LENDING CLUB

CASE STUDY

Name: Suhani Jaiswal,Sumanjit Moshat

Lending Club – Risk Management

Lending Club , a consumer finance company, is looking to analyze credit risk.

The overall goal is to find out the consumer and loan attributes which are the primary influencer for a loan default.

Business Understanding of Loan Status:

• Loan Status is the attribute which implies the status of the loan

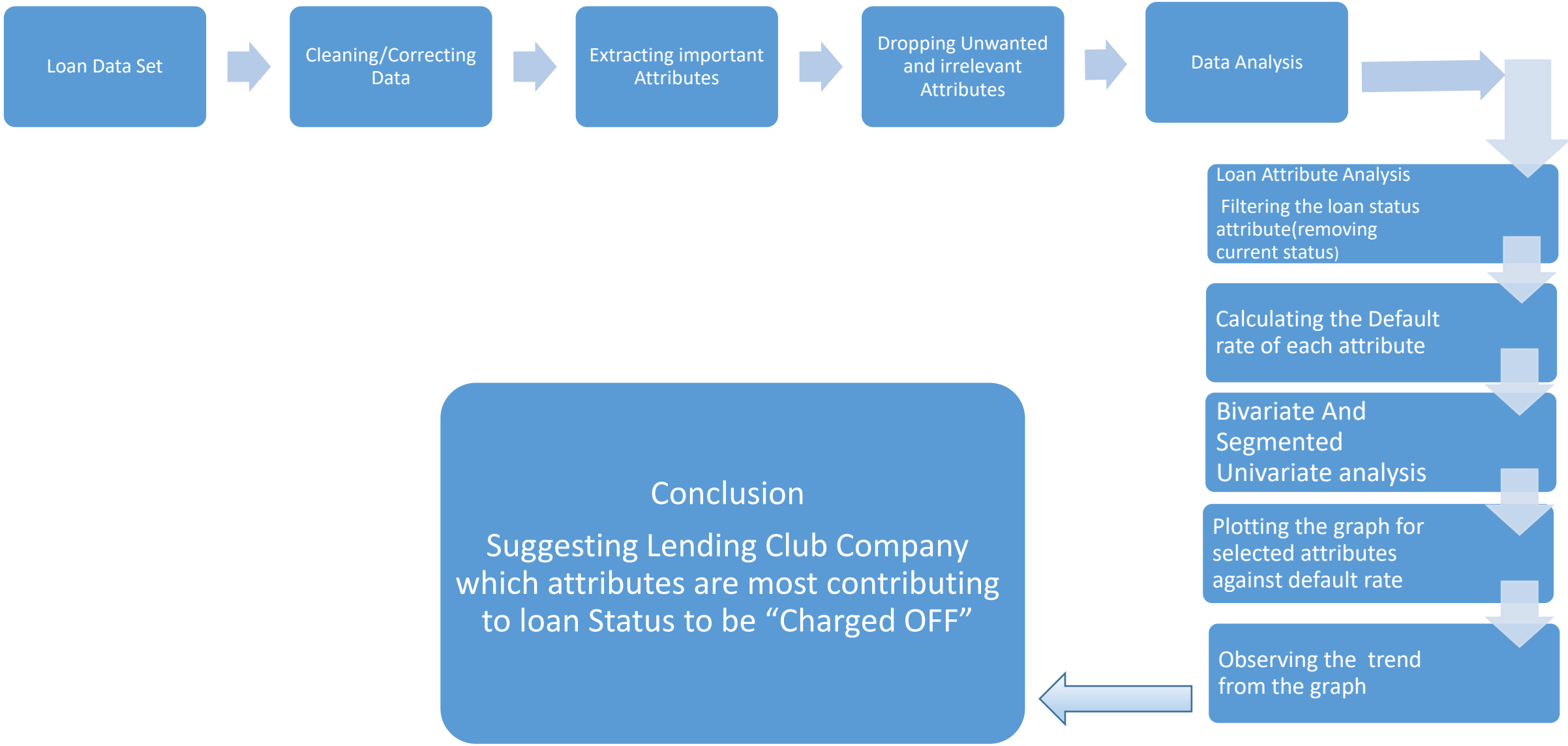
1. Fully Paid
2. Current
3. Charged Off

Business objective: To identify the loans which are likely to default .

Goals of data analysis:

• Finding out the attributes which influences a loan to be defaulted .

Problem Solving Methodology



Data For Analysis

1. Loan Data, A file with data of loans issued.
2. Data Dictionary, A file with the meaning of all the columns.

Assumptions:

- All the information provided within the 2 files is a trusted source for analysis.

Data cleaning performed:

- Removing all columns which one unique value throughout
- Removing rows with loan status current.
- There are many variables in the data set which is not relevant for our analysis, also the data contains the loan status and information about the people who have taken loan so we are also dropping those which are generated after the loan is approved such revolving balance, next payment date etc.
- Removing columns having null values which are not required for the current analysis.
- Removing rows having null value in the required analysis fields.
- Removing outliers.

Loan Characteristics Analysis

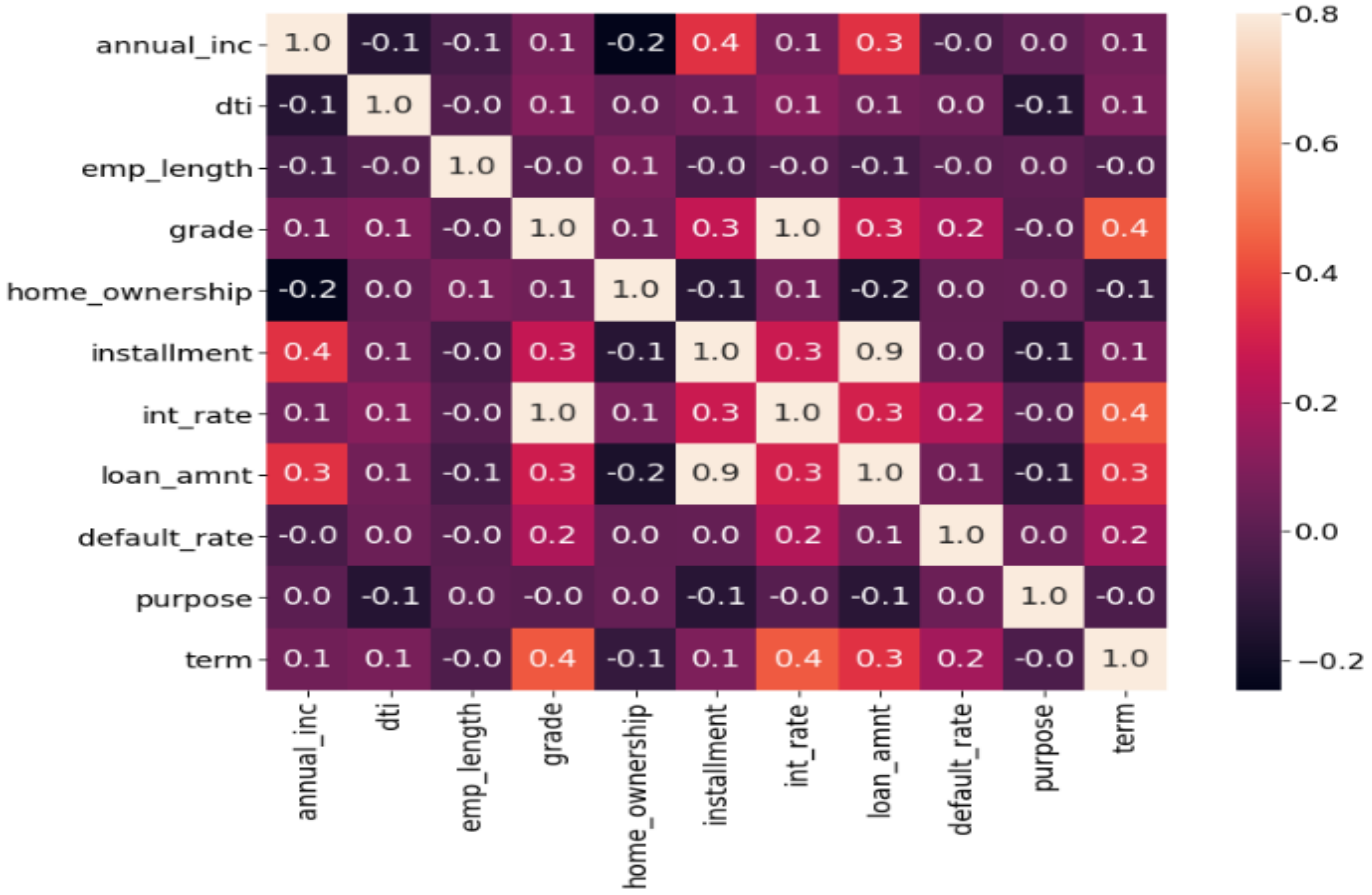
Default Rate – Number of Defaults / Total Loans

After all sorts of cleaning we got 18 attributes which has the potential to influence the default rate. After analyzing each of them with univariate , bivariate and segmented univariate we produced 6 attributes which have significant influence over the default rate . Please find them below -

1) Interest Rate
2) Terms in Months
3) Loan Amount
4) Grade
5) DTI
6) Verification Status

In the next few slides we will describe how these attributes vary with default rate and how are they distributed in the whole dataset.

Correlation Map among important attributes



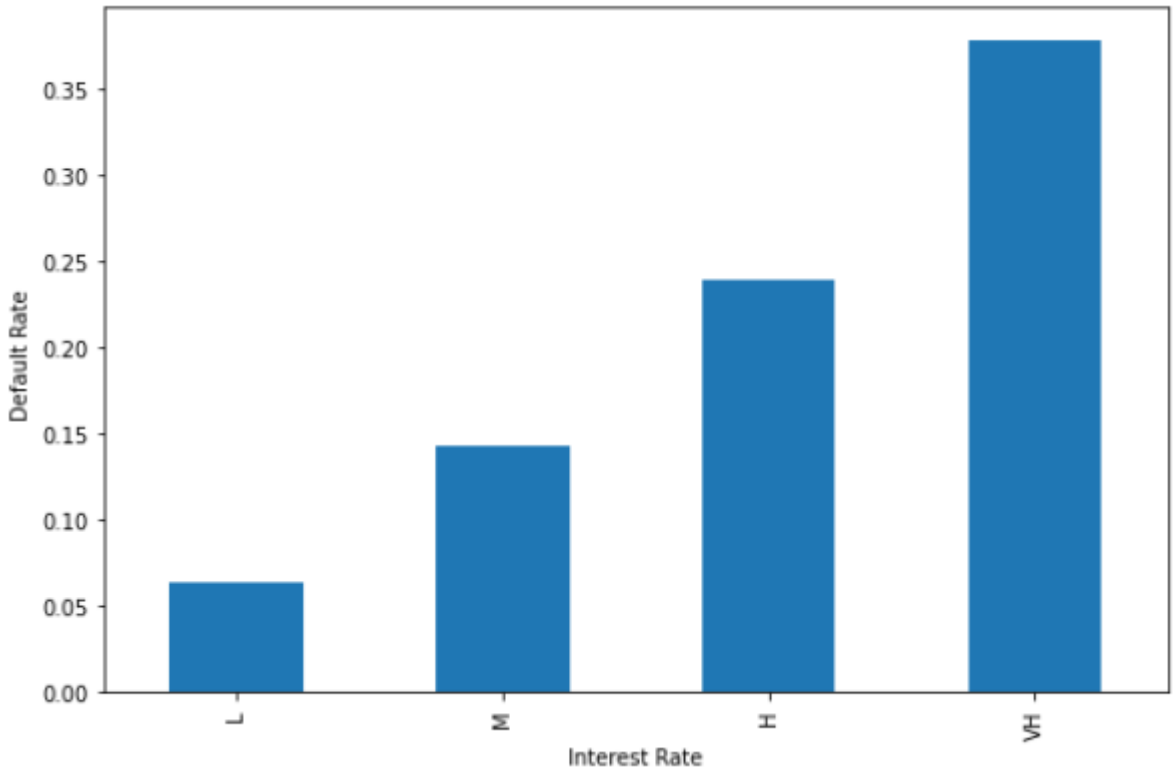
Annual income is highly correlated with loan amount, which can be concluded as for higher annual income the loan amount demanded is also high.

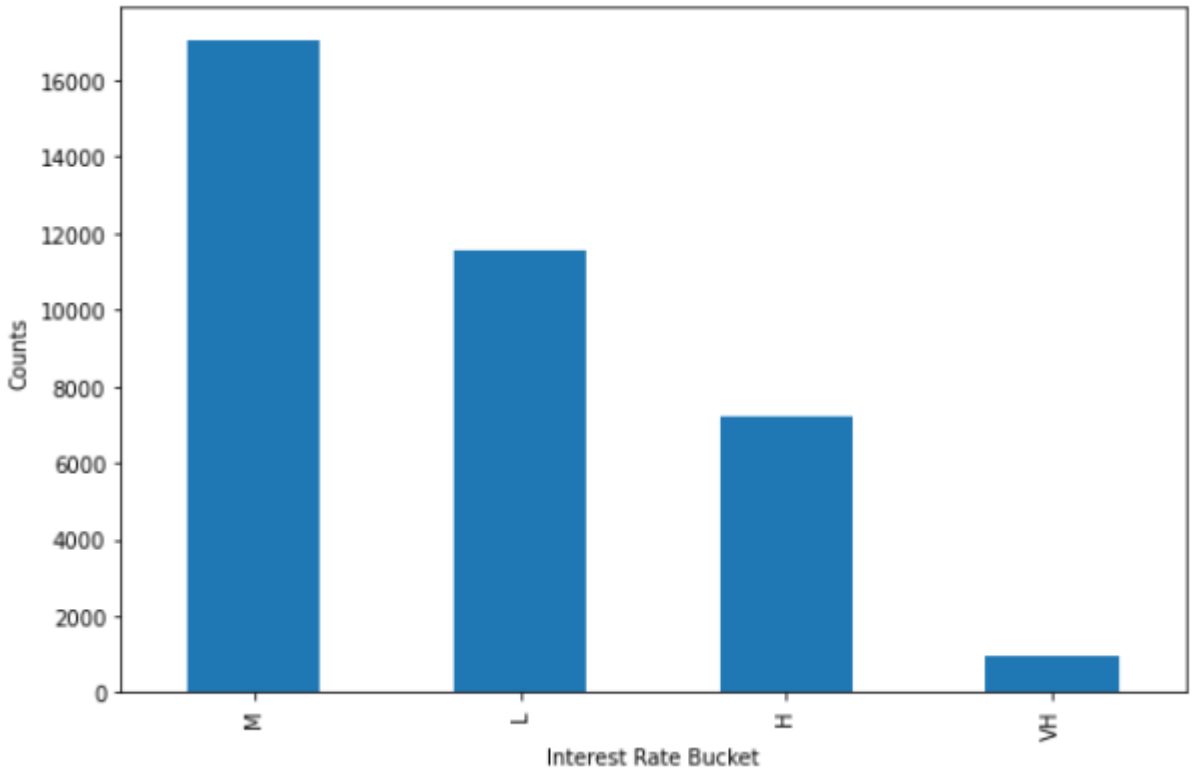
Grade and Interest rate is highly correlated with term so when term is long the interest rate is more and grade is low.

