requests_library_solution

March 19, 2023

In [2]:
```
# Import libraries
from bs4 import BeautifulSoup
import requests
import pandas as pd
import numpy as np

# Create a Response object
r = requests.get('https://en.wikipedia.org/wiki/Amazon_(company)')

# Get HTML data
html_data = r.text

# Create a BeautifulSoup Object
page_content = BeautifulSoup(html_data, 'html.parser')

# Find financial table
wikitable = page_content.find('table', {'class': 'wikitable float-left'})

# Find all column titles
wikicolumns = wikitable.tbody.findAll('tr')[0].findAll('th')

# Loop through column titles and store into Python array
df_columns = []
for column in wikicolumns:
    text = column.get_text(strip=True, separator=" ")
    df_columns.append(text)

# Loop through the data rows and store into Python array
df_data = []
for row in wikitable.tbody.findAll('tr')[1:]:
    row_data = []
    for td in row.findAll('td'):
        text = td.get_text(strip=True, separator=" ")
        row_data.append(text)
    df_data.append(np.array(row_data))

# Print financial data in DataFrame format and set `Year` as index
```

```
dataframe = pd.DataFrame(data=df_data, columns=df_columns)
dataframe.set_index('Year', inplace=True)
dataframe
```

Out[2]:

Year	Revenue in mil. USD$	Net income in mil. USD$	Total Assets in mil. USD$	Employees
2007 [135]	14,835	476	6,485	17,000
2008 [136]	19,166	645	8,314	20,700
2009 [137]	24,509	902	13,813	24,300
2010 [138]	34,204	1,152	18,797	33,700
2011 [139]	48,077	631	25,278	56,200
2012 [140]	61,093	39	32,555	88,400
2013 [141]	74,452	274	40,159	117,300
2014 [142]	88,988	241	54,505	154,100
2015 [143]	107,006	596	64,747	230,800
2016 [144]	135,987	2,371	83,402	341,400
2017 [145]	177,866	3,033	131,310	566,000
2018 [146]	232,887	10,073	162,648	647,500
2019 [147]	280,522	11,588	225,248	798,000

In [ ]: