

wrangle_act

December 21, 2018

```
In [50]: import pandas as pd
import requests
import tweepy
import json
import matplotlib.pyplot as plt
import glob
import re
```

1 Gathering

```
In [51]: WeRateDogs_archive = pd.read_csv('twitter-archive-enhanced.csv')
WeRateDogs_archive.head(50)
```

```
Out[51]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	NaN	NaN	
1	892177421306343426	NaN	NaN	
2	891815181378084864	NaN	NaN	
3	891689557279858688	NaN	NaN	
4	891327558926688256	NaN	NaN	
5	891087950875897856	NaN	NaN	
6	890971913173991426	NaN	NaN	
7	890729181411237888	NaN	NaN	
8	890609185150312448	NaN	NaN	
9	890240255349198849	NaN	NaN	
10	890006608113172480	NaN	NaN	
11	889880896479866881	NaN	NaN	
12	889665388333682689	NaN	NaN	
13	889638837579907072	NaN	NaN	
14	889531135344209921	NaN	NaN	
15	889278841981685760	NaN	NaN	
16	888917238123831296	NaN	NaN	
17	888804989199671297	NaN	NaN	
18	888554962724278272	NaN	NaN	
19	888202515573088257	NaN	NaN	
20	888078434458587136	NaN	NaN	
21	887705289381826560	NaN	NaN	
22	887517139158093824	NaN	NaN	

23	887473957103951883	NaN	NaN
24	887343217045368832	NaN	NaN
25	887101392804085760	NaN	NaN
26	886983233522544640	NaN	NaN
27	886736880519319552	NaN	NaN
28	886680336477933568	NaN	NaN
29	886366144734445568	NaN	NaN
30	886267009285017600	8.862664e+17	2.281182e+09
31	886258384151887873	NaN	NaN
32	886054160059072513	NaN	NaN
33	885984800019947520	NaN	NaN
34	885528943205470208	NaN	NaN
35	885518971528720385	NaN	NaN
36	885311592912609280	NaN	NaN
37	885167619883638784	NaN	NaN
38	884925521741709313	NaN	NaN
39	884876753390489601	NaN	NaN
40	884562892145688576	NaN	NaN
41	884441805382717440	NaN	NaN
42	884247878851493888	NaN	NaN
43	884162670584377345	NaN	NaN
44	883838122936631299	NaN	NaN
45	883482846933004288	NaN	NaN
46	883360690899218434	NaN	NaN
47	883117836046086144	NaN	NaN
48	882992080364220416	NaN	NaN
49	882762694511734784	NaN	NaN

	timestamp \
0	2017-08-01 16:23:56 +0000
1	2017-08-01 00:17:27 +0000
2	2017-07-31 00:18:03 +0000
3	2017-07-30 15:58:51 +0000
4	2017-07-29 16:00:24 +0000
5	2017-07-29 00:08:17 +0000
6	2017-07-28 16:27:12 +0000
7	2017-07-28 00:22:40 +0000
8	2017-07-27 16:25:51 +0000
9	2017-07-26 15:59:51 +0000
10	2017-07-26 00:31:25 +0000
11	2017-07-25 16:11:53 +0000
12	2017-07-25 01:55:32 +0000
13	2017-07-25 00:10:02 +0000
14	2017-07-24 17:02:04 +0000
15	2017-07-24 00:19:32 +0000
16	2017-07-23 00:22:39 +0000
17	2017-07-22 16:56:37 +0000
18	2017-07-22 00:23:06 +0000

19 2017-07-21 01:02:36 +0000
20 2017-07-20 16:49:33 +0000
21 2017-07-19 16:06:48 +0000
22 2017-07-19 03:39:09 +0000
23 2017-07-19 00:47:34 +0000
24 2017-07-18 16:08:03 +0000
25 2017-07-18 00:07:08 +0000
26 2017-07-17 16:17:36 +0000
27 2017-07-16 23:58:41 +0000
28 2017-07-16 20:14:00 +0000
29 2017-07-15 23:25:31 +0000
30 2017-07-15 16:51:35 +0000
31 2017-07-15 16:17:19 +0000
32 2017-07-15 02:45:48 +0000
33 2017-07-14 22:10:11 +0000
34 2017-07-13 15:58:47 +0000
35 2017-07-13 15:19:09 +0000
36 2017-07-13 01:35:06 +0000
37 2017-07-12 16:03:00 +0000
38 2017-07-12 00:01:00 +0000
39 2017-07-11 20:47:12 +0000
40 2017-07-11 00:00:02 +0000
41 2017-07-10 15:58:53 +0000
42 2017-07-10 03:08:17 +0000
43 2017-07-09 21:29:42 +0000
44 2017-07-09 00:00:04 +0000
45 2017-07-08 00:28:19 +0000
46 2017-07-07 16:22:55 +0000
47 2017-07-07 00:17:54 +0000
48 2017-07-06 15:58:11 +0000
49 2017-07-06 00:46:41 +0000

source
0 Twitter for iPhone
1 Twitter for iPhone
2 Twitter for iPhone
3 Twitter for iPhone
4 Twitter for iPhone
5 Twitter for iPhone
6 Twitter for iPhone
7 Twitter for iPhone
8 Twitter for iPhone
9 Twitter for iPhone
10 Twitter for iPhone
11 Twitter for iPhone
12 Twitter for iPhone
13 Twitter for iPhone
14 Twitter for iPhone

15 Twitter for iPhone
16 Twitter for iPhone
17 Twitter for iPhone
18 Twitter for iPhone
19 Twitter for iPhone
20 Twitter for iPhone
21 Twitter for iPhone
22 Twitter for iPhone
23 Twitter for iPhone
24 Twitter for iPhone
25 Twitter for iPhone
26 Twitter for iPhone
27 Twitter for iPhone
28 Twitter for iPhone
29 Twitter for iPhone
30 Twitter for iPhone
31 Twitter for iPhone
32 Twitter for iPhone
33 Twitter for iPhone
34 Twitter for iPhone
35 Twitter for iPhone
36 Twitter for iPhone
37 Twitter for iPhone
38 Twitter for iPhone
39 Twitter for iPhone
40 Twitter for iPhone
41 Twitter for iPhone
42 Twitter for iPhone
43 Twitter for iPhone
44 Twitter for iPhone
45 Twitter for iPhone
46 Twitter for iPhone
47 Twitter for iPhone
48 Twitter for iPhone
49 Twitter for iPhone

0 This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/
1 This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she's
2 This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall grass. You
3 This is Darla. She commenced a snooze mid meal. 13/10 happens to the best of us htt
4 This is Franklin. He would like you to stop calling him "cute." He is a very fierce
5 Here we have a majestic great white breaching off South Africa's coast. Absolutely
6 Meet Jax. He enjoys ice cream so much he gets nervous around it. 13/10 help Jax enj
7 When you watch your owner call another dog a good boy but then they turn back to yo
8 This is Zoey. She doesn't want to be one of the scary sharks. Just wants to be a sn
9 This is Cassie. She is a college pup. Studying international doggo communication an
10 This is Koda. He is a South Australian deckshark. Deceptively deadly. Frighteningly

11 This is Bruno. He is a service shark. Only gets out of the water to assist you. 13/
 12 Here's a puppo that seems to be on the fence about something haha no but seriously
 13 This is Ted. He does his best. Sometimes that's not enough. But it's ok. 12/10 would
 14 This is Stuart. He's sporting his favorite fanny pack. Secretly filled with bones o
 15 This is Oliver. You're witnessing one of his many brutal attacks. Seems to be playi
 16 This is Jim. He found a fren. Taught him how to sit like the good boys. 12/10 for b
 17 This is Zeke. He has a new stick. Very proud of it. Would like you to throw it for
 18 This is Ralphus. He's powering up. Attempting maximum borkdrive. 13/10 inspirationa
 19 RT @dog_rates: This is Canela. She attempted some fancy porch pics. They were unsuo
 20 This is Gerald. He was just told he didn't get the job he interviewed for. A h*ckin
 21 This is Jeffrey. He has a monopoly on the pool noodles. Currently running a 'boop f
 22 I've yet to rate a Venezuelan Hover Wiener. This is such an honor. 14/10 paw-inspir
 23 This is Canela. She attempted some fancy porch pics. They were unsuccessful. 13/10
 24 You may not have known you needed to see this today. 13/10 please enjoy (IG: emmylo
 25 This... is a Jubilant Antarctic House Bear. We only rate dogs. Please only send dog
 26 This is Maya. She's very shy. Rarely leaves her cup. 13/10 would find her an enviro
 27 This is Mingus. He's a wonderful father to his smol pup. Confirmed 13/10, but he ne
 28 This is Derek. He's late for a dog meeting. 13/10 pet...al to the metal <https://t.co/1r4>
 29 This is Roscoe. Another pupper fallen victim to spontaneous tongue ejections. Get t
 30 @NonWhiteHat @MayhewMayhem omg hello tanner you are a scary good boy 12/10 would pe
 31 This is Waffles. His doggles are pupside down. Unsure how to fix. 13/10 someone ass
 32 RT @Athletics: 12/10 #BATP <https://t.co/WxwJmvjfxo>
 33 Viewer discretion advised. This is Jimbo. He will rip ur finger right h*ckin off. C
 34 This is Maisey. She fell asleep mid-excavation. Happens to the best of us. 13/10 wo
 35 I have a new hero and his name is Howard. 14/10 <https://t.co/gzLHboL7Sk>
 36 RT @dog_rates: This is Lilly. She just parallel barked. Kindly requests a reward no
 37 Here we have a corgi undercover as a malamute. Pawbably doing important investigati
 38 This is Earl. He found a hat. Nervous about what you think of it. 12/10 it's deligh
 39 This is Lola. It's her first time outside. Must test the earth and taste the atmo
 40 This is Kevin. He's just so happy. 13/10 what is your secret Kevin <https://t.co/1r4>
 41 I present to you, Pup in Hat. Pup in Hat is great for all occasions. Extremely vers
 42 OMG HE DIDN'T MEAN TO HE WAS JUST TRYING A LITTLE BARKOUR HE'S SUPER SORRY 13/10 WO
 43 Meet Yogi. He doesn't have any important dog meetings today he just enjoys looking
 44 This is Noah. He can't believe someone made this mess. Got the vacuum out for you t
 45 This is Bella. She hopes her smile made you smile. If not, she is also offering you
 46 Meet Grizzwald. He may be the floofiest floofer I ever did see. Lost eyes saving a
 47 Please only send dogs. We don't rate mechanics, no matter how h*ckin good. Thank yo
 48 This is Rusty. He wasn't ready for the first pic. Clearly puppared for the second.
 49 This is Gus. He's quite the cheeky pupper. Already perfected the disinterested wink

	retweeted_status_id	retweeted_status_user_id	retweeted_status_timestamp	\
0	NaN	NaN	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	
5	NaN	NaN	NaN	
6	NaN	NaN	NaN	

7	NaN	NaN	NaN
8	NaN	NaN	NaN
9	NaN	NaN	NaN
10	NaN	NaN	NaN
11	NaN	NaN	NaN
12	NaN	NaN	NaN
13	NaN	NaN	NaN
14	NaN	NaN	NaN
15	NaN	NaN	NaN
16	NaN	NaN	NaN
17	NaN	NaN	NaN
18	NaN	NaN	NaN
19	8.874740e+17	4.196984e+09	2017-07-19 00:47:34 +0000
20	NaN	NaN	NaN
21	NaN	NaN	NaN
22	NaN	NaN	NaN
23	NaN	NaN	NaN
24	NaN	NaN	NaN
25	NaN	NaN	NaN
26	NaN	NaN	NaN
27	NaN	NaN	NaN
28	NaN	NaN	NaN
29	NaN	NaN	NaN
30	NaN	NaN	NaN
31	NaN	NaN	NaN
32	8.860537e+17	1.960740e+07	2017-07-15 02:44:07 +0000
33	NaN	NaN	NaN
34	NaN	NaN	NaN
35	NaN	NaN	NaN
36	8.305833e+17	4.196984e+09	2017-02-12 01:04:29 +0000
37	NaN	NaN	NaN
38	NaN	NaN	NaN
39	NaN	NaN	NaN
40	NaN	NaN	NaN
41	NaN	NaN	NaN
42	NaN	NaN	NaN
43	NaN	NaN	NaN
44	NaN	NaN	NaN
45	NaN	NaN	NaN
46	NaN	NaN	NaN
47	NaN	NaN	NaN
48	NaN	NaN	NaN
49	NaN	NaN	NaN

0 https://twitter.com/dog_rates/status/892420643555336193/photo/1
1 https://twitter.com/dog_rates/status/892177421306343426/photo/1
2 https://twitter.com/dog_rates/status/891815181378084864/photo/1

3 https://twitter.com/dog_rates/status/891689557279858688/photo/1
4 https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://twitter.com/dog_rates/status/891327558926688256/photo/1
5 https://twitter.com/dog_rates/status/891087950875897856/photo/1
6 <https://gofundme.com/ydvmve-surgery-for-jax>,https://twitter.com/dog_rates/status/891087950875897856/photo/1
7 https://twitter.com/dog_rates/status/890729181411237888/photo/1,https://twitter.com/dog_rates/status/890729181411237888/photo/1
8 https://twitter.com/dog_rates/status/890609185150312448/photo/1
9 https://twitter.com/dog_rates/status/890240255349198849/photo/1
10 https://twitter.com/dog_rates/status/890006608113172480/photo/1,https://twitter.com/dog_rates/status/890006608113172480/photo/1
11 https://twitter.com/dog_rates/status/889880896479866881/photo/1
12 https://twitter.com/dog_rates/status/889665388333682689/photo/1
13 https://twitter.com/dog_rates/status/889638837579907072/photo/1,https://twitter.com/dog_rates/status/889638837579907072/photo/1
14 https://twitter.com/dog_rates/status/889531135344209921/photo/1
15 https://twitter.com/dog_rates/status/889278841981685760/video/1
16 https://twitter.com/dog_rates/status/888917238123831296/photo/1
17 https://twitter.com/dog_rates/status/888804989199671297/photo/1,https://twitter.com/dog_rates/status/888804989199671297/photo/1
18 https://twitter.com/dog_rates/status/888554962724278272/photo/1,https://twitter.com/dog_rates/status/888554962724278272/photo/1
19 https://twitter.com/dog_rates/status/887473957103951883/photo/1,https://twitter.com/dog_rates/status/887473957103951883/photo/1
20 https://twitter.com/dog_rates/status/888078434458587136/photo/1,https://twitter.com/dog_rates/status/888078434458587136/photo/1
21 https://twitter.com/dog_rates/status/887705289381826560/photo/1
22 https://twitter.com/dog_rates/status/887517139158093824/video/1
23 https://twitter.com/dog_rates/status/887473957103951883/photo/1,https://twitter.com/dog_rates/status/887473957103951883/photo/1
24 https://twitter.com/dog_rates/status/887343217045368832/video/1
25 https://twitter.com/dog_rates/status/887101392804085760/photo/1
26 https://twitter.com/dog_rates/status/886983233522544640/photo/1,https://twitter.com/dog_rates/status/886983233522544640/photo/1
27 <https://www.gofundme.com/mingusneedsus>,https://twitter.com/dog_rates/status/88673680336477933568/photo/1
28 https://twitter.com/dog_rates/status/886680336477933568/photo/1
29 https://twitter.com/dog_rates/status/886366144734445568/photo/1,https://twitter.com/dog_rates/status/886366144734445568/photo/1
30 NaN
31 https://twitter.com/dog_rates/status/886258384151887873/photo/1
32 https://twitter.com/dog_rates/status/886053434075471873,https://twitter.com/dog_rates/status/886053434075471873
33 https://twitter.com/dog_rates/status/885984800019947520/photo/1
34 https://twitter.com/dog_rates/status/885528943205470208/photo/1
35 <https://twitter.com/4bonds2carbon/status/885517367337512960>
36 https://twitter.com/dog_rates/status/830583320585068544/photo/1,https://twitter.com/dog_rates/status/830583320585068544/photo/1
37 https://twitter.com/dog_rates/status/885167619883638784/photo/1,https://twitter.com/dog_rates/status/885167619883638784/photo/1
38 https://twitter.com/dog_rates/status/884925521741709313/photo/1
39 https://twitter.com/dog_rates/status/884876753390489601/photo/1,https://twitter.com/dog_rates/status/884876753390489601/photo/1
40 https://twitter.com/dog_rates/status/884562892145688576/photo/1
41 https://twitter.com/dog_rates/status/884441805382717440/photo/1
42 https://twitter.com/kaijohnson_19/status/883965650754039809
43 https://twitter.com/dog_rates/status/884162670584377345/photo/1
44 https://twitter.com/dog_rates/status/883838122936631299/photo/1
45 https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.com/dog_rates/status/883482846933004288/photo/1
46 https://twitter.com/dog_rates/status/883360690899218434/photo/1
47 https://twitter.com/dog_rates/status/883117836046086144/photo/1,https://twitter.com/dog_rates/status/883117836046086144/photo/1
48 https://twitter.com/dog_rates/status/882992080364220416/photo/1,https://twitter.com/dog_rates/status/882992080364220416/photo/1
49 https://twitter.com/dog_rates/status/882762694511734784/photo/1

	rating_numerator	rating_denominator	name	doggo	floofer	pupper \
0	13	10	Phineas	None	None	None
1	13	10	Tilly	None	None	None
2	12	10	Archie	None	None	None
3	13	10	Darla	None	None	None
4	12	10	Franklin	None	None	None
5	13	10	None	None	None	None
6	13	10	Jax	None	None	None
7	13	10	None	None	None	None
8	13	10	Zoey	None	None	None
9	14	10	Cassie	doggo	None	None
10	13	10	Koda	None	None	None
11	13	10	Bruno	None	None	None
12	13	10	None	None	None	None
13	12	10	Ted	None	None	None
14	13	10	Stuart	None	None	None
15	13	10	Oliver	None	None	None
16	12	10	Jim	None	None	None
17	13	10	Zeke	None	None	None
18	13	10	Ralphus	None	None	None
19	13	10	Canela	None	None	None
20	12	10	Gerald	None	None	None
21	13	10	Jeffrey	None	None	None
22	14	10	such	None	None	None
23	13	10	Canela	None	None	None
24	13	10	None	None	None	None
25	12	10	None	None	None	None
26	13	10	Maya	None	None	None
27	13	10	Mingus	None	None	None
28	13	10	Derek	None	None	None
29	12	10	Roscoe	None	None	pupper
30	12	10	None	None	None	None
31	13	10	Waffles	None	None	None
32	12	10	None	None	None	None
33	12	10	Jimbo	None	None	None
34	13	10	Maisey	None	None	None
35	14	10	None	None	None	None
36	13	10	Lilly	None	None	None
37	13	10	None	None	None	None
38	12	10	Earl	None	None	None
39	13	10	Lola	None	None	None
40	13	10	Kevin	None	None	None
41	14	10	None	None	None	None
42	13	10	None	None	None	None
43	12	10	Yogi	doggo	None	None
44	12	10	Noah	None	None	None
45	5	10	Bella	None	None	None
46	13	10	Grizzwald	None	floofer	None

47	13	10	None	None	None	None
48	13	10	Rusty	None	None	None
49	12	10	Gus	None	None	pupper

	puppo
0	None
1	None
2	None
3	None
4	None
5	None
6	None
7	None
8	None
9	None
10	None
11	None
12	puppo
13	None
14	puppo
15	None
16	None
17	None
18	None
19	None
20	None
21	None
22	None
23	None
24	None
25	None
26	None
27	None
28	None
29	None
30	None
31	None
32	None
33	None
34	None
35	None
36	None
37	None
38	None
39	None
40	None
41	None
42	None

