Analysis Report IP5 Evan Gousha

In this assignment, we were to create regression models based off of our Self Awareness Data Set in which we documented variables such as hours of sleep, hours of exercise, hours of study, hours of DS study, hours of work, what day of the week it is, and how stressed we were at the end of the day. The goal of these models was to predict how we felt at the end of the day based off the independent variables. Now going into this, I'd have to say I live a pretty stress free and easy life, relatively regardless of the variables that I listed above, so I didn't have much faith in either the python or R models considering my feeling at the end of the day stayed pretty consistent. First, looking at the model that was generated by python, the r-squared came out to about 0.206 which shows little positive association between the variables and my feeling at the end of the day. This is backed up by the p values from all the variables, with only the day being Tuesday being statistically significant, with a p-value of 0.045. However, with this being said the regression model was relatively accurate when predicting what my feeling would be, with most being off by less than 1 with the exception of the one day I had a 5 as my feeling. On to the model generated by R, where my r-squared value was even lower with a value of 0.13, further indicating little association between variables and my feeling. Again, this is also shown in the p-values, with none being statistically significant, with only the day being Tuesday or Wednesday coming close with values of 0.06. With the predictions, this model was actually close to my actual values, with a mean squared error of about 0.81 compared to python's MSE of 1.57. I think this goes to show that for me personally, my stress isn't really tied to these variables, and most likely controlled by something outside of school and exercise, if controlled by anything at all. Despite the models coming close in their predictions, I think it came as a result of coincidence rather than actual accuracy, given that stress level stayed relatively consistent as it is. Anyways, I find this method of statistical forecasting interesting, especially when it can accurately guess data, even if I'm not sold that it accurately guessed mine based off of the regression models.