Lending Club Case Study

By
Mangai Maharajan
Manthi Ramesh

Lending Club Case Study

Description :

This study is based on a consumer finance company which specializes in lending various types of loans to urban customers.

Company has to take decision to approve the loan application based on data of past loan applicants to reduce/avoid defaulters hence reduce the financial loses to the company.

Task :

Our aim is to study **consumer attributes** and **loan attributes** with the help of data from past loan applicants using EDA then identify patterns, which leads to a person is likely to default.

Libraries and data set used for case study :

Libraries:	Data set:
• Numpy • Pandas • Matplotlib.pyplot • Seaborn	Loan.csv Number of Rows in the Dataset= 39717 Number of Columns in the Dataset= 111

Data Cleaning and Manipulation :

- After dropping null values left out Rows = 39717, Columns = 57
- After removing single valued columns, which are not useful for analysis remaining Rows = 39717, Columns = 48
- Columns such as **id, member_id, url, zip_code, last_credit_pull_d,** and **desc** do not contribute to loan risk analysis. Hence, removed these columns from the dataset.
- Columns with **more than 60% missing values** cannot be imputed without introducing bias into our analysis. Therefore, removed these columns.

  Final shape of the data set is 39717 rows and 27 columns.

Filtering Loan Status for Analysis:

The objective of this analysis is to determine the likelihood of loan default. This can only be assessed for loans that are either fully paid or charged off. Current loans cannot provide meaningful insights into default risk. Therefore, we excluded records with a current loan status from our dataset.

```
## exclude records with a current loan status from our dataset
loan_data=loan_data[~(loan_data['loan_status']=='Current')]
loan_data.loan_status.value_counts()
```

```
loan_status
Fully Paid     32947
Charged Off     5627
Name: count, dtype: int64
```

Checking and Imputing Missing Values:

To ensure the quality of our data, we need to identify and handle missing values appropriately.

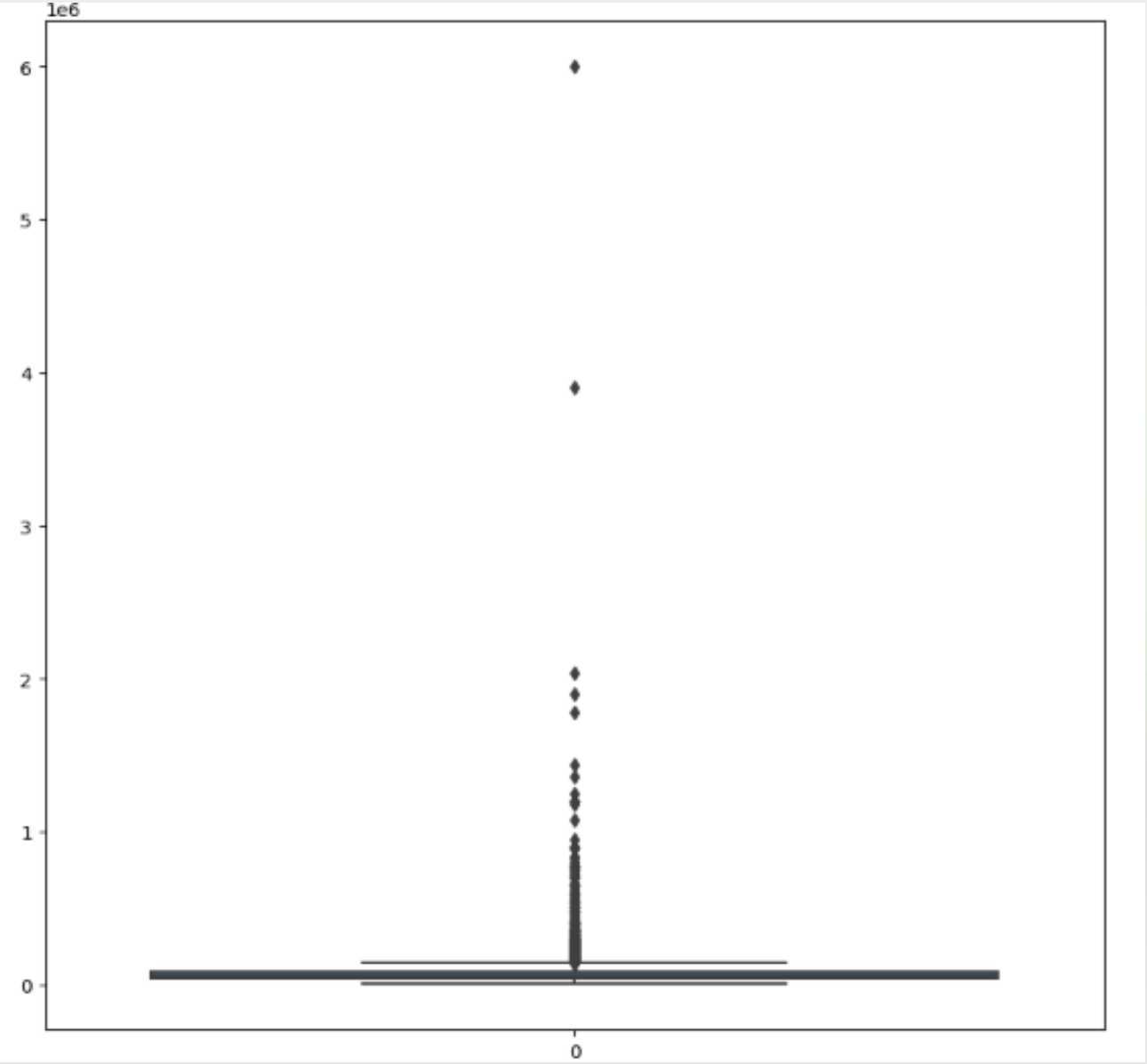
Imputing missing values helps maintain the integrity of the dataset without introducing bias.

```
missing_values=round(loan_data.isnull().mean()*100,2)
missing_values[missing_values>0]
```

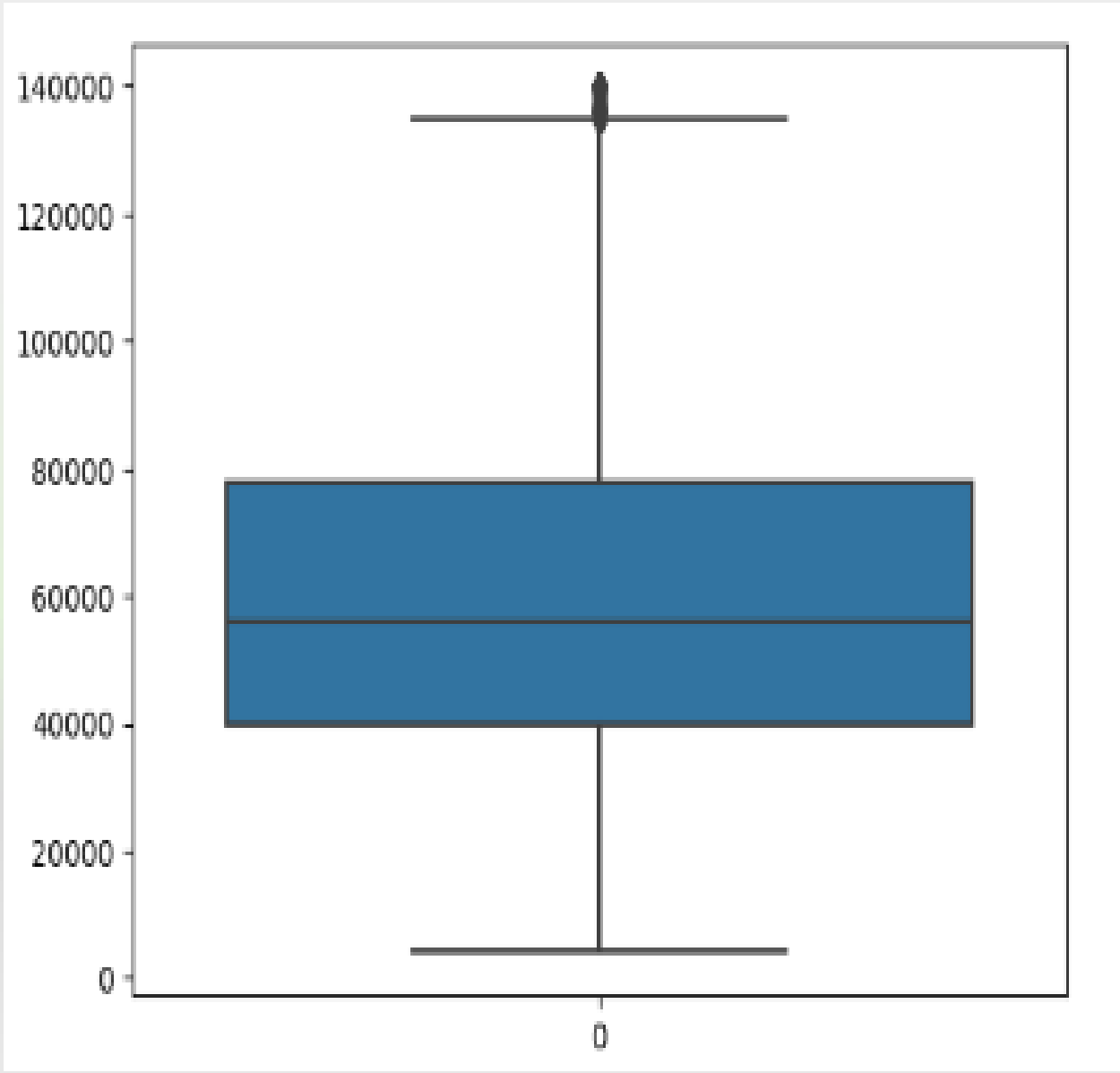
```
emp_title                6.19
emp_length               2.68
title                    0.03
revol_util               0.13
pub_rec_bankruptcies     1.81
dtype: float64
```

Imputed missing values with Mode to maintain integrity of data set.

Outlier Detection and Removal Using Boxplots :

