

Lending Club Case Study

By:

Manu Sijaria

Aryan Baruah

Overview

- About Company : Lending Club 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.
- Problem statement: Analyze data to identify
 - The driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
 - Consumer and loan attributes that influence the tendency of default.

Methodology



Data Analysis and Cleaning

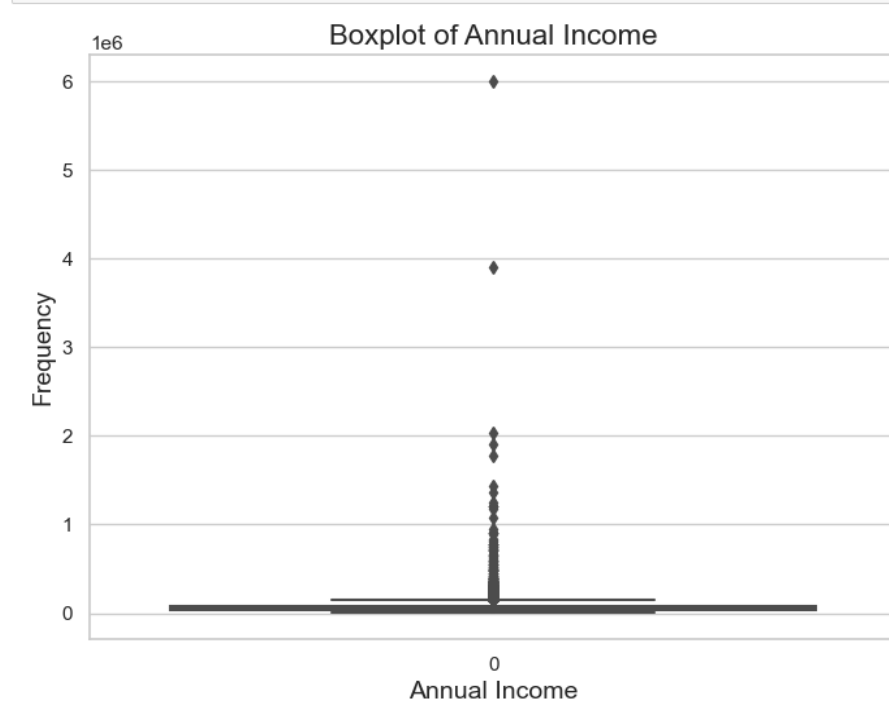
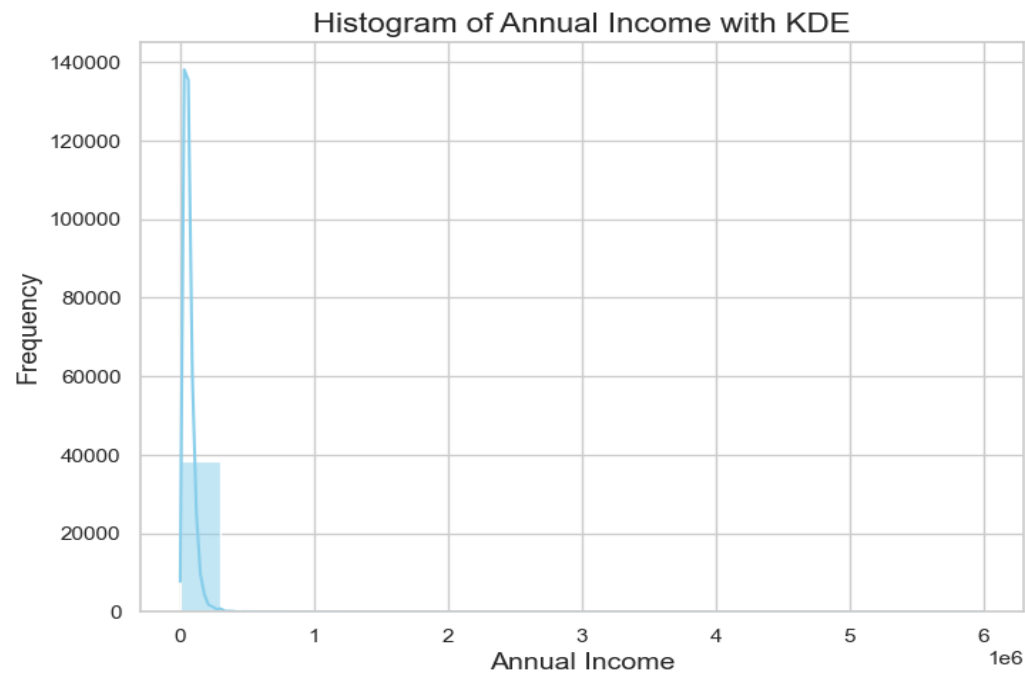
- Identifying and dealing with missing values in the dataset. Removing the columns with all Null values.
- Check for duplicate rows and remove them.
- Removing columns with unique values after analysis.
- Removing fields which are not useful in terms of the problem statement
- Remove fields not adding valuable meaning to the problem statement wrt loan domain.
- Outliers can significantly impact the results of statistical analysis. Identifying and deciding whether to remove, transform, or keep outliers is essential.

Data Analysis and Cleaning

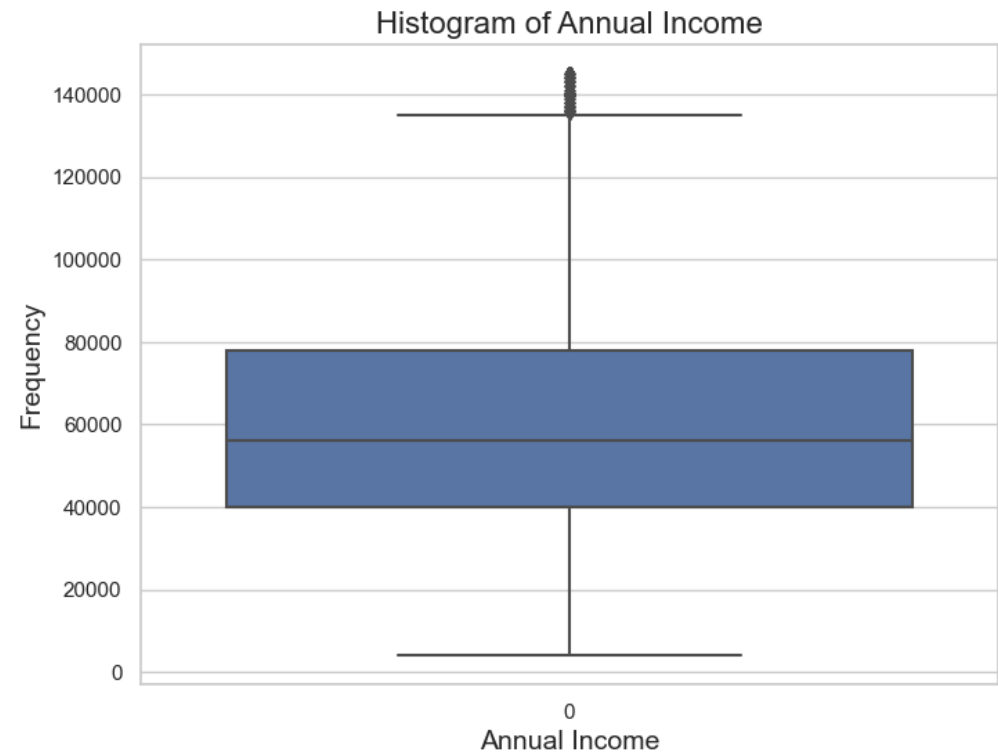
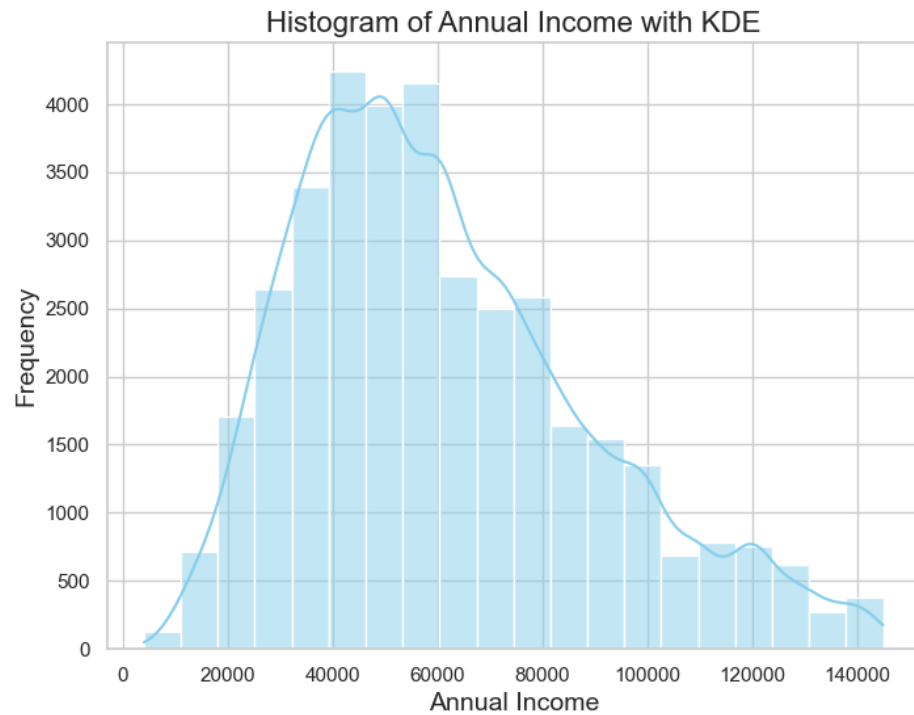
- Correcting Inconsistencies by identifying and rectifying inconsistencies in the data.
- Data Transformation by analysing and convert data into proper types eg int, float date etc
- If categorical variables have too many levels, they might be consolidated into broader categories for better analysis and interpretation.

Univariate Analysis

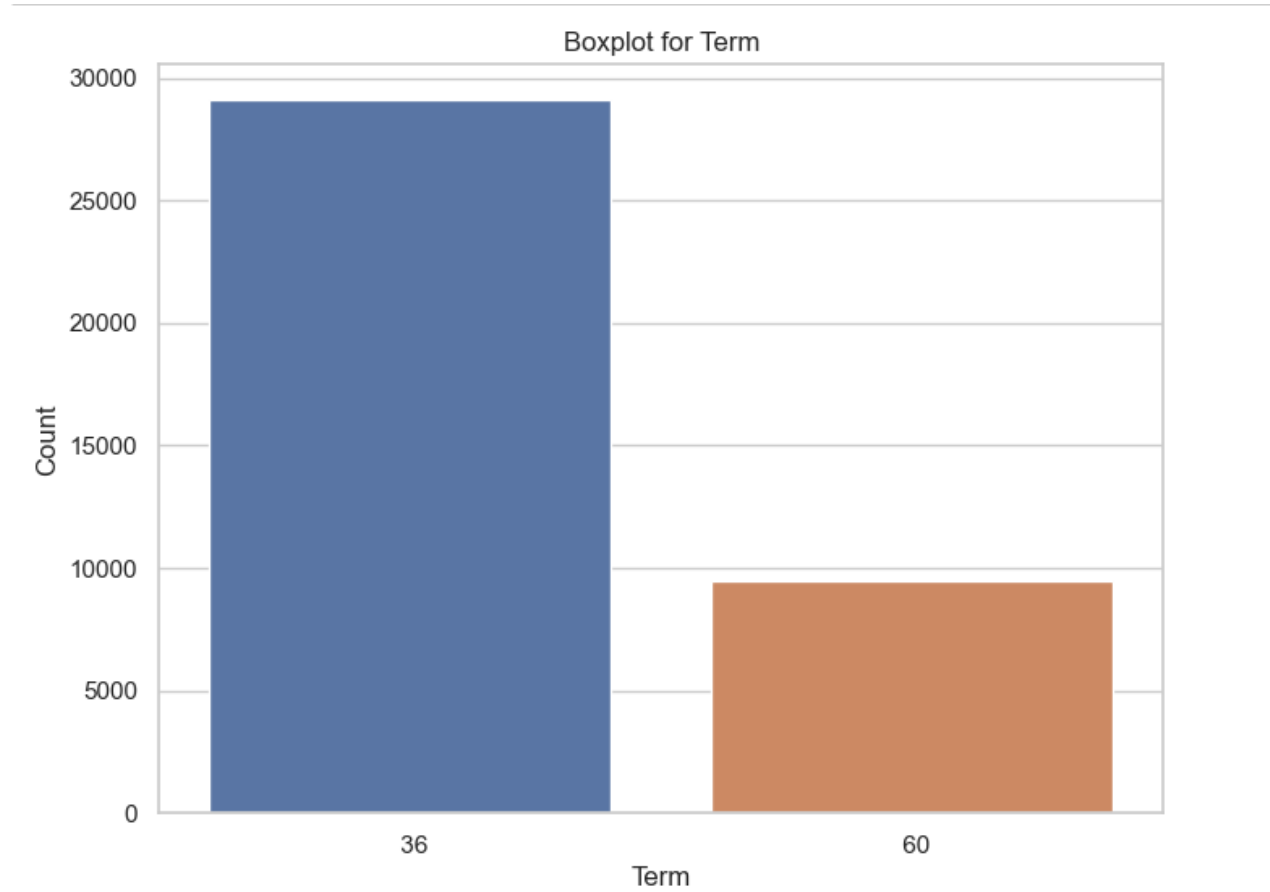
Observations : While plotting the annual income, outliers are observed.



Observation : Inter Quartile Range is used to remove the outliers. The distribution is normal.

