Lending Club Project

Muralidhar G ; Lalitha U

Project Overview

Business Context : Analysis of loan applications of Urban borrowers by Consumer finance company. Company is looking for a model to minimize the risk of default or write offs.

Problem Statement : Identifying drivers to minimize Risky Loan Applications which can be used by Analyst

Solution Approach

Start

Requirements Phase : Domain Expertise

Understand the Business Context and Problem Statement → Learn more about Domain and Business Environment → Convert Business Problem To Analytics Problem

Analytical Phase: Statistics/ Data Expertise

Collect Required Data → Analyze Raw Data → Cleanse available data → Prepare data by adding derived variables

Univariate Analysis → Bivariate Analysis → Multivariable Analysis → Collect Observations

Conclusion Phase

Convert Observation and inferences to Business terms → Prepare checklist to identify the Risk Loans

Finish

Analysis

- ~14.7% (5418) of loans are charged off.
- Loans with 36 months are more popular than long term loans of 5 years.
- In histogram we see there are two towers (normal distributions). Probably data can be analyzed for Interest rate between 5% to 8% and 9% to 25% separately.
- Majority of loans were taken by Borrowers with 10+ years experience.
- Most Borrowers live in rented accommodation or on mortgaged properties.
- Employee verification status is only 50%.
- Majority of the borrowers are from the large urban cities like california, new york, texas, florida etc.
- Majority of the borrowers have no record of Public Recorded Bankruptcy.

Intrest Rate Histogram


barplot of emp_length
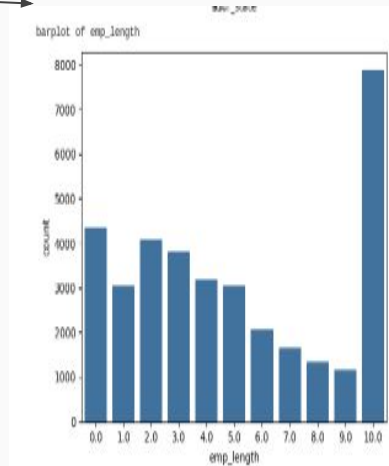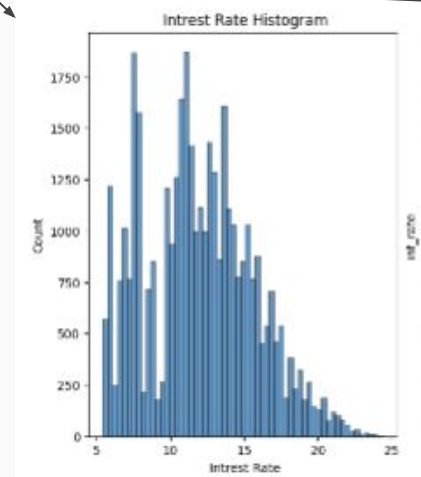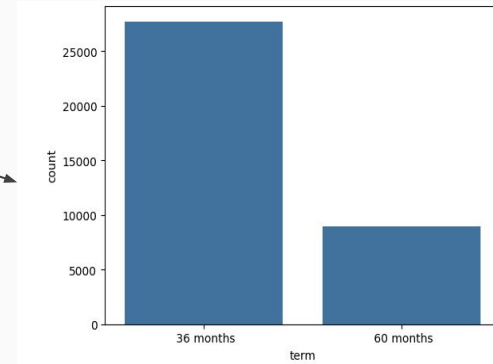
Bivariate Analysis

- Loans with Higher Interest Rates led to charge off (not paying up).
- Charged Off loans Interest Rate median is higher than median of Fully Paid loans. **Inference : Higher Interest loan rates are Risky**
- Borrowers with Higher % of revolving debt utilization need to be considered as risky.
- Default rate is higher for long term loans.
- Default rate of Borrowers with Own house is lower than borrowers staying in rented house or mortgaged house.
- Education Loans are less risky
- Borrowers of 'No record of Public Recorded Bankruptcy' are safe choice fo issue loan.
- Interest rate increases as Grade increases. I.e A Grade Interests are 3 times less than Garde G borrowers.
- Debt consolidation is the single largest factor driving volume of loans.
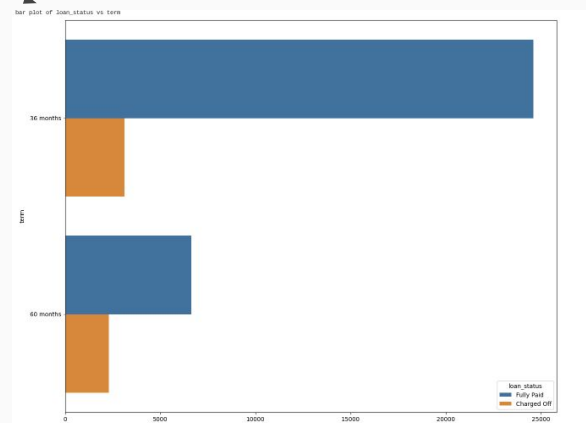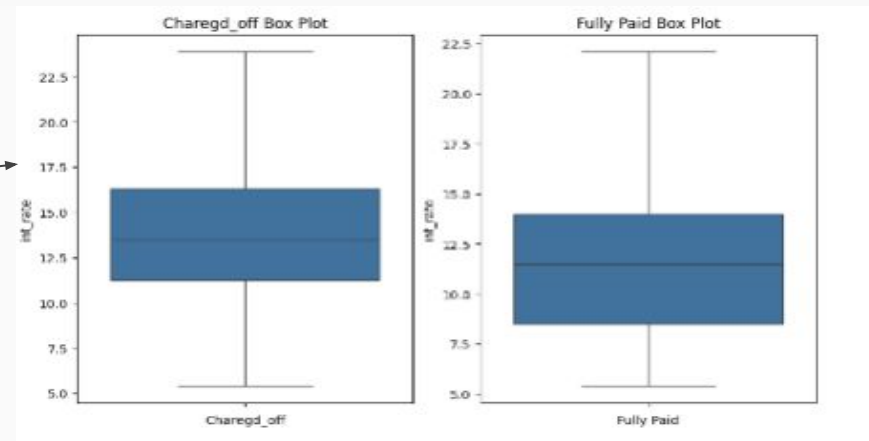
Bivariate Analysis

- Default rate of Borrowers with Own house is lower than borrowers staying in rented house or mortgaged house.
- Education Loans are less risky
- Borrowers of 'No record of Public Recorded Bankruptcy' are safe choice fo issue loan.
- Interest rate increases as Grade increases. I.e A Grade Interests are 3 times less than Garde G borrowers.
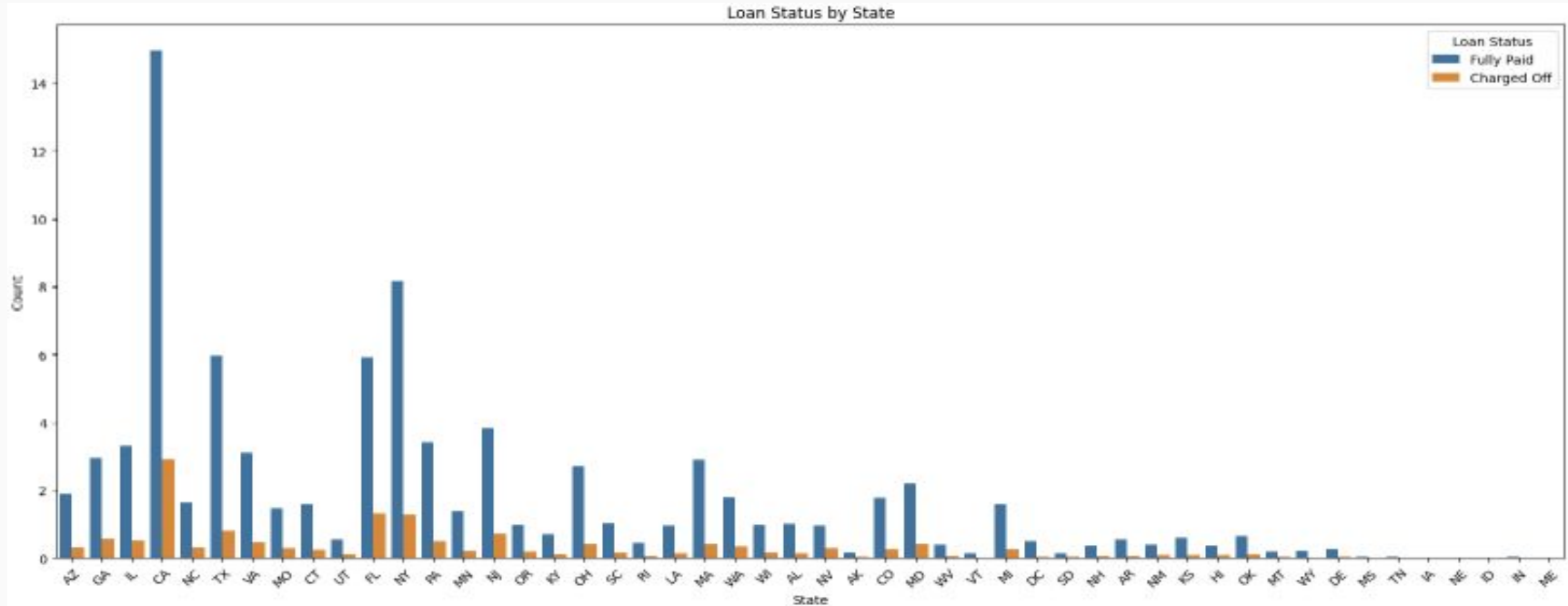- Debt consolidation is the single largest factor driving volume of loans.

bar plot of loan_status vs home_ownership
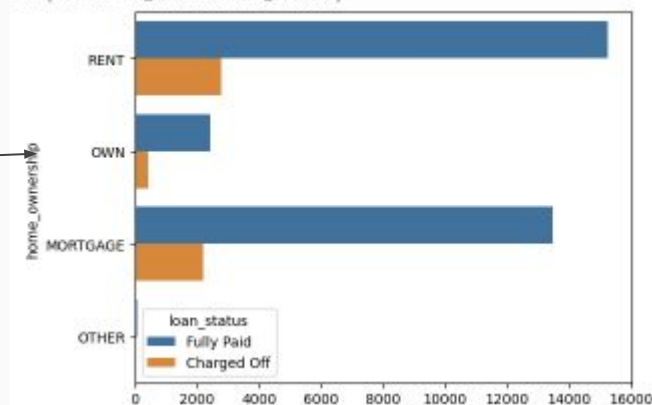
bar plot of loan_status vs purpose

Bivariate Analysis

Loan Status by State
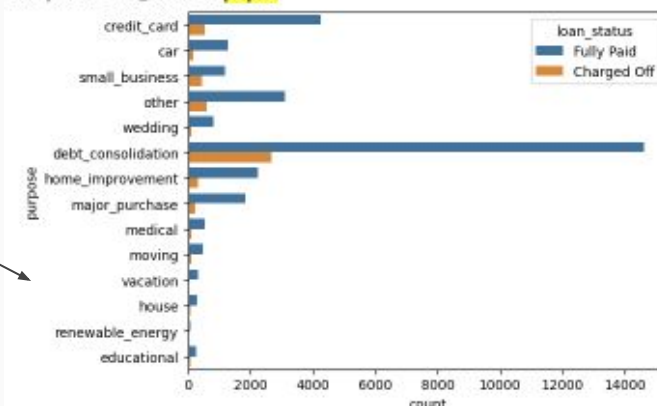
- California, Florida and New York have the highest number of defaults. % of loans defaulted in Florida is higher among these 3 states.

Conclusion

- Loans with higher interest rates are like to be risky.
- Borrowers staying in renting accommodation or mortgaged house are risky applications
- Loans taken for Debt consolidation are risky for the lenders
- Borrowers of 'No record of Public Recorded Bankruptcy' are safe choice to issue loan.
- California, Florida and New York applications are risky.
- Borrowers with Higher % of revolving debt utilization need to be considered as risky.

Factor need to be considered for evaluating Risk of loan are :
         1) Rate of Interest **2)** Borrowers accommodation Status **3)** Purpose stated for taking loan
         4) Record of Bankruptcy **5)** State of Residence **6)** % of revolving Debt