Problem Solving Using Python and R Lab

Name: ANNAPOORNIMA S

Roll No.: 225229101

Question1. Retrieve data from web page using URLLIB and print the frequency of words from that page.

In [48]:
```
# import urllib.request
counts=dict()
web=urllib.request.urlopen('https://en.wikipedia.org/wiki/Periyar')
for line in web:
    words=line.decode().split()
    for word in words:
        counts[word]=counts.get(word,0)+1
for i in counts.items():
    print(i)
```

```
('<!DOCTYPE', 1)
('html>', 1)
('<html', 1)
('class="client-nojs"', 1)
('lang="en"', 3)
('dir="ltr">', 1)
('<head>', 1)
('<meta', 18)
('charset="UTF-8"/>', 1)
('<title>Periyar', 1)
('-', 7)
('Wikipedia</title>', 1)
('<script>document.documentElement.className="client-js";RLCONF={"wgBreakFram
es":false,"wgSeparatorTransformTable":["",""],"wgDigitTransformTable":
["",""],"wgDefaultDateFormat":"dmy","wgMonthNames":["","January","Februar
y","March","April","May","June","July","August","September","October","Novemb
er","December"],"wgRequestId":"9b81c665-8222-42f5-b9da-4b2fc4274e98","wgCSPNo
nce":false,"wgCanonicalNamespace":"","wgCanonicalSpecialPageName":false,"wgNa
mespaceNumber":0,"wgPageName":"Periyar","wgTitle":"Periyar","wgCurRevisionI
```

Question2. Retrieve and display all hyperlinks (ie., HREF attribute) from a webpage using BeautifulSoup.

In [3]:
```
import urllib.request,urllib.parse,urllib.error
from bs4 import BeautifulSoup
import ssl
ctx=ssl.create_default_context()
ctx.check_hostname=False
ctx.verify_mode=ssl.CERT_NONE
url=('https://en.wikipedia.org/wiki/Periyar')
html=urllib.request.urlopen(url,context=ctx).read()
soup=BeautifulSoup(html,'html.parser')
tags=soup('a')
for tag in tags:
    print(tag.get('href',None))
```

```
None
/wiki/Wikipedia:Good_articles
/wiki/Wikipedia:Protection_policy#semi
#mw-head
#searchInput
/wiki/Periyar_(disambiguation)
/wiki/File:PperiyarEVRStamp.jpg
/wiki/Dravidar_Kazhagam
/wiki/Annai_E._V._R._Maniammai
/wiki/Justice_Party_(India)
/wiki/C._Natesa_Mudaliar
/wiki/Ramakrishna_Ranga_Rao_of_Bobbili
/wiki/P._T._Rajan
/wiki/Erode
/wiki/Coimbatore_District_(Madras_Presidency)
/wiki/Madras_Presidency
/wiki/British_Raj
/wiki/Erode_District
/wiki/Tamil_Nadu
/wiki/Vellore
```

Question3. Create a HTML file for the following Student Marks and print the number of students and their names and marks.

In [1]:
```
from IPython.core.display import HTML
que3='''

<table>
    <tr>
        <td>Id</td>
        <td>Name</td>
        <td>Mark1</td>
        <td>Mark2</td>
        <td>Mark3</td>
    </tr>
    <tr>
        <td>DS01</td>
        <td>rex</td>
        <td>87</td>
        <td>57</td>
        <td>74</td>
    </tr>
    <tr>
        <td>DS02</td>
        <td>peter</td>
        <td>68</td>
        <td>98</td>
        <td>55</td>
    </tr>
</table>

'''

HTML(que3)
```

Out[1]:

Id	Name	Mark1	Mark2	Mark3
DS01	rex	87	57	74
DS02	peter	68	98	55

Question4. Create a JSON file for the following Students Marks and print the number of students and their names and marks.

In [28]:
```
import json
data='''
[
{"id":"DS01","Name":"rex","semester1":"80,55","semester2":"50,70,82"},
{"id":"DS02","Name":"peter","semester1":"92,75","semester2":" "}
]'''
info=json.loads(data)
for item in info:
    print('ID:',item['id'],'\t','Name:',item['Name'])
    print('\t\t','semester1:',item['semester1'])
    print('\t\t','semester2:',item['semester2'])
```

```
ID: DS01        Name: rex
                semester1: 80,55
                semester2: 50,70,82
ID: DS02        Name: peter
                semester1: 92,75
                semester2:
```

Question5. Crawl Weather of a City and Display

In [55]:
```
import requests
from bs4 import BeautifulSoup
page = requests.get('http://www.weather.com')
page.content
bs4 = BeautifulSoup(page.content,'html.parser')
bs4.find_all('p')
(bs4.find_all(class_='DetailSummary--DetailsSummary--QpFD'))
days=[bs4.find_all('h2')[day].get_text() for day in range(len(bs4.find_all('h2'))
tempr=[bs4.find_all(class_="DetailsSummary--temperature--3FMlw")[temp].get_text()
days = days[1:12]
tempr = tempr[1:12]
from IPython.display import display
import pandas as pd
b = {"Days":days,"Temperature":tempr}
weather = pd.DataFrame.from_dict(b, orient = 'index')
weather = weather.transpose()
weather
```

Out[55]:

	Days	Temperature
0	New delhi, DL के लिए आज का पूर्वानुमान	None
1	New delhi, DL में आज का मौसम	None
2	घंटेवार पूर्वानुमान	None
3	दैनिक पूर्वानुमान	None
4	राडार	None
5	वायु गुणवत्ता सूचकांक	None
6	स्वास्थ्य और गतिविधियाँ	None

Question6. Real Time Stock Prices Crawling and Display of a specified Company

In [40]:
```
!pip install yfinance
!pip install pandas.datareader
import pandas_datareader as pdr
import yfinance as yf
yf.pdr_override()
df_info = pdr.get_data_yahoo("INFY", start="2018-01-01").reset_index()
df_info.to_csv('INFY.csv',index=False)
df_info.head()
```

```
Collecting yfinance
  Downloading yfinance-0.1.74-py2.py3-none-any.whl (27 kB)
Requirement already satisfied: pandas>=0.24.0 in c:\users\arulk\anaconda3\lib\s
ite-packages (from yfinance) (1.4.2)
Requirement already satisfied: lxml>=4.5.1 in c:\users\arulk\anaconda3\lib\site
-packages (from yfinance) (4.8.0)
Requirement already satisfied: numpy>=1.15 in c:\users\arulk\anaconda3\lib\site
-packages (from yfinance) (1.21.5)
Collecting multitasking>=0.0.7
  Downloading multitasking-0.0.11-py3-none-any.whl (8.5 kB)
Requirement already satisfied: requests>=2.26 in c:\users\arulk\anaconda3\lib\s
ite-packages (from yfinance) (2.27.1)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\arulk\anacond
a3\lib\site-packages (from pandas>=0.24.0->yfinance) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\arulk\anaconda3\lib\sit
e-packages (from pandas>=0.24.0->yfinance) (2021.3)
Requirement already satisfied: six>=1.5 in c:\users\arulk\anaconda3\lib\site-pa
ckages (from python-dateutil>=2.8.1->pandas>=0.24.0->yfinance) (1.16.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\arulk\anaconda3\l
ib\site-packages (from requests>=2.26->yfinance) (2021.10.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\arulk\anac
onda3\lib\site-packages (from requests>=2.26->yfinance) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\arulk\anaconda3\lib\sit
e-packages (from requests>=2.26->yfinance) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\arulk\anaconda
3\lib\site-packages (from requests>=2.26->yfinance) (1.26.9)
Installing collected packages: multitasking, yfinance
Successfully installed multitasking-0.0.11 yfinance-0.1.74
Collecting pandas.datareader
  Downloading pandas_datareader-0.10.0-py3-none-any.whl (109 kB)
Requirement already satisfied: pandas>=0.23 in c:\users\arulk\anaconda3\lib\sit
e-packages (from pandas.datareader) (1.4.2)
Requirement already satisfied: lxml in c:\users\arulk\anaconda3\lib\site-packag
es (from pandas.datareader) (4.8.0)
Requirement already satisfied: requests>=2.19.0 in c:\users\arulk\anaconda3\lib
\site-packages (from pandas.datareader) (2.27.1)
Requirement already satisfied: pytz>=2020.1 in c:\users\arulk\anaconda3\lib\sit
e-packages (from pandas>=0.23->pandas.datareader) (2021.3)
Requirement already satisfied: numpy>=1.18.5 in c:\users\arulk\anaconda3\lib\si
te-packages (from pandas>=0.23->pandas.datareader) (1.21.5)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\arulk\anacond
a3\lib\site-packages (from pandas>=0.23->pandas.datareader) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\arulk\anaconda3\lib\site-pa
ckages (from python-dateutil>=2.8.1->pandas>=0.23->pandas.datareader) (1.16.0)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\arulk\anac
onda3\lib\site-packages (from requests>=2.19.0->pandas.datareader) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\arulk\anaconda3\l
ib\site-packages (from requests>=2.19.0->pandas.datareader) (2021.10.8)
```

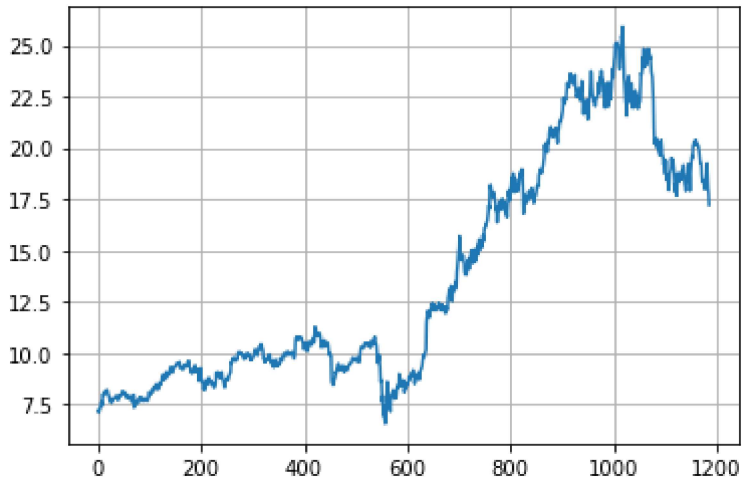
```
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\arulk\anaconda
3\lib\site-packages (from requests>=2.19.0->pandas.datareader) (1.26.9)
Requirement already satisfied: idna<4,>=2.5 in c:\users\arulk\anaconda3\lib\sit
e-packages (from requests>=2.19.0->pandas.datareader) (3.3)
Installing collected packages: pandas.datareader
Successfully installed pandas.datareader-0.10.0
```

Out[40]:

	Date	High	Low	Open	Close	Volume	Adj Close
0	2018-01-02	8.195	8.115	8.135	8.145	12298200.0	7.208184
1	2018-01-03	8.135	8.050	8.120	8.075	10250800.0	7.146235
2	2018-01-04	8.100	8.010	8.100	8.025	16272000.0	7.101985
3	2018-01-05	8.190	8.075	8.085	8.175	9813600.0	7.234734
4	2018-01-08	8.260	8.170	8.190	8.240	11198200.0	7.292257

In [41]:
```
import matplotlib.pyplot as plt
%matplotlib inline
df_info["Adj Close"].plot(grid=True)
```

Out[41]: <AxesSubplot:>



In [42]:
```
!pip install yfinance
!pip install pandas_datareader
import pandas_datareader as pdr
import yfinance as yf
yf.pdr_override()
df_cts = pdr.get_data_yahoo("CTS", start="2018-01-01").reset_index()
df_cts.to_csv('CTS.csv', index=False)
df_cts.head()
```

```
Requirement already satisfied: yfinance in c:\users\arulk\anaconda3\lib\site-pa
ckages (0.1.74)
Requirement already satisfied: numpy>=1.15 in c:\users\arulk\anaconda3\lib\site
-packages (from yfinance) (1.21.5)
Requirement already satisfied: requests>=2.26 in c:\users\arulk\anaconda3\lib\s
ite-packages (from yfinance) (2.27.1)
Requirement already satisfied: pandas>=0.24.0 in c:\users\arulk\anaconda3\lib\s
ite-packages (from yfinance) (1.4.2)
Requirement already satisfied: multitasking>=0.0.7 in c:\users\arulk\anaconda3
\lib\site-packages (from yfinance) (0.0.11)
Requirement already satisfied: lxml>=4.5.1 in c:\users\arulk\anaconda3\lib\site
-packages (from yfinance) (4.8.0)
Requirement already satisfied: pytz>=2020.1 in c:\users\arulk\anaconda3\lib\sit
e-packages (from pandas>=0.24.0->yfinance) (2021.3)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\arulk\anacond
a3\lib\site-packages (from pandas>=0.24.0->yfinance) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\arulk\anaconda3\lib\site-pa
ckages (from python-dateutil>=2.8.1->pandas>=0.24.0->yfinance) (1.16.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\arulk\anaconda
3\lib\site-packages (from requests>=2.26->yfinance) (1.26.9)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\arulk\anac
onda3\lib\site-packages (from requests>=2.26->yfinance) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\arulk\anaconda3\lib\sit
e-packages (from requests>=2.26->yfinance) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\arulk\anaconda3\l
ib\site-packages (from requests>=2.26->yfinance) (2021.10.8)
Requirement already satisfied: pandas_datareader in c:\users\arulk\anaconda3\li
b\site-packages (0.10.0)
Requirement already satisfied: pandas>=0.23 in c:\users\arulk\anaconda3\lib\sit
e-packages (from pandas_datareader) (1.4.2)
Requirement already satisfied: lxml in c:\users\arulk\anaconda3\lib\site-packag
es (from pandas_datareader) (4.8.0)
Requirement already satisfied: requests>=2.19.0 in c:\users\arulk\anaconda3\lib
\site-packages (from pandas_datareader) (2.27.1)
Requirement already satisfied: numpy>=1.18.5 in c:\users\arulk\anaconda3\lib\si
te-packages (from pandas>=0.23->pandas_datareader) (1.21.5)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\arulk\anacond
a3\lib\site-packages (from pandas>=0.23->pandas_datareader) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\arulk\anaconda3\lib\sit
e-packages (from pandas>=0.23->pandas_datareader) (2021.3)
Requirement already satisfied: six>=1.5 in c:\users\arulk\anaconda3\lib\site-pa
ckages (from python-dateutil>=2.8.1->pandas>=0.23->pandas_datareader) (1.16.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\arulk\anaconda
3\lib\site-packages (from requests>=2.19.0->pandas_datareader) (1.26.9)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\arulk\anaconda3\l
ib\site-packages (from requests>=2.19.0->pandas_datareader) (2021.10.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\arulk\anac
onda3\lib\site-packages (from requests>=2.19.0->pandas_datareader) (2.0.4)
```

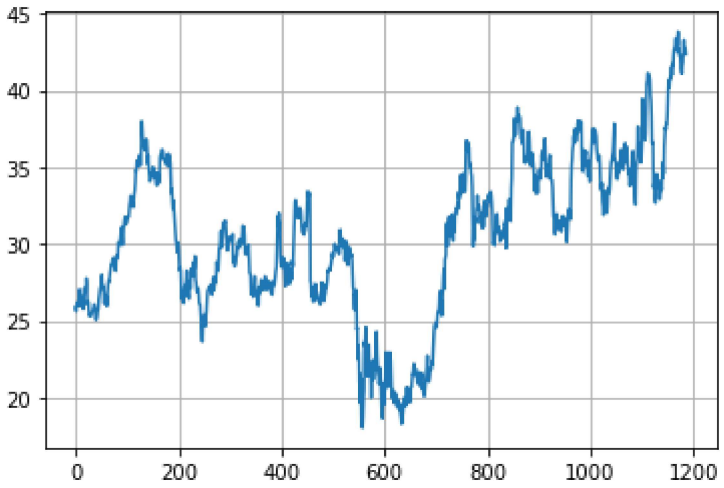
```
Requirement already satisfied: idna<4,>=2.5 in c:\users\arulk\anaconda3\lib\sit
e-packages (from requests>=2.19.0->pandas_datareader) (3.3)
```

Out[42]:

	Date	High	Low	Open	Close	Volume	Adj Close
0	2018-01-02	26.950001	25.850000	25.900000	26.500000	117900.0	25.864653
1	2018-01-03	26.700001	26.200001	26.549999	26.250000	62600.0	25.620646
2	2018-01-04	27.150000	26.450001	26.500000	26.750000	55800.0	26.108658
3	2018-01-05	27.049999	26.600000	26.750000	26.850000	39700.0	26.206263
4	2018-01-08	26.900000	26.549999	26.750000	26.700001	40100.0	26.059858

In [43]:
```
import matplotlib.pyplot as plt
%matplotlib inline
df_cts["Adj Close"].plot(grid=True)
```

Out[43]: <AxesSubplot:>



In [ ]: