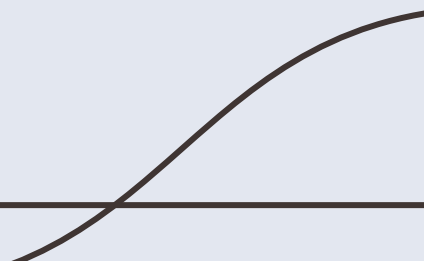




# Lending Club Case Study

A EDA Group Study By Bhavish Arora and Dolly Chakraborty



# Problem Statement

This analysis aims to address the critical issue of loan default for the lending company. Specifically, we seek to identify patterns and factors that contribute to the likelihood of a loan applicant defaulting on their repayment obligations. By understanding these underlying trends, we can develop more effective risk assessment to minimize financial losses for lending company

The primary objective of this study is to:

- **Identify key predictors of loan default:** Determine which customer and loan attributes have the strongest correlation with default.
- **Provide actionable insights:** Offer recommendations to lending institutions to mitigate credit risk and improve their decision-making processes.

**Data Source:** Historical data of loan applicants and their repayment statuses.

**Approach:** Minimize credit loss by identifying risky applicants through EDA.

# Process Flow

**01**

## Data Understanding

Understanding the imperfections in data

**02**

## Data Cleaning

Removing non relevant columns and filling missing values

**03**

## Univariate Analysis

Analysing the distribution of relevant columns

**04**

## Bivariate Analysis

Analysing relationship of various variables with defaults

**05**

## Multivariate Analysis

Analysing correlation between different numerical variables

**06**

## Recommendations

Our Findings and Recommendations

# Data Understanding

**Dataset Overview:** The dataset “Loan.csv” Contains loan application data from past applicants.

**Total Rows** = 39717 and **Columns** = 111

**Key Variables:**

- **Loan Status:** The repayment status of the loan (Fully Paid, Charged Off, Current).
- **Loan Amount:** The amount of money borrowed.
- **Interest Rate** (int\_rate): The annual interest rate for the loan.
- **Grade:** The borrower's credit grade (A-G), a measure of creditworthiness.
- **Annual Income** (annual\_inc): The borrower's self-reported annual income.
- **Debt-to-Income Ratio** (dti): A ratio measuring the borrower's debt compared to income.
- **Loan Purpose:** The stated reason for taking the loan (e.g., debt consolidation, small business).
- **Zip Code** :The first 3 numbers of the zip code provided by the borrower in the loan application.
- **Home Ownership** :The home ownership status provided by the borrower during registration.
- **Verification Status:**Indicates if income was verified by LC, not verified, or if the income source was verified

**Target Variable: Loan Status** (loan\_status): Identifying whether the loan was fully paid, charged off (defaulted), or still being paid (current).

**Missing Values:** Some columns had missing values, particularly employment-related data and financial ratios, which will be cleaned during the data preparation step.

# Data Cleaning

## Handling Missing Values:

- Columns having more than **75% missing values** were dropped. After dropping a total of **53 columns** were remaining
- Based on Columns, Remaining missing values were imputed with median, mode and “unknown”

## Handling Data Types

- **Interest Rate:** Converted the (int\_rate) column from string format (e.g., '13.56%') to numerical and stripping '%' sign for accurate analysis.
- **Loan Term:** Cleaned the (term) column to extract the loan term as a numerical value (e.g., '36 months' → 36.00).
- **revol\_util:** Converted the (revol\_util) column from string format (e.g., '13.56%') to numerical and stripping '%' sign for accurate analysis.

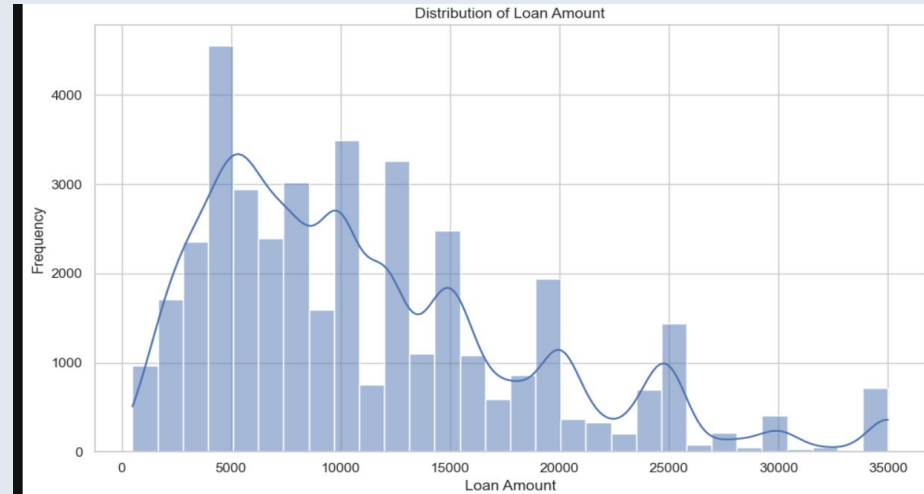
The slide features a light blue background with two horizontal dark blue lines. The top line has a curved end on the left side, and the bottom line has a curved end on the right side.

# UNIVARIATE ANALYSIS

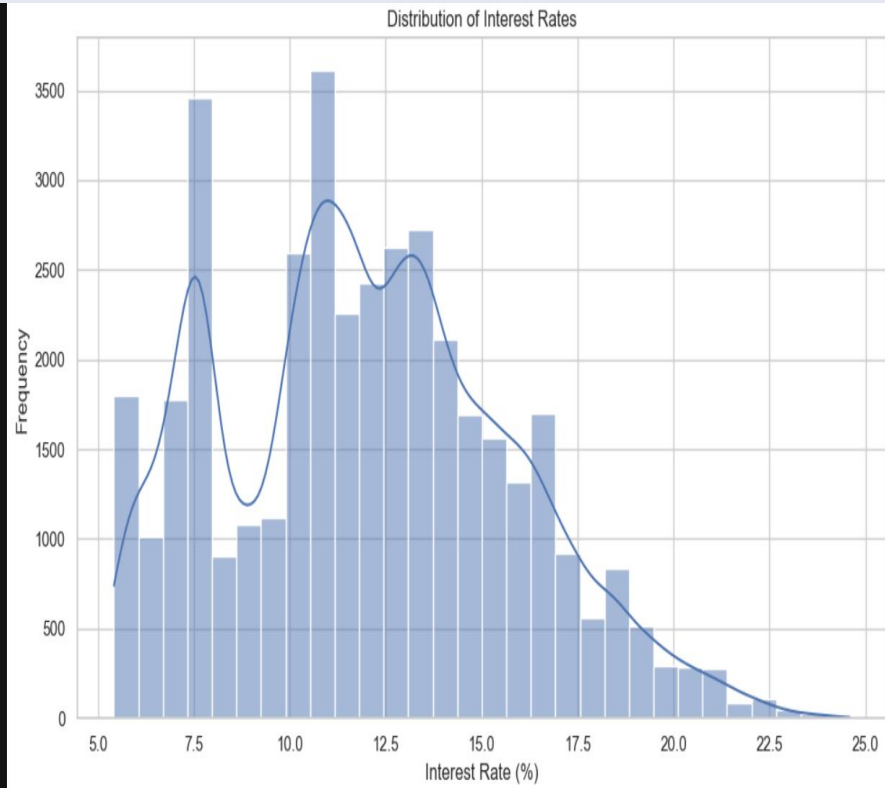
# Distribution of loan amount

## Observations:

1. Median loan amount application is 10K.
2. Max count of loan amount ranges between 5000 to 10000 and as the loan amount increases, the count decreases.



# Distribution of Interest rates

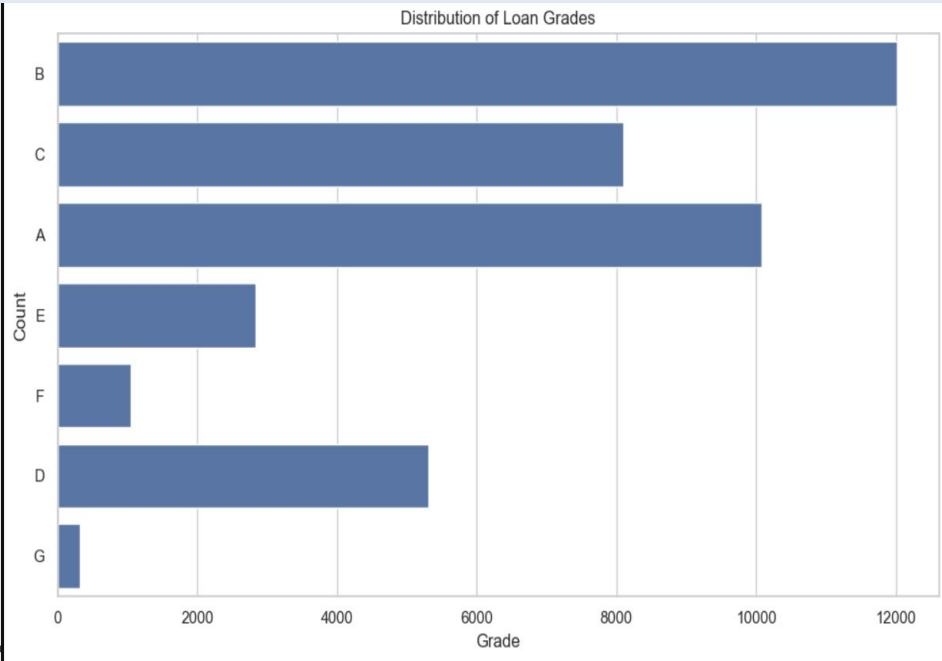


## Observations

Maximum interest rate count is between 10 to 12.5%



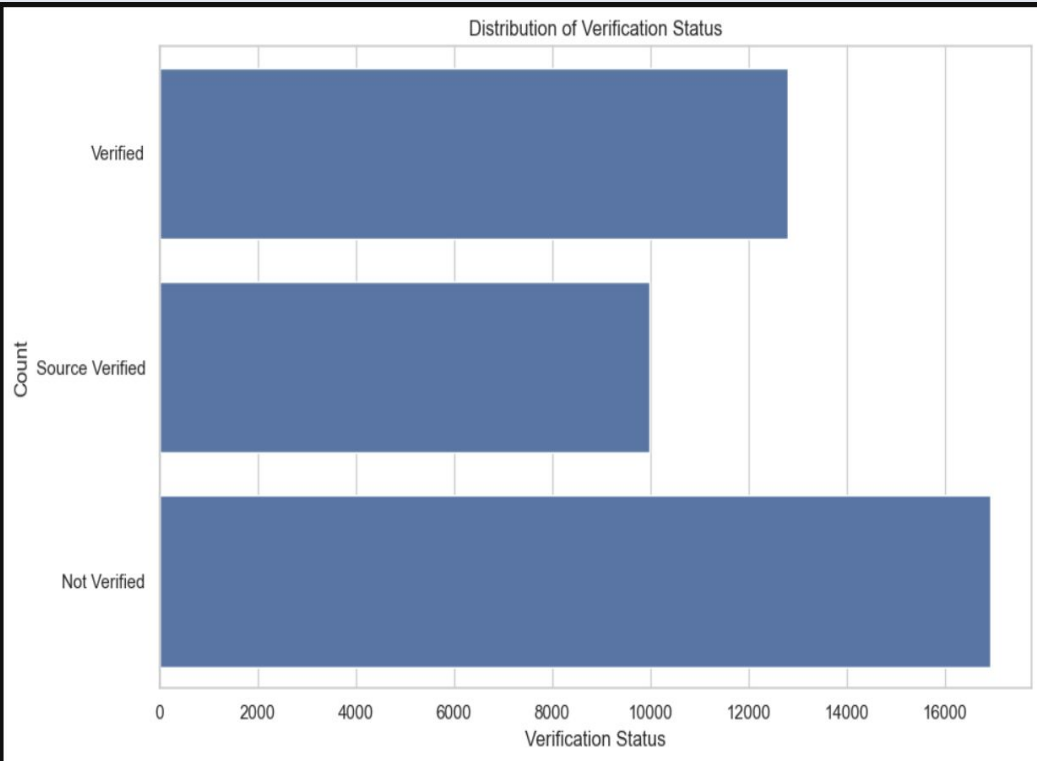
# Distribution of loan grades



## Observations

Most of the loan accounts comprise of the top 4 grades A, B, C & D

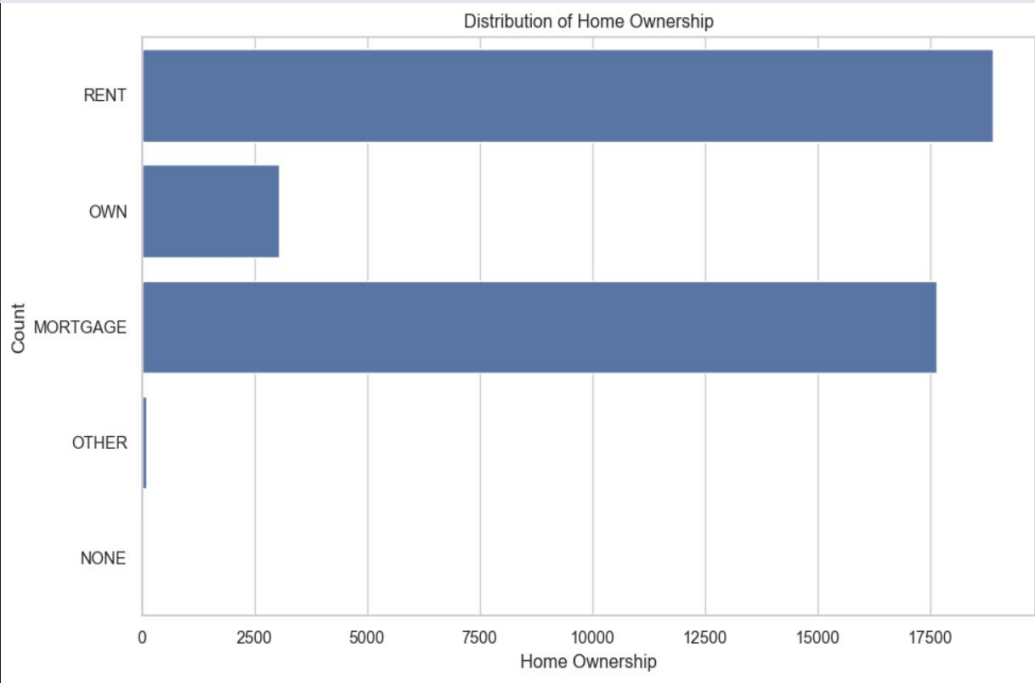
# Distribution of Verification Status



## Observations

A large chunk of the loan applications appear to be in the unverified status and represents risk of posing data quality and reliability of loan data to accurately evaluate members ability to pay up the loan

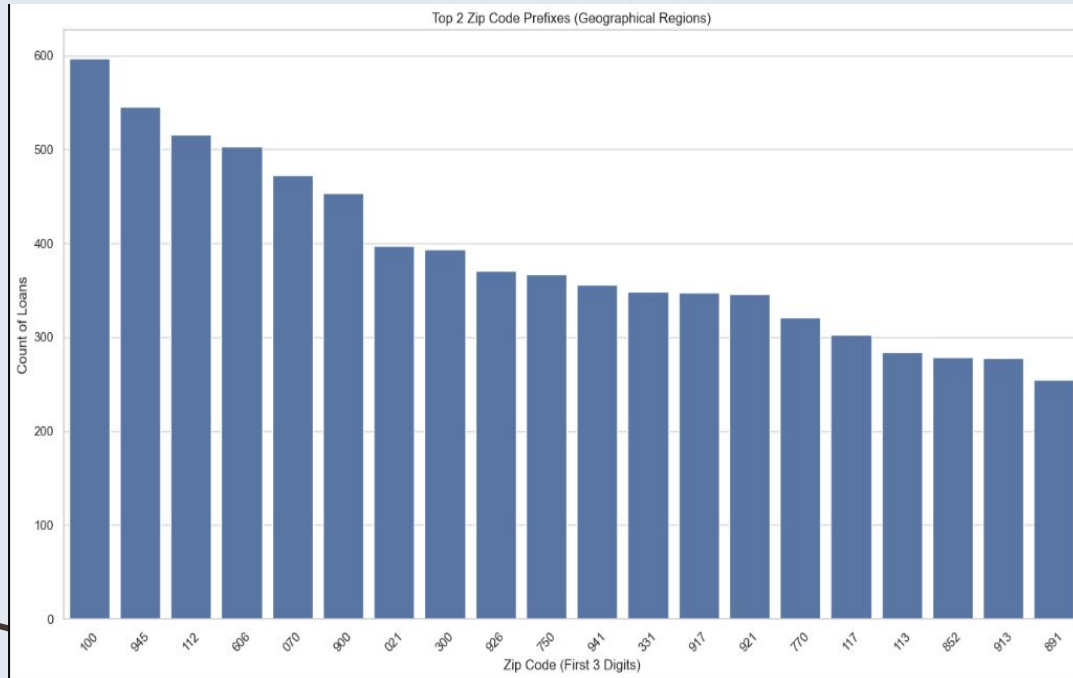
# Distribution of home ownership



## Observations

About 90% of the loan applicants either have a mortgage or are on rent and about 6% are home owners

# Distribution of Zip Code Prefixes



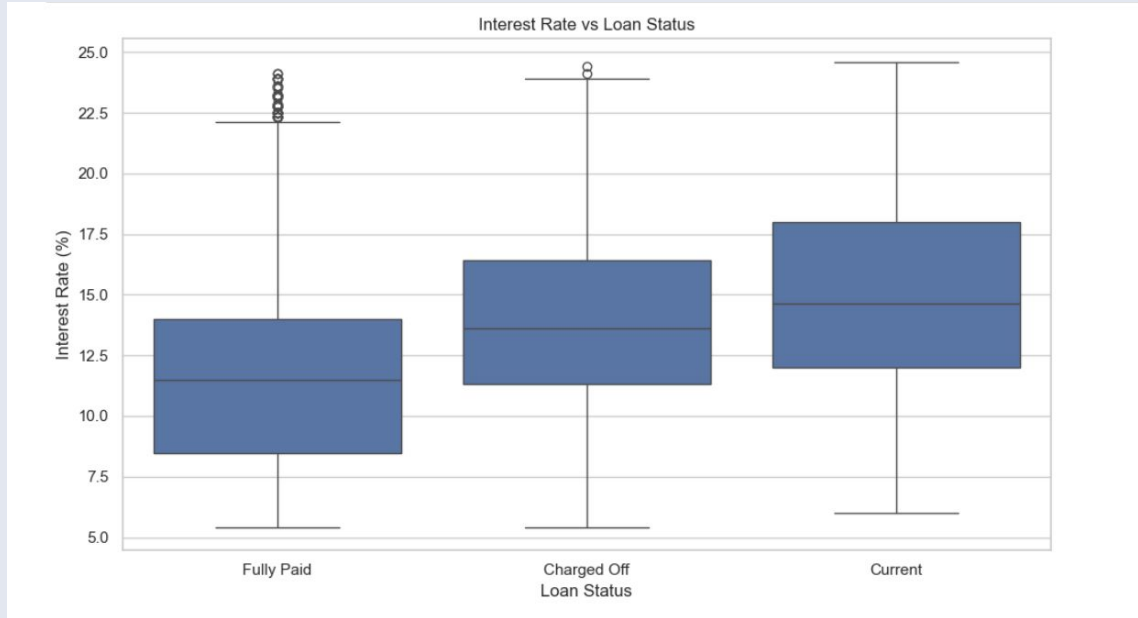
## Observations

Specific zip codes tend to have higher consumption. This might be because of the difference in income levels.



# Bivariate Analysis

# Interest Rate VS Loan Status



## Higher Interest

Higher the Interest, Higher is the tendency to default

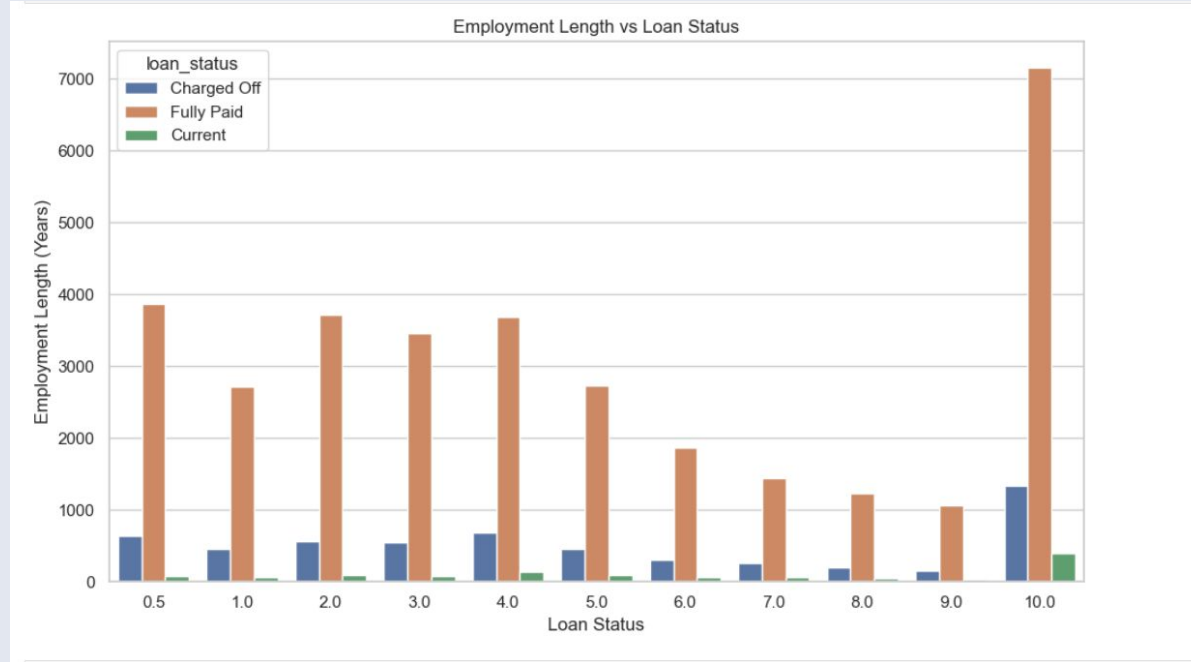
## Fully Paid

Most of Fully Paid Loans are between 7.5% and 15%, median being around 11 %

## Charged off

Clear indication as we go higher interest, above 10% and below 17.5%. More is Tendency to default. Median around 14%

# Employment Length VS Loan Status



## Increasing Length

As the Work- Ex Increases, Tendency to default decreases until 10 years after that there is a increase defaults.

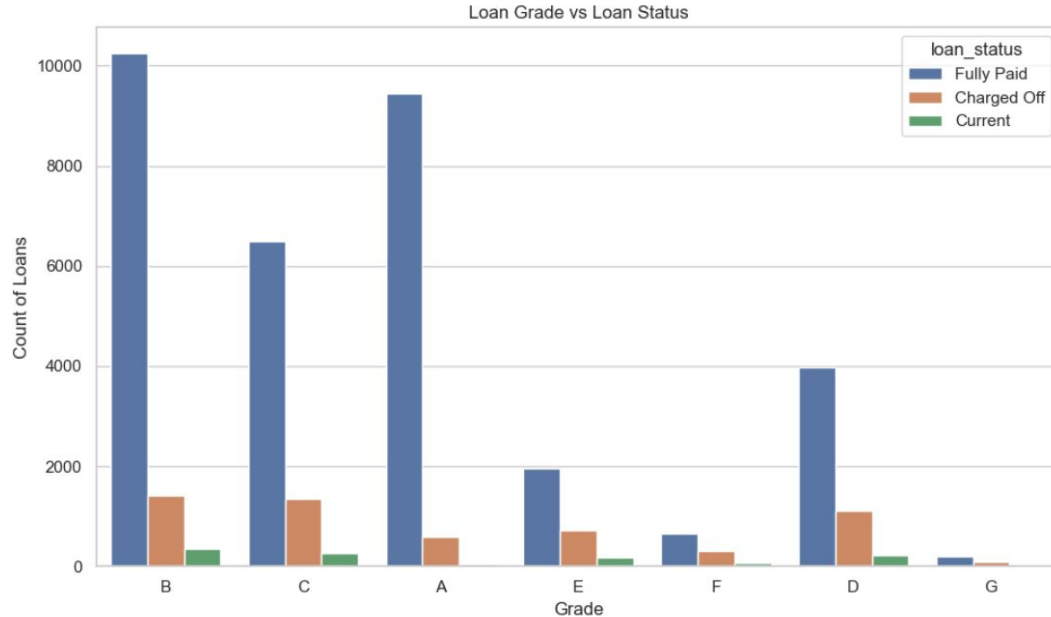
### Before 10 years

Amounts of defaults decreases steadily until 10 years of work ex

### After 10 years

After 10 years of work ex, there is increase in defaults

# Loan Grade VS Loan Status



## High Risk

Loan Grade C, D, E, F possess High risk of default. E being the Riskiest and B being the Safest

## Low Risk

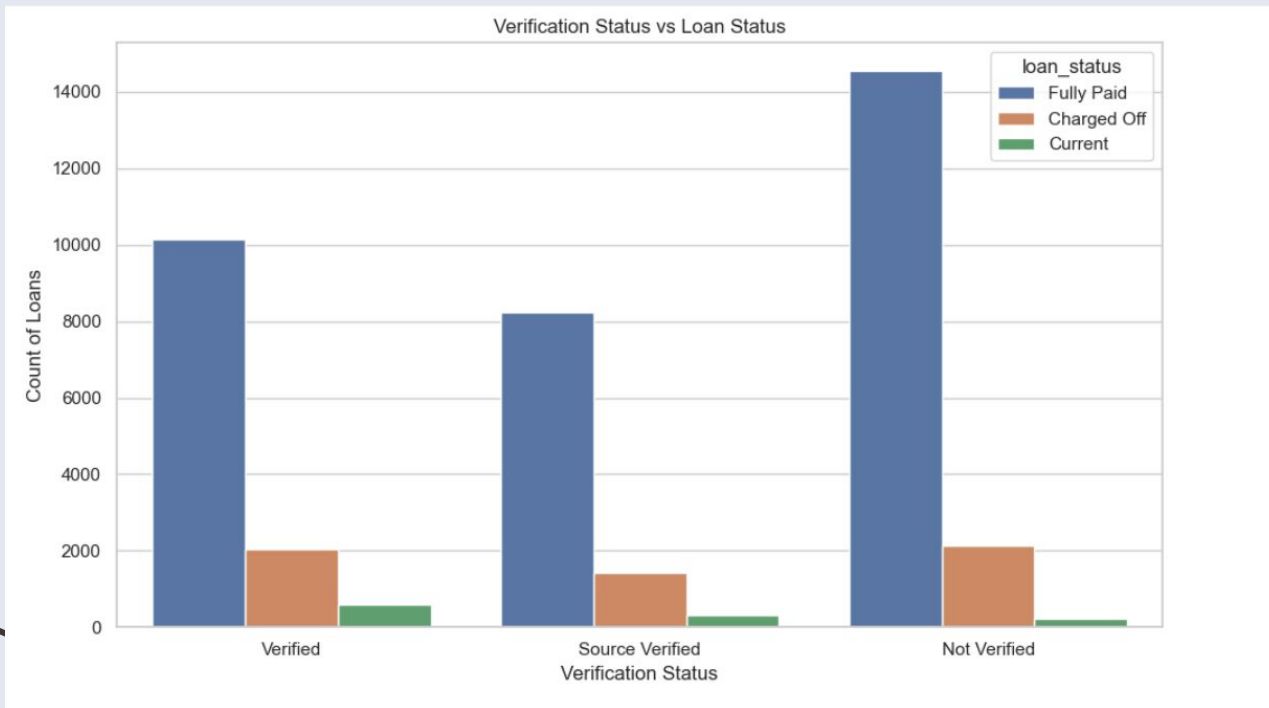
Loan grades A, B, G have Low Risk of Default out of which A being Safest.

## Grade A

Grade A outperforms every other grade to be the safest with least defaults and high fully paid



# Verification Status VS Loan Status



## Verification Failure

Verified and Unverified Loans have same number of Defaults.

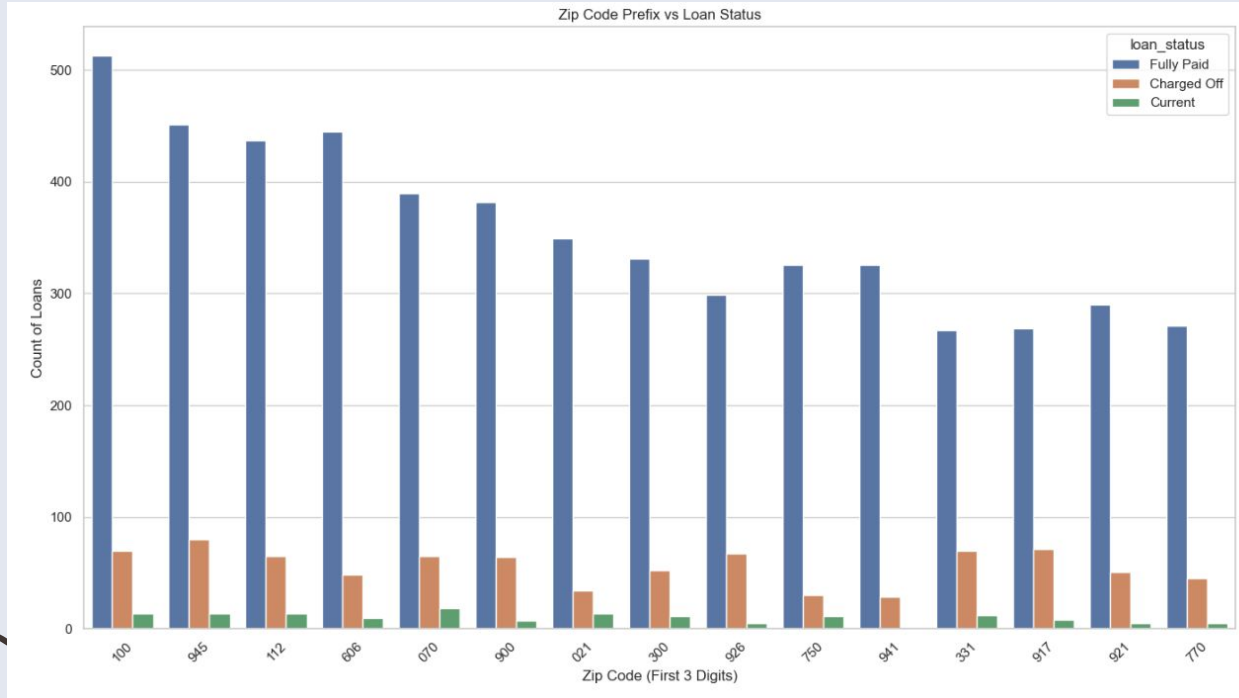
## Verified

Source Verified have less amount of defaults than the verified

## Not Verified

Not Verified Loans have same defaults as verified, but if we see ratio of fully paid to charged off, unverified seem to take a lead

# Zip Code VS Loan Status



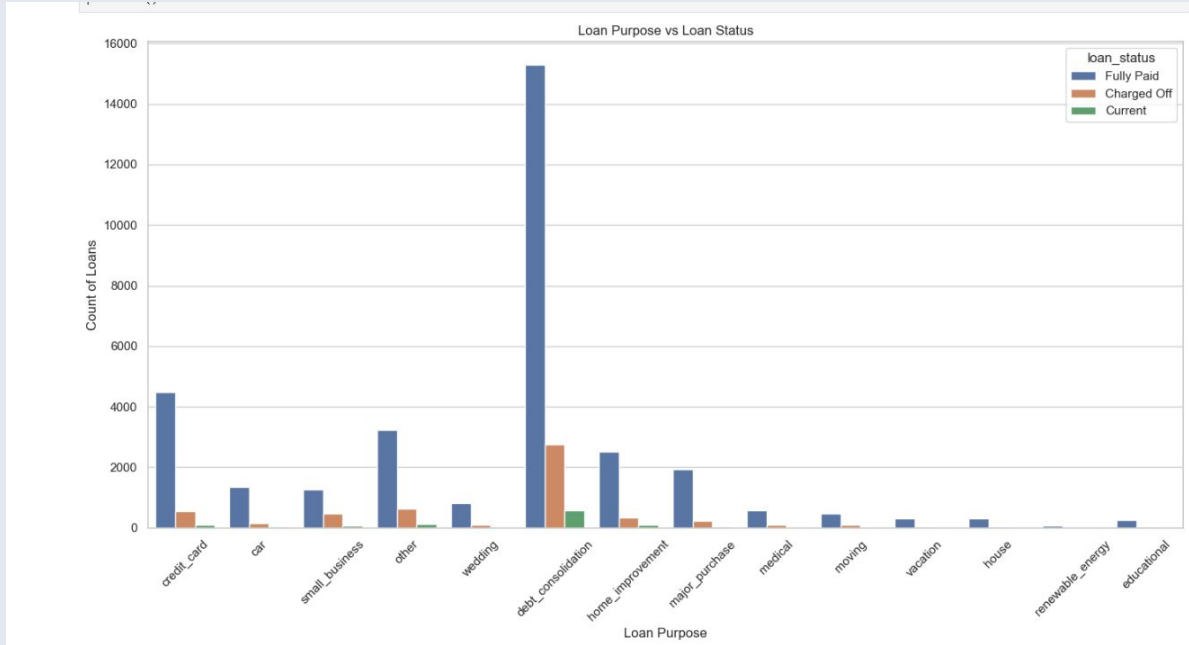
## Safe Zip Codes

606,021,750,941 Zip codes fall under safe zip codes having less defaults in relation to the fully paid.

## Unsafe Zip Codes

917,331,921,945 are few examples of unsafe /risky zip codes

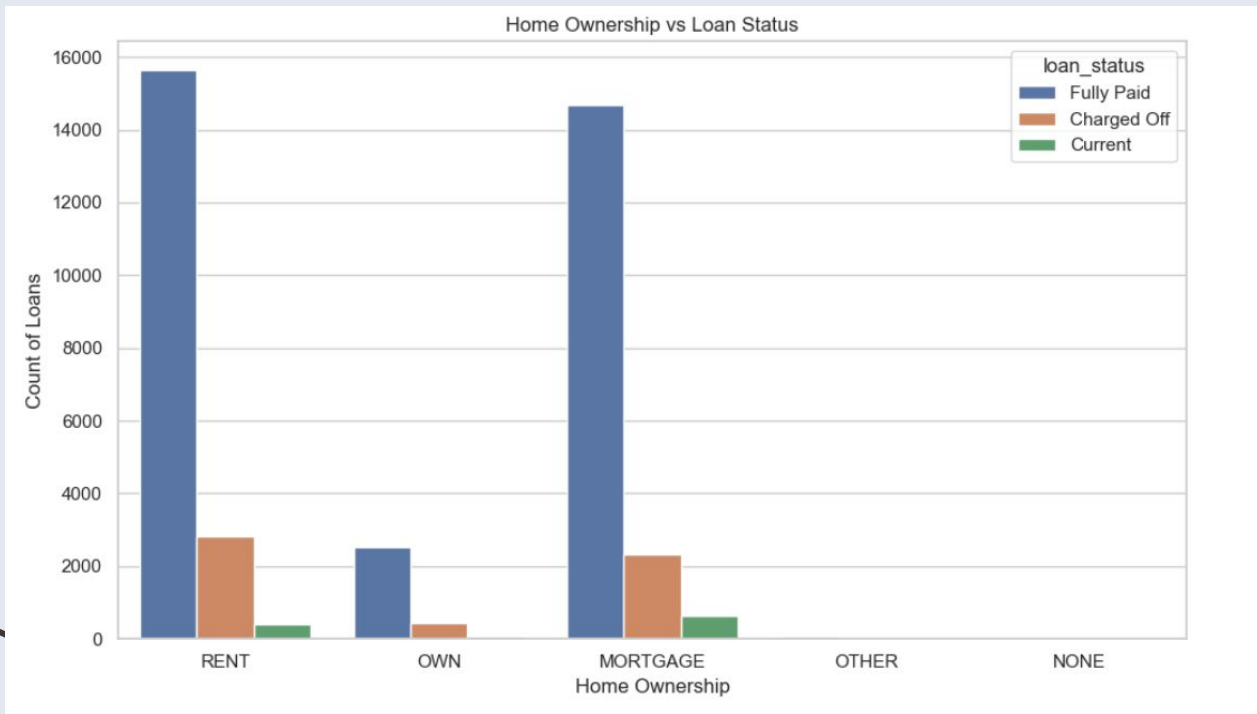
# Loan Purpose VS Loan Status



## High Risk Purposes

Loan Purpose Like Credit Card, Debt Consolidation, Small Business have High Risk of Defaults

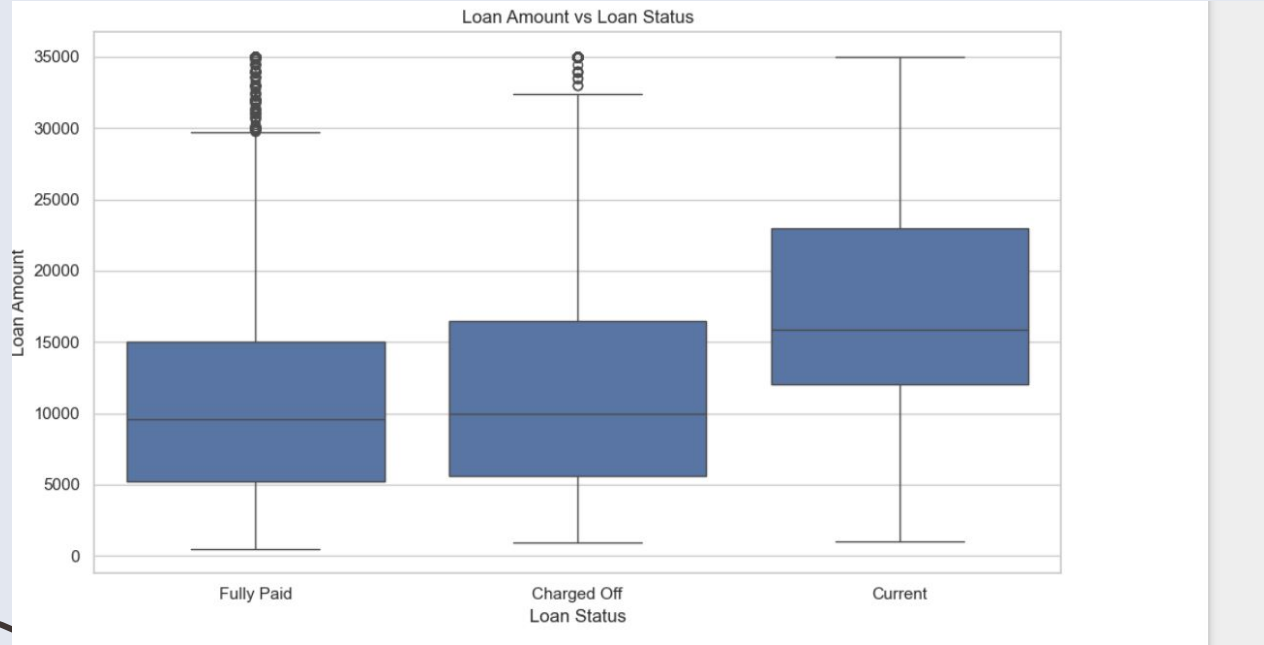
# Home Ownership VS Loan Status



## High Risk

Rented and Mortgage home owners are more likely to default

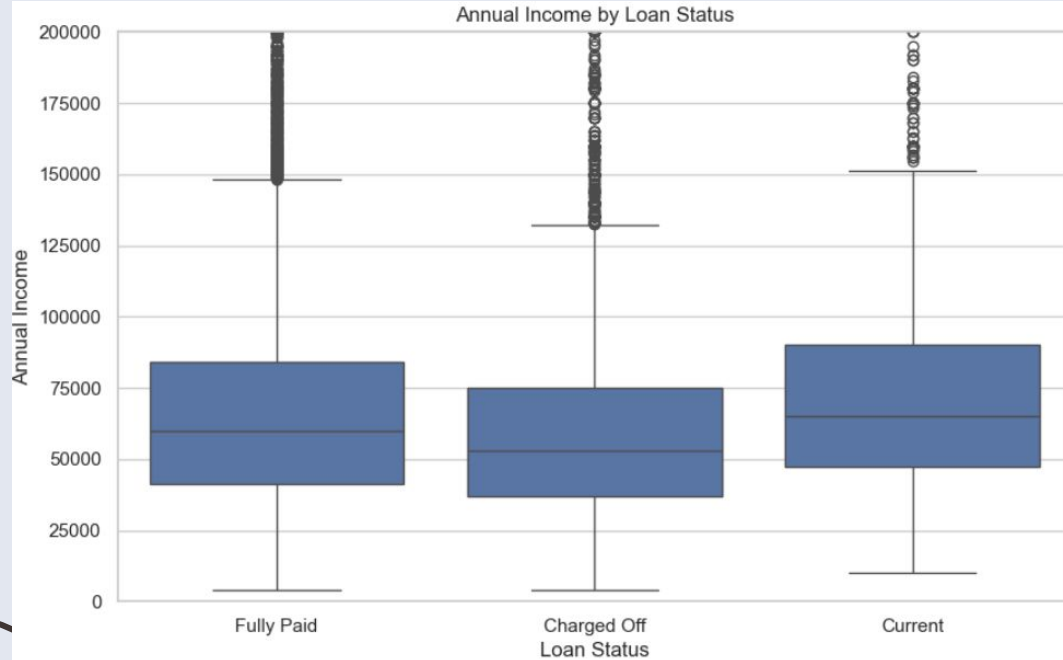
# Loan Amount VS Loan Status



## Weak Relation

Loan amount does not seem to affect probability of default. But after 15000 more is probability to default

# Annual Income VS Loan Status



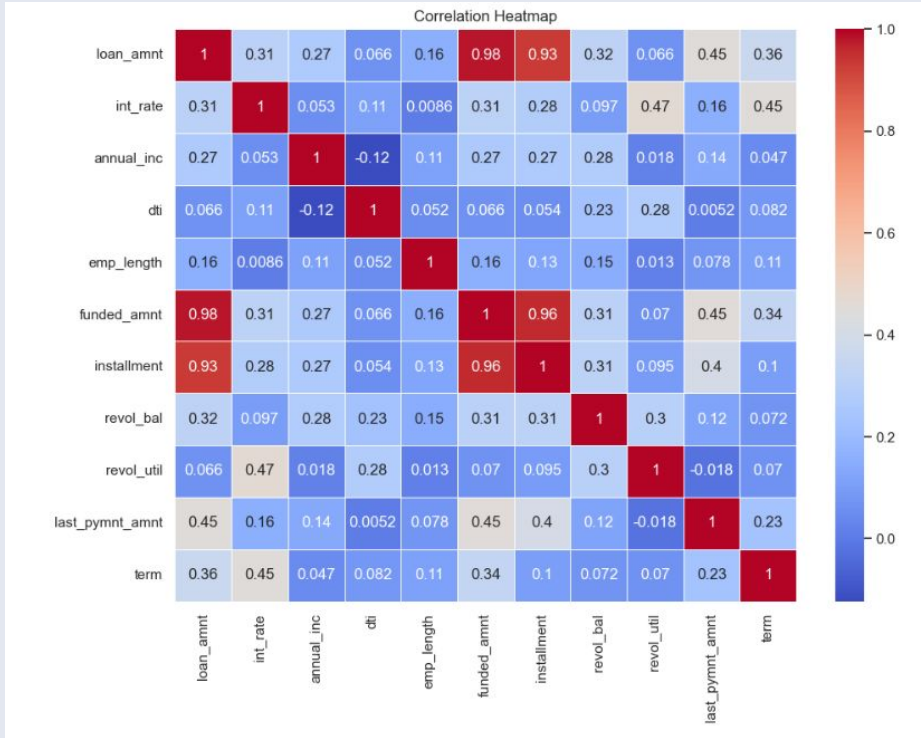
## Low income

With Decrease in Income level  
Probability of Default increases..



# Multivariate Analysis

# Correlation Heatmap



## Summary

**Strong positive correlations:** Loan amount and funded amount have a very strong positive correlation, as expected.

**Moderate positive correlations:** Interest rate and term have a moderate positive correlation, indicating that longer loans tend to have higher interest rates.

**Weak positive correlations:** There are weak positive correlations between loan amount and annual income, as well as between loan amount and installment.

**Negative correlation:** Debt-to-income ratio (dti) has a weak negative correlation with annual income, suggesting that people with higher incomes may have lower dti ratios.

**No significant correlations:** There are no significant correlations between revolving balance and other variables, indicating that revolving balance is not strongly related to other factors in the data.



# Insights & Recommendations

## Insights :

**Interest Rates:** Applicants with interest rates above 14% are much more likely to default.

**Grade and Defaults:** Borrowers with credit grades D, C, E, F have significantly higher default rates, making them riskier.

**Loan Purpose:** Borrowers taking loans for debt consolidation, small business, credit card are more likely to default.

**Employment Length:** Shorter employment tenure (less than 1 year) or after 10 years increases the likelihood of default.

**Verification:** Verification Does not seem to have effect on the Likelihood of the Default, Only Source verified seem to be having results.

## Recommendations:

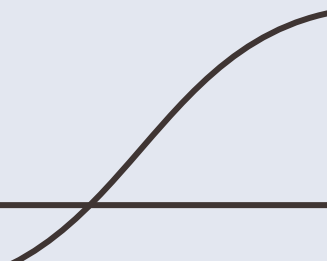
1. Ensure verification Process if Stringent and is carried out properly.
2. Engage more with employees with experience over 1 year and lesser than 10 to expand the pool by learning about their needs
3. Scrutinize purpose and public bankruptcy record of the loan applicants to prevent risk of default.
4. Keep the Interest Rates in Favourable range of 7.5 to 14%.
5. Focus on Homeowners
6. Be cautious with high default zip codes



---

# Thank You

Submitted By Bhavish Arora  
And Dolly Chakraborty



---