

wrangle_act

April 27, 2020

0.1 Importing libraries and getting data into jupyter

```
In [4]: #importting libraries
import requests
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

In [2]: #TSV url
url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictio

In [3]: # request to tsu url and store data into file
response = requests.get(url)
with open (url.split('/')[ -1], mode='wb') as file :
    file.write(response.content)

In [4]: #checking file list
os.listdir()

Out[4]: ['wrangle_act.ipynb',
        '.ipynb_checkpoints',
        'twitter-archive-enhanced.csv',
        'image-predictions.tsv']

In [5]: # make list of tweets id
twitter_archive_enh = pd.read_csv('twitter-archive-enhanced.csv')
tweet_ids = list(twitter_archive_enh.tweet_id)
len(tweet_ids)

Out[5]: 2356

In [5]: import tweepy
from timeit import default_timer as timer
import json

In [ ]: consumer_key = 'bvzJAYUMWhqQTmZid8yqnVXYA'
consumer_secret = 'jTrHT7bDVhAShmaW1Z3BRXHrVj3fixyYG4FGOEpgUirWBqKPCk'
access_token = '372271092-vHKFE2yNEf8JM3nj8RftPq3kREFzcku0Xtq74gIF'
```

```

access_secret = 'yZpupERQaYD2f6If0b1obi3l8hf03eK3VZaBS3uMeygii'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)

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

count = 0
fails_dict = {}
start = timer()
# Save each tweet's returned JSON as a new line in a .txt file
with open('tweet_json.txt', 'w') as outfile:
    # This loop will likely take 20-30 minutes to run because of Twitter's rate limit
    for tweet_id in tweet_ids:
        count += 1
        print(str(count) + ": " + str(tweet_id))
        try:
            tweet = api.get_status(tweet_id, tweet_mode='extended')
            print("Success")
            json.dump(tweet._json, outfile)
            outfile.write('\n')
        except tweepy.TweepError as e:
            print("Fail")
            fails_dict[tweet_id] = e
            pass
end = timer()
print(end - start)
print(fails_dict)

```

```

In [7]: #checking file list
        os.listdir()

```

```

Out[7]: ['wrangle_act.ipynb',
         '.ipynb_checkpoints',
         'twitter-archive-enhanced.csv',
         'image-predictions.tsv']

```

0.2 See overview of all data

```

In [6]: # importing data to pandas dataframe
        twitter_archive_enh = pd.read_csv('twitter-archive-enhanced.csv')
        image_predictions = pd.read_csv('image-predictions.tsv', sep='\t')

In [7]: #open json data
        json_data = [json.loads(line) for line in open('tweet_json.txt', 'r')]

In [8]: #get list of keys in dictionary
        list(json_data[0].keys())

```

```
Out[8]: ['created_at',
        'id',
        'id_str',
        'full_text',
        'truncated',
        'display_text_range',
        'entities',
        'extended_entities',
        'source',
        'in_reply_to_status_id',
        'in_reply_to_status_id_str',
        'in_reply_to_user_id',
        'in_reply_to_user_id_str',
        'in_reply_to_screen_name',
        'user',
        'geo',
        'coordinates',
        'place',
        'contributors',
        'is_quote_status',
        'retweet_count',
        'favorite_count',
        'favorited',
        'retweeted',
        'possibly_sensitive',
        'possibly_sensitive_appealable',
        'lang']
```

```
In [9]: json_data[0]['full_text']
```

```
Out[9]: "This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/10 h
```

```
In [10]: # take several parameter from dictionary data into list
         json_conv = []
```

```
         for datum in json_data:
             json_conv.append([datum['id'], datum['full_text'], datum['in_reply_to_status_id'], d
```

```
In [11]: #change list into df
         json_df = pd.DataFrame(json_conv, columns=['tweet_id', 'full_text', 'in_reply_to_status
```

0.2.1 twitter_archive_enhanced

```
In [12]: # see head and tail of twitter archive enhanced
         twitter_archive_enh
```

```
Out[12]:
```

	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
...
2326	666411507551481857	NaN	NaN
2327	666407126856765440	NaN	NaN
2328	666396247373291520	NaN	NaN
2329	666373753744588802	NaN	NaN
2330	666362758909284353	NaN	NaN
2331	666353288456101888	NaN	NaN
2332	666345417576210432	NaN	NaN
2333	666337882303524864	NaN	NaN
2334	666293911632134144	NaN	NaN
2335	666287406224695296	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

2345	666063827256086533	NaN	NaN
2346	666058600524156928	NaN	NaN
2347	666057090499244032	NaN	NaN
2348	666055525042405380	NaN	NaN
2349	666051853826850816	NaN	NaN
2350	666050758794694657	NaN	NaN
2351	666049248165822465	NaN	NaN
2352	666044226329800704	NaN	NaN
2353	666033412701032449	NaN	NaN
2354	666029285002620928	NaN	NaN
2355	666020888022790149	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
...	...
2326	2015-11-17 00:24:19 +0000
2327	2015-11-17 00:06:54 +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
 2333 2015-11-16 19:31:45 +0000
 2334 2015-11-16 16:37:02 +0000
 2335 2015-11-16 16:11:11 +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
 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
 2351 2015-11-16 00:24:50 +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
 2355 2015-11-15 22:32:08 +0000

source \
 0 <a href="http://twitter.com/download/iphone" r...
 1 <a href="http://twitter.com/download/iphone" r...
 2 <a href="http://twitter.com/download/iphone" r...
 3 <a href="http://twitter.com/download/iphone" r...
 4 <a href="http://twitter.com/download/iphone" r...
 5 <a href="http://twitter.com/download/iphone" r...
 6 <a href="http://twitter.com/download/iphone" r...
 7 <a href="http://twitter.com/download/iphone" r...
 8 <a href="http://twitter.com/download/iphone" r...
 9 <a href="http://twitter.com/download/iphone" r...
 10 <a href="http://twitter.com/download/iphone" r...
 11 <a href="http://twitter.com/download/iphone" r...
 12 <a href="http://twitter.com/download/iphone" r...
 13 <a href="http://twitter.com/download/iphone" r...
 14 <a href="http://twitter.com/download/iphone" r...
 15 <a href="http://twitter.com/download/iphone" r...
 16 <a href="http://twitter.com/download/iphone" r...
 17 <a href="http://twitter.com/download/iphone" r...
 18 <a href="http://twitter.com/download/iphone" r...
 19 <a href="http://twitter.com/download/iphone" r...

20 <a href="http://twitter.com/download/iphone" r...
 21 <a href="http://twitter.com/download/iphone" r...
 22 <a href="http://twitter.com/download/iphone" r...
 23 <a href="http://twitter.com/download/iphone" r...
 24 <a href="http://twitter.com/download/iphone" r...
 25 <a href="http://twitter.com/download/iphone" r...
 26 <a href="http://twitter.com/download/iphone" r...
 27 <a href="http://twitter.com/download/iphone" r...
 28 <a href="http://twitter.com/download/iphone" r...
 29 <a href="http://twitter.com/download/iphone" r...

 2326 <a href="http://twitter.com/download/iphone" r...
 2327 <a href="http://twitter.com/download/iphone" r...
 2328 <a href="http://twitter.com/download/iphone" r...
 2329 <a href="http://twitter.com/download/iphone" r...
 2330 <a href="http://twitter.com/download/iphone" r...
 2331 <a href="http://twitter.com/download/iphone" r...
 2332 <a href="http://twitter.com/download/iphone" r...
 2333 <a href="http://twitter.com/download/iphone" r...
 2334 <a href="http://twitter.com/download/iphone" r...
 2335 <a href="http://twitter.com/download/iphone" r...
 2336 <a href="http://twitter.com/download/iphone" r...
 2337 <a href="http://twitter.com/download/iphone" r...
 2338 <a href="http://twitter.com/download/iphone" r...
 2339 <a href="http://twitter.com/download/iphone" r...
 2340 <a href="http://twitter.com/download/iphone" r...
 2341 <a href="http://twitter.com/download/iphone" r...
 2342 <a href="http://twitter.com/download/iphone" r...
 2343 <a href="http://twitter.com/download/iphone" r...
 2344 <a href="http://twitter.com/download/iphone" r...
 2345 <a href="http://twitter.com/download/iphone" r...
 2346 <a href="http://twitter.com/download/iphone" r...
 2347 <a href="http://twitter.com/download/iphone" r...
 2348 <a href="http://twitter.com/download/iphone" r...
 2349 <a href="http://twitter.com/download/iphone" r...
 2350 <a href="http://twitter.com/download/iphone" r...
 2351 <a href="http://twitter.com/download/iphone" r...
 2352 <a href="http://twitter.com/download/iphone" r...
 2353 <a href="http://twitter.com/download/iphone" r...
 2354 <a href="http://twitter.com/download/iphone" r...
 2355 <a href="http://twitter.com/download/iphone" r...

	text	retweeted_status_id \
0	This is Phineas. He's a mystical boy. Only eve...	NaN
1	This is Tilly. She's just checking pup on you...	NaN
2	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN

5	Here we have a majestic great white breaching ...	NaN
6	Meet Jax. He enjoys ice cream so much he gets ...	NaN
7	When you watch your owner call another dog a g...	NaN
8	This is Zoey. She doesn't want to be one of th...	NaN
9	This is Cassie. She is a college pup. Studying...	NaN
10	This is Koda. He is a South Australian decksha...	NaN
11	This is Bruno. He is a service shark. Only get...	NaN
12	Here's a puppo that seems to be on the fence a...	NaN
13	This is Ted. He does his best. Sometimes that'...	NaN
14	This is Stuart. He's sporting his favorite fan...	NaN
15	This is Oliver. You're witnessing one of his m...	NaN
16	This is Jim. He found a fren. Taught him how t...	NaN
17	This is Zeke. He has a new stick. Very proud o...	NaN
18	This is Ralphus. He's powering up. Attempting ...	NaN
19	RT @dog_rates: This is Canela. She attempted s...	8.874740e+17
20	This is Gerald. He was just told he didn't get...	NaN
21	This is Jeffrey. He has a monopoly on the pool...	NaN
22	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN
23	This is Canela. She attempted some fancy porch...	NaN
24	You may not have known you needed to see this ...	NaN
25	This... is a Jubilant Antarctic House Bear. We...	NaN
26	This is Maya. She's very shy. Rarely leaves he...	NaN
27	This is Mingus. He's a wonderful father to his...	NaN
28	This is Derek. He's late for a dog meeting. 13...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
...
2326	This is quite the dog. Gets really excited whe...	NaN
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN
2328	Oh goodness. A super rare northeast Qdoba kang...	NaN
2329	Those are sunglasses and a jean jacket. 11/10 ...	NaN
2330	Unique dog here. Very small. Lives in containe...	NaN
2331	Here we have a mixed Asiago from the Galápagos...	NaN
2332	Look at this jokester thinking seat belt laws ...	NaN
2333	This is an extremely rare horned Parthenon. No...	NaN
2334	This is a funny dog. Weird toes. Won't come do...	NaN
2335	This is an Albanian 3 1/2 legged Episcopalian...	NaN
2336	Can take selfies 11/10 https://t.co/ws2AMaWpPW	NaN
2337	Very concerned about fellow dog trapped in com...	NaN
2338	Not familiar with this breed. No tail (weird)...	NaN
2339	Oh my. Here you are seeing an Adobe Setter giv...	NaN
2340	Can stand on stump for what seems like a while...	NaN
2341	This appears to be a Mongolian Presbyterian mi...	NaN
2342	Here we have a well-established sunblockerspan...	NaN
2343	Let's hope this flight isn't Malaysian (lol). ...	NaN
2344	Here we have a northern speckled Rhododendron...	NaN
2345	This is the happiest dog you will ever see. Ve...	NaN
2346	Here is the Rand Paul of retrievers folks! He'...	NaN
2347	My oh my. This is a rare blond Canadian terrie...	NaN

2348	Here is a Siberian heavily armored polar bear ...	NaN
2349	This is an odd dog. Hard on the outside but lo...	NaN
2350	This is a truly beautiful English Wilson Staff...	NaN
2351	Here we have a 1949 1st generation vulpix. Enj...	NaN
2352	This is a purebred Piers Morgan. Loves to Netf...	NaN
2353	Here is a very happy pup. Big fan of well-main...	NaN
2354	This is a western brown Mitsubishi terrier. Up...	NaN
2355	Here we have a Japanese Irish Setter. Lost eye...	NaN

	retweeted_status_user_id	retweeted_status_timestamp	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	
7	NaN	NaN	
8	NaN	NaN	
9	NaN	NaN	
10	NaN	NaN	
11	NaN	NaN	
12	NaN	NaN	
13	NaN	NaN	
14	NaN	NaN	
15	NaN	NaN	
16	NaN	NaN	
17	NaN	NaN	
18	NaN	NaN	
19	4.196984e+09	2017-07-19 00:47:34	+0000
20	NaN	NaN	
21	NaN	NaN	
22	NaN	NaN	
23	NaN	NaN	
24	NaN	NaN	
25	NaN	NaN	
26	NaN	NaN	
27	NaN	NaN	
28	NaN	NaN	
29	NaN	NaN	
...	
2326	NaN	NaN	
2327	NaN	NaN	
2328	NaN	NaN	
2329	NaN	NaN	
2330	NaN	NaN	
2331	NaN	NaN	
2332	NaN	NaN	

2333	NaN	NaN
2334	NaN	NaN
2335	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
2345	NaN	NaN
2346	NaN	NaN
2347	NaN	NaN
2348	NaN	NaN
2349	NaN	NaN
2350	NaN	NaN
2351	NaN	NaN
2352	NaN	NaN
2353	NaN	NaN
2354	NaN	NaN
2355	NaN	NaN

	expanded_urls	rating_numerator \
0	https://twitter.com/dog_rates/status/892420643...	13
1	https://twitter.com/dog_rates/status/892177421...	13
2	https://twitter.com/dog_rates/status/891815181...	12
3	https://twitter.com/dog_rates/status/891689557...	13
4	https://twitter.com/dog_rates/status/891327558...	12
5	https://twitter.com/dog_rates/status/891087950...	13
6	https://gofundme.com/ydvmve-surgery-for-jax,ht...	13
7	https://twitter.com/dog_rates/status/890729181...	13
8	https://twitter.com/dog_rates/status/890609185...	13
9	https://twitter.com/dog_rates/status/890240255...	14
10	https://twitter.com/dog_rates/status/890006608...	13
11	https://twitter.com/dog_rates/status/889880896...	13
12	https://twitter.com/dog_rates/status/889665388...	13
13	https://twitter.com/dog_rates/status/889638837...	12
14	https://twitter.com/dog_rates/status/889531135...	13
15	https://twitter.com/dog_rates/status/889278841...	13
16	https://twitter.com/dog_rates/status/888917238...	12
17	https://twitter.com/dog_rates/status/888804989...	13
18	https://twitter.com/dog_rates/status/888554962...	13
19	https://twitter.com/dog_rates/status/887473957...	13
20	https://twitter.com/dog_rates/status/888078434...	12
21	https://twitter.com/dog_rates/status/887705289...	13
22	https://twitter.com/dog_rates/status/887517139...	14

23	https://twitter.com/dog_rates/status/887473957...	13
24	https://twitter.com/dog_rates/status/887343217...	13
25	https://twitter.com/dog_rates/status/887101392...	12
26	https://twitter.com/dog_rates/status/886983233...	13
27	https://www.gofundme.com/mingusneedsus , https://...	13
28	https://twitter.com/dog_rates/status/886680336...	13
29	https://twitter.com/dog_rates/status/886366144...	12
...
2326	https://twitter.com/dog_rates/status/666411507...	2
2327	https://twitter.com/dog_rates/status/666407126...	7
2328	https://twitter.com/dog_rates/status/666396247...	9
2329	https://twitter.com/dog_rates/status/666373753...	11
2330	https://twitter.com/dog_rates/status/666362758...	6
2331	https://twitter.com/dog_rates/status/666353288...	8
2332	https://twitter.com/dog_rates/status/666345417...	10
2333	https://twitter.com/dog_rates/status/666337882...	9
2334	https://twitter.com/dog_rates/status/666293911...	3
2335	https://twitter.com/dog_rates/status/666287406...	1
2336	https://twitter.com/dog_rates/status/666273097...	11
2337	https://twitter.com/dog_rates/status/666268910...	10
2338	https://twitter.com/dog_rates/status/666104133...	1
2339	https://twitter.com/dog_rates/status/666102155...	11
2340	https://twitter.com/dog_rates/status/666099513...	8
2341	https://twitter.com/dog_rates/status/666094000...	9
2342	https://twitter.com/dog_rates/status/666082916...	6
2343	https://twitter.com/dog_rates/status/666073100...	10
2344	https://twitter.com/dog_rates/status/666071193...	9
2345	https://twitter.com/dog_rates/status/666063827...	10
2346	https://twitter.com/dog_rates/status/666058600...	8
2347	https://twitter.com/dog_rates/status/666057090...	9
2348	https://twitter.com/dog_rates/status/666055525...	10
2349	https://twitter.com/dog_rates/status/666051853...	2
2350	https://twitter.com/dog_rates/status/666050758...	10
2351	https://twitter.com/dog_rates/status/666049248...	5
2352	https://twitter.com/dog_rates/status/666044226...	6
2353	https://twitter.com/dog_rates/status/666033412...	9
2354	https://twitter.com/dog_rates/status/666029285...	7
2355	https://twitter.com/dog_rates/status/666020888...	8

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10	Phineas	None	None	None	None
1	10	Tilly	None	None	None	None
2	10	Archie	None	None	None	None
3	10	Darla	None	None	None	None
4	10	Franklin	None	None	None	None
5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None

8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
19	10	Canela	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None
2343	10	None	None	None	None	None
2344	10	None	None	None	None	None
2345	10	the	None	None	None	None
2346	10	the	None	None	None	None
2347	10	a	None	None	None	None
2348	10	a	None	None	None	None
2349	10	an	None	None	None	None
2350	10	a	None	None	None	None

2351	10	None	None	None	None	None
2352	10	a	None	None	None	None
2353	10	a	None	None	None	None
2354	10	a	None	None	None	None
2355	10	None	None	None	None	None

[2356 rows x 17 columns]

- find wrong column name, floof instead of floofer
- timestamp has date and time in one column

```
In [13]: twitter_archive_enh.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
```

timestamp and retweeted_status_timestamp must be in date type in_reply_to_status_id and in_reply_to_user_id must be in int type

```
In [14]: twitter_archive_enh[twitter_archive_enh['retweeted_status_id'].notnull()]
```

```
Out[14]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
19	888202515573088257	NaN	NaN	
32	886054160059072513	NaN	NaN	
36	885311592912609280	NaN	NaN	
68	879130579576475649	NaN	NaN	
73	878404777348136964	NaN	NaN	
74	878316110768087041	NaN	NaN	

78	877611172832227328	NaN	NaN
91	874434818259525634	NaN	NaN
95	873697596434513921	NaN	NaN
97	873337748698140672	NaN	NaN
101	872668790621863937	NaN	NaN
109	871166179821445120	NaN	NaN
118	869988702071779329	NaN	NaN
124	868639477480148993	NaN	NaN
130	867072653475098625	NaN	NaN
132	866816280283807744	NaN	NaN
137	866094527597207552	NaN	NaN
146	863471782782697472	NaN	NaN
155	861769973181624320	NaN	NaN
159	860981674716409858	NaN	NaN
160	860924035999428608	NaN	NaN
165	860177593139703809	NaN	NaN
171	858860390427611136	NaN	NaN
180	857062103051644929	NaN	NaN
182	856602993587888130	NaN	NaN
185	856330835276025856	NaN	NaN
194	855245323840757760	NaN	NaN
195	855138241867124737	NaN	NaN
204	852936405516943360	NaN	NaN
211	851953902622658560	NaN	NaN
...
784	775096608509886464	NaN	NaN
794	773336787167145985	NaN	NaN
800	772615324260794368	NaN	NaN
811	771171053431250945	NaN	NaN
815	771004394259247104	NaN	NaN
818	770743923962707968	NaN	NaN
822	770093767776997377	NaN	NaN
826	769335591808995329	NaN	NaN
829	768909767477751808	NaN	NaN
833	768554158521745409	NaN	NaN
841	766864461642756096	NaN	NaN
847	766078092750233600	NaN	NaN
860	763167063695355904	NaN	NaN
868	761750502866649088	NaN	NaN
872	761371037149827077	NaN	NaN
885	760153949710192640	NaN	NaN
890	759566828574212096	NaN	NaN
895	759159934323924993	NaN	NaN
908	757729163776290825	NaN	NaN
911	757597904299253760	NaN	NaN
926	754874841593970688	NaN	NaN
937	753298634498793472	NaN	NaN
943	752701944171524096	NaN	NaN

949	752309394570878976	NaN	NaN
1012	747242308580548608	NaN	NaN
1023	746521445350707200	NaN	NaN
1043	743835915802583040	NaN	NaN
1242	711998809858043904	NaN	NaN
2259	667550904950915073	NaN	NaN
2260	667550882905632768	NaN	NaN

	timestamp \
19	2017-07-21 01:02:36 +0000
32	2017-07-15 02:45:48 +0000
36	2017-07-13 01:35:06 +0000
68	2017-06-26 00:13:58 +0000
73	2017-06-24 00:09:53 +0000
74	2017-06-23 18:17:33 +0000
78	2017-06-21 19:36:23 +0000
91	2017-06-13 01:14:41 +0000
95	2017-06-11 00:25:14 +0000
97	2017-06-10 00:35:19 +0000
101	2017-06-08 04:17:07 +0000
109	2017-06-04 00:46:17 +0000
118	2017-05-31 18:47:24 +0000
124	2017-05-28 01:26:04 +0000
130	2017-05-23 17:40:04 +0000
132	2017-05-23 00:41:20 +0000
137	2017-05-21 00:53:21 +0000
146	2017-05-13 19:11:30 +0000
155	2017-05-09 02:29:07 +0000
159	2017-05-06 22:16:42 +0000
160	2017-05-06 18:27:40 +0000
165	2017-05-04 17:01:34 +0000
171	2017-05-01 01:47:28 +0000
180	2017-04-26 02:41:43 +0000
182	2017-04-24 20:17:23 +0000
185	2017-04-24 02:15:55 +0000
194	2017-04-21 02:22:29 +0000
195	2017-04-20 19:16:59 +0000
204	2017-04-14 17:27:40 +0000
211	2017-04-12 00:23:33 +0000
...	...
784	2016-09-11 22:20:06 +0000
794	2016-09-07 01:47:12 +0000
800	2016-09-05 02:00:22 +0000
811	2016-09-01 02:21:21 +0000
815	2016-08-31 15:19:06 +0000
818	2016-08-30 22:04:05 +0000
822	2016-08-29 03:00:36 +0000
826	2016-08-27 00:47:53 +0000

829 2016-08-25 20:35:48 +0000
 833 2016-08-24 21:02:45 +0000
 841 2016-08-20 05:08:29 +0000
 847 2016-08-18 01:03:45 +0000
 860 2016-08-10 00:16:21 +0000
 868 2016-08-06 02:27:27 +0000
 872 2016-08-05 01:19:35 +0000
 885 2016-08-01 16:43:19 +0000
 890 2016-07-31 01:50:18 +0000
 895 2016-07-29 22:53:27 +0000
 908 2016-07-26 00:08:05 +0000
 911 2016-07-25 15:26:30 +0000
 926 2016-07-18 03:06:01 +0000
 937 2016-07-13 18:42:44 +0000
 943 2016-07-12 03:11:42 +0000
 949 2016-07-11 01:11:51 +0000
 1012 2016-06-27 01:37:04 +0000
 1023 2016-06-25 01:52:36 +0000
 1043 2016-06-17 16:01:16 +0000
 1242 2016-03-21 19:31:59 +0000
 2259 2015-11-20 03:51:52 +0000
 2260 2015-11-20 03:51:47 +0000

source \
 19 <a href="http://twitter.com/download/iphone" r...
 32 <a href="http://twitter.com/download/iphone" r...
 36 <a href="http://twitter.com/download/iphone" r...
 68 <a href="http://twitter.com/download/iphone" r...
 73 <a href="http://twitter.com/download/iphone" r...
 74 <a href="http://twitter.com/download/iphone" r...
 78 <a href="http://twitter.com/download/iphone" r...
 91 <a href="http://twitter.com/download/iphone" r...
 95 <a href="http://twitter.com/download/iphone" r...
 97 <a href="http://twitter.com/download/iphone" r...
 101 <a href="http://twitter.com/download/iphone" r...
 109 <a href="http://twitter.com/download/iphone" r...
 118 <a href="http://twitter.com/download/iphone" r...
 124 <a href="http://twitter.com/download/iphone" r...
 130 <a href="http://twitter.com/download/iphone" r...
 132 <a href="http://twitter.com/download/iphone" r...
 137 <a href="http://twitter.com/download/iphone" r...
 146 <a href="http://twitter.com/download/iphone" r...
 155 <a href="http://twitter.com/download/iphone" r...
 159 <a href="http://twitter.com/download/iphone" r...
 160 <a href="http://twitter.com/download/iphone" r...
 165 <a href="http://twitter.com/download/iphone" r...
 171 <a href="http://twitter.com/download/iphone" r...
 180 <a href="http://twitter.com/download/iphone" r...


```

182 <a href="http://twitter.com/download/iphone" r...
185 <a href="http://twitter.com/download/iphone" r...
194 <a href="http://twitter.com/download/iphone" r...
195 <a href="http://twitter.com/download/iphone" r...
204 <a href="http://twitter.com/download/iphone" r...
211 <a href="http://twitter.com/download/iphone" r...
...
784 <a href="http://twitter.com/download/iphone" r...
794 <a href="http://twitter.com/download/iphone" r...
800 <a href="http://twitter.com/download/iphone" r...
811 <a href="http://twitter.com/download/iphone" r...
815 <a href="http://twitter.com/download/iphone" r...
818 <a href="http://twitter.com/download/iphone" r...
822 <a href="http://twitter.com/download/iphone" r...
826 <a href="http://twitter.com/download/iphone" r...
829 <a href="http://twitter.com/download/iphone" r...
833 <a href="http://twitter.com/download/iphone" r...
841 <a href="http://twitter.com/download/iphone" r...
847 <a href="http://twitter.com/download/iphone" r...
860 <a href="http://twitter.com/download/iphone" r...
868 <a href="http://twitter.com/download/iphone" r...
872 <a href="http://twitter.com/download/iphone" r...
885 <a href="http://twitter.com/download/iphone" r...
890 <a href="http://twitter.com/download/iphone" r...
895 <a href="http://twitter.com/download/iphone" r...
908 <a href="http://twitter.com/download/iphone" r...
911 <a href="http://twitter.com/download/iphone" r...
926 <a href="http://twitter.com/download/iphone" r...
937 <a href="http://twitter.com/download/iphone" r...
943 <a href="http://twitter.com/download/iphone" r...
949 <a href="http://twitter.com/download/iphone" r...
1012 <a href="http://twitter.com/download/iphone" r...
1023 <a href="http://twitter.com/download/iphone" r...
1043 <a href="http://twitter.com/download/iphone" r...
1242 <a href="http://twitter.com/download/iphone" r...
2259 <a href="http://twitter.com" rel="nofollow">Tw...
2260 <a href="http://twitter.com" rel="nofollow">Tw...

```

	text	retweeted_status_id \
19	RT @dog_rates: This is Canela. She attempted s...	8.874740e+17
32	RT @Athletics: 12/10 #BATP https://t.co/WxwJmv...	8.860537e+17
36	RT @dog_rates: This is Lilly. She just paralle...	8.305833e+17
68	RT @dog_rates: This is Emmy. She was adopted t...	8.780576e+17
73	RT @dog_rates: Meet Shadow. In an attempt to r...	8.782815e+17
74	RT @dog_rates: Meet Terrance. He's being yelle...	6.690004e+17
78	RT @rachel2195: @dog_rates the boyfriend and h...	8.768508e+17
91	RT @dog_rates: This is Coco. At first I though...	8.663350e+17
95	RT @dog_rates: This is Walter. He won't start ...	8.688804e+17

97	RT @dog_rates: This is Sierra. She's one preci...	8.732138e+17
101	RT @loganamnosis: Penelope here is doing me qu...	8.726576e+17
109	RT @dog_rates: This is Dawn. She's just checki...	8.410770e+17
118	RT @dog_rates: We only rate dogs. This is quit...	8.591970e+17
124	RT @dog_rates: Say hello to Cooper. His expres...	8.685523e+17
130	RT @rachaeleasler: these @dog_rates hats are 1...	8.650134e+17
132	RT @dog_rates: This is Jamesy. He gives a kiss...	8.664507e+17
137	RT @dog_rates: Here's a pupper before and afte...	8.378202e+17
146	RT @dog_rates: Say hello to Quinn. She's quite...	8.630625e+17
155	RT @dog_rates: "Good afternoon class today we'...	8.066291e+17
159	RT @dog_rates: Meet Lorenzo. He's an avid nift...	8.605638e+17
160	RT @tallylott: h*ckin adorable promposal. 13/1...	8.609145e+17
165	RT @dog_rates: Ohboyohboyohboyohboyohboyohboy...	7.616730e+17
171	RT @dog_rates: Meet Winston. He knows he's a l...	8.395493e+17
180	RT @AaronChewning: First time wearing my @dog_...	8.570611e+17
182	RT @dog_rates: This is Luna. It's her first ti...	8.447048e+17
185	RT @Jenna_Marbles: @dog_rates Thanks for ratin...	8.563302e+17
194	RT @dog_rates: Meet George. He looks slightly ...	8.421635e+17
195	RT @frasercampbell_: oh my... what's that... b...	8.551225e+17
204	RT @dog_rates: I usually only share these on F...	8.316501e+17
211	RT @dog_rates: This is Astrid. She's a guide d...	8.293743e+17
...
784	RT @dog_rates: After so many requests, this is...	7.403732e+17
794	RT @dog_rates: Meet Fizz. She thinks love is a...	7.713808e+17
800	RT @dog_rates: This is Gromit. He's pupset bec...	7.652221e+17
811	RT @dog_rates: This is Frankie. He's wearing b...	6.733201e+17
815	RT @katieornah: @dog_rates learning a lot at c...	7.710021e+17
818	RT @dog_rates: Here's a doggo blowing bubbles...	7.392382e+17
822	RT @dog_rates: This is just downright precious...	7.410673e+17
826	RT @dog_rates: Ever seen a dog pet another dog...	7.069045e+17
829	RT @dog_rates: When it's Janet from accounting...	7.001438e+17
833	RT @dog_rates: This is Nollie. She's waving at...	7.399792e+17
841	RT @dog_rates: We only rate dogs... this is a ...	7.599238e+17
847	RT @dog_rates: This is Colby. He's currently r...	7.258423e+17
860	RT @dog_rates: Meet Eve. She's a raging alcoho...	6.732953e+17
868	RT @dog_rates: "Tristan do not speak to me wit...	6.853251e+17
872	RT @dog_rates: Oh. My. God. 13/10 magical af h...	7.116948e+17
885	RT @hownottodraw: The story/person behind @dog...	7.601538e+17
890	RT @dog_rates: This... is a Tyrannosaurus rex...	7.395441e+17
895	RT @dog_rates: AT DAWN...\nWE RIDE\n\n11/10 ht...	6.703191e+17
908	RT @dog_rates: This is Chompsky. He lives up t...	6.790626e+17
911	RT @jon_hill1987: @dog_rates There is a cunning...	7.575971e+17
926	RT @dog_rates: This is Rubio. He has too much ...	6.791584e+17
937	RT @dog_rates: This is Carly. She's actually 2...	6.815232e+17
943	RT @dog_rates: HEY PUP WHAT'S THE PART OF THE ...	6.835159e+17
949	RT @dog_rates: Everyone needs to watch this. 1...	6.753544e+17
1012	RT @dog_rates: This pupper killed this great w...	7.047611e+17
1023	RT @dog_rates: This is Shaggy. He knows exactl...	6.678667e+17

1043	RT @dog_rates: Extremely intelligent dog here...	6.671383e+17
1242	RT @twitter: @dog_rates Awesome Tweet! 12/10. ...	7.119983e+17
2259	RT @dogratingrating: Exceptional talent. Origi...	6.675487e+17
2260	RT @dogratingrating: Unoriginal idea. Blatant ...	6.675484e+17

	retweeted_status_user_id	retweeted_status_timestamp	\
19	4.196984e+09	2017-07-19 00:47:34	+0000
32	1.960740e+07	2017-07-15 02:44:07	+0000
36	4.196984e+09	2017-02-12 01:04:29	+0000
68	4.196984e+09	2017-06-23 01:10:23	+0000
73	4.196984e+09	2017-06-23 16:00:04	+0000
74	4.196984e+09	2015-11-24 03:51:38	+0000
78	5.128045e+08	2017-06-19 17:14:49	+0000
91	4.196984e+09	2017-05-21 16:48:45	+0000
95	4.196984e+09	2017-05-28 17:23:24	+0000
97	4.196984e+09	2017-06-09 16:22:42	+0000
101	1.547674e+08	2017-06-08 03:32:35	+0000
109	4.196984e+09	2017-03-13 00:02:39	+0000
118	4.196984e+09	2017-05-02 00:04:57	+0000
124	4.196984e+09	2017-05-27 19:39:34	+0000
130	7.874618e+17	2017-05-18 01:17:25	+0000
132	4.196984e+09	2017-05-22 00:28:40	+0000
137	4.196984e+09	2017-03-04 00:21:08	+0000
146	4.196984e+09	2017-05-12 16:05:02	+0000
155	4.196984e+09	2016-12-07 22:38:52	+0000
159	4.196984e+09	2017-05-05 18:36:06	+0000
160	3.638908e+08	2017-05-06 17:49:42	+0000
165	4.196984e+09	2016-08-05 21:19:27	+0000
171	4.196984e+09	2017-03-08 18:52:12	+0000
180	5.870972e+07	2017-04-26 02:37:47	+0000
182	4.196984e+09	2017-03-23 00:18:10	+0000
185	6.669901e+07	2017-04-24 02:13:14	+0000
194	4.196984e+09	2017-03-16 00:00:07	+0000
195	7.475543e+17	2017-04-20 18:14:33	+0000
204	4.196984e+09	2017-02-14 23:43:18	+0000
211	4.196984e+09	2017-02-08 17:00:26	+0000
...
784	4.196984e+09	2016-06-08 02:41:38	+0000
794	4.196984e+09	2016-09-01 16:14:48	+0000
800	4.196984e+09	2016-08-15 16:22:20	+0000
811	4.196984e+09	2015-12-06 01:56:44	+0000
815	1.732729e+09	2016-08-31 15:10:07	+0000
818	4.196984e+09	2016-06-04 23:31:25	+0000
822	4.196984e+09	2016-06-10 00:39:48	+0000
826	4.196984e+09	2016-03-07 18:09:06	+0000
829	4.196984e+09	2016-02-18 02:24:13	+0000
833	4.196984e+09	2016-06-07 00:36:02	+0000
841	4.196984e+09	2016-08-01 01:28:46	+0000

847	4.196984e+09	2016-04-29 00:21:01 +0000
860	4.196984e+09	2015-12-06 00:17:55 +0000
868	4.196984e+09	2016-01-08 05:00:14 +0000
872	4.196984e+09	2016-03-20 23:23:54 +0000
885	1.950368e+08	2016-08-01 16:42:51 +0000
890	4.196984e+09	2016-06-05 19:47:03 +0000
895	4.196984e+09	2015-11-27 19:11:49 +0000
908	4.196984e+09	2015-12-21 22:15:18 +0000
911	2.804798e+08	2016-07-25 15:23:28 +0000
926	4.196984e+09	2015-12-22 04:35:49 +0000
937	4.196984e+09	2015-12-28 17:12:42 +0000
943	4.196984e+09	2016-01-03 05:11:12 +0000
949	4.196984e+09	2015-12-11 16:40:19 +0000
1012	4.196984e+09	2016-03-01 20:11:59 +0000
1023	4.196984e+09	2015-11-21 00:46:50 +0000
1043	4.196984e+09	2015-11-19 00:32:12 +0000
1242	7.832140e+05	2016-03-21 19:29:52 +0000
2259	4.296832e+09	2015-11-20 03:43:06 +0000
2260	4.296832e+09	2015-11-20 03:41:59 +0000

	expanded_urls	rating_numerator \
19	https://twitter.com/dog_rates/status/887473957...	13
32	https://twitter.com/dog_rates/status/886053434...	12
36	https://twitter.com/dog_rates/status/830583320...	13
68	https://twitter.com/dog_rates/status/878057613...	14
73	https://www.gofundme.com/3yd6y1c,https://twitt...	13
74	https://twitter.com/dog_rates/status/669000397...	11
78	https://twitter.com/rachel2195/status/87685077...	14
91	https://twitter.com/dog_rates/status/866334964...	12
95	https://twitter.com/dog_rates/status/868880397...	14
97	https://www.gofundme.com/help-my-baby-sierra-g...	12
101	https://twitter.com/loganamnosis/status/872657...	14
109	https://twitter.com/dog_rates/status/841077006...	12
118	https://twitter.com/dog_rates/status/859196978...	12
124	https://www.gofundme.com/3ti3nps,https://twitt...	12
130	https://twitter.com/rachaeleasler/status/86501...	13
132	https://twitter.com/dog_rates/status/866450705...	13
137	https://twitter.com/dog_rates/status/837820167...	12
146	https://www.gofundme.com/helpquinny,https://tw...	13
155	https://twitter.com/dog_rates/status/806629075...	13
159	https://www.gofundme.com/help-lorenzo-beat-can...	13
160	https://twitter.com/tallylott/status/860914485...	13
165	https://twitter.com/dog_rates/status/761672994...	10
171	https://twitter.com/dog_rates/status/839549326...	12
180	https://twitter.com/AaronChewning/status/85706...	13
182	https://twitter.com/dog_rates/status/844704788...	13
185	NaN	14
194	https://twitter.com/dog_rates/status/842163532...	12

195	https://twitter.com/frasercampbell_/status/855...	14
204	http://www.gofundme.com/bluethewhitehusky,http...	13
211	https://twitter.com/dog_rates/status/829374341...	13
...
784	https://twitter.com/dog_rates/status/740373189...	9
794	https://twitter.com/dog_rates/status/771380798...	11
800	https://twitter.com/dog_rates/status/765222098...	10
811	https://twitter.com/dog_rates/status/673320132...	11
815	https://twitter.com/katieornah/status/77100213...	12
818	https://twitter.com/dog_rates/status/739238157...	13
822	https://twitter.com/dog_rates/status/741067306...	12
826	https://vine.co/v/iXQAm5Lrgrh,https://vine.co/...	13
829	https://twitter.com/dog_rates/status/700143752...	10
833	https://twitter.com/dog_rates/status/739979191...	12
841	https://twitter.com/dog_rates/status/759923798...	10
847	https://twitter.com/dog_rates/status/725842289...	12
860	https://twitter.com/dog_rates/status/673295268...	8
868	https://twitter.com/dog_rates/status/685325112...	10
872	https://twitter.com/dog_rates/status/711694788...	13
885	https://weratedogs.com/pages/about-us,https://...	11
890	https://twitter.com/dog_rates/status/739544079...	10
895	https://twitter.com/dog_rates/status/670319130...	11
908	https://twitter.com/dog_rates/status/679062614...	11
911	https://twitter.com/jon_hill1987/status/7575971...	11
926	https://twitter.com/dog_rates/status/679158373...	11
937	https://twitter.com/dog_rates/status/681523177...	12
943	https://vine.co/v/ibvnzrauFuV,https://vine.co/...	11
949	https://twitter.com/dog_rates/status/675354435...	13
1012	https://twitter.com/dog_rates/status/704761120...	13
1023	https://twitter.com/dog_rates/status/667866724...	10
1043	https://twitter.com/dog_rates/status/667138269...	10
1242	https://twitter.com/twitter/status/71199827977...	12
2259	https://twitter.com/dogratingrating/status/667...	12
2260	https://twitter.com/dogratingrating/status/667...	5

	rating_denominator	name	doggo	floofer	pupper	puppo
19	10	Canela	None	None	None	None
32	10	None	None	None	None	None
36	10	Lilly	None	None	None	None
68	10	Emmy	None	None	None	None
73	10	Shadow	None	None	None	None
74	10	Terrance	None	None	None	None
78	10	None	None	None	pupper	None
91	10	Coco	None	None	None	None
95	10	Walter	None	None	None	None
97	10	Sierra	None	None	pupper	None
101	10	None	None	None	None	None
109	10	Dawn	None	None	None	None

118	10	quite	None	None	None	None
124	10	Cooper	None	None	None	None
130	10	None	None	None	None	None
132	10	Jamesy	None	None	pupper	None
137	10	None	None	None	pupper	None
146	10	Quinn	None	None	None	None
155	10	None	None	None	None	None
159	10	Lorenzo	None	None	None	None
160	10	None	None	None	None	None
165	10	None	None	None	None	None
171	10	Winston	None	None	None	None
180	10	None	None	None	None	None
182	10	Luna	None	None	None	None
185	10	None	None	None	None	None
194	10	George	None	None	None	None
195	10	None	None	None	None	None
204	10	None	None	None	None	None
211	10	Astrid	doggo	None	None	None
...
784	11	None	None	None	None	None
794	10	Fizz	None	None	None	None
800	10	Gromit	None	None	None	None
811	10	Frankie	None	None	None	None
815	10	None	None	None	pupper	None
818	10	None	doggo	None	None	None
822	10	just	doggo	None	pupper	None
826	10	None	None	None	None	None
829	10	None	None	None	pupper	None
833	10	Nollie	None	None	None	None
841	10	None	None	None	None	None
847	10	Colby	None	None	None	None
860	10	Eve	None	None	pupper	None
868	10	None	None	None	None	None
872	10	None	None	None	None	None
885	10	None	None	None	None	None
890	10	None	None	None	None	None
895	10	None	None	None	None	None
908	10	Chompsky	None	None	None	None
911	10	None	None	None	pupper	None
926	10	Rubio	None	None	None	None
937	10	Carly	None	None	None	None
943	10	None	None	None	None	None
949	10	None	None	None	None	None
1012	10	None	None	None	pupper	None
1023	10	Shaggy	None	None	None	None
1043	10	None	None	None	None	None
1242	10	None	None	None	None	None
2259	10	None	None	None	None	None

