Lending Club Case Study

Name:
Kaustubh Santosh Choudhari

Problem Statement:

- Company Profile: The Lending Club is one of the largest online marketplace for the loans. Consumer can easily apply for the loans in various categories. The loans are funded by Investors based on the Credit ratings of the Customers.

- Problem: Lots of loans are Defaulted and Charged Off. The company wants to understand the driving factors behind loan default. The company can utilise this knowledge for its portfolio and risk assessment to make better decisions.

- Objective: As a Data Analyst, perform EDA to understand how consumer attributes and loan attributes influence the tendency of loans default.

Abstract

- In this presentation, we will dive into how data analytics can address this challenge, providing valuable insights into risk assessment and lending strategies.

- It faces a dual challenge: approving loans to creditworthy borrowers to grow its business, while avoiding potential defaults that could lead to financial losses.

- The objective of analysis is to use the information about past loan applicants and find whether they 'defaulted' or not based on their history data.

Data Availability

- The data available contains the information about past loan applicants and whether they 'defaulted' or not.
- The dataset does not contain any information for the loans which were 'NOT Approved'.

A. Datasets provided:

1. loan.csv
- The historical data for the approved loans which have either been fully-paid or Charged Off. It is in comma separated value (csv) file format.

2. Data_Dictionary.xlsx
- Contains information about the various columns. It is required for Data Understanding. It is in excel format.

Analysis Approach :

Data Cleaning

Removing the null valued columns, unnecessary variables and checking the null value percentage and removing the respective rows.

Data Understanding

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

Univariate Analysis

Analysing each column, plotting the distributions of each column.

Segmented Univariate Analysis

Analysing the continuous data columns with respect to the categorical column

Bivariate Analysis

Analysing the two variable behaviour like term and loan status with respect to loan amount.

Recommendations

Analysing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.
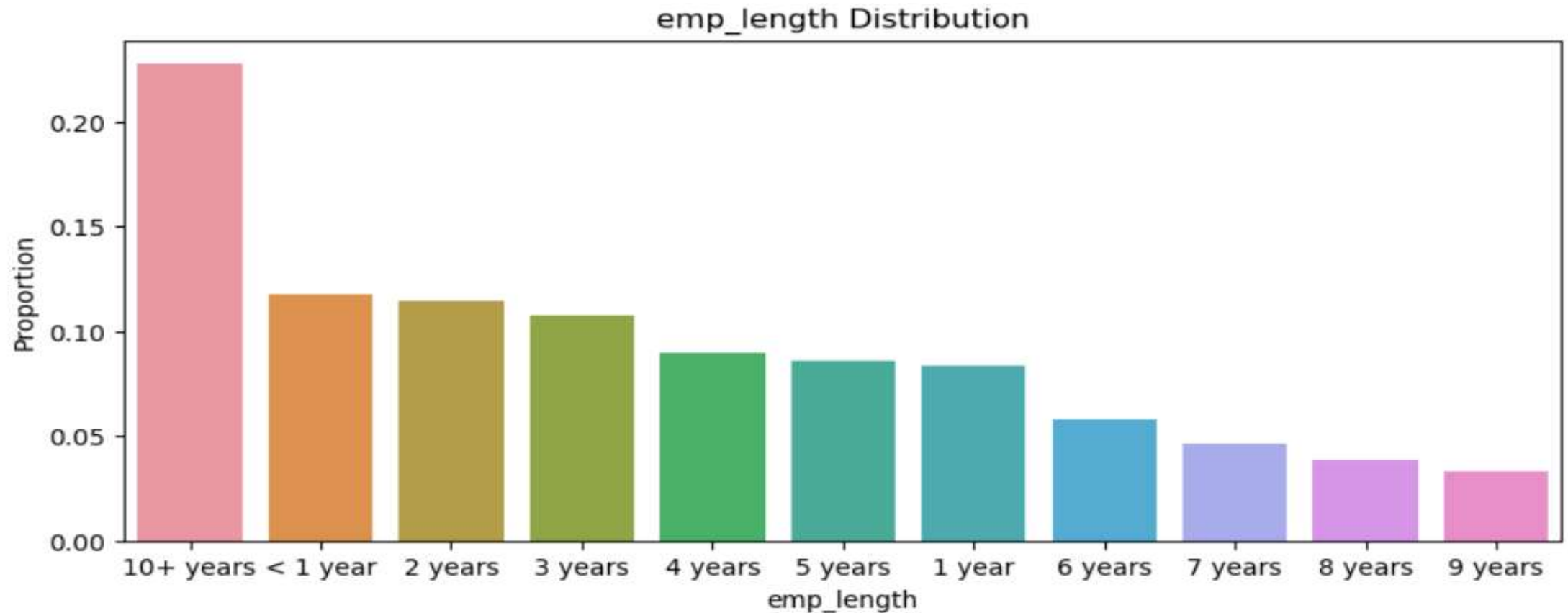
Analysis
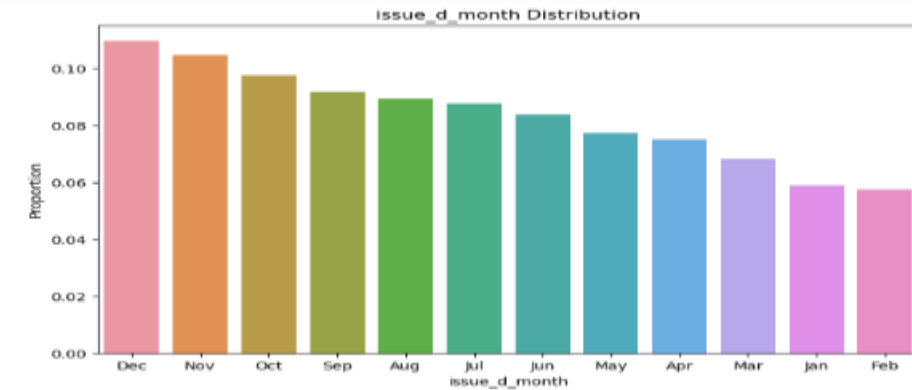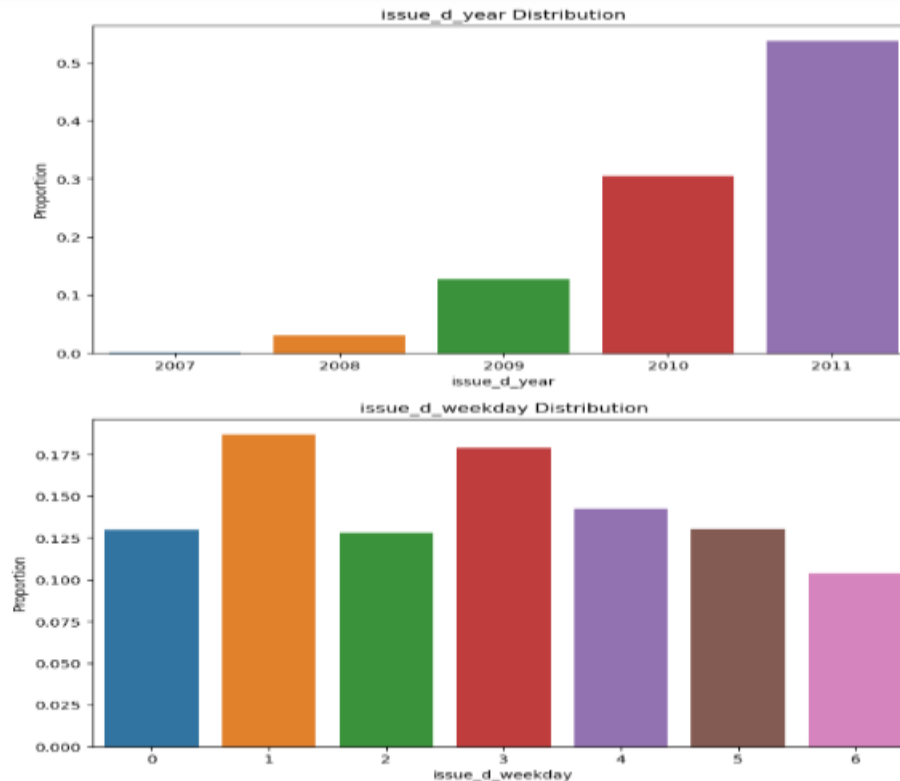
- Lending club has really expanded year by year, the number of loan issued are doubled every year.
- Also the issued month of loans is also increasing from January to December. In the final quarter of year there are more loans issued this could be because of vacation and christmas.
- Bank has issued more loans on tuesday and wednessday than other week days.

Analysis
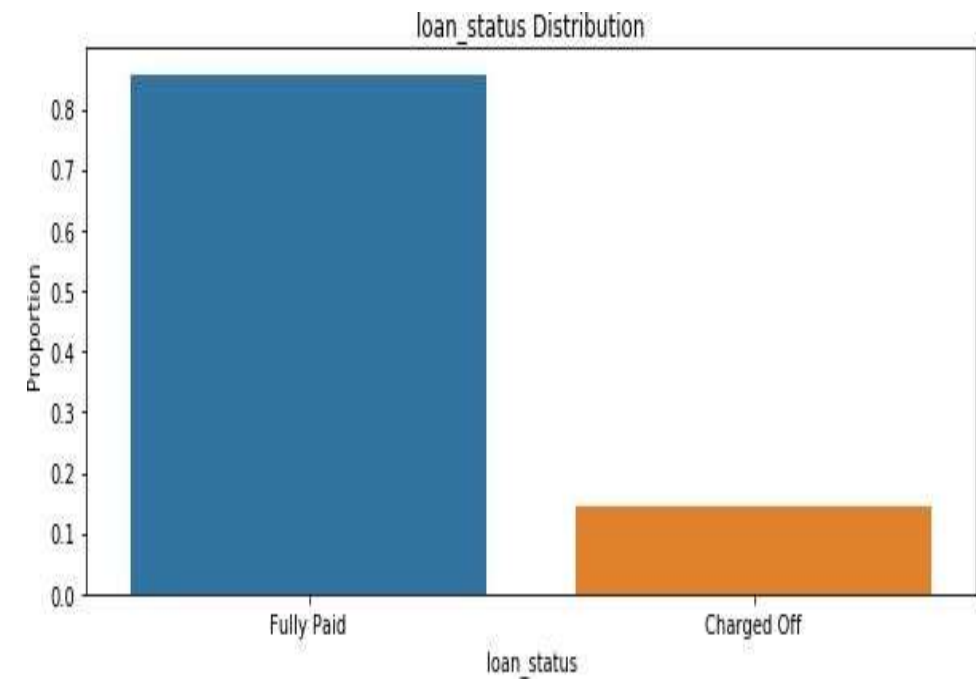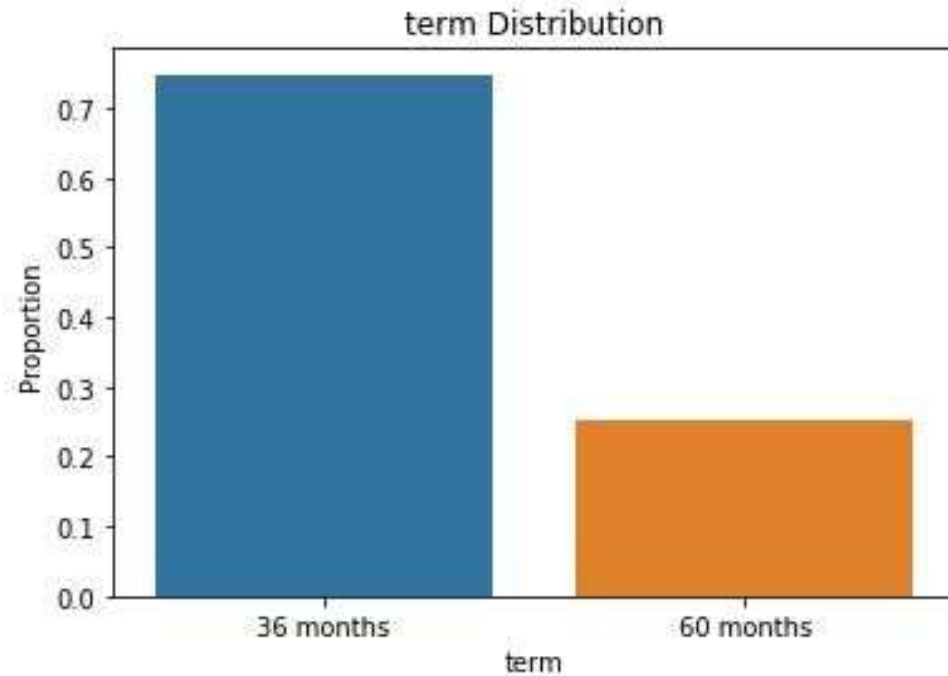
- There are only two loan terms 36 and 60 months. Around 75% borrowers took loans with 36 months term.
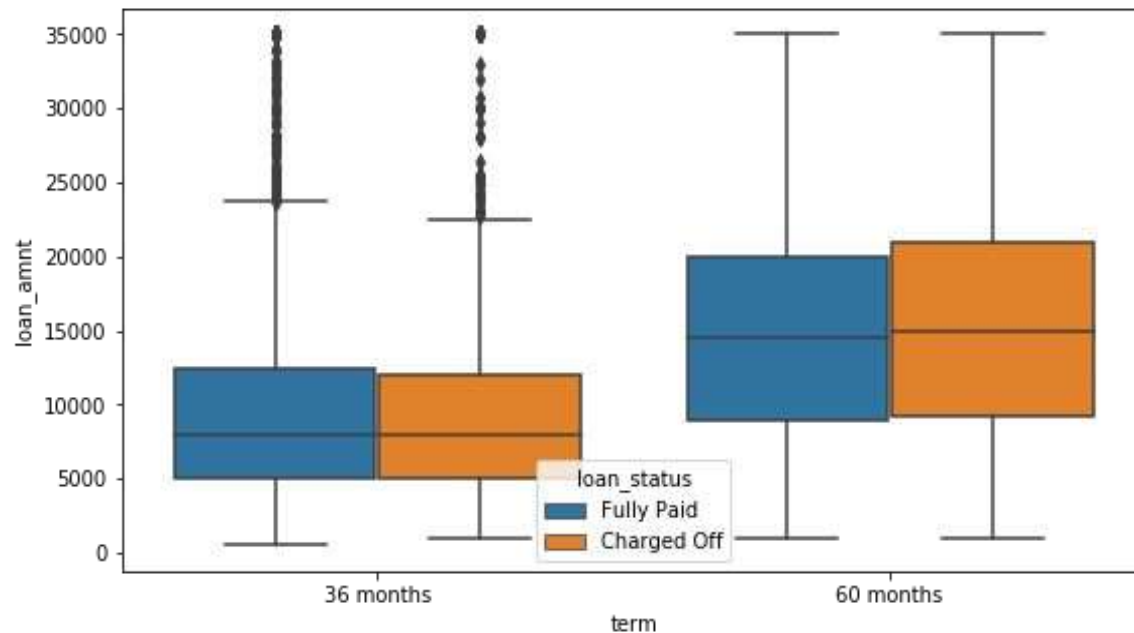- The charged off borrowers are around 15% and fully paid is around 85% in the given data set.

emp_length Distribution

- It is observed as people who has employee length of more than 10+ years are mostly taking loan.

home_ownership Distribution
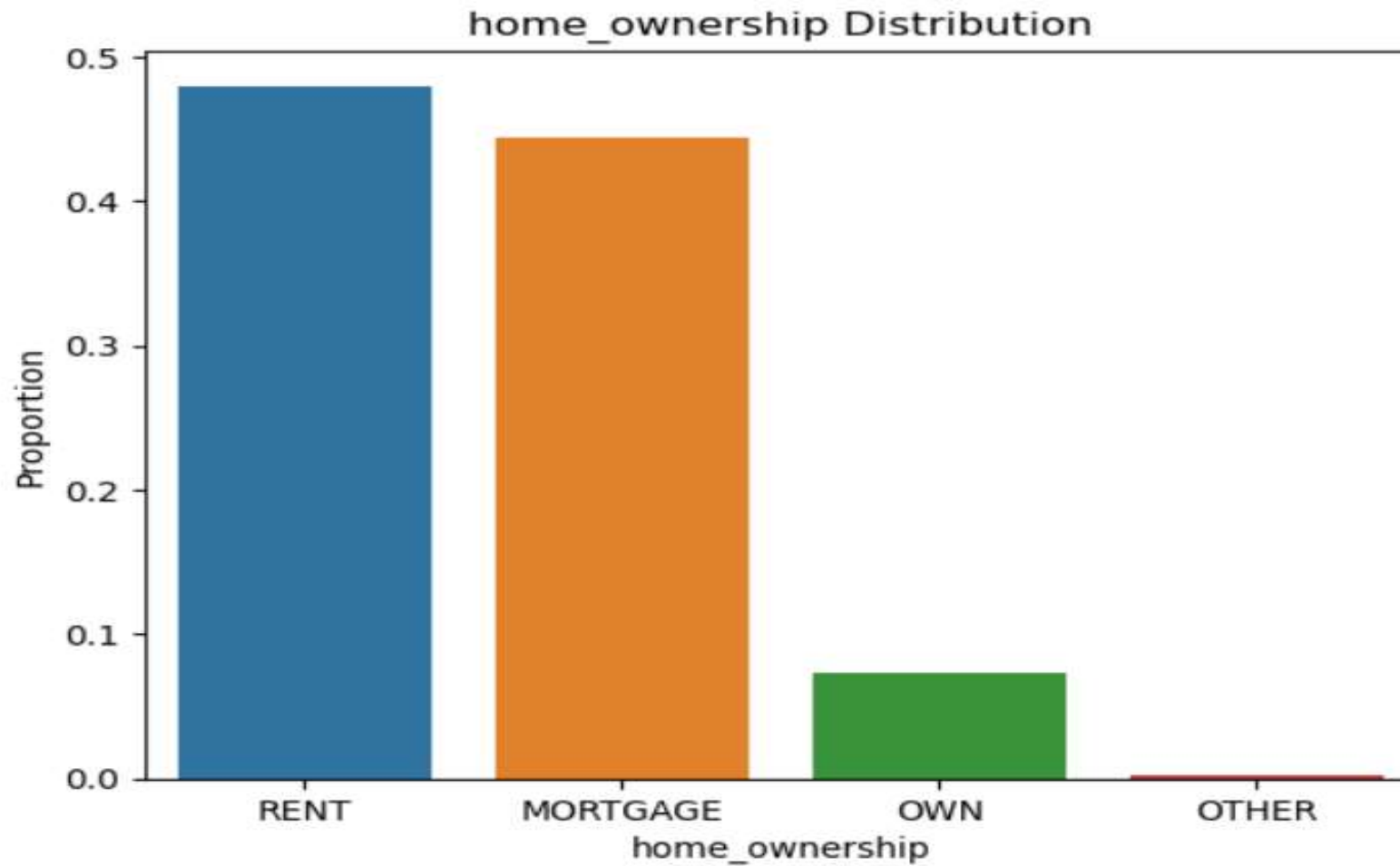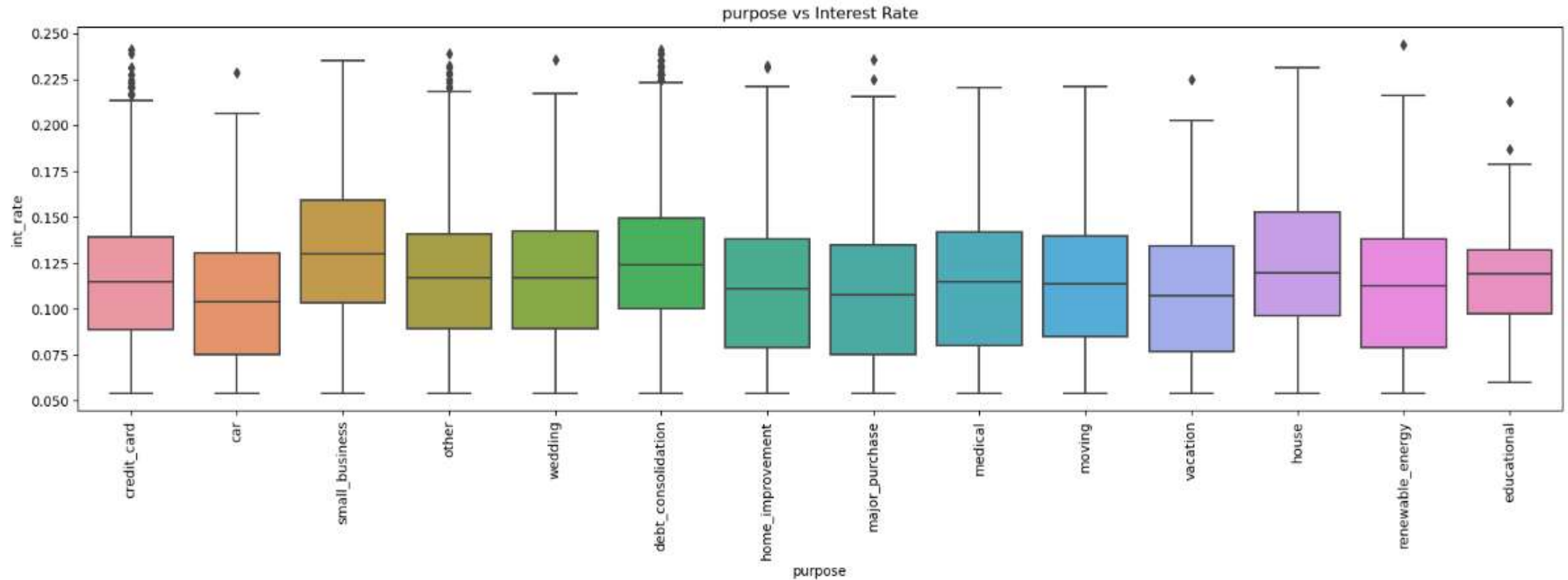
- Loan borrowers are mostly having Rented & Mortage house

Analysis
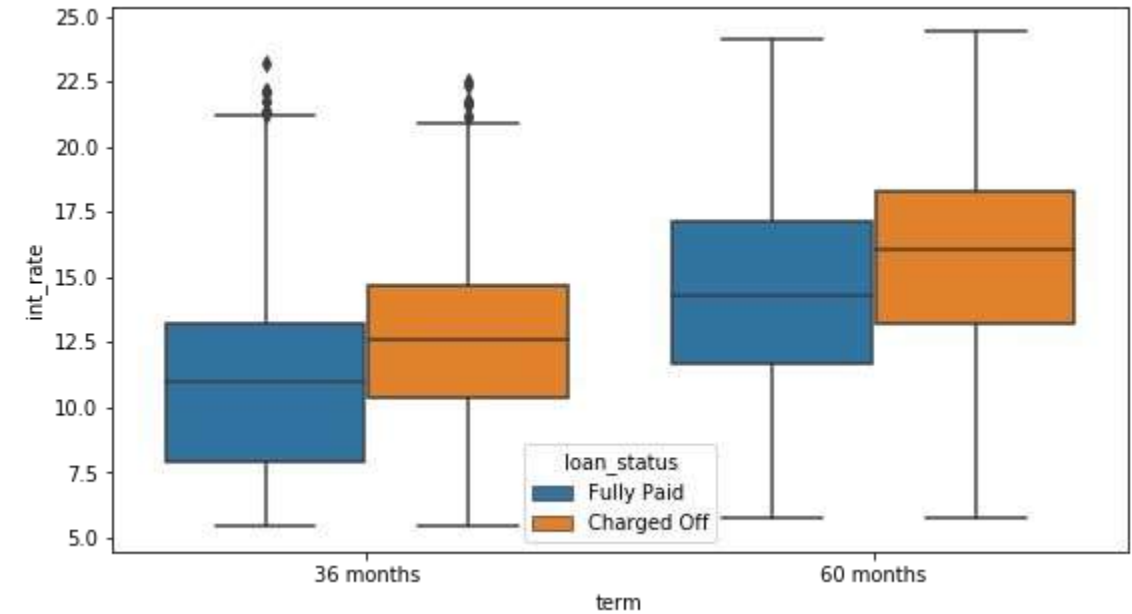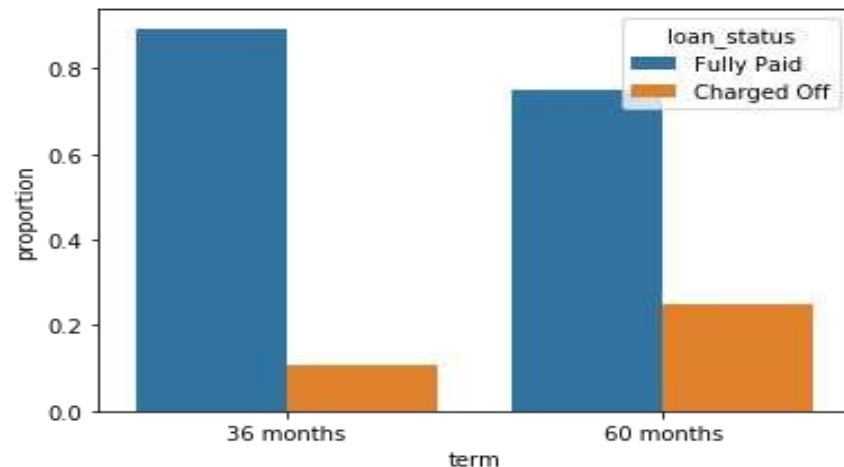
The default rate is high in 60 months tenure because most people took high loan amount with high interest rate in it and they faced difficulties in returning the sum to bank.

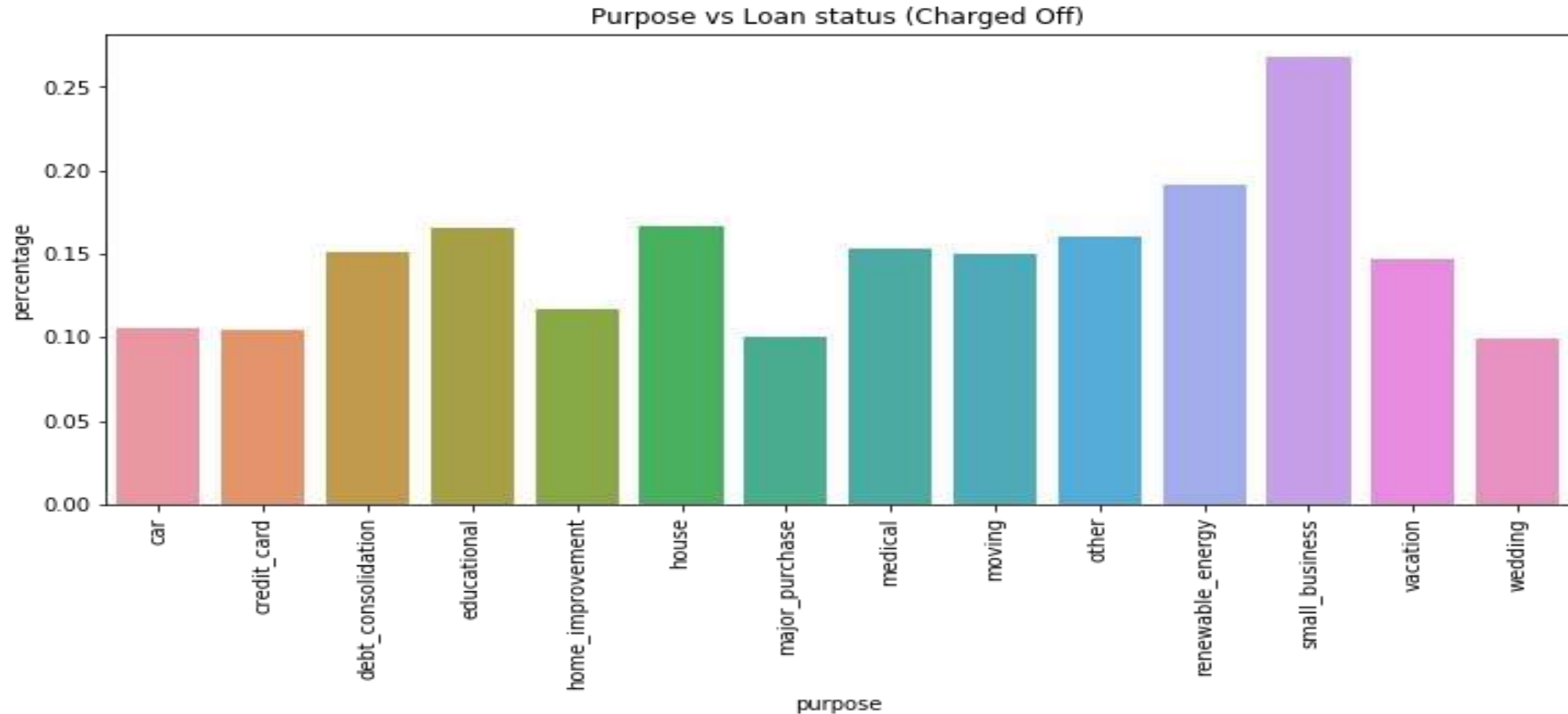# Analysis

purpose vs Interest Rate

- Small Business, Debt Consolidation and House loans are given with more interest rates comparatively others.
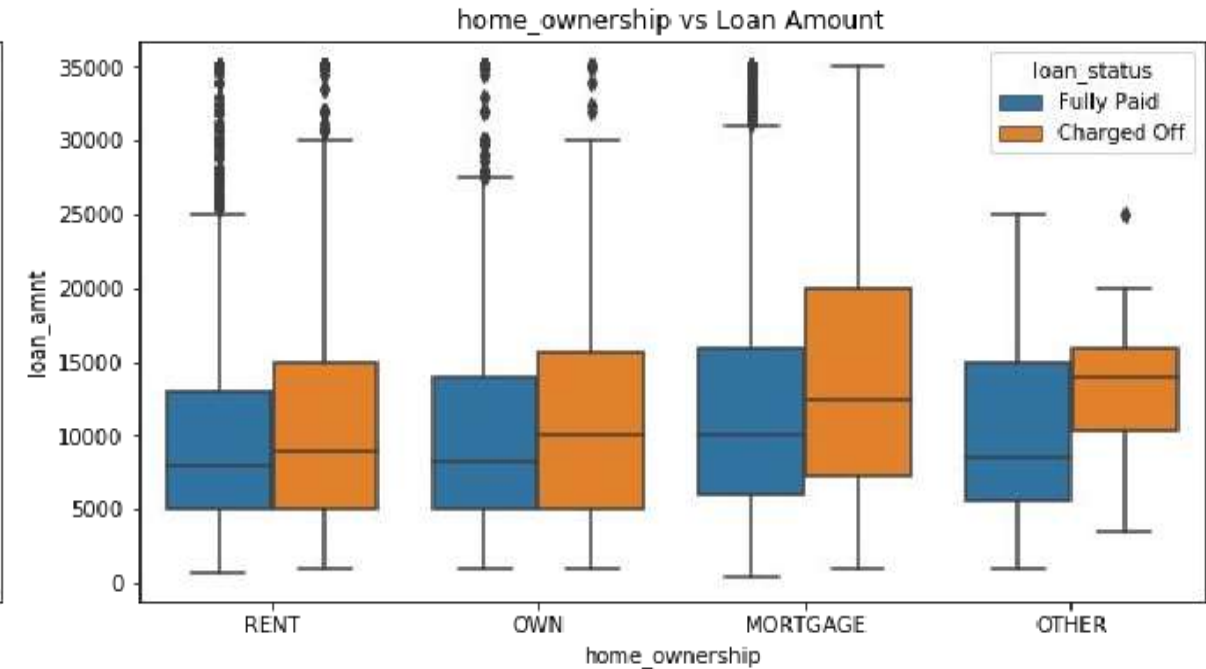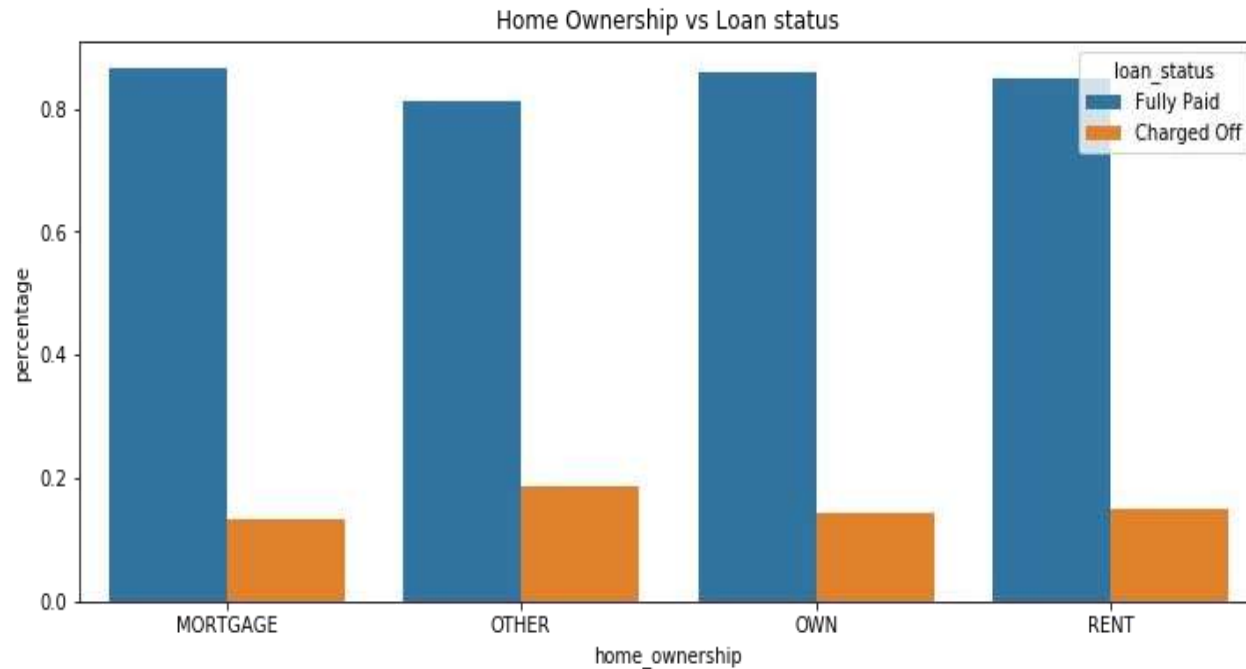
Analysis
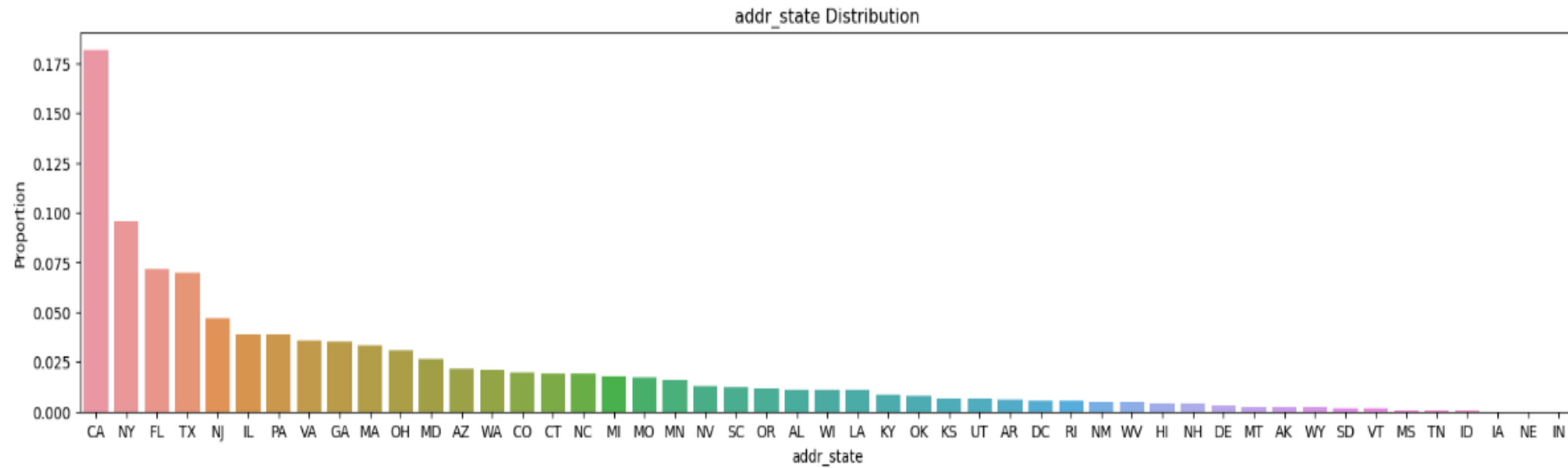
- Most of the borrowers are from CA, NY, FL and TX state.

Analysis

Purpose vs Loan status (Charged Off)

- Borrower's who took loans for small business purpose have defaulted more.

Analysis

- There is around 20% chance of loan default in each home ownership category.
- From the 2nd plot we can see the people with higher loan amounts in **mortgage** home ownership has high default rate than others.

Correlation Matrices

Correlation between Columns
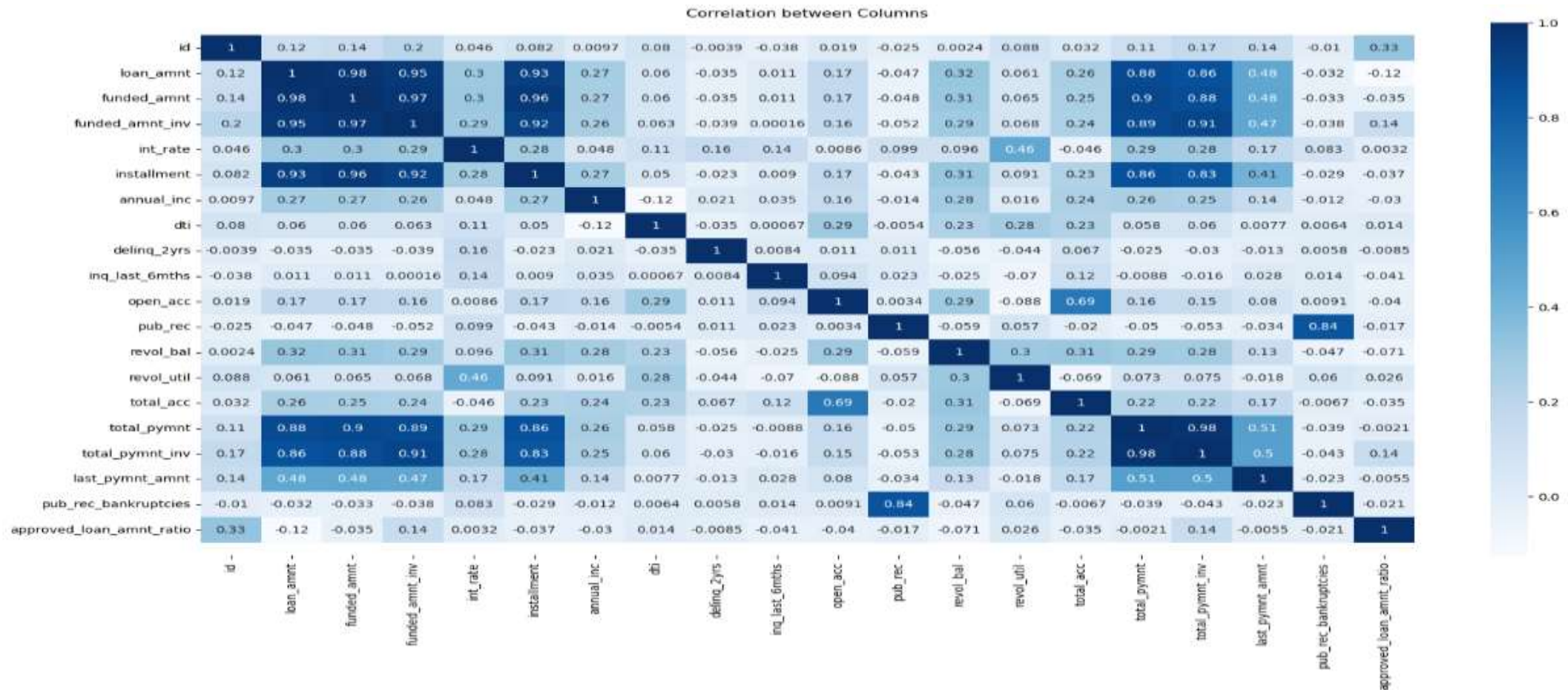
Observation : The public derogatory records column is highly (+ve) correlated with public bankruptcies records.

Key learnings

- Data-driven decision-making is crucial in financial services.
- Data analysis enables risk assessment and identification of high-risk borrowers.
- Reduced default rates lead to increased profitability and customer satisfaction.
- Continuous monitoring and compliance with regulations are essential.
- Ethical data usage is paramount.
- Data analytics optimizes portfolio management and investor relations.
- Customer-centric lending practices enhance long-term relationships.
- Data analytics is a strategic tool for sustainable growth and risk mitigation in consumer finance.

Recommendations

- Income-Linked Loan Limits: Restrict loans over $30,000 to high-income individuals.
- Selective Loan Categories: Reduce approvals for Small Business, Renewable Energy, and Education loans.
- Grade-Based Decisions: Don't approve loans with poor credit grades (F and G).
- Income-Based Limits: Be cautious with individuals earning < $20,000 annually.
- Debt-to-Income Ratio Assessment: Avoid loans for those with DTI ratios > 20%-25%.
- Financial Education: Provide resources for borrower financial literacy.
- Borrowers with mortgage home ownership are taking higher loans and defaulting the approved loans. Lending club should stop giving loans to this category when loan amount requested is more than 12000.