Data scraping:

url: https://www.basketball-reference.com/draft/BAA\_1947.html

Scraped all the data from 1947 to 2019 ( 73 years in total ).

attachment of original data set: basketball\_dataset.csv)

Parsehub project file see attachment: ludi\_assignment0916.phj

Data cleaning:

1. Null value data processing: replace the value "null" with "0"

2. Rename the columns

3. Delete the Pk column in the original data

4. Transfer Rk data type to numeric

5. Delete all data except 1966 to 2014

For the openrefine project file, please see the attachment: basketball\_dataset-csv.openrefine.tar

attachment of dataset after cleaning: basketball\_dataset-csv(cleaned).xls