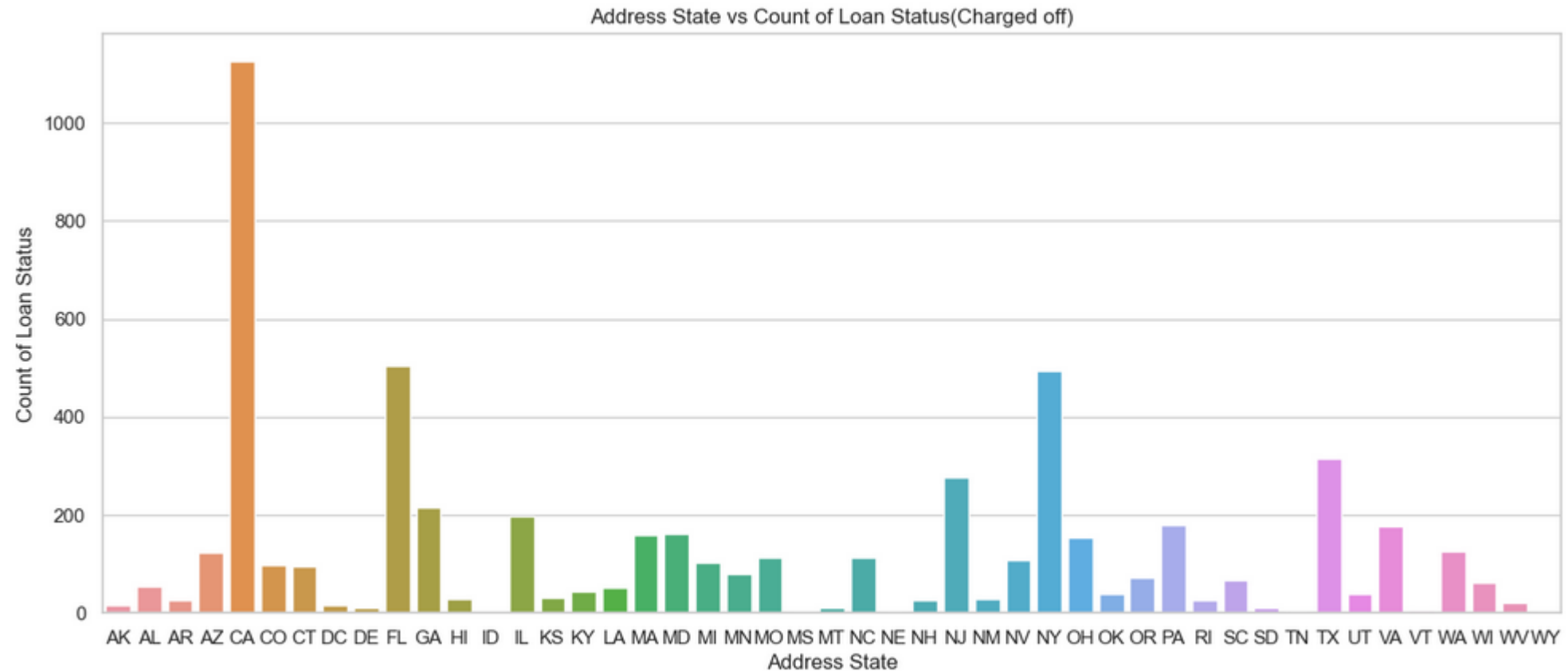


Observations: Applicants take more loans for 36 months

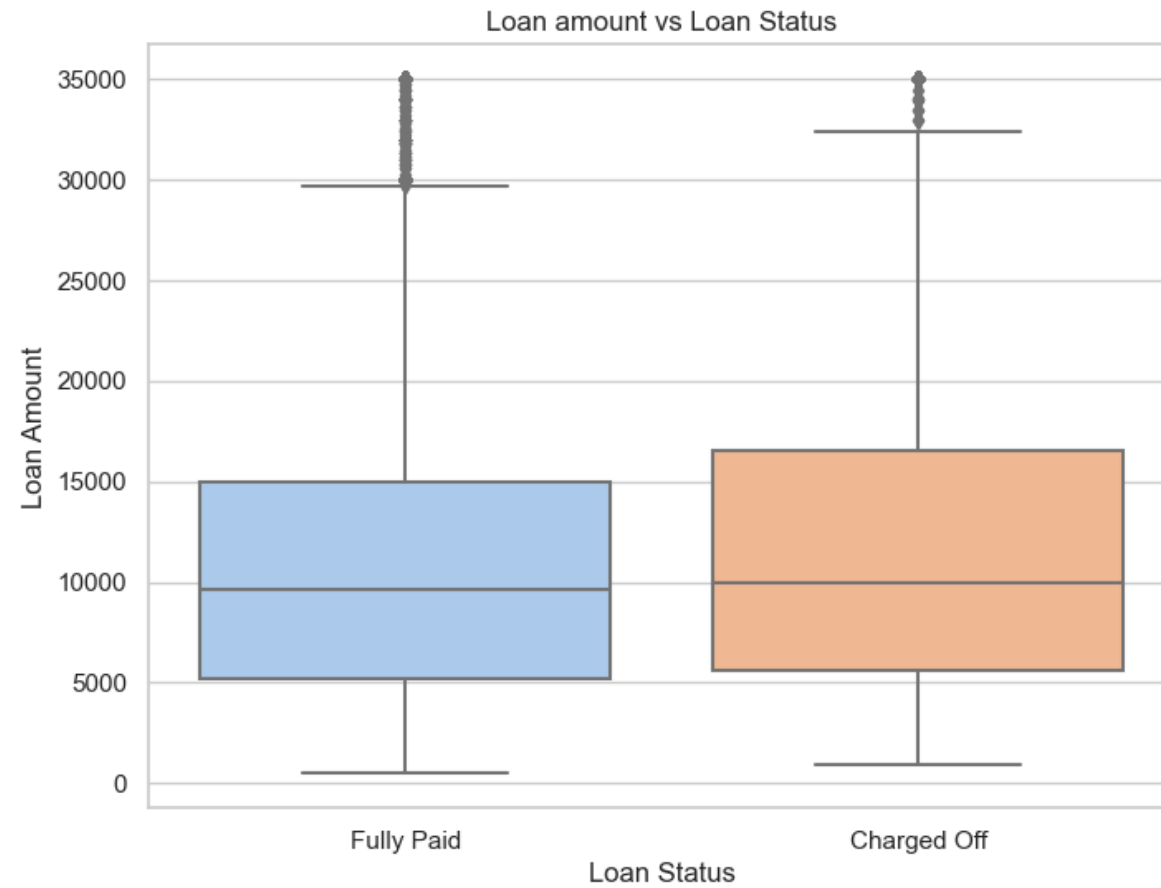


Bivariate Analysis

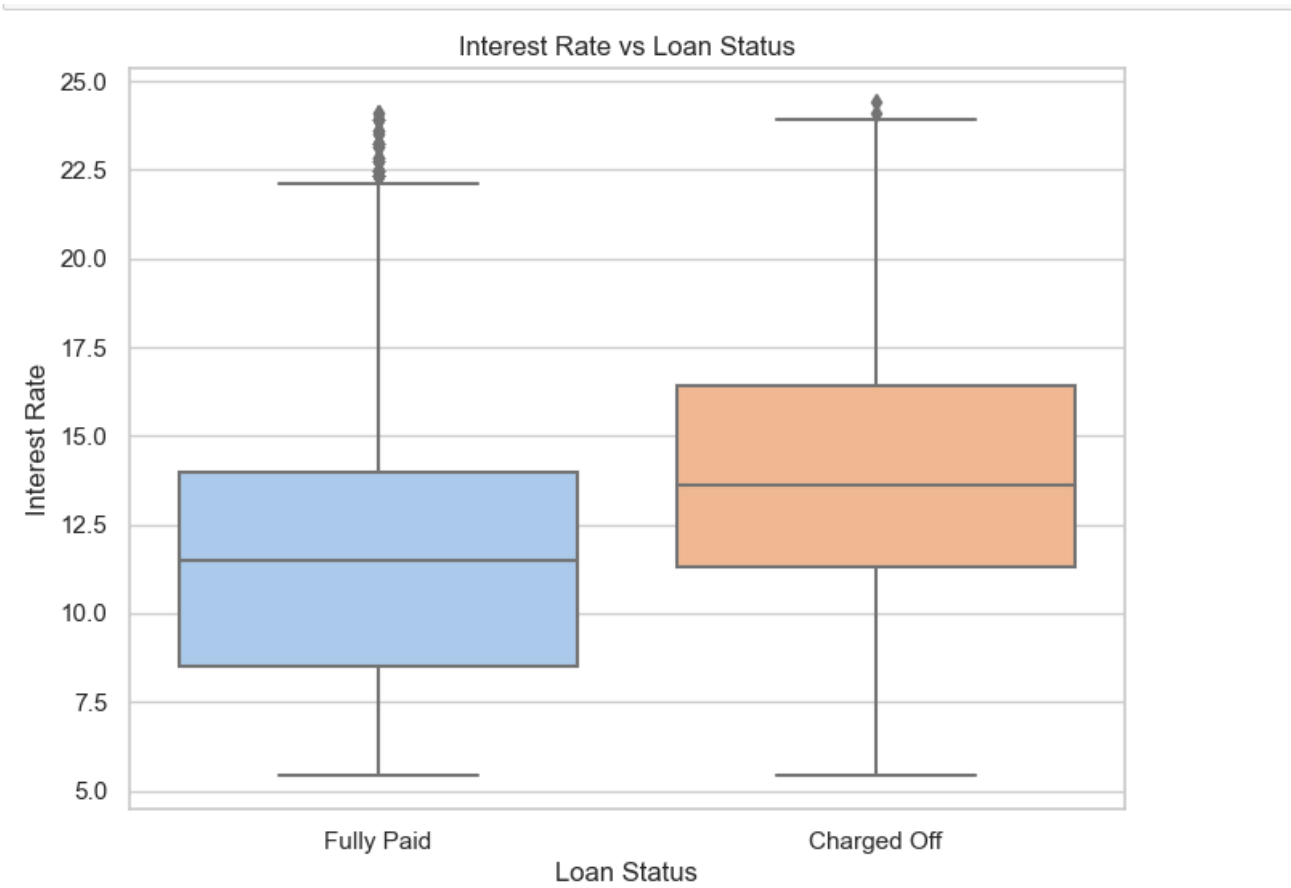
Observation: Maximum number of defaulters are from the state of CA



