

Web Scraping of yahoo Financial stats "<https://uk.finance.yahoo.com/>" by using **pandas** **datereader** and **yahoo\_fin** packages

## ▼ Step 1) Import libraries

```
## Import all the libraries
import pandas as pd
from pandas_datareader import data
```

##Now requires headers, pandas and pandas-datareader must be updated. Other libraries work

```
!pip install --upgrade pandas-datareader
```

```
## Create variables for the sources
```

```
start_date = '2010-01-01'
```

```
end_date = '2020-01-01'
```


```
## Create dataframe
```

```
## name extracted from yahoo finance for Apple
```

```
df = data.DataReader('aapl', data_source='yahoo' , start=start_date, end=end_date )
```

```
## View the data -earliest
```

```
df.head()
```

	High	Low	Open	Close	Volume	Adj Close	
Date							
<b>2010-01-04</b>	7.660714	7.585000	7.622500	7.643214	493729600.0	6.526021	
<b>2010-01-05</b>	7.699643	7.616071	7.664286	7.656429	601904800.0	6.537303	
<b>2010-01-06</b>	7.686786	7.526786	7.656429	7.534643	552160000.0	6.433319	
<b>2010-01-07</b>	7.571429	7.466071	7.562500	7.520714	477131200.0	6.421426	
<b>2010-01-08</b>	7.571429	7.466429	7.510714	7.570714	447610800.0	6.464118	

```
## View the data -latest
```

```
df.tail()
```

	High	Low	Open	Close	Volume	Adj Close
Date						
2019-12-24	71.222504	70.730003	71.172501	71.067497	48478800.0	69.738724
2019-12-26	72.495003	71.175003	71.205002	72.477501	93121200.0	71.122368

## ▼ 2) Yahoo\_Fin open source packages to extract stock info

```
##Use yahoo_fin packages
!pip install yahoo_fin
```

```
## install requests_html packages
! pip install requests_html
```

```
#load library
from yahoo_fin import stock_info as si
```

```
#Pull data from yahoo finance
#get the live price
si.get_live_price('aapl')
```

```
140.67010498046875
```

```
#round the price
round(si.get_live_price('aapl'))
```

```
141
```

```
#round the price with one decimal
round(si.get_live_price('aapl'), 0)
```

```
141.0
```

```
#Extract income statement
si.get_income_statement('aapl', yearly=True)
```

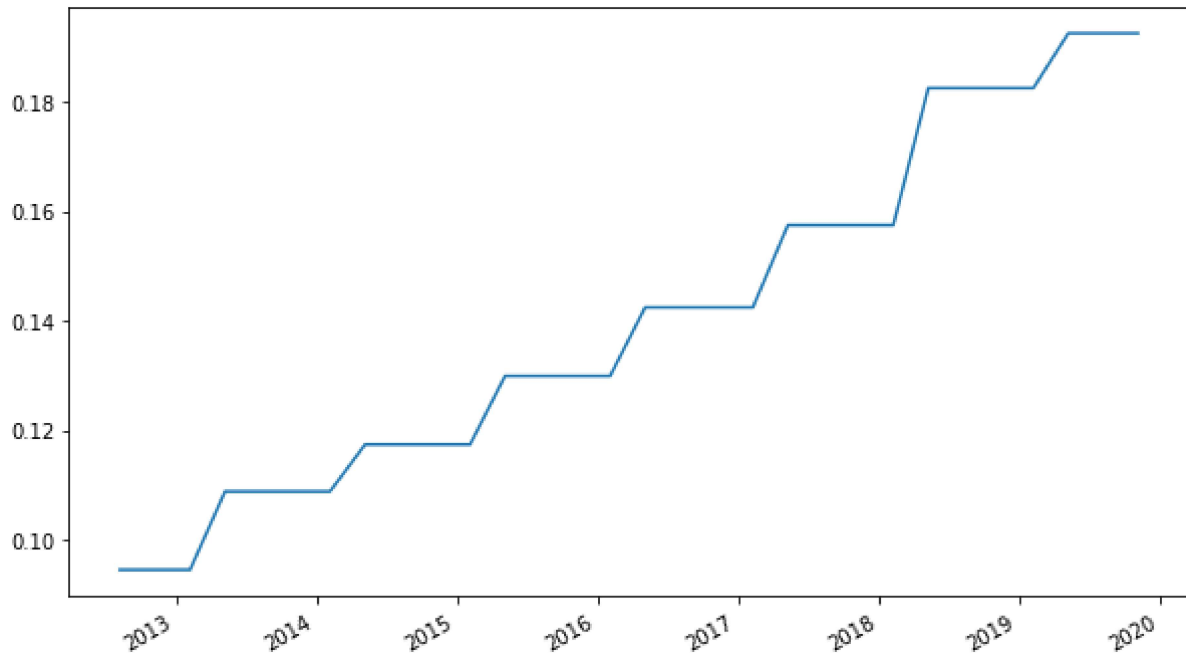
```
##Extract analyst info
si.get_analysts_info('aapl')
```

```
## extract dividends info
si.get_dividends('aapl', start_date=start_date, end_date=end_date)
```

```
## create a dataframe for visualisation
df = si.get_dividends('aapl', start_date=start_date, end_date=end_date)
```

```
##plot chart
df['dividend'].plot(kind='line', figsize=(10,6))
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f02a2e31390>



```
## Extract holders info
si.get_holders('aapl')
```

```
{'Major Holders':      0      1
0  0.07%      % of Shares Held by All Insider
1  59.58%      % of Shares Held by Institutions
2  59.62%      % of Float Held by Institutions
3  5455  Number of Institutions Holding Shares,
'Direct Holders (Forms 3 and 4)':
Reported % Out      Value      Holder      Shares Date
0  Vanguard Group, Inc. (The) 1277319054 Jun 29, 2022 7.95% 179271737025
1  Blackrock Inc. 1028688317 Jun 29, 2022 6.40% 144376411569
2  Berkshire Hathaway, Inc 894802319 Jun 29, 2022 5.57% 125585510933
3  State Street Corporation 598178524 Jun 29, 2022 3.72% 83954359494
4  FMR, LLC 344317974 Jun 29, 2022 2.14% 48325029752
5  Geode Capital Management, LLC 278256192 Jun 29, 2022 1.73% 39053258245
6  Price (T.Rowe) Associates Inc 237910783 Jun 29, 2022 1.48% 33390779846
7  Morgan Stanley 182450565 Jun 29, 2022 1.14% 25606937911
8  Northern Trust Corporation 179828922 Jun 29, 2022 1.12% 25238990300
9  Bank of America Corporation 149133915 Jun 29, 2022 0.93% 20930945880,
'Top Institutional Holders':
...      Value
0  Vanguard Total Stock Market Index Fund ... 63550026626
1  Vanguard 500 Index Fund ... 47765805012
2  Fidelity 500 Index Fund ... 23145638941
3  SPDR S&P 500 ETF Trust ... 23249292895
4  Invesco ETF Tr-Invesco QQQ Tr, Series 1 ETF ... 19902069459
5  iShares Core S&P 500 ETF ... 19111146227
6  Vanguard Growth Index Fund ... 17400728913
7  Vanguard Institutional Index Fund-Institutiona... ... 15988056418
8  Vanguard Information Technology Index Fund ... 10355263589
9  Select Sector SPDR Fund-Technology ... 8437296124
```

[10 rows x 5 columns]}

[Colab paid products](#) - [Cancel contracts here](#)

✓ 1s completed at 4:59 PM

