LENDING CLUB CASE STUDY

SUBMISSION

Group Members:

1) Kaustubh Chepe

2) Shreyas Vadhula Sainath

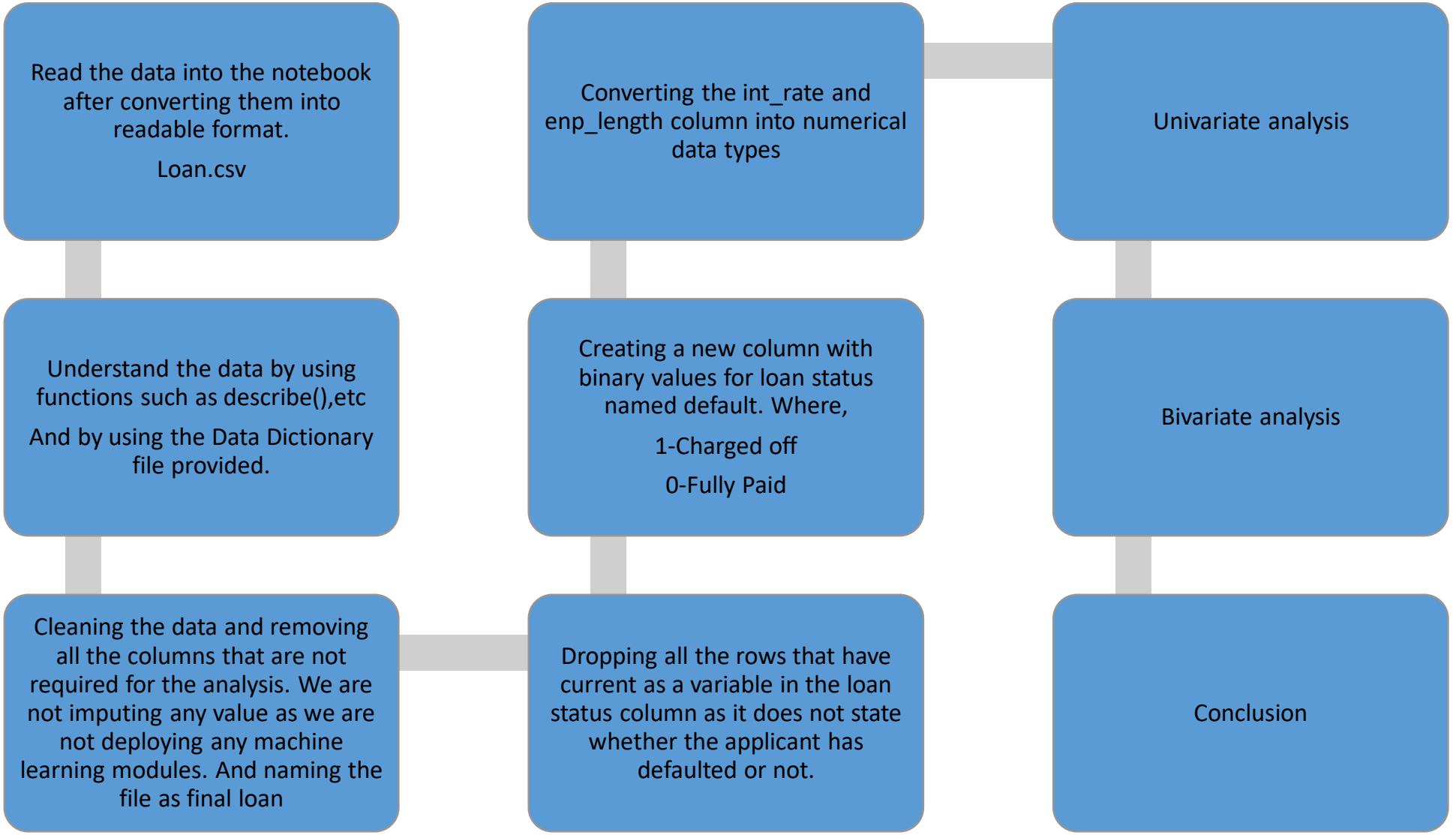
Abstract :

Lending Club 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. However leaning loans is a risky in nature, it will lead to financial loss to the company if the loan get sanctioned to a wrong customer. If one is able to identify these risky loan applicants before sanctioning the loan to them, the company can reduce its losses.

Business Objectives :

Lending Club wants to identify the risky applicants and understand the driving factors behind the loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge foe its portfolio and risk assessment.

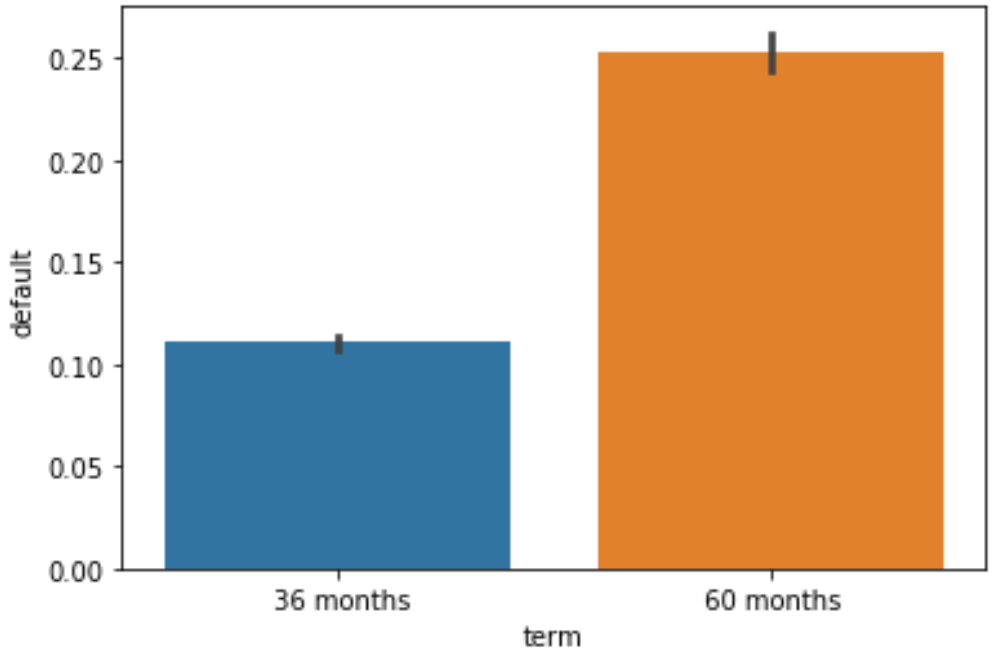
Problem solving methodology :



