

In [1]:

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

headers = {
    "User-Agent": "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/75.0.3770.100 Safari/537.36"
}

urls = [
    "https://www.icc-cricket.com/rankings/mens/player-rankings/test/batting",
    "https://www.icc-cricket.com/rankings/mens/player-rankings/test/bowling",
    "https://www.icc-cricket.com/rankings/mens/player-rankings/odi/batting",
    "https://www.icc-cricket.com/rankings/mens/player-rankings/odi/bowling",
    "https://www.icc-cricket.com/rankings/mens/player-rankings/t20i/batting",
    "https://www.icc-cricket.com/rankings/mens/player-rankings/t20i/bowling",
    "https://www.icc-cricket.com/rankings/womens/player-rankings/odi/batting",
    "https://www.icc-cricket.com/rankings/womens/player-rankings/t20i/batting",
    "https://www.icc-cricket.com/rankings/womens/player-rankings/odi/bowling",
    "https://www.icc-cricket.com/rankings/womens/player-rankings/t20i/bowling",
]

final_result_file_name = "All Ranking List.csv"
final_column_names = ["Ranking Type", "Position", "Player Name", "Team Name", "Rating", "Career Best Rating", "Crawl URL"]
pd.DataFrame(columns=final_column_names).to_csv(final_result_file_name, sep="\t", index=False, encoding="utf-8")

for url in urls:
    request_object = requests.get(url, headers=headers)
    html_content = request_object.text
    print(request_object.status_code, "->", url)
    soup_object = BeautifulSoup(html_content, "lxml")
    for element in soup_object.select('[class="ranking-pos up"]', [class="ranking-pos down"]):
        element.replace_with(BeautifulSoup("<" + element.name + ">/" + element.name + ">", "html.parser"))

    ranking_type = soup_object.select_one(".rankings-block__title-container > h4").text

    result_file_name = ranking_type + ".csv"
    column_names = ["Position", "Player Name", "Team Name", "Rating", "Career Best Rating", "Crawl URL"]
    pd.DataFrame(columns=column_names).to_csv(result_file_name, sep="\t", index=False, encoding="utf-8")
```

```

for element in soup_object.select('table[class="table rankings-table"] tr'):
    if(element.find("th")):
        continue
    data_dict = dict()
    data_dict["Crawl URL"] = url
    data_dict["Ranking Type"] = ranking_type
    if(element.select_one('[class*="position"]')):
        data_dict["Position"] = element.select_one('[class*="position"]').text
    for player_name in (element.select('a[href*="/player-rankings"]')):
        if(player_name.text.strip()):
            data_dict["Player Name"] = player_name.text
    if(element.select_one('[class^="flag-15"]')):
        data_dict["Team Name"] = element.select_one('[class^="flag-15"]')['class'][-1]
    if(element.select_one('[class$="rating"]')):
        data_dict["Rating"] = element.select_one('[class$="rating"]').text
    if(element.select_one('td.u-hide-phablet')):
        data_dict["Career Best Rating"] = element.select_one('td.u-hide-phablet').text
    for key in data_dict.keys():
        data_dict[key] = re.sub(r"\s+", " ", data_dict[key])
        data_dict[key] = data_dict[key].strip()
    pd.DataFrame([data_dict], columns=column_names).to_csv(result_file_name, sep="\t", index=False, header=False, mode='a')
    pd.DataFrame([data_dict], columns=final_column_names).to_csv(final_result_file_name, sep="\t", index=False, mode='a')

```

```

200 -> https://www.icc-cricket.com/rankings/mens/player-rankings/test/batting
200 -> https://www.icc-cricket.com/rankings/mens/player-rankings/test/bowling
200 -> https://www.icc-cricket.com/rankings/mens/player-rankings/odi/batting
200 -> https://www.icc-cricket.com/rankings/mens/player-rankings/odi/bowling
200 -> https://www.icc-cricket.com/rankings/mens/player-rankings/t20i/batting
200 -> https://www.icc-cricket.com/rankings/mens/player-rankings/t20i/bowling
200 -> https://www.icc-cricket.com/rankings/womens/player-rankings/odi/batting
200 -> https://www.icc-cricket.com/rankings/womens/player-rankings/t20i/batting
200 -> https://www.icc-cricket.com/rankings/womens/player-rankings/odi/bowling
200 -> https://www.icc-cricket.com/rankings/womens/player-rankings/t20i/bowling

```

In []:

In []: