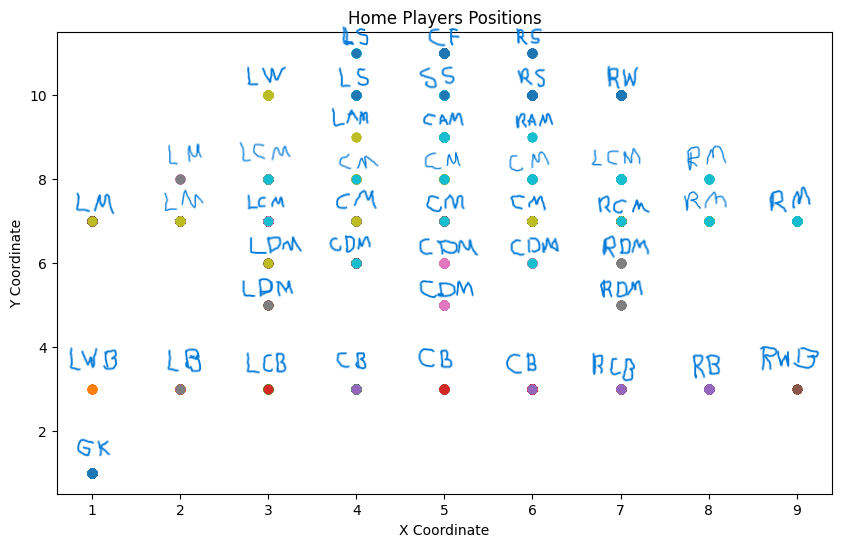
Which are the 10 most and least valuable players for the session 2014-2015?

The files for this tasks are: “map\_of\_positions\_14-15.ipynb”, “most\_valueble\_player\_list\_submission.csv”, “least\_valueble\_player\_list\_submission.csv”

An extra file where I worked on first: “map\_of\_positions\_all\_seasons.ipynb”

Metrics I used and steps I took to find my answer:

* I started by converting the “player\_X#” and “player\_Y#” to their original field position by using my knowledge of football and the help of some friends. I ended up with this as a result:



* After this I also grouped them into 7 different groups: Goalkeeper, Defender, Defensive Midfielder, Midfielder, Attacking Midfielder, Forwarder and Striker respectively. This I put into a new column in the “epl\_matches” dataset.
* Then I made sure that all the “player\_id” values were intergers because some of them were floats. And I need them to be all the same before I connect them to the other dataset.
* Then I merged the 2 datasets together with the “player\_id” as the main connecter. Mainly keeping the years, “position\_name” “position\_category”, season, “match\_id” and all the individual player stats.
* Then I made sure that I only select the rows that have the season == “2014/2015”.
* After this I calculated per group category their rating (which all need to be done differently, because defending does not matter for an striker for example). Here is a more deep explanation on how I did it:
* I figured these coefficients out based on a mix of common football knowledge, real-world player traits, and my own observations from football analytics and coaching. I realized that, generally, attributes like **vision** and **dribbling** are far more influential for attacking midfielders because they play a crucial role in creating opportunities and progressing the ball forward. On the other hand**, agility** and **ball control** are still important but aren't as vital as vision or dribbling.
* I also took some inspiration from football management simulations like [Football Manager](https://www.reddit.com/r/footballmanagergames/comments/13gfru4/attributes_weights_theorycrafting/) which reflects how attributes are weighted in real-life football. A lot of football analysts and coaches agree on the importance of these key attributes for attacking midfielders.
* If I wanted to validate or tweak these coefficients further, I could experiment with statistical analysis on a large dataset of player performances, which would show me which attributes correlate most strongly with success on the pitch for attacking midfielders. I could also dive into more detailed football analytics platforms, like [Opta](https://optaplayerstats.statsperform.com/en_GB/soccer), [StatsBomb](file:///C:\Users\thijn_6wrysbu\Documents\StatsBomb) or [Whoscored](https://www.whoscored.com), to get a better understanding of how player stats break down by position, or check forums like [FM Scout](https://www.fmscout.com) to see what the community thinks about player ratings.
* Lastly I took the mean of all the matches that each individual player scored rating points on and then I was left of with this result:

|  |  |
| --- | --- |
| Top 10 most valuable players | |
| player\_id | rank |
| 30613 | 1 |
| 37412 | 2 |
| 30627 | 3 |
| 107417 | 4 |
| 37459 | 5 |
| 39027 | 6 |
| 30859 | 7 |
| 26181 | 8 |
| 31306 | 9 |
| 30829 | 10 |

|  |  |
| --- | --- |
| Top 10 least valuable players | |
| player\_id | rank |
| 38994 | 1 |
| 280653 | 2 |
| 139671 | 3 |
| 214570 | 4 |
| 316688 | 5 |
| 35327 | 6 |
| 427438 | 7 |
| 570760 | 8 |
| 172841 | 9 |
| 538112 | 10 |

Can I predict the match results for season 2015-2016 ?

I was able to end up with a prediction\_submission.csv file with predictions based of the team stats, goal stats, player stats, and their previous matches from the training set. This resulted in a result with a lot of losses somehow. This is probably not correct but at least I tried. I also have 7 total different approaches that I took in 7 different ‘ipynb’ files. I also have 3 different “csv” files with different values but I ended up with the last one. The “Win”/”Lose”/”Draw” represents what the home playing team did.

All the “ipynb” files are named: “predict\_matches\_v1.ipynb”, “predict\_matches\_v2.ipynb”, respectively.

All the “csv” files are named: “predicted\_results\_v1.csv”, “predicted\_results\_v2.csv”, respectively.

To get to my final answer I used: RandomForestClassifier & RandomizedSearchCV with a lot of different hyperparameters. Which in the end got me to my great result.

The rest of the steps I used are all mentioned and visible in my ipynb notebooks