Lending Club Case Study

Abhishek Kaushik

Harsha

Overview

Goal:

Lending club, the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures wants to identify factors that indicate a risky customer, which might turn default causing immense loss to investor.

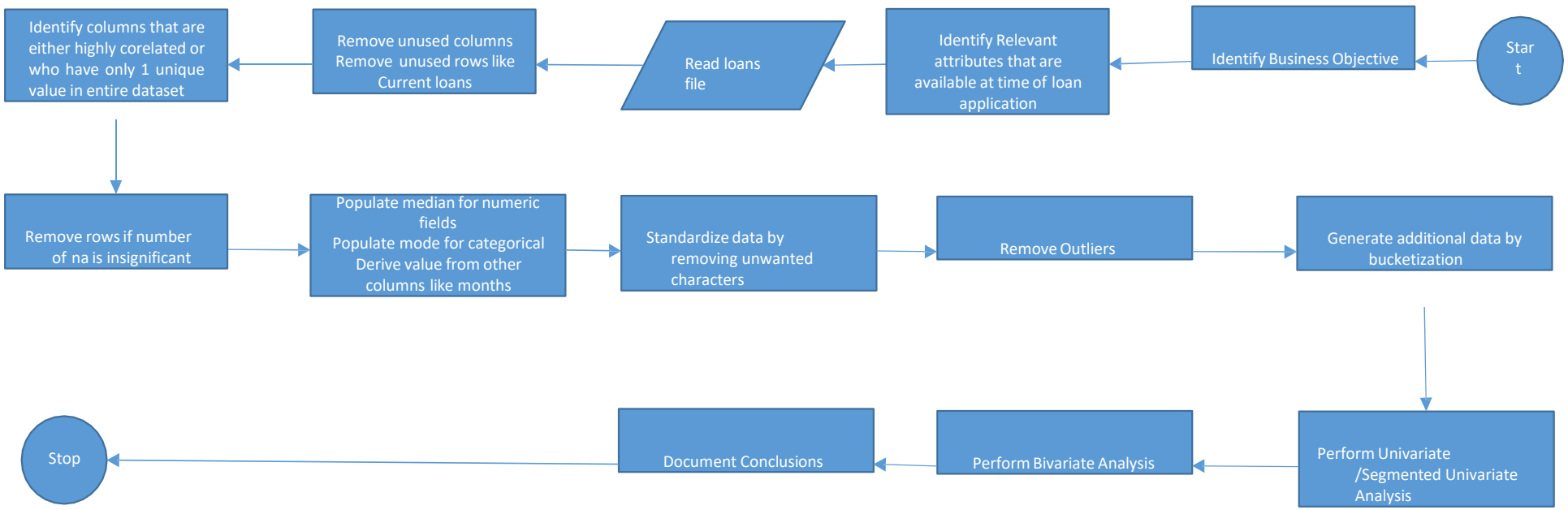
Key Requirements

- Identify factors that indicate a default by performing univariate/bivariate analysis

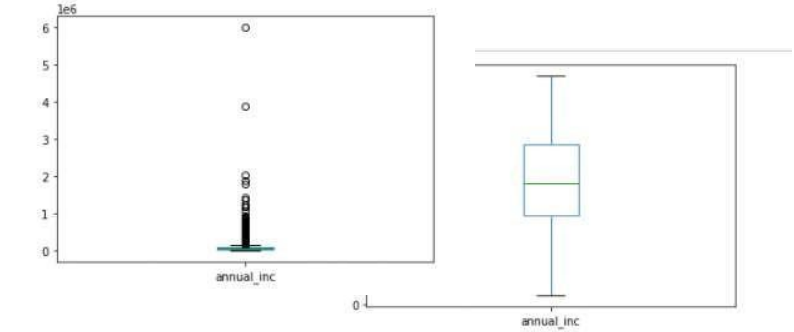
Key Inputs

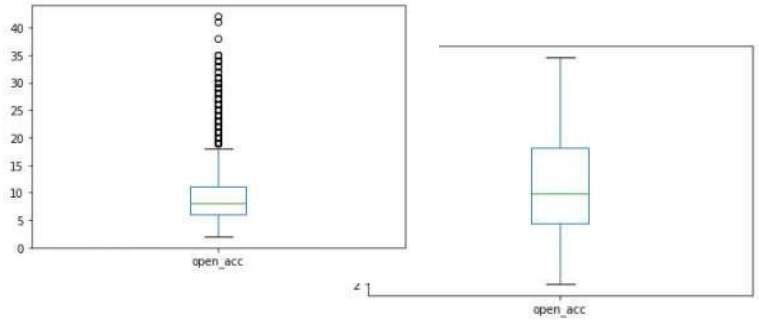
- Schema file containing details for loan table
- Loan table itself

Approach


Start
Identify Business Objective
Identify Relevant attributes that are available at time of loan application
Read loans file
Remove unused columns
Remove unused rows like Current loans
Identify columns that are either highly corelated or who have only 1 unique value in entire dataset
Remove rows if number of na is insignificant
Populate median for numeric fields
Populate mode for categorical
Derive value from other columns like months
Standardize data by removing unwanted characters
Remove Outliers
Generate additional data by bucketization
Perform Univariate /Segmented Univariate Analysis
Perform Bivariate Analysis
Document Conclusions
Stop

Outlier Detection & Bucketing of Data*





Annual_inc	Annual_inc_range
0-25000	25000 bucket
25000-50000	50000 bucket
…	….

int_rate	int_rate_range
0-3	0
3-6	3
6-9	6

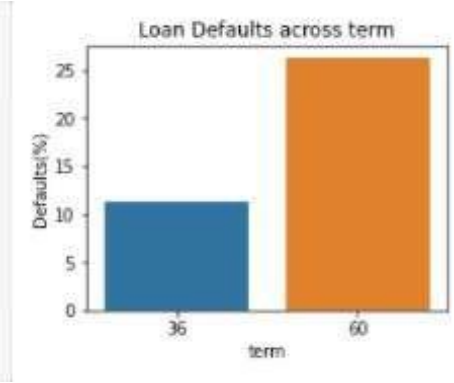
- Data was skewed with high income borrows.
- There were few individuals with high number of open accounts.

** These are only some examples*

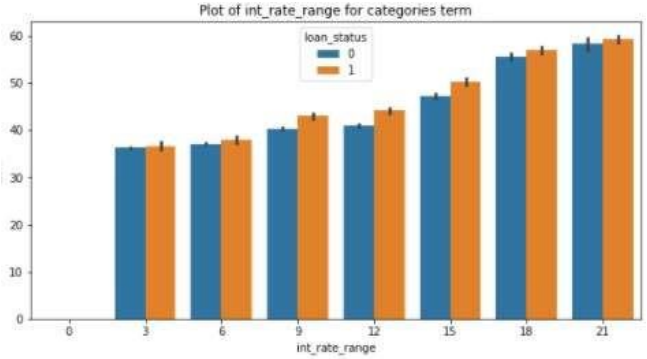
- Bucketized data into specific buckets
- We could have gone with text buckets but would have lost some advantage that we get with integers

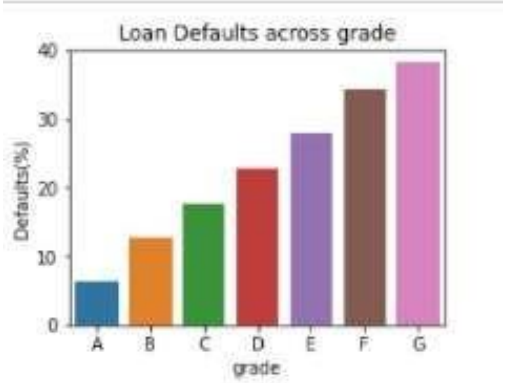
Analysis & Findings

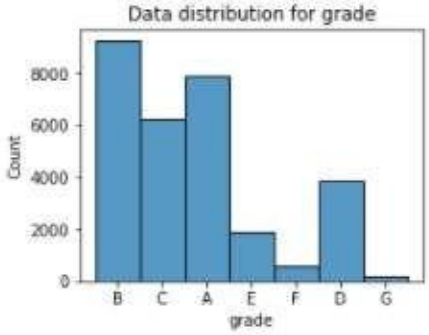
Term & Grade



- Ratio for defaults in 60 month loan is more than 36 month loan
- Defaults also increase with increases in interest rate.
- As the grade lowers the chances of default are more. Based on current data most loans are issued for grades up to C.

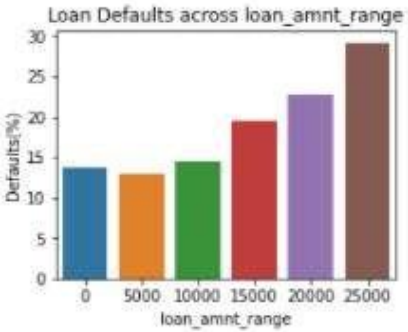




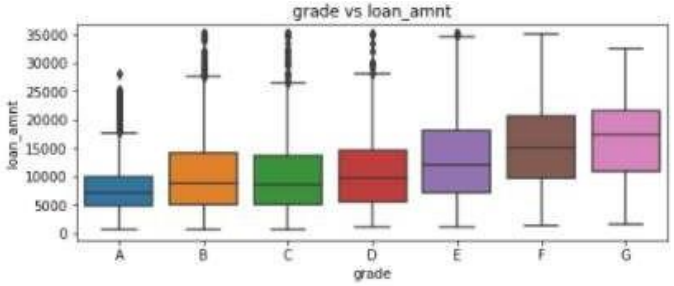


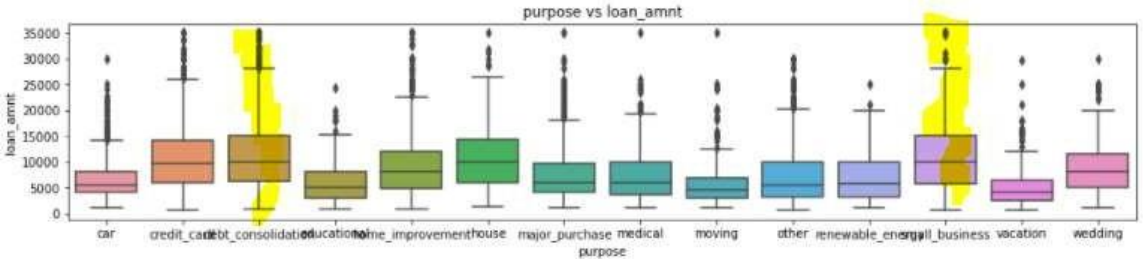
Analysis & Findings

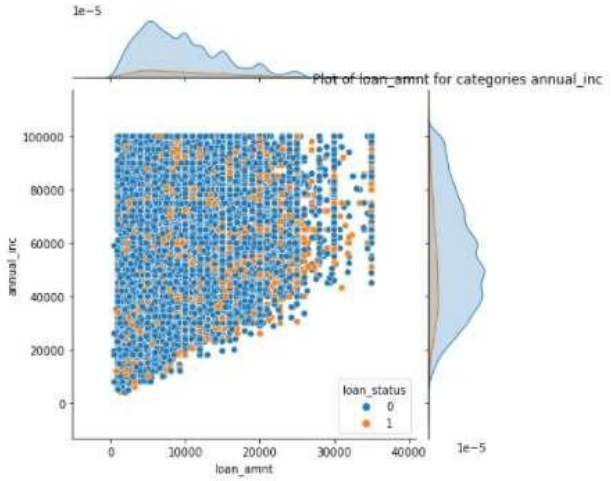
Loan amount



- People tend to take loans in range of 8k. We do see a tendency of people to default if loans are above 15k.
- From above it is evident that loan amount is more for lower grade and higher tenure.
- Looks more loan is granted to small business and debt consolidation
- Looks loans disbursement has increased significantly from 2009. This is likely year we had a recession starting



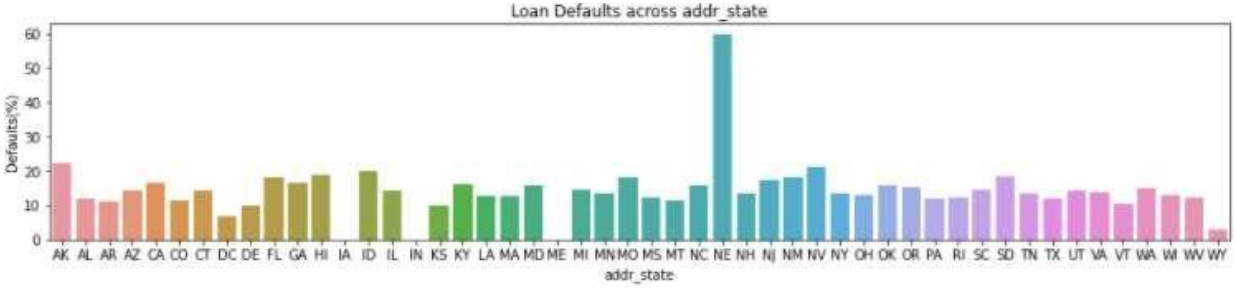


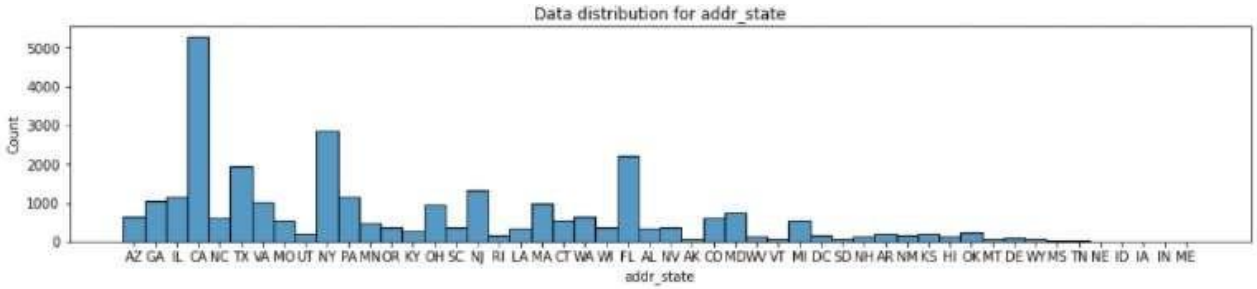


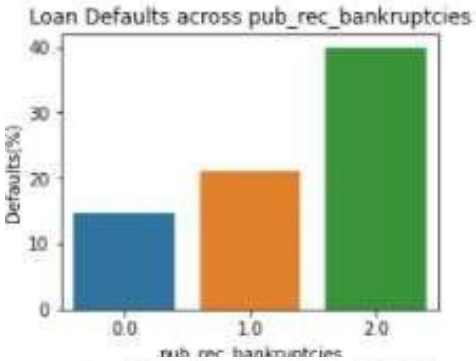
Analysis & Findings

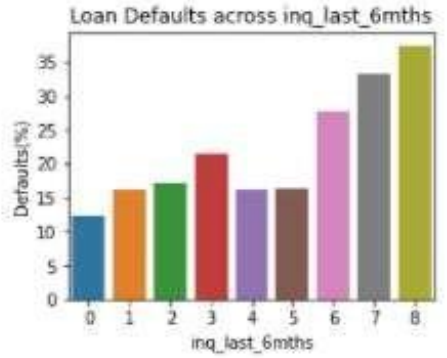
State, Inquiries in last 6 months &Public Recorded Bankruptcies

- NE state has maximum default of around 60%. Loans issued over there are risky. CA has asymmetrically high number of loans issued
- Chronic defaulters tend to default again.
- Increase in loan inquiries tend to increase chances of default



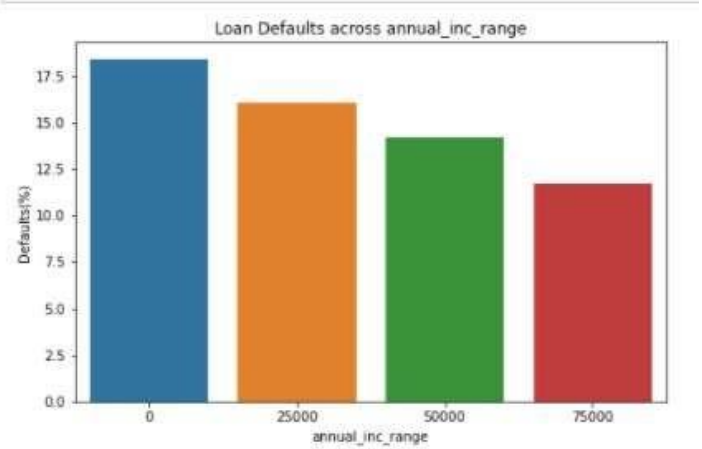




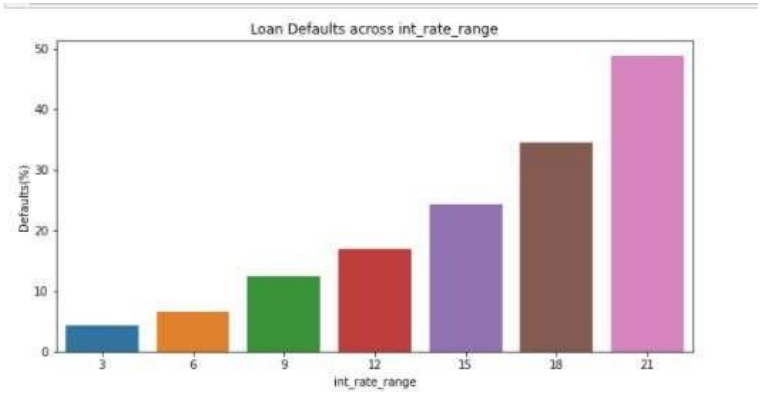


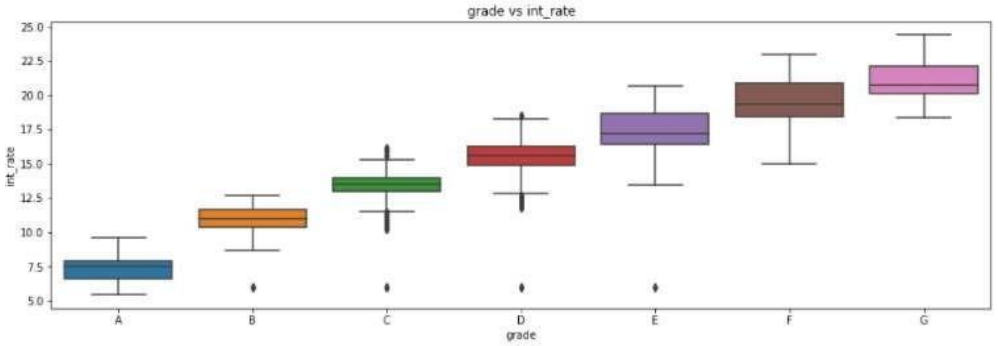
Analysis & Findings

Annual Income & Interest Rate

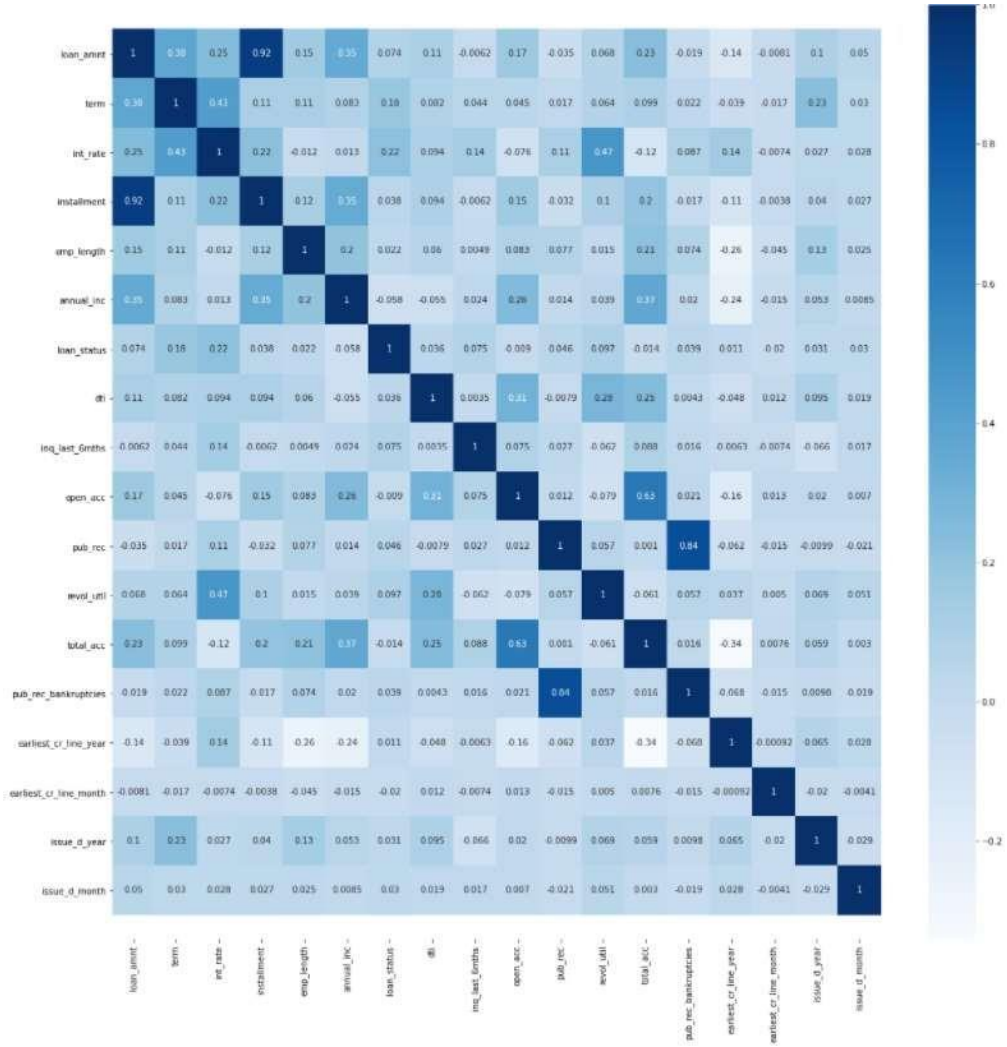


- Lower income group have higher chances of default
- An increase in interest rate leads to more default
- As expected lower grade have more interest rate than lower grade loans





Conclusions - 1



- there is strong correlation between bankruptcies and public records
- there is fair correlation between total_acc and open_acc
- there is strong correlation between installment and loan_amt

Conclusion - 2

- There is a higher chance of default for loan amount(*loan_amnt*) greater than 15000
- Higher the *term* the more chances of failure
- Lower the *grade* the higher the default rate
- Lower the income*(annual_inc)* the higher the default rate
- More the *int_rate* the higher chances of default
- NE state(*addr_state*) seems to be leading in defaults by 60% default rate. In terms of volume CA leads the pack
- More than 6 enquires(*inq_last_6mths*) are likely to cause default
- Bankruptcies (*pub_rec_bankruptcies*) has impact on default rate