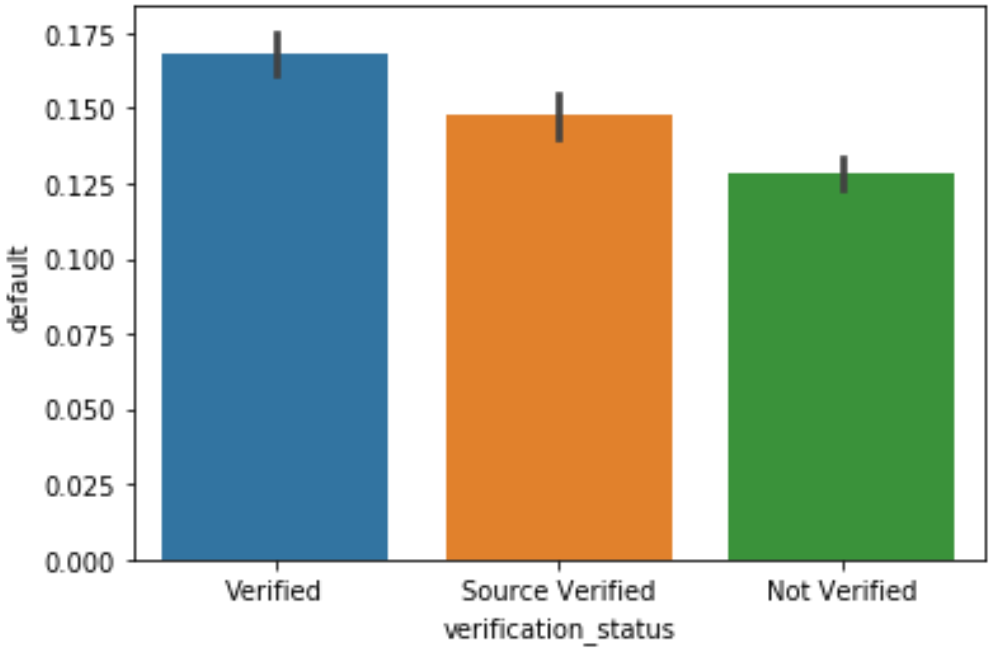
Analysis Approach :

- The main reason for doing this analysis is to figure out the variables that directly affect the out come of our target variable, which in this data is loan status.
- By converting this variable into a binary form where, 1- charged off (defaulted) and 0- fully paid (non-defaulted), it will become easy for analysis.
- On performing Univariate and Bivariate analysis and comparing the target variable with all the other variables we can arrive at a conclusions based on the outcomes obtained from them.

Univariate Analysis :

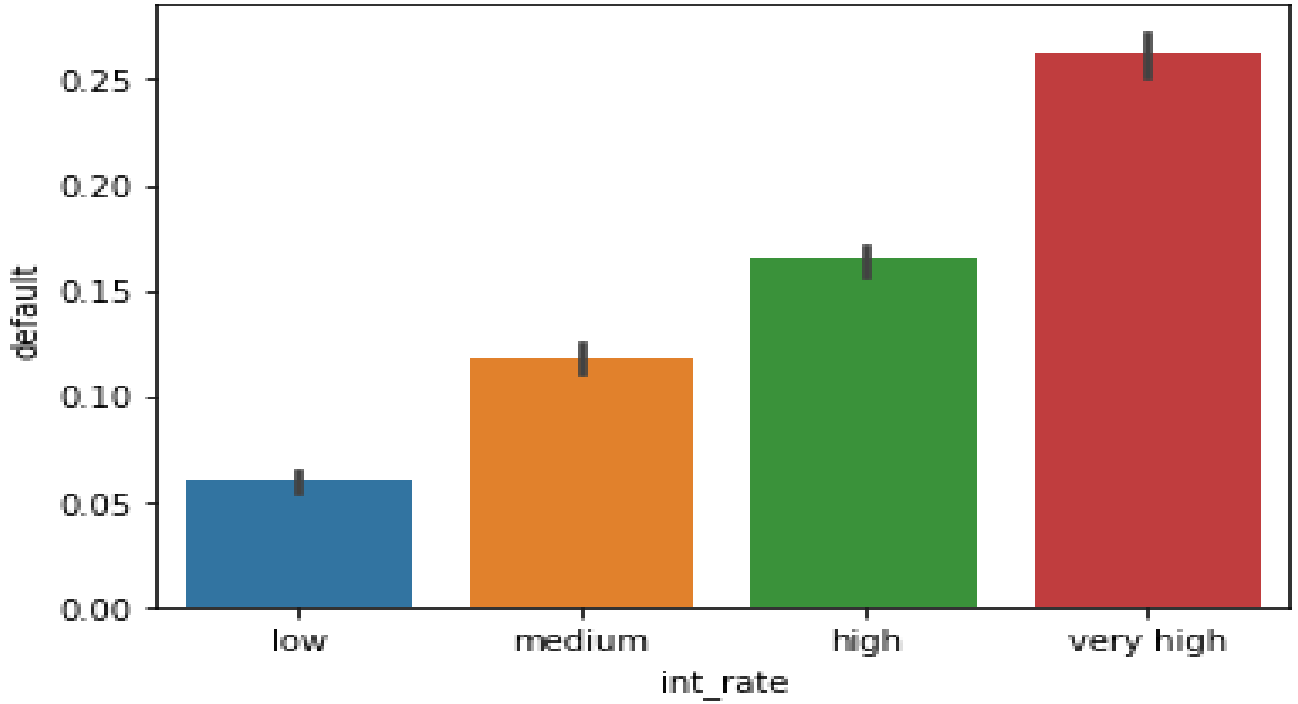


a) Plot of term vs default
From above graph we can assume that the applicants who applies for 60 months plan are more likely to default than 36 months.



b) Plot of verification vs default
From above graph we can assume that the applicants who applications are verified tend to default more but this cannot be taken into consideration as a term affecting the target variable.

Univariate Plots :

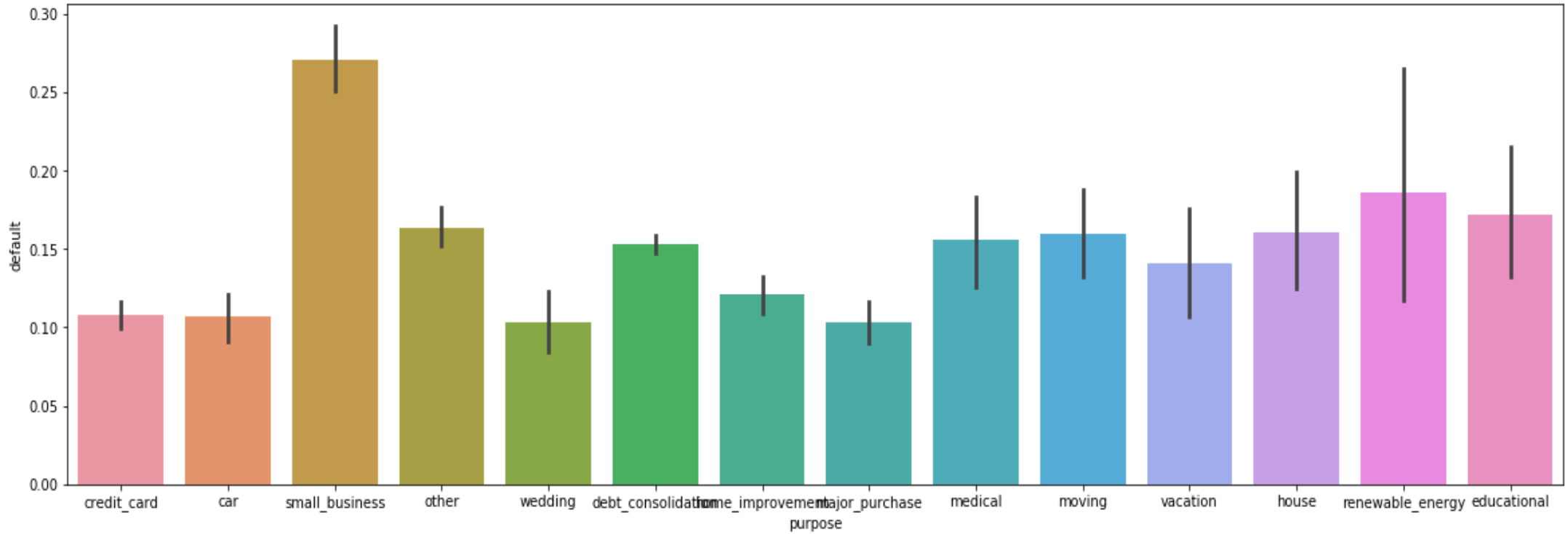


c) Plot of int_rate vs default

Here the int_rate column is binned into four groups.

From above we can say that with vary high interest rate default rate is also high.

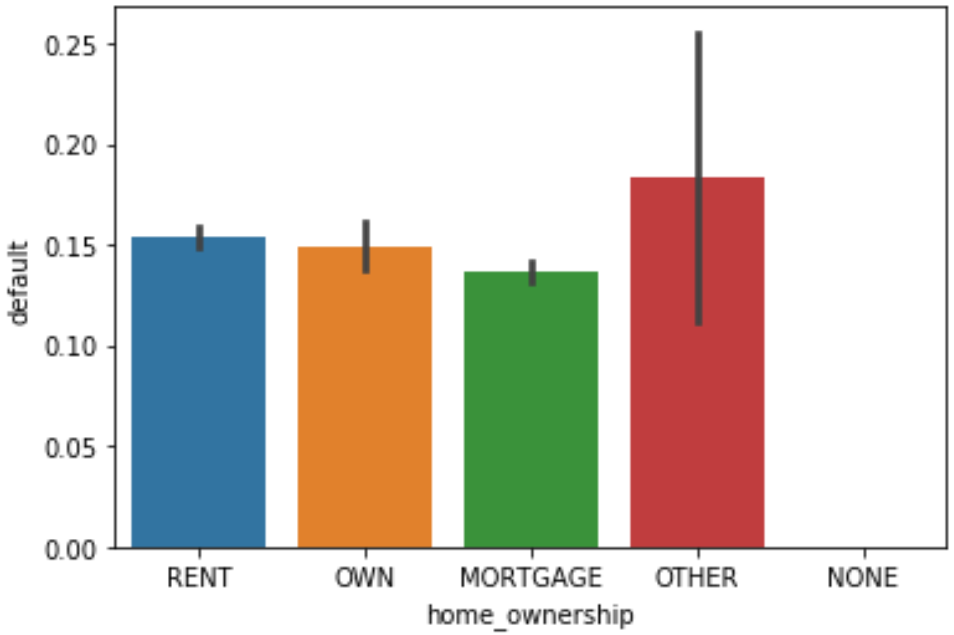
Univariate Plots :



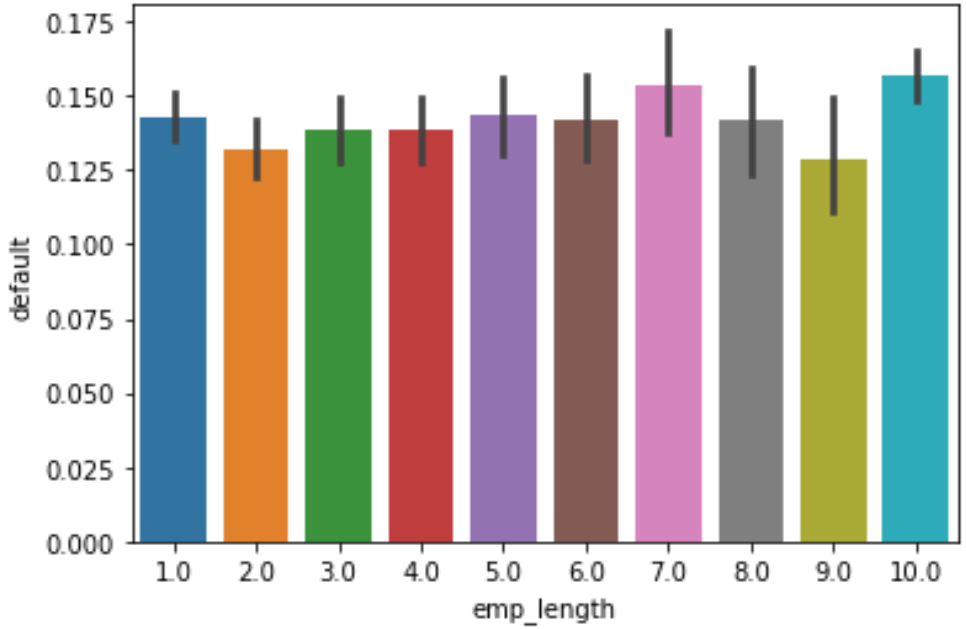
d) Plot of purpose of loan vs default

From above figure almost all the loan purpose have almost same default rate except for purpose with small business, thus we assume that there are little more chance for applicants get default with loan purpose of small business.

