

tweets_search_api_coffee_island

June 25, 2021

0.0.1 Finding your Python Path

Your Python path can be displayed using the built-in `os` module. The `OS` module is for operating system dependent functionality into Python programs and scripts.

To find your current working directory, the function required is `os.getcwd()`. The `os.listdir()` function can be used to display all files in a directory, which is a good check to see if the CSV file you are loading is in the directory as expected.

```
# Find out your current working directory
import os
print(os.getcwd())

# Out: /Users/shane/Documents/blog

# Display all of the files found in your current working directory
print(os.listdir(os.getcwd()))

# Out: ['test_delimited.ssv', 'CSV Blog.ipynb', 'test_data.csv']
```

Instead of moving the required data files to your working directory, you can also change your current working directory to the directory where the files reside using `os.chdir()`.

0.0.2 Advanced Read CSV Files

There are some additional flexible parameters in the Pandas `read_csv()` function that are useful to have in your arsenal of data science techniques:

Specifying Data Types As mentioned before, CSV files do not contain any type information for data. Data types are inferred through examination of the top rows of the file, which can lead to errors. To manually specify the data types for different columns, the `dtype` parameter can be used with a dictionary of column names and data types to be applied, for example:

```
dtype={"name": str, "age": np.int32}
```

Note that for dates and date times, the format, columns, and other behaviour can be adjusted using `parse_dates`, `date_parser`, `dayfirst`, `keep_date` parameters.

There is a `parse_dates` parameter for `read_csv` which allows you to define the names of the columns you want treated as dates or datetimes:

```
date_cols = ['col1', 'col2']
```

```
pd.read_csv(file, sep='\t', header=None, names=headers, parse_dates=date_cols)
```

```
[27]: import pandas as pd
```

```
pdf2 = pd.read_csv('output3.csv')
pdf2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6510 entries, 0 to 6509
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   text                   6510 non-null   object
1   favourite_count        6510 non-null   int64
2   retweet_count          6510 non-null   int64
3   created_at             6510 non-null   object
dtypes: int64(2), object(2)
memory usage: 203.6+ KB
```

My workaround was to load as its default type, then use `pandas.to_datetime()` function one line down.

```
df[target_col] = pd.to_datetime(df[target_col])
```

```
[128]: pdf2['created_at'] = pd.to_datetime(pdf2['created_at'])
pdf2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6510 entries, 0 to 6509
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   text                   6510 non-null   object
1   favourite_count        6510 non-null   int64
2   retweet_count          6510 non-null   int64
3   created_at             6510 non-null   datetime64[ns]
dtypes: datetime64[ns](1), int64(2), object(1)
memory usage: 203.6+ KB
```

```
[118]: import pandas as pd
```

```
pdf2 = pd.read_csv('output3.csv', parse_dates = ['created_at',])
pdf2.head(5)
```

```
[118]:
```

| | text | favourite_count | \ |
|---|---|-----------------|---|
| 0 | @breakfasttv I received the Astrazeneca vaccin... | 0 | |
| 1 | #Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJo... | 3 | |
| 2 | RT @TimWattsMP: Can you believe this?\n\nThe M... | 0 | |
| 3 | RT @DrEricDing: 5) One dose of the vaccine is ... | 0 | |

```
4 RT @DrEricDing: 5) One dose of the vaccine is ... 0
```

| | retweet_count | created_at |
|---|---------------|---------------------|
| 0 | 0 | 2021-06-20 23:59:37 |
| 1 | 5 | 2021-06-20 23:59:33 |
| 2 | 212 | 2021-06-20 23:58:46 |
| 3 | 270 | 2021-06-20 23:57:54 |
| 4 | 270 | 2021-06-20 23:56:29 |

```
[119]: pdf2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6510 entries, 0 to 6509
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   text            6510 non-null   object
1   favourite_count  6510 non-null   int64
2   retweet_count   6510 non-null   int64
3   created_at      6510 non-null   datetime64[ns]
dtypes: datetime64[ns](1), int64(2), object(1)
memory usage: 203.6+ KB
```

1 Get Tweets

This script extracts all the tweets with hashtag #covid-19 related to the day before today (yesterday) and saves them into a .csv file. We use the `tweepy` library, which can be installed with the command `pip install tweepy`.

Firstly, we import the configuration file, called `config.py`, which is located in the same directory of this script.

```
[14]: import os
      jv = os.environ.get('JAVA_HOME', None)

      import findspark
      findspark.init()

      os.environ['PYSPARK_SUBMIT_ARGS'] = \
      '--packages com.johnsnowlabs.nlp:spark-nlp_2.12:3.0.0 pyspark-shell'
      # '--packages org.postgresql:postgresql:42.1.1 pyspark-shell'

      sys.path
```

```
[14]: ['/opt/spark/python',
      '/opt/spark/python/lib/py4j-0.10.9-src.zip',
      '/opt/spark/python',
```

```

'/tmp/spark-b26658bd-6681-46ae-8cbe-
ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-bd30-efe7fbc1b256/org.joda_joda-
convert-1.8.1.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-
ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-bd30-efe7fbc1b256/joda-time_joda-
time-2.9.5.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/dk.brics.automaton_automaton-1.11-8.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.google.code.gson_gson-2.3.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.google.code.findbugs_jsr305-3.0.1.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-
ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-bd30-efe7fbc1b256/net.jcip_jcip-
annotations-1.0.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/org.slf4j_slf4j-api-1.7.21.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/org.projectlombok_lombok-1.16.8.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/it.unimi.dsi_fastutil-7.0.12.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.google.protobuf_protobuf-java-3.0.0-beta-3.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.google.protobuf_protobuf-java-util-3.0.0-beta-3.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.google.code.findbugs_annotations-3.0.1.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/net.sf.trove4j_trove4j-3.0.3.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.johnsnowlabs.nlp_tensorflow-cpu_2.12-0.2.2.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/org.json4s_json4s-ext_2.12-3.5.3.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.navigamez_greex-1.0.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.github.universal-automata_liblevenshtein-3.0.0.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-
ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-bd30-efe7fbc1b256/com.amazonaws_aws-
java-sdk-bundle-1.11.603.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/org.rocksdb_rocksdbjni-6.5.3.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.typesafe_config-1.3.0.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-b
d30-efe7fbc1b256/com.johnsnowlabs.nlp_spark-nlp_2.12-3.0.0.jar',
'/tmp/spark-b26658bd-6681-46ae-8cbe-

```

```

ba891f867721/userFiles-b8b7c7f9-a77b-4ea4-bd30-efe7fbc1b256',
  '/opt/spark/python/lib/py4j-0.10.9-src.zip',
  '/home/hadoopuser/Documents/twitter_search_api',
  '/home/hadoopuser/.vscode/extensions/ms-
toolsai.jupyter-2021.6.999662501/pythonFiles/vscode_datascience_helpers',
  '/home/hadoopuser/.vscode/extensions/ms-
toolsai.jupyter-2021.6.999662501/pythonFiles',
  '/home/hadoopuser/.vscode/extensions/ms-
toolsai.jupyter-2021.6.999662501/pythonFiles/lib/python',
  '/opt/anaconda/envs/pyspark_env/lib/python37.zip',
  '/opt/anaconda/envs/pyspark_env/lib/python3.7',
  '/opt/anaconda/envs/pyspark_env/lib/python3.7/lib-dynload',
  '',
  '/opt/anaconda/envs/pyspark_env/lib/python3.7/site-packages',
  '/opt/anaconda/envs/pyspark_env/lib/python3.7/site-
packages/IPython/extensions',
  '/home/hadoopuser/.ipython',
  '/home/hadoopuser/.ivy2/jars/net.sf.trove4j_trove4j-3.0.3.jar',
  '/home/hadoopuser/.ivy2/jars/org.mongodb.spark_mongo-spark-
connector_2.11-2.4.3.jar',
  '/home/hadoopuser/.ivy2/jars/joda-time_joda-time-2.9.5.jar',
  '/home/hadoopuser/.ivy2/jars/org.joda_joda-convert-1.8.1.jar',
  '/home/hadoopuser/.ivy2/jars/org.mongodb_mongo-java-driver-3.12.5.jar',
  '/home/hadoopuser/.ivy2/jars/com.amazonaws_aws-java-sdk-bundle-1.11.603.jar',
  '/home/hadoopuser/.ivy2/jars/org.slf4j_slf4j-api-1.7.21.jar',
  '/home/hadoopuser/.ivy2/jars/org.rocksdb_rocksdbjni-6.5.3.jar',
  '/home/hadoopuser/.ivy2/jars/com.typesafe_config-1.3.0.jar',
  '/home/hadoopuser/.ivy2/jars/com.google.code.findbugs_annotations-3.0.1.jar',
  '/home/hadoopuser/.ivy2/jars/org.projectlombok_lombok-1.16.8.jar',
  '/home/hadoopuser/.ivy2/jars/com.github.universal-
automata_liblevenshtein-3.0.0.jar',
  '/home/hadoopuser/.ivy2/jars/net.jcip_jcip-annotations-1.0.jar',
  '/home/hadoopuser/.ivy2/jars/com.johnsnowlabs.nlp_tensorflow-
cpu_2.12-0.2.2.jar',
  '/home/hadoopuser/.ivy2/jars/com.google.code.findbugs_jsr305-3.0.1.jar',
  '/home/hadoopuser/.ivy2/jars/dk.brics.automaton_automaton-1.11-8.jar',
  '/home/hadoopuser/.ivy2/jars/com.google.code.gson_gson-2.3.jar',
  '/home/hadoopuser/.ivy2/jars/org.json4s_json4s-ext_2.12-3.5.3.jar',
  '/home/hadoopuser/.ivy2/jars/com.navigamez_greex-1.0.jar',
  '/home/hadoopuser/.ivy2/jars/org.mongodb_mongodb-driver-core-4.0.5.jar',
  '/home/hadoopuser/.ivy2/jars/com.google.protobuf_protobuf-java-
util-3.0.0-beta-3.jar',
  '/home/hadoopuser/.ivy2/jars/com.google.protobuf_protobuf-
java-3.0.0-beta-3.jar',
  '/home/hadoopuser/.ivy2/jars/org.mongodb.spark_mongo-spark-
connector_2.12-3.0.0.jar',
  '/home/hadoopuser/.ivy2/jars/org.mongodb_mongodb-driver-sync-4.0.5.jar',

```

```

'/home/hadoopuser/.ivy2/jars/org.mongodb_bson-4.0.5.jar',
'/home/hadoopuser/.ivy2/jars/it.unimi.dsi_fastutil-7.0.12.jar',
'/home/hadoopuser/.ivy2/jars/com.johnsnowlabs.nlp_spark-nlp_2.12-3.0.0.jar']

```

```

[15]: import sys, glob, os
sys.path.extend(glob.glob(os.path.join(os.path.expanduser("~"), ".ivy2/jars/*.
↪jar")))

```

```

[17]: from pyspark.sql import SparkSession
import pyspark.sql.functions as F

spark=SparkSession.builder \
    .appName("Spark_and_Pandas_twitter_dfs") \
    .master("local[*]") \
    .getOrCreate()

```

```

2021-06-25 02:59:45,198 - DEBUG - Command to send: c
o29
sessionState
e

```

```

2021-06-25 02:59:45,201 - DEBUG - Answer received: !yro42
2021-06-25 02:59:45,203 - DEBUG - Command to send: c
o42
conf
e

```

```

2021-06-25 02:59:45,205 - DEBUG - Answer received: !yro43
2021-06-25 02:59:45,207 - DEBUG - Command to send: c
o43
setConfString
sspark.app.name
sSpark_and_Pandas_twitter_dfs
e

```

```

2021-06-25 02:59:45,210 - DEBUG - Answer received: !yv
2021-06-25 02:59:45,212 - DEBUG - Command to send: c
o29
sessionState
e

```

```

2021-06-25 02:59:45,213 - DEBUG - Answer received: !yro44
2021-06-25 02:59:45,217 - DEBUG - Command to send: c
o44
conf
e

```

2021-06-25 02:59:45,219 - DEBUG - Answer received: !yro45

2021-06-25 02:59:45,230 - DEBUG - Command to send: c

o45

setConfString

sspark.master

slocal[*]

e

2021-06-25 02:59:45,235 - DEBUG - Answer received: !yv

[18]: spark

2021-06-25 02:59:53,695 - DEBUG - Command to send: c

o35

get

sspark.sql.catalogImplementation

e

2021-06-25 02:59:53,697 - DEBUG - Answer received: !ysin-memory

2021-06-25 02:59:53,698 - DEBUG - Command to send: c

o19

sc

e

2021-06-25 02:59:53,699 - DEBUG - Answer received: !yro46

2021-06-25 02:59:53,700 - DEBUG - Command to send: c

o46

uiWebUrl

e

2021-06-25 02:59:53,701 - DEBUG - Answer received: !yro47

2021-06-25 02:59:53,702 - DEBUG - Command to send: c

o47

get

e

2021-06-25 02:59:53,703 - DEBUG - Answer received: !yshttp://spark-client:4040

2021-06-25 02:59:53,704 - DEBUG - Command to send: c

o19

version

e

2021-06-25 02:59:53,707 - DEBUG - Answer received: !ys3.0.0

[18]: <pyspark.sql.session.SparkSession at 0x7fb7840e9990>

```
[19]: from config import *
import tweepy
import datetime

import sys
import logging

logger = logging.getLogger('tweets_search')
```

```
[20]: import pandas as pd
# import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

```
[21]: print(f"logger.root.level = {logger.root.level}, logger.root.name = {logger.
↪root.name}")
print(f"logger.name = {logger.name}")
```

```
logger.root.level = 10, logger.root.name = root
logger.name = tweets_search
```

```
[22]: format = "%(asctime)s - %(levelname)s - %(message)s"
logging.basicConfig(format=format, stream=sys.stdout, level = logging.DEBUG)
# logging.basicConfig(format=format, stream=sys.stdout, level = logging.INFO)
```

```
[9]: print(logger.root.level)
```

```
10
```

```
[17]: # logger.root.level = 10
```

```
[18]: # print(logger.root.level)
```

```
10
```

We setup the connection to our Twitter App by using the OAuthHandler() class and its access_token() function. Then we call the Twitter API through the API() function.

```
[23]: auth = tweepy.OAuthHandler(TWITTER_CONSUMER_KEY, TWITTER_CONSUMER_SECRET)
auth.set_access_token(TWITTER_ACCESS_TOKEN, TWITTER_ACCESS_TOKEN_SECRET)
api = tweepy.API(auth,wait_on_rate_limit=True, wait_on_rate_limit_notify = True)
```

```
2021-06-25 03:00:27,051 - DEBUG - Command to send: m
d
o38
e
```

```
2021-06-25 03:00:27,053 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,054 - DEBUG - Command to send: m
```


d
o39
e

2021-06-25 03:00:27,055 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,055 - DEBUG - Command to send: m

d
o40
e

2021-06-25 03:00:27,056 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,057 - DEBUG - Command to send: m

d
o41
e

2021-06-25 03:00:27,057 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,058 - DEBUG - Command to send: m

d
o42
e

2021-06-25 03:00:27,059 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,059 - DEBUG - Command to send: m

d
o43
e

2021-06-25 03:00:27,060 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,061 - DEBUG - Command to send: m

d
o44
e

2021-06-25 03:00:27,062 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,063 - DEBUG - Command to send: m

d
o45
e

2021-06-25 03:00:27,066 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,067 - DEBUG - Command to send: m

d
o46
e

2021-06-25 03:00:27,072 - DEBUG - Answer received: !yv
2021-06-25 03:00:27,079 - DEBUG - Command to send: m

d
o47
e

2021-06-25 03:00:27,083 - DEBUG - Answer received: !yv

[11]: api.me()

```
2021-06-24 23:14:44,448 - DEBUG - PARAMS: {}
2021-06-24 23:14:44,450 - DEBUG - Signing request <PreparedRequest [GET]> using
client <Client client_key=pw0ihLFxH3nwDrd4HBd7pqUrc, client_secret=****,
resource_owner_key=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz,
resource_owner_secret=****, signature_method=HMAC-SHA1,
signature_type=AUTH_HEADER, callback_uri=None, rsa_key=None, verifier=None,
realm=None, encoding=utf-8, decoding=None, nonce=None, timestamp=None>
2021-06-24 23:14:44,451 - DEBUG - Including body in call to sign: False
2021-06-24 23:14:44,455 - DEBUG - Collected params: [('oauth_nonce',
'111288902889639547901624565684'), ('oauth_timestamp', '1624565684'),
('oauth_version', '1.0'), ('oauth_signature_method', 'HMAC-SHA1'),
('oauth_consumer_key', 'pw0ihLFxH3nwDrd4HBd7pqUrc'), ('oauth_token',
'1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz')]
2021-06-24 23:14:44,458 - DEBUG - Normalized params: oauth_consumer_key=pw0ihLFx
H3nwDrd4HBd7pqUrc&oauth_nonce=111288902889639547901624565684&oauth_signature_met
hod=HMAC-SHA1&oauth_timestamp=1624565684&oauth_token=1360011857969479682-iLrxBUL
qdtExwkqiN9iZsHYDXIFTZz&oauth_version=1.0
2021-06-24 23:14:44,461 - DEBUG - Normalized URI:
https://api.twitter.com/1.1/account/verify_credentials.json
2021-06-24 23:14:44,462 - DEBUG - Signing: signature base string: GET&https%3A%2
F%2Fapi.twitter.com%2F1.1%2Faccount%2Fverify_credentials.json&oauth_consumer_key
%3Dpw0ihLFxH3nwDrd4HBd7pqUrc%26oauth_nonce%3D111288902889639547901624565684%26oa
uth_signature_method%3DHMAC-SHA1%26oauth_timestamp%3D1624565684%26oauth_token%3D
1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz%26oauth_version%3D1.0
2021-06-24 23:14:44,475 - DEBUG - Signature: OPJvHp+wfn64s6RCBDDmvG6Xev8=
2021-06-24 23:14:44,477 - DEBUG - Updated url:
https://api.twitter.com/1.1/account/verify_credentials.json
2021-06-24 23:14:44,478 - DEBUG - Updated headers: {'Host': 'api.twitter.com',
'Authorization': 'OAuth oauth_nonce="111288902889639547901624565684",
oauth_timestamp="1624565684", oauth_version="1.0", oauth_signature_method="HMAC-
SHA1", oauth_consumer_key="pw0ihLFxH3nwDrd4HBd7pqUrc",
oauth_token="1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz",
oauth_signature="OPJvHp%2Bwfn64s6RCBDDmvG6Xev8%3D"' }
2021-06-24 23:14:44,478 - DEBUG - Updated body: None
2021-06-24 23:14:44,483 - DEBUG - Starting new HTTPS connection (1):
api.twitter.com:443
2021-06-24 23:14:46,587 - DEBUG - https://api.twitter.com:443 "GET
/1.1/account/verify_credentials.json HTTP/1.1" 200 1295
2021-06-24 23:14:46,601 - DEBUG - PARAMS: {'screen_name': b'vasiange'}
2021-06-24 23:14:46,608 - DEBUG - Signing request <PreparedRequest [GET]> using
```

```

client <Client client_key=pw0ihLFxH3nwDrd4HBd7pqUrc, client_secret=****,
resource_owner_key=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz,
resource_owner_secret=****, signature_method=HMAC-SHA1,
signature_type=AUTH_HEADER, callback_uri=None, rsa_key=None, verifier=None,
realm=None, encoding=utf-8, decoding=None, nonce=None, timestamp=None>
2021-06-24 23:14:46,609 - DEBUG - Including body in call to sign: False
2021-06-24 23:14:46,615 - DEBUG - Collected params: [('screen_name',
'vasiange'), ('oauth_nonce', '85808766550224443101624565686'),
('oauth_timestamp', '1624565686'), ('oauth_version', '1.0'),
('oauth_signature_method', 'HMAC-SHA1'), ('oauth_consumer_key',
'pw0ihLFxH3nwDrd4HBd7pqUrc'), ('oauth_token',
'1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz')]
2021-06-24 23:14:46,615 - DEBUG - Normalized params: oauth_consumer_key=pw0ihLFx
H3nwDrd4HBd7pqUrc&oauth_nonce=85808766550224443101624565686&oauth_signature_meth
od=HMAC-SHA1&oauth_timestamp=1624565686&oauth_token=1360011857969479682-iLrxBULq
dtExwkqiN9iZsHYDXIFTZz&oauth_version=1.0&screen_name=vasiange
2021-06-24 23:14:46,616 - DEBUG - Normalized URI:
https://api.twitter.com/1.1/users/show.json
2021-06-24 23:14:46,617 - DEBUG - Signing: signature base string: GET&https%3A%2
F%2Fapi.twitter.com%2F1.1%2Fusers%2Fshow.json&oauth_consumer_key%3Dpw0ihLFxH3nwD
rd4HBd7pqUrc%26oauth_nonce%3D85808766550224443101624565686%26oauth_signature_met
hod%3DHMAC-SHA1%26oauth_timestamp%3D1624565686%26oauth_token%3D13600118579694796
82-iLrxBULqdtExwkqiN9iZsHYDXIFTZz%26oauth_version%3D1.0%26screen_name%3Dvasiange
2021-06-24 23:14:46,617 - DEBUG - Signature: nUeyDo6knkuU6DKEjksoueoDrS0=
2021-06-24 23:14:46,618 - DEBUG - Updated url:
https://api.twitter.com/1.1/users/show.json?screen_name=vasiange
2021-06-24 23:14:46,622 - DEBUG - Updated headers: {'Host': 'api.twitter.com',
'Authorization': 'OAuth oauth_nonce="85808766550224443101624565686",
oauth_timestamp="1624565686", oauth_version="1.0", oauth_signature_method="HMAC-
SHA1", oauth_consumer_key="pw0ihLFxH3nwDrd4HBd7pqUrc",
oauth_token="1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz",
oauth_signature="nUeyDo6knkuU6DKEjksoueoDrS0%3D"}
2021-06-24 23:14:46,623 - DEBUG - Updated body: None
2021-06-24 23:14:46,626 - DEBUG - Starting new HTTPS connection (1):
api.twitter.com:443
2021-06-24 23:14:47,602 - DEBUG - https://api.twitter.com:443 "GET
/1.1/users/show.json?screen_name=vasiange HTTP/1.1" 200 1319

```

```

[11]: User(_api=<tweepy.api.API object at 0x7f6253980850>, _json={'id':
1360011857969479682, 'id_str': '1360011857969479682', 'name': 'vasiange',
'screen_name': 'vasiange', 'location': '', 'profile_location': None,
'description': '', 'url': None, 'entities': {'description': {'urls': []}},
'protected': False, 'followers_count': 0, 'friends_count': 15, 'listed_count':
0, 'created_at': 'Thu Feb 11 23:45:06 +0000 2021', 'favourites_count': 0,
'utc_offset': None, 'time_zone': None, 'geo_enabled': False, 'verified': False,
'statuses_count': 0, 'lang': None, 'contributors_enabled': False,
'is_translator': False, 'is_translation_enabled': False,

```

```
'profile_background_color': 'F5F8FA', 'profile_background_image_url': None,
'profile_background_image_url_https': None, 'profile_background_tile': False,
'profile_image_url':
'http://abs.twimg.com/sticky/default_profile_images/default_profile_normal.png',
'profile_image_url_https': 'https://abs.twimg.com/sticky/default_profile_images/
default_profile_normal.png', 'profile_link_color': '1DA1F2',
'profile_sidebar_border_color': 'CODEED', 'profile_sidebar_fill_color':
'DDEEF6', 'profile_text_color': '333333', 'profile_use_background_image': True,
'has_extended_profile': True, 'default_profile': True, 'default_profile_image':
True, 'following': False, 'follow_request_sent': False, 'notifications': False,
'translator_type': 'none', 'withheld_in_countries': [], 'suspended': False,
'needs_phone_verification': False}, id=1360011857969479682,
id_str='1360011857969479682', name='vasiange', screen_name='vasiange',
location='', profile_location=None, description='', url=None,
entities={'description': {'urls': []}}, protected=False, followers_count=0,
friends_count=15, listed_count=0, created_at=datetime.datetime(2021, 2, 11, 23,
45, 6), favourites_count=0, utc_offset=None, time_zone=None, geo_enabled=False,
verified=False, statuses_count=0, lang=None, contributors_enabled=False,
is_translator=False, is_translation_enabled=False,
profile_background_color='F5F8FA', profile_background_image_url=None,
profile_background_image_url_https=None, profile_background_tile=False, profile_
image_url='http://abs.twimg.com/sticky/default_profile_images/default_profile_no
rmal.png', profile_image_url_https='https://abs.twimg.com/sticky/default_profile
_images/default_profile_normal.png', profile_link_color='1DA1F2',
profile_sidebar_border_color='CODEED', profile_sidebar_fill_color='DDEEF6',
profile_text_color='333333', profile_use_background_image=True,
has_extended_profile=True, default_profile=True, default_profile_image=True,
following=False, follow_request_sent=False, notifications=False,
translator_type='none', withheld_in_countries=[], suspended=False,
needs_phone_verification=False)
```

```
[12]: api.rate_limit_status()
```

```
2021-06-24 23:14:55,395 - DEBUG - PARAMS: {}
2021-06-24 23:14:55,396 - DEBUG - Signing request <PreparedRequest [GET]> using
client <Client client_key=pw0ihLFxH3nwDrd4HBd7pqUrc, client_secret=****,
resource_owner_key=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz,
resource_owner_secret=****, signature_method=HMAC-SHA1,
signature_type=AUTH_HEADER, callback_uri=None, rsa_key=None, verifier=None,
realm=None, encoding=utf-8, decoding=None, nonce=None, timestamp=None>
2021-06-24 23:14:55,398 - DEBUG - Including body in call to sign: False
2021-06-24 23:14:55,399 - DEBUG - Collected params: [('oauth_nonce',
'126286861049097242631624565695'), ('oauth_timestamp', '1624565695'),
('oauth_version', '1.0'), ('oauth_signature_method', 'HMAC-SHA1'),
('oauth_consumer_key', 'pw0ihLFxH3nwDrd4HBd7pqUrc'), ('oauth_token',
'1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz')]
2021-06-24 23:14:55,400 - DEBUG - Normalized params: oauth_consumer_key=pw0ihLFx
```

```

H3nwDrd4HBd7pqUrc&oauth_nonce=126286861049097242631624565695&oauth_signature_met
hod=HMAC-SHA1&oauth_timestamp=1624565695&oauth_token=1360011857969479682-iLrxBUL
qdtExwkqiN9iZsHYDXIFTZz&oauth_version=1.0
2021-06-24 23:14:55,401 - DEBUG - Normalized URI:
https://api.twitter.com/1.1/application/rate_limit_status.json
2021-06-24 23:14:55,406 - DEBUG - Signing: signature base string: GET&https%3A%2
F%2Fapi.twitter.com%2F1.1%2Fapplication%2Frate_limit_status.json&oauth_consumer_
key%3Dpw0ihLFxH3nwDrd4HBd7pqUrc%26oauth_nonce%3D126286861049097242631624565695%2
6oauth_signature_method%3DHMAC-SHA1%26oauth_timestamp%3D1624565695%26oauth_token
%3D1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz%26oauth_version%3D1.0
2021-06-24 23:14:55,407 - DEBUG - Signature: pKwVTEhN0+b6fx7rHgc2g6A84Ls=
2021-06-24 23:14:55,407 - DEBUG - Updated url:
https://api.twitter.com/1.1/application/rate_limit_status.json
2021-06-24 23:14:55,408 - DEBUG - Updated headers: {'Host': 'api.twitter.com',
'Authorization': 'OAuth oauth_nonce="126286861049097242631624565695",
oauth_timestamp="1624565695", oauth_version="1.0", oauth_signature_method="HMAC-
SHA1", oauth_consumer_key="pw0ihLFxH3nwDrd4HBd7pqUrc",
oauth_token="1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz",
oauth_signature="pKwVTEhN0%2Bb6fx7rHgc2g6A84Ls%3D"'}}
2021-06-24 23:14:55,412 - DEBUG - Updated body: None
2021-06-24 23:14:55,417 - DEBUG - Starting new HTTPS connection (1):
api.twitter.com:443
2021-06-24 23:14:56,715 - DEBUG - https://api.twitter.com:443 "GET
/1.1/application/rate_limit_status.json HTTP/1.1" 200 18959

```

```

[12]: reset': 1624566596}},
      'guide': {'/guide': {'limit': 180, 'remaining': 180, 'reset': 1624566596},
                 '/guide/get_explore_locations': {'limit': 100,
                 'remaining': 100,
                 'reset': 1624566596},
                 '/guide/explore_locations_with_autocomplete': {'limit': 200,
                 'remaining': 200,
                 'reset': 1624566596}},
      'auth': {'/auth/csrf_token': {'limit': 15,
                                     'remaining': 15,
                                     'reset': 1624566596}},
      'blocks': {'/blocks/list': {'limit': 15,
                                   'remaining': 15,
                                   'reset': 1624566596},
                 '/blocks/ids': {'limit': 15, 'remaining': 15, 'reset': 1624566596}},
      'geo': {'/geo/similar_places': {'limit': 15,
                                       'remaining': 15,
                                       'reset': 1624566596},
              '/geo/place_page': {'limit': 75, 'remaining': 75, 'reset': 1624566596},
              '/geo/id/:place_id': {'limit': 75, 'remaining': 75, 'reset': 1624566596},
              '/geo/reverse_geocode': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
              '/geo/search': {'limit': 15, 'remaining': 15, 'reset': 1624566596}},

```

```

'users': {'/users/': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
'/users/:id': {'limit': 900, 'remaining': 899, 'reset': 1624566587},
'/users/report_spam': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
'/users/:source_user_id/blocking/:target_user_id&DELETE': {'limit': 50,
'remaining': 50,
'reset': 1624566596},
'/users/contributors/pending': {'limit': 2000,
'remaining': 2000,
'reset': 1624566596},
'/users/show/:id': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
'/users/:source_user_id/following&POST': {'limit': 50,
'remaining': 50,
'reset': 1624566596},
'/users/:id/tweets': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
'/users/:id/retweets/:source_tweet_id&DELETE': {'limit': 50,
'remaining': 50,
'reset': 1624566596},
'/users/search': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
'/users/:id/likes&POST': {'limit': 50,
'remaining': 50,
'reset': 1624566596},
'/users/suggestions/:slug': {'limit': 15,
'remaining': 15,
'reset': 1624566596},
'/users/contributees/pending': {'limit': 200,
'remaining': 200,
'reset': 1624566596},
'/users/:id/retweets&POST': {'limit': 50,
'remaining': 50,
'reset': 1624566596},
'/users/profile_banner': {'limit': 180,
'remaining': 180,
'reset': 1624566596},
'/users/by/username/:source_username/following/:target_user_name&DELETE':
{'limit': 50,
'remaining': 50,
'reset': 1624566596},
'/users/by/username/:handle/tweets': {'limit': 900,
'remaining': 900,
'reset': 1624566596},
'/users/derived_info': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
'/users/:id/blocking&POST': {'limit': 50,
'remaining': 50,
'reset': 1624566596},
'/users/by/username/:source_username/following&POST': {'limit': 50,
'remaining': 50,
'reset': 1624566596},

```

```

'/users/:id/followers': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
'/users/suggestions/:slug/members': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
'/users/:id/muting': {'limit': 50, 'remaining': 50, 'reset': 1624566596},
'/users/:id/following': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
'/users/:id/mentions': {'limit': 180,
  'remaining': 180,
  'reset': 1624566596},
'/users/:source_user_id/following/:target_user_id&DELETE': {'limit': 50,
  'remaining': 50,
  'reset': 1624566596},
'/users/by/username/:username': {'limit': 900,
  'remaining': 900,
  'reset': 1624566596},
'/users/by/username/:username/followers': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
'/users/lookup': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
'/users/:id/blocking': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
'/users/suggestions': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
'/users/:id/likes/:tweet_id&DELETE': {'limit': 50,
  'remaining': 50,
  'reset': 1624566596},
'/users/by/username/:username/following': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
'/users/by/username/:handle/mentions': {'limit': 180,
  'remaining': 180,
  'reset': 1624566596},
'/users/by': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
'/users/:id/liked_tweets': {'limit': 75,
  'remaining': 75,
  'reset': 1624566596}},
'teams': {'/teams/authorize': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596}},
'followers': {'/followers/ids': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/followers/list': {'limit': 15, 'remaining': 15, 'reset': 1624566596}},
'collections': {'/collections/list': {'limit': 1000,
  'remaining': 1000,
  'reset': 1624566596},
  '/collections/entries': {'limit': 1000,
  'remaining': 1000,
  'reset': 1624566596},

```

```

    '/collections/show': {'limit': 1000,
        'remaining': 1000,
        'reset': 1624566596}},
    'permissions': {'/permissions/user_permissions/admin_email_verification':
{'limit': 3,
    'remaining': 3,
    'reset': 1624566596},
    '/permissions/user_permissions': {'limit': 3,
        'remaining': 3,
        'reset': 1624566596}},
    'statuses': {'/statuses/retweeters/ids': {'limit': 75,
        'remaining': 75,
        'reset': 1624566596},
        '/statuses/retweets_of_me': {'limit': 75,
            'remaining': 75,
            'reset': 1624566596},
        '/statuses/home_timeline': {'limit': 15,
            'remaining': 15,
            'reset': 1624566596},
        '/statuses/show/:id': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
        '/statuses/user_timeline': {'limit': 900,
            'remaining': 900,
            'reset': 1624566596},
        '/statuses/friends': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
        '/statuses/retweets/:id': {'limit': 75,
            'remaining': 75,
            'reset': 1624566596},
        '/statuses/mentions_timeline': {'limit': 75,
            'remaining': 75,
            'reset': 1624566596},
        '/statuses/oembed': {'limit': 180, 'remaining': 180, 'reset': 1624566596},
        '/statuses/lookup': {'limit': 900, 'remaining': 900, 'reset': 1624566596}},
    'custom_profiles': {'/custom_profiles/list': {'limit': 180,
        'remaining': 180,
        'reset': 1624566596},
        '/custom_profiles/show': {'limit': 180,
            'remaining': 180,
            'reset': 1624566596}},
    'webhooks': {'/webhooks/subscriptions/direct_messages': {'limit': 15,
        'remaining': 15,
        'reset': 1624566596},
        '/webhooks': {'limit': 15, 'remaining': 15, 'reset': 1624566596}},
    'contacts': {'/contacts/uploaded_by': {'limit': 300,
        'remaining': 300,
        'reset': 1624566596},
        '/contacts/users': {'limit': 300, 'remaining': 300, 'reset': 1624566596},
        '/contacts/addressbook': {'limit': 300,

```



```

    'remaining': 300,
    'reset': 1624566596},
  '/contacts/users_and_uploaded_by': {'limit': 300,
    'remaining': 300,
    'reset': 1624566596},
  '/contacts/delete/status': {'limit': 300,
    'remaining': 300,
    'reset': 1624566596}},
'labs': {'/labs/2/platform_manipulation/reports': {'limit': 5,
  'remaining': 5,
  'reset': 1624566596},
  '/labs/:version/tweets/:id/hidden&PUT': {'limit': 10,
  'remaining': 10,
  'reset': 1624566596},
  '/labs/:version/tweets/stream/filter/': {'limit': 50,
  'remaining': 50,
  'reset': 1624566596},
  '/labs/:version/users/:id/tweets': {'limit': 225,
  'remaining': 225,
  'reset': 1624566596},
  '/labs/2/reports': {'limit': 5, 'remaining': 5, 'reset': 1624566596},
  '/labs/:version/tweets/stream/filter/rules&POST': {'limit': 450,
  'remaining': 450,
  'reset': 1624566596},
  '/labs/:version/tweets/stream/sample': {'limit': 50,
  'remaining': 50,
  'reset': 1624566596},
  '/labs/:version/users/by/username/:handle/tweets': {'limit': 225,
  'remaining': 225,
  'reset': 1624566596},
  '/labs/:version/tweets/metrics/private': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/labs/:version/tweets/stream/filter/rules/:instance_name': {'limit': 450,
  'remaining': 450,
  'reset': 1624566596},
  '/labs/:version/tweets/*': {'limit': 900,
  'remaining': 900,
  'reset': 1624566596},
  '/labs/:version/users/*': {'limit': 900,
  'remaining': 900,
  'reset': 1624566596},
  '/labs/:version/tweets/stream/filter/:instance_name': {'limit': 50,
  'remaining': 50,
  'reset': 1624566596},
  '/labs/:version/tweets/stream/filter/rules/': {'limit': 450,
  'remaining': 450,

```

