**Data Preparation**

“My explanation of how I addressed the task 1, including the steps of dealing with the potential issues/errors, and I will create one sub-section for each type of errors”

For task one, I took a brief look at the initial 50 records to identify any obvious errors before

**Error 1: White space**

The first error I look to fix is any extra whitespace in each value. This error is easy to spot using value\_counts() as these values will be displayed separately from the properly formatted value. Below is an example of the value\_counts() of the column ‘Pos’, ‘ PG’ is displayed separately from ‘PG’.



This error can be fixed by using strip() which will replace whitespace before and after the string. However, it is important to examine the valid data and ensure that there are no values where leading and trailing whitespace is apart of the valid string. In the case of ‘Pos’, there was no valid whitespace values and this held true for all columns.

**Error 2: Letter capitalisation**

The next most obvious and simple error was mistakes in the capitalisation of the string.

**Error 3: Duplicate players**

**Error 4: Individual errors**

There were multiple errors that only appeared in very low amounts, mostly a single record, and were not related to any of the previous errors and therefore needed specific attention to be rectified. These errors were fixed by targeting the value with a .replace() and changing it to the fixed value.

One example of this is the value ‘SGA’ and ‘PFA’ in the column ‘Pos’. These values were not apart of the values that the specification specified that were apart of the column, and therefore needed to be fixed. These were remedied by removing the trailing ‘A’, as no position value contains an A and by removing the letter, they now had values that were valid. A similar error is the value ‘SF.’ in the same column, which was valid once the trailing period was removed.

Another error is the value ‘SFPF’, which was missing a middle dash (-) to become a valid value which

....

**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”