Lending Club Case Study

Submitted by:

1. Ashutosh Singh
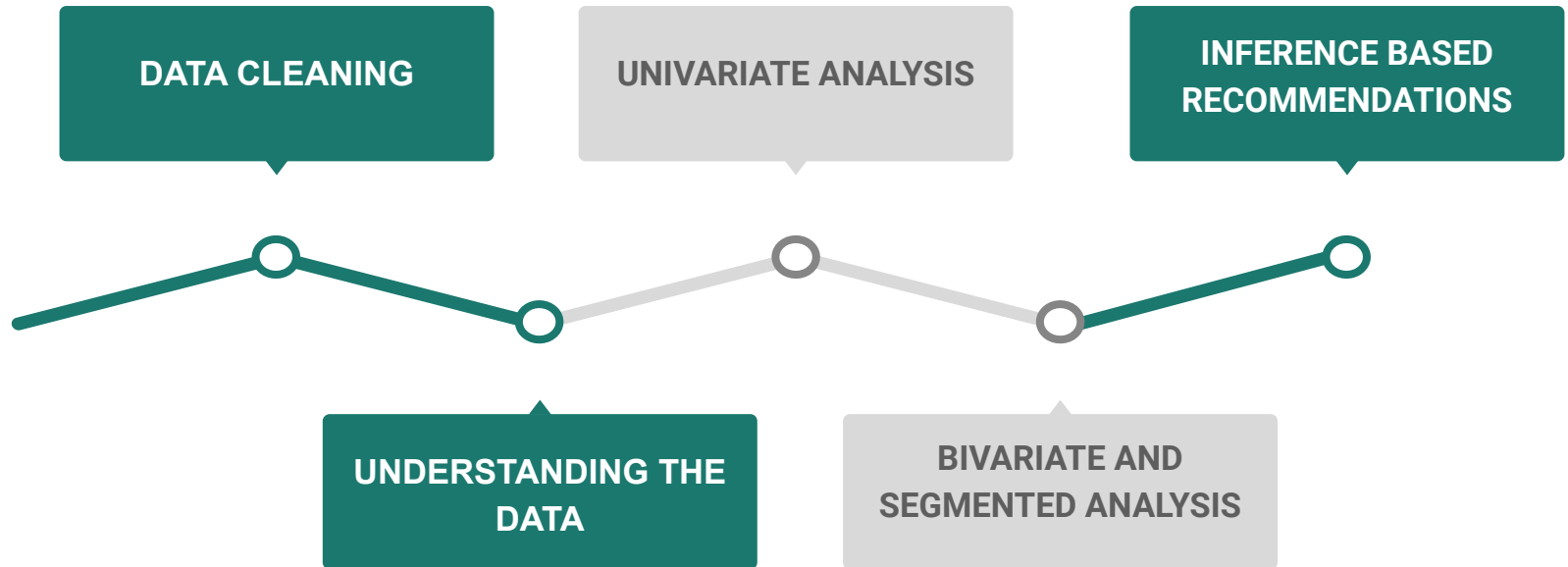2. Lynette Linzbuoy

Business Objective

To identify variables that indicate what makes an applicant more likely to default on loan payments and to use this analysis to aid in decision making regarding approval or rejection of applicants.

Analyse the data to identify which loan applicants are likely to default.

Analyse and present the variables affecting the increase in charge off rates

This analysis is to aid the LC in taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Methodology:

DATA CLEANING

UNIVARIATE ANALYSIS

INFERENCE BASED RECOMMENDATIONS

UNDERSTANDING THE DATA

BIVARIATE AND SEGMENTED ANALYSIS

Cleaning the data:

1. Removing columns with null values above 30%
2. Deleting all columns not applicable to the analysis
3. Removing rows with missing values
4. Standardising values
5. Splitting year & month columns
6. Removing all entries where loan status is listed as = Current
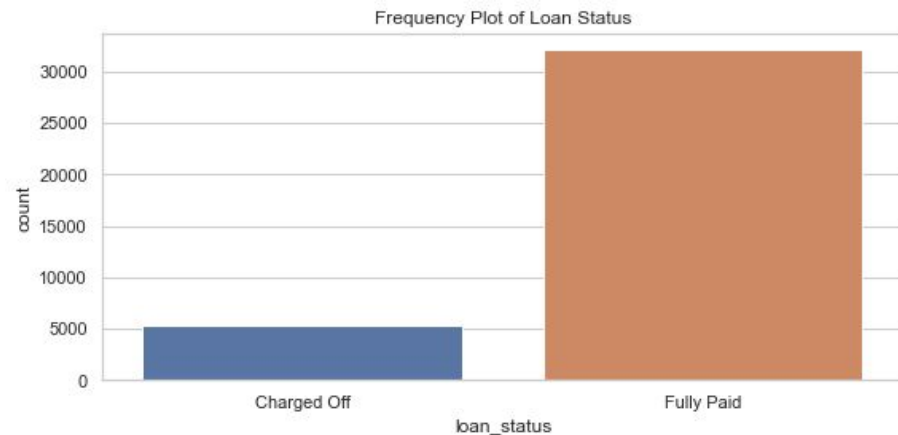7. Removing outliers

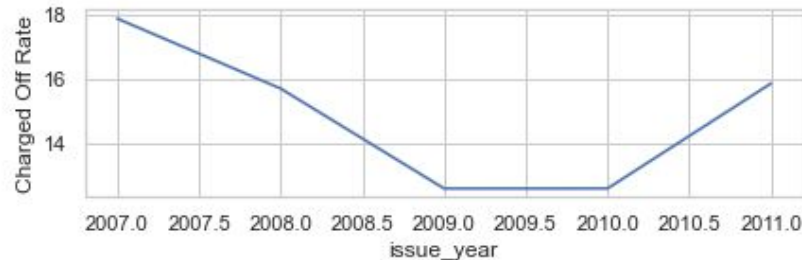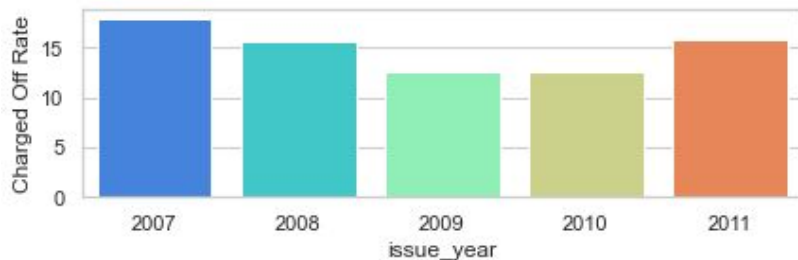Distribution Plot After removing outlier
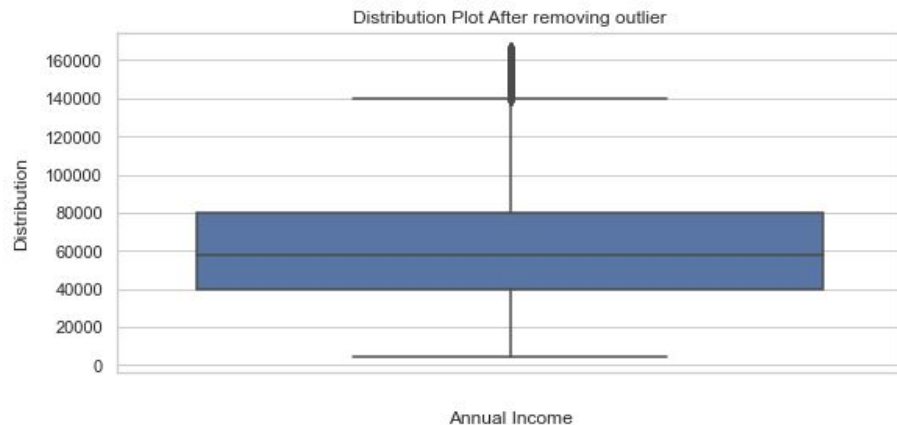
ANALYSIS

% of loans are getting Charged Off =

$$\frac{\text{No. Of Charged Off Loans} \times 100}{\text{Total number}}$$

14.35% loans are getting charged Off.

Frequency Plot of Loan Status

Charged off rate is high for starting and reduced in subsequent year but increased in last year 2011.
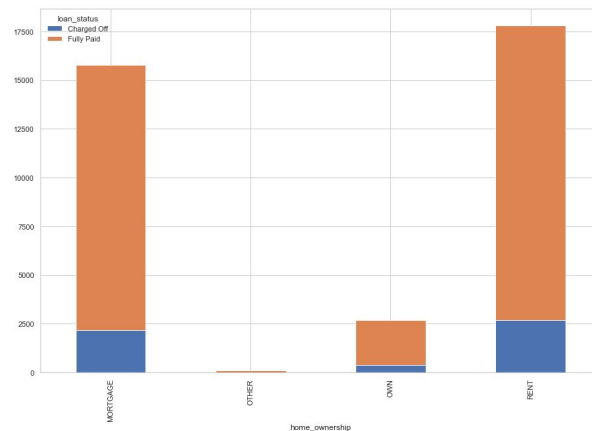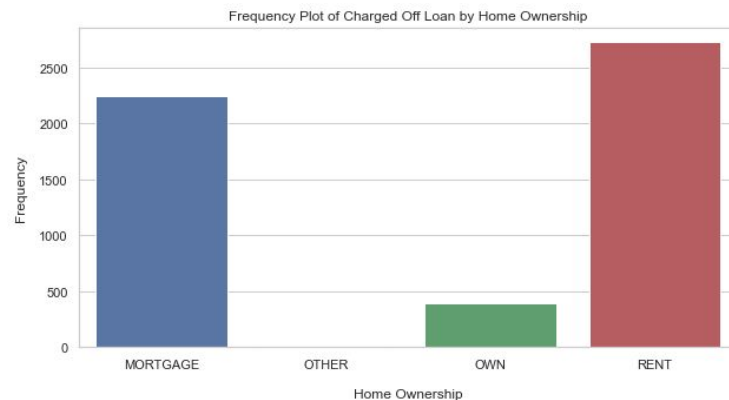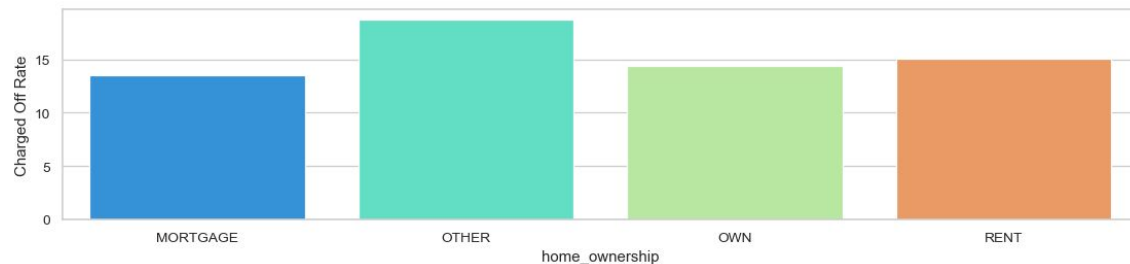
Charged Off Loans by Home Ownership

Most number of loan applications are from those who have a mortgage or are living on rent.

Applicants who have mentioned OTHER in their property section are most risky as they are most likely to default on the payment
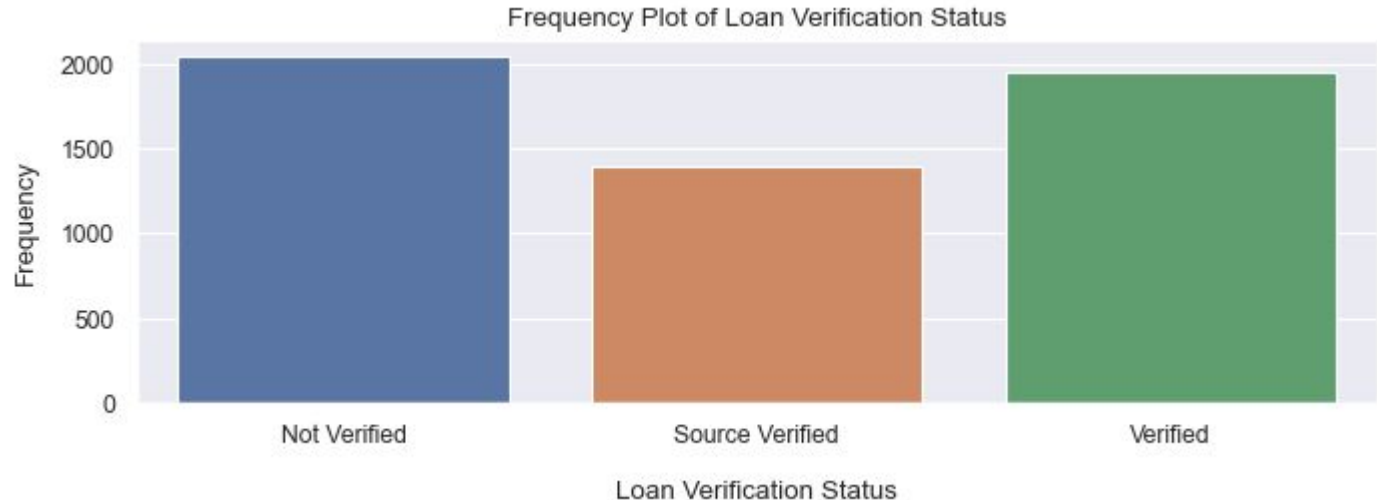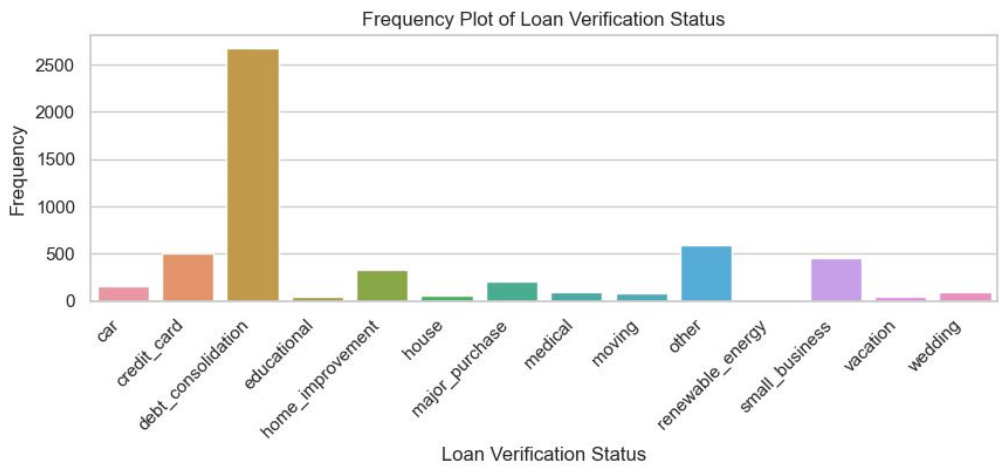
Charged Off Loans by Loan Verification status

