Predict the results of 2018/2019 soccer season for Italian Seria A League.

1. 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/>)

A1) 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.

A2) 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 the prediction that is more accurate?

1. 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-Italian) players (can be found at <https://www.transfermarkt.us/>).

B1) 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.

B2) 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 the prediction that is more accurate?

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

Submit the following files:

1. Source code written in **Python only**
2. Excel or *scv* file with the data used by your script

Please, archive these files, and then add an extension *.txt* to your archive.

For example, you have archived your files of the source code into *SuperStudent.zip*

Then, you have to rename it as ***SuperStudent.zip.txt*** and upload on iCollege.

Please, **NO google drive links this time**!

The report as a *pdf* or *docx* file. **Do NOT archive the report file**!

1. which must include
   * For A1 and A2, plots of the real and the predicted results
   * Answer to the question which method (A1 or A2) gives more accurate prediction
   * For B1 and B2, plots of the real and the predicted results
   * Answer to the question which method (B1 or B2) gives more accurate prediction
   * Comparison of the models A and B

In total, there must be only 2 (TWO) files in the submission:

1. Renamed archive (simply add *.txt* to the full name of your archive) with the source code written in **Python** and *Excel* or *scv* file with your data.
2. Report. Again, **do NOT archive it**!

**All deviations from this scheme will be penalized!**

**Do not be late either; you will have -20 points for each extra day!**