

wrangle_act

May 20, 2020

1 Project Objective:

Gathering Data from three sources, and assess them visually and programmatically for quality and tidiness issues. Detect and document at least eight (8) quality issues and two (2) tidiness issues in your `wrangle_act.ipynb` Jupyter Notebook

Clean each of the issues you documented while assessing. Perform this cleaning in **wrangle_act.ipynb** as well. The result should be a high quality and tidy master pandas DataFrame (or DataFrames, if appropriate)

Store the clean DataFrame(s) in a CSV file with the main one named **twitter_archive_master.csv**.

Analyze and visualize your wrangled data in your **wrangle_act.ipynb** Jupyter Notebook. At least **three (3) insights and one (1) visualization** must be produced.

Create a 300-600 word written report called **wrangle_report.pdf** or **wrangle_report.html** that briefly describes your wrangling efforts. This is to be framed as an internal document.

Create a 250-word-minimum written report called **act_report.pdf** or **act_report.html** that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

1.0.1 Final deliverables:

1. `wrangle_act.ipynb`,
2. `twitter_archive_master.csv`
3. `wrangle_report.pdf`
4. `wrangle_report.html`

1.0.2 Importing Data and packages

```
In [1]: #importing packages
import pandas as pd
import numpy as np
import json
import matplotlib.pyplot as plt
% matplotlib inline
```

```

In [2]: #importing twitter-archive-enhanced data
        twitter_arch = pd.read_csv("twitter-archive-enhanced.csv")

In [3]: #verifying if data is imported
        twitter_arch.head(1)

Out[3]:
          tweet_id  in_reply_to_status_id  in_reply_to_user_id \
0  892420643555336193                NaN                NaN

          timestamp \
0  2017-08-01 16:23:56 +0000

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

          text  retweeted_status_id \
0  This is Phineas. He's a mystical boy. Only eve...      NaN

          retweeted_status_user_id retweeted_status_timestamp \
0                NaN                NaN

          expanded_urls  rating_numerator \
0  https://twitter.com/dog_rates/status/892420643...      13

          rating_denominator  name doggo floofer pupper puppo
0                10  Phineas  None  None  None  None

In [4]: # Download the image prediction file using the link provided to Udacity students
import requests
url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictio
image_request = requests.get(url, allow_redirects=True)

open('image_predictions.tsv', 'wb').write(image_request.content)

Out[4]: 335079

In [5]: #importing image_predictions.tsv file
        image = pd.read_csv('image_predictions.tsv', sep='\t')

In [6]: #verifying if data is imported
        image.head(1)

Out[6]:
          tweet_id                jpg_url \
0  666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg

          img_num          p1  p1_conf  p1_dog          p2  p2_conf \
0           1  Welsh_springer_spaniel  0.465074    True  collie  0.156665

          p2_dog          p3  p3_conf  p3_dog
0    True  Shetland_sheepdog  0.061428    True

```

```

In [7]: #importing tweet_json.txt (json file)
        tweet_json=pd.read_json('tweet_json.txt',lines=True)

In [8]: #verifying if data is imported
        tweet_json.head(1)

Out[8]: contributors coordinates created_at display_text_range \
0          NaN          NaN 2017-08-01 16:23:56          [0, 85]

        entities \
0 {'hashtags': [], 'symbols': [], 'user_mentions...

        extended_entities favorite_count \
0 {'media': [{'id': 892420639486877696, 'id_str'... 39467

        favorited full_text geo \
0      False This is Phineas. He's a mystical boy. Only eve... NaN

        ... \
0      ...

        possibly_sensitive_appealable quoted_status quoted_status_id \
0                                0.0          NaN          NaN

        quoted_status_id_str retweet_count retweeted retweeted_status \
0          NaN          8853      False          NaN

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

        user
0 {'id': 4196983835, 'id_str': '4196983835', 'na...

[1 rows x 31 columns]

```

2 Assessing Data

My Goal in this section is to assess three Dataframes (twitter_arch , image , tweet_json) for Quality and Tidiness issues. After analysing in this section I will clean those issues in Cleaning Section

```
In [9]: twitter_arch.shape
```

```
Out[9]: (2356, 17)
```

```
In [10]: twitter_arch.head()
```

```
Out[10]: 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 \
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

	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...

	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	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	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

	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

In [11]: twitter_arch.sample(5)

```

Out[11]:
      tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
718    783466772167098368                NaN                NaN
1369   702332542343577600                NaN                NaN
942    752917284578922496                NaN                NaN
1650   683515932363329536                NaN                NaN
1749   679047485189439488                NaN                NaN

      timestamp  \
718    2016-10-05 00:40:09 +0000
1369   2016-02-24 03:21:41 +0000
942    2016-07-12 17:27:23 +0000
1650   2016-01-03 05:11:12 +0000
1749   2015-12-21 21:15:11 +0000

      source  \
718    <a href="http://twitter.com/download/iphone" r...
1369   <a href="http://vine.co" rel="nofollow">Vine -...
942    <a href="http://twitter.com/download/iphone" r...
1650   <a href="http://vine.co" rel="nofollow">Vine -...
1749   <a href="http://twitter.com/download/iphone" r...

      text  retweeted_status_id  \
718    This is Finn. He likes eavesdropping from fili...      NaN
1369   This is Rudy. He's going to be a star. 13/10 t...      NaN
942    This is Grizzzie. She's a semi-submerged Bahrai...      NaN
1650   HEY PUP WHAT'S THE PART OF THE HUMAN BODY THAT...      NaN
1749   This dog doesn't know how to stairs. Quite tra...      NaN

      retweeted_status_user_id  retweeted_status_timestamp  \
718                        NaN                        NaN
1369                       NaN                        NaN
942                        NaN                        NaN
1650                       NaN                        NaN
1749                       NaN                        NaN

      expanded_urls  rating_numerator  \
718    https://twitter.com/dog_rates/status/783466772...      11
1369                https://vine.co/v/irlDujgw0jd          13
942    https://twitter.com/dog_rates/status/752917284...      11
1650                https://vine.co/v/ibvnzrauFuV          11
1749    https://twitter.com/dog_rates/status/679047485...      9

      rating_denominator  name  doggo  floofer  pupper  puppo
718                    10    Finn   None    None    None    None
1369                   10    Rudy   None    None    None    None
942                    10  Grizzzie  None    None    None    None
1650                   10     None   None    None    None    None
1749                   10     None   None    None    None    None

```

```
In [12]: twitter_arch.tail()
```

```
Out[12]:
```

	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
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

Assessing twitter_arch

```
In [13]: twitter_arch.info()
```

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

```
In [14]: twitter_arch.describe()
```

```
Out[14]:
```

	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

```
In [15]: #verifying if tweet ids are unique
         twitter_arch["tweet_id"].nunique()
```

```
Out[15]: 2356
```

```
In [16]: twitter_arch["timestamp"].nunique()
```

```
Out[16]: 2356
```

```
In [17]: twitter_arch["source"].unique()
```

```
Out[17]: array([ '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
                '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>',
                '<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>',
                '<a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck
```

```
In [18]: twitter_arch.text.nunique()
```

```
Out[18]: 2356
```

```
In [19]: twitter_arch[twitter_archretweeted_status_id.isnull()]
```

```
Out[19]:
```

	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
20	888078434458587136	NaN	NaN
21	887705289381826560	NaN	NaN
22	887517139158093824	NaN	NaN
23	887473957103951883	NaN	NaN
24	887343217045368832	NaN	NaN
25	887101392804085760	NaN	NaN
26	886983233522544640	NaN	NaN
27	886736880519319552	NaN	NaN
28	886680336477933568	NaN	NaN
29	886366144734445568	NaN	NaN
30	886267009285017600	8.862664e+17	2.281182e+09
...
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
20	2017-07-20 16:49:33 +0000
21	2017-07-19 16:06:48 +0000
22	2017-07-19 03:39:09 +0000
23	2017-07-19 00:47:34 +0000
24	2017-07-18 16:08:03 +0000
25	2017-07-18 00:07:08 +0000
26	2017-07-17 16:17:36 +0000
27	2017-07-16 23:58:41 +0000
28	2017-07-16 20:14:00 +0000
29	2017-07-15 23:25:31 +0000
30	2017-07-15 16:51:35 +0000
...	...
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...
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...
  
```

```

30      <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
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/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
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
30	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
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	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
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

[2175 rows x 17 columns]

In [20]: twitter_arch[twitter_arch.in_reply_to_status_id.isnull()]

```

Out[20]:
      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

[2278 rows x 17 columns]

In [21]: twitter_arch[twitter_arch.retweeted_status_id.notnull()]

Out[21]:

	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...

[illegible]

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 Tw...
 2260 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_hill987: @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/loganammosis/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]

All rows corresponding to retweet can be deleted as they do not have tweet. these rows just have retweeted information

In [22]: `twitter_arch.expanded_urls.nunique()`

Out[22]: 2218

In [23]: `twitter_arch.expanded_urls.isnull().sum()`

Out[23]: 59

In [24]: `twitter_arch.expanded_urls.head()`

Out[24]: 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...
Name: expanded_urls, dtype: object

In [25]: `twitter_arch[twitter_arch.expanded_urls.isnull()]`

Out[25]:

	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
179	857214891891077121	8.571567e+17	1.806710e+08
185	856330835276025856	NaN	NaN
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
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
375	828361771580813312	NaN	NaN
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
570	801854953262350336	8.018543e+17	1.185634e+07
576	800859414831898624	8.008580e+17	2.918590e+08
611	797165961484890113	7.971238e+17	2.916630e+07
701	786051337297522688	7.727430e+17	7.305050e+17
707	785515384317313025	NaN	NaN
843	766714921925144576	7.667118e+17	4.196984e+09
857	763956972077010945	7.638652e+17	1.584641e+07
967	750381685133418496	7.501805e+17	4.717297e+09
1005	747651430853525504	7.476487e+17	4.196984e+09
1080	738891149612572673	7.384119e+17	3.589728e+08
1295	707983188426153984	7.079801e+17	2.319108e+09
1345	704491224099647488	7.044857e+17	2.878549e+07
1445	696518437233913856	NaN	NaN
1446	696490539101908992	6.964887e+17	4.196984e+09
1474	693644216740769793	6.936422e+17	4.196984e+09
1479	693582294167244802	6.935722e+17	1.198989e+09
1497	692423280028966913	6.924173e+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
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

1844	675849018447167488	6.758457e+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
2038	671550332464455680	6.715449e+17	4.196984e+09
2149	669684865554620416	6.693544e+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
179	2017-04-26 12:48:51 +0000
185	2017-04-24 02:15:55 +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
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
375	2017-02-05 21:56:51 +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
570	2016-11-24 18:28:13 +0000
576	2016-11-22 00:32:18 +0000
611	2016-11-11 19:55:50 +0000
701	2016-10-12 03:50:17 +0000
707	2016-10-10 16:20:36 +0000
843	2016-08-19 19:14:16 +0000
857	2016-08-12 04:35:10 +0000
967	2016-07-05 17:31:49 +0000
1005	2016-06-28 04:42:46 +0000
1080	2016-06-04 00:32:32 +0000
1295	2016-03-10 17:35:20 +0000
1345	2016-03-01 02:19:31 +0000

1445 2016-02-08 02:18:30 +0000
 1446 2016-02-08 00:27:39 +0000
 1474 2016-01-31 03:57:23 +0000
 1479 2016-01-30 23:51:19 +0000
 1497 2016-01-27 19:05:49 +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
 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
 1844 2015-12-13 01:25:37 +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
 2038 2015-12-01 04:44:10 +0000
 2149 2015-11-26 01:11:28 +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...
 179 <a href="http://twitter.com/download/iphone" r...
 185 <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...
 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...
 375 Tw...
 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...
 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...
 701 <a href="http://twitter.com/download/iphone" r...
 707 <a href="http://twitter.com/download/iphone" r...
 843 <a href="http://twitter.com/download/iphone" r...
 857 <a href="http://twitter.com/download/iphone" r...
 967 <a href="http://twitter.com/download/iphone" r...
 1005 <a href="http://twitter.com/download/iphone" r...
 1080 <a href="http://twitter.com/download/iphone" r...
 1295 <a href="http://twitter.com/download/iphone" r...
 1345 <a href="http://twitter.com/download/iphone" r...
 1445 <a href="http://twitter.com/download/iphone" r...
 1446 <a href="http://twitter.com/download/iphone" r...
 1474 <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...
 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...
 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...
 1844 <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...
 2038 <a href="http://twitter.com/download/iphone" r...
 2149 <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
179	@Marc_IRL pixelated af 12/10	NaN
185	RT @Jenna_Marbles: @dog_rates Thanks for ratin...	8.563302e+17
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
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	@jonny_sun @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
375	Beebop and Doobert should start a band 12/10 w...	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
570	.@NBCSports OMG THE TINY HAT I'M GOING TO HAVE...	NaN
576	@SkyWilliams doggo simply protecting you from ...	NaN
611	@JODYHiGROLLER it may be an 11/10 but what do...	NaN
701	13/10 for breakdancing puppo @shibbnbot	NaN
707	Today, 10/10, should be National Dog Rates Day	NaN
843	His name is Charley and he already has a new s...	NaN
857	@TheEllenShow I'm not sure if you know this bu...	NaN
967	13/10 such a good doggo\n@spaghemily	NaN
1005	Other pupper asked not to have his identity sh...	NaN
1080	@mount_alex3 13/10	NaN
1295	@serial @MrRoles OH MY GOD I listened to all o...	NaN
1345	13/10 hero af\n@ABC	NaN
1445	Oh my god 10/10 for every little hot dog pupper	NaN
1446	After reading the comments I may have overesti...	NaN
1474	BREAKING PUPDATE: I've just been notified that...	NaN
1479	Personally I'd give him an 11/10. Not sure why...	NaN
1497	PUPDATE: just noticed this dog has some extra ...	NaN
1523	12/10 @LightningHolt	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
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
1844	This dog is being demoted to a 9/10 for not we...	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
2038	After 22 minutes of careful deliberation this ...	NaN
2149	After countless hours of research and hundreds...	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	expanded_urls	\
30	NaN	NaN	NaN	
55	NaN	NaN	NaN	
64	NaN	NaN	NaN	
113	NaN	NaN	NaN	
148	NaN	NaN	NaN	
179	NaN	NaN	NaN	
185	66699013.0	2017-04-24 02:13:14 +0000	NaN	
186	NaN	NaN	NaN	
188	NaN	NaN	NaN	
189	NaN	NaN	NaN	
218	NaN	NaN	NaN	
228	NaN	NaN	NaN	
234	NaN	NaN	NaN	
274	NaN	NaN	NaN	
290	NaN	NaN	NaN	
291	NaN	NaN	NaN	
313	NaN	NaN	NaN	
342	NaN	NaN	NaN	
346	NaN	NaN	NaN	
375	NaN	NaN	NaN	
387	NaN	NaN	NaN	
409	NaN	NaN	NaN	
427	NaN	NaN	NaN	
498	NaN	NaN	NaN	
513	NaN	NaN	NaN	
570	NaN	NaN	NaN	
576	NaN	NaN	NaN	
611	NaN	NaN	NaN	
701	NaN	NaN	NaN	
707	NaN	NaN	NaN	
843	NaN	NaN	NaN	
857	NaN	NaN	NaN	
967	NaN	NaN	NaN	
1005	NaN	NaN	NaN	
1080	NaN	NaN	NaN	
1295	NaN	NaN	NaN	
1345	NaN	NaN	NaN	
1445	NaN	NaN	NaN	
1446	NaN	NaN	NaN	
1474	NaN	NaN	NaN	
1479	NaN	NaN	NaN	
1497	NaN	NaN	NaN	
1523	NaN	NaN	NaN	
1598	NaN	NaN	NaN	
1605	NaN	NaN	NaN	
1618	NaN	NaN	NaN	

1663	NaN	NaN	NaN
1689	NaN	NaN	NaN
1774	NaN	NaN	NaN
1819	NaN	NaN	NaN
1844	NaN	NaN	NaN
1895	NaN	NaN	NaN
1905	NaN	NaN	NaN
1914	NaN	NaN	NaN
1940	NaN	NaN	NaN
2038	NaN	NaN	NaN
2149	NaN	NaN	NaN
2189	NaN	NaN	NaN
2298	NaN	NaN	NaN

	rating_numerator	rating_denominator	name	doggo	floofer	pupper	puppo
30	12	10	None	None	None	None	None
55	17	10	None	None	None	None	None
64	14	10	None	None	None	None	None
113	10	10	None	None	None	None	None
148	12	10	None	None	None	None	None
179	12	10	None	None	None	None	None
185	14	10	None	None	None	None	None
186	14	10	None	None	None	None	None
188	420	10	None	None	None	None	None
189	666	10	None	None	None	None	None
218	13	10	None	None	None	None	None
228	11	10	None	None	None	None	None
234	13	10	None	None	None	None	None
274	10	10	None	None	None	None	None
290	182	10	None	None	None	None	None
291	15	10	None	None	None	None	None
313	960	0	None	None	None	None	None
342	11	15	None	None	None	None	None
346	12	10	None	None	None	None	None
375	12	10	None	None	None	None	None
387	7	10	None	None	None	None	None
409	13	10	None	None	None	None	None
427	13	10	None	None	None	None	None
498	12	10	None	None	None	None	None
513	11	10	None	None	None	None	None
570	11	10	None	None	None	None	None
576	11	10	None	doggo	None	None	None
611	11	10	None	None	None	None	None
701	13	10	None	None	None	None	puppo
707	10	10	None	None	None	None	None
843	13	10	None	None	None	None	None
857	12	10	None	doggo	None	None	None
967	13	10	None	doggo	None	None	None

1005	12	10	None	None	None	pupper	None
1080	13	10	None	None	None	None	None
1295	12	10	None	None	None	None	None
1345	13	10	None	None	None	None	None
1445	10	10	None	None	None	pupper	None
1446	1	10	None	None	None	None	None
1474	10	10	None	None	None	None	None
1479	11	10	None	None	None	None	None
1497	9	10	None	None	None	None	None
1523	12	10	None	None	None	None	None
1598	4	20	None	None	None	None	None
1605	14	10	None	None	None	None	None
1618	5	10	None	None	None	None	None
1663	20	16	None	None	None	None	None
1689	5	10	None	None	None	None	None
1774	13	10	None	None	None	None	None
1819	7	10	None	None	None	None	None
1844	9	10	None	None	None	None	None
1895	11	10	None	None	None	None	None
1905	13	10	None	None	None	None	None
1914	13	10	None	None	None	None	None
1940	1	10	None	None	None	None	None
2038	1	10	None	None	None	None	None
2149	11	10	None	None	None	None	None
2189	12	10	None	None	None	None	None
2298	10	10	None	None	None	None	None

In [26]: twitter_arch.rating_numerator.value_counts()

```
Out[26]: 12      558
         11      464
         10      461
         13      351
          9      158
          8      102
          7       55
         14       54
          5       37
          6       32
          3       19
          4       17
          1        9
          2        9
        420        2
          0        2
         15        2
         75        2
         80        1
```

```

20      1
24      1
26      1
44      1
50      1
60      1
165     1
84      1
88      1
144     1
182     1
143     1
666     1
960     1
1776    1
17      1
27      1
45      1
99      1
121     1
204     1
Name: rating_numerator, dtype: int64

```

```
In [27]: twitter_arch.rating_numerator.max()
```

```
Out[27]: 1776
```

```
In [28]: sum(twitter_arch.rating_numerator > 20)
```

```
Out[28]: 24
```

```
In [29]: twitter_arch.rating_denominator.max()
```

```
Out[29]: 170
```

```
In [30]: twitter_arch[twitter_arch.rating_denominator == 170]
```

```

Out[30]:
      tweet_id  in_reply_to_status_id  in_reply_to_user_id \
1120  731156023742988288              NaN                NaN

      timestamp \
1120  2016-05-13 16:15:54 +0000

      source \
1120  <a href="http://twitter.com/download/iphone" r...

      text  retweeted_status_id \
1120  Say hello to this unbelievably well behaved sq...      NaN

```

	retweeted_status_user_id	retweeted_status_timestamp	\
1120	NaN	NaN	

	expanded_urls	rating_numerator	\
1120	https://twitter.com/dog_rates/status/731156023...	204	

	rating_denominator	name	doggo	floofer	pupper	puppo
1120	170	this	None	None	None	None

In [31]: twitter_arch.floofer.value_counts()

Out[31]: None 2346
floofer 10
Name: floofer, dtype: int64

In [32]: twitter_arch.name.value_counts()

Out[32]: None 745
a 55
Charlie 12
Lucy 11
Cooper 11
Oliver 11
Lola 10
Tucker 10
Penny 10
Bo 9
Winston 9
Sadie 8
the 8
an 7
Daisy 7
Toby 7
Bailey 7
Buddy 7
Milo 6
Stanley 6
Dave 6
Oscar 6
Koda 6
Scout 6
Leo 6
Jax 6
Bella 6
Jack 6
Rusty 6
Finn 5
...
Lassie 1

Ralphé	1
Kota	1
Rose	1
Sweets	1
Tug	1
Vixen	1
Holly	1
Charl	1
Skye	1
Jay	1
Philippe	1
old	1
Mosby	1
Luther	1
Eleanor	1
Brian	1
Bones	1
Naphaniel	1
Ambrose	1
Grey	1
Dale	1
Dunkin	1
Gunner	1
Mo	1
Snoop	1
Henry	1
Cal	1
Brownie	1
Griswold	1

Name: name, Length: 957, dtype: int64

```
In [33]: twitter_arch[twitter_arch.name.str.islower()]["name"]
```

```
Out[33]: 22      such
56      a
118     quite
169     quite
193     quite
335     not
369     one
542  incredibly
649     a
682     mad
759     an
773     very
801     a
819     very
822     just
```

852	my
924	one
988	not
992	his
993	one
1002	a
1004	a
1017	a
1025	an
1031	very
1040	actually
1049	a
1063	just
1071	getting
1095	mad
	...
2191	a
2198	a
2204	an
2211	a
2212	the
2218	a
2222	a
2235	a
2249	a
2255	a
2264	a
2273	a
2287	a
2304	a
2311	a
2314	a
2326	quite
2327	a
2333	an
2334	a
2335	an
2345	the
2346	the
2347	a
2348	a
2349	an
2350	a
2352	a
2353	a
2354	a

Name: name, Length: 109, dtype: object

3 Notes for Understanding

3.0.1 2356 rows and 17 columns

- tweet_id :Unique Tweet id
- in_reply_to_status_id :78 Non Null - Will delete this column
- in_reply_to_user_id :78 Non Null - Will delete this column
- timestamp :time of tweet - unique
- source :Source of Tweet (Twitter for iPhone, Twitter Web Client, Vine - Make a Scene, Tweet-Deck)
- text :text of tweet
- retweeted_status_id :181 Non Null - Will delete this column
- retweeted_status_user_id :181 Non Null - Will delete this column
- retweeted_status_timestamp :181 Non Null - Will delete this column
- expanded_urls : Tweet Url
- rating_numerator :Rating for dogs (min = 0, max= 1776)
- rating_denominator :Rating out of (min = 0, max = 170)
- name :Given Name of Dog
- doggo :Dog stage (Adult)
- floofer :Dog stage (fluffy)
- pupper :Dog stage (young)
- puppo :Dog stage (transitioning from young to adult)

```
In [34]: twitter_arch.duplicated().sum()
```

```
Out[34]: 0
```

```
In [35]: twitter_arch.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
```

twitter_arch have 2356 rows and 17 columns. This dataset does not have any duplicate columns, Although some columns do have null values in them. In this data I am going to focus on some cleaning issues as given below:

1. in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp have null values, but these columns just have retweet and reply information which I do not need in Analysis so I will delete rows corresponding to this columns.
2. Delete columns from above point.
3. Some rating_numerator are very high values so I will clean them
4. Standardize rating_denominator to 10.
5. Change datatype of timeStamp
6. Changing Datatype of tweet_id
7. clean dog name column - dog name have many non-dog name such as such, 'a','quite','quite','not','one','incredibly', etc and are in lower case
8. Source Type has weblink. I am going to extract source type from this column

3.0.2 Assessing Image Dataframe

```
In [36]: image.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 [37]: image.head()
```

```
Out[37]:
```

	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

```

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

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

```
In [38]: image.duplicated().sum()
```

```
Out[38]: 0
```

```
In [39]: image.p1.value_counts()
```

```

Out[39]: golden_retriever      150
         Labrador_retriever    100
         Pembroke              89
         Chihuahua             83
         pug                   57
         chow                  44
         Samoyed               43
         toy_poodle            39
         Pomeranian            38
         malamute              30
         cocker_spaniel        30
         French_bulldog        26
         miniature_pinscher    23
         Chesapeake_Bay_retriever 23
         seat_belt              22
         Siberian_husky        20
         German_shepherd       20
         Staffordshire_bullterrier 20
         Cardigan              19
         web_site               19
         teddy                  18
         beagle                 18
         Eskimo_dog             18
         Shetland_sheepdog      18
         Maltese_dog            18

```

Shih-Tzu	17
Lakeland_terrier	17
Rottweiler	17
Italian_greyhound	16
kuvasz	16
...	
teapot	1
loupe	1
tiger_shark	1
cowboy_boot	1
suit	1
zebra	1
four-poster	1
sandbar	1
coho	1
cliff	1
studio_couch	1
marmot	1
shopping_basket	1
lawn_mower	1
grey_fox	1
damselfly	1
maze	1
killer_whale	1
hammer	1
earthstar	1
school_bus	1
bearskin	1
military_uniform	1
dining_table	1
barbell	1
park_bench	1
prayer_rug	1
hay	1
boathouse	1
rapeseed	1

Name: p1, Length: 378, dtype: int64

```
In [40]: image.duplicated().sum()
```

```
Out[40]: 0
```

```
In [41]: len(image.query('p1_dog == False or p2_dog == False or p3_dog == False'))
```

```
Out[41]: 832
```

Delet rows in image where prediction doest not match to dog

- tweet_id: unique tweet identifier

- jpg_url: image of the dog
- img_num: number out of 4 possible images (most are 1)
- p1: the algorithm's first prediction for the image in the tweet
- p1_conf: is how confident the algorithm is in its first prediction
- p1_dog: is whether or not the first prediction is a breed of dog
- p2: the algorithm's second most likely prediction
- p3: the algorithm's third most likely prediction

```
In [42]: image.duplicated().sum()
```

```
Out[42]: 0
```

Image Data Frame has 2075 rows and 12 columns. It does not have any duplicate or null values in it. 1. In this Dataframe my goal is to delete rows where predictions does not match with dog in all the three algorithms 2. Image, Update p1, p2, p3 to title text and remove underscores

```
In [43]: tweet_json.sample(5)
```

```
Out[43]:
```

	contributors	coordinates	created_at	display_text_range	\
449	NaN	NaN	2017-01-10 02:30:30	[0, 119]	
342	NaN	NaN	2017-02-16 01:34:34	[0, 139]	
2097	NaN	NaN	2015-11-28 21:34:09	[0, 95]	
1491	NaN	NaN	2016-01-28 16:53:37	[0, 92]	
1623	NaN	NaN	2016-01-06 20:16:44	[0, 99]	

	entities	\
449	{'hashtags': [], 'symbols': [], 'user_mentions': []}	
342	{'hashtags': [], 'symbols': [], 'user_mentions': []}	
2097	{'hashtags': [], 'symbols': [], 'user_mentions': []}	
1491	{'hashtags': [], 'symbols': [], 'user_mentions': []}	
1623	{'hashtags': [], 'symbols': [], 'user_mentions': []}	

	extended_entities	favorite_count	\
449	NaN	0	
342	{'media': [{'id': 769940407350362112, 'id_str': ...}]}	0	
2097	{'media': [{'id': 670717326967205888, 'id_str': ...}]}	1301	
1491	{'media': [{'id': 692752274041507842, 'id_str': ...}]}	7496	
1623	NaN	38551	

	favorited	full_text	geo	\
449	False	RT @dog_rates: This is Seamus. He's very bad a...	NaN	
342	False	RT @dog_rates: This is Klein. These pics were ...	NaN	
2097	False	*screams for a little bit and then crumples to...	NaN	
1491	False	"Hello yes could I get one pupper to go please...	NaN	
1623	False	This little fella really hates stairs. Prefers...	NaN	

	...	\
449	...	
342	...	

2097	...
1491	...
1623	...

	possibly_sensitive_appealable	quoted_status	quoted_status_id \
449	0.0	NaN	NaN
342	0.0	NaN	NaN
2097	0.0	NaN	NaN
1491	0.0	NaN	NaN
1623	0.0	NaN	NaN

	quoted_status_id_str	retweet_count	retweeted \
449	NaN	24069	False
342	NaN	11131	False
2097	NaN	547	False
1491	NaN	4124	False
1623	NaN	24514	False

	retweeted_status \
449	{'created_at': 'Mon Dec 12 16:16:49 +0000 2016...
342	{'created_at': 'Sun Aug 28 16:51:16 +0000 2016...
2097	NaN
1491	NaN
1623	NaN

	source truncated \
449	<a href="http://twitter.com/download/iphone" r... False
342	<a href="http://twitter.com/download/iphone" r... False
2097	<a href="http://twitter.com/download/iphone" r... False
1491	<a href="http://twitter.com/download/iphone" r... False
1623	Vine -... False

	user
449	{'id': 4196983835, 'id_str': '4196983835', 'na...
342	{'id': 4196983835, 'id_str': '4196983835', 'na...
2097	{'id': 4196983835, 'id_str': '4196983835', 'na...
1491	{'id': 4196983835, 'id_str': '4196983835', 'na...
1623	{'id': 4196983835, 'id_str': '4196983835', 'na...

[5 rows x 31 columns]

In []:

In [44]: tweet_json.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 31 columns):
```

```

contributors          0 non-null float64
coordinates            0 non-null float64
created_at            2354 non-null datetime64[ns]
display_text_range    2354 non-null object
entities              2354 non-null object
extended_entities     2073 non-null object
favorite_count        2354 non-null int64
favorited             2354 non-null bool
full_text             2354 non-null object
geo                   0 non-null float64
id                    2354 non-null int64
id_str                2354 non-null int64
in_reply_to_screen_name 78 non-null object
in_reply_to_status_id 78 non-null float64
in_reply_to_status_id_str 78 non-null float64
in_reply_to_user_id    78 non-null float64
in_reply_to_user_id_str 78 non-null float64
is_quote_status        2354 non-null bool
lang                  2354 non-null object
place                 1 non-null object
possibly_sensitive     2211 non-null float64
possibly_sensitive_appealable 2211 non-null float64
quoted_status         28 non-null object
quoted_status_id       29 non-null float64
quoted_status_id_str   29 non-null float64
retweet_count          2354 non-null int64
retweeted              2354 non-null bool
retweeted_status       179 non-null object
source                2354 non-null object
truncated              2354 non-null bool
user                  2354 non-null object
dtypes: bool(4), datetime64[ns](1), float64(11), int64(4), object(11)
memory usage: 505.8+ KB

```

3.0.3 Keeping columns that are required. I am doing it here to make this dataframe more assessible. If I need any column I will come back and include here again.

3.0.4 1. cleaning step - including only columns required for analysis using dataframe

```
In [45]: tweet_jason_clean = tweet_json[['id', 'favorite_count', 'retweet_count']]
```

```
In [46]: tweet_jason_clean.head()
```

```

Out[46]:
   id  favorite_count  retweet_count
0  892420643555336193         39467         8853
1  892177421306343426         33819         6514
2  891815181378084864         25461         4328

```

3	891689557279858688	42908	8964
4	891327558926688256	41048	9774

```
In [47]: tweet_jason_clean.shape
```

```
Out[47]: (2354, 3)
```

```
In [48]: tweet_jason_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 3 columns):
id                2354 non-null int64
favorite_count    2354 non-null int64
retweet_count     2354 non-null int64
dtypes: int64(3)
memory usage: 55.2 KB
```

```
In [49]: tweet_jason_clean.id.nunique()
```

```
Out[49]: 2354
```

- id: unique tweet identifier
- retweet_count: number of retweets a tweet received
- favorite_count: number of favorites a tweets received

```
In [50]: tweet_jason.favorite_count.value_counts()
```

```
Out[50]: 0          179
        610          3
        345          3
        2918         3
        1691         3
        2176         3
        2768         3
        1339         3
        2706         3
        522          2
        3134         2
        1618         2
        250          2
        2250         2
        2660         2
        2262         2
        2305         2
        1111         2
        784          2
        4878         2
```


346	2
14685	2
780	2
6923	2
6515	2
2433	2
3603	2
13518	2
3593	2
1536	2
...	
4681	1
523	1
559	1
802	1
527	1
27154	1
6676	1
535	1
537	1
6682	1
8731	1
23074	1
21029	1
667	1
6696	1
2608	1
35400	1
21041	1
4659	1
10804	1
4099	1
68152	1
10812	1
573	1
6718	1
33345	1
814	1
23108	1
2630	1
8143	1

Name: favorite_count, Length: 2007, dtype: int64

```
In [51]: tweet_json.retweet_count.value_counts()
```

```
Out[51]: 1972    5
         3652    5
         83     5
```

146	4
61	4
748	4
2243	4
336	4
183	4
179	4
1207	4
265	4
115	4
71	4
1124	4
542	4
819	4
577	4
516	4
397	3
619	3
661	3
2511	3
261	3
431	3
482	3
403	3
557	3
572	3
576	3
	..
2088	1
1271	1
2030	1
43	1
5365	1
4143	1
3316	1
1263	1
16439	1
2104	1
4125	1
27	1
4121	1
4119	1
4079	1
1285	1
10226	1
8183	1
2042	1
11524	1

```

6148      1
7         1
1281      1
2060      1
1825      1
8209      1
19        1
2068      1
30742     1
0         1
Name: retweet_count, Length: 1724, dtype: int64

```

3.1 Tidiness - Messy Data Untidy

1. doggo, floofer, pupper, puppo must make one column
2. Merge three dataframe together to make one dataframe and extract it to .CSV file format

3.2 Quality

1. First Cleaning Step was already performed on Tweet_json to include only Columns required for Analysis. I am going to keep this step as iterative and include column later if required
2. Datatypes of rating_numerator must be float as some of the rating in decimal. I will make denominator-rating also float
3. in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp have null values, and these columns just have retweet and reply information which I do not need in Analysis so I will delete rows corresponding to this columns.
4. Making Twitter_arch Dataframe well structured by deleting columns for which rows are deleted in point 1.
5. Twitter_arch - convert timestamp to datetime object as it may be required for Timeseries Analysis
6. Changing Datatype of tweet_id to str in all Dataframes
7. There is unexpected High values in numerator (1776)
8. Values in Denominator and denominator is more than 10. Standardize rating_denominator to 10.
9. in Twitter archive clean source link to extract source
10. clean dog name column - dog name have many non-dog name such as such', 'a', 'quite', 'quite', 'not', 'one', 'incredibly', etc and are in lower case
11. Image, Update p1, p2, p3 to title text and remove underscores
12. image_predictions, Remove non-dogs, the 324 rows where p1, p2, and p3 are false

4 Cleaning

In this Step I am going to Clean and resolve the issues identified in Assessing Phase

```

In [52]: # Copying twitter_arch to twitter_arch_clean
twitter_arch_clean = twitter_arch.copy()
twitter_arch_clean.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

```

4.1 2. changing rating_numerator, rating_denominator to float using astype()

In this step using `.astype('float')`, I am going to convert datatype of `rating_numerator` and `rating_denominator` from int to float

4.1.1 code

```
In [53]: twitter_arch_clean[["rating_numerator","rating_denominator"]] = twitter_arch_clean[["ra
```

4.1.2 Test

```
In [54]: twitter_arch_clean.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 float64
rating_denominator            2356 non-null float64
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(6), int64(1), object(10)
memory usage: 313.0+ KB

```

4.2 3. Changing Datatype of timestamp using astype("datetime64[ns])

4.2.1 This is required for time series analysis. So using as.type('datetime64[ns]') I will convert datatype to date time format

4.2.2 Code

```
In [55]: twitter_arch_clean["timestamp"] = twitter_arch_clean["timestamp"].astype('datetime64[ns])
```

```
In [56]: twitter_arch_clean.sample()
```

```

Out[56]:
      tweet_id  in_reply_to_status_id  in_reply_to_user_id \
22  887517139158093824                NaN                NaN

      timestamp                                     source \
22  2017-07-19 03:39:09  <a href="http://twitter.com/download/iphone" r...

      text  retweeted_status_id \
22  I've yet to rate a Venezuelan Hover Wiener. Th...          NaN

      retweeted_status_user_id  retweeted_status_timestamp \
22                NaN                NaN

      expanded_urls  rating_numerator \
22  https://twitter.com/dog_rates/status/887517139...          14.0

      rating_denominator  name  doggo  floofer  pupper  puppo
22                10.0  such  None    None    None    None

```

4.2.3 Test

```
In [57]: twitter_arch_clean["timestamp"].head()
```

```

Out[57]: 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
Name: timestamp, dtype: datetime64[ns]
```

4.3 4. Changing Datatype of tweet_id

In this step using `.astype('str')`, I am going to convert datatype of `tweet_id` from int to string

4.3.1 Code

```
In [58]: twitter_arch_clean["tweet_id"] = twitter_arch_clean["tweet_id"].astype('str')
```

4.3.2 Test

```
In [59]: twitter_arch_clean.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                2356 non-null object
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 object
expanded_urls            2297 non-null object
rating_numerator         2356 non-null float64
rating_denominator       2356 non-null float64
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: datetime64[ns](1), float64(6), object(10)
memory usage: 313.0+ KB
```

5 5. Deleting RetweeD rows using drop function in pandas

As mentioned earlier `in_reply_to_status_id`, `in_reply_to_user_id`, `retweeted_status_id`, `retweeted_status_user_id`, `retweeted_status_timestamp` have null values, and these columns just have retweet and reply information which I do not need in Analysis so I will delete rows corresponding to this columns. I will drop rows corresponding to this column using idea to drop rows where values are not null

5.0.1 Code

```
In [60]: twitter_arch_clean.drop(twitter_arch_clean[twitter_arch_clean.retweeted_status_id.notnull()])
```

```
In [61]: twitter_arch_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                2175 non-null object
in_reply_to_status_id    78 non-null float64
in_reply_to_user_id      78 non-null float64
timestamp               2175 non-null datetime64[ns]
source                  2175 non-null object
text                    2175 non-null object
retweeted_status_id      0 non-null float64
retweeted_status_user_id 0 non-null float64
retweeted_status_timestamp 0 non-null object
expanded_urls            2117 non-null object
rating_numerator         2175 non-null float64
rating_denominator       2175 non-null float64
name                     2175 non-null object
doggo                    2175 non-null object
floofer                  2175 non-null object
pupper                  2175 non-null object
puppo                    2175 non-null object
dtypes: datetime64[ns](1), float64(6), object(10)
memory usage: 305.9+ KB
```

5.0.2 Iterating Assessing to see if I can delete rows based on in_reply_to_status_id

Below assessing shows that we can delete rows where in_reply_to_status_id is not null as on digging into it I find out that they are just reply to tweet and does not have any information regarding dog rating. Only one dog "Tessa" is present in not null columns. So I am deleting rows where in_reply_to_status_id is not null

```
In [62]: twitter_arch_clean[twitter_arch_clean.in_reply_to_status_id.isnull()]
```

```
Out[62]:
```

	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
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
31	886258384151887873	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	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...
31	2017-07-15 16:17:19	<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
31	This is Waffles. His doggles are pupside down...	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
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
31	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...	13.0
1	https://twitter.com/dog_rates/status/892177421...	13.0
2	https://twitter.com/dog_rates/status/891815181...	12.0
3	https://twitter.com/dog_rates/status/891689557...	13.0
4	https://twitter.com/dog_rates/status/891327558...	12.0
5	https://twitter.com/dog_rates/status/891087950...	13.0
6	https://gofundme.com/ydvmve-surgery-for-jax,ht...	13.0
7	https://twitter.com/dog_rates/status/890729181...	13.0
8	https://twitter.com/dog_rates/status/890609185...	13.0
9	https://twitter.com/dog_rates/status/890240255...	14.0
10	https://twitter.com/dog_rates/status/890006608...	13.0
11	https://twitter.com/dog_rates/status/889880896...	13.0
12	https://twitter.com/dog_rates/status/889665388...	13.0
13	https://twitter.com/dog_rates/status/889638837...	12.0
14	https://twitter.com/dog_rates/status/889531135...	13.0
15	https://twitter.com/dog_rates/status/889278841...	13.0
16	https://twitter.com/dog_rates/status/888917238...	12.0
17	https://twitter.com/dog_rates/status/888804989...	13.0
18	https://twitter.com/dog_rates/status/888554962...	13.0
20	https://twitter.com/dog_rates/status/888078434...	12.0
21	https://twitter.com/dog_rates/status/887705289...	13.0
22	https://twitter.com/dog_rates/status/887517139...	14.0
23	https://twitter.com/dog_rates/status/887473957...	13.0
24	https://twitter.com/dog_rates/status/887343217...	13.0
25	https://twitter.com/dog_rates/status/887101392...	12.0
26	https://twitter.com/dog_rates/status/886983233...	13.0
27	https://www.gofundme.com/mingusneedsus,https:/...	13.0
28	https://twitter.com/dog_rates/status/886680336...	13.0
29	https://twitter.com/dog_rates/status/886366144...	12.0
31	https://twitter.com/dog_rates/status/886258384...	13.0
...
2326	https://twitter.com/dog_rates/status/666411507...	2.0
2327	https://twitter.com/dog_rates/status/666407126...	7.0
2328	https://twitter.com/dog_rates/status/666396247...	9.0
2329	https://twitter.com/dog_rates/status/666373753...	11.0
2330	https://twitter.com/dog_rates/status/666362758...	6.0
2331	https://twitter.com/dog_rates/status/666353288...	8.0
2332	https://twitter.com/dog_rates/status/666345417...	10.0
2333	https://twitter.com/dog_rates/status/666337882...	9.0
2334	https://twitter.com/dog_rates/status/666293911...	3.0
2335	https://twitter.com/dog_rates/status/666287406...	1.0
2336	https://twitter.com/dog_rates/status/666273097...	11.0
2337	https://twitter.com/dog_rates/status/666268910...	10.0
2338	https://twitter.com/dog_rates/status/666104133...	1.0
2339	https://twitter.com/dog_rates/status/666102155...	11.0

2340	https://twitter.com/dog_rates/status/666099513...	8.0
2341	https://twitter.com/dog_rates/status/666094000...	9.0
2342	https://twitter.com/dog_rates/status/666082916...	6.0
2343	https://twitter.com/dog_rates/status/666073100...	10.0
2344	https://twitter.com/dog_rates/status/666071193...	9.0
2345	https://twitter.com/dog_rates/status/666063827...	10.0
2346	https://twitter.com/dog_rates/status/666058600...	8.0
2347	https://twitter.com/dog_rates/status/666057090...	9.0
2348	https://twitter.com/dog_rates/status/666055525...	10.0
2349	https://twitter.com/dog_rates/status/666051853...	2.0
2350	https://twitter.com/dog_rates/status/666050758...	10.0
2351	https://twitter.com/dog_rates/status/666049248...	5.0
2352	https://twitter.com/dog_rates/status/666044226...	6.0
2353	https://twitter.com/dog_rates/status/666033412...	9.0
2354	https://twitter.com/dog_rates/status/666029285...	7.0
2355	https://twitter.com/dog_rates/status/666020888...	8.0

