Python: Scrape a List of largest public companies in Canada by profit

Objective: put one dataset together using webscraping and provide summary statistics

In [11]:

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

web_url = 'https://en.wikipedia.org/wiki/List_of_largest_public_companies_in_Canada_by_profit'
response = requests.get(web_url)

print('Status code\n', response.status_code)
print('\n--\n')
print('Content of the website\n', response.content[:2000])

soup_object = BeautifulSoup(response.content)

data_table = soup_object.find_all('table', 'wikitable sortable')[1]

all_values = data_table.find_all('tr')
all_values[:5] # Prints the first 5 captured tag elements

#Create an empty dataframe
pubic_companies = pd.DataFrame(columns = ['Rank', 'Rank Change', 'Company','Profit','Headquarter'])

#Use for loop to iterate each element into the empty dataframe
for row in all_values[1:]:
    values = row.find_all('td') # Extract all elements with tag <td>
    # Pick only the text part from the <td> tag
    rank = values[0].text.replace('\n', '')
    rank_change = values[1].img["alt"]+ values[1].text.replace('\n', '')
    company = values[2].text.replace('\n', '')
    profit = values[3].text.replace('\n', '')
    headquarter = values[4].text.replace('\n', '')

    pubic_companies.loc[ix] = [rank, rank_change, company, profit, headquarter]
    ix += 1

#export it to csv format
pubic_companies.to_csv('largest public companies in Canada by profit', index=False)
```

```
Status code
 200

--

Content of the website
 b'<!DOCTYPE html>\n<html class="client-nojs" lang="en" dir="ltr">\n<head>\n<meta charset="UTF-8"/>\
n<title>List of largest public companies in Canada by profit - Wikipedia</title>\n<script>document.d
ocumentElement.className="client-js";RLCONF={"wgBreakFrames":!1,"wgSeparatorTransformTable":["",""],
"wgDigitTransformTable":["",""],"wgDefaultDateFormat":"dmy","wgMonthNames":["","January","February",
"March","April","May","June","July","August","September","October","November","December"],"wgRequest
Id":"ce491fe2-aa33-4688-9f7e-4b149a3dc025","wgCSPNonce":!1,"wgCanonicalNamespace":"","wgCanonicalSpe
cialPageName":!1,"wgNamespaceNumber":0,"wgPageName":"List_of_largest_public_companies_in_Canada_by_p
rofit","wgTitle":"List of largest public companies in Canada by profit","wgCurRevisionId":982509858,
"wgRevisionId":982509858,"wgArticleId":36374850,"wgIsArticle":!0,"wgIsRedirect":!1,"wgAction":"view"
,"wgUserName":null,"wgUserGroups":["*"],"wgCategories":["Articles with short description","Short des
cription matches Wikidata","Wikipedia articles in need of updating from February 2019","All Wikipedi
a articles in need of updating","Lists of companies of Canada",\n"Publicly traded companies of Canad
a"],"wgPageContentLanguage":"en","wgPageContentModel":"wikitext","wgRelevantPageName":"List_of_large
st_public_companies_in_Canada_by_profit","wgRelevantArticleId":36374850,"wgIsProbablyEditable":!0,"w
gRelevantPageIsProbablyEditable":!0,"wgRestrictionEdit":[],"wgRestrictionMove":[],"wgMediaViewerOnCl
ick":!0,"wgMediaViewerEnabledByDefault":!0,"wgPopupsReferencePreviews":!1,"wgPopupsConflictsWithNavP
opupGadget":!1,"wgVisualEditor":{"pageLanguageCode":"en","pageLanguageDir":"ltr","pageVariantFallbac
ks":"en"},"wgMFDisplayWikibaseDescriptions":{"search":!0,"nearby":!0,"watchlist":!0,"tagline":!1},"w
gWMESchemaEditAttemptStepOversample":!1,"wgULSCurrentAutonym":"English","wgNoticeProject":"wikipedia
","wgCentralAuthMobileDomain":!1,"wgEditSubmitButtonLabelPublish":!0,"wgULSPosition":"interlanguage"
,"wgWikib'
```

In [12]:

```
pubic_companies.head(10)
```

Out[12]:

	Rank	Rank Change	Company	Profit	Headquarter
500	1	Increase 2	Royal Bank of Canada	7,442,000	Toronto & Montreal
501	2	Steady	Bank of Nova Scotia	6,466,000	Toronto
502	3	Decrease 2	Toronto-Dominion Bank	6,367,000	Toronto
503	4	Increase 3	Bank of Montreal	4,115,000	Montreal & Toronto
504	5	Increase 1	Imperial Oil	3,766,000	Calgary
505	6	Increase 2	Canadian Imperial Bank of Commerce	3,339,000	Toronto
506	7	Decrease 2	Suncor Energy	2,783,000	Calgary
507	8	Increase 5	BCE Inc.	2,763,000	Montreal
508	9	Increase 3	Canadian National Railway	2,680,000	Montreal
509	10	Decrease 2	Potash Corporation of Saskatchewan	2,668,000[1]	Saskatoon

In [12]:

```
pubic_companies.head(10)
```

Out[12]:

	Rank	Rank Change	Company	Profit	Headquarter