



Lending Club Case Study

Loan data analysis

By: Soumyadeep Dutta

The problem

Company

Consumer finance company which specialises in lending various types of loans to urban customers.

Context

The loans can have two fates

- Customer likely to repay the loan which brings profit to the company
- Customer is not likely to repay the loan which in turn bring a financial loss to the company.

Problem statement

Analyse and understand the consumer attributes and loan attributes influence on the tendency of default.

Challenges deep-dive

Attributes Analysis

Data Loading

The company has a huge customer base and all the necessary information needs to be loaded properly to analyse

Analyse

- All the information provided regarding a loan application needs to be checked properly as some data is not relevant for further analysis.
- Data needs to be modified in a proper format which is helpful for a insightful analysis.

Data Cleanup

- Irrelevant Loan/Consumer attributes removal.
- Empty attributes removal.
- Data shaping for a good analysis outcome.
- Removal of Out of the box data which may impact the analysis otherwise.
- Populate missing values if possible.
- Derive new attributes as required.

Analysis

Univariate Analysis

To understand the loan and consumer specific attributes

Bivariate Analysis

To analysis the dependency of the attributes on loan status

Solution

Attribute level analysis is done to identify the attributes which are affecting the loan status.

Univariate and bivariate analysis on cleaned and shaped data brings us the valuable information regarding the loan data.

Implementation

Data Attributes Analysis

Data Shape: 111 columns/Attributes & 39717 records.

- Attributes were analysed properly with attribute definition to understand the importance of each attribute and what value it brings to the data set.
- Attributes with unique values captured as it is not adding values to data set.
- Introduce new attributes
 - LoanToIncomeRatio: This is a ration between approved loan amount and income amount of customers.



Data Cleanup

- Attributes having no or very little amount of data is removed.
- Attributes having no meaningful information are removed from dataset.
- Attributes with unique values are removed
- Attribute with very few missing values are populated through proper data analysis.
Depending on the data variance, we have substituted the null values either by mean or median value.
- Change the datatype of the columns as required.



Analysis

Analysis is done on different attribute levels to identify the importance of each attribute and how it is associated with the corresponding loan status.

Attributes are categorized into two types.

1. Categorized attributes
2. Numeric attributes

Different data analysis graphical methods are used based on the type of the attributes.

- Univariate Analysis
 - Categorical attributes
 - Countplot Graph
 - Numeric or Continuous attributes
 - Histogram
 - Boxplot
 - Bivariate Analysis
 - Between different attribute types.
 - Boxplot
 - Stacked Bar Graph
-



Univariate Analysis Insights

Univariate Analysis

- Loan Amount:
 - LC generally disbursed loan amount between 10-10.5K dollars for max number of people.
- Interest Rate:
 - People on average got the loan around 12% interest rate.
 - Highest number of people got at 7.5% interest rate.
- Installment:
 - Maximum people are paying installment between 171-185 dollars.
 - Lot of borrower are paying very high amount of interest rate which is more than 700 dollars.
 - On average borrower pays 285 dollars on installment.
- Annual income of Loan Takers:
 - On average 50% of loan takers are earning 51k dollars.
 - Highest proportion of borrower who took loan are on salary range of 58 to 60k.
- DTI:
 - Average dti is around 13
- Public record bankruptcies:
 - Max people do not have public record of bankruptcies.



Univariate Analysis Cont....

- Delinquency in last 2 years:
 - Very few people have 1-2 delinquency in past 2 years.
 - Highest 11 delinquencies are visible in dataset for a borrower.
- Open Accounts:
 - On average 8 credit lines are open in borrower's credit file.
 - Maximum borrower are having 3-15 credit lines open
- Total credit revolving balance:
 - Maximum borrower have 0 credit revolving balance.
 - 50% borrower have around 7855 credit revolving balance.
- Revolving line utilization rate
 - 50% borrower have around 48% utilization rate
- Total Accounts:
 - On Average every borrower have 20 credit lines in credit file.
- Inquiries in Last 6 months:
 - Max borrower did not enquire about it in last 6 months.
- Number of derogatory public records:
 - Max people have 0 derogatory public records



Univariate Analysis Cont....

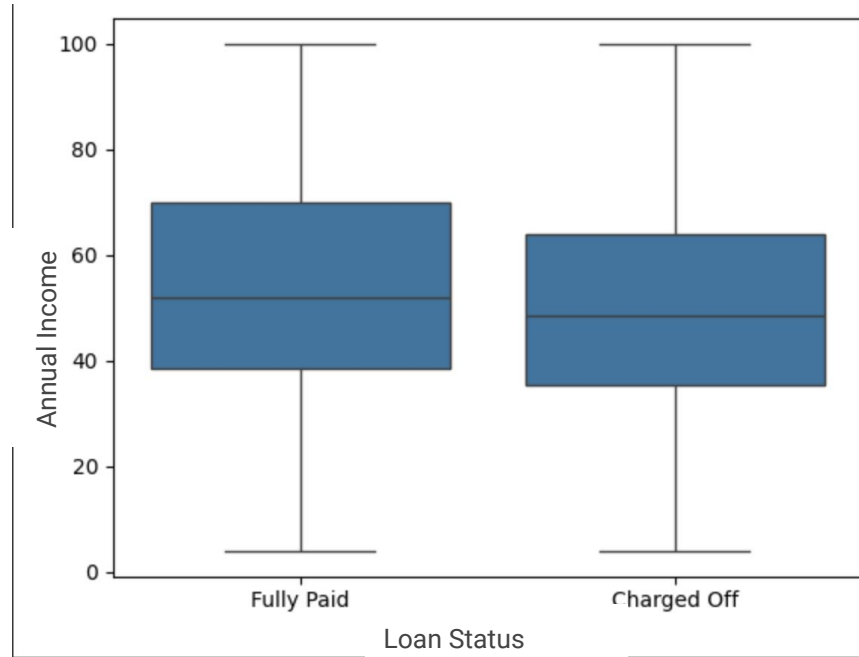
- Loan Term:
 - Maximum people took 36 months of loan term.
- Loan Grade:
 - Maximum loans are disbursed on B grade.
- Loan Sub Grade:
 - Maximum loans are disbursed on A4 sub grade.
- Home Ownership
 - Maximum borrower are staying on rented home.
- Loan Verification Status:
 - Maximum borrowers are not verified.
- Purpose of Loan:
 - Maximum people took loan for debt consolidation.



Bivariate Analysis Insights

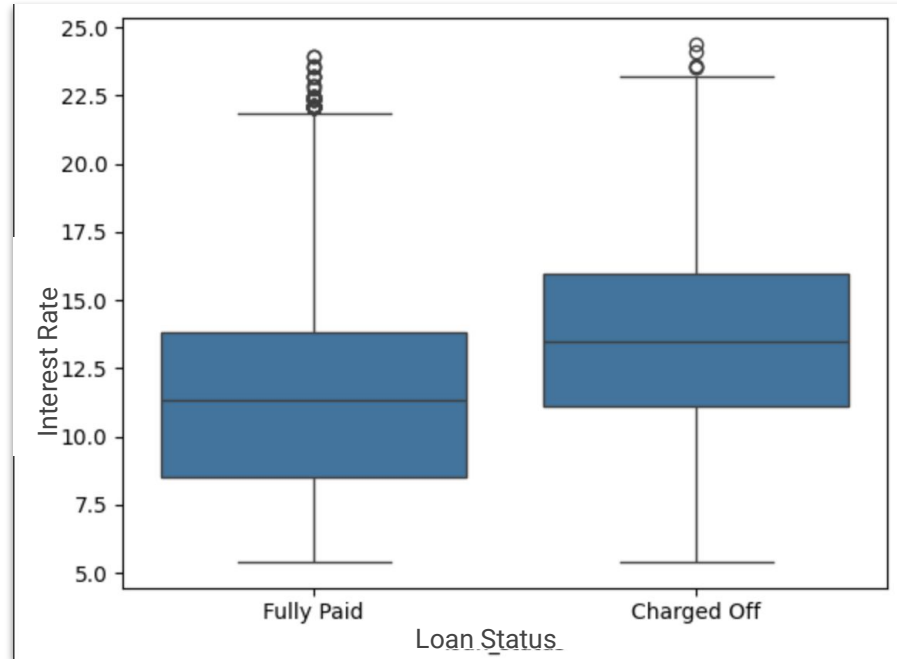
Dependency of Annual Income on Loan status

Borrowers who got charged off have less annual income than the borrowers who paid the loan fully.



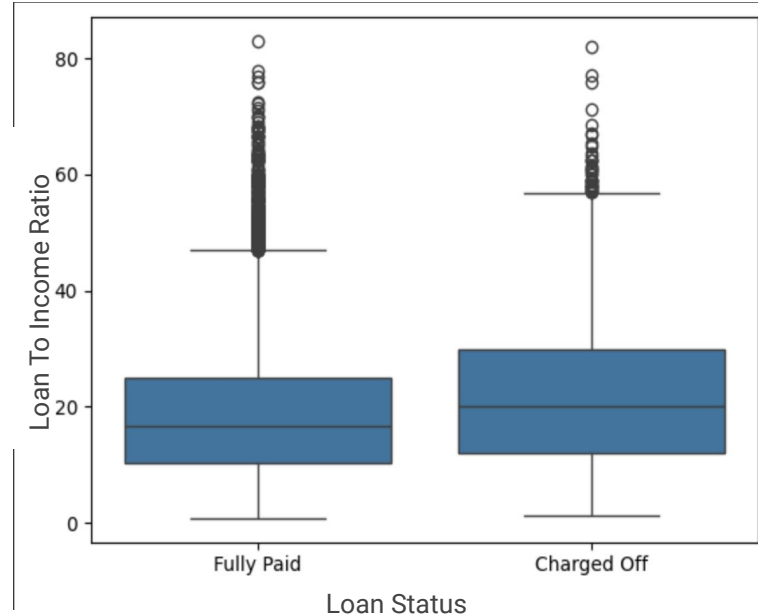
Influence of Interest Rate on Loan status

Borrowers who took loan on high interest rate have higher chances of Charge off



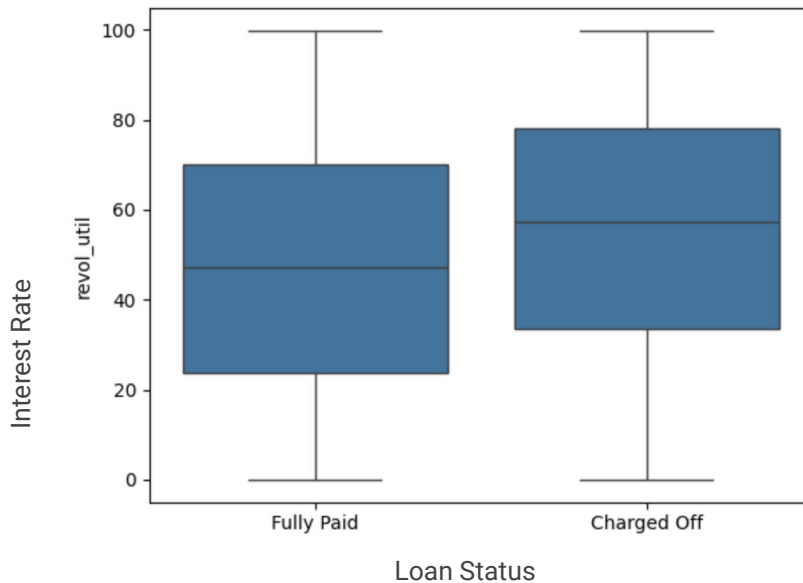
Influence of Loan To Income Ratio on Loan status

It is observed that the borrowers of charged off loans have on average high **loan to income ratio** than the borrowers who paid the loan amount fully.



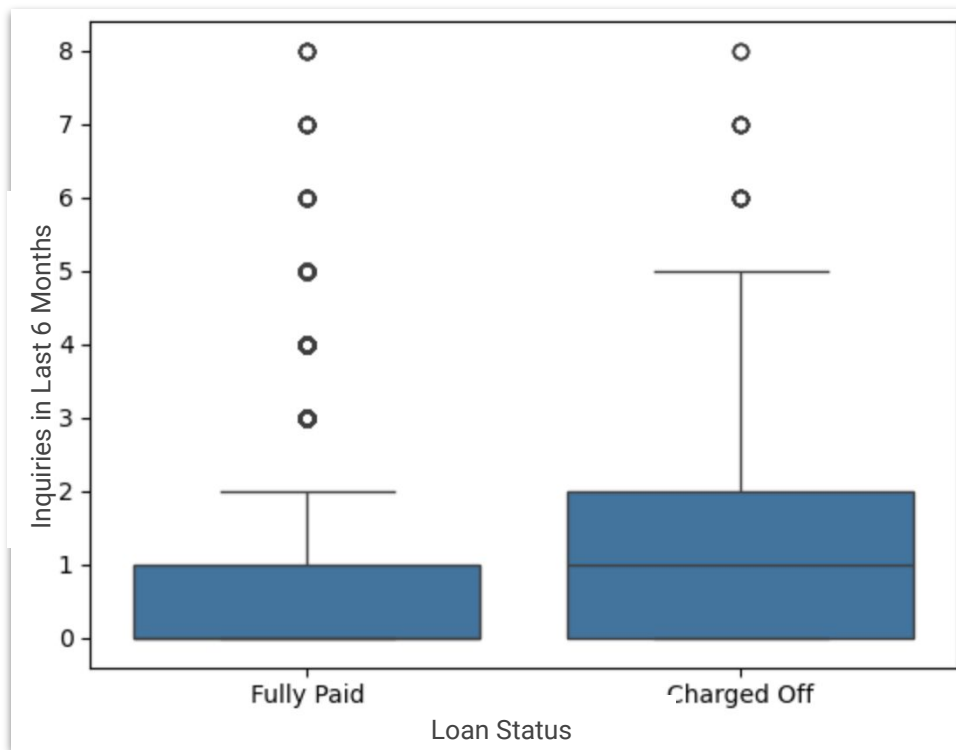
Influence of Revolving line utilization rate on Loan status

It is observed that the charged off borrowers have high revolving line utilization rate.



Dependency of Inquiries in past 6 months on Loan status

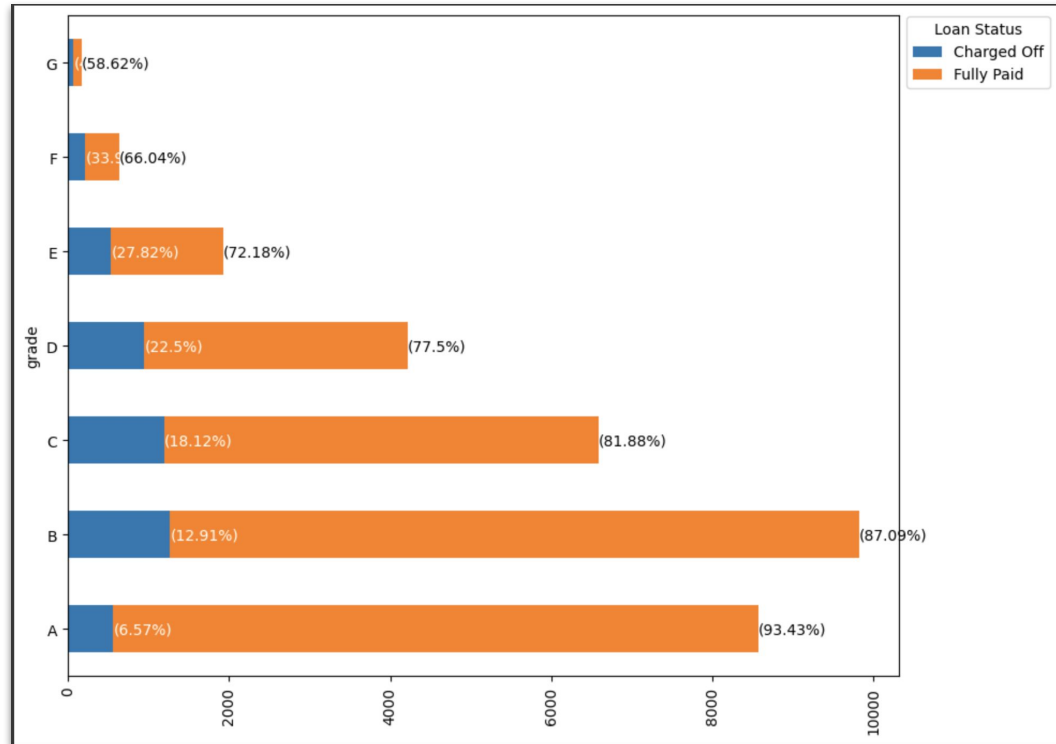
Borrowers who got charged off enquired more in last 6 months.



Influence of Loan Grades on Loan status

Loans which got disbursed on G grade has max chances of Default.

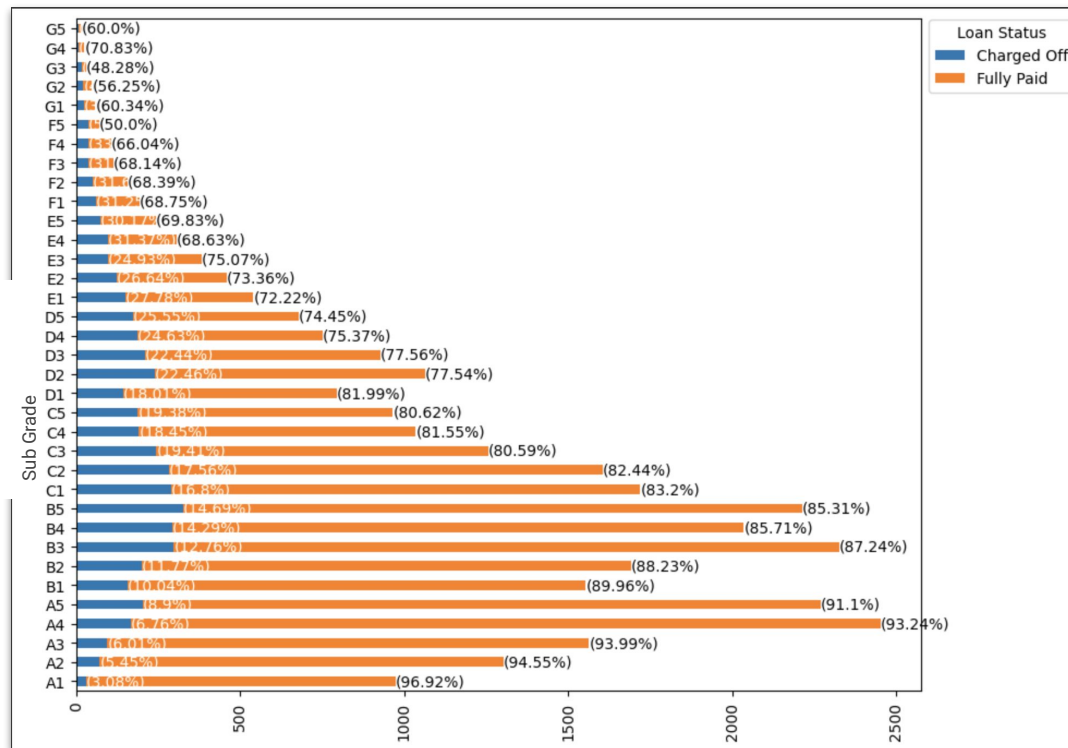
41.38% loans disbursed under G grade got charged off.



Influence of Loan Subgrades on Loan status

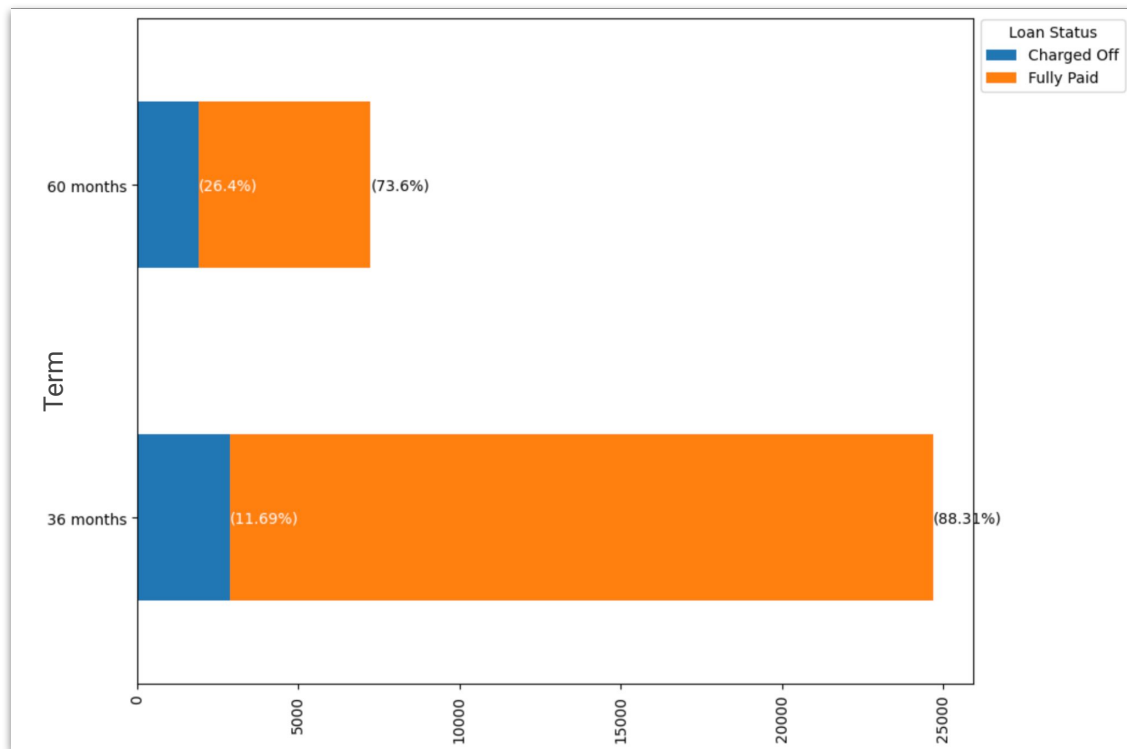
Loans which got disbursed on G grade and very specific to G3 subgrade has **48.38%** success ratio.

That means 51.62% borrower who got loan under G3 does not pay the full amount.



Influence of Loan Term on Loan status

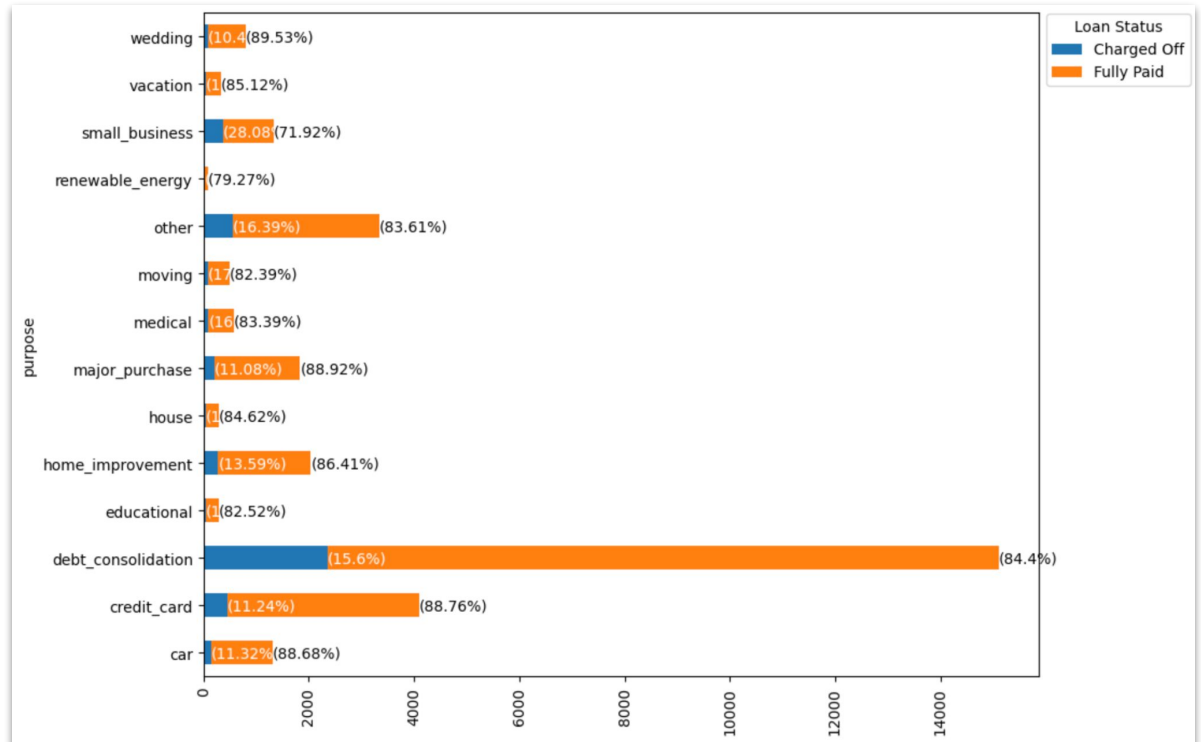
It is observed that 26.4% borrower defaulted who took loan on 60 months term whereas only 11.69% people defaulted on 36 months loan term.



Influence of Purpose on Loan status

Loan taken related to small businesses has higher chances of charge off.

Almost 28.08% loans were charged off according to the analysis.

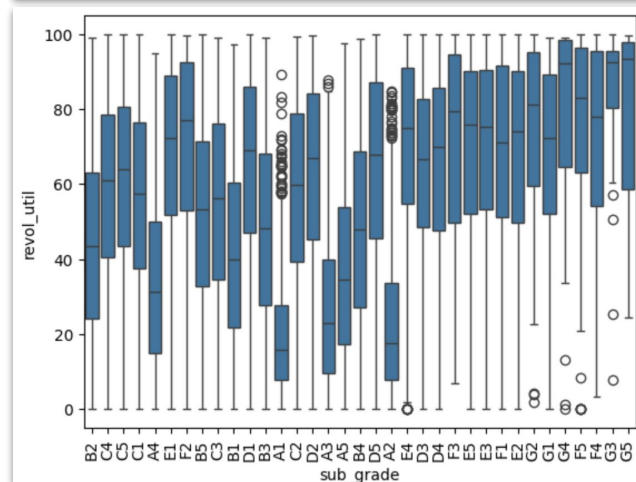
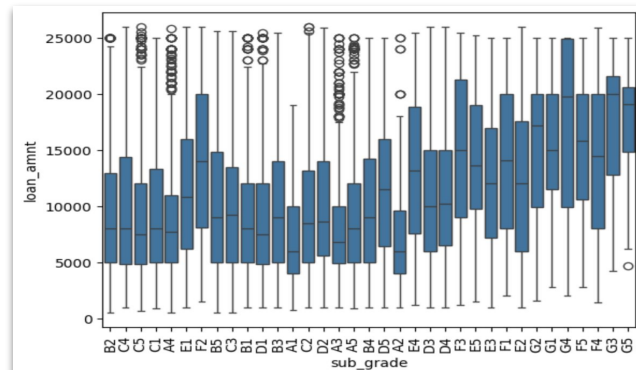
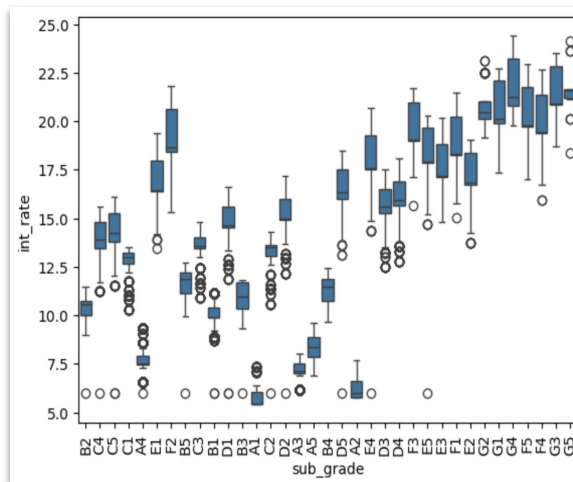
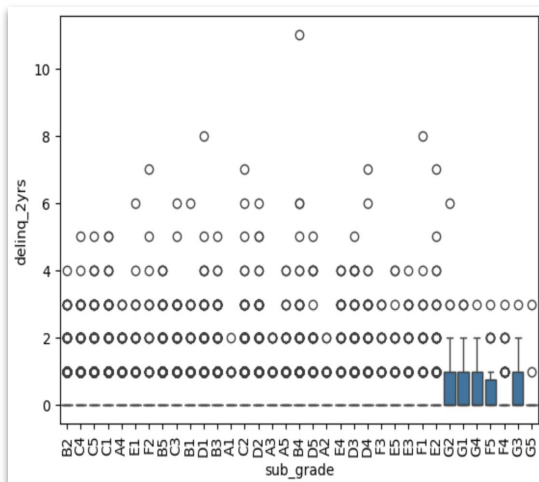


Detailed Findings

Influence of Subgrade, Loan amount, Interest Rate, Revolving line utilization rate on Loan status

As it is already observed that loans on subgroup g3 has highest rate of charge off.

1. G3 subgroup has few of the highest approved loans
2. Has the 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
3. Has one of the highest amount of interest rate.
4. Loans on G3, has one of the highest revolving line utilization rate.



Overall Insights

Borrowers

1. who have less annual income and takes loan on high interest rate
2. who have enquired more than one time in average in last 6 months
3. Who took loan for small business purpose

have higher chances of charge off.

Generally these types of loans are disbursed under "G" grade only.

