Most Runs Records in Test Cricket

In [1]:

```python
import requests
from bs4 import BeautifulSoup
import pandas as pd
```

- requests library will give us the content from the given Url.
- By Using Beautiful Soup library we can extract and parse data from HTML files.
- Use requests and Beautiful Soup for scraping and parsing data from the Web.
- Pandas makes it easy to scrape a table (tag) on a web page.
- After obtaining it as a DataFrame, we can save it as an Excel file or csv file.

In [2]:

```python
url = "https://www.espncricinfo.com/records/most-runs-in-career-223646"
```

- we will use the above url for scrapping the data with the help of requests library.
- If we get '200' response then we can scrap the page

In [3]:

```python
response = requests.get(url)
print(response)
```

```
<Response [200]>
```

In [4]:

```python
soup = BeautifulSoup(response.text)
soup.prettify
```

```
[346BF]|3[AC-F]|40|5[BDE]|60|7[BCD])/g,decodeURIComponent)}},{path:"/"});return t}));
        </script><script type="text/javascript">
          var appInfoValue = Cookies.get('ci-app');
          var appInfo = null;
          var isApp = false;
          if (appInfoValue) {
            appInfo = JSON.parse(appInfoValue);

            // android || ios
            if (appInfo['pf'] === 'android' || appInfo['pf'] === 'ios') {
              isApp = true;
            }
          }
          window.hsci = window.hsci || {
            consent: {
              region: 'unknown',
              enabled: false,
              taken: true,
            },
            did: null,
```

- BeautifulSoup will parse the above document.
- The prettify() method will turn a Beautiful Soup parse tree into a nicely formatted Unicode string, with a separate line for each tag and each string.

In [5]:

```python
Cric = soup.find('table')
Cric
```

Out[5]:

```
<table class="ds-w-full ds-table ds-table-xs ds-table-auto ds-w-full ds-overflow-scroll ds-scrollbar-hide"><thead class="ds-b
g-fill-content-alternate ds-text-left"><tr class=""><td class="ds-min-w-max"><div class="ds-popper-wrapper"><span class="ds-c
ursor-pointer">Player</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-
cursor-pointer">Span</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-c
ursor-pointer">Mat</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-cur
sor-pointer">Inns</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-curs
or-pointer">NO</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-cursor-
pointer"><strong>Runs</strong></span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span c
lass="ds-cursor-pointer">HS</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span clas
s="ds-cursor-pointer">Ave</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class
="ds-cursor-pointer">BF</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="d
s-cursor-pointer">SR</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-c
ursor-pointer">100</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-cur
sor-pointer">50</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-cursor
-pointer">0</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-cursor-poi
nter">4s</span></div></td><td class="ds-min-w-max ds-text-right"><div class="ds-popper-wrapper"><span class="ds-cursor-pointe
r">6s</span></div></td></tr></thead><tbody class=""><tr class=""><td class="ds-min-w-max"><div class="ds-popper-wrapper"><spa
n class="ds-cursor-pointer"><a class="ds-inline-flex ds-items-start ds-leading-none" href="/cricketers/sachin-tendulkar-3532
0" title="SR Tendulkar (IND)"><span class="ds-text-tight-s ds-font-regular ds-text-typo-primary hover:ds-text-typo-primary-ho
```

