LENDING CLUB CASE STUDY

Submitted by

-Tazeen Shaikh – Group Facilitator

-Sameer Inamdar - Team Member

Batch

AIMLC53

Date 05-July-2023

AGENDA

- Purpose of Lending Club Case Study
- Analysis Approach
- Loan Data Set Analysis
- Recommendations
- Assumptions

PURPOSE OF LENDING CLUB CASE STUDY

Apply the techniques learnt in EDA Course to understand how consumer attributes and loan attributes influence the tendency of default.

Develop a basic understanding of risk analytics in banking and financial services.

Understand how data is used to minimize the risk of losing money while lending to customers.

OUR ANALYSIS APPROACH

3 fold approach to Analyse Loan data set

DATA UNDERSTANDING

- The shape of the Data Set Loan was 39717,111
- Description is available only for 87 columns out of the 111 columns
- 56 out 111 columns are blanks
- 5 rows have more than 5 values missing for the columns
- Some values like emp length , loan amount is string instead of numeric

DATA CLEANSING

- Dropped all missing columns
- Removed all rows will null values > 5
- Interest Rate , emp length converted to numeric and dropping '%' ,'+' and '>' characters
- Converted annual income to numeric
- Converted issue Date to datetime
- Removed outliers from annual income
- Loan status converted to 0 and 1 for better analysis
- Loan Status 'current' has been ignored

DATA ANALYSIS

- Identified year wise charged off loans
- States where loans are highest
- no of loan and home ownership wise
- no of loan applications and years of employment
- year wise distribution of loans that are charged off and fully paid
- Correlate grade and loan status
- correlation between home ownership and loan status
- term-wise distribution of charged-off and fully paid loans vs loan amount
- Plotting loan amount and loan status
- Annual Income and Loan Status

LOAN DATA SET ANALYSIS

How we met the evaluation Criteria

UNIVARIATE ANALYSIS

1. No of loans charged off = 1 VS no of loans fully paid = 0
2. States where loans are highest
3. no of loans home ownership wise
4. no of loans and years of employment

SEGMENTED UNIVARIATE

1. Loan status conversion to low medium high very high to plot loan amount
2. Annual Income conversion to low medium high very high to plot loan amount

5 INDICATORS OF DEFAULT

I Borrower applicant attributes:

1.Annual Income

2.Employment length

II Loan attributes:

3. Interest Rate
4. Loan Term
5. Loan Amount

NEW METRICS DERIVED

1. Charged off Loans
2. Fully Paid Loans
3. Loan status – low to very high
4. Annual income – low to very high

BIVARIATE ANALYSIS

1.Correlate grade and loan status

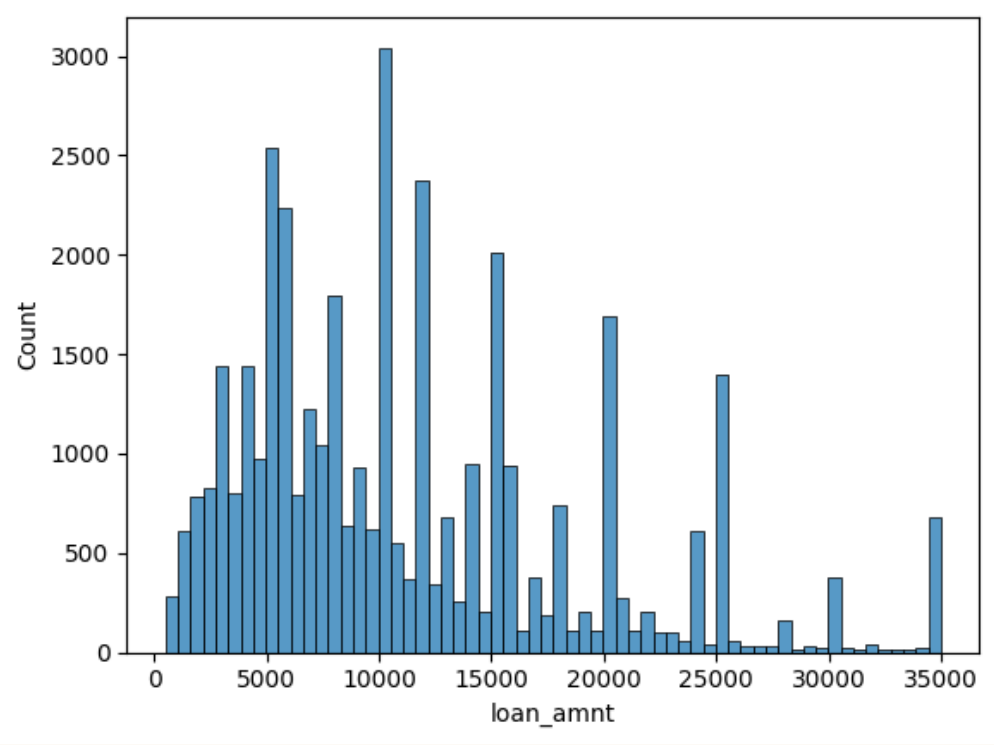
2.Correlation between home ownership and loan status

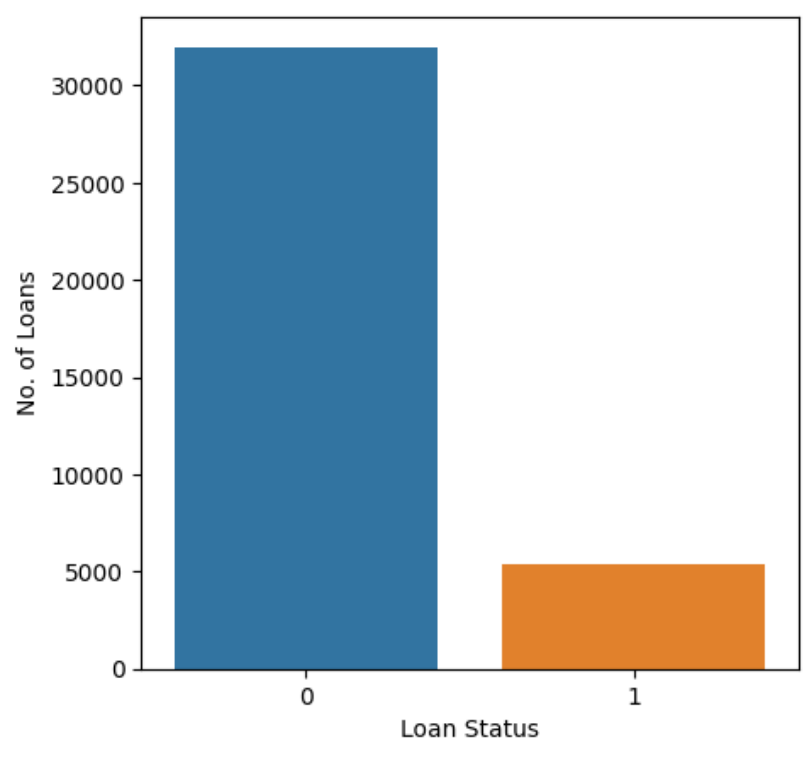
3.term-wise distribution of charged-off and fully paid loans vs loan amount

4.Plotting loan amount and loan status

5.Annual Income and Loan Status

UNIVARIATE ANALYSIS -I

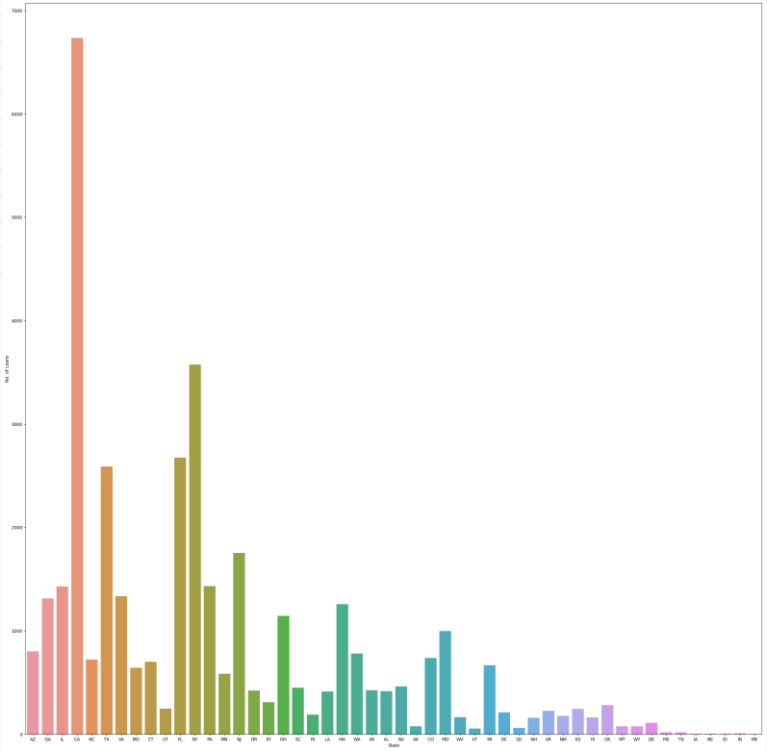




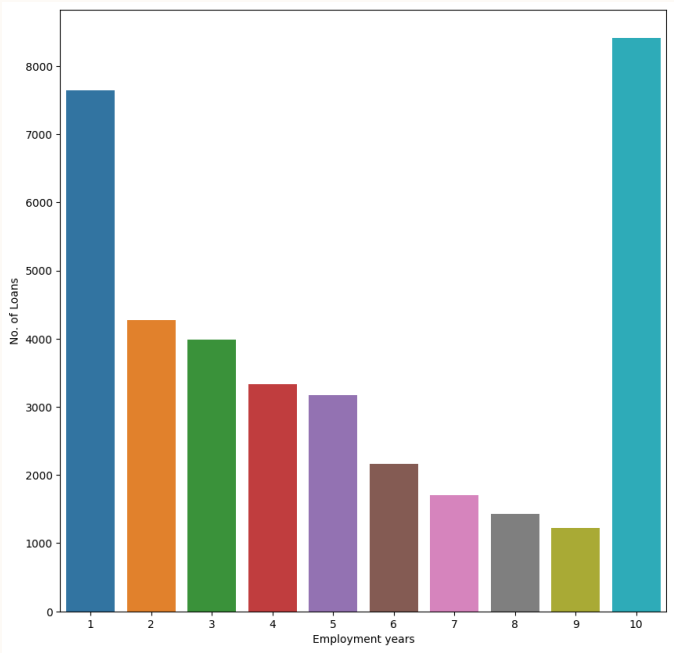
Analysing distribution of loan amount : loan amount of 10k is the highest

Analysis of the above count plot is that there are 30000 and above fully paid loans as compared to approx. 5000 default loans . Which is approximately 4% of the total loan count.

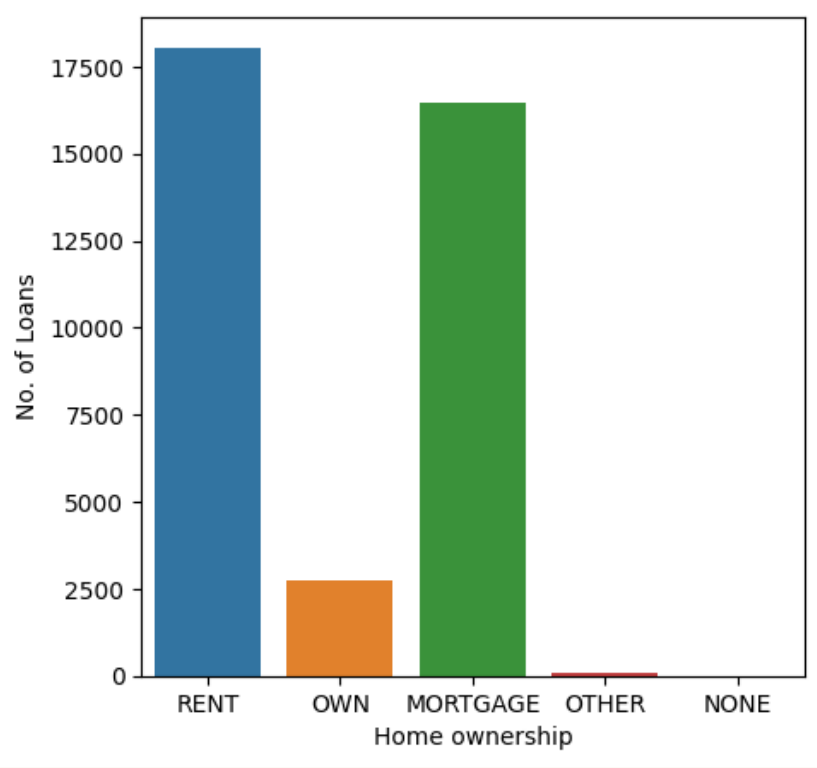
UNIVARIATE ANALYSIS -II



The states and number of applications does not really infer much about the defaulter states. California has the highest number of loans

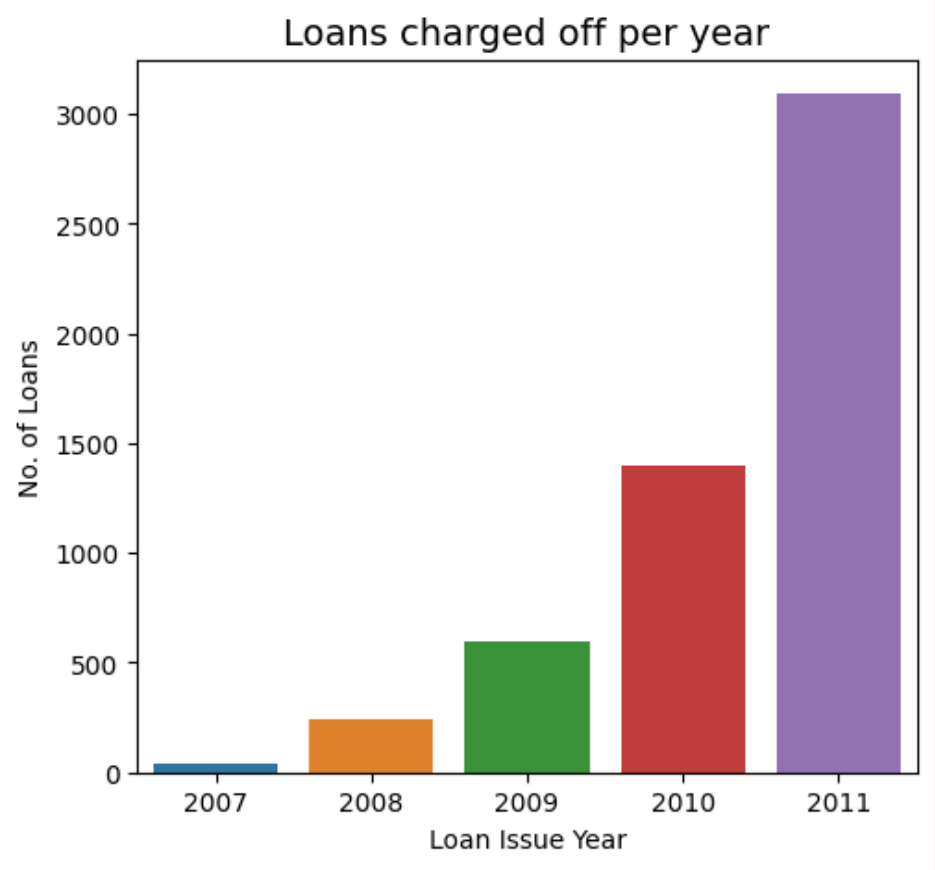


In the above bar graph we can infer that the bank is giving approximately equal number of loans to borrowers with 1 year employment and 10 year employment.

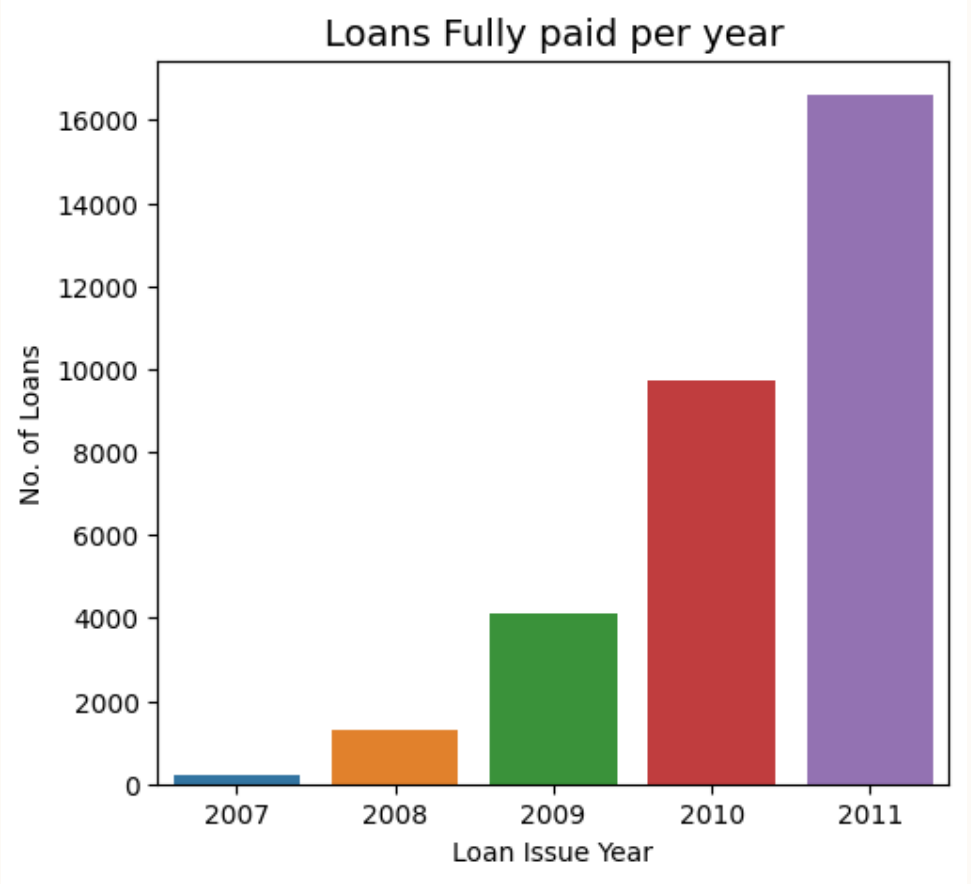


Borrowers with rented homes have highest loans

NEW METRICS

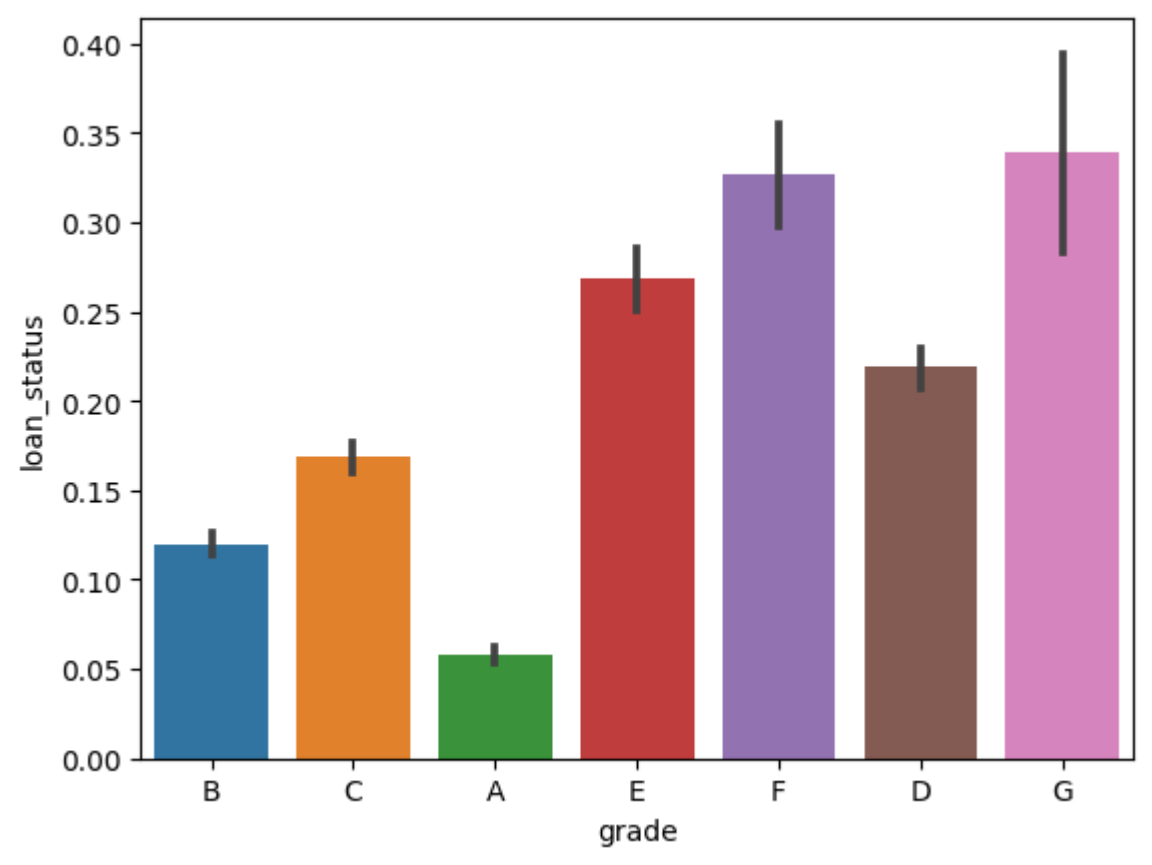


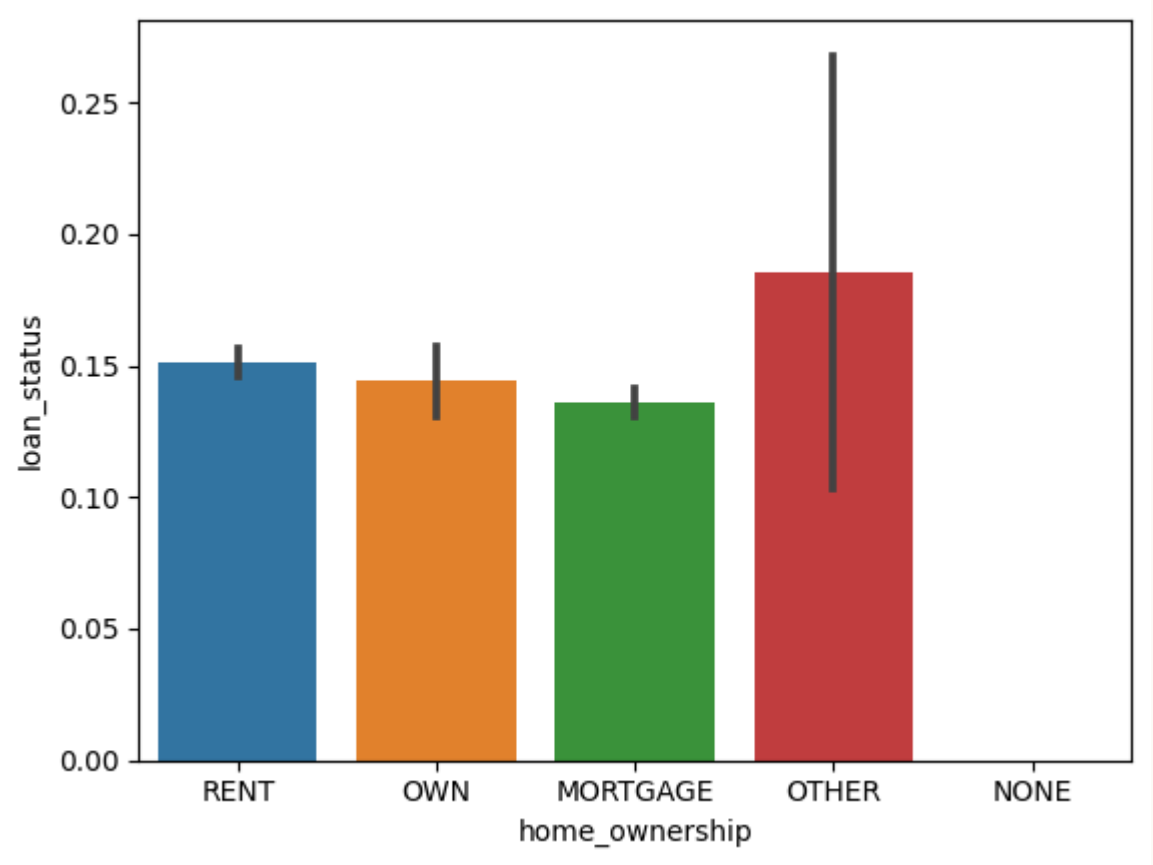
The number of default loans is increasing exponentially from 2007 to 2011 almost double each year which should be a cause of worry for the lending club



Fully paid loans are also increasing each year

BIVARIATE ANALYSIS - I

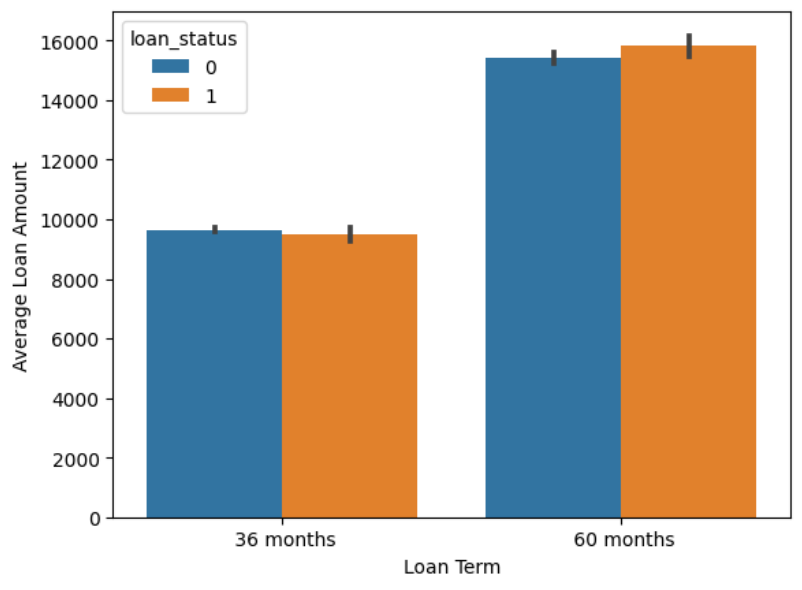




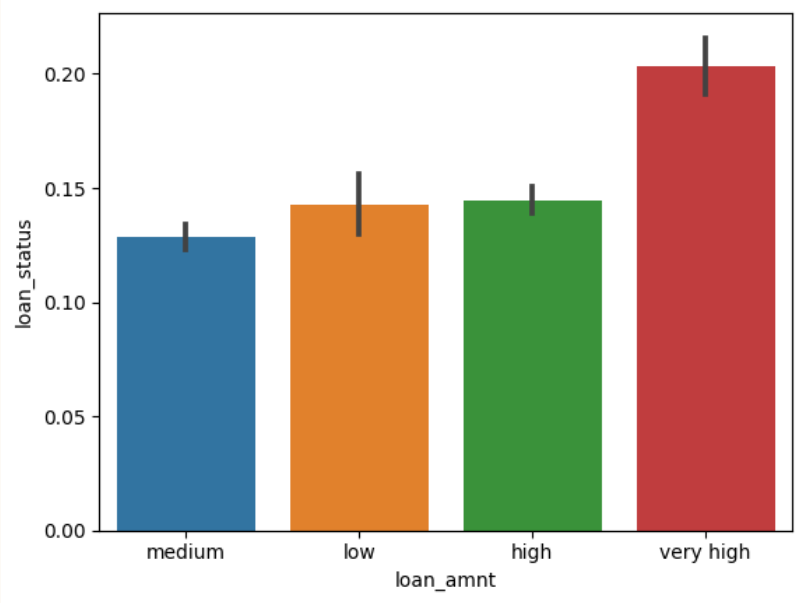
grade and loan status indicates that for A ,B grades the loan status is fully paid as compared to F , G E and D.

from the above barplot we can infer there is not much variation on loan status based on the ownership type. Therefore home ownership should not be considered as a variable to identity potentially default applicants.

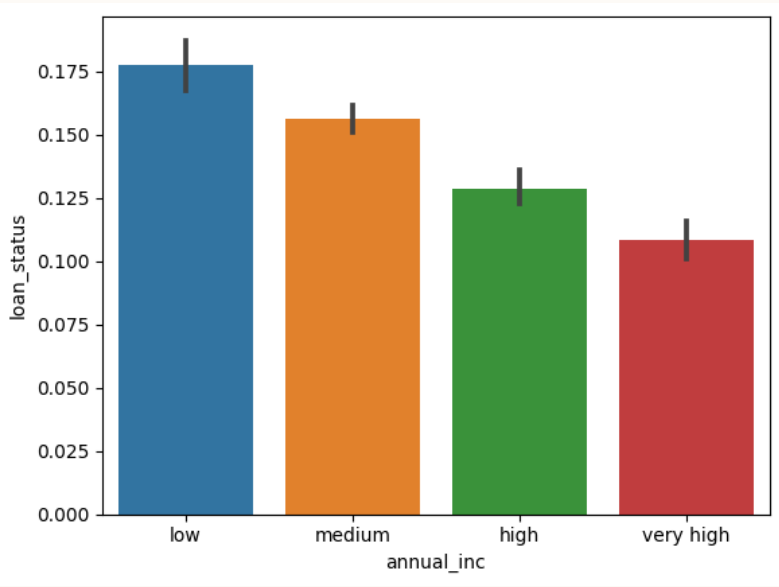
BIVARIATE ANALYSIS - II



As we can see from the above bar plot 36 months or 60 months term period does not have much impact on charged off and fully paid ratio.



very high loan amount is creating more charged off loan status



lower the annual income higher is the default rate

RECOMMENDATIONS

1.The data provided to provide clear default profile parameters is insufficient. Out of the 111 columns only 57 columns had meaningful data and finally ended up using only 54 columns which is approximately 50% of the columns only. Recommendation is to invest in a proper data gathering exercise and redo the Analysis.

2.The following variables were analyzed and compared:
For the borrower : Annual Income , Employment length and home ownership For the Loan : Loan Amount , term of loan , grade , interest rate , Loan status. If the data for other attributes provided then Analysis can be more accurate

3.out of the total loan count of 38420 for Lending Club Bank only approximately 4% loans are charged off which indicates there is a robust mechanism in place to identify defaulter borrower profiles

4.grade and loan status indicates that for A ,B grades the loan status is fully paid as compared to F , G E and D. so bank should give more of A,B and C graded loans.

5. Even though the percentage of total Charged off loans is 4% , it is increasing only a YOY basis . Therefore Lending Club Bank should invest in thorough Data Analysis with good quality data.

6.Home ownership of the applicant could be a good indicator for identifying defaulters. But there are many 'other' types which does not provide correct insight into the defaulter profile.

7. Loan amount and Loan status are directly correlated. As the loan amount increases charged off count also goes up so bank should avoid giving higher value loans.

8. 36 months or 60 months term period does not have much impact on charged off and fully paid ratio. So bank can go up to 60 month terms period

9. Annual income of the borrower applicant if high is able to pay off the loan and does not default. So bank should target high net worth individuals for loans.

10. The bank is giving approximately equal number of loans with 1 year employment and 10 year employment. Bank can look into increasing loans to applicants with more number of employment years as we have noticed more defaulters with less employment years.

ASSUMPTIONS

1.The data provided is of the bank at which the applicant has applied for loan. We do not have other bank data.

1.The current loan status has been ignored as the assumption here is that a person already having a loan with the bank will not be given another loan.

THANK YOU

Tazeen Shaikh

Sameer Inamdar

https://github.com/tazshaikh/LendingClubCaseStudy