Lending Club Case Study

Objective:

We work in Lending club, a consumer finance company which specializes in lending various types of loans to urban customers.

As it lends money, there is always a risk of defaulting which results in major losses.

This case study provides a behavior pattern of defaulters which can be used to reduce the losses caused by defaulting.
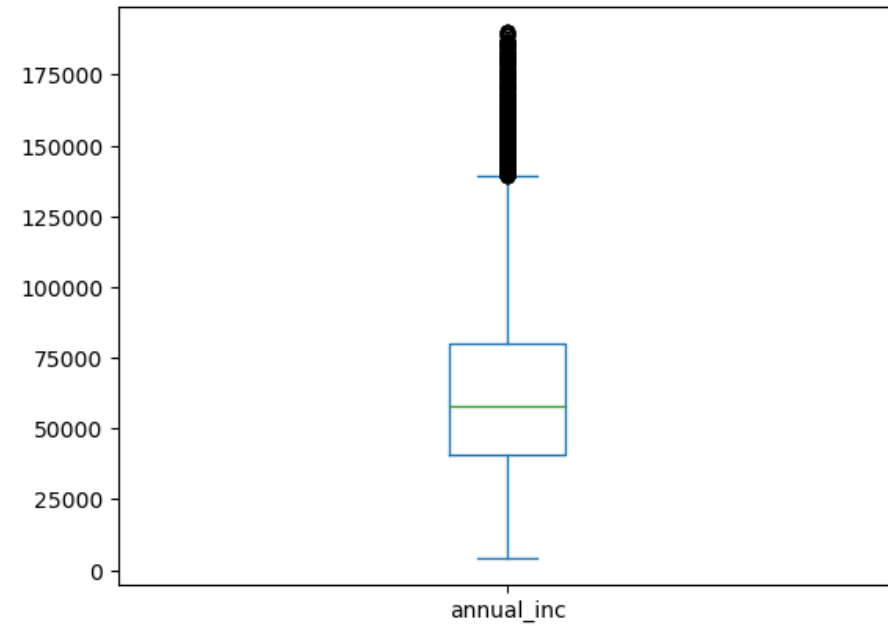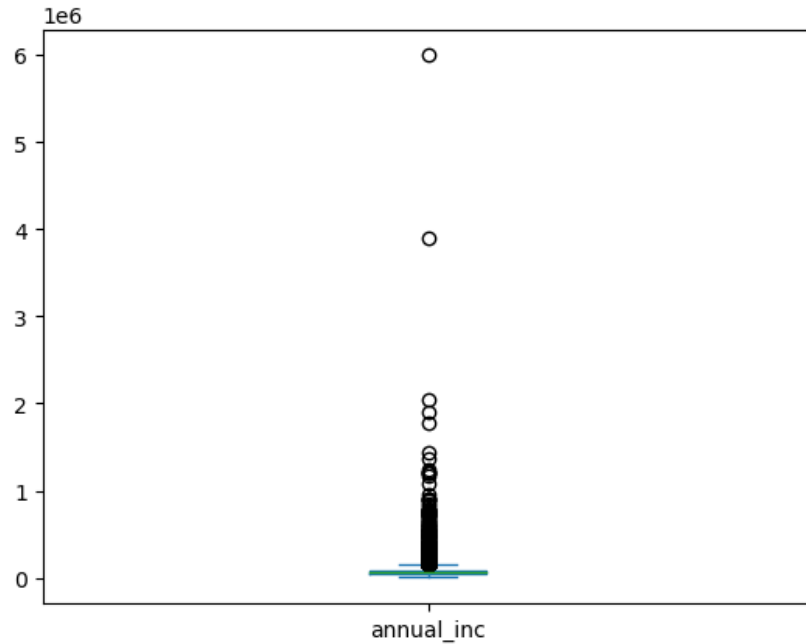
Contributors:

Purna

Vani

The process of the Exploratory Data Analysis followed in this case study is

- ➢ Data Sourcing

- ➢ Data Cleaning

- ➢ Univariate Analysis

- ➢ Bivariate Analysis

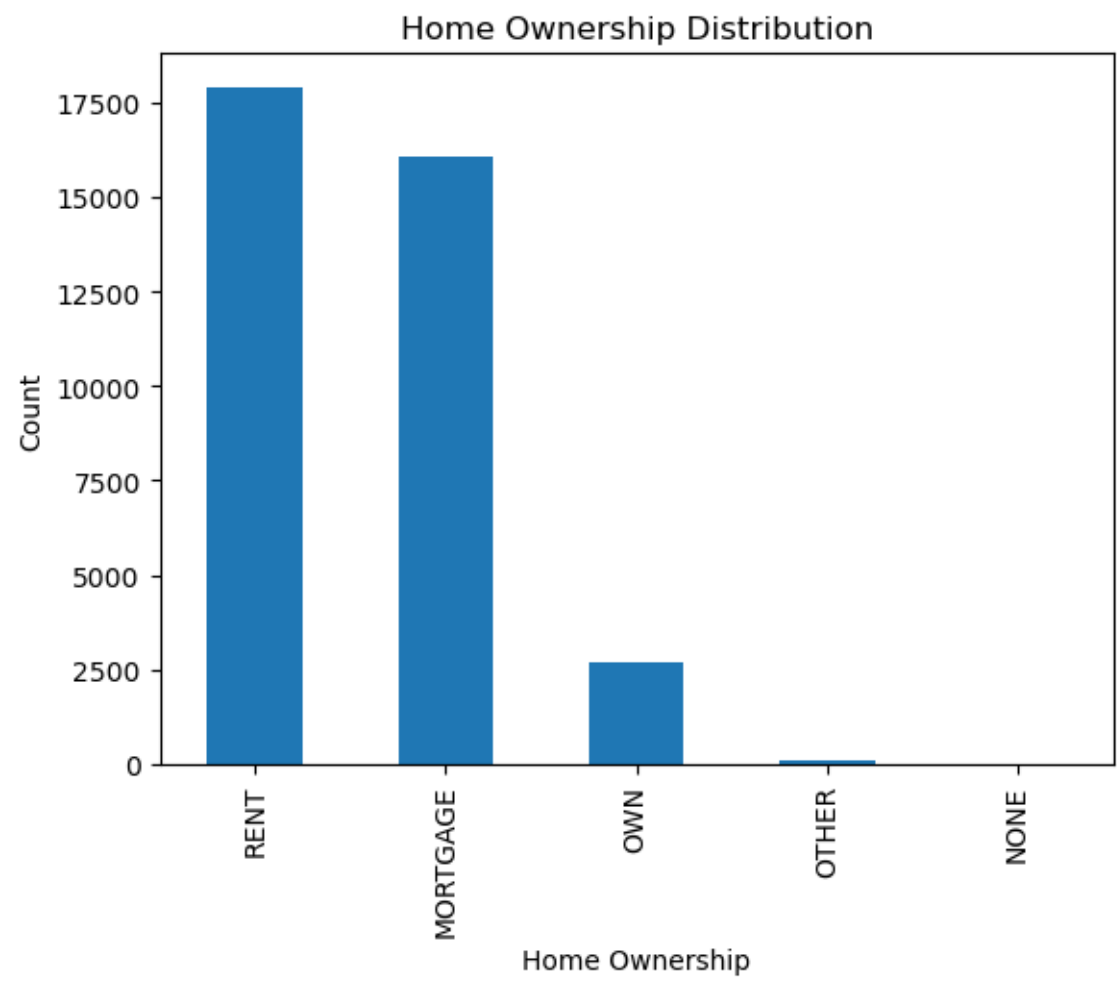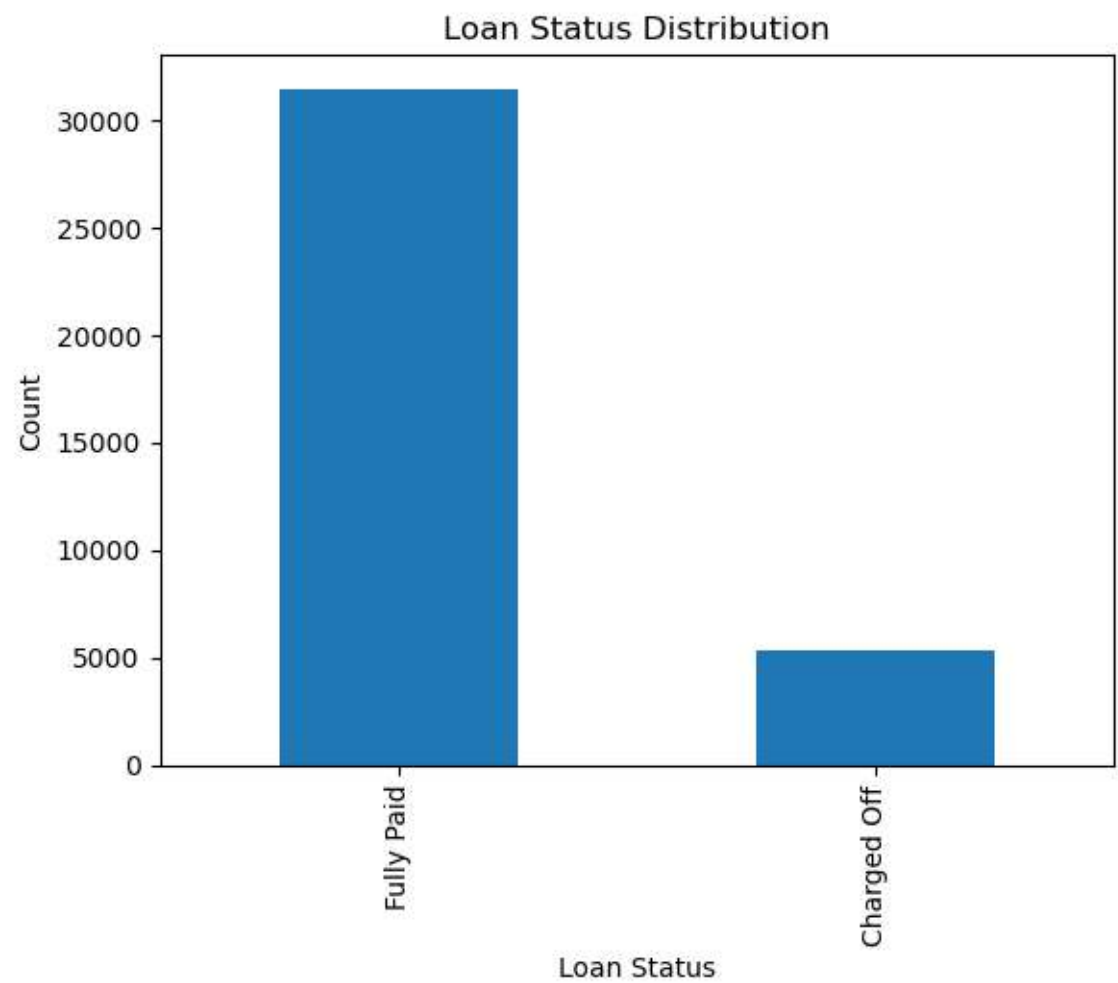- ➢ Multivariate Analysis

- ➢ Observation

Data Cleaning

➢ Removed the columns with all null values as they are of no use.

➢ Removed the columns with only one unique value as it will not be useful in analysis.

➢ Removed the column with null values more than 2000 as filling those values could give a false analysis of the data.

➢ Removed the columns with unique values more than 10000 as it would make the data complex and causing scaling issue.

➢ Removed the columns that will not be useful for the defaulting analysis as these columns are data after the loan is issued and some of the columns don't contribute to the analysis.

➢ Removed the rows with current in loan status as they have already got the loan and the data is not helpful in finding defaulter behavior analysis.

➢ Only two columns have missing data emp_length and revol_util and missing percentage is less than 3%. So we can remove the rows with missing data as it wouldn't affect much in the analysis.
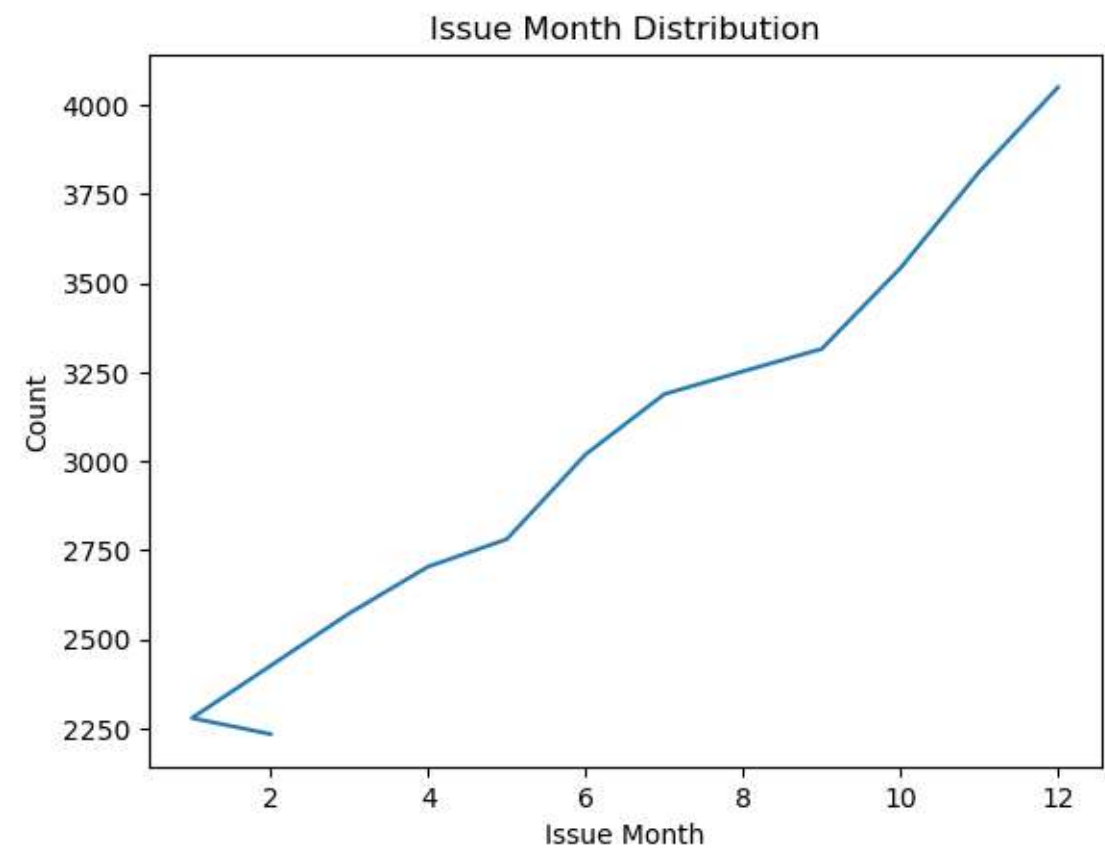
➢ Standardization of Data Types

➢ Creating New Columns from Existing Columns

➢ The outliers in annual income was high causing the scale to move to higher value but by using an percentile of 0.98, the outliers has been drastically reduced. Outliers in other measures
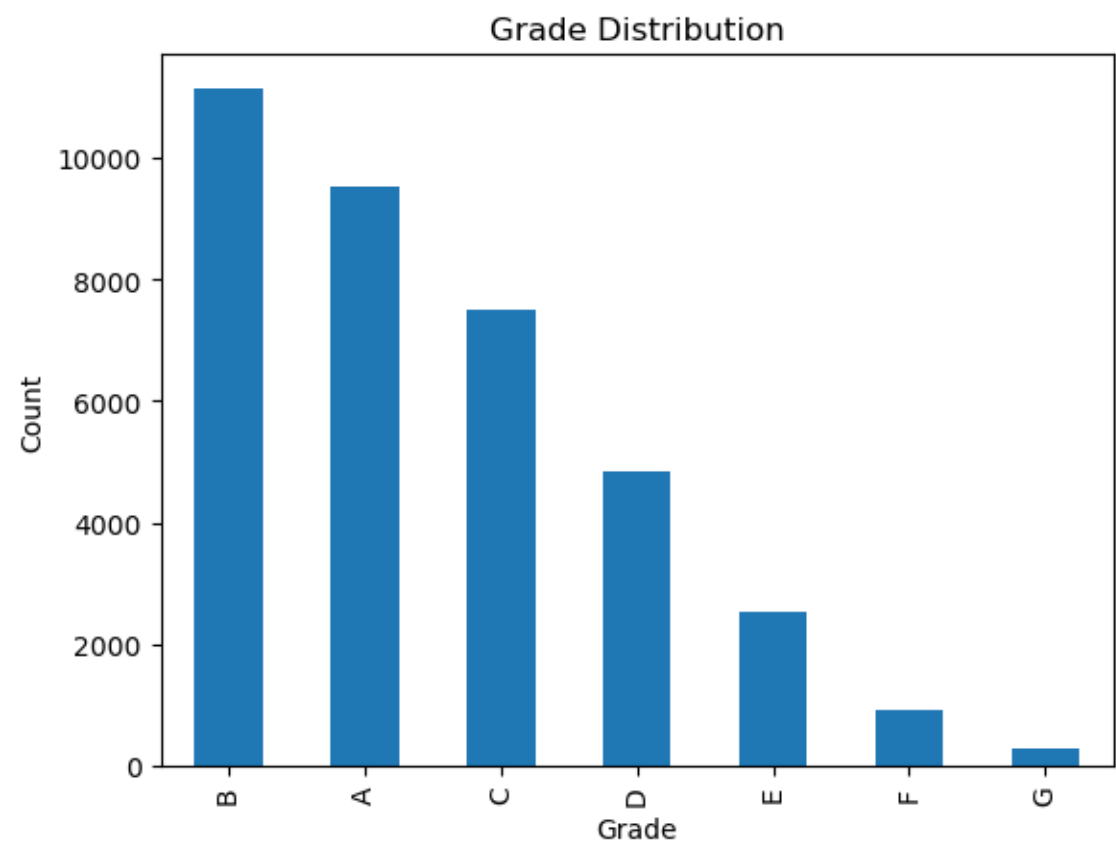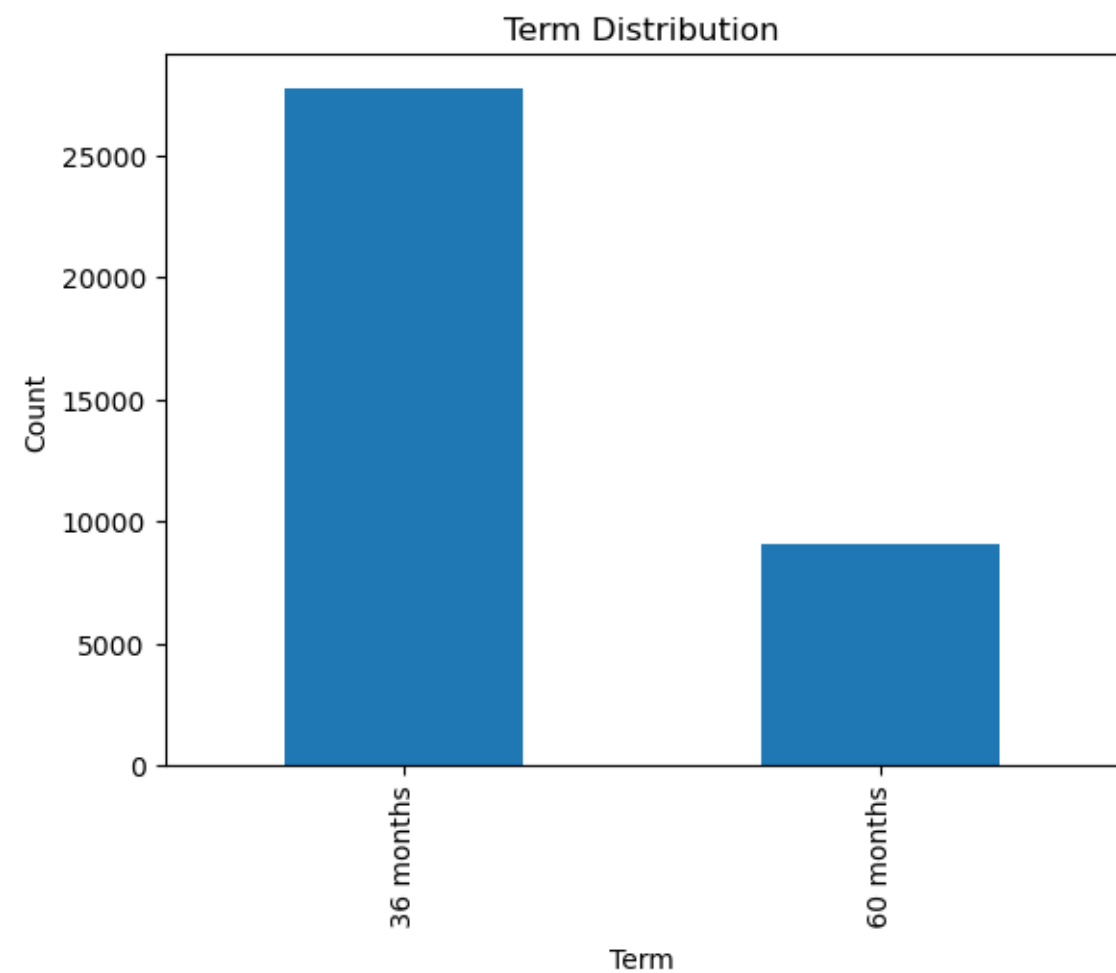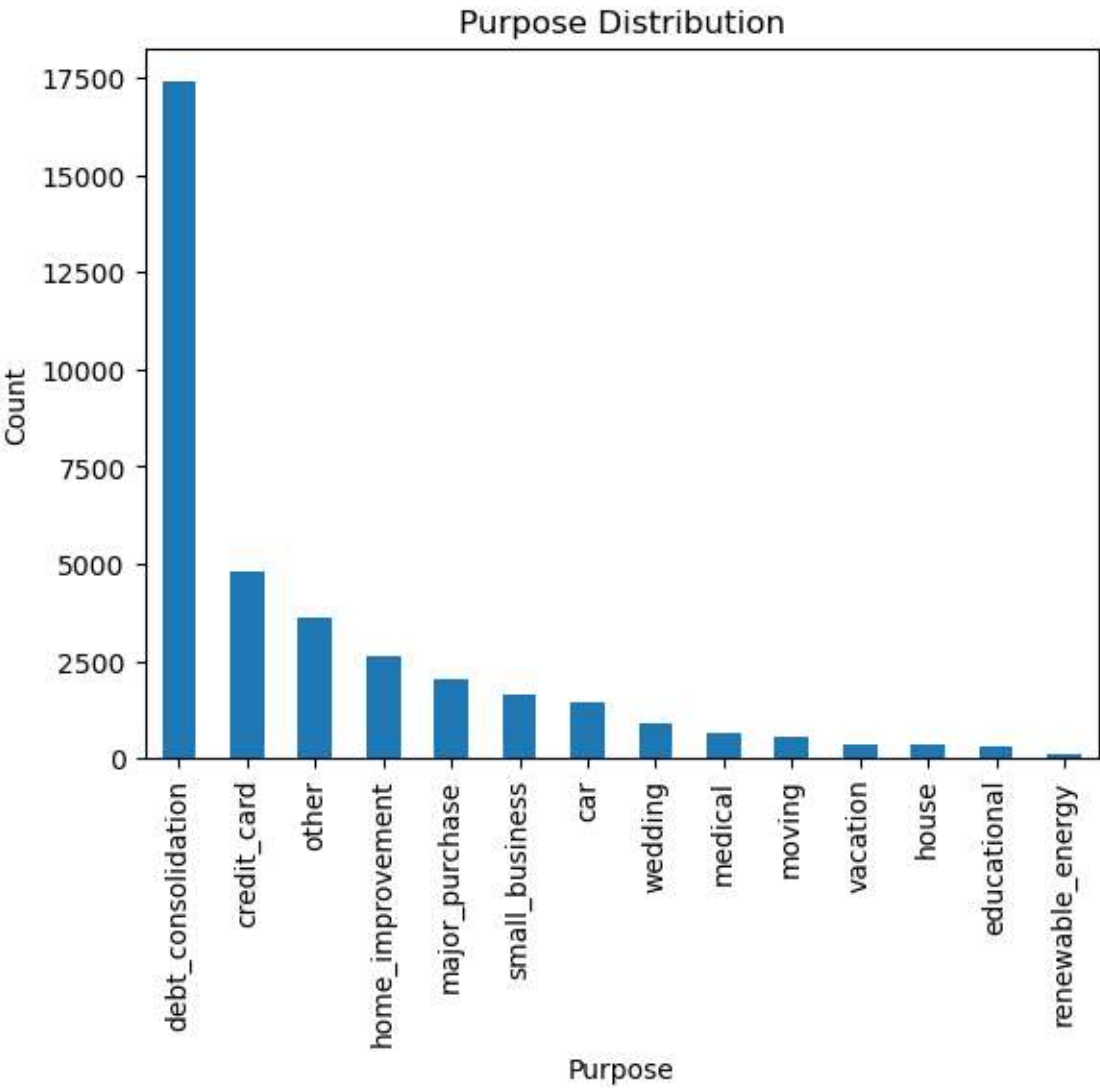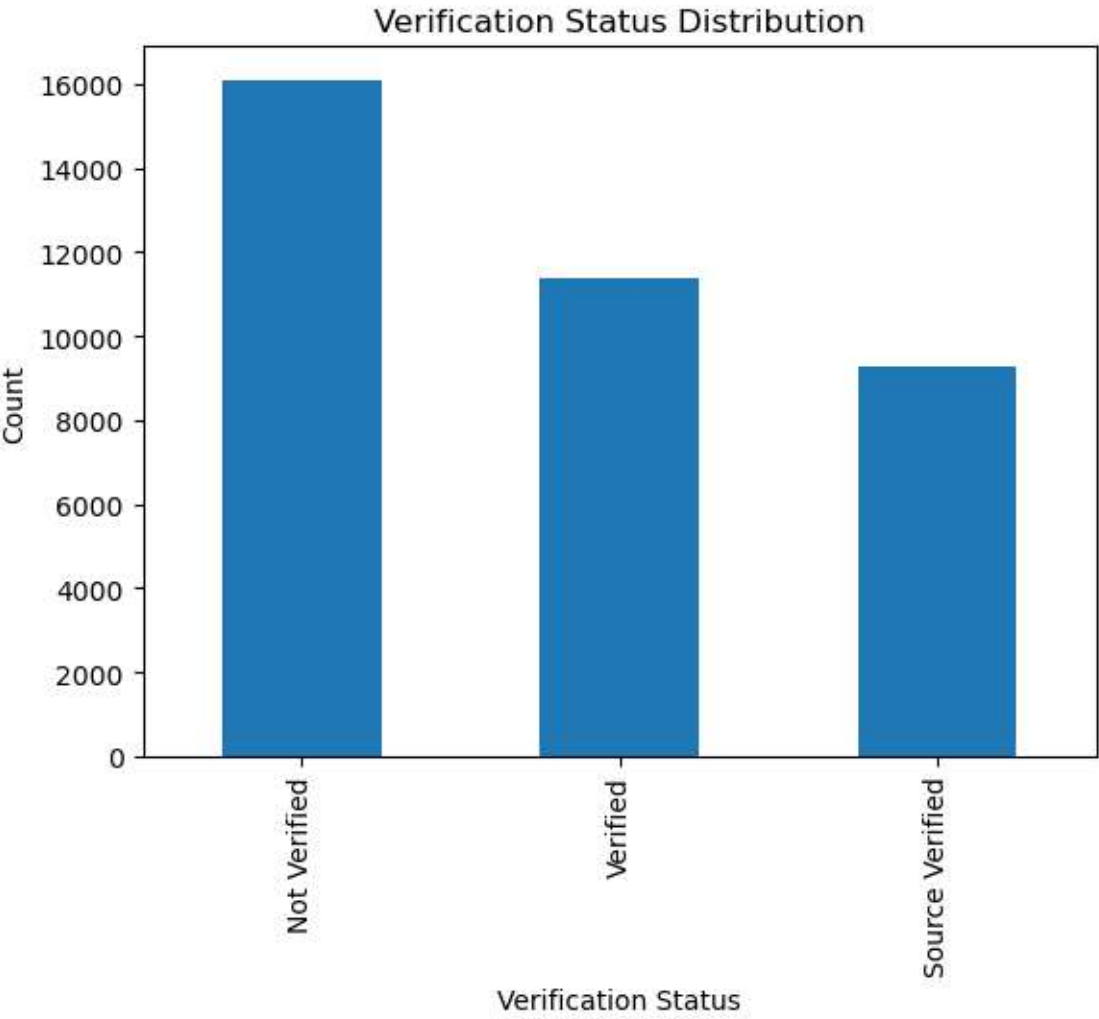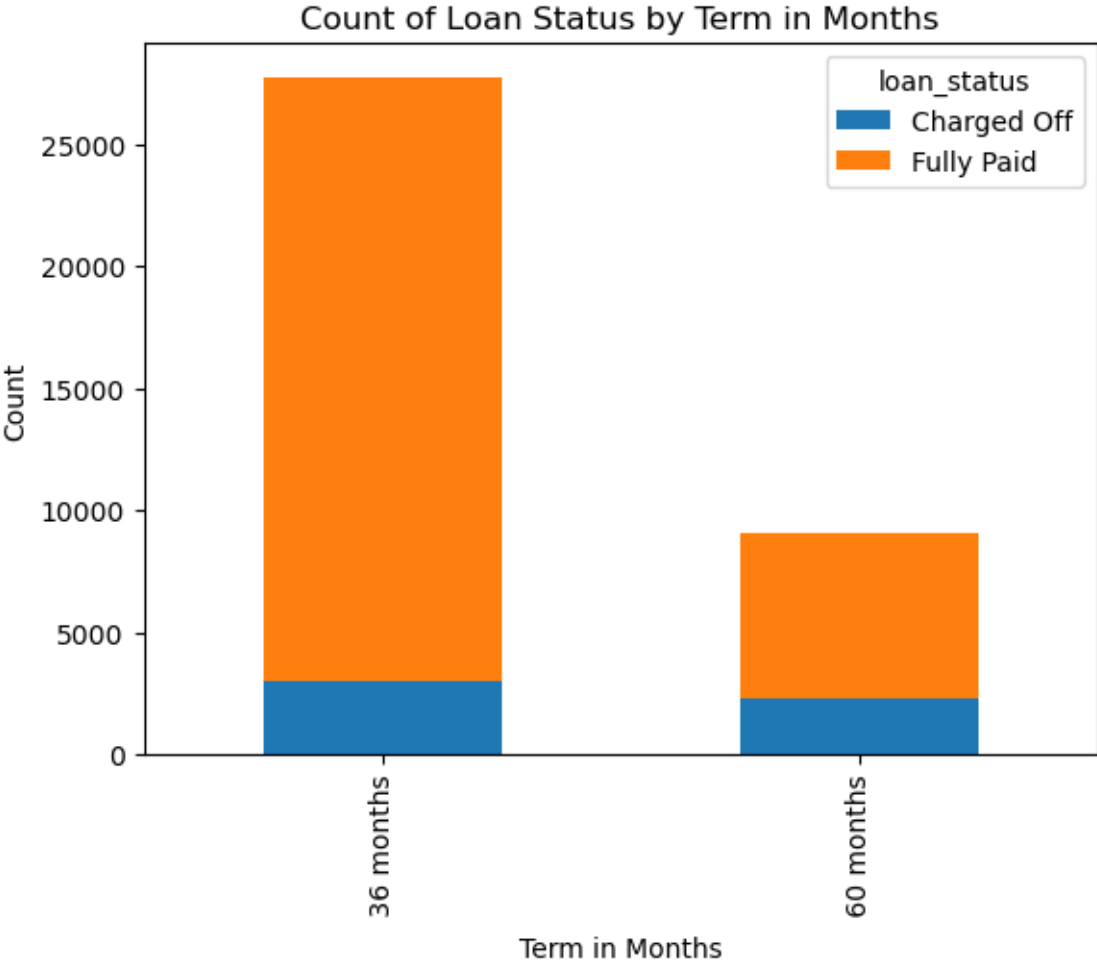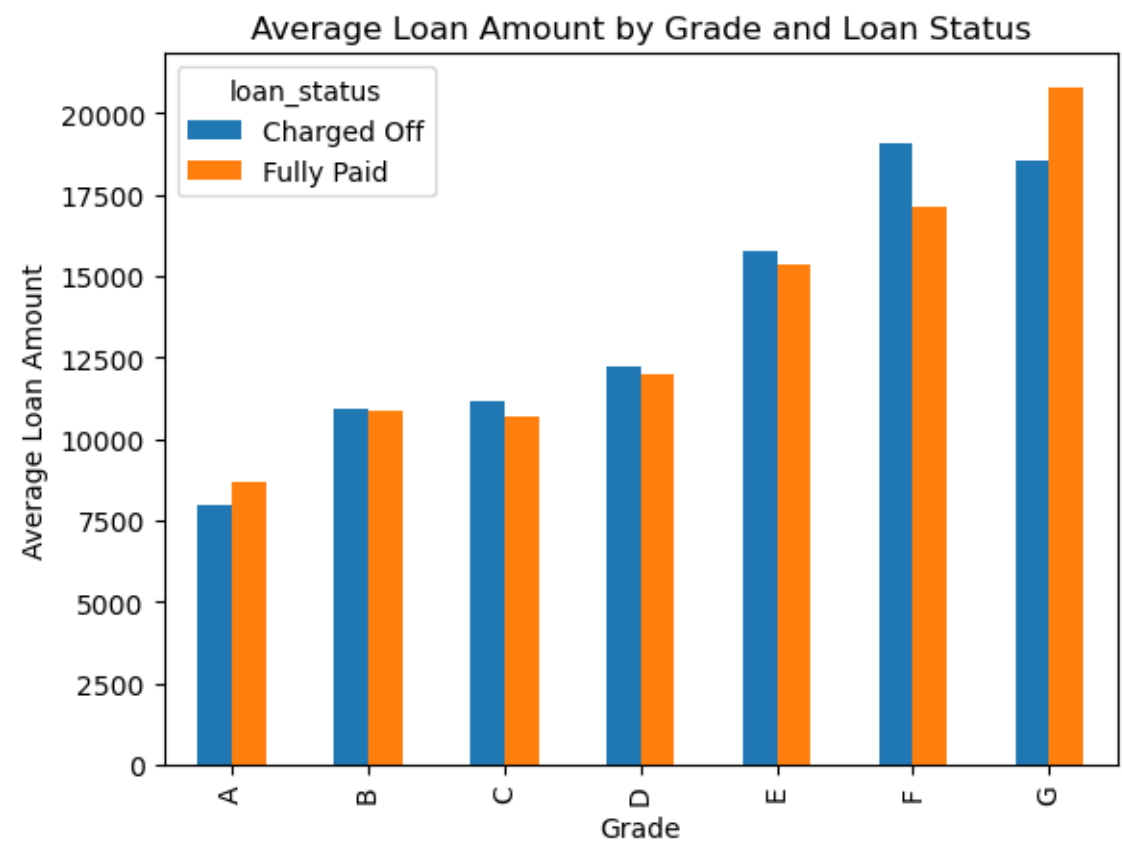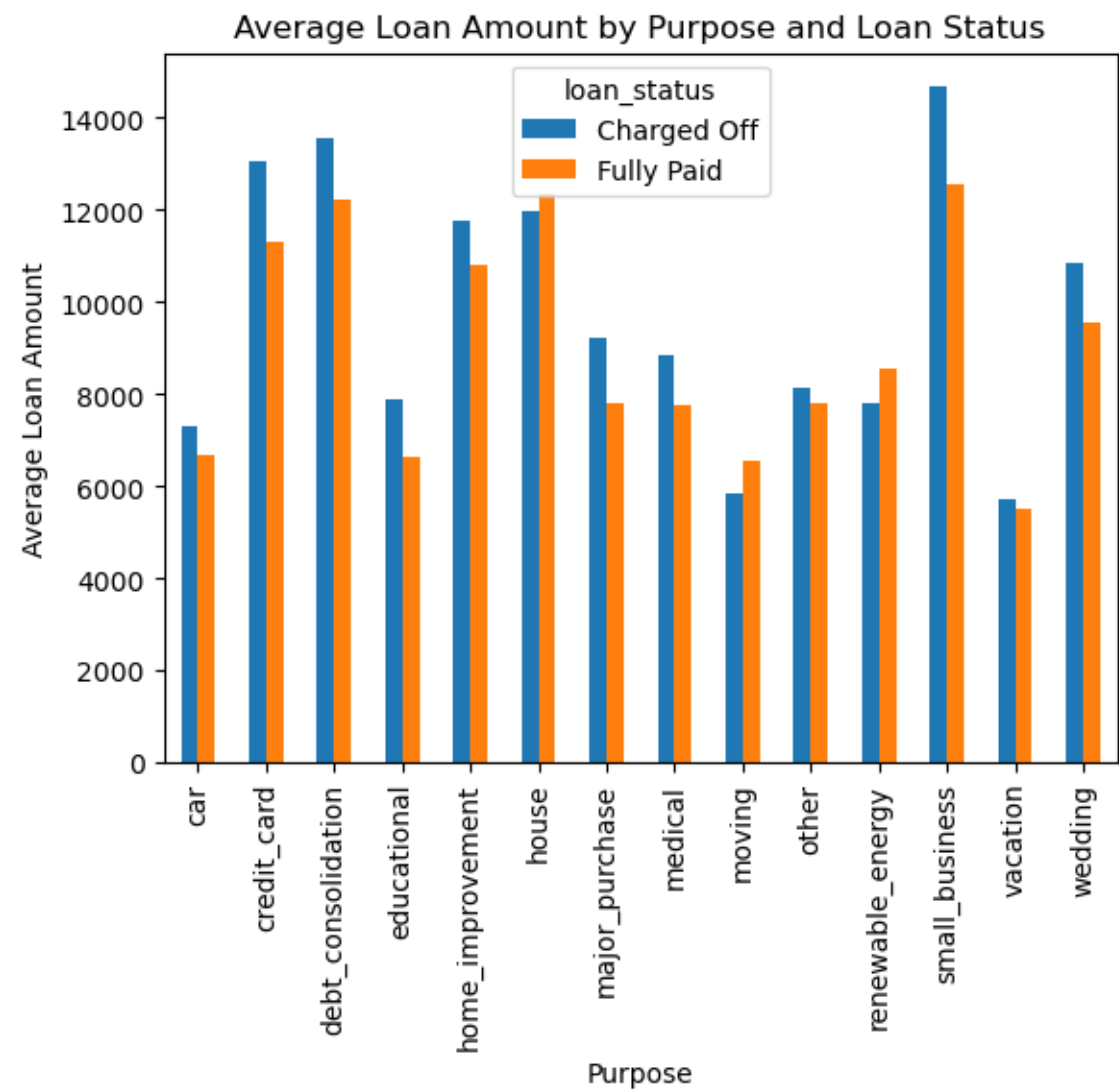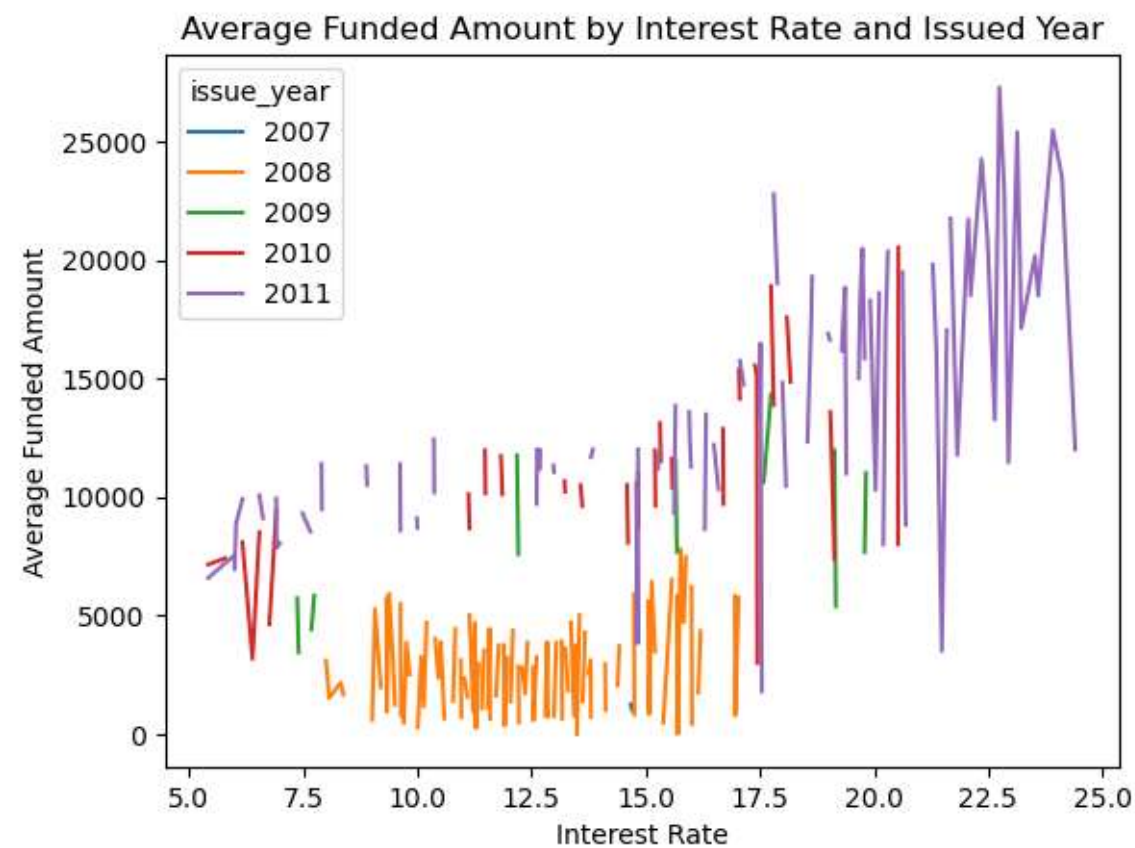
Univariate Analysis

Univariate Analysis
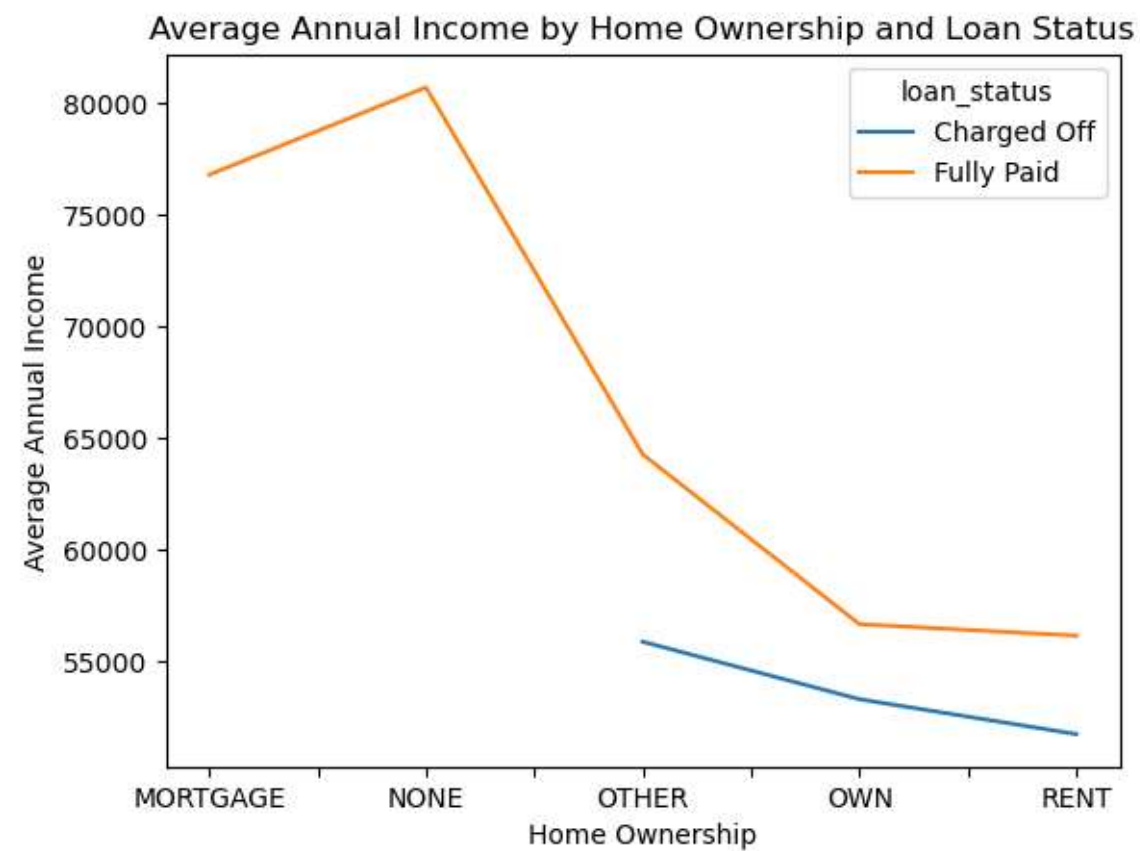
Univariate Analysis

Univariate Analysis
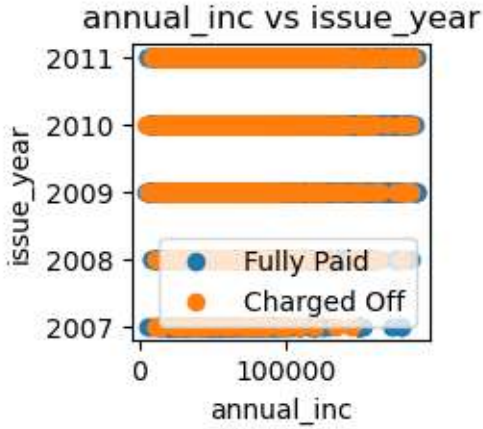
Bivariate Analysis
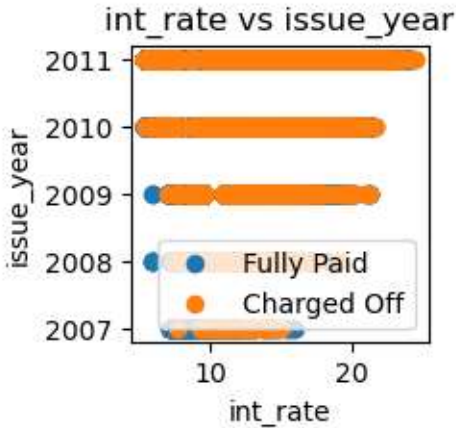

Count of Loan Status by Term in Months

Bivariate Analysis

Bivariate Analysis

Average Loan Amount by Purpose and Loan Status

Bivariate Analysis

Average Funded Amount by Interest Rate and Issued Year

Bivariate Analysis

Average Annual Income by Home Ownership and Loan Status
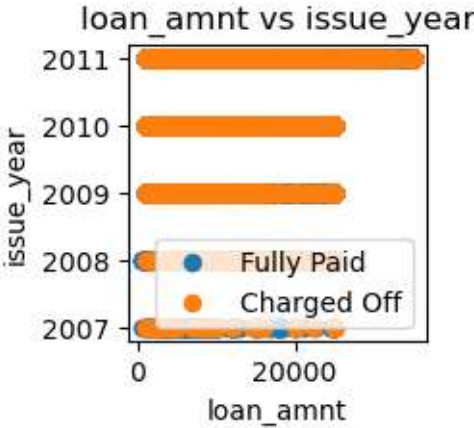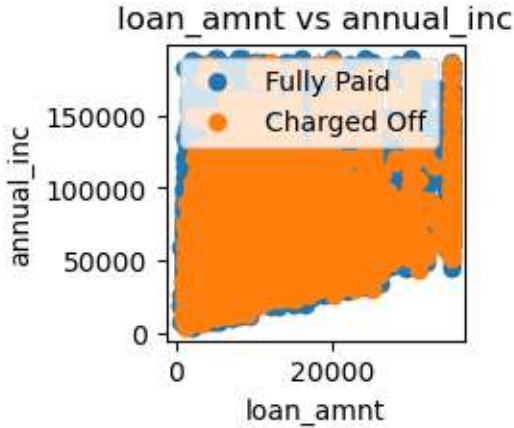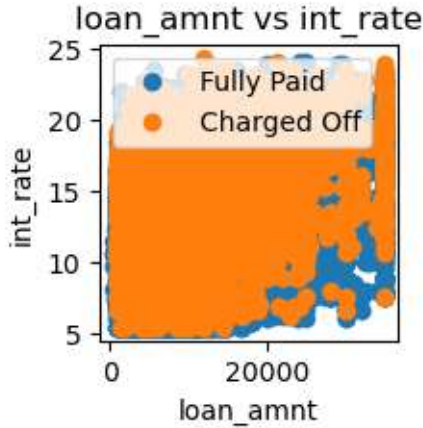
Multivariate Analysis

Observation

➢ People who rent and mortgage are more prone to defaulting than those who own a house.

➢ The loans provided at the later stages of months of the year are prone to defaulting.

➢ Loans provided for short term is more prone to defaulting than long term.

➢ Loads provided to G Grade is prone to defaulting.

➢ Loans provided to non verified cases are more prone to defaulting.

➢ Loans taken for debt consolidation is prone to defaulting.

➢ The annual income being less than 60K and those who stay in rented house are prone to defaulting.

➢ The defaulted loan amount and interest rate has been increasing over years, can try reducing interest to defaulting.

Observation

➢ Small Business are more prone to defaulting.

➢ Loan amount less than 20000 and interest rate less than 15% is mostly fully paid.

➢ Annual income more than 150000 and amount less than 20000 has been mostly dully paid.

➢ Loan amount charged off has increased a lot in 2011 along with interest higher than 15% .

➢ Loans given to people with salary less than 100000 is prone to defaulting over years.