- We will find the table tag using soup.find required table and store it in 'table' variable.

```
In [6]: Cric = Cric.find('tbody')
        Cric
```

```
Out[6]: <tbody class=""><tr class=""><td class="ds-min-w-max"><div class="ds-popper-wrapper"><span class="ds-cursor-pointer"><a class
        ="ds-inline-flex ds-items-start ds-leading-none" href="/cricketers/sachin-tendulkar-35320" title="SR Tendulkar (IND)"><span c
        lass="ds-text-tight-s ds-font-regular ds-text-typo-primary hover:ds-text-typo-primary-hover ds-block">SR Tendulkar (IND)</spa
        n></a></span></div></td><td class="ds-min-w-max ds-text-right"><span class="">1989-2013</span></td><td class="ds-min-w-max ds
        -text-right"><span class="">200</span></td><td class="ds-min-w-max ds-text-right"><span class="">329</span></td><td class="ds
        -min-w-max ds-text-right"><span class="">33</span></td><td class="ds-min-w-max ds-text-right"><span class=""><strong>15921</s
        trong></span></td><td class="ds-min-w-max ds-text-right"><span class="ds-relative ds--right-1.5">248*</span></td><td class="d
        s-min-w-max ds-text-right"><span class="">53.78</span></td><td class="ds-min-w-max ds-text-right"><span class="ds-relative ds
        --right-1.5">29437+</span></td><td class="ds-min-w-max ds-text-right"><span class="">54.04</span></td><td class="ds-min-w-max
        ds-text-right"><span class="">51</span></td><td class="ds-min-w-max ds-text-right"><span class="">68</span></td><td class="ds
        -min-w-max ds-text-right"><span class="">14</span></td><td class="ds-min-w-max ds-text-right"><span class="ds-relative ds--ri
        ght-1.5">2058+</span></td><td class="ds-min-w-max ds-text-right"><span class="">69</span></td></tr><tr class=""><td class="ds
        -min-w-max"><div class="ds-popper-wrapper"><span class="ds-cursor-pointer"><a class="ds-inline-flex ds-items-start ds-leading
        -none" href="/cricketers/ricky-ponting-7133" title="RT Ponting (AUS)"><span class="ds-text-tight-s ds-font-regular ds-text-ty
        po-primary hover:ds-text-typo-primary-hover ds-block">RT Ponting (AUS)</span></a></span></div></td><td class="ds-min-w-max ds
        -text-right"><span class="">1995-2012</span></td><td class="ds-min-w-max ds-text-right"><span class="">168</span></td><td cla
        ss="ds-min-w-max ds-text-right"><span class="">287</span></td><td class="ds-min-w-max ds-text-right"><span class="">29</span>
        </td><td class="ds-min-w-max ds-text-right"><span class=""><strong>13378</strong></span></td><td class="ds-min-w-max ds-text-
        right"><span class="">257</span></td><td class="ds-min-w-max ds-text-right"><span class="">51.85</span></td><td class="ds-min
```

- Now we will find 'tbody' tag because required content is present in 'tbody' tag

```
In [7]: rec = Cric.find_all("tr")
        rec
```

```
Out[7]: [<tr class=""><td class="ds-min-w-max"><div class="ds-popper-wrapper"><span class="ds-cursor-pointer"><a class="ds-inline-fle
        x ds-items-start ds-leading-none" href="/cricketers/sachin-tendulkar-35320" title="SR Tendulkar (IND)"><span class="ds-text-t
        ight-s ds-font-regular ds-text-typo-primary hover:ds-text-typo-primary-hover ds-block">SR Tendulkar (IND)</span></a></span></
        div></td><td class="ds-min-w-max ds-text-right"><span class="">1989-2013</span></td><td class="ds-min-w-max ds-text-right"><s
        pan class="">200</span></td><td class="ds-min-w-max ds-text-right"><span class="">329</span></td><td class="ds-min-w-max ds-t
        ext-right"><span class="">33</span></td><td class="ds-min-w-max ds-text-right"><span class=""><strong>15921</strong></span></
        td><td class="ds-min-w-max ds-text-right"><span class="ds-relative ds--right-1.5">248*</span></td><td class="ds-min-w-max ds-
        text-right"><span class="">53.78</span></td><td class="ds-min-w-max ds-text-right"><span class="ds-relative ds--right-1.5">29
        437+</span></td><td class="ds-min-w-max ds-text-right"><span class="">54.04</span></td><td class="ds-min-w-max ds-text-righ
        t"><span class="">51</span></td><td class="ds-min-w-max ds-text-right"><span class="">68</span></td><td class="ds-min-w-max d
        s-text-right"><span class="">14</span></td><td class="ds-min-w-max ds-text-right"><span class="ds-relative ds--right-1.5">205
        8+</span></td><td class="ds-min-w-max ds-text-right"><span class="">69</span></td></tr>,
         <tr class=""><td class="ds-min-w-max"><div class="ds-popper-wrapper"><span class="ds-cursor-pointer"><a class="ds-inline-fle
        x ds-items-start ds-leading-none" href="/cricketers/ricky-ponting-7133" title="RT Ponting (AUS)"><span class="ds-text-tight-s
        ds-font-regular ds-text-typo-primary hover:ds-text-typo-primary-hover ds-block">RT Ponting (AUS)</span></a></span></div></td>
        <td class="ds-min-w-max ds-text-right"><span class="">1995-2012</span></td><td class="ds-min-w-max ds-text-right"><span class
        ="">168</span></td><td class="ds-min-w-max ds-text-right"><span class="">287</span></td><td class="ds-min-w-max ds-text-righ
        t"><span class="">29</span></td><td class="ds-min-w-max ds-text-right"><span class=""><strong>13378</strong></span></td><td c
        lass="ds-min-w-max ds-text-right"><span class="">257</span></td><td class="ds-min-w-max ds-text-right"><span class="">51.85</
```

