Financial Data Science - Assignment 1: 14^{th} February, 2023

Author: Gavin Connolly
Student Number: 18308483

Module: Financial Data Science

Academic Year - 2022/23
Submitted: 14^{th} February, 2023

Part 1: Yahoo Finance

In the following code, we create a candlestick plot showing the performance of the Boeing Company (Ticker: BA) over the past year.

We first read the assigned ticker from the classlist csv file corresponding to my student number. We then use ticker value to search for the previous years' performance. Once we have loaded in our data, we can then proceed to create the candlestick plot.

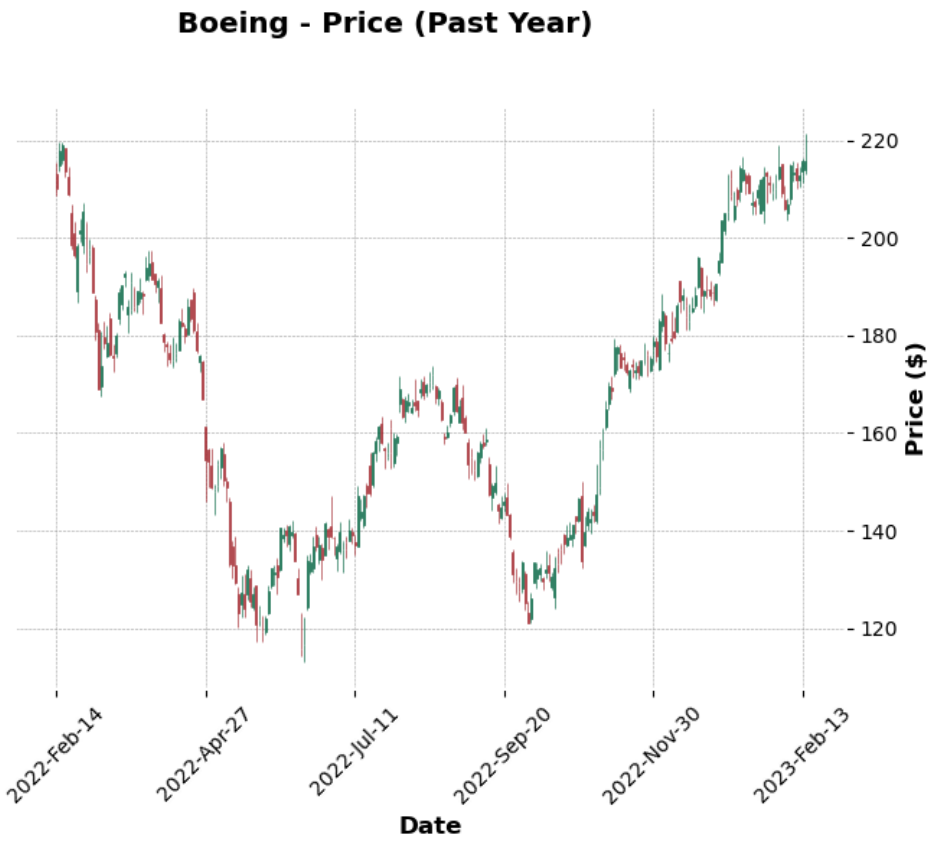

Figure 1 Candlestick plot of Boeing's stock performance over the past year.

```
import pandas as pd
import yfinance as yf
import mplfinance as mpf

my_student_number = 18308483

class_list = pd.read_csv("ClassList1.csv", sep=",")
ticker_name = class_list.loc[class_list["Student ID"] == my_student_number, "Assigned Ticker"].item()

yticker = yf.Ticker(ticker_name)
BA = yticker.history(period="1y")

mpf.plot(BA, type='candle', style = 'charles',
    title='Boeing - Price (Past Year)', xlabel = 'Date', ylabel='Price ($)')
```

Part 2: pdf Scraping

In this question we are looking to read the data from a table in pdf format as a python dataframe using tabula. We then output this table to Latex.

Country	Ranking	Economy	GDP
USA	1	United States	20,936,600
CHN	2	China	14,722,731
JPN	3	Japan	5,064,873
DEU	4	Germany	3,806,060
GBR	5	United Kingdom	2,707,744
IND	6	India	2,622,984
FRA	7	France	2,603,004
ITA	8	Italy	1,886,445
CAN	9	Canada	1,643,408
KOR	10	Korea, Rep.	1,630,525
RUS	11	Russian Federation	1,483,498
BRA	12	Brazil	1,444,733
AUS	13	Australia	1,330,901
ESP	14	Spain	1,281,199
MEX	15	Mexico	1,076,163
IDN	16	Indonesia	1,058,424
NLD	17	Netherlands	912,242
CHE	18	Switzerland	747,969
TUR	19	Turkey	720,101
SAU	20	Saudi Arabia	700,118
POL	21	Poland	594,165
SWE	22	Sweden	537,610
BEL	23	Belgium	515,332
THA	24	Thailand	501,795
NGA	25	Nigeria	432,294
AUT	26	Austria	428,965
ARE	27	United Arab Emirates	421,142
IRL	28	Ireland	418,622
ISR	29	Israel	401,954
ARG	30	Argentina	383,067
EGY	31	Egypt, Arab Rep.	363,069
NOR	32	Norway	362,009
PHL	33	Philippines	361,489
DNK	34	Denmark	355,184
HKG	35	Hong Kong SAR, China	346,586
SGP	36	Singapore	339,998
MYS	37	Malaysia	336,664
BGD	38	Bangladesh	324,239
ZAF	39	South Africa	301,924
COL	40	Colombia	271,347
FIN	41	Finland	271,234
VNM	42	Vietnam	271,158
PAK	43	Pakistan	263,687
CHL	44	Chile	252,940
ROU	45	Romania	248,716
CZE	46	Czech Republic	243,530
PRT	47	Portugal	231,256
NZL	48	New Zealand	212,482
PER	49	Peru	202,014
IRN	50	Iran, Islamic Rep.	191,718
GRC	51	Greece	189,410
KAZ	52	Kazakhstan	169,835
IRQ	53	Iraq	167,224
UKR	54	Ukraine	155,582
HUN	55	Hungary	155,013
QAT	56	Qatar	146,374
DZA	57	Algeria	145,164
KWT	58	Kuwait	136,197
MAR	59	Morocco	112,871
ETH	60	Ethiopia	107,645
SVK	61	Slovak Republic	104,574
PRI	62	Puerto Rico	103,138
CUB	63	Cuba	103,131
KEN	64	Kenya	98,843
ECU	65	Ecuador	98,808

```python
import tabula as tb
import pandas as pd

gdp_lst = tb.read_pdf("GDP11.pdf", pages='1', pandas_options={'header': None})
gdp_df = pd.DataFrame(gdp_lst[0])
gdp_df.drop(columns=gdp_df.columns[-1], axis=1, inplace=True)
gdp_df.columns = ["Country", "Ranking", "Economy", "GDP"]

print(gdp_df.to_latex(index = False))
```

Part 3: Crypto Punk Web Scraping

We begin by reading the class number associated with my student number, which gives us a value of 4. We then make a request from the corresponding cryptopunks URL to obtain the HTML file defining the structure of the webpage.

This HTML file is then read using the BeautifulSoup package, which then allows us to begin parsing the file for the attributes associated with the cryptopunk. The attributes are all stored within containers of the class 'col-md-4', meaning we can locate these attributes in the HTML files by searching through our BeatifulSoup object for divs of this class.

Once we have all the divs of class 'col-md-4', we then iterate through each of the elements of the result set, searching for the hyperlink tags containing the cryptopunk's attributes, and store the resulting descriptors in a list.

Running the below script, we get that the attributes associated with our assigned cryptopunk are: 'Big Shades', 'Wild Hair', 'Earring' & 'Goat'

```
import pandas as pd
import requests
from bs4 import BeautifulSoup

my_student_number = 18308483
BaseStr = "https://www.larvalabs.com/cryptopunks/details/"

class_list = pd.read_csv("ClassList1.csv", sep=",")
class_number = class_list.loc[class_list["Student ID"] == my_student_number, "Class Number"].item()

page = requests.get(BaseStr + str(class_number))
soup = BeautifulSoup(page.content, 'html.parser')

result_set = soup.find_all("div", class_="col-md-4")
attributes = []

for link in result_set:
    attributes.extend(link.find('a'))

attributes = attributes[1:]
print(attributes)
```