# Project -1

# Scraping and analysis of statistics of basketball players from real-time web page

## **OVERVIEW**: A project that gives you a better understanding of scraping data from websites and how to analyse them.Usage of various libraries as Numpy,MatPlot,Pandas.

In the course of completing the project,you use the web scraping function,converting the extracted data into a pandas data frame,and storing the analysed data.

## **Problem statement**

Web scrape basketball statistics from wikipedia of 10 basketball players and export it as a csv file format.

## **Software requirements**

1. Programming language : Python
2. Environment: Jupyter Notebooks\ Google Collab
3. Database : CSV(export type)
4. Operation System: Windows XP or above
5. Libraries Used : Beautiful soup,Requests, Pandas,Numpy,boto3, Matplotlib,Display

## **Open a New Notebook and import the required libraries**

import bs4

import requests

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from IPython.display import display

**Description**: required libraries are imported in this block of code

1. **Reading the webpage**

def get\_basketball\_stats(link='https://en.wikipedia.org/wiki/Michael\_Jordan'):

response = requests.get(link)

soup = bs4.BeautifulSoup(response.text, 'html.parser')

**Description** : The webpage which has to be read is given here

1. **Main Function Process**

table = soup.find(class\_='wikitable sortable')

headers = table.tr

titles = headers.find\_all('abbr')

data = {title['title']: [] for title in titles}

for row in table.find\_all('tr')[1:]:

for key, a in zip(data.keys(),row.find\_all('td')[2:]):

data[key].append(''.join(c for c in a.text if (c.isdigit() or c == '.')))

Min = min([len(x) for x in data.values()])

for key in data.keys():

data[key] = list(map(lambda x: float(x), data[key][:Min]))

return data

**Description :** In this we create and pass table header as the keys

We iterate over each table row by fining each table tag tr and assign it .

as each cell corresponds to a different column we all obtain the corresponding key corresponding the column n

## **4**.**Declaring links and names of the basketball players to scrap the data**

links = ['https://en.wikipedia.org/wiki/Michael\_Jordan'\

,'https://en.wikipedia.org/wiki/Kobe\_Bryant'\

,'https://en.wikipedia.org/wiki/LeBron\_James'\

,'https://en.wikipedia.org/wiki/Kevin\_Durant'\

,'https://en.wikipedia.org/wiki/Magic\_Johnson'\

,'https://en.wikipedia.org/wiki/James\_Harden'\

,'https://en.wikipedia.org/wiki/Dwyane\_Wade'\

,'https://en.wikipedia.org/wiki/Anthony\_Davis'\

,'https://en.wikipedia.org/wiki/Paul\_George'\

,'https://en.wikipedia.org/wiki/Stephen\_Curry']

names = ['Michael Jordan','Kobe Bryant','Lebron James','kevin Durant','Magic Johnson','James Harden','Dwyane Wade','Anthony Davis','Paul George','Stephen Curry']

Michael\_jordan\_dict = get\_basketball\_stats(links[0])

kobe\_bryant\_dict = get\_basketball\_stats(links[1])

lebron\_james\_dict = get\_basketball\_stats(links[2])

Kevin\_Durant\_dict = get\_basketball\_stats(links[3])

Magic\_Johnson\_dict = get\_basketball\_stats(links[4])

James\_Harden\_dict = get\_basketball\_stats(links[5])

Dwyane\_Wade\_dict = get\_basketball\_stats(links[6])

Anthony\_Davis\_dict = get\_basketball\_stats(links[7])

Paul\_George\_dict = get\_basketball\_stats(links[8])

stephen\_curry\_dict = get\_basketball\_stats(links[9])

mj\_table = pd.DataFrame(michael\_jordan\_dict)

kb\_table = pd.DataFrame(kobe\_bryant\_dict)

lj\_table = pd.DataFrame(lebron\_james\_dict)

kd\_table = pd.DataFrame(Kevin\_Durant\_dict)

maj\_table = pd.DataFrame(Magic\_Johnson\_dict)

jh\_table = pd.DataFrame(James\_Harden\_dict)

dw\_table = pd.DataFrame(Dwyane\_Wade\_dict)

ad\_table = pd.DataFrame(Anthony\_Davis\_dict)

pg\_table = pd.DataFrame(Paul\_George\_dict)

sc\_table = pd.DataFrame(stephen\_curry\_dict)

list\_table =[mj\_table, kb\_table, lj\_table,kd\_table,maj\_table,jh\_table,dw\_table,ad\_table,pg\_table, sc\_table]

i = 0

for name in names:

print(name)

display(list\_table[i].head())

i += 1

**OUTPUT**

Michael Jordan

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 82.0 | 82.0 | 38.3 | 0.515 | 0.173 | 0.845 | 6.5 | 5.9 | 2.4 | 0.8 | 28.2 |
| **1** | 18.0 | 7.0 | 25.1 | 0.457 | 0.167 | 0.840 | 3.6 | 2.9 | 2.1 | 1.2 | 22.7 |
| **2** | 82.0 | 82.0 | 40.0 | 0.482 | 0.182 | 0.857 | 5.2 | 4.6 | 2.9 | 1.5 | 37.1 |
| **3** | 82.0 | 82.0 | 40.4 | 0.535 | 0.132 | 0.841 | 5.5 | 5.9 | 3.2 | 1.6 | 35.0 |
| **4** | 81.0 | 81.0 | 40.2 | 0.538 | 0.276 | 0.850 | 8.0 | 8.0 | 2.9 | 0.8 | 32.5 |

Kobe Bryant

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 71.0 | 6.0 | 15.5 | 0.417 | 0.375 | 0.819 | 1.9 | 1.3 | 0.7 | 0.3 | 7.6 |
| **1** | 79.0 | 1.0 | 26.0 | 0.428 | 0.341 | 0.794 | 3.1 | 2.5 | 0.9 | 0.5 | 15.4 |
| **2** | 50.0 | 50.0 | 37.9 | 0.465 | 0.267 | 0.839 | 5.3 | 3.8 | 1.4 | 1.0 | 19.9 |
| **3** | 66.0 | 62.0 | 38.2 | 0.468 | 0.319 | 0.821 | 6.3 | 4.9 | 1.6 | 0.9 | 22.5 |
| **4** | 68.0 | 68.0 | 40.9 | 0.464 | 0.305 | 0.853 | 5.9 | 5.0 | 1.7 | 0.6 | 28.5 |

Lebron James

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 79.0 | 79.0 | 39.5 | 0.417 | 0.290 | 0.754 | 5.5 | 5.9 | 1.6 | 0.7 | 20.9 |
| **1** | 80.0 | 80.0 | 42.4 | 0.472 | 0.351 | 0.750 | 7.4 | 7.2 | 2.2 | 0.7 | 27.2 |
| **2** | 79.0 | 79.0 | 42.5 | 0.480 | 0.335 | 0.738 | 7.0 | 6.6 | 1.6 | 0.8 | 31.4 |
| **3** | 78.0 | 78.0 | 40.9 | 0.476 | 0.319 | 0.698 | 6.7 | 6.0 | 1.6 | 0.7 | 27.3 |
| **4** | 75.0 | 74.0 | 40.4 | 0.484 | 0.315 | 0.712 | 7.9 | 7.2 | 1.8 | 1.1 | 30.0 |

kevin Durant

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 80.0 | 80.0 | 34.6 | 0.430 | 0.288 | 0.873 | 4.4 | 2.4 | 1.0 | 0.9 | 20.3 |
| **1** | 74.0 | 74.0 | 39.0 | 0.476 | 0.422 | 0.863 | 6.5 | 2.8 | 1.3 | 0.7 | 25.3 |
| **2** | 82.0 | 82.0 | 39.5 | 0.476 | 0.365 | 0.900 | 7.6 | 2.8 | 1.4 | 1.0 | 30.1 |
| **3** | 78.0 | 78.0 | 38.9 | 0.462 | 0.350 | 0.880 | 6.8 | 2.7 | 1.1 | 1.0 | 27.7 |
| **4** | 66.0 | 66.0 | 38.6 | 0.496 | 0.387 | 0.860 | 8.0 | 3.5 | 1.3 | 1.2 | 28.0 |

Magic Johnson

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 77.0 | 72.0 | 36.3 | 0.530 | 0.226 | 0.81 | 7.7 | 7.3 | 2.4 | 0.5 | 18.0 |
| **1** | 37.0 | 35.0 | 37.1 | 0.532 | 0.176 | 0.76 | 8.6 | 8.6 | 3.4 | 0.7 | 21.6 |
| **2** | 78.0 | 77.0 | 38.3 | 0.537 | 0.207 | 0.76 | 9.6 | 9.5 | 2.7 | 0.4 | 18.6 |
| **3** | 79.0 | 79.0 | 36.8 | 0.548 | 0.000 | 0.80 | 8.6 | 10.5 | 2.2 | 0.6 | 16.8 |
| **4** | 67.0 | 66.0 | 38.3 | 0.565 | 0.207 | 0.81 | 7.3 | 13.1 | 2.2 | 0.7 | 17.6 |

James Harden

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 76.0 | 0.0 | 22.9 | 0.403 | 0.375 | 0.808 | 3.2 | 1.8 | 1.1 | 0.3 | 9.9 |
| **1** | 82.0 | 5.0 | 26.7 | 0.436 | 0.349 | 0.843 | 3.1 | 2.1 | 1.1 | 0.3 | 12.2 |
| **2** | 62.0 | 2.0 | 31.4 | 0.491 | 0.390 | 0.846 | 4.1 | 3.7 | 1.0 | 0.2 | 16.8 |
| **3** | 78.0 | 78.0 | 38.3 | 0.438 | 0.368 | 0.851 | 4.9 | 5.8 | 1.8 | 0.5 | 25.9 |
| **4** | 73.0 | 73.0 | 38.0 | 0.456 | 0.366 | 0.866 | 4.7 | 6.1 | 1.6 | 0.4 | 25.4 |

Dwyane Wade

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 61.0 | 56.0 | 34.9 | 0.465 | 0.302 | 0.747 | 4.0 | 4.5 | 1.4 | 0.6 | 16.2 |
| **1** | 77.0 | 77.0 | 38.6 | 0.478 | 0.289 | 0.762 | 5.2 | 6.8 | 1.6 | 1.1 | 24.1 |
| **2** | 75.0 | 75.0 | 38.6 | 0.495 | 0.171 | 0.783 | 5.7 | 6.7 | 1.9 | 0.8 | 27.2 |
| **3** | 51.0 | 50.0 | 37.9 | 0.491 | 0.266 | 0.807 | 4.7 | 7.5 | 2.1 | 1.2 | 27.4 |
| **4** | 51.0 | 49.0 | 38.3 | 0.469 | 0.286 | 0.758 | 4.2 | 6.9 | 1.7 | 0.7 | 24.6 |

Anthony Davis

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 64.0 | 60.0 | 28.8 | 0.516 | 0.000 | 0.751 | 8.2 | 1.0 | 1.2 | 1.8 | 13.5 |
| **1** | 67.0 | 66.0 | 35.2 | 0.519 | 0.222 | 0.791 | 10.0 | 1.6 | 1.3 | 2.8 | 20.8 |
| **2** | 68.0 | 68.0 | 36.1 | 0.535 | 0.083 | 0.805 | 10.2 | 2.2 | 1.5 | 2.9 | 24.4 |
| **3** | 61.0 | 61.0 | 35.5 | 0.493 | 0.324 | 0.758 | 10.3 | 1.9 | 1.3 | 2.0 | 24.3 |
| **4** | 75.0 | 75.0 | 36.1 | 0.505 | 0.299 | 0.802 | 11.8 | 2.1 | 1.3 | 2.2 | 28.0 |

Paul George

|  | **Games played** | **Games started** | **Minutes per game** | **Field goal percentage** | **3-point field-goal percentage** | **Free-throw percentage** | **Rebounds per game** | **Assists per game** | **Steals per game** | **Blocks per game** | **Points per game** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 61.0 | 19.0 | 20.7 | 0.453 | 0.297 | 0.762 | 3.7 | 1.1 | 1.0 | 0.4 | 7.8 |
| **1** | 66.0 | 66.0 | 29.7 | 0.440 | 0.385 | 0.802 | 5.6 | 2.4 | 1.6 | 0.6 | 12.1 |
| **2** | 79.0 | 79.0 | 37.6 | 0.419 | 0.362 | 0.807 | 7.6 | 4.1 | 1.8 | 0.6 | 17.4 |
| **3** | 80.0 | 80.0 | 36.2 | 0.424 | 0.364 | 0.864 | 6.8 | 3.5 | 1.9 | 0.3 | 21.7 |
| **4** | 6.0 | 0.0 | 15.2 | 0.367 | 0.409 | 0.727 | 3.7 | 1.0 | 0.8 | 0.2 | 8.8 |

**Description**: The data is collected and put in the form of tables

**5. Making a plot using Matplotlib**

j = 0

for name in names:

plt.plot(list\_table[j][['Points per game']],label=name)

plt.legend()

plt.xlabel('years')

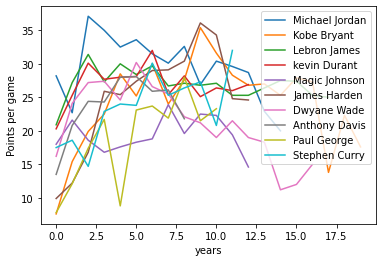
plt.ylabel('Points per game')

j += 1

csv\_name = 'MJ1.csv'

mj\_table.to\_csv(csv\_name)

**OUTPUT**

****

**Description:** All the values of points and years played is plot

**Conclusion :** In This project we have learned to import libraries which are used in particular code then data from a particular website is taken and the data is cleaned ,processed and a graph is plotted comparing this values