



Lending Club Case Study

Submitted by: Leena Barua



Problem Statement

- Lending Club is a consumer finance company which specializes in lending various types of loans to customers.
- If the company approves the loan for an applicant who is not likely to repay the loan, it will cause a financial loss for the company.
- However, if the company doesn't approve the loan for an applicant who is likely to repay the loan, it will result in a loss of business.
- As such it is vital to identify risky loan applicants



Objective

The case study aims to identify the driving factors behind loan default, i.e. the variables which are strong indicators of default and thus identify risky loan applicants.

Data

We have been given complete loan data for all approved loans issued through the time period 2007 to 2011.



Data Cleaning: Part 1

Columns:

- Discard columns having high percentage of missing values (>40%) as they can skew the analysis
- Discard columns having only 1 unique value as they contribute nothing to analysis
- Discard columns that are not related to pre-approval of loan
- Discard columns containing too many unique values in texts like title, emp_title, url
- Discard funded_amnt_inv as funded_amnt is already present

Rows:

- Discard rows with loan status as current as they contribute nothing to analysis



Data Cleaning: Part 2

- **Treat missing values:**

1. By imputation (if there are many)
2. By dropping the rows (if they are too few)

- **Datatype correction:**

1. Correct data types for the columns
2. Analyze columns with object datatype and lots of unique values

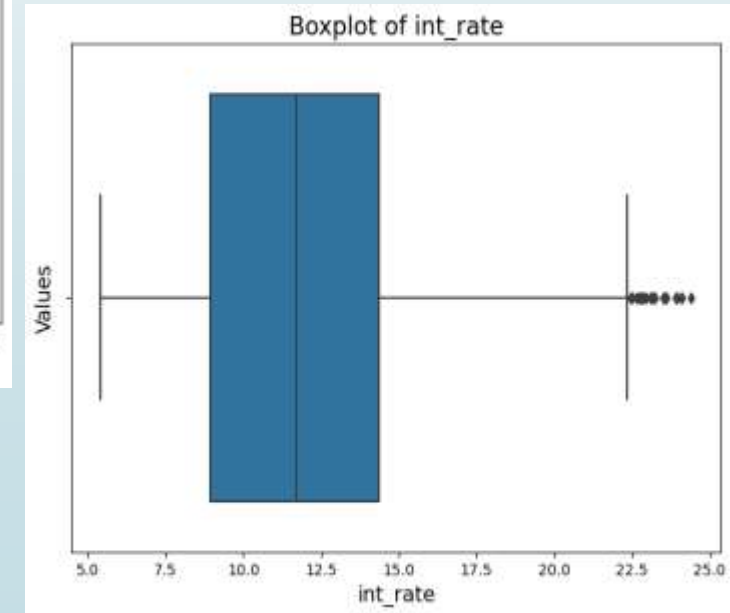
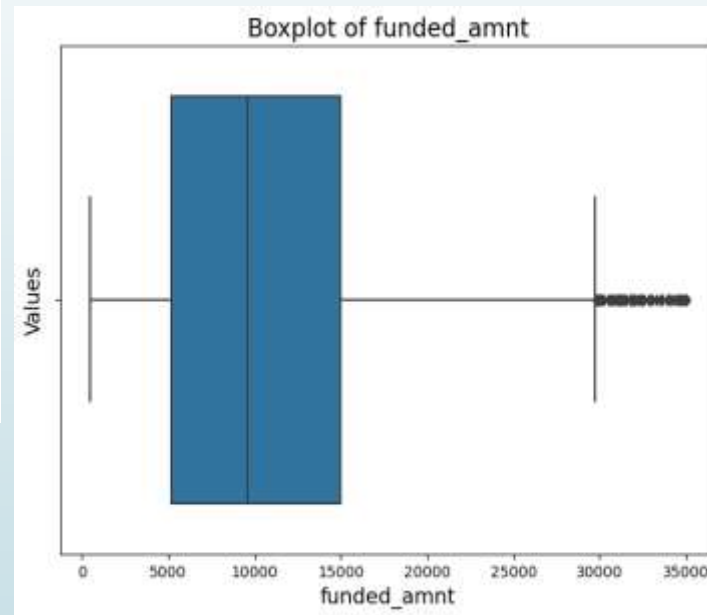
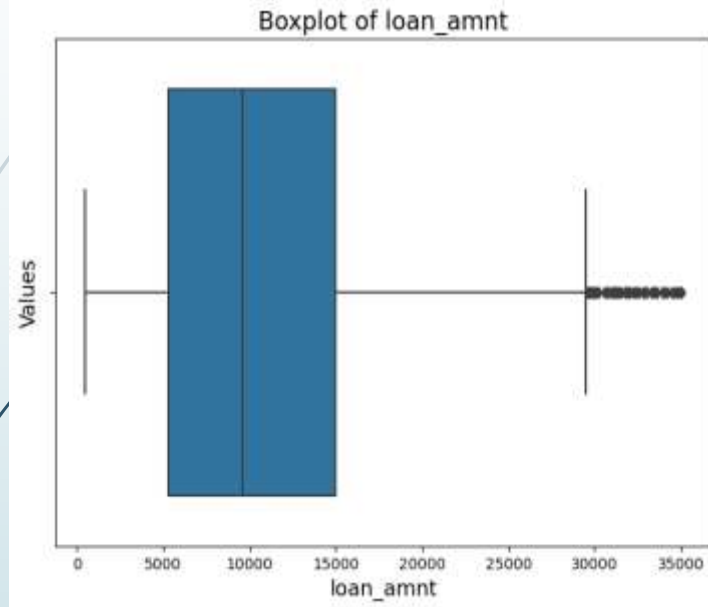
- **Column Data Correction**

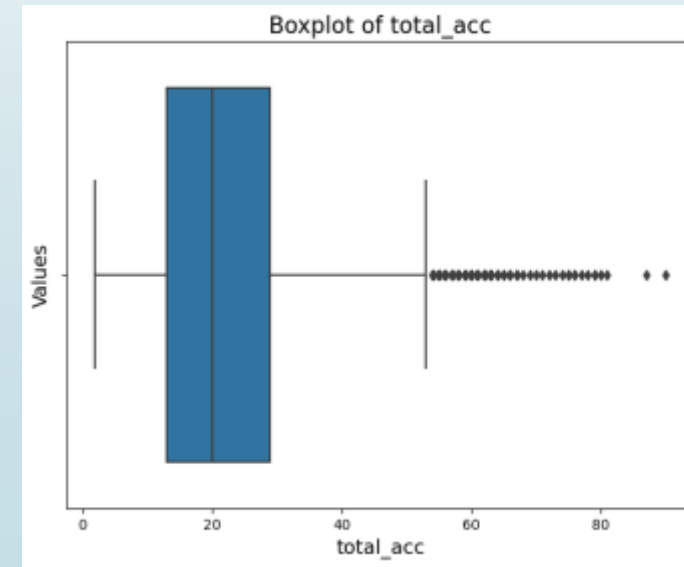
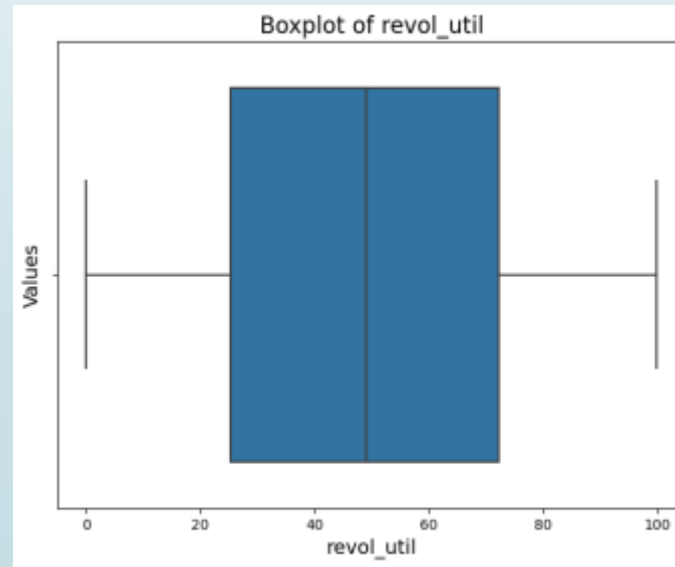
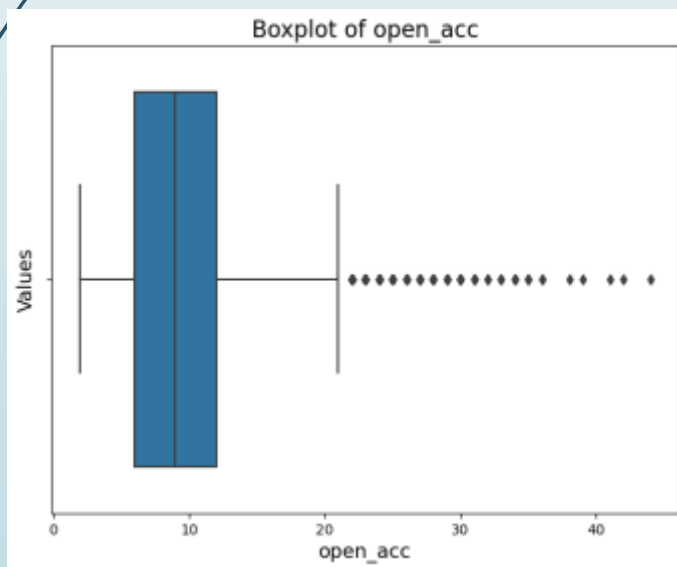
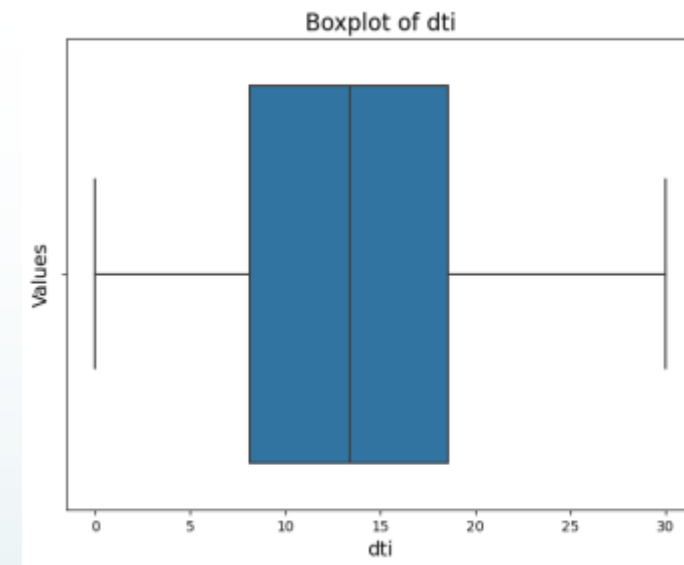
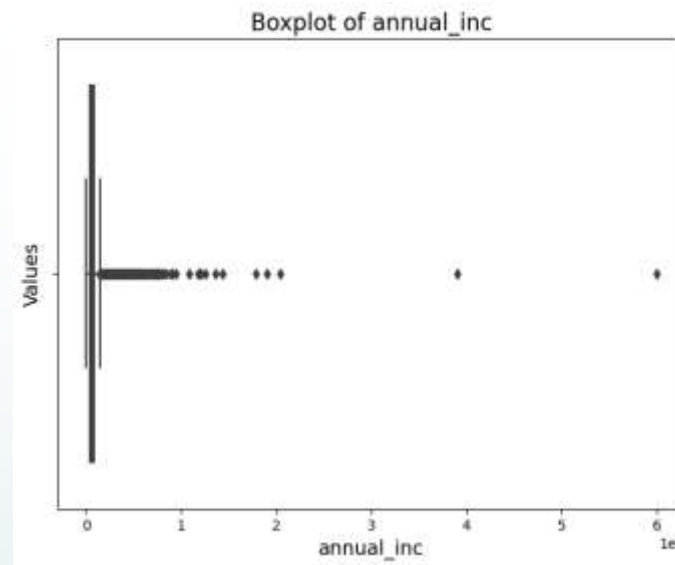
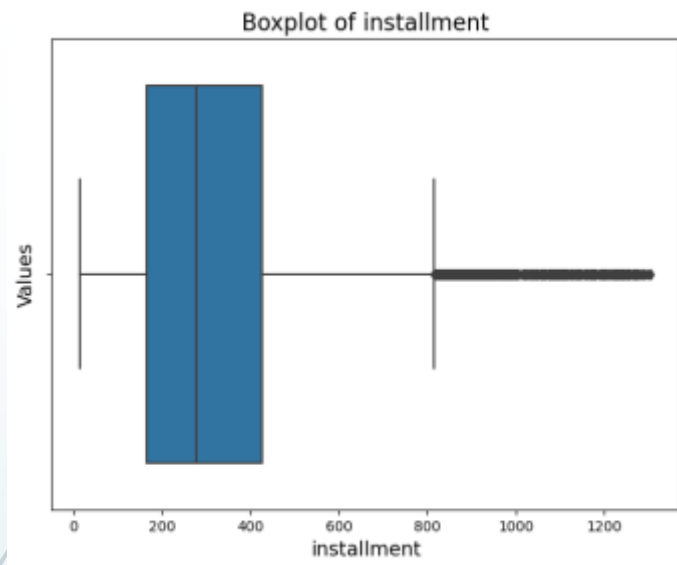
1. Replace 'Source Verified' values with 'Verified' as both have the same meaning

- **Separate Numerical and Categorical columns**

Outlier Analysis

Create boxplots for numerical columns





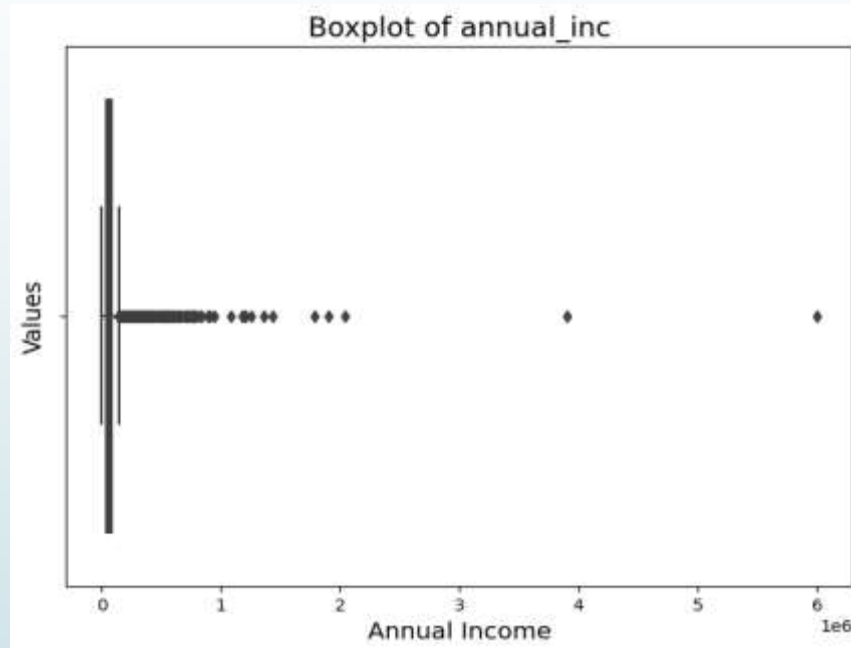


Analysis Results:

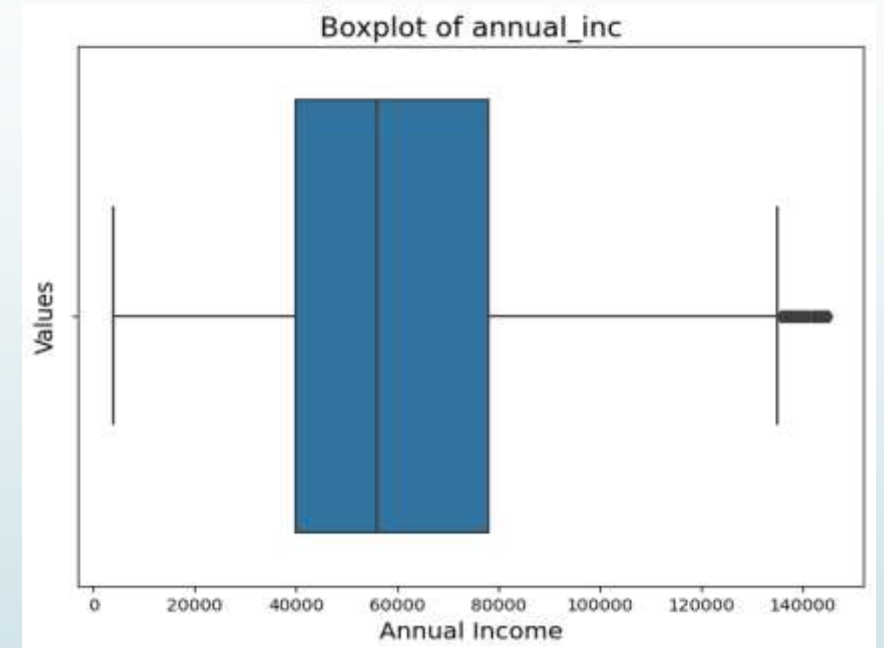
- Loan amount of most of the loan applicants is between 5000 to 15000 USD
- The interest rate on the loan is mostly between 8 percent to 14 percent
- The monthly installment amount on the loan is mostly between 160 to 450
- **Annual income has too many values outside the Inter Quartile range**
- The debt to income ratio is mostly between 8 to 18
- Number of open accounts are mostly between 5 to 12
- Revolving line utilization rate is mostly between 22 to 78
- Total accounts are mostly between 16 to 30

Outlier Treatment for Annual Income

Before filtering outliers



After filtering outliers

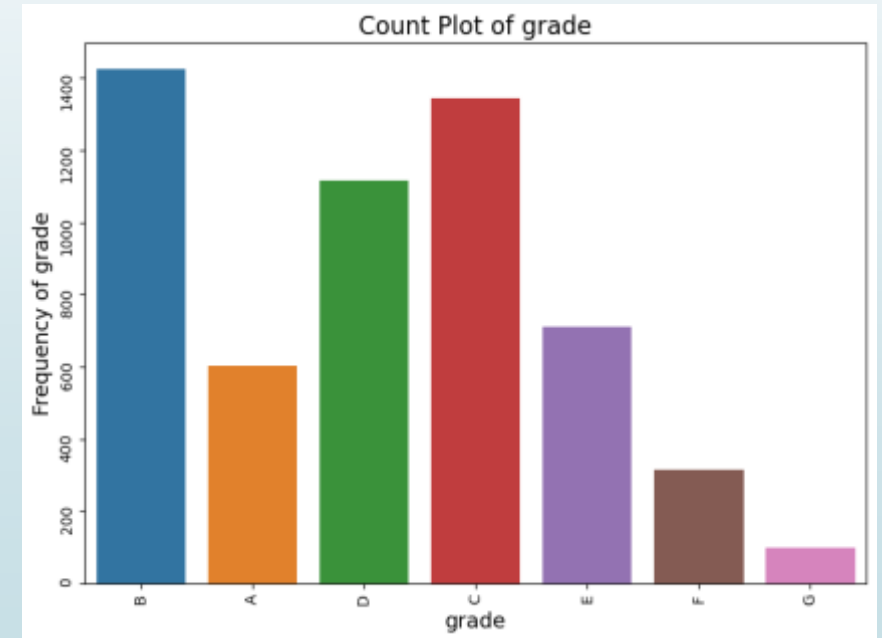
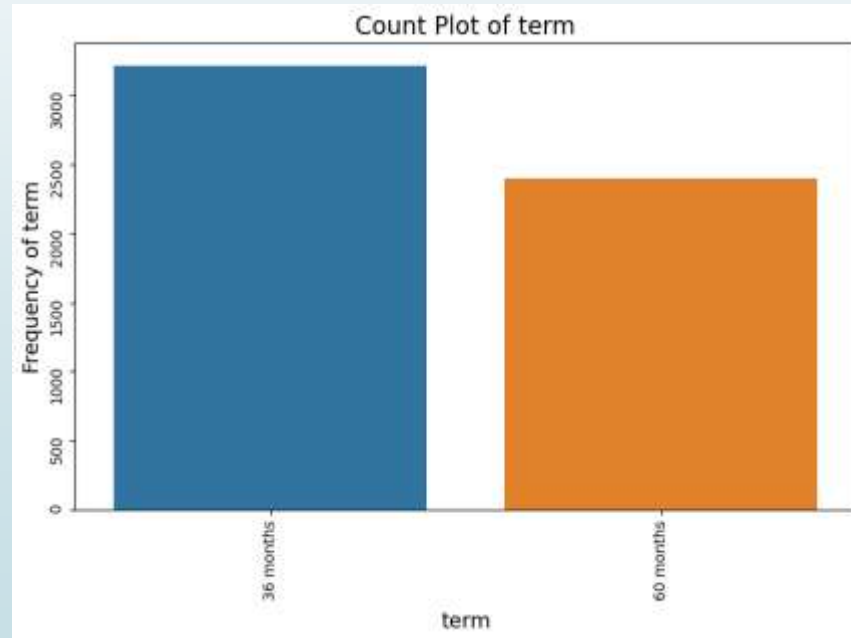


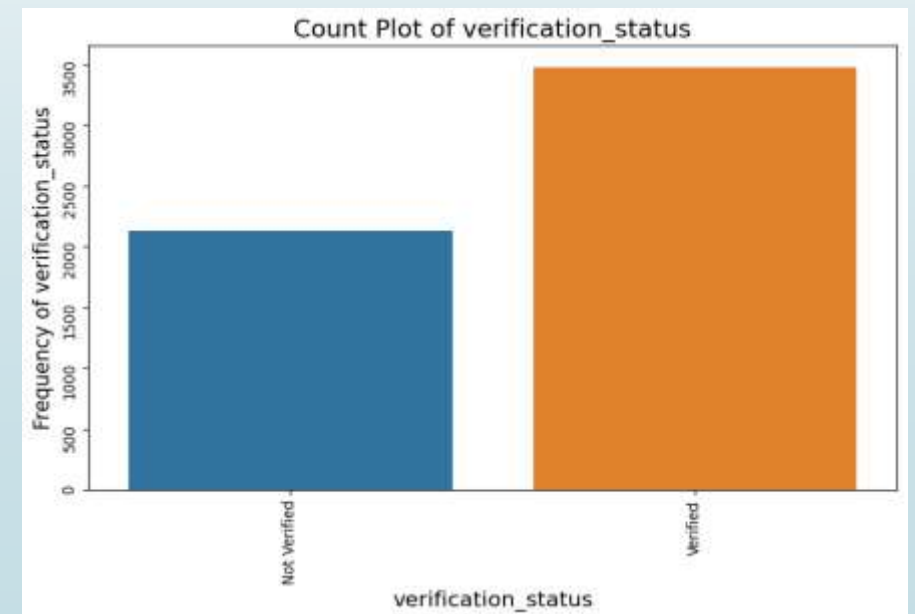
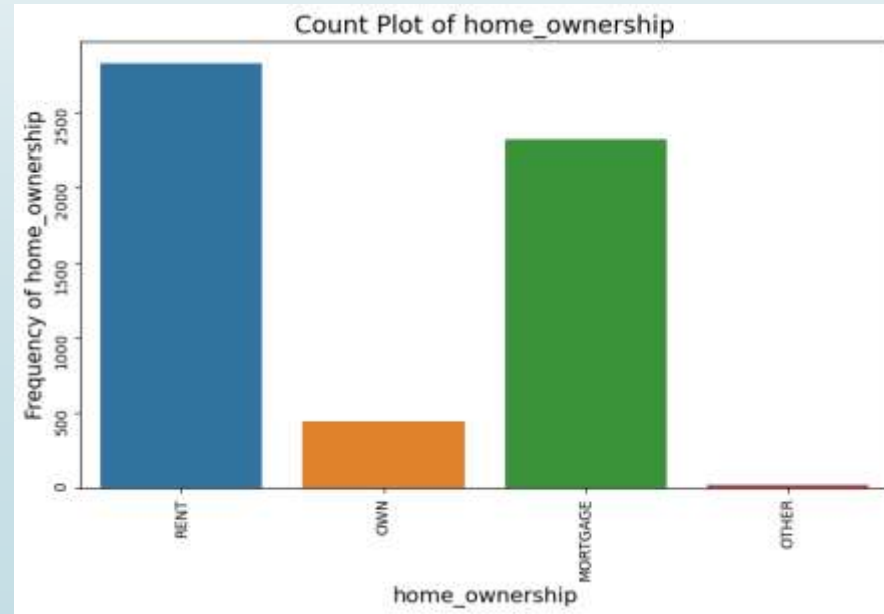
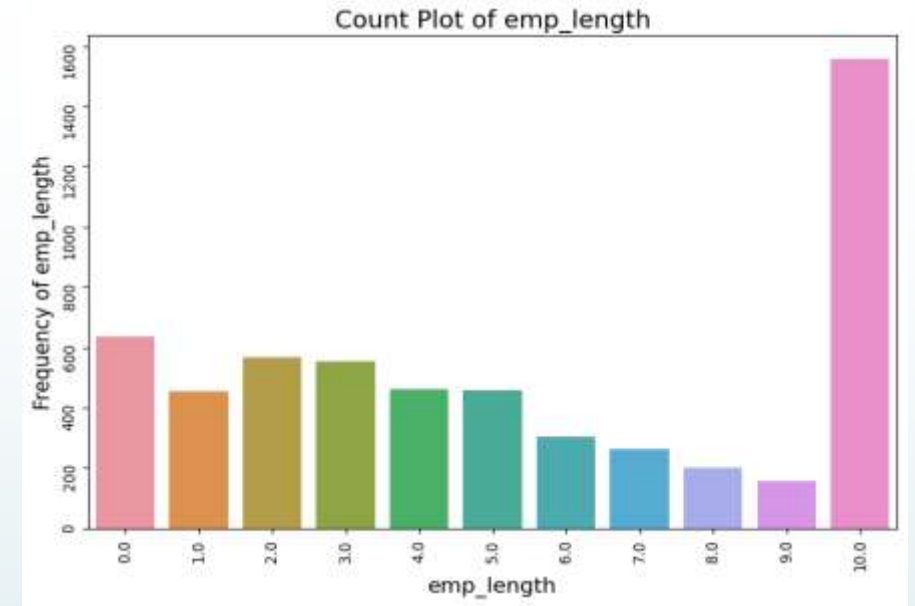
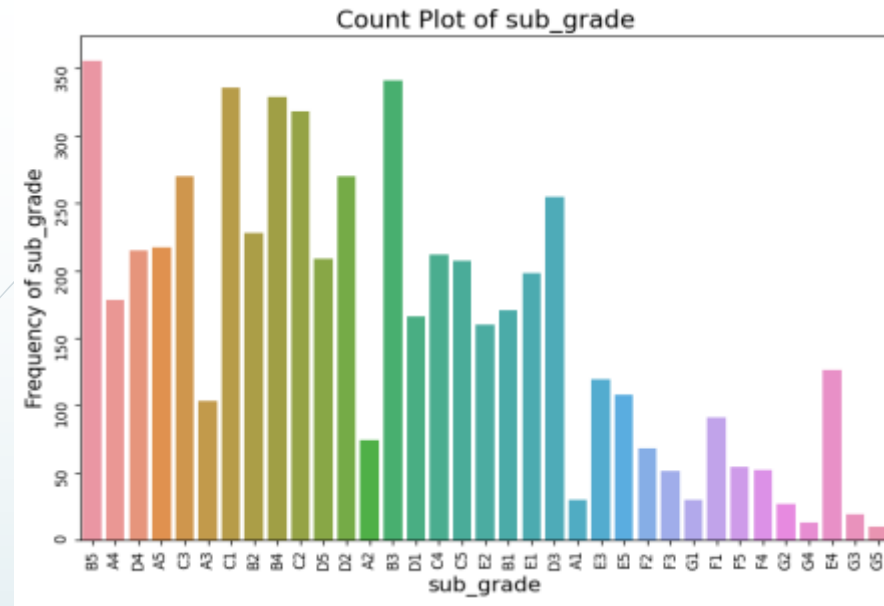
Analysis Results:

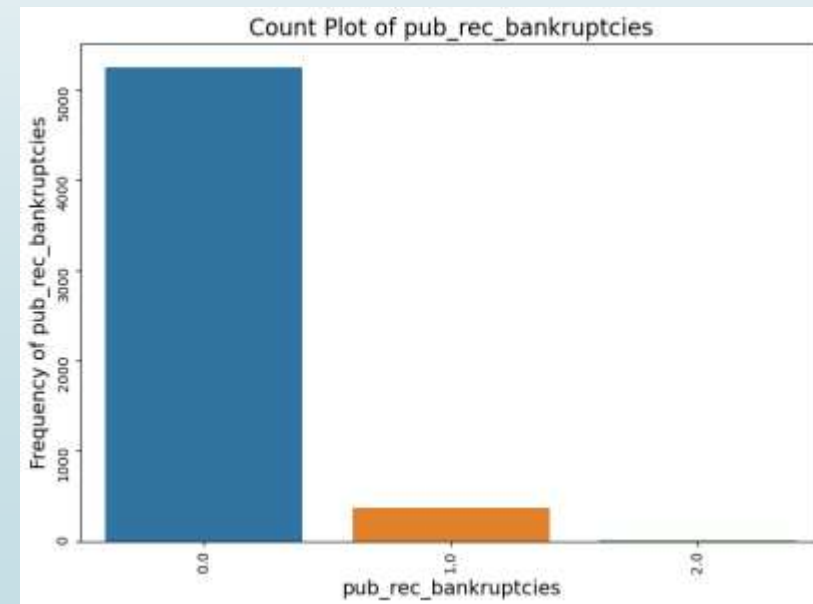
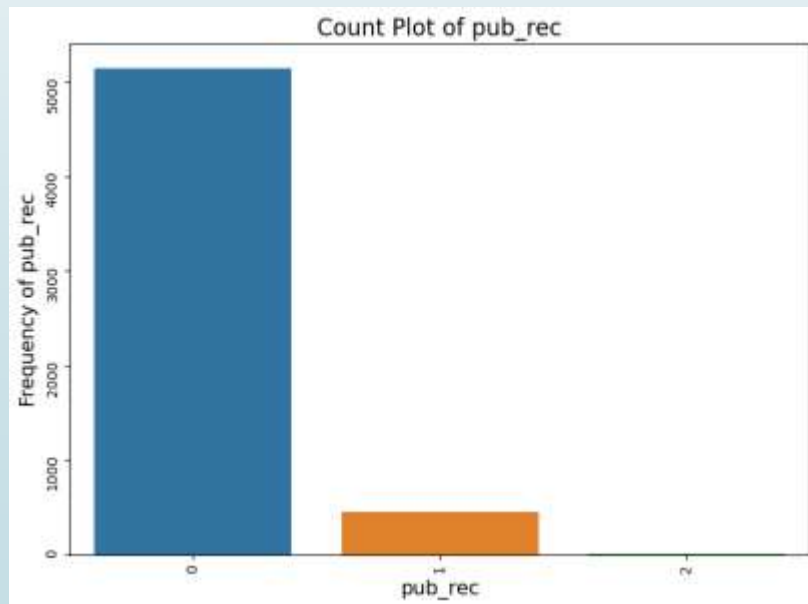
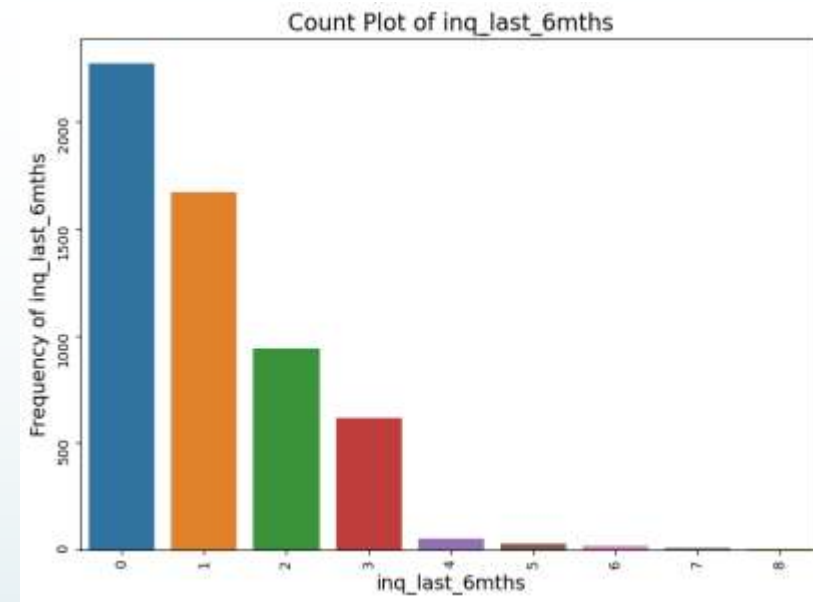
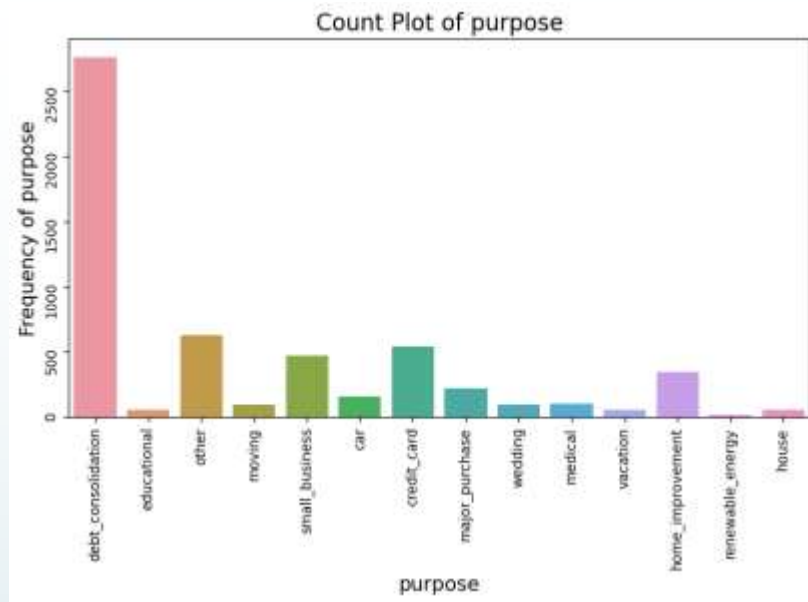
- Annual income of most of the loan applicants is between 40000 to 80000 USD
- Even though annual income has lots of values outside the Interquartile range, those values will still be considered for the analysis as Annual Income can logically have those values. We'll analyze them by binning later.

Segmented Univariate Analysis

- Create dataframe for the loan defaulters
- Plot countplots for the categorical columns in the loan defaulters dataframe









Analysis Results:

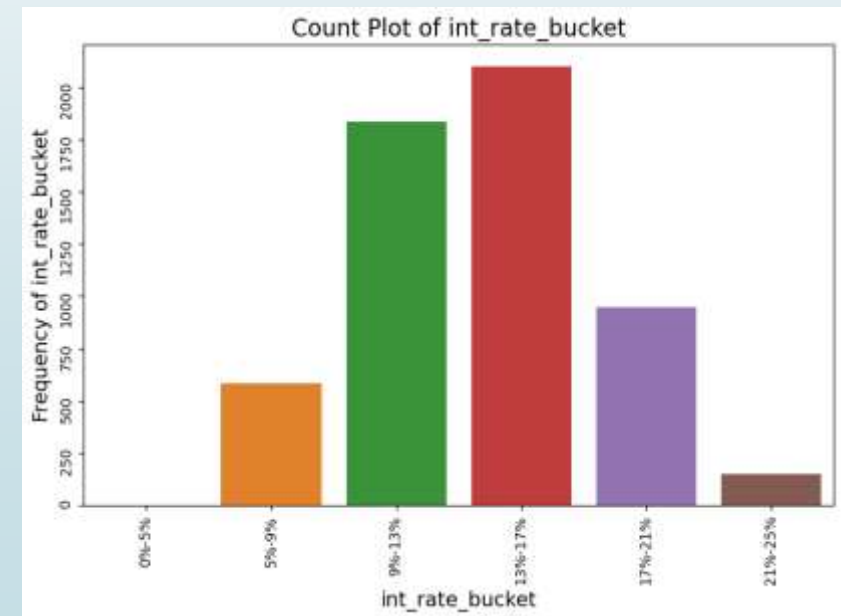
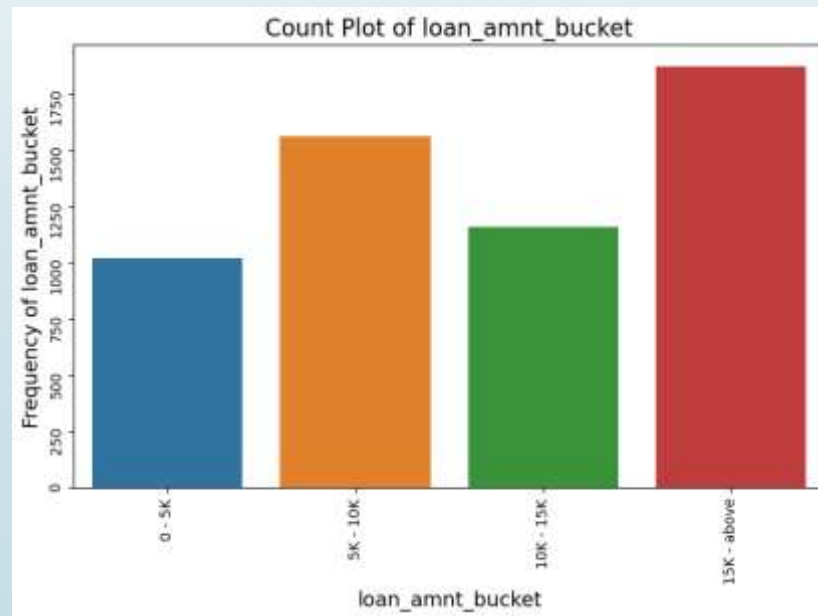
Most defaulters had the following characteristics

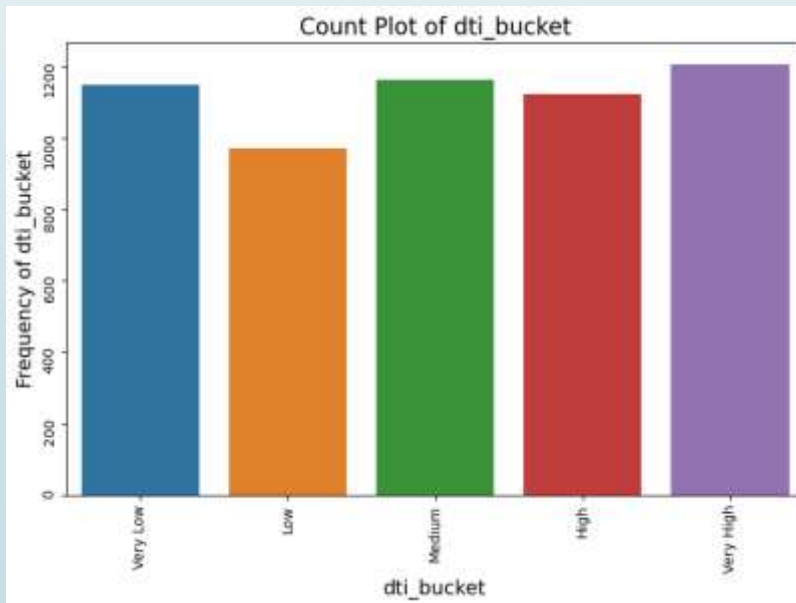
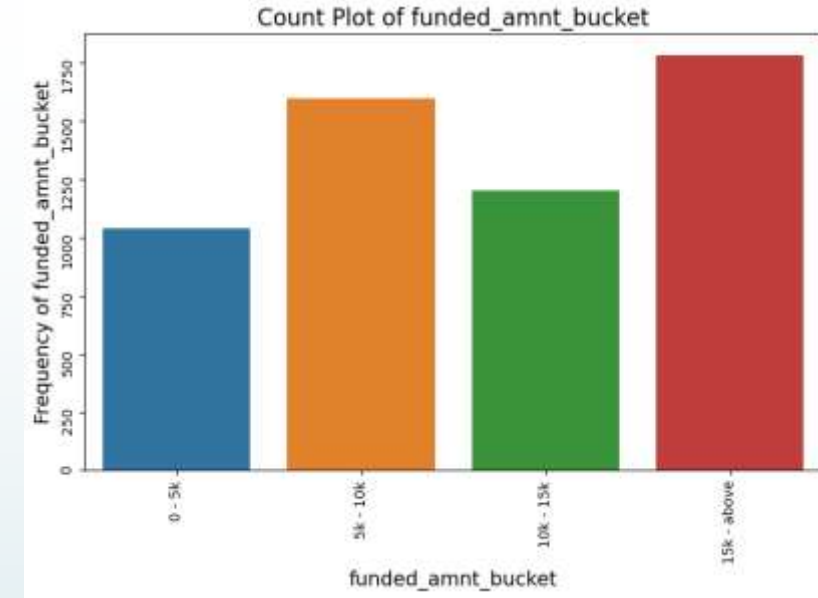
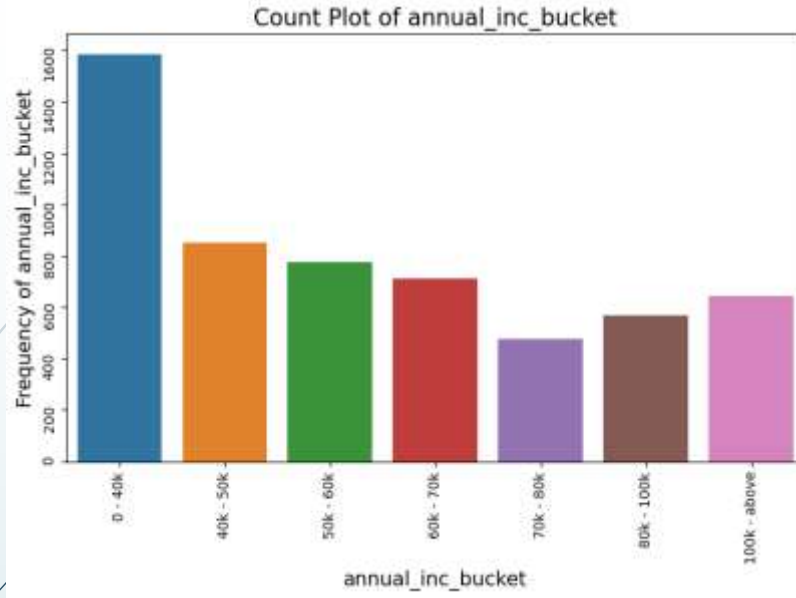
- Applied for Short-term loans with a duration of 36 months
- Loan grade was B or C
- Loan subgrade was B5, C1, B4 or B3
- Employed for more than 10 years
- Lived in rented houses
- Had verified income sources
- Made no inquiries in past 6 months(excluding auto and mortgage inquiries)
- Debt consolidation was the primary loan purpose
- No derogatory public records
- No public record bankruptcies

Binning numerical columns of Defaulter's Data

Group continuous numerical values into discrete categories called buckets

- Create a function to bucket Loan Amount, Interest Rate, Annual Income, total funded amount, Debt to Income
- Plot countplots for the buckets



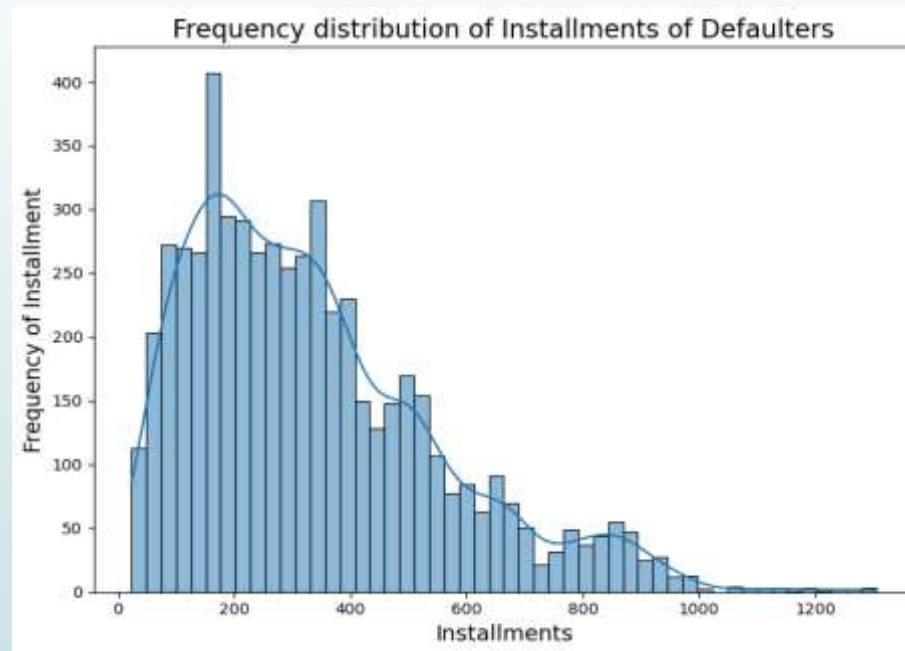


Analysis Results:

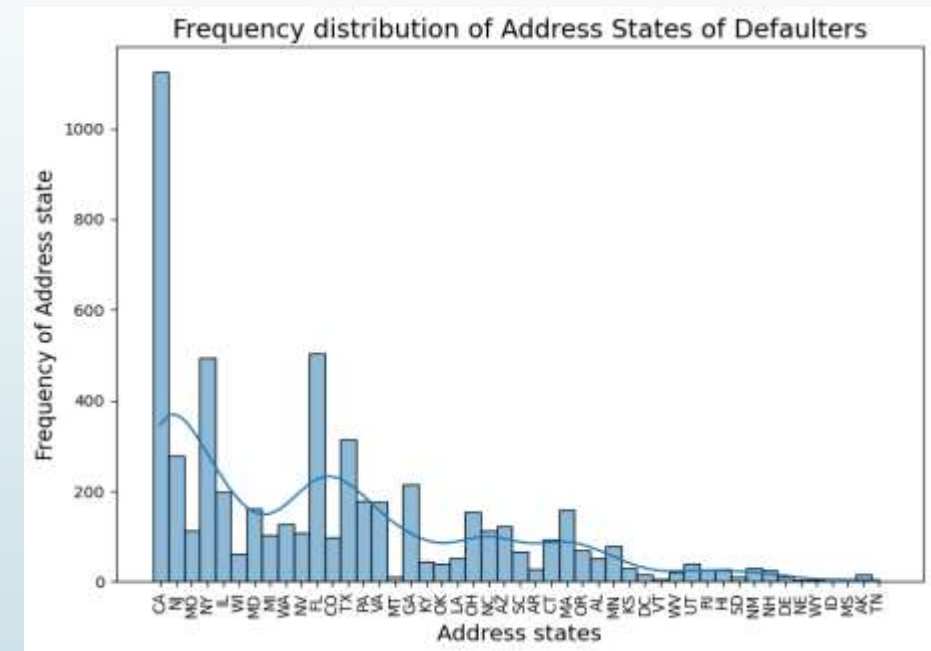
- Most defaulters had the following characteristics:
- Annual salaries less than 40,000 USD
- Interest rate between 13% to 17%
- Loan amounts of 15000 USD and above
- Amounts funded of 15,000 USD and above
- Very high debt-to-income ratios

Frequency Distribution of Defaulter's Installments and Address States

Installments



Address State



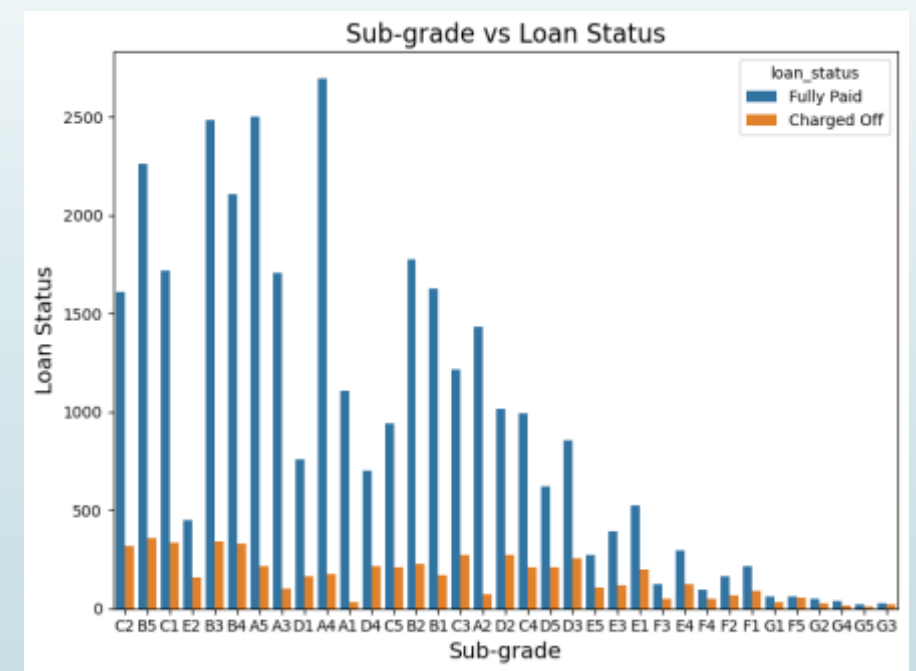
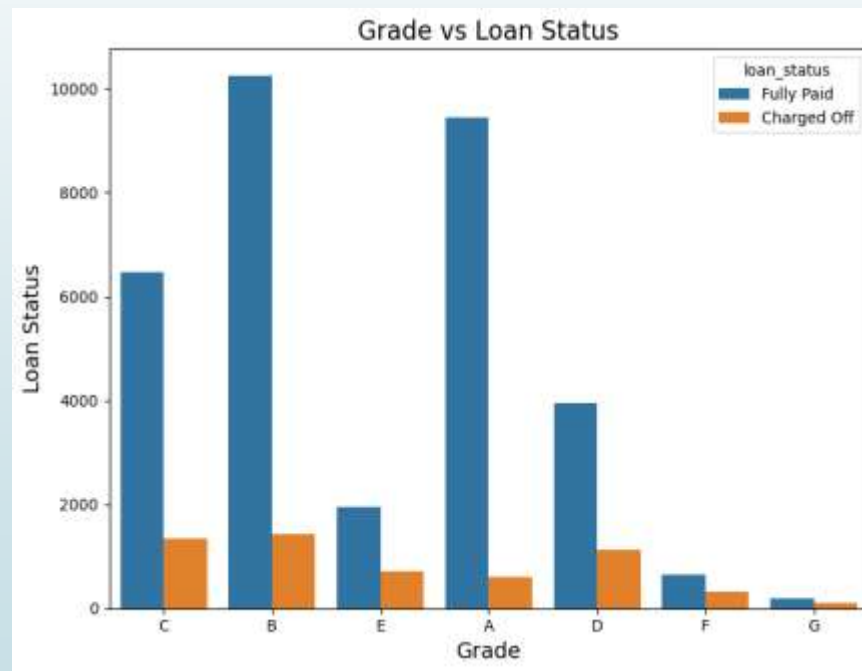
Analysis Results:

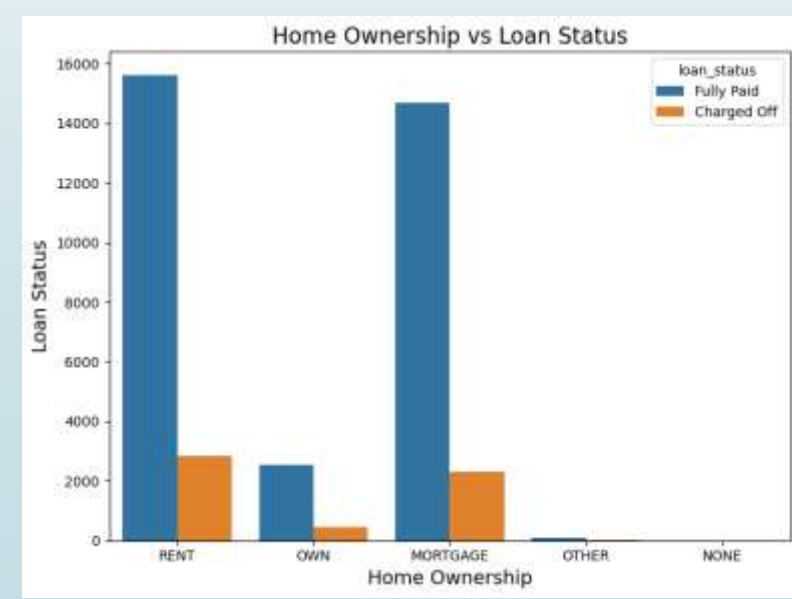
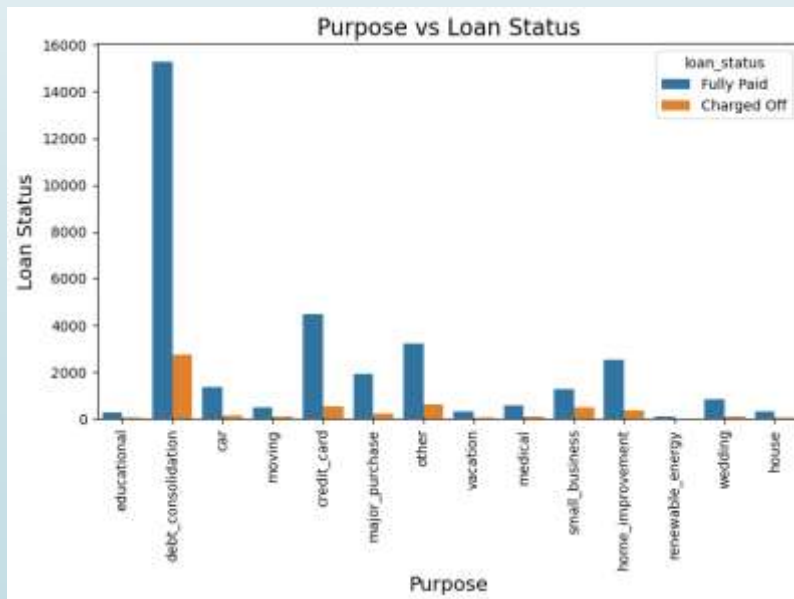
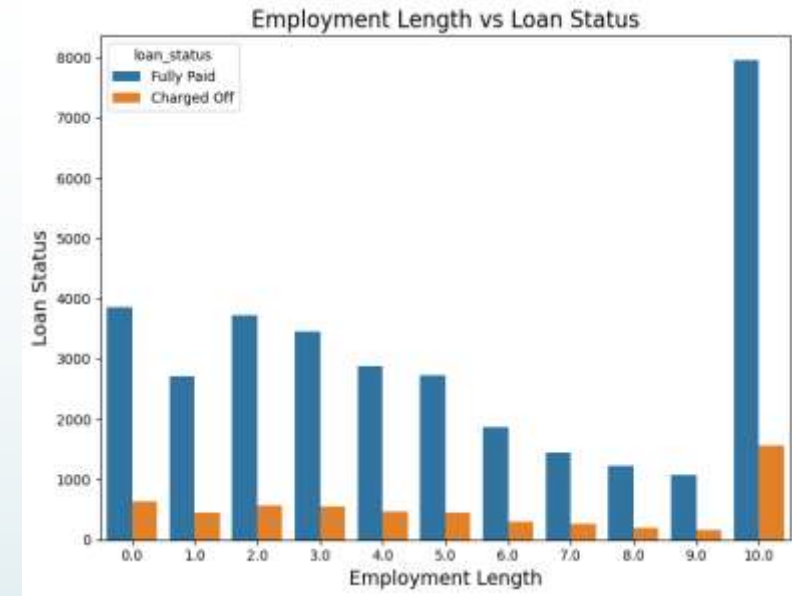
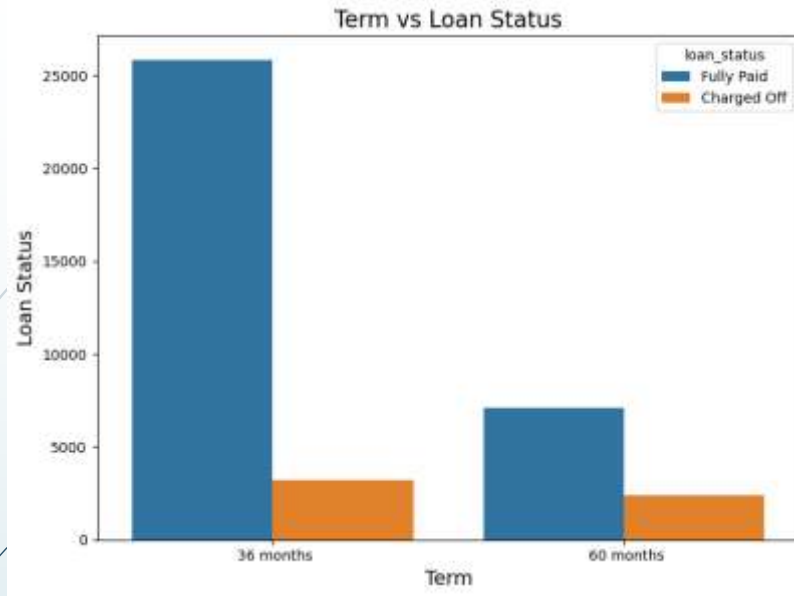
Most defaulters had:

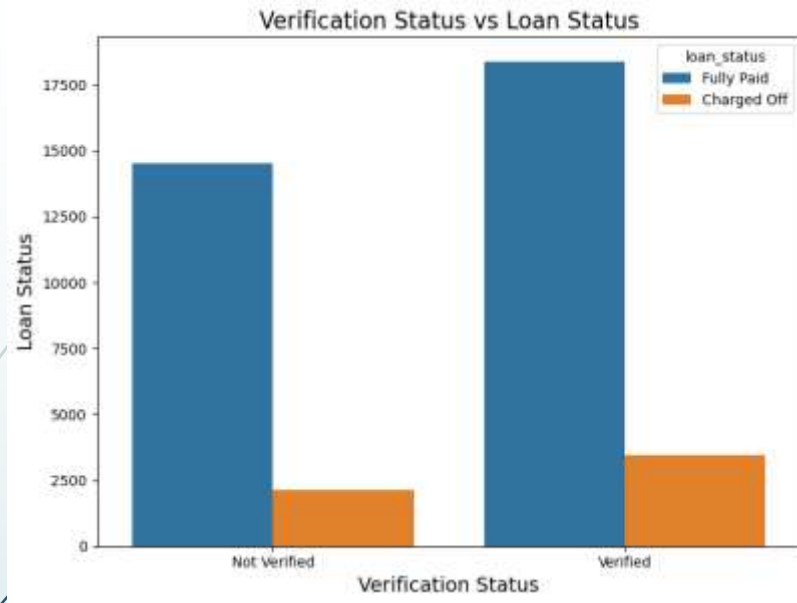
- Monthly installment amounts in the range of 160 to 440 USD
- Most lived in California, Florida or New York

Bivariate Analysis

Analyze categorical columns : Grade, Sub-Grade, Term, Employment Length, Purpose, Home Ownership, Verification Status against the Loan Status column by creating countplots.



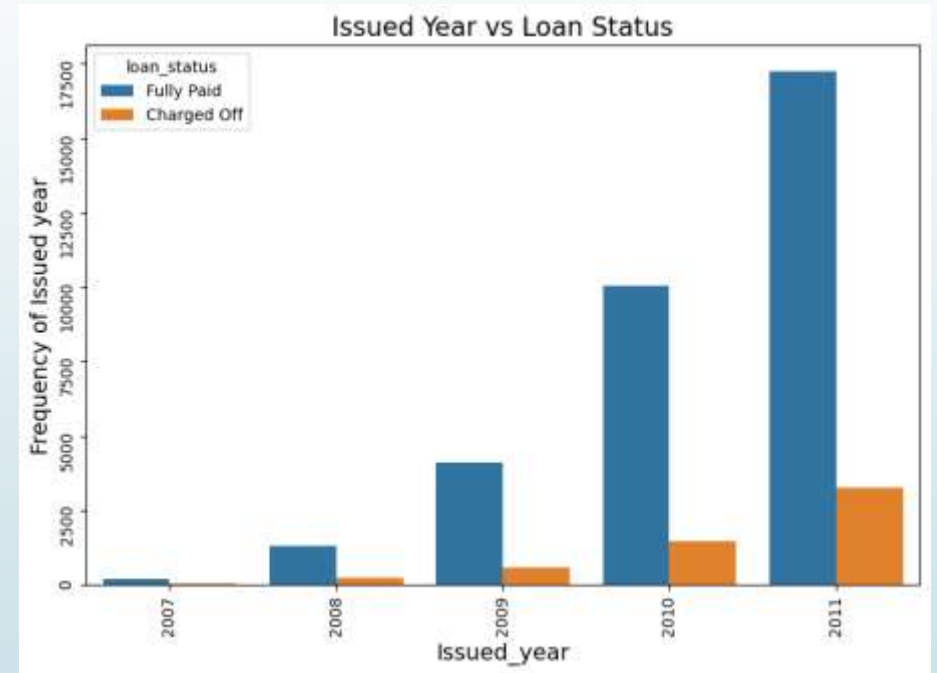
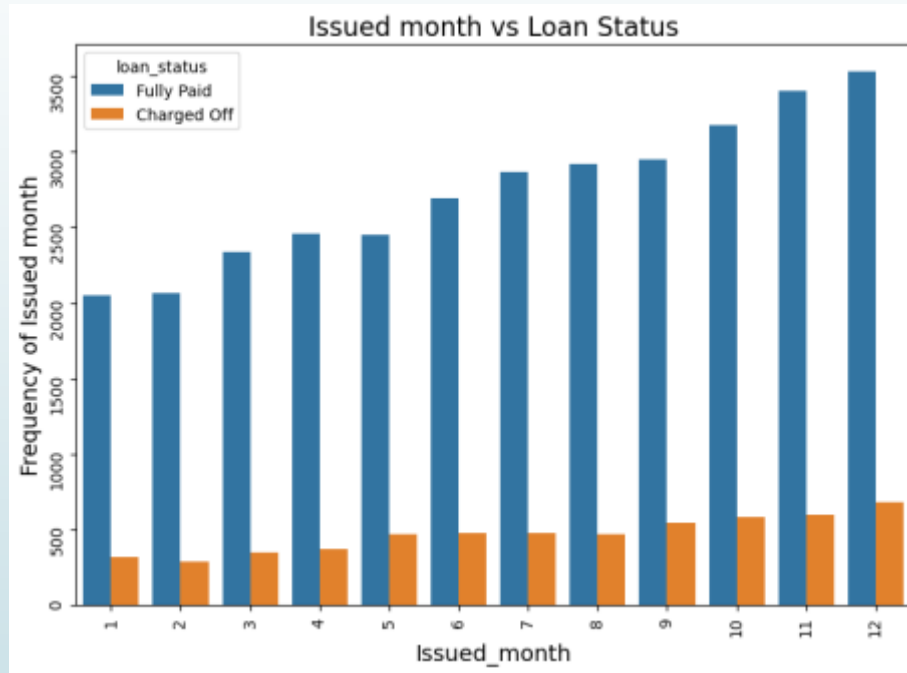




Analysis Results:

- Loan applicants of Grades B and C contribute most to "Charged Off" loans
- Loan applicants from Sub-grades B5, B3, C1 and B4 contribute most to "Charged Off" loans
- Short-term loans with a duration of 36 months contribute most to "Charged Off" loans
- Loan applicants with more than 10 years of experience are more likely to apply for loans as well as most likely to default
- Most defaulters have Debt consolidation as purpose of the loan
- Most defaulters lived in a rented or mortgaged house
- Most loan applicants had verified income sources and most defaulters also had verified income sources

Bivariate Analysis of Derived metrics from Date column



Analysis Results:

- The number of loans taken increased over the months, as did the number of defaulted loans.
- The number of loans taken increased from 2007 to 2011, as did the number of defaulted loans.



Summary

The company should be careful when giving loans of the following nature:

- Short-term loans with a duration of 36 months
- Loan grade B or C
- Loan subgrade B5, C1, B4 or B3
- Debt consolidation is the primary loan purpose
- Interest rate between 13% to 17%
- Applied for Loan amounts of 15000 USD and above
- Loans taken towards the end of the year

The company should be careful when giving loans to applicants with the following characteristics:

- Employed for more than 10 years
- Live in rented or mortgaged houses
- Made no inquiries in past 6 months(excluding auto and mortgage inquiries)
- With annual salary less than 40000 USD
- Live in California, Florida or New York
- Have very high debt-to-income ratios