

```
In [1]: import pandas as pd
```

```
In [2]: import numpy as np
```

```
In [11]: import matplotlib.pyplot as plt
```

```
In [17]: dataset = pd.read_csv(r'C:\Users\Dell\dsbda\tweets.csv')
```

```
In [6]: dataset.head()
```

Out[6]:

	id	user_name	user_location	user_description	user_created	user_followe
0	1340539111971516416	Rachel Roh	La Crescenta-Montrose, CA	Aggregator of Asian American news; scanning di...	2009-04-08 17:52:46	4
1	1338158543359250433	Albert Fong	San Francisco, CA	Marketing dude, tech geek, heavy metal & '80s ...	2009-09-21 15:27:30	8
2	1337858199140118533	eliLTEU👉	Your Bed	heil, hydra 🙌👉	2020-06-25 23:30:28	
3	1337855739918835717	Charles Adler	Vancouver, BC - Canada	Hosting "CharlesAdlerTonight" Global News Radi...	2008-09-10 11:28:53	491
4	1337854064604966912	Citizen News Channel	NaN	Citizen News Channel bringing you an alternati...	2020-04-23 17:58:42	1

In [7]: `dataset.tail()`

Out[7]:

	id	user_name	user_location	user_description	user_created
11015	1460932683630977025	theSun	Petaling Jaya	Breaking news updates from the website of Mala...	2009-07-29 01:58:44
11016	1460809350792896516	Sujan	NaN	•Napricorn• •It's jus...	2013-01-28 11:48:20
11017	1460581918295031809	theSun	Petaling Jaya	Breaking news updates from the website of Mala...	2009-07-29 01:58:44
11018	1460344277951782922	Aleksejs Jackovs	Ireland	#Scientificresearcher in #InformationTechnolog...	2012-08-13 13:47:03
11019	1460208097956225028	theSun	Petaling Jaya	Breaking news updates from the website of Mala...	2009-07-29 01:58:44

In [8]: `dataset.describe()`

Out[8]:

	id	user_followers	user_friends	user_favourites	retweets	favorites
count	1.102000e+04	1.102000e+04	11020.000000	1.102000e+04	11020.000000	11020.000000
mean	1.372977e+18	3.518827e+04	1170.470236	1.533289e+04	1.360436	7.510617
std	3.311622e+16	2.942509e+05	2798.429673	4.944401e+04	11.364732	51.192608
min	1.337728e+18	0.000000e+00	0.000000	0.000000e+00	0.000000	0.000000
25%	1.347623e+18	1.110000e+02	164.000000	4.360000e+02	0.000000	0.000000
50%	1.360274e+18	4.790000e+02	460.000000	2.323500e+03	0.000000	1.000000
75%	1.395234e+18	2.171250e+03	1221.000000	1.114075e+04	1.000000	4.000000
max	1.463244e+18	1.371493e+07	103226.000000	1.213794e+06	678.000000	2315.000000

