Lending Club Casestudy

Presentation

Sameep Mishra (Group Facilitator)
Saurav Ujjain (Group Member)

Project Objectives

Project Brief

We analyze the historical loan data of a large online loan marketplace specializing in lending various kind of loans to urban customers.

Problem statement

We analyze the historical loan data of a large online loan marketplace specializing in lending various kind of loans to urban customers.

Data Understanding, Cleaning and Manipulation

Identify and correct quality issues , Interpret the correct meaning of variables (Columns)

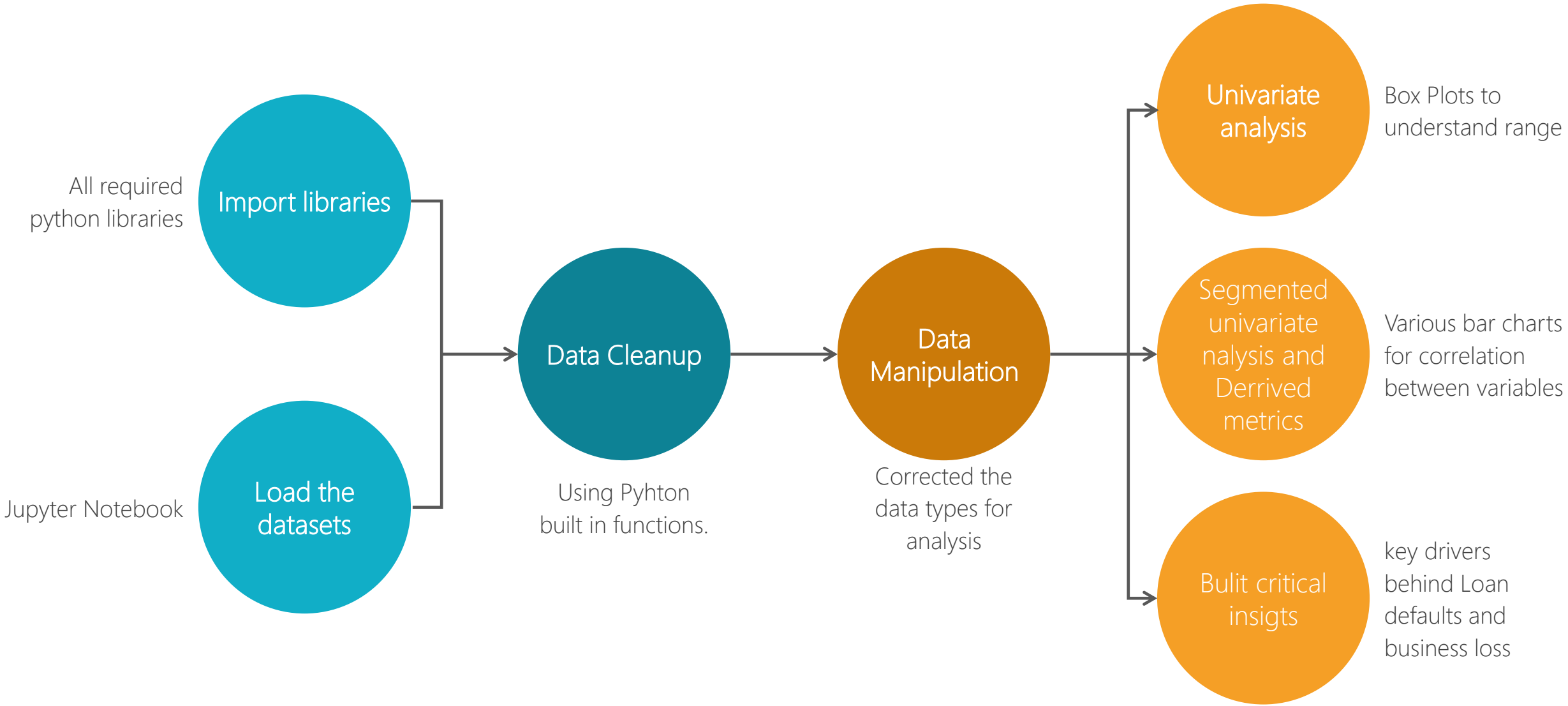
Data Analysis

Univariate & Segmented univariate analysis, Derrived metrics Bi and Multi variate analysis with plots to draw business insights

Presentation

summary in a presentation format , explaining the insgihts drawn with clear actionable suggestions for improvement

Project Progress



Step 1

Imported Required Libraries

Python Libraries

1. *Numpy: Math library*
2. *Pandas : To work with dataset*
3. *Matplotlib: to plot some parameters*
4. *Seaborn: Graph library that use matplot in background*

Step 2

Loaded Datasets

Jupyter Notebook

‘Loan.csv’ was loaded into a jupyter notebook to commence analysis

Step 3

Data Cleaning and Manipulation

Data Cleaning

Remove all the up blank columns

Identify the columns with more than 50% Null values , remove them

Found how many unique values are in these columns for further filtering

We deleted all the columns with just 1 (Unique) value as they would not add any value in analysis.

Identified duplicate or very close Columns, Deleted the duplicates.

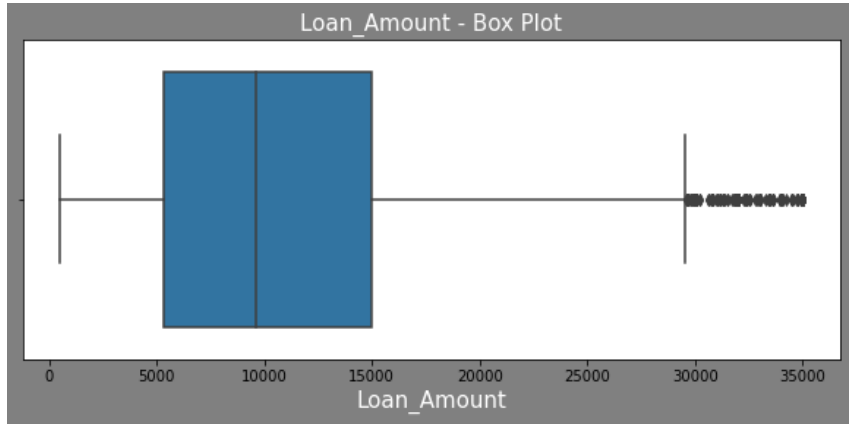
Data Correction

- Removed Months from term column
- Rmoved non number characters from emp_length column
- *Removed non number characters from ‘int_rate’ & ‘revol_util’ columns and change to int type*
- *Changed 'NONE' entry in 'home_ownership' with 'OTHER'*
- *Replacing Blank employer name value with Unknown*
- Replacing NAN value for public bankruptcies with 0 bankruptcies
- *Only the 'Fully Paid' and 'charged off' loans are relevant to analysis from the ‘loan_status’ column. So we deleted all the rows with 'current' loans*

