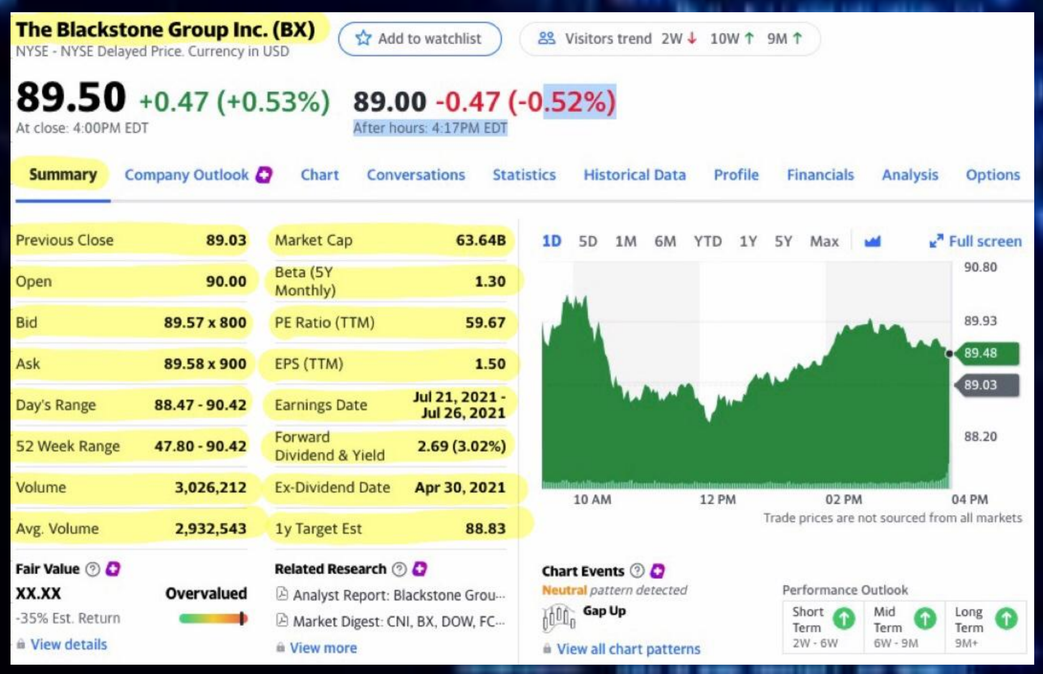
Stock Web Scraper

Nathan Chung - Computer Science
Science, Discover, and the Universe
Email: nathanchung80@gmail.com

Research Question: How could you implement a web scraper for stocks?

- ➔ Web scraping consists of gathering data from web sources for extraction and analysis.
- ➔ Implementation for program was built using Python 3, HTML, Python LXML.
- ➔ Using target URL to extract attributes for focus in datasource webpage.

Stock Metric Data Fields from Yahoo Finance that will be scraped (Data Collected)



- ➔ Running the web scraper, giving a ticker argument to reference a stock abbreviation. (In this case: TSLA)
- ➔ Information in red is the output file that shows all data that was pulled.

```
python3 yahoofinance.py -h

usage: yahoo_finance.py [-h] TSLA
positional arguments: TSLA optional arguments: -h, --help\
                      show this help message and exit
```


```
{
    "Previous Close": "293.16",
    "Open": "295.06",
    "Bid": "298.51 x 800",
    "Ask": "298.88 x 900",
    "Day's Range": "294.48 - 301.00",
    "52 Week Range": "170.27 - 327.85",
    "Volume": "36,263,602",
    "Avg. Volume": "50,925,925",
    "Market Cap": "1.29T",
    "Beta (5Y Monthly)": "1.17",
    "PE Ratio (TTM)": "23.38",
    "EPS (TTM)": 12.728,
    "Earnings Date": "2020-07-28 to 2020-08-03",
    "Forward Dividend & Yield": "3.28 (1.13%)",
    "Ex-Dividend Date": "May 08, 2020",
    "1y Target Est": 308.91,
    "ticker": "AAPL",
    "url": "http://finance.yahoo.com/quote/AAPL?p=AAPL"
}
```


METHODOLOGY

- ➔ Download target content from web source.
- ➔ Extract unstructured data to reformat into structured data.
- ➔ Python requests are used to download HTML content from web source.
- ➔ LXML is used for parsing structure of HTML.

```
In [ ]: from lxml import html
        import requests
        import json
        import argparse
        from collections import OrderedDict


        def get_headers():
            return {"accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,\
                    image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9",
                    "accept-encoding": "gzip, deflate, br",
                    "accept-language": "en-GB,en;q=0.9,en-US;q=0.8,ml;q=0.7",
                    "cache-control": "max-age=0",
                    "dnt": "1",
                    "sec-fetch-dest": "document",
                    "sec-fetch-mode": "navigate",
                    "sec-fetch-site": "none",
                    "sec-fetch-user": "?1",
                    "upgrade-insecure-requests": "1",
                    "user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 \
                    (KHTML, like Gecko) Chrome/81.0.4044.122 Safari/537.36"}
```

ANALYSIS OF DATA COLLECTED

- ➔ Stock data that was scraped is exported onto a table for more comprehensive analysis (easier to read).
- ➔ Competitive stock's metrics are laid out in same columns which can be visually more efficient when comparing the performances of stocks.

				Account Balance		
Account	Symbol	Name	Category	Shares	Price	Value
Vanguard Rollover IRA	VTSAX	Vanguard Total Stock Market Index Fund Admiral Shares	US Stocks	3,000.000	53.67	$161,010.00
Vanguard Rollover IRA	VTIAX	Vanguard Total International Stock Index Fund Admiral Shares	International Stocks	1,800.000	28.67	$51,606.00
Vanguard Rollover IRA	VBTLX	Vanguard Total Bond Market Index Fund Admiral Shares	Bonds	2,000.000	10.81	$21,620.00
Vanguard Rollover IRA	VGSLX	Vanguard REIT Index Fund Admiral Shares	Alternatives	200.000	113.48	$22,696.00
Vanguard Rollover IRA	VTI	Vanguard Total Stock Market ETF	US Stocks	200.000	110.3	$22,060.00
Vanguard Roth IRA	VTSAX	Vanguard Total Stock Market Index Fund Admiral Shares	US Stocks	2,000.000	53.67	$107,340.00
Vanguard Roth IRA	VTIAX	Vanguard Total International Stock Index Fund Admiral Shares	International Stocks	1,400.000	28.67	$40,138.00
Vanguard Roth IRA	VBMFX	Vanguard Total Bond Market Index Fund Investor Shares	Bonds	1,300.000	10.81	$14,053.00
Vanguard Roth IRA	VGSIX	Vanguard REIT Index Fund Investor Shares	Alternatives	500.000	26.59	$13,295.00
PCS 401K	VTSAX	Vanguard Total Stock Market Index Fund Admiral Shares	US Stocks	700.000	53.67	$37,569.00
PCS 401K	VTIAX	Vanguard Total International Stock Index Fund Admiral Shares	International Stocks	300.000	28.67	$8,601.00
PCS 401K	VBTLX	Vanguard Total Bond Market Index Fund Admiral Shares	Bonds	400.000	10.81	$4,324.00
PCS 401K	VGSLX	Vanguard REIT Index Fund Admiral Shares	Alternatives	40.000	113.48	$4,539.20
HSA Bank	VTI	Vanguard Total Stock Market ETF	US Stocks	150.000	110.3	$16,545.00
HSA Bank	VEU	Vanguard FTSE All-World ex-US ETF	International Stocks	80.000	51.7	$4,136.00
HSA Bank	BND	Vanguard Total Bond Market ETF	Bonds	20.000	82.06	$1,641.20
HSA Bank	VNQ	Vanguard REIT Index Fund	Alternatives	25.000	80.05	$2,001.25
HSA Bank	TDAXX	TDAM Money Market Class A	Cash	30.000	1	$30.00
						$533,204.65

Suggestions for Future Research

- ➔ Adding more functionality than simple web scraping for relevant articles/news about a stock.
- ➔ Could integrate Python SKLearn Linear Regression Models to implement a stock prediction tool.
- ➔ Explore other company API's like JPMorgan's Perspective to add a live data graph for stocks.
- ➔ Widen target databases that are used to get stock info. (Google, Bloomberg, Wall Street)

Citations:

- ➔ https://realpython.com/beautiful-soup-web-scraper-python/
- ➔ https://www.edureka.co/blog/web-scraping-with-python/

Mentor: James Wang