Lending Club Case Study

Group Facilitator: Aaradhya Bendre

Objective:

Data Cleaning

The given data is not always suitable for a direct analysis.

May contain empty spaces or irrelevant data

Data Cleaning can be addition or deletion of unwanted spaces

Data Balancing

The formats of the data may differ from one row to the other.

Such variation is difficult to handle when working with the dataset

Data balancing/ standardization is a must

Analysis

Once we have a clean and standard data, we are performing analysis as follow:

- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis

Business Understanding

A bank/ lending company receives applications for loans, these applications are then scrutinised and loan is approved based of the data in the application.

General risk with the decision:

- If applicant is likely to repay and the loan is not approved
- If the applicant is not likely to pay but the loan is approved

___

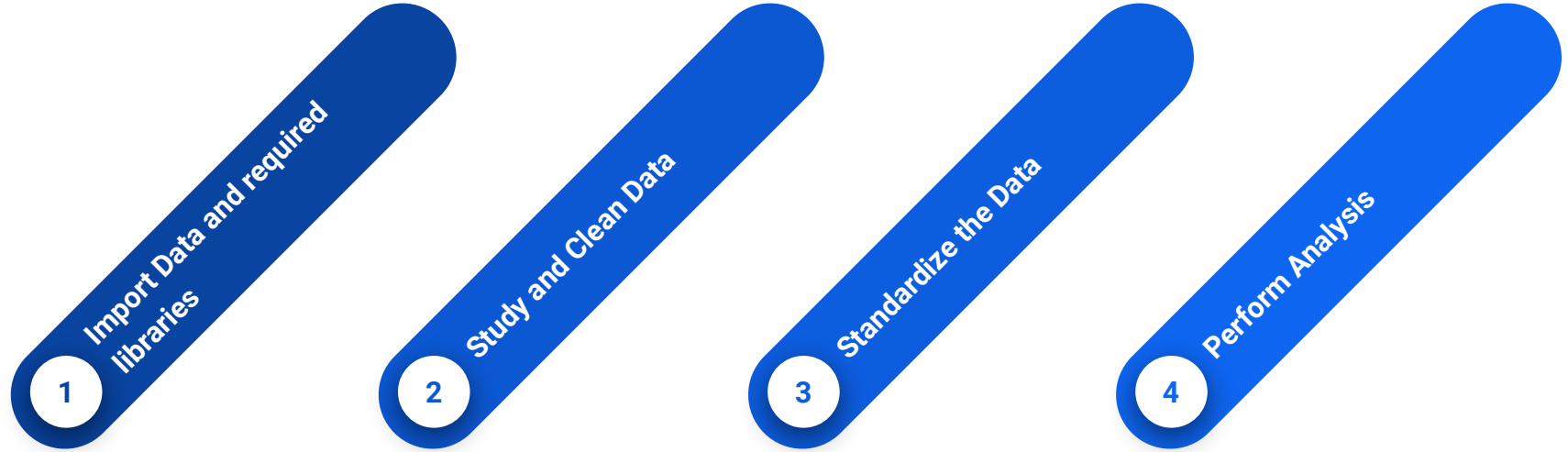# Case Study Consideration:

-

Current on going loans are not considered as they are neither default not non default

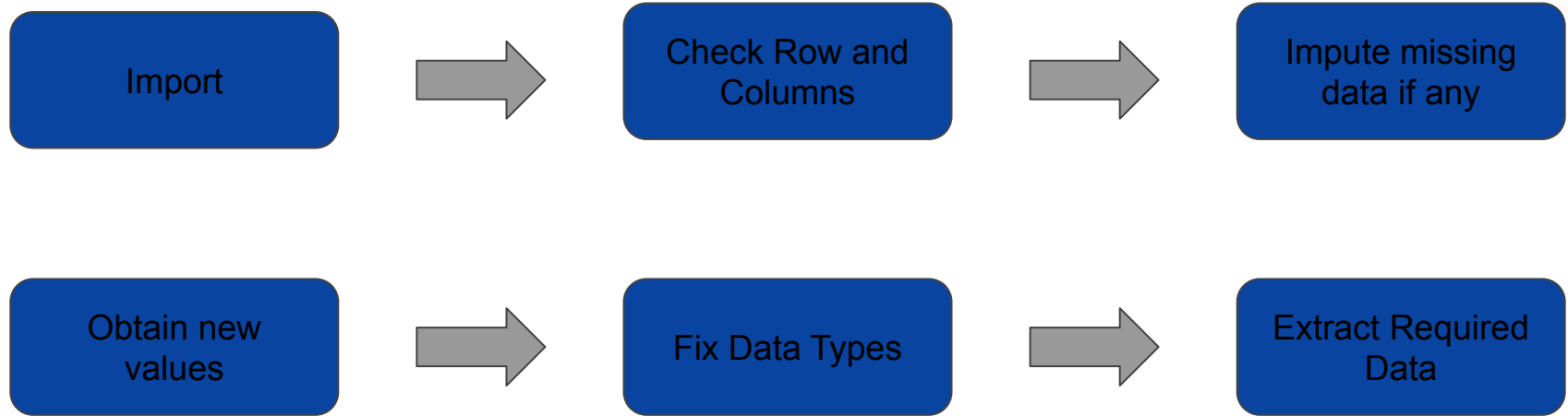Loans rejected do not have any role to play in this study

Columns with all empty spaces are taken away

Columns with very high empty spaces are taken away

Process Followed During the Case Study

1. Import Data and required libraries
2. Study and Clean Data
3. Standardize the Data
4. Perform Analysis

Data Clean Up Process

Import → Check Row and Columns → Impute missing data if any

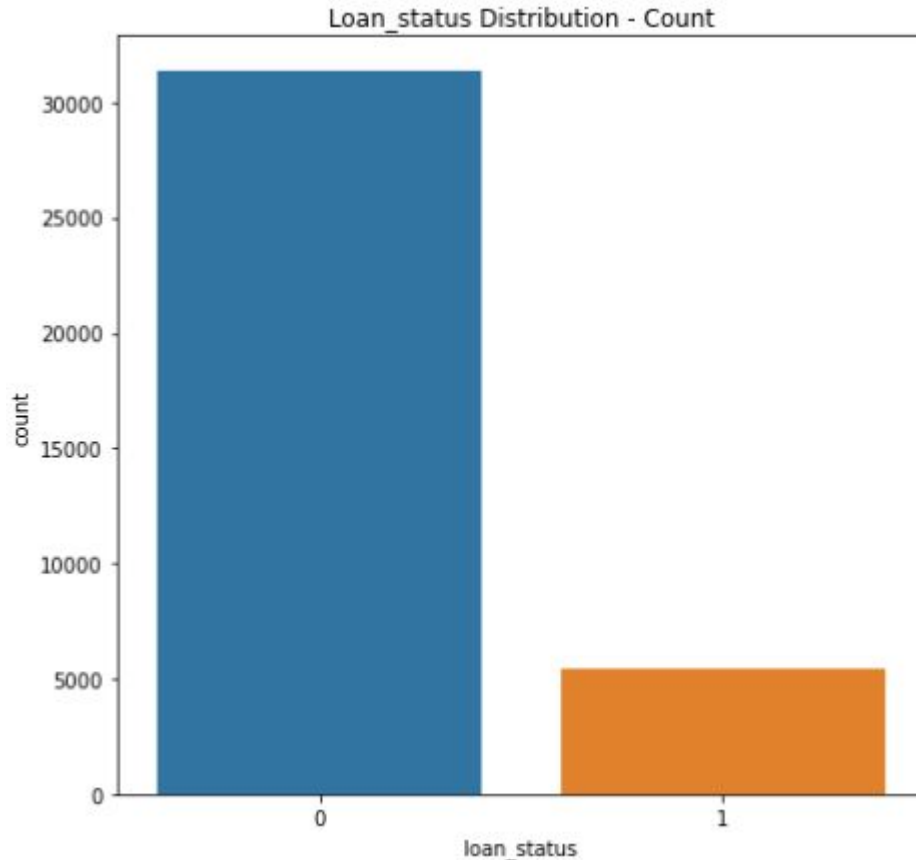Obtain new values → Fix Data Types → Extract Required Data

Analysis Performed

- Univariate Analysis

- Segmented Univariate Analysis

- Bivariate Analysis

Univariate Analysis

**Analysis of a single variable** - Descriptive in Nature

Loan Status Check

O - Represents the Full paid Loans
1 - Represents charged off loans

As per the data and the graph plotted, we can clearly see that majority of the loans are paid in full
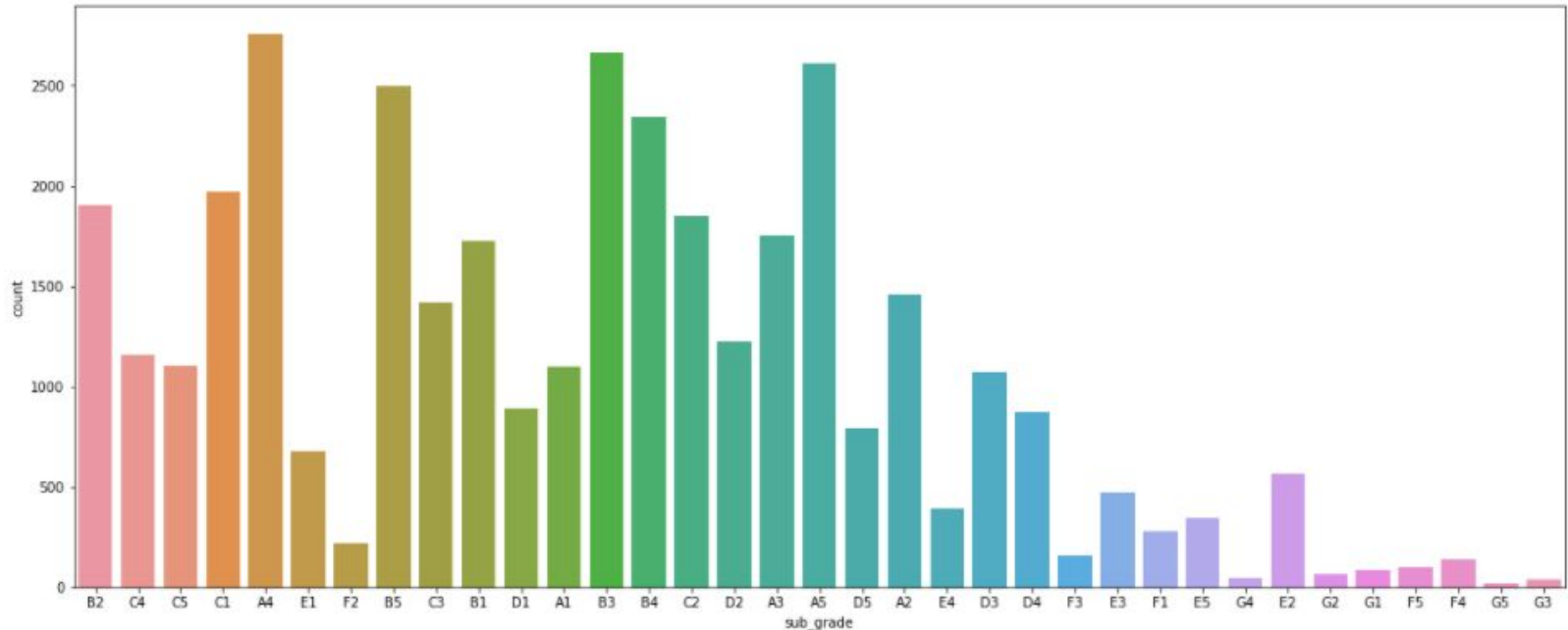
Loan Grade and Majority

- Here, we can see that Grade A loans are very less in terms of approval/applications
- Grade G is the one in demand amongst the people applying
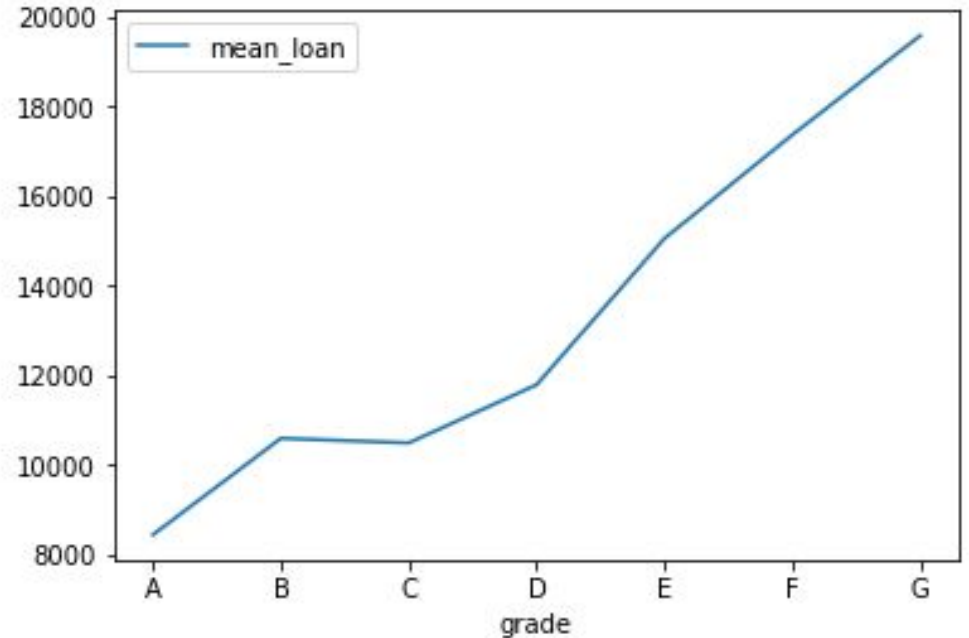
Loan Distribution along Sub grades

- With respect to the previous and the below graph we see that majority of the loans fall in A to B loan grades
- Specifically A4 to B5

Income Groups likely to Take loans

- Incomes groups that take highest loans are in the income range of 40k-50k
- Reason for income group taking less loans below 40K and above 60k can be very different
- Possibility, below 40k low repayment capacity, less savings
- Possibility, above 60k can handle without loans due to high savings/investments
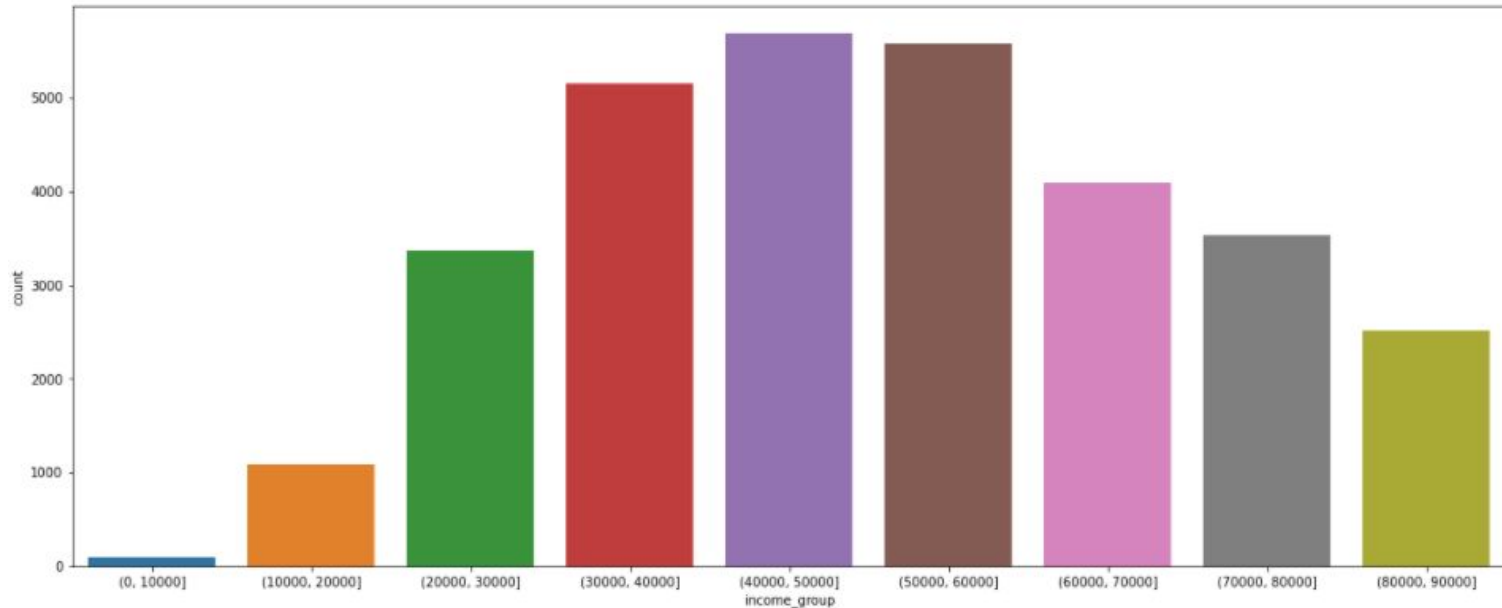
Segmented Univariate Analysis

Analysis of a single variable -
No involvement with other variable
Dataset is analyzed in subsets

Loan Status vs Emp Length vs Loan Count

- This graph gives us information that how long is the candidate employed, which category of emp length takes more loans and what is the status of their loans
- We see that people with **10+ years** of employment take more loans and also are the top at full payment of the same
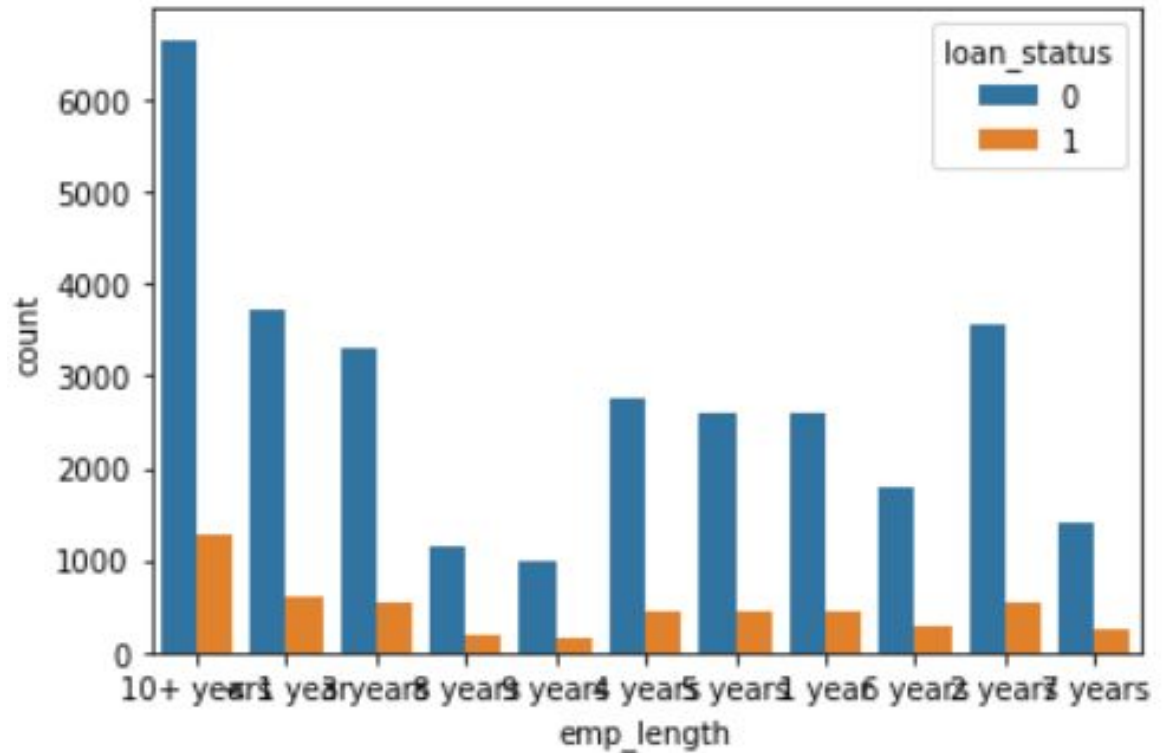
Loan Status vs Home ownership condition vs Loan Count

- This graph gives us information about the candidate's home ownership condition, who takes more loans and what is the status of their loans
- We see that people with **rented house** take more loans, are the top at full payment of the same, but also max. Number of charged off loans come from the same category
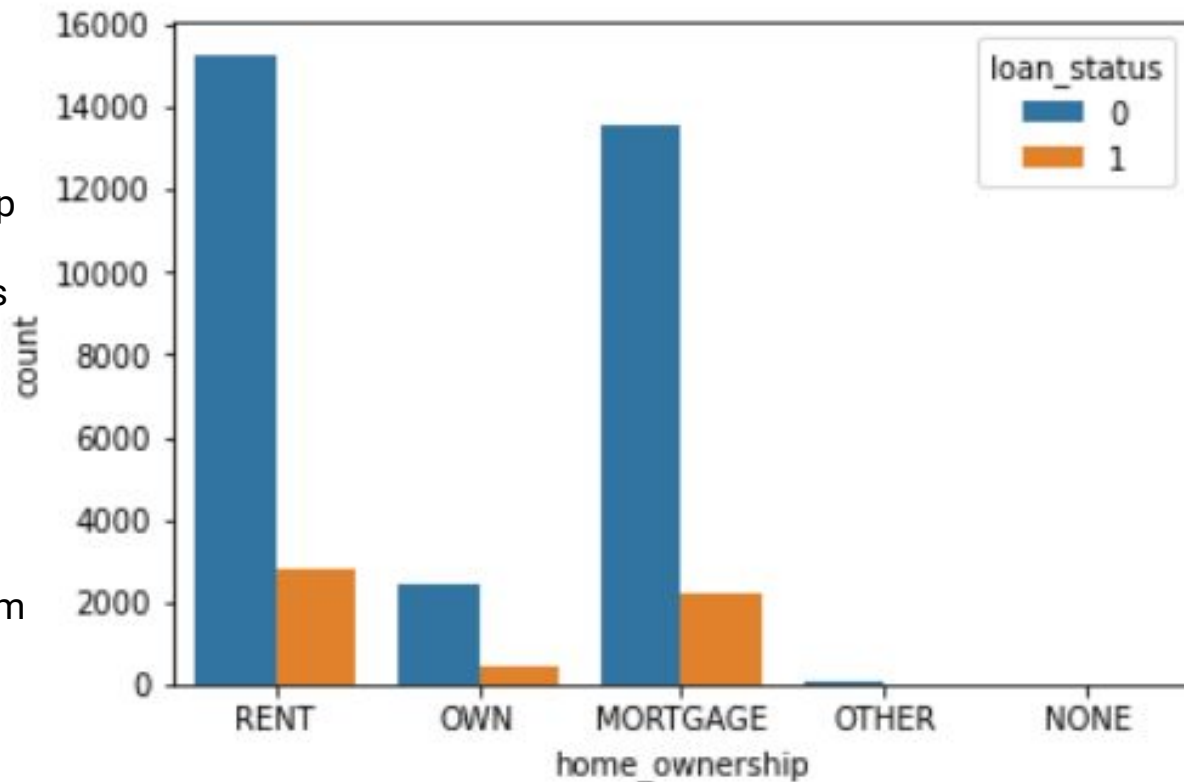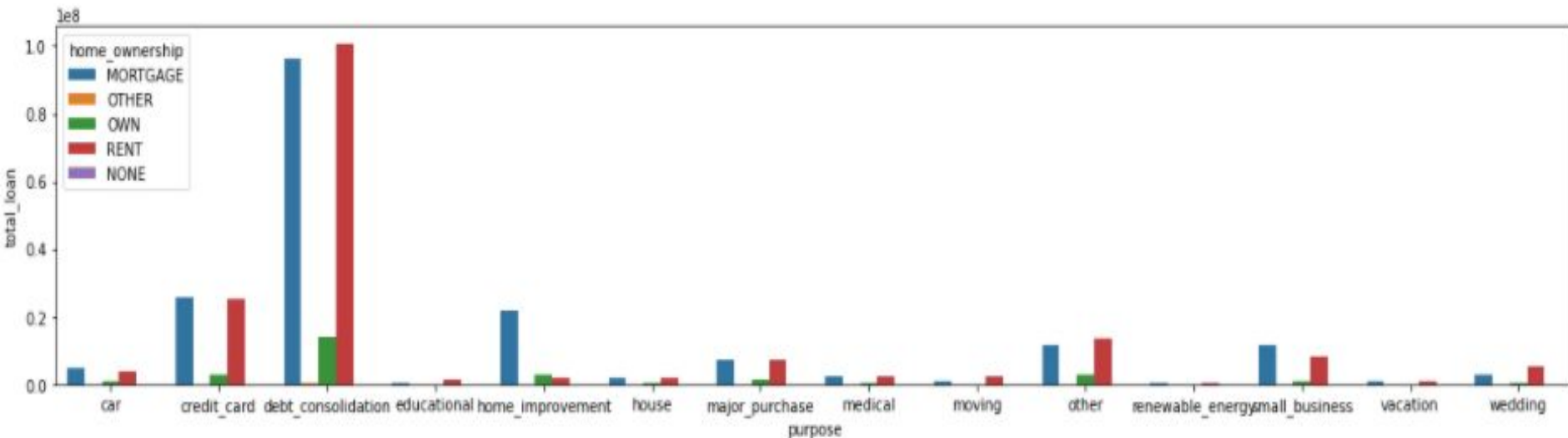
Bivariate Analysis

Analysis of a double variable -
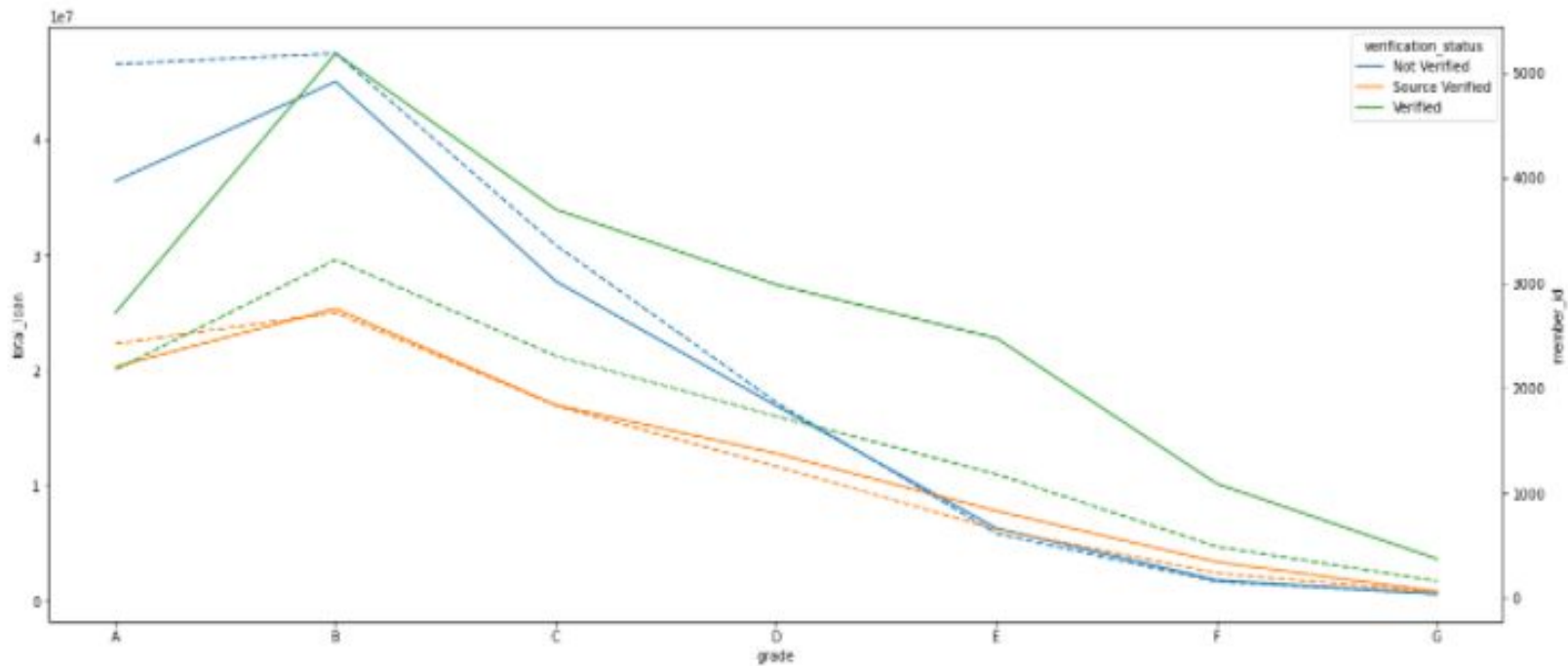Involvement with other variable

Purpose vs House Ownership

- Huge spike in total loans is seen for people who do not own a house
- The purpose of most of the loans can be concluded as for debt consolidation

- Max number of charged off loans are between A and C

Summary:

Data Summary:
- The loan amount ranges from 0 to 35000, mean is 10000
- Dataset was containing some outliers which were removed
- Some columns of the dataset were of not use to the analysis hence they were discarded
- Loans were divided into different grades from A to G
- Max. loans were in between A to C

Analysis Summary:
- It was observed that maximum loans are paid off in full
- Max loans are taken by candidate in the income range of 40k-50k
- People who do not own a house are the once that the the max. Loans, repay their loans in full but also get there loans charged off
- People working for more than 10 years take the highest loans and also are the ones who repay the most