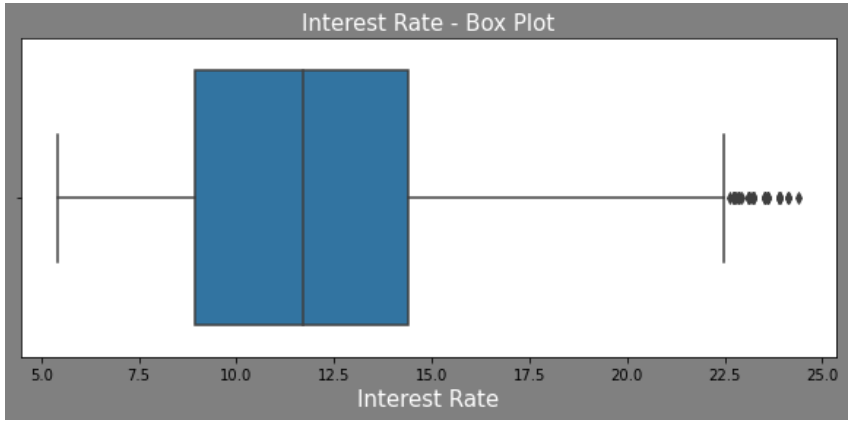
Step 3

Univariate Analysis

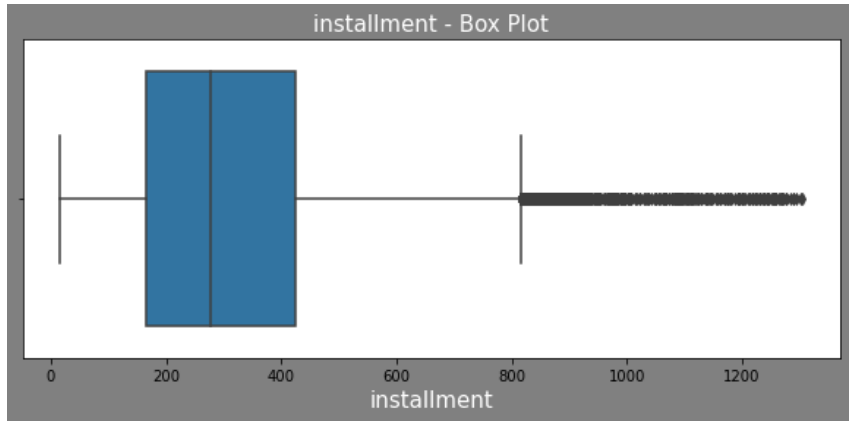
Box Plots



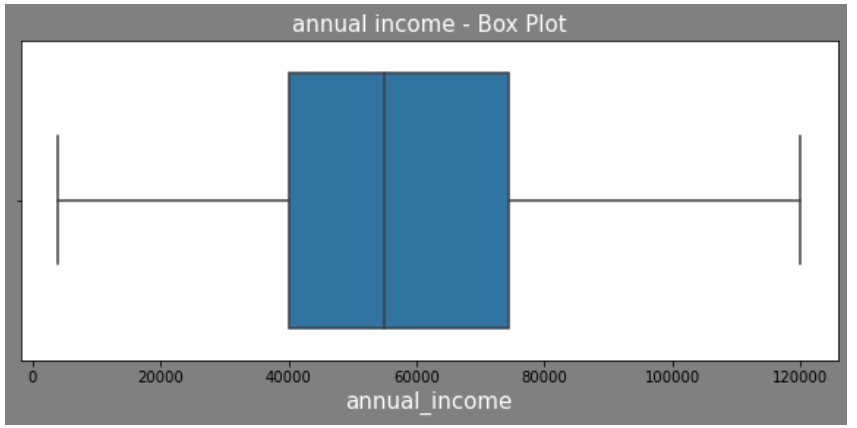
Majority of loans issued are in the range of USD 5000 and USD 15000 with the median loan amount a little less than USD 10,000



Majority of interest rates (Inter quartile range) issued are in the range of 9% to 15% and the median rate is at 12%



Majority of iinstallments(Inter quartile range) fall in the range of USD 200 to USD 400 range and the median rate is around USD 250

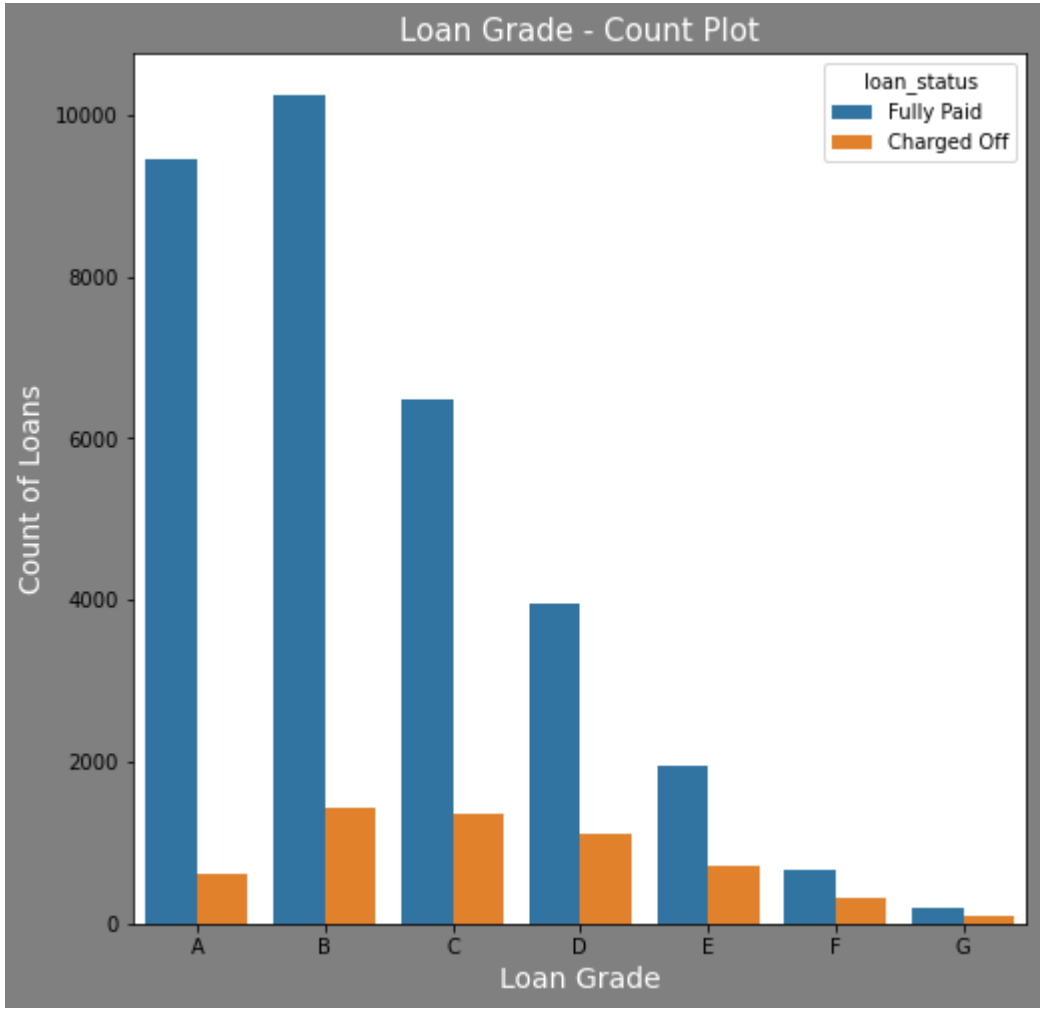


Majority of yearly income(Inter quartile range) fall in the range of USD 40000 to USD 70000 range and the median income is around USD 55000

Step 3

Segmented Univariate Analysis

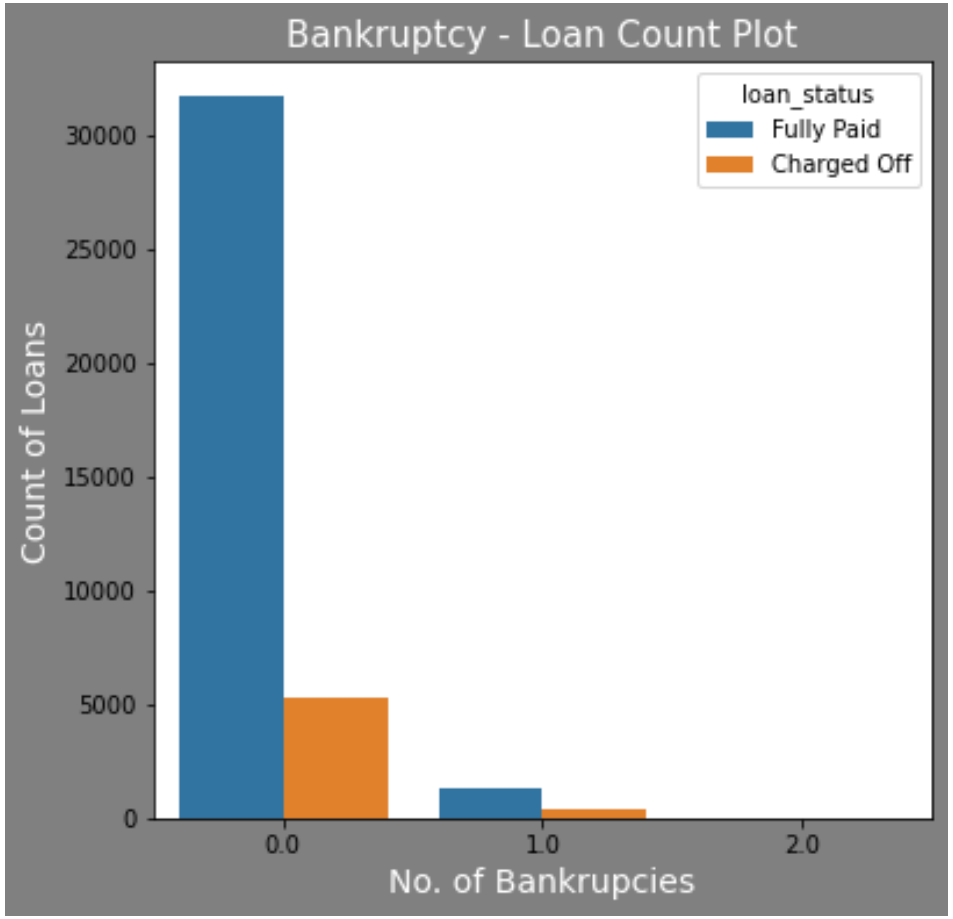
Categorical Variables



Loan Grade with Loan Status

We can see that a relatively higher ratio of Grade C (approx 15%) of total loans are being charged off compared to (approx 10%) grade A loans.
Though it's clear that the default rate is much worse as the grade of loan goes lesser with almost half (50%) of Grade E,F and G loans going into default.

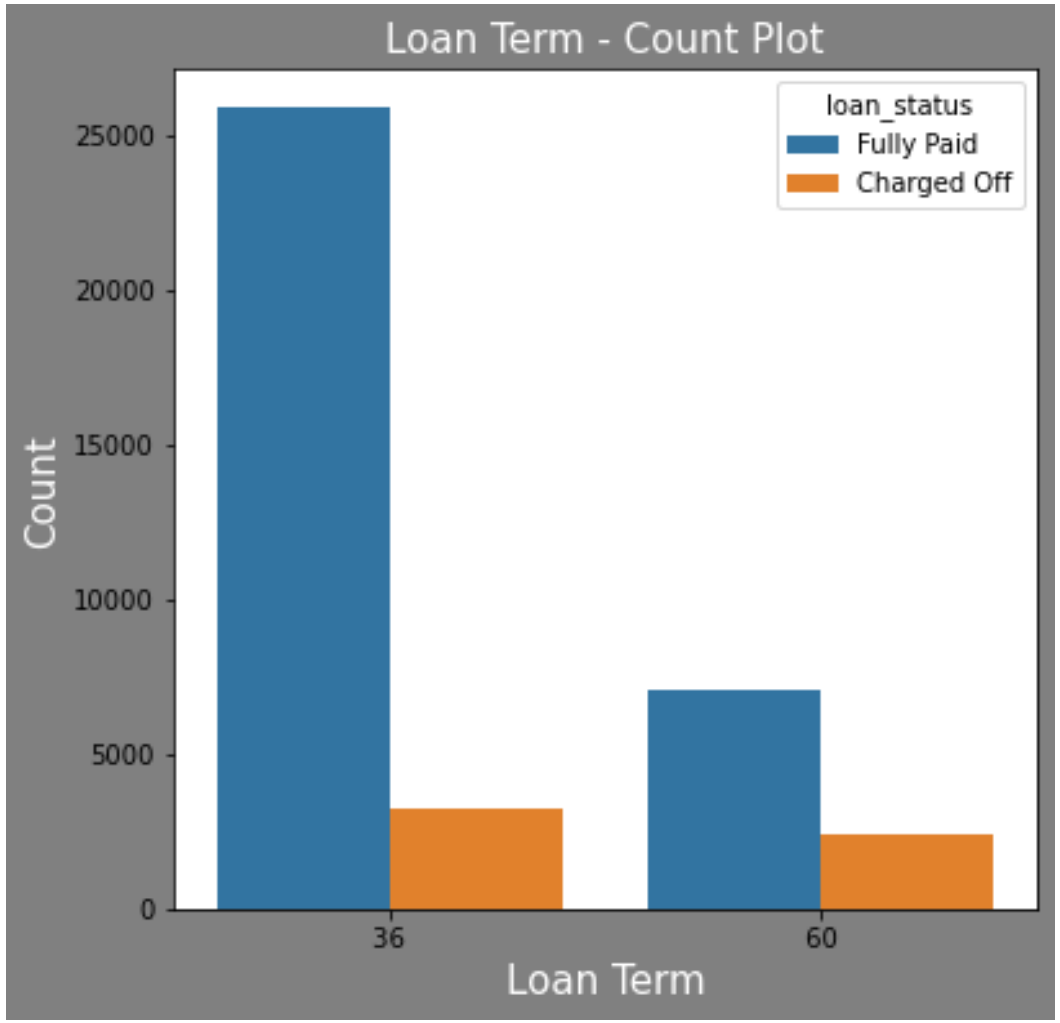
Categorical Variables



Declared Bankruptcy with Loan Status

There are hardly any loan given to people with declared bankruptcies.

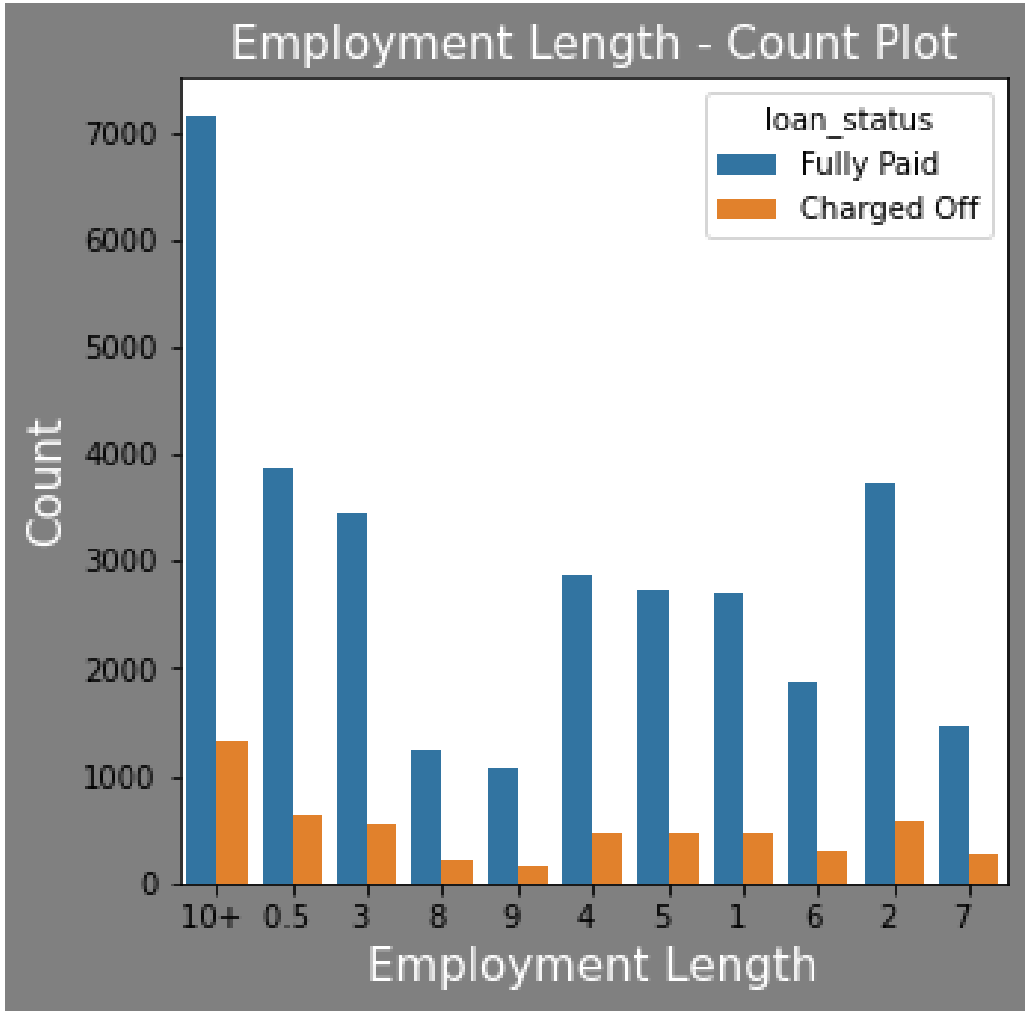
Categorical Variables



Loan term with loan status

Rate of defaults are significantly higher for shorter 60 month loan terms

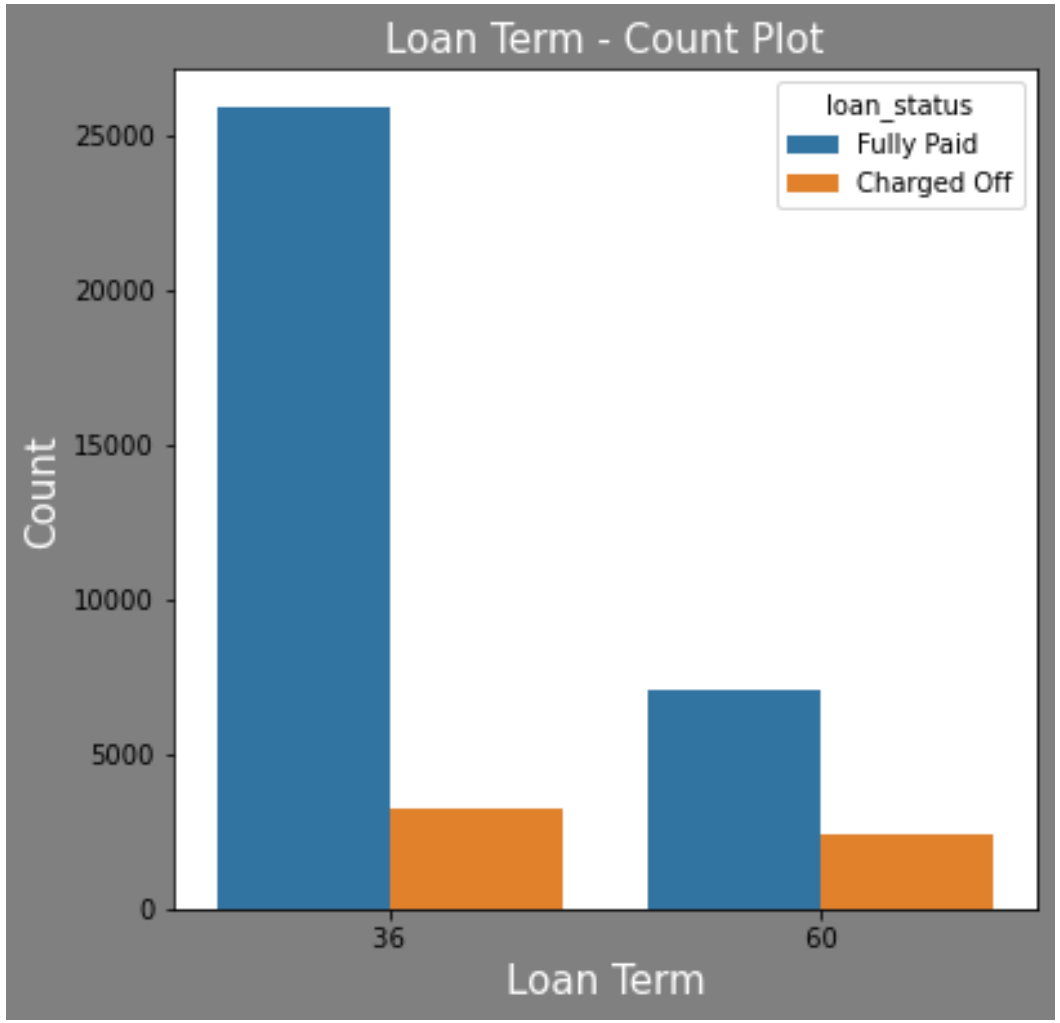
Categorical Variables



Employment duration with Loan status

Rate of loan default is clearly lower for people with 10+ years of employment, whereas it is higher for 0-5 years range

