LENDING CLUB CASE STUDY

Group Members:

1. **Komati Sathish (Facilitator)**
2. **Vanama Charmila(Member)**

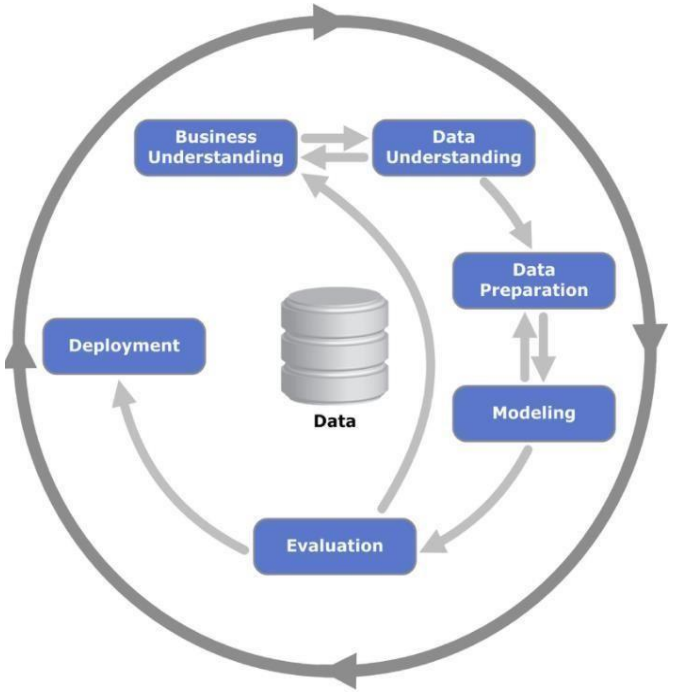
OBJECTIVES

Identification of the risky loan applicants.

Understand the driver variables or driving factors behind the risky loan default.

Lending companies can utilize this knowledge for its portfolio and risk assessment

FLOW CHART FOR DATA ANALYSIS



Understanding Dataset
(Data Cleaning)

Data Cleaning
(Imputed the NA values for all the driver variables)

Data Preparation

Exploratory Data Analysis
(Performed Univariate, Univariate Segmented & Multivariate analysis)

DATA CLEANING AND MANIPULATION

Redundant data

- *For this analysis, we don't need items that are "current" in loan status as these are for people who are still in the process of paying loan so no decision can be taken using this data. So, there is only need to analyze the patterns for people who's status is "fully paid" and "charged off“*

Converting data to suitable format

- *For calculation purposes, it is better to remove unnecessary things and keep all items in a column consistent throughout*

Derived Data

- *Manipulating strings and dates*

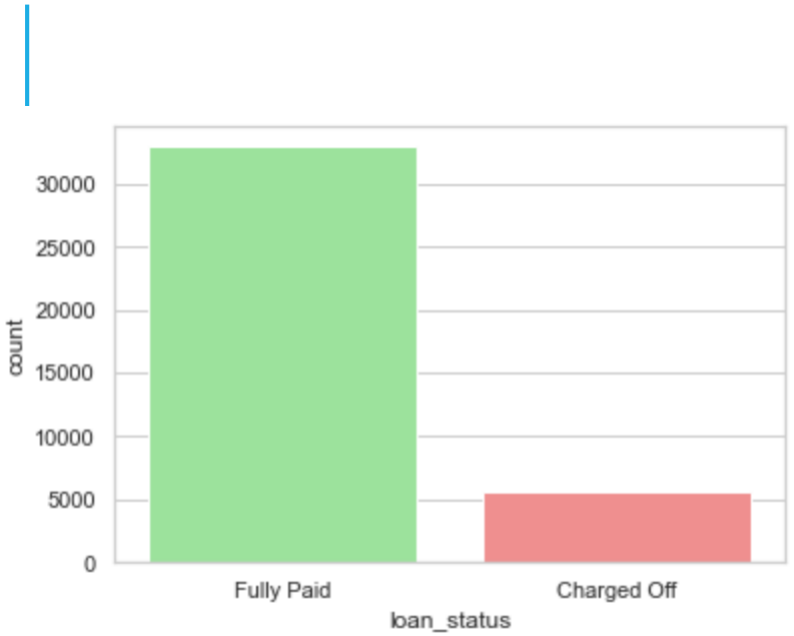
DATA ANALYSIS

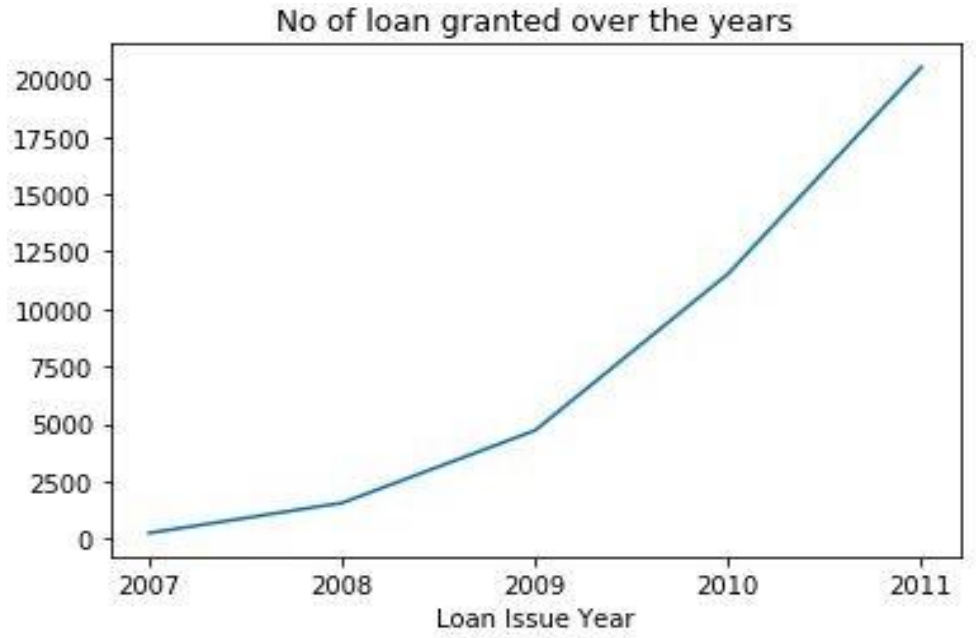
- In this case study we analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting loan .

- We will do data cleaning and data preparation on the Loan dataset:
  - Imputed the NA values for all the variables
  - Created two new columns:
    - Profit and Loss column
    - Ratio of funded amount and annual income

- univariate analysis :
    - *It means analyzing one variable at a time*
    - *Single variable needs to be analyzed clearly before further analyzing the multiple variables or whole dataset.*
    - The Ratio of funded amount and annual income

