Muhammad Satrio Pinoto Negoro Linkedin: https://www.linkedin.com/in/satriopino/ Webscrapping The World's Largest Companies by Forbes.com (2023) using BeautifulSoup Web scraping, also known as web harvesting or web data extraction, is the process of automatically extracting data from websites. It involves fetching a web page and extracting data from it. The data can be parsed, searched, reformatted, and copied into a spreadsheet or loaded into a database. Web scraping can be done manually, but in most cases, automated tools are preferred as they can be less costly and work at a faster rate. Web scraping is used for various purposes, including lead generation, price monitoring, market research, and content aggregation. However, some websites use methods to prevent web scraping, such as detecting and disallowing bots from crawling their pages. In response, there are web scraping systems that rely on using techniques in DOM parsing, computer vision, and natural language processing to simulate human browsing to enable gathering web page content for offline parsing. **Dependencies** Actually to follow this module you only need to install beautifulsoup4 with pip install beautifulsoup4 and you are good to go. But here some libraries that needed to be installed first that I use at bis module: beautifulSoup4 pandas Background At this project I try to scrap Company Name, Country, Sales, Profit, and Assets from The World's Largest Companies by Forbes.com (2023) data. Forbes, founded in 1917, is a renowned American business magazine and online platform. Known for its influential annual lists like the Forbes 400 and Forbes Global 2000, Forbes provides comprehensive coverage of business, technology, and lifestyle. Undergoing a digital transformation, it engages a global audience through insightful articles and expert contributors. The Forbes brand extends beyond media, including exclusive communities like Forbes Councils and impactful events. With a legacy of family ownership, Forbes remains a trusted source of authoritative business journalism, shaping conversations in the ever-evolving landscape of finance and entrepreneurship. A lot of you might ask why I need to scrap this data from the Forbes sites. The data cannot be downloaded and I want to make a report and gain some insight from that data and maybe can be useful for others. To do that I need to have the data, and scrapping is a good way to collect the data I don't have from the public. I will scrap 5 points from this sites. That is Company Name, Country, Sales, Profit, and Assets. What is BeautifulSoup Beautiful Soup is a Python library for pulling data out of HTML and XML files. Beautiful Soup 3 only works on Python 2.x, but Beautiful Soup 4 also works on Python 3.x. Beautiful Soup 4 is faster, has more features, and works with third-party parsers like lxml and html5lib. Since beautifulsoup used to pull the data out of a HTML, so first we need to pull out the html first. How we do it? We will use default library request. So all this code is doing is sending a GET request to spesific address we give. This is the same type of request your browser sent to view this page, but the only difference is that Requests can't actually render the HTML, so instead you will just get the raw HTML and the other response information. I'm using the .get() function here, but Requests allows you to use other functions like .post() and .put() to send those requests as well. At this case we will going to the Global 2000 ranking the world's largest companies by Forbes.com page, you can click here to follow what exactly that link goes to. import requests from bs4 import BeautifulSoup 'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/119.0.0.0 Safari/537.36' In [3]: | url = 'https://www.forbes.com/lists/global2000/?sh=710d4f0c5ac0' response = requests.get(url, headers=headers) response.content[:500] b'<!DOCTYPE html><html lang="en"><head><link rel="preload" as="font" href="https://i.forbesimg.com/assets/fonts/merriweather/merriweather-bold-webf Out[4]: ont.woff2" type="font/woff2" crossorigin><link rel="preload" as="font" href="https://i.forbesimg.com/assets/fonts/work-sans/worksans-regular-webfon t.woff2" type="font/woff2" crossorigin><link rel="preload" as="font" href="https://i.forbesimg.com/assets/fonts/merriweather/merriweather-regular-w ebfont.woff2" type="font/woff2" crossorigin><link rel="preloa' soup = BeautifulSoup(response.content, 'html.parser') print(type(soup)) <class 'bs4.BeautifulSoup'> print(soup.prettify()[:500]) <!DOCTYPE html> <html lang="en"> <head> <link as="font" crossorigin="" href="https://i.forbesimg.com/assets/fonts/merriweather/merriweather-bold-webfont.woff2" rel="preload" type="font/</pre> <link as="font" crossorigin="" href="https://i.forbesimg.com/assets/fonts/work-sans/worksans-regular-webfont.woff2" rel="preload" type="font/woff</pre> <link as="font" crossorigin="" href="https://i.forbesimg.com/assets/fonts/merriweather/merriweather-regular-webfont.woff2" rel="preload" type="fo</pre> nt/w soup.find("div", attrs={'class':'table-row-group'}).prettify() '<div class="table-row-group">\n <a aria-label="JPMorgan Chase" canexpand="" class="table-row active premiumProfile-jpmorgan-chase" href="https://w ww.forbes.com/companies/jpmorgan-chase/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="jpmorgan-chase"> \n <div class="rank first table-cell rank">\n 1.\n </div>\n <div class="organizationName second table-cell name">\n JPMorgan Chase\n </div> \n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$179.93 B\n </div>\n < div class="profits table-cell profit (\$)">\n \$41.8 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$3,744.3 B\n </div >\n <\div class="assets table-cell assets (\$)">\n \$3,744.3 B\n <\div class=\div ss="marketValue table-cell market value (\$)">\n \$399.59 B\n </div>\n <a>\n <a aria-label="Saudi Arabian Oil Company (Saudi Aramco)" canexpand ="" class="table-row active premiumProfile-saudi-arabian-oil-company-saudi-aramco" href="https://www.forbes.com/companies/saudi-arabian-oil-company -saudi-aramco/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="saudi-arabian-oil-company-saudi-aramco"> \n <div class="rank first table-cell rank">\n 2.\n </div>\n <div class="organizationName second table-cell name">\n Saudi Arabian Oil Compan y (Saudi Aramco)\n </div>\n <div class="country table-cell country">\n Saudi Arabia\n </div>\n <div class="revenue table-cell sales (\$)">\n \$589.47 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$156.36 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$660.9 9 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$2,055.22 B\n </div>\n </a>\n <a aria-label="ICBC" canexpand="" class ="table-row active premiumProfile-icbc" href="https://www.forbes.com/companies/icbc/?list=global2000" rel="noopener noreferrer" style="background-c olor:" target="#\_blank" uri="icbc">\n <div class="rank first table-cell rank">\n </div>\n <div class="organizationName second table-cell" rank">\n </div> name">\n ICBC\n </div>\n <div class="country table-cell country">\n China\n </div>\n <div class="revenue table-cell sales (\$)">\n \$216.77 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$52.47 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$6,116.82 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$203.01 B\n </div>\n </a>\n <a aria-label="China Construction Bank" canexpand ="" class="table-row active premiumProfile-china-construction-bank" href="https://www.forbes.com/companies/china-construction-bank/?list=global200 0" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="china-construction-bank">\n <div class="rank first table-cell rank"> \n 4.\n </div>\n <div class="organizationName second table-cell name">\n China Construction Bank\n </div>\n <div class="country table-cell name">\n China Construction Bank\n <dout table name = country">\n China\n </div>\n <div class="revenue table-cell sales (\$)">\n \$203.08 B\n </div>\n <div class="profits table-cell profit (\$)"> \n \$48.25 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$4,977.48 B\n </div>\n <div class="marketValue table-cell market value" (\$)">\n \$172.99 B\n </div>\n <a>\n <a aria-label="Agricultural Bank of China" canexpand="" class="table-row active premiumProfile-agriculturalbank-of-china" href="https://www.forbes.com/companies/agricultural-bank-of-china/?list=global2000" rel="noopener noreferrer" style="background-colo r:" target="#\_blank" uri="agricultural-bank-of-china">\n <div class="rank first table-cell rank">\n 5.\n </div>\n <div class="organizationName">\n 5.\n </div>\n = 1. second table-cell name">\n Agricultural Bank of China\n </div>\n <div class="country table-cell country">\n China\n </div>\n <div class="re venue table-cell sales (\$)">\n \$186.14 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$37.92 B\n </div>\n <div class="assets ta ble-cell assets (\$)">\n \$5,356.86 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$141.82 B\n </div>\n <a aria-l abel="Bank of America" canexpand="" class="table-row active premiumProfile-bank-of-america" href="https://www.forbes.com/companies/bank-of-americ a/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="bank-of-america">\n <div class="rank first table-cel l rank">\n 6.\n </div>\n <div class="organizationName second table-cell name">\n Bank of America\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$133.84 B\n </div>\n <div class="profits table-cell profi t (\$)">\n \$28.62 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$3,194.66 B\n </div>\n <div class="marketValue table-cell market" value (\$)">\n \$220.82 B\n </div>\n </a>\n <a aria-label="Alphabet" canexpand="" class="table-row active premiumProfile-alphabet" href="https://w ww.forbes.com/companies/alphabet/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="alphabet">\n <div cla untry table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$282.85 B\n </div>\n <div class="profits" table-cell profit (\$)">\n \$58.59 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$369.49 B\n </div>\n <div class="marketValue tab le-cell market value (\$)">\n \$1,340.53 B\n </div>\n <a aria-label="ExxonMobil" canexpand="" class="table-row active premiumProfile-exxonmobil" href="https://www.forbes.com/companies/exxon-mobil/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" ur i="exxon-mobil">\n <div class="rank first table-cell rank">\n 8.\n </div>\n <div class="organizationName second table-cell name">\n ExxonMob il\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$393.16 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$61.69 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$369.37 B\n </div> \n <div class="marketValue table-cell market value (\$)">\n \$439.39 B\n </div>\n <a aria-label="Microsoft" canexpand="" class="table-row" active premiumProfile-microsoft" href="https://www.forbes.com/companies/microsoft/?list=global2000" rel="noopener noreferrer" style="background-col or:" target="#\_blank" uri="microsoft">\n <div class="rank first table-cell rank">\n 9.\n </div>\n <div class="organizationName second table-ce ll name">\n Microsoft\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales"  $(\$)">\n $207.59 B\n </div>\n <div class="profits table-cell profit (\$)">\n $69.02 B\n </div>\n <div class="assets table-cell assets (\$)">\n $69.02 B\n </div>\n <div class="assets table-cell assets (\$)">\n $69.02 B\n </div>\n <div class="assets table-cell assets (\$)">\n $69.02 B\n </div>$ \$380.09 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$2,309.84 B\n </div>\n <a aria-label="Apple" canexpand="" class="table-row active premiumProfile-apple" href="https://www.forbes.com/companies/apple/?list=global2000" rel="noopener noreferrer" style="backg round-color: "target="#\_blank" uri="apple">\n <div class="rank first table-cell rank">\n 10.\n </div>\n <div class="organizationName second ta ble-cell name">\n Apple\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$385.1 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$94.32 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$332.16 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$2,746.21 B\n </div>\n <a aria-label="Shell" canexpand="" class="table-row active premiumProfile-shell" href="https://www.forbes.com/companies/shell/?list=global2000" rel="noopener noreferrer" style="backg round-color:" target="#\_blank" uri="shell">\n <div class="rank first table-cell rank">\n 11.\n </div>\n <div class="organizationName second ta ble-cell name">\n Shell\n </div>\n <div class="country table-cell country">\n United Kingdom\n </div>\n <div class="revenue table-cell sale s (\$)">\n \$365.89 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$43.51 B\n </div>\n <div class="assets table-cell assets (\$)"> \n \$429.15 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$205.45 B\n </div>\n <a aria-label="Bank of China" ca nexpand="" class="table-row active premiumProfile-bank-of-china" href="https://www.forbes.com/companies/bank-of-china/?list=global2000" rel="noopen er noreferrer" style="background-color:" target="#\_blank" uri="bank-of-china">\n <div class="rank first table-cell rank">\n 12.\n </div>\n <di v class="organizationName second table-cell name">\n Bank of China\n </div>\n <div class="country table-cell country">\n China\n </div>\n < div class="revenue table-cell sales (\$)">\n \$158.23 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$33.23 B\n </div>\n <\div class="profits table-cell profit (\$)">\n \$33.23 B\n </div>\n <\div class="profits table-cell profit (\$)">\n \$33.23 B\n </div>\n <\div class="profits table-cell profit (\$)">\n \$33.23 B\n <\div class="profits table-cell profits table-cell prof ss="assets table-cell assets (\$)">\n \$4,421.76 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$122.67 B\n </div>\n </a >\n <a aria-label="Toyota Motor" canexpand="" class="table-row active premiumProfile-toyota-motor" href="https://www.forbes.com/companies/toyota-mo tor/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="toyota-motor">\n <div class="rank first table-cell rank">\n 13.\n </div>\n <div class="organizationName second table-cell name">\n Toyota Motor\n </div>\n <div class="country table-cell country tab try">\n Japan\n </div>\n <div class="revenue table-cell sales (\$)">\n \$270.49 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$18.5 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$542.5 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$188.2 B\n </div>\n </a>\n <a aria-label="Samsung Electronics" canexpand="canExpand" class="table-row active premiumProfile expandable premiumProf ile-samsung-electronics" rel="noopener noreferrer" style="background-color:" uri="samsung-electronics">\n <div class="rank first table-cell rank"> \n 14.\n </div>\n <div class="organizationName second table-cell name">\n Samsung Electronics\n </div>\n <div class="country table-cell cou ntry">\n South Korea\n </div>\n <div class="revenue table-cell sales (\$)">\n \$220.07 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$34.49 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$348.81 B\n </div>\n <div class="marketValue table-cell canExpand" market value (\$)">\n \$334.31 B\n </div>\n <a aria-label="UnitedHealth Group" canexpand="" class="table-row active premiumProfile-unitedhe alth-group" href="https://www.forbes.com/companies/unitedhealth-group/?list=global2000" rel="noopener noreferrer" style="background-color:" target ="#\_blank" uri="unitedhealth-group">\n <div class="rank first table-cell rank">\n 15.\n </div>\n <div class="organizationName second table-cel l name">\n UnitedHealth Group\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$335.94 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$20.7 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$283.68 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$460.19 B\n </div>\n <a aria-label="Ping An Insu rance Group" canexpand="" class="table-row active premiumProfile-ping-an-insurance" href="https://www.forbes.com/companies/ping-an-insurance/?list= global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="ping-an-insurance">\n <div class="rank first table-cell ran k">\n 16.\n </div>\n <div class="organizationName second table-cell name">\n Ping An Insurance Group\n </div>\n <div class="country table-c ell country">\n China\n </div>\n <div class="revenue table-cell sales (\$)">\n \$166.37 B\n </div>\n <div class="profits table-cell profit table-cell profit table-cell sales (\$)">\n \$166.37 B\n </div>\n <div class="profits table-cell profit table-cell profi (\$)">\n \$12.64 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$1,598.49 B\n </div>\n <div class="marketValue table-cell market v alue (\$)">\n \$138.56 B\n </div>\n <a aria-label="Wells Fargo" canexpand="" class="table-row active premiumProfile-wells-fargo" href="http s://www.forbes.com/companies/wells-fargo/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="wells-fargo"> \n <div class="rank first table-cell rank">\n 17.