# Lending Club

## A Case Study Using EDA

**Harshal Bhatkhar & Sunkanna Basa**

# The Problem Statement

To identify the driving variables for loan defaulting from the fresh loan applicants in Banking and Financial sector, and to minimise the loan risk of losing money while lending money to the customers

# The Approach

Using the Exploratory Data Analysis (EDA), we took a shot at the loan data to dug and probe around it for facts. We get to the bottom of the data to find out the driving variables for loan defaulting.

- Understanding and cleaning data
- Applying Univariate Analysis to elicit facts from columns
- Applying Bivariate Analysis to see how two variables help us identify the driving variables of loan defaulting

The driving variables or factors are recommend at the end, after the analysis.

# The Approach

Summary

- The data ‘loans’ is the data after cleaning and standardising
- The data ‘loan_less’ is the segregated data, used in Bivariate Analysis, for the funded amount by investors, which is equal to or less than 10000
- The data ‘loan_more’ is the segregated data, used in Bivariate Analysis, for the funded amount by investors, which is equal to or greater than 10000
- The data ‘loan_overall’ is the segregated data for loan status ‘Charged Off’, and is used to analyse the driving factors behind a loan defaulting

Assumption

The value ‘10000’ is just an assumption as it is close to the 50th percentile for the ‘funded amount by investors’ variable

# Univariate Analysis

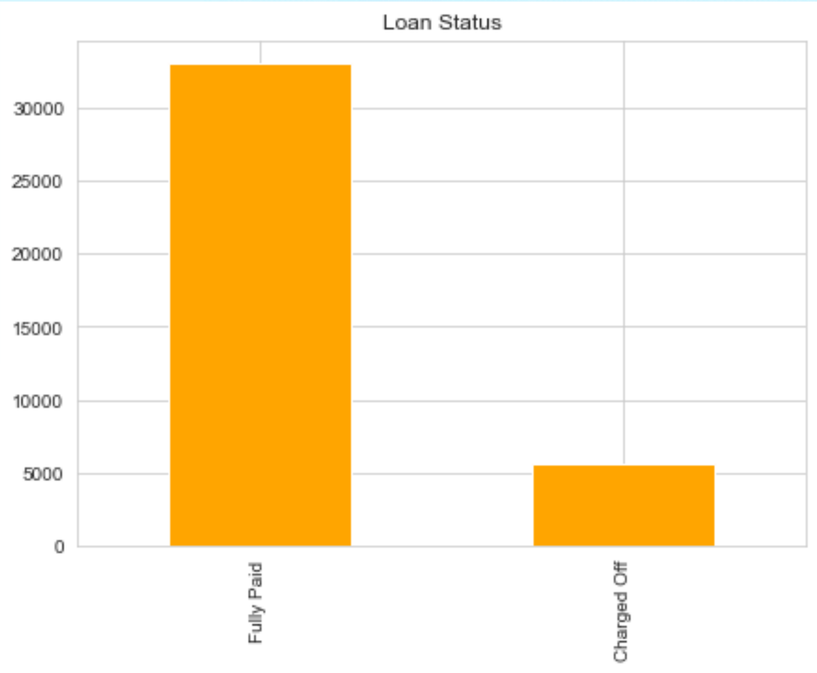
The results from analysing a single variable from ‘loans’ data

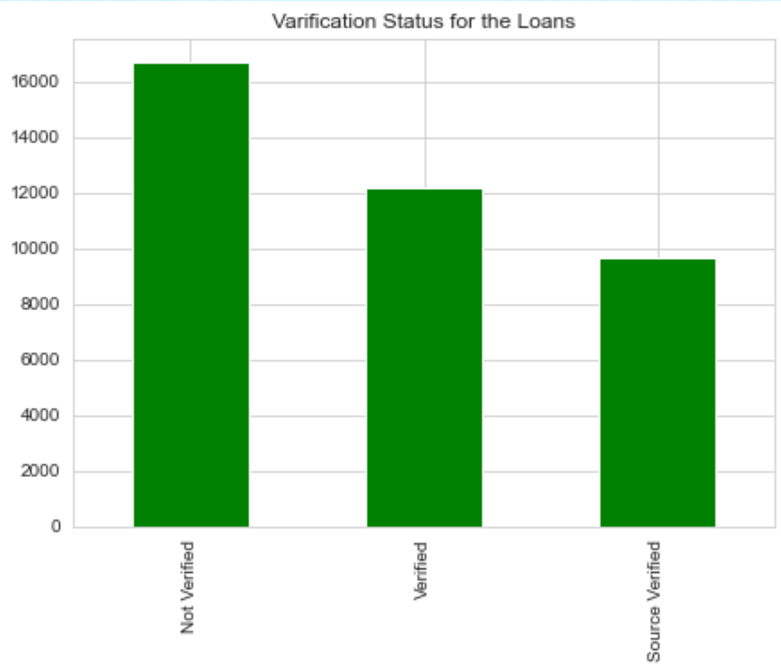
Here on the loans data, we have:

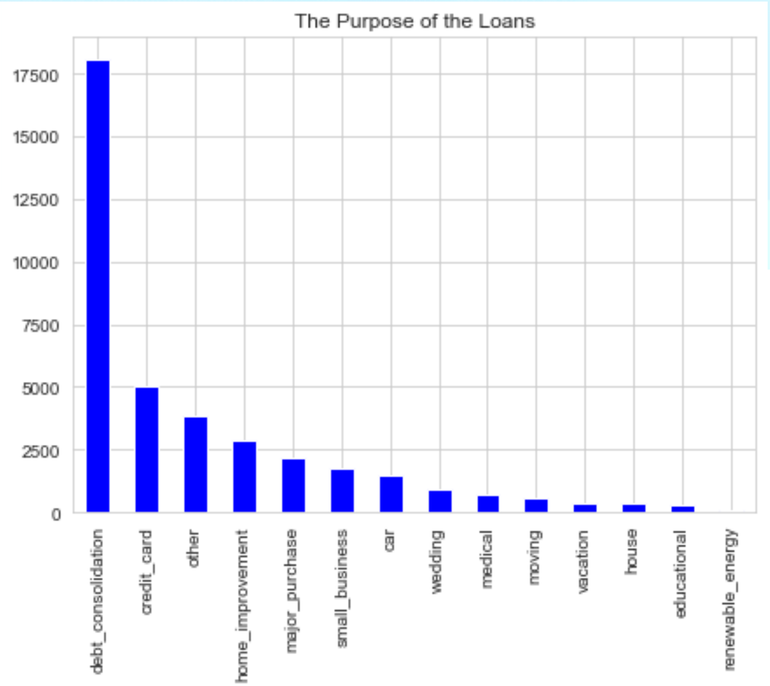
- Unordered Categorical Variables: Loan Status, Verification Status, Address, Home Ownership
- Ordered Categorical Variables: Grade, Sub-grade, Issue Date & Employment Length

# Univariate Analysis

## Loan Status, Verification Status & Purpose



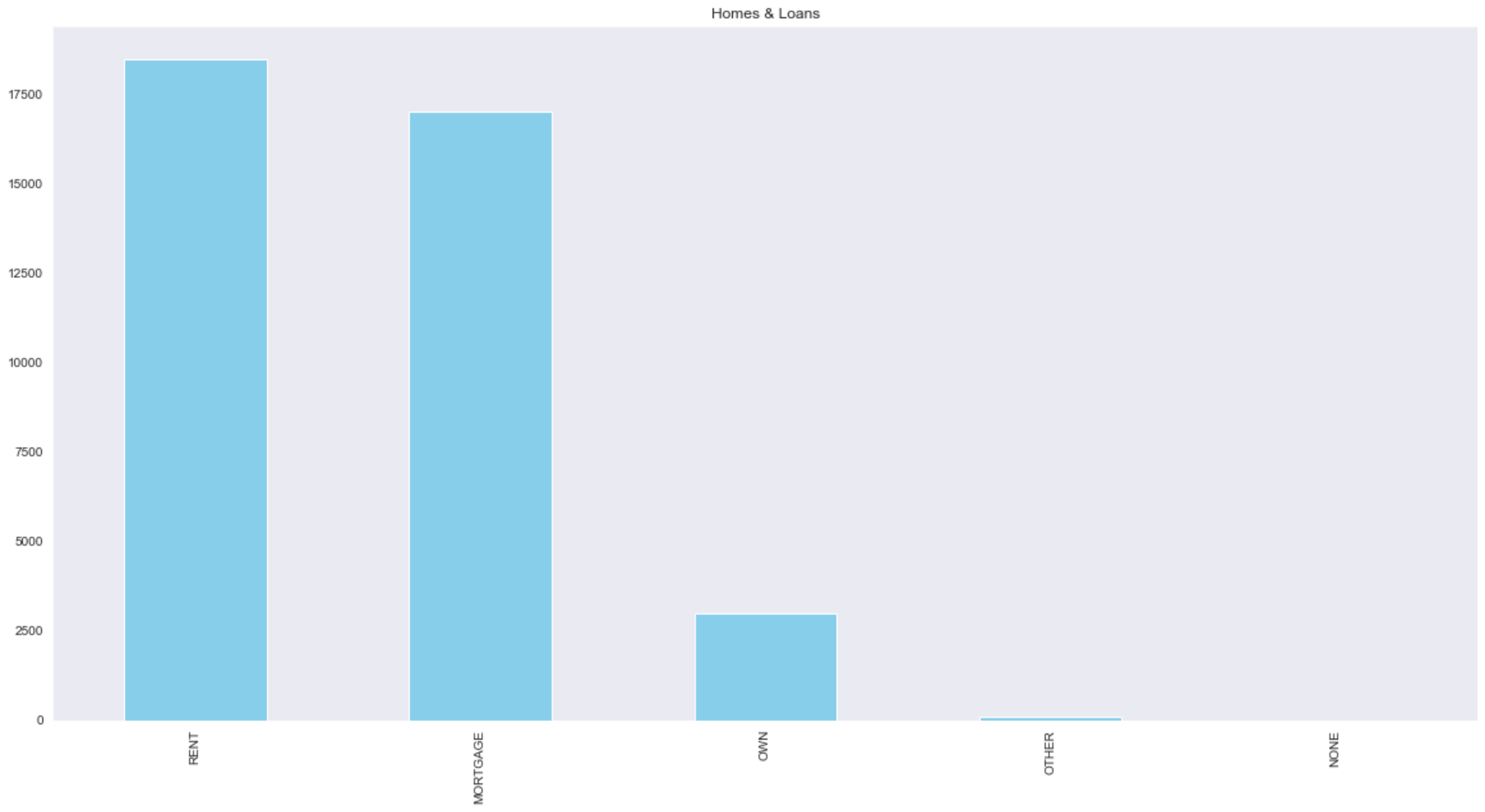
# Univariate Analysis

## Insights

- About 13 percent of the loans have turned out to be bad loans (Loan Status)
- Majority of the loans have not been verified (Verification Status). Unverified loans maybe one reason forloan defaulting.
- Most loans(especially the top three) have no signficant purppose. Debt consolidation is to get a single, bigger to clear multiple loans. Credit card spends and unspecified purpose (Other) loans. Poor financial management may turn out to be one of one of the driving factors for the loan defaulting. We have to see that in the Bivariate Analysis. (From The Purpose of the Loans)

# Univariate Analysis

## Loans Across Home Ownership



# Univariate Analysis

## Insights on Home Ownership

Even Mortgage has almost equalled to Rent column in the chart. In reality, the Mortgage loans are more defaulted between 2008 and 2011, which resulted in Subprime Mortgage Crisis in the United States.

