Lending Club Case Study

By Manikanta Vedula

Problem Statement

A consumer finance company which specialises 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 data given below contains the information about past loan applicants and whether they 'defaulted' or not. 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.

Data Cleaning and Derived Metrics

Data Cleaning

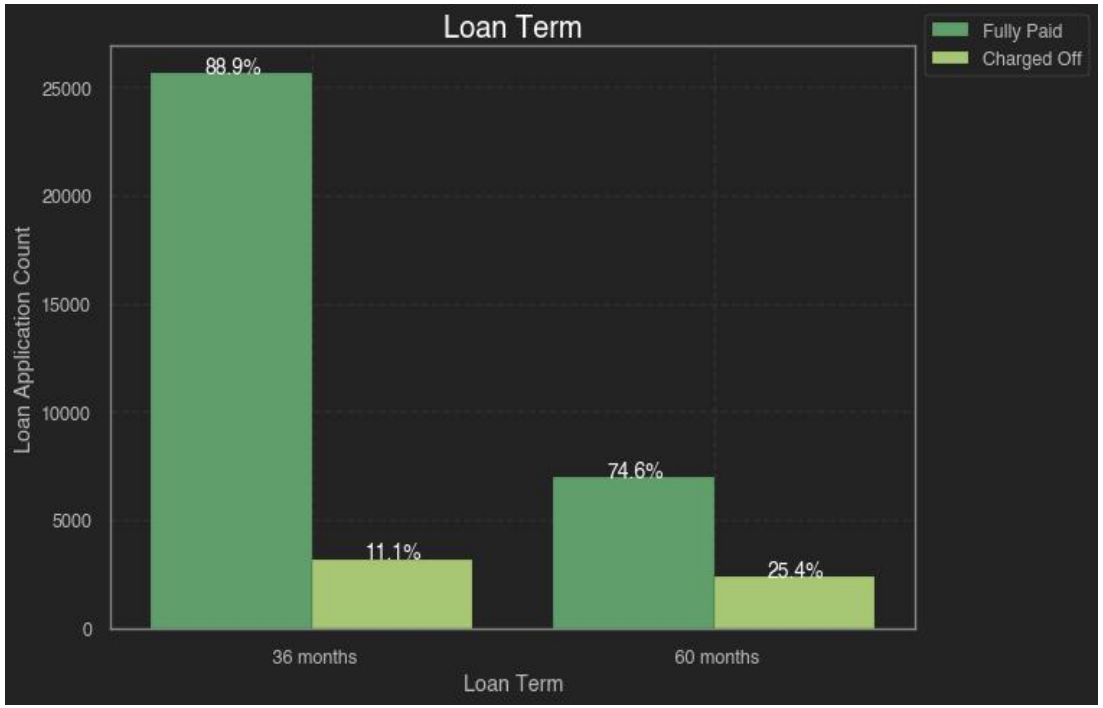
1. Dropped columns with highest Null Values and single Unique Values.
2. Dropped unnecessary columns that are not required for this objective.
3. Fixed Missing and Null Values and Standardized DataTypes and Fixed Invalid Values.
4. Removed Outliers with 0.99 quantile.
5. Filtered data and removed all records with Loan Status as 'Current'.

Derived Metrics

1. Created new columns based on Type Driven, Business Driven and Data Driven Metrics.
2. Created Categorical Variables as Buckets from Continuous Variables - Loan Amount, Interest Rate, Annaul Income, Debt to Income Ratio, Installment, Revolving Balance and Revolving Utilization Columns.

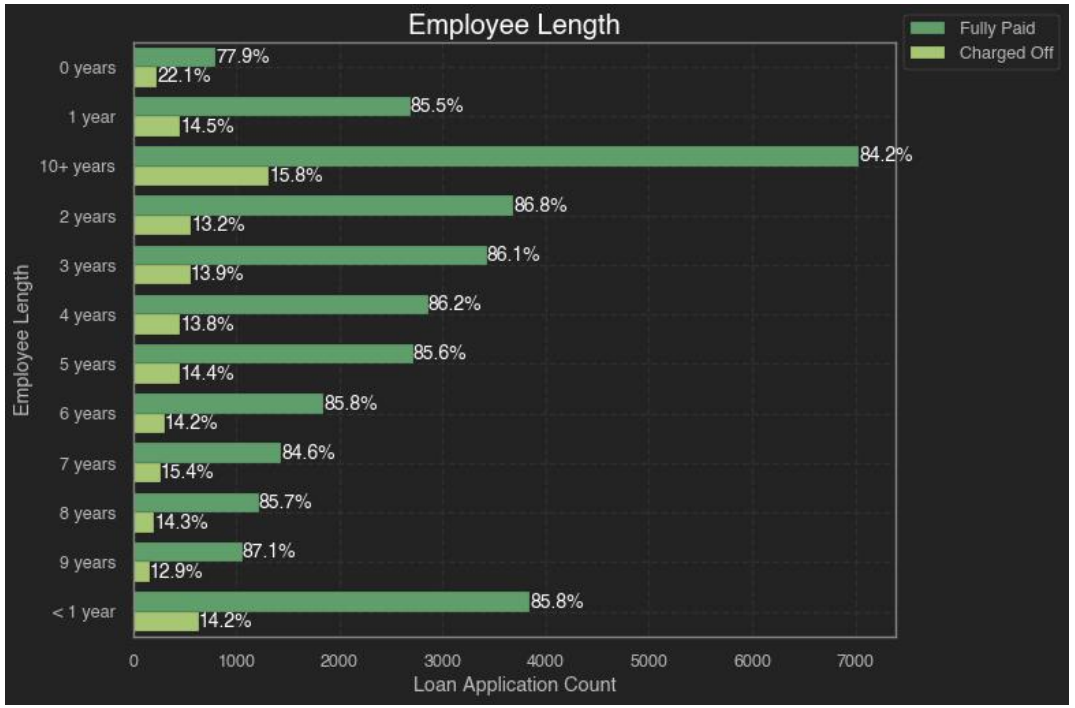
Univariate and Segmented Univariate Analysis

Loan Term:



1. 3 years Loan Term has 11.1% Defaulters and 5 years Loan Term has 25.4% Defaulters.
2. Hence, Higher the Loan Term, Higher the Default Percentage.

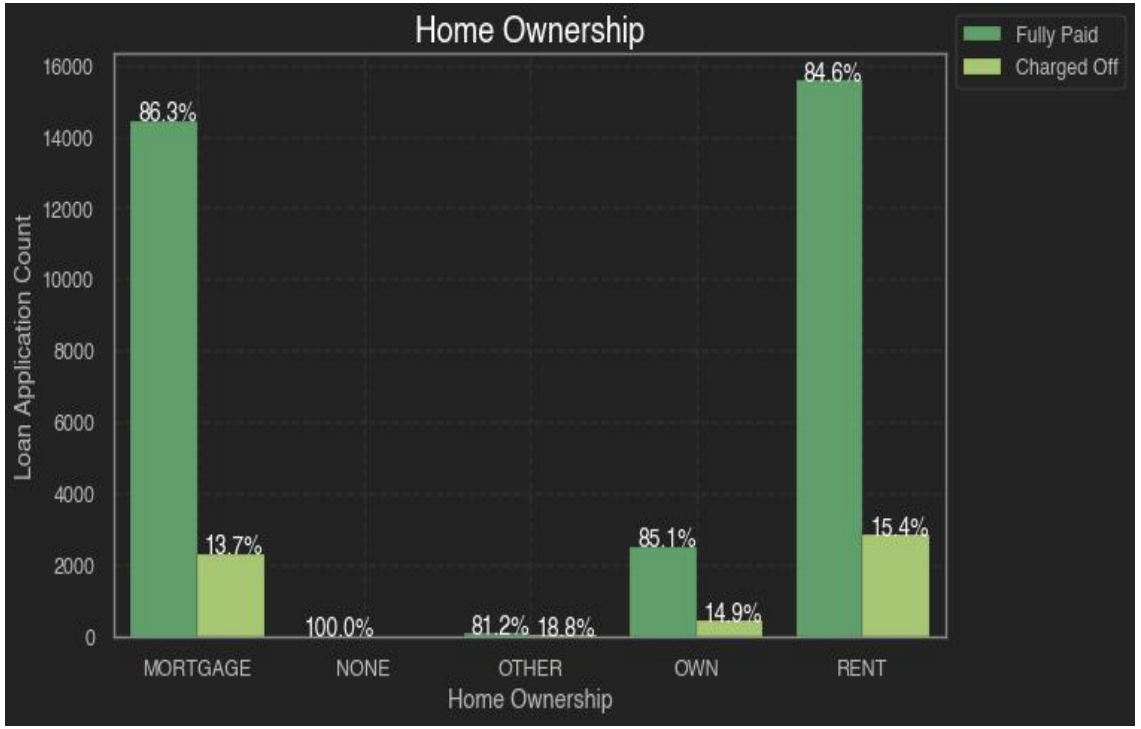
Employment Length:



1. Borrowers with No Experience and Not mentioned any Experience have the Highest Default Percentage (22.1%).
2. Hence, Lower the Employment Experience, Higher the Default Percentage.

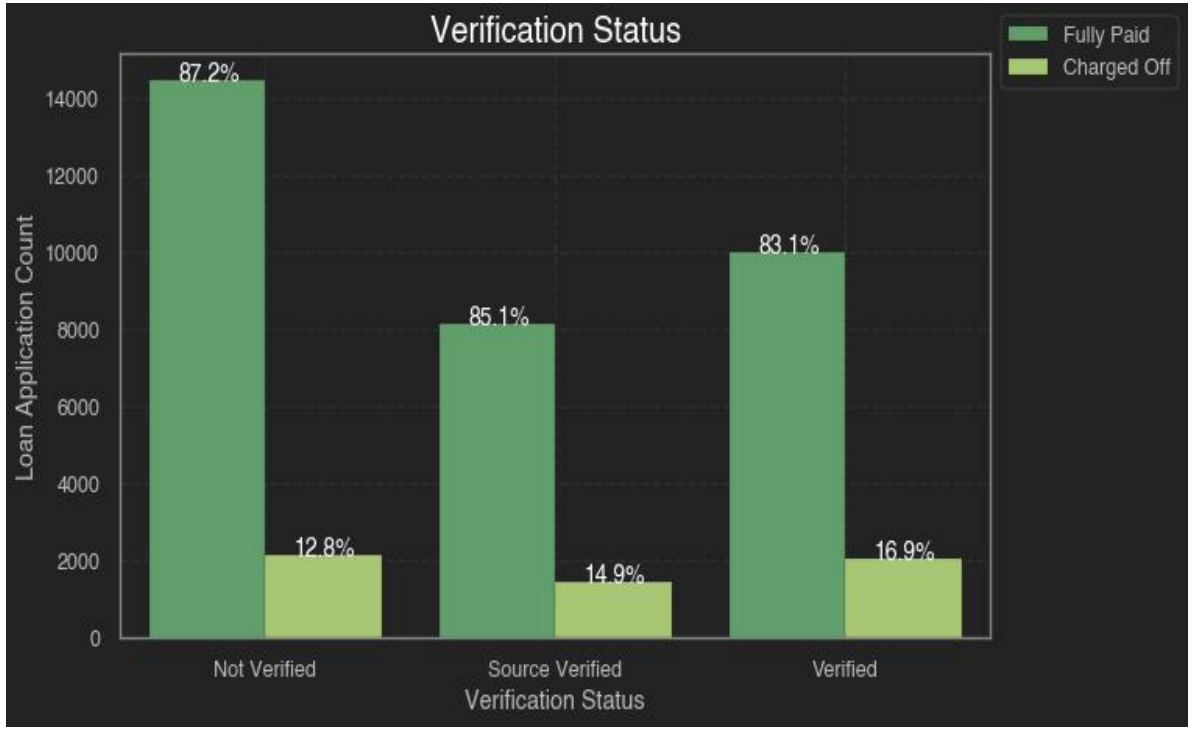
Univariate and Segmented Univariate Analysis

Home Ownership:



1. Borrowers with Rent and Mortage Home Ownership have Default Percentages as 15.4% and 13.7% respectively.
2. Hence, Home Ownership also plays crucial role for the Loan Default.

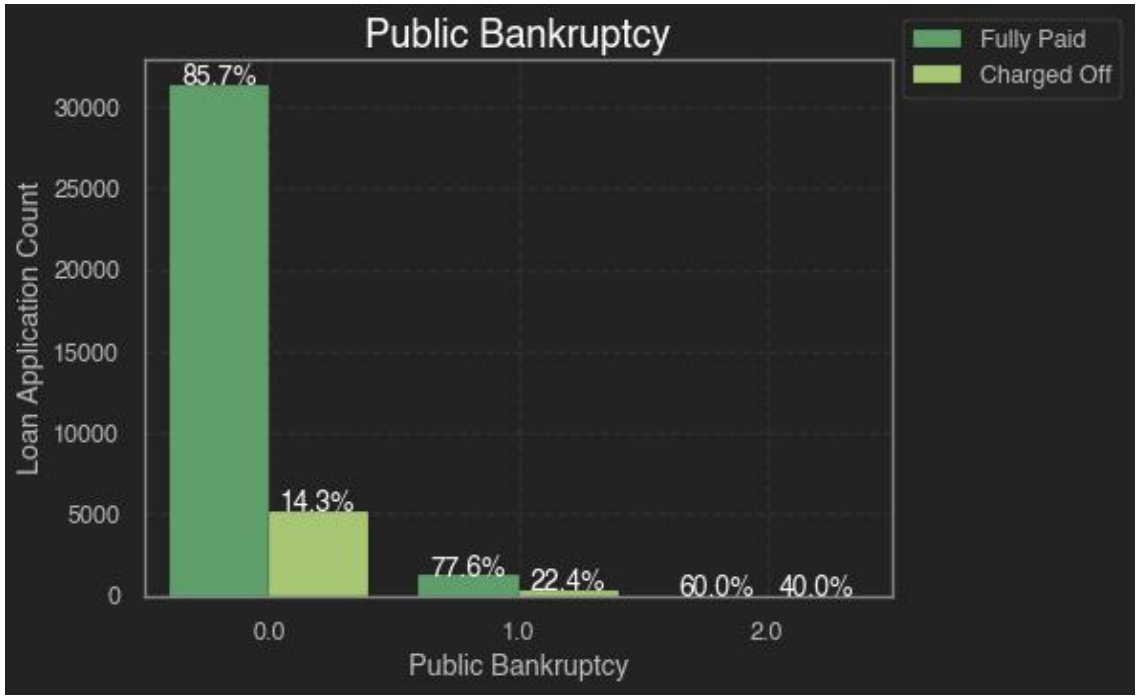
Verification Status:



1. Verified Status and Source Verified have the Highest Default Percentage. It means Income should be Verified correctly by Lending Club.
2. Verification Status plays an important role in Loan Default.

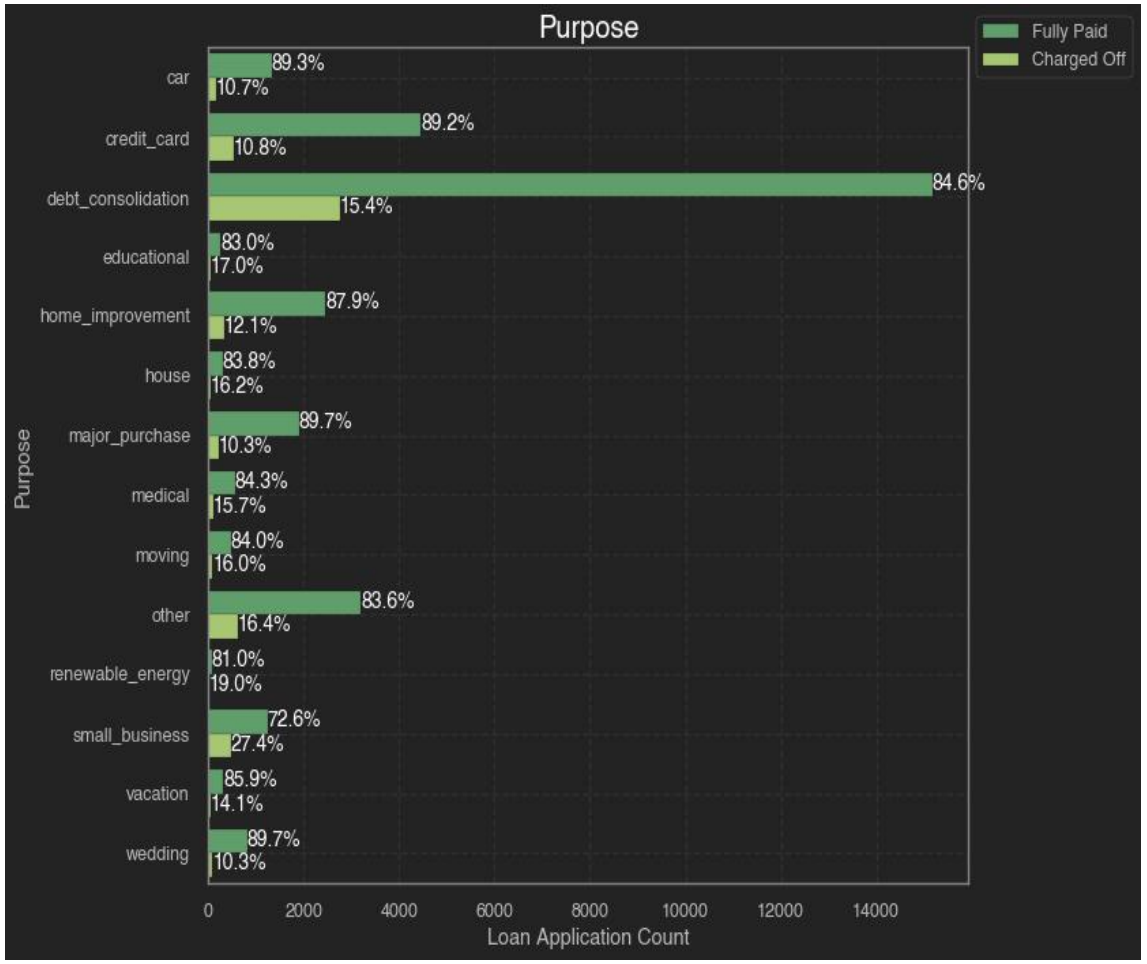
Univariate and Segmented Univariate Analysis

Public Bankruptcy:



1. Charged Off Percentage is increasing with having Public Bankruptcy Records.
2. Hence, Lower the Public Bankruptcies, Lower the Default Percentage.

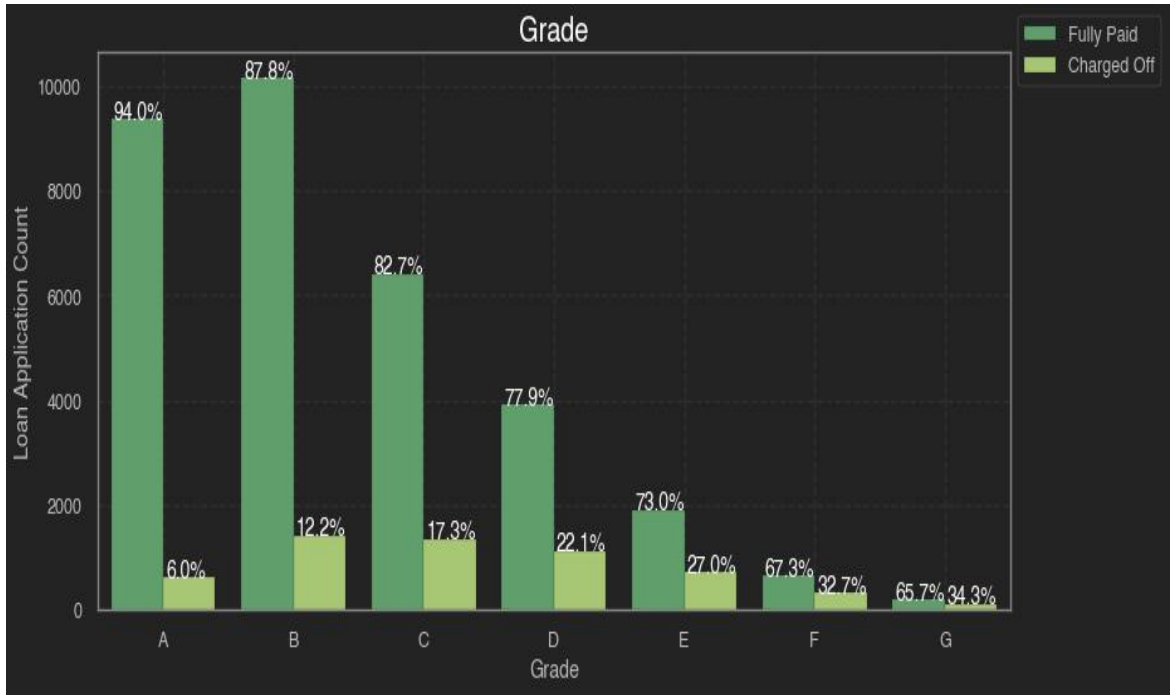
Purpose:



1. Small Business, Renewable Energy, Other Purpose, Educational, House and Debt Consolidation have the highest Charged Off Percentage.

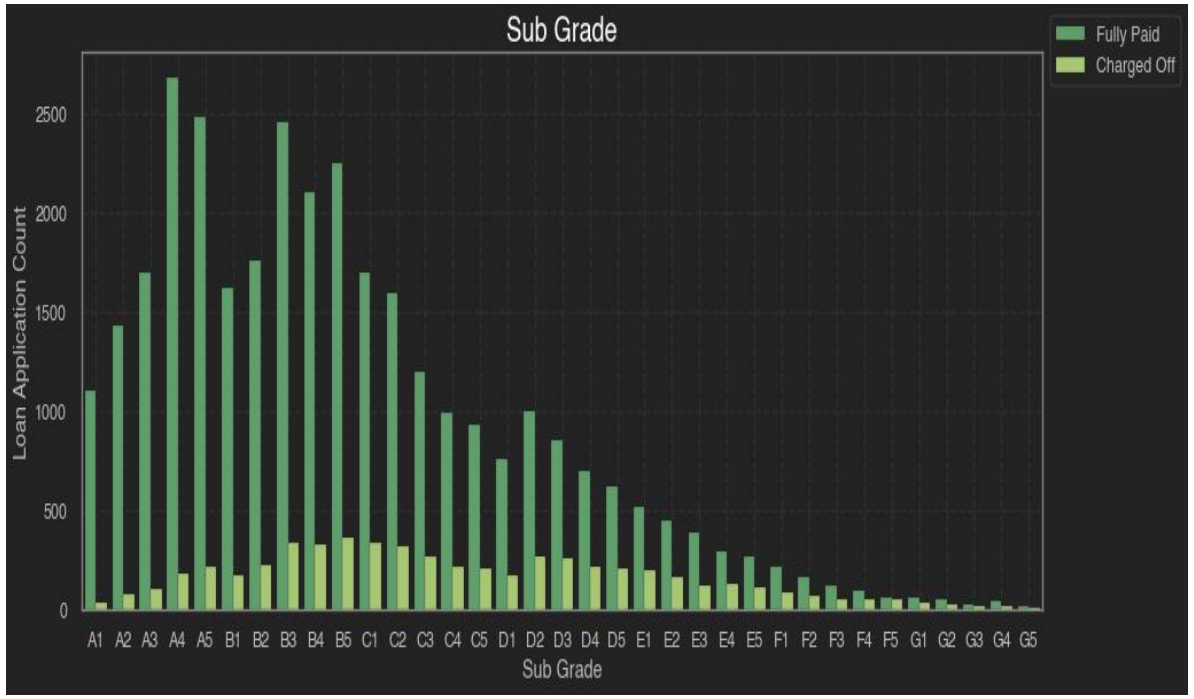
Univariate and Segmented Univariate Analysis

Grade:



1. With the decrease in Grade Rank, there is an increase in Charged Off Percentage.
2. Hence, Higher the Loan Grade Rank, Lower the Default Percentage.

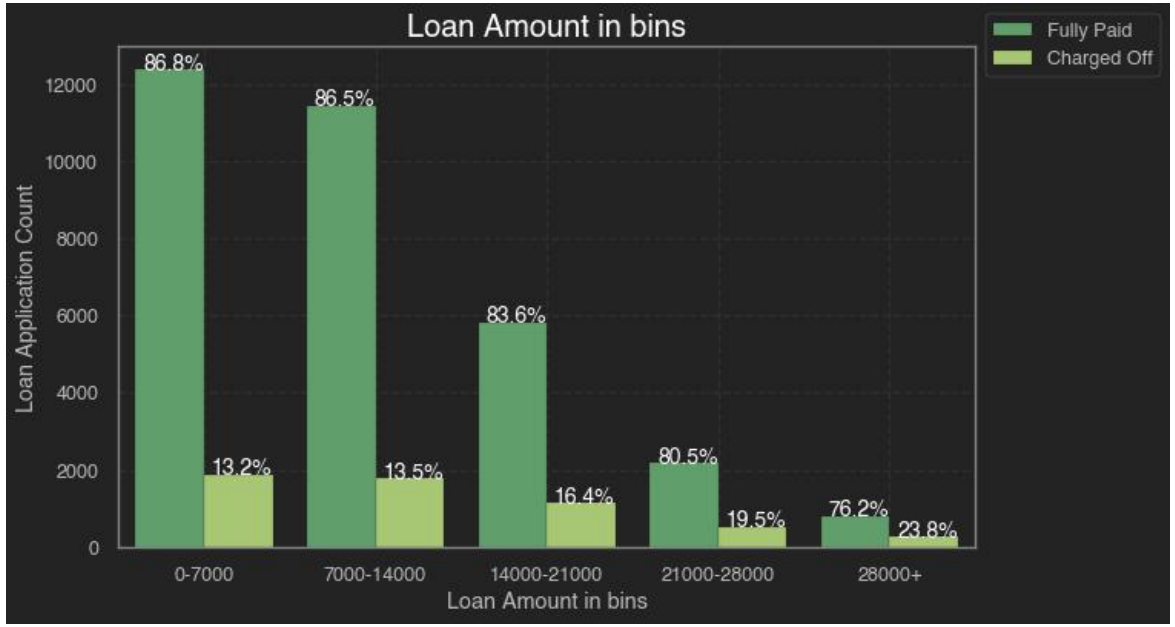
Sub-Grade:



1. With the decrease in Sub Grade Rank, there is an increase in Charged Off Percentage.
2. Hence, Higher the Loan Sub Grade, Lower the Default Percentage.

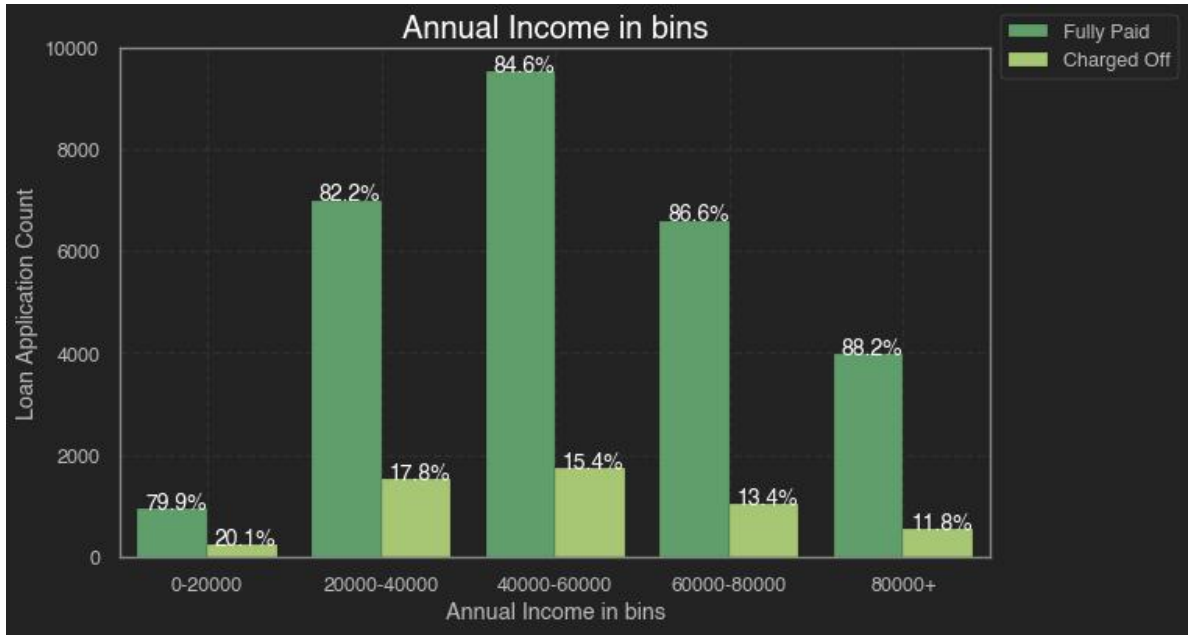
Univariate and Segmented Univariate Analysis

Loan Amount in Bins:



1. Charged Off Percentage is increasing with the Loan Amount.
2. Hence, Lower the Loan Amount, Lower the Default Percentage.

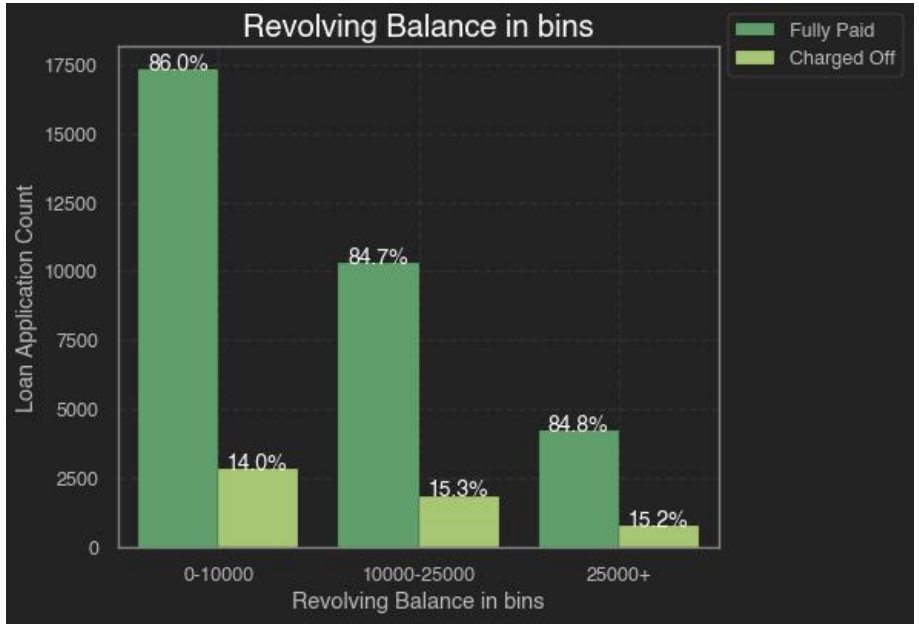
Annual Income in Bins:



1. Charged Off Percentage is increasing with the decrease in Annual Income.
2. Hence, Higher the Annual Income, Lower the Default Percentage.

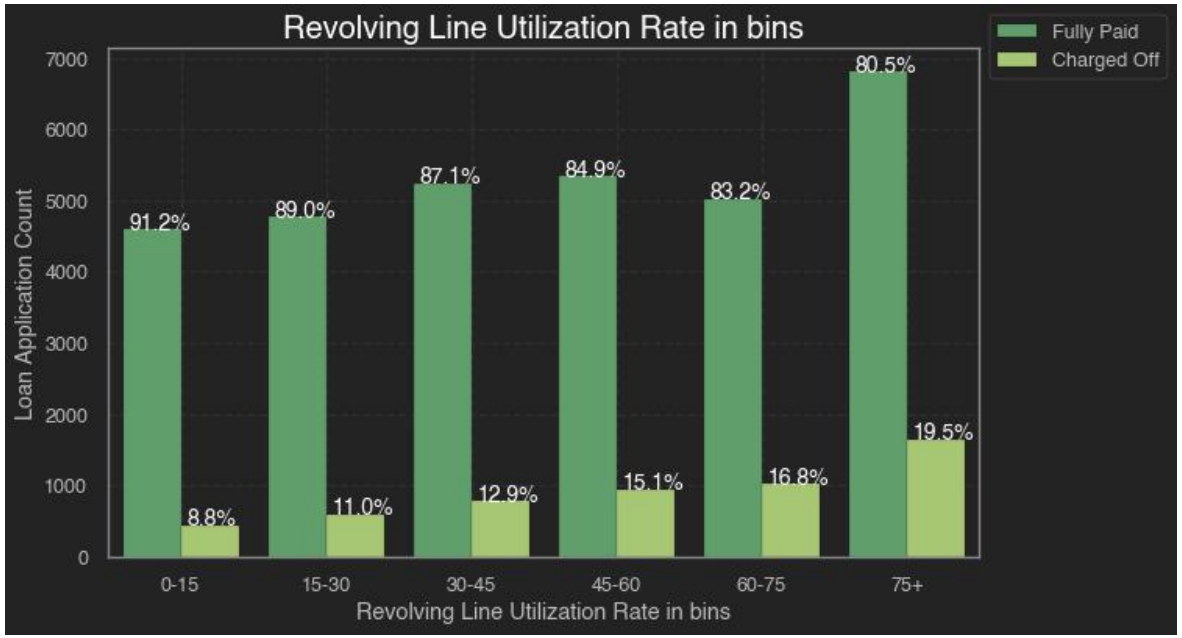
Univariate and Segmented Univariate Analysis

Revolving Balance in Bins:



1. Charged Off Percentage is increasing with the increase in Revolving Balance.
2. Hence, Higher the Revolving Balance, Higher the Default Percentage.

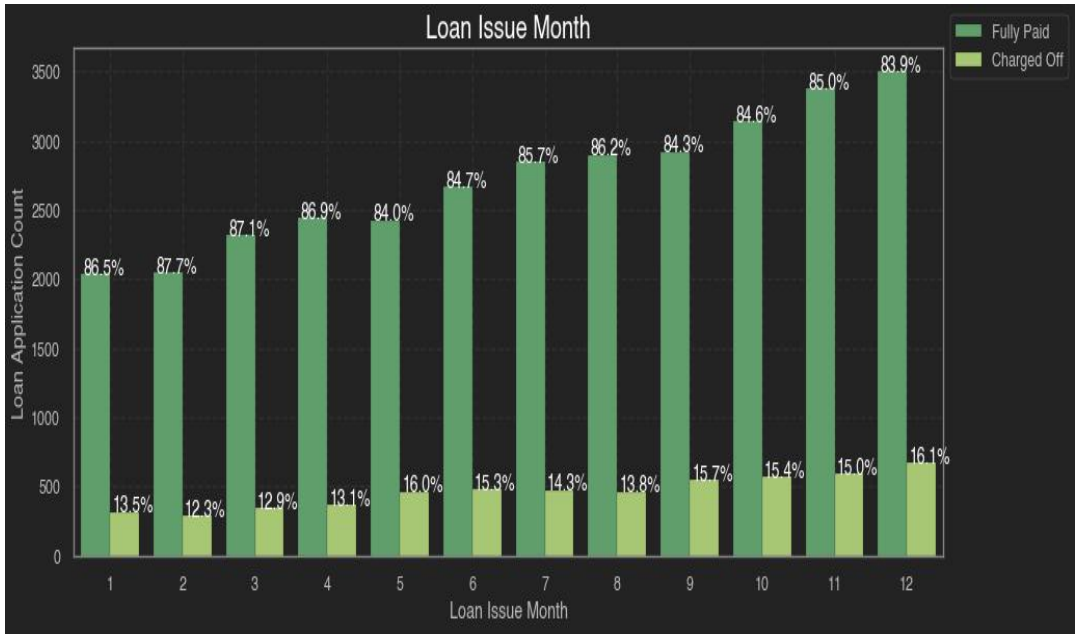
Revolving Line Utilization Rate:



1. Charged Off Percentage is increasing with the increase in Revolving Line Utilization Rate.
2. Hence, Higher the Revolving Line Utilization Rate, Higher the Default Percentage.

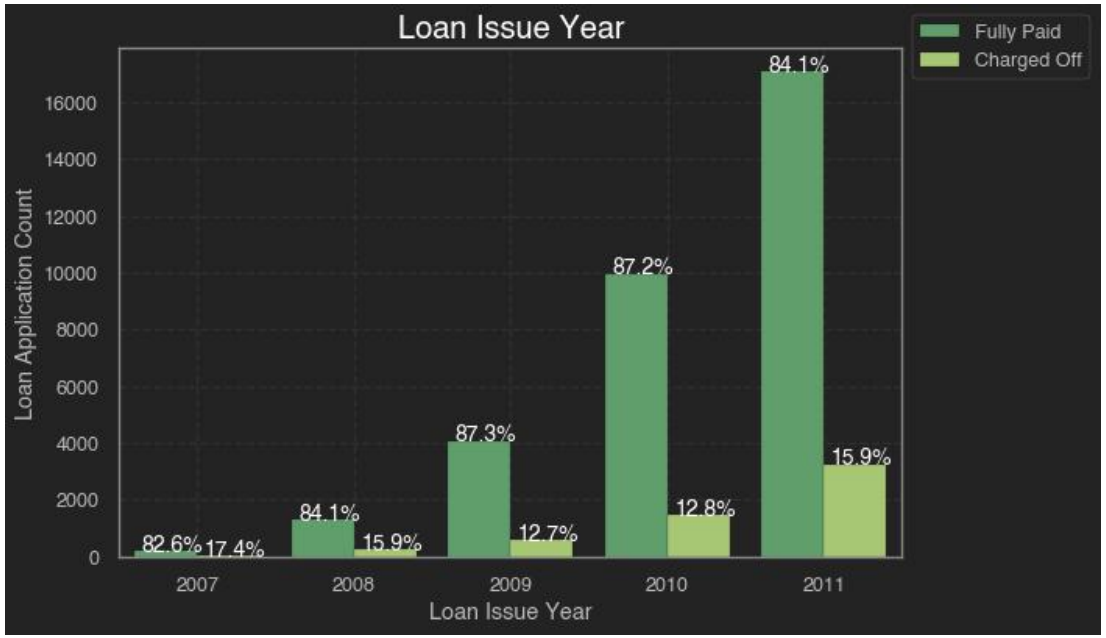
Univariate and Segmented Univariate Analysis

Loan Issue Month:



1. Loan Applications are increasing with Month
2. May and December Months have the Highest Loan Defaulters.

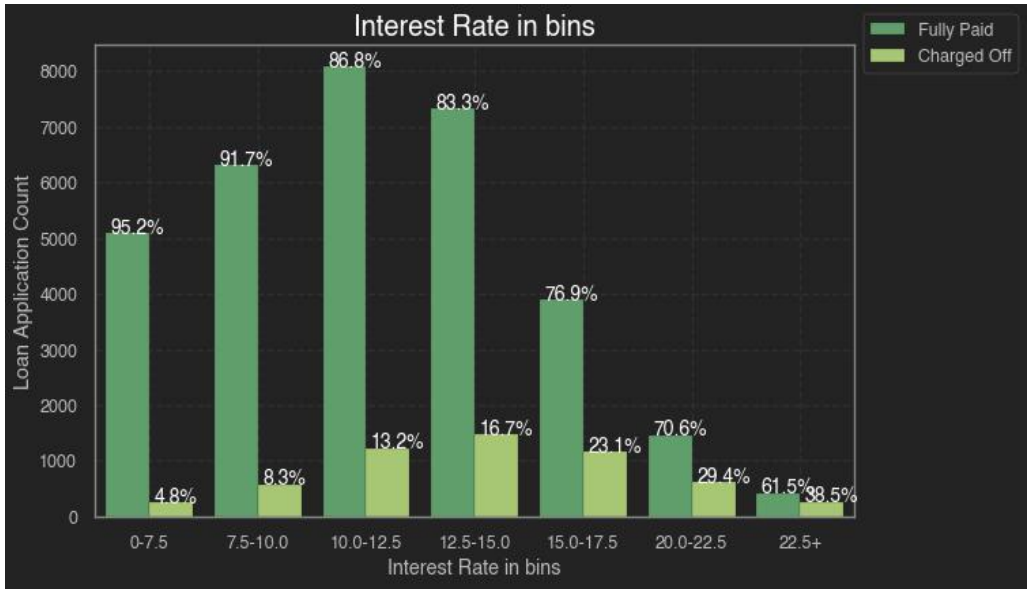
Loan Issue Year:



1. Loan Applications are increasing with Year

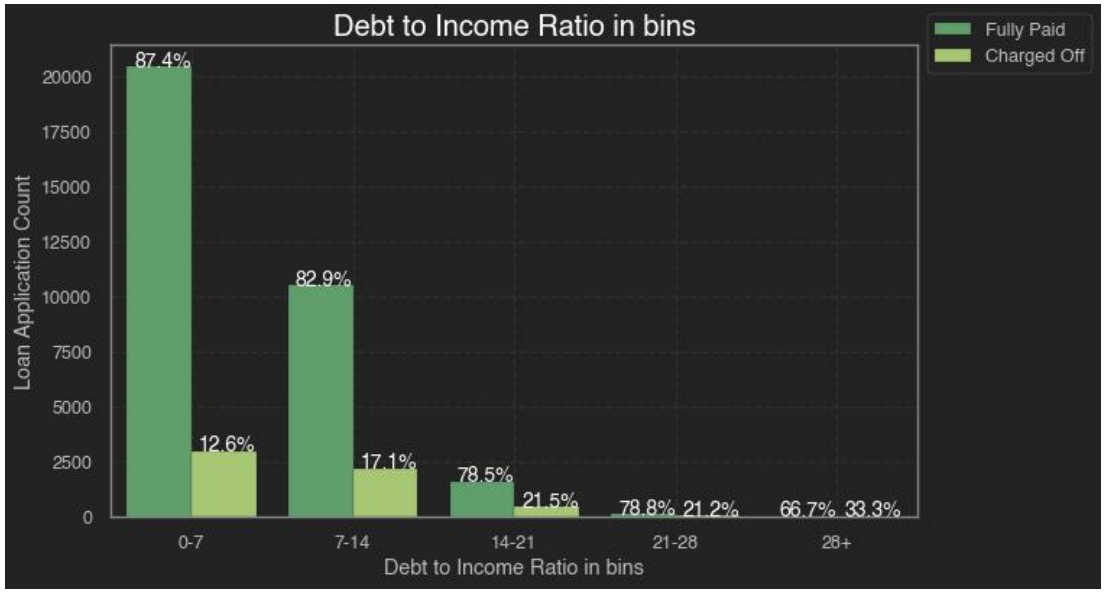
Univariate and Segmented Univariate Analysis

Interest Rate in Bins:



1. Charged Off Percentage is increasing with Interest Rate.
2. Hence, Lower the Interest Rate, Lower the Default Percentage.

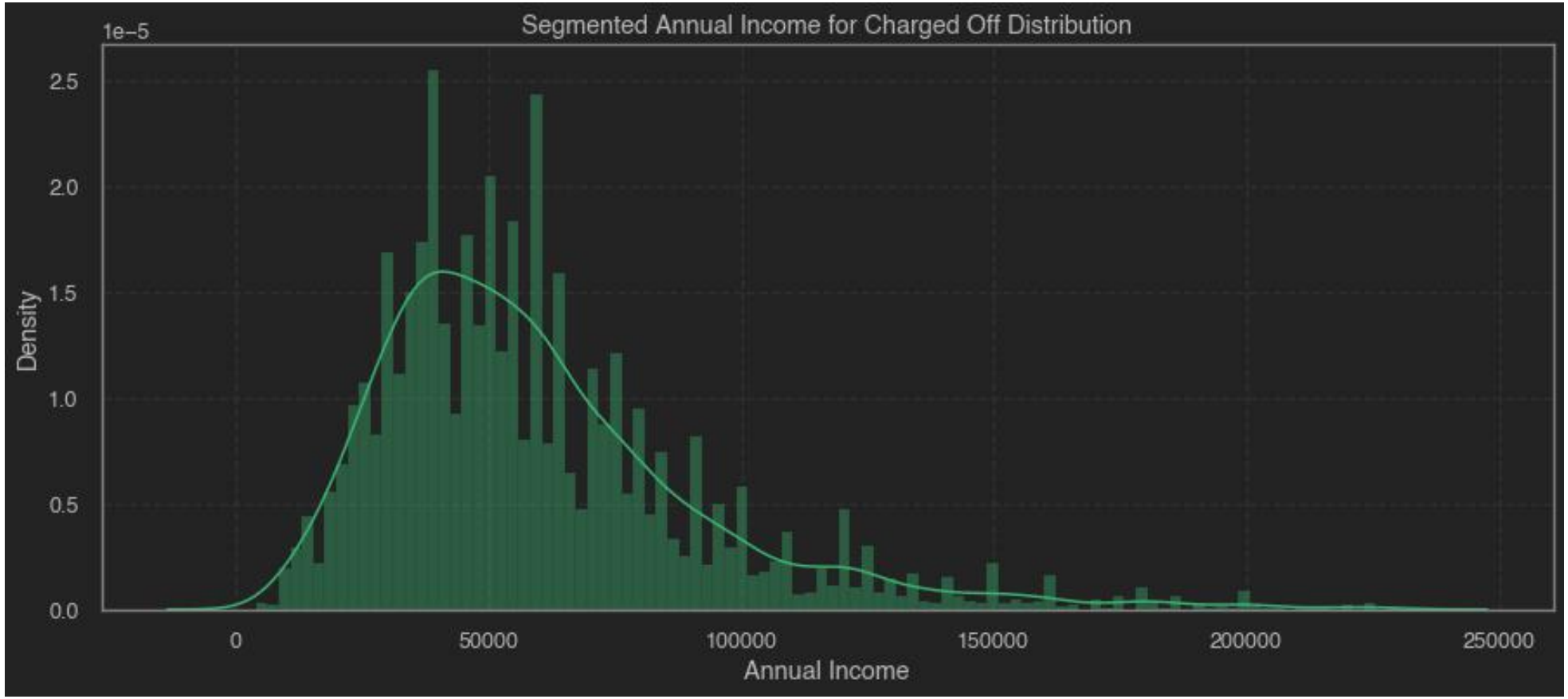
Debt to Income Ratio in Bins:



1. Charged Off Percentage is increasing with the Debt to Income Ratio.
2. Hence, Lower the Debt to Income Ratio, Lower the Default Percentage.

Univariate and Segmented Univariate Analysis

Segmented Annual Income for Loan Status as Charged Off:

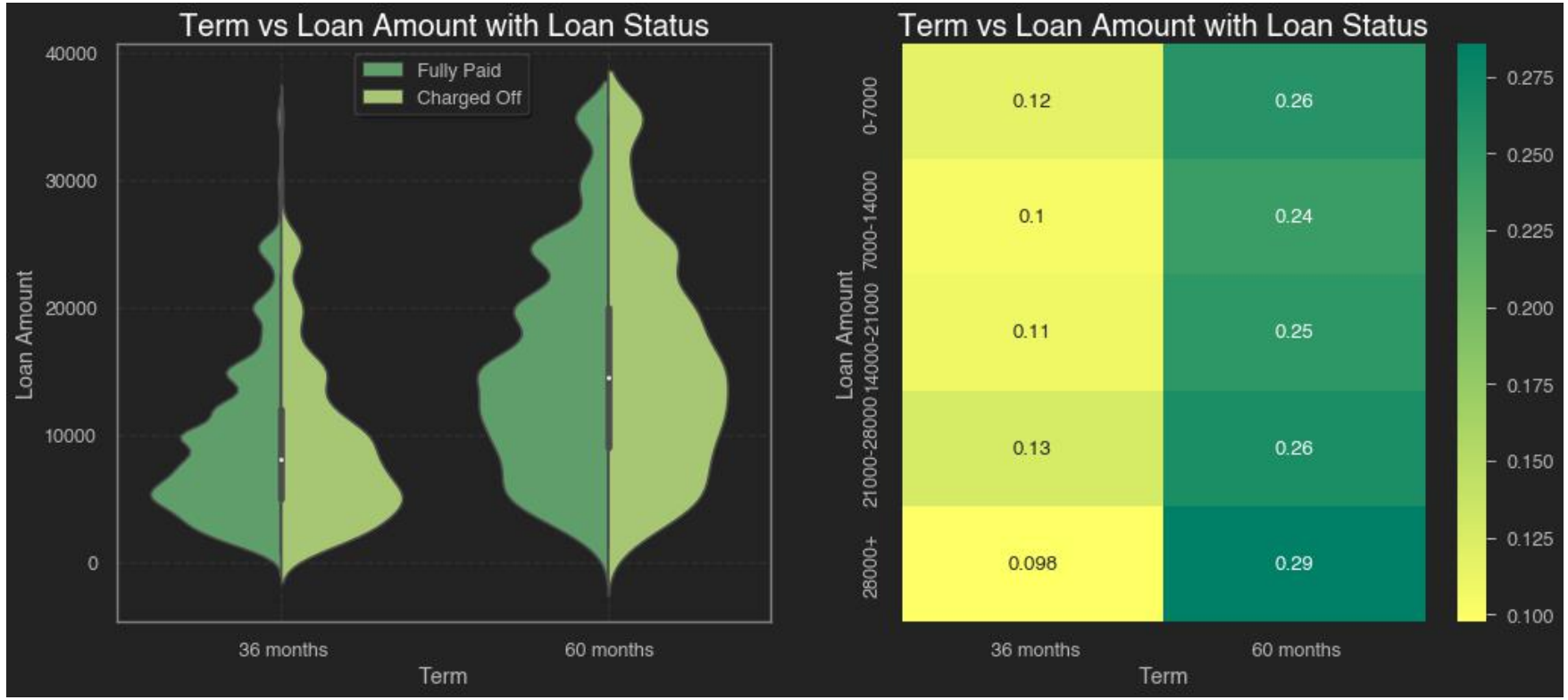


Observation: Segemnted Data for Loan Status 'Charged Off'

1. Annual Income for most of the Defaulters is low.
2. Hence, Lower the Annual Income, Higher the Default Percentage.

Bivariate Analysis

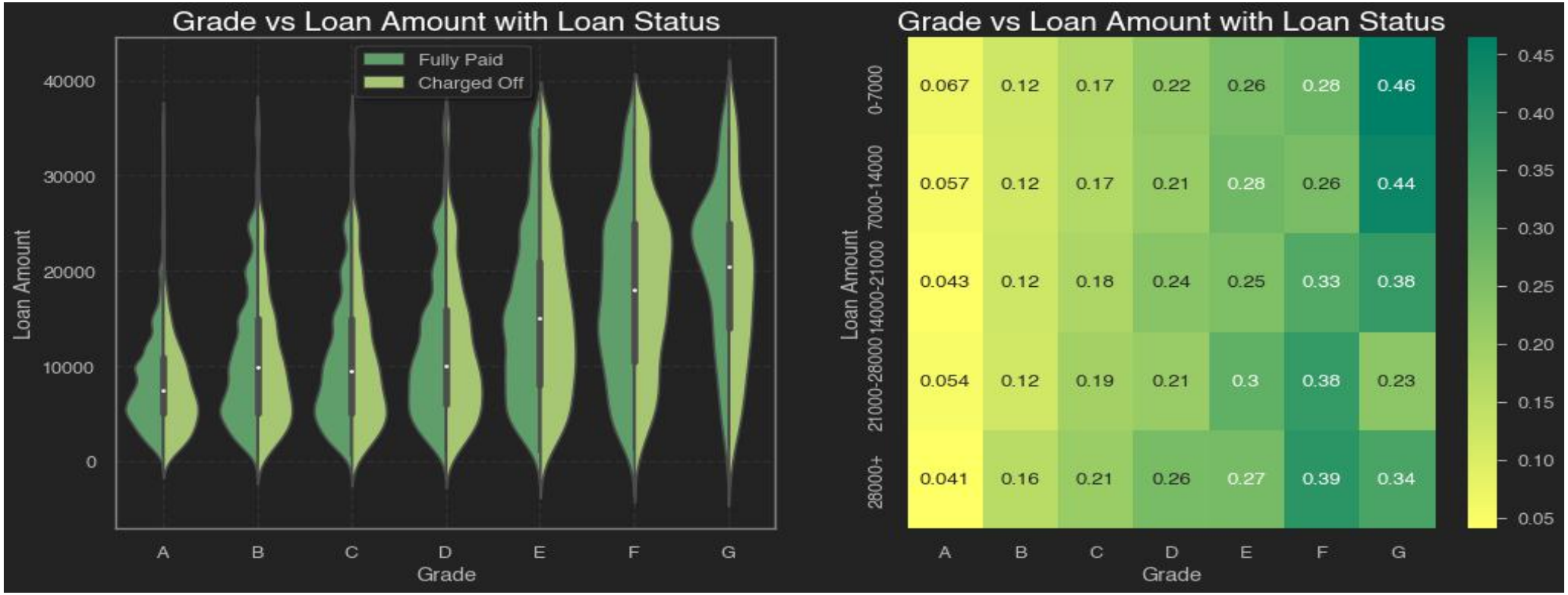
Term vs Loan Amount with Loan Status as Hue:



1. Loan Repayment Term is increasing with the increase in Loan Amount.
2. Lower the Loan Amount, Lower the Loan Repayment Term.
3. Hence, Lower the Default Percentage.

Bivariate Analysis

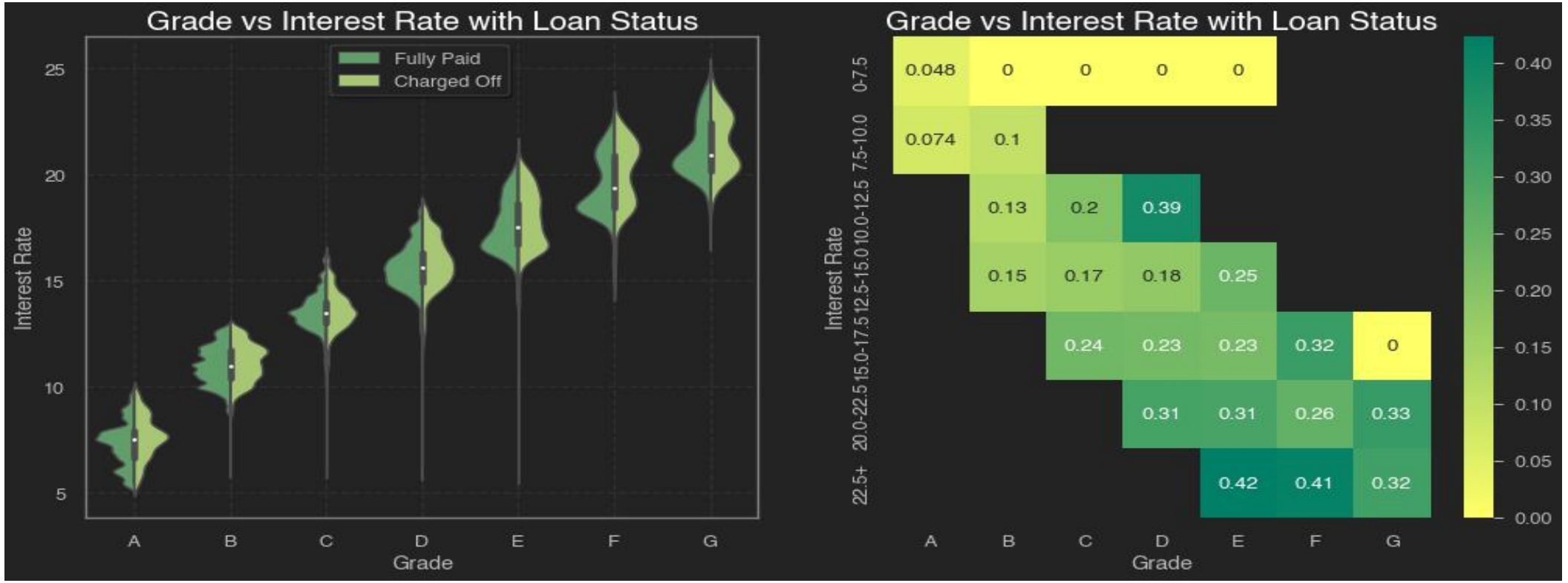
Grade vs Loan Amount with Loan Status as Hue:



1. Loan Amount is increasing with the increase in Grade Rank.
2. Higher the Grade Rank, Lower the Loan Amount.
3. Hence, Lower the Default Percentage.

Bivariate Analysis

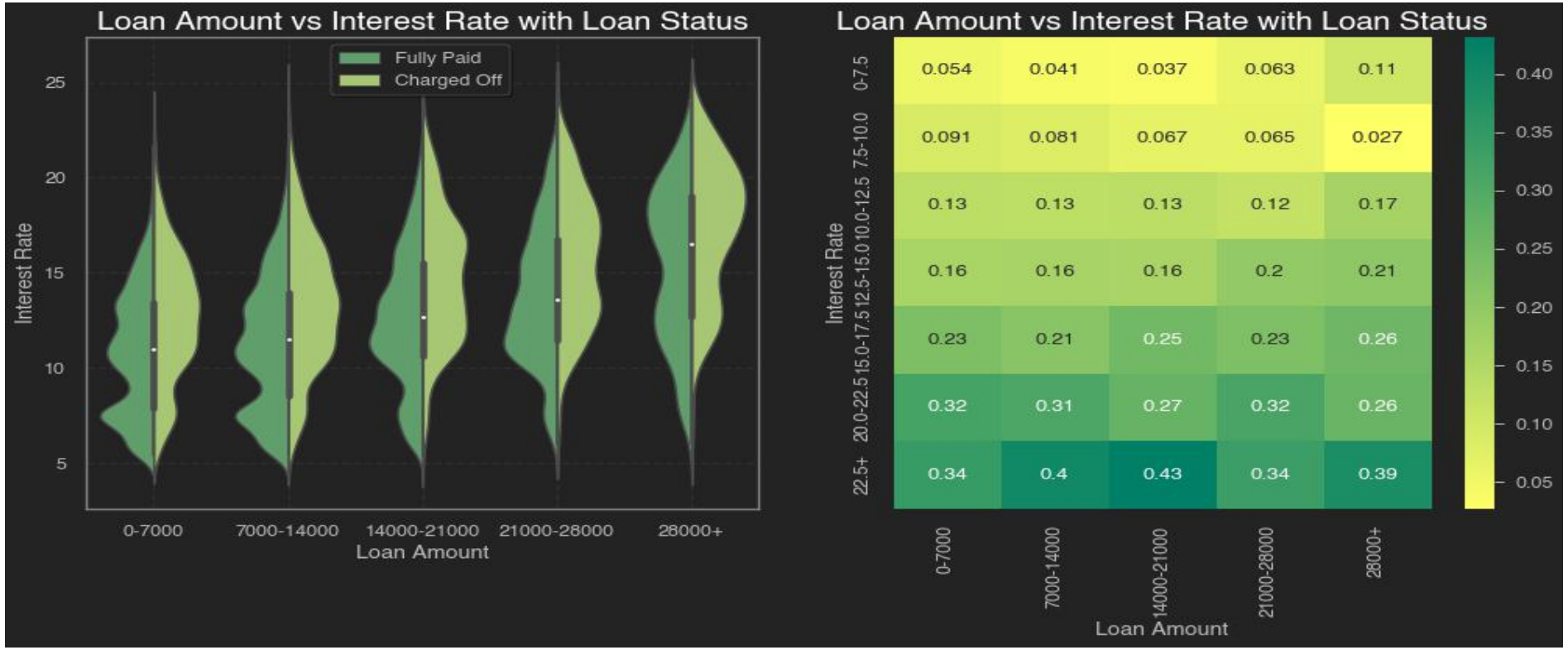
Grade vs Interest Rate with Loan Status as Hue:



1. Interest Rate is increasing with the decrease in Grade Rank.
2. Higher the Grade Rank, Lower the Interest Rate.
3. Hence, Lower the Default Percentage.

Bivariate Analysis

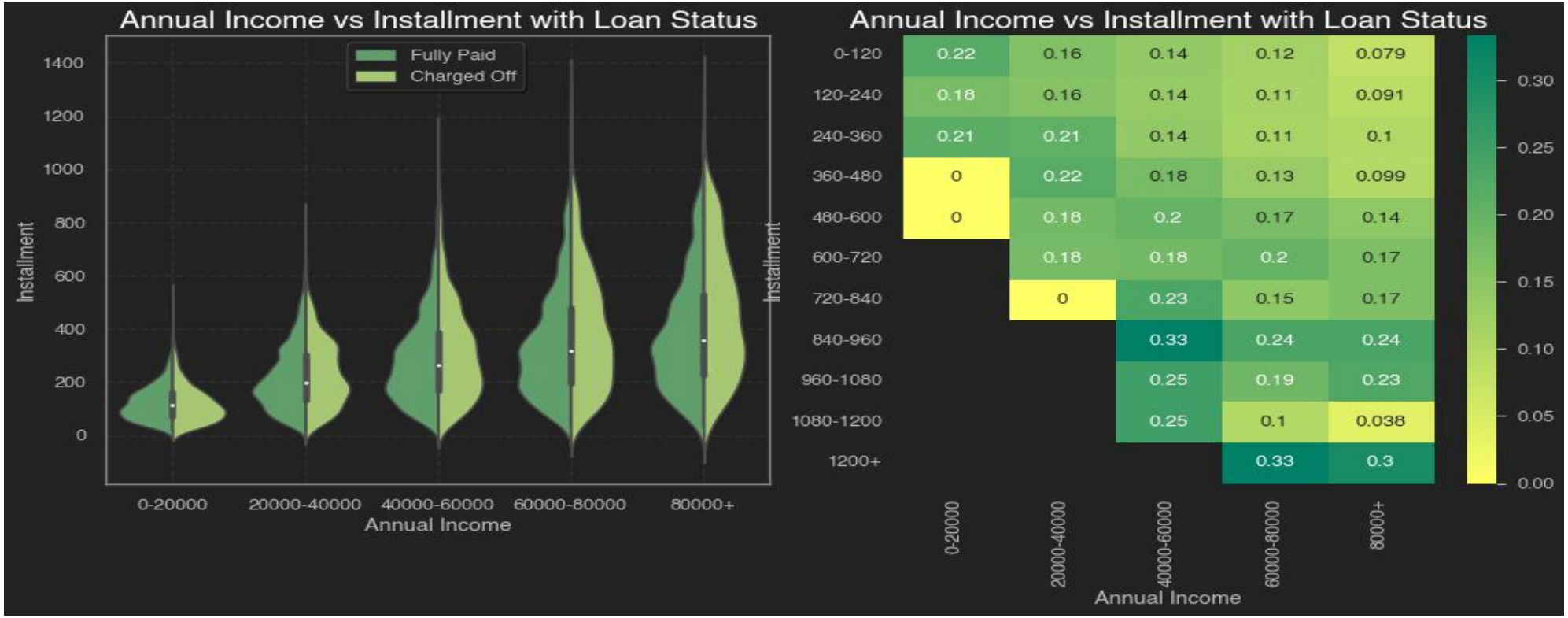
Loan Amount vs Interest Rate with Loan Status as Hue:



1. Interest Rate is increasing with the increase in Loan Amount.
2. Lower the Loan Amount, Lower the Interest Rate.
3. Hence, Lower the Default Percentage.

Bivariate Analysis

Annual Income vs Installment with Loan Status as Hue:



1. Lower Installments to Higher Annual Income.
2. Hence, Lower the Default Percentage.

Conclusion

1. Borrowers with Lower Annual Income is more likely to be defaulted.
2. Borrowers with Higher Debt to Income Ratio is more likely to be defaulted.
3. Loan Amount with longer Loan Repayment Term has more likely to be defaulted.
4. Loan Amount with Lower Grade Rank has more likely to be defaulted.
5. Lower Grade Rank with Higher Interest Rate has more likely to be defaulted.
6. Loan Amount with Higher Interest Rate has more likely to be defaulted.
7. Higher Installments with Lower Annual Income have more likely to be defaulted
8. No Employment Experience or Not mentioned any Experience is more likely to be defaulted.