\n </div>\n <div class="organizationName second table-cell name">\n Wells Fargo\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$108.93 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$14.5 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$1,886.4 B\n </div>\n <div class="assets table-cell assets" ="marketValue table-cell market value (\$)">\n \$142.36 B\n </div>\n <a aria-label="Chevron" canexpand="" class="table-row active premiumPr ofile-chevron" href="https://www.forbes.com/companies/chevron/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blan k" uri="chevron">\n <div class="rank first table-cell rank">\n </div>\n <div class="organizationName second table-cell name">\n Chevro n\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$232.12 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$35.78 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$255.89 B\n </div> \n <div class="marketValue table-cell market value (\$)">\n \$303.54 B\n </div>\n </a>\n <a aria-label="PetroChina" canexpand="" class="table-row" active premiumProfile-petrochina" href="https://www.forbes.com/companies/petrochina/?list=global2000" rel="noopener noreferrer" style="background-c olor:" target="#\_blank" uri="petrochina">\n <div class="rank first table-cell rank">\n 18.\n </div>\n <div class="organizationName second tabl e-cell name">\n PetroChina\n </div>\n <div class="country table-cell country">\n China\n </div>\n <div class="revenue table-cell sales" (\$)">\n \$457.4 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$22.18 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$384.58 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$122.91 B\n </div>\n <a aria-label="HSBC Holdings" canexpa nd="" class="table-row active premiumProfile-hsbc-holdings" href="https://www.forbes.com/companies/hsbc-holdings/?list=global2000" rel="noopener no ss="organizationName second table-cell name">\n HSBC Holdings\n </div>\n <div class="country table-cell country">\n United Kingdom\n </div> class="assets table-cell assets (\$)">\n \$2,989.7 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$151.19 B\n </div>\n </a>\n <a aria-label="TotalEnergies" canexpand="" class="table-row active premiumProfile-totalenergies" href="https://www.forbes.com/companies/tota lenergies/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="totalenergies">\n <div class="rank first tab le-cell rank">\n 21.\n </div>\n <div class="organizationName second table-cell name">\n TotalEnergies\n </div>\n <div class="country tablecell country">\n France\n </div>\n <div class="revenue table-cell sales (\$)">\n \$257.59 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$21.12 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$293.03 B\n </div>\n <div class="marketValue table-cell market val ue (\$)">\n \$151.64 B\n </div>\n <a aria-label="Verizon Communications" canexpand="" class="table-row active premiumProfile-verizon-commun ications" href="https://www.forbes.com/companies/verizon-communications/?list=global2000" rel="noopener noreferrer" style="background-color:" targe t="#\_blank" uri="verizon-communications">\n <div class="rank first table-cell rank">\n </div>\n <div class="organizationName second table" e-cell name">\n Verizon Communications\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue t able-cell sales (\$)">\n \$136.19 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$21.59 B\n </div>\n <div class="assets table-cell profit (\$)">\n \$21.59 B\n </div>\n <div class="assets table-cell profit (\$)">\n \$21.59 B\n </div>\n <div class="assets table-cell profit (\$)">\n \$21.59 B\n </div> l assets (\$)">\n \$377.72 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$159.04 B\n </div>\n </a>\n <a aria-label="Wal mart" canexpand="" class="table-row active premiumProfile-walmart" href="https://www.forbes.com/companies/walmart/?list=global2000" rel="noopener n oreferrer" style="background-color:" target="#\_blank" uri="walmart">\n <div class="rank first table-cell rank">\n 23.\n </div>\n <div class="o s="revenue table-cell sales (\$)">\n \$611.29 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$11.68 B\n </div>\n <div class="asse" ts table-cell assets (\$)">\n \$243.2 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$409.12 B\n </div>\n <a aria -label="Citigroup" canexpand="" class="table-row active premiumProfile-citigroup" href="https://www.forbes.com/companies/citigroup/?list=global200 0" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="citigroup">\n <div class="rank first table-cell rank">\n </div class="rank first table-cell rank">\n </dr> iv>\n <div class="organizationName second table-cell name">\n Citigroup\n </div>\n <div class="country table-cell country">\n United States \n </div>\n <div class="revenue table-cell sales (\$)">\n \$120.42 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$15.02 B\n </div iv>\n <div class="assets table-cell assets (\$)">\n \$2,455.11 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$90.03 B\n </div>\n </a>\n <a aria-label="China Mobile" canexpand="" class="table-row active premiumProfile-china-mobile" href="https://www.forbes.com/compani es/china-mobile/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="china-mobile">\n <div class="rank firs t table-cell rank">\n 25.\n </div>\n <div class="organizationName second table-cell name">\n China Mobile\n </div>\n <div class="country ta ble-cell country">\n Hong Kong\n </div>\n <div class="revenue table-cell sales (\$)">\n \$140.53 B\n </div>\n <div class="profits table-cell sales" profit (\$)">\n \$18.76 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$290 B\n </div>\n <div class="marketValue table-cell market" value (\$)">\n \$186.41 B\n </div>\n <a aria-label="China Merchants Bank" canexpand="" class="table-row active premiumProfile-china-merchan ts-bank" href="https://www.forbes.com/companies/china-merchants-bank/?list=global2000" rel="noopener noreferrer" style="background-color:" target ="#\_blank" uri="china-merchants-bank">\n <div class="rank first table-cell rank">\n <div class="organizationName second table-cell rank">\n <d ell name">\n China Merchants Bank\n </div>\n <div class="country table-cell country">\n China\n </div>\n <div class="revenue table-cell sal es (\$)">\n \$72.44 B\n  $</div>\n <div class="profits table-cell profit (<math>\$$ )">\n \$20.56 B\n  $</div>\n <div class="assets table-cell assets (<math>\$$ )"> \n \$1,458.48 B\n </div>\n <div>\n <div>\n <div>\n </div>\n </a>\n <a aria-label="Postal Savings" Bank Of China (PSBC)" canexpand="" class="table-row active premiumProfile-postal-savings-bank-of-china" href="https://www.forbes.com/companies/post al-savings-bank-of-china/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="postal-savings-bank-of-china"> \n <div class="rank first table-cell rank">\n 27.\n </div>\n <div class="organizationName second table-cell name">\n Postal Savings Bank Of </div>\n <div class="profits table-cell profit (\$)">\n \$12.78 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$2,023.57 B\n </div >\n <div class="marketValue table-cell market value (\$)">\n \$110.18 B\n </div>\n <a aria-label="BP" canexpand="" class="table-row active" premiumProfile-bp" href="https://www.forbes.com/companies/bp/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="bp">\n <div class="rank first table-cell rank">\n 28.\n </div>\n <div class="organizationName second table-cell name">\n BP\n </div>\n <div class="country table-cell country">\n United Kingdom\n </div>\n <div class="revenue table-cell sales (\$)">\n \$248.11 B\n </div>\n <div</pre> class="profits table-cell profit (\$)">\n \$25.89 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$278.66 B\n </div>\n <div class="assets table-cell assets" ="marketValue table-cell market value (\$)">\n \$108.3 B\n </div>\n <a aria-label="Volkswagen Group" canexpand="" class="table-row active p remiumProfile-volkswagen-group" href="https://www.forbes.com/companies/volkswagen-group/?list=global2000" rel="noopener noreferrer" style="backgroup" https://www.forbes.com/companies/volkswagen-group/?list=global2000" rel="noopener noreferrer" style="backgroup" https://www.forbes.com/companies/volkswagen-group" https://www.forbes.com/companies/volkswagen-group https://www.forbes.com/companies/wagen-group https://www.forbes.com/companies/wagen-group https://www.forbes.com/companies/wagen-group https://www.forbes.com/com/companies/wagen-group https://www.forbes.com/companies/wagen-gr nd-color:" target="#\_blank" uri="volkswagen-group">\n <div class="rank first table-cell rank">\n </div>\n </div>\n <div class="organizationName s econd table-cell name">\n Volkswagen Group\n </div>\n <div class="country table-cell country">\n Germany\n </div>\n <div class="revenue table-cell country">\n Germany\n <\div <\div >\n <\div le-cell sales (\$)">\n \$293.47 B\n  $</div>\n <div class="profits table-cell profit (<math>\$$ )">\n \$15.63 B\n  $</div>\n <div class="assets table-cell profit">\n <math>$15.63$  B\n \$15.63 B\n assets (\$)">\n \$633.78 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$70.16 B\n </div>\n </a>\n <a aria-label="Morgan" Stanley" canexpand="" class="table-row active premiumProfile-morgan-stanley" href="https://www.forbes.com/companies/morgan-stanley/?list=global200 0" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="morgan-stanley">\n <div class="rank first table-cell rank">\n </div>\n <div class="organizationName second table-cell name">\n Morgan Stanley\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$74.16 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$10.34 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$1,199.9 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$141.76 B\n </div>\n </a>\n <a aria-label="Meta Platforms" canexpand="" class="table-row active premiumProfile-meta-platforms" href="https://www.forbes.co m/companies/meta-platforms/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="meta-platforms">\n <div cla ss="rank first table-cell rank">\n 31.