Final_Assignment_Webscraping

June 16, 2021

Extracting Stock Data Using a Web Scraping

Not all stock data is available via API in this assignment; you will use web-scraping to obtain financial data. You will be quizzed on your results.
Using beautiful soup we will extract historical share data from a web-page.

Table of Contents

```
<ul>
    <li>Downloading the Webpage Using Requests Library</li>
    <li>Parsing Webpage HTML Using BeautifulSoup</li>
    <li>Extracting Data and Building DataFrame</li>
</ul>
```

Estimated Time Needed: 30 min

[ ]:
```
#!pip install pandas
#!pip install requests
!pip install bs4
#!pip install plotly
```

[ ]:
```
import pandas as pd
import requests
from bs4 import BeautifulSoup
```

0.1 Using Webscraping to Extract Stock Data

Use the `requests` library to download the webpage https://finance.yahoo.com/quote/AMZN/history?period1=145
Save the text of the response as a variable named `html_data`.

[ ]:
```
r = requests.get('https://finance.yahoo.com/quote/AMZN/history?
↪period1=1451606400\&period2=1612137600\&interval=1mo\&filter=history\&frequency=1mo\&includ
↪allow_redirects=True).text
```

Parse the html data using `beautiful_soup`.

[ ]:
```
soup = BeautifulSoup(r, 'html.parser')
```

Question 1 what is the content of the title attribute:

[ ]:
```
soup.head.title
```

Using beautiful soup extract the table with historical share prices and store it into a dataframe named `amazon_data`. The dataframe should have columns Date, Open, High, Low, Close, Adj Close, and Volume. Fill in each variable with the correct data from the list `col`.

Hint: Print the `col` list to see what data to use

[ ]:
```
amazon_data = pd.DataFrame(columns=["Date", "Open", "High", "Low", "Close",␣
↪"Volume"])

for row in soup.find("tbody").find_all("tr"):
    col = row.find_all("td")
    date =col[0].text
    Open = col[1].text
    high = col[2].text
    low = col[3].text
    close = col[4].text
    adj_close = col[5].text
    volume = col[6].text

    amazon_data = amazon_data.append({"Date":date, "Open":Open, "High":high,␣
↪"Low":low, "Close":close, "Adj Close":adj_close, "Volume":volume},␣
↪ignore_index=True)
```

Print out the first five rows of the `amazon_data` dataframe you created.

[ ]:
```
amazon_data.head(10)
```

Question 2 What is the name of the columns of the dataframe

[ ]:
```
Date      Open      High       Low      Close     Volume Adj Close
```

Question 3 What is the `Open` of `Jun 01, 2019` of the dataframe?

[ ]:
```
open_value = amazon_data["Open"].where(amazon_data["Date"] == 'Jun 16, 2021')
```

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

0.2 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

##