In [9]: `print(dataset)`

```

      id          user_name      user_location \
0    1340539111971516416      Rachel Roh  La Crescenta-Montrose, CA
1    1338158543359250433      Albert Fong    San Francisco, CA
2    1337858199140118533      eliLTEU👉      Your Bed
3    1337855739918835717      Charles Adler  Vancouver, BC - Canada
4    1337854064604966912  Citizen News Channel      NaN
...
11015 1460932683630977025      theSun      Petaling Jaya
11016 1460809350792896516      Sujan      NaN
11017 1460581918295031809      theSun      Petaling Jaya
11018 1460344277951782922      Aleksejs Jackovs      Ireland
11019 1460208097956225028      theSun      Petaling Jaya

      user_description      user_created
\
0    Aggregator of Asian American news; scanning di... 2009-04-08 17:52:46
1    Marketing dude, tech geek, heavy metal & '80s ... 2009-09-21 15:27:30
2    heil, hydra 🙌🏼👉 2020-06-25 23:30:28
3    Hosting "CharlesAdlerTonight" Global News Radi... 2008-09-10 11:28:53
4    Citizen News Channel bringing you an alternati... 2020-04-23 17:58:42
...
11015 Breaking news updates from the website of Mala... 2009-07-29 01:58:44
11016 _____ •Napricorn• •It's jus... 2013-01-28 11:48:20
11017 Breaking news updates from the website of Mala... 2009-07-29 01:58:44
11018 #Scientificresearcher in #InformationTechnolog... 2012-08-13 13:47:03
11019 Breaking news updates from the website of Mala... 2009-07-29 01:58:44

      user_followers  user_friends  user_favourites  user_verified \
0                405          1692             3247          False
1                834           666              178          False
2                 10           88               155          False
3             49165         3933             21853           True
4                 152          580             1473          False
...
11015          151937           465              593          False
11016           4259           622             156941          False
11017          151937           465              593          False
11018                 1           38               99          False
11019          151937           465              593          False

      date      text
\
0    2020-12-20 06:06:44  Same folks said daikon paste could treat a cyt...
1    2020-12-13 16:27:13  While the world has been on the wrong side of ...
2    2020-12-12 20:33:45  #coronavirus #SputnikV #AstraZeneca #PfizerBio...
3    2020-12-12 20:23:59  Facts are immutable, Senator, even when you're...
4    2020-12-12 20:17:19  Explain to me again why we need a vaccine @Bor...
...
11015 2021-11-17 11:27:49  The number of Covid-19 cases today, 17 Novembe...
11016 2021-11-17 03:17:44      First dose ✅\n#PfizerBioNTech
11017 2021-11-16 12:14:00  The number of Covid-19 cases today, 16 Novembe...
11018 2021-11-15 20:29:42  #PfizerBioNTech has developed new pill #Paxlov...
11019 2021-11-15 11:28:34  The number of Covid-19 cases today, 15 Novembe...

```

	hashtags	source
\		
0	['PfizerBioNTech']	Twitter for Android
1	NaN	Twitter Web App
2	['coronavirus', 'SputnikV', 'AstraZeneca', 'Pf...]	Twitter for Android
3	NaN	Twitter Web App
4	['whereareallthesickpeople', 'PfizerBioNTech']	Twitter for iPhone
...
11015	NaN	Twitter Web App
11016	['PfizerBioNTech']	Twitter for Android
11017	NaN	Twitter Web App
11018	['PfizerBioNTech', 'Paxlovid', 'COVID19']	Twitter Web App
11019	NaN	Twitter Web App

	retweets	favorites	is_retweet
0	0	0	False
1	1	1	False
2	0	0	False
3	446	2129	False
4	0	0	False
...
11015	2	6	False
11016	0	4	False
11017	3	5	False
11018	1	1	False
11019	2	6	False

[11020 rows x 16 columns]



In [10]: dataset.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11020 entries, 0 to 11019
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                     11020 non-null  int64
1   user_name              11020 non-null  object
2   user_location          8750 non-null   object
3   user_description       10341 non-null  object
4   user_created           11020 non-null  object
5   user_followers         11020 non-null  int64
6   user_friends           11020 non-null  int64
7   user_favourites        11020 non-null  int64
8   user_verified          11020 non-null  bool
9   date                   11020 non-null  object
10  text                   11020 non-null  object
11  hashtags               8438 non-null   object
12  source                 11019 non-null  object
13  retweets               11020 non-null  int64
14  favorites               11020 non-null  int64
15  is_retweet             11020 non-null  bool
dtypes: bool(2), int64(6), object(8)
memory usage: 1.2+ MB
```

In [19]: dataset.isnull().sum()

```
Out[19]: id                     0
user_name                     0
user_location                2270
user_description              679
user_created                  0
user_followers                0
user_friends                  0
user_favourites               0
user_verified                 0
date                          0
text                          0
hashtags                     2582
source                        1
retweets                      0
favorites                     0
is_retweet                    0
dtype: int64
```

In [20]: dataset.columns

```
Out[20]: Index(['id', 'user_name', 'user_location', 'user_description', 'user_created',
               'user_followers', 'user_friends', 'user_favourites', 'user_verified',
               'date', 'text', 'hashtags', 'source', 'retweets', 'favorites',
               'is_retweet'],
              dtype='object')
```

In []:

In []: