LENDING CLUB CASE STUDY SUBMISSION

Group Members:
1. Palak Jindal

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

Problem solving methodology using CRISP-DM
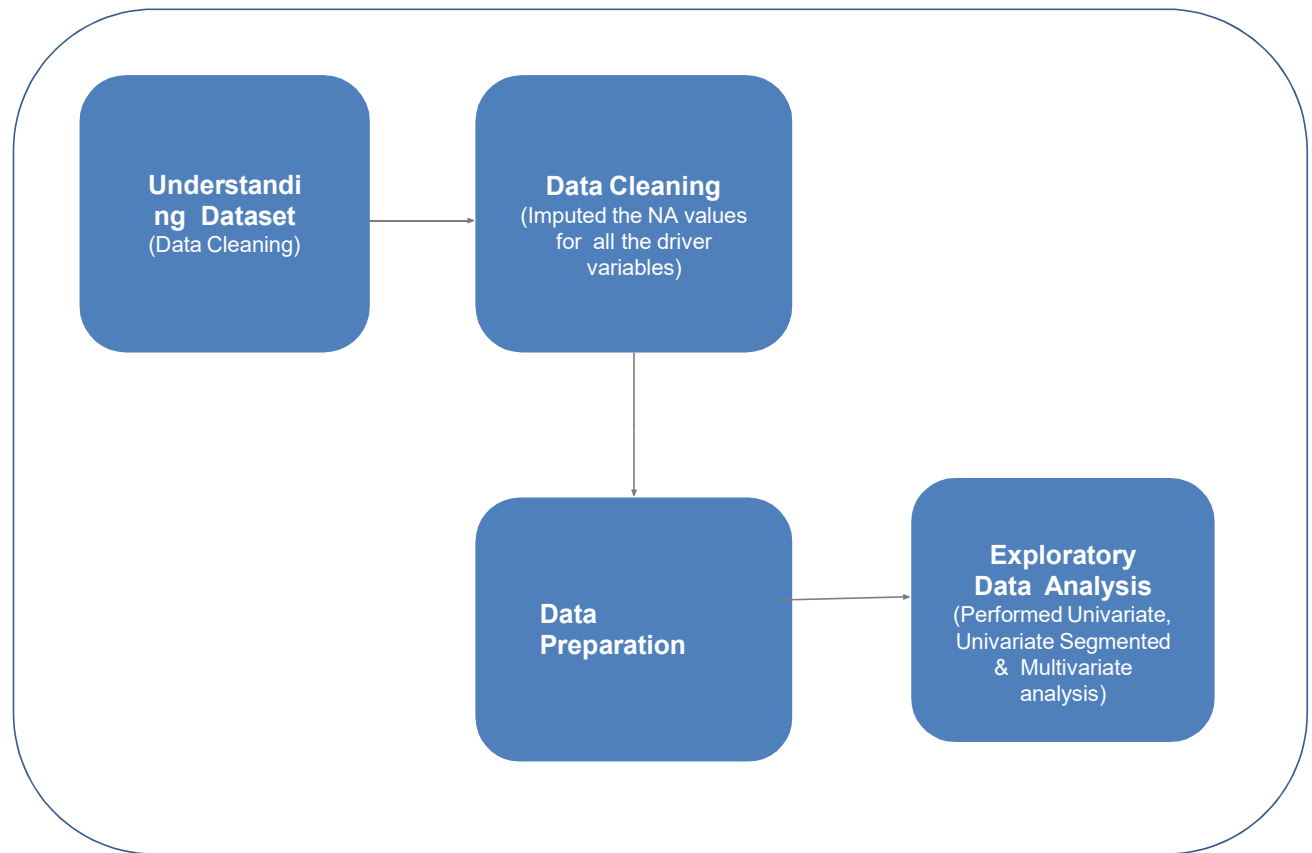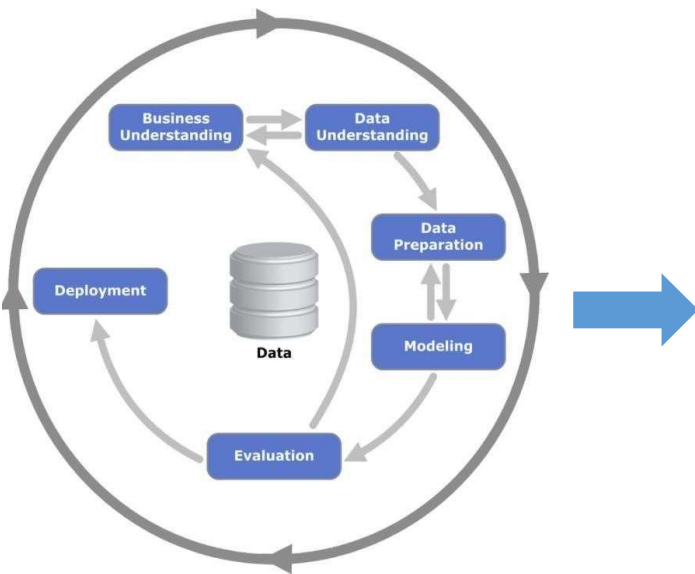
Understanding Dataset

- Read the loan dataset given into a pandas dataframe

- Go through the Data dictionary given in order to understand the significance of each attribute.

