Lending Club Case Study

Created by

Sagar Ambalam

&

Sandeepan Roy

Steps involved in EDA

- Schema understanding
- Data cleaning
- Data visualization

Schema understanding

- There were 111 columns initially in the dataset provided for analysis.

- The number of data points in the dataset were 39K

- There were many columns which had missing values which had to be handled in appropriate manner.

- There were 54 columns where all the values in the columns were null.

- There were also 3 columns where the column consisted of only 1 unique value. These columns would not add any value to the analysis and hence were to be removed.

Data Cleaning

- The steps involved in data cleaning are handling outliers and selecting appropriate column for the analysis and dropping the redundant ones.

- We found that there were many columns that were redundant based on our business understanding. The columns that contained credit bureau information were important part of the analysis but as per the business, the date at which the bureau information was extracted was not matching with the loan issue date. Hence all these variables had to be dropped.

Data Cleaning

- The following are the columns selected for the analysis
  - Loan amount
  - Interest rate
  - Grade
  - Sub grade
  - Term
  - Dti
  - Annual_income
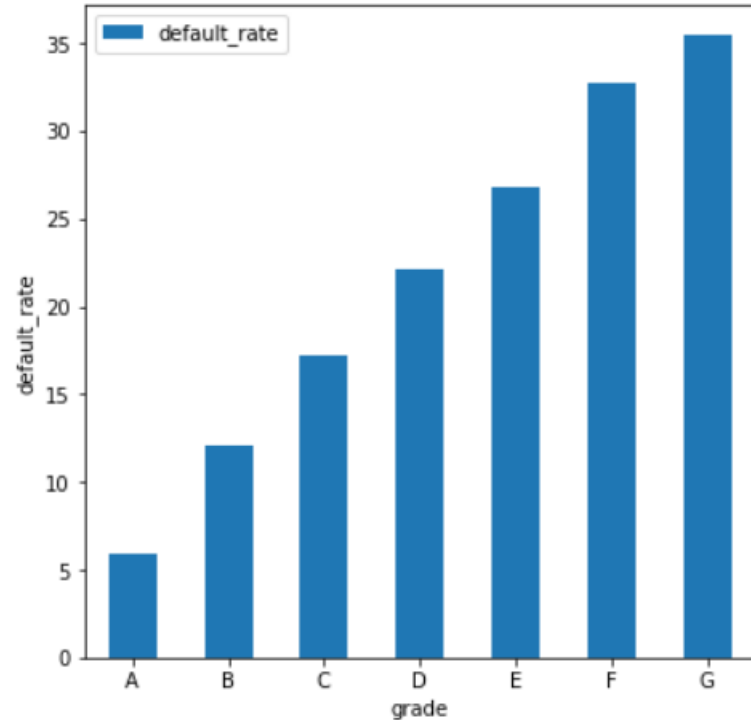 - Purpose
  - Emp_length
  - Home_ownership

Data Visualization & Recommendation
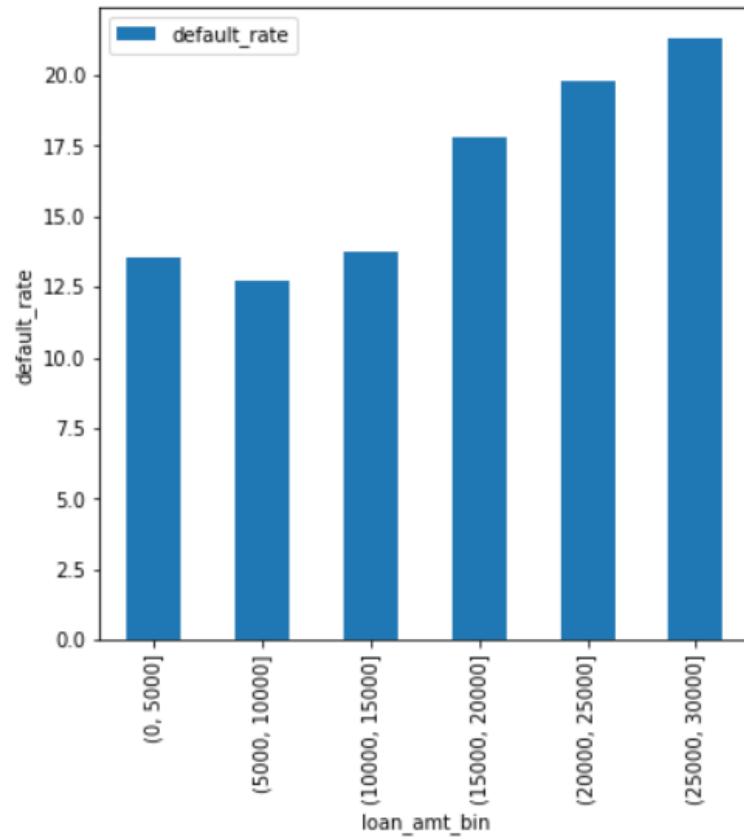
Loan Amount

- In the given dataset we found that, as the loan amount increases so has the default rate of the customer.

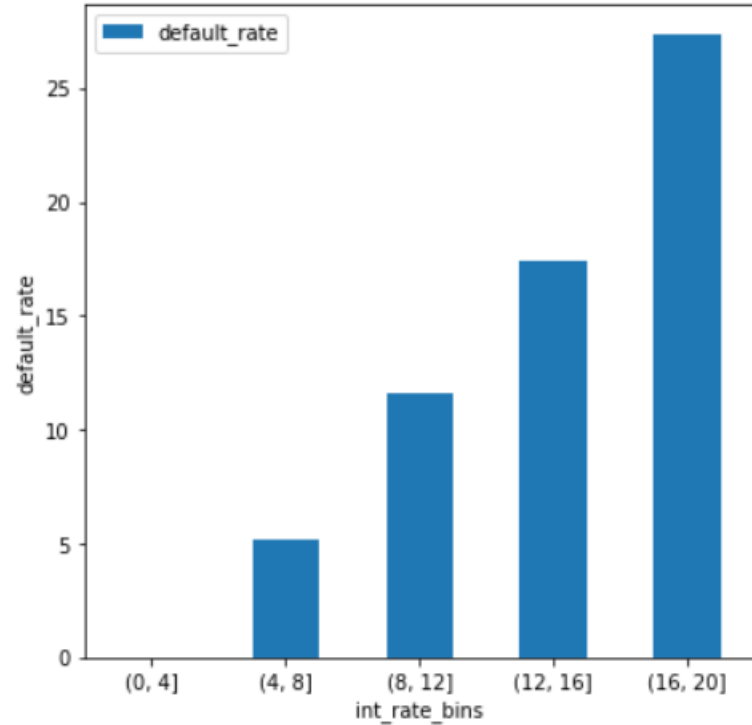- Hence, the loan amount should be decided carefully.

Data Visualization & Recommendation

Grade

- Grade of the customer is good predictor of defaulting customer

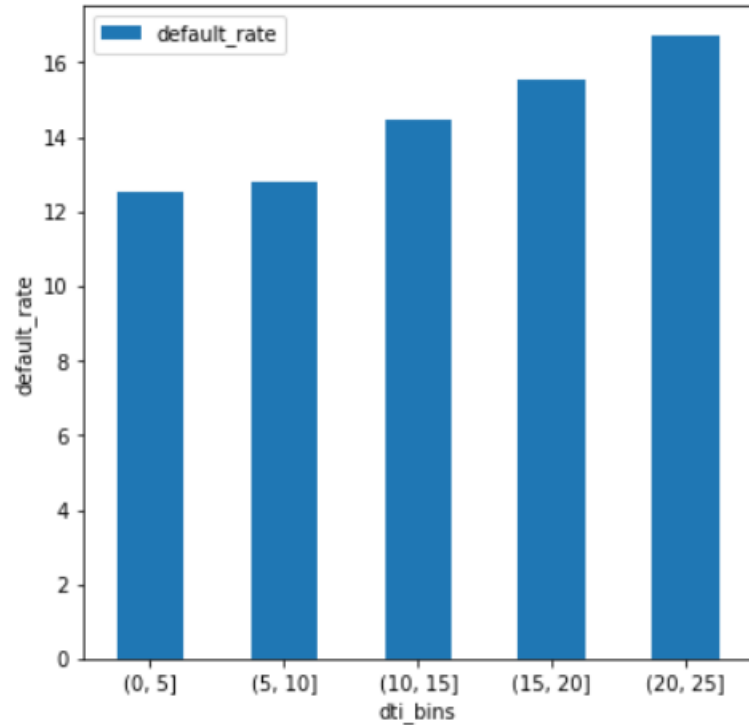- As the grade increases from A to G, the default rate also increased.

Data Visualization & Recommendation

Interest Rate

- Interest rate is also a good driver variable in determining if the customer is likely to default.

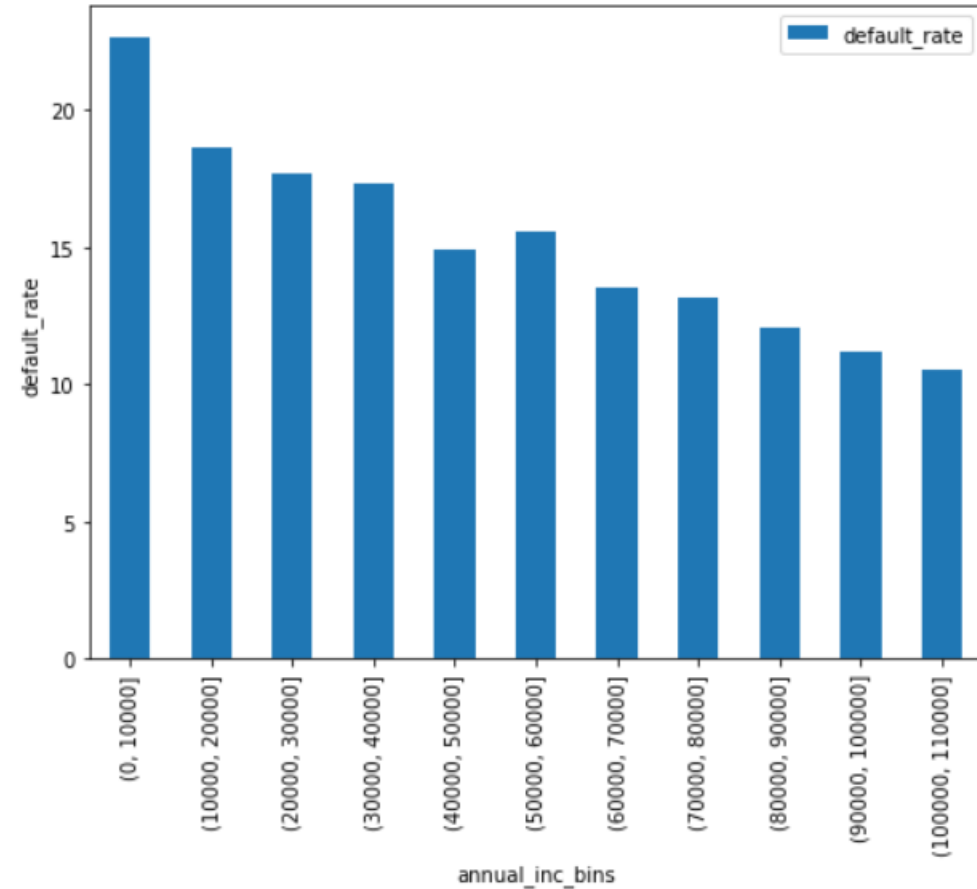- Higher the interest rate, higher chance of default.

Data Visualization & Recommendation

Debt To Income Ratio
- dti(debt to income ratio) shows a good trend of increasing default with increase in dti.

- As debt increases or income decreases, the customer is more likely to default

Data Visualization & Recommendation

Annual income

- Annual income of the customer has a high influence on determining whether the customer is likely to default or not.
- Lower the annual income, higher the chance of default.