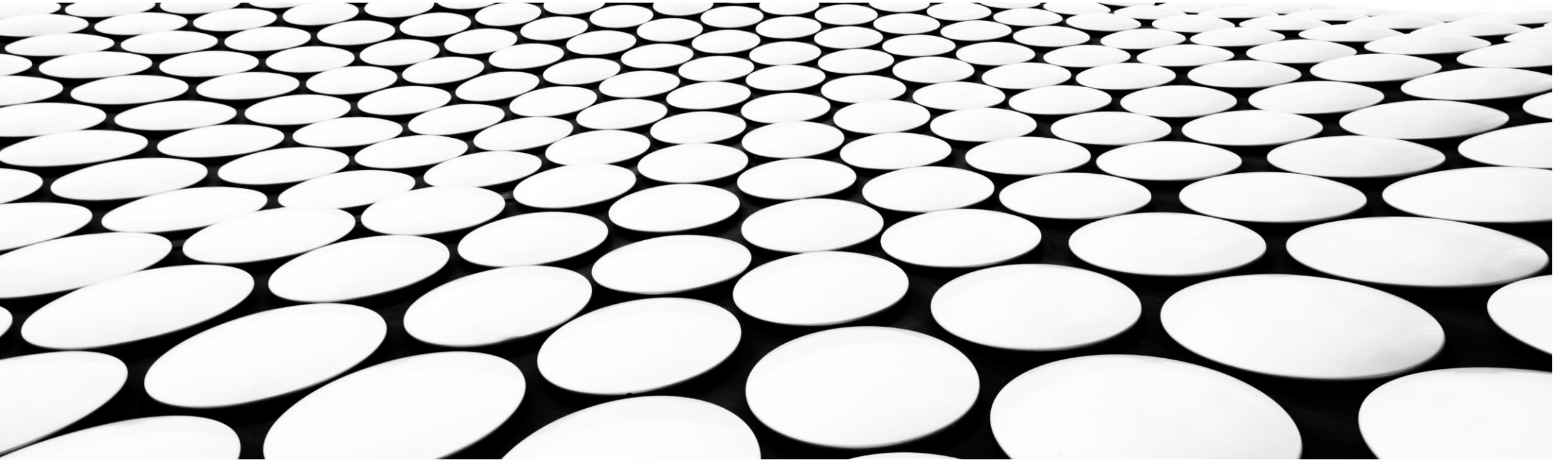# LENDING CLUB CASE STUDY

- NITIN KATIYAR & RAHUL NANWANI

AGENDA

- Introduction

- Problem Statement

- Approach

- Conclusion

OBJECTIVES

Borrowers can easily access lower interest rate loans through a fast online interface.

understand the driving factors (or driver variables) behind loan default

The company can utilize this knowledge for its portfolio and risk assessment.

Background:

- Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

Problem Statement:

- 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:

  - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

  - 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

APPROACH

| Data Sourcing	Data Cleaning	Data Analysis	Conclusion
We already have the dataset provided. We need to understand this dataset that has 39717 rows and 111 columns.	We removed irrelevant data, handle the missing values, standardize and clean the data and handled the outliers.	We need to conduct data analysis using univariate, bivariate analysis, deriving new metrics from the data, etc.	From the data we have analysed making decisions would be easier just by looking at the plots.

UNIVARIATE ANALYSIS

Categorical Variables

- There are more high grade loans in compare to low grade loans.

- People with own home tend to take less loans in compare to people living in Rent or Mortgage.
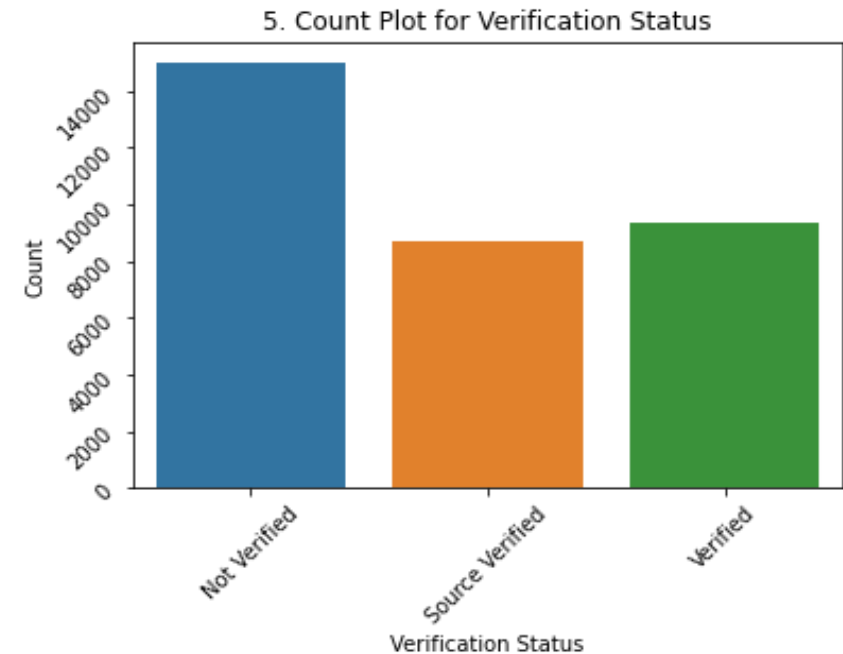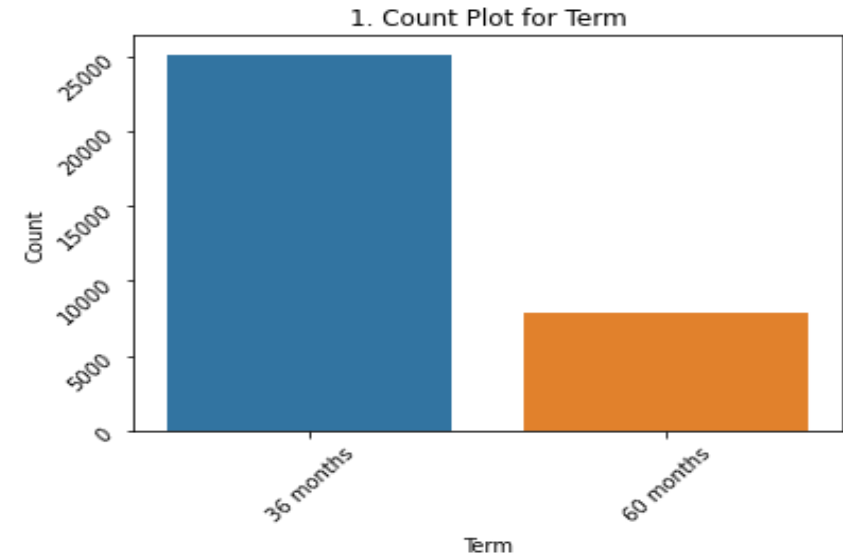

2. Count Plot for Grade
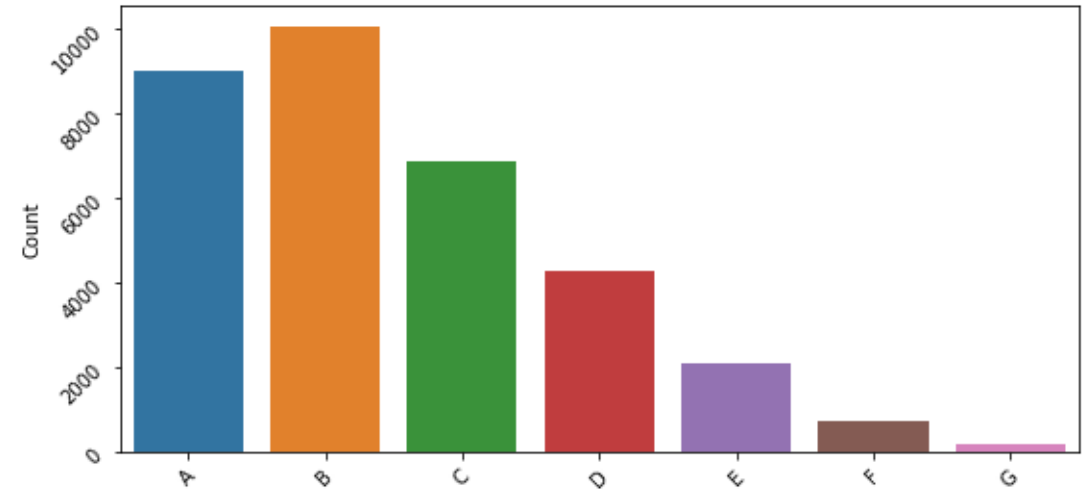
4. Count Plot for Home Ownership
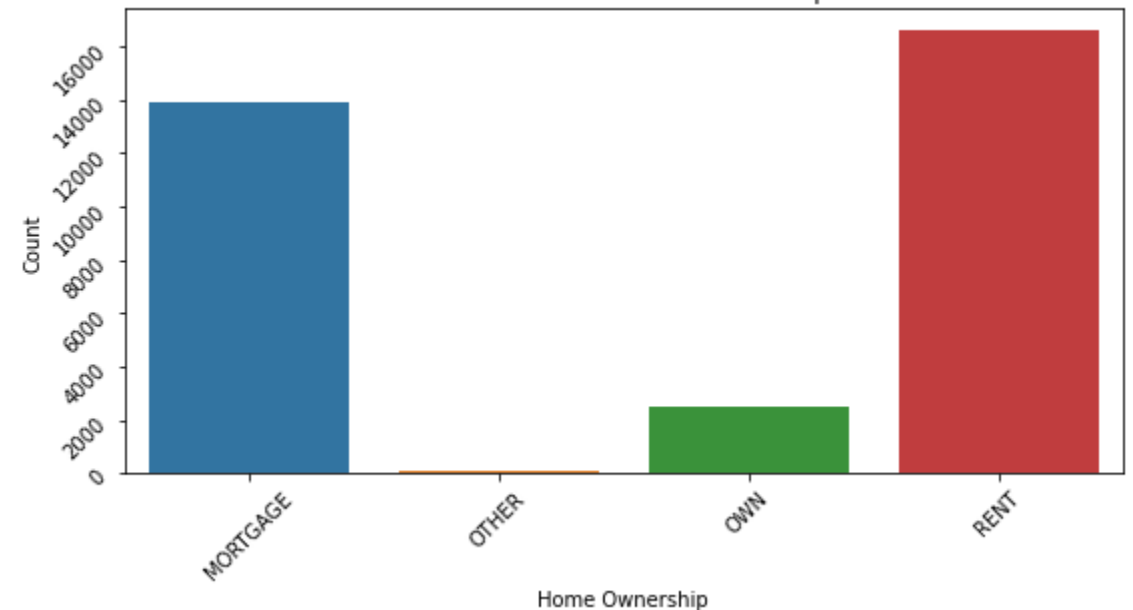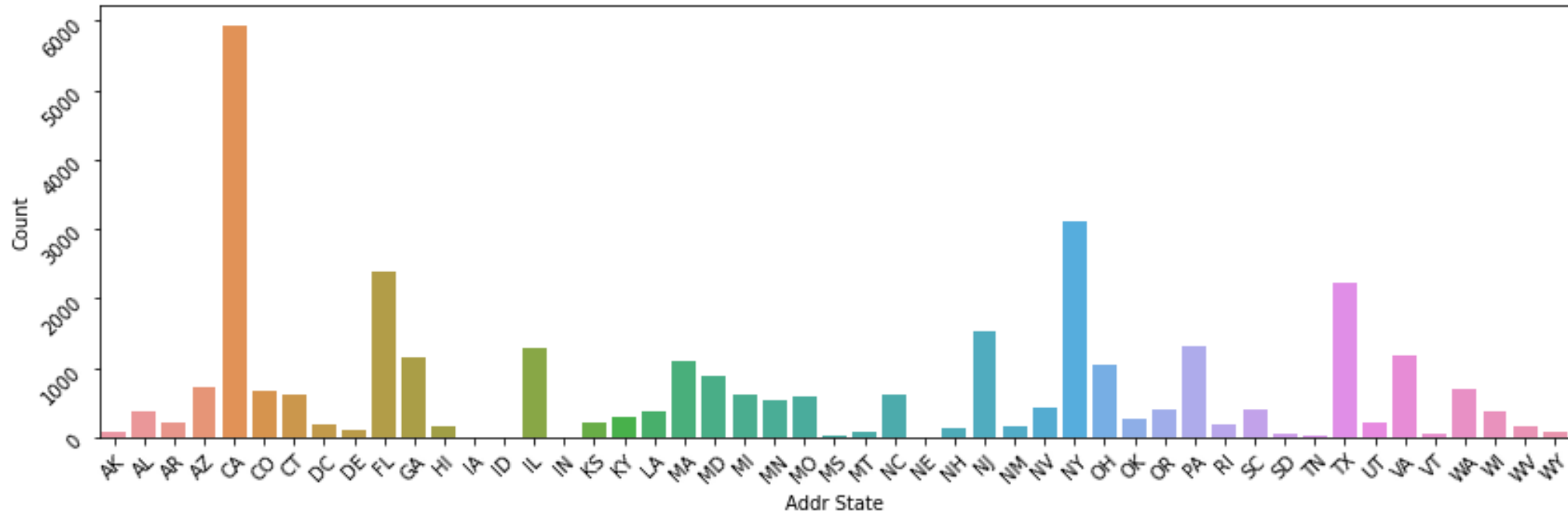
UNIVARIATE ANALYSIS

Categorical Variables

- Most loans are of 36 Months

- The percentage of non-verified loans is also very high. We need to analyse if that contributes to loan default.

UNIVARIATE ANALYSIS
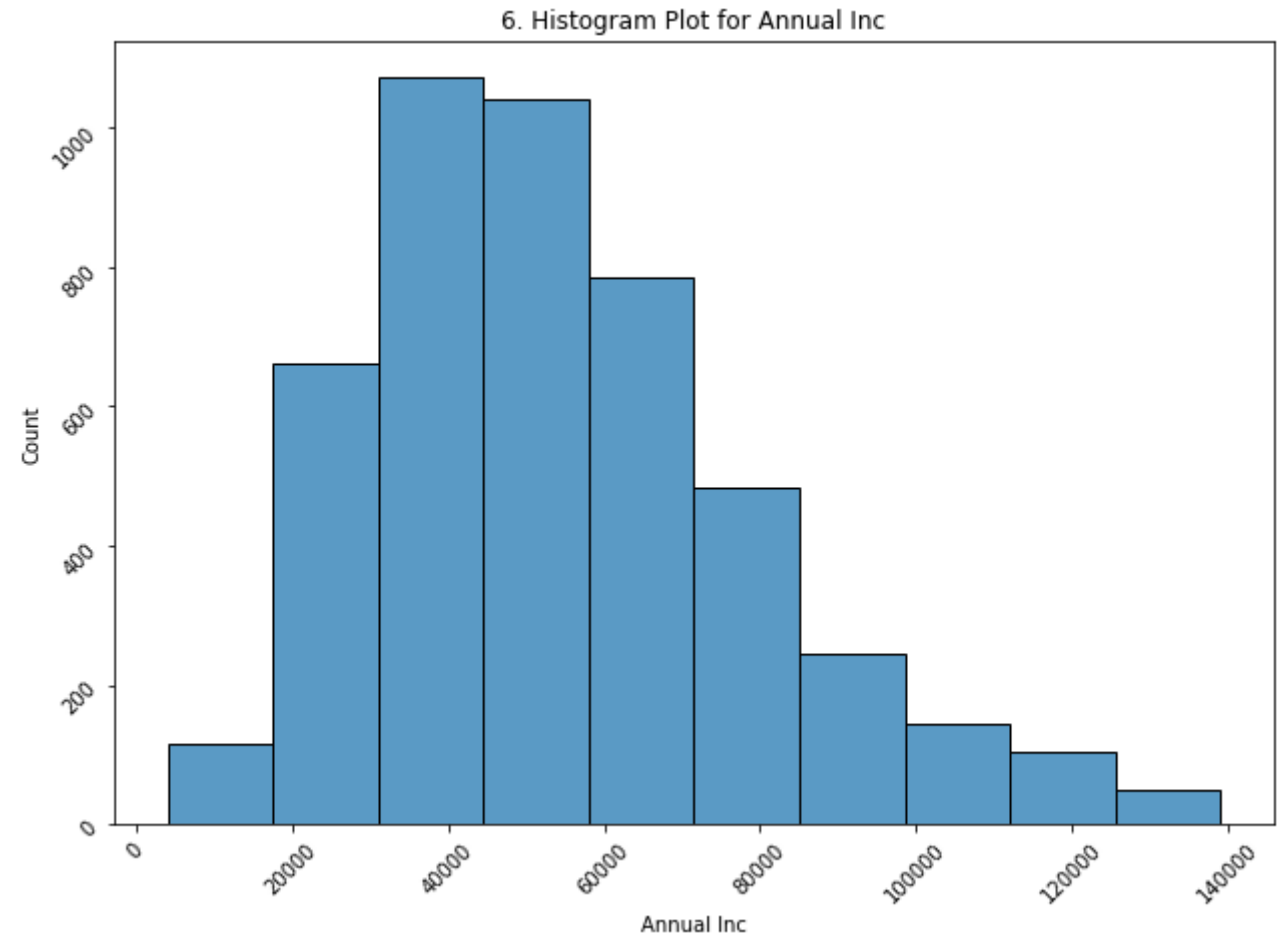

7. Count Plot for Addr State

- People from State CA have very high number of loans in contrast to IA, ID or TN state
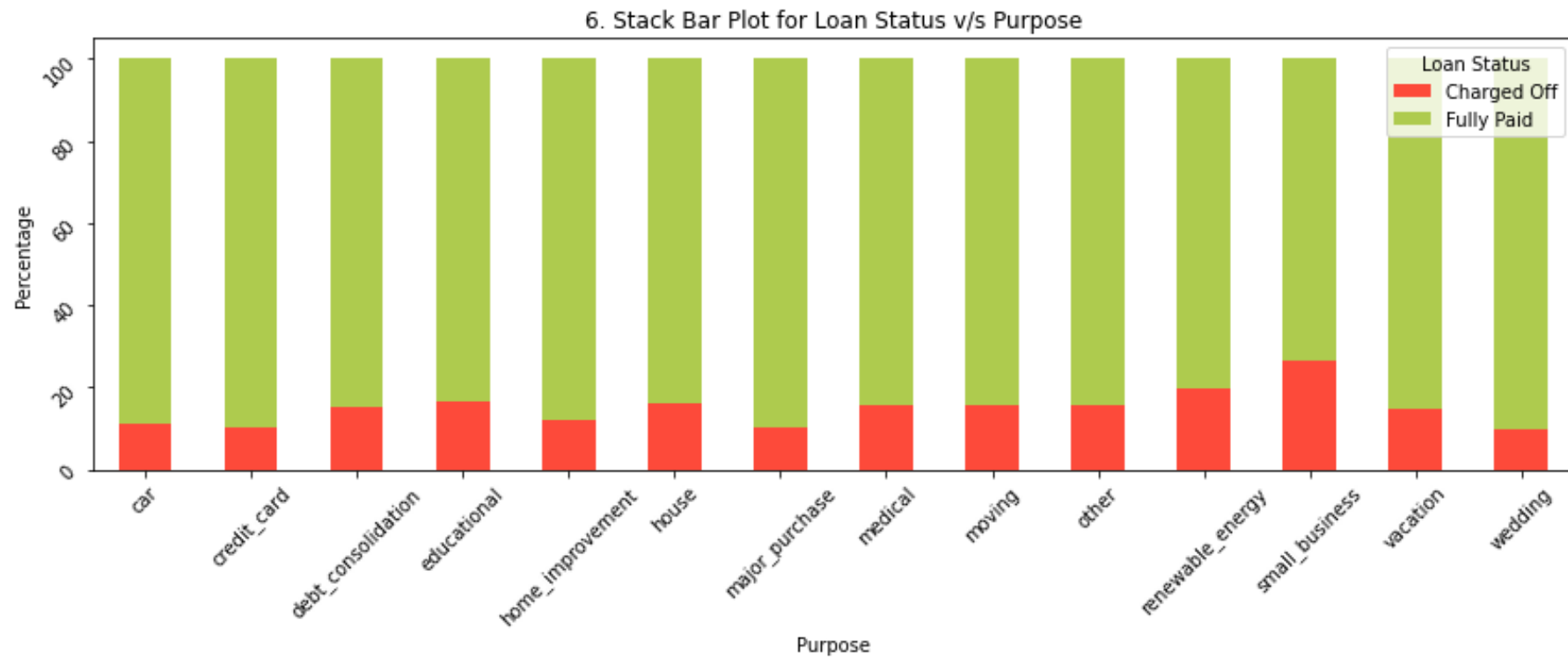
UNIVARIATE ANALYSIS

Numerical variables

- Most loan takers have income between 40k to 60k

- Defaulted loans are also in this range.

6. Histogram Plot for Annual Inc

BIVARIATE ANALYSIS
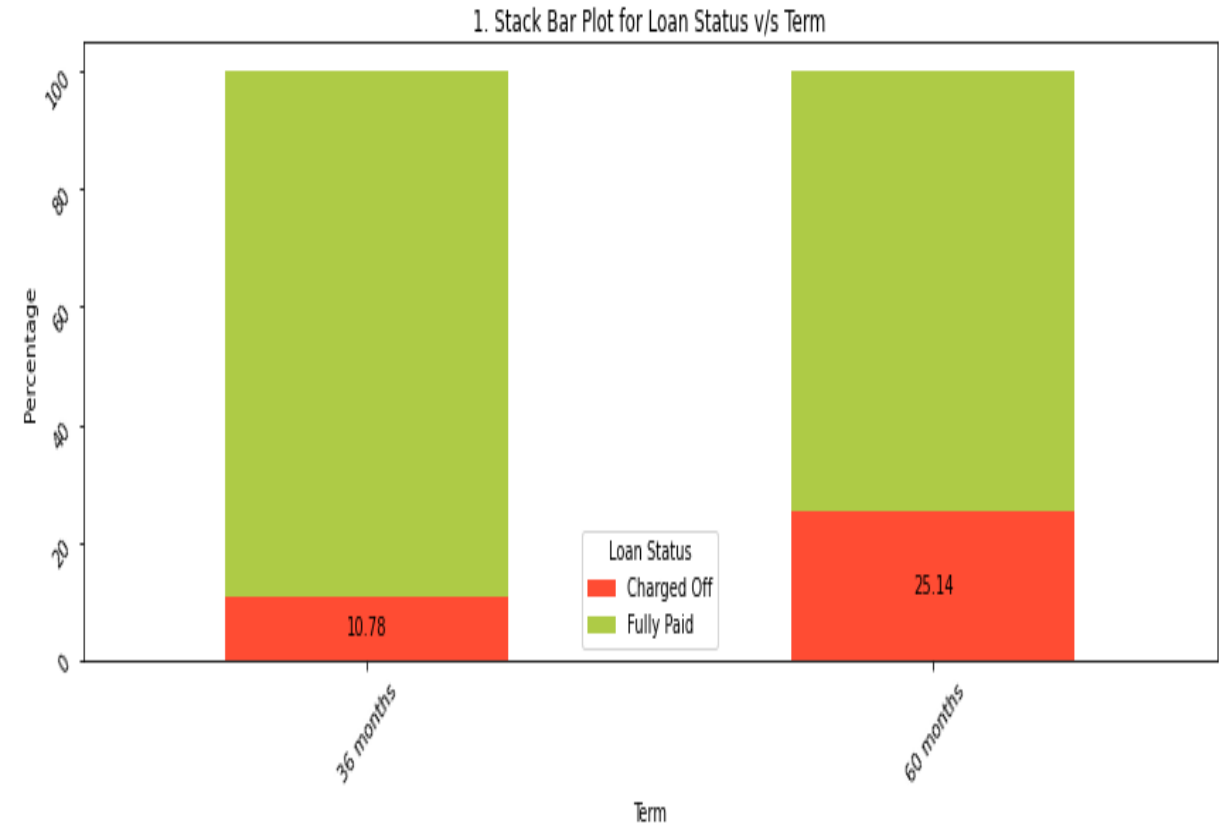

6. Stack Bar Plot for Loan Status v/s Purpose

- Loans taken for Small business has highest percentage of default so they are most likely to default.

# BIVARIATE ANALYSIS

Loan status vs Term Categorical Variables

- Loans taken for 60 months are almost 2.5x more likely to default as compared to loans taken for 36 months.

1. Stack Bar Plot for Loan Status v/s Term

BIVARIATE ANALYSIS

Loan status vs Categorical Variables

- It can be clearly seen that loans with high interest rates are more likely to get charged off.

- Loans with low interest are more likely to get fully paid.

4. Box Plot for Loan Status v/s Int Rate
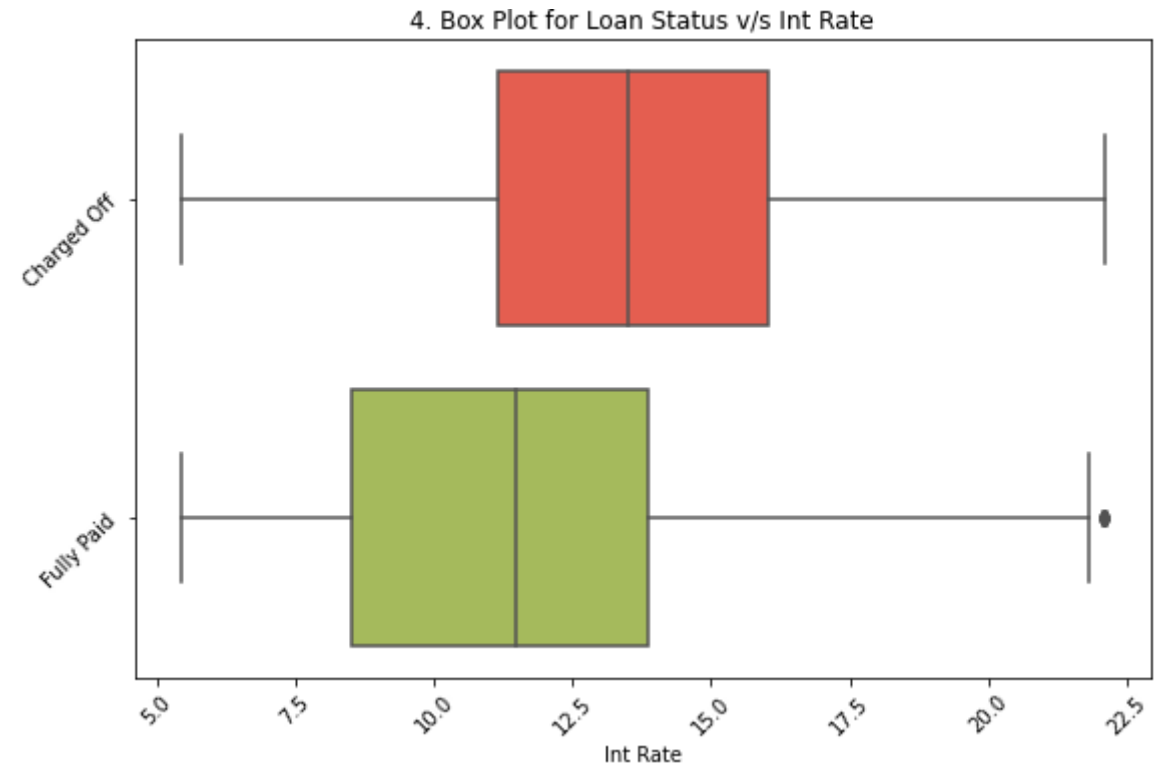
CONCLUSION

- Loan for Small business are most likely to be default.

- Term has the most impact on default rate, loan term being 60 months are 2.5x more likely to default than 36 months.

- Loan with highest interest rate in all buckets are more likely to charge off.

- Borrower within grade E,F,G are more likely to charge off.

- Higher annual income may very slightly reduce the risk of defaulting the loan.