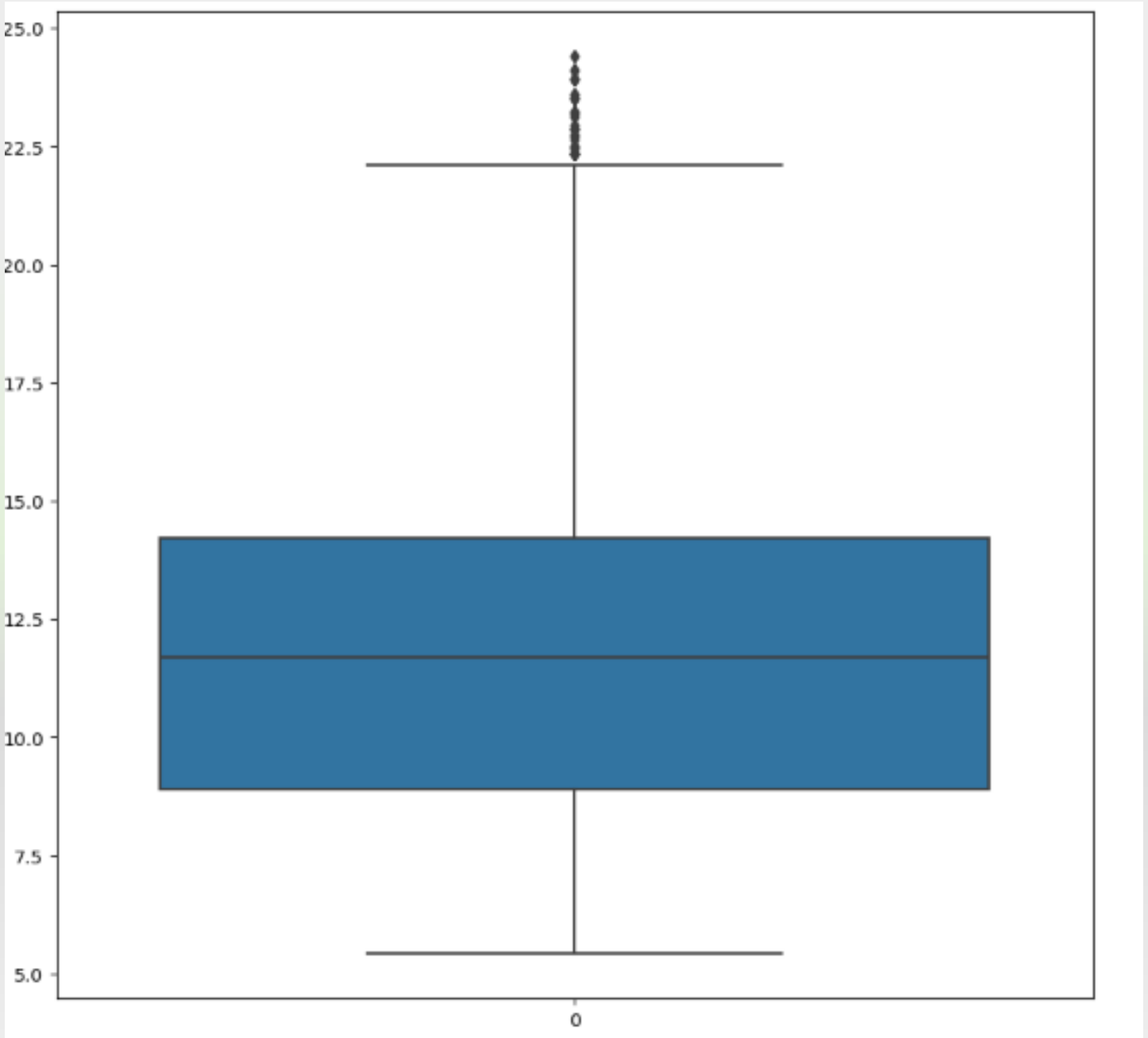


Annual Income Box Plot at 100%

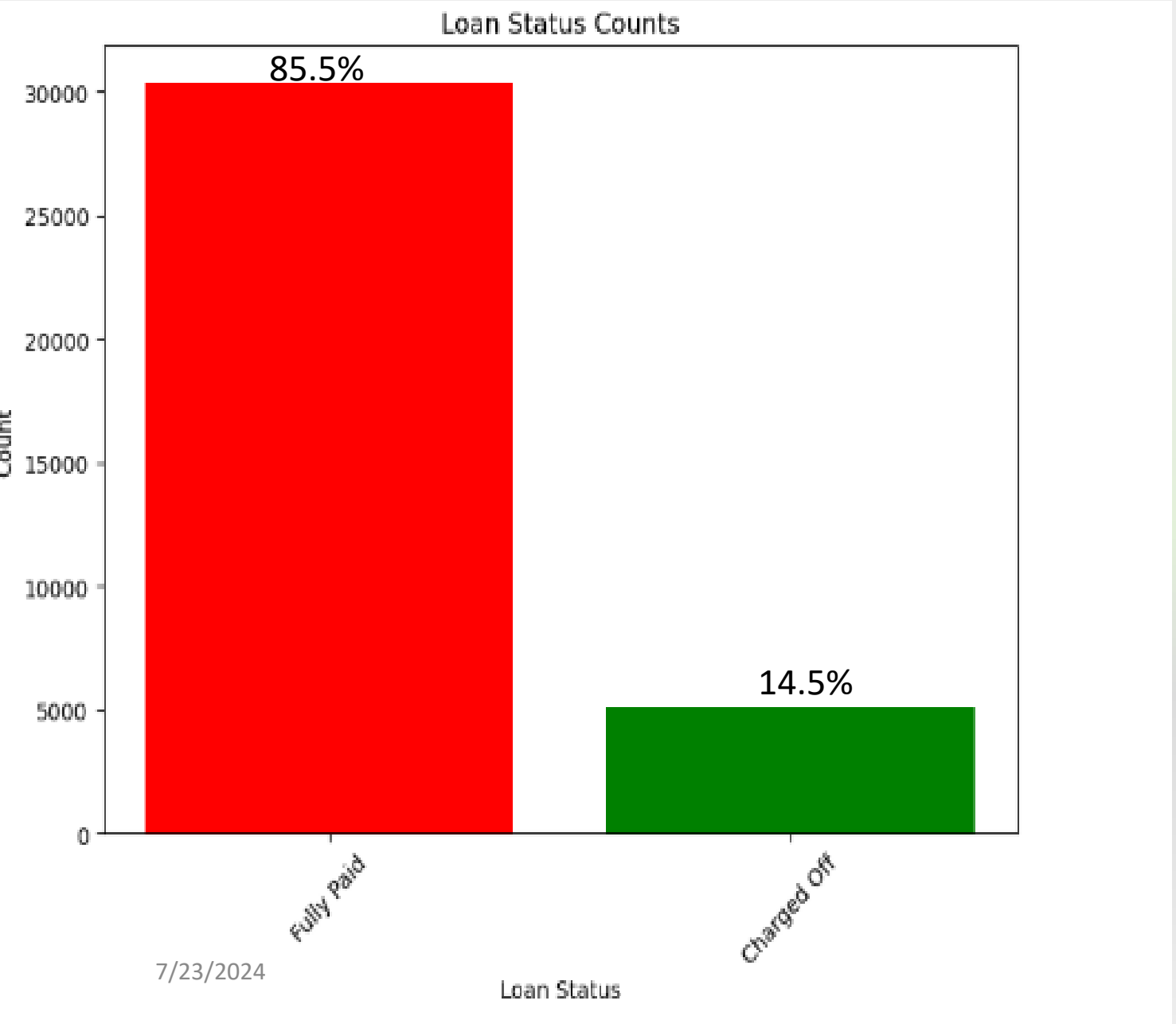


Annual Income Box Plot at 95% quantile

Interest Rate Box Plot



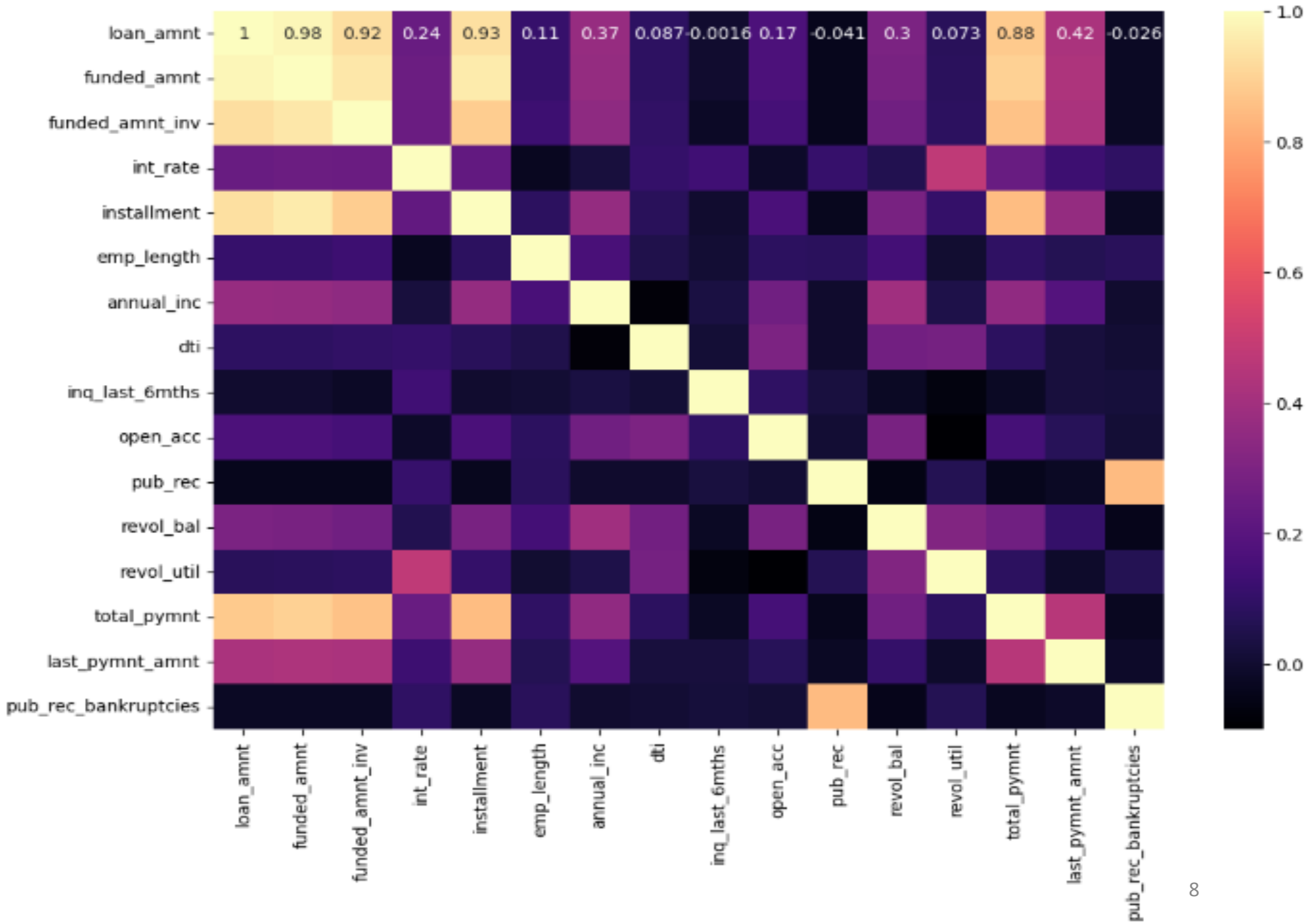
Bar Plot of Loan Status



Percentage of customers fully paid =85.5

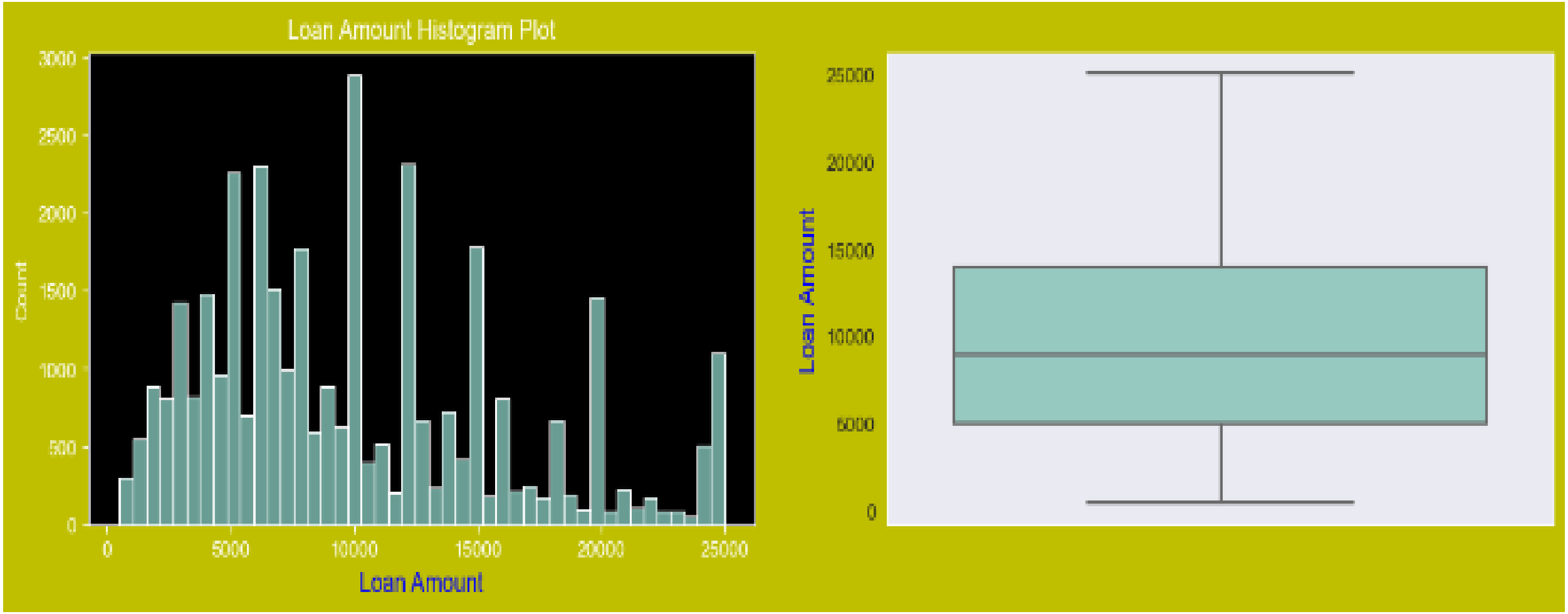
Percentage of customers Charged off =14.5

Heat Map


loan_amnt
funded_amnt
funded_amnt_inv
int_rate
installment
emp_length
annual_inc
dti
inq_last_6mths
open_acc
pub_rec
revol_bal
revol_util
total_pymnt
last_pymnt_amnt
pub_rec_bankruptcies
1 0.98 0.92 0.24 0.93 0.11 0.37 0.087 -0.0016 0.17 -0.041 0.3 0.073 0.88 0.42 -0.026
1.0
0.8
0.6
0.4
0.2
0.0

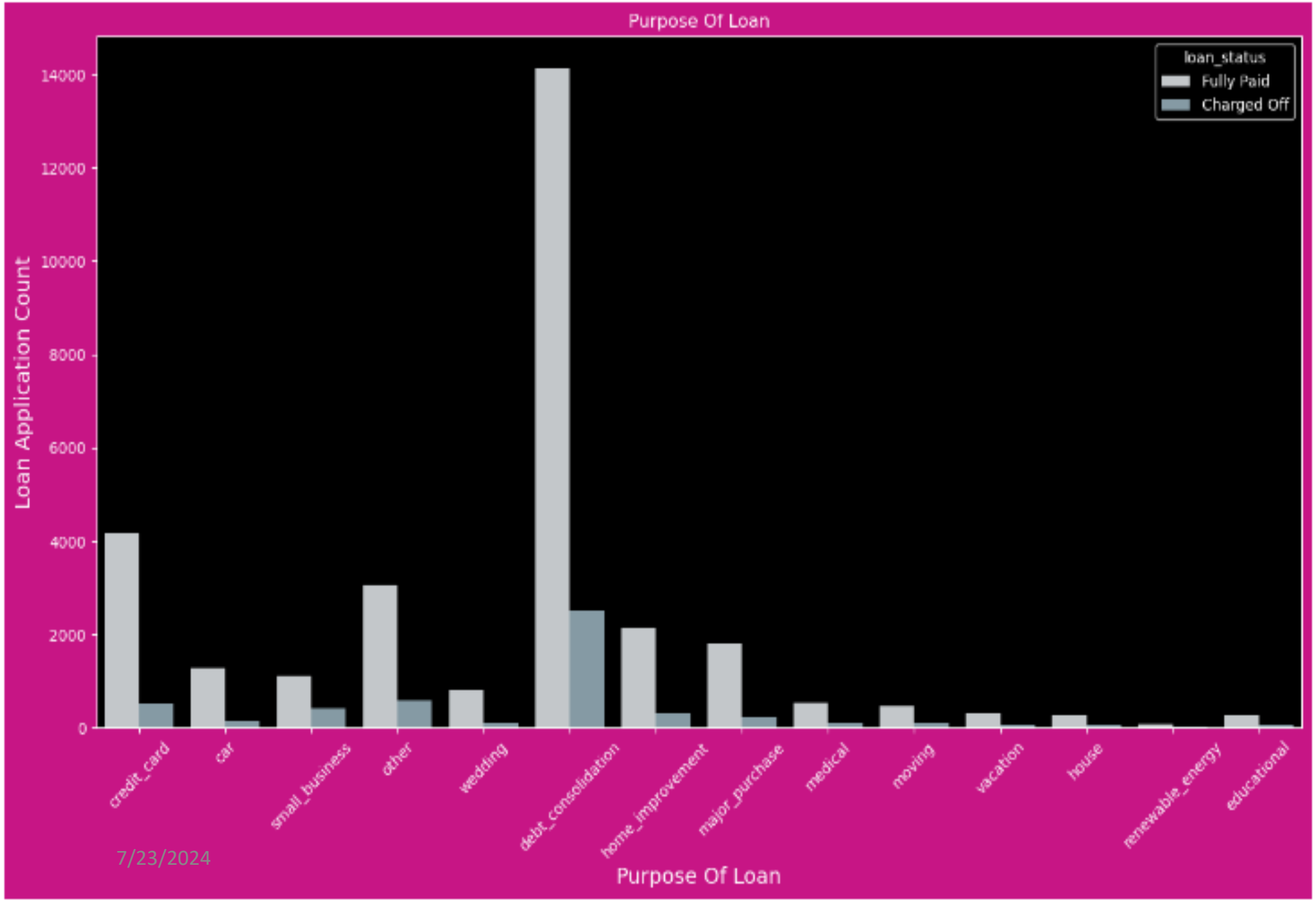
Univariate Analysis

Loan Amount Histogram and Box Plot



Observations : The majority of loan amounts fall within the range of 5000 to 15000

Histogram Plot for Purpose of Loan

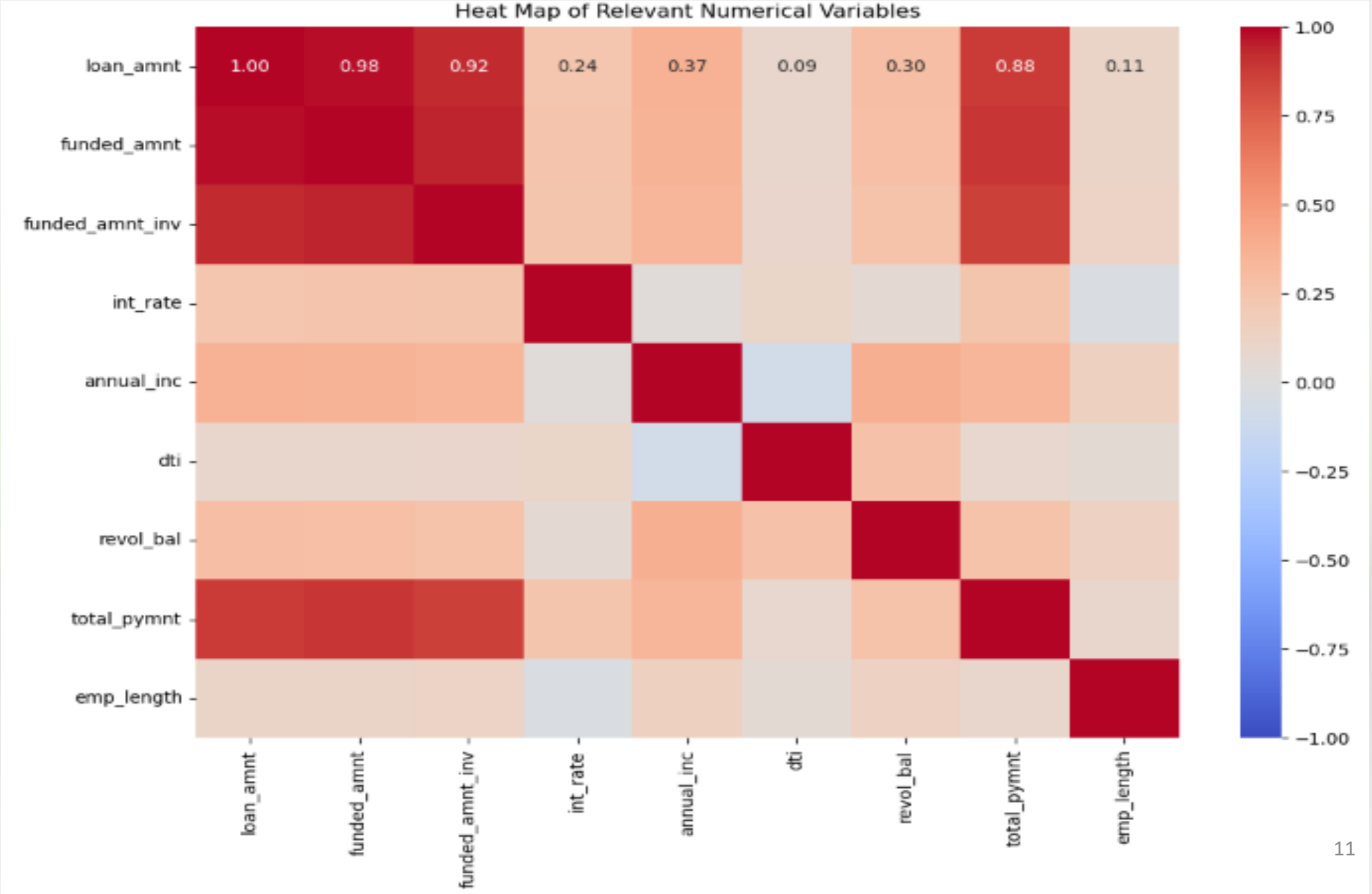


Observations :

The count plot shows that debt consolidation accounted for the majority of loans taken out. Furthermore, a very high number of loans that have been charged off are connected to these uses.

Bivariate Analysis

Heat Map



Observations

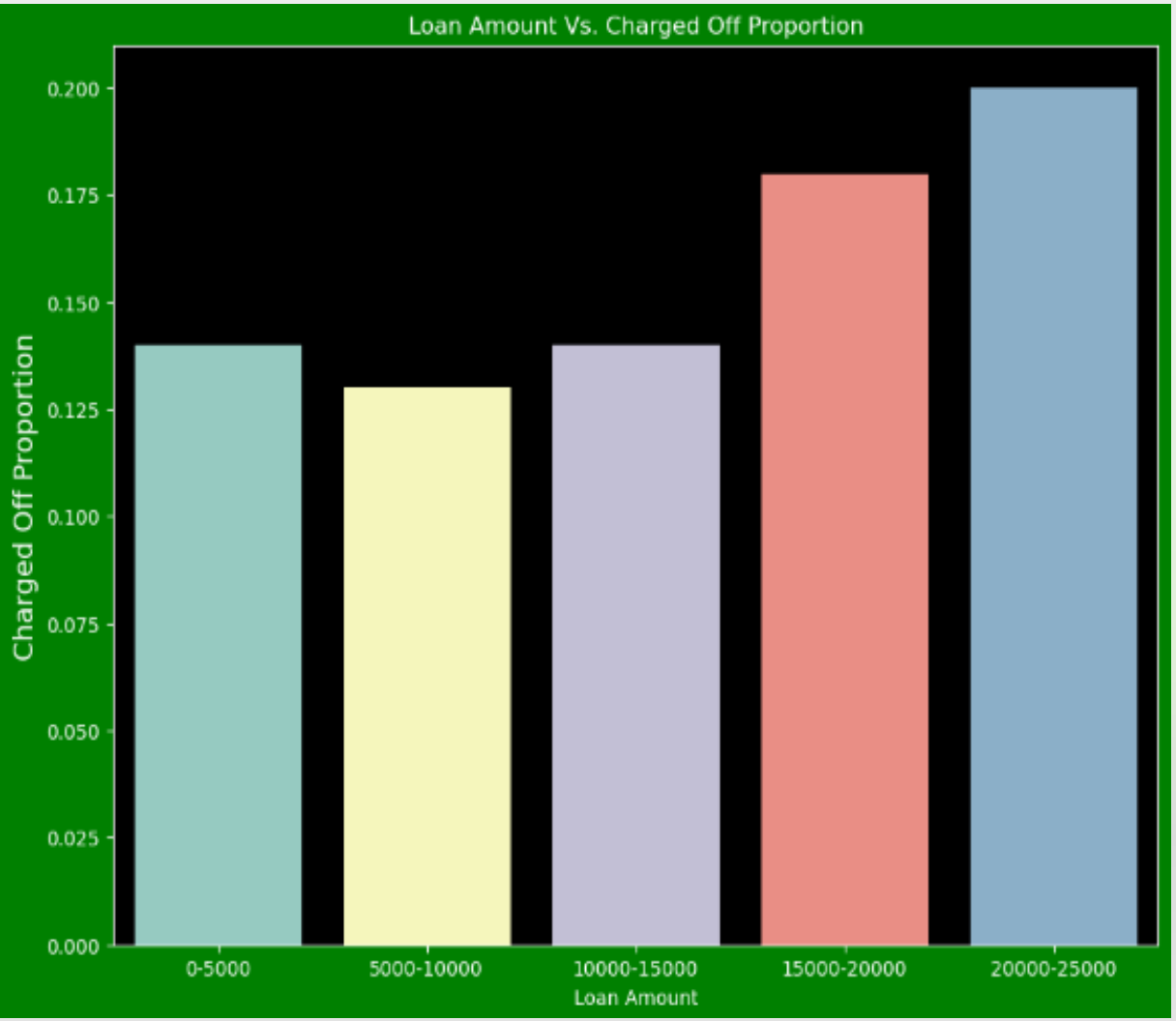
- **Loan amount, funding amount, and funded amount invested are strongly correlated**: These three variables are closely related, indicating that as one increases, the others tend to increase as well.
- **Interest rate and employee length are negatively correlated**: This suggests that longer employee tenure is associated with lower interest rates, possibly reflecting lower risk for long-term employed individuals.
- **Delinquency in the past 2 years and inquiries in the last 6 months are positively correlated with the interest rate**: This suggests a tendency towards higher interest rates for borrowers with recent delinquencies and numerous recent credit inquiries, indicating a higher risk of default.
- **Annual income is positively correlated with loan amount, revol_bal, and total payment**: Higher income individuals tend to take out larger loans, have higher revolving balances, and make higher total payments.
- **Debt-to-income (DTI) ratio is positively correlated with revolving balance**: This indicates that individuals with higher revolving balances tend to have higher debt-to-income ratios.
- **Total payment is strongly correlated with loan amount, funded amount, and funded amount invested**: This shows that larger loans lead to higher total payments over time.

