ReadMe

How to run the program:

The cells are listed in sequential order with a brief synopsis of functionality.

1. Installs the uci repo api
2. Uses the uci api and loads in the data into an X and y dataframe
3. Performs nan checking

4a. Shows distribution for binary features

4b. Shows distribution for non-binary features

5a. Standardizes the data

5b. Normalizes the data (CHOOSE 1 not both)

6. Creates correlation heatmap to find desired features
7. Drops undesired features
8. Splits data into 80/20
9. Defines model. The top is the default model which should be used to find optimal hyperparameters. The bottom is the tuned regressor which is set to the optimized hyperparameters.
10. The automated hyperparameter tuning that finds optimal parameters. The code block below it prints them.
11. Fits and shows performance metrics for SGDRegressor
12. Fits and shows performance metrics for OLS model