Lending club case study

Created by : Jagdish & Mahesh

Problem statement

A consumer finance company wants to accesses the loan application and find out the chances of default.

Past data with many parameters are available for analysis.

Based on this dataset important parameters driving the loan defaults needs to be identified.

Solution approach

- Read the data in python file
- Data Exploration and Cleaning:
  - ✓Check the missing values
  - ✓Handle missing values through imputation or data removal
  - ✓Check for data types and convert if required
  - ✓Analyse the data and identify the parameters impacting defaults
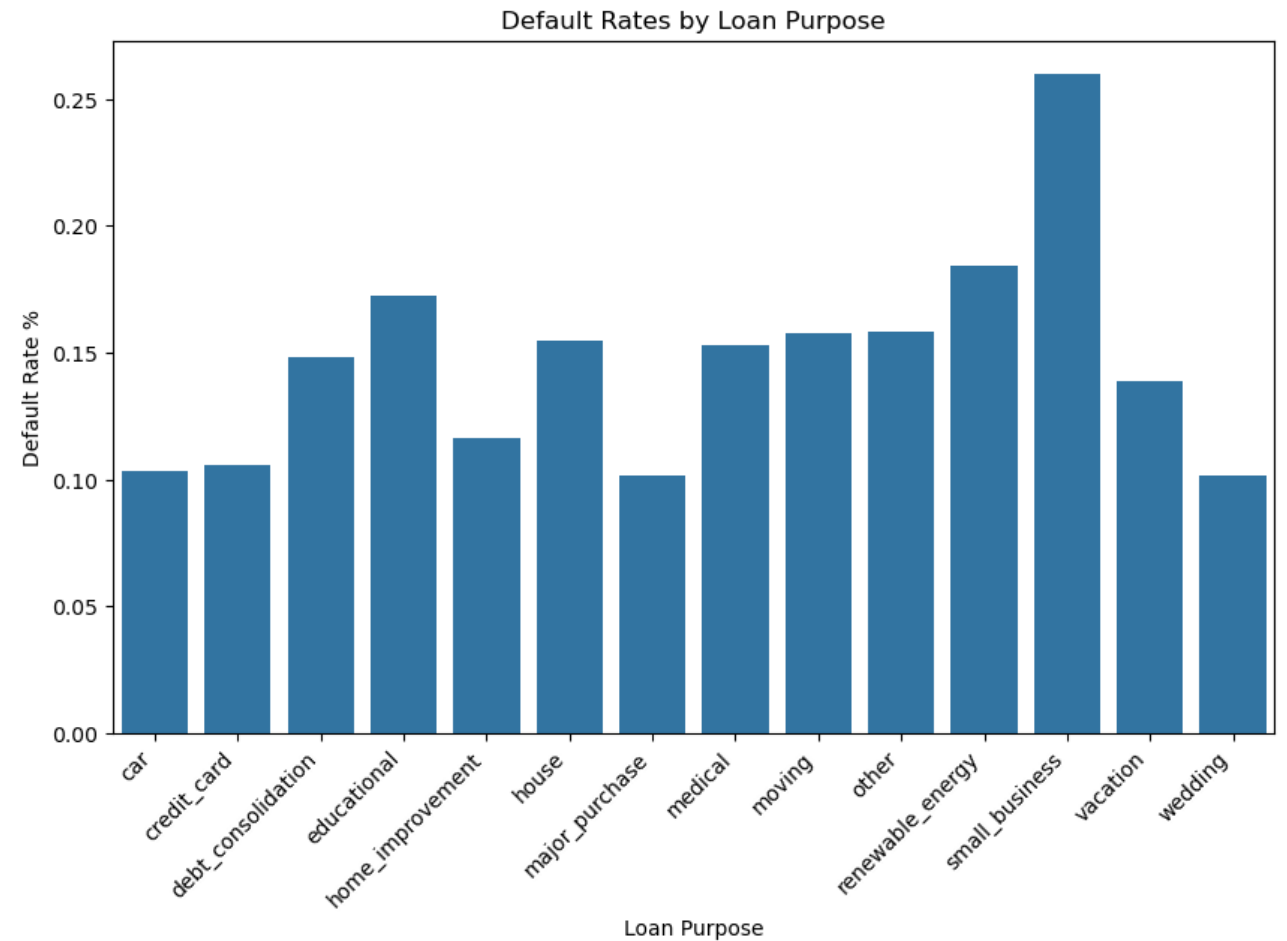- Plot the data for visual representation

Purpose of loans

Customers request loan for various purposes . Graph shows the count of loan applications based on purpose
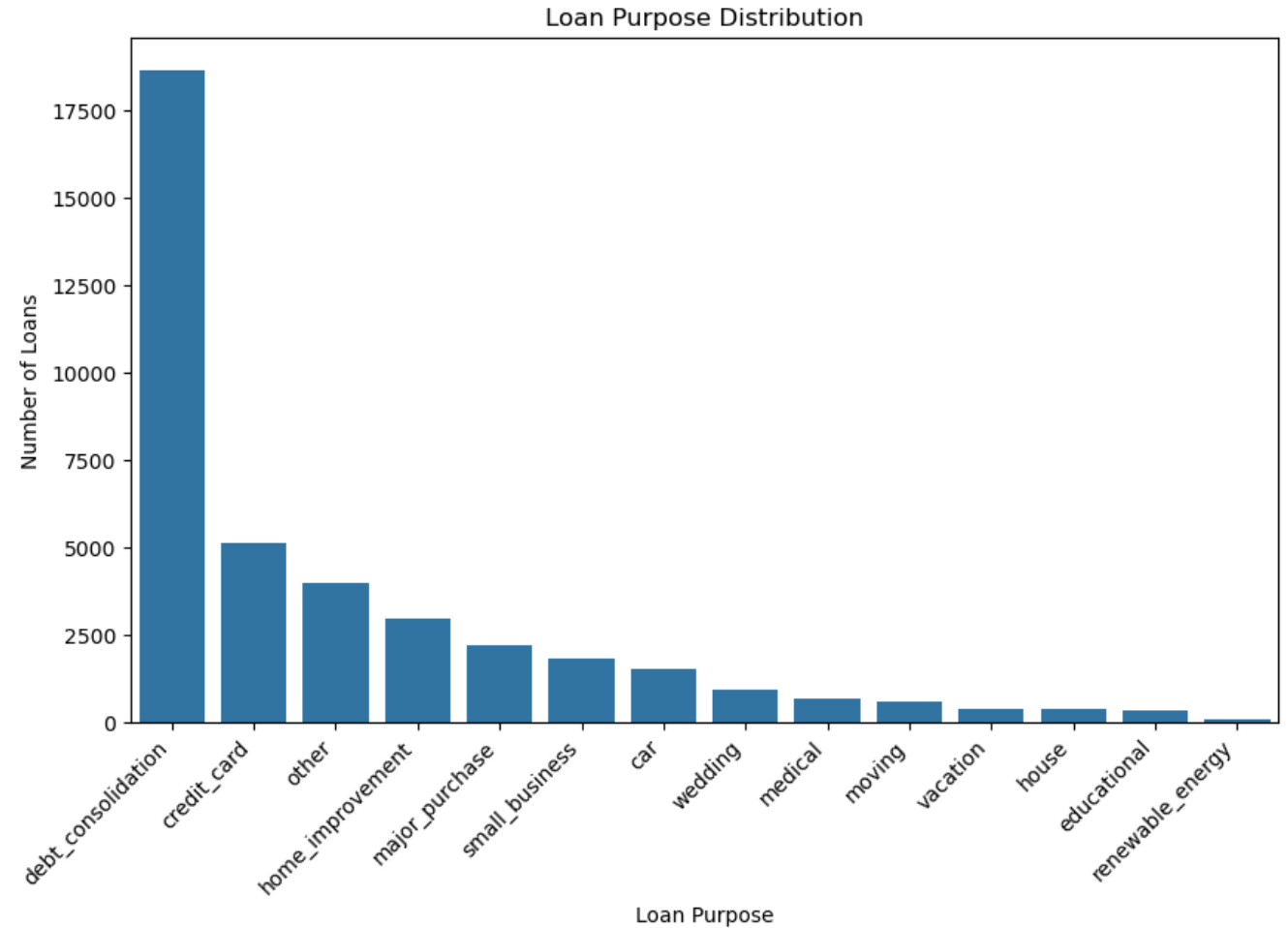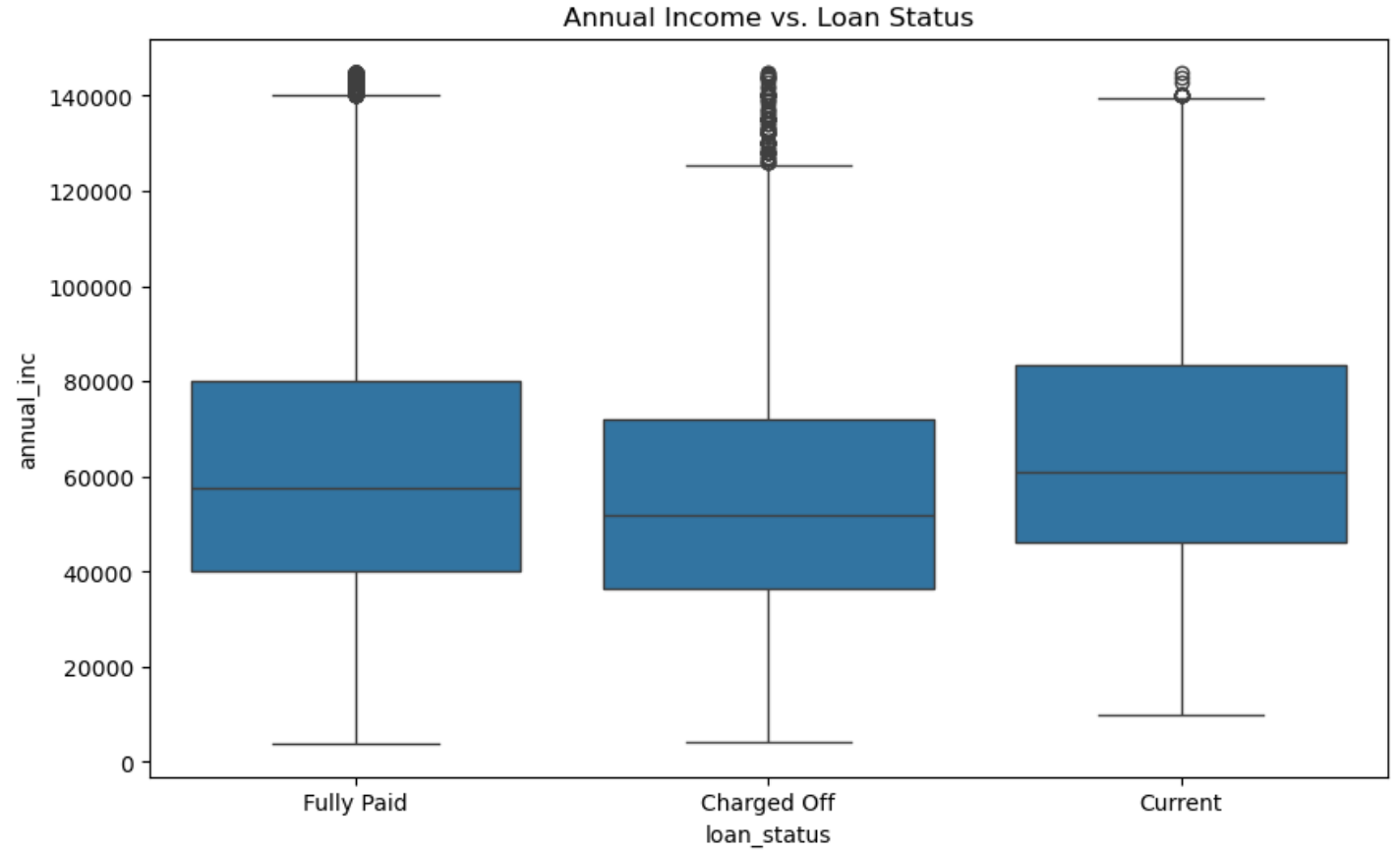
Default rates based on loan purpose

Loans taken for small_business carries higher risk Of defaulting.

Default Rates by Loan Purpose

Loan status vs annual income

Defaulters have average annual income lesser than fully paid category by 1000$.
Lower annual income indicates higher default rate
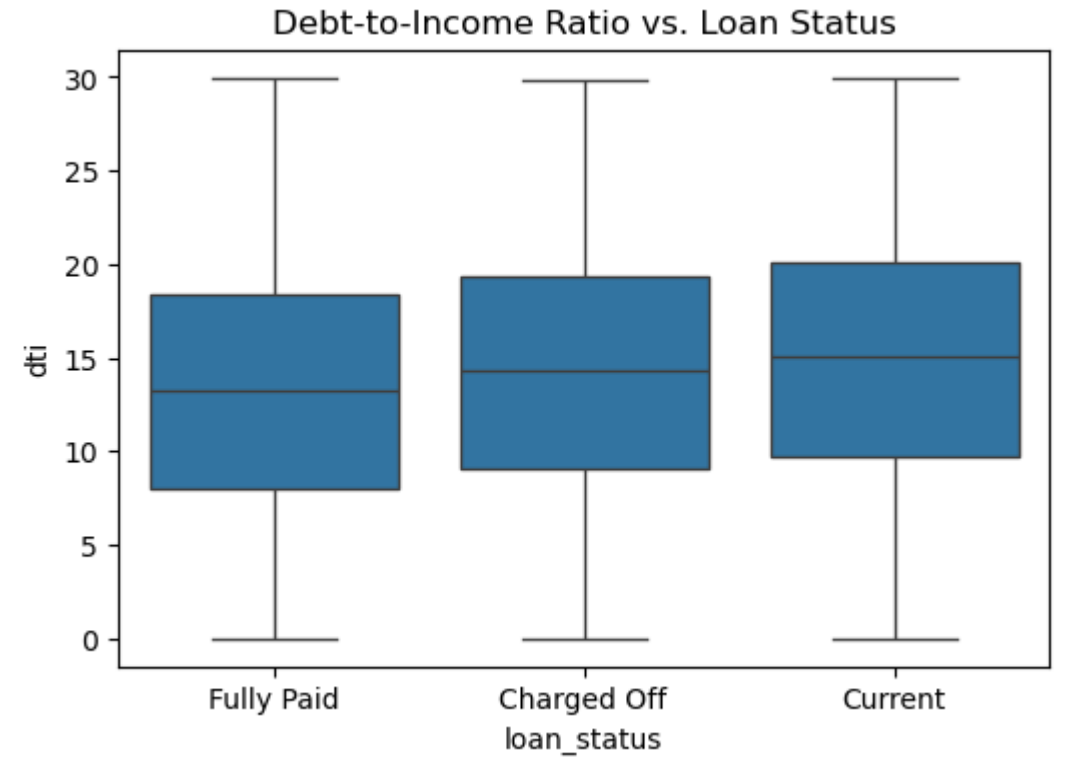