Data Cleaning Steps

- Perform the Sanity checks like (shape, column names, data) on the data extracted in the dataframe from csv file

- Remove the columns which have missing values > 50%

- Remove the columns with unique values (like 'desc', 'member_id', 'id', 'url','emp_title') as they are of no use in the analysis

- Remove columns with the same values for all the records(like 'delinq_amnt', 'acc_now_delinq', 'application_type','policy_code', 'pymnt_plan', 'initial_list_status', 'chargeoff_within_12_mths', 'tax_liens', 'collections_12_mths_ex_med') in the dataset. As these columns will not create any difference in the analysis.

- Clean redundant keywords from the data like 'months', 'years', '+','<','%'

- Create derived variables for better and granular level analysis like issue_d_month, issue_d_year etc.

- Change the data type of columns as per their values.

Analysis

- The essence of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting. Therefore, removed records with loan_status as Current.

- We performed **data cleaning and preparation** on the Loan dataset:
  - Imputed the NA values for all the variables
  - Created two new columns:
    - Profit and Loss column
    - Ratio of funded amount and annual income

- During **univariate analysis** we have created:
  - Histograms and Bar charts to check out the distribution of all the driver variables
  - Box plots to detect the Outliers
  - Performed the Multivariate analysis to understand how different variables interact with each other.
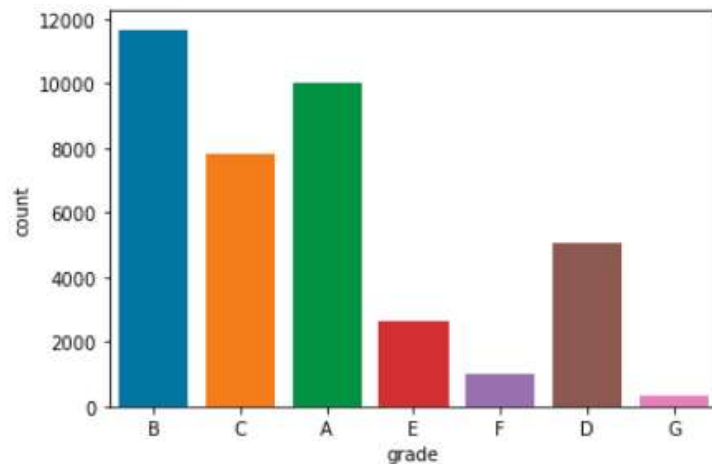
Loan Status Analysis
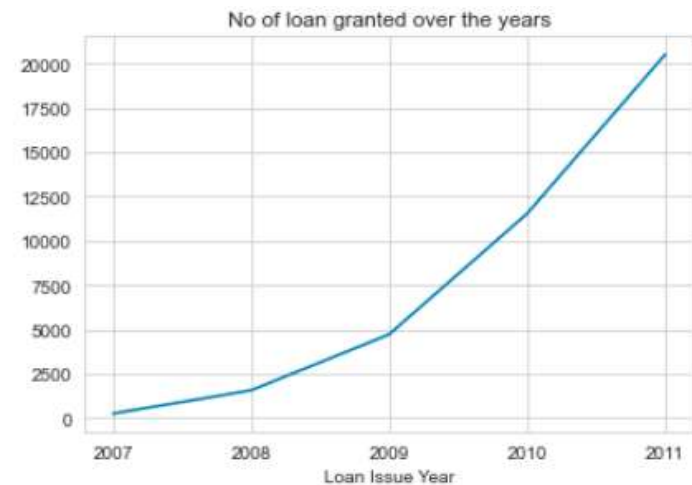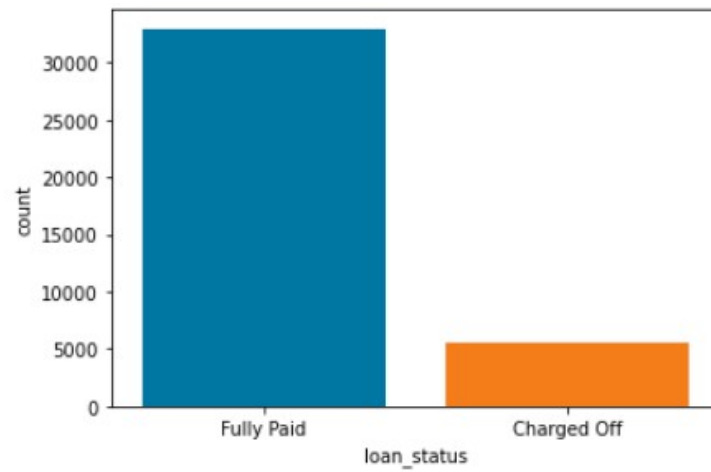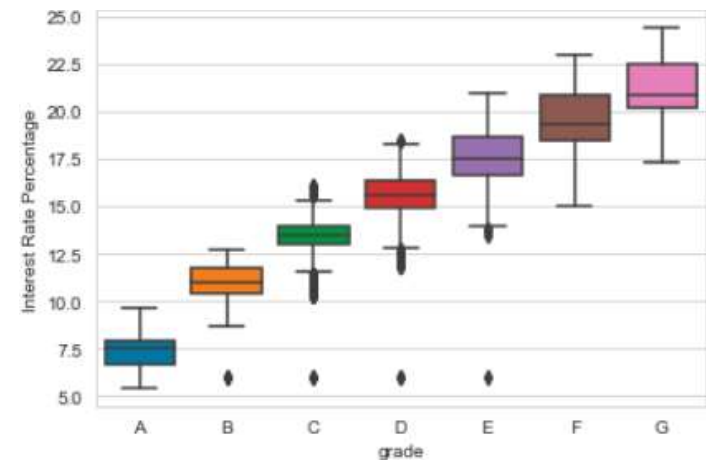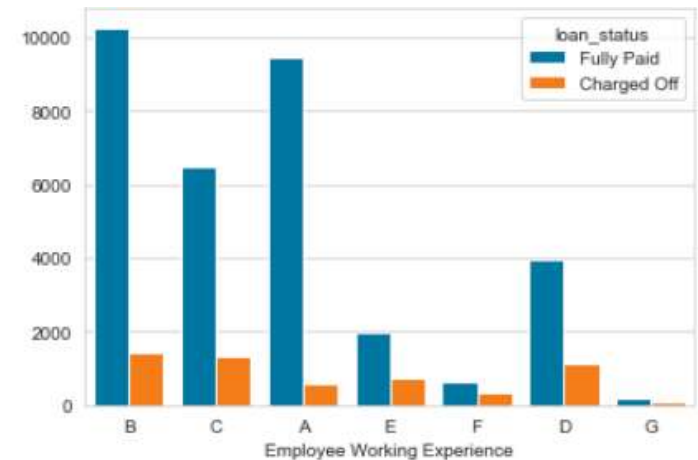

Observations:
1. Most of the loans are Fully Paid.
2. About 14% of loan are having status as defaulters.
3. The loan has been increasing exponentially over the years.
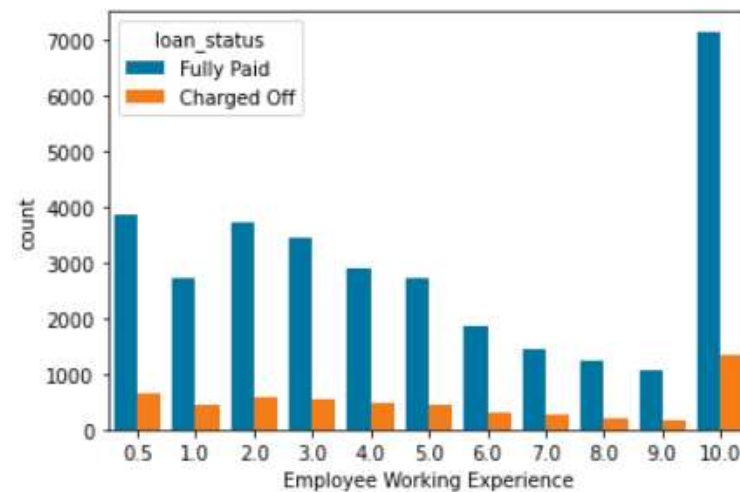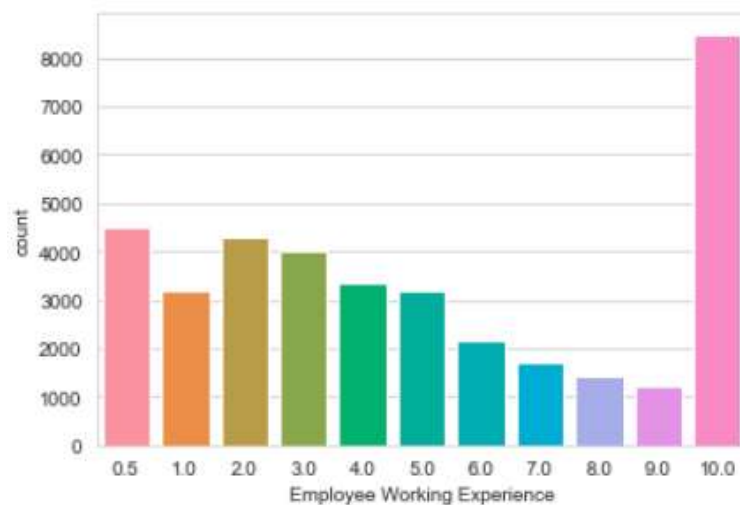
Loan Grade Analysis

Observation:

1. Most of the loans have grade of A and B. Therefore stating most of the loans are high graded loans.
2. High Quality loans have low interest rate.
3. Which depicts, higher interest rate have higher tendency to default the loan.

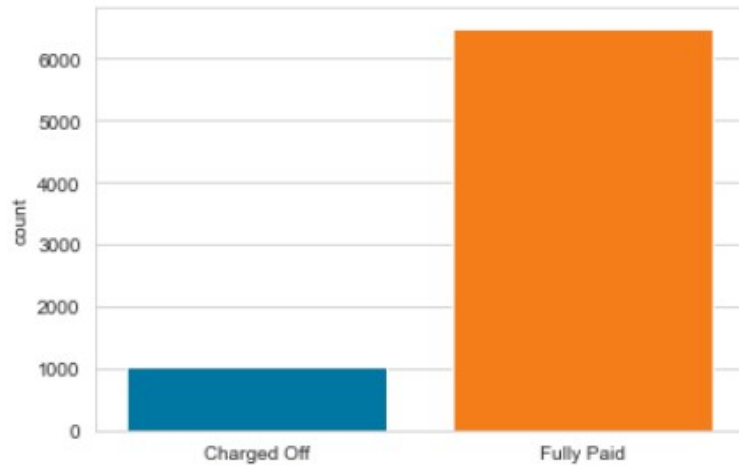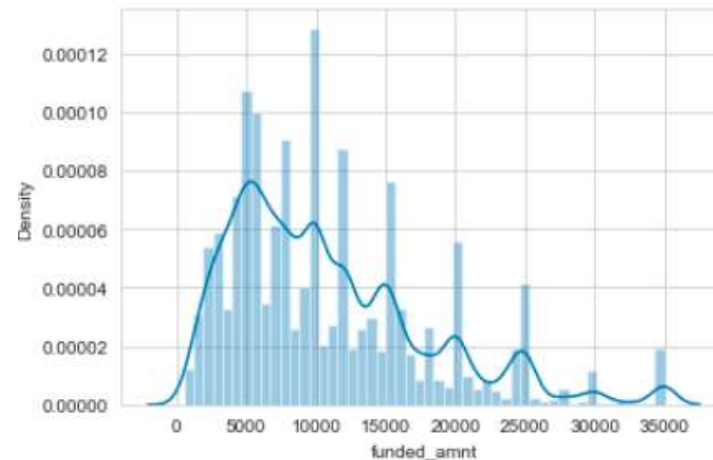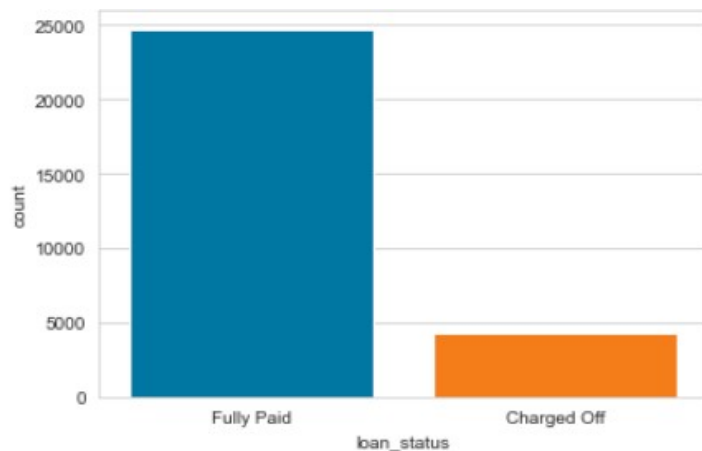# Loan Applicants work Experience Analysis

Observation:
1. Majority of employees applying for the loan have more than 10 years of experience
2. Tendency of person to default the loan with 10 years of experience is also high. So company need to be careful when granting loan.

Loan Amount Analysis
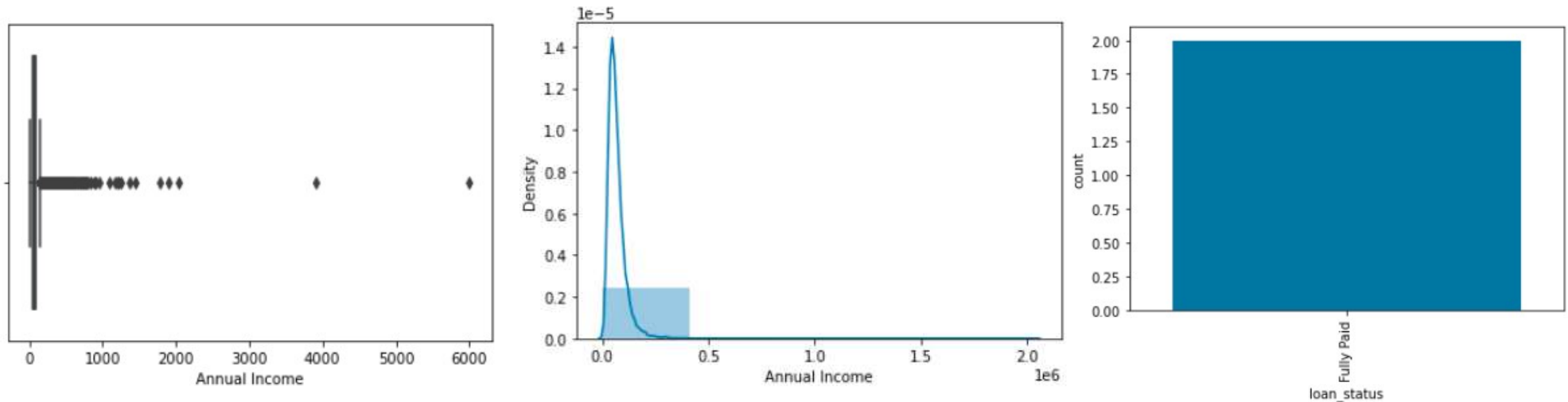
Funded amount <= 5000