\n </div>\n <div class="organizationName second table-cell name">\n Meta Platforms\n </div>\n <div cl ass="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$117.35 B\n </div>\n <div class ="profits table-cell profit (\$)">\n \$21.44 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$184.49 B\n </div>\n <div class="marke tValue table-cell market value (\$)">\n \$599.82 B\n </div>\n <a aria-label="Sinopec" canexpand="" class="table-row active premiumProfile-s" inopec" href="https://www.forbes.com/companies/sinopec/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri ="sinopec">\n <div class="rank first table-cell rank">\n 32.\n </div>\n <div class="organizationName second table-cell name">\n Sinopec\n </div>\n <div class="country table-cell country">\n China\n </div>\n <div class="revenue table-cell sales (\$)">\n \$453.56 B\n </div>\n <div v class="profits table-cell profit (\$)">\n \$9.85 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$283.3 B\n </div>\n <div class="assets table-cell assets" ="marketValue table-cell market value (\$)">\n \$114.32 B\n </div>\n <a aria-label="BNP Paribas" canexpand="" class="table-row active premi umProfile-bnp-paribas" href="https://www.forbes.com/companies/bnp-paribas/?list=global2000" rel="noopener noreferrer" style="background-color:" tar get="#\_blank" uri="bnp-paribas">\n <div class="rank first table-cell rank">\n 33.\n </div>\n <div class="organizationName second table-cell na me">\n BNP Paribas\n </div>\n <div class="country table-cell country">\n France\n </div>\n <div class="revenue table-cell sales (\$)">\n \$99.47 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$10.11 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$2,845.6 9 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$78.59 B\n </div>\n <a aria-label="Goldman Sachs Group" canexpan d="" class="table-row active premiumProfile-goldman-sachs-group" href="https://www.forbes.com/companies/goldman-sachs-group/?list=global2000" rel ="noopener noreferrer" style="background-color:" target="#\_blank" uri="goldman-sachs-group">\n <div class="rank first table-cell rank">\n 33.\n </div>\n <div class="organizationName second table-cell name">\n Goldman Sachs Group\n </div>\n <div class="country table-cell country">\n U nited States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$79.45 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$10. 56 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$1,538.35 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$108.72 B\n </div>\n </a>\n <a aria-label="Tencent Holdings" canexpand="" class="table-row active premiumProfile-tencent-holdings" href="https://w ww.forbes.com/companies/tencent-holdings/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="tencent-holding" target="tencent-holding" target="#\_blank" uri="tencent-holding" target="tencent-holding" target gs">\n <div class="rank first table-cell rank">\n 35.\n </div>\n <div class="organizationName second table-cell name">\n Tencent Holdings\n </div>\n <div class="country table-cell country">\n China\n </div>\n <div class="revenue table-cell sales (\$)">\n \$82.41 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$27.26 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$227.01 B\n </div>\n <div class="assets table-cell assets" ="marketValue table-cell market value (\$)">\n \$415.36 B\n </div>\n <a aria-label="Amazon" canexpand="" class="table-row active premiumPro file-amazon" href="https://www.forbes.com/companies/amazon/?list=global2000" rel="noopener noreferrer" style="background-color:" target="# blank" u ri="amazon">\n <div class="rank first table-cell rank">\n 36.\n </div>\n <div class="organizationName second table-cell name">\n Amazon\n </div>\n <div class="revenue table-cell sales (\$)">\n \$524.9 B\n </div> \n <div class="profits table-cell profit (\$)">\n \$4.29 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$464.38 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$1,084.06 B\n </div>\n <a aria-label="Allianz" canexpand="" class="table-row active pr emiumProfile-allianz" href="https://www.forbes.com/companies/allianz/?list=global2000" rel="noopener noreferrer" style="background-color:" target uri="allianz">\n <div class="rank first table-cell rank">\n 37.\n </div>\n <div class="organizationName second table-cell name">\n Allianz\n </div>\n <div class="country table-cell country">\n Germany\n </div>\n <div class="revenue table-cell sales (\$)">\n \$134.26 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$7.08 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$1,139.49 B\n </div> \n <div class="marketValue table-cell market value (\$)">\n \$95.45 B\n </div>\n <a aria-label="RBC" canexpand="" class="table-row active premiumProfile-rbc" href="https://www.forbes.com/companies/rbc/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blan k" uri="rbc">\n <div class="rank first table-cell rank">\n 38.\n </div>\n <div class="organizationName second table-cell name">\n RBC\n </div> iv>\n <div class="country table-cell country">\n Canada\n </div>\n <div class="revenue table-cell sales (\$)">\n \$60.22 B\n </div>\n <div c lass="profits table-cell profit (\$)">\n \$11.4 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$1,448.66 B\n </div>\n <div class="assets table-cell assets" ="marketValue table-cell market value (\$)">\n \$135.15 B\n </div>\n <a aria-label="Pfizer" canexpand="canExpand" class="table-row active p remiumProfile expandable premiumProfile-pfizer" rel="noopener noreferrer" style="background-color:" uri="pfizer">\n <div class="rank first table-c ell rank">\n 39.\n </div>\n <div class="organizationName second table-cell name">\n Pfizer\n </div>\n <div class="country table-cell country" y">\n United States\n </div>\n <div class="revenue table-cell sales (\$)">\n \$92.95 B\n </div>\n <div class="profits table-cell profit (\$)"> \n \$29.04 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$197.21 B\n </div>\n <div class="marketValue table-cell canExpand marke t value (\$)">\n \$216.28 B\n </div>\n <a>\n <a aria-label="Johnson & amp; Johnson" canexpand="" class="table-row active premiumProfile-johnson-johnson-johnson" canexpand="" class="table-row active premiumProfile-johnson-johnson-johnson" canexpand="" class="table-row active premiumProfile-johnson-johnso hnson" href="https://www.forbes.com/companies/johnson-johnson/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blan k" uri="johnson-johnson">\n <div class="rank first table-cell rank">\n </div>\n </div>\n <div class="organizationName second table-cell name">\n Johnson & amp; Johnson\n </div>\n <div class="country table-cell country">\n United States\n </div>\n <div class="revenue table-cell sales" (\$)">\n \$96.31 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$12.72 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$195.97 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$422.76 B\n </div>\n <a aria-label="Deutsche Telekom" cane xpand="" class="table-row active premiumProfile-deutsche-telekom" href="https://www.forbes.com/companies/deutsche-telekom/?list=global2000" rel="no opener noreferrer" style="background-color:" target="#\_blank" uri="deutsche-telekom">\n <div class="rank first table-cell rank">\n </div> \n <div class="organizationName second table-cell name">\n Deutsche Telekom\n </div>\n <div class="country table-cell country">\n Germany\n  $</div>\n <div class="revenue table-cell sales ($)">\n $120.72 B\n </div>\n <div class="profits table-cell profit ($)">\n $8.41 B\n </div>\n$ <div class="assets table-cell assets (\$)">\n \$342.62 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$117.18 B\n </div> \n </a>\n <a aria-label="Mercedes-Benz Group" canexpand="" class="table-row active premiumProfile-mercedes-benz-group" href="https://www.forbes.co m/companies/mercedes-benz-group/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="mercedes-benz-group">\n <div class="rank first table-cell rank">\n 42.\n </div>\n <div class="organizationName second table-cell name">\n Mercedes-Benz Group\n </di>

v>\n <div class="country table-cell country">\n Germany\n </div>\n <div class="revenue table-cell sales (\$)">\n \$158.78 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$15.55 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$286 B\n </div>\n <div class="mar ketValue table-cell market value (\$)">\n \$78.49 B\n </div>\n <a aria-label="TD Bank Group" canexpand="" class="table-row active premiumPr ofile-td-bank-group" href="https://www.forbes.com/companies/td-bank-group/?list=global2000" rel="noopener noreferrer" style="background-color:" tar get="#\_blank" uri="td-bank-group">\n <div class="rank first table-cell rank">\n </div>\n </div>\n <div class="organizationName second table-cell rank">\n </div> name">\n TD Bank Group\n </div>\n <div class="country table-cell country">\n Canada\n </div>\n <div class="revenue table-cell sales (\$)">\n  $56.72 \text{ B} < \text{div} < \text{div$ 1 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$112.92 B\n </div>\n <a aria-label="Taiwan Semiconductor" canexp and="" class="table-row active premiumProfile-taiwan-semiconductor" href="https://www.forbes.com/companies/taiwan-semiconductor/?list=global2000" r el="noopener noreferrer" style="background-color:" target="#\_blank" uri="taiwan-semiconductor">\n <div class="rank first table-cell rank">\n 4 4.\n </div>\n <div class="organizationName second table-cell name">\n Taiwan Semiconductor\n </div>\n <div class="country table-cell country"  $y">\n$  Taiwan $\n$  </div>\n <div class="revenue table-cell sales (\$)">\n \$75.03 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$3 3.57 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$165.72 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$423.17 B\n </div>\n </a>\n <a aria-label="Reliance Industries" canexpand="" class="table-row active premiumProfile-reliance-industries" href="htt ps://www.forbes.com/companies/reliance-industries/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="relia nce-industries">\n <div class="rank first table-cell rank">\n </div>\n <div class="organizationName second table-cell name">\n Relianc e Industries\n </div>\n <div class="country table-cell country">\n India\n </div>\n <div class="revenue table-cell sales (\$)">\n \$109.43 B \n </div>\n <div class="profits table-cell profit (\$)">\n \$8.3 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$208.5 B\n </div> \n <div class="marketValue table-cell market value (\$)">\n \$189.19 B\n </div>\n <a aria-label="BMW Group" canexpand="" class="table-row" active premiumProfile-bmw-group" href="https://www.forbes.com/companies/bmw-group/?list=global2000" rel="noopener noreferrer" style="background-col or:" target="#\_blank" uri="bmw-group">\n <div class="rank first table-cell rank">\n 46.\n </div>\n <div class="organizationName second table-cell rank">\n 46.\n </div>\n <div class="organizationName second table-cell rank">\n 46.\n </div>\n <div class="organizationName second table-cell rank">\n 46.\n </div> ell name">\n BMW Group\n </div>\n <div class="country table-cell country">\n Germany\n </div>\n <div class="revenue table-cell sales (\$)"> \$154.25 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$11.67 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$2

68.98 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$75.77 B\n </div>\n <a aria-label="LVMH Moët Hennessy Louis" nessy-louis-vuitton/?list=global2000" rel="noopener noreferrer" style="background-color:" target="#\_blank" uri="lvmh-moet-hennessy-louis-vuitton"> \n <div class="rank first table-cell rank">\n 47.\n </div>\n <div class="organizationName second table-cell name">\n LVMH Moët Hennessy Loui s Vuitton\n </div>\n <div class="country table-cell country">\n France\n </div>\n <div class="revenue table-cell sales (\$)">\n \$83.22 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$14.8 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$143.7 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$482.45 B\n </div>\n <a aria-label="AXA Group" canexpand="" class="table-row acti</pre> ve premiumProfile-axa-group" href="https://www.forbes.com/companies/axa-group/?list=global2000" rel="noopener noreferrer" style="background-color:" target="# blank" uri="axa-group">\n <div class="rank first table-cell rank">\n </div>\n </div>\n <div class="organizationName second table-cell n ame">\n AXA Group\n </div>\n <div class="country table-cell country">\n France\n </div>\n <div class="revenue table-cell sales (\$)">\n \$1 10.92 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$6.83 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$701.25 B \n </div>\n <div class="marketValue table-cell market value (\$)">\n \$70.07 B\n </div>\n </a>\n <a aria-label="Santander" canexpand="" class="t able-row active premiumProfile-santander" href="https://www.forbes.com/companies/santander/?list=global2000" rel="noopener noreferrer" style="backg round-color: "target="#\_blank" uri="santander">\n <div class="rank first table-cell rank">\n </div>\n </div>\n <div class="organizationName secon d table-cell name">\n Santander\n </div>\n <div class="country table-cell country">\n Spain\n </div>\n <div class="revenue table-cell sales (\$)">\n \$87.74 B\n </div>\n <div class="profits table-cell profit (\$)">\n \$10.09 B\n </div>\n <div class="assets table-cell assets (\$)">\n \$1,900.64 B\n </div>\n <div class="marketValue table-cell market value (\$)">\n \$56.18 B\n </div>\n <a aria-label="Nestlé" canexpand="" class="table-row active premiumProfile-nestle" href="https://www.forbes.com/companies/nestle/?list=global2000" rel="noopener noreferrer" style="bac kground-color:" target="#\_blank" uri="nestle">\n <div class="rank first table-cell rank">\n <0iv class="organizationName second" table-cell name">\n Nestlé\n </div>\n <div class="country table-cell country">\n Switzerland\n </div>\n <div class="revenue table-cell sale s (\$)">\n \$98.87 B\n  $</div>\n <math><div$  class="profits table-cell profit (\$)">\n \$9.71 B\n </div>\n <div class="assets table-cell assets (\$)">\n

