

Lending Club: Exploratory Data Analysis



-
-
-
-
-
-

Presenters:

Ananth Kumar Kambhampati
Venu Madhav Kanamatareddy

Agenda

- Business Problem
- Business Understanding and Objective
- Data Description
- Explanatory Data Analysis
- Conclusions



Business Problem

To decide for loan approval based on the applicant's profile.

Two **types of risks** are associated here :

1.If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company.

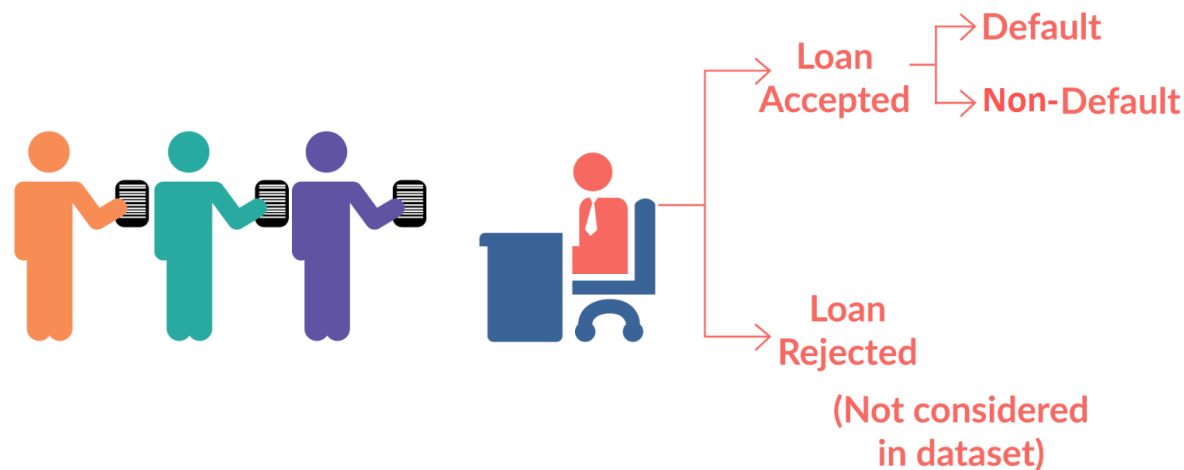
2.If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company.



Business Understanding and Objective

- There are **two types of decisions** that could be taken by the company from the provided past dataset. : **Loan Accepted** and **Loan Rejected** .
- Here we are interested in identifying the pattern for **Default** loan.

LOAN DATASET



Data Description

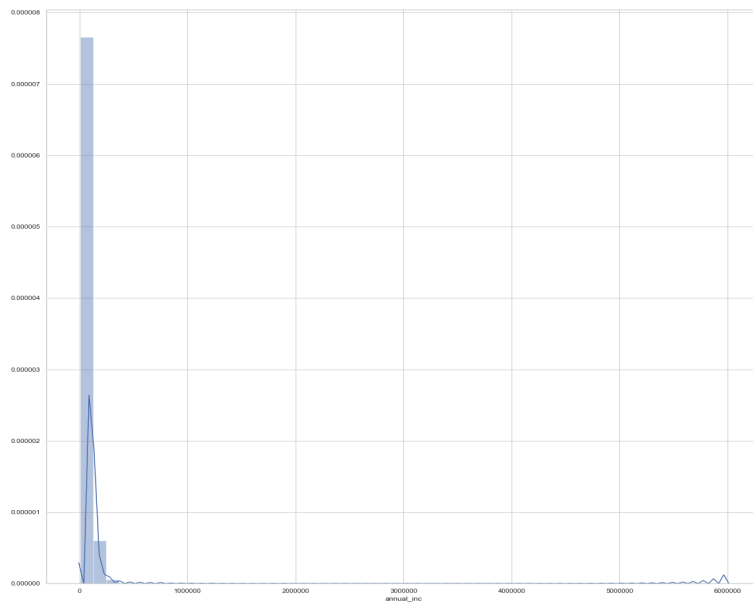
- The Data set has ~ **40,000** records .
- The variables which are used in the dataset provide ample amount of information which we could use to predict loan **Default** likelihood.
- We considered only attributes which have direct impact to borrower's potential to Default. We used feature selection techniques and business knowledge for choosing relevant variables categorized as :
 - Demographics attributes
 - Loan Characteristics attributes
- This data is structured data with lot of missing/null values. Includes continuous, ordinal and nominal feature types.
- We did not consider the “**Current**” category of loan payers.



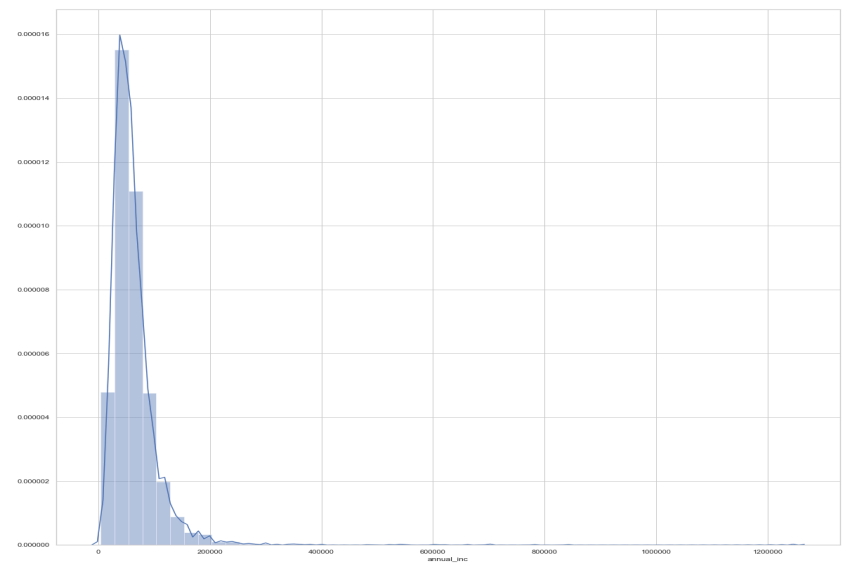
Data Explanatory Analysis

- The annual income falls under same normal distribution curve between for paid and defaulters

Paid

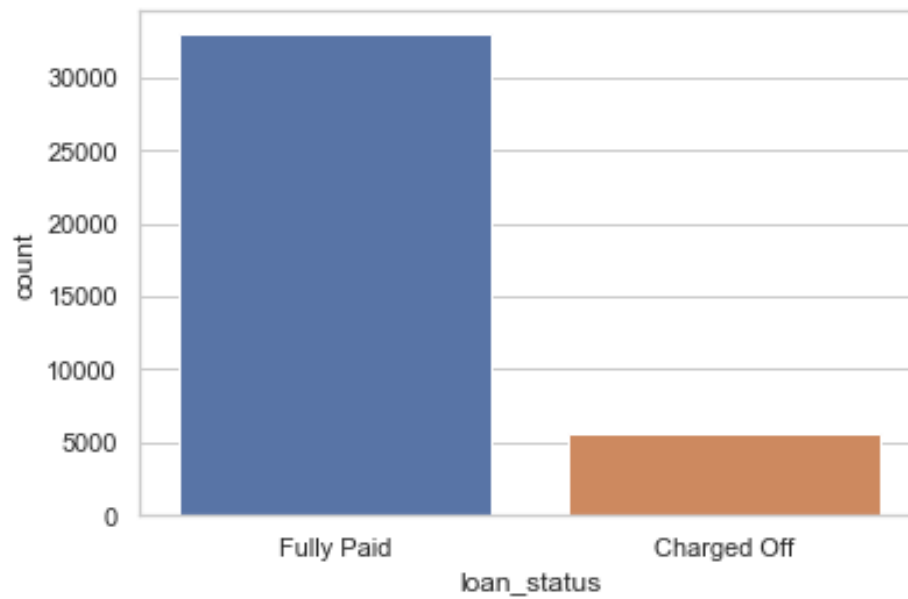


Defaulters



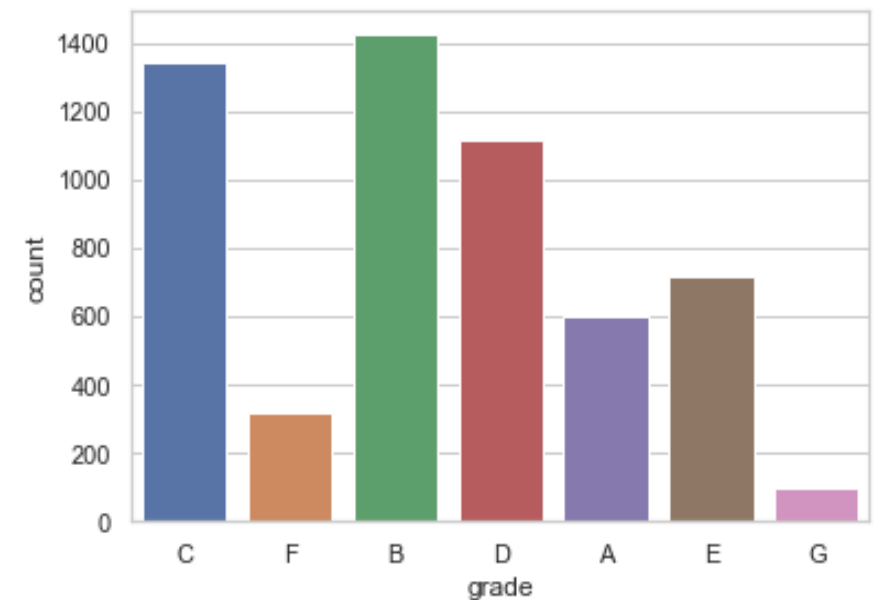
Data Explanatory Analysis

- Fully Paid and Charged Off Loan status count Bar plot



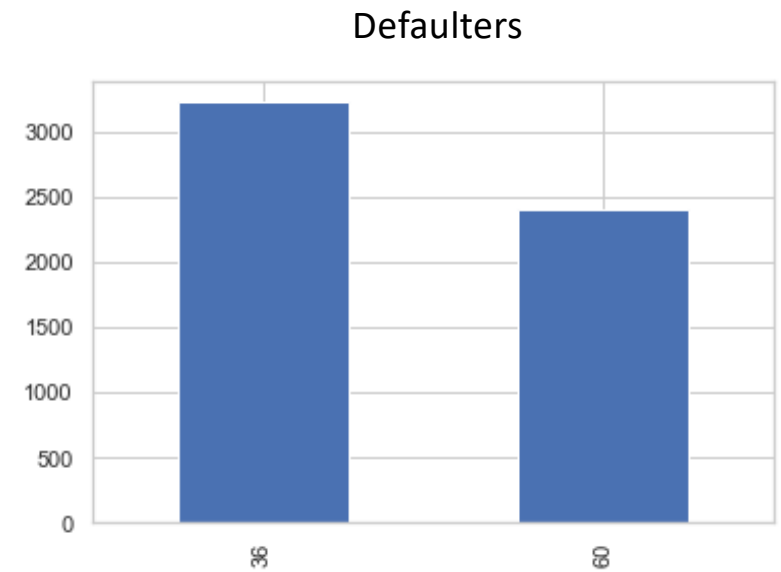
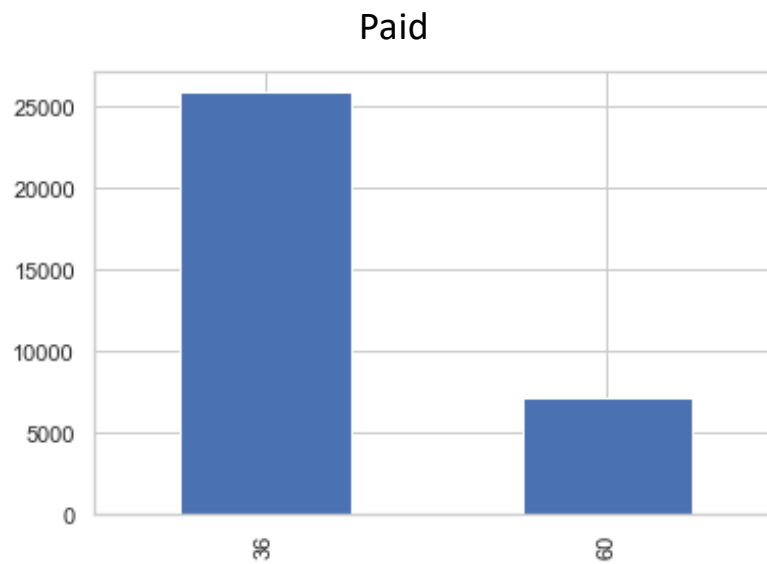
Data Explanatory Analysis

- Defaulters come from the RENT category of home ownership followed by mortgage
- Defaulters come from Grade B followed by C and D



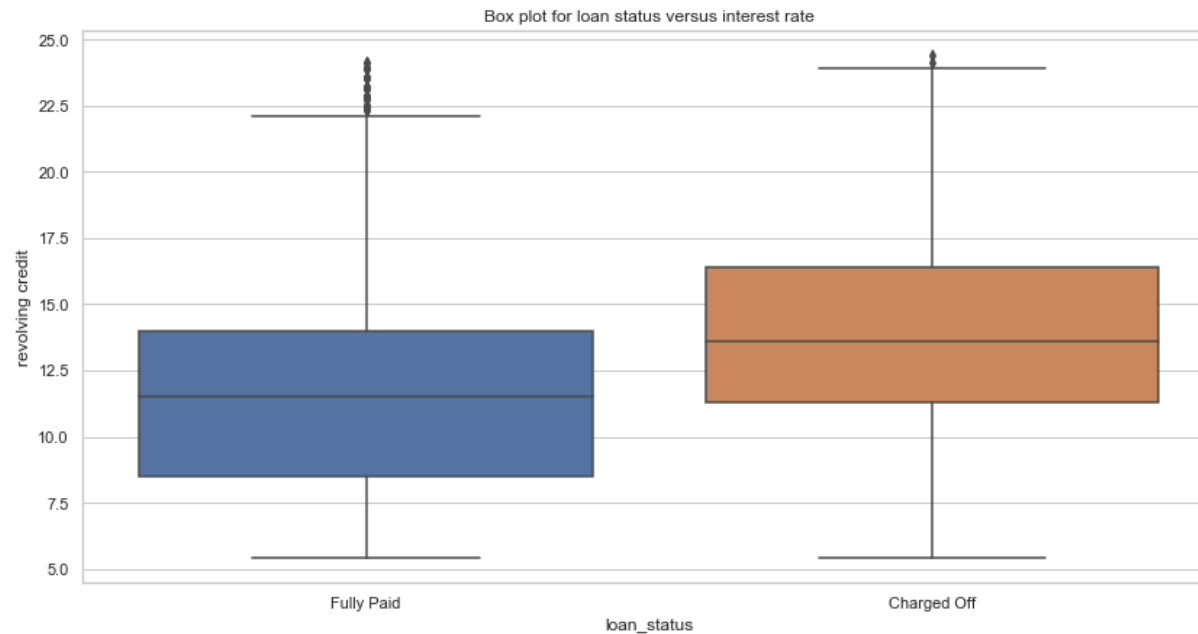
Data Explanatory Analysis

- Defaulters have more percentage of 60month term loans compared to paid borrowers



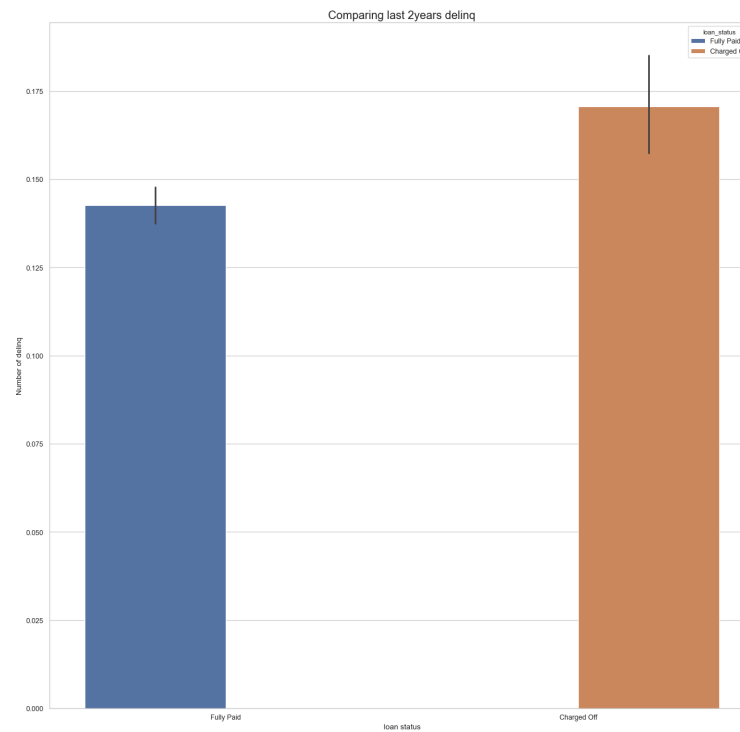
Data Explanatory Analysis

- Defaulters have the highest interest rate compared to the paid borrowers
- We can see from the above plot that higher the interest rate, higher the chances of loan getting charged off



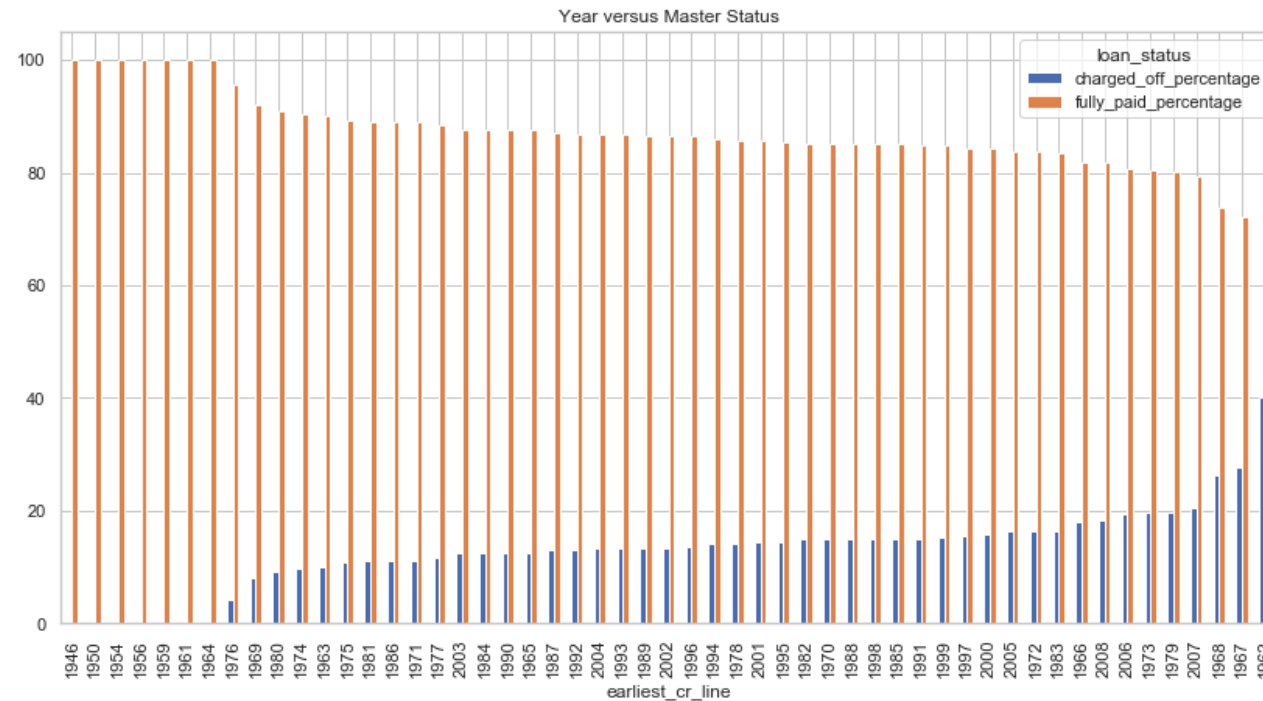
Data Explanatory Analysis

- Defaulters have high missed payments compared to paid borrowers.
- Higher the DTI(Debit-to-Inco) higher the chances for charge off



Data Explanatory Analysis

- There is some significance that people having longer credit line are more reliable.



Conclusions

- Below Applicant attributes are playing significant role in Defaulter.
 - Bankruptcy
 - Interest rate
 - Term
 - Grade – B, C, D
 - Age of creditline