Default rate is highly correlated with grade , interest rate and terms

Univariate, Bivariate, Segmented Univariate Analysis

Interest Rate

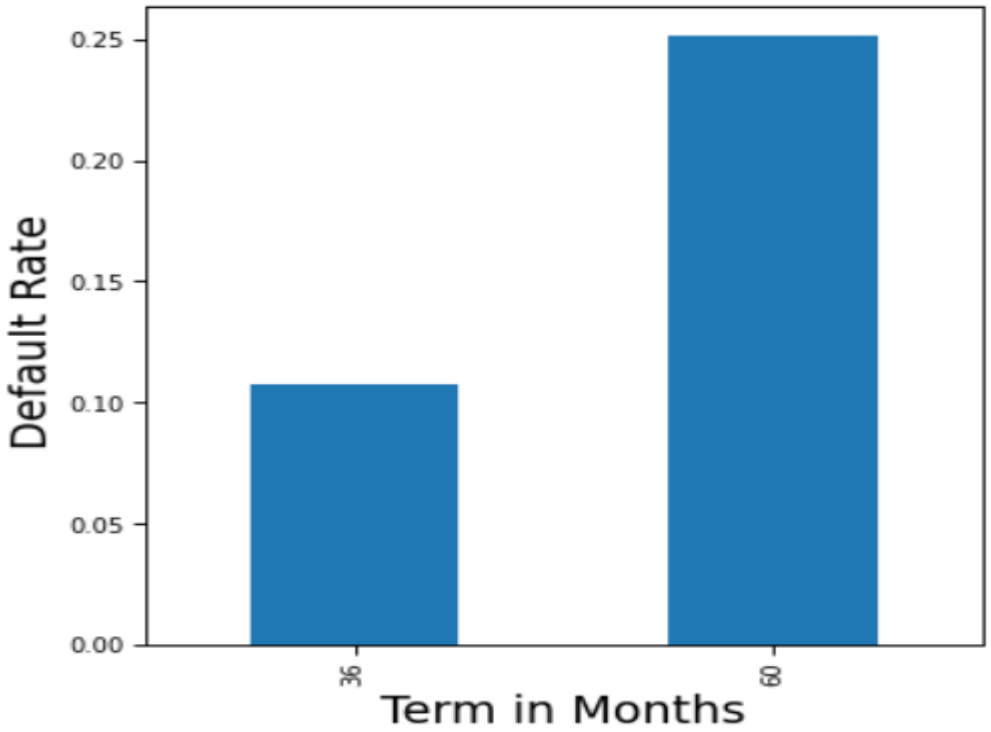


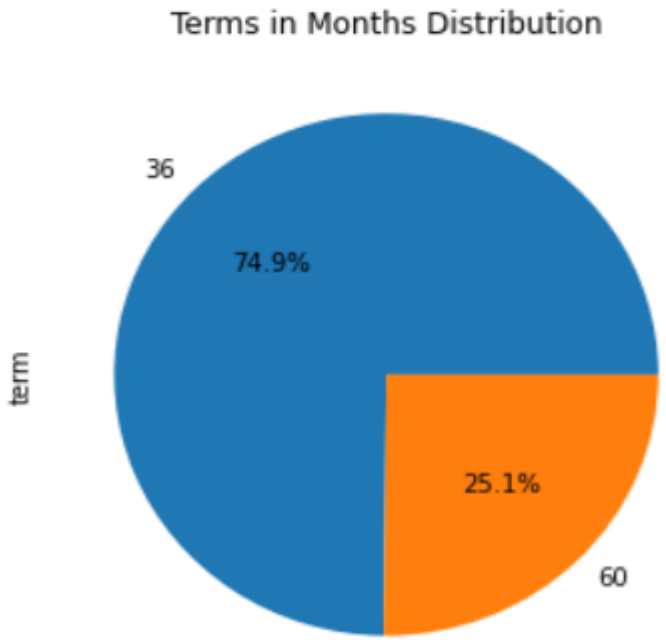


Bivariate - Interest Rate is inversely proportional with Default Rate

Univariate - Interest rate is moderately well distributed

Term

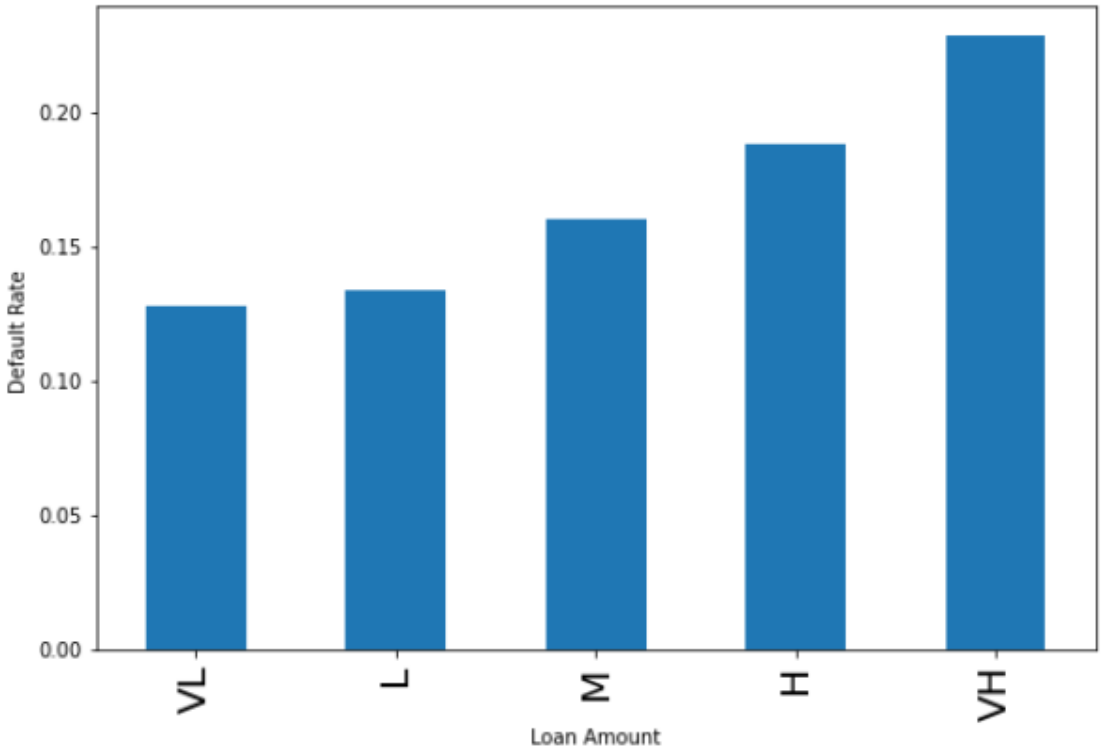


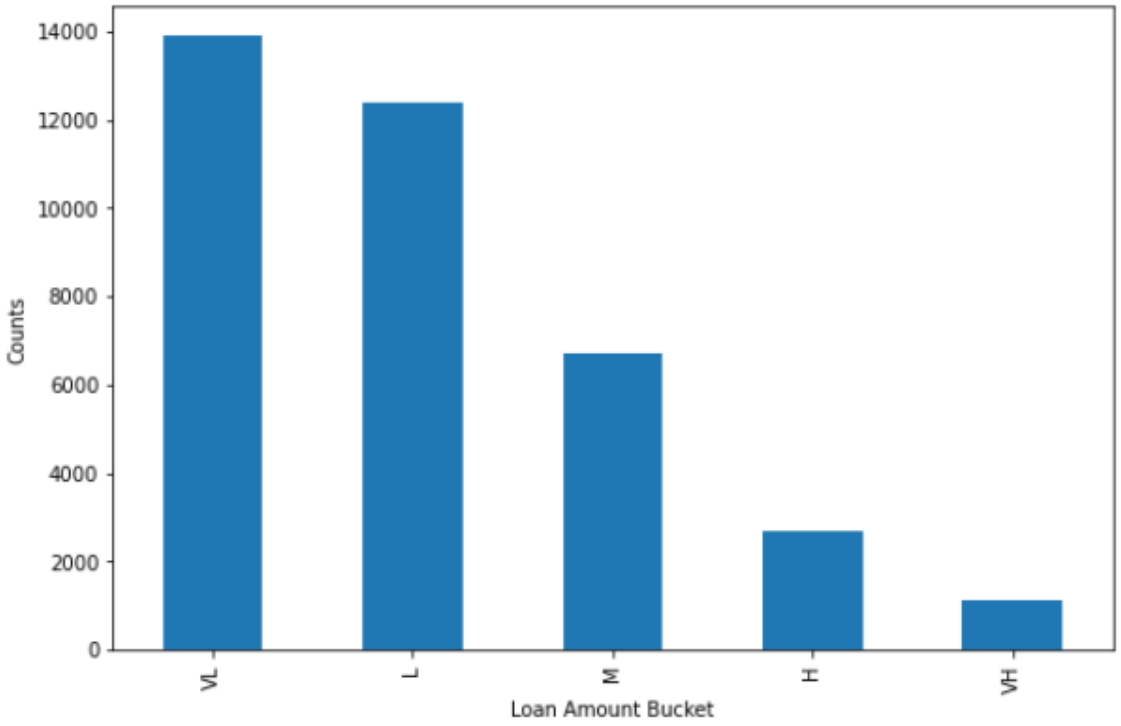


Loans with 60 months term are highly likely to default whereas loans with 36 months term are less likely to default.

Terms with 36 and 60 is distributed approximately in 3:1 ratio.

Loan Amount

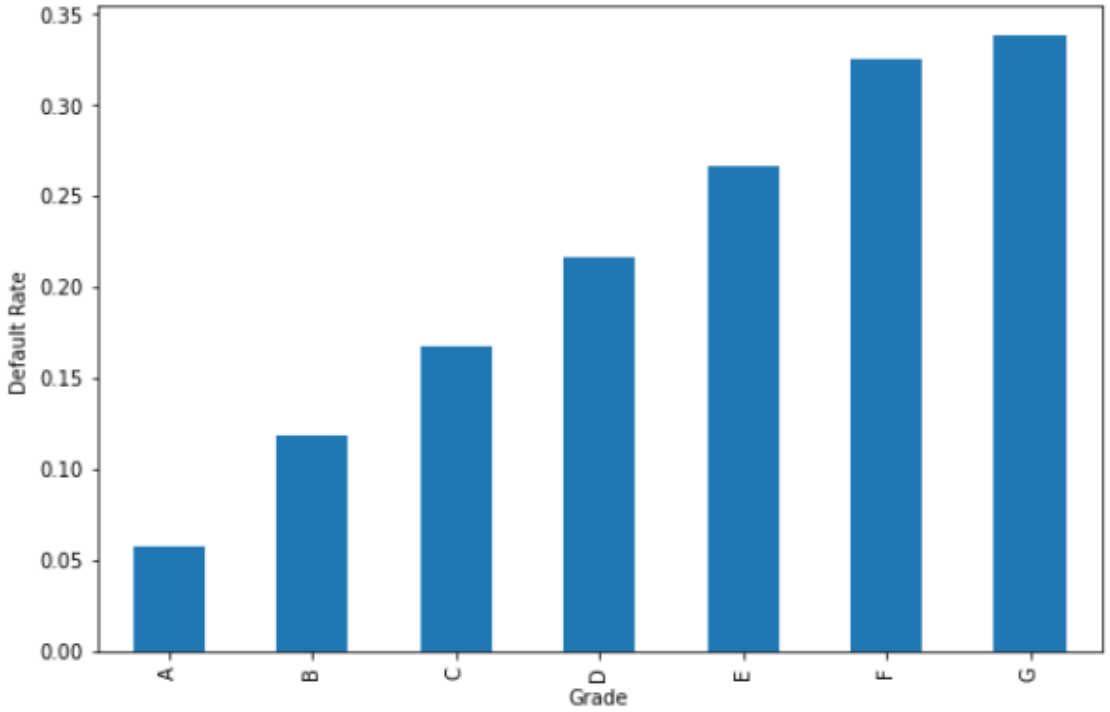


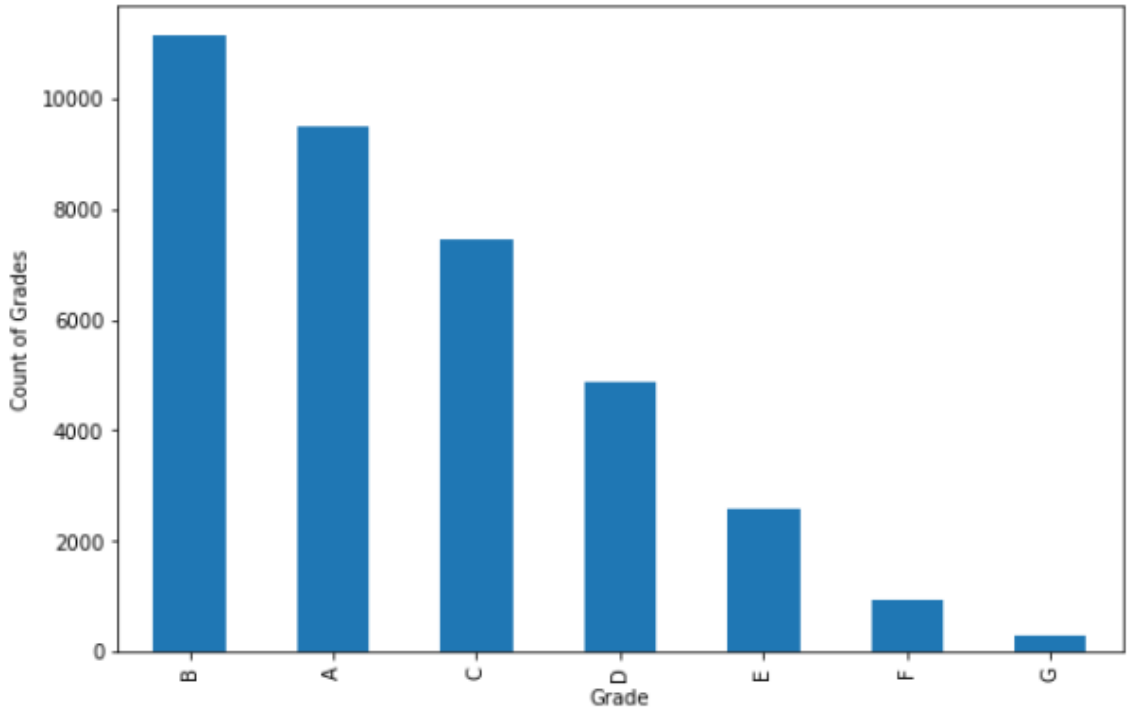


Loan Amount is directly proportional to Default Rate

Loan Amount is moderately well distributed .

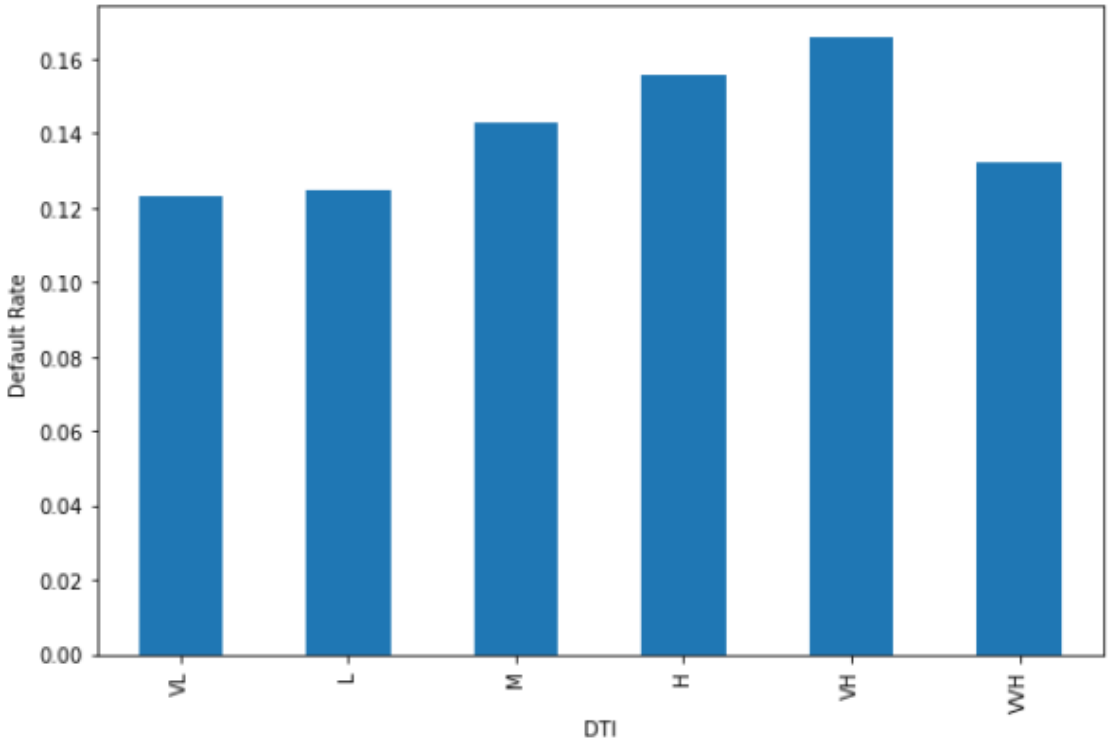
Grade

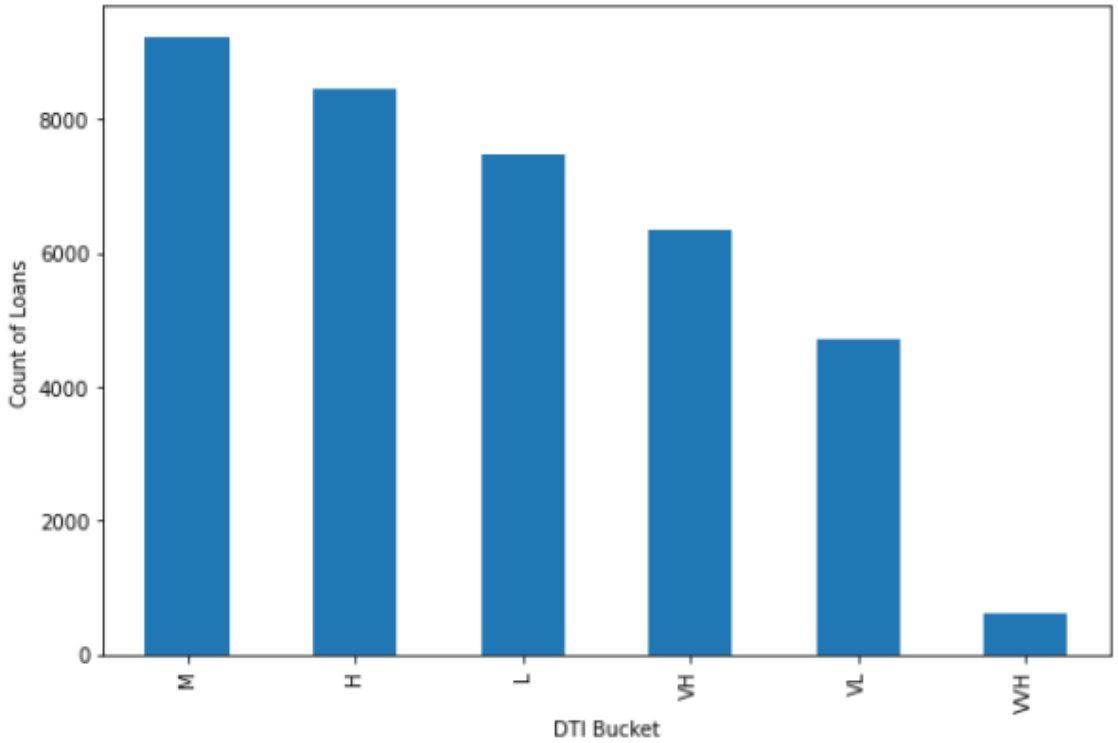




From A to G as the grade decreases the default rate also increases.
Grade is moderately well distributed

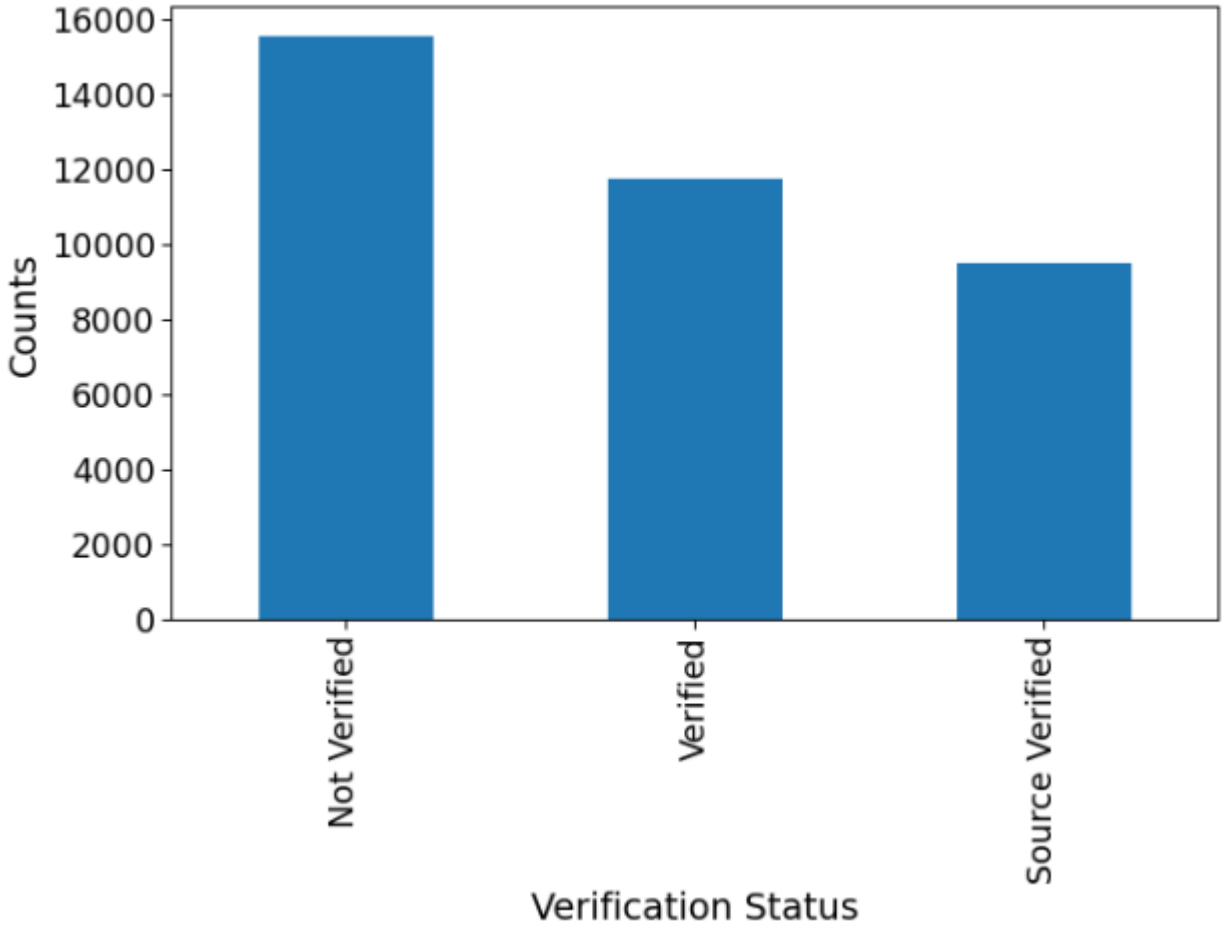
DTI





DTI is mostly proportional with default rate . But at very high DTI the default rare slightly decreases. Although the increase is very minor.
DTI is also well distributed .

Verification Status



Not Verified , Source Verified , Verified – Default rate increases in this order

Verification Status is well distributed .

Miscellaneous Attributes

While analyzing we also found out some attributes which seemed important. But they had issues. Such as they weren't well distributed, or they didn't have a continues or conclusive pattern in their plots against **Default Rate**. We are noting down the most important information we got from these attributes below.

- Loans issued in 2007 has the highest rate of default but onwards the default rate has remained similar y-o-y. Number of loans has increased with from 2007 to 2011 with time in a large number.
- People with annual income in the range of 4000 and 252000 are more likely to take loans . And people with annual income in the range of 752000 and 948000 have highest default rate.
- The maximum number of loans have installment in the range of 16 and 273. And the default rate is highest in the loans with installment between 790 and 1043.

Conclusion

After doing exploratory data analysis on the loan dataset we can conclude that Interest Rate , Loan Amount , Term and Grade are directly proportional to default rate . DTI is also mostly directly proportional to default rate but there is a sudden decrease in Default Rate for the very high values of DTI. And Not verified , Source verified , Verified - the default rate increases in the same order for Verification Status.

Lending Club can consider these attributes to figure out which loans are most likely to default and reduce their risks.