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10.0	Phineas	None	None	None	None
1	10.0	Tilly	None	None	None	None
2	10.0	Archie	None	None	None	None
3	10.0	Darla	None	None	None	None
4	10.0	Franklin	None	None	None	None
5	10.0	None	None	None	None	None
6	10.0	Jax	None	None	None	None
7	10.0	None	None	None	None	None
8	10.0	Zoey	None	None	None	None
9	10.0	Cassie	doggo	None	None	None
10	10.0	Koda	None	None	None	None
11	10.0	Bruno	None	None	None	None
12	10.0	None	None	None	None	puppo
13	10.0	Ted	None	None	None	None
14	10.0	Stuart	None	None	None	puppo
15	10.0	Oliver	None	None	None	None
16	10.0	Jim	None	None	None	None
17	10.0	Zeke	None	None	None	None
18	10.0	Ralphus	None	None	None	None
20	10.0	Gerald	None	None	None	None
21	10.0	Jeffrey	None	None	None	None
22	10.0	such	None	None	None	None
23	10.0	Canela	None	None	None	None
24	10.0	None	None	None	None	None
25	10.0	None	None	None	None	None
26	10.0	Maya	None	None	None	None
27	10.0	Mingus	None	None	None	None
28	10.0	Derek	None	None	None	None
29	10.0	Roscoe	None	None	pupper	None
31	10.0	Waffles	None	None	None	None

...
2326	10.0	quite	None	None	None	None
2327	10.0	a	None	None	None	None
2328	10.0	None	None	None	None	None
2329	10.0	None	None	None	None	None
2330	10.0	None	None	None	None	None
2331	10.0	None	None	None	None	None
2332	10.0	None	None	None	None	None
2333	10.0	an	None	None	None	None
2334	10.0	a	None	None	None	None
2335	2.0	an	None	None	None	None
2336	10.0	None	None	None	None	None
2337	10.0	None	None	None	None	None
2338	10.0	None	None	None	None	None
2339	10.0	None	None	None	None	None
2340	10.0	None	None	None	None	None
2341	10.0	None	None	None	None	None
2342	10.0	None	None	None	None	None
2343	10.0	None	None	None	None	None
2344	10.0	None	None	None	None	None
2345	10.0	the	None	None	None	None
2346	10.0	the	None	None	None	None
2347	10.0	a	None	None	None	None
2348	10.0	a	None	None	None	None
2349	10.0	an	None	None	None	None
2350	10.0	a	None	None	None	None
2351	10.0	None	None	None	None	None
2352	10.0	a	None	None	None	None
2353	10.0	a	None	None	None	None
2354	10.0	a	None	None	None	None
2355	10.0	None	None	None	None	None

[2097 rows x 17 columns]

In [63]: twitter_arch_clean[twitter_arch_clean.in_reply_to_status_id.notnull()]

Out[63]:

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

1842 2015-12-13 02:51:51 <a href="http://twitter.com/download/iphone" r...
1844 2015-12-13 01:25:37 <a href="http://twitter.com/download/iphone" r...
1852 2015-12-12 16:02:36 <a href="http://twitter.com/download/iphone" r...
1866 2015-12-11 16:20:15 <a href="http://twitter.com/download/iphone" r...
1882 2015-12-10 17:11:09 <a href="http://twitter.com/download/iphone" r...
1885 2015-12-10 03:30:58 <a href="http://twitter.com/download/iphone" r...
1892 2015-12-10 00:54:28 <a href="http://twitter.com/download/iphone" r...
1895 2015-12-10 00:08:50 <a href="http://twitter.com/download/iphone" r...
1905 2015-12-09 15:09:55 <a href="http://twitter.com/download/iphone" r...
1914 2015-12-08 20:53:11 <a href="http://twitter.com/download/iphone" r...
1940 2015-12-07 04:11:02 <a href="http://twitter.com/download/iphone" r...
2036 2015-12-01 16:37:44 <a href="http://twitter.com/download/iphone" r...
2038 2015-12-01 04:44:10 <a href="http://twitter.com/download/iphone" r...
2149 2015-11-26 01:11:28 <a href="http://twitter.com/download/iphone" r...
2169 2015-11-25 03:14:30 <a href="http://twitter.com/download/iphone" r...
2189 2015-11-24 01:42:25 <a href="http://twitter.com/download/iphone" r...
2298 2015-11-18 20:02:51 <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	@jonnysun @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 @LightningHolt	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 puppies 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\nABC7	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@wanel15	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	12.0
55	NaN	17.0
64	NaN	14.0
113	NaN	10.0
148	NaN	12.0
149	https://twitter.com/dog_rates/status/863079547...	14.0
179	NaN	12.0
184	https://twitter.com/dog_rates/status/856526610...	14.0
186	NaN	14.0
188	NaN	420.0
189	NaN	666.0
218	NaN	13.0
228	NaN	11.0
234	NaN	13.0
251	https://twitter.com/dog_rates/status/844979544...	13.0
274	NaN	10.0
290	NaN	182.0
291	NaN	15.0
313	NaN	960.0
342	NaN	11.0
346	NaN	12.0
387	NaN	7.0
409	NaN	13.0
427	NaN	13.0
498	NaN	12.0
513	NaN	11.0
565	https://twitter.com/dog_rates/status/802265048...	11.0
570	NaN	11.0
576	NaN	11.0
611	NaN	11.0
...
1479	NaN	11.0
1497	NaN	9.0
1501	https://twitter.com/dog_rates/status/692142790...	13.0
1523	NaN	12.0
1598	NaN	4.0
1605	NaN	14.0
1618	NaN	5.0
1630	https://twitter.com/dog_rates/status/684538444...	12.0
1634	https://twitter.com/dog_rates/status/684225744...	143.0
1663	NaN	20.0
1689	NaN	5.0
1774	NaN	13.0
1819	NaN	7.0
1842	https://twitter.com/dog_rates/status/675870721...	11.0
1844	NaN	9.0
1852	https://twitter.com/dog_rates/status/675707330...	11.0

1866	https://twitter.com/dog_rates/status/675349384...	13.0
1882	https://twitter.com/dog_rates/status/674999807...	13.0
1885	https://twitter.com/dog_rates/status/674793399...	13.0
1892	https://twitter.com/dog_rates/status/674754018...	12.0
1895	NaN	11.0
1905	NaN	13.0
1914	NaN	13.0
1940	NaN	1.0
2036	https://twitter.com/dog_rates/status/671729906...	13.0
2038	NaN	1.0
2149	NaN	11.0
2169	https://twitter.com/dog_rates/status/669353438...	10.0
2189	NaN	12.0
2298	NaN	10.0

	rating_denominator	name	doggo	floofer	pupper	puppo
30	10.0	None	None	None	None	None
55	10.0	None	None	None	None	None
64	10.0	None	None	None	None	None
113	10.0	None	None	None	None	None
148	10.0	None	None	None	None	None
149	10.0	None	None	None	None	None
179	10.0	None	None	None	None	None
184	10.0	None	None	None	None	None
186	10.0	None	None	None	None	None
188	10.0	None	None	None	None	None
189	10.0	None	None	None	None	None
218	10.0	None	None	None	None	None
228	10.0	None	None	None	None	None
234	10.0	None	None	None	None	None
251	10.0	None	None	None	None	None
274	10.0	None	None	None	None	None
290	10.0	None	None	None	None	None
291	10.0	None	None	None	None	None
313	0.0	None	None	None	None	None
342	15.0	None	None	None	None	None
346	10.0	None	None	None	None	None
387	10.0	None	None	None	None	None
409	10.0	None	None	None	None	None
427	10.0	None	None	None	None	None
498	10.0	None	None	None	None	None
513	10.0	None	None	None	None	None
565	10.0	None	doggo	None	pupper	None
570	10.0	None	None	None	None	None
576	10.0	None	doggo	None	None	None
611	10.0	None	None	None	None	None
...
1479	10.0	None	None	None	None	None

1497	10.0	None	None	None	None	None
1501	10.0	None	None	None	None	None
1523	10.0	None	None	None	None	None
1598	20.0	None	None	None	None	None
1605	10.0	None	None	None	None	None
1618	10.0	None	None	None	None	None
1630	10.0	None	None	None	None	None
1634	130.0	None	None	None	None	None
1663	16.0	None	None	None	None	None
1689	10.0	None	None	None	None	None
1774	10.0	None	None	None	None	None
1819	10.0	None	None	None	None	None
1842	10.0	None	None	None	None	None
1844	10.0	None	None	None	None	None
1852	10.0	None	None	None	None	None
1866	10.0	None	None	None	None	None
1882	10.0	None	None	None	None	None
1885	10.0	None	None	None	None	None
1892	10.0	None	None	None	None	None
1895	10.0	None	None	None	None	None
1905	10.0	None	None	None	None	None
1914	10.0	None	None	None	None	None
1940	10.0	None	None	None	None	None
2036	10.0	None	None	None	None	None
2038	10.0	None	None	None	None	None
2149	10.0	None	None	None	None	None
2169	10.0	Tessa	None	None	None	None
2189	10.0	None	None	None	None	None
2298	10.0	None	None	None	None	None

[78 rows x 17 columns]

In [64]: twitter_arch_clean.head()

```
Out[64]:
```

	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	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	expanded_urls	rating_numerator \
0	https://twitter.com/dog_rates/status/892420643...	13.0
1	https://twitter.com/dog_rates/status/892177421...	13.0
2	https://twitter.com/dog_rates/status/891815181...	12.0
3	https://twitter.com/dog_rates/status/891689557...	13.0
4	https://twitter.com/dog_rates/status/891327558...	12.0

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10.0	Phineas	None	None	None	None
1	10.0	Tilly	None	None	None	None
2	10.0	Archie	None	None	None	None
3	10.0	Darla	None	None	None	None
4	10.0	Franklin	None	None	None	None

6 6. Deleting replied rows using drop function in pandas

These rows does not have real tweet info, so we can delete them. As done above here also I will drop rows corresponding to this column using idea to drop rows where values are not null

6.0.1 Code

```
In [65]: twitter_arch_clean.drop(twitter_arch_clean[twitter_arch_clean.in_reply_to_status_id.not
```

6.0.2 Test

```
In [66]: twitter_arch_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2097 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                2097 non-null object
in_reply_to_status_id    0 non-null float64
in_reply_to_user_id      0 non-null float64
timestamp               2097 non-null datetime64[ns]
```



```

source                2097 non-null object
text                  2097 non-null object
retweeted_status_id    0 non-null float64
retweeted_status_user_id 0 non-null float64
retweeted_status_timestamp 0 non-null object
expanded_urls          2094 non-null object
rating_numerator        2097 non-null float64
rating_denominator      2097 non-null float64
name                   2097 non-null object
doggo                  2097 non-null object
floofer                2097 non-null object
pupper                 2097 non-null object
puppo                  2097 non-null object
dtypes: datetime64[ns](1), float64(6), object(10)
memory usage: 294.9+ KB

```

7 7. I am going to delete Columns - in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp

As these Columns are null, so it make more sense to delet them from Dataframe. Using .drop function in Pnadas I will delete them

```
In [67]: twitter_arch_clean=twitter_arch_clean.drop(['in_reply_to_status_id', 'in_reply_to_user_id', 'retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timestamp'])
```

```
In [68]: twitter_arch_clean.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2097 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id                2097 non-null object
timestamp                2097 non-null datetime64[ns]
source                  2097 non-null object
text                    2097 non-null object
expanded_urls           2094 non-null object
rating_numerator         2097 non-null float64
rating_denominator       2097 non-null float64
name                    2097 non-null object
doggo                   2097 non-null object
floofer                 2097 non-null object
pupper                  2097 non-null object
puppo                   2097 non-null object
dtypes: datetime64[ns](1), float64(2), object(9)
memory usage: 213.0+ KB

```

8 8. Cleaning Dog name column and converting to Title case

dog name have many non-dog name such as such', 'a', 'quite', 'quite', 'not', 'one', 'incredibly', etc. most of them are in lower case. So I am going to use `.str.islower()` function to grab all these lower case name and replace them with null using `np.nan`. As most of the non-dog name are in lower case so this method will work very well. Secondly, the names also have None in them, which I will replace with `np.nan` too. Thirdly there is a dog name O, in text I found visually that it must be "O'Malley", so I will do replace it here. Finally I will change all name to Title case

```
In [69]: mask = twitter_arch_clean.name.str.islower()
         column_name = 'name'
         twitter_arch_clean.loc[mask, column_name] = np.nan
```

```
In [70]: # Replace all 'None's with a NaN
         twitter_arch_clean['name'].replace('None',
                                           np.nan,
                                           inplace = True)
```

```
In [71]: # Replace the name 'O' with "O'Malley"
         twitter_arch_clean['name'].replace('O',
                                           "O'Malley",
                                           inplace = True)
```

```
In [72]: # Convert names to title case
         twitter_arch_clean.name = twitter_arch_clean.name.str.title()
```

```
In [73]: twitter_arch_clean.name.unique()
```

```
Out[73]: array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', nan, 'Jax',
                'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver', 'Jim',
                'Zeke', 'Ralphus', 'Gerald', 'Jeffrey', 'Canela', 'Maya', 'Mingus',
                'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Earl', 'Lola',
                'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald', 'Rusty', 'Gus',
                'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'Elliot', 'Louis',
                'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack', 'Steven', 'Beau',
                'Snoopy', 'Shadow', 'Emmy', 'Aja', 'Penny', 'Dante', 'Nelly',
                'Ginger', 'Benedict', 'Venti', 'Goose', 'Nugget', 'Cash', 'Jed',
                'Sebastian', 'Sierra', 'Monkey', 'Harry', 'Kody', 'Lassie', 'Rover',
                'Napolean', 'Boomer', 'Cody', 'Rumble', 'Clifford', 'Dewey',
                'Scout', 'Gizmo', 'Walter', 'Cooper', 'Harold', 'Shikha', 'Lili',
                'Jamesy', 'Coco', 'Sammy', 'Meatball', 'Paisley', 'Albus',
                'Neptune', 'Belle', 'Quinn', 'Zooey', 'Dave', 'Jersey', 'Hobbes',
                'Burt', 'Lorenzo', 'Carl', 'Jordy', 'Milky', 'Trooper', 'Sophie',
                'Wyatt', 'Rosie', 'Thor', 'Oscar', 'Callie', 'Cermet', 'Marlee',
                'Arya', 'Einstein', 'Alice', 'Rumpole', 'Benny', 'Aspen', 'Jarod',
                'Wiggles', 'General', 'Sailor', 'Iggy', 'Snoop', 'Kyle', 'Leo',
                'Riley', 'Noosh', 'Odin', 'Jerry', 'Georgie', 'Rontu', 'Cannon',
                'Furzey', 'Daisy', 'Tuck', 'Barney', 'Vixen', 'Jarvis', 'Mimosa',
                'Pickles', 'Brady', 'Luna', 'Charlie', 'Margo', 'Sadie', 'Hank',
```

'Tycho', 'Indie', 'Winnie', 'George', 'Bentley', 'Max', 'Dawn',
 'Maddie', 'Monty', 'Sojourner', 'Winston', 'Odie', 'Arlo',
 'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera', 'Ava', 'Eli', 'Ash',
 'Tucker', 'Tobi', 'Chester', 'Wilson', 'Sunshine', 'Lipton',
 'Bronte', 'Poppy', 'Gidget', 'Rhino', 'Willow', 'Orion', 'Eevee',
 'Smiley', 'Miguel', 'Emanuel', 'Kuyu', 'Dutch', 'Pete', 'Scooter',
 'Reggie', 'Lilly', 'Samson', 'Mia', 'Astrid', 'Malcolm', 'Dexter',
 'Alfie', 'Fiona', 'Mutt', 'Bear', 'Doobert', 'Beebop', 'Alexander',
 'Sailer', 'Brutus', 'Kona', 'Boots', 'Ralphie', 'Loki', 'Cupid',
 'Pawnd', 'Pilot', 'Ike', 'Mo', 'Toby', 'Sweet', 'Pablo', 'Nala',
 'Crawford', 'Gabe', 'Jimison', 'Duchess', 'Harlso', 'Sundance',
 'Luca', 'Flash', 'Sunny', 'Howie', 'Jazzy', 'Anna', 'Finn', 'Bo',
 'Wafer', 'Tom', 'Florence', 'Autumn', 'Buddy', 'Dido', 'Eugene',
 'Ken', 'Strudel', 'Tebow', 'Chloe', 'Timber', 'Binky', 'Moose',
 'Dudley', 'Comet', 'Akumi', 'Titan', 'Olivia', 'Alf', 'Oshie',
 'Chubbs', 'Sky', 'Atlas', 'Eleanor', 'Layla', 'Rocky', 'Baron',
 'Tyr', 'Bauer', 'Swagger', 'Brandi', 'Mary', 'Moe', 'Halo', 'Augie',
 'Craig', 'Sam', 'Hunter', 'Pavlov', 'Phil', 'Kyro', 'Wallace',
 'Ito', 'Seamus', 'Ollie', 'Stephan', 'Lennon', 'Major', 'Duke',
 'Sansa', 'Shooter', 'Django', 'Diogi', 'Sonny', 'Marley', 'Severus',
 'Ronnie', 'Milo', 'Bones', 'Mauve', 'Chef', 'Doc', 'Peaches',
 'Sobe', 'Longfellow', 'Mister', 'Iroh', 'Pancake', 'Snicku', 'Ruby',
 'Brody', 'Mack', 'Nimbus', 'Laika', 'Maximus', 'Dobby', 'Moreton',
 'Juno', 'Maude', 'Lily', 'Newt', 'Benji', 'Nida', 'Robin',
 'Monster', 'Bebe', 'Remus', 'Levi', 'Mabel', 'Misty', 'Betty',
 'Mosby', 'Maggie', 'Bruce', 'Happy', 'Ralphy', 'Brownie', 'Rizzy',
 'Stella', 'Butter', 'Frank', 'Tonks', 'Lincoln', 'Rory', 'Logan',
 'Dale', 'Rizzo', 'Arnie', 'Mattie', 'Pinot', 'Dallas', 'Hero',
 'Frankie', 'Stormy', 'Reginald', 'Balto', 'Mairi', 'Loomis', 'Godi',
 'Cali', 'Deacon', 'Timmy', 'Sampson', 'Chipson', 'Combo', 'Oakley',
 'Dash', 'Hercules', 'Jay', 'Mya', 'Strider', 'Wesley', 'Solomon',
 'Huck', 'O'Malley', 'Blue', 'Anakin', 'Finley', 'Sprinkles',
 'Heinrich', 'Shakespeare', 'Chelsea', 'Bungalo', 'Chip', 'Grey',
 'Roosevelt', 'Willem', 'Davey', 'Dakota', 'Fizz', 'Dixie', 'Al',
 'Jackson', 'Carbon', 'Klein', 'Dondon', 'Kirby', 'Lou', 'Chevy',
 'Tito', 'Philbert', 'Louie', 'Rupert', 'Rufus', 'Brudge', 'Shadoe',
 'Angel', 'Brat', 'Tove', 'Gromit', 'Aubie', 'Kota', 'Leela',
 'Glenn', 'Shelby', 'Sephie', 'Bonaparte', 'Albert', 'Wishes',
 'Rose', 'Theo', 'Rocco', 'Fido', 'Emma', 'Spencer', 'Lilli',
 'Boston', 'Brandonald', 'Corey', 'Leonard', 'Beckham', 'Devón',
 'Gert', 'Watson', 'Keith', 'Dex', 'Ace', 'Tayzie', 'Grizzie',
 'Fred', 'Gilbert', 'Meyer', 'Zoe', 'Stewie', 'Calvin', 'Lilah',
 'Spanky', 'Jameson', 'Piper', 'Atticus', 'Blu', 'Dietrich',
 'Divine', 'Tripp', 'Cora', 'Huxley', 'Keurig', 'Bookstore', 'Linus',
 'Abby', 'Shiloh', 'Gustav', 'Arlen', 'Percy', 'Lenox', 'Sugar',
 'Harvey', 'Blanket', 'Geno', 'Stark', 'Beya', 'Kilo', 'Kayla',
 'Maxaroni', 'Bell', 'Doug', 'Edmund', 'Aqua', 'Theodore', 'Baloo',
 'Chase', 'Nollie', 'Rorie', 'Simba', 'Charles', 'Bayley', 'Axel',

'Storkson', 'Remy', 'Chadrick', 'Kellogg', 'Buckley', 'Livvie',
 'Terry', 'Hermione', 'Ralpher', 'Aldrick', 'Larry', 'Rooney',
 'Crystal', 'Ziva', 'Stefan', 'Pupcasso', 'Puff', 'Flurpson',
 'Coleman', 'Enchilada', 'Raymond', 'Rueben', 'Cilantro', 'Karll',
 'Sprout', 'Blitz', 'Bloop', 'Colby', 'Lillie', 'Ashleigh',
 'Kreggory', 'Sarge', 'Luther', 'Ivar', 'Jangle', 'Schnitzel',
 'Panda', 'Berkeley', 'Ralphé', 'Charleson', 'Clyde', 'Harnold',
 'Sid', 'Pippa', 'Otis', 'Carper', 'Bowie', 'Alexanderson', 'Suki',
 'Barclay', 'Skittle', 'Ebby', 'Flávio', 'Smokey', 'Link',
 'Jennifur', 'Ozzy', 'Bluebert', 'Stephanus', 'Bubbles', 'Zeus',
 'Bertson', 'Nico', 'Michelangelo', 'Siba', 'Calbert', 'Curtis',
 'Travis', 'Thumas', 'Kanu', 'Lance', 'Opie', 'Stubert', 'Kane',
 'Olive', 'Chuckles', 'Stanuel', 'Sora', 'Beemo', 'Gunner', 'Lacy',
 'Tater', 'Olaf', 'Cecil', 'Vince', 'Karma', 'Billy', 'Walker',
 'Rodney', 'Klevin', 'Malikai', 'Bobbie', 'River', 'Jebberson',
 'Remington', 'Farfle', 'Jiminus', 'Harper', 'Clarkus', 'Finnegus',
 'Cupcake', 'Kathmandu', 'Ellie', 'Katie', 'Kara', 'Adele', 'Zara',
 'Ambrose', 'Jimothy', 'Bode', 'Terrenth', 'Reese', 'Chesterson',
 'Lucia', 'Bisquick', 'Ralphson', 'Socks', 'Rambo', 'Rudy', 'Fiji',
 'Rilo', 'Bilbo', 'Coopson', 'Yoda', 'Millie', 'Chet', 'Crouton',
 'Daniel', 'Kaia', 'Murphy', 'Dotsy', 'Eazy', 'Coops', 'Fillup',
 'Miley', 'Charl', 'Reagan', 'Yukon', 'Cece', 'Cuddles', 'Claude',
 'Jessiga', 'Carter', 'Ole', 'Pherb', 'Blipson', 'Reptar', 'Trevith',
 'Berb', 'Bob', 'Colin', 'Brian', 'Olivier', 'Grady', 'Kobe',
 'Freddery', 'Bodie', 'Dunkin', 'Wally', 'Tupawc', 'Amber',
 'Herschel', 'Edgar', 'Teddy', 'Kingsley', 'Brockly', 'Richie',
 'Molly', 'Vinscent', 'Cedrick', 'Hazel', 'Lolo', 'Eriq', 'Phred',
 'Oddie', 'Maxwell', 'Geoff', 'Covach', 'Durg', 'Fynn', 'Ricky',
 'Herald', 'Lucky', 'Ferg', 'Trip', 'Clarence', 'Hamrick', 'Brad',
 'Pubert', 'Frönq', 'Derby', 'Lizzie', 'Ember', 'Blakely', 'Opal',
 'Marq', 'Kramer', 'Barry', 'Tyrone', 'Gordon', 'Baxter', 'Mona',
 'Horace', 'Crimson', 'Birf', 'Hammond', 'Lorelei', 'Marty',
 'Brooks', 'Petrick', 'Hubertson', 'Gerbald', 'Oreo', 'Bruiser',
 'Perry', 'Bobby', 'Jeph', 'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe',
 'Tiger', 'Jiminy', 'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Darrel',
 'Taco', 'Joey', 'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses',
 'Toffee', 'Apollo', 'Carly', 'Asher', 'Glacier', 'Chuck', 'Champ',
 'Ozzie', 'Griswold', 'Cheesy', 'Moofasa', 'Hector', 'Goliath',
 'Kawhi', 'Emmie', 'Penelope', 'Willie', 'Rinna', 'Mike', 'William',
 'Dwight', 'Evy', 'Hurley', 'Rubio', 'Chompsky', 'Rascal', 'Linda',
 'Tug', 'Tango', 'Grizz', 'Jerome', 'Crumpet', 'Jessifer', 'Izzy',
 'Ralph', 'Sandy', 'Humphrey', 'Tassy', 'Juckson', 'Chuq', 'Tyrus',
 'Karl', 'Godzilla', 'Vinnie', 'Kenneth', 'Herm', 'Bert', 'Striker',
 'Donny', 'Pepper', 'Bernie', 'Buddah', 'Lenny', 'Arnold', 'Zuzu',
 'Mollie', 'Laela', 'Tedders', 'Superpup', 'Rufio', 'Jeb', 'Rodman',
 'Jonah', 'Chesney', 'Kenny', 'Henry', 'Bobbay', 'Mitch', 'Kaiya',
 'Acro', 'Aiden', 'Obie', 'Dot', 'Shnuggles', 'Kendall', 'Jeffri',
 'Steve', 'Eve', 'Mac', 'Fletcher', 'Kenzie', 'Pumpkin', 'Schnozz',

```
'Gustaf', 'Cheryl', 'Ed', 'Leonidas', 'Norman', 'Caryl', 'Scott',
'Taz', 'Darby', 'Jackie', 'Jazz', 'Franq', 'Pippin', 'Rolf',
'Snickers', 'Ridley', 'Cal', 'Bradley', 'Bubba', 'Tuco', 'Patch',
'Mojo', 'Batdog', 'Dylan', 'Mark', 'Jd', 'Alejandro', 'Scruffers',
'Pip', 'Julius', 'Tanner', 'Sparky', 'Anthony', 'Holly', 'Jett',
'Amy', 'Sage', 'Andy', 'Mason', 'Trigger', 'Antony', 'Creg',
'Traviss', 'Gin', 'Jeffrie', 'Danny', 'Ester', 'Pluto', 'Bloo',
'Edd', 'Paull', 'Willy', 'Herb', 'Damon', 'Peanut', 'Nigel',
'Butters', 'Sandra', 'Fabio', 'Randall', 'Liam', 'Tommy', 'Ben',
'Raphael', 'Julio', 'Andru', 'Kloey', 'Shawwn', 'Skye', 'Kollin',
'Ronduh', 'Billl', 'Saydee', 'Dug', 'Sully', 'Kirk', 'Ralf',
'Clarq', 'Jaspers', 'Samsom', 'Terrance', 'Harrison', 'Chaz',
'Jeremy', 'Jaycob', 'Lambeau', 'Ruffles', 'Amélie', 'Bobb',
'Banditt', 'Kevon', 'Winifred', 'Hanz', 'Churlie', 'Zeek', 'Timofy',
'Maks', 'Jomathan', 'Kallie', 'Marvin', 'Spark', 'Gòrdón', 'Jo',
'Dayz', 'Jareld', 'Torque', 'Ron', 'Skittles', 'Cleopatricia',
'Erik', 'Stu', 'Tedrick', 'Shaggy', 'Filup', 'Kial', 'Naphaniel',
'Dook', 'Hall', 'Philippe', 'Biden', 'Fwed', 'Genevieve', 'Joshwa',
'Timison', 'Bradlay', 'Pipsy', 'Clybe', 'Keet', 'Carll', 'Jockson',
'Josep', 'Lugan', 'Christopher'], dtype=object)
```

9 9. Extracting Source Type from Source using str.extract

Extracting different sources can help us in analyse what is the most used source by users. Using `.str.extract` a function of pandas I am going to achieve it

9.0.1 Code

```
In [74]: twitter_arch_clean['source'] = twitter_arch_clean['source'].str.extract(r'>(.)<', expand=True)
```

9.0.2 Test

```
In [75]: twitter_arch_clean['source'].unique()
```

```
Out[75]: array(['Twitter for iPhone', 'Twitter Web Client', 'Vine - Make a Scene',
                'TweetDeck'], dtype=object)
```

```
In [76]: twitter_arch_clean['source'].value_counts()
```

```
Out[76]: Twitter for iPhone      1964
         Vine - Make a Scene      91
         Twitter Web Client      31
         TweetDeck                11
         Name: source, dtype: int64
```

10 10. In Image dataframe the some names start with lower case and some have upper case. Visually I have seen that some name have "-", "- " abd " " in them #### so I am going to standardize them by replacing "-" and " " with "" and converting them to Title case

```
In [77]: image_clean = image.copy()
```

```
In [78]: image_clean.head()
         image.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
```

10.0.1 Code

```
In [79]: image_clean[['p1','p2','p3']] = image_clean[['p1','p2','p3']].applymap(str.title)
         image_clean[['p1','p2','p3']] = image_clean[['p1','p2','p3']].replace(' ','_').replace('-', '_')
```

10.0.2 Test

```
In [80]: image_clean[['p1','p2','p3']].sample(10)
```

```
Out[80]:
```

	p1	p2	p3
342	Ostrich	Arabian_Camel	Llama
499	Pomeranian	Chow	Pekinese
1275	Saluki	Borzoi	Labrador_Retriever
1471	Sundial	Cash_Machine	Maze
1891	Fountain	American_Black_Bear	Sundial
1528	Chihuahua	Pekinese	Pembroke
620	Lakeland_Terrier	West_Highland_White_Terrier	Norwich_Terrier
143	Crash_Helmet	Toaster	Mouse
276	Shih-Tzu	Lhasa	Guinea_Pig
1845	Border_Terrier	Brabancon_Griffon	Yorkshire_Terrier

11 11 image_predictions, Remove non-dogs, the 324 rows where p1, p2, and p3 are false

in this step using query function, I will only import the rows where all three prediction predict that the picture is of dog only

