LENDING CLUB CASE STUDY

- EXPLORATORY DATA ANALYSIS (EDA)

Group members:

Vennesh Kari

Rajesh Draksharapu

PROBLEM STATEMENT

- Problem:

You are working in a Consumer finance company which is lending various types od loans to urban customers. When the company receives loan application it has to decide based on applicant`s profile. But there are lot of loans which are being as charged off which is a loss to the company.

- Objective:

We need to do EDA (exploratory data analysis) of the data to identify which terms are causing increase in charged off loans and also we can provide some recommendations such that company can reduce financial loss.

AVAILABLE DATA SET & APPROACH

- Dataset:

  - Loan.csv:

    This file has all the completed loan details in time period of 2007 to 2011.

  - Data Dictionary:

    Contains details of each column present in loan data set.

- Approach

 Data Cleaning ➜ Derived Metrics ➜ Univariate Analysis ➜ Segmented Univariate Analysis ➜ Bi Variate Analysis / Multi Variate Analysis

DATA CLEANING

- As part of Data Cleaning we have done following:
  - Checking Null values and if any column have more than 60% of null values we are dropping that column
  - Checking single valued columns and dropping those columns.
  - Removing unnecessary columns such as total , title, description, URL etc.,
  - Standardizing the int_rate and revol_util columns like removing % symbol and convert them from object to numeric.
  - Standardizing the term and emp_length columns converting them from object to numeric.
  - Perform Sanitation checks like whether installment is less than loan_amount and funded_amount

DERIVED METRICS

- As part of derived metrics we have done following:
  - We have derived year and month column from issue_d column for more insights.
  - We have derived charged_off_num column to find the defaulters percentages.
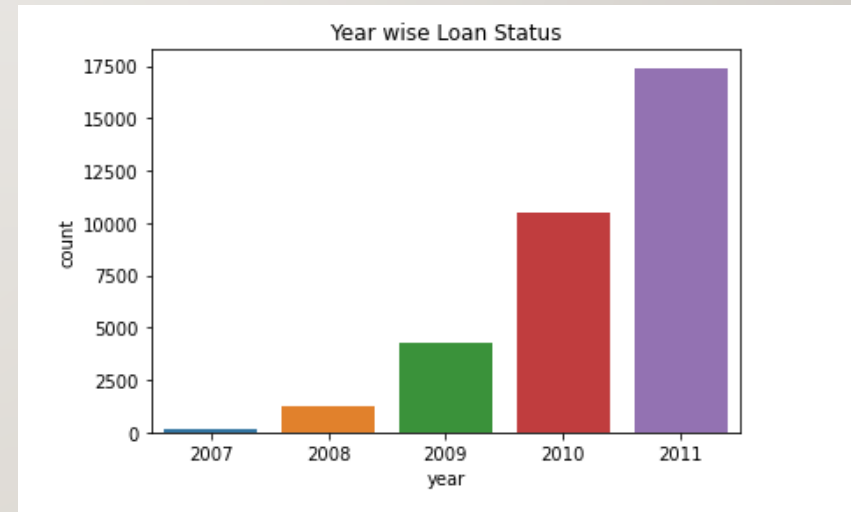
UNIVARIATE ANALYSIS

- As part of univariate analysis we have done following:
    - We have considered grade, subgrade, month as ordered variables.
    - We have considered purpose, home_ownership as unordered variables.
    - We have considered annual_inc, total_acc, open_acc as quantitative variables.
    - Removed outliers by calculating IQR for Quantitative variables and plotted box plots.

UNIVARIATE ANALYSIS CONTD.
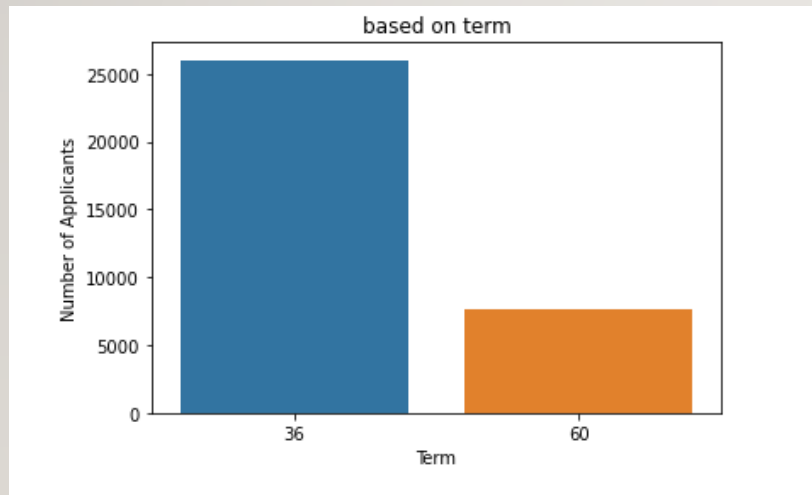
Out of all loans approx., 15% of them are charged-off.
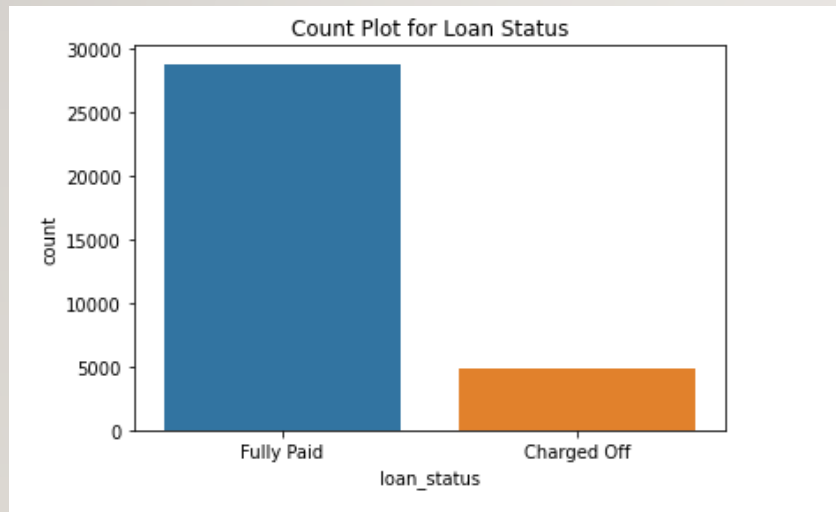
Number of applicants are increasing year by year.

UNIVARIATE ANALYSIS CONTD.

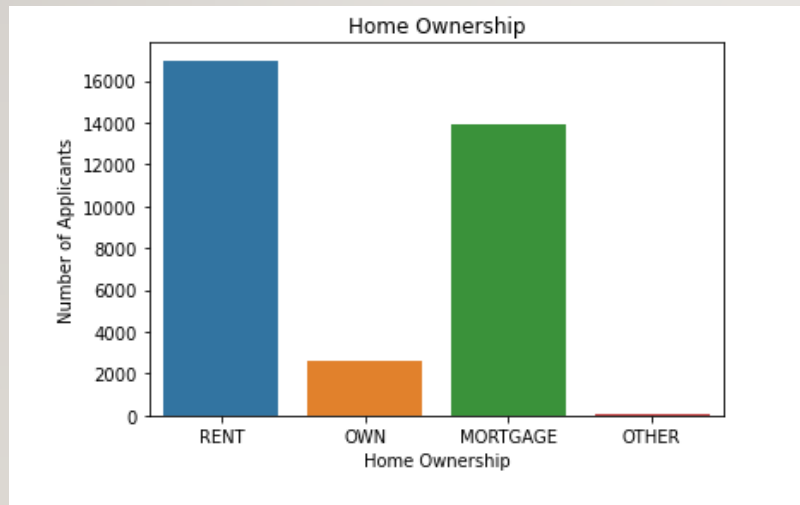Out of all loans we can see there are more of term 36 than 60

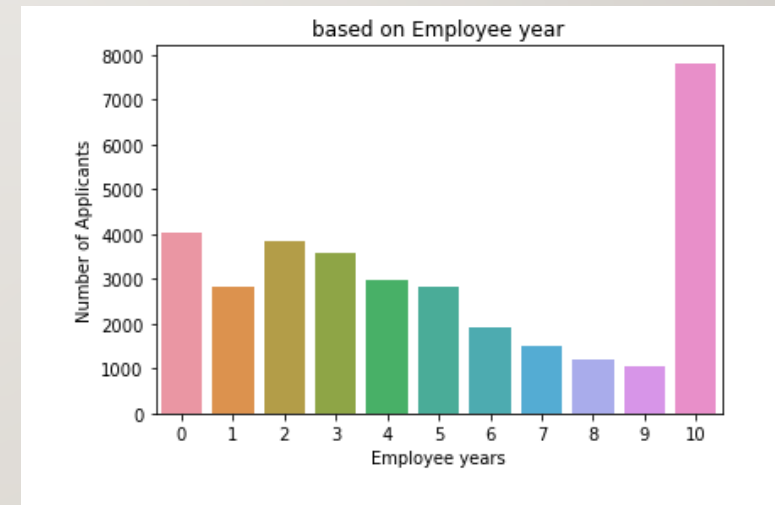Out of all loans there are more not verified loans than others.
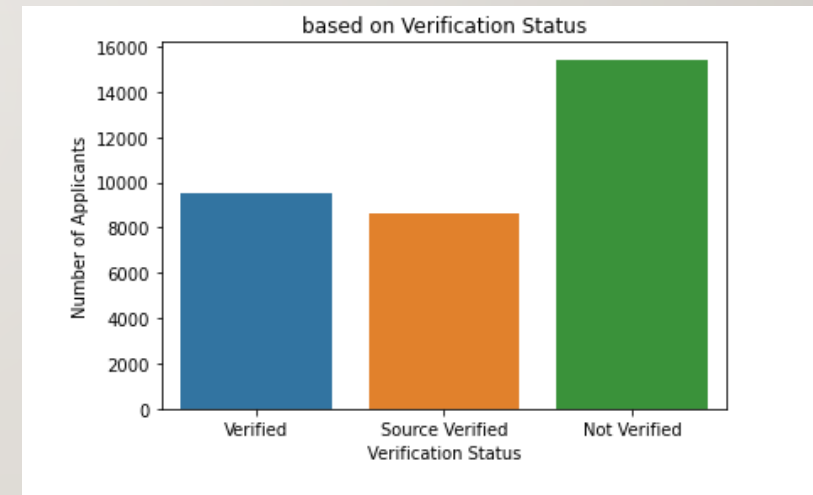
UNIVARIATE ANALYSIS CONTD.

Out of all loans there are more from rented property

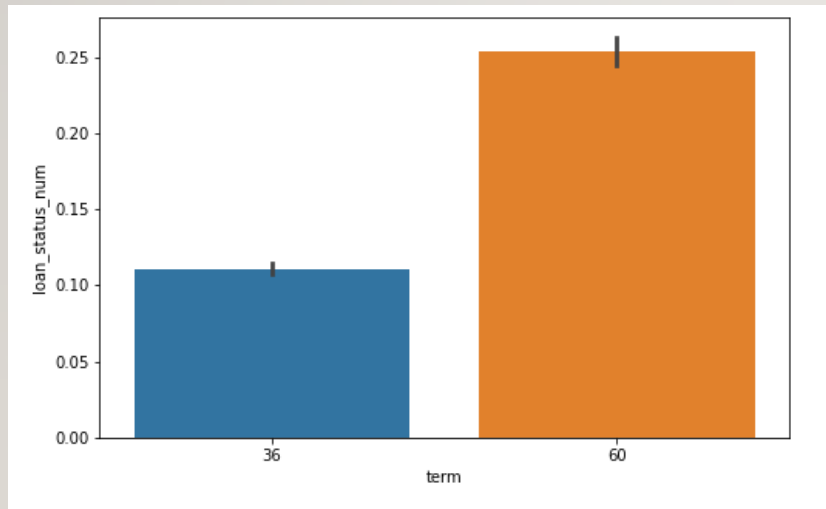Employees who are applying loans are more from 10+ years
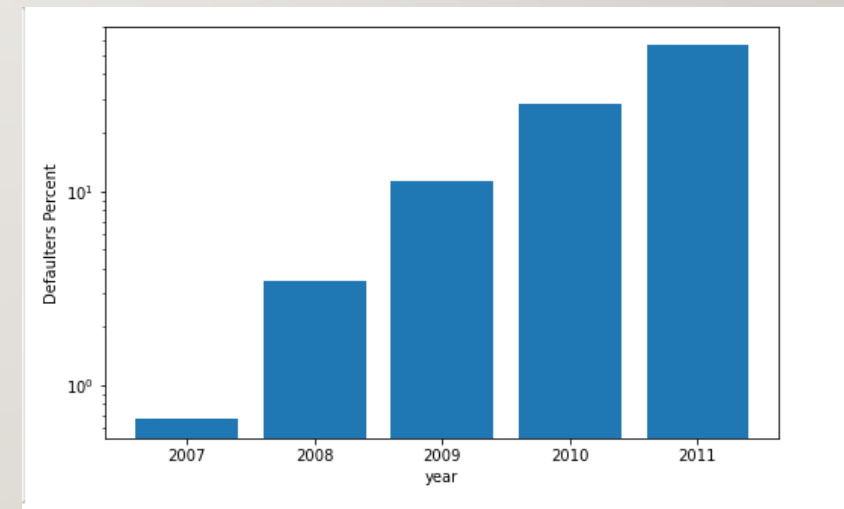
SEGMENTED UNIVARIATE ANALYSIS