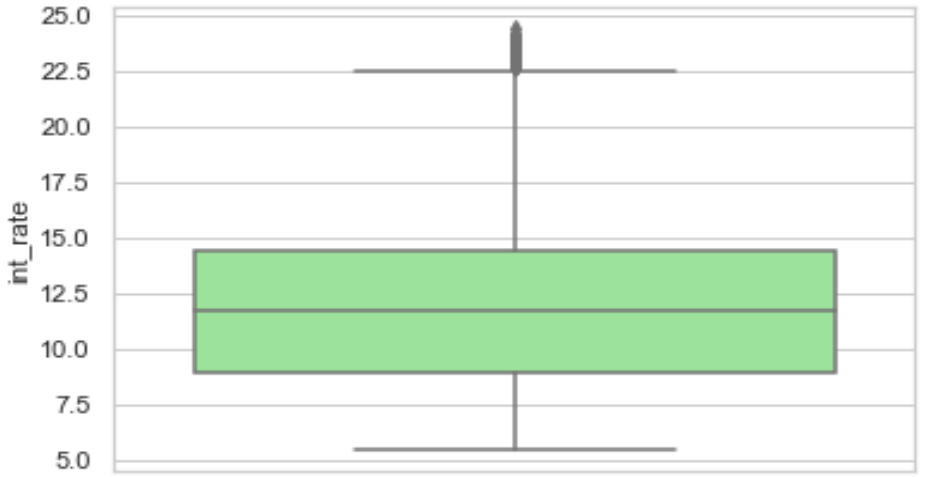
UNIVARIATE ANALYSIS



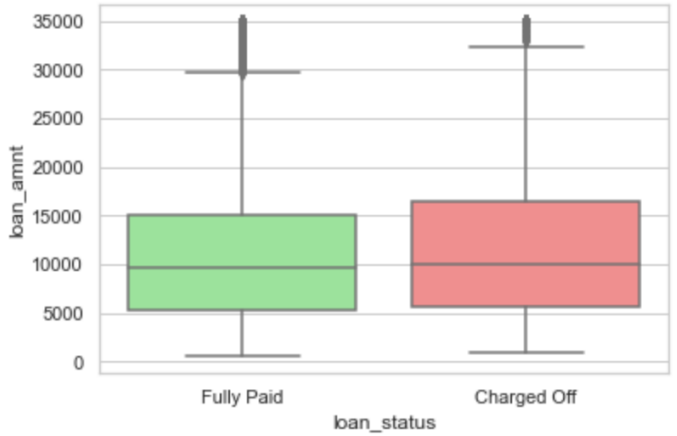


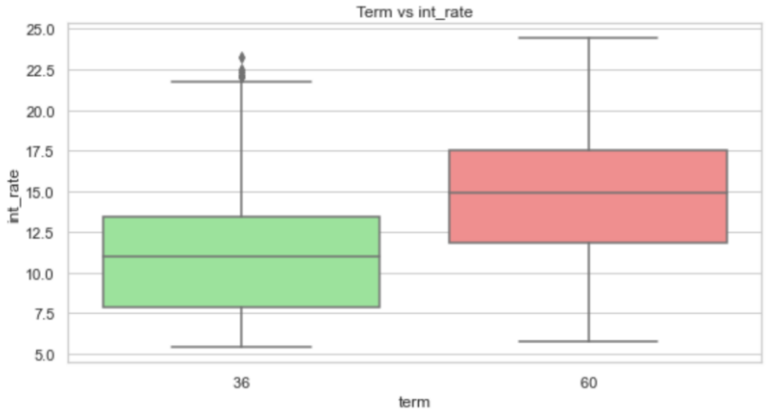
Observation:

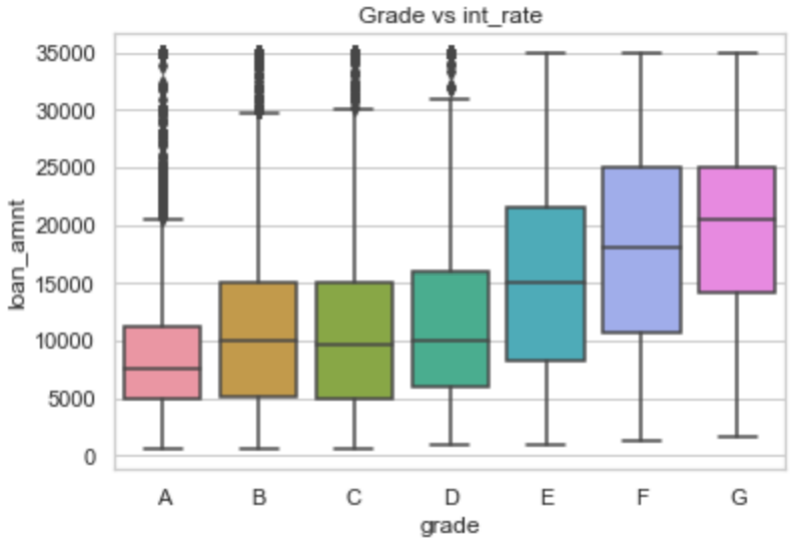
1. There are very few members who comes under charged off category, that is around 14.6%
2. The loan has been increasing exponentially over the years
3. The interest rate is between 5 to 25



SEGMENTED UNIVARIATE ANALYSIS







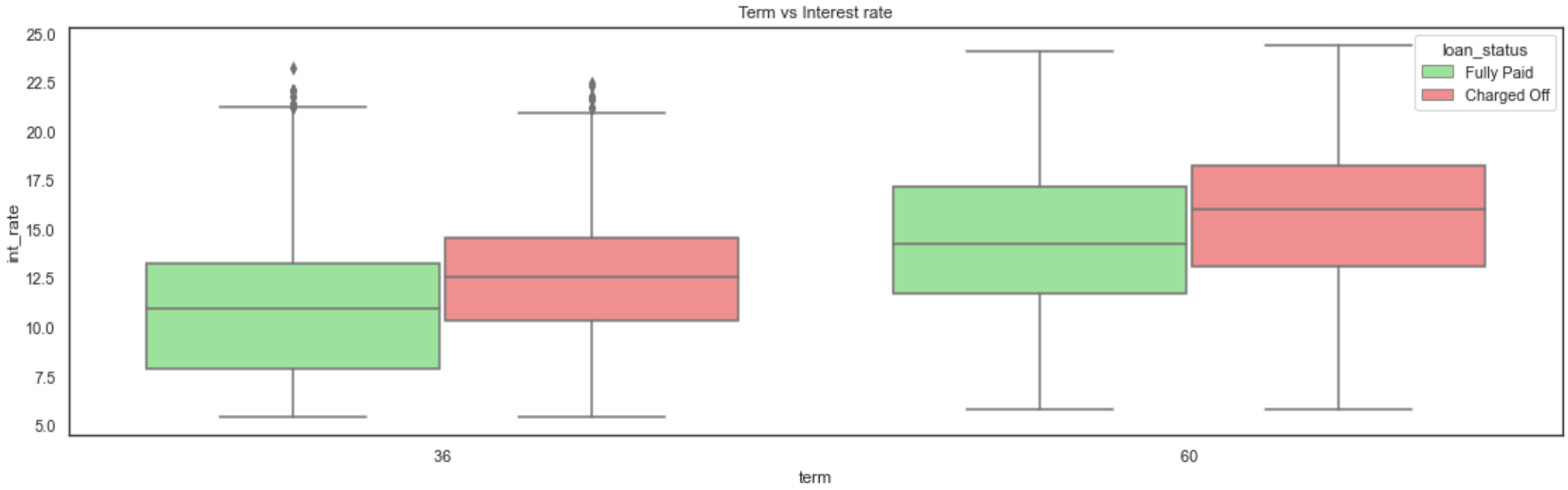
Observation:

1. Loan amount is increased from grade A to grade G
2. Loan status is charged off, when there is high interest rate, high loan amount and high dti

Bivariate Analysis

Bivariate analysis is essentially an extension of the segmented univariate analysis to another categorical variable.

term vs int_rate vs loan_status

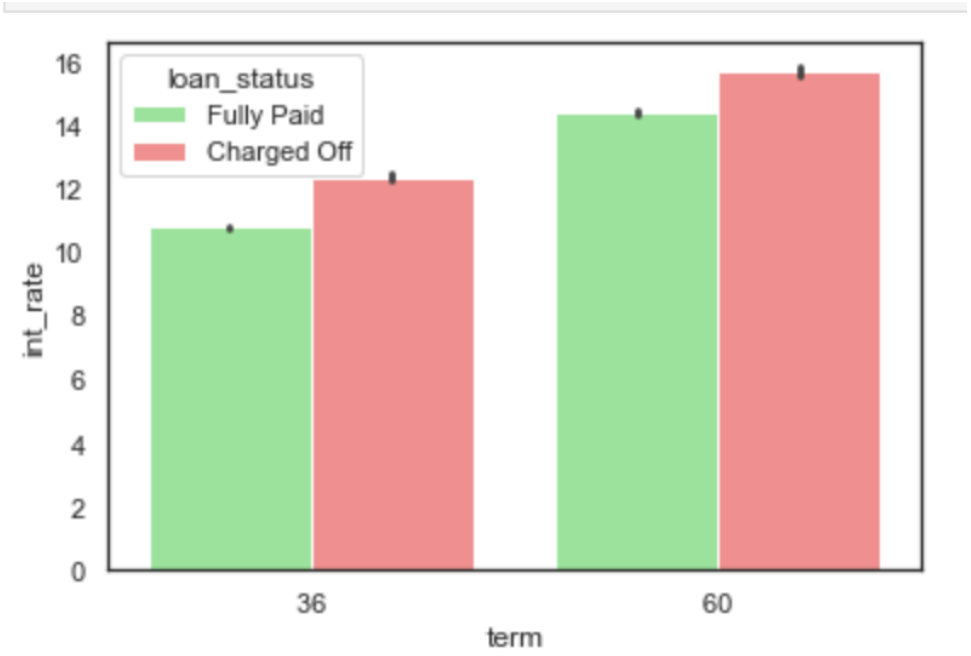


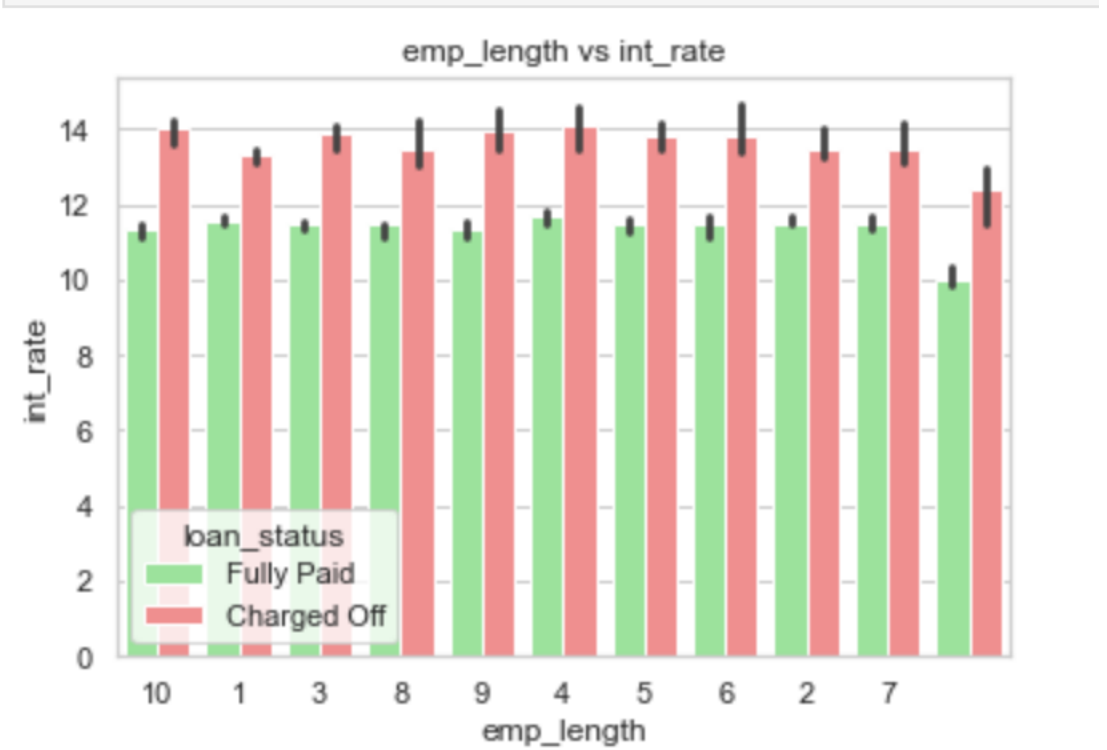
Observation *:*

From above box plot, it is clear that the people are getting charged off at the higher interest rate

Bivariate Analysis

term vs int_rate





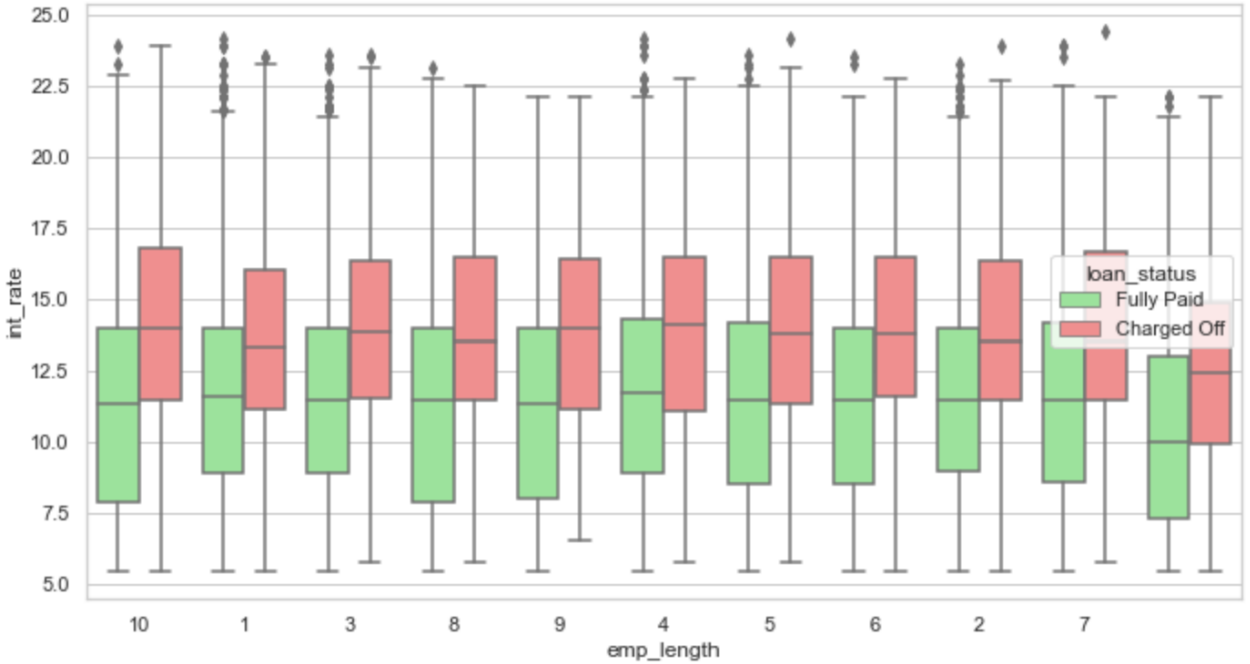
Observation *:*

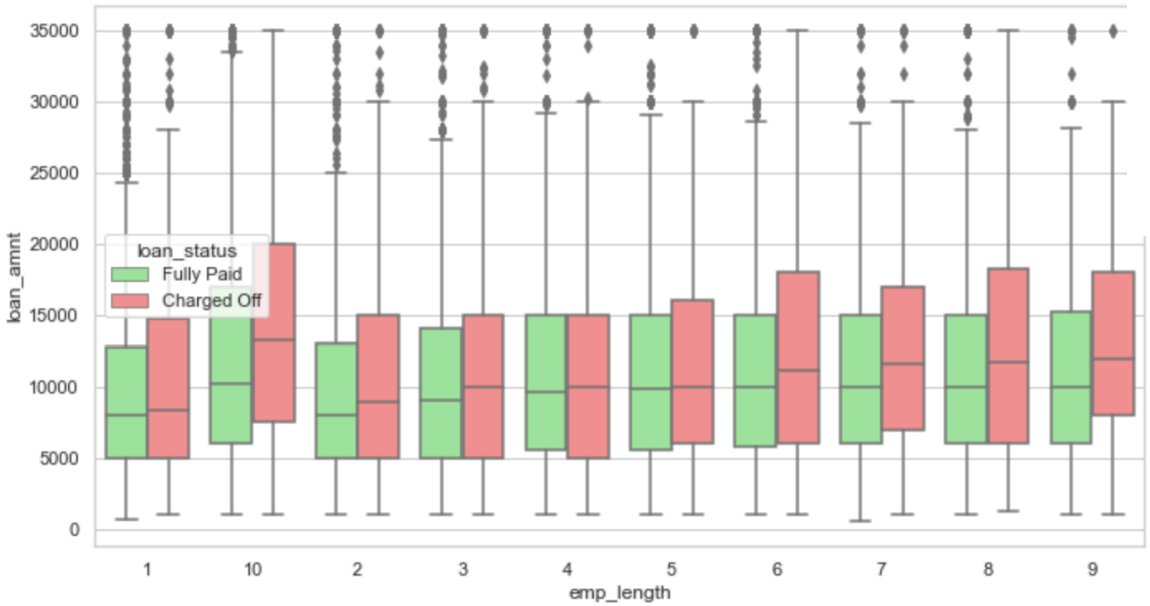
From above box plot, it is clear that the people are getting charged off at the higher interest rate

Bivariate Analysis

Observation *:*

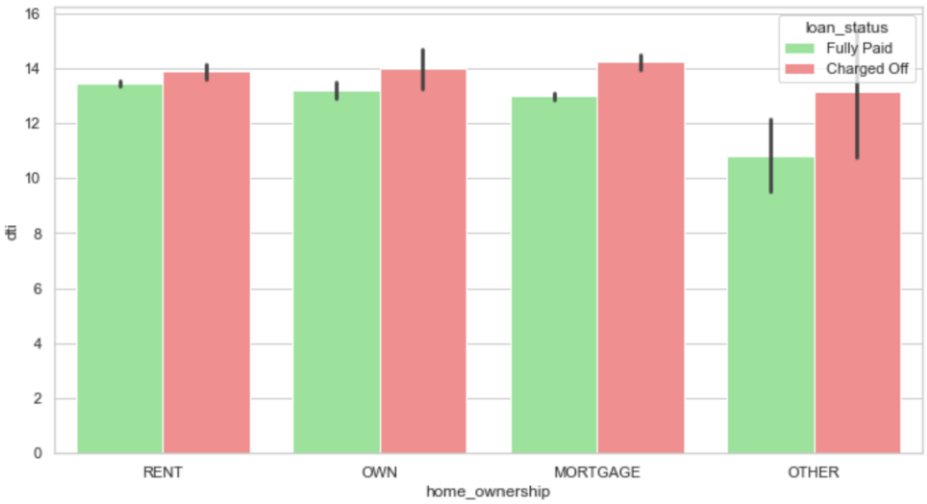
Charged off rate is higher with interest rate, across all emp_length and loan_amount

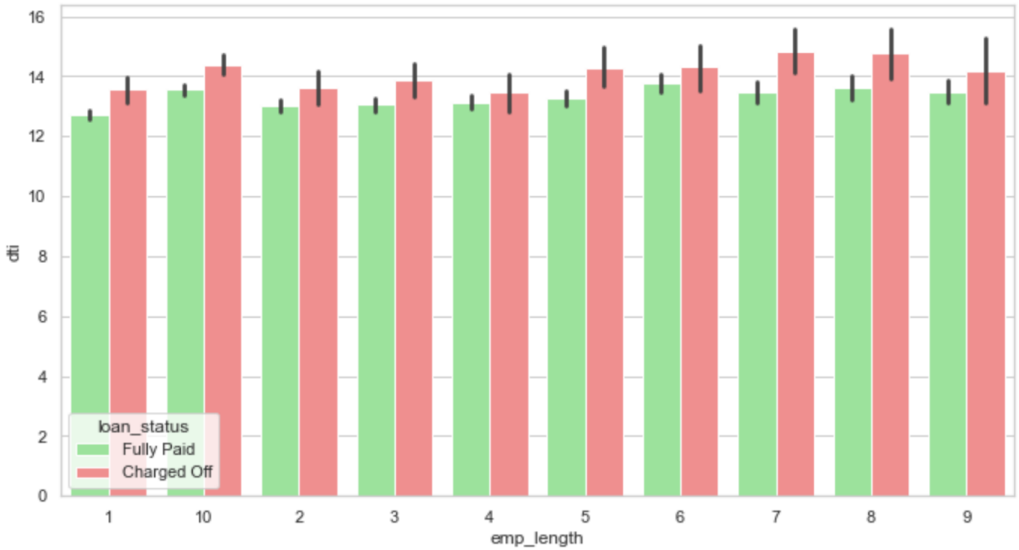


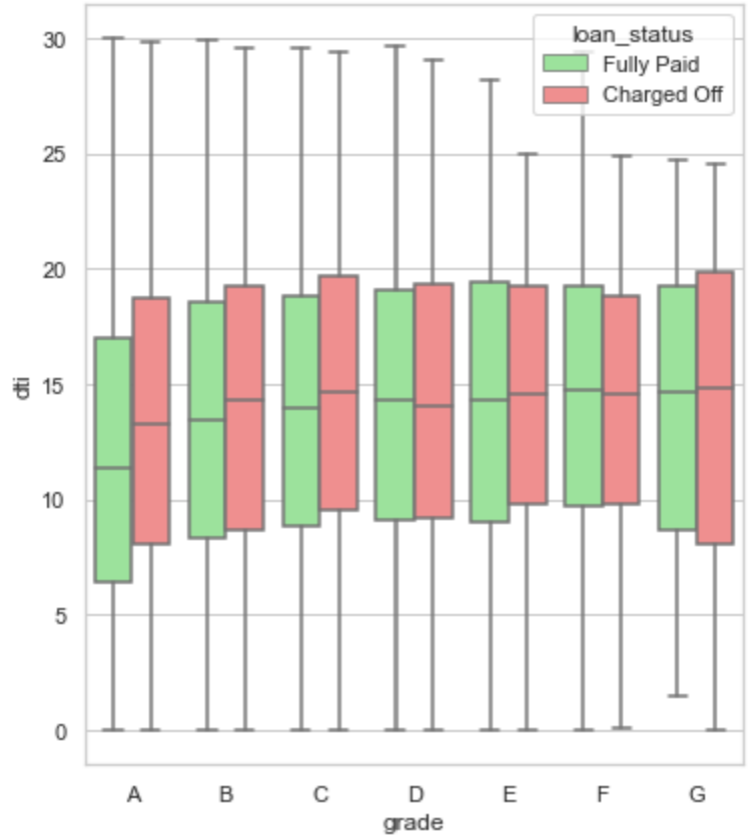


Observations: For all the experience levels, the fully paid members have less loan amount as compared to charged off members. From this graph, we can say that members who have high loan amount have high chances of getting charged off

Bivariate Analysis







Observation *:*

Charged off rate is higher with dti, across all emp_length, home ownership and grade.

SUMMARY

- There are more defaulters in loan in scenarios where there is higher interest rate. If interest rate is high, check the background of applicant thoroughly
- Extra scrutiny is necessary required for the applicants are having high DTI. This may lead to higher chances to default
- Default rate for grades is increasing from grade A to G
- Highest loan amount is offered to small business purpose, hence, high interest rate is also offered to small business. Hence, bank will face loss if charged off decision is not correctly decided in this category
- If member mentioned purpose as debt consolidation then check thoroughly as it has high chances to default