11.0.1 Code

```
In [81]: len(image_clean.query('p1_dog == False & p2_dog == False & p3_dog == False'))
```

```
Out[81]: 324
```

```
In [82]: image_clean = image_clean.query('p1_dog == True or p2_dog == True or p3_dog == True')
```

11.0.2 test

```
In [83]: len(image_clean.query('p1_dog == False & p2_dog == False & p3_dog == False'))
```

```
Out[83]: 0
```

11.1 12. Changing name of id to tweet_id and changing Datatype of tweet_id using astype('str') in tweet_json Dataframe and changing Datatype of tweet_id using astype('str') in image_clean Dataframe

As i have to combine all dataframe together, so I will change the datatype of it to string using .astype('str')

11.2 code

```
In [84]: image_clean["tweet_id"] = image_clean["tweet_id"].astype('str')
```

```
In [85]: image_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1751 entries, 0 to 2073
Data columns (total 12 columns):
tweet_id      1751 non-null object
jpg_url       1751 non-null object
img_num       1751 non-null int64
p1            1751 non-null object
p1_conf       1751 non-null float64
p1_dog        1751 non-null bool
p2            1751 non-null object
p2_conf       1751 non-null float64
p2_dog        1751 non-null bool
p3            1751 non-null object
p3_conf       1751 non-null float64
p3_dog        1751 non-null bool
dtypes: bool(3), float64(3), int64(1), object(5)
memory usage: 141.9+ KB
```

```
In [86]: tweet_jason_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 3 columns):
id                2354 non-null int64
favorite_count    2354 non-null int64
retweet_count     2354 non-null int64
dtypes: int64(3)
memory usage: 55.2 KB
```

```
In [88]: tweet_jason_clean.rename({'id': 'tweet_id'}, axis=1, inplace=True)
```

```
/opt/conda/lib/python3.6/site-packages/pandas/core/frame.py:3781: 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#
    return super(DataFrame, self).rename(**kwargs)
```

```
In [89]: tweet_jason_clean["tweet_id"] = tweet_jason_clean["tweet_id"].astype('str')
```

```
/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.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
    """Entry point for launching an IPython kernel.
```

```
In [90]: tweet_jason_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 3 columns):
tweet_id          2354 non-null object
favorite_count    2354 non-null int64
retweet_count     2354 non-null int64
dtypes: int64(2), object(1)
memory usage: 55.2+ KB
```

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2097 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id          2097 non-null object
```



```

timestamp          2097 non-null datetime64[ns]
source             2097 non-null object
text               2097 non-null object
expanded_urls      2094 non-null object
rating_numerator   2097 non-null float64
rating_denominator 2097 non-null float64
name               1390 non-null object
doggo              2097 non-null object
floofer            2097 non-null object
pupper             2097 non-null object
puppo              2097 non-null object
dtypes: datetime64[ns](1), float64(2), object(9)
memory usage: 213.0+ KB

```

12 13 Cleaning Numerator and denominator ratings

1. First, I will extract all rating from the text column using Regex

2. Second I assume the rating benchmark to be 20. So, First, I will extract rating_numerator and text column to see the mismatch in ratings and if mismatch is there I will replace ratings in numerator_rating with rating from text

3. I will also replace very high ratings such as 1776, 666 with median of ratings

12.0.1 4. I will also replace Null values with mean rating

5. I will change rating_denominator to 10. This will standardize it

```
In [92]: twitter_arch_clean = twitter_arch_clean.copy()
```

12.0.2 Code

```
In [93]: twitter_arch_clean['rating_numerator'],twitter_arch_clean['rating_denominator'] = twitt
```

```

twitter_arch_clean['rating_numerator'] = twitter_arch_clean['rating_numerator'].astype(
twitter_arch_clean['rating_denominator'] = twitter_arch_clean['rating_denominator'].ast

```

```
In [94]: from pandas import option_context
```

```

with option_context('display.max_colwidth', 400):
    display(twitter_arch_clean[twitter_arch_clean['rating_numerator']>20][['tweet_id', '

```

```

        tweet_id \
433    820690176645140481
902    758467244762497024
979    749981277374128128

```

1120 731156023742988288
 1202 716439118184652801
 1228 713900603437621249
 1254 710658690886586372
 1274 709198395643068416
 1351 704054845121142784
 1433 697463031882764288
 1635 684222868335505415
 1779 677716515794329600
 1843 675853064436391936
 2074 670842764863651840

433 The floofs have been released I repeat the
 902 Why does this never ha
 979 This is Atticus. He's
 1120 Say hello to this unbelievably well behaved squad of doggos. 204/
 1202 This is Bluebert. He just saw that both #FinalFur match ups a
 1228 Happy Saturday here's 9 puppies on a
 1254 Here's a brigade of puppies. All look very prepared
 1274 From left to right:\nCletus, Jerome, Alejandro, Burp, & Titson\nNone know where camera
 1351 Here is a whole flock of p
 1433 Happy Wednesday here's a bucket of p
 1635 Someone help the girl is being mugged. Several are distracting her while two steal h
 1779 IT'S PUPPERGED
 1843 Here we have an entire platoon of puppies. Total sc
 2074 After so many requests... h

	rating_numerator
433	84.0
902	165.0
979	1776.0
1120	204.0
1202	50.0
1228	99.0
1254	80.0
1274	45.0
1351	60.0
1433	44.0
1635	121.0
1779	144.0
1843	88.0
2074	420.0

313,433,902,1120 1202,1228,1254,1274,1351,1433,1634,1635,1779,1843,

```
In [95]: twitter_arch_clean.loc[313,'rating_numerator'] = 13.00
         twitter_arch_clean.loc[433,'rating_numerator'] = 8.40
```

```

twitter_arch_clean.loc[902, 'rating_numerator'] = 11.00
twitter_arch_clean.loc[1120, 'rating_numerator'] = 12.00
twitter_arch_clean.loc[1202, 'rating_numerator'] = 11.00
twitter_arch_clean.loc[1228, 'rating_numerator'] = 11.00
twitter_arch_clean.loc[1254, 'rating_numerator'] = 10.00
twitter_arch_clean.loc[1274, 'rating_numerator'] = 9.00
twitter_arch_clean.loc[1351, 'rating_numerator'] = 12.00
twitter_arch_clean.loc[1433, 'rating_numerator'] = 11.00
twitter_arch_clean.loc[1634, 'rating_numerator'] = 11.00
twitter_arch_clean.loc[1635, 'rating_numerator'] = 11.00
twitter_arch_clean.loc[1779, 'rating_numerator'] = 12.00
twitter_arch_clean.loc[1843, 'rating_numerator'] = 11.00

```

```
In [96]: twitter_arch_clean['rating_numerator'].median()
```

```
Out[96]: 11.0
```

replacing 188,189,290,342,979,2074 with median

```
In [97]: twitter_arch_clean.loc[188, 'rating_numerator'] =twitter_arch_clean['rating_numerator'].m
twitter_arch_clean.loc[189, 'rating_numerator']=twitter_arch_clean['rating_numerator'].m
twitter_arch_clean.loc[290, 'rating_numerator']=twitter_arch_clean['rating_numerator'].m
twitter_arch_clean.loc[342, 'rating_numerator']=twitter_arch_clean['rating_numerator'].m
twitter_arch_clean.loc[979, 'rating_numerator']=twitter_arch_clean['rating_numerator'].m
twitter_arch_clean.loc[2074, 'rating_numerator']=twitter_arch_clean['rating_numerator'].m

```

12.0.3 Test

```
In [98]: from pandas import option_context
```

```

with option_context('display.max_colwidth', 400):
    display(twitter_arch_clean[twitter_arch_clean['rating_numerator']>20][['tweet_id', '

```

Empty DataFrame

Columns: [tweet_id, text, rating_numerator]

Index: []

```
In [99]: twitter_arch_clean[twitter_arch_clean['rating_numerator'].isnull()]
```

```

Out[99]:
      tweet_id      timestamp      source \
516  810984652412424192  2016-12-19 23:06:23  Twitter for iPhone

      text \
516  Meet Sam. She smiles 24/7 & secretly aspir...

      expanded_urls  rating_numerator \

```

```
516 https://www.gofundme.com/sams-smile,https://tw... NaN
```

```
rating_denominator name doggo floofer pupper puppo
516 NaN Sam None None None None
```

12.0.4 Code

```
In [100]: #replacing null value with mean
twitter_arch_clean.loc[516, 'rating_numerator']=twitter_arch_clean['rating_numerator'].
```

12.0.5 Test

```
In [101]: twitter_arch_clean[twitter_arch_clean['rating_numerator'].isnull()]
```

```
Out[101]: Empty DataFrame
Columns: [tweet_id, timestamp, source, text, expanded_urls, rating_numerator, rating_d
Index: []
```

```
In [102]: len(twitter_arch_clean[twitter_arch_clean['rating_numerator']>20])
```

```
Out[102]: 0
```

```
In [103]: twitter_arch_clean.rating_denominator.unique()
```

```
Out[103]: array([ 10.,  70.,  nan, 150., 170.,  20.,  50.,  90.,  80.,
                  40., 110., 120.])
```

```
In [104]: twitter_arch_clean.rating_denominator =10
twitter_arch_clean.sample(5)
```

```
Out[104]:
```

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator
2131	670086499208155136	2015-11-27 03:47:25	Twitter for iPhone	"Hi yes this is dog. I can't help with that s-...	https://twitter.com/dog_rates/status/670086499...	10.0
788	774314403806253056	2016-09-09 18:31:54	Twitter for iPhone	I WAS SENT THE ACTUAL DOG IN THE PROFILE PIC B...	https://twitter.com/dog_rates/status/774314403...	14.0
2023	671896809300709376	2015-12-02 03:40:57	Twitter for iPhone	This is Lola. She fell asleep on a piece of pi...	https://twitter.com/dog_rates/status/671896809...	10.0
1021	746726898085036033	2016-06-25 15:29:00	Twitter for iPhone	Meet Abby. She's incredibly distracting. Just ...	https://twitter.com/dog_rates/status/746726898...	12.0
894	759197388317847553	2016-07-30 01:22:17	Twitter for iPhone	This is Luna. She's just heckin precious af I ...		

894 https://twitter.com/dog_rates/status/759197388...

12.0

	rating_denominator	name	doggo	floofer	pupper	puppo
2131	10	NaN	None	None	None	None
788	10	NaN	None	None	None	None
2023	10	Lola	None	None	None	None
1021	10	Abby	None	None	None	None
894	10	Luna	None	None	None	None

Define: Combine dog stage columns (doggo, floofer, pupper, puppo) into one 'dog_stage' column. Delete the separate dog stage categories after visually inspecting a random sample to ensure the combination worked accurately. Convert entries in this column title case.

12.1 Tidiness - Messy Data Untidy

In this section I will work on Two tidiness Issues that I identified in Assessing process

1. doggo, floofer, pupper, puppo must make one column - Varioable Stage in 4 Columns
2. Merge three dataframe together to make one dataframe and extract it to .CSV file format

12.1.1 1. doggo, floofer, pupper, puppo must make one column - Varioable Stage in 4 Columns

1. Combine stage columns (doggo, floofer, pupper, puppo) into one 'stage' column. Delete the separate dog stage categories after visually inspecting a random sample to ensure the combination worked accurately

12.1.2 Code

In [105]: *# making one column for all*

```
twitter_arch_clean['stage'] = twitter_arch_clean['doggo']+twitter_arch_clean['floofer']
twitter_arch_clean['stage'] = twitter_arch_clean['stage'].str.replace('None','')
twitter_arch_clean = twitter_arch_clean.replace(({ 'stage': {'':np.nan}}))
twitter_arch_clean.drop(['doggo', 'floofer', 'pupper', 'puppo'],axis=1,inplace=True)
```

12.1.3 Test

In [106]: twitter_arch_clean['stage'].value_counts()

```
Out[106]: pupper      221
doggo      72
puppo      23
floofer      9
doggopupper  9
doggofloofer  1
doggopuppo   1
Name: stage, dtype: int64
```

12.2 2. Merge three dataframe together to make one dataframe and extract it to .CSV file format

12.2.1 Merging all Three DataFrame into 1 DataFrame. In this I am going to merge twitter_arch_clean and image_clean based on Inner join and lastly merge tweet_jason_clean based on left join on tweet-id

12.2.2 Code

```
In [107]: twitter_archive_master = pd.merge(twitter_arch_clean,image_clean,how='inner',on='tweet
```

12.2.3 Test

```
In [108]: twitter_archive_master.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1666 entries, 0 to 1665
Data columns (total 22 columns):
tweet_id          1666 non-null object
timestamp         1666 non-null datetime64[ns]
source            1666 non-null object
text              1666 non-null object
expanded_urls     1666 non-null object
rating_numerator  1666 non-null float64
rating_denominator 1666 non-null int64
name              1186 non-null object
stage             257 non-null object
jpg_url           1666 non-null object
img_num           1666 non-null int64
p1                1666 non-null object
p1_conf           1666 non-null float64
p1_dog            1666 non-null bool
p2                1666 non-null object
p2_conf           1666 non-null float64
p2_dog            1666 non-null bool
p3                1666 non-null object
p3_conf           1666 non-null float64
p3_dog            1666 non-null bool
favorite_count    1666 non-null int64
retweet_count     1666 non-null int64
dtypes: bool(3), datetime64[ns](1), float64(4), int64(4), object(10)
memory usage: 265.2+ KB
```

```
In [109]: twitter_archive_master.sample(5)
```

```
Out[109]:
```

	tweet_id	timestamp	source	\
780	724046343203856385	2016-04-24 01:24:35	Twitter for iPhone	
442	788412144018661376	2016-10-18 16:11:17	Twitter for iPhone	
502	776218204058357768	2016-09-15 00:36:55	Twitter for iPhone	

```

619 752519690950500352 2016-07-11 15:07:30 Twitter for iPhone
1235 678334497360859136 2015-12-19 22:02:01 Twitter for iPhone

```

```

                                text \
780 This is Sarge. Not even he knows what his tong...
442 This is Dexter. He breaks hearts for a living...
502 Atlas rolled around in some chalk and now he's...
619 Hopefully this puppo on a swing will help get ...
1235 This is Rosie. She has a snazzy bow tie and a ...

```

```

                                expanded_urls rating_numerator \
780 https://twitter.com/dog_rates/status/724046343... 10.0
442 https://twitter.com/dog_rates/status/788412144... 11.0
502 https://twitter.com/dog_rates/status/776218204... 13.0
619 https://twitter.com/dog_rates/status/752519690... 11.0
1235 https://twitter.com/dog_rates/status/678334497... 10.0

```

```

rating_denominator name stage \
780 10 Sarge NaN
442 10 Dexter NaN
502 10 NaN floofer
619 10 NaN puppo
1235 10 Rosie NaN

```

```

                                jpg_url ... \
780 https://pbs.twimg.com/media/CgxUTS_XEAAC0pv.jpg ...
442 https://pbs.twimg.com/media/CvEAqQoWgAADj5K.jpg ...
502 https://pbs.twimg.com/media/CsWuVEdWcAAqbe9.jpg ...
619 https://pbs.twimg.com/media/CnF8qVDWYAAhOg1.jpg ...
1235 https://pbs.twimg.com/media/CWntoDVWcAE13NB.jpg ...

```

```

p1_conf p1_dog p2 p2_conf p2_dog p3 \
780 0.826272 True Bull_Mastiff 0.158595 True Great_Dane
442 0.805238 True Labrador_Retriever 0.113798 True Brittany_Spaniel
502 0.940326 True Pomeranian 0.055527 True Keeshond
619 0.999984 False Labrador_Retriever 0.000010 True Eskimo_Dog
1235 0.378643 True Golden_Retriever 0.095594 True Kelpie

```

```

p3_conf p3_dog favorite_count retweet_count
780 0.011859 True 2901 620
442 0.038559 True 16060 5990
502 0.002226 True 33345 18497
619 0.000001 True 8157 3917
1235 0.085309 True 1427 304

```

```
[5 rows x 22 columns]
```

```
In [110]: #Importing CSV file named - twitter_archive_master.csv
```

```
twitter_archive_master.to_csv('twitter_archive_master.csv')
```

12.3

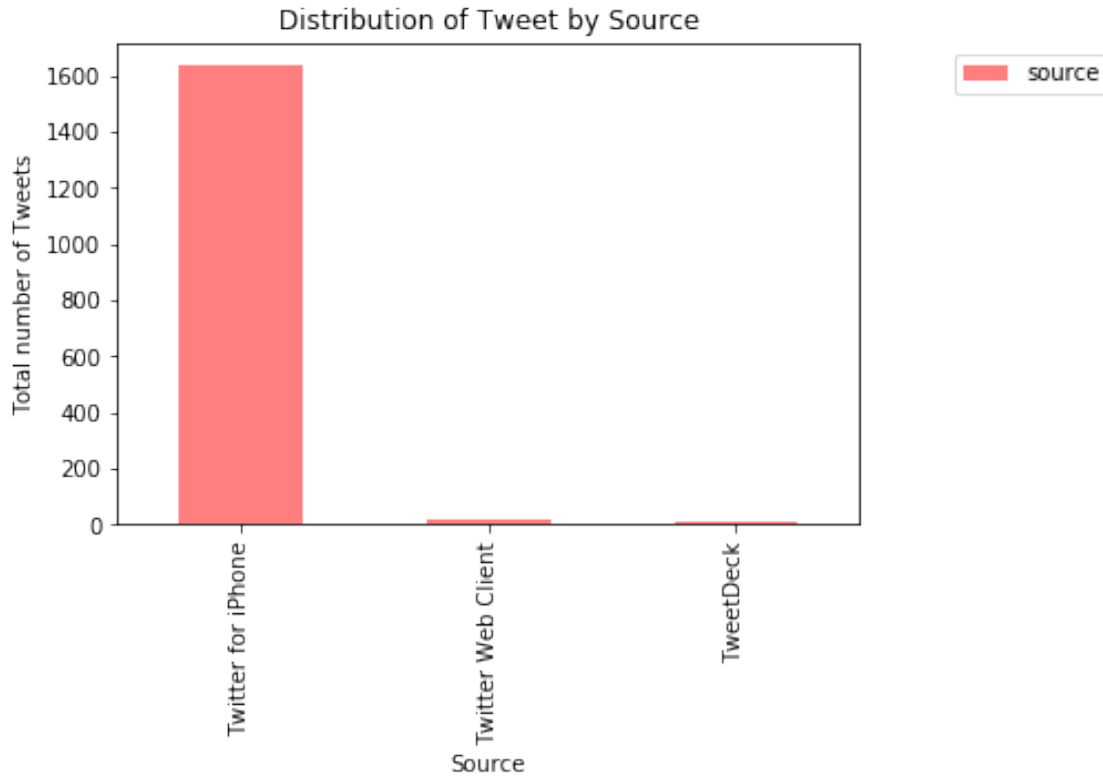
13 Analysis and Visualization

```
In [111]: twitter_archive_master.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1666 entries, 0 to 1665
Data columns (total 22 columns):
tweet_id          1666 non-null object
timestamp         1666 non-null datetime64[ns]
source            1666 non-null object
text              1666 non-null object
expanded_urls     1666 non-null object
rating_numerator  1666 non-null float64
rating_denominator 1666 non-null int64
name              1186 non-null object
stage             257 non-null object
jpg_url           1666 non-null object
img_num           1666 non-null int64
p1                1666 non-null object
p1_conf           1666 non-null float64
p1_dog            1666 non-null bool
p2                1666 non-null object
p2_conf           1666 non-null float64
p2_dog            1666 non-null bool
p3                1666 non-null object
p3_conf           1666 non-null float64
p3_dog            1666 non-null bool
favorite_count    1666 non-null int64
retweet_count     1666 non-null int64
dtypes: bool(3), datetime64[ns](1), float64(4), int64(4), object(10)
memory usage: 265.2+ KB
```

13.1 1. Most used Source Type

```
In [112]: ax=twitter_archive_master['source'].value_counts().plot(kind='bar',color="red",alpha=.
ax.set_title("Distribution of Tweet by Source")
ax.set_xlabel("Source")
ax.set_ylabel("Total number of Tweets")
ax.legend(bbox_to_anchor=(1.35, 1));
```

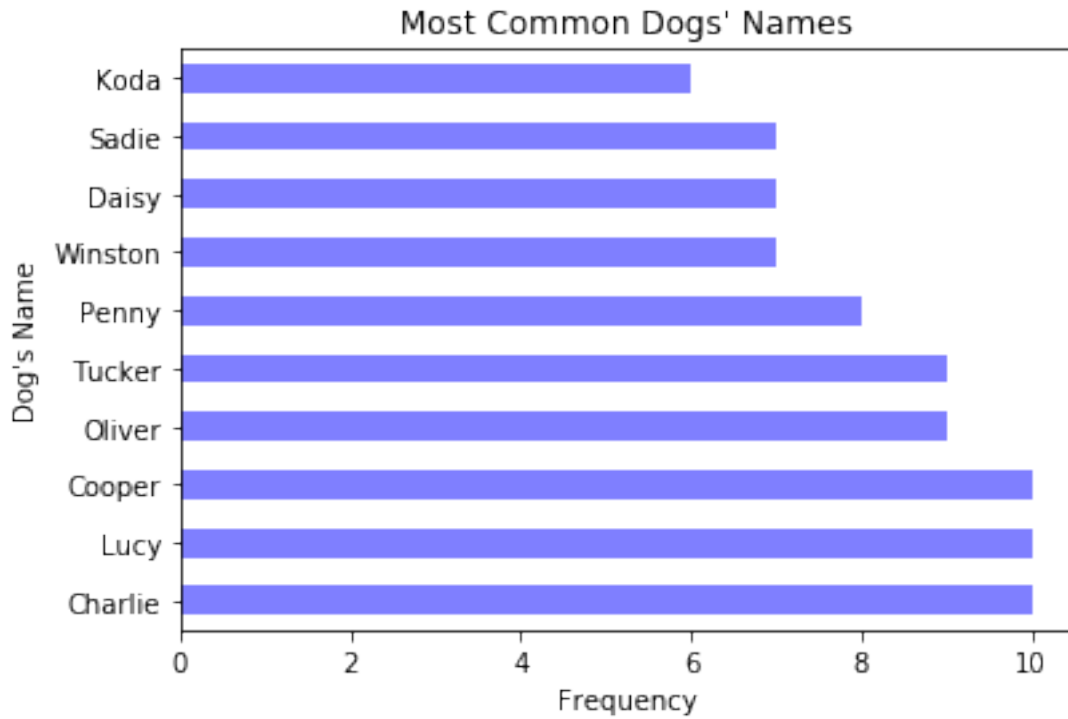
```
In [113]: twitter_archive_master['source'].value_counts()
```

```
Out[113]: Twitter for iPhone    1635
          Twitter Web Client     22
          TweetDeck              9
          Name: source, dtype: int64
```

13.1.1 Twitter is the most used source for Tweeting to rate dogs in this dataset

13.2 2. Most common Name

```
In [114]: # rank the names frequency in a descending order
          ax = twitter_archive_master.name.value_counts().sort_values(ascending =False)[:10].plot()
          ax.set_title("Most Common Dogs' Names")
          ax.set_xlabel('Frequency')
          ax.set_ylabel("Dog's Name");
```



13.2.1 Lucy, Cooper, Charlie are the most used names for dog

13.3 3. Most rate dog with highest Number of favourite count and retweet count

In [115]: `twitter_archive_master['rating_numerator'].value_counts().sort_values(ascending = False)`

```
Out[115]: 12.00    423
          10.00    360
          11.00    359
          13.00    221
          9.00     133
          8.00     68
          7.00     31
          14.00     22
          6.00     16
          5.00     13
          4.00      7
          3.00      5
          2.00      2
          11.26      1
          9.75      1
          11.27      1
          0.00      1
          13.50      1
```

```

8.40      1
Name: rating_numerator, dtype: int64

In [116]: highest_rating = twitter_archive_master[twitter_archive_master['rating_numerator']==14]

In [117]: len(highest_rating)

Out[117]: 22

In [118]: highest_rating[highest_rating.retweet_count==highest_rating.retweet_count.max()]

Out[118]:
```

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator	name	stage	jpg_url	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog	favorite_count	retweet_count
296	819004803107983360	2017-01-11 02:15:36	Twitter for iPhone	This is Bo. He was a very good First Doggo. 14...	https://twitter.com/dog_rates/status/819004803...	14.0	10	Bo	doggo	https://pbs.twimg.com/media/C12whDoVEAALRxa.jpg	0.351308	Toy_Poodle	0.271929	True	Tibetan_Terrier	0.094759	True	95450	42228	

```

[1 rows x 22 columns]

In [119]: highest_rating[highest_rating.favorite_count==highest_rating.favorite_count.max()][["p1", "p2", "p3"]]

Out[119]:
```

	p1	p2	p3
296	Standard_Poodle	Toy_Poodle	Tibetan_Terrier

```

In [120]: from IPython.display import Image
          from IPython.core.display import HTML
          Image(url= "https://pbs.twimg.com/media/C12whDoVEAALRxa.jpg",width=400, height=400)
          #https://stackoverflow.com/questions/32370281/how-to-embed-image-or-picture-in-jupyter

Out[120]: <IPython.core.display.Image object>

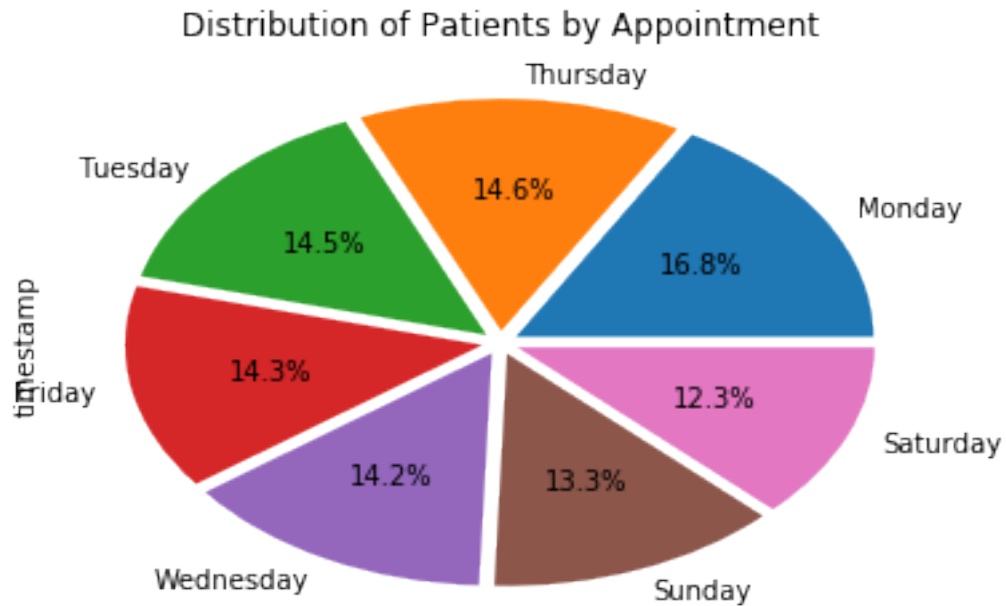
```

13.3.1 Dog Named Bo which seems to be Standard Poodle is one of the highest rated dog, but with highest Retweet and and favourite count of 42228, 95450 respectively

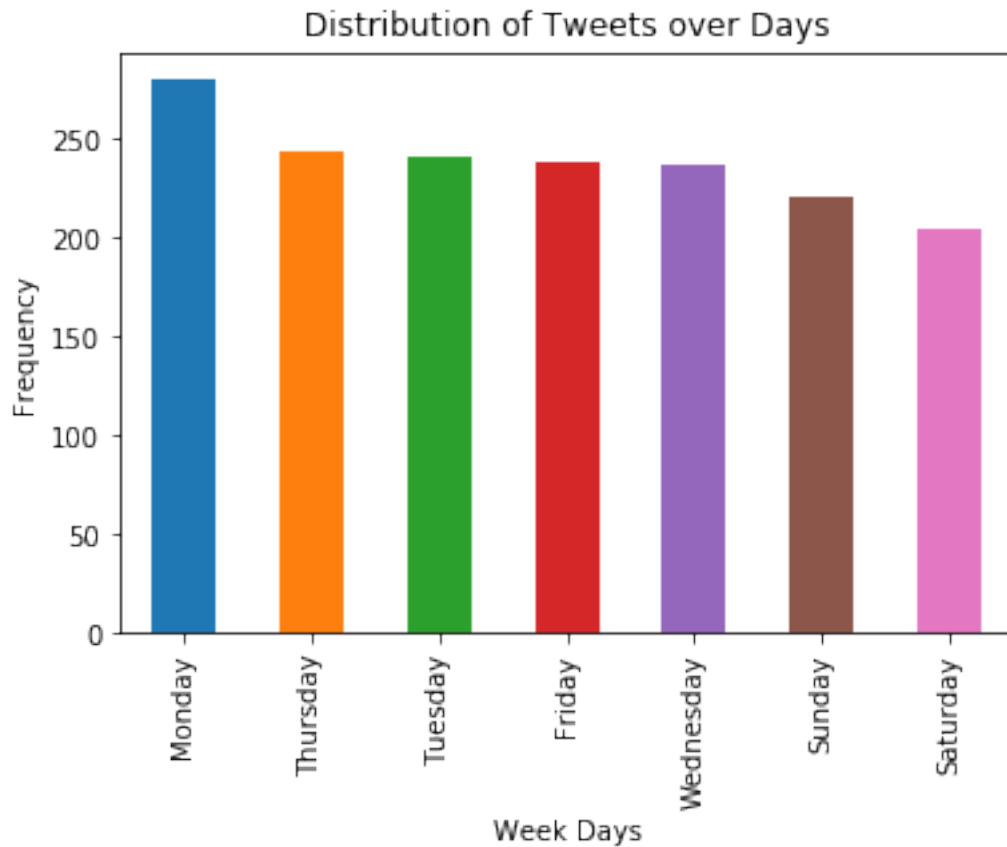
<https://www.traditionsvetcenters.com/services/dogs/breeds/standard-poodle>

13.4 4. Most Tweeted Day to rate dogs

```
In [121]: ax=twitter_archive_master.timestamp.dt.weekday_name.value_counts().plot.pie(explode =  
ax.set_title("Distribution of Patients by Appointment");
```



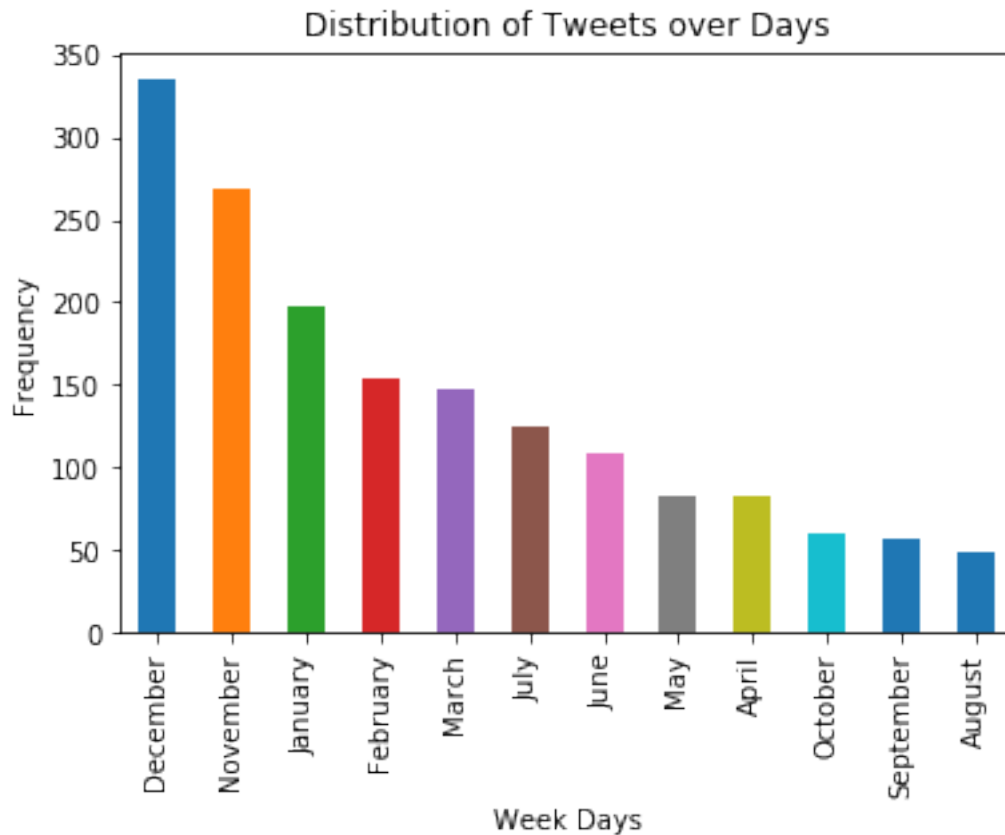
```
In [122]: twitter_archive_master.timestamp.dt.weekday_name.value_counts().plot('bar')  
plt.title("Distribution of Tweets over Days")  
plt.xlabel('Week Days')  
plt.ylabel('Frequency');
```



On Monday People Tweeted the most

13.5 5. Most Tweeted Month to rate dogs

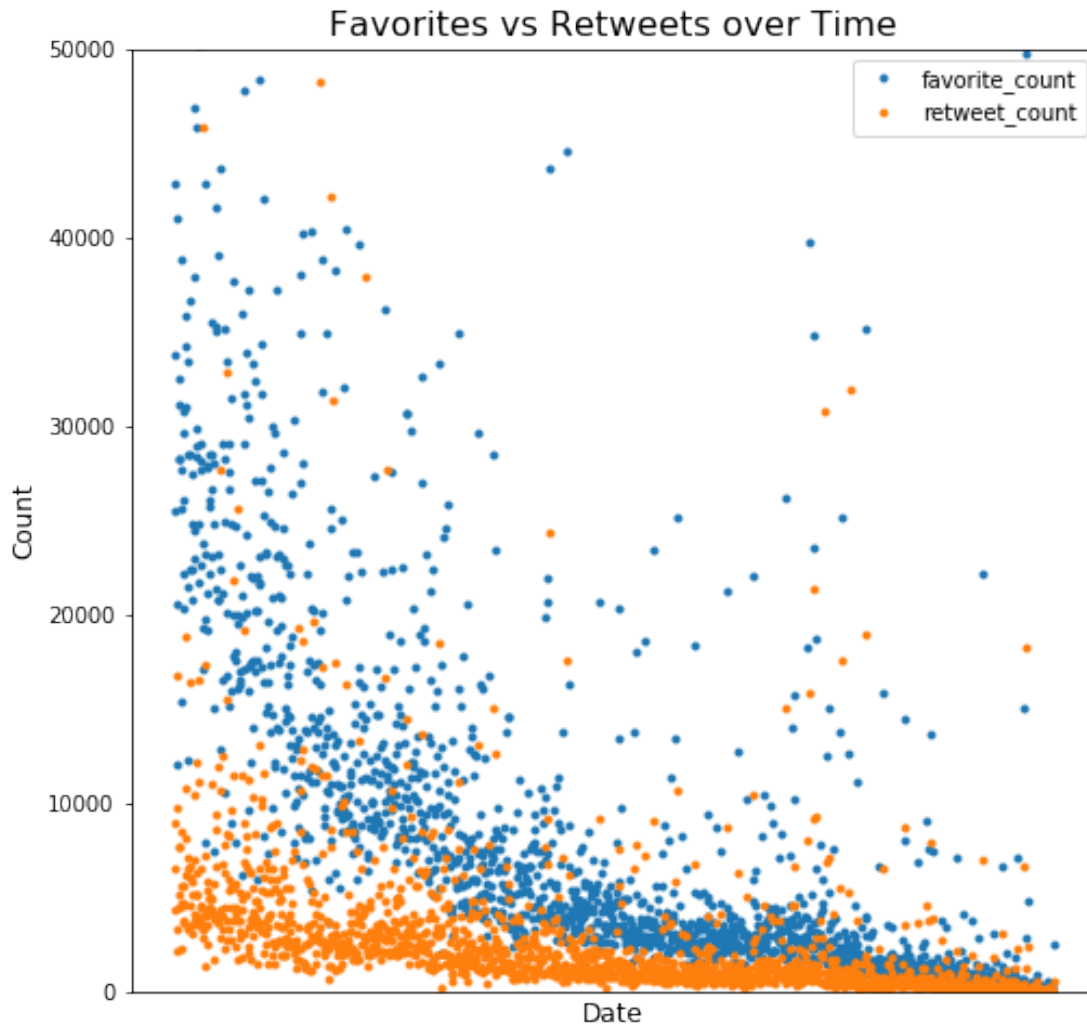
```
In [123]: twitter_archive_master.timestamp.dt.month_name().value_counts().plot('bar')
plt.title("Distribution of Tweets over Days")
plt.xlabel('Week Days')
plt.ylabel('Frequency');
```



13.5.1 December is the most Tweeted mont. I assume people are in the festive mood, so they visit this page to refresh. And it interesting to note that Summer and Fall is the least tweeted season, I think people usually go out on this day and does not spend much time on twitter

13.6 6. Favourite vs Retweet over the time

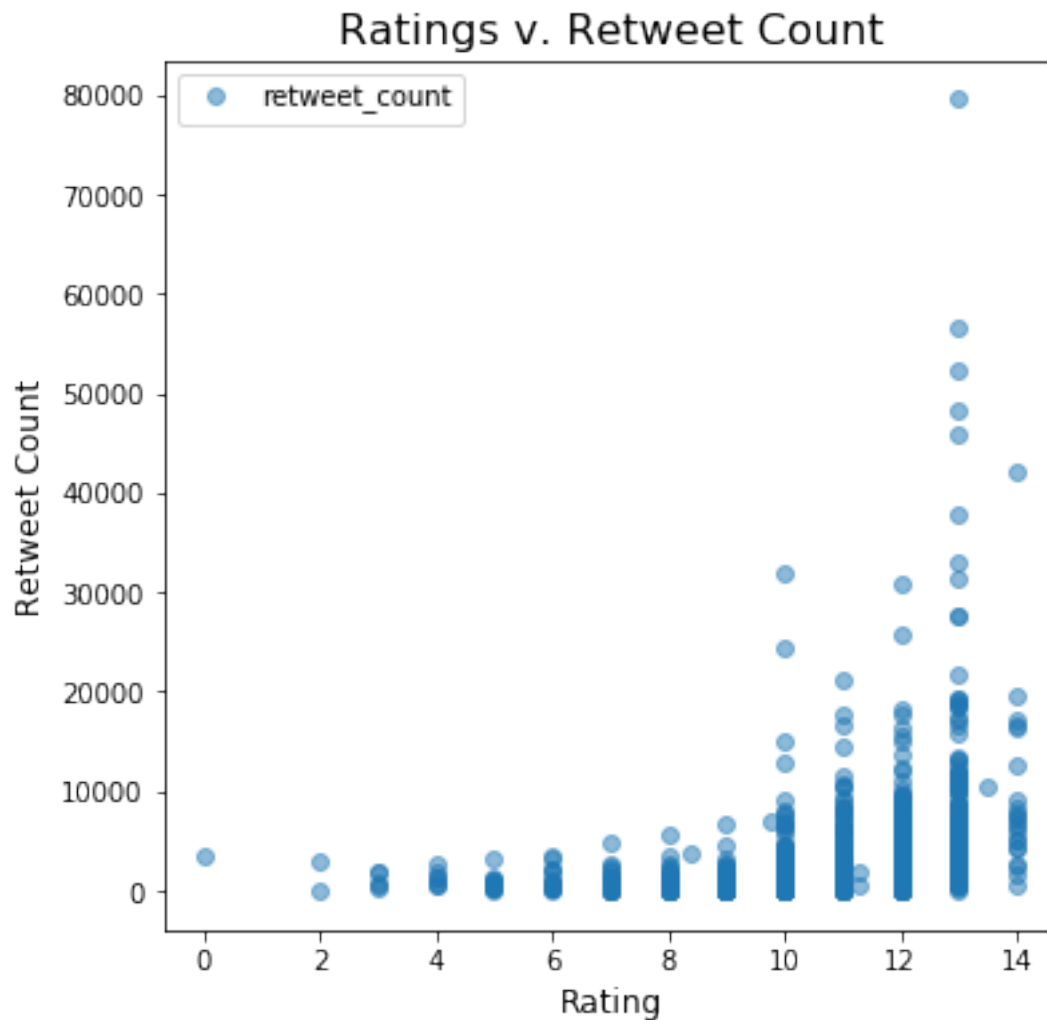
```
In [124]: twitter_archive_master[['favorite_count', 'retweet_count']].plot(style =
      '.', ylim=[0, 50000], figsize=(8,8))
plt.title('Favorites vs Retweets over Time', size=16)
plt.xlabel('Date', size=12)
plt.xticks([], [])
plt.ylabel('Count', size=12)
plt.legend(ncol=1, loc='upper right')
plt.savefig('retweets-favorites-time.png');
```



13.6.1 Favorites are more popular than retweets. Both decreasing over time, and retweets even more so.

13.7 7. Rating vs Retweet Count

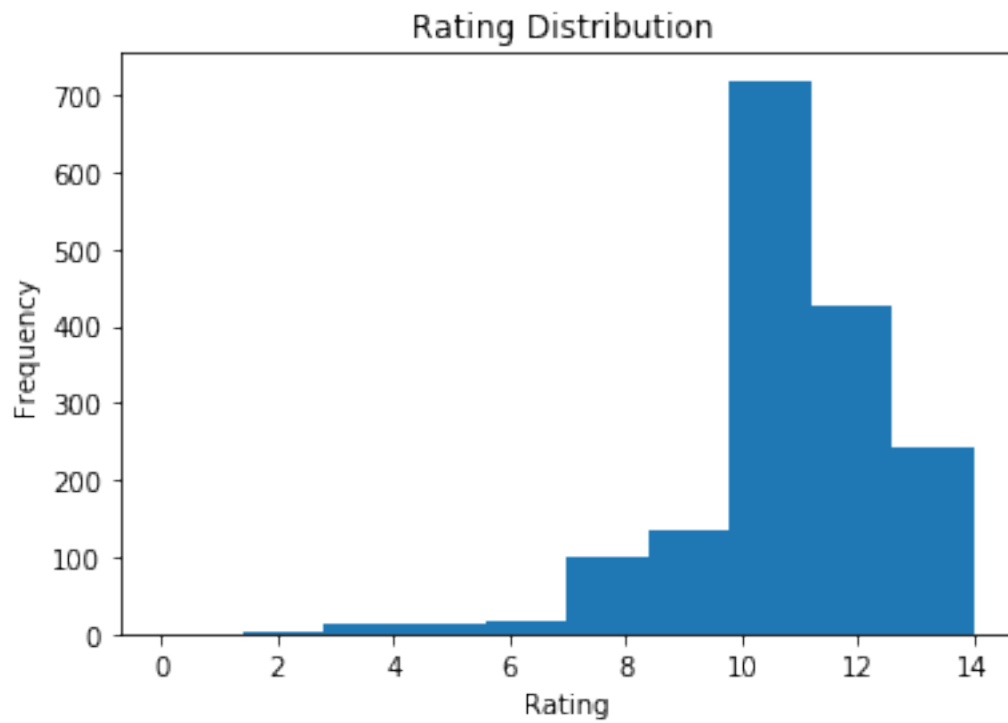
```
In [125]: ax = twitter_archive_master.plot(x = 'rating_numerator', y = 'retweet_count', style =
        'o', alpha=.5, figsize=(6,6));
ax.set_title('Ratings v. Retweet Count', size=16)
ax.set_xlabel('Rating', size=12)
ax.set_ylabel('Retweet Count', size=12);
```



13.7.1 Higher rated dogs have more retweets.

13.8 8. Rating Distribution

```
In [126]: plt.hist(twitter_archive_master.rating_numerator)
plt.title("Rating Distribution ")
plt.xlabel("Rating")
plt.ylabel("Frequency");
```

13.8.1 Most of the ratings are between 10 and 11

In []: