Lending Club Case Study

By

Ayush Gupta
Saibi Singhai

Lending Club Case Study

COMPANY INTRODUCTION

LendingClub was an American peer-to-peer lending company, headquartered in San Francisco, California. It was the first peer-to-peer lender to register its offerings as securities with the Securities and Exchange Commission, and to offer loan trading on a secondary market

It is the largest online marketplace, facilitating personal loans, business loans and financing of medical procedures. It enables peer to peer lending of money between lenders and borrowers.

Borrowers are the individuals who come to Lending Club to avail loans at lower interest rate.
Investors are the individuals who lend on the Lending Club get solid returns based on the interest rate they choose to lend.

Lending loans to risky applicants is the largest source of financial or credit loss. Credit loss is the amount of money lost by lender when the borrower refuses to pay or runs away with the money owed. The borrowers in this case are labelled as Defaulters.

BUSINESS OBJECTIVE

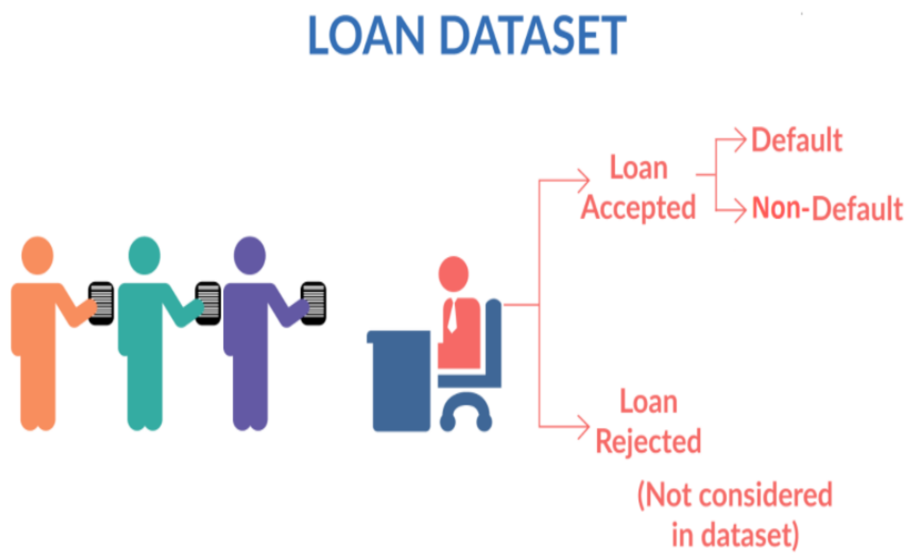
To figure out the driving factors (or driver variables) behind loan default, i.e. variables which are strong indicators of default and potentially use the insights in approving or rejecting the loans to avoid credit loss.

Lending Club Case Study

BUSINESS UNDERSTANDING

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 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



When a person applies for a loan, there are two types of decisions that could be taken by the company:

1. Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:
   1. Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
   2. Current: Applicant is in the process of paying the installments, i.e. the tenure of the loan is not yet completed.
   3. Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

2. Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.)

Lending Club Case Study

DATA UNDERSTANDING

This step includes sourcing or importing our dataset, understanding the semantics of variables and determining Target Variable

DATA CLEANING

Drop columns with null or monotonous values.
Drop or impute values for invalid or missing values in rows.
Perform datatype conversion.
Perform Sanity Check.

UNIVARIATE ANALYSIS

Plotting and observing numerical and categorical data with graphs.
Understanding the value distribution and other stats of variable

BIVARIATE ANALYSIS

Plotting and analyzing relation of multiple variables, among each other and with with target variables.

INFERENCE

Finding out significant variables. Bring out the insights.

Lending Club Case Study

Scrutiny of Data

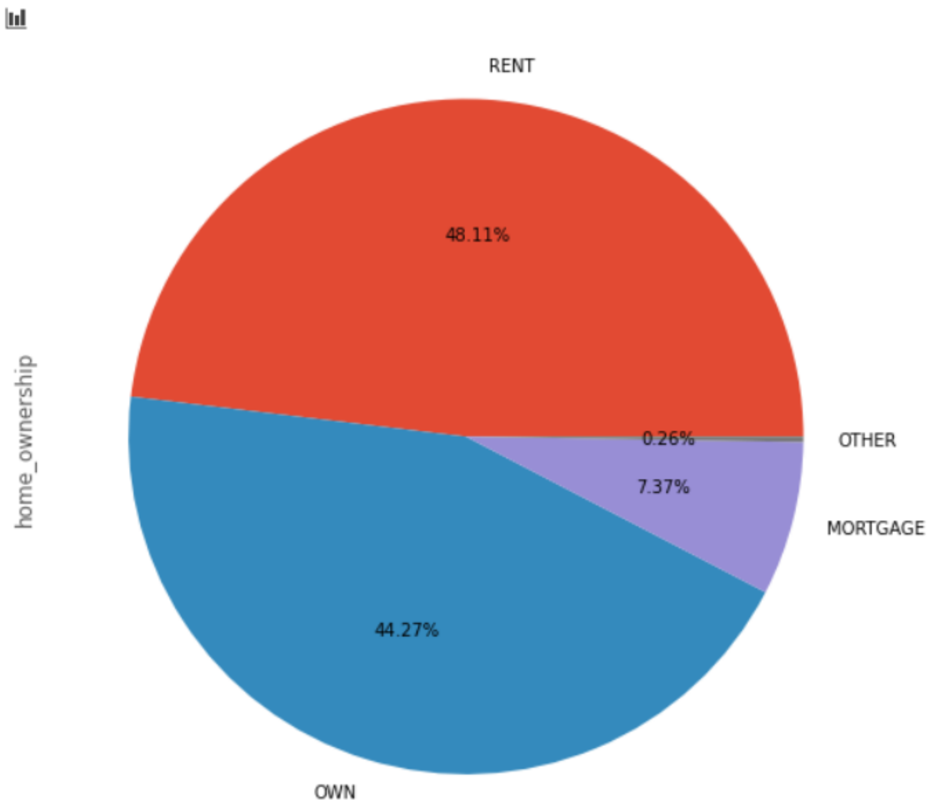
Data Understanding:

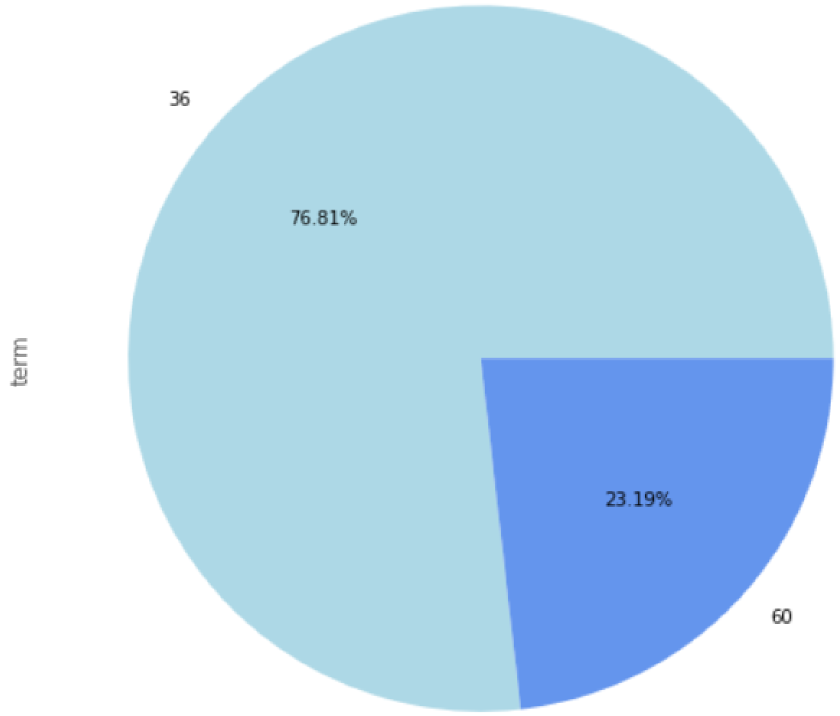
1. The dataset is present in the loan.csv file.
2. Initially there are 111 columns and 39717 rows in the dataset.
3. There are 54 columns in the dataset having completely null values.
4. There are 9 columns with monotonous values
5. Loan Status is the target variable.
6. Out of 39717 records, 5627 were with charged off status, 32950 were fully paid status and rest were current status.

Data Cleaning:

1. 55 columns were remaining in the dataset after removing all the columns with atleast 80% null values
2. Rows for the columns with atmost 10% of missing values were removed.
3. Rows with missing values for public record bankruptcies were imputed.
4. After cleanup, total of 31 columns and 38465 rows were present in the dataset
5. After datatype conversion, the size of the data set reduced to 7.3mb from 33.6mb.
6. After data cleaning, there are 23 numeric and 10 categorical variables present in our dataset

Lending Club Case Study





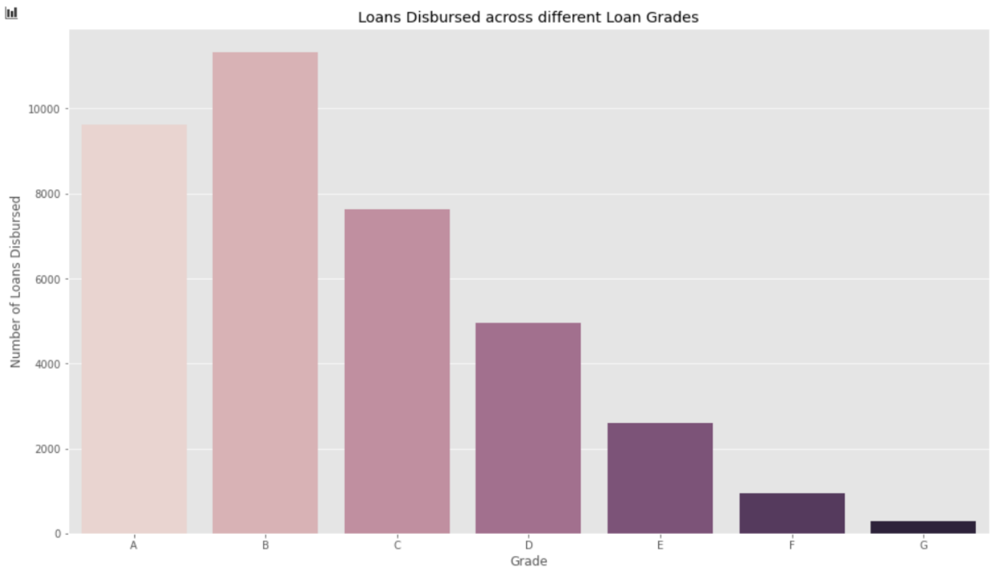
Analysis of home ownership of applicants

1. 4 kinds of ownership
2. Most of the people live in a rented house, seconded by people who are having their own house.

Analysis of tenure for which loan is preferred

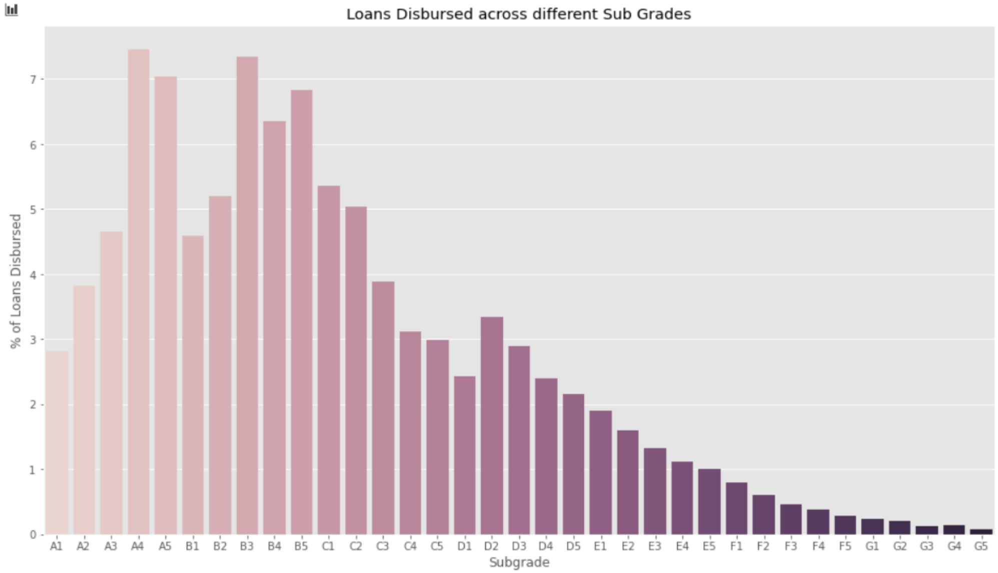
From the chart, it is evident that LC favors giving loans to borrowers for lesser term i.e. 36 months.

Lending Club Case Study



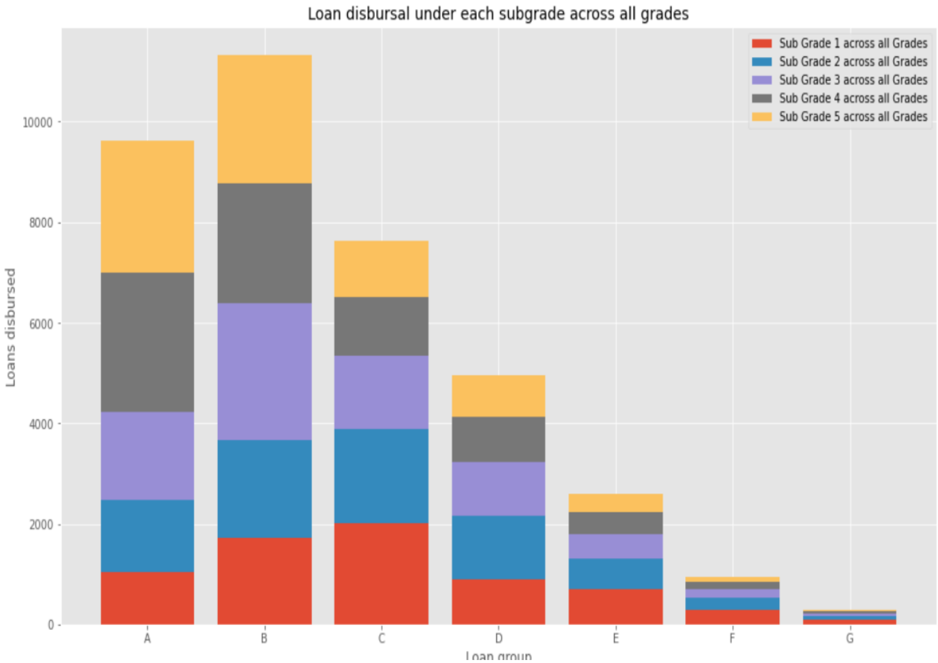
Analysis of loans disbursed for different grades

From Grade A to G, risk increases. Hence, people with lower risk are provided loans. Grade B falls into the sweet spot of both and hence it is the category within which most borrowers are lent loans.



Analysis of relationship between grades and subgrades

Number of Loans dispersed in each Subgrade across all Grades accumulate to the number of loans disbursed in each Grade analyzed independently. Hence, they are highly correlated variables

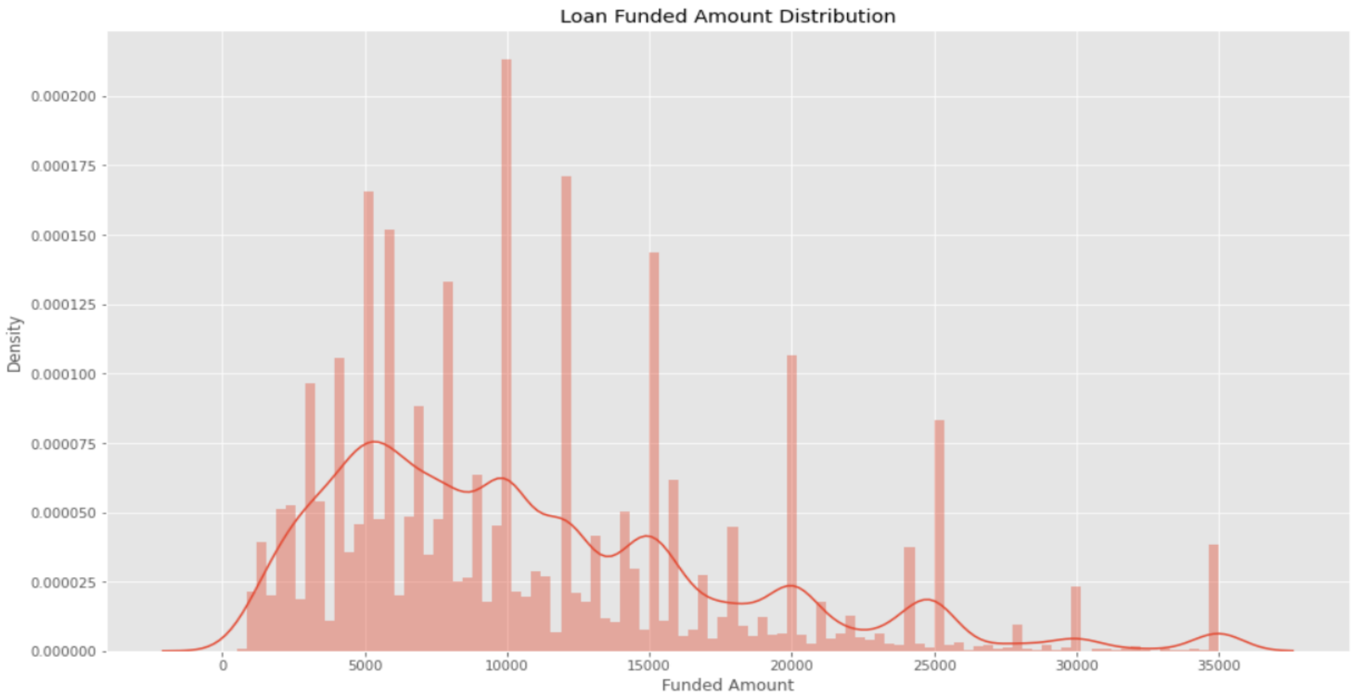


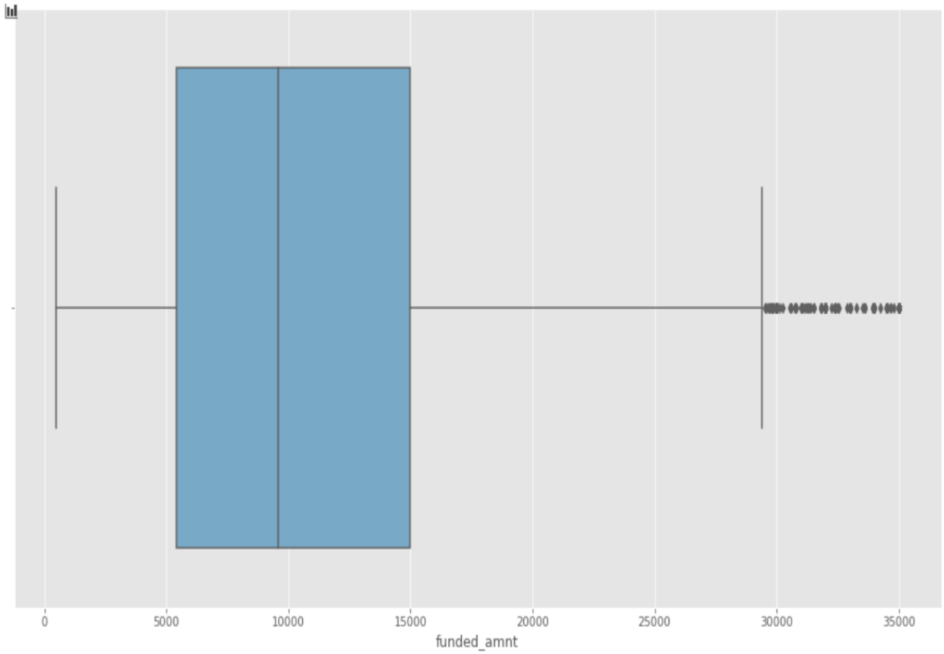
Analysis of loans disbursed for sub grades

Loan Subgrade, as the title suggests, follows pattern the same as loan grade.

