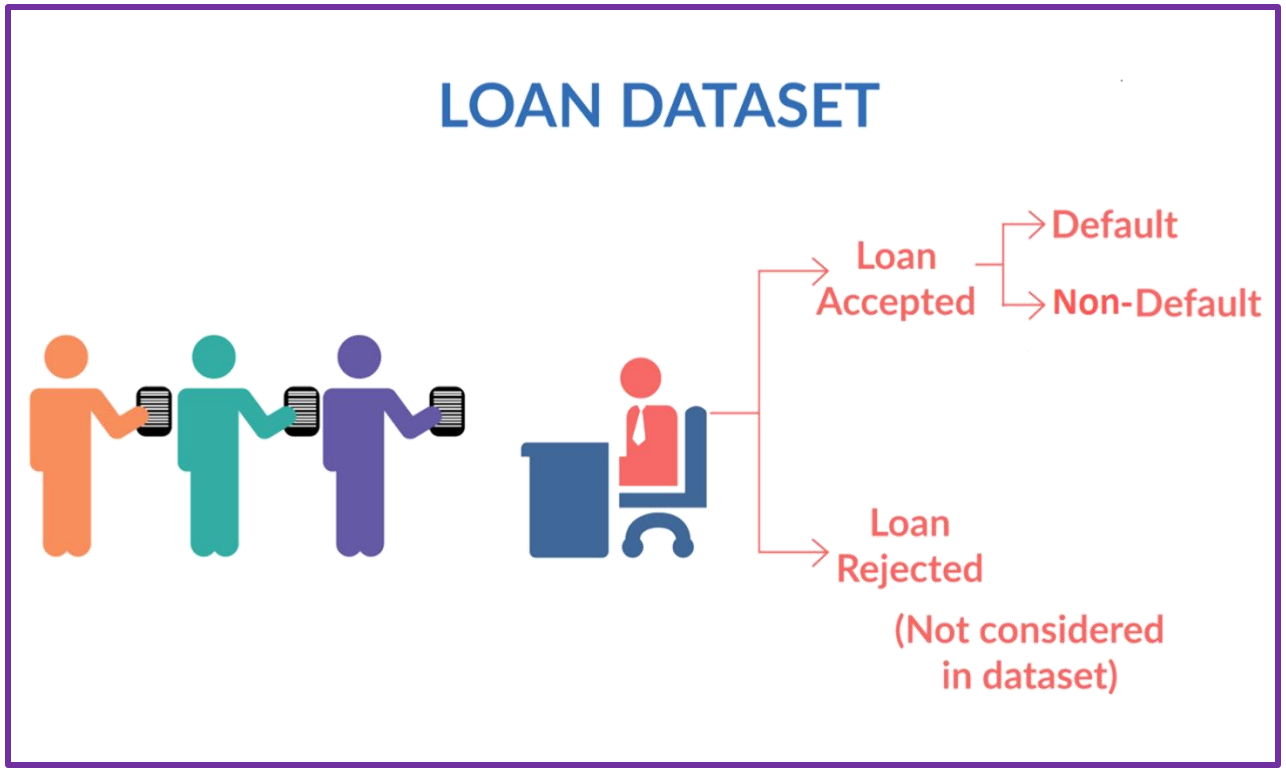
LENDING CLUB CASE STUDY

Group Members:

1. Chandrasekaran B
2. Sankar Thulasimani

Case Study Objectives



- Identification of Loan Applicant traits that tend to 'default' paying back
- Understand the 'Driving Factors' or 'Driver Variables' behind Loan Default phenomena
- Loan Lending Organizations may choose to utilize this knowledge for its portfolio and risk assessment of new loan applicants

EDA ON Lending Club Data

Understanding Dataset

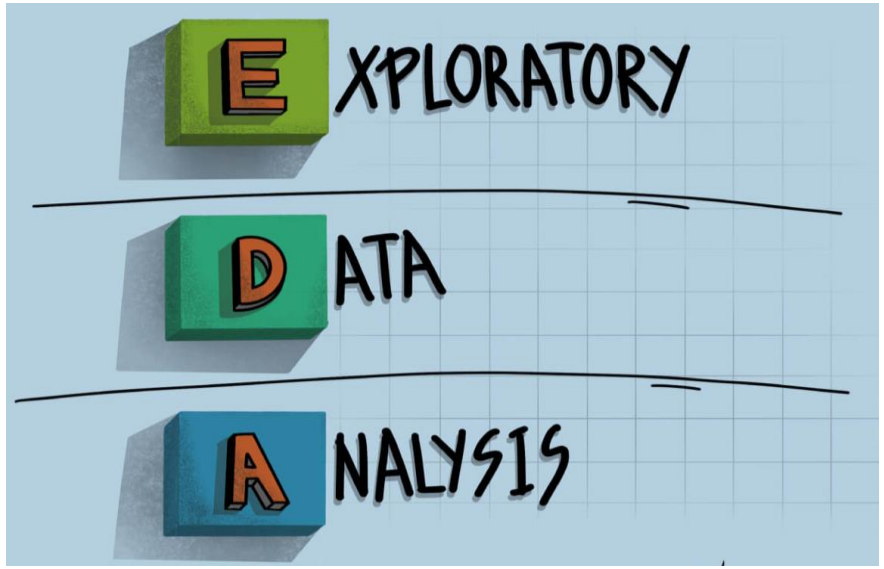
Data Cleaning

(Removing Null values and imputing null values)

Data Preparation

(Deriving new columns, Sanity checks)

Exploratory Data Analysis



- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

Exploratory Data Analysis – Approach Followed

- The purpose of this case study is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting which resulting the loss to the company

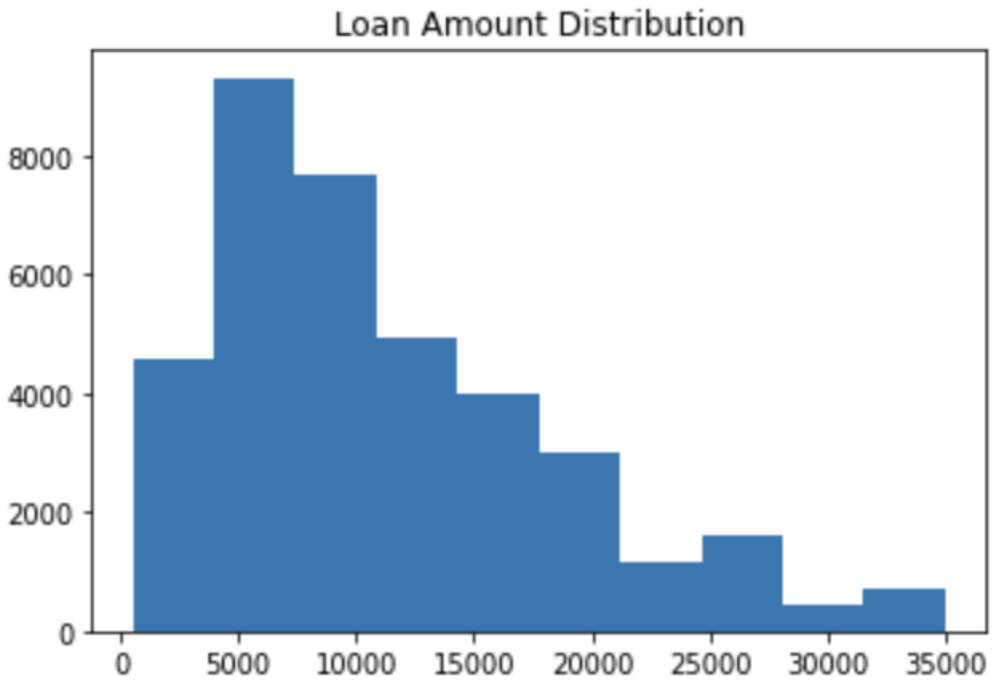
- We performed **data cleaning and preparation** on the Loan dataset:
  - Removed the NA values for all the variables (above 30%)
  - Derived multiple columns from the given data
    - Segregated date column into day, month and year
    - Created total amount column from term and installment
    - Created income range, interest range, dti range , total amount range , etc

