# Dataset Loading

The dataset provided is a feather file. Saving files in feather extension is convenient for several reasons:

* Very lower time in I/O (read and write) operation, if compared to csv for example
* Pandas dataframe types are preserved in the I/O operations, differently from csv where the types are guessed when the file is loaded into python
* Much lower storage

Pandas offers a super easy way to read and write feather files. Note: you must pip install pyarrow before using feather files!

# Dataset content

Inside the dataset you will find match data for each player in Premier League (EN), Serie A (IT) and La Liga (ES) for the current season (season = 2023) and the previous one (season = 2022).

Competitions id:

* 8: Premier League
* 21: Serie A
* 23: La Liga

For each player you will find a lot of stats! Not all of them are relevant and you should not use all of them! Help yourself with the provided documentation with the explanations of metrics to better understand the dataset content. (To visualize all the stats in a notebook you can set the max visible cols to None in this way:

pd**.**set\_option**(**'display.max\_columns'**,** None**)**

****

Each player has a cluster\_label that represents his player function based on Soccerment proprietary clustering model.

If you need team aggregated data, you can derive them from the single players applying appropriate grouping.

You can derive the seasonal data from the single match ones for both players and teams.

Note: you may find a lot of Null or Nan values, which are stats without a value assigned. You can assume them as 0.