Extracting and Visualizing Stock Data

Description

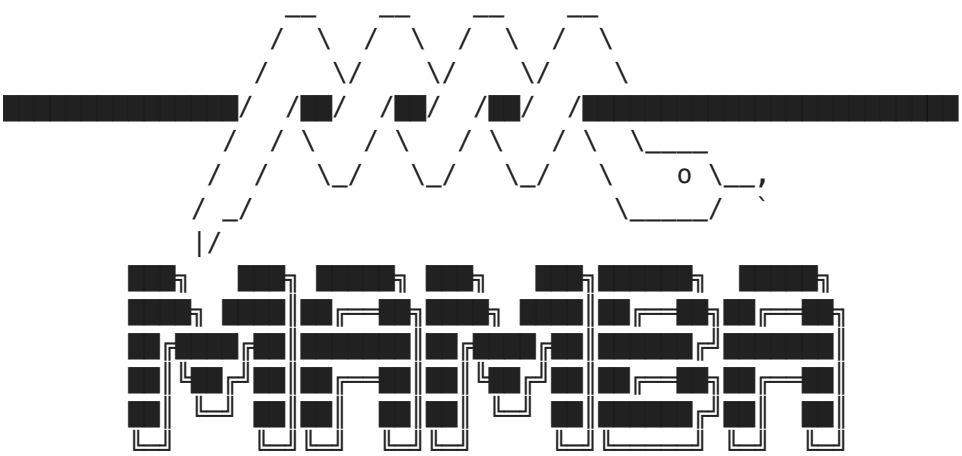
Extracting essential data from a dataset and displaying it is a necessary part of data science; therefore individuals can make correct decisions based on the data. In this assignment, you will extract some stock data, you will then display this data in a graph.