- We will find the all the rows which are present in tr tag

```
In [8]: columns = []

        for i in rec:
            col=i.find_all("td")
            print(col)
            runs = [c.text for c in col]
            print(runs)
            columns.append(runs)
        print(columns)
```

- Here,We will take an empty list 'columns' and then we will take for loop in tablerows which contains all the tr tags.
- Then we will find all td tags for every tr tags and then we will append the created empty list 'columns'

[9]:

```
player = []
span = []
mat = []
inns = []
no = []
runs = []
hs = []
ave = []
bf = []
sr = []
hun = []
fif = []
zer = []
fours = []
six = []

for i in columns:
    player.append(i[0])
    span.append(i[1])
    mat.append(i[2])
    inns.append(i[3])
    no.append(i[4])
    runs.append(i[5])
    hs.append(i[6])
    ave.append(i[7])
    bf.append(i[8])
    sr.append(i[9])
    hun.append(i[10])
    fif.append(i[11])
    zer.append(i[12])
    fours.append(i[13])
    six.append(i[14])
```

- Here we will create empty lists for every column.
- Now required content is stored in 'columns' list so we will iterate every item of list using for loop.
- Then we will append each item of list using indexing.

In [10]:

```
Record_Runs = {"Player Name" : player, "Playing Span" : span, "Matches Played" : mat, "Innings Battled" : inns, "Not Outs" : no,
               "Runs Scored" : runs, "Highest Innings Scores" : hs, "Batting Average" : ave, "Balls Faced" : bf, "Strike Rate" :
              "100s Scored" : hun, "50s Scored" : fif, "Ducks" : zer, "4s" : fours, "6s" : six}
```

- A dictionary named Record_Runs is created, where the column names are used as keys, and the corresponding lists of values are used as values.

In [11]:

```
Most_Runs = pd.DataFrame(Record_Runs)
Most_Runs
```

Out[11]:

	Player Name	Playing Span	Matches Played	Innings Battled	Not Outs	Runs Scored	Highest Innings Scores	Batting Average	Balls Faced	Strike Rate	100s Scored	50s Scored	Ducks	4s	6s
0	SR Tendulkar (IND)	1989-2013	200	329	33	15921	248*	53.78	29437+	54.04	51	68	14	2058+	69
1	RT Ponting (AUS)	1995-2012	168	287	29	13378	257	51.85	22782	58.72	41	62	17	1509	73
2	JH Kallis (ICC/SA)	1995-2013	166	280	40	13289	224	55.37	28903	45.97	45	58	16	1488	97
3	R Dravid (ICC/IND)	1996-2012	164	286	32	13288	270	52.31	31258	42.51	36	63	8	1654	21
4	AN Cook (ENG)	2006-2018	161	291	16	12472	294	45.35	26562	46.95	33	57	9	1442	11
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
102	Tamim Iqbal (BAN)	2008-2023	70	134	2	5134	206	38.89	8852	57.99	10	31	11	655	41
103	A Ranatunga (SL)	1982-2000	93	155	12	5105	135*	35.69	8672+	49.96	4	38	12	534+	40
104	AM Rahane (IND)	2013-2023	85	144	12	5077	188	38.46	10256	49.50	12	26	10	578	35
105	Zaheer Abbas (PAK)	1969-1985	78	124	11	5062	274	44.79	6935+	55.22	12	20	10	458+	22
106	UT Khawaja (AUS)	2011-2023	66	117	11	5004	195*	47.20	10213	48.99	15	24	6	543	25

107 rows × 15 columns

- The 'Most_Runs' dictionary is passed to the pd.DataFrame() function to create a Pandas DataFrame named 'Most_Runs'.

```
In [12]: Most_Runs.to_csv("Most_Runs_Records", index = False)
```

- Here, We will convert 'Most_Runs' dataframe into csv file
- The index parameter is set to False to exclude the index column from the saved CSV file.