Extracting and Visualizing Stock Data

Description

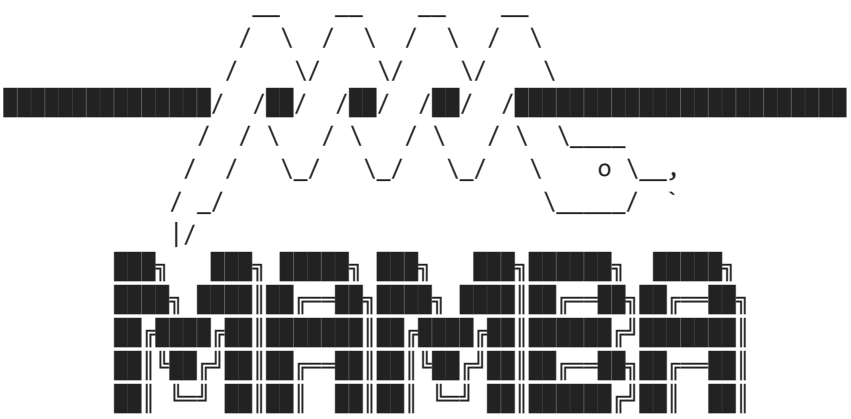
Extracting essential data from a dataset and displaying it is a necessary part of data science; therefore individuals can make correct decisions based on the data. In this assignment, you will extract some stock data, you will then display this data in a graph.

Table of Contents

Estimated Time Needed: **30 min**

In [1]:
```
!pip install yfinance
!mamba install bs4==4.10.0 -y
#!pip install bs4
!pip install nbformat==4.2.0
!pip install --upgrade nbformat
```

```
Requirement already satisfied: yfinance in /opt/conda/envs/Python-RT23.1/lib/python3.10/sit
e-packages (0.2.4)
Requirement already satisfied: pandas>=1.3.0 in /opt/conda/envs/Python-RT23.1/lib/python3.1
0/site-packages (from yfinance) (1.5.3)
Requirement already satisfied: numpy>=1.16.5 in /opt/conda/envs/Python-RT23.1/lib/python3.1
0/site-packages (from yfinance) (1.23.5)
Requirement already satisfied: requests>=2.26 in /opt/conda/envs/Python-RT23.1/lib/python3.
10/site-packages (from yfinance) (2.31.0)
Requirement already satisfied: multitasking>=0.0.7 in /opt/conda/envs/Python-RT23.1/lib/pyt
hon3.10/site-packages (from yfinance) (0.0.11)
Requirement already satisfied: lxml>=4.9.1 in /opt/conda/envs/Python-RT23.1/lib/python3.10/
site-packages (from yfinance) (4.9.2)
Requirement already satisfied: appdirs>=1.4.4 in /opt/conda/envs/Python-RT23.1/lib/python3.
10/site-packages (from yfinance) (1.4.4)
Requirement already satisfied: pytz>=2022.5 in /opt/conda/envs/Python-RT23.1/lib/python3.10
/site-packages (from yfinance) (2022.7)
Requirement already satisfied: frozendict>=2.3.4 in /opt/conda/envs/Python-RT23.1/lib/pytho
n3.10/site-packages (from yfinance) (2.4.0)
Requirement already satisfied: cryptography>=3.3.2 in /opt/conda/envs/Python-RT23.1/lib/pyt
hon3.10/site-packages (from yfinance) (41.0.7)
Requirement already satisfied: beautifulsoup4>=4.11.1 in /opt/conda/envs/Python-RT23.1/lib/
python3.10/site-packages (from yfinance) (4.12.3)
Requirement already satisfied: html5lib>=1.1 in /opt/conda/envs/Python-RT23.1/lib/python3.1
0/site-packages (from yfinance) (1.1)
Requirement already satisfied: soupsieve>1.2 in /opt/conda/envs/Python-RT23.1/lib/python3.1
0/site-packages (from beautifulsoup4>=4.11.1->yfinance) (2.4)
Requirement already satisfied: cffi>=1.12 in /opt/conda/envs/Python-RT23.1/lib/python3.10/s
ite-packages (from cryptography>=3.3.2->yfinance) (1.15.1)
Requirement already satisfied: six>=1.9 in /opt/conda/envs/Python-RT23.1/lib/python3.10/sit
e-packages (from html5lib>=1.1->yfinance) (1.16.0)
Requirement already satisfied: webencodings in /opt/conda/envs/Python-RT23.1/lib/python3.10
/site-packages (from html5lib>=1.1->yfinance) (0.5.1)
Requirement already satisfied: python-dateutil>=2.8.1 in /opt/conda/envs/Python-RT23.1/lib/
python3.10/site-packages (from pandas>=1.3.0->yfinance) (2.8.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/envs/Python-RT23.1/li
b/python3.10/site-packages (from requests>=2.26->yfinance) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/envs/Python-RT23.1/lib/python3.10
/site-packages (from requests>=2.26->yfinance) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/envs/Python-RT23.1/lib/pyth
on3.10/site-packages (from requests>=2.26->yfinance) (1.26.18)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python-RT23.1/lib/pyth
on3.10/site-packages (from requests>=2.26->yfinance) (2023.11.17)
Requirement already satisfied: pycparser in /opt/conda/envs/Python-RT23.1/lib/python3.10/si
te-packages (from cffi>=1.12->cryptography>=3.3.2->yfinance) (2.21)
```



```
        ▌▌     ▌▌▌▌  ▌▌▌▌     ▌▌▌▌▌▌▌▌ ▌▌   ▌▌

        mamba (1.4.2) supported by @QuantStack

        GitHub:  https://github.com/mamba-org/mamba
        Twitter: https://twitter.com/QuantStack

████████████████████████████████████████████████████████████

Looking for: ['bs4==4.10.0']

[+] 0.0s
opt/ibm/custom-channels/meta-wscloud/linux-64 ━━━━━━━━━━━━━━━━━━━━━━━━   0.0 B  0.0sopt/ibm/cus
tom-channels/meta-wscloud/linux-64       0.0s
opt/ibm/custom-channels/meta-wscloud/noarch         13.3kB @ 122.4MB/s  0.0s
opt/ibm/custom-channels/placebo-20230406-noarch/..  ??.?MB @  ??.?MB/s 0 failed  0.0s
opt/ibm/custom-channels/placebo-20230406-noarch/..   5.9kB @  82.9MB/s  0.0s
[+] 0.1s
pkgs/main/linux-64 ━━━━━━━━━━━━━━━━━━━━━━━━   0.0 B /  ??.?MB @  ??.?MB/s  0.1s
pkgs/main/noarch   ━━━━━━━━━━━━━━━━━━━━━━━━   0.0 B /  ??.?MB @  ??.?MB/s  0.1spkgs/main/l
inux-64                                          No change
pkgs/main/noarch                                           No change

Pinned packages:
  - python 3.10.*
  - python 3.10.*
  - widgetsnbextension 4.0.5.*


Transaction

  Prefix: /opt/conda/envs/Python-RT23.1

  All requested packages already installed

Collecting nbformat==4.2.0
  Using cached nbformat-4.2.0-py2.py3-none-any.whl (153 kB)
Requirement already satisfied: ipython-genutils in /opt/conda/envs/Python-RT23.1/lib/python
3.10/site-packages (from nbformat==4.2.0) (0.2.0)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /opt/conda/envs/Python-RT23.1/lib
/python3.10/site-packages (from nbformat==4.2.0) (4.17.3)
Requirement already satisfied: jupyter-core in /opt/conda/envs/Python-RT23.1/lib/python3.10
/site-packages (from nbformat==4.2.0) (5.3.0)
Requirement already satisfied: traitlets>=4.1 in /opt/conda/envs/Python-RT23.1/lib/python3.
10/site-packages (from nbformat==4.2.0) (5.7.1)
Requirement already satisfied: attrs>=17.4.0 in /opt/conda/envs/Python-RT23.1/lib/python3.1
0/site-packages (from jsonschema!=2.5.0,>=2.4->nbformat==4.2.0) (23.1.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /opt/conda/
envs/Python-RT23.1/lib/python3.10/site-packages (from jsonschema!=2.5.0,>=2.4->nbformat==4.
2.0) (0.18.0)
Requirement already satisfied: platformdirs>=2.5 in /opt/conda/envs/Python-RT23.1/lib/pytho
n3.10/site-packages (from jupyter-core->nbformat==4.2.0) (2.5.2)
Installing collected packages: nbformat
  Attempting uninstall: nbformat
    Found existing installation: nbformat 5.9.2
    Uninstalling nbformat-5.9.2:
      Successfully uninstalled nbformat-5.9.2
```

```
Successfully installed nbformat-4.2.0
Requirement already satisfied: nbformat in /opt/conda/envs/Python-RT23.1/lib/python3.10/sit
e-packages (4.2.0)
Collecting nbformat
  Using cached nbformat-5.9.2-py3-none-any.whl.metadata (3.4 kB)
Requirement already satisfied: fastjsonschema in /opt/conda/envs/Python-RT23.1/lib/python3.
10/site-packages (from nbformat) (2.16.2)
Requirement already satisfied: jsonschema>=2.6 in /opt/conda/envs/Python-RT23.1/lib/python
3.10/site-packages (from nbformat) (4.17.3)
Requirement already satisfied: jupyter-core in /opt/conda/envs/Python-RT23.1/lib/python3.10
/site-packages (from nbformat) (5.3.0)
Requirement already satisfied: traitlets>=5.1 in /opt/conda/envs/Python-RT23.1/lib/python3.
10/site-packages (from nbformat) (5.7.1)
Requirement already satisfied: attrs>=17.4.0 in /opt/conda/envs/Python-RT23.1/lib/python3.1
0/site-packages (from jsonschema>=2.6->nbformat) (23.1.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /opt/conda/
envs/Python-RT23.1/lib/python3.10/site-packages (from jsonschema>=2.6->nbformat) (0.18.0)
Requirement already satisfied: platformdirs>=2.5 in /opt/conda/envs/Python-RT23.1/lib/pytho
n3.10/site-packages (from jupyter-core->nbformat) (2.5.2)
Using cached nbformat-5.9.2-py3-none-any.whl (77 kB)
Installing collected packages: nbformat
  Attempting uninstall: nbformat
    Found existing installation: nbformat 4.2.0
    Uninstalling nbformat-4.2.0:
      Successfully uninstalled nbformat-4.2.0
Successfully installed nbformat-5.9.2
```

In [2]:
```
import yfinance as yf
import pandas as pd
import requests
from bs4 import BeautifulSoup
import plotly.graph_objects as go
from plotly.subplots import make_subplots
```

Define Graphing Function

In this section, we define the function `make_graph`. You don't have to know how the function works, you should only care about the inputs. It takes a dataframe with stock data (dataframe must contain Date and Close columns), a dataframe with revenue data (dataframe must contain Date and Revenue columns), and the name of the stock.

In [3]:
```
def make_graph(stock_data, revenue_data, stock):
    fig = make_subplots(rows=2, cols=1, shared_xaxes=True, subplot_titles=("Historical Shar
    stock_data_specific = stock_data[stock_data.Date <= '2021--06-14']
    revenue_data_specific = revenue_data[revenue_data.Date <= '2021-04-30']
    fig.add_trace(go.Scatter(x=pd.to_datetime(stock_data_specific.Date, infer_datetime_form
    fig.add_trace(go.Scatter(x=pd.to_datetime(revenue_data_specific.Date, infer_datetime_fo
    fig.update_xaxes(title_text="Date", row=1, col=1)
    fig.update_xaxes(title_text="Date", row=2, col=1)
    fig.update_yaxes(title_text="Price ($US)", row=1, col=1)
    fig.update_yaxes(title_text="Revenue ($US Millions)", row=2, col=1)
    fig.update_layout(showlegend=False,
    height=900,
    title=stock,
    xaxis_rangeslider_visible=True)
    fig.show()
```

Question 1: Use yfinance to Extract Stock Data

Using the `Ticker` function enter the ticker symbol of the stock we want to extract data on to create a ticker object. The stock is Tesla and its ticker symbol is `TSLA`.

In [4]:
```
tesla=yf.Ticker("TSLA")
```

Using the ticker object and the function `history` extract stock information and save it in a dataframe named `tesla_data`. Set the `period` parameter to `max` so we get information for the maximum amount of time.

In [5]:
```
tesla_data=tesla.history(period="max")
```

Reset the index using the `reset_index(inplace=True)` function on the tesla_data DataFrame and display the first five rows of the `tesla_data` dataframe using the `head` function. Take a screenshot of the results and code from the beginning of Question 1 to the results below.

In [6]:
```
tesla_data.reset_index(inplace=True)
tesla_data.head(5)
```

Out[6]:

	Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
0	2010-06-29 00:00:00-04:00	1.266667	1.666667	1.169333	1.592667	281494500	0.0	0.0
1	2010-06-30 00:00:00-04:00	1.719333	2.028000	1.553333	1.588667	257806500	0.0	0.0
2	2010-07-01 00:00:00-04:00	1.666667	1.728000	1.351333	1.464000	123282000	0.0	0.0
3	2010-07-02 00:00:00-04:00	1.533333	1.540000	1.247333	1.280000	77097000	0.0	0.0
4	2010-07-06 00:00:00-04:00	1.333333	1.333333	1.055333	1.074000	103003500	0.0	0.0

Question 2: Use Webscraping to Extract Tesla Revenue Data

Use the `requests` library to download the webpage https://www.macrotrends.net/stocks/charts/TSLA/tesla/revenue. Save the text of the response as a variable named `html_data`.

In [7]:
```
url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkill
html_data = requests.get(url).text
```

Parse the html data using `beautiful_soup`.

In [8]:
```
soup = BeautifulSoup(html_data)
```

Using `BeautifulSoup` or the `read_html` function extract the table with `Tesla Quarterly Revenue` and store it into a dataframe named `tesla_revenue`. The dataframe should have columns `Date` and `Revenue`.

▸ Click here if you need help locating the table

In [9]:
```
data = []
for table in soup.find_all("table"):

    if any(["Tesla Quarterly Revenue".lower() in th.text.lower() for th in table.find_all("
        for row in table.find("tbody").find_all("tr"):
            date_col, rev_col = [col for col in row.find_all("td")]
            data.append({
                "Date": date_col.text,
                "Revenue": rev_col.text
            })

tesla_revenue = pd.DataFrame(data)
```

Execute the following line to remove the comma and dollar sign from the `Revenue` column.

In [10]:
```
tesla_revenue["Revenue"] = tesla_revenue['Revenue'].str.replace(',|\$',"")
```

```
/tmp/wsuser/ipykernel_1114/349343550.py:1: FutureWarning: The default value of regex will c
hange from True to False in a future version.
  tesla_revenue["Revenue"] = tesla_revenue['Revenue'].str.replace(',|\$',"")
```

Execute the following lines to remove an null or empty strings in the Revenue column.

In [11]:
```
tesla_revenue.dropna(inplace=True)

tesla_revenue = tesla_revenue[tesla_revenue['Revenue'] != ""]
```

Display the last 5 row of the `tesla_revenue` dataframe using the `tail` function. Take a screenshot of the results.

In [12]:
```
tesla_revenue.tail(5)
```

Out[12]:

	Date	Revenue
48	2010-09-30	31
49	2010-06-30	28
50	2010-03-31	21
52	2009-09-30	46
53	2009-06-30	27

Question 3: Use yfinance to Extract Stock Data

Using the `Ticker` function enter the ticker symbol of the stock we want to extract data on to create a ticker object. The stock is GameStop and its ticker symbol is `GME`.

In [13]:
```
gme = yf.Ticker("GME")
```

Using the ticker object and the function `history` extract stock information and save it in a dataframe named `gme_data`. Set the `period` parameter to `max` so we get information for the maximum amount of time.

In [14]:
```
gme_data = gme.history(period="max")
```

Reset the index using the `reset_index(inplace=True)` function on the gme_data DataFrame and display the first five rows of the `gme_data` dataframe using the `head` function. Take a screenshot of the results and code from the beginning of Question 3 to the results below.

In [15]:
```
gme_data.reset_index(inplace=True)
gme_data.head()
```

Out[15]:

	Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
0	2002-02-13 00:00:00-05:00	1.620129	1.693350	1.603296	1.691667	76216000	0.0	0.0
1	2002-02-14 00:00:00-05:00	1.712707	1.716074	1.670626	1.683251	11021600	0.0	0.0
2	2002-02-15 00:00:00-05:00	1.683250	1.687458	1.658002	1.674834	8389600	0.0	0.0
3	2002-02-19 00:00:00-05:00	1.666417	1.666417	1.578047	1.607504	7410400	0.0	0.0
4	2002-02-20 00:00:00-05:00	1.615921	1.662210	1.603296	1.662210	6892800	0.0	0.0

Question 4: Use Webscraping to Extract GME Revenue Data

Use the `requests` library to download the webpage https://www.macrotrends.net/stocks/charts/GME/gamestop/revenue. Save the text of the response as a variable named `html_data`.

In [16]:
```
url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkill
#url = "https://www.macrotrends.net/stocks/charts/GME/gamestop/revenue"
html_data = requests.get(url).text
```

Parse the html data using `beautiful_soup` .

In [17]:
```
soup = BeautifulSoup(html_data)
```

Using `BeautifulSoup` or the `read_html` function extract the table with `GameStop Quarterly Revenue` and store it into a dataframe named `gme_revenue` . The dataframe should have columns `Date` and `Revenue` . Make sure the comma and dollar sign is removed from the `Revenue` column using a method similar to what you did in Question 2.

▸ Click here if you need help locating the table

In [18]:
```
data = []
for table in soup.find_all("table"):

    if any(["GameStop Quarterly Revenue".lower() in th.text.lower() for th in table.find_al
        for row in table.find("tbody").find_all("tr"):
            date_col, rev_col = [col for col in row.find_all("td")]
            data.append({
                "Date": date_col.text,
                "Revenue": rev_col.text.replace("$", "").replace(",", "")
            })

gme_revenue = pd.DataFrame(data)
```

Display the last five rows of the `gme_revenue` dataframe using the `tail` function. Take a screenshot of the results.

In [19]:
```
gme_revenue.tail()
```

Out[19]:

	Date	Revenue
57	2006-01-31	1667
58	2005-10-31	534
59	2005-07-31	416
60	2005-04-30	475
61	2005-01-31	709

Question 5: Plot Tesla Stock Graph

Use the `make_graph` function to graph the Tesla Stock Data, also provide a title for the graph. The structure to call the `make_graph` function is `make_graph(tesla_data, tesla_revenue, 'Tesla')` . Note the graph will only show data upto June 2021.

In [ ]:
```
make_graph(tesla_data, tesla_revenue, 'Tesla')
```

Question 6: Plot GameStop Stock Graph

Use the `make_graph` function to graph the GameStop Stock Data, also provide a title for the graph. The structure to call the `make_graph` function is `make_graph(gme_data, gme_revenue, 'GameStop')` . Note the graph will only show data upto June 2021.

In [ ]:
```
make_graph(gme_data, gme_revenue, 'GameStop')
```

About the Authors:

Joseph Santarcangelo has a PhD in Electrical Engineering, his research focused on using machine learning, signal processing, and computer vision to determine how videos impact human cognition. Joseph has been working for IBM since he completed his PhD.

Azim Hirjani

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-11-10	1.1	Malika Singla	Deleted the Optional part
2020-08-27	1.0	Malika Singla	Added lab to GitLab