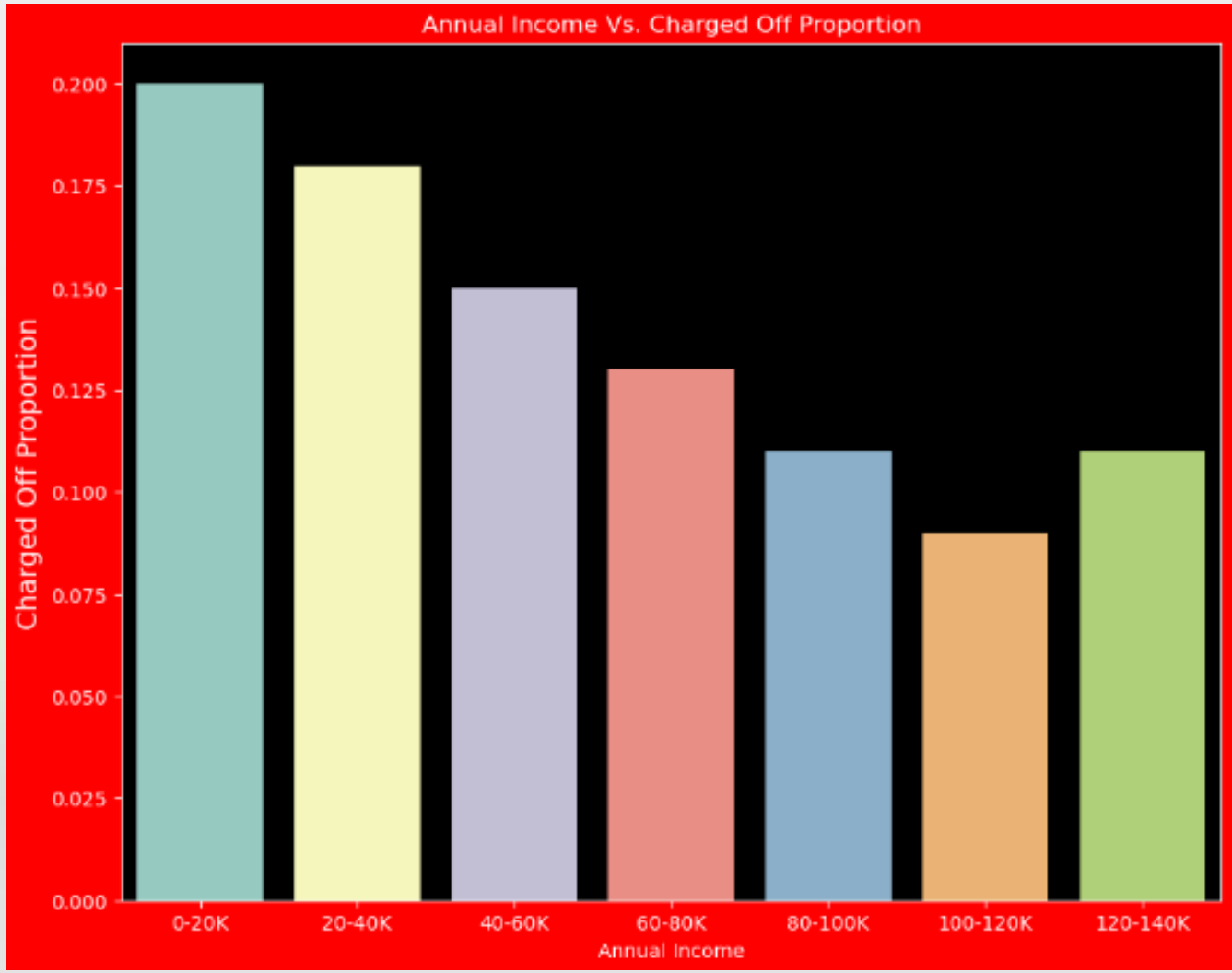
Bi variate Analysis, Part One with loan status

In this section, we will compare the loan status to certain important columns that might have an impact on debts that have been charged off.

- LOAN AMOUNT
- INTEREST RATE
- ANNUAL INCOME
- PURPOSE OF LOAN
- GRADE
- SUB GRADE
- EMP LENGTH (EMPLOYEE EXPERIENCE)
- DTI
- HOME OWNERSHIP
- VERIFICATION STATUS
- ADDRESS STATE
- PUB_REC_BANKRUPTICIES (BANKRUPTICIES)

