LENDING CLUB CASE STUDY

Raghavendra Srinivasan

Amit Kumar Pradhan

PROBLEM STATEMENT

To understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Two **types of risks** are associated with the bank's decision:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a loss of business to the company

- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

OUR OBJECTIVE

The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

PROBLEM SOLVING METHODOLOGY

- Data Cleaning
- Data Understanding
- Univariate Analysis
- Bi-variate Analysis
- Multi-variate Analysis
- Conclusion

DATA CLEANING

Initially, we had data with 39717 records and 111 columns.

Steps involving data cleaning -
- Removal of columns with null values
- Removal of irrelevant columns
- Converting to relevant data type in columns
- Deleting of correlated columns
- Null value imputation
- Detecting outliers
- Treating outliers

After applying all the above data cleaning methods, we are left with 36654 records and 22 columns.

IMPORTANCE OF EDA

- A simple approach to summarize the Main Characteristics of the dataset

- A very important aspect of Descriptive Statistics

- Gives an exact idea about what has happened

- Helps us in a better understanding of Statistical tools

- Maximum time spent at EDA results into a better insight of the Data and how it should be handled while model building

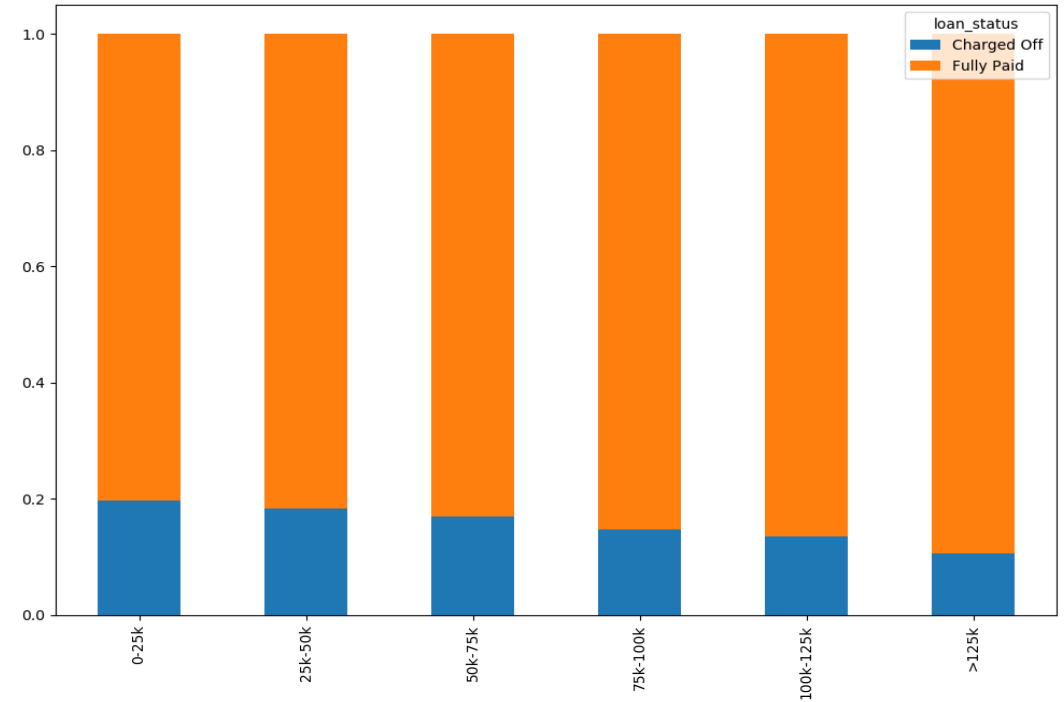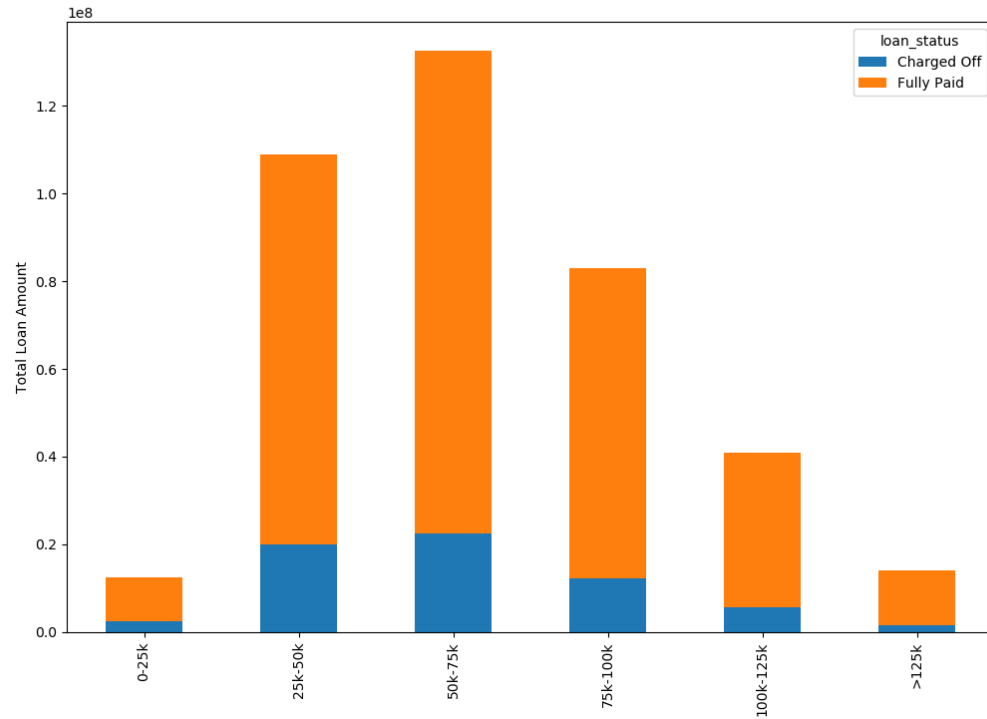# DISTRIBUTION OF TARGET VARIABLE - LOAN STATUS
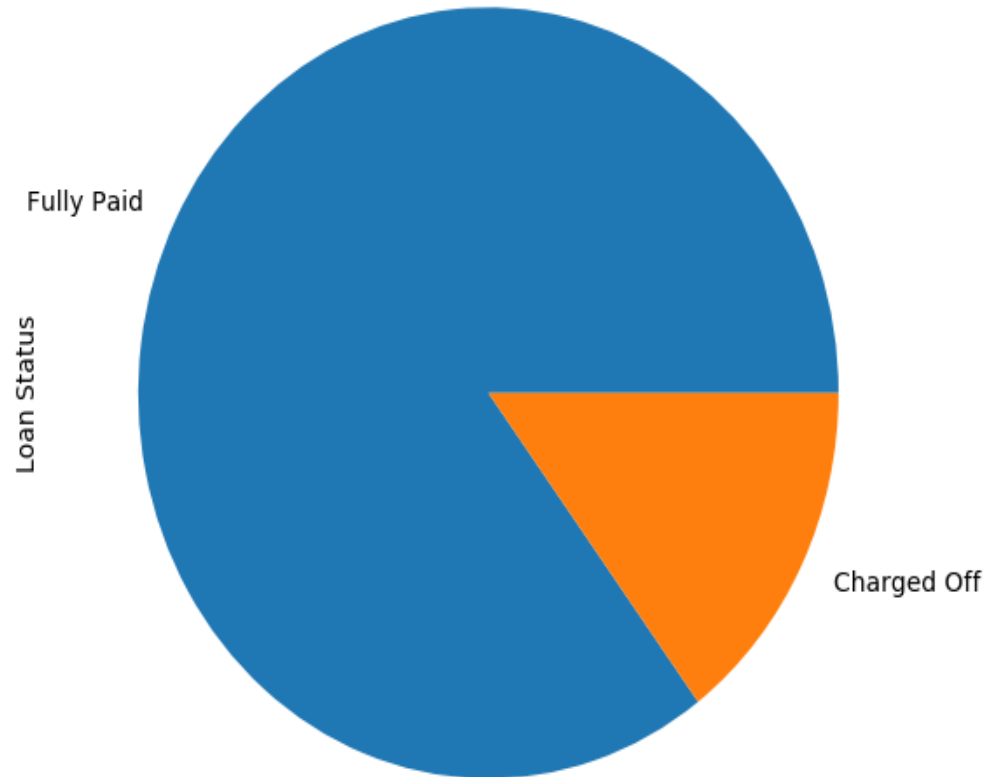
The data consists of non defaulter (**Fully Paid**) more than defaulter (**Charged Off**).
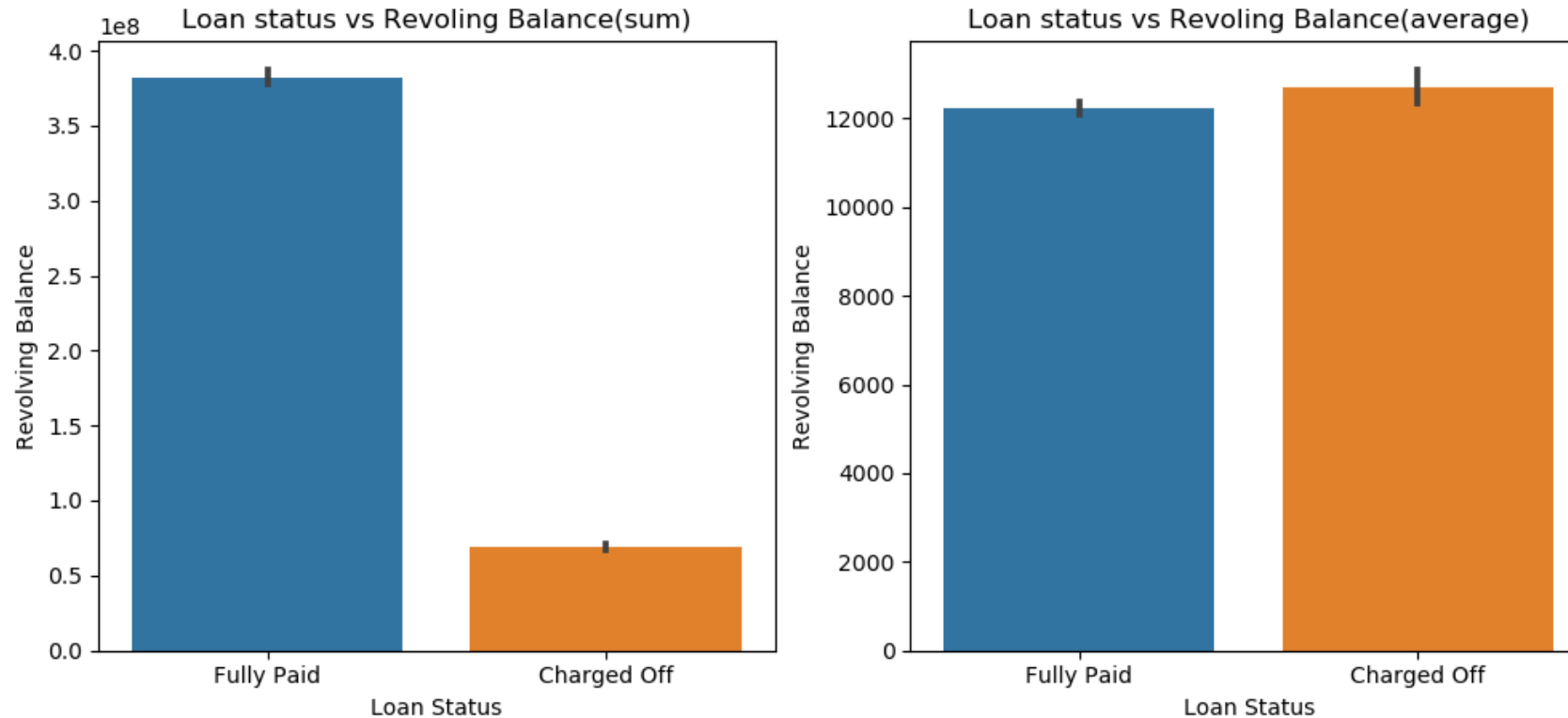
ACCORDING TO GRADE
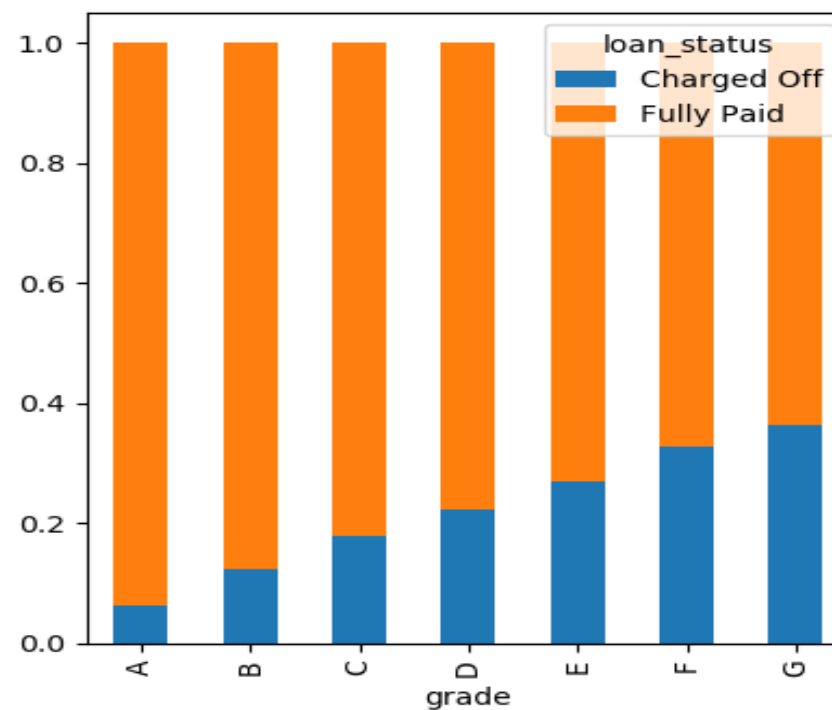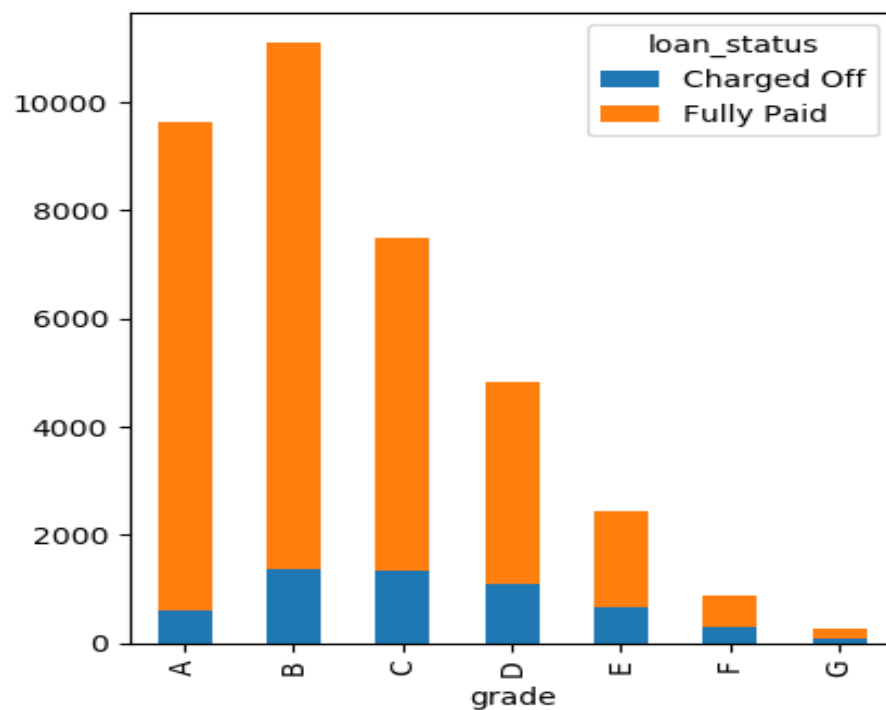
We can see that as the grade is getting lower the percentage of applicant being defaulter is more.

ACCORDING TO ANNUAL INCOME

We can see a trend that as the annual income is increasing, the possibility of applicant being defaulter is getting lesser and lesser.
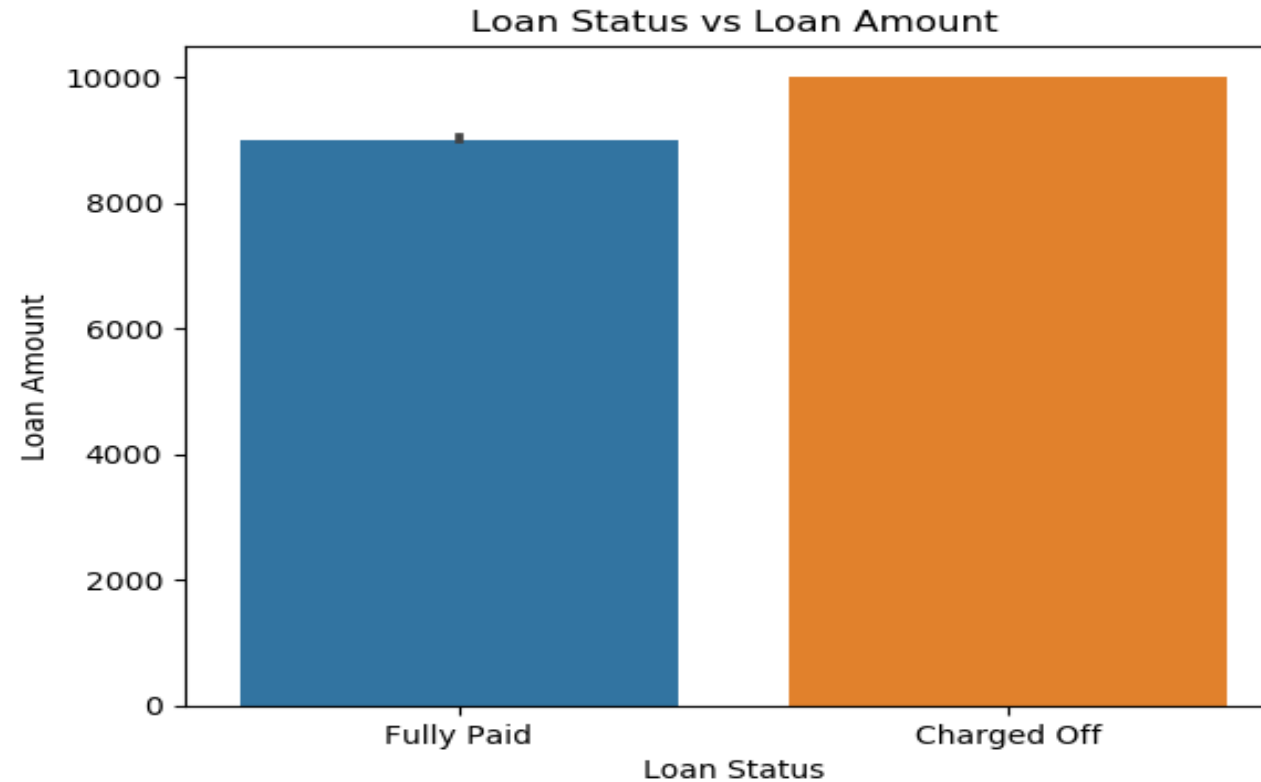
ACCORDING TO REVOLVING BALANCE

From 1st graph, we can see that total revolving balance is more for fully paid applicant and it is due to the distribution of data.

When plotted the same data with average , we get the clear picture that if the revolving balance is more, then the applicant can be defaulter.

LOAN AMOUNT

The median value of charged off is greater than fully paid customer. We can say that whenever applicant is asking for more loan we should scrutinize them more cautiously compared to others.

