Lending Club

Case Study

Problem statment

Risk Assessment of loosing money while lending to borrowers.

- The online loan company has to decide whether to lend the money to applicant or not based on historical data collected

Associated Risks are:

- Company should not miss on possible profit by rejecting the loan application for the borrower who is likely to repay the complete loan
- Company must not lend to the borrower with high likelihood of default
- In both of the above scenarios company will suffer financial loss.

Analysis approach

Data:

- Dataset containing information about previous lenders with approved loan approval is provided in a CSV file.
- Target column that is our column of interest will be the column named loan_status where there are 3 types of loan status as mentioned below:
  - **Charged-off** – Borrowers who have defaulted on their loan.
  - **Fully-paid** – Borrowers who have fully paid back the loan.
  - **Current** – Borrowers still paying their loan and not defaulted yet.
- Besides the target column there are other columns present in the csv file which we will use for univariate and bivariate analysis along with the target column that will help us to make a conclusion that which borrower could possibly default and which borrower will fully payback the loan amount.

Data Cleaning And Manipulation

## Steps involved in Data Cleaning:

- We will be dropping all those columns which have null values more than 40% from the dataset.
- After executing step 1 above we will be checking the percentage of null values for each column and if the percentage of null value is less than 5% then we will drop all the rows with any null value.
- For the loan_amnt column we will also be plotting a boxplot and then calculate the upper whisker value and then drop all the rows which have loan_amnt values above the upper whisker.

Steps Involved in Data Manipulation:

- We will be converting all the required columns into their correct datatype, for eg the int_rate column has % appended after the values so we will remove the % and convert the int_rate into float datatype.
- We will also extract month from issue_d (loan issue date) column and create a new column which will contain the month in which loan was issued.
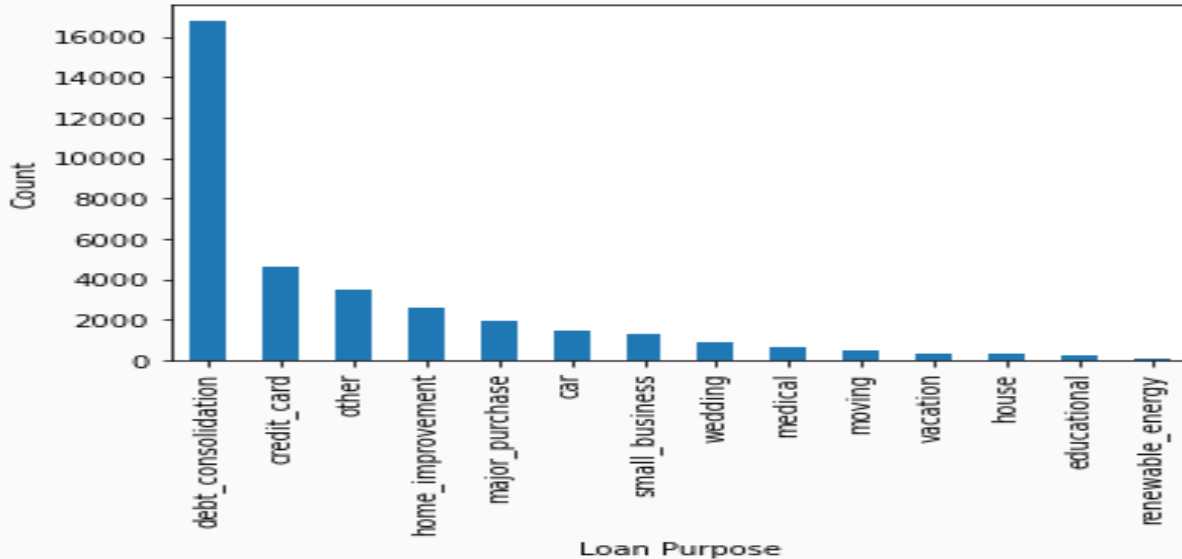
Univariate Analysis

This Bar Chart tells us that most of the rows in the csv file for loan_type column contain Fully Paid value followed by Charged Off followed by Current.
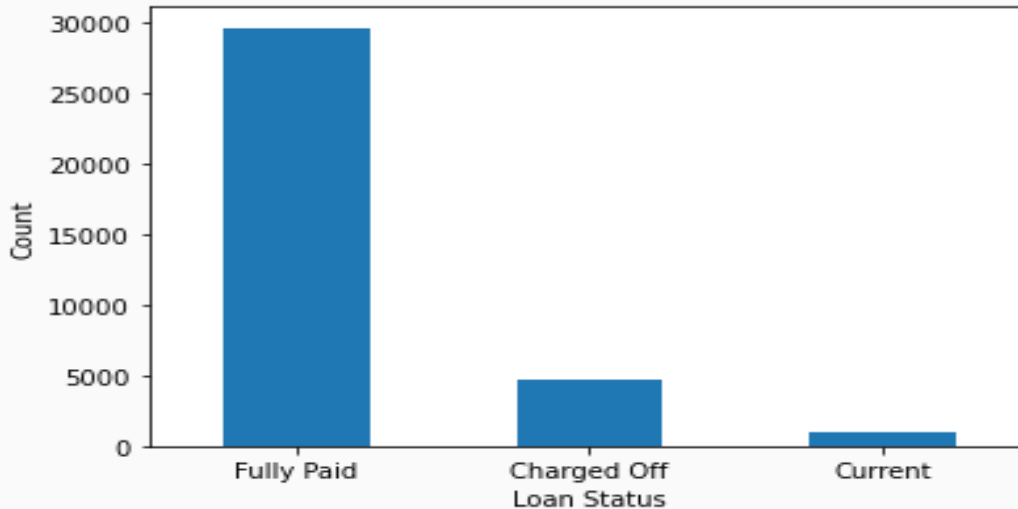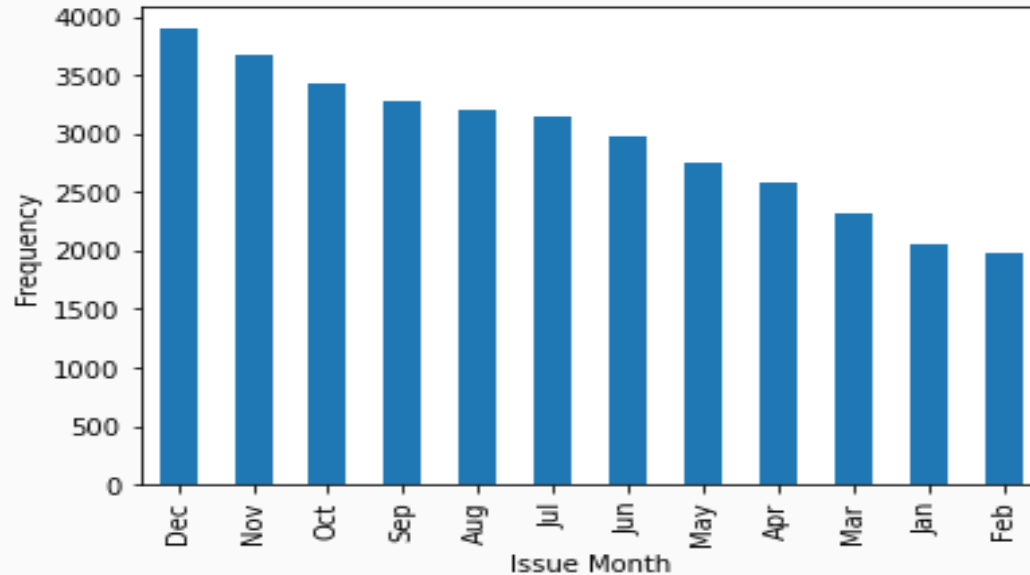
Univariate Analysis

This Bar Chart tells us that most of the rows in the csv file for purpose column contain debt-consolidation value, this means that most of the loans are taken for debt_consolidation purpose.
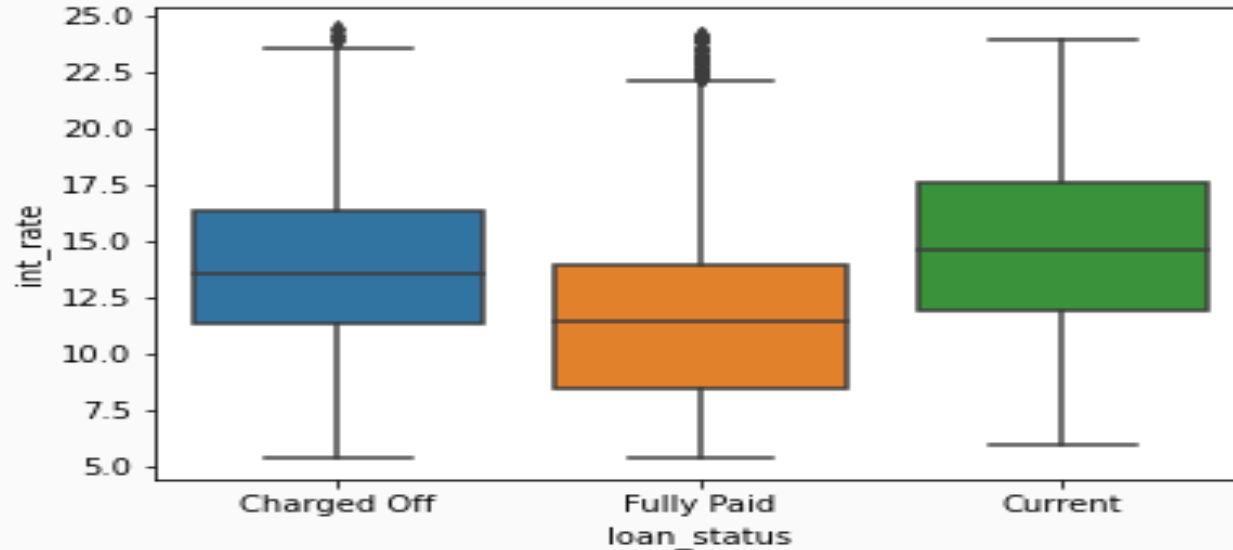
Univariate Analysis

This Bar Chart tells us that most of the loans are issued in the month of December followed by November.
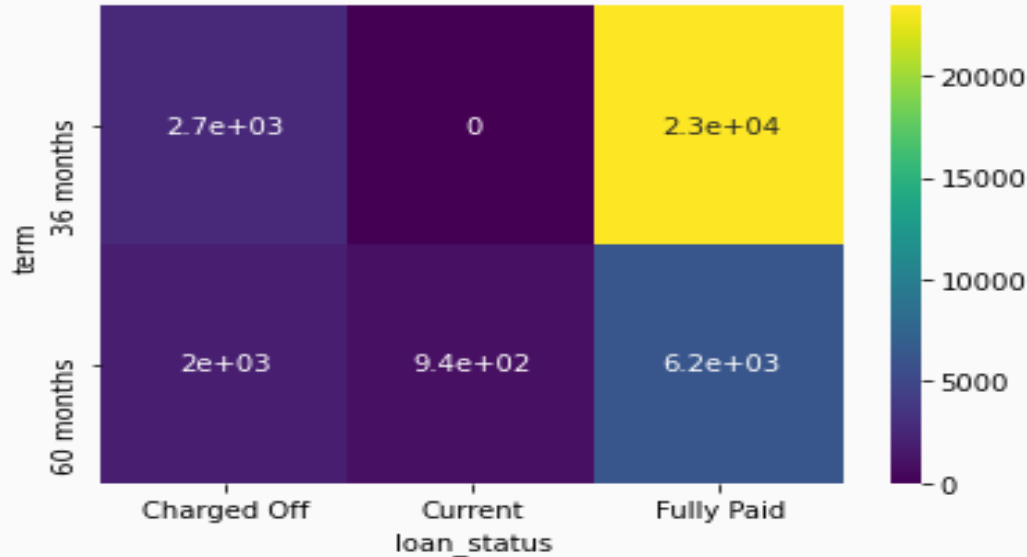
Bivariate Analysis

This Box Plot tells us that with increase in interest rate the chances of loan getting charged off increases and if the interest rate is low then the chances of loan getting fully paid is more.

Bivariate Analysis

This Heatmap tells us that the loans provided for longer term i.e. 60 months have a very high chance of being charged off (defaulted) than loan given for a shorter term i.e. 36 months.
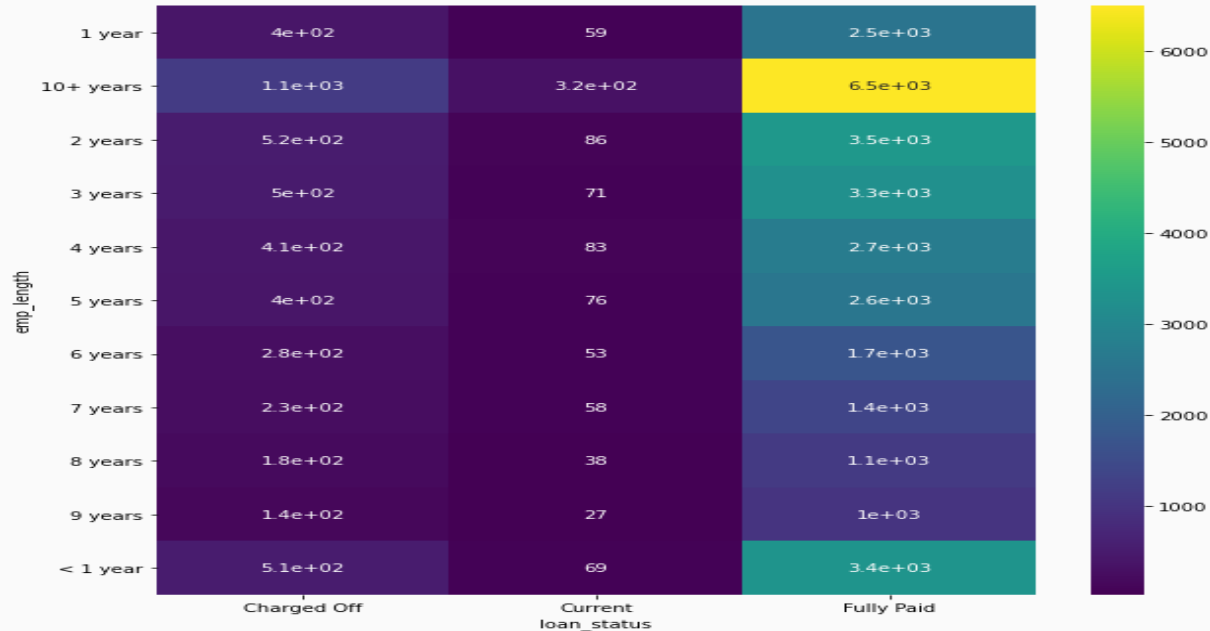
Bivariate Analysis

This Heatmap tells us that people living on rent and mortgage have a very good probability of fully paying back the loan in comparison to people living on owned home.

Bivariate Analysis

This Heatmap says that people who are doing job for more than 10 years have the highest probability of paying back the loan.

Bivariate Analysis

From below piecharts we can say that loan taken for the purpose of small businesses have higher chances of getting charged off whereas loan taken for major_purchase have good chances of getting fully paid.
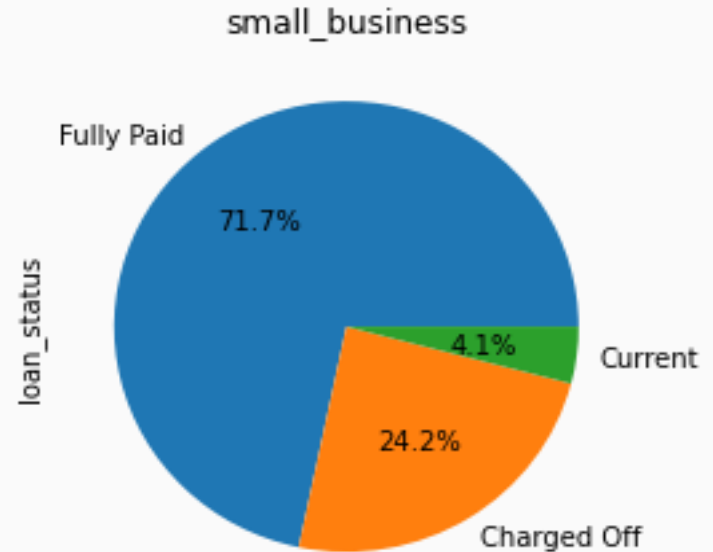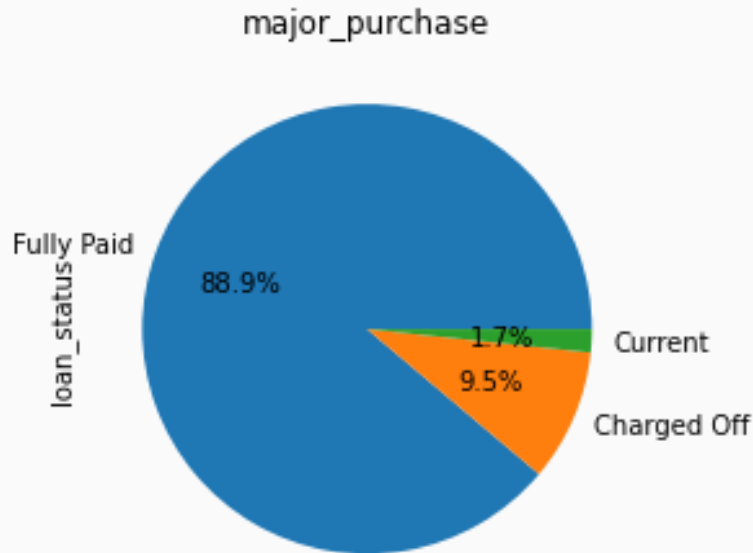
Conclusion

By doing the EDA on the dataset provided to us we have come to a conclusion that there are some points which online loan company should keep in mind while granting the loan:

- For the loans given at higher interest rate chances of loan getting charged off or not getting paid are higher compared to the loans given at low interest rate.
- Loans given for shorter terms have higher chances of getting fully paid in comparison to the loans with higher term.
- Another interesting fact that we came across was that people living on rented and mortgaged property have most of times paid back the loan fully.
- Loans given to people who have been employed for more than 10 years have great probability of getting fully paid.
- Loans given for major_purchase purpose have the highest chances of getting fully paid whereas loans given for small businesses purpose have higher chances of getting charged off.