LENDING CLUB CASE STUDY

Yash Joshi & Ravi Babu Arepalli

OBJECTIVE

The objective of case study is to apply our learnings of EDA on Loan data (a real-world problem) and present our key observations from the Univariate and Bivariate analysis

PROBLEM STATEMENT

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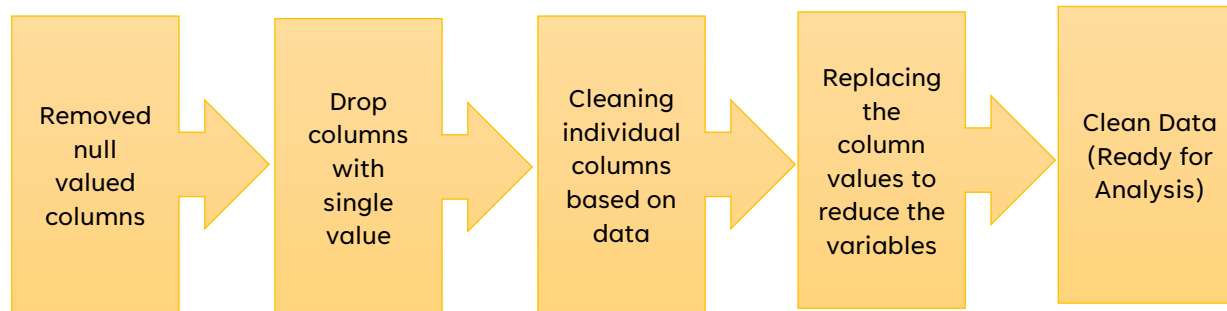
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

- If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company

The data given below contains information about past loan applicants and whether they 'defaulted' or not. 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 amount of loan, lending (to risky applicants) at a higher interest rate, etc.

DATA UNDERSTANDING AND DATA CLEANING

As we received the load data for case study.

We started with understanding the data variable in different columns and importance of each column with the help of Data_Dictionary provided with Data.

Removed null valued columns → Drop columns with single value → Cleaning individual columns based on data → Replacing the column values to reduce the variables → Clean Data (Ready for Analysis)

COLUMNS REMOVED

- Remove several columns in the data, as they have values as NA, which are not useful for analysis

- We have also removed multiple columns that can be used to uniquely identify the row data , but since there are multiples so we have worked with id.

- We have not used the description [desc]. column in our analysis.

- we have also not used customer behavioral variables.

- We dropped the columns that have missing values of more than 70%

MANIPULATIONS

- We saw that int_rate is not coming correctly as it should be float but it was object

- On further analysis for 10 records, we saw that there is a % sign with the data so we have manipulated the column and stripped the % from the column.

- We have also merged None and Other together as difference was to less.
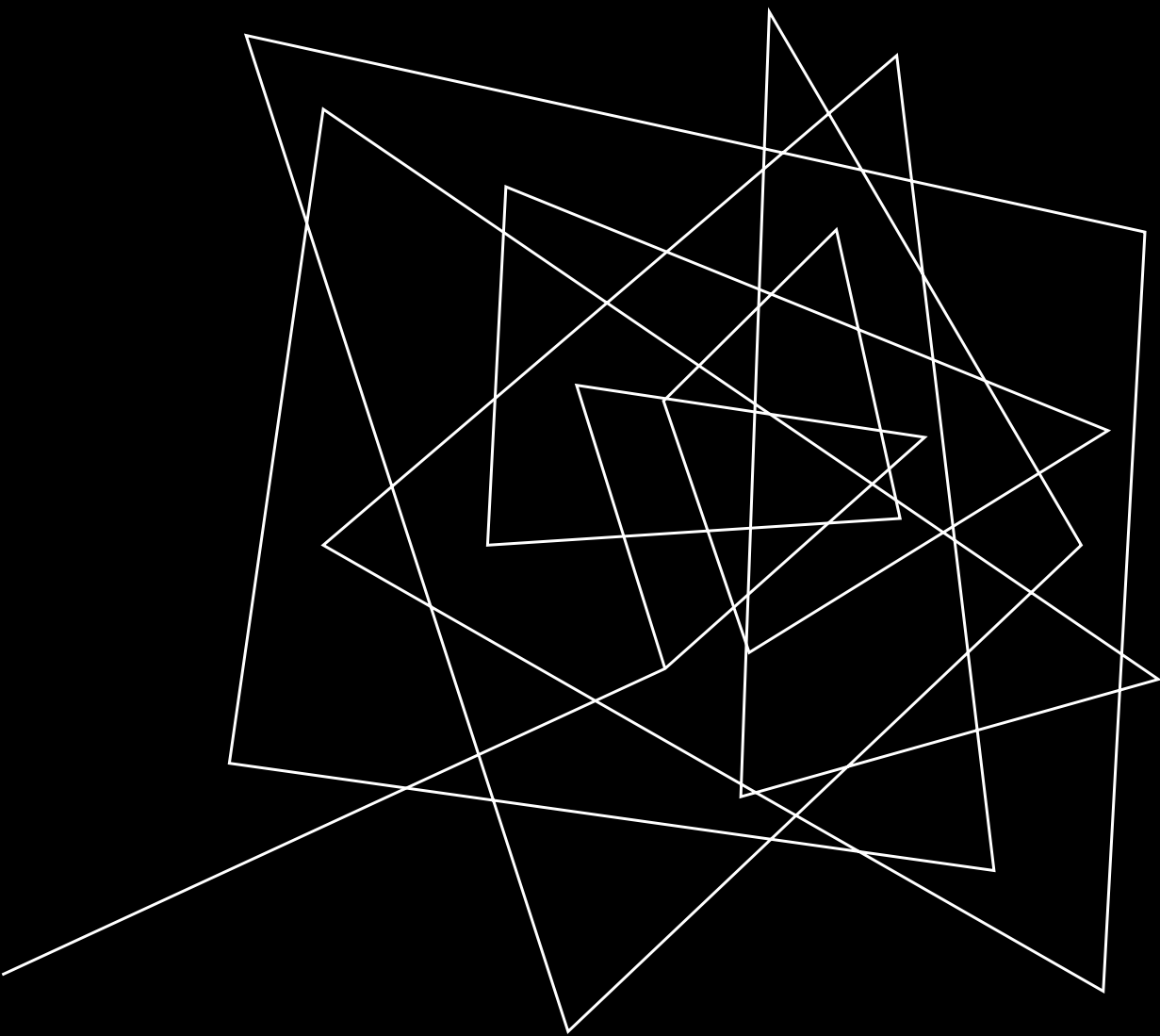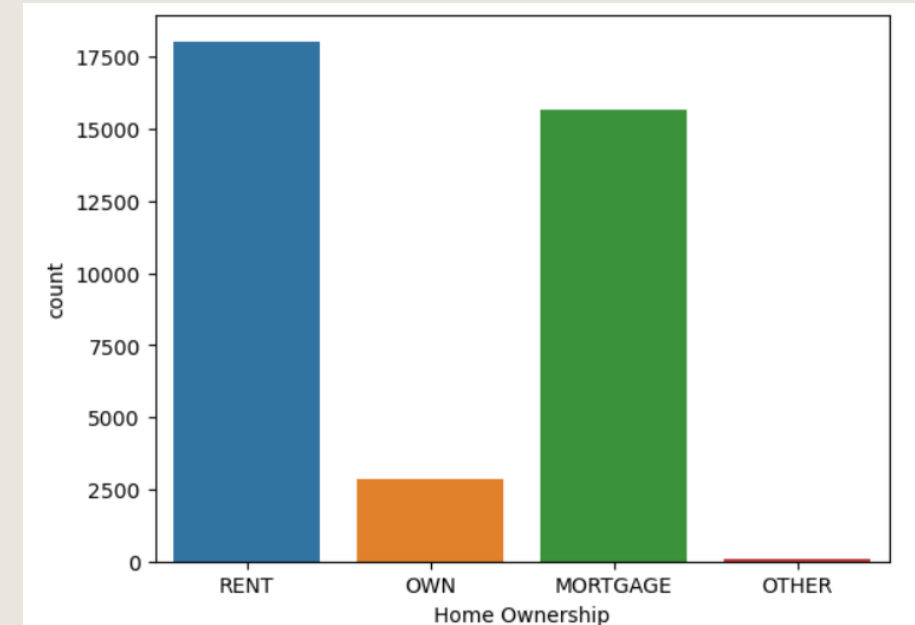
DATA ANALYSIS

Univariate

LOAN PAYMENT STATUS

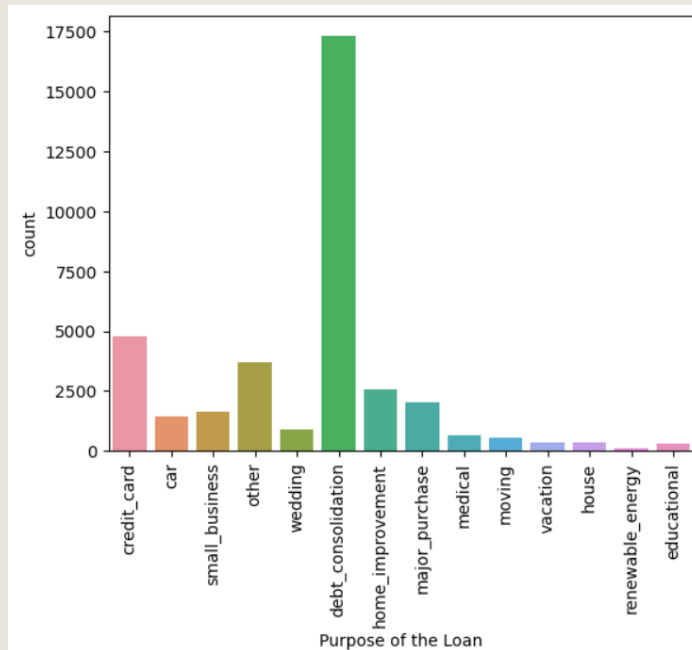Identified loan_status, Home Ownership as categorical variable and analysed data
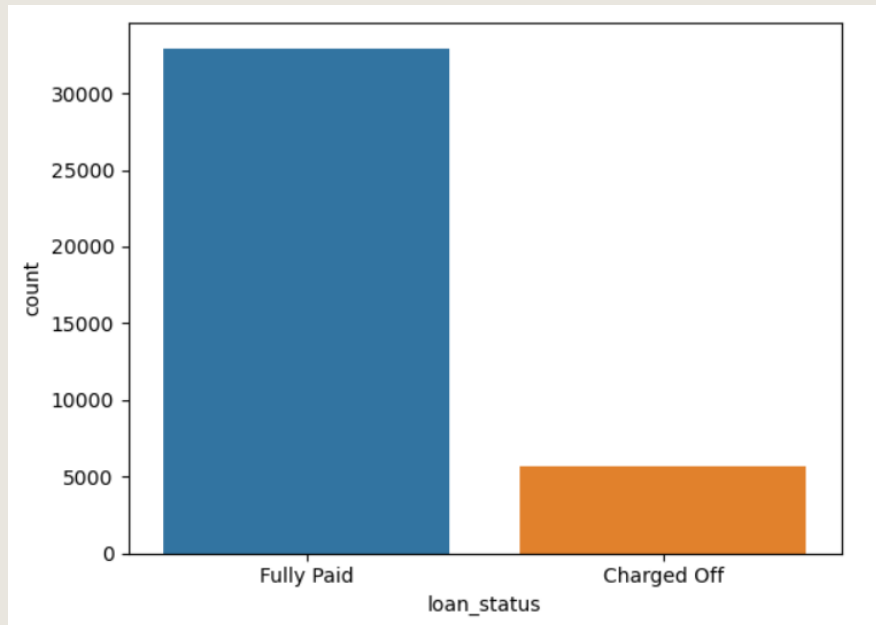
About 85.4% of loans are fully paid and 14.6% loans are charged off

Most of the people staying in rented house have taken the home loans
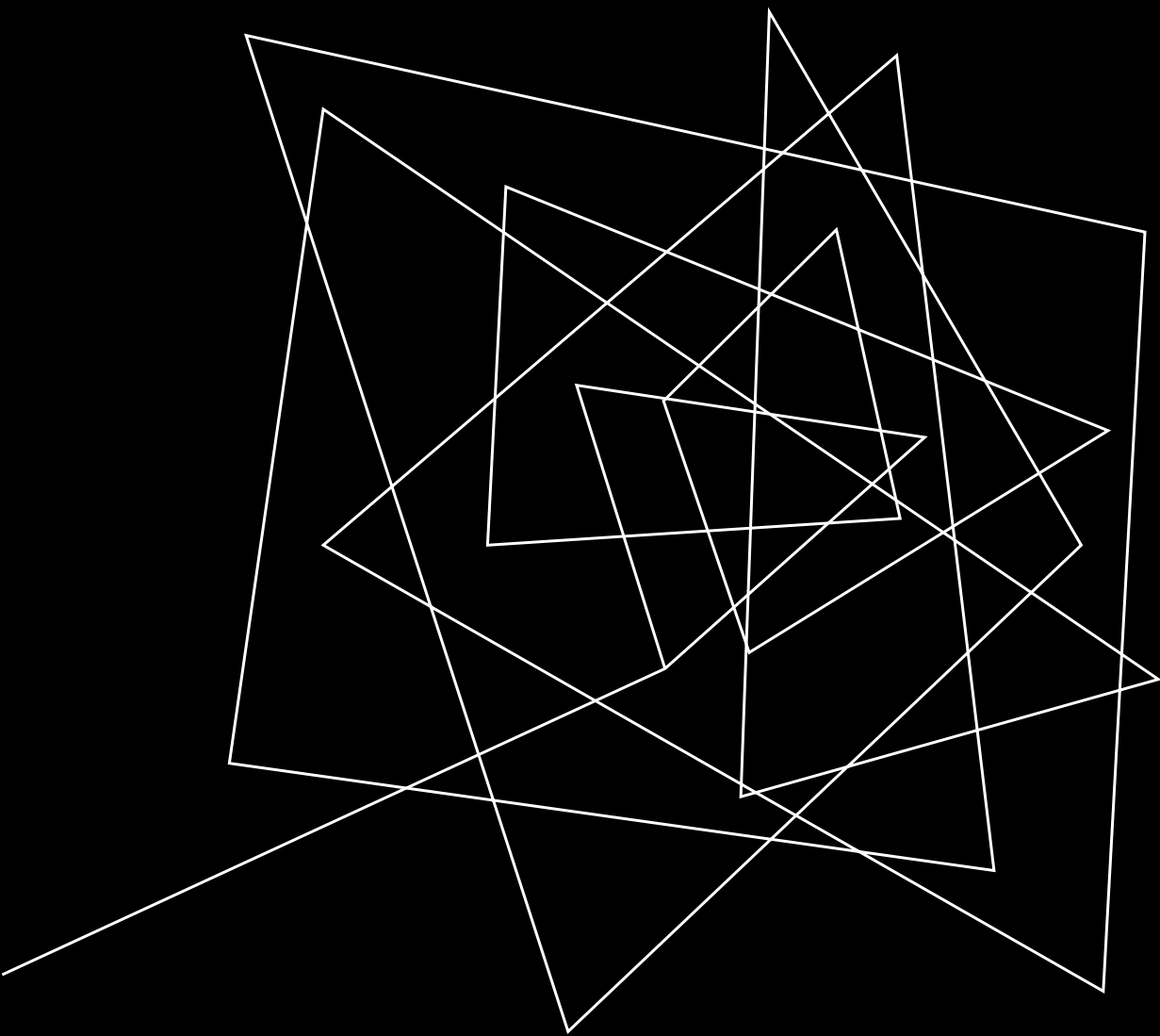
LOAN PAYMENT STATUS

Identified Purpose of the Loan, Duration of the Loan as categorical variable and analysed data

Most of the loans are taken for Debt consolidation
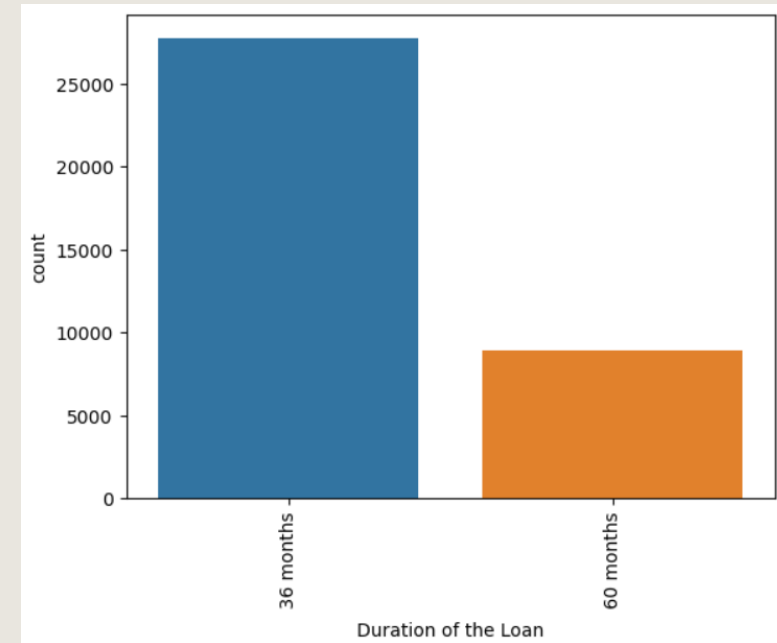
The Loans taken for 36 month term are much more than 60 months and have lower chance of defaulting.
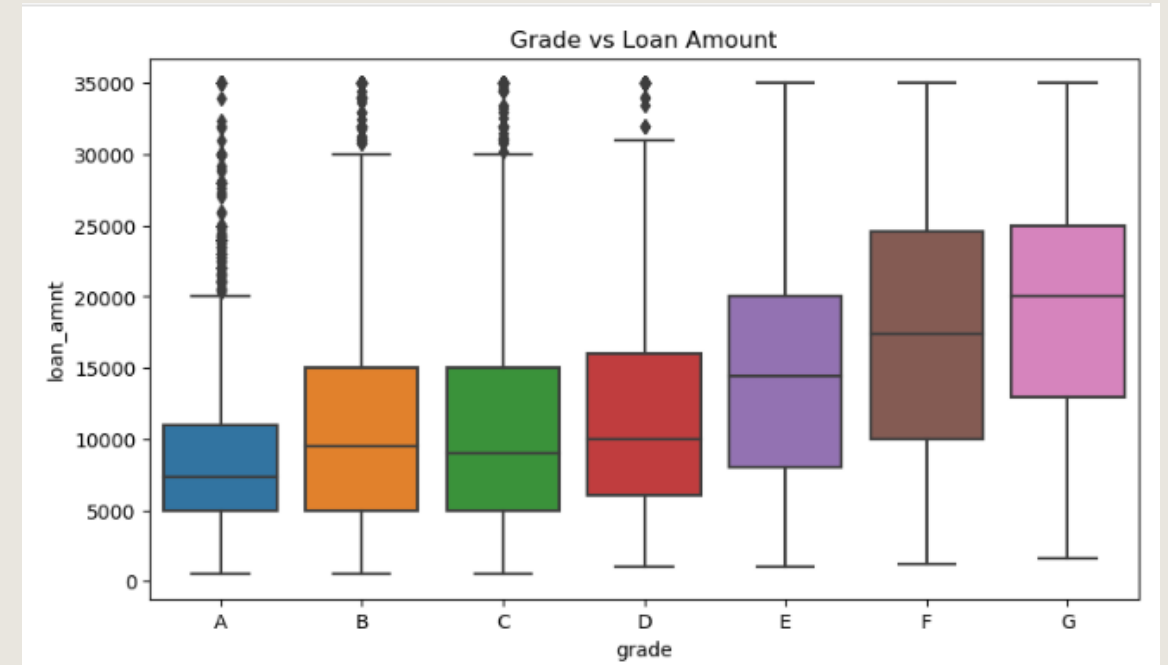
DATA ANALYSIS