```

    'reset': 1624566596},
    '/labs/:version/tweets/stream/compliance': {'limit': 500,
    'remaining': 500,
    'reset': 1624566596},
    '/labs/:version/tweets/search': {'limit': 225,
    'remaining': 225,
    'reset': 1624566596}},
    'i': {'/i/config': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
    '/i/tfb/v1/smb/web/:account_id/payment/save': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596}},
    'tweet_prompts': {'/tweet_prompts/report_interaction': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
    '/tweet_prompts/show': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596}},
    'moments': {'/moments/statuses/update': {'limit': 5,
    'remaining': 5,
    'reset': 1624566596},
    '/moments/create': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
    '/moments/permissions': {'limit': 300,
    'remaining': 300,
    'reset': 1624566596}},
    'limiter_scalding_report_creation': {'/limiter_scalding_report_creation':
{'limit': 4500,
    'remaining': 4500,
    'reset': 1624566596}},
    'fleets': {'/fleets/:version/mutes/create': {'limit': 100,
    'remaining': 100,
    'reset': 1624566596},
    '/fleets/:version/viewers': {'limit': 100,
    'remaining': 100,
    'reset': 1624566596},
    '/fleets/:version/delete': {'limit': 50,
    'remaining': 50,
    'reset': 1624566596},
    '/fleets/:version/avatar_content': {'limit': 100,
    'remaining': 100,
    'reset': 1624566596},
    '/fleets/:version/create': {'limit': 50,
    'remaining': 50,
    'reset': 1624566596},
    '/fleets/:version/user_fleets': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
    '/fleets/:version/fleetline': {'limit': 100,

```

```

    'remaining': 100,
    'reset': 1624566596},
  '/fleets/:version/track_events': {'limit': 100,
    'remaining': 100,
    'reset': 1624566596},
  '/fleets/:version/update': {'limit': 50,
    'remaining': 50,
    'reset': 1624566596},
  '/fleets/:version/fleet_threads': {'limit': 1000,
    'remaining': 1000,
    'reset': 1624566596},
  '/fleets/:version/mutes/list': {'limit': 100,
    'remaining': 100,
    'reset': 1624566596},
  '/fleets/:version/mutes/destroy': {'limit': 100,
    'remaining': 100,
    'reset': 1624566596},
  '/fleets/:version/home_timeline': {'limit': 100,
    'remaining': 100,
    'reset': 1624566596},
  '/fleets/:version/feedback/create': {'limit': 500,
    'remaining': 500,
    'reset': 1624566596},
  '/fleets/:version/mark_read': {'limit': 1000,
    'remaining': 1000,
    'reset': 1624566596}},
  'help': {'/help/tos': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
    '/help/configuration': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
    '/help/settings': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
    '/help/privacy': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
    '/help/languages': {'limit': 15, 'remaining': 15, 'reset': 1624566596}},
  'feedback': {'/feedback/show/:id': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
    '/feedback/events': {'limit': 1000,
    'remaining': 1000,
    'reset': 1624566596}},
  'business_experience': {'/business_experience/dashboard_settings/destroy':
{'limit': 450,
    'remaining': 450,
    'reset': 1624566596},
    '/business_experience/dashboard_features': {'limit': 450,
    'remaining': 450,
    'reset': 1624566596},
    '/business_experience/keywords': {'limit': 450,
    'remaining': 450,
    'reset': 1624566596},

```

```

'/business_experience/dashboard_settings/update': {'limit': 450,
'remaining': 450,
'reset': 1624566596},
'/business_experience/dashboard_settings/show': {'limit': 450,
'remaining': 450,
'reset': 1624566596}},
'graphql&POST': {'/graphql&POST': {'limit': 15,
'remaining': 15,
'reset': 1624566596}},
'friends': {'/friends/following/ids': {'limit': 15,
'remaining': 15,
'reset': 1624566596},
'/friends/following/list': {'limit': 15,
'remaining': 15,
'reset': 1624566596},
'/friends/list': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
'/friends/ids': {'limit': 15, 'remaining': 15, 'reset': 1624566596}},
'sandbox': {'/sandbox/account_activity/webhooks/:id/subscriptions': {'limit':
500,
'remaining': 500,
'reset': 1624566596}},
'drafts': {'/drafts/statuses/update': {'limit': 450,
'remaining': 450,
'reset': 1624566596},
'/drafts/statuses/destroy': {'limit': 450,
'remaining': 450,
'reset': 1624566596},
'/drafts/statuses/ids': {'limit': 450,
'remaining': 450,
'reset': 1624566596},
'/drafts/statuses/list': {'limit': 450,
'remaining': 450,
'reset': 1624566596},
'/drafts/statuses/show': {'limit': 450,
'remaining': 450,
'reset': 1624566596},
'/drafts/statuses/create': {'limit': 450,
'remaining': 450,
'reset': 1624566596}},
'direct_messages': {'/direct_messages/sent': {'limit': 300,
'remaining': 300,
'reset': 1624566596},
'/direct_messages/broadcasts/list': {'limit': 60,
'remaining': 60,
'reset': 1624566596},
'/direct_messages/subscribers/lists/members/show': {'limit': 1000,
'remaining': 1000,

```

```

    'reset': 1624566596},
  '/direct_messages/mark_read': {'limit': 1000,
    'remaining': 1000,
    'reset': 1624566596},
  '/direct_messages/subscribers/ids': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
  '/direct_messages/sent_and_received': {'limit': 300,
    'remaining': 300,
    'reset': 1624566596},
  '/direct_messages/broadcasts/statuses/list': {'limit': 60,
    'remaining': 60,
    'reset': 1624566596},
  '/direct_messages': {'limit': 300, 'remaining': 300, 'reset': 1624566596},
  '/direct_messages/subscribers/lists/members/ids': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
  '/direct_messages/subscribers/show': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
  '/direct_messages/broadcasts/show': {'limit': 60,
    'remaining': 60,
    'reset': 1624566596},
  '/direct_messages/broadcasts/statuses/show': {'limit': 60,
    'remaining': 60,
    'reset': 1624566596},
  '/direct_messages/subscribers/lists/list': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
  '/direct_messages/show': {'limit': 300,
    'remaining': 300,
    'reset': 1624566596},
  '/direct_messages/subscribers/lists/show': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
  '/direct_messages/events/list': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596},
  '/direct_messages/events/show': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596}},
  'media': {'/media/upload': {'limit': 500,
    'remaining': 500,
    'reset': 1624566596}},
  'traffic': {'/traffic/map': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596}},

```

```

'strato': {'/strato/column/None/:id/cms/*': {'limit': 150,
  'remaining': 150,
  'reset': 1624566596}},
'account_activity': {'/account_activity/all/webhooks': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/account_activity/all/:instance_name/subscriptions': {'limit': 500,
  'remaining': 500,
  'reset': 1624566596},
  '/account_activity/direct_messages/webhooks': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/account_activity/webhooks/:id/subscriptions/direct_messages/list':
{'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/account_activity/webhooks/:id/subscriptions/all': {'limit': 500,
  'remaining': 500,
  'reset': 1624566596},
  '/account_activity/direct_messages/:instance_name/webhooks': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/account_activity/webhooks/:id/subscriptions/all/list': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/account_activity/webhooks/:id/subscriptions/direct_messages': {'limit':
500,
  'remaining': 500,
  'reset': 1624566596},
  '/account_activity/webhooks': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/account_activity/direct_messages/:instance_name/subscriptions': {'limit':
15,
  'remaining': 15,
  'reset': 1624566596},
  '/account_activity/webhooks/:id/subscriptions': {'limit': 500,
  'remaining': 500,
  'reset': 1624566596},
  '/account_activity/all/:instance_name/webhooks': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596}},
'account': {'/account/login_verification_enrollment': {'limit': 15,
  'remaining': 15,
  'reset': 1624566596},
  '/account/update_profile': {'limit': 15,
  'remaining': 15,

```

```

    'reset': 1624566596},
  '/account/authenticate_web_view': {'limit': 50,
    'remaining': 50,
    'reset': 1624566596},
  '/account/verify_credentials': {'limit': 75,
    'remaining': 74,
    'reset': 1624566586},
  '/account/settings': {'limit': 15, 'remaining': 15, 'reset': 1624566596},
  '/account/change_password': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596}},
  'safety': {'/safety/detection_feedback': {'limit': 450000,
    'remaining': 450000,
    'reset': 1624566596}},
  'favorites': {'/favorites/list': {'limit': 75,
    'remaining': 75,
    'reset': 1624566596}},
  'device': {'/device/token': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596}},
  'tweets': {'/tweets/search/all': {'limit': 900,
    'remaining': 900,
    'reset': 1624566596},
    '/tweets/search/stream/rules': {'limit': 450,
    'remaining': 450,
    'reset': 1624566596},
    '/tweets/search/recent': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596},
    '/tweets/sample/stream': {'limit': 50,
    'remaining': 50,
    'reset': 1624566596},
    '/tweets/': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
    '/tweets/counts/all': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
    '/tweets/search/stream': {'limit': 50,
    'remaining': 50,
    'reset': 1624566596},
    '/tweets/search/:product/:label': {'limit': 1800,
    'remaining': 1800,
    'reset': 1624566596},
    '/tweets/search/stream/rules/validation&POST': {'limit': 450,
    'remaining': 450,
    'reset': 1624566596},
    '/tweets/search/:product/:instance/counts': {'limit': 900,
    'remaining': 900,
    'reset': 1624566596},
    '/tweets/:tweet_id/liking_users': {'limit': 75,

```

```

    'remaining': 75,
    'reset': 1624566596},
    '/tweets/search/stream/rules&DELETE': {'limit': 450,
    'remaining': 450,
    'reset': 1624566596},
    '/tweets/:id': {'limit': 900, 'remaining': 900, 'reset': 1624566596},
    '/tweets/counts/recent': {'limit': 900,
    'remaining': 900,
    'reset': 1624566596},
    '/tweets/search/stream/rules&POST': {'limit': 450,
    'remaining': 450,
    'reset': 1624566596},
    '/tweets/:id/hidden&PUT': {'limit': 50,
    'remaining': 50,
    'reset': 1624566596}},
    'saved_searches': {'/saved_searches/destroy/:id': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596},
    '/saved_searches/show/:id': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596},
    '/saved_searches/list': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596}},
    'oauth': {'/oauth/revoke': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596},
    '/oauth/invalidate_token': {'limit': 450,
    'remaining': 450,
    'reset': 1624566596},
    '/oauth/revoke_html': {'limit': 15, 'remaining': 15, 'reset': 1624566596}},
    'search': {'/search/tweets': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596}},
    'trends': {'/trends/closest': {'limit': 75,
    'remaining': 75,
    'reset': 1624566596},
    '/trends/available': {'limit': 75, 'remaining': 75, 'reset': 1624566596},
    '/trends/place': {'limit': 75, 'remaining': 75, 'reset': 1624566596}},
    'live_pipeline': {'/live_pipeline/events': {'limit': 180,
    'remaining': 180,
    'reset': 1624566596}},
    'graphql': {'/graphql': {'limit': 15,
    'remaining': 15,
    'reset': 1624566596}}}}

```

Now we setup dates. We need to setup today and yesterday.


```
[24]: today = datetime.date.today()
      since= today - datetime.timedelta(days=7)
      until= today
      until, since
      # (datetime.date(2021, 6, 7), datetime.date(2021, 6, 6))
```

```
[24]: (datetime.date(2021, 6, 25), datetime.date(2021, 6, 18))
```

```
[67]: logger.debug(f"full_text: '{until, since}')
```

```
2021-06-25 01:50:41,724 - DEBUG - full_text: '(datetime.date(2021, 6, 25),
datetime.date(2021, 6, 18))'
```

We search for tweets on Twitter by using the `Cursor()` function. We pass the `api.search` parameter to the cursor, as well as the query string, which is specified through the `q` parameter of the cursor. The query string can receive many parameters, such as the following (not mandatory) ones: `* from:` - to specify a specific Twitter user profile `* since:` - to specify the beginning date of search `* until:` - to specify the ending date of search The cursor can also receive other parameters, such as the language and the `tweet_mode`. If `tweet_mode='extended'`, all the text of the tweet is returned, otherwise only the first 140 characters.

```
[ ]: # # example
      # code tweets = tweepy.Cursor(api.search, tweet_mode='extended')
      # for tweet in tweets:
      #     content = tweet.full_text
```

```
[ ]: # tweets_list = tweepy.Cursor(api.search, q="#Covid-19 since:" +
      ↪str(yesterday)+ " until:" + str(today),tweet_mode='extended', lang='en').
      ↪items()
```

```
[ ]: # tweets_list = tweepy.Cursor(api.search, q=f"#Covid-19 since:{str(yesterday)}
      ↪until:{str(today)}",tweet_mode='extended', lang='en').items()
```

```
[29]: # tweets_list = tweepy.Cursor(api.search, q=['astrazeneca', 'pfizer'],since=
      ↪str(since), until=str(until),tweet_mode='extended', lang='en').items()

      # Greek Language = el
      # tweets_list = tweepy.Cursor(api.search, q=['coffee island'],since=
      ↪str(since), until=str(until),tweet_mode='extended', lang='el').items()

      # English Language = en
      tweets_list = tweepy.Cursor(api.search, q=['coffee island OR
      ↪CoffeeIsland'],since= str(since), until=str(until),tweet_mode='extended',
      ↪lang='en').items()
```

Now we loop across the `tweets_list`, and, for each tweet, we extract the text, the creation date, the number of retweets and the favourite count. We store every tweet into a list, called `output`.

```
[26]: import time
seconds = 5
start = time.time()
time.sleep(seconds)
end = time.time()
logger.info(f"elapsed_time: '{end - start}'")
```

2021-06-25 03:00:56,957 - INFO - elapsed_time: '5.004919528961182'

2 TEST

```
[17]: # tweets_list2 = tweepy.Cursor(api.search, q=['pfizer', 'astrazeneca'], since=
↳ str(since), until=str(until), tweet_mode='extended', lang='en').items(2)

import time
start = time.time()
output = []
for tweet in tweets_list:
    # text = tweet._json["full_text"]
    # print(text)
    # https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/
↳ api-reference/get-search-tweets
    # "geo": null, "coordinates": null, "place": null, "contributors": null,
    # "is_quote_status": false, "retweet_count": 988, "favorite_count": 3875,
    # "favorited": false, "retweeted": false, "possibly_sensitive": false, "lang":
↳ "en"
    # https://developer.twitter.com/en/docs/twitter-ids
    logger.info(f"tweet_id_str: {'-'*30}")
    logger.info(f"created_at: {tweet.created_at}")
    logger.info(f"full_text: {tweet._json['full_text']}")
    logger.info(f"tweet_id: {tweet.id}")
    logger.info(f"tweet_id_str: {tweet.id_str}")
    logger.info(f"user: {tweet._json['user']['name']}")
    logger.info(f"user_id: {tweet._json['user']['id']}")
    # favourite_count = tweet.favorite_count
    # retweet_count = tweet.retweet_count
    # created_at = tweet.created_at

#     line = {'text' : text, 'favourite_count' : favourite_count,
↳ 'retweet_count' : retweet_count, 'created_at' : created_at}
#     output.append(line)
#     logger.info(f"Append list length : { len(output)}")
# end = time.time()
```

```
# logger.info(f"elapsed_time: '{end - start}')
```

```
2021-06-24 23:18:06,420 - DEBUG - PARAMS: {'q': b"['coffee island']", 'since':  
b'2021-06-22', 'until': b'2021-06-24', 'tweet_mode': b'extended', 'lang': b'gr'}  
2021-06-24 23:18:06,428 - DEBUG - Signing request <PreparedRequest [GET]> using  
client <Client client_key=pw0ihLFxH3nwDrd4HBd7pqUrc, client_secret=****,  
resource_owner_key=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz,  
resource_owner_secret=****, signature_method=HMAC-SHA1,  
signature_type=AUTH_HEADER, callback_uri=None, rsa_key=None, verifier=None,  
realm=None, encoding=utf-8, decoding=None, nonce=None, timestamp=None>  
2021-06-24 23:18:06,429 - DEBUG - Including body in call to sign: False  
2021-06-24 23:18:06,442 - DEBUG - Collected params: [('q', "['coffee island']"),  
('since', '2021-06-22'), ('until', '2021-06-24'), ('tweet_mode', 'extended'),  
('lang', 'gr'), ('oauth_nonce', '89097120203846973971624565886'),  
('oauth_timestamp', '1624565886'), ('oauth_version', '1.0'),  
('oauth_signature_method', 'HMAC-SHA1'), ('oauth_consumer_key',  
'pw0ihLFxH3nwDrd4HBd7pqUrc'), ('oauth_token',  
'1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz')]  
2021-06-24 23:18:06,448 - DEBUG - Normalized params: lang=gr&oauth_consumer_key=  
pw0ihLFxH3nwDrd4HBd7pqUrc&oauth_nonce=89097120203846973971624565886&oauth_signat  
ure_method=HMAC-SHA1&oauth_timestamp=1624565886&oauth_token=1360011857969479682-  
iLrxBULqdtExwkqiN9iZsHYDXIFTZz&oauth_version=1.0&q=%5B%27coffee%20island%27%5D&s  
ince=2021-06-22&tweet_mode=extended&until=2021-06-24  
2021-06-24 23:18:06,449 - DEBUG - Normalized URI:  
https://api.twitter.com/1.1/search/tweets.json  
2021-06-24 23:18:06,450 - DEBUG - Signing: signature base string: GET&https%3A%2  
F%2Fapi.twitter.com%2F1.1%2Fsearch%2Ftweets.json&lang%3Dgr%26oauth_consumer_key%  
3Dpw0ihLFxH3nwDrd4HBd7pqUrc%26oauth_nonce%3D89097120203846973971624565886%26aut  
h_signature_method%3DHMAC-SHA1%26oauth_timestamp%3D1624565886%26oauth_token%3D13  
60011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz%26oauth_version%3D1.0%26q%3D%25  
5B%2527coffee%2520island%2527%255D%26since%3D2021-06-22%26tweet_mode%3Dextended%  
26until%3D2021-06-24  
2021-06-24 23:18:06,457 - DEBUG - Signature: 0aK/4MeMjbf3kLgJW6Twy4Veu10=  
2021-06-24 23:18:06,460 - DEBUG - Updated url: https://api.twitter.com/1.1/searc  
h/tweets.json?q=%5B%27coffee+island%27%5D&since=2021-06-22&until=2021-06-24&tweet  
_mode=extended&lang=gr  
2021-06-24 23:18:06,461 - DEBUG - Updated headers: {'Host': 'api.twitter.com',  
'Authorization': 'OAuth oauth_nonce="89097120203846973971624565886",  
oauth_timestamp="1624565886", oauth_version="1.0", oauth_signature_method="HMAC-  
SHA1", oauth_consumer_key="pw0ihLFxH3nwDrd4HBd7pqUrc",  
oauth_token="1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz",  
oauth_signature="0aK%2F4MeMjbf3kLgJW6Twy4Veu10%3D"' }  
2021-06-24 23:18:06,462 - DEBUG - Updated body: None  
2021-06-24 23:18:06,473 - DEBUG - Starting new HTTPS connection (1):  
api.twitter.com:443  
2021-06-24 23:18:07,527 - DEBUG - https://api.twitter.com:443 "GET /1.1/search/t  
weets.json?q=%5B%27coffee+island%27%5D&since=2021-06-22&until=2021-06-24&tweet_m
```

```
ode=extended&lang=gr HTTP/1.1" 200 382
2021-06-24 23:18:07,535 - DEBUG - PARAMS: {'create': b'True'}
```

```
[30]: import time
start = time.time()
output = []
for tweet in tweets_list:
    text = tweet._json["full_text"]
    #print(text)
    # https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/
    ↪ api-reference/get-search-tweets
    # "geo": null, "coordinates": null, "place": null, "contributors": null,
    # "is_quote_status": false, "retweet_count": 988, "favorite_count": 3875,
    # "favorited": false, "retweeted": false, "possibly_sensitive": false, "lang":
    ↪ "en"
    logger.debug(f"full_text: '{text}'")
    favourite_count = tweet.favorite_count
    retweet_count = tweet.retweet_count
    created_at = tweet.created_at

    line = {'text' : text, 'favourite_count' : favourite_count, 'retweet_count':
    ↪ retweet_count, 'created_at' : created_at}
    output.append(line)
    logger.info(f"Append list length : { len(output)}")
end = time.time()
logger.info(f"elapsed_time: '{end - start}'")
```

```
full_text: '@jessicahodlr DM use to reserve Satoshi's Selection coffee. its
Bitcoin coffee from the enchanted island of Puerto Rico    Available in 2
weeks for shipping'
```

```
2021-06-25 03:18:30,485 - INFO - Append list length : 604
```

```
2021-06-25 03:18:30,486 - DEBUG - full_text: 'Colourful handmade chocolates
being boxed up from Arlene's Island Artisan Chocolate
```

```
Available here    https://t.co/yT7USQMTcx https://t.co/Bzs0iRjgoh'
```

```
2021-06-25 03:18:30,486 - INFO - Append list length : 605
```

```
2021-06-25 03:18:30,487 - DEBUG - full_text: '90% of socializing in Rhode Island
involves iced coffee'
```

```
2021-06-25 03:18:30,488 - INFO - Append list length : 606
```

```
2021-06-25 03:18:30,489 - DEBUG - full_text: '
[Friends+]Treasure Hunt Island #VRCinHere 21:37:10'
```

```
2021-06-25 03:18:30,490 - INFO - Append list length : 607
```

```
2021-06-25 03:18:30,490 - DEBUG - full_text: '@island_iverson This tweet made me
spill hot coffee on my d region. Good stuff.'
```

```
2021-06-25 03:18:30,491 - INFO - Append list length : 608
```

```
2021-06-25 03:18:30,491 - DEBUG - full_text: 'RT @CatherineSher10: A statue of
```

Hawaiian King Kamehameha the Great, Wailoa State Park in #Hilo #Hawaii on the Big Island.
 #BigIsland #CL...'
 2021-06-25 03:18:30,492 - INFO - Append list length : 609
 2021-06-25 03:18:30,492 - DEBUG - full_text: 'RT @cathsherman: Afternoon at the Eighteenth Hole, Kiawah Island, South Carolina #KiawahIsland #Golf #SouthCarolina Clubhouse #SOLD...'
 2021-06-25 03:18:30,493 - INFO - Append list length : 610
 2021-06-25 03:18:30,493 - DEBUG - full_text: 'RT @cathsherman: Afternoon at the Eighteenth Hole, Kiawah Island, South Carolina #KiawahIsland #Golf #SouthCarolina Clubhouse #SOLD...'
 2021-06-25 03:18:30,494 - INFO - Append list length : 611
 2021-06-25 03:18:30,494 - DEBUG - full_text: 'Good morning from Hilton Head Island, SC! Coffee and a great book by @JulieDiCaro on the balcony. https://t.co/Z8R2VdSubX'
 2021-06-25 03:18:30,495 - INFO - Append list length : 612
 2021-06-25 03:18:30,495 - DEBUG - full_text: '@LorinOnTV Found a great Long Island coffee company. https://t.co/GMdm06ndPv'
 2021-06-25 03:18:30,496 - INFO - Append list length : 613
 2021-06-25 03:18:30,496 - DEBUG - full_text: 'RT @visitcyprus: Morning coffee by the sea at charming locations, here, at your favourite island destination! Where are you having your cof...'
 2021-06-25 03:18:30,496 - INFO - Append list length : 614
 2021-06-25 03:18:30,497 - DEBUG - PARAMS: {'max_id': b'1405849986219261951', 'q': b"['coffee island OR CoffeeIsland']", 'since': b'2021-06-18', 'until': b'2021-06-25', 'tweet_mode': b'extended', 'lang': b'en'}
 2021-06-25 03:18:30,498 - DEBUG - Signing request <PreparedRequest [GET]> using client <Client client_key=pw0ihLFxH3nwDrd4HBd7pqUrc, client_secret=****, resource_owner_key=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz, resource_owner_secret=****, signature_method=HMAC-SHA1, signature_type=AUTH_HEADER, callback_uri=None, rsa_key=None, verifier=None, realm=None, encoding=utf-8, decoding=None, nonce=None, timestamp=None>
 2021-06-25 03:18:30,499 - DEBUG - Including body in call to sign: False
 2021-06-25 03:18:30,500 - DEBUG - Collected params: [('max_id', '1405849986219261951'), ('q', "['coffee island OR CoffeeIsland']"), ('since', '2021-06-18'), ('until', '2021-06-25'), ('tweet_mode', 'extended'), ('lang', 'en'), ('oauth_nonce', '147693561906163895201624580310'), ('oauth_timestamp', '1624580310'), ('oauth_version', '1.0'), ('oauth_signature_method', 'HMAC-SHA1'), ('oauth_consumer_key', 'pw0ihLFxH3nwDrd4HBd7pqUrc'), ('oauth_token', '1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz')]
 2021-06-25 03:18:30,501 - DEBUG - Normalized params: lang=en&max_id=1405849986219261951&oauth_consumer_key=pw0ihLFxH3nwDrd4HBd7pqUrc&oauth_nonce=147693561906163895201624580310&oauth_signature_method=HMAC-SHA1&oauth_timestamp=1624580310&oauth_token=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz&oauth_version=1.0&q=%5B%27coffee%20island%20OR%20CoffeeIsland%27%5D&since=2021-06-18&tweet_mode=exten

ded&until=2021-06-25

2021-06-25 03:18:30,502 - DEBUG - Normalized URI:
https://api.twitter.com/1.1/search/tweets.json

2021-06-25 03:18:30,503 - DEBUG - Signing: signature base string: GET&https%3A%2F%2Fapi.twitter.com%2F1.1%2Fsearch%2Ftweets.json&lang%3Den%26max_id%3D1405849986219261951%26oauth_consumer_key%3Dpw0ihLFxH3nwDrd4HBd7pqUrc%26oauth_nonce%3D147693561906163895201624580310%26oauth_signature_method%3DHMAC-SHA1%26oauth_timestamp%3D1624580310%26oauth_token%3D1360011857969479682-iLrxBUlqdtExwkqiN9iZsHYDXIFTZz%26oauth_version%3D1.0%26q%3D%255B%2527coffee%2520island%2520OR%2520CoffeeIsland%2527%255D%26since%3D2021-06-18%26tweet_mode%3Dextended%26until%3D2021-06-25

2021-06-25 03:18:30,503 - DEBUG - Signature: /CZb47lwHs3NFdMybfdok4fUjKc=

2021-06-25 03:18:30,505 - DEBUG - Updated url: https://api.twitter.com/1.1/search/tweets.json?max_id=1405849986219261951&q=%5B%27coffee+island+OR+CoffeeIsland%27%5D&since=2021-06-18&until=2021-06-25&tweet_mode=extended&lang=en

2021-06-25 03:18:30,508 - DEBUG - Updated headers: {'Host': 'api.twitter.com', 'Cookie': 'guest_id=v1%3A162458023776513966; personalization_id="v1_xu/E8D37HQfaLp3okgVYng=="', 'Authorization': 'OAuth oauth_nonce="147693561906163895201624580310", oauth_timestamp="1624580310", oauth_version="1.0", oauth_signature_method="HMAC-SHA1", oauth_consumer_key="pw0ihLFxH3nwDrd4HBd7pqUrc", oauth_token="1360011857969479682-iLrxBUlqdtExwkqiN9iZsHYDXIFTZz", oauth_signature="%2FCZb47lwHs3NFdMybfdok4fUjKc%3D"%}'}

2021-06-25 03:18:30,509 - DEBUG - Updated body: None

2021-06-25 03:18:30,512 - DEBUG - Starting new HTTPS connection (1):
api.twitter.com:443

2021-06-25 03:18:31,437 - DEBUG - https://api.twitter.com:443 "GET /1.1/search/tweets.json?max_id=1405849986219261951&q=%5B%27coffee+island+OR+CoffeeIsland%27%5D&since=2021-06-18&until=2021-06-25&tweet_mode=extended&lang=en HTTP/1.1" 200 80282

2021-06-25 03:18:31,531 - DEBUG - PARAMS: {'create': b'True'}

2021-06-25 03:18:31,538 - DEBUG - full_text: 'Here's your OTD trivia for 6/18... Join us tonight at Amelia Island Coffee, and have a great Friday! #trivia30 #wakeupyourbrain #OnThisDay #Monterey #California #JimiHendrix #JanisJoplin #TheWho #OtisRedding #USSR #FastAndFurious #Spain #CarolKane #KurtBrowning #BlakeShelton https://t.co/K6IMU3DNGy'

2021-06-25 03:18:31,543 - INFO - Append list length : 615

2021-06-25 03:18:31,548 - DEBUG - full_text: 'Join us tonight for an evening of TRIVIA:30 at Amelia Island Coffee! The fun starts at 6:00... Free answer is *Ensenada* See you there! #whattimeisit #TRIVIA30 #wakeupyourbrain #trivia #fridaynight #AmeliaIsland #FernandinaBeach #LoveAmelia #Fernandina #trivia #fridayfun https://t.co/T2JENWOhlT'

2021-06-25 03:18:31,550 - INFO - Append list length : 616

2021-06-25 03:18:31,551 - DEBUG - full_text: 'Good Morning, I've got the coffee on for you.'

2021-06-25 03:18:31,552 - INFO - Append list length : 617

2021-06-25 03:18:31,556 - DEBUG - full_text: 'RT @TheW0lfpup: Coffee island'

2021-06-25 03:18:31,557 - INFO - Append list length : 618

2021-06-25 03:18:31,570 - DEBUG - full_text: 'RT @GySgt2: @justmeagain34 Good

morning my friend, it is 73 degrees here on Parris island, the coffee is strong and black, the spiders wh...

2021-06-25 03:18:31,572 - INFO - Append list length : 619

2021-06-25 03:18:31,572 - DEBUG - full_text: '@justmeagain34 Good morning my friend, it is 73 degrees here on Parris island, the coffee is strong and black, the spiders where All over me this morning leaving my house I guess I am just going to have to get a flame thrower for the armory be fir I go home tonight, TGIF '

2021-06-25 03:18:31,573 - INFO - Append list length : 620

2021-06-25 03:18:31,573 - DEBUG - full_text: '@mochis_island_ I'm watching tv and kissing coffee but I should be working'

2021-06-25 03:18:31,574 - INFO - Append list length : 621

2021-06-25 03:18:31,574 - DEBUG - full_text: '@SarkySage coffee followed by Russell Island tap water vintage 2021'

2021-06-25 03:18:31,574 - INFO - Append list length : 622

2021-06-25 03:18:31,575 - DEBUG - full_text: 'Bonjour! Let's make a sandwich to start a happy day! Open the fridge, and grab two pieces of toasts. Take ham as the protein, tomato and spinach as veggies. Finally, add some thousand island dressing and cheddar cheese. Have it with a cup of black coffee. Enjoy your morning! <https://t.co/lzew0c0HhG>'

2021-06-25 03:18:31,577 - INFO - Append list length : 623

2021-06-25 03:18:31,578 - DEBUG - full_text: 'Read Treasure Island and enjoy some kombucha out of a irish coffee glass.'

2021-06-25 03:18:31,581 - INFO - Append list length : 624

2021-06-25 03:18:31,582 - DEBUG - full_text: 'Magical island #Howth View from #Blackrock #Dublin

Morning dip & coffee with pals

#Reset #VitaminSea

@janeconan @TheGinaLondon <https://t.co/LQd53AoUz4>'

2021-06-25 03:18:31,583 - INFO - Append list length : 625

2021-06-25 03:18:31,584 - DEBUG - full_text: 'RT @TheW0lfpup: Coffee island'

2021-06-25 03:18:31,589 - INFO - Append list length : 626

2021-06-25 03:18:31,590 - DEBUG - full_text: 'RT @TheW0lfpup: Coffee island <https://t.co/52WId63NXo>'

2021-06-25 03:18:31,590 - INFO - Append list length : 627

2021-06-25 03:18:31,591 - DEBUG - full_text: 'RT @TheW0lfpup: Coffee island'

2021-06-25 03:18:31,591 - INFO - Append list length : 628

2021-06-25 03:18:31,593 - DEBUG - full_text: 'RT @TheW0lfpup: Coffee island'

2021-06-25 03:18:31,594 - INFO - Append list length : 629

2021-06-25 03:18:31,601 - DEBUG - PARAMS: {'max_id': b'1405771752257449985', 'q': b"['coffee island OR CoffeeIsland']", 'since': b'2021-06-18', 'until': b'2021-06-25', 'tweet_mode': b'extended', 'lang': b'en'}

2021-06-25 03:18:31,621 - DEBUG - Signing request <PreparedRequest [GET]> using client <Client client_key=pw0ihLFxH3nwDrd4HBd7pqUrc, client_secret=****, resource_owner_key=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz, resource_owner_secret=****, signature_method=HMAC-SHA1,

