Lending Club Case Study

Submission By
- Satya Prakash
- Mythili Kandula

Problem Statement

Lending club is a consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Solution Approach

Data Loading and Cleaning

- Import Libraries.
- Read CSV file in Dataframe.
- Drop/Impute Columns.

Fix Data Type & Categorize

- Fix data types.
- Add derived columns.
- Categorize columns for analysis.

Univariate Analysis

- Numerical
- Categorical

Bivariate Analysis

- Numerical vs Numerical
- Numerical vs Categorical
- Categorical vs Categorical

Multivariate Analysis

- Correlation Analysis

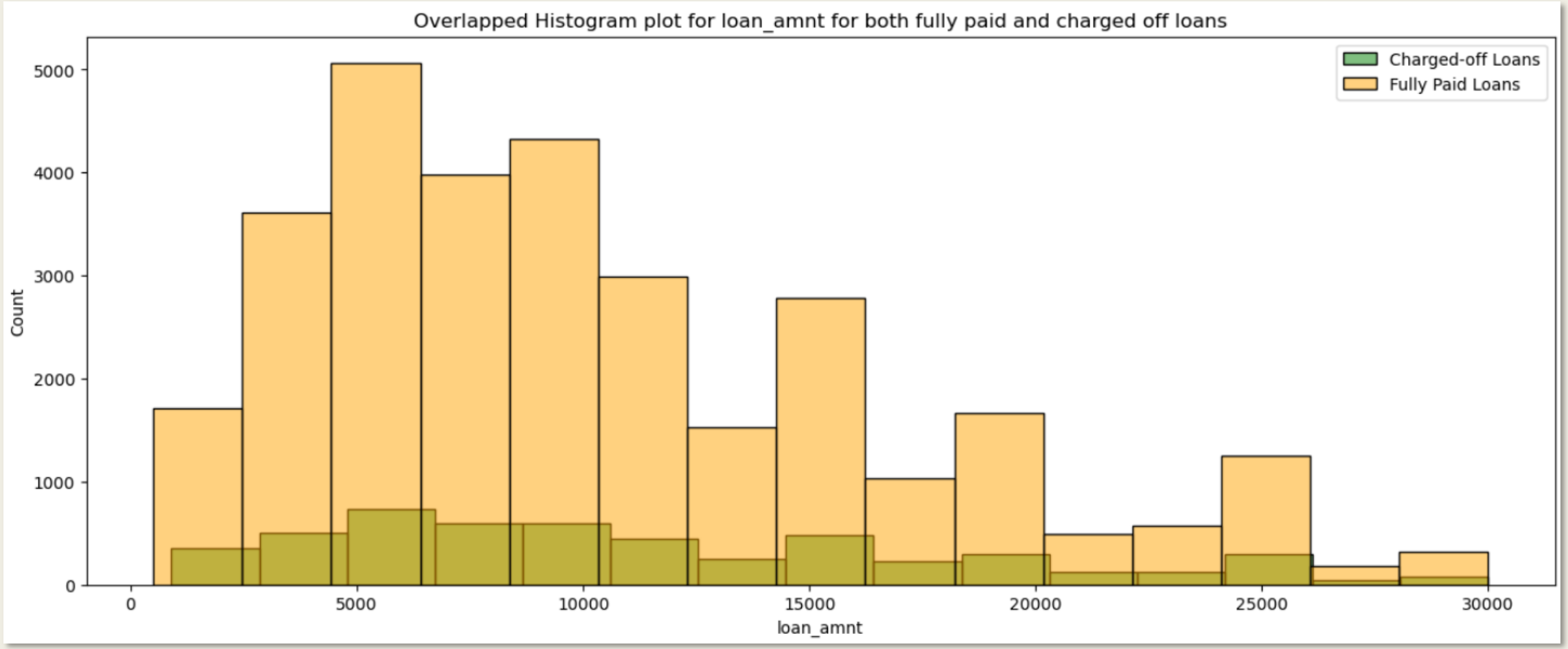
Analysis

Univariate

(Numerical)

Bivariate

Multivariate



Most of the loans approved are in the range of 5000 to 10000. Pattern of fully paid and charged off is almost same.

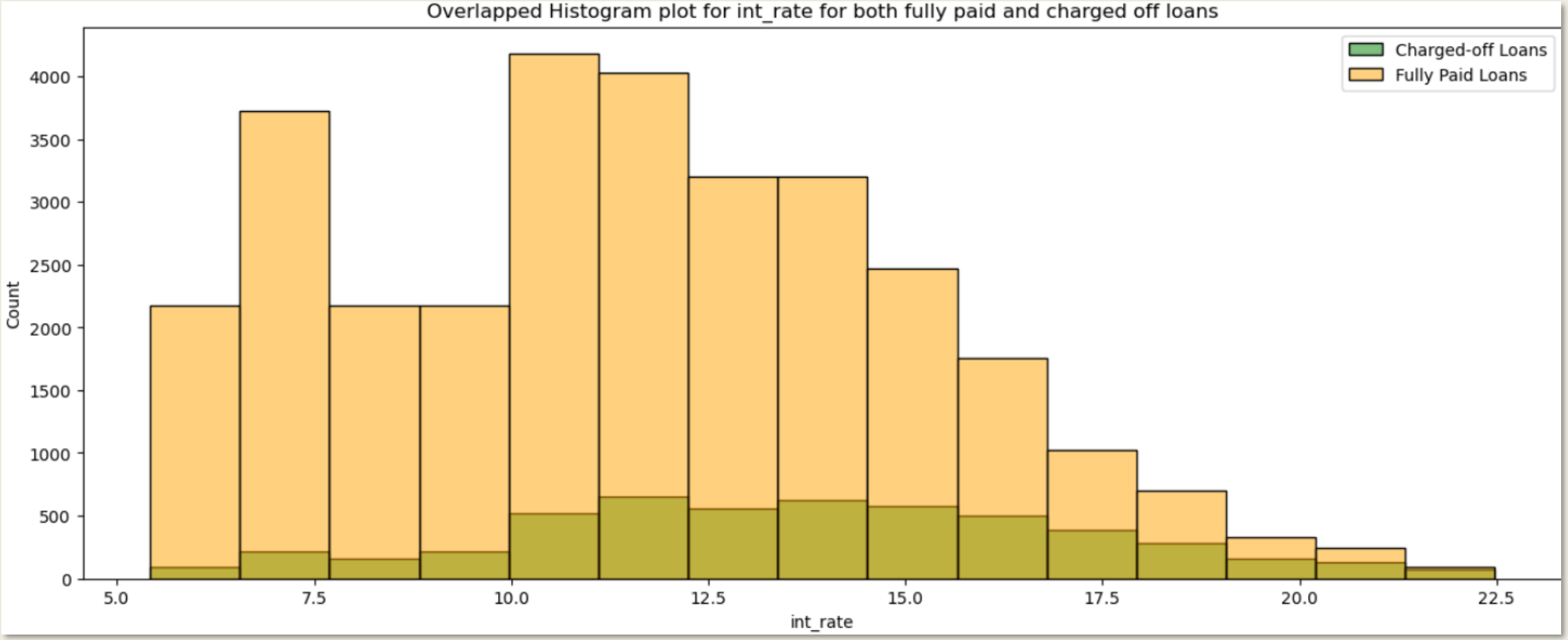
Analysis

Univariate

(Numerical)

Bivariate

Multivariate



The higher the interest rate, more is the probability of defaulting. proportionally most of the loans with interest rates less than 10% are paid off fully.

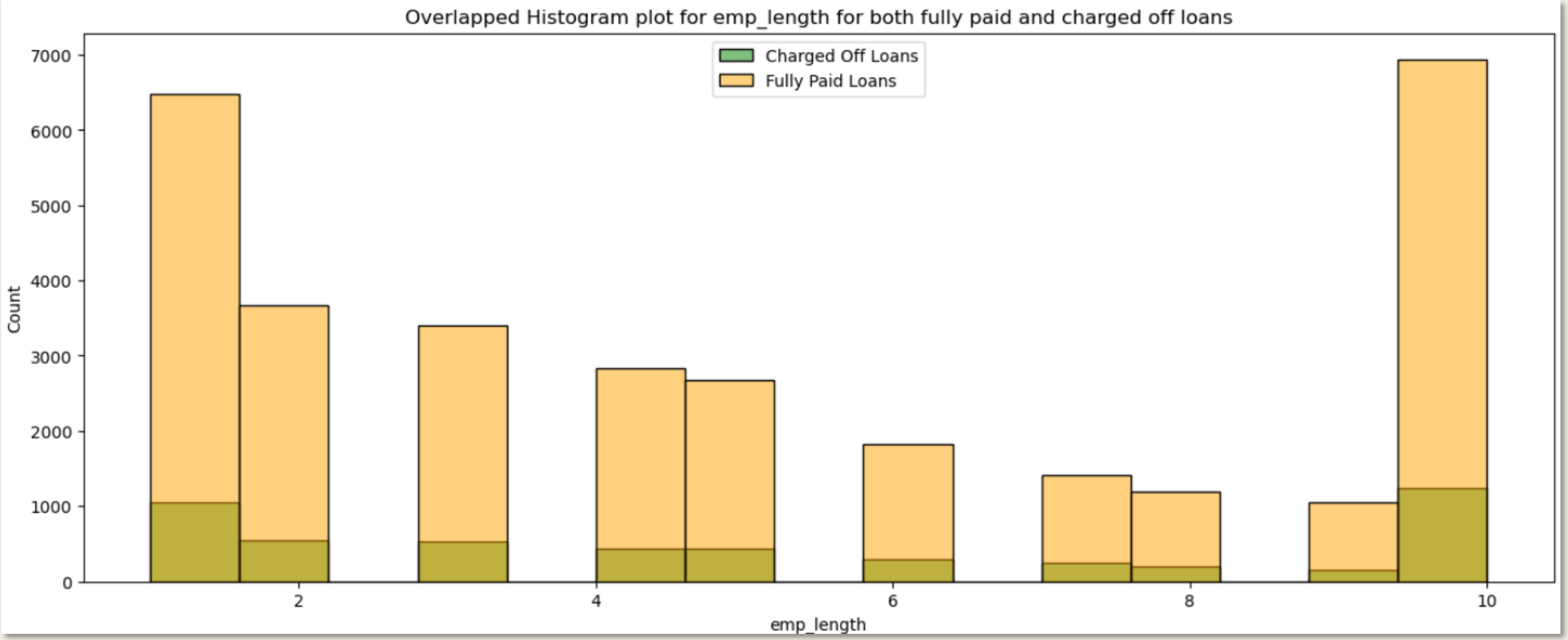
Analysis

Univariate

(Numerical)

Bivariate

Multivariate



Surprisingly most loans defaulters have greater than 10 years of employment history.

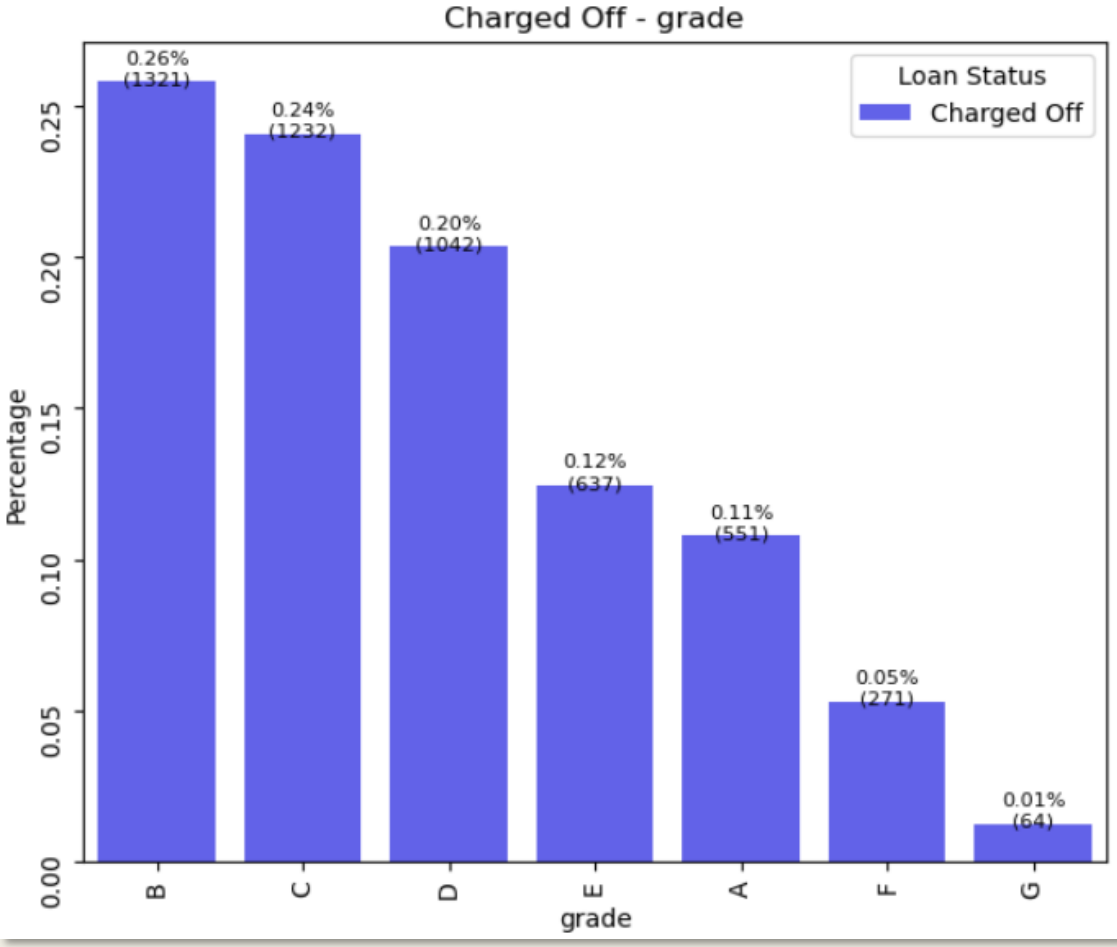
Analysis

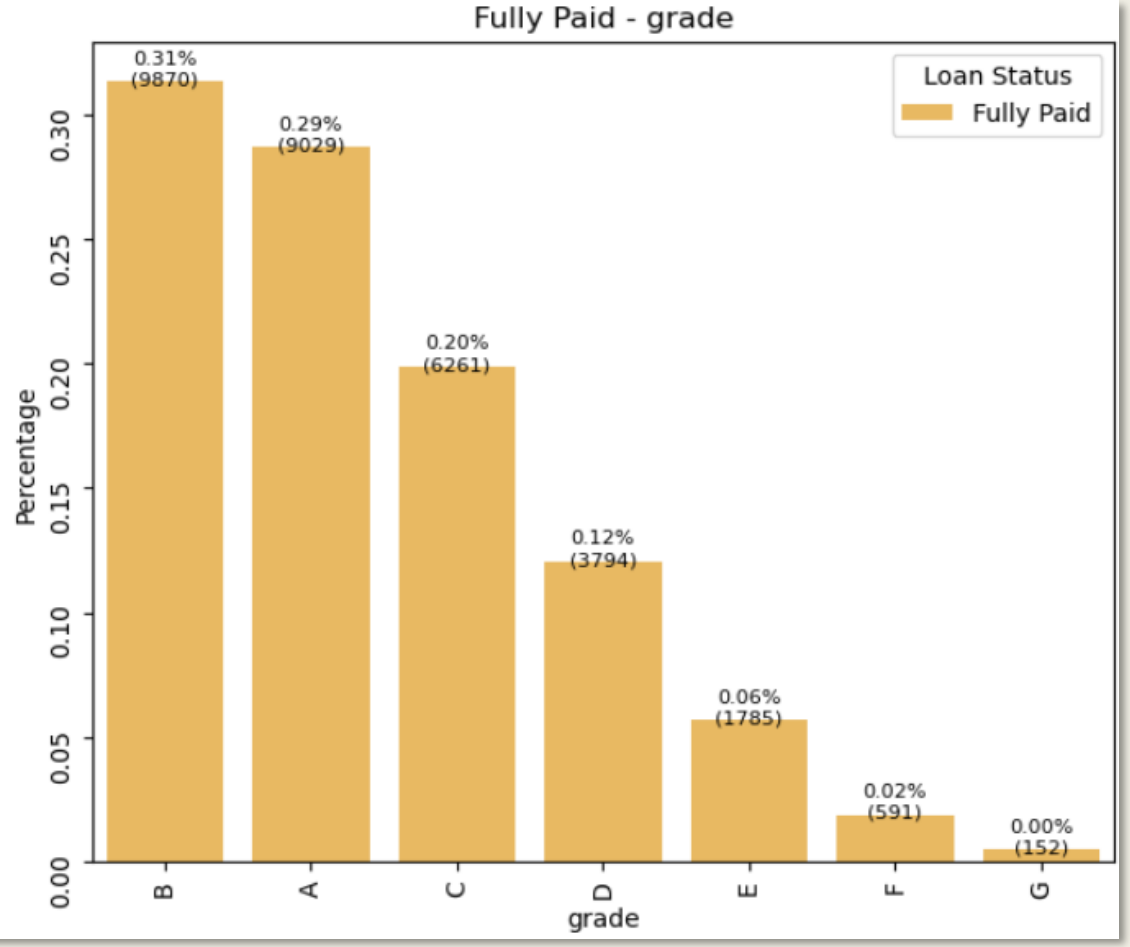
Univariate

(Categorical)

Bivariate

Multivariate





B grade loans have higher probability of defaulting followed by grades C and D. When comparing fully paid and charged off, we can observe that A grade loans have highest probability(29% vs 11%) of being fully paid.

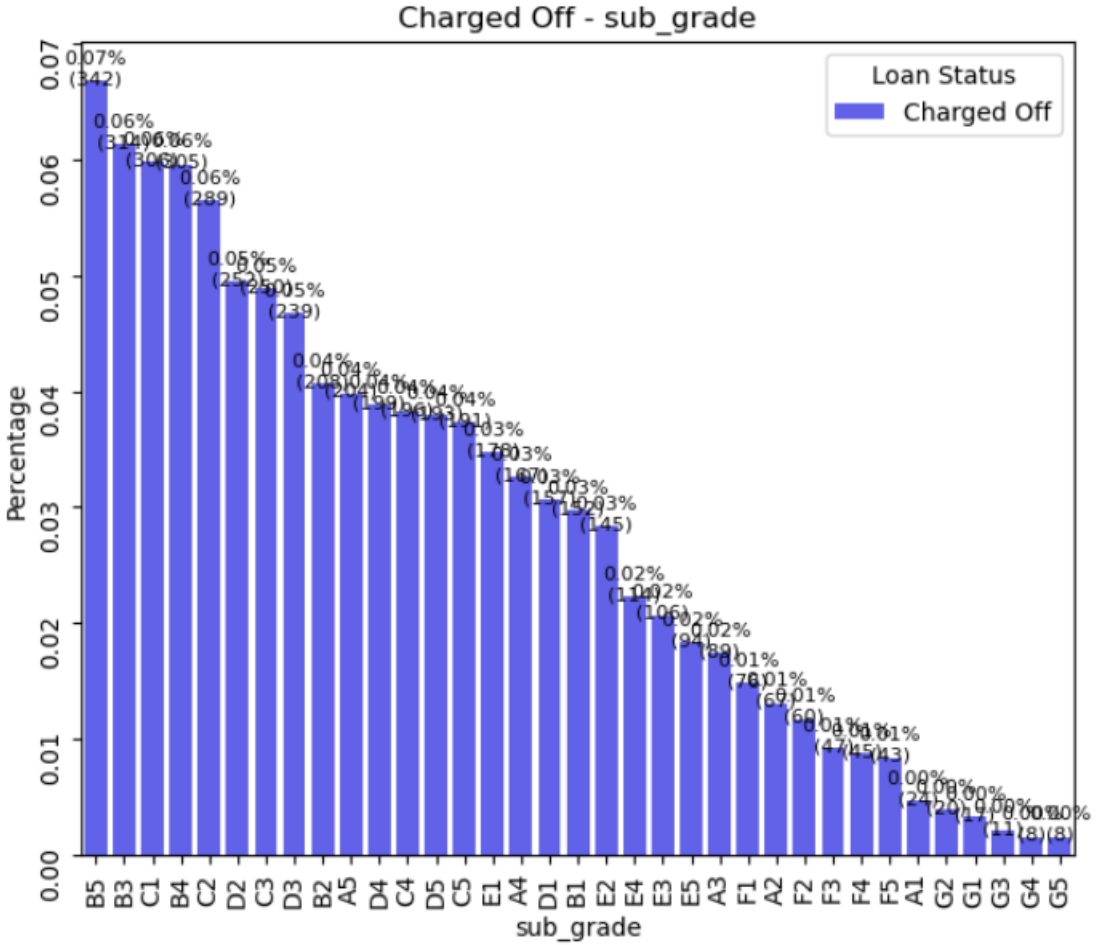
Analysis

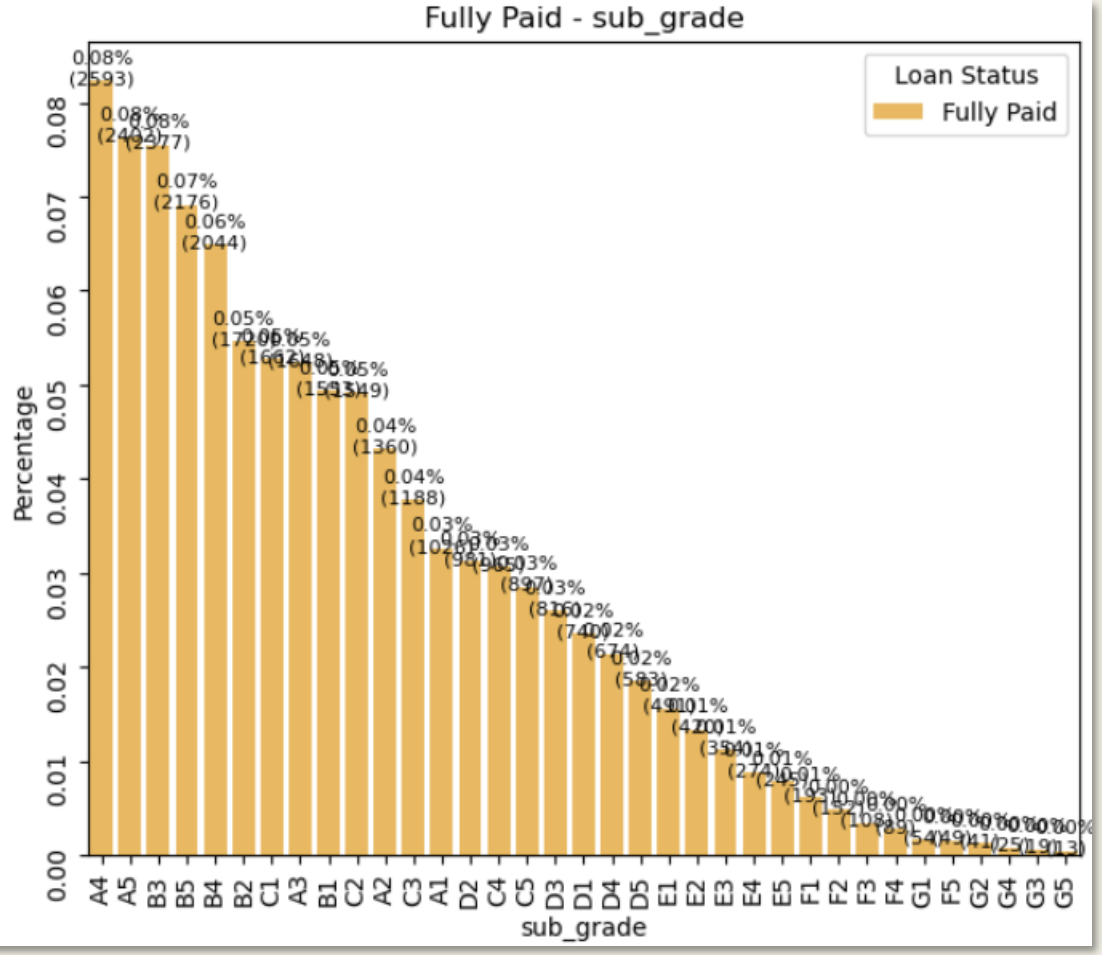
Univariate

(Categorical)

Bivariate

Multivariate





B5 subgrade loans have highest probability of default followed by B3 and C1. A4, A5 and B3 sub grade loans have highest probability of being fully paid off

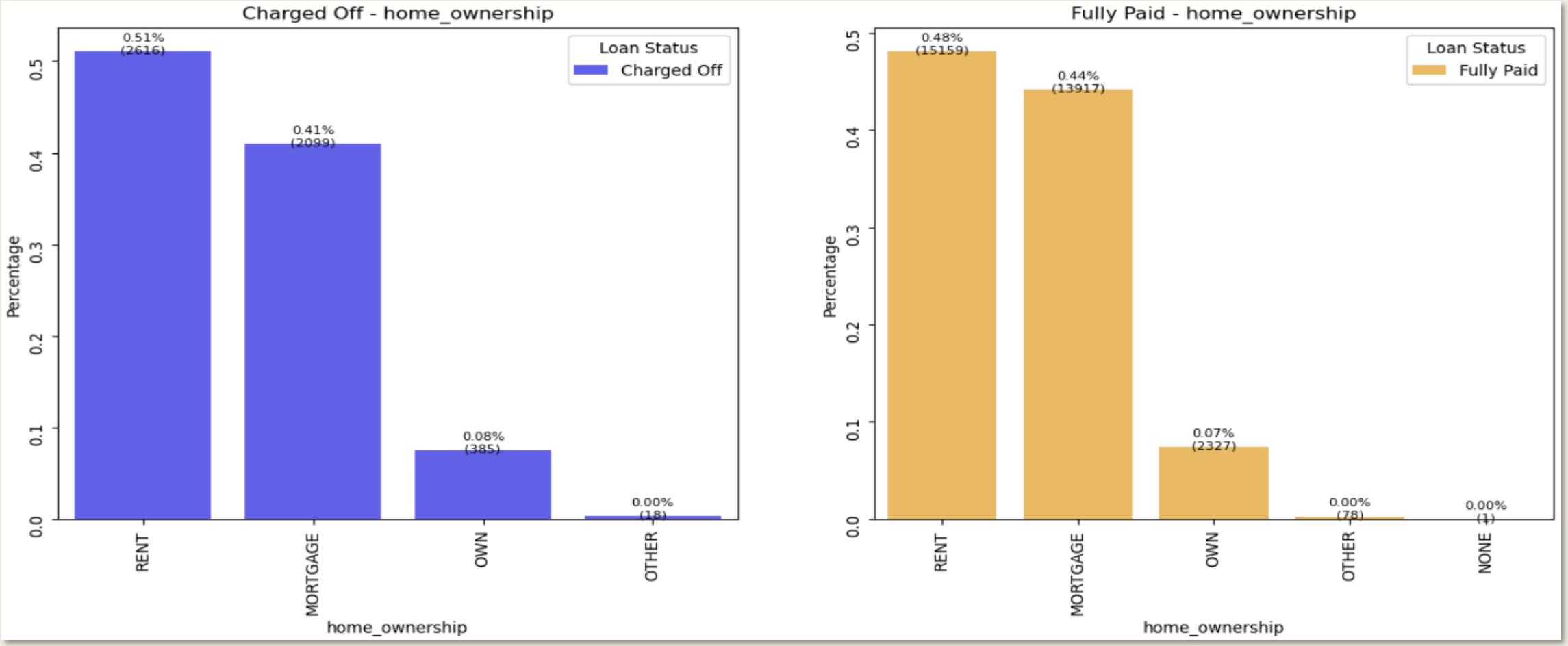
Analysis

Univariate

(Categorical)

Bivariate

Multivariate



People living in rented house have highest probability of defaulting compared to mortgage and own.

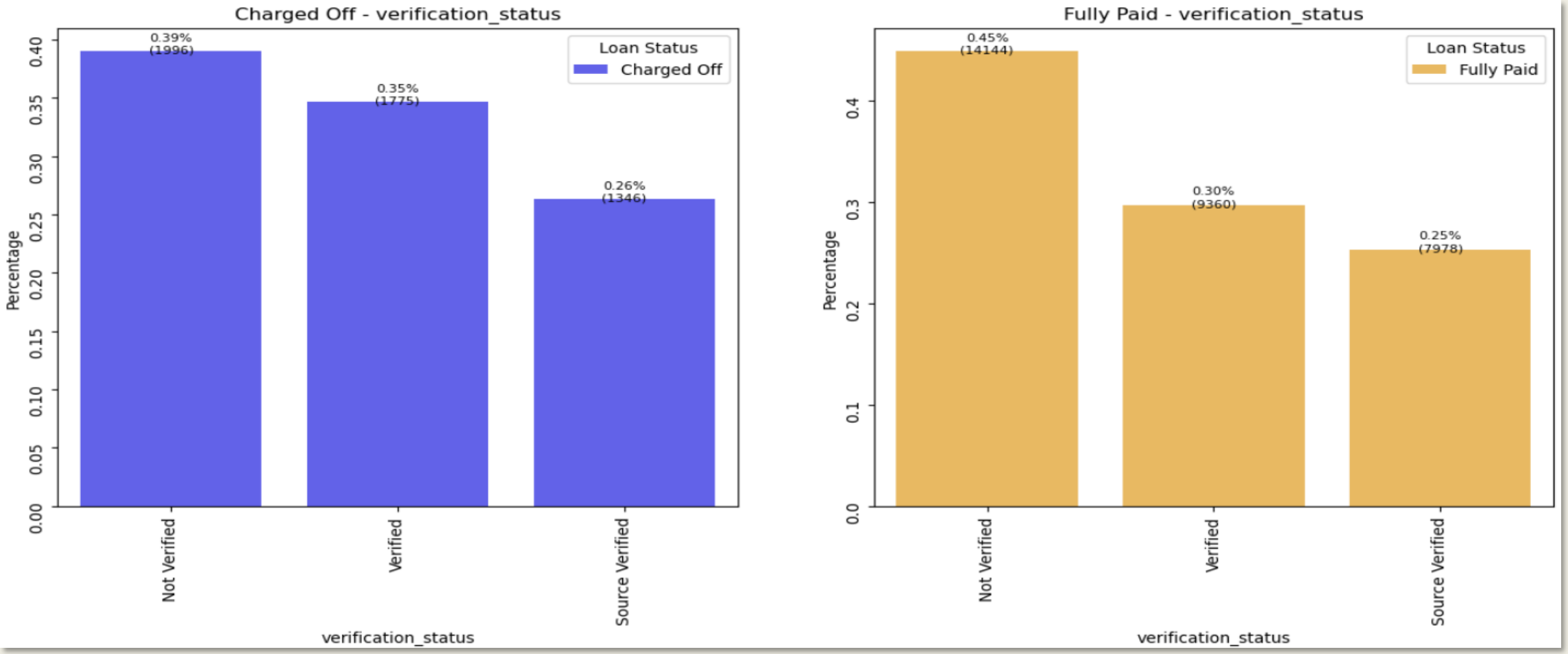
Analysis

Univariate

(Categorical)

Bivariate

Multivariate



Surprisingly the loans with income source not verified are the ones that is getting mostly fully paid

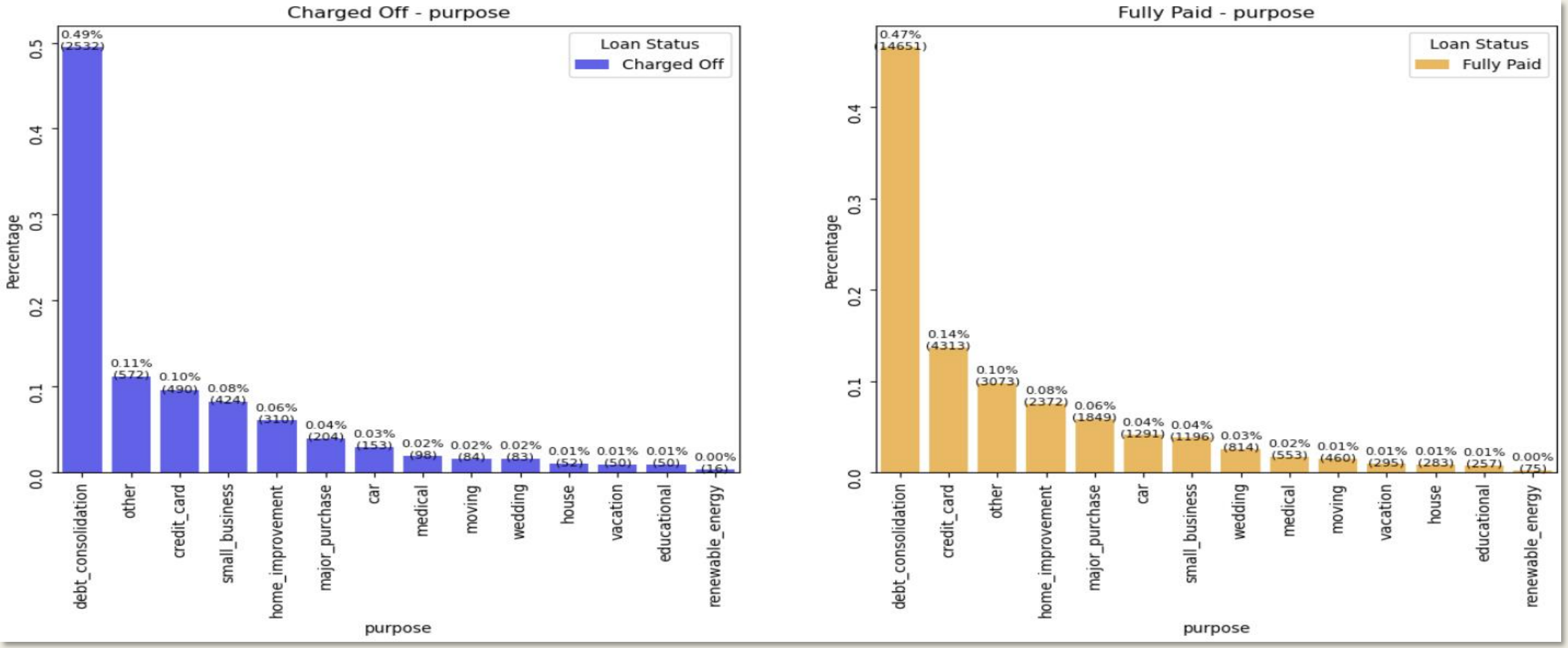
Analysis

Univariate

(Categorical)

Bivariate

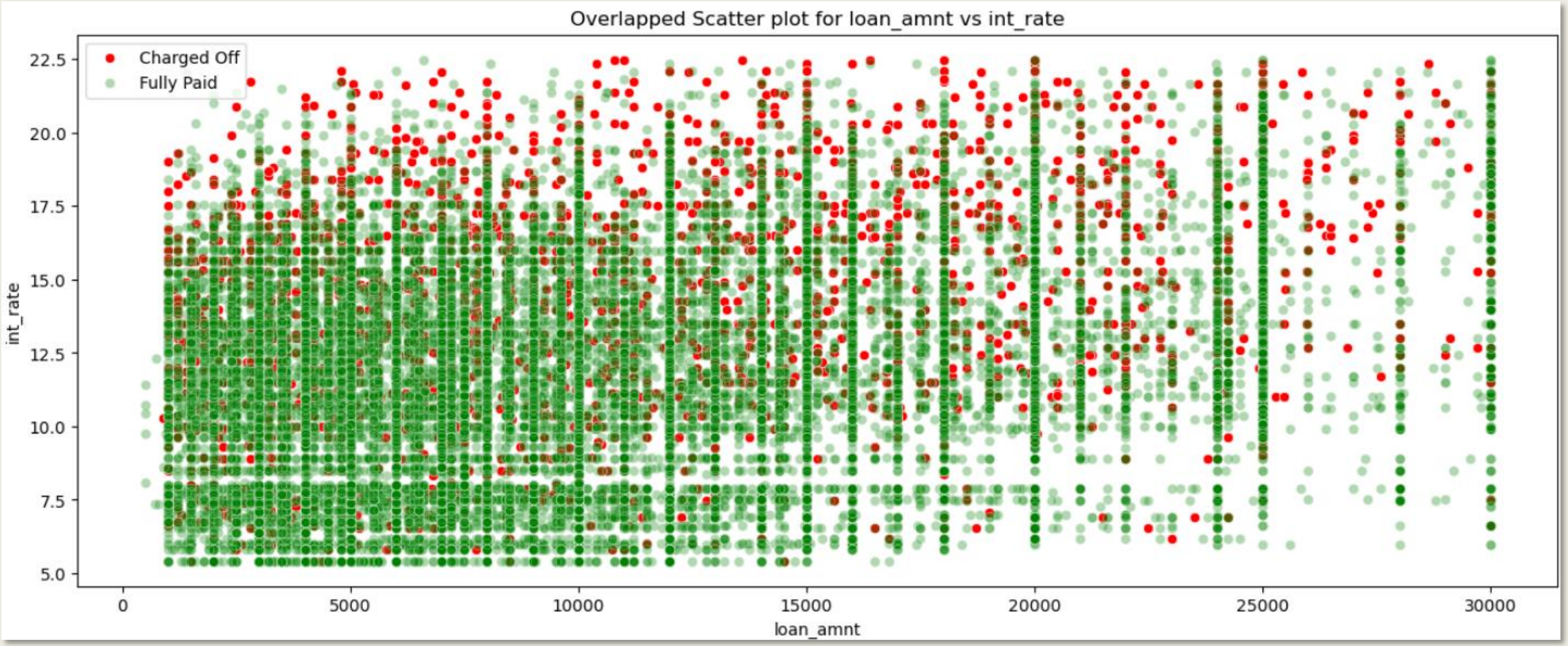
Multivariate



Out of all the defaulted loans, around 50% were taken for the purpose of debt consolidation followed by other and credit card.

Analysis

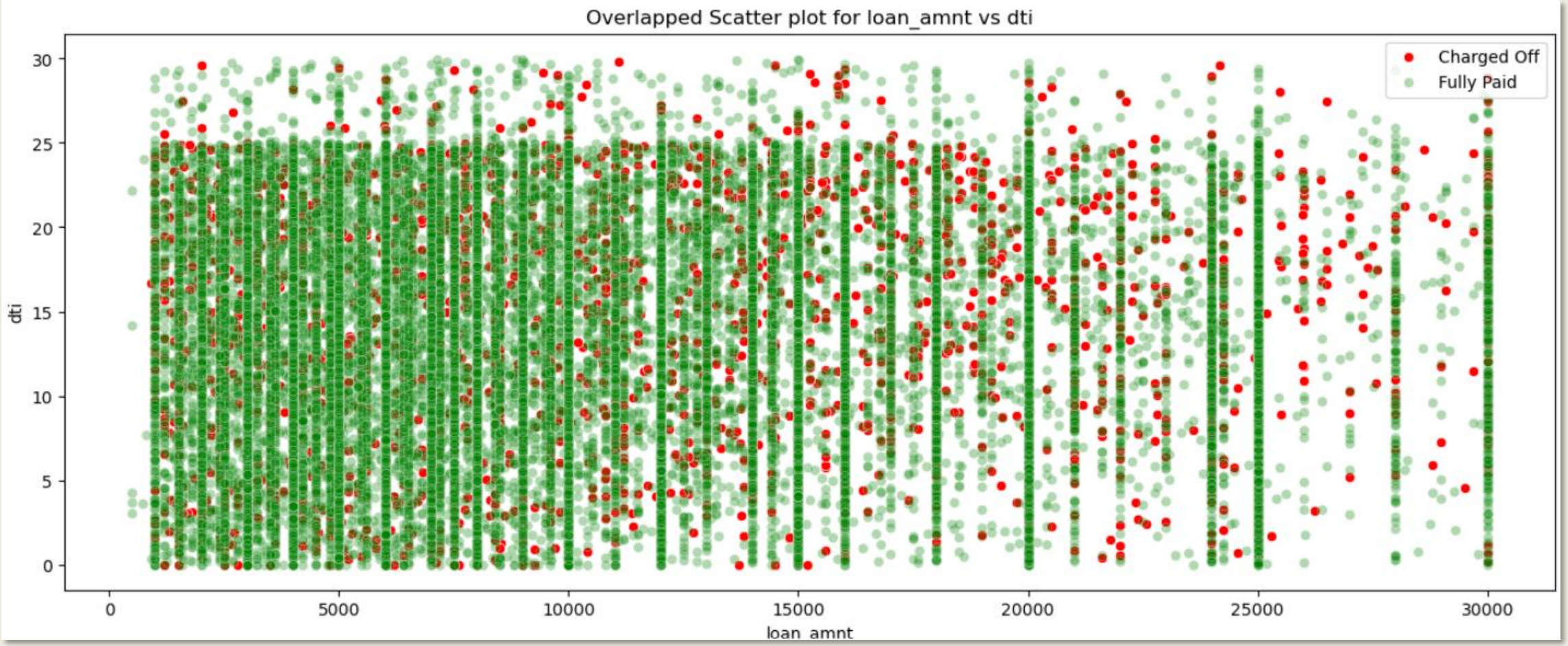
Univariate | **Bivariate** (Numerical vs Numerical) | Multivariate



It is observed that irrespective of loan amount, the customer is more likely to default if the interest rate is higher.

Analysis

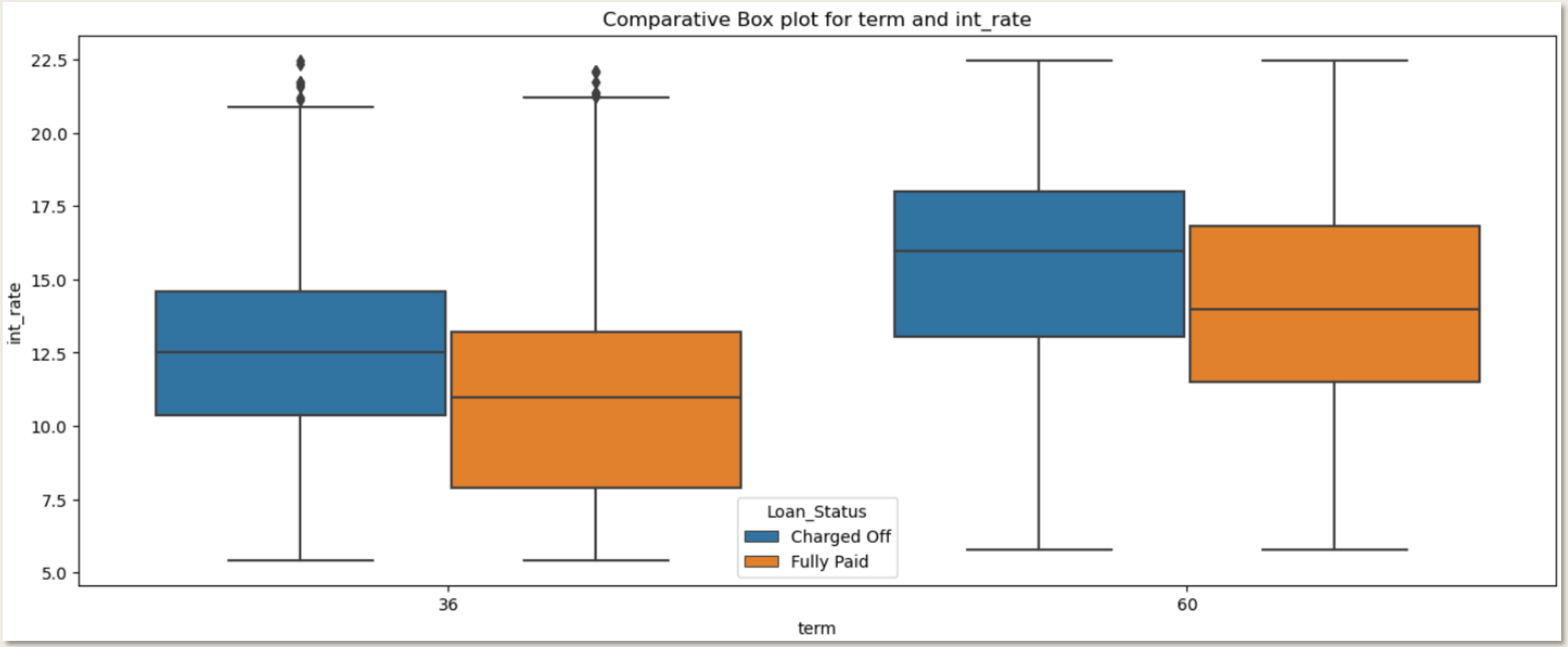
Univariate | **Bivariate** (Numerical vs Numerical) | Multivariate



It is observed that those customers are more likely to default whose debt to income ratio is more than 25 or the loan amount is more than 25000(Irrespective of DTI value)

Analysis

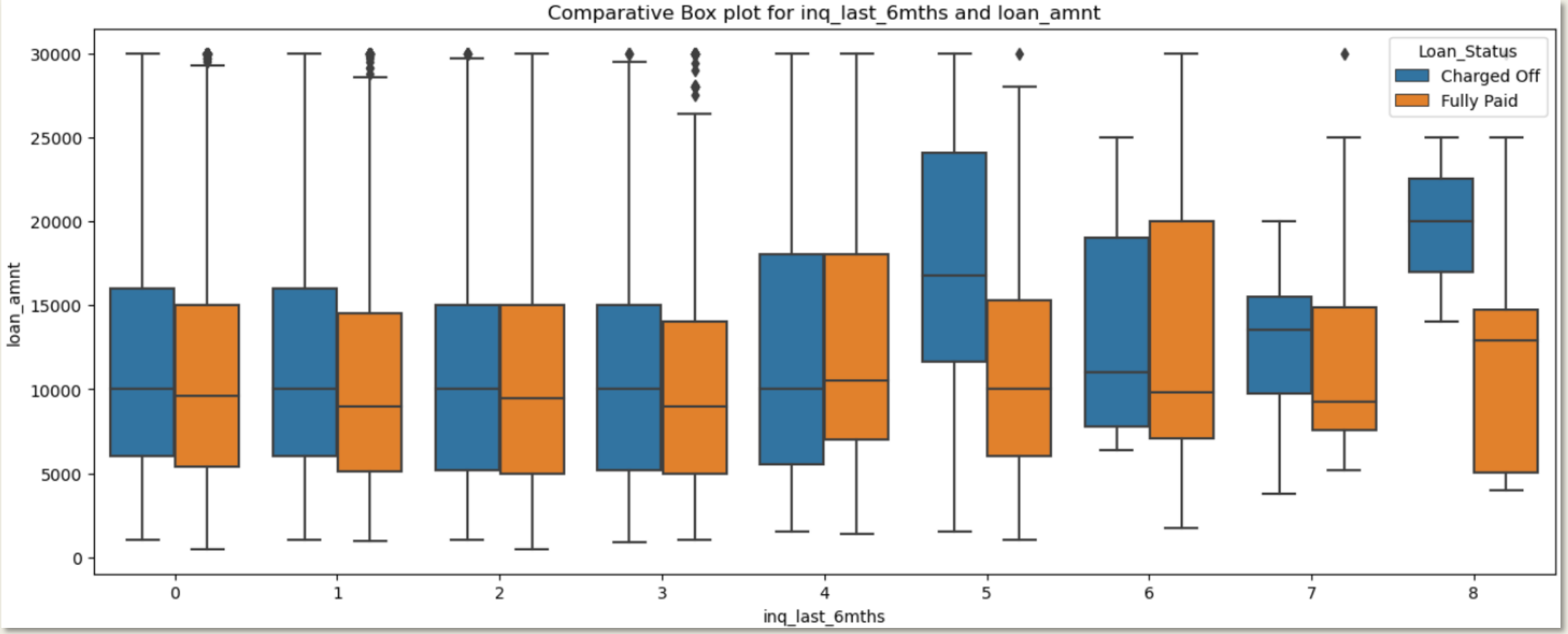
Univariate | **Bivariate** (Numerical vs Categorical) | Multivariate



It is evident that for both 36 month and 60 months loan term, the median of interest rate has been higher for defaulted loans vs fully paid loans.

Analysis

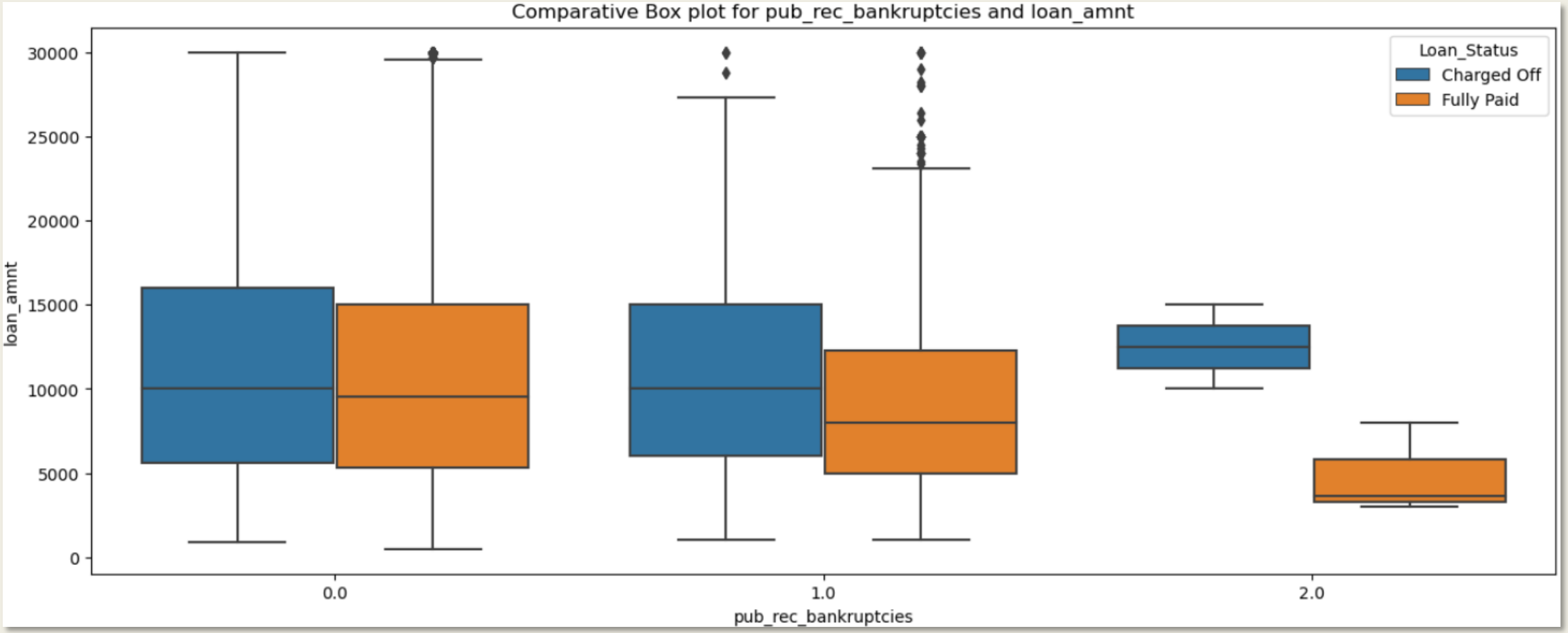
Univariate | **Bivariate** (Numerical vs Categorical) | Multivariate



It is strongly evident that any customer who has made 8 inquiries and had loan amount greater than 15000 is almost likely to default. This can be a good indicator for a prediction model.

Analysis

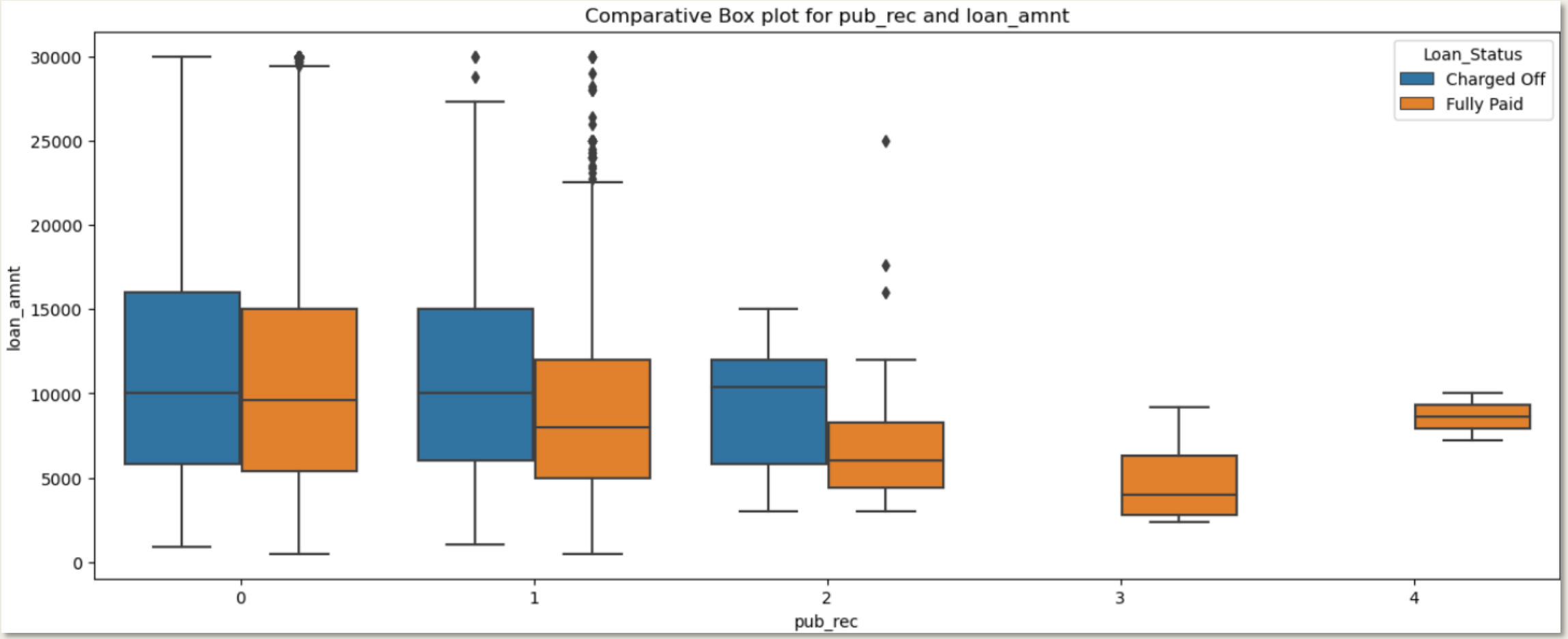
Univariate | Bivariate (Numerical vs Categorical) | Multivariate



It is strongly evident that any customer with 2 bankruptcy records with a loan amount greater than 10000 is almost likely to default. This can be a good indicator for a prediction model.

Analysis

Univariate | Bivariate (Numerical vs Categorical) | Multivariate



It is strongly evident that any customer with more than 2 public records will not default irrespective of the loan amount. This can be a good indicator for a prediction model.

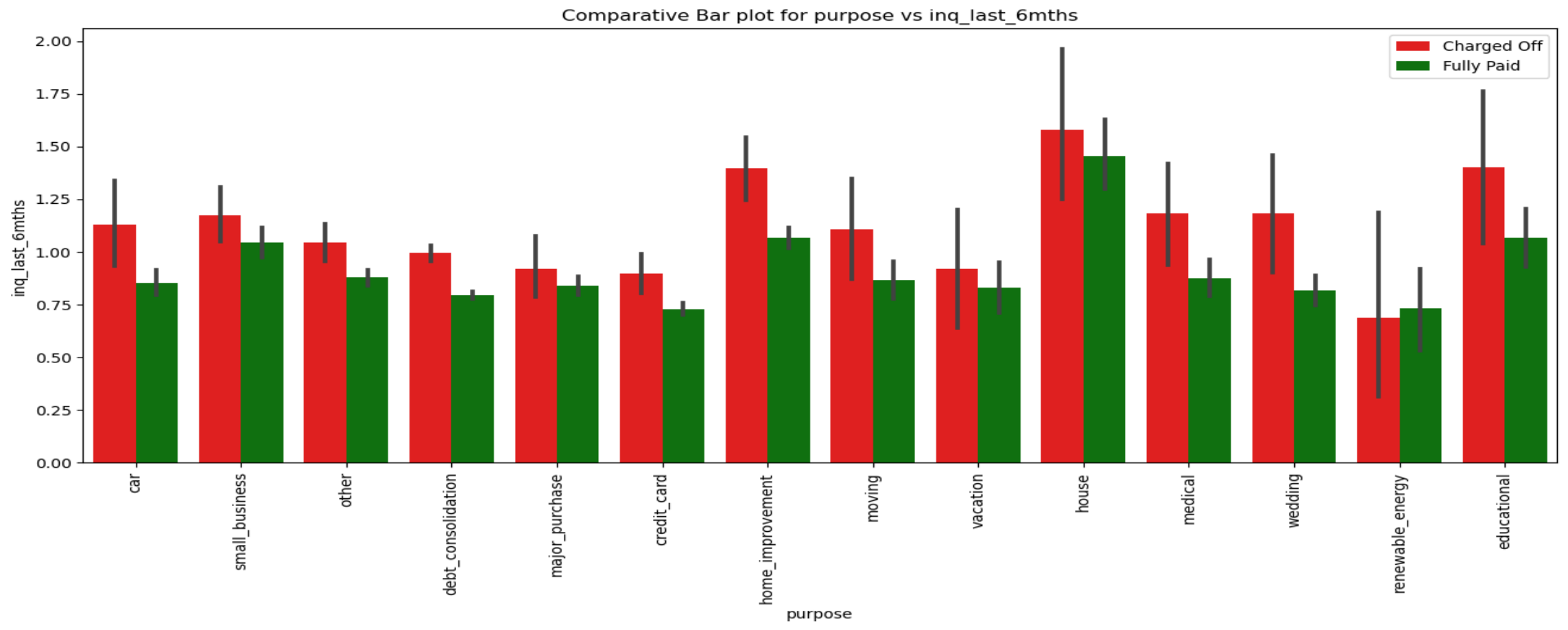
Analysis

Univariate

Bivariate

(Categorical vs Categorical)

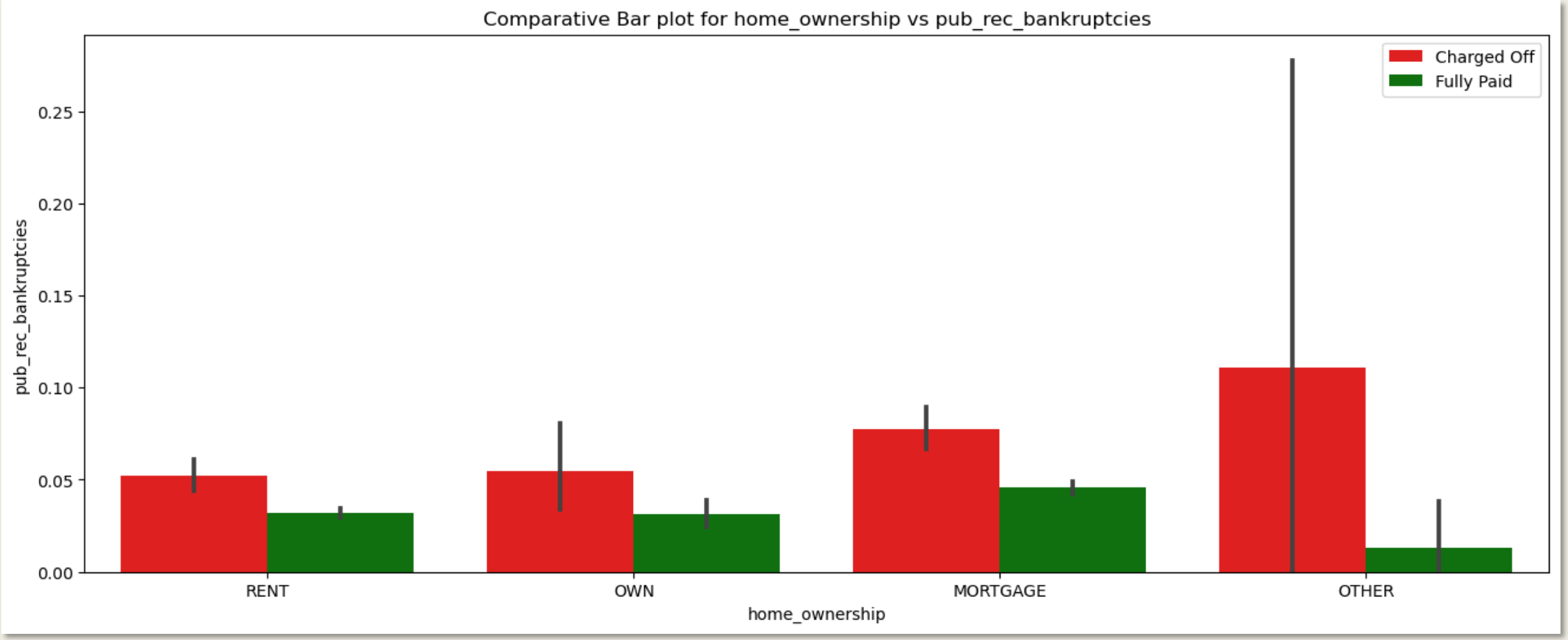
Multivariate



Customers who inquired in last 6 months for the purpose of house, home improvement or education purposes have more defaulters when compared to other purposes.

Analysis

Univariate | **Bivariate** (Categorical vs Categorical) | Multivariate



Customers with home ownership of OTHER followed by MORTGAGE have higher public record bankruptcies.

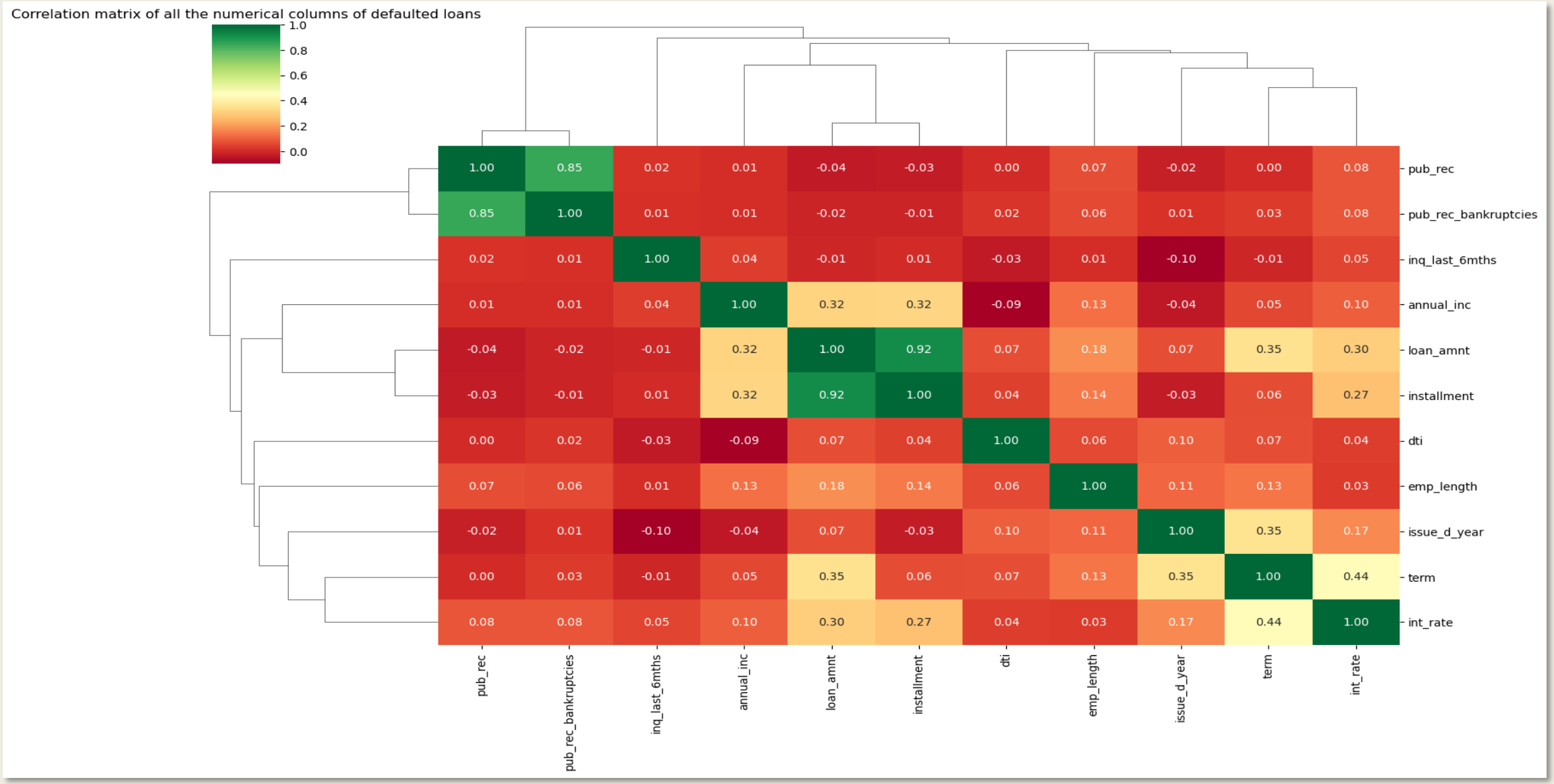
Analysis

Univariate

Bivariate

Multivariate

(Correlation Matrix)



Observations from Correlation Matrix

- Public record and public record bankruptcies are highly correlated
- Annual income and debt to income ratio is least correlated.
- Annual income have low correlation to the customers public record bankruptcies.
- Loan amount and installment have high correlation.
- Employment length and inquiry in the last 6 months have low correlation.
- Interest rate and emp length have low correlation. Employees are not offered the interest rates based on the employment length.
- Interest rate have strong correlation with the loan term. Most customers with higher interest rate have taken higher loan term.

Conclusions

Charged Off

- It is observed that irrespective of loan amount, the customer is more likely to default if the interest rate is higher.
- Customers are more likely to default whose debt to income ratio is more than 25 or the loan amount is more than 25000(Irrespective of DTI value). Lending club should consider DTI metric to be less than 25 while issuing the loans.
- Grades and sub-grades are good metric for detecting defaulters. B grade loans have highest probability of defaulting followed by grades C and D. Sub-grades B5, B3 and C1 have highest probability of defaulting.
- Lending Club should consider giving lesser loan amount for customers with 2 bankruptcy public records. It has been observed that almost all the customers with 2 bankruptcy cases have defaulted and their median loan amount has been around 13000.
- To predict any upcoming default, lending company can use below indicators:
  - *Customers has made 8 inquiries and has a loan amount greater than 15000.*
  - *Customer has 2 bankruptcy records and has a loan amount greater than 10000.*

Fully Paid

- A4, A5 and B3 sub grade loans have highest probability of being fully paid off. So, lending club can consider issuing more of these.
- Looking at the data it is strongly evident that any customer with more than 2 public records will not default irrespective of the loan amount. Lending club can use this information to grant loans with confidence.

THANK YOU