Univariate Plots :

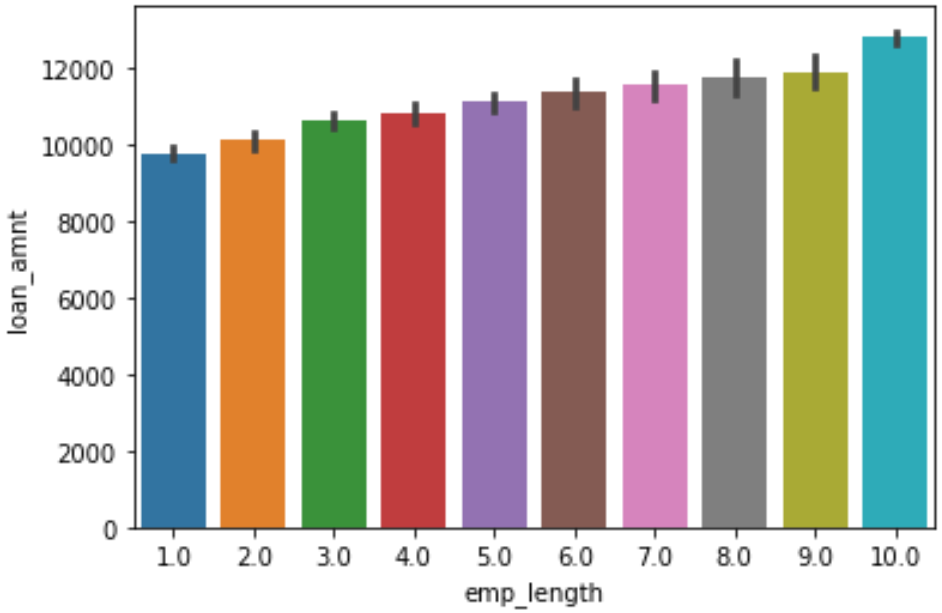


e) Plot of home_ownership vs default
From above figure we cannot say much about it as default rate is almost same for all categories of home ownership..

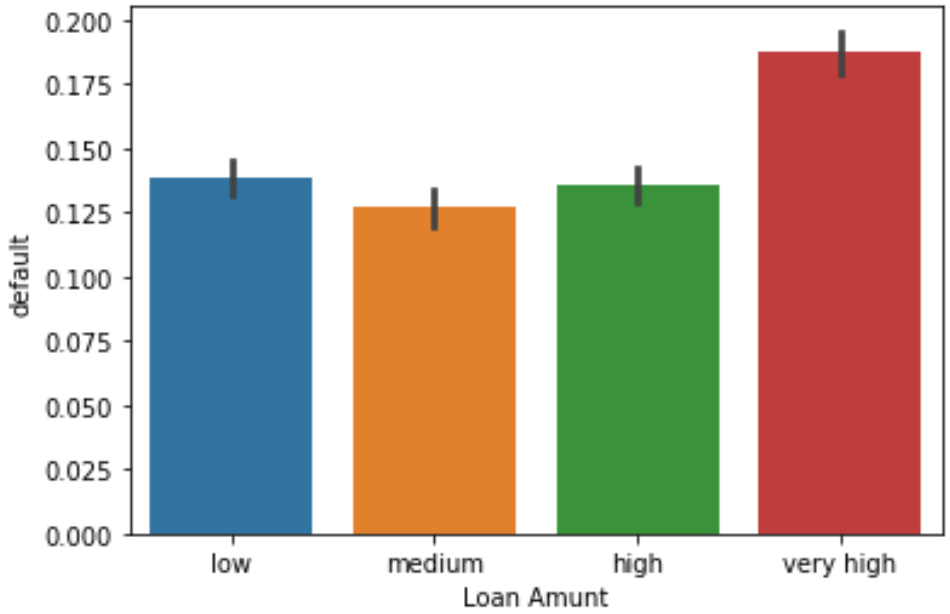


f) Plot of emp_length vs default
In above figure default rate is almost equally spread out for all categories.

Univariate Plots :

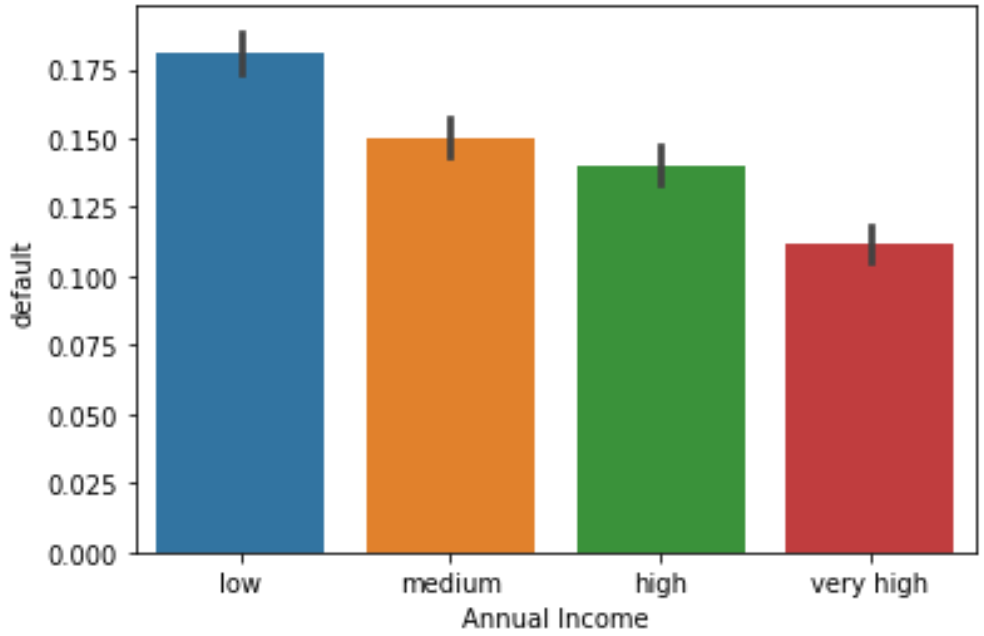


g) Plot of emp_length vs loan_amnt
In above figure we can notice that loan amount goes on increasing as the applicant's annual income increases