 $146.11 \, \text{h} </\text{div} < \text{div} < \text{d$ 

soup.find\_all("div", attrs={'class':'organizationName second table-cell name'})[1950].text.strip()

soup.find\_all("div", attrs={'class':'country table-cell country'})[1999].text.strip()

soup.find\_all("div", attrs={'class':'revenue table-cell sales (\$)'})[1999].text.strip()

soup.find\_all("div", attrs={'class':'assets table-cell assets (\$)'})[1999].text.strip()

soup.find\_all("div", attrs={'class':'marketValue table-cell market value (\$)'})[1950].text

rowlength = len(soup.find\_all("div", attrs={'class':'organizationName second table-cell name'}))

name = soup.find\_all("div", attrs={'class':'organizationName second table-cell name'})[i].text.strip()

country = soup.find\_all("div", attrs={'class':'country table-cell country'})[i].text.strip()
sales = soup.find\_all("div", attrs={'class':'revenue table-cell sales (\$)'})[i].text.strip()
profits = soup.find\_all("div", attrs={'class':'profits table-cell profit (\$)'})[i].text.strip()
assets = soup.find\_all("div", attrs={'class':'assets table-cell assets (\$)'})[i].text.strip()

In [11]: | soup.find\_all("div", attrs={'class':'profits table-cell profit (\$)'})[1999].text.strip()

'United Therapeutics'

'France'

'\$3.18 B'

'\$681.7 M'

'\$551 M'

rowlength

data = []

for i in range(0, rowlength):

'Saudi Arabia', '\$589.47 B', '\$156.36 B', '\$660.99 B'),

import pandas as pd

data\_company.head(10)

data\_company.info()

Column

Country

Sales

Profit Assets

dtypes: object(5)
memory usage: 78.2+ KB

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 5 columns):

Company Name 2000 non-null

def convert\_total\_votes(votes\_str):

if votes\_str[-1] == 'M':
 multiplier = 1e6
elif votes\_str[-1] == 'B':
 multiplier = 1e9

1 Saudi Arabian Oil Company (Saudi Aramco)

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 5 columns):

Company Name 2000 non-null object

dtypes: category(1), float64(3), object(1)

1 Saudi Arabian Oil Company (Saudi Aramco)

2000 non-null

2000 non-null

2000 non-null

2000 non-null

pd.read\_csv("The World's Largest Companies.csv").head()

**Company Name** 

China Construction Bank
Agricultural Bank of China

Non-Null Count Dtype

category

float64

float64

float64

Country

data company.to csv("The World's Largest Companies.csv", index=False)

multiplier = 1

data\_company.head(10)

data\_company.info()

Column

Country

-----

Sales

Profit

Assets

memory usage: 67.1+ KB

In [82]:

Out[82]:

In [83]:

In [84]:

Out[85]:

2

6

7

8

0

3

0

2

3

4

'China', '\$186.14 B', '\$37.92 B', '\$5,356.86 B'),

('Agricultural Bank of China',

data.append((name, country, sales, profits, assets))

('ICBC', 'China', '\$216.77 B', '\$52.47 B', '\$6,116.82 B'),

**Company Name** 

China Construction Bank

Agricultural Bank of China

1 Saudi Arabian Oil Company (Saudi Aramco) Saudi Arabia \$589.47 B \$156.36 B

**ICBC** 

Non-Null Count Dtype

2000 non-null

2000 non-null

2000 non-null

2000 non-null

return int(float(votes\_str[:-1]) \* multiplier)

**Company Name** 

China Construction Bank Agricultural Bank of China

data\_company['Sales'] = data\_company['Sales'].apply(convert\_total\_votes)
data\_company['Profit'] = data\_company['Profit'].apply(convert\_total\_votes)
data\_company['Assets'] = data\_company['Assets'].apply(convert\_total\_votes)

data\_company['Country'] = data\_company['Country'].astype('category')

Country

('Saudi Arabian Oil Company (Saudi Aramco)',

[('JPMorgan Chase', 'United States', '\$179.93 B', '\$41.8 B', '\$3,744.3 B'),

('China Construction Bank', 'China', '\$203.08 B', '\$48.25 B', '\$4,977.48 B'),

('Bank of America', 'United States', '\$133.84 B', '\$28.62 B', '\$3,194.66 B'),

Country

JPMorgan Chase United States \$179.93 B

Bank of America United States \$133.84 B

Apple United States

Alphabet United States \$282.85 B

Microsoft United States \$207.59 B

ExxonMobil United States \$393.16 B

object

object

object

object

object

data\_company = pd.DataFrame(data, columns=('Company Name', 'Country', 'Sales', 'Profit', 'Assets'))

China \$216.77 B

China \$203.08 B

China \$186.14 B

\$385.1 B

Profit

\$41.8 B

\$58.59 B

\$61.69 B

\$69.02 B

\$94.32 B

data\_company['Sales'] = data\_company['Sales'].str.replace('\$', '').str.replace(' ', '').str.replace(',', '')
data\_company['Profit'] = data\_company['Profit'].str.replace('\$', '').str.replace(' ', '').str.replace(',', '')
data\_company['Assets'] = data\_company['Assets'].str.replace('\$', '').str.replace(' ', '').str.replace(',', '')

data\_company['Sales'] = data\_company['Sales'].str.replace('\$', '').str.replace(' ', '').str.replace(',', '')

data\_company[['Sales', 'Profit', 'Assets']] = data\_company[['Sales', 'Profit', 'Assets']].astype('float')

Sales

Saudi Arabia 5.894700e+11 1.563600e+11 6.609900e+11

China 2.167700e+11 5.247000e+10 6.116820e+12 China 2.030800e+11 4.825000e+10 4.977480e+12

China 1.861400e+11 3.792000e+10 5.356860e+12

JPMorgan Chase United States 1.799300e+11 4.180000e+10 3.744300e+12

Bank of America United States 1.338400e+11 2.862000e+10 3.194660e+12

Alphabet United States 2.828500e+11 5.859000e+10 3.694900e+11

ExxonMobil United States 3.931600e+11 6.169000e+10 3.693700e+11

Microsoft United States 2.075900e+11 6.902000e+10 3.800900e+11

Apple United States 3.851000e+11 9.432000e+10 3.321600e+11

Sales

Saudi Arabia 5.894700e+11 1.563600e+11 6.609900e+11

China 2.167700e+11 5.247000e+10 6.116820e+12 China 2.030800e+11 4.825000e+10 4.977480e+12

China 1.861400e+11 3.792000e+10 5.356860e+12

JPMorgan Chase United States 1.799300e+11 4.180000e+10 3.744300e+12

**Profit** 

**Profit** 

**Assets** 

re version. In addition, single character regular expressions will \*not\* be treated as literal strings when regex=True.

re version. In addition, single character regular expressions will \*not\* be treated as literal strings when regex=True. data\_company['Profit'] = data\_company['Profit'].str.replace('\$', '').str.replace(' ', '').str.replace(',', '')

re version. In addition, single character regular expressions will \*not\* be treated as literal strings when regex=True. data\_company['Assets'] = data\_company['Assets'].str.replace('\$', '').str.replace(' ', '').str.replace(',', '')

**Assets** 

\$3,744.3 B

\$660.99 B

\$369.49 B

\$369.37 B

\$380.09 B

\$332.16 B

C:\Users\pinos\AppData\Local\Temp\ipykernel\_2304\236497370.py:1: FutureWarning: The default value of regex will change from True to False in a futu

C:\Users\pinos\AppData\Local\Temp\ipykernel\_2304\236497370.py:2: FutureWarning: The default value of regex will change from True to False in a futu

C:\Users\pinos\AppData\Local\Temp\ipykernel\_2304\236497370.py:3: FutureWarning: The default value of regex will change from True to False in a futu

**Assets** 

\$52.47 B \$6,116.82 B

\$48.25 B \$4,977.48 B

\$37.92 B \$5,356.86 B

\$28.62 B \$3,194.66 B

('Alphabet', 'United States', '\$282.85 B', '\$58.59 B', '\$369.49 B'), ('ExxonMobil', 'United States', '\$393.16 B', '\$61.69 B', '\$369.37 B'), ('Microsoft', 'United States', '\$207.59 B', '\$69.02 B', '\$380.09 B'), ('Apple', 'United States', '\$385.1 B', '\$94.32 B', '\$332.16 B')]

2000

In [18]: | data[:10]

Out[8]:

Out[9]:

Out[10]:

Out[11]:

In [12]:

Out[12]:

In [13]:

Out[13]:

In [14]:

Out[14]:

In [15]:

In [16]:

In [63]:

Out[63]:

2

4

6

8

0

1

3

In [64]: