LENDING CLUB CASE STUDY

Jane Saldanha

Anant

AGENDA

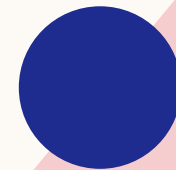# INTRODUCTION

You work for a consumer finance company which specializes in lending various types of loans to customers. When the company receives a loan application , the company has to make a decision for loan approval based on applicant's profile. Two types of risks are associated with bank's decision:
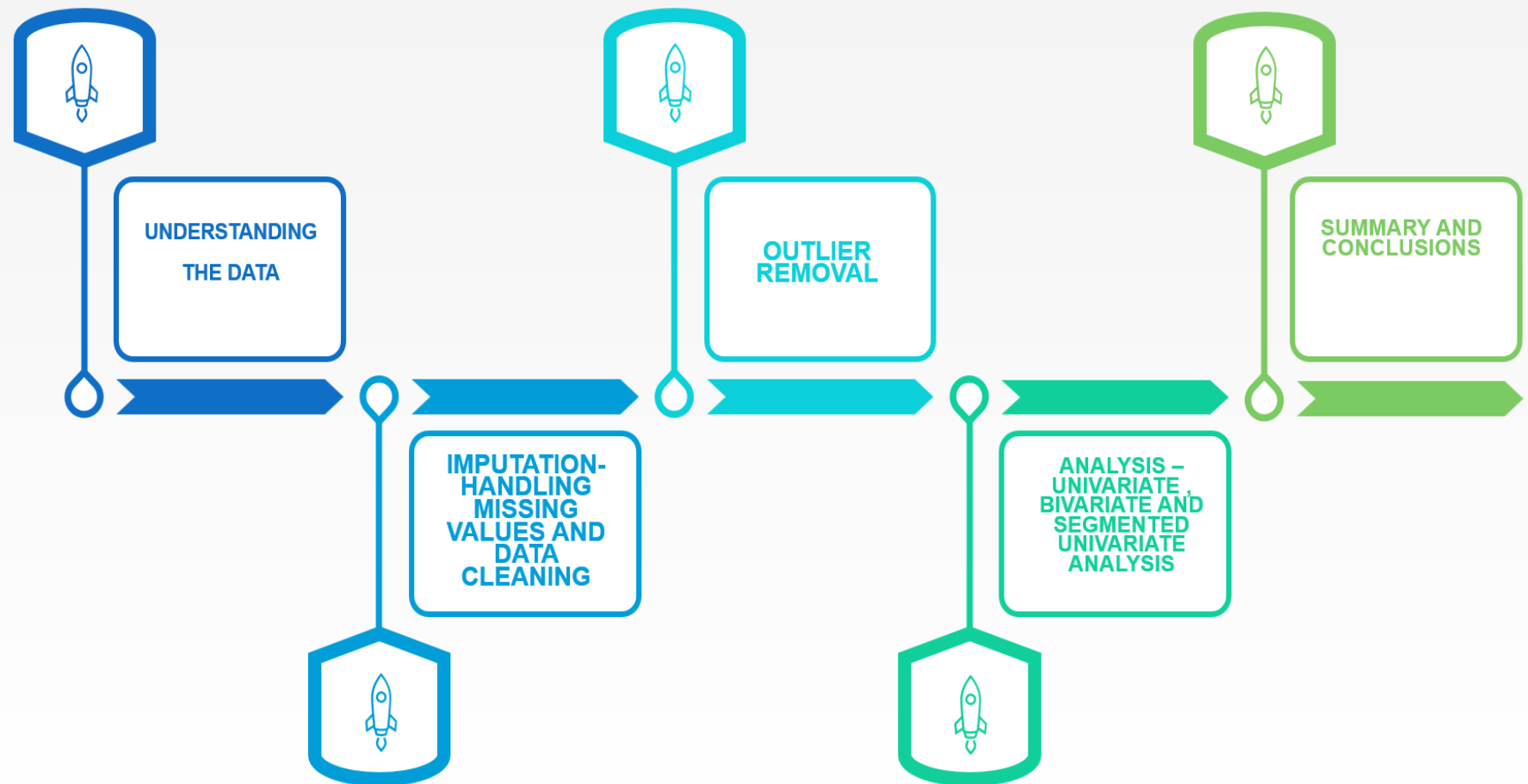
1. If applicant is likely to repay the loan then not approving the loan , could lead to loss

2. If the applicant is not likey to repay the loan then she or he is likely to default then approving the loan may lead to financial loss.
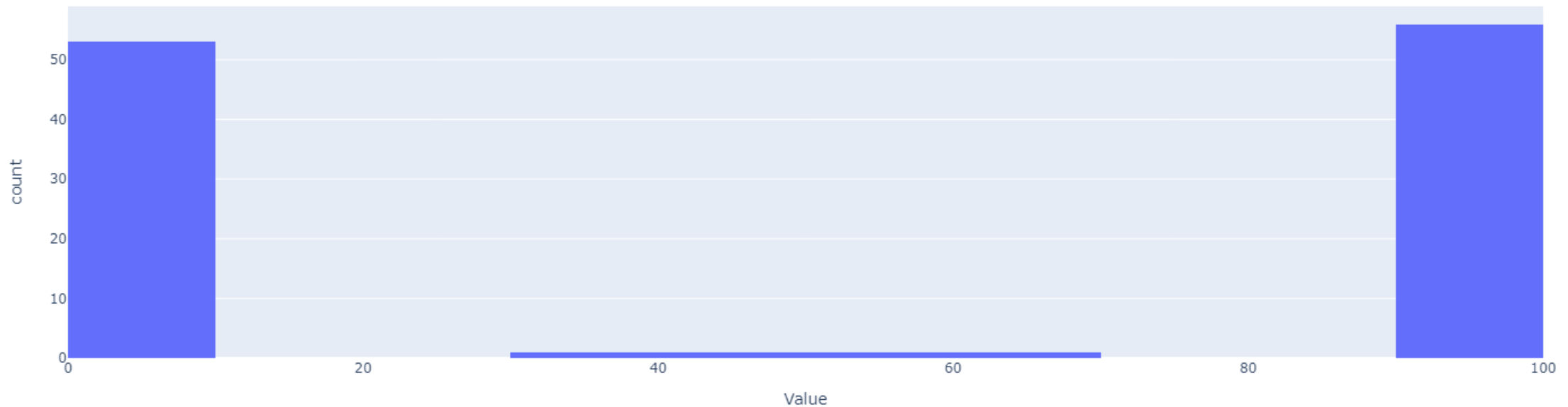
PRIMARY GOALS

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 loan amount, lending at a higher interest rate etc.

HOW WE ACHIEVE OUR GOAL

UNDERSTANDING THE DATA

IMPUTATION-HANDLING MISSING VALUES AND DATA CLEANING

OUTLIER REMOVAL

ANALYSIS – UNIVARIATE BIVARIATE AND SEGMENTED UNIVARIATE ANALYSIS

SUMMARY AND CONCLUSIONS

A LITTLE PEAK ON THE DATA CLEANING PROCESS..

The graph above shows the percentage of null values in each of the columns and there are around 56 columns above 90% null. So as a business logic all the values with null values more than 50% were dropped and the rest of the columns were filled using appropriate values.

DISTRIBUTION OF LOAN AMOUNT

Frequency of the Loan Amount
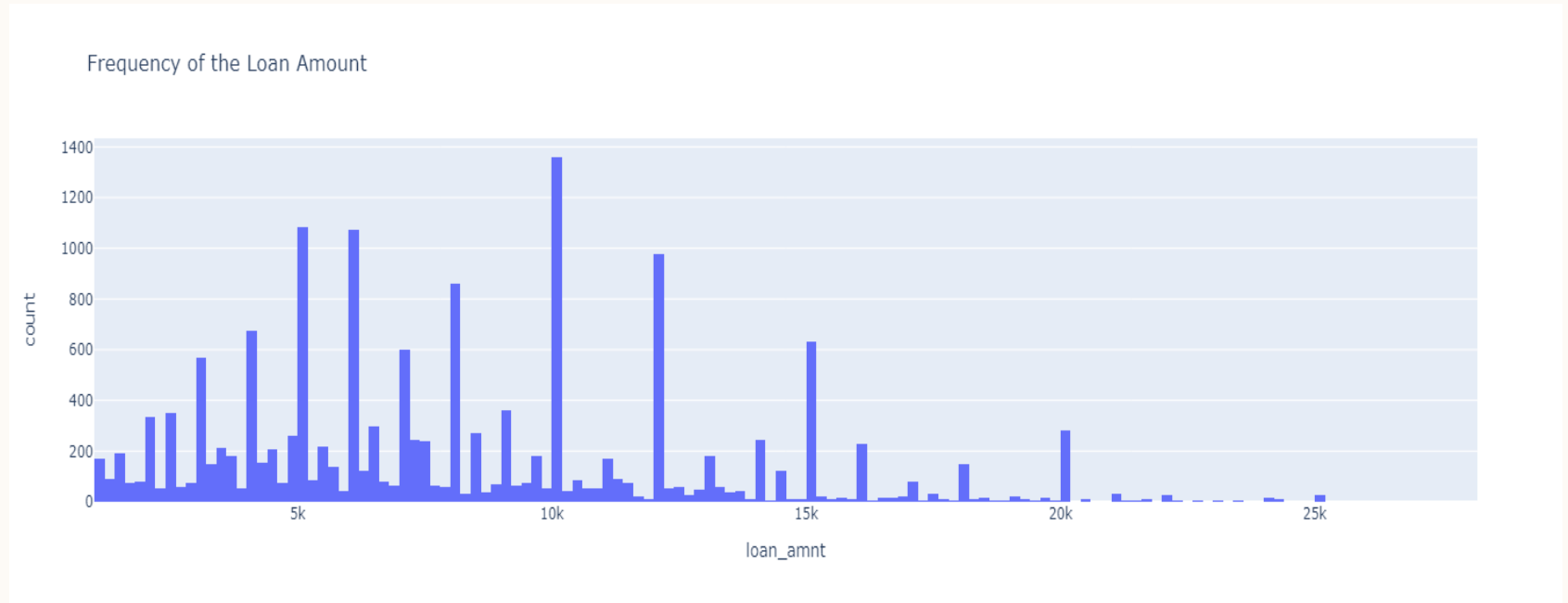
As we can see that majority of the applicants have applied for loans in the range of 10k USD.
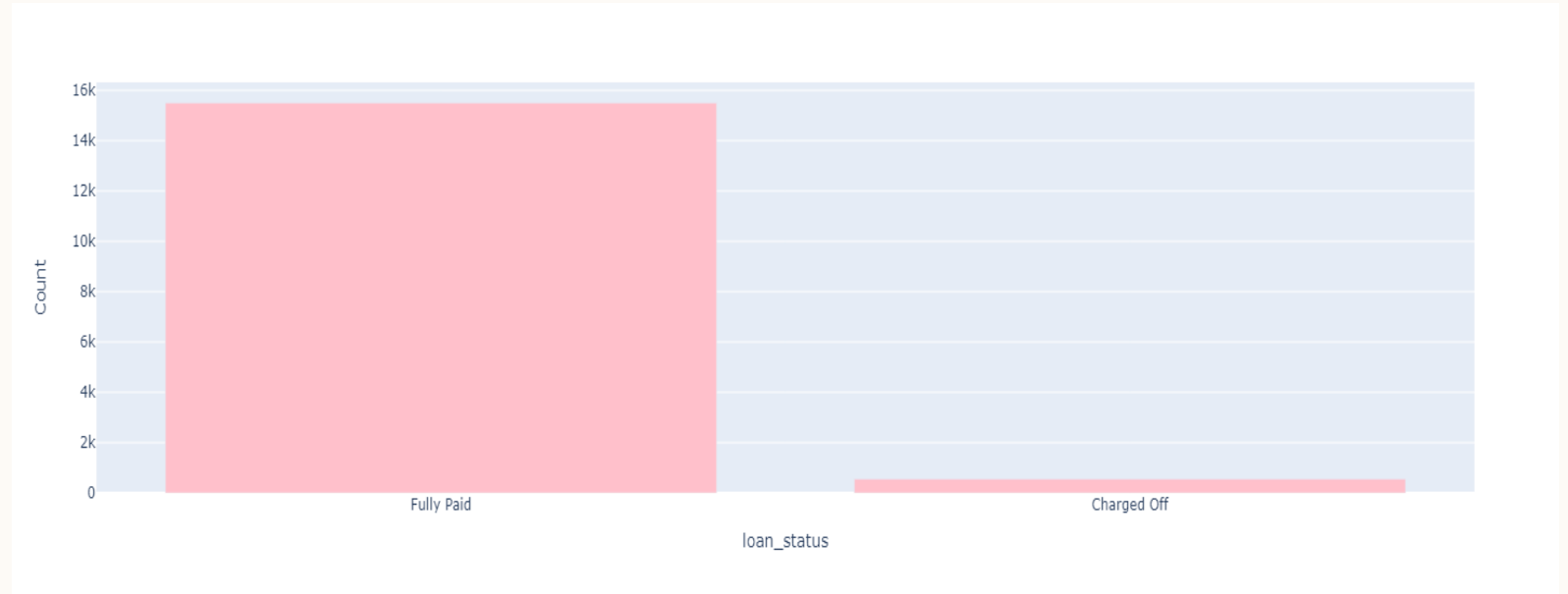The max loan applied to the bank is recorded as 28k USD.

TERM OF THE LOAN GIVEN OUT BY THE BANK

As we can see majority of the bank loans approved for are in the term of 36 months.

FREQUENCY OF THE LOANS STATUS

As we can see that majority of the applicants have paid off their loans and there are around 500 applicants who have defaulted.

UNDERSTANDING THE PATTERN OF THE LOAN AMOUNT APPLIED FOR BY THE DEFAULTERS AND THE FULLY PAID CUSTOMERS

If you see the violin plot the loan amount is slightly higher for the defaulter than those who have fully paid off. The median of the full paid customers is 7200 and for defaulter it is 9000 usd. Loans applied for by the full paid people are on the lower end than those who defaulted.
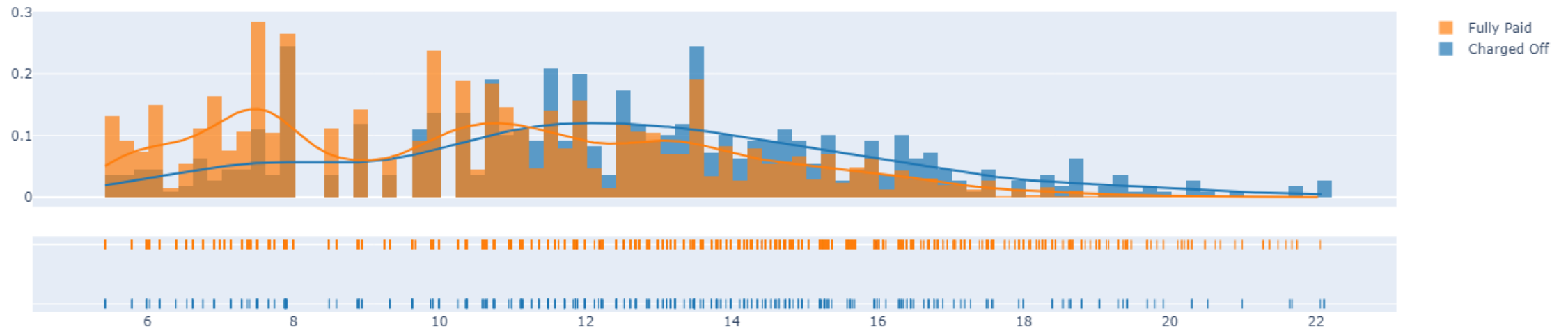
UNDERSTANDING THE TERM APPLIED FOR BY DEFAULTER AND NON-DEFAULTER
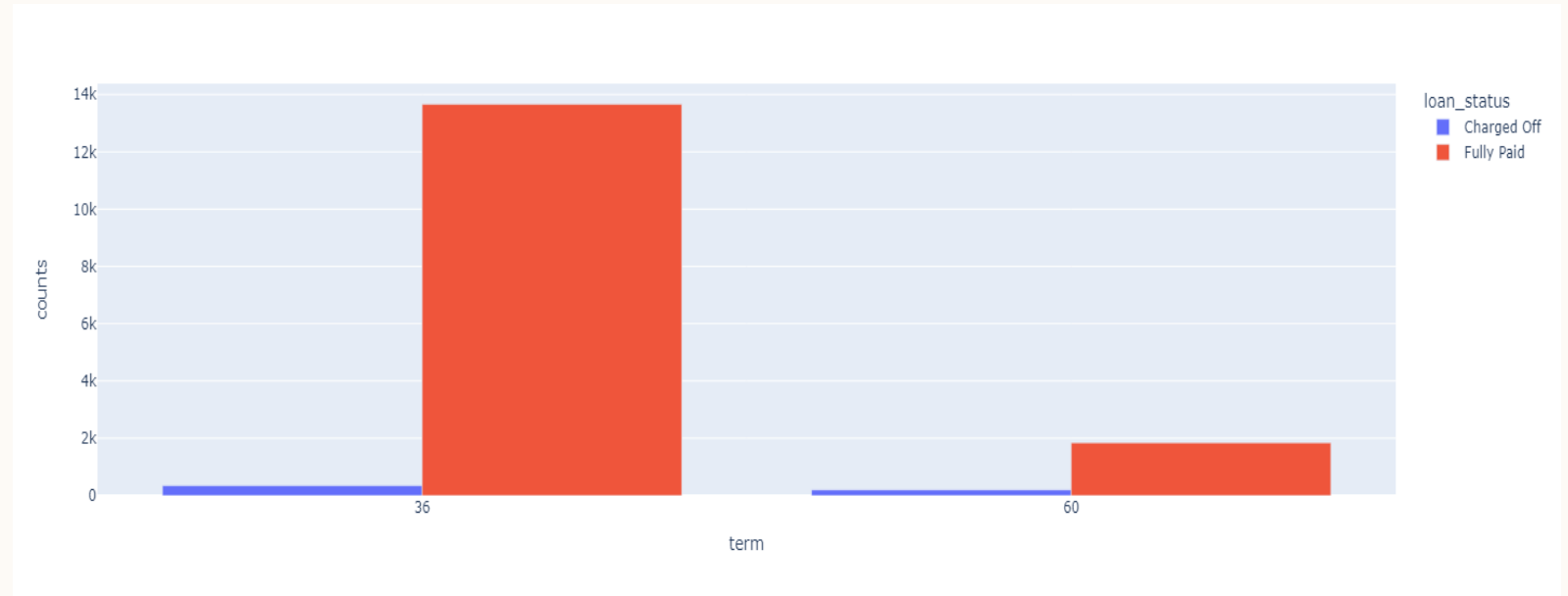
The people who have fully paid off their loans prefer the 36 months term compared to the 60 months but for the defaulters we can see that they prefer 60 months time periods over 36.

UNDERSTANDING THE INTEREST RATE OF DEFAULTERS AND PEOPLE WHO HAVE FULLY PAID OFF THEIR LOANS


This is a distribution plot for interest rate and these are the following conclusions:

1. The frequency of the rate of interest is high for people who have paid off their loans ie people who paid off their loans have cheaper rate of interest
2. The interests almost intersects at 9% for both the categories and then it is overtaken for the charged off people after 11 %
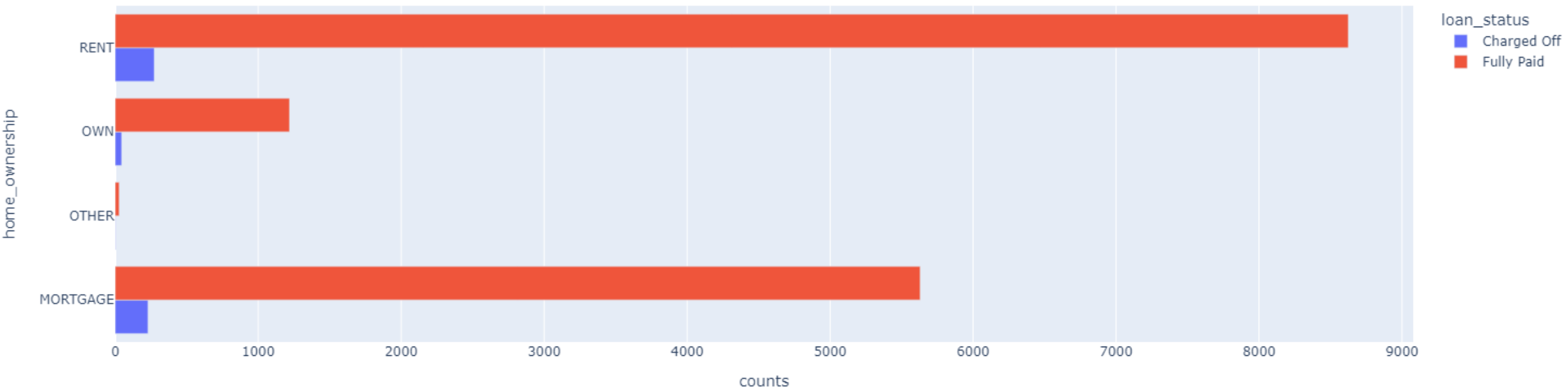
UNDERSTANDING THE INSTALLMENTS GIVEN BY THE DEFAULTER AND FULLY PAID CUSTOMERS
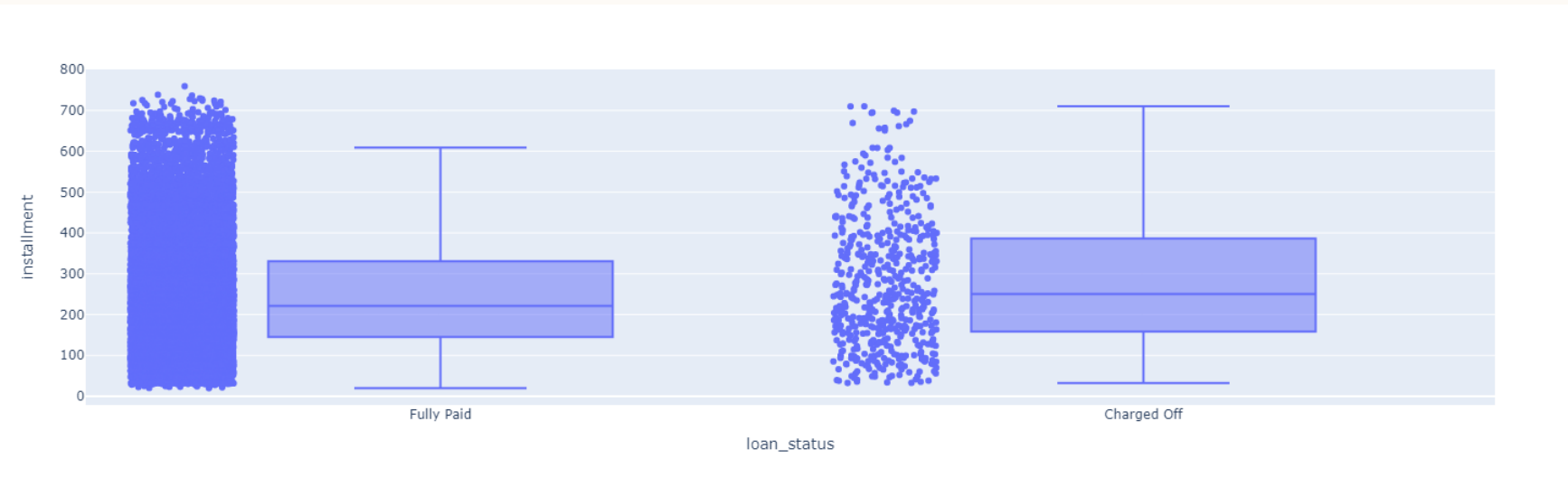
The median installment to be paid by the defaulters and those who fully have paid their loan is almost similar. The installment is slightly higher end for the charged off customers.

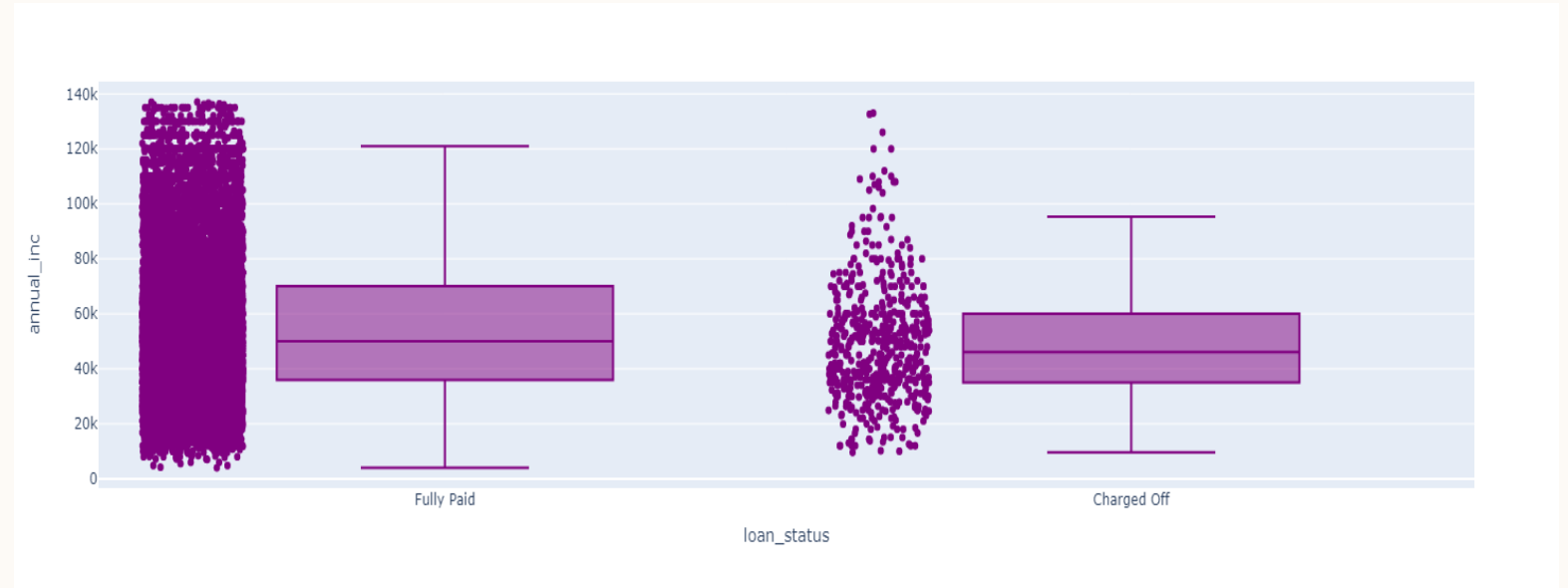# UNDERSTANDING THE HOME OWNERSHIP STATUS DEPENDING ON LOAN STATUS

For fully paid customers, majority of the people live in rent homes followed by mortgaged homes and few people owning homes.
This could be the case probably because they invest their money in some other forms like stocks rather than live in real estate.

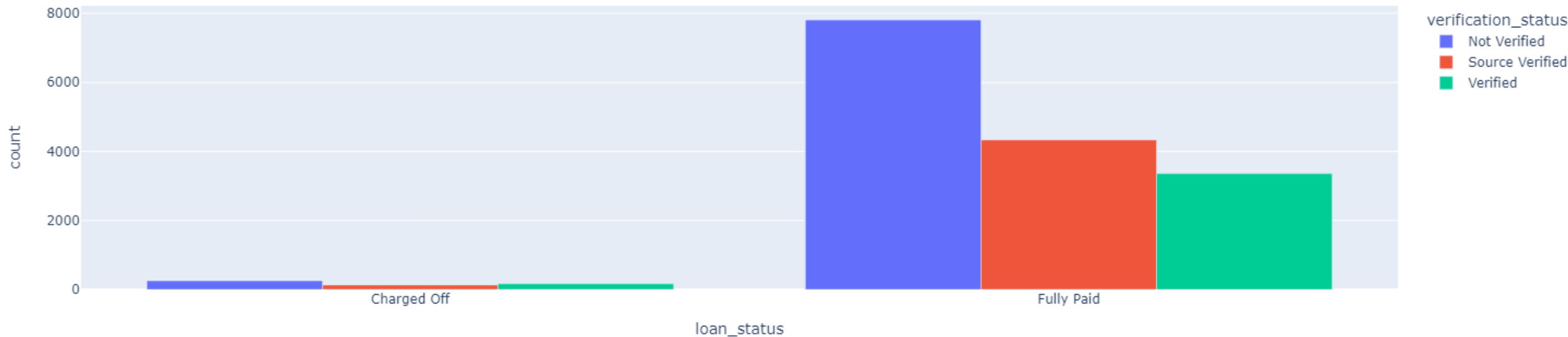# UNDERSTANDING THE ANNUAL INCOME OF THE APPLICANTS

From the above plot we can see that the annual income e of the people who have fully paid off their loans are on the higher end compared to the people who have been charged off.

UNDERSTANDING THE VERIFICATION STATUS DEPENDING ON LOAN STATUS

Comparision of verification status with respect to loan_status

An observation: Lot of the loan applications are not verified which could be risky for the bank. Hence this is something the bank should work on in order to avoid risk.
As for pattern, for fully paid loans the source verified is more than verified and this is opposite in the case of charged off.

UNDERSTANDING THE LOCATION OF VARIOUS LOAN APPLICANTS
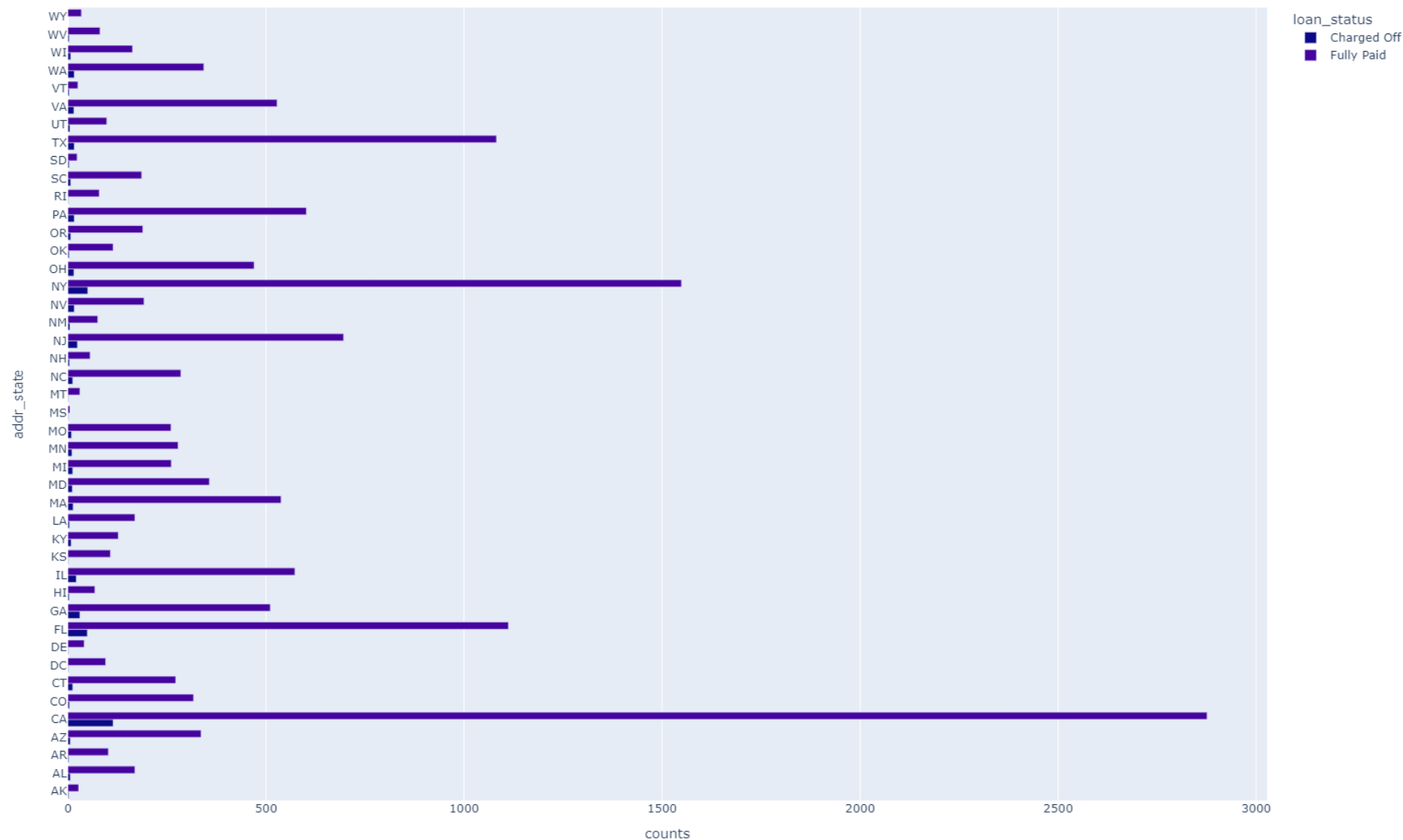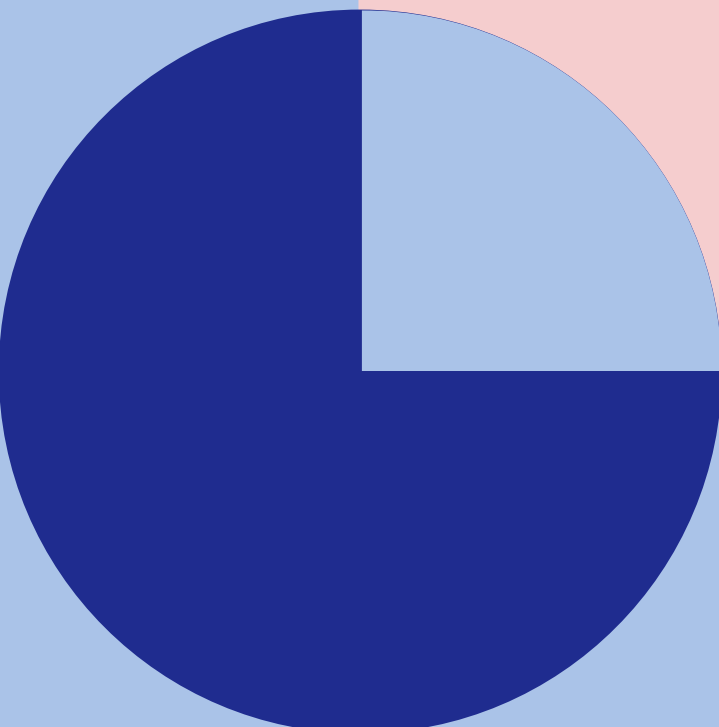
Majority of the loan applicants are from the Californian region – both defaulted as well as full paid. California is a good place for bank business. There are few states in which there are no defaulters at all probably because the loan was never sanctioned. The defaulters are mostly in developed cities. People who are middle class and lower middle class in those cities.
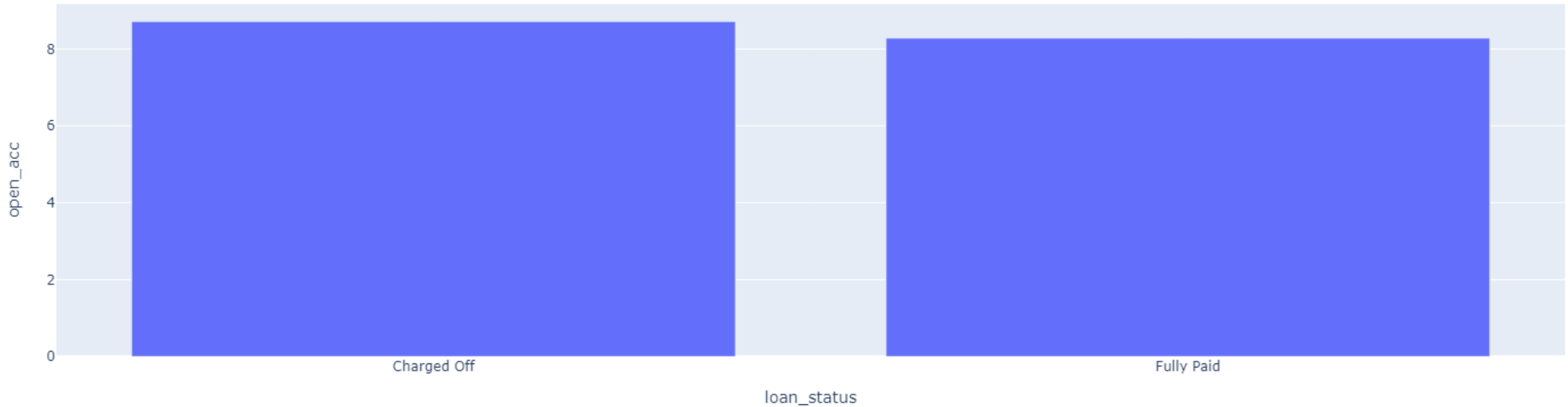
UNDERSTANDING THE RATE OF INQUIRY IN LAST 6 MONTHS
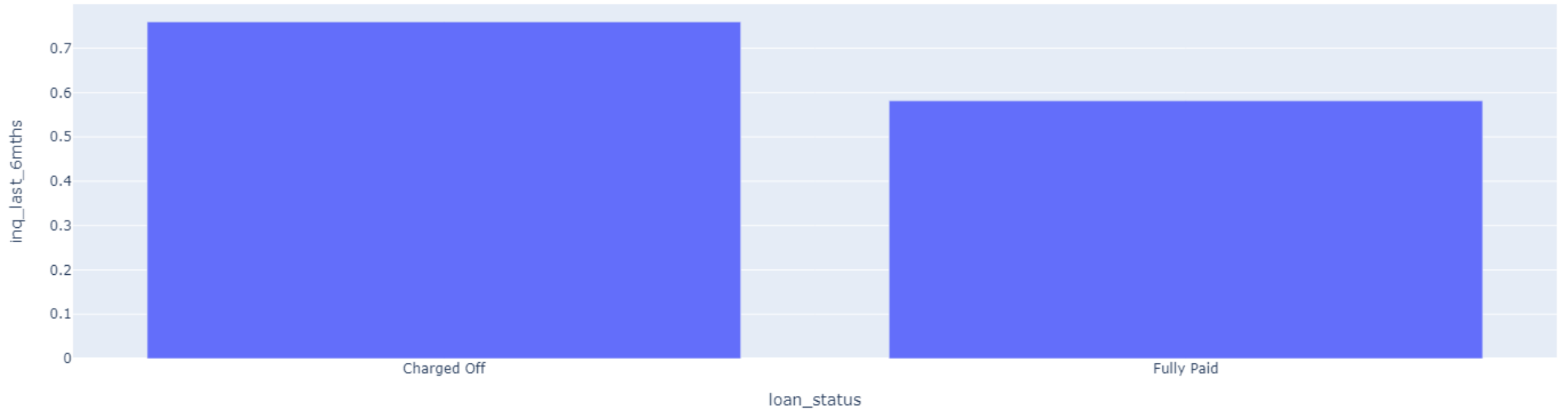
Charged off people have had higher number of inquries than fully paid people which can be a flag.

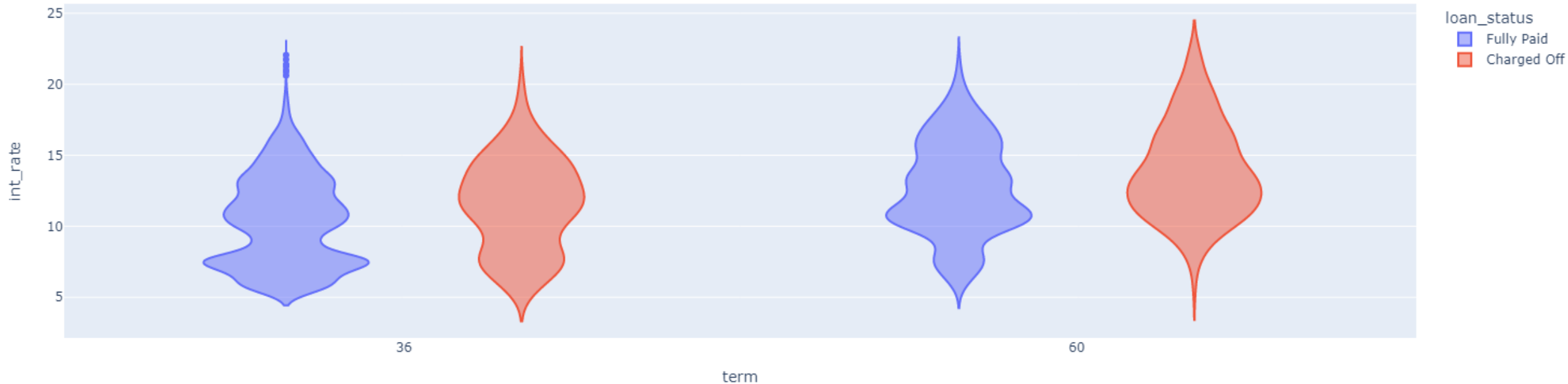# UNDERSTANDING THE PATTERN OF OPEN CREDIT LINES

People with more open credit lines are likely to be defaulters

UNDERSTANDING THE PATTERN OF THE TERM AND INTEREST RATE RELATION FOR DIFFERENT CUSTOMER TYPES


For customers in 36 months term loan, we can see that the interest rate for the majority fully paid is in the range of 5 to 10 where as for defaulters it is in the range of 10 to 15

For customers in the term of 60 months the interest rate for fully paid off people is higher compared to 36 months and is slightly lower to defaulters (majority) in the 60 months category

# UNDERSTANDING THE PATTERN OF THE HOME OWNERSHIP AND RATE OF INTEREST
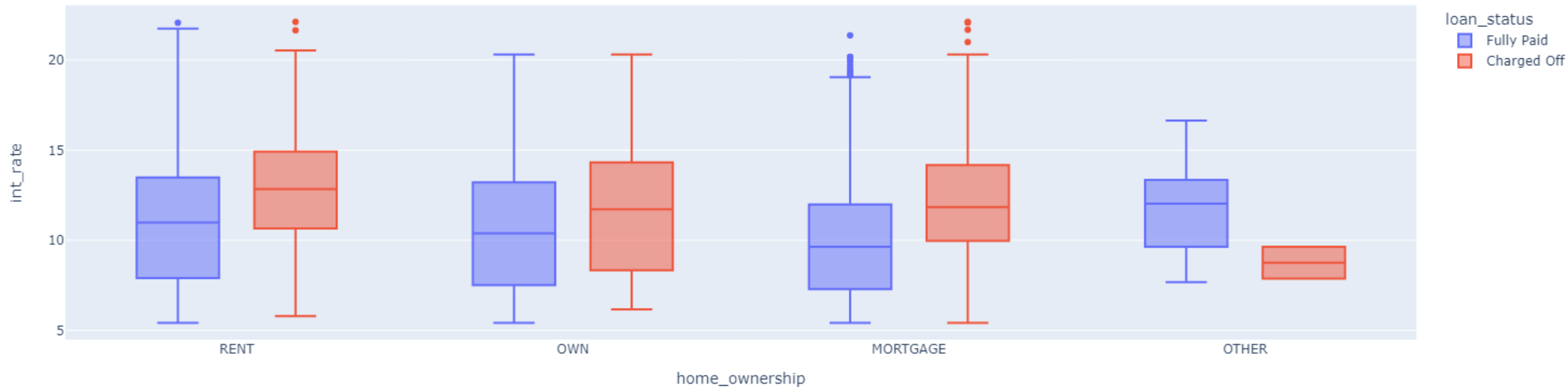
For fully paid off people, the interest rate is higher for Other category followed by rent, own and mortgage being the lowest

For charged off customers, the interest rate is higher in their category but it is the lowest for other, people who own the houses followed by mortgage and then rent.

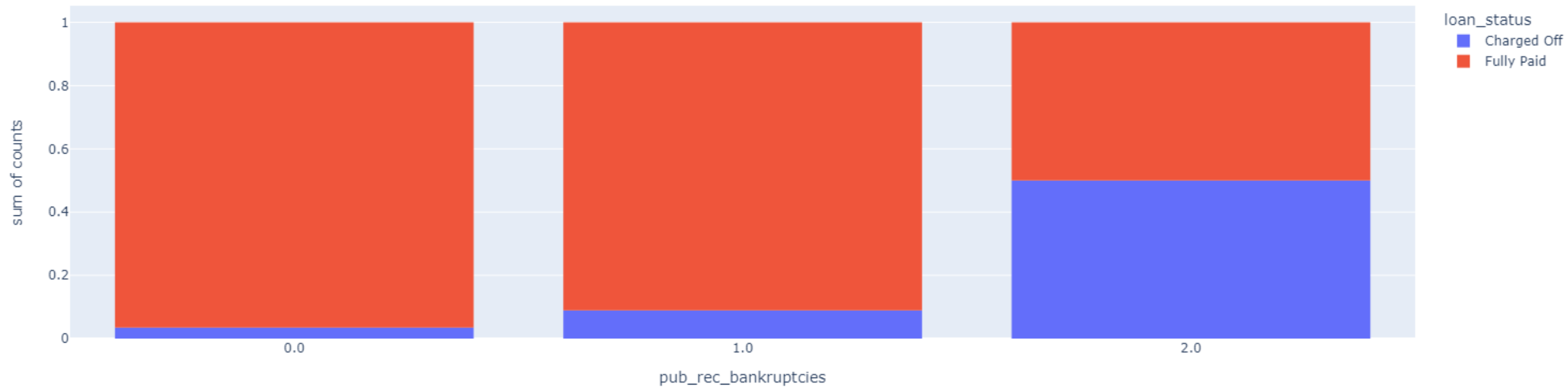# UNDERSTANDING THE PATTERN OF THE HOME OWNERSHIP AND LOAN AMOUNT
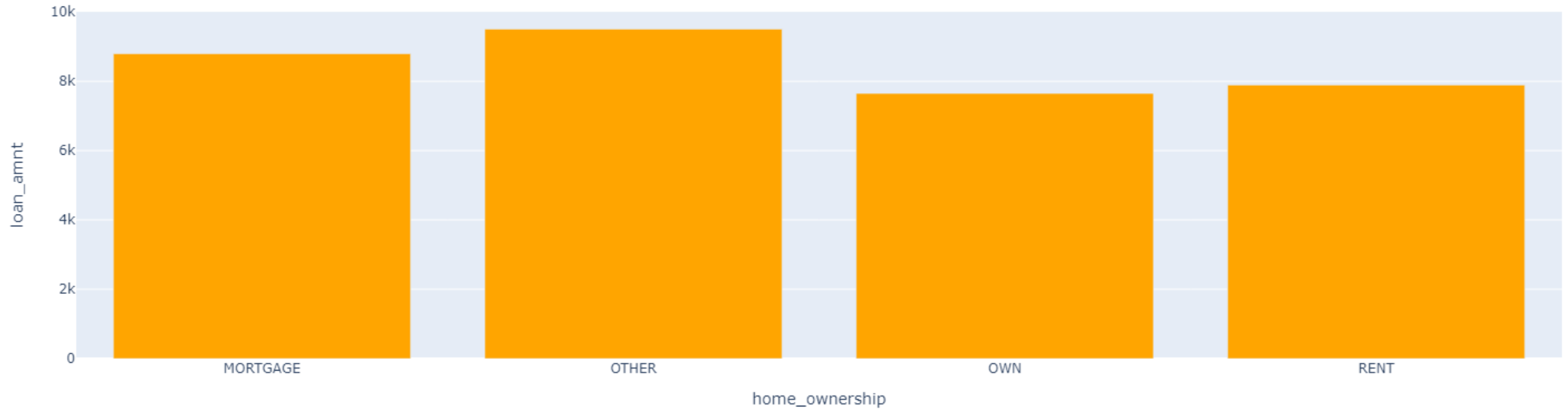
People who either had a mortgage or were in other category had higher loans in compared to people with own homes and rent.

People with own homes have more stability and are financial well off hence the loans could be lower.

UNDERSTANDING THE PATTERN OF THE BANKRUPTCIES AND LOAN STATUS

More the bankruptcies more the possibility of defaulting.

SUMMARY

Following are the metrics the bank should look out for to identify a defaulter:

1. Loan Amount: People who default have a relatively higher loan amount compared to those who fully pay off their loans

2. Employee Length: The bank approves loan easily for the people who have 10 or more than 10 years of experience but they also tend to be huge defaulters. The next highest number of defaulters are the people with less than 1 years of experience.

3. Term: People who default tend to take loans on 60 months than 36 months

4. Interest rate: People with higher rate of interest are more likely to be defaulters.

5. Grade: For fully paid customers, the highest number of applications are for A type loan and it reduces in a fashionable manner - which means that there are few customers whose loans are risky. A being the least riskiest. For Charged off customers, Grade Type B seems to be the highest after which the number of applications per risk type keep reducing probably because those loans that looked too risky were not approved by the bank to begin with.

6. Home ownership: People who own houses have more stability and the defaulters in this case are lower.

7. Inquiry in Last 6 months: People who have had large number of inquries are more likely to default

8. Open Credit Lines: Charged off people tend to have more open credit lines.

9. Number of Bankruptcies: More the bankruptcies; more the chance of defaulting.

THANK YOU