```
2260          10      None      None      None      None      None
```

```
[181 rows x 17 columns]
```

several Retweet data in this dataframe

```
In [15]: twitter_archive_enh.describe()
```

```
Out[15]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
count	2.356000e+03	7.800000e+01	7.800000e+01	
mean	7.427716e+17	7.455079e+17	2.014171e+16	
std	6.856705e+16	7.582492e+16	1.252797e+17	
min	6.660209e+17	6.658147e+17	1.185634e+07	
25%	6.783989e+17	6.757419e+17	3.086374e+08	
50%	7.196279e+17	7.038708e+17	4.196984e+09	
75%	7.993373e+17	8.257804e+17	4.196984e+09	
max	8.924206e+17	8.862664e+17	8.405479e+17	

	retweeted_status_id	retweeted_status_user_id	rating_numerator	\
count	1.810000e+02	1.810000e+02	2356.000000	
mean	7.720400e+17	1.241698e+16	13.126486	
std	6.236928e+16	9.599254e+16	45.876648	
min	6.661041e+17	7.832140e+05	0.000000	
25%	7.186315e+17	4.196984e+09	10.000000	
50%	7.804657e+17	4.196984e+09	11.000000	
75%	8.203146e+17	4.196984e+09	12.000000	
max	8.874740e+17	7.874618e+17	1776.000000	

	rating_denominator
count	2356.000000
mean	10.455433
std	6.745237
min	0.000000
25%	10.000000
50%	10.000000
75%	10.000000
max	170.000000

Strange value in rating_numerator and rating_denominator since number of max is very far from mean data and 75% data

```
In [16]: # adding new column will help the investigation
# calculate normalize rating by dividing rating_numerator and rating/denominator
twitter_archive_enh['normalize_rating'] = twitter_archive_enh['rating_numerator']/twitter_archive_enh['rating_denominator']
```

```
In [20]: twitter_archive_enh[['rating_numerator', 'rating_denominator', 'normalize_rating']].describe()
```

```
Out[20]:
```

	rating_numerator	rating_denominator	normalize_rating
count	2356.000000	2356.000000	2356.000000

mean	13.126486	10.455433	inf
std	45.876648	6.745237	NaN
min	0.000000	0.000000	0.000000
25%	10.000000	10.000000	1.000000
50%	11.000000	10.000000	1.100000
75%	12.000000	10.000000	1.200000
max	1776.000000	170.000000	inf

there are inf value in normalize rating, which is really strange value since median is 1.10. It is probably occure because of 0 value in rating denominator. Deep exploration needed in this value.

```
In [22]: twitter_archive_enh.normalize_rating.sort_values(ascending=False)
```

```
Out[22]: 313      inf
          979    177.600000
          189    66.600000
         2074    42.000000
          188    42.000000
          290    18.200000
          340     7.500000
          695     7.500000
          516     3.428571
          763     2.700000
         1712     2.600000
           55     1.700000
          285     1.500000
          291     1.500000
           64     1.400000
          418     1.400000
          924     1.400000
           68     1.400000
          186     1.400000
          117     1.400000
          123     1.400000
          404     1.400000
          395     1.400000
          187     1.400000
          490     1.400000
           76     1.400000
           56     1.400000
          141     1.400000
          866     1.400000
          426     1.400000
          ...
         2202     0.300000
         2305     0.300000
          1601     0.300000
          1189     0.300000
```



```

1249      0.300000
1692      0.300000
2239      0.300000
2136      0.300000
1314      0.300000
2310      0.200000
1761      0.200000
2349      0.200000
1764      0.200000
1598      0.200000
1165      0.200000
2246      0.200000
1920      0.200000
2326      0.200000
2079      0.200000
2237      0.200000
2091      0.100000
2038      0.100000
605       0.100000
2338      0.100000
1869      0.100000
1940      0.100000
2261      0.100000
1446      0.100000
315       0.000000
1016      0.000000
Name: normalize_rating, Length: 2356, dtype: float64

```

```

In [23]: # see data when in_reply_to_status_id is not null
twitter_archive_enh[twitter_archive_enh.in_reply_to_status_id.notnull()]

```

```

Out[23]:
   tweet_id  in_reply_to_status_id  in_reply_to_user_id \
30  886267009285017600      8.862664e+17      2.281182e+09
55  881633300179243008      8.816070e+17      4.738443e+07
64  879674319642796034      8.795538e+17      3.105441e+09
113 870726314365509632      8.707262e+17      1.648776e+07
148 863427515083354112      8.634256e+17      7.759620e+07
149 863079547188785154      6.671522e+17      4.196984e+09
179 857214891891077121      8.571567e+17      1.806710e+08
184 856526610513747968      8.558181e+17      4.196984e+09
186 856288084350160898      8.562860e+17      2.792810e+08
188 855862651834028034      8.558616e+17      1.943518e+08
189 855860136149123072      8.558585e+17      1.361572e+07
218 850333567704068097      8.503288e+17      2.195506e+07
228 848213670039564288      8.482121e+17      4.196984e+09
234 847617282490613760      8.476062e+17      4.196984e+09
251 844979544864018432      7.590995e+17      4.196984e+09
274 840698636975636481      8.406983e+17      8.405479e+17

```

290	838150277551247360	8.381455e+17	2.195506e+07
291	838085839343206401	8.380855e+17	2.894131e+09
313	835246439529840640	8.352460e+17	2.625958e+07
342	832088576586297345	8.320875e+17	3.058208e+07
346	831926988323639298	8.319030e+17	2.068372e+07
387	826598799820865537	8.265984e+17	4.196984e+09
409	823333489516937216	8.233264e+17	1.582854e+09
427	821153421864615936	8.211526e+17	1.132119e+08
498	813130366689148928	8.131273e+17	4.196984e+09
513	811647686436880384	8.116272e+17	4.196984e+09
565	802265048156610565	7.331095e+17	4.196984e+09
570	801854953262350336	8.018543e+17	1.185634e+07
576	800859414831898624	8.008580e+17	2.918590e+08
611	797165961484890113	7.971238e+17	2.916630e+07
...
1479	693582294167244802	6.935722e+17	1.198989e+09
1497	692423280028966913	6.924173e+17	4.196984e+09
1501	692142790915014657	6.920419e+17	4.196984e+09
1523	690607260360429569	6.903413e+17	4.670367e+08
1598	686035780142297088	6.860340e+17	4.196984e+09
1605	685681090388975616	6.855479e+17	4.196984e+09
1618	684969860808454144	6.849598e+17	4.196984e+09
1630	684538444857667585	6.844811e+17	4.196984e+09
1634	684225744407494656	6.842229e+17	4.196984e+09
1663	682808988178739200	6.827884e+17	4.196984e+09
1689	681340665377193984	6.813394e+17	4.196984e+09
1774	678023323247357953	6.780211e+17	4.196984e+09
1819	676590572941893632	6.765883e+17	4.196984e+09
1842	675870721063669760	6.757073e+17	4.196984e+09
1844	675849018447167488	6.758457e+17	4.196984e+09
1852	675707330206547968	6.754971e+17	4.196984e+09
1866	675349384339542016	6.749998e+17	4.196984e+09
1882	674999807681908736	6.747934e+17	4.196984e+09
1885	674793399141146624	6.717299e+17	4.196984e+09
1892	674754018082705410	6.747522e+17	4.196984e+09
1895	674742531037511680	6.747400e+17	4.196984e+09
1905	674606911342424069	6.744689e+17	4.196984e+09
1914	674330906434379776	6.658147e+17	1.637468e+07
1940	673716320723169284	6.737159e+17	4.196984e+09
2036	671729906628341761	6.715610e+17	4.196984e+09
2038	671550332464455680	6.715449e+17	4.196984e+09
2149	669684865554620416	6.693544e+17	4.196984e+09
2169	669353438988365824	6.678065e+17	4.196984e+09
2189	668967877119254528	6.689207e+17	2.143566e+07
2298	667070482143944705	6.670655e+17	4.196984e+09

timestamp \

30 2017-07-15 16:51:35 +0000

55	2017-07-02	21:58:53	+0000
64	2017-06-27	12:14:36	+0000
113	2017-06-02	19:38:25	+0000
148	2017-05-13	16:15:35	+0000
149	2017-05-12	17:12:53	+0000
179	2017-04-26	12:48:51	+0000
184	2017-04-24	15:13:52	+0000
186	2017-04-23	23:26:03	+0000
188	2017-04-22	19:15:32	+0000
189	2017-04-22	19:05:32	+0000
218	2017-04-07	13:04:55	+0000
228	2017-04-01	16:41:12	+0000
234	2017-03-31	01:11:22	+0000
251	2017-03-23	18:29:57	+0000
274	2017-03-11	22:59:09	+0000
290	2017-03-04	22:12:52	+0000
291	2017-03-04	17:56:49	+0000
313	2017-02-24	21:54:03	+0000
342	2017-02-16	04:45:50	+0000
346	2017-02-15	18:03:45	+0000
387	2017-02-01	01:11:25	+0000
409	2017-01-23	00:56:15	+0000
427	2017-01-17	00:33:26	+0000
498	2016-12-25	21:12:41	+0000
513	2016-12-21	19:01:02	+0000
565	2016-11-25	21:37:47	+0000
570	2016-11-24	18:28:13	+0000
576	2016-11-22	00:32:18	+0000
611	2016-11-11	19:55:50	+0000
...			...
1479	2016-01-30	23:51:19	+0000
1497	2016-01-27	19:05:49	+0000
1501	2016-01-27	00:31:15	+0000
1523	2016-01-22	18:49:36	+0000
1598	2016-01-10	04:04:10	+0000
1605	2016-01-09	04:34:45	+0000
1618	2016-01-07	05:28:35	+0000
1630	2016-01-06	00:54:18	+0000
1634	2016-01-05	04:11:44	+0000
1663	2016-01-01	06:22:03	+0000
1689	2015-12-28	05:07:27	+0000
1774	2015-12-19	01:25:31	+0000
1819	2015-12-15	02:32:17	+0000
1842	2015-12-13	02:51:51	+0000
1844	2015-12-13	01:25:37	+0000
1852	2015-12-12	16:02:36	+0000
1866	2015-12-11	16:20:15	+0000
1882	2015-12-10	17:11:09	+0000

1885 2015-12-10 03:30:58 +0000
 1892 2015-12-10 00:54:28 +0000
 1895 2015-12-10 00:08:50 +0000
 1905 2015-12-09 15:09:55 +0000
 1914 2015-12-08 20:53:11 +0000
 1940 2015-12-07 04:11:02 +0000
 2036 2015-12-01 16:37:44 +0000
 2038 2015-12-01 04:44:10 +0000
 2149 2015-11-26 01:11:28 +0000
 2169 2015-11-25 03:14:30 +0000
 2189 2015-11-24 01:42:25 +0000
 2298 2015-11-18 20:02:51 +0000

source \

30 <a href="http://twitter.com/download/iphone" r...
 55 <a href="http://twitter.com/download/iphone" r...
 64 <a href="http://twitter.com/download/iphone" r...
 113 <a href="http://twitter.com/download/iphone" r...
 148 <a href="http://twitter.com/download/iphone" r...
 149 <a href="http://twitter.com/download/iphone" r...
 179 <a href="http://twitter.com/download/iphone" r...
 184 <a href="http://twitter.com/download/iphone" r...
 186 <a href="http://twitter.com/download/iphone" r...
 188 <a href="http://twitter.com/download/iphone" r...
 189 <a href="http://twitter.com/download/iphone" r...
 218 <a href="http://twitter.com/download/iphone" r...
 228 <a href="http://twitter.com/download/iphone" r...
 234 <a href="http://twitter.com/download/iphone" r...
 251 <a href="http://twitter.com/download/iphone" r...
 274 <a href="http://twitter.com/download/iphone" r...
 290 <a href="http://twitter.com/download/iphone" r...
 291 <a href="http://twitter.com/download/iphone" r...
 313 <a href="http://twitter.com/download/iphone" r...
 342 <a href="http://twitter.com/download/iphone" r...
 346 <a href="http://twitter.com/download/iphone" r...
 387 <a href="http://twitter.com/download/iphone" r...
 409 <a href="http://twitter.com/download/iphone" r...
 427 <a href="http://twitter.com/download/iphone" r...
 498 <a href="http://twitter.com/download/iphone" r...
 513 <a href="http://twitter.com/download/iphone" r...
 565 <a href="http://twitter.com/download/iphone" r...
 570 <a href="http://twitter.com/download/iphone" r...
 576 <a href="http://twitter.com/download/iphone" r...
 611 <a href="http://twitter.com/download/iphone" r...

 1479 <a href="http://twitter.com/download/iphone" r...
 1497 <a href="http://twitter.com/download/iphone" r...
 1501 <a href="http://twitter.com/download/iphone" r...

1523 <a href="http://twitter.com/download/iphone" r...
 1598 <a href="http://twitter.com/download/iphone" r...
 1605 <a href="http://twitter.com/download/iphone" r...
 1618 <a href="http://twitter.com/download/iphone" r...
 1630 <a href="http://twitter.com/download/iphone" r...
 1634 <a href="http://twitter.com/download/iphone" r...
 1663 <a href="http://twitter.com/download/iphone" r...
 1689 <a href="http://twitter.com/download/iphone" r...
 1774 <a href="http://twitter.com/download/iphone" r...
 1819 <a href="http://twitter.com/download/iphone" r...
 1842 <a href="http://twitter.com/download/iphone" r...
 1844 <a href="http://twitter.com/download/iphone" r...
 1852 <a href="http://twitter.com/download/iphone" r...
 1866 <a href="http://twitter.com/download/iphone" r...
 1882 <a href="http://twitter.com/download/iphone" r...
 1885 <a href="http://twitter.com/download/iphone" r...
 1892 <a href="http://twitter.com/download/iphone" r...
 1895 <a href="http://twitter.com/download/iphone" r...
 1905 <a href="http://twitter.com/download/iphone" r...
 1914 <a href="http://twitter.com/download/iphone" r...
 1940 <a href="http://twitter.com/download/iphone" r...
 2036 <a href="http://twitter.com/download/iphone" r...
 2038 <a href="http://twitter.com/download/iphone" r...
 2149 <a href="http://twitter.com/download/iphone" r...
 2169 <a href="http://twitter.com/download/iphone" r...
 2189 <a href="http://twitter.com/download/iphone" r...
 2298 <a href="http://twitter.com/download/iphone" r...

	text	retweeted_status_id \
30	@NonWhiteHat @MayhewMayhem omg hello tanner yo...	NaN
55	@roushfenway These are good dogs but 17/10 is ...	NaN
64	@RealKentMurphy 14/10 confirmed	NaN
113	@ComplicitOwl @ShopWeRateDogs >10/10 is res...	NaN
148	@Jack_Septic_Eye I'd need a few more pics to p...	NaN
149	Ladies and gentlemen... I found Pipsy. He may ...	NaN
179	@Marc_IRL pixelated af 12/10	NaN
184	THIS IS CHARLIE, MARK. HE DID JUST WANT TO SAY...	NaN
186	@xianmcguire @Jenna_Marbles Kardashians wouldn...	NaN
188	@dhmontgomery We also gave snoop dogg a 420/10...	NaN
189	@s8n You tried very hard to portray this good ...	NaN
218	@markhoppus MARK THAT DOG HAS SEEN AND EXPERIE...	NaN
228	Jerry just apuppologized to me. He said there ...	NaN
234	.@breaannanicolee PUPDATE: Cannon has a heart ...	NaN
251	PUPDATE: I'm proud to announce that Toby is 23...	NaN
274	@O_kelvin_0 >10/10 is reserved for puppos s...	NaN
290	@markhoppus 182/10	NaN
291	@bragg6of8 @Andy_Pace_ we are still looking fo...	NaN
313	@jonnyisun @Lin_Manuel ok jomny I know you're e...	NaN

342	@docmisterio account started on 11/15/15	NaN
346	@UNC can confirm 12/10	NaN
387	I was going to do 007/10, but the joke wasn't ...	NaN
409	@HistoryInPics 13/10	NaN
427	@imgur for a polar bear tho I'd say 13/10 is a...	NaN
498	I've been informed by multiple sources that th...	NaN
513	PUPDATE: I've been informed that Augie was act...	NaN
565	Like doggo, like pupper version 2. Both 11/10 ...	NaN
570	.@NBCSports OMG THE TINY HAT I'M GOING TO HAVE...	NaN
576	@SkyWilliams doggo simply protecting you from ...	NaN
611	@JODYHiGHROLLER it may be an 11/10 but what do...	NaN
...
1479	Personally I'd give him an 11/10. Not sure why...	NaN
1497	PUPDATE: just noticed this dog has some extra ...	NaN
1501	These are some pictures of Teddy that further ...	NaN
1523	12/10 @LightningHoltt	NaN
1598	Yes I do realize a rating of 4/20 would've bee...	NaN
1605	Jack deserves another round of applause. If yo...	NaN
1618	For those who claim this is a goat, u are wron...	NaN
1630	After watching this video, we've determined th...	NaN
1634	Two sneaky puppers were not initially seen, mo...	NaN
1663	I'm aware that I could've said 20/16, but here...	NaN
1689	I've been told there's a slight possibility he...	NaN
1774	After getting lost in Reese's eyes for several...	NaN
1819	After some outrage from the crowd. Bubbles is ...	NaN
1842	& this is Yoshi. Another world record cont...	NaN
1844	This dog is being demoted to a 9/10 for not we...	NaN
1852	We've got ourselves a battle here. Watch out R...	NaN
1866	Yea I lied. Here's more. All 13/10 https://t.c...	NaN
1882	Ok last one of these. I may try to make some m...	NaN
1885	I have found another. 13/10 https://t.co/HwroP...	NaN
1892	Just received another perfect photo of dogs an...	NaN
1895	Some clarification is required. The dog is sin...	NaN
1905	The 13/10 also takes into account this impecca...	NaN
1914	13/10\n@ABC7	NaN
1940	The millennials have spoken and we've decided ...	NaN
2036	I'm just going to leave this one here as well...	NaN
2038	After 22 minutes of careful deliberation this ...	NaN
2149	After countless hours of research and hundreds...	NaN
2169	This is Tessa. She is also very pleased after ...	NaN
2189	12/10 good shit Bubka\n@wane15	NaN
2298	After much debate this dog is being upgraded t...	NaN
	retweeted_status_user_id retweeted_status_timestamp \	
30	NaN	NaN
55	NaN	NaN
64	NaN	NaN
113	NaN	NaN

148	NaN	NaN
149	NaN	NaN
179	NaN	NaN
184	NaN	NaN
186	NaN	NaN
188	NaN	NaN
189	NaN	NaN
218	NaN	NaN
228	NaN	NaN
234	NaN	NaN
251	NaN	NaN
274	NaN	NaN
290	NaN	NaN
291	NaN	NaN
313	NaN	NaN
342	NaN	NaN
346	NaN	NaN
387	NaN	NaN
409	NaN	NaN
427	NaN	NaN
498	NaN	NaN
513	NaN	NaN
565	NaN	NaN
570	NaN	NaN
576	NaN	NaN
611	NaN	NaN
...
1479	NaN	NaN
1497	NaN	NaN
1501	NaN	NaN
1523	NaN	NaN
1598	NaN	NaN
1605	NaN	NaN
1618	NaN	NaN
1630	NaN	NaN
1634	NaN	NaN
1663	NaN	NaN
1689	NaN	NaN
1774	NaN	NaN
1819	NaN	NaN
1842	NaN	NaN
1844	NaN	NaN
1852	NaN	NaN
1866	NaN	NaN
1882	NaN	NaN
1885	NaN	NaN
1892	NaN	NaN
1895	NaN	NaN

1905	NaN	NaN
1914	NaN	NaN
1940	NaN	NaN
2036	NaN	NaN
2038	NaN	NaN
2149	NaN	NaN
2169	NaN	NaN
2189	NaN	NaN
2298	NaN	NaN

	expanded_urls	rating_numerator \
30	NaN	12
55	NaN	17
64	NaN	14
113	NaN	10
148	NaN	12
149	https://twitter.com/dog_rates/status/863079547...	14
179	NaN	12
184	https://twitter.com/dog_rates/status/856526610...	14
186	NaN	14
188	NaN	420
189	NaN	666
218	NaN	13
228	NaN	11
234	NaN	13
251	https://twitter.com/dog_rates/status/844979544...	13
274	NaN	10
290	NaN	182
291	NaN	15
313	NaN	960
342	NaN	11
346	NaN	12
387	NaN	7
409	NaN	13
427	NaN	13
498	NaN	12
513	NaN	11
565	https://twitter.com/dog_rates/status/802265048...	11
570	NaN	11
576	NaN	11
611	NaN	11
...
1479	NaN	11
1497	NaN	9
1501	https://twitter.com/dog_rates/status/692142790...	13
1523	NaN	12
1598	NaN	4
1605	NaN	14

1618	NaN	5
1630	https://twitter.com/dog_rates/status/684538444...	12
1634	https://twitter.com/dog_rates/status/684225744...	143
1663	NaN	20
1689	NaN	5
1774	NaN	13
1819	NaN	7
1842	https://twitter.com/dog_rates/status/675870721...	11
1844	NaN	9
1852	https://twitter.com/dog_rates/status/675707330...	11
1866	https://twitter.com/dog_rates/status/675349384...	13
1882	https://twitter.com/dog_rates/status/674999807...	13
1885	https://twitter.com/dog_rates/status/674793399...	13
1892	https://twitter.com/dog_rates/status/674754018...	12
1895	NaN	11
1905	NaN	13
1914	NaN	13
1940	NaN	1
2036	https://twitter.com/dog_rates/status/671729906...	13
2038	NaN	1
2149	NaN	11
2169	https://twitter.com/dog_rates/status/669353438...	10
2189	NaN	12
2298	NaN	10

	rating_denominator	name	doggo	floofer	pupper	puppo
30	10	None	None	None	None	None
55	10	None	None	None	None	None
64	10	None	None	None	None	None
113	10	None	None	None	None	None
148	10	None	None	None	None	None
149	10	None	None	None	None	None
179	10	None	None	None	None	None
184	10	None	None	None	None	None
186	10	None	None	None	None	None
188	10	None	None	None	None	None
189	10	None	None	None	None	None
218	10	None	None	None	None	None
228	10	None	None	None	None	None
234	10	None	None	None	None	None
251	10	None	None	None	None	None
274	10	None	None	None	None	None
290	10	None	None	None	None	None
291	10	None	None	None	None	None
313	0	None	None	None	None	None
342	15	None	None	None	None	None
346	10	None	None	None	None	None
387	10	None	None	None	None	None

409	10	None	None	None	None	None
427	10	None	None	None	None	None
498	10	None	None	None	None	None
513	10	None	None	None	None	None
565	10	None	doggo	None	pupper	None
570	10	None	None	None	None	None
576	10	None	doggo	None	None	None
611	10	None	None	None	None	None
...
1479	10	None	None	None	None	None
1497	10	None	None	None	None	None
1501	10	None	None	None	None	None
1523	10	None	None	None	None	None
1598	20	None	None	None	None	None
1605	10	None	None	None	None	None
1618	10	None	None	None	None	None
1630	10	None	None	None	None	None
1634	130	None	None	None	None	None
1663	16	None	None	None	None	None
1689	10	None	None	None	None	None
1774	10	None	None	None	None	None
1819	10	None	None	None	None	None
1842	10	None	None	None	None	None
1844	10	None	None	None	None	None
1852	10	None	None	None	None	None
1866	10	None	None	None	None	None
1882	10	None	None	None	None	None
1885	10	None	None	None	None	None
1892	10	None	None	None	None	None
1895	10	None	None	None	None	None
1905	10	None	None	None	None	None
1914	10	None	None	None	None	None
1940	10	None	None	None	None	None
2036	10	None	None	None	None	None
2038	10	None	None	None	None	None
2149	10	None	None	None	None	None
2169	10	Tessa	None	None	None	None
2189	10	None	None	None	None	None
2298	10	None	None	None	None	None

[78 rows x 17 columns]

In [25]: twitter_archive_enh.query('normalize_rating==inf')

```
Out[25]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
313	835246439529840640	8.352460e+17	26259576.0	
				timestamp \

```

313 2017-02-24 21:54:03 +0000

source \
313 <a href="http://twitter.com/download/iphone" r...

text retweeted_status_id \
313 @jonnyusun @Lin_Manuel ok jomny I know you're e... NaN

retweeted_status_user_id retweeted_status_timestamp expanded_urls \
313 NaN NaN NaN

rating_numerator rating_denominator name doggo floofer pupper puppo \
313 960 0 None None None None None

normalize_rating
313 inf

```

rating_numerator is very big but rating_denominator is 0. let's check the value in text.

```

In [29]: # get index of infinite number
twitter_archive_enh.query('normalize_rating==inf').index

```

```

Out[29]: Int64Index([313], dtype='int64')

```

```

In [27]: list(twitter_archive_enh.query('normalize_rating==inf').text)

```

```

Out[27]: ["@jonnyusun @Lin_Manuel ok jomny I know you're excited but 960/00 isn't a valid rating,

```

in fact, value stored in rating_numerator and rating_denominator is wrong. It is 13/10 instead of 960/0. So rating_numerator and rating denominator in row 313 need to be fixed into 13/10

```

In [ ]: twitter_archive_enh['text'][263]

```

```

In [26]: twitter_archive_enh['text'][571]

```

```

Out[26]: "This is Wallace. He'll be your chau-fur this evening. 12/10 eyes on the road Wallace h

```

text column including content and link

0.2.2 image_prediction

```

In [30]: image_predictions.head(20)

```

```

Out[30]:
   tweet_id  jpg_url \
0  666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.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
5  666050758794694657  https://pbs.twimg.com/media/CT5Jof1WUAEvVxN.jpg

```

6	666051853826850816	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7	666055525042405380	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8	666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9	666058600524156928	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10	666063827256086533	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11	666071193221509120	https://pbs.twimg.com/media/CT5cN_3WEAA1OoZ.jpg
12	666073100786774016	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13	666082916733198337	https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg
14	666094000022159362	https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg
15	666099513787052032	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16	666102155909144576	https://pbs.twimg.com/media/CT54YGiwUAZnoK.jpg
17	666104133288665088	https://pbs.twimg.com/media/CT56LSZWAA1Jj2.jpg
18	666268910803644416	https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
19	666273097616637952	https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.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
5	1	Bernese_mountain_dog	0.651137	True	English_springer
6	1	box_turtle	0.933012	False	mud_turtle
7	1	chow	0.692517	True	Tibetan_mastiff
8	1	shopping_cart	0.962465	False	shopping_basket
9	1	miniature_poodle	0.201493	True	komondor
10	1	golden_retriever	0.775930	True	Tibetan_mastiff
11	1	Gordon_setter	0.503672	True	Yorkshire_terrier
12	1	Walker_hound	0.260857	True	English_foxhound
13	1	pug	0.489814	True	bull_mastiff
14	1	bloodhound	0.195217	True	German_shepherd
15	1	Lhasa	0.582330	True	Shih-Tzu
16	1	English_setter	0.298617	True	Newfoundland
17	1	hen	0.965932	False	cock
18	1	desktop_computer	0.086502	False	desk
19	1	Italian_greyhound	0.176053	True	toy_terrier

	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
5	0.263788	True	Greater_Swiss_Mountain_dog	0.016199	True
6	0.045885	False	terrapin	0.017885	False
7	0.058279	True	fur_coat	0.054449	False
8	0.014594	False	golden_retriever	0.007959	True
9	0.192305	True	soft-coated_wheaten_terrier	0.082086	True

10	0.093718	True	Labrador_retriever	0.072427	True
11	0.174201	True	Pekinese	0.109454	True
12	0.175382	True	Ibizan_hound	0.097471	True
13	0.404722	True	French_bulldog	0.048960	True
14	0.078260	True	malinois	0.075628	True
15	0.166192	True	Dandie_Dinmont	0.089688	True
16	0.149842	True	borzoi	0.133649	True
17	0.033919	False	partridge	0.000052	False
18	0.085547	False	bookcase	0.079480	False
19	0.111884	True	basenji	0.111152	True

In [31]: 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 [32]: image_predictions.describe()

```
Out[32]:
```

	tweet_id	img_num	p1_conf	p2_conf	p3_conf
count	2.075000e+03	2075.000000	2075.000000	2.075000e+03	2.075000e+03
mean	7.384514e+17	1.203855	0.594548	1.345886e-01	6.032417e-02
std	6.785203e+16	0.561875	0.271174	1.006657e-01	5.090593e-02
min	6.660209e+17	1.000000	0.044333	1.011300e-08	1.740170e-10
25%	6.764835e+17	1.000000	0.364412	5.388625e-02	1.622240e-02
50%	7.119988e+17	1.000000	0.588230	1.181810e-01	4.944380e-02
75%	7.932034e+17	1.000000	0.843855	1.955655e-01	9.180755e-02
max	8.924206e+17	4.000000	1.000000	4.880140e-01	2.734190e-01

In [33]: image_predictions.jpg_url.value_counts()

```
Out[33]: https://pbs.twimg.com/media/C2oRbOuWEAAbVS1.jpg
https://pbs.twimg.com/media/CvJCabcWgAIoUxW.jpg
https://pbs.twimg.com/media/CUN4Or5UAAAAa5K4.jpg
```

https://pbs.twimg.com/media/Cs_DYr1XEAA54Pu.jpg
<https://pbs.twimg.com/media/CiibOMzUYAA9Mxz.jpg>
<https://pbs.twimg.com/media/CVM0lMiWwAA4Yxl.jpg>
<https://pbs.twimg.com/media/CYLDikFWEAAIy1y.jpg>
<https://pbs.twimg.com/media/CmoPdmHW8AAi8BI.jpg>
<https://pbs.twimg.com/media/Ct2q05PXEA6eB0.jpg>
<https://pbs.twimg.com/media/CtKHLuCWYAA2TTs.jpg>
https://pbs.twimg.com/tweet_video_thumb/CeBym7oXEAEWbEg.jpg
<https://pbs.twimg.com/media/CtzKC7zXEAALfSo.jpg>
<https://pbs.twimg.com/media/CcG07BYW0AErrC9.jpg>
<https://pbs.twimg.com/media/CW88XN4WsAAlo8r.jpg>
<https://pbs.twimg.com/media/CsV07ljW8AAckRD.jpg>
<https://pbs.twimg.com/media/CVuQ2LeUsAAIe3s.jpg>
https://pbs.twimg.com/ext_tw_video_thumb/807106774843039744/pu/img/8XZg1xW35Xp2J6JW.jpg
<https://pbs.twimg.com/media/Ct72q9jWcAAhlnw.jpg>
<https://pbs.twimg.com/media/CWyD2HGUYAQ1Xa7.jpg>
<https://pbs.twimg.com/media/CzG425nWgAAAnP7P.jpg>
<https://pbs.twimg.com/media/Cx5R8wPVEAALa9r.jpg>
<https://pbs.twimg.com/media/CxqsX-8XUAAEvjD.jpg>
https://pbs.twimg.com/media/Crwx5yWgAAX5P_.jpg
<https://pbs.twimg.com/media/CU3mITUWIAAfyQS.jpg>
https://pbs.twimg.com/ext_tw_video_thumb/675354114423808004/pu/img/qL1R_nGLqa6lmk0x.jpg
<https://pbs.twimg.com/media/CtVAvX-WIAAcGTf.jpg>
<https://pbs.twimg.com/media/CrXhIqBW8AA6Bse.jpg>
https://pbs.twimg.com/ext_tw_video_thumb/817423809049493505/pu/img/50FW0yueFu9oTUiQ.jpg
<https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg>
<https://pbs.twimg.com/media/CwiuEJmW8AAZnit.jpg>

<https://pbs.twimg.com/media/CeaLlAPUMAIcC7U.jpg>
<https://pbs.twimg.com/media/CWPUB9TWwAALPPx.jpg>
https://pbs.twimg.com/media/CYZvRttWYAE_RXc.jpg
<https://pbs.twimg.com/media/CyJtSmDUAAA2F9x.jpg>
https://pbs.twimg.com/media/C_KVJjDXsAEUCWn.jpg
<https://pbs.twimg.com/media/CW-loUBWYAAAn2Cb.jpg>
<https://pbs.twimg.com/media/CWcrAVQWEAA6QMp.jpg>
<https://pbs.twimg.com/media/CUbrDWOWcAEyMdm.jpg>
<https://pbs.twimg.com/media/CUtw9SAVEAAAtFUN.jpg>
<https://pbs.twimg.com/media/CySpCSHXcAAN-qC.jpg>
<https://pbs.twimg.com/media/COuXObSXUAAIzmV.jpg>
<https://pbs.twimg.com/media/C1IT6rVXUAIvwYT.jpg>
<https://pbs.twimg.com/media/CVkJgsxU8AA50YX.jpg>
<https://pbs.twimg.com/media/CU3NE8EWUAEVdPD.jpg>
<https://pbs.twimg.com/media/CUroc7QW4AATIff.jpg>
https://pbs.twimg.com/ext_tw_video_thumb/688211379870806016/pu/img/0IOVTwPYnsScQg8R.jpg
<https://pbs.twimg.com/media/C54DS1kXQAEU5pS.jpg>
<https://pbs.twimg.com/media/CZTLeBuWIAAFkeR.jpg>
<https://pbs.twimg.com/media/CwHj-jGWAAAnsny.jpg>
<https://pbs.twimg.com/media/CUnrN7vUcAAfGvN.jpg>

```

https://pbs.twimg.com/media/CVt49k_WsAAAtNYC.jpg
https://pbs.twimg.com/media/CbYmRHyWEAASNzm.jpg
https://pbs.twimg.com/media/CW4UtmYWsAAEjqA.jpg
https://pbs.twimg.com/media/C9kq_bbVwAAuRZd.jpg
https://pbs.twimg.com/media/C9QEqZ7XYAIR7fS.jpg
https://pbs.twimg.com/media/CZiO7mWUEAAa4zo.jpg
https://pbs.twimg.com/media/CVhEoq4WcAE8pBm.jpg
https://pbs.twimg.com/media/CaTNMUGUYAAB6vs.jpg
https://pbs.twimg.com/media/C52pYJXWgAA2BEf.jpg
https://pbs.twimg.com/media/CU7SW39WwAAL8Rw.jpg
Name: jpg_url, Length: 2009, dtype: int64

```

several jpg_url has duplication, it needs further inspection to make sure there are a duplication of image or not

```
In [34]: image_predictions[image_predictions['jpg_url']=='https://pbs.twimg.com/media/CsGnz64WYA
```

```

Out[34]:
      tweet_id      jpg_url \
1441  775085132600442880  https://pbs.twimg.com/media/CsGnz64WYAEIDHJ.jpg
1601  799774291445383169  https://pbs.twimg.com/media/CsGnz64WYAEIDHJ.jpg

      img_num  p1  p1_conf  p1_dog      p2  p2_conf  p2_dog \
1441         1  chow  0.316565   True  golden_retriever  0.241929   True
1601         1  chow  0.316565   True  golden_retriever  0.241929   True

      p3  p3_conf  p3_dog
1441  Pomeranian  0.157524   True
1601  Pomeranian  0.157524   True

```

there are duplication data with different tweet_id

```
In [35]: dup = image_predictions[image_predictions.duplicated(subset=['jpg_url', 'img_num'], keep=
dup.sort_values('jpg_url')
```

```

Out[35]:
      tweet_id      jpg_url \
480  675354435921575936  https://pbs.twimg.com/ext_tw_video_thumb/67535...
1297  752309394570878976  https://pbs.twimg.com/ext_tw_video_thumb/67535...
1864  842892208864923648  https://pbs.twimg.com/ext_tw_video_thumb/80710...
1641  807106840509214720  https://pbs.twimg.com/ext_tw_video_thumb/80710...
1703  817181837579653120  https://pbs.twimg.com/ext_tw_video_thumb/81596...
1691  815966073409433600  https://pbs.twimg.com/ext_tw_video_thumb/81596...
1705  817423860136083457  https://pbs.twimg.com/ext_tw_video_thumb/81742...
1858  841833993020538882  https://pbs.twimg.com/ext_tw_video_thumb/81742...
1715  819004803107983360  https://pbs.twimg.com/media/C12whDoVEAALRxa.jpg
1718  819015337530290176  https://pbs.twimg.com/media/C12whDoVEAALRxa.jpg
1716  819006400881917954  https://pbs.twimg.com/media/C12x-JTVIAAzdf1.jpg
1717  819015331746349057  https://pbs.twimg.com/media/C12x-JTVIAAzdf1.jpg
1746  823269594223824897  https://pbs.twimg.com/media/C2kzTGxWEAE0pPL.jpg
1738  822244816520155136  https://pbs.twimg.com/media/C2kzTGxWEAE0pPL.jpg

```

1740	822489057087389700	https://pbs.twimg.com/media/C2oRb0uWEAAbVS1.jpg
1742	822647212903690241	https://pbs.twimg.com/media/C2oRb0uWEAAbVS1.jpg
1767	826958653328592898	https://pbs.twimg.com/media/C3nygbBWQAAjwcW.jpg
1789	829878982036299777	https://pbs.twimg.com/media/C3nygbBWQAAjwcW.jpg
1785	829374341691346946	https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg
1903	851953902622658560	https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg
2041	885311592912609280	https://pbs.twimg.com/media/C4bTH6nWMAAX_bJ.jpg
1791	830583320585068544	https://pbs.twimg.com/media/C4bTH6nWMAAX_bJ.jpg
224	670319130621435904	https://pbs.twimg.com/media/CU1zsMSUAAAS0qW.jpg
1345	759159934323924993	https://pbs.twimg.com/media/CU1zsMSUAAAS0qW.jpg
1590	798644042770751489	https://pbs.twimg.com/media/CU3mITUWIAAfyQS.jpg
241	670444955656130560	https://pbs.twimg.com/media/CU3mITUWIAAfyQS.jpg
1589	798628517273620480	https://pbs.twimg.com/media/CUN40r5UAAAa5K4.jpg
85	667509364010450944	https://pbs.twimg.com/media/CUN40r5UAAAa5K4.jpg
327	671896809300709376	https://pbs.twimg.com/media/CVM01MiWwAA4Yxl.jpg
1591	798665375516884993	https://pbs.twimg.com/media/CVM01MiWwAA4Yxl.jpg
...
1571	794983741416415232	https://pbs.twimg.com/media/CvT6IV6WEAQhhV5.jpg
1530	789530877013393408	https://pbs.twimg.com/media/CvT6IV6WEAQhhV5.jpg
1538	790723298204217344	https://pbs.twimg.com/media/CvaYgDOWgAEfjls.jpg
1533	789986466051088384	https://pbs.twimg.com/media/CvaYgDOWgAEfjls.jpg
1683	813944609378369540	https://pbs.twimg.com/media/Cveg1-NXgAASaaT.jpg
1534	790277117346975746	https://pbs.twimg.com/media/Cveg1-NXgAASaaT.jpg
1539	790946055508652032	https://pbs.twimg.com/media/CvoBPWRWgAA4het.jpg
1699	816829038950027264	https://pbs.twimg.com/media/CvoBPWRWgAA4het.jpg
1544	791672322847637504	https://pbs.twimg.com/media/CvyVxQRWEAAaSZS.jpg
1564	793614319594401792	https://pbs.twimg.com/media/CvyVxQRWEAAaSZS.jpg
1804	832215726631055365	https://pbs.twimg.com/media/CwJR1okWIAA6XMP.jpg
1561	793286476301799424	https://pbs.twimg.com/media/CwJR1okWIAA6XMP.jpg
1566	793962221541933056	https://pbs.twimg.com/media/CwS4aqZXUAAe3I0.jpg
1652	809808892968534016	https://pbs.twimg.com/media/CwS4aqZXUAAe3I0.jpg
1572	795076730285391872	https://pbs.twimg.com/media/CwiuEJmW8AAZnit.jpg
1755	824796380199809024	https://pbs.twimg.com/media/CwiuEJmW8AAZnit.jpg
1579	796177847564038144	https://pbs.twimg.com/media/Cwx99rpW8AMk_Ie.jpg
1578	796149749086875649	https://pbs.twimg.com/media/Cwx99rpW8AMk_Ie.jpg
1610	801167903437357056	https://pbs.twimg.com/media/Cx5R8wPVEAALa9r.jpg
1645	808134635716833280	https://pbs.twimg.com/media/Cx5R8wPVEAALa9r.jpg
1603	800141422401830912	https://pbs.twimg.com/media/CxqsX-8XUAAEvjD.jpg
1727	820446719150292993	https://pbs.twimg.com/media/CxqsX-8XUAAEvjD.jpg
1944	861769973181624320	https://pbs.twimg.com/media/CzG425nWgAAAnP7P.jpg
1638	806629075125202948	https://pbs.twimg.com/media/CzG425nWgAAAnP7P.jpg
1970	868880397819494401	https://pbs.twimg.com/media/DA7iHL5U0AA10Qo.jpg
1992	873697596434513921	https://pbs.twimg.com/media/DA7iHL5U0AA10Qo.jpg
2051	887473957103951883	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2055	888202515573088257	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
1033	711694788429553666	https://pbs.twimg.com/tweet_video_thumb/CeBym7...
1364	761371037149827077	https://pbs.twimg.com/tweet_video_thumb/CeBym7...

	img_num	p1	p1_conf	p1_dog \
480	1	upright	0.303415	False
1297	1	upright	0.303415	False
1864	1	Chihuahua	0.505370	True
1641	1	Chihuahua	0.505370	True
1703	1	Tibetan_mastiff	0.506312	True
1691	1	Tibetan_mastiff	0.506312	True
1705	1	ice_bear	0.336200	False
1858	1	ice_bear	0.336200	False
1715	1	standard_poodle	0.351308	True
1718	1	standard_poodle	0.351308	True
1716	4	prison	0.907083	False
1717	4	prison	0.907083	False
1746	1	Samoyed	0.585441	True
1738	1	Samoyed	0.585441	True
1740	1	Samoyed	0.416769	True
1742	1	Samoyed	0.416769	True
1767	1	golden_retriever	0.617389	True
1789	1	golden_retriever	0.617389	True
1785	1	Staffordshire_bullterrier	0.757547	True
1903	1	Staffordshire_bullterrier	0.757547	True
2041	1	Labrador_retriever	0.908703	True
1791	1	Labrador_retriever	0.908703	True
224	1	Irish_terrier	0.254856	True
1345	1	Irish_terrier	0.254856	True
1590	1	English_springer	0.403698	True
241	1	English_springer	0.403698	True
1589	1	beagle	0.636169	True
85	1	beagle	0.636169	True
327	1	chow	0.243529	True
1591	1	chow	0.243529	True
...
1571	3	schipperke	0.363272	True
1530	3	schipperke	0.363272	True
1538	1	tub	0.479477	False
1533	1	tub	0.479477	False
1683	1	Labrador_retriever	0.427742	True
1534	1	Labrador_retriever	0.427742	True
1539	1	dishwasher	0.700466	False
1699	1	dishwasher	0.700466	False
1544	1	golden_retriever	0.705092	True
1564	1	golden_retriever	0.705092	True
1804	1	Afghan_hound	0.274637	True
1561	1	Afghan_hound	0.274637	True
1566	1	Labrador_retriever	0.861651	True
1652	1	Labrador_retriever	0.861651	True
1572	2	gas_pump	0.676439	False
1755	2	gas_pump	0.676439	False

1579	1	golden_retriever	0.600276	True
1578	1	golden_retriever	0.600276	True
1610	1	cocker_spaniel	0.740220	True
1645	1	cocker_spaniel	0.740220	True
1603	3	golden_retriever	0.938048	True
1727	3	golden_retriever	0.938048	True
1944	2	Arabian_camel	0.366248	False
1638	2	Arabian_camel	0.366248	False
1970	1	laptop	0.153718	False
1992	1	laptop	0.153718	False
2051	2	Pembroke	0.809197	True
2055	2	Pembroke	0.809197	True
1033	1	brown_bear	0.713293	False
1364	1	brown_bear	0.713293	False

		p2	p2_conf	p2_dog	\
480		golden_retriever	0.181351	True	
1297		golden_retriever	0.181351	True	
1864		Pomeranian	0.120358	True	
1641		Pomeranian	0.120358	True	
1703		Tibetan_terrier	0.295690	True	
1691		Tibetan_terrier	0.295690	True	
1705		Samoyed	0.201358	True	
1858		Samoyed	0.201358	True	
1715		toy_poodle	0.271929	True	
1718		toy_poodle	0.271929	True	
1716		palace	0.020089	False	
1717		palace	0.020089	False	
1746		Pomeranian	0.193654	True	
1738		Pomeranian	0.193654	True	
1740		malamute	0.252706	True	
1742		malamute	0.252706	True	
1767		Labrador_retriever	0.337053	True	
1789		Labrador_retriever	0.337053	True	
1785	American_Staffordshire_terrier		0.149950	True	
1903	American_Staffordshire_terrier		0.149950	True	
2041		seat_belt	0.057091	False	
1791		seat_belt	0.057091	False	
224		briard	0.227716	True	
1345		briard	0.227716	True	
1590		Brittany_spaniel	0.347609	True	
241		Brittany_spaniel	0.347609	True	
1589		Labrador_retriever	0.119256	True	
85		Labrador_retriever	0.119256	True	
327		hamster	0.227150	False	
1591		hamster	0.227150	False	
...		
1571		kelpie	0.197021	True	

1530	kelpie	0.197021	True
1538	bathtub	0.325106	False
1533	bathtub	0.325106	False
1683	Great_Dane	0.190503	True
1534	Great_Dane	0.190503	True
1539	golden_retriever	0.245773	True
1699	golden_retriever	0.245773	True
1544	Labrador_retriever	0.219721	True
1564	Labrador_retriever	0.219721	True
1804	borzoi	0.142204	True
1561	borzoi	0.142204	True
1566	golden_retriever	0.044462	True
1652	golden_retriever	0.044462	True
1572	harvester	0.049995	False
1755	harvester	0.049995	False
1579	Labrador_retriever	0.140798	True
1578	Labrador_retriever	0.140798	True
1610	Dandie_Dinmont	0.061604	True
1645	Dandie_Dinmont	0.061604	True
1603	kuvasz	0.025119	True
1727	kuvasz	0.025119	True
1944	house_finch	0.209852	False
1638	house_finch	0.209852	False
1970	French_bulldog	0.099984	True
1992	French_bulldog	0.099984	True
2051	Rhodesian_ridgeback	0.054950	True
2055	Rhodesian_ridgeback	0.054950	True
1033	Indian_elephant	0.172844	False
1364	Indian_elephant	0.172844	False

	p3	p3_conf	p3_dog
480	Brittany_spaniel	0.162084	True
1297	Brittany_spaniel	0.162084	True
1864	toy_terrier	0.077008	True
1641	toy_terrier	0.077008	True
1703	otterhound	0.036251	True
1691	otterhound	0.036251	True
1705	Eskimo_dog	0.186789	True
1858	Eskimo_dog	0.186789	True
1715	Tibetan_terrier	0.094759	True
1718	Tibetan_terrier	0.094759	True
1716	umbrella	0.007850	False
1717	umbrella	0.007850	False
1746	Arctic_fox	0.071648	False
1738	Arctic_fox	0.071648	False
1740	kuvasz	0.157028	True
1742	kuvasz	0.157028	True
1767	tennis_ball	0.008554	False

1789	tennis_ball	0.008554	False
1785	Chesapeake_Bay_retriever	0.047523	True
1903	Chesapeake_Bay_retriever	0.047523	True
2041	pug	0.011933	True
1791	pug	0.011933	True
224	soft-coated_wheaten_terrier	0.223263	True
1345	soft-coated_wheaten_terrier	0.223263	True
1590	Welsh_springer_spaniel	0.137186	True
241	Welsh_springer_spaniel	0.137186	True
1589	golden_retriever	0.082549	True
85	golden_retriever	0.082549	True
327	Pomeranian	0.056057	True
1591	Pomeranian	0.056057	True
...
1571	Norwegian_elkhound	0.151024	True
1530	Norwegian_elkhound	0.151024	True
1538	golden_retriever	0.078530	True
1533	golden_retriever	0.078530	True
1683	curly-coated_retriever	0.146427	True
1534	curly-coated_retriever	0.146427	True
1539	chow	0.039012	True
1699	chow	0.039012	True
1544	kuvasz	0.015965	True
1564	kuvasz	0.015965	True
1804	doormat	0.109677	False
1561	doormat	0.109677	False
1566	Staffordshire_bullterrier	0.016497	True
1652	Staffordshire_bullterrier	0.016497	True
1572	swing	0.044660	False
1755	swing	0.044660	False
1579	seat_belt	0.087355	False
1578	seat_belt	0.087355	False
1610	English_setter	0.041331	True
1645	English_setter	0.041331	True
1603	Labrador_retriever	0.022977	True
1727	Labrador_retriever	0.022977	True
1944	cocker_spaniel	0.046403	True
1638	cocker_spaniel	0.046403	True
1970	printer	0.077130	False
1992	printer	0.077130	False
2051	beagle	0.038915	True
2055	beagle	0.038915	True
1033	water_buffalo	0.038902	False
1364	water_buffalo	0.038902	False

[132 rows x 12 columns]

dog prediction will be more clear without '_' and '-' separator and all character must re-cast

into lowercase

0.2.3 Json_dataframe

```
In [36]: json_df
```

```
Out[36]:
```

	tweet_id	full_text \
0	892420643555336193	This is Phineas. He's a mystical boy. Only eve...
1	892177421306343426	This is Tilly. She's just checking pup on you...
2	891815181378084864	This is Archie. He is a rare Norwegian Pouncin...
3	891689557279858688	This is Darla. She commenced a snooze mid meal...
4	891327558926688256	This is Franklin. He would like you to stop ca...
5	891087950875897856	Here we have a majestic great white breaching ...
6	890971913173991426	Meet Jax. He enjoys ice cream so much he gets ...
7	890729181411237888	When you watch your owner call another dog a g...
8	890609185150312448	This is Zoey. She doesn't want to be one of th...
9	890240255349198849	This is Cassie. She is a college pup. Studying...
10	890006608113172480	This is Koda. He is a South Australian decksha...
11	889880896479866881	This is Bruno. He is a service shark. Only get...
12	889665388333682689	Here's a puppo that seems to be on the fence a...
13	889638837579907072	This is Ted. He does his best. Sometimes that'...
14	889531135344209921	This is Stuart. He's sporting his favorite fan...
15	889278841981685760	This is Oliver. You're witnessing one of his m...
16	888917238123831296	This is Jim. He found a fren. Taught him how t...
17	888804989199671297	This is Zeke. He has a new stick. Very proud o...
18	888554962724278272	This is Ralphus. He's powering up. Attempting ...
19	888078434458587136	This is Gerald. He was just told he didn't get...
20	887705289381826560	This is Jeffrey. He has a monopoly on the pool...
21	887517139158093824	I've yet to rate a Venezuelan Hover Wiener. Th...
22	887473957103951883	This is Canela. She attempted some fancy porch...
23	887343217045368832	You may not have known you needed to see this ...
24	887101392804085760	This... is a Jubilant Antarctic House Bear. We...
25	886983233522544640	This is Maya. She's very shy. Rarely leaves he...
26	886736880519319552	This is Mingus. He's a wonderful father to his...
27	886680336477933568	This is Derek. He's late for a dog meeting. 13...
28	886366144734445568	This is Roscoe. Another pupper fallen victim t...
29	886267009285017600	@NonWhiteHat @MayhewMayhem omg hello tanner yo...
...
2301	666411507551481857	This is quite the dog. Gets really excited whe...
2302	666407126856765440	This is a southern Vesuvius bumblegruff. Can d...
2303	666396247373291520	Oh goodness. A super rare northeast Qdoba kang...
2304	666373753744588802	Those are sunglasses and a jean jacket. 11/10 ...
2305	666362758909284353	Unique dog here. Very small. Lives in containe...
2306	666353288456101888	Here we have a mixed Asiago from the Galápagos...
2307	666345417576210432	Look at this jokester thinking seat belt laws ...
2308	666337882303524864	This is an extremely rare horned Parthenon. No...
2309	666293911632134144	This is a funny dog. Weird toes. Won't come do...
2310	666287406224695296	This is an Albanian 3 1/2 legged Episcopalian...

2311 666273097616637952 Can take selfies 11/10 <https://t.co/ws2AMaNwPW>
 2312 666268910803644416 Very concerned about fellow dog trapped in com...
 2313 666104133288665088 Not familiar with this breed. No tail (weird)...
 2314 666102155909144576 Oh my. Here you are seeing an Adobe Setter giv...
 2315 666099513787052032 Can stand on stump for what seems like a while...
 2316 666094000022159362 This appears to be a Mongolian Presbyterian mi...
 2317 666082916733198337 Here we have a well-established sunblockerspan...
 2318 666073100786774016 Let's hope this flight isn't Malaysian (lol). ...
 2319 666071193221509120 Here we have a northern speckled Rhododendron...
 2320 666063827256086533 This is the happiest dog you will ever see. Ve...
 2321 666058600524156928 Here is the Rand Paul of retrievers folks! He'...
 2322 666057090499244032 My oh my. This is a rare blond Canadian terrie...
 2323 666055525042405380 Here is a Siberian heavily armored polar bear ...
 2324 666051853826850816 This is an odd dog. Hard on the outside but lo...
 2325 666050758794694657 This is a truly beautiful English Wilson Staff...
 2326 666049248165822465 Here we have a 1949 1st generation vulpix. Enj...
 2327 666044226329800704 This is a purebred Piers Morgan. Loves to Netf...
 2328 666033412701032449 Here is a very happy pup. Big fan of well-main...
 2329 666029285002620928 This is a western brown Mitsubishi terrier. Up...
 2330 666020888022790149 Here we have a Japanese Irish Setter. Lost eye...

	in_reply_to_status_id	in_reply_to_user_id	retweet_count	\
0	NaN	NaN	7740	
1	NaN	NaN	5730	
2	NaN	NaN	3791	
3	NaN	NaN	7902	
4	NaN	NaN	8524	
5	NaN	NaN	2852	
6	NaN	NaN	1859	
7	NaN	NaN	17268	
8	NaN	NaN	3934	
9	NaN	NaN	6730	
10	NaN	NaN	6721	
11	NaN	NaN	4569	
12	NaN	NaN	9165	
13	NaN	NaN	4115	
14	NaN	NaN	2063	
15	NaN	NaN	4883	
16	NaN	NaN	4124	
17	NaN	NaN	3889	
18	NaN	NaN	3190	
19	NaN	NaN	3171	
20	NaN	NaN	4932	
21	NaN	NaN	10784	
22	NaN	NaN	16495	
23	NaN	NaN	9587	
24	NaN	NaN	5457	
25	NaN	NaN	7025	

26	NaN	NaN	2947
27	NaN	NaN	4090
28	NaN	NaN	2900
29	8.862664e+17	2.281182e+09	4
...
2301	NaN	NaN	302
2302	NaN	NaN	32
2303	NaN	NaN	76
2304	NaN	NaN	85
2305	NaN	NaN	522
2306	NaN	NaN	66
2307	NaN	NaN	128
2308	NaN	NaN	84
2309	NaN	NaN	322
2310	NaN	NaN	62
2311	NaN	NaN	73
2312	NaN	NaN	32
2313	NaN	NaN	6011
2314	NaN	NaN	11
2315	NaN	NaN	61
2316	NaN	NaN	68
2317	NaN	NaN	42
2318	NaN	NaN	147
2319	NaN	NaN	54
2320	NaN	NaN	199
2321	NaN	NaN	54
2322	NaN	NaN	128
2323	NaN	NaN	222
2324	NaN	NaN	781
2325	NaN	NaN	51
2326	NaN	NaN	41
2327	NaN	NaN	132
2328	NaN	NaN	41
2329	NaN	NaN	42
2330	NaN	NaN	465

	favorite_count
0	36366
1	31326
2	23598
3	39669
4	37867
5	19069
6	11109
7	61269
8	26251
9	29986
10	28879

11	26242
12	45215
13	25450
14	14234
15	23710
16	27406
17	24073
18	18598
19	20501
20	28459
21	43665
22	64746
23	31676
24	28826
25	32912
26	11290
27	21157
28	19918
29	113
...	...
2301	414
2302	100
2303	159
2304	174
2305	729
2306	200
2307	275
2308	182
2309	469
2310	138
2311	161
2312	96
2313	13712
2314	72
2315	143
2316	154
2317	103
2318	298
2319	136
2320	450
2321	105
2322	270
2323	411
2324	1130
2325	124
2326	96
2327	273
2328	112


```

2329          121
2330          2422

```

```
[2331 rows x 6 columns]
```

```
In [37]: json_df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2331 entries, 0 to 2330
Data columns (total 6 columns):
tweet_id          2331 non-null int64
full_text         2331 non-null object
in_reply_to_status_id  77 non-null float64
in_reply_to_user_id  77 non-null float64
retweet_count      2331 non-null int64
favorite_count      2331 non-null int64
dtypes: float64(2), int64(3), object(1)
memory usage: 109.3+ KB

```

```
In [38]: json_df.describe()
```

```

Out[38]:

```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id \
count	2.331000e+03	7.700000e+01	7.700000e+01
mean	7.419079e+17	7.440692e+17	2.040329e+16
std	6.823170e+16	7.524295e+16	1.260797e+17
min	6.660209e+17	6.658147e+17	1.185634e+07
25%	6.782670e+17	6.757073e+17	3.589728e+08
50%	7.182469e+17	7.032559e+17	4.196984e+09
75%	7.986692e+17	8.233264e+17	4.196984e+09
max	8.924206e+17	8.862664e+17	8.405479e+17

	retweet_count	favorite_count
count	2331.00000	2331.000000
mean	2716.42042	7594.847276
std	4594.59356	11784.709902
min	1.00000	0.000000
25%	548.50000	1324.500000
50%	1275.00000	3307.000000
75%	3156.00000	9306.000000
max	78210.00000	156972.000000

max value in retweet_count and favourite_count is unusual, deeper inspection is needed

```

In [39]: # adding favorite_retweet_ratio that calculated from favorite_count/retweet_count
         json_df['favorite_retweet_ration'] = json_df['favorite_count']/json_df['retweet_count']

```

```
In [40]: json_df.describe()
```

```
Out[40]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id \
count	2.331000e+03	7.700000e+01	7.700000e+01
mean	7.419079e+17	7.440692e+17	2.040329e+16
std	6.823170e+16	7.524295e+16	1.260797e+17
min	6.660209e+17	6.658147e+17	1.185634e+07
25%	6.782670e+17	6.757073e+17	3.589728e+08
50%	7.182469e+17	7.032559e+17	4.196984e+09
75%	7.986692e+17	8.233264e+17	4.196984e+09
max	8.924206e+17	8.862664e+17	8.405479e+17

	retweet_count	favorite_count	favorite_retweet_ratio
count	2331.00000	2331.000000	2331.000000
mean	2716.42042	7594.847276	3.563748
std	4594.59356	11784.709902	4.175131
min	1.00000	0.000000	0.000000
25%	548.50000	1324.500000	2.471211
50%	1275.00000	3307.000000	3.226345
75%	3156.00000	9306.000000	4.072335
max	78210.00000	156972.000000	140.000000

```
In [41]: json_df.query('(retweet_count>3156) & (favorite_count<3307)')
```

```
Out[41]:
```

	tweet_id	full_text \
35	885311592912609280	RT @dog_rates: This is Lilly. She just paralle...
67	879130579576475649	RT @dog_rates: This is Emmy. She was adopted t...
73	878316110768087041	RT @dog_rates: Meet Terrance. He's being yelle...
90	874434818259525634	RT @dog_rates: This is Coco. At first I though...
105	871166179821445120	RT @dog_rates: This is Dawn. She's just checki...
131	866094527597207552	RT @dog_rates: Here's a pupper before and afte...
158	860177593139703809	RT @dog_rates: Ohboyohboyohboyohboyohboyohboy...
164	858860390427611136	RT @dog_rates: Meet Winston. He knows he's a l...
186	855245323840757760	RT @dog_rates: Meet George. He looks slightly ...
213	849668094696017920	RT @dog_rates: This is Gidget. She's a spy pup...
221	847978865427394560	RT @dog_rates: This is Charlie. He fell asleep...
240	845098359547420673	RT @dog_rates: This is Bungalo. She uses that ...
254	841833993020538882	RT @dog_rates: This is Ken. His cheeks are mag...
261	840728873075638272	RT @dog_rates: This is Pipsy. He is a fluffbal...
274	838831947270979586	RT @dog_rates: This is Riley. His owner put a ...
277	838201503651401729	RT @dog_rates: Meet Sunny. He can take down a ...
289	836397794269200385	RT @dog_rates: This is Buddy. He ran into a gl...
293	835685285446955009	RT @dog_rates: This is Rory. He's got an inter...
296	835309094223372289	RT @dog_rates: So this just changed my life. 1...
305	834477809192075265	RT @dog_rates: This is Leo. He was a skater pu...
326	832215909146226688	RT @dog_rates: This is Logan, the Chow who liv...
327	832215726631055365	RT @dog_rates: This is Moreton. He's the Good ...
329	832040443403784192	RT @dog_rates: This is Klein. These pics were ...
343	830173239259324417	RT @dog_rates: This is Kyro. He's a Stratocumu...
345	829878982036299777	RT @dog_rates: This is Loki. He smiles like El...

351 828801551087042563 RT @dog_rates: This is Gus. He likes to be clo...
 370 826615380357632002 RT @dog_rates: Please only send in dogs. We on...
 383 824796380199809024 RT @dog_rates: This is Bailey. She loves going...
 390 823719002937630720 RT @dog_rates: This is Balto. He's very conten...
 395 823269594223824897 RT @dog_rates: We only rate dogs. Please don't...

 651 790227638568808452 RT @dog_rates: This is Leela. She's a Fetty Wo...
 653 789960241177853952 RT @dog_rates: This is Buddy. His father was a...
 659 789280767834746880 RT @dog_rates: This is Meyer. He has to hold s...
 664 788552643979468800 RT @dog_rates: Say hello to mad pupper. You kn...
 668 788070120937619456 RT @dog_rates: This is Bo and Ty. Bo eats pape...
 674 787111942498508800 RT @dog_rates: This is Carl. He's very powerfu...
 684 786036967502913536 RT @dog_rates: This is Scout. He really wants ...
 702 783347506784731136 RT @dog_rates: This is Kenny. He just wants to...
 710 782021823840026624 RT @dog_rates: This is Harper. She scraped her...
 723 780496263422808064 RT @dog_rates: This is Bell. She likes holding...
 731 779124354206535695 RT @dog_rates: Meet Hurley. He's the curly one...
 734 778774459159379968 RT @dog_rates: In case you haven't seen the mo...
 740 778396591732486144 RT @dog_rates: This is an East African Chalupa...
 745 777953400541634568 RT @dog_rates: Meet Gerald. He's a fairly exot...
 748 777641927919427584 RT @dog_rates: This is Arnie. He's a Nova Scot...
 751 776819012571455488 RT @dog_rates: Everybody look at this beautifu...
 754 776249906839351296 RT @dog_rates: We only rate dogs. Pls stop sen...
 759 775898661951791106 RT @dog_rates: Like father (doggo), like son (...
 774 773336787167145985 RT @dog_rates: Meet Fizz. She thinks love is a...
 780 772615324260794368 RT @dog_rates: This is Gromit. He's pupset bec...
 791 771171053431250945 RT @dog_rates: This is Frankie. He's wearing b...
 804 769335591808995329 RT @dog_rates: Ever seen a dog pet another dog...
 811 768554158521745409 RT @dog_rates: This is Nollie. She's waving at...
 819 766864461642756096 RT @dog_rates: We only rate dogs... this is a ...
 846 761750502866649088 RT @dog_rates: "Tristan do not speak to me wit...
 850 761371037149827077 RT @dog_rates: Oh. My. God. 13/10 magical af h...
 885 757729163776290825 RT @dog_rates: This is Chompsky. He lives up t...
 903 754874841593970688 RT @dog_rates: This is Rubio. He has too much ...
 913 753298634498793472 RT @dog_rates: This is Carly. She's actually 2...
 925 752309394570878976 RT @dog_rates: Everyone needs to watch this. 1...

	in_reply_to_status_id	in_reply_to_user_id	retweet_count \
35	NaN	NaN	16975
67	NaN	NaN	6243
73	NaN	NaN	6123
90	NaN	NaN	13401
105	NaN	NaN	5299
131	NaN	NaN	7886
158	NaN	NaN	29468
164	NaN	NaN	7765
186	NaN	NaN	5777
213	NaN	NaN	4886

221	NaN	NaN	3218
240	NaN	NaN	7597
254	NaN	NaN	15074
261	NaN	NaN	16140
274	NaN	NaN	11035
277	NaN	NaN	10090
289	NaN	NaN	27648
293	NaN	NaN	8177
296	NaN	NaN	21093
305	NaN	NaN	10728
326	NaN	NaN	6263
327	NaN	NaN	9425
329	NaN	NaN	9717
343	NaN	NaN	5756
345	NaN	NaN	5089
351	NaN	NaN	3457
370	NaN	NaN	4040
383	NaN	NaN	5583
390	NaN	NaN	11393
395	NaN	NaN	10062
..
651	NaN	NaN	3665
653	NaN	NaN	4231
659	NaN	NaN	5087
664	NaN	NaN	7372
668	NaN	NaN	3516
674	NaN	NaN	6912
684	NaN	NaN	7206
702	NaN	NaN	5674
710	NaN	NaN	6351
723	NaN	NaN	3784
731	NaN	NaN	18725
734	NaN	NaN	10115
740	NaN	NaN	12497
745	NaN	NaN	3595
748	NaN	NaN	4314
751	NaN	NaN	13962
754	NaN	NaN	9326
759	NaN	NaN	16854
774	NaN	NaN	5177
780	NaN	NaN	3439
791	NaN	NaN	7613
804	NaN	NaN	7847
811	NaN	NaN	5921
819	NaN	NaN	5720
846	NaN	NaN	3965
850	NaN	NaN	18285
885	NaN	NaN	8037

903	NaN	NaN	8003
913	NaN	NaN	5756
925	NaN	NaN	16416

	favorite_count	favorite_retweet_ratio
35	0	0.0
67	0	0.0
73	0	0.0
90	0	0.0
105	0	0.0
131	0	0.0
158	0	0.0
164	0	0.0
186	0	0.0
213	0	0.0
221	0	0.0
240	0	0.0
254	0	0.0
261	0	0.0
274	0	0.0
277	0	0.0
289	0	0.0
293	0	0.0
296	0	0.0
305	0	0.0
326	0	0.0
327	0	0.0
329	0	0.0
343	0	0.0
345	0	0.0
351	0	0.0
370	0	0.0
383	0	0.0
390	0	0.0
395	0	0.0
..
651	0	0.0
653	0	0.0
659	0	0.0
664	0	0.0
668	0	0.0
674	0	0.0
684	0	0.0
702	0	0.0
710	0	0.0
723	0	0.0
731	0	0.0
734	0	0.0

740	0	0.0
745	0	0.0
748	0	0.0
751	0	0.0
754	0	0.0
759	0	0.0
774	0	0.0
780	0	0.0
791	0	0.0
804	0	0.0
811	0	0.0
819	0	0.0
846	0	0.0
850	0	0.0
885	0	0.0
903	0	0.0
913	0	0.0
925	0	0.0

[112 rows x 7 columns]

It is weird to see favorite_retweet_ratio is 0, in fact that tweet has big retweet count. With average favorite_retweet_ratio around 3.22, favorite count should be bigger than retweet count

In [42]: json_df.query('favorite_retweet_ration>4.07').sort_values('favorite_retweet_ration', as

```
Out[42]:
```

	tweet_id	full_text \
279	838085839343206401	@bragg6of8 @Andy_Pace_ we are still looking fo...
262	840698636975636481	@O_kelvin_0 >10/10 is reserved for puppos s...
109	870726314365509632	@ComplicitOwl @ShopWeRateDogs >10/10 is res...
178	856288084350160898	@xianmcguire @Jenna_Marbles Kardashians wouldn...
328	832088576586297345	@docmisterio account started on 11/15/15
63	879674319642796034	@RealKentMurphy 14/10 confirmed
299	835246439529840640	@jonnyysun @Lin_Manuel ok jomny I know you're e...
393	823333489516937216	@HistoryInPics 13/10
29	886267009285017600	@NonWhiteHat @MayhewMayhem omg hello tanner yo...
411	821153421864615936	@imgur for a polar bear tho I'd say 13/10 is a...
1271	707983188426153984	@serial @MrRoles OH MY GOD I listened to all o...
943	750381685133418496	13/10 such a good doggo\n@spaghemily
142	863427515083354112	@Jack_Septic_Eye I'd need a few more pics to p...
371	826598799820865537	I was going to do 007/10, but the joke wasn't ...
1056	738891149612572673	@mount_alex3 13/10
1499	690607260360429569	12/10 @LightningHolt
54	881633300179243008	@rroushfenway These are good dogs but 17/10 is ...
225	847617282490613760	.@breaannanicolee PUPDATE: Cannon has a heart ...
1581	685681090388975616	Jack deserves another round of applause. If yo...
835	763956972077010945	@TheEllenShow I'm not sure if you know this bu...
180	855862651834028034	@dhmontgomery We also gave snoop dogg a 420/10...

172 857214891891077121 @Marc_IRL pixelated af 12/10
360 828361771580813312 Beebop and Doobert should start a band 12/10 w...
263 840696689258311684 I didn't even have to intervene. Took him 4 mi...
758 776088319444877312 I can't tap the screen to make the hearts appe...
161 859607811541651456 Sorry for the lack of posts today. I came home...
1178 716439118184652801 This is Bluebert. He just saw that both #Final...
219 848213670039564288 Jerry just apuppologized to me. He said there ...
1450 693644216740769793 BREAKING PUPDATE: I've just been notified that...
1639 682808988178739200 I'm aware that I could've said 20/16, but here...
... ..
458 816091915477250048 This is Comet. He's a Wild Estonian Poofer. Su...
1394 698355670425473025 This is Jessiga. She's a Tasmanian McCringlebe...
1588 685321586178670592 This is Rocky. He sleeps like a psychopath. 10...
1234 710283270106132480 This is Gunner. He's a Figamus Newton. King of...
797 770655142660169732 We only rate dogs. Pls stop sending in non-can...
678 786664955043049472 "Honestly Kathleen I just want more Ken Bone" ...
497 811627233043480576 This is Augie. He's a savage. Doesn't give a h...
1144 721001180231503872 This is Oliver. Bath time is upon him. His fea...
852 761292947749015552 Meet Bonaparte. He's pupset because it's cloud...
2267 667160273090932737 This is Bradlay. He is a Ronaldinho Matsuyama ...
2032 671355857343524864 This is Lou. He's a Petrarch Sunni Pinto. Well...
2081 670468609693655041 This is Edd. He's a Czechoslovakian Googolplex...
501 810657578271330305 This is Pavlov. His floatation device has fail...
887 757611664640446465 This is Cooper. He tries to come across as fei...
1662 681610798867845120 *collapses* 12/10 <https://t.co/C7M8mnzHIK>
555 801167903437357056 This is Milo. I would do terrible things for M...
1427 695794761660297217 This is Wyatt. His throne is modeled after him...
1453 693622659251335168 When you keepin the popcorn bucket in your lap...
1906 674036086168010753 Meet Daisy. She has no eyes & her face has...
795 770787852854652928 This is Winston. His tongue has gone rogue. Do...
1418 696754882863349760 This is Reptar. He specifically asked for his ...
492 812466873996607488 This is Mary. She's desperately trying to recr...
1406 697575480820686848 This is Ole. He's not sure how to gravity. 8/1...
224 847842811428974592 This is Rontu. He is described as a pal, cuddl...
361 828046555563323392 This is Sailer. He waits on the roof for his o...
889 757596066325864448 Here's another picture without a dog in it. Id...
1167 717841801130979328 This is Barclay. His father was a banana. 11/1...
1389 698907974262222848 This dog is never sure if he's doing the right...
1109 728751179681943552 This is Flurpson. He can't believe it's not bu...
1145 720785406564900865 This is Archie. He hears everything you say. D...

	in_reply_to_status_id	in_reply_to_user_id	retweet_count	\
279	8.380855e+17	2.894131e+09	1	
262	8.406983e+17	8.405479e+17	2	
109	8.707262e+17	1.648776e+07	3	
178	NaN	NaN	15	
328	8.320875e+17	3.058208e+07	2	
63	8.795538e+17	3.105441e+09	10	

299	8.352460e+17	2.625958e+07	73
393	8.233264e+17	1.582854e+09	25
29	8.862664e+17	2.281182e+09	4
411	8.211526e+17	1.132119e+08	10
1271	7.079801e+17	2.319108e+09	2
943	7.501805e+17	4.717297e+09	28
142	8.634256e+17	7.759620e+07	91
371	8.265984e+17	4.196984e+09	257
1056	7.384119e+17	3.589728e+08	6
1499	6.903413e+17	4.670367e+08	17
54	8.816070e+17	4.738443e+07	7
225	8.476062e+17	4.196984e+09	410
1581	6.855479e+17	4.196984e+09	124
835	7.638652e+17	1.584641e+07	56
180	8.558616e+17	1.943518e+08	25
172	8.571567e+17	1.806710e+08	17
360	NaN	NaN	173
263	NaN	NaN	974
758	NaN	NaN	154
161	NaN	NaN	1489
1178	NaN	NaN	207
219	8.482121e+17	4.196984e+09	722
1450	6.936422e+17	4.196984e+09	123
1639	6.827884e+17	4.196984e+09	169
...
458	NaN	NaN	2192
1394	NaN	NaN	449
1588	NaN	NaN	641
1234	NaN	NaN	517
797	NaN	NaN	1797
678	NaN	NaN	2622
497	NaN	NaN	3161
1144	NaN	NaN	608
852	NaN	NaN	1097
2267	NaN	NaN	60
2032	NaN	NaN	114
2081	NaN	NaN	84
501	NaN	NaN	2721
887	NaN	NaN	1116
1662	NaN	NaN	464
555	NaN	NaN	6095
1427	NaN	NaN	768
1453	NaN	NaN	375
1906	NaN	NaN	202
795	NaN	NaN	1218
1418	NaN	NaN	361
492	NaN	NaN	1987
1406	NaN	NaN	498

224	NaN	NaN	1318
361	NaN	NaN	2913
889	NaN	NaN	1071
1167	NaN	NaN	603
1389	NaN	NaN	625
1109	NaN	NaN	679
1145	NaN	NaN	763

	favorite_count	favorite_retweet_ratio
279	140	140.000000
262	175	87.500000
109	116	38.666667
178	504	33.600000
328	64	32.000000
63	297	29.700000
299	2117	29.000000
393	710	28.400000
29	113	28.250000
411	261	26.100000
1271	51	25.500000
943	705	25.178571
142	2127	23.373626
371	5184	20.171206
1056	111	18.500000
1499	294	17.294118
54	120	17.142857
225	6943	16.934146
1581	2037	16.427419
835	767	13.696429
180	332	13.280000
172	219	12.882353
360	2220	12.832370
263	12306	12.634497
758	1878	12.194805
161	17983	12.077233
1178	2371	11.454106
219	8106	11.227147
1450	1346	10.943089
1639	1792	10.603550
...
458	9104	4.153285
1394	1864	4.151448
1588	2661	4.151326
1234	2145	4.148936
797	7451	4.146355
678	10852	4.138825
497	13082	4.138564
1144	2516	4.138158

852	4538	4.136737
2267	248	4.133333
2032	471	4.131579
2081	347	4.130952
501	11239	4.130467
887	4607	4.128136
1662	1915	4.127155
555	25111	4.119934
1427	3163	4.118490
1453	1543	4.114667
1906	831	4.113861
795	5006	4.110016
1418	1482	4.105263
492	8156	4.104680
1406	2042	4.100402
224	5396	4.094082
361	11910	4.088568
889	4377	4.086835
1167	2461	4.081260
1389	2550	4.080000
1109	2767	4.075110
1145	3109	4.074705

[583 rows x 7 columns]

```
In [43]: cek_rt_json = json_df.full_text.str.extract(r'(RT @.*)')
cek_rt_json[cek_rt_json[0].notnull()]
```

```
Out[43]:
```

31	RT @Athletics: 12/10 #BATP https://t.co/WxwJmv...	0
35	RT @dog_rates: This is Lilly. She just paralle...	
67	RT @dog_rates: This is Emmy. She was adopted t...	
72	RT @dog_rates: Meet Shadow. In an attempt to r...	
73	RT @dog_rates: Meet Terrance. He's being yelle...	
77	RT @rachael2195: @dog_rates the boyfriend and h...	
90	RT @dog_rates: This is Coco. At first I though...	
95	RT @dog_rates: This is Sierra. She's one preci...	
105	RT @dog_rates: This is Dawn. She's just checki...	
119	RT @dog_rates: Say hello to Cooper. His expres...	
125	RT @rachaeleasler: these @dog_rates hats are 1...	
131	RT @dog_rates: Here's a pupper before and afte...	
140	RT @dog_rates: Say hello to Quinn. She's quite...	
152	RT @dog_rates: Meet Lorenzo. He's an avid nift...	
153	RT @tallylott: h*ckin adorable promposal. 13/1...	
158	RT @dog_rates: Ohboyohboyohboyohboyohboyohboy...	
164	RT @dog_rates: Meet Winston. He knows he's a l...	
173	RT @AaronChewning: First time wearing my @dog_...	
177	RT @Jenna_Marbles: @dog_rates Thanks for ratin...	

```

186 RT @dog_rates: Meet George. He looks slightly ...
187 RT @frasercampbell_: oh my... what's that... b...
196 RT @dog_rates: I usually only share these on F...
203 RT @eddie_coe98: Thanks @dog_rates completed m...
213 RT @dog_rates: This is Gidget. She's a spy pup...
221 RT @dog_rates: This is Charlie. He fell asleep...
222 RT @basic_vacek_: I love my new mug easy 13/10...
240 RT @dog_rates: This is Bungalo. She uses that ...
254 RT @dog_rates: This is Ken. His cheeks are mag...
260 RT @dog_rates: Say hello to Maddie and Gunner...
261 RT @dog_rates: This is Pipsy. He is a fluffbal...
...
748 RT @dog_rates: This is Arnie. He's a Nova Scot...
751 RT @dog_rates: Everybody look at this beautifu...
754 RT @dog_rates: We only rate dogs. Pls stop sen...
759 RT @dog_rates: Like father (doggo), like son (...
774 RT @dog_rates: Meet Fizz. She thinks love is a...
780 RT @dog_rates: This is Gromit. He's pupset bec...
791 RT @dog_rates: This is Frankie. He's wearing b...
800 RT @dog_rates: This is just downright precious...
804 RT @dog_rates: Ever seen a dog pet another dog...
807 RT @dog_rates: When it's Janet from accounting...
811 RT @dog_rates: This is Nollie. She's waving at...
819 RT @dog_rates: We only rate dogs... this is a ...
825 RT @dog_rates: This is Colby. He's currently r...
838 RT @dog_rates: Meet Eve. She's a raging alchoh...
846 RT @dog_rates: "Tristan do not speak to me wit...
850 RT @dog_rates: Oh. My. God. 13/10 magical af h...
863 RT @hownottodraw: The story/person behind @dog...
872 RT @dog_rates: AT DAWN...
885 RT @dog_rates: This is Chompsky. He lives up t...
888 RT @jon_hill987: @dog_rates There is a cunning...
903 RT @dog_rates: This is Rubio. He has too much ...
913 RT @dog_rates: This is Carly. She's actually 2...
919 RT @dog_rates: HEY PUP WHAT'S THE PART OF THE ...
925 RT @dog_rates: Everyone needs to watch this. 1...
988 RT @dog_rates: This pupper killed this great w...
999 RT @dog_rates: This is Shaggy. He knows exactl...
1019 RT @dog_rates: Extremely intelligent dog here...
1218 RT @twitter: @dog_rates Awesome Tweet! 12/10. ...
2234 RT @dogratingrating: Exceptional talent. Origi...
2235 RT @dogratingrating: Unoriginal idea. Blatant ...

```

[163 rows x 1 columns]

found Retweeted tweets in json dataframe

0.3 Assess data note

0.4 Quality

twitter atchive enhanced

- timestamp and retweeted status timestamp has type string, which will good in date type
- wrong column name, floofer must be floof
- in_reply_to_status_id and in_reply_to_user_id type is in float, but it must be in int
- Retweet data is in dataframe must be deleted
- rating_numerator and rating_denominator in row 313 value on normalize_rating is 13/10 instead of 960/0

0.4.1 image prediction

- dog prediction will be more clear by replacing '_' separator into space
- Several data duplication stored with different tweet_id
- name of all dogs predictions are variate with upper and lower case

0.4.2 Json data

- retweeted tweets is in data

0.5 Tidy

0.5.1 twitter archive enhanced

- timestamp has date and time in one column
- text column value has two informations, there are tweet content and link

0.6 Cleaning data process in quality

0.6.1 copy dataframe

```
In [73]: twit_arc_copy = twitter_archive_enh.copy()
```

```
In [74]: img_pred_copy = image_predictions.copy()
```

```
In [75]: json_copy = json_df.copy()
```

0.6.2 timestamp and retweet_timestamp has string type, which must be date

0.6.3 Define

timestamp and retweet_timestamp column type will change into date type

0.6.4 Code

```
In [76]: #inspect value in timestamp
        twit_arc_copy.timestamp.head()
```

```
Out[76]: 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
        Name: timestamp, dtype: object
```

```
In [77]: #change data type in timestamp into datetime64
        twit_arc_copy['timestamp'] = twit_arc_copy.timestamp.astype('datetime64')
```

```
In [78]: #change data type in retweet_status_timestamp into datetime64
        twit_arc_copy['retweeted_status_timestamp'] = twit_arc_copy.retweeted_status_timestamp.
```

```
In [79]: twit_arc_copy[['timestamp', 'retweeted_status_timestamp']]
```

```
Out[79]:
```

	timestamp	retweeted_status_timestamp
0	2017-08-01 16:23:56	NaT
1	2017-08-01 00:17:27	NaT
2	2017-07-31 00:18:03	NaT
3	2017-07-30 15:58:51	NaT
4	2017-07-29 16:00:24	NaT
5	2017-07-29 00:08:17	NaT
6	2017-07-28 16:27:12	NaT
7	2017-07-28 00:22:40	NaT
8	2017-07-27 16:25:51	NaT
9	2017-07-26 15:59:51	NaT
10	2017-07-26 00:31:25	NaT
11	2017-07-25 16:11:53	NaT
12	2017-07-25 01:55:32	NaT
13	2017-07-25 00:10:02	NaT
14	2017-07-24 17:02:04	NaT
15	2017-07-24 00:19:32	NaT
16	2017-07-23 00:22:39	NaT
17	2017-07-22 16:56:37	NaT
18	2017-07-22 00:23:06	NaT
19	2017-07-21 01:02:36	2017-07-19 00:47:34
20	2017-07-20 16:49:33	NaT
21	2017-07-19 16:06:48	NaT
22	2017-07-19 03:39:09	NaT
23	2017-07-19 00:47:34	NaT
24	2017-07-18 16:08:03	NaT
25	2017-07-18 00:07:08	NaT
26	2017-07-17 16:17:36	NaT
27	2017-07-16 23:58:41	NaT

28	2017-07-16 20:14:00	NaT
29	2017-07-15 23:25:31	NaT
...
2326	2015-11-17 00:24:19	NaT
2327	2015-11-17 00:06:54	NaT
2328	2015-11-16 23:23:41	NaT
2329	2015-11-16 21:54:18	NaT
2330	2015-11-16 21:10:36	NaT
2331	2015-11-16 20:32:58	NaT
2332	2015-11-16 20:01:42	NaT
2333	2015-11-16 19:31:45	NaT
2334	2015-11-16 16:37:02	NaT
2335	2015-11-16 16:11:11	NaT
2336	2015-11-16 15:14:19	NaT
2337	2015-11-16 14:57:41	NaT
2338	2015-11-16 04:02:55	NaT
2339	2015-11-16 03:55:04	NaT
2340	2015-11-16 03:44:34	NaT
2341	2015-11-16 03:22:39	NaT
2342	2015-11-16 02:38:37	NaT
2343	2015-11-16 01:59:36	NaT
2344	2015-11-16 01:52:02	NaT
2345	2015-11-16 01:22:45	NaT
2346	2015-11-16 01:01:59	NaT
2347	2015-11-16 00:55:59	NaT
2348	2015-11-16 00:49:46	NaT
2349	2015-11-16 00:35:11	NaT
2350	2015-11-16 00:30:50	NaT
2351	2015-11-16 00:24:50	NaT
2352	2015-11-16 00:04:52	NaT
2353	2015-11-15 23:21:54	NaT
2354	2015-11-15 23:05:30	NaT
2355	2015-11-15 22:32:08	NaT

[2356 rows x 2 columns]

0.6.5 Test

```
In [80]: # check data type of changed column
         twit_arc_copy.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 18 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 datetime64[ns]
```

```

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 datetime64[ns]
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
normalize_rating        2356 non-null float64
dtypes: datetime64[ns](2), float64(5), int64(3), object(8)
memory usage: 331.4+ KB

```

0.6.6 wrong column name, floof insted of floofer

0.6.7 Define

rename floofer column into floof

0.6.8 Code

```

In [81]: # Change column named 'floofer' become 'floof'
         twit_arc_copy.rename(columns={'floofer':'floof'}, inplace = True)

```

```

In [82]: twit_arc_copy.head()

```

```

Out[82]:
   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

   timestamp                                     source \
0  2017-08-01 16:23:56  <a href="http://twitter.com/download/iphone" r...
1  2017-08-01 00:17:27  <a href="http://twitter.com/download/iphone" r...
2  2017-07-31 00:18:03  <a href="http://twitter.com/download/iphone" r...
3  2017-07-30 15:58:51  <a href="http://twitter.com/download/iphone" r...
4  2017-07-29 16:00:24  <a href="http://twitter.com/download/iphone" r...

   text  retweeted_status_id \
0  This is Phineas. He's a mystical boy. Only eve...      NaN
1  This is Tilly. She's just checking pup on you...      NaN
2  This is Archie. He is a rare Norwegian Pouncin...      NaN

```

3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN

	retweeted_status_user_id	retweeted_status_timestamp \
0	NaN	NaT
1	NaN	NaT
2	NaN	NaT
3	NaN	NaT
4	NaN	NaT

	expanded_urls	rating_numerator \
0	https://twitter.com/dog_rates/status/892420643...	13
1	https://twitter.com/dog_rates/status/892177421...	13
2	https://twitter.com/dog_rates/status/891815181...	12
3	https://twitter.com/dog_rates/status/891689557...	13
4	https://twitter.com/dog_rates/status/891327558...	12

	rating_denominator	name	doggo	floof	pupper	puppo	normalize_rating
0	10	Phineas	None	None	None	None	1.3
1	10	Tilly	None	None	None	None	1.3
2	10	Archie	None	None	None	None	1.2
3	10	Darla	None	None	None	None	1.3
4	10	Franklin	None	None	None	None	1.2

0.6.9 Test

```
In [83]: # listing all column names in twit archive clean
list (twit_arc_copy.columns)
```

```
Out[83]: ['tweet_id',
          'in_reply_to_status_id',
          'in_reply_to_user_id',
          'timestamp',
          'source',
          'text',
          'retweeted_status_id',
          'retweeted_status_user_id',
          'retweeted_status_timestamp',
          'expanded_urls',
          'rating_numerator',
          'rating_denominator',
          'name',
          'doggo',
          'floof',
          'pupper',
          'puppo',
          'normalize_rating']
```


0.6.10 `in_reply_to_status_id` and `in_reply_to_user_id` should be in `int`, present in `float`

0.6.11 Define

Change `in_reply_to_status_id` type into `int`, either do `in_reply_to_user_id` type

0.6.12 Code

```
In [84]: twitter_archive_enh.tail()
```

```
Out[84]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
2351	666049248165822465	NaN	NaN	
2352	666044226329800704	NaN	NaN	
2353	666033412701032449	NaN	NaN	
2354	666029285002620928	NaN	NaN	
2355	666020888022790149	NaN	NaN	

	timestamp	\
2351	2015-11-16 00:24:50 +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	
2355	2015-11-15 22:32:08 +0000	

	source	\
2351	<a href="http://twitter.com/download/iphone" r...	
2352	<a href="http://twitter.com/download/iphone" r...	
2353	<a href="http://twitter.com/download/iphone" r...	
2354	<a href="http://twitter.com/download/iphone" r...	
2355	<a href="http://twitter.com/download/iphone" r...	

	text	retweeted_status_id	\
2351	Here we have a 1949 1st generation vulpix. Enj...	NaN	
2352	This is a purebred Piers Morgan. Loves to Netf...	NaN	
2353	Here is a very happy pup. Big fan of well-main...	NaN	
2354	This is a western brown Mitsubishi terrier. Up...	NaN	
2355	Here we have a Japanese Irish Setter. Lost eye...	NaN	

	retweeted_status_user_id	retweeted_status_timestamp	\
2351	NaN	NaN	
2352	NaN	NaN	
2353	NaN	NaN	
2354	NaN	NaN	
2355	NaN	NaN	

	expanded_urls	rating_numerator	\
2351	https://twitter.com/dog_rates/status/666049248...	5	
2352	https://twitter.com/dog_rates/status/666044226...	6	

```

2353 https://twitter.com/dog_rates/status/666033412... 9
2354 https://twitter.com/dog_rates/status/666029285... 7
2355 https://twitter.com/dog_rates/status/666020888... 8

```

	rating_denominator	name	doggo	floofer	pupper	puppo	normalize_rating
2351	10	None	None	None	None	None	0.5
2352	10	a	None	None	None	None	0.6
2353	10	a	None	None	None	None	0.9
2354	10	a	None	None	None	None	0.7
2355	10	None	None	None	None	None	0.8

```
In [85]: # change null value into 0
```

```

twit_arc_copy.in_reply_to_status_id = twit_arc_copy.in_reply_to_status_id.fillna(0)
twit_arc_copy.in_reply_to_user_id = twit_arc_copy.in_reply_to_user_id.fillna(0)

```

```
In [86]: twit_arc_copy[twit_arc_copy.in_reply_to_status_id.isnull()]
```

```
Out[86]: Empty DataFrame
```

```

Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text]
Index: []

```

```
In [87]: # change in_reply_to_status_id and in_reply_to_user_id data type to int
```

```

twit_arc_copy.in_reply_to_status_id = twit_arc_copy.in_reply_to_status_id.astype(int)
twit_arc_copy.in_reply_to_user_id = twit_arc_copy.in_reply_to_user_id.astype(int)

```

```
In [88]: twit_arc_copy.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 18 columns):
tweet_id                2356 non-null int64
in_reply_to_status_id   2356 non-null int64
in_reply_to_user_id     2356 non-null int64
timestamp               2356 non-null datetime64[ns]
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 datetime64[ns]
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
floof                   2356 non-null object
pupper                  2356 non-null object
puppo                   2356 non-null object
normalize_rating         2356 non-null float64
dtypes: datetime64[ns](2), float64(3), int64(5), object(8)

```

memory usage: 331.4+ KB

In [89]: `twit_arc_copy.tail()`

```
Out[89]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
2351	666049248165822465	0	0	
2352	666044226329800704	0	0	
2353	666033412701032449	0	0	
2354	666029285002620928	0	0	
2355	666020888022790149	0	0	

	timestamp	source	\
2351	2015-11-16 00:24:50	<a href="http://twitter.com/download/iphone" r...	
2352	2015-11-16 00:04:52	<a href="http://twitter.com/download/iphone" r...	
2353	2015-11-15 23:21:54	<a href="http://twitter.com/download/iphone" r...	
2354	2015-11-15 23:05:30	<a href="http://twitter.com/download/iphone" r...	
2355	2015-11-15 22:32:08	<a href="http://twitter.com/download/iphone" r...	

	text	retweeted_status_id	\
2351	Here we have a 1949 1st generation vulpix. Enj...	NaN	
2352	This is a purebred Piers Morgan. Loves to Netf...	NaN	
2353	Here is a very happy pup. Big fan of well-main...	NaN	
2354	This is a western brown Mitsubishi terrier. Up...	NaN	
2355	Here we have a Japanese Irish Setter. Lost eye...	NaN	

	retweeted_status_user_id	retweeted_status_timestamp	\
2351	NaN	NaT	
2352	NaN	NaT	
2353	NaN	NaT	
2354	NaN	NaT	
2355	NaN	NaT	

	expanded_urls	rating_numerator	\
2351	https://twitter.com/dog_rates/status/666049248...	5	
2352	https://twitter.com/dog_rates/status/666044226...	6	
2353	https://twitter.com/dog_rates/status/666033412...	9	
2354	https://twitter.com/dog_rates/status/666029285...	7	
2355	https://twitter.com/dog_rates/status/666020888...	8	

	rating_denominator	name	doggo	floof	pupper	puppo	normalize_rating
2351	10	None	None	None	None	None	0.5
2352	10	a	None	None	None	None	0.6
2353	10	a	None	None	None	None	0.9
2354	10	a	None	None	None	None	0.7
2355	10	None	None	None	None	None	0.8

0.6.13 Test

```
In [90]: twit_arc_copy['in_reply_to_user_id'].unique()
```

```
Out[90]: array([          0,          2281181600,          47384430,
                3105440746,          16487760,          77596200,
                4196983835,          180670967,          279280991,
                194351775,          13615722,          21955058,
                840547864354918400,          2894131180,          26259576,
                30582082,          20683724,          1582853809,
                113211856,          11856342,          291859009,
                29166305, 730505014150582272,          15846407,
                4717297476,          358972768,          2319108198,
                28785486,          1198988510,          467036706,
                16374678,          21435658])
```

```
In [91]: twit_arc_copy.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 18 columns):
tweet_id                2356 non-null int64
in_reply_to_status_id   2356 non-null int64
in_reply_to_user_id     2356 non-null int64
timestamp               2356 non-null datetime64[ns]
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 datetime64[ns]
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
floof                   2356 non-null object
pupper                  2356 non-null object
puppo                   2356 non-null object
normalize_rating        2356 non-null float64
dtypes: datetime64[ns](2), float64(3), int64(5), object(8)
memory usage: 331.4+ KB
```

```
In [92]: twit_arc_copy[twit_arc_copy['in_reply_to_status_id'].notnull()]
```

```
Out[92]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	0	0	
1	892177421306343426	0	0	
2	891815181378084864	0	0	

3	891689557279858688	0	0
4	891327558926688256	0	0
5	891087950875897856	0	0
6	890971913173991426	0	0
7	890729181411237888	0	0
8	890609185150312448	0	0
9	890240255349198849	0	0
10	890006608113172480	0	0
11	889880896479866881	0	0
12	889665388333682689	0	0
13	889638837579907072	0	0
14	889531135344209921	0	0
15	889278841981685760	0	0
16	888917238123831296	0	0
17	888804989199671297	0	0
18	888554962724278272	0	0
19	888202515573088257	0	0
20	888078434458587136	0	0
21	887705289381826560	0	0
22	887517139158093824	0	0
23	887473957103951883	0	0
24	887343217045368832	0	0
25	887101392804085760	0	0
26	886983233522544640	0	0
27	886736880519319552	0	0
28	886680336477933568	0	0
29	886366144734445568	0	0
...
2326	666411507551481857	0	0
2327	666407126856765440	0	0
2328	666396247373291520	0	0
2329	666373753744588802	0	0
2330	666362758909284353	0	0
2331	666353288456101888	0	0
2332	666345417576210432	0	0
2333	666337882303524864	0	0
2334	666293911632134144	0	0
2335	666287406224695296	0	0
2336	666273097616637952	0	0
2337	666268910803644416	0	0
2338	666104133288665088	0	0
2339	666102155909144576	0	0
2340	666099513787052032	0	0
2341	666094000022159362	0	0
2342	666082916733198337	0	0
2343	666073100786774016	0	0
2344	666071193221509120	0	0
2345	666063827256086533	0	0

2346	666058600524156928	0	0
2347	666057090499244032	0	0
2348	666055525042405380	0	0
2349	666051853826850816	0	0
2350	666050758794694657	0	0
2351	666049248165822465	0	0
2352	666044226329800704	0	0
2353	666033412701032449	0	0
2354	666029285002620928	0	0
2355	666020888022790149	0	0

	timestamp	source \
0	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone" r...
1	2017-08-01 00:17:27	<a href="http://twitter.com/download/iphone" r...
2	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...
3	2017-07-30 15:58:51	<a href="http://twitter.com/download/iphone" r...
4	2017-07-29 16:00:24	<a href="http://twitter.com/download/iphone" r...
5	2017-07-29 00:08:17	<a href="http://twitter.com/download/iphone" r...
6	2017-07-28 16:27:12	<a href="http://twitter.com/download/iphone" r...
7	2017-07-28 00:22:40	<a href="http://twitter.com/download/iphone" r...
8	2017-07-27 16:25:51	<a href="http://twitter.com/download/iphone" r...
9	2017-07-26 15:59:51	<a href="http://twitter.com/download/iphone" r...
10	2017-07-26 00:31:25	<a href="http://twitter.com/download/iphone" r...
11	2017-07-25 16:11:53	<a href="http://twitter.com/download/iphone" r...
12	2017-07-25 01:55:32	<a href="http://twitter.com/download/iphone" r...
13	2017-07-25 00:10:02	<a href="http://twitter.com/download/iphone" r...
14	2017-07-24 17:02:04	<a href="http://twitter.com/download/iphone" r...
15	2017-07-24 00:19:32	<a href="http://twitter.com/download/iphone" r...
16	2017-07-23 00:22:39	<a href="http://twitter.com/download/iphone" r...
17	2017-07-22 16:56:37	<a href="http://twitter.com/download/iphone" r...
18	2017-07-22 00:23:06	<a href="http://twitter.com/download/iphone" r...
19	2017-07-21 01:02:36	<a href="http://twitter.com/download/iphone" r...
20	2017-07-20 16:49:33	<a href="http://twitter.com/download/iphone" r...
21	2017-07-19 16:06:48	<a href="http://twitter.com/download/iphone" r...
22	2017-07-19 03:39:09	<a href="http://twitter.com/download/iphone" r...
23	2017-07-19 00:47:34	<a href="http://twitter.com/download/iphone" r...
24	2017-07-18 16:08:03	<a href="http://twitter.com/download/iphone" r...
25	2017-07-18 00:07:08	<a href="http://twitter.com/download/iphone" r...
26	2017-07-17 16:17:36	<a href="http://twitter.com/download/iphone" r...
27	2017-07-16 23:58:41	<a href="http://twitter.com/download/iphone" r...
28	2017-07-16 20:14:00	<a href="http://twitter.com/download/iphone" r...
29	2017-07-15 23:25:31	<a href="http://twitter.com/download/iphone" r...
...
2326	2015-11-17 00:24:19	<a href="http://twitter.com/download/iphone" r...
2327	2015-11-17 00:06:54	<a href="http://twitter.com/download/iphone" r...
2328	2015-11-16 23:23:41	<a href="http://twitter.com/download/iphone" r...
2329	2015-11-16 21:54:18	<a href="http://twitter.com/download/iphone" r...
2330	2015-11-16 21:10:36	<a href="http://twitter.com/download/iphone" r...

2331 2015-11-16 20:32:58 <a href="http://twitter.com/download/iphone" r...
2332 2015-11-16 20:01:42 <a href="http://twitter.com/download/iphone" r...
2333 2015-11-16 19:31:45 <a href="http://twitter.com/download/iphone" r...
2334 2015-11-16 16:37:02 <a href="http://twitter.com/download/iphone" r...
2335 2015-11-16 16:11:11 <a href="http://twitter.com/download/iphone" r...
2336 2015-11-16 15:14:19 <a href="http://twitter.com/download/iphone" r...
2337 2015-11-16 14:57:41 <a href="http://twitter.com/download/iphone" r...
2338 2015-11-16 04:02:55 <a href="http://twitter.com/download/iphone" r...
2339 2015-11-16 03:55:04 <a href="http://twitter.com/download/iphone" r...
2340 2015-11-16 03:44:34 <a href="http://twitter.com/download/iphone" r...
2341 2015-11-16 03:22:39 <a href="http://twitter.com/download/iphone" r...
2342 2015-11-16 02:38:37 <a href="http://twitter.com/download/iphone" r...
2343 2015-11-16 01:59:36 <a href="http://twitter.com/download/iphone" r...
2344 2015-11-16 01:52:02 <a href="http://twitter.com/download/iphone" r...
2345 2015-11-16 01:22:45 <a href="http://twitter.com/download/iphone" r...
2346 2015-11-16 01:01:59 <a href="http://twitter.com/download/iphone" r...
2347 2015-11-16 00:55:59 <a href="http://twitter.com/download/iphone" r...
2348 2015-11-16 00:49:46 <a href="http://twitter.com/download/iphone" r...
2349 2015-11-16 00:35:11 <a href="http://twitter.com/download/iphone" r...
2350 2015-11-16 00:30:50 <a href="http://twitter.com/download/iphone" r...
2351 2015-11-16 00:24:50 <a href="http://twitter.com/download/iphone" r...
2352 2015-11-16 00:04:52 <a href="http://twitter.com/download/iphone" r...
2353 2015-11-15 23:21:54 <a href="http://twitter.com/download/iphone" r...
2354 2015-11-15 23:05:30 <a href="http://twitter.com/download/iphone" r...
2355 2015-11-15 22:32:08 <a href="http://twitter.com/download/iphone" r...

	text	retweeted_status_id \
0	This is Phineas. He's a mystical boy. Only eve...	NaN
1	This is Tilly. She's just checking pup on you...	NaN
2	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN
5	Here we have a majestic great white breaching ...	NaN
6	Meet Jax. He enjoys ice cream so much he gets ...	NaN
7	When you watch your owner call another dog a g...	NaN
8	This is Zoey. She doesn't want to be one of th...	NaN
9	This is Cassie. She is a college pup. Studying...	NaN
10	This is Koda. He is a South Australian decksha...	NaN
11	This is Bruno. He is a service shark. Only get...	NaN
12	Here's a puppo that seems to be on the fence a...	NaN
13	This is Ted. He does his best. Sometimes that'...	NaN
14	This is Stuart. He's sporting his favorite fan...	NaN
15	This is Oliver. You're witnessing one of his m...	NaN
16	This is Jim. He found a fren. Taught him how t...	NaN
17	This is Zeke. He has a new stick. Very proud o...	NaN
18	This is Ralphus. He's powering up. Attempting ...	NaN
19	RT @dog_rates: This is Canela. She attempted s...	8.874740e+17
20	This is Gerald. He was just told he didn't get...	NaN

21	This is Jeffrey. He has a monopoly on the pool...	NaN
22	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN
23	This is Canela. She attempted some fancy porch...	NaN
24	You may not have known you needed to see this ...	NaN
25	This... is a Jubilant Antarctic House Bear. We...	NaN
26	This is Maya. She's very shy. Rarely leaves he...	NaN
27	This is Mingus. He's a wonderful father to his...	NaN
28	This is Derek. He's late for a dog meeting. 13...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
...
2326	This is quite the dog. Gets really excited whe...	NaN
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN
2328	Oh goodness. A super rare northeast Qdoba kang...	NaN
2329	Those are sunglasses and a jean jacket. 11/10 ...	NaN
2330	Unique dog here. Very small. Lives in containe...	NaN
2331	Here we have a mixed Asiago from the Galápagos...	NaN
2332	Look at this jokester thinking seat belt laws ...	NaN
2333	This is an extremely rare horned Parthenon. No...	NaN
2334	This is a funny dog. Weird toes. Won't come do...	NaN
2335	This is an Albanian 3 1/2 legged Episcopalian...	NaN
2336	Can take selfies 11/10 https://t.co/ws2AMaWpPW	NaN
2337	Very concerned about fellow dog trapped in com...	NaN
2338	Not familiar with this breed. No tail (weird)...	NaN
2339	Oh my. Here you are seeing an Adobe Setter giv...	NaN
2340	Can stand on stump for what seems like a while...	NaN
2341	This appears to be a Mongolian Presbyterian mi...	NaN
2342	Here we have a well-established sunblockerspan...	NaN
2343	Let's hope this flight isn't Malaysian (lol). ...	NaN
2344	Here we have a northern speckled Rhododendron...	NaN
2345	This is the happiest dog you will ever see. Ve...	NaN
2346	Here is the Rand Paul of retrievers folks! He'...	NaN
2347	My oh my. This is a rare blond Canadian terrie...	NaN
2348	Here is a Siberian heavily armored polar bear ...	NaN
2349	This is an odd dog. Hard on the outside but lo...	NaN
2350	This is a truly beautiful English Wilson Staff...	NaN
2351	Here we have a 1949 1st generation vulpix. Enj...	NaN
2352	This is a purebred Piers Morgan. Loves to Netf...	NaN
2353	Here is a very happy pup. Big fan of well-main...	NaN
2354	This is a western brown Mitsubishi terrier. Up...	NaN
2355	Here we have a Japanese Irish Setter. Lost eye...	NaN

	retweeted_status_user_id	retweeted_status_timestamp \
0	NaN	NaT
1	NaN	NaT
2	NaN	NaT
3	NaN	NaT
4	NaN	NaT
5	NaN	NaT

6	NaN	NaN
7	NaN	NaN
8	NaN	NaN
9	NaN	NaN
10	NaN	NaN
11	NaN	NaN
12	NaN	NaN
13	NaN	NaN
14	NaN	NaN
15	NaN	NaN
16	NaN	NaN
17	NaN	NaN
18	NaN	NaN
19	4.196984e+09	2017-07-19 00:47:34
20	NaN	NaN
21	NaN	NaN
22	NaN	NaN
23	NaN	NaN
24	NaN	NaN
25	NaN	NaN
26	NaN	NaN
27	NaN	NaN
28	NaN	NaN
29	NaN	NaN
...
2326	NaN	NaN
2327	NaN	NaN
2328	NaN	NaN
2329	NaN	NaN
2330	NaN	NaN
2331	NaN	NaN
2332	NaN	NaN
2333	NaN	NaN
2334	NaN	NaN
2335	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
2345	NaN	NaN
2346	NaN	NaN
2347	NaN	NaN
2348	NaN	NaN

2349	NaN	NaT
2350	NaN	NaT
2351	NaN	NaT
2352	NaN	NaT
2353	NaN	NaT
2354	NaN	NaT
2355	NaN	NaT

	expanded_urls	rating_numerator	\
0	https://twitter.com/dog_rates/status/892420643...	13	
1	https://twitter.com/dog_rates/status/892177421...	13	
2	https://twitter.com/dog_rates/status/891815181...	12	
3	https://twitter.com/dog_rates/status/891689557...	13	
4	https://twitter.com/dog_rates/status/891327558...	12	
5	https://twitter.com/dog_rates/status/891087950...	13	
6	https://gofundme.com/ydvmve-surgery-for-jax,ht...	13	
7	https://twitter.com/dog_rates/status/890729181...	13	
8	https://twitter.com/dog_rates/status/890609185...	13	
9	https://twitter.com/dog_rates/status/890240255...	14	
10	https://twitter.com/dog_rates/status/890006608...	13	
11	https://twitter.com/dog_rates/status/889880896...	13	
12	https://twitter.com/dog_rates/status/889665388...	13	
13	https://twitter.com/dog_rates/status/889638837...	12	
14	https://twitter.com/dog_rates/status/889531135...	13	
15	https://twitter.com/dog_rates/status/889278841...	13	
16	https://twitter.com/dog_rates/status/888917238...	12	
17	https://twitter.com/dog_rates/status/888804989...	13	
18	https://twitter.com/dog_rates/status/888554962...	13	
19	https://twitter.com/dog_rates/status/887473957...	13	
20	https://twitter.com/dog_rates/status/888078434...	12	
21	https://twitter.com/dog_rates/status/887705289...	13	
22	https://twitter.com/dog_rates/status/887517139...	14	
23	https://twitter.com/dog_rates/status/887473957...	13	
24	https://twitter.com/dog_rates/status/887343217...	13	
25	https://twitter.com/dog_rates/status/887101392...	12	
26	https://twitter.com/dog_rates/status/886983233...	13	
27	https://www.gofundme.com/mingusneedsus,https:/...	13	
28	https://twitter.com/dog_rates/status/886680336...	13	
29	https://twitter.com/dog_rates/status/886366144...	12	
...	
2326	https://twitter.com/dog_rates/status/666411507...	2	
2327	https://twitter.com/dog_rates/status/666407126...	7	
2328	https://twitter.com/dog_rates/status/666396247...	9	
2329	https://twitter.com/dog_rates/status/666373753...	11	
2330	https://twitter.com/dog_rates/status/666362758...	6	
2331	https://twitter.com/dog_rates/status/666353288...	8	
2332	https://twitter.com/dog_rates/status/666345417...	10	
2333	https://twitter.com/dog_rates/status/666337882...	9	

2334	https://twitter.com/dog_rates/status/666293911...	3
2335	https://twitter.com/dog_rates/status/666287406...	1
2336	https://twitter.com/dog_rates/status/666273097...	11
2337	https://twitter.com/dog_rates/status/666268910...	10
2338	https://twitter.com/dog_rates/status/666104133...	1
2339	https://twitter.com/dog_rates/status/666102155...	11
2340	https://twitter.com/dog_rates/status/666099513...	8
2341	https://twitter.com/dog_rates/status/666094000...	9
2342	https://twitter.com/dog_rates/status/666082916...	6
2343	https://twitter.com/dog_rates/status/666073100...	10
2344	https://twitter.com/dog_rates/status/666071193...	9
2345	https://twitter.com/dog_rates/status/666063827...	10
2346	https://twitter.com/dog_rates/status/666058600...	8
2347	https://twitter.com/dog_rates/status/666057090...	9
2348	https://twitter.com/dog_rates/status/666055525...	10
2349	https://twitter.com/dog_rates/status/666051853...	2
2350	https://twitter.com/dog_rates/status/666050758...	10
2351	https://twitter.com/dog_rates/status/666049248...	5
2352	https://twitter.com/dog_rates/status/666044226...	6
2353	https://twitter.com/dog_rates/status/666033412...	9
2354	https://twitter.com/dog_rates/status/666029285...	7
2355	https://twitter.com/dog_rates/status/666020888...	8

	rating_denominator	name	doggo	floof	pupper	puppo	\
0	10	Phineas	None	None	None	None	
1	10	Tilly	None	None	None	None	
2	10	Archie	None	None	None	None	
3	10	Darla	None	None	None	None	
4	10	Franklin	None	None	None	None	
5	10	None	None	None	None	None	
6	10	Jax	None	None	None	None	
7	10	None	None	None	None	None	
8	10	Zoey	None	None	None	None	
9	10	Cassie	doggo	None	None	None	
10	10	Koda	None	None	None	None	
11	10	Bruno	None	None	None	None	
12	10	None	None	None	None	puppo	
13	10	Ted	None	None	None	None	
14	10	Stuart	None	None	None	puppo	
15	10	Oliver	None	None	None	None	
16	10	Jim	None	None	None	None	
17	10	Zeke	None	None	None	None	
18	10	Ralphus	None	None	None	None	
19	10	Canela	None	None	None	None	
20	10	Gerald	None	None	None	None	
21	10	Jeffrey	None	None	None	None	
22	10	such	None	None	None	None	
23	10	Canela	None	None	None	None	

24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None
2343	10	None	None	None	None	None
2344	10	None	None	None	None	None
2345	10	the	None	None	None	None
2346	10	the	None	None	None	None
2347	10	a	None	None	None	None
2348	10	a	None	None	None	None
2349	10	an	None	None	None	None
2350	10	a	None	None	None	None
2351	10	None	None	None	None	None
2352	10	a	None	None	None	None
2353	10	a	None	None	None	None
2354	10	a	None	None	None	None
2355	10	None	None	None	None	None

	normalize_rating
0	1.3
1	1.3
2	1.2
3	1.3
4	1.2
5	1.3
6	1.3
7	1.3
8	1.3

9	1.4
10	1.3
11	1.3
12	1.3
13	1.2
14	1.3
15	1.3
16	1.2
17	1.3
18	1.3
19	1.3
20	1.2
21	1.3
22	1.4
23	1.3
24	1.3
25	1.2
26	1.3
27	1.3
28	1.3
29	1.2
...	...
2326	0.2
2327	0.7
2328	0.9
2329	1.1
2330	0.6
2331	0.8
2332	1.0
2333	0.9
2334	0.3
2335	0.5
2336	1.1
2337	1.0
2338	0.1
2339	1.1
2340	0.8
2341	0.9
2342	0.6
2343	1.0
2344	0.9
2345	1.0
2346	0.8
2347	0.9
2348	1.0
2349	0.2
2350	1.0
2351	0.5

2352	0.6
2353	0.9
2354	0.7
2355	0.8

[2356 rows x 18 columns]

0.6.14 rating_numerator and rating_denominator in row 313 is 13/10 instead of 960/0

0.6.15 Define

- change rating_numerator and rating_denominator in row 313 change into 13 and 10

0.6.16 Code

```
In [105]: # find row of inf normalized_rating
          twit_arc_copy.query('normalize_rating==inf').index
```

```
Out[105]: Int64Index([313], dtype='int64')
```

```
In [106]: twit_arc_copy.rating_numerator[313] = 13
          twit_arc_copy.rating_denominator[313] = 10
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: 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#
"""Entry point for launching an IPython kernel.
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:2: 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#
```

```
In [107]: twit_arc_copy.normalize_rating[313] = twit_arc_copy.rating_numerator[313]/twit_arc_copy.rating_denominator[313]
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: 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#
"""Entry point for launching an IPython kernel.
```

```
In [108]: # recheck denominator condition
          twit_arc_copy[twit_arc_copy['rating_denominator']==0]
```

```
Out[108]: Empty DataFrame
          Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text]
          Index: []
```

0.6.17 Test

```
In [109]: twit_arc_copy.describe()
```

```
Out[109]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
count	2.356000e+03	2.356000e+03	2.356000e+03	
mean	7.427716e+17	2.468150e+16	6.668307e+14	
std	6.856705e+16	1.341142e+17	2.293821e+16	
min	6.660209e+17	0.000000e+00	0.000000e+00	
25%	6.783989e+17	0.000000e+00	0.000000e+00	
50%	7.196279e+17	0.000000e+00	0.000000e+00	
75%	7.993373e+17	0.000000e+00	0.000000e+00	
max	8.924206e+17	8.862664e+17	8.405479e+17	

	retweeted_status_id	retweeted_status_user_id	rating_numerator	\
count	1.810000e+02	1.810000e+02	2356.000000	
mean	7.720400e+17	1.241698e+16	12.724533	
std	6.236928e+16	9.599254e+16	41.518626	
min	6.661041e+17	7.832140e+05	0.000000	
25%	7.186315e+17	4.196984e+09	10.000000	
50%	7.804657e+17	4.196984e+09	11.000000	
75%	8.203146e+17	4.196984e+09	12.000000	
max	8.874740e+17	7.874618e+17	1776.000000	

	rating_denominator	normalize_rating
count	2356.000000	2356.000000
mean	10.459677	1.222065
std	6.741801	4.082618
min	2.000000	0.000000
25%	10.000000	1.000000
50%	10.000000	1.100000
75%	10.000000	1.200000
max	170.000000	177.600000

0.6.18 Retweet data in dataframe must be deleted

0.6.19 Define

- find retweet data in dataframe, get index and delete it

0.6.20 Code

```
In [110]: #find Retweet tweet id
rt_tweet = twit_arc_copy[twit_arc_copy.retweeted_status_id.notnull()]['tweet_id'].values
len(rt_tweet)
```

```
Out[110]: 181
```

```
In [111]: test_cp = twit_arc_copy.copy()
```

```
In [112]: #delete rows based on tweet id
         for x in rt_tweet:
             rt_ind = twit_arc_copy[twit_arc_copy['tweet_id']== x].index
             twit_arc_copy.drop(rt_ind, inplace=True)
```

0.6.21 Test

```
In [113]: twit_arc_copy[twit_arc_copy.retweeted_status_id.notnull()]
```

```
Out[113]: Empty DataFrame
          Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text]
          Index: []
```

0.6.22 dog prediction will be more clear by replacing '_' with space

0.6.23 Define

- replace '_' with ' ' in column p_1, p_2 and p_3

0.6.24 Code

```
In [114]: # col name checking
         list(img_pred_copy.columns)
```

```
Out[114]: ['tweet_id',
          'jpg_url',
          'img_num',
          'p1',
          'p1_conf',
          'p1_dog',
          'p2',
          'p2_conf',
          'p2_dog',
          'p3',
          'p3_conf',
          'p3_dog']
```

```
In [115]: # change '_' and '-' with space
         img_pred_copy['p1'] = img_pred_copy['p1'].str.replace('_', ' ', regex=True)
         img_pred_copy['p2'] = img_pred_copy['p2'].str.replace('_', ' ', regex=True)
         img_pred_copy['p3'] = img_pred_copy['p3'].str.replace('_', ' ', regex=True)
         img_pred_copy['p1'] = img_pred_copy['p1'].str.replace('-', ' ', regex=True)
         img_pred_copy['p2'] = img_pred_copy['p2'].str.replace('-', ' ', regex=True)
         img_pred_copy['p3'] = img_pred_copy['p3'].str.replace('-', ' ', regex=True)
```

0.6.25 Test

```
In [116]: img_pred_copy
```


Out[116]:	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/CT4521TWwAAEvMyu.jpg
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
5	666050758794694657	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6	666051853826850816	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7	666055525042405380	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8	666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9	666058600524156928	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10	666063827256086533	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11	666071193221509120	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
12	666073100786774016	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13	666082916733198337	https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg
14	666094000022159362	https://pbs.twimg.com/media/CT5w9gUW4AAAsBNN.jpg
15	666099513787052032	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16	666102155909144576	https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg
17	666104133288665088	https://pbs.twimg.com/media/CT56LSZWAA1Jj2.jpg
18	666268910803644416	https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
19	666273097616637952	https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
20	666287406224695296	https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
21	666293911632134144	https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22	666337882303524864	https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg
23	666345417576210432	https://pbs.twimg.com/media/CT9Vn7PWAA_ZCM.jpg
24	666353288456101888	https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25	666362758909284353	https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg
26	666373753744588802	https://pbs.twimg.com/media/CT9vZEYWUAA1Z05.jpg
27	666396247373291520	https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28	666407126856765440	https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.jpg
29	666411507551481857	https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...
2045	886366144734445568	https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046	886680336477933568	https://pbs.twimg.com/media/DE4fEDzWAAAYHMM.jpg
2047	886736880519319552	https://pbs.twimg.com/media/DE5Se8FXcAAAJFx4.jpg
2048	886983233522544640	https://pbs.twimg.com/media/DE8yicJW0AAAABJ.jpg
2049	887101392804085760	https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050	887343217045368832	https://pbs.twimg.com/ext_tw_video_thumb/88734...
2051	887473957103951883	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2052	887517139158093824	https://pbs.twimg.com/ext_tw_video_thumb/88751...
2053	887705289381826560	https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054	888078434458587136	https://pbs.twimg.com/media/DFMWn56WsAAKA7B.jpg
2055	888202515573088257	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056	888554962724278272	https://pbs.twimg.com/media/DFTH_0-UQAAACu20.jpg
2057	888804989199671297	https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058	888917238123831296	https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg
2059	889278841981685760	https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060	889531135344209921	https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg

2061	889638837579907072	https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062	889665388333682689	https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063	889880896479866881	https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg
2064	890006608113172480	https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065	890240255349198849	https://pbs.twimg.com/media/DFrEyVuWAAA03t9.jpg
2066	890609185150312448	https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067	890729181411237888	https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068	890971913173991426	https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg
2069	891087950875897856	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070	891327558926688256	https://pbs.twimg.com/media/DF6hr6BUMAAZzgT.jpg
2071	891689557279858688	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072	891815181378084864	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073	892177421306343426	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074	892420643555336193	https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

	img_num		p1	p1_conf	p1_dog	\
0	1	Welsh springer spaniel	0.465074	True		
1	1	redbone	0.506826	True		
2	1	German shepherd	0.596461	True		
3	1	Rhodesian ridgeback	0.408143	True		
4	1	miniature pinscher	0.560311	True		
5	1	Bernese mountain dog	0.651137	True		
6	1	box turtle	0.933012	False		
7	1	chow	0.692517	True		
8	1	shopping cart	0.962465	False		
9	1	miniature poodle	0.201493	True		
10	1	golden retriever	0.775930	True		
11	1	Gordon setter	0.503672	True		
12	1	Walker hound	0.260857	True		
13	1	pug	0.489814	True		
14	1	bloodhound	0.195217	True		
15	1	Lhasa	0.582330	True		
16	1	English setter	0.298617	True		
17	1	hen	0.965932	False		
18	1	desktop computer	0.086502	False		
19	1	Italian greyhound	0.176053	True		
20	1	Maltese dog	0.857531	True		
21	1	three toed sloth	0.914671	False		
22	1	ox	0.416669	False		
23	1	golden retriever	0.858744	True		
24	1	malamute	0.336874	True		
25	1	guinea pig	0.996496	False		
26	1	soft coated wheaten terrier	0.326467	True		
27	1	Chihuahua	0.978108	True		
28	1	black and tan coonhound	0.529139	True		
29	1	coho	0.404640	False		
...		
2045	1	French bulldog	0.999201	True		

2046	1	convertible	0.738995	False
2047	1	kuvasz	0.309706	True
2048	2	Chihuahua	0.793469	True
2049	1	Samoyed	0.733942	True
2050	1	Mexican hairless	0.330741	True
2051	2	Pembroke	0.809197	True
2052	1	limousine	0.130432	False
2053	1	basset	0.821664	True
2054	1	French bulldog	0.995026	True
2055	2	Pembroke	0.809197	True
2056	3	Siberian husky	0.700377	True
2057	1	golden retriever	0.469760	True
2058	1	golden retriever	0.714719	True
2059	1	whippet	0.626152	True
2060	1	golden retriever	0.953442	True
2061	1	French bulldog	0.991650	True
2062	1	Pembroke	0.966327	True
2063	1	French bulldog	0.377417	True
2064	1	Samoyed	0.957979	True
2065	1	Pembroke	0.511319	True
2066	1	Irish terrier	0.487574	True
2067	2	Pomeranian	0.566142	True
2068	1	Appenzeller	0.341703	True
2069	1	Chesapeake Bay retriever	0.425595	True
2070	2	basset	0.555712	True
2071	1	paper towel	0.170278	False
2072	1	Chihuahua	0.716012	True
2073	1	Chihuahua	0.323581	True
2074	1	orange	0.097049	False

	p2	p2_conf	p2_dog	p3 \
0	collie	0.156665	True	Shetland sheepdog
1	miniature pinscher	0.074192	True	Rhodesian ridgeback
2	malinois	0.138584	True	bloodhound
3	redbone	0.360687	True	miniature pinscher
4	Rottweiler	0.243682	True	Doberman
5	English springer	0.263788	True	Greater Swiss Mountain dog
6	mud turtle	0.045885	False	terrapin
7	Tibetan mastiff	0.058279	True	fur coat
8	shopping basket	0.014594	False	golden retriever
9	komondor	0.192305	True	soft coated wheaten terrier
10	Tibetan mastiff	0.093718	True	Labrador retriever
11	Yorkshire terrier	0.174201	True	Pekinese
12	English foxhound	0.175382	True	Ibizan hound
13	bull mastiff	0.404722	True	French bulldog
14	German shepherd	0.078260	True	malinois
15	Shih Tzu	0.166192	True	Dandie Dinmont
16	Newfoundland	0.149842	True	borzoi

17	cock	0.033919	False	partridge
18	desk	0.085547	False	bookcase
19	toy terrier	0.111884	True	basenji
20	toy poodle	0.063064	True	miniature poodle
21	otter	0.015250	False	great grey owl
22	Newfoundland	0.278407	True	groenendael
23	Chesapeake Bay retriever	0.054787	True	Labrador retriever
24	Siberian husky	0.147655	True	Eskimo dog
25	skunk	0.002402	False	hamster
26	Afghan hound	0.259551	True	briard
27	toy terrier	0.009397	True	papillon
28	bloodhound	0.244220	True	flat coated retriever
29	barracouta	0.271485	False	gar
...
2045	Chihuahua	0.000361	True	Boston bull
2046	sports car	0.139952	False	car wheel
2047	Great Pyrenees	0.186136	True	Dandie Dinmont
2048	toy terrier	0.143528	True	can opener
2049	Eskimo dog	0.035029	True	Staffordshire bullterrier
2050	sea lion	0.275645	False	Weimaraner
2051	Rhodesian ridgeback	0.054950	True	beagle
2052	tow truck	0.029175	False	shopping cart
2053	redbone	0.087582	True	Weimaraner
2054	pug	0.000932	True	bull mastiff
2055	Rhodesian ridgeback	0.054950	True	beagle
2056	Eskimo dog	0.166511	True	malamute
2057	Labrador retriever	0.184172	True	English setter
2058	Tibetan mastiff	0.120184	True	Labrador retriever
2059	borzoi	0.194742	True	Saluki
2060	Labrador retriever	0.013834	True	redbone
2061	boxer	0.002129	True	Staffordshire bullterrier
2062	Cardigan	0.027356	True	basenji
2063	Labrador retriever	0.151317	True	muzzle
2064	Pomeranian	0.013884	True	chow
2065	Cardigan	0.451038	True	Chihuahua
2066	Irish setter	0.193054	True	Chesapeake Bay retriever
2067	Eskimo dog	0.178406	True	Pembroke
2068	Border collie	0.199287	True	ice lolly
2069	Irish terrier	0.116317	True	Indian elephant
2070	English springer	0.225770	True	German short haired pointer
2071	Labrador retriever	0.168086	True	spatula
2072	malamute	0.078253	True	kelpie
2073	Pekinese	0.090647	True	papillon
2074	bagel	0.085851	False	banana

	p3_conf	p3_dog
0	0.061428	True
1	0.072010	True

2	0.116197	True
3	0.222752	True
4	0.154629	True
5	0.016199	True
6	0.017885	False
7	0.054449	False
8	0.007959	True
9	0.082086	True
10	0.072427	True
11	0.109454	True
12	0.097471	True
13	0.048960	True
14	0.075628	True
15	0.089688	True
16	0.133649	True
17	0.000052	False
18	0.079480	False
19	0.111152	True
20	0.025581	True
21	0.013207	False
22	0.102643	True
23	0.014241	True
24	0.093412	True
25	0.000461	False
26	0.206803	True
27	0.004577	True
28	0.173810	True
29	0.189945	False
...
2045	0.000076	True
2046	0.044173	False
2047	0.086346	True
2048	0.032253	False
2049	0.029705	True
2050	0.134203	True
2051	0.038915	True
2052	0.026321	False
2053	0.026236	True
2054	0.000903	True
2055	0.038915	True
2056	0.111411	True
2057	0.073482	True
2058	0.105506	True
2059	0.027351	True
2060	0.007958	True
2061	0.001498	True
2062	0.004633	True
2063	0.082981	False

2064	0.008167	True
2065	0.029248	True
2066	0.118184	True
2067	0.076507	True
2068	0.193548	False
2069	0.076902	False
2070	0.175219	True
2071	0.040836	False
2072	0.031379	True
2073	0.068957	True
2074	0.076110	False

[2075 rows x 12 columns]

0.6.26 duplication data stored with different tweet_id

0.6.27 Define

- find duplicate data and drop it from dataframe

0.6.28 Code

```
In [117]: #delete rows based on tweet id
          for x in rt_tweet:
              rt_ind = img_pred_copy[img_pred_copy['tweet_id']== x].index
              img_pred_copy.drop(rt_ind, inplace=True)
```

0.6.29 Test

```
In [118]: # check if it still duplication jpg_url or retweeted tweet in dataframe
          img_pred_copy[img_pred_copy['jpg_url'].duplicated()]
```

Out[118]: Empty DataFrame

Columns: [tweet_id, jpg_url, img_num, p1, p1_conf, p1_dog, p2, p2_conf, p2_dog, p3, p3_conf, p3_dog]
Index: []

0.6.30 name of all dogs predictions are variate with upper and lower case

0.6.31 Define

Change all character in p1,p2, and p3 in lowercase

0.6.32 Code

```
In [119]: img_pred_copy['p1'] = img_pred_copy['p1'].str.lower()
          img_pred_copy['p2'] = img_pred_copy['p2'].str.lower()
          img_pred_copy['p3'] = img_pred_copy['p3'].str.lower()
```

0.6.33 Test

```
In [120]: img_pred_copy[['p1', 'p2', 'p3']]
```

```
Out[120]:
```

	p1	p2 \
0	welsh springer spaniel	collie
1	redbone	miniature pinscher
2	german shepherd	malinois
3	rhodesian ridgeback	redbone
4	miniature pinscher	rottweiler
5	bernese mountain dog	english springer
6	box turtle	mud turtle
7	chow	tibetan mastiff
8	shopping cart	shopping basket
9	miniature poodle	komondor
10	golden retriever	tibetan mastiff
11	gordon setter	yorkshire terrier
12	walker hound	english foxhound
13	pug	bull mastiff
14	bloodhound	german shepherd
15	lhasa	shih tzu
16	english setter	newfoundland
17	hen	cock
18	desktop computer	desk
19	italian greyhound	toy terrier
20	maltese dog	toy poodle
21	three toed sloth	otter
22	ox	newfoundland
23	golden retriever	chesapeake bay retriever
24	malamute	siberian husky
25	guinea pig	skunk
26	soft coated wheaten terrier	afghan hound
27	chihuahua	toy terrier
28	black and tan coonhound	bloodhound
29	coho	barracouta
...
2044	pug	shower cap
2045	french bulldog	chihuahua
2046	convertible	sports car
2047	kuvasz	great pyrenees
2048	chihuahua	toy terrier
2049	samoyed	eskimo dog
2050	mexican hairless	sea lion
2051	pembroke	rhodesian ridgeback
2052	limousine	tow truck
2053	basset	redbone
2054	french bulldog	pug
2056	siberian husky	eskimo dog
2057	golden retriever	labrador retriever

2058	golden retriever	tibetan mastiff
2059	whippet	borzoi
2060	golden retriever	labrador retriever
2061	french bulldog	boxer
2062	pembroke	cardigan
2063	french bulldog	labrador retriever
2064	samoyed	pomeranian
2065	pembroke	cardigan
2066	irish terrier	irish setter
2067	pomeranian	eskimo dog
2068	appenzeller	border collie
2069	chesapeake bay retriever	irish terrier
2070	basset	english springer
2071	paper towel	labrador retriever
2072	chihuahua	malamute
2073	chihuahua	pekinese
2074	orange	bagel

p3

0	shetland sheepdog
1	rhodesian ridgeback
2	bloodhound
3	miniature pinscher
4	doberman
5	greater swiss mountain dog
6	terrapin
7	fur coat
8	golden retriever
9	soft coated wheaten terrier
10	labrador retriever
11	pekinese
12	ibizan hound
13	french bulldog
14	malinois
15	dandie dinmont
16	borzoi
17	partridge
18	bookcase
19	basenji
20	miniature poodle
21	great grey owl
22	groenendael
23	labrador retriever
24	eskimo dog
25	hamster
26	briard
27	papillon
28	flat coated retriever


```

29                gar
...                ...
2044             siamese cat
2045             boston bull
2046             car wheel
2047             dandie dinmont
2048             can opener
2049  staffordshire bullterrier
2050             weimaraner
2051             beagle
2052             shopping cart
2053             weimaraner
2054             bull mastiff
2056             malamute
2057             english setter
2058             labrador retriever
2059             saluki
2060             redbone
2061  staffordshire bullterrier
2062             basenji
2063             muzzle
2064             chow
2065             chihuahua
2066  chesapeake bay retriever
2067             pembroke
2068             ice lolly
2069             indian elephant
2070  german short haired pointer
2071             spatula
2072             kelpie
2073             papillon
2074             banana

```

```
[1994 rows x 3 columns]
```

0.6.34 Test

0.6.35 retweeted tweets in json data

0.6.36 Define

find retweeted tweets and delete it

0.6.37 Code

```
In [121]: json_copy.head()
```

```

Out[121]:      tweet_id      full_text \
0  892420643555336193  This is Phineas. He's a mystical boy. Only eve...

```

```

1 892177421306343426 This is Tilly. She's just checking pup on you...
2 891815181378084864 This is Archie. He is a rare Norwegian Pouncin...
3 891689557279858688 This is Darla. She commenced a snooze mid meal...
4 891327558926688256 This is Franklin. He would like you to stop ca...

```

	in_reply_to_status_id	in_reply_to_user_id	retweet_count	favorite_count	\
0	NaN	NaN	7740	36366	
1	NaN	NaN	5730	31326	
2	NaN	NaN	3791	23598	
3	NaN	NaN	7902	39669	
4	NaN	NaN	8524	37867	

	favorite_retweet_ratio
0	4.698450
1	5.467016
2	6.224743
3	5.020121
4	4.442398

```

In [122]: #delete rows based on tweet id
          for x in rt_tweet:
              rt_ind = json_copy[json_copy['tweet_id']== x].index
              json_copy.drop(rt_ind, inplace=True)

```

0.6.38 Test

```

In [123]: # check if it still 'RT @' in full_text
          cek_rt_cp = json_copy.full_text.str.extract(r'(RT @.*)')
          cek_rt_cp[cek_rt_cp[0].notnull()]

```

```

Out[123]: Empty DataFrame
          Columns: [0]
          Index: []

```

0.7 Cleaning data in Tidy

0.7.1 timestamp has date and time in one column

0.7.2 Define

split timestamp into date and time

0.7.3 Code

```

In [124]: twit_arc_copy.timestamp

```

```

Out[124]: 0      2017-08-01 16:23:56
          1      2017-08-01 00:17:27
          2      2017-07-31 00:18:03
          3      2017-07-30 15:58:51

```

4	2017-07-29 16:00:24
5	2017-07-29 00:08:17
6	2017-07-28 16:27:12
7	2017-07-28 00:22:40
8	2017-07-27 16:25:51
9	2017-07-26 15:59:51
10	2017-07-26 00:31:25
11	2017-07-25 16:11:53
12	2017-07-25 01:55:32
13	2017-07-25 00:10:02
14	2017-07-24 17:02:04
15	2017-07-24 00:19:32
16	2017-07-23 00:22:39
17	2017-07-22 16:56:37
18	2017-07-22 00:23:06
20	2017-07-20 16:49:33
21	2017-07-19 16:06:48
22	2017-07-19 03:39:09
23	2017-07-19 00:47:34
24	2017-07-18 16:08:03
25	2017-07-18 00:07:08
26	2017-07-17 16:17:36
27	2017-07-16 23:58:41
28	2017-07-16 20:14:00
29	2017-07-15 23:25:31
30	2017-07-15 16:51:35
	...
2326	2015-11-17 00:24:19
2327	2015-11-17 00:06:54
2328	2015-11-16 23:23:41
2329	2015-11-16 21:54:18
2330	2015-11-16 21:10:36
2331	2015-11-16 20:32:58
2332	2015-11-16 20:01:42
2333	2015-11-16 19:31:45
2334	2015-11-16 16:37:02
2335	2015-11-16 16:11:11
2336	2015-11-16 15:14:19
2337	2015-11-16 14:57:41
2338	2015-11-16 04:02:55
2339	2015-11-16 03:55:04
2340	2015-11-16 03:44:34
2341	2015-11-16 03:22:39
2342	2015-11-16 02:38:37
2343	2015-11-16 01:59:36
2344	2015-11-16 01:52:02
2345	2015-11-16 01:22:45
2346	2015-11-16 01:01:59

```

2347    2015-11-16 00:55:59
2348    2015-11-16 00:49:46
2349    2015-11-16 00:35:11
2350    2015-11-16 00:30:50
2351    2015-11-16 00:24:50
2352    2015-11-16 00:04:52
2353    2015-11-15 23:21:54
2354    2015-11-15 23:05:30
2355    2015-11-15 22:32:08
Name: timestamp, Length: 2175, dtype: datetime64[ns]

```

```

In [125]: # extract date and time from timestamp
twit_arc_copy['date'] = pd.DatetimeIndex(twit_arc_copy['timestamp']).date
twit_arc_copy['time'] = pd.DatetimeIndex(twit_arc_copy['timestamp']).time

```

0.7.4 Test

```

In [126]: # check date and time validity
twit_arc_copy

```

```

Out[126]:
      tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
0      892420643555336193              0                  0
1      892177421306343426              0                  0
2      891815181378084864              0                  0
3      891689557279858688              0                  0
4      891327558926688256              0                  0
5      891087950875897856              0                  0
6      890971913173991426              0                  0
7      890729181411237888              0                  0
8      890609185150312448              0                  0
9      890240255349198849              0                  0
10     890006608113172480              0                  0
11     889880896479866881              0                  0
12     889665388333682689              0                  0
13     889638837579907072              0                  0
14     889531135344209921              0                  0
15     889278841981685760              0                  0
16     888917238123831296              0                  0
17     888804989199671297              0                  0
18     888554962724278272              0                  0
20     888078434458587136              0                  0
21     887705289381826560              0                  0
22     887517139158093824              0                  0
23     887473957103951883              0                  0
24     887343217045368832              0                  0
25     887101392804085760              0                  0
26     886983233522544640              0                  0
27     886736880519319552              0                  0

```

28	886680336477933568	0	0
29	886366144734445568	0	0
30	886267009285017600	886266357075128320	2281181600
...
2326	666411507551481857	0	0
2327	666407126856765440	0	0
2328	666396247373291520	0	0
2329	666373753744588802	0	0
2330	666362758909284353	0	0
2331	666353288456101888	0	0
2332	666345417576210432	0	0
2333	666337882303524864	0	0
2334	666293911632134144	0	0
2335	666287406224695296	0	0
2336	666273097616637952	0	0
2337	666268910803644416	0	0
2338	666104133288665088	0	0
2339	666102155909144576	0	0
2340	666099513787052032	0	0
2341	666094000022159362	0	0
2342	666082916733198337	0	0
2343	666073100786774016	0	0
2344	666071193221509120	0	0
2345	666063827256086533	0	0
2346	666058600524156928	0	0
2347	666057090499244032	0	0
2348	666055525042405380	0	0
2349	666051853826850816	0	0
2350	666050758794694657	0	0
2351	666049248165822465	0	0
2352	666044226329800704	0	0
2353	666033412701032449	0	0
2354	666029285002620928	0	0
2355	666020888022790149	0	0

	timestamp	source \
0	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone" r...
1	2017-08-01 00:17:27	<a href="http://twitter.com/download/iphone" r...
2	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...
3	2017-07-30 15:58:51	<a href="http://twitter.com/download/iphone" r...
4	2017-07-29 16:00:24	<a href="http://twitter.com/download/iphone" r...
5	2017-07-29 00:08:17	<a href="http://twitter.com/download/iphone" r...
6	2017-07-28 16:27:12	<a href="http://twitter.com/download/iphone" r...
7	2017-07-28 00:22:40	<a href="http://twitter.com/download/iphone" r...
8	2017-07-27 16:25:51	<a href="http://twitter.com/download/iphone" r...
9	2017-07-26 15:59:51	<a href="http://twitter.com/download/iphone" r...
10	2017-07-26 00:31:25	<a href="http://twitter.com/download/iphone" r...
11	2017-07-25 16:11:53	<a href="http://twitter.com/download/iphone" r...

12	2017-07-25 01:55:32	<a href="http://twitter.com/download/iphone" r...
13	2017-07-25 00:10:02	<a href="http://twitter.com/download/iphone" r...
14	2017-07-24 17:02:04	<a href="http://twitter.com/download/iphone" r...
15	2017-07-24 00:19:32	<a href="http://twitter.com/download/iphone" r...
16	2017-07-23 00:22:39	<a href="http://twitter.com/download/iphone" r...
17	2017-07-22 16:56:37	<a href="http://twitter.com/download/iphone" r...
18	2017-07-22 00:23:06	<a href="http://twitter.com/download/iphone" r...
20	2017-07-20 16:49:33	<a href="http://twitter.com/download/iphone" r...
21	2017-07-19 16:06:48	<a href="http://twitter.com/download/iphone" r...
22	2017-07-19 03:39:09	<a href="http://twitter.com/download/iphone" r...
23	2017-07-19 00:47:34	<a href="http://twitter.com/download/iphone" r...
24	2017-07-18 16:08:03	<a href="http://twitter.com/download/iphone" r...
25	2017-07-18 00:07:08	<a href="http://twitter.com/download/iphone" r...
26	2017-07-17 16:17:36	<a href="http://twitter.com/download/iphone" r...
27	2017-07-16 23:58:41	<a href="http://twitter.com/download/iphone" r...
28	2017-07-16 20:14:00	<a href="http://twitter.com/download/iphone" r...
29	2017-07-15 23:25:31	<a href="http://twitter.com/download/iphone" r...
30	2017-07-15 16:51:35	<a href="http://twitter.com/download/iphone" r...
...
2326	2015-11-17 00:24:19	<a href="http://twitter.com/download/iphone" r...
2327	2015-11-17 00:06:54	<a href="http://twitter.com/download/iphone" r...
2328	2015-11-16 23:23:41	<a href="http://twitter.com/download/iphone" r...
2329	2015-11-16 21:54:18	<a href="http://twitter.com/download/iphone" r...
2330	2015-11-16 21:10:36	<a href="http://twitter.com/download/iphone" r...
2331	2015-11-16 20:32:58	<a href="http://twitter.com/download/iphone" r...
2332	2015-11-16 20:01:42	<a href="http://twitter.com/download/iphone" r...
2333	2015-11-16 19:31:45	<a href="http://twitter.com/download/iphone" r...
2334	2015-11-16 16:37:02	<a href="http://twitter.com/download/iphone" r...
2335	2015-11-16 16:11:11	<a href="http://twitter.com/download/iphone" r...
2336	2015-11-16 15:14:19	<a href="http://twitter.com/download/iphone" r...
2337	2015-11-16 14:57:41	<a href="http://twitter.com/download/iphone" r...
2338	2015-11-16 04:02:55	<a href="http://twitter.com/download/iphone" r...
2339	2015-11-16 03:55:04	<a href="http://twitter.com/download/iphone" r...
2340	2015-11-16 03:44:34	<a href="http://twitter.com/download/iphone" r...
2341	2015-11-16 03:22:39	<a href="http://twitter.com/download/iphone" r...
2342	2015-11-16 02:38:37	<a href="http://twitter.com/download/iphone" r...
2343	2015-11-16 01:59:36	<a href="http://twitter.com/download/iphone" r...
2344	2015-11-16 01:52:02	<a href="http://twitter.com/download/iphone" r...
2345	2015-11-16 01:22:45	<a href="http://twitter.com/download/iphone" r...
2346	2015-11-16 01:01:59	<a href="http://twitter.com/download/iphone" r...
2347	2015-11-16 00:55:59	<a href="http://twitter.com/download/iphone" r...
2348	2015-11-16 00:49:46	<a href="http://twitter.com/download/iphone" r...
2349	2015-11-16 00:35:11	<a href="http://twitter.com/download/iphone" r...
2350	2015-11-16 00:30:50	<a href="http://twitter.com/download/iphone" r...
2351	2015-11-16 00:24:50	<a href="http://twitter.com/download/iphone" r...
2352	2015-11-16 00:04:52	<a href="http://twitter.com/download/iphone" r...
2353	2015-11-15 23:21:54	<a href="http://twitter.com/download/iphone" r...
2354	2015-11-15 23:05:30	<a href="http://twitter.com/download/iphone" r...

2355 2015-11-15 22:32:08 <a href="http://twitter.com/download/iphone" r...

	text	retweeted_status_id \
0	This is Phineas. He's a mystical boy. Only eve...	NaN
1	This is Tilly. She's just checking pup on you...	NaN
2	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN
5	Here we have a majestic great white breaching ...	NaN
6	Meet Jax. He enjoys ice cream so much he gets ...	NaN
7	When you watch your owner call another dog a g...	NaN
8	This is Zoey. She doesn't want to be one of th...	NaN
9	This is Cassie. She is a college pup. Studying...	NaN
10	This is Koda. He is a South Australian decksha...	NaN
11	This is Bruno. He is a service shark. Only get...	NaN
12	Here's a puppo that seems to be on the fence a...	NaN
13	This is Ted. He does his best. Sometimes that'...	NaN
14	This is Stuart. He's sporting his favorite fan...	NaN
15	This is Oliver. You're witnessing one of his m...	NaN
16	This is Jim. He found a fren. Taught him how t...	NaN
17	This is Zeke. He has a new stick. Very proud o...	NaN
18	This is Ralphus. He's powering up. Attempting ...	NaN
20	This is Gerald. He was just told he didn't get...	NaN
21	This is Jeffrey. He has a monopoly on the pool...	NaN
22	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN
23	This is Canela. She attempted some fancy porch...	NaN
24	You may not have known you needed to see this ...	NaN
25	This... is a Jubilant Antarctic House Bear. We...	NaN
26	This is Maya. She's very shy. Rarely leaves he...	NaN
27	This is Mingus. He's a wonderful father to his...	NaN
28	This is Derek. He's late for a dog meeting. 13...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
30	@NonWhiteHat @MayhewMayhem omg hello tanner yo...	NaN
...
2326	This is quite the dog. Gets really excited whe...	NaN
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN
2328	Oh goodness. A super rare northeast Qdoba kang...	NaN
2329	Those are sunglasses and a jean jacket. 11/10 ...	NaN
2330	Unique dog here. Very small. Lives in containe...	NaN
2331	Here we have a mixed Asiago from the Galápagos...	NaN
2332	Look at this jokester thinking seat belt laws ...	NaN
2333	This is an extremely rare horned Parthenon. No...	NaN
2334	This is a funny dog. Weird toes. Won't come do...	NaN
2335	This is an Albanian 3 1/2 legged Episcopalian...	NaN
2336	Can take selfies 11/10 https://t.co/ws2AMaNPW	NaN
2337	Very concerned about fellow dog trapped in com...	NaN
2338	Not familiar with this breed. No tail (weird)...	NaN
2339	Oh my. Here you are seeing an Adobe Setter giv...	NaN

2340	Can stand on stump for what seems like a while...	NaN
2341	This appears to be a Mongolian Presbyterian mi...	NaN
2342	Here we have a well-established sunblockerspan...	NaN
2343	Let's hope this flight isn't Malaysian (lol). ...	NaN
2344	Here we have a northern speckled Rhododendron...	NaN
2345	This is the happiest dog you will ever see. Ve...	NaN
2346	Here is the Rand Paul of retrievers folks! He'...	NaN
2347	My oh my. This is a rare blond Canadian terrie...	NaN
2348	Here is a Siberian heavily armored polar bear ...	NaN
2349	This is an odd dog. Hard on the outside but lo...	NaN
2350	This is a truly beautiful English Wilson Staff...	NaN
2351	Here we have a 1949 1st generation vulpix. Enj...	NaN
2352	This is a purebred Piers Morgan. Loves to Netf...	NaN
2353	Here is a very happy pup. Big fan of well-main...	NaN
2354	This is a western brown Mitsubishi terrier. Up...	NaN
2355	Here we have a Japanese Irish Setter. Lost eye...	NaN

	retweeted_status_user_id	retweeted_status_timestamp \
0	NaN	NaT
1	NaN	NaT
2	NaN	NaT
3	NaN	NaT
4	NaN	NaT
5	NaN	NaT
6	NaN	NaT
7	NaN	NaT
8	NaN	NaT
9	NaN	NaT
10	NaN	NaT
11	NaN	NaT
12	NaN	NaT
13	NaN	NaT
14	NaN	NaT
15	NaN	NaT
16	NaN	NaT
17	NaN	NaT
18	NaN	NaT
20	NaN	NaT
21	NaN	NaT
22	NaN	NaT
23	NaN	NaT
24	NaN	NaT
25	NaN	NaT
26	NaN	NaT
27	NaN	NaT
28	NaN	NaT
29	NaN	NaT
30	NaN	NaT

...
2326	NaN	NaT
2327	NaN	NaT
2328	NaN	NaT
2329	NaN	NaT
2330	NaN	NaT
2331	NaN	NaT
2332	NaN	NaT
2333	NaN	NaT
2334	NaN	NaT
2335	NaN	NaT
2336	NaN	NaT
2337	NaN	NaT
2338	NaN	NaT
2339	NaN	NaT
2340	NaN	NaT
2341	NaN	NaT
2342	NaN	NaT
2343	NaN	NaT
2344	NaN	NaT
2345	NaN	NaT
2346	NaN	NaT
2347	NaN	NaT
2348	NaN	NaT
2349	NaN	NaT
2350	NaN	NaT
2351	NaN	NaT
2352	NaN	NaT
2353	NaN	NaT
2354	NaN	NaT
2355	NaN	NaT

	expanded_urls	rating_numerator \
0	https://twitter.com/dog_rates/status/892420643...	13
1	https://twitter.com/dog_rates/status/892177421...	13
2	https://twitter.com/dog_rates/status/891815181...	12
3	https://twitter.com/dog_rates/status/891689557...	13
4	https://twitter.com/dog_rates/status/891327558...	12
5	https://twitter.com/dog_rates/status/891087950...	13
6	https://gofundme.com/ydvmve-surgery-for-jax,ht...	13
7	https://twitter.com/dog_rates/status/890729181...	13
8	https://twitter.com/dog_rates/status/890609185...	13
9	https://twitter.com/dog_rates/status/890240255...	14
10	https://twitter.com/dog_rates/status/890006608...	13
11	https://twitter.com/dog_rates/status/889880896...	13
12	https://twitter.com/dog_rates/status/889665388...	13
13	https://twitter.com/dog_rates/status/889638837...	12
14	https://twitter.com/dog_rates/status/889531135...	13

15	https://twitter.com/dog_rates/status/889278841...	13
16	https://twitter.com/dog_rates/status/888917238...	12
17	https://twitter.com/dog_rates/status/888804989...	13
18	https://twitter.com/dog_rates/status/888554962...	13
20	https://twitter.com/dog_rates/status/888078434...	12
21	https://twitter.com/dog_rates/status/887705289...	13
22	https://twitter.com/dog_rates/status/887517139...	14
23	https://twitter.com/dog_rates/status/887473957...	13
24	https://twitter.com/dog_rates/status/887343217...	13
25	https://twitter.com/dog_rates/status/887101392...	12
26	https://twitter.com/dog_rates/status/886983233...	13
27	https://www.gofundme.com/mingusneedsus , https://...	13
28	https://twitter.com/dog_rates/status/886680336...	13
29	https://twitter.com/dog_rates/status/886366144...	12
30	NaN	12
...
2326	https://twitter.com/dog_rates/status/666411507...	2
2327	https://twitter.com/dog_rates/status/666407126...	7
2328	https://twitter.com/dog_rates/status/666396247...	9
2329	https://twitter.com/dog_rates/status/666373753...	11
2330	https://twitter.com/dog_rates/status/666362758...	6
2331	https://twitter.com/dog_rates/status/666353288...	8
2332	https://twitter.com/dog_rates/status/666345417...	10
2333	https://twitter.com/dog_rates/status/666337882...	9
2334	https://twitter.com/dog_rates/status/666293911...	3
2335	https://twitter.com/dog_rates/status/666287406...	1
2336	https://twitter.com/dog_rates/status/666273097...	11
2337	https://twitter.com/dog_rates/status/666268910...	10
2338	https://twitter.com/dog_rates/status/666104133...	1
2339	https://twitter.com/dog_rates/status/666102155...	11
2340	https://twitter.com/dog_rates/status/666099513...	8
2341	https://twitter.com/dog_rates/status/666094000...	9
2342	https://twitter.com/dog_rates/status/666082916...	6
2343	https://twitter.com/dog_rates/status/666073100...	10
2344	https://twitter.com/dog_rates/status/666071193...	9
2345	https://twitter.com/dog_rates/status/666063827...	10
2346	https://twitter.com/dog_rates/status/666058600...	8
2347	https://twitter.com/dog_rates/status/666057090...	9
2348	https://twitter.com/dog_rates/status/666055525...	10
2349	https://twitter.com/dog_rates/status/666051853...	2
2350	https://twitter.com/dog_rates/status/666050758...	10
2351	https://twitter.com/dog_rates/status/666049248...	5
2352	https://twitter.com/dog_rates/status/666044226...	6
2353	https://twitter.com/dog_rates/status/666033412...	9
2354	https://twitter.com/dog_rates/status/666029285...	7
2355	https://twitter.com/dog_rates/status/666020888...	8

rating_denominator name doggo floof pupper puppo \

0	10	Phineas	None	None	None	None
1	10	Tilly	None	None	None	None
2	10	Archie	None	None	None	None
3	10	Darla	None	None	None	None
4	10	Franklin	None	None	None	None
5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None
8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
30	10	None	None	None	None	None
...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None

2343	10	None	None	None	None	None
2344	10	None	None	None	None	None
2345	10	the	None	None	None	None
2346	10	the	None	None	None	None
2347	10	a	None	None	None	None
2348	10	a	None	None	None	None
2349	10	an	None	None	None	None
2350	10	a	None	None	None	None
2351	10	None	None	None	None	None
2352	10	a	None	None	None	None
2353	10	a	None	None	None	None
2354	10	a	None	None	None	None
2355	10	None	None	None	None	None

	normalize_rating	date	time
0	1.3	2017-08-01	16:23:56
1	1.3	2017-08-01	00:17:27
2	1.2	2017-07-31	00:18:03
3	1.3	2017-07-30	15:58:51
4	1.2	2017-07-29	16:00:24
5	1.3	2017-07-29	00:08:17
6	1.3	2017-07-28	16:27:12
7	1.3	2017-07-28	00:22:40
8	1.3	2017-07-27	16:25:51
9	1.4	2017-07-26	15:59:51
10	1.3	2017-07-26	00:31:25
11	1.3	2017-07-25	16:11:53
12	1.3	2017-07-25	01:55:32
13	1.2	2017-07-25	00:10:02
14	1.3	2017-07-24	17:02:04
15	1.3	2017-07-24	00:19:32
16	1.2	2017-07-23	00:22:39
17	1.3	2017-07-22	16:56:37
18	1.3	2017-07-22	00:23:06
20	1.2	2017-07-20	16:49:33
21	1.3	2017-07-19	16:06:48
22	1.4	2017-07-19	03:39:09
23	1.3	2017-07-19	00:47:34
24	1.3	2017-07-18	16:08:03
25	1.2	2017-07-18	00:07:08
26	1.3	2017-07-17	16:17:36
27	1.3	2017-07-16	23:58:41
28	1.3	2017-07-16	20:14:00
29	1.2	2017-07-15	23:25:31
30	1.2	2017-07-15	16:51:35
...
2326	0.2	2015-11-17	00:24:19
2327	0.7	2015-11-17	00:06:54

2328	0.9	2015-11-16	23:23:41
2329	1.1	2015-11-16	21:54:18
2330	0.6	2015-11-16	21:10:36
2331	0.8	2015-11-16	20:32:58
2332	1.0	2015-11-16	20:01:42
2333	0.9	2015-11-16	19:31:45
2334	0.3	2015-11-16	16:37:02
2335	0.5	2015-11-16	16:11:11
2336	1.1	2015-11-16	15:14:19
2337	1.0	2015-11-16	14:57:41
2338	0.1	2015-11-16	04:02:55
2339	1.1	2015-11-16	03:55:04
2340	0.8	2015-11-16	03:44:34
2341	0.9	2015-11-16	03:22:39
2342	0.6	2015-11-16	02:38:37
2343	1.0	2015-11-16	01:59:36
2344	0.9	2015-11-16	01:52:02
2345	1.0	2015-11-16	01:22:45
2346	0.8	2015-11-16	01:01:59
2347	0.9	2015-11-16	00:55:59
2348	1.0	2015-11-16	00:49:46
2349	0.2	2015-11-16	00:35:11
2350	1.0	2015-11-16	00:30:50
2351	0.5	2015-11-16	00:24:50
2352	0.6	2015-11-16	00:04:52
2353	0.9	2015-11-15	23:21:54
2354	0.7	2015-11-15	23:05:30
2355	0.8	2015-11-15	22:32:08

[2175 rows x 20 columns]

0.7.5 text column contain text and url

0.7.6 Define

extract text then store it to content column and extract url then store it to url column

0.7.7 Code

```
In [127]: # get example of text in row 0
         twit_arc_copy.text[30]
```

```
Out[127]: '@NonWhiteHat @MayhewMayhem omg hello tanner you are a scary good boy 12/10 would pet
```

```
In [128]: ##### extract content in each row
         content_raw = twit_arc_copy.text.str.extract(r'(.*)(http|\\n|$)')
```

```
In [129]: # dropping https that is filtered by regex
         content_raw[0][464]
```

```
Out[129]: "Meet Strudel. He's rather h*ckin pupset that your clothes clash. 11/10 click the link
```

```
In [130]: # extract url in content
```

```
content_url = twit_arc_copy.text.str.extract(r'(https://t.co/\w+)\ ?(https.*)?')
```

```
In [131]: # change column name
```

```
content_url.rename(columns={0: 'url_content_1', 1: 'url_content_2'}, inplace = True)
```

```
In [132]: # checkking are there any 2 urls in text column
```

```
twit_arc_copy.text[6]
```

```
Out[132]: 'Meet Jax. He enjoys ice cream so much he gets nervous around it. 13/10 help Jax enjoy
```

```
In [133]: # check content_url
```

```
content_url
```

```
Out[133]:
```

	url_content_1	url_content_2
0	https://t.co/MgUWQ76dJU	NaN
1	https://t.co/OXxu71qeIV	NaN
2	https://t.co/wUnZnhtVJB	NaN
3	https://t.co/tD36da7qLQ	NaN
4	https://t.co/AtUZn91f7f	NaN
5	https://t.co/kQ04fDDRmh	NaN
6	https://t.co/Zr4hWfAs1H	https://t.co/tVJBRMnhxl
7	https://t.co/v0nONBcwq	NaN
8	https://t.co/9TwLuAGH0b	NaN
9	https://t.co/t1bfwz5S2A	NaN
10	https://t.co/dVPWOBOMme	NaN
11	https://t.co/u1XPQM129g	NaN
12	https://t.co/BxvuXkOUCm	NaN
13	https://t.co/f8dEDcrKSR	NaN
14	https://t.co/y70o6h3isq	NaN
15	https://t.co/WpHvrQedPb	NaN
16	https://t.co/chxruIOUJN	NaN
17	https://t.co/HTQ77yNQ5K	NaN
18	https://t.co/YnYAFCTTiK	NaN
20	https://t.co/DK7iDPfuRX	NaN
21	https://t.co/PhrUk20Q64	NaN
22	https://t.co/20VrLAA8ba	NaN
23	https://t.co/cLyzpcUcMX	NaN
24	https://t.co/WZqNqygEyV	NaN
25	https://t.co/4Ad1jzJSdp	NaN
26	https://t.co/I6oNyOCgiT	NaN
27	https://t.co/bVi0Yr4Cff	https://t.co/ISvK0Skd5b
28	https://t.co/BCoWue0abA	NaN
29	https://t.co/RGE08MIJox	NaN
30	NaN	NaN
...
2326	https://t.co/aMCTNW094t	NaN

2327	https://t.co/LopTBkKa8h	NaN
2328	https://t.co/Dc7b0E8qFE	NaN
2329	https://t.co/uHXrPkUEyl	NaN
2330	https://t.co/XMD9CwjEnM	NaN
2331	https://t.co/tltQ5w9aUO	NaN
2332	https://t.co/VFKG1vxGjB	NaN
2333	https://t.co/QpRjllzWAL	NaN
2334	https://t.co/IIXis0ztaO	NaN
2335	https://t.co/d9NcXFKwLv	NaN
2336	https://t.co/ws2AMaWpPW	NaN
2337	https://t.co/OyxApIikpk	NaN
2338	https://t.co/Asgdc6kuLX	NaN
2339	https://t.co/11LvqN4WLq	NaN
2340	https://t.co/Ri4nMTLq5C	NaN
2341	https://t.co/mnioXo3IfP	NaN
2342	https://t.co/3RU6x0vHB7	NaN
2343	https://t.co/Yk6GHE9t0Y	NaN
2344	https://t.co/ZoL8kq2XFx	NaN
2345	https://t.co/RhUEAloehK	NaN
2346	https://t.co/pYAJkAe76p	NaN
2347	https://t.co/yWBqbzry80	NaN
2348	https://t.co/rdivxLiqEt	NaN
2349	https://t.co/v5A4vzSDdc	NaN
2350	https://t.co/fvIbQfHjIe	NaN
2351	https://t.co/4B7c0c1EDq	NaN
2352	https://t.co/DWnyCjf2mx	NaN
2353	https://t.co/y671yMhoiR	NaN
2354	https://t.co/r7m0b2mOUI	NaN
2355	https://t.co/BLDqew2Ijj	NaN

[2175 rows x 2 columns]

```
In [134]: # add parsed data into dataframe
          twit_arc_copy['content'] = content_raw[0]
          twit_arc_copy[['content_url_1', 'content_url_2']] = content_url
```

```
In [135]: twit_arc_copy
```

```
Out[135]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	0	0	
1	892177421306343426	0	0	
2	891815181378084864	0	0	
3	891689557279858688	0	0	
4	891327558926688256	0	0	
5	891087950875897856	0	0	
6	890971913173991426	0	0	
7	890729181411237888	0	0	
8	890609185150312448	0	0	

9	890240255349198849	0	0
10	890006608113172480	0	0
11	889880896479866881	0	0
12	889665388333682689	0	0
13	889638837579907072	0	0
14	889531135344209921	0	0
15	889278841981685760	0	0
16	888917238123831296	0	0
17	888804989199671297	0	0
18	888554962724278272	0	0
20	888078434458587136	0	0
21	887705289381826560	0	0
22	887517139158093824	0	0
23	887473957103951883	0	0
24	887343217045368832	0	0
25	887101392804085760	0	0
26	886983233522544640	0	0
27	886736880519319552	0	0
28	886680336477933568	0	0
29	886366144734445568	0	0
30	886267009285017600	886266357075128320	2281181600
...
2326	666411507551481857	0	0
2327	666407126856765440	0	0
2328	666396247373291520	0	0
2329	666373753744588802	0	0
2330	666362758909284353	0	0
2331	666353288456101888	0	0
2332	666345417576210432	0	0
2333	666337882303524864	0	0
2334	666293911632134144	0	0
2335	666287406224695296	0	0
2336	666273097616637952	0	0
2337	666268910803644416	0	0
2338	666104133288665088	0	0
2339	666102155909144576	0	0
2340	666099513787052032	0	0
2341	666094000022159362	0	0
2342	666082916733198337	0	0
2343	666073100786774016	0	0
2344	666071193221509120	0	0
2345	666063827256086533	0	0
2346	666058600524156928	0	0
2347	666057090499244032	0	0
2348	666055525042405380	0	0
2349	666051853826850816	0	0
2350	666050758794694657	0	0
2351	666049248165822465	0	0

2352	666044226329800704	0	0
2353	666033412701032449	0	0
2354	666029285002620928	0	0
2355	666020888022790149	0	0

	timestamp	source \
0	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone" r...
1	2017-08-01 00:17:27	<a href="http://twitter.com/download/iphone" r...
2	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...
3	2017-07-30 15:58:51	<a href="http://twitter.com/download/iphone" r...
4	2017-07-29 16:00:24	<a href="http://twitter.com/download/iphone" r...
5	2017-07-29 00:08:17	<a href="http://twitter.com/download/iphone" r...
6	2017-07-28 16:27:12	<a href="http://twitter.com/download/iphone" r...
7	2017-07-28 00:22:40	<a href="http://twitter.com/download/iphone" r...
8	2017-07-27 16:25:51	<a href="http://twitter.com/download/iphone" r...
9	2017-07-26 15:59:51	<a href="http://twitter.com/download/iphone" r...
10	2017-07-26 00:31:25	<a href="http://twitter.com/download/iphone" r...
11	2017-07-25 16:11:53	<a href="http://twitter.com/download/iphone" r...
12	2017-07-25 01:55:32	<a href="http://twitter.com/download/iphone" r...
13	2017-07-25 00:10:02	<a href="http://twitter.com/download/iphone" r...
14	2017-07-24 17:02:04	<a href="http://twitter.com/download/iphone" r...
15	2017-07-24 00:19:32	<a href="http://twitter.com/download/iphone" r...
16	2017-07-23 00:22:39	<a href="http://twitter.com/download/iphone" r...
17	2017-07-22 16:56:37	<a href="http://twitter.com/download/iphone" r...
18	2017-07-22 00:23:06	<a href="http://twitter.com/download/iphone" r...
20	2017-07-20 16:49:33	<a href="http://twitter.com/download/iphone" r...
21	2017-07-19 16:06:48	<a href="http://twitter.com/download/iphone" r...
22	2017-07-19 03:39:09	<a href="http://twitter.com/download/iphone" r...
23	2017-07-19 00:47:34	<a href="http://twitter.com/download/iphone" r...
24	2017-07-18 16:08:03	<a href="http://twitter.com/download/iphone" r...
25	2017-07-18 00:07:08	<a href="http://twitter.com/download/iphone" r...
26	2017-07-17 16:17:36	<a href="http://twitter.com/download/iphone" r...
27	2017-07-16 23:58:41	<a href="http://twitter.com/download/iphone" r...
28	2017-07-16 20:14:00	<a href="http://twitter.com/download/iphone" r...
29	2017-07-15 23:25:31	<a href="http://twitter.com/download/iphone" r...
30	2017-07-15 16:51:35	<a href="http://twitter.com/download/iphone" r...
...
2326	2015-11-17 00:24:19	<a href="http://twitter.com/download/iphone" r...
2327	2015-11-17 00:06:54	<a href="http://twitter.com/download/iphone" r...
2328	2015-11-16 23:23:41	<a href="http://twitter.com/download/iphone" r...
2329	2015-11-16 21:54:18	<a href="http://twitter.com/download/iphone" r...
2330	2015-11-16 21:10:36	<a href="http://twitter.com/download/iphone" r...
2331	2015-11-16 20:32:58	<a href="http://twitter.com/download/iphone" r...
2332	2015-11-16 20:01:42	<a href="http://twitter.com/download/iphone" r...
2333	2015-11-16 19:31:45	<a href="http://twitter.com/download/iphone" r...
2334	2015-11-16 16:37:02	<a href="http://twitter.com/download/iphone" r...
2335	2015-11-16 16:11:11	<a href="http://twitter.com/download/iphone" r...
2336	2015-11-16 15:14:19	<a href="http://twitter.com/download/iphone" r...

2337 2015-11-16 14:57:41 <a href="http://twitter.com/download/iphone" r...
 2338 2015-11-16 04:02:55 <a href="http://twitter.com/download/iphone" r...
 2339 2015-11-16 03:55:04 <a href="http://twitter.com/download/iphone" r...
 2340 2015-11-16 03:44:34 <a href="http://twitter.com/download/iphone" r...
 2341 2015-11-16 03:22:39 <a href="http://twitter.com/download/iphone" r...
 2342 2015-11-16 02:38:37 <a href="http://twitter.com/download/iphone" r...
 2343 2015-11-16 01:59:36 <a href="http://twitter.com/download/iphone" r...
 2344 2015-11-16 01:52:02 <a href="http://twitter.com/download/iphone" r...
 2345 2015-11-16 01:22:45 <a href="http://twitter.com/download/iphone" r...
 2346 2015-11-16 01:01:59 <a href="http://twitter.com/download/iphone" r...
 2347 2015-11-16 00:55:59 <a href="http://twitter.com/download/iphone" r...
 2348 2015-11-16 00:49:46 <a href="http://twitter.com/download/iphone" r...
 2349 2015-11-16 00:35:11 <a href="http://twitter.com/download/iphone" r...
 2350 2015-11-16 00:30:50 <a href="http://twitter.com/download/iphone" r...
 2351 2015-11-16 00:24:50 <a href="http://twitter.com/download/iphone" r...
 2352 2015-11-16 00:04:52 <a href="http://twitter.com/download/iphone" r...
 2353 2015-11-15 23:21:54 <a href="http://twitter.com/download/iphone" r...
 2354 2015-11-15 23:05:30 <a href="http://twitter.com/download/iphone" r...
 2355 2015-11-15 22:32:08 <a href="http://twitter.com/download/iphone" r...

	text	retweeted_status_id \
0	This is Phineas. He's a mystical boy. Only eve...	NaN
1	This is Tilly. She's just checking pup on you...	NaN
2	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN
5	Here we have a majestic great white breaching ...	NaN
6	Meet Jax. He enjoys ice cream so much he gets ...	NaN
7	When you watch your owner call another dog a g...	NaN
8	This is Zoey. She doesn't want to be one of th...	NaN
9	This is Cassie. She is a college pup. Studying...	NaN
10	This is Koda. He is a South Australian decksha...	NaN
11	This is Bruno. He is a service shark. Only get...	NaN
12	Here's a puppo that seems to be on the fence a...	NaN
13	This is Ted. He does his best. Sometimes that'...	NaN
14	This is Stuart. He's sporting his favorite fan...	NaN
15	This is Oliver. You're witnessing one of his m...	NaN
16	This is Jim. He found a fren. Taught him how t...	NaN
17	This is Zeke. He has a new stick. Very proud o...	NaN
18	This is Ralphus. He's powering up. Attempting ...	NaN
20	This is Gerald. He was just told he didn't get...	NaN
21	This is Jeffrey. He has a monopoly on the pool...	NaN
22	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN
23	This is Canela. She attempted some fancy porch...	NaN
24	You may not have known you needed to see this ...	NaN
25	This... is a Jubilant Antarctic House Bear. We...	NaN
26	This is Maya. She's very shy. Rarely leaves he...	NaN
27	This is Mingus. He's a wonderful father to his...	NaN

28	This is Derek. He's late for a dog meeting. 13...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
30	@NonWhiteHat @MayhewMayhem omg hello tanner yo...	NaN
...
2326	This is quite the dog. Gets really excited whe...	NaN
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN
2328	Oh goodness. A super rare northeast Qdoba kang...	NaN
2329	Those are sunglasses and a jean jacket. 11/10 ...	NaN
2330	Unique dog here. Very small. Lives in containe...	NaN
2331	Here we have a mixed Asiago from the Galápagos...	NaN
2332	Look at this jokester thinking seat belt laws ...	NaN
2333	This is an extremely rare horned Parthenon. No...	NaN
2334	This is a funny dog. Weird toes. Won't come do...	NaN
2335	This is an Albanian 3 1/2 legged Episcopalian...	NaN
2336	Can take selfies 11/10 https://t.co/ws2AMaWpW	NaN
2337	Very concerned about fellow dog trapped in com...	NaN
2338	Not familiar with this breed. No tail (weird)...	NaN
2339	Oh my. Here you are seeing an Adobe Setter giv...	NaN
2340	Can stand on stump for what seems like a while...	NaN
2341	This appears to be a Mongolian Presbyterian mi...	NaN
2342	Here we have a well-established sunblockerspan...	NaN
2343	Let's hope this flight isn't Malaysian (lol). ...	NaN
2344	Here we have a northern speckled Rhododendron...	NaN
2345	This is the happiest dog you will ever see. Ve...	NaN
2346	Here is the Rand Paul of retrievers folks! He'...	NaN
2347	My oh my. This is a rare blond Canadian terrie...	NaN
2348	Here is a Siberian heavily armored polar bear ...	NaN
2349	This is an odd dog. Hard on the outside but lo...	NaN
2350	This is a truly beautiful English Wilson Staff...	NaN
2351	Here we have a 1949 1st generation vulpix. Enj...	NaN
2352	This is a purebred Piers Morgan. Loves to Netf...	NaN
2353	Here is a very happy pup. Big fan of well-main...	NaN
2354	This is a western brown Mitsubishi terrier. Up...	NaN
2355	Here we have a Japanese Irish Setter. Lost eye...	NaN

	retweeted_status_user_id	retweeted_status_timestamp \
0	NaN	NaT
1	NaN	NaT
2	NaN	NaT
3	NaN	NaT
4	NaN	NaT
5	NaN	NaT
6	NaN	NaT
7	NaN	NaT
8	NaN	NaT
9	NaN	NaT
10	NaN	NaT
11	NaN	NaT

12	NaN	NaT
13	NaN	NaT
14	NaN	NaT
15	NaN	NaT
16	NaN	NaT
17	NaN	NaT
18	NaN	NaT
20	NaN	NaT
21	NaN	NaT
22	NaN	NaT
23	NaN	NaT
24	NaN	NaT
25	NaN	NaT
26	NaN	NaT
27	NaN	NaT
28	NaN	NaT
29	NaN	NaT
30	NaN	NaT
...
2326	NaN	NaT
2327	NaN	NaT
2328	NaN	NaT
2329	NaN	NaT
2330	NaN	NaT
2331	NaN	NaT
2332	NaN	NaT
2333	NaN	NaT
2334	NaN	NaT
2335	NaN	NaT
2336	NaN	NaT
2337	NaN	NaT
2338	NaN	NaT
2339	NaN	NaT
2340	NaN	NaT
2341	NaN	NaT
2342	NaN	NaT
2343	NaN	NaT
2344	NaN	NaT
2345	NaN	NaT
2346	NaN	NaT
2347	NaN	NaT
2348	NaN	NaT
2349	NaN	NaT
2350	NaN	NaT
2351	NaN	NaT
2352	NaN	NaT
2353	NaN	NaT
2354	NaN	NaT

```

                                expanded_urls \
0      https://twitter.com/dog_rates/status/892420643...
1      https://twitter.com/dog_rates/status/892177421...
2      https://twitter.com/dog_rates/status/891815181...
3      https://twitter.com/dog_rates/status/891689557...
4      https://twitter.com/dog_rates/status/891327558...
5      https://twitter.com/dog_rates/status/891087950...
6      https://gofundme.com/ydvmve-surgery-for-jax,ht...
7      https://twitter.com/dog_rates/status/890729181...
8      https://twitter.com/dog_rates/status/890609185...
9      https://twitter.com/dog_rates/status/890240255...
10     https://twitter.com/dog_rates/status/890006608...
11     https://twitter.com/dog_rates/status/889880896...
12     https://twitter.com/dog_rates/status/889665388...
13     https://twitter.com/dog_rates/status/889638837...
14     https://twitter.com/dog_rates/status/889531135...
15     https://twitter.com/dog_rates/status/889278841...
16     https://twitter.com/dog_rates/status/888917238...
17     https://twitter.com/dog_rates/status/888804989...
18     https://twitter.com/dog_rates/status/888554962...
20     https://twitter.com/dog_rates/status/888078434...
21     https://twitter.com/dog_rates/status/887705289...
22     https://twitter.com/dog_rates/status/887517139...
23     https://twitter.com/dog_rates/status/887473957...
24     https://twitter.com/dog_rates/status/887343217...
25     https://twitter.com/dog_rates/status/887101392...
26     https://twitter.com/dog_rates/status/886983233...
27     https://www.gofundme.com/mingusneedsus,https:/...
28     https://twitter.com/dog_rates/status/886680336...
29     https://twitter.com/dog_rates/status/886366144...
30                                           NaN
...
2326   https://twitter.com/dog_rates/status/666411507...
2327   https://twitter.com/dog_rates/status/666407126...
2328   https://twitter.com/dog_rates/status/666396247...
2329   https://twitter.com/dog_rates/status/666373753...
2330   https://twitter.com/dog_rates/status/666362758...
2331   https://twitter.com/dog_rates/status/666353288...
2332   https://twitter.com/dog_rates/status/666345417...
2333   https://twitter.com/dog_rates/status/666337882...
2334   https://twitter.com/dog_rates/status/666293911...
2335   https://twitter.com/dog_rates/status/666287406...
2336   https://twitter.com/dog_rates/status/666273097...
2337   https://twitter.com/dog_rates/status/666268910...
2338   https://twitter.com/dog_rates/status/666104133...
2339   https://twitter.com/dog_rates/status/666102155...

```

2340 https://twitter.com/dog_rates/status/666099513...
 2341 https://twitter.com/dog_rates/status/666094000...
 2342 https://twitter.com/dog_rates/status/666082916...
 2343 https://twitter.com/dog_rates/status/666073100...
 2344 https://twitter.com/dog_rates/status/666071193...
 2345 https://twitter.com/dog_rates/status/666063827...
 2346 https://twitter.com/dog_rates/status/666058600...
 2347 https://twitter.com/dog_rates/status/666057090...
 2348 https://twitter.com/dog_rates/status/666055525...
 2349 https://twitter.com/dog_rates/status/666051853...
 2350 https://twitter.com/dog_rates/status/666050758...
 2351 https://twitter.com/dog_rates/status/666049248...
 2352 https://twitter.com/dog_rates/status/666044226...
 2353 https://twitter.com/dog_rates/status/666033412...
 2354 https://twitter.com/dog_rates/status/666029285...
 2355 https://twitter.com/dog_rates/status/666020888...

	...	doggo	floof	pupper	puppo	normalize_rating	\
0	...	None	None	None	None	1.3	
1	...	None	None	None	None	1.3	
2	...	None	None	None	None	1.2	
3	...	None	None	None	None	1.3	
4	...	None	None	None	None	1.2	
5	...	None	None	None	None	1.3	
6	...	None	None	None	None	1.3	
7	...	None	None	None	None	1.3	
8	...	None	None	None	None	1.3	
9	...	doggo	None	None	None	1.4	
10	...	None	None	None	None	1.3	
11	...	None	None	None	None	1.3	
12	...	None	None	None	puppo	1.3	
13	...	None	None	None	None	1.2	
14	...	None	None	None	puppo	1.3	
15	...	None	None	None	None	1.3	
16	...	None	None	None	None	1.2	
17	...	None	None	None	None	1.3	
18	...	None	None	None	None	1.3	
20	...	None	None	None	None	1.2	
21	...	None	None	None	None	1.3	
22	...	None	None	None	None	1.4	
23	...	None	None	None	None	1.3	
24	...	None	None	None	None	1.3	
25	...	None	None	None	None	1.2	
26	...	None	None	None	None	1.3	
27	...	None	None	None	None	1.3	
28	...	None	None	None	None	1.3	
29	...	None	None	pupper	None	1.2	
30	...	None	None	None	None	1.2	

...
2326	...	None	None	None	0.2
2327	...	None	None	None	0.7
2328	...	None	None	None	0.9
2329	...	None	None	None	1.1
2330	...	None	None	None	0.6
2331	...	None	None	None	0.8
2332	...	None	None	None	1.0
2333	...	None	None	None	0.9
2334	...	None	None	None	0.3
2335	...	None	None	None	0.5
2336	...	None	None	None	1.1
2337	...	None	None	None	1.0
2338	...	None	None	None	0.1
2339	...	None	None	None	1.1
2340	...	None	None	None	0.8
2341	...	None	None	None	0.9
2342	...	None	None	None	0.6
2343	...	None	None	None	1.0
2344	...	None	None	None	0.9
2345	...	None	None	None	1.0
2346	...	None	None	None	0.8
2347	...	None	None	None	0.9
2348	...	None	None	None	1.0
2349	...	None	None	None	0.2
2350	...	None	None	None	1.0
2351	...	None	None	None	0.5
2352	...	None	None	None	0.6
2353	...	None	None	None	0.9
2354	...	None	None	None	0.7
2355	...	None	None	None	0.8

	date	time	content \
0	2017-08-01	16:23:56	This is Phineas. He's a mystical boy. Only eve...
1	2017-08-01	00:17:27	This is Tilly. She's just checking pup on you...
2	2017-07-31	00:18:03	This is Archie. He is a rare Norwegian Pouncin...
3	2017-07-30	15:58:51	This is Darla. She commenced a snooze mid meal...
4	2017-07-29	16:00:24	This is Franklin. He would like you to stop ca...
5	2017-07-29	00:08:17	Here we have a majestic great white breaching ...
6	2017-07-28	16:27:12	Meet Jax. He enjoys ice cream so much he gets ...
7	2017-07-28	00:22:40	When you watch your owner call another dog a g...
8	2017-07-27	16:25:51	This is Zoey. She doesn't want to be one of th...
9	2017-07-26	15:59:51	This is Cassie. She is a college pup. Studying...
10	2017-07-26	00:31:25	This is Koda. He is a South Australian decksha...
11	2017-07-25	16:11:53	This is Bruno. He is a service shark. Only get...
12	2017-07-25	01:55:32	Here's a puppo that seems to be on the fence a...
13	2017-07-25	00:10:02	This is Ted. He does his best. Sometimes that'...
14	2017-07-24	17:02:04	This is Stuart. He's sporting his favorite fan...

15	2017-07-24	00:19:32	This is Oliver. You're witnessing one of his m...
16	2017-07-23	00:22:39	This is Jim. He found a fren. Taught him how t...
17	2017-07-22	16:56:37	This is Zeke. He has a new stick. Very proud o...
18	2017-07-22	00:23:06	This is Ralphus. He's powering up. Attempting ...
20	2017-07-20	16:49:33	This is Gerald. He was just told he didn't get...
21	2017-07-19	16:06:48	This is Jeffrey. He has a monopoly on the pool...
22	2017-07-19	03:39:09	I've yet to rate a Venezuelan Hover Wiener. Th...
23	2017-07-19	00:47:34	This is Canela. She attempted some fancy porch...
24	2017-07-18	16:08:03	You may not have known you needed to see this ...
25	2017-07-18	00:07:08	This... is a Jubilant Antarctic House Bear. We...
26	2017-07-17	16:17:36	This is Maya. She's very shy. Rarely leaves he...
27	2017-07-16	23:58:41	This is Mingus. He's a wonderful father to his...
28	2017-07-16	20:14:00	This is Derek. He's late for a dog meeting. 13...
29	2017-07-15	23:25:31	This is Roscoe. Another pupper fallen victim t...
30	2017-07-15	16:51:35	@NonWhiteHat @MayhewMayhem omg hello tanner yo...
...
2326	2015-11-17	00:24:19	This is quite the dog. Gets really excited whe...
2327	2015-11-17	00:06:54	This is a southern Vesuvius bumblegruff. Can d...
2328	2015-11-16	23:23:41	Oh goodness. A super rare northeast Qdoba kang...
2329	2015-11-16	21:54:18	Those are sunglasses and a jean jacket. 11/10 ...
2330	2015-11-16	21:10:36	Unique dog here. Very small. Lives in containe...
2331	2015-11-16	20:32:58	Here we have a mixed Asiago from the Galápagos...
2332	2015-11-16	20:01:42	Look at this jokester thinking seat belt laws ...
2333	2015-11-16	19:31:45	This is an extremely rare horned Parthenon. No...
2334	2015-11-16	16:37:02	This is a funny dog. Weird toes. Won't come do...
2335	2015-11-16	16:11:11	This is an Albanian 3 1/2 legged Episcopalian...
2336	2015-11-16	15:14:19	Can take selfies 11/10 https://t.co/ws2AMaWpW
2337	2015-11-16	14:57:41	Very concerned about fellow dog trapped in com...
2338	2015-11-16	04:02:55	Not familiar with this breed. No tail (weird)...
2339	2015-11-16	03:55:04	Oh my. Here you are seeing an Adobe Setter giv...
2340	2015-11-16	03:44:34	Can stand on stump for what seems like a while...
2341	2015-11-16	03:22:39	This appears to be a Mongolian Presbyterian mi...
2342	2015-11-16	02:38:37	Here we have a well-established sunblockerspan...
2343	2015-11-16	01:59:36	Let's hope this flight isn't Malaysian (lol). ...
2344	2015-11-16	01:52:02	Here we have a northern speckled Rhododendron...
2345	2015-11-16	01:22:45	This is the happiest dog you will ever see. Ve...
2346	2015-11-16	01:01:59	Here is the Rand Paul of retrievers folks! He'...
2347	2015-11-16	00:55:59	My oh my. This is a rare blond Canadian terrie...
2348	2015-11-16	00:49:46	Here is a Siberian heavily armored polar bear ...
2349	2015-11-16	00:35:11	This is an odd dog. Hard on the outside but lo...
2350	2015-11-16	00:30:50	This is a truly beautiful English Wilson Staff...
2351	2015-11-16	00:24:50	Here we have a 1949 1st generation vulpix. Enj...
2352	2015-11-16	00:04:52	This is a purebred Piers Morgan. Loves to Netf...
2353	2015-11-15	23:21:54	Here is a very happy pup. Big fan of well-main...
2354	2015-11-15	23:05:30	This is a western brown Mitsubishi terrier. Up...
2355	2015-11-15	22:32:08	Here we have a Japanese Irish Setter. Lost eye...

content_url_1

content_url_2

0	https://t.co/MgUWQ76dJU	NaN
1	https://t.co/OXxu71qeIV	NaN
2	https://t.co/wUnZnhtVJB	NaN
3	https://t.co/tD36da7qLQ	NaN
4	https://t.co/AtUZn91f7f	NaN
5	https://t.co/kQ04fDDRmh	NaN
6	https://t.co/Zr4hWfAs1H	https://t.co/tVJBRMnhx1
7	https://t.co/v0nONBcwXq	NaN
8	https://t.co/9TwLuAGH0b	NaN
9	https://t.co/t1bfwz5S2A	NaN
10	https://t.co/dVPWOBOMme	NaN
11	https://t.co/u1XPQM129g	NaN
12	https://t.co/BxvuXk0UCm	NaN
13	https://t.co/f8dEDcrKSR	NaN
14	https://t.co/y70o6h3isq	NaN
15	https://t.co/WpHvrQedPb	NaN
16	https://t.co/chxruIOUJN	NaN
17	https://t.co/HTQ77yNQ5K	NaN
18	https://t.co/YnYAFCTTiK	NaN
20	https://t.co/DK7iDPfuRX	NaN
21	https://t.co/PhrUk20Q64	NaN
22	https://t.co/20VrLAA8ba	NaN
23	https://t.co/cLyzpcUcMX	NaN
24	https://t.co/WZqNqygEyV	NaN
25	https://t.co/4Ad1jzJSdp	NaN
26	https://t.co/I6oNy0CgiT	NaN
27	https://t.co/bVi0Yr4Cff	https://t.co/ISvK0Skd5b
28	https://t.co/BCoWue0abA	NaN
29	https://t.co/RGE08MIJox	NaN
30	NaN	NaN
...
2326	https://t.co/aMCTNW094t	NaN
2327	https://t.co/LopTBkKa8h	NaN
2328	https://t.co/Dc7b0E8qFE	NaN
2329	https://t.co/uHXrPkUEyl	NaN
2330	https://t.co/XMD9CwjEnM	NaN
2331	https://t.co/tltQ5w9aU0	NaN
2332	https://t.co/VFKG1vxGjB	NaN
2333	https://t.co/QpRjllzWAL	NaN
2334	https://t.co/IIXis0zta0	NaN
2335	https://t.co/d9NcXFKwLv	NaN
2336	https://t.co/ws2AMaWpPW	NaN
2337	https://t.co/OyxApIikpk	NaN
2338	https://t.co/Asgdc6kuLX	NaN
2339	https://t.co/11LvqN4WLq	NaN
2340	https://t.co/Ri4nMTLq5C	NaN
2341	https://t.co/mnioXo3IfP	NaN
2342	https://t.co/3RU6x0vHB7	NaN

2343	https://t.co/Yk6GHE9t0Y	NaN
2344	https://t.co/ZoL8kq2XFx	NaN
2345	https://t.co/RhUEAloehK	NaN
2346	https://t.co/pYAJkAe76p	NaN
2347	https://t.co/yWBqbrzy80	NaN
2348	https://t.co/rdivxLiqEt	NaN
2349	https://t.co/v5A4vzSDdc	NaN
2350	https://t.co/fvIbQfHjIe	NaN
2351	https://t.co/4B7c0c1EDq	NaN
2352	https://t.co/DWnyCjf2mx	NaN
2353	https://t.co/y671yMhoiR	NaN
2354	https://t.co/r7m0b2mOUI	NaN
2355	https://t.co/BLDqew2Ijj	NaN

[2175 rows x 23 columns]

0.7.8 Test

```
In [136]: # overview check parsed text
          twit_arc_copy[['text', 'content', 'content_url_1', 'content_url_2']]
```

```
Out[136]:
```

	text \
0	This is Phineas. He's a mystical boy. Only eve...
1	This is Tilly. She's just checking pup on you...
2	This is Archie. He is a rare Norwegian Pouncin...
3	This is Darla. She commenced a snooze mid meal...
4	This is Franklin. He would like you to stop ca...
5	Here we have a majestic great white breaching ...
6	Meet Jax. He enjoys ice cream so much he gets ...
7	When you watch your owner call another dog a g...
8	This is Zoey. She doesn't want to be one of th...
9	This is Cassie. She is a college pup. Studying...
10	This is Koda. He is a South Australian decksha...
11	This is Bruno. He is a service shark. Only get...
12	Here's a puppo that seems to be on the fence a...
13	This is Ted. He does his best. Sometimes that'...
14	This is Stuart. He's sporting his favorite fan...
15	This is Oliver. You're witnessing one of his m...
16	This is Jim. He found a fren. Taught him how t...
17	This is Zeke. He has a new stick. Very proud o...
18	This is Ralphus. He's powering up. Attempting ...
20	This is Gerald. He was just told he didn't get...
21	This is Jeffrey. He has a monopoly on the pool...
22	I've yet to rate a Venezuelan Hover Wiener. Th...
23	This is Canela. She attempted some fancy porch...
24	You may not have known you needed to see this ...
25	This... is a Jubilant Antarctic House Bear. We...
26	This is Maya. She's very shy. Rarely leaves he...

27 This is Mingus. He's a wonderful father to his...
 28 This is Derek. He's late for a dog meeting. 13...
 29 This is Roscoe. Another pupper fallen victim t...
 30 @NonWhiteHat @MayhewMayhem omg hello tanner yo...
 ...
 2326 This is quite the dog. Gets really excited whe...
 2327 This is a southern Vesuvius bumblegruff. Can d...
 2328 Oh goodness. A super rare northeast Qdoba kang...
 2329 Those are sunglasses and a jean jacket. 11/10 ...
 2330 Unique dog here. Very small. Lives in containe...
 2331 Here we have a mixed Asiago from the Galápagos...
 2332 Look at this jokester thinking seat belt laws ...
 2333 This is an extremely rare horned Parthenon. No...
 2334 This is a funny dog. Weird toes. Won't come do...
 2335 This is an Albanian 3 1/2 legged Episcopalian...
 2336 Can take selfies 11/10 <https://t.co/ws2AMaNPW>
 2337 Very concerned about fellow dog trapped in com...
 2338 Not familiar with this breed. No tail (weird)...
 2339 Oh my. Here you are seeing an Adobe Setter giv...
 2340 Can stand on stump for what seems like a while...
 2341 This appears to be a Mongolian Presbyterian mi...
 2342 Here we have a well-established sunblockerspan...
 2343 Let's hope this flight isn't Malaysian (lol). ...
 2344 Here we have a northern speckled Rhododendron...
 2345 This is the happiest dog you will ever see. Ve...
 2346 Here is the Rand Paul of retrievers folks! He'...
 2347 My oh my. This is a rare blond Canadian terrie...
 2348 Here is a Siberian heavily armored polar bear ...
 2349 This is an odd dog. Hard on the outside but lo...
 2350 This is a truly beautiful English Wilson Staff...
 2351 Here we have a 1949 1st generation vulpix. Enj...
 2352 This is a purebred Piers Morgan. Loves to Netf...
 2353 Here is a very happy pup. Big fan of well-main...
 2354 This is a western brown Mitsubishi terrier. Up...
 2355 Here we have a Japanese Irish Setter. Lost eye...

content \

0 This is Phineas. He's a mystical boy. Only eve...
 1 This is Tilly. She's just checking pup on you...
 2 This is Archie. He is a rare Norwegian Pouncin...
 3 This is Darla. She commenced a snooze mid meal...
 4 This is Franklin. He would like you to stop ca...
 5 Here we have a majestic great white breaching ...
 6 Meet Jax. He enjoys ice cream so much he gets ...
 7 When you watch your owner call another dog a g...
 8 This is Zoey. She doesn't want to be one of th...
 9 This is Cassie. She is a college pup. Studying...
 10 This is Koda. He is a South Australian decksha...

11 This is Bruno. He is a service shark. Only get...
 12 Here's a puppo that seems to be on the fence a...
 13 This is Ted. He does his best. Sometimes that'...
 14 This is Stuart. He's sporting his favorite fan...
 15 This is Oliver. You're witnessing one of his m...
 16 This is Jim. He found a fren. Taught him how t...
 17 This is Zeke. He has a new stick. Very proud o...
 18 This is Ralphus. He's powering up. Attempting ...
 20 This is Gerald. He was just told he didn't get...
 21 This is Jeffrey. He has a monopoly on the pool...
 22 I've yet to rate a Venezuelan Hover Wiener. Th...
 23 This is Canela. She attempted some fancy porch...
 24 You may not have known you needed to see this ...
 25 This... is a Jubilant Antarctic House Bear. We...
 26 This is Maya. She's very shy. Rarely leaves he...
 27 This is Mingus. He's a wonderful father to his...
 28 This is Derek. He's late for a dog meeting. 13...
 29 This is Roscoe. Another pupper fallen victim t...
 30 @NonWhiteHat @MayhewMayhem omg hello tanner yo...
 ...
 2326 This is quite the dog. Gets really excited whe...
 2327 This is a southern Vesuvius bumblegruff. Can d...
 2328 Oh goodness. A super rare northeast Qdoba kang...
 2329 Those are sunglasses and a jean jacket. 11/10 ...
 2330 Unique dog here. Very small. Lives in containe...
 2331 Here we have a mixed Asiago from the Galápagos...
 2332 Look at this jokester thinking seat belt laws ...
 2333 This is an extremely rare horned Parthenon. No...
 2334 This is a funny dog. Weird toes. Won't come do...
 2335 This is an Albanian 3 1/2 legged Episcopalian...
 2336 Can take selfies 11/10 <https://t.co/ws2AMaNwPW>
 2337 Very concerned about fellow dog trapped in com...
 2338 Not familiar with this breed. No tail (weird)...
 2339 Oh my. Here you are seeing an Adobe Setter giv...
 2340 Can stand on stump for what seems like a while...
 2341 This appears to be a Mongolian Presbyterian mi...
 2342 Here we have a well-established sunblockerspan...
 2343 Let's hope this flight isn't Malaysian (lol). ...
 2344 Here we have a northern speckled Rhododendron...
 2345 This is the happiest dog you will ever see. Ve...
 2346 Here is the Rand Paul of retrievers folks! He'...
 2347 My oh my. This is a rare blond Canadian terrie...
 2348 Here is a Siberian heavily armored polar bear ...
 2349 This is an odd dog. Hard on the outside but lo...
 2350 This is a truly beautiful English Wilson Staff...
 2351 Here we have a 1949 1st generation vulpix. Enj...
 2352 This is a purebred Piers Morgan. Loves to Netf...
 2353 Here is a very happy pup. Big fan of well-main...

2354 This is a western brown Mitsubishi terrier. Up...
 2355 Here we have a Japanese Irish Setter. Lost eye...

	content_url_1	content_url_2
0	https://t.co/MgUWQ76dJU	NaN
1	https://t.co/OXxu71qeIV	NaN
2	https://t.co/wUnZnhtVJB	NaN
3	https://t.co/tD36da7qLQ	NaN
4	https://t.co/AtUZn91f7f	NaN
5	https://t.co/kQ04fDDRmh	NaN
6	https://t.co/Zr4hWfAs1H	https://t.co/tVJBRMnhxl
7	https://t.co/vOnONBcwXq	NaN
8	https://t.co/9TwLuAGH0b	NaN
9	https://t.co/t1bfwz5S2A	NaN
10	https://t.co/dVPWOBOMme	NaN
11	https://t.co/u1XPQM129g	NaN
12	https://t.co/BxvuXk0UCm	NaN
13	https://t.co/f8dEDcrKSR	NaN
14	https://t.co/y70o6h3isq	NaN
15	https://t.co/WpHvrQedPb	NaN
16	https://t.co/chxruIOUJN	NaN
17	https://t.co/HTQ77yNQ5K	NaN
18	https://t.co/YnYAFCTTiK	NaN
20	https://t.co/DK7iDPfuRX	NaN
21	https://t.co/PhrUk20Q64	NaN
22	https://t.co/20VrLAA8ba	NaN
23	https://t.co/cLyzpcUcMX	NaN
24	https://t.co/WZqNqygEyV	NaN
25	https://t.co/4Ad1jzJSdp	NaN
26	https://t.co/I6oNyOCgiT	NaN
27	https://t.co/bVi0Yr4Cff	https://t.co/ISvK0Skd5b
28	https://t.co/BCoWue0abA	NaN
29	https://t.co/RGE08MIJox	NaN
30	NaN	NaN
...
2326	https://t.co/aMCTNW094t	NaN
2327	https://t.co/LopTBkKa8h	NaN
2328	https://t.co/Dc7b0E8qFE	NaN
2329	https://t.co/uHXrPkuUEyl	NaN
2330	https://t.co/XMD9CwjEnM	NaN
2331	https://t.co/tltQ5w9aU0	NaN
2332	https://t.co/VFKG1vxGjB	NaN
2333	https://t.co/QpRjllzWAL	NaN
2334	https://t.co/IIXis0zta0	NaN
2335	https://t.co/d9NcXFKwLv	NaN
2336	https://t.co/ws2AMaNwPW	NaN
2337	https://t.co/OyxApIikpk	NaN
2338	https://t.co/Asgdc6kuLX	NaN

2339	https://t.co/11LvqN4WLq	NaN
2340	https://t.co/Ri4nMTLq5C	NaN
2341	https://t.co/mnioXo3IfP	NaN
2342	https://t.co/3RU6x0vHB7	NaN
2343	https://t.co/Yk6GHE9t0Y	NaN
2344	https://t.co/ZoL8kq2XFx	NaN
2345	https://t.co/RhUEAl0ehK	NaN
2346	https://t.co/pYAJkAe76p	NaN
2347	https://t.co/yWBqbrzy80	NaN
2348	https://t.co/rdivxLiqEt	NaN
2349	https://t.co/v5A4vzSDdc	NaN
2350	https://t.co/fvIbQfHjIe	NaN
2351	https://t.co/4B7c0c1EDq	NaN
2352	https://t.co/DWnyCjf2mx	NaN
2353	https://t.co/y671yMhoiR	NaN
2354	https://t.co/r7m0b2m0UI	NaN
2355	https://t.co/BLDqew2Ijj	NaN

[2175 rows x 4 columns]

```
In [137]: # make sure there are no unparsed content
          len(twit_arc_copy[twit_arc_copy.content.isnull()]['text'])
```

Out[137]: 0

0.7.9 Dropping unused data column

0.7.10 Twitter archive

```
In [138]: # dropping timestamp and text
          twit_arc_copy.drop(['timestamp', 'text'], axis = 1, inplace = True)
```

timestamp and text are parsed, and the parsed data has been join with tweeter archive df. Dropping timestamp and text will reduce storage memory when data is stored.

```
In [139]: twit_arc_copy.drop(['retweeted_status_id', 'retweeted_status_user_id', 'retweeted_stat
```

tweeter archive dataframe has been cleaned from retweeted tweet. In addition all column that have corelation with retweet only hold null value. retweeted_status_id, retweeted_status_user_id and retweeted_status_timestamp become unnecessary column based on previous reason.

0.7.11 json data

```
In [140]: json_copy.head()
```

	tweet_id	full_text \
0	892420643555336193	This is Phineas. He's a mystical boy. Only eve...
1	892177421306343426	This is Tilly. She's just checking pup on you...
2	891815181378084864	This is Archie. He is a rare Norwegian Pouncin...
3	891689557279858688	This is Darla. She commenced a snooze mid meal...

```
4 891327558926688256 This is Franklin. He would like you to stop ca...
```

	in_reply_to_status_id	in_reply_to_user_id	retweet_count	favorite_count	\
0	NaN	NaN	7740	36366	
1	NaN	NaN	5730	31326	
2	NaN	NaN	3791	23598	
3	NaN	NaN	7902	39669	
4	NaN	NaN	8524	37867	

	favorite_retweet_ratio
0	4.698450
1	5.467016
2	6.224743
3	5.020121
4	4.442398

```
In [141]: json_copy.drop(['full_text', 'in_reply_to_status_id', 'in_reply_to_user_id'], axis = 1)
```

full_text, in_reply_to_status_id and in_reply_to_user_id has stored in twitter archive data, storing it is unnecessary.

0.8 Storing data

```
In [142]: # store data from df to csv and tsv
twit_arc_copy.to_csv('twitter_archive_master.csv', index = False)
img_pred_copy.to_csv('image_prediction_filtered.tsv', sep='\t', index = False)
json_copy.to_csv('favorite_retweet_filtered.csv', index = False)
```

```
In [143]: os.listdir()
```

```
Out[143]: ['wrangle_act.ipynb',
'.ipynb_checkpoints',
'twitter-archive-enhanced.csv',
'Untitled1.ipynb',
'favorite_retweet_filtered.csv',
'image-predictions.tsv',
'Untitled.ipynb',
'act_report.ipynb',
'wrangle_report.ipynb',
'twitter_archive_master.csv',
'image_prediction_filtered.tsv',
'tweet_json.txt']
```

0.9 Recheck by opening data

```
In [144]: twit_ac = pd.read_csv('twitter_archive_master.csv')
image_filt = pd.read_csv('image_prediction_filtered.tsv', sep = '\t')
rt_fav = pd.read_csv('favorite_retweet_filtered.csv')
```

```
In [145]: twit_ac.head(10)
```

```
Out[145]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	0	0	
1	892177421306343426	0	0	
2	891815181378084864	0	0	
3	891689557279858688	0	0	
4	891327558926688256	0	0	
5	891087950875897856	0	0	
6	890971913173991426	0	0	
7	890729181411237888	0	0	
8	890609185150312448	0	0	
9	890240255349198849	0	0	

	source	\
0	<a href="http://twitter.com/download/iphone" r...	
1	<a href="http://twitter.com/download/iphone" r...	
2	<a href="http://twitter.com/download/iphone" r...	
3	<a href="http://twitter.com/download/iphone" r...	
4	<a href="http://twitter.com/download/iphone" r...	
5	<a href="http://twitter.com/download/iphone" r...	
6	<a href="http://twitter.com/download/iphone" r...	
7	<a href="http://twitter.com/download/iphone" r...	
8	<a href="http://twitter.com/download/iphone" r...	
9	<a href="http://twitter.com/download/iphone" r...	

	expanded_urls	rating_numerator	\
0	https://twitter.com/dog_rates/status/892420643...	13	
1	https://twitter.com/dog_rates/status/892177421...	13	
2	https://twitter.com/dog_rates/status/891815181...	12	
3	https://twitter.com/dog_rates/status/891689557...	13	
4	https://twitter.com/dog_rates/status/891327558...	12	
5	https://twitter.com/dog_rates/status/891087950...	13	
6	https://gofundme.com/ydvmve-surgery-for-jax,ht...	13	
7	https://twitter.com/dog_rates/status/890729181...	13	
8	https://twitter.com/dog_rates/status/890609185...	13	
9	https://twitter.com/dog_rates/status/890240255...	14	

	rating_denominator	name	doggo	floof	pupper	puppo	normalize_rating	\
0	10	Phineas	None	None	None	None	1.3	
1	10	Tilly	None	None	None	None	1.3	
2	10	Archie	None	None	None	None	1.2	
3	10	Darla	None	None	None	None	1.3	
4	10	Franklin	None	None	None	None	1.2	
5	10	None	None	None	None	None	1.3	
6	10	Jax	None	None	None	None	1.3	
7	10	None	None	None	None	None	1.3	
8	10	Zoey	None	None	None	None	1.3	

9	10	Cassie	doggo	None	None	None	1.4
---	----	--------	-------	------	------	------	-----

	date	time	content \
0	2017-08-01	16:23:56	This is Phineas. He's a mystical boy. Only eve...
1	2017-08-01	00:17:27	This is Tilly. She's just checking pup on you...
2	2017-07-31	00:18:03	This is Archie. He is a rare Norwegian Pouncin...
3	2017-07-30	15:58:51	This is Darla. She commenced a snooze mid meal...
4	2017-07-29	16:00:24	This is Franklin. He would like you to stop ca...
5	2017-07-29	00:08:17	Here we have a majestic great white breaching ...
6	2017-07-28	16:27:12	Meet Jax. He enjoys ice cream so much he gets ...
7	2017-07-28	00:22:40	When you watch your owner call another dog a g...
8	2017-07-27	16:25:51	This is Zoey. She doesn't want to be one of th...
9	2017-07-26	15:59:51	This is Cassie. She is a college pup. Studying...

	content_url_1	content_url_2
0	https://t.co/MgUWQ76dJU	NaN
1	https://t.co/0Xxu71qeIV	NaN
2	https://t.co/wUnZnhtVJB	NaN
3	https://t.co/tD36da7qLQ	NaN
4	https://t.co/AtUZn91f7f	NaN
5	https://t.co/kQ04fDDRmh	NaN
6	https://t.co/Zr4hWfAs1H	https://t.co/tVJBRMnhxl
7	https://t.co/vOnONBcwqx	NaN
8	https://t.co/9TwLuAGH0b	NaN
9	https://t.co/t1bfwz5S2A	NaN

In [146]: image_filt.head(50)

Out[146]:

	tweet_id	jpg_url \
0	666020888022790149	https://pbs.twimg.com/media/CT4udn0WwAA0aMy.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
5	666050758794694657	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6	666051853826850816	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7	666055525042405380	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8	666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9	666058600524156928	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10	666063827256086533	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11	666071193221509120	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
12	666073100786774016	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13	666082916733198337	https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg
14	666094000022159362	https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg
15	666099513787052032	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16	666102155909144576	https://pbs.twimg.com/media/CT54YGiWUAEZnOk.jpg
17	666104133288665088	https://pbs.twimg.com/media/CT56LSZW0AALJj2.jpg
18	666268910803644416	https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg

19 666273097616637952 <https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg>
20 666287406224695296 <https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg>
21 666293911632134144 <https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg>
22 666337882303524864 <https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg>
23 666345417576210432 https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg
24 666353288456101888 https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25 666362758909284353 <https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg>
26 666373753744588802 <https://pbs.twimg.com/media/CT9vZEYUAA1ZO5.jpg>
27 666396247373291520 <https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg>
28 666407126856765440 <https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.jpg>
29 666411507551481857 <https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg>
30 666418789513326592 <https://pbs.twimg.com/media/CT-YWb7U8AA7QnN.jpg>
31 666421158376562688 <https://pbs.twimg.com/media/CT-aggCXAAMfT3.jpg>
32 666428276349472768 <https://pbs.twimg.com/media/CT-g-ODUwAEQdSn.jpg>
33 666430724426358785 <https://pbs.twimg.com/media/CT-jNYqW4AAPi2M.jpg>
34 666435652385423360 <https://pbs.twimg.com/media/CT-nsTQWEAEkyDn.jpg>
35 666437273139982337 <https://pbs.twimg.com/media/CT-pKmrWIAAxUWj.jpg>
36 666447344410484738 <https://pbs.twimg.com/media/CT-yU5QWwAEjLX5.jpg>
37 666454714377183233 <https://pbs.twimg.com/media/CT-5Bs-WUAA2JeC.jpg>
38 666444823164719104 https://pbs.twimg.com/media/CUB16IwVAAA9_zT.jpg
39 666449482315059201 <https://pbs.twimg.com/media/CUBqKnLwWAA50QB.jpg>
40 666691418707132416 <https://pbs.twimg.com/media/CUCQTpEWEAA7EDz.jpg>
41 666701168228331520 <https://pbs.twimg.com/media/CUCZLH1UAAAeAig.jpg>
42 666739327293083650 <https://pbs.twimg.com/media/CUC74aTWoAInZey.jpg>
43 666776908487630848 <https://pbs.twimg.com/media/CUDeDoWUYAAD-EM.jpg>
44 666781792255496192 https://pbs.twimg.com/media/CUDigRXXIAATI_H.jpg
45 666786068205871104 <https://pbs.twimg.com/media/CUDmZikWcAAIPPe.jpg>
46 666804364988780544 <https://pbs.twimg.com/media/CUD3A7YW0AA82N0.jpg>
47 666817836334096384 <https://pbs.twimg.com/media/CUEDSMEEWAAuXVZ.jpg>
48 666826780179869698 <https://pbs.twimg.com/media/CUELa0NUkAAscGC.jpg>
49 666835007768551424 <https://pbs.twimg.com/media/CUES51dXIAEahyG.jpg>

	img_num		p1	p1_conf	p1_dog	\
0	1	welsh springer spaniel	0.465074	True		
1	1	redbone	0.506826	True		
2	1	german shepherd	0.596461	True		
3	1	rhodesian ridgeback	0.408143	True		
4	1	miniature pinscher	0.560311	True		
5	1	bernese mountain dog	0.651137	True		
6	1	box turtle	0.933012	False		
7	1	chow	0.692517	True		
8	1	shopping cart	0.962465	False		
9	1	miniature poodle	0.201493	True		
10	1	golden retriever	0.775930	True		
11	1	gordon setter	0.503672	True		
12	1	walker hound	0.260857	True		
13	1	pug	0.489814	True		
14	1	bloodhound	0.195217	True		

15	1	lhasa	0.582330	True
16	1	english setter	0.298617	True
17	1	hen	0.965932	False
18	1	desktop computer	0.086502	False
19	1	italian greyhound	0.176053	True
20	1	maltese dog	0.857531	True
21	1	three toed sloth	0.914671	False
22	1	ox	0.416669	False
23	1	golden retriever	0.858744	True
24	1	malamute	0.336874	True
25	1	guinea pig	0.996496	False
26	1	soft coated wheaten terrier	0.326467	True
27	1	chihuahua	0.978108	True
28	1	black and tan coonhound	0.529139	True
29	1	coho	0.404640	False
30	1	toy terrier	0.149680	True
31	1	blenheim spaniel	0.906777	True
32	1	pembroke	0.371361	True
33	1	llama	0.505184	False
34	1	chesapeake bay retriever	0.184130	True
35	1	chihuahua	0.671853	True
36	1	curly coated retriever	0.322084	True
37	1	dalmatian	0.278954	True
38	1	ibizan hound	0.044333	True
39	1	border collie	0.447803	True
40	1	german shepherd	0.975401	True
41	1	labrador retriever	0.887707	True
42	1	miniature poodle	0.546933	True
43	1	seat belt	0.375057	False
44	1	italian greyhound	0.618316	True
45	1	snail	0.999888	False
46	1	english setter	0.328792	True
47	1	miniature schnauzer	0.496953	True
48	1	maltese dog	0.359383	True
49	1	airedale	0.448459	True

	p2	p2_conf	p2_dog	p3 \
0	collie	0.156665	True	shetland sheepdog
1	miniature pinscher	0.074192	True	rhodesian ridgeback
2	malinois	0.138584	True	bloodhound
3	redbone	0.360687	True	miniature pinscher
4	rottweiler	0.243682	True	doberman
5	english springer	0.263788	True	greater swiss mountain dog
6	mud turtle	0.045885	False	terrapin
7	tibetan mastiff	0.058279	True	fur coat
8	shopping basket	0.014594	False	golden retriever
9	komondor	0.192305	True	soft coated wheaten terrier
10	tibetan mastiff	0.093718	True	labrador retriever

11	yorkshire terrier	0.174201	True	pekinese
12	english foxhound	0.175382	True	ibizan hound
13	bull mastiff	0.404722	True	french bulldog
14	german shepherd	0.078260	True	malinois
15	shih tzu	0.166192	True	dandie dinmont
16	newfoundland	0.149842	True	borzoi
17	cock	0.033919	False	partridge
18	desk	0.085547	False	bookcase
19	toy terrier	0.111884	True	basenji
20	toy poodle	0.063064	True	miniature poodle
21	otter	0.015250	False	great grey owl
22	newfoundland	0.278407	True	groenendael
23	chesapeake bay retriever	0.054787	True	labrador retriever
24	siberian husky	0.147655	True	eskimo dog
25	skunk	0.002402	False	hamster
26	afghan hound	0.259551	True	briard
27	toy terrier	0.009397	True	papillon
28	bloodhound	0.244220	True	flat coated retriever
29	barracouta	0.271485	False	gar
30	papillon	0.148258	True	chihuahua
31	cocker spaniel	0.090346	True	shih tzu
32	chow	0.249394	True	pomeranian
33	irish terrier	0.104109	True	dingo
34	chain saw	0.056775	False	power drill
35	beagle	0.124680	True	saluki
36	giant schnauzer	0.287955	True	labrador retriever
37	labrador retriever	0.237612	True	great pyrenees
38	pembroke	0.043209	True	west highland white terrier
39	english springer	0.170497	True	collie
40	beagle	0.008687	True	bloodhound
41	chihuahua	0.029307	True	french bulldog
42	cocker spaniel	0.165255	True	toy poodle
43	miniature pinscher	0.167175	True	chihuahua
44	weimaraner	0.151363	True	vizsla
45	slug	0.000055	False	acorn
46	brittany spaniel	0.283545	True	ibizan hound
47	standard schnauzer	0.285276	True	giant schnauzer
48	teddy	0.148759	False	west highland white terrier
49	toy poodle	0.124030	True	teddy

	p3_conf	p3_dog
0	0.061428	True
1	0.072010	True
2	0.116197	True
3	0.222752	True
4	0.154629	True
5	0.016199	True
6	0.017885	False

7	0.054449	False
8	0.007959	True
9	0.082086	True
10	0.072427	True
11	0.109454	True
12	0.097471	True
13	0.048960	True
14	0.075628	True
15	0.089688	True
16	0.133649	True
17	0.000052	False
18	0.079480	False
19	0.111152	True
20	0.025581	True
21	0.013207	False
22	0.102643	True
23	0.014241	True
24	0.093412	True
25	0.000461	False
26	0.206803	True
27	0.004577	True
28	0.173810	True
29	0.189945	False
30	0.142860	True
31	0.001117	True
32	0.241878	True
33	0.062071	False
34	0.036763	False
35	0.044094	True
36	0.166331	True
37	0.171106	True
38	0.038906	True
39	0.139206	True
40	0.005394	True
41	0.020756	True
42	0.095959	True
43	0.086951	True
44	0.085989	True
45	0.000026	False
46	0.057461	True
47	0.073764	True
48	0.106007	True
49	0.110183	False

In [147]: rt_fav.head()

Out[147]:	tweet_id	retweet_count	favorite_count	favorite_retweet_ratio
0	892420643555336193	7740	36366	4.698450

1	892177421306343426	5730	31326	5.467016
2	891815181378084864	3791	23598	6.224743
3	891689557279858688	7902	39669	5.020121
4	891327558926688256	8524	37867	4.442398

0.10 data visualisation

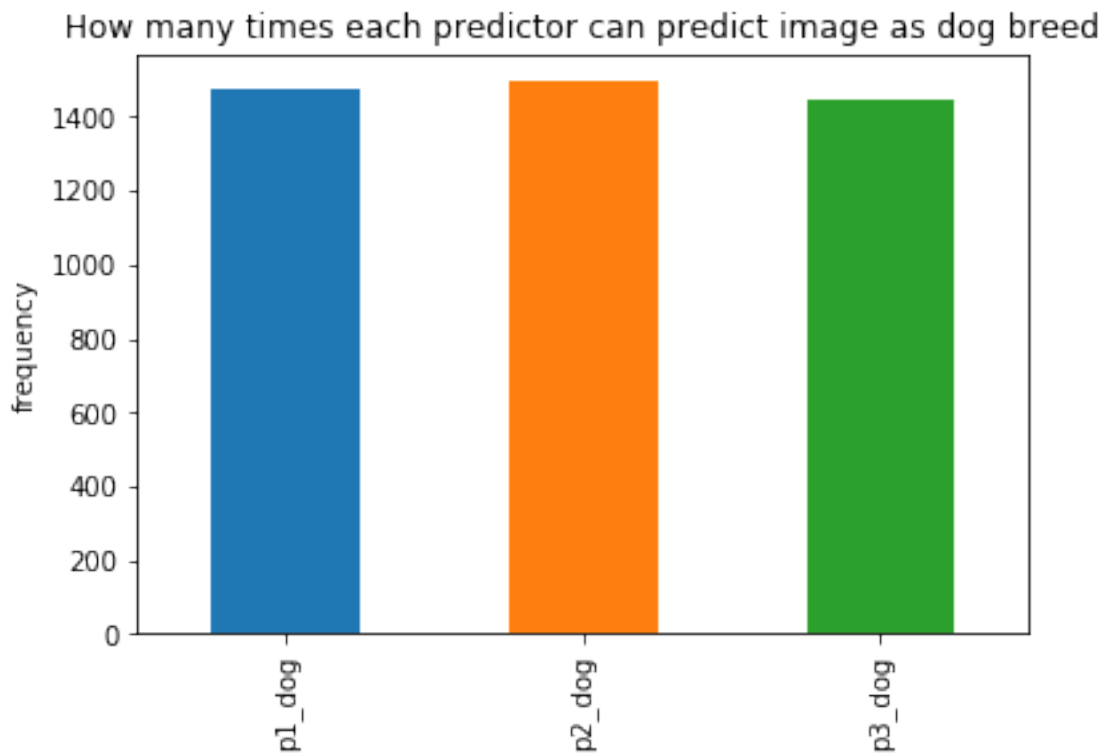
0.10.1 visualizing how often each predictor can predict image as a breed dog

```
In [148]: import matplotlib.pyplot as plt
          %matplotlib inline
```

```
In [149]: breed_predict = image_filt[['p1_dog', 'p2_dog', 'p3_dog']]
```

```
In [150]: breed_predict.sum().plot(kind='bar');
          plt.title('How many times each predictor can predict image as dog breed')
          plt.ylabel('frequency')
```

```
Out[150]: Text(0,0.5,'frequency')
```



The best three predictors can predict given image as dog breed with similar performance with small differences.

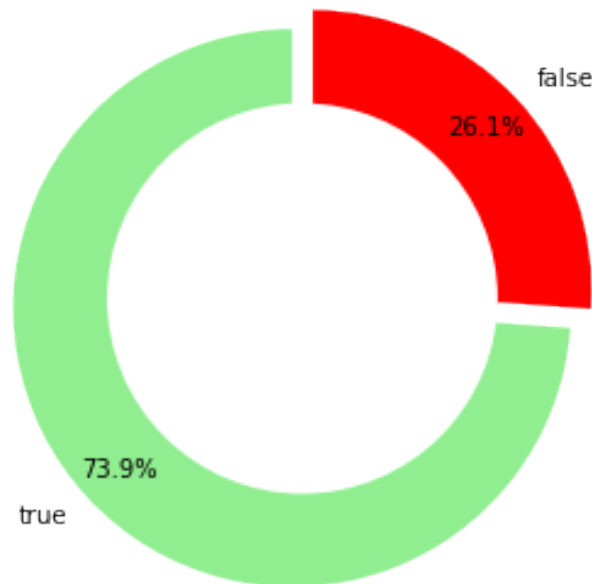
0.10.2 Visualizing mean true proportion of dog breed prediction

```
In [151]: #calculationg true and false prediction proportion
true_prop = breed_predict.sum().mean()/breed_predict.shape[0]
false_prop = 1 - true_prop

In [152]: # Pie chart of proporsion
labels = ['true', 'false']
sizes = [true_prop, false_prop]
colors = ['#90EE90', '#FF0000']
explode = (0.05, 0.05)
fig1, ax1 = plt.subplots()

plt.pie(sizes, colors = colors, labels=labels, autopct='%1.1f%%', pctdistance=0.85,
centre_circle = plt.Circle((0,0),0.70,fc='white'))
fig = plt.gcf()
fig.gca().add_artist(centre_circle)
ax1.axis('equal')
plt.tight_layout()
plt.title('true/false prediction of dog breed proporsion aggregation of all predictors')
plt.show()
```

true/false prediction of dog breed proporsion aggregation of all predictors



0.10.3 Normalize tweet rating

```
In [153]: # ensure there are no 0 value in rating_denominator
twit_ac[twit_ac['rating_denominator']==0]
```

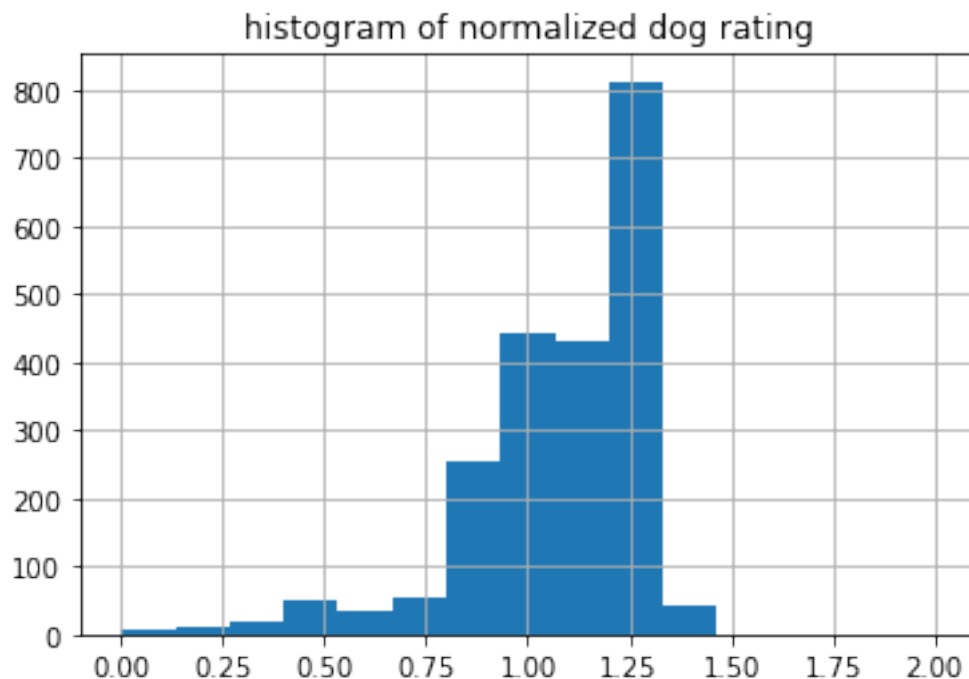
```
Out[153]: Empty DataFrame
Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, source, expanded_urls,
Index: []
```

Seeing that dog rating has different scales, normalization becomes very important to compare each dog rating. Then, normalization can be calculated by formula = rating_numerator / rating_denominator.

```
In [154]: twit_ac['normalize_rating'].describe()
```

```
Out[154]: count      2175.000000
mean         1.223433
std          4.246755
min          0.000000
25%          1.000000
50%          1.100000
75%          1.200000
max         177.600000
Name: normalize_rating, dtype: float64
```

```
In [155]: twit_ac['normalize_rating'].hist(range=(0,2), bins=15);
plt.title('histogram of normalized dog rating')
plt.show()
```



Normalized dog's rating data spreading is almost in normal distribution. There are some data outline that is proved by seeing describe data results. The max value of this parameter is 177.6 that is very far from mean data 1.22. Checking the data outline condition is needed to make data cleaner.


```
In [156]: twit_ac['normalize_rating'].sort_values(ascending=False)
```

```
Out[156]: 804      177.600000
          163      66.600000
          162      42.000000
          1895     42.000000
          246     18.200000
          558       7.500000
          429      3.428571
          617      2.700000
          1533     2.600000
           52      1.700000
          247      1.500000
           38      1.400000
           53      1.400000
          529      1.400000
          875      1.400000
          310      1.400000
          313      1.400000
          753      1.400000
           98      1.400000
          357      1.400000
          352      1.400000
          159      1.400000
          183      1.400000
           61      1.400000
          105      1.400000
          180      1.400000
          160      1.400000
          161      1.400000
           70      1.400000
          507      1.400000
          ...
          2124     0.300000
          1070     0.300000
           618     0.300000
          1011     0.300000
          2060     0.300000
          2023     0.300000
          1957     0.300000
          2004     0.300000
          1422     0.300000
          1513     0.300000
          1900     0.200000
          2067     0.200000
          2168     0.200000
          2129     0.200000
          1741     0.200000
```

```

2058      0.200000
2145      0.200000
1419      0.200000
987       0.200000
1585      0.200000
1582      0.200000
2157      0.100000
1267      0.100000
1859      0.100000
2080      0.100000
1690      0.100000
1912      0.100000
1761      0.100000
265       0.000000
840       0.000000
Name: normalize_rating, Length: 2175, dtype: float64

```

```
In [157]: twit_ac.iloc[804,:]
```

```

Out[157]: tweet_id      749981277374128128
in_reply_to_status_id      0
in_reply_to_user_id      0
source      <a href="https://about.twitter.com/products/tw...
expanded_urls      https://twitter.com/dog_rates/status/749981277...
rating_numerator      1776
rating_denominator      10
name      Atticus
doggo      None
floof      None
pupper      None
puppo      None
normalize_rating      177.6
date      2016-07-04
time      15:00:45
content      This is Atticus. He's quite simply America af...
content_url_1      https://t.co/GRXwMxLBkh
content_url_2      NaN
Name: 804, dtype: object

```

rating numerator is very big, let see the tweet content to prve this value.

```
In [158]: twit_ac.loc[804,'content']
```

```
Out[158]: "This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkh"
```

it is true that user has given that kind of score for the dog. Then to prove it in the unusual value, we need to print content on row 163, 162, 1895

```
In [159]: rows_sel = [162, 163, 1895]
```

```
for i in rows_sel:
    print(twit_ac.content[i])
```

@dhmontgomery We also gave snoop dogg a 420/10 but I think that predated your research
@s8n You tried very hard to portray this good boy as not so good, but you have ultimately failed
After so many requests... here you go.

All data is valid by seeing content.

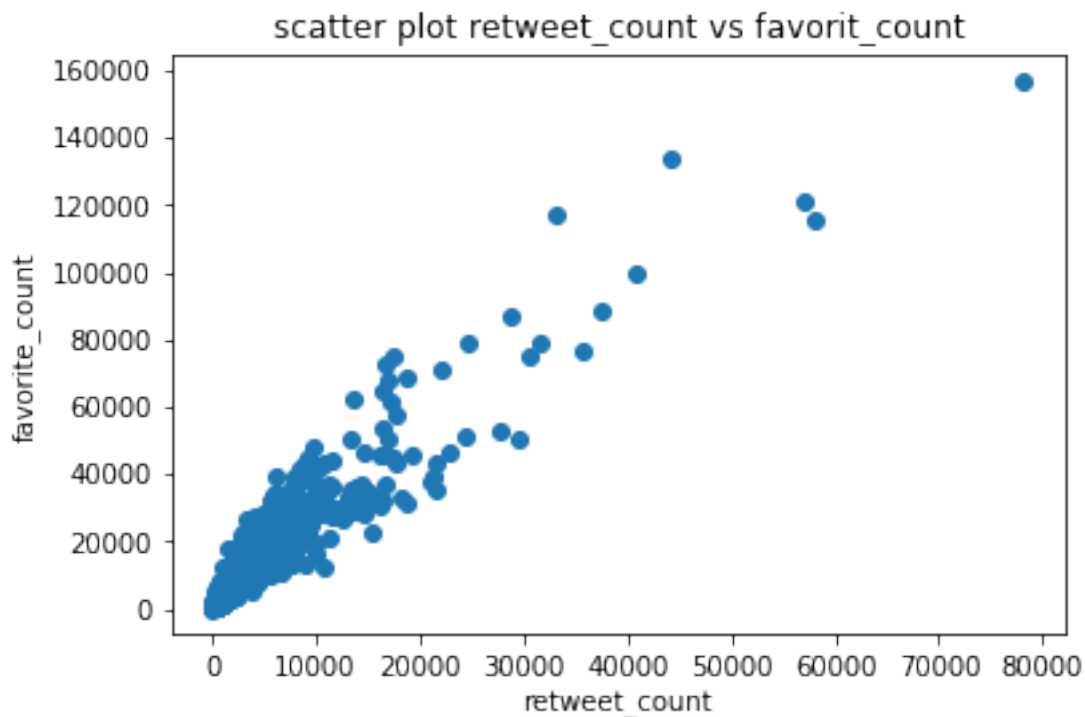
0.10.4 RT and favorite count correlation

```
In [160]: rt_fav.head()
```

```
Out[160]:
```

	tweet_id	retweet_count	favorite_count	favorite_retweet_ratio
0	892420643555336193	7740	36366	4.698450
1	892177421306343426	5730	31326	5.467016
2	891815181378084864	3791	23598	6.224743
3	891689557279858688	7902	39669	5.020121
4	891327558926688256	8524	37867	4.442398

```
In [161]: plt.scatter(rt_fav['retweet_count'], rt_fav['favorite_count']);
plt.title('scatter plot retweet_count vs favorit_count')
plt.xlabel('retweet_count')
plt.ylabel('favorite_count')
plt.show()
```



it seems that retweet_count and favorite_count has positive correlation. This assumption is based on scatter plot above. It can be seen that the more big retweet_count the more value favorite_count is. The correlation between two variables can be calculated in calculation method below.

```
In [162]: rt_fav[['retweet_count', 'favorite_count']].corr(method='pearson')
```

```
Out[162]:
```

	retweet_count	favorite_count
retweet_count	1.000000	0.926974
favorite_count	0.926974	1.000000

Computed correlation between those variables is 0.927, which is proved that those variables has high enough positive correlation.

```
In [ ]:
```