```

signature_type=AUTH_HEADER, callback_uri=None, rsa_key=None, verifier=None,
realm=None, encoding=utf-8, decoding=None, nonce=None, timestamp=None>
2021-06-25 03:18:31,622 - DEBUG - Including body in call to sign: False
2021-06-25 03:18:31,624 - DEBUG - Collected params: [('max_id',
'1405771752257449985'), ('q', "['coffee island OR CoffeeIsland']"), ('since',
'2021-06-18'), ('until', '2021-06-25'), ('tweet_mode', 'extended'), ('lang',
'en'), ('oauth_nonce', '80464792822679196621624580311'), ('oauth_timestamp',
'1624580311'), ('oauth_version', '1.0'), ('oauth_signature_method', 'HMAC-
SHA1'), ('oauth_consumer_key', 'pw0ihLFxH3nwDrd4HBd7pqUrc'), ('oauth_token',
'1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz')]
2021-06-25 03:18:31,626 - DEBUG - Normalized params: lang=en&max_id=140577175225
7449985&oauth_consumer_key=pw0ihLFxH3nwDrd4HBd7pqUrc&oauth_nonce=804647928226791
96621624580311&oauth_signature_method=HMAC-SHA1&oauth_timestamp=1624580311&oauth
_token=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz&oauth_version=1.0&q=%5
B%27coffee%20island%20OR%20CoffeeIsland%27%5D&since=2021-06-18&tweet_mode=extend
ed&until=2021-06-25
2021-06-25 03:18:31,628 - DEBUG - Normalized URI:
https://api.twitter.com/1.1/search/tweets.json
2021-06-25 03:18:31,629 - DEBUG - Signing: signature base string: GET&https%3A%2
F%2Fapi.twitter.com%2F1.1%2Fsearch%2Ftweets.json&lang%3Den%26max_id%3D1405771752
257449985%26oauth_consumer_key%3Dpw0ihLFxH3nwDrd4HBd7pqUrc%26oauth_nonce%3D80464
792822679196621624580311%26oauth_signature_method%3DHMAC-SHA1%26oauth_timestamp%
3D1624580311%26oauth_token%3D1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz%
26oauth_version%3D1.0%26q%3D%255B%2527coffee%2520island%2520OR%2520CoffeeIsland%
2527%255D%26since%3D2021-06-18%26tweet_mode%3Dextended%26until%3D2021-06-25
2021-06-25 03:18:31,630 - DEBUG - Signature: GwKp/JTMZq5qEQry7n7ufpU5tp8=
2021-06-25 03:18:31,632 - DEBUG - Updated url: https://api.twitter.com/1.1/searc
h/tweets.json?max_id=1405771752257449985&q=%5B%27coffee+island+OR+CoffeeIsland%2
7%5D&since=2021-06-18&until=2021-06-25&tweet_mode=extended&lang=en
2021-06-25 03:18:31,634 - DEBUG - Updated headers: {'Host': 'api.twitter.com',
'Cookie': 'guest_id=v1%3A162458023776513966;
personalization_id=v1_xu/E8D37HQfaLp3okgVYng==', 'Authorization': 'OAuth
oauth_nonce="80464792822679196621624580311", oauth_timestamp="1624580311",
oauth_version="1.0", oauth_signature_method="HMAC-SHA1",
oauth_consumer_key="pw0ihLFxH3nwDrd4HBd7pqUrc",
oauth_token="1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz",
oauth_signature="GwKp%2FJTMZq5qEQry7n7ufpU5tp8%3D"}
2021-06-25 03:18:31,635 - DEBUG - Updated body: None
2021-06-25 03:18:31,641 - DEBUG - Starting new HTTPS connection (1):
api.twitter.com:443
2021-06-25 03:18:33,361 - DEBUG - https://api.twitter.com:443 "GET /1.1/search/t
weets.json?max_id=1405771752257449985&q=%5B%27coffee+island+OR+CoffeeIsland%27%5
D&since=2021-06-18&until=2021-06-25&tweet_mode=extended&lang=en HTTP/1.1" 200
80405
2021-06-25 03:18:33,473 - DEBUG - PARAMS: {'create': b'True'}
2021-06-25 03:18:33,478 - DEBUG - full_text: 'RT @TheW0lfup: Coffee island
https://t.co/52WId63NXo'
2021-06-25 03:18:33,487 - INFO - Append list length : 630

```


2021-06-25 03:18:33,488 - DEBUG - full_text: 'RT @visitcyprus: Morning coffee by the sea at charming locations, here, at your favourite island destination! Where are you having your cof...'

2021-06-25 03:18:33,489 - INFO - Append list length : 631

2021-06-25 03:18:33,491 - DEBUG - full_text: 'RT @TheW0lfpup: Coffee island'

2021-06-25 03:18:33,491 - INFO - Append list length : 632

2021-06-25 03:18:33,494 - DEBUG - full_text: 'Coffee Island
<https://t.co/yM2zbp9hp3>'

2021-06-25 03:18:33,502 - INFO - Append list length : 633

2021-06-25 03:18:33,510 - DEBUG - full_text: 'DDDaniel just supported @cretetravel on @buymeacoffee!'

@cretetravel is ... creating memorable experiences in lovely Crete Island! . You can support by buying a coffee here - <https://t.co/Qo8h8xTPAN>'

2021-06-25 03:18:33,513 - INFO - Append list length : 634

2021-06-25 03:18:33,515 - DEBUG - full_text: 'RT @TheW0lfpup: Coffee island'

2021-06-25 03:18:33,520 - INFO - Append list length : 635

2021-06-25 03:18:33,523 - DEBUG - full_text: '@SerpicoDet That's incredible! It would have been in Nanaimo, my family spoke very highly of you. If you ever end up on the island you're more than welcome over for coffee again. Thanks for confirming an old story!'

2021-06-25 03:18:33,527 - INFO - Append list length : 636

2021-06-25 03:18:33,528 - DEBUG - full_text: 'RT @TheW0lfpup: Coffee island'

2021-06-25 03:18:33,529 - INFO - Append list length : 637

2021-06-25 03:18:33,529 - DEBUG - full_text: 'RT @visitcyprus: Morning coffee by the sea at charming locations, here, at your favourite island destination! Where are you having your cof...'

2021-06-25 03:18:33,537 - INFO - Append list length : 638

2021-06-25 03:18:33,538 - DEBUG - full_text: 'Read Treasure Island and drink some chilled kombucha from a irish coffee glass.'

2021-06-25 03:18:33,540 - INFO - Append list length : 639

2021-06-25 03:18:33,541 - DEBUG - full_text: '@ARLESVENUSX the coffee ice cream slander here is disturbing. i am here to say you would not survive in rhode island '

2021-06-25 03:18:33,544 - INFO - Append list length : 640

2021-06-25 03:18:33,545 - DEBUG - full_text: '"She also loves to travel around the world and sip coffee along the way, so if we are ever able to connect in real life, I think we'd get along splendidly."

Truth, Dr. Hasty!

Thanks for a fun #bookfeature with #IslandCharm from @PelicanBookGrp.

<https://t.co/Ef0H61FEky>'

2021-06-25 03:18:33,560 - INFO - Append list length : 641

2021-06-25 03:18:33,567 - DEBUG - full_text: 'RT @lizwadsworth65: @quinncy Exile and Duke, a used bookstore/cafe specializing in historical romances with unhappy endings. It's located i...'

2021-06-25 03:18:33,569 - INFO - Append list length : 642

2021-06-25 03:18:33,570 - DEBUG - full_text: '@quinncy Exile and Duke, a used bookstore/cafe specializing in historical romances with unhappy endings. It's located in a refurbished island prison in which all the cells are private reading rooms with French press coffee makers and complimentary macarons.'

2021-06-25 03:18:33,570 - INFO - Append list length : 643

2021-06-25 03:18:33,571 - DEBUG - full_text: '@bananzers I wanna visit the big island and drink Kona coffee from the source '

2021-06-25 03:18:33,572 - INFO - Append list length : 644

2021-06-25 03:18:33,574 - DEBUG - PARAMS: {'max_id': b'1405681100459028483', 'q': b"['coffee island OR CoffeeIsland']", 'since': b'2021-06-18', 'until': b'2021-06-25', 'tweet_mode': b'extended', 'lang': b'en'}

2021-06-25 03:18:33,578 - DEBUG - Signing request <PreparedRequest [GET]> using client <Client client_key=pw0ihLFxH3nwDrd4HBd7pqUrc, client_secret=****, resource_owner_key=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz, resource_owner_secret=****, signature_method=HMAC-SHA1, signature_type=AUTH_HEADER, callback_uri=None, rsa_key=None, verifier=None, realm=None, encoding=utf-8, decoding=None, nonce=None, timestamp=None>

2021-06-25 03:18:33,579 - DEBUG - Including body in call to sign: False

2021-06-25 03:18:33,580 - DEBUG - Collected params: [('max_id', '1405681100459028483'), ('q', "['coffee island OR CoffeeIsland']"), ('since', '2021-06-18'), ('until', '2021-06-25'), ('tweet_mode', 'extended'), ('lang', 'en'), ('oauth_nonce', '113879326453028059981624580313'), ('oauth_timestamp', '1624580313'), ('oauth_version', '1.0'), ('oauth_signature_method', 'HMAC-SHA1'), ('oauth_consumer_key', 'pw0ihLFxH3nwDrd4HBd7pqUrc'), ('oauth_token', '1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz')]

2021-06-25 03:18:33,582 - DEBUG - Normalized params: lang=en&max_id=1405681100459028483&oauth_consumer_key=pw0ihLFxH3nwDrd4HBd7pqUrc&oauth_nonce=113879326453028059981624580313&oauth_signature_method=HMAC-SHA1&oauth_timestamp=1624580313&oauth_token=1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz&oauth_version=1.0&q=%5B%27coffee%20island%20OR%20CoffeeIsland%27%5D&since=2021-06-18&tweet_mode=extended&until=2021-06-25

2021-06-25 03:18:33,583 - DEBUG - Normalized URI: https://api.twitter.com/1.1/search/tweets.json

2021-06-25 03:18:33,584 - DEBUG - Signing: signature base string: GET&https%3A%2F%2Fapi.twitter.com%2F1.1%2Fsearch%2Ftweets.json&lang%3Den%26max_id%3D1405681100459028483%26oauth_consumer_key%3Dpw0ihLFxH3nwDrd4HBd7pqUrc%26oauth_nonce%3D113879326453028059981624580313%26oauth_signature_method%3DHMAC-SHA1%26oauth_timestamp%3D1624580313%26oauth_token%3D1360011857969479682-iLrxBULqdtExwkqiN9iZsHYDXIFTZz%26oauth_version%3D1.0%26q%3D%255B%2527coffee%2520island%2520OR%2520CoffeeIsland%2527%255D%26since%3D2021-06-18%26tweet_mode%3Dextended%26until%3D2021-06-25

2021-06-25 03:18:33,585 - DEBUG - Signature: oANPns3UzMayAeo0sh0Wx9E/naA=

2021-06-25 03:18:33,588 - DEBUG - Updated url: https://api.twitter.com/1.1/search/tweets.json?max_id=1405681100459028483&q=%5B%27coffee+island+OR+CoffeeIsland%27%5D&since=2021-06-18&until=2021-06-25&tweet_mode=extended&lang=en

2021-06-25 03:18:33,592 - DEBUG - Updated headers: {'Host': 'api.twitter.com', 'Cookie': 'guest_id=v1%3A162458023776513966; personalization_id="v1_xu/E8D37HQfaLp3okgVYng=="', 'Authorization': 'OAuth

```

oauth_nonce="113879326453028059981624580313", oauth_timestamp="1624580313",
oauth_version="1.0", oauth_signature_method="HMAC-SHA1",
oauth_consumer_key="pw0ihLFxH3nwDrd4HBd7pqUrc",
oauth_token="1360011857969479682-iLrxBUlqdtExwkqiN9iZsHYDXIFTZz",
oauth_signature="oANPns3UzMayAeo0sh0Wx9E%2FnaA%3D"}
2021-06-25 03:18:33,594 - DEBUG - Updated body: None
2021-06-25 03:18:33,603 - DEBUG - Starting new HTTPS connection (1):
api.twitter.com:443
2021-06-25 03:18:35,089 - DEBUG - https://api.twitter.com:443 "GET /1.1/search/t
weets.json?max_id=1405681100459028483&q=%5B%27coffee+island+OR+CoffeeIsland%27%5
D&since=2021-06-18&until=2021-06-25&tweet_mode=extended&lang=en HTTP/1.1" 200
417
2021-06-25 03:18:35,092 - DEBUG - PARAMS: {'create': b'True'}
2021-06-25 03:18:35,095 - INFO - elapsed_time: '78.96849727630615'

```

[31]: output

```

[31]: ,
      {'text': 'RT @mir_gifs: June 19th bunny and animal crossing island coffee
status https://t.co/cHdGkmEMN1',
       'favourite_count': 0,
       'retweet_count': 4,
       'created_at': datetime.datetime(2021, 6, 19, 3, 4, 41)},
      {'text': 'June 19th bunny and animal crossing island coffee status
https://t.co/cHdGkmEMN1',
       'favourite_count': 14,
       'retweet_count': 4,
       'created_at': datetime.datetime(2021, 6, 19, 2, 55, 57)},
      {'text': 'RT @cathsherman: Coral and Limestone Beach, Great Abaco Island, The
Bahamas #Bahamas #FineArtAmerica #CLS \n#ArtForSale\nhttps://t.co/5Uk1lc1...',
       'favourite_count': 0,
       'retweet_count': 39,
       'created_at': datetime.datetime(2021, 6, 19, 2, 37, 42)},
      {'text': "@BjornIronsights At one point in time I had one overnight a month
(developmentally disabled kiddo.)\nI drove 2 or 3 hours to a scruffy little
seaside motel in Long Island. I'd bring a journal and my guitar. No TV. Fried
clams. Coffee. Pancakes. Room 66 every time.",
       'favourite_count': 0,
       'retweet_count': 0,
       'created_at': datetime.datetime(2021, 6, 19, 2, 13, 55)},
      {'text': 'RT @TheW0lfup: Coffee island',
       'favourite_count': 0,
       'retweet_count': 20,
       'created_at': datetime.datetime(2021, 6, 19, 1, 31, 15)},
      {'text': '@Merkabagaming Man before we got it we'd go to the local coffee shop.
Now we have even better coffee!\n\nWe loved the Big Island coffee roasters you
recommended, and are on I think our 3rd or 4th bag now ',

```

```

    'favourite_count': 1,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 19, 1, 21, 35)},
{'text': 'RT @CatherineSher10: Dauphin Island Sunset\nClouds streak across the
setting sun on #DauphinIsland #Alabama #SunsetCapital #Sunset #GulfofMe...',
 'favourite_count': 0,
 'retweet_count': 62,
 'created_at': datetime.datetime(2021, 6, 19, 1, 12, 39)},
{'text': 'RT @CatherineSher10: A statue of Hawaiian King Kamehameha the Great,
Wailoa State Park in #Hilo #Hawaii on the Big Island. \n#BigIsland #CL...',
 'favourite_count': 0,
 'retweet_count': 28,
 'created_at': datetime.datetime(2021, 6, 19, 0, 20, 52)},
{'text': 'Small Town Coffee! Loving the island life! https://t.co/6SghIKp8sg',
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 19, 0, 19, 21)},
{'text': 'RT @TheArtist1329: Strolling Great Egret on #GreatAbaco Island in The
#Bahamas #BeachLife #FineArtAmerica #CLS \n#ArtForSale \nhttps://t.co/...',
 'favourite_count': 0,
 'retweet_count': 47,
 'created_at': datetime.datetime(2021, 6, 18, 22, 54, 11)},
{'text': 'RT @CatherineSher10: A statue of Hawaiian King Kamehameha the Great,
Wailoa State Park in #Hilo #Hawaii on the Big Island. \n#BigIsland #CL...',
 'favourite_count': 0,
 'retweet_count': 28,
 'created_at': datetime.datetime(2021, 6, 18, 22, 53, 1)},
{'text': 'RT @cathsherman: A great blue heron stands near mossy rocks near
Pelican Point on Dauphin Island #Alabama \nIn the background, a brown pelic...',
 'favourite_count': 0,
 'retweet_count': 85,
 'created_at': datetime.datetime(2021, 6, 18, 22, 42, 10)},
{'text': 'RT @cathsherman: Afternoon at the Eighteenth Hole, Kiawah Island,
South Carolina #KiawahIsland #Golf #SouthCarolina Clubhouse \n#SOLD...',
 'favourite_count': 0,
 'retweet_count': 49,
 'created_at': datetime.datetime(2021, 6, 18, 22, 39, 29)},
{'text': 'For the next month #Nanaimo artist Melissa Anderson presents her oil
painting exhibit 'Seascapes' at White Rabbit Coffee Co.
https://t.co/KuMV7e5r4o',
 'favourite_count': 2,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 22, 18, 34)},
{'text': 'RT @KauaiCoffeeCo: Guaranteed to get you out of bed faster than your
alarm clock. Try 100% Kauai Coffee with a hint of delicious island-ins...',
 'favourite_count': 0,
 'retweet_count': 3,

```

```

    'created_at': datetime.datetime(2021, 6, 18, 22, 7, 30)},
    {'text': 'RT @cathsherman: Afternoon at the Eighteenth Hole, Kiawah Island,
South Carolina #KiawahIsland #Golf      #SouthCarolina Clubhouse \n#SOLD...',
    'favourite_count': 0,
    'retweet_count': 49,
    'created_at': datetime.datetime(2021, 6, 18, 21, 51, 20)},
    {'text': 'RT @cathsherman: Afternoon at the Eighteenth Hole, Kiawah Island,
South Carolina #KiawahIsland #Golf      #SouthCarolina Clubhouse \n#SOLD...',
    'favourite_count': 0,
    'retweet_count': 49,
    'created_at': datetime.datetime(2021, 6, 18, 20, 52, 59)},
    {'text': 'RT @cathsherman: Afternoon at the Eighteenth Hole, Kiawah Island,
South Carolina #KiawahIsland #Golf      #SouthCarolina Clubhouse \n#SOLD...',
    'favourite_count': 0,
    'retweet_count': 49,
    'created_at': datetime.datetime(2021, 6, 18, 20, 52, 31)},
    {'text': '@JadahParadox No but I won a set of hand painted coffee mugs at the
St. George Island Fire Station Bingo Night!  ',
    'favourite_count': 1,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 20, 2, 46)},
    {'text': '@wd3ncha does this explain my island (i should clarify i played for
an additional month-ish before kenneth spilled coffee on my switch and i had to
get a new one and start over)',
    'favourite_count': 1,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 19, 47, 13)},
    {'text': 'RT @TheW0lfup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 19, 11, 10)},
    {'text': 'RT @TheW0lfup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 19, 1, 57)},
    {'text': 'RT @TheW0lfup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 18, 58, 23)},
    {'text': 'RT @TheW0lfup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 18, 23, 19)},
    {'text': 'Read Treasure Island and drink some chilled lemonade from a irish
coffee glass.',
    'favourite_count': 0,
    'retweet_count': 0,

```

```

    'created_at': datetime.datetime(2021, 6, 18, 18, 6, 4)},
    {'text': 'honestly you can leave me in an island alone but i would need
thousand of books and my own coffee maker. if you're kind enough to give me a
non-ending supply of coffee beans. it would be deeply appreciated',
    'favourite_count': 4,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 18, 4, 29)},
    {'text': 'RT @visitcyprus: Morning coffee by the sea at charming locations,
here, at your favourite island destination! Where are you having your cof...',
    'favourite_count': 0,
    'retweet_count': 24,
    'created_at': datetime.datetime(2021, 6, 18, 17, 40, 19)},
    {'text': '@nikkimwalls There was a coffee shop in Rhode Island that was
adjacent to this jewelry store-really great coffee & work vibes, but the
catch was they could *only* play the jewelry store approved Pandora radio
station that cycled through the same 20 songs endlessly...',
    'favourite_count': 1,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 17, 39, 48)},
    {'text': 'RT @premierhop: & Online\n-\n@mikkelleruk x Seven Island x
@WarPigs_Brewpub / Beer Greek Breakfast / 12.5% / 375ml\n-\nCoffee from
Cafetierra...',
    'favourite_count': 0,
    'retweet_count': 1,
    'created_at': datetime.datetime(2021, 6, 18, 16, 41, 20)},
    {'text': ' \n[Friends+]Treasure Hunt Island #VRCinHere 0:19:10',
    'favourite_count': 0,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 15, 16, 59)},
    {'text': ' & Online\n-\n@mikkelleruk x Seven Island x @WarPigs_Brewpub /
Beer Greek Breakfast / 12.5% / 375ml\n-\nCoffee from Cafetierra roastery based
on Corfu Island; Mahleb from Thessaloniki as well as Carob flour from Chios
Island which can be described as a Mediterranean Cocoa.
https://t.co/S57qKRjkyG',
    'favourite_count': 0,
    'retweet_count': 1,
    'created_at': datetime.datetime(2021, 6, 18, 14, 54, 21)},
    {'text': '@The_MoBiz Got to admit I haven't taken the ferry off the Island in
about 2 years now. May do that just for a relaxing there and back trip; sit on
the deck, drink coffee, watch the scenery go by. Are you doing the Tsawassen-
Swartz Bay route?',
    'favourite_count': 2,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 14, 39, 30)},
    {'text': ' \n[Friends+]Treasure Hunt Island #VRCinHere 23:40:46',
    'favourite_count': 0,
    'retweet_count': 0,

```

```

    'created_at': datetime.datetime(2021, 6, 18, 14, 38, 35)},
{'text': ' \n[Friends+]Treasure Hunt Island #VRCinHere 23:29:42',
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 14, 27, 31)},
{'text': 'Coffee with David (@ Cat Island Coffeehouse/Pass Christian Books in
Pass Christian, MS) https://t.co/4kMXKMLuAP',
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 14, 25, 17)},
{'text': ' \n[Friends+]Treasure Hunt Island #VRCinHere 23:11:51',
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 14, 9, 54)},
{'text': 'Spread over five acres along the cruise port in Ocho Rios, this
shopping center has around 30 duty-free shops selling souvenirs, clothing,
sunglasses, Blue Mountain coffee, sauces, books, and Bob Marley memorabilia.
#walkthevillage #islandvillage https://t.co/htUz8w4wsu',
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 14, 2, 10)},
{'text': '@AlexaGagosz Understandably! haha\n\nIt's very cool to see them have
grown from a coffee truck to now have three locations. Love their
drinks/food/service/vibes & their seasonal Maple Sea Salt Latte from fall
2019 remains the best non-alcoholic coffee drink I've had in Rhode Island',
 'favourite_count': 1,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 13, 51, 15)},
{'text': 'I showed up on time for my train to Fire Island with an iced coffee
in hand. This is my biggest accomplishment since getting a master's degree.',
 'favourite_count': 45,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 13, 20, 28)},
{'text': "@jessicahodlr DM use to reserve Satoshi's Selection coffee. its
Bitcoin coffee from the enchanted island of Puerto Rico Available in 2
weeks for shipping",
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 13, 18, 12)},
{'text': "Colourful handmade chocolates being boxed up from Arlene's Island
Artisan Chocolate \n\nAvailable here https://t.co/yT7USQMTcx
https://t.co/Bzs0iRjgoh",
 'favourite_count': 2,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 12, 55, 2)},
{'text': '90% of socializing in Rhode Island involves iced coffee',
 'favourite_count': 2,

```

```

    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 12, 52, 35)},
{'text': ' \n[Friends+]Treasure Hunt Island #VRCinHere 21:37:10',
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 12, 35, 14)},
{'text': '@island_iverson This tweet made me spill hot coffee on my d region.
Good stuff.',
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 12, 24, 47)},
{'text': 'RT @CatherineSher10: A statue of Hawaiian King Kamehameha the Great,
Wailoa State Park in #Hilo #Hawaii on the Big Island. \n#BigIsland #CL...',
 'favourite_count': 0,
 'retweet_count': 28,
 'created_at': datetime.datetime(2021, 6, 18, 12, 24, 11)},
{'text': 'RT @cathsherman: Afternoon at the Eighteenth Hole, Kiawah Island,
South Carolina #KiawahIsland #Golf #SouthCarolina Clubhouse \n#SOLD...',
 'favourite_count': 0,
 'retweet_count': 49,
 'created_at': datetime.datetime(2021, 6, 18, 12, 24)},
{'text': 'RT @cathsherman: Afternoon at the Eighteenth Hole, Kiawah Island,
South Carolina #KiawahIsland #Golf #SouthCarolina Clubhouse \n#SOLD...',
 'favourite_count': 0,
 'retweet_count': 49,
 'created_at': datetime.datetime(2021, 6, 18, 12, 14, 57)},
{'text': 'Good morning from Hilton Head Island, SC! Coffee and a great book by
@JulieDiCaro on the balcony. https://t.co/Z8R2VdSubX',
 'favourite_count': 6,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 12, 9, 3)},
{'text': '@LorinOnTV Found a great Long Island coffee
company.\nhttps://t.co/GMdM06ndPv',
 'favourite_count': 1,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 11, 30, 46)},
{'text': 'RT @visitcyprus: Morning coffee by the sea at charming locations,
here, at your favourite island destination! Where are you having your cof...',
 'favourite_count': 0,
 'retweet_count': 24,
 'created_at': datetime.datetime(2021, 6, 18, 11, 28, 50)},
{'text': "Here's your OTD trivia for 6/18... Join us tonight at Amelia Island
Coffee, and have a great Friday! #trivia30 #wakeupyourbrain #OnThisDay #Monterey
#California #JimiHendrix #JanisJoplin #TheWho #OtisRedding #USSR #FastAndFurious
#Spain #CarolKane #KurtBrowning #BlakeShelton https://t.co/K6IMU3DNGy",
 'favourite_count': 0,
 'retweet_count': 0,

```



```

    'created_at': datetime.datetime(2021, 6, 18, 11, 10, 7)},
    {'text': 'Join us tonight for an evening of TRIVIA:30 at Amelia Island Coffee!
The fun starts at 6:00... Free answer is *Ensenada* See you there! #whattimeisit
#TRIVIA30 #wakeupyourbrain #trivia #fridaynight #AmeliaIsland #FernandinaBeach
#LoveAmelia #Fernandina #trivia #fridayfun https://t.co/T2JENWOhlT',
    'favourite_count': 0,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 11, 5, 5)},
    {'text': "Good Morning, I've got the coffee on for you.",
    'favourite_count': 0,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 11, 0, 42)},
    {'text': 'RT @TheW0lfpup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 10, 37, 3)},
    {'text': 'RT @GySgt2: @justmeagain34 Good morning my friend, it is 73 degrees
here on Parris island, the coffee is strong and black, the spiders wh...',
    'favourite_count': 0,
    'retweet_count': 2,
    'created_at': datetime.datetime(2021, 6, 18, 10, 29, 26)},
    {'text': '@justmeagain34 Good morning my friend, it is 73 degrees here on
Parris island, the coffee is strong and black, the spiders where All over me
this morning leaving my house I guess I am just going to have to get a flame
thrower for the armory be fir I go home tonight, TGIF ',
    'favourite_count': 7,
    'retweet_count': 2,
    'created_at': datetime.datetime(2021, 6, 18, 10, 15, 7)},
    {'text': '@mochis_island_ I'm watching tv and kissing coffee but I should
be working',
    'favourite_count': 1,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 9, 47, 31)},
    {'text': '@SarkySage coffee followed by Russell Island tap water vintage 2021',
    'favourite_count': 1,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 8, 50, 14)},
    {'text': "Bonjour! Let's make a sandwich to start a happy day! Open the
fridge, and grab two pieces of toasts. Take ham as the protein, tomato and
spinach as veggies. Finally, add some thousand island dressing and cheddar
cheese. Have it with a cup of black coffee. Enjoy your morning!
https://t.co/lzew0cOHhG",
    'favourite_count': 0,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 8, 20, 55)},
    {'text': 'Read Treasure Island and enjoy some kombucha out of a irish coffee
glass.',

```

```

    'favourite_count': 0,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 8, 15, 39)},
    {'text': 'Magical island #Howth \nView from #Blackrock #Dublin \n\nMorning
dip & coffee with pals \u200d \n#Reset #VitaminSea \n\n
\u2066@janeconan\u2069 \u2066@TheGinaLondon\u2069 https://t.co/LQd53AoUz4',
    'favourite_count': 29,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 7, 6, 5)},
    {'text': 'RT @TheW0lfup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 6, 47, 43)},
    {'text': 'RT @TheW0lfup: Coffee island https://t.co/52WId63NXo',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 6, 39, 20)},
    {'text': 'RT @TheW0lfup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 6, 36, 32)},
    {'text': 'RT @TheW0lfup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 6, 17, 58)},
    {'text': 'RT @TheW0lfup: Coffee island https://t.co/52WId63NXo',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 6, 15, 42)},
    {'text': 'RT @visitcyprus: Morning coffee by the sea at charming locations,
here, at your favourite island destination! Where are you having your cof...',
    'favourite_count': 0,
    'retweet_count': 24,
    'created_at': datetime.datetime(2021, 6, 18, 5, 37, 19)},
    {'text': 'RT @TheW0lfup: Coffee island',
    'favourite_count': 0,
    'retweet_count': 20,
    'created_at': datetime.datetime(2021, 6, 18, 4, 32, 7)},
    {'text': 'Coffee Island https://t.co/yM2zbp9hp3',
    'favourite_count': 2,
    'retweet_count': 0,
    'created_at': datetime.datetime(2021, 6, 18, 4, 6, 35)},
    {'text': 'DDDDaniel just supported @cretetravel on @buymeacoffee!
\n\n@cretetravel is ... creating memorable experiences in lovely Crete Island!
. You can support by buying a coffee here - https://t.co/Qo8h8xTPAN',
    'favourite_count': 1,
    'retweet_count': 0,

```

```

    'created_at': datetime.datetime(2021, 6, 18, 4, 2, 26)},
{'text': 'RT @TheWolfpup: Coffee island',
 'favourite_count': 0,
 'retweet_count': 20,
 'created_at': datetime.datetime(2021, 6, 18, 3, 57, 17)},
{'text': "@SerpicoDet That's incredible! It would have been in Nanaimo, my
family spoke very highly of you. If you ever end up on the island you're more
than welcome over for coffee again. Thanks for confirming an old story!",
 'favourite_count': 3,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 3, 55, 4)},
{'text': 'RT @TheWolfpup: Coffee island',
 'favourite_count': 0,
 'retweet_count': 20,
 'created_at': datetime.datetime(2021, 6, 18, 3, 41, 52)},
{'text': 'RT @visitcyprus: Morning coffee by the sea at charming locations,
here, at your favourite island destination! Where are you having your cof...',
 'favourite_count': 0,
 'retweet_count': 24,
 'created_at': datetime.datetime(2021, 6, 18, 3, 35, 29)},
{'text': 'Read Treasure Island and drink some chilled kombucha from a irish
coffee glass.',
 'favourite_count': 0,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 3, 26, 41)},
{'text': '@ARLESVENUSX the coffee ice cream slander here is disturbing. i am
here to say you would not survive in rhode island ',
 'favourite_count': 1,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 1, 58, 12)},
{'text': '"She also loves to travel around the world and sip coffee along the
way, so if we are ever able to connect in real life, I think we'd get along
splendidly." \n\nTruth, Dr. Hasty! \n\nThanks for a fun #bookfeature with
#IslandCharm from @PelicanBookGrp. \n\nhttps://t.co/EfOH61FEky',
 'favourite_count': 5,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 1, 40, 34)},
{'text': "RT @lizwadsworth65: @quinncy Exile and Duke, a used bookstore/cafe
specializing in historical romances with unhappy endings. It's located i...",
 'favourite_count': 0,
 'retweet_count': 1,
 'created_at': datetime.datetime(2021, 6, 18, 1, 1, 30)},
{'text': "@quinncy Exile and Duke, a used bookstore/cafe specializing in
historical romances with unhappy endings. It's located in a refurbished island
prison in which all the cells are private reading rooms with French press coffee
makers and complimentary macarons.",
 'favourite_count': 25,

```

```

        'retweet_count': 1,
        'created_at': datetime.datetime(2021, 6, 18, 0, 59, 47)}},
{'text': '@bananzers I wanna visit the big island and drink Kona coffee from
the source ',
 'favourite_count': 2,
 'retweet_count': 0,
 'created_at': datetime.datetime(2021, 6, 18, 0, 17, 45)}}]

```

```
[32]: len(output)
```

```
[32]: 644
```

2.0.1 create pdf from list

Finally, we convert the output list to a `pandas DataFrame` and we store results.

```
[33]: pdf_cof_island = pd.DataFrame(output)
```

```
[34]: pdf_cof_island.shape
```

```
[34]: (644, 4)
```

```
[35]: pdf_cof_island.head(2)
```

```
[35]:
```

| | text | favourite_count | retweet_count | created_at |
|---|---|-----------------|---------------|---------------------|
| 0 | Stop by the @nespressousa store after my docto... | 0 | 0 | 2021-06-24 23:53:05 |
| 1 | RT @witch_mote: Legend has is that if you fini... | 0 | 143 | 2021-06-24 23:48:56 |

```
[40]: pdf_cof_island.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 644 entries, 0 to 643
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   text                   644 non-null   object
1   favourite_count        644 non-null   int64
2   retweet_count          644 non-null   int64
3   created_at             644 non-null   datetime64[ns]
dtypes: datetime64[ns](1), int64(2), object(1)
memory usage: 20.2+ KB

```

```
[37]: pdf_cof_island[['text', 'created_at']]
```

```
[37]:
```

| | | text | created_at |
|-----|--|---|---------------------|
| 0 | | Stop by the @nespressousa store after my docto... | 2021-06-24 23:53:05 |
| 1 | | RT @witch_mote: Legend has is that if you fini... | 2021-06-24 23:48:56 |
| 2 | | RT @SafemoonWarrior: #SAFEMOON \nCan some... | 2021-06-24 23:45:59 |
| 3 | | @orangepillpod @maxkeiser @stacyherbert DM us ... | 2021-06-24 23:34:34 |
| 4 | | RT @SafemoonWarrior: #SAFEMOON \nCan some... | 2021-06-24 23:24:31 |
| ... | | ... | ... |
| 639 | | @ARLESVENUSX the coffee ice cream slander here... | 2021-06-18 01:58:12 |
| 640 | | "She also loves to travel around the world and... | 2021-06-18 01:40:34 |
| 641 | | RT @lizwadsworth65: @quinncy Exile and Duke, a... | 2021-06-18 01:01:30 |
| 642 | | @quinncy Exile and Duke, a used bookstore/cafe... | 2021-06-18 00:59:47 |
| 643 | | @bananzers I wanna visit the big island and dr... | 2021-06-18 00:17:45 |

[644 rows x 2 columns]

```
[39]: pdf_cof_island[['text', 'created_at']].groupby('created_at').first()
```

```
[39]:
```

| created_at | text |
|---------------------|---|
| 2021-06-18 00:17:45 | @bananzers I wanna visit the big island and dr... |
| 2021-06-18 00:59:47 | @quinncy Exile and Duke, a used bookstore/cafe... |
| 2021-06-18 01:01:30 | RT @lizwadsworth65: @quinncy Exile and Duke, a... |
| 2021-06-18 01:40:34 | "She also loves to travel around the world and... |
| 2021-06-18 01:58:12 | @ARLESVENUSX the coffee ice cream slander here... |
| ... | ... |
| 2021-06-24 23:24:31 | RT @SafemoonWarrior: #SAFEMOON \nCan some... |
| 2021-06-24 23:34:34 | @orangepillpod @maxkeiser @stacyherbert DM us ... |
| 2021-06-24 23:45:59 | RT @SafemoonWarrior: #SAFEMOON \nCan some... |
| 2021-06-24 23:48:56 | RT @witch_mote: Legend has is that if you fini... |
| 2021-06-24 23:53:05 | Stop by the @nespressousa store after my docto... |

[641 rows x 1 columns]

2.0.2 save and read pdf without header

```
[28]: pdf_cof_island.to_csv('output_cof_island.csv', mode='a', header = False, index_
      ↪= False )
      #df.to_csv('output.csv')
```

```
[29]: pdf_cof_island = pd.read_csv('output3.csv',
      ↪names=['text', 'favourite_count', 'retweet_count', 'created_at'],
      ↪parse_dates=['created_at'],)
      df2.head(5)
```

```
[29]:
```

| | text | favourite_count | \ |
|---|---|-----------------|---|
| 0 | | | |
| 1 | @breakfasttv I received the Astrazeneca vaccin... | 0 | |
| 2 | #Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJo... | 3 | |
| 3 | RT @TimWattsMP: Can you believe this?\n\nThe M... | 0 | |
| 4 | RT @DrEricDing: 5) One dose of the vaccine is ... | 0 | |

| | retweet_count | created_at |
|---|---------------|---------------------|
| 0 | retweet_count | created_at |
| 1 | 0 | 2021-06-20 23:59:37 |
| 2 | 5 | 2021-06-20 23:59:33 |
| 3 | 212 | 2021-06-20 23:58:46 |
| 4 | 270 | 2021-06-20 23:57:54 |

```
[30]: pdf=df2
pdf.shape(
```

```
-----
TypeError                                Traceback (most recent call last)
<ipython-input-30-7e9cd4b4d6a5> in <module>
      1 pdf=df2
      2
----> 3 pdf.shape()

TypeError: 'tuple' object is not callable
```

```
[31]: pdf_cof_island.shape
```

```
[31]: (13089, 4)
```

2.0.3 save a pdf to a csv file with header

```
[46]: #df = pd.DataFrame(output)
pdf_cof_island.to_csv('output_cof_island.csv', mode='a', header=True, index =_
↪False)
```

3 Create a pdf and a sdf from a csv file with header

3.0.1 pdf

```
[47]: pdf_cof_island2 = pd.read_csv('output_cof_island.csv',  
    ↳ parse_dates=['created_at'],  
pdf_cof_island2.head(5)
```

```
[47]:
```

| | | | | text | favourite_count | \ |
|---|--------------------|---------------|---------------|------|-----------------|---|
| 0 | RT @NikosPachilas: | Mikel | Coffee Island | ... | 0 | |
| 1 | RT @NikosPachilas: | Mikel | Coffee Island | ... | 0 | |
| 2 | RT @RealTime_eu: | Mikel | Coffee Island | ... | 0 | |
| 3 | RT @RealTime_eu: | Mikel | Coffee Island | ... | 0 | |
| 4 | Mikel | Coffee Island | ... | | 2 | |

| | retweet_count | created_at |
|---|---------------|---------------------|
| 0 | 2 | 2021-06-24 12:26:01 |
| 1 | 2 | 2021-06-24 12:20:35 |
| 2 | 2 | 2021-06-24 12:20:27 |
| 3 | 2 | 2021-06-24 12:19:07 |
| 4 | 2 | 2021-06-24 12:15:01 |

```
[ ]:
```

```
[48]: pdf_cof_island2.shape
```

```
[48]: (14, 4)
```

```
[113]: pdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 6510 entries, 0 to 6509  
Data columns (total 8 columns):  
#   Column                Non-Null Count  Dtype  
---  -  
0   text                   6510 non-null   object  
1   favourite_count        6510 non-null   int64  
2   retweet_count          6510 non-null   int64  
3   created_at             6510 non-null   object  
4   negative_nltk           6510 non-null   float64  
5   positive_nltk           6510 non-null   float64  
6   neutral_nltk            6510 non-null   float64  
7   compound_nltk           6510 non-null   float64  
dtypes: float64(4), int64(2), object(2)  
memory usage: 407.0+ KB
```

```
[111]: pdf.describe()
```

```
[111]:
```

| | favourite_count | retweet_count | negative_nltk | positive_nltk | \ |
|-------|-----------------|---------------|---------------|---------------|---|
| count | 6510.000000 | 6510.000000 | 6510.000000 | 6510.000000 | |
| mean | 1.997081 | 903.088172 | 0.047635 | 0.038732 | |
| std | 54.727334 | 1714.038097 | 0.083419 | 0.070628 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.000000 | 8.000000 | 0.000000 | 0.000000 | |
| 50% | 0.000000 | 200.000000 | 0.000000 | 0.000000 | |
| 75% | 0.000000 | 1307.000000 | 0.096000 | 0.073000 | |
| max | 3482.000000 | 19068.000000 | 0.375000 | 0.526000 | |

| | neutral_nltk | compound_nltk |
|-------|--------------|---------------|
| count | 6510.000000 | 6510.000000 |
| mean | 0.913632 | -0.021985 |
| std | 0.106266 | 0.358451 |
| min | 0.474000 | -0.957800 |
| 25% | 0.847000 | -0.077200 |
| 50% | 0.960500 | 0.000000 |
| 75% | 1.000000 | 0.000000 |
| max | 1.000000 | 0.940200 |

3.1 sdf

```
[18]: # spark.read.csv(
#       "some_input_file.csv",
#       header=True,
#       mode="DROPMALFORMED",
#       schema=schema
# )

# or

# (
#     spark.read
#     .schema(schema)
#     .option("header", "true")
#     .option("mode", "DROPMALFORMED")
#     .csv("some_input_file.csv")
# )

from pyspark.sql.types import StructType, StructField
from pyspark.sql.types import DoubleType, IntegerType, StringType, DateType

schema = StructType([
    StructField("text", StringType()),
    StructField("favourite_count", IntegerType()),
```



```

    StructField("retweet_count", IntegerType()),
    StructField("created_at", DateType())
])
# https://datascience.stackexchange.com/questions/12727/
# ↳reading-csvs-with-new-lines-in-fields-with-spark
sdf = spark.read.csv('output3.csv', schema=schema, header=True, escape = '"',
↳multiLine=True)

```

```

[19]: type(sdf)
# pyspark.sql.dataframe.DataFrame
sdf.printSchema()

```

```

root
|-- text: string (nullable = true)
|-- favourite_count: integer (nullable = true)
|-- retweet_count: integer (nullable = true)
|-- created_at: date (nullable = true)

```

```

[20]: sdf.count()

```

```

[20]: 6510

```

```

[21]: sdf.columns

```

```

[21]: ['text', 'favourite_count', 'retweet_count', 'created_at']

```

```

[22]: len(sdf.columns)

```

```

[22]: 4

```

```

[ ]: # Add this to the your code:

import pyspark
def spark_shape(self):
    return (self.count(), len(self.columns))
pyspark.sql.dataframe.DataFrame.shape = spark_shape
# Then you can do

# >>> df.shape()
# (10000, 10)

```

```

[23]: sdf.toPandas().head(5)

```

```

[23]:
          text  favourite_count \
0  @breakfasttv I received the Astrazeneca vaccin...      0
1  #Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJo...      3

```

```

2 RT @TimWattsMP: Can you believe this?\n\nThe M... 0
3 RT @DrEricDing: 5) One dose of the vaccine is ... 0
4 RT @DrEricDing: 5) One dose of the vaccine is ... 0

```

```

    retweet_count  created_at
0                0  2021-06-20
1                5  2021-06-20
2             212  2021-06-20
3             270  2021-06-20
4             270  2021-06-20

```

```

[29]: s = sdf.toPandas()

s.shape

```

```

[29]: (6510, 4)

```

3.1.1 def sentiment_scores & sentiment_scoresUDF

```

[30]: import sys
from pyspark.sql.functions import udf
# DoubleType, FloatType, ByteType, IntegerType, LongType, ShortType,
↳ArrayType, StructField, StructType, Row
from pyspark.sql.functions import pandas_udf, PandasUDFType
import pyspark.sql.types as Types
from nltk.sentiment.vader import SentimentIntensityAnalyzer

def sentiment_scores(sentence: str) -> dict :
    # Create a SentimentIntensityAnalyzer object.
    sid = SentimentIntensityAnalyzer()
    # polarity_scores method of SentimentIntensityAnalyzer
    # object gives a sentiment dictionary.
    # which contains pos, neg, neu, and compound scores.
    r = sid.polarity_scores(sentence);
    return r
    # You can optionally set the return type of your UDF. The default return
↳type, is StringType.
    # udf factorial_p = udf(factorial_p, LongType())

sentiment_scoresUDF = udf(sentiment_scores, Types.MapType(Types.
↳StringType(), Types.DoubleType()))

```

3.1.2 pdf

```
[56]: #df3['rating'] = df3['text'].apply(sid.polarity_scores)

pdf['rating'] = pdf['text'].apply(sentiment_scores)
```

```
[57]: pdf.head(2)
```

```
[57]:
```

| | text | favourite_count | \ |
|---|---|-----------------|---|
| 0 | @breakfasttv I received the Astrazeneca vaccin... | 0 | |
| 1 | #Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJo... | 3 | |

| | retweet_count | created_at | \ |
|---|---------------|---------------------|---|
| 0 | 0 | 2021-06-20 23:59:37 | |
| 1 | 5 | 2021-06-20 23:59:33 | |

| | rating |
|---|---|
| 0 | {'neg': 0.049, 'neu': 0.752, 'pos': 0.199, 'co... |
| 1 | {'neg': 0.0, 'neu': 0.603, 'pos': 0.397, 'comp... |

```
[33]: pdf.tail(2)
```

```
[33]:
```

| | text | favourite_count | \ |
|------|---|-----------------|---|
| 6508 | RT @JohnRHewson: For the record, how many of o... | 0 | |
| 6509 | RT @DrEricDing: 3) "After years of reading res... | 0 | |

| | retweet_count | created_at | \ |
|------|---------------|---------------------|---|
| 6508 | 890 | 2021-06-19 00:00:27 | |
| 6509 | 55 | 2021-06-19 00:00:08 | |

| | rating |
|------|---|
| 6508 | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... |
| 6509 | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... |

```
[58]: pdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6510 entries, 0 to 6509
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   text             6510 non-null   object
1   favourite_count  6510 non-null   int64
2   retweet_count    6510 non-null   int64
3   created_at       6510 non-null   object
4   rating           6510 non-null   object
dtypes: int64(2), object(3)
memory usage: 254.4+ KB
```

3.1.3 sdf

```
[59]: from pyspark.sql.functions import col,sqrt,log,reverse
      sdf = sdf.withColumn("rating", sentiment_scoresUDF(sdf.text))
      # t.show()
      sdf.toPandas().style.set_properties(subset=['text'], **{'width': '300px'})
```

```
[59]: <pandas.io.formats.style.Styler at 0x7f2a49c571d0>
```

```
[60]: sdf.toPandas().head(2).style.set_properties(subset=['text'], **{'width': '300px'})
```

```
[60]: <pandas.io.formats.style.Styler at 0x7f2a56ca5c10>
```

```
[61]: sdf.toPandas().tail(2).style.set_properties(subset=['text'], **{'width': '300px'})
```

```
[61]: <pandas.io.formats.style.Styler at 0x7f2a4876bb90>
```

```
[62]: sdf.show(2, truncate = 30)
```

```
+-----+-----+-----+-----+-----+
|          text|favourite_count|retweet_count|created_at|negative_nltk|positive_nltk|neutral_nltk|compound_nltk|rating|
+-----+-----+-----+-----+-----+
|@breakfasttv I received the...|0|0|2021-06-20|0.049|0.199|0.752|0.7793|[neg -> 0.049, pos -> 0.199...|
|#Pfizer
#AstraZeneca
#Moder...|3|5|2021-06-20|0.0|0.397|
0.603|0.7639|[neg -> 0.0, pos -> 0.397, ...|
+-----+-----+-----+-----+
only showing top 2 rows
```

```
[49]: # Returns the first num rows as a list of Row.
      # This method should only be used if the resulting array is expected to be
      # small, as all the data is loaded into the driver's memory.
      sdf.head(2)
```

```
[49]: [Row(text='@breakfasttv I received the Astrazeneca vaccine on April 9th. No, I
do not regret it! I look forward to getting the Pfizer vaccine on the 28th of
this month. I feel confident we will be fully protected. Even if it is
```

```
unconventional. ', favourite_count=0, retweet_count=0,
created_at=datetime.date(2021, 6, 20), rating={'neg': 0.049, 'pos': 0.199,
'compound': 0.7793, 'neu': 0.752}),
Row(text="#Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJohnson \n\nWhat's up !
Safe and effective !\n\n\n https://t.co/fvw1H8cmXT", favourite_count=3,
retweet_count=5, created_at=datetime.date(2021, 6, 20), rating={'neg': 0.0,
'pos': 0.397, 'compound': 0.7639, 'neu': 0.603})]
```

```
[50]: # Returns the last num rows as a list of Row.
sdf.tail(2)
```

```
[50]: [Row(text='RT @JohnRHewson: For the record, how many of our political leaders
had Pfizer rather than AstraZeneca?', favourite_count=0, retweet_count=890,
created_at=datetime.date(2021, 6, 19), rating={'neg': 0.0, 'pos': 0.0,
'compound': 0.0, 'neu': 1.0}),
Row(text='RT @DrEricDing: 3) "After years of reading research on mixing vaccine
types -- known as heterologous prime-boosting -- Morgon concluded tha...',
favourite_count=0, retweet_count=55, created_at=datetime.date(2021, 6, 19),
rating={'neg': 0.0, 'pos': 0.0, 'compound': 0.0, 'neu': 1.0})]
```

```
[63]: sdf.printSchema()
```

```
root
|-- text: string (nullable = true)
|-- favourite_count: integer (nullable = true)
|-- retweet_count: integer (nullable = true)
|-- created_at: date (nullable = true)
|-- negative_nltk: double (nullable = true)
|-- positive_nltk: double (nullable = true)
|-- neutral_nltk: double (nullable = true)
|-- compound_nltk: double (nullable = true)
|-- rating: map (nullable = true)
|   |-- key: string
|   |-- value: double (valueContainsNull = true)
```

```
[64]: from pyspark.sql.functions import col

sdf = sdf.withColumn('negative_nltk', col('rating')['neg']) \
.withColumn('positive_nltk', col('rating')['pos']) \
.withColumn('neutral_nltk', col('rating')['neu']) \
.withColumn('compound_nltk', col('rating')['compound']) \
.drop('rating')
```

```
[65]: sdf.toPandas().head(2).style.set_properties(subset=['text'], **{'width': 300px})
```

```
[65]: <pandas.io.formats.style.Styler at 0x7f2a4d28b0d0>
```

```
[66]: sdf.toPandas().tail(2).style.set_properties(subset=['text'], **{'width': 300px})
```

```
[66]: <pandas.io.formats.style.Styler at 0x7f2a4d305750>
```

```
[228]: sdf_from_list_of_rows = spark.createDataFrame(sdf.tail(2))

sdf_from_list_of_rows.show()
```

```
+-----+-----+-----+-----+-----+-----+
|          text|favourite_count|retweet_count|created_at|negative_nltk|positive_nltk|neutral_nltk|compound_nltk|
+-----+-----+-----+-----+-----+-----+
|RT @JohnRHewson: ...|          0|          890|2021-06-19|          0.0|          0.0|          1.0|          0.0|
|RT @DrEricDing: 3...|          0|          55|2021-06-19|          0.0|          0.0|          1.0|          0.0|
+-----+-----+-----+-----+-----+-----+
```

```
[229]: sdf_from_pdf = spark.createDataFrame(pdf)

sdf_from_list_of_rows.show(2)
```

```
+-----+-----+-----+-----+-----+-----+
|          text|favourite_count|retweet_count|created_at|negative_nltk|positive_nltk|neutral_nltk|compound_nltk|
+-----+-----+-----+-----+-----+-----+
|RT @JohnRHewson: ...|          0|          890|2021-06-19|          0.0|          0.0|          1.0|          0.0|
|RT @DrEricDing: 3...|          0|          55|2021-06-19|          0.0|          0.0|          1.0|          0.0|
+-----+-----+-----+-----+-----+-----+
```

<https://stackoverflow.com/questions/61608057/output-vader-sentiment-scores-in-columns-based-on-dataframe-rows-of-tweets>

```
[67]: pdf
```

```
[67]:
```

| | text | favourite_count | \ |
|------|---|-----------------|---|
| 0 | @breakfasttv I received the Astrazeneca vaccin... | 0 | |
| 1 | #Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJo... | 3 | |
| 2 | RT @TimWattsMP: Can you believe this?\n\nThe M... | 0 | |
| 3 | RT @DrEricDing: 5) One dose of the vaccine is ... | 0 | |
| 4 | RT @DrEricDing: 5) One dose of the vaccine is ... | 0 | |
| ... | ... | ... | |
| 6505 | @Fab485617452 @JohnRHewson I'm 73 and my wife ... | 6 | |
| 6506 | RT @kyle_minogue: @bjornradstrom Pfizer - 12-1... | 0 | |
| 6507 | RT @JohnRHewson: For the record, how many of o... | 0 | |
| 6508 | RT @JohnRHewson: For the record, how many of o... | 0 | |
| 6509 | RT @DrEricDing: 3) "After years of reading res... | 0 | |

| | retweet_count | created_at | \ |
|------|---------------|---------------------|---|
| 0 | 0 | 2021-06-20 23:59:37 | |
| 1 | 5 | 2021-06-20 23:59:33 | |
| 2 | 212 | 2021-06-20 23:58:46 | |
| 3 | 270 | 2021-06-20 23:57:54 | |
| 4 | 270 | 2021-06-20 23:56:29 | |
| ... | ... | ... | |
| 6505 | 0 | 2021-06-19 00:02:33 | |
| 6506 | 2 | 2021-06-19 00:02:21 | |
| 6507 | 890 | 2021-06-19 00:01:32 | |
| 6508 | 890 | 2021-06-19 00:00:27 | |
| 6509 | 55 | 2021-06-19 00:00:08 | |

| | rating |
|------|---|
| 0 | {'neg': 0.049, 'neu': 0.752, 'pos': 0.199, 'co... |
| 1 | {'neg': 0.0, 'neu': 0.603, 'pos': 0.397, 'comp... |
| 2 | {'neg': 0.124, 'neu': 0.876, 'pos': 0.0, 'comp... |
| 3 | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... |
| 4 | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... |
| ... | ... |
| 6505 | {'neg': 0.0, 'neu': 0.927, 'pos': 0.073, 'comp... |
| 6506 | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... |
| 6507 | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... |
| 6508 | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... |
| 6509 | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... |

[6510 rows x 5 columns]

```
[68]: pdf['negative_nltk']=i['neg'] for i in pdf.rating]
pdf['positive_nltk']=i['pos'] for i in pdf.rating]
pdf['neutral_nltk']=i['neu'] for i in pdf.rating]
pdf['ncompound_nltk']=i['compound'] for i in pdf.rating]

pdf.head(2)
```

```
[68]:
```

| | text | favourite_count | \ |
|---|---|-----------------|---|
| 0 | @breakfasttv I received the Astrazeneca vaccin... | 0 | |
| 1 | #Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJo... | 3 | |

| | retweet_count | created_at | \ |
|---|---------------|---------------------|---|
| 0 | 0 | 2021-06-20 23:59:37 | |
| 1 | 5 | 2021-06-20 23:59:33 | |

| | rating | negative_nltk | \ |
|---|---|---------------|---|
| 0 | {'neg': 0.049, 'neu': 0.752, 'pos': 0.199, 'co... | 0.049 | |
| 1 | {'neg': 0.0, 'neu': 0.603, 'pos': 0.397, 'comp... | 0.000 | |

| | positive_nltk | neutral_nltk | ncompound_nltk |
|---|---------------|--------------|----------------|
| 0 | 0.199 | 0.752 | 0.7793 |
| 1 | 0.397 | 0.603 | 0.7639 |

```
[69]: pdf['negative_nltk'] = pdf['rating'].apply(lambda x : x['neg'])
pdf['positive_nltk'] = pdf['rating'].apply(lambda x : x['pos'])
pdf['neutral_nltk'] = pdf['rating'].apply(lambda x : x['neu'])
pdf['compound_nltk'] = pdf['rating'].apply(lambda x : x['compound'])

pdf = pdf.drop('rating', axis=1)
pdf.head(2)
```

```
[69]:
```

| | text | favourite_count | \ |
|---|---|-----------------|---|
| 0 | @breakfasttv I received the Astrazeneca vaccin... | 0 | |
| 1 | #Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJo... | 3 | |

| | retweet_count | created_at | negative_nltk | positive_nltk | \ |
|---|---------------|---------------------|---------------|---------------|---|
| 0 | 0 | 2021-06-20 23:59:37 | 0.049 | 0.199 | |
| 1 | 5 | 2021-06-20 23:59:33 | 0.000 | 0.397 | |

| | neutral_nltk | ncompound_nltk |
|---|--------------|----------------|
| 0 | 0.752 | 0.7793 |
| 1 | 0.603 | 0.7639 |

```
[70]: pdf['created_at'][pdf.shape[0]-1]
```

```
[70]: '2021-06-19 00:00:08'
```

```
[71]: pdf['text'].count()
```

```
[71]: 6510
```

```
[72]: pdf.count()
```



```
[72]: text                6510
      favourite_count    6510
      retweet_count      6510
      created_at         6510
      negative_nltk       6510
      positive_nltk       6510
      neutral_nltk        6510
      ncompound_nltk      6510
      dtype: int64
```

```
[73]: pdf.count(axis=0)
```

```
[73]: text                6510
      favourite_count    6510
      retweet_count      6510
      created_at         6510
      negative_nltk       6510
      positive_nltk       6510
      neutral_nltk        6510
      ncompound_nltk      6510
      dtype: int64
```

```
[74]: pdf.shape[0]
```

```
[74]: 6510
```

```
[75]: len(pdf.index)
```

```
[75]: 6510
```

4 Spark GroupBy and Aggregate Functions

GroupBy allows you to group rows together based off some column value, for example, you could group together sales data by the day the sale occurred, or group repeat customer data based off the name of the customer.

Once you've performed the GroupBy operation you can use an aggregate function off that data. An **aggregate function** aggregates multiple rows of data into a single output, such as taking the sum of inputs, or counting the number of inputs.

Dataframe Aggregation

A set of methods for aggregations on a DataFrame:

```
agg
avg
count
```

```
max
mean
min
pivot
sum
```

<https://hendra-herviawan.github.io/pyspark-groupby-and-aggregate-functions.html>

```
[77]: sdf.columns
```

```
[77]: ['text',
      'favourite_count',
      'retweet_count',
      'created_at',
      'negative_nltk',
      'positive_nltk',
      'neutral_nltk',
      'compound_nltk']
```

```
[108]: sdf.select('created_at')
```

```
[108]: DataFrame[created_at: date]
```

```
5 sdf_agg_byDate = sdf.groupBy('created_at').agg({'negative_nltk':'sum',
      {'positive_nltk':'sum'},      {'neutral_nltk':'sum'},      {'com-
      pound_nltk':'sum'})
```

```
import pyspark.sql.functions as f from pyspark.sql.functions import col
```

```
sdf_agg_byDate = sdf.groupBy('created_at').agg({'negative_nltk':'sum','positive_nltk':'sum','neutral_nltk':'sum',
```

```
sdf_agg_byDate.count()
```

```
[79]: # sdf.groupBy('created_at').count().select('created_at', f.col('count').
      ↪alias('tweets')).show()
```

```
[81]: (sdf
      .groupBy('created_at')
      .agg(f.count('created_at').alias('tweets'))
      .show())
```

```
+-----+-----+
|created_at|tweets|
+-----+-----+
|2021-06-20|  2407|
|2021-06-19|  4103|
+-----+-----+
```

```
[82]: sdf_agg_byDate.show()
```

```
+-----+-----+-----+-----+-----+
+-----+-----+
|created_at|
sum(compound_nltk)|sum(positive_nltk)|count(created_at)|sum(negative_nltk)|
sum(neutral_nltk)|
+-----+-----+-----+-----+-----+
+-----+-----+
|2021-06-20| -36.288300000000009| 112.9250000000001|          2407|
133.614|2160.4989999999916|
|2021-06-19|-106.83589999999874| 139.2199999999993|
4103|176.492000000000073|3787.2440000000083|
+-----+-----+-----+-----+-----+
+-----+-----+
```

```
[ ]: # sdf_agg_byDate = sdf_agg_byDate.withColumn("sum",
↳ col("sum(negative_nltk)") + col("science_score"))
#      df1.show()
```

```
[83]: # How to delete columns in pyspark dataframe
columns_to_drop = ['sum(compound_nltk)', 'sum(positive_nltk)',
↳ 'sum(negative_nltk)', 'sum(neutral_nltk)']
```

```
sdf_agg_byDate = (sdf_agg_byDate
    .withColumn('compound_nltk', f.col('sum(compound_nltk)')/f.
↳ col('count(created_at)'))
    .withColumn('positive_nltk', f.col('sum(positive_nltk)')/f.
↳ col('count(created_at)'))
    .withColumn('negative_nltk', f.col('sum(negative_nltk)')/f.
↳ col('count(created_at)'))
    .withColumn('neutral_nltk', f.col('sum(neutral_nltk)')/f.
↳ col('count(created_at)'))
    .withColumnRenamed('count(created_at)', 'tweets')).drop(*columns_to_drop)

sdf_agg_byDate.show()
```

```
+-----+-----+-----+-----+-----+
+-----+-----+
|created_at|tweets|      compound_nltk|      positive_nltk|
negative_nltk|      neutral_nltk|
+-----+-----+-----+-----+-----+
+-----+-----+
|2021-06-20|  2407|-0.01507615288741...|0.046915247195679306|0.05551059410054009
4|0.8975899459908565|
|2021-06-19|
```

```
4103|-0.02603848403607...|0.033931269802583305|0.043015354618571956|
0.923042651718257|
```

```
+-----+-----+-----+-----+-----+-----+
-+-----+-----+-----+-----+-----+-----+
```

```
[97]: sdf_agg_byDate.toPandas()
      # .drop(*columns_to_drop)
```

```
[97]:   created_at  count(created_at)  compound_nltk  positive_nltk  negativen_ltk  \
0   2021-06-20             2407      -0.015076      0.046915      0.055511
1   2021-06-19             4103      -0.026038      0.033931      0.043015

      neutral_nltk
0      0.897590
1      0.923043
```

```
[101]: sdf.created_at[0]
```

```
[101]: Column<b'created_at[0] '>
```

```
[84]: # The below statement changes column 'count(created_at)' to 'tweets' on PySpark
      ↪ DataFrame.
sdf_agg_byDate = (sdf_agg_byDate
                  .withColumnRenamed('count(created_at)', 'tweets'))
```

```
[85]: sdf_agg_byDate.toPandas()
```

```
[85]:   created_at  tweets  compound_nltk  positive_nltk  negativen_ltk  \
0   2021-06-20     2407      -0.015076      0.046915      0.055511
1   2021-06-19     4103      -0.026038      0.033931      0.043015

      neutral_nltk
0      0.897590
1      0.923043
```

```
[218]: # DataFrame.sort(*cols, **kwargs) - Returns a new DataFrame sorted by the
      ↪ specified column(s).
sdf_agg_byDate = sdf_agg_byDate.sort('created_at')

sdf_agg_byDate.toPandas()
```

```
[218]:   created_at  tweets  compound_nltk  positive_nltk  negativen_ltk  \
0   2021-06-19     4103      -0.026038      0.033931      0.043015
1   2021-06-20     2407      -0.015076      0.046915      0.055511

      neutral_nltk
```

```
0      0.923043
1      0.897590
```

```
[224]: #
sdf_agg_byDate.dtypes
```

```
[224]: [('created_at', 'date'),
        ('tweets', 'bigint'),
        ('compound_nltk', 'double'),
        ('positive_nltk', 'double'),
        ('negativen_ltk', 'double'),
        ('neutral_nltk', 'double')]
```

```
[225]: sdf_agg_byDate.head(1)
```

```
[225]: Row(created_at=datetime.date(2021, 6, 19), tweets=4103,
           compound_nltk=-0.026038484036070862, positive_nltk=0.033931269802583305,
           negativen_ltk=0.043015354618571956, neutral_nltk=0.923042651718257)
```

```
[226]: sdf_agg_byDate.limit(1).show()
```

```
+-----+-----+-----+-----+-----+
+-----+
|created_at|tweets|      compound_nltk|      positive_nltk|
negativen_ltk|      neutral_nltk|
+-----+-----+-----+-----+-----+
+-----+
|2021-06-19|  4103|-0.02603848403607...|0.033931269802583305|0.04301535461857195
6|0.923042651718257|
+-----+-----+-----+-----+-----+
+-----+
```

```
[ ]:
```

```
[223]: # last, head, tail
sdf_agg_byDate.first()
```

```
[223]: Row(created_at=datetime.date(2021, 6, 19), tweets=4103,
           compound_nltk=-0.026038484036070862, positive_nltk=0.033931269802583305,
           negativen_ltk=0.043015354618571956, neutral_nltk=0.923042651718257)
```

```
[ ]: import datetime, time
# dates = ("2021-06-19 00:00:00", "2021-06-20 00:00:00")
date_from, date_to = ("2021-06-19", "2021-06-20")

timestamps = (
```

```
time.mktime(datetime.datetime.strptime(s, "%Y-%m-%d %H:%M:%S").timetuple())
for s in dates)
```

```
sdf_agg_byDate.filter(col('created_at') > )
```

```
[ ]:
```

```
[220]: # sqlContext.sql("set spark.sql.shuffle.partitions=10")
```

```
spark.sql.functions.shuffle.partitions
```

```
-----
AttributeError                                Traceback (most recent call last)
<ipython-input-220-bfd28aba44c4> in <module>
      2
      3
----> 4 spark.sql.shuffle.partitions

AttributeError: 'function' object has no attribute 'shuffle'
```

5.1 Pandas groupBy and aggregate functions

```
[88]: # Change name of a specific column
# pdf = pdf.rename(columns={'ncompound_nltk': 'compound_nltk'})
# pdf.head(2)
```

```
[88]:
```

| | | text | favourite_count | \ |
|---|---|--------------------------------------|-----------------|---|
| 0 | @breakfasttv | I received the Astrazeneca vaccin... | 0 | |
| 1 | #Pfizer\n#AstraZeneca\n#Moderna\n#JohnsonAndJo... | | 3 | |

| | retweet_count | created_at | negative_nltk | positive_nltk | \ |
|---|---------------|---------------------|---------------|---------------|---|
| 0 | 0 | 2021-06-20 23:59:37 | 0.049 | 0.199 | |
| 1 | 5 | 2021-06-20 23:59:33 | 0.000 | 0.397 | |

| | neutral_nltk | compound_nltk |
|---|--------------|---------------|
| 0 | 0.752 | 0.7793 |
| 1 | 0.603 | 0.7639 |

```
[202]: pdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6510 entries, 0 to 6509
```

```
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   text                   6510 non-null  object
1   favourite_count        6510 non-null  int64
2   retweet_count          6510 non-null  int64
3   created_at             6510 non-null  object
4   negative_nltk          6510 non-null  float64
5   positive_nltk          6510 non-null  float64
6   neutral_nltk           6510 non-null  float64
7   compound_nltk          6510 non-null  float64
dtypes: float64(4), int64(2), object(2)
memory usage: 407.0+ KB
```

```
[203]: pdf['created_at'] = pd.to_datetime(pdf['created_at'])
```

```
[204]: pdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6510 entries, 0 to 6509
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   text                   6510 non-null  object
1   favourite_count        6510 non-null  int64
2   retweet_count          6510 non-null  int64
3   created_at             6510 non-null  datetime64[ns]
4   negative_nltk          6510 non-null  float64
5   positive_nltk          6510 non-null  float64
6   neutral_nltk           6510 non-null  float64
7   compound_nltk          6510 non-null  float64
dtypes: datetime64[ns](1), float64(4), int64(2), object(1)
memory usage: 407.0+ KB
```

```
[165]: pdf.head(2)
```

```
[165]:
```

| | | text | favourite_count | \ |
|---|--------------|--|-----------------|---|
| 0 | @breakfasttv | I received the Astrazeneca vaccin... | 0 | |
| 1 | #Pfizer | #AstraZeneca\n#Moderna\n#JohnsonAndJo... | 3 | |

| | retweet_count | created_at | negative_nltk | positive_nltk | neutral_nltk | \ |
|---|---------------|------------|---------------|---------------|--------------|---|
| 0 | 0 | 2021-06-20 | 0.049 | 0.199 | 0.752 | |
| 1 | 5 | 2021-06-20 | 0.000 | 0.397 | 0.603 | |

| | compound_nltk |
|---|---------------|
| 0 | 0.7793 |
| 1 | 0.7639 |

```
[205]: pdf['created_at'] = pdf['created_at'].dt.strftime('%Y-%m-%d')
```

```
[206]: pdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6510 entries, 0 to 6509
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   text                   6510 non-null   object
1   favourite_count        6510 non-null   int64
2   retweet_count          6510 non-null   int64
3   created_at             6510 non-null   object
4   negative_nltk           6510 non-null   float64
5   positive_nltk           6510 non-null   float64
6   neutral_nltk            6510 non-null   float64
7   compound_nltk           6510 non-null   float64
dtypes: float64(4), int64(2), object(2)
memory usage: 407.0+ KB
```

```
[174]: # For a DataFrame, by default the aggregates return results within each column:
# pdf['created_at'].dt.date => Converts Datetime to Date in Pandas df
pdf_agg_byDate = pdf.groupby('created_at').agg({'negative_nltk':
↪ 'sum', 'positive_nltk': 'sum', 'neutral_nltk': 'sum', 'compound_nltk': 'sum',
↪ 'created_at': 'size'})

pdf_agg_byDate.count()
```

```
[174]: negative_nltk      2
positive_nltk      2
neutral_nltk       2
compound_nltk      2
created_at         2
dtype: int64
```

```
[175]: pdf_agg_byDate
```

```
[175]:      negative_nltk  positive_nltk  neutral_nltk  compound_nltk  \
created_at
2021-06-19      176.492      139.220      3787.244      -106.8359
2021-06-20      133.614      112.925      2160.499       -36.2883

      created_at
created_at
2021-06-19      4103
2021-06-20      2407
```



```
[209]: pdf_agg_byDate = pdf.groupby('created_at').
      ↪agg(negative_nltk=('negative_nltk','sum'),positive_nltk=('positive_nltk','sum'),neutral_nltk=('neutral_nltk','sum'))
      ↪tweets = ('created_at','size'))

pdf_agg_byDate.reset_index(level=0, inplace=True)
pdf_agg_byDate.count()
```

```
[209]: created_at      2
      negative_nltk    2
      positive_nltk    2
      neutral_nltk     2
      compound_nltk     2
      tweets          2
      dtype: int64
```

```
[210]: # row 0
      type(pdf.created_at[0])
```

```
[210]: str
```

```
[211]: # row 5
      type(pdf.created_at[5])
```

```
[211]: str
```

```
[212]: pdf_agg_byDate
```

```
[212]:   created_at  negative_nltk  positive_nltk  neutral_nltk  compound_nltk  \
0  2021-06-19         176.492         139.220         3787.244         -106.8359
1  2021-06-20         133.614         112.925         2160.499          -36.2883

      tweets
0         4103
1         2407
```

```
[193]: pdf_agg_byDate.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2 entries, 0 to 1
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   created_at      2 non-null     object
1   negative_nltk    2 non-null     float64
2   positive_nltk    2 non-null     float64
3   neutral_nltk     2 non-null     float64
4   compound_nltk    2 non-null     float64
```

```

5    tweets          2 non-null      int64
dtypes: float64(4), int64(1), object(1)
memory usage: 224.0+ bytes

```

5.1.1 Divide multiple columns by another column in pandas

```

[216]: # pdf_agg_byDate = (pdf_agg_byDate.apply(['compound_nltk']/
→pdf_agg_byDate['tweets']
#       .withColumn('positive_nltk',f.col('sum(positive_nltk)')/f.
→col('count(created_at)'))
#       .withColumn('negative_nltk',f.col('sum(negative_nltk)')/f.
→col('count(created_at)'))
#       .withColumn('neutral_nltk',f.col('sum(neutral_nltk)')/f.
→col('count(created_at)'))
#       .withColumnRenamed('count(created_at)', 'tweets')).drop(*columns_to_drop

pdf_agg_byDate[['negative_nltk','positive_nltk','neutral_nltk','compound_nltk']] = \
    ↵
→(pdf_agg_byDate[['negative_nltk','positive_nltk','neutral_nltk','compound_nltk']]
→divide(pdf_agg_byDate['tweets'], axis = 'index'))

pdf_agg_byDate

```

```

[216]:   created_at  negative_nltk  positive_nltk  neutral_nltk  compound_nltk  \
0   2021-06-19         0.043015         0.033931         0.923043        -0.026038
1   2021-06-20         0.055511         0.046915         0.897590        -0.015076

      tweets
0         4103
1         2407

```

```

[208]: pdf_agg_byDate

```

```

[208]:   negative_nltk  positive_nltk  neutral_nltk  compound_nltk
0         0.043015         0.033931         0.923043        -0.026038
1         0.055511         0.046915         0.897590        -0.015076

```

6 perfplot

```

[68]: import numpy as np
import pandas as pd
import perfplot

perfplot.save(
    "out.png",

```

```
setup=lambda n: pd.DataFrame(np.arange(n * 3).reshape(n, 3)),
n_range=[2**k for k in range(25)],
kernels=[
    lambda df: len(df.index),
    lambda df: df.shape[0],
    lambda df: df[df.columns[0]].count(),
],
labels=["len(df.index)", "df.shape[0]", "df[df.columns[0]].count()"],
xlabel="Number of rows",
)
```

Output()

[]: