Lending Club Case Study

Nasir

Paras Paul

Agenda

Problem Statement

Data Understanding

Data Cleaning

Data Analysis

Recommendations

Problem Statement

Lending Club company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Approach

Data Understanding	Data Cleaning	Univariate Analysis	Bivariate Analysis	Recommendations
Objective of this section is to understand data using data definitions, identify data quality issues and report them.	1. Fix rows and columns 2. Fix missing values 3. Standardise values 4. Fix invalid values 5. Filter data	The primary goal of univariate analysis is to describe and summarize the characteristics of a single variable.	It involves the simultaneous analysis of two variables to understand the relationships between them.	Based on the analysis we will formulate the recommendations.

Data Understanding

The data given 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.

The overall loan process is represented by three steps

- Potential borrower requests for loan amount (loan_amnt)
- The approver approves/rejects an amount based on past history/risk (funded_amnt)
- The final amount offered as loan by the investor (funded_amnt_inv)

DATA

CLEANING

1. Fix rows and columns
2. Fix missing values
3. Standardise values
4. Fix invalid values
5. Filter data

Data Cleaning

Fix Rows

- No summary, no header or footer rows in the dataset
- No column number, indicators etc. found in the dataset
- Rows where the loan status is current have been dropped as current loans are in progress and will not contribute in the decision making of pass or fail of the loan.

Fix Columns

- Columns with more than 75% NA values have been dropped.
- Columns with zero values have been dropped.
- Columns like employee title, url are descriptive and have been dropped.
- Unique columns like member id have been dropped.
- Customer behaviour columns have been dropped.

Data Conversions

- Convert term column to integer.
- Convert issue date to date time format.

Data Cleaning

Fix missing values

- Replaced zero with missing values in term column.
- Replaced zero with missing values in emp_length column.

Extract new columns

- Extracted issue year and issue month from issue month.
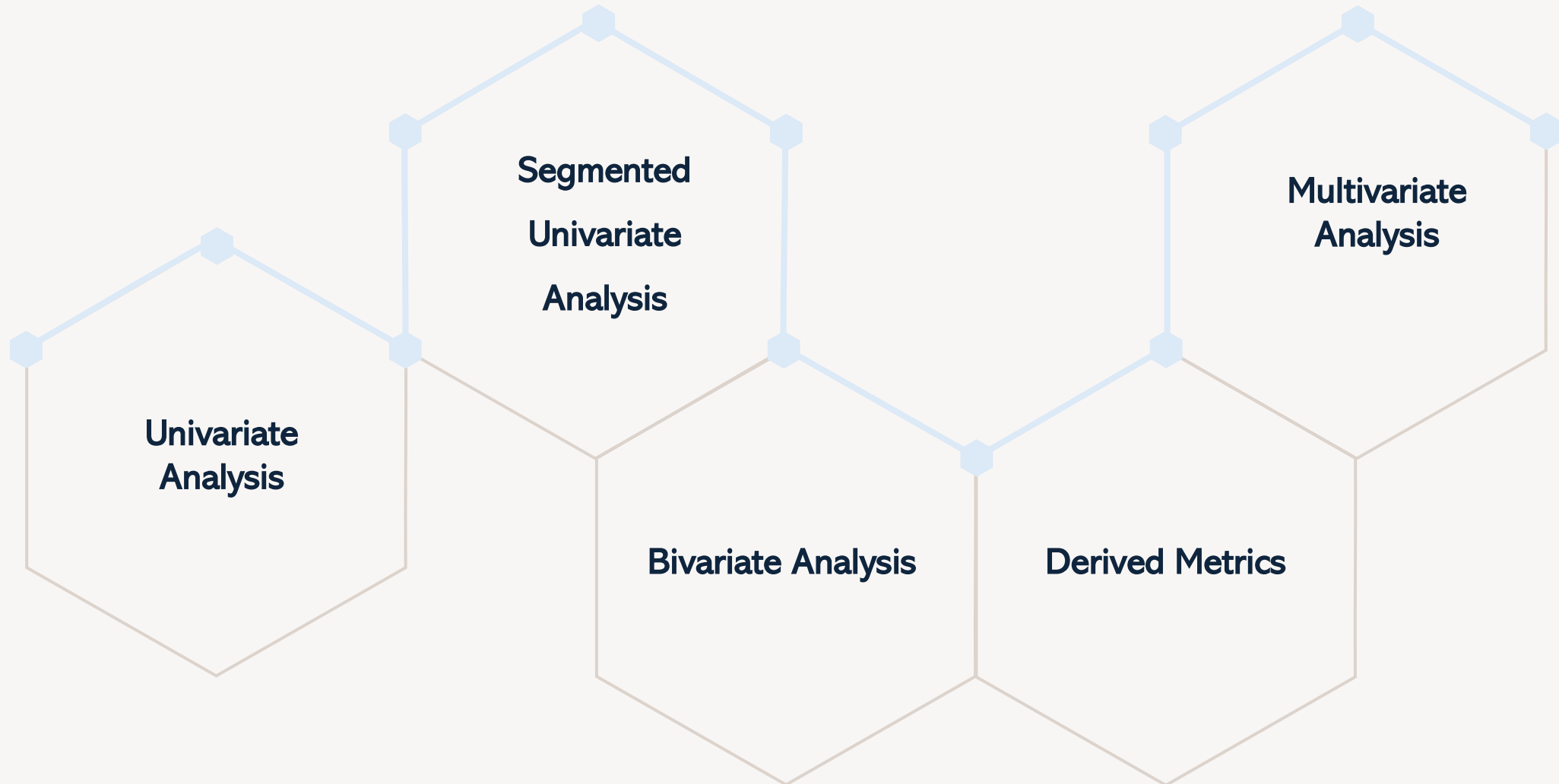
Finding Outliers
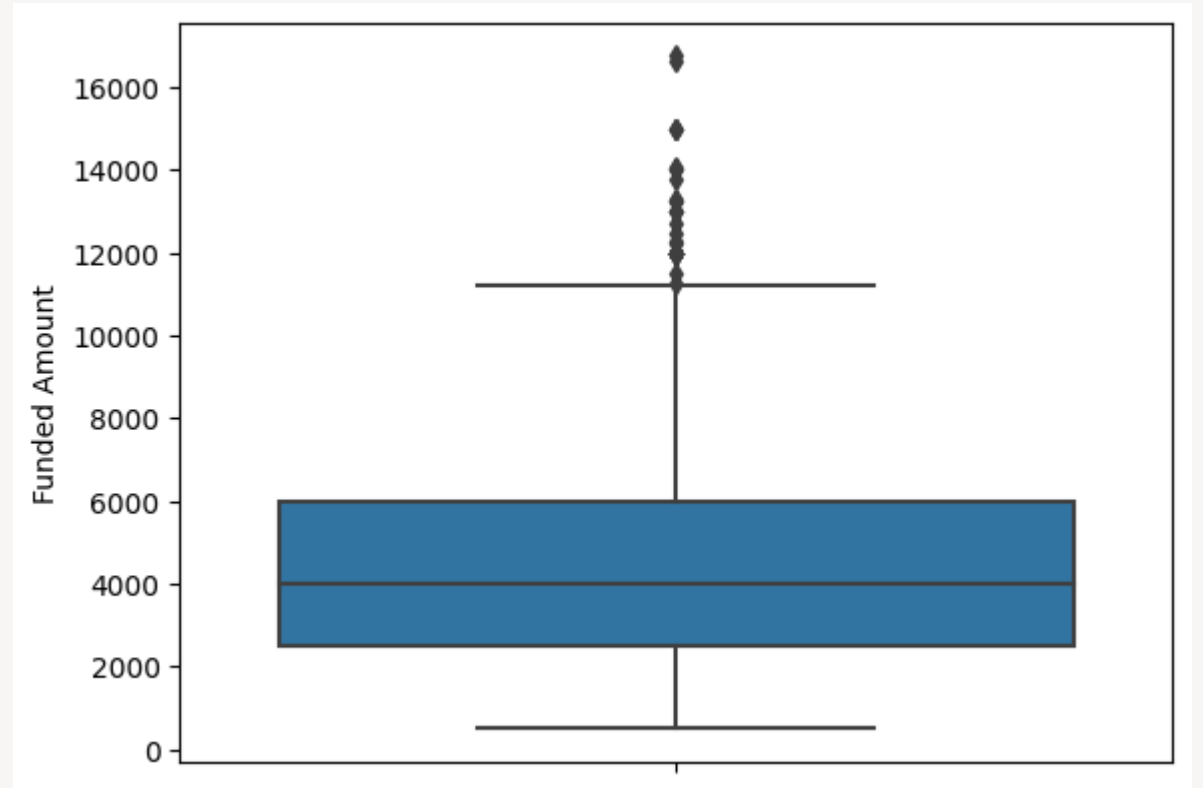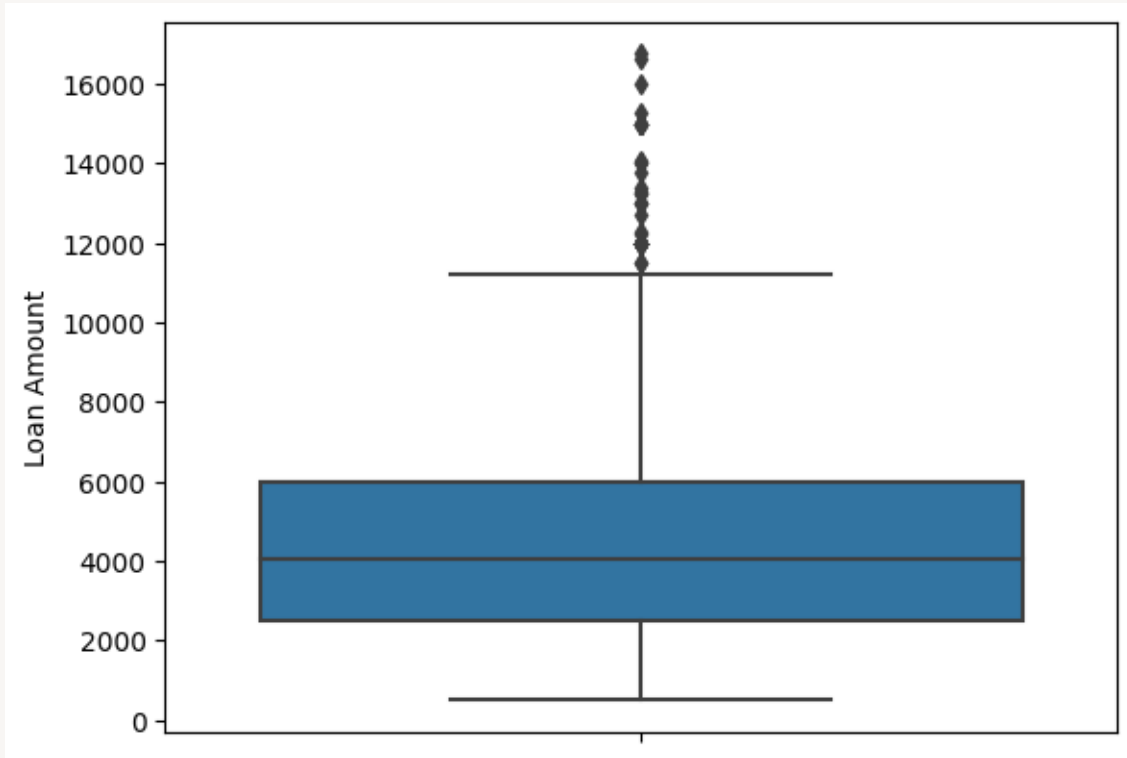
Finding Outliers
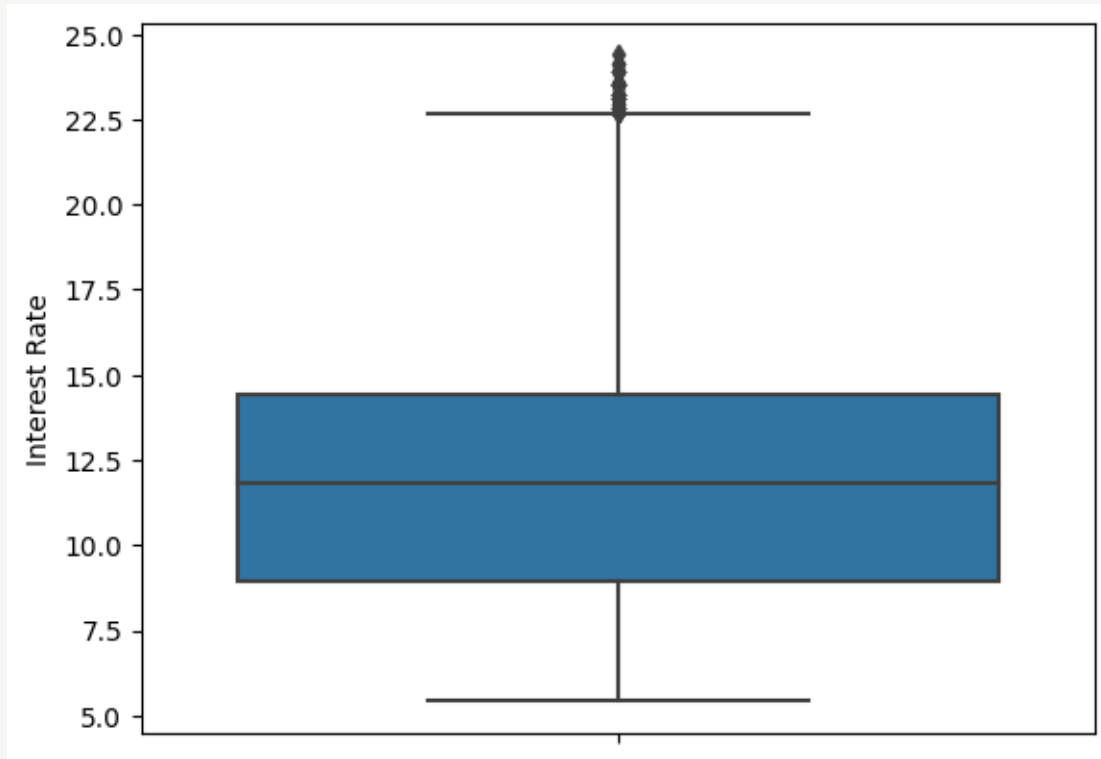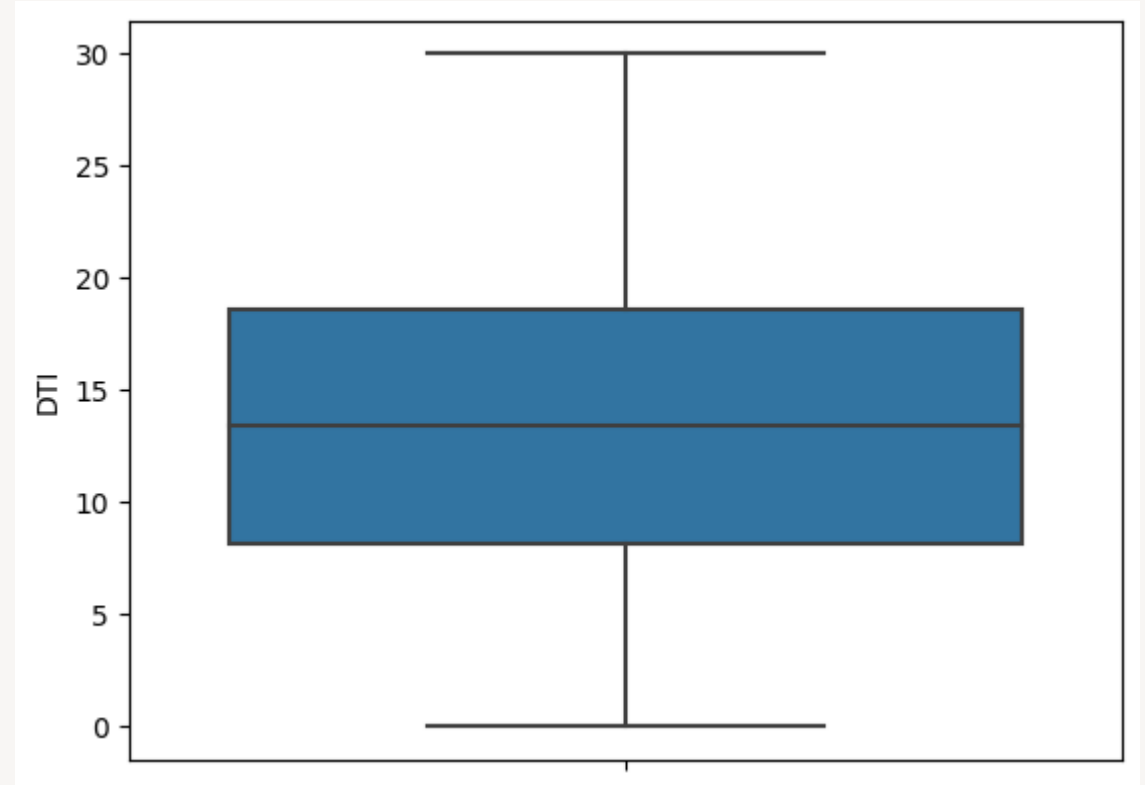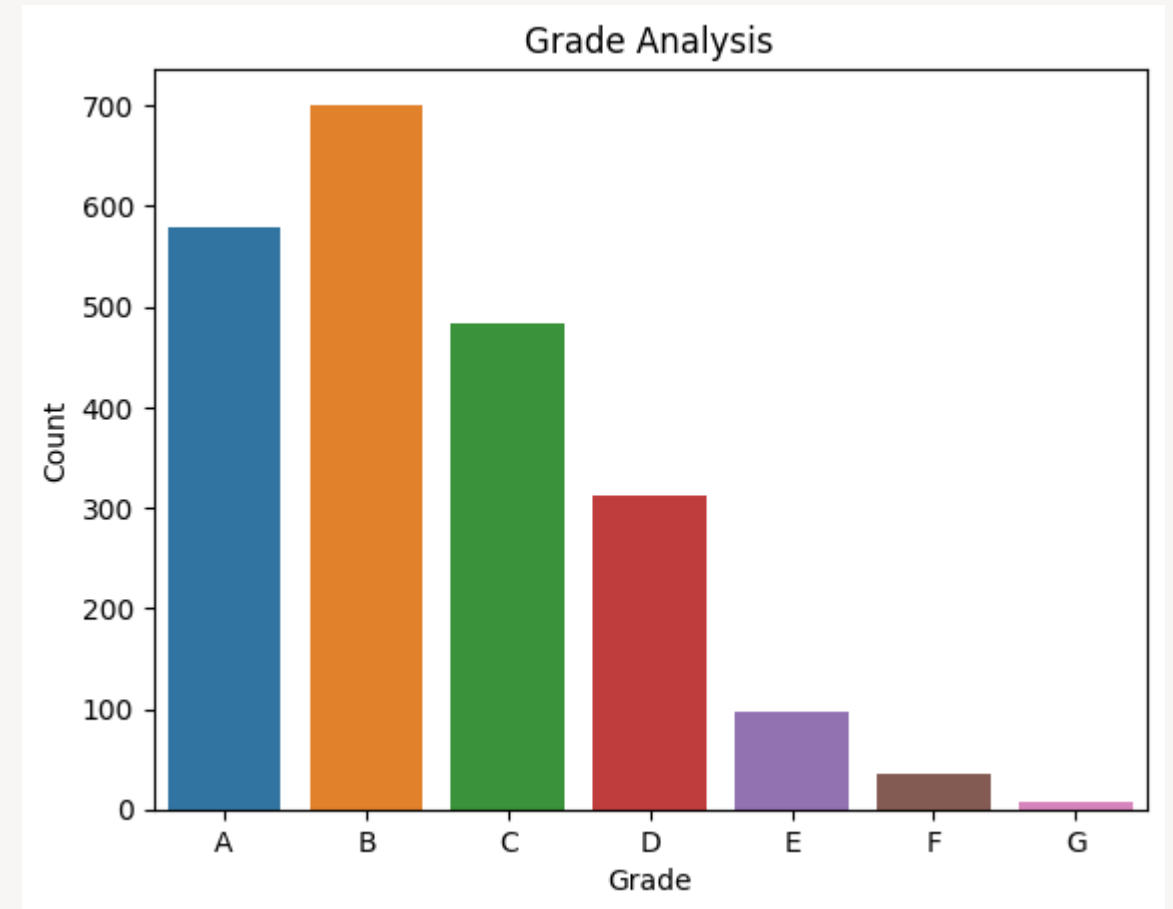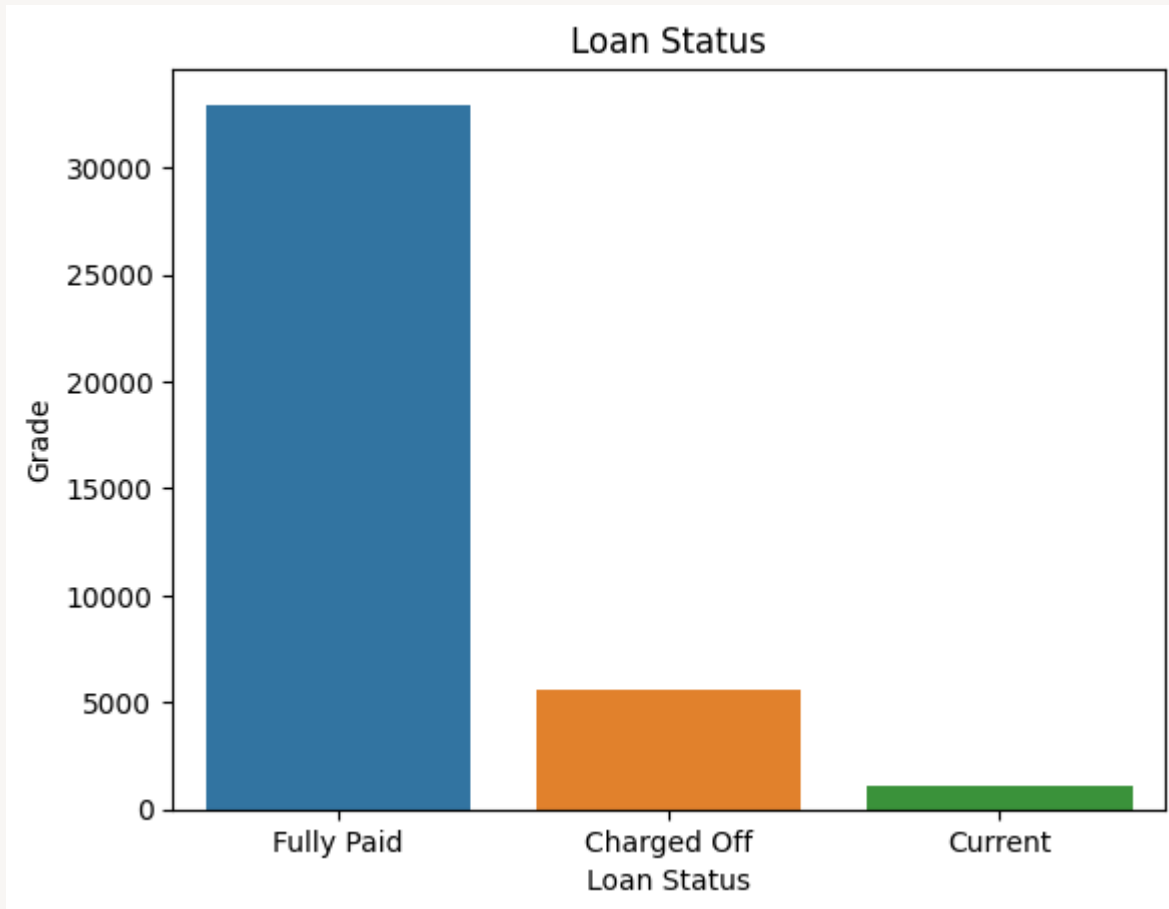

