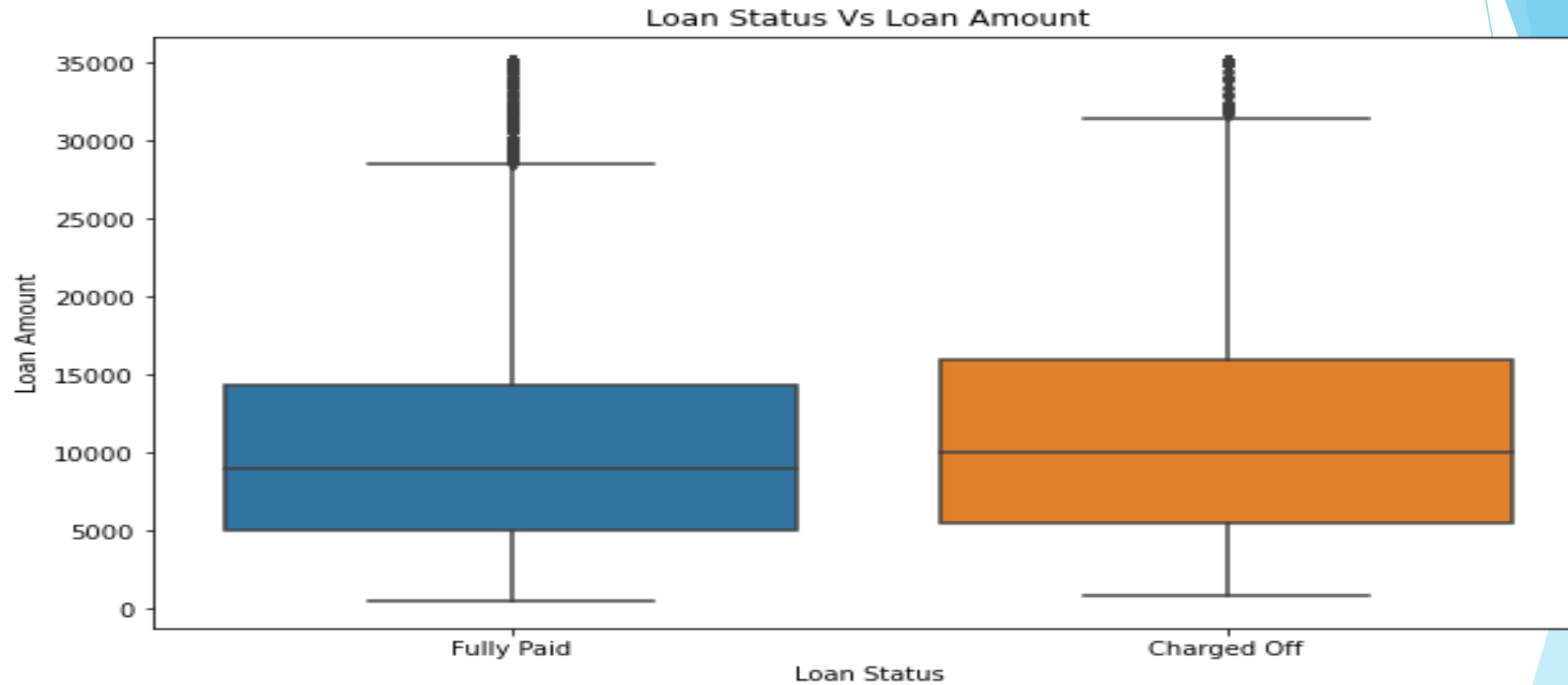


We can see from above plot that Nebraska (NE) state borrowers are more prone to defaulters



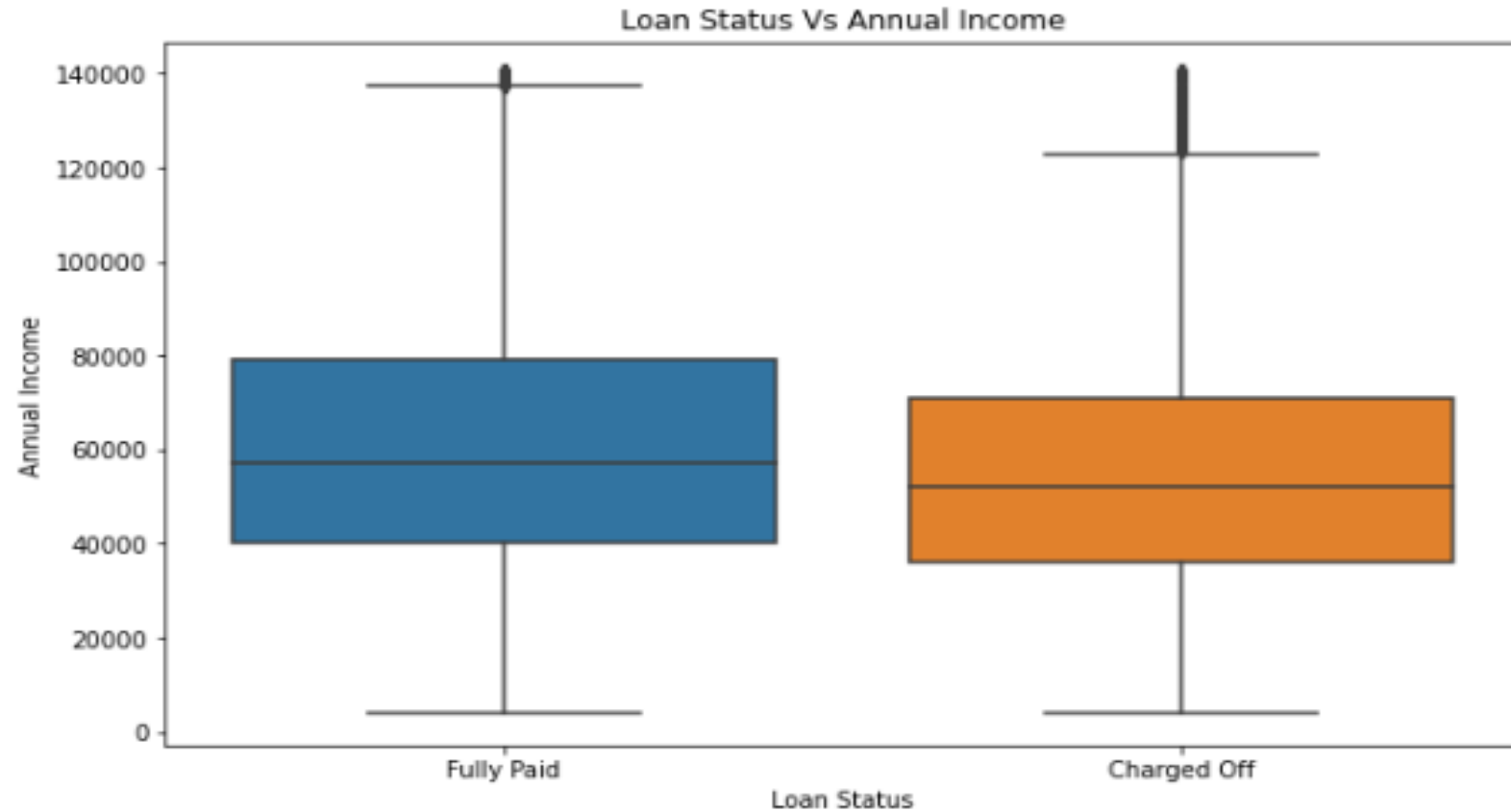
- Observation is that count of loan application is increasing every passing year.
- Increase in number of loan applications are adding more to number of charged off applications.
- Number of loans issued in 2008(May-October) got dipped, may be due to Recession.

