Lending Club Case Study

Group Members:
Pallab Bhattacharya
Anilan M

Date 24-Sep-2024

Business Objective

- We are assisting a consumer finance company in identifying key characteristics of applicants who are likely to default (fail to repay the loan) and those who are not, using Exploratory Data Analysis (EDA) methods.

- To achieve this, we will analyze past data to determine which types of applicants are likely to default and which are not, thereby simplifying the loan approval or rejection process.

Solution Approach Design

| Data loading	Data Cleaning	Data Understanding	Univariate Analysis	Segmented Univariate	Bivariate	Conclusion
Activities Performed						
Reading the loan.csv	Removing nan values, extra columns, categorial columns with 1 value	Creating derived metrics, binning, and understanding the various columns as per the data dictionary	Univariate is analysing single column as a time, like plotting the frequency or count of the variables	In Segmented univariate we analyse one column per section (using binning) or per category	In Bivariate analysis we understand two columns at a time and make inference how those are related	From the various plots we provide recommendation which will decrease the defaulters. Business-objective

Univariate Analysis -1

- December is the preferred month and maximum loan was disbursed

- The loan amount is increasing every year starting 2008 to 2011

- 36-month term is preferred term

- More than 80% of loan were fully paid, less than 20% were charged off.
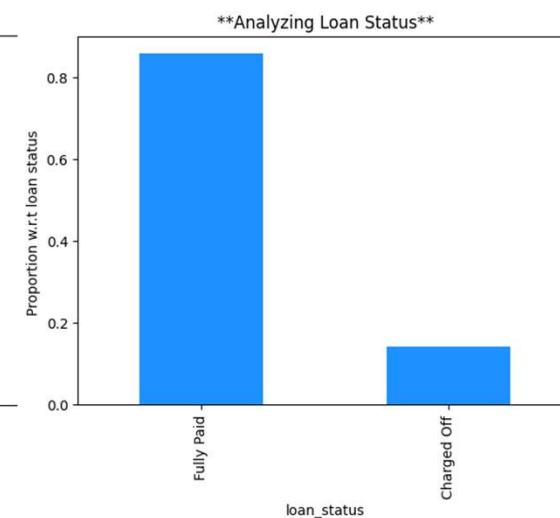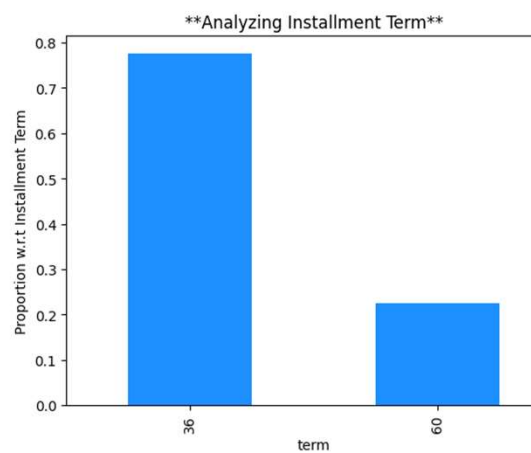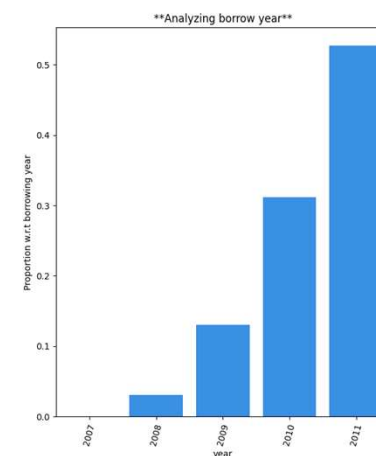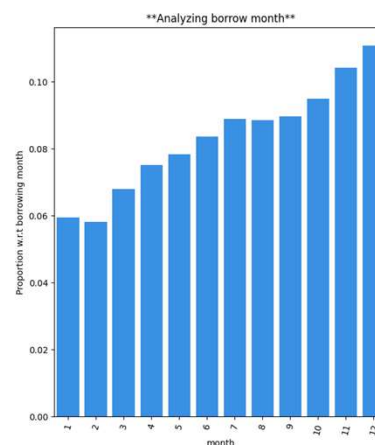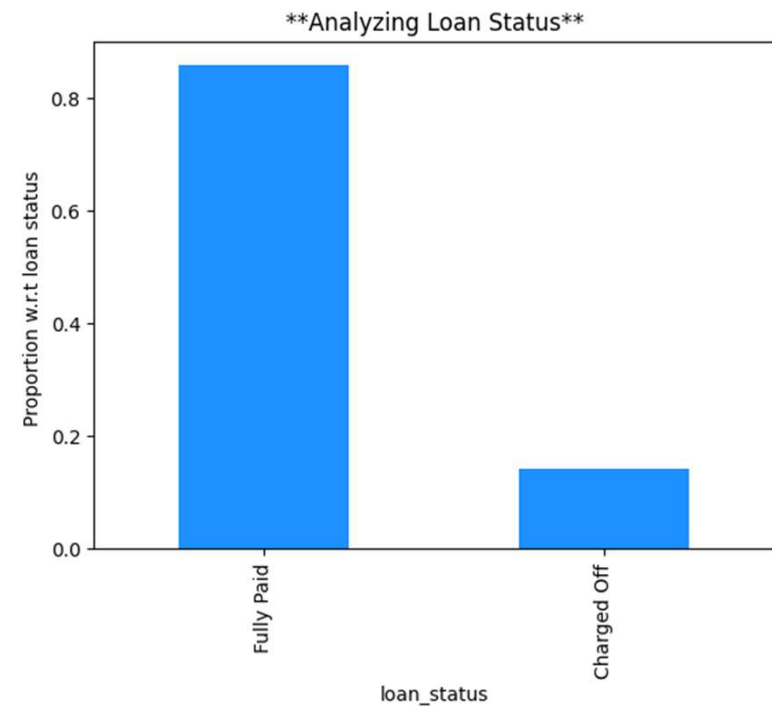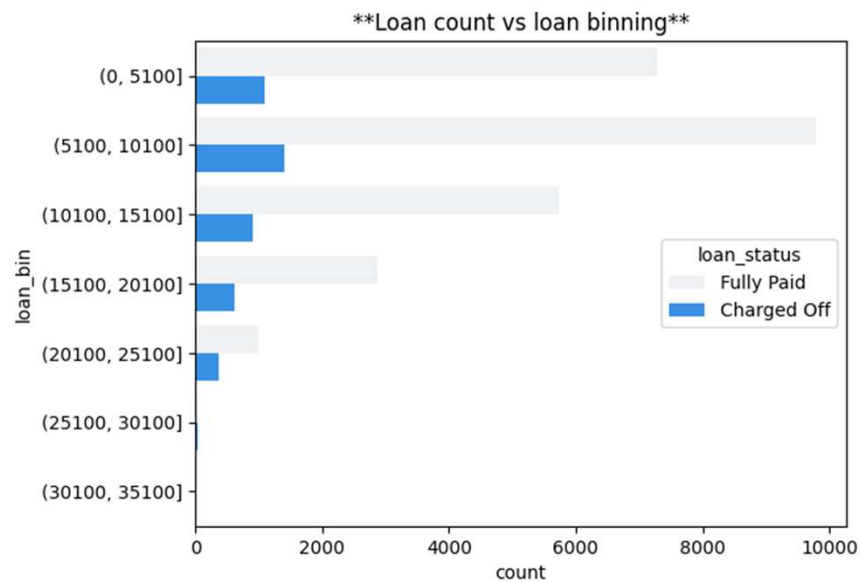
Univariate Analysis -2

