

Lending Club Case Study

Team: Souvik Sen & Manish Gautam



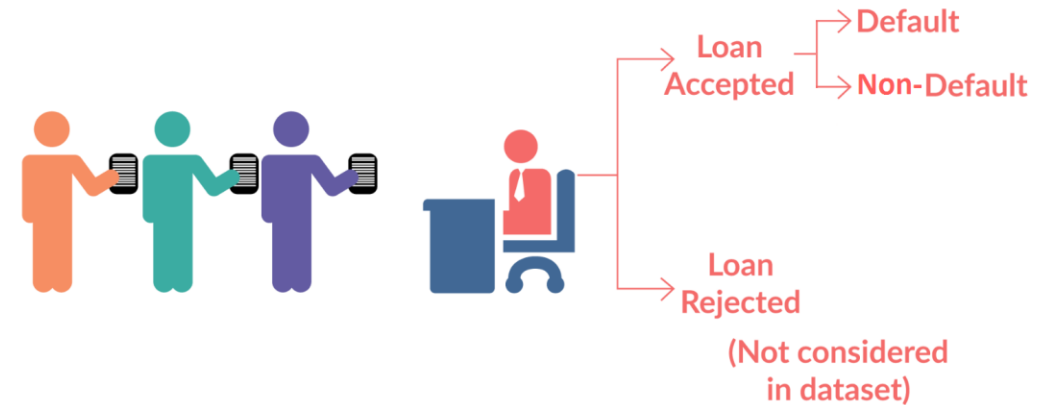
Problem Statement

A consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. 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

Provided data contains the information about past loan applicants and whether they 'defaulted' or not. 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 (to risky applicants) at a higher interest rate, etc.

LOAN DATASET



EDA Analysis

Performed all steps of EDA analysis :

- Data Sourcing
- Data Cleanup
- Univariate Analysis
- Bivariate Analysis
- Derive Metrics



Data Clean-up

- 1 Removed columns where we just have one value for all rows
- 2 Removed column "Emp Title" which has lots of missing values and is not adding any quality to the analysis.



Univariate Analysis

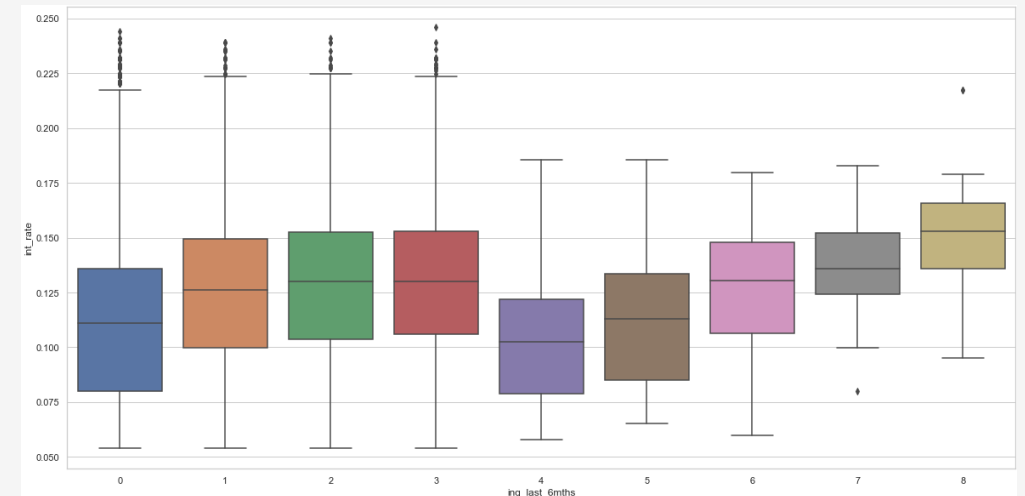
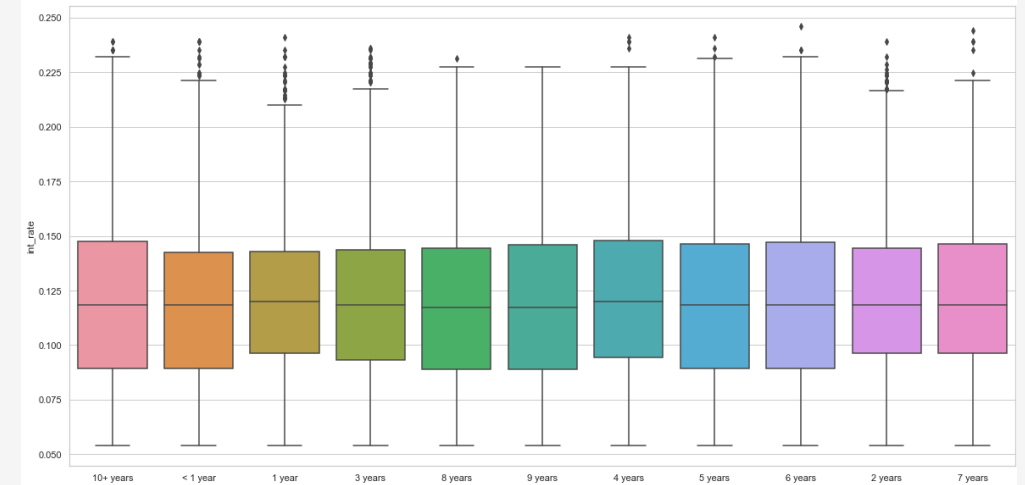
Divided the data into 3 types of data columns:

- CUSTOMER DEMOGRAPHICS, which will include all the customer related information
- LOAN CONSTRAINTS, that lending company applies on the loan like interest rate
- LOAN OUTCOME, like payment, dues, etc.

Since our goal is to understand the effect of customer application on the loan outcome, our primary focus will be on the relationship between CUSTOMER DEMOGRAPHICS and LOAN OUTCOME

Conclusion:

- Unordered categorical variables follows power law for but are not sufficient to give insightful information
- Applications with 5 years term is riskier than with 3 years term
- Emp length and verification status does not have significant impact on risk
- The higher the no. of delinquencies in last 2 years, higher is the risk
- An optimum # of inquiries in last 6 months in best. Too many inquiries or too less inquiries are both signs of higher risk



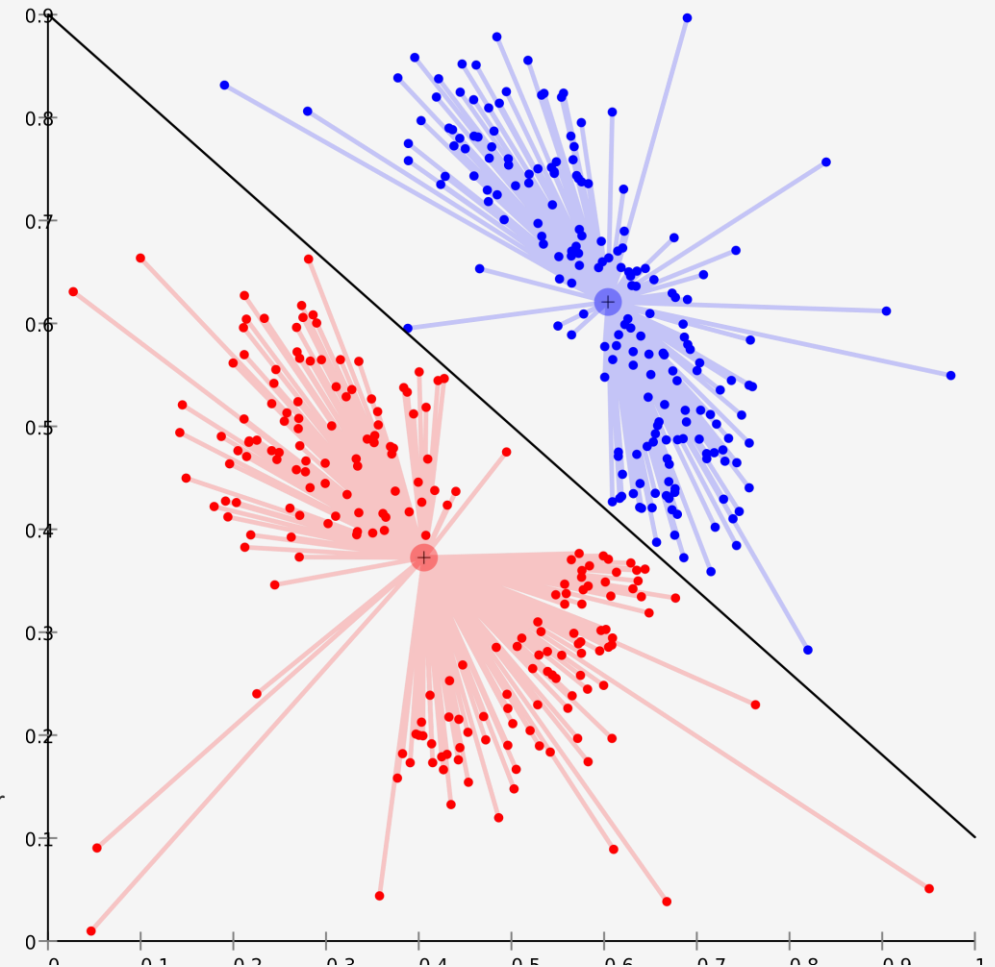
Bivariate Analysis

Did bivariate analysis on below data points:

- home_ownership & term, emp_length, verification_status, delinq_2yrs
- How interest rates (and segments) are related to loan status (mainly defaulting %)
- How loan terms are related to loan status (mainly defaulting %)
- How employment terms are related to loan status (mainly defaulting %)
- How grade and sub-grades are related to loan status (mainly defaulting %)
- How interest rates (and segments) are related to loan terms
- How home ownership is related to loan status (mainly defaulting %)
- How verification status are related to loan status (mainly defaulting %)
- How annual income (and segments) are related to loan status (mainly defaulting %)
- How state (location) are related to loan status (mainly defaulting %)
- How bankruptcy number is related to loan status (mainly defaulting %)
- How loan purpose is related to loan status (mainly defaulting %)
- How ratio of loan amount with annual income is related to loan status (mainly defaulting %)

Conclusion:

- Applicants with purpose as 'education' and term '60 months' are riskier
- Loans for small business are risky however, small business and 60 months term are even riskier
- Applicants with purpose as 'education' and emp_length '3' are riskier
- Applicant with purpose 'house' and emp_length of 6 years are riskier
- Applicant with home_ownership as 'other' and purpose as 'small business' are riskier



Derived Metrics

Derived multiple metrics from existing data:

- Segment for interest rates : Low, Medium, High and Very high
- Loan amt / Annual Income Ratio
- Segment for Loan amt / Annual Income Ratio :Low, Medium, High and Very high
- Segment for annual income :Very Low, Low, Medium and High

Conclusion:

- Applicants with 1 or more Bankruptcy or very low income (<25000 annually) or unemployed or when loan amount is more than 50% of the annual income are very riskier
- Applicants with Low income (<100000 annually) or when loan amount is between 25% to 50% of the annual income or loan purpose is for setting up Small Business and riskier



Thank you!

