Lending Club

Case study

Problem Statement

- *As part of Consumer finance company it is required to analyze the exist data to gain insight about possible defaulters.*
- *It is required to understand what are the possible factors affect the customers who have fully repaid the loan acquired.*
- *It is required to identify the factors which might indicate the customer who will default on their loan.*
- *By the analyzing the above factors we need to identify which customer would be good for the business and which customer needs to rejected.(i.e., profitable or incurring loss)*

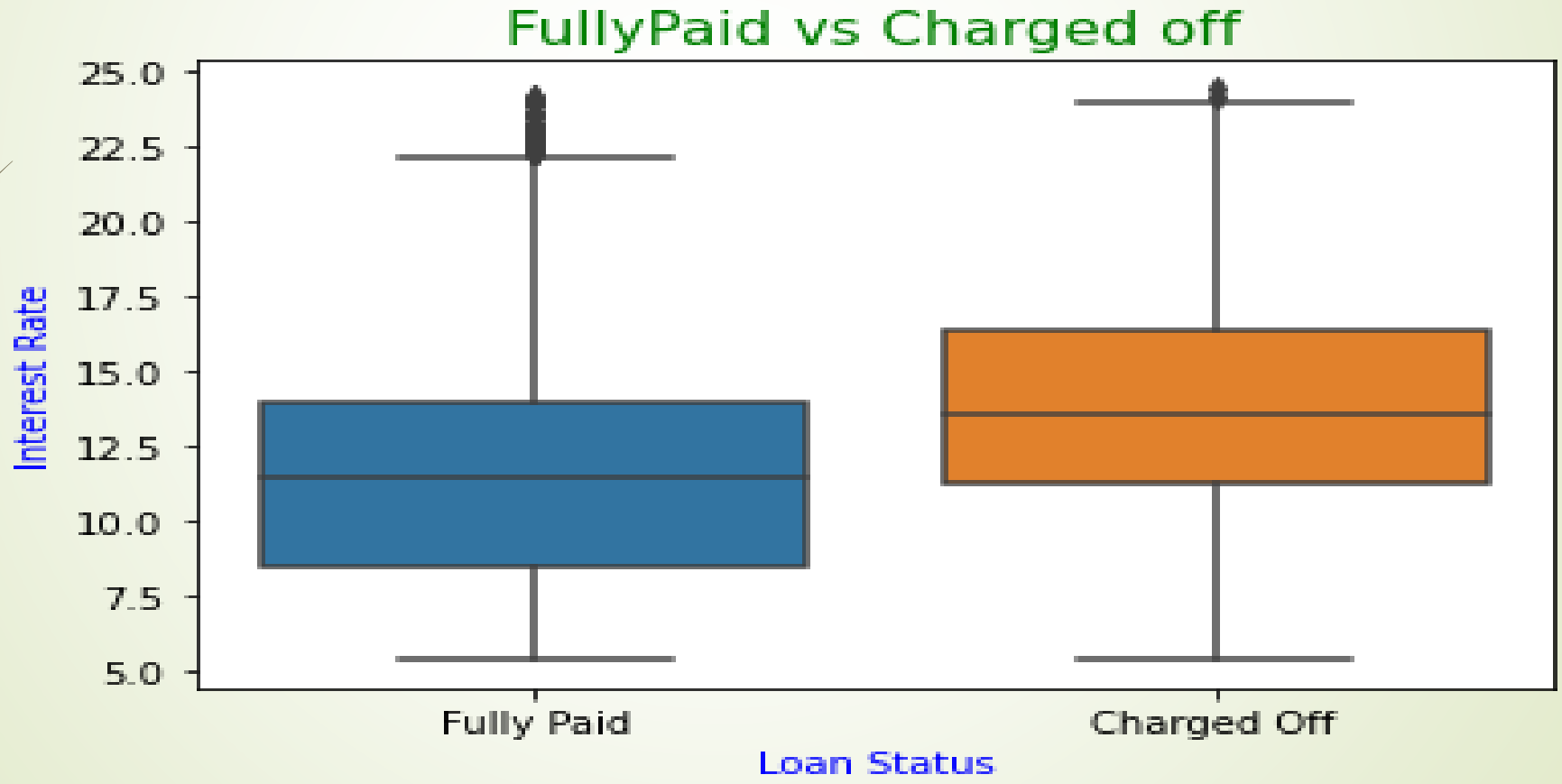
Assumption

- *For the analysis we need to focus on the customer who have fully repaid the loan or charged off. As the customer who in current state might not help to analyze the possibility of defaulting on their loan. In our case assumed to be completely repaying their loans in due course of time.*

Approach

- *We need to clean the dataset loan.csv to identify the columns which have 100% missing values.*
- *Columns having more than 30% of missing value were rejected.*
- *The int_rate (interest rate) column was cleaned to get rid of '%' .*
- *The column emp_length (Employment length in years) has a '+' which causes an issue while plot the graphs so it was stripped from the series.*
- *Column with unnecessary data were dropped(URL, Zip code etc)*

EDA Graphs and Analysis
Interest Rate



Interest Rate - Fully Paid

- *It was observed that the customers who fully repaid their loan has the interest rate between :*
- *count 32950.000000*
- *mean 11.609761*
- *std 3.599964*
- *min 5.420000*
- *25% 8.490000*
- *50% 11.490000*
- *75% 13.990000*
- *max 24.110000*
- *Name: int_rate, dtype: float64*

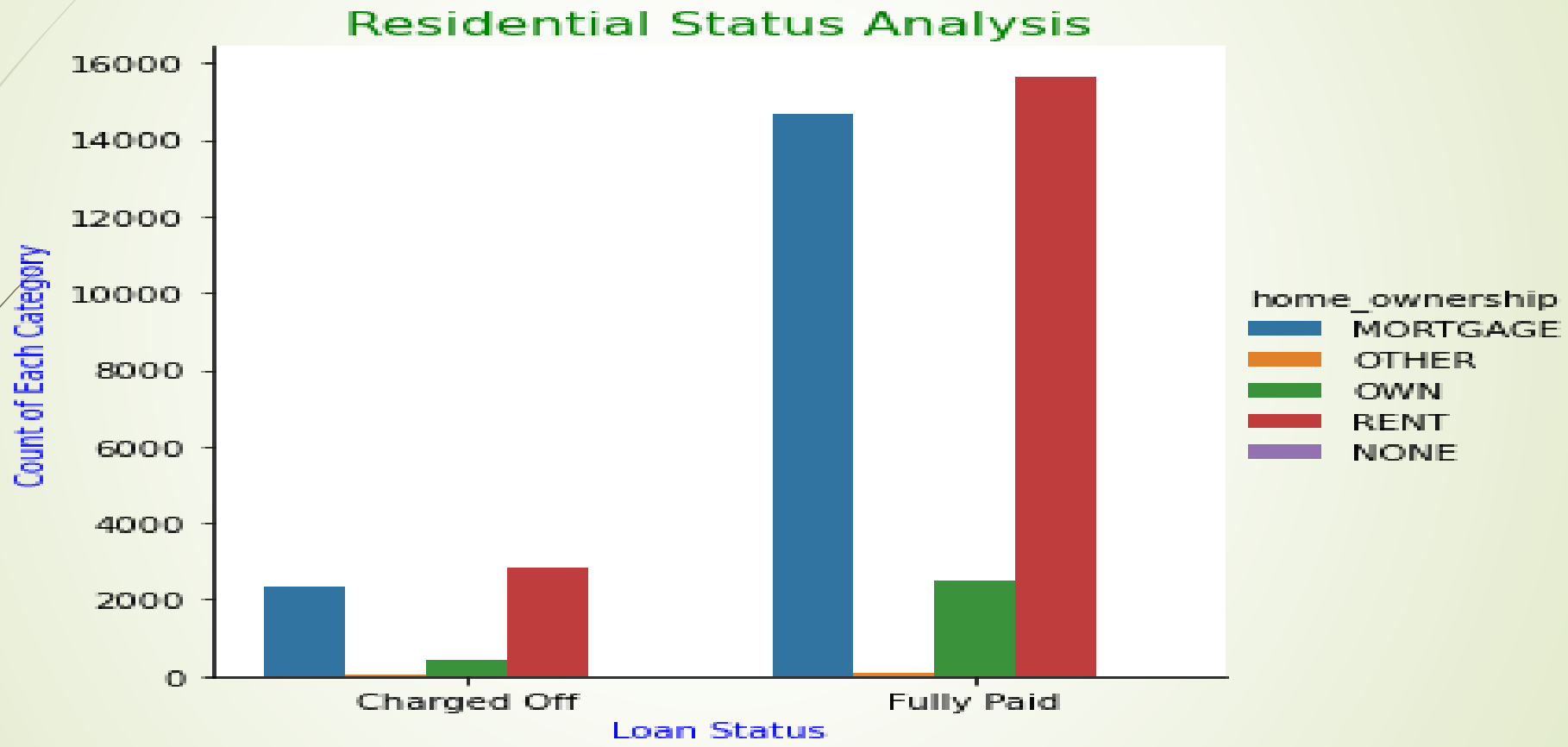
Interest Rate - Charged Off

- It was observed that the customers who fully repaid their loan has the interest rate between :
- count 5627.000000
- mean 13.820432
- std 3.654413
- min 5.420000
- 25% 11.310000
- 50% 13.610000
- 75% 16.400000
- max 24.400000
- Name: int_rate, dtype: float64

Fully Paid vs Charged Off(Interest Rate)

- *The customer who repaid the loan has considerably lesser rate of interest than those who had defaulted on their loan repayment.*

Residential Status

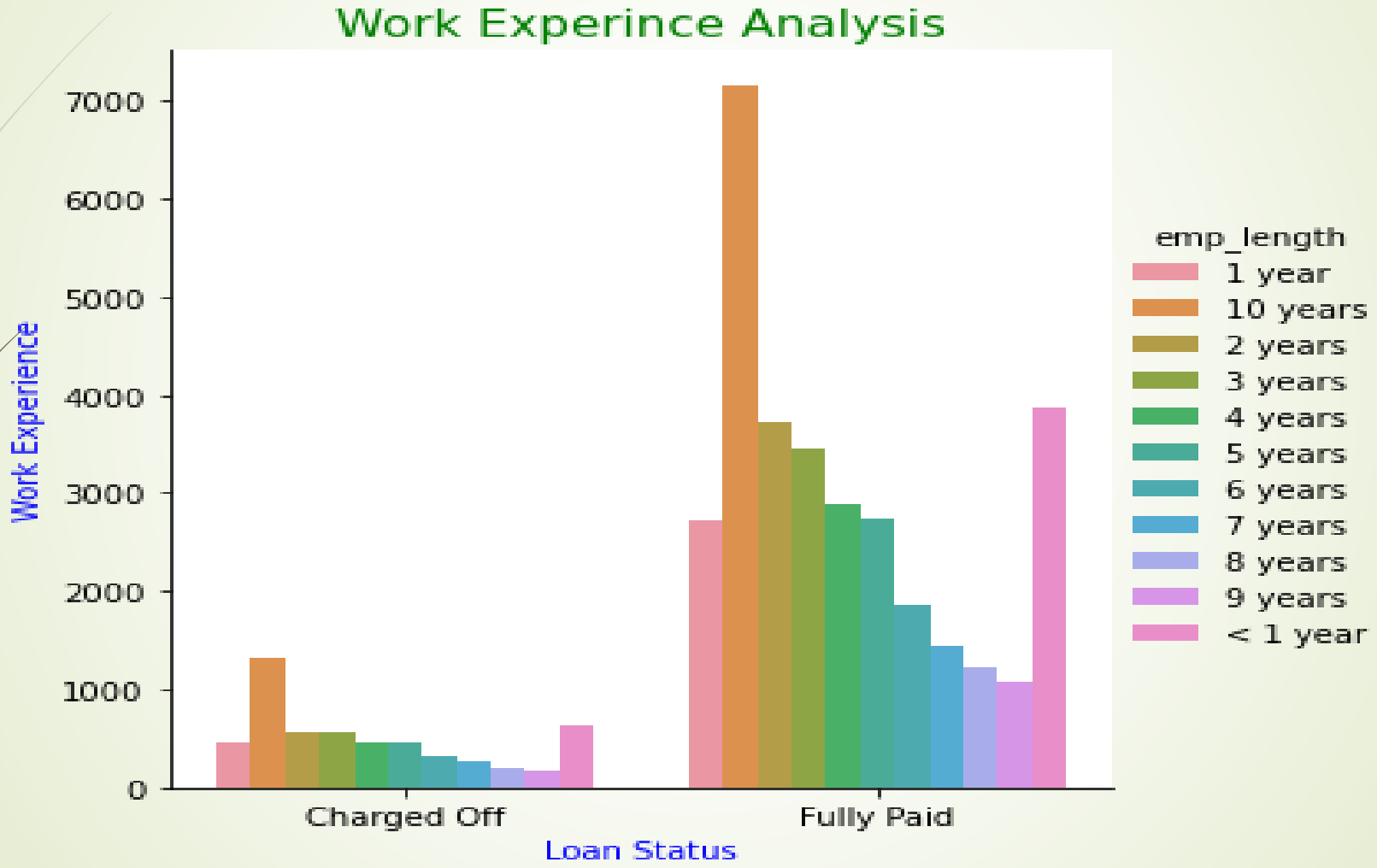


Residential Status – Fully paid vs Charged Off

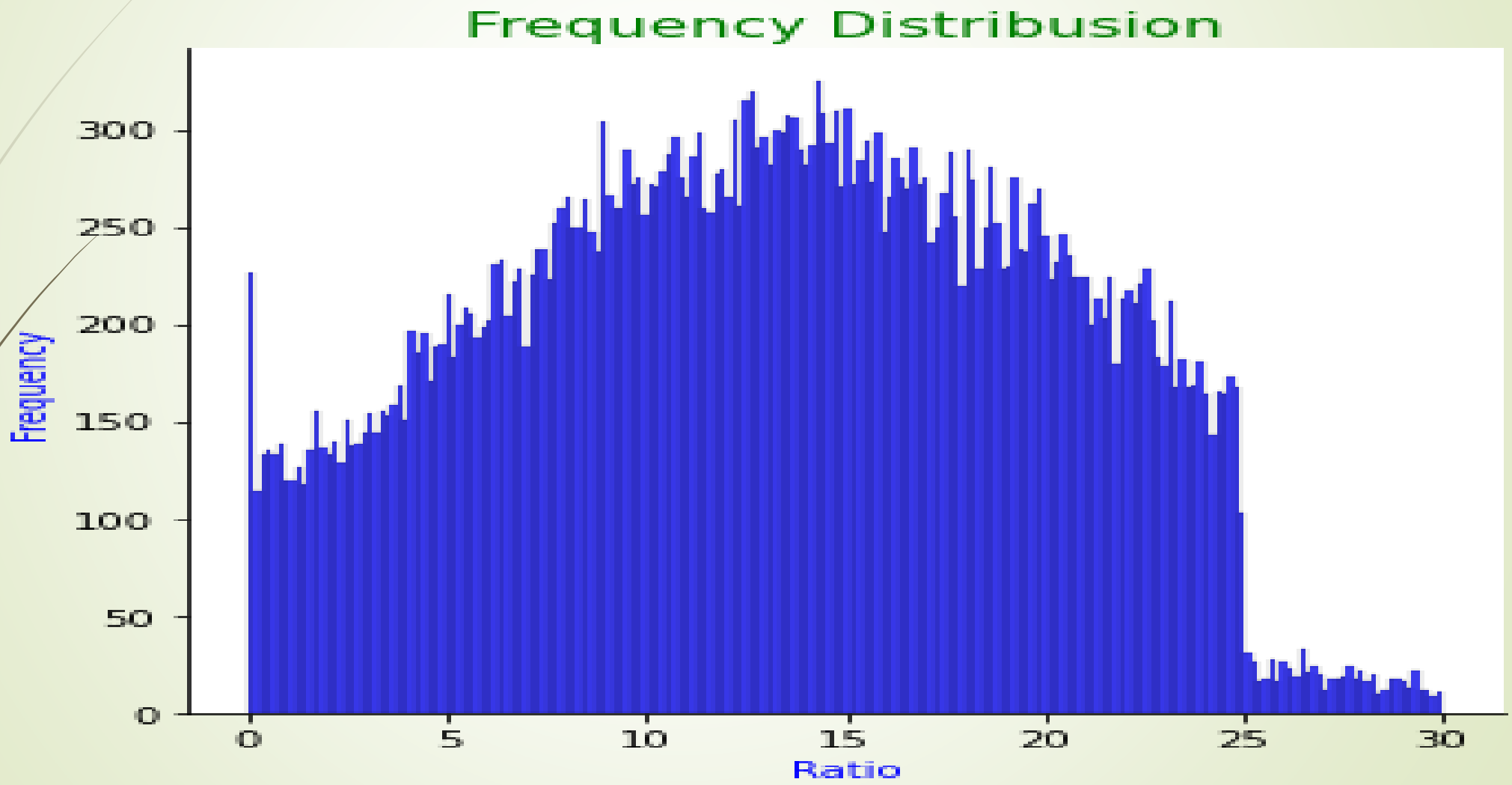
- *41% of Charged off customers had Mortgage on their house. Compared to 44.6% customers who repaid the loan.*
- *50% of Customers who defaulted lived in rented house. Compared to 47% of customers who fully repaid.*
- *7.8% of customers who owned a house defaulted in comparison to 7.7% who repaid.*

- *Above data indicated that customers who had been living in a rented house were prone to default on their loans when compared other 2 categories.*

Work Experience(Employment Duration)



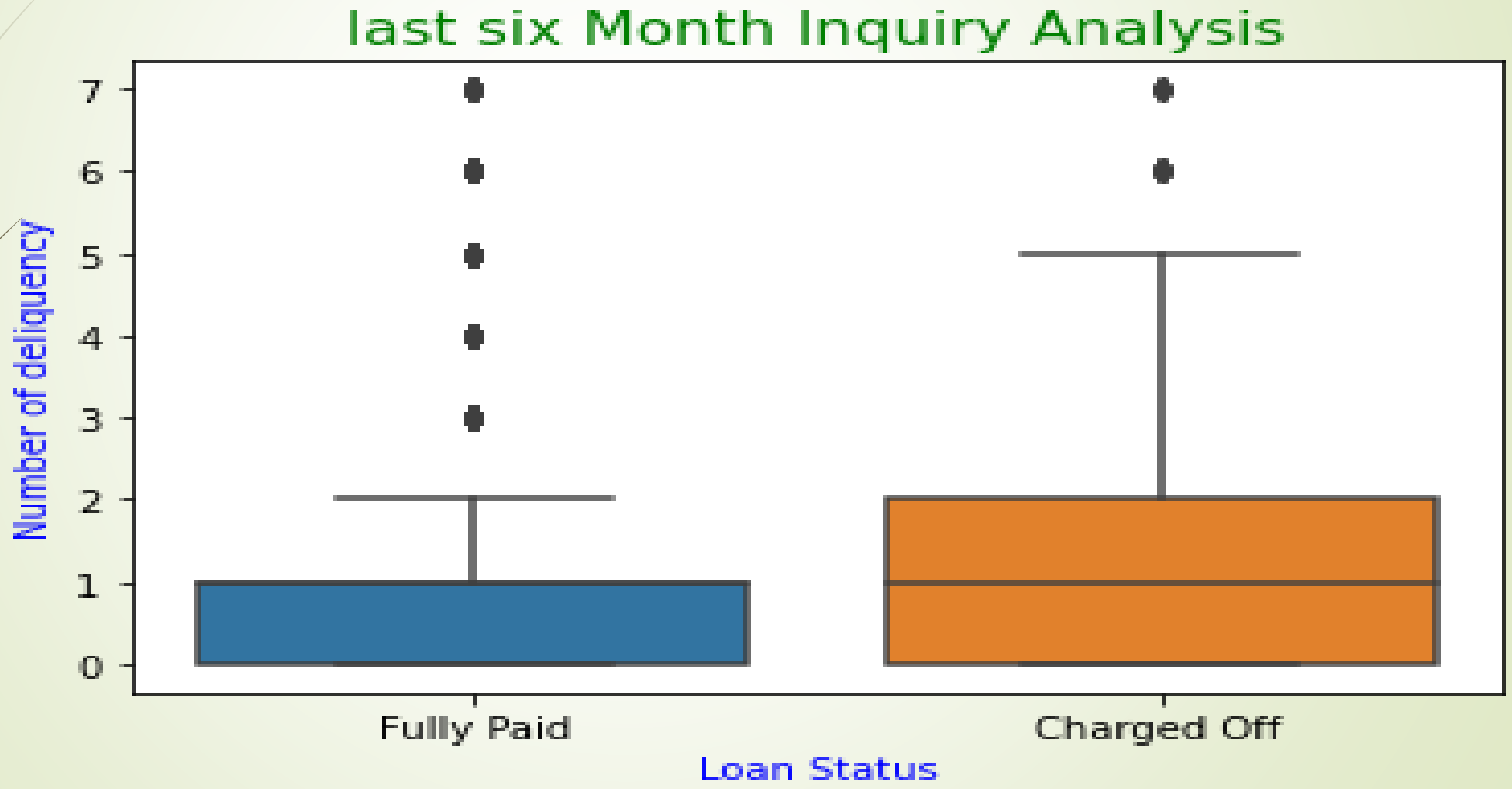
dti


Frequency Distribusion
Frequency
300
250
200
150
100
50
0
0
5
10
15
20
25
30
Ratio

dti - fully paid vs charged off

- *The dti was observed to be having maximum density around 15 when displayed on distribution plot*

Inquiry since last six months



Inquiry since last six months –fully paid vs charged off

- *It is observed that customer who had frequent inquiry is last since six months were more prone to default on their loans.*

Conclusion

- *After analyzing the data wrt to fully paid customer and charged off defaulters the following inferences were drawn.*
- *Interest rate played a key role when the customer defaulted that is the loans with higher rate of interest were observed to be defaulted.*
- *The residential status was a factor to consider while the customers defaulted like, the customers with rented house were more prone to be charged off.*
- *Inquiry since last six month was the major factor which has to be considered as it was observed that the defaulter had almost the double amount of inquiries compared to fully paid customers.*