Lending club case study

Loan Dataset

Santosh K U 9/2/2022

Overview :

Basic understanding and expectation :

- The Case study is provided with a dataset consisting details of the various loan transactions

- The dataset consists loan details of ongoing accounts, completely paid and charged off accounts.

- Expectation is to find the possible reasons for defaulting. (for charged off cases)

Assumption :

- The expectation is to find the possible reasons for defaulting.

- The loans which are currently active, cannot provide the insight for defaulting.

- Hence the data related to active loan accounts is removed for the analysis.

Data Cleaning :

- Data cleaning techniques such as null values handling and outlier detection is applied.

- Columns with more than 85%null values removed from analysis.

- For certain columns of interest the outliers are detected and removed.

Univariate Analysis :

Key details:

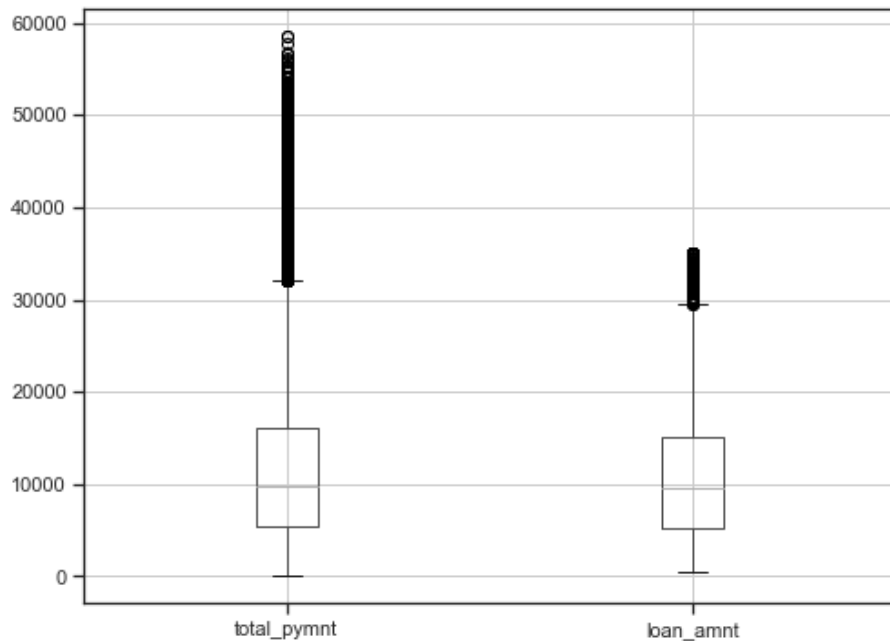- About 85% of the loan is fully paid and 15% of the loan is Defaulted
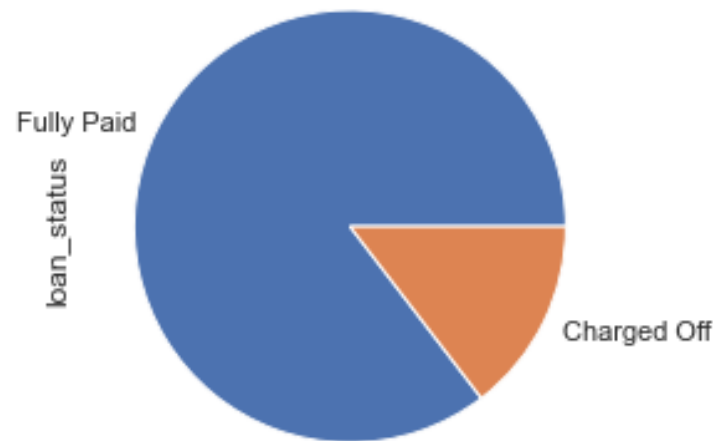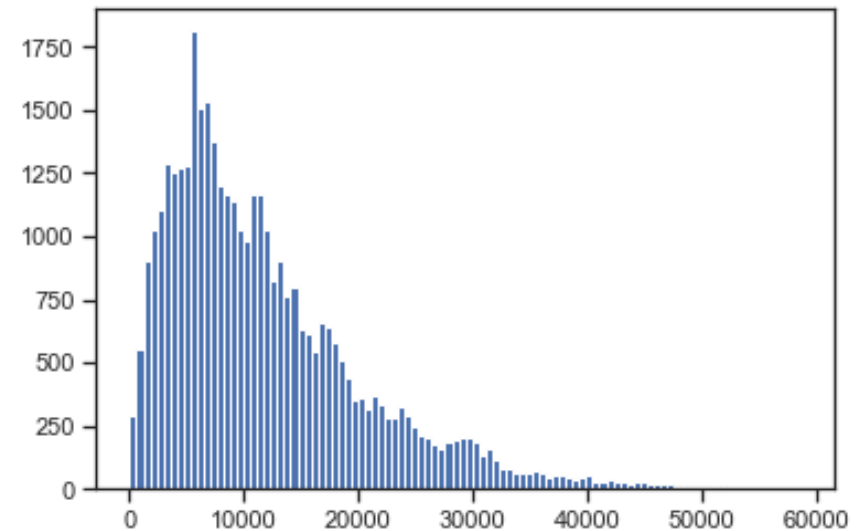

Pie chart depicting the loan status

Univariate Analysis :

Key details:

- In the analysis, we see lot of outliers,

- From the graph we can infer that majority of the loan amounts resides below 1.5 lakh.

Loan and total payment comparison

Histogram of total payment

Bivariate Analysis :

Key Details:

- When loan Status was checked with loan grade, majority of the loans are from grade A, B and C.

- But there is strong possibility that defaulting

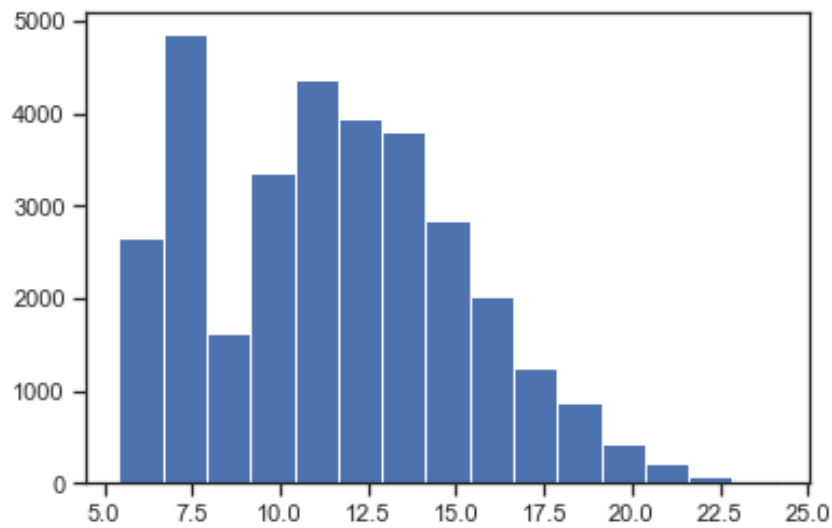- for loan types G,F,E and D respectively

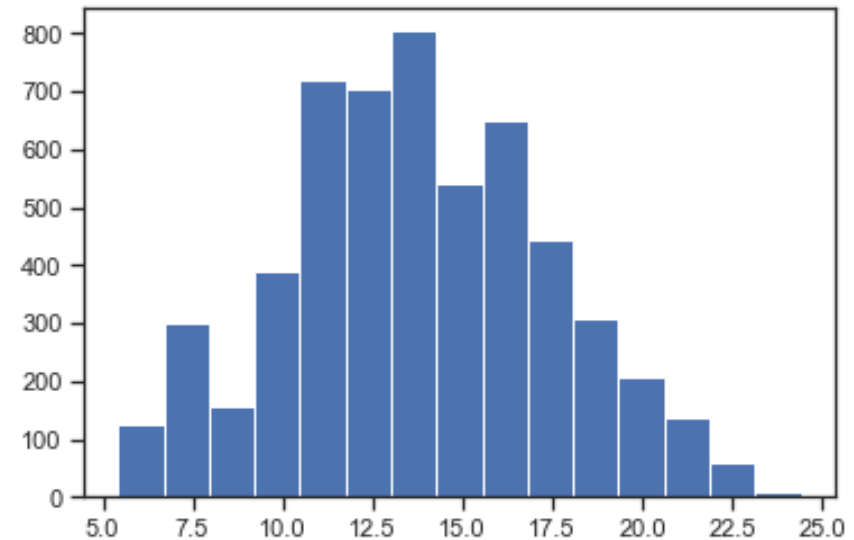| loan_status	grade	id
Charged Off	A	602
	B	1425
	C	1347
	D	1118
	E	715
	F	319
	G	101
Fully Paid	A	9443
	B	10250
	C	6487
	D	3967
	E	1948
	F	657
	G	198

Grade vs Loan status

Bivariate Analysis :

Key Details:

- Defaulting on loan has increasing trend when interest percentage is more than 10%
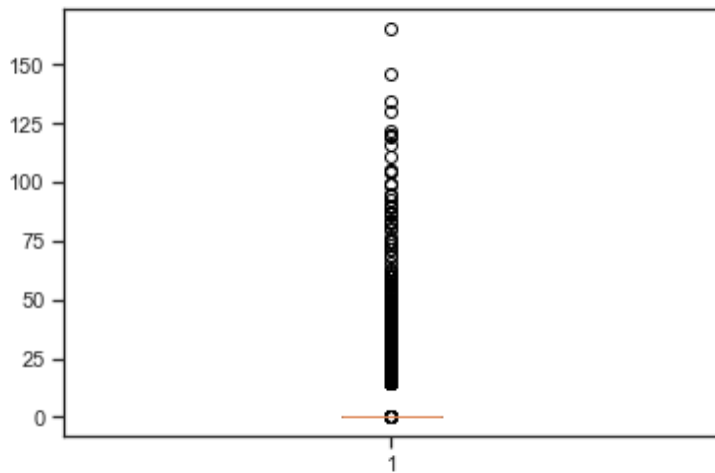

Histogram of interest rates of fully paid loans

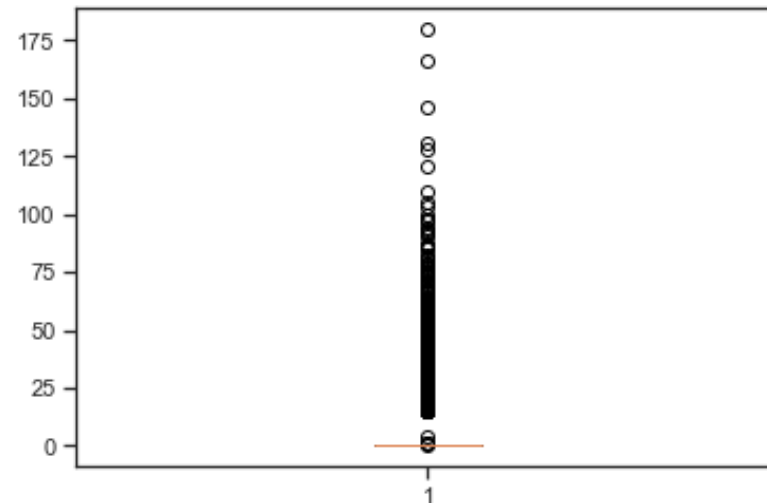Histogram of interest rates of charged off loans

Bivariate Analysis :

Key Details:

- Average late fees is high for charged off accounts, this indicates that customers with late fees have high probability of defaulting.

- Note : since the outliers are similar average are compared.

Box plot of late fees of fully paid accounts
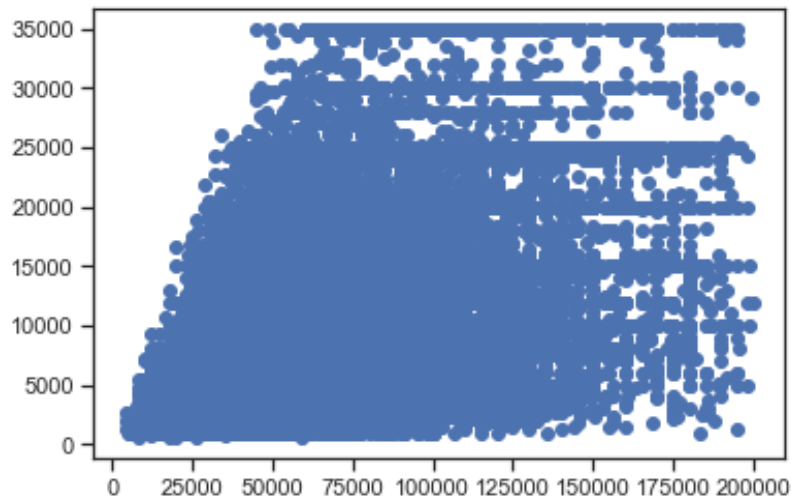Average late fee for fully paid = 0.86

Box plot of late fees of Charged off accounts
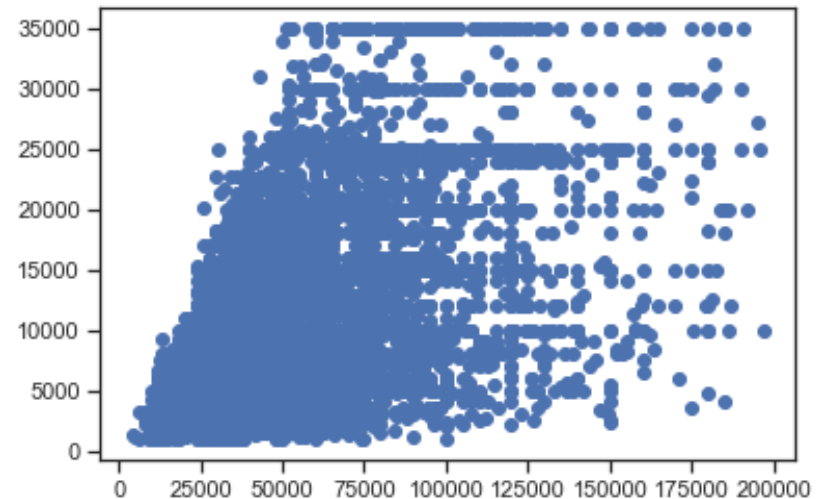Average late fee for fully paid = 4.27

Bivariate Analysis :

Key Details:

- The loan amounts close to 75000 with annual income less that 25000, show more instances of defaulting.

Scatter plot between annual inc vs loan amount of full paid accounts

Scatter plot between annual inc vs loan amount of charged off accounts

Conclusion :

- Percentage of defaulted accounts is 15.

- Majority of loan amount resides below 1.5lakhs.

- Defaulting of loan has increasing trend as interest rate increase greater then 10%

- Average late fees is high for charged off accounts, this indicates that customers with late fees have high probability of defaulting.

- The loan amounts close to 75000 with annual income less that 25000, show more instances of defaulting.

Additional Details :

Points:

- Majority of loans are taken by members whose houses are on mortgage or they stay in rented house. Percentage of defaulting between the categories is similar.

- Generally people having work experience greater than 10 years have most number of loans, there is no significant pattern with defaulting based on years of work experience.

Thank you