



Lending Club Case Study

Pankaj Dixit



UNDERSTANDING
LENDING CLUB



PROBLEM TO SOLVE

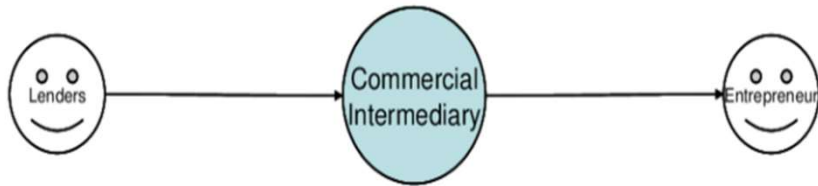


DATASET
PREPARATION



DATA ANALYSIS
&
CONCLUSION

Understanding Lending Club



- Lending club 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 Club assigns each borrower a grade based on their credit history.
- Investors select loan request and fund accordingly.

Problem to Solve

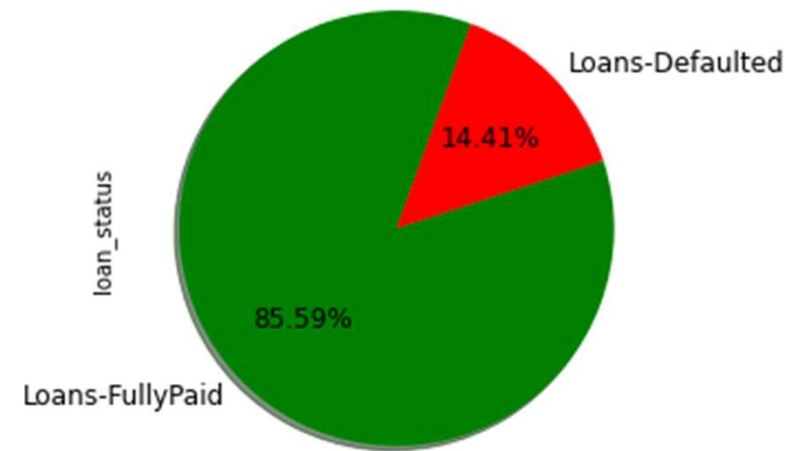
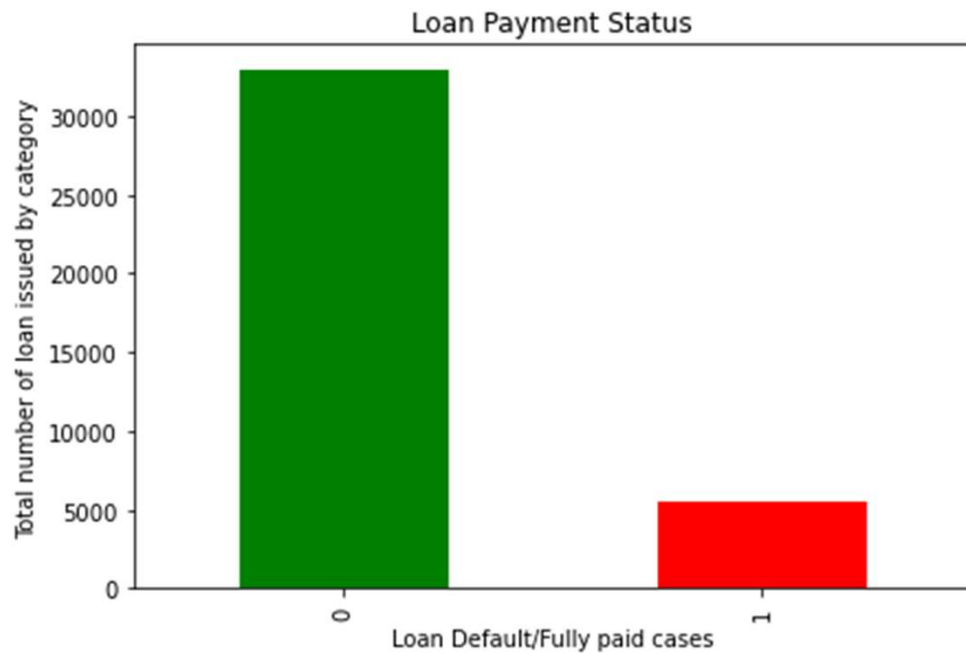
- Identify the key indicators for loan default
- Reduce the bad loan and increase the loan recovery rate

DATASET PREPARATION

- Dataset contains data in the form of multiple columns which can't be used as is due to below reasons and hence data cleaning is an extensive activity that is needed.
 - Multiple columns having “NA”
 - Strings are part of data values
 - Multiple missing values under various headers
 - Multiple columns have negligible impact on the analysis procedure
- So, to make the data ready for analysis data cleaning is being done which makes the data ready for plotting and further analysis.

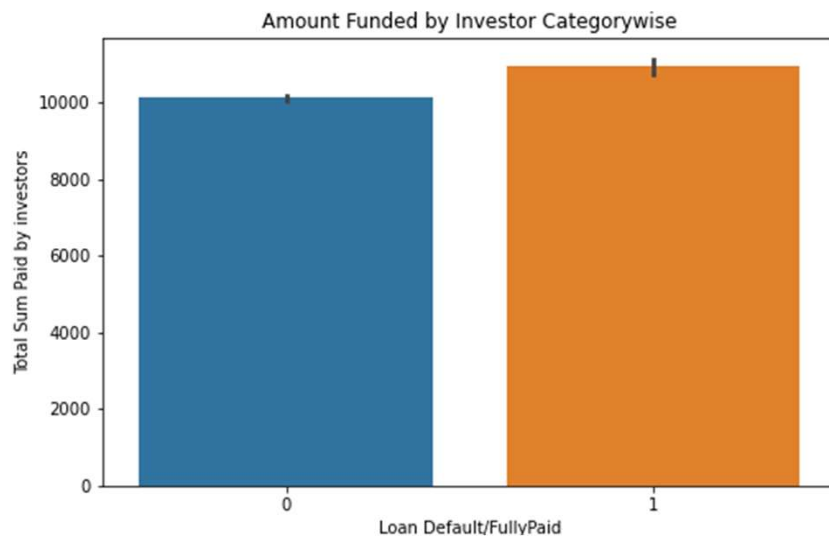
Data Analysis

Percentage of loan default

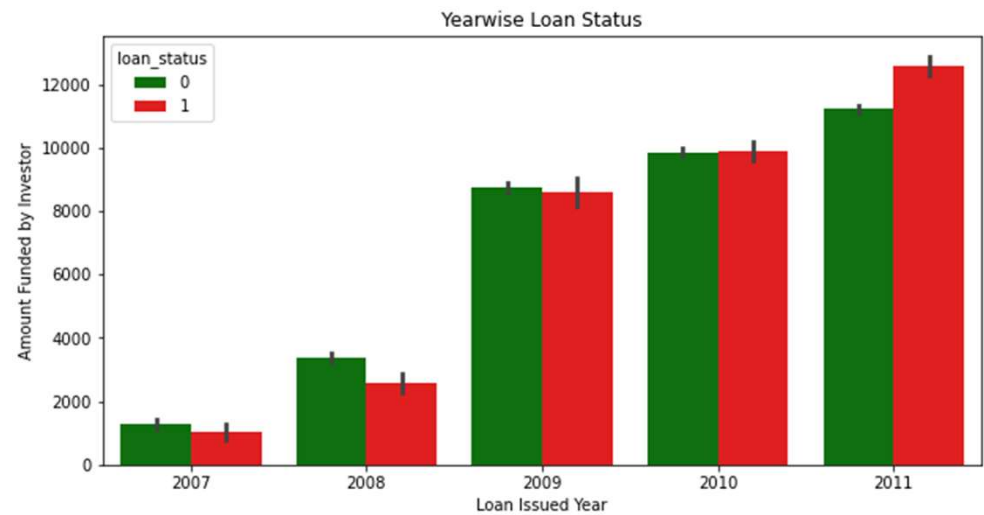


From the data it is clear that 14.41% of the loan is defaulted

Loan Status



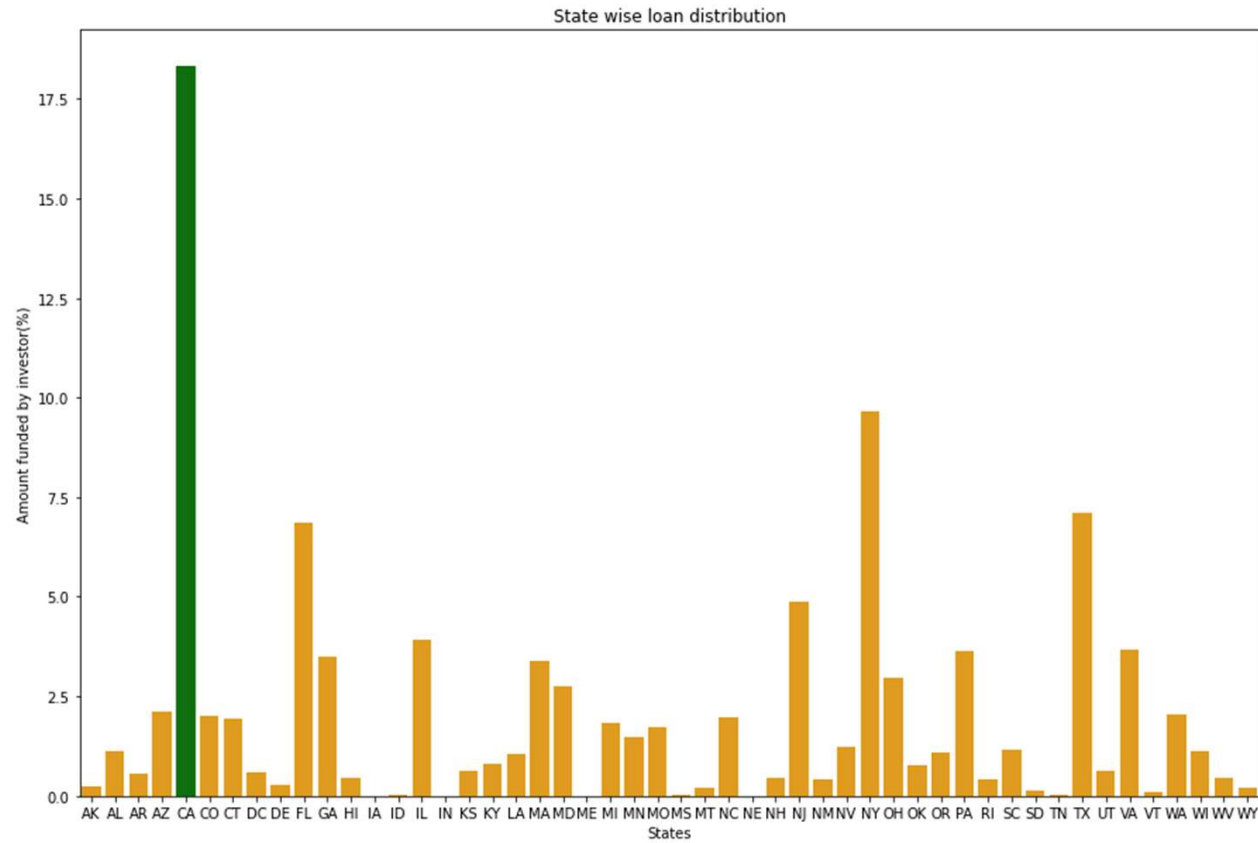
Loan Status Vs Amount Funded by investor



Year wise amount funding

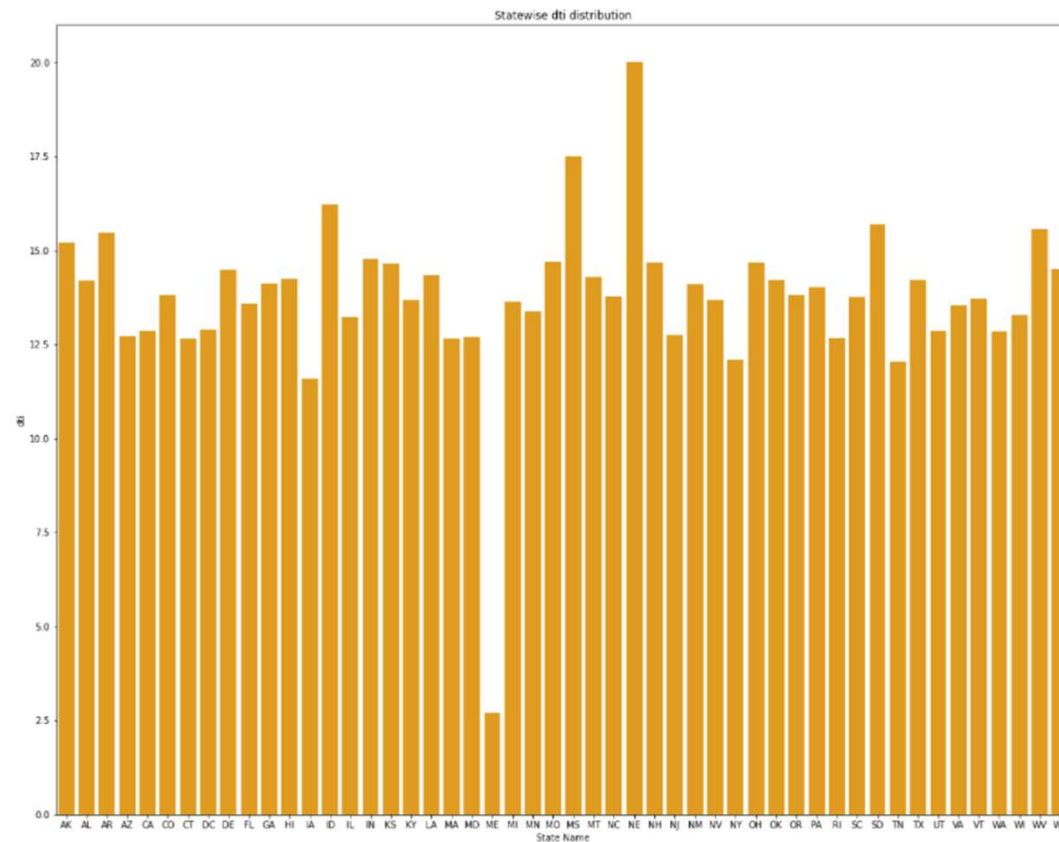
- Loan amount more than 10K has more no. of defaulters.
- The funding has increased YoY and the default as well.
- In 2011 the default was more as compared to other years.

State wise loan distribution



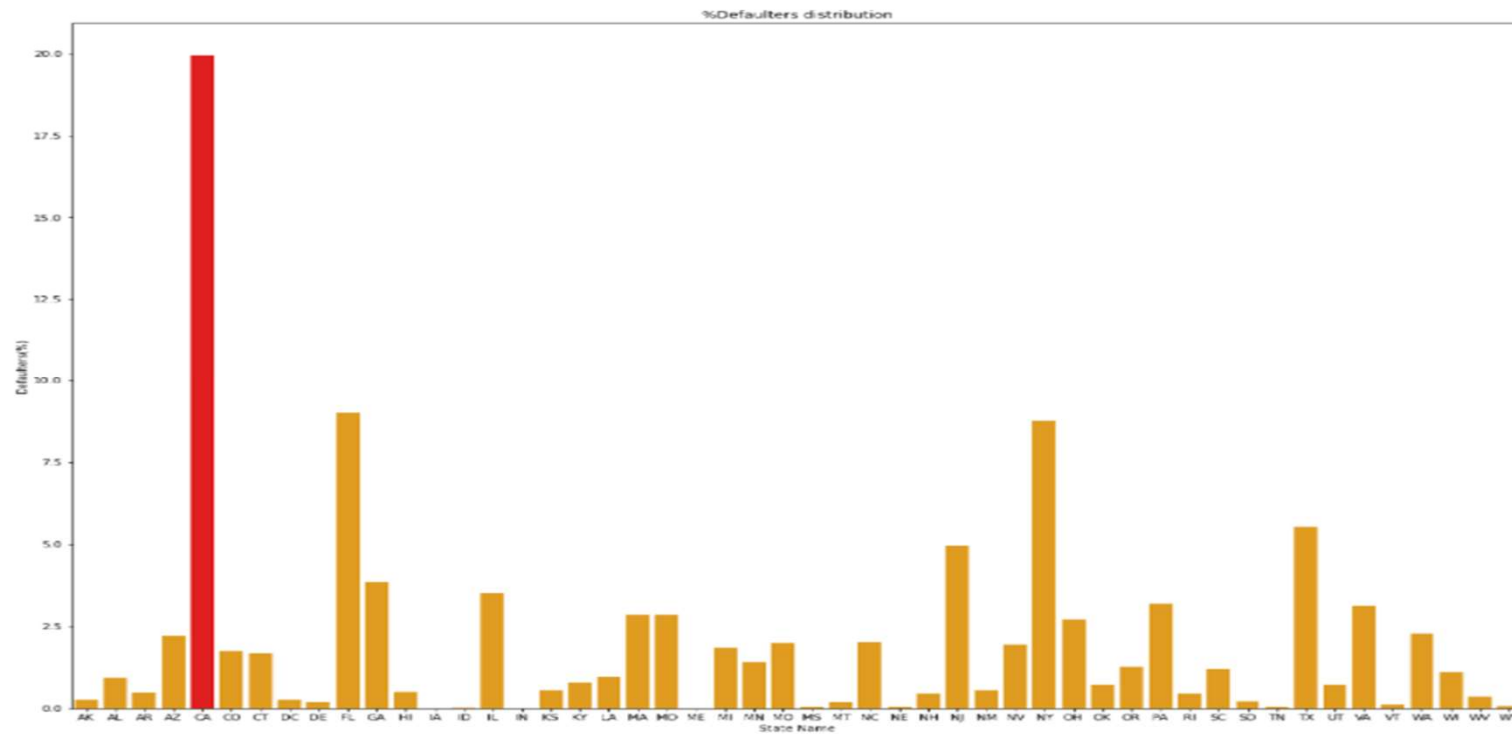
CA state receives more funding compared to other states

State vs Avg debt to income



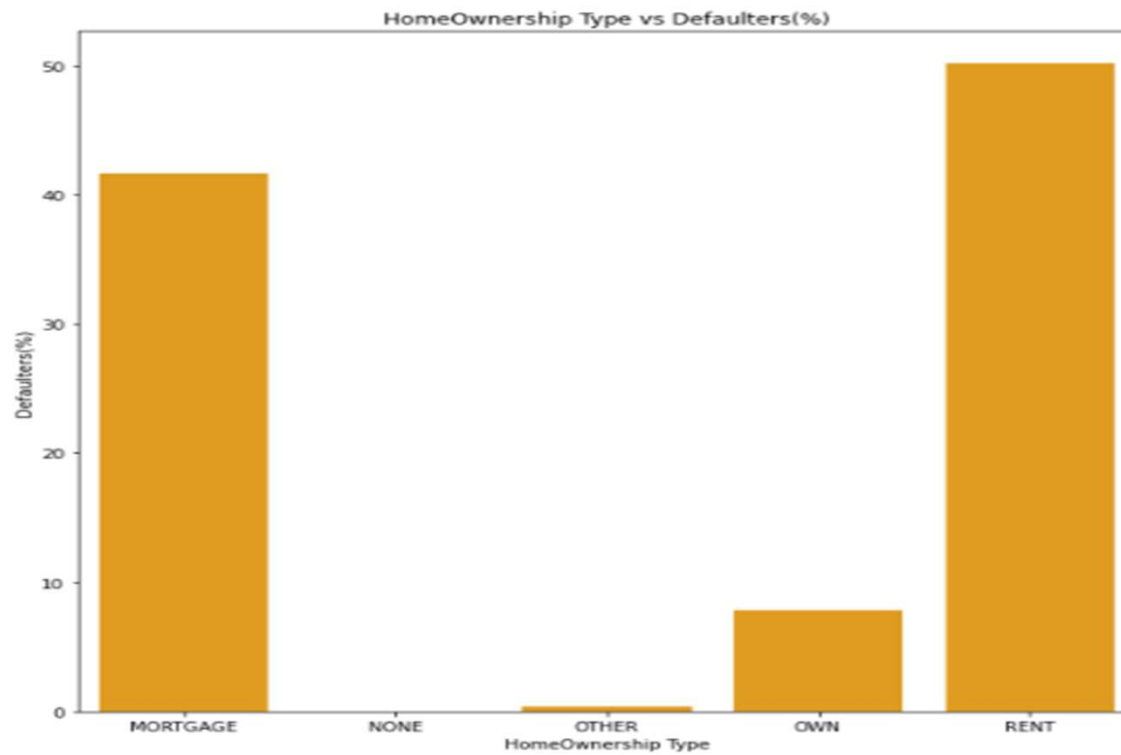
ME has the least DTI so it is the preferable state while NE is highest so the company may avoid lending in this state.

State wise Loan Defaulters



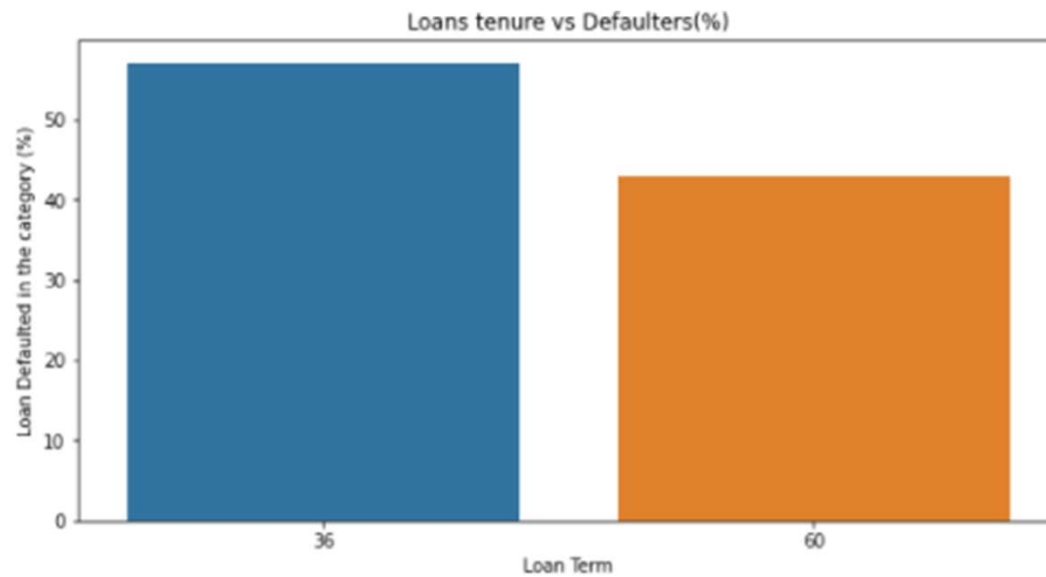
CA state has the greatest number of defaulters so company should be more cautious in this state.

Homeownership status Vs Defaulters



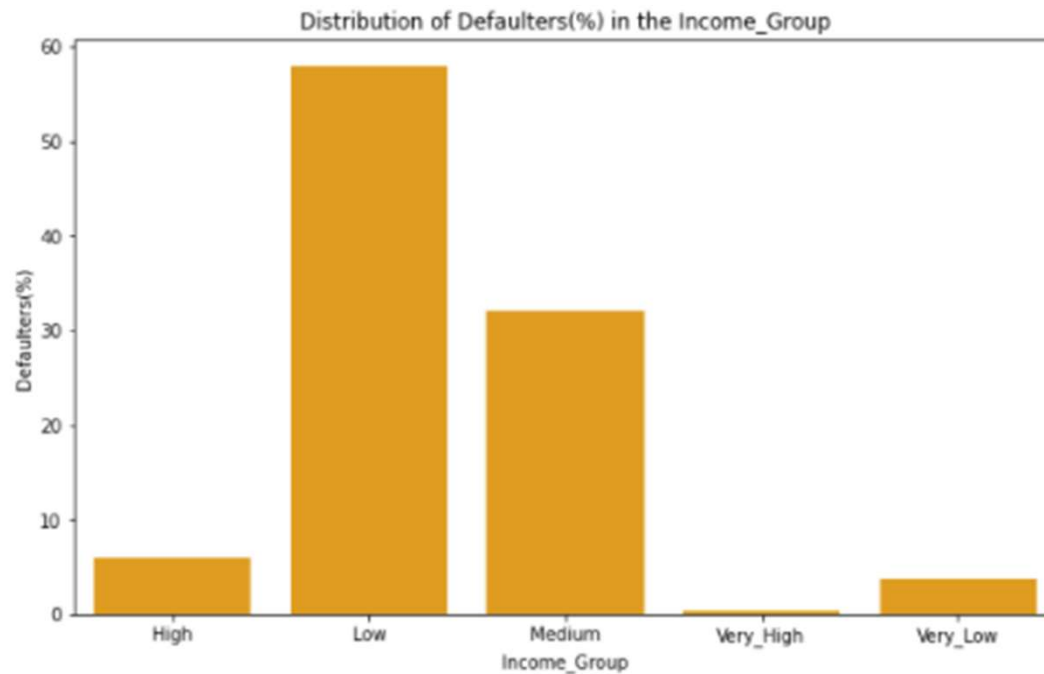
Person living in Rent are more likely to default

Loans tenure vs Defaulters



Loan default is seen in 36-month term

Income Group vs Defaulters

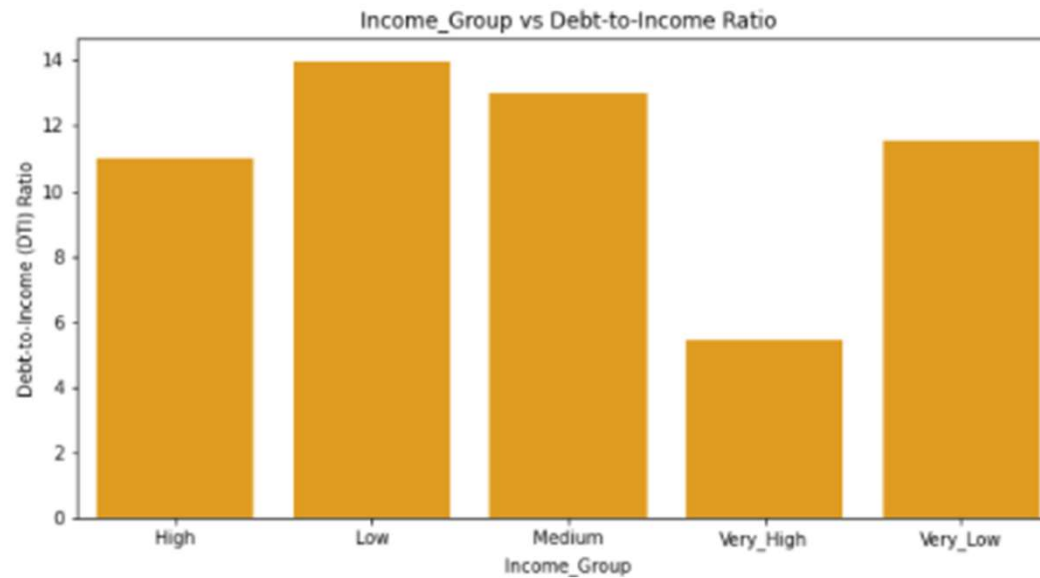


Clients are categorized as below in terms of Income Group

- a. income less than 20000 --> Very_Low
- b. income more than 20000 and less than 60000 --> Low
- c. income more than 60000 and less than 120000 --> Medium
- d. income more than 120000 and less than 300000 --> High
- e. income more than 300000 --> Very_High

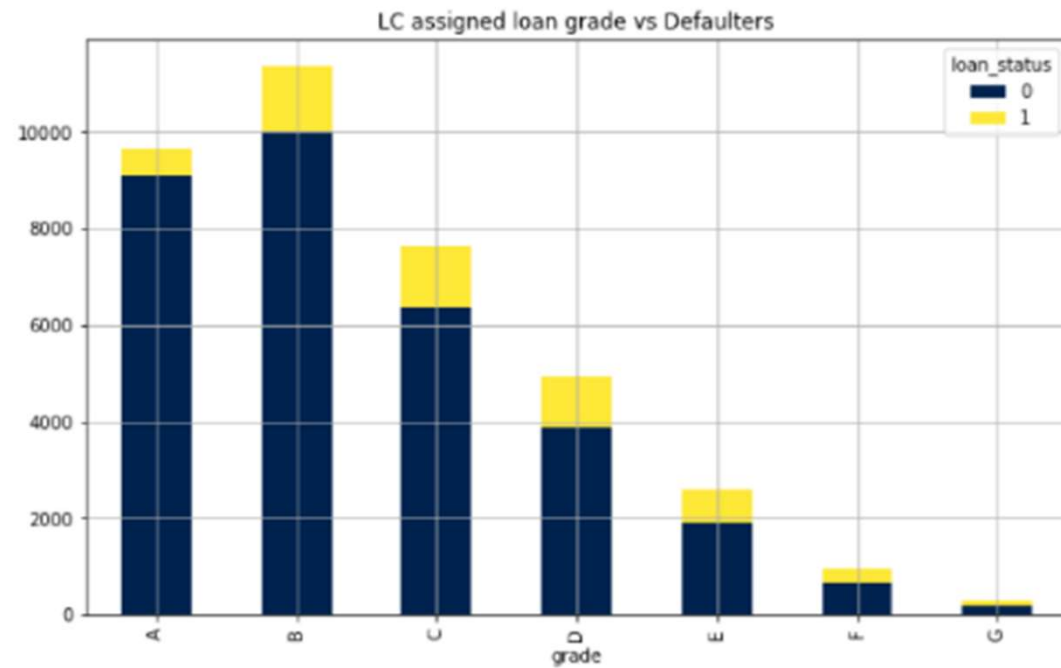
Clients with Low Income Group are more likely to default

Income Group vs Debt-to-Income (DTI) Ratio



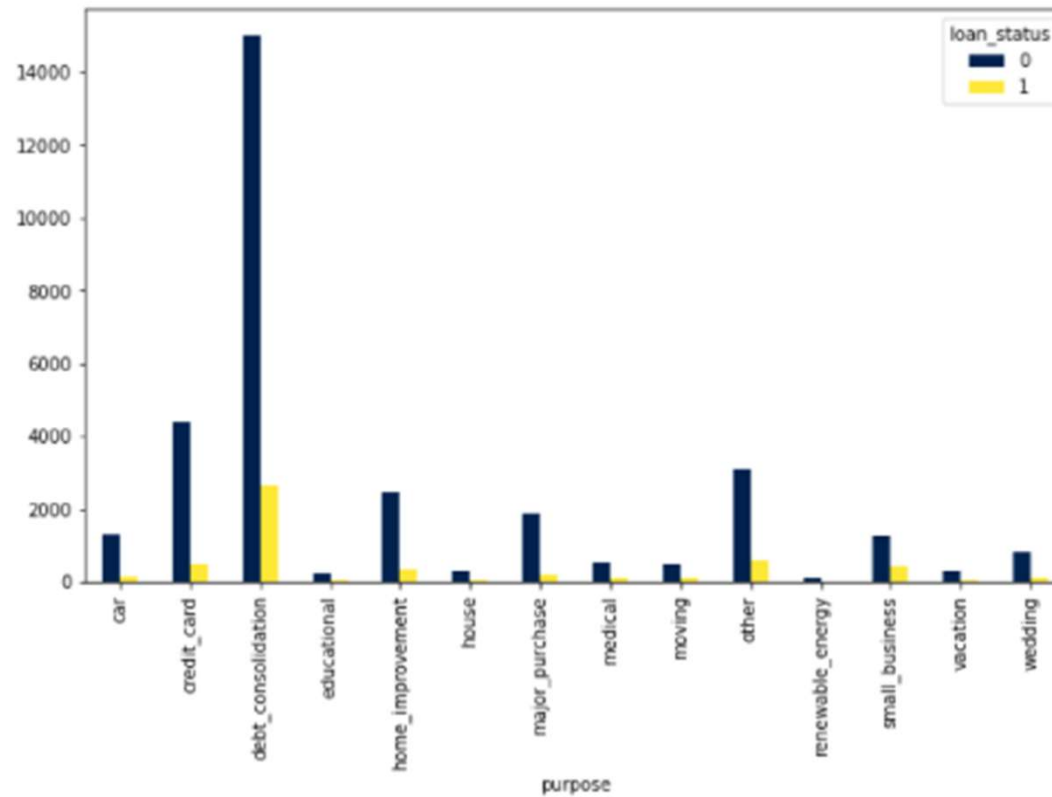
- Low-income group has highest DTI so there is a risk of default
- Very_High income group has lowest DTI so have less risk of default

LC assigned loan grade vs Defaulters



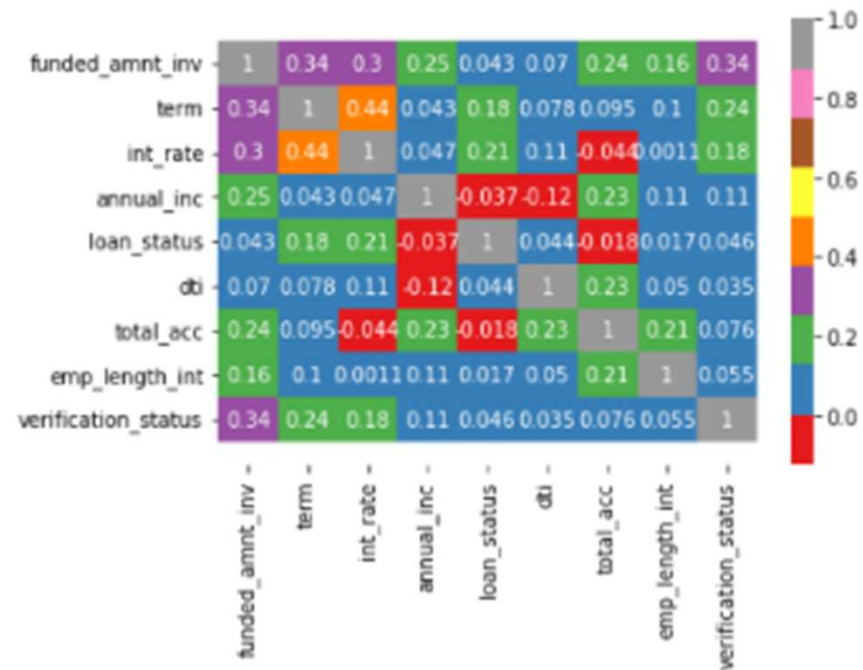
Loan grade B has highest number of defaulters

Purpose of loan Vs Defaulters



Loan taken for debt consolidation has highest number of defaulters

Correlation Matrix



Summary

- Average annual income and loan status have negative correlation so lower the income higher the default. This is evident from the graph as well as correlation matrix.
- DTI and annual income have negative correlation so increase in salary will decrease DTI and hence increase chances of loan approval.
- Clients with Low Income Group are more likely to default.
- Loan taken for debt consolidation has highest number of defaulters.
- Loan grade B has highest number of defaulters.
- Loan default is seen the most in 36-month term.
- CA state has most of the defaulters.