When the SOP dilemma occured, I was blown out of my mind. i thought I failed as an analyst, but I think I may have come with something okayish:

I arranged the data in ascending oredr of Chances and calculated the coefficients for 1st 70% of the people(after sorting, ofcourse):

This gave me a +ve coefficient for SOP: +0.000896

In the same run, the max coeff belonged to CGPA: 0.103356 (look at the difference)

The 70th% student in the chances column(The last student for this run) had a CGPA of: 9.0

SO, SOP till this point is a +ve factor, but negligible

Then, I ran the test for 80% of students, with the same sorted csv:

Coef for SOP was: -0.000354 coef for Highest contributing factor (CGPA): 0.101128 student with Highest CGPA is: 9.06

I saw almost every data point from 70% range uptill the end as chances increases, the CGPA increased but people also got poor SOP scores

MY regressor is a fool.

Since it is linear regression, it has evaluated coefficients for each factor independently because, th sop scores fall drastically sometimes even with high enrty chances, it imposes a penalty for high SOP score.

Now, something I want to recall: remember seeing a good enough streamlinig and clustering in the heatmap of CGPA when chances increased but how ugly the “spread” was before that? There were some people with splendid chances also whogot a low CGPA

WELL, THIS IS IT…!!!

When your CGPA isn’t good, you make a good SOP for compensation but when your CGPA is sky-rocketing, SOP is just out of question.

The independent assignment of coefficients, I believe, is what motivated AI developers to move ahead of LR and reach neural networks, where almost every factor is inter-related. Looking forward to that…!!!