

LENDING CLUB CASE STUDY

Problem statement

- From the given data we need to understand the Borrower's behavior.
- There are three type of Borrower
 1. Fully Paid (Non-Defaulted)
 2. Charged Off (Defaulted)
 3. Current
- With the help of EDA we need to analyse Borrower's personal, financial and other factors which effects loan return.
- Analysis with respect to business prospective which Borrower will be non defaulter & which one will be defaulter.

Approach to solve case study

- Data understanding
- Data cleaning
- Exploratory Data Analysis
 1. Univariate analysis
 2. Bivariate analysis
 3. Segmented univariate analysis
 4. Multivariate analysis

Data understanding

- Removed columns which contained more than 50% null values.
- Removed rows which contained more than 50% null values.
- The remaining data was studied with respect to business and on the basis of that irrelevant data was removed.
- Finally the data set ready with following columns
 1. Loan amount
 2. Interest rate
 3. Employee length
 4. Grade and sub grade
 5. Home ownership
 6. dti
 7. Annual income
 8. The number of delinquency in past 2 years
 9. Number of public record bankruptcies
 10. Loan status

Data cleaning

- Converted data in proper format.
- Replaced NaN values with zeros.
- Removed unnecessary symbols like '%', '<', 'year', '>', '+' etc from int_rate and emp_length for analysis.
- Converted numeric data from object to float/int.

