

Capstone Project

DATASET GENERATE

SYSTEM info:

MS WINDOWS 11

PYTHON VERSION: 3.10

DEEPNOTE PAID WEB SERVICE

Data scraping, also known as web scraping, is the process of importing information from a website into a spreadsheet or local file saved on your computer.

It's one of the most efficient ways to get data from the web, and in some cases to channel that data to another website.



Here I will show just basic example of how I created my dataset for my capstone project which took almost 52 hours.

```
In [ ]: # %%capture
!pip3 install snsrape

##Just Installing snsscrape
```

```
In [16]: import snsrape.modules.twitter as sntwitter
import pandas as pd
from datetime import datetime
```

```
In [17]: data = pd.DataFrame((sntwitter.TwitterSearchScraper(
    "#tsla since:2022-01-01 until:2021-01-07").get_items()))
```

```
In [20]: !snsrape twitter-search "#tsla since:2022-01-01 until:2022-01-07" > tsla.txt
```

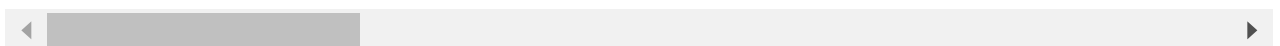
```
In [23]: data = pd.DataFrame((sntwitter.TwitterSearchScraper("#tsla since:2022-01-01 until:2022-
```

```
In [24]: data
```

Out[24]:

	url	date	content	rendered
0	https://twitter.com/teslaCupertino/status/1479...	2022-01-06 23:54:33+00:00	IMHO all stocks including #tsla will be availa...	IMHO including #t
1	https://twitter.com/teslaCupertino/status/1479...	2022-01-06 23:52:07+00:00	@bevedoni Agree. Hope to get #tsla at discount...	@bevedo Hope to g
2	https://twitter.com/10trillionBTC/status/14792...	2022-01-06 23:51:59+00:00	Did #TSLA just bought another \$1.5bn worth of ...	Did #TSLA ju another \$1..
3	https://twitter.com/yubirunsny/status/1479237...	2022-01-06 23:44:54+00:00	@KewalTony @squawksquare Tomorrow #tsla will b...	@f @squa Tomorrow #t
4	https://twitter.com/Moshkelgosha_M/status/1479...	2022-01-06 23:24:15+00:00	#TSLA do you know the one who sits at the top ...	#TSLA do the one who
...
1569	https://twitter.com/Lowcarb_Style/status/14770...	2022-01-01 00:37:39+00:00	テスラのモデル3とモデルSのリコール、「セキュリティリスク」として中国に拡大! https:...	テスラのモデルSのリ「セキュリティク」として大!
1570	https://twitter.com/tumtmum/status/14770759697...	2022-01-01 00:35:28+00:00	テスラ、47万台リコール、株価下がるかな? \n\n#テスラ \n\n#NASDAQ\n#TSLA\...	テスラ、47万ル、株価下か \n \n#NASDAQ\
1571	https://twitter.com/white55605421/status/14770...	2022-01-01 00:24:42+00:00	2022年 01.01\n#VOO ▲0.27% YTD27.02%\n#QQQ ▲0.62...	2022年 01.0 . YTD27.02%\
1572	https://twitter.com/NorthernCow_Boy/status/147...	2022-01-01 00:06:05+00:00	@HeartDocTesla 290,000 deliveries and the roof...	@Hea 290,000 deli
1573	https://twitter.com/JesseCochard/status/147706...	2022-01-01 00:02:02+00:00	@kchau @garyblack00 @WholeMarsBlog Elon increa...	@kchau @ga @WholeMars

1574 rows × 27 columns



In [29]:

```
#!snsrape --jsonl --since '2021-01-01' twitter-search "#Covid19#Gaming until:2021-06-2
```

```
##THIS IS THE CONSTRAINT I USED IT.
```

```
In [ ]:
```

```
## Now we get the dataset we will further progress to basic knowledge of our dataset.
```