

LENDING CLUB CASE STUDY

Team

Anshu Dhiman

Renganathan Rajagopal

About Lending Club

The company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

Problem Statement

Borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as '**charged-off**' are the '**defaulters**'

Objective

The company wants to understand **the driving factors behind loan default**, i.e. the variables which are strong indicators of default.



Approach Done

Data Understanding

Imported the data into a Pandas Dataframe and looked for the total columns and understood about the data

Data Cleaning

We could see there are null values are present for the most of the columns so did proper analysis and removed Null Values and unnecessary values & column

Univariate Analysis

we Plotted count plots and analyzed the data distribution for various columns and come up with the Observation.

Bivariate Analysis

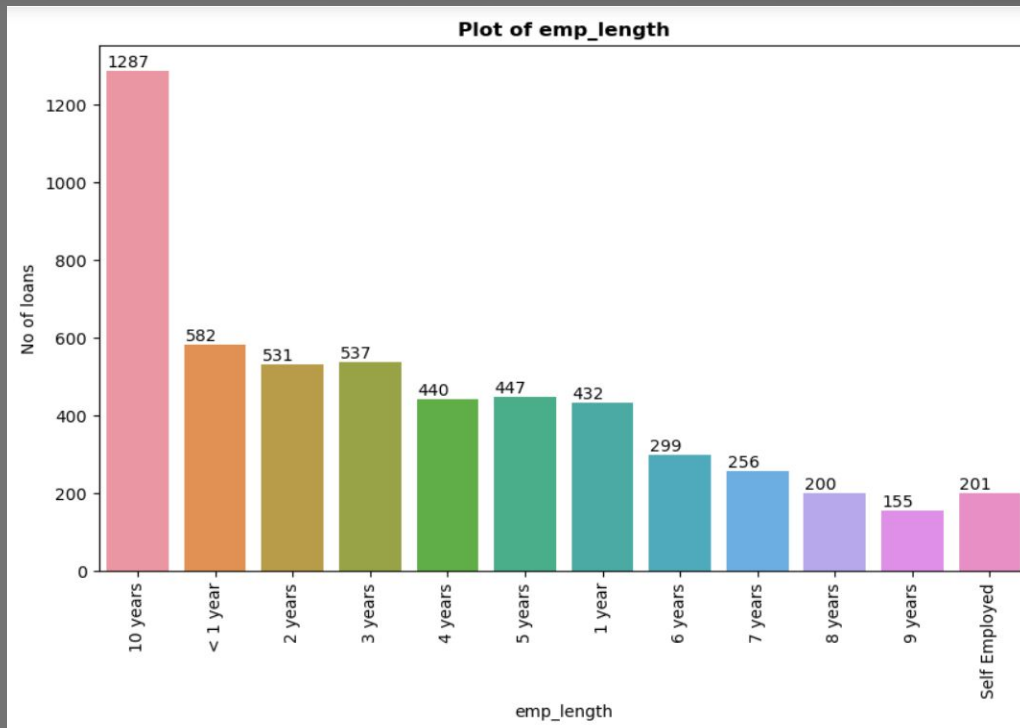
We Analysed the data with respective to another columns and observed how it is impacting the loan status

Recommendation

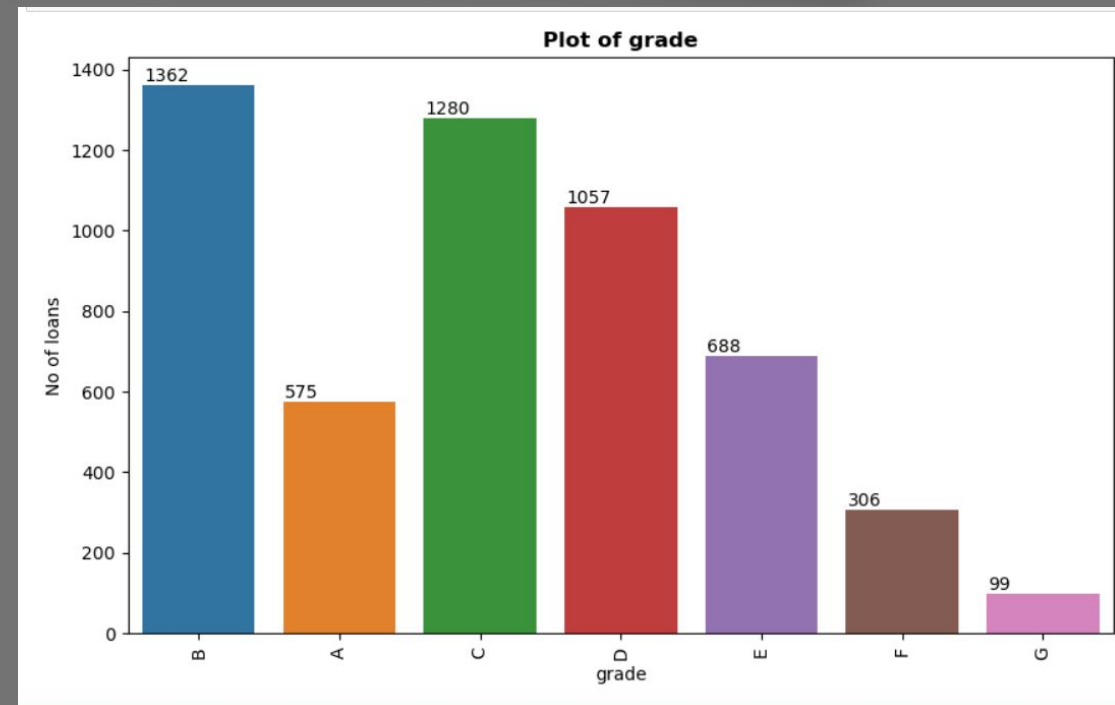
Using the Previous Observations from the Univariate and Bivariate Analysis we have come up with some Recommendations to reduce the defaulters

NOTE : The following plots are created taking the data for Loan_status = Charged_off

Univariate Analysis



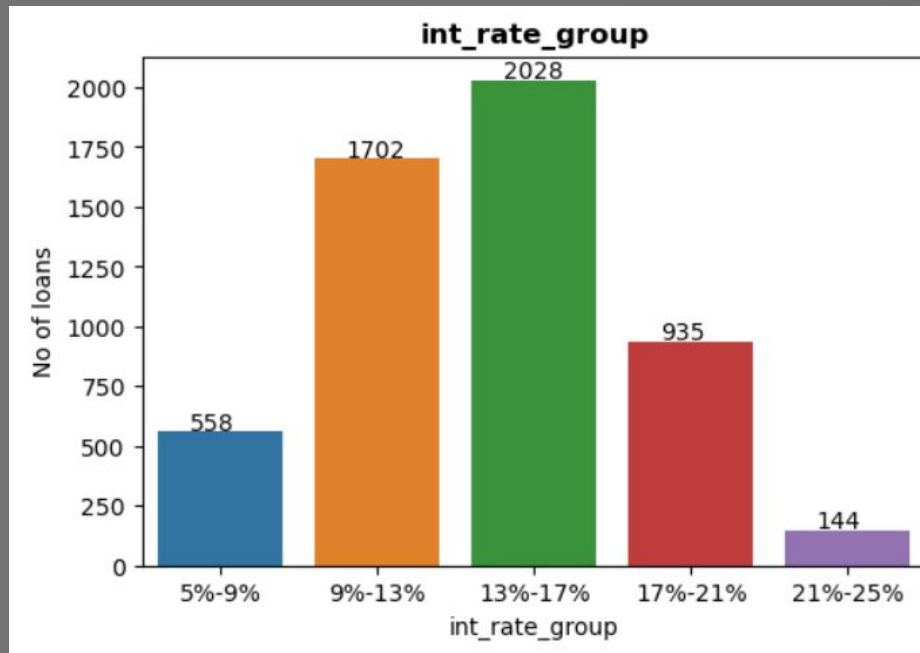
- Count Plot of **emp_length** shows ,employees with 10 years experience have more defaulted on loan compared with lesser experience



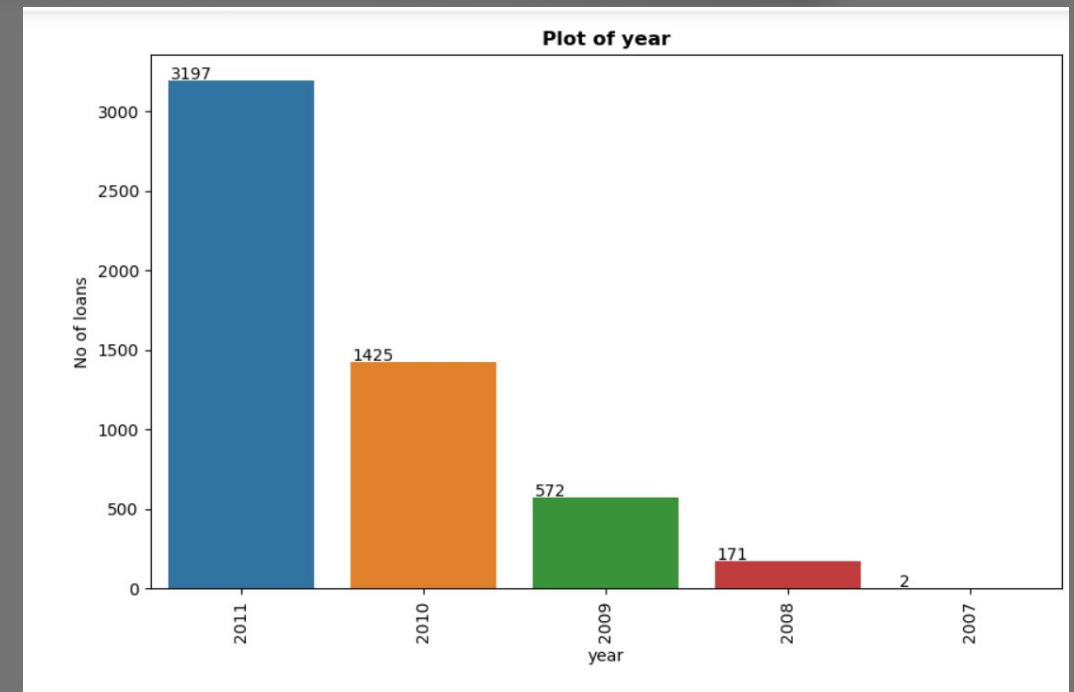
- Count Plot of **Grade** shows Grade B ,C, D are given more likely to charged off as compared to other grades

NOTE : The following plots are created taking the data for Loan_status = Charged_off

Univariate Analysis

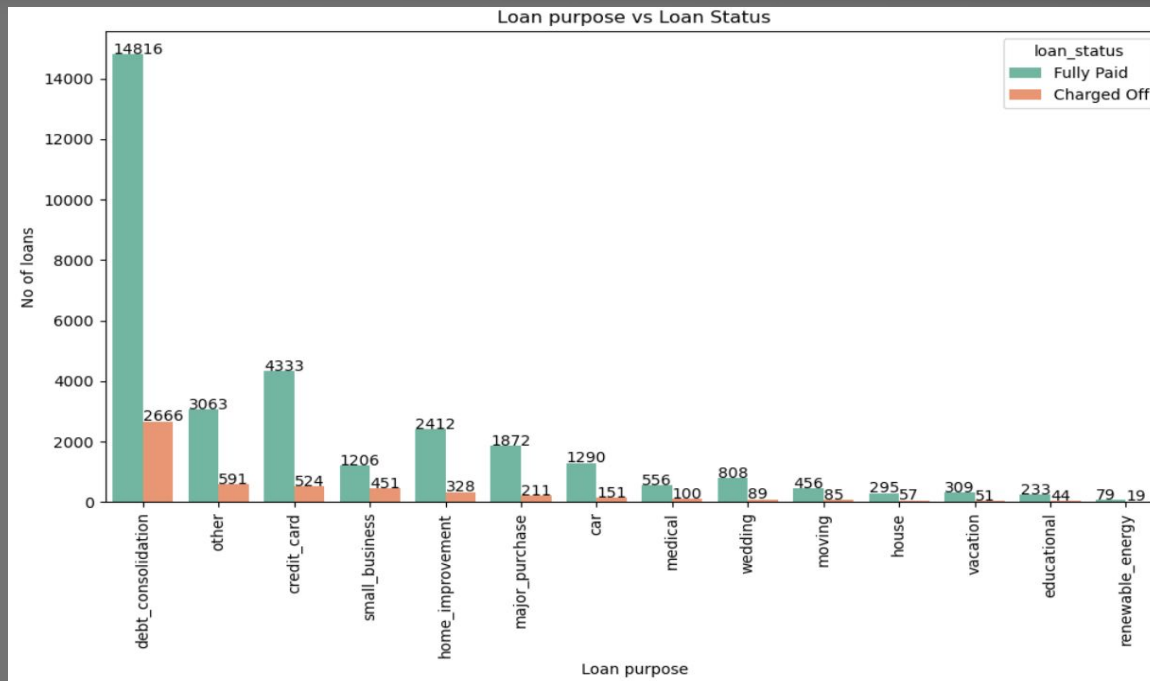


- Interest rate range 13% - 17% is the range with maximum loan defaults
- If we notice the % of defaults in each groups we can state that the interest range 21%-25% has maximum chances of defaults (as high as 44%)
- P.S. (Percent here means no. of charged off divided by the total no. of loan issues in a particular interest rate group)

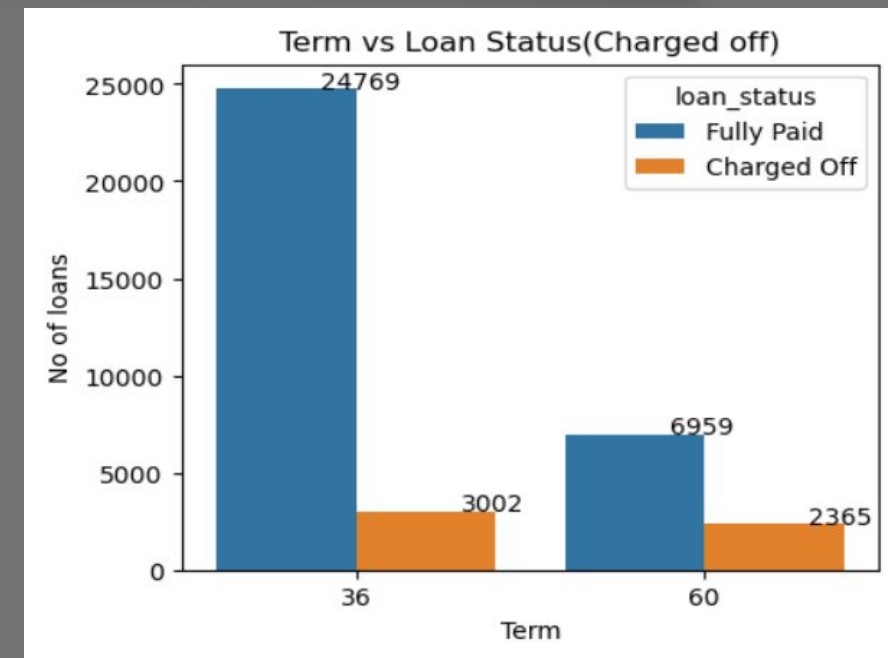


- Count Plot of issue **year** shows the no. of defaulted loans have increased in the year 2011. The trend is increasing with the increase in the year

Univariate Analysis



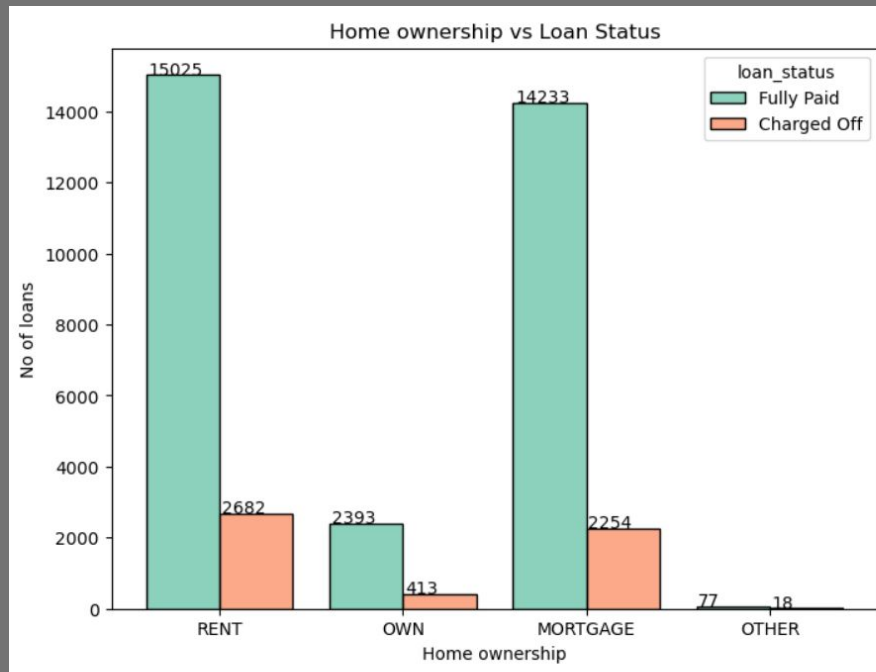
- Loans with purpose debt consolidation, other, credit card and home improvement categories have failed to pay the loan compared with education / renewable energy
- Small business have highest % of charged off



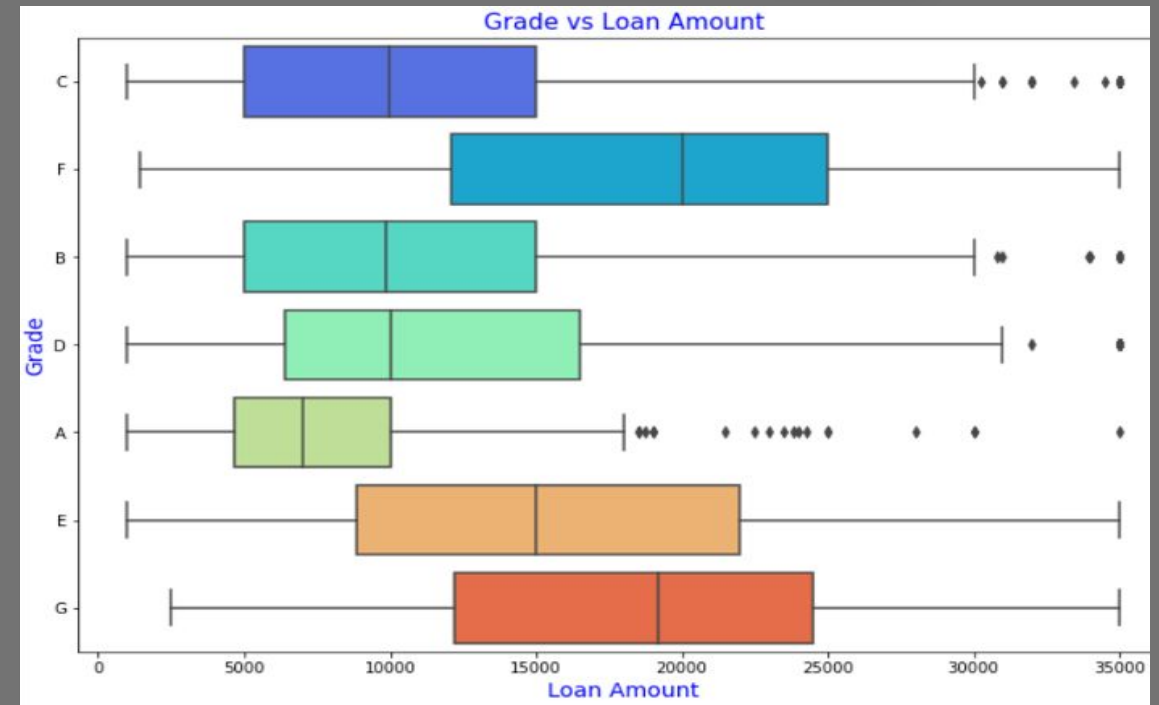
- Calculate % charged off in 60 months it is 25% of the total loans issued with a 60 months which is much higher as compared to 36 month tenure so it is more likely for a 60 months loan issued to be a default

NOTE : The box plot is created taking the data for Loan_status = Charged_off

Bivariate Analysis

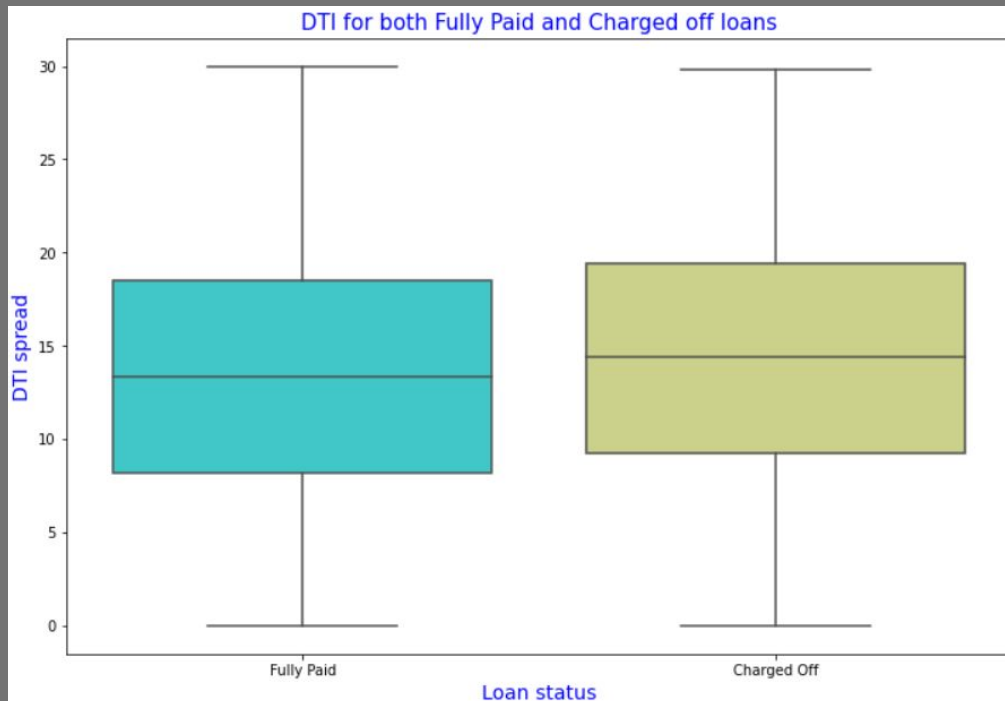


- People who are in Rent or Mortgage have failed to pay the loan compared with people in Own house

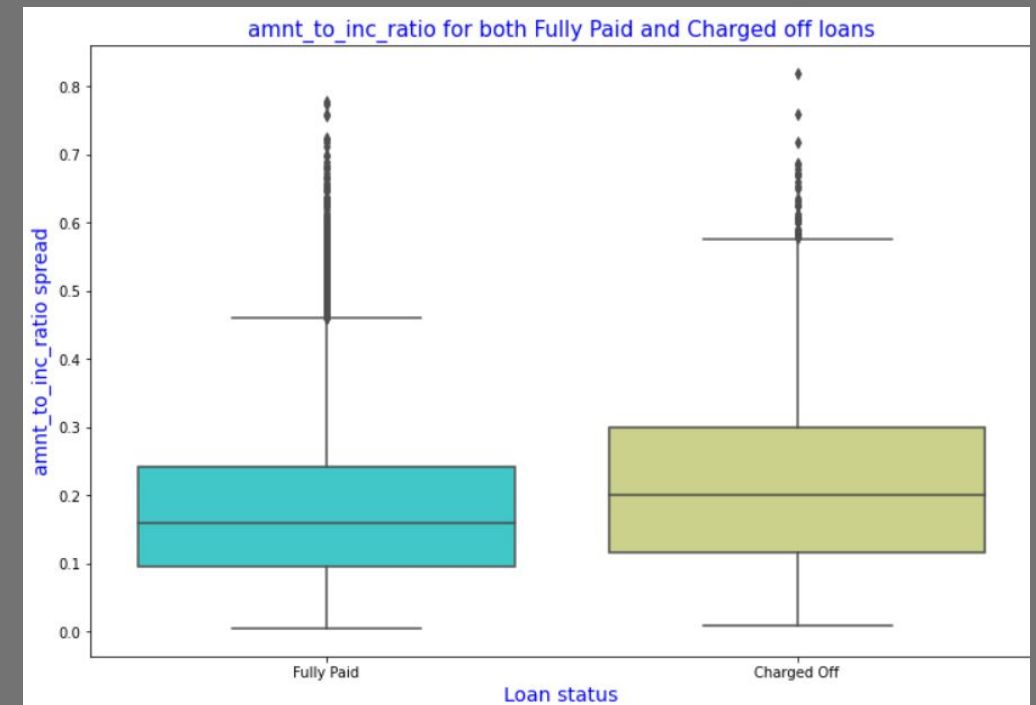


- Grade F, G and E are the three category which has higher charged off
- The median of F and G is around 20k and Q3 at 25k
- Grade A has a median at 7.5k and is at the least

Bivariate Analysis



- DTI is not a significant factor because the median is close to each other for Fully Paid and Charged Off



- Amnt_to_int_ratio is an indicator of bad loans as for charged off the median and Q3 is higher

Recommendations:

- Lending club should reduce the high interest loans for **60 months tenure**, they are prone to loan default.
- **Small business loans** are defaulted more. Lending club should cautiously give the loans to them.
- Borrowers with **mortgage home ownership** are taking higher loans and defaulting the approved loans. High Alert for Lending club when loan amount asked is 12000 or more.
- Charging an **Interest rate** ranging between 21-25% can lead to a list of defaulters so we can minimize giving loans to the profile on which interest rate is calculated high.
- **Grades** can be one of the indicator for detecting defaulters. When grade is F or G and loan amount is between 15k-20k Those are more likely to be defaulters
- Lending Club should keep a check to the number of loan issues to borrowers who are from CA, FL and NY to avoid risks.