Lending Club Case Study

Prepared By :
Group Facilitator – Shubham Gupta
Team Member – Abhijit Banerjee

Introduction

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

You work for a consumer finance company which specializes in lending various types of loans to urban customers. 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:

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
2. 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.

This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface. Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss).

Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

Objective of the Project

The data contains the information about past loan applicants and whether they ‘defaulted’ or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

If one is able to **identify these risky loan applicants**, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Analysis Approach

1. Programming Language used – Python

2. Platform used – Jupyter Notebook

3. Libraries used:

    Numpy for Linear Algebra
    Pandas for Data Processing
    Matplotlib for Data Visualization
    Seaborn for Data Visualization

After loading the data it is observed that there are in total **39717 Rows** & **111 Columns** in the Dataset

Data Cleaning

- Verifying the presence of duplicate rows in the dataset based on the Column "id“.
- Filtering out the rows with status “current” from the Column "loan_status“ since current status data cannot be used to find the driving factors behind loan default
- Checking and dropping the columns having :
  - all values as NULL from the Dataset.
  - having more than equal to 30% of NULL values from the dataset
- Dropping out those columns that has only 1 unique value.
- Dropping irrelevant Columns that could not be used during data analyzation
- Performing Data imputation for columns having NULL values less than 30%
- Standardising required columns by removing special characters , rounding off to 2 decimal places and updating the data format
- Creating derived metrics for column “issue_d”
- Checking and removing outliers from the Dataset

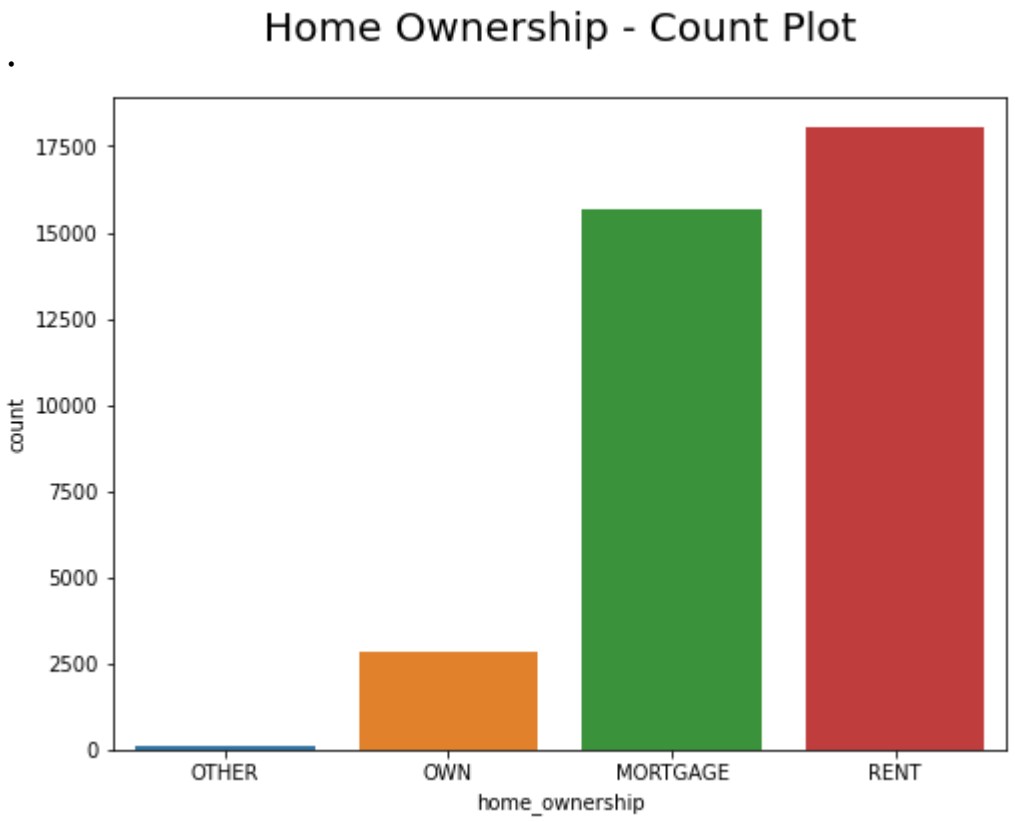
After preforming all the above tasks it is observed that there are in total **38577 Rows & 17 Columns** in the Dataset

Data Understanding

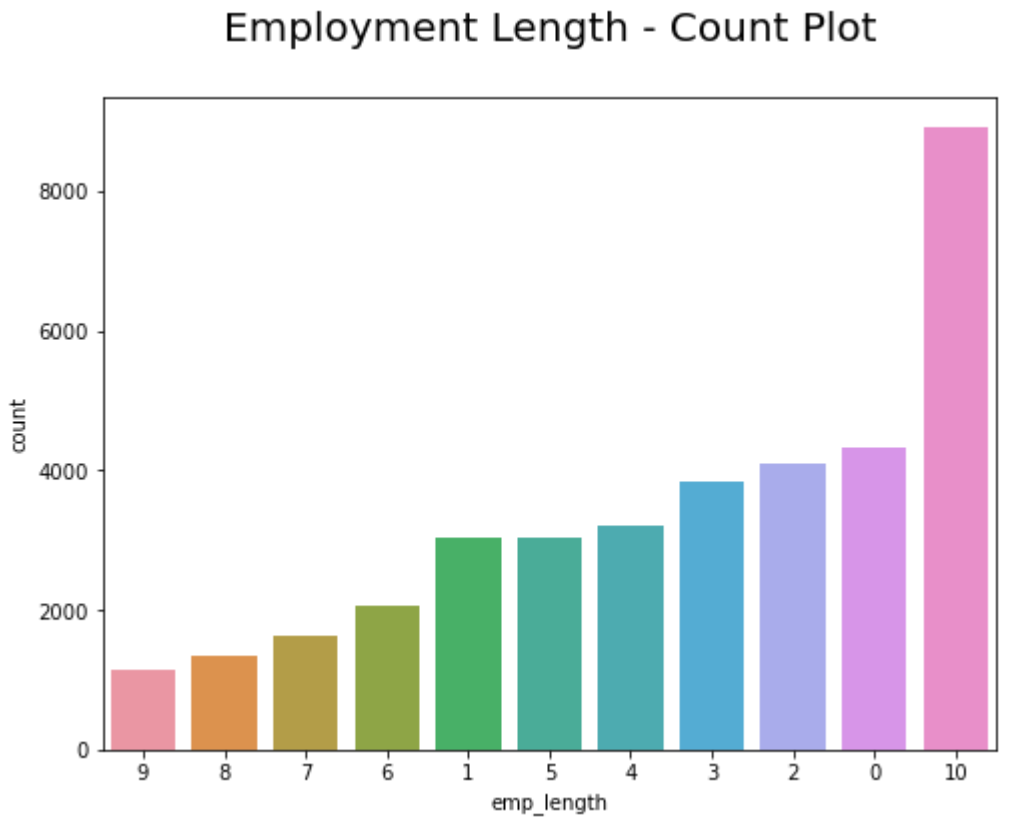
Below are the 17 columns , based on which the Data Analysis will be performed :

1. loan_amnt : The listed amount of the loan applied for by the borrower.
2. term : The number of payments on the loan. Values are in months and can be either 36 or 60.
3. int_rate : Interest Rate on the loan
4. grade : The grade of the loan. Values are A,B,C,D,E,F and G
5. sub_grade : The subgrade of the loan. The values are A1 to A5,B1 to B5 and so on.
6. emp_length : The number of years the loan applicant is employed.
7. home_ownership : The home ownership status provided by the borrower. The values are: RENT, OWN, MORTGAGE, OTHER.
8. annual_inc : The self-reported annual income provided by the borrower during registration.
9. verification_status : Indicates if income was verified by LC, not verified, or if the income source was verified
10. issue_month : The month which the loan was funded.
11. loan_status : Current status of the loan. The values are fully paid, current and charged-off
12. purpose : A category provided by the borrower for the loan request.
13. addr_state : The state provided by the borrower in the loan application
14. dti : A ratio of borrower’s total monthly debt and the borrower’s self-reported monthly income.
15. funded_amnt : The total amount committed to that loan at that point in time.
16. funded_amnt_inv : The total amount committed by investors for that loan at that point in time.
17. installment: The monthly payment owed by the borrower if the loan originates.

Univariate Analysis

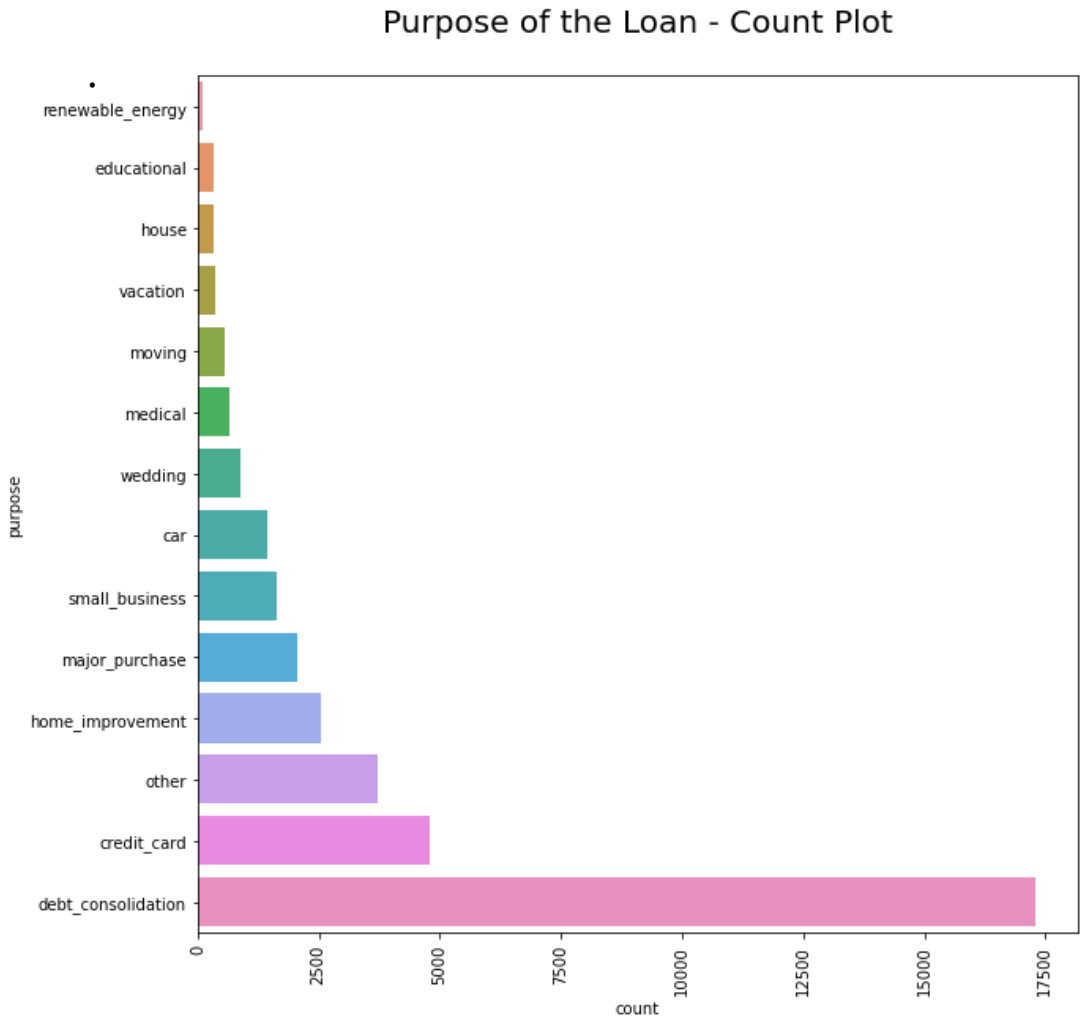


The above plot shows that the maximum borrowers were staying in rented home or their home was under mortgage and very few borrowers have their own house.

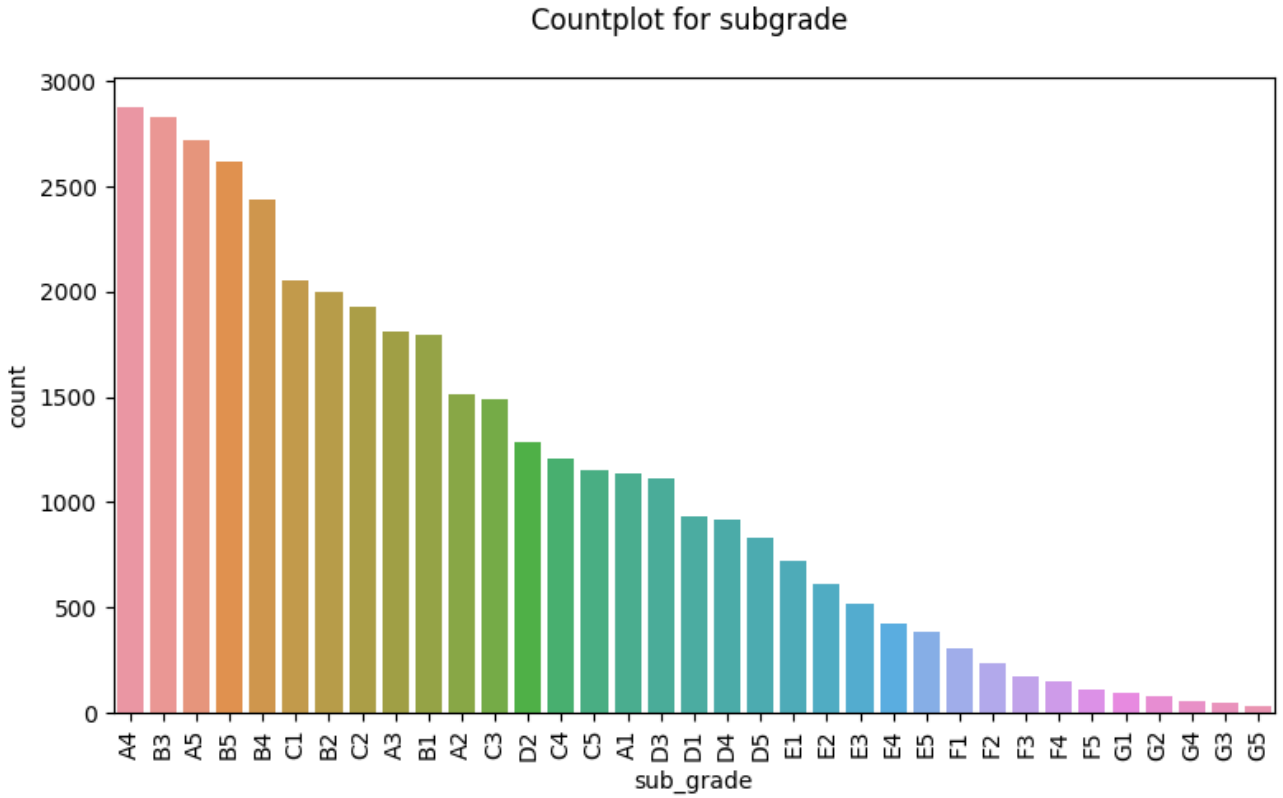


The above plots shows that the maximum borrowers had a employment length of either 10 years or 0 years that is they are freshers

Univariate Analysis



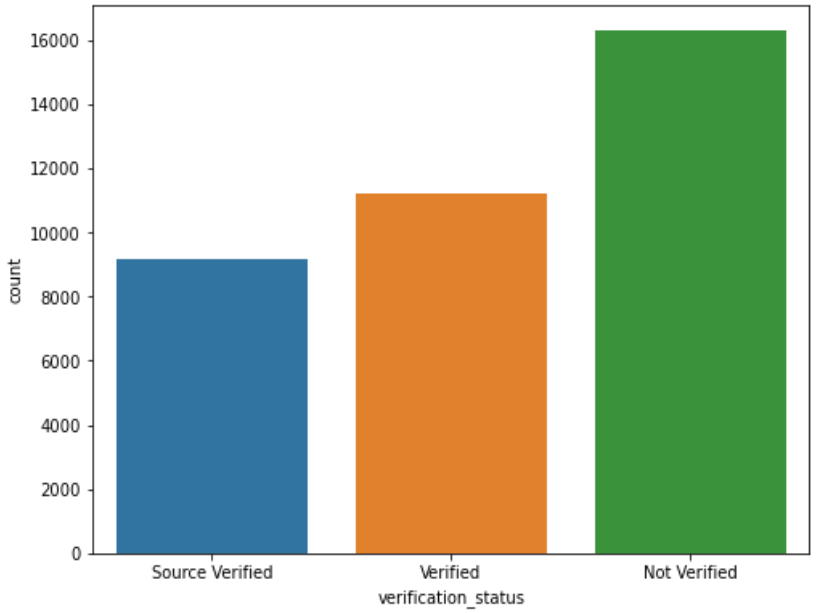
Above plots shows that the maximum borrowers had taken the loan for debt consolidation that is they are taking a new loan to pay off another loan.



Above plots shows that the maximum borrowers had a grade of either A or B and very few borrowers had a grade of G

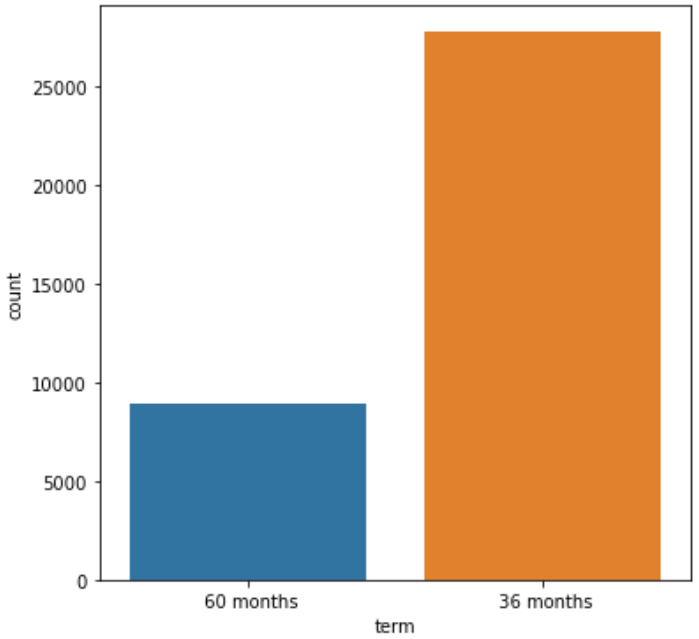
Univariate Analysis

- Verification Status - Count Plot



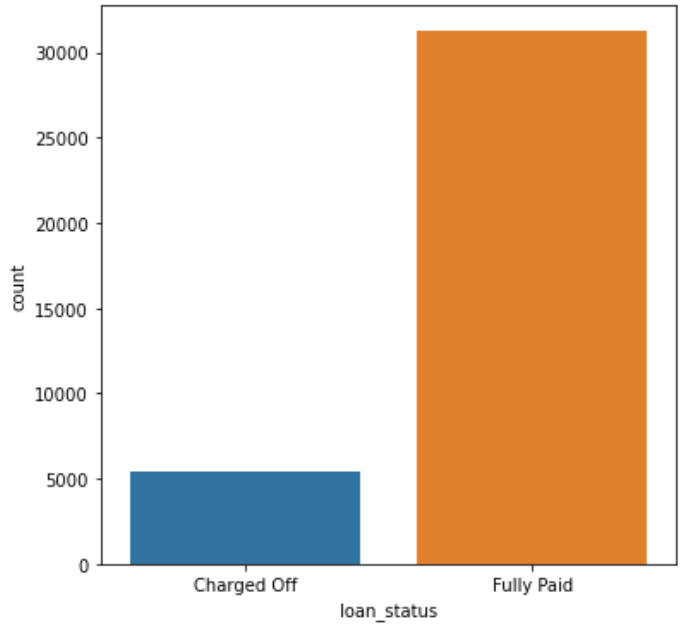
Above plots shows that the maximum borrowers income was not verified

Term of the Loan - Count Plot



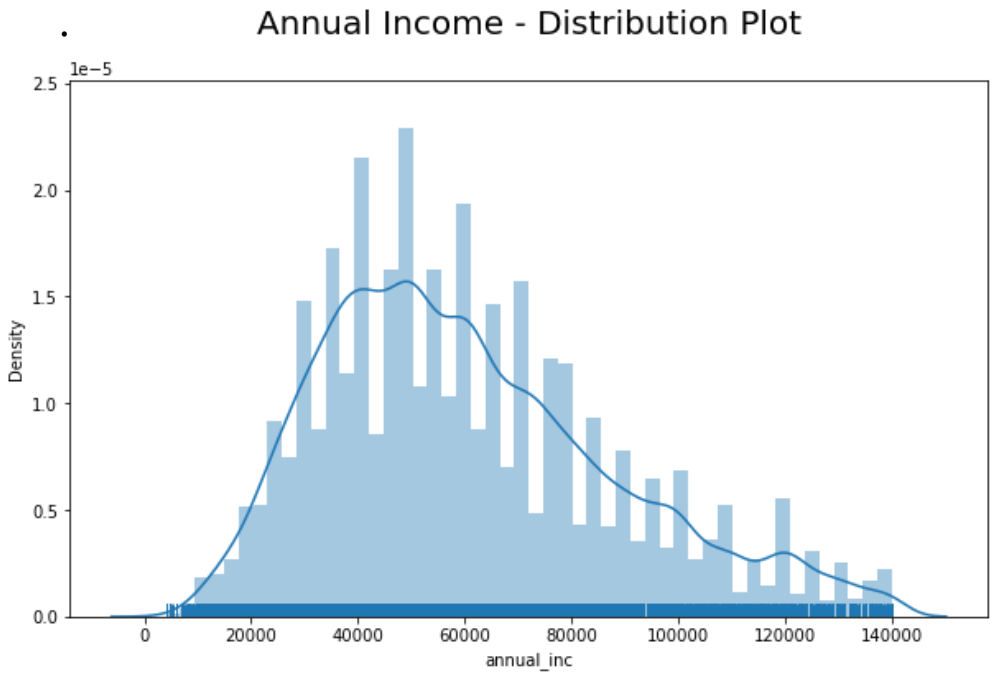
Above plots shows that the maximum borrowers had taken the loan for a period of 36 months

Loan Status - Count Plot

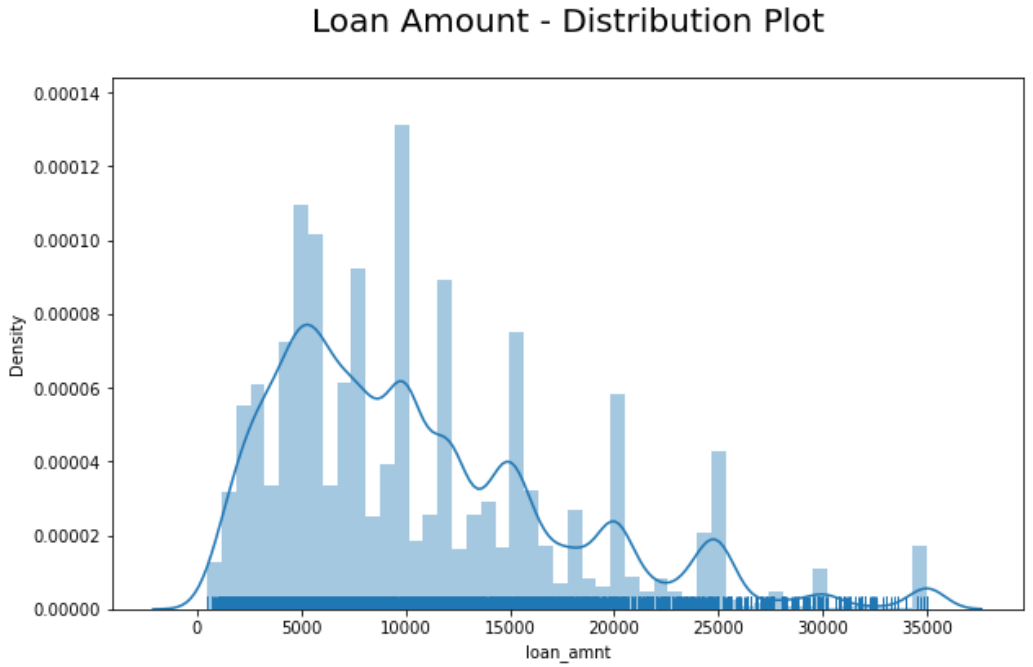


Above plots shows that the maximum borrowers had fully paid their loan

Univariate Analysis



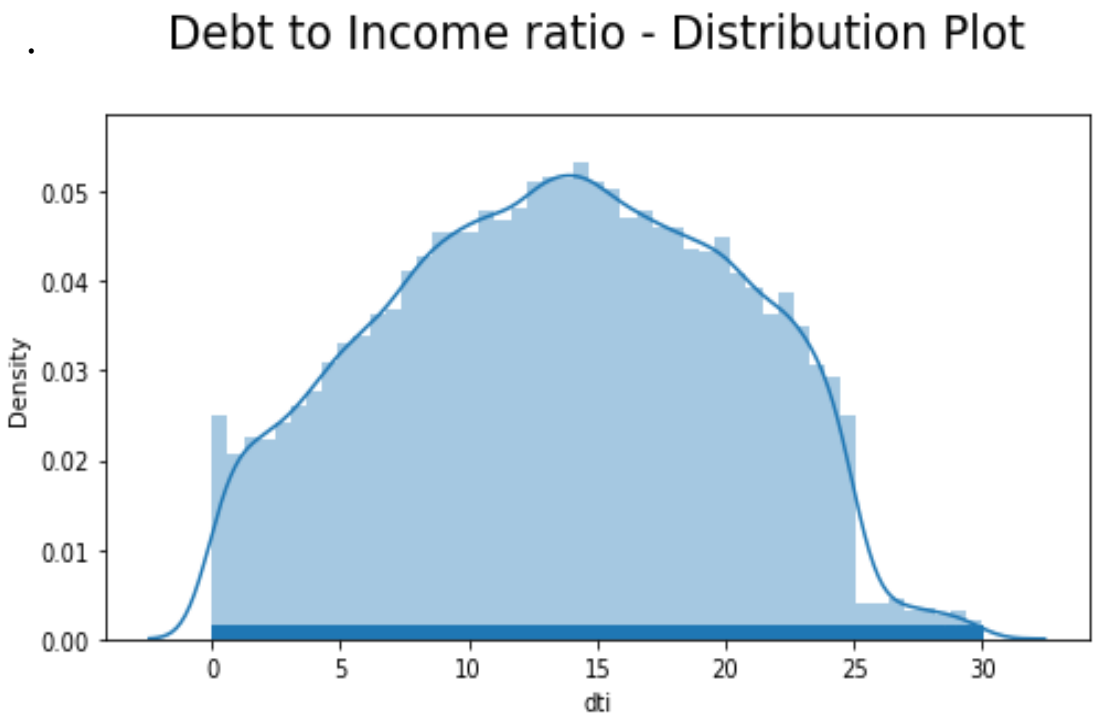
Above plots shows that maximum borrowers had an annual income between 40000 to 60000



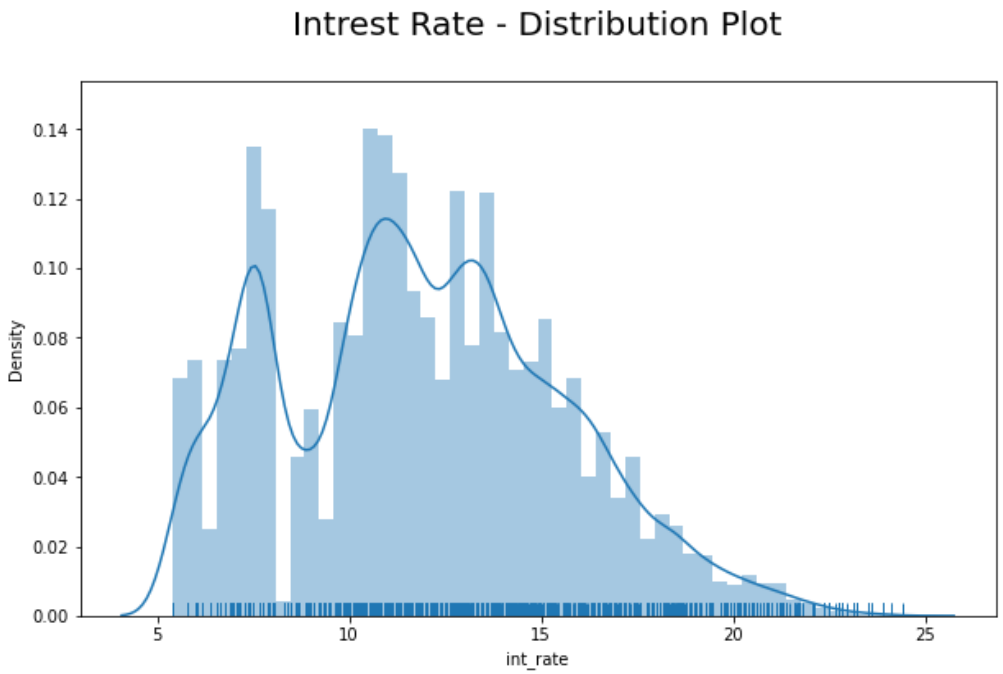
Above plots shows that the maximum borrowers had taken loan between the range of 5000 to 15000

Univariate Analysis

.

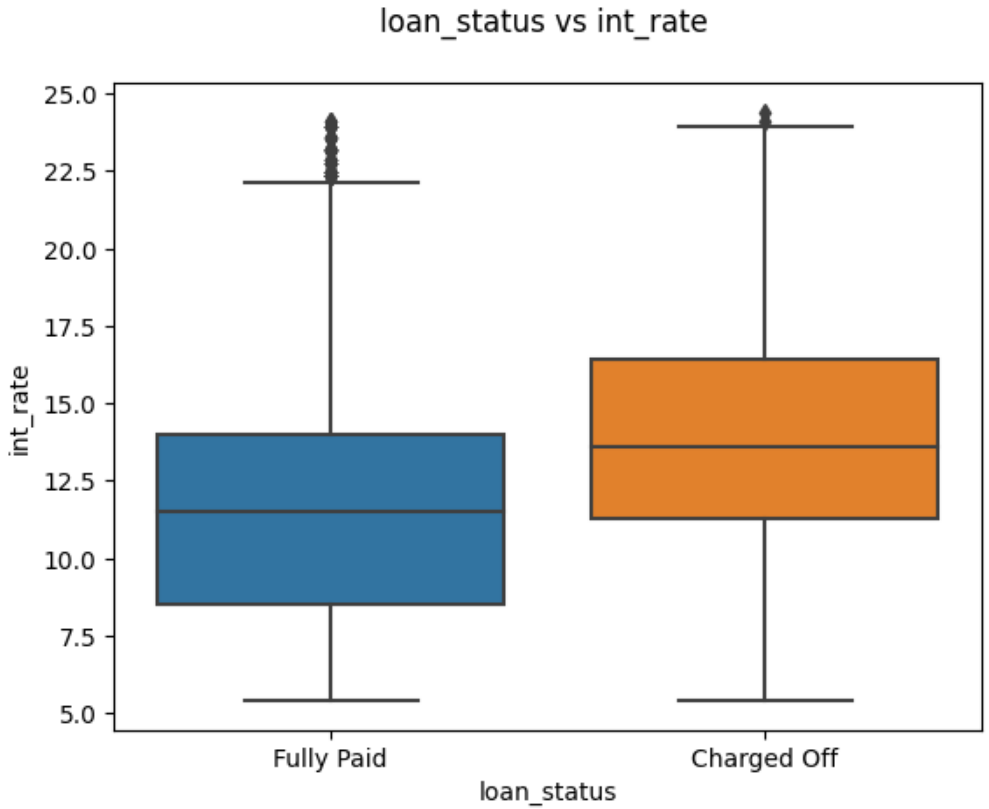


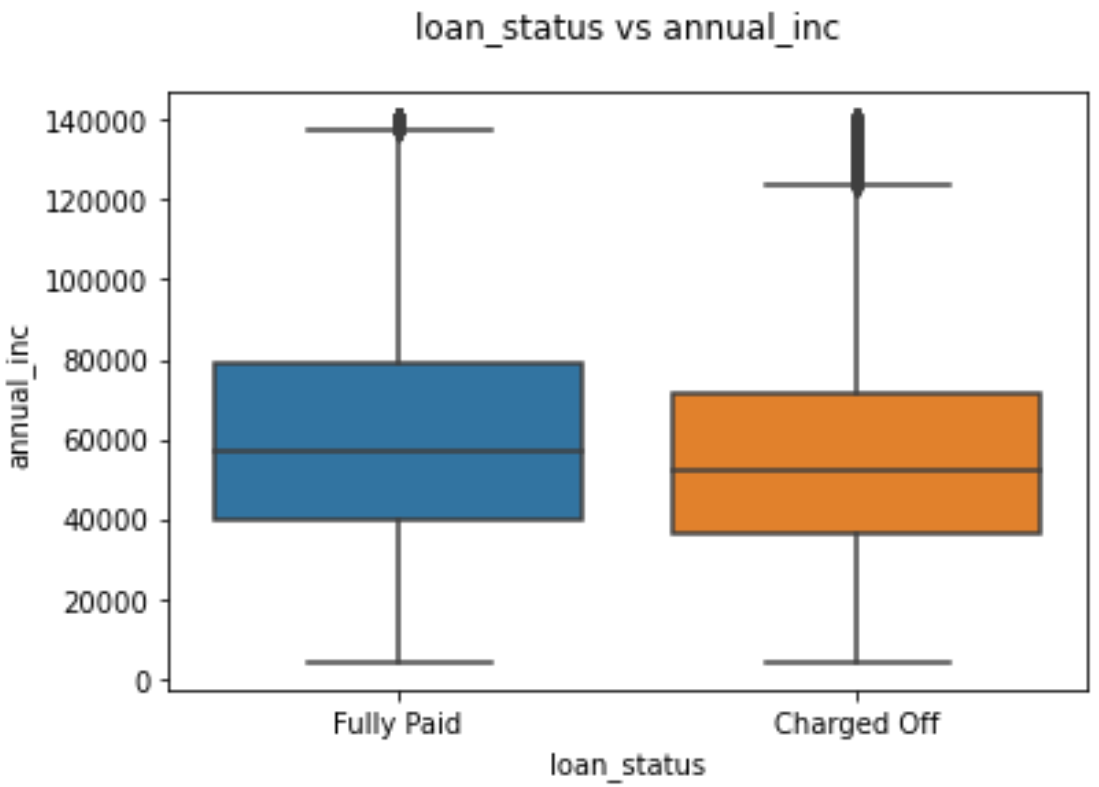
Above plots shows that maximum borrowers had an annual income between 40000 to 60000



Above plots shows that maximum borrowers had an interest rates in range of 10% to 15%

Bivariate Analysis

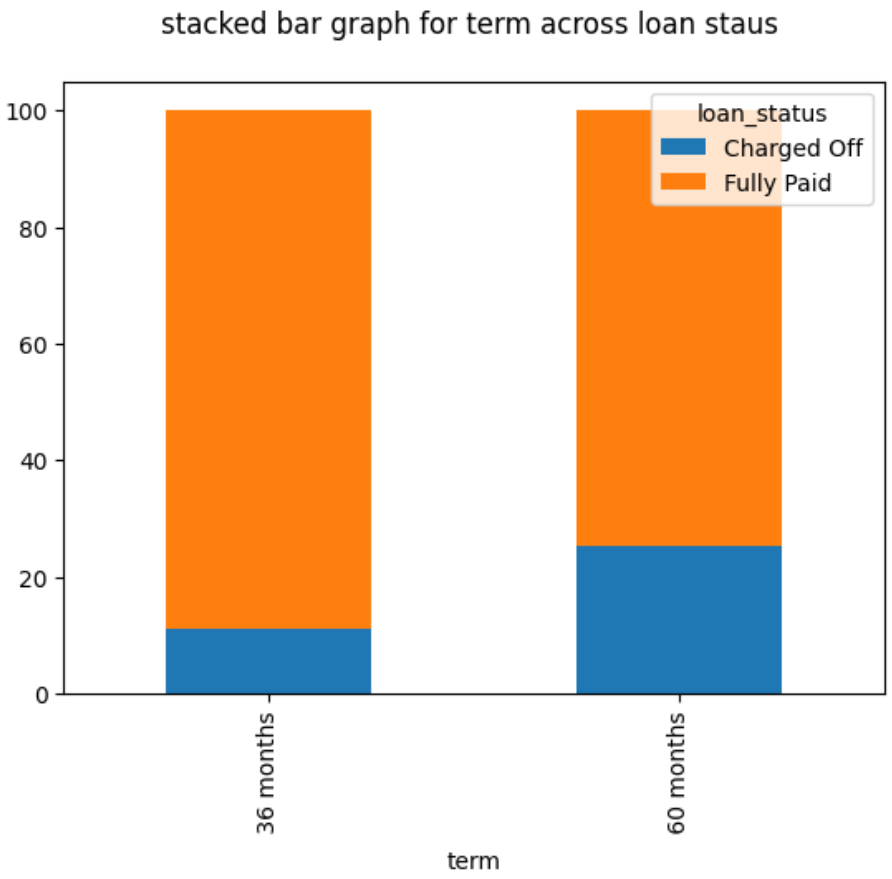




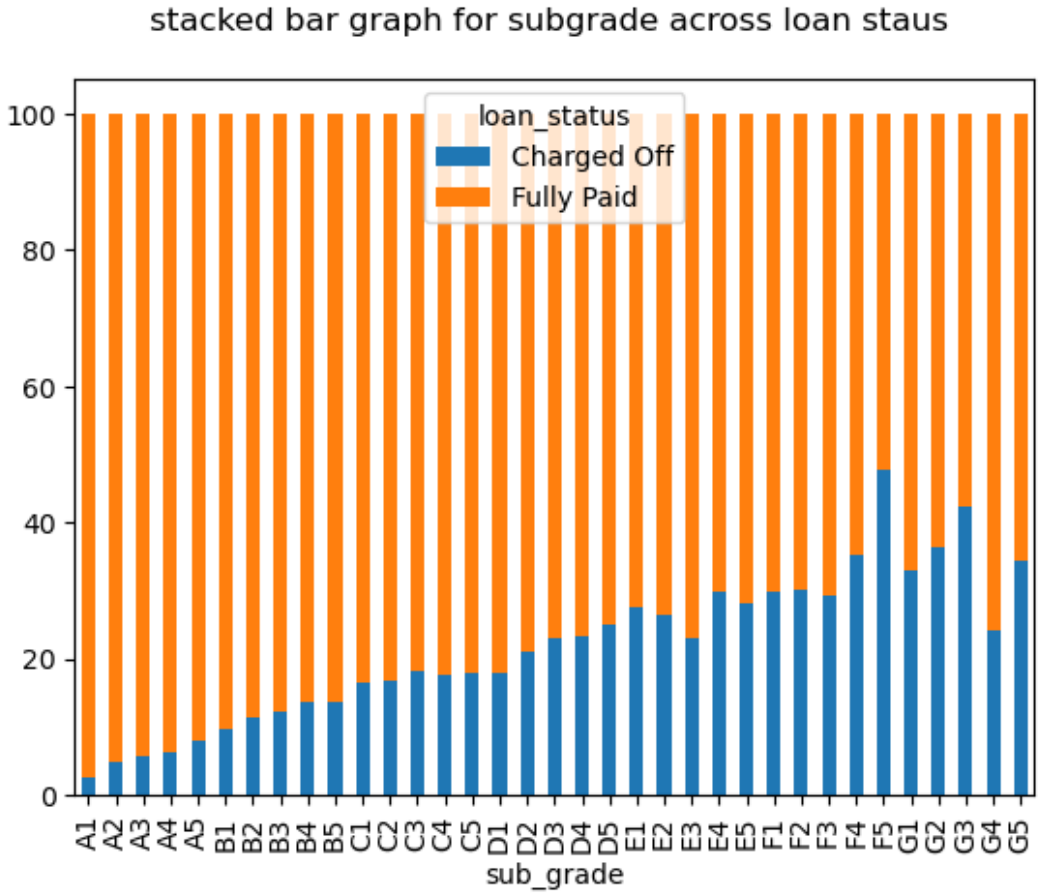
Above plots shows that the :

- Applicant who have charged higher interest rate have a very high changes to get default
- Applicant who are had low annual income had a higher chances to get default

Bivariate Analysis

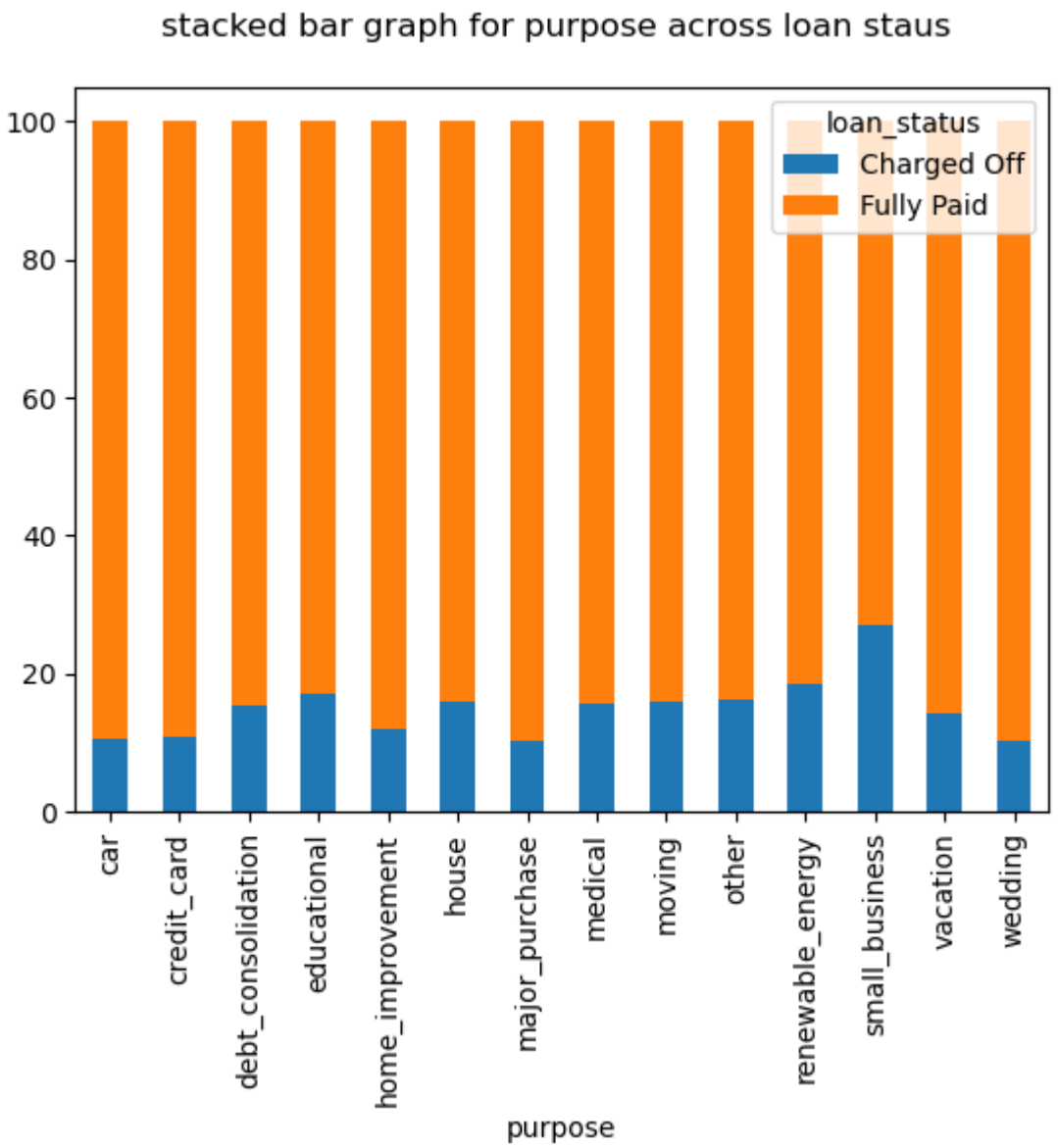


Above plots shows that the applicant with 60 month loan term are more likely to default compare to 36 month term.



Above plots shows that there is very high chance of default if applicant fall under F and G categories especially F5 and G3.

Bivariate Analysis



Above plots shows that there is significant difference in default percentage across purpose categories, with highest value in small business

Observation

The 5 driving factors that are strong indicators of default are:

1. Interest Rate: It is observed from the analysis that applicant who have been charged higher interest rate are more likely to default.
2. Annual Income : It is observed from the analysis that applicant who have lower annual income are more likely to default.
3. Term : It is observed that applicant with 60 month loan term are more likely to default compare to 36 month term.
4. Purpose : There is significant difference in default percentage across purpose categories, with highest value in small business.
5. Subgrade : There is very high chance of default if applicant fall under f and g categories especially f5 and g3.

So from above observation it is recommended that if any applicant fall under these criteria, bank should not give them loan directly. They should do more investigation before approving loan and thus it will reduce credit loss.

Thank You