Annual Income vs. Loan Status
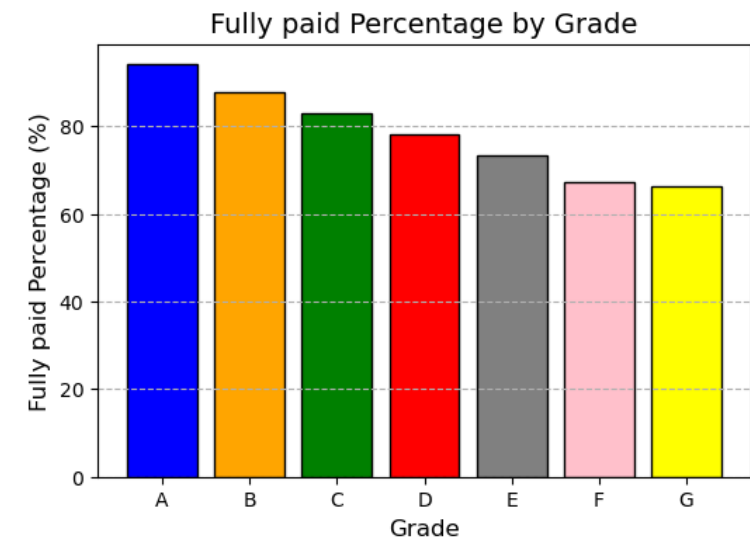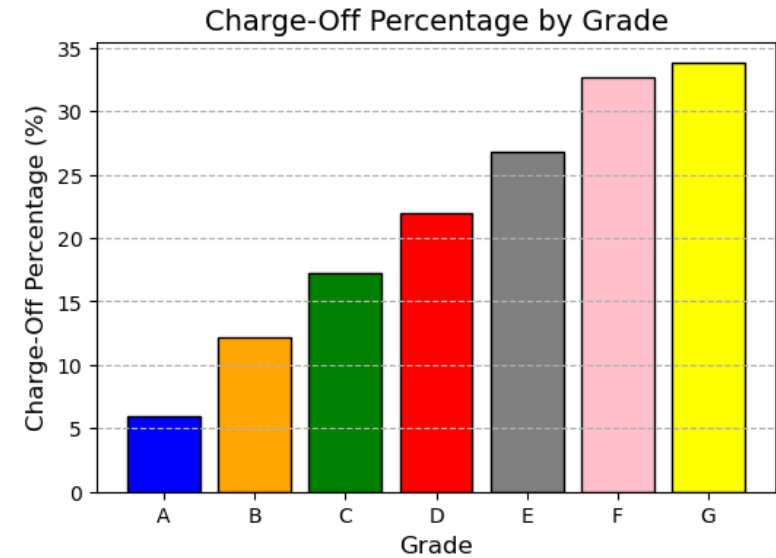
Debt to income ratio

Higher debt to income ratio indicates more financial stress and leads to defaults as shown in figure

Parameter : dti

Default rates based on LC assigned grades

From both the graphs Grade A has low chance of default compared to other higher grades. Risk of default increases towards higher grades as we go on to B..C........G
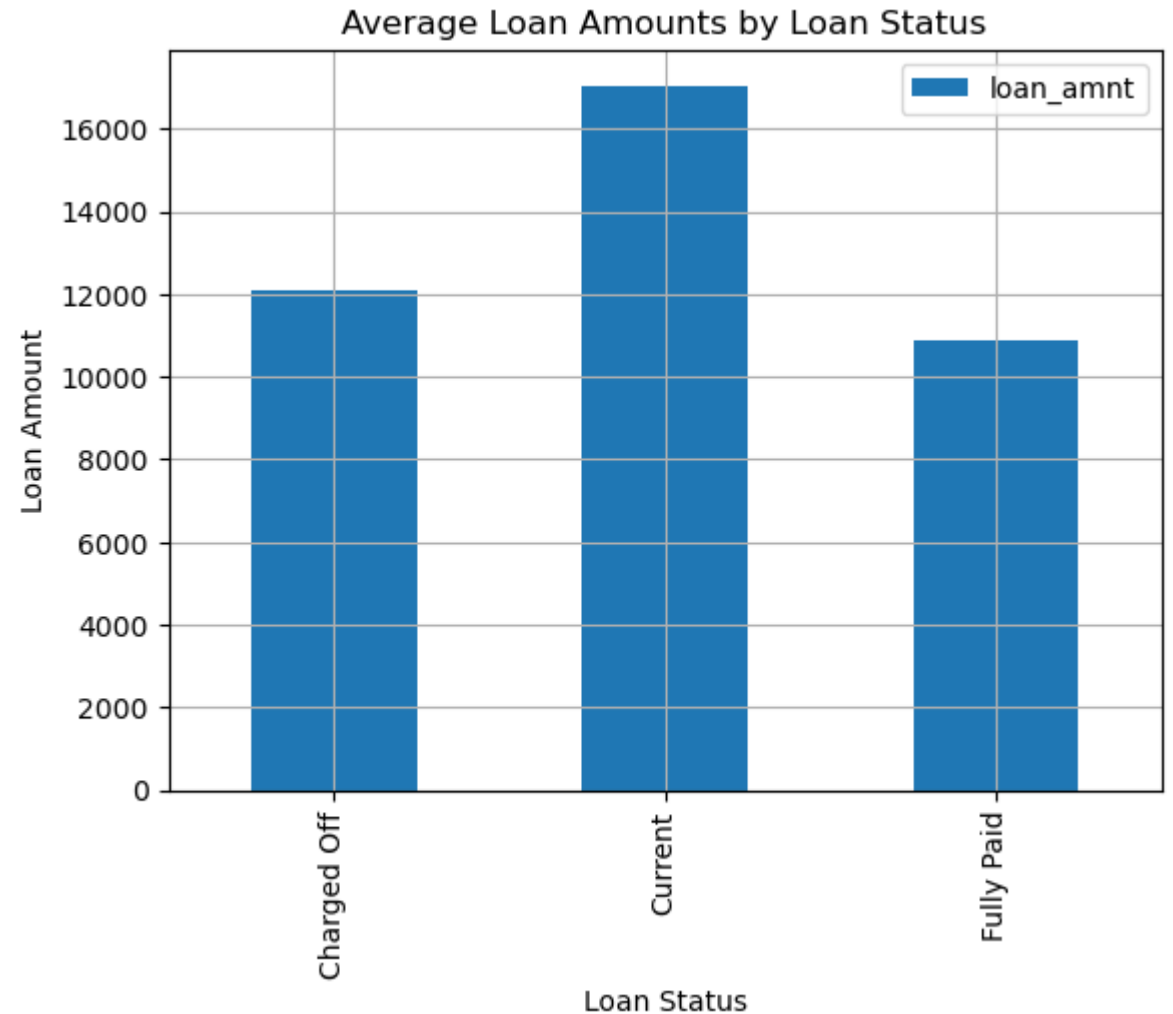
Loan status vs average loan amount

Average loan amount different categories is as follows

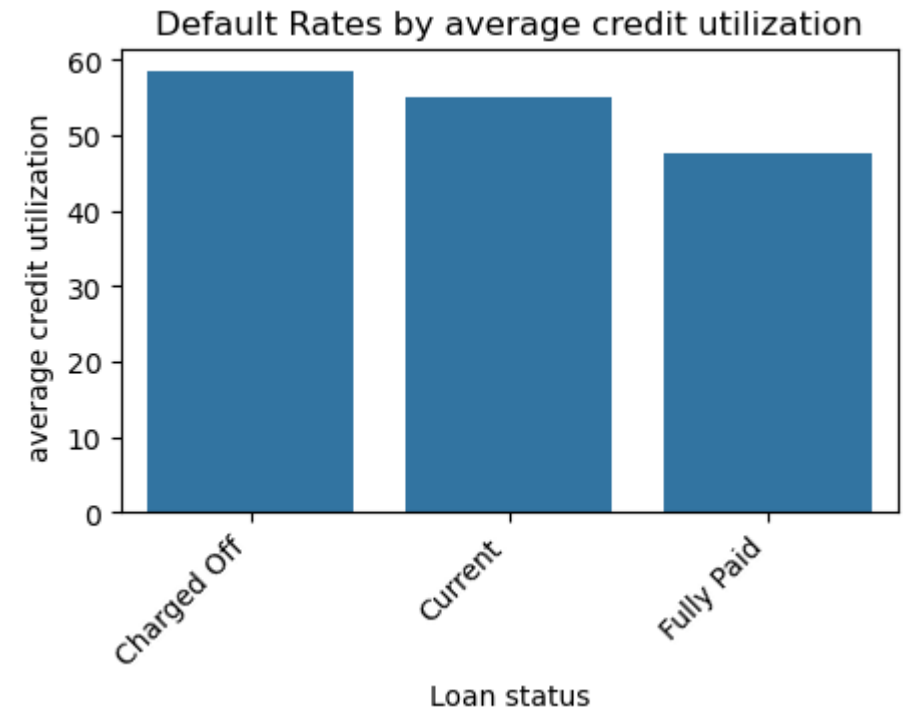Charged off loans have ~1500 $ higher value compared to
Fully Paid loans.

Note : Current is ignored in the analysis as the loan is not closed yet

Default rates based on credit utilization

Higher defaults seen for higher average credit utilization %

Parameter : revol_uti

Default Rates by average credit utilization

0 Charged Off 58.40
1 Current 54.95
2 Fully Paid 47.60

Default rates based on delinq_2yrs

Delinq_2yrs talks about number of times a borrower missed the scheduled payment date in 2 years. Higer rate indicates more financial stress.

Higher average rate for this parameter indicates higher chances of defaults a shown in the graph for Charged off
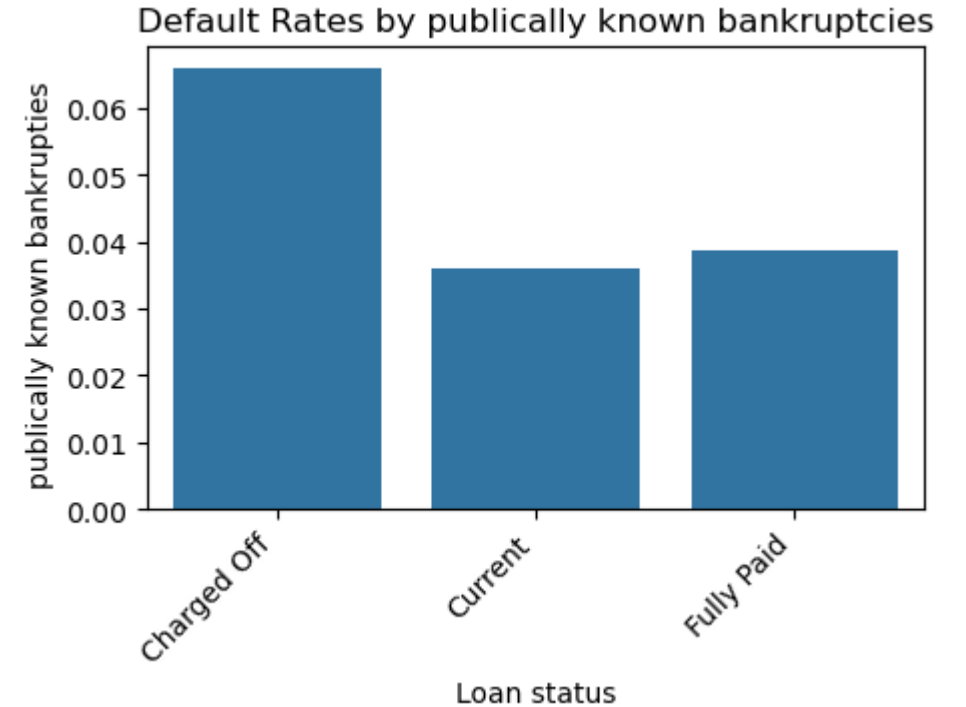

Default Rates by delinq_rate
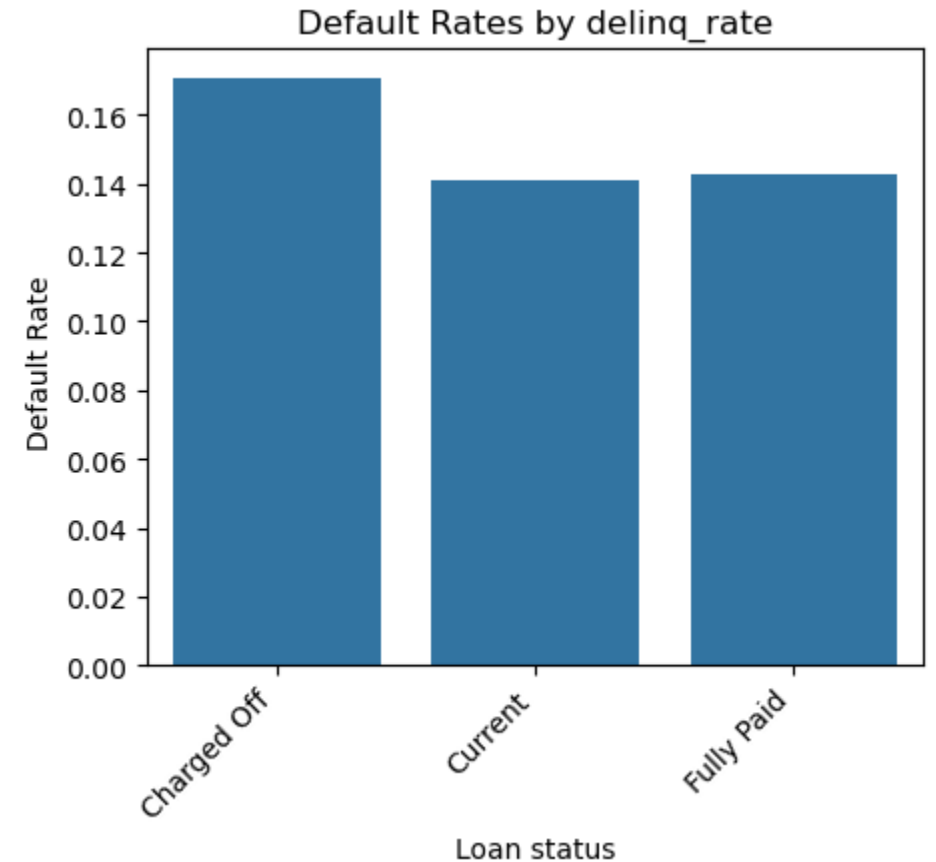
Defaults based on public records

Chances of defaults are higher in case some one has
Higher public records
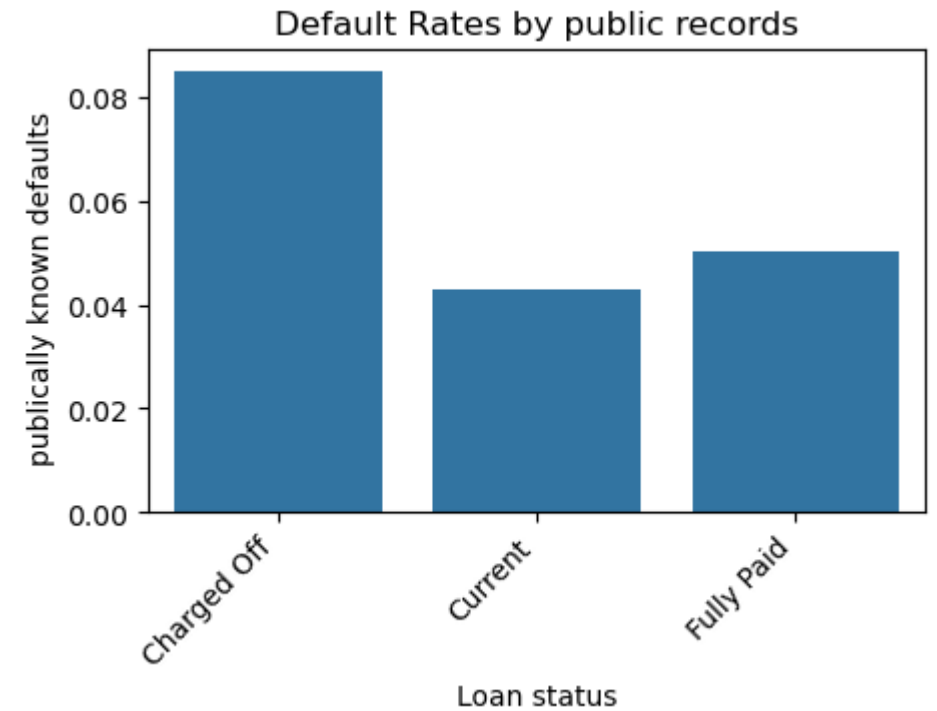
Defaults based on publicly known bankruptcies

Chances of defaults are higher in where a borrower has
Undergone bankruptcies

Correlation matrix

1. loan_amnt & int_rate: There's a moderate positive correlation. This suggests that larger loan amounts are generally associated with higher interest rates.
2. loan_amnt & annual_inc: There's a moderate positive correlation. This indicates that individuals with higher annual incomes tend to borrow larger amounts.
3. int_rate & dti: A weak positive correlation is observed. This suggests a slight tendency for higher interest rates to be associated with higher debt-to-income ratios.
4. annual_inc & dti: A weak negative correlation is present. This implies that individuals with higher annual incomes might have slightly lower debt-to-income ratios.

Correlation Matrix
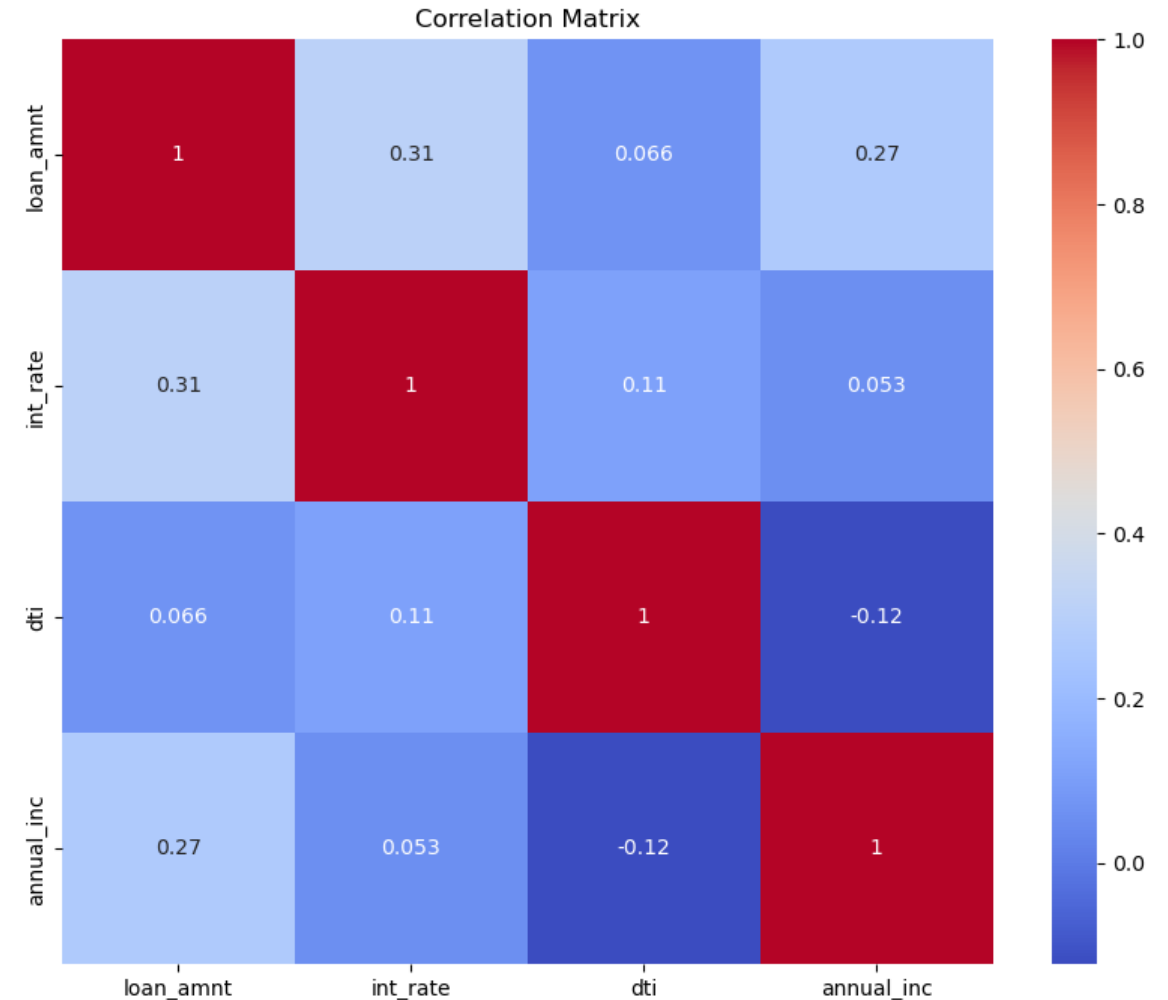
Summary

Parameters influencing the default situation

Parameters	Remarks
Grade	LC assigned grade A – lowest risk ….G- highest risk
Loan purpose	Small_business has higher risk
dti (debt to income ratio)	Higher debt to income ratio , higher risk
Credit utilization	Higer credit utilization leads to higher chance of defaults
Delequncy_2yr	Delinq_2yrs talks about number of times a borrower missed the scheduled payment date in 2 years. Higer rate indicates more financial stress.
Publicly known bankruptcies, and issues	Defaults known in public space has shown impacts of defaults

THANK YOU