```

43 None
44 None
45 None
46 None
47 None
48 None
49 None

```

We have 2356 entries in WeRateDogs Archive. 17 columns are provided as part of the dataset.

```
In [52]: WeRateDogs_archive.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                2356 non-null int64
in_reply_to_status_id    78 non-null float64
in_reply_to_user_id      78 non-null float64
timestamp               2356 non-null object
source                  2356 non-null object
text                    2356 non-null object
retweeted_status_id       181 non-null float64
retweeted_status_user_id  181 non-null float64
retweeted_status_timestamp 181 non-null object
expanded_urls            2297 non-null object
rating_numerator         2356 non-null int64
rating_denominator       2356 non-null int64
name                    2356 non-null object
doggo                    2356 non-null object
floofer                  2356 non-null object
pupper                  2356 non-null object
puppo                    2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB

```

```
In [53]: url=" https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions-2017-08-17/image_predictions.tsv"
response = requests.get(url)
```

```

with open('image_predictions.tsv', 'wb') as file:
    file.write(response.content)

```

```
image_predictions = pd.read_csv('image_predictions.tsv', sep='\t')
```

```
image_predictions.head()
```

```

Out[53]:
   tweet_id                jpg_url \
0  666020888022790149  https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg
1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg

```

```

2 666033412701032449 https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3 666044226329800704 https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4 666049248165822465 https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

```

	img_num	p1	p1_conf	p1_dog	p2 \
0	1	Welsh_springer_spaniel	0.465074	True	collie
1	1	redbone	0.506826	True	miniature_pinscher
2	1	German_shepherd	0.596461	True	malinois
3	1	Rhodesian_ridgeback	0.408143	True	redbone
4	1	miniature_pinscher	0.560311	True	Rottweiler

	p2_conf	p2_dog	p3	p3_conf	p3_dog
0	0.156665	True	Shetland_sheepdog	0.061428	True
1	0.074192	True	Rhodesian_ridgeback	0.072010	True
2	0.138584	True	bloodhound	0.116197	True
3	0.360687	True	miniature_pinscher	0.222752	True
4	0.243682	True	Doberman	0.154629	True

Gather data from twitter API using Python's Tweepy library and store data

```

In [54]: consumer_key = ''
        consumer_secret = ''
        access_token = ''
        access_secret = ''

        auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
        auth.set_access_token(access_token, access_secret)
        api = tweepy.API(auth, wait_on_rate_limit=True , wait_on_rate_limit_notify=True)

        '''
        txt_file=open('tweet-json.txt','w')
        #txt_file.write('tweet_id retweet_count favorite_count'\n')
        for tweet_id in df1.tweet_id:
            try:
                tweet_info=api.get_status(str(tweet_id),tweet_mode='extended')
                retweet_count=tweet_info.retweet_count
                print(tweet_id)
                favorite_count=tweet_info.favorite_count
                txt_file.write(tweet_id+' '+str(retweet_count)+' '+str(favorite_count)+'\n')
            except Exception as e:
                print('error')
        txt_file.close()
        '''

Out[54]: "\ntxt_file=open('tweet-json.txt','w')\n#txt_file.write('tweet_id retweet_count favorit

In [55]: # Reading tweet-json.txt file, line by line and append it to "status" list
        with open('tweet-json.txt') as file:
            status = []

```

```

    for line in file:
        status.append(json.loads(line))

#using json_normalize to convert json keys to columns
from pandas.io.json import json_normalize
df_json = pd.DataFrame.from_dict(json_normalize(status), orient='columns')

# Refining the the json dataframe to the columns we need for analysis
columns_needed = ['id' , 'retweet_count' , 'favorite_count']
df_json_cleaned = df_json[columns_needed]

df_json_cleaned.rename(columns = {'id':'tweet_id'} , inplace=True)

df_json_cleaned.head()

```

/opt/conda/lib/python3.6/site-packages/pandas/core/frame.py:2746: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#**kwargs)

```

Out[55]:
      tweet_id  retweet_count  favorite_count
0  892420643555336193      8853           39467
1  892177421306343426      6514           33819
2  891815181378084864      4328           25461
3  891689557279858688      8964           42908
4  891327558926688256      9774           41048

```

2 Assessing

```
In [56]: WeRateDogs_archive.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                2356 non-null int64
in_reply_to_status_id    78 non-null float64
in_reply_to_user_id      78 non-null float64
timestamp                2356 non-null object
source                  2356 non-null object
text                    2356 non-null object
retweeted_status_id      181 non-null float64
retweeted_status_user_id 181 non-null float64
retweeted_status_timestamp 181 non-null object
expanded_urls            2297 non-null object

```

```

rating_numerator      2356 non-null int64
rating_denominator    2356 non-null int64
name                   2356 non-null object
doggo                  2356 non-null object
floofer                2356 non-null object
pupper                2356 non-null object
puppo                  2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB

```

```

In [57]: WeRateDogs_archive.tweet_id.nunique()
         print("All IDs are unique")

```

All IDs are unique

```

In [58]: print (WeRateDogs_archive.source.value_counts())
         print("\n No twitter for Android ?!! Weird Hah :) ")

```

```

<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>      2221
<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>                        91
<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>                     33
<a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>    11
Name: source, dtype: int64

```

No twitter for Android ?!! Weird Hah :)

```

In [59]: #This is to see full text and assessing the data deeply ;)
         pd.set_option('display.max_colwidth', -1)

```

```

#Here we want to see if there is another way of calling a dog name other than "This is"
WeRateDogs_archive[~WeRateDogs_archive['text'].str.contains("This is" , case=False)]

```

```

Out[59]:
   tweet_id  in_reply_to_status_id  in_reply_to_user_id \
5      891087950875897856      NaN                      NaN
6      890971913173991426      NaN                      NaN
7      890729181411237888      NaN                      NaN
12     889665388333682689      NaN                      NaN
24     887343217045368832      NaN                      NaN
25     887101392804085760      NaN                      NaN
30     886267009285017600  8.862664e+17      2.281182e+09
32     886054160059072513      NaN                      NaN
35     885518971528720385      NaN                      NaN
37     885167619883638784      NaN                      NaN
41     884441805382717440      NaN                      NaN
42     884247878851493888      NaN                      NaN
43     884162670584377345      NaN                      NaN

```

46	883360690899218434	NaN	NaN
47	883117836046086144	NaN	NaN
55	881633300179243008	8.816070e+17	4.738443e+07
56	881536004380872706	NaN	NaN
57	881268444196462592	NaN	NaN
59	880872448815771648	NaN	NaN
61	880221127280381952	NaN	NaN
62	880095782870896641	NaN	NaN
64	879674319642796034	8.795538e+17	3.105441e+09
72	878604707211726852	NaN	NaN
73	878404777348136964	NaN	NaN
74	878316110768087041	NaN	NaN
75	878281511006478336	NaN	NaN
78	877611172832227328	NaN	NaN
80	877316821321428993	NaN	NaN
83	876537666061221889	NaN	NaN
85	876120275196170240	NaN	NaN
...
2301	667044094246576128	NaN	NaN
2305	666837028449972224	NaN	NaN
2306	666835007768551424	NaN	NaN
2307	666826780179869698	NaN	NaN
2310	666786068205871104	NaN	NaN
2316	666649482315059201	NaN	NaN
2320	666437273139982337	NaN	NaN
2321	666435652385423360	NaN	NaN
2322	666430724426358785	NaN	NaN
2323	666428276349472768	NaN	NaN
2324	666421158376562688	NaN	NaN
2328	666396247373291520	NaN	NaN
2329	666373753744588802	NaN	NaN
2330	666362758909284353	NaN	NaN
2331	666353288456101888	NaN	NaN
2332	666345417576210432	NaN	NaN
2336	666273097616637952	NaN	NaN
2337	666268910803644416	NaN	NaN
2338	666104133288665088	NaN	NaN
2339	666102155909144576	NaN	NaN
2340	666099513787052032	NaN	NaN
2341	666094000022159362	NaN	NaN
2342	666082916733198337	NaN	NaN
2343	666073100786774016	NaN	NaN
2344	666071193221509120	NaN	NaN
2346	666058600524156928	NaN	NaN
2348	666055525042405380	NaN	NaN
2351	666049248165822465	NaN	NaN
2353	666033412701032449	NaN	NaN
2355	666020888022790149	NaN	NaN

	timestamp \
5	2017-07-29 00:08:17 +0000
6	2017-07-28 16:27:12 +0000
7	2017-07-28 00:22:40 +0000
12	2017-07-25 01:55:32 +0000
24	2017-07-18 16:08:03 +0000
25	2017-07-18 00:07:08 +0000
30	2017-07-15 16:51:35 +0000
32	2017-07-15 02:45:48 +0000
35	2017-07-13 15:19:09 +0000
37	2017-07-12 16:03:00 +0000
41	2017-07-10 15:58:53 +0000
42	2017-07-10 03:08:17 +0000
43	2017-07-09 21:29:42 +0000
46	2017-07-07 16:22:55 +0000
47	2017-07-07 00:17:54 +0000
55	2017-07-02 21:58:53 +0000
56	2017-07-02 15:32:16 +0000
57	2017-07-01 21:49:04 +0000
59	2017-06-30 19:35:32 +0000
61	2017-06-29 00:27:25 +0000
62	2017-06-28 16:09:20 +0000
64	2017-06-27 12:14:36 +0000
72	2017-06-24 13:24:20 +0000
73	2017-06-24 00:09:53 +0000
74	2017-06-23 18:17:33 +0000
75	2017-06-23 16:00:04 +0000
78	2017-06-21 19:36:23 +0000
80	2017-06-21 00:06:44 +0000
83	2017-06-18 20:30:39 +0000
85	2017-06-17 16:52:05 +0000
...	...
2301	2015-11-18 18:17:59 +0000
2305	2015-11-18 04:35:11 +0000
2306	2015-11-18 04:27:09 +0000
2307	2015-11-18 03:54:28 +0000
2310	2015-11-18 01:12:41 +0000
2316	2015-11-17 16:09:56 +0000
2320	2015-11-17 02:06:42 +0000
2321	2015-11-17 02:00:15 +0000
2322	2015-11-17 01:40:41 +0000
2323	2015-11-17 01:30:57 +0000
2324	2015-11-17 01:02:40 +0000
2328	2015-11-16 23:23:41 +0000
2329	2015-11-16 21:54:18 +0000
2330	2015-11-16 21:10:36 +0000
2331	2015-11-16 20:32:58 +0000

2332 2015-11-16 20:01:42 +0000
 2336 2015-11-16 15:14:19 +0000
 2337 2015-11-16 14:57:41 +0000
 2338 2015-11-16 04:02:55 +0000
 2339 2015-11-16 03:55:04 +0000
 2340 2015-11-16 03:44:34 +0000
 2341 2015-11-16 03:22:39 +0000
 2342 2015-11-16 02:38:37 +0000
 2343 2015-11-16 01:59:36 +0000
 2344 2015-11-16 01:52:02 +0000
 2346 2015-11-16 01:01:59 +0000
 2348 2015-11-16 00:49:46 +0000
 2351 2015-11-16 00:24:50 +0000
 2353 2015-11-15 23:21:54 +0000
 2355 2015-11-15 22:32:08 +0000

			source
5	Twitter for iPhone</a	
6	Twitter for iPhone</a	
7	Twitter for iPhone</a	
12	Twitter for iPhone</a	
24	Twitter for iPhone</a	
25	Twitter for iPhone</a	
30	Twitter for iPhone</a	
32	Twitter for iPhone</a	
35	Twitter for iPhone</a	
37	Twitter for iPhone</a	
41	Twitter for iPhone</a	
42	Twitter for iPhone</a	
43	Twitter for iPhone</a	
46	Twitter for iPhone</a	
47	Twitter for iPhone</a	
55	Twitter for iPhone</a	
56	Twitter for iPhone</a	
57	Twitter for iPhone</a	
59	Twitter for iPhone</a	
61	Twitter for iPhone</a	
62	Twitter for iPhone</a	
64	Twitter for iPhone</a	
72	Twitter for iPhone</a	
73	Twitter for iPhone</a	
74	Twitter for iPhone</a	
75	Twitter for iPhone</a	
78	Twitter for iPhone</a	
80	Twitter for iPhone</a	
83	Twitter for iPhone</a	
85	Twitter for iPhone</a	
...			..

2301 Twitter for iPhone
 2305 Twitter for iPhone
 2306 Twitter for iPhone
 2307 Twitter for iPhone
 2310 Twitter for iPhone
 2316 Twitter for iPhone
 2320 Twitter for iPhone
 2321 Twitter for iPhone
 2322 Twitter for iPhone
 2323 Twitter for iPhone
 2324 Twitter for iPhone
 2328 Twitter for iPhone
 2329 Twitter for iPhone
 2330 Twitter for iPhone
 2331 Twitter for iPhone
 2332 Twitter for iPhone
 2336 Twitter for iPhone
 2337 Twitter for iPhone
 2338 Twitter for iPhone
 2339 Twitter for iPhone
 2340 Twitter for iPhone
 2341 Twitter for iPhone
 2342 Twitter for iPhone
 2343 Twitter for iPhone
 2344 Twitter for iPhone
 2346 Twitter for iPhone
 2348 Twitter for iPhone
 2351 Twitter for iPhone
 2353 Twitter for iPhone
 2355 Twitter for iPhone

5 Here we have a majestic great white breaching off South Africa's coast. Absolutel
 6 Meet Jax. He enjoys ice cream so much he gets nervous around it. 13/10 help Jax e
 7 When you watch your owner call another dog a good boy but then they turn back to
 12 Here's a puppo that seems to be on the fence about something haha no but seriousl
 24 You may not have known you needed to see this today. 13/10 please enjoy (IG: emmy
 25 This... is a Jubilant Antarctic House Bear. We only rate dogs. Please only send d
 30 @NonWhiteHat @MayhewMayhem omg hello tanner you are a scary good boy 12/10 would
 32 RT @Athletics: 12/10 #BATP https://t.co/WxwJmvjfxo
 35 I have a new hero and his name is Howard. 14/10 https://t.co/gzLHboL7Sk
 37 Here we have a corgi undercover as a malamute. Pawbably doing important investiga
 41 I present to you, Pup in Hat. Pup in Hat is great for all occasions. Extremely ve
 42 OMG HE DIDN'T MEAN TO HE WAS JUST TRYING A LITTLE BARKOUR HE'S SUPER SORRY 13/10
 43 Meet Yogi. He doesn't have any important dog meetings today he just enjoys lookin
 46 Meet Grizzwald. He may be the floofiest floofer I ever did see. Lost eyes saving
 47 Please only send dogs. We don't rate mechanics, no matter how h*ckin good. Thank
 55 @rroushfenway These are good dogs but 17/10 is an emotional impulse rating. More 1

56 Here is a pupper approaching maximum borkdrive. Zooming at never before seen speed.
 57 Meet Elliot. He's a Canadian Forrest Pup. Unusual number of antlers for a dog. Sn
 59 Ugh not again. We only rate dogs. Please don't send in well-dressed floppy-tongue
 61 Meet Jesse. He's a Fetty Woof. His tongue ejects without warning. A true bleptoma
 62 Please don't send in photos without dogs in them. We're not @porch_rates. Insubor
 64 @RealKentMurphy 14/10 confirmed
 72 Martha is stunning how h*ckin dare you. 13/10 <https://t.co/9uABQXgjwa>
 73 RT @dog_rates: Meet Shadow. In an attempt to reach maximum zooming borkdrive, he
 74 RT @dog_rates: Meet Terrance. He's being yelled at because he stapled the wrong s
 75 Meet Shadow. In an attempt to reach maximum zooming borkdrive, he tore his ACL. S
 78 RT @rachel2195: @dog_rates the boyfriend and his soaking wet pupper h*cking love
 80 Meet Dante. At first he wasn't a fan of his new raincoat, then he saw his reflect
 83 I can say with the pupmost confidence that the doggos who assisted with this sear
 85 Meet Venti, a seemingly caffeinated puppoccino. She was just informed the weekend
 ...
 2301 12/10 gimme now <https://t.co/QZAnwgnOMB>
 2305 My goodness. Very rare dog here. Large. Tail dangerous. Kinda fat. Only eats leav
 2306 These are Peruvian Feldspars. Their names are Cupit and Prencer. Both resemble Ra
 2307 12/10 simply brilliant pup <https://t.co/V6ZzG45zzG>
 2310 Unfamiliar with this breed. Ears pointy af. Won't let go of seashell. Won't eat k
 2316 Cool dog. Enjoys couch. Low monotone bark. Very nice kicks. Pisses milk (must be
 2320 Here we see a lone northeastern Cumberbatch. Half ladybug. Only builds with brick
 2321 "Can you behave? You're ruining my wedding day"\nDOG: idgaf this flashlight taste
 2322 Oh boy what a pup! Sunglasses take this one to the next level. Weirdly folds from
 2323 Here we have an Austrian Pulitzer. Collectors edition. Levitates (?). 7/10 would
 2324 *internally screaming* 12/10 <https://t.co/YMcrcXC2Y6R>
 2328 Oh goodness. A super rare northeast Qdoba kangaroo mix. Massive feet. No pouch (d
 2329 Those are sunglasses and a jean jacket. 11/10 dog cool af <https://t.co/uHXrPkUEyl>
 2330 Unique dog here. Very small. Lives in container of Frosted Flakes (?). Short legs
 2331 Here we have a mixed Asiago from the Galápagos Islands. Only one ear working. Big
 2332 Look at this jokester thinking seat belt laws don't apply to him. Great tongue th
 2336 Can take selfies 11/10 <https://t.co/ws2AMaWpW>
 2337 Very concerned about fellow dog trapped in computer. 10/10 <https://t.co/OyxApIkp>
 2338 Not familiar with this breed. No tail (weird). Only 2 legs. Doesn't bark. Surpris
 2339 Oh my. Here you are seeing an Adobe Setter giving birth to twins!!! The world is
 2340 Can stand on stump for what seems like a while. Built that birdhouse? Impressive.
 2341 This appears to be a Mongolian Presbyterian mix. Very tired. Tongue slip confirme
 2342 Here we have a well-established sunblockerspaniel. Lost his other flip-flop. 6/10
 2343 Let's hope this flight isn't Malaysian (lol). What a dog! Almost completely camou
 2344 Here we have a northern speckled Rhododendron. Much sass. Gives 0 fucks. Good ton
 2346 Here is the Rand Paul of retrievers folks! He's probably good at poker. Can drink
 2348 Here is a Siberian heavily armored polar bear mix. Strong owner. 10/10 I would do
 2351 Here we have a 1949 1st generation vulpix. Enjoys sweat tea and Fox News. Cannot
 2353 Here is a very happy pup. Big fan of well-maintained decks. Just look at that ton
 2355 Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big fan of relaxin

	retweeted_status_id	retweeted_status_user_id	\
5	NaN	NaN	

6	NaN	NaN
7	NaN	NaN
12	NaN	NaN
24	NaN	NaN
25	NaN	NaN
30	NaN	NaN
32	8.860537e+17	1.960740e+07
35	NaN	NaN
37	NaN	NaN
41	NaN	NaN
42	NaN	NaN
43	NaN	NaN
46	NaN	NaN
47	NaN	NaN
55	NaN	NaN
56	NaN	NaN
57	NaN	NaN
59	NaN	NaN
61	NaN	NaN
62	NaN	NaN
64	NaN	NaN
72	NaN	NaN
73	8.782815e+17	4.196984e+09
74	6.690004e+17	4.196984e+09
75	NaN	NaN
78	8.768508e+17	5.128045e+08
80	NaN	NaN
83	NaN	NaN
85	NaN	NaN
...
2301	NaN	NaN
2305	NaN	NaN
2306	NaN	NaN
2307	NaN	NaN
2310	NaN	NaN
2316	NaN	NaN
2320	NaN	NaN
2321	NaN	NaN
2322	NaN	NaN
2323	NaN	NaN
2324	NaN	NaN
2328	NaN	NaN
2329	NaN	NaN
2330	NaN	NaN
2331	NaN	NaN
2332	NaN	NaN
2336	NaN	NaN
2337	NaN	NaN

2338	NaN	NaN
2339	NaN	NaN
2340	NaN	NaN
2341	NaN	NaN
2342	NaN	NaN
2343	NaN	NaN
2344	NaN	NaN
2346	NaN	NaN
2348	NaN	NaN
2351	NaN	NaN
2353	NaN	NaN
2355	NaN	NaN

	retweeted_status_timestamp \
5	NaN
6	NaN
7	NaN
12	NaN
24	NaN
25	NaN
30	NaN
32	2017-07-15 02:44:07 +0000
35	NaN
37	NaN
41	NaN
42	NaN
43	NaN
46	NaN
47	NaN
55	NaN
56	NaN
57	NaN
59	NaN
61	NaN
62	NaN
64	NaN
72	NaN
73	2017-06-23 16:00:04 +0000
74	2015-11-24 03:51:38 +0000
75	NaN
78	2017-06-19 17:14:49 +0000
80	NaN
83	NaN
85	NaN
...	...
2301	NaN
2305	NaN
2306	NaN

2307 NaN
 2310 NaN
 2316 NaN
 2320 NaN
 2321 NaN
 2322 NaN
 2323 NaN
 2324 NaN
 2328 NaN
 2329 NaN
 2330 NaN
 2331 NaN
 2332 NaN
 2336 NaN
 2337 NaN
 2338 NaN
 2339 NaN
 2340 NaN
 2341 NaN
 2342 NaN
 2343 NaN
 2344 NaN
 2346 NaN
 2348 NaN
 2351 NaN
 2353 NaN
 2355 NaN

5 https://twitter.com/dog_rates/status/891087950875897856/photo/1
 6 <https://gofundme.com/ydvmve-surgery-for-jax>,https://twitter.com/dog_rates/status/890729181411237888/photo/1,https://twitter.com/dog_rates/status/889665388333682689/photo/1
 7 https://twitter.com/dog_rates/status/887343217045368832/video/1
 12 https://twitter.com/dog_rates/status/887101392804085760/photo/1
 24 https://twitter.com/dog_rates/status/886053434075471873,https://twitter.com/dog_rates/status/885517367337512960
 25 https://twitter.com/dog_rates/status/885167619883638784/photo/1,https://twitter.com/dog_rates/status/884441805382717440/photo/1
 30 NaN
 32 https://twitter.com/dog_rates/status/884162670584377345/photo/1
 35 https://twitter.com/dog_rates/status/883360690899218434/photo/1
 37 https://twitter.com/dog_rates/status/883117836046086144/photo/1,https://twitter.com/dog_rates/status/881536004380872706/video/1
 41 https://twitter.com/dog_rates/status/881268444196462592/photo/1
 42 https://twitter.com/dog_rates/status/880872448815771648/photo/1
 43 https://twitter.com/dog_rates/status/880872448815771648/photo/1
 46 https://twitter.com/dog_rates/status/880872448815771648/photo/1
 47 https://twitter.com/dog_rates/status/880872448815771648/photo/1
 49 NaN
 56 https://twitter.com/dog_rates/status/880872448815771648/photo/1
 57 https://twitter.com/dog_rates/status/880872448815771648/photo/1
 59 https://twitter.com/dog_rates/status/880872448815771648/photo/1

61 https://twitter.com/dog_rates/status/880221127280381952/photo/1,<https://twitter.c>
 62 https://twitter.com/dog_rates/status/880095782870896641/photo/1
 64 NaN
 72 <https://twitter.com/bbcworld/status/878599868507402241>
 73 <https://www.gofundme.com/3yd6y1c>,https://twitter.com/dog_rates/status/87828151100
 74 https://twitter.com/dog_rates/status/669000397445533696/photo/1
 75 <https://www.gofundme.com/3yd6y1c>,https://twitter.com/dog_rates/status/87828151100
 78 <https://twitter.com/rachel2195/status/876850772322988033/photo/1>,<https://twitter.c>
 80 https://twitter.com/dog_rates/status/877316821321428993/photo/1,<https://twitter.c>
 83 <https://twitter.com/mpstowerham/status/876162994446753793>
 85 https://twitter.com/dog_rates/status/876120275196170240/photo/1

 2301 https://twitter.com/dog_rates/status/667044094246576128/photo/1
 2305 https://twitter.com/dog_rates/status/666837028449972224/photo/1
 2306 https://twitter.com/dog_rates/status/666835007768551424/photo/1
 2307 https://twitter.com/dog_rates/status/666826780179869698/photo/1
 2310 https://twitter.com/dog_rates/status/666786068205871104/photo/1
 2316 https://twitter.com/dog_rates/status/666649482315059201/photo/1
 2320 https://twitter.com/dog_rates/status/666437273139982337/photo/1
 2321 https://twitter.com/dog_rates/status/666435652385423360/photo/1
 2322 https://twitter.com/dog_rates/status/666430724426358785/photo/1
 2323 https://twitter.com/dog_rates/status/666428276349472768/photo/1
 2324 https://twitter.com/dog_rates/status/666421158376562688/photo/1
 2328 https://twitter.com/dog_rates/status/666396247373291520/photo/1
 2329 https://twitter.com/dog_rates/status/666373753744588802/photo/1
 2330 https://twitter.com/dog_rates/status/666362758909284353/photo/1
 2331 https://twitter.com/dog_rates/status/666353288456101888/photo/1
 2332 https://twitter.com/dog_rates/status/666345417576210432/photo/1
 2336 https://twitter.com/dog_rates/status/666273097616637952/photo/1
 2337 https://twitter.com/dog_rates/status/666268910803644416/photo/1
 2338 https://twitter.com/dog_rates/status/666104133288665088/photo/1
 2339 https://twitter.com/dog_rates/status/666102155909144576/photo/1
 2340 https://twitter.com/dog_rates/status/666099513787052032/photo/1
 2341 https://twitter.com/dog_rates/status/666094000022159362/photo/1
 2342 https://twitter.com/dog_rates/status/666082916733198337/photo/1
 2343 https://twitter.com/dog_rates/status/666073100786774016/photo/1
 2344 https://twitter.com/dog_rates/status/666071193221509120/photo/1
 2346 https://twitter.com/dog_rates/status/666058600524156928/photo/1
 2348 https://twitter.com/dog_rates/status/666055525042405380/photo/1
 2351 https://twitter.com/dog_rates/status/666049248165822465/photo/1
 2353 https://twitter.com/dog_rates/status/666033412701032449/photo/1
 2355 https://twitter.com/dog_rates/status/666020888022790149/photo/1

	rating_numerator	rating_denominator	name	doggo	floofer	pupper	\
5	13	10	None	None	None	None	
6	13	10	Jax	None	None	None	
7	13	10	None	None	None	None	
12	13	10	None	None	None	None	

24	13	10	None	None	None	None
25	12	10	None	None	None	None
30	12	10	None	None	None	None
32	12	10	None	None	None	None
35	14	10	None	None	None	None
37	13	10	None	None	None	None
41	14	10	None	None	None	None
42	13	10	None	None	None	None
43	12	10	Yogi	doggo	None	None
46	13	10	Grizzwald	None	floofer	None
47	13	10	None	None	None	None
55	17	10	None	None	None	None
56	14	10	a	None	None	pupper
57	12	10	Elliot	None	None	None
59	12	10	None	None	None	None
61	12	10	Jesse	None	None	None
62	11	10	None	None	None	None
64	14	10	None	None	None	None
72	13	10	None	None	None	None
73	13	10	Shadow	None	None	None
74	11	10	Terrance	None	None	None
75	13	10	Shadow	None	None	None
78	14	10	None	None	None	pupper
80	13	10	Dante	None	None	None
83	14	10	None	None	None	None
85	13	10	Venti	None	None	None
...
2301	12	10	None	None	None	None
2305	3	10	None	None	None	None
2306	10	10	None	None	None	None
2307	12	10	None	None	None	None
2310	2	10	None	None	None	None
2316	4	10	None	None	None	None
2320	7	10	None	None	None	None
2321	10	10	None	None	None	None
2322	6	10	None	None	None	None
2323	7	10	None	None	None	None
2324	12	10	None	None	None	None
2328	9	10	None	None	None	None
2329	11	10	None	None	None	None
2330	6	10	None	None	None	None
2331	8	10	None	None	None	None
2332	10	10	None	None	None	None
2336	11	10	None	None	None	None
2337	10	10	None	None	None	None
2338	1	10	None	None	None	None
2339	11	10	None	None	None	None
2340	8	10	None	None	None	None

2341	9	10	None	None	None	None
2342	6	10	None	None	None	None
2343	10	10	None	None	None	None
2344	9	10	None	None	None	None
2346	8	10	the	None	None	None
2348	10	10	a	None	None	None
2351	5	10	None	None	None	None
2353	9	10	a	None	None	None
2355	8	10	None	None	None	None

	puppo
5	None
6	None
7	None
12	puppo
24	None
25	None
30	None
32	None
35	None
37	None
41	None
42	None
43	None
46	None
47	None
55	None
56	None
57	None
59	None
61	None
62	None
64	None
72	None
73	None
74	None
75	None
78	None
80	None
83	None
85	None
...	...
2301	None
2305	None
2306	None
2307	None
2310	None
2316	None


```

2320 None
2321 None
2322 None
2323 None
2324 None
2328 None
2329 None
2330 None
2331 None
2332 None
2336 None
2337 None
2338 None
2339 None
2340 None
2341 None
2342 None
2343 None
2344 None
2346 None
2348 None
2351 None
2353 None
2355 None

```

```
[1069 rows x 17 columns]
```

2.1 People usually (from tweets) call their dogs on the following ways:

- This is
- Meet
- Here is
- Here's
- Named
- His name is

Let's see if the previous calls exist in a tweet text and no name is stated for that

```
In [60]: print("Approximate number of tweets with name in a text is")
         (WeRateDogs_archive[WeRateDogs_archive['text'].str.contains("This is|meet|here is|here'")
```

Approximate number of tweets with name in a text is

```
Out[60]: 1609
```

```
In [61]: print("Number of tweets with names (Not equal to \"None\")")
         (WeRateDogs_archive[~WeRateDogs_archive['name'].str.contains("None" , case=False)]).count()
```

Number of tweets with names (Not equal to "None")

Out[61]: 1611

```
In [62]: print("Number of tweets with name less than 2 characters long, not real name ")

         (WeRateDogs_archive[WeRateDogs_archive['name'].str.len() < 2]).count()[0]
```

Number of tweets with name less than 2 characters long, not real name

Out[62]: 56

```
In [63]: print("Sort data by names\n we can see more unmeaningful names ( Not real, propably :\n
              WeRateDogs_archive.name.sort_values())
```

Sort data by names

we can see more unmeaningful names (Not real, propably : \)

```
Out[63]: 1035   Abby
          1021   Abby
           938   Ace
          1933   Acro
          1327   Adele
          1934   Aiden
           77    Aja
          480   Akumi
          820    Al
          875   Albert
          1954   Albert
          144   Albus
          412   Albus
          1115   Aldrick
          2046   Alejandro
          374   Alexander
          1189   Alexanderson
          486    Alf
          2238   Alfie
          661   Alfie
          367   Alfie
          1616   Alfie
          858   Alfie
           51   Alfy
          201   Alice
          1701   Alice
          1495   Amber
          1334   Ambrose
```

```

2078    Amy
2195    Amélie
      ...
852     my
988     not
335     not
1747    officially
1206    old
993     one
1936    one
924     one
369     one
2326    quite
169     quite
118     quite
193     quite
2030    space
22      such
2345    the
2346    the
1603    the
2037    the
2212    the
1815    the
1797    the
1527    the
1120    this
1121    unacceptable
1031    very
773     very
1097    very
819     very
1385    very

```

Name: name, Length: 2356, dtype: object

```

In [64]: print("Names with lower case, not real name as well")
         (WeRateDogs_archive.loc[(WeRateDogs_archive['name'].str.islower())]).count()[0]

```

Names with lower case, not real name as well

Out[64]: 109

```

In [65]: print("Showing text with (Name call) phrases and has lower case name")
         WeRateDogs_archive.loc[(WeRateDogs_archive['name'].str.islower()) & (WeRateDogs_archive

```

Showing text with (Name call) phrases and has lower case name

