For your Lab of 4/20/2022

Please submit the following.

A screenshot of your submission to kaggle.

Your code that you used to build out the model for Kaggle.

Lab Expectations:

Jupyter File Code(10pts)

5 Points: Quality of Data Cleaning

5: Nulls and Outliers are being handled appropriately. The selected features have reason to them, and they are being scaled and normalized if need be.

3-4 : The Data set has been cleaned, but there is still some variation of values between rows and columns. Nulls and NaNs were replaced or taken out, but no justification was given as to why. The Data can be worked, but it is difficult and time consuming.

1-2 : There was little to no effort made to clean the data or make it easier to work with. Working through the dataset was difficult because steps were not taken here to simplify it.

5 Points : Clean logical application of the Data Science Libraries.

5 : Pandas methods and functions were invoked with purpose, and redundant code was taken out of the project and replaced with more efficient code. Code that was going to be used again was encapsulated in functions, and called throughout the program. Use of Vectorization was included when the situation called for it, as well as mapping and applying. When applying regression we are using the correct loss function as well and tuning the hyperparameters for the optimal regression line if SGD was used. We are tuning our parameters and models using the appropriate methods.

3-4 : The libraries were applied, but the application was difficult to read, and it was hard to justify the uses of the methods and functions being used. An attempt was made to refactor code, but improvements were still able to be made.

1-2 : No reasoning was used when calling methods and functions, and the libraries were misused.

Jupyter File Code Structure(10pts)

3 Points : Sense of flow within the Code.

3 : There was a clear flow to the program, and following through it was easy because logical steps were being followed from one cell to another, and the reader was being facilitated through the cells.

2: The flow was difficult to follow at times, but overall it was clear enough to follow alongside the project.

1: There was difficulty following through the project, and there was no attempt made to provide clarity and justification as to what was being done, when it was being done, and why it was being done.

3 Points : Segmented and Commented Code

3 : Lines of code were clearly commented on, and thoroughly explained what was happening in the cells. Cells were divided by logic, and each cell encapsulates a single change happening without in our project.

2: There was an attempt to comment on the code, but not enough clarity was provided, and there was ambiguity as to what was going on in a cell. Cells are overloaded and contain different applications of logic and code, and are not properly encapsulated.

1: There was little to no attempt to comment, and the cells are hard to follow along logically.

2 Points : Clear and labeled Visualizations

2: The visualizations are clearly labeled and sized, and it is clear as to what the visualization is meant to represent. A Write up is provided when needed.

1: The visualizations are not labeled and sized, and there is little to no understanding gained from the visualizations that were generated.