Predict the results of 2017/2018 soccer season for English Premier League.

A. Build a linear regression model which predicts the total season points of each team based on the average market value of its players (can be found at https://www.transfermarkt.us/)

1) Use the results of top 10 teams as a training set. Plot the data points and the best-fitting regression line. Apply the trained model to the bottom 10 teams and plot the real and predicted results.

2) Use the results of randomly selected 10 teams as a training set and apply the trained model to the remaining 10 teams. Repeat this experiment 20 times and calculate the predicted result of each team as an average result over all experiments. Plot the real and predicted results.

Which method gives more accurate prediction?

B. Build a linear regression model which predicts the total season points of each team based on 2 features: average age and number of foreign (non-English) players (can be found at https://www.transfermarkt.us/).

1) Use the results of top 10 teams as a training set. Apply the trained model to the bottom 10 teams and plot the real and predicted results.

2) Use the results of randomly selected 10 teams as a training set and apply the trained model to the remaining 10 teams. Repeat this experiment 20 times and calculate the predicted result of each team as an average result over all experiments. Plot the real and predicted results.

Compare models A and B. In your opinion, which of them gives the better prediction?

Submit the following files:

- source code of your script
- Excel or scv file with the data used by your script