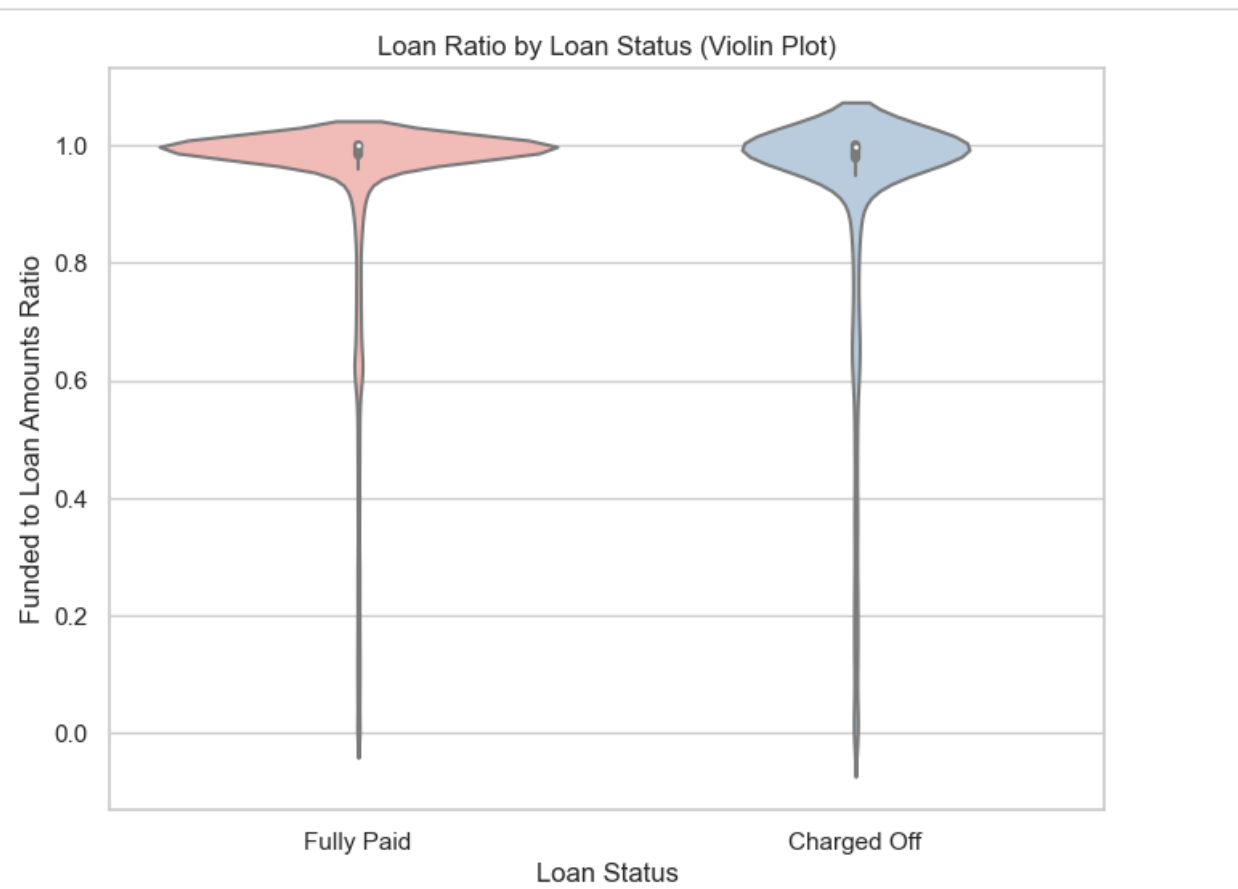
Observation: Charged off loans have higher loan amounts than fully paid loans.



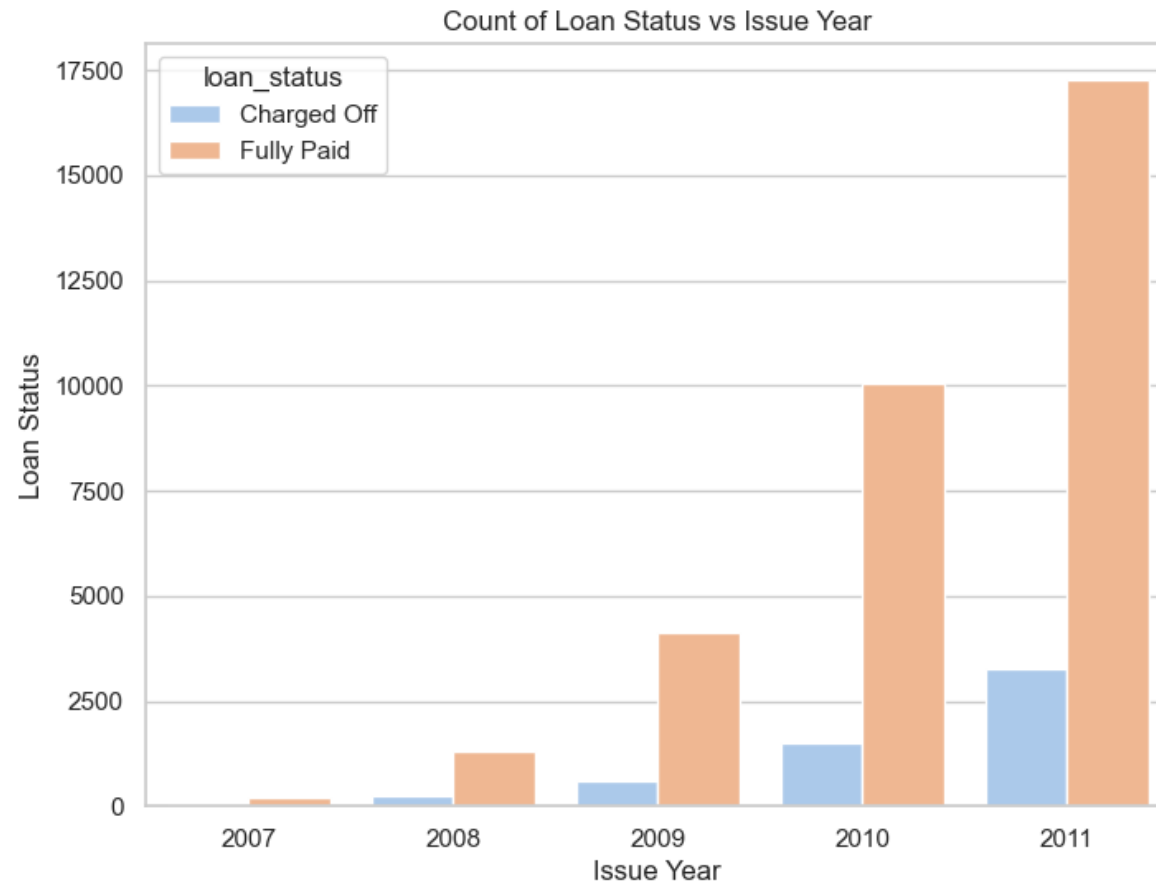
Observation: Higher rate of interest causes more defaulters.



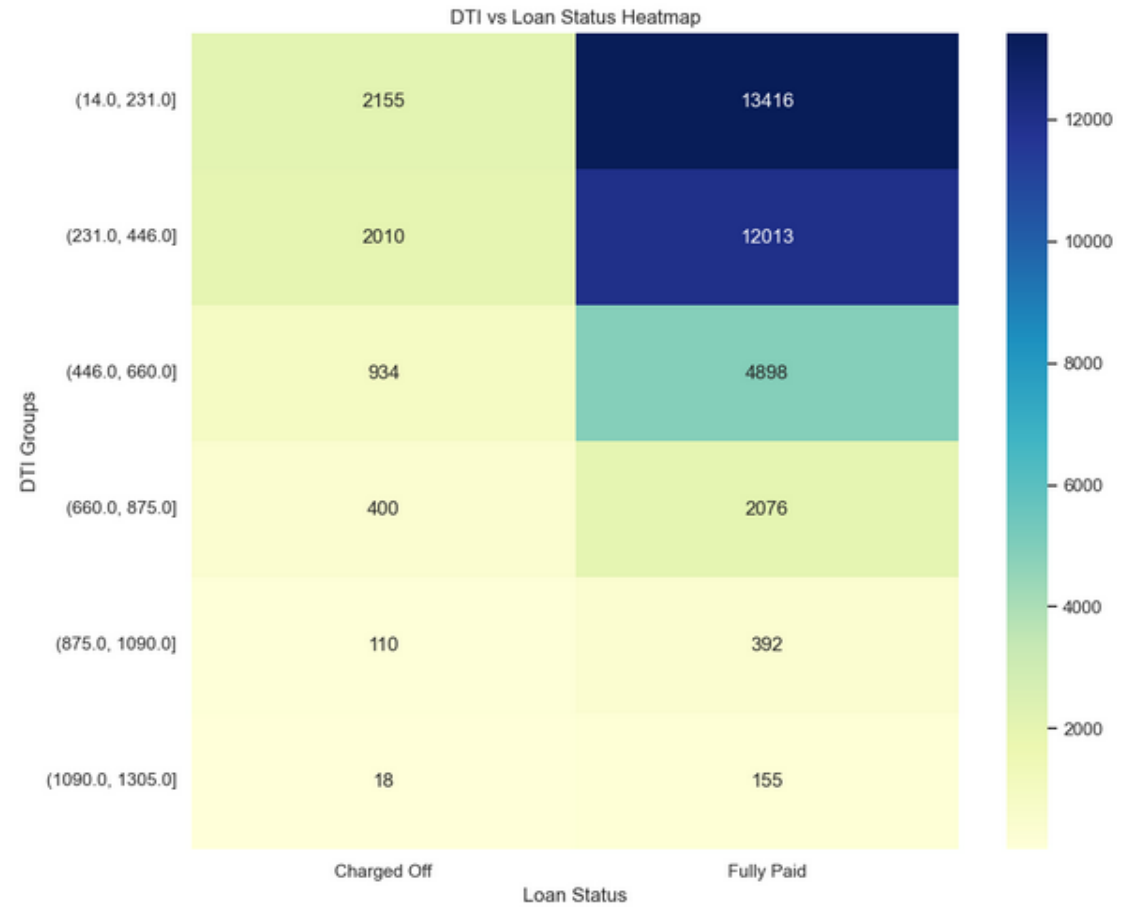
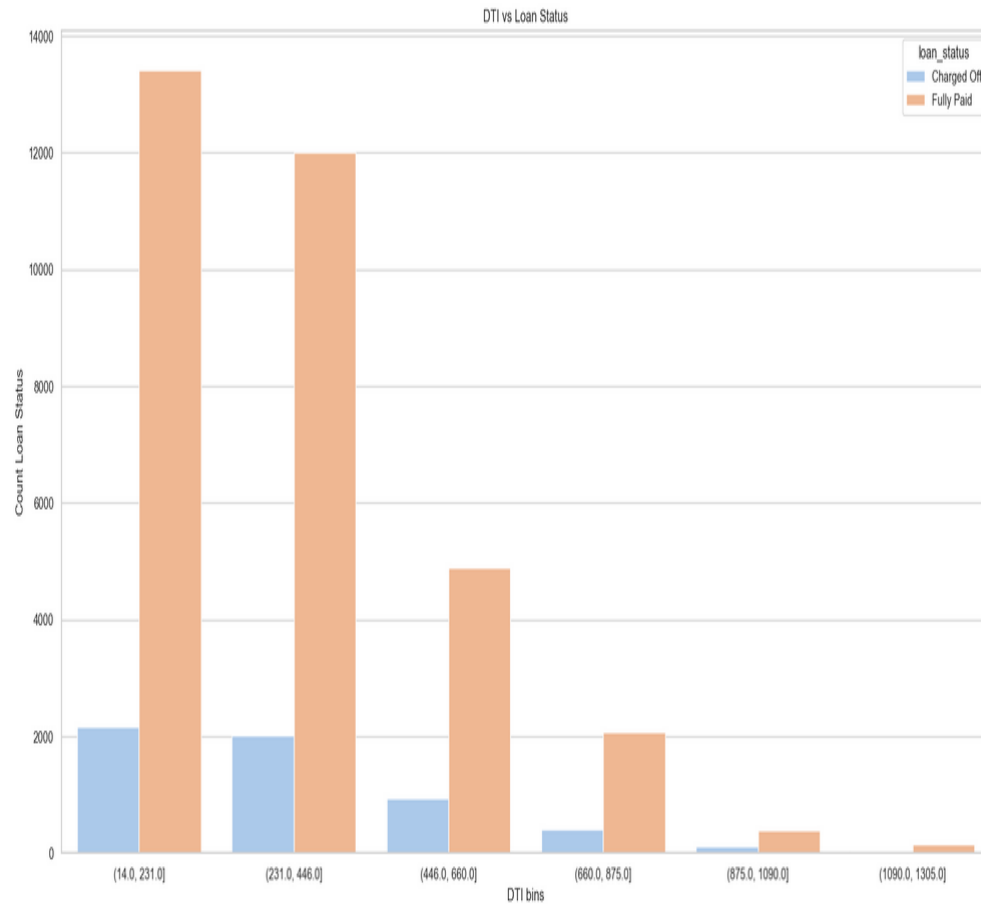
Observations: There is a higher probability of loan getting fully paid.



Observations: Even though loan applicants are increasing with the years, there is a much higher chance of the loan getting paid off.



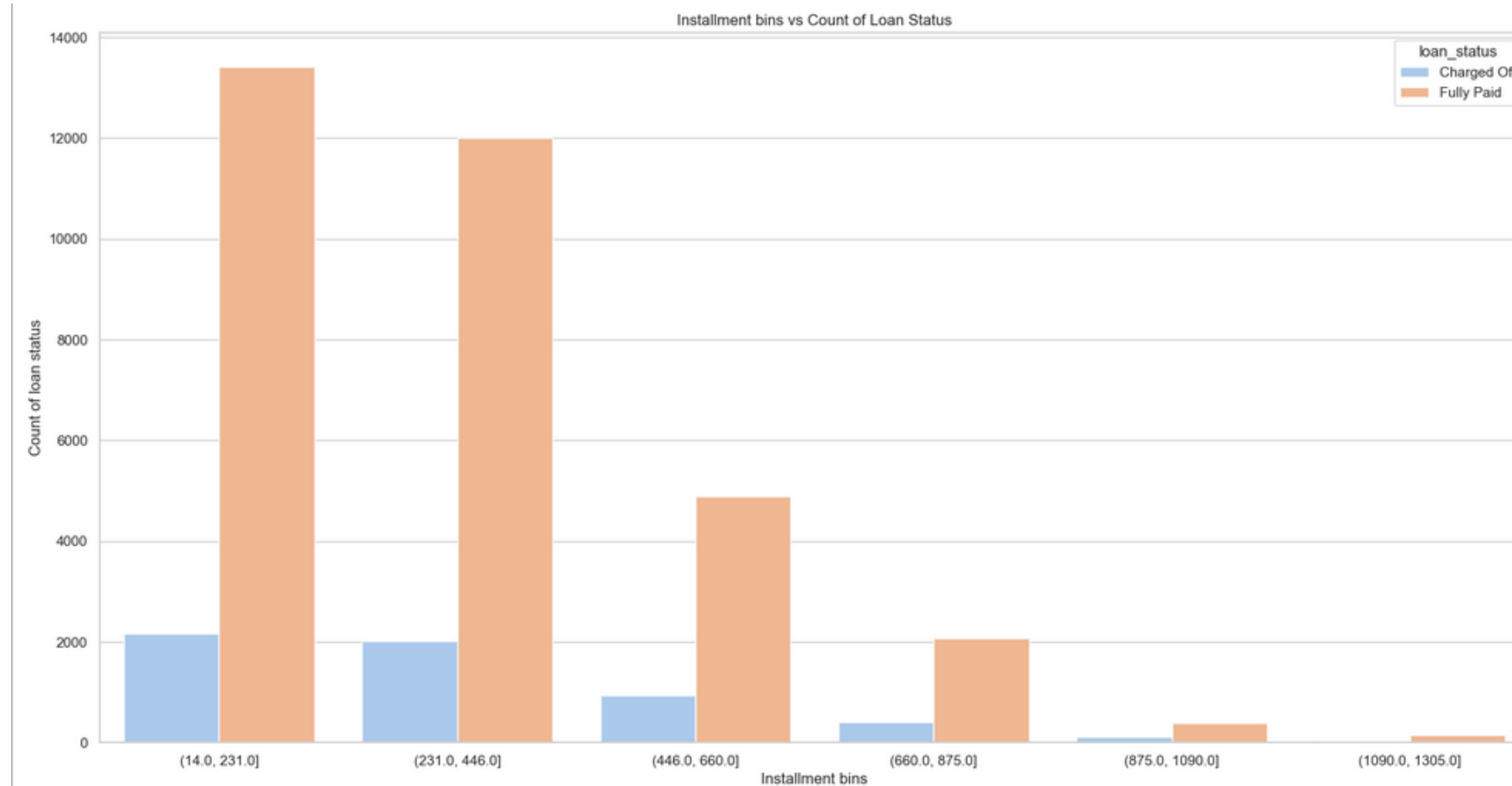
Observations: Lower DTI has significantly higher loan status of fully paid.



