Lending Club Case Study

Exploratory Data Analysis

Pratik Thombre - MLC50

Understanding Data
Columns Study

1.**auunal_inc** = It contains the annual income of the customer.

2.monthly_**inc** = It contains the monthly income of the customer.

3.**verification_status** = It contains the info regarding if the source of income of customer

   was verified or not.

4.**emp_length** = It contains the info about the years a customer is working.

5.**home_ownership** = It contains the info about the home ownership, like if the cusomer

   owns the house or live in rent, etc.

6.**grade** = It contains the info about the LC assigned grade.

7.**sub_grade** = It contains the info about the LC assigned sub-grade.

8.**loan_amnt** = How much the loan was alloted.

9.**installment** = How much the installment was.

10.**int_rate** = At what interest rate was the loan given.

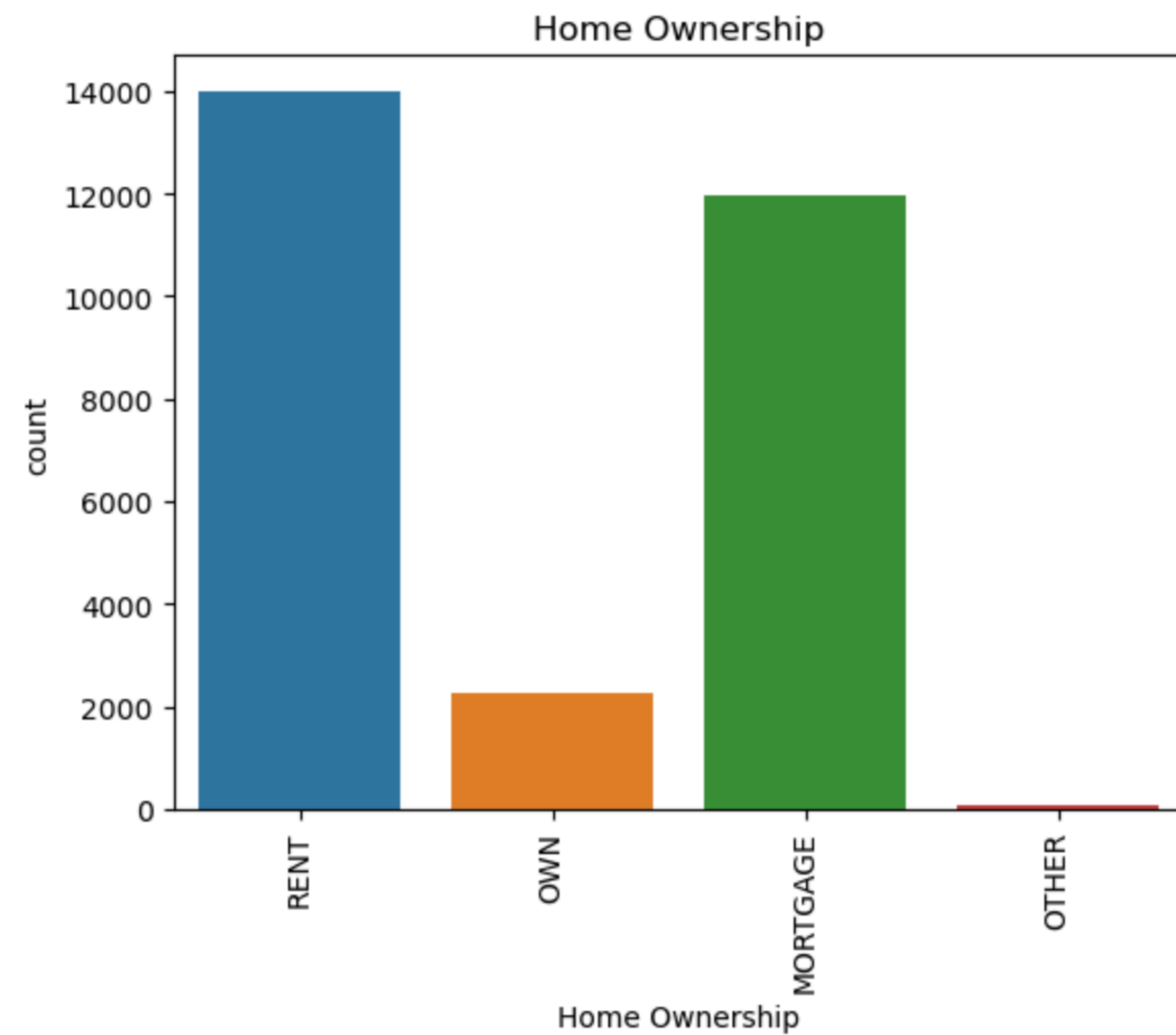11.**loan_status** = Status of loan, if the loan was paid or not.

Categorical
- verification_status
- emp_length
- home_ownership
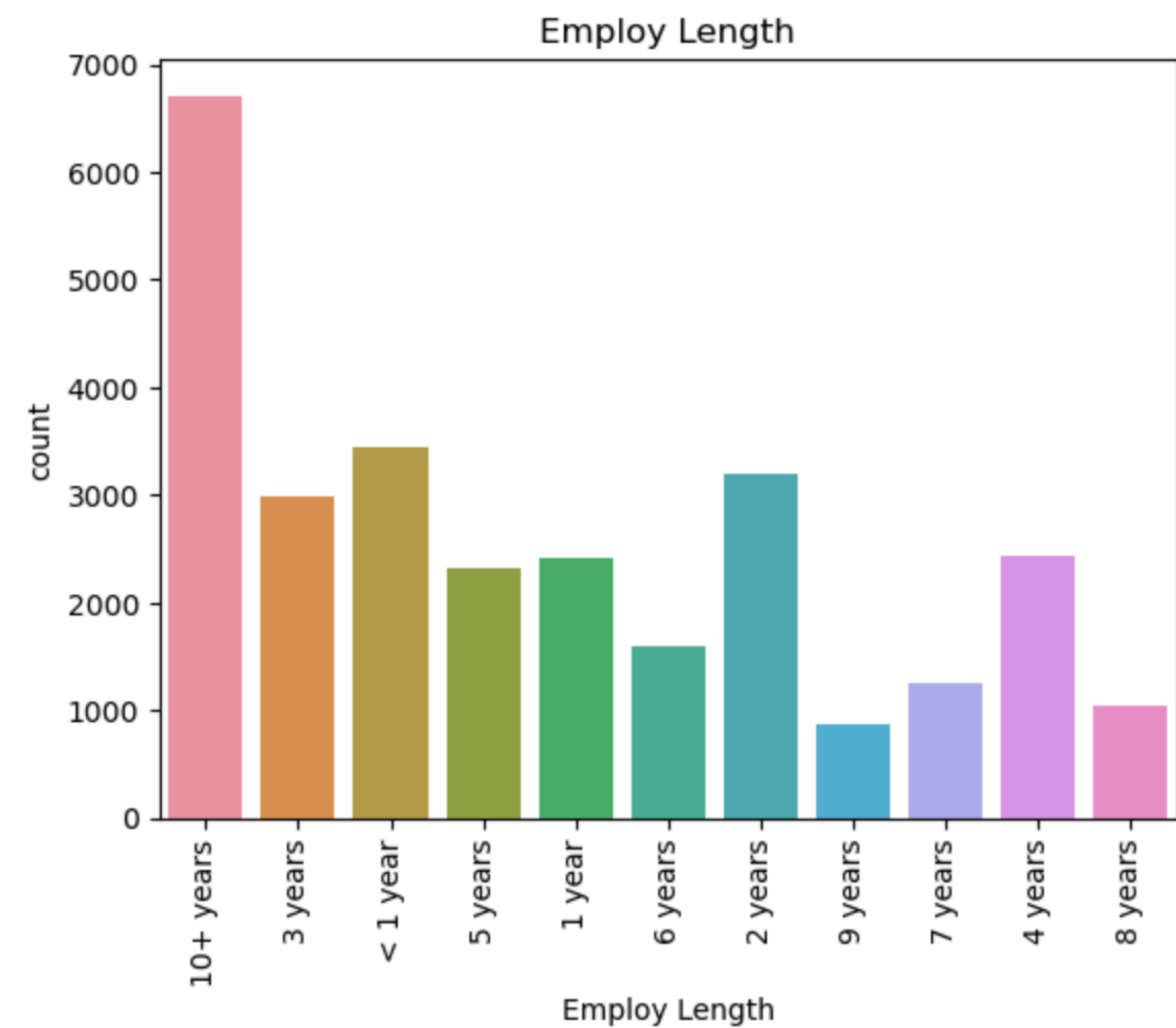- grade
- sub_grade
- loan_status

Quantative
- annual_inc
- monthly_inc
- loan_amnt
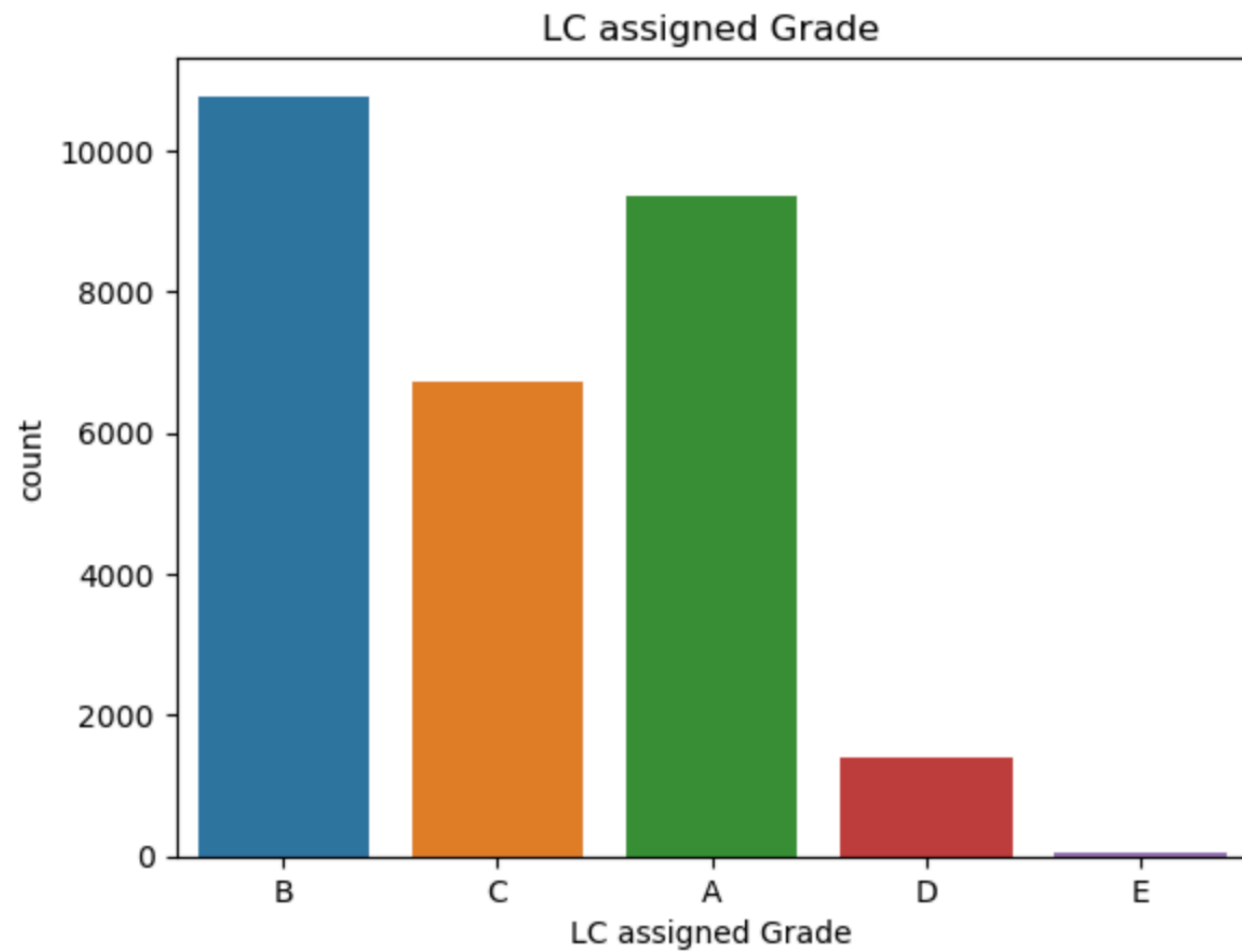- installment
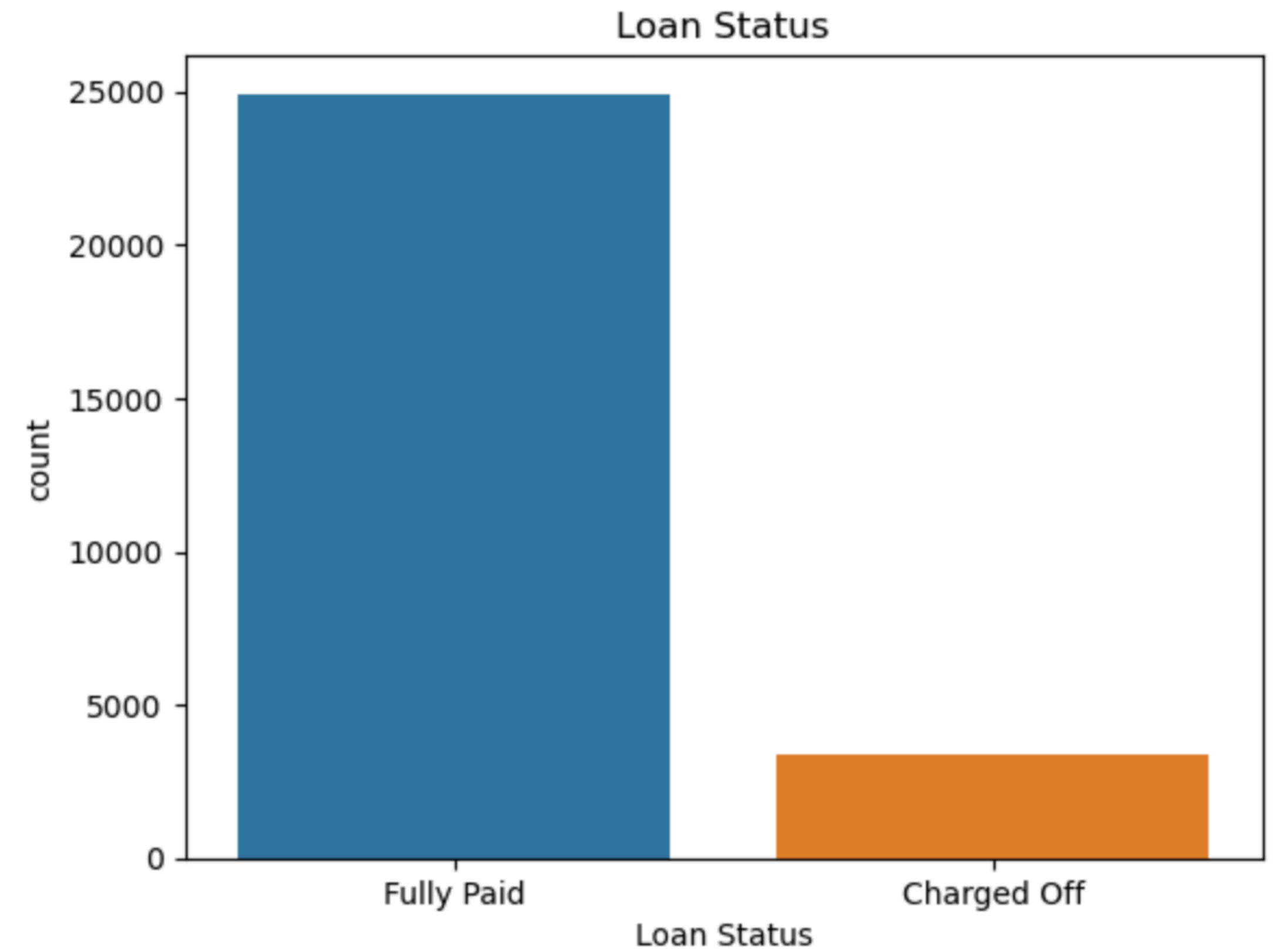- int_rate

Univariate Analysis

It was very clear that, most of the loan takers are living on rent or had mortgaged their home.

Most of the loan is already given to the customers with 10+ years experience.

As per the LC grades, most of the loan takers stands in grade A and B which is why the amount of Fully paid customers are more then the Charged Off.