```

Out[65]:
      tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
22    887517139158093824  NaN                               NaN
56    881536004380872706  NaN                               NaN
118   869988702071779329  NaN                               NaN
169   859196978902773760  NaN                               NaN
193   855459453768019968  NaN                               NaN
335   832645525019123713  NaN                               NaN
369   828650029636317184  NaN                               NaN
542   806219024703037440  NaN                               NaN
649   792913359805018113  NaN                               NaN
759   778396591732486144  NaN                               NaN
773   776249906839351296  NaN                               NaN
801   772581559778025472  NaN                               NaN
819   770655142660169732  NaN                               NaN
822   770093767776997377  NaN                               NaN
852   765395769549590528  NaN                               NaN
924   755206590534418437  NaN                               NaN
988   748977405889503236  NaN                               NaN
992   748692773788876800  NaN                               NaN
993   748575535303884801  NaN                               NaN
1002  747885874273214464  NaN                               NaN
1004  747816857231626240  NaN                               NaN
1017  746872823977771008  NaN                               NaN
1025  746369468511756288  NaN                               NaN
1031  745422732645535745  NaN                               NaN
1040  744223424764059648  NaN                               NaN
1049  743222593470234624  NaN                               NaN
1063  741067306818797568  NaN                               NaN
1071  740214038584557568  NaN                               NaN
1097  736225175608430592  NaN                               NaN
1121  730924654643314689  NaN                               NaN
...      ...      ..
2191  668955713004314625  NaN                               NaN
2198  668815180734689280  NaN                               NaN
2204  668636665813057536  NaN                               NaN
2211  668614819948453888  NaN                               NaN
2212  668587383441514497  NaN                               NaN
2218  668507509523615744  NaN                               NaN
2222  668466899341221888  NaN                               NaN
2235  668171859951755264  NaN                               NaN
2249  667861340749471744  NaN                               NaN
2255  667773195014021121  NaN                               NaN
2264  667538891197542400  NaN                               NaN
2273  667470559035432960  NaN                               NaN
2287  667177989038297088  NaN                               NaN
2304  666983947667116034  NaN                               NaN
2311  666781792255496192  NaN                               NaN
2314  666701168228331520  NaN                               NaN

```

2326	666411507551481857	NaN	NaN
2327	666407126856765440	NaN	NaN
2333	666337882303524864	NaN	NaN
2334	666293911632134144	NaN	NaN
2335	666287406224695296	NaN	NaN
2345	666063827256086533	NaN	NaN
2346	666058600524156928	NaN	NaN
2347	666057090499244032	NaN	NaN
2348	666055525042405380	NaN	NaN
2349	666051853826850816	NaN	NaN
2350	666050758794694657	NaN	NaN
2352	666044226329800704	NaN	NaN
2353	666033412701032449	NaN	NaN
2354	666029285002620928	NaN	NaN

	timestamp \
22	2017-07-19 03:39:09 +0000
56	2017-07-02 15:32:16 +0000
118	2017-05-31 18:47:24 +0000
169	2017-05-02 00:04:57 +0000
193	2017-04-21 16:33:22 +0000
335	2017-02-17 17:38:57 +0000
369	2017-02-06 17:02:17 +0000
542	2016-12-06 19:29:28 +0000
649	2016-10-31 02:17:31 +0000
759	2016-09-21 00:53:04 +0000
773	2016-09-15 02:42:54 +0000
801	2016-09-04 23:46:12 +0000
819	2016-08-30 16:11:18 +0000
822	2016-08-29 03:00:36 +0000
852	2016-08-16 03:52:26 +0000
924	2016-07-19 01:04:16 +0000
988	2016-07-01 20:31:43 +0000
992	2016-07-01 01:40:41 +0000
993	2016-06-30 17:54:50 +0000
1002	2016-06-28 20:14:22 +0000
1004	2016-06-28 15:40:07 +0000
1017	2016-06-26 01:08:52 +0000
1025	2016-06-24 15:48:42 +0000
1031	2016-06-22 01:06:43 +0000
1040	2016-06-18 17:41:06 +0000
1049	2016-06-15 23:24:09 +0000
1063	2016-06-10 00:39:48 +0000
1071	2016-06-07 16:09:13 +0000
1097	2016-05-27 15:58:54 +0000
1121	2016-05-13 00:56:32 +0000
...	...
2191	2015-11-24 00:54:05 +0000

2198 2015-11-23 15:35:39 +0000
 2204 2015-11-23 03:46:18 +0000
 2211 2015-11-23 02:19:29 +0000
 2212 2015-11-23 00:30:28 +0000
 2218 2015-11-22 19:13:05 +0000
 2222 2015-11-22 16:31:42 +0000
 2235 2015-11-21 20:59:20 +0000
 2249 2015-11-21 00:25:26 +0000
 2255 2015-11-20 18:35:10 +0000
 2264 2015-11-20 03:04:08 +0000
 2273 2015-11-19 22:32:36 +0000
 2287 2015-11-19 03:10:02 +0000
 2304 2015-11-18 14:18:59 +0000
 2311 2015-11-18 00:55:42 +0000
 2314 2015-11-17 19:35:19 +0000
 2326 2015-11-17 00:24:19 +0000
 2327 2015-11-17 00:06:54 +0000
 2333 2015-11-16 19:31:45 +0000
 2334 2015-11-16 16:37:02 +0000
 2335 2015-11-16 16:11:11 +0000
 2345 2015-11-16 01:22:45 +0000
 2346 2015-11-16 01:01:59 +0000
 2347 2015-11-16 00:55:59 +0000
 2348 2015-11-16 00:49:46 +0000
 2349 2015-11-16 00:35:11 +0000
 2350 2015-11-16 00:30:50 +0000
 2352 2015-11-16 00:04:52 +0000
 2353 2015-11-15 23:21:54 +0000
 2354 2015-11-15 23:05:30 +0000

22 Twitter for iPhone
 56 Twitter for iPhone
 118 Twitter for iPhone
 169 Twitter for iPhone
 193 Twitter for iPhone
 335 Twitter Web Client
 369 Twitter for iPhone
 542 Twitter for iPhone
 649 Twitter for iPhone
 759 Twitter for iPhone
 773 Twitter for iPhone
 801 Twitter for iPhone
 819 Twitter for iPhone
 822 Twitter for iPhone
 852 Twitter for iPhone
 924 Twitter for iPhone
 988 Twitter for iPhone

992 Twitter for iPhone
 993 Twitter for iPhone
 1002 Twitter for iPhone
 1004 Twitter for iPhone
 1017 Twitter for iPhone
 1025 Twitter for iPhone
 1031 Twitter for iPhone
 1040 Twitter for iPhone
 1049 Twitter for iPhone
 1063 Twitter for iPhone
 1071 Twitter for iPhone
 1097 Twitter for iPhone
 1121 Twitter for iPhone
 ...
 2191 Twitter for iPhone
 2198 Twitter for iPhone
 2204 Twitter for iPhone
 2211 Twitter for iPhone
 2212 Vine - Make a Scene
 2218 Twitter for iPhone
 2222 Twitter for iPhone
 2235 Twitter for iPhone
 2249 Twitter for iPhone
 2255 Twitter Web Client
 2264 Twitter Web Client
 2273 Twitter Web Client
 2287 Twitter for iPhone
 2304 Twitter for iPhone
 2311 Twitter for iPhone
 2314 Twitter for iPhone
 2326 Twitter for iPhone
 2327 Twitter for iPhone
 2333 Twitter for iPhone
 2334 Twitter for iPhone
 2335 Twitter for iPhone
 2345 Twitter for iPhone
 2346 Twitter for iPhone
 2347 Twitter for iPhone
 2348 Twitter for iPhone
 2349 Twitter for iPhone
 2350 Twitter for iPhone
 2352 Twitter for iPhone
 2353 Twitter for iPhone
 2354 Twitter for iPhone

22 I've yet to rate a Venezuelan Hover Wiener. This is such an honor. 14/10 paw-insp
 56 Here is a pupper approaching maximum borkdrive. Zooming at never before seen speed

118 RT @dog_rates: We only rate dogs. This is quite clearly a smol broken polar bear.
 169 We only rate dogs. This is quite clearly a smol broken polar bear. We'd appreciat
 193 Guys, we only rate dogs. This is quite clearly a bulbasaur. Please only send dogs
 335 There's going to be a dog terminal at JFK Airport. This is not a drill. 10/10 \n
 369 Occasionally, we're sent fantastic stories. This is one of them. 14/10 for Grace
 542 We only rate dogs. Please stop sending in non-canines like this Freudian Poof Lic
 649 Here is a perfect example of someone who has their priorities in order. 13/10 for
 759 RT @dog_rates: This is an East African Chalupa Seal. We only rate dogs. Please on
 773 RT @dog_rates: We only rate dogs. Pls stop sending in non-canines like this Mongo
 801 Guys this is getting so out of hand. We only rate dogs. This is a Galapagos Speed
 819 We only rate dogs. Pls stop sending in non-canines like this Arctic Floof Kangaro
 822 RT @dog_rates: This is just downright precious af. 12/10 for both pupper and dogg
 852 This is my dog. Her name is Zoey. She knows I've been rating other dogs. She's no
 924 This is one of the most inspirational stories I've ever come across. I have no wo
 988 What jokester sent in a pic without a dog in it? This is not @rock_rates. This is
 992 That is Quizno. This is his beach. He does not tolerate human shenanigans on his
 993 This is one of the most reckless puppers I've ever seen. How she got a license in
 1002 This is a mighty rare blue-tailed hammer sherk. Human almost lost a limb trying t
 1004 Viewer discretion is advised. This is a terrible attack in progress. Not even in
 1017 This is a carrot. We only rate dogs. Please only send in dogs. You all really sho
 1025 This is an Iraqi Speed Kangaroo. It is not a dog. Please only send in dogs. I'm v
 1031 We only rate dogs. Pls stop sending in non-canines like this Jamaican Flop Seal.
 1040 This is actually a pupper and I'd pet it so well. 12/10\nhttps://t.co/RNqS7C4Y4N
 1049 This is a very rare Great Alaskan Bush Pupper. Hard to stumble upon without spook
 1063 This is just downright precious af. 12/10 for both pupper and doggo https://t.co/
 1071 This is getting incredibly frustrating. This is a Mexican Golden Beaver. We only
 1097 We only rate dogs. Please stop sending in non-canines like this Alaskan Flop Turt
 1121 We only rate dogs. Pls stop sending non-canines like this Bulgarian Eyeless Porch
 ...
 2191 This is a Slovakian Helter Skelter Feta named Leroi. Likes to skip on roofs. Good
 2198 This is a wild Toblerone from Papua New Guinea. Mouth always open. Addicted to ha
 2204 This is an Irish Rigatoni terrier named Berta. Completely made of rope. No eyes.
 2211 Here is a horned dog. Much grace. Can jump over moons (dam!). Paws not soft. Bad
 2212 Never forget this vine. You will not stop watching for at least 15 minutes. This
 2218 This is a Birmingham Quagmire named Chuk. Loves to relax and watch the game while
 2222 Here is a mother dog caring for her pups. Snazzy red mohawk. Doesn't wag tail. Pu
 2235 This is a Trans Siberian Kellogg named Alfonso. Huge ass eyeballs. Actually Dobby
 2249 This is a Shotokon Macadamia mix named Cheryl. Sophisticated af. Looks like a dis
 2255 This is a rare Hungarian Pinot named Jessiga. She is either mid-stroke or got stu
 2264 This is a southwest Coriander named Klint. Hat looks expensive. Still on house ar
 2273 This is a northern Wahoo named Kohl. He runs this town. Chases tumbleweeds. Draws
 2287 This is a Dasani Kingfisher from Maine. His name is Daryl. Daryl doesn't like bei
 2304 This is a curly Ticonderoga named Pepe. No feet. Loves to jet ski. 11/10 would hu
 2311 This is a purebred Bacardi named Octaviath. Can shoot spaghetti out of mouth. 10/
 2314 This is a golden Buckminsterfullerene named Johm. Drives trucks. Lumberjack (?).
 2326 This is quite the dog. Gets really excited when not in water. Not very soft tho.
 2327 This is a southern Vesuvius bumblegruff. Can drive a truck (wow). Made friends wi
 2333 This is an extremely rare horned Parthenon. Not amused. Wears shoes. Overall very

2334 This is a funny dog. Weird toes. Won't come down. Loves branch. Refuses to eat hi
 2335 This is an Albanian 3 1/2 legged Episcopalian. Loves well-polished hardwood floo
 2345 This is the happiest dog you will ever see. Very committed owner. Nice couch. 10/
 2346 Here is the Rand Paul of retrievers folks! He's probably good at poker. Can drink
 2347 My oh my. This is a rare blond Canadian terrier on wheels. Only \$8.98. Rather doc
 2348 Here is a Siberian heavily armored polar bear mix. Strong owner. 10/10 I would do
 2349 This is an odd dog. Hard on the outside but loving on the inside. Petting still f
 2350 This is a truly beautiful English Wilson Staff retriever. Has a nice phone. Privi
 2352 This is a purebred Piers Morgan. Loves to Netflix and chill. Always looks like he
 2353 Here is a very happy pup. Big fan of well-maintained decks. Just look at that tom
 2354 This is a western brown Mitsubishi terrier. Upset about leaf. Actually 2 dogs her

	retweeted_status_id	retweeted_status_user_id \
22	NaN	NaN
56	NaN	NaN
118	8.591970e+17	4.196984e+09
169	NaN	NaN
193	NaN	NaN
335	NaN	NaN
369	NaN	NaN
542	NaN	NaN
649	NaN	NaN
759	7.030419e+17	4.196984e+09
773	7.007478e+17	4.196984e+09
801	NaN	NaN
819	NaN	NaN
822	7.410673e+17	4.196984e+09
852	NaN	NaN
924	NaN	NaN
988	NaN	NaN
992	NaN	NaN
993	NaN	NaN
1002	NaN	NaN
1004	NaN	NaN
1017	NaN	NaN
1025	NaN	NaN
1031	NaN	NaN
1040	NaN	NaN
1049	NaN	NaN
1063	NaN	NaN
1071	NaN	NaN
1097	NaN	NaN
1121	NaN	NaN
...
2191	NaN	NaN
2198	NaN	NaN
2204	NaN	NaN
2211	NaN	NaN

2212	NaN	NaN
2218	NaN	NaN
2222	NaN	NaN
2235	NaN	NaN
2249	NaN	NaN
2255	NaN	NaN
2264	NaN	NaN
2273	NaN	NaN
2287	NaN	NaN
2304	NaN	NaN
2311	NaN	NaN
2314	NaN	NaN
2326	NaN	NaN
2327	NaN	NaN
2333	NaN	NaN
2334	NaN	NaN
2335	NaN	NaN
2345	NaN	NaN
2346	NaN	NaN
2347	NaN	NaN
2348	NaN	NaN
2349	NaN	NaN
2350	NaN	NaN
2352	NaN	NaN
2353	NaN	NaN
2354	NaN	NaN

	retweeted_status_timestamp \
22	NaN
56	NaN
118	2017-05-02 00:04:57 +0000
169	NaN
193	NaN
335	NaN
369	NaN
542	NaN
649	NaN
759	2016-02-26 02:20:37 +0000
773	2016-02-19 18:24:26 +0000
801	NaN
819	NaN
822	2016-06-10 00:39:48 +0000
852	NaN
924	NaN
988	NaN
992	NaN
993	NaN
1002	NaN

1004	NaN
1017	NaN
1025	NaN
1031	NaN
1040	NaN
1049	NaN
1063	NaN
1071	NaN
1097	NaN
1121	NaN
...	...
2191	NaN
2198	NaN
2204	NaN
2211	NaN
2212	NaN
2218	NaN
2222	NaN
2235	NaN
2249	NaN
2255	NaN
2264	NaN
2273	NaN
2287	NaN
2304	NaN
2311	NaN
2314	NaN
2326	NaN
2327	NaN
2333	NaN
2334	NaN
2335	NaN
2345	NaN
2346	NaN
2347	NaN
2348	NaN
2349	NaN
2350	NaN
2352	NaN
2353	NaN
2354	NaN

22	https://twitter.com/dog_rates/status/887517139158093824/video/1
56	https://twitter.com/dog_rates/status/881536004380872706/video/1
118	https://twitter.com/dog_rates/status/859196978902773760/video/1
169	https://twitter.com/dog_rates/status/859196978902773760/video/1
193	https://twitter.com/dog_rates/status/855459453768019968/photo/1 , https://twitter.com/dog_rates/status/855459453768019968/photo/1

335 <http://us.blastingnews.com/news/2017/02/jfk-announces-its-first-ever-ark-oasis-an>
 369 https://twitter.com/dog_rates/status/828650029636317184/photo/1,https://twitter.com/dog_rates/status/806219024703037440/photo/1
 542 https://twitter.com/dog_rates/status/806219024703037440/photo/1
 649 https://twitter.com/dog_rates/status/792913359805018113/photo/1,https://twitter.com/dog_rates/status/703041949650034688/photo/1,https://twitter.com/dog_rates/status/700747788515020802/photo/1
 759 https://twitter.com/dog_rates/status/703041949650034688/photo/1,https://twitter.com/dog_rates/status/772581559778025472/photo/1,https://twitter.com/dog_rates/status/770655142660169732/photo/1
 773 https://twitter.com/dog_rates/status/772581559778025472/photo/1,https://twitter.com/dog_rates/status/741067306818797568/photo/1,https://twitter.com/dog_rates/status/765395769549590528/photo/1
 801 https://twitter.com/dog_rates/status/772581559778025472/photo/1,https://twitter.com/dog_rates/status/755206590534418437/photo/1,https://twitter.com/dog_rates/status/748977405889503236/photo/1
 819 https://twitter.com/dog_rates/status/770655142660169732/photo/1
 822 https://twitter.com/dog_rates/status/741067306818797568/photo/1,https://twitter.com/dog_rates/status/748692773788876800/photo/1
 852 https://twitter.com/dog_rates/status/765395769549590528/photo/1
 924 https://twitter.com/dog_rates/status/755206590534418437/photo/1,https://twitter.com/dog_rates/status/748575535303884801/photo/1
 988 https://twitter.com/dog_rates/status/748977405889503236/photo/1
 992 https://twitter.com/dog_rates/status/748692773788876800/photo/1
 993 https://twitter.com/dog_rates/status/748575535303884801/photo/1
 1002 https://twitter.com/dog_rates/status/747885874273214464/photo/1,https://twitter.com/dog_rates/status/747816857231626240/photo/1
 1004 https://twitter.com/dog_rates/status/747816857231626240/photo/1
 1017 https://twitter.com/dog_rates/status/746872823977771008/photo/1,https://twitter.com/dog_rates/status/746369468511756288/photo/1
 1025 https://twitter.com/dog_rates/status/746369468511756288/photo/1
 1031 https://twitter.com/dog_rates/status/745422732645535745/photo/1
 1040 https://twitter.com/strange_animals/status/672108316018024452
 1049 https://twitter.com/dog_rates/status/743222593470234624/photo/1
 1063 https://twitter.com/dog_rates/status/741067306818797568/photo/1
 1071 https://twitter.com/dog_rates/status/740214038584557568/photo/1
 1097 https://twitter.com/dog_rates/status/736225175608430592/photo/1
 1121 https://twitter.com/dog_rates/status/730924654643314689/photo/1
 ...
 2191 https://twitter.com/dog_rates/status/668955713004314625/photo/1
 2198 https://twitter.com/dog_rates/status/668815180734689280/photo/1
 2204 https://twitter.com/dog_rates/status/668636665813057536/photo/1
 2211 https://twitter.com/dog_rates/status/668614819948453888/photo/1
 2212 <https://vine.co/v/ea00wvPTx9l>
 2218 https://twitter.com/dog_rates/status/668507509523615744/photo/1
 2222 https://twitter.com/dog_rates/status/668466899341221888/photo/1
 2235 https://twitter.com/dog_rates/status/668171859951755264/photo/1
 2249 https://twitter.com/dog_rates/status/667861340749471744/photo/1
 2255 https://twitter.com/dog_rates/status/667773195014021121/photo/1
 2264 https://twitter.com/dog_rates/status/667538891197542400/photo/1
 2273 https://twitter.com/dog_rates/status/667470559035432960/photo/1
 2287 https://twitter.com/dog_rates/status/667177989038297088/photo/1
 2304 https://twitter.com/dog_rates/status/666983947667116034/photo/1
 2311 https://twitter.com/dog_rates/status/666781792255496192/photo/1
 2314 https://twitter.com/dog_rates/status/666701168228331520/photo/1
 2326 https://twitter.com/dog_rates/status/666411507551481857/photo/1
 2327 https://twitter.com/dog_rates/status/666407126856765440/photo/1
 2333 https://twitter.com/dog_rates/status/666337882303524864/photo/1
 2334 https://twitter.com/dog_rates/status/666293911632134144/photo/1
 2335 https://twitter.com/dog_rates/status/666287406224695296/photo/1
 2345 https://twitter.com/dog_rates/status/666063827256086533/photo/1

2346 https://twitter.com/dog_rates/status/666058600524156928/photo/1
 2347 https://twitter.com/dog_rates/status/666057090499244032/photo/1
 2348 https://twitter.com/dog_rates/status/666055525042405380/photo/1
 2349 https://twitter.com/dog_rates/status/666051853826850816/photo/1
 2350 https://twitter.com/dog_rates/status/666050758794694657/photo/1
 2352 https://twitter.com/dog_rates/status/666044226329800704/photo/1
 2353 https://twitter.com/dog_rates/status/666033412701032449/photo/1
 2354 https://twitter.com/dog_rates/status/666029285002620928/photo/1

	rating_numerator	rating_denominator	name	doggo	floofer	\
22	14	10	such	None	None	
56	14	10	a	None	None	
118	12	10	quite	None	None	
169	12	10	quite	None	None	
193	12	10	quite	None	None	
335	10	10	not	None	None	
369	14	10	one	None	None	
542	11	10	incredibly	None	None	
649	13	10	a	None	None	
759	10	10	an	None	None	
773	11	10	very	None	None	
801	10	10	a	None	None	
819	11	10	very	None	None	
822	12	10	just	doggo	None	
852	13	10	my	None	None	
924	14	10	one	doggo	None	
988	10	10	not	None	None	
992	10	10	his	doggo	None	
993	6	10	one	None	None	
1002	8	10	a	None	None	
1004	4	10	a	None	None	
1017	11	10	a	None	None	
1025	9	10	an	None	None	
1031	9	10	very	None	None	
1040	12	10	actually	None	None	
1049	12	10	a	None	None	
1063	12	10	just	doggo	None	
1071	10	10	getting	None	None	
1097	10	10	very	None	None	
1121	9	10	unacceptable	None	None	
...	
2191	10	10	a	None	None	
2198	7	10	a	None	None	
2204	10	10	an	None	None	
2211	7	10	a	None	None	
2212	13	10	the	None	None	
2218	10	10	a	None	None	
2222	4	10	a	None	None	

2235	7	10	a	None	None
2249	9	10	a	None	None
2255	8	10	a	None	None
2264	9	10	a	None	None
2273	11	10	a	None	None
2287	8	10	a	None	None
2304	11	10	a	None	None
2311	10	10	a	None	None
2314	8	10	a	None	None
2326	2	10	quite	None	None
2327	7	10	a	None	None
2333	9	10	an	None	None
2334	3	10	a	None	None
2335	1	2	an	None	None
2345	10	10	the	None	None
2346	8	10	the	None	None
2347	9	10	a	None	None
2348	10	10	a	None	None
2349	2	10	an	None	None
2350	10	10	a	None	None
2352	6	10	a	None	None
2353	9	10	a	None	None
2354	7	10	a	None	None

	pupper	puppo
22	None	None
56	pupper	None
118	None	None
169	None	None
193	None	None
335	None	None
369	None	None
542	None	None
649	None	None
759	None	None
773	None	None
801	None	None
819	None	None
822	pupper	None
852	None	None
924	None	None
988	None	None
992	None	None
993	None	None
1002	None	None
1004	None	None
1017	None	None
1025	None	None

1031	None	None
1040	pupper	None
1049	pupper	None
1063	pupper	None
1071	None	None
1097	None	None
1121	None	None
...
2191	None	None
2198	None	None
2204	None	None
2211	None	None
2212	None	None
2218	None	None
2222	None	None
2235	None	None
2249	None	None
2255	None	None
2264	None	None
2273	None	None
2287	None	None
2304	None	None
2311	None	None
2314	None	None
2326	None	None
2327	None	None
2333	None	None
2334	None	None
2335	None	None
2345	None	None
2346	None	None
2347	None	None
2348	None	None
2349	None	None
2350	None	None
2352	None	None
2353	None	None
2354	None	None

[106 rows x 17 columns]

```
In [66]: print("To show tweets with rating in decimals, we can see nominator is not matching the
rating_columns = ['text' , 'rating_numerator' , 'rating_denominator']
WeRateDogs_archive[WeRateDogs_archive.text.str.contains(r"(\d+\.\d*/\d+)")][rating_col
```

To show tweets with rating in decimals, we can see nominator is not matching the actual text rat

/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:3: UserWarning: This pattern has ma

This is separate from the ipykernel package so we can avoid doing imports until

Out[66]:

```
45      This is Bella. She hopes her smile made you smile. If not, she is also offering y
340      RT @dog_rates: This is Logan, the Chow who lived. He solemnly swears he's up to l
695      This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
763      This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
1689     I've been told there's a slight possibility he's checking his mirror. We'll bump
1712     Here we have uncovered an entire battalion of holiday puppers. Average of 11.26/1
```

	rating_numerator	rating_denominator
45	5	10
340	75	10
695	75	10
763	27	10
1689	5	10
1712	26	10

```
In [67]: print("From the below results, 10 is not the only scale we have in the data. We have bi
          WeRateDogs_archive[WeRateDogs_archive.text.str.contains(r"(\d+\/\d+)") & (WeRateDogs_ar
```

From the below results, 10 is not the only scale we have in the data. We have big numbers which

/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:2: UserWarning: This pattern has ma

Out[67]:

```
342      @docmisterio account started on 11/15/15
433      The floofs have been released I repeat the floofs have been released. 84/70 https
784      RT @dog_rates: After so many requests, this is Bretagne. She was the last survivi
902      Why does this never happen at my front door... 165/150 https://t.co/HmwrdfEfUE
1068     After so many requests, this is Bretagne. She was the last surviving 9/11 search
1120     Say hello to this unbelievably well behaved squad of doggos. 204/170 would try to
1165     Happy 4/20 from the squad! 13/10 for all https://t.co/eV1diwds8a
1202     This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. Amaz
1228     Happy Saturday here's 9 puppers on a bench. 99/90 good work everybody https://t.c
1254     Here's a brigade of puppers. All look very prepared for whatever happens next. 80
1274     From left to right:\nCletus, Jerome, Alejandro, Burp, & Titson\nNone know whe
1351     Here is a whole flock of puppers. 60/50 I'll take the lot https://t.co/9dpcw6MdW
1433     Happy Wednesday here's a bucket of pups. 44/40 would pet all at once https://t.co
1598     Yes I do realize a rating of 4/20 would've been fitting. However, it would be unj
1634     Two sneaky puppers were not initially seen, moving the rating to 143/130. Please
1635     Someone help the girl is being mugged. Several are distracting her while two stea
1662     This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was ju
1663     I'm aware that I could've said 20/16, but here at WeRateDogs we are very professi
1779     IT'S PUPPERGEDDON. Total of 144/120 ...I think https://t.co/ZanVtAtvIq
```


1843 Here we have an entire platoon of puppies. Total score: 88/80 would pet all at on

	rating_numerator	rating_denominator
342	11	15
433	84	70
784	9	11
902	165	150
1068	9	11
1120	204	170
1165	4	20
1202	50	50
1228	99	90
1254	80	80
1274	45	50
1351	60	50
1433	44	40
1598	4	20
1634	143	130
1635	121	110
1662	7	11
1663	20	16
1779	144	120
1843	88	80

2.2 Now let's check image prediction dataframe

In [68]: image_predictions.head()

Out[68]:

	tweet_id	jpg_url	\
0	666020888022790149	https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg	
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg	
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg	
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg	

	img_num	p1	p1_conf	p1_dog	p2	\
0	1	Welsh_springer_spaniel	0.465074	True	collie	
1	1	redbone	0.506826	True	miniature_pinscher	
2	1	German_shepherd	0.596461	True	malinois	
3	1	Rhodesian_ridgeback	0.408143	True	redbone	
4	1	miniature_pinscher	0.560311	True	Rottweiler	

	p2_conf	p2_dog	p3	p3_conf	p3_dog
0	0.156665	True	Shetland_sheepdog	0.061428	True
1	0.074192	True	Rhodesian_ridgeback	0.072010	True
2	0.138584	True	bloodhound	0.116197	True
3	0.360687	True	miniature_pinscher	0.222752	True
4	0.243682	True	Doberman	0.154629	True

```
In [69]: image_predictions.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id      2075 non-null int64
jpg_url       2075 non-null object
img_num       2075 non-null int64
p1            2075 non-null object
p1_conf       2075 non-null float64
p1_dog        2075 non-null bool
p2            2075 non-null object
p2_conf       2075 non-null float64
p2_dog        2075 non-null bool
p3            2075 non-null object
p3_conf       2075 non-null float64
p3_dog        2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

```
In [70]: print("Number of tweets in WeRateDogs that doesn't have an image in image_predictions d
         WeRateDogs_archive.count()[0] - pd.merge(WeRateDogs_archive, image_predictions, on=['tw
```

Number of tweets in WeRateDogs that doesn't have an image in image_predictions dataframe is:

```
Out[70]: 281
```

```
In [71]: print("No image URL is null")
         image_predictions[image_predictions.jpg_url.isnull()]
```

No image URL is null

```
Out[71]: Empty DataFrame
         Columns: [tweet_id, jpg_url, img_num, p1, p1_conf, p1_dog, p2, p2_conf, p2_dog, p3, p3_
         Index: []
```

```
In [72]: print("Expanded URLs means tweet image. The number of tweets with no expanded URLs prov
         WeRateDogs_archive.expanded_urls.isnull().sum()
```

Expanded URLs means tweet image. The number of tweets with no expanded URLs provided is:

```
Out[72]: 59
```

2.3 Quality issues:

- Tweets with wrong rating numerator because of decimal number in text.
- Dog names with less than 2 letter.
- Dog names with lower-case (Not real or wrong).
- Dog names with inapplied names (e.g. "The" , "a").
- Missing dog stage values (All None).
- inconsistent denominator rating (150 , 120 , 130).
- inadequate datatypes (timestamp)
- Tweets with no expanded URLs.
- There are retweets records.

2.4 Tidiness issues:

- Dogs stages in different columns, better to be unified in one columns called Dog_Stage.
- Merge all 3 dataframes into one.

3 Cleaning

```
In [73]: # Don't play with the original data, let's make a copy
```

```
copy_WeRateDogs_archive = WeRateDogs_archive.copy()
copy_image_predictions = image_predictions.copy()
copy_retweet_like = df_json_cleaned.copy()
```

3.1 Tidiness issue 1 - One column for Dog_stage:

```
In [74]: #First, I will start with the dog stages and collapse them all into one columns.
```

```
copy_WeRateDogs_archive['Dog_stage'] = copy_WeRateDogs_archive['text'].str.extract('(dog|cat|bird|fish|insect|other)')
copy_WeRateDogs_archive.Dog_stage.isnull().count()
```

```
Out[74]: 2356
```

3.2 Tidiness issue 2 - Merge needed columns from the 3 dataframes into copy_WeRateDogs_archive :

```
In [75]: copy_image_predictions.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id    2075 non-null int64
jpg_url     2075 non-null object
img_num     2075 non-null int64
p1          2075 non-null object
p1_conf     2075 non-null float64
```

```

p1_dog      2075 non-null bool
p2          2075 non-null object
p2_conf     2075 non-null float64
p2_dog      2075 non-null bool
p3          2075 non-null object
p3_conf     2075 non-null float64
p3_dog      2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB

```

```
In [ ]:
```

```

In [76]: copy_WeRateDogs_archive = pd.merge(copy_WeRateDogs_archive, copy_image_predictions, on=

        copy_WeRateDogs_archive = pd.merge(copy_WeRateDogs_archive, copy_retweet_like, on='tweet

        copy_WeRateDogs_archive.info()

```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2073 entries, 0 to 2072
Data columns (total 31 columns):
tweet_id      2073 non-null int64
in_reply_to_status_id    23 non-null float64
in_reply_to_user_id      23 non-null float64
timestamp     2073 non-null object
source        2073 non-null object
text          2073 non-null object
retweeted_status_id      79 non-null float64
retweeted_status_user_id  79 non-null float64
retweeted_status_timestamp 79 non-null object
expanded_urls  2073 non-null object
rating_numerator  2073 non-null int64
rating_denominator  2073 non-null int64
name          2073 non-null object
doggo         2073 non-null object
floofer       2073 non-null object
pupper       2073 non-null object
puppo        2073 non-null object
Dog_stage     349 non-null object
jpg_url       2073 non-null object
img_num       2073 non-null int64
p1            2073 non-null object
p1_conf       2073 non-null float64
p1_dog        2073 non-null bool
p2            2073 non-null object
p2_conf       2073 non-null float64
p2_dog        2073 non-null bool

```

```

p3                2073 non-null object
p3_conf           2073 non-null float64
p3_dog            2073 non-null bool
retweet_count     2073 non-null int64
favorite_count    2073 non-null int64
dtypes: bool(3), float64(7), int64(6), object(15)
memory usage: 475.7+ KB

```

```

In [77]: columns_of_interest = ['tweet_id' , 'timestamp' , 'source' , 'text' , 'expanded_urls' ,
    copy_WeRateDogs_archive[columns_of_interest].head()

```

```

Out[77]:
      tweet_id      timestamp \
0  892420643555336193  2017-08-01 16:23:56 +0000
1  892177421306343426  2017-08-01 00:17:27 +0000
2  891815181378084864  2017-07-31 00:18:03 +0000
3  891689557279858688  2017-07-30 15:58:51 +0000
4  891327558926688256  2017-07-29 16:00:24 +0000

      source
0  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
1  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
3  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
4  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

0  This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/1
1  This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she's
2  This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall grass. You
3  This is Darla. She commenced a snooze mid meal. 13/10 happens to the best of us http
4  This is Franklin. He would like you to stop calling him "cute." He is a very fierce

0  https://twitter.com/dog_rates/status/892420643555336193/photo/1
1  https://twitter.com/dog_rates/status/892177421306343426/photo/1
2  https://twitter.com/dog_rates/status/891815181378084864/photo/1
3  https://twitter.com/dog_rates/status/891689557279858688/photo/1
4  https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://twitter.com/

      rating_numerator  rating_denominator      name  Dog_stage  retweet_count  \
0      13              10      Phineas   NaN          8853
1      13              10      Tilly     NaN          6514
2      12              10      Archie     NaN          4328
3      13              10      Darla      NaN          8964
4      12              10      Franklin   NaN          9774

      favorite_count

```

```
0 39467
1 33819
2 25461
3 42908
4 41048
```

Define Missing tweets images Code

```
In [78]: print("Drop empty expanded URLs")
        copy_WeRateDogs_archive.dropna(subset=['expanded_urls'] , inplace=True)
```

Drop empty expanded URLs

Test

```
In [79]: copy_WeRateDogs_archive.expanded_urls.isnull().sum()
```

```
Out[79]: 0
```

Define check and clean rating outliers Code

```
In [80]: print("check and clean rating outliers")
        wrong_deno_indexes = copy_WeRateDogs_archive[copy_WeRateDogs_archive.rating_denominator
        wrong_deno_indexes
```

check and clean rating outliers

```
Out[80]: Int64Index([ 345,  415,  734,  876,  924,  967, 1001, 1022, 1047, 1065, 1131,
                    1207, 1379, 1380, 1405, 1512, 1571, 2052],
                    dtype='int64')
```

Test

```
In [81]: print(copy_WeRateDogs_archive.tweet_id.count())
        copy_WeRateDogs_archive.drop(wrong_deno_indexes , inplace=True)

        print(copy_WeRateDogs_archive.tweet_id.count())
```

```
2073
2055
```

Define removing retweets Code

```
In [82]: print("removing retweets tweets")
        print("We have {} retweets in the dataframe".format ( (copy_WeRateDogs_archive[~copy_WeRateDogs_archiveretweeted_status])

        copy_WeRateDogs_archive = copy_WeRateDogs_archive[copy_WeRateDogs_archiveretweeted_status == False]

removing retweets tweets
We have 79 retweets in the dataframe
```

Test

```
In [83]: print("Now, we have {} retweets in the dataframe".format ( (copy_WeRateDogs_archive[~copy_WeRateDogs_archiveretweeted_status])

Now, we have 0 retweets in the dataframe
```

Define removing lower-case names Code

```
In [84]: print("There are {} tweets with lower-case dog name ".format( (copy_WeRateDogs_archive[copy_WeRateDogs_archive.name.str.islower()])

        copy_WeRateDogs_archive.drop( (copy_WeRateDogs_archive.loc[(copy_WeRateDogs_archive.name.str.islower())])

There are 95 tweets with lower-case dog name
```

Test

```
In [85]: print("Now, there are {} tweets with lower-case dog name ".format( (copy_WeRateDogs_archive[copy_WeRateDogs_archive.name.str.islower()])
        print(copy_WeRateDogs_archive.name.sort_values())
```

```
Now, there are 0 tweets with lower-case dog name
836      Abby
847      Abby
768      Ace
1656     Acro
1109     Adele
1657     Aiden
66      Aja
383     Akumi
663     Al
711     Albert
1676     Albert
328     Albus
120     Albus
919     Aldrick
1767    Alejandro
298     Alexander
991     Alexanderson
388     Alf
```

```

1956    Alfie
291     Alfie
695     Alfie
1364    Alfie
46      Alfy
161     Alice
1441    Alice
1259    Amber
1116    Ambrose
1799    Amy
1914    Amélie
629     Anakin
...
636     Winnie
212     Winnie
1558    Winston
808     Winston
324     Winston
1854    Winston
1036    Winston
660     Winston
225     Winston
452     Winston
713     Wishes
1221    Wyatt
145     Wyatt
326     Wyatt
1155    Yoda
38      Yogi
497     Yogi
1114    Zara
1925    Zeek
442     Zeke
17      Zeke
150     Zeke
1007    Zeus
928     Ziva
790     Zoe
1862    Zoey
8       Zoey
96      Zoey
125     Zooey
1602    Zuzu
Name: name, Length: 1881, dtype: object

```

Define removing lower-case names Code

```
In [86]: print("There are/is {} name(s) with less than 2 characters ".format( (copy_WeRateDogs
```



```
copy_WeRateDogs_archive.drop( (copy_WeRateDogs_archive[copy_WeRateDogs_archive['name']
```

There are/is 1 name(s) with less than 2 characters

Test

```
In [87]: print("Now, there are/is {} name(s) with less than 2 characters ".format( (copy_WeRat
```

Now, there are/is 0 name(s) with less than 2 characters

```
In [88]: decimal_texts = (copy_WeRateDogs_archive[copy_WeRateDogs_archive.text.str.contains(r"(\n
decimal_texts
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: UserWarning: This pattern has ma
    """Entry point for launching an IPython kernel.
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: FutureWarning: currently extract
    """Entry point for launching an IPython kernel.
```

```
Out[88]: 40      13.5/10
         558      9.75/10
         614     11.27/10
        1451     11.26/10
         Name: text, dtype: object
```

Define Correct decimal rating issue Code & Test

```
In [89]: print("Let's correct decimal rating values.")
         for key,value in decimal_texts.items():
             #print("value {} {}".format( key, value.split("/")[0] ))
             print("Rating numerator before {}".format( copy_WeRateDogs_archive.loc[key]['rati
             copy_WeRateDogs_archive.loc[key , 'rating_numerator'] = value.split("/")[0]
             print("Rating numerator After {}".format( copy_WeRateDogs_archive.loc[key]['ratin
```

Let's correct decimal rating values.

Rating numerator before 5

Rating numerator After 13.5

Rating numerator before 75

Rating numerator After 9.75

Rating numerator before 27

Rating numerator After 11.27

Rating numerator before 26

Rating numerator After 11.26

Define Change timestamp column type Code

```
In [90]: print("Change timestamp column type")
         copy_WeRateDogs_archive['timestamp'] = pd.to_datetime(copy_WeRateDogs_archive['timestamp'])
```

Change timestamp column type

```
In [91]: print("Change timestamp column type")
         copy_WeRateDogs_archive['timestamp'] = pd.to_datetime(copy_WeRateDogs_archive['timestamp'])

         copy_WeRateDogs_archive['rating_numerator'] = copy_WeRateDogs_archive['rating_numerator']
         copy_WeRateDogs_archive['rating_denominator'] = copy_WeRateDogs_archive['rating_denominator']
```

Change timestamp column type

Test

```
In [92]: copy_WeRateDogs_archive.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1880 entries, 0 to 2072
Data columns (total 31 columns):
tweet_id                1880 non-null int64
in_reply_to_status_id   22 non-null float64
in_reply_to_user_id     22 non-null float64
timestamp               1880 non-null datetime64[ns]
source                  1880 non-null object
text                    1880 non-null object
retweeted_status_id     0 non-null float64
retweeted_status_user_id 0 non-null float64
retweeted_status_timestamp 0 non-null object
expanded_urls           1880 non-null object
rating_numerator         1880 non-null float64
rating_denominator       1880 non-null float64
name                    1880 non-null object
doggo                   1880 non-null object
floofer                 1880 non-null object
pupper                  1880 non-null object
puppo                   1880 non-null object
Dog_stage                314 non-null object
jpg_url                  1880 non-null object
```

```

img_num          1880 non-null int64
p1               1880 non-null object
p1_conf          1880 non-null float64
p1_dog           1880 non-null bool
p2              1880 non-null object
p2_conf          1880 non-null float64
p2_dog           1880 non-null bool
p3              1880 non-null object
p3_conf          1880 non-null float64
p3_dog           1880 non-null bool
retweet_count    1880 non-null int64
favorite_count   1880 non-null int64
dtypes: bool(3), datetime64[ns](1), float64(9), int64(4), object(14)
memory usage: 511.4+ KB

```

Saving the cleaned dataframe to csv file:

```

In [93]: print("Save the cleaned dataframe to csv file")
         copy_WeRateDogs_archive.to_csv('copy_WeRateDogs_archive.csv')

```

Save the cleaned dataframe to csv file

In []:

4 Analysing

```

In [94]: columns_of_interest = ['timestamp' , 'source', 'rating_numerator' , 'rating_denominator']
         copy_WeRateDogs_archive[columns_of_interest].info()

```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1880 entries, 0 to 2072
Data columns (total 8 columns):
timestamp          1880 non-null datetime64[ns]
source             1880 non-null object
rating_numerator    1880 non-null float64
rating_denominator  1880 non-null float64
name               1880 non-null object
Dog_stage          314 non-null object
retweet_count       1880 non-null int64
favorite_count      1880 non-null int64
dtypes: datetime64[ns](1), float64(2), int64(2), object(3)
memory usage: 212.2+ KB

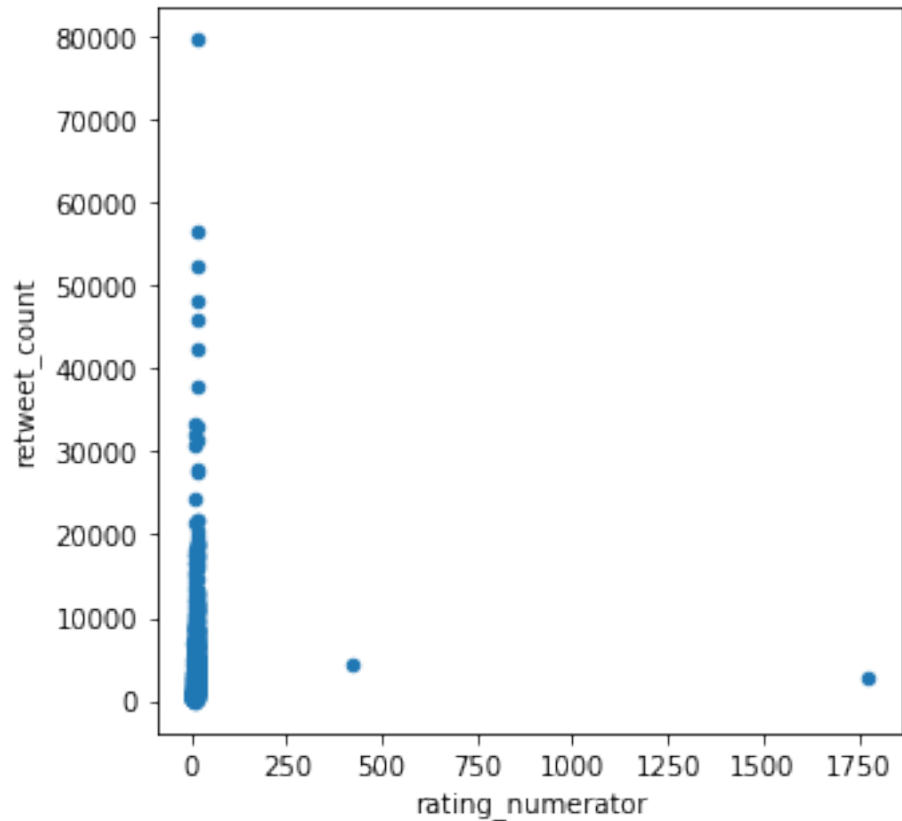
```

Plotting a scatter plot with rating and retweet_count

```

In [95]: copy_WeRateDogs_archive.plot(title='', x='rating_numerator', y='retweet_count' , kind='

```



It looks like that we found couple of rating_numerators outliers .. thanks to the scatter plot.
Let's check and clean that to avoid missing our analysis.

```
In [96]: copy_WeRateDogs_archive.rating_numerator.sort_values(ascending=False).head()
```

```
Out[96]: 802      1776.0
         1795      420.0
         173       14.0
         65       14.0
         431      14.0
         Name: rating_numerator, dtype: float64
```

```
In [97]: copy_WeRateDogs_archive[(copy_WeRateDogs_archive.index == 802) | (copy_WeRateDogs_archive.index == 1795)]
```

```
Out[97]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp
802	749981277374128128	NaN	NaN	2016-07-04 15:00:45
1795	670842764863651840	NaN	NaN	2015-11-29 05:52:33

```

802     <a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
1795     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

802     This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkh
1795     After so many requests... here you go.\n\nGood dogg. 420/10 https://t.co/yfAAo1gd

      retweeted_status_id  retweeted_status_user_id  \
802  NaN                      NaN
1795 NaN                      NaN

      retweeted_status_timestamp  \
802  NaN
1795  NaN

      expanded_urls  \
802  https://twitter.com/dog_rates/status/749981277374128128/photo/1
1795  https://twitter.com/dog_rates/status/670842764863651840/photo/1

      ...      p1_conf  p1_dog      p2  p2_conf  p2_dog  \
802      ...      0.533941  False  sunglasses  0.080822  False
1795      ...      0.096063  False  accordion  0.094075  False

      p3  p3_conf  p3_dog  retweet_count  favorite_count
802  sunglass  0.050776  False  2772          5569
1795  drumstick  0.061113  False  4324          7989

[2 rows x 31 columns]

```

We should drop the two tweets with the unreasonable rating values.

```
In [98]: copy_WeRateDogs_archive.drop([802,1795], inplace=True)
```

```
In [99]: copy_WeRateDogs_archive.rating_numerator.sort_values(ascending=False).head()
```

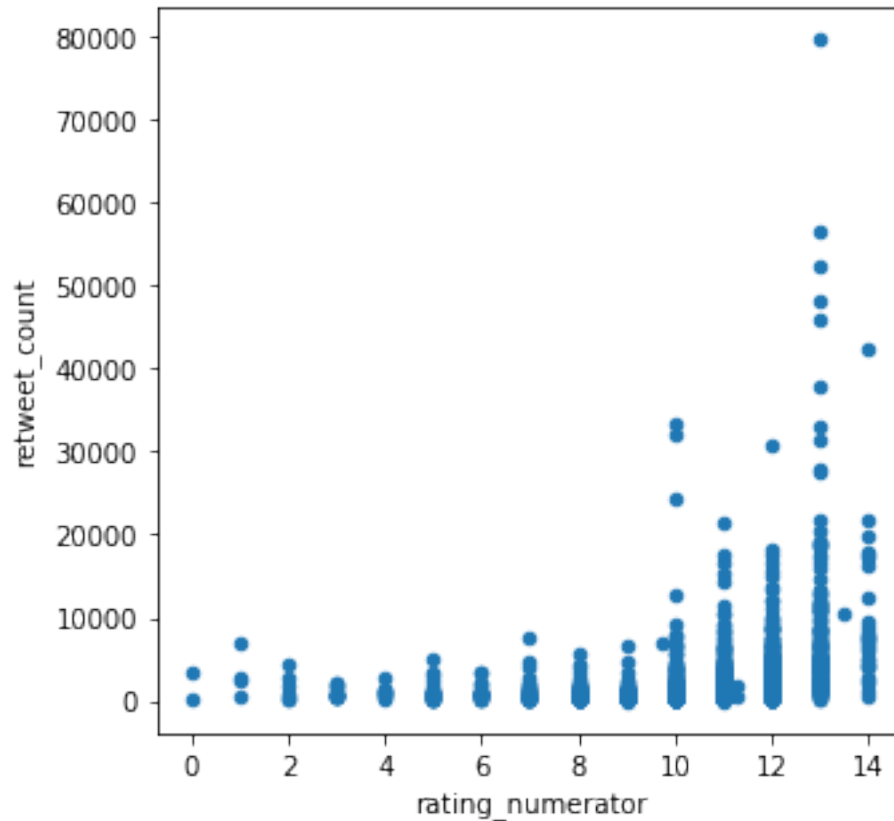
```

Out[99]: 702      14.0
         117      14.0
         153      14.0
         152      14.0
         103      14.0
         Name: rating_numerator, dtype: float64

```

Now, the scatter plot points distributions is clearer.

```
In [100]: copy_WeRateDogs_archive.plot(title='', x='rating_numerator', y='retweet_count', kind=
```



Showing counts of tweets per day of week. Useful in observing the activities volumn per day.
Overall, the highest activities take place on Monday.

```
In [101]: copy_WeRateDogs_archive_filtered_columns = copy_WeRateDogs_archive[columns_of_interest]
          copy_WeRateDogs_archive_filtered_columns.groupby(copy_WeRateDogs_archive_filtered_columns)
```

```
Out[101]:
```

	timestamp	source	rating_numerator	rating_denominator	name \
	count	count	count	count	count
timestamp					
Friday	260	260	260	260	260
Monday	305	305	305	305	305
Saturday	245	245	245	245	245
Sunday	242	242	242	242	242
Thursday	272	272	272	272	272
Tuesday	275	275	275	275	275
Wednesday	279	279	279	279	279

	Dog_stage	retweet_count	favorite_count
	count	count	count
timestamp			
Friday	37	260	260
Monday	55	305	305

Saturday	44	245	245
Sunday	32	242	242
Thursday	48	272	272
Tuesday	52	275	275
Wednesday	46	279	279

The highest retweet and favorite counts are 79,515.00 and 132,810.00 respectively.

```
In [102]: copy_WeRateDogs_archive.loc[:,['retweet_count','favorite_count']].max()
```

```
Out[102]: retweet_count      79515
          favorite_count    132810
          dtype: int64
```

```
In [103]: copy_WeRateDogs_archive.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1878 entries, 0 to 2072
Data columns (total 31 columns):
tweet_id                1878 non-null int64
in_reply_to_status_id   22 non-null float64
in_reply_to_user_id     22 non-null float64
timestamp               1878 non-null datetime64[ns]
source                  1878 non-null object
text                    1878 non-null object
retweeted_status_id     0 non-null float64
retweeted_status_user_id 0 non-null float64
retweeted_status_timestamp 0 non-null object
expanded_urls           1878 non-null object
rating_numerator        1878 non-null float64
rating_denominator      1878 non-null float64
name                    1878 non-null object
doggo                   1878 non-null object
floofer                1878 non-null object
pupper                 1878 non-null object
puppo                  1878 non-null object
Dog_stage               314 non-null object
jpg_url                 1878 non-null object
img_num                 1878 non-null int64
p1                      1878 non-null object
p1_conf                 1878 non-null float64
p1_dog                  1878 non-null bool
p2                      1878 non-null object
p2_conf                 1878 non-null float64
p2_dog                  1878 non-null bool
p3                      1878 non-null object
p3_conf                 1878 non-null float64
p3_dog                  1878 non-null bool
retweet_count           1878 non-null int64
```

```

favorite_count          1878 non-null int64
dtypes: bool(3), datetime64[ns](1), float64(9), int64(4), object(14)
memory usage: 511.0+ KB

```

We want to analyze the dog stage rating, retweet and favorite totals. As a result, we will see which dog stage is most popular and high rating across the data.

Unfortunately, we found another issue that needs to be cleaned. Upper and lower case duplicate records of Dog_Stage column. Let's correct that and unify that to lower-case.

```

In [104]: Dog_stage_columns = ['Dog_stage' , 'rating_numerator' , 'rating_denominator' , 'retweet_count' , 'favorite_count']
copy_WeRateDogs_archive_filtered_columns = copy_WeRateDogs_archive[Dog_stage_columns]
copy_WeRateDogs_archive_filtered_columns.groupby(copy_WeRateDogs_archive_filtered_columns['Dog_stage']).sum()

```

```

Out[104]:
           rating_numerator rating_denominator retweet_count favorite_count
           sum              sum              sum              sum
Dog_stage
Doggo      51.00             40.0             68380             156749
PUPPER     23.00             20.0             3789             9260
Pupper     60.27             50.0             14674             46949
Puppo      13.00             10.0             4319             19763
doggo      789.00            680.0            415014            1162562
pupper     2176.26           2060.0            476367            1464767
puppo      340.00            280.0            189760             637250

```

```

In [105]: lower_dog_stages = copy_WeRateDogs_archive.loc[:, 'Dog_stage'].str.lower()

```

```

In [106]: copy_WeRateDogs_archive.loc[:, ('Dog_stage')] = lower_dog_stages

```

Let's do the analysis again after fixing lower/upper dog stages issue.

```

In [107]: Dog_stage_columns = ['Dog_stage' , 'rating_numerator' , 'rating_denominator' , 'retweet_count' , 'favorite_count']
copy_WeRateDogs_archive_filtered_columns = copy_WeRateDogs_archive[Dog_stage_columns]
copy_WeRateDogs_archive_filtered_columns.groupby(copy_WeRateDogs_archive_filtered_columns['Dog_stage']).sum()

```

```

Out[107]:
           rating_numerator rating_denominator retweet_count favorite_count
           sum              sum              sum              sum
Dog_stage
doggo      840.00             720.0            483394            1319311
pupper     2259.53            2130.0            494830            1520976
puppo      353.00             290.0            194079             657013

```

The most popular dog stage is **Pupper**

```

In [ ]:

```