In [2]:
```
#web scraping world population data
from bs4 import BeautifulSoup
import requests
```

In [3]:
```
url = "https://en.wikipedia.org/wiki/List_of_countries_by_population_(United_Nations)"
page = requests.get(url)

#print(page.status_code): This is to confirm if you can access the url's html
#print(page.text): This helps to show the whole url's HTML
```

In [4]:
```
soup = BeautifulSoup(page.text,'html.parser')
```

In [5]:
```
#to get the column headers
Population_column = soup.find_all('th')[0:4]
print(Population_column)
#table
```

```
[<th>Location
</th>, <th>Population<br/>(1 July 2022)
</th>, <th>Population<br/>(1 July 2023)
</th>, <th>Change
</th>]
```

In [6]:
```
Population_headers = [title.text.strip() for title in Population_column]
Population_headers
```

Out[6]:
```
['Location', 'Population(1 July 2022)', 'Population(1 July 2023)', 'Change']
```

In [7]:
```
import pandas as pd
df = pd.DataFrame(columns = Population_headers)

#Appointing new columns to a new value
Additional_Columns = ['UN Continental Region','UN Statistical Subregion']
#Adding new columns to the dataframe and confirming there is no value inside
df[Additional_Columns[0]] = None
df[Additional_Columns[1]] = None
df
```

Out[7]:

Location	Population(1 July 2022)	Population(1 July 2023)	Change	UN Continental Region	UN Statistical Subregion

In [8]:
```
rows_data = soup.find_all("tr")[2:] #examined the html code to pick where the data I need sta
                    #, class_ = 'wikitable sortable')
#print(table)[0:4]
#print(table)
```

In [10]:
```
for row in rows_data[:238]:
    all_row_data = row.find_all('td')
    the_row_data = [data.text.strip() for data in all_row_data]

    #print(the_row_data)
    length = len(df)
    df.loc[len(df)] = the_row_data
```

In [11]:
```
df
```

Out[11]:

	Location	Population(1 July 2022)	Population(1 July 2023)	Change	UN Continental Region	UN Statistical Subregion
0	India	1,417,173,173	1,428,627,663	+0.81%	Asia	Southern Asia
1	China[a]	1,425,887,337	1,425,671,352	−0.02%	Asia	Eastern Asia
2	United States	338,289,857	339,996,564	+0.50%	Americas	Northern America
3	Indonesia	275,501,339	277,534,123	+0.74%	Asia	South-eastern Asia
4	Pakistan	235,824,862	240,485,658	+1.98%	Asia	Southern Asia
...	...	...	...	...	...	...
471	Falkland Islands (United Kingdom)	3,780	3,791	+0.29%	Americas	South America
472	Niue	1,934	1,935	+0.05%	Oceania	Polynesia
473	Tokelau (New Zealand)	1,871	1,893	+1.18%	Oceania	Polynesia
474	Vatican City[x]	510	518	N/A	Europe	Southern Europe
475	Pitcairn Islands (United Kingdom)	55	45	-18.1818%	Oceania	Polynesia

476 rows × 6 columns

In [ ]:
```
df.to_excel (r'filepath\World_Population(2024).xlsx', index=False)
```