Most of the them has fully paid as mostly the loan was given to the grade 'A' and 'B' customers.

Observations from univariate analysis

1. Most of the Loans were given to the customers who has 10* years of working period.

2. Most of the Loan takers were living in rent or had mortgaged their homes.

3. Most of the Loans were given to the grade 'A' and 'B' customers.

4. 75% quantile Monthly income of customers is around 7,000rs.

5. 75% quantile Loan amount of customers is around 14,500rs.

6. 75% quantile Installment of customers is around 400rs.

7. 75% quantile Interest Rate of customers is around 15%.

Bivariate Analysis

Count of loan status as per other categorial variables with respect to Loan Status.

Count of Loan Status as per Employ Length
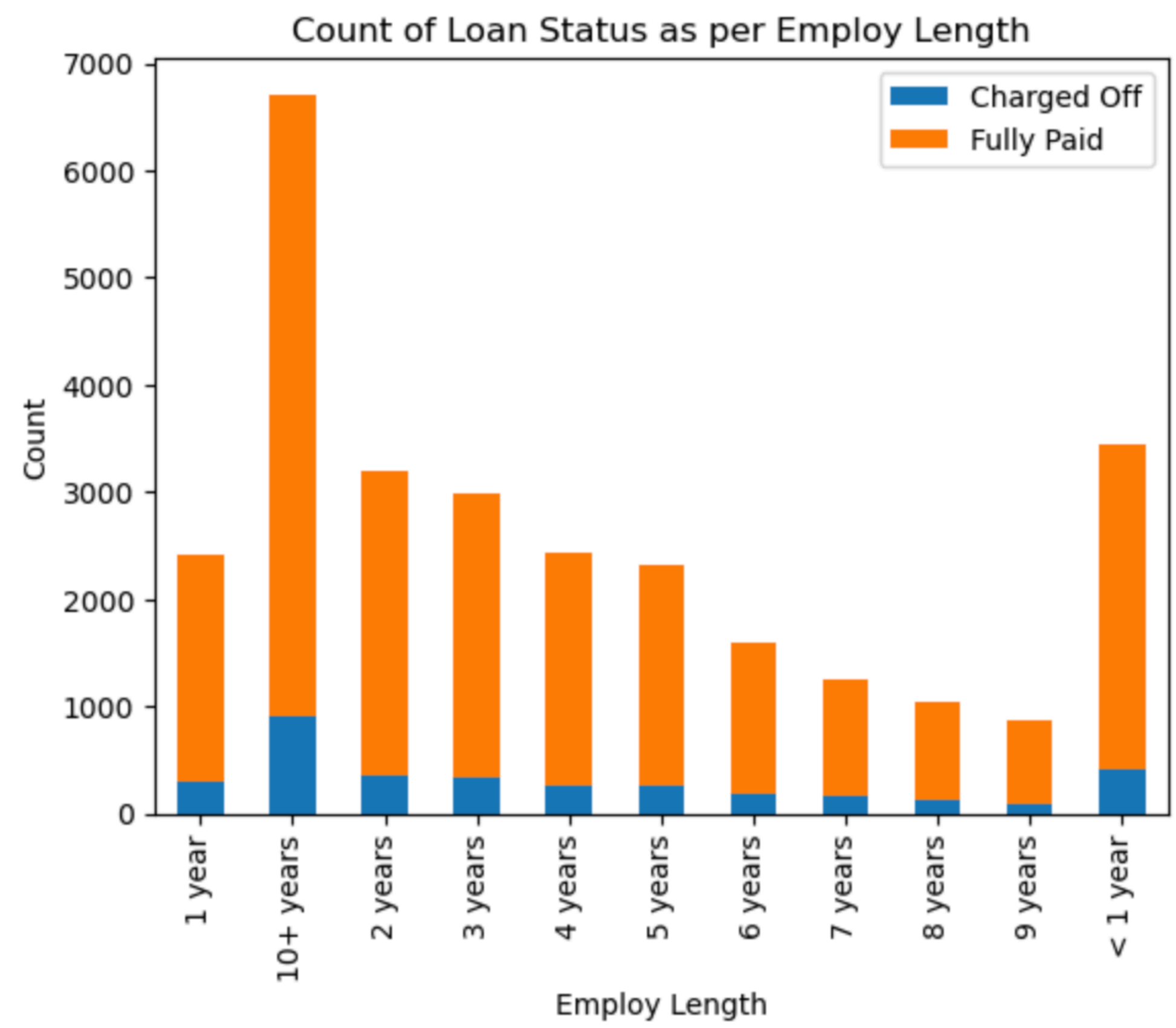
1. Here we can see that, around 12.15% customers who were "< 1 Year" had Charged Off.
2. Here we can see that, around 12.27% customers who were "1 Year" had Charged Off

```
10+ years
Charged Off : 13.55%, Fully Paid : 86.45%

3 years
Charged Off : 11.06%, Fully Paid : 88.94%

< 1 year
Charged Off : 12.15%, Fully Paid : 87.85%

5 years
Charged Off : 11.34%, Fully Paid : 88.66%

1 year
Charged Off : 12.27%, Fully Paid : 87.73%

6 years
Charged Off : 11.06%, Fully Paid : 88.94%

2 years
Charged Off : 10.84%, Fully Paid : 89.16%

9 years
Charged Off : 9.92%, Fully Paid : 90.08%

7 years
Charged Off : 12.44%, Fully Paid : 87.56%

4 years
Charged Off : 10.78%, Fully Paid : 89.22%

8 years
Charged Off : 11.44%, Fully Paid : 88.56%
```
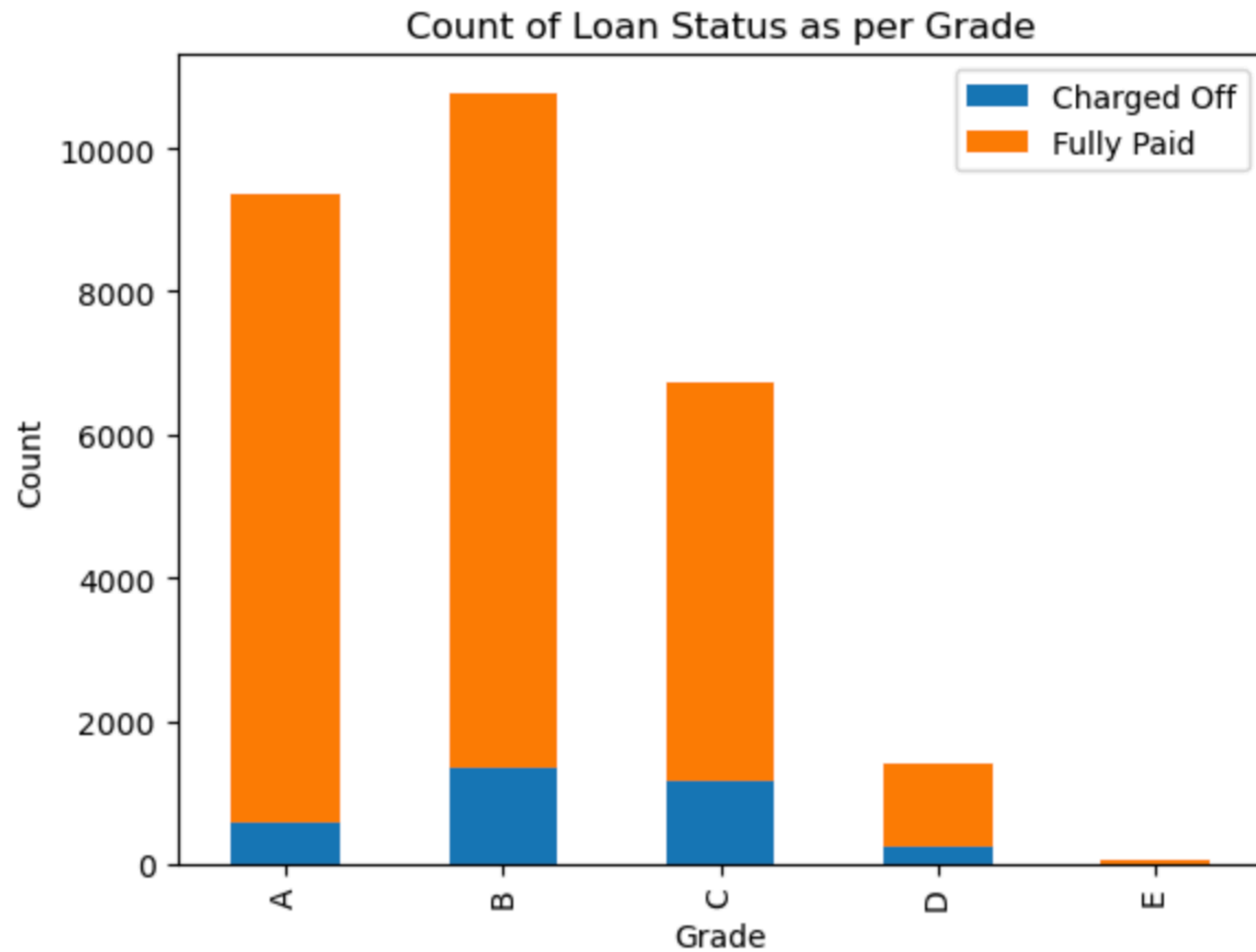
Count of Loan Status as per Grade

B
Charged Off : 12.57%, Fully Paid : 87.43%

C
Charged Off : 17.29%, Fully Paid : 82.710000000000001%

A
Charged Off : 6.23%, Fully Paid : 93.77%

D
Charged Off : 18.36%, Fully Paid : 81.64%

E
Charged Off : 21.57%, Fully Paid : 78.43%

1. Here we can see that, around 17.29% customers who were "C grade" had Charged Off.
2. Here we can see that, around 18.36% customers who were "D grade" had Charged Off.
3. Here we can see that, around 21.57% customers who were "E grade" had Charged Off.

Count of Loan Status as per Home Ownership
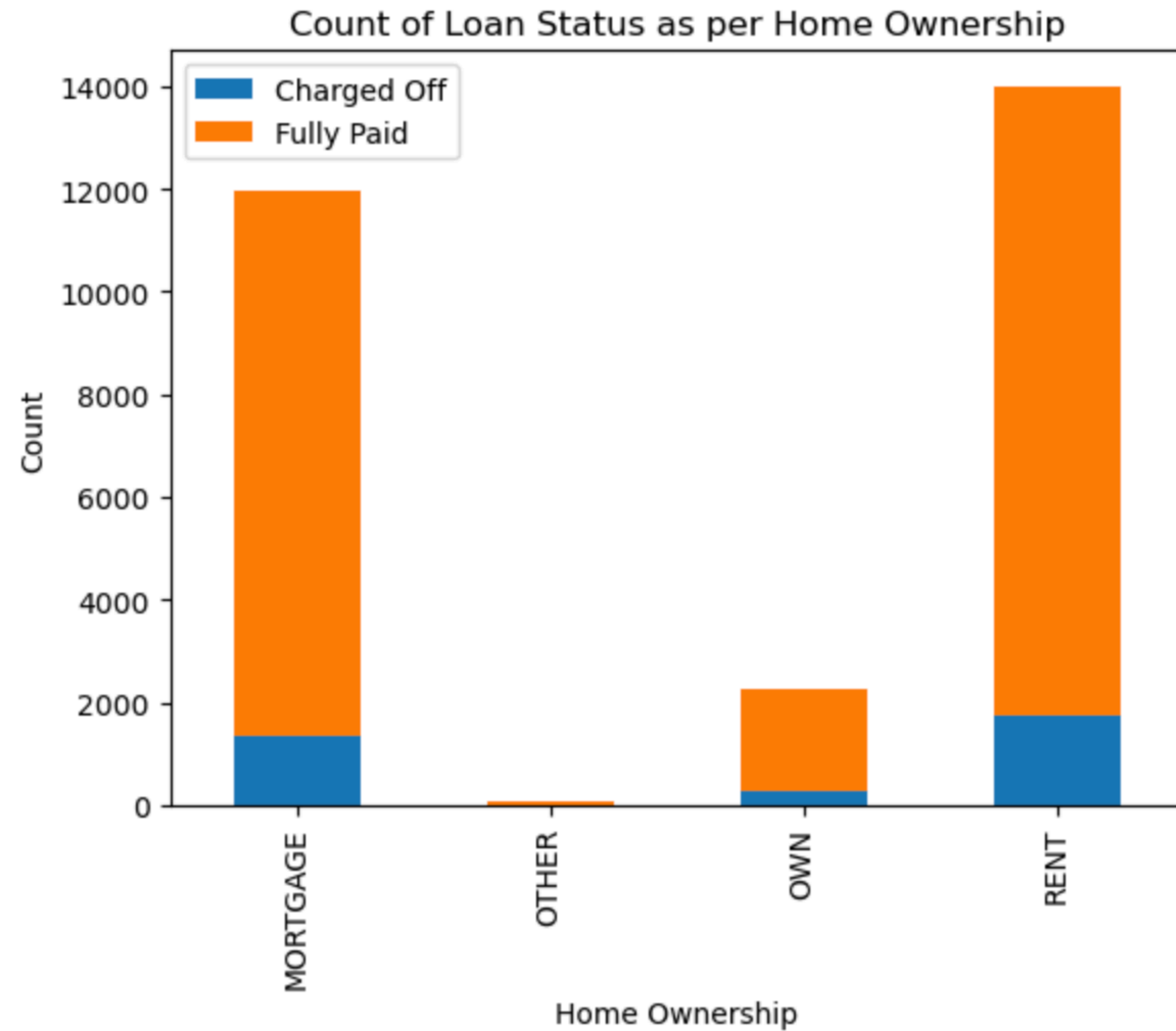
RENT
Charged Off : 12.47%, Fully Paid : 87.53%

OWN
Charged Off : 12.18%, Fully Paid : 87.82%

MORTGAGE
Charged Off : 11.17%, Fully Paid : 88.83%

OTHER
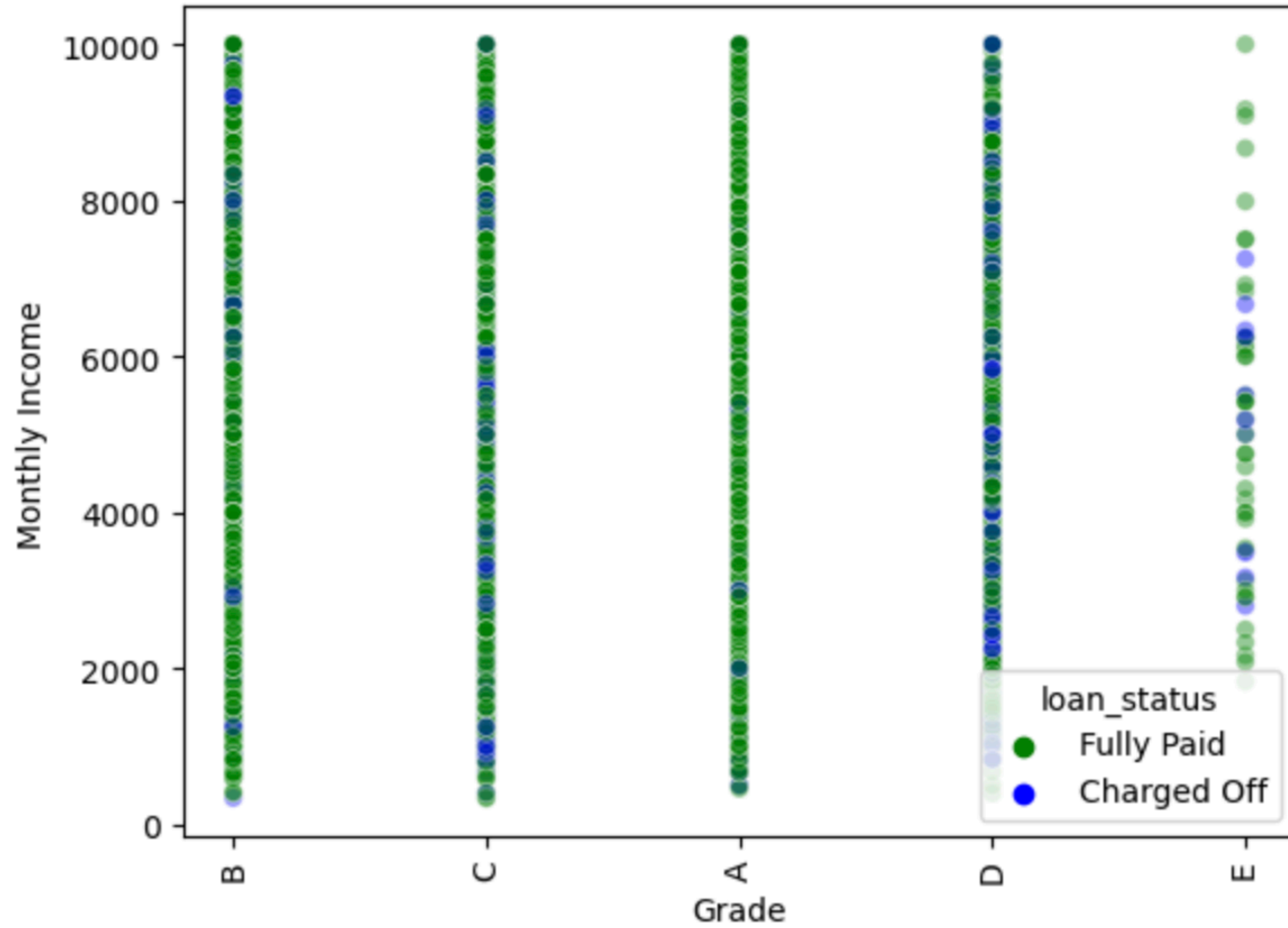Charged Off : 13.89%, Fully Paid : 86.11%

Here we can see that, around 13.89% customers whose home ownership is not known which is "Other" had Charged Off
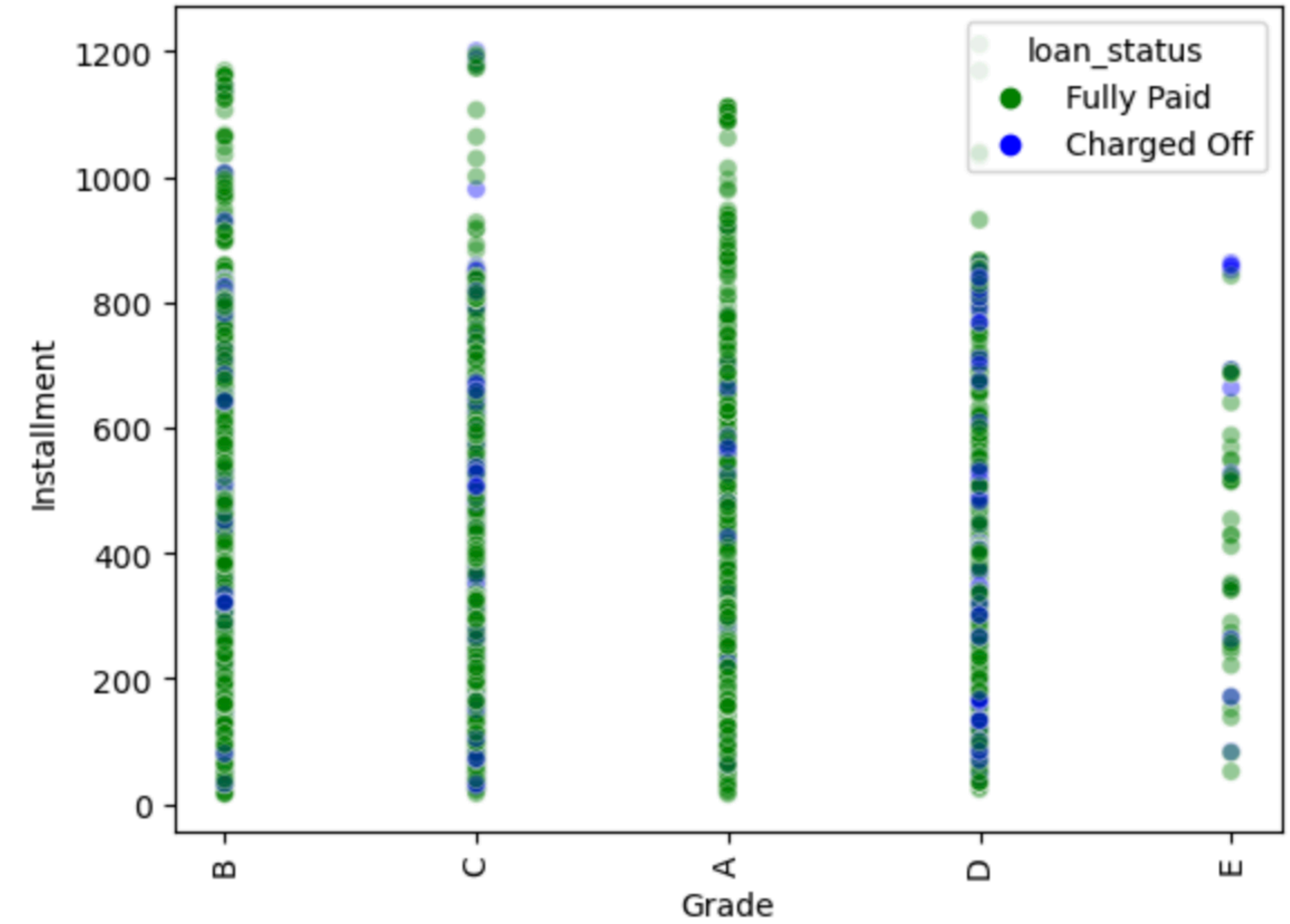
Observations from bivariate analysis

1. If the customer has 1 or less than 1 year working experience then it is morelikely that he will default.

2. If the customer has LC grade of 'C', 'D', 'E'. are more likely to default.

3. If the customers home ownership is not clear, so he is morelikely to default.
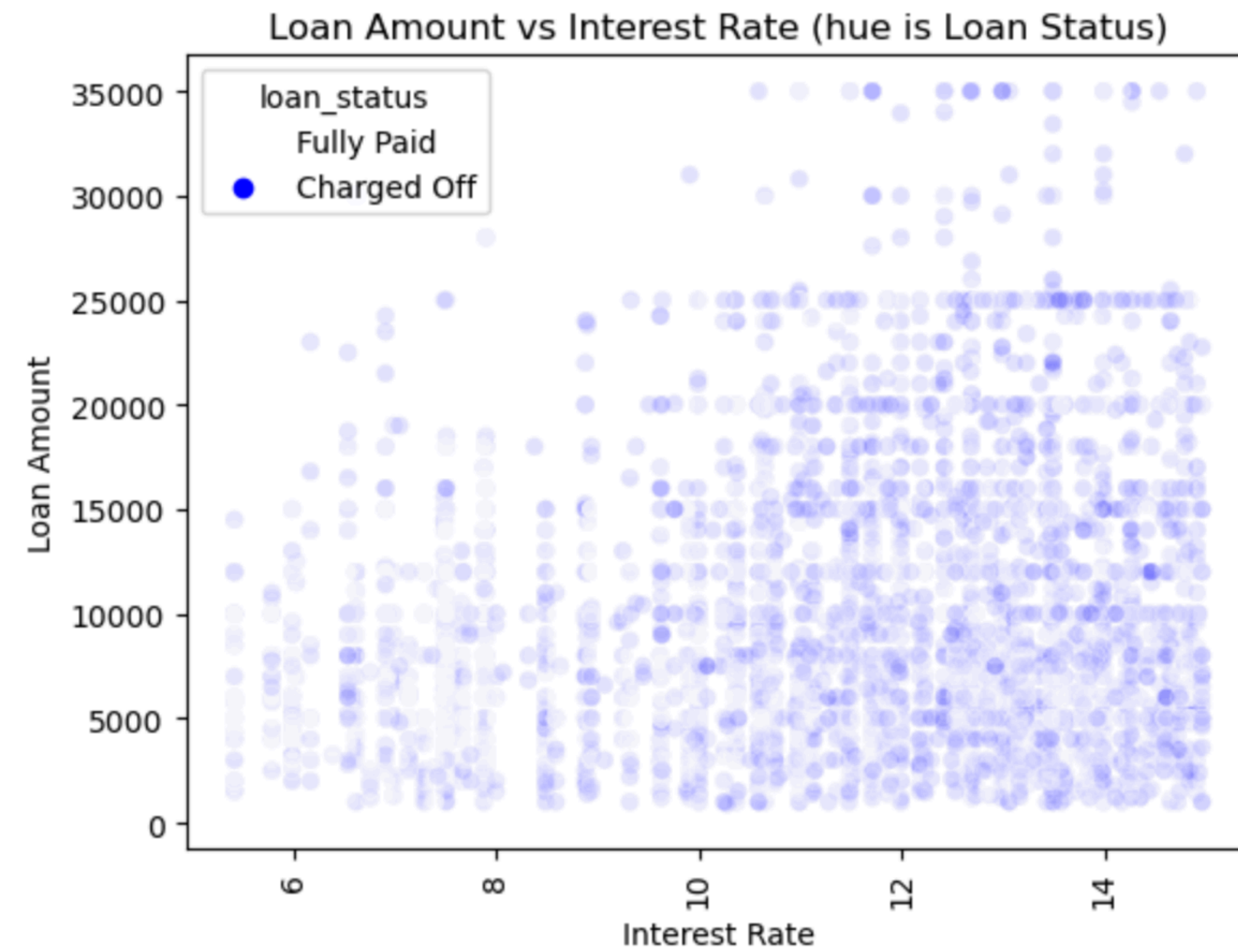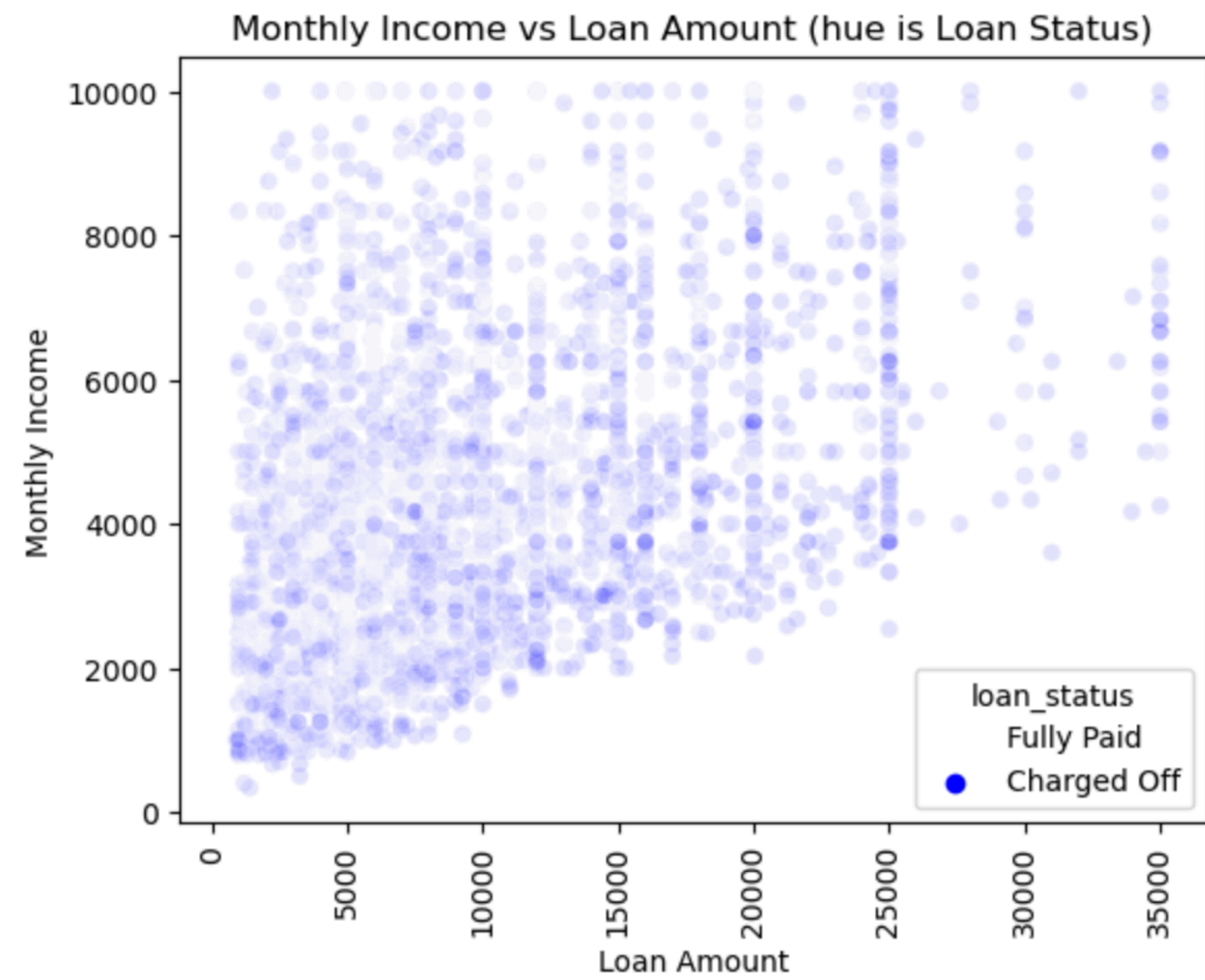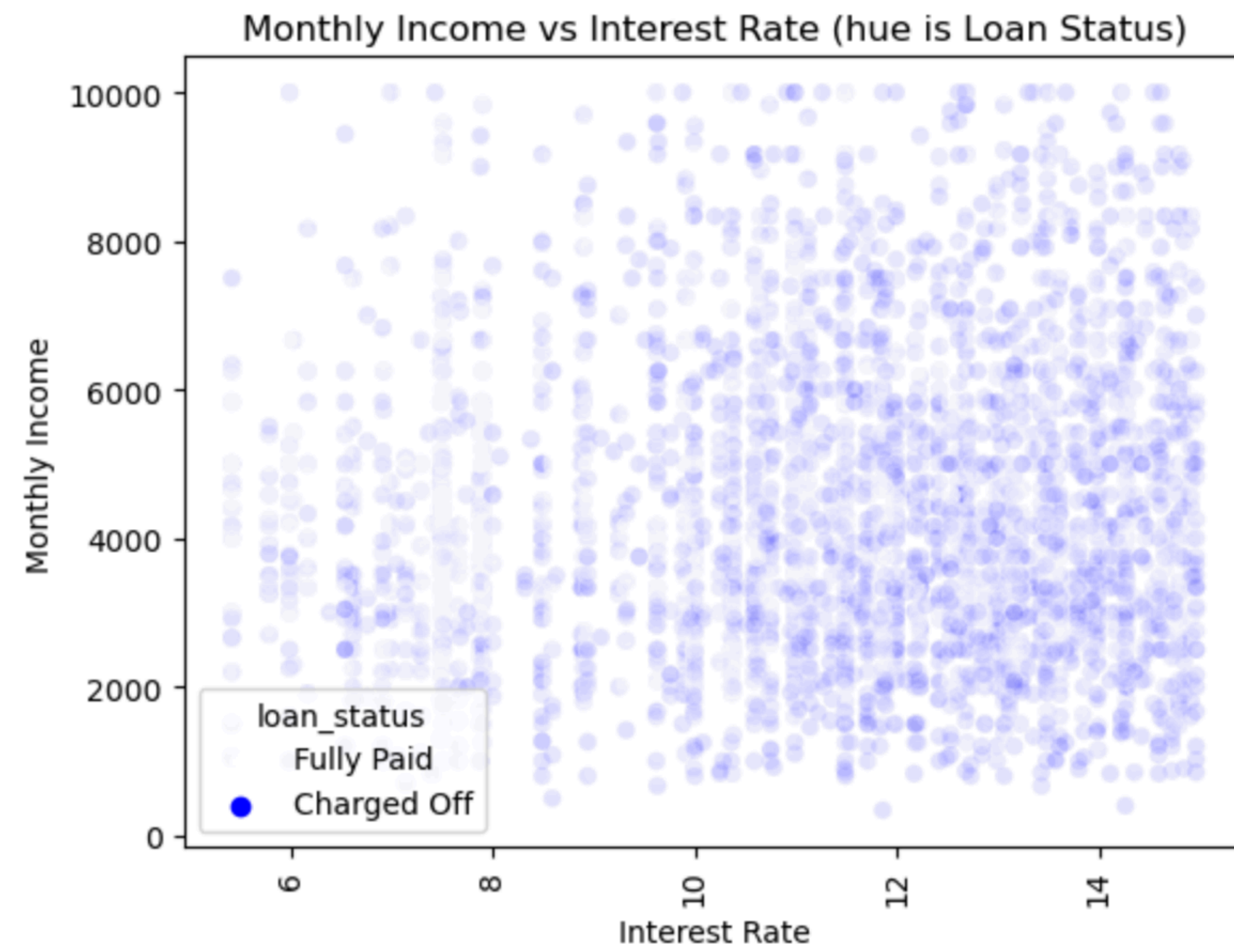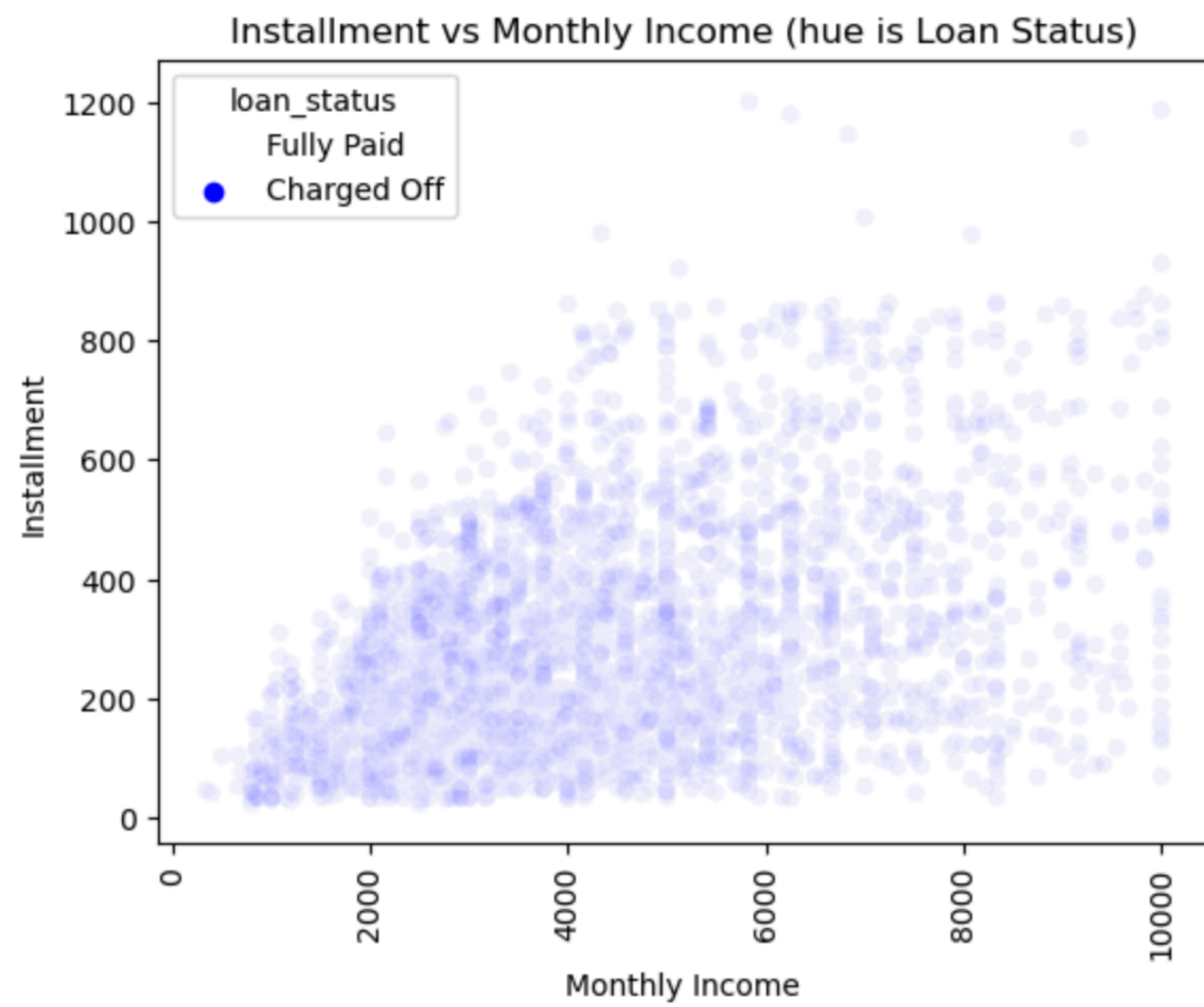
Multivariate Analysis

It is very clear that the 'C', 'D', 'E' or greater then these categories are not good to lend money.
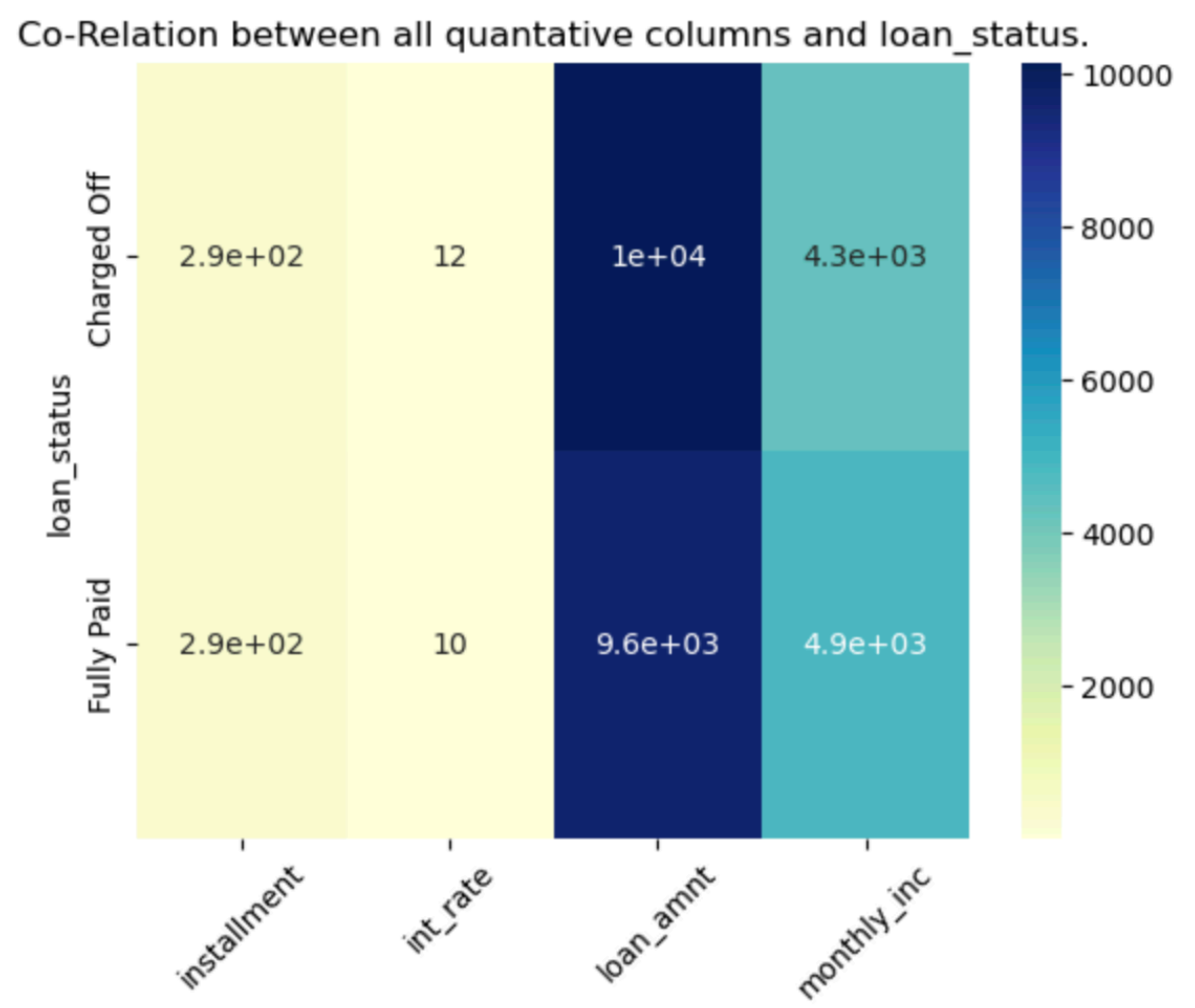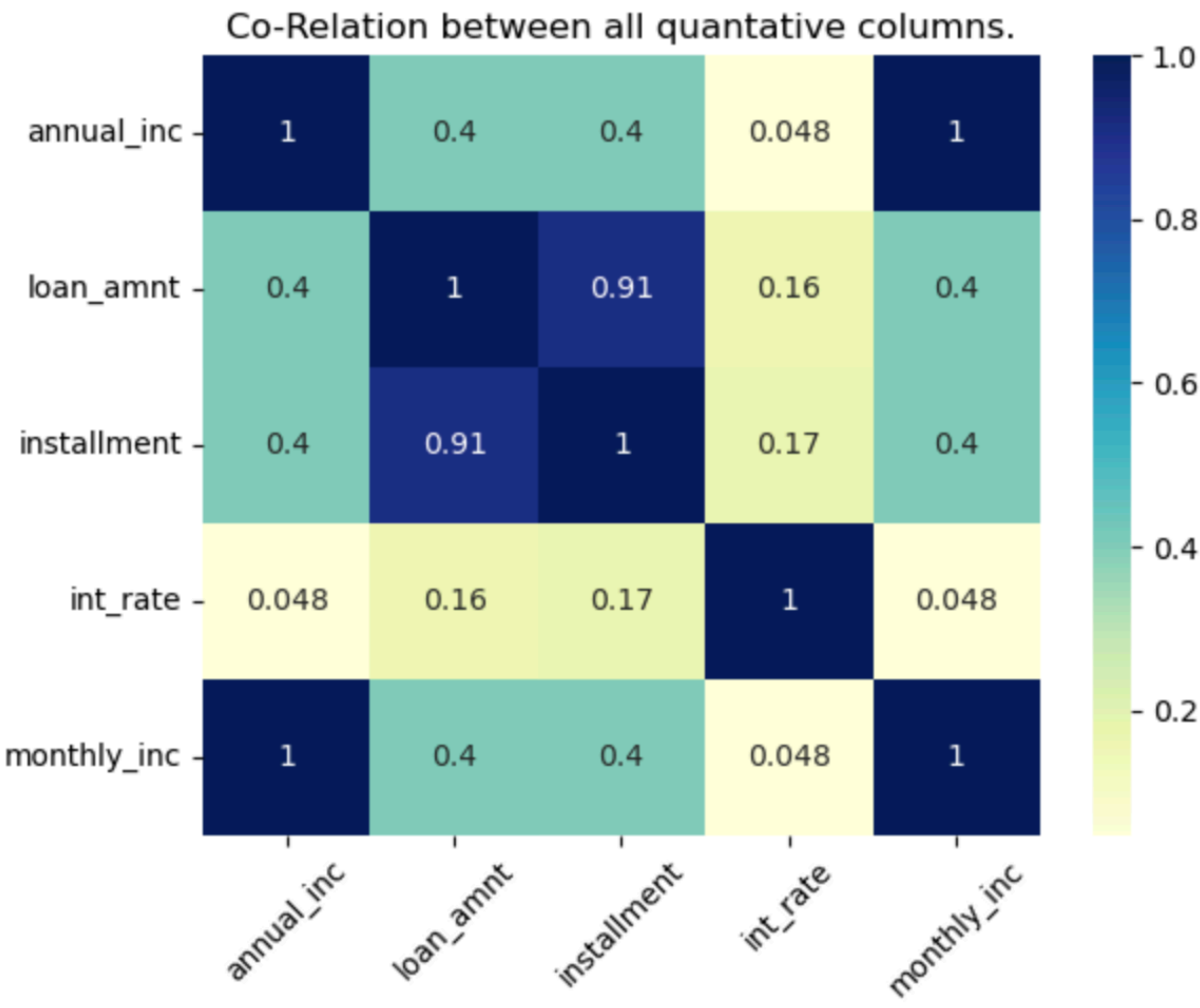
Installment vs Home Ownership (hue is Loan Status)
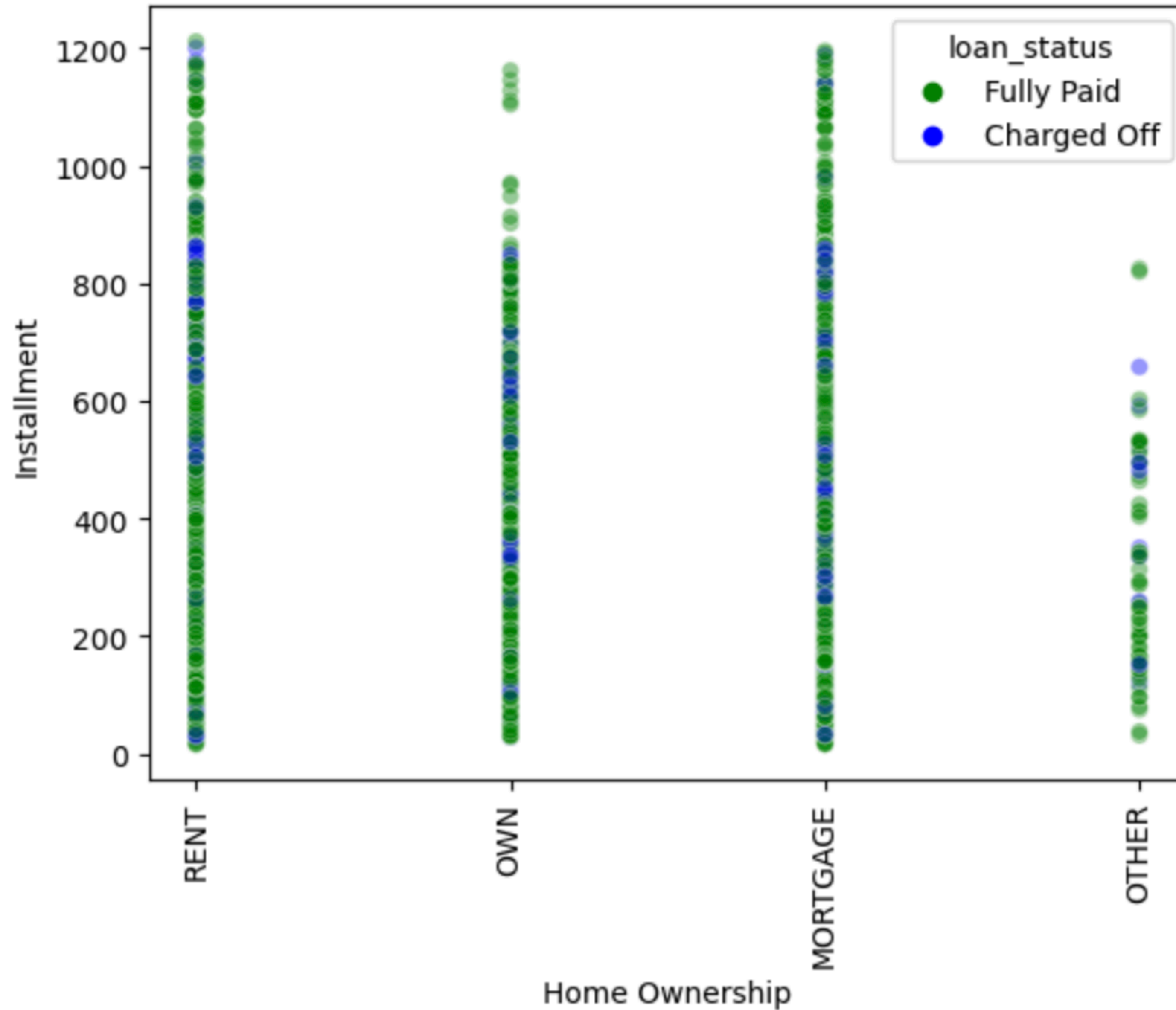
It is very clear that the 'Mortgage' and 'Other' Home ownership categories are not so well in paying loans

Installment vs Monthly Income (hue is Loan Status)

Monthly Income vs Interest Rate (hue is Loan Status)

Monthly Income vs Loan Amount (hue is Loan Status)

Loan Amount vs Interest Rate (hue is Loan Status)

1. Major co-relations are seen in Interest Rate and monthly income and installments.

2. It shown that, more the interest rate: more the chances of default.

Co-Relation between all quantative columns.

Co-Relation between all quantative columns and loan_status.

1. Monthly Income and Loan amount shares good co-relation with each other.

2. Monthly inc and Charged OFF also shares good relations.

Observations from multivariate analysis

1. It can be clearly seen here that lesser the monthly income, it is more likely be charged off.

2. And more the loan, more the chances of getting charged off if the monthly income is less as monthly income and loan

    amount shares good co relation with each other which is shown in the above co-relation map.