Table of Contents

Estimated Time Needed: **30 min**

In [35]:
```
!pip install yfinance==0.1.67
#!pip install pandas==1.3.3
#!pip install requests==2.26.0
!mamba install bs4==4.10.0 -y
#!pip install plotly==5.3.1
!pip install html5lib
!pip install --upgrade nbformat
```

```
Requirement already satisfied: yfinance==0.1.67 in /home/jupyterlab/conda/
envs/python/lib/python3.7/site-packages (0.1.67)
Requirement already satisfied: pandas>=0.24 in /home/jupyterlab/conda/env
s/python/lib/python3.7/site-packages (from yfinance==0.1.67) (1.3.5)
Requirement already satisfied: numpy>=1.15 in /home/jupyterlab/conda/envs/
python/lib/python3.7/site-packages (from yfinance==0.1.67) (1.21.6)
Requirement already satisfied: requests>=2.20 in /home/jupyterlab/conda/en
vs/python/lib/python3.7/site-packages (from yfinance==0.1.67) (2.29.0)
Requirement already satisfied: multitasking>=0.0.7 in /home/jupyterlab/con
da/envs/python/lib/python3.7/site-packages (from yfinance==0.1.67) (0.0.1
1)
Requirement already satisfied: lxml>=4.5.1 in /home/jupyterlab/conda/envs/
python/lib/python3.7/site-packages (from yfinance==0.1.67) (4.6.4)
Requirement already satisfied: python-dateutil>=2.7.3 in /home/jupyterlab/
conda/envs/python/lib/python3.7/site-packages (from pandas>=0.24->yfinance
==0.1.67) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in /home/jupyterlab/conda/env
s/python/lib/python3.7/site-packages (from pandas>=0.24->yfinance==0.1.67)
(2023.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /home/jupyterla
b/conda/envs/python/lib/python3.7/site-packages (from requests>=2.20->yfin
ance==0.1.67) (3.1.0)
Requirement already satisfied: idna<4,>=2.5 in /home/jupyterlab/conda/env
s/python/lib/python3.7/site-packages (from requests>=2.20->yfinance==0.1.6
7) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /home/jupyterlab/c
onda/envs/python/lib/python3.7/site-packages (from requests>=2.20->yfinanc
e==0.1.67) (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in /home/jupyterlab/cond
a/envs/python/lib/python3.7/site-packages (from requests>=2.20->yfinance==
0.1.67) (2023.5.7)
Requirement already satisfied: six>=1.5 in /home/jupyterlab/conda/envs/pyth
on/lib/python3.7/site-packages (from python-dateutil>=2.7.3->pandas>=0.24
->yfinance==0.1.67) (1.16.0)
```



```
        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
pkgs/main/linux-64 ──────────────────────   0.0 B /  ??.?MB @  ??.?MB/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
0.1s
pkgs/r/linux-64    ──────────────────────   0.0 B /  ??.?MB @  ??.?MB/s
0.1s
pkgs/r/noarch      ──────────────────────   0.0 B /  ??.?MB @  ??.?MB/s
0.1spkgs/main/noarch                                          No chang
e
pkgs/main/linux-64                                         No change
pkgs/r/noarch                                              No change
pkgs/r/linux-64                                            No change

Pinned packages:
  - python 3.7.*


Transaction

  Prefix: /home/jupyterlab/conda/envs/python

  All requested packages already installed

Requirement already satisfied: html5lib in /home/jupyterlab/conda/envs/pyt
hon/lib/python3.7/site-packages (1.1)
Requirement already satisfied: six>=1.9 in /home/jupyterlab/conda/envs/pyt
hon/lib/python3.7/site-packages (from html5lib) (1.16.0)
Requirement already satisfied: webencodings in /home/jupyterlab/conda/env
s/python/lib/python3.7/site-packages (from html5lib) (0.5.1)
Requirement already satisfied: nbformat in /home/jupyterlab/conda/envs/pyt
hon/lib/python3.7/site-packages (4.2.0)
Collecting nbformat
  Using cached nbformat-5.8.0-py3-none-any.whl (77 kB)
Requirement already satisfied: fastjsonschema in /home/jupyterlab/conda/en
vs/python/lib/python3.7/site-packages (from nbformat) (2.16.3)
Requirement already satisfied: importlib-metadata>=3.6 in /home/jupyterla
b/conda/envs/python/lib/python3.7/site-packages (from nbformat) (4.11.4)
Requirement already satisfied: jsonschema>=2.6 in /home/jupyterlab/conda/e
nvs/python/lib/python3.7/site-packages (from nbformat) (4.17.3)
Requirement already satisfied: jupyter-core in /home/jupyterlab/conda/env
s/python/lib/python3.7/site-packages (from nbformat) (4.12.0)
Requirement already satisfied: traitlets>=5.1 in /home/jupyterlab/conda/en
vs/python/lib/python3.7/site-packages (from nbformat) (5.9.0)
Requirement already satisfied: zipp>=0.5 in /home/jupyterlab/conda/envs/py
thon/lib/python3.7/site-packages (from importlib-metadata>=3.6->nbformat)
(3.15.0)
Requirement already satisfied: typing-extensions>=3.6.4 in /home/jupyterla
b/conda/envs/python/lib/python3.7/site-packages (from importlib-metadata>=
3.6->nbformat) (4.5.0)
Requirement already satisfied: attrs>=17.4.0 in /home/jupyterlab/conda/env
s/python/lib/python3.7/site-packages (from jsonschema>=2.6->nbformat) (23.
1.0)
Requirement already satisfied: importlib-resources>=1.4.0 in /home/jupyter
lab/conda/envs/python/lib/python3.7/site-packages (from jsonschema>=2.6->n
bformat) (5.12.0)
Requirement already satisfied: pkgutil-resolve-name>=1.3.10 in /home/jupyt
erlab/conda/envs/python/lib/python3.7/site-packages (from jsonschema>=2.6-
```

```
>nbformat) (1.3.10)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.1
4.0 in /home/jupyterlab/conda/envs/python/lib/python3.7/site-packages (fro
m jsonschema>=2.6->nbformat) (0.19.3)
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-5.8.0
```

In [36]:
```
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 [37]:
```
def make_graph(stock_data, revenue_data, stock):
    fig = make_subplots(rows=2, cols=1, shared_xaxes=True, subplot_titles
    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, i
    fig.add_trace(go.Scatter(x=pd.to_datetime(revenue_data_specific.Date,
    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 [38]:
```
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 [39]:
```
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 [40]:
```
tesla_data.reset_index(inplace=True)
tesla_data.head()
```

Out[40]:

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

Question 2: Use Webscraping to Extract Tesla Revenue Data

Use the `requests` library to download the webpage https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-PY0220EN-SkillsNetwork/labs/project/revenue.htm Save the text of the response as a variable named `html_data`.

In [41]:
```
url = "https://www.macrotrends.net/stocks/charts/TSLA/tesla/revenue"
html_data  = requests.get(url).text
```

Parse the html data using `beautiful_soup`.

In [42]:
```
soup = BeautifulSoup(html_data,"html5lib")
```

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 [43]:
```
tesla_revenue = pd.DataFrame(columns=['Date', 'Revenue'])

for table in soup.find_all('table'):

    if ('Tesla Quarterly Revenue' in table.find('th').text):
        rows = table.find_all('tr')

        for row in rows:
            col = row.find_all('td')

            if col != []:
                date = col[0].text
                revenue = col[1].text.replace(',','').replace('$','')

                tesla_revenue = tesla_revenue.append({"Date":date, "Reven
```

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

In [44]:
```
tesla_revenue["Revenue"] = tesla_revenue['Revenue'].str.replace(',|\$',""
```

```
/home/jupyterlab/conda/envs/python/lib/python3.7/site-packages/ipykernel_l
auncher.py:1: FutureWarning: The default value of regex will change from T
rue to False in a future version.
  """Entry point for launching an IPython kernel.
```

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

In [45]:
```
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 [46]:
```
tesla_revenue.tail()
```

Out[46]:

	Date	Revenue
50	2010-09-30	31
51	2010-06-30	28
52	2010-03-31	21
54	2009-09-30	46
55	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 [47]:
```
GameStop = 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 [49]:
```
gme_data = GameStop.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 [50]:
```
gme_data.reset_index(inplace=True)
gme_data.head()
```

Out[50]:

	Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
0	2002-02-13	1.620128	1.693350	1.603296	1.691666	76216000	0.0	0.0
1	2002-02-14	1.712707	1.716073	1.670626	1.683250	11021600	0.0	0.0
2	2002-02-15	1.683251	1.687459	1.658002	1.674834	8389600	0.0	0.0
3	2002-02-19	1.666418	1.666418	1.578047	1.607504	7410400	0.0	0.0
4	2002-02-20	1.615920	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://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-PY0220EN-SkillsNetwork/labs/project/stock.html. Save the text of the response as a variable named `html_data`.

In [51]:
```
url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud

html_data  = requests.get(url).text
```

Parse the html data using `beautiful_soup`.

In [52]:
```
soup = BeautifulSoup(html_data,"html5lib")
```

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 [53]:
```
gme_revenue = pd.DataFrame(columns=['Date', 'Revenue'])

for table in soup.find_all('table'):

    if ('GameStop Quarterly Revenue' in table.find('th').text):
        rows = table.find_all('tr')

        for row in rows:
            col = row.find_all('td')

            if col != []:
                date = col[0].text
                revenue = col[1].text.replace(',','').replace('$','')

                gme_revenue = gme_revenue.append({"Date":date, "Revenue":
```

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

In [54]:
```

```

Out[54]:

	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

In [58]:
```
def make_graph(gme_revenue):
    fig = make_subplots(rows=2, cols=1, shared_xaxes=True, subplot_titles
    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, i
    fig.add_trace(go.Scatter(x=pd.to_datetime(revenue_data_specific.Date,
    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()
```

In [59]:
```
def make_graph(tesla_revenue):
    fig = make_subplots(rows=2, cols=1, shared_xaxes=True, subplot_titles
```

```
    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, i
    fig.add_trace(go.Scatter(x=pd.to_datetime(revenue_data_specific.Date,
    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 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.

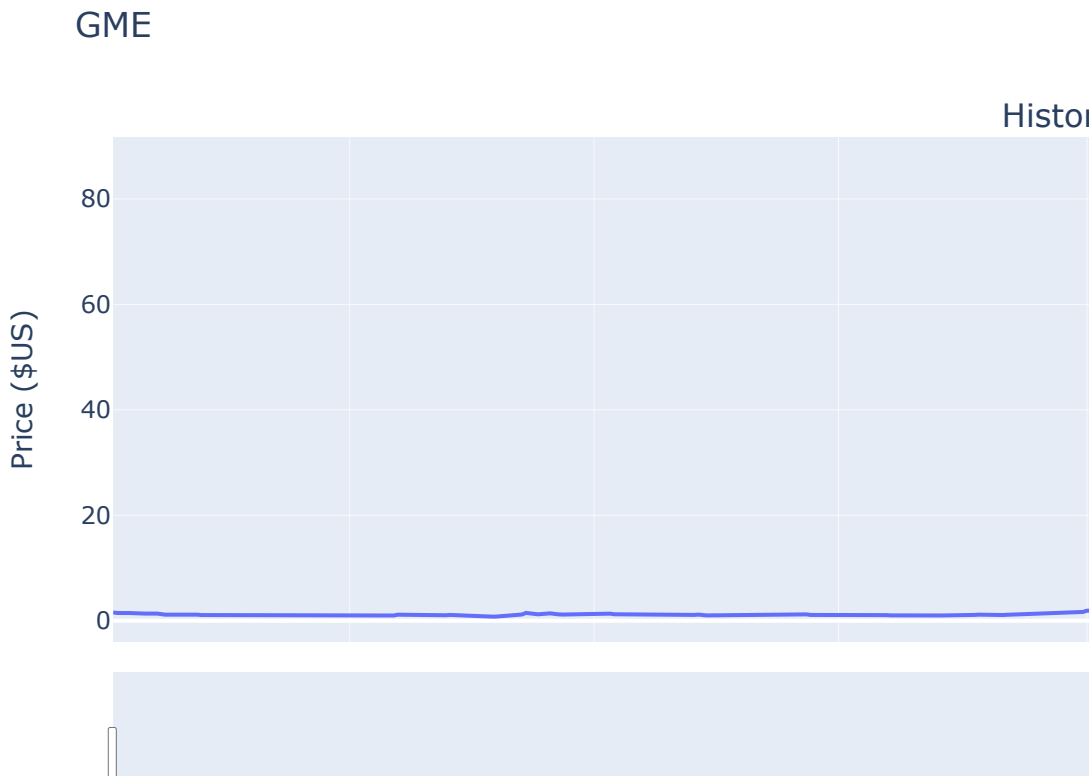
In [69]:
```
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import yfinance as yf
import pandas as pd

stock_data = yf.download("TSLA", start="2020-01-01", end="2021-09-30", pr
revenue_data = yf.download("TSLA", start="2020-01-01", end="2021-09-30",
stock_data.reset_index(inplace=True)
revenue_data.reset_index(inplace=True)

def make_graph(stock_data, revenue_data, stock):
    fig = make_subplots(rows=2, cols=1,
                        shared_xaxes=True,
                        subplot_titles=("Historical Share Price", "Histor
                        vertical_spacing=.3)

    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_format=
        y=stock_data_specific.Close.astype("float"), name="Share Price"),

    fig.add_trace(go.Scatter(x=pd.to_datetime(revenue_data_specific.Date,
                             y=revenue_data_specific.Volume.astype("float
                             name="Volume"), row=2, col=1)

    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()

make_graph(stock_data, revenue_data, 'TSLA')
```

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 [71]:
```
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import yfinance as yf
import pandas as pd

stock_data = yf.download("GME", start="2020-01-01", end="2021-09-30", pro
revenue_data = yf.download("GME", start="2020-01-01", end="2021-09-30", p
stock_data.reset_index(inplace=True)
revenue_data.reset_index(inplace=True)

def make_graph(stock_data, revenue_data, stock):
    fig = make_subplots(rows=2, cols=1,
                        shared_xaxes=True,
                        subplot_titles=("Historical Share Price", "Histor
                        vertical_spacing=.3)

    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_format=
        y=stock_data_specific.Close.astype("float"), name="Share Price"),

    fig.add_trace(go.Scatter(x=pd.to_datetime(revenue_data_specific.Date,
                             y=revenue_data_specific.Volume.astype("float
                             name="Volume"), row=2, col=1)

    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()

make_graph(stock_data, revenue_data, 'GME')
```



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
2022-02-28	1.2	Lakshmi Holla	Changed the URL of GameStop
2020-11-10	1.1	Malika Singla	Deleted the Optional part
2020-08-27	1.0	Malika Singla	Added lab to GitLab