Funded amount > 5000

Observation:

1. Funded amount is left skewed. Most of the loan amount given is below 5000.
2. Probability of people with loan amount greater than 5000 tends to default is more than the people with less than 5000
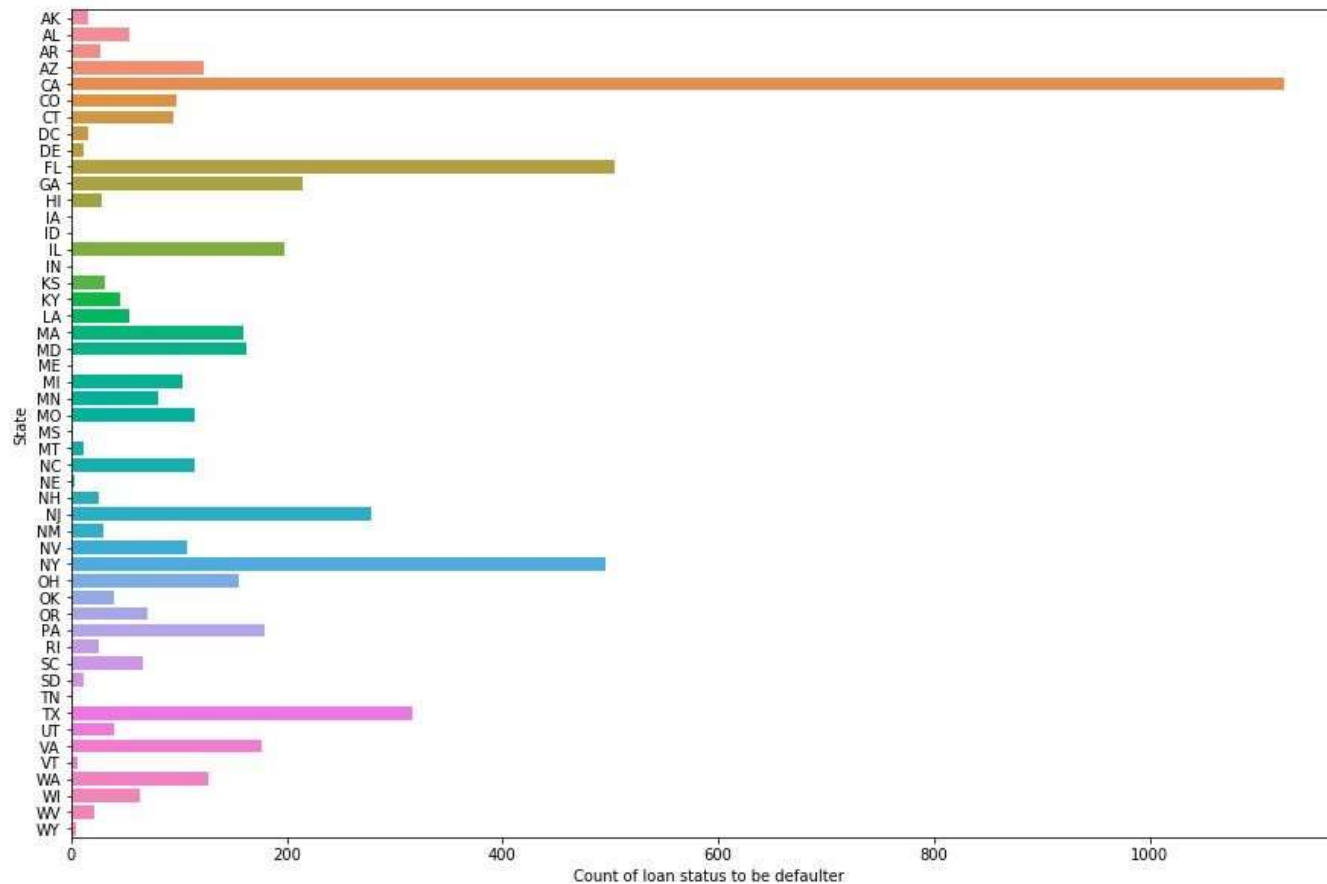
Income Analysis for the loan applicants

Observation:

1. Most of the applicants have income between 0 to 5 lakhs.
2. There are two applicants having income above 30 lakhs
3. Probability of people with annual income less than 3 lakhs tends to default is more than the people with more than 3 lakhs
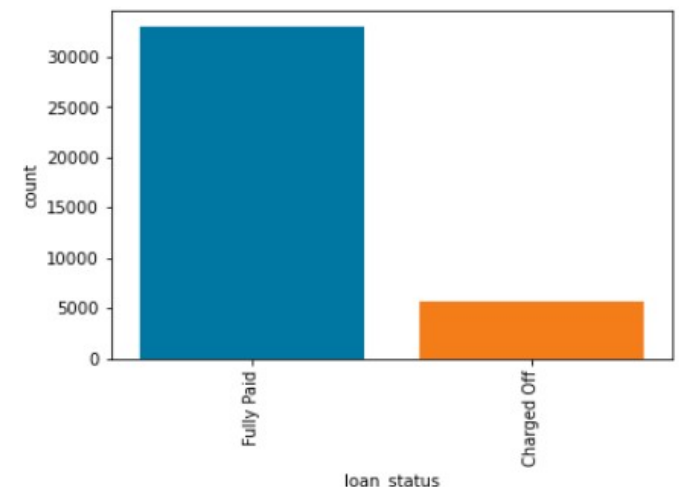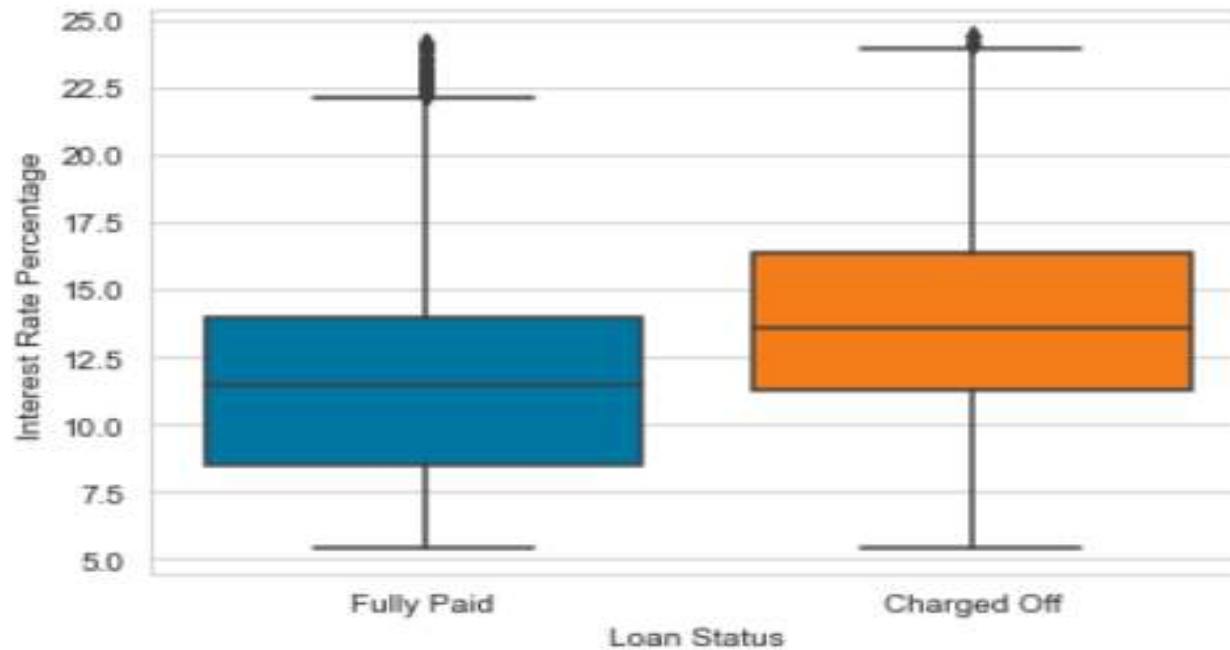
Loan Defaulters by State Analysis

Observation:
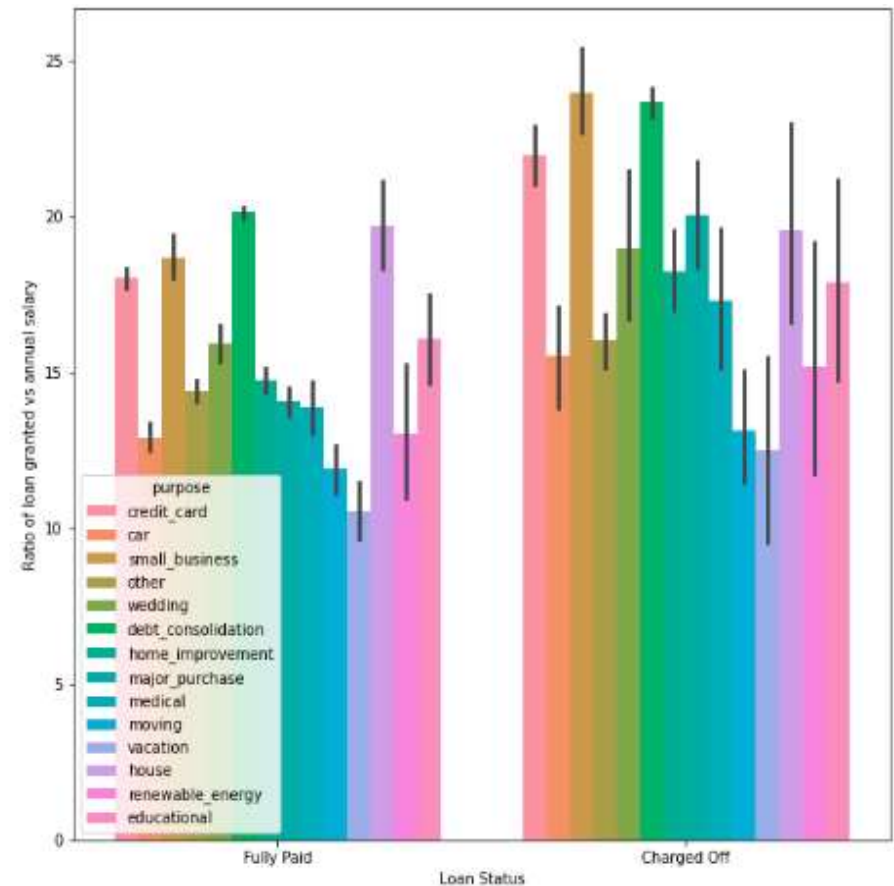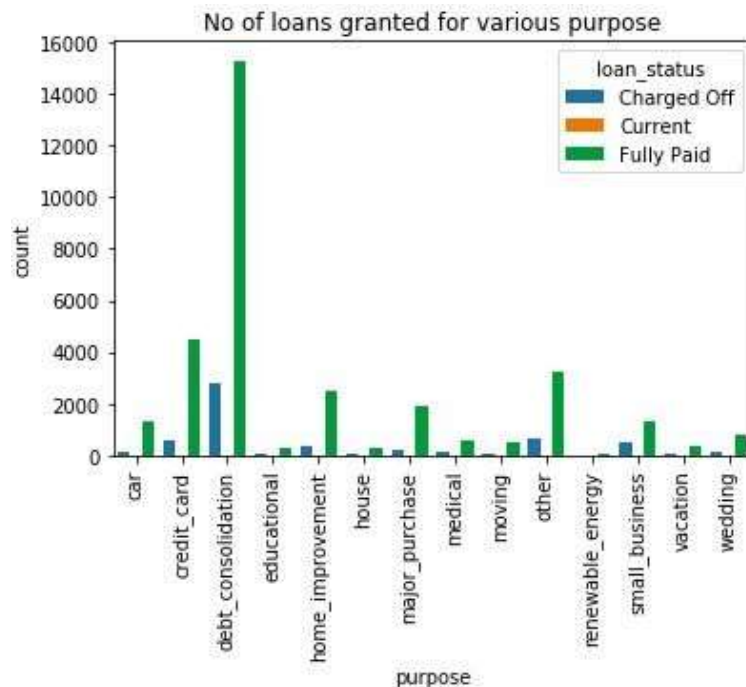Applicants from state CA have high tendency to default the loan

Loan Interest Rate Analysis

Observation:

1. Most of the loans with higher interest rate tends to default.

Purpose for applying Loan Analysis

Observation:

1. Most of the loans are granted were for debt consolidation.
2. Applicants having higher loan amount and low annual income tends to default the loan.

Loan Status vs Loan Tenure

Observation:
1. Most of the loans granted were of 36 months.
2. And loans granted for 36 months have slightly high tendency to default.
3. Additionally, Debt Consolidation is the most common purpose and loans are granted for 36 months and 60 months.
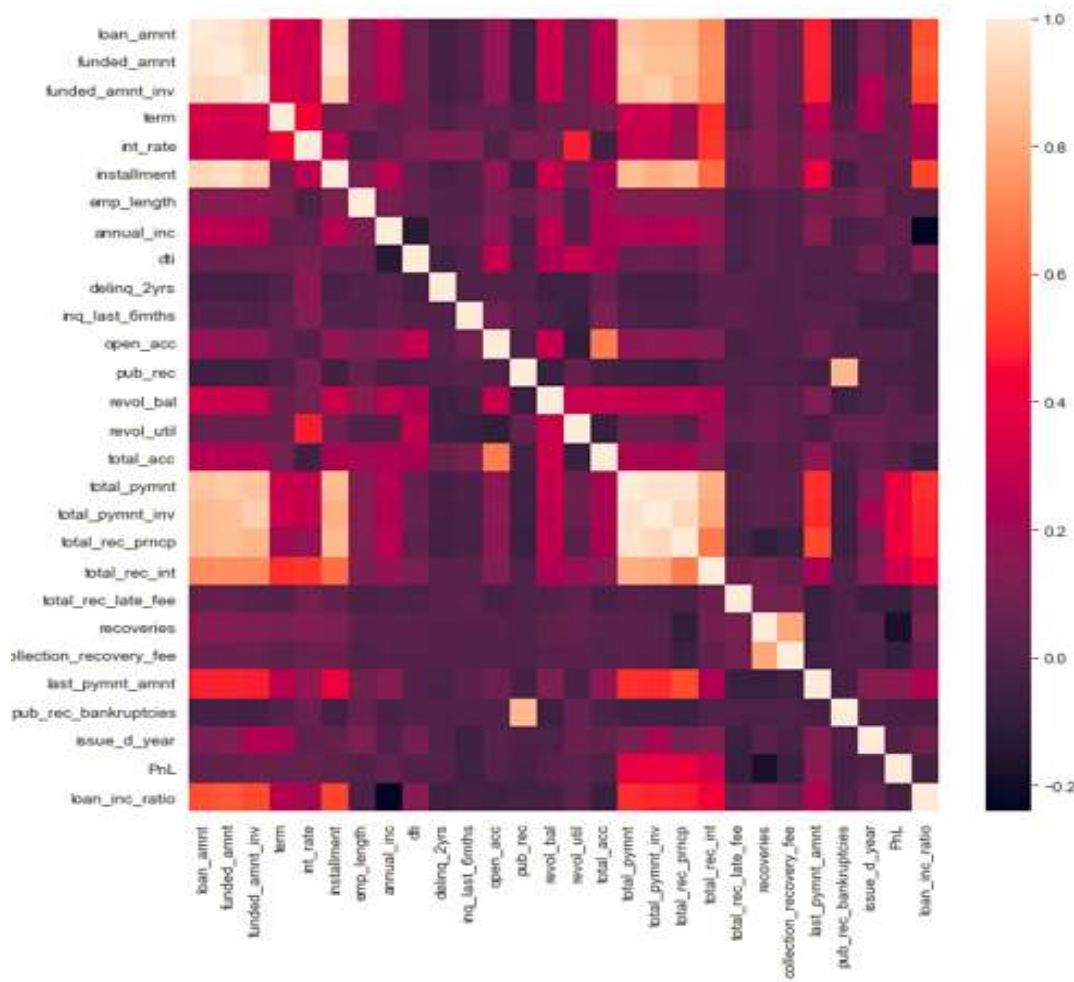
Loan Status by Home Ownership
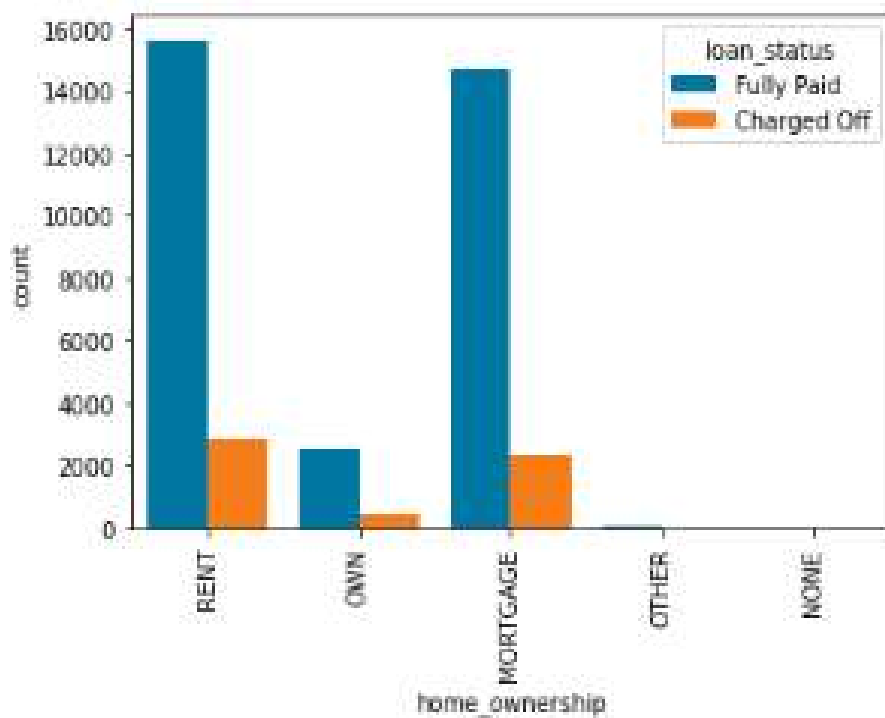
Observation:
Applicants having there home rented or on mortgage have almost equal tendency to default the loan.

Multivariate Analysis

Observation: Below variable analysis tends to default -

- Higher interest rate.
- Higher revolving line utilization rate
- Longer Repayment term (5 years)
- Derogatory public records (1 or 2)
- Public bankruptcy records (1 or 2)
- High loan amount & interest rate for lower income group
- High installment and longer repayment term

Conclusion

1. Low grade loans have high tendency to default. Grading system is working as expected.

2. Loans having higher interest rate have more defaulters. Check the background of applicant thoroughly if interest rate is high.

3. Extra scrutiny must be done for the applicants belonging to CA state, as tendency to default is high.

4. When the purpose is debt consolidation check applicant thoroughly as it has high tendency to default.