

LENDING CLUB CASE STUDY

Submitted By
Ahmed Hussain
NV Uma Reddy

AGENDA

- Problem Statement
- Data Understanding
- Data Cleaning and Manipulation
- Data Analysis
- Data Visualization using Graphs
- Insights and Recommendations

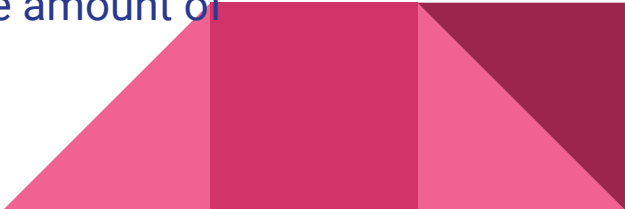


PROBLEM STATEMENT

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return. 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

1) If the applicant is likely to repay the loan, then not approving the loan results in a loss of business for 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
The aim is to identify patterns indicating if a person is likely to default, which may be used for actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.



DATA UNDERSTANDING

- We have understood the problem statement and understood the business requirement.
- Understood the metadata of the provided dataset using the provided data dictionary and loan data
- Identified the fields required to analyze and relationship between the fields.
- Removed the fields with high null values from our analysis
- Selected a set of columns which are suitable as target variables and analyzed data around these columns.
- Sliced the data based on uniqueness of the data.
- Filtered out special characters and spaces which blocks data transforming.



Data Cleaning and Manipulation

- Identified and removed the fields with complete Null and NaN values, identified numerical and character variables.
- Dropping the columns having unique data
- Finding the columns having unique values
- Sorted Unique Values based on the columns
- Dropped the columns not relevant as per the column's data uniqueness
- Dropped the rows having loan status as Current.
- Removing the % from int_rate column as all records store data in percentage
- The term column is converted to integers by truncating month
- Identified the columns with missing values and imputed the missing values, for example emp_length field is imputed with 'ffill' method.



Data Analysis

The objective of the case study is to determine the factors responsible for Defaulting loan so that this analysis can be implemented to Approve/Reject Loans with minimal or zero impact to Business

Based on the provided data, we have performed the below 3 variables including the correlation between variables

- 1) Univariate Analysis
- 2) Segmented Univariate Analysis
- 3) Multivariate Analysis

There are broadly 3 types of variables

Variables related to the applicant (Demographic variables such as age, occupation, employment details etc.)

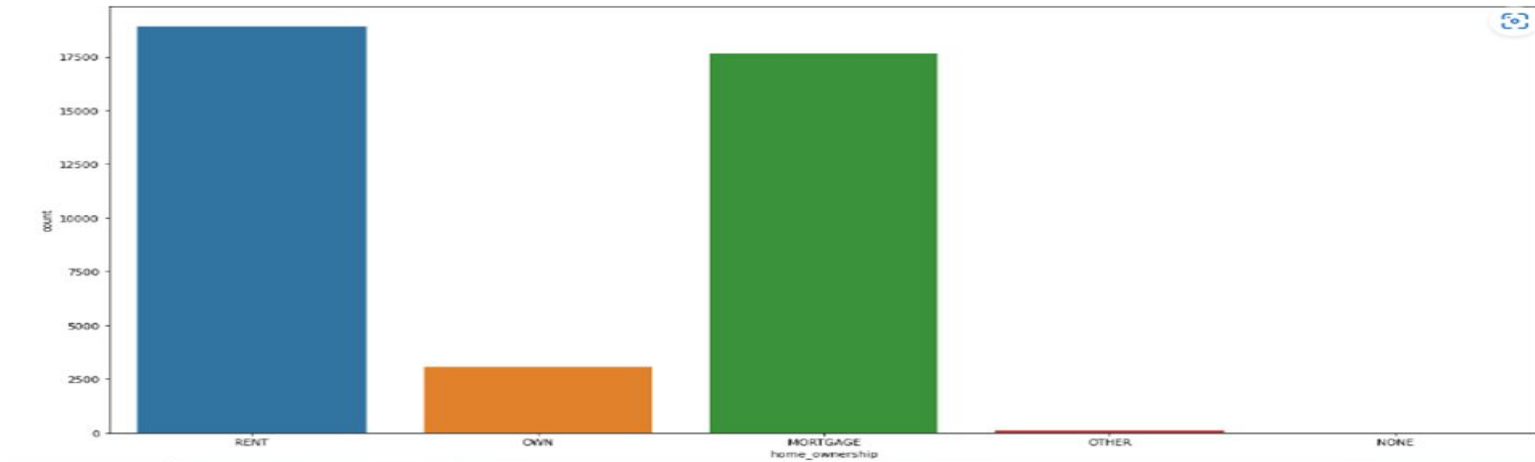
Loan Characteristics (loan amount, interest rate ,purpose of loan etc.) and
Customer Behavior Variables (those which are generated after the loan is approved such as delinquent 2 years, revolving balance,next payment date etc.)



Univariate

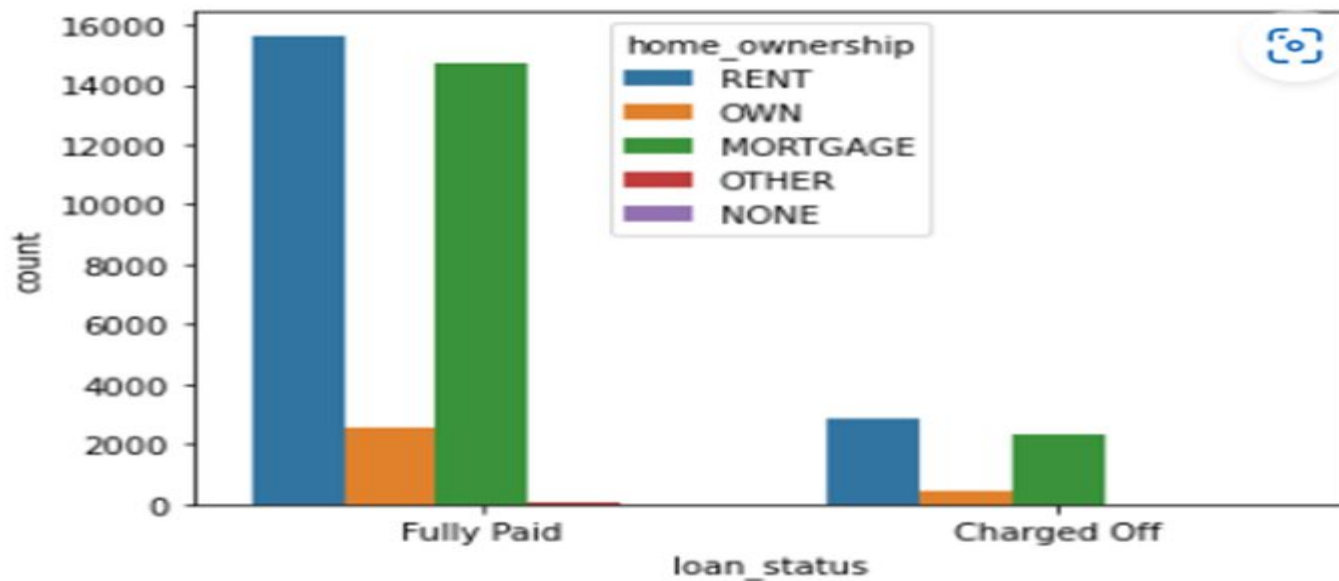
Loans taken based on Home Ownership

Approximately 92% percent of the total loans are taken by home ownership as Rented and Mortgage.



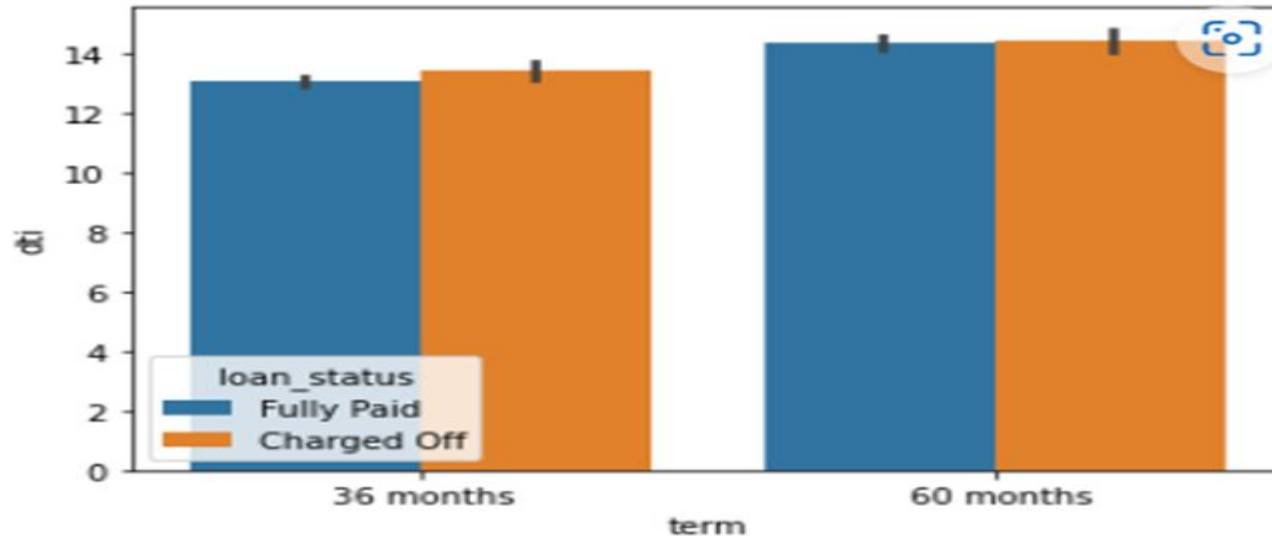
Bi-variate analysis

More than half defaulters for Rented and Mortgage applicants compared to homeowners.



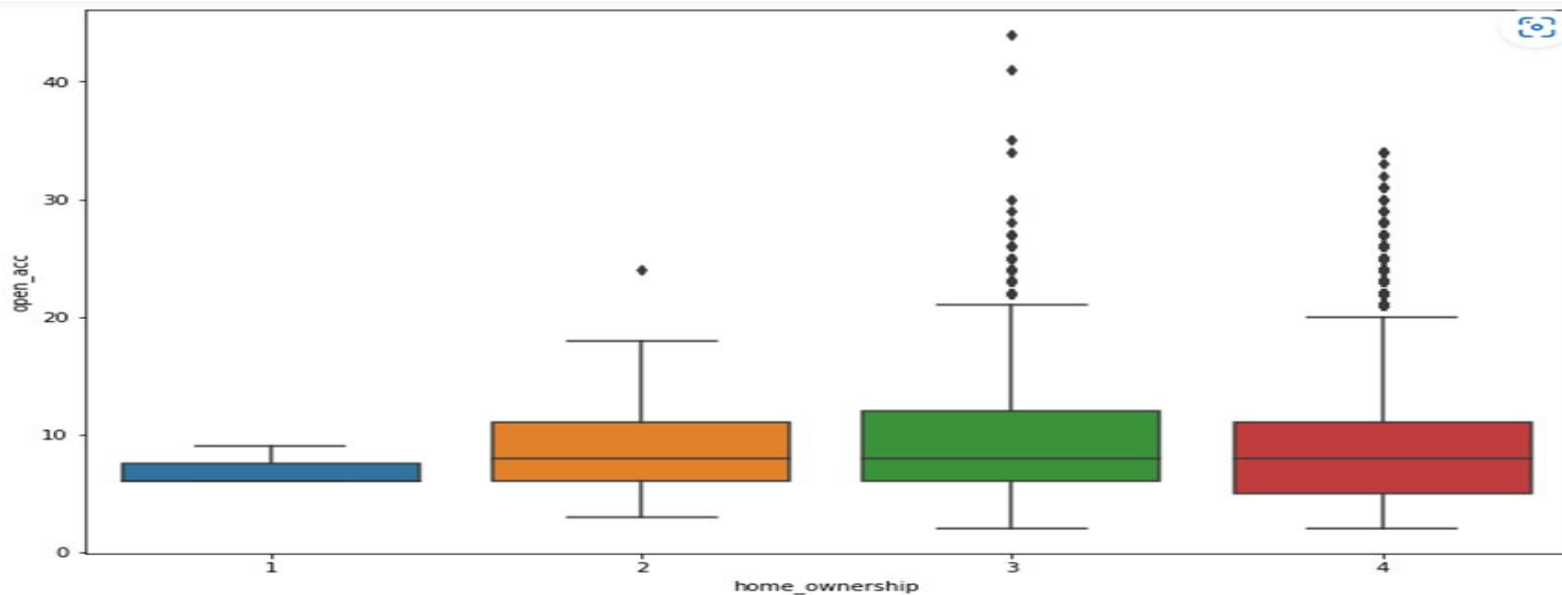
Debit to Income (DTI)

In Debit to Income(DTI) correlation with Loan term, charged off loans are slightly more compared to Fully Paid for 36 month term whereas in 60 months term both Fully paid and charged off are approximately equal.

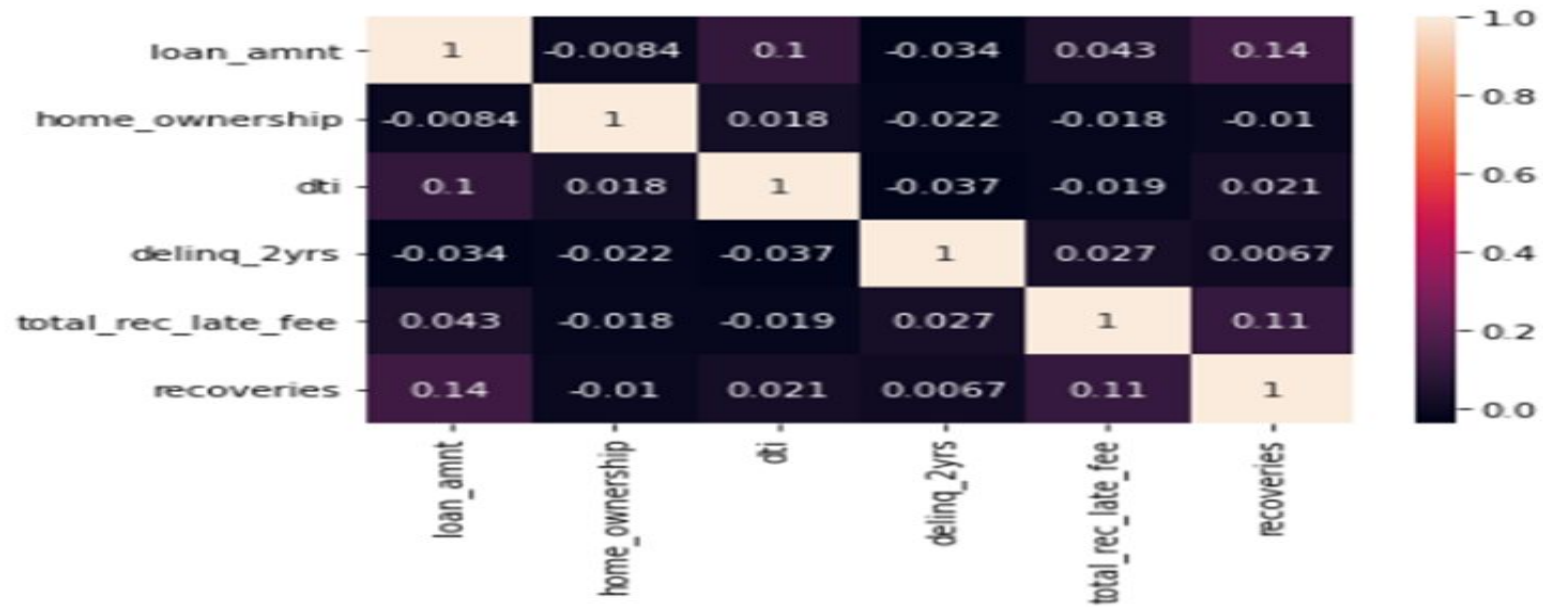


Credit Lines

There are around 7 open credit lines and it is same across home ownership, whereas Mortgage and others have maximum open credit lines of around 10.



Heat Map Correlation



Recommendations

- People staying in mortgage houses have higher defaulters compared to other means home ownership.
- Reduce Loan approval for Rent Category for home ownership as it has higher rate of charge off.
- Verification of employment proof of loan applicant.
- People staying in mortgage houses have taken less loan amount compared to people staying in rent and own houses
- Charged Off is more if the term plan is 36 months
- Annual income with 90k-120k having high interest rate have more charged off
- Emp length is more than 10 years charged off is more
- Defaulters are more if Interest Rate is between 10%-15%.
- Reduce Loan approval for Rent Category for home ownership as it has higher rate of charge off.
- Lending Club should consider accepting more loans from applicants whose annual income is > 100000 as their probability of charge off is minimal



THANK YOU

