Lending Club Case Study

Mayank Kumar & Shashwata Chattopadhyay

Outline

1. Problem Statement
2. Problem solving approach
3. Understanding the Dataset
4. Cleaning the Dataset
5. Data Analysis
6. Summary of Driving Factors for Loan Defaults

Problem Statement

- Our company is into consumer finance and gives loans to customers with varying interest rates & loan periods and always there is always a risk of defaulting on the loan by the borrower.
- Our objective is to identify the driving factors behind such non-repayment of loans by borrowers, so that our company can take appropriate measures to minimise risk of default.

Problem-Solving Approach

The problem has been solved using the Exploratory Data Analysis (EDA) process as follows:

1. Understand the dataset and the domain
2. Clean the dataset
3. Analyse the data using appropriate techniques:
    1. Segmented Univariate analysis
    2. Bivariate analysis on categorical variables
4. Identify driving factors and make recommendations

Understanding the Dataset

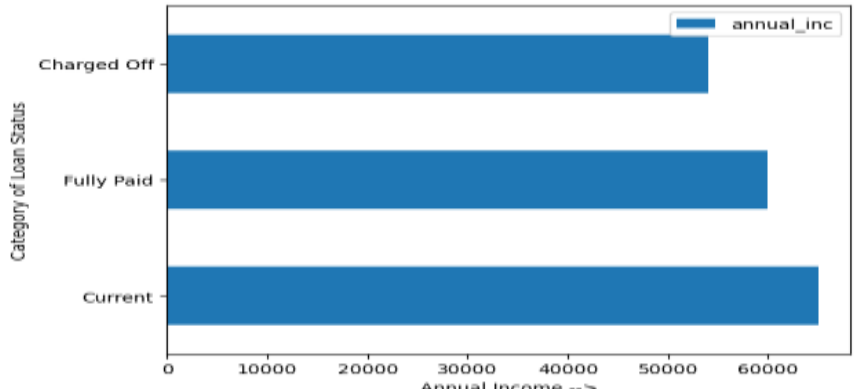
1. The dataset provided by the company has different attributes of the borrowers, starting with the Customer ID.
2. The values in the dataset are of mixed type, meaning some columns have integer data, some have floating point data while more others are data in string format.
3. There are a large number of columns having Null data (NA).
4. The dataset contains total 39717 rows, containing unique customer records.

Cleaning the Dataset

1. The columns having more than 90% Null records (“NA”) are removed from the dataset.
2. The rows having NaN records are removed.
3. The columns that do not contain relevant information for this EDA are removed:
    1. Customer specific information like the customer ID/member ID.
    2. Columns with single value repeated in all rows (=0) like **acc_now_delinq, collections_12_mths_ex_med, tax_liens, delinq_amnt, etc.**

Data Analysis – Effect of Income

- People will be forced to default on loans when they do not have the money to pay for the loan (due to low income level as compared to expenditure).
- So, our first target variable is the income of the borrower. Since the income column has outliers, we consider the **MEDIAN values** for the income in the different groups – "Charged Off" and "Fully Paid". We neglect the group "Current" in all our analysis as these loans are presently running and will not give the full picture.
- We created a PIVOT TABLE grouped by loan status for the 3 factors – annual income, interest rate and loan amount. It can be observed that the annual income was low while the loan amount was high in the "Charged Off" group as compared to the "Fully Paid" group. Also the interest rate is found to be higher for the "Charged Off" group as compared to the "Fully Paid".



loan_status	annual_inc	int_rate	loan_amnt
Charged Off	54000.0	13.61	10000
Current	65000.0	14.65	16000
Fully Paid	60000.0	11.49	9600

Data Analysis – Effect of “dti”

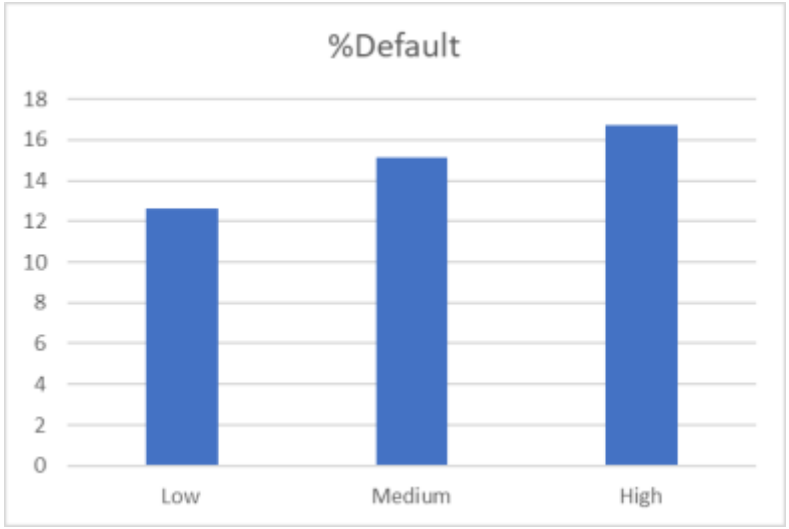
- Since a combined trend is observed for the Annual income and the loan amount, we also looked at the ratio between the two variables which is captured in the “dti” column of the dataset.
- The data was grouped into 3 categories based on the “dti” values (Low, Medium and High) and the percentage of defaults was calculated:

  %defaults = Nco / (Nco + Nfp) x 100

  where,

  Nco = no. of charged Off,

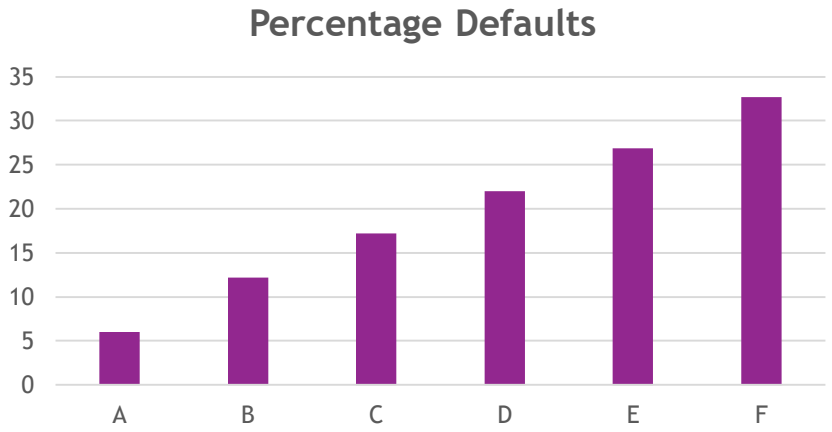
  Nfp = no. of Fully Paid



- A clear trend has emerged for the effect of the “dti” on percentage of default, so ***it is recommended to check the “dti” of the applicant before passing the loan application***.

Data Analysis – Effect of “Loan Grade”

