EDA GRAMENER

CASE STUDY

(Upgrad Project)

BY: AYUSHI SHARMA

:

The data given below contains information about past loan applicants and whether get "defaulted".

The aim is to identify patterns indicating that a person is likely to default which may be used to deny the loan, reduce the loan amount ,lend(torisk applicants) at a higher interest rate,etc.

The main objective of the analysis is to determine the conditions and situations that leads to an applicant being charged off or default.

UNDERSTANDING THE BUSINESS

- I work for a **consumer finance company** that specializes in providing various types of loans to urban customers. When the company receives a loan application, it has to decide whether to approve or reject it 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., they are likely to default, then approving the loan may lead to **financial loss** for the company.
-

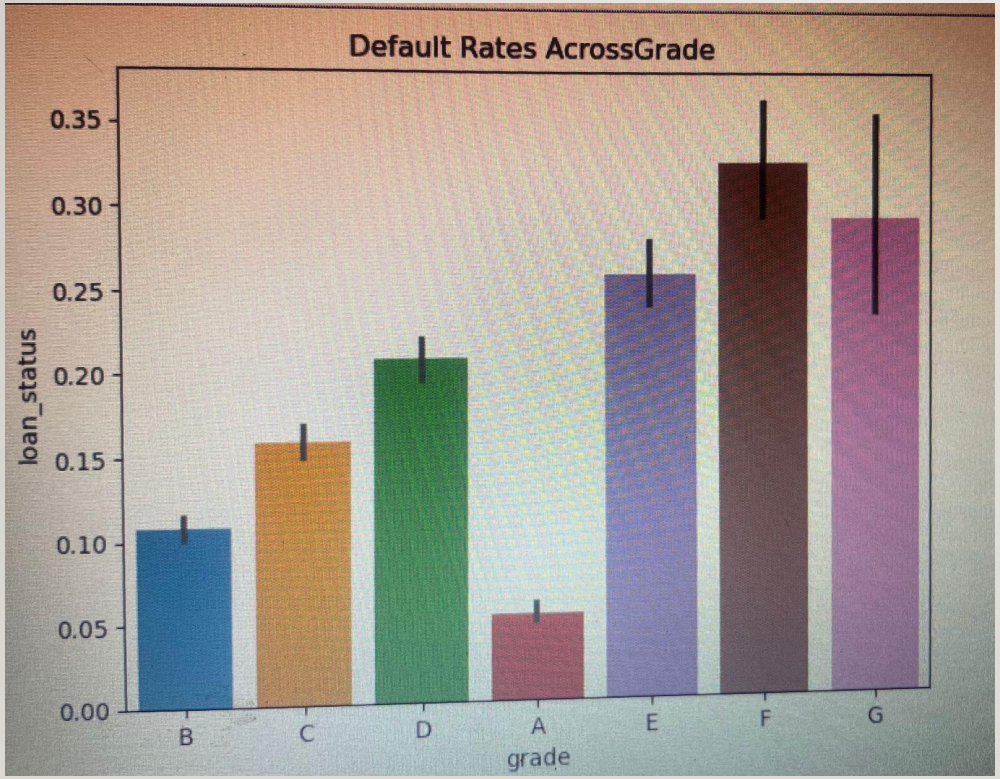
PROCEDURE

The analysis of the given dataset is divided into 4 main parts

1) Data understanding
2) Data cleaning
3) Data analysis
4) Conclusion

The task is done by using univariate and bivariate analysis of different columns of dataset

OBSERVATION FROM ANALISIS



Now of the 28 columns we need to find the ones which effect the target variable "loan status". We will do this by comparing it with other columns and by analysing each of these columns on their own.

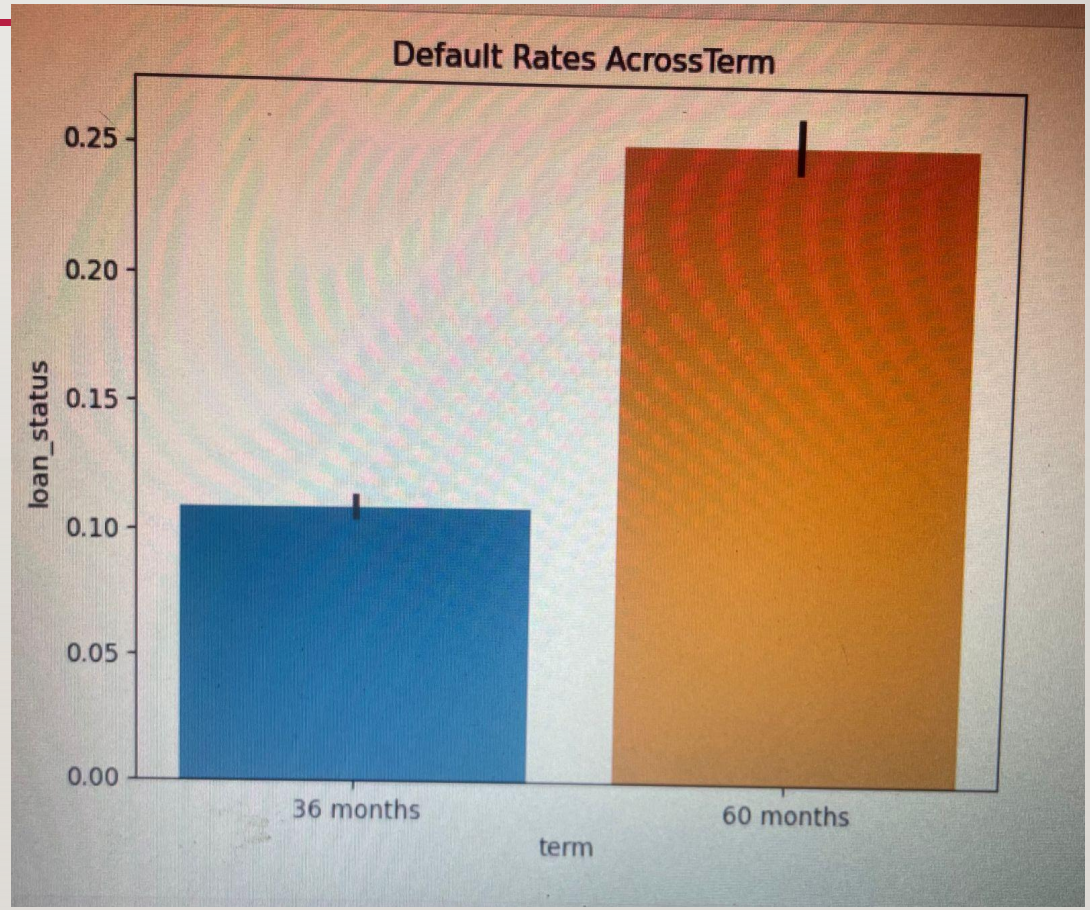
On analysing our loan_status we find out that the overall default rate is up to 14 percent

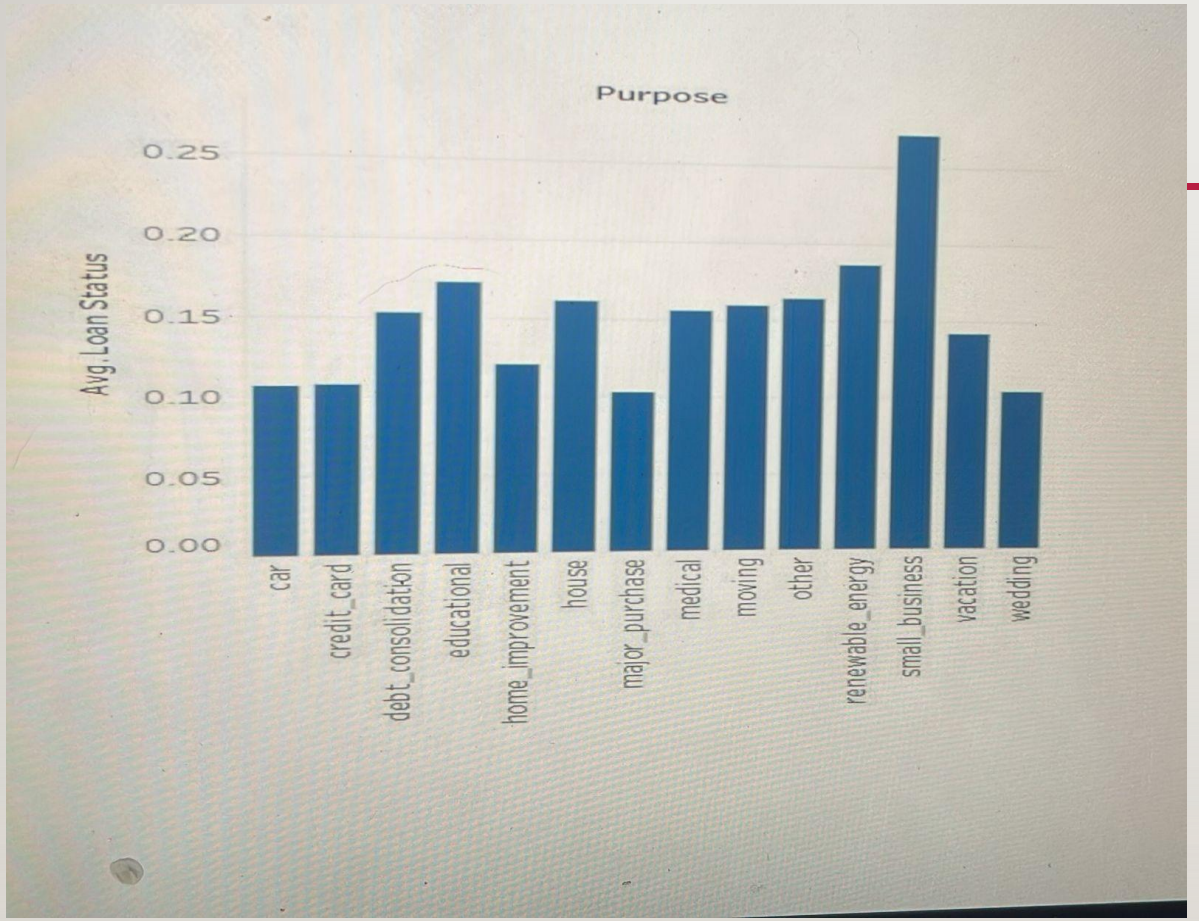
To start things off, lets look at all the categorical columns first.

It can be clearly seen that the risk of loan increases as we go from grade A to G, which is expected because of LC guidelines of assigning the grade.

DEFAULT RATE ACROSS TERM

From this it can be observed that loans of 60 months term tend to default more than 36 month term loans.

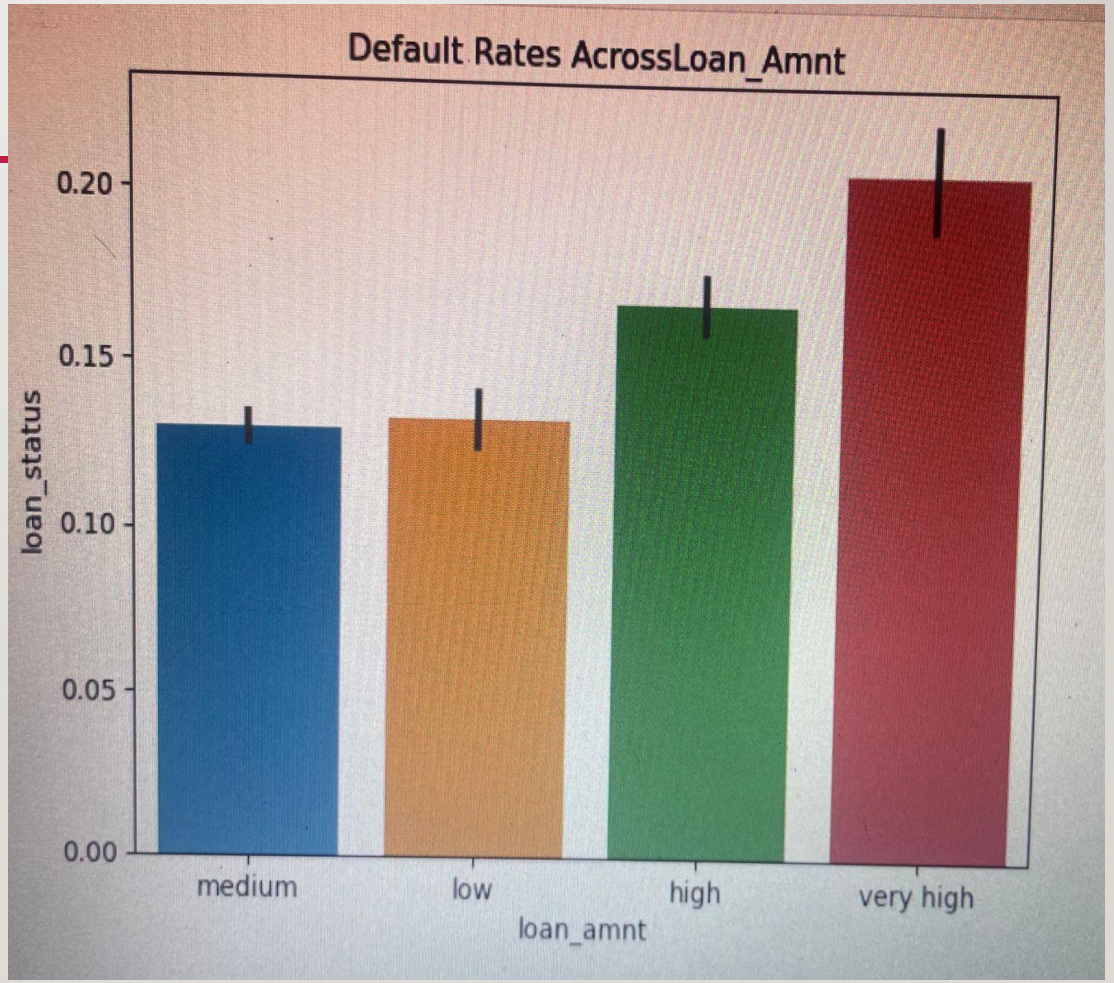


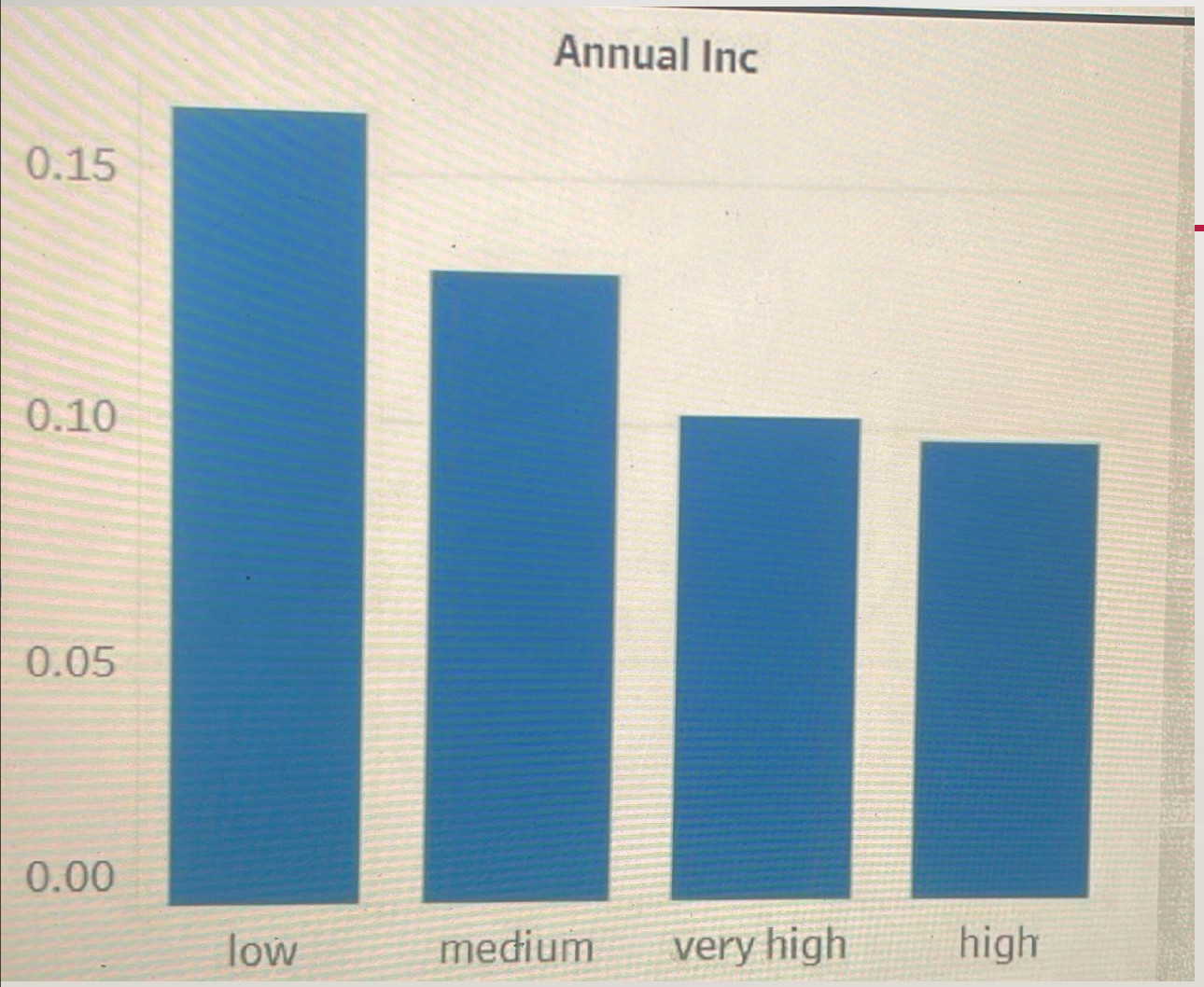


Plotting the purpose of loan shows that small business ,debt consolidation, education and renewable energy loans default more than any other category

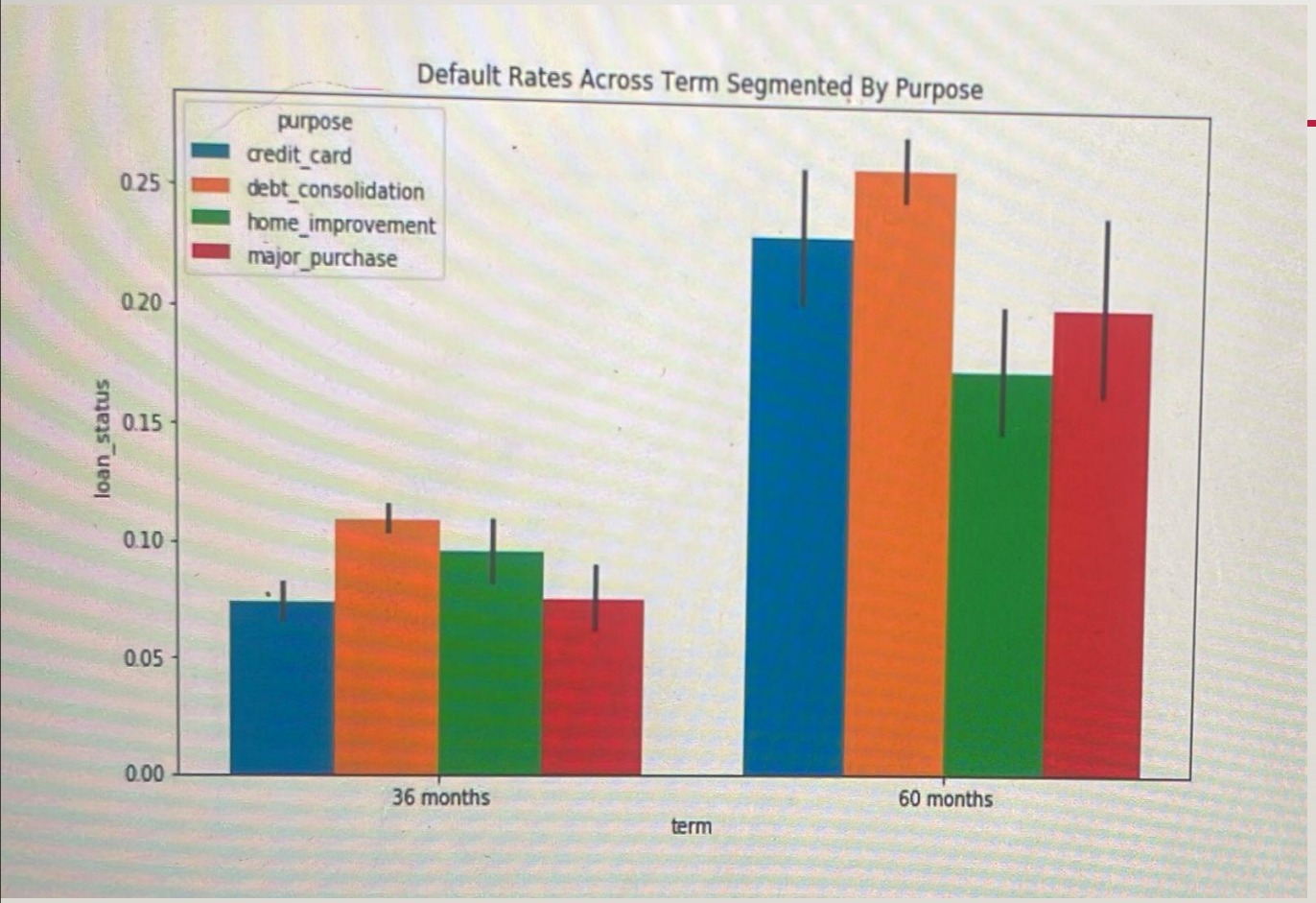
After analyzing categorical variables lets now look at continuous variables. We will bin these variables into different categories to plot them better.

As loan amount increases loan tend to default more.





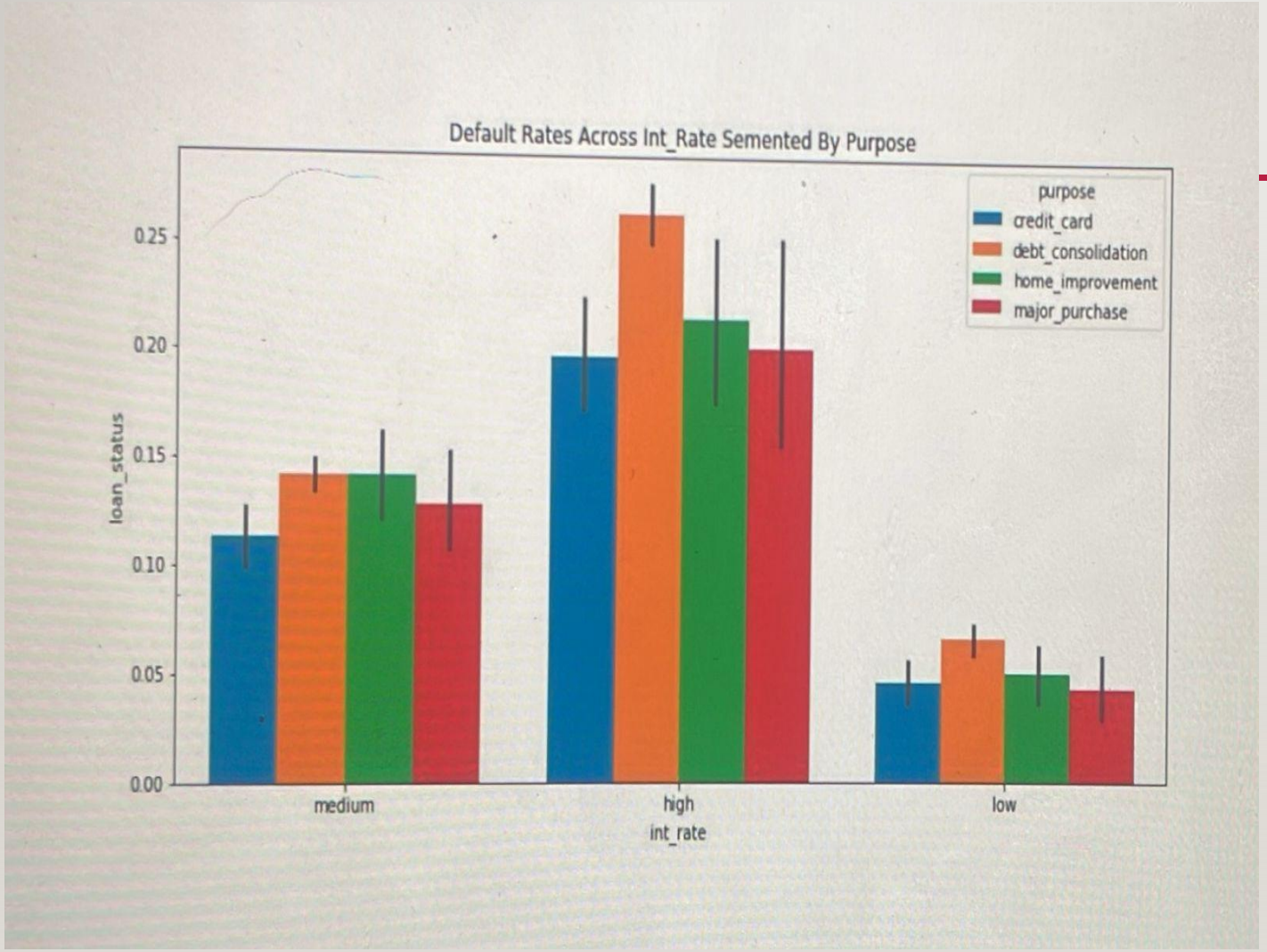
Annual income seems to inversely affect the default rate.



Now lets do some segmented univariate analysis

Lets analyse various categories with loan status, segmenting it with purpose.

First lets compare it with term



From above plots, a general trend can be observed that the debt consolidation loans have higher default rates in almost every other category

CONCLUSION

- The grade of the loan goes from A to G , the default rate increases.This is expected because the grade is decided by Lending Club based on riskiness. of the loan
- Term : 60 months loan default more than 36 months loan
- Verification status: Verified loans default more than not verified
- default rate increases the amount and funded amount increases
- Segmented univariate analysis doesnt yield much insight but it can be seen that debt consolidation loans tends to default frequently