



LENDING CLUB CASE STUDY

BY:
AMRITA SINGH
MADHURIMA MUKHERJEE

Objective:

- The objective of analysis is to use the information about past loan applicants and find whether they are likely to 'default' or not based on driver variables so that loan lending organisations can use this analysis to assess risk associated with new loan applicants.

Steps used for analysis:

Data cleaning:

Removed the null valued columns and unnecessary variables, checked the null value percentage and removed the respective rows.

Data understanding:

Worked with the Data Dictionary and understood all the columns and their domain specific uses

Univariate/ Segmented Univariate analysis:

Plotted distribution after analysing every column and driver variables.

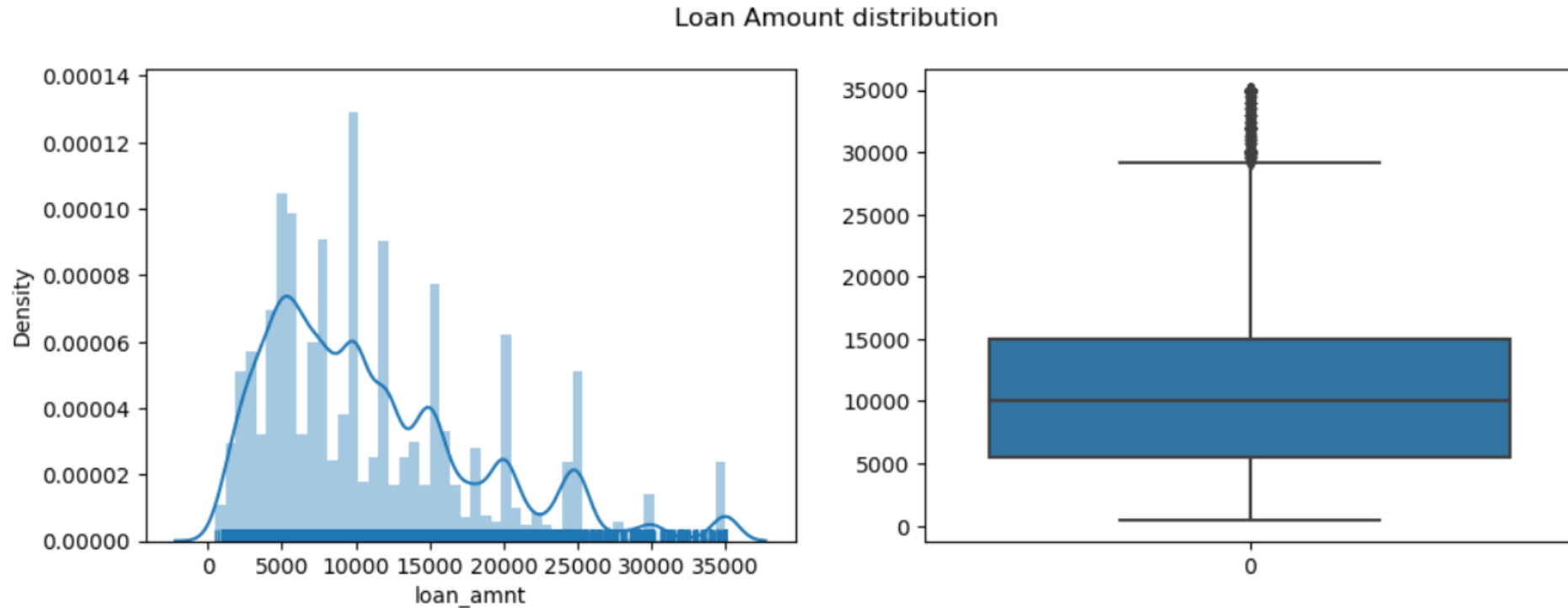
Bivariate/ Multivariate analysis:

Performed bivariate and multivariate analysis to understand how different variables interact with each other.

Recommendation:

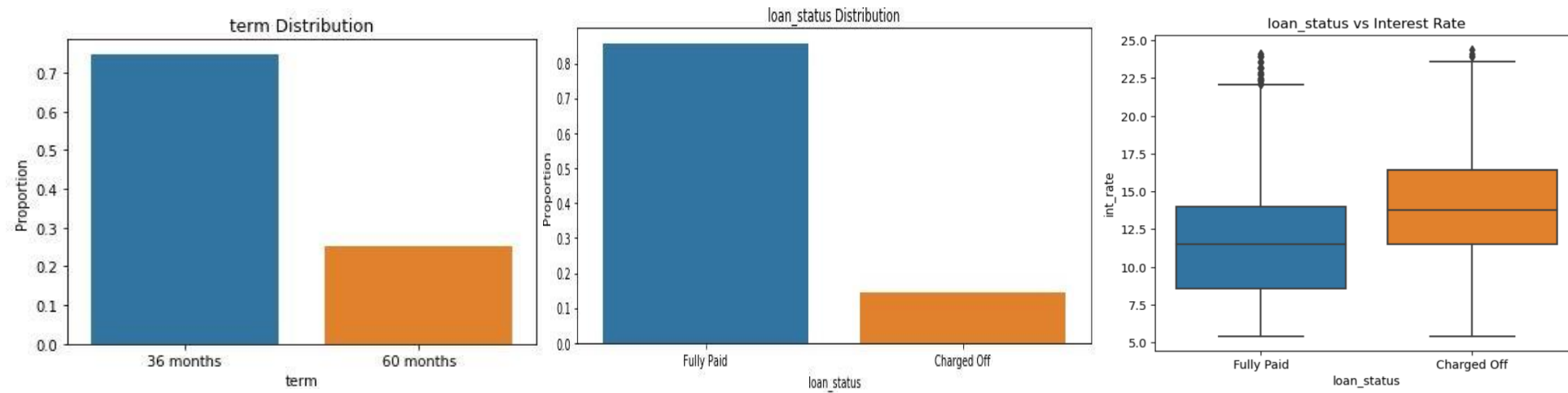
Recommendations for reducing the loss of business by analysing and detecting columns which contribute to loan defaulters.

Distribution curves based on analysis:



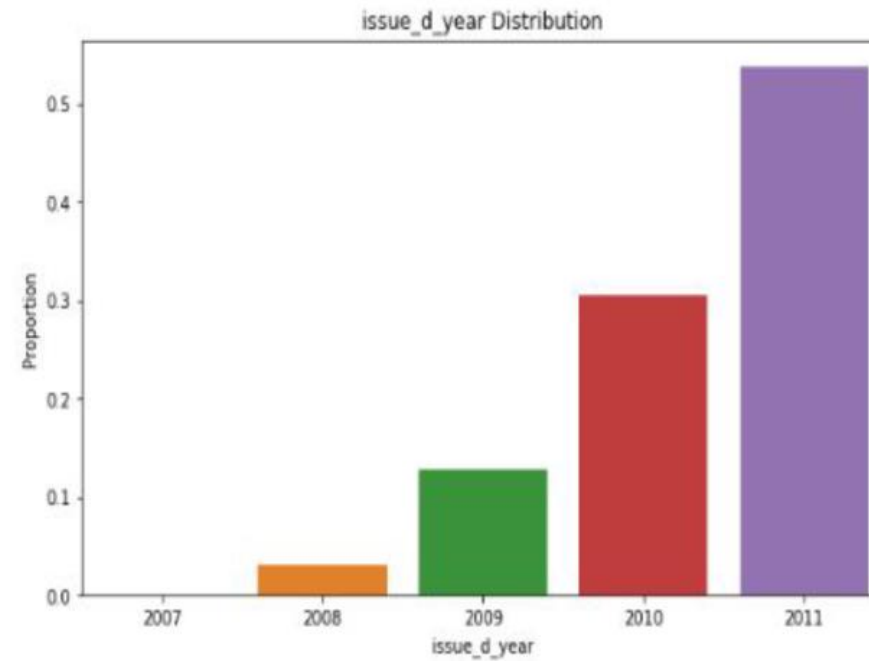
- More number of people took loan amount of 10000 while very few people took more than 30000 loan amount.

Distribution curves based on analysis:



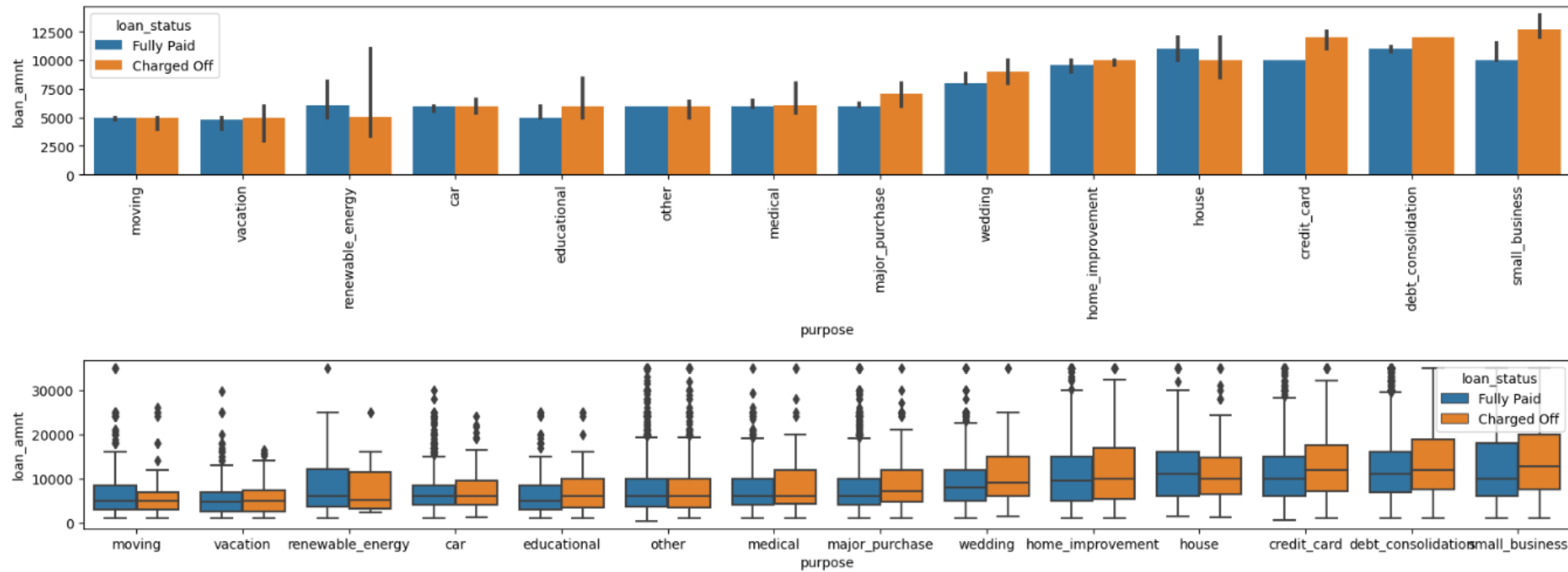
- There are only two loan terms 36 and 60 months. Around 75% borrowers took loans with 36 months term.
- The charged off borrowers are around 15% and fully paid is around 85% in the given data set.
- The greater the interest rate, higher the chances of defaulting loan

Distribution curves based on analysis:



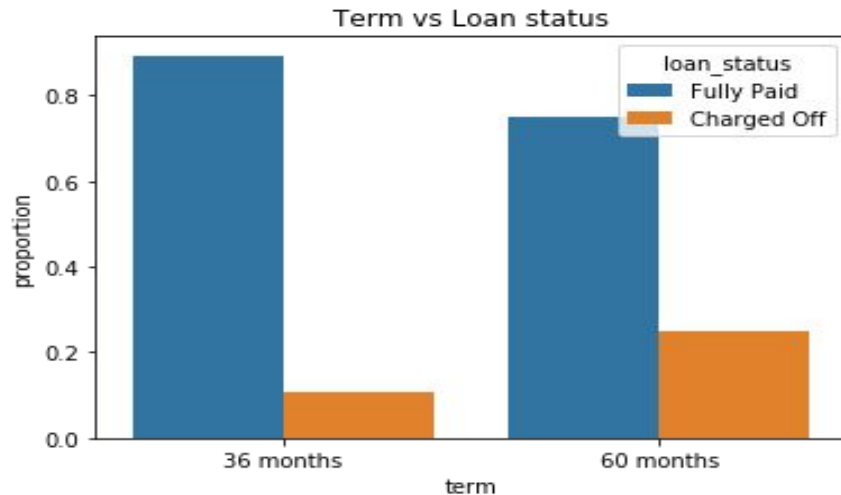
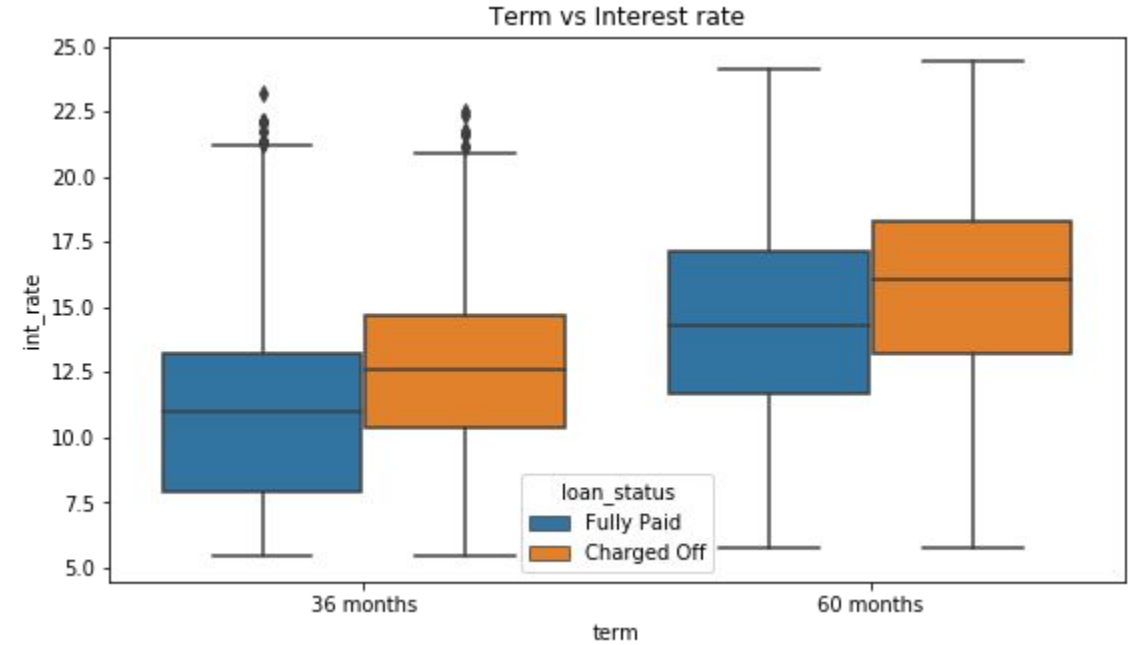
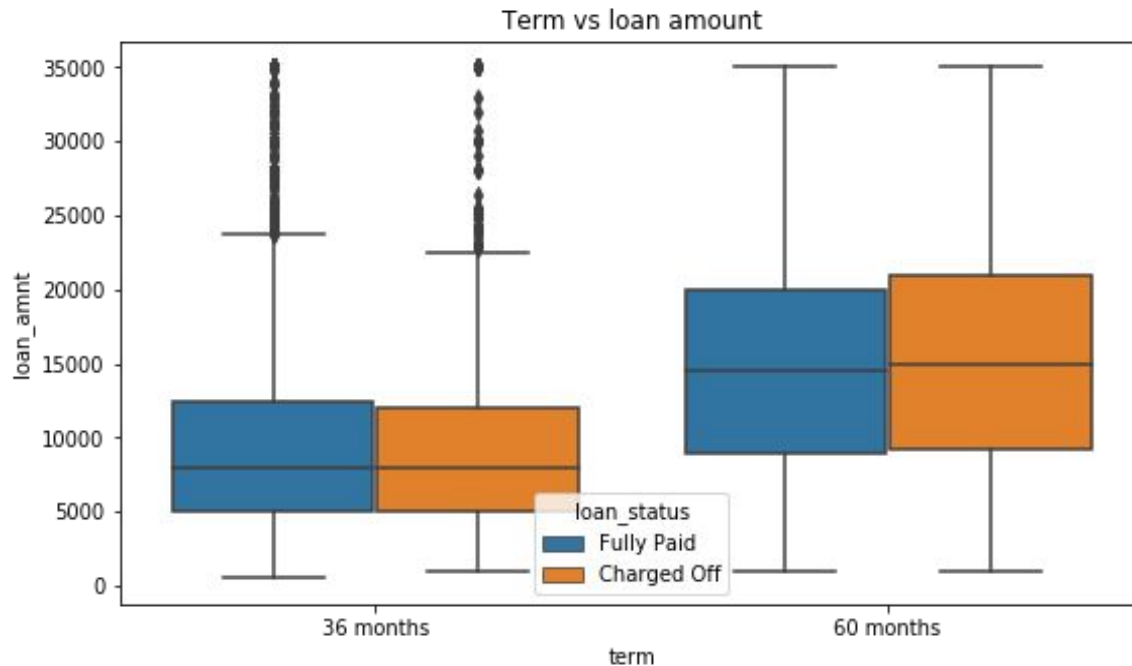
- The number of loans issued by lending club has doubled every year.

Distribution curves based on analysis:



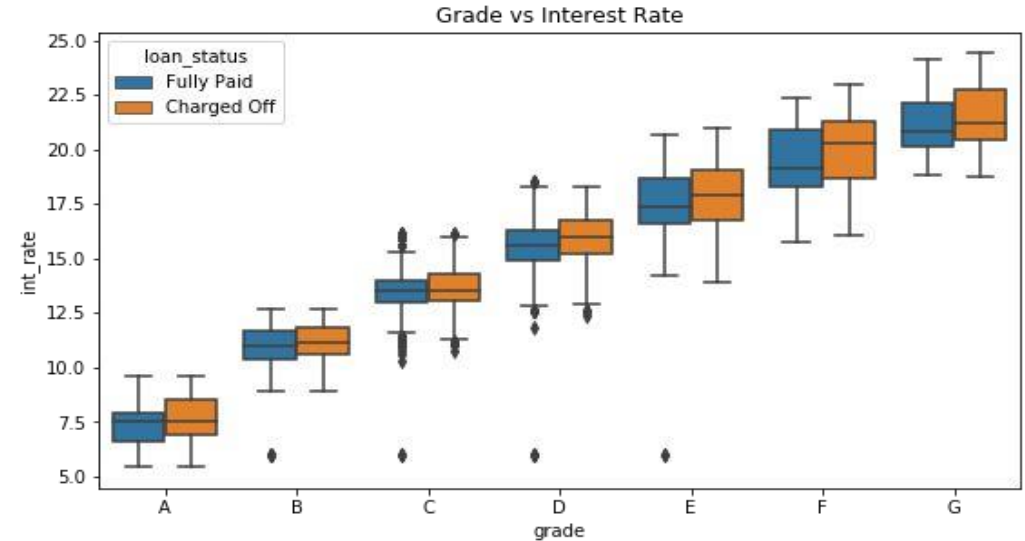
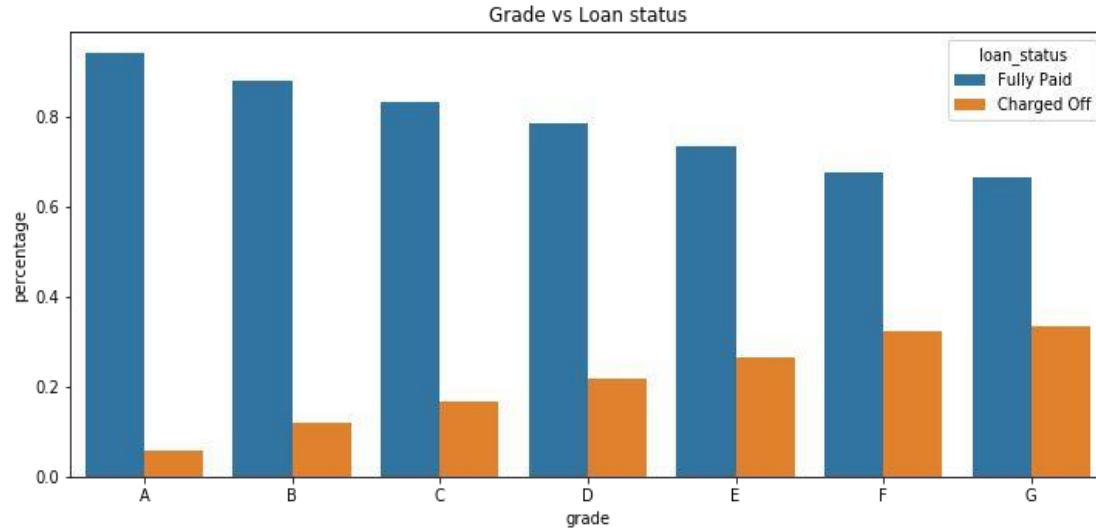
➤ Small Businesses have more defaults when the loan amount is high.

Distribution curves based on analysis:



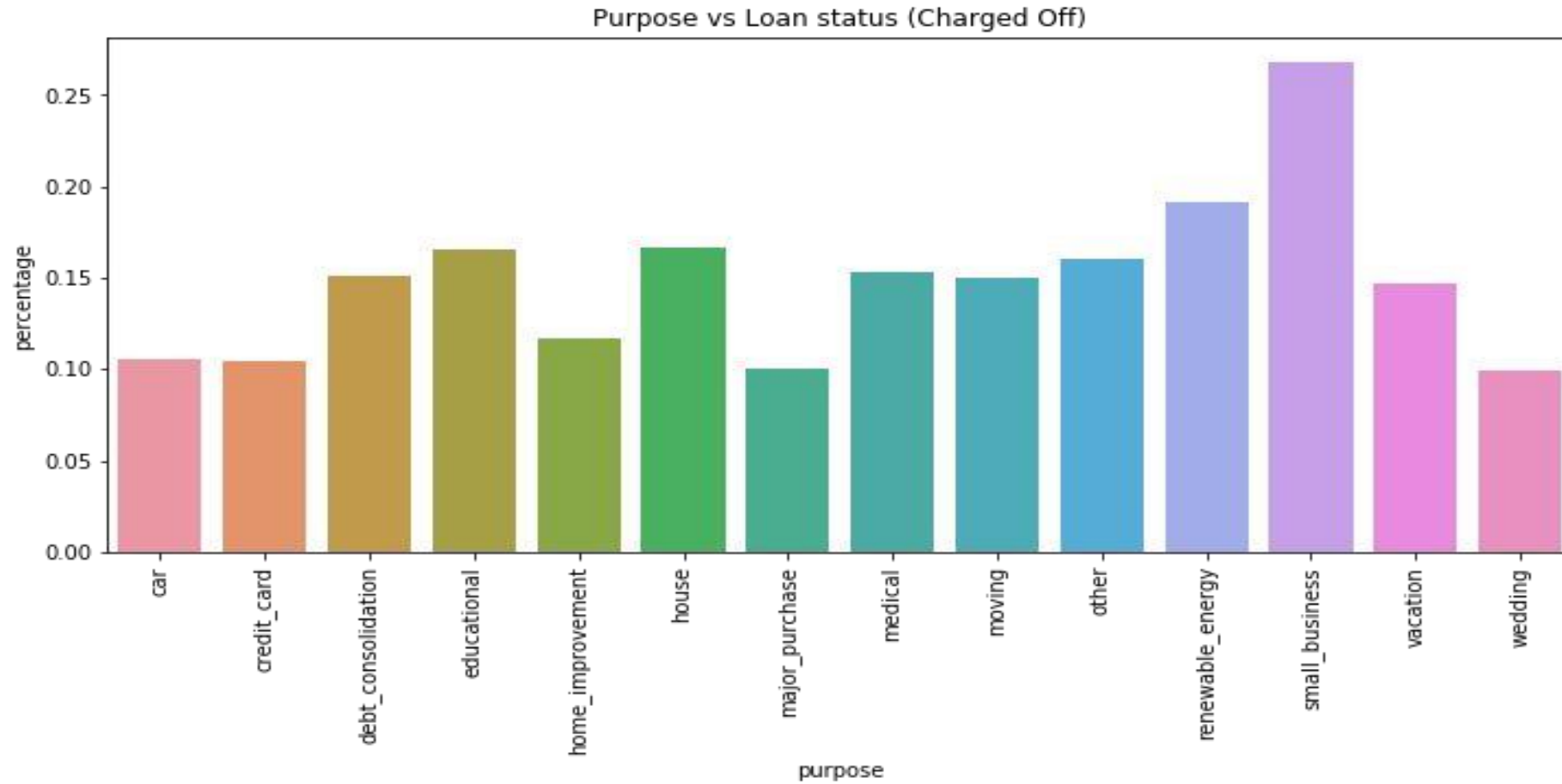
- The default rate is high in 60 months tenure because most people took high loan amount with high interest rate in it and they faced difficulties in returning the sum to bank.

Distribution curves based on analysis:



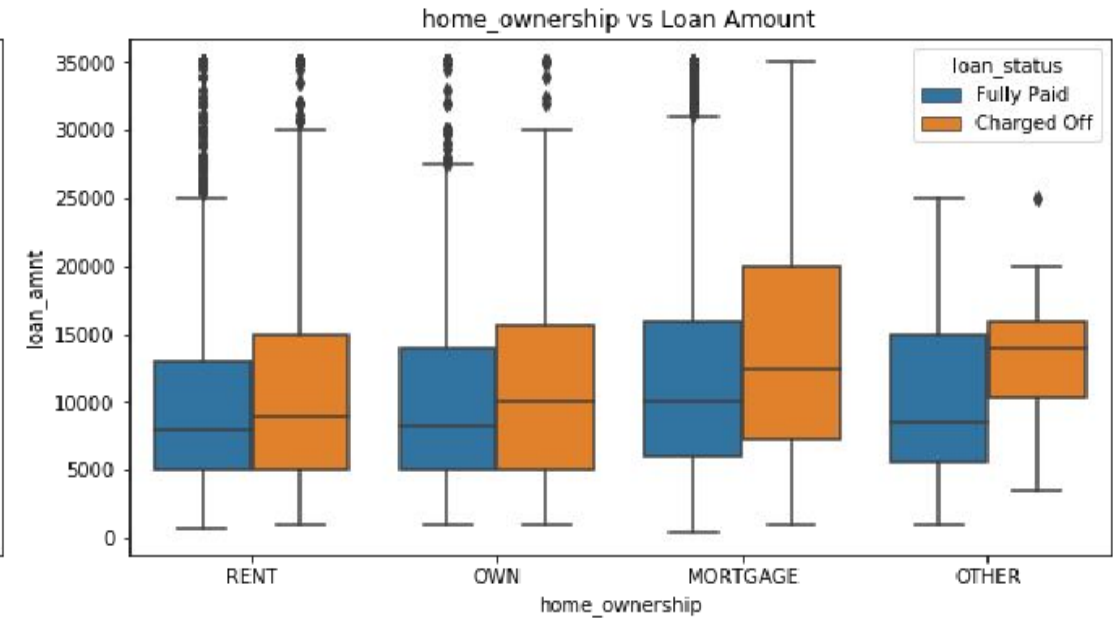
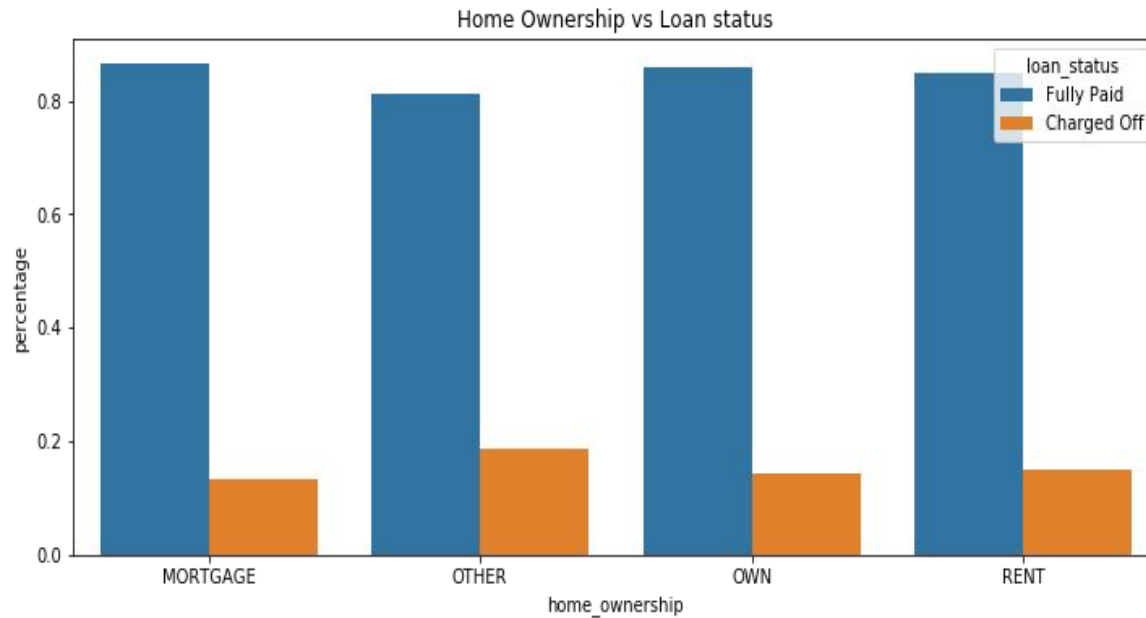
- Grades are very good category to tell the borrower probability of defaulting the loan.
- The Lower grades(E,F,G) have higher chances of defaulting the loan than Higher ones(A,B)
- Also the Lower grades are getting loans for higher interest rates which might be the cause for loan default.

Distribution curves based on analysis:



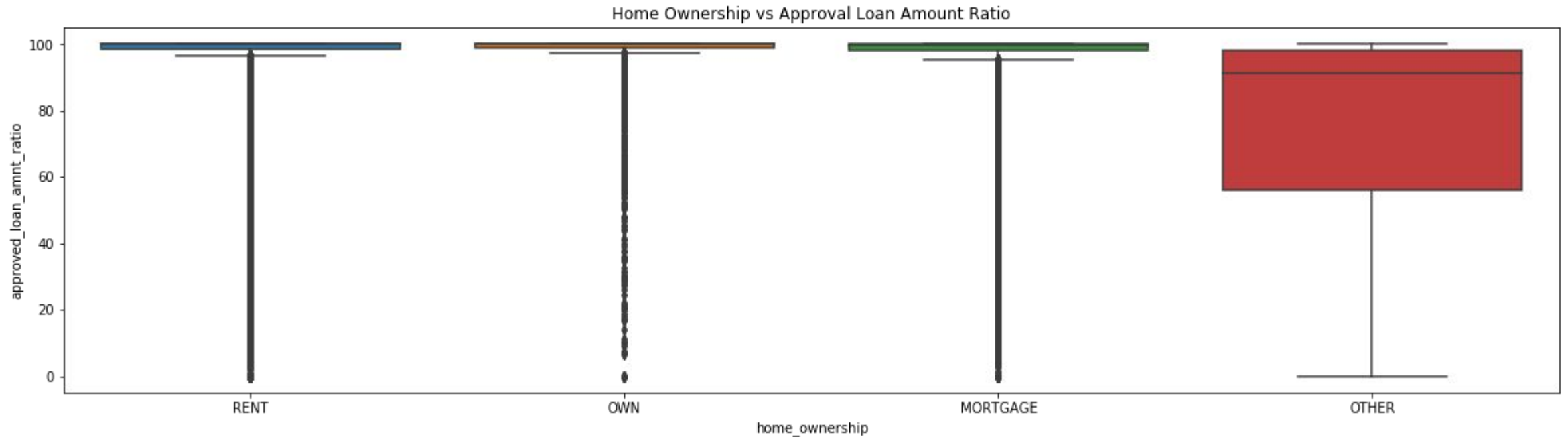
➤ Borrower's who took loans for small business purpose have defaulted more.

Distribution curves based on analysis:



- There is around 20% chance of loan default in each home ownership category.
- From the 2nd plot we can see the people with higher loan amounts in **mortgage** home ownership has high default rate than others.

Distribution curves based on analysis:



- Approved loan (Funded Amount by investor) is less than than the requested loanamount by borrowers for Other Home ownership category.

Conclusions and recommendations:

- Lending club should reduce the high interest loans for 60 months tenure, they are prone to loan default.
- Lending club should stop/reduce issuing the loans to small businesses as they are likely to default more.
- Lending club should examine more information from borrowers before issuing loans to Low grade (G to A) as it is a good metric for finding defaulters.
- Lending club should stop giving loans to borrowers with mortgage home ownership when loan amount requested is more than 12000 as they are likely to default approved loans after taking higher amount of loans.
- Lending club should make sure there are no public derogatory records since they are more prone to bankruptcy for borrower.