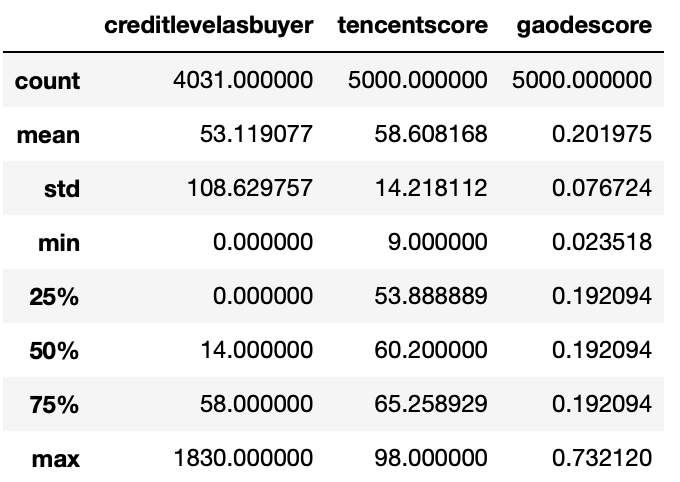
FIN3210 Fintech Theory and Practice

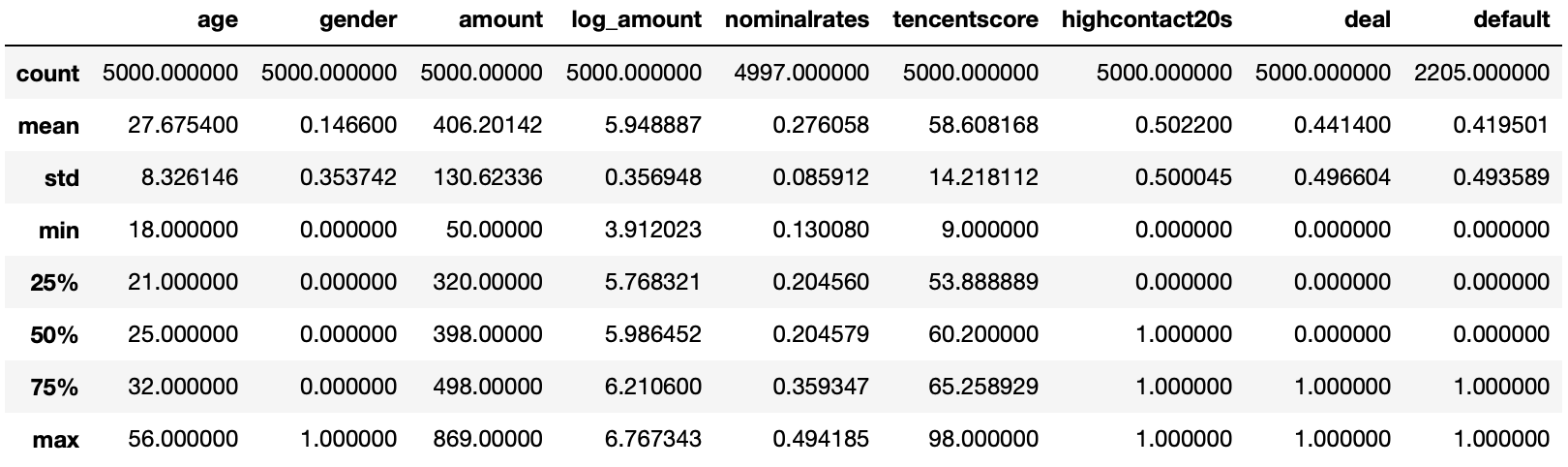
Report of Assignment 1

1. Present a table of summary statistics for the key variables including the borrower’s age, gender, loan amount, interest rate, credit scores, a dummy whether the borrower has a frequent contact, approval dummy, and delinquency dummy

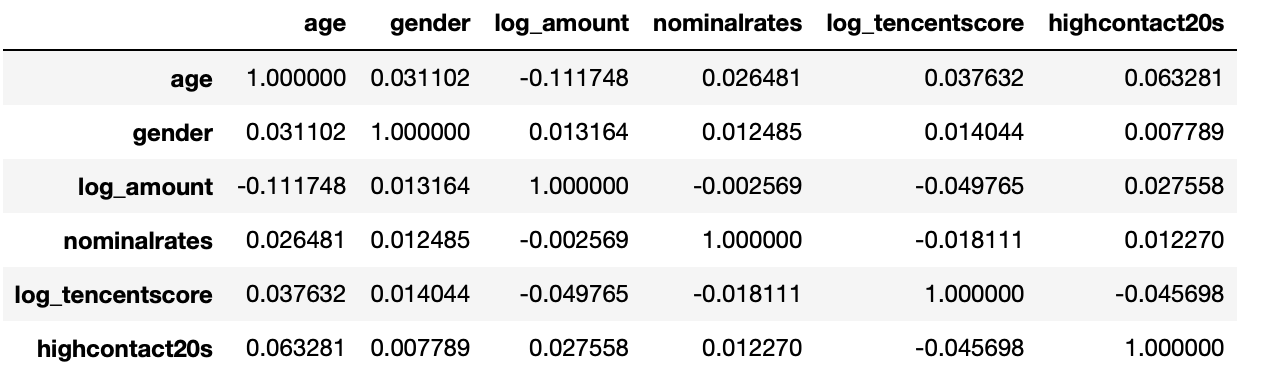
According to the given data, there are 3 credit scores, “*creditlevelasbuyer*”, “tencentscore”, and “*gaodescore”*. From the table below, there are too many missing value of “*creditlevelasbuyer”* and the standard deviation is quiet large. Also, the differences among values of “*gaodescore”* are not distinguished. In this way, I prefer to choose *“*tencentscore” as the credit scores.



The summary statistics are shown as followed. The descriptions of variables are shown in the footnote.[[1]](#footnote-0) From the statistics, we find that there exists some imbalance in the distribution of genders, which may cause some issues in the following regression.

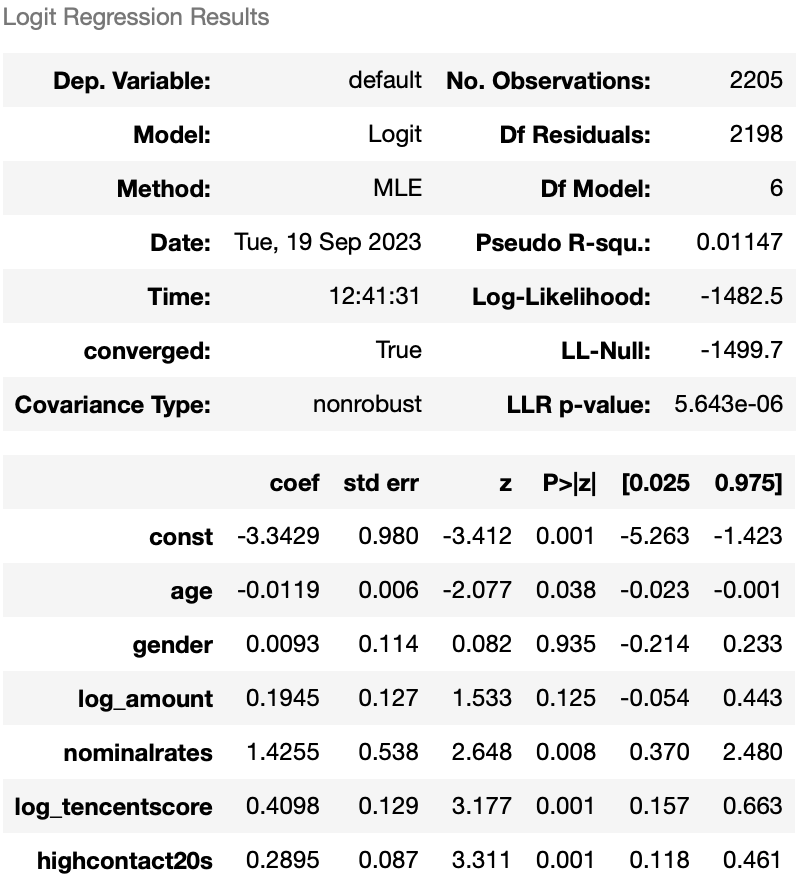
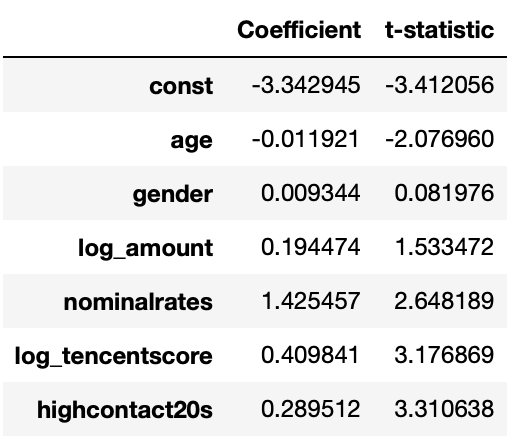


In the next 3 question, I take 'age', 'gender', 'log\_amount', 'nominalrates', 'log\_tencentscore', and 'highcontact20s' as the independent variables, after checking their correlation. Then, I fill missing values in 'nominalrates' with its median. And for the credit score and loan amount, the logarithm value is take in regression analysis.



1. Perform a logit regression and examine the relation between the delinquency likelihood and credit scores.

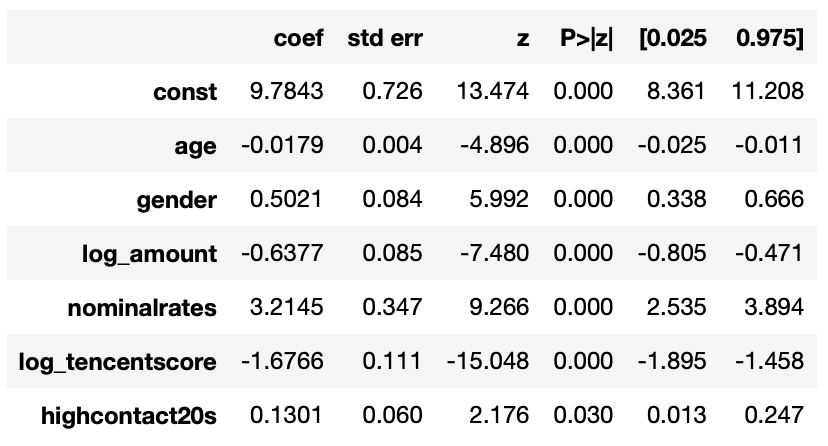
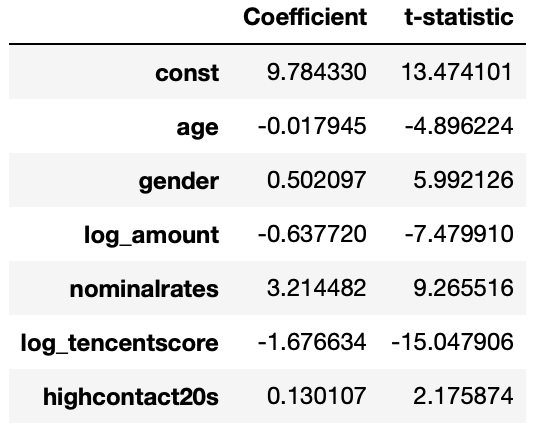
The size of sample is 2205, after dropping the NaN value in “default”. And the result of logit regression is showing below:

The coefficient on logarithm value of Tencent score is 0.4098 and *t-statistics* is 3.1769, which is significantly positive in the model, aligned with the notion that risky borrowers tend to engage in loan delinquency.

1. Perform a logit regression and examine the relation between the loan approval likelihood and credit scores.

The size of sample is 5000. The result of logit regression is showing below:

A borrower with a higher risk profile (*log\_tencentscore*) is less likely to be approved for his/her loan application. The coefficients on credit score is -1.6766(outlined in red), and is significant at one percent level (*t-statistics*=-15.0480*)*.

1. Perform a logit regression and examine the relation between the loan approval likelihood and the dummy whether the borrower has a frequent contact.

The regression model is same as that in question 3. We find that the coefficients on dummy variable of frequent contact is 0.1301(outlined in green), and is significant at the five percent level .

1. (1)Gender: 1 for Female, 0 for male; (2)Amount: of loan principal in thousands of Chinese Yuan; (3)Log\_amount: take the logarithm value of amount; (4)Nominalrates: interest rate of loan on annual basis; (5)Tencentscore: the credit score, and logarithm value is taken in regression analysis; (6)Highcontact20s: An indicator variable equal to one if a borrower has at least 3 frequent contact (last for more than 20 seconds), and zero otherwise. [↑](#footnote-ref-0)