Lending Club Default Analysis

Group Facilitator: Naresh Kumar Ailaveni (chanti_36@yahoo.co.in/+91-9972493009)
Team Member: Abhishek Arora (mail2abhishekarora@gmail.com/+917503995004)

Problem Statement

Understand how **consumer attributes** and **loan attributes** influence the tendency of default using the EDA.

Data Understanding/Cleaning

- Based on available information for the Data Set and Data Dictionary, the **driving factors (or driver variables)** behind loan default are understood.
- Used those variables to understand the risk associated to loan default.
- Understood the insights of the variables that can influence the loan default using live session on Lending Case study conducted on 5th Mar 2023.
- Used below key variables for the study using EDA i.e., 'loan_amnt','funded_amnt','funded_amnt_inv','term','int_rate','grade','sub_grade','annual_inc','dti','emp_length','purpose','loan_status','issue_d','home_ownership','verification_status'
- Cleaned the data to separate the string and numeric values in the variables of interest.
- Derived new variable for Month and Year using "pd.to_datetime" function in Pandas.
- Converted numeric columns into numeric data to perform correlation.

Outlier Identification

- Used “describe()” function and “boxplot” option to understand the outliers in the variables.
- Removed outliers in annual income variable.

Correlation

- Correlation for the key variables are performed and concluded that the distribution in loan Amount, Funded Amount, Funded Amount by Investor has high correlation.
- Below mentioned are the observations:
  - Observation is that Loan amount, investor amount, funding amount are strongly correlated. So only “Loan Amount” variable is considered for further analysis and dropped “Funded Amount’ and ‘Funded Amount by Investor’ variable.
  - Annual income and Interest rate is positively correlated
  - Annual income with dti is negatively correlated
  - Term and Interest Rate is positively correlated
  - Annual Income and Employment length (Years) is positively correlated

Univariant Analysis

- Univariant analysis is performed on Unordered Categorical Variables i.e., Purpose, Home-Ownership.
- Univariant analysis is performed on ordered Categorical Variables i.e., Loan Status, Annual Income, Term, employee Service, Issue Month and Issue Year.
- Grouped the variables i.e., loan amount, annual income, interest rate and dti matching to “Percentile” of 25, 50, 75 and 100 to understand the variation and influence of the group to loan default.
- Below mentioned are the observations from the existing variables, derived variables and grouped variables:
  - Most of the loans were taken for the purpose of debt consolidation followed by paying credit card bill. Also, charged off count is high for debt consolidation followed by paying credit card bill.
  - Most of the loans taken for the home ownership falls under rent and mortgage. Also, charged off count is high for rent and mortgage.
  - With increase in loan applications every month and year, there is increase in charged off.
  - With increase in loan amount there is increase in charged off.
  - Most of the loans were taken for the purpose of debt consolidation.
  - With increase in annual income there is decrease in charged off.
  - With increase in interest rate there is increase in charged off.
  - With increase in dti there is increase in charged off.
  - Grade B,C,D,E have high risk of charged off.
  - Grade B1:B5,C1:C5,D1:D5,E1:E5 have high risk of charged off.

Bivariant Analysis

- Bivariant analysis is performed on Unordered Categorical variables and Categorical variables.
- Below mentioned are the observations for the Bivariant analysis:
  - Considering the volume of applications for 36 and 60 month term, applicants taken loan for 60 months had more percentage of getting charged off as compared to applicants who had taken loan for 36 months.
  - For vacation purpose, when high experienced employee applies for loan, there is a high risk of loan default. Similarly, when low experienced employee applies for education loan, there is risk of loan default.
  - Applicants taking loan for Mortgage under home ownership and 60 months term has high risk of default.
  - Applicants with high DTI, under OTHER in home ownership with 60 months term has high risk of default.
  - Loans taken for the purpose of debt consolidation followed by paying credit card bill possess risk of loan default.
  - With increase in annual income there is decrease in charged off and with increase in interest rate there is increase in charged off.
  - Applicants living in rented home or mortgaged their home possess risk of loan default
  - Grade F with high annual income choosing 60 months term has high risk of default.

PairPlots using Seaborn

- Pair plots are loan amount, annual income, term, interest rate and dti variables.
- Below mentioned are the observations from pair plots:
  - Higher the interest rate the higher charged off ratio.
  - Higher the annual income, higher the loan amount, interest rate & dti and vice-versa.
  - Interest rate is increasing with increase in loan amount and is leading to loan default.

Conclusion

The conclusion from the EDA are

- Loans with 60 month term had more percentage of getting charged off.
- Loan on vacation by experienced employee possess high risk of loan default.
- Loan on mortgage with 60 months term has high risk of default.
- Debit to income for unknow purpose by house owners with 60 months term has high risk of default.
- Grade F with high annual income choosing 60 months term has high risk of default.
- Loans taken for the purpose of debt consolidation possess risk of loan default.
- With increase in annual income there is decrease in charged off and with increase in interest rate there is increase in charged off.
- Applicants living in rented home or mortgage possess risk of loan default