Segmented Univariate & Bivariate Analysis

LOAN PAYMENT STATUS

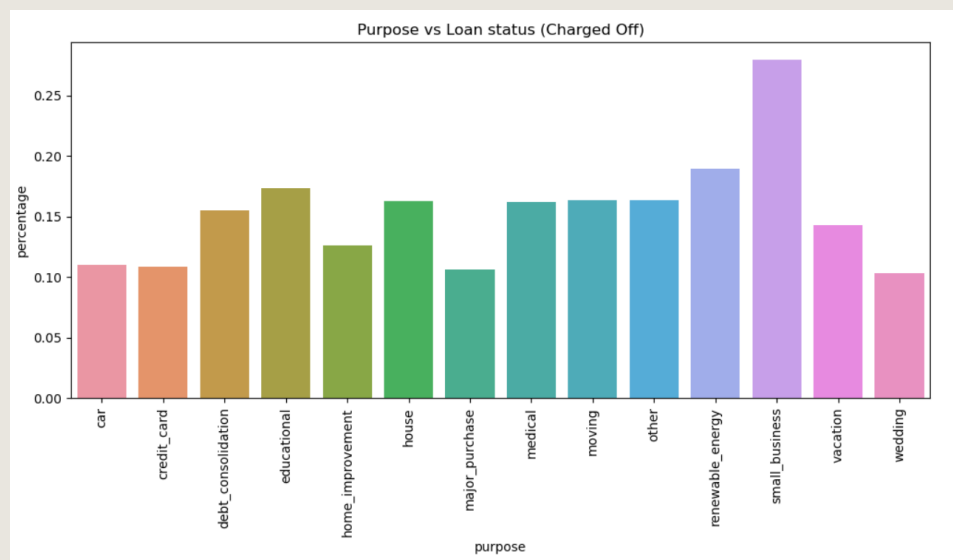Loan amount Vs Loan Grade & Loan amount Vs Loan Grade
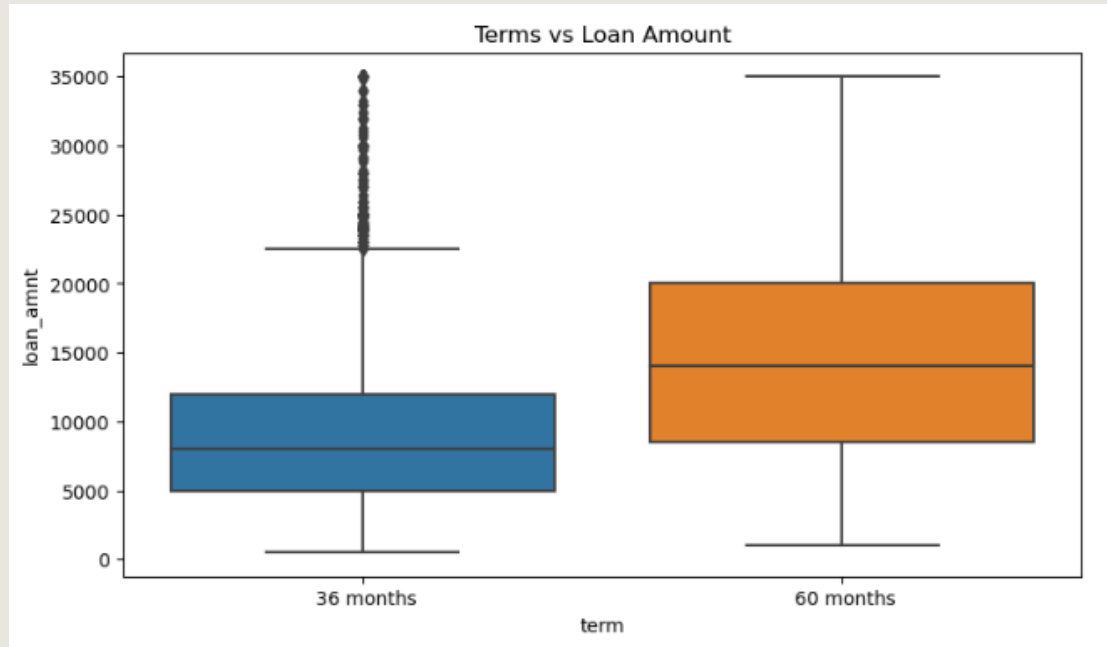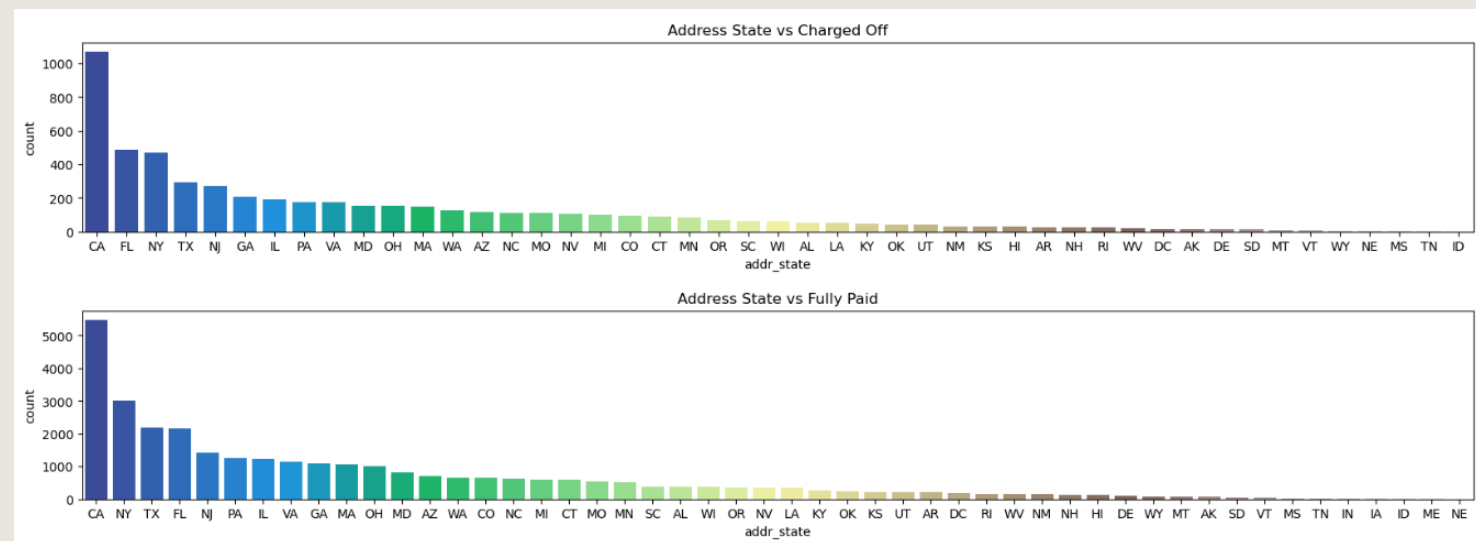

Bigger amounts have longer term

Loan amount is increased based on the grade, higher the grade loan amount is higher

REGION/LOAN PURPOSE-BASED ANALYSIS

As Charged off data is most important for future plans our analysis mostly focus on Charged off customers
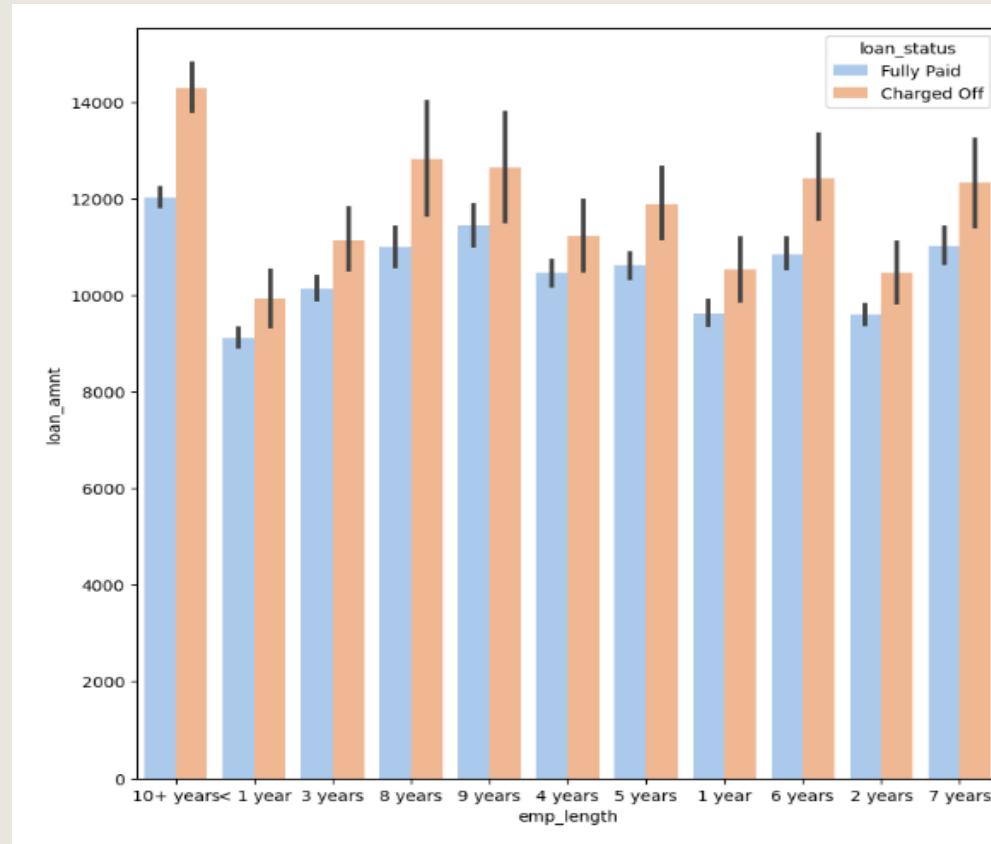

Loans taken for small business purpose are higher % of charged off
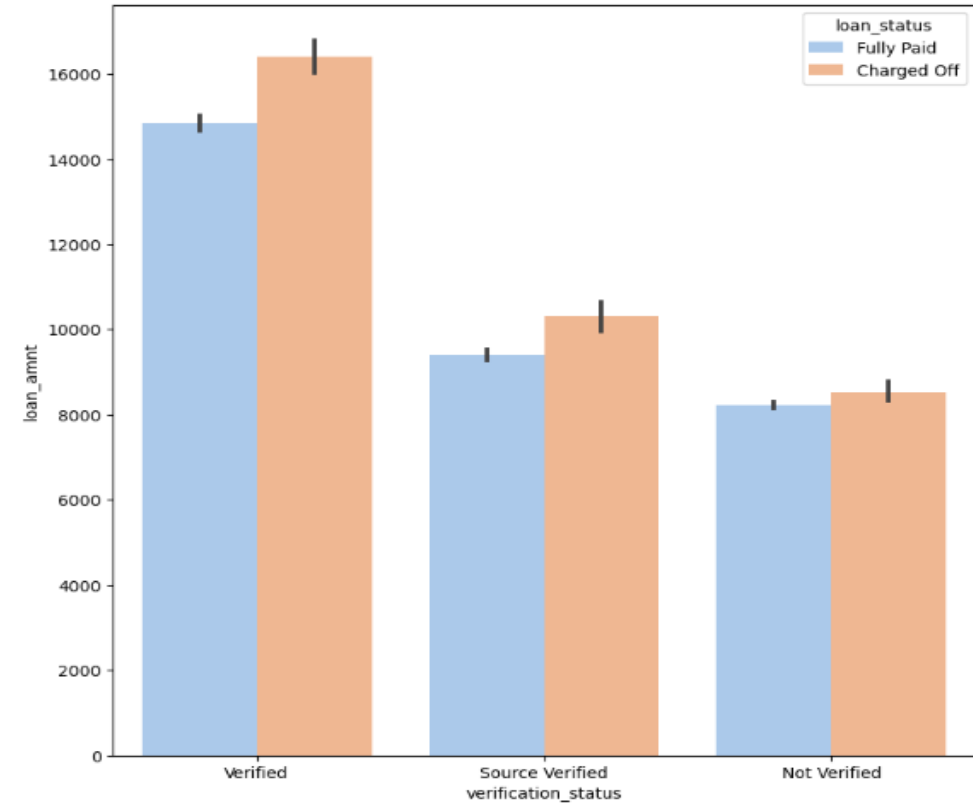
In CA region we observed that Loan Lending rate and defaulters' rate is higher
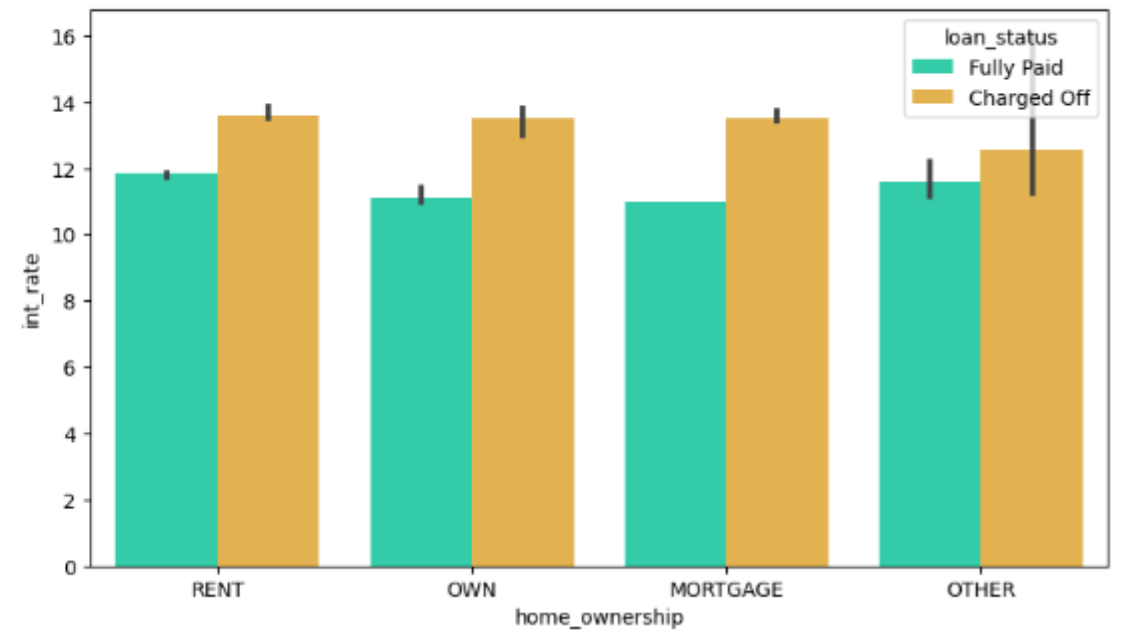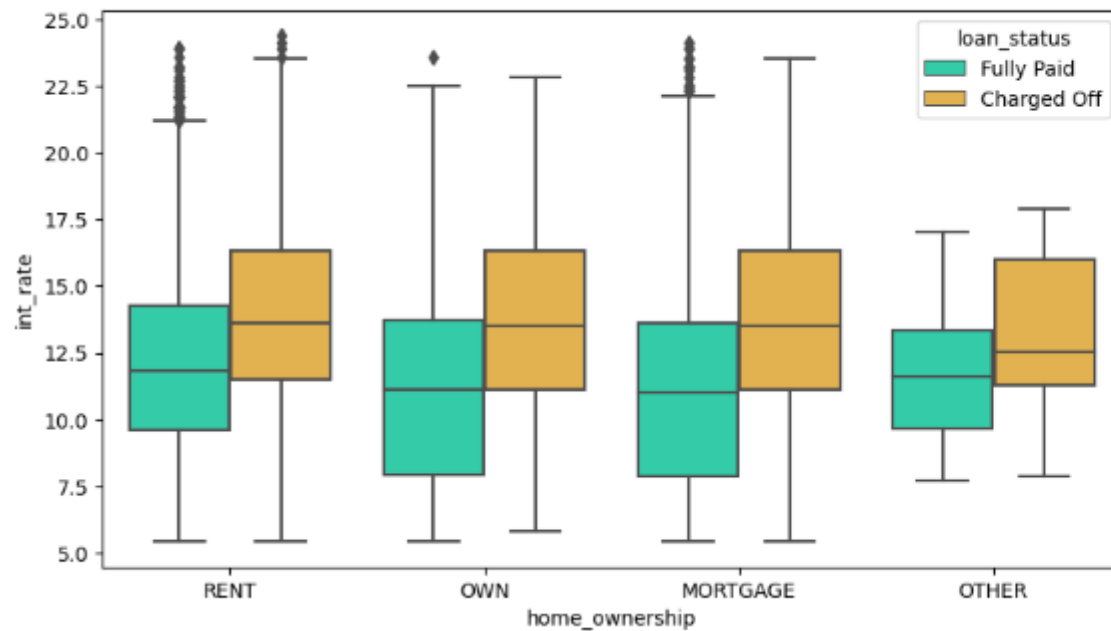
DEFAULTERS ANALYSIS

Employees of more experience are eligible to get higher loan amount, it is observed the defaulters list is also high

Verified borrowers are eligible to get the more loan amount. It is observed that there are more defaulters with the verified status.
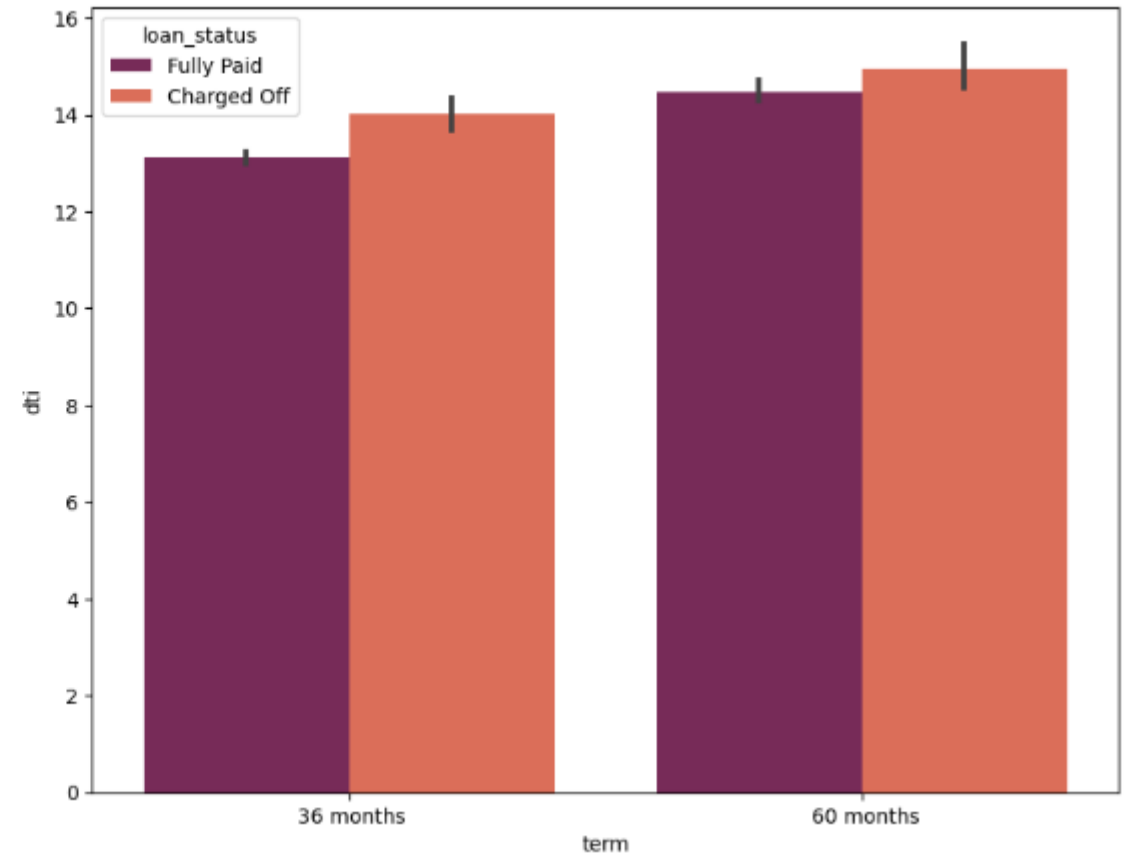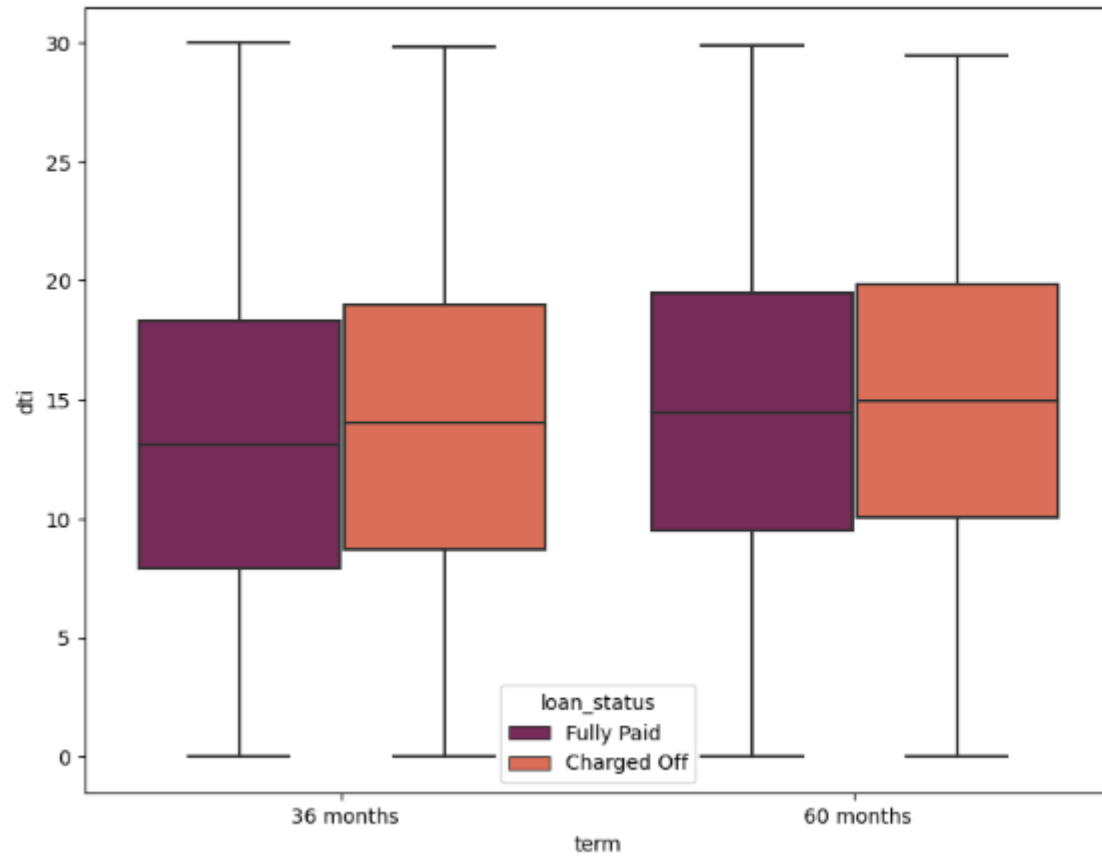
DEFAULTERS ANALYSIS
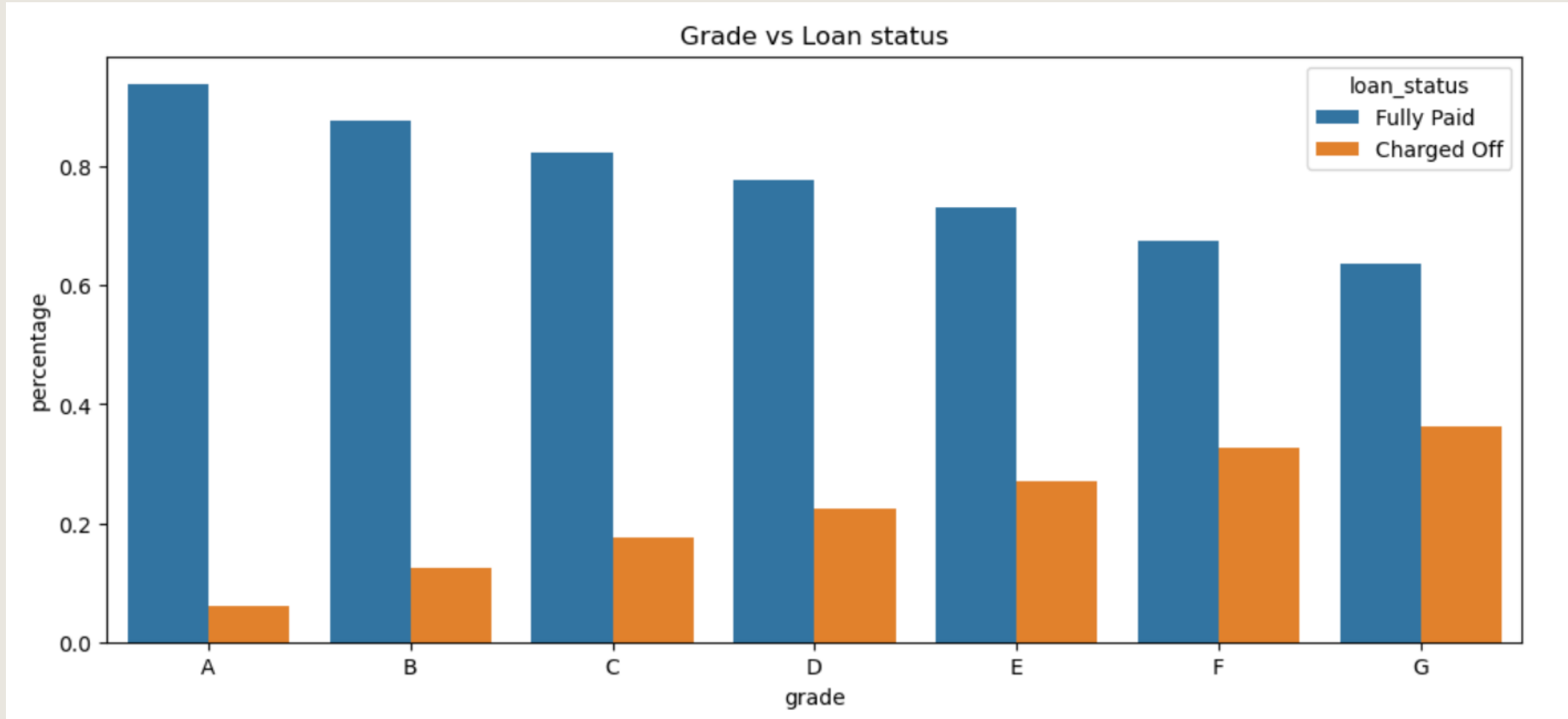

home_ownership vs Interest Rate

Observed that Changed off % is high, when the interest rates goes up

DEBT-TO-INCOME(DTI) BASED ANALYSIS

DTI of >13 for 36 months are charged off & DTI of >14 and 60 months are charged off

DEFAULTERS BASED ON GRADE OF THE LOAN

We observed that as the grade increases, % of charged off borrowers increases

SUMMARY

Following are the parameters to be considered while giving the loans to avoid the Losses to Lending Clubs

- Verification Status

- Purpose of the loan

- Borrower's Employee experience

- Borrowers' current resident status

- Interest Rate

- Borrowers with High DTI value

THANK YOU