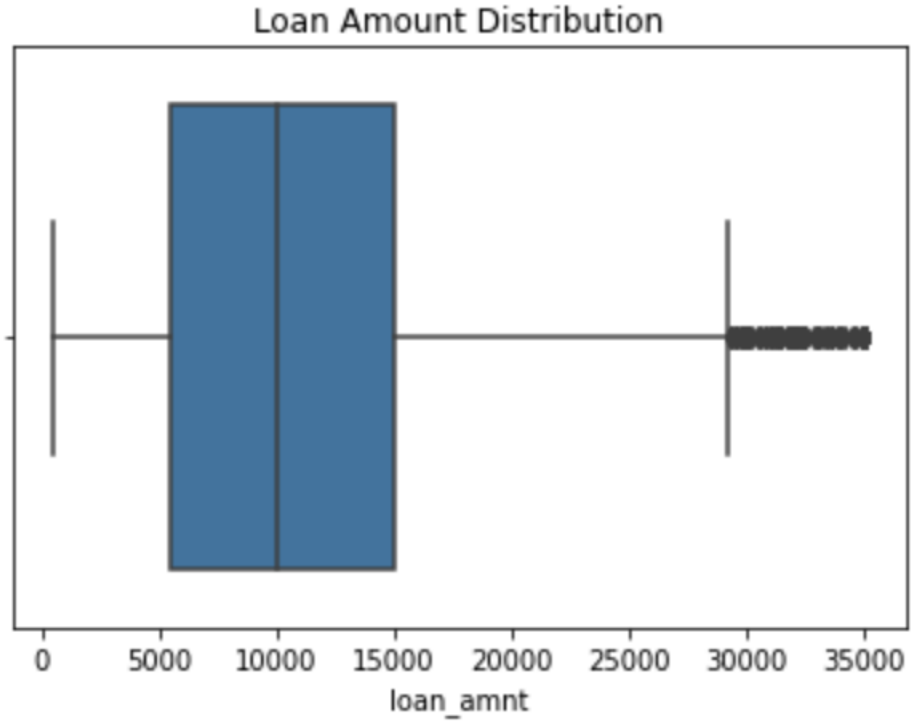
- **Univariate Analysis:**
  - Histograms and Bar charts to check out the distribution of all the driver variables
  - Box plots to detect the Outliers
  - Performed Segmented Univariate analysis for certain set of data

- **Bivariant Analysis:**
  - Performed the Bivariant analysis to understand how different variables interact with charging off of loans.

Univariate Analysis

Loan Distribution Analysis

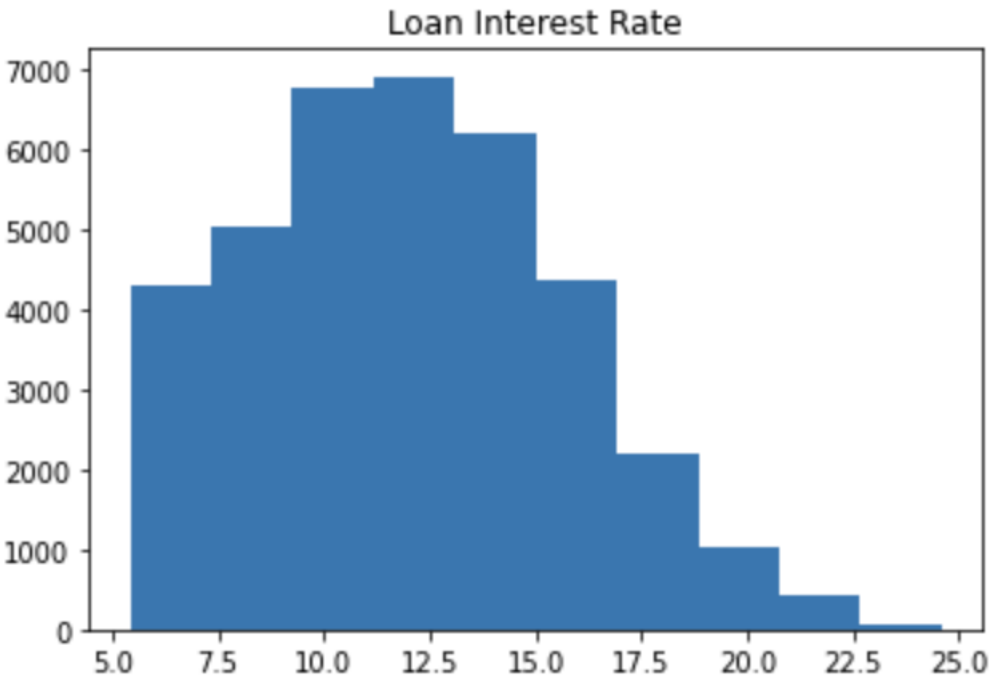


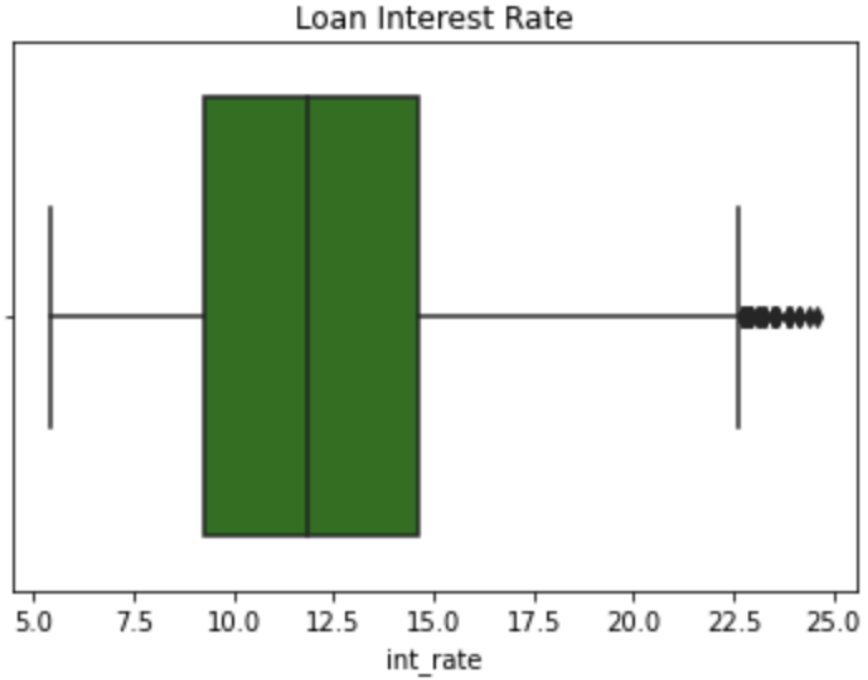


Observations:

1. The above plots indicate Very few people opted for higher amount loans, most of the people were borrowing from 10000-20000
2. This Observation can help the management to provide many offers to those category to make them opt for higher amount

Loan Interest Rate Analysis

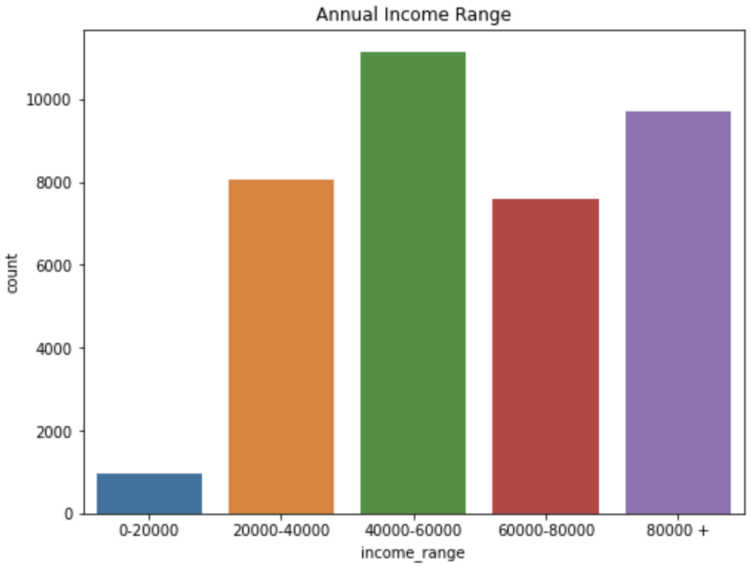




Observation:

1. The above plots indicate that interest rate distribution was higher for 10-15%, many people have borrowed at this interest rate
2. 2. There were few outliers in interest rate above 22.5%, which indicated few people have borrowed more than the usual interest rates

Loan Applicants Income & Debt Analysis

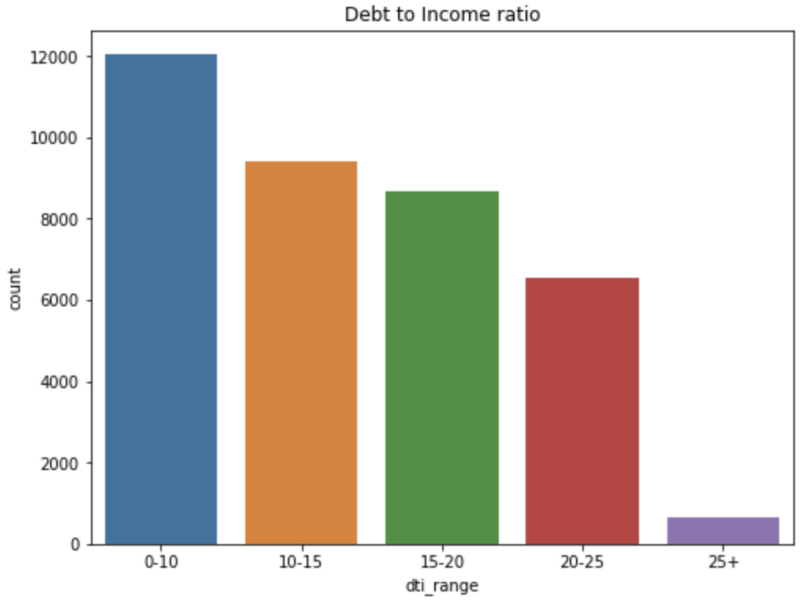


Observation:

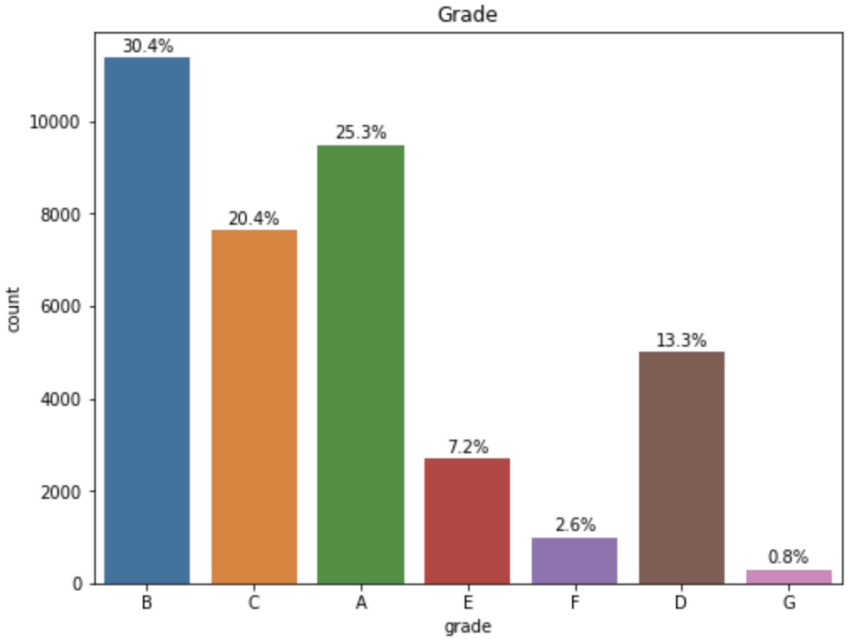
1. The left plot indicate that majority of the people are earning salary of around 40000 – 60000
2. There were few people who receive very less amount of below 20000

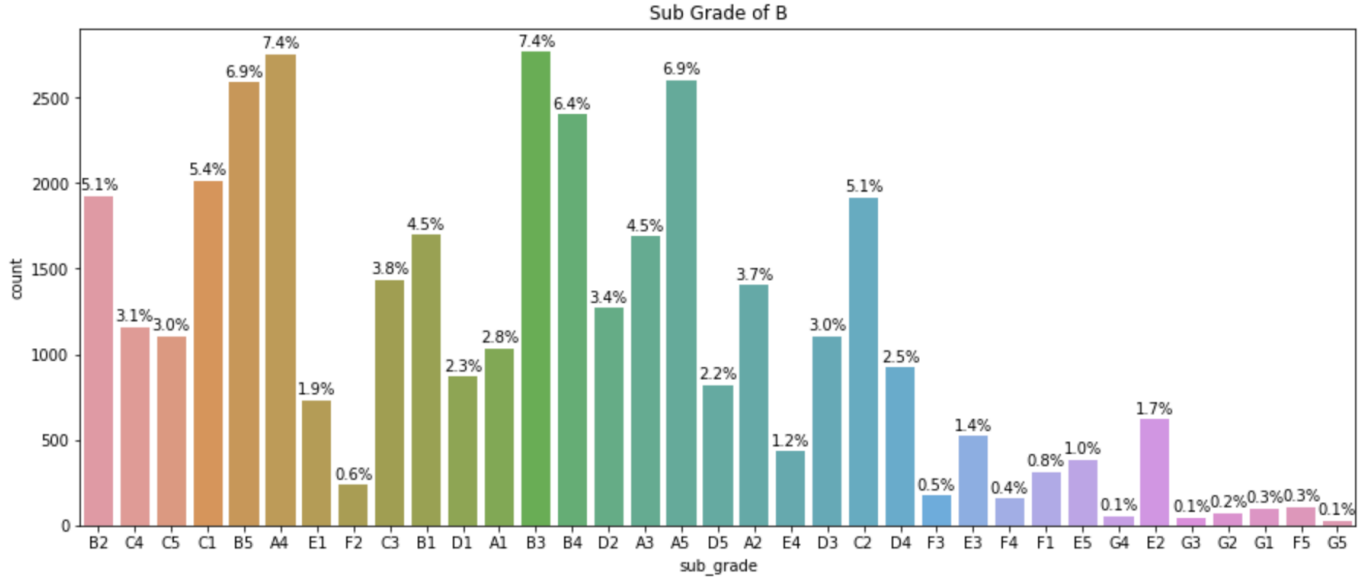
Observation

1. The right plot indicate that many people were having less than 10% of debt-to-income ratio
2. There were no major outliers in debt-to-income ratio



Loan Applicants Grade Analysis

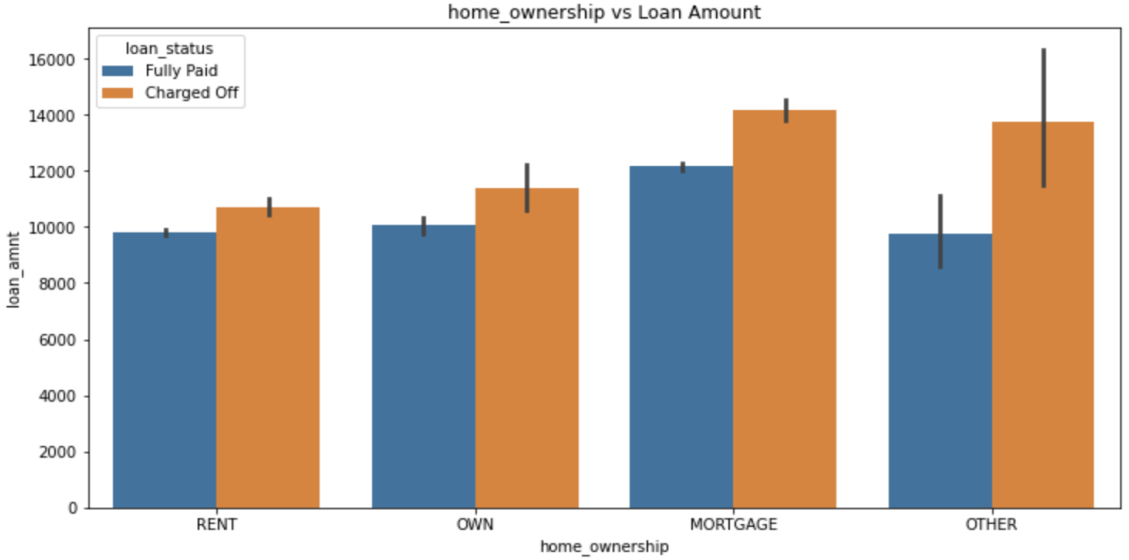




Observation:

1. The above plots indicate Grade B has the highest loan applications with 30.4%
2. Grade G has the least application count with 0.8%
3. The above plots indicate sub–Grades B has the highest loan applications, of these sub grade B3 has the highest applications
4. Grade G5 has the least application count with 0.1%

Purpose for applying Loan Vs Asset ownership

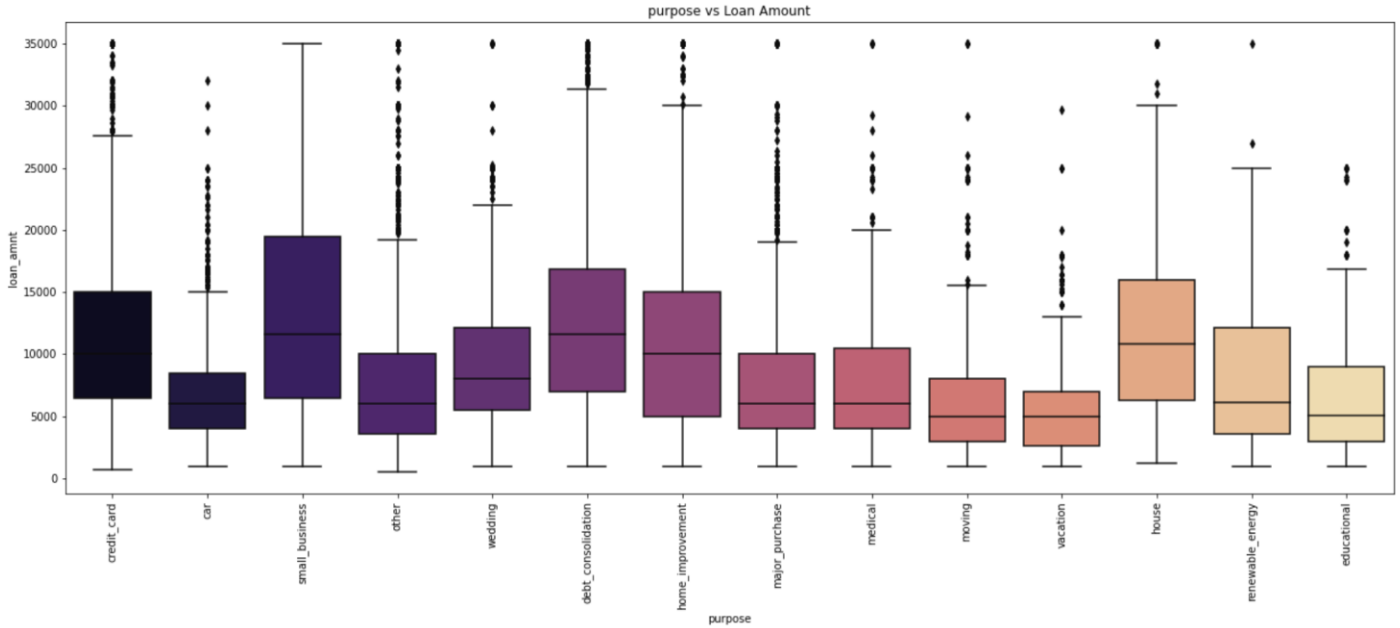


Observation:

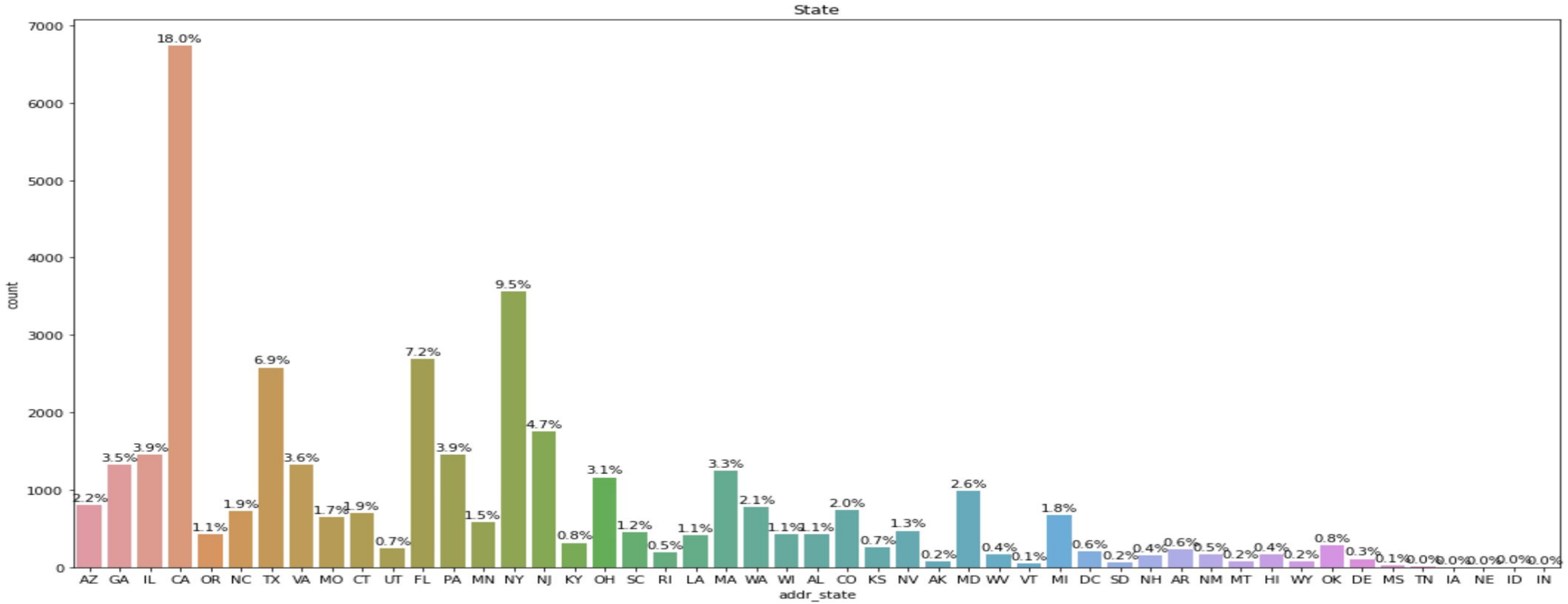
1. The Left plots indicate higher amount of loans are taken for small business
2. Very less amount of loans are taken for vacation

Observation:

1. The Right plots indicate that People who have mortgage on homes tend to take loans on higher amount.
2. And the same people of mortgage category tend to default loan amount more compared to others



Loan Defaulters by State Analysis

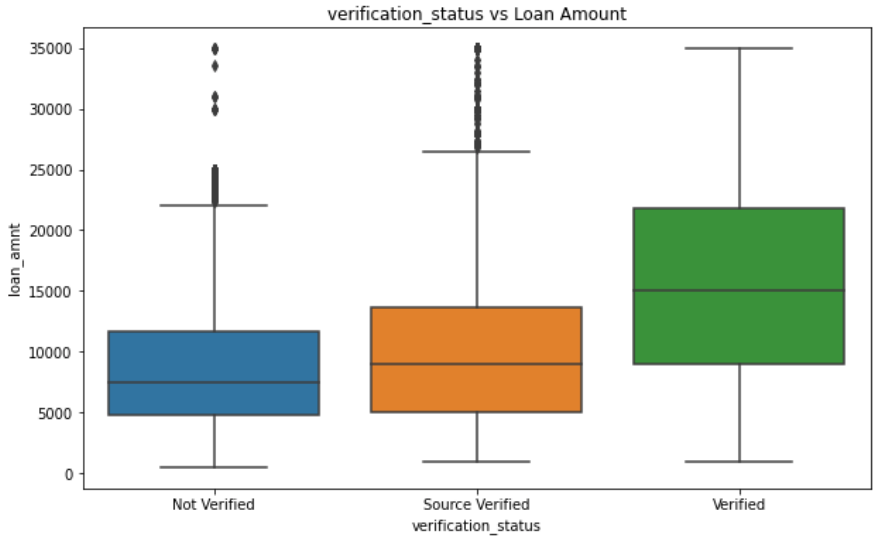


Observation:

1. The above plots indicate that CA state people have more loan applications than any other state
2. Most of the states have less than 1000 applications

Bivariate Analysis

Loan Verification Analysis

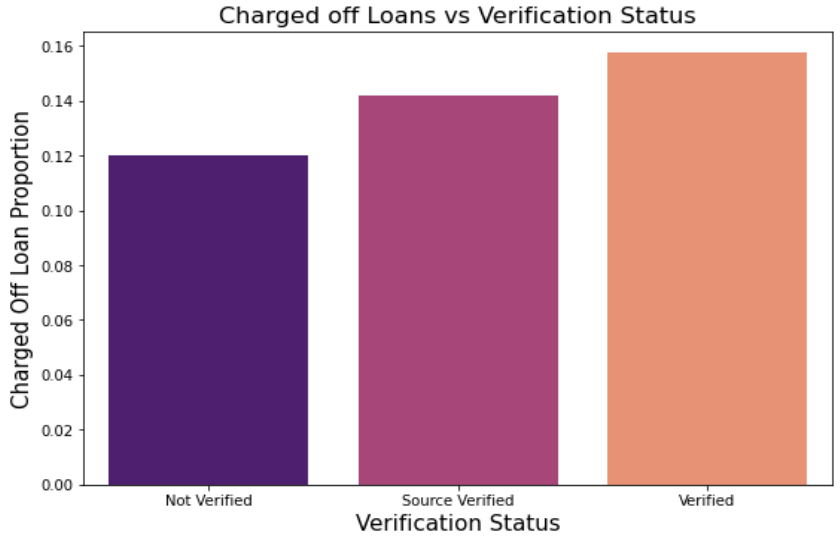


Observation:

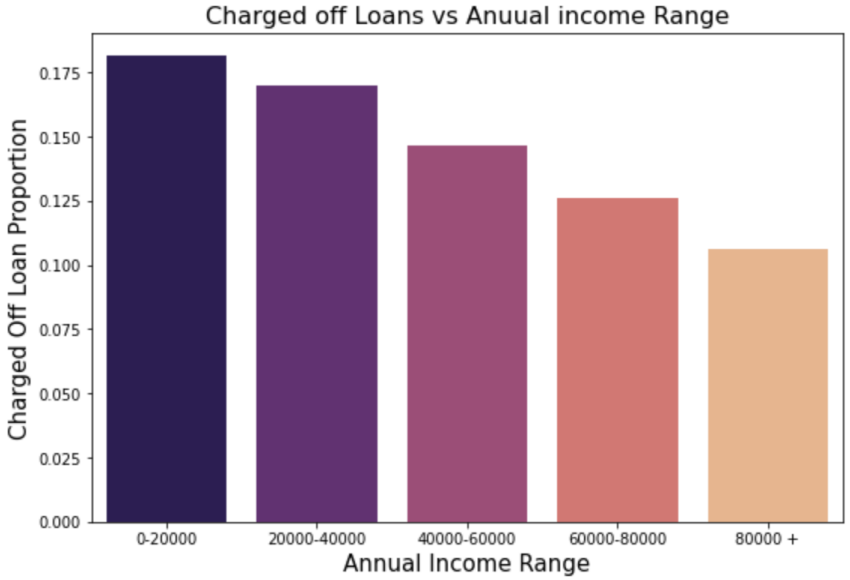
1. The above plots indicate most of the loans above 10000 are verified
2. Loans less than 10000 are mostly not verified

Observation:

1. The above plots indicate that loan charging off does not depend much on verification
2. There is no much inference with this data of verification



Annual Income vs Charged off Loans

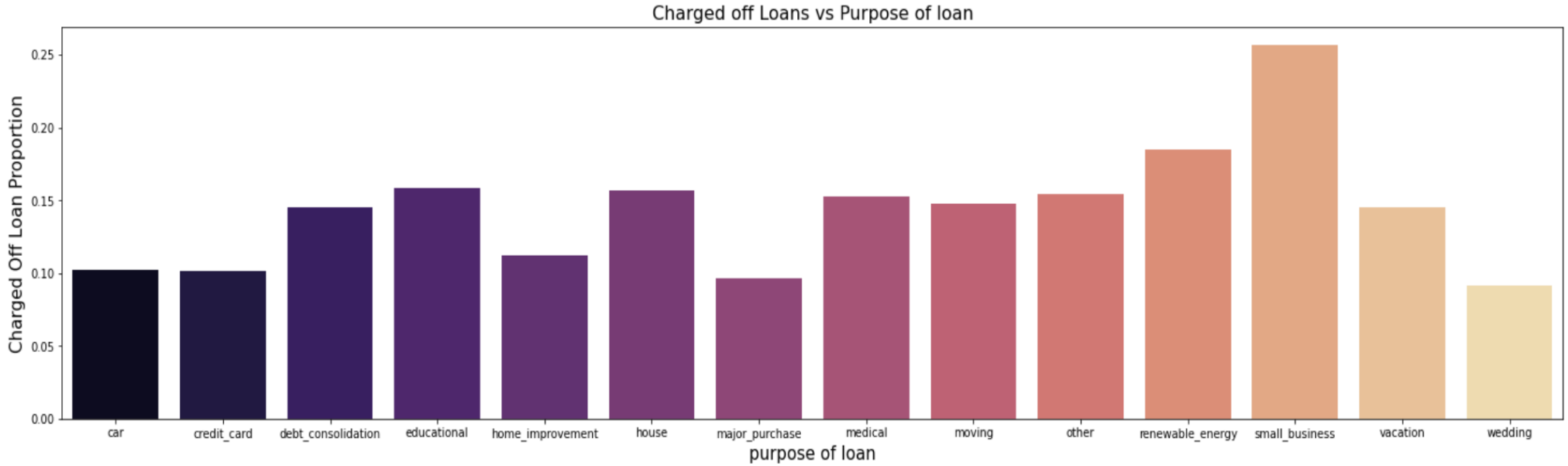


loan_status	income_range	Charged Off	Current	Fully Paid	chargedoff_prop
0	0-20000	173	6	774	0.181532
1	20000-40000	1369	156	6539	0.169767
2	40000-60000	1630	327	9167	0.146530
3	60000-80000	957	237	6400	0.126021
4	80000 +	1033	358	8316	0.106418

Observation:

1. The above plots indicate that increase in income range play a high influence on charge off of loans
2. Number of charged off loans decreases with increase in income

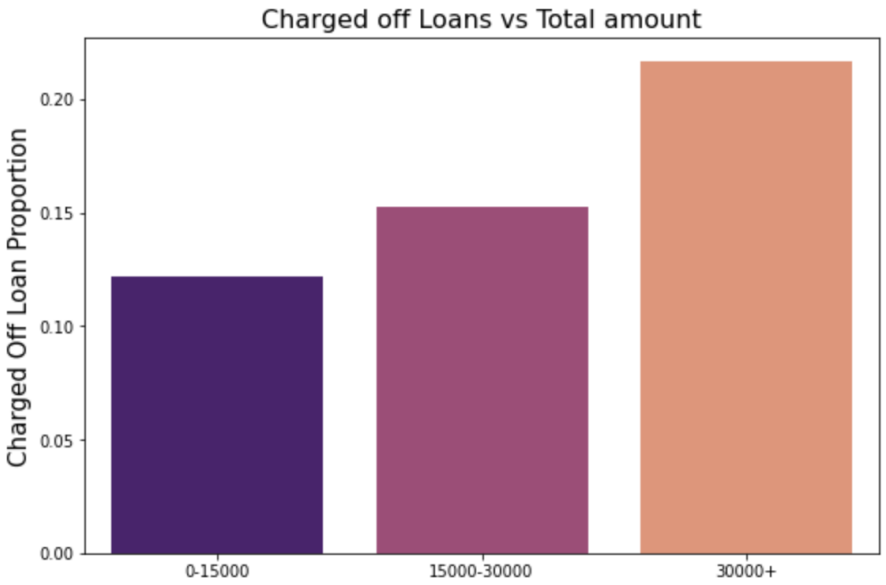
Purpose vs Charged off Loans

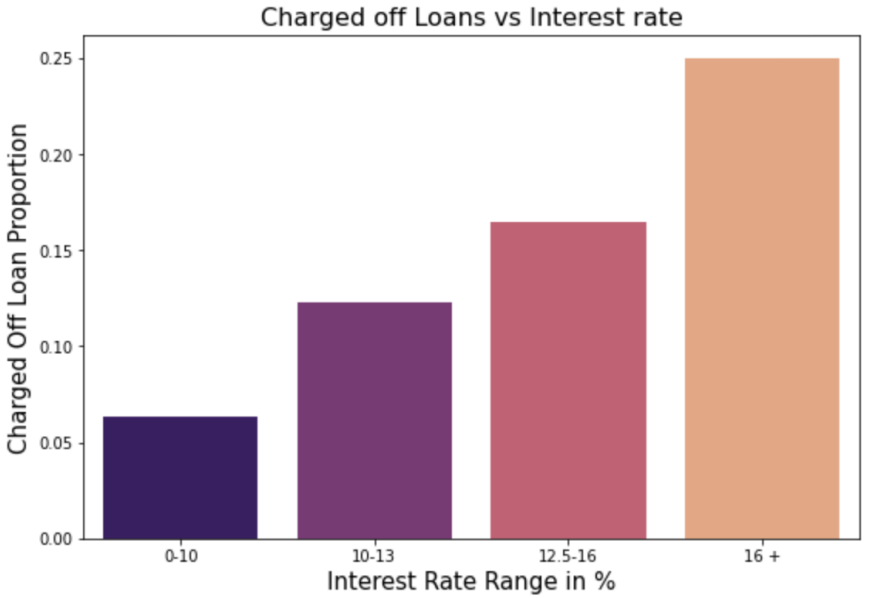


Observation:

1. The above plots indicate that people who took loans for small business were the highest defaulters
2. Consumer loans such as car, credit card, other purchases were among the less defaulting loans

Total Amount/Interest Rate vs Charged off Loans

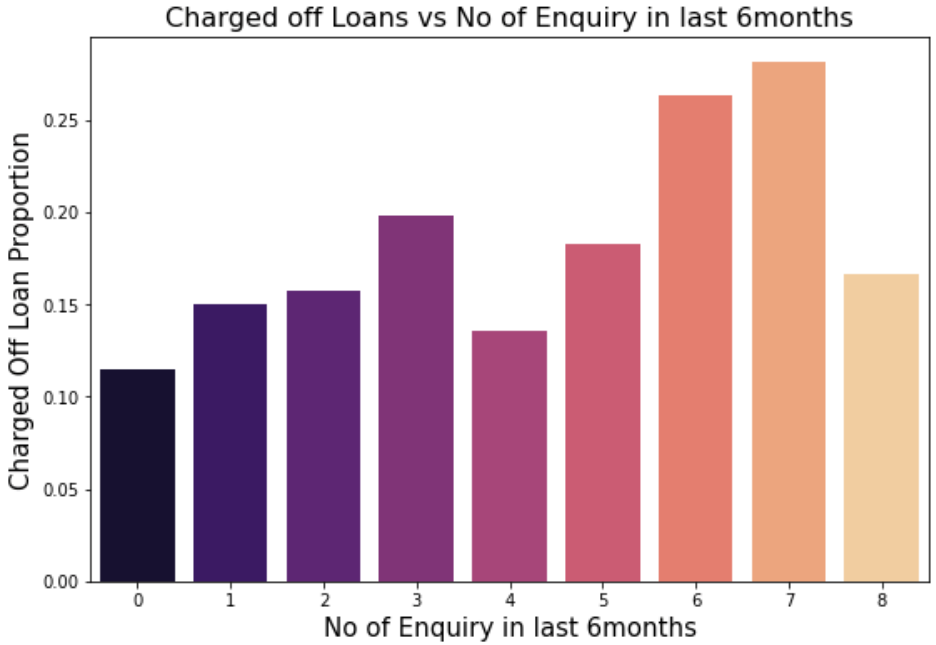




Observation:

1. The above plots in the left indicate that loan charging off increases with increase in Interest rate
2. Interest rate of less than 10% has less chances of defaulting, whereas interest rate of 16+ has more chances for defaulting

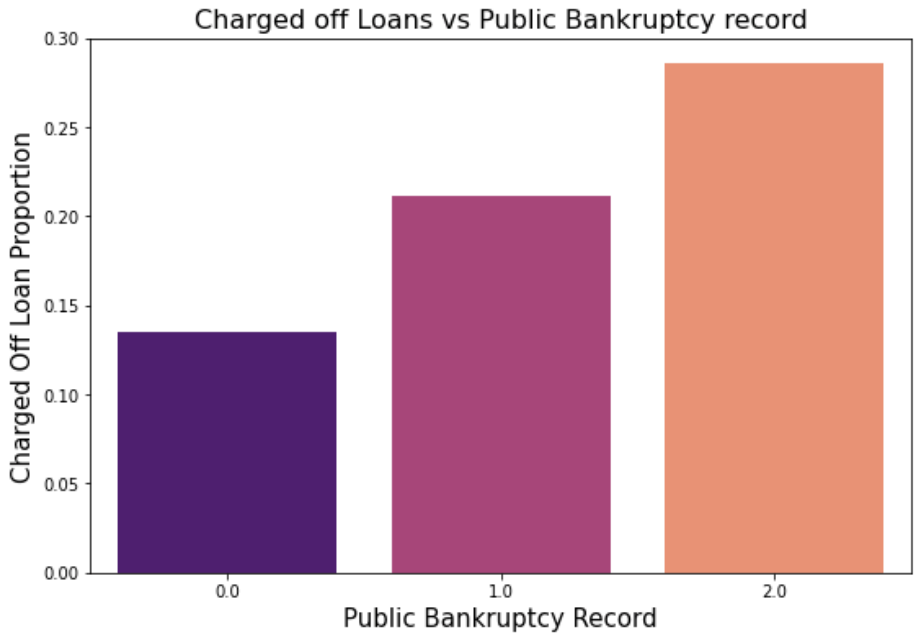
Enquiry Vs Charged Off Loans



Observation:

1. The above plots indicate that loan charging off increase with increase in number of enquires
2. People who made enquiries of more than 1 tend to default more, and people with more than 7 enquiry have the highest defaulting probability

