Lending Club Case Study

-by Raghu Kalyan

Table of contents:

- Introduction
- Business Understanding
- Problem Statement
- Business Objective
- Data Understanding
- Univariate Analysis: Insights
- Bivariate Analysis: Insights
- Conclusion

Introduction

Solving this assignment will give an idea about how real business problems are solved using EDA. In this case study, techniques are applied to understand of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Business Understanding

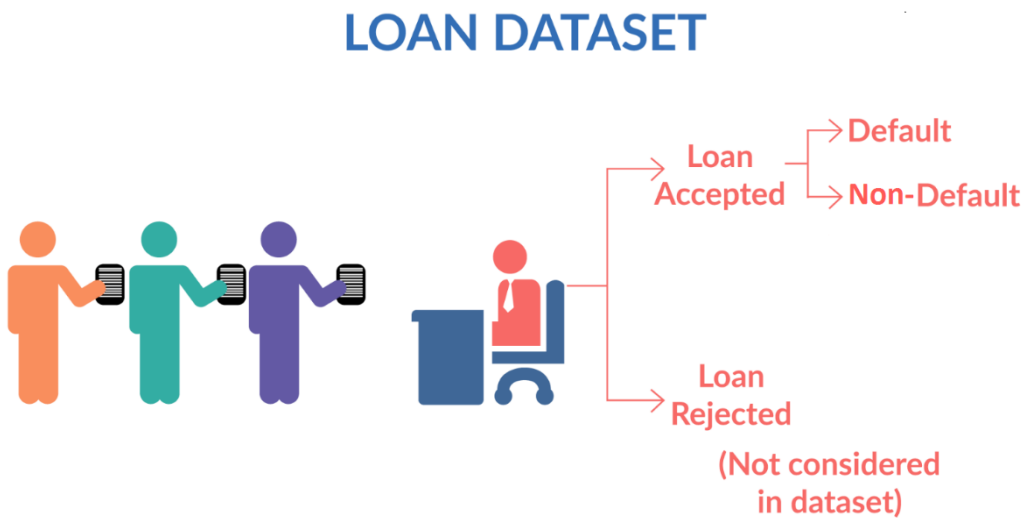
Data explains about lending various types of loans to urban customers. When the company receives a loan application, the company must decide for loan approval based on the applicant's profile.

Problem Statement

Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company.

The data contains information about past loan applicants and whether they 'defaulted' or 'not defaulted'. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (too risky applicants) at a higher interest rate, etc.



In this case study, by applying various EDA techniques to understand how consumer attributes and loan attributes influence the tendency of default.

When a person applies for a loan, there are two types of decisions that could be taken by the company:

- Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:
  - Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
  - Current: Applicant is in the process of paying the instalments, i.e., the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
  - Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan.

- Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Business Objective

This company 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.

Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

If one can identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicant's using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

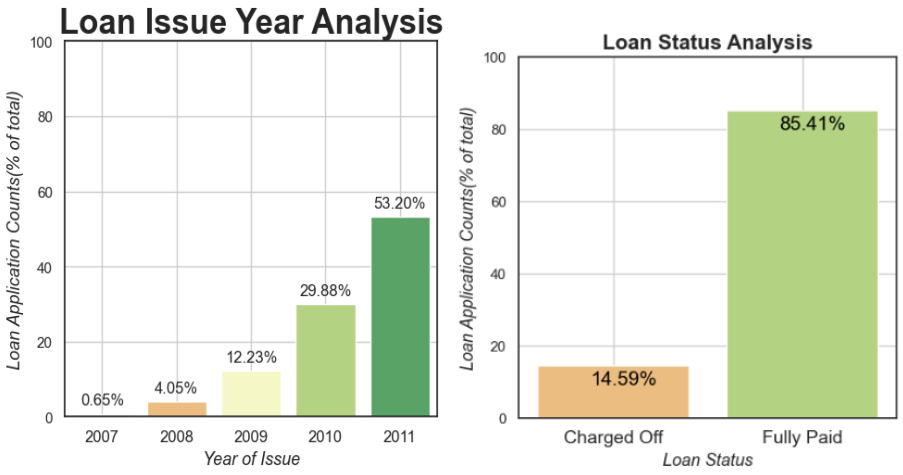
Data Understanding

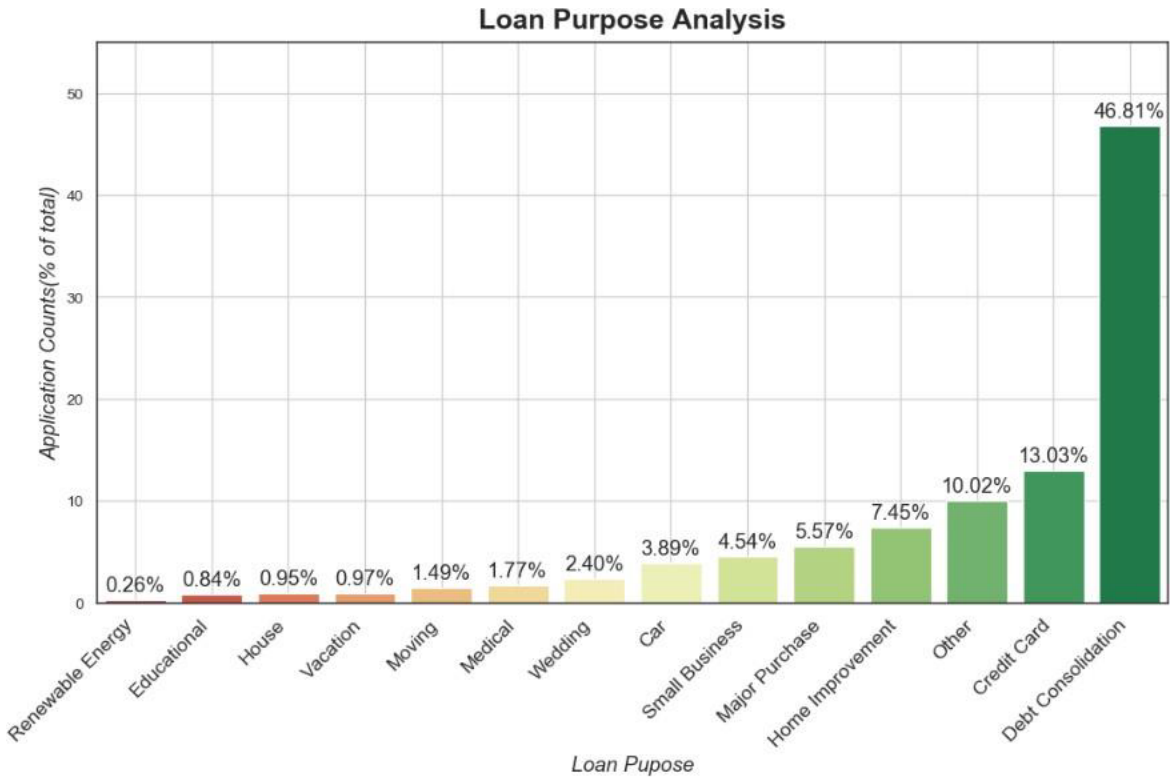
EDA

With the help of shape function noticed 39K+ rows and 111 columns in the data set and with the help of info function it was understood that many columns are not relevant for analysis and used drop function to remove columns from the dataset and derived final data set with 39K+ rows and 23columns for further analysis.

Univariate Analysis: Insights

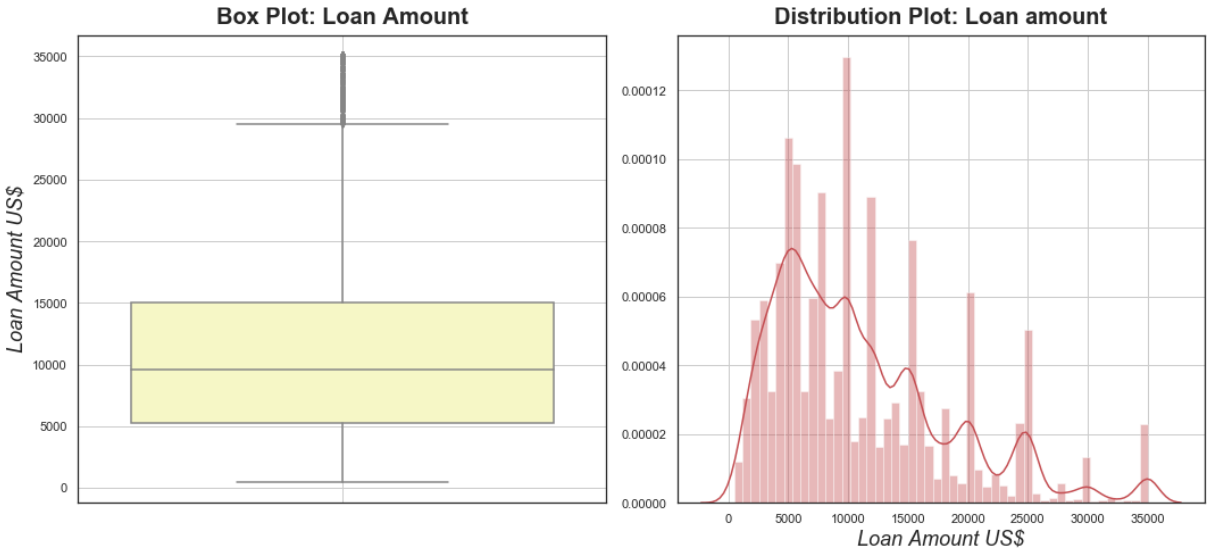
- Overall Default Rate stands at 14.59%.
- Approval rate of loansjumpedbyaround78%from the year 2010to2011.
- Approved Loan amount almost doubled between years'2010and2011.

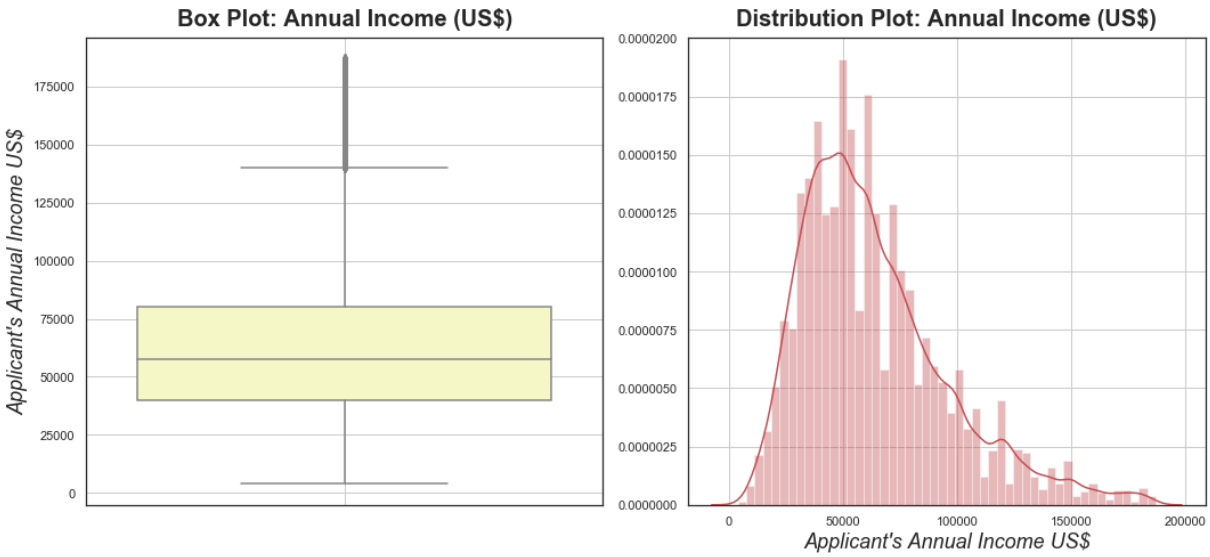




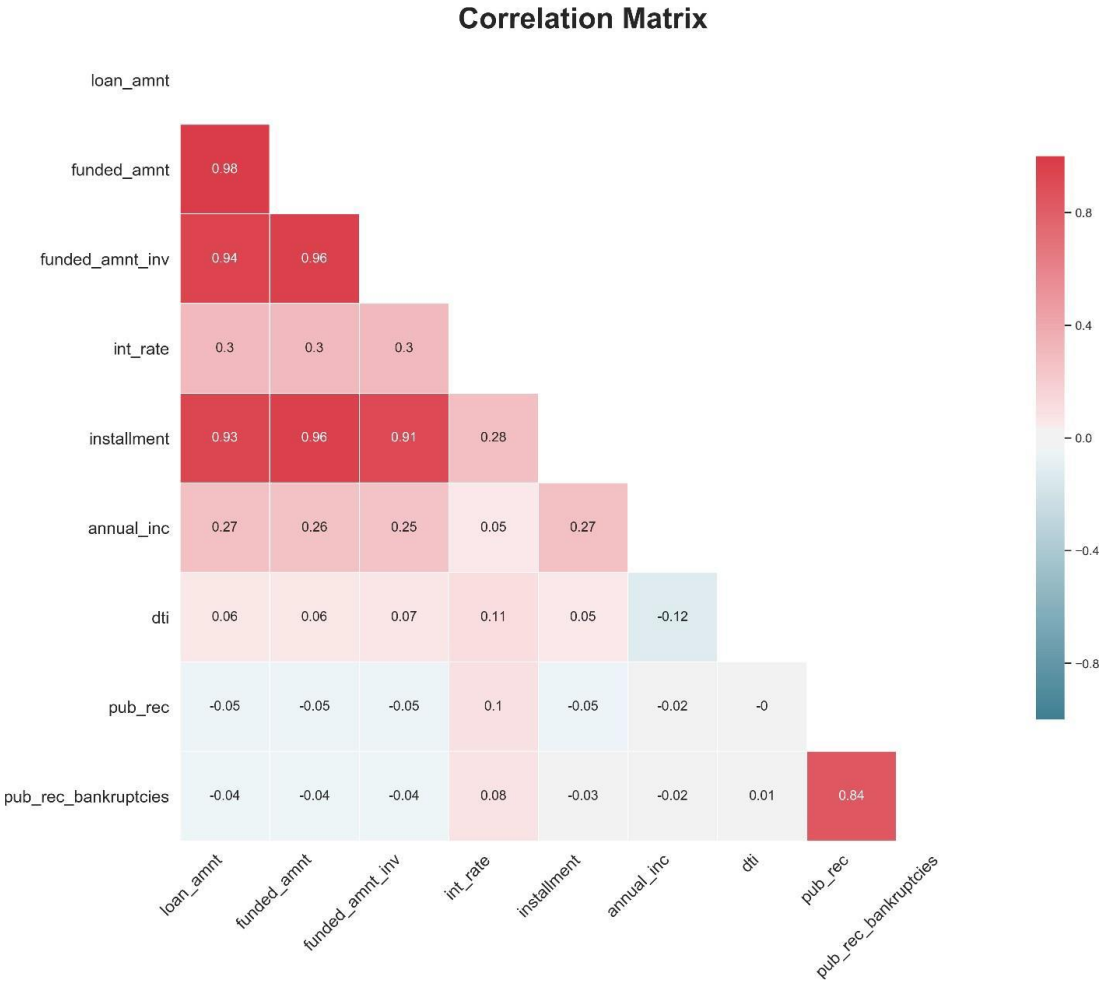
'Debt Consolidation' followed by 'Credit card' are the top two reasons for loan applications amongst approved loans.

- 50% of loan applicants request a loan amount between 5.3K USD and 15K USD.
- Annual Income was highly skewed as expected ; ignoring the outliers, the average salary of applicants is 68.78K USD.
- 50% of the applicants earn between 40K USD and 82K USD annually.

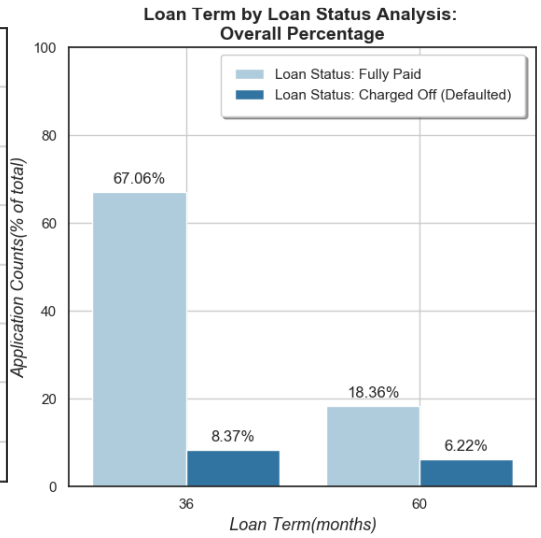




Bivariate Analysis: Insights

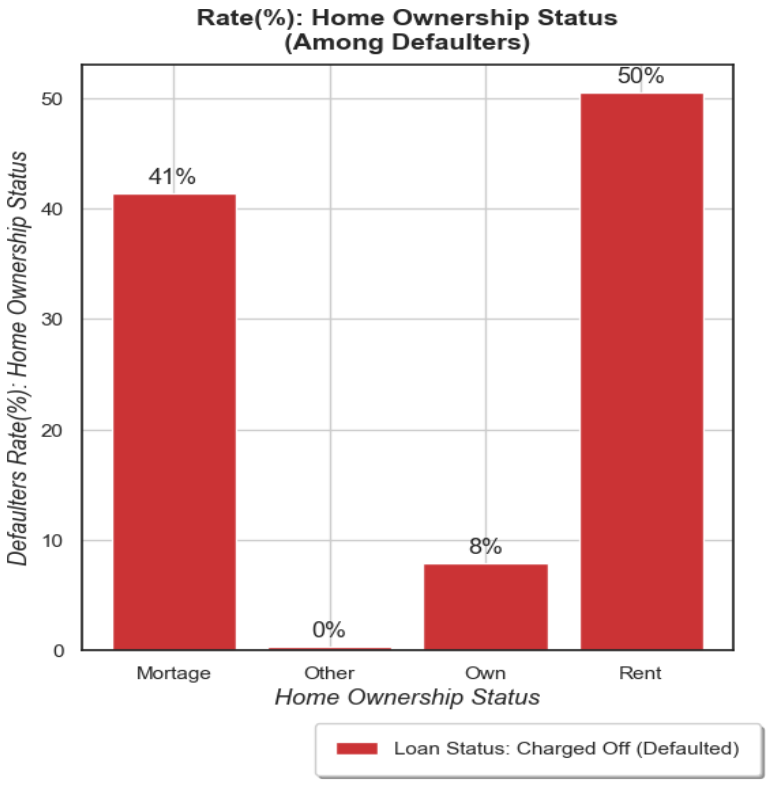


- Pearson Correlation plot between numeric/continuous variables retained.
- Numbers are indicative of the strength of correlation between variables.

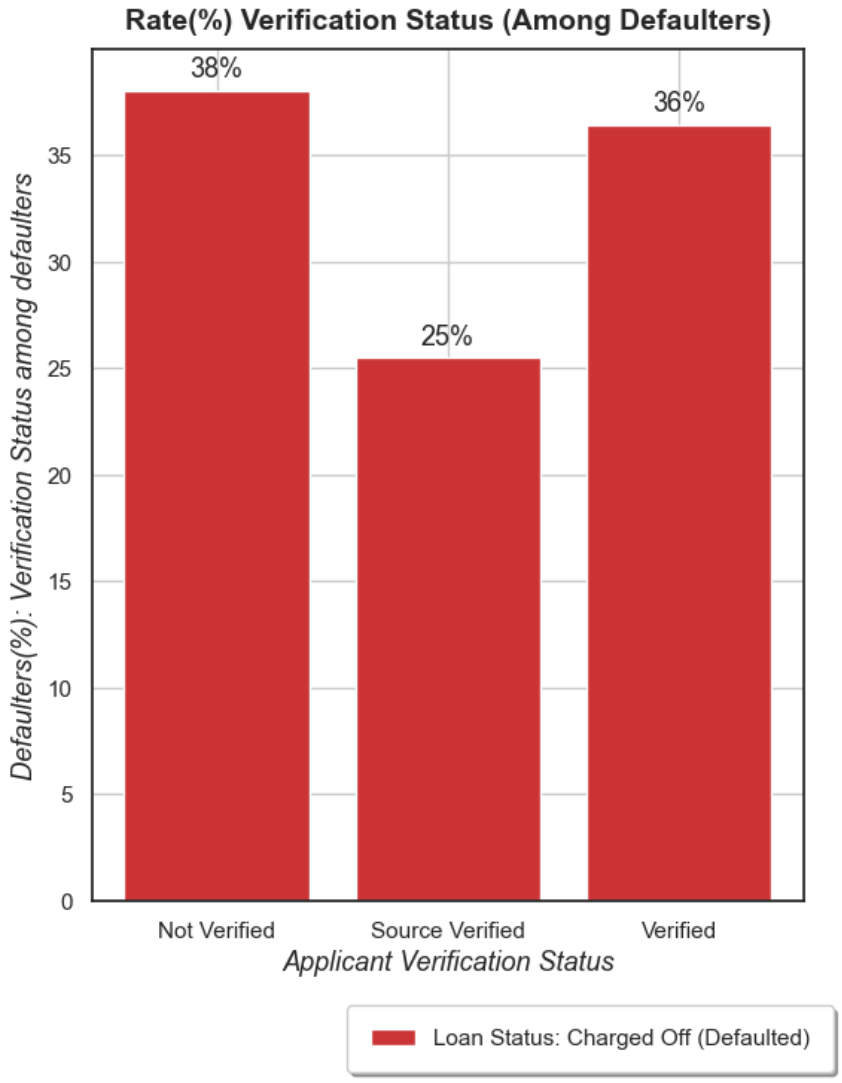


- Average Interest rates increases the risk associated with the applicants increase, which can be denoted by the grades. Assigned to applications. It is an effective process to follow.
- Loan approved for 36 months seems to have a greater % of defaulters as compared to loans approved for 60 months term.

91% of the defaulters already have a mortgage taken out or live on rent.



38% of defaulters are not verified.



Conclusion

Driving Factors (or driver variables):

Grade: Default Rate is high in high-risk loan applicants. It would be important for Lending Club to thoroughly vet high risk loan applications.

Annual Income: Applicants from 'Low'(<=45K USD) and 'Medium'(45K-90K USD) income group have a greater share of defaulted loans.

Employment Length: Maximum number of defaulters have 10/10+ years of experience and 0 to 2 years of experience. Hence, Lending Club should be taking this aspect into consideration while lending loans.

Loan Purpose: The top two reasons for loans are debt consolidation and credit card. Such applications should be carefully assessed.