- As part of this we have identified defaulters percentage for each column which we have analyzed as part of univariate analysis
  - Year
  - Term
  - Grade
  - Home_ownership
  - Emp_length

SEGMENTED UNIVARIATE ANALYSIS CONTD.

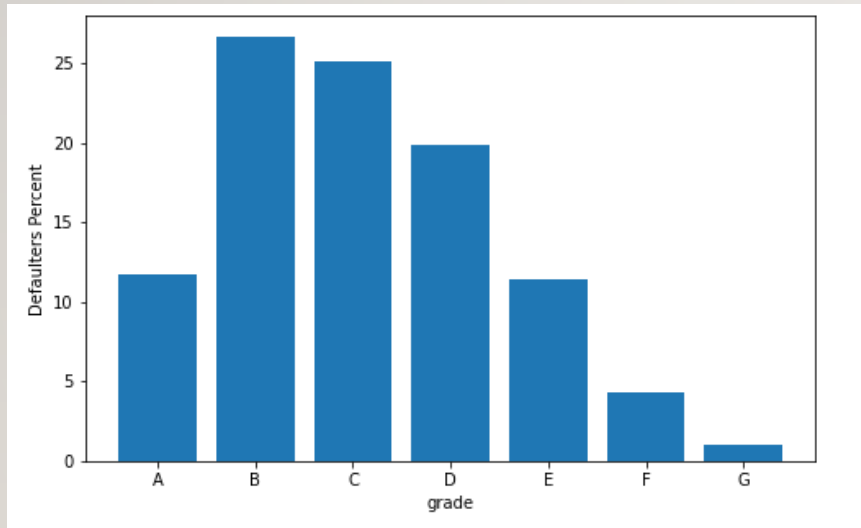Defaulters are increasing when the term is increasing

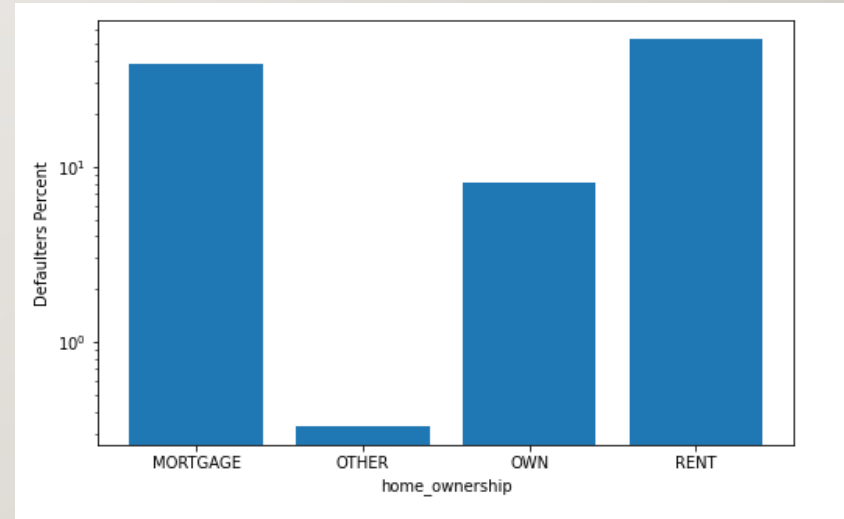
Defaulters are increasing year by year

SEGMENTED UNIVARIATE ANALYSIS CONTD.

Out of all loans defaulters are more in B and C grades.

Defaulters are more in rented property applicants.

BIVARIATE ANALYSIS / MULTI VARIATE ANALYSIS

- As part of this we have plotted correlation plot and scatter plots for both positive and negative correlations.

  - We have also plotted box plots with respect to int_rate and loan_status with other columns like term, year, grade and purpose.

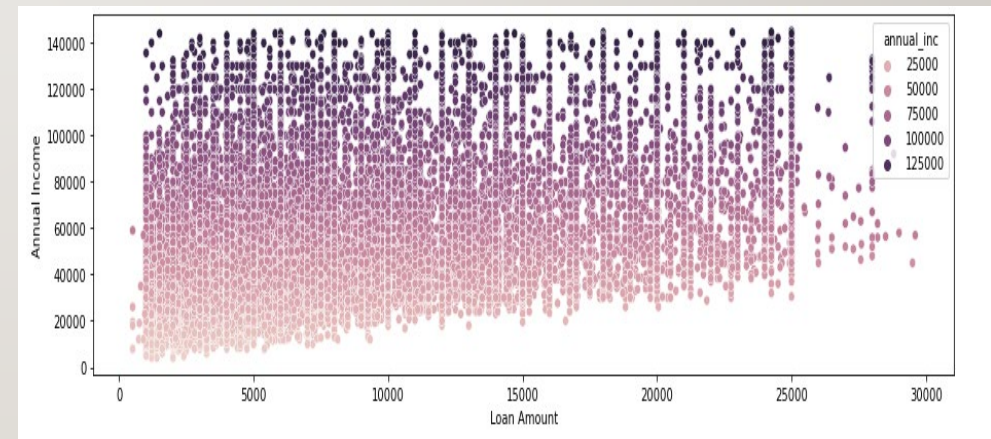   loan_amount, interest_rate and Installment can be considered as Cluster.
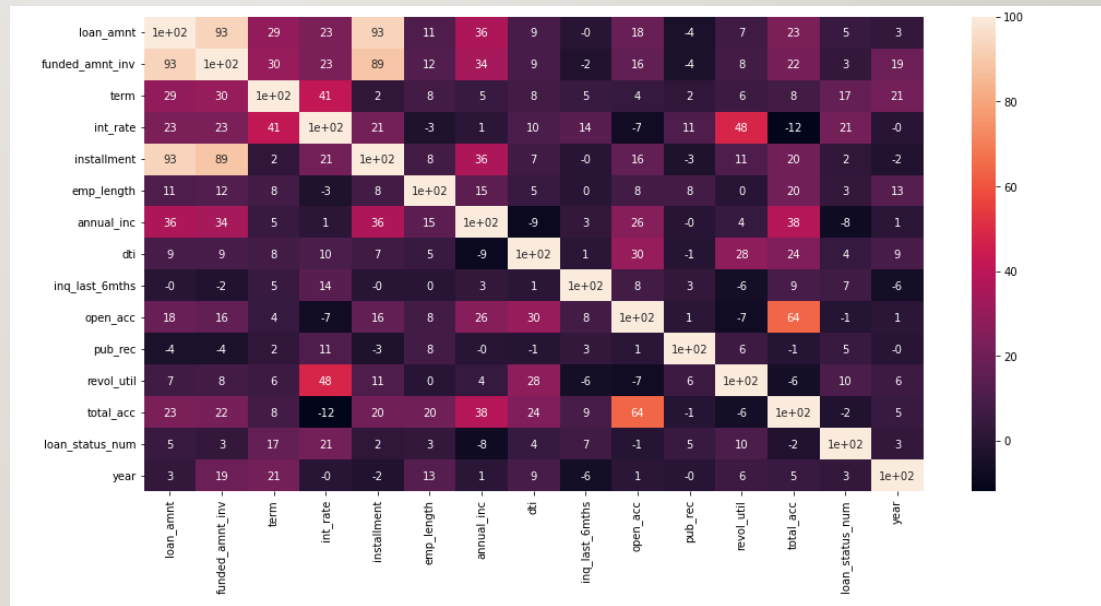
BIVARIATE ANALYSIS / MULTI VARIATE ANALYSIS CONTD.

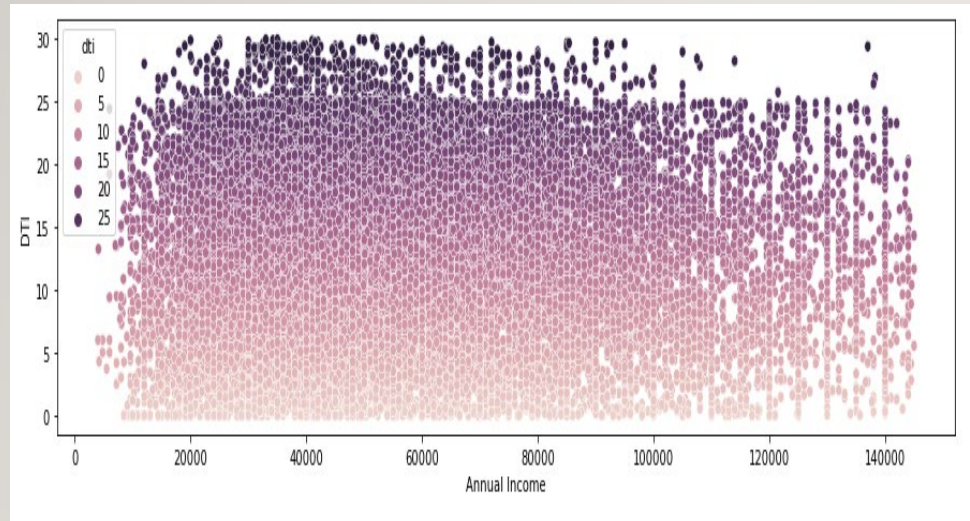We can see positive relation between loan amount and installment
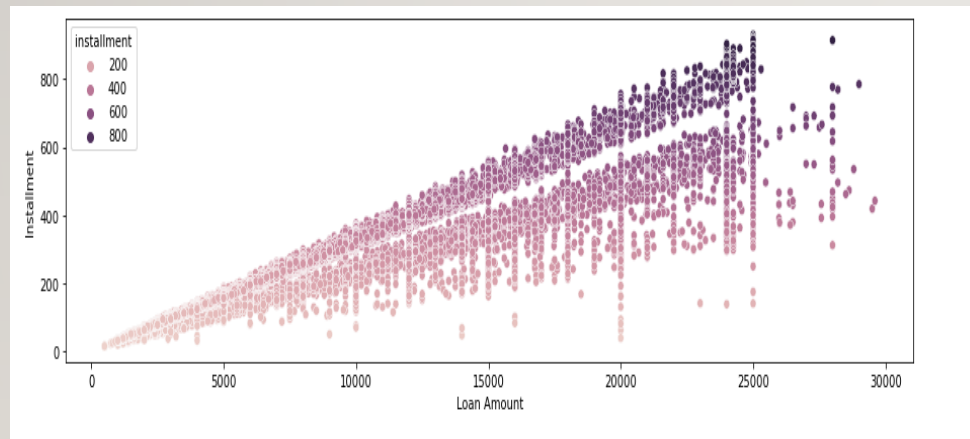
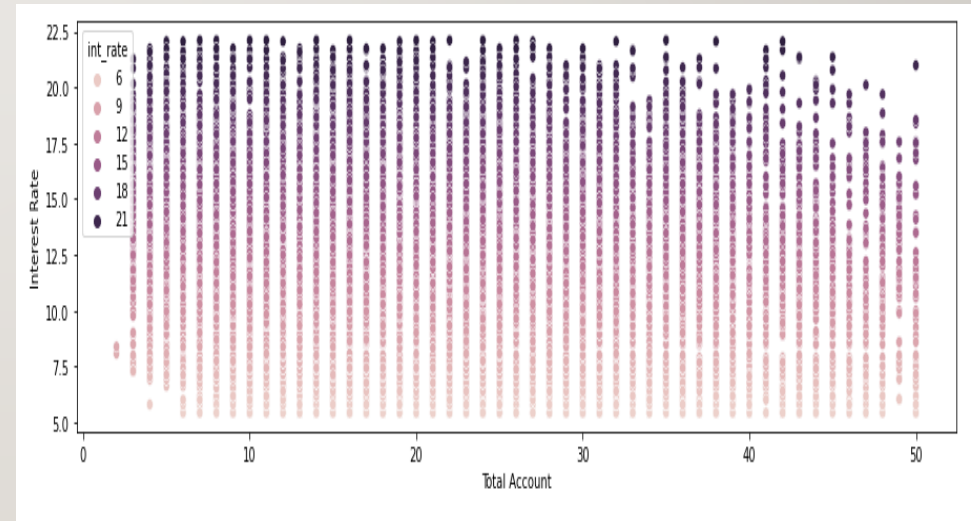Positive relation between annual income and Loan Amount

BIVARIATE ANALYSIS / MULTI VARIATE ANALYSIS CONTD.

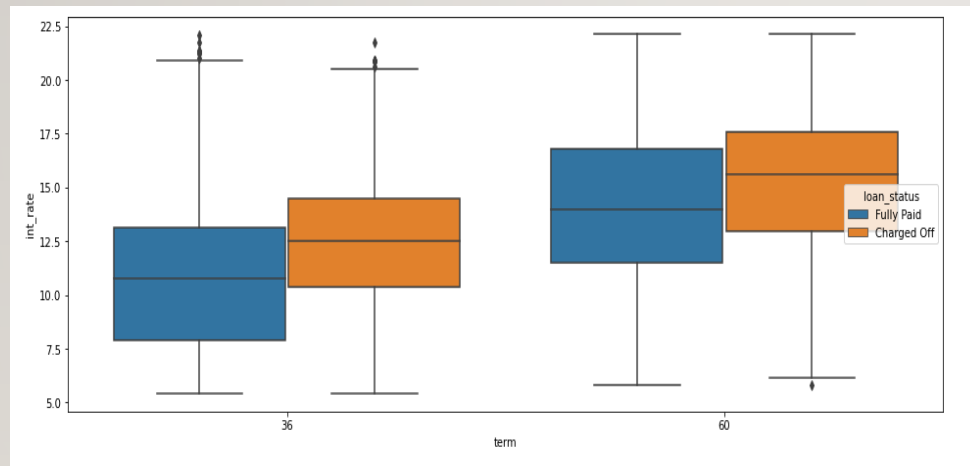We can see negative relation between annual income and dti

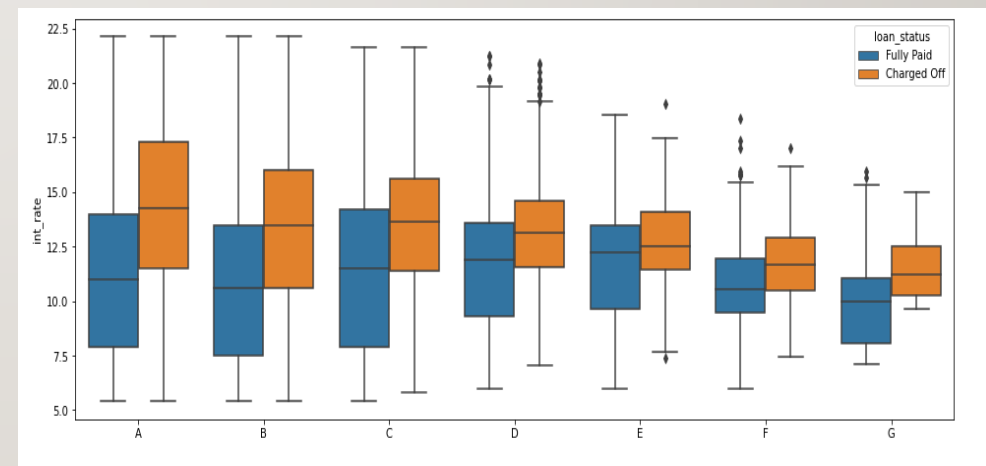Negative relation between total credit line and interest rate

BIVARIATE ANALYSIS / MULTI VARIATE ANALYSIS CONTD.

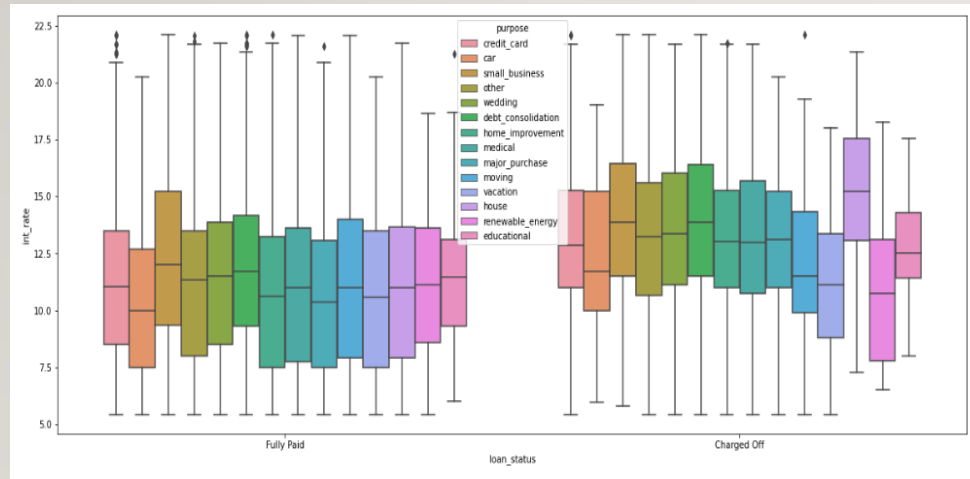Defaulters are increasing when term and interest rate increases.

We can see irrespective of grade when interest increase defaulters are increasing.
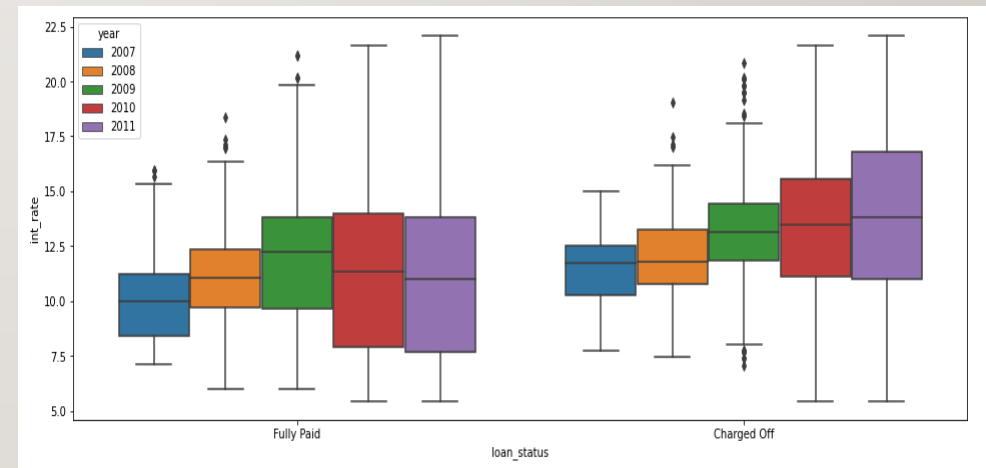
BIVARIATE ANALYSIS / MULTI VARIATE ANALYSIS CONTD.

We can see that each purpose defaulters are increasing when interest rate increase

We can see that each year defaulters are increasing when interest rate increases but 2011 has more charged offs.

CONCLUSION

- Columns or variables which are having chance to increase defaulters are:
  - Term
    - We can see from analysis charged-off rate increases when term increases.
  - Annual income
    - We can see a negative relation between annual income and charged-off rate.i.e., lesser annual income more defaulters
  - Interest rate
    - We can see when interest rate increases the defaulters also increases.
  - Employee length
    - We can see when employee length increases then the defaulters also increases.
  - Verification status
    - We can observe that defaulters are more when not verified.
  - Home ownership
    - Defaulters are increased incase when ownership is rented

RECOMMENDATIONS

- Reduce the loans for employees who have lesser annual income.

- Consider employees who are greater than 1 year and less than 10+ years to avoid more charged offs.

- Reduce interest rate below 20% to reduce defaulters.

- Approve loans only to verified employees.