

wrangle_act

November 17, 2018

```
In [1]: import pandas as pd
import numpy as np
import requests
import tweepy
import os
import json
import time
import re
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
from skimage import io
```

0.1 Gather Data

0.1.1 Twitter archive enhanced file

```
In [2]: # read csv as a pandas dataframe
twitter_archive = pd.read_csv('twitter-archive-enhanced.csv')
twitter_archive.head()
```

```
Out[2]:
```

	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 [3]: twitter_archive.info()
```

```

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

```

```

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

```

0.1.2 Image prediction file download

In [4]: *# Download the image predictions file using requests library*

```

url="https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions
response = requests.get(url)

```

Save the response content to a tab-separated (tsv) file on our current root folder

```

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

```

Read and display the image predictions file content

```

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

```

In [5]: `image_predictions.head()`

```

Out[5]:
      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

```

0.1.3 Twitter Tweets API

```
In [6]: CONSUMER_KEY = "eStzV76oj11xEDwYhk3Jzy2F7"
        CONSUMER_SECRET = "mjFR2ALQ6eFUIFlk2A1UhEPkfCtde7L2MksSxi6P7TLw7mNdF9"
        OAUTH_TOKEN = "2810107749-xJhCR7hL3gszeaizqDnheIKOYd56TeaWDK1xjqD"
        OAUTH_TOKEN_SECRET = "iMFm7vzyecTx06YumRjQ059gPLi8QRGAC6TDfwcQlLntP"
        auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
        auth.set_access_token(OAUTH_TOKEN, OAUTH_TOKEN_SECRET)
        api = tweepy.API(auth)
```

```
In [7]: # List of the error tweets
        error_list = []
        # List of tweets
        df_list = []
        # Calculate the time of execution
        start = time.time()

        # For loop which will add each available tweet json to df_list
        for tweet_id in twitter_archive['tweet_id']:
            try:
                tweet = api.get_status(tweet_id, tweet_mode='extended',
                                       wait_on_rate_limit = True, wait_on_rate_limit_notify = True)
                favorites = tweet['favorite_count'] # How many favorites the tweet had
                retweets = tweet['retweet_count'] # Count of the retweet
                user_followers = tweet['user']['followers_count'] # How many followers the user has
                user_favourites = tweet['user']['favourites_count'] # How many favorites the user has
                date_time = tweet['created_at'] # The date and time of the creation

                df_list.append({'tweet_id': int(tweet_id),
                              'favorites': int(favorites),
                              'retweets': int(retweets),
                              'user_followers': int(user_followers),
                              'user_favourites': int(user_favourites),
                              'date_time': pd.to_datetime(date_time)})
            except Exception as e:
                print(str(tweet_id)+ " _ " + str(e))
                error_list.append(tweet_id)
        # Calculate the time of execution
        end = time.time()
        print(end - start)
```

```
888202515573088257 _ [{'code': 144, 'message': 'No status found with that ID.'}]
873697596434513921 _ [{'code': 144, 'message': 'No status found with that ID.'}]
872668790621863937 _ [{'code': 144, 'message': 'No status found with that ID.'}]
869988702071779329 _ [{'code': 144, 'message': 'No status found with that ID.'}]
866816280283807744 _ [{'code': 144, 'message': 'No status found with that ID.'}]
861769973181624320 _ [{'code': 144, 'message': 'No status found with that ID.'}]
845459076796616705 _ [{'code': 144, 'message': 'No status found with that ID.'}]
842892208864923648 _ [{'code': 144, 'message': 'No status found with that ID.'}]
```

```

837012587749474308 _ [{'code': 144, 'message': 'No status found with that ID.'}]
827228250799742977 _ [{'code': 144, 'message': 'No status found with that ID.'}]
802247111496568832 _ [{'code': 144, 'message': 'No status found with that ID.'}]
775096608509886464 _ [{'code': 144, 'message': 'No status found with that ID.'}]
770743923962707968 _ [{'code': 144, 'message': 'No status found with that ID.'}]
Rate limit reached. Sleeping for: 739
754011816964026368 _ [{'code': 144, 'message': 'No status found with that ID.'}]
Rate limit reached. Sleeping for: 739
1909.8017165660858

```

```

In [8]: # length of the result list
        print("The length of the result", len(df_list))

        # length of the error list
        print("The length of the errors", len(error_list))

```

```

The length of the result 2342
The length of the errors 14

```

The length of the result 2342 & the length of the errors 14

```

In [9]: # Create DataFrames from list of dictionaries
        json_tweets = pd.DataFrame(df_list, columns = ['tweet_id', 'favorites', 'retweets',
                                                    'user_followers', 'user_favourites', 'data'])

        # Save the dataframe in file
        json_tweets.to_csv('tweet_json.txt', encoding = 'utf-8', index=False)

```

0.2 Assess Data

Check tweet_json.txt file content

```

In [10]: # Read the saved tweet_json.txt file into a dataframe
         tweet_data = pd.read_csv('tweet_json.txt', encoding = 'utf-8')
         tweet_data

```

```

Out[10]:
   tweet_id  favorites  retweets  user_followers \
0  892420643555336193    38221     8374      7431629
1  892177421306343426    32767     6182      7431629
2  891815181378084864    24674     4091      7431629
3  891689557279858688    41551     8512      7431630
4  891327558926688256    39737     9218      7431630
5  891087950875897856    19946     3065      7431630
6  890971913173991426    11670     2034      7431630
7  890729181411237888    64506    18583      7431630
8  890609185150312448    27415     4207      7431630
9  890240255349198849    31455     7275      7431630
10 890006608113172480    30242     7221      7431630

```

11	889880896479866881	27397	4900	7431630
12	889665388333682689	47439	9904	7431630
13	889638837579907072	26784	4477	7431630
14	889531135344209921	14888	2208	7431630
15	889278841981685760	24898	5300	7431630
16	888917238123831296	28676	4427	7431630
17	888804989199671297	25204	4234	7431630
18	888554962724278272	19556	3498	7431630
19	888078434458587136	21450	3445	7431630
20	887705289381826560	29745	5308	7431630
21	887517139158093824	45619	11493	7431630
22	887473957103951883	68114	17913	7431630
23	887343217045368832	33208	10247	7431630
24	887101392804085760	30132	5875	7431630
25	886983233522544640	34671	7657	7431630
26	886736880519319552	11879	3232	7431630
27	886680336477933568	22120	4396	7431630
28	886366144734445568	20885	3152	7431630
29	886267009285017600	116	4	7431630
...
2312	666411507551481857	440	323	7431653
2313	666407126856765440	106	40	7431653
2314	666396247373291520	167	84	7431653
2315	666373753744588802	186	92	7431653
2316	666362758909284353	769	564	7431653
2317	666353288456101888	218	71	7431653
2318	666345417576210432	290	133	7431653
2319	666337882303524864	197	90	7431653
2320	666293911632134144	494	352	7431653
2321	666287406224695296	146	64	7431653
2322	666273097616637952	173	76	7431653
2323	666268910803644416	102	35	7431653
2324	666104133288665088	14167	6486	7431653
2325	666102155909144576	80	12	7431653
2326	666099513787052032	153	67	7431653
2327	666094000022159362	163	72	7431653
2328	666082916733198337	116	44	7431653
2329	666073100786774016	318	162	7431653
2330	666071193221509120	144	59	7431653
2331	666063827256086533	471	215	7431653
2332	666058600524156928	111	57	7431653
2333	666057090499244032	293	141	7431653
2334	666055525042405380	431	244	7431653
2335	666051853826850816	1207	841	7431653
2336	666050758794694657	131	59	7431653
2337	666049248165822465	107	40	7431653
2338	666044226329800704	295	139	7431653
2339	666033412701032449	125	44	7431653

2340	666029285002620928	129	47	7431653
2341	666020888022790149	2543	508	7431653

	user_favourites	date_time
0	139602	2017-08-01 16:23:56
1	139602	2017-08-01 00:17:27
2	139602	2017-07-31 00:18:03
3	139602	2017-07-30 15:58:51
4	139602	2017-07-29 16:00:24
5	139602	2017-07-29 00:08:17
6	139602	2017-07-28 16:27:12
7	139602	2017-07-28 00:22:40
8	139602	2017-07-27 16:25:51
9	139602	2017-07-26 15:59:51
10	139602	2017-07-26 00:31:25
11	139602	2017-07-25 16:11:53
12	139602	2017-07-25 01:55:32
13	139602	2017-07-25 00:10:02
14	139602	2017-07-24 17:02:04
15	139602	2017-07-24 00:19:32
16	139602	2017-07-23 00:22:39
17	139602	2017-07-22 16:56:37
18	139602	2017-07-22 00:23:06
19	139602	2017-07-20 16:49:33
20	139602	2017-07-19 16:06:48
21	139602	2017-07-19 03:39:09
22	139602	2017-07-19 00:47:34
23	139602	2017-07-18 16:08:03
24	139602	2017-07-18 00:07:08
25	139602	2017-07-17 16:17:36
26	139602	2017-07-16 23:58:41
27	139602	2017-07-16 20:14:00
28	139602	2017-07-15 23:25:31
29	139602	2017-07-15 16:51:35
...
2312	139602	2015-11-17 00:24:19
2313	139602	2015-11-17 00:06:54
2314	139602	2015-11-16 23:23:41
2315	139602	2015-11-16 21:54:18
2316	139602	2015-11-16 21:10:36
2317	139602	2015-11-16 20:32:58
2318	139602	2015-11-16 20:01:42
2319	139602	2015-11-16 19:31:45
2320	139602	2015-11-16 16:37:02
2321	139602	2015-11-16 16:11:11
2322	139602	2015-11-16 15:14:19
2323	139602	2015-11-16 14:57:41
2324	139602	2015-11-16 04:02:55

2325	139602	2015-11-16	03:55:04
2326	139602	2015-11-16	03:44:34
2327	139602	2015-11-16	03:22:39
2328	139602	2015-11-16	02:38:37
2329	139602	2015-11-16	01:59:36
2330	139602	2015-11-16	01:52:02
2331	139602	2015-11-16	01:22:45
2332	139602	2015-11-16	01:01:59
2333	139602	2015-11-16	00:55:59
2334	139602	2015-11-16	00:49:46
2335	139602	2015-11-16	00:35:11
2336	139602	2015-11-16	00:30:50
2337	139602	2015-11-16	00:24:50
2338	139602	2015-11-16	00:04:52
2339	139602	2015-11-15	23:21:54
2340	139602	2015-11-15	23:05:30
2341	139602	2015-11-15	22:32:08

[2342 rows x 6 columns]

Check twitter-archive-enhanced.csv file content

In [11]: twitter_archive

Out[11]:

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

23	887473957103951883	NaN	NaN
24	887343217045368832	NaN	NaN
25	887101392804085760	NaN	NaN
26	886983233522544640	NaN	NaN
27	886736880519319552	NaN	NaN
28	886680336477933568	NaN	NaN
29	886366144734445568	NaN	NaN
...
2326	666411507551481857	NaN	NaN
2327	666407126856765440	NaN	NaN
2328	666396247373291520	NaN	NaN
2329	666373753744588802	NaN	NaN
2330	666362758909284353	NaN	NaN
2331	666353288456101888	NaN	NaN
2332	666345417576210432	NaN	NaN
2333	666337882303524864	NaN	NaN
2334	666293911632134144	NaN	NaN
2335	666287406224695296	NaN	NaN
2336	666273097616637952	NaN	NaN
2337	666268910803644416	NaN	NaN
2338	666104133288665088	NaN	NaN
2339	666102155909144576	NaN	NaN
2340	666099513787052032	NaN	NaN
2341	666094000022159362	NaN	NaN
2342	666082916733198337	NaN	NaN
2343	666073100786774016	NaN	NaN
2344	666071193221509120	NaN	NaN
2345	666063827256086533	NaN	NaN
2346	666058600524156928	NaN	NaN
2347	666057090499244032	NaN	NaN
2348	666055525042405380	NaN	NaN
2349	666051853826850816	NaN	NaN
2350	666050758794694657	NaN	NaN
2351	666049248165822465	NaN	NaN
2352	666044226329800704	NaN	NaN
2353	666033412701032449	NaN	NaN
2354	666029285002620928	NaN	NaN
2355	666020888022790149	NaN	NaN

	timestamp \
0	2017-08-01 16:23:56 +0000
1	2017-08-01 00:17:27 +0000
2	2017-07-31 00:18:03 +0000
3	2017-07-30 15:58:51 +0000
4	2017-07-29 16:00:24 +0000
5	2017-07-29 00:08:17 +0000
6	2017-07-28 16:27:12 +0000
7	2017-07-28 00:22:40 +0000

8	2017-07-27	16:25:51	+0000
9	2017-07-26	15:59:51	+0000
10	2017-07-26	00:31:25	+0000
11	2017-07-25	16:11:53	+0000
12	2017-07-25	01:55:32	+0000
13	2017-07-25	00:10:02	+0000
14	2017-07-24	17:02:04	+0000
15	2017-07-24	00:19:32	+0000
16	2017-07-23	00:22:39	+0000
17	2017-07-22	16:56:37	+0000
18	2017-07-22	00:23:06	+0000
19	2017-07-21	01:02:36	+0000
20	2017-07-20	16:49:33	+0000
21	2017-07-19	16:06:48	+0000
22	2017-07-19	03:39:09	+0000
23	2017-07-19	00:47:34	+0000
24	2017-07-18	16:08:03	+0000
25	2017-07-18	00:07:08	+0000
26	2017-07-17	16:17:36	+0000
27	2017-07-16	23:58:41	+0000
28	2017-07-16	20:14:00	+0000
29	2017-07-15	23:25:31	+0000
...			...
2326	2015-11-17	00:24:19	+0000
2327	2015-11-17	00:06:54	+0000
2328	2015-11-16	23:23:41	+0000
2329	2015-11-16	21:54:18	+0000
2330	2015-11-16	21:10:36	+0000
2331	2015-11-16	20:32:58	+0000
2332	2015-11-16	20:01:42	+0000
2333	2015-11-16	19:31:45	+0000
2334	2015-11-16	16:37:02	+0000
2335	2015-11-16	16:11:11	+0000
2336	2015-11-16	15:14:19	+0000
2337	2015-11-16	14:57:41	+0000
2338	2015-11-16	04:02:55	+0000
2339	2015-11-16	03:55:04	+0000
2340	2015-11-16	03:44:34	+0000
2341	2015-11-16	03:22:39	+0000
2342	2015-11-16	02:38:37	+0000
2343	2015-11-16	01:59:36	+0000
2344	2015-11-16	01:52:02	+0000
2345	2015-11-16	01:22:45	+0000
2346	2015-11-16	01:01:59	+0000
2347	2015-11-16	00:55:59	+0000
2348	2015-11-16	00:49:46	+0000
2349	2015-11-16	00:35:11	+0000
2350	2015-11-16	00:30:50	+0000

2351 2015-11-16 00:24:50 +0000
2352 2015-11-16 00:04:52 +0000
2353 2015-11-15 23:21:54 +0000
2354 2015-11-15 23:05:30 +0000
2355 2015-11-15 22:32:08 +0000

```

                                source \
0      <a href="http://twitter.com/download/iphone" r...
1      <a href="http://twitter.com/download/iphone" r...
2      <a href="http://twitter.com/download/iphone" r...
3      <a href="http://twitter.com/download/iphone" r...
4      <a href="http://twitter.com/download/iphone" r...
5      <a href="http://twitter.com/download/iphone" r...
6      <a href="http://twitter.com/download/iphone" r...
7      <a href="http://twitter.com/download/iphone" r...
8      <a href="http://twitter.com/download/iphone" r...
9      <a href="http://twitter.com/download/iphone" r...
10     <a href="http://twitter.com/download/iphone" r...
11     <a href="http://twitter.com/download/iphone" r...
12     <a href="http://twitter.com/download/iphone" r...
13     <a href="http://twitter.com/download/iphone" r...
14     <a href="http://twitter.com/download/iphone" r...
15     <a href="http://twitter.com/download/iphone" r...
16     <a href="http://twitter.com/download/iphone" r...
17     <a href="http://twitter.com/download/iphone" r...
18     <a href="http://twitter.com/download/iphone" r...
19     <a href="http://twitter.com/download/iphone" r...
20     <a href="http://twitter.com/download/iphone" r...
21     <a href="http://twitter.com/download/iphone" r...
22     <a href="http://twitter.com/download/iphone" r...
23     <a href="http://twitter.com/download/iphone" r...
24     <a href="http://twitter.com/download/iphone" r...
25     <a href="http://twitter.com/download/iphone" r...
26     <a href="http://twitter.com/download/iphone" r...
27     <a href="http://twitter.com/download/iphone" r...
28     <a href="http://twitter.com/download/iphone" r...
29     <a href="http://twitter.com/download/iphone" r...
...
2326  <a href="http://twitter.com/download/iphone" r...
2327  <a href="http://twitter.com/download/iphone" r...
2328  <a href="http://twitter.com/download/iphone" r...
2329  <a href="http://twitter.com/download/iphone" r...
2330  <a href="http://twitter.com/download/iphone" r...
2331  <a href="http://twitter.com/download/iphone" r...
2332  <a href="http://twitter.com/download/iphone" r...
2333  <a href="http://twitter.com/download/iphone" r...
2334  <a href="http://twitter.com/download/iphone" r...
2335  <a href="http://twitter.com/download/iphone" r...
```

2336 <a href="http://twitter.com/download/iphone" r...
 2337 <a href="http://twitter.com/download/iphone" r...
 2338 <a href="http://twitter.com/download/iphone" r...
 2339 <a href="http://twitter.com/download/iphone" r...
 2340 <a href="http://twitter.com/download/iphone" r...
 2341 <a href="http://twitter.com/download/iphone" r...
 2342 <a href="http://twitter.com/download/iphone" r...
 2343 <a href="http://twitter.com/download/iphone" r...
 2344 <a href="http://twitter.com/download/iphone" r...
 2345 <a href="http://twitter.com/download/iphone" r...
 2346 <a href="http://twitter.com/download/iphone" r...
 2347 <a href="http://twitter.com/download/iphone" r...
 2348 <a href="http://twitter.com/download/iphone" r...
 2349 <a href="http://twitter.com/download/iphone" r...
 2350 <a href="http://twitter.com/download/iphone" r...
 2351 <a href="http://twitter.com/download/iphone" r...
 2352 <a href="http://twitter.com/download/iphone" r...
 2353 <a href="http://twitter.com/download/iphone" r...
 2354 <a href="http://twitter.com/download/iphone" r...
 2355 <a href="http://twitter.com/download/iphone" r...

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

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

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

11	NaN	NaN
12	NaN	NaN
13	NaN	NaN
14	NaN	NaN
15	NaN	NaN
16	NaN	NaN
17	NaN	NaN
18	NaN	NaN
19	4.196984e+09	2017-07-19 00:47:34 +0000
20	NaN	NaN
21	NaN	NaN
22	NaN	NaN
23	NaN	NaN
24	NaN	NaN
25	NaN	NaN
26	NaN	NaN
27	NaN	NaN
28	NaN	NaN
29	NaN	NaN
...
2326	NaN	NaN
2327	NaN	NaN
2328	NaN	NaN
2329	NaN	NaN
2330	NaN	NaN
2331	NaN	NaN
2332	NaN	NaN
2333	NaN	NaN
2334	NaN	NaN
2335	NaN	NaN
2336	NaN	NaN
2337	NaN	NaN
2338	NaN	NaN
2339	NaN	NaN
2340	NaN	NaN
2341	NaN	NaN
2342	NaN	NaN
2343	NaN	NaN
2344	NaN	NaN
2345	NaN	NaN
2346	NaN	NaN
2347	NaN	NaN
2348	NaN	NaN
2349	NaN	NaN
2350	NaN	NaN
2351	NaN	NaN
2352	NaN	NaN
2353	NaN	NaN

2354	NaN	NaN
2355	NaN	NaN

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

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

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10	Phineas	None	None	None	None
1	10	Tilly	None	None	None	None
2	10	Archie	None	None	None	None
3	10	Darla	None	None	None	None
4	10	Franklin	None	None	None	None
5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None
8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
19	10	Canela	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None

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

[2356 rows x 17 columns]

```
In [12]: twitter_archive['rating_numerator'].value_counts()
```

```
Out[12]: 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 [13]: twitter_archive['rating_denominator'].value_counts()
```

```
Out[13]: 10      2333
         11       3
         50       3
         80       2
         20       2
          2       1
         16       1
         40       1
         70       1
         15       1
         90       1
        110       1
        120       1
        130       1
        150       1
```

```

170      1
7        1
0        1
Name: rating_denominator, dtype: int64

```

```
In [13]: twitter_archive['name'].value_counts()
```

```

Out[13]: None      745
a              55
Charlie       12
Lucy          11
Oliver        11
Cooper        11
Penny         10
Tucker        10
Lola           10
Winston        9
Bo             9
the            8
Sadie          8
Bailey         7
Daisy          7
Buddy          7
Toby           7
an             7
Scout          6
Jack           6
Leo            6
Oscar          6
Bella          6
Jax            6
Dave           6
Rusty          6
Koda           6
Stanley        6
Milo           6
Sunny          5
...
Harnold        1
Loomis         1
Spencer        1
Dido           1
Jersey         1
Skye           1
Alfy           1
Dobby          1
Pavlov         1
Aja            1

```

Adele	1
Bertson	1
Jeb	1
Terrenth	1
Jomathan	1
Koko	1
Kody	1
Alf	1
Anna	1
Carper	1
Lance	1
Reptar	1
Edgar	1
Brady	1
Glacier	1
Akumi	1
BeBe	1
Murphy	1
Jazzy	1
Bonaparte	1

Name: name, Length: 957, dtype: int64

```
In [14]: # Look up number of retweets usually starting with RT
         len((twitter_archive[twitter_archive['text'].str.startswith('RT') == True]))
```

Out[14]: 183

```
In [15]: # Look for tweets with no image
         twitter_archive['expanded_urls'].isnull().value_counts()
```

```
Out[15]: False    2297
         True      59
         Name: expanded_urls, dtype: int64
```

Check image_predictions.tsv file content

```
In [16]: image_predictions.info()
         image_predictions
```

```
<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

```

```

Out[16]:
      tweet_id      jpg_url \
0    666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
1    666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2    666033412701032449  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3    666044226329800704  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4    666049248165822465  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
5    666050758794694657  https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6    666051853826850816  https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7    666055525042405380  https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8    666057090499244032  https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9    666058600524156928  https://pbs.twimg.com/media/CT5Qw94XAAA-2dP.jpg
10   666063827256086533  https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11   666071193221509120  https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
12   666073100786774016  https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13   666082916733198337  https://pbs.twimg.com/media/CT5m4VGWEAAKc8.jpg
14   666094000022159362  https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg
15   666099513787052032  https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16   666102155909144576  https://pbs.twimg.com/media/CT54YGiWUAENZnoK.jpg
17   666104133288665088  https://pbs.twimg.com/media/CT56LSZWAA1Jj2.jpg
18   666268910803644416  https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
19   666273097616637952  https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
20   666287406224695296  https://pbs.twimg.com/media/CT8g3BpUEAAUfjg.jpg
21   666293911632134144  https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22   666337882303524864  https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg
23   666345417576210432  https://pbs.twimg.com/media/CT9Vn7PWAA_ZCM.jpg
24   666353288456101888  https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25   666362758909284353  https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg
26   666373753744588802  https://pbs.twimg.com/media/CT9vZEYWUAA1Z05.jpg
27   666396247373291520  https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28   666407126856765440  https://pbs.twimg.com/media/CT-NvwmW4AAUGZ.jpg
29   666411507551481857  https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...
2045 886366144734445568  https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046 886680336477933568  https://pbs.twimg.com/media/DE4fEDzWAAAYHMM.jpg
2047 886736880519319552  https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
2048 886983233522544640  https://pbs.twimg.com/media/DE8yicJW0AAAABJ.jpg
2049 887101392804085760  https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050 887343217045368832  https://pbs.twimg.com/ext_tw_video_thumb/88734...
2051 887473957103951883  https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg

```

2052	887517139158093824	https://pbs.twimg.com/ext_tw_video_thumb/88751...
2053	887705289381826560	https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054	888078434458587136	https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
2055	888202515573088257	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056	888554962724278272	https://pbs.twimg.com/media/DFTH_0-UQAAACu20.jpg
2057	888804989199671297	https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058	888917238123831296	https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg
2059	889278841981685760	https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060	889531135344209921	https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg
2061	889638837579907072	https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062	889665388333682689	https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063	889880896479866881	https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg
2064	890006608113172480	https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065	890240255349198849	https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg
2066	890609185150312448	https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067	890729181411237888	https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068	890971913173991426	https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg
2069	891087950875897856	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070	891327558926688256	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071	891689557279858688	https://pbs.twimg.com/media/DF_q7IAWsAEuN8.jpg
2072	891815181378084864	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073	892177421306343426	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074	892420643555336193	https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

	img_num		p1	p1_conf	p1_dog	\
0	1	Welsh_springer_spaniel	0.465074	True		
1	1	redbone	0.506826	True		
2	1	German_shepherd	0.596461	True		
3	1	Rhodesian_ridgeback	0.408143	True		
4	1	miniature_pinscher	0.560311	True		
5	1	Bernese_mountain_dog	0.651137	True		
6	1	box_turtle	0.933012	False		
7	1	chow	0.692517	True		
8	1	shopping_cart	0.962465	False		
9	1	miniature_poodle	0.201493	True		
10	1	golden_retriever	0.775930	True		
11	1	Gordon_setter	0.503672	True		
12	1	Walker_hound	0.260857	True		
13	1	pug	0.489814	True		
14	1	bloodhound	0.195217	True		
15	1	Lhasa	0.582330	True		
16	1	English_setter	0.298617	True		
17	1	hen	0.965932	False		
18	1	desktop_computer	0.086502	False		
19	1	Italian_greyhound	0.176053	True		
20	1	Maltese_dog	0.857531	True		
21	1	three-toed_sloth	0.914671	False		
22	1	ox	0.416669	False		

23	1	golden_retriever	0.858744	True
24	1	malamute	0.336874	True
25	1	guinea_pig	0.996496	False
26	1	soft-coated_wheaten_terrier	0.326467	True
27	1	Chihuahua	0.978108	True
28	1	black-and-tan_coonhound	0.529139	True
29	1	coho	0.404640	False
...
2045	1	French_bulldog	0.999201	True
2046	1	convertible	0.738995	False
2047	1	kuvasz	0.309706	True
2048	2	Chihuahua	0.793469	True
2049	1	Samoyed	0.733942	True
2050	1	Mexican_hairless	0.330741	True
2051	2	Pembroke	0.809197	True
2052	1	limousine	0.130432	False
2053	1	basset	0.821664	True
2054	1	French_bulldog	0.995026	True
2055	2	Pembroke	0.809197	True
2056	3	Siberian_husky	0.700377	True
2057	1	golden_retriever	0.469760	True
2058	1	golden_retriever	0.714719	True
2059	1	whippet	0.626152	True
2060	1	golden_retriever	0.953442	True
2061	1	French_bulldog	0.991650	True
2062	1	Pembroke	0.966327	True
2063	1	French_bulldog	0.377417	True
2064	1	Samoyed	0.957979	True
2065	1	Pembroke	0.511319	True
2066	1	Irish_terrier	0.487574	True
2067	2	Pomeranian	0.566142	True
2068	1	Appenzeller	0.341703	True
2069	1	Chesapeake_Bay_retriever	0.425595	True
2070	2	basset	0.555712	True
2071	1	paper_towel	0.170278	False
2072	1	Chihuahua	0.716012	True
2073	1	Chihuahua	0.323581	True
2074	1	orange	0.097049	False

	p2	p2_conf	p2_dog	p3 \
0	collie	0.156665	True	Shetland_sheepdog
1	miniature_pinscher	0.074192	True	Rhodesian_ridgeback
2	malinois	0.138584	True	bloodhound
3	redbone	0.360687	True	miniature_pinscher
4	Rottweiler	0.243682	True	Doberman
5	English_springer	0.263788	True	Greater_Swiss_Mountain_dog
6	mud_turtle	0.045885	False	terrapin
7	Tibetan_mastiff	0.058279	True	fur_coat

8	shopping_basket	0.014594	False	golden_retriever
9	komondor	0.192305	True	soft-coated_wheaten_terrier
10	Tibetan_mastiff	0.093718	True	Labrador_retriever
11	Yorkshire_terrier	0.174201	True	Pekinese
12	English_foxhound	0.175382	True	Ibizan_hound
13	bull_mastiff	0.404722	True	French_bulldog
14	German_shepherd	0.078260	True	malinois
15	Shih-Tzu	0.166192	True	Dandie_Dinmont
16	Newfoundland	0.149842	True	borzoi
17	cock	0.033919	False	partridge
18	desk	0.085547	False	bookcase
19	toy_terrier	0.111884	True	basenji
20	toy_poodle	0.063064	True	miniature_poodle
21	otter	0.015250	False	great_grey_owl
22	Newfoundland	0.278407	True	groenendael
23	Chesapeake_Bay_retriever	0.054787	True	Labrador_retriever
24	Siberian_husky	0.147655	True	Eskimo_dog
25	skunk	0.002402	False	hamster
26	Afghan_hound	0.259551	True	briard
27	toy_terrier	0.009397	True	papillon
28	bloodhound	0.244220	True	flat-coated_retriever
29	barracouta	0.271485	False	gar
...
2045	Chihuahua	0.000361	True	Boston_bull
2046	sports_car	0.139952	False	car_wheel
2047	Great_Pyrenees	0.186136	True	Dandie_Dinmont
2048	toy_terrier	0.143528	True	can_opener
2049	Eskimo_dog	0.035029	True	Staffordshire_bullterrier
2050	sea_lion	0.275645	False	Weimaraner
2051	Rhodesian_ridgeback	0.054950	True	beagle
2052	tow_truck	0.029175	False	shopping_cart
2053	redbone	0.087582	True	Weimaraner
2054	pug	0.000932	True	bull_mastiff
2055	Rhodesian_ridgeback	0.054950	True	beagle
2056	Eskimo_dog	0.166511	True	malamute
2057	Labrador_retriever	0.184172	True	English_setter
2058	Tibetan_mastiff	0.120184	True	Labrador_retriever
2059	borzoi	0.194742	True	Saluki
2060	Labrador_retriever	0.013834	True	redbone
2061	boxer	0.002129	True	Staffordshire_bullterrier
2062	Cardigan	0.027356	True	basenji
2063	Labrador_retriever	0.151317	True	muzzle
2064	Pomeranian	0.013884	True	chow
2065	Cardigan	0.451038	True	Chihuahua
2066	Irish_setter	0.193054	True	Chesapeake_Bay_retriever
2067	Eskimo_dog	0.178406	True	Pembroke
2068	Border_collie	0.199287	True	ice_lolly
2069	Irish_terrier	0.116317	True	Indian_elephant

2070	English_springer	0.225770	True	German_short-haired_pointer
2071	Labrador_retriever	0.168086	True	spatula
2072	malamute	0.078253	True	kelpie
2073	Pekinese	0.090647	True	papillon
2074	bagel	0.085851	False	banana

	p3_conf	p3_dog
0	0.061428	True
1	0.072010	True
2	0.116197	True
3	0.222752	True
4	0.154629	True
5	0.016199	True
6	0.017885	False
7	0.054449	False
8	0.007959	True
9	0.082086	True
10	0.072427	True
11	0.109454	True
12	0.097471	True
13	0.048960	True
14	0.075628	True
15	0.089688	True
16	0.133649	True
17	0.000052	False
18	0.079480	False
19	0.111152	True
20	0.025581	True
21	0.013207	False
22	0.102643	True
23	0.014241	True
24	0.093412	True
25	0.000461	False
26	0.206803	True
27	0.004577	True
28	0.173810	True
29	0.189945	False
...
2045	0.000076	True
2046	0.044173	False
2047	0.086346	True
2048	0.032253	False
2049	0.029705	True
2050	0.134203	True
2051	0.038915	True
2052	0.026321	False
2053	0.026236	True
2054	0.000903	True

2055	0.038915	True
2056	0.111411	True
2057	0.073482	True
2058	0.105506	True
2059	0.027351	True
2060	0.007958	True
2061	0.001498	True
2062	0.004633	True
2063	0.082981	False
2064	0.008167	True
2065	0.029248	True
2066	0.118184	True
2067	0.076507	True
2068	0.193548	False
2069	0.076902	False
2070	0.175219	True
2071	0.040836	False
2072	0.031379	True
2073	0.068957	True
2074	0.076110	False

[2075 rows x 12 columns]

0.2.1 Access Summary

Quality

- Some entries in twitter_archive name column is not actual name. For example "a" is not a name.
- Columns "in_reply_to_status_id", "in_reply_to_user_id", "retweeted_status_id", "retweeted_status_user_id" are with wrong data type. It is float and it should be integers/strings.
- There are incorrect entries for rating_numerator and rating_denominator.
- We need to consider only original ratings with images and no retweets.
- We see that in some columns the null values are not treated as null values.
- image_predictions table has only 2075 entries instead of 2356.(missing data).
- image_predictions table contains retweets and some tweet_ids have the same jpg_url.
- There are only 2342 entries instead of 2356 in tweet_data table.

0.2.2 Tidiness

- To join "image_predictions" and "tweet_info" to "twitter_archive"
- Images predictions table need to be made tidy by condensing various columns.
- Dog "stage" variable in four columns i.e doggo, floofer, pupper, puppo.

0.2.3 Clean Data

Task 1 : To combine the DataFrames

```
In [17]: # Make copies of dataframes to clean
```

```
df_clean = twitter_archive.copy()
df1_clean = image_predictions.copy()
df2_clean = tweet_data.copy()
```

Define Merge image_predictions and tweet_info to twitter_archive table.

Code

```
In [18]: # Join df_clean & df3_clean on tweet_id
```

```
df3 = pd.merge(df_clean, df2_clean, left_on = 'tweet_id', right_on = 'tweet_id', how =
```

```
# Join df4 with df2_clean (Image Predictions)
```

```
df4 = pd.merge(df3, df1_clean, left_on = 'tweet_id', right_on = 'tweet_id', how = 'inne
```

Test

```
In [19]: df4.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 2068 entries, 0 to 2067
```

```
Data columns (total 33 columns):
```

tweet_id	2068 non-null int64
in_reply_to_status_id	23 non-null float64
in_reply_to_user_id	23 non-null float64
timestamp	2068 non-null object
source	2068 non-null object
text	2068 non-null object
retweeted_status_id	75 non-null float64
retweeted_status_user_id	75 non-null float64
retweeted_status_timestamp	75 non-null object
expanded_urls	2068 non-null object
rating_numerator	2068 non-null int64
rating_denominator	2068 non-null int64
name	2068 non-null object
doggo	2068 non-null object
floofer	2068 non-null object
pupper	2068 non-null object
puppo	2068 non-null object
favorites	2068 non-null int64
retweets	2068 non-null int64
user_followers	2068 non-null int64
user_favourites	2068 non-null int64
date_time	2068 non-null object
jpg_url	2068 non-null object
img_num	2068 non-null int64
p1	2068 non-null object

```

p1_conf          2068 non-null float64
p1_dog           2068 non-null bool
p2               2068 non-null object
p2_conf          2068 non-null float64
p2_dog           2068 non-null bool
p3               2068 non-null object
p3_conf          2068 non-null float64
p3_dog           2068 non-null bool
dtypes: bool(3), float64(7), int64(8), object(15)
memory usage: 506.9+ KB

```

```

In [20]: # Save new dataframe for next sections
         df4.to_csv('twitter_archive_master.csv', index=False)

```

```

In [21]: df_master = pd.read_csv('twitter_archive_master.csv')

```

```

In [22]: df_master.sample(5)

```

```

Out[22]:
      tweet_id  in_reply_to_status_id  in_reply_to_user_id \
1123  704364645503647744              NaN              NaN
1556  676191832485810177              NaN              NaN
801    749395845976588288              NaN              NaN
1156  700796979434098688              NaN              NaN
894    737310737551491075              NaN              NaN

      timestamp \
1123  2016-02-29 17:56:32 +0000
1556  2015-12-14 00:07:50 +0000
801    2016-07-03 00:14:27 +0000
1156  2016-02-19 21:39:54 +0000
894    2016-05-30 15:52:33 +0000

      source \
1123  <a href="http://twitter.com/download/iphone" r...
1556  <a href="http://twitter.com/download/iphone" r...
801    <a href="http://twitter.com/download/iphone" r...
1156  <a href="http://twitter.com/download/iphone" r...
894    <a href="http://twitter.com/download/iphone" r...

      text  retweeted_status_id \
1123  Say hello to Bisquick. He's a Beneplop Cumbers...      NaN
1556  These two pups just met and have instantly bon...      NaN
801    This is George. He just remembered that bees a...      NaN
1156  This is Daniel. He's a neat pup. Exotic af. Cu...      NaN
894    Everybody stop what you're doing and watch the...      NaN

      retweeted_status_user_id  retweeted_status_timestamp \
1123                          NaN                          NaN

```

1556	NaN	NaN
801	NaN	NaN
1156	NaN	NaN
894	NaN	NaN

	expanded_urls	...	img_num	\
1123	https://twitter.com/dog_rates/status/704364645...	...	1	
1556	https://twitter.com/dog_rates/status/676191832...	...	2	
801	https://twitter.com/dog_rates/status/749395845...	...	1	
1156	https://twitter.com/dog_rates/status/700796979...	...	1	
894	https://twitter.com/dog_rates/status/737310737...	...	1	

	p1	p1_conf	p1_dog		p2	p2_conf	p2_dog	\
1123	Pembroke	0.980695	True		Cardigan	0.018504	True	
1556	Chihuahua	0.376741	True	Italian_greyhound	0.173114	True		
801	Pomeranian	0.973715	True		chow	0.020758	True	
1156	tailed_frog	0.652712	False		tree_frog	0.280212	False	
894	cliff	0.439077	False		lakeside	0.062899	False	

	p3	p3_conf	p3_dog
1123	Chihuahua	0.000215	True
1556	muzzle	0.071485	False
801	keeshond	0.003784	True
1156	bullfrog	0.040177	False
894	valley	0.039758	False

[5 rows x 33 columns]

Task 2 : Melting Doggo, Floofer, Pupper, Puppo columns into one column

Define We will create one column "dog_stages" by melting 4 different columns of Doggo, Floofer, Pupper & Puppo

Code

```
In [23]: # Create and merge dog stages into one column
#df_master['dog_stages'] = df_master.text.str.extract('(doggo/floofer/pupper/puppo)', e

# Drop the unnecessary columns
#df_master.drop(['doggo', 'floofer', 'pupper', 'puppo'], axis=1, inplace = True)

dog_stages = []

x = ['pupper', 'puppo', 'doggo', 'floof']
y = ['pupper', 'puppo', 'doggo', 'floof']

for row in df_master['text']:
```

```

        row = row.lower()
        for word in x:
            if word in str(row):
                dog_stages.append(y[x.index(word)])
                break
        else:
            dog_stages.append('None')

df_master['dog_stages'] = dog_stages

In [24]: df_master['dog_stages'].value_counts()

Out[24]: None      1683
pupper      247
doggo       73
floof       35
puppo       30
Name: dog_stages, dtype: int64

In [25]: # removing extra columns
df_master.drop(['doggo', 'floofer', 'pupper', 'puppo'], axis=1, inplace=True)

```

Test

```

In [26]: df_master.columns

Out[26]: Index(['tweet_id', 'in_reply_to_status_id', 'in_reply_to_user_id', 'timestamp',
'source', 'text', 'retweeted_status_id', 'retweeted_status_user_id',
'retweeted_status_timestamp', 'expanded_urls', 'rating_numerator',
'rating_denominator', 'name', 'favorites', 'retweets', 'user_followers',
'user_favourites', 'date_time', 'jpg_url', 'img_num', 'p1', 'p1_conf',
'p1_dog', 'p2', 'p2_conf', 'p2_dog', 'p3', 'p3_conf', 'p3_dog',
'dog_stages'],
dtype='object')

In [27]: #Checking if any 'null' value in column dog_stages
sum(df_master.dog_stages.isnull())

```

Out[27]: 0

Task 3 : Creating Column 'breed' and 'confidence and removing 'p1', 'p1_conf', 'p1_dog', 'p2', 'p2_conf', 'p2_dog', 'p3', 'p3_conf', 'p3_dog'

Define For making our analysis simpler, we will condense above columns and will categorise in 2 columns - breed and confidence.

Code

```
In [28]: #Creating 2 lists
        breed = []
        conf= []

        #Defining function
        def breed_conf(row):
            if row['p1_dog']:
                breed.append(row['p1'])
                conf.append(row['p1_conf'])
            elif row['p2_dog']:
                breed.append(row['p2'])
                conf.append(row['p2_conf'])
            elif row['p3_dog']:
                breed.append(row['p3'])
                conf.append(row['p3_conf'])
            else:
                breed.append('Unidentifiable')
                conf.append(0)

        #Calling the function
        df_master.apply(breed_conf, axis = 1)

        df_master['breed'] = breed
        df_master['confidence'] = conf

In [29]: #removing the processed columns
        df_master.drop(['p1', 'p1_conf', 'p1_dog', 'p2', 'p2_conf', 'p2_dog', 'p3', 'p3_conf',
```

Test

```
In [30]: df_master.breed.value_counts()

Out[30]: Unidentifiable      323
         golden_retriever    173
         Labrador_retriever  113
         Pembroke            95
         Chihuahua           93
         pug                 65
         toy_poodle          52
         chow                51
         Samoyed             46
         Pomeranian          42
         malamute            34
         cocker_spaniel       33
         Chesapeake_Bay_retriever 31
         French_bulldog       30
         miniature_pinscher   26
```

Cardigan	23
Staffordshire_bullterrier	22
Eskimo_dog	22
German_shepherd	21
beagle	21
Siberian_husky	20
Shih-Tzu	20
Shetland_sheepdog	19
Lakeland_terrier	19
Maltese_dog	19
Rottweiler	19
kuvasz	19
Italian_greyhound	17
basset	17
West_Highland_white_terrier	16
...	
Afghan_hound	4
Rhodesian_ridgeback	4
Weimaraner	4
Scottish_deerhound	4
Welsh_springer_spaniel	4
bluetick	4
keeshond	4
Leonberg	3
komondor	3
cairn	3
briard	3
toy_terrier	3
curly-coated_retriever	3
Brabancon_griffon	3
Irish_water_spaniel	3
Greater_Swiss_Mountain_dog	3
Australian_terrier	2
black-and-tan_coonhound	2
groenendael	2
Appenzeller	2
Sussex_spaniel	2
wire-haired_fox_terrier	2
EntleBucher	1