REPORT: CampusPulse

Level 1. This was a rather easy task, but for OCD purposes I attempted it AFTER I solve data integrity and null fills
(luckily data was very integrous): To deduce feature I looked at their other top correlations.
Feature 2 seemed intuitive, correlated with grades and slightly with sex; looked at it, saw how more feature 2 increased avg
grade and decreased stdev but that you could have high grades even studying less. (IQ factor clue). Also girls (studied)
more than boys (stereotype) so... study time
Then I visualized mini decision trees, Looked at Feature 3, immediate hint for it to be Walc, very corr with Dalc and gout,
same scale
Feature 1 a bt confusion from decision tree but the graph made everything clearer (and values ranging 16 to 22 helped a lot)

Level 2. Luckily data was very very hygenic, verified more when I casually mapped all the string categorical data ordinal
number wise and it all hunky dory. also there were like 8 columns which had non zero NaN in rows so... easy task.
I... just look: I did think whether Fedu not listed wld be 0 or G2 not listed wld be 0 but then- NO there were over 15 G1s
with NaN G2 so I filled EVERY null col with a LinearRegr model.

Level 3. my ipynb shows this better...

Level 4. oof...
tried all models: randomforest, xgb, lgbm, catboost, logistic regr: it was like 60% F1/acc is the speed of light...
tried adjusting threshold thru prec rec but no use: often it reduced metrics.
tried changing number of estimators/iterations (100 to 1000 by step 100 then 10 to 100 by step 10): again little use, but I did find the maximum by a tiny margin:
F1 score: 20 estimator random forest (0.62)
Accuracy: 40 estimator catboost (0.66)

Level 5. ipynb.