

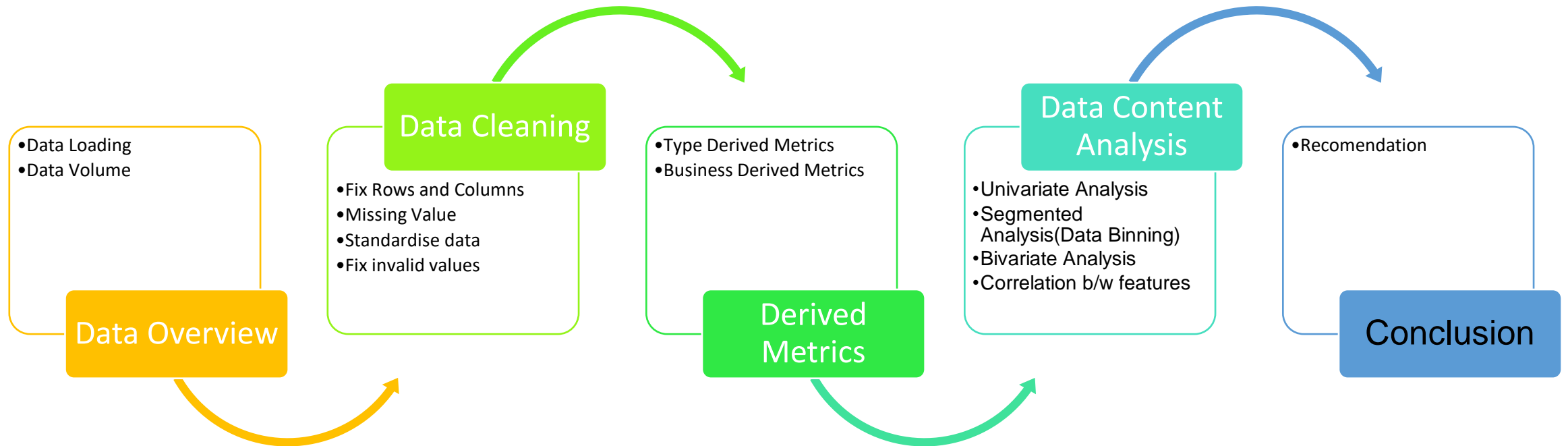
Lending Club Case Study

By
Sheena Mathew

Business Objectives

- 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. and With the help of **loan, consumer attributes** and **loan attributes** influence the tendency of default.
- The lending club company can clearly understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. Lending company can utilize this knowledge for its portfolio and risk assessment.

Problem solving methodology



Insights on Data Cleaning

1. **Dropping** below list of columns , since the below features do not add any value to the analysis

- ❖ Columns which have full null values.
- ❖ Columns which have only one unique value.
- ❖ Checking for duplicate records taking id as subset, no duplicate rows found.
- ❖ Columns where all values are unique. for instance id,member_id,url.
- ❖ .Columns that do not contribute to check the analysis of defaulters list , it is mostly for post loan approval stage
[title, ip_code,addr_state,desc,last_credit_pull_d,out_prncp_inv,,total_pymnt,total_pymnt_inv,revol_bal,total_rec_prncp
total_rec_int,total_rec_late_fee,recoveries,collection_recovery_fee,last_pymnt_d,last_pymnt_amnt,next_pymnt_d,delinq_2yrs,pub_
rec_bankruptcies'
- ❖ funded_amnt is removed as it is amount that is approved by the lending club and not by the investors.
- ❖ Columns which has more 50% null values are dropped. for instance::mths_since_last_record and mths_since_last_delin
- ❖ Records pertaining to Current are removed, as it is ambiguous whether they would repay the amount or not
- ❖ emp_title is dropped as it has lot of unique values

2. Missing values

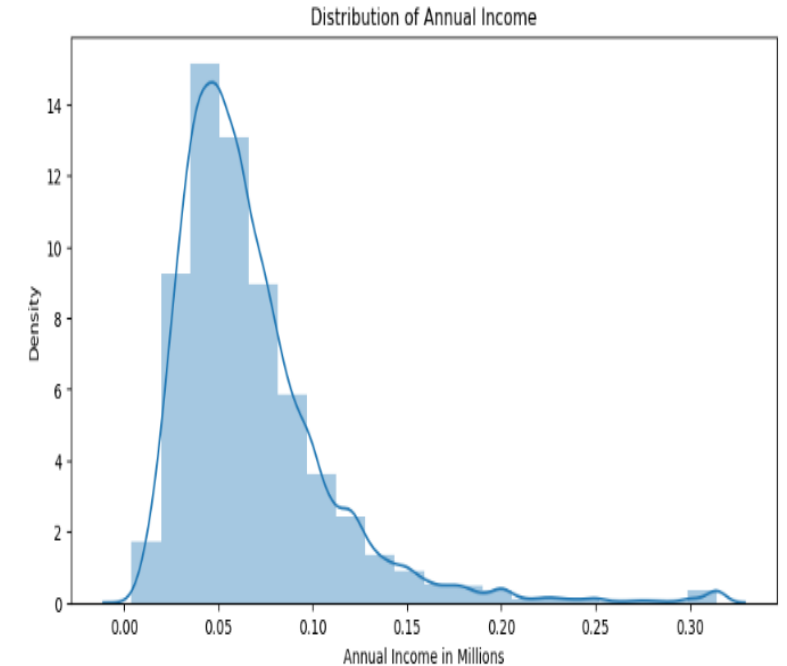
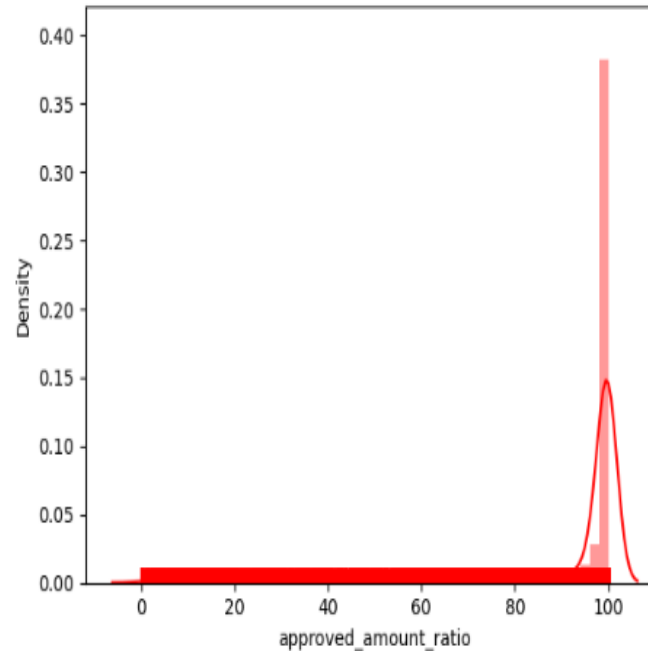
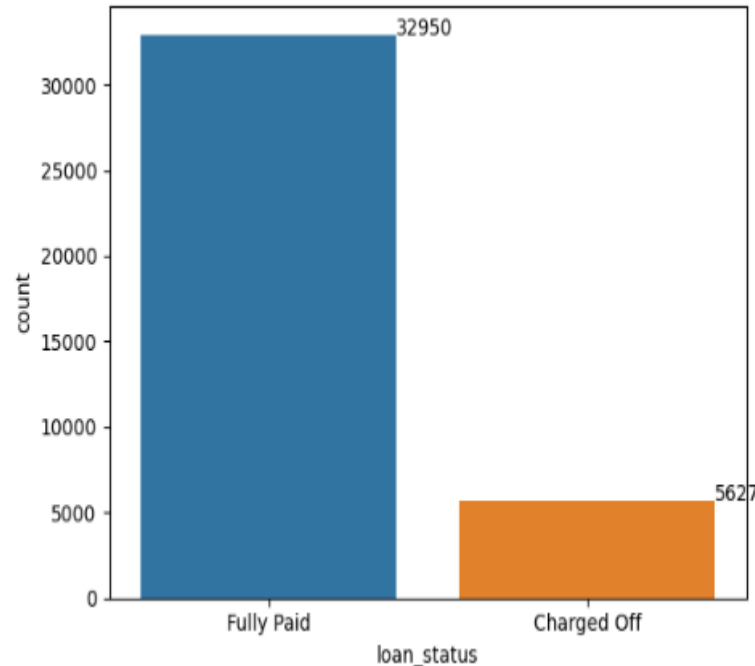
- ❖ Based on the missingno analysis , 2 columns were found with more than 0% null values
[emp_length,revol_util]
- ❖ emp_length was imputed with mode value
- ❖ revol_util were filled with median since it had skewness

3 . Standardize Data

- ❖ Incorrect data types are replaced with proper data type for the below columns
 - int_rate is converted to float object from string object
 - issue_d and earliest_cr_line is converted to datetime object
- ❖ Detecting and Removing Outliers.
- ❖ Data is categorized into numerical variable and categorical variable.
- ❖ outliers found in ['annual_inc','inq_last_6mths','open_acc'] were capped with min and max value.

After Data cleaning , Size of the **Data Frame is (38577, 21)**

Data Analysis

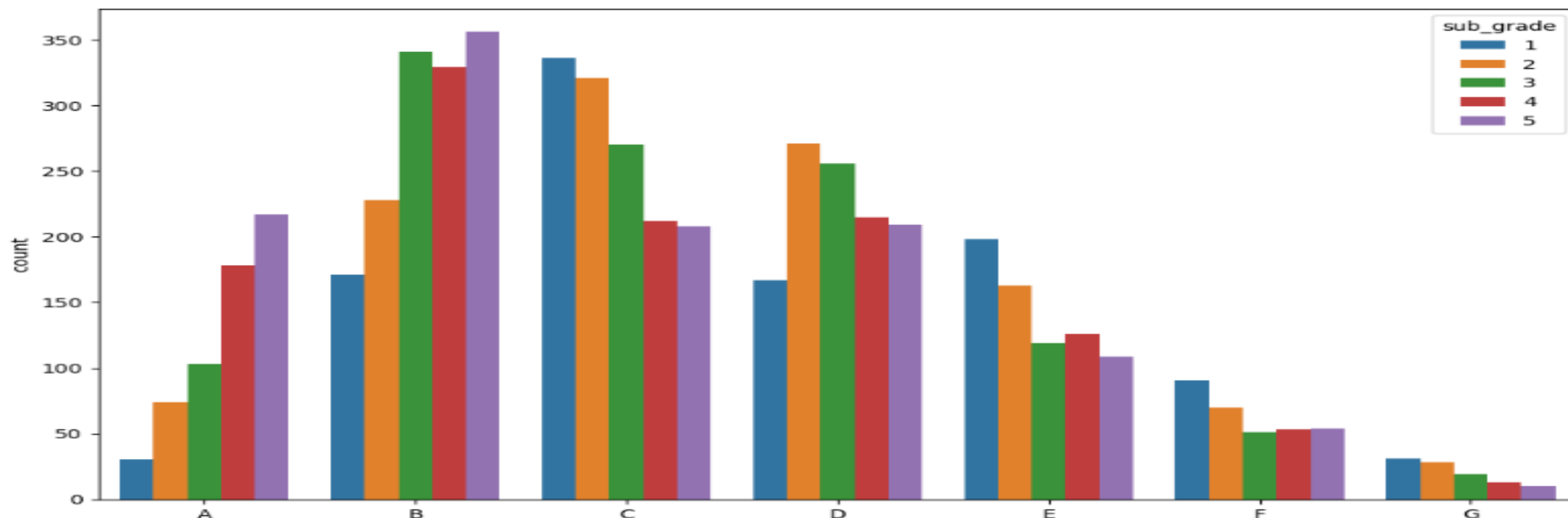


- only **13%** loan applicants have defaulted the loan
- **90%** of the applicants loan was approved by the investor
- Annual Income is **left skewed** normal distribution ,therefore majority burrowers have relatively low income than others

Data Analysis

Loan status(charged off) is taken as target variable

1.Grades vs Loan Status

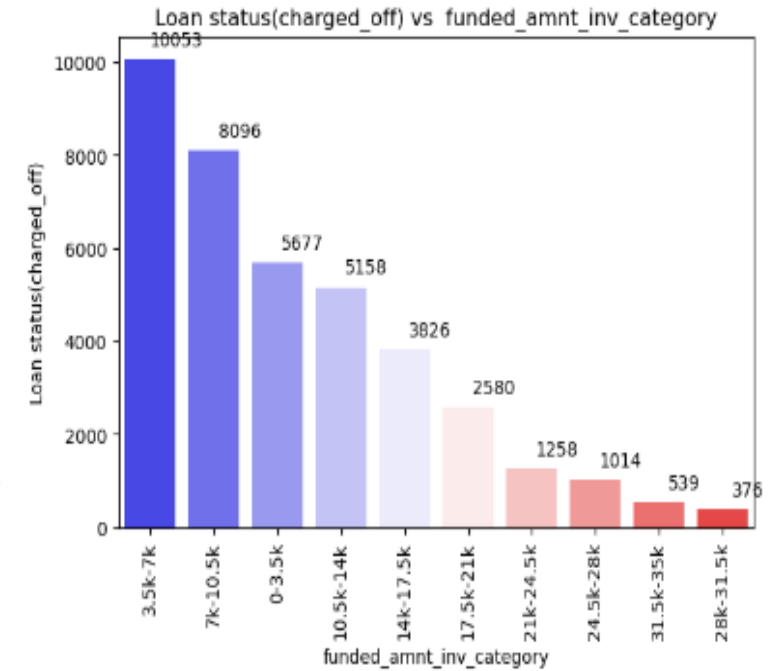
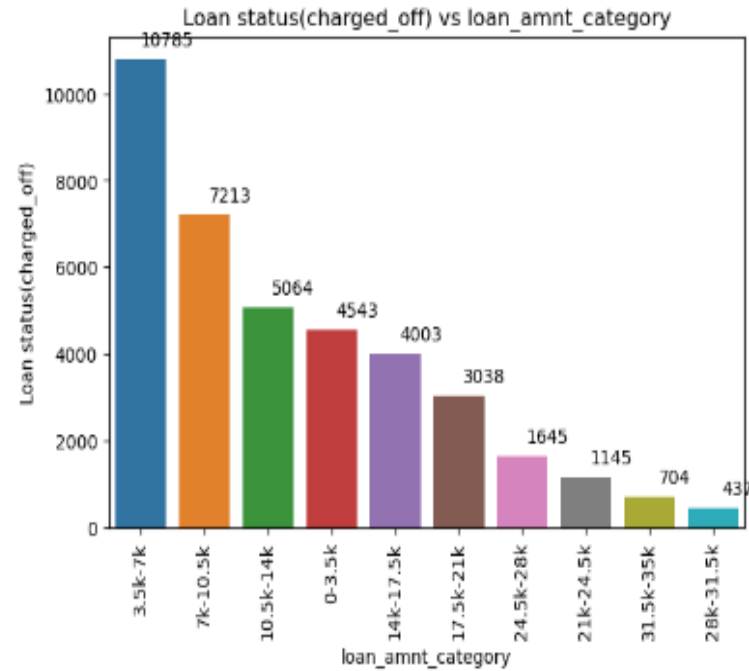
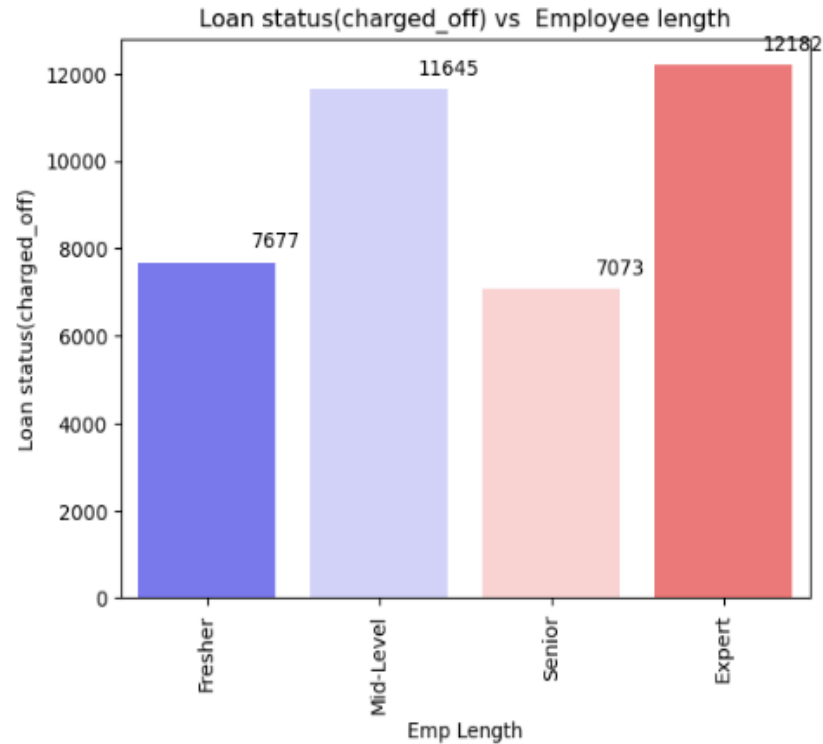


- Majority of the loan defaulters belongs to **A and B grades**
- for Grade C,E, F and G. sub grade 1 has defaulted the most compared to other subgrades but for A and B subgrade 1 has defaulted the least

2.Employee length vs Loan Status

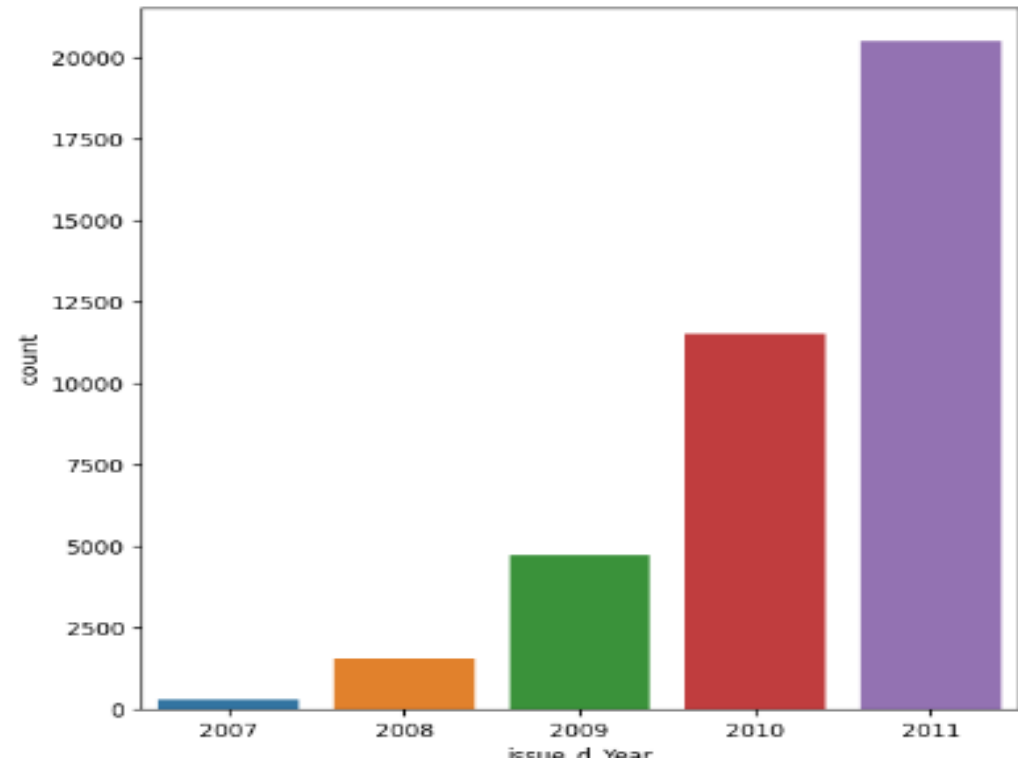
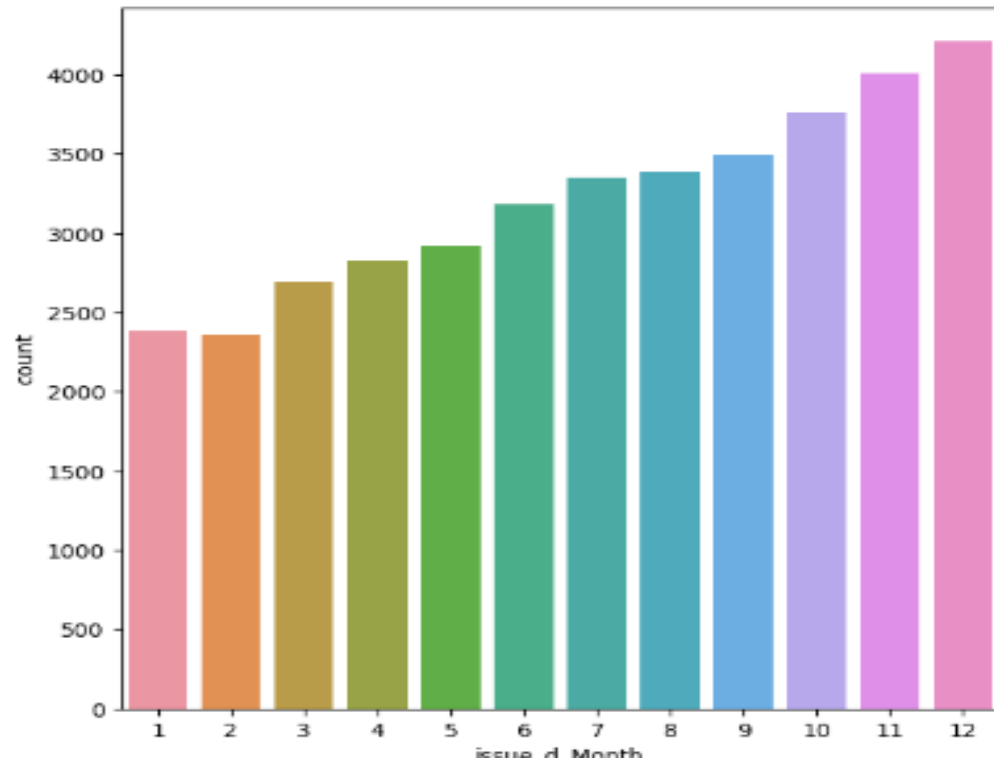
3.Loan Amount Category vs Loan Status

4.Funded Amount inv category vs Loan Status



- Applicants who are charged off belongs to **Expert Category**(i.e they have experience more than 7 years)
- funded amount is in the range of **3500-7000**

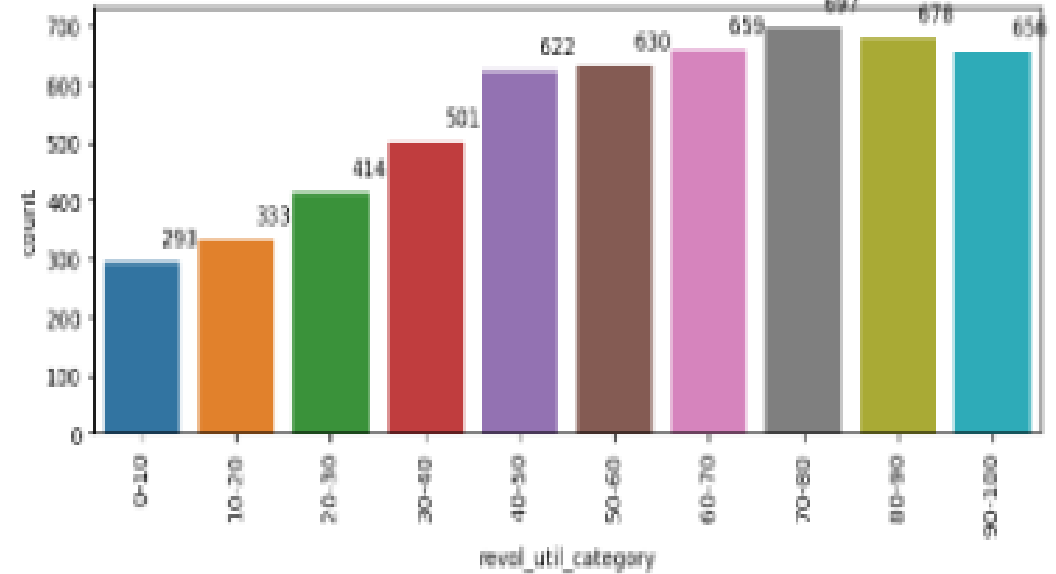
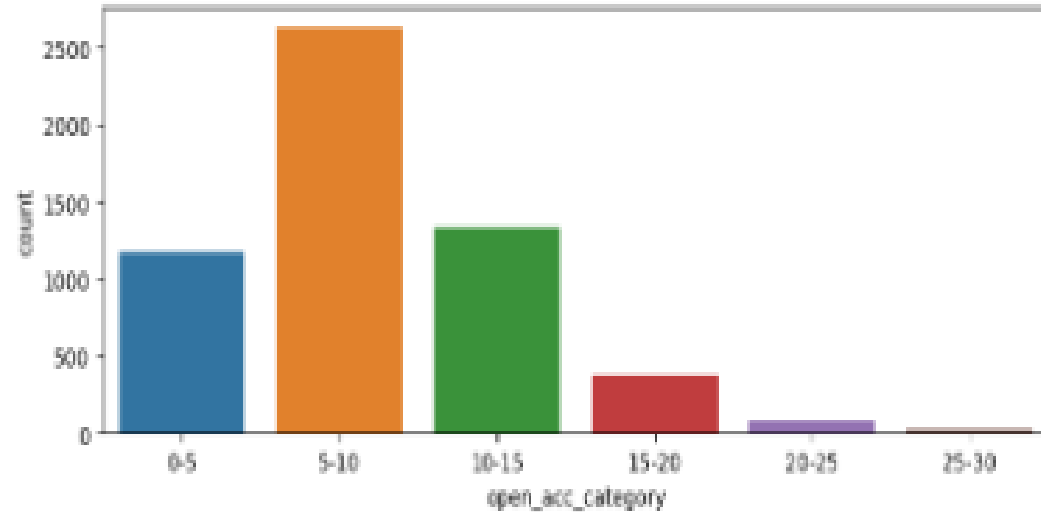
5.Loan status vs Loan Issued Month and Year



Applicant whose loan was issued in the year **2011** and in the month of **December** have defaulted the most. It might be because of the 2011 financial Crisis that took place in US

6.Loan Status vs open account category

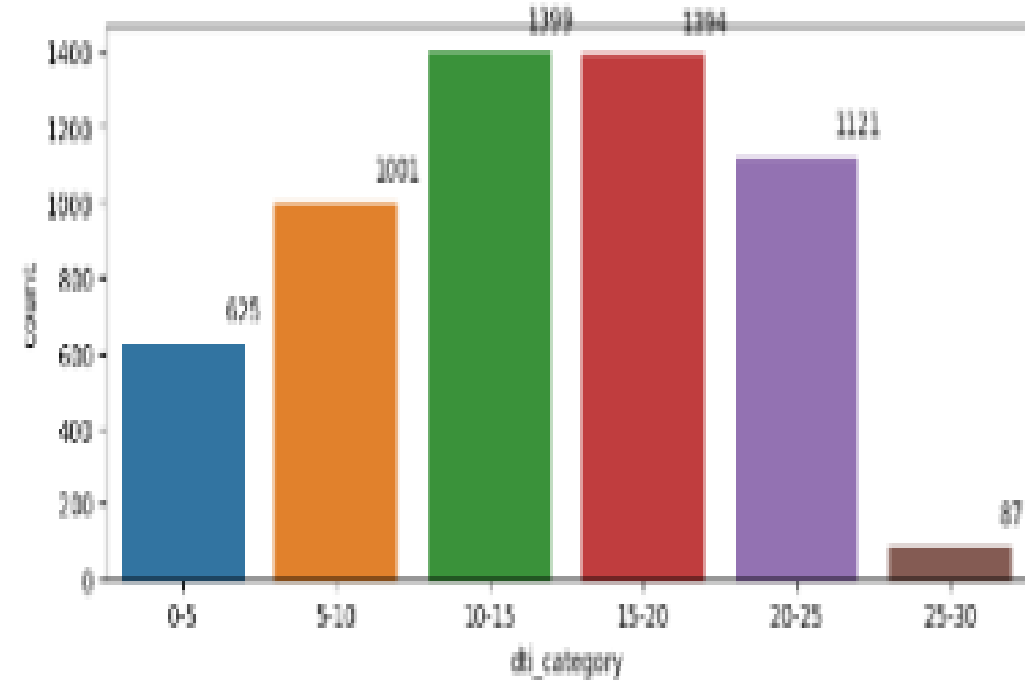
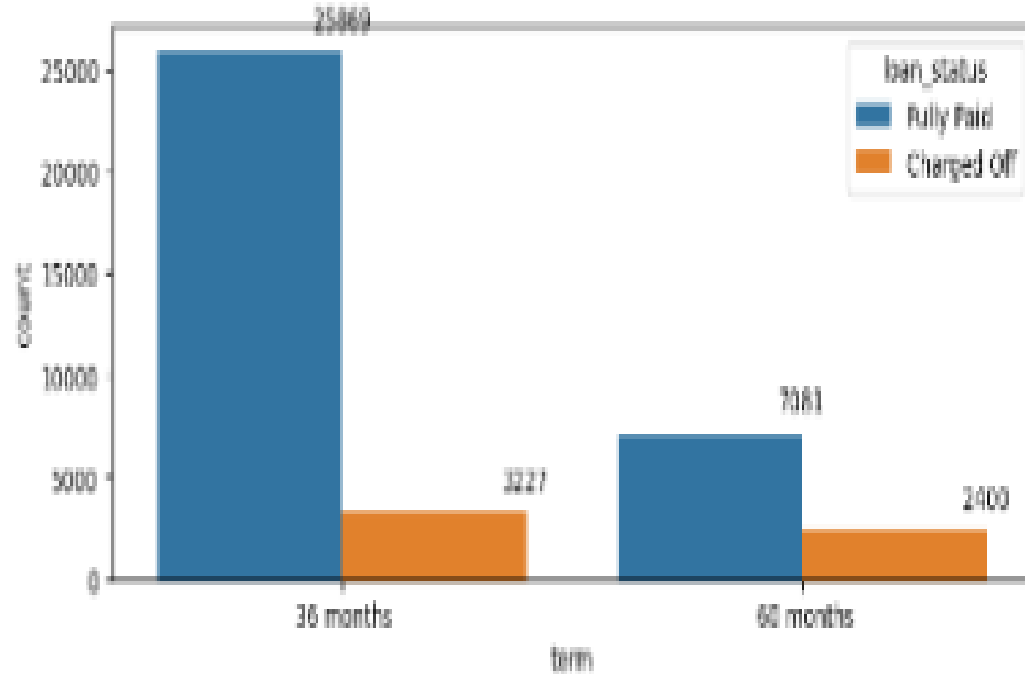
7.Loan Status vs Credit utilization Ratio



- **Higher the credit utilization rate** , higher the default rate
- Applicants who have **5-10 open_acc** , have defaulted the most

8. Loan status vs Term

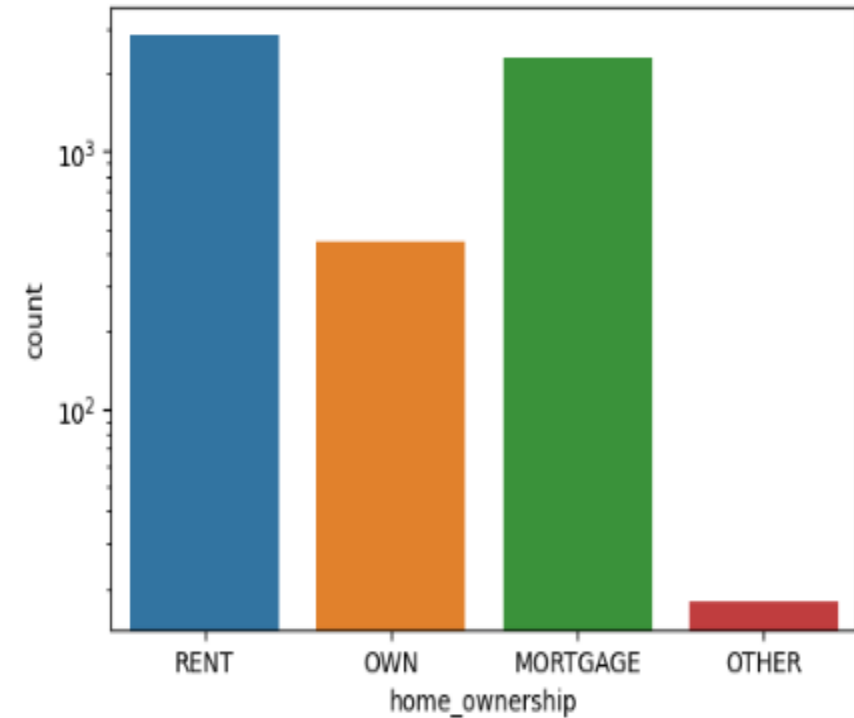
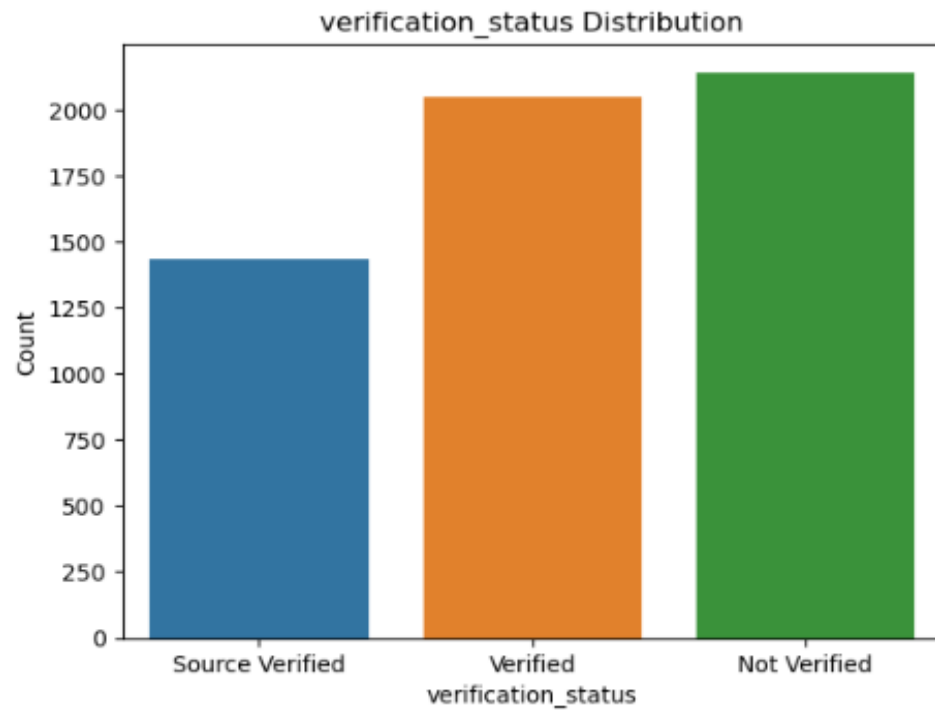
9.Loan Status(Charged off) vs DTI Category



- Most of the Applicants have taken **36 month** term to pay the loan off
- **Majority defaulters** are those who have taken **60 months** term ,As there is **only 11%** default case in **36 month** as compared to the ones who have taken loan for 60 months period.
- Maximum defaulters have DTI is in the range of 10-15 .

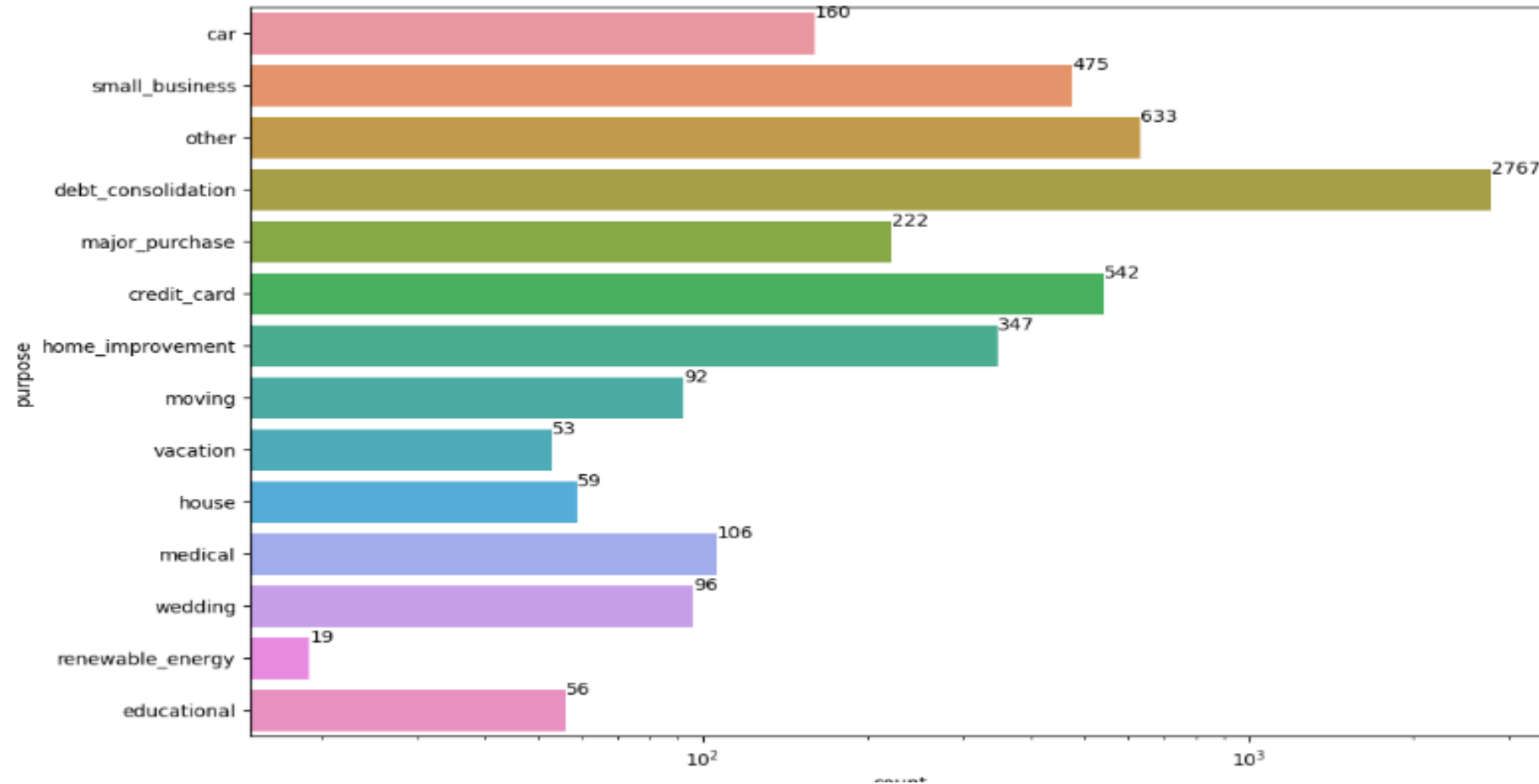
10.Loan Status(Charged off) vs Verification status

11. Loan Status(Charged off) vs Home Ownership



- Annual Income and other document were **not verified** for the vast majority of the defaulters
- Applicants having house_ownership as '**RENT**' have defaulted the most

12.Loan Status(Charged off) vs Purpose

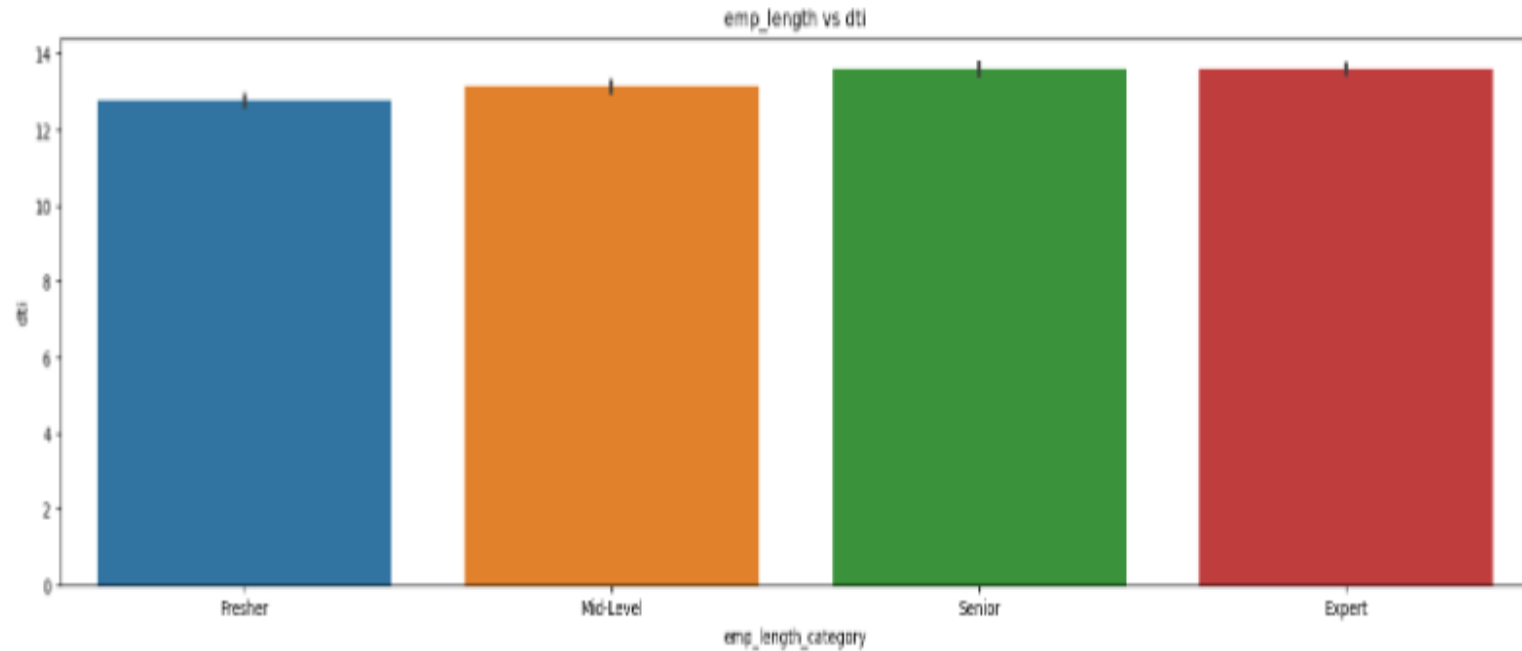
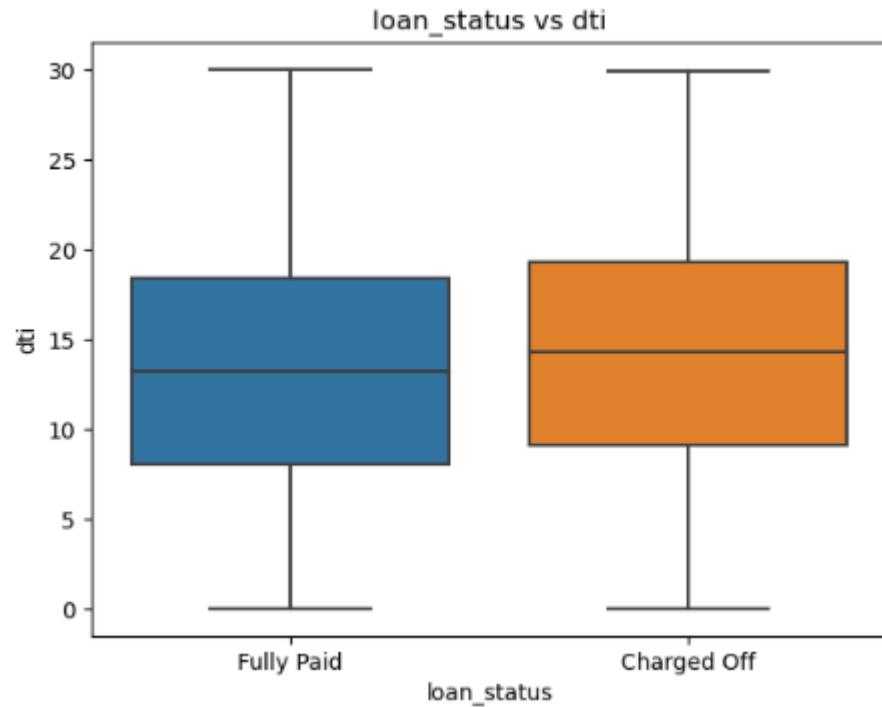


- Maximum Defaulters have taken loans based on following purpose: **debt_consolidation, credit card, home improvement and small business**

DTI as Target Variable

1.DTI vs Home Ownership

2. DTI vs emp length



- Applicants having high DTI
- Dti is almost same for the loan applicants across all the experience category

Recommendation

Lending club ought to reduce the term from 60 months to 36 months. They should have close watch on the borrowers who fall in below categories

- * Applicants who fall under **A and B grades**
- * Applicants who has more than **7 years experience**
- * Borrowers with **mortgage home ownership** .
- * People with more derogatory records and the one who have **high credit utilization rate**.
- * Borrowers who take loan for the following purposes:-
debt_consolidation, credit card, home improvement and small business