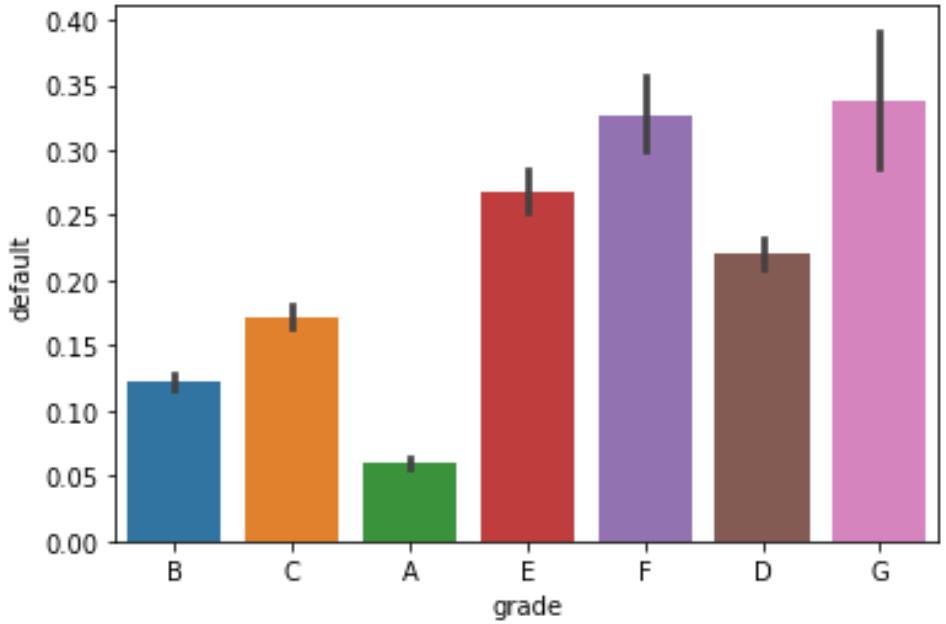


h) Plot of loan_amnt vs default
After binning the loan amount into 4 groups based on the percentile and then analysing it we obtained the above plot. From above figure we can say that applicant having high loan amount tends to default more as compared to others.

Univariate Plots :

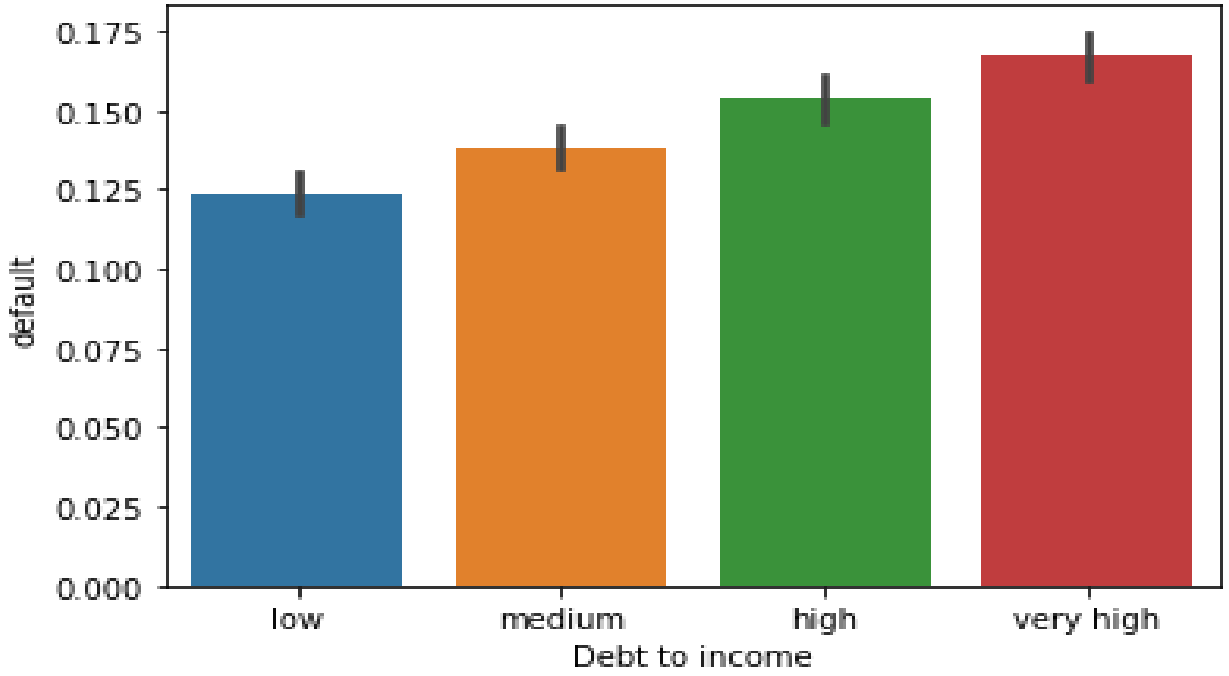


i) Plot of annual income vs default
After binning the loan annual income into 4 groups based on the percentile and then analysing it we obtained the above plot. Here in above figure the applicants having low annual income are more likely to default than others.



j) Plot of grade vs default
Here from above figure its clear that default rate varies from low to hight with respect to grades A to G respectively.

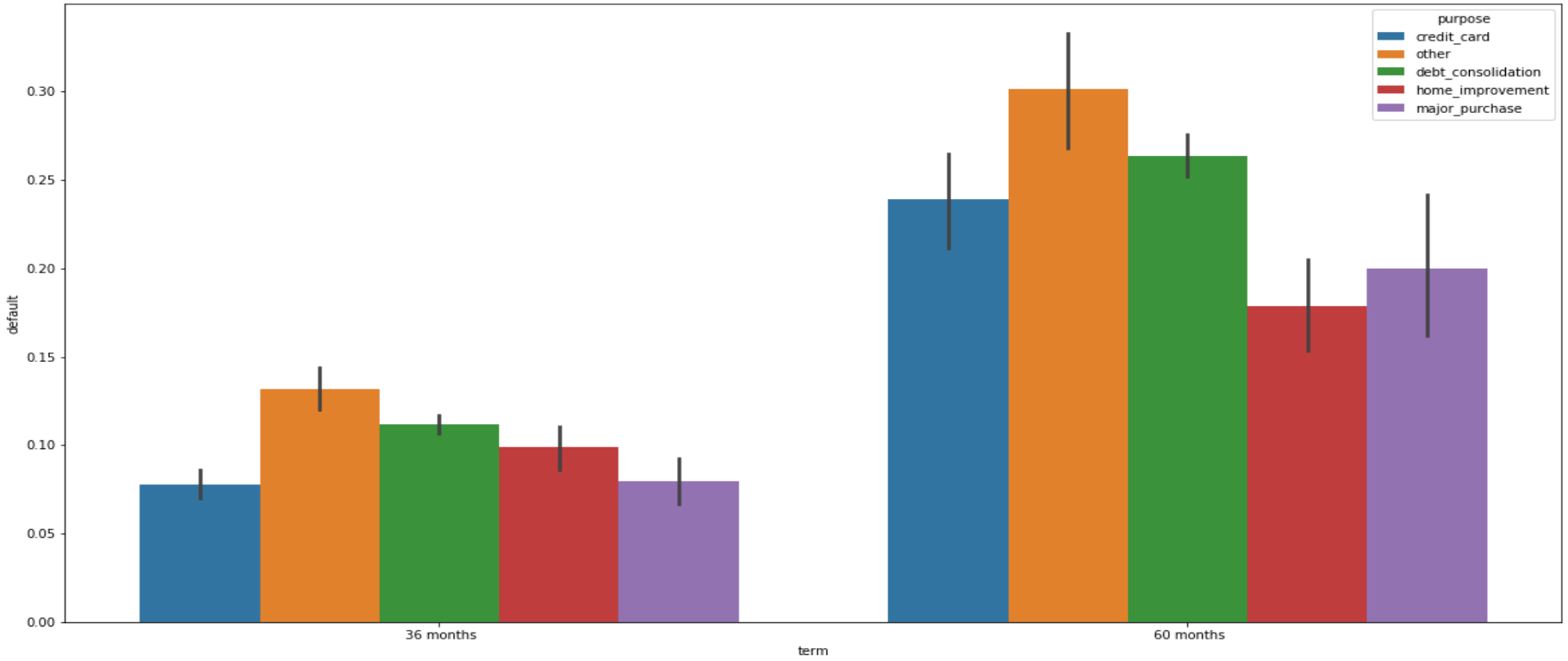
Univariate Plots :



j) Plot of term vs default

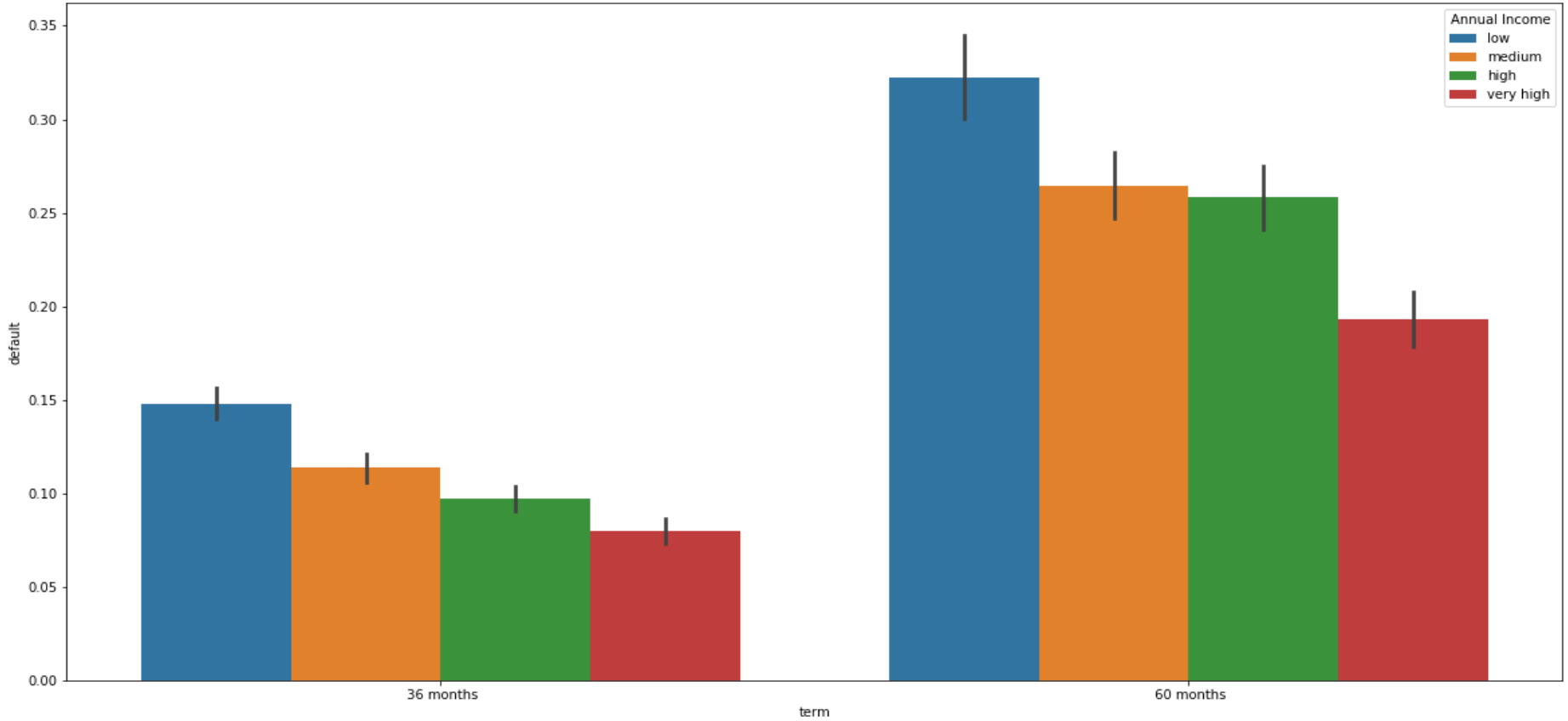
After binning the loan annual income into 4 groups based on the percentile and then analysing it we obtained the above plot. From above figure we can say that the applicants having higher debts are more likely to get default.

Bivariate Analysis :



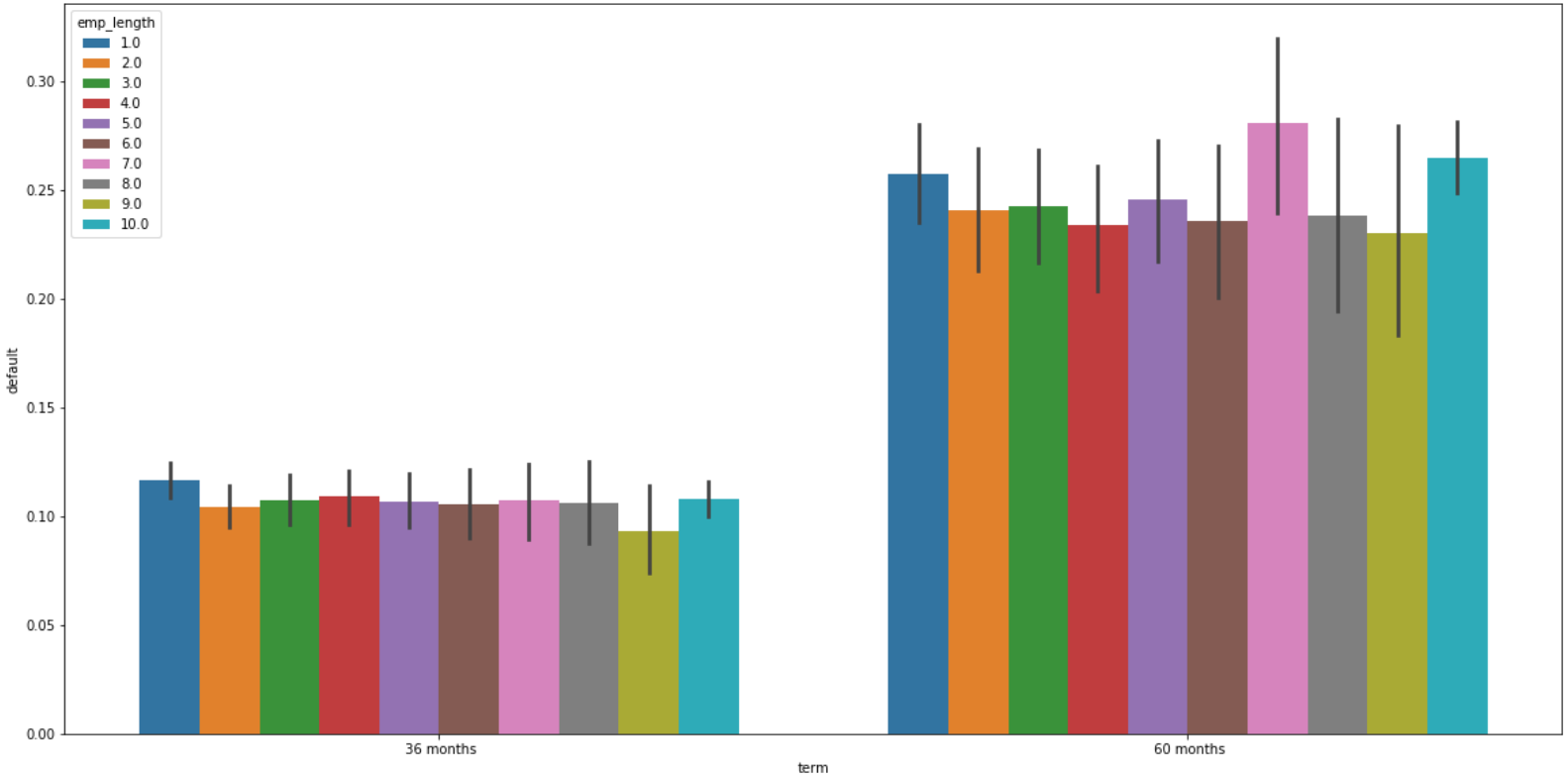
a) Plot of term vs default with hue as purpose
Here in above figure we can observe that the applicants having loan purpose of other and debt consolidation are likely to get default than other either in 36 or 60 months plan.

Bivariate Plots



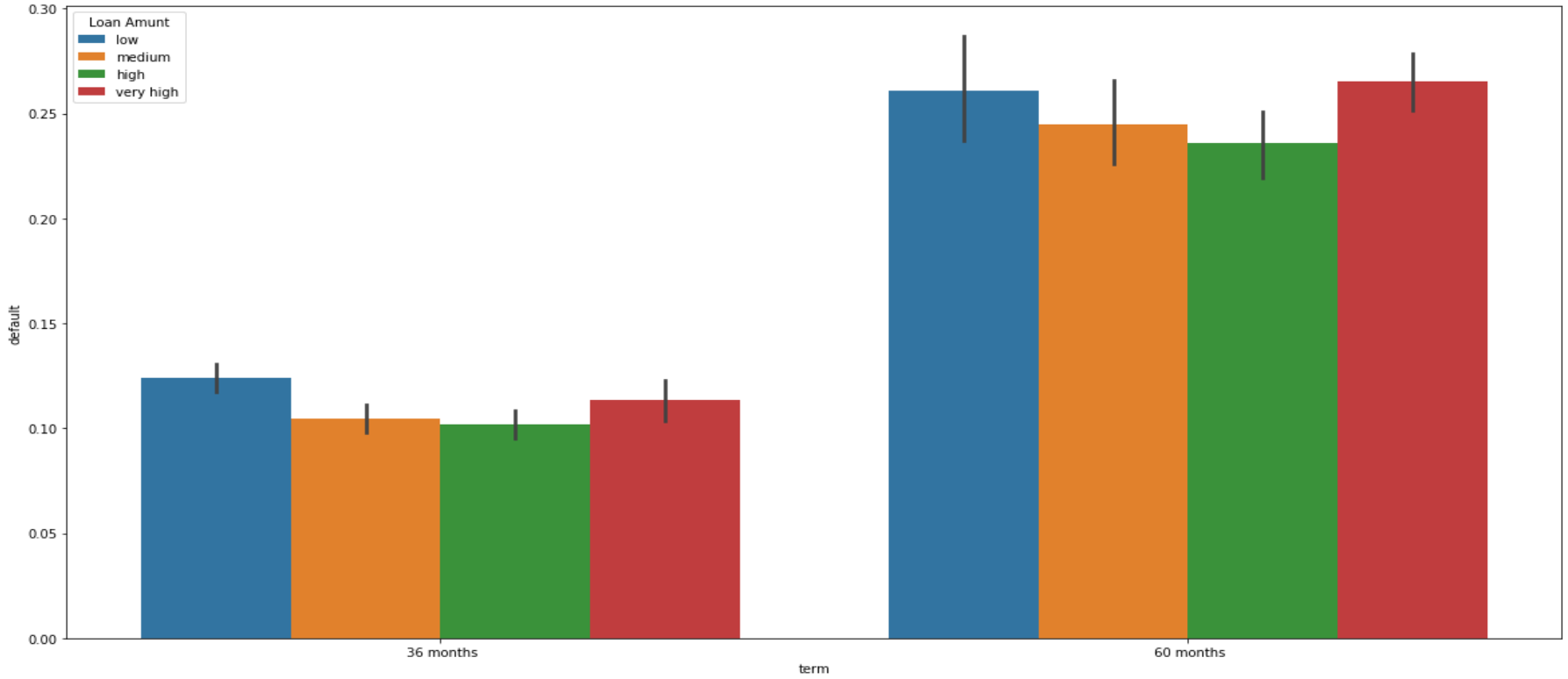
b) Plot of term vs default with hue as annual income
From above figure for both plan either 36 or 60 months loan plan applicants having low annual income are tends to default more.

Bivariate Plots :



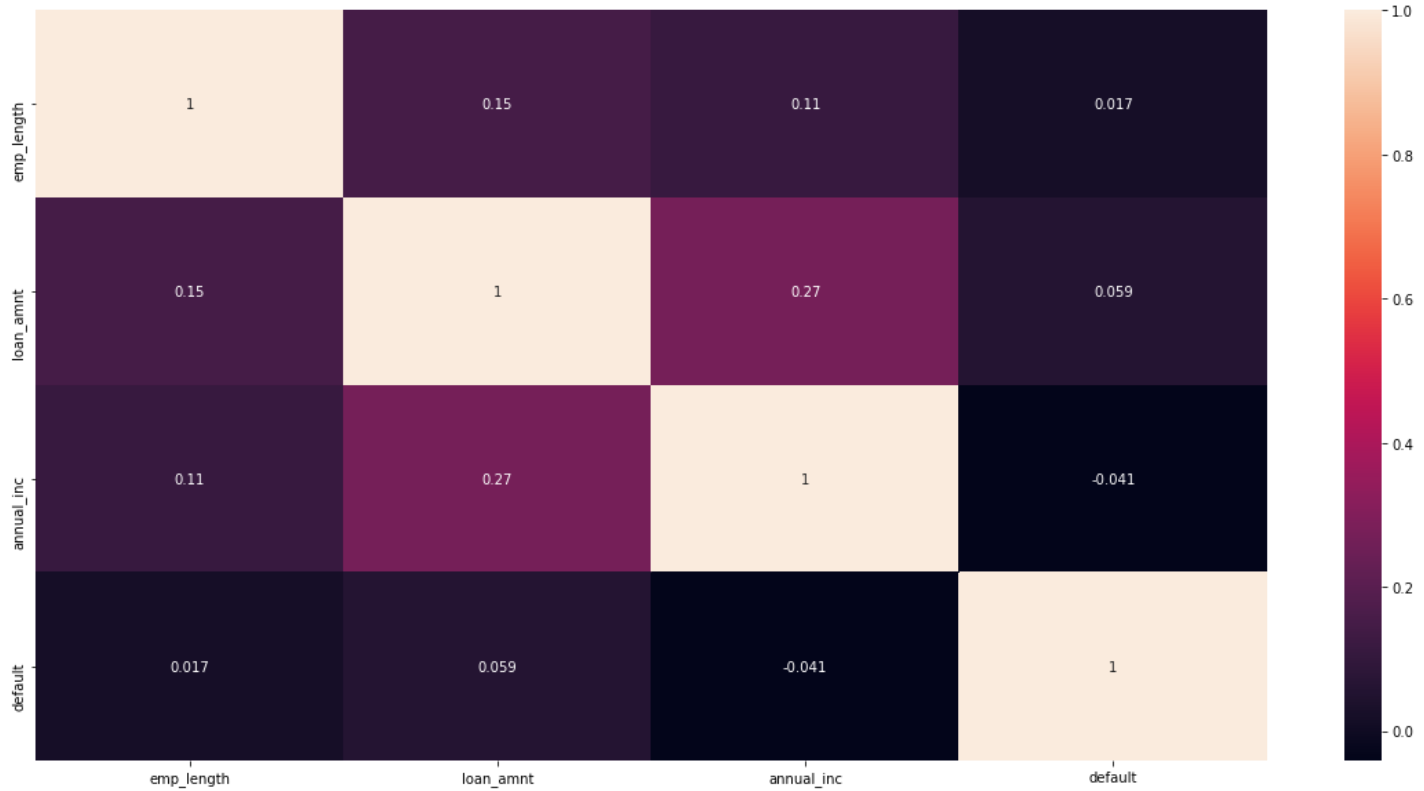
c) Plot of term vs default with hue as emp_length
In above figure default rate is equally spread for all employment length.

Bivariate Plots :



d) Plot of term vs default with hue as loan amount
Here in above figure the observation that is made is that there is not much of a difference that can be seen.

Heat-map of correlation between affecting variables



e) Plotting a Heat map taking few variables and getting a correlation.
From the above figure we can infer that annual_inc is the least correlated and loan_amnt is the highest correlated variables to the default variable.

Conclusion :

As per the analysis of the data provided, it can be concluded that the loan applicant tends to get default depends mainly on following factors :

1. Term : 60 Months
2. High interest rate
3. Low annual income
4. High loan amount
5. High Debt to Income Ratio