**Group 5: Additional description on running Code files and which datasets to use**

For Manchester\_City.ipynb:

* We have used 'male\_players (legacy).csv' for Manchester City analysis. This has been directly imported from the Kaggle dataset called 'male\_players (legacy).csv' which can be accessed here: https://www.kaggle.com/datasets/stefanoleone992/fifa-23-complete-player-dataset?select=male\_players+%28legacy%29.csv. For the second part of this problem, we used man.csv which is a dataframe created from the first part of the analysis. (See code for more details).

For Manchester\_United.ipynb:

* We used 'players\_dataset.csv' for Manchester United analysis of the main model. We downloaded the Kaggle dataset called called 'male\_players (legacy).csv'. It can be accessed here: <https://www.kaggle.com/datasets/stefanoleone992/fifa-23-complete-player-dataset?select=male_players+%28legacy%29.csv>. After exploring the variables, we did some data pre-processing based on the initial dataset, and then we got the 'players\_dataset.csv' as our main dataset.
* After we selected the 23 players, we generated an additional dataset called ‘starting\_lineup.csv’ to do our following optimization model: the formation of 4-2-3-1 and 4-3-3, which is the 2nd and 3rd part of our analysis.

For two Germany file: 2018 Germany.ipynb and 2020 Germany.ipynb:

* We only used 'male\_players (legacy).csv' for whole German team analysis.