The loans applicants with verified sources have less chance to be charged off.

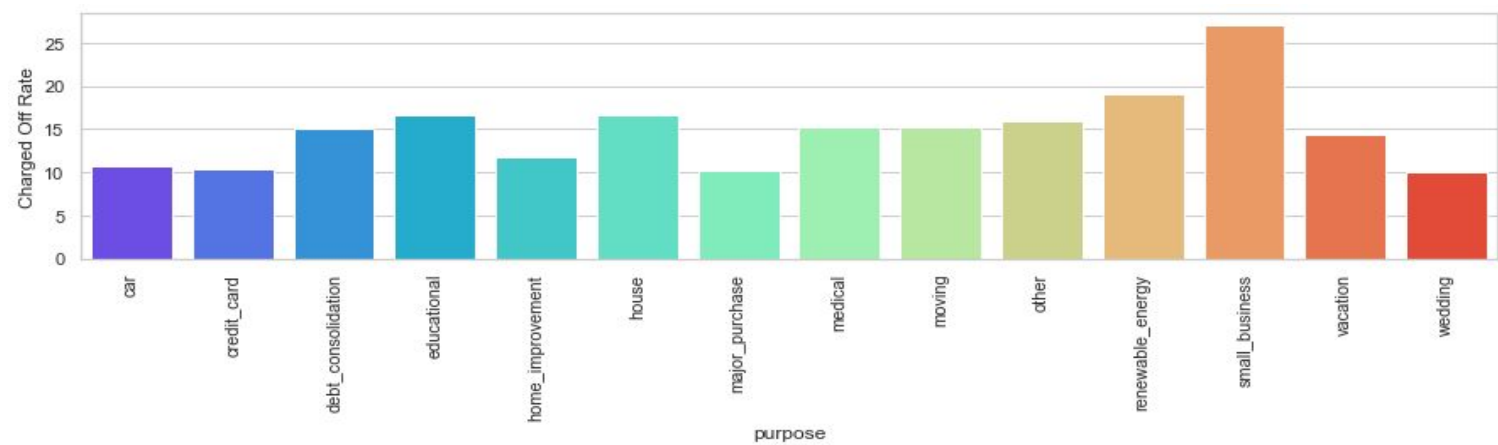Maximum loans were provided without verification

Frequency Plot of Loan Verification Status

Charged Off Loans by Loan Purpose

Frequency Plot of Loan Verification Status

Most of the number of loans is given for debt_consolidation purpose.

loan given for small_business (27.0334%) and renewable_energy (19.1489) purpose is a risk of getting defaulted.
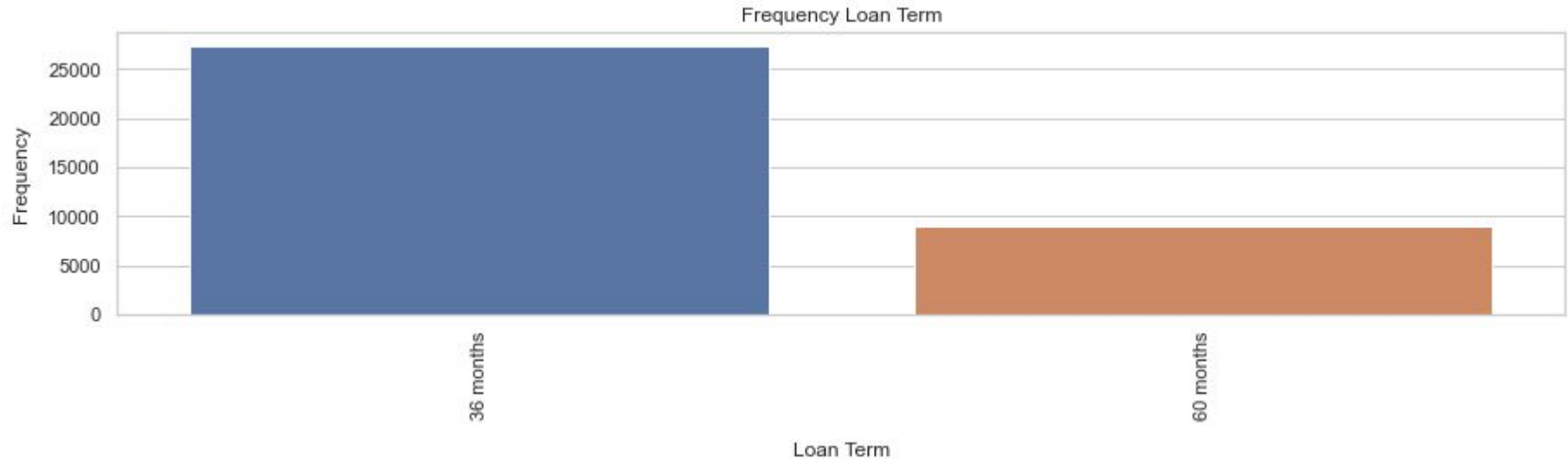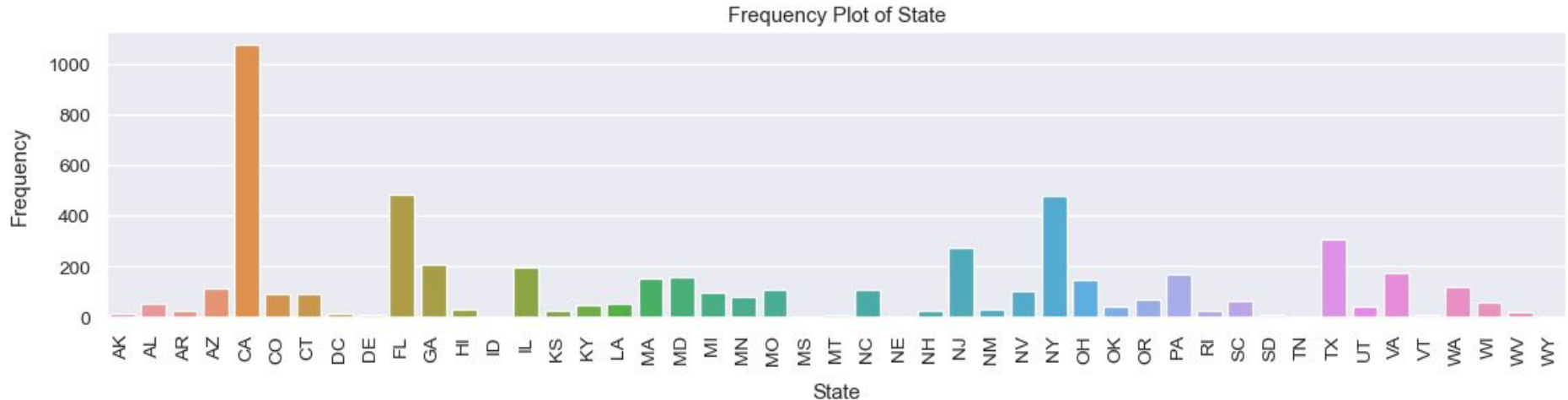
Charged Off Loans by term

Most people opted for 36 months as their loan term and whereas people who have opted for 60 months have high percentage of getting defaulted (22.70%)
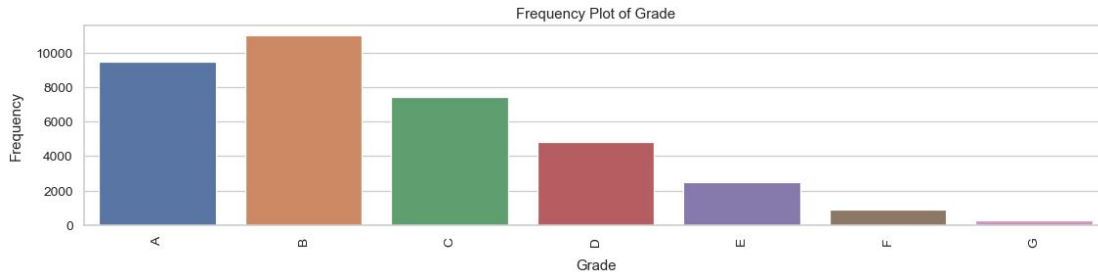
Charged Off Loans by State

Most of the loans are given to "CA" state whereas Loan given to "NE" state is risky as 60% time loans are getting defaulted. (NE = 60.00%, NV = 21.68%, SD = 18.33%)


Frequency Plot of State

Loans by Grade

Loan grade A and B are mostly given to customers and looks safe whereas applicants who apply for Loan Grade G, F and E are seems to be risky as they have high rate of charged off 33.7% , 32.6% and 26.7% respectively. loans which are of Sub Grade B3,B4,B5,C1 and C2 have high chance to Charged Off.


Frequency Plot of Grade

loans which are of Sub Grade B3,B4,B5,C1 and C2 have high chance to Charged Off.


Frequency Plot of State
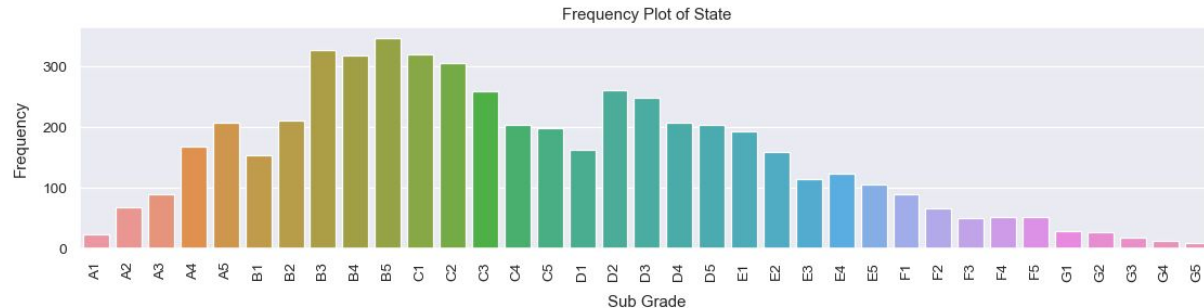
Charged Off Loans by Employment Length

Employees who are working for more than 10+ years are having high chance to Charged Off. But this data also include all people who are having work experience above 10+ i.e. 11, 12,13 etc. So this number is not certain. So we can't conclude anything on 10+ years but certainly we can say that people who are in initial phase of their employment they have a chance to Charge off.
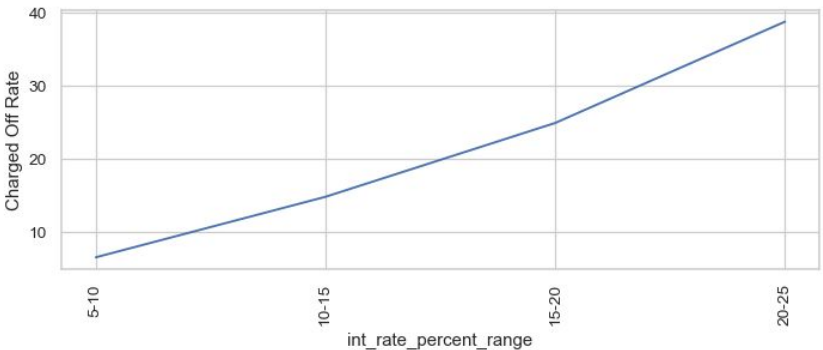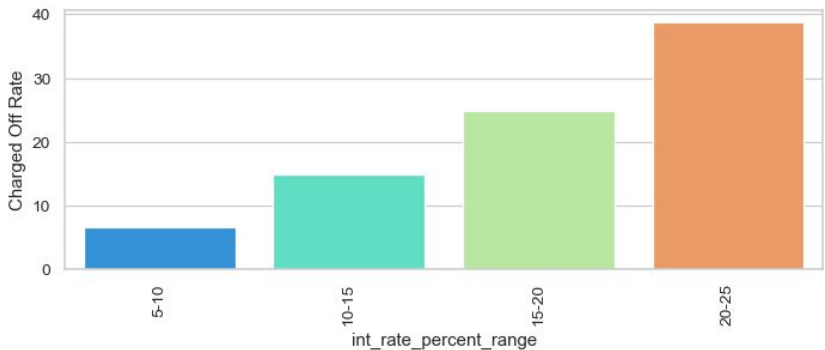
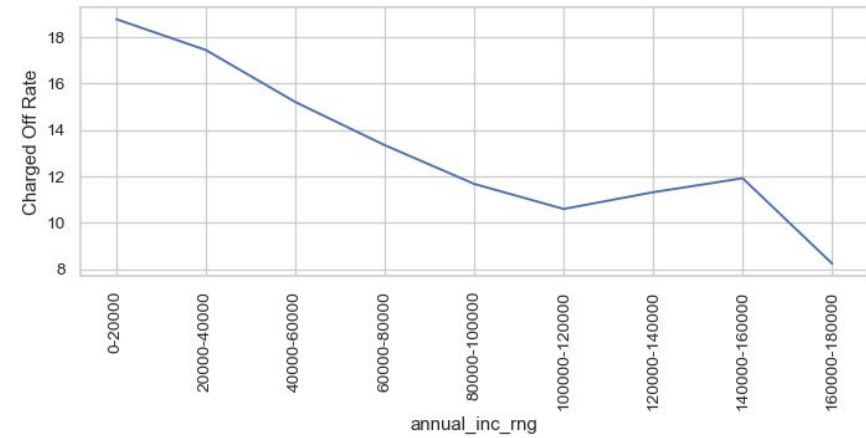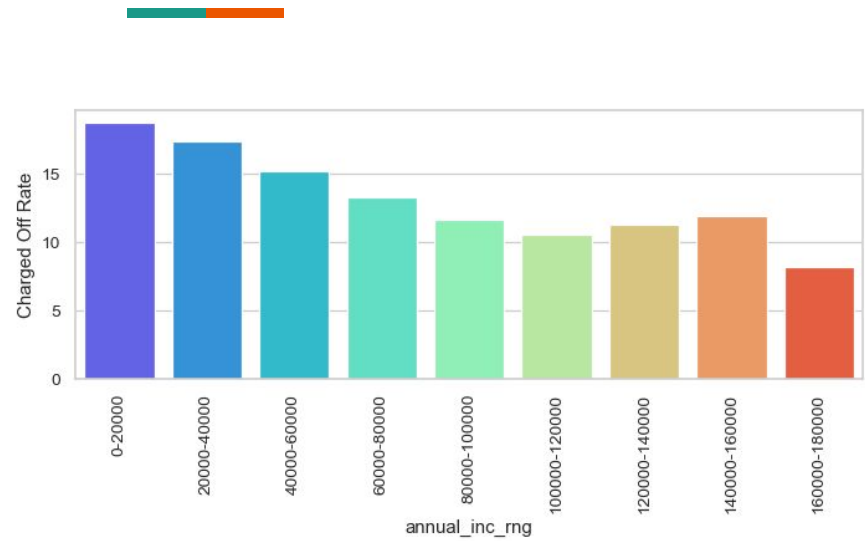

Frequency Plot of Employment Length
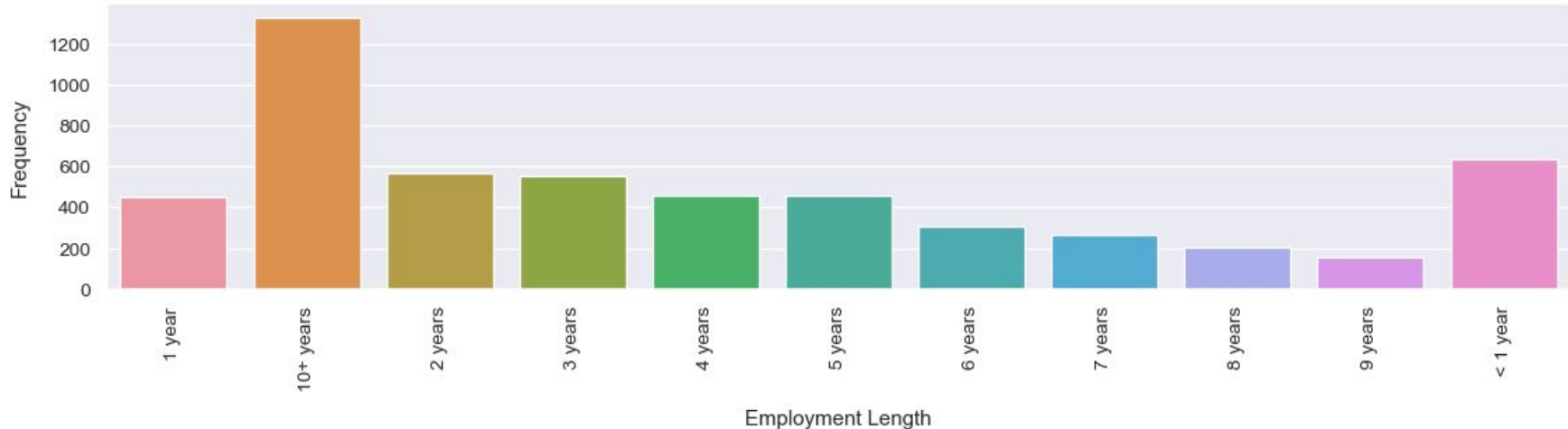
As loan amount increases , charged off percentage also increases.
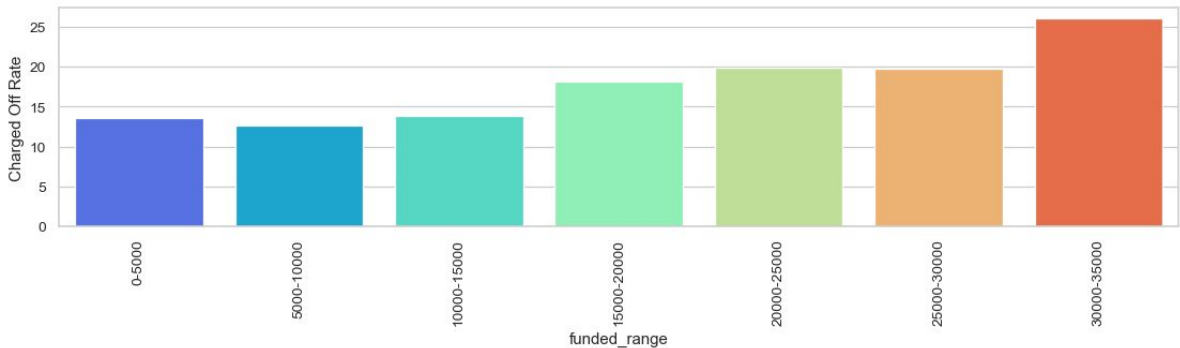
As rate of interest increases there are high chance of getting Charged Off.
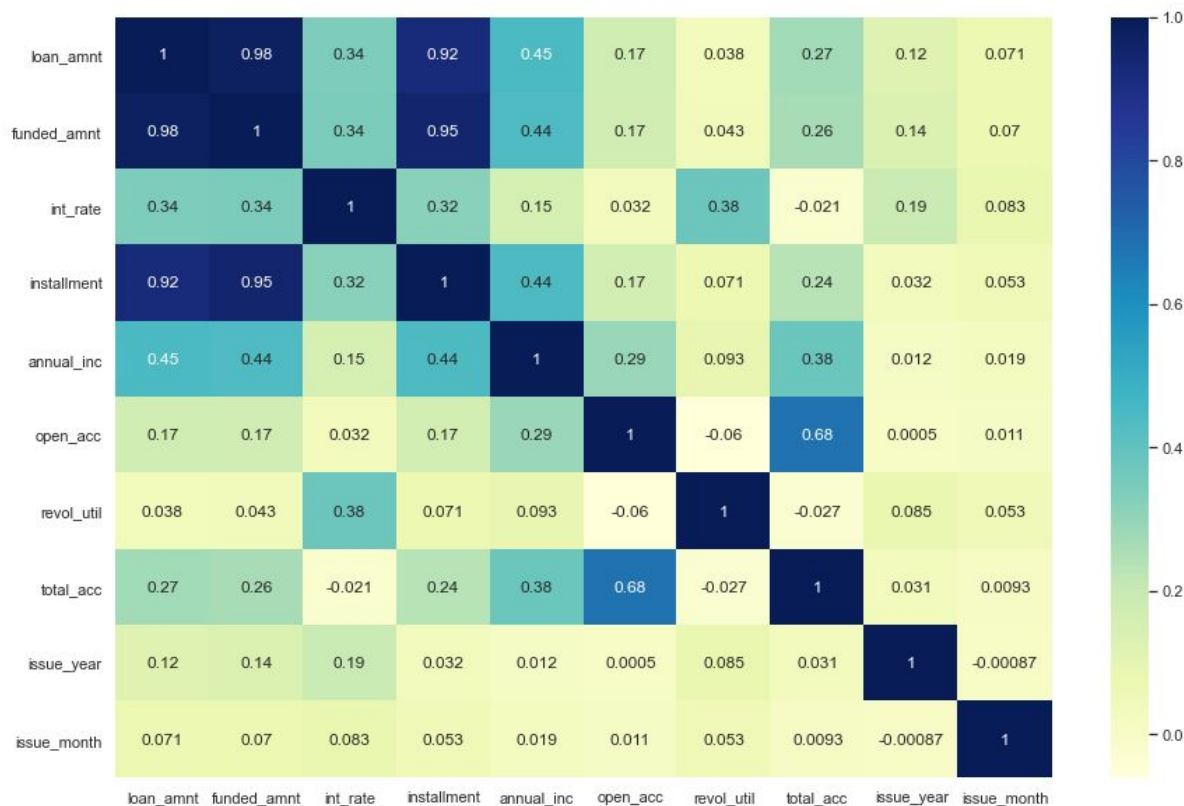
People whose income is less than 60K , are having high rate of getting defaulted

Bivariate Analysis: Continuous Variables

Loan_amnt and installment is highly correlated

INFERENCE:

- The background of applicants applying for the purpose of debt consolidation should be checked thoroughly as this category has the highest defaulters.
- People applying under the category of small businesses should be scrutinized before approval of loans
- People applying from the state of CA must be evaluated thoroughly
- When home ownership is listed as Other, chances of defaulting on paying back the loan are high
- The lower the loan grade, higher the chance for defaulting