CBD 3335 6 [B123]

Data Mining and Analysis

Assignment-1

Submitted To - Mohammad Islam

Sai Srikanth Raju	**C0846551**
Shweta Yadav	**C0854479**
Thakshak Revanth	**C846100**
Sandra Nicolas	**C0851356**

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Collecting tweets related to the stock market

*Note: Sub-titles are not captured in Xplore and should not be used

line 1: 1st Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 2nd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 3rd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 4th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 5th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 6th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

***Abstract*—This assignment is based on Twitter data analysis. We are downloading tweets of some specific keywords.**

1. Collecting data: In this assignment, we are collecting data related to the stock market from Twitter for one week. On Twitter, ticker symbols are used for stocks and companies. We are using keywords Altcoin, Bitcoin, Coindesk, Cryptocurrency,Gold, APPL, GOOG and YHOO

2. Saving data: You need to save the requested data into csv format of 8 files where data related to each keyword is saved. Each file consist of four columns: tweet id, time of tweet, user id, and text.

3. Cleaning data: remove duplication, remove punctuations, remove numbers in tweets, and remove words with lengths less than 2.

4. Visualizing data: You need to present the daily number of tweets for each keyword as well as the daily number of users. Use Clustering of similar tweets if feasible and applicable

I. Introduction

Python is our preferred programming language, which we use for various purposes (API connections, Modelling, Data engineering). We are creating a Twitter API call in Python to download stock market keywords as a hashtag. We downloaded data and cleaned it with Python (library functions). In addition, we have included a basic visualization of the cleaned data.

For twitter analysis, Information is collected by either the user, the access point, what's in the post, and how users view or use your post. Using this information, we can understand demographics, total view on your profile or how many people have seen a person's Tweet.

.

II. Fetching data from Twitter

A. Importing libraries

```python
import datetime
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import re
import time
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import nltk
import tweepy
import configparser
```

Fig. 1. Importing libraries

Tweepy**:** Tweepy is an open-source, user-friendly Python library for interacting with the Twitter API.

Time**:** This module offers a variety of time-related functions.

Pandas**:** Pandas is an open-source data analysis and manipulation tool that is fast, powerful, flexible, and straightforward. It is built on top of the Python programming language.

Textblob**:** Textblob is a text processing library. It provides a straightforward API for delving into standard natural language processing (NLP) tasks like part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

Wordcloud: a data visualization technique for representing text data in which the size of each word indicates the frequency or importance of that word.

Numpy**:** Numpy is a Python programming language open-source library. It is used in scientific computing and array manipulation.

Re: The re module provides a set of powerful regular expression facilities that allow you to quickly check whether a given string matches or contains a given pattern (via the match function) (using the search function).

B. Using key essentials

Using Keys from the Twitter developer account,we can download tweets related to our keywords.

```
api_key = "APEDWnbf8cO64ui02UB6YXuEx"
api_key_secret = "WXk3iBxdOaaQspDDhfWSXdYQoBZoqnHgsCKT6yyNFlxgV6xTUW"

access_token = "1526337205337202689-aMQw7oSLYmx0lo1szlJ9wDU2JHsV9q"
access_token_secret = "1KxurDw0K9z8c8HH6rJSpon1HAmuhirELXvp8Nbupfq9L"

# authentication
auth = tweepy.OAuthHandler(api_key, api_key_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth, wait_on_rate_limit=True,wait_on_rate_limit_notify=True)
```

Fig. 2. Using keys for fetching tweets

We are creating the access token and access token secret and making the API object while passing in the auth information.

C.Fetch market data

We are mining stock market data here by creating a function fetch_data (). We obtain tweets for all eight tickers by storing them in a list and iterating over it. We get the tweets' creation date, username, tweet and location, and keywords, i.e., Tickers.

```
In [7]: def fetch_data():

    # Tickers on what we search the tweets
    tickers = ['#bitcoin', '#gold','#Altcoin', '#Coindesk','#Cryptocurrency','#APPL','#GOOG','#YHOO']
    # Present the limit is set to 10 to save the computational time
    limit=10

    # Creating a datafram in wich the tweets extracted are stored
    df = pd.DataFrame(columns = ['Created At', 'User', 'Tweet', 'Location', 'Tickers'])
    data = []
    # Specifing the duration of time span in which the tweets are extracted
    date_list=["2022-07-01","2022-07-02","2022-07-03","2022-07-04","2022-07-05","2022-07-06","2022-07-07"]
    #Extracting the tweets
    for ticker in tickers:
        for date in date_list:

            tweets = tweepy.Cursor(api.search, q=ticker,until=date,lang="en", tweet_mode='extended').items(limit)
            # Defining the columns
            columns = ['Created At', 'User', 'Tweet', 'Location']

            for tweet in tweets:
                # Append all tweets for the ticker
                data.append([tweet.created_at, tweet.user.screen_name,tweet.full_text,tweet.user.location])
                print(tweet.created_at)
            print(data)
            ticker_df = pd.DataFrame(data, columns=columns) # All tweets for 1 ticker

            # Generate new column to classify tweets and ticker
            query_labels = pd.Series([ticker]).repeat(len(ticker_df))
            query_labels_df = query_labels.to_frame(name="Tickers").reset_index(drop=True)

            intermed_ticker_df = pd.concat([ticker_df, query_labels_df], axis=1)
            time.sleep(30)
            # Append to final df
            df = pd.concat([df, intermed_ticker_df], axis=0)
    return df
```

Fig. 3. Fetching tweets for specific keywords

III. Data Cleaning

Data is cleaned by removing duplicates, punctuation marks, and words of length less than 2.

```
# Create a function to clean the tweets
def cleanTxt(text):
    text = re.sub('@[A-Za-z0-9]+', '', text) #Removing @mentions
    text = re.sub('#', '', text) # Removing '#' hash tag
    text = re.sub(r'\W*\b\w{1,2}\b', '', text) # Removing wordls less than length of 2
    text = re.sub('RT[\s]+', '', text) # Removing RT
    text = re.sub('https?:\/\/\S+', '', text) # Removing hyperlink
    return text
df3["Tweet"] = df3["Tweet"].apply(cleanTxt)
df3
```

```
tickers = ['#bitcoin', '#gold','#Altcoin', '#Coindesk','#Cryptocurrency','#APPL','#GOOG','#YHOO']
df3['Created At'] = pd.to_datetime(df3['Created At'])

start_date='2022-07-01'

end_date='2022-07-07'

tweets_in_past_day = []

for ticker in tickers:

    mask = (df3['Created At'] > start_date) & (df3['Created At'] <= end_date) & (df3['Tickers'] == ticker)
    num_tweets = len(df3.loc[mask])
    tweets_in_past_day.append(num_tweets)
```

```
C:\Users\thaks\AppData\Local\Temp/ipykernel_676/2577241891.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy
  df3['Created At'] = pd.to_datetime(df3['Created At'])
```

Fig. 4. Data cleaning

IV. Data Visualization

A. The number of tweets in the past day

```
tweets_in_past_day
```

```
[60, 120, 180, 240, 295, 325, 381, 381]
```

```
visualization_df = pd.DataFrame({'Tickers':tickers, 'Number of tweets':tweets_in_past_day})
ax = visualization_df.plot.bar(x='Tickers', y='Number of tweets', rot=90)
```

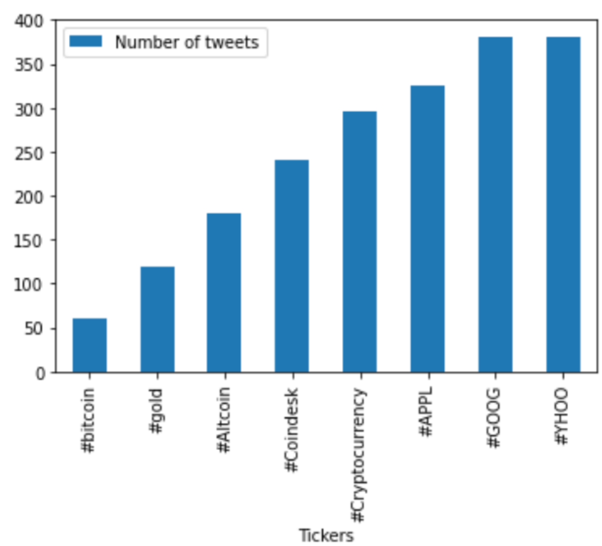

Fig. 5. Number of tweets in the past days

B.Number of users tweeting in a day/hour for ticker symbol

```
start_date='2022-07-01'

end_date='2022-07-07'

distinct_users_in_past_day = []

for ticker in tickers:
    mask = (df3['Created At'] > start_date) & (df3['Created At'] <= end_date) & (df3['Tickers'] == ticker)
    num_users = df3.loc[mask]
    n = len(pd.unique(num_users['User']))
    distinct_users_in_past_day.append(n)
```

```
visualization2 = pd.DataFrame({'Tickers':tickers, 'Number of distinct users':distinct_users_in_past_day})
ax = visualization2.plot.bar(x='Tickers', y='Number of distinct users', rot=90)
```

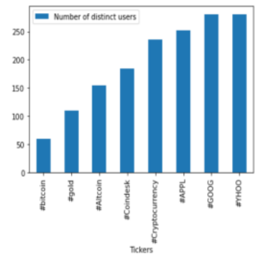

Fig. 6. Number of users tweeting a day

V. DOWNLOADING INDIVIDUAL DATA FRAME

Now, we are going to download individual dataset files for each and every keyword by converting it into CSV files.

A. *Individual Bitcoin Dataset*

```
bitcoin=df3[df3['Tickers']=="#bitcoin"]
```

```
bitcoin
```

	Created At	User	Tweet	Location	Tickers	Created_Date
0	2022-06-30 23:59:59	EngieFord	RT @IncomeSharks: From July 1st to November 1s...		#bitcoin	2022-06-30
1	2022-06-30 23:59:59	DeepStar22	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Delhi, India	#bitcoin	2022-06-30
2	2022-06-30 23:59:57	nerivansouza01	RT @milkshake_io: MicroStrategy chief executiv...		#bitcoin	2022-06-30
3	2022-06-30 23:59:56	EngieFord	RT @CedYoungelman: Your home is not worth more...		#bitcoin	2022-06-30
4	2022-06-30 23:59:56	AndersTintin	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Norway, Alberta	#bitcoin	2022-06-30
...	...	...	...	...	...	...
65	2022-07-06 23:59:51	HMitke	RT @saylor: "Bitcoin is a monetary instrument ...		#bitcoin	2022-07-06
66	2022-07-06 23:59:50	stillbtc	RT @aiproenergy: Immersion fluid is more effic...	Thinking Green	#bitcoin	2022-07-06
67	2022-07-06 23:59:48	hann_marie22	RT @MartiniGuyYT: If #Bitcoin doesn't hit $26,...	507 proofs	#bitcoin	2022-07-06
68	2022-07-06 23:59:46	daleperozzo	RT @TheCryptoLark: How Investing In Crypto Act...	Earth	#bitcoin	2022-07-06
69	2022-07-06 23:59:45	nftaddie	RT @cz_binance: Banks now use #bitcoin for res...	Salt Lake City	#bitcoin	2022-07-06

70 rows × 6 columns

```
bitcoin=df3[df3['Tickers']=="#bitcoin"]
bitcoin.to_csv('bitcoin.csv')
```

Fig. 7. Bitcoin dataset

B. *Individual dataset for Gold*

	Created At	User	Tweet	Location	Tickers	Created_Date
0	2022-06-30 23:59:59+00:00	EngieFord	RT @IncomeSharks: From July 1st to November 1s...		#gold	2022-06-30
1	2022-06-30 23:59:59+00:00	DeepStar22	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Delhi, India	#gold	2022-06-30
2	2022-06-30 23:59:57+00:00	nerivansouza01	RT @milkshake_io: MicroStrategy chief executiv...		#gold	2022-06-30
3	2022-06-30 23:59:56+00:00	EngieFord	RT @CedYoungelman: Your home is not worth more...		#gold	2022-06-30
4	2022-06-30 23:59:56+00:00	AndersTintin	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Norway, Alberta	#gold	2022-06-30
...	...	...	...	...	...	...
135	2022-07-06 23:55:56+00:00	wojtekcrypto08	#Gold: Barrick Gold Corp (GOLD) closed today a...		#gold	2022-07-06
136	2022-07-06 23:55:44+00:00	muzzleloaderman	RT @roxiewin7: Excited to share this item from...	Oregon, USA	#gold	2022-07-06
137	2022-07-06 23:55:03+00:00	TalkMarkets	Rare #Gold-#Silver Crystal Sighting GLDSLV ...		#gold	2022-07-06
138	2022-07-06 23:54:05+00:00	BHC_Valentine	#Hopping into some #hydroneer #earlyaccess #Ga...		#gold	2022-07-06
139	2022-07-06 23:53:01+00:00	SalvadorMaurice	$STK \n\nMULTIPLE NEW #GOLD ZONES AND HIGH-GRA...	Bali Indonesia	#gold	2022-07-06

140 rows × 6 columns

Fig. 8. Gold dataset

C. *Altcoin*

```
Altcoin=df3[df3['Tickers']=="#Altcoin"]
Altcoin.to_csv('Altcoin.csv')
Altcoin
```

	Created At	User	Tweet	Location	Tickers	Created_Date
0	2022-06-30 23:59:59+00:00	EngieFord	RT @IncomeSharks: From July 1st to November 1s...		#Altcoin	2022-06-30
1	2022-06-30 23:59:59+00:00	DeepStar22	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Delhi, India	#Altcoin	2022-06-30
2	2022-06-30 23:59:57+00:00	nerivansouza01	RT @milkshake_io: MicroStrategy chief executiv...		#Altcoin	2022-06-30
3	2022-06-30 23:59:56+00:00	EngieFord	RT @CedYoungelman: Your home is not worth more...		#Altcoin	2022-06-30
4	2022-06-30 23:59:56+00:00	AndersTintin	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Norway, Alberta	#Altcoin	2022-06-30
...	...	...	...	...	...	...
205	2022-07-06 23:58:24+00:00	Rosemar97946025	RT @metaboxmarket: May will be bullish? \nTell...		#Altcoin	2022-07-06
206	2022-07-06 23:57:41+00:00	Cryptoinsight14	@cryptojack The only #altcoin I would buy from...		#Altcoin	2022-07-06
207	2022-07-06 23:55:15+00:00	RyiPay	RT @RyiPay: The #Cryptocurrency of Choice for ...	Los Angeles, California, USA	#Altcoin	2022-07-06
208	2022-07-06 23:55:08+00:00	ErsoganRecep	RT @ENQ_enecuum: We remind you that the Enecuu...		#Altcoin	2022-07-06
209	2022-07-06 23:53:28+00:00	H2bRojas	RT @AAdsabdo: @MrBigWhaleREAL #Saitama is the ...		#Altcoin	2022-07-06

210 rows × 6 columns

Fig. 9. Altcoin dataset

D. *Coindesk*

```
Coindesk=df3[df3['Tickers']=="#Coindesk"]
Coindesk.to_csv('Coindesk.csv')
Coindesk
```

	Created At	User	Tweet	Location	Tickers	Created_Date
0	2022-06-30 23:59:59+00:00	EngieFord	RT @IncomeSharks: From July 1st to November 1s...		#Coindesk	2022-06-30
1	2022-06-30 23:59:59+00:00	DeepStar22	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Delhi, India	#Coindesk	2022-06-30
2	2022-06-30 23:59:57+00:00	nerivansouza01	RT @milkshake_io: MicroStrategy chief executiv...		#Coindesk	2022-06-30
3	2022-06-30 23:59:56+00:00	EngieFord	RT @CedYoungelman: Your home is not worth more...		#Coindesk	2022-06-30
4	2022-06-30 23:59:56+00:00	AndersTintin	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Norway, Alberta	#Coindesk	2022-06-30
...	...	...	...	...	...	...
275	2022-07-06 13:57:37+00:00	codedtrader	BTC/USDT 4H symmetrical triangle. A breakout i...		#Coindesk	2022-07-06
276	2022-07-06 13:57:04+00:00	QvrseDAO	We recommend this article about #DAO #Web3 #La...	Republic of Slovenia	#Coindesk	2022-07-06
277	2022-07-06 13:26:53+00:00	The_Cryptostan	Tuesday saw the return of Brazilian real depos...		#Coindesk	2022-07-06
278	2022-07-06 12:00:03+00:00	BTCticker	One Bitcoin now worth $20086.836. MarketCap$...		#Coindesk	2022-07-06
279	2022-07-06 11:07:43+00:00	CryptoServu	Bitcoin Recovers Above $20K as Small ETF Sees ...		#Coindesk	2022-07-06

280 rows × 6 columns

Fig. 10. Coindesk dataset

D. *APPL*

```
APPL=df3[df3['Tickers']=="#APPL"]
APPL.to_csv('APPL.csv')
APPL
```

	Created At	User	Tweet	Location	Tickers	Created_Date
0	2022-06-30 23:59:59+00:00	EngieFord	RT @IncomeSharks: From July 1st to November 1s...		#APPL	2022-06-30
1	2022-06-30 23:59:59+00:00	DeepStar22	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Delhi, India	#APPL	2022-06-30
2	2022-06-30 23:59:57+00:00	nerivansouza01	RT @milkshake_io: MicroStrategy chief executiv...		#APPL	2022-06-30
3	2022-06-30 23:59:56+00:00	EngieFord	RT @CedYoungelman: Your home is not worth more...		#APPL	2022-06-30
4	2022-06-30 23:59:56+00:00	AndersTintin	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Norway, Alberta	#APPL	2022-06-30
...	...	...	...	...	...	...
409	2022-07-06 06:07:27+00:00	im_ba1tazar	@gandalfcryptto Love ❤ #AITX #ai #APPL #MSFT ...		#APPL	2022-07-06
410	2022-07-06 05:06:15+00:00	PyScaleLLC	RT @im_ba1tazar: @BabyCatCoin @cz_binance Love...	Miami, FL	#APPL	2022-07-06
411	2022-07-06 05:06:14+00:00	PyScaleLLC	RT @im_ba1tazar: Love ❤ #AITX #ai #APPL #MSFT...	Miami, FL	#APPL	2022-07-06
412	2022-07-06 05:06:00+00:00	PyScaleLLC	RT @im_ba1tazar: @jesmotril @CheemsInu Same! L...	Miami, FL	#APPL	2022-07-06
413	2022-07-06 05:01:38+00:00	im_ba1tazar	@jesmotril @CheemsInu Same! Love ❤ #AITX #ai ...		#APPL	2022-07-06

378 rows × 6 columns

Fig. 11. APPL dataset

E. *Cryptocurrency*

```
Cryptocurrency=df3[df3['Tickers']=="#Cryptocurrency"]
Cryptocurrency.to_csv('Cryptocurrency.csv')
Cryptocurrency
```

	Created At	User	Tweet	Location	Tickers	Created_Date
0	2022-06-30 23:59:59+00:00	EngieFord	RT @IncomeSharks: From July 1st to November 1s...		#Cryptocurrency	2022-06-30
1	2022-06-30 23:59:59+00:00	DeepStar22	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Delhi, India	#Cryptocurrency	2022-06-30
2	2022-06-30 23:59:57+00:00	nerivansouza01	RT @milkshake_io: MicroStrategy chief executiv...		#Cryptocurrency	2022-06-30
3	2022-06-30 23:59:56+00:00	EngieFord	RT @CedYoungelman: Your home is not worth more...		#Cryptocurrency	2022-06-30
4	2022-06-30 23:59:56+00:00	AndersTintin	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Norway, Alberta	#Cryptocurrency	2022-06-30
...	...	...	...	...	...	...
345	2022-07-06 23:59:44+00:00	Fatih66864011	RT @metablackmoon: 10.000$ BUSD GIVEAWAY \n\...		#Cryptocurrency	2022-07-06
346	2022-07-06 23:59:43+00:00	WARWOLF62189157	RT @ta64945087: #Saitama development team abou...		#Cryptocurrency	2022-07-06
347	2022-07-06 23:59:37+00:00	nourimounir409	RT @PlayWizardia: All this talk of listings, g...		#Cryptocurrency	2022-07-06
348	2022-07-06 23:59:37+00:00	KenanGlolu16	RT @coinlivetoday: 🚀🚀🚀 ‼ Breaking News ‼ Mas...		#Cryptocurrency	2022-07-06
349	2022-07-06 23:59:34+00:00	pastibigwinkok	RT @GoCryptoMeCoin: We did it https://t.co/j6A...	where?	#Cryptocurrency	2022-07-06

344 rows × 6 columns

Fig. 12. Cryptocurrency dataset

F.GOOG

```
GOOG=df3[df3['Tickers']=="#GOOG"]
GOOG.to_csv('GOOG.csv')
GOOG
```

	Created At	User	Tweet	Location	Tickers	Created_Date
0	2022-06-30 23:59:59+00:00	EngieFord	RT @IncomeSharks: From July 1st to November 1s...		#GOOG	2022-06-30
1	2022-06-30 23:59:59+00:00	DeepStar22	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Delhi, India	#GOOG	2022-06-30
2	2022-06-30 23:59:57+00:00	nerivansouza01	RT @milkshake_io: MicroStrategy chief executiv...		#GOOG	2022-06-30
3	2022-06-30 23:59:56+00:00	EngieFord	RT @CedYoungelman: Your home is not worth more...		#GOOG	2022-06-30
4	2022-06-30 23:59:56+00:00	AndersTintin	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Norway, Alberta	#GOOG	2022-06-30
...	...	...	...	...	...	...
479	2022-07-06 13:23:47+00:00	leoventuresllc	Google is the internet !! #google #googl #goog...	Chicago, IL	#GOOG	2022-07-06
480	2022-07-06 12:53:46+00:00	leoventuresllc	Google...What Can I Say !! #google #googl #sto...	Chicago, IL	#GOOG	2022-07-06
481	2022-07-06 09:29:54+00:00	iob_fi	RT @IOB_llc: #Google #GOOG #TechnicalIndicator...	global	#GOOG	2022-07-06
482	2022-07-06 06:23:46+00:00	leoventuresllc	Google...What Can I Say !! #google #googl #sto...	Chicago, IL	#GOOG	2022-07-06
483	2022-07-06 02:30:14+00:00	IOB_llc	#Google #GOOG #TechnicalIndicator Market Types...	Shanghai	#GOOG	2022-07-06

444 rows × 6 columns

Fig. 13. GOOG dataset

G.YAHOO

```
YHOO=df3[df3['Tickers']=="#YHOO"]
YHOO.to_csv('YHOO.csv')
YHOO
```

	Created At	User	Tweet	Location	Tickers	Created_Date
0	2022-06-30 23:59:59+00:00	EngieFord	RT @IncomeSharks: From July 1st to November 1s...		#YHOO	2022-06-30
1	2022-06-30 23:59:59+00:00	DeepStar22	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Delhi, India	#YHOO	2022-06-30
2	2022-06-30 23:59:57+00:00	nerivansouza01	RT @milkshake_io: MicroStrategy chief executiv...		#YHOO	2022-06-30
3	2022-06-30 23:59:56+00:00	EngieFord	RT @CedYoungelman: Your home is not worth more...		#YHOO	2022-06-30
4	2022-06-30 23:59:56+00:00	AndersTintin	RT @OfficialTravlad: What tf is this #Bitcoin ...	New Norway, Alberta	#YHOO	2022-06-30
...	...	...	...	...	...	...
479	2022-07-06 13:23:47+00:00	leoventuresllc	Google is the internet !! #google #googl #goog...	Chicago, IL	#YHOO	2022-07-06
480	2022-07-06 12:53:46+00:00	leoventuresllc	Google...What Can I Say !! #google #googl #sto...	Chicago, IL	#YHOO	2022-07-06
481	2022-07-06 09:29:54+00:00	iob_fi	RT @IOB_llc: #Google #GOOG #TechnicalIndicator...	global	#YHOO	2022-07-06
482	2022-07-06 06:23:46+00:00	leoventuresllc	Google...What Can I Say !! #google #googl #sto...	Chicago, IL	#YHOO	2022-07-06
483	2022-07-06 02:30:14+00:00	IOB_llc	#Google #GOOG #TechnicalIndicator Market Types...	Shanghai	#YHOO	2022-07-06

444 rows × 6 columns

Fig. 14. YAHOO dataset

References

[1] https://dev.twitter.com/overview/documentation
[2] https://www.python.org/doc/