

Lending Club Case Study

Group Members

Prashant Dahiya

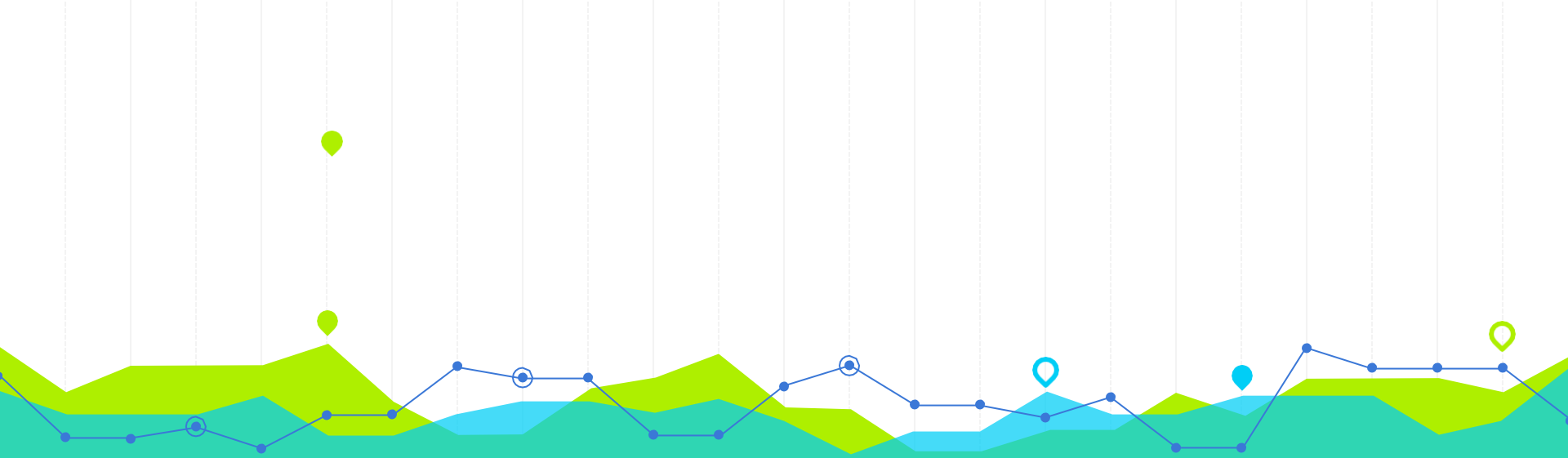
(prashantdahiya77@gmail.com)

Yatharth Tripathi

(yatharthtripathi@gmail.com)







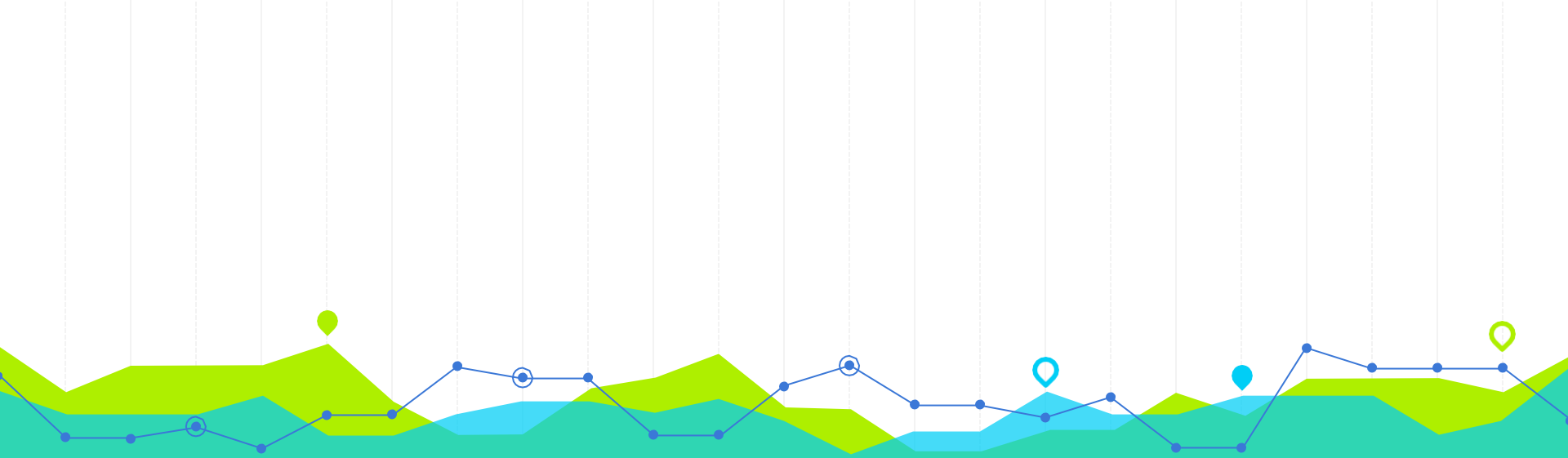
Imported Python Libraries and Case Study

1

1. Numpy
2. Pandas
3. Matplotlib
4. Seaborn

Lending Club Case Study

- Lending Club provides loans to borrowers, but a lot of borrowers' default, that is they do not return the borrowed amount along with accumulated interest back to the lenders.
- Such defaults results in monetary loss to the businesses or lenders.
- Hence, the purpose of this case study is to identify some of the prominent factors which can help understanding borrowers who have higher probability of defaulting.
- In this project, some of the factors have been analyzed and identified from the data and have been reported.



Datasets Load in Dataframe

2

These are the two datasets we imported

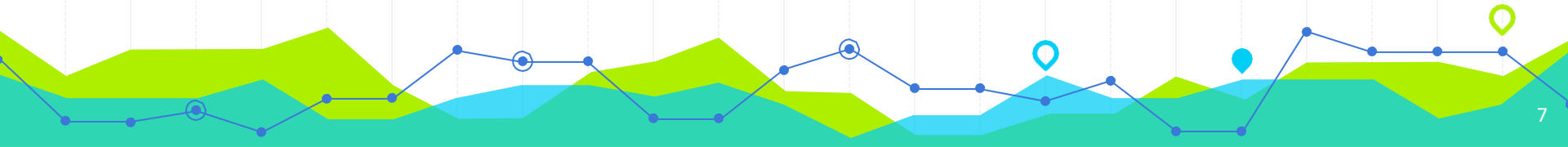
Assumptions: Dataset present in the same directory of code