Bi variate Analysis, Part One with loan status

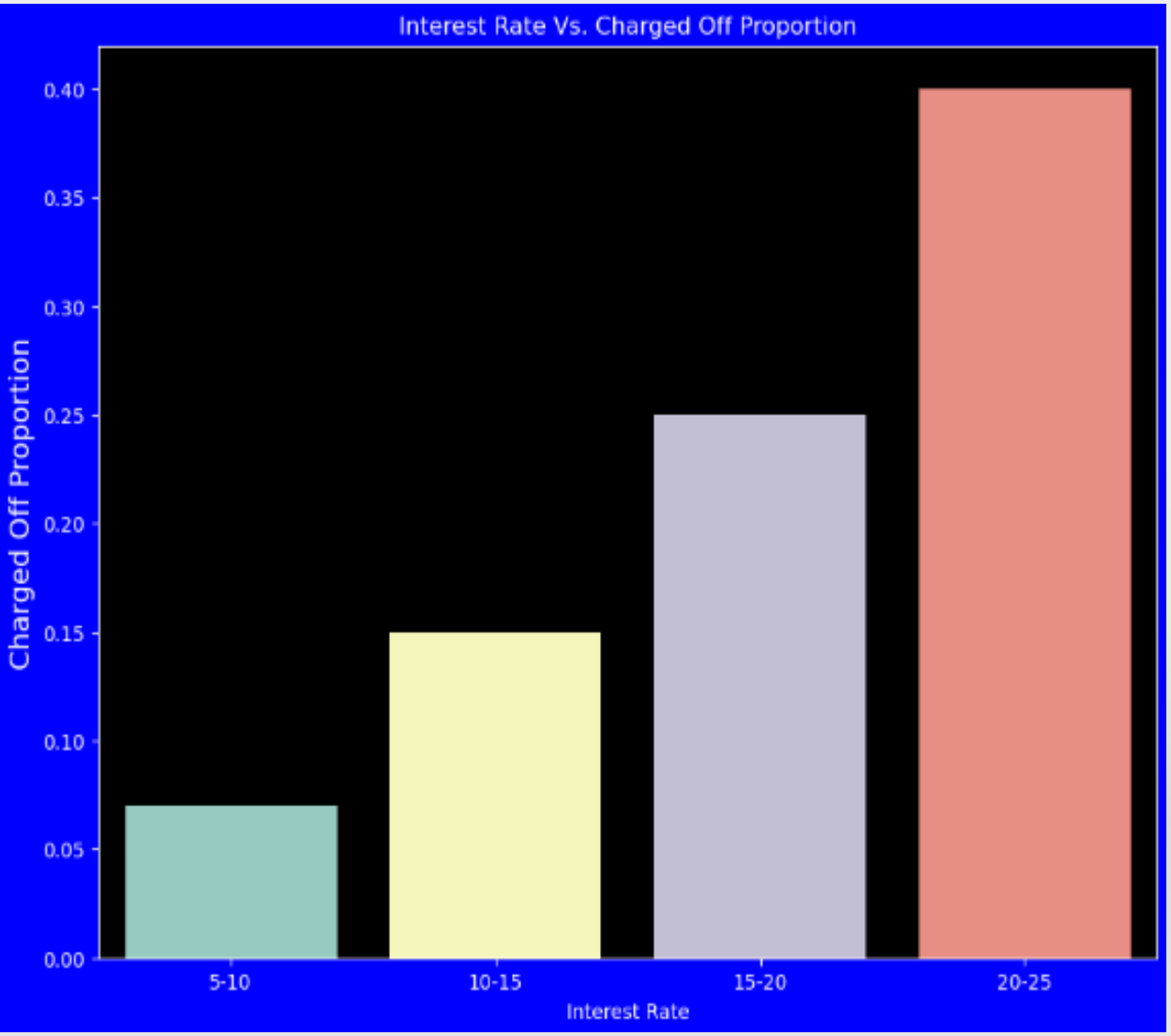


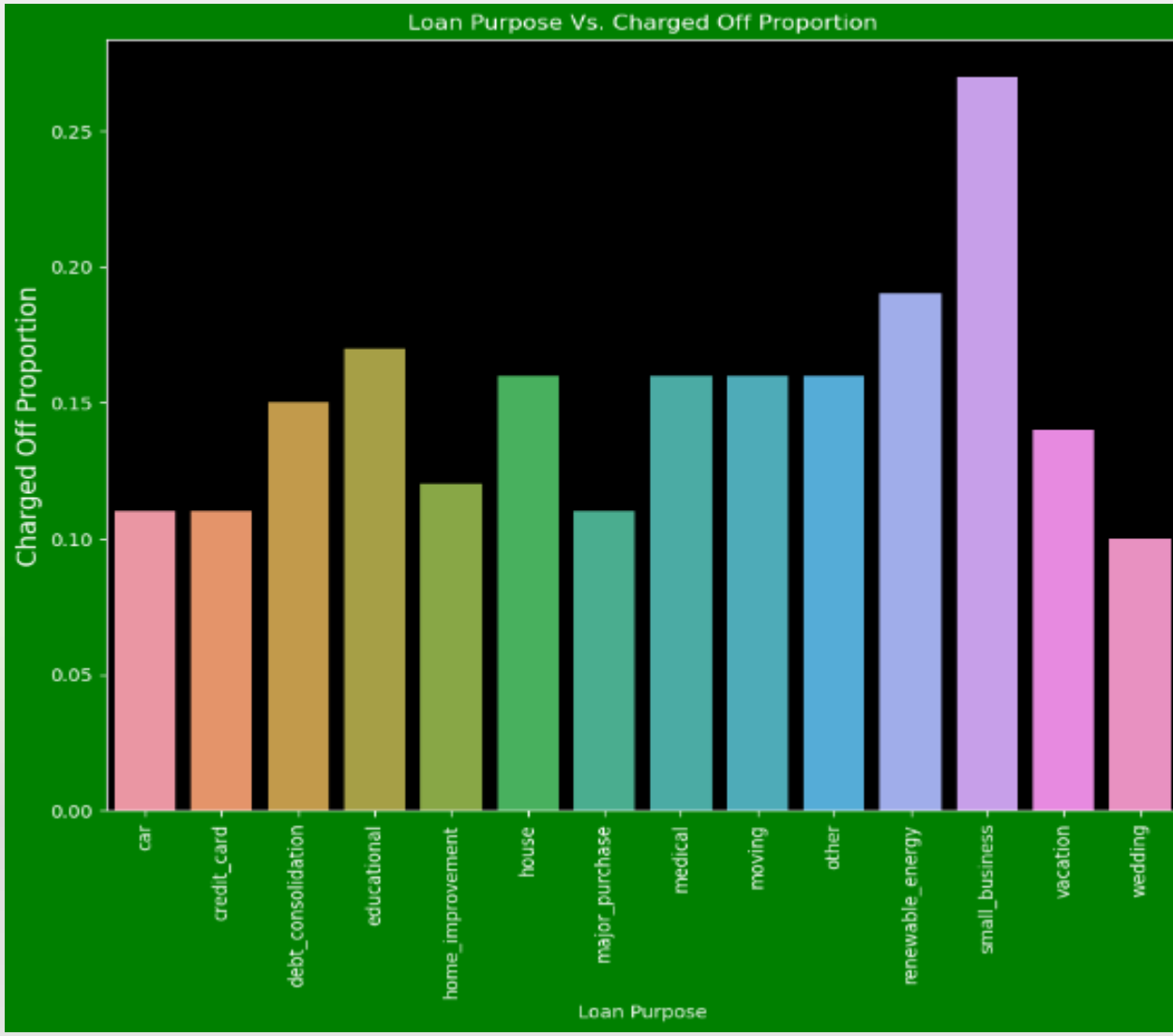


Loan Amount Vs Charged off Proportion

Annual Income Vs Charged off Proportion

Bi variate Analysis, Part One with loan status

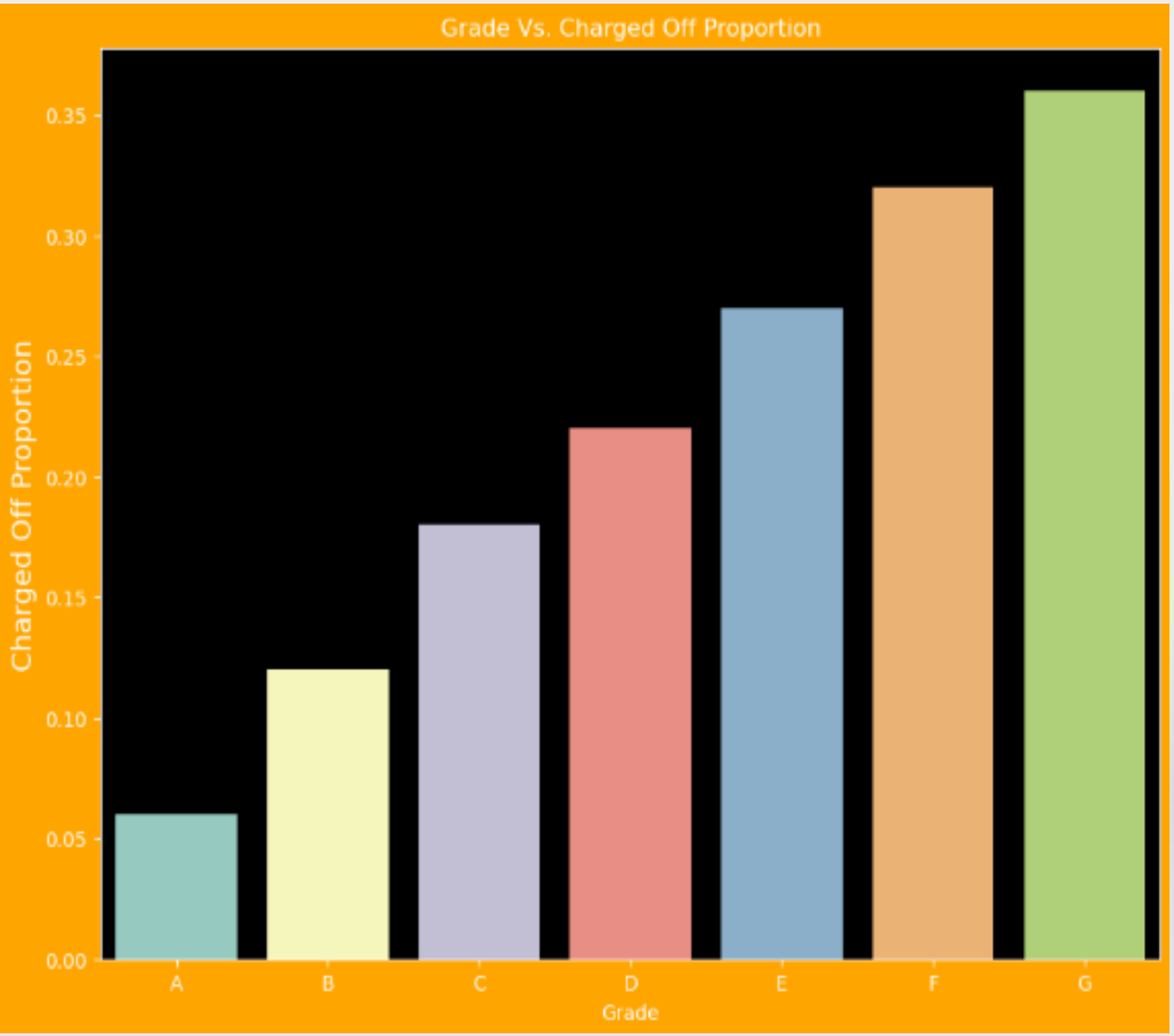




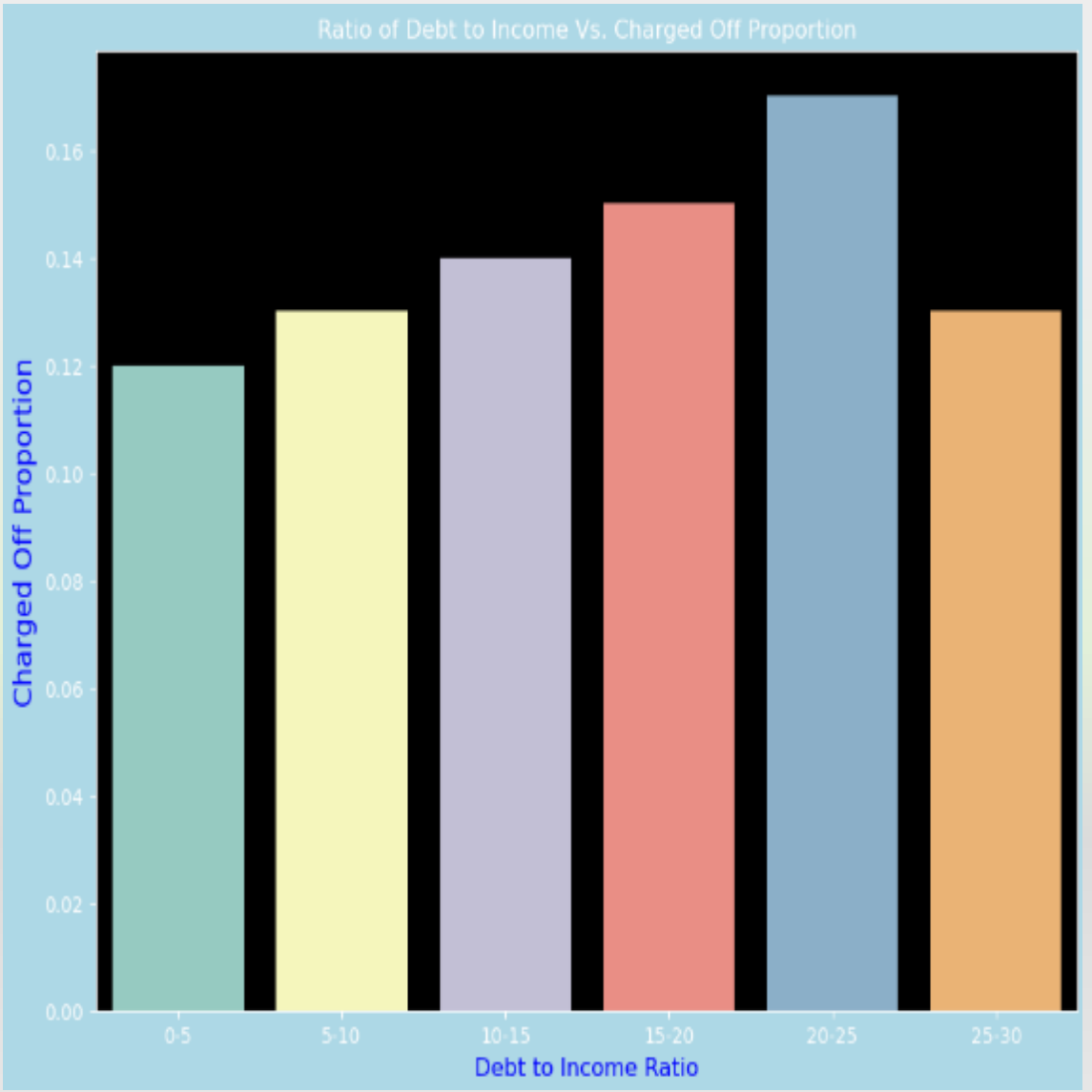
Interest Rate Vs Charged off Proportion

Loan Purpose Vs Charged off Proportion

Bi variate Analysis, Part One with loan status

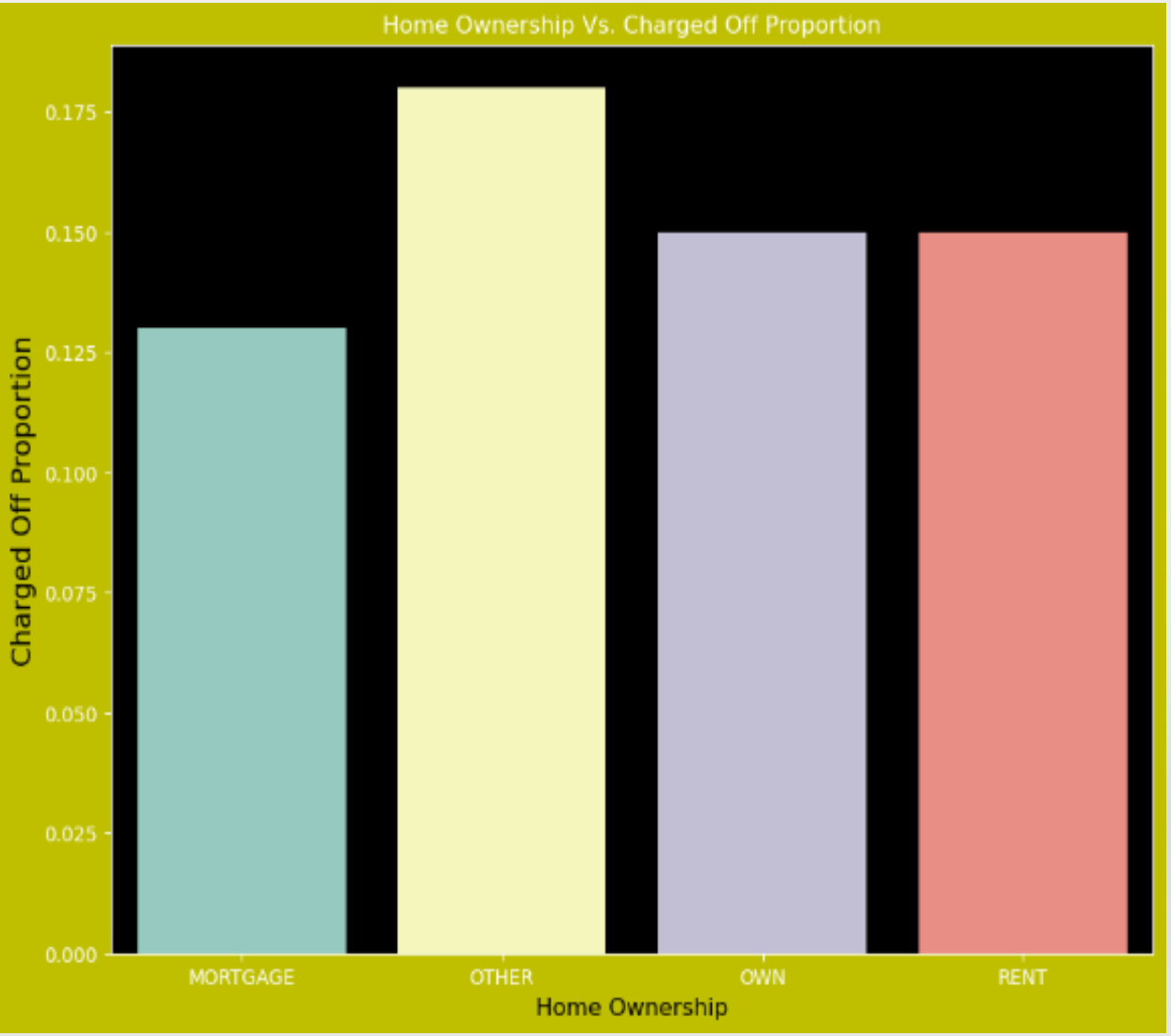


Grade Vs Charged off Proportion

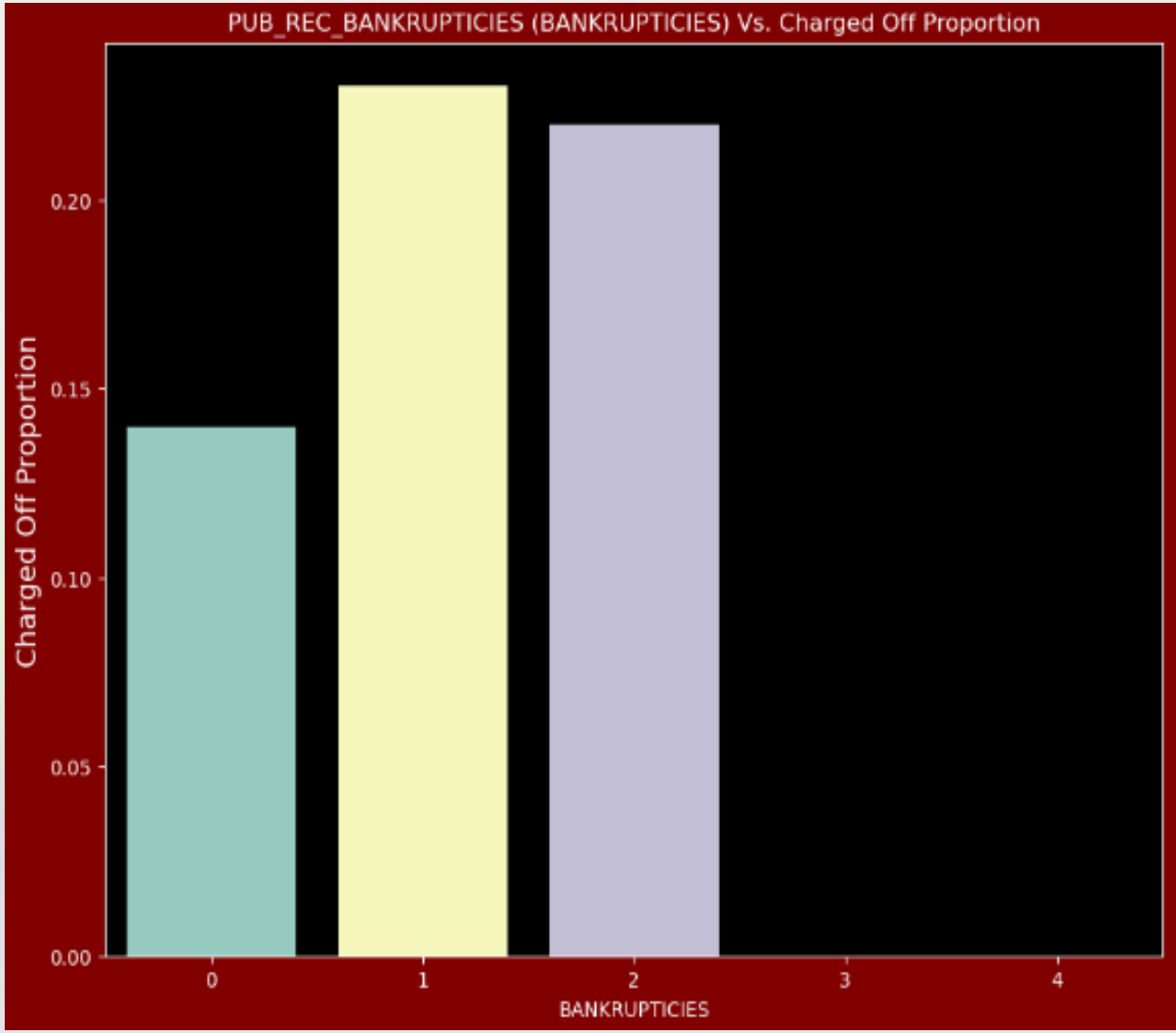


Debt to Income Ratio Vs Charged off Proportion

Bi variate Analysis, Part One with loan status

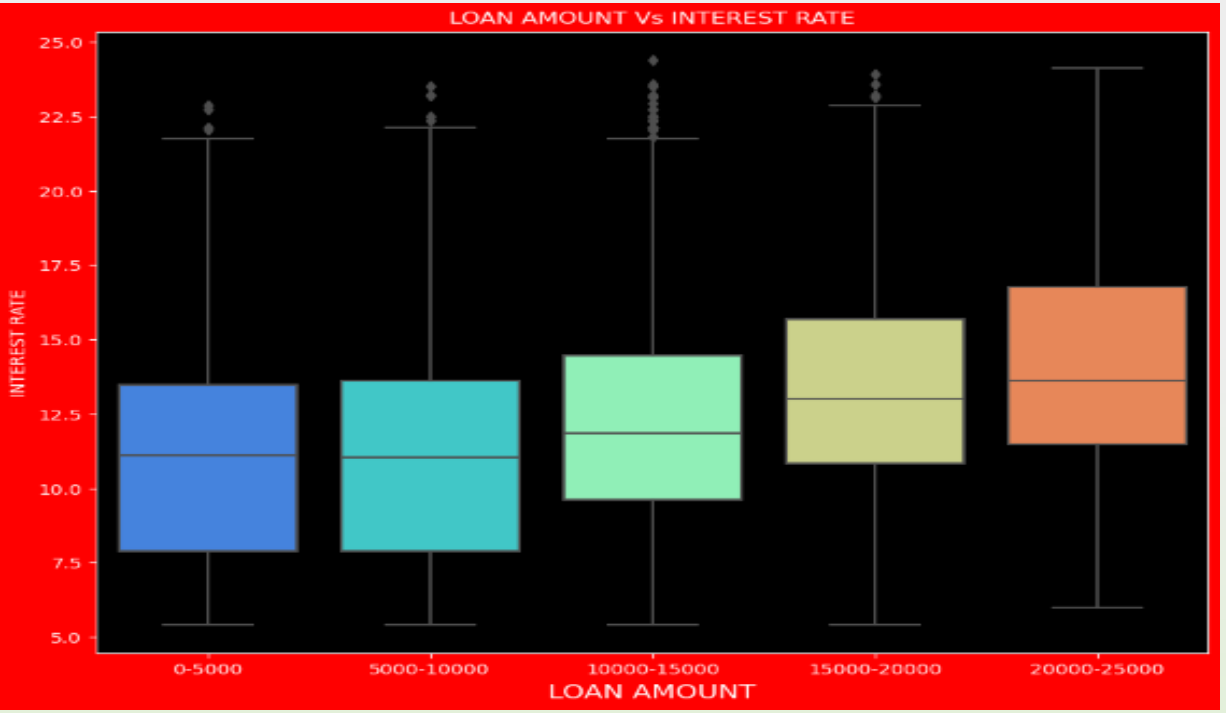


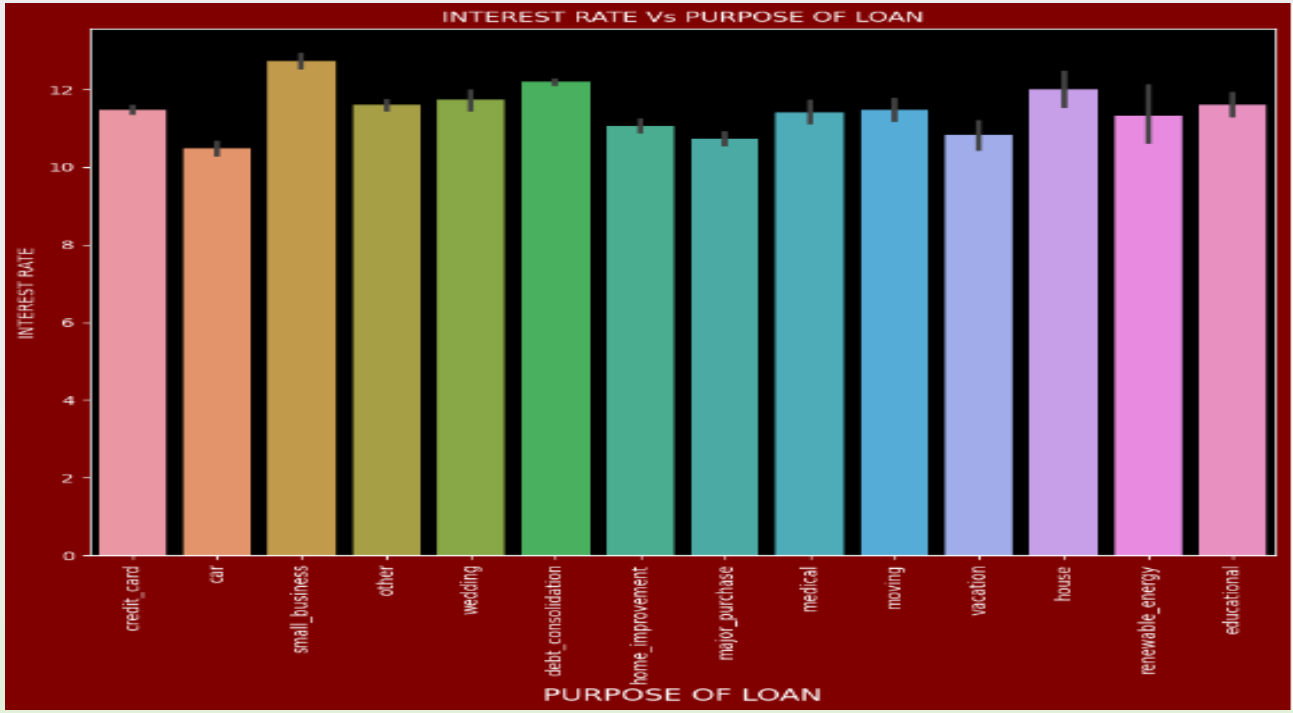
Home Ownership Vs Charged off Proportion



Bankruptcies Vs Charged off Proportion

Bi variate Analysis with Two Different components, Part Two



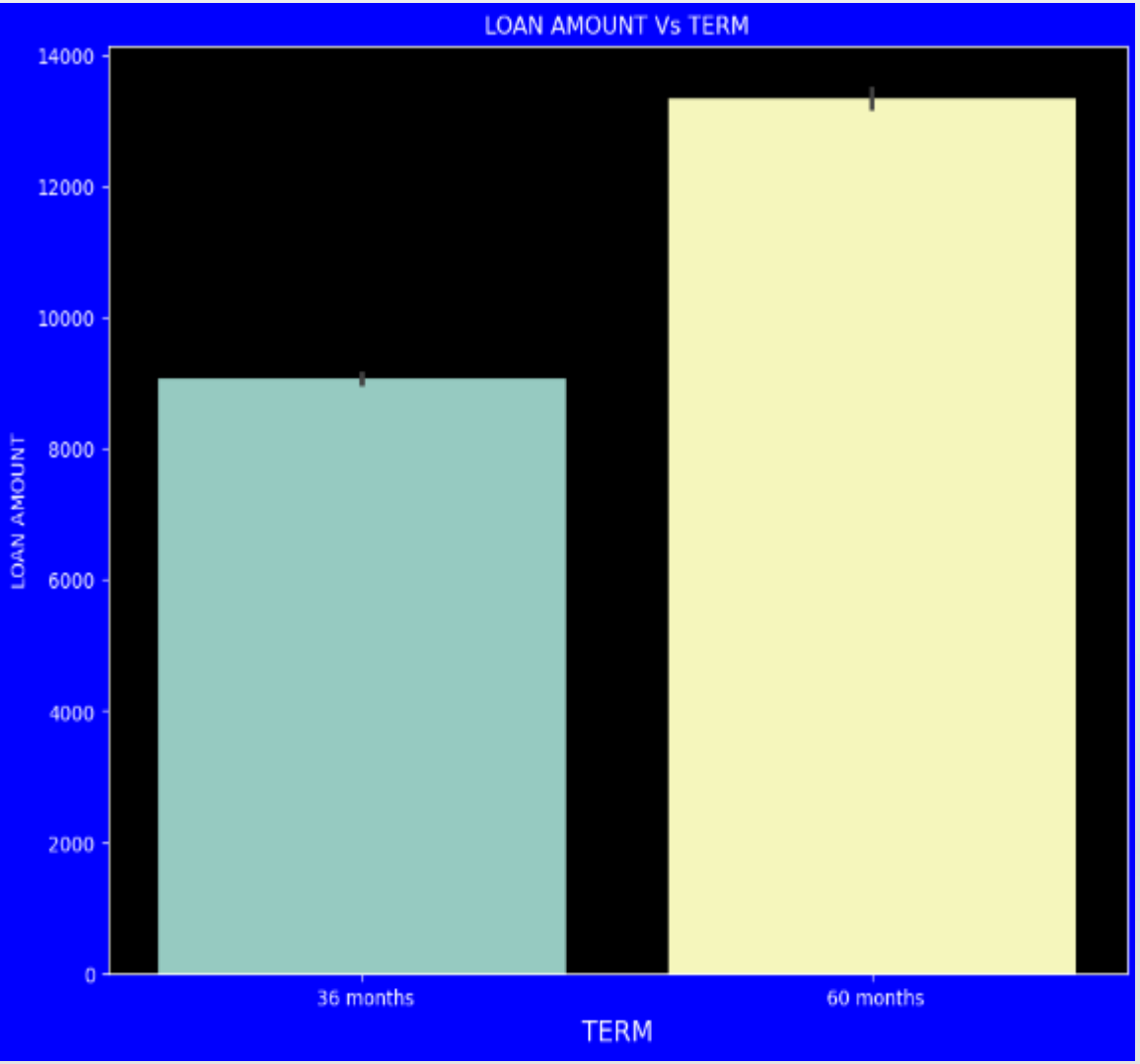


Loan Amount Vs Rate of Interest

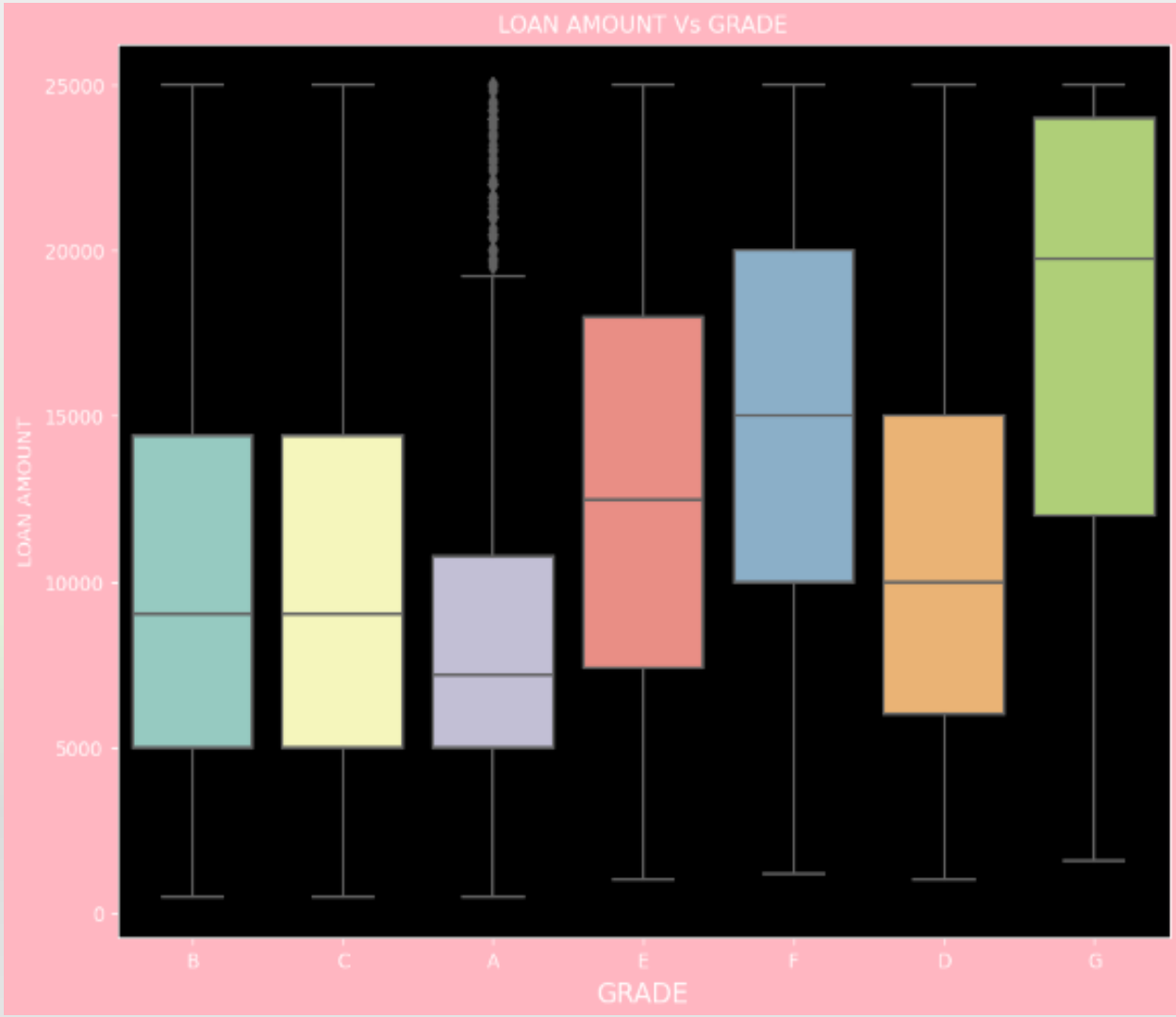
Loan Purpose Vs Rate of Interest

Debt to Income Vs Rate of Interest

Bi variate Analysis with Two Different components, Part Two

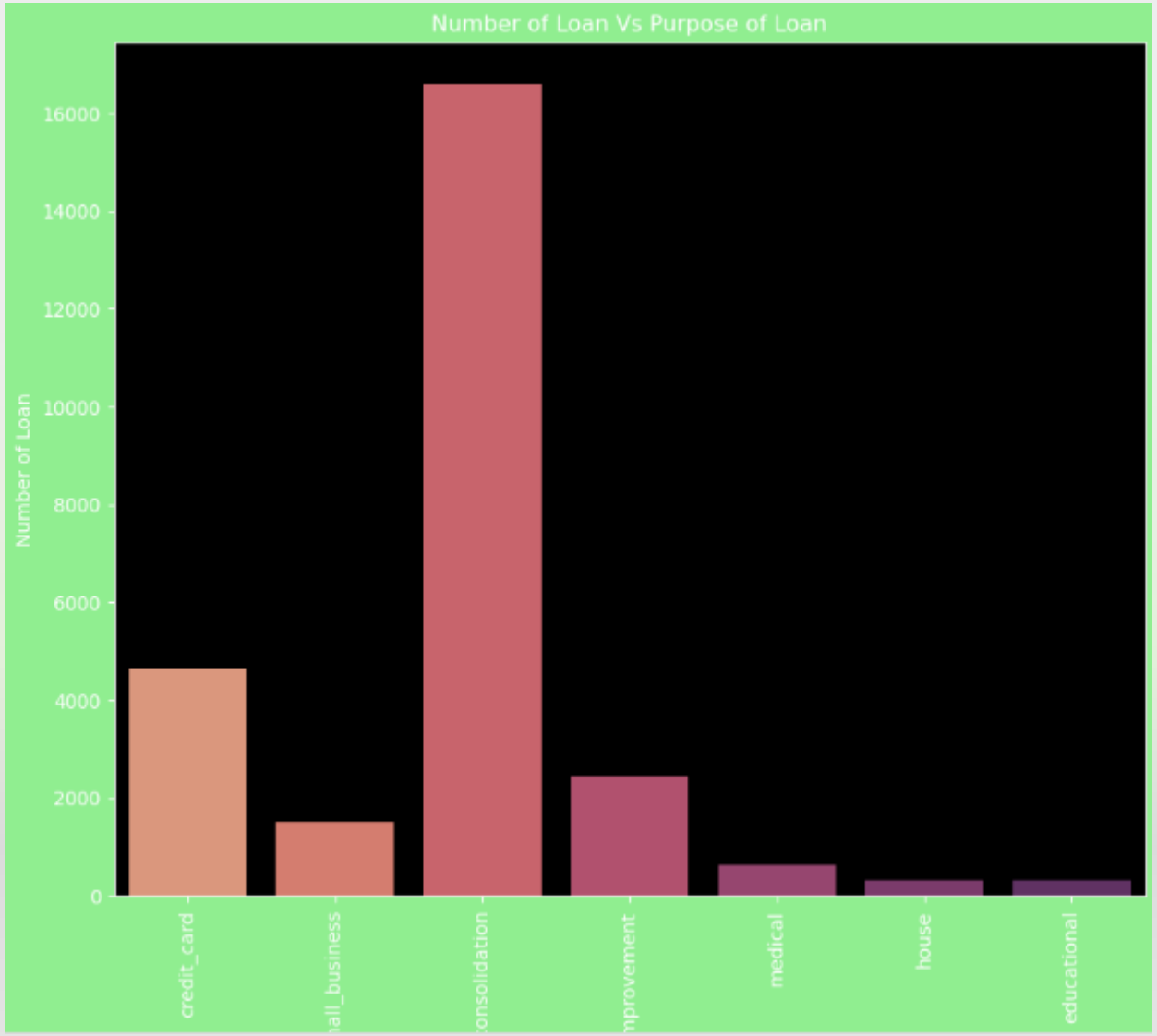


Loan Amount Vs Term

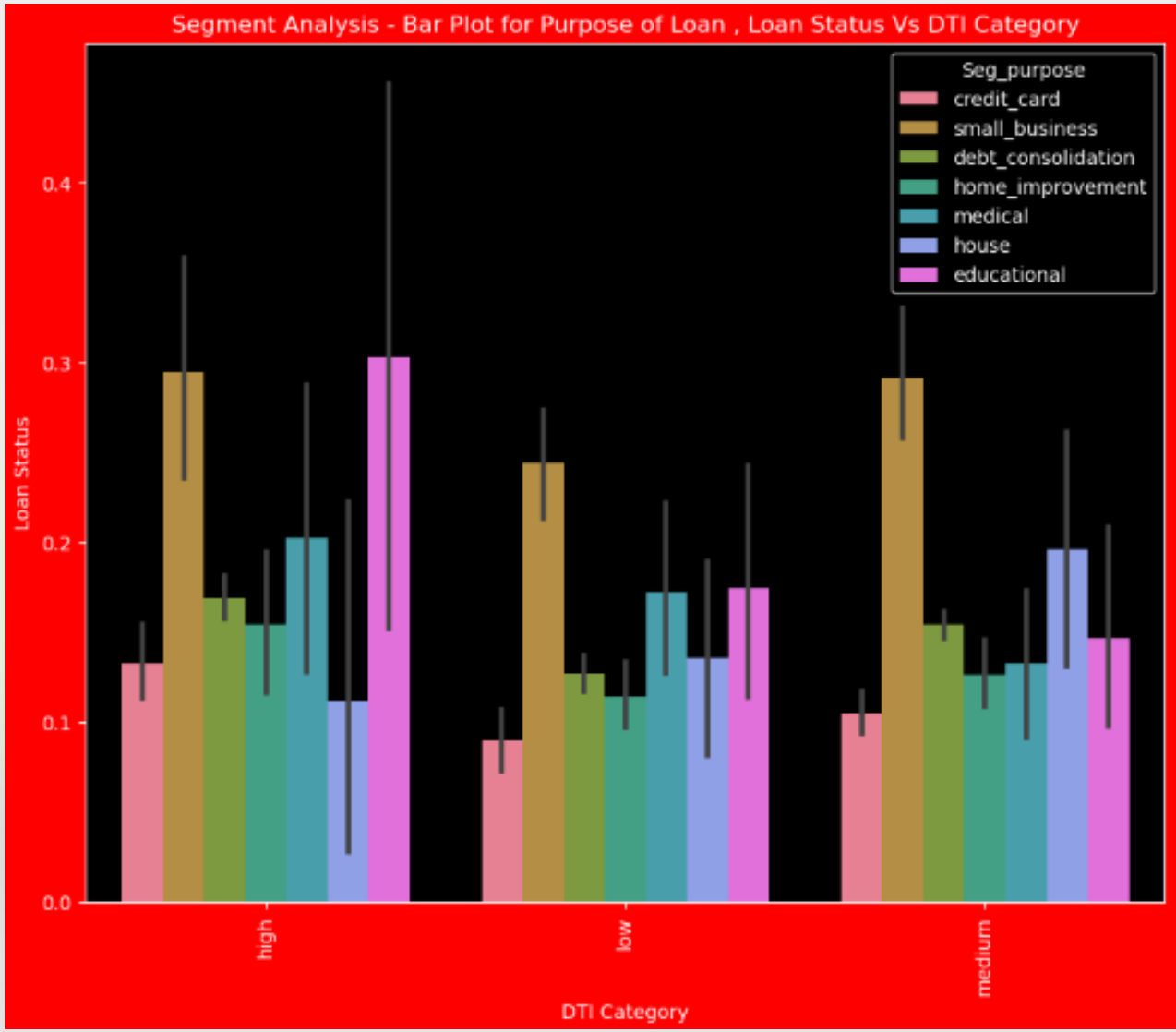


Loan Amount Vs Grade

Segmented Analysis



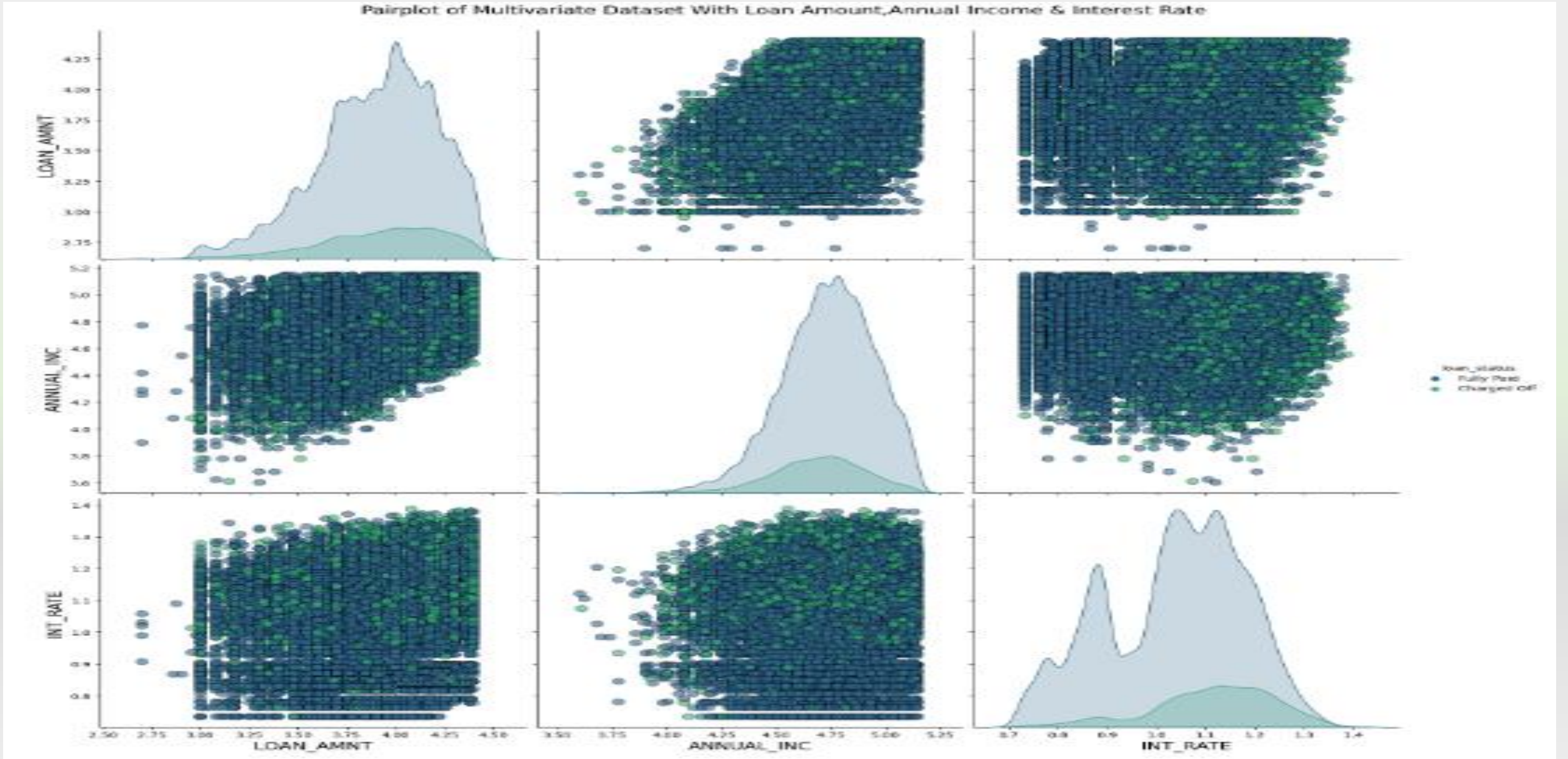
Purpose Vs No of Loans



Purpose Vs DTI

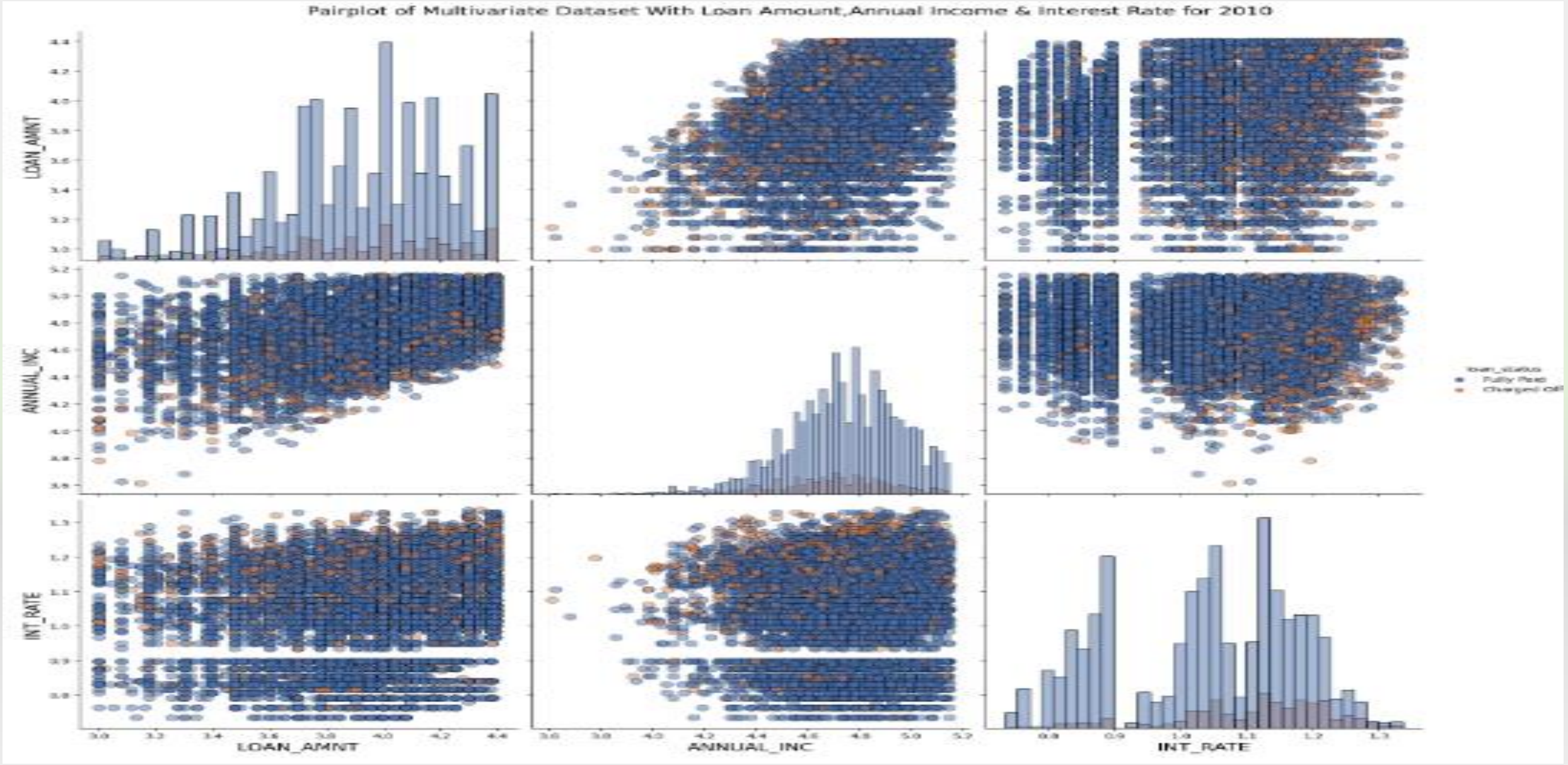
MULTIVARIATE ANALYSIS

Pair plot of Multivariate Dataset with Loan Amount, Annual Income and Interest Rate



MULTIVARIATE ANALYSIS

Pair plot of Multivariate Dataset with Loan Amount, Annual Income and Interest Rate for 2010



Observations

Multivariate Analysis -2010 Pair plot of Multivariate Dataset With Loan Amount, Annual Income & Interest Rate

- A higher charged-off ratio indicates a higher interest rate.
- As annual income increases, the loan amount increases slightly.
- The interest rate increases with the loan amount, and as it does, there is a significant charge off.

Key Findings:

Amount Categories Distribution

•All four amount categories exhibit similar distribution shapes.
•Most loan amounts range from 5,000to15,000.

Annual Income

•Annual incomes mostly range from 40,000to80,000.
•Higher annual incomes correlate with:
- Larger loans
- Higher revolving balances
- Higher total payments

Interest Rates

•Interest rates primarily fall between 9-14%.
•Higher interest rates (20%-25%) result in more charge-offs.
•Interest rates decrease with longer employee tenure.
•Higher delinquencies and recent credit inquiries are linked to higher interest rates.

Debt Consolidation

•Debt consolidation is the most common loan purpose, with many charged-off loans linked to it.
•Loans for debt consolidation are the most common.

Loan Amounts and Correlations

•Loan amount, funding amount, and funded amount invested are highly correlated.
•Larger loans lead to higher total payments.
•Loans between 20,000 and 25,000 have higher charge-off rates.
•Higher loan amounts and interest rates are linked to higher charge-offs.
•Longer loan terms (60 months) have higher loan amounts.
•More inquiries are associated with larger loan amounts.

Debt-to-Income Ratio (DTI)

•High debt-to-income (DTI) ratios are linked to high revolving balances.
•Higher DTI loans are more likely to be charged off.
•Higher DTI loans have more defaulters, especially in educational and small business loans.

Loan Purpose and Defaults

•Small business loans have the highest charge-off rates.
•Wedding loans have the least charge-offs.
•Small business loans have the highest number of defaulters, followed by educational loans.
•House loan defaulters are higher in rented houses.
•Small business loan defaulters are higher among those who own homes.
•Education and housing loans are fewer.

Loan Grades and Default Risks

•Loans with Grade A and Sub Grade A have the lowest charge-off risks.
•Grades/Sub Grades F and G have the highest charge-off risks.

Employment and Charge-Offs

•Unemployed individuals or those with less than one year of experience are more likely to be charged off.

Geographic and Bankruptcy Insights

•States like New York, California, and Florida have high charge-off rates.
•Those with public record bankruptcies have higher charge-off percentages.

Interest Rate and Loan Types

- Small Business and Debt Consolidation loans charge higher interest rates.
- Car loans charge less interest.
- Higher charged-off ratios indicate higher interest rates.

Historical and Income Group Insights

- Defaulters are more common in years like 2007 and 2009.
- Small business loan defaulters span all annual income groups.
- Higher-income groups see more house and education loan defaulters.
- Higher loan amounts see more defaulters across categories.
- All loan defaulters are fewer when interest rates are low and higher when rates are high.

Charge-Off Trends

- As annual income increases, loan amounts and interest rates also increase, leading to more charge-offs.