

Lending club Case study

Swathi Mamillapalli

Problem Statement

- Lending Club is Online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- Lending loans to 'risky' applicants is the largest source of credit loss.
- Identification of such risky applicants using EDA is the aim of this case study.
- What are the driving factors that decides applicant is risky applicant or not?

Analysis Approach

- Understood Data Dictionary of Loan dataset
- Data Cleaning
 - Removed columns where more than 40% of data is missing
 - Removed columns related to post approval process
 - Removed columns which has single value
 - Removed customer identification information like member id, id, url, etc.,
 - Removed/Fixed null values or missing values
 - Removed “current” applicant’s data
- Removed Outliers
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis

Understanding Features

- Quantitative Variables:

Numerical variables, which can be added / multiplied etc. Below are some fields identified:

- Loan Amount
- Approved funded amount
- Approved funded amount by investor
- Annual Income
- Loan Interest Rate
- DTI

- Ordered Categorical Variables:

Variables where there can be some kind of ordering enforced. Below are some fields identified.

- Term
- Grade
- Subgrade

Understanding Features contd..

- Unordered Categorical Variables:

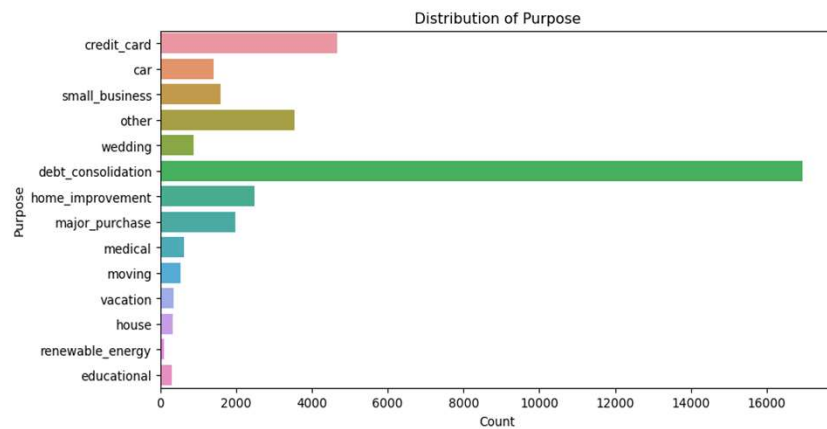
Variables where there cannot be any ordering. Below are some fields identified.

- Loan Status
- Purpose
- Home Ownership

- Target Variable:

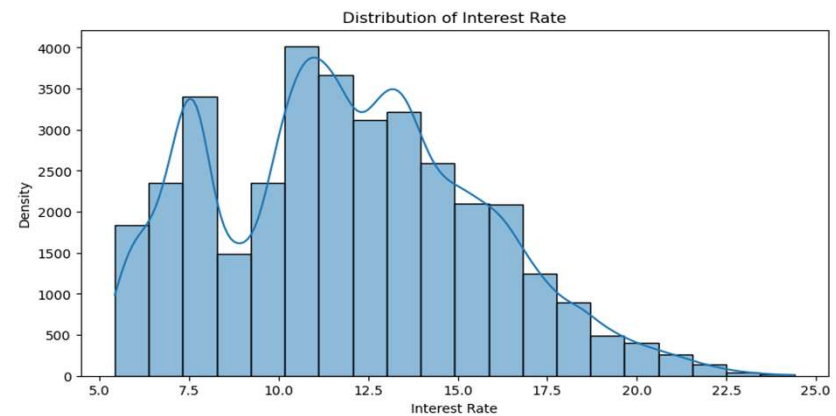
- Loan Status

Univariate Analysis



Observation :

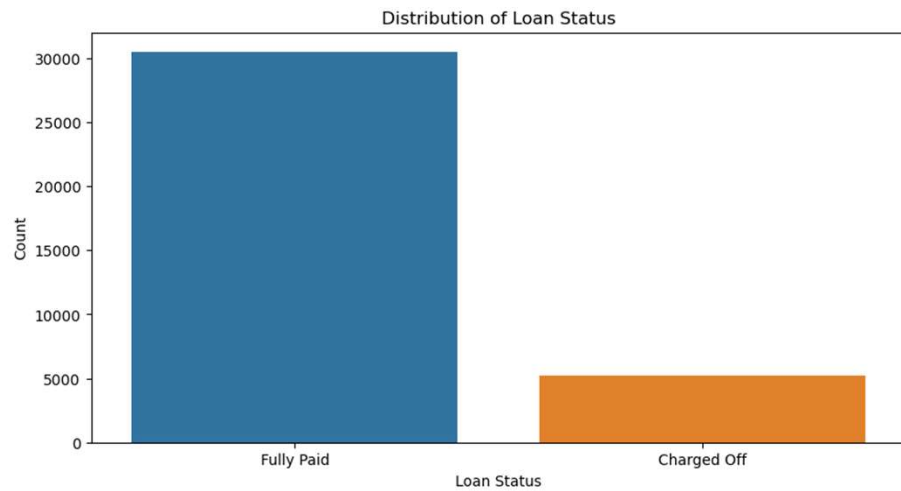
Most of the loans are taken for Debt consolidation purpose



Observation :

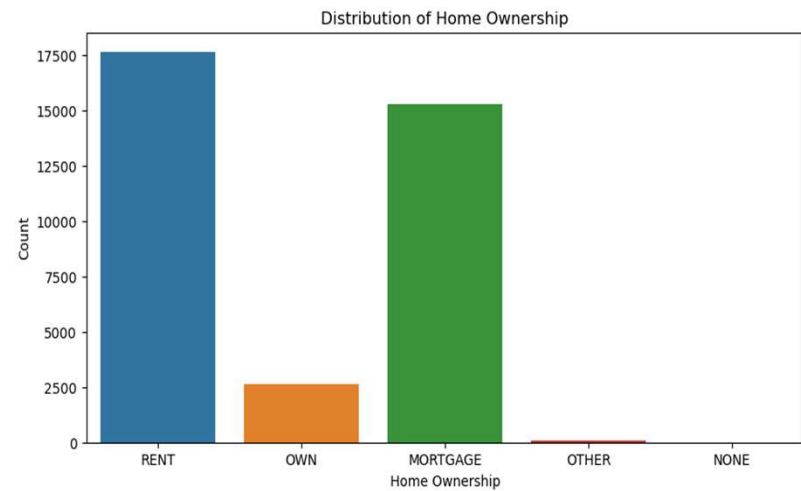
Most of loans are given at 5 - 10% and 10 - 15 % interest rates

Univariate Analysis contd..



Observation :

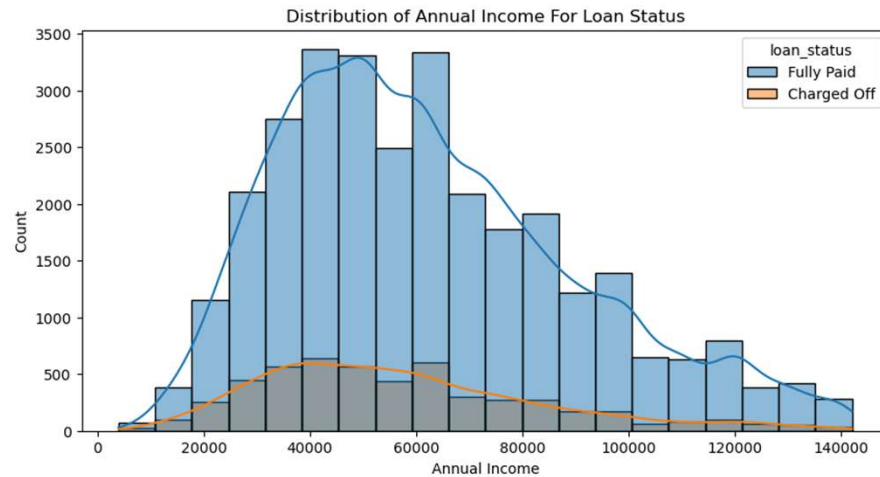
There are more "Fully Paid" customers than "Charged Off"



Observation :

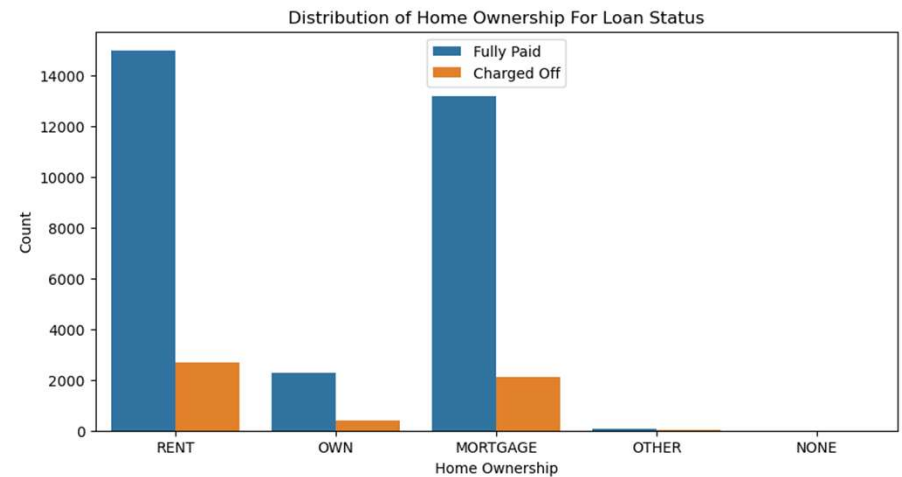
Very few customers owns house. Most of them are either on Rent or Mortgage

Segmented Univariate Analysis



Observation:

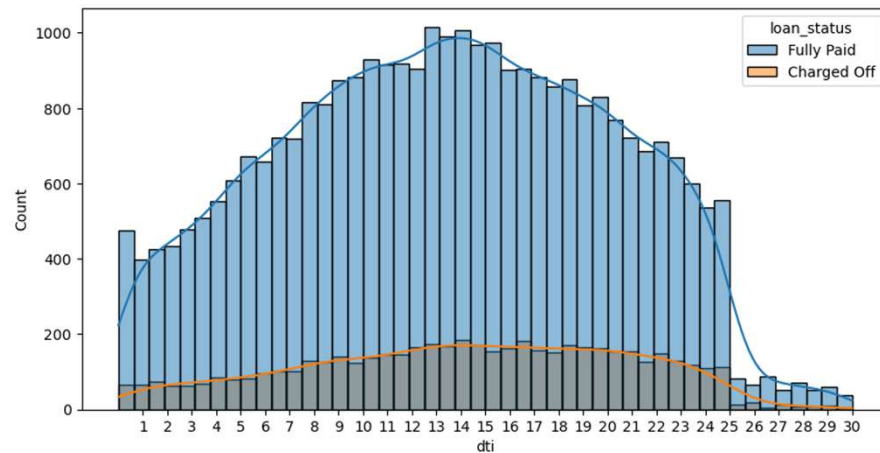
Most of the customer's Annual income in range of 40K to 60K. Customer's with Annual income more than 60K less likely to default



Observation:

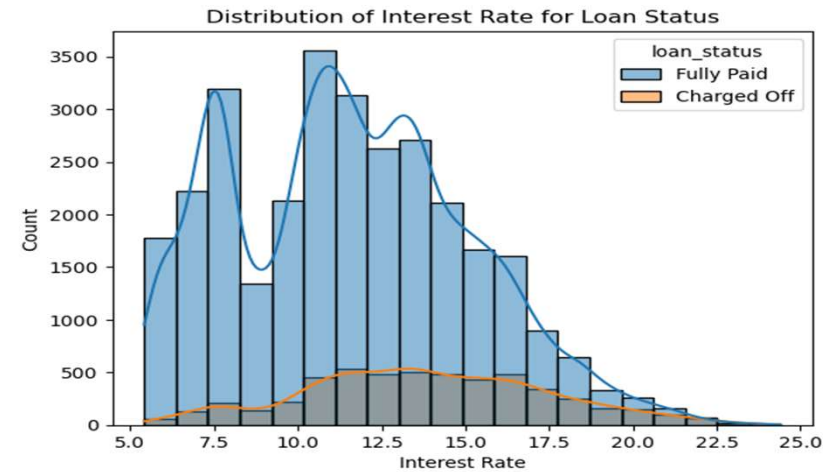
Defaulted loan are lower for the borrower which own their property compared to mortgage or rent.

Segmented Univariate Analysis contd



Observation :

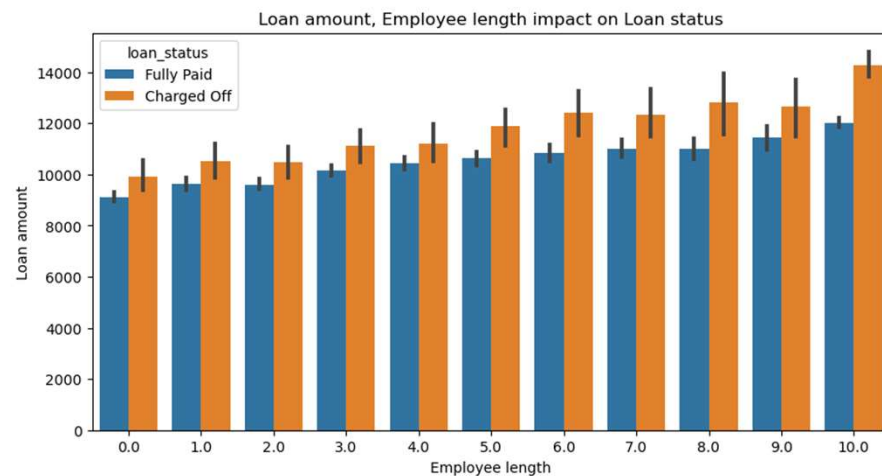
For DTI (Debit to Income Ratio) of 12-20% has high chances of charging off as compared to others. For DTI > 25% there are very less chances of charging off.



Observation:

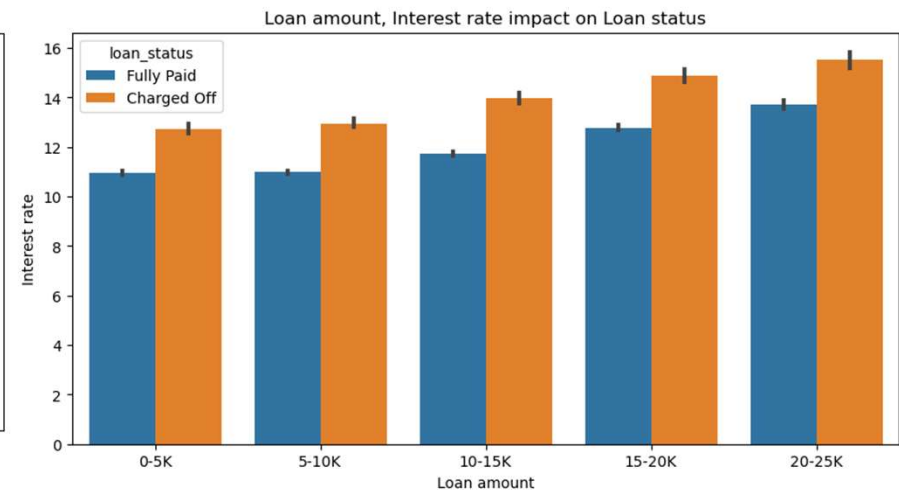
Customers who took with high interest rate are mostly likely to default. For ex: Customers who took loan at 22.5% almost 100% defaulted. Customers who took loan at 20-21% more than 50% defaulted

Bivariate Analysis



Observation:

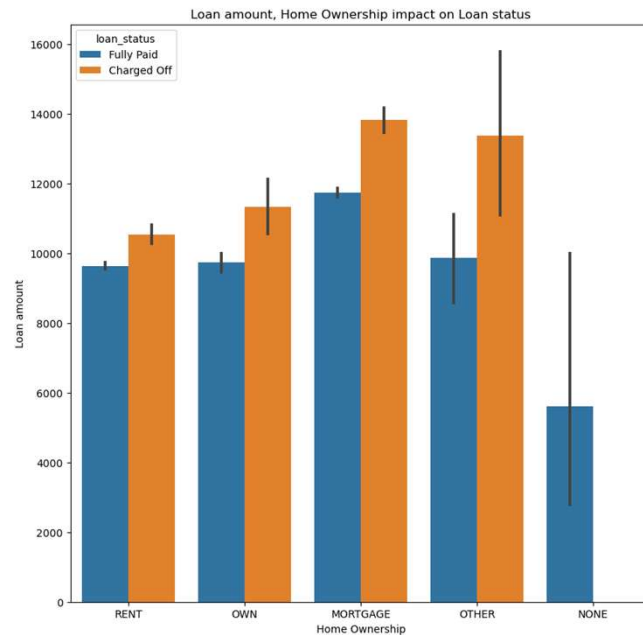
Employees with more experience have requested for large loan amount, and in this case the defaulters are more as the employee experience and loan amount increases. For loan amount > 10K, and experience, greater than 6 years the defaulter percentage is relatively higher.



Observations:

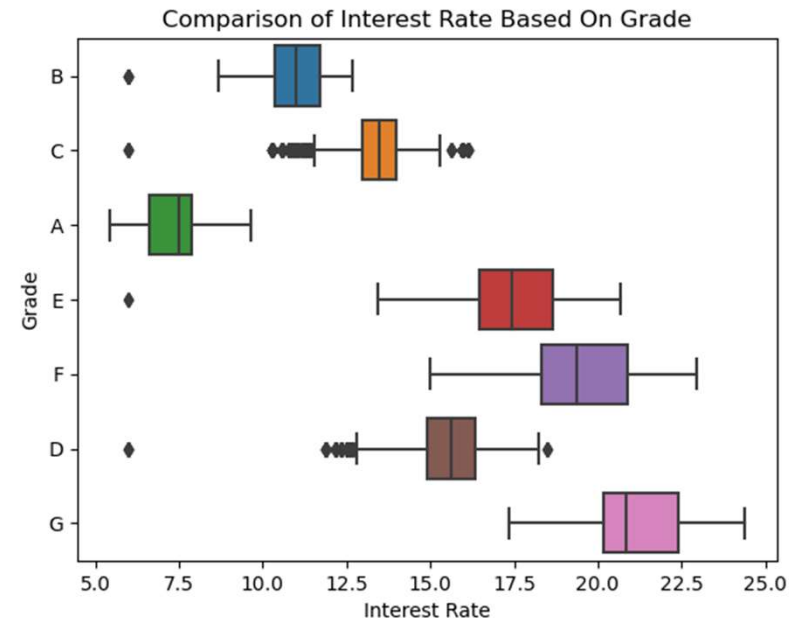
As per above graph, for any loan amount > 15000, and interest rate 14% or above there are higher chances of charging off.

Bivariate Analysis contd..



Observation:

People with "MORTGAGE" or "OTHER" home ownership category has taken large loan amounts(\$10000+), and chances of charged off is more in those home ownership categories.



Observation:

The Grade represent risk factor thus we can say interest rate increases with the risk. For example: G Grade have more interest rate whereas A grade is having less interest rate.

Insights:

From the analysis so far, there are higher chances of defaulting when:

- Borrower loan interest rate is 14% and loan amount > 15K or above
- Borrower loan term is longer (60 months)
- Borrower employee experience is > 6 years and loan amount > 10K
- Loans issued in last quarter in the year 2011
- Loans issued for the purposes of Debt Consolidation, Other
- Borrower has house on MORTGAGE or OTHER home ownership category has taken loan amounts \$10K+
- Less number of loans are completely approved after 20K loan amount
- Partially approved loans there is more % of defaulters
- loan amount 30K - 35K more customers are tend to default.
- The Grade represent risk factor thus we can say interest rate increases with the risk.
- Borrower with less than 60000 annual income.
- DTI (Debit to Income Ratio) of 12-20% has high chances of charging off as compared to others.
- Borrower from urban cities.

Recommendations

- Borrowers with home_ownership category of "OWN"
- Borrowers with Annual income > 60000
- Borrowers with loan interest rate of < 14%
- Borrowers with DTI ratio > 25%
- Borrowers with years of exp between 3 years to 6 years
- Borrowers with less term i.e 36 months.
- Borrowers whose status is verified.

Driving factors

Major Driving factors which can be used to predict the chance of defaulting and avoid Credit Loss:

- Annual income
- Interest rate
- DTI ratio
- Loan Amount
- Home Ownership
- Grade
- Term