- Little more than 20% have an employment more than 10+ year, followed by less than 1 Year.

- There is a significant amount of unverified application, later we will see this has strong indicator of loan defaulting

- Around 52% have a rented house followed by mortgage. Later we will find that it is also has driving variable.
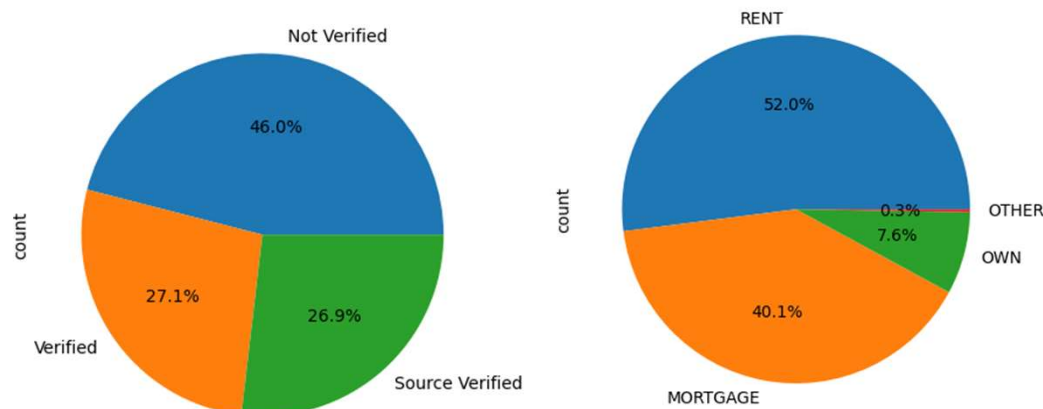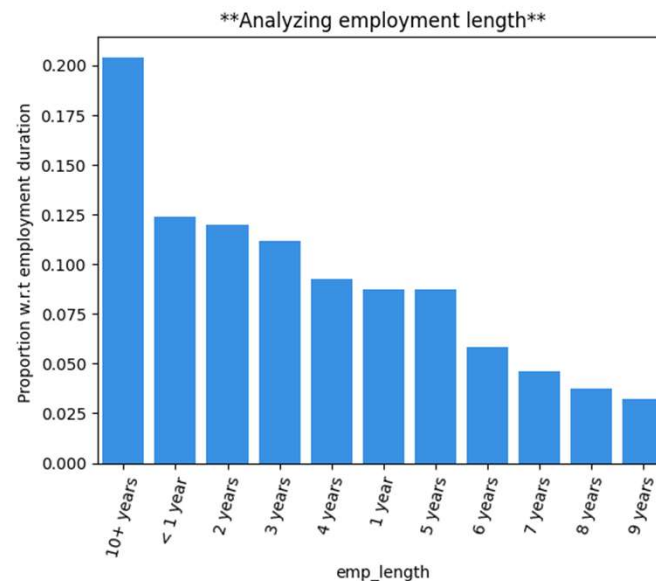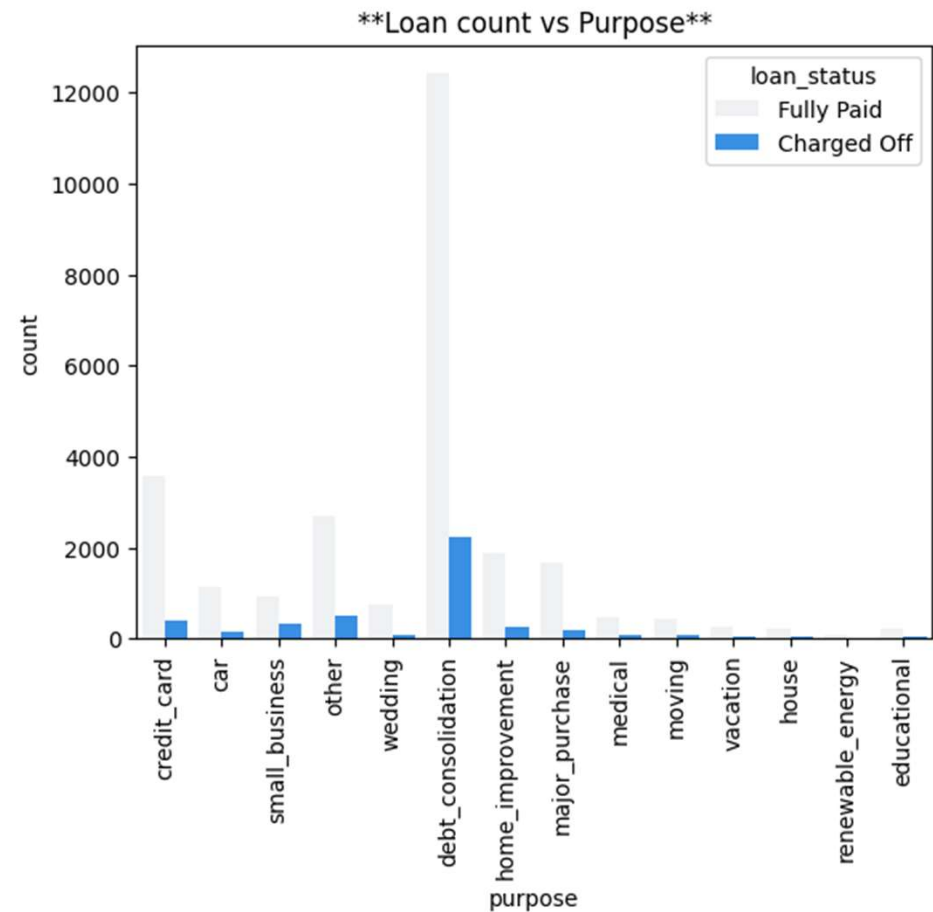
Segmented Univariate Analysis -1