# Univariate Analysis

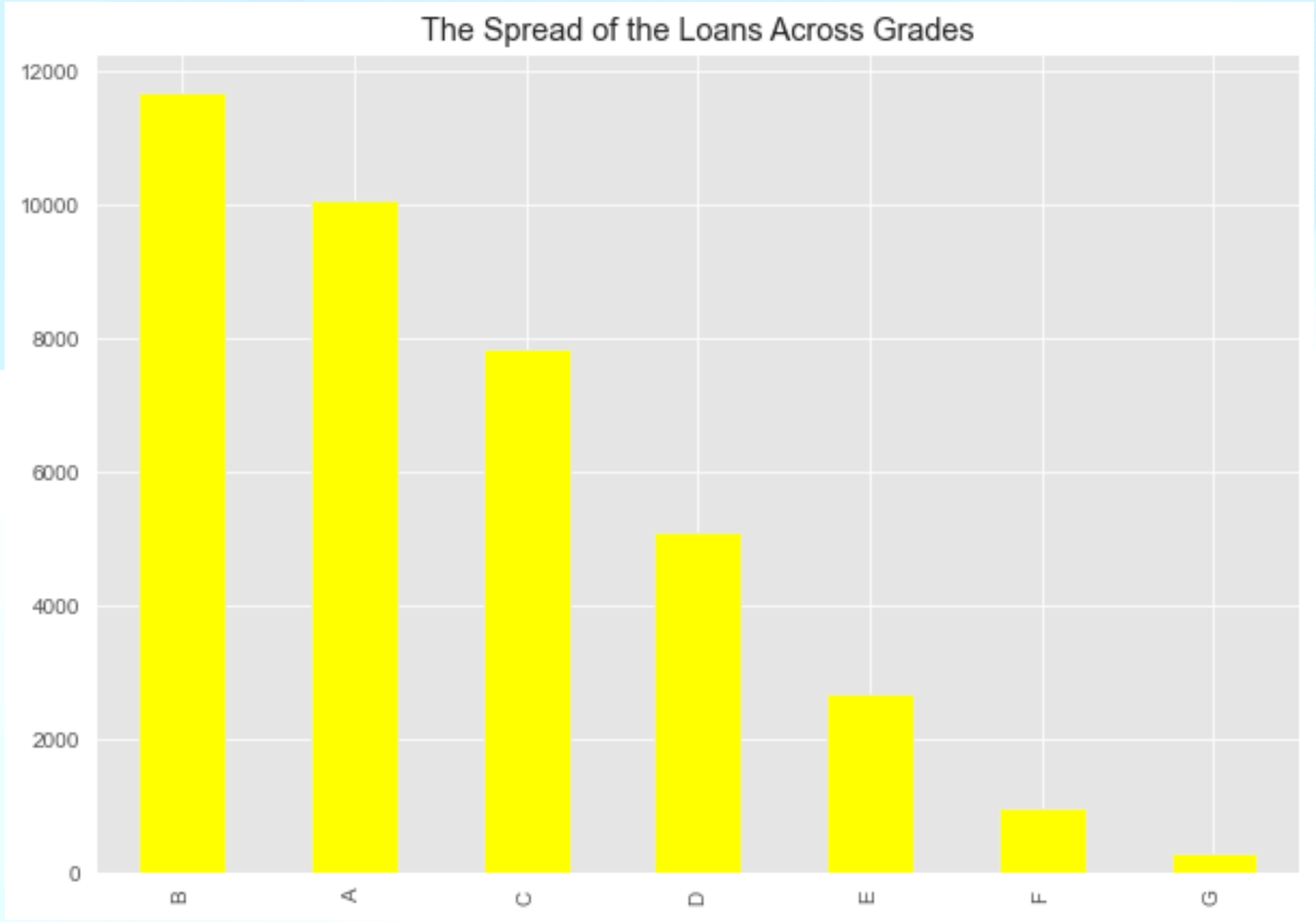
## Ordered Categorical Variables

Let's try to understand what these variables tells through the insights from them

- We will new columns such as Year and Month and try to elicit insights from them
- We will also have a look at the Grade and Subgrade columns and see if they become part of the factors for loans defaulting

# Univariate Analysis

## Grades




The first three Grades contribute to the more number of loans. So a chunk of loan defaults may come from these grades combined with Mortgage in Home Ownership or Unverified from Verification Status or insignificant purpose from the Purpose column.

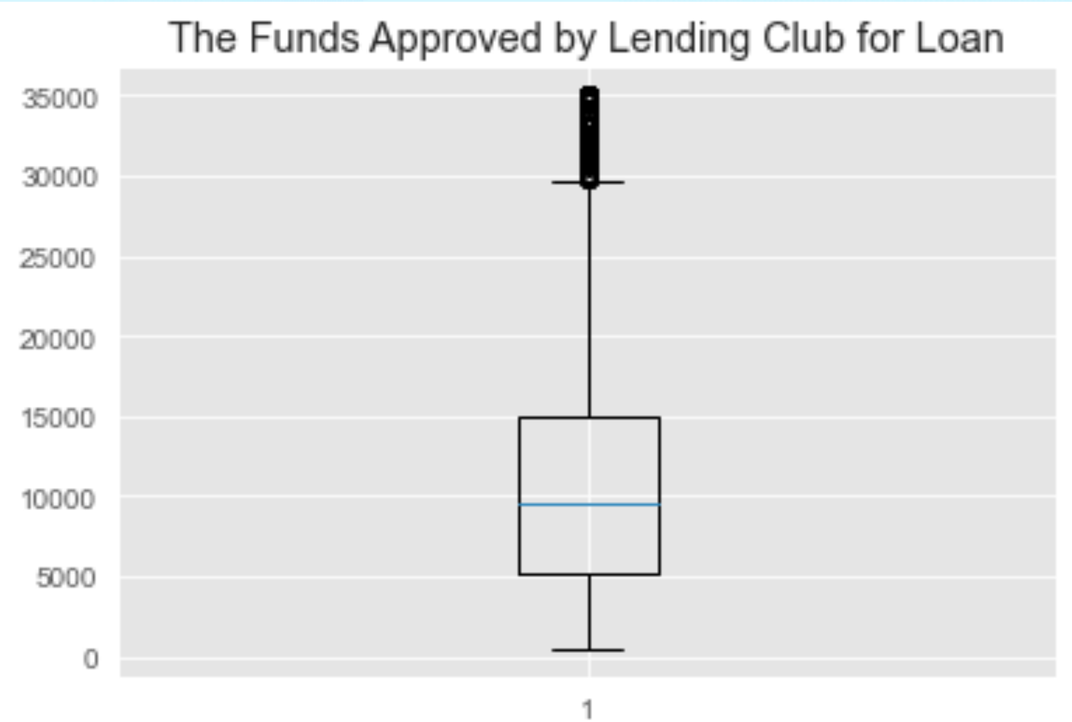
# Univariate Analysis

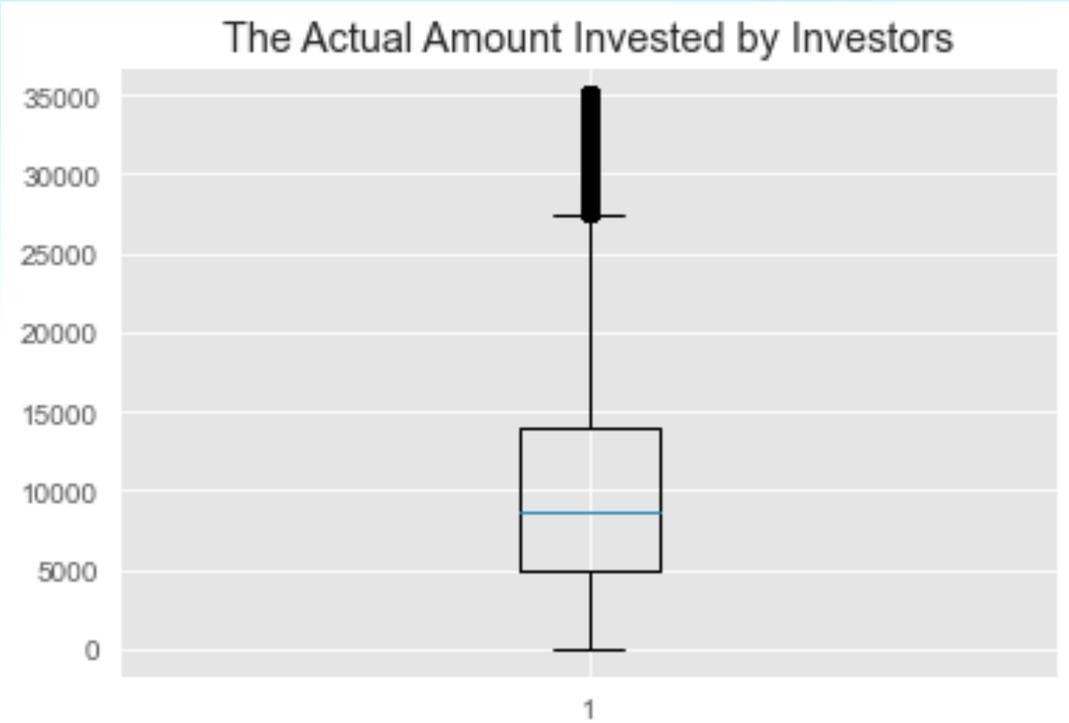
## Quantitative Variable

- There are some quantitative variables like Interest Rate, Debt to Income (dti), Income, Funded Amount By Investors, Funded Amount, and the Loan Amount.
- The dti or FICO Score(which is not here, and is similar to CIBIL Score in India) is one simple and quick way to get sight of loan defaulters. Let's try if the ‘dti’ helps us any way out.

# Univariate Analysis

## Funded Amount & Funded Amount by Investors



# Univariate Analysis

## Comments | Funded Amount & Funded Amount by Investors

- There are outliers in the loans data, which we are not manipulating here. Three-fourths of the loans are under 15000. The outliers with 15000 to 20000 more, on the upward-side, are influencing the data.
- There are about 25 percent of loans above the 75th percentile. These outliers influence the data analysis overwhelmingly.
- On the otherhand, the outliers on the lower side is quite a few

# Bivariate Analysis

## For Decision-Making

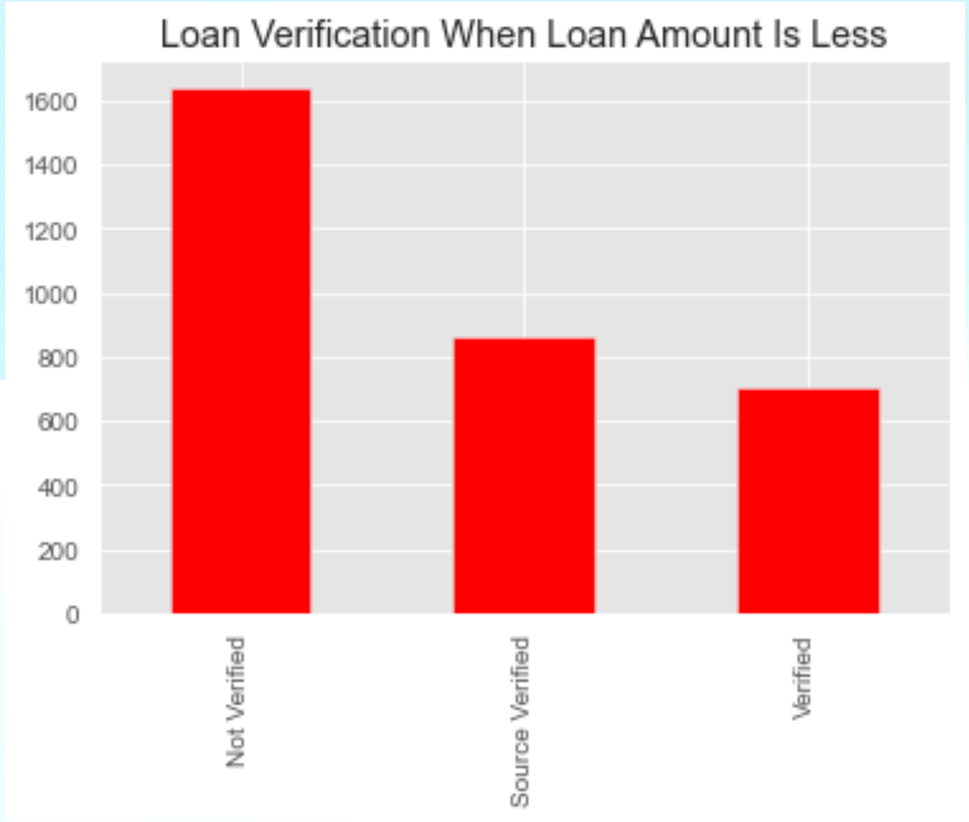
So far we tried to understand what a different column says and tried to draw insights from them as much as possible. Now let's segment the columns into two groups and try to understand the driving factors for a loan default.

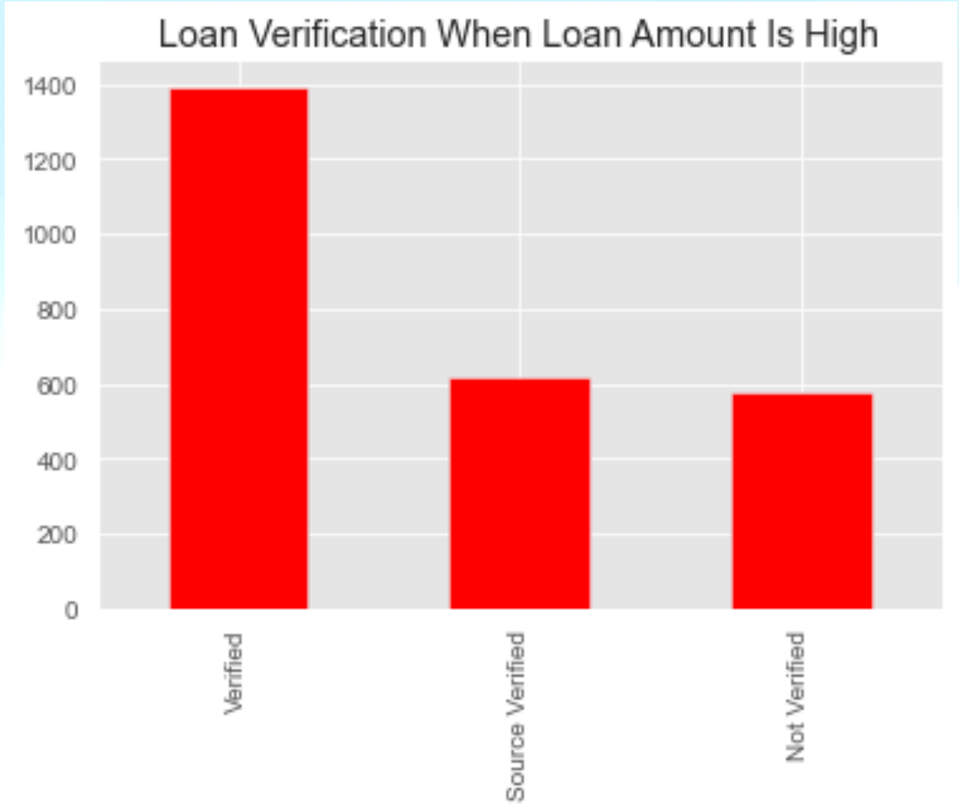
we have:

- A table for Funded Amound by Investors below 10000, but the loan status is Charged Off. The question to ask is why the borrowers defaulted the loans when the funded amount is less?
- Another table for Funded Amound by Investors above 10000, but the loan status is Charged Off. We ask the same question here to: why they charged off when the funded amount is more?

# Bivariate Analysis

## Verification Status When Loan Is Less, And When Loan Is More



# Bivariate Analysis

## Verification Status When Loan Is Less, And When Loan Is More

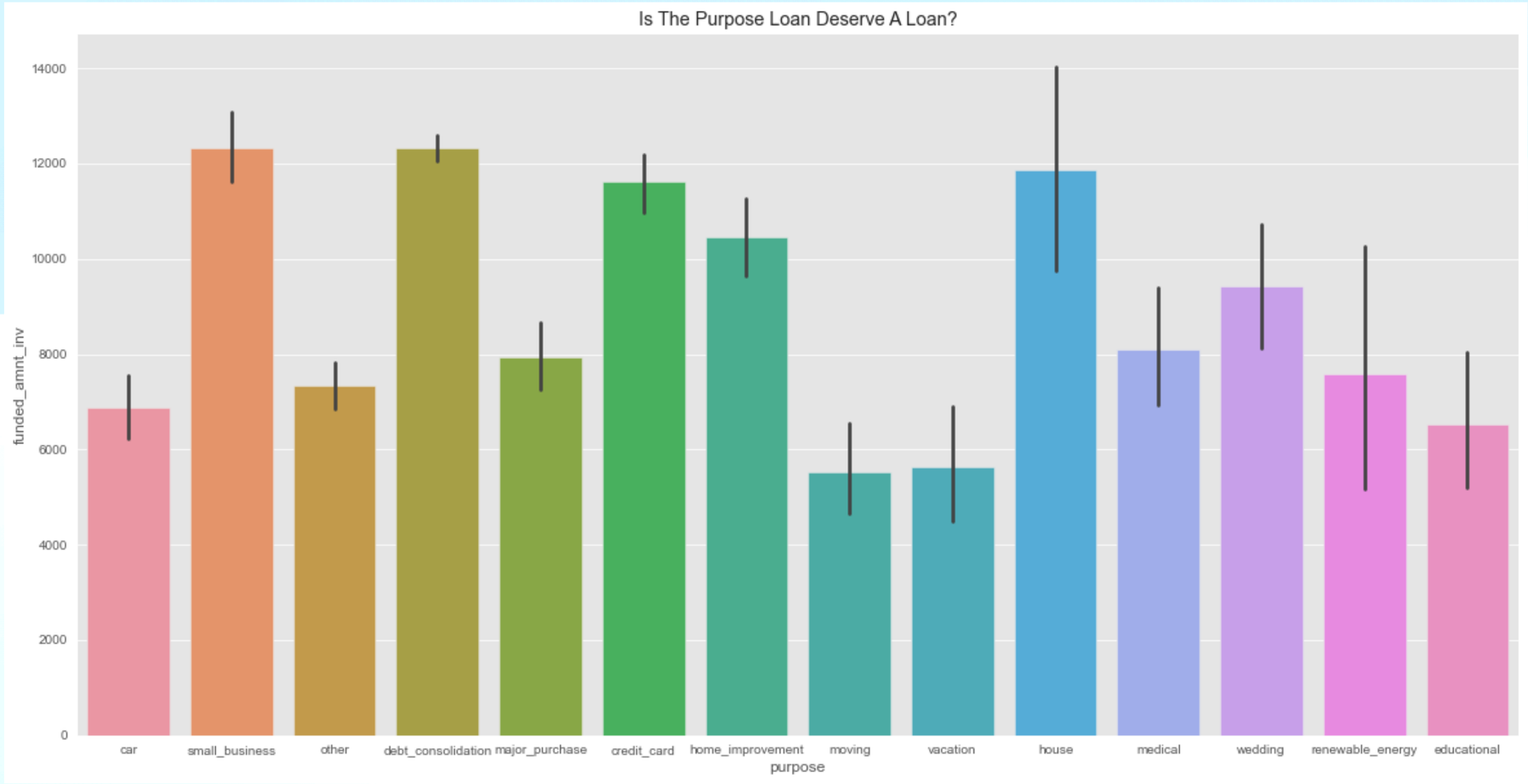
Insights:

- There are more Charged Off loans when the loan application is not verified and the amount is less (<10000)
- There are more Charged Off loans when the loan application is verified and the amount is more (>10000)

Overall, the Not Verified loans are slightly more than the Verified loans in the Charged Off status

# Bivariate Analysis

## Purpose Against Funded Amount



# Bivariate Analysis

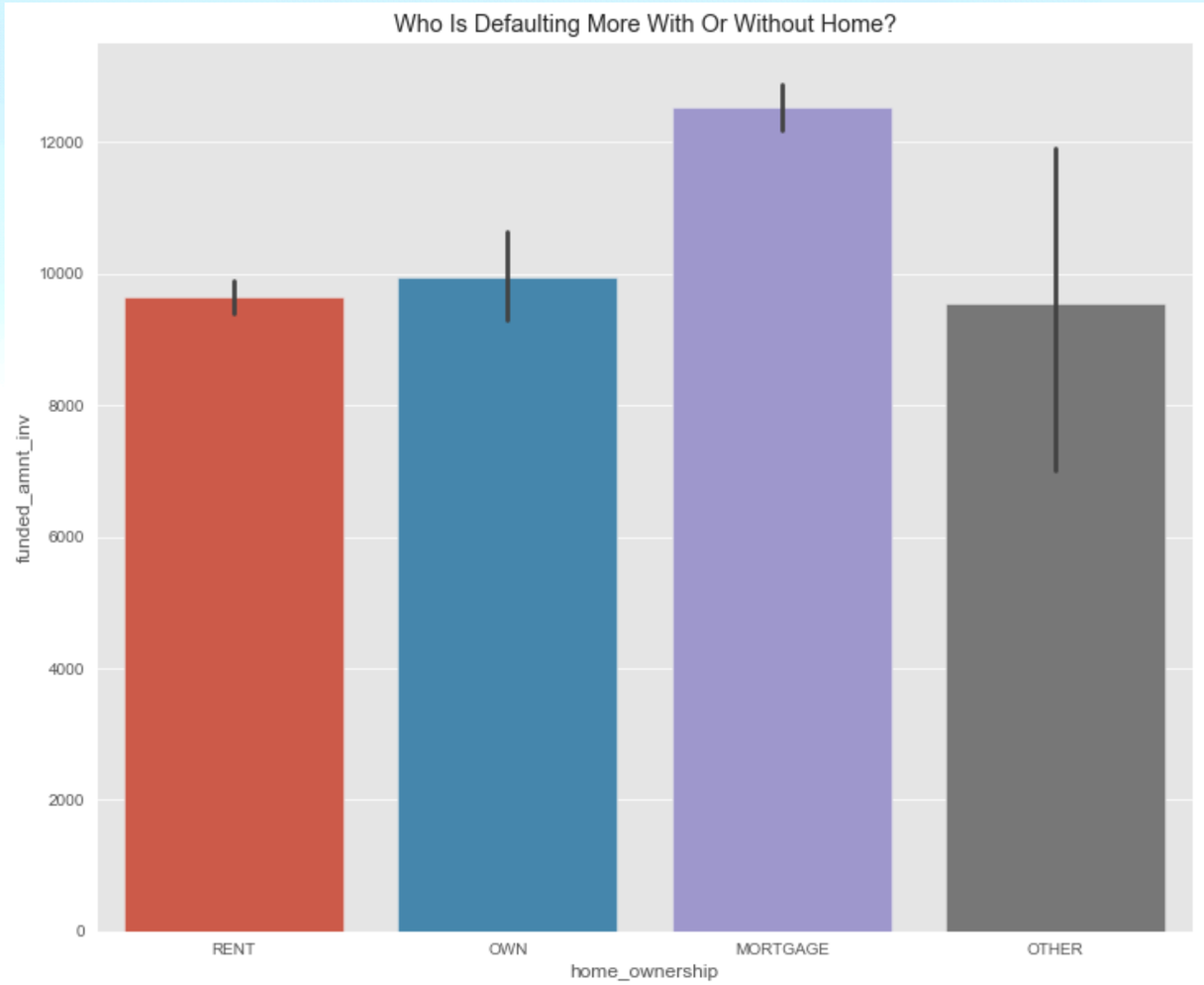
## Insights | Purpose Against Funded Amount

- The Small Business loans are more Charged Off loans. Secondly, the Debt Consolidation loans bagged the second spot. And the next is Credit Card. We can't ignore Home Improvement and House as they are accounted for the Subprime Lending Crisis.
- The Small Businesses borrowers may pay back or may not. The Debt Consolidation and Credit Card loans are risky and unsecured loans, as the applicants are Serial Debtors (arent't they?).

So the purpose of the loan borrower is going to be a driving factor for a loan default.

# Bivariate Analysis

## Home Ownership & Funded Amount by Investors

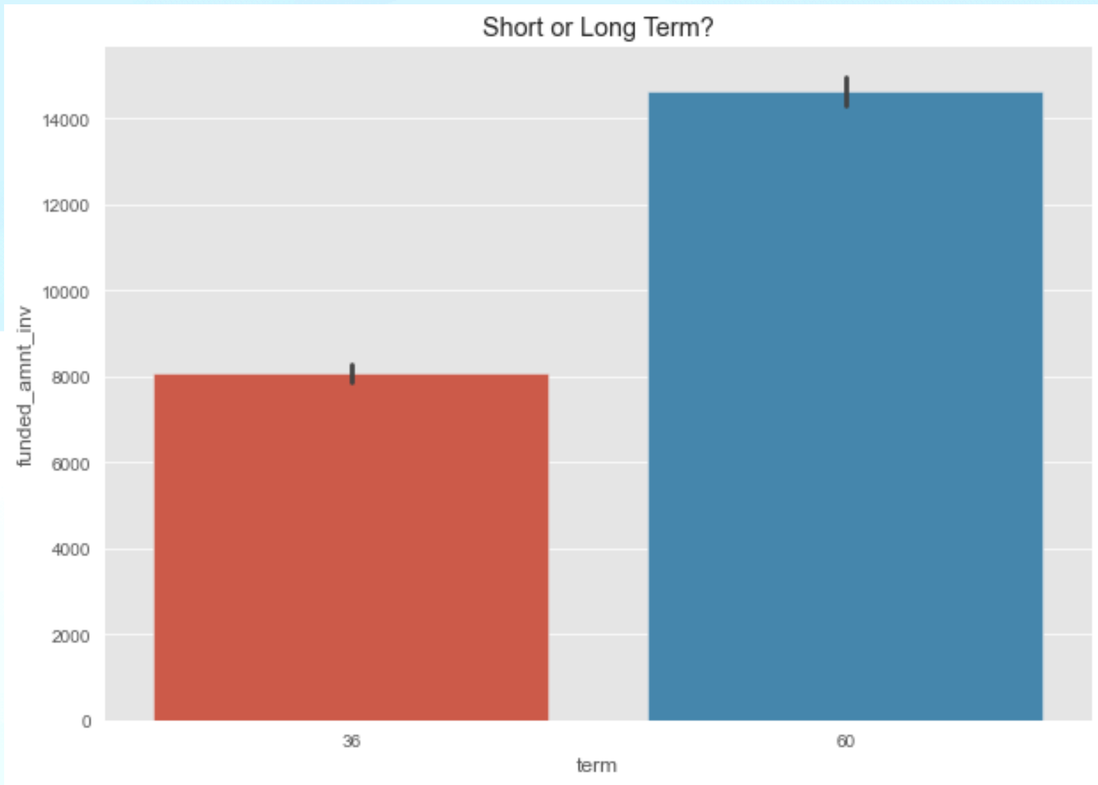


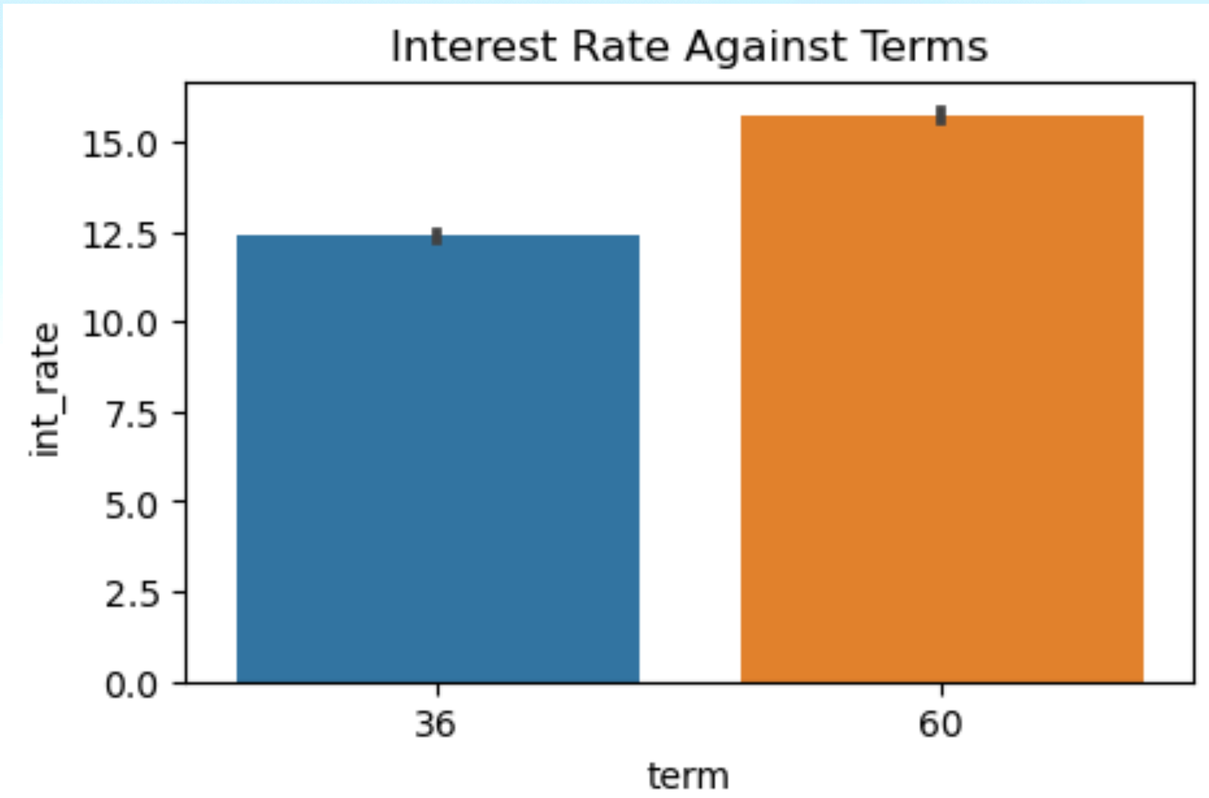

Insights:

If the borrowers are mortgaging, they are more likely to default their loans

# Bivariate Analysis

## Term Vs Funded Amount by Investors & Interest Rate



# Bivariate Analysis

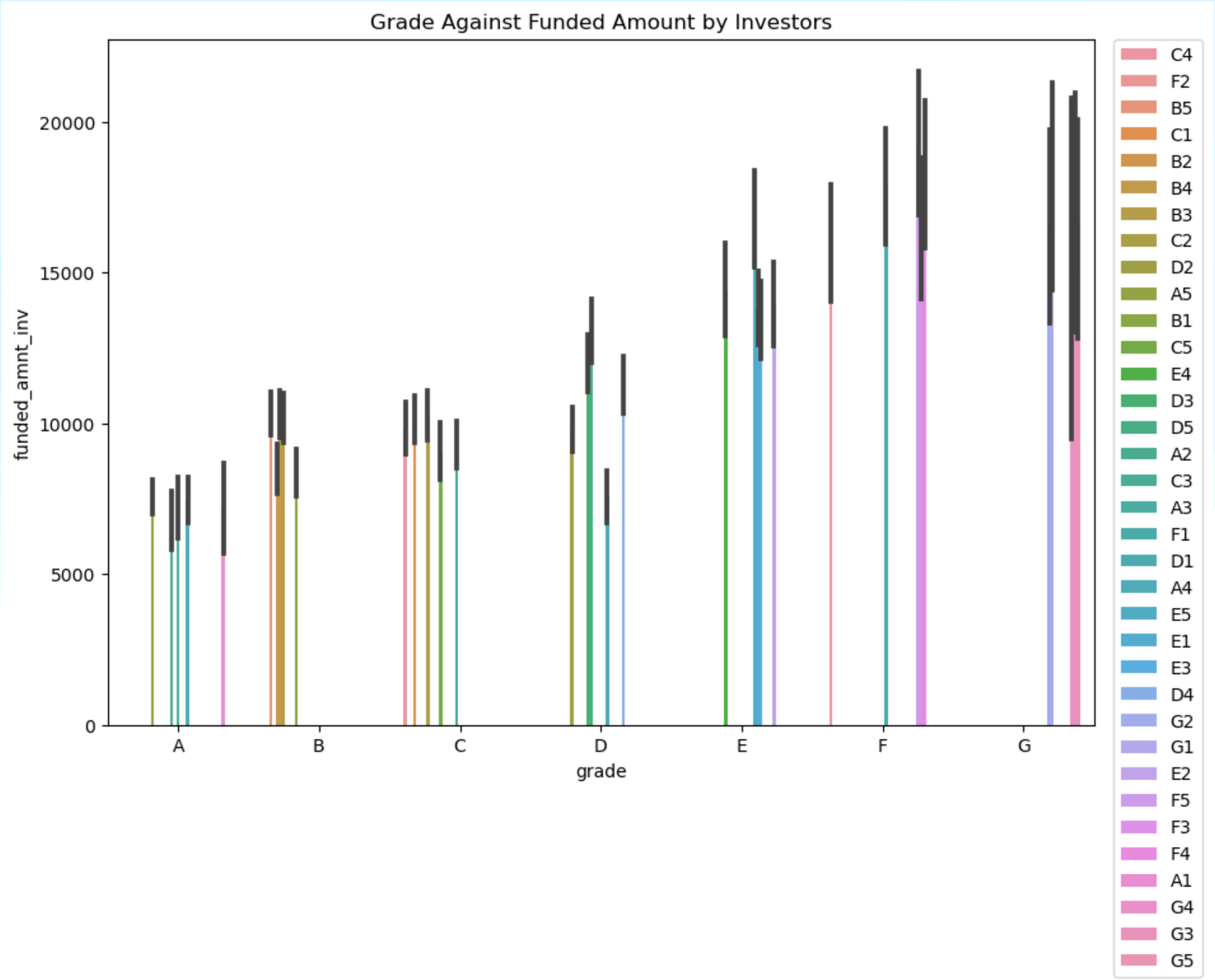
## Term Vs Funded Amount by Investors & Interest Rate

Insight:

The longer a loan term is, the more likely a borrower is gonna default it. Even the Interest Rate against Term prove it all. The long term loans with high interest rates are one of the driving factors for loans defaulting.

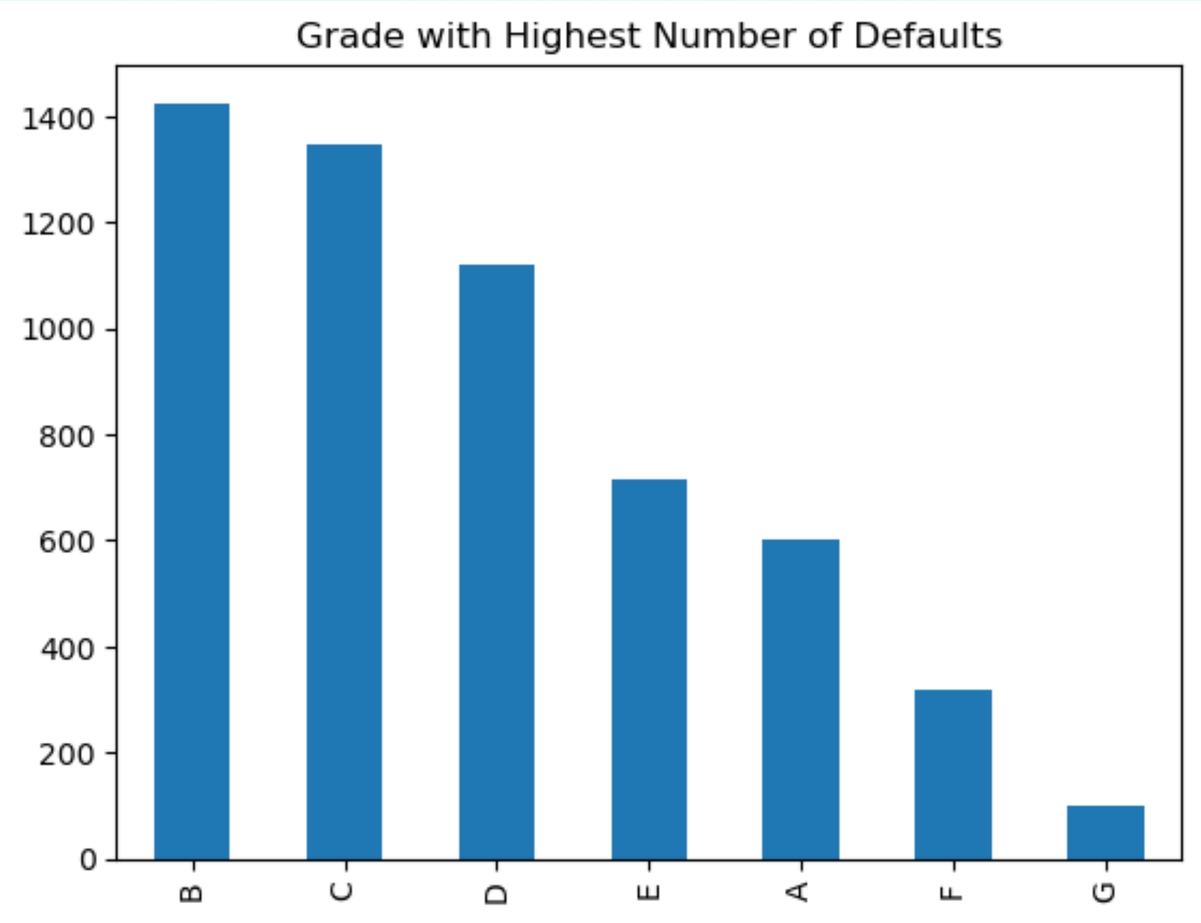
# Bivariate Analysis

## Grades Vs Funded Amount by Investors



# Bivariate Analysis

## Grades Against ‘Charged Off’ Loan Status




Insights:

This is what we can understand from Grade Against Funded Amount by Investors and from the Grade with Highest Number of Defaulters.

- Smaller amount of loans with Not Verified status contribute to the loan defaulting

# Recommendations

## Final Facts | Driving Factors for Loan Defaulting

These are recommendations we can make from the Analysis of the above loans data:

1. Verify the loans if the loan amount is less (Verification Status and Funded Amount by Investors)
2. Give loans to the right purpose. The loans for the debt consolidation, small business, and credit card types are riskier. (Purpose).
3. Make changes to the duration of a loan term. (Term)
4. Mortgage Loans bring you loss. So avoid them. (Home Ownership)

# Thank You