

Lending Club Case Study

Group Members:

- Rajender Kanchu
- Swapnil Raut

Problem Statement

- About Company

Lending Club company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

- Background

Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

- Requirement

Need to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicant using EDA is the aim of this case study.

Approach for Analysis

- Clean data
 - i. Drop columns with null, single values, non relevant values like URL.
 - ii. Convert values to numeric/integer values. (Removal of % , >, < characters)
 - iii. Remove columns where data is missing for more than 50 % rows.

```
#Remove columns that aren't relevant
```

```
not_required_columns = ["id","member_id","url","zip_code"]  
loan.drop(labels = not_required_columns, axis =1, inplace=True)  
print("So now we are left with",loan.shape ,"rows & columns.")
```

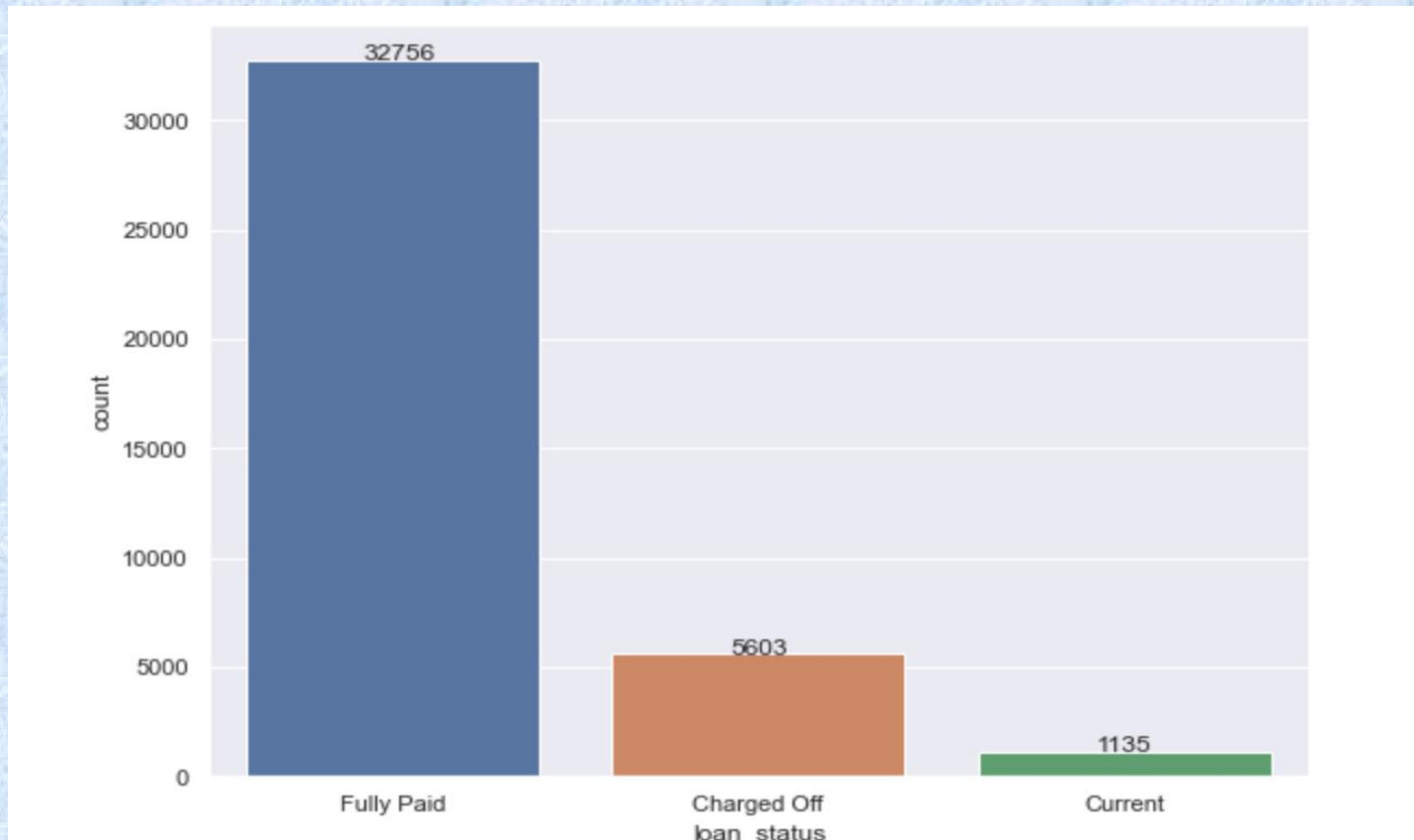
```
So now we are left with (39394, 40) rows & columns.
```

```
loan.describe()
```

	loan_amnt	annual_inc
count	39394.000000	39394.000000
mean	11178.877494	66378.442227
std	7419.880033	37410.922236
min	500.000000	4000.000000
25%	5500.000000	40002.000000
50%	10000.000000	58758.000000
75%	15000.000000	81996.000000
max	35000.000000	294000.000000

Univariate Analysis

- i Check occurrence of various categorical and numerical values.
- ii Create derived variable.

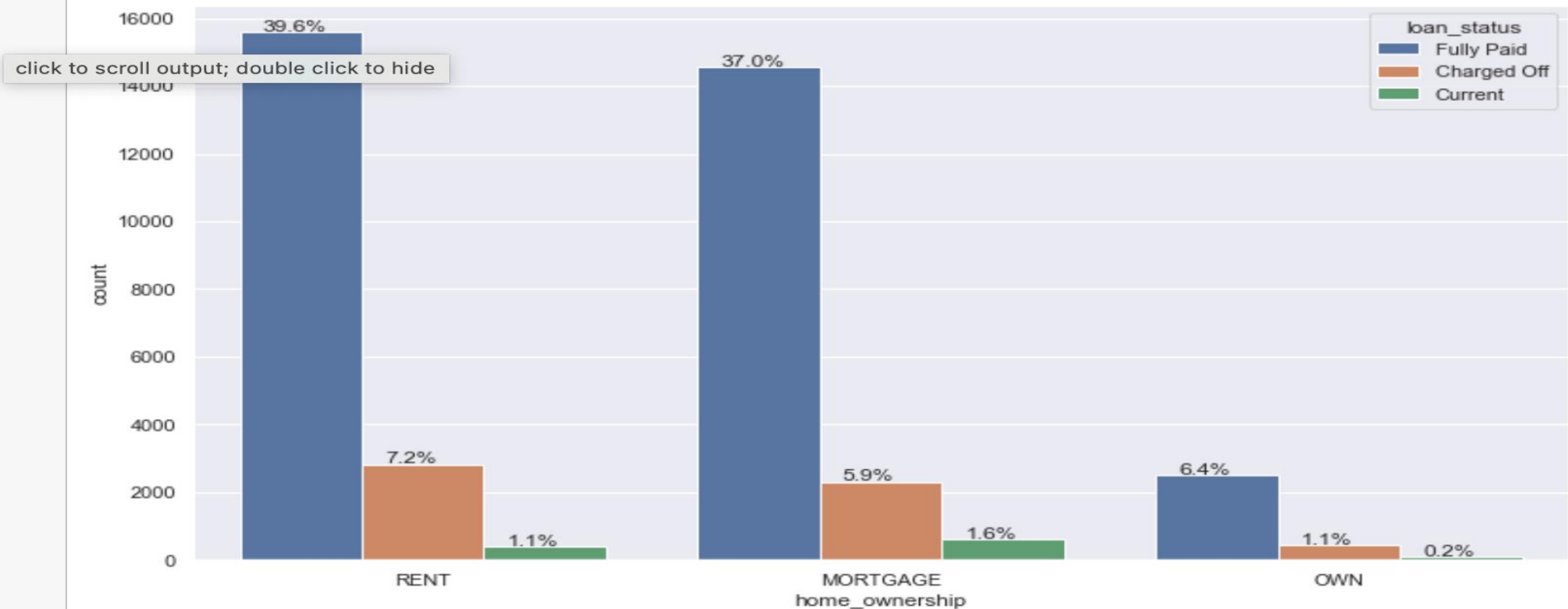


i

Segmented Univariate Analysis

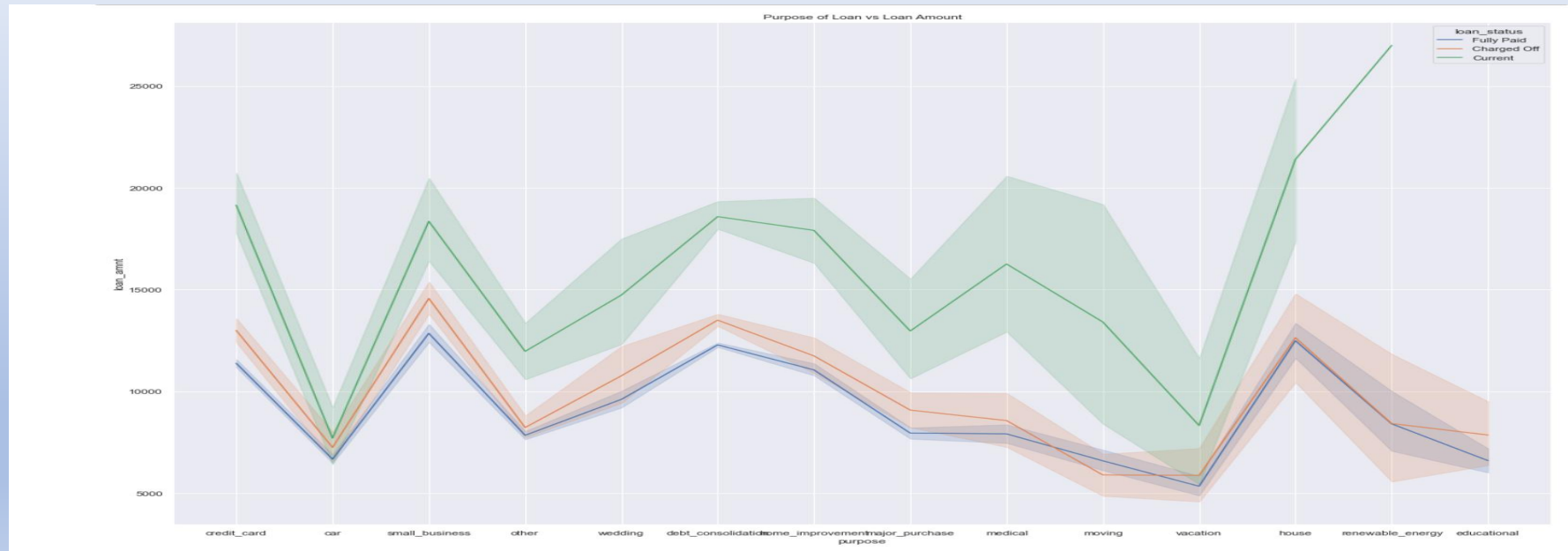
Analysis of variable against other segmented variables.
Create derived variable.

```
In [33]: univariate(df=loan,col='home_ownership',vartype=1,hue='loan_status')
```



Approach for Analysis continue ...

- Bivariate analysis
 - i. Correlation analysis to check two variable impact.
 - ii. Analysis of joint distribution.



Analysis Loan Allocation status

- Summary

Visualizations and summaries the finding from detailed analysis.

Provide some recommendation to reduce credit loss and improve profit.

So, here we can see this all different aspect to consider Target Variable is Loan Status And top 5 major variables to consider for loan prediction: Purpose of Loan, Employment Length, Grade, Interest Rate, Term. Now we are ready to Train our model and prediction