Lending Club Case Study

Authors: Neeraj Dhiman and Anu Thomas

Problem Statement

A finance company which specialises in lending various types of loans to urban customers wants to analyze the **driving factors** that lead to **loan default**.

A dataset on borrowers for 5 years (2007-2011) is provided. Different attributes of the dataset include amount of loan taken, number of time the person has defaulted in any account, payment of installments etc. The Company aims to understand patterns in the data so that it can utilise this knowledge for its portfolio and risk assessment. This includes approval of loans, providing the loan at a higher loan rate for those likely to default etc. A data dictionary is also available to understand the fields in the dataset.

About the dataset

- There are about 40K records with over 111 attributes.
- Many of the columns are empty. Hence can be removed from the dataset.
- Some of the important categorical columns are loan_status, purpose, verification_status, grade, state, term.
- Some of the important numeric columns are loan_amnt, int_rate, debt to income ratio, annual income.

Important Industry metrics

The default rate is calculated using the following formula:

$$\text{Default Rate} = \frac{\text{Number of Defaulted Loans}}{\text{Total Number of Loans}} \times 100$$

Univariate Analysis
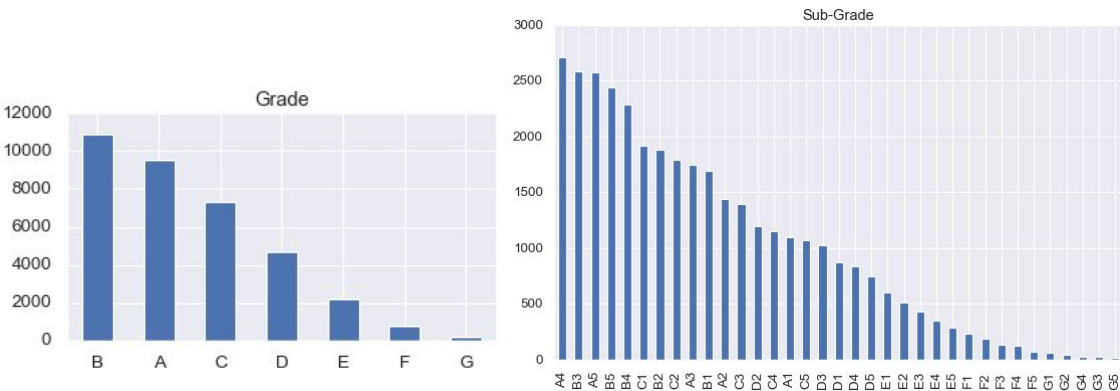
A histogram is a graphical representation of the distribution of a data set. It is also used to understand the frequency of the data. By looking at a histogram, we can understand if the data is skewed towards one extreme, if it is symmetrical, if it has many outliers, and so on. Here we are analyzing the **categorical** variables

Grade: Represented as A/B/C/D/E/F/G. Lot of loans provided in the A/B/C grades. Very few in F and G

Sub-Grade : There is lot of variations here. Looking similar to Grade and can be dropped from further evaluation
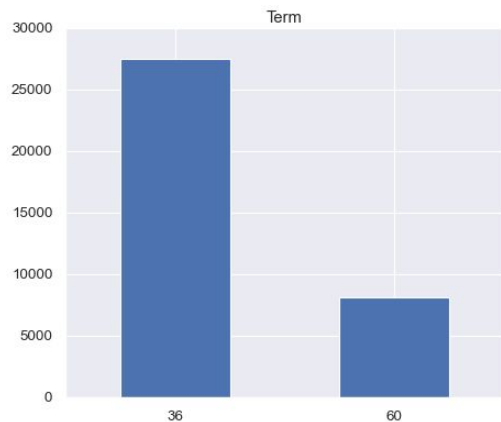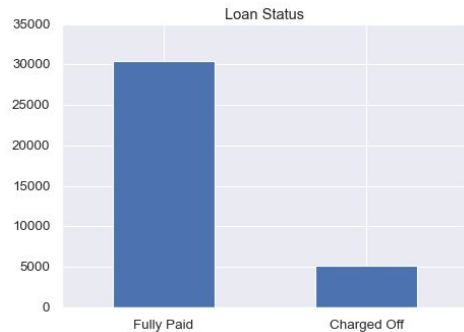
*Term*: There are only 2 category of loan duration - 3 or 5 yrs. Most loans provided for 3 yrs term.
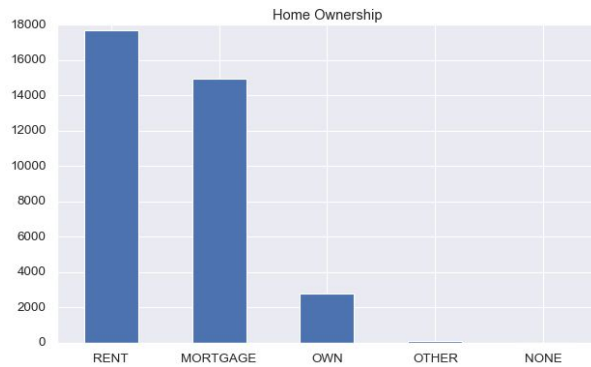
Home Ownership

Home Ownership: Majority of the borrowers are on rented or mortgaged properties. Very less own homes.

Verification Status

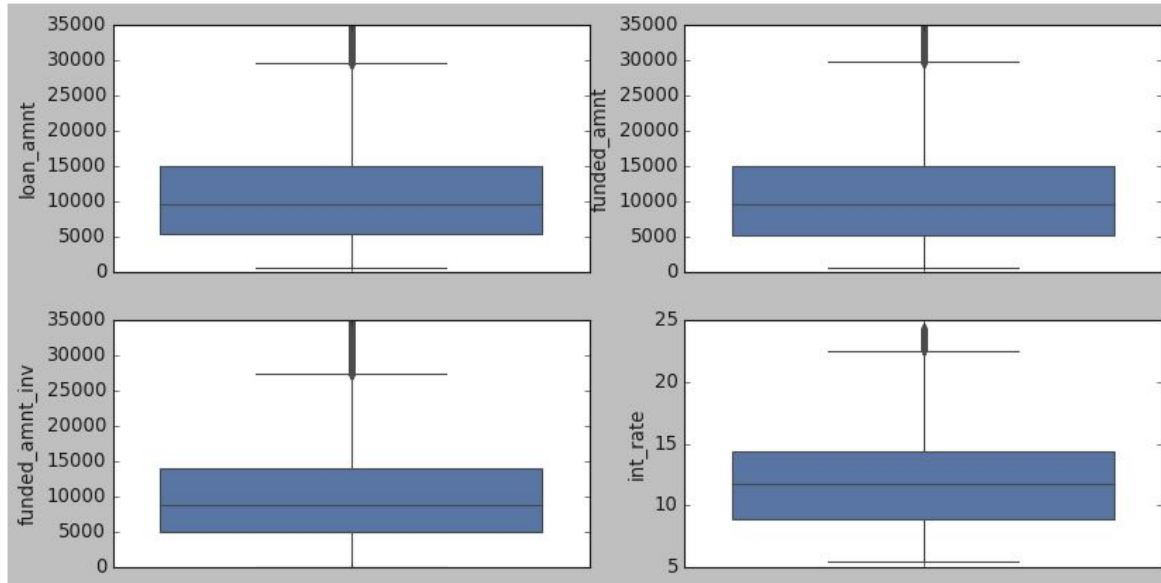**Verification Status**: Indicates if the income was verified. Majority of the borrower's income is not verified.

Loan Status

Loan Status: Most of the loans are settled. Around 6K loans have been defaulted.
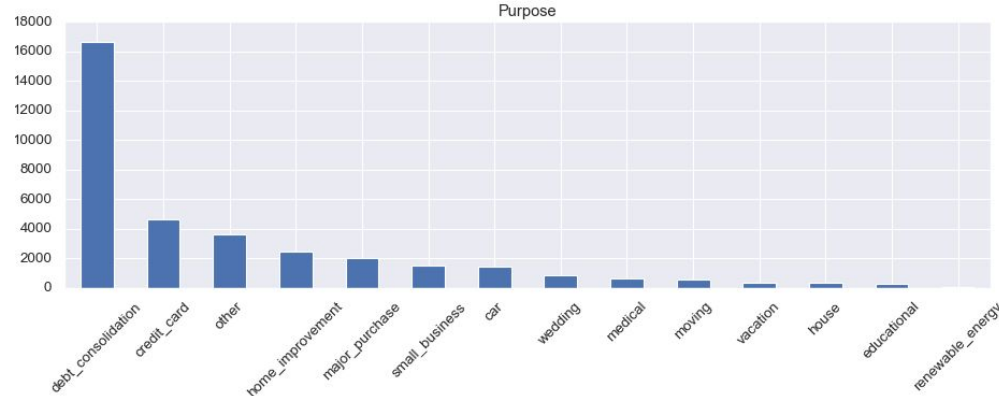
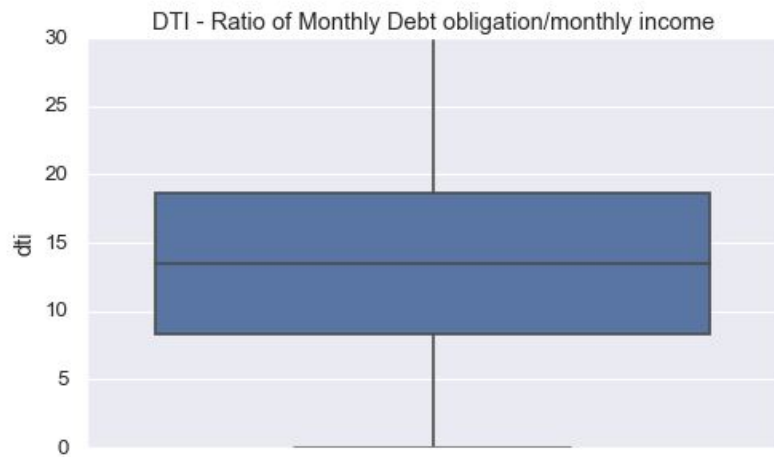**Purpose**: Majority of the borrowers have taken the loan to consolidate their debt.

Numeric Variables for outliers:

Loan Amount / Funded Amount/ Interest Rates: All has outliers beyond the upper fence.
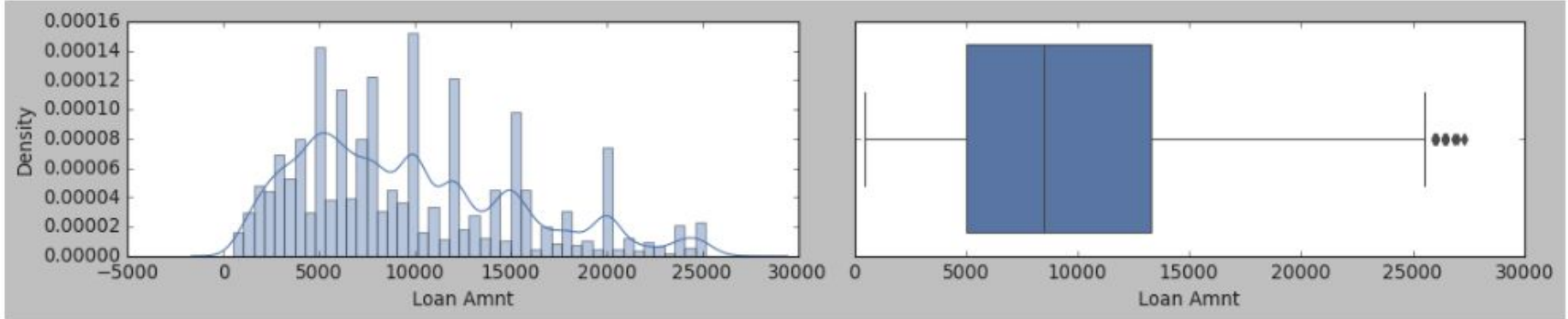
Annual Income

Annual Income: There is a lot of variations in the income, hence have drawn it on a log scale.

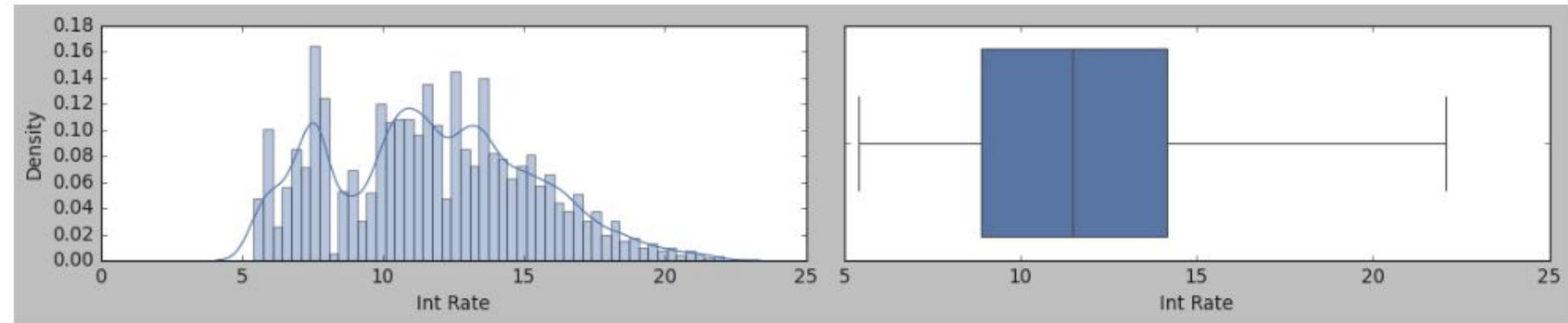DTI - Ratio of Monthly Debt obligation/monthly income

DTI : No outliers around the fact DTI..

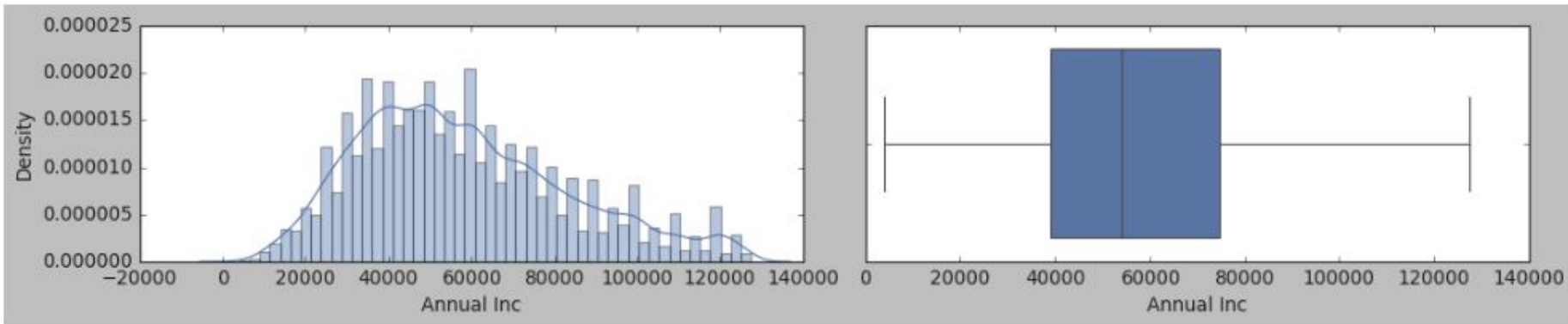# After removal of outliers:

Loan Amount

Interest Rate:
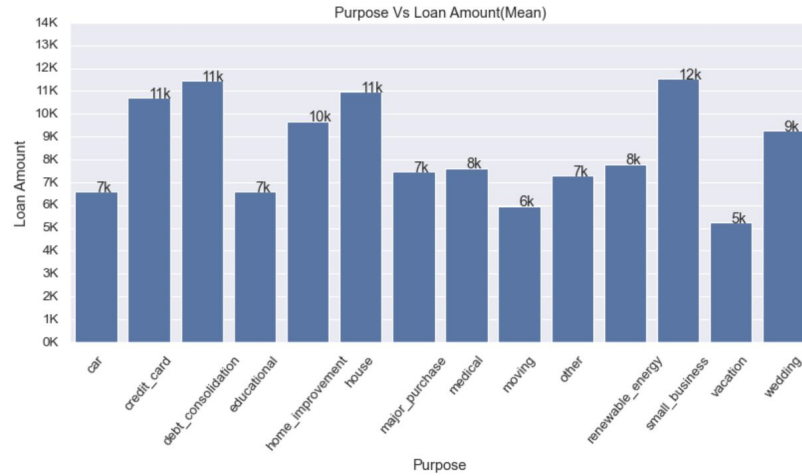

Most of the loan amounts range from from 5-14k.

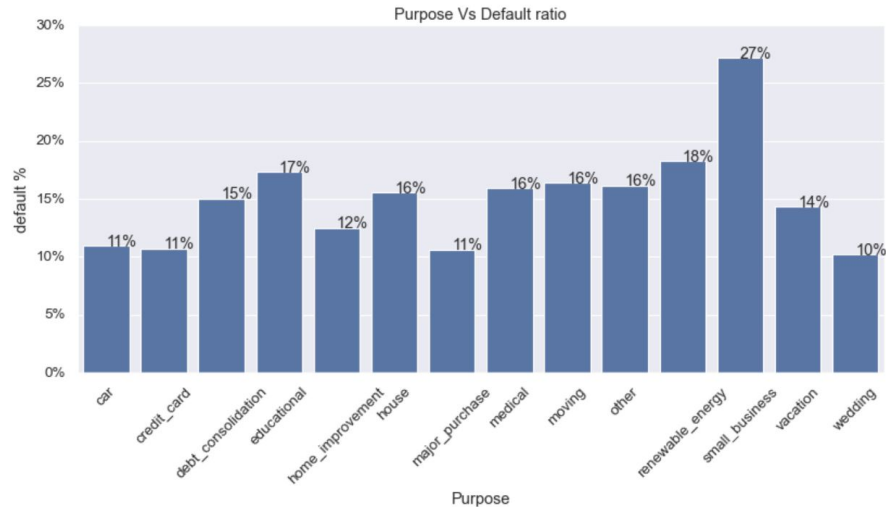**Most of the loans applied had interest rates between 9%-14%.**

Annual Income

Annual income of 50% of the borrowers are between 40k-75k

Bivariate Analysis
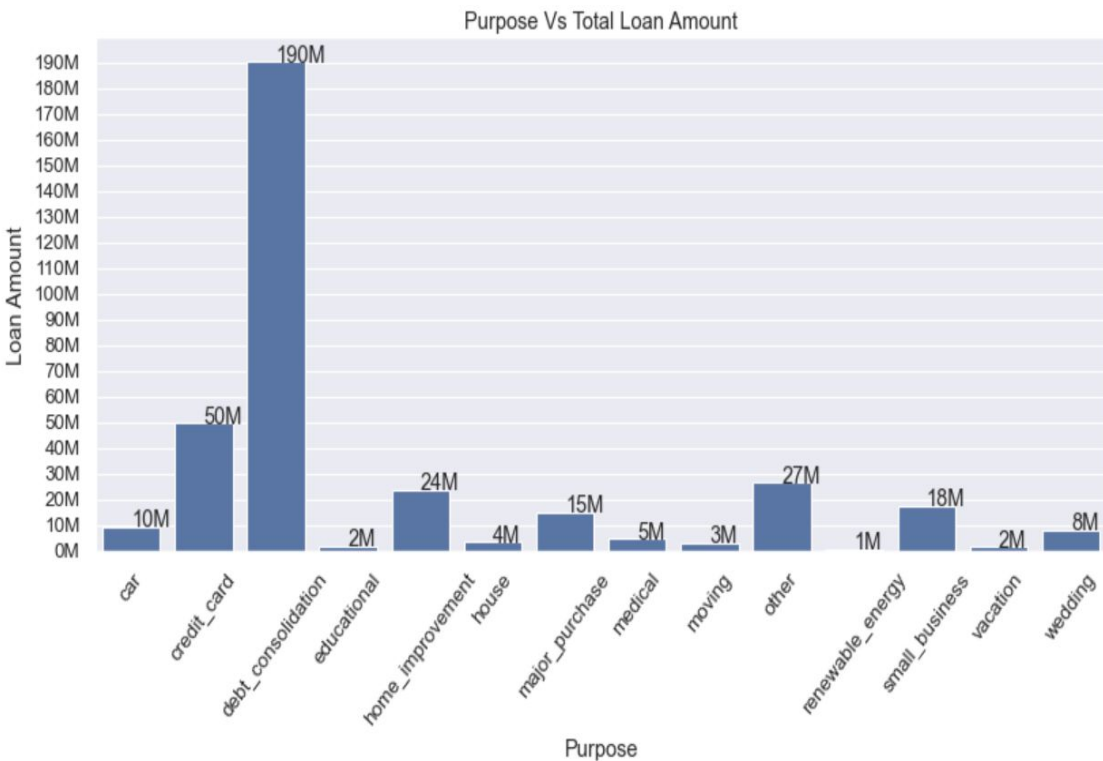
Purpose vs Avg. Loan Amount:
Loan amount of around $12K mostly availed for Small Business. For House, Debt Consolidation and Credit Card the average amount is around 11k.

Purpose vs Loan Default %:
Most defaults are in Small business
- Small Business - 27 %
- Renewable Energy - 18%
- Educational   - 17%
- House/Moving/Medical - 16%
- Debt Consolidation - 15%

Purpose Vs Total Loan Amount

Purpose vs Loan Amount:
At a cumulative level, there is $**190M** worth of loans provided for Debt consolidation
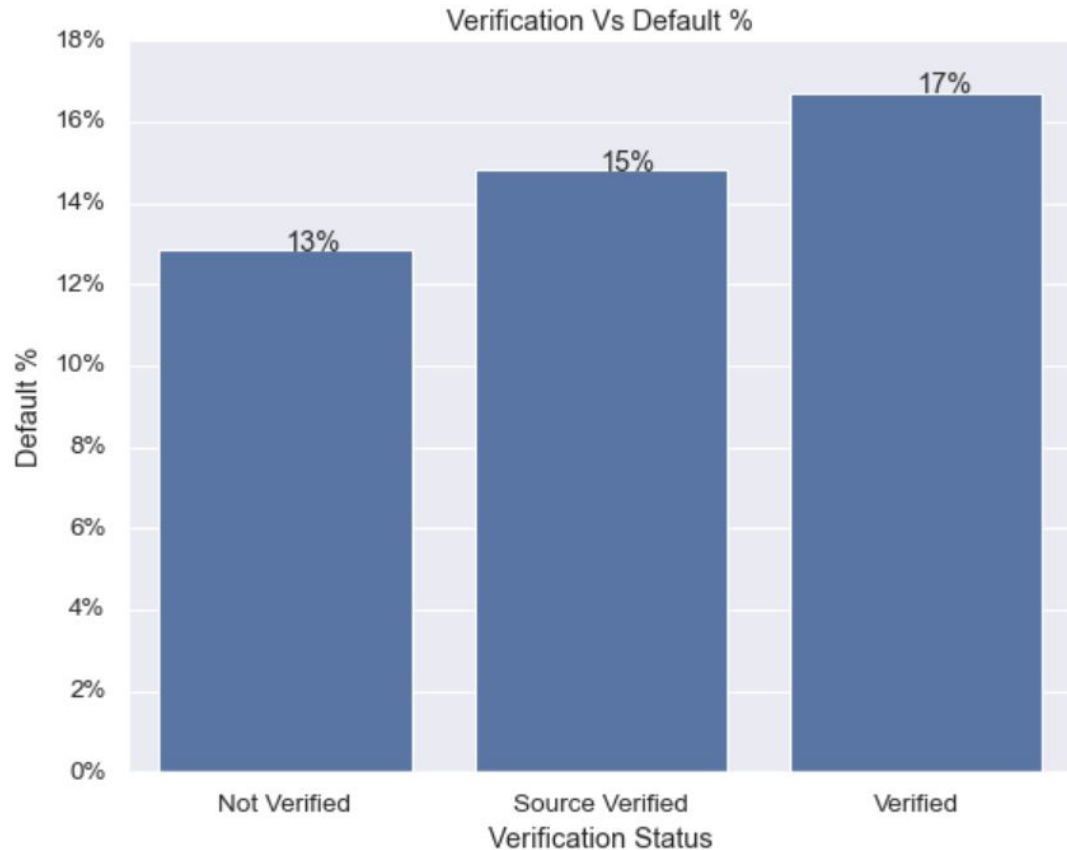Debt consolidation - 190M
Credit Card - 50M
Others - 27M
Home Improvement - 24M
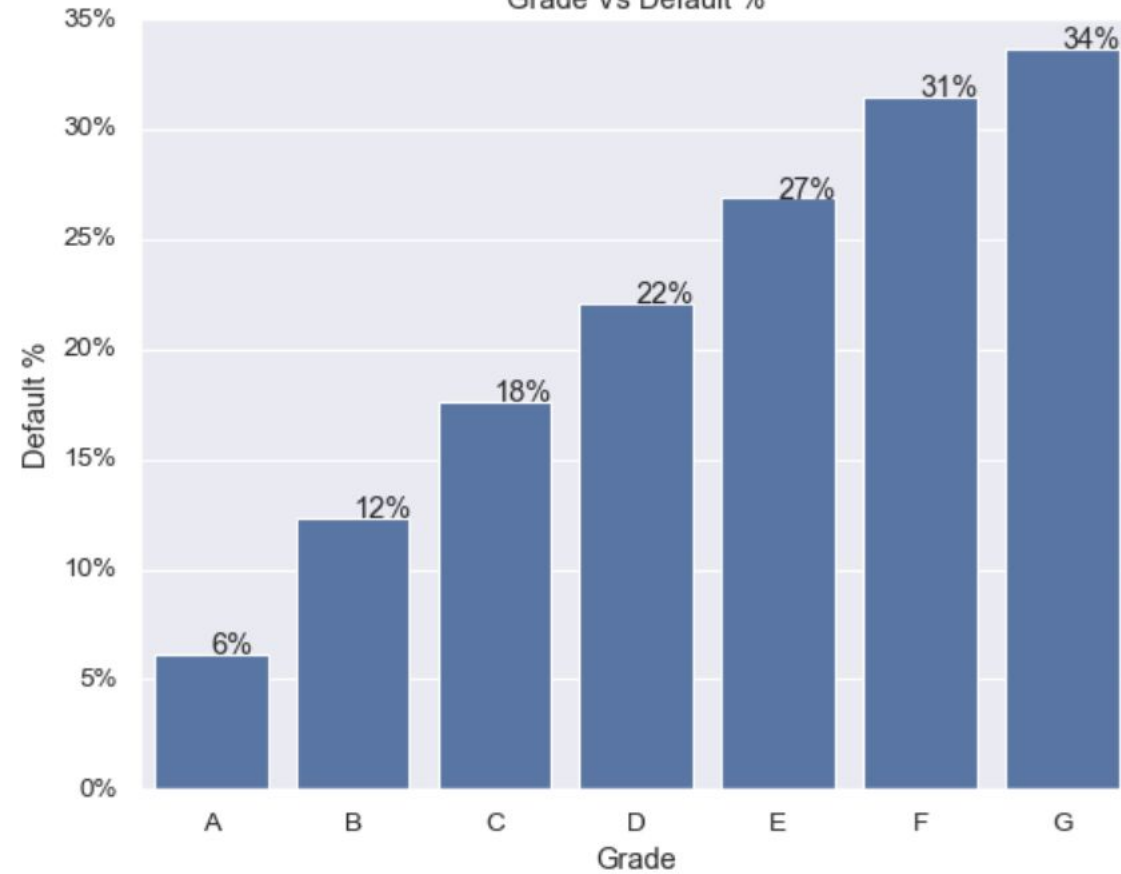Small Business - 18M
Car -10M

Verification Vs Default %

Verification vs Loan Default Rate

The loan verification process needs to be relooked at as most defaulters are borrowers whose income has been verified.

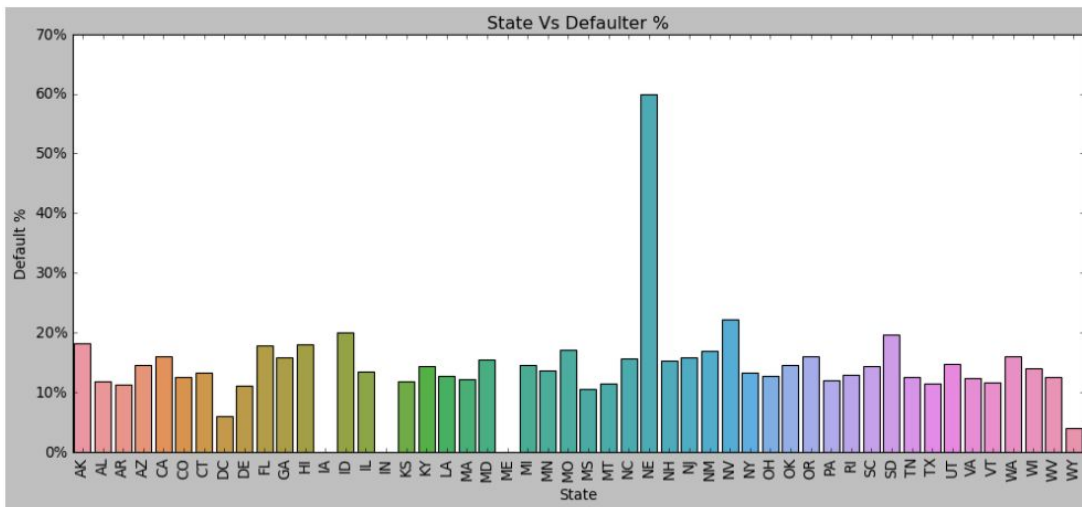Verified - **32**% (verified + Income Source verified)

Grade vs Loan Default Rate

Most of the loans provided in grade F and G has defaulted.

Default in Grade F and G approx **65**%

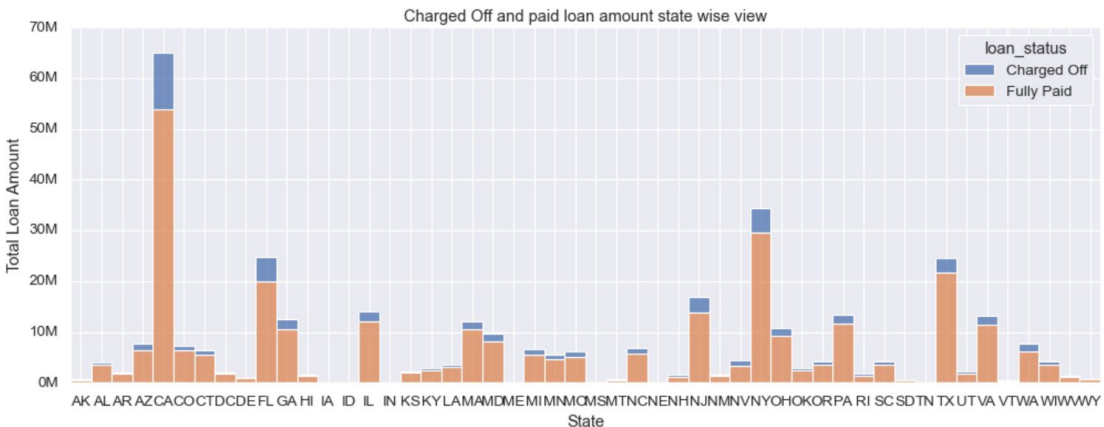Default in Grade E approx **27**%

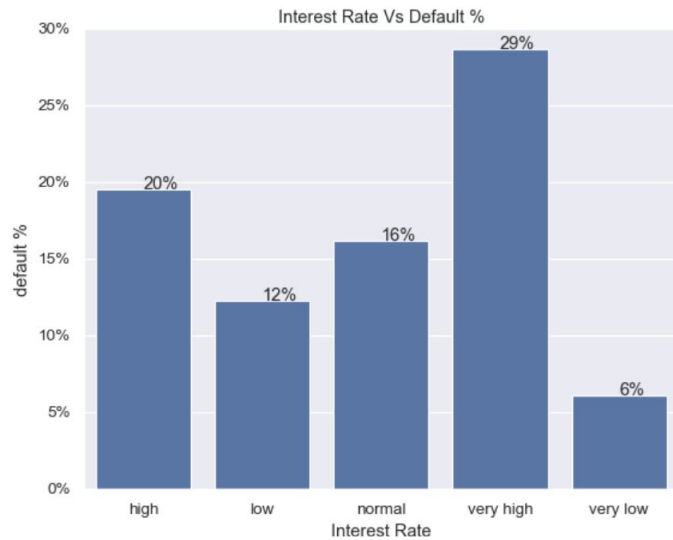Default in Grade D approx **22**%

State vs Loan Default Rate

Nebraska has the highest default rate - 60%

Indiana(IN), Iowa(IA) and Maine(ME) has zero default cases

California gets highest loan amount around $65M and highest in the loan default amount $11M approximately.

Interest Rate Vs Default %

Interest Rate vs Default Rate

Most Defaults are in the high interest rate loans where the rate is more than 14% → **50**%

Interest Range
Very low - <9%
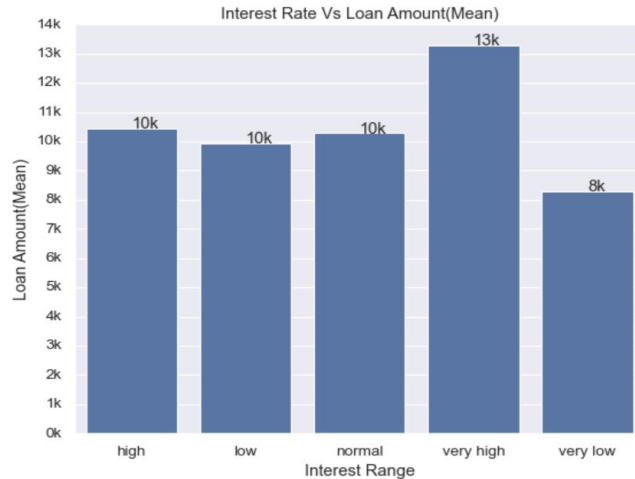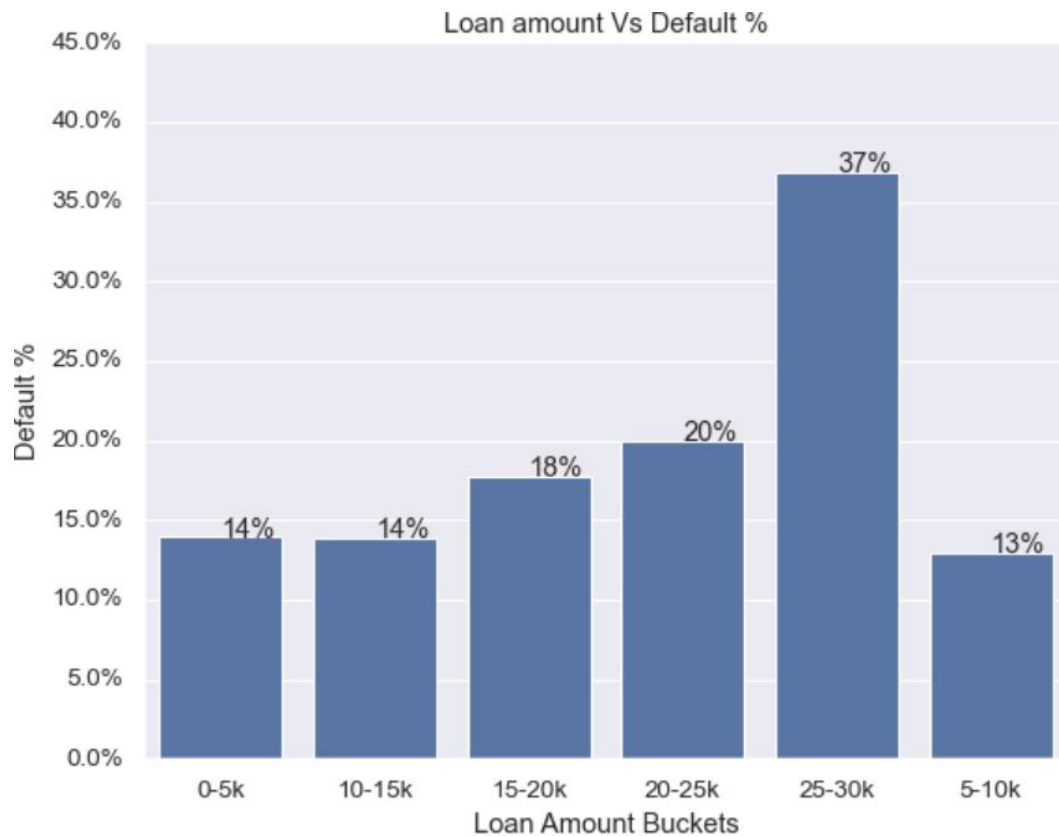Low   - 9 to 12%
Normal   - 12 to 14%
High   - 14 to 16%
Very High - >16%

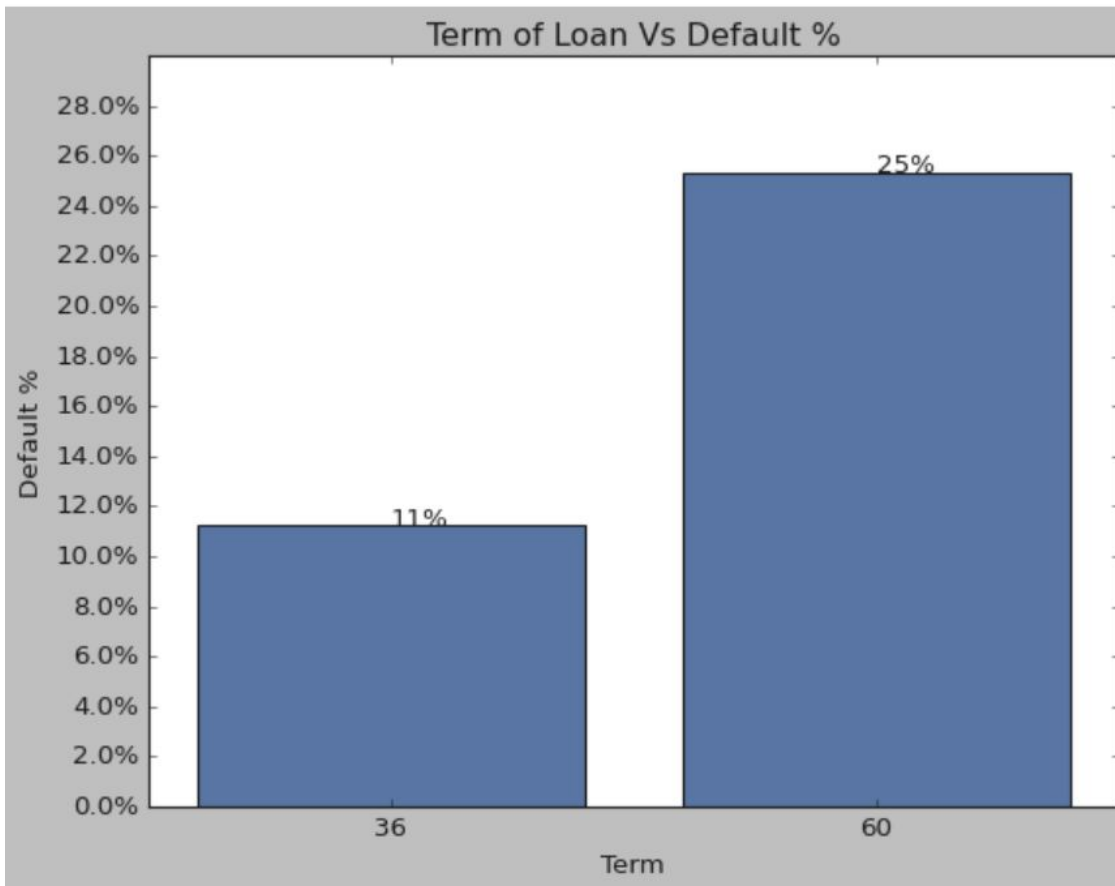Interest Rate Vs Loan Amount(Mean)

Interest Rate vs Avg. Loan Amount

The Avg. loan amount is highest in the loans where interest rate is "very high" - $13K

Loan Amount vs Loan Default Rate
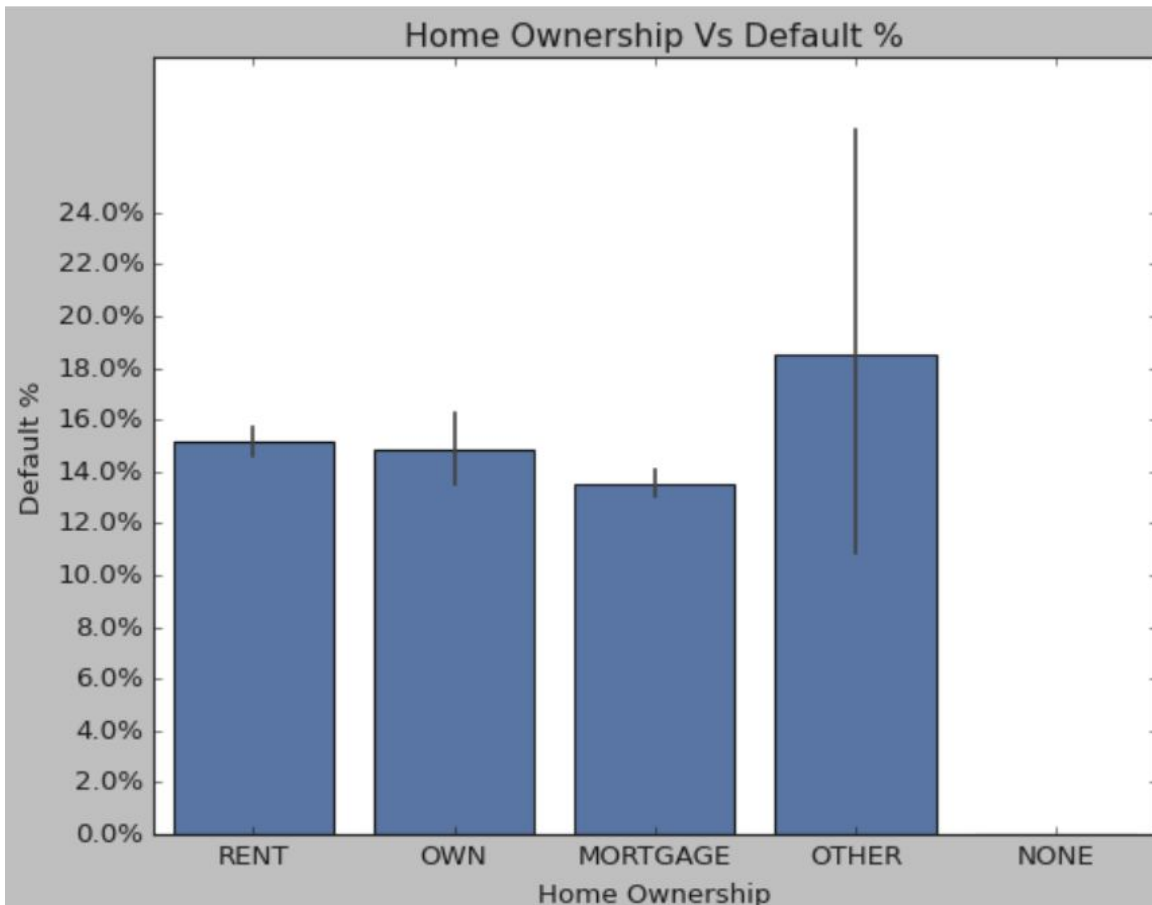
Most defaulters those borrowers who has taken a greater loan amount. All loan amount greater than **25K** has more than **37%** of default rate.

Term of Loan Vs Default %

Term vs Loan Default Rate

Most default is happening on loans which have a term of 5 years - **25%**.

Home Ownership vs Loan Default Rate

There is not much impact of the home ownership status on the loan payback.

Driving Indicators of Default

1. **Purpose of the loan** -
   a. If the loan is for small business, then it needs to be approved after proper scrutiny.
   b. In terms of loan amount, there is $190M dispersed in debt consolidation, out of which 15% has defaulted. So more stringent measures needed for such loans
2. **Verification Process** - The Lending club needs to assess the methods adopted to evaluate the income document submitted by the borrower.
3. Higher **Loan amounts** which are provided at higher **interest rates** are defaulting the most. Hence to reduce risk, it is better to increase the number of loans provided below $20K.
4. Loans allocated in **Grade** A, B, C can be further increased. Additional scrutiny and validation needs to be introduced for loans in Grade F and G.
5. Status of **Home ownership** is not having any impact on the ability of a borrower to pay back the loan.
6. The **Term** of the loan is having an impact. Upto 25% of Loans provided for 5 years terms are defaulting.
7. Most loans(both dollar amount and number) are provided in the **state** of **California** and so are the number of default cases. Hence, additional measure need to be in place for loans approved in the state.