Lending Club Case Study

Satyanarayana k

&

Vinoth k

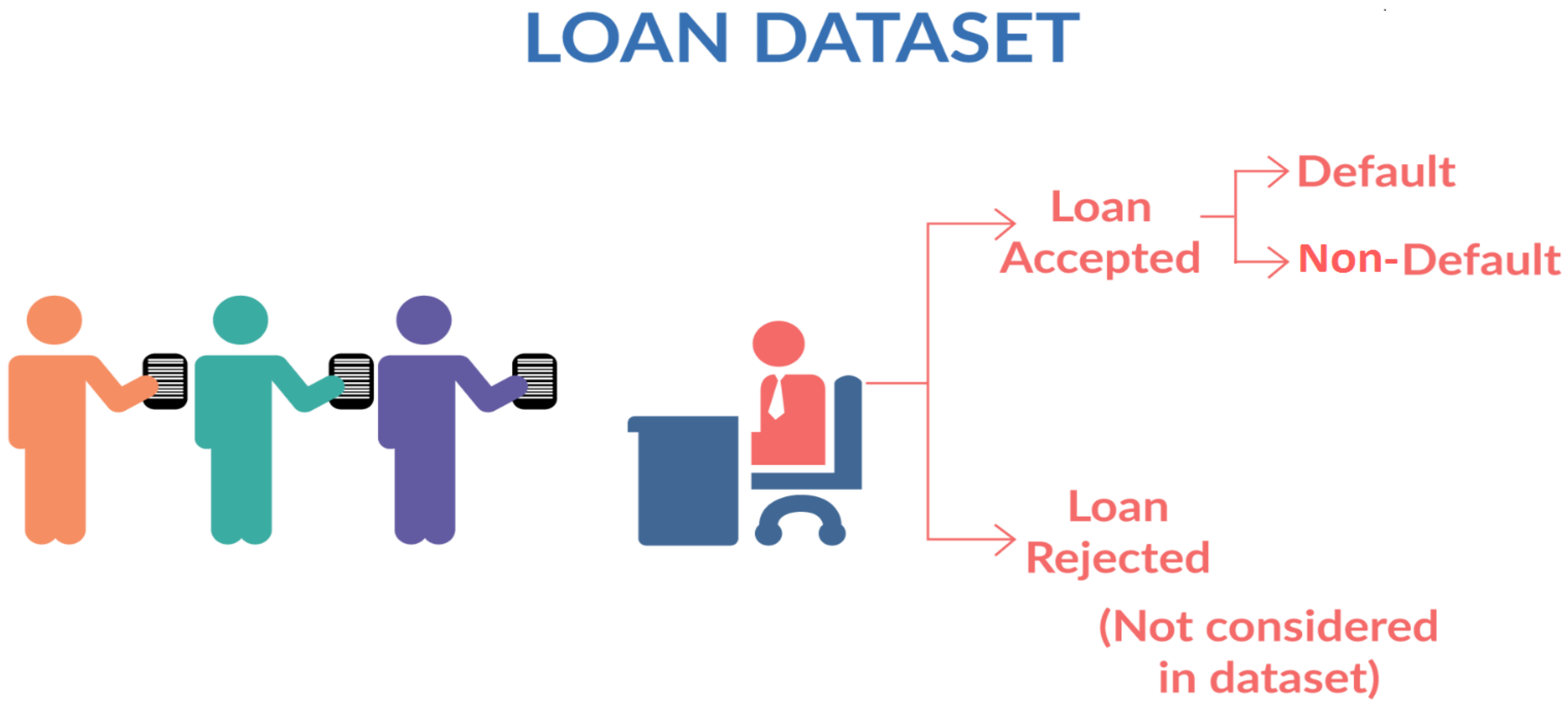
Business Understanding

You work for a **consumer finance company** which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

1. If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
2. 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
3. The data given below contains the information about past loan applicants and whether they 'defaulted' or not. 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.

Business Understanding

- In this case study, you will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.



Problem Statement

1. This company is fintech company online loan marketplace for smaller loan like personal loan, car loan , credit card loan etc. Borrowers can easily access lower interest rate loans through a fast online services.
2. Lending loans to risky applicants is the huge financial loss called credit loss for the company.
3. To filter out such applicants we are going to deploy a EDA model to understand who are going to be charged-off are the defaulters.
4. We are in need to find-out the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
5. To develop our understanding of the domain we are going to analysis the data set and driving the useful variable from it to provide the possible solution for it.

Problem Statement

We are going to do the following steps to get insights from the dataset:

1. Data sourcing
2. Data cleaning
3. Univariate analysis
4. Segmented Univariate analysis
5. Bivariate analysis

Data sourcing

1. The dataset by Fintech company . It contains the complete loan data for all loans issued through the time period 2007 to 2011.
2. The dataset is in .CSV format.
3. The fintech company provide the data dictionary which describes the meaning of these variables.

In [3]:
```
df = pd.read_csv("loan.csv", low_memory=False)
```

In [4]:
```
df.head()
```

Out[4]:

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_past_
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	
3	1076863	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	C1	...	NaN	
4	1075358	1311748	3000	3000	3000.0	60 months	12.69%	67.79	B	B5	...	NaN	