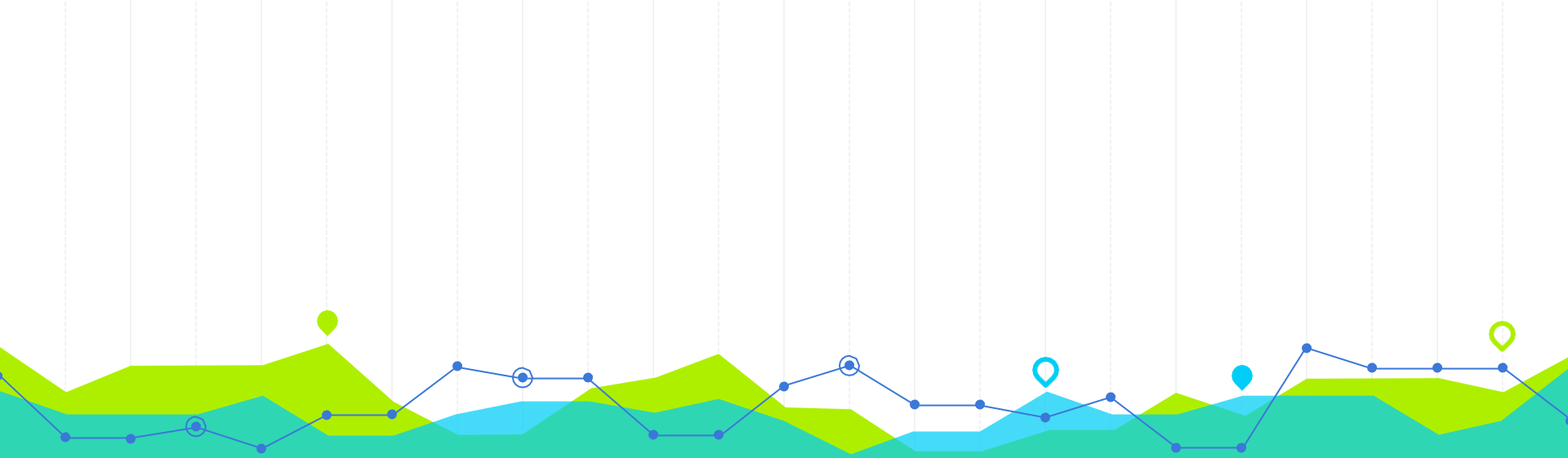
- Loan Data

```
df_loan = pd.read_csv("loan.csv", low_memory=False))
```

- Column Definitions

```
df_columns = pd.read_excel("Data_Dictionary.xlsx")
```





Cleaning the data

3

Data Cleaning

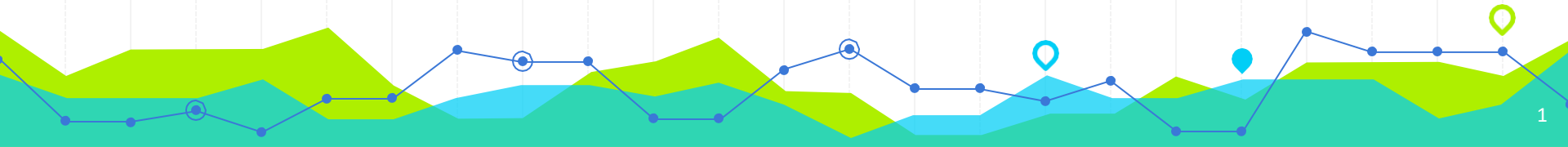
- Removed all the columns with Null values
- Removed all the rows with majority Null values
- Removed Columns which are not defined in the Column Definition dataset

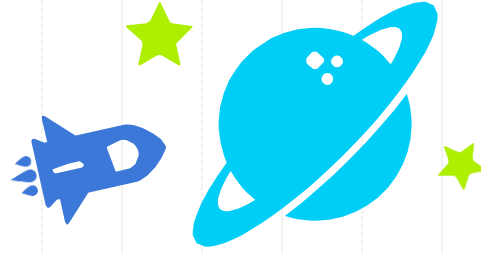


Processing the DataFrame 4

Note:

- 1. First, we understood the columns*
- 2. After that we removed outliers*
- 3. Then we performed Univariate*
- 4. Analysis wherever required*





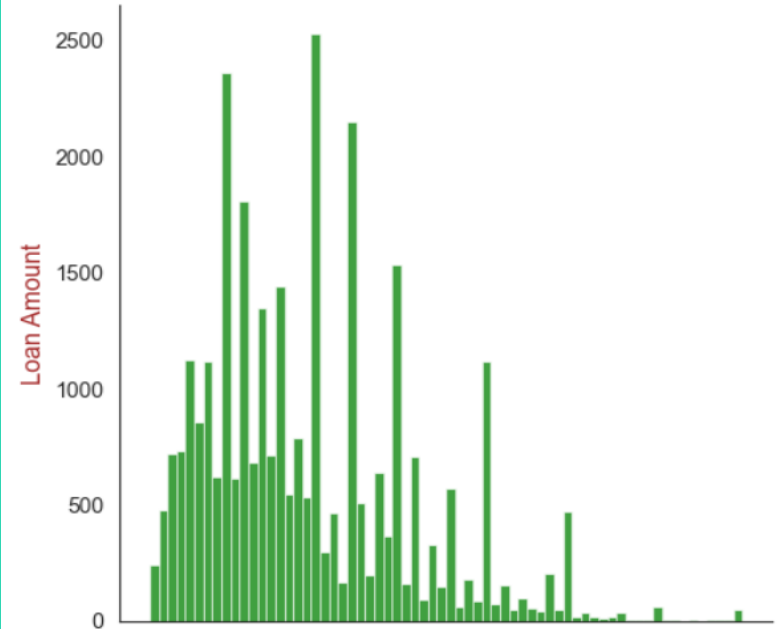
Univariate Analysis



LOAN AMOUNT DISTRIBUTION

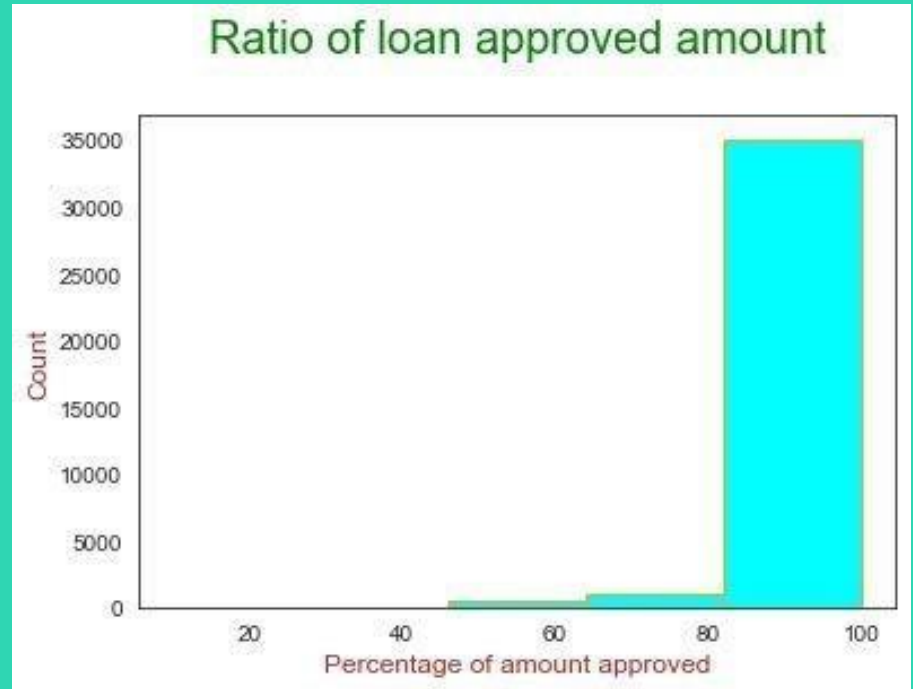
- Loan applied lies between 500 to 35000
- By this graph we got to understand that most of the loan borrowers need money under 15000
- Majority of borrowers asked for rounded number that is 5000, 10,000 as observed from the Distribution graph

Distribution of loan amount



PERCENTAGE OF LOAN APPROVAL

- Majority of the people were funded

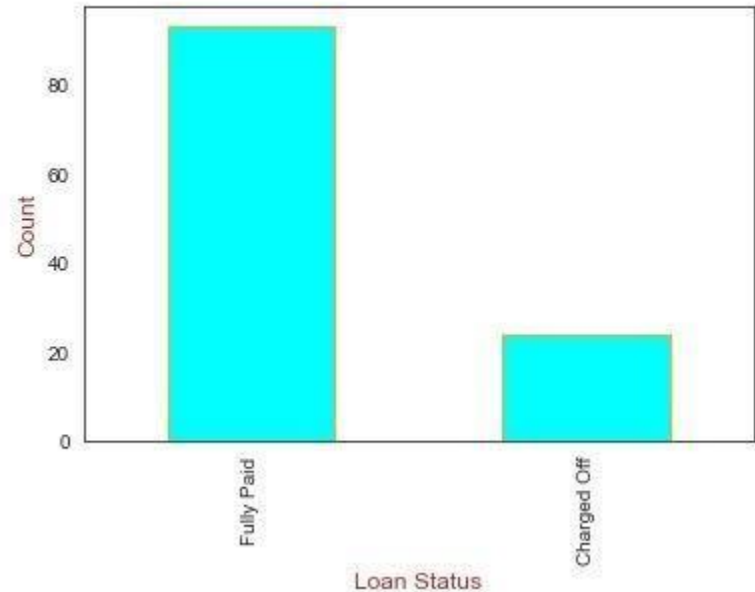


LOAN STATUS

- `funded_amnt_inv` lies between 0 to 35000 but there are 129 records with 0 this might be because investor did want to invest

In the 129 records majority of the people paid off the loan so it is better to provide them loans and reduce the business loss

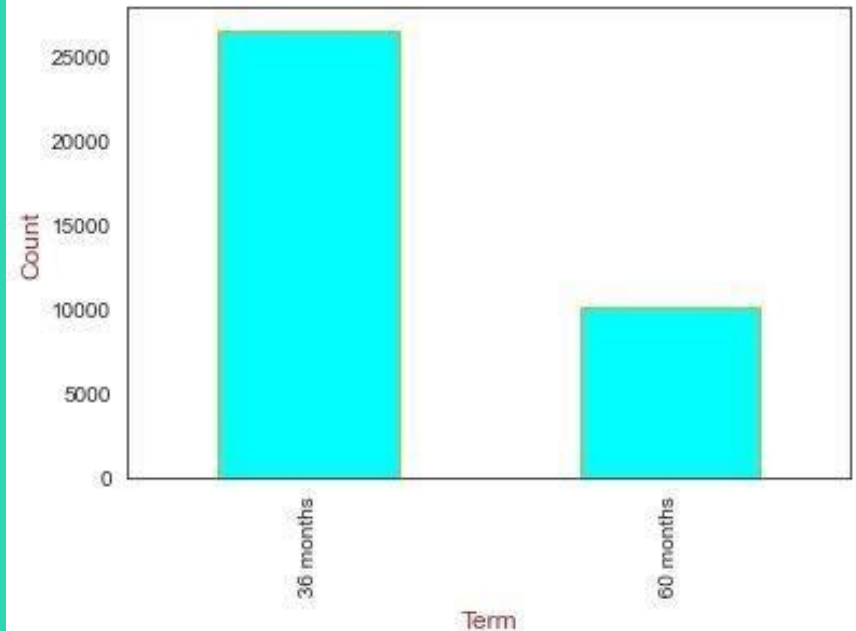
People who were not accepted to fund by investor



LOAN TERM

- + There are two values 36 and 60 months
- + By this we get to understand that majority of the people have taken the 36 months tenure period

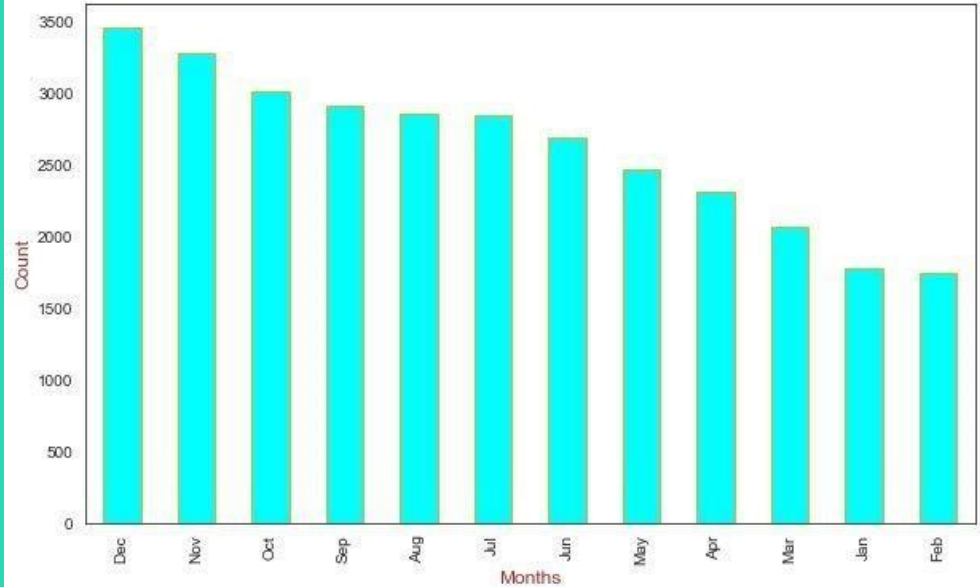
Understanding how people took the terms



LOANS TAKEN IN EACH MONTH

- + Majority of the people have taken loan in the end of the year
- + Need to promote other months also to increase profits

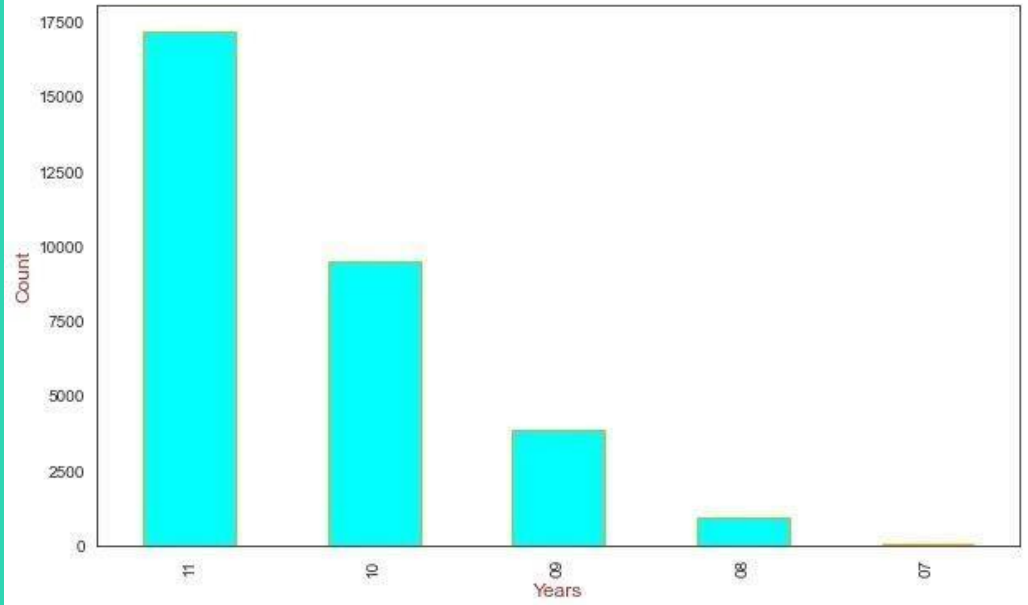
Understanding Distribution of loan taken by Montly



LOANS TAKEN IN EACH YEAR

+ Majority of the people have taken loan in 2011 and after every year, the number of loan borrowers have increased

Understanding Distribution of loan taken by Year

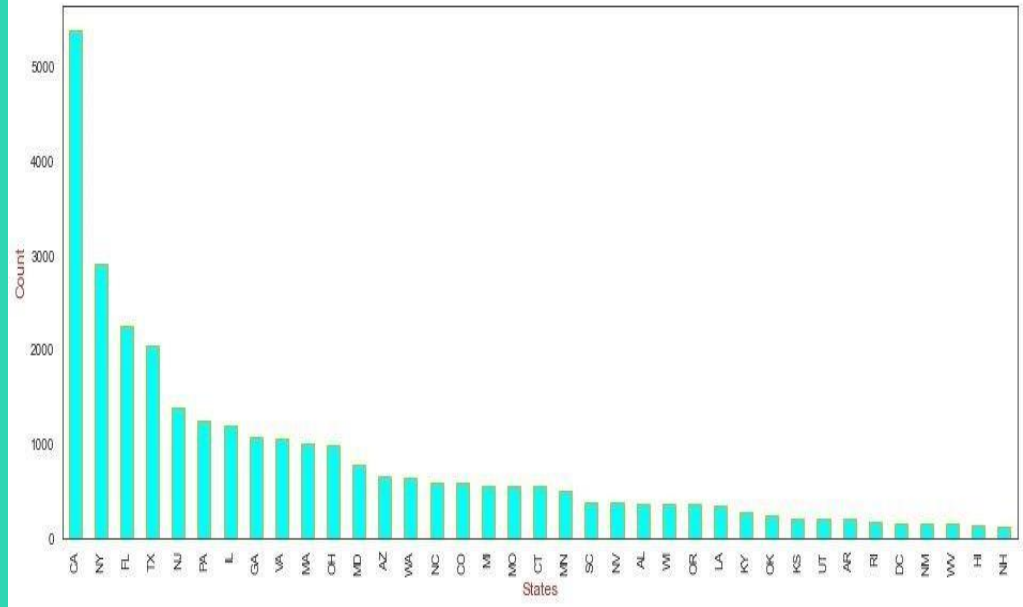


LOANS TAKEN IN EACH STATES

+ California state has highest number of loan count

+ Need to promote loans in other states as well to increase profit

Understanding Distribution of loan taken by States

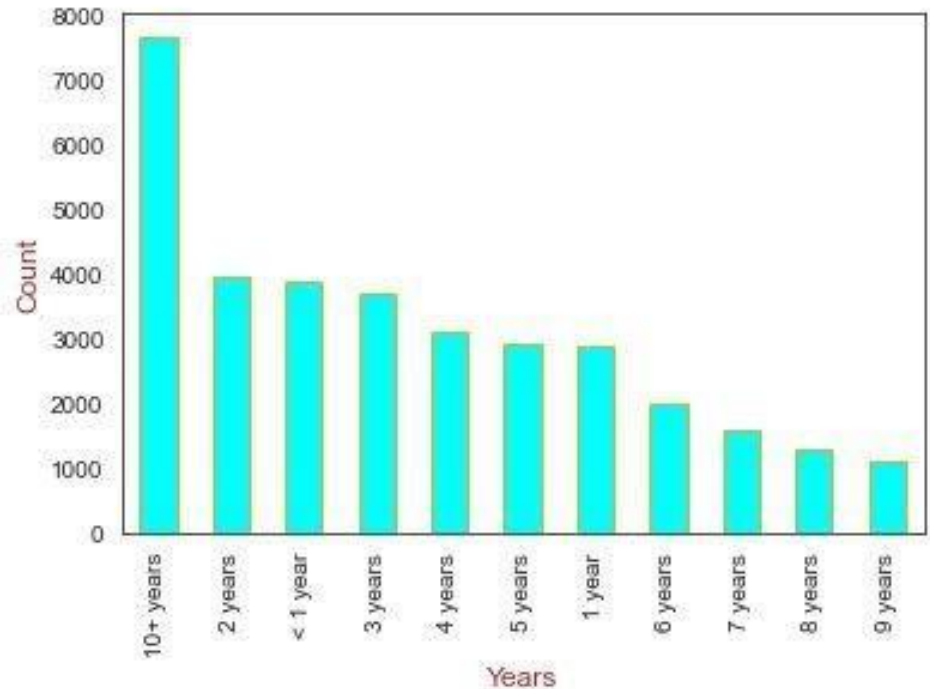


EMPLOYMENT LENGTH AND LOAN COUNT

+ Possible values are between 0 and 10 where 0 denotes less than one year and 10 denotes ten or more years.

+ From the graph we can see that people having 10+ year experience are majority of applicants of the loan

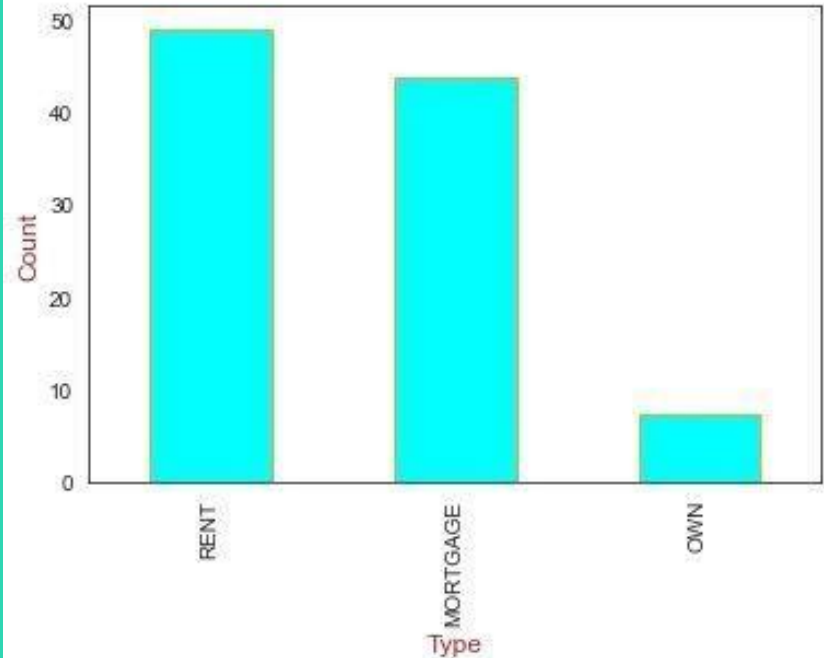
Understanding Employment length Data

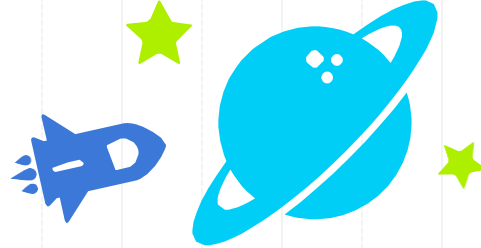


HOME OWNERSHIP DATA WITH LOAN COUNT

- + Majority of the people who applied for the loan are living on rent or mortgage
- + This implies that 80% are not having full home ownership

Understanding Home Ownership Data



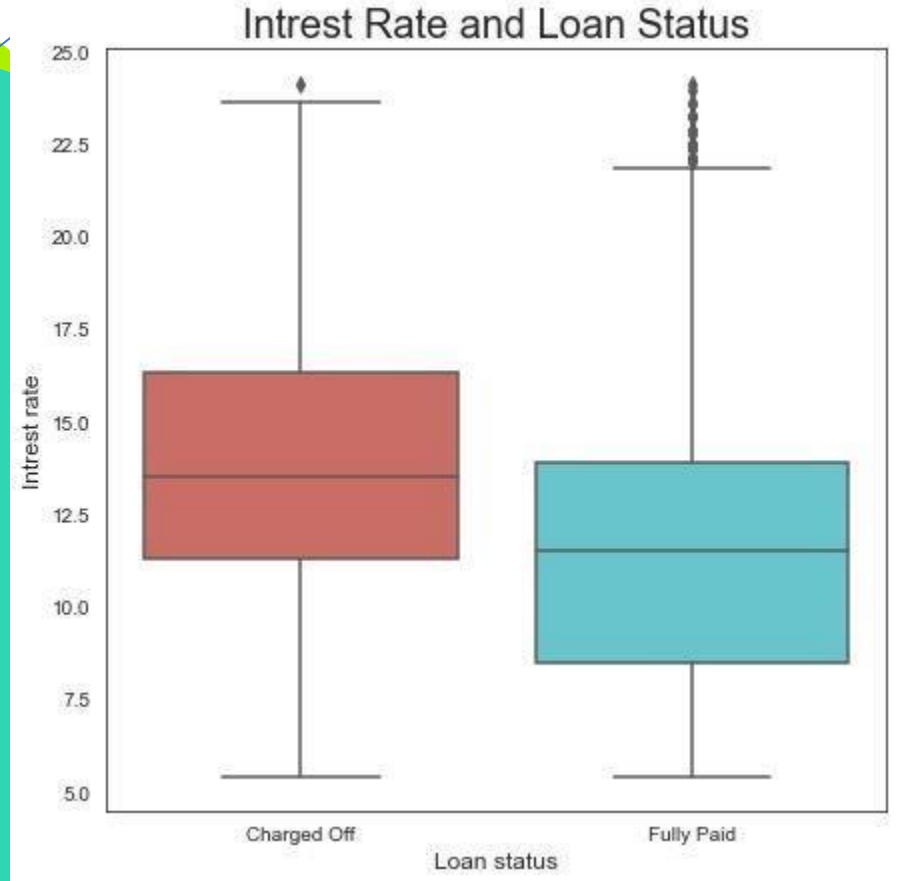


Bivariate Analysis



INTEREST RATE VS LOAN STATUS

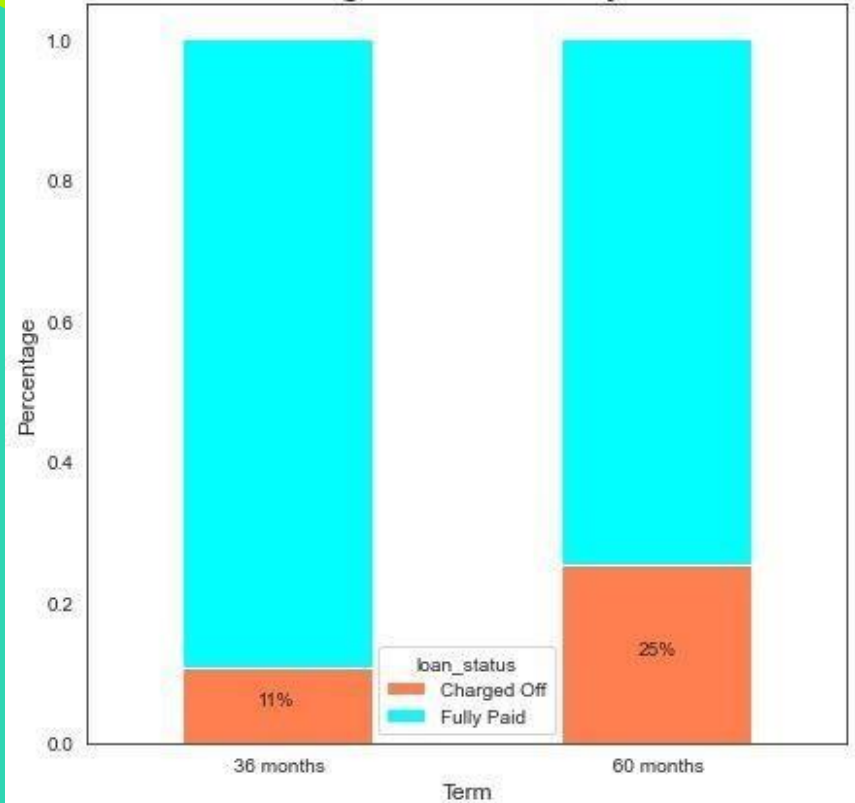
+ The interest rate in charged off were higher than that of fully paid which may infer that higher interest rate could result in higher charge offs



PERCENTAGE VS TERM

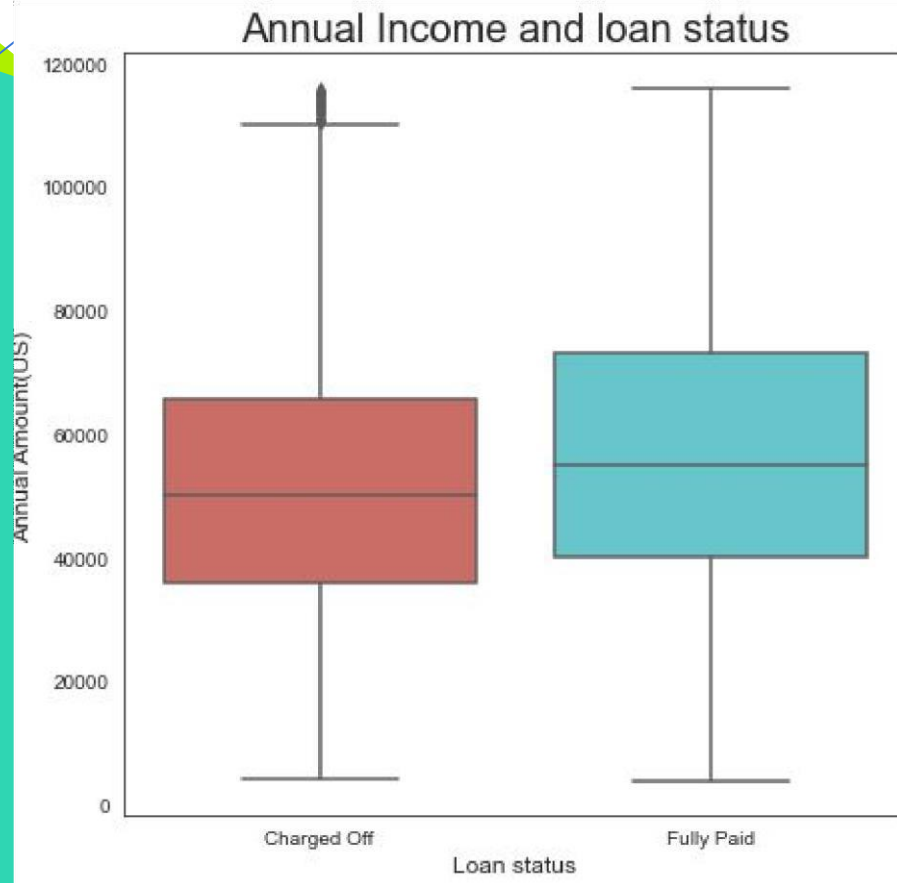
- + The percentage of defaulters are higher in 60 months term with about 25% charge-off whereas it is just 11% for 36 month term
- + So, there is a possibility that longer term loans are more risky and prone to higher default

Checking Loan status by terms



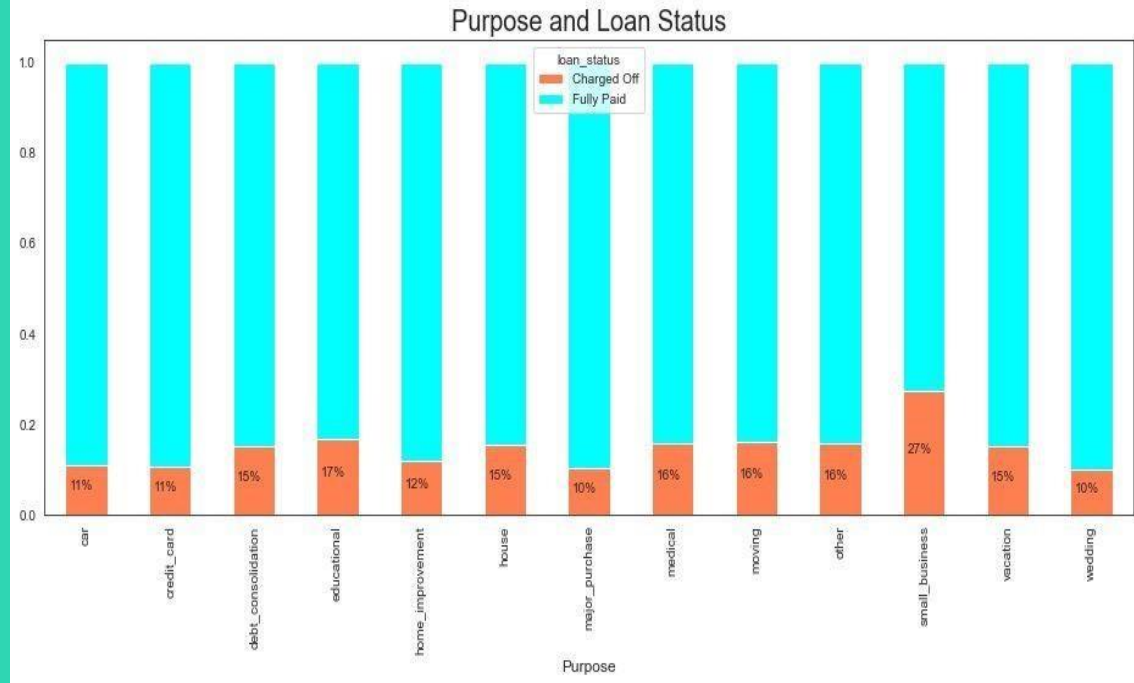
ANNUAL INCOME VS LOAN STATUS

- + The median annual income is less for charged off persons
- + People having less income are more likely to default.



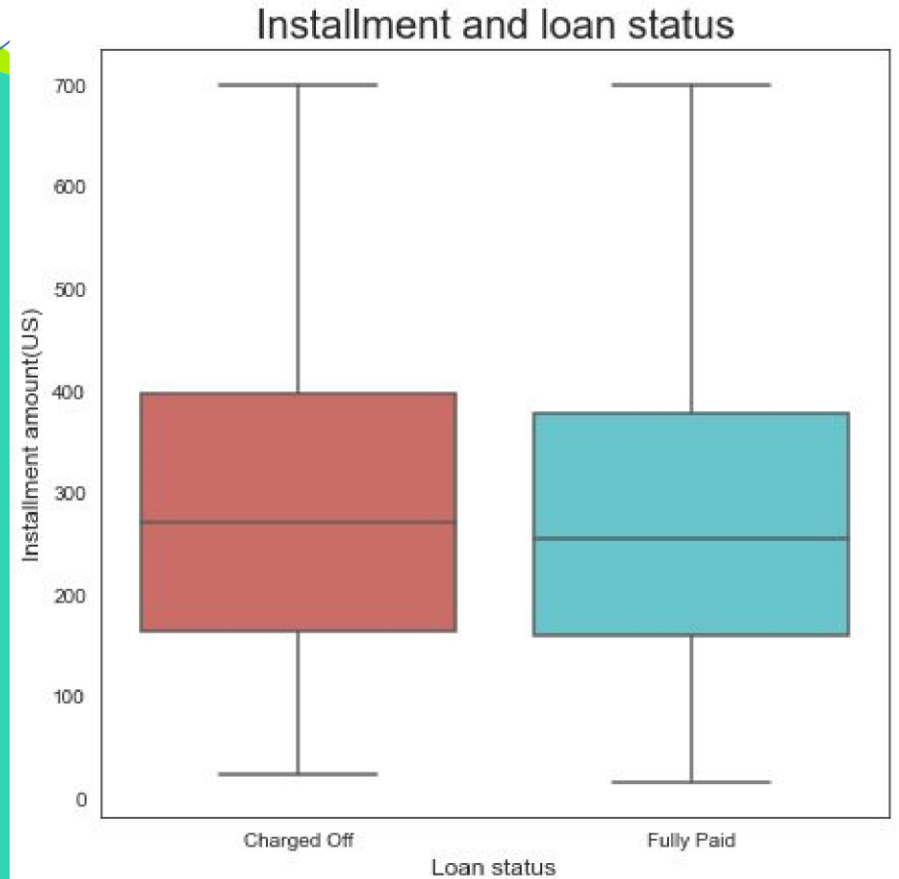
LOAN STATUS VS PURPOSE

+ Small business are more likely to charge off compared to other purpose as they are the 28% of people who charged off



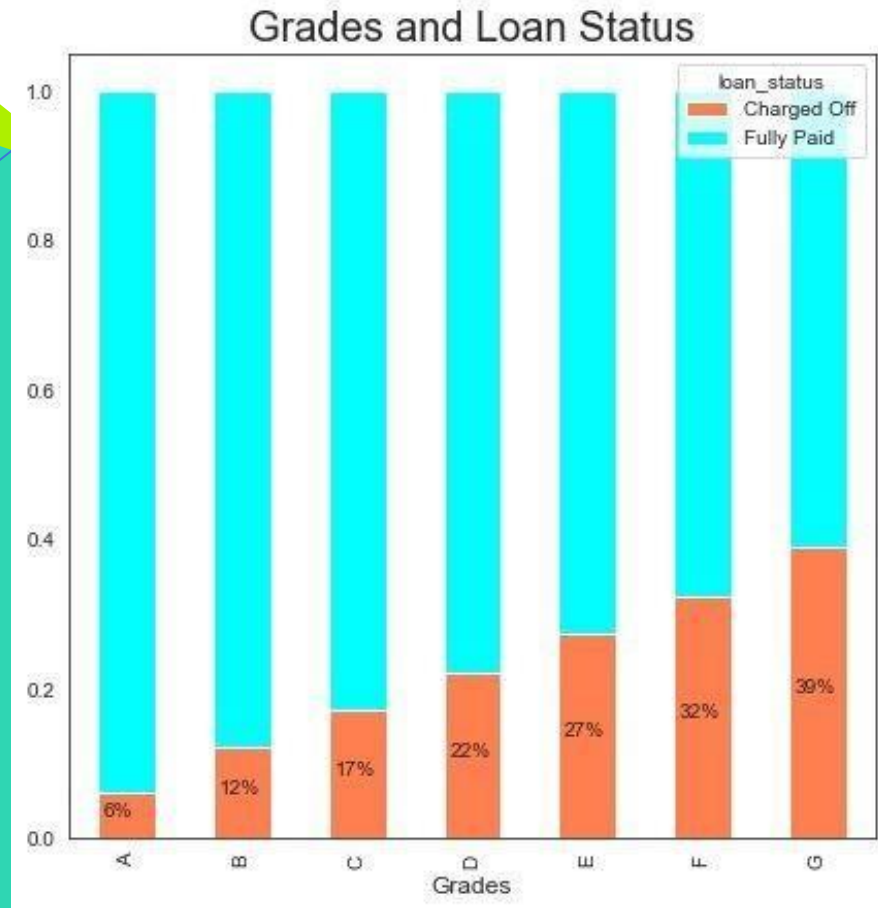
INSTALLMENT AMOUNT VS LOAN STATUS

+ By observing the median and 3rd quartile people who have taken higher installment amount are likely to charge off



LOAN STATUS VS LOAN GRADES

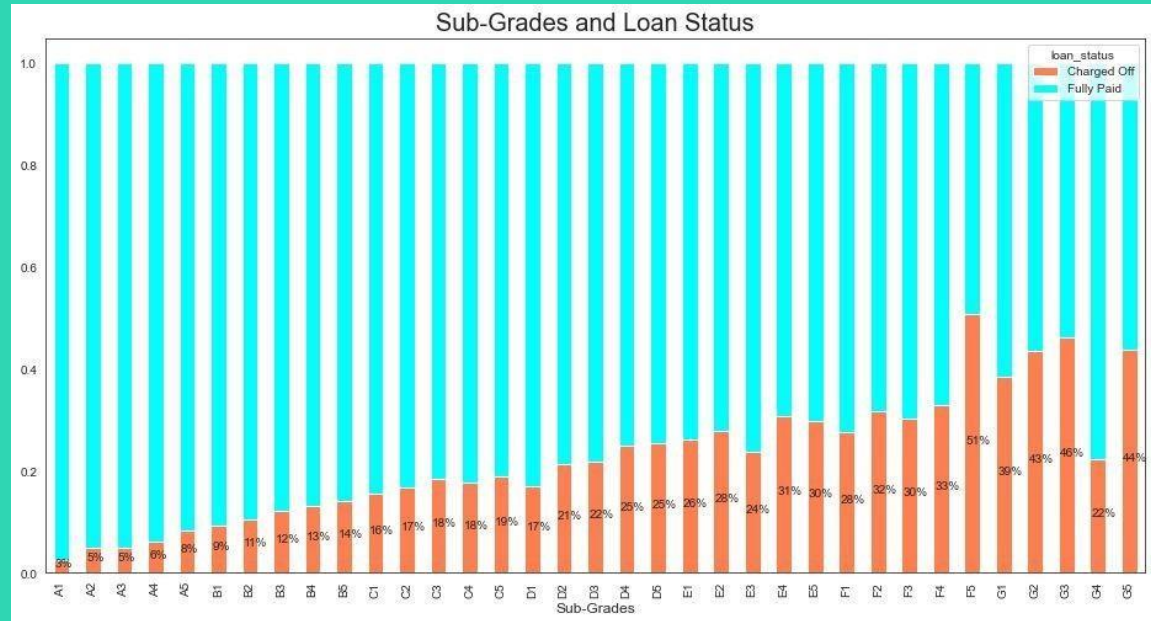
- + Lower grade loans results in higher charge offs.
- + We can see as grades gets lower, charge off rate increases proportionately



LOAN STATUS VS SUB-GRADES OF LOAN

+ We can see as sub-grades gets lower, charge off rate increases proportionately

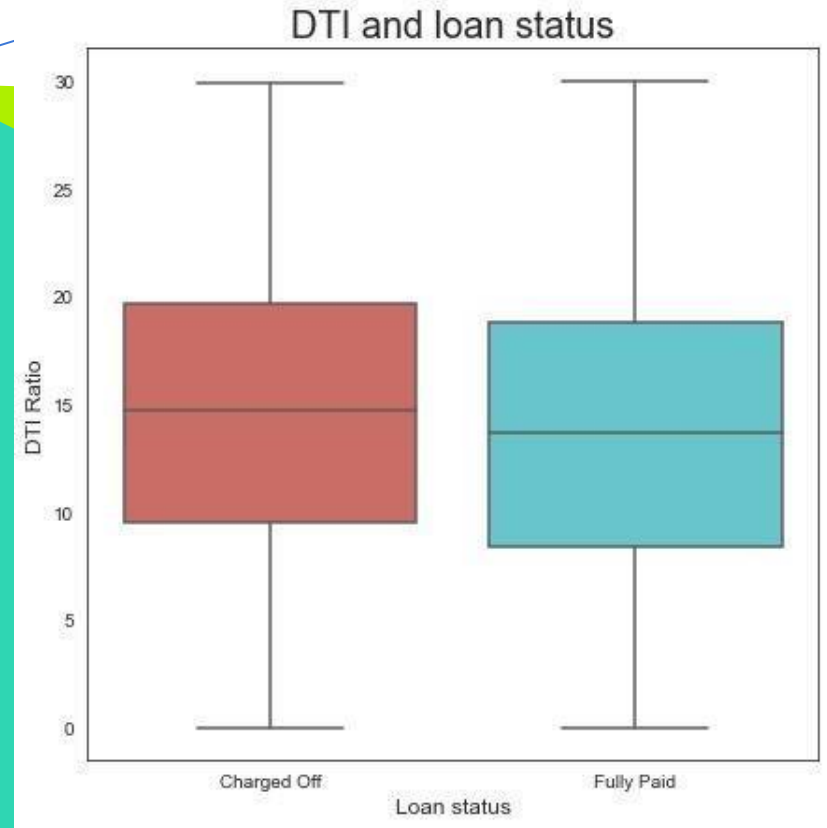
+ A1 grade is one of the best grade to give the loan



DTI VS LOAN STATUS

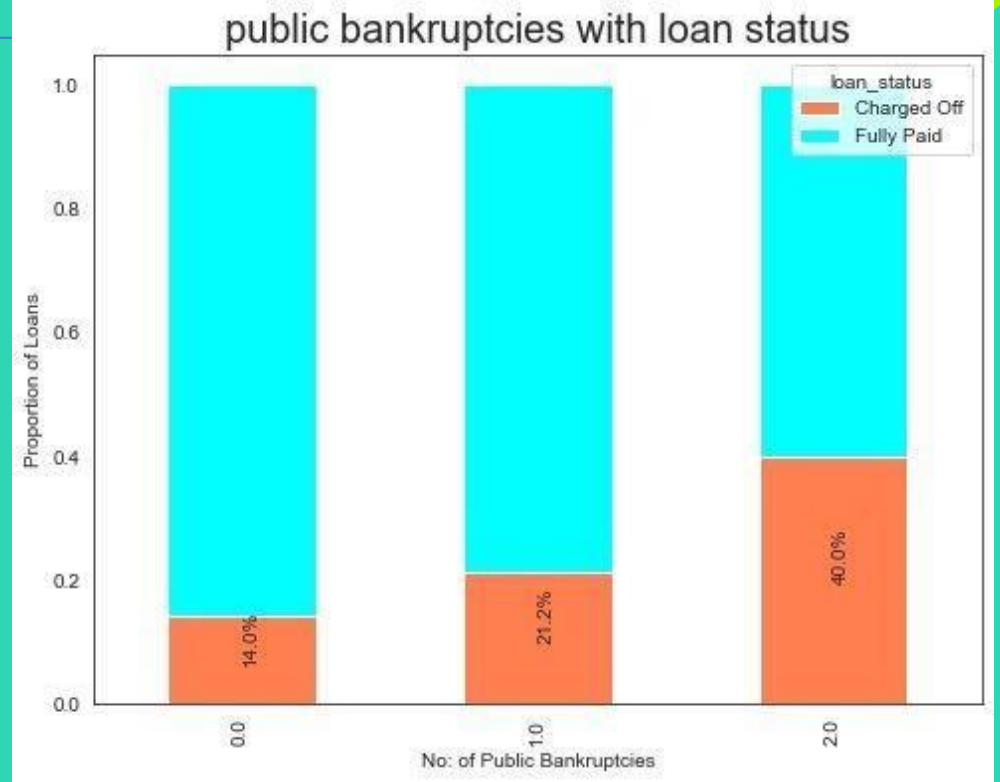
+ DTI is a ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.

+ Higher dti are more likely to be charged off



PERCENTAGE OF LOANS VS NO: OF PUBLIC BANKRUPTCIES

+ As number of bankruptcies increases
the charge off rate also increases



THANKS!

