# **Data Preparation**

In task 1, I went through the dataset column by column looking for errors. First, I did a quick manual inspection of the dataset, mainly keeping an eye out on missing values or NaN or if there are incorrect data types ie an int rather than a string. Then, I proceeded to use python to check for data type inconsistencies, missing values, impossible values and other errors as outlined in detail below.

## Error 1 – Incorrect data type (Player, Pos, Tm)

The initial errors located is that pandas dataframe sees ‘Player’, ‘Pos’ and ‘Tm’ columns as an object data type. I have corrected this by using ‘pandas.DataFrame.astype’ to manually set those columns as string data. This allows for usage of string methods in python later.

## Error 2 – Redundant white space (Pos)

Some data entries in the Position column contained redundant white space, extra letters / symbols on the values or values were not capitalised. To fix this, I have used ‘.str.lstrip()’ and ‘.str.rstrip()’ to remove redundant white space, removed extra letters and symbols from the data entries then capitalised any values that were in lower case.

## Error 3 – Impossible values (Age)

# **Data Exploration**

## Task 2.1

“My answer to Task 2.1”

## Task 2.2

“My answer to Task 2.2”

## Task 2.3

“My answer to Task 2.3”