SEGMENTED UNIVARIATE ANALYSIS ON CONTINUOUS VARIABLES



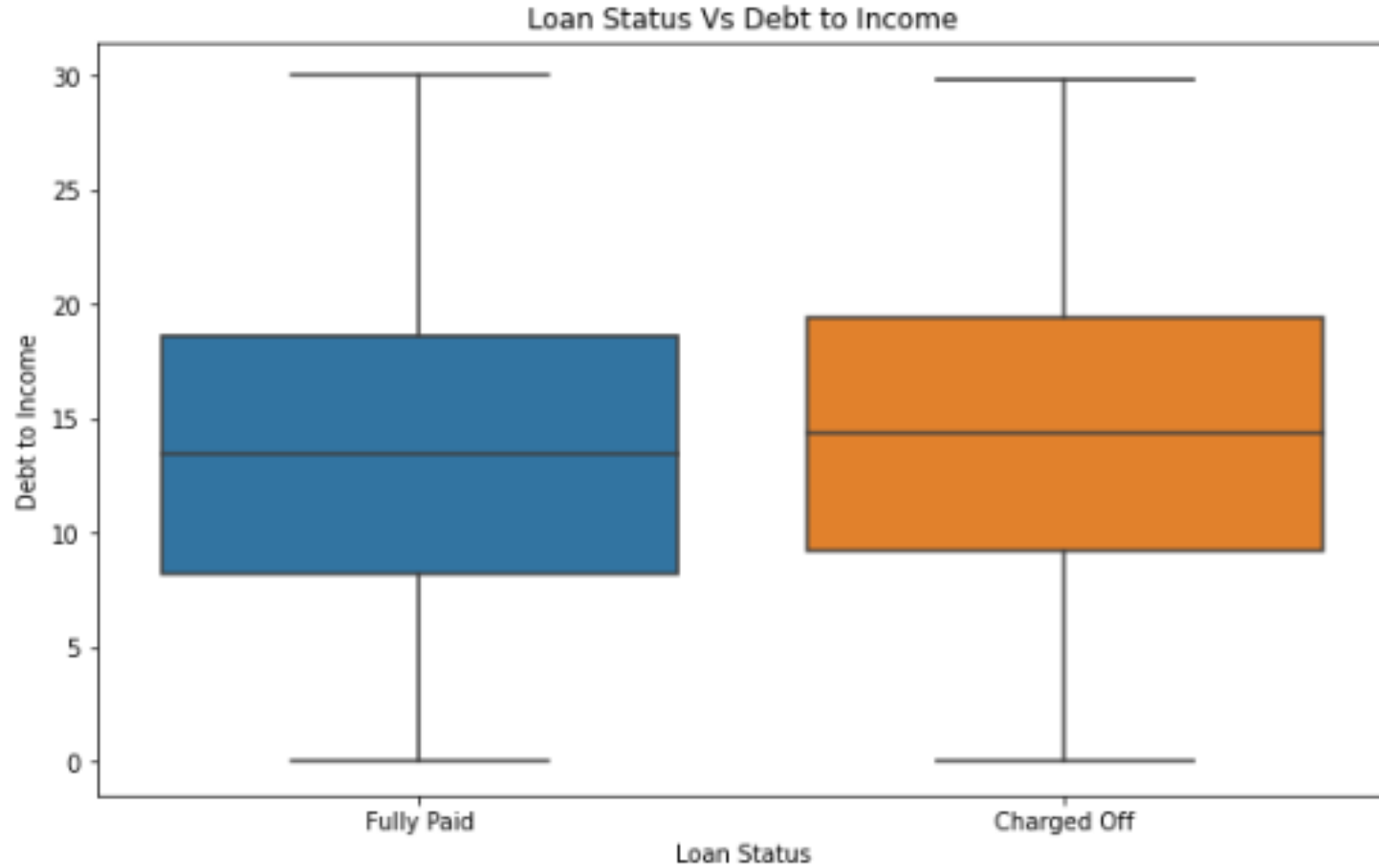
Impact of Loan amount: Higher the loan amount ,higher the chance of default

SEGMENTED UNIVARIATE ANALYSIS ON CONTINUOUS VARIABLES



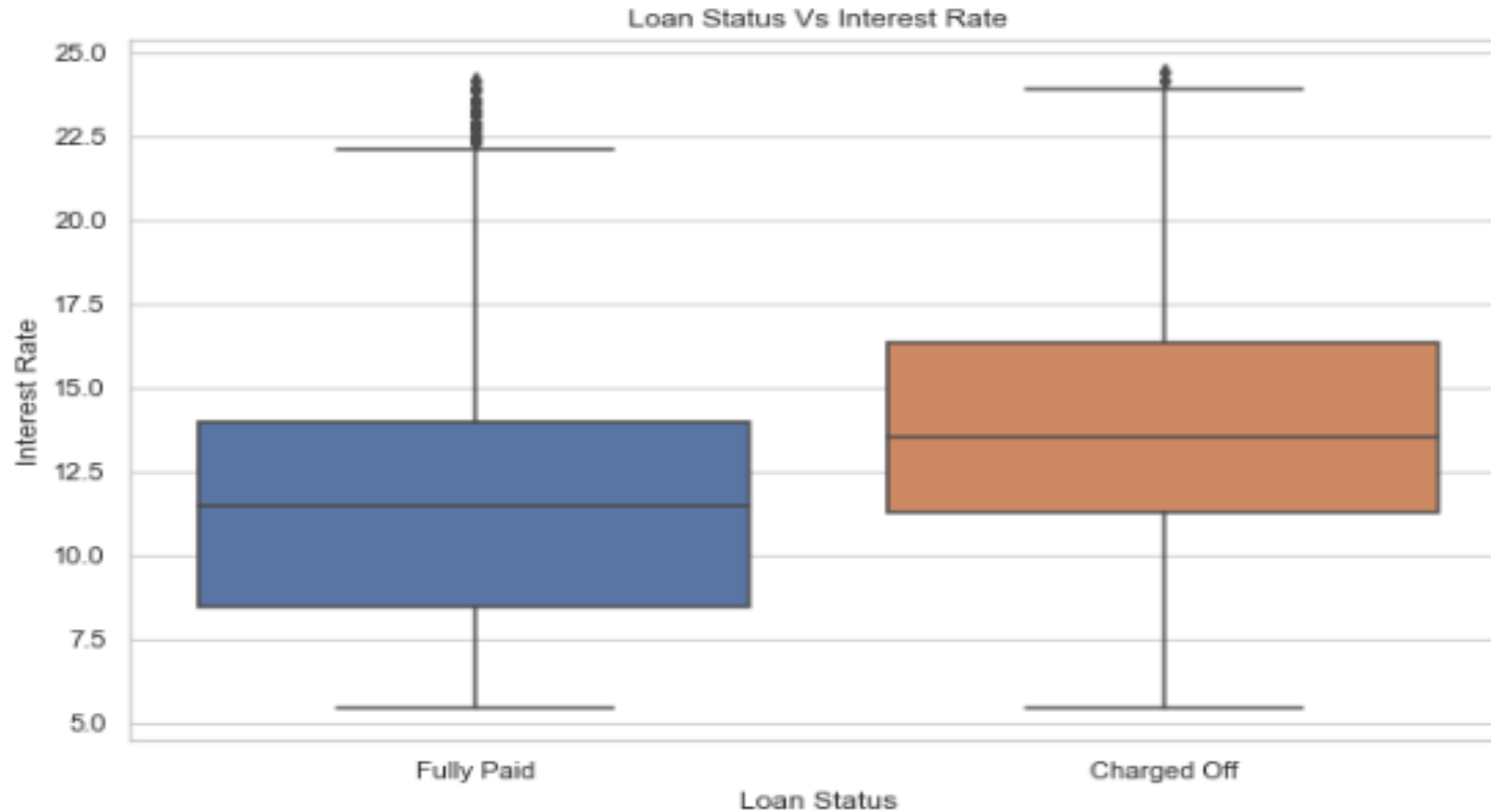
Impact of Annual Income: Lower the annual income, higher the chance of default

SEGMENTED UNIVARIATE ANALYSIS ON CONTINUOUS VARIABLES



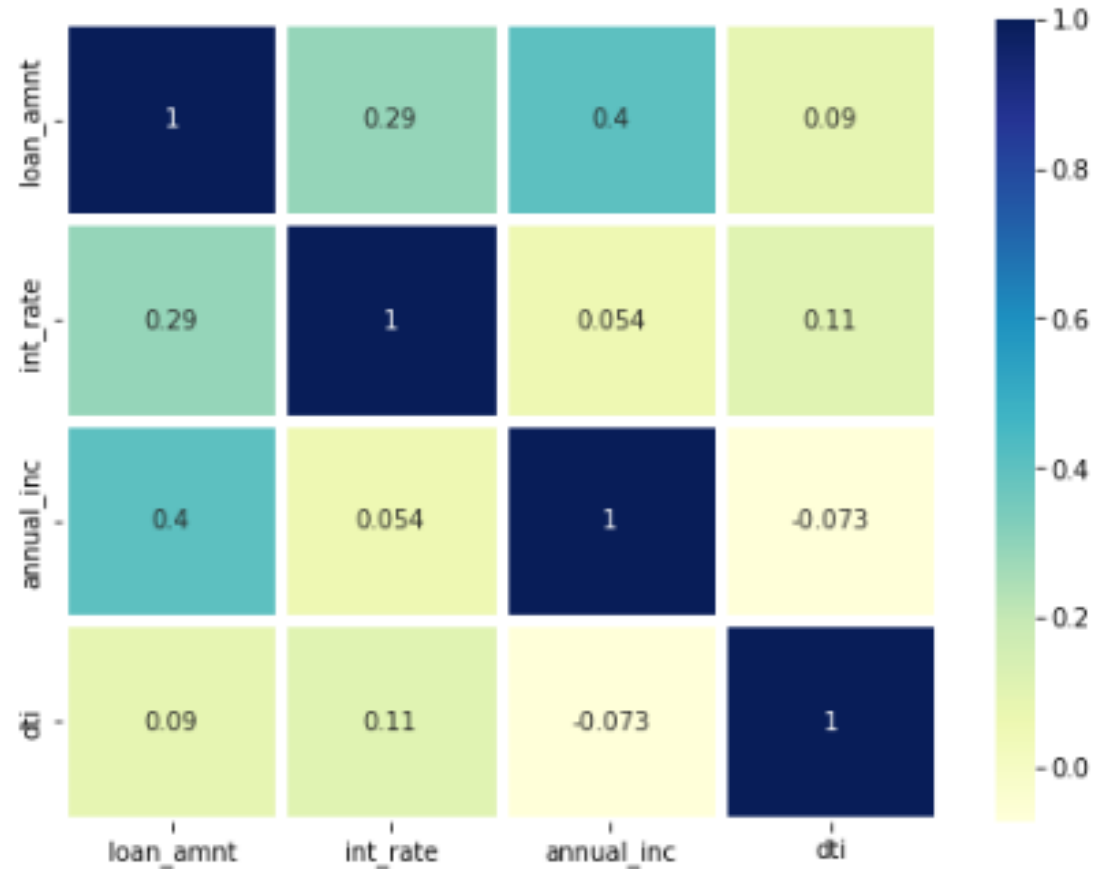
Impact of DTI: Low debt to income is slight less prone to default

