Lending Club Case Study

Ashlesha Dipak Ghule

Problem Statement

The aim of this lending club case study 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

Business objective

This company 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.

lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In this case, the customers labelled as 'charged-off' are the 'defaulters’.

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.

Business understanding

The main objective of this case study is to take a decision when the company receives a loan application whether to approve loan or not based on applicant’s profile.

Data attributes :

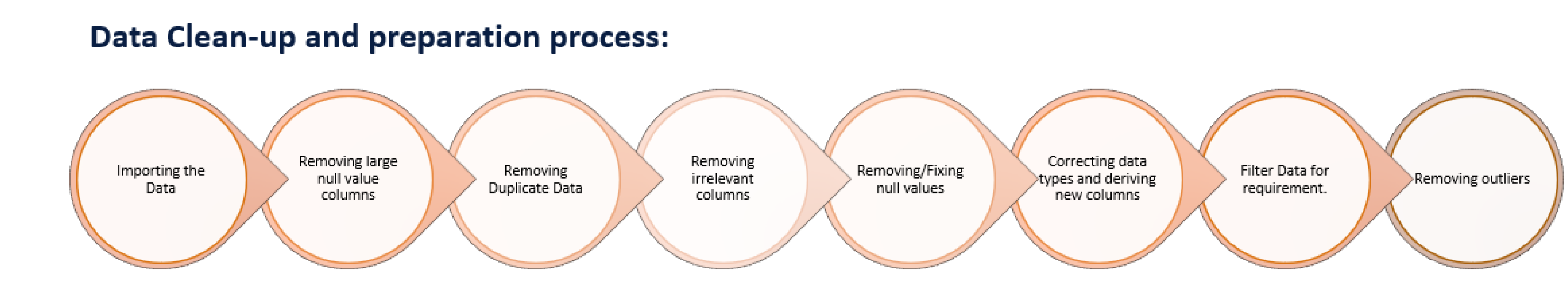
The loan data contains the data that describs if the applications received in past were defaulters or not. Also they are many attributes which helps in analysis purpose.

If the company approves the loan, there are 3 possible scenarios described below:

- **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
- **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
- **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan

Data preprocessing

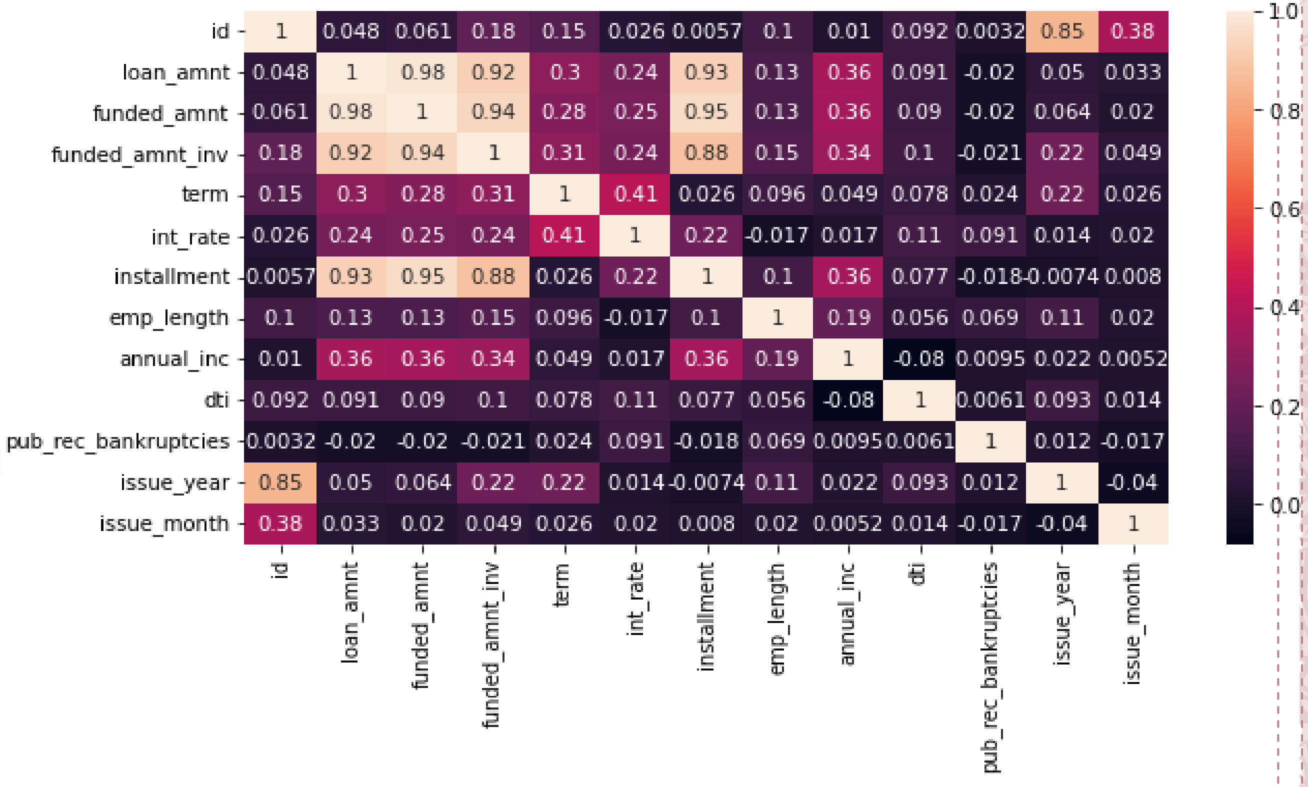
- Removing Null values >30 percent
- Dropping irrelevant columns for analysis
- Handling Null values
- Correcting the data and data types
- Treating outliers



Univariate and Multivariate analysis inferences

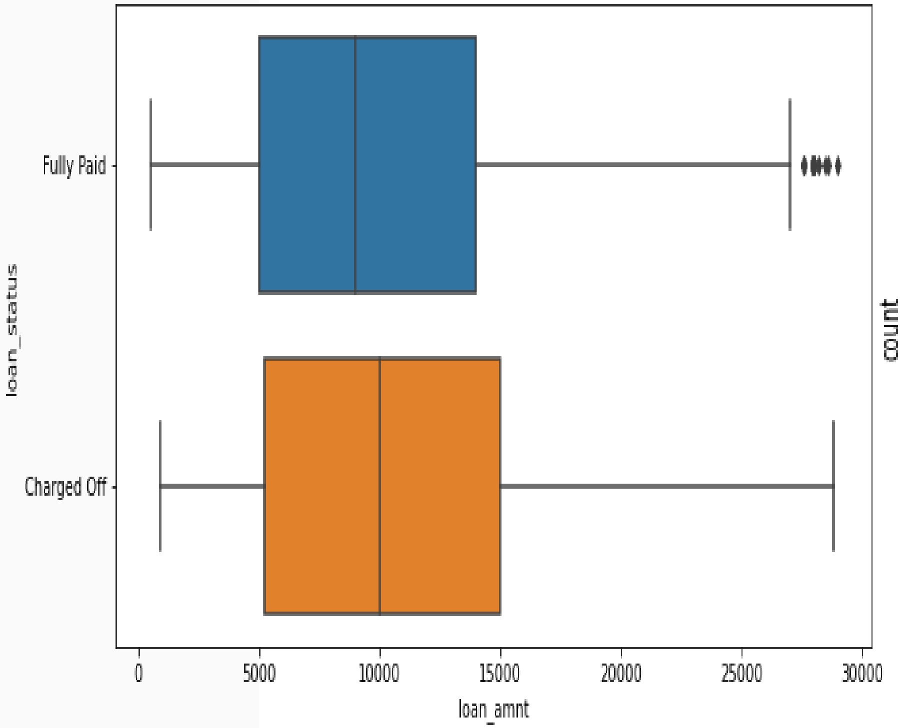
```
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   id                    34456 non-null  int64
 1   loan_amnt             34456 non-null  int64
 2   funded_amnt           34456 non-null  int64
 3   funded_amnt_inv       34456 non-null  float64
 4   term                  34456 non-null  int32
 5   int_rate              34456 non-null  float64
 6   installment           34456 non-null  float64
 7   grade                 34456 non-null  object
 8   sub_grade             34456 non-null  object
 9   emp_length            34456 non-null  float64
 10  home_ownership        34456 non-null  object
 11  annual_inc            34456 non-null  float64
 12  verification_status   34456 non-null  object
 13  issue_d               34456 non-null  datetime64[ns]
 14  loan_status           34456 non-null  object
 15  purpose               34456 non-null  object
 16  addr_state            34456 non-null  object
 17  dti                   34456 non-null  float64
 18  pub_rec_bankruptcies  34456 non-null  float64
 19  issue_year            34456 non-null  int64
 20  issue_month           34456 non-null  int64
```

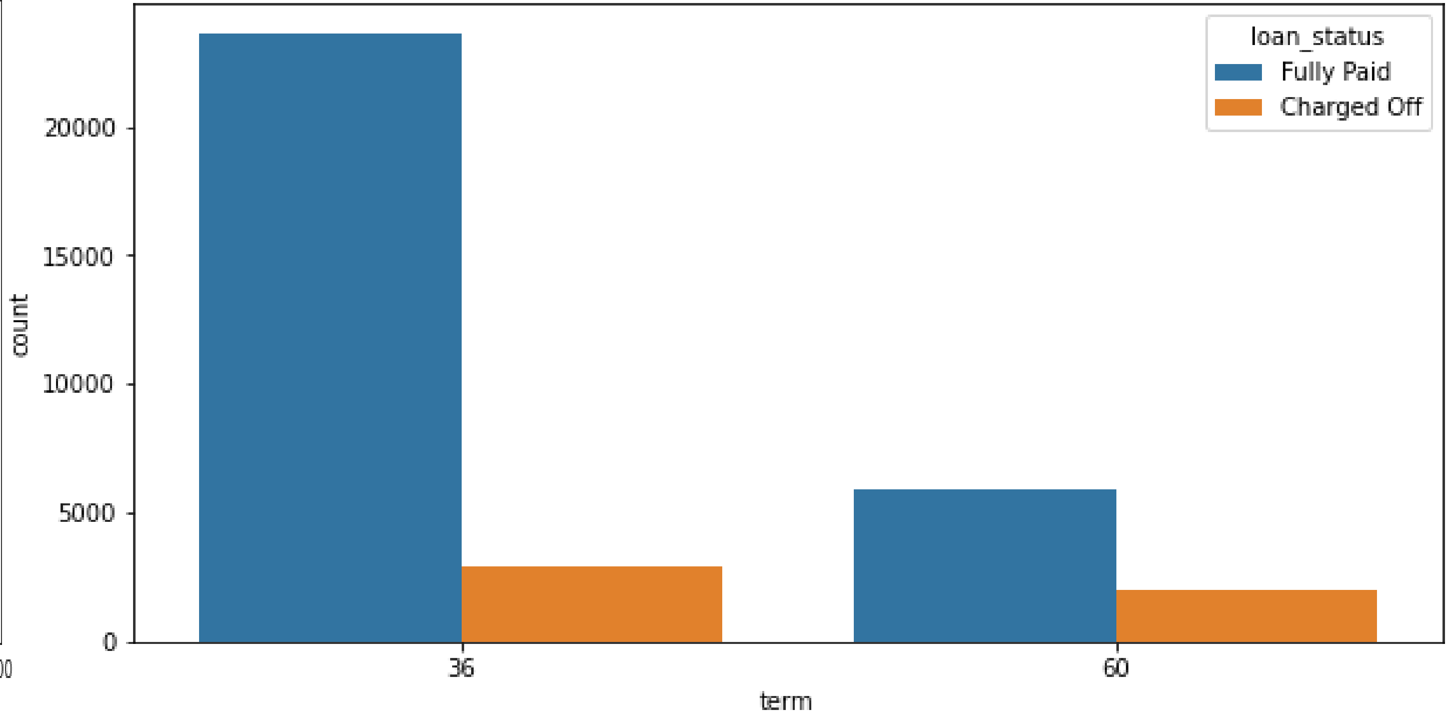
Final columns after the data cleaning



Correlation Matrix

loan status vs loan amount and term

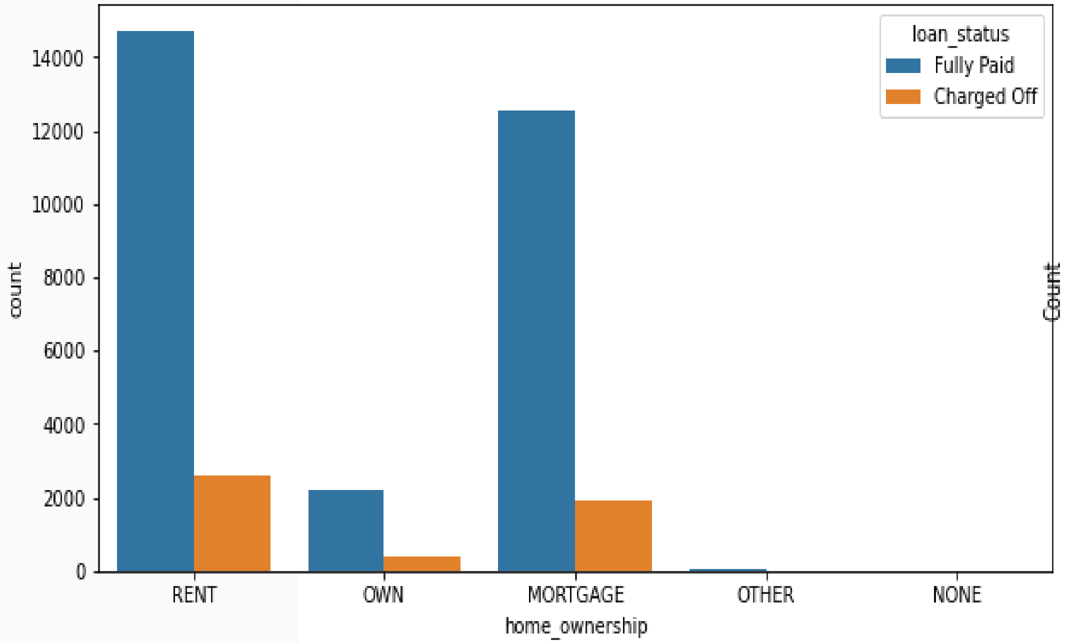


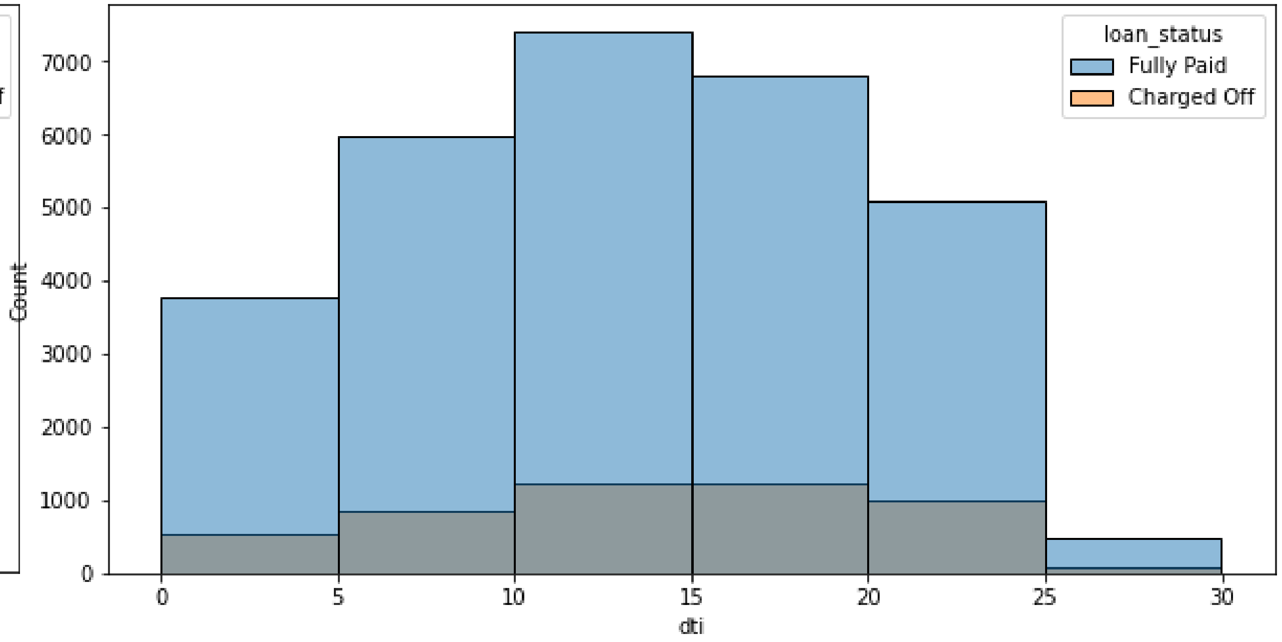


Higher amount of loan have high chances of defaulting.

Compared to the term month as 60, there are more defaulters in the term as 36 months.

Loan status vs DTI and home ownership

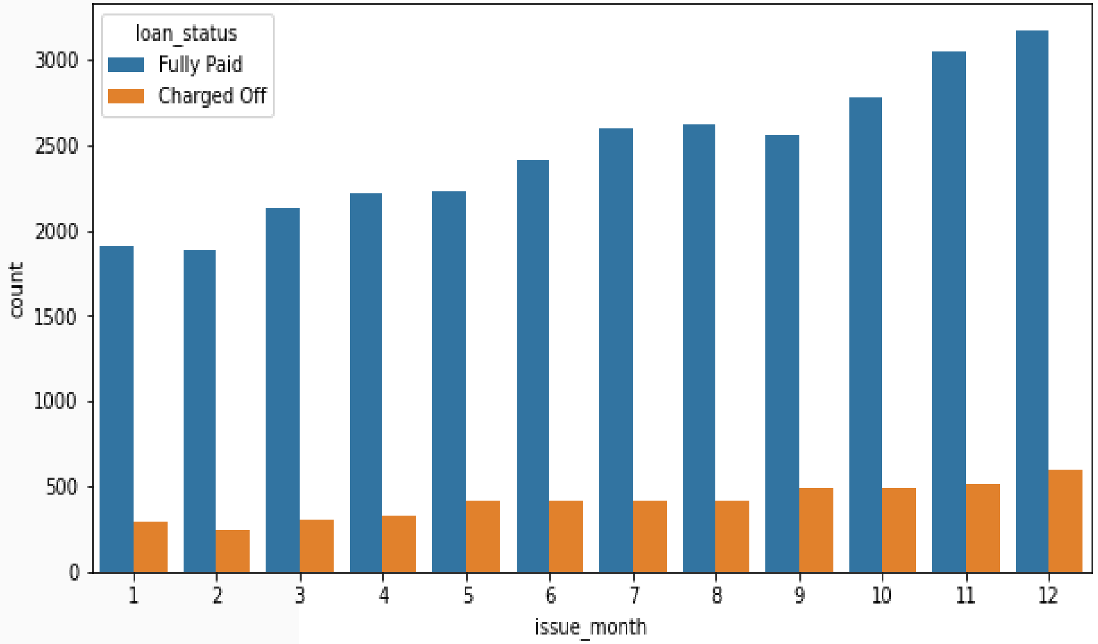


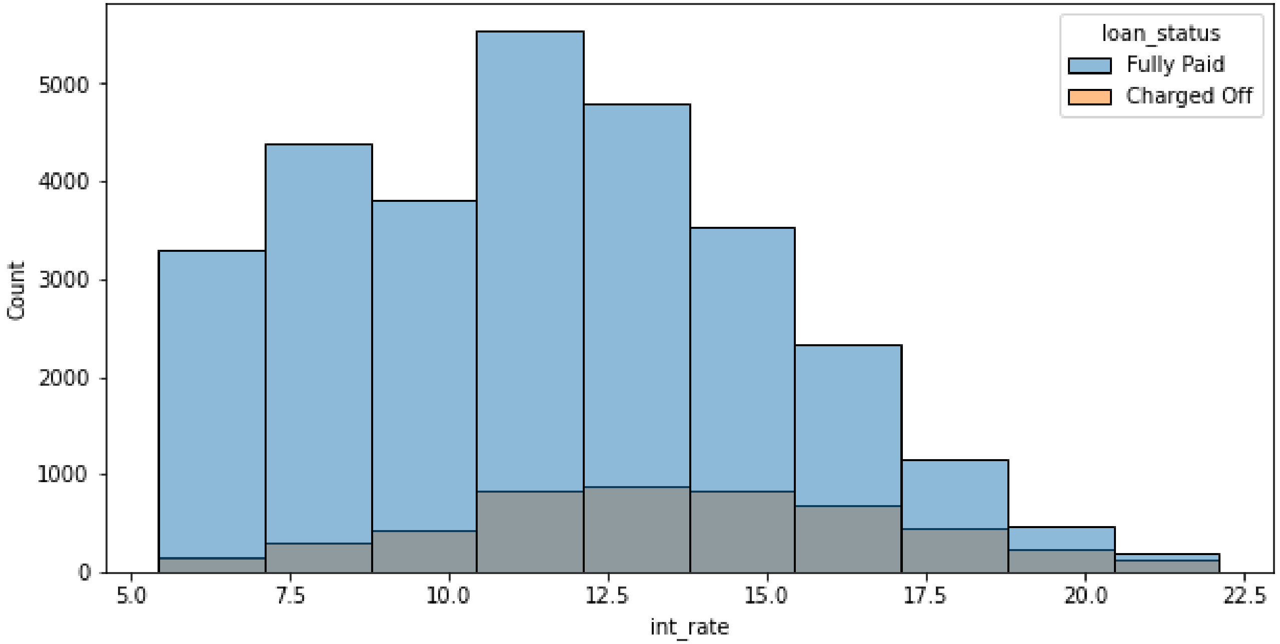


Borrowers with home ownership as rent and mortgage are more likely to not repay the loan compared to the others.

Borrowers with DTI ratio between 10 to 20 are more likely to be a defaulter who may not repay loan. In the other side the less dti ratio can be preferred.

Loan Status vs issue month and interest rate

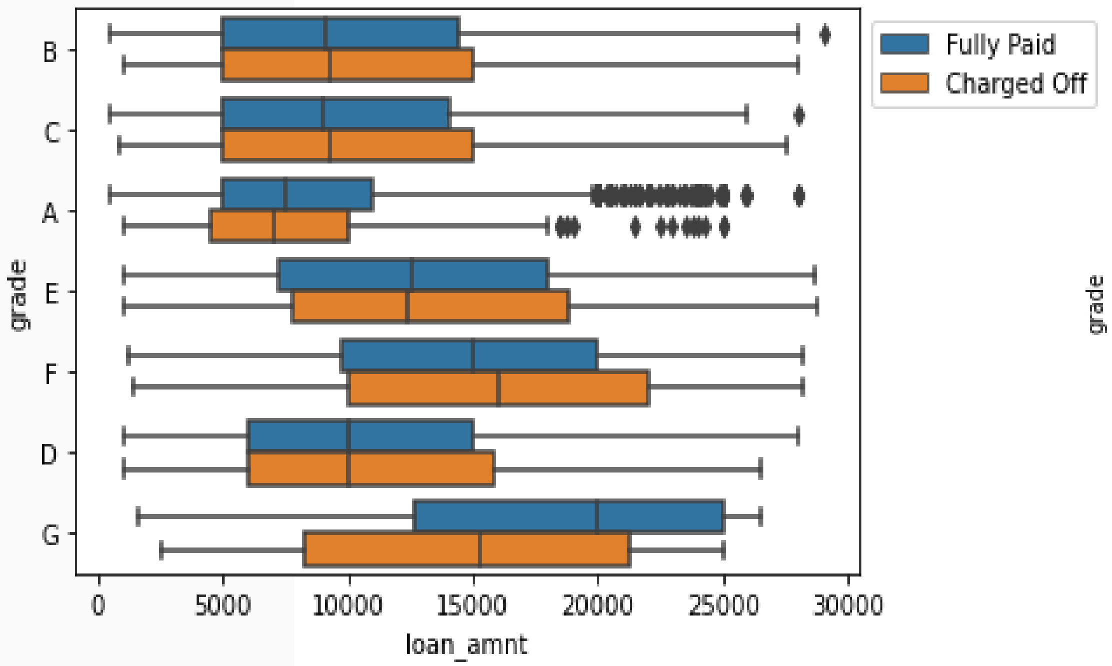


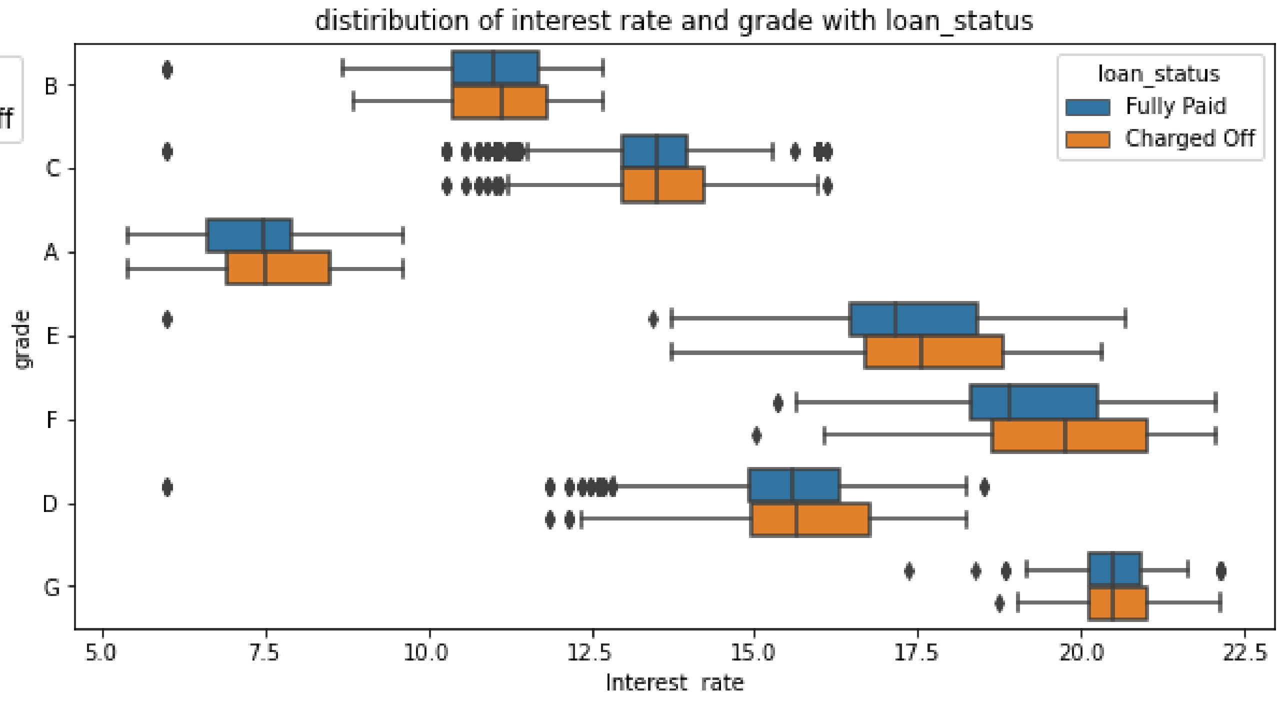


Majority of loan which are given in last quarter are having more number of borrowers as defaulters

If the interest rate is higher the charged off rate is also higher.

Analysis for loan amount, grade, loan status and interest rates

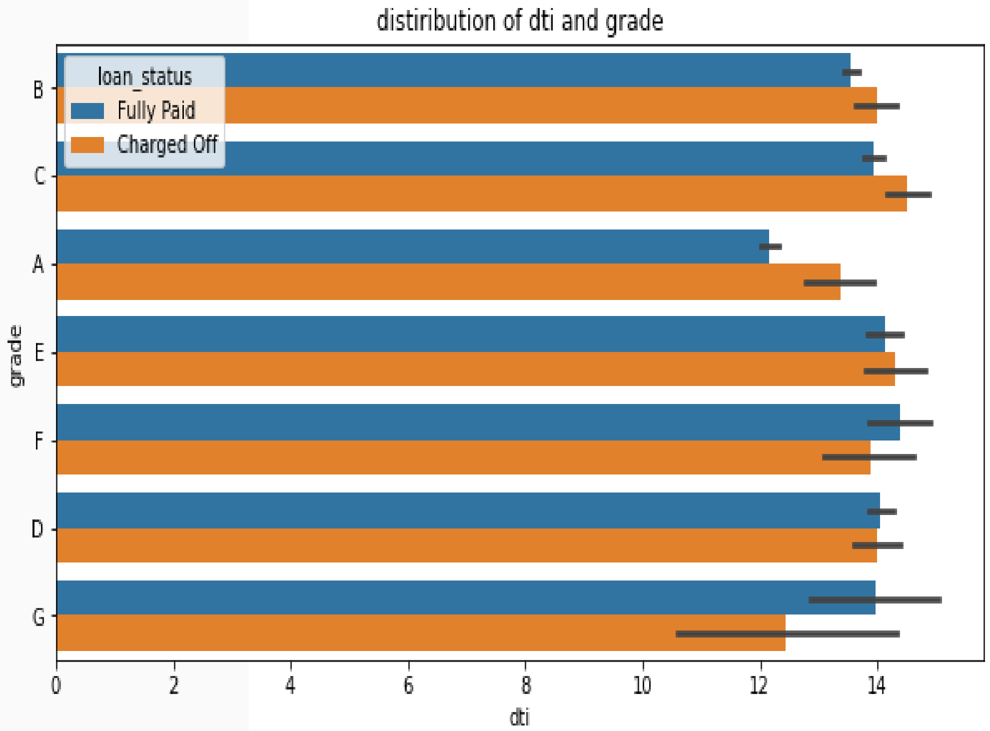


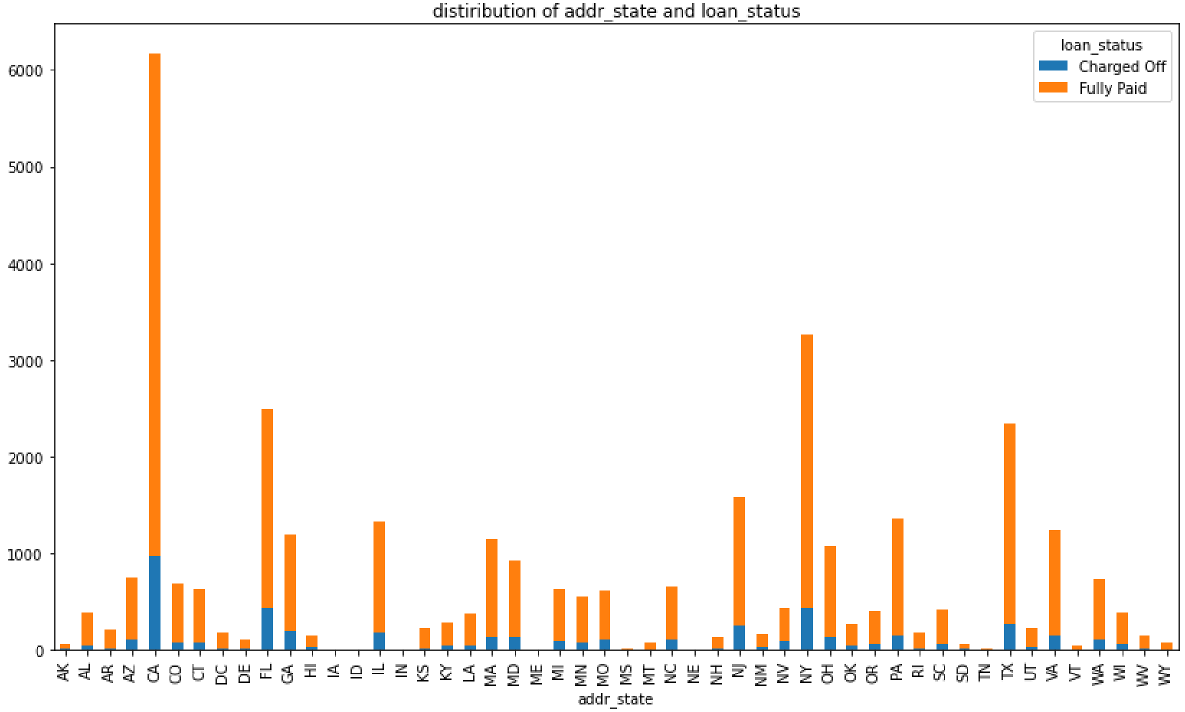


With higher grade & higher loan amount and with lower grade & lower loan amount, the borrower are less likedly to charged off.

As the grade is increasing from A to G the interest rate also increases and with the chances of charged off also increases.

Analysis of grade, DTI, address state and loan status

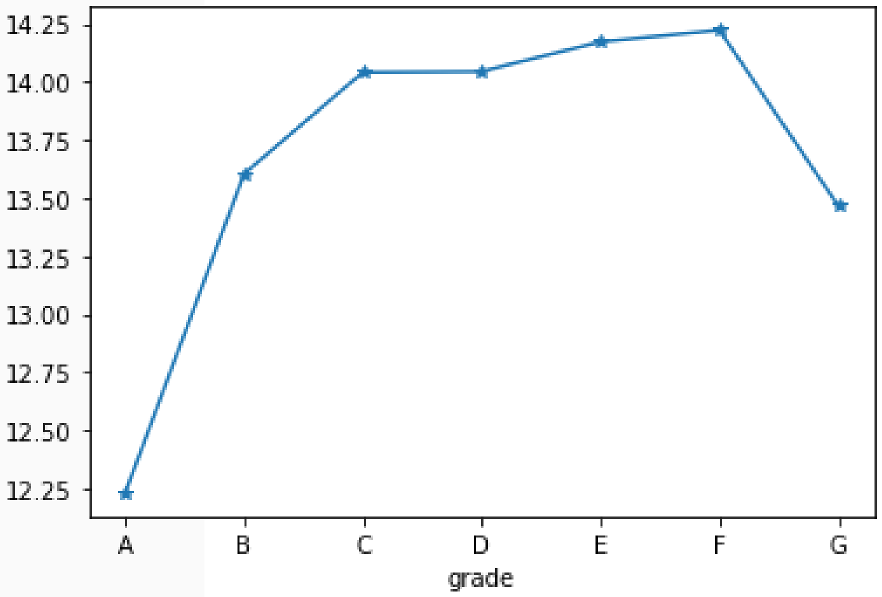


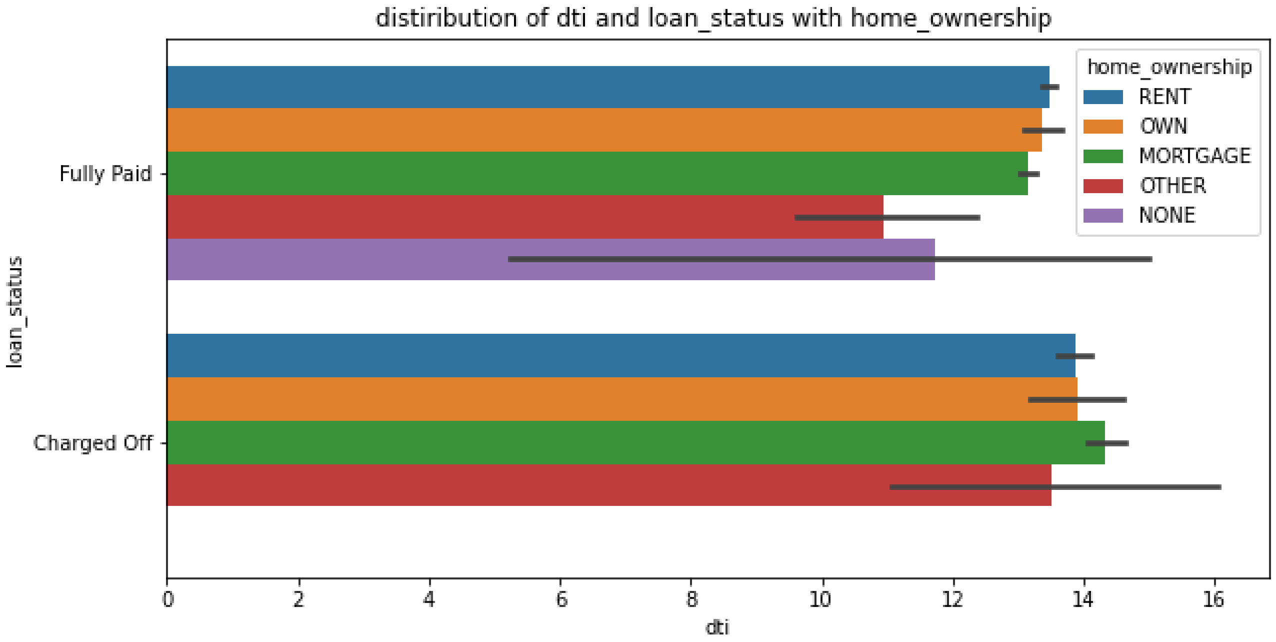


C grade is having maximum chances of charged off borrowers with highest DTI ratio and lowest of charged off borrowers in G grade

California(CA) state shows more number of charged off borrowers and also fully paid borrowers.

Average DTI VS grade and loan status vs homeownership

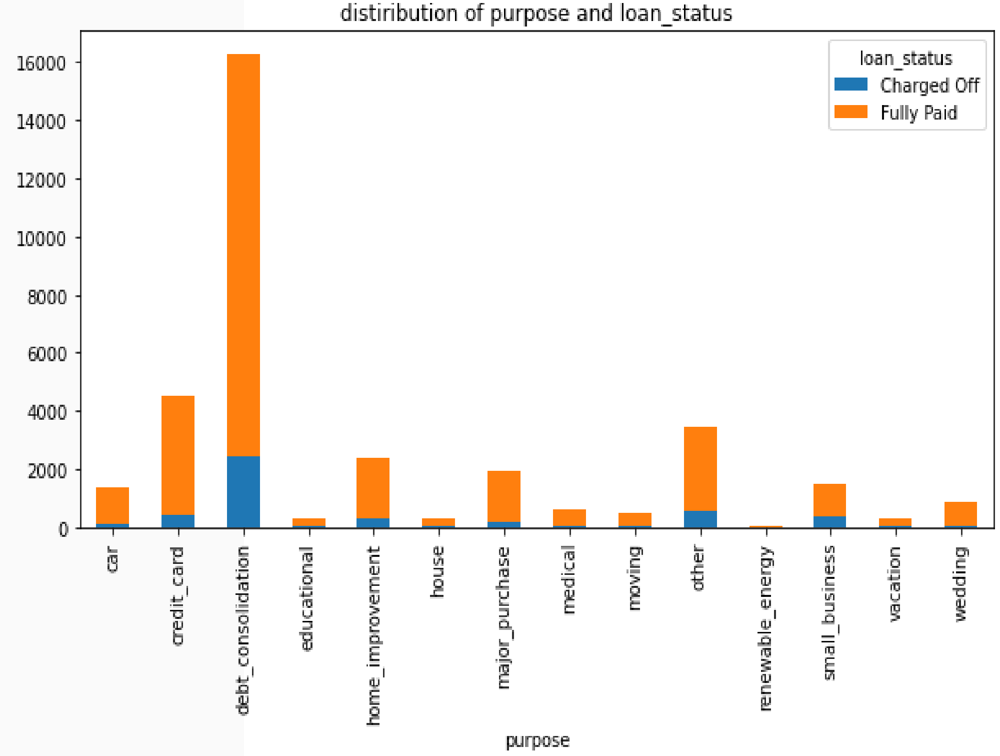


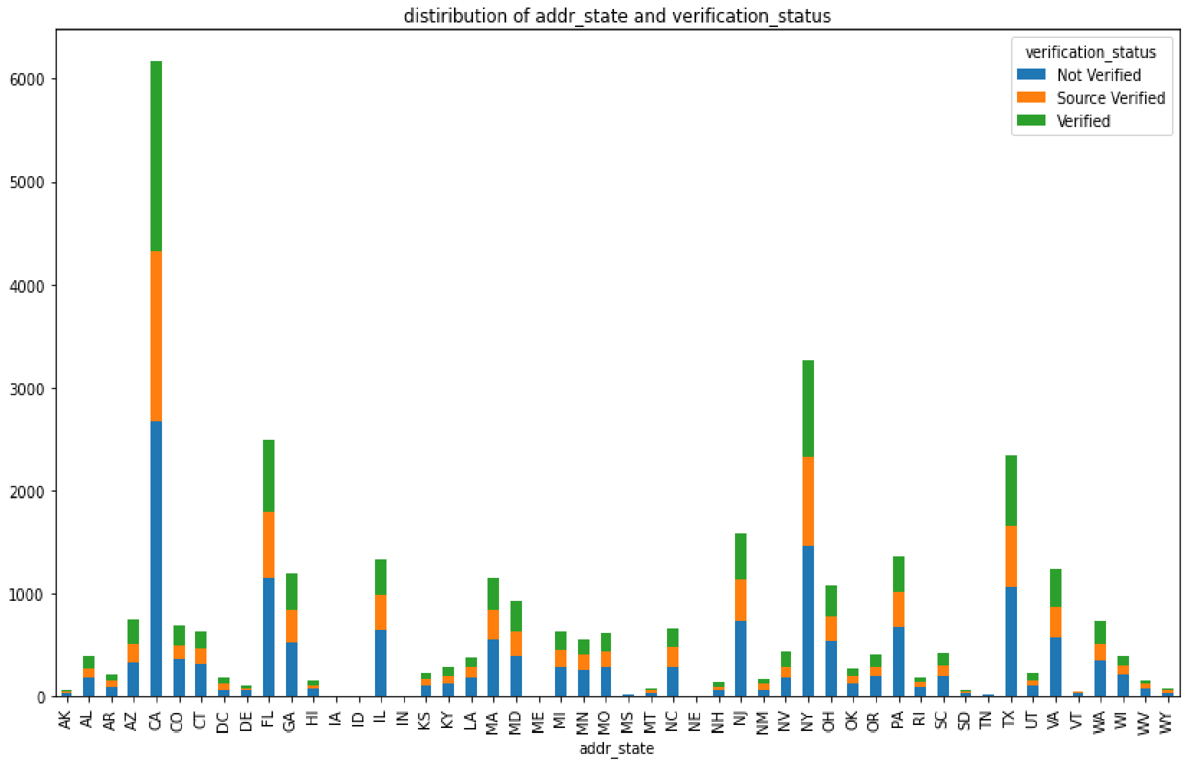


The avg DTI is increasing with each grade till F but decreasing from F to G. As the grade increases from A to higher the risk also increases

Defaulters are more likely be from the mortgage home ownership and have high DTI ratio

Loan Status vs address state and purpose





Defaulters are taking loans for the smaller bussiness and debt_consolidation purpose.

In California(CA),Florida(FL)Newyork(NY),Texas(TX) , most of the borrowers are not verified.

Recommendations to reduce defaulters

Driving Factors
DTI
Verification status
Interest rates
pub_rec_bankrupties
Grades

- Applicants should not have high DTI ratio.
- Applicants should not be from California(CA), Florida(FL), New York (NY) and Texas (TX) states and applications should be verified properly.
- Interest rate should not be higher as that increases charged off.
- public record of bankruptcies for applicant’s should be checked properly.
- Grade should be higher for the loan approval as that means low risk.

Thank You