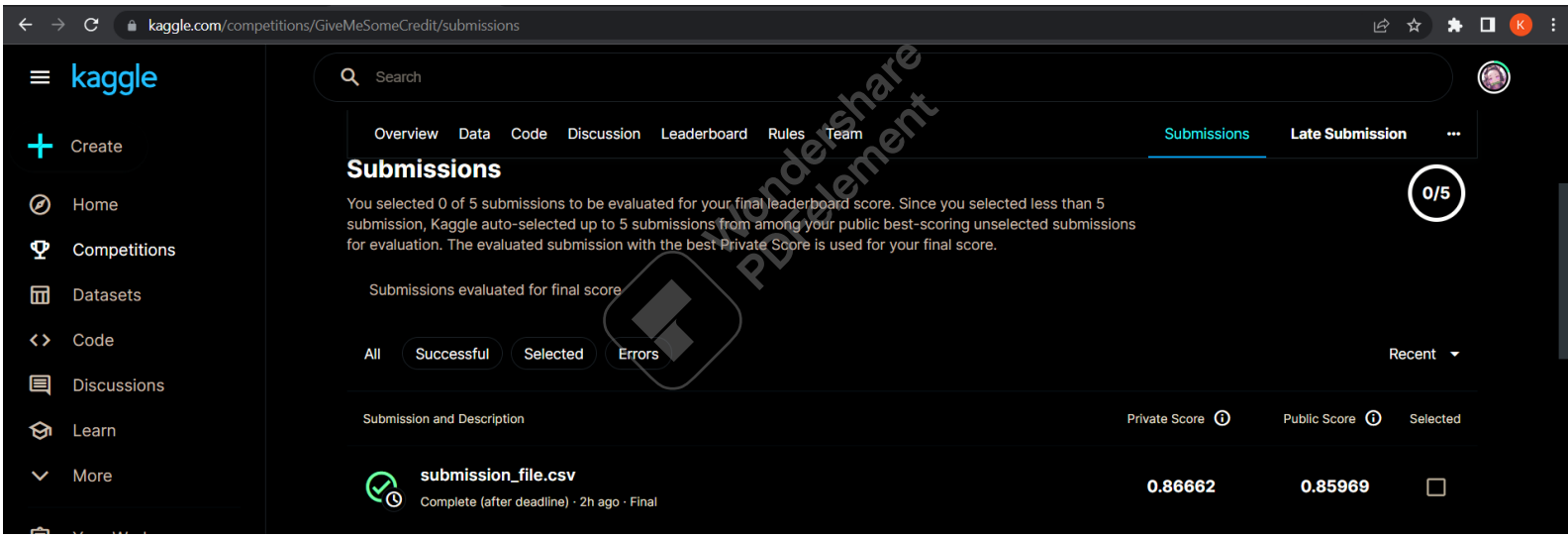
Binance DS Assessment

Give me some credit

1) Data Analysis/Screening/Cleaning -
   i) named the unnamed column as id
   ii) Visualised distribution of SeriousDlqin2yrs and relations with other fields
   iii) found some very unusually high debt ratio so analyzed it further and consequently removed those outliers
   iv) plotted pairplot to show relations between different fields wrt to target variable
   v) Saw correlation between the target and other fields to find further insights
   vi) found and imputed null values with median values
   vii)normalized the data
   viii)detected and removed outliers
2) Model -
   i) Tried implementing various models but XGboost gave the best results
   ii) Predicted the output after inserting in the trained xgegressor
   iii) Performed hyperparameter tuning using gridsearch to optimize the max_depth,n_estimators,lr

Insights -
1)NumberOfTimes90DaysLate, NumberOfTime60-89 DaysPastDueNotWorse, NumberOfTime30-59DaysPastDueNotWorse share the same values, specifically 96 and 98.
We can see that sharing the same values of 96 and 98 respectively is not logical since trivial calculations can reveal that being 30 days past due for 96 times for a single person within a timespan of 2 years is not possible.
2)There seems to be more younger people defaulting and the distribution seems fine on the whole
3)Distribution of monthlyincome values is skewed, we can consider imputation with median.
4)2.5% of clients owe around 3490 or more times what they own

Results -
i) After a lot of tuning and making changes in the model I was able to get a score of 0.86662
ii) The first on the leaderboard was at a score 0.86955
iii) I was unable to get my rank as submissions were not allowed for ranking anymore as this contest was 11y ago
iv) I believe my score would have landed me around the rank of 100
v) This was an incredible opportunity for me to learn and take my passion in Data science to the nect level :)



My Score