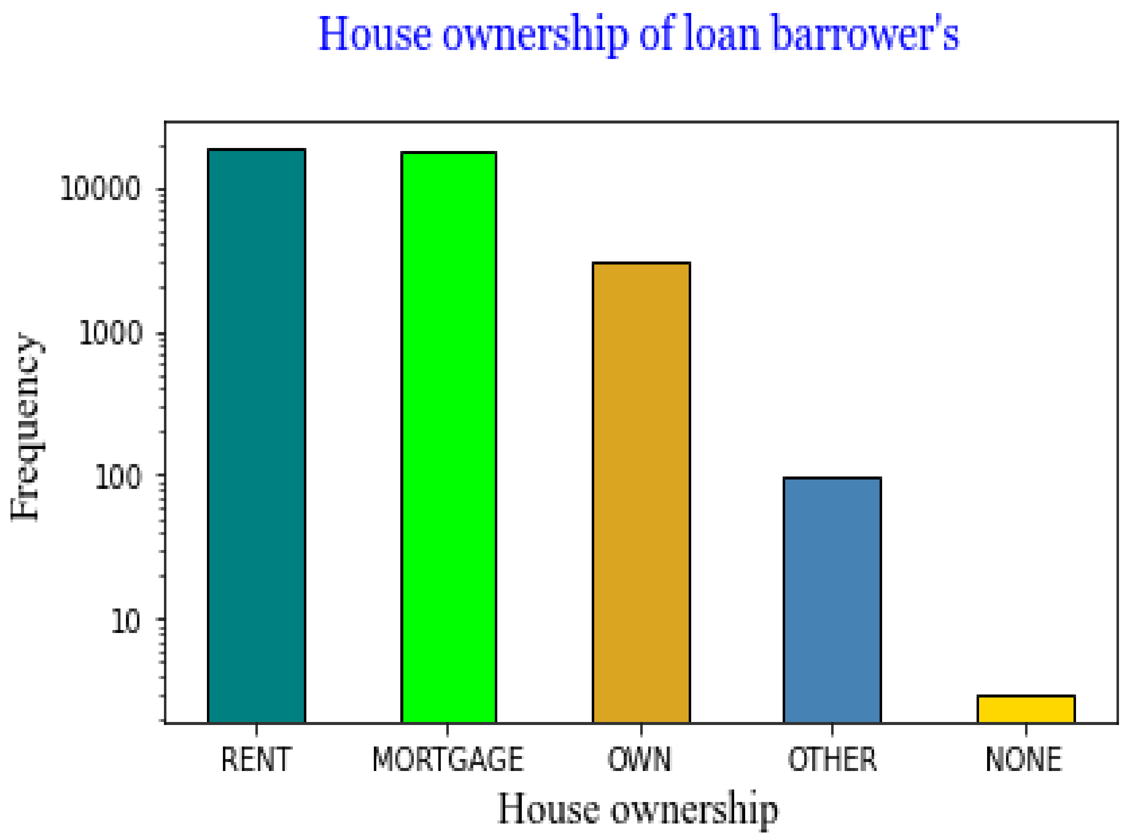
5 rows × 111 columns

Data cleaning

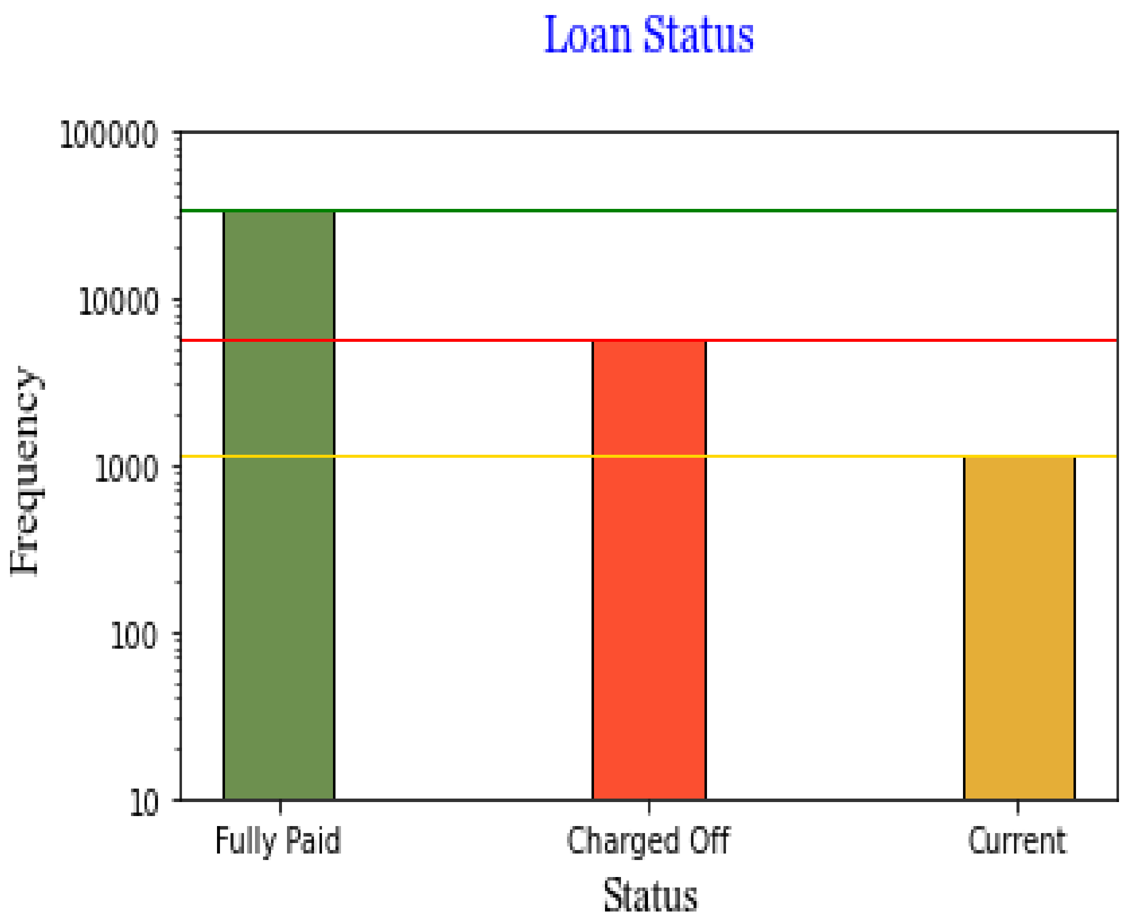
1. Loaded the dataset into the pandas Data Frame by using jupyter notebook.
2. Check sanity checks
3. Started the Data cleaning for data set
4. Fixing the Rows and Columns
5. Dropping all the columns having NaN in all of the rows
6. Taken case the missing values
7. Dropped the columns as majority of the rows are NaN columns

Univariate Analysis

As we can see Rent and Mortgage owner are taking more loan than others.

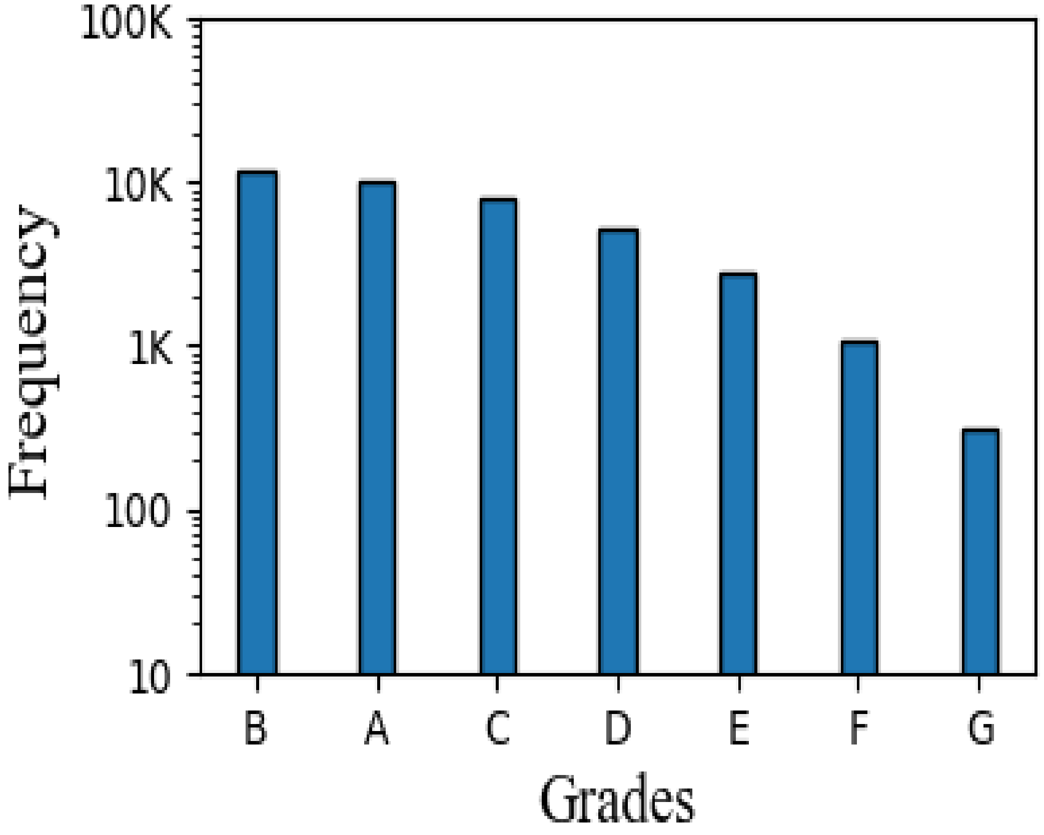


Now we need concentrate on charged off status and current only Than fully paid.

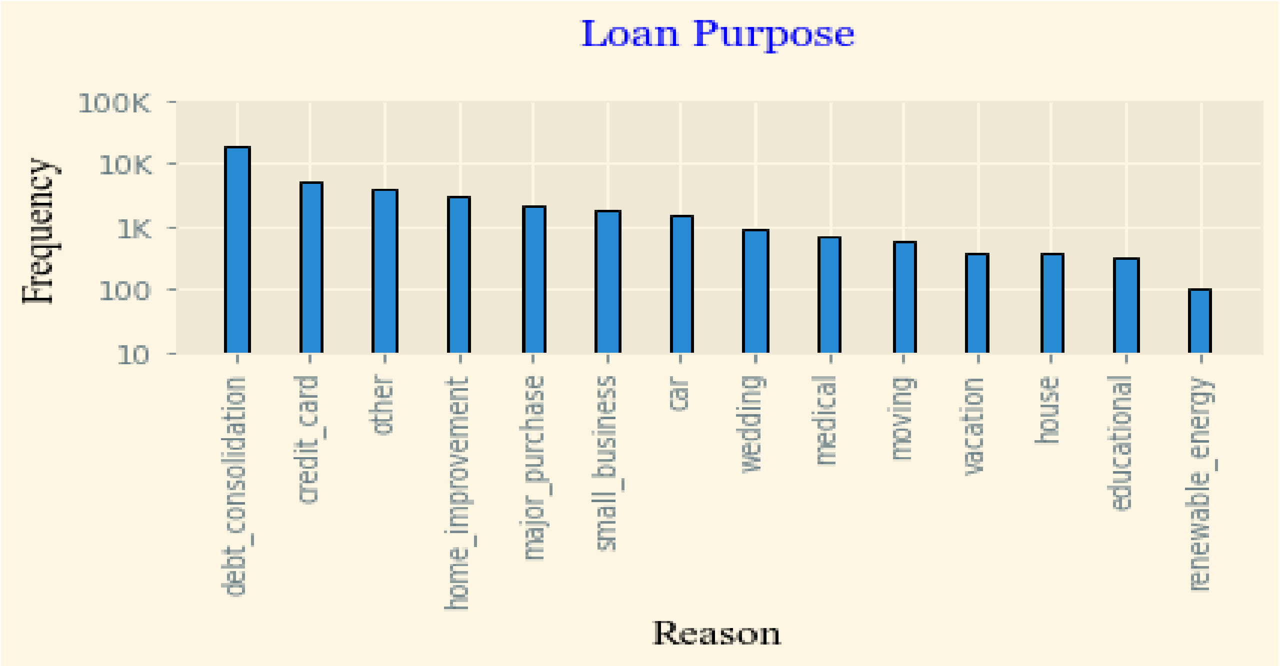


Univariate Analysis

The below plot shows the frequency of borrowers falls under each grades

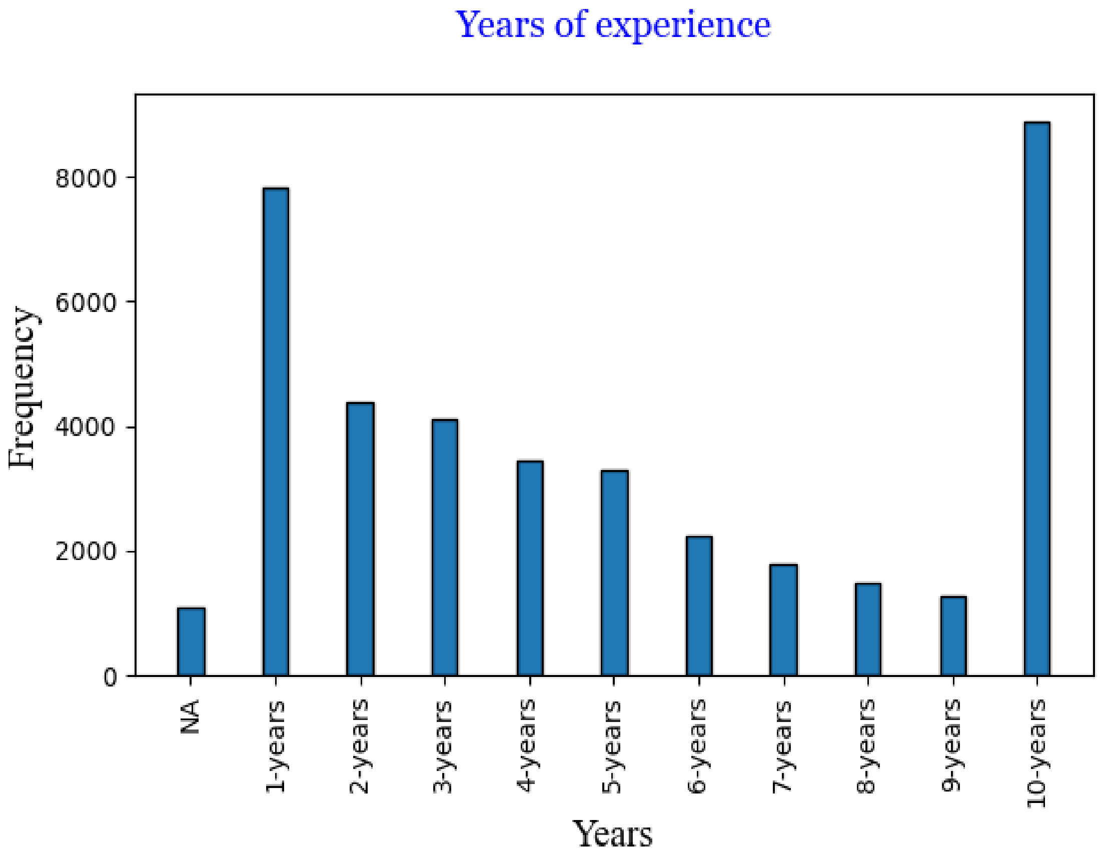


The below plot shows the frequency of borrowers who falls under different loan purpose. It seems that more number of people are in debt.

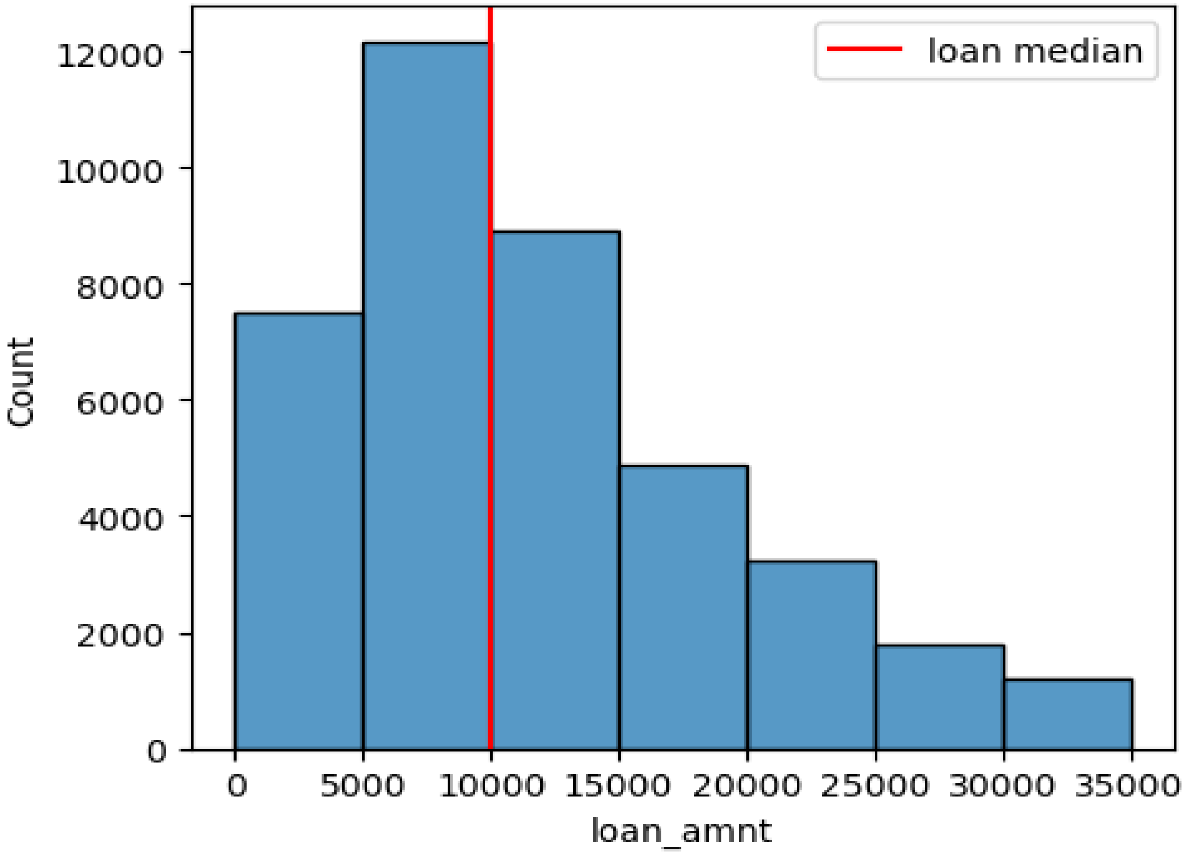


Univariate Analysis

It is very obvious that borrower's are more in 10 years of experience as well as in 1 year of experience



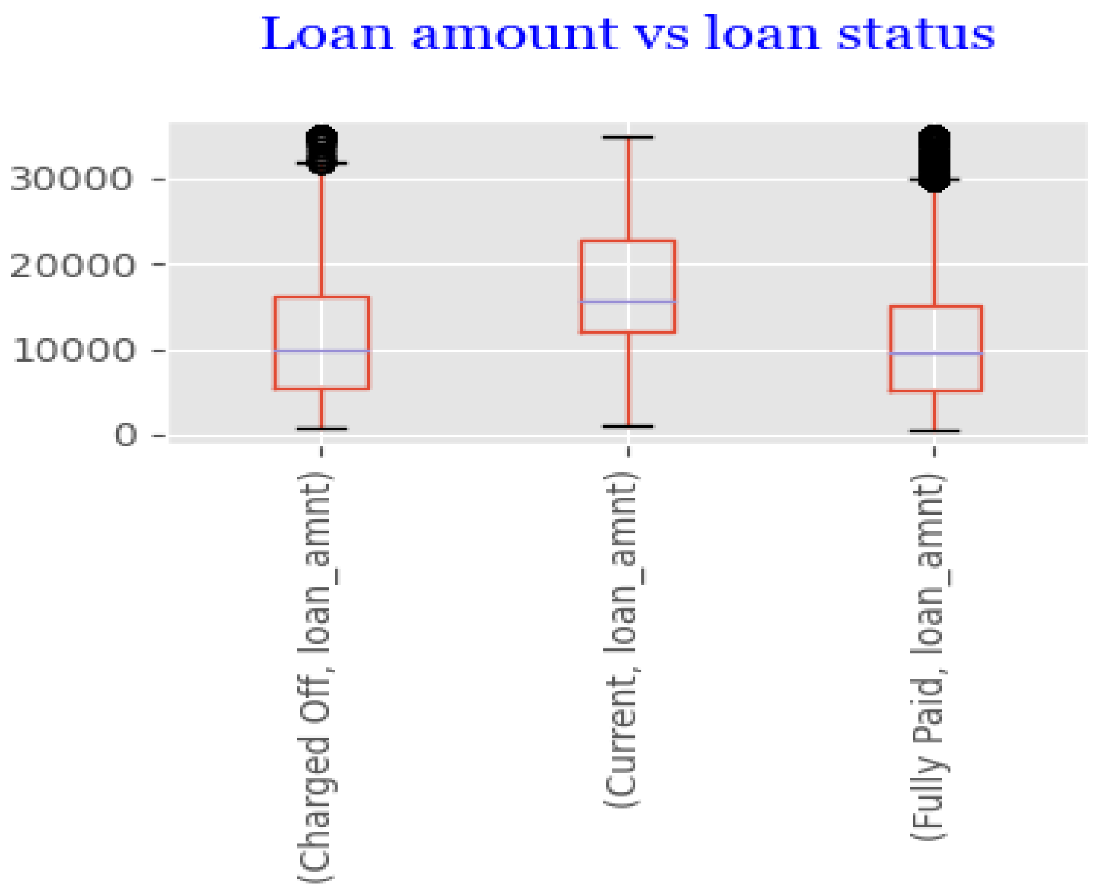
Loan amount of the borrowers and median is around 10k as we can see in the below chart.

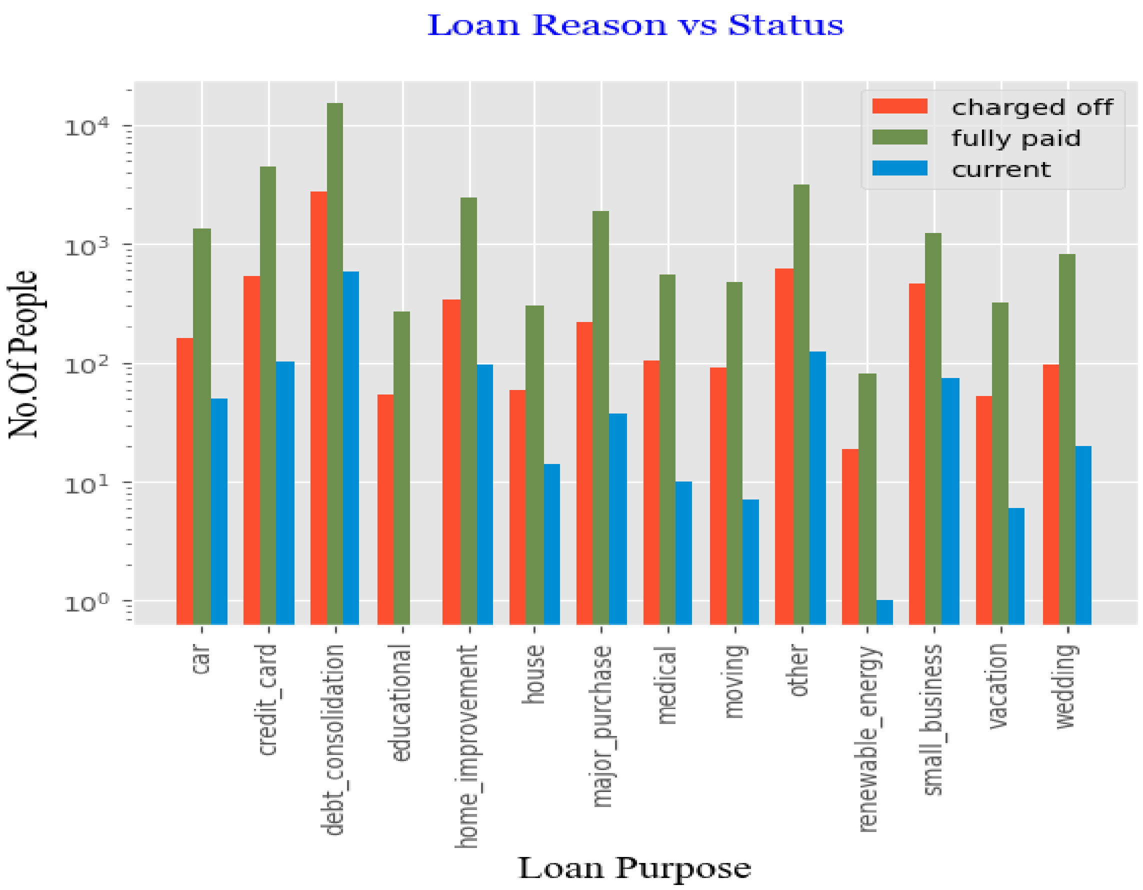


Segmented Univariate Analysis

Here we can see in box plot the mean and median and outliers for the loan amount vs loan status.

- The defaulters are spreader across the almost all the categories of the loan purpose.
- However they are very more in number for debt.



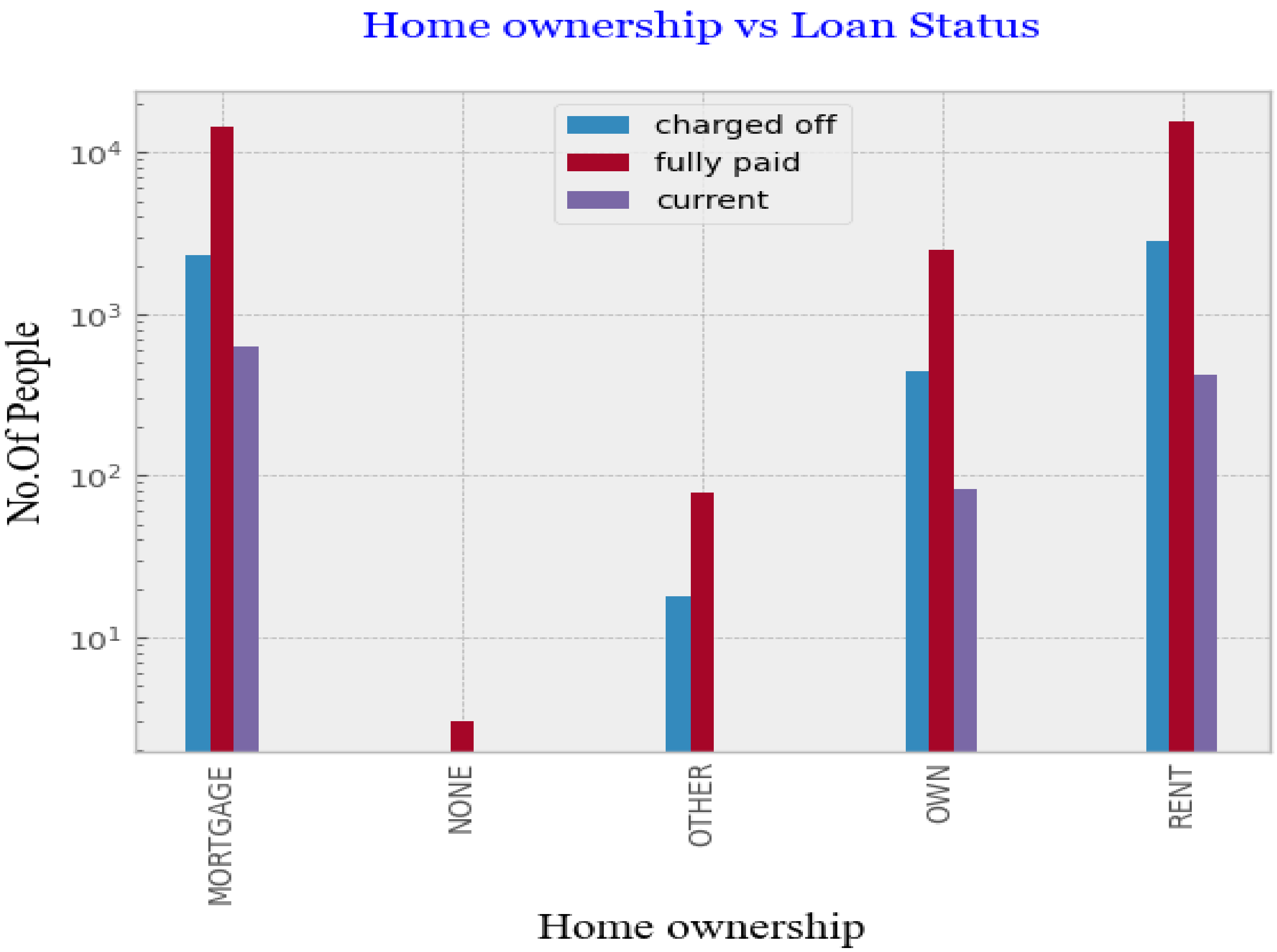


Segmented Univariate Analysis

The below graph shows the distribution of barrowers with respect to the house ownership and loan status category.

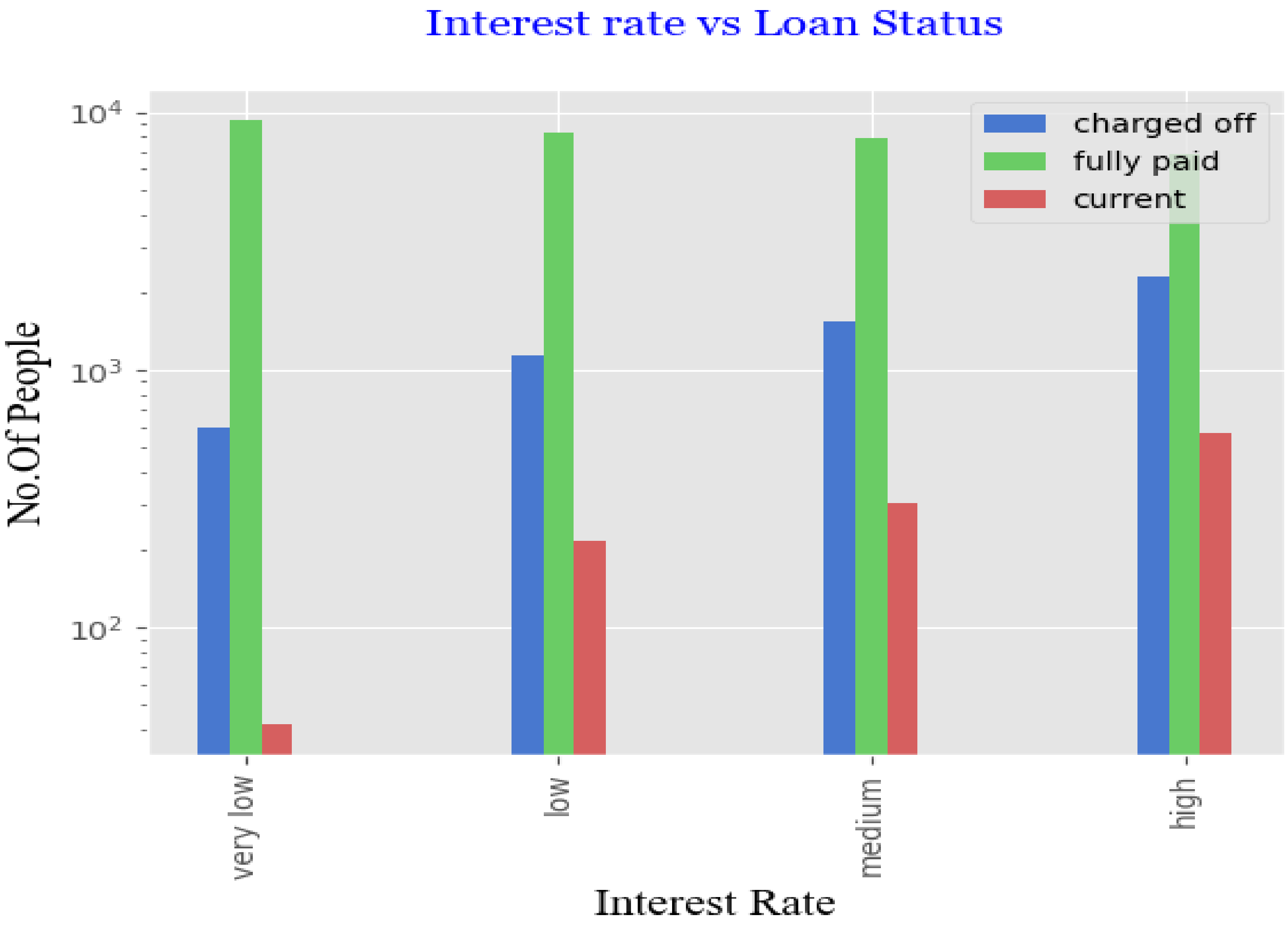
- It seems that the people who don’t have their own house are those who are in high need of loan.

- That is may be their loan purpose can be for home loan as well.



Segmented Univariate Analysis

The below graph shows the distribution of barrowers with respect to the interest rate and loan status category.

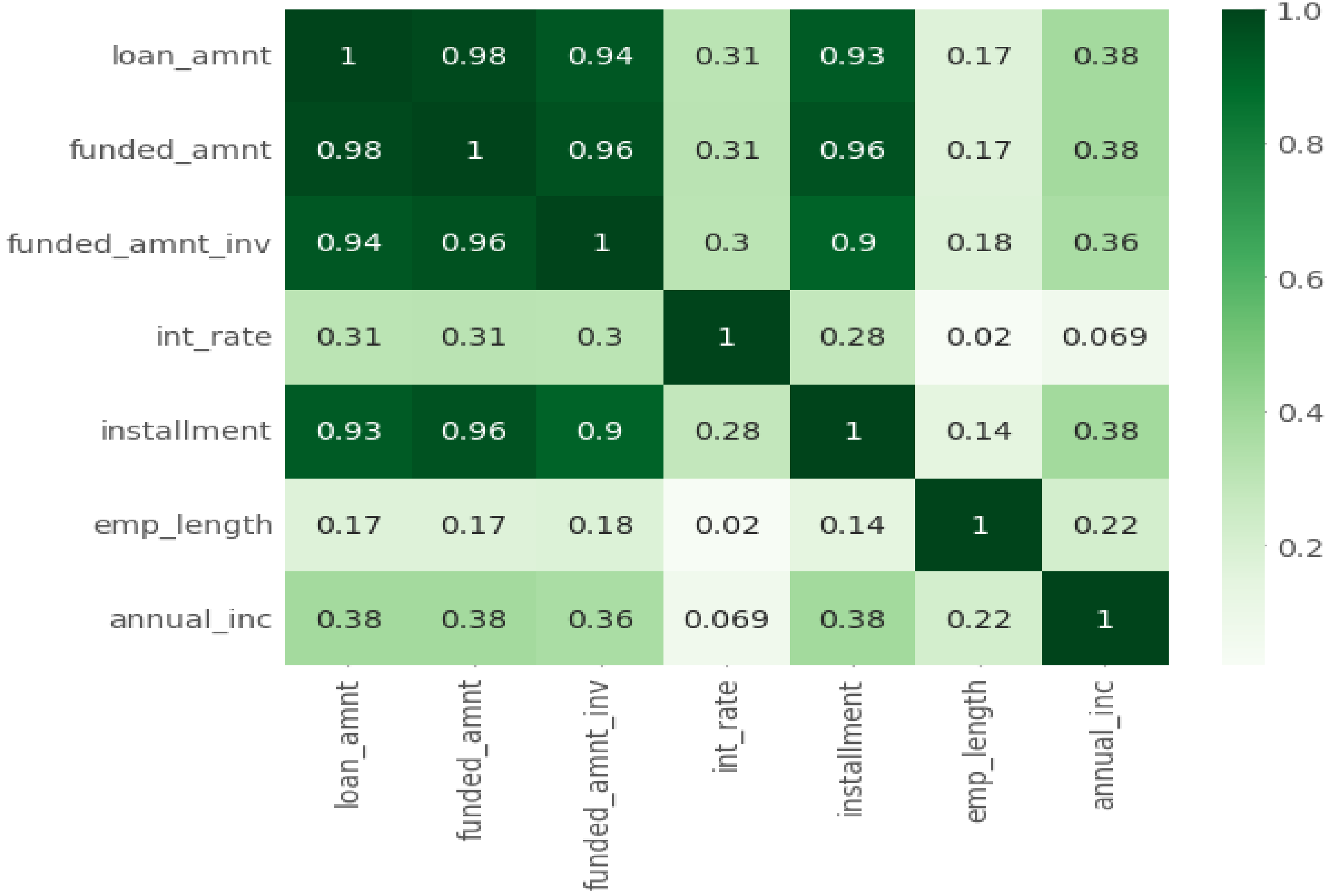


- The scale in which the graph is plotted is logarithmic.
- Interest rate clearly shows that there is linear increase in the no of people falling under loan rates varying from lowest to the highest.

Bivariate Analysis

The heat map is plotted against the very most important continuous variables from the data set.

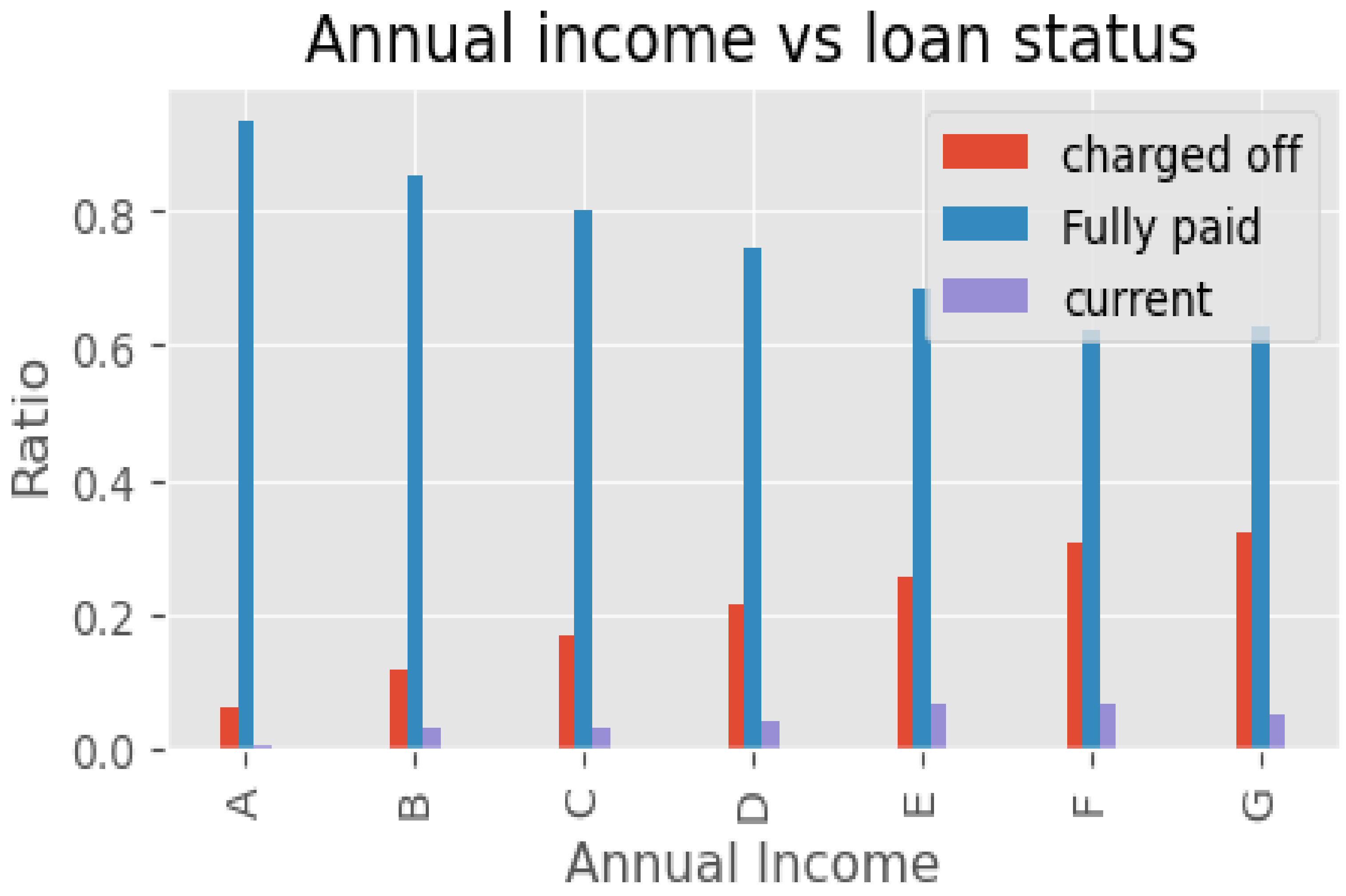
- Loan amount is highly correlated with the funded amount and funded amount inv which states that almost every borrower's are getting their loan amount for which they are quoted for.
- The installment is also highly correlated with funded amt and loan amt.
- Interest rate, employee length and annual income are not having so much correlated with anything.



Bivariate Analysis

The loan provided to the borrower's were almost falling under in all aspects of salary categories.

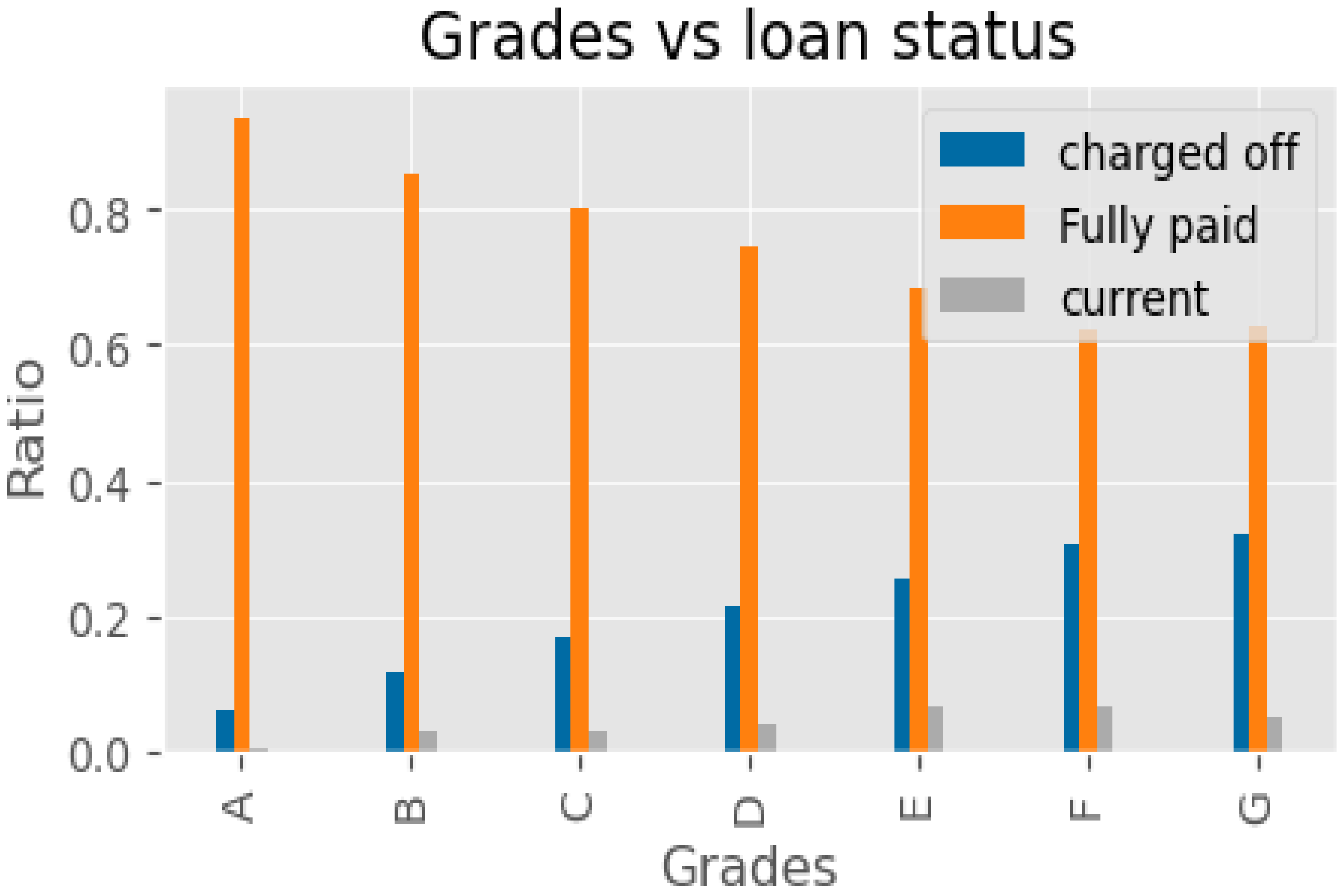
- The above statement is true for the defaulters as well.
- Here we can see that some interesting well common fact becomes true.
- More the higher salary, less the defaulters. Lesser the salary, higher the defaulters.
- Even though it's a common fact, the plot proves it as true.



Bivariate Analysis

Grade is one among the most important driving factor for charged off categories.

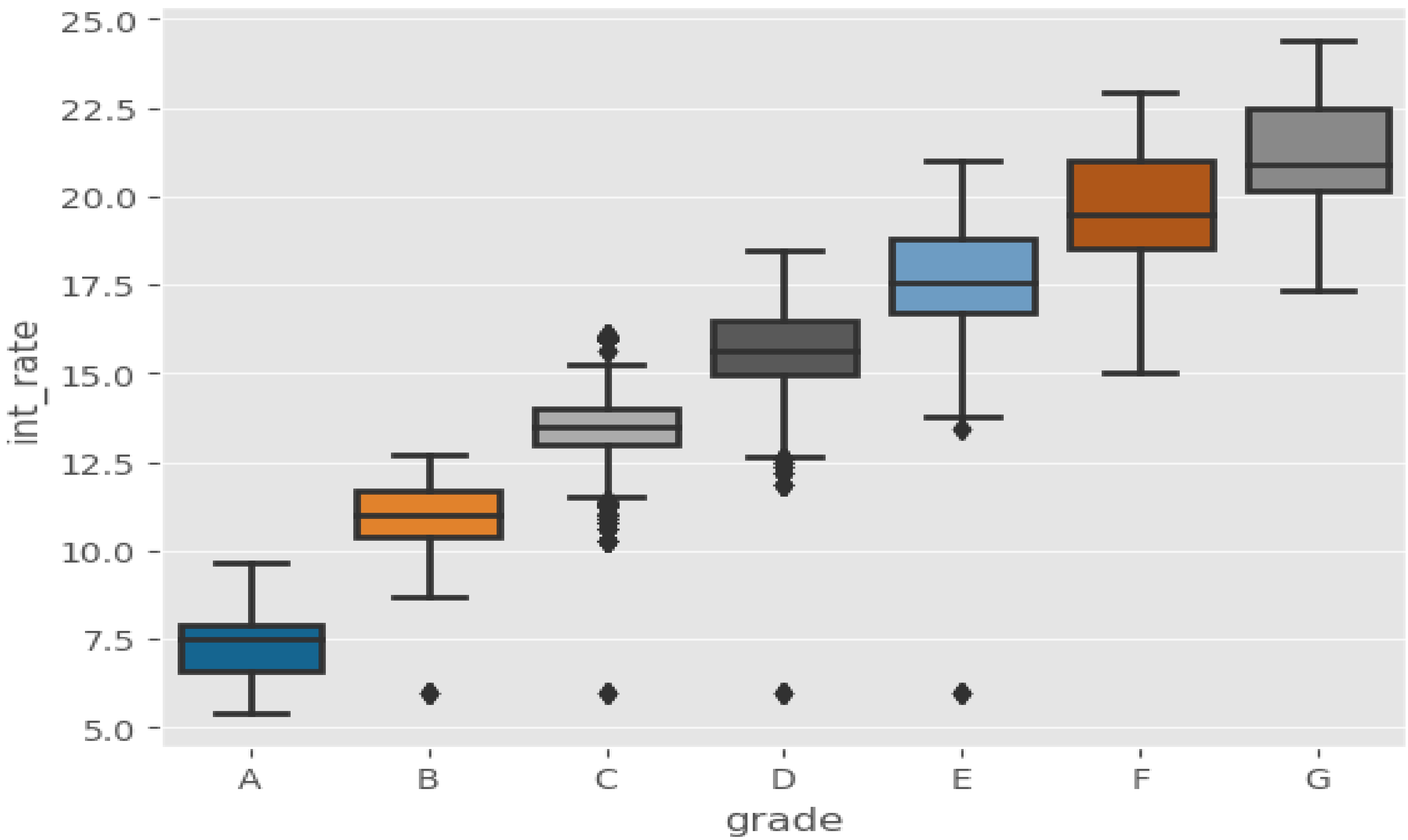
- It is very obvious that increase in the grade category from A to G also drives the charged off ratio from lower to higher.
- When the Grades are categorized from A-G. Its shows the linearity between the Charged off ratio and the Grades.
- Grades are highly correlated with charged off ratio.



Bivariate Analysis

We have seen in the previous slide charged off ratio is higher for higher grades. Here its obvious that interest rate is varying in accordance with the LC grades.

- Hence the grade is directly proportional to the interest.
- When the grade increases, the interest rate is also increasing with respect it.
- Thus, higher the interest rate higher the charged off ratio.



Conclusion

- The interest rate is one of the highly impacting factor for getting charged off.
- The borrower’s employment length (i.e. their experience) which is the second most driving variable towards being end up as a defaulter.
- House ownership is also playing an important role in driving the charged off ratio.
- The borrower whose annual income is higher are lesser becoming defaulter.