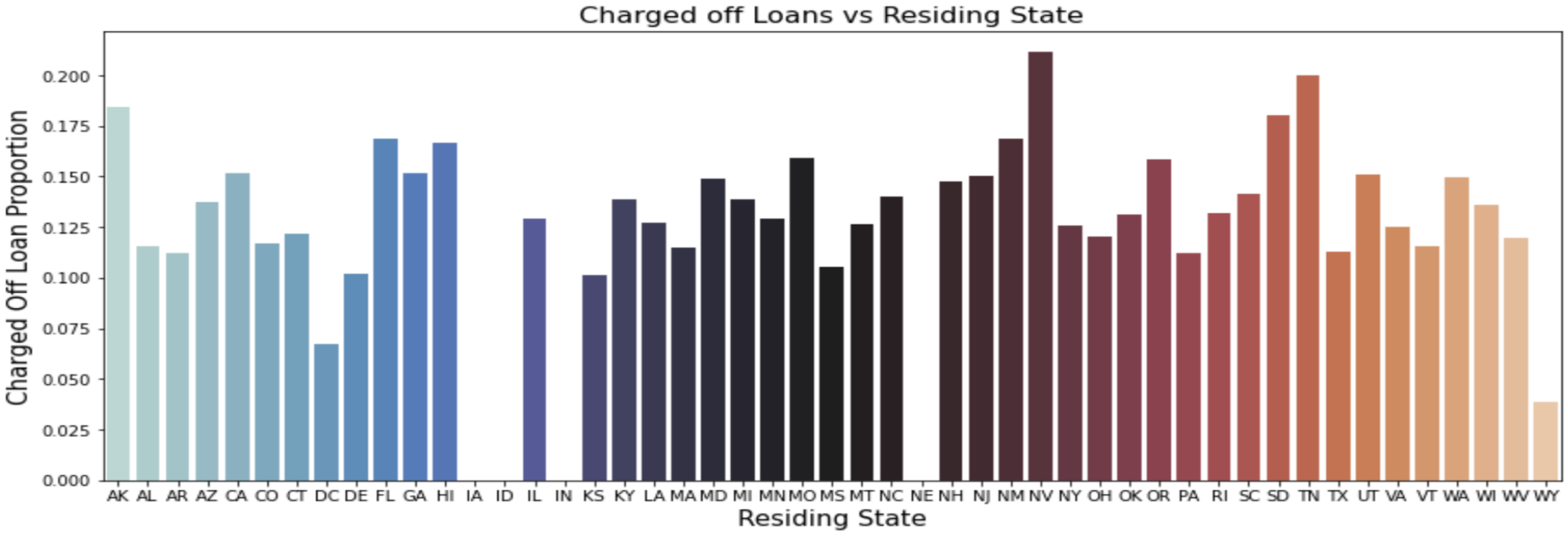
Public Bankruptcy Vs Charged Off Loans



Observation:

1. The above plots indicate that loan charging off increases with increase in number of bankruptcy
2. people with bankruptcy record of 2 tend to default more compared with 1 or no bankruptcy

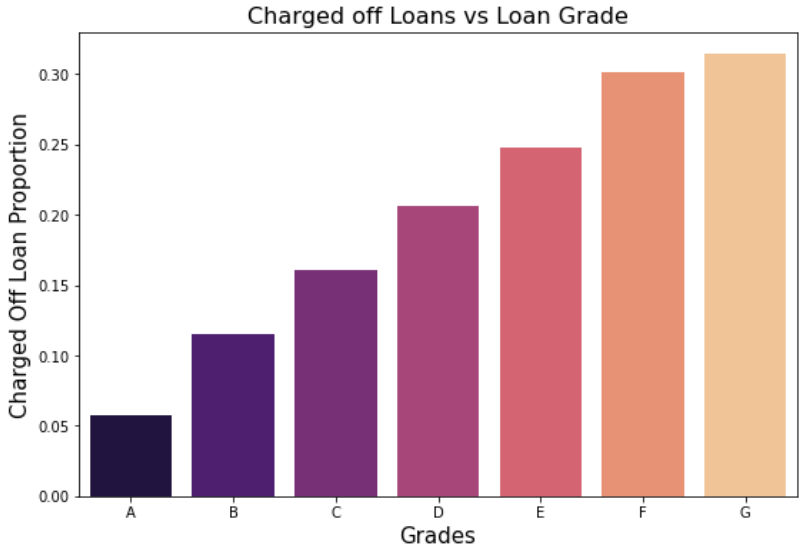
Residing City Vs Charged Off Loans



Observation:

1. The above plots indicate that loan charging off depend slightly on residing state, there is only a minimal difference
2. Few States have zero charged off applications, NV has the most charged off application

Grades vs Charged Off Loans

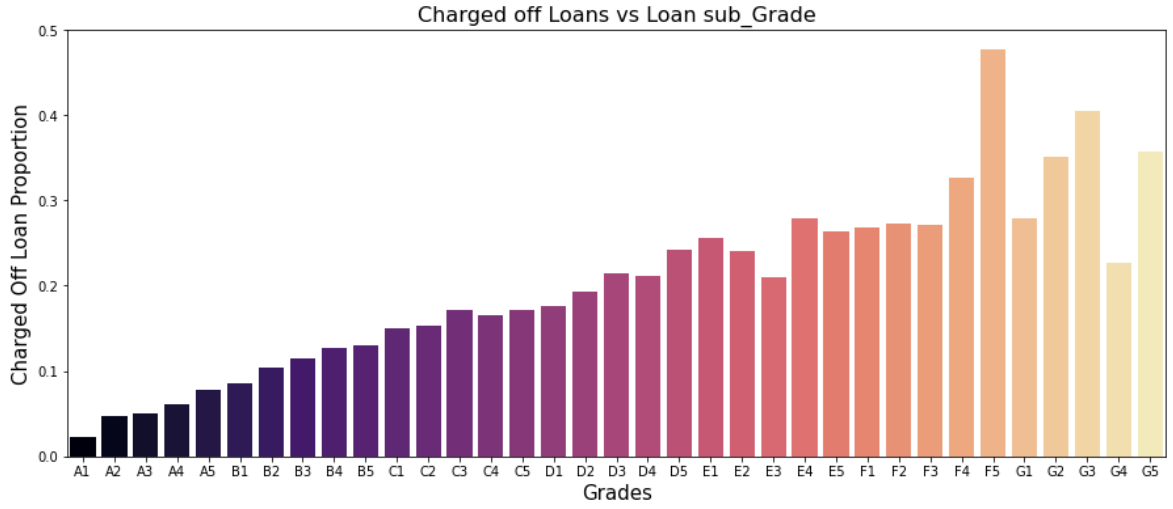


Observation:

1. The above plots indicate that loan charging off increases with increase in Grade alphabets
2. Grade A has less defaulters and whereas the last grade "G" has the greatest number of charged off loans

Observation:

1. The above plots indicate the same which grade vs loans status plot was indicating, loan defaults increases with increase in grade A being minimum
2. Plots indicate F5 which is a sub grade of F has the highest number of charged off loans



Conclusive Insight

1. **Loans with higher interest rate** are having more defaulters. Backgroud verificaiton has to be throughly as a deciding factor before lending the loan
2. **Lower graded loans** has more defaulters which shows the grading system is one of the important deciding factor for lending loans, lower graded loans should be handled with more verification.
3. **CA state applicants** should have addition verification as the tendency to default is higher. This is also for other states where defaulting of loans are high.
4. **People with past history of bankruptcy** must be handled with more verification
5. **High Debt to Income ratio leads to Defaulting**, this ratio must be verified for high risky applicants
6. **Credit hungry applicants**(>5 enquiry in last 6months) should be verified with additional documents.
7. **small businesses loans has more defaulters**, business loans should be validated through with the business plan and profitability