The binned loan amount shows higher number
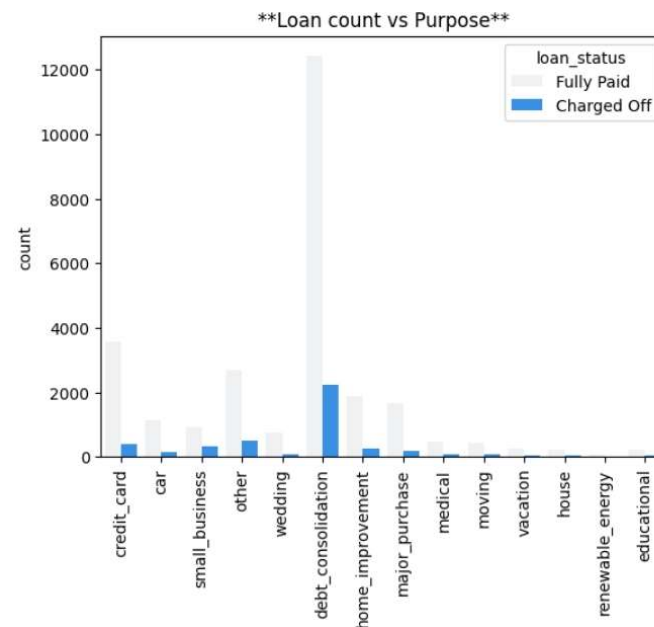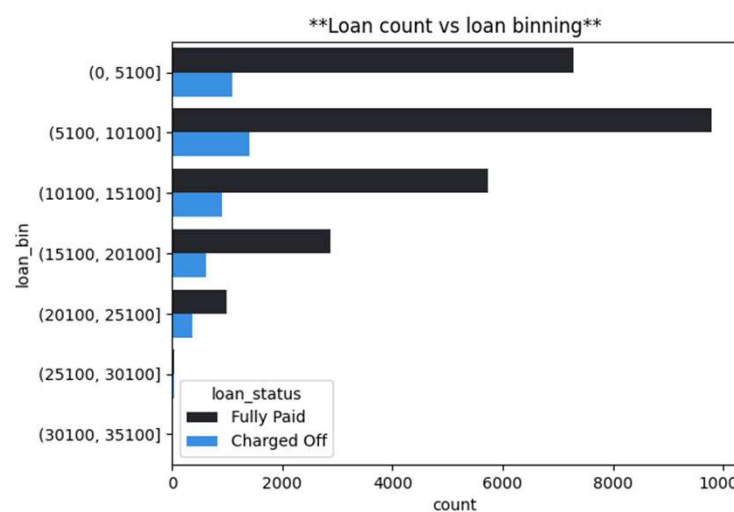of charge off at 5k to 10k range

Total number of charged off cases are little less
than 20 % whereas the fully paid is more than
80%

Segmented Univariate Analysis -2

This plot shows that the debt consolidation has higher number of charge off compared to other purposes. Hence this will be a driving factor for the finding the defaulters.


Loan count vs Purpose

Bivariate Analysis -1

Loan count vs loan binning

Loan count vs Purpose
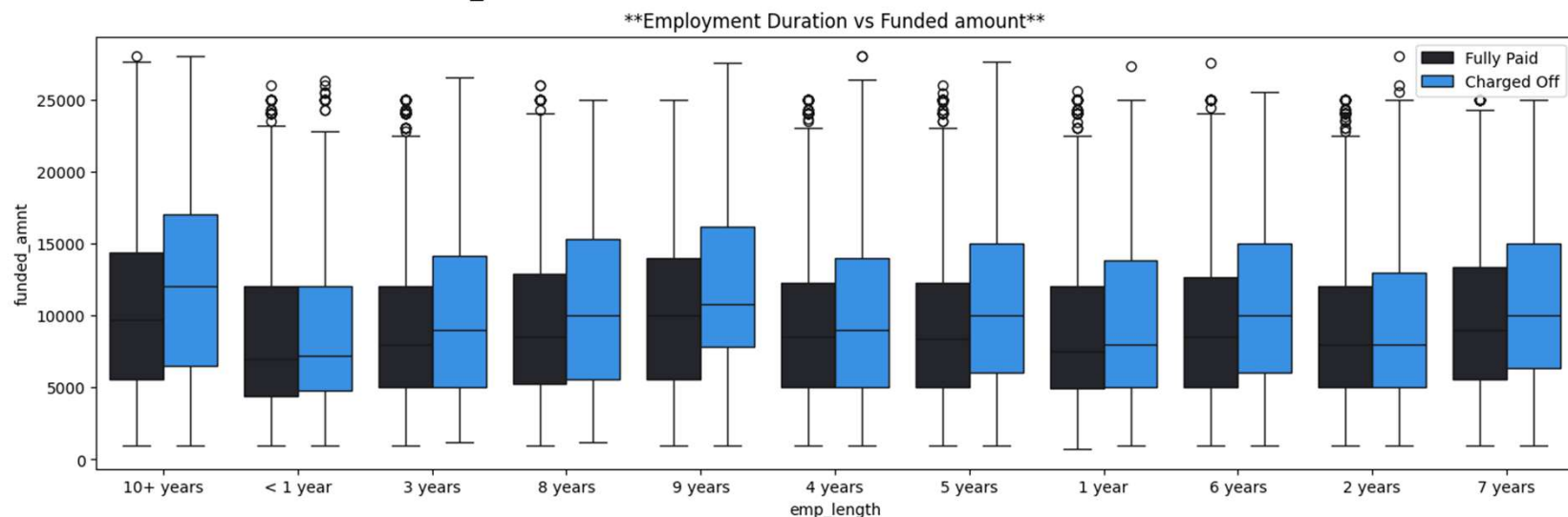
Public record bankruptcies vs Loan status

Binning the loan amount shows that 5k to 10k range has higher amount of charge off.

The greater number of public record of bankruptcies the chances of charge off increases. Debt consolidation is one of the areas where charge off happened to a larger extent
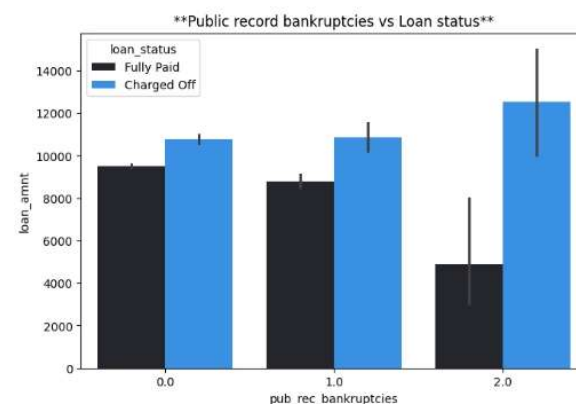
Bivariate Analysis -2


Employment Duration vs Funded amount

Employment > 10+ years shows more charged off cases.
Non verified application adds unto charge off cases.
As soon as the interest rate increases it leads to charge off.
The fully paid have little more than 12.% as 75th percentile.

Public record bankruptcies vs Loan status

Loan status vs Interest rate
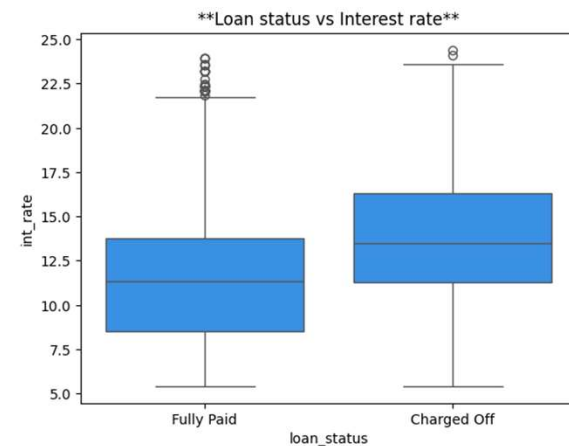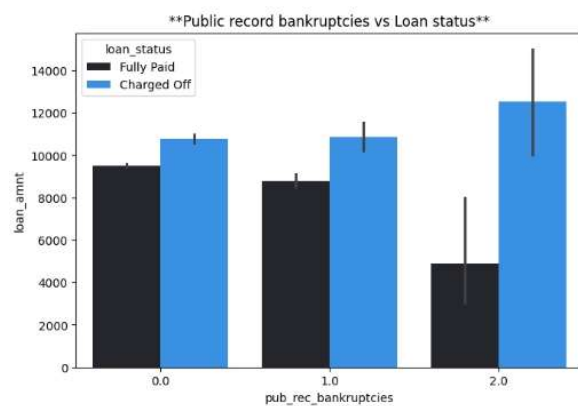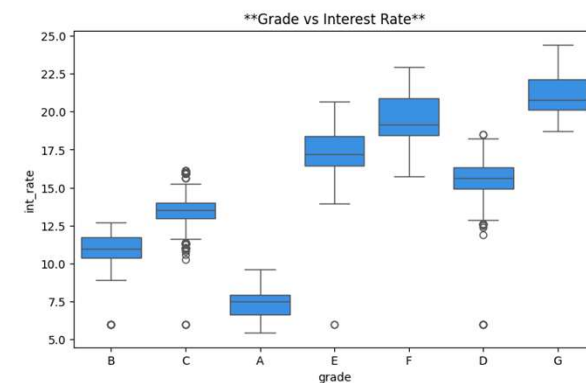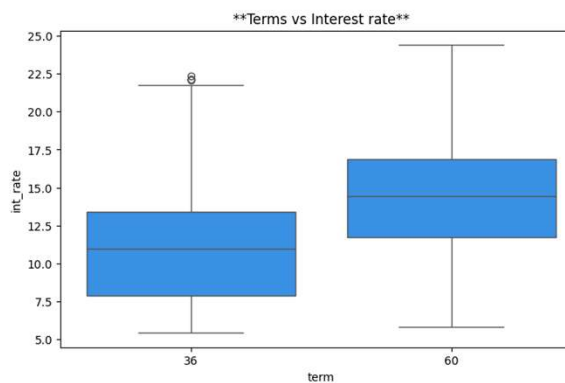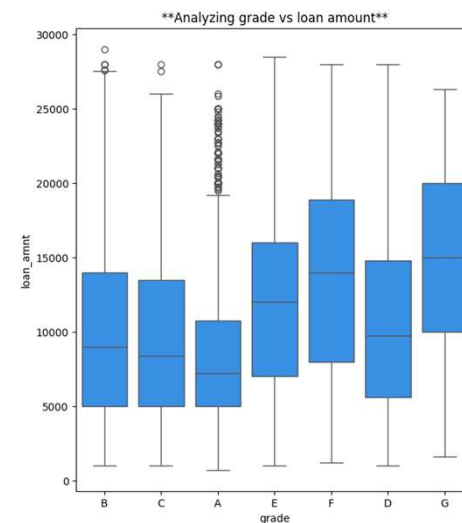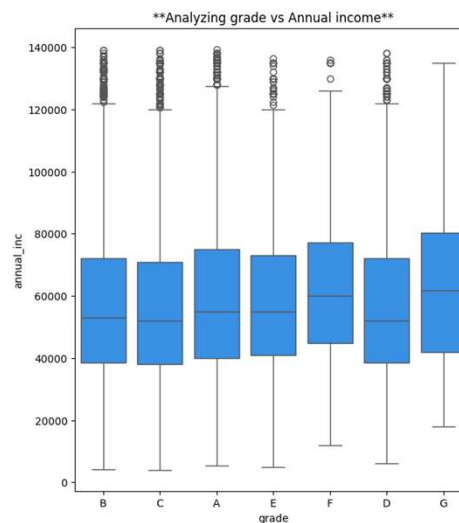
Bivariate Analysis -3

- Q4 quarter has shown high value of loan and even charge off.

- Mortgage shows higher proportion charge off then followed by rent.

- CA shows higher about of charge off followed by FL and NY.

  .
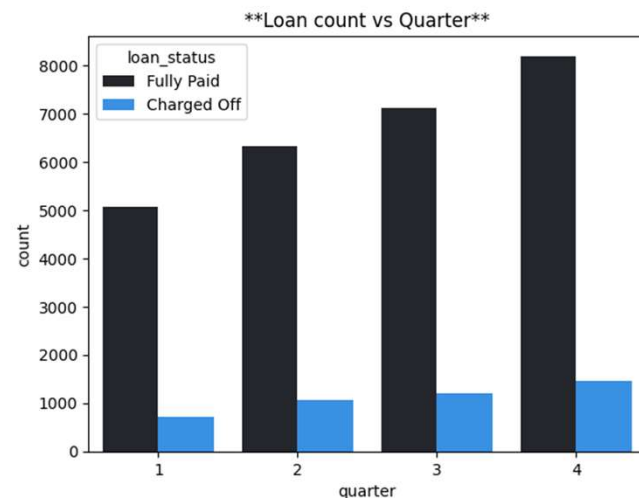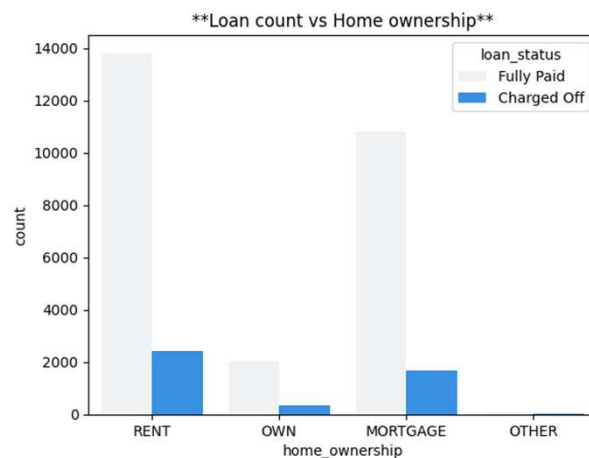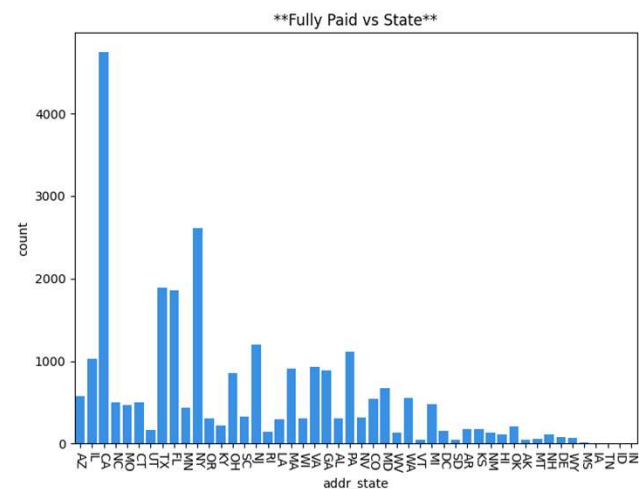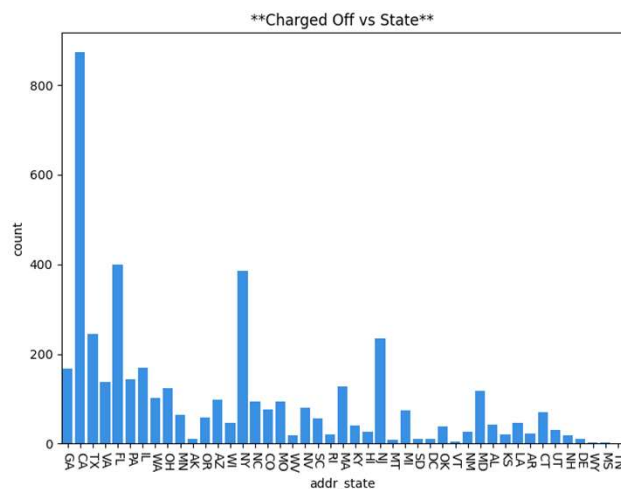
Bivariate Analysis -4

• Grade F and G has on the higher end of borrowing money and, they are in the higher interest rate bucket leading to more charge off.

• Interest rate is higher for 60-month term when compared with 36-month term.
• The interest rate is also higher for Grade G and F which led to defaulting
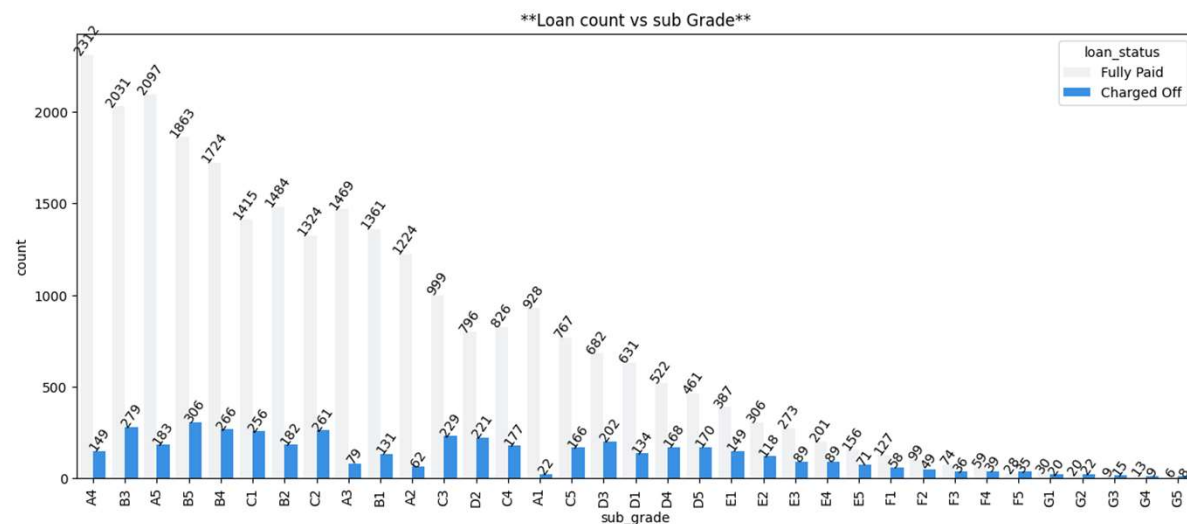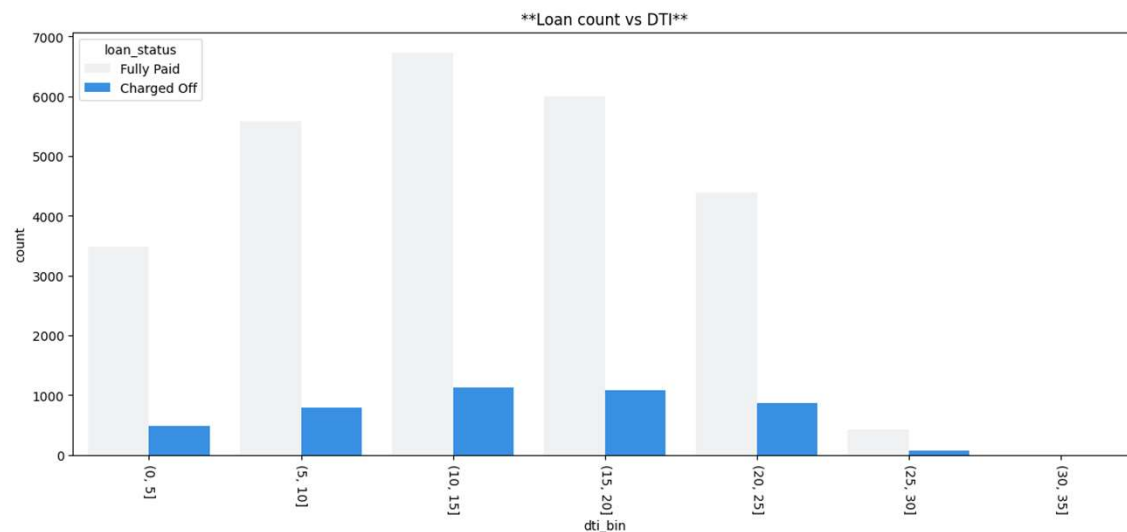
Bivariate Analysis -5
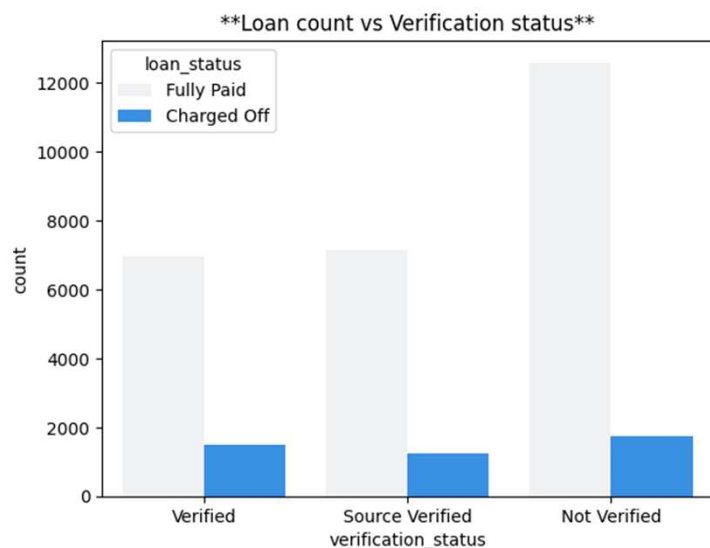
Not verified source has higher amount of charge off.

Binning result shows the 10-15% DTI has higher number of charge off compared to others.
Subgrades B5. B3, C2 shows higher amount of charge off compared to others.

Bivariate Analysis -6

- Grade F and G shows higher amount of charge off. The lowest is Grade A.

- The zip code analysis shows 945xx followed by 313xx has higher amount of charge off.
- 60 month has a higher proportion of charge off

Analyzing Interest Rate with Grade along with loan status

Zip Code for Top 5 Charge off

Term Period by Loan Status

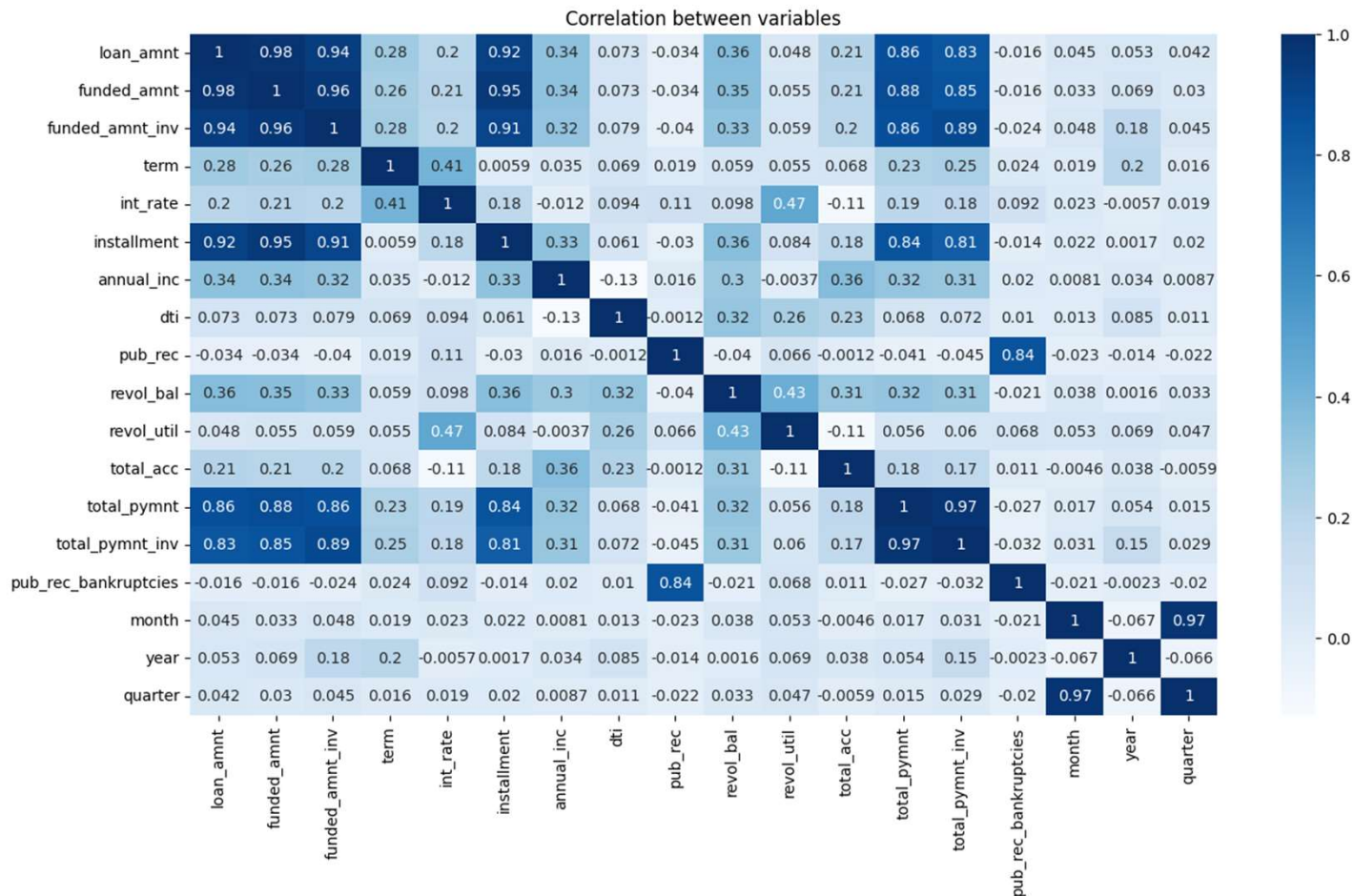# Multivariate Analysis
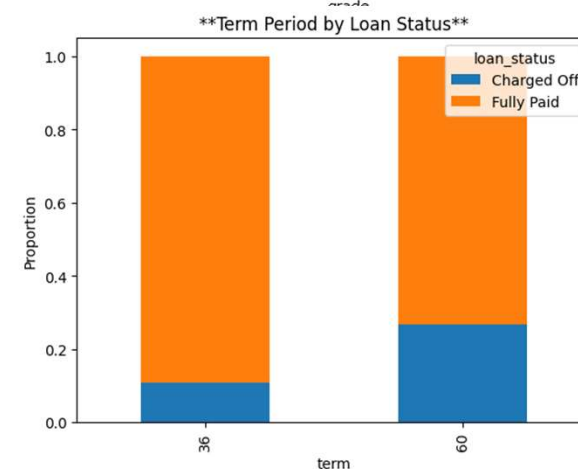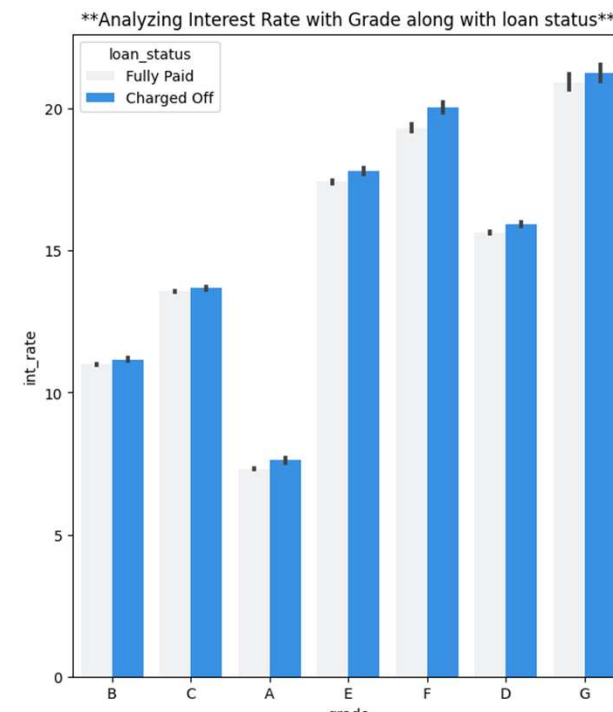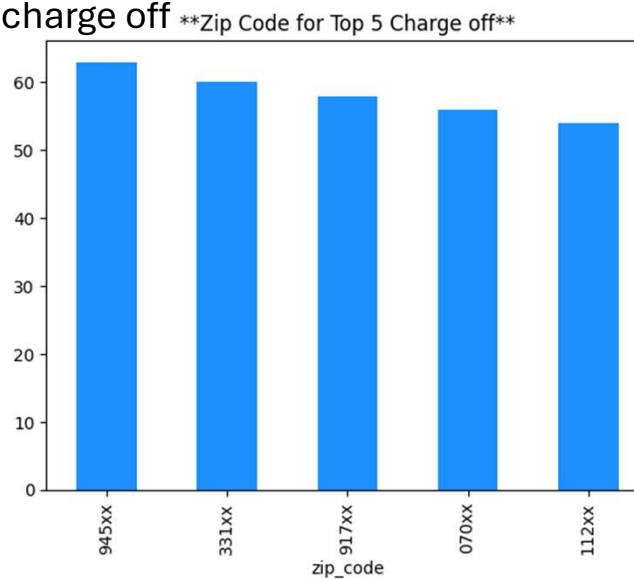
Correlation between variables

- Public record bankruptcies are highly correlated with public records, other variables are not correlated or negatively correlated.
- Total payment is highly correlated with loan amount, funded amount and funded amount investor.
- Installment shows high correlation with total payment and total payment investor
- Month, year and quarter are less correlated with other variables.
- Revolving balance is not highly correlated with other variables.

Conclusions

- Grade F and G has on the higher end of borrowing money also, they are in the higher interest rate bucket leading to more charge off.

- Home ownership having rent, or mortgage has paid fully compared to other categories

- Higher interest rate also risks towards charge off.

- Public derogatory records leads to more charge off.

- In the zip code analysis is the top 3 charge off happened from zip codes 945xx, 331xx, 917xx whereas the top 3 for full payment was from 112xx, 606xx and 100xx.

- Not verified status has a greater number of charge off, so verification is mandatory going forward for new applicants.

- CA state has more than 800 charged off situation followed by FL which is close to 400.

- Higher DTI leads to charge off.

- 60-month term is more prone to charge off.

Thank you