SEGMENTED UNIVARIATE ANALYSIS ON CONTINUOUS VARIABLES



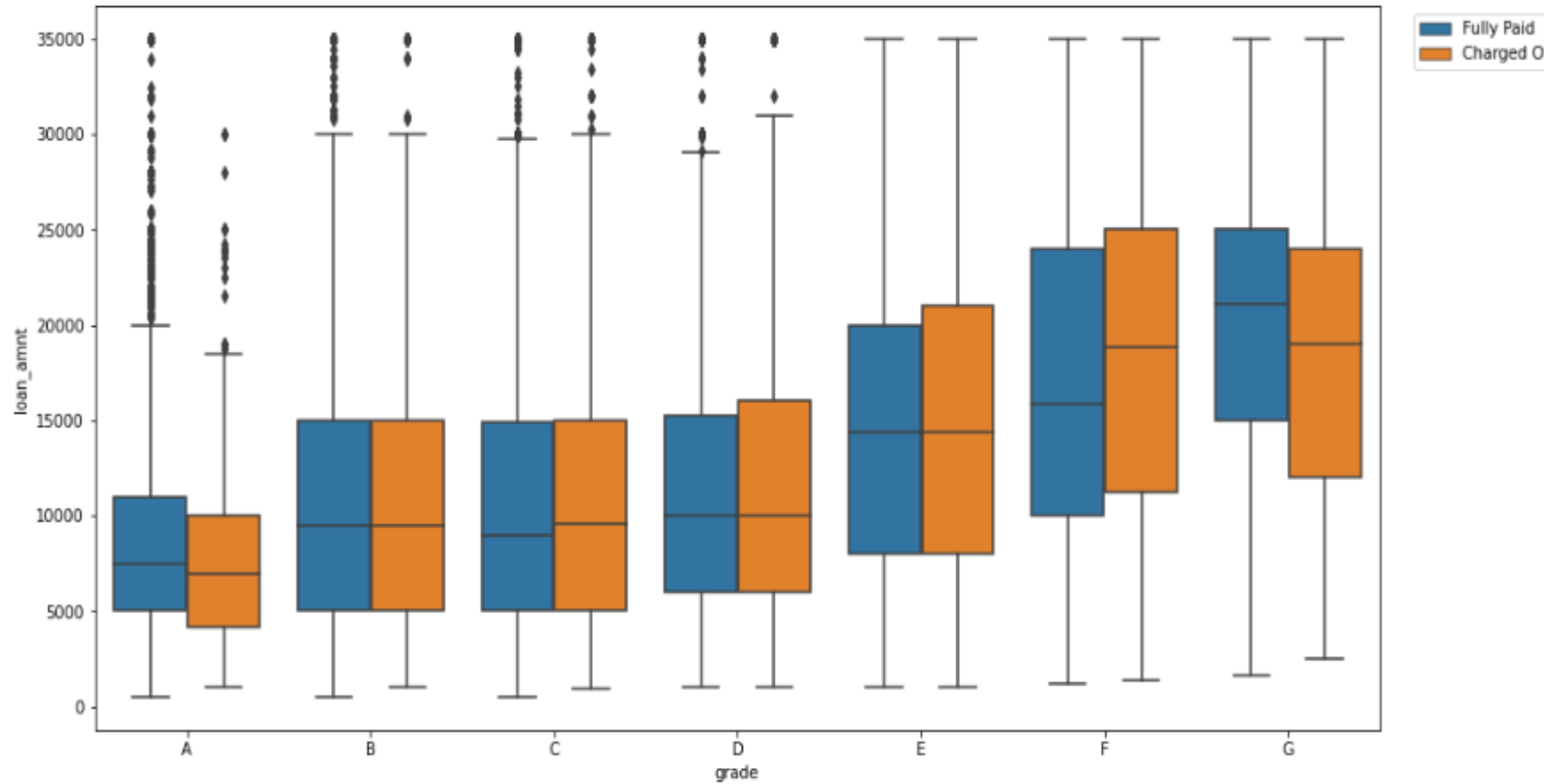
Impact of Interest rate: Higher interest rates are more prone to default

BIVARIATE ON CONTINUOUS VARIABLE



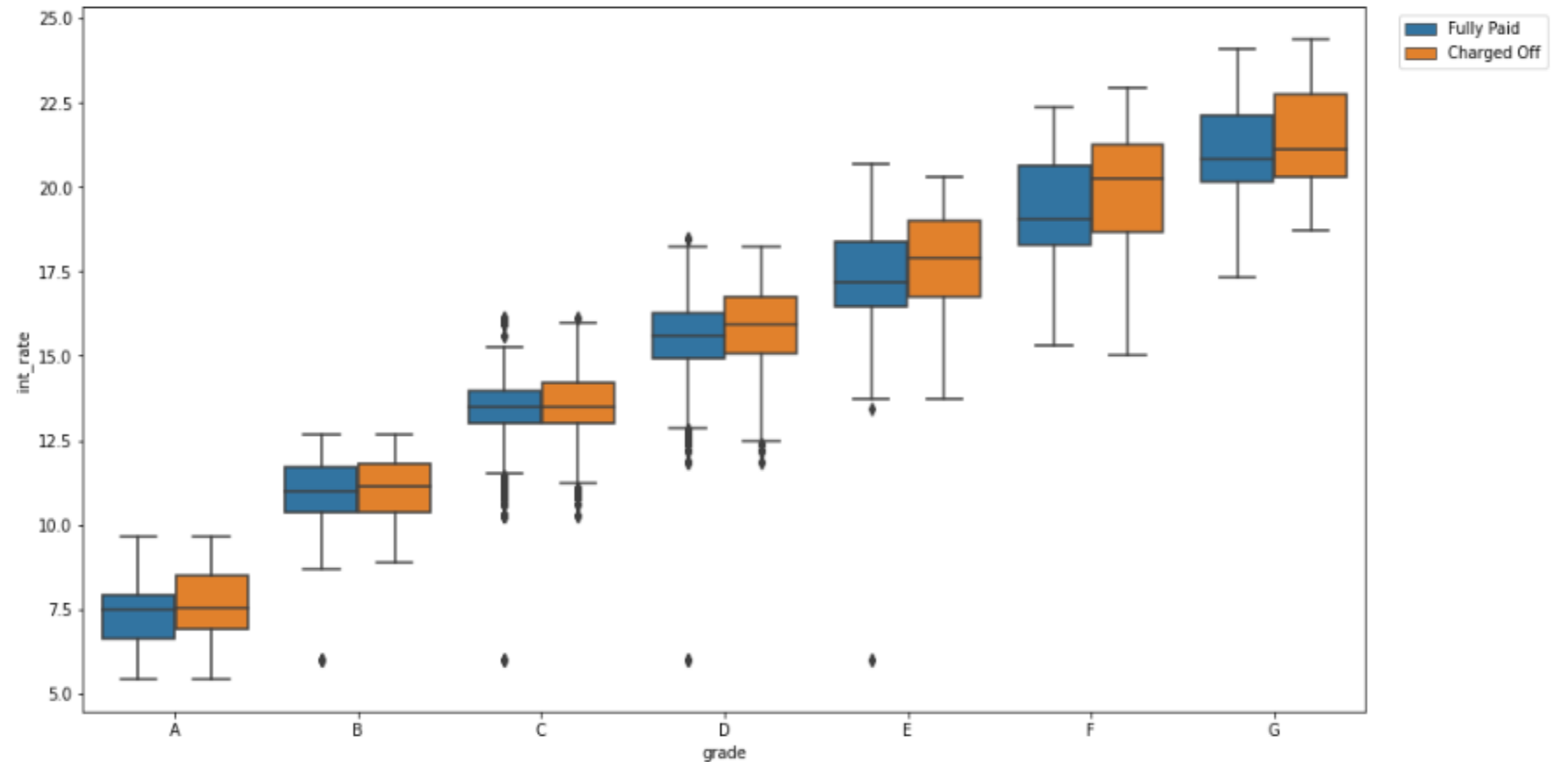
Loan amount and Annual Income is positively correlated with the correlation value of 0.4
Other correlation values are very small

BIVARIATE ON CATEGORICAL VARIABLE



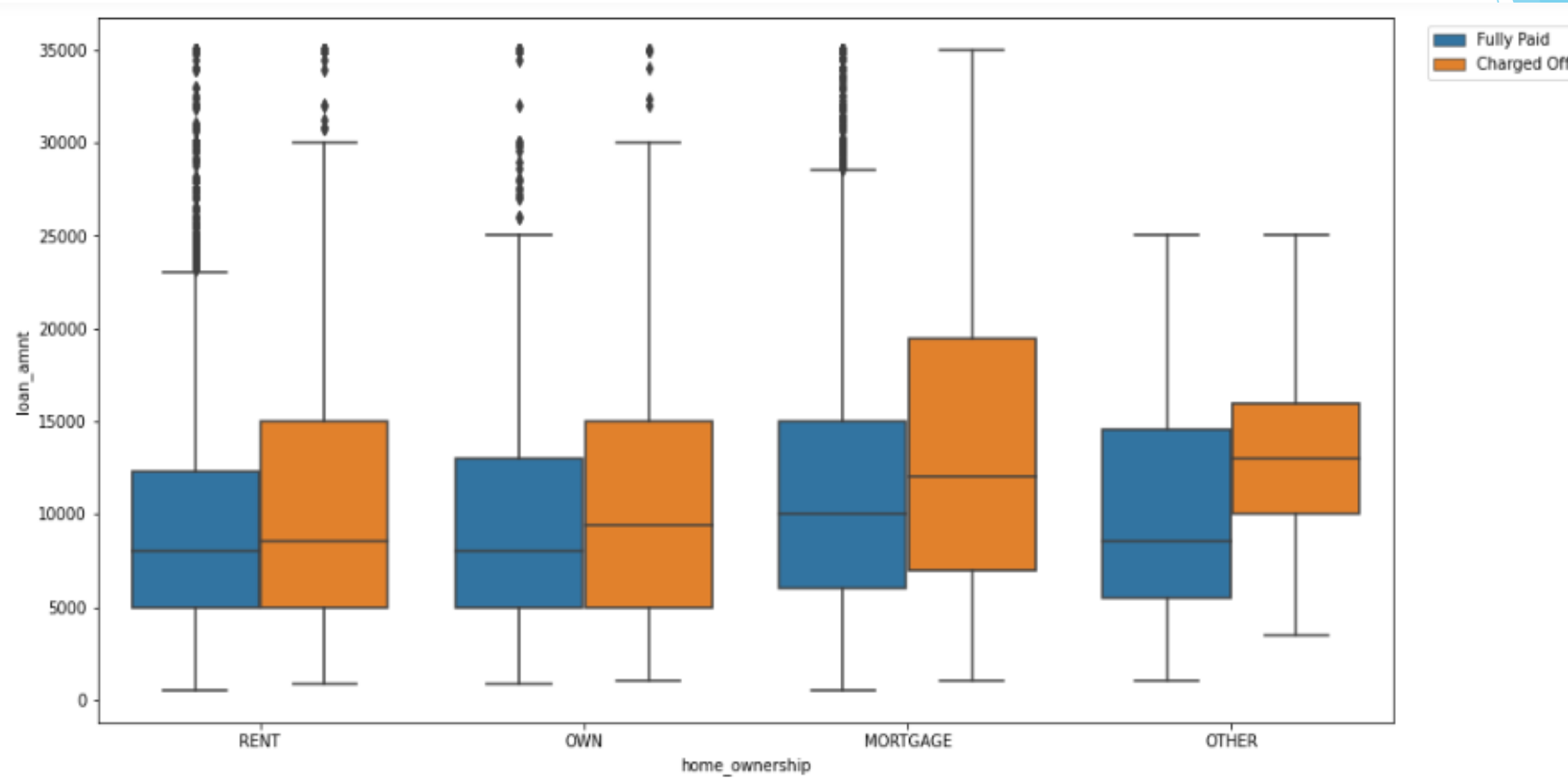
We can see from the plot that for Grade F, higher the loan amount higher are the chances of default

BIVARIATE ON CATEGORICAL VARIABLE



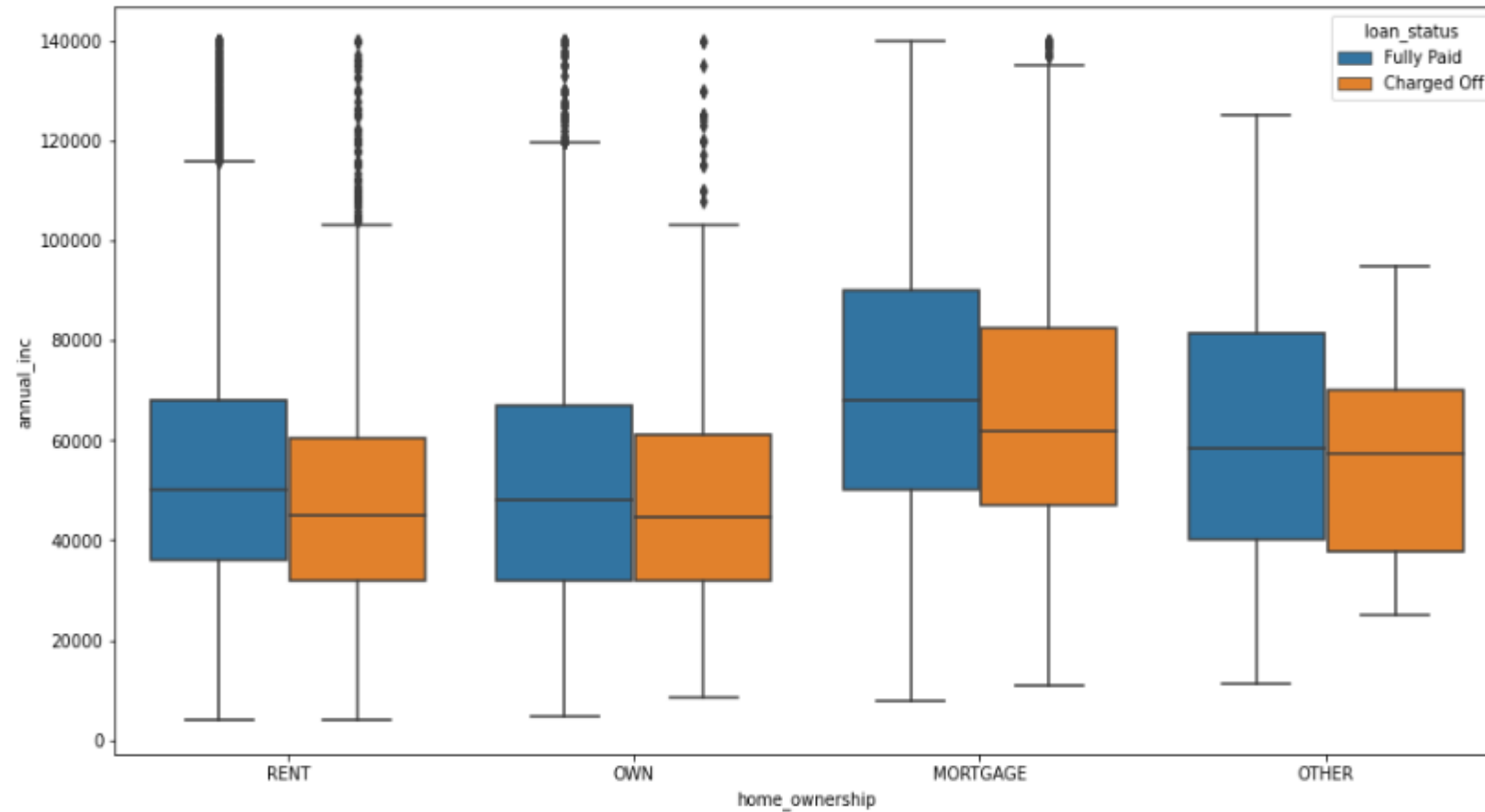
1. We can see as the grade is moving from A to G, the interest rate is also increasing.
2. We can see from the plot for Grade F, G, higher the interest amount, higher the chances of default

BIVARIATE ON CATEGORICAL VARIABLE



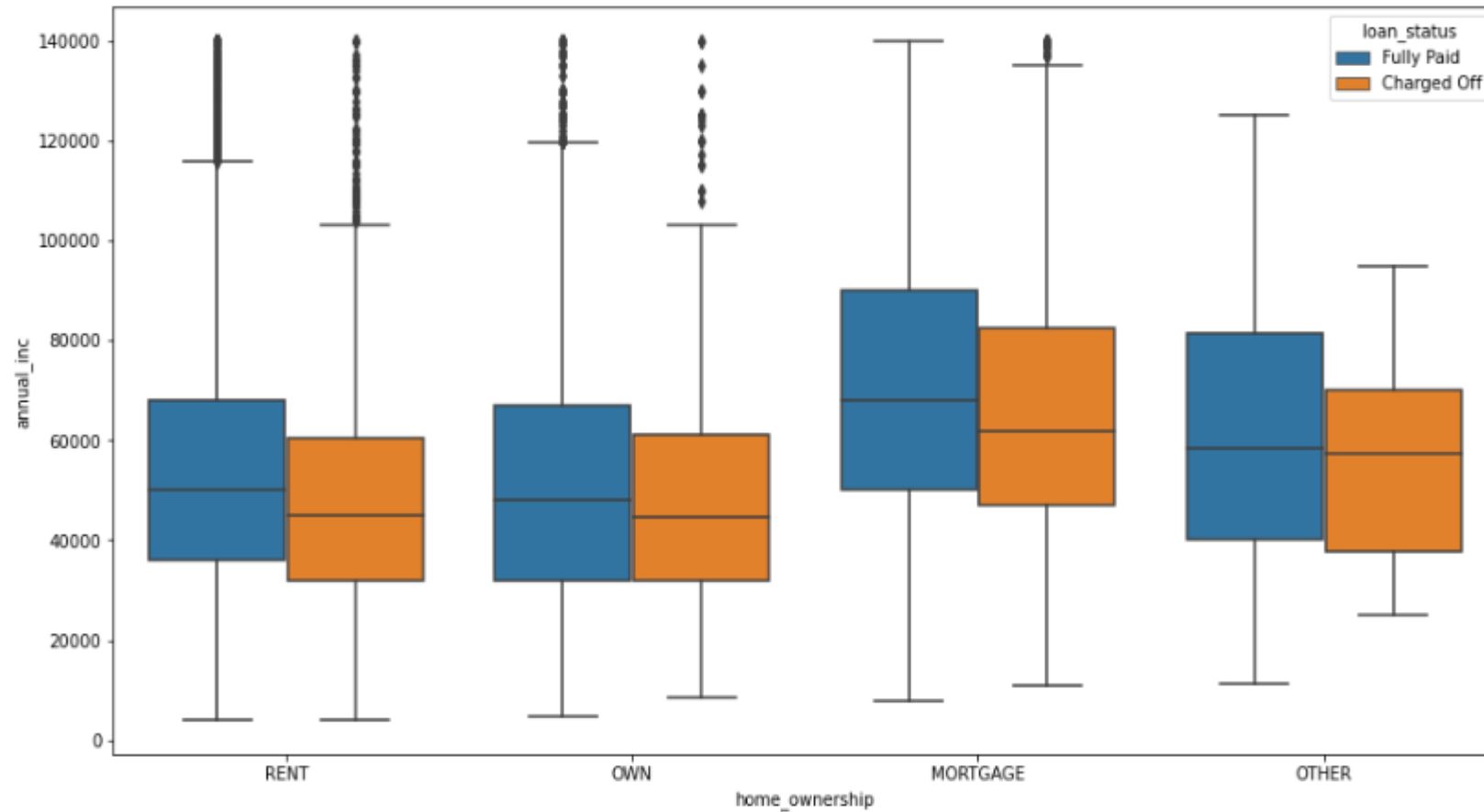
For Mortgage homes, higher the loan amount, higher the chances of default

BIVARIATE ON CATEGORICAL VARIABLE



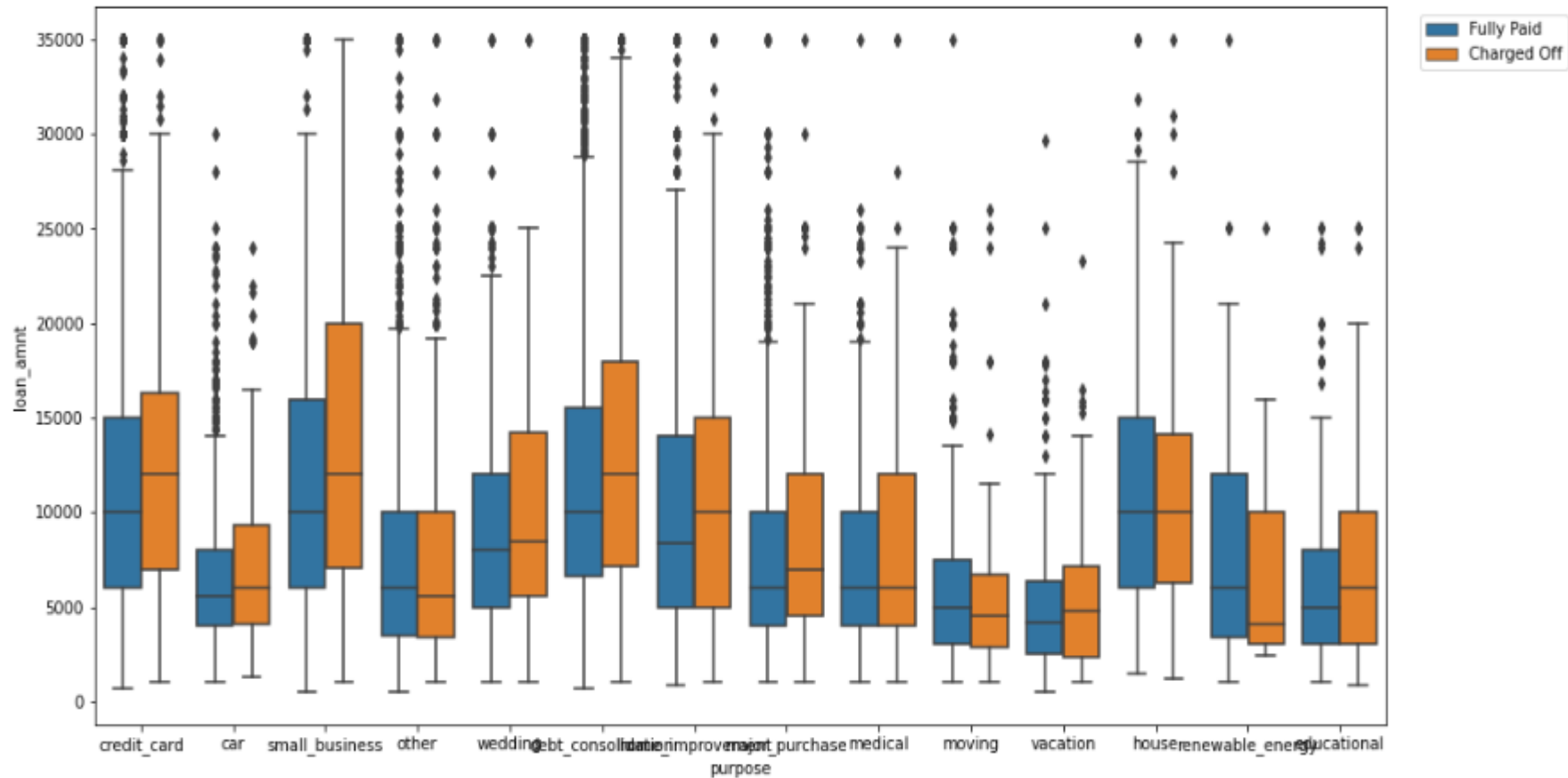
We cannot comment here anything as it shows same for all types ownership, lesser the income, higher the chance of default

BIVARIATE ON CATEGORICAL VARIABLE



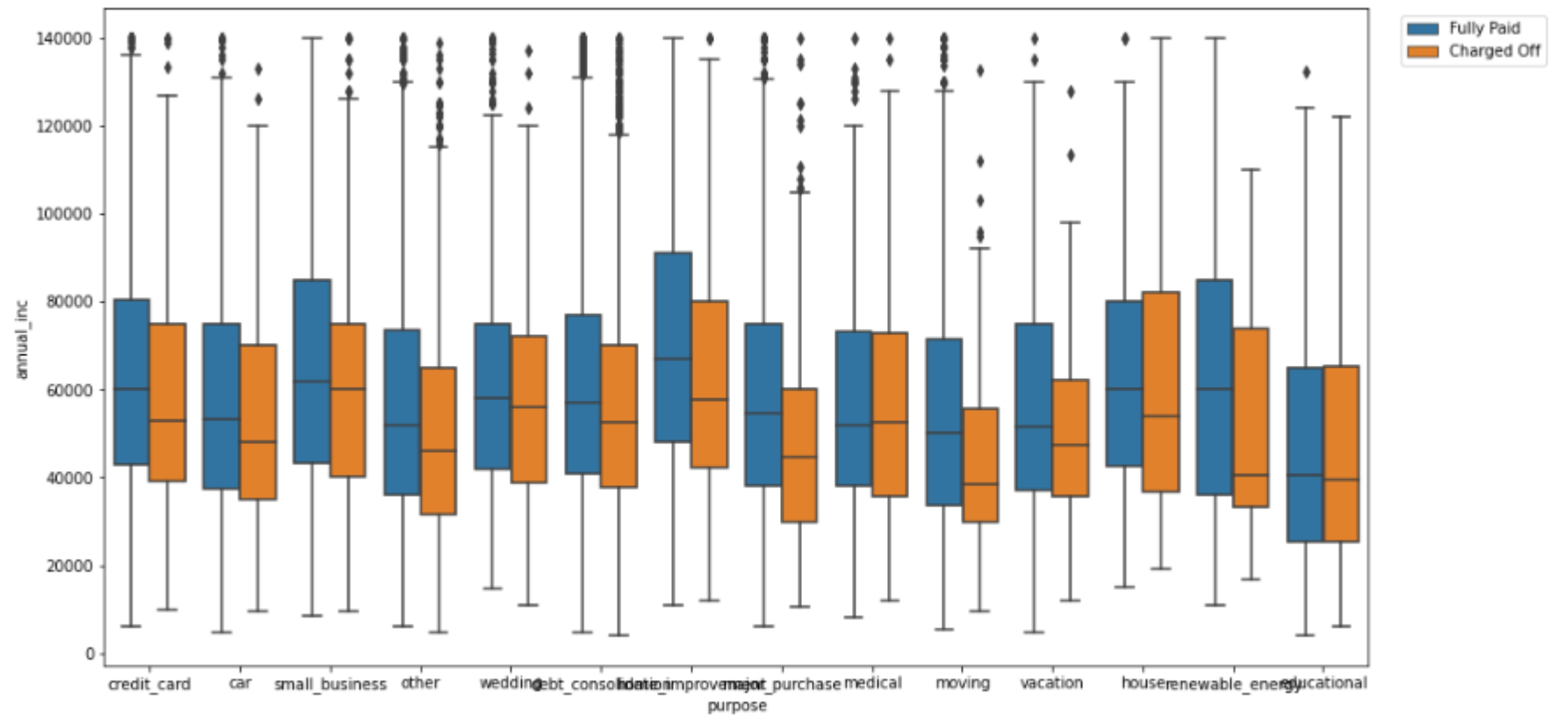
We cannot comment here anything as it shows same for all types ownership, lesser the income, higher the chance of default

BIVARIATE ON CATEGORICAL VARIABLE



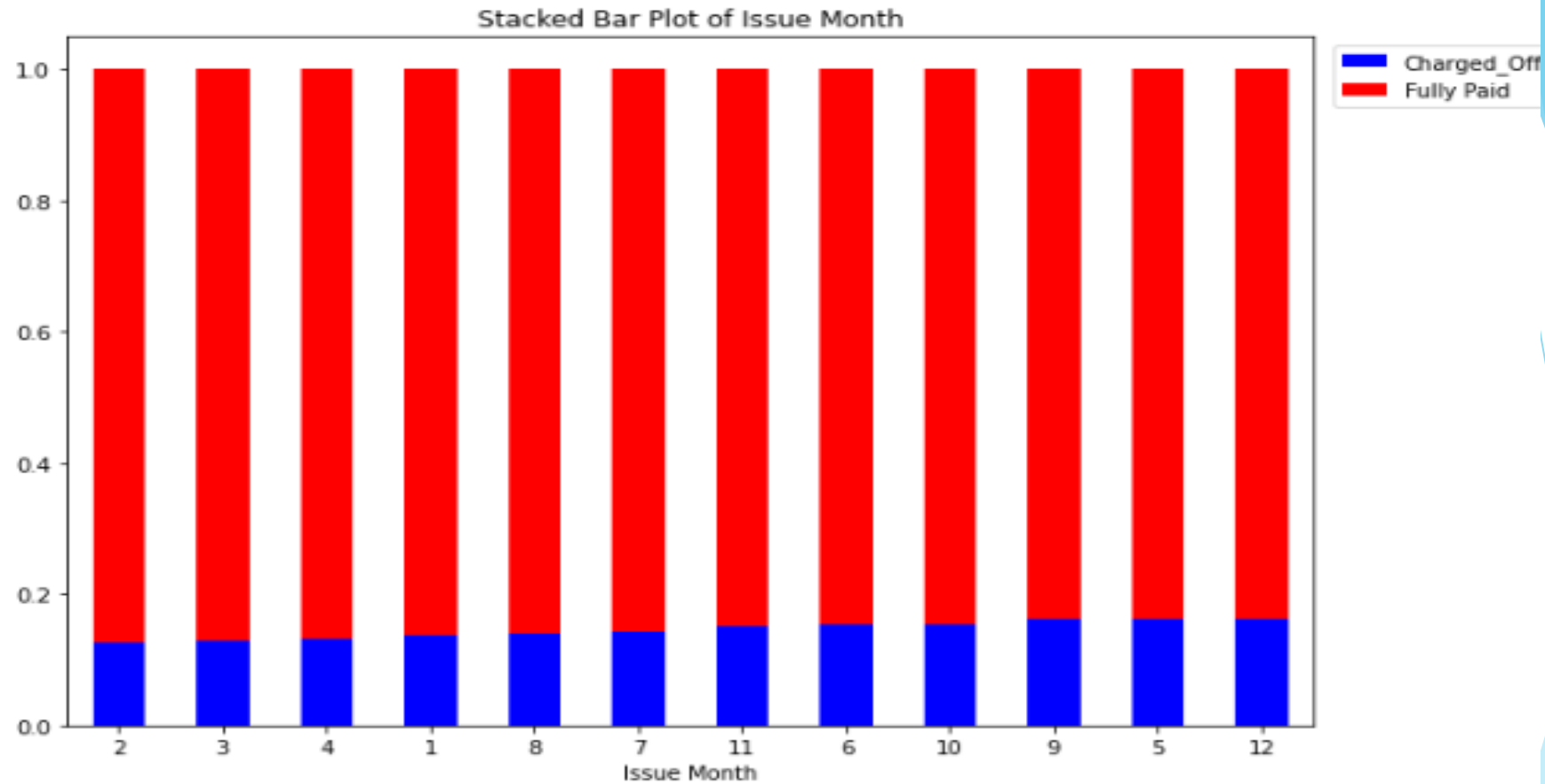
We can see from the plot that for Small business and Debt consolidation if the loan amount is high. higher the chance of default.

BIVARIATE ON CATEGORICAL VARIABLE



We can see from the graph for the renewable energy if the annual income is low, higher the chance of default.

DERIVED MATRIC ANALYSIS



Above plot shows 'May', 'September' and 'December' has the high possibility of defaulting

RECOMMENDATIONS BASED ON UNIVARIATE ANALYSIS

- ▶ Borrowers under Grade G has highest chances of defaulters : 36%
- ▶ Borrowers under sub grade F5 is more prone to defaulters : 51%
- ▶ Most of the defaulters are the one who has term of 60 months : 25%
- ▶ Nebraska (NE) state borrowers are more likely to be defaulters: 60%
- ▶ Verified customers are prone to defaulters but the impact is less but impact is not too high
- ▶ Applicants taking loan for debt consolidation, paying credit card bill and small business has more prone to defaulters
- ▶ Applicants with rented house or with mortgages on their homes are more prone to default. Number of loans issued in 2008(May-October) got dipped, may be due to Recession
- ▶ Count of loan application is increasing every passing year.
- ▶ Higher the loan amount, higher the chances of defaulter
- ▶ Lower the income, higher the chances of being defaulter
- ▶ Higher interest rates are more prone to become defaulters

RECOMMENDATIONS BASED ON BIVARIATE ANALYSIS

- ▶ Borrowers under Grade G and F has higher loan amount with high interest resulting in higher chances of defaulters
- ▶ Lower annual income leads to charge off in each grade category
- ▶ For Mortgage homes, higher the loan amount, higher the chances of default.
- ▶ Loan taken for the purpose of small business and debt consolidation, if the loan amount is high there are higher chance of charged off
- ▶ Higher the Loan amount, higher the Annual Income. They are positively correlated with the correlation value of 0.4
- ▶ In the month of 'May', 'September' and 'December', there is high possibility of defaulting