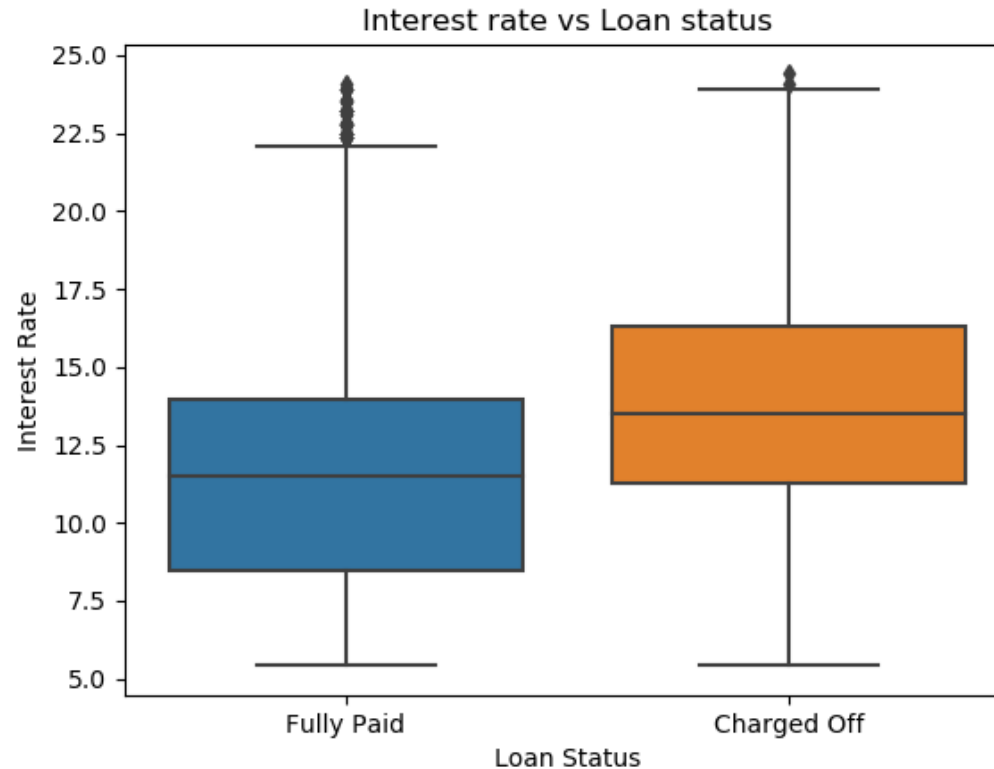# ACCORDING TO INQUIRES IN 6MONTHS
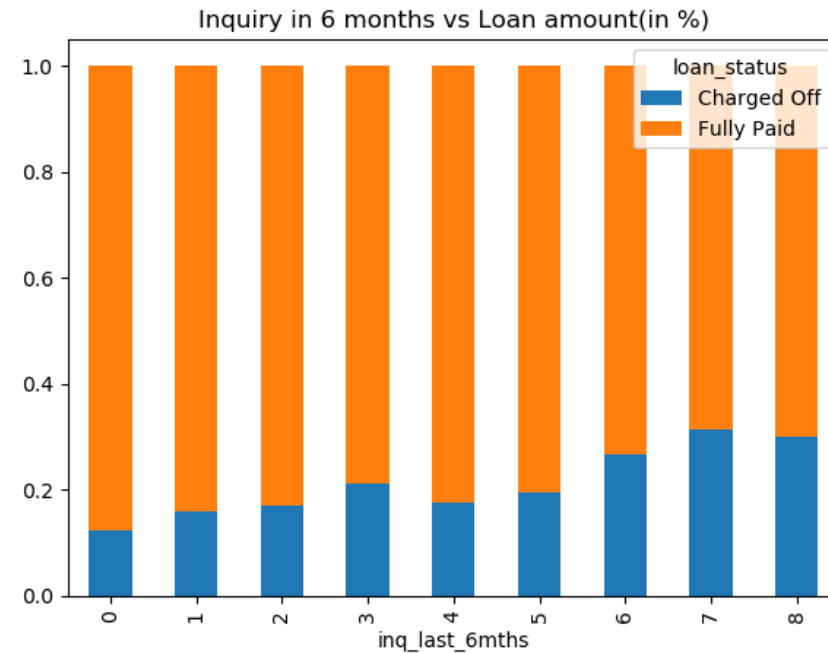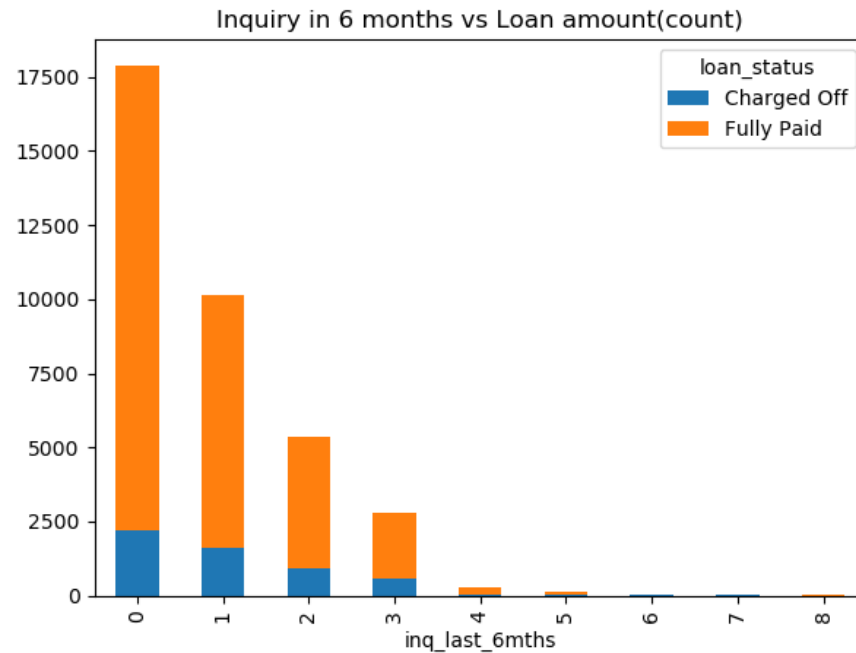
We can see that as the inquiry done by applicant increases, the defaulter rate is also increasing.

# ACCORDING TO REVOLVING LINE UTILIZATION RATE
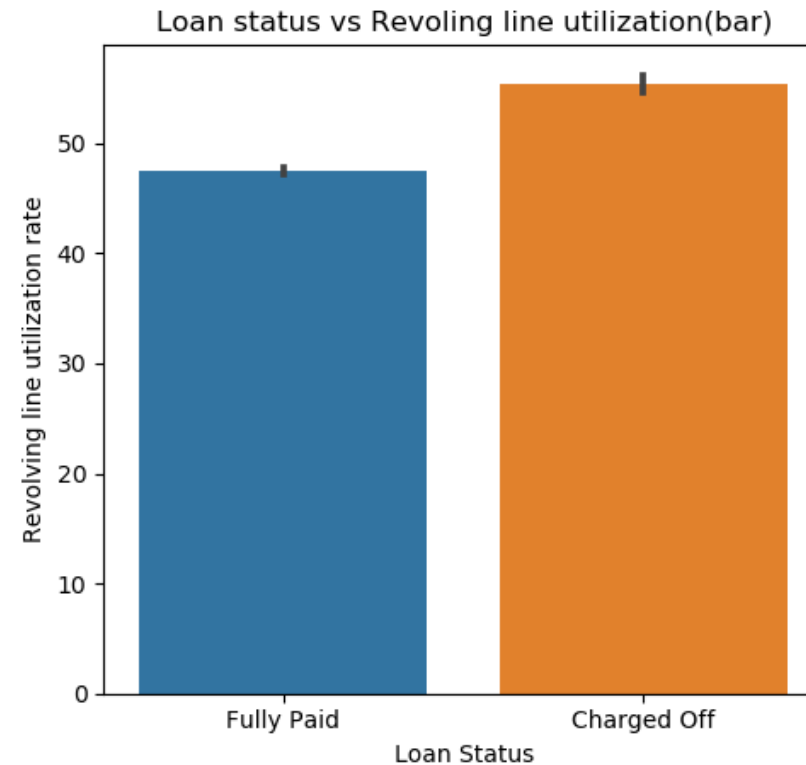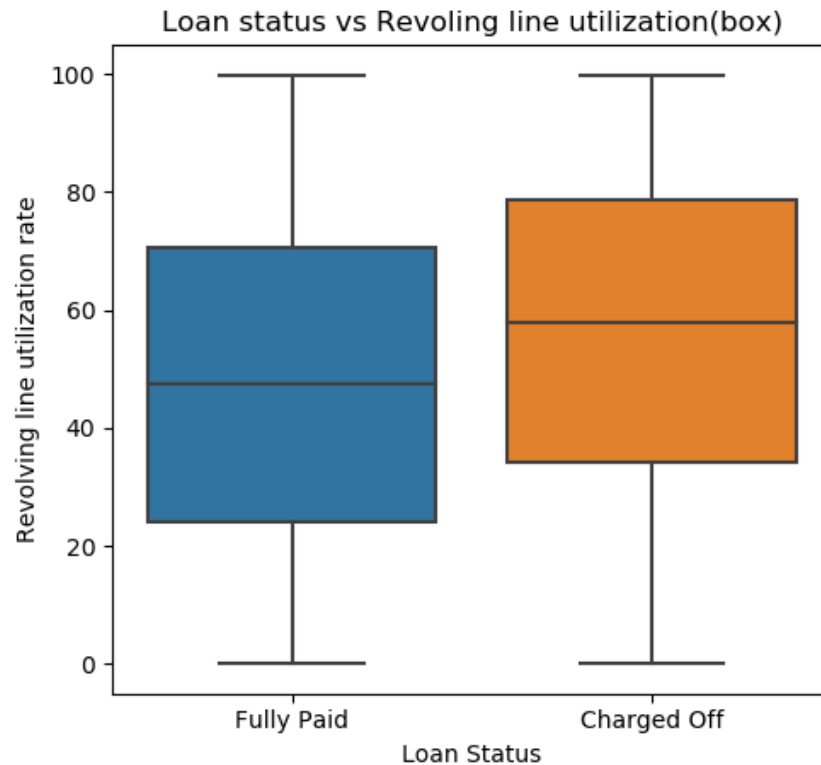
We can see that if revolving utilization rate is more than 50% than the possibility of applicant to be a defaulter is high.

ACCORDING TO INTEREST RATE

We can interpret that if the interest rate is high then the possibility of applicant becoming defaulter is high. Interest rate is decided according to grade of the applicant.

ACCORDING TO DTI

Interest rate vs Loan status
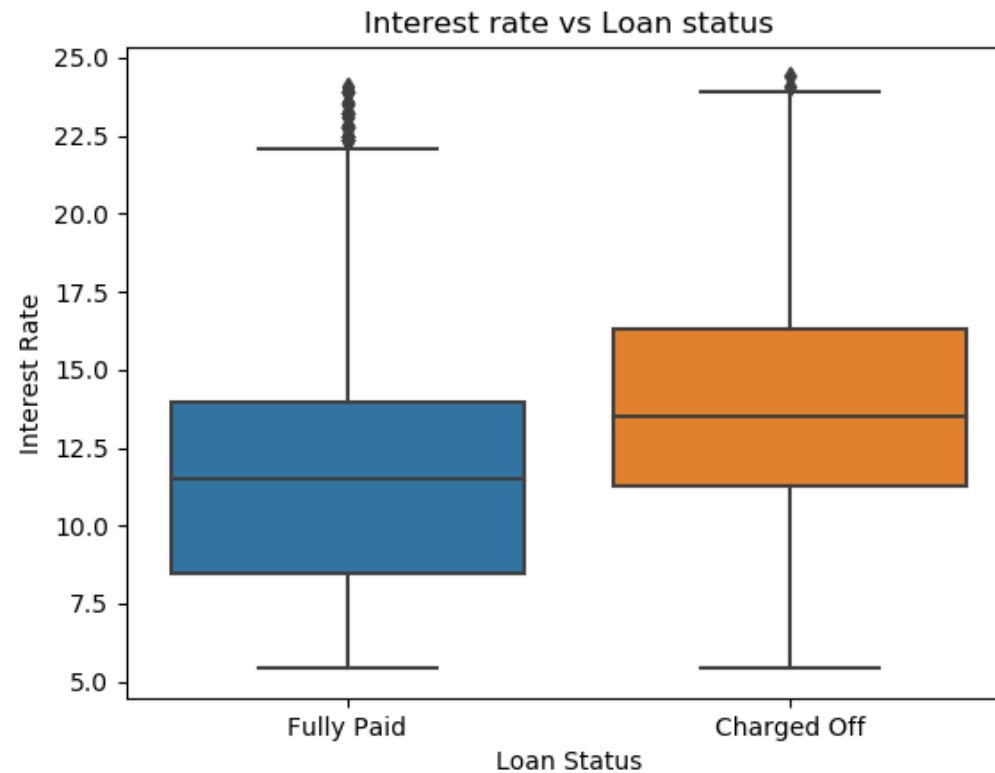
We can say that dti of charged-off applicant is higher than fully paid applicant.

CONCLUSION

We have done Data cleaning and Exploratory data analysis to get better insights of data. We found following driving variable which will give us inference about the applicant can be default or not.

Following are the driving variables -

- **Grade** - We have seen that as the 'grade' is lower then the chance of applicant to be defaulter increases.
- **Annual Income** - We have seen that as the annual income is increasing, the chances of applicant being a defaulter decreases.
- **Revolving Balance** - We have seen that if the revolving balance is higher then the chance of applicant being defaulter is also high.
- **Loan Amount** - We have seen that as the loan amount increases, the possibility of applicant being defaulter also increases.
- **Revolving line utilization rate** - We have seen that if revolving line utilization rate is higher then there is a higher chance of applicant being a defaulter.
- **DTI** - We have seen that higher the dti, higher is the chance of applicant being defaulter.
- **Interest Rate** - We have seen that as the interest rate increases, then the possibility of applicant being defaulter also increases.
- **Inquiry in last 6months** - We have seen that if the no. of inquires is more, then defaulter rate is also high.

Thank you