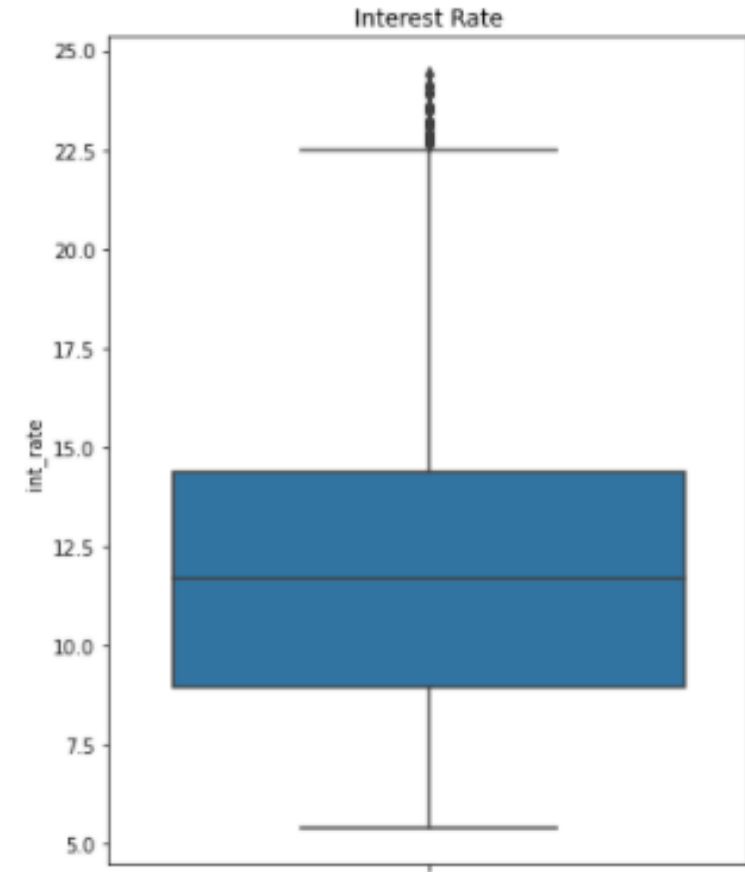
EDA Analysis

- Checked basic characteristics of data

	id	loan_amnt	int_rate	emp_length	annual_inc	dti	delinq_2yrs	pub_rec_bankruptcies
count	3.857700e+04	38577.000000	38577.000000	37544.000000	3.857700e+04	38577.000000	38577.000000	37880.000000
mean	6.763787e+05	11047.025430	11.932219	5.061048	6.877797e+04	13.272727	0.146668	0.043479
std	2.092639e+05	7348.441646	3.691327	3.401750	6.421868e+04	6.673044	0.492271	0.204583
min	5.473400e+04	500.000000	5.420000	1.000000	4.000000e+03	0.000000	0.000000	0.000000
25%	5.120330e+05	5300.000000	8.940000	2.000000	4.000000e+04	8.130000	0.000000	0.000000
50%	6.564230e+05	9600.000000	11.710000	4.000000	5.886800e+04	13.370000	0.000000	0.000000
75%	8.291460e+05	15000.000000	14.380000	9.000000	8.200000e+04	18.560000	0.000000	0.000000
max	1.077501e+06	35000.000000	24.400000	10.000000	6.000000e+06	29.990000	11.000000	2.000000

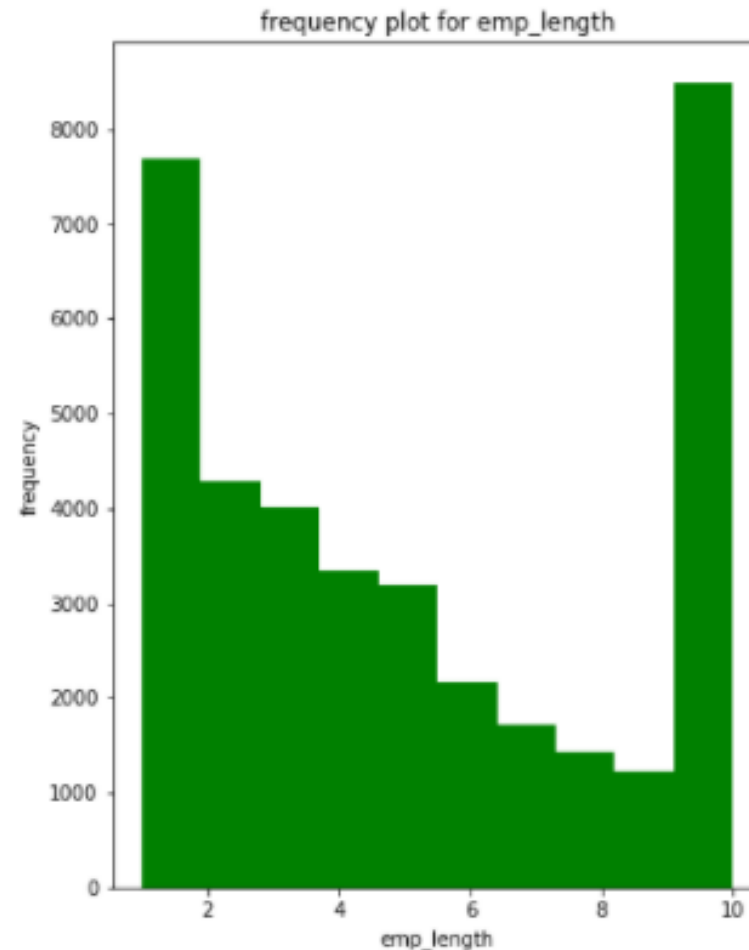
Univariate analysis

- Analysis is carried out on single variable, which gives basic information which helps to understand borrower's nature
- Interest rate: most of the Borrowers fall in the interest rate (9-14% range)
- maximum interest rate - 22.5
- minimum interest rate - 6
- mean interest rate - around 11.5
- Outliers - Above 22.5



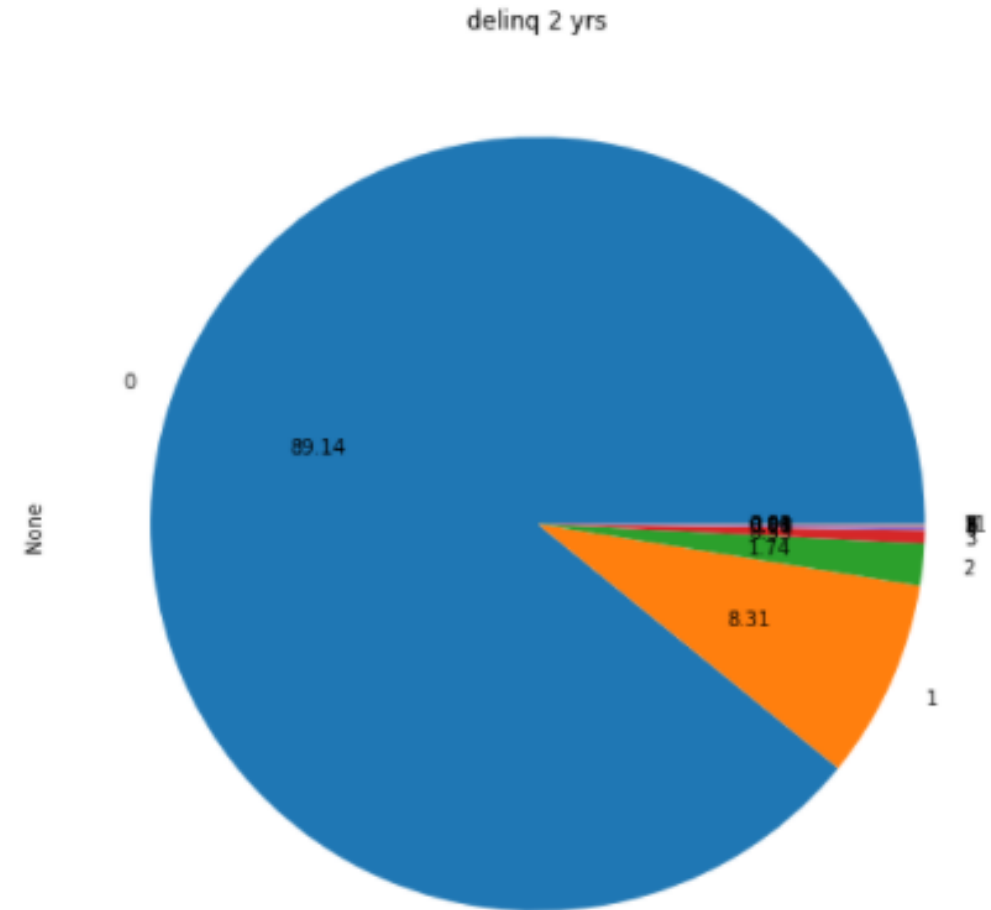
Univariate analysis

- Employee service length
- Employee with 10+ years of experience are highest in number.
- Employee with less than 1 year experience are also higher in number.



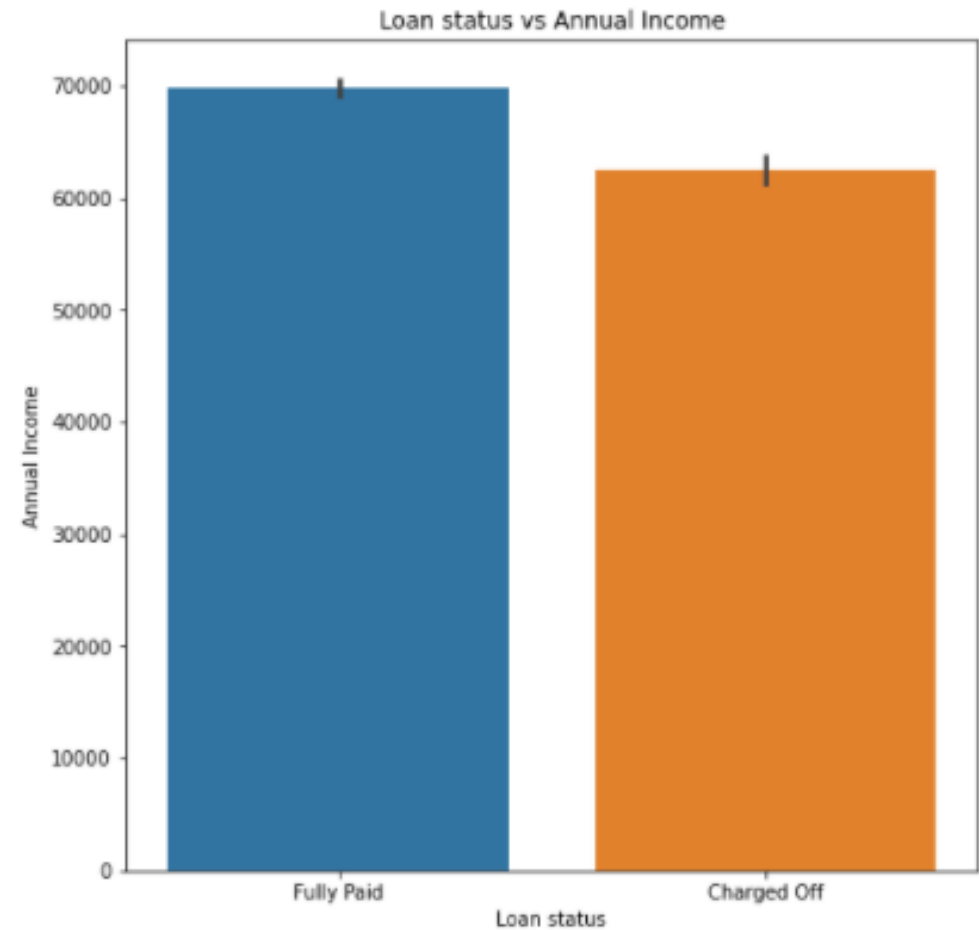
Univariate analysis

- The number of delinquency in past 2 years
- Maximum borrowers have no delinquency



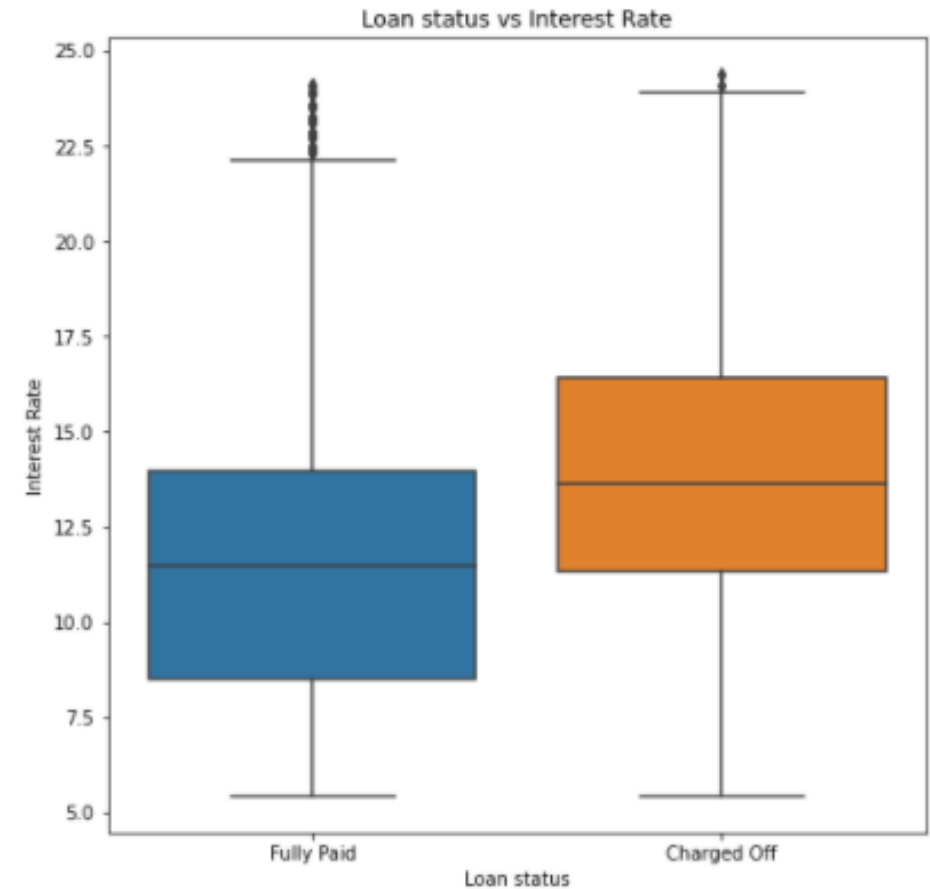
Bivariate analysis

- Loan status vs annual income:
- Non defaulter has more annual Income than defaulter.



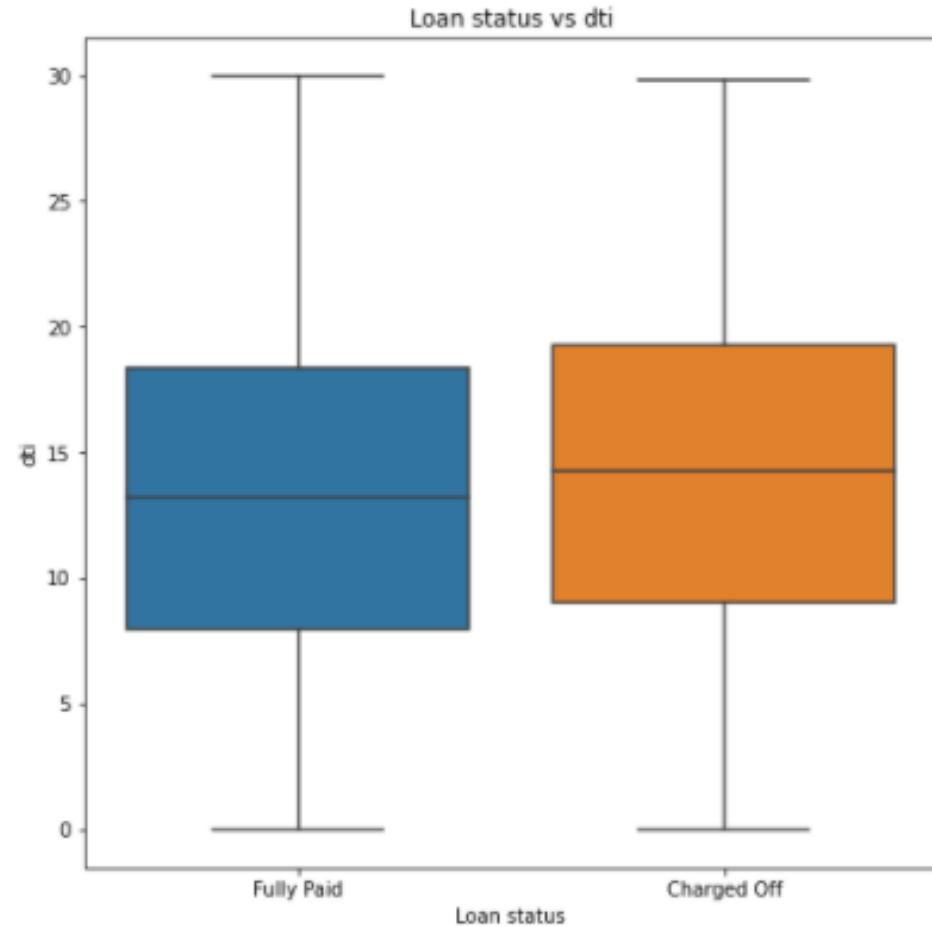
Bivariate analysis

- Loan status vs interest rate:
- For higher interest rate the number of defaulter is more in number.



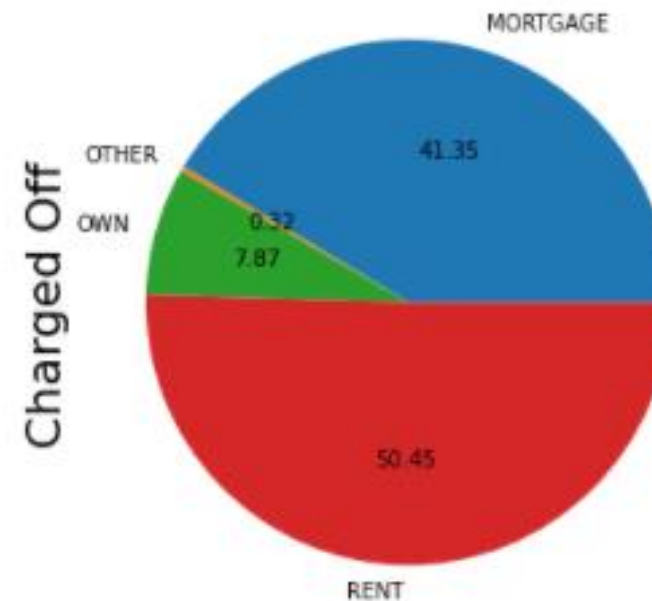
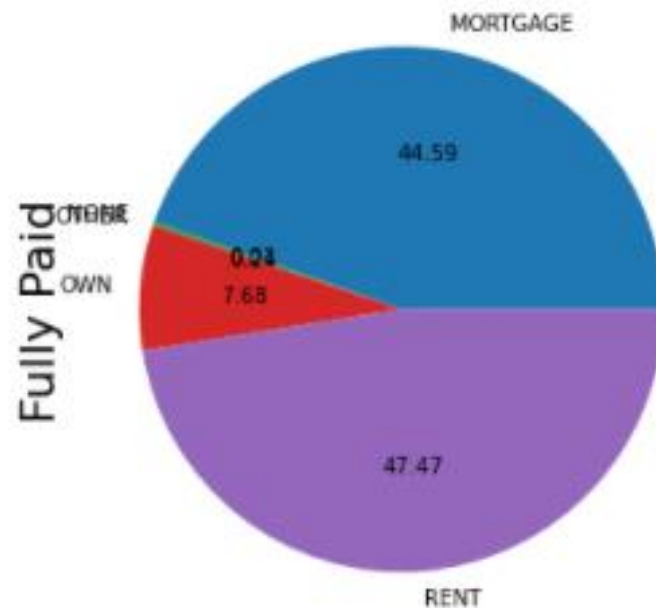
Bivariate analysis

- Loan status vs dti (debt to income ratio):
- Non defaulter has lower dti than defaulter.



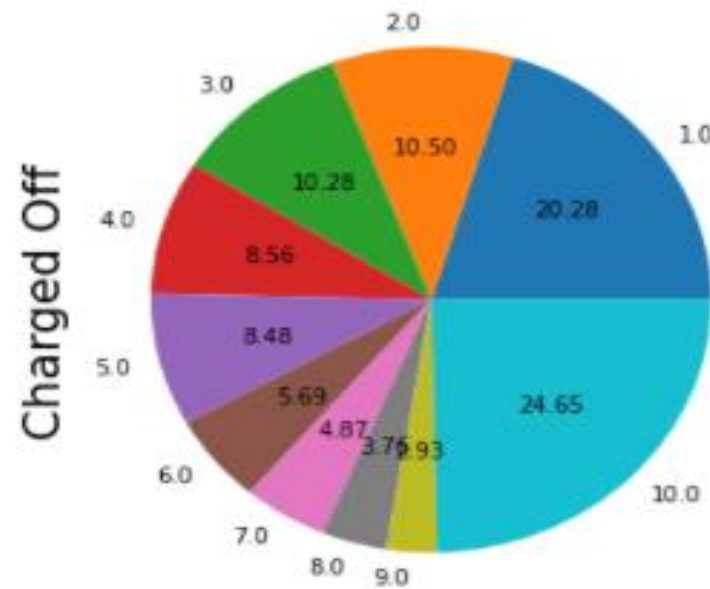
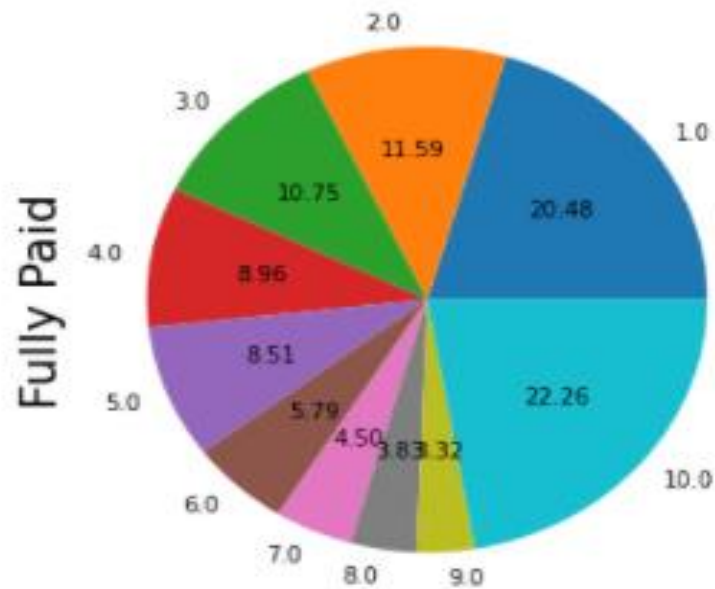
Segmented univariate analysis

- Loan status vs home ownership:
- In defaulter class most people live in rented house.



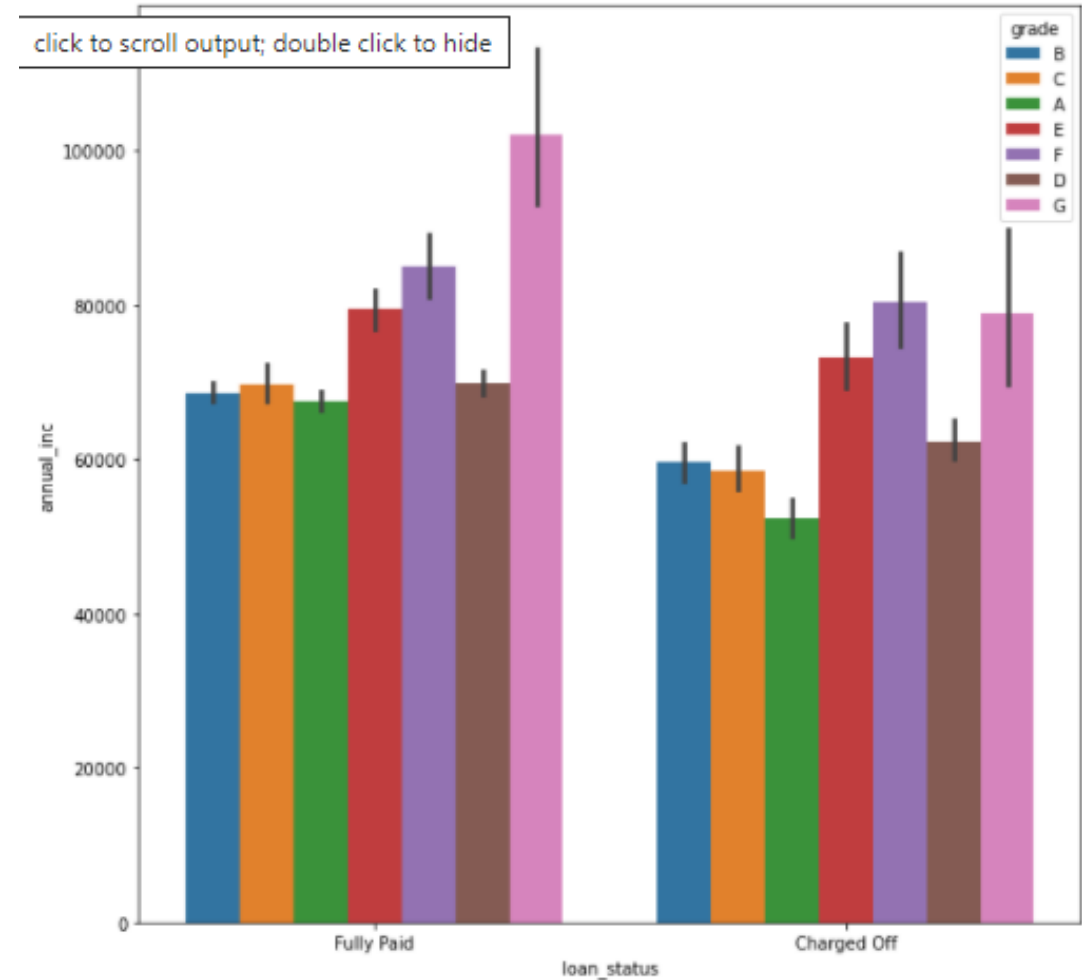
Segmented univariate analysis

- Loan status vs Employee service length
- Employee service length does not have much effect on loan status.



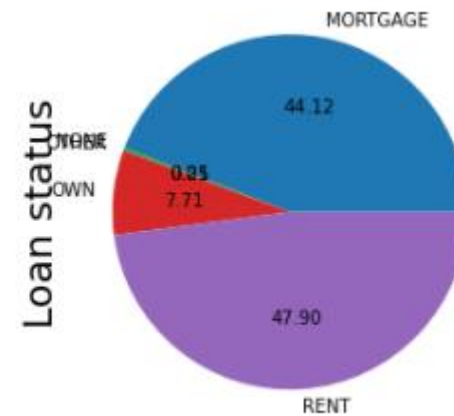
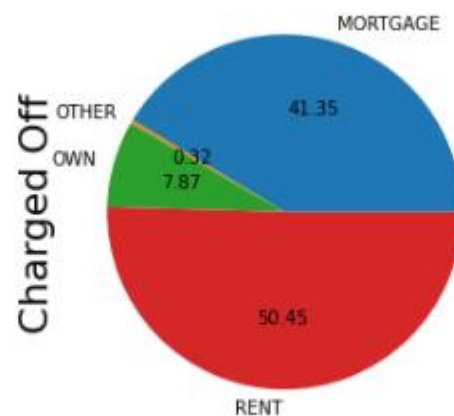
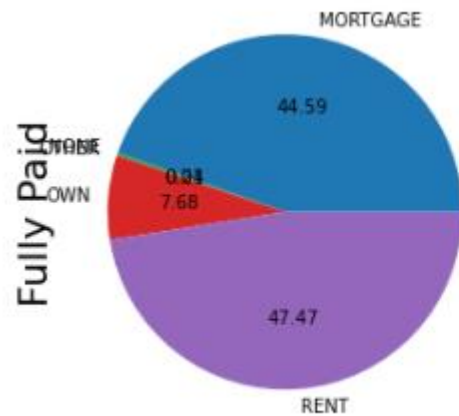
Multivariate analysis

- Comparison of annual income, loan status and grade
- For each grade category fully paid have more annual income as compared to charged off ones.



Multivariate analysis

- Comparison of home ownership vs loan status
- Borrowers who are charged off mostly live on rent and constitute almost 50% strength which is more as compared to that of fully paid



Recommendations

- Borrower's having higher income tend to pay their loans completely hence they are non defaulters.
- With less interest rate the chances of recovery of loan is more.
- Most of the non defaulter shave their own home
- Lower the value of dti, higher is the chance of getting loan back.
- Non defaults have zero delinquency in the past years.

Lending club case study submitted by

Rashmi Jennefer

Control Engineer

TMIEC India private Limited, Hyderabad

Harshal Jamode

SME Mathematics

Infinity learn, Hyderabad

Thank you!!!