Lending Club Case Study - Submission

Name: Ravi Ramchandran, Sreejith Sreekumaran

22-Feb-2021

Abstract of the problem

Background :

Lending Club is a **consumer finance company** which specialises in lending various types of loans to urban customers

Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss), like all Lending companies

LC wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.

Objective :

LC wants to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss

Use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

Problem Solving Methodology

Data Cleansing

- Drop Columns with Null Values
- Remove Anomalies like %, Months etc

Data Preparation

- Define Target Variable
- Create Derived Variable like Issue Year, Month etc

Data Exploration

- Create distributions for Categorical and Numeric Variables
- Overall View of Data using Charts to Narrow List of Features

Univariate Analysis/Segmented Univariate

- Analyse One Feature with Target Variable to see Correlations
- Create Segmented Distributions to visualize the same

Bi Variate Analysis

- Create Derived Variables
- Create joint distribution to see impact of see relations between Derived Variables/2 variables wrt to Target Variables
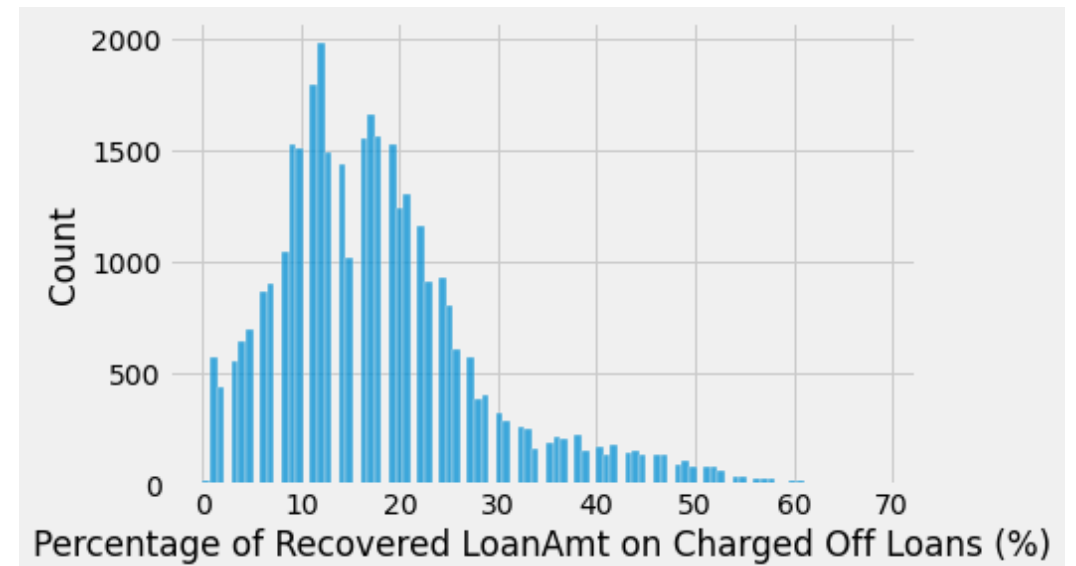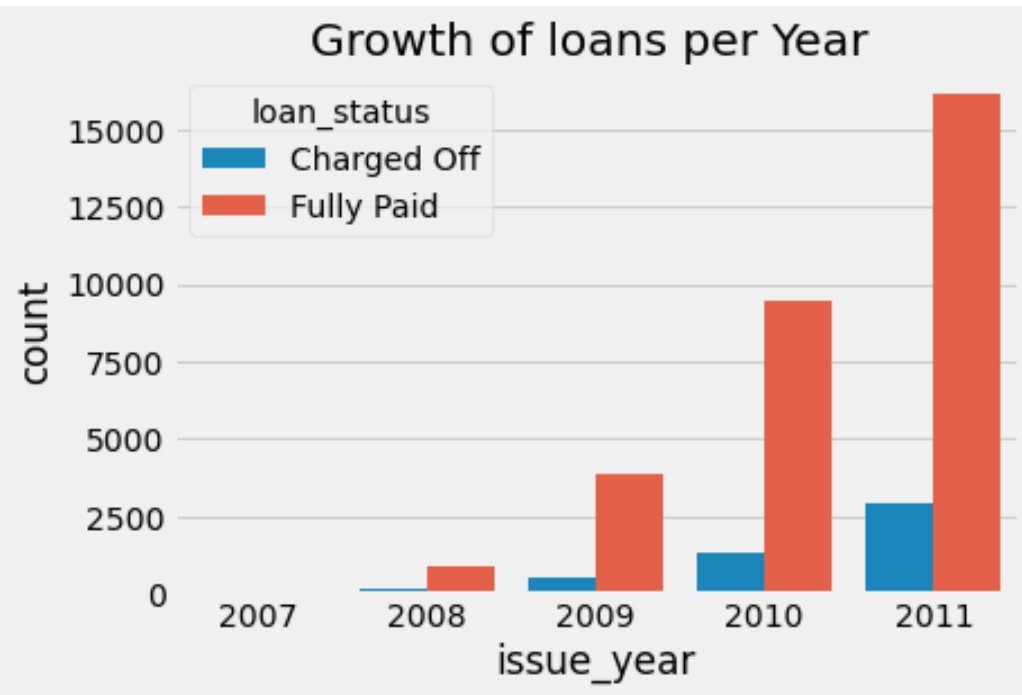
Analysis Recommendations

- Publish Observations and Insights

Data Analysis

- Loans are Increasing on a Year on Year Basis which is a good sign
- However Defaults are also increasing, at a similar rate which should be reduced
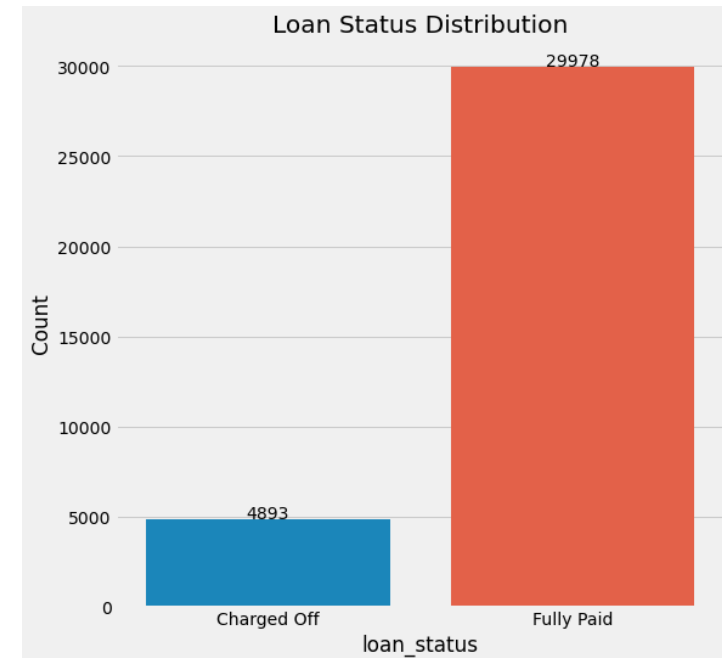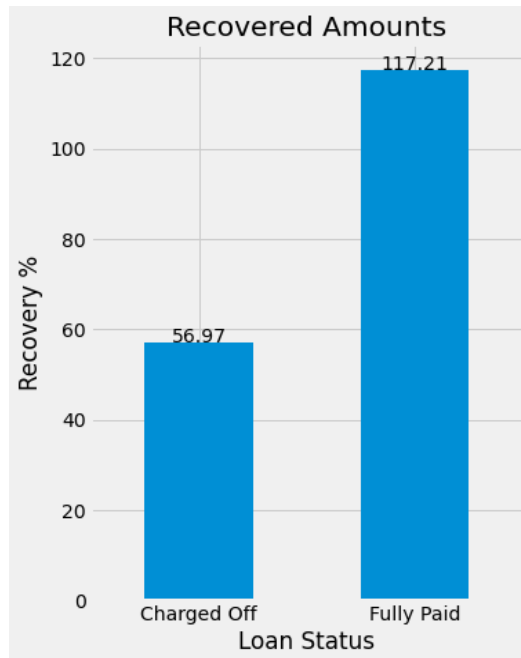- Recovery Rate is still relatively around 30% which needs to be improved

Data Analysis

- Around 14% of the loans are Charged off
- LC made 17% profit in Fully paid loans while it could recover only 57% from Charge off Loans
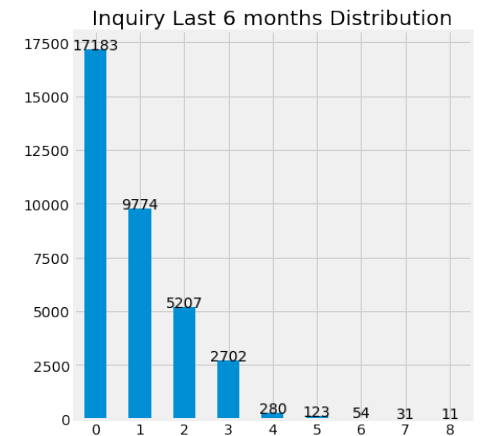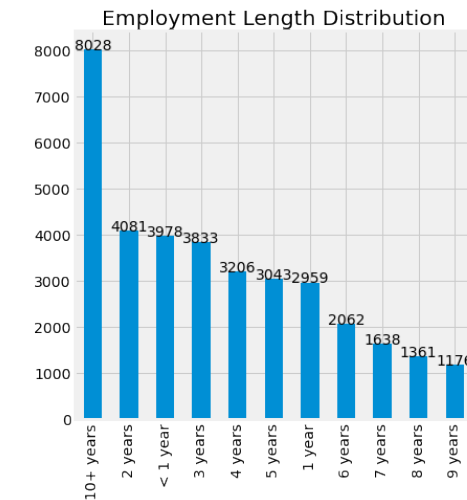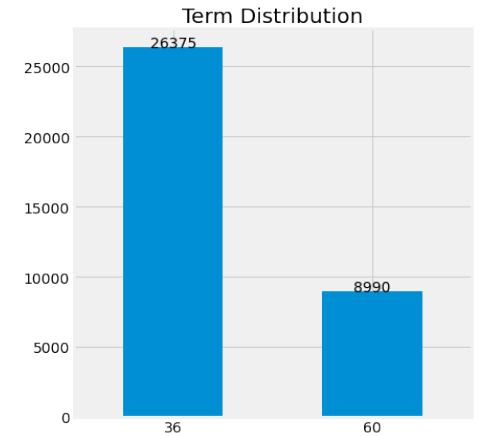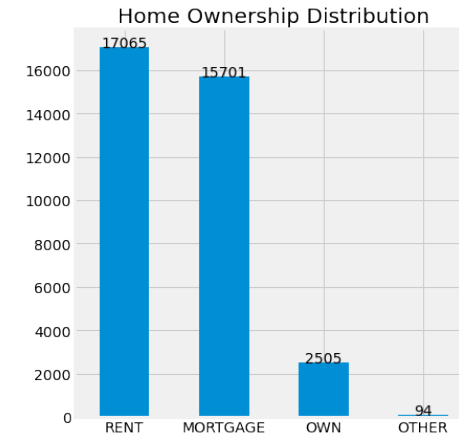
Understanding Data

Higher proportion of Loans are sought by people with 10+ years of employment.

Most loans have term of 36 months.

Maximum loans are taken by people staying on rent or homes on mortagage.

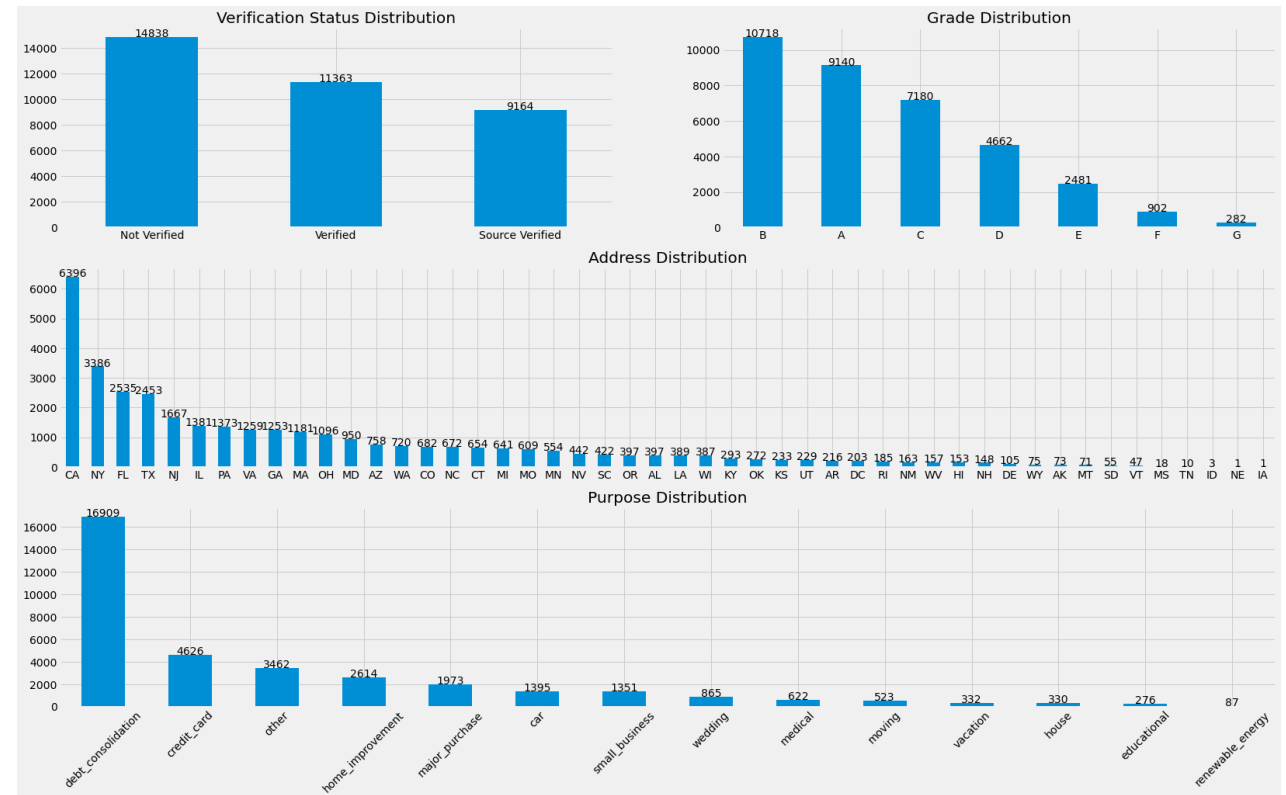Repeat enquiries have been low in the last six months.

Understanding Data

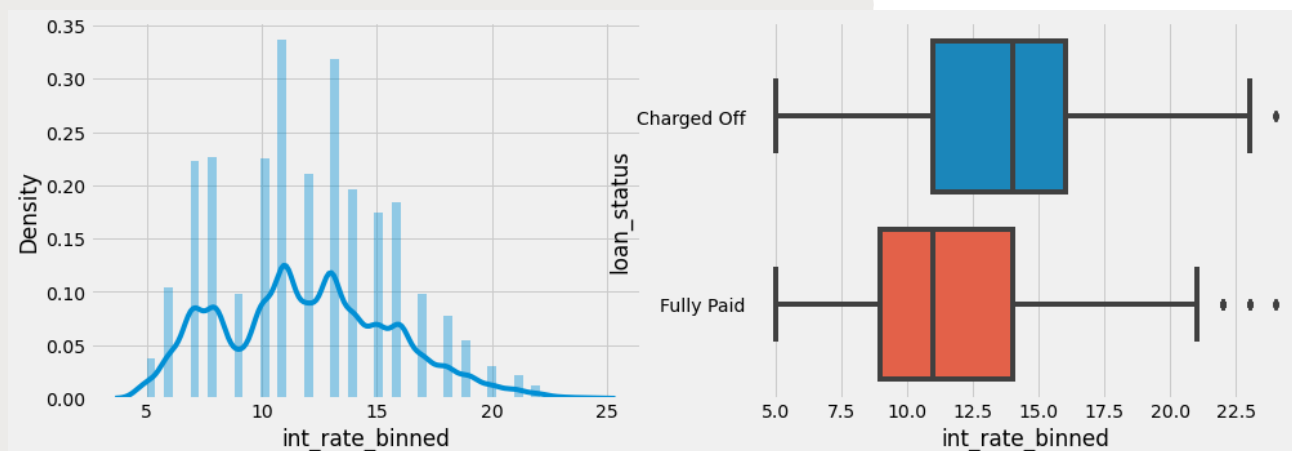Seems there is a higher tendency to give of loans without verification.

Most loans are taken from California state.

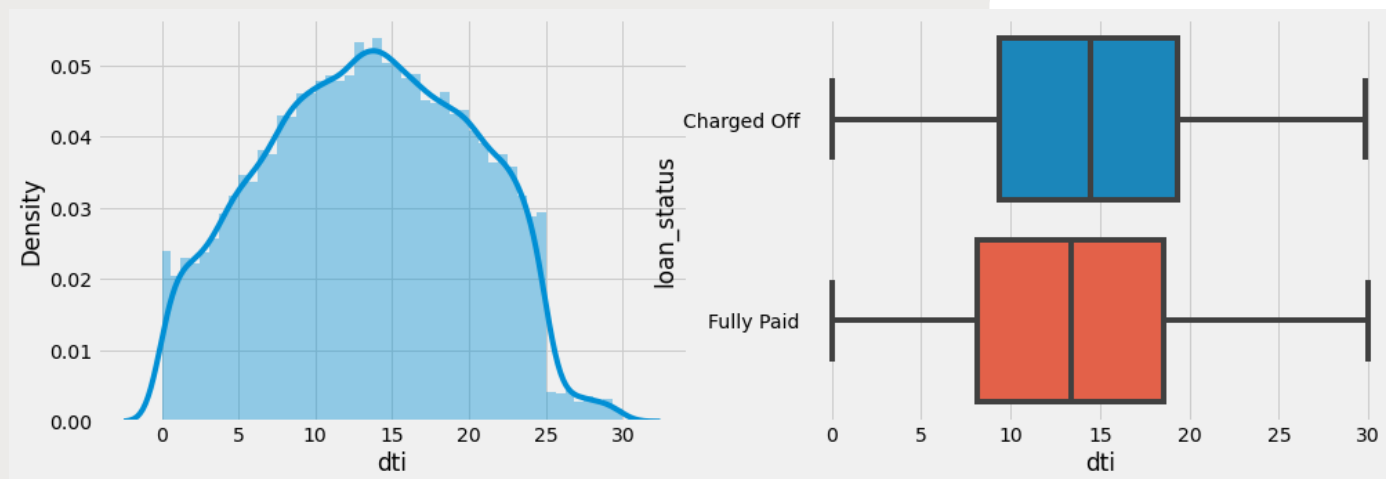Most of loans are taken for debt consolidation.

Majority of the loans belong to grade 'B'.

Univariate Analysis

Distribution of Interest Rate across Loan Status

Distribution of DTI across Loan Status

Interest Rate

- Charged off loans are 2% points higher than Fully paid loans. Looks like -

- LC is assessing risk correctly and charging riskier loans higher

DTI

- As can be seen the 75% of all Fully Paid Loans is around 17, which is less than that of Charged of Loans.

- We can infer that the DTI above 18 is at higher risk of being "Charged Off" State

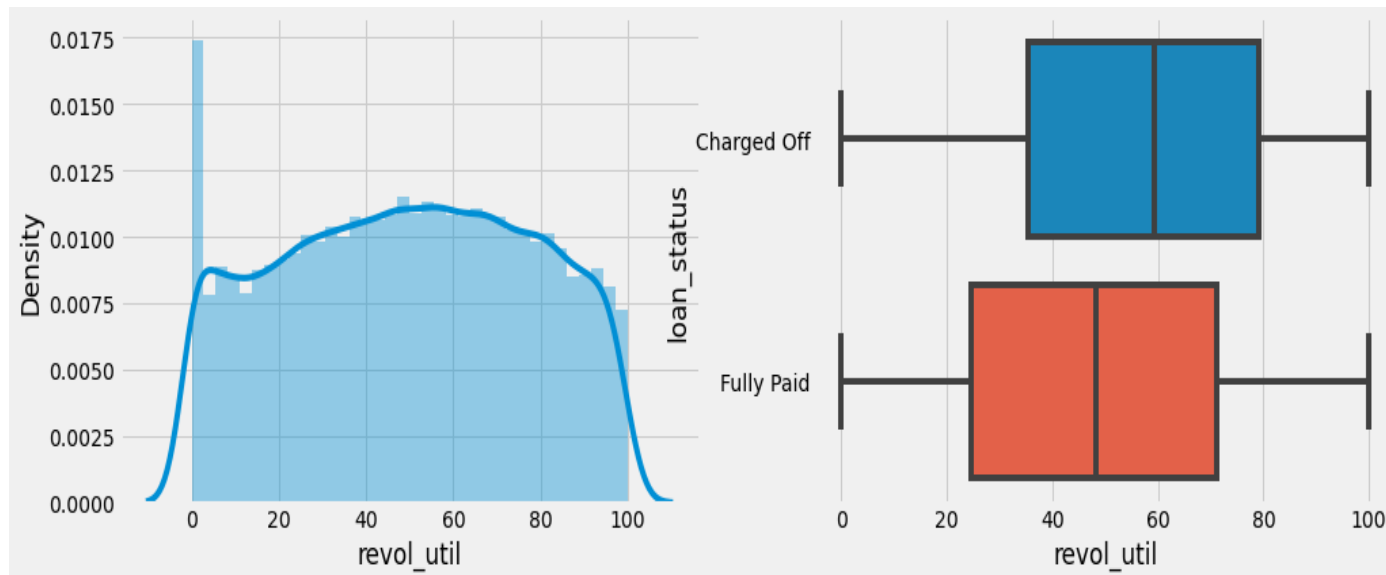- LC should restrict lending to applicants having DTI >17

Univariate Analysis

Distribution of Enquiry across Loan Status
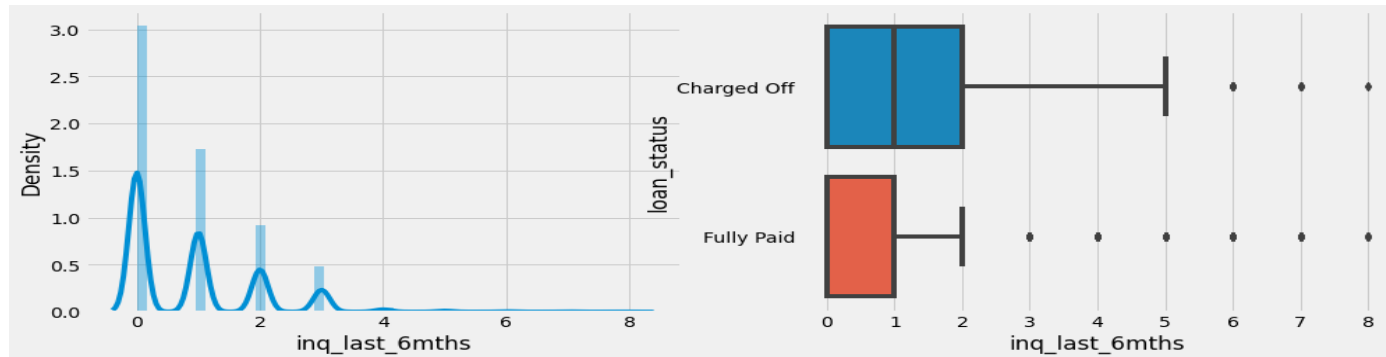
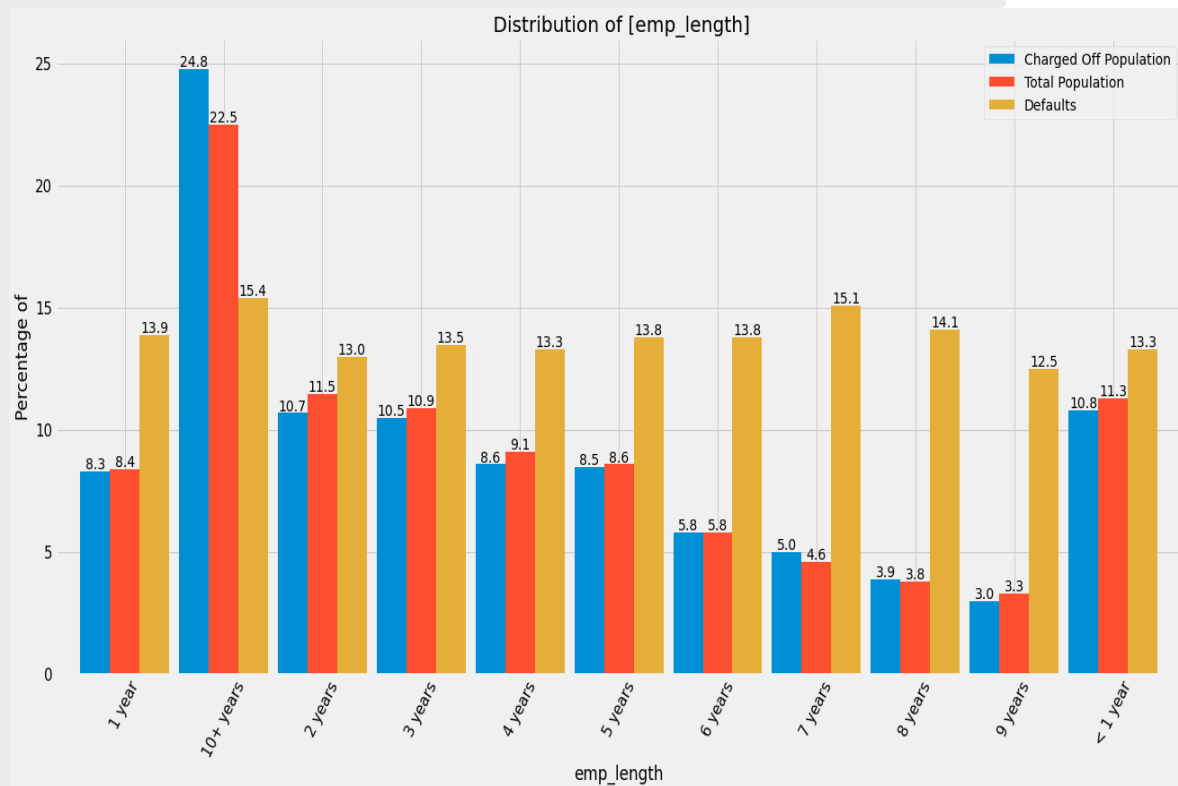Distribution of Revolving Line of Credit across Loan Status

Repeat Enquiry
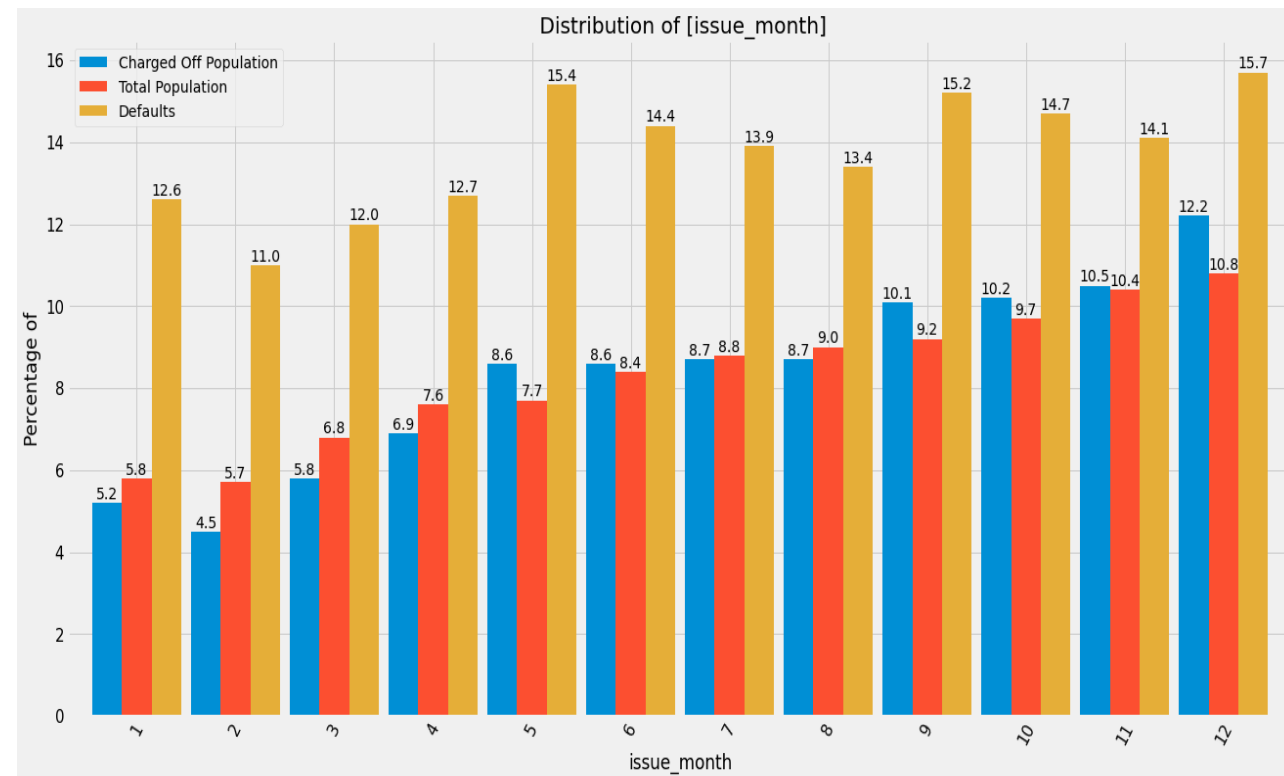
- People with repeat enquiry tend to default more.

Revol Util

- Revolving Line of Credit is higher for Charged Off loans

Univariate Analysis

Distribution of [emp_length]

Distribution of [issue_month]

Employment Length

- Max Loans are given to tenured employees, which is working well
- Defaults ratios are same across categories
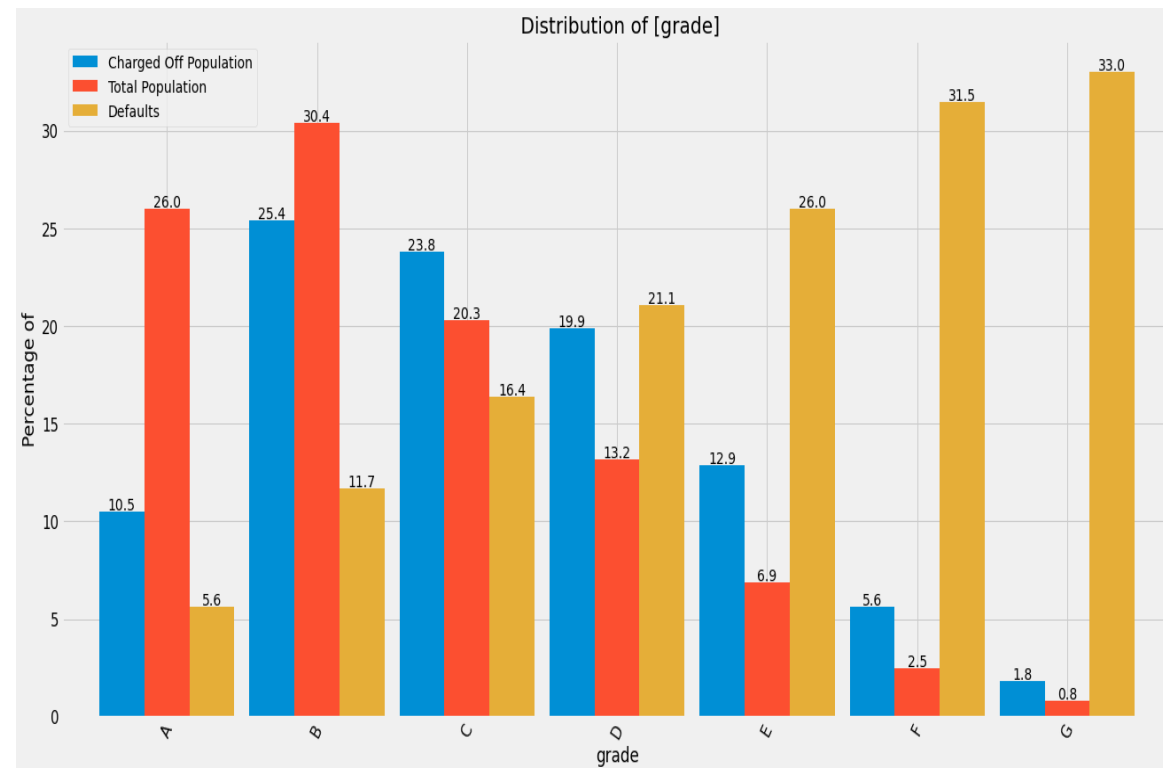- LC should continue the same

Month

- LC has been giving out maximum loans in December
- There is a high distribution of Loans as the year is going to end. Any targets being met by LC ?
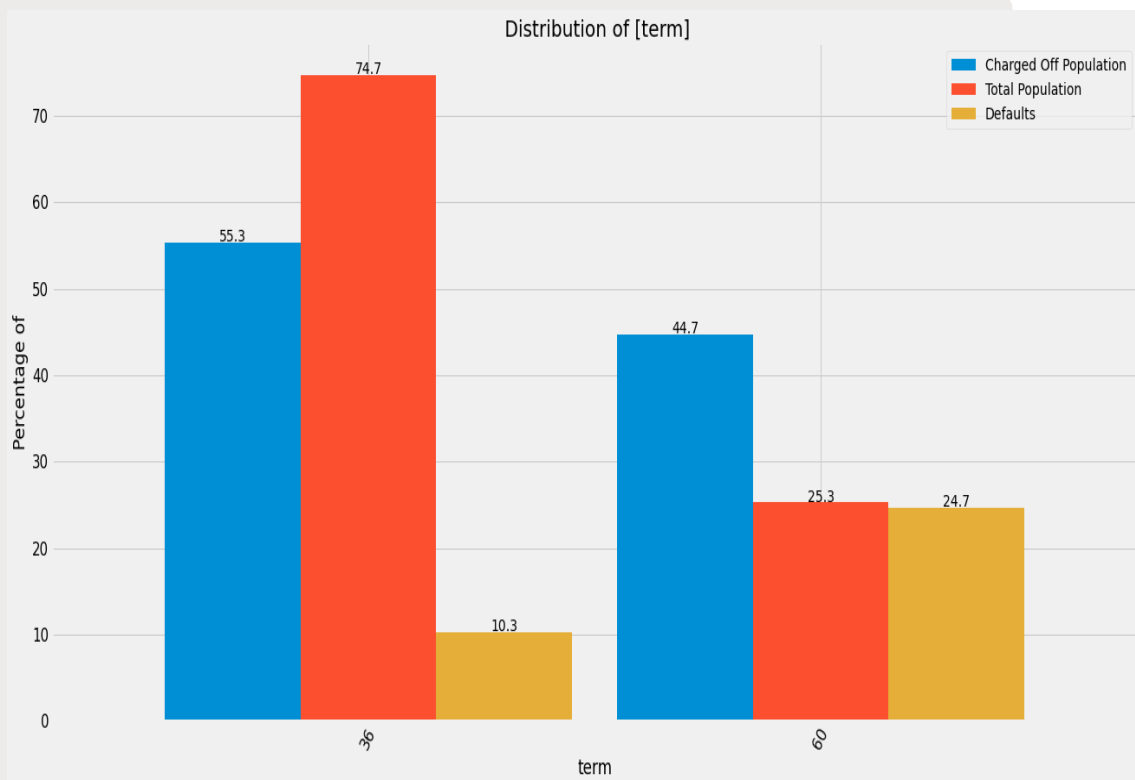
Univariate Analysis

Distribution of [purpose]

Distribution of [grade]
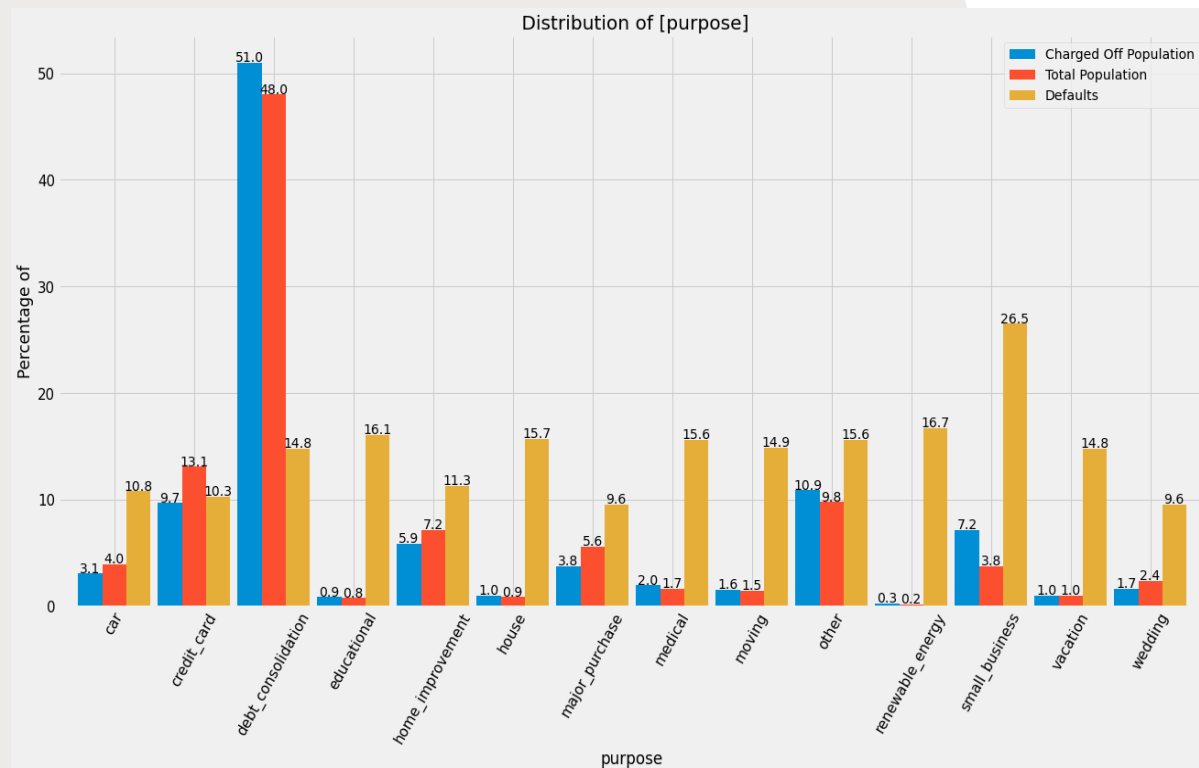
Purpose

- Small Business has default rate of 26%.

Grade

- Nearly 30% of all loans in Grades F and G see a default.
- Grade E onwards are risky and less numerous.

Univariate Analysis

Distribution of [term]

Distribution of [verification_status]

Term

- Almost 25% of all charged off Loans have duration of 60 months
- Higher the term means higher the probability of Default
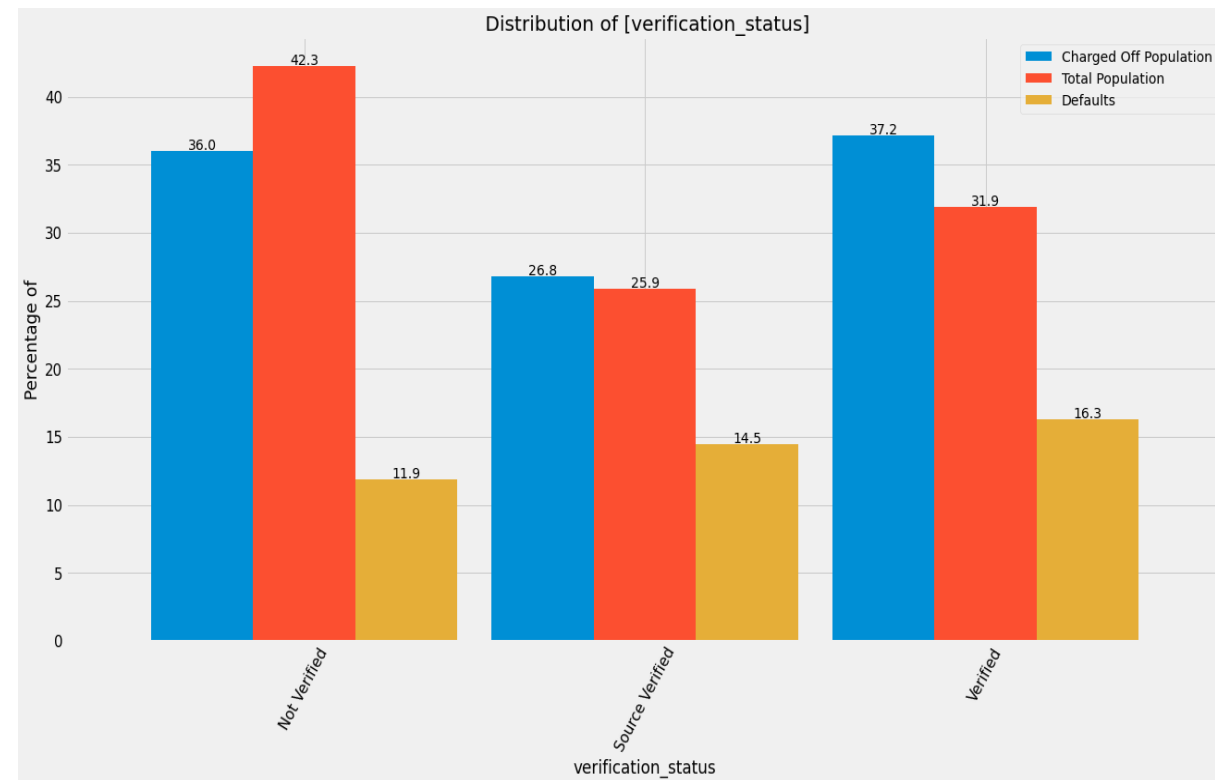
Verification Status

- 'Non Verified' loans is very high. LC should take steps to verify these loans.
- Verified loans show more charged Off percentage as compared to Not verified loans.

Univariate Analysis

Distribution of [loan_income_ratio_bin]

Distribution of [annual_inc_range]
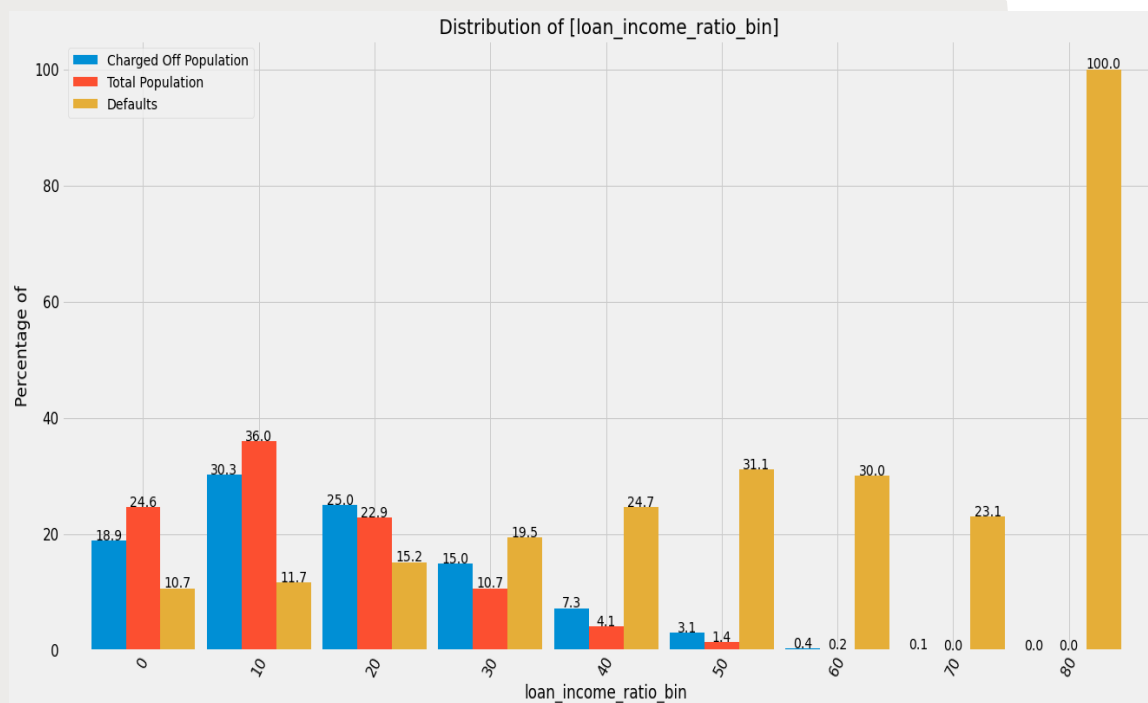
Loan to Income Ratio

- Loan amount of 30% and above of the total income is having probability to default.
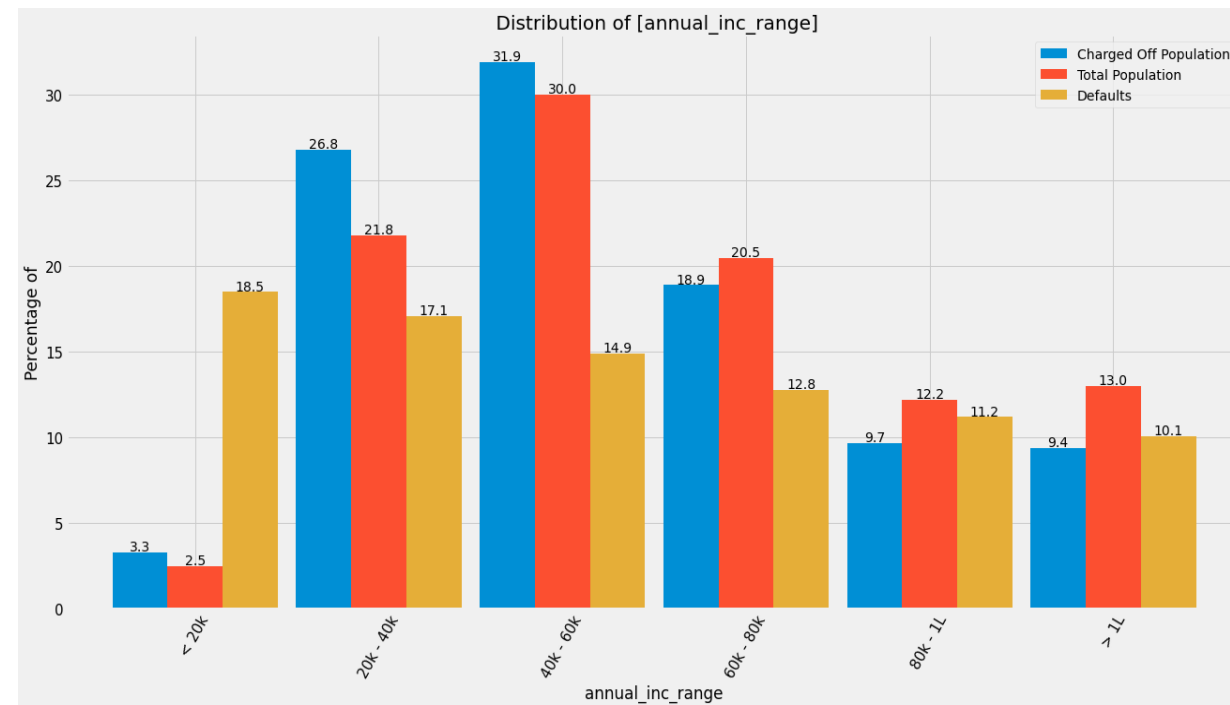
Annual Income

- Loan defaults are higher at lower income and reduces as the income goes up

Bi Variate Analysis

Grade and Probability of Default

- LC has a Grading System from A to E
- Defaults are increasing as per grade, so LC must continue with the system
- LC should restrict loans in grade E, F, G as the cost benefit of risk is not justified
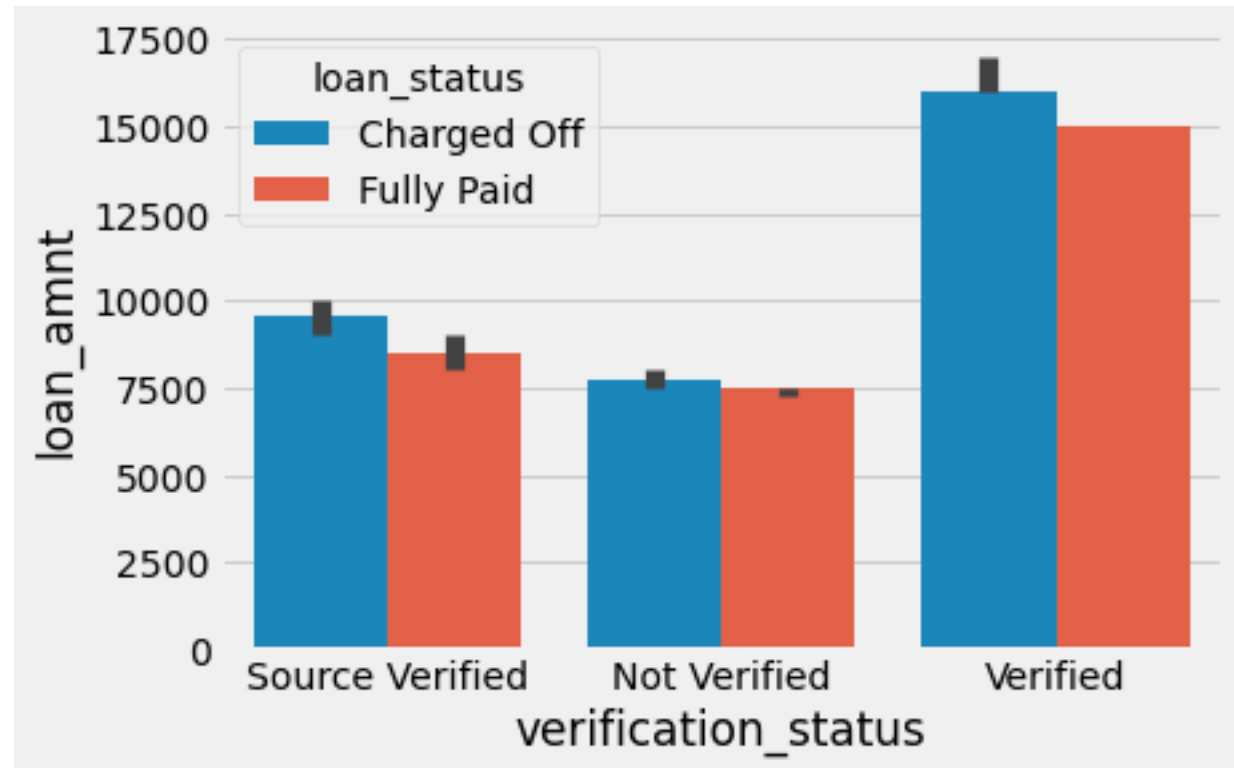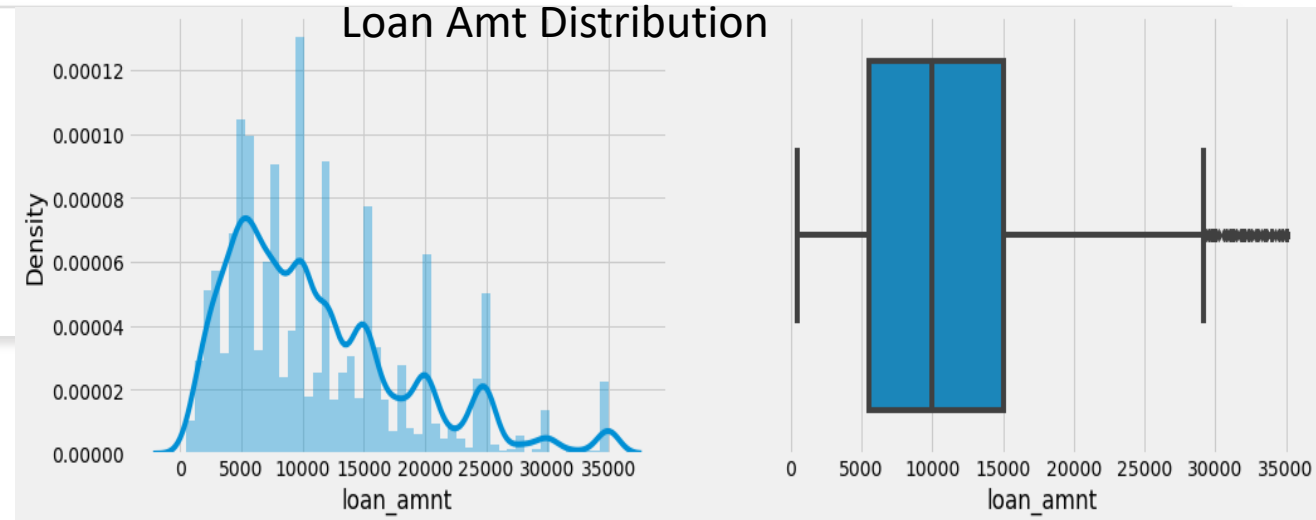
Bi Variate Analysis


Loan Amt Distribution
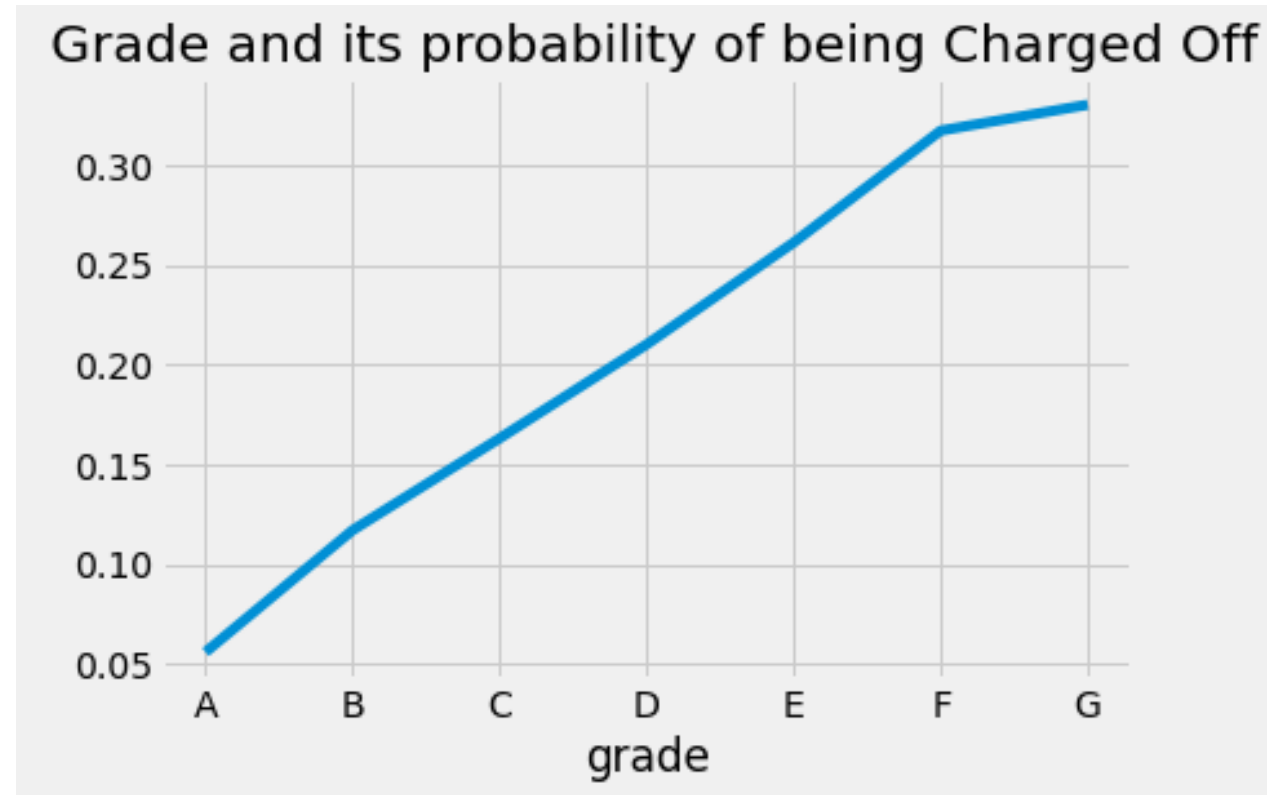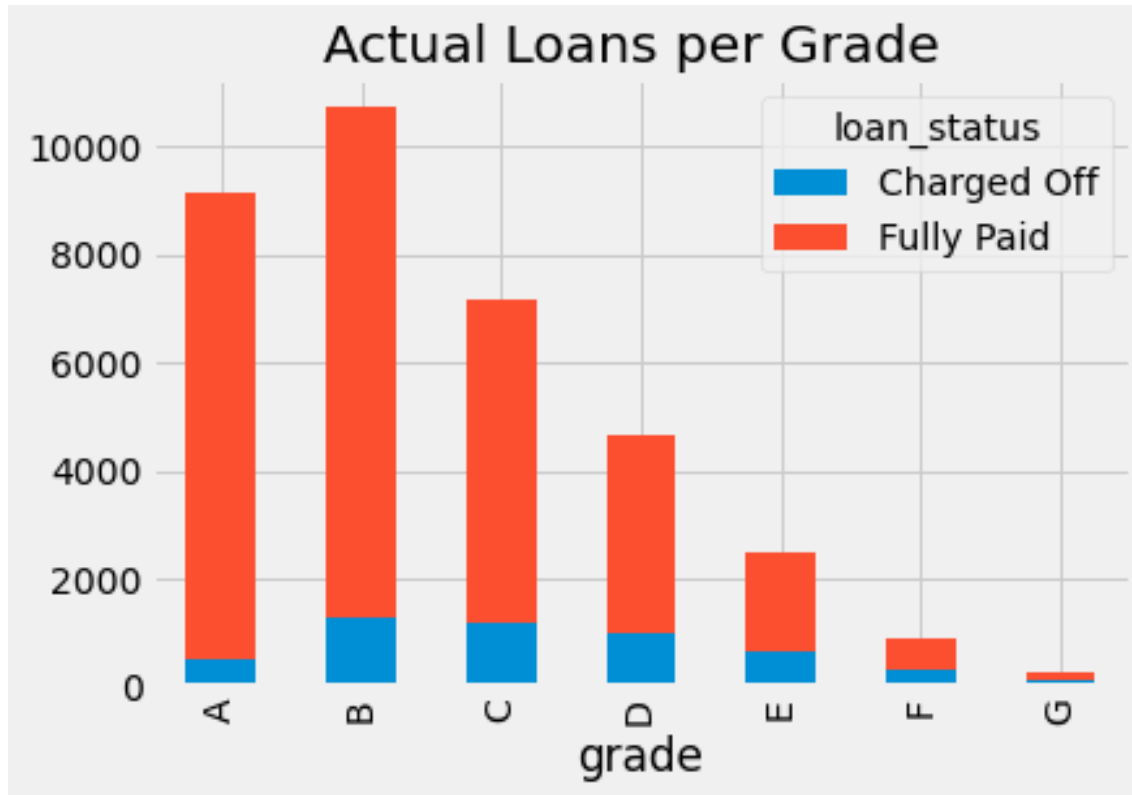
Verification Status with Loan Amount – Loan Status

- Verification is done for higher amounts but defaults are higher at larger amounts too.
- LC should continue doing the same
- However Verification process should be tightened further as Verification is not providing right results

Bi Variate Analysis

Actual Loans per House Ownership
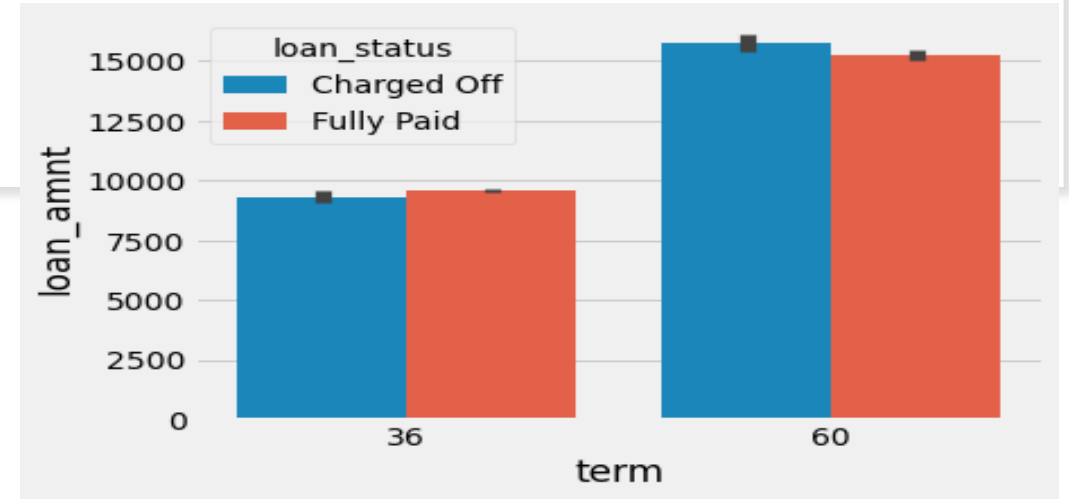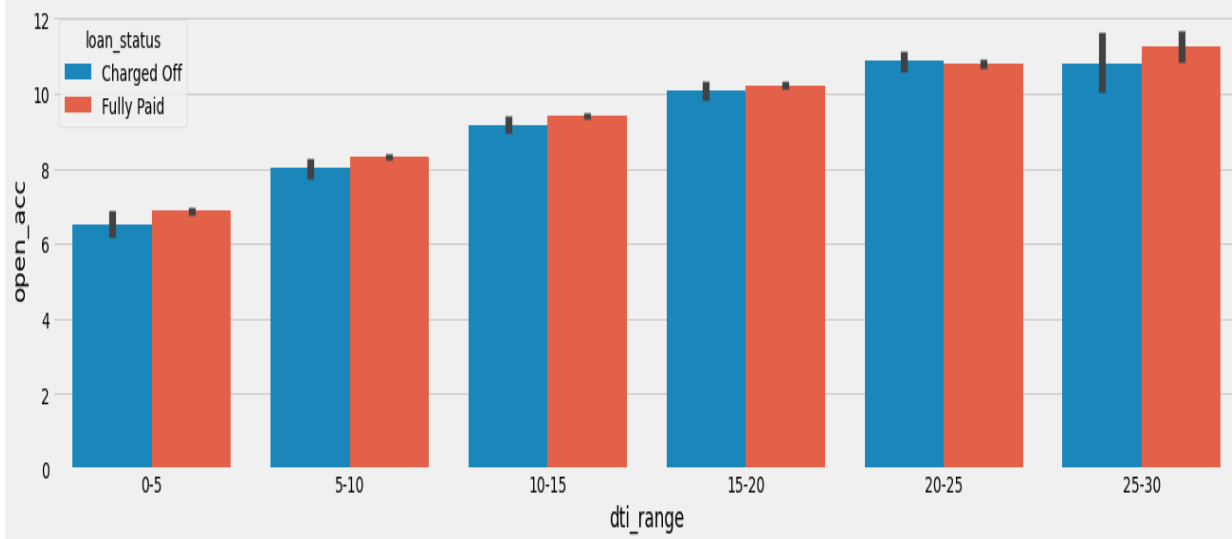

House Ownership and its probability of being Charged Off
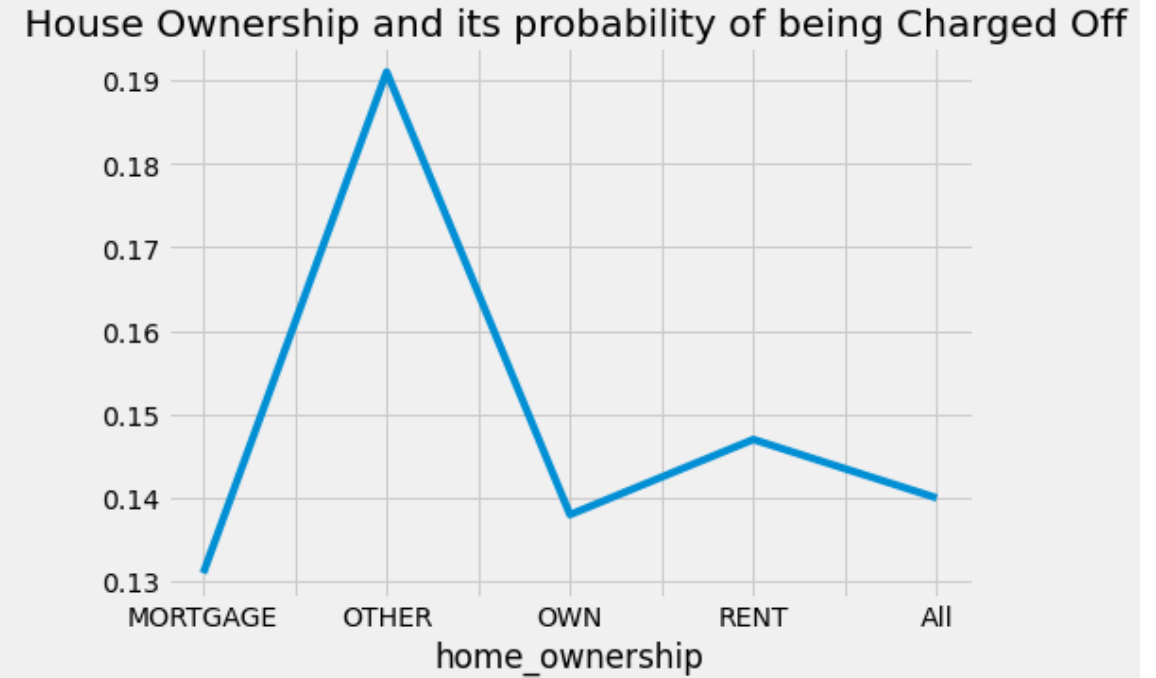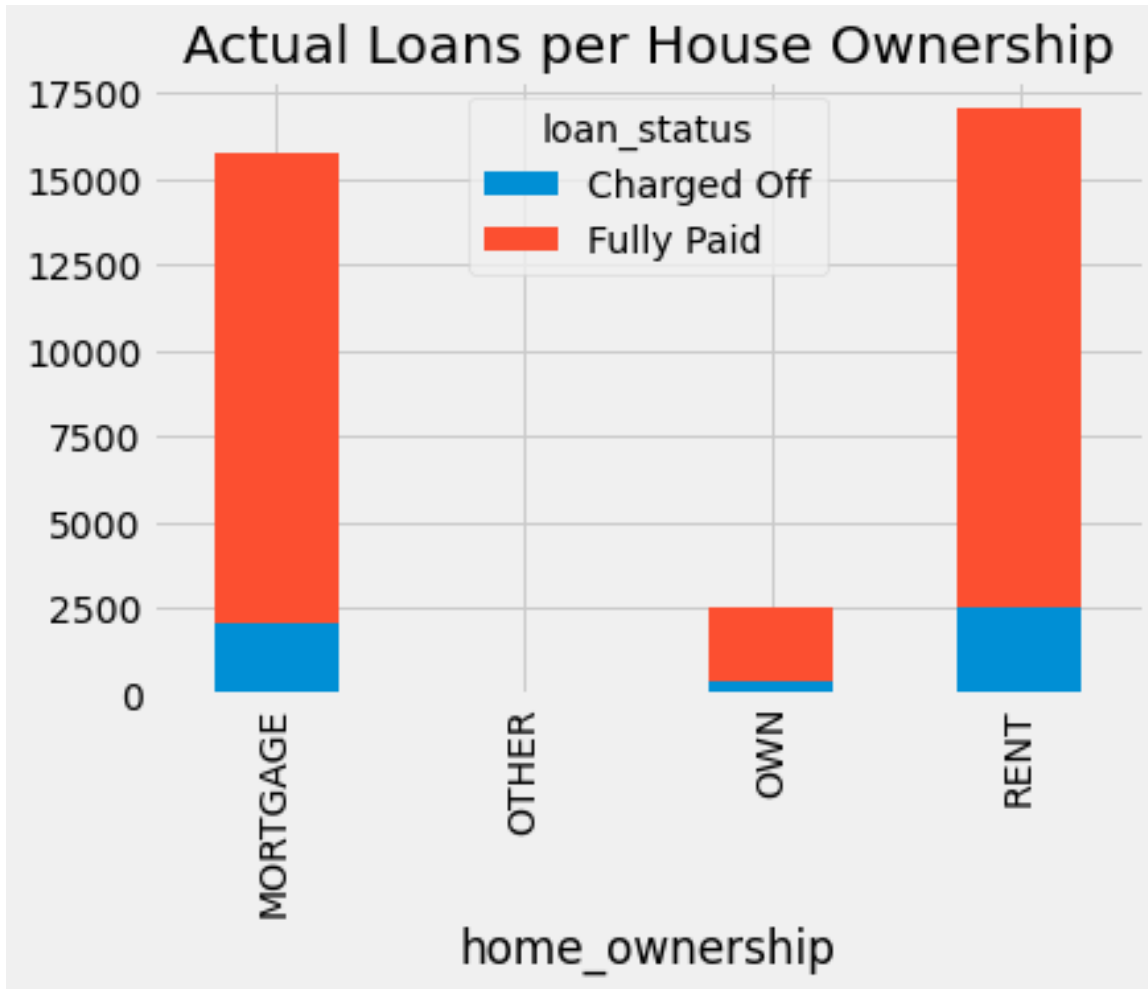
House Ownership and Probability of Default

- There is a bias towards Rented and Mortgaged Loan Seekers
- LC should prefer Mortgaged Owners over Rented ones as the probability is marginally lower on Mortgaged Seekes
- Other Category should be strictly avoided

Bi Variate Analysis
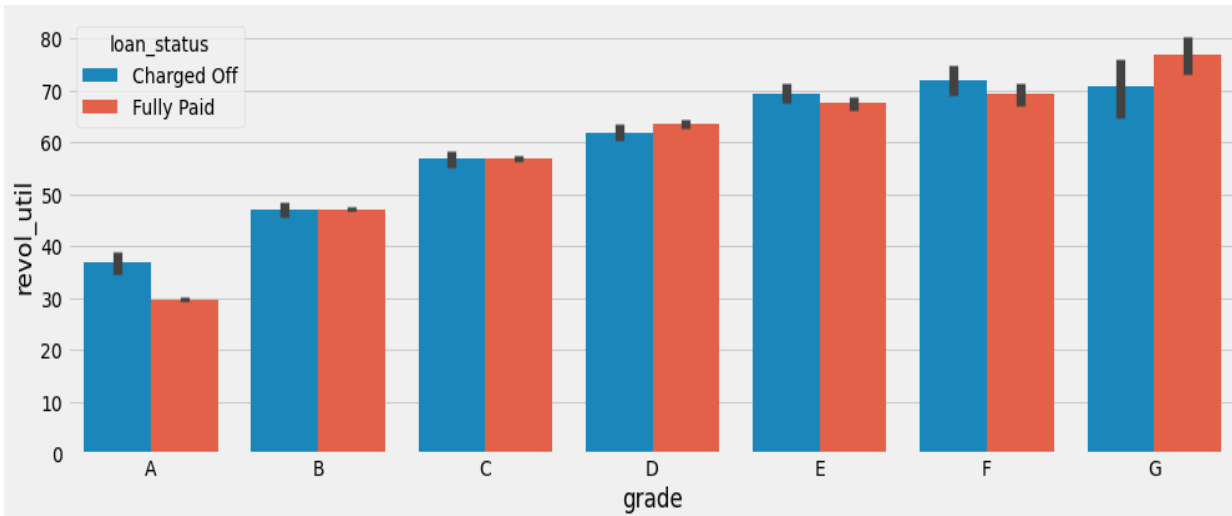
Loan Amount, Revol Util, Grade, Open Acc and DTI range

- Higher loan amount is associated with longer terms and high defaults.
- Grade decreases as revolving credit line increases.
- DTI increases as the open accounts increases

Analysis Recommendations

Continue Doing

- Continue Charging Higher Loan Rates
- Continue giving loans to Tenured Employees
- Continue Lending Irrespective of Months
- Continue the Grading System
- Continue Verification for Higher Value Loans
- Continue Preferring Rented/Mortgaged Owners

Start Doing

- Avoid Higher Terms of 60 months
- Consider higher margin on Riskier loans
- Reduce lending where purpose is "Small Business"
- LC should restrict lending to applicants having DTI >17
- Avoid grade categories E,F,G
- Tighten the Verification process which has holes
- Avoid Other Category Home Owners