- Outliers.

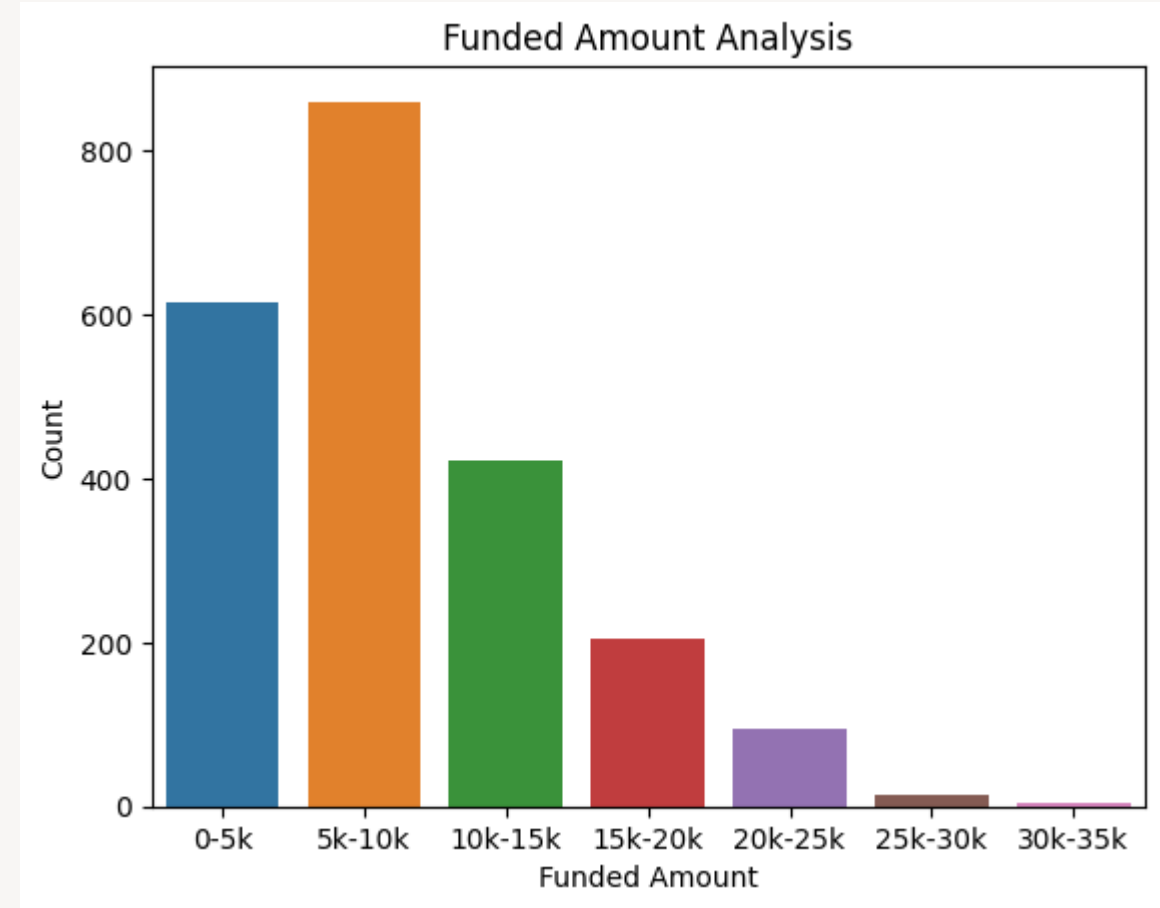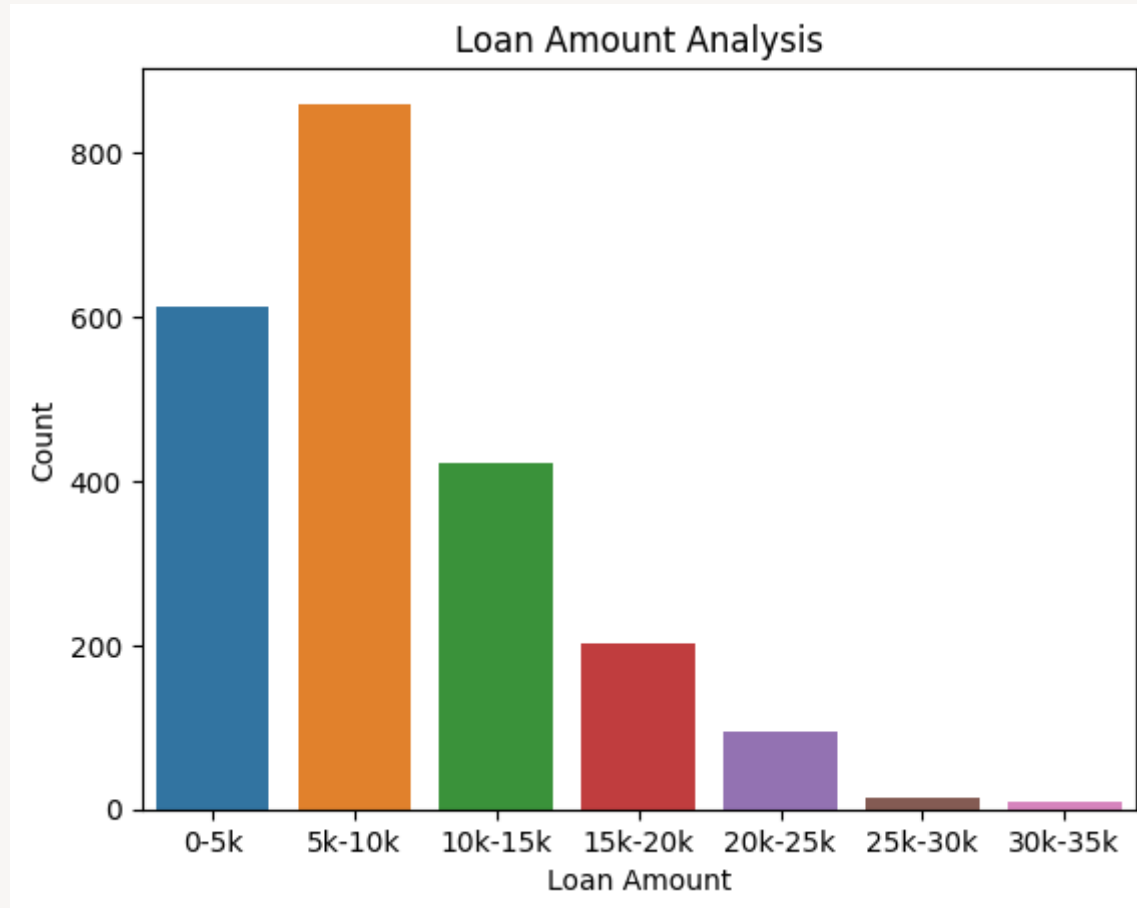# Finding Outliers

- Outliers.

Data Analysis

Segmented
Univariate
Analysis

Multivariate
Analysis

Univariate
Analysis

Bivariate Analysis

Derived Metrics

Univariate Analysis
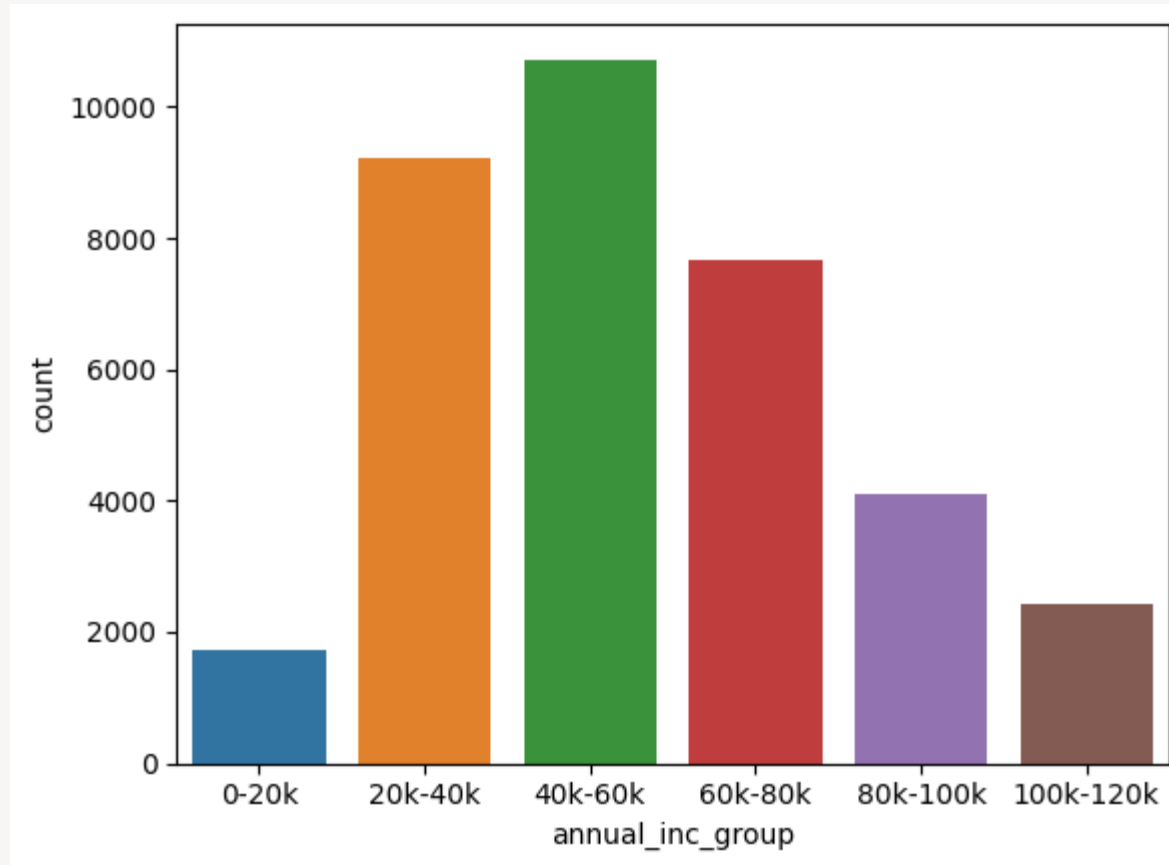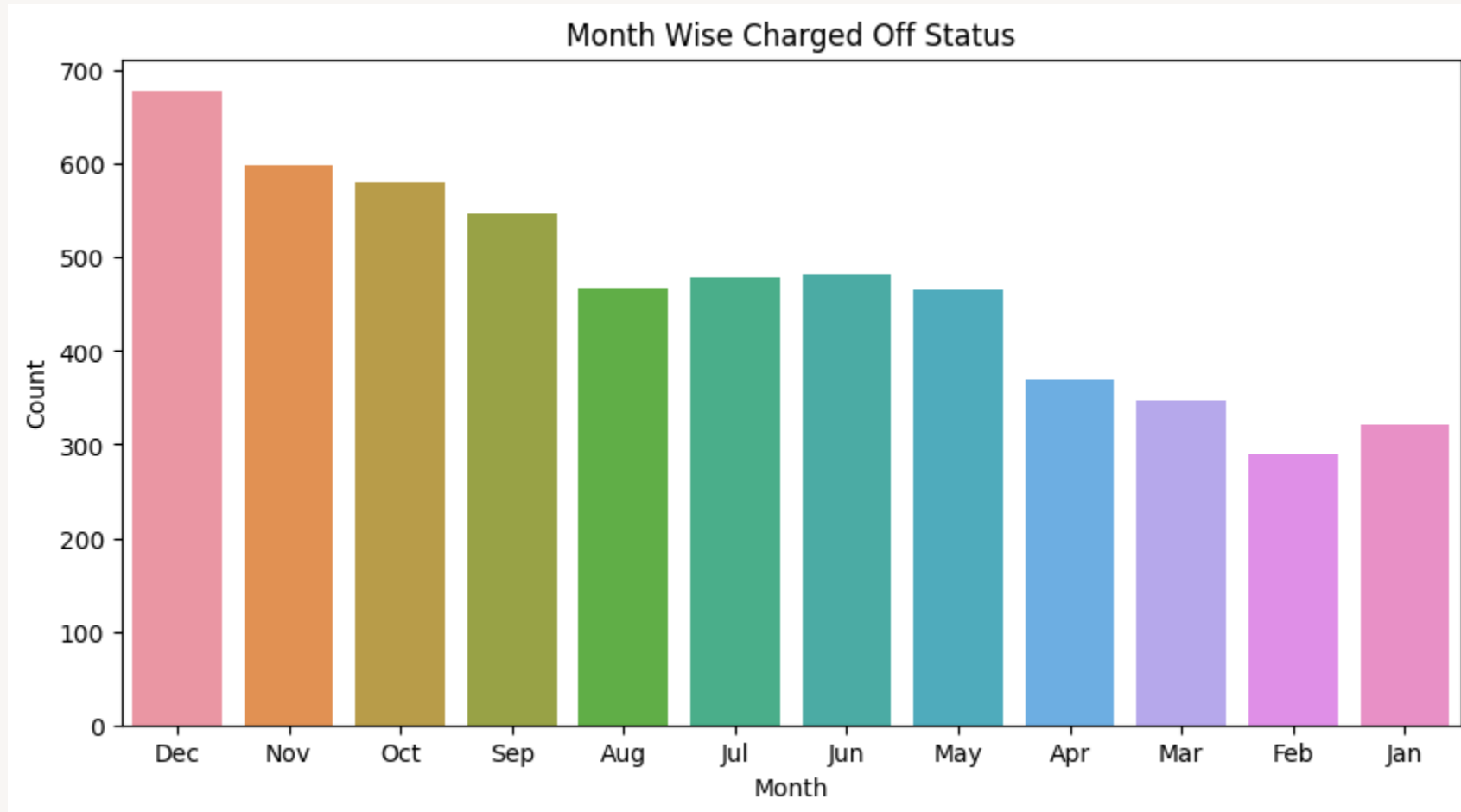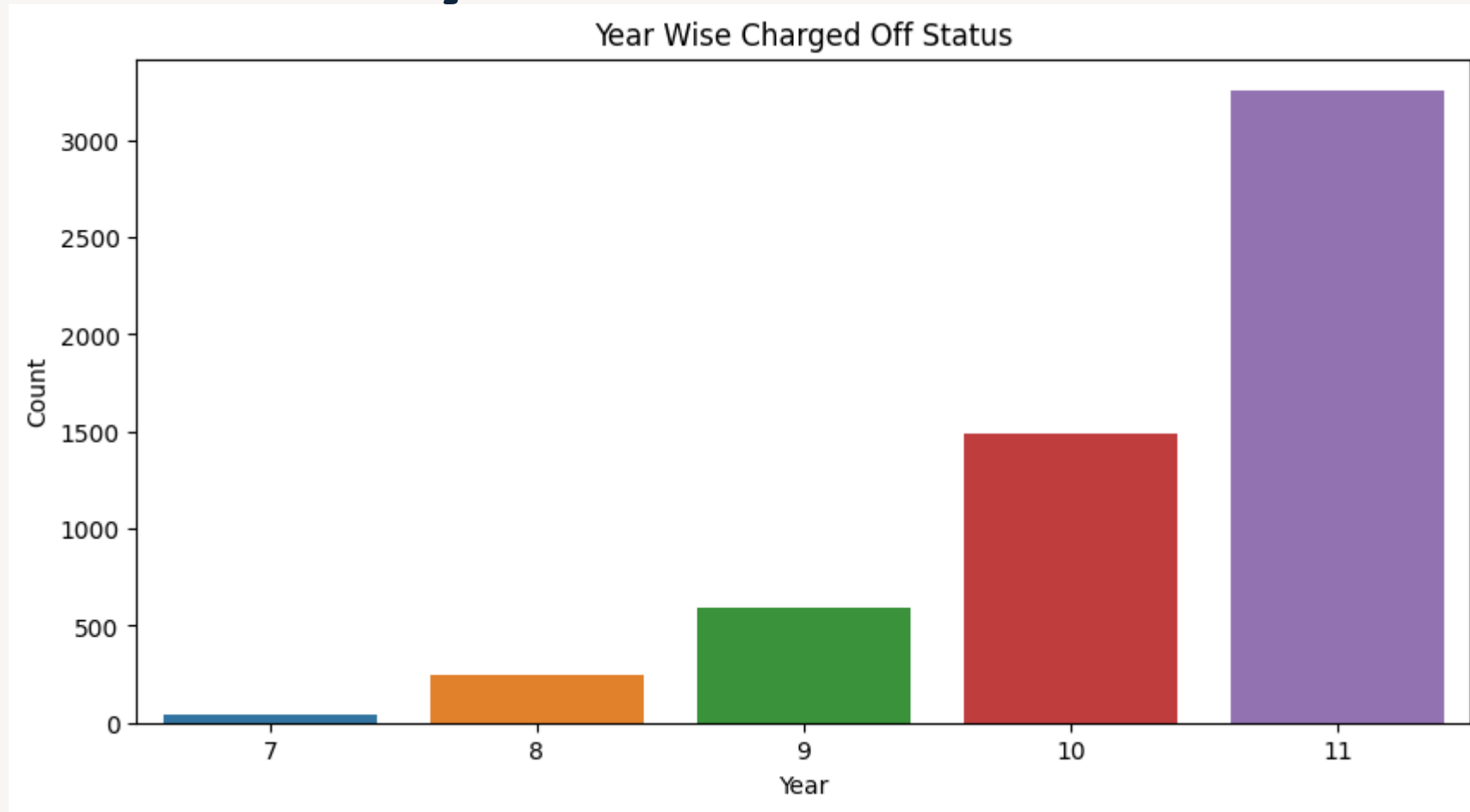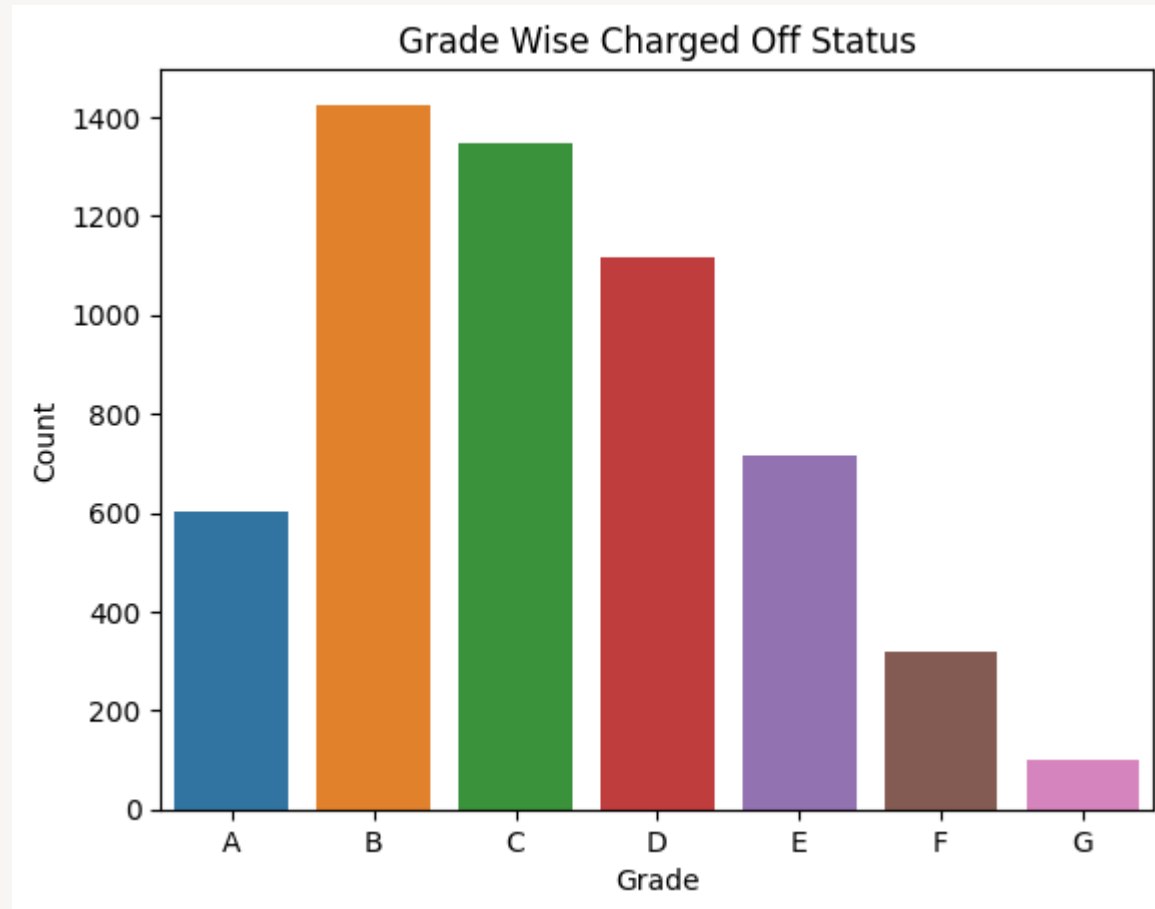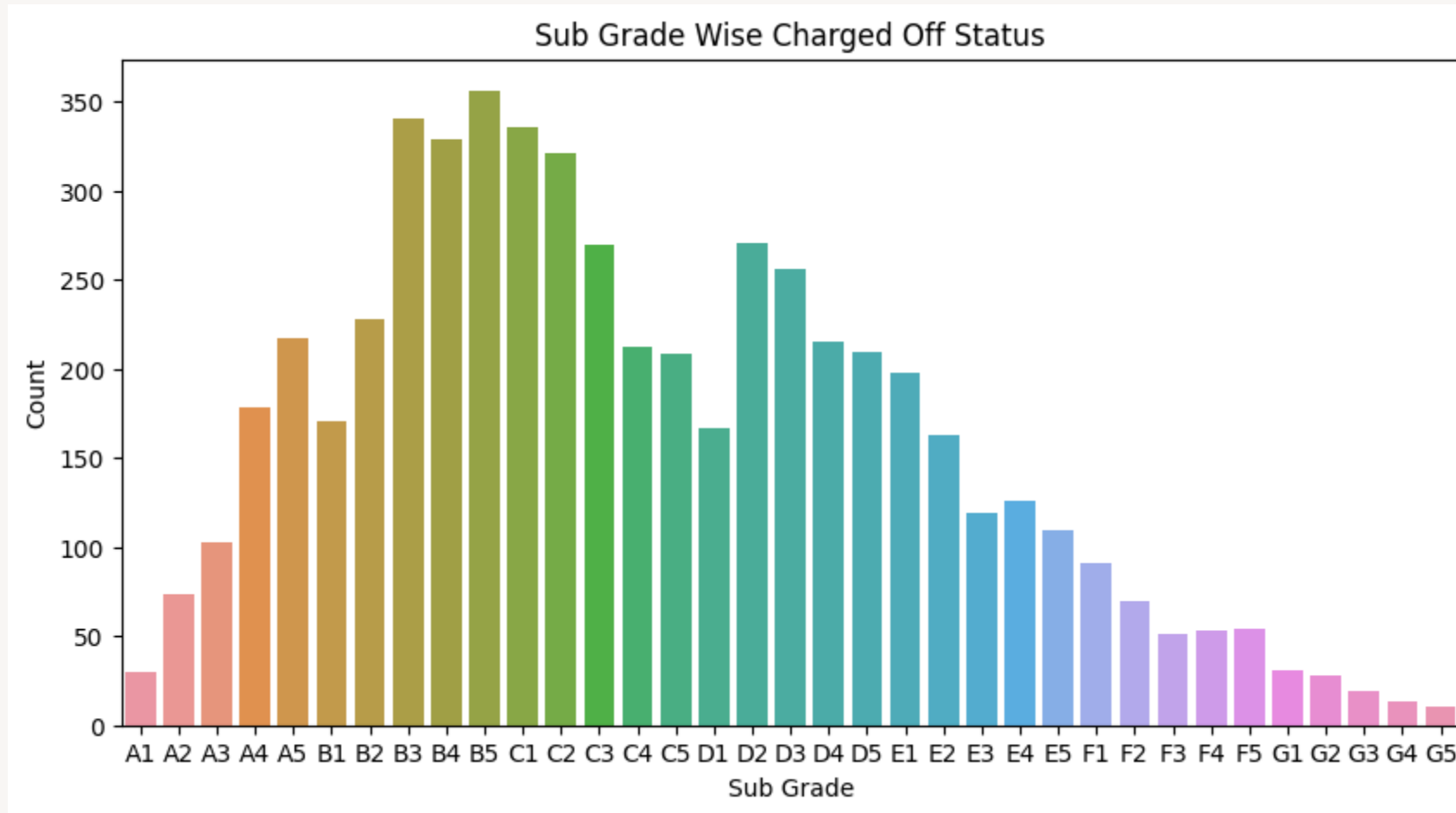

- Loan

Univariate Analysis
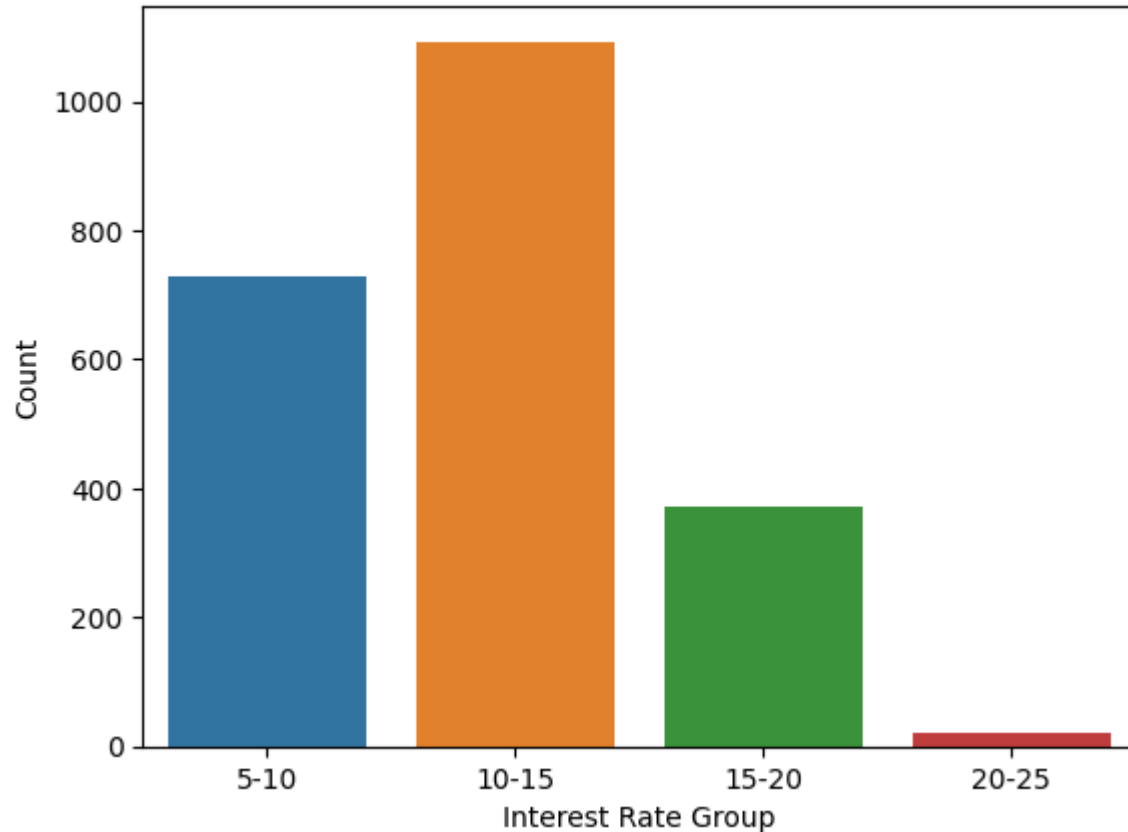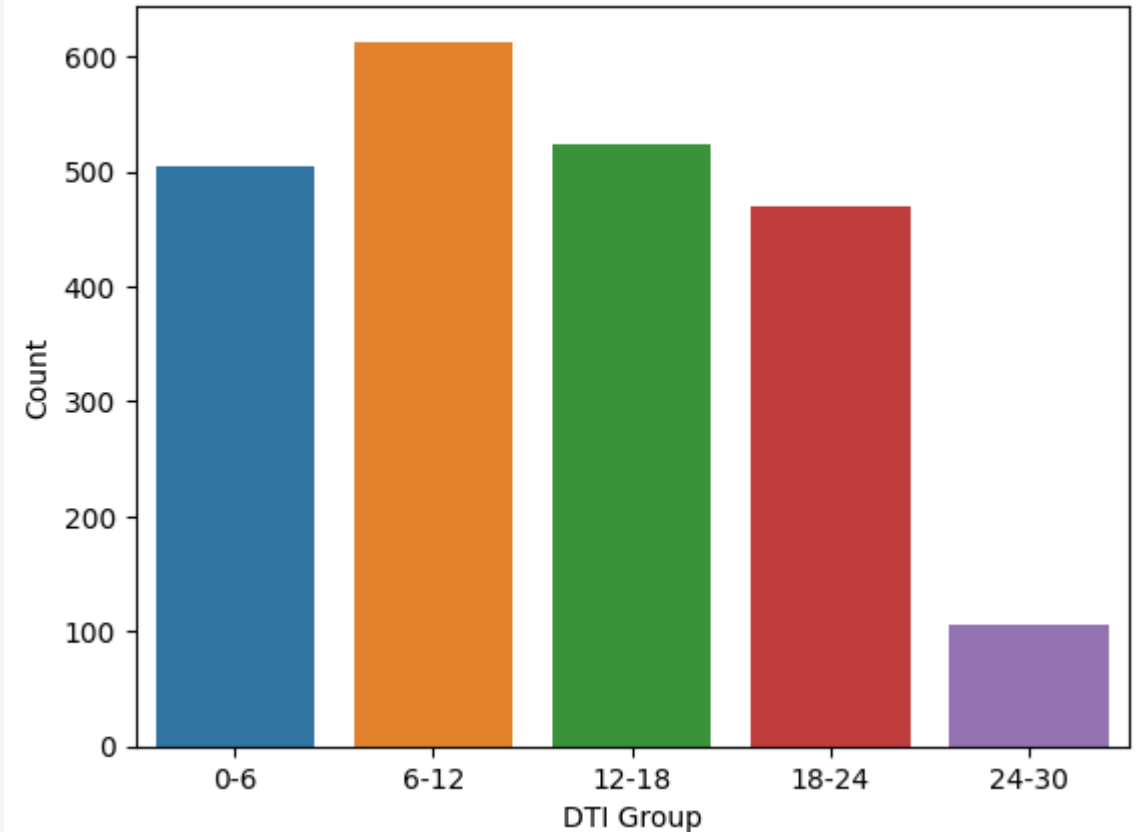
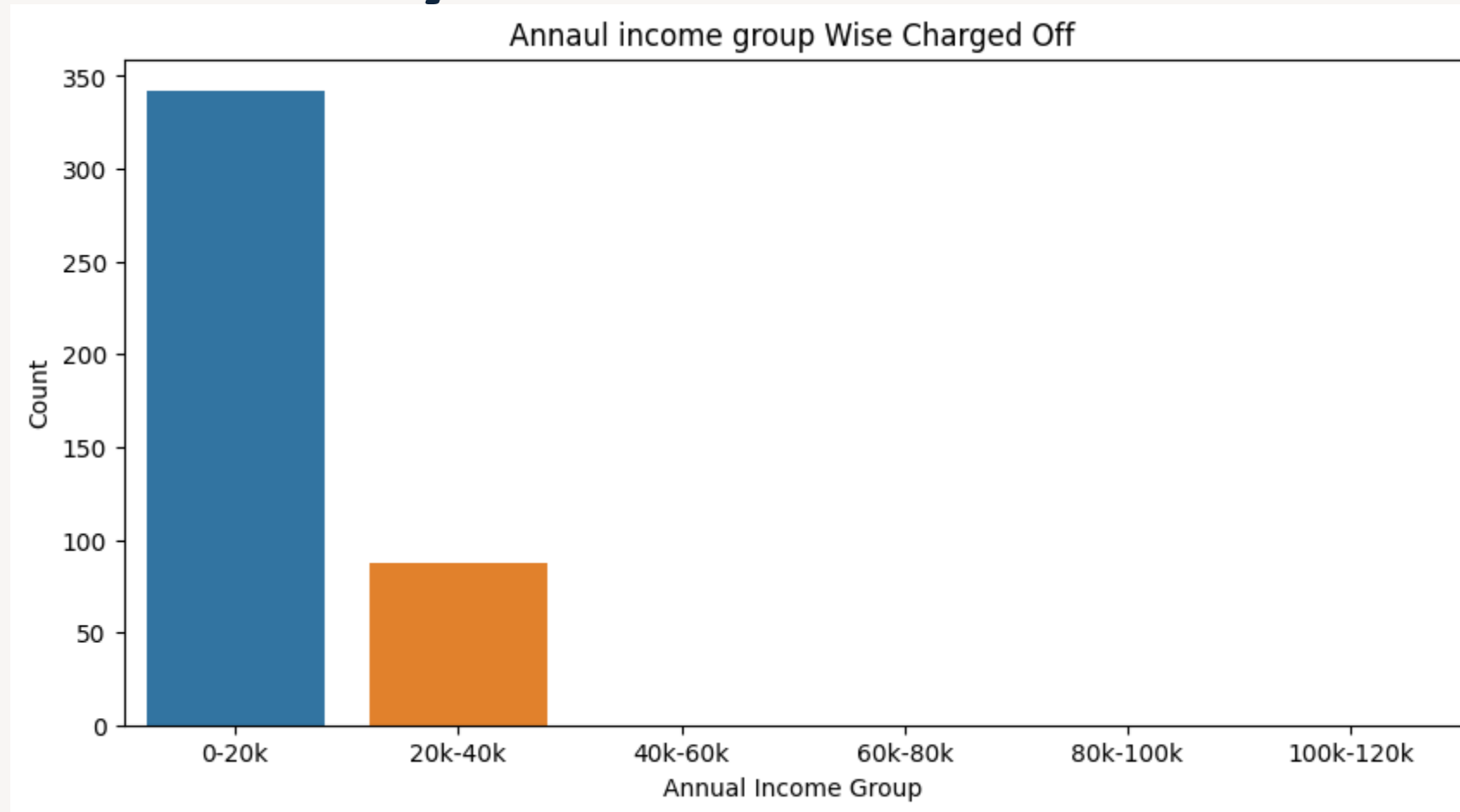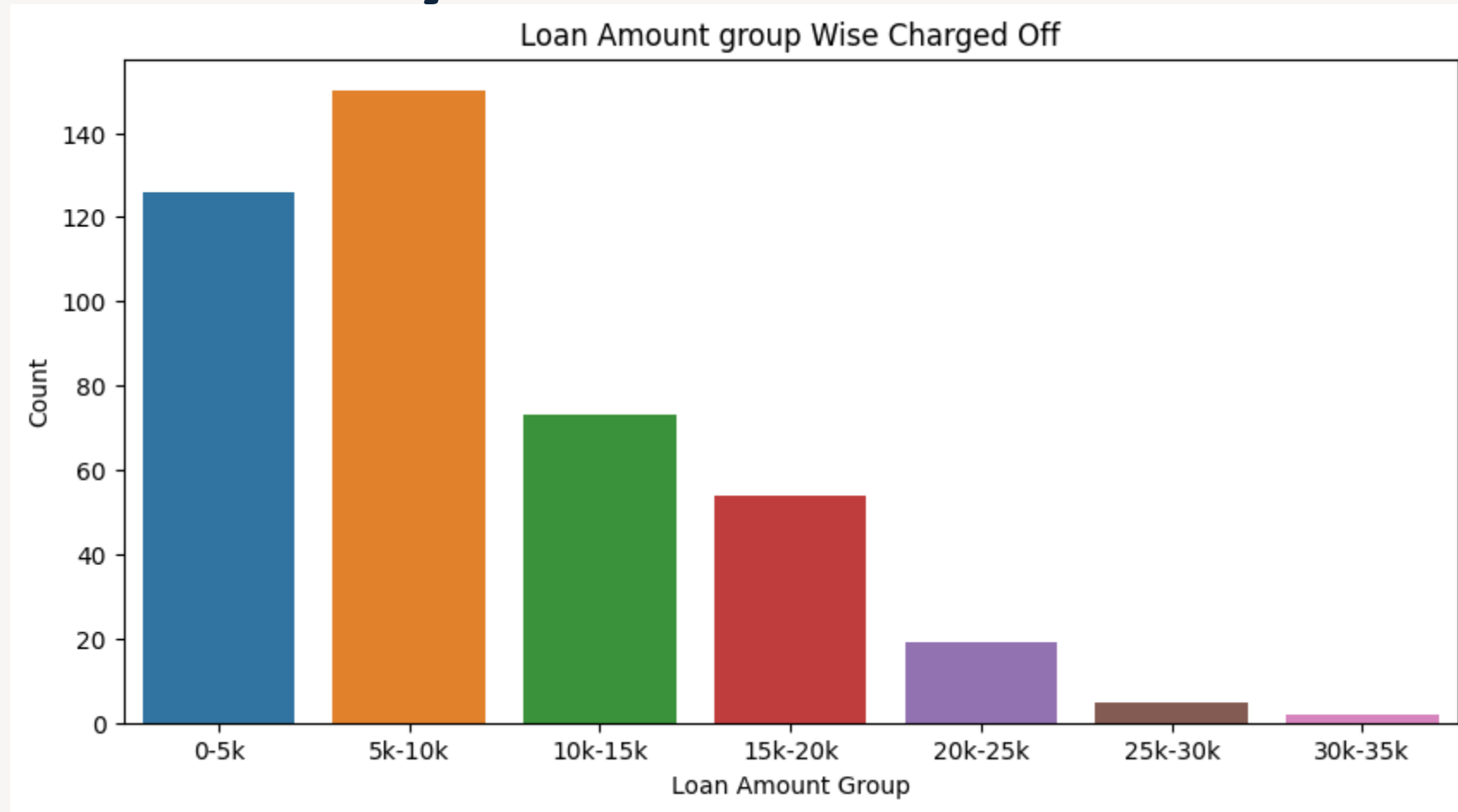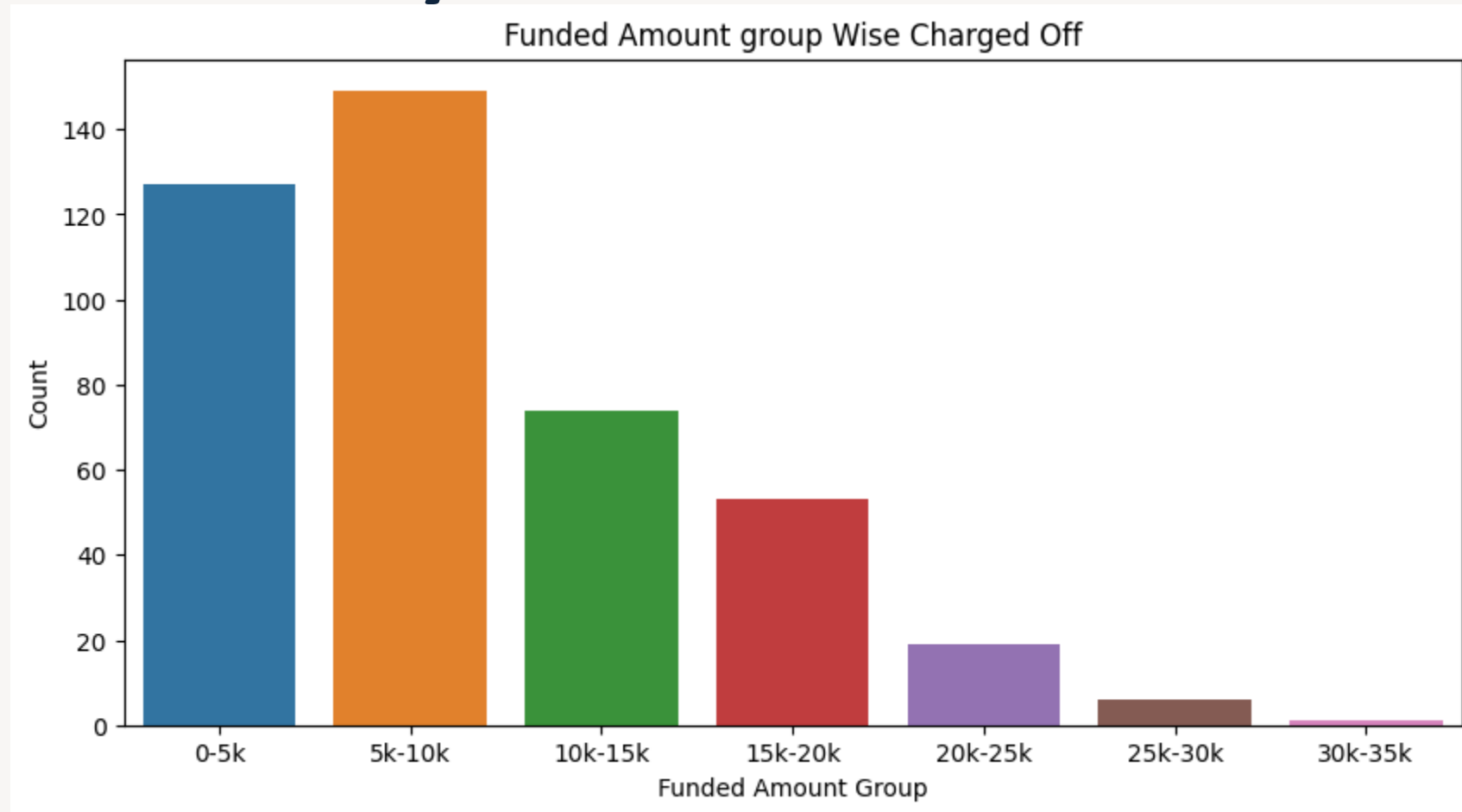# Univariate Analysis

Univariate Analysis

Univariate Analysis

Univariate Analysis
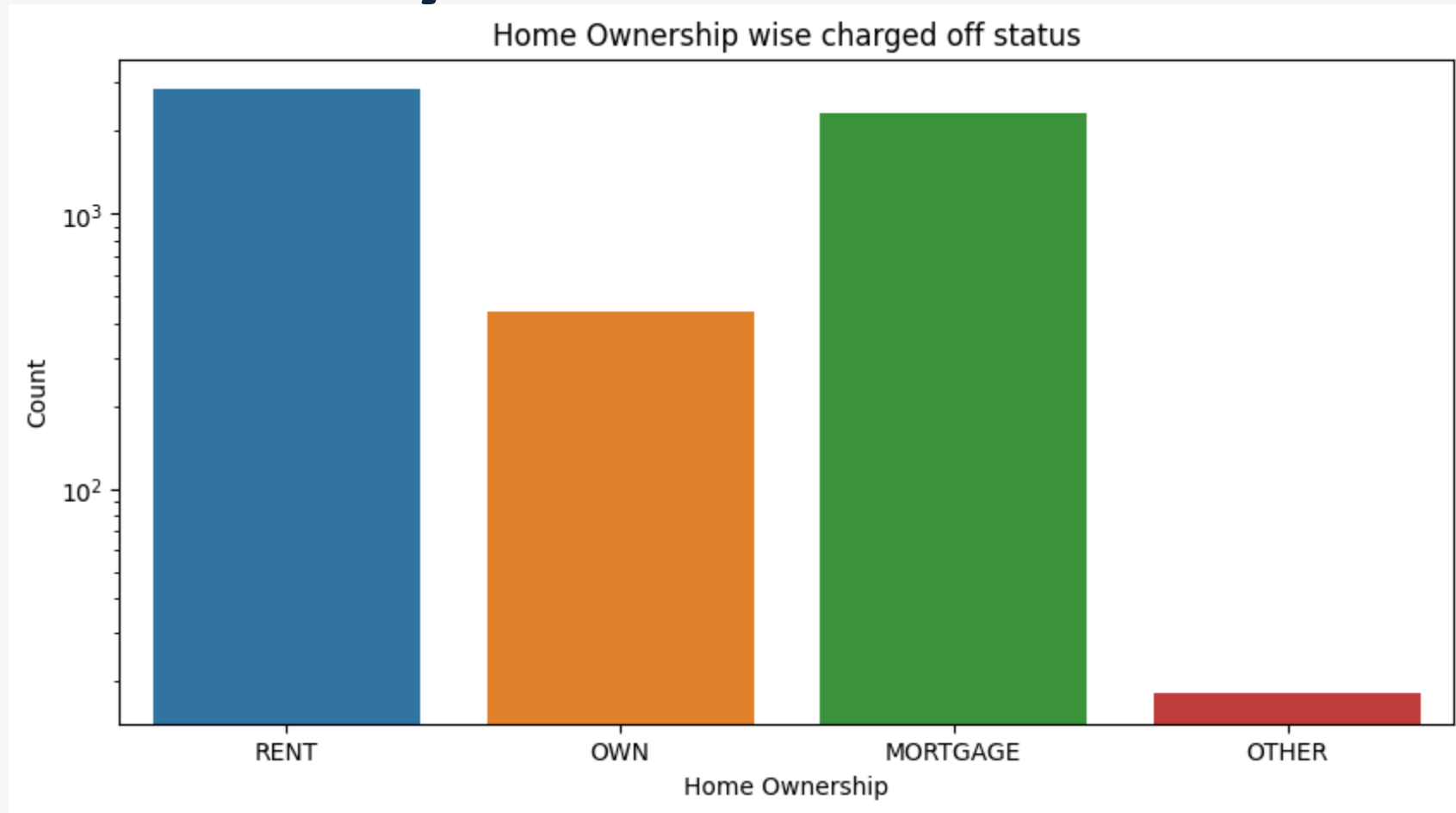

Year Wise Charged Off Status

Univariate Analysis

Grade Wise Charged Off Status

Univariate Analysis

Sub Grade Wise Charged Off Status

Univariate Analysis

Univariate Analysis

Loan Amount group Wise Charged Off

Univariate Analysis

Univariate Analysis

Home Ownership wise charged off status

Univariate Analysis

Purpose wise charged off status

Univariate Analysis

Univariate Analysis

Term wise charged off status

Univariate Analysis

State wise charged off status

Bivariate Analysis

Verification Status wise charged off status

Bivariate Analysis

Annual Income and Purpose wise charged off status

Bivariate Analysis

Bivariate Analysis

Loan Amount and Purpose wise Loan Status

Bivariate Analysis

Bivariate Analysis

Loan amount and verification status wise loan status
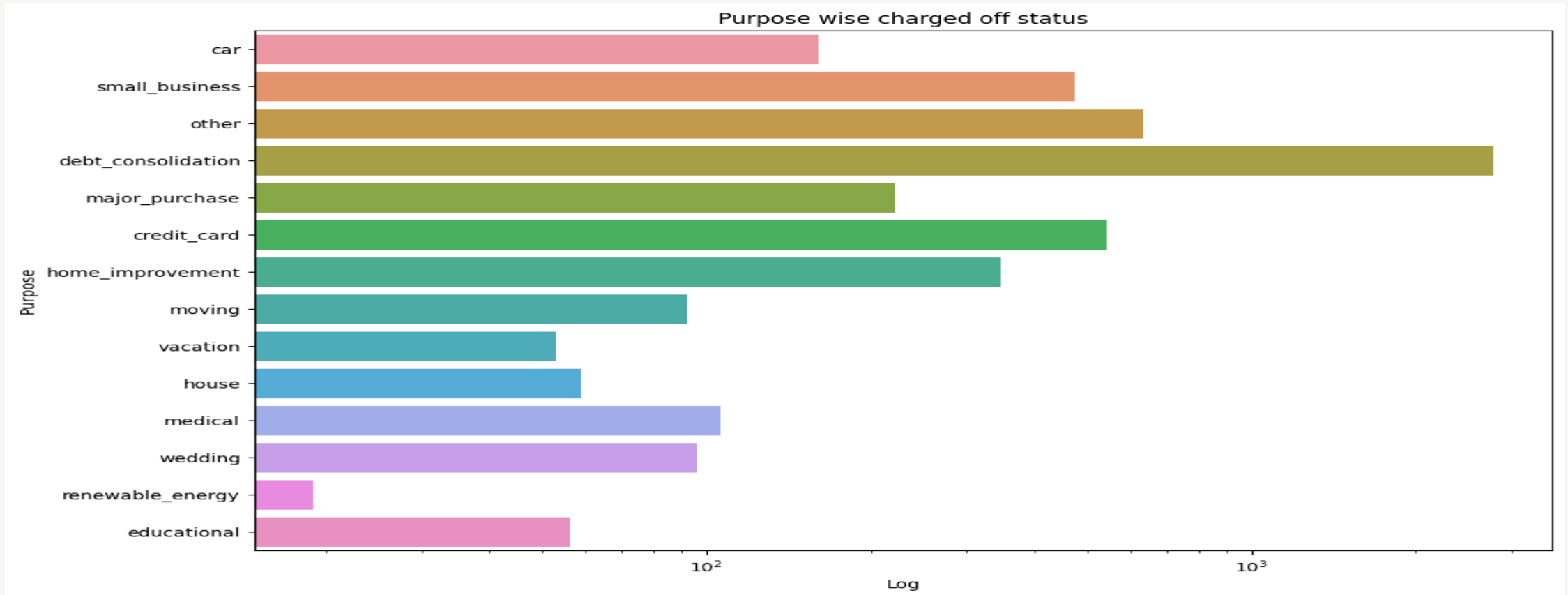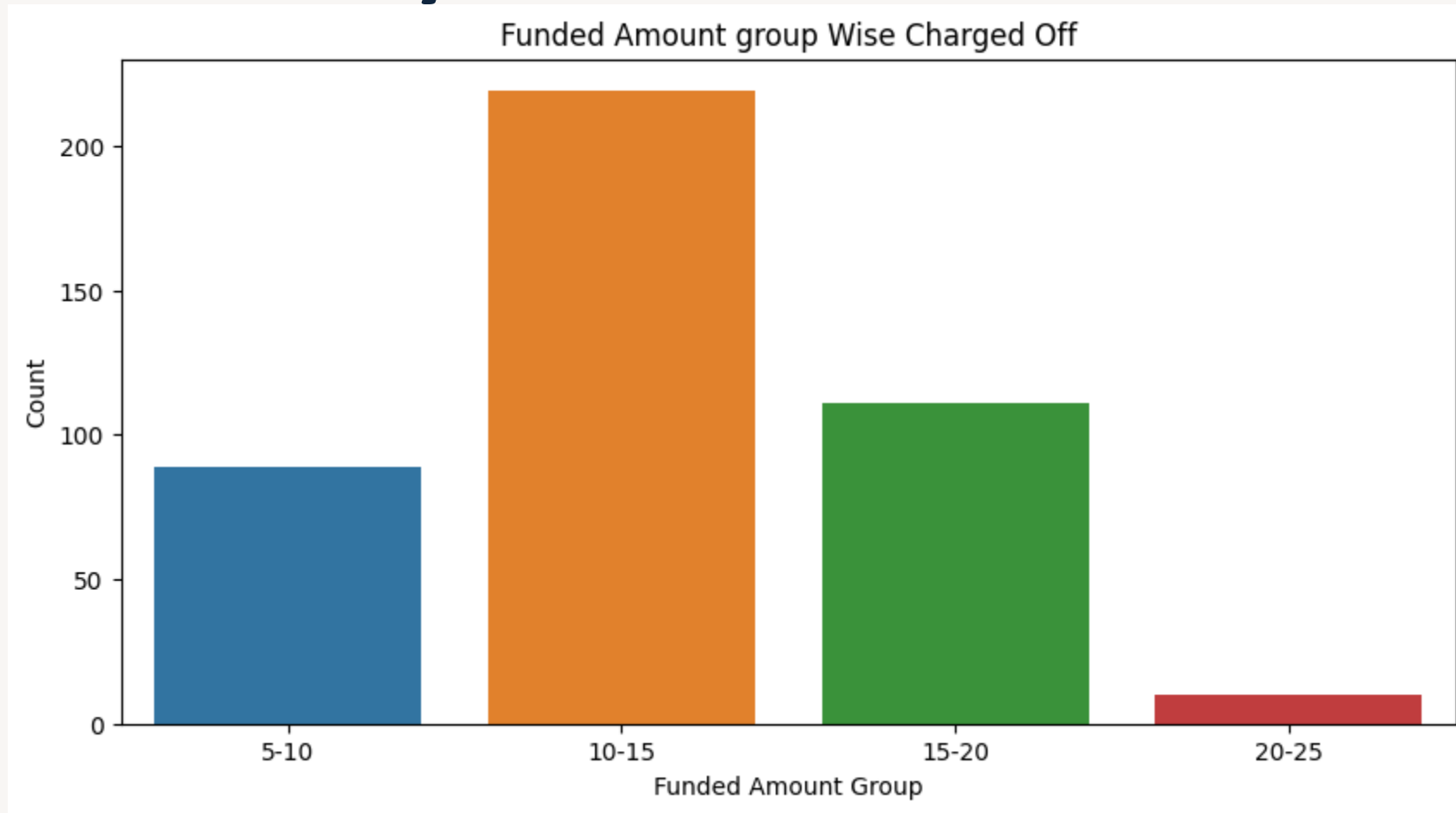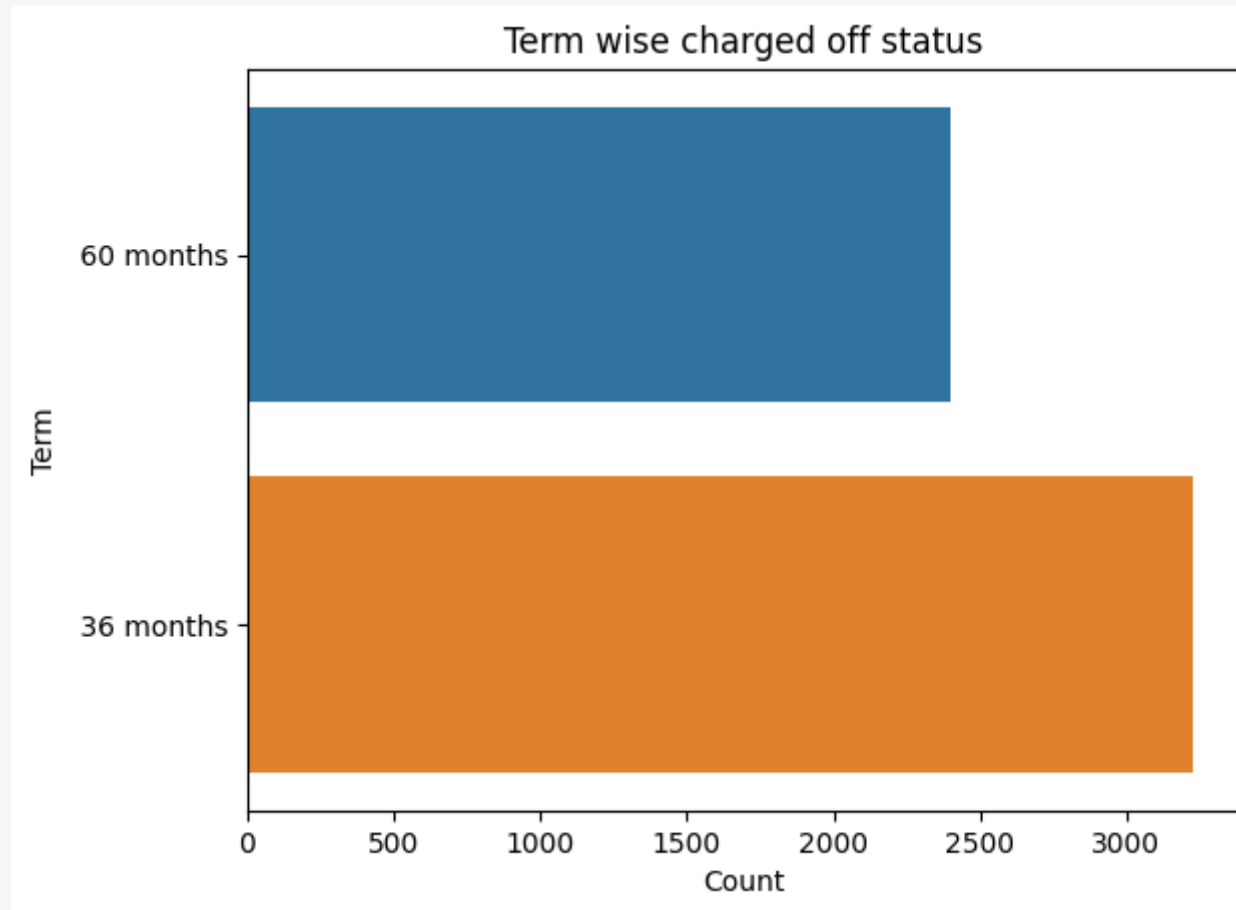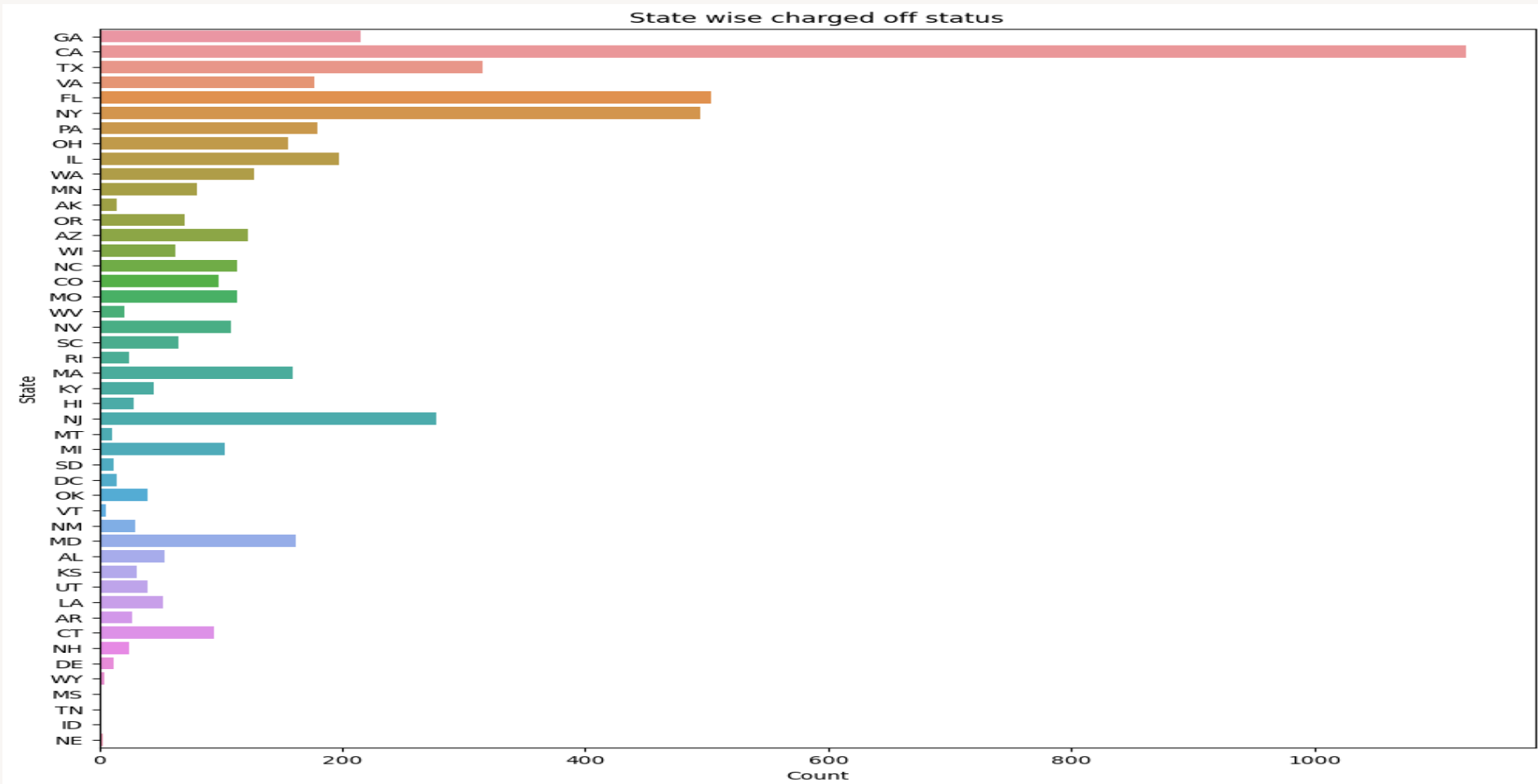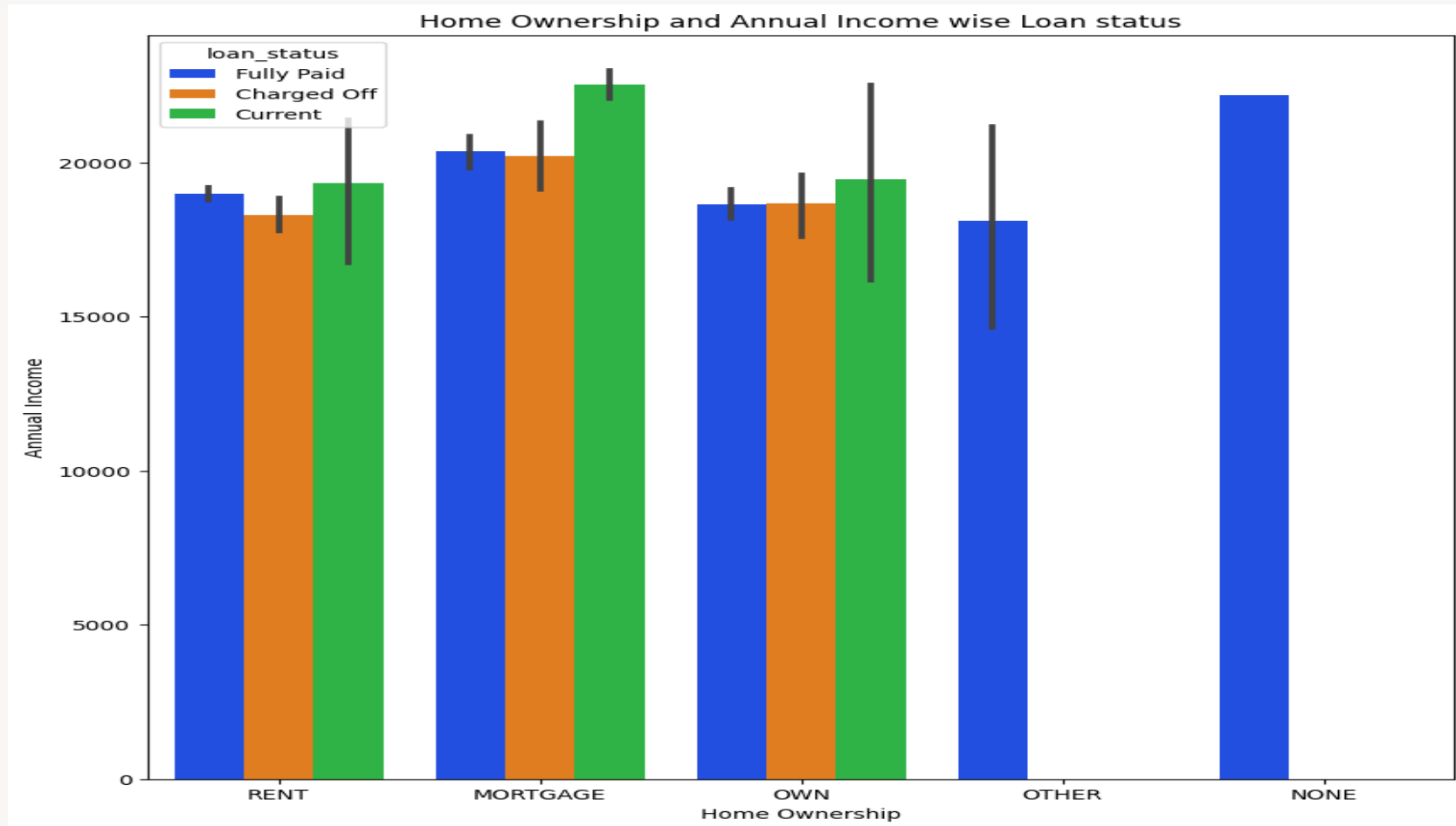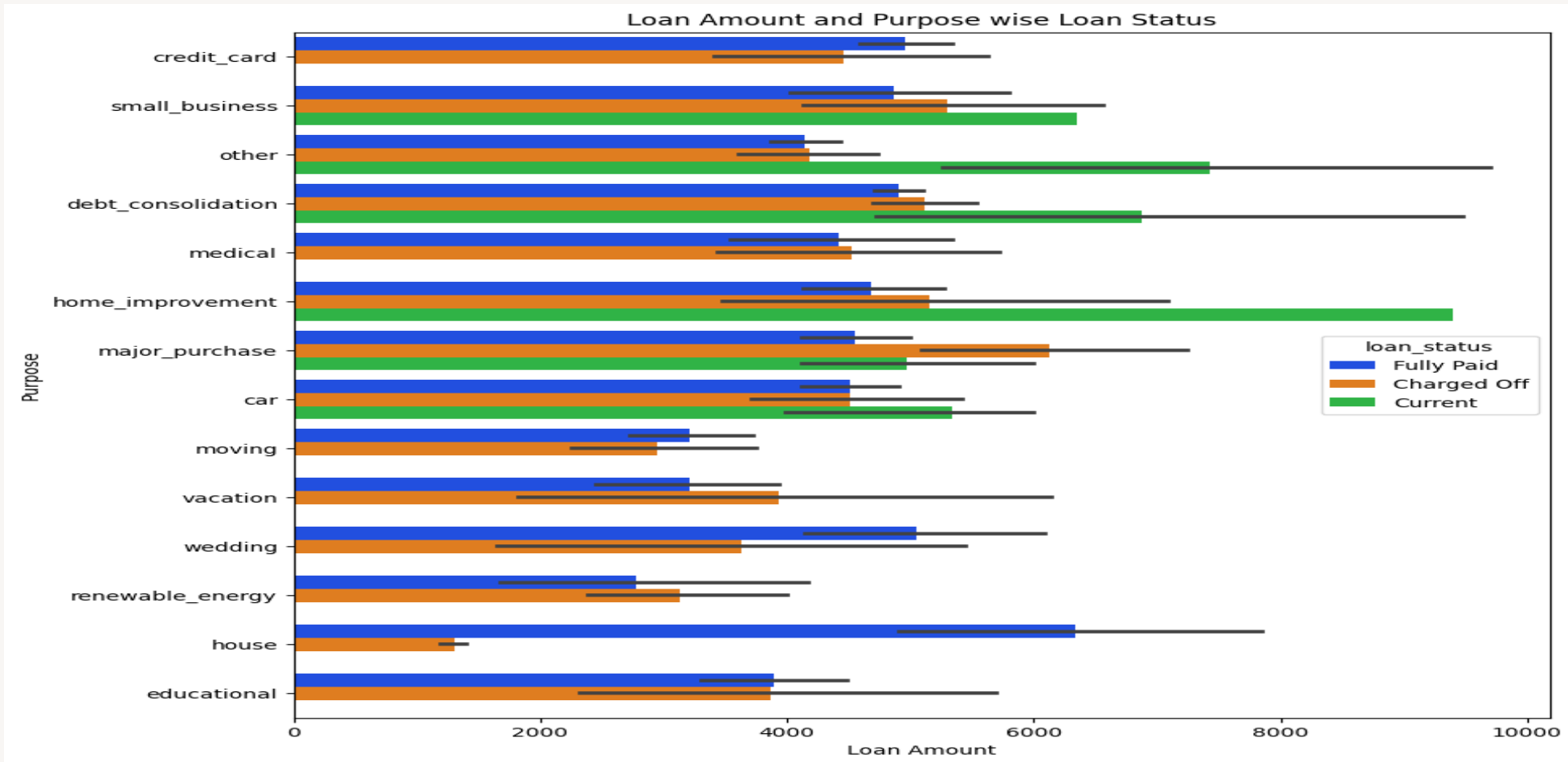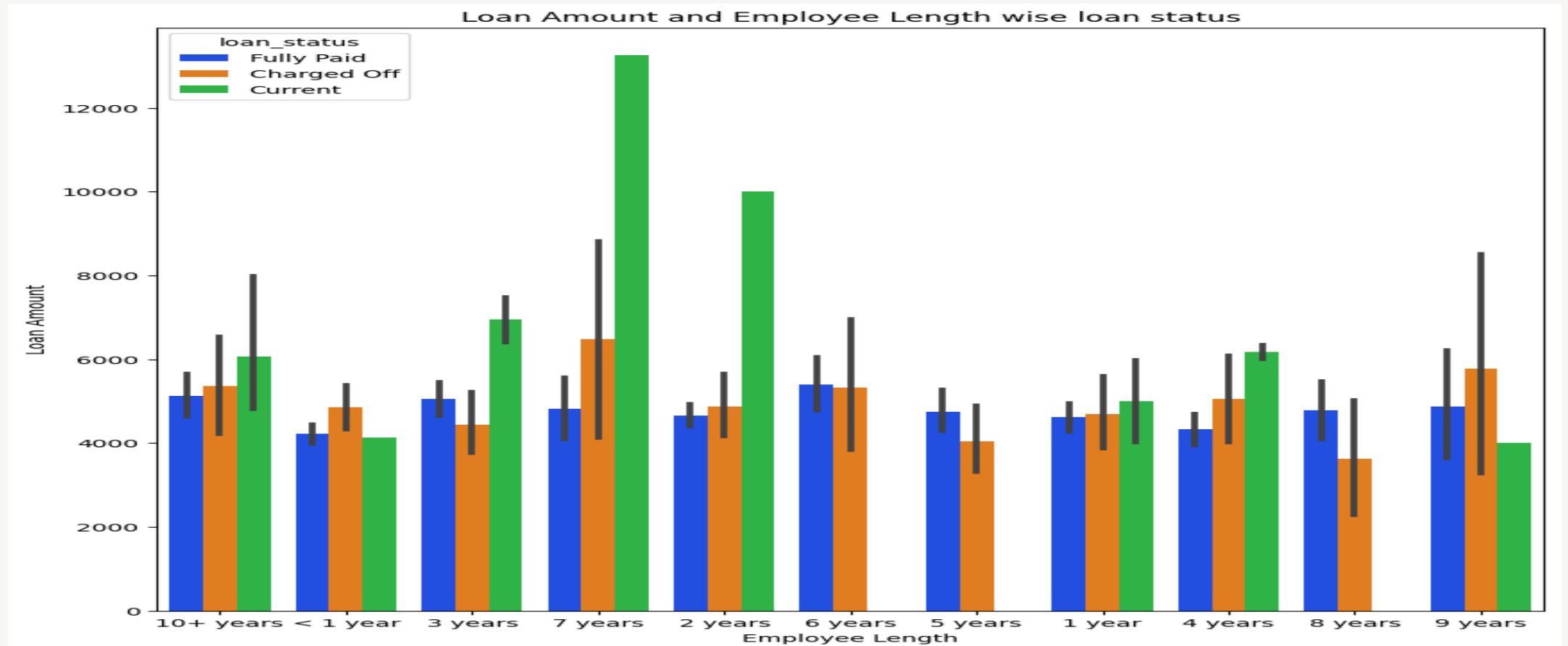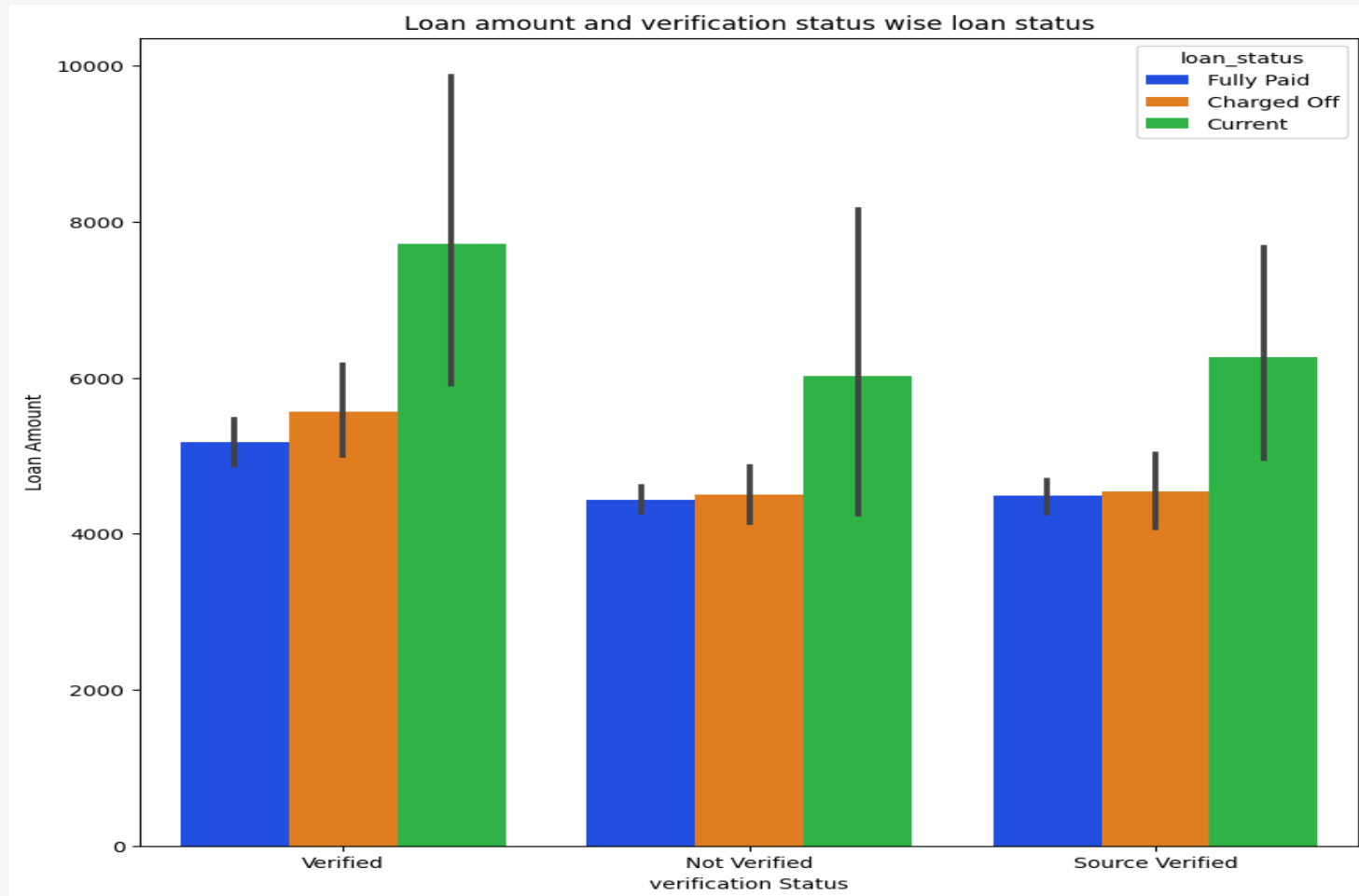
Bivariate Analysis
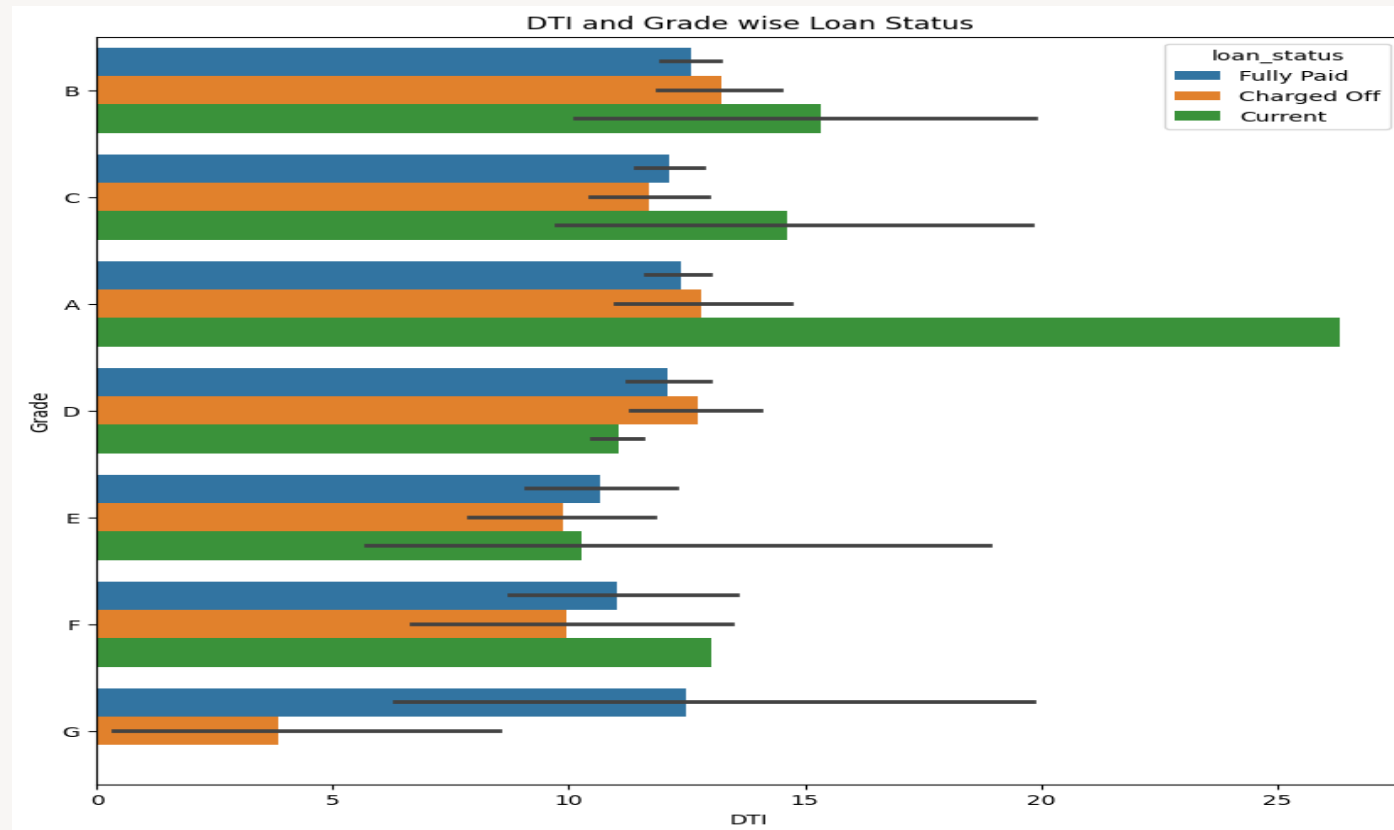
Correlation Matrix
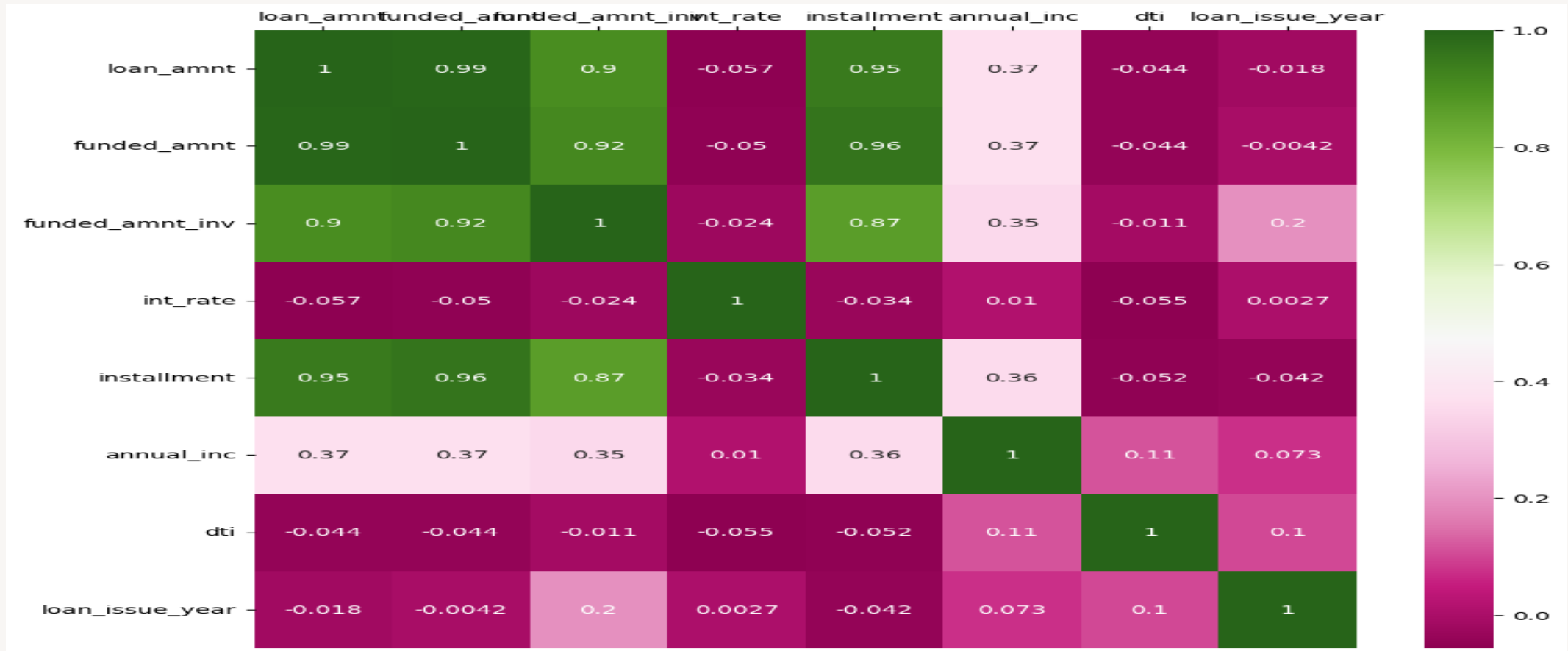
Observations

Univariate Analysis

- Fully loan paid average is more than defaulter.
- There are A and B grade where most of the loans are charged off and least loans are charged off in Grade F and G.
- Most of the employee want loan between 0-10k where 5-10k are most highest number of loan asking group.
- Mostly loans are funded between 0-10k where 5-10k are most highest number of loan funded group.
- Most of the loans are provided between 5-15% interest rate.
- 6-12 is highest group of DTI where loan is provided where 24-30 group is lowest area where loan is provided.
- Most of the loans are provided between 40k to 60k annual income group.
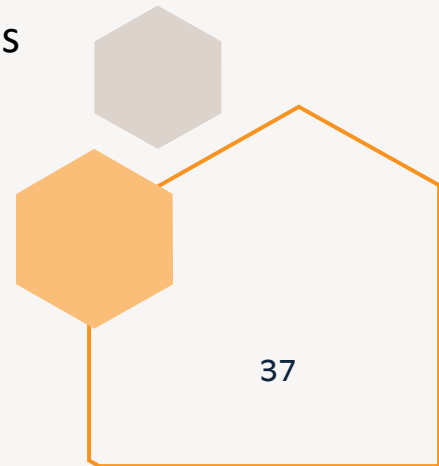
Observations

Bivariate Analysis

- The loan issued in December is the highest defaulter and loan issue in February is the lowest defaulter.
- The loan issued in 2011 is the highest defaulter and loan issued in 2007 is the lowest defaulter.
- Loan provided in Grade B is highest defaulter where loan provided in Grade G is lowest defaulter.
- Sub Grade Group B3 to C3 is highest sub group where loan is defaulted.
- The loan provided between 0-20k is the highest defaulter.
- Loan asking 0-10k group have highest number of defaulter.
- Loan funded in group 0-10k have highest number of defaulter.
- Rent and Mortgage is highest number of defaulter.
- Taken loan for Debt Consolidation purpose is highest number of defaulter where taken loan for renewable energy has least number of defaulter.
- Interest rate between 10-15% have highest number of defaulter where highest interest rate loan have less defaulters.
- 36 months loan have more defaulter than 60 months.
- California state has highest number of defaulter.
- Not verified have highest number of loan defaulters but verified have also large defaulters.
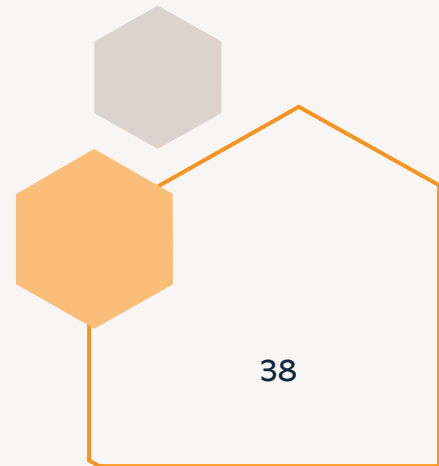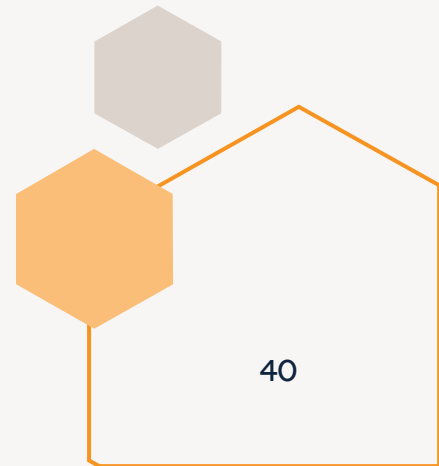
Observations

Multivariate Analysis

- Nearest 200k annual income and purpose is Home improvement and credit card is highest number of defaulters.
- Nearest 200k annual income and Mortgage is highest number of defaulters.
- Loan amount is 6k-8k and purpose is Major Purchase have highest number of defaulters.
- Loan amount is 6k-8k and employee length 7 year have number of defaulters.
- Between dti 10-15 and grade A to D have highest defaulters.
- Wallmart employee is highest number of defaulter.
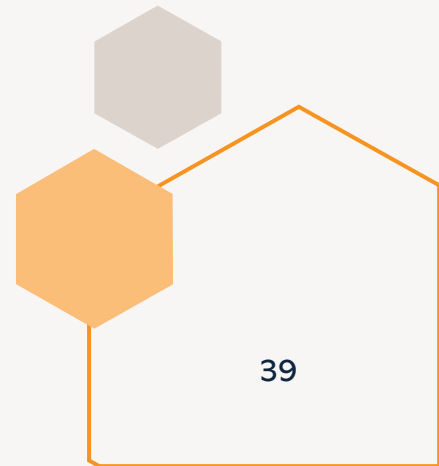
Recommendations

- Applications from lower grades and with public record bankruptcies should not be approved.
- Applications where grades are lower and interest rate is high tend to default more.
- Debt Consolidation and Small business loans are high defaulters and lending club should be careful in granting loans for these purpose.
- Applications from California must be scrutinized more.
- Applicants with higher annual income but with low debt to income ratio should be ideal candidates.

Thank you

Nasir

Paras Paul