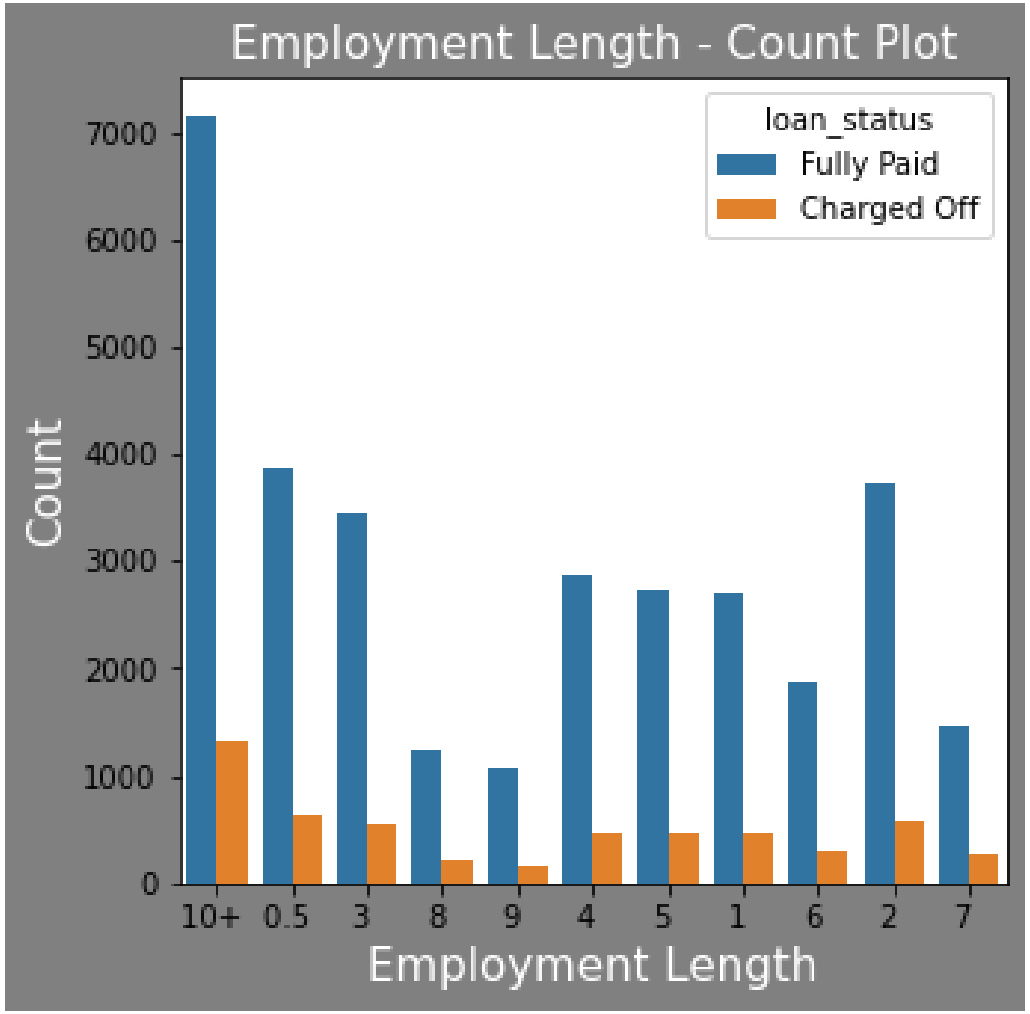
Categorical Variables



Loan term with loan status

Rate of defaults are significantly higher for shorter 60 month loan terms

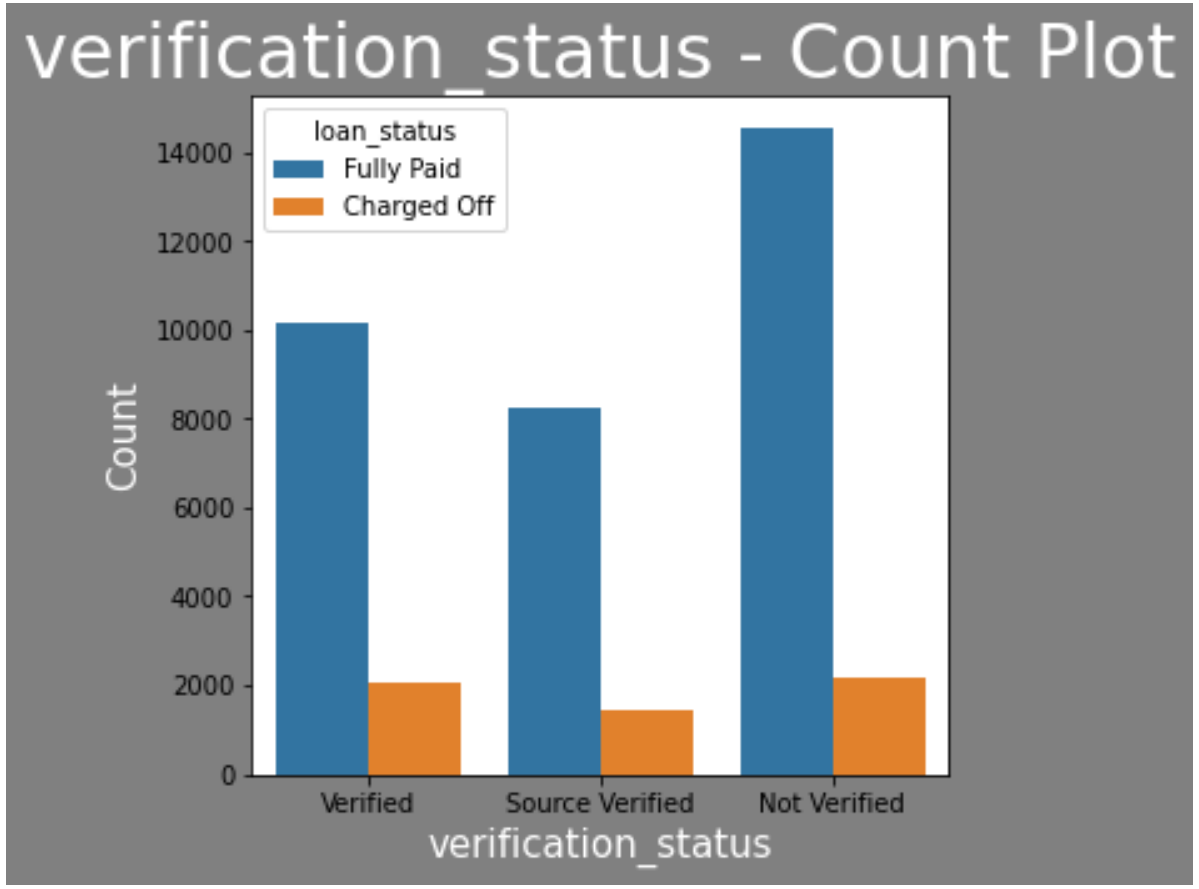
Categorical Variables



Loan purpose analysis with loan status

Most of the loans are for direct Debt consolidation or indirect debt conslidation (Credit_card)

Categorical Variables



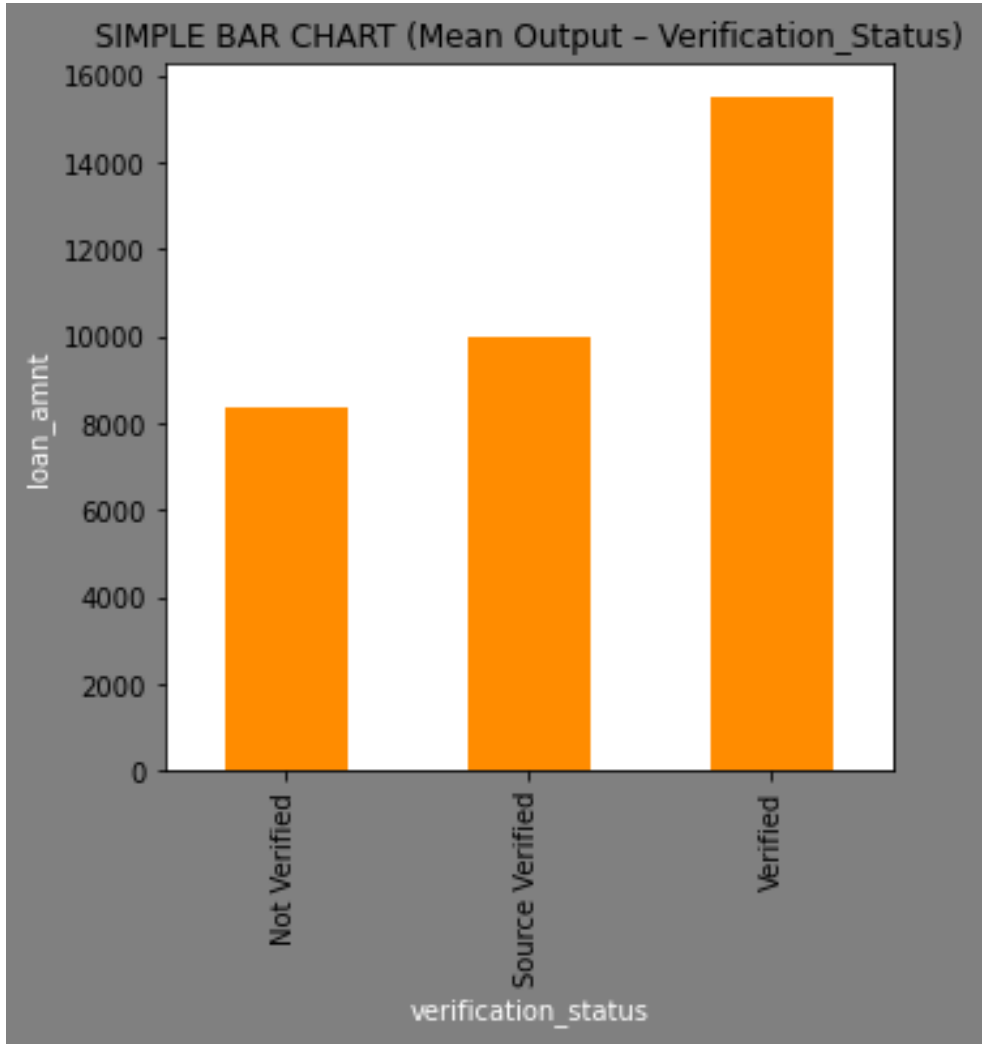
Loan verification analysis with loan status

The above bar chart shows comparison between 'Not Verified' & 'Verified', the risk portion has been balanced and the verified applicants get significantly higher avg loan value then not verified.

Step 3

Bivariate Analysis

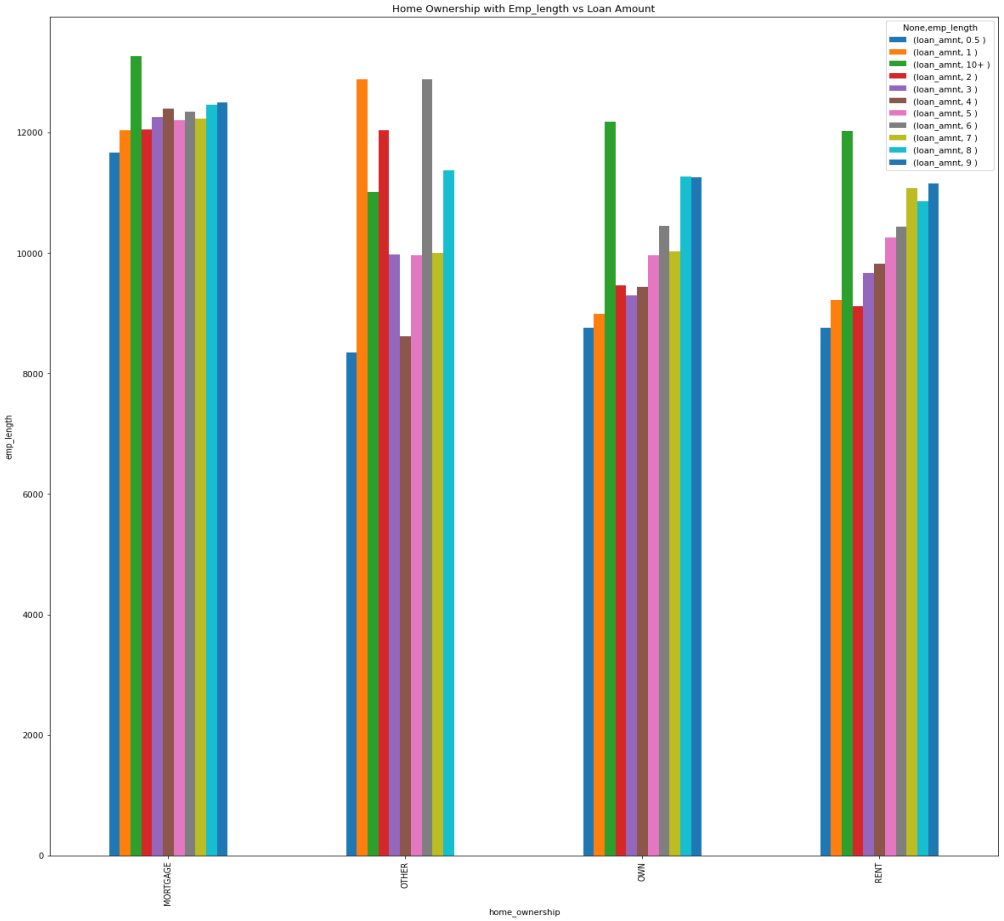
Bivariate Analysis



Amount of money invested based on customer verification status

The above bar chart shows comparison between 'Not Verified' & 'Verified', the risk portion has been balanced and the verified applicants get significantly higher avg loan value then not verified.

Multivariate Analysis



Loan distribution with home ownership vs experience.

Below Multiple Bar Chart has Loan distribution with home ownership vs experience.

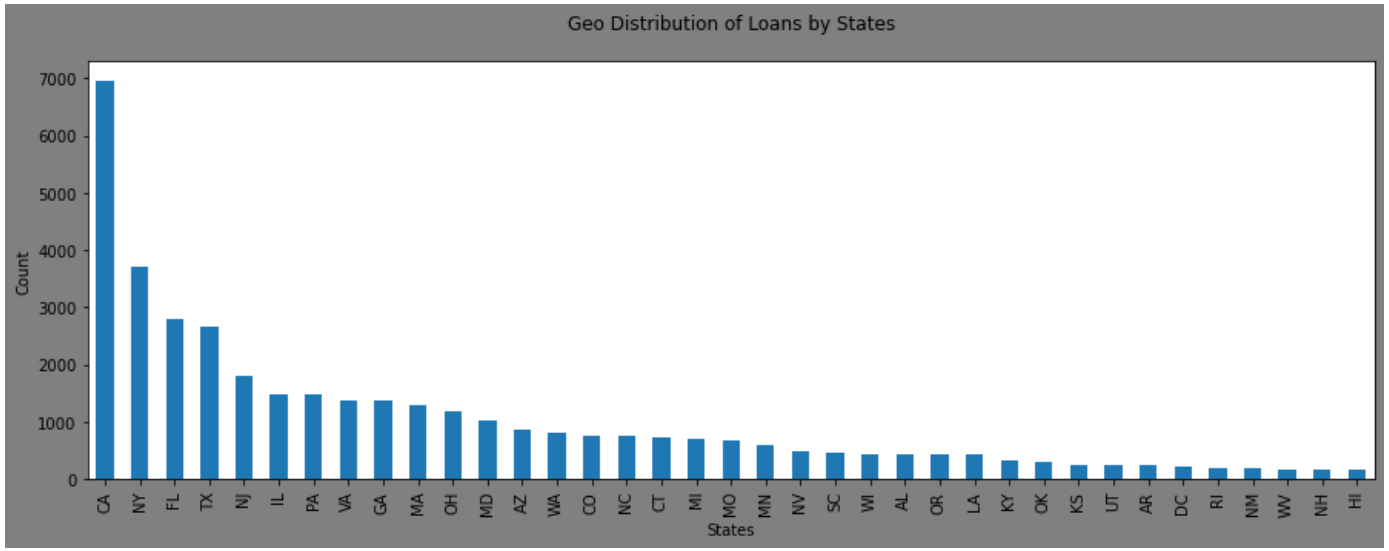
1. The loan distribution is higher on people with Mortage, high risk.

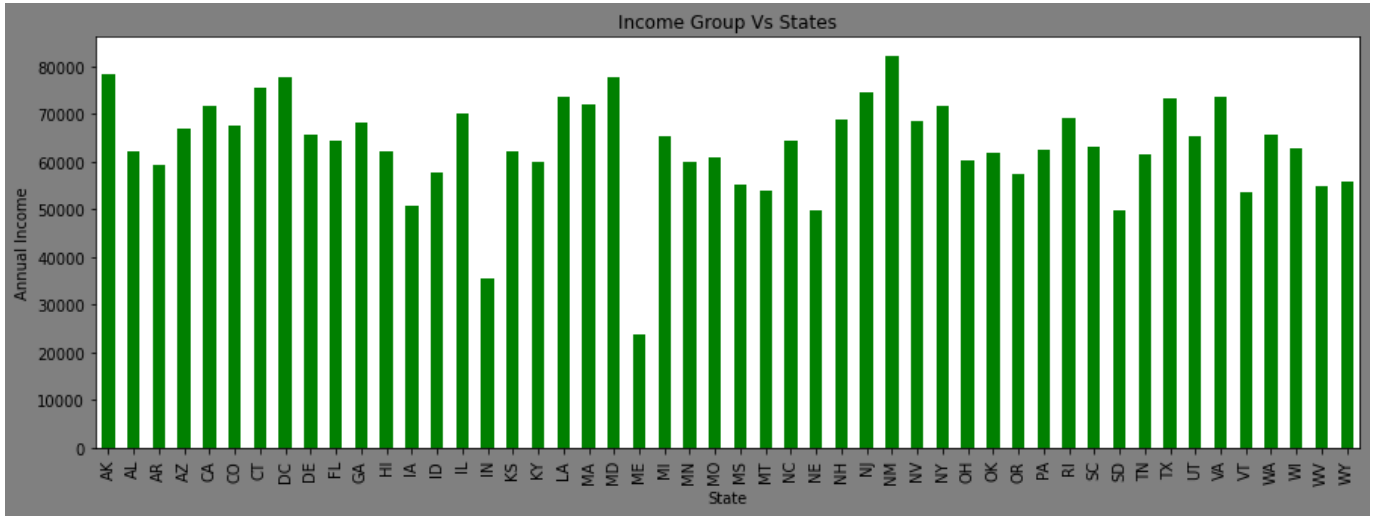
- Customer with existing obligation(Mortage), it puts additional burden to customer during repay.

2. Instead, we see there a huge opportunity space available on people with Own property.

- They are always reachable as their properly state is stable.

Multivariate Analysis



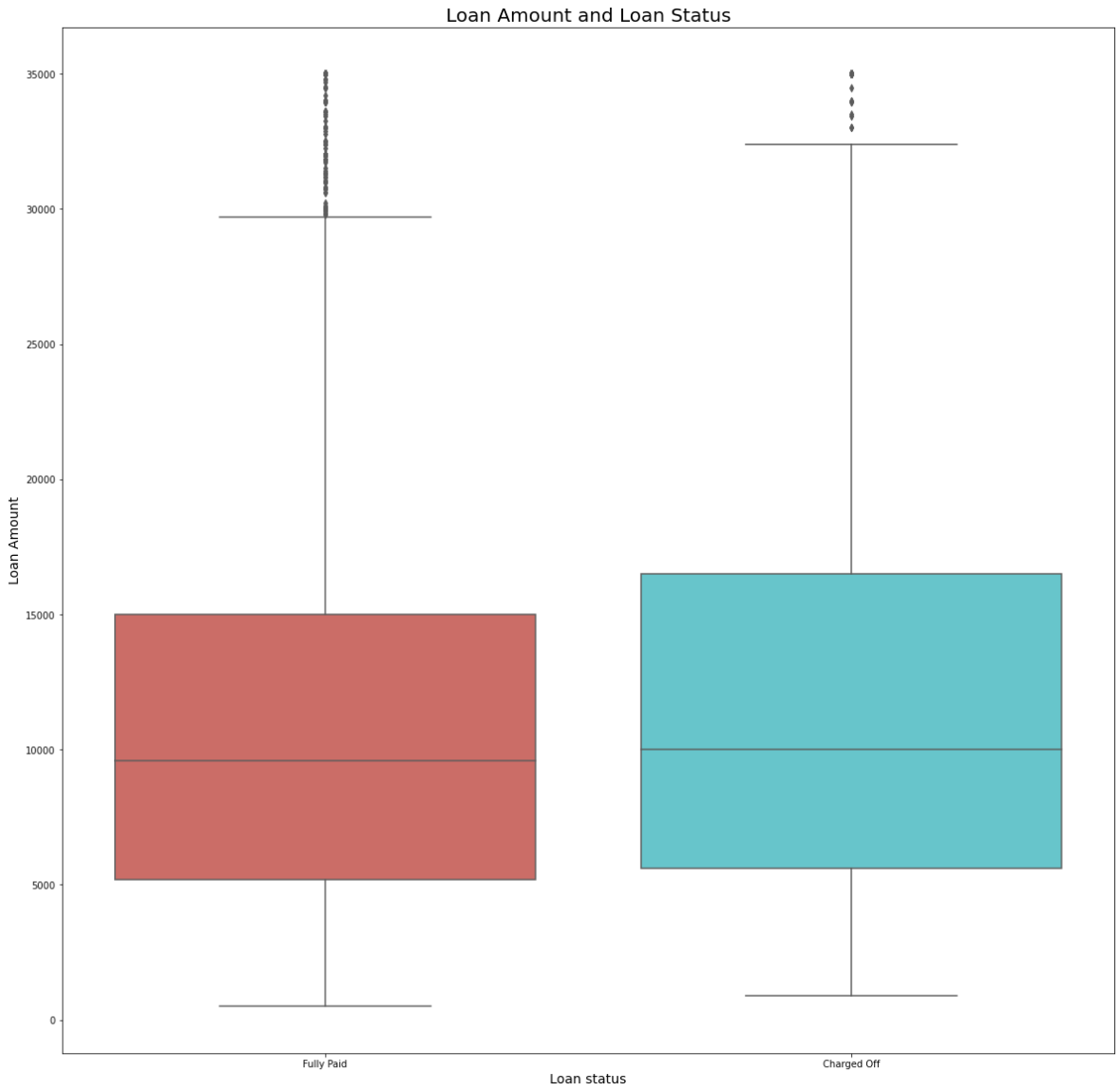


Geo Distribution of states shows, the majority of customers are in CA

Derived Business Insight:

Organization should consider expanding in other states as there is not much variation in annual income between CA and other states which is a good indicator of repayment ability

Multivariate Analysis

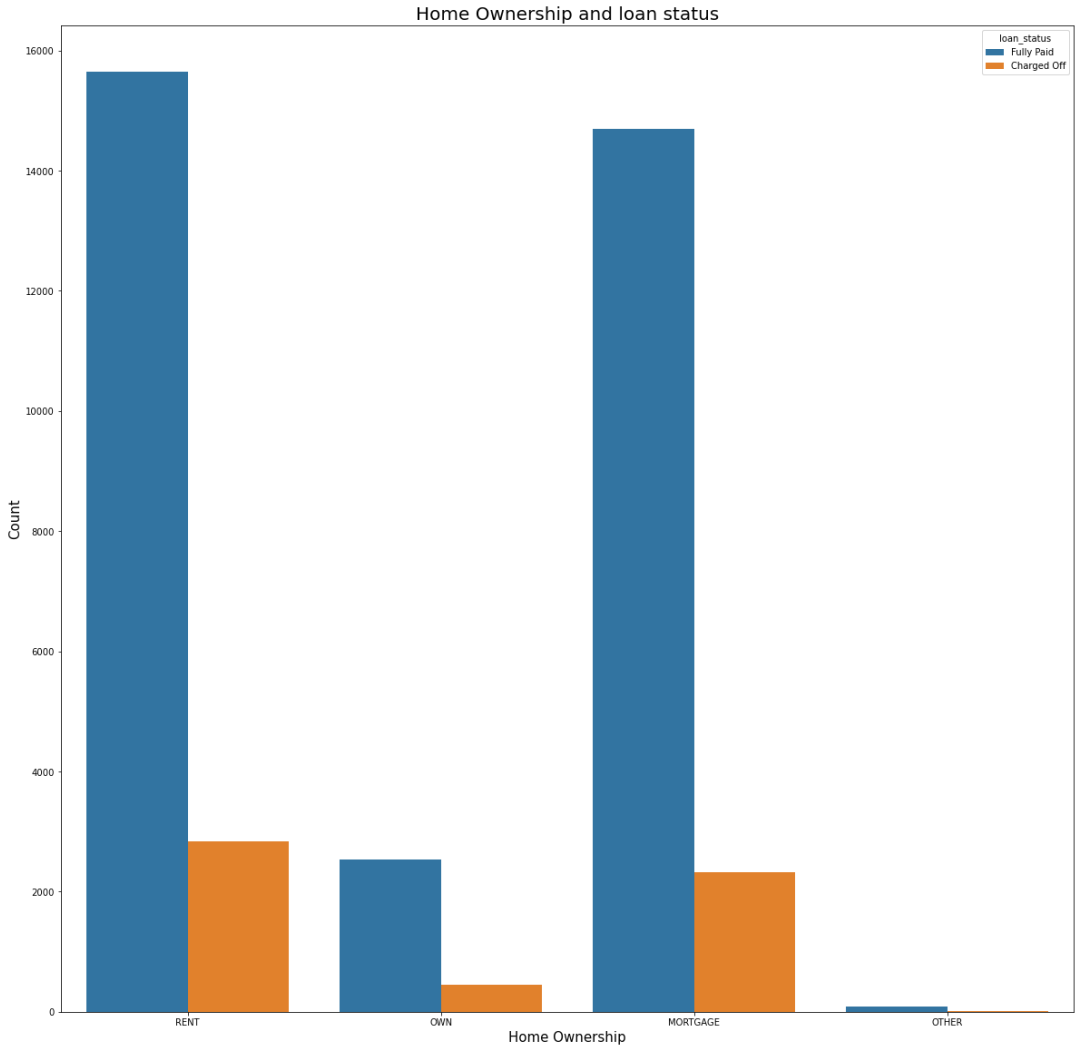


Box Plot comparing the historical data with current¶

Insight:

There is not much to choose from between the average loan amounts which gets paid and gets charged off as both of them are venely spread with close medians

Multivariate Analysis

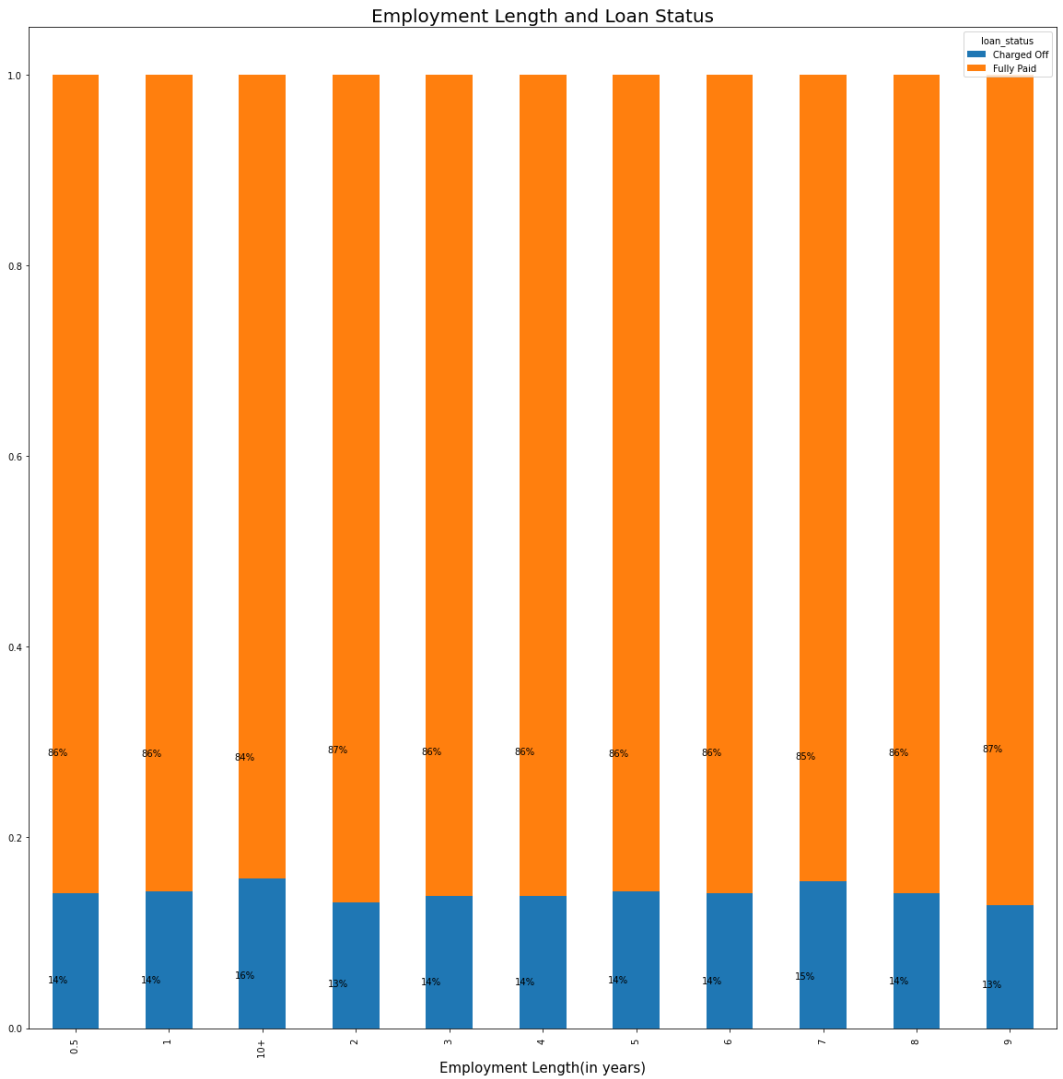


Home ownership with loan status¶

Insights:

1.People with Own homes like to default very less.
2.People with Rent are high on risk.
3.Sales focus should be more on people with Property own followed by Mortages
4.Multiple sales techniques like lower interest rate for people with own property to increase the loan distrubution in this section.

Multivariate Analysis



Stacked Bar Chart shows comparison of experience aginst Loan status¶

¶

insights:

It indicates the default percentage ranges around 15% on all experience buckets.