from bs4 import BeautifulSoup

import requests

import re

import pandas as pd

urls = [

    "<https://www.icc-cricket.com/rankings/mens/team-rankings/odi>",

    "<https://www.icc-cricket.com/rankings/mens/player-rankings/odi>",

    "<https://www.icc-cricket.com/rankings/mens/player-rankings/odi>",

    "<https://www.icc-cricket.com/rankings/womens/team-rankings/odi>",

    "<https://www.icc-cricket.com/rankings/womens/player-rankings/odi/batting>",

    "<https://www.icc-cricket.com/rankings/womens/player-rankings/odi/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, "1xml")

    for element in soup\_object.select('[class="ranking-pop up"], [class="ranking-pop down"]'):

        element.replace\_with(BeautifulSoup("<" + [element.name](http://element.name/) + "<>/" + [element.name](http://element.name/) +">", html.parser ))

    ranking\_type = soup\_object.select\_one(".rankings-block\_title\_container > h4.text").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()

        columns=column\_names).to\_csv(result\_file\_name, sep="/t", index=False, header= False, encoding="utf-8", mode"a")

pd.DataFrame([data\_dict], columns=final\_column\_names).to\_csv(final\_result\_file\_name, sep="/t", index=False, header= False, encoding="utf-8", mode"a")