

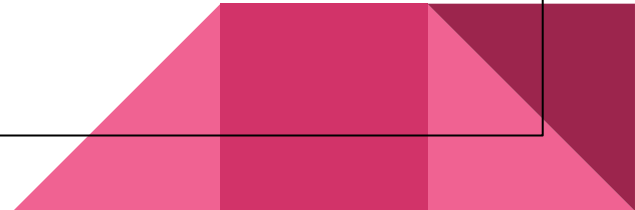
Lending Club Case Study

Anil Turlapati

Objectives

The objective of this case study is to apply the EDA concepts acquired and apply to the real world data

- This helps in basic understanding of data , analyzing of data
- Feel of real world data and data cleansing activities
- Learn to understand data and how it can help business and change the perspective
- Visualization of data with the help of python libraries



Problem statement

Provide a resolution on approve or reject loan applications based on the variables or dimensions available

Dataset Details:

The data contains information about data from past loan applicants the status of 'defaulted' and paid. It has three main status of loan which is Fully Paid, Current and Charged-Off. For the current case study we can ignore current status as this is still in process and could not be useful for evaluations



Data process

Understanding of Data

Removal of Null values and
Duplicate values

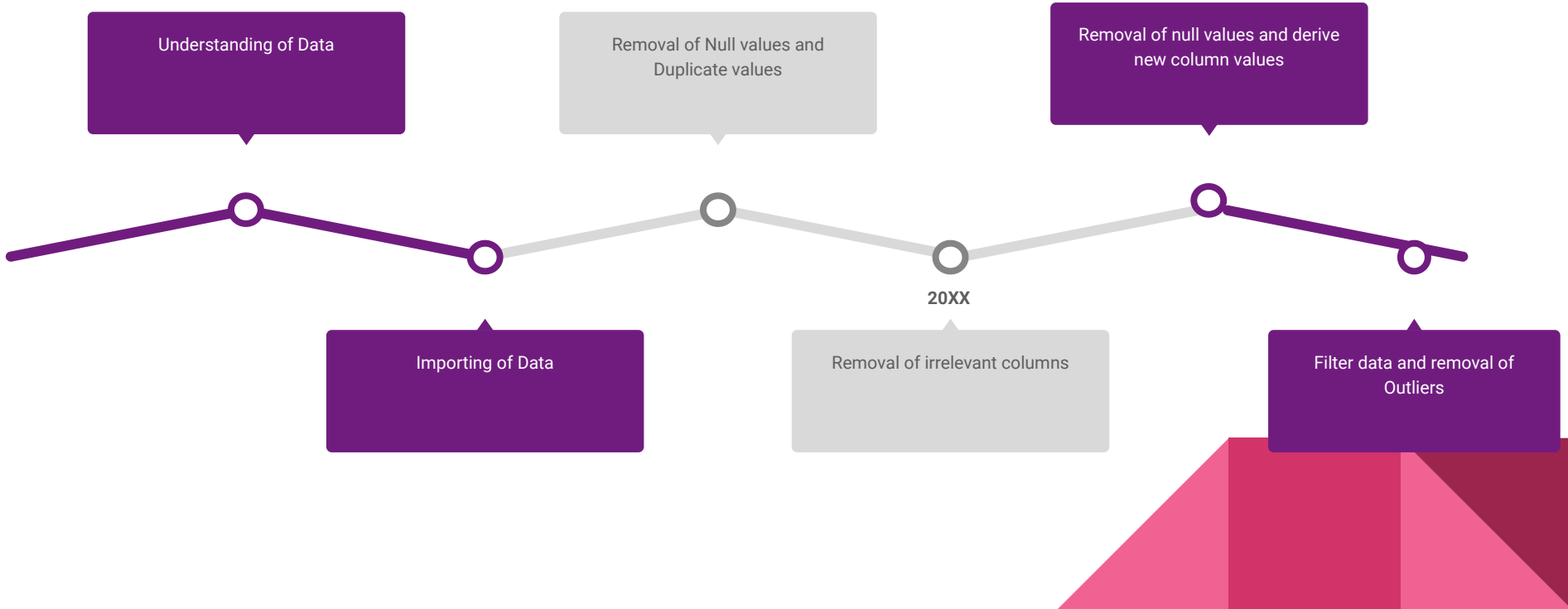
Removal of null values and derive
new column values

Importing of Data

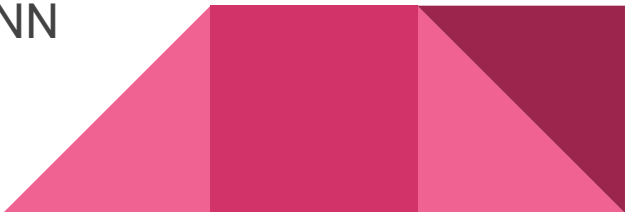
Removal of irrelevant columns

Filter data and removal of
Outliers

20XX



Data understanding

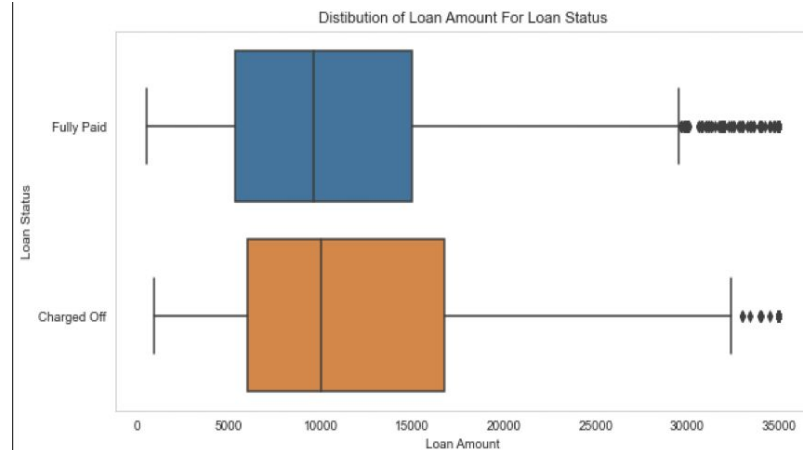
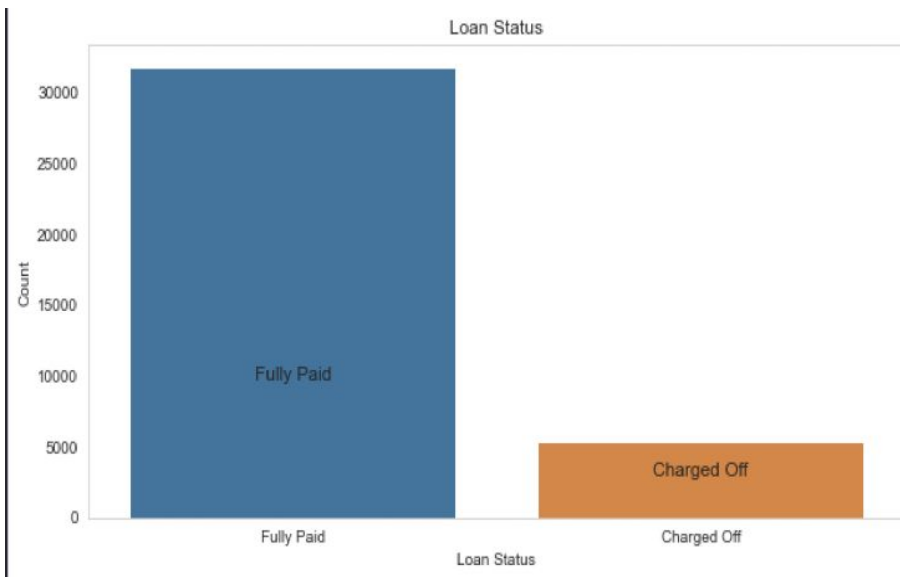
- Initial data is of 39717 records of data with 111 columns
 - The data types are with mixture of Int and float data types
 - The data set does not have any duplicate records
 - The set has huge number of missing values and unique values which has to be dealt with
 - Majority of data has irrelevant columns and majorities of columns with null values
 - Assumption made is to remove columns with null values and add values for required data variable using different methods like KNN
- 

Data cleaning

- As a part of data cleaning activity , initially the duplicates are checked and removed
- The columns with unique and null values based on % has be checked
- Deleted the columns where the null values are more than 50% as this skew the analysis
- Based on the case study, we do not require all the columns and so deleted the column not required for analysis(ex: description)
- Next would be deletion on unique value columns with less count

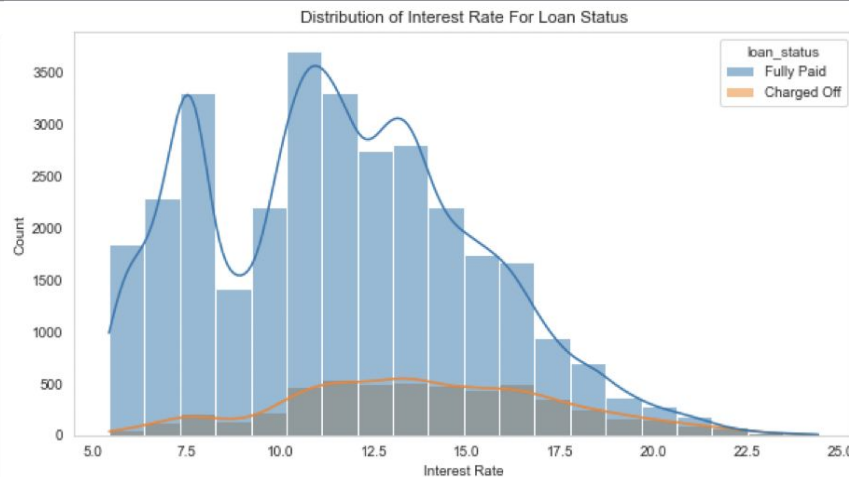
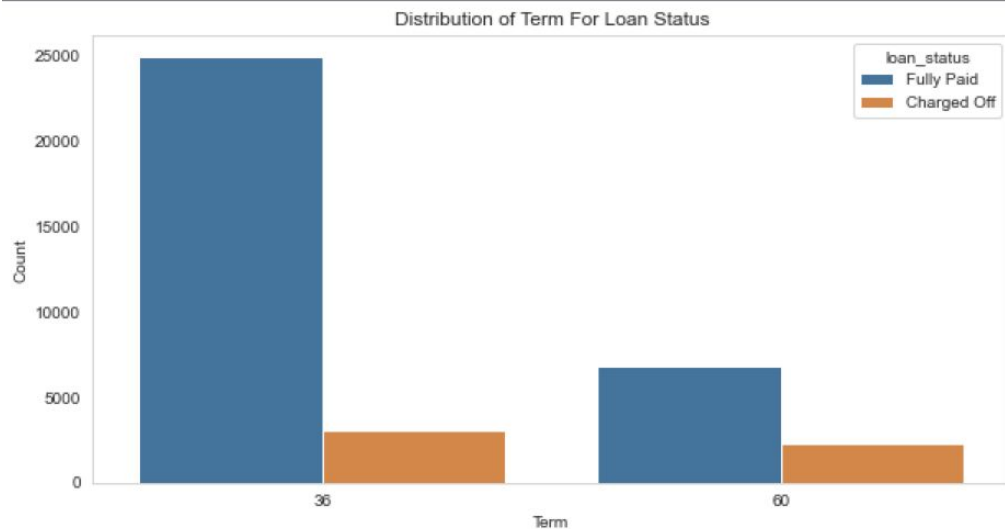


Data Analysis - Loan Status vs Amount



The Defaulted loan is way smaller than the paid off loan,
Fully paid loans are 7 times to Defaulted loans

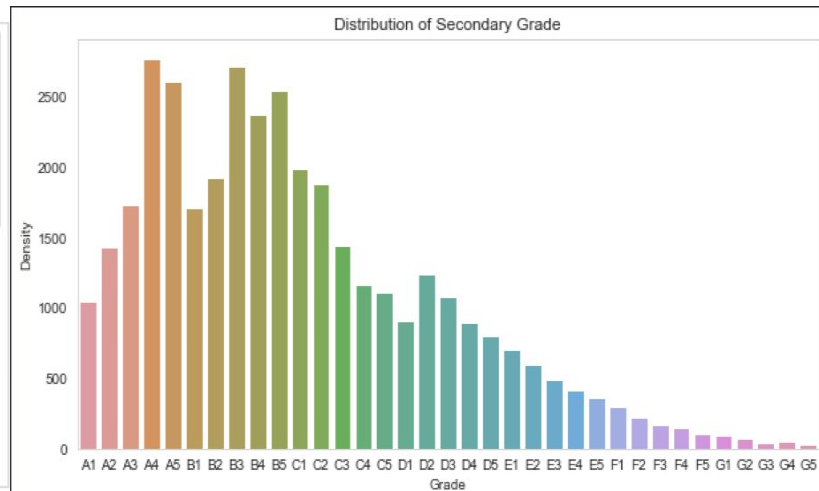
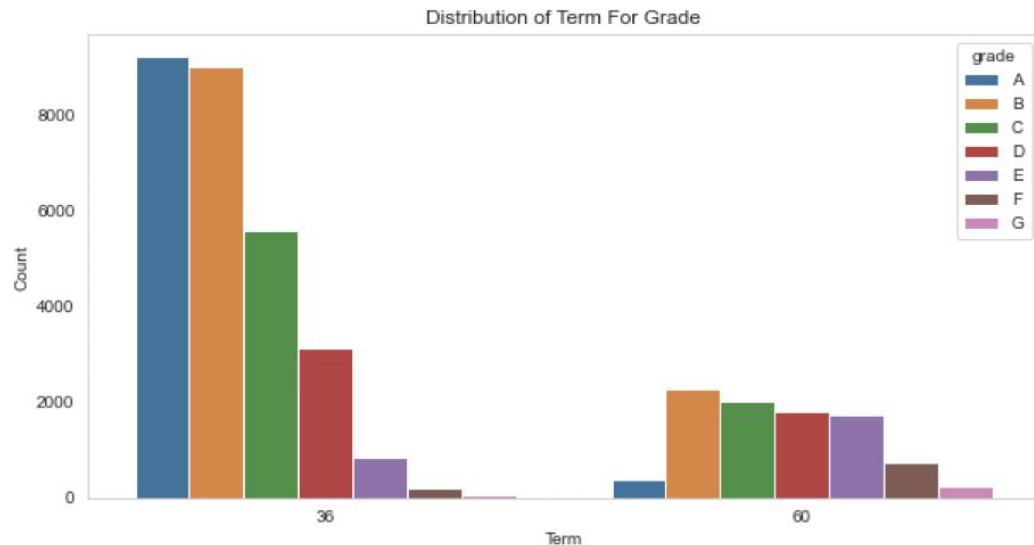
Term and Interest rate



majority loans applied for 36 months than 60 months

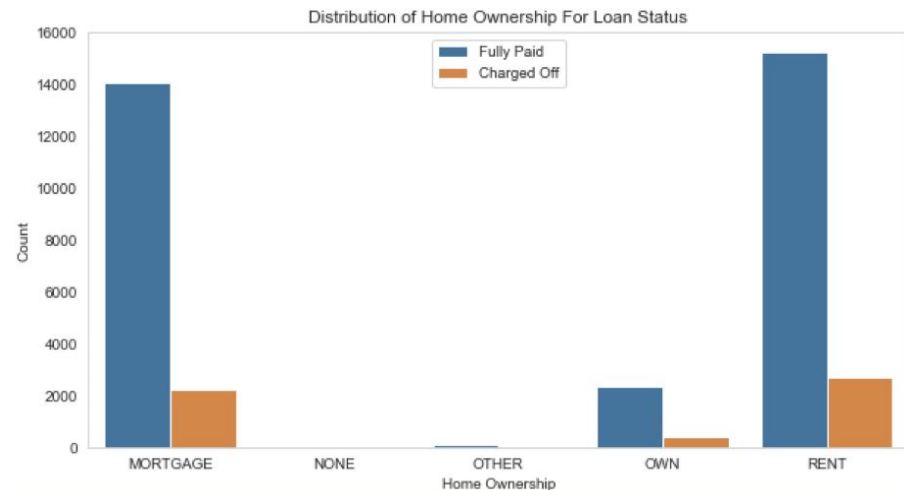
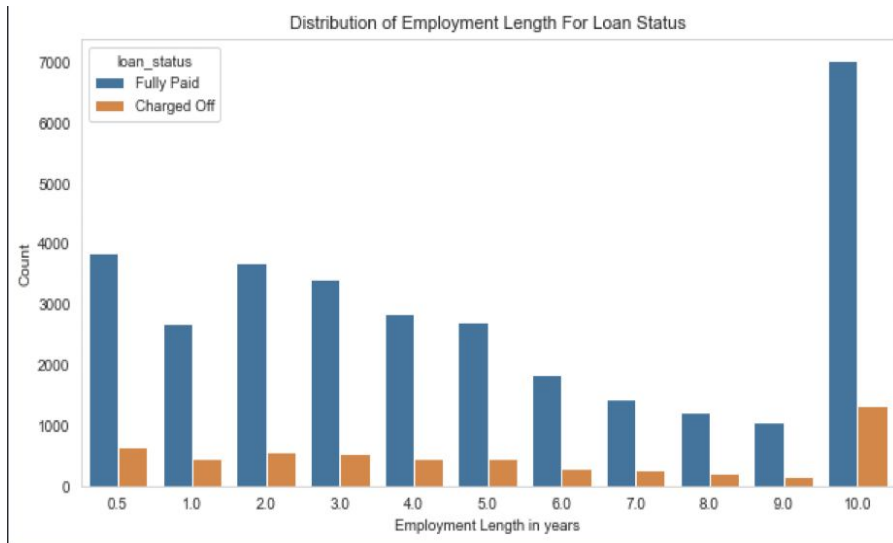
The interest rate is more crowded around 5-10 and 10-15 with a drop near 10.

Grade and subgrade



A large amount of loans are with grade 'A' and 'B' compared to others.

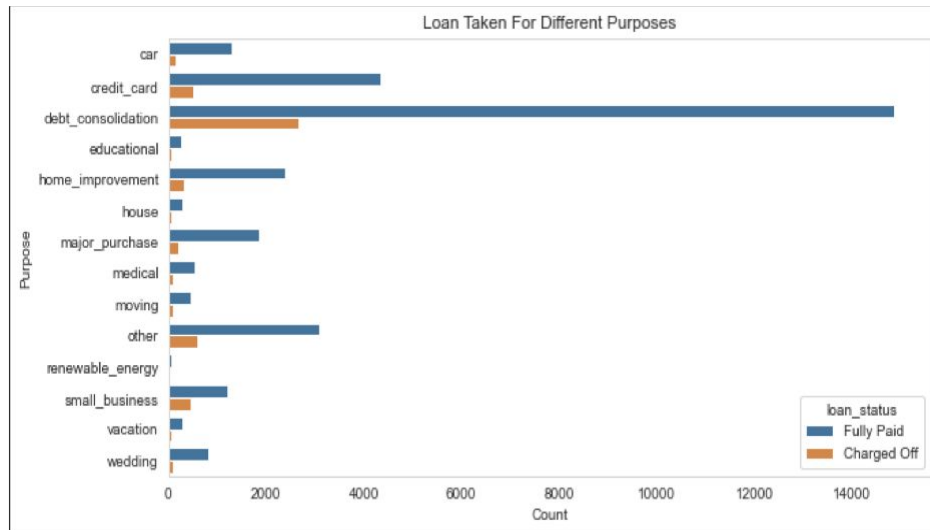
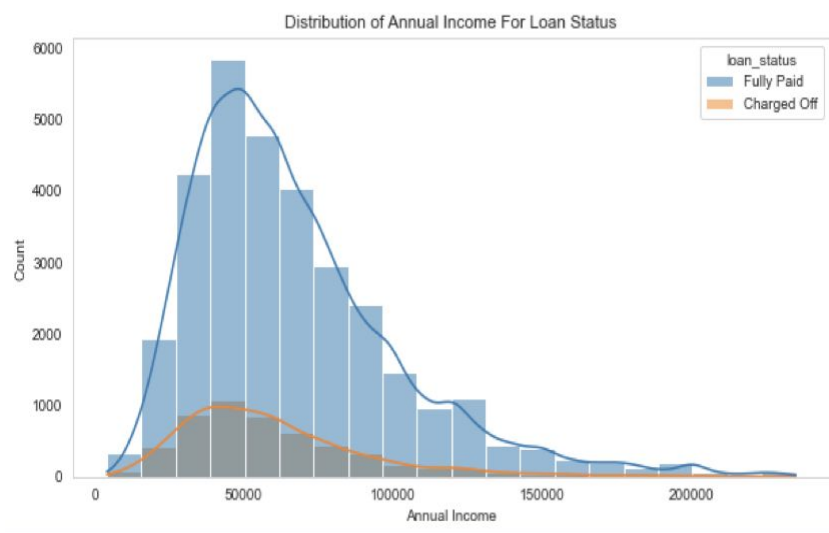
Employment length and home owner



Majority of Lenddes have working experience more than 10 years.

Majority of Lendees are on Rent residency or on Mortagages with few having own houses.

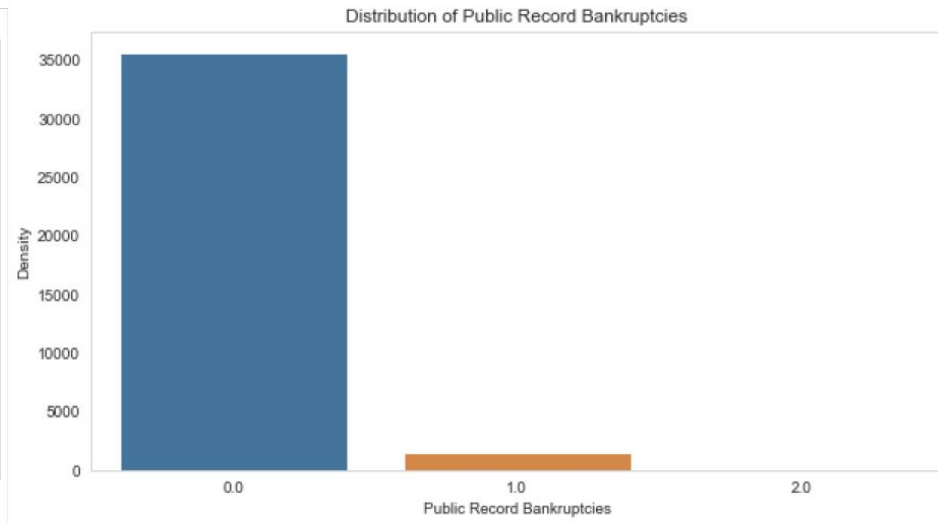
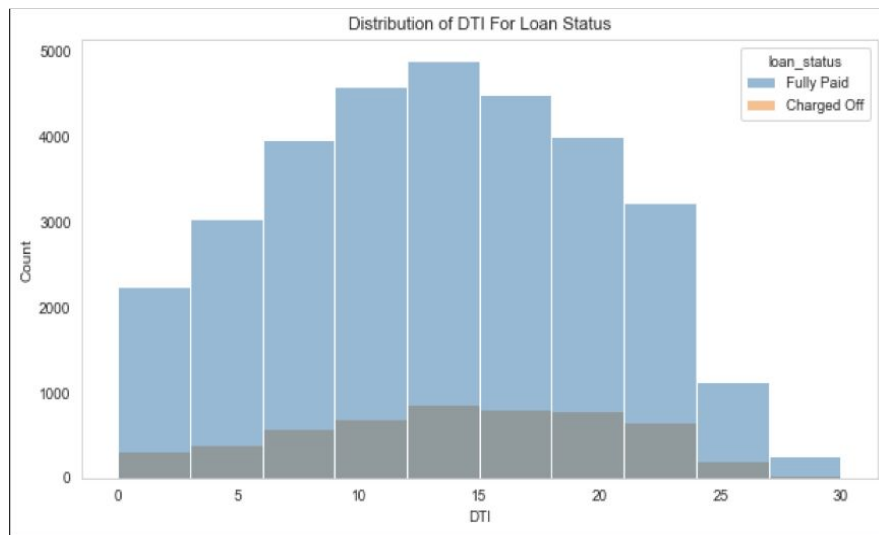
Annual Income and purpose of the loan



About 50% of the Lendees are verified by the company or have source verified.

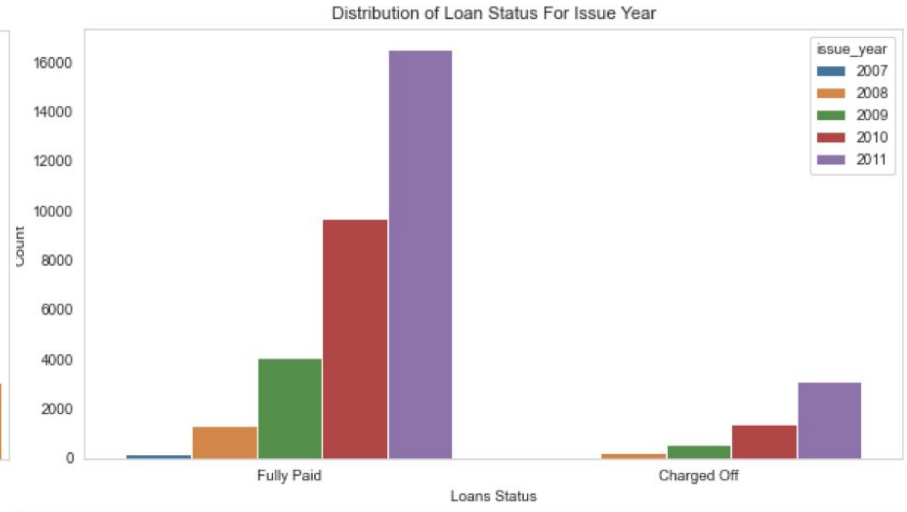
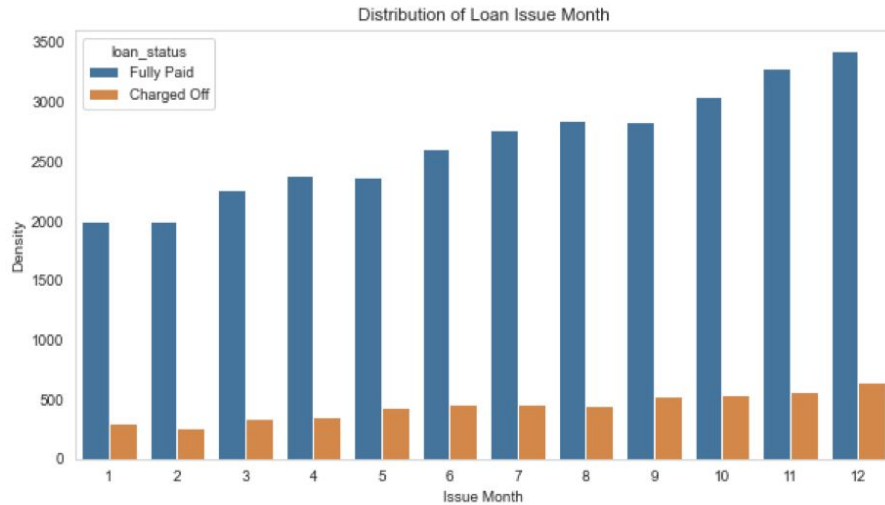
Annual Income is left skewed giving indication that majority is earning low annual income

Debt to Income with Bankruptcy



Majority of lendees has high DTI which is a risk and at the same time lendees who are not filed Bankruptcy has high rate of repaying the loan amount

Loan tenure for longer term



A gradual increase in loan approved through the year has lesser defaulting rate in April ,August, December quarter wise and better more late in year.

As passing year loan taken are increasing exponentially which indicate we are seeing large increase in DTI ratio and decrease in defaulting rate.

Impacts and Recommendations

Below factors has lot of impact on the repayment of the loans, these impacts should be considered

- High loan amount and at the same time interest rate for individuals with lower income group has possibility of defaulting
- High installment with longer repayment term i.e loan tenure or duration is risky the shorter the duration the better chances of repayment
- Home ownership plays critical role and risk free
- loan purpose and timing of the loans usually Q4 applied loans are for previous loans consolidations and
- Income group and other important factor to be considered is Employment tenure



Impacts and recommendations

Other factors to be considered while giving loans are

The Lendee current Debt to Income and existing loans

The employemnt status and his Annual income level or band

The current interest rate at which the loan is applied and with duration.