Lending Club Case Study

Analysis of funded amount

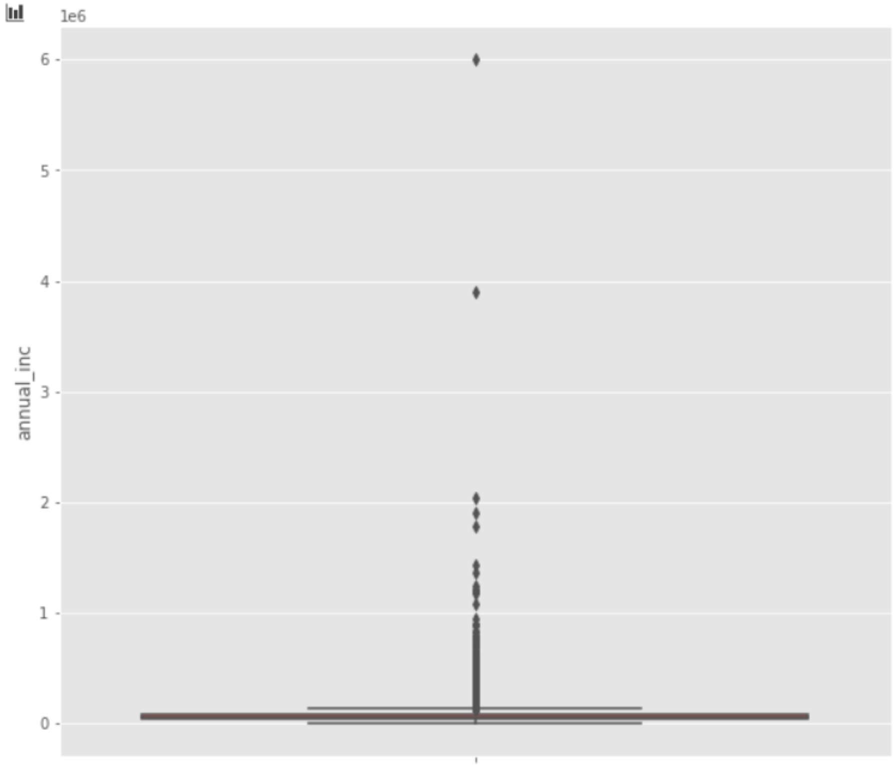


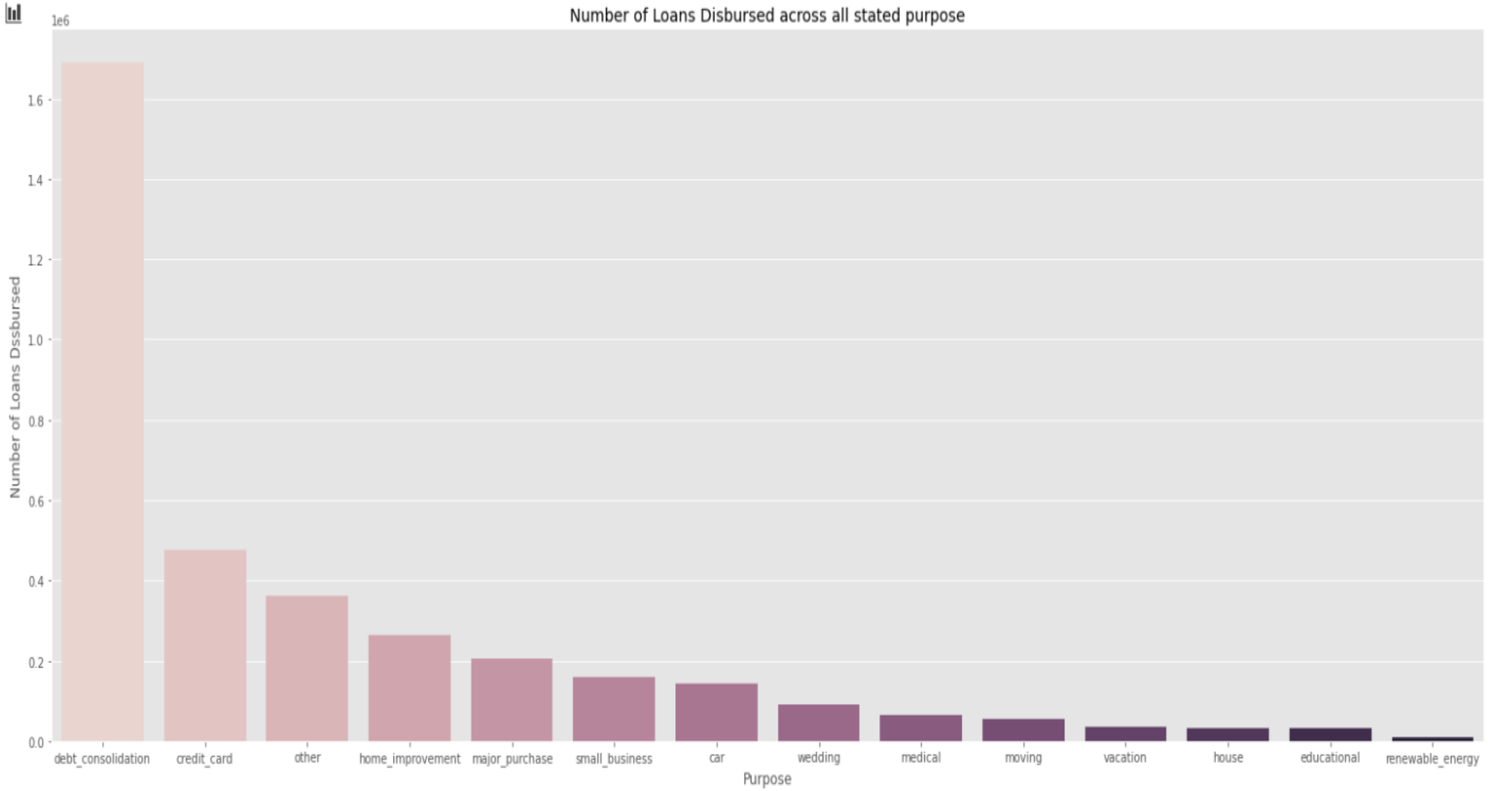


1. Loans for Lower Loan Amount are funded more than with higher amount.
2. Borrowers and Investors prefer to take/grant loans for an amount which is rounded off to the nearest 5K/10K multiple.

It is inferred from the above box plot, that there is a presence of outliers with loan amount.

Lending Club Case Study





Analysis of annual income of applicant

It is inferred from the above box plot, that there is a presence of outliers with loan amount.

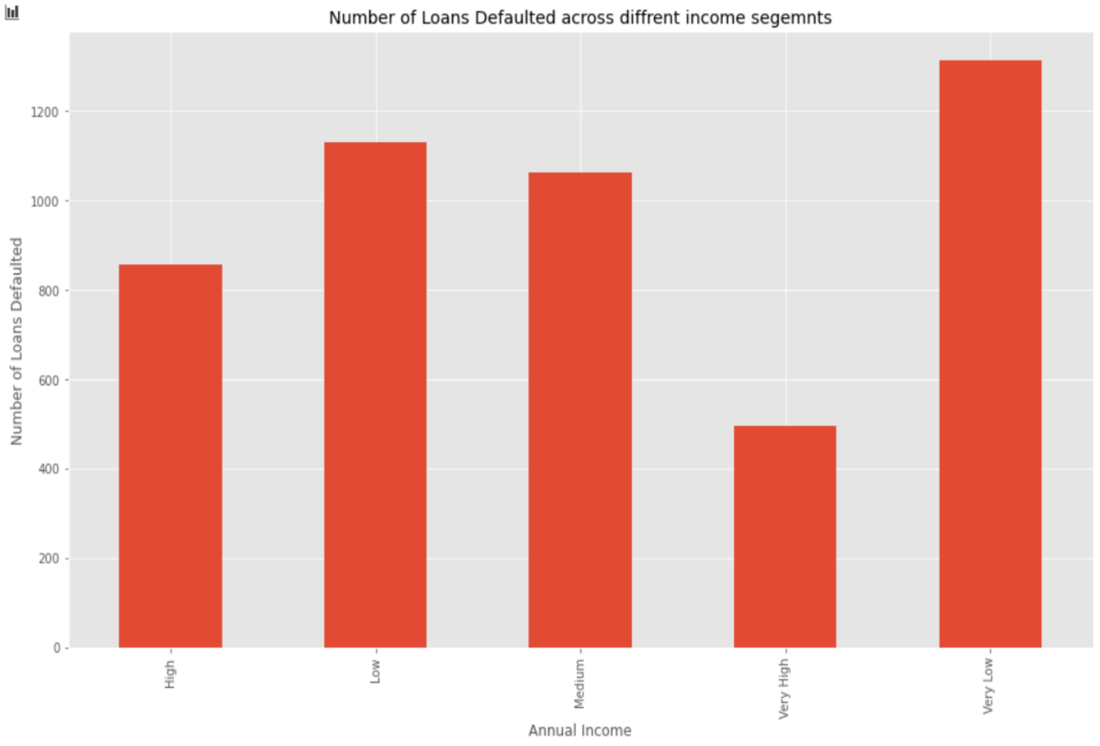
Analysis of purpose of loan

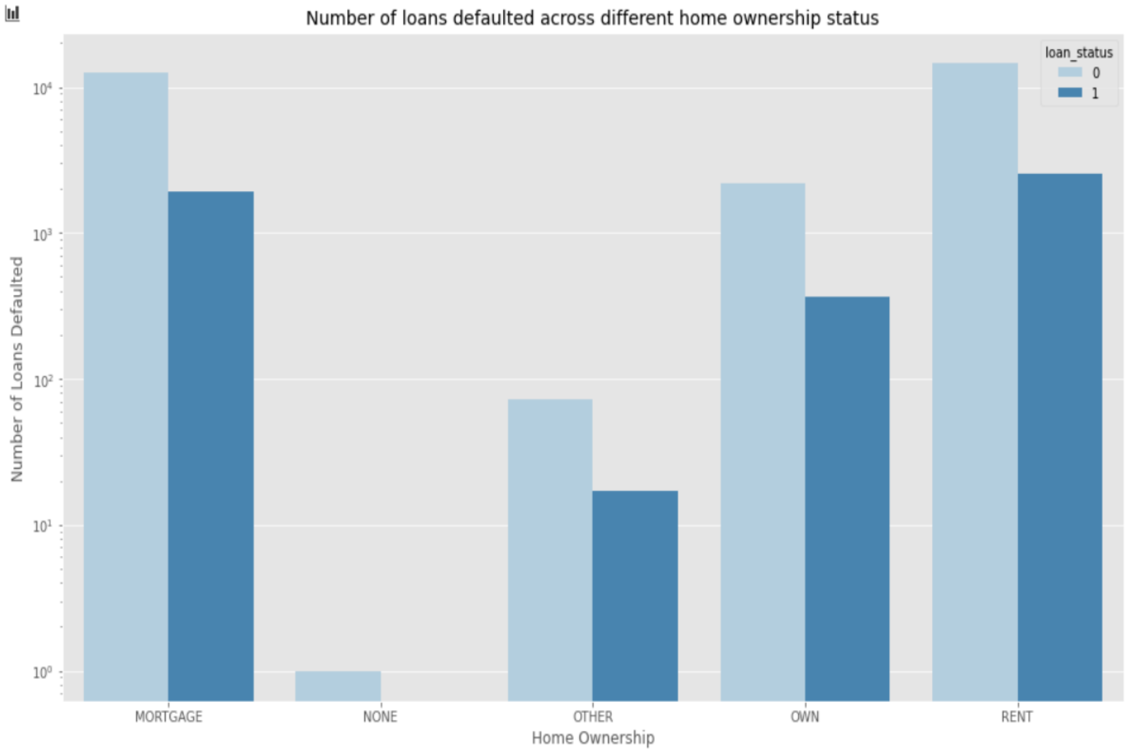
It is inferred from the above plot, mostly the loans have been taken for debt_consolidation.

Lending Club Case Study

Analysis for annual income of defaulted cases

Annual income is having an impact over loan repayment capacity of Borrowers. Borrowers with higher income are 'safe' for loan disbursal.

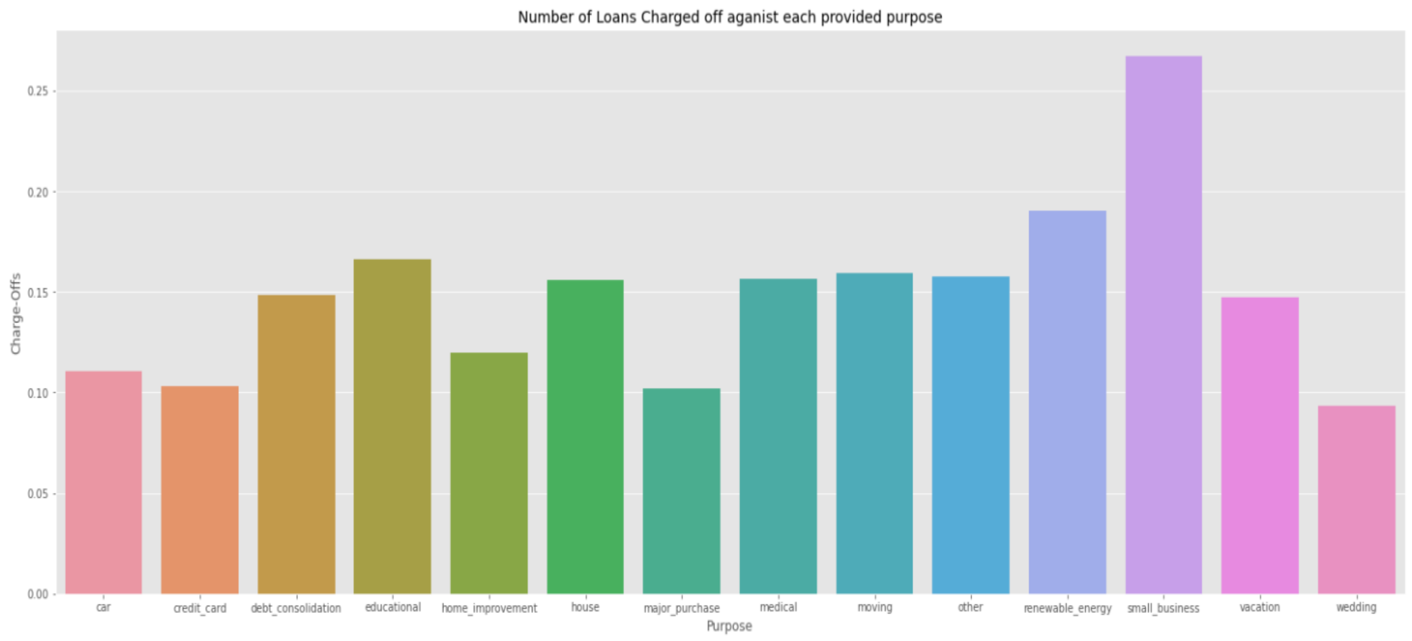


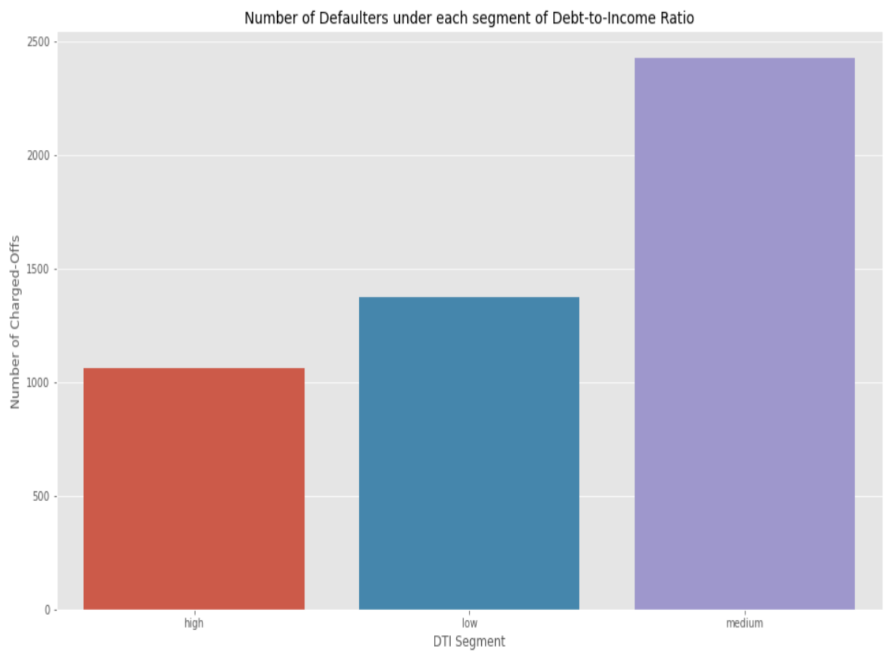


Analysis for home ownership of default and fully paid cases

1. People with rented/mortgaged homes are more likely to default.
2. Borrowers with their own homes have relatively much lower number of defaults

Lending Club Case Study





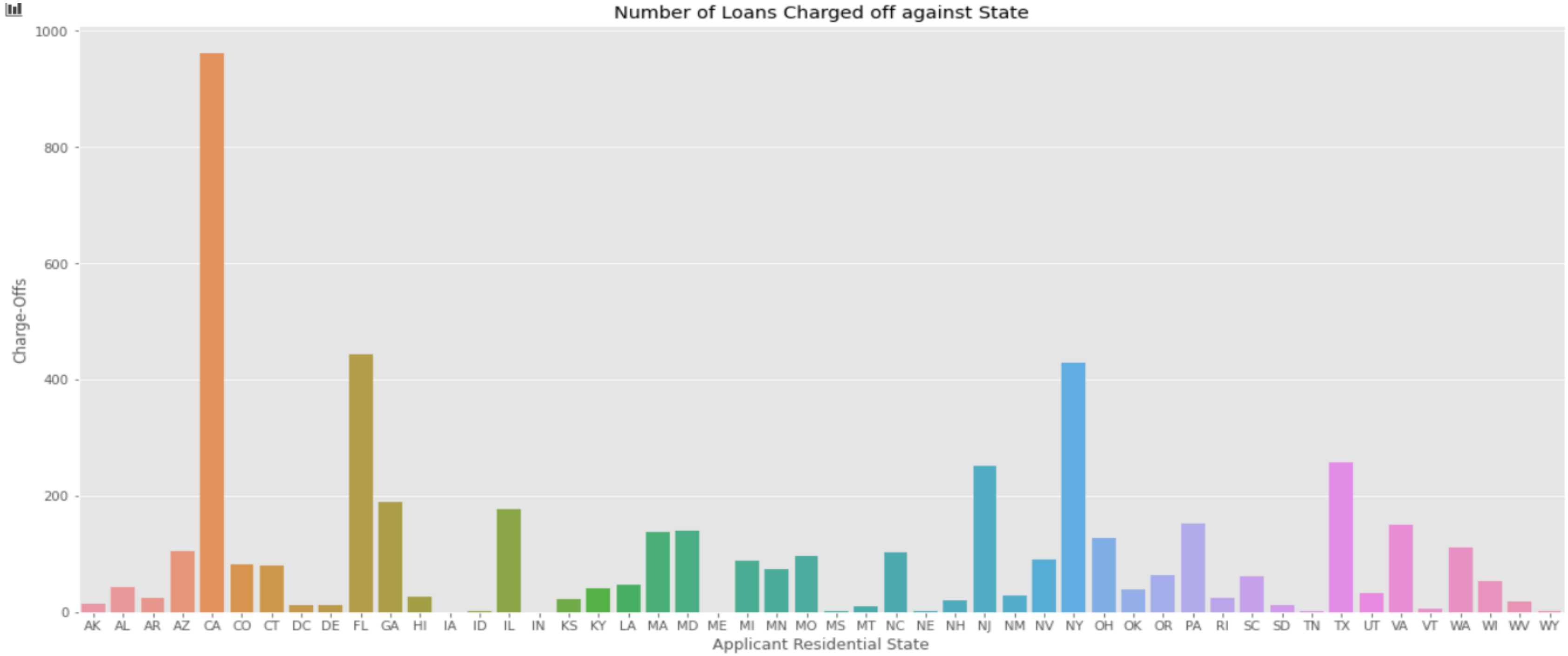
Analysis for loan purpose of charged off applicants

1. Small Businesses has highest risk associated with them followed by Renewal Energy sector and then Education.
2. Loans taken for luxuries like Wedding, Home Improvement or car have lower default rate than taken for necessity like medical, education and small businesses.

Analysis for debt to income ration of defaulted cases

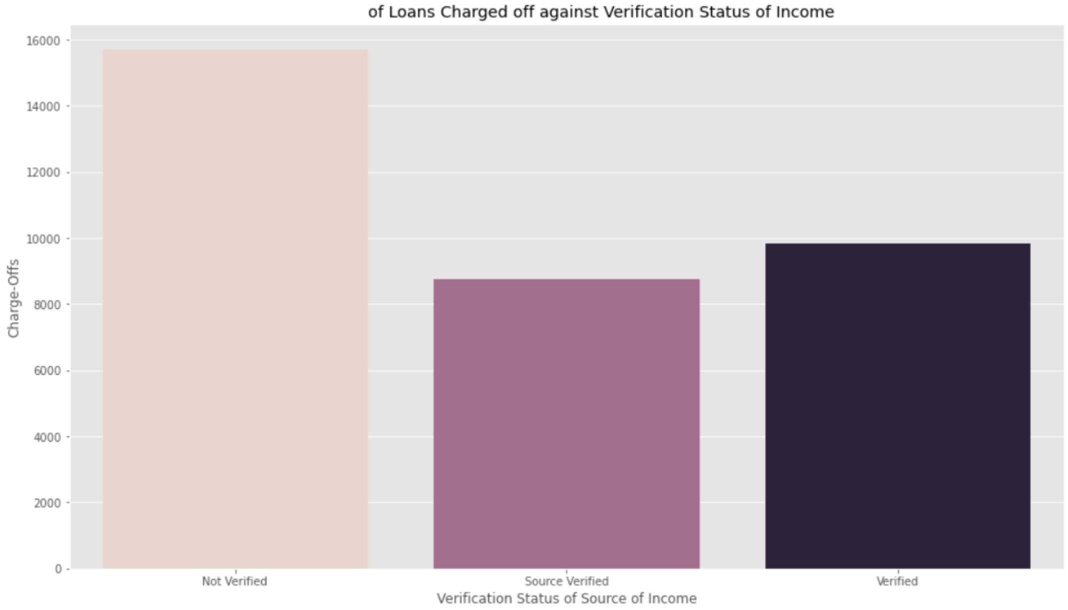
Applicants with high DTI are safe to lend out loans to.

Lending Club Case Study



Analysis for state of charged off applicants

CA state is the state with highest number of loan default cases

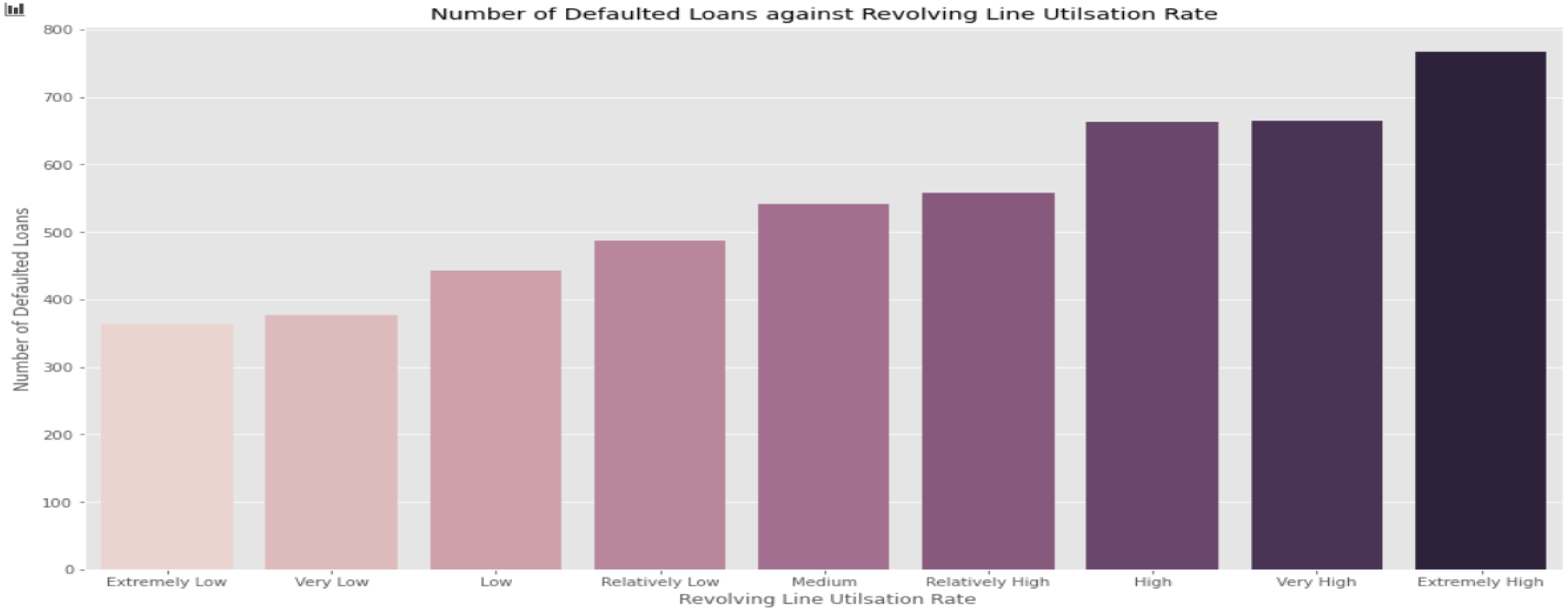


Analysis for verification status of defaulted cases

Verification is an important part of loan disbursal process to mitigate the number of default cases.

Lending Club Case Study

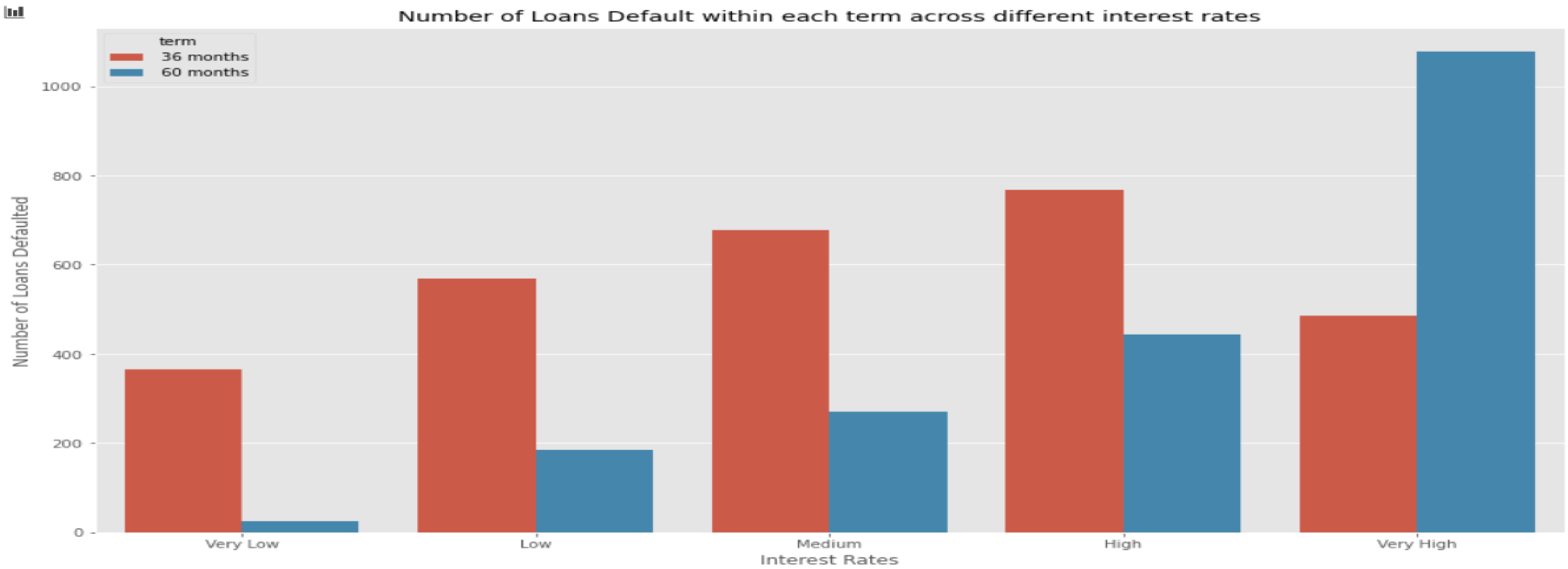
Analysis for revolving utilization rate of charged off applicants



With increase in revolving line utilization, the risk of loan being defaulted increases.

Lending Club Case Study

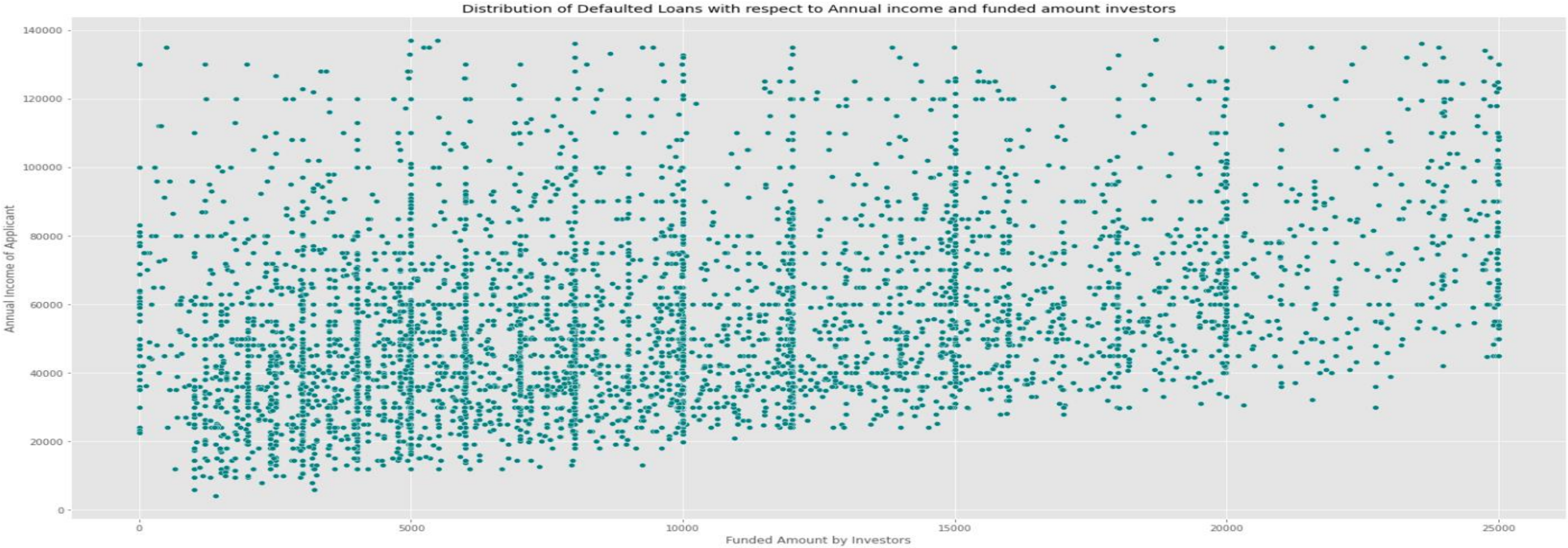
Analysis for default cases with term and interest rates



1. Generally speaking, with increase in interest rate, the probability of loans being defaulted increases.
2. For the same interest rate range, 36 months term has a higher default rate than 60 months.
3. Exceptional case for very high interest rate is present. LC should look into this, what could be the reason behind this.

Lending Club Case Study

Analysis for default cases with annual income of applicant and amount funded by investors



Applicants with lower side of income apply *and are approved by LC* for lower fund amount (<10000). These people are likely to default. However, on the contrary, as seen previously, more and more number of loans are taken by the borrower whose annual income is at the lower end of spectrum. And most of these loans DO get fully paid. Hence, it cannot be said concretely that people with lower end of annual income would likely default. But it can be said that the people at higher end of income are very safe to invest in and such loans would not convert to Charge-Offs.

Lending Club Case Study

CONCLUSION

Many variables in the given dataset can impact the defaulting of a loan and hence are really important driving factor for business. Some of the variables as per inference are listed below:

1. Borrowers with **Verification Status** as not verified are the most risky. Hence verification is very necessary.

2. Loans with higher **interest rates** have higher chances of default.

3. **Purpose** is stated as Small Businesses are the most risky borrowers due to uncertainty in business flourishing, followed by Renewable energy sector as it is budding.

4. Applicants with high **Debt to Income Ratio** should be preferred.

5. **Annual Income** of the borrower is one of the most important factor. People with low annual income are having higher chances of defaulting and should be avoided or charged higher interest rates.

6. Applicants with high **utilization of revolving line** are having more risk of defaulting.

7. Residential State is a factor which affects the default rate, but nothing can be said conclusively about states being risky. Each state has a different default rate, however, the most risky states are CA,NY and FL.

NOTE: The highlighted variables above are the driving factors.

Lending Club Case Study

REFERENCES

1. Lending Club: https://help.lendingclub.com/hc/en-us
2. Data set from Upgrad: https://learn.upgrad.com/v/course/1120/session/110398/segment/613656
3. Data dictionary from Upgrad: https://learn.upgrad.com/v/course/1120/session/110398/segment/613656
4. LC Grading System: https://www.lendingclub.com/foliofn/rateDetail.action
5. What are the terms of loans?- LendingClub : https://help.lendingclub.com/hc/en-us/articles/215466718-What-are-the-terms-of-the-loans-#:~:text=Loans%20have%20either%203%2Dyear,include%20both%20interest%20and%20principal

Thank You