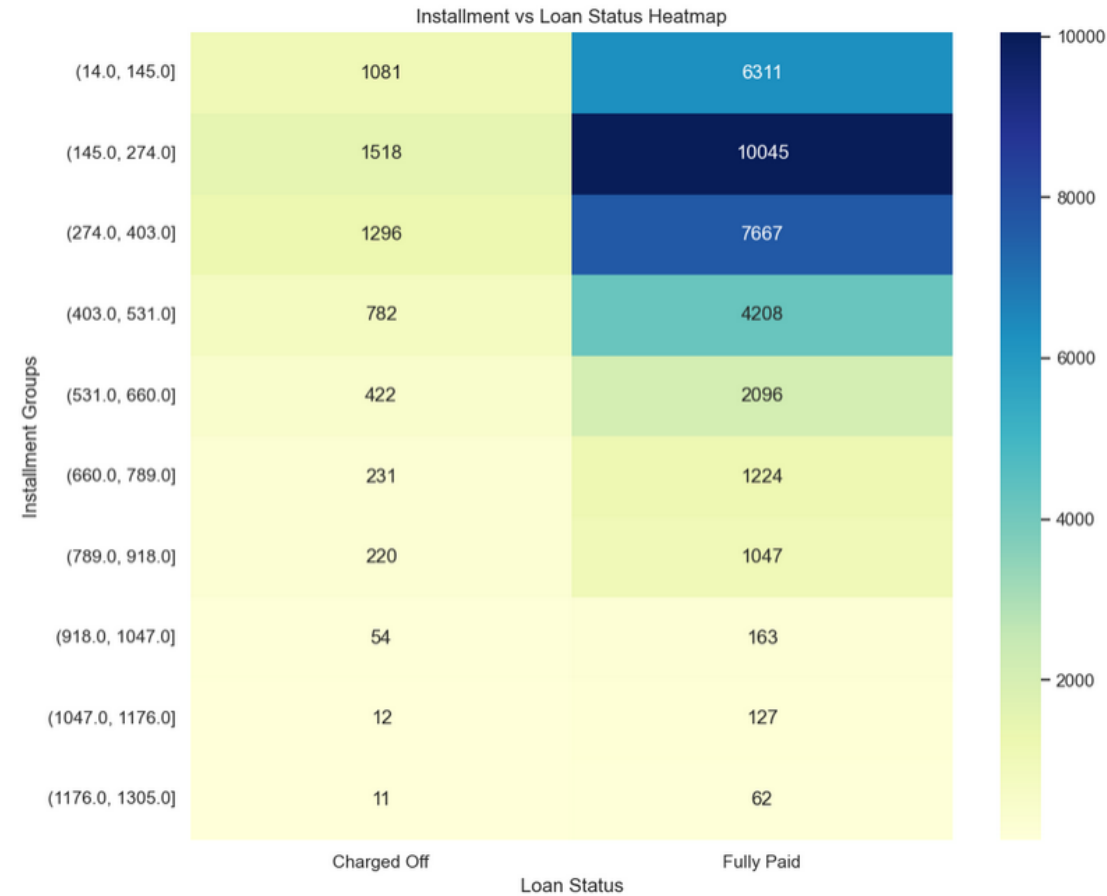
Observation:

If the installment is less, there is higher fully paid loans.

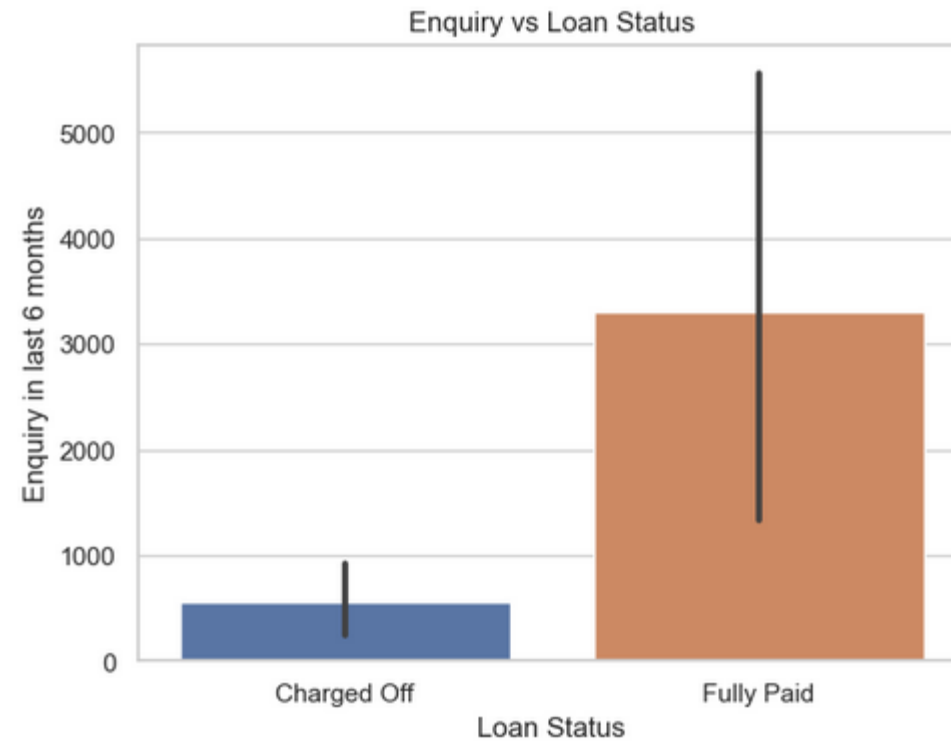
In each installment bin, it has been observed that fully paid loan status has a greater number of records wrt charged off status.



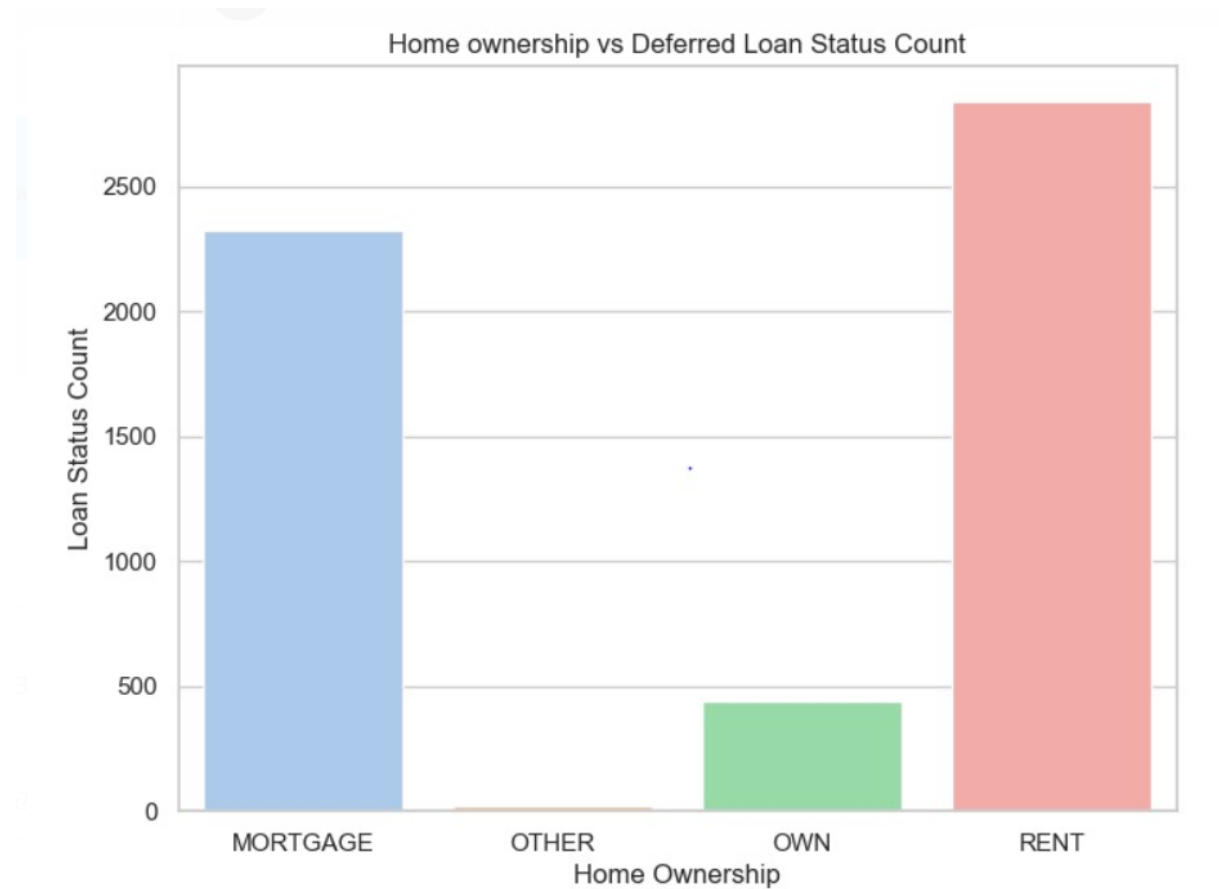
Observations: Less number of installment has more number of fully paid loan status records.



Observations: In the last 6 months it can be seen that People who fully paid the loan call more often to the bank then the people who charged off



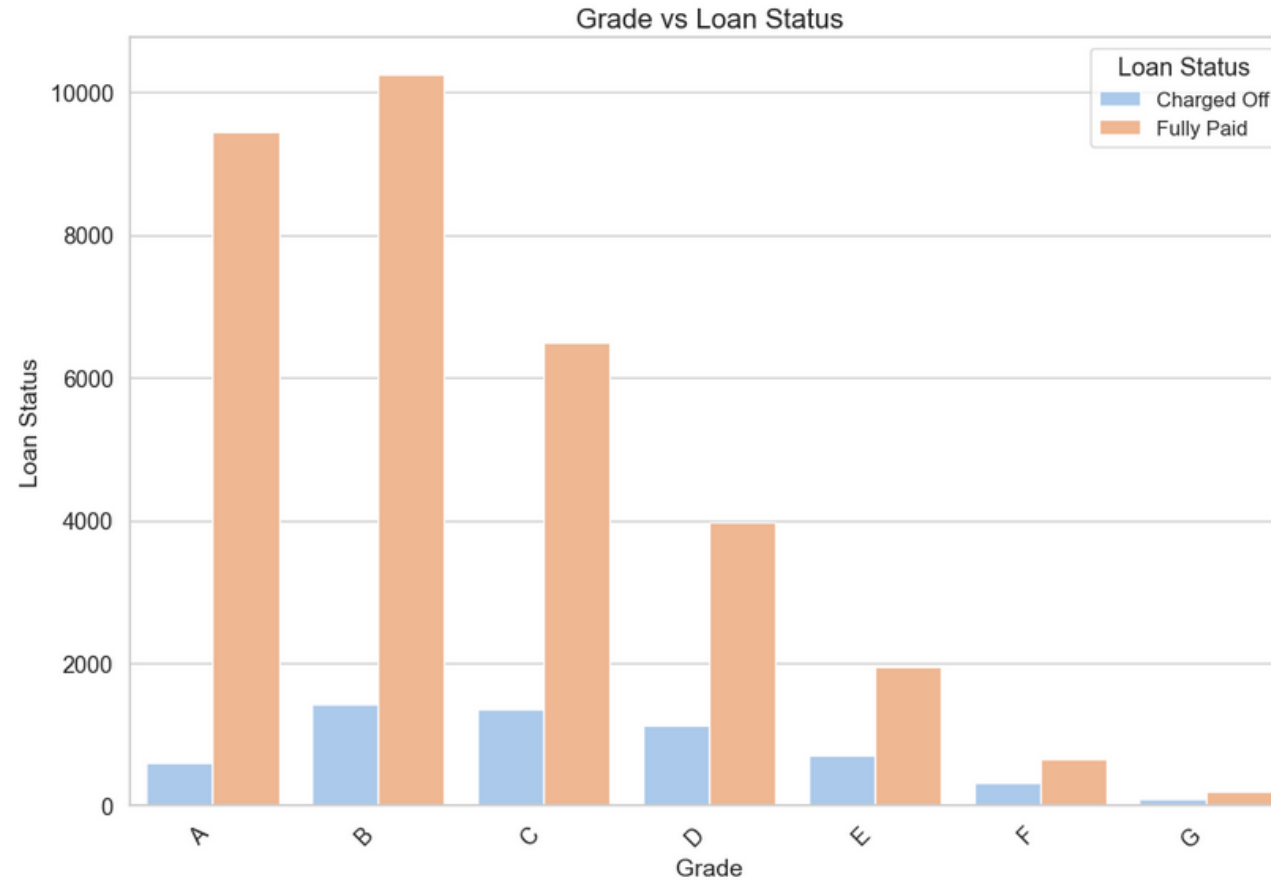
Observations: Applicants who rent would defer loans more.



Observations:

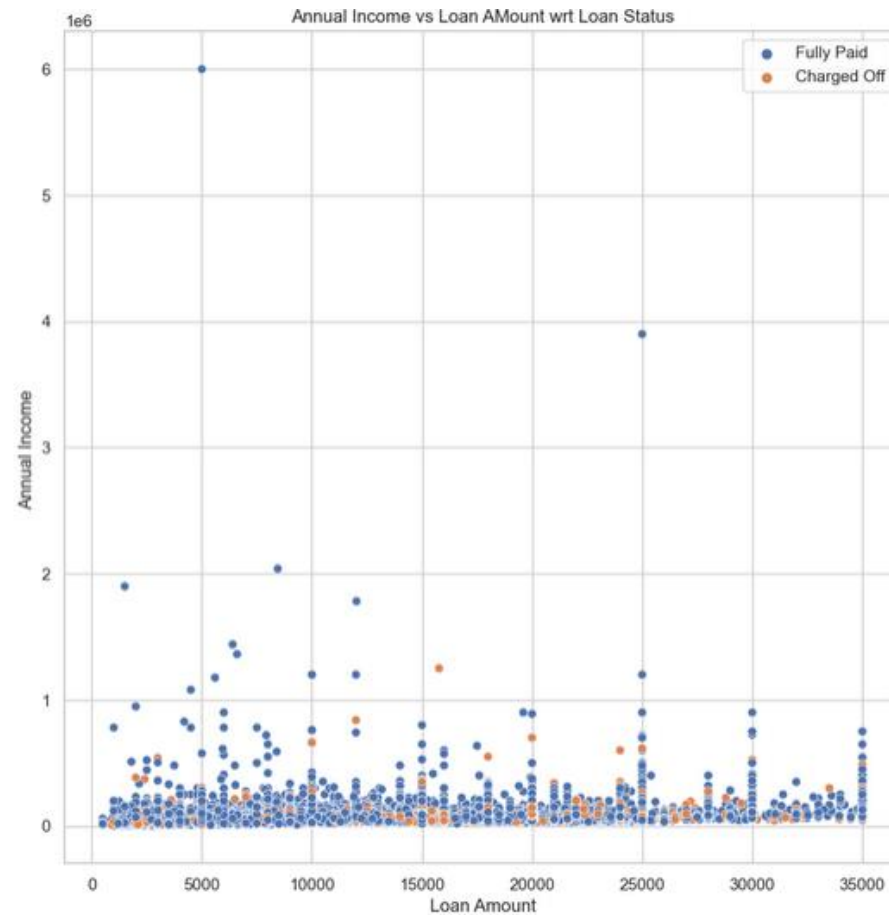
Applicants with grade A, B,C are most likely to pay off the loan.

Applicants with grade G is most likely to default the loan.



Multivariate Analysis

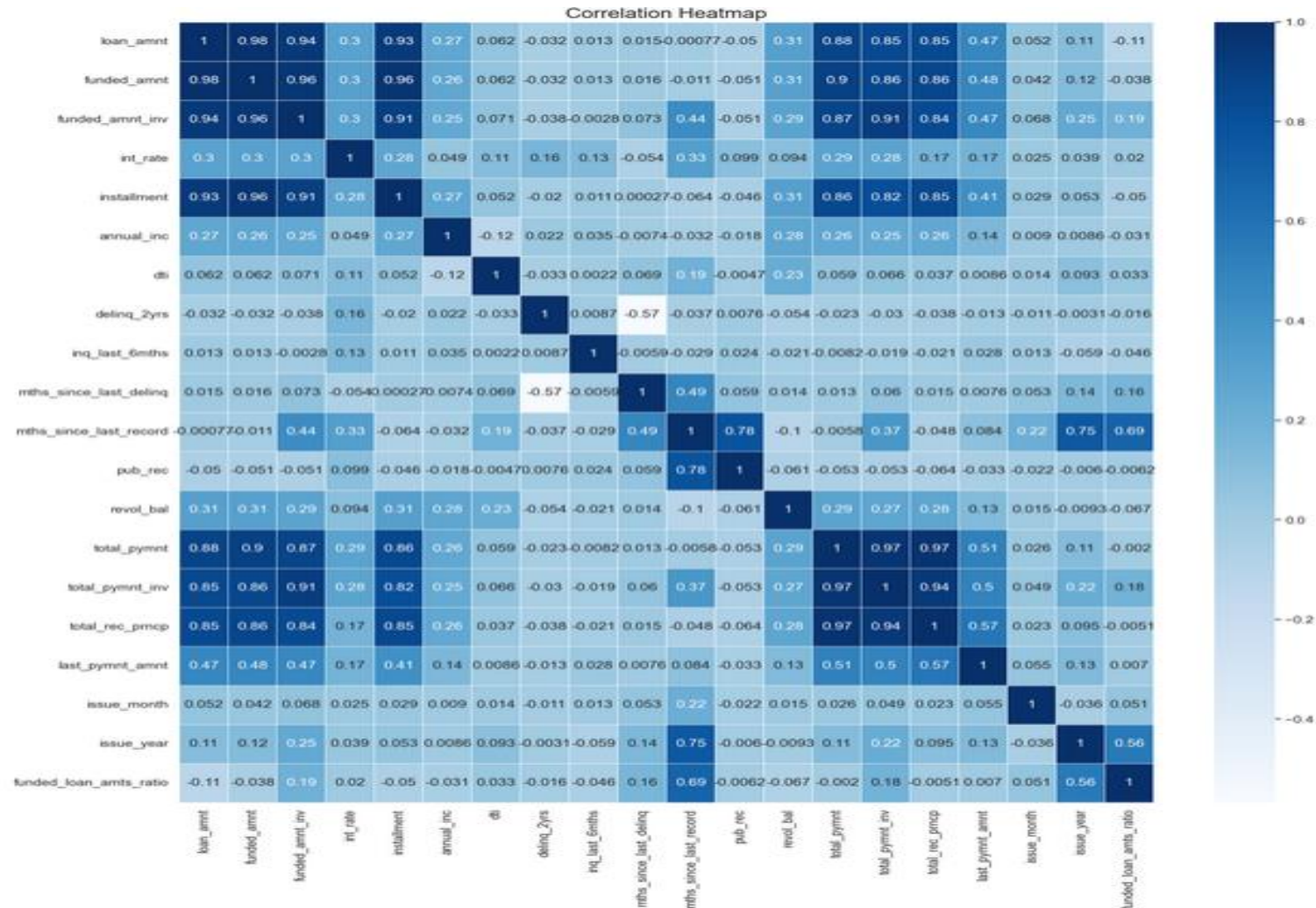
Observations: Applicants with more annual income has paid off their loans more.



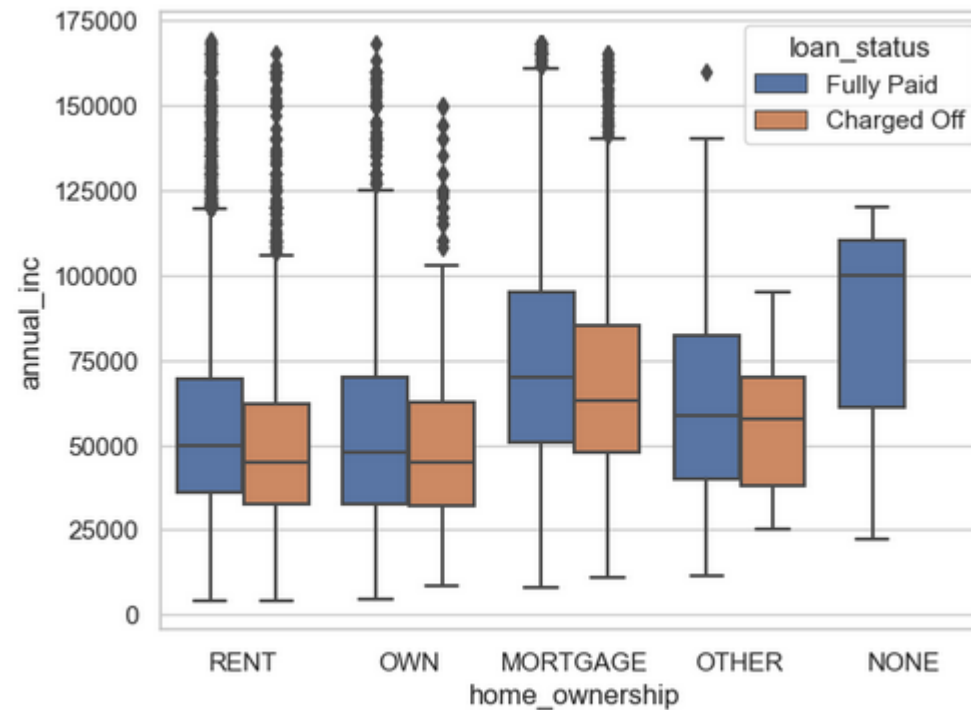
Observations:

There is a very high correlation between number of installment and total payments. Also observed in Funded amount and total payment.

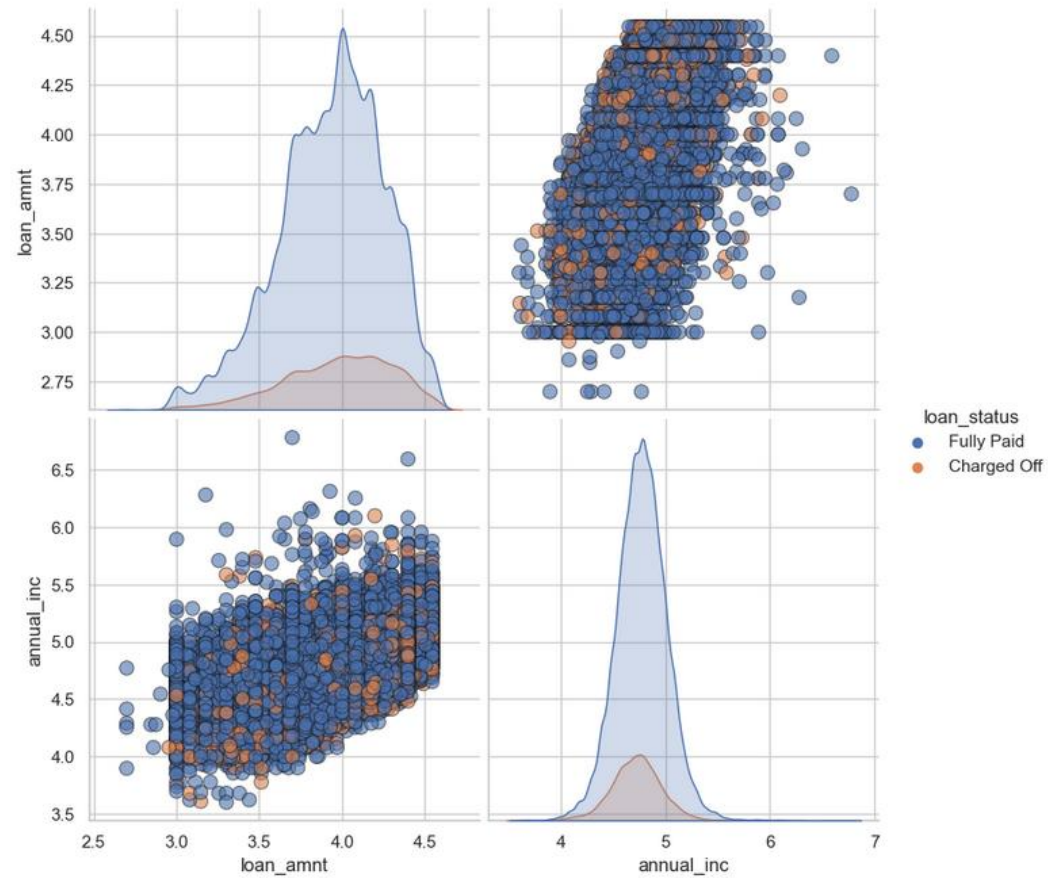
DTI and Annual income has negative correlation.



Observations: Applicants who rent and have less annual income & applications with higher annual income and mortgage are more likely to have fully paid loan status.



Observation: Applicants with more annual income have higher rate of paid off.



Summary

- Applicant with more Annual Income have paid off their loans fully.
- Applicant with more number of funded amount has paid off their loan fully.
- Applicant with home mortgage and high annual income have paid off their loans fully.
- Applicant with home ownership and less income have max number of fully paid off loans .
- Applicants with grade A, B,C are most likely to pay off the loan.
- Applicants with grade G is most likely to default the loan.
- Less number of installment has more number of fully paid loan status records.
- In each installment bin, it has been observed that fully paid loan status has a greater number of records wrt charged off status.
- Maximum number of defaulters are from the state of CA