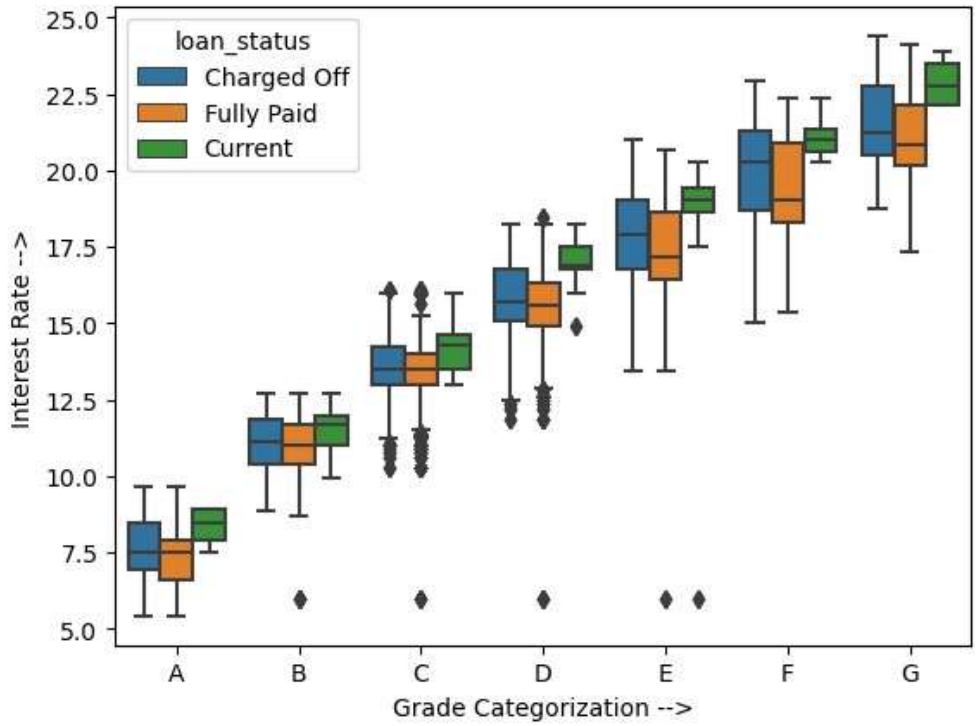
- Loan company has assigned Loan Grades for the different borrowers, so we checked for the relation between the Percentage of defaults and the Loan Grade.
- The percentage of defaults is found to increase as the loan grade increased from “A” to “F” and the trend is found to be significant.



- ***The company is doing a good job in assigning the Loan Grades as it is a good predictor of the chances of default.***

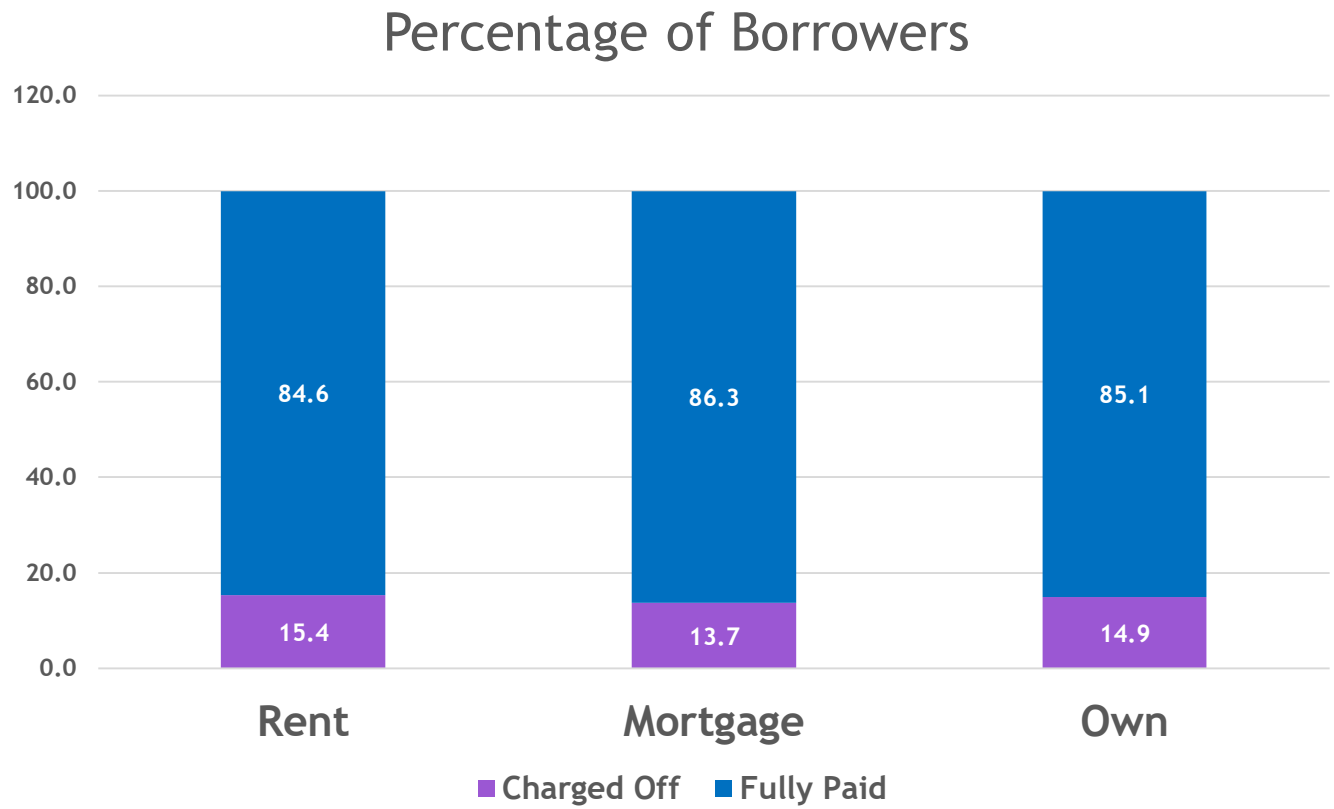
Bivariate Analysis of Interest rate and Loan Grade

- The interest rate set by the company has been plotted against the Loan Grade in below figure and positive correlation can be found.



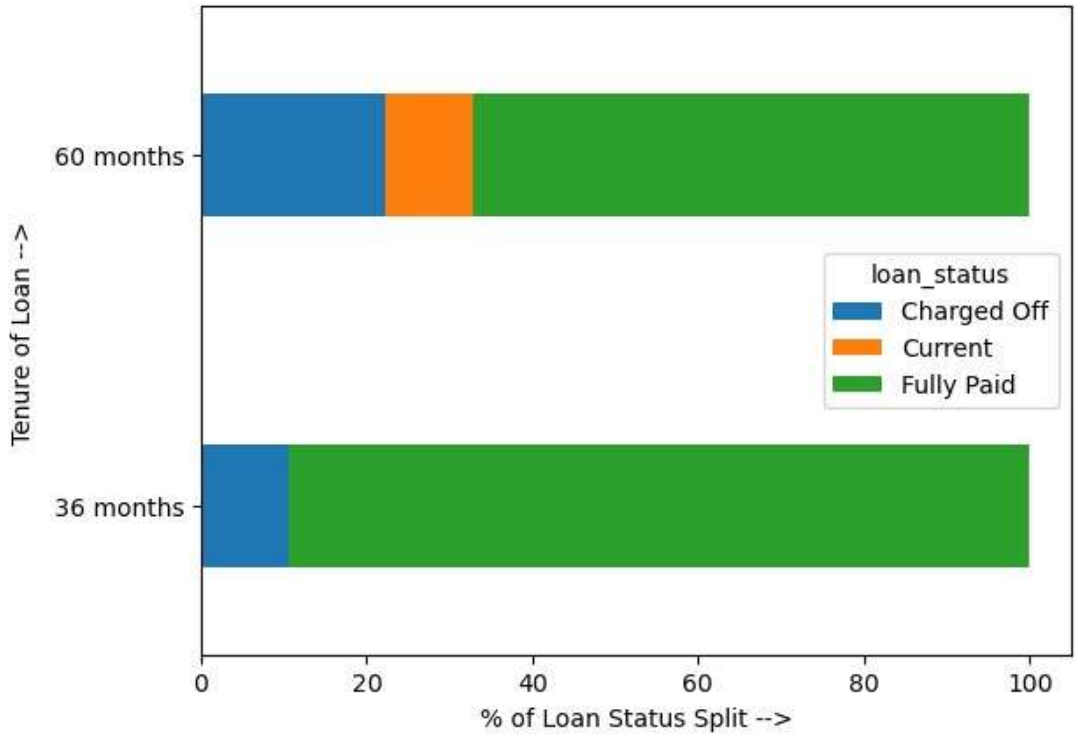
Data Analysis – Effect of House Ownership

- The effect of house ownership (rent/Mortgage/Own) on the percentage of defaults (charged off) was found to be too small to be relevant.



Data Analysis – Effect of Loan Tenure

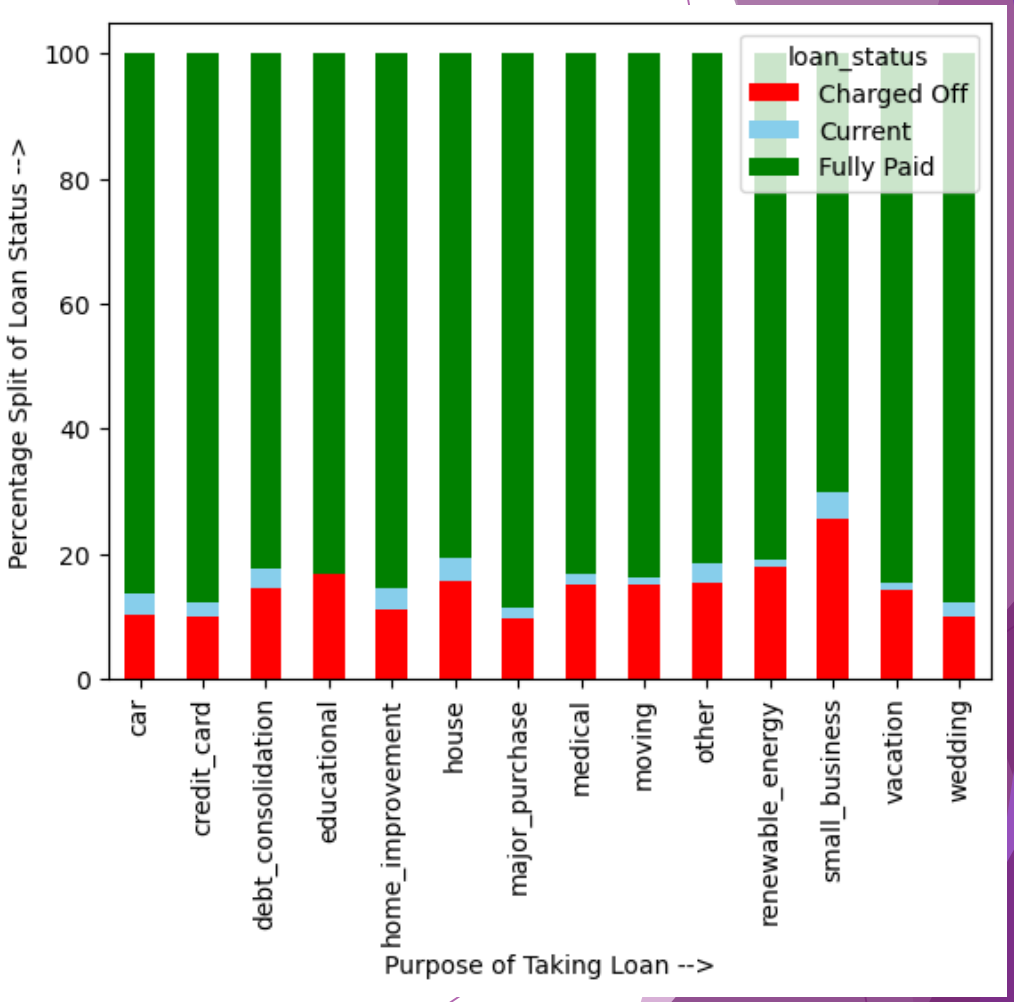
- The effect of the loan tenure on the percentage of defaults (charged off) was found to be significant.
- The interest rate was found to be higher for higher loan tenure (which is usually done by the company as the longer loan tenure carries more risk of default).



loan_status	term	annual_inc	installment	int_rate	loan_amnt
Charged Off	36 months	50000.0	262.330	12.53	8000
	60 months	59020.0	348.385	16.02	15000
Current	60 months	65000.0	369.210	14.65	16000
Fully Paid	36 months	58000.0	264.440	10.99	8000
	60 months	65000.0	318.220	14.27	14500

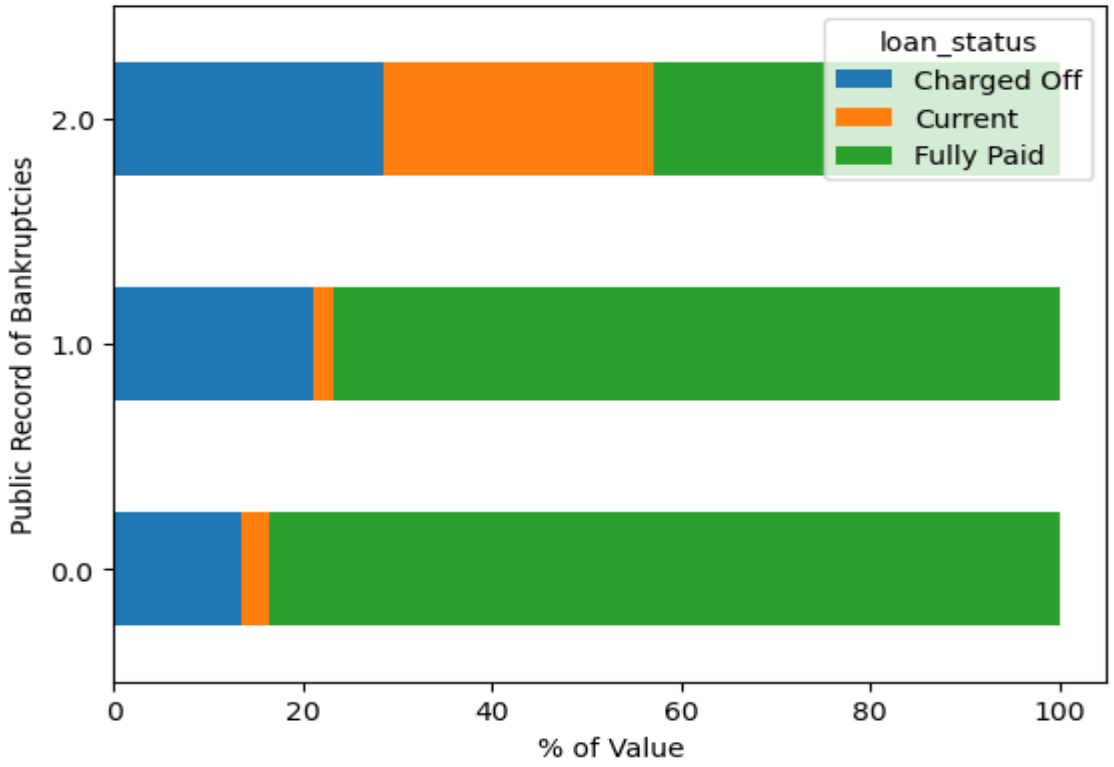
Data Analysis – Effect of Loan Purpose

- The effect of the purpose of taking loan on the percentage of defaults has been analysed in the adjoining figure.
- The “small business” category has the maximum percentage of defaults (charged off), followed by “renewable energy”, so it is recommended to carefully analyse such loan applications to avoid defaults.
- “Major purchases” and “Wedding” have emerged as the safest categories with lowest percentage of defaults.



Data Analysis – Effect of Bankruptcy Record

- People with previous record of bankruptcy are found to have a higher tendency for defaulting on the loans as evident from the below figure.



- ***It is recommended that the company avoids giving loans to persons who have previous records of bankruptcies.***

Summary of Driving Factors for Loan Defaults

- The following driving factors for loan defaults have been identified in our analysis:
  1. Persons with high ratio of loan amount to monthly income (dti)
  2. Persons with high loan tenures
  3. Persons assigned a higher loan grade
  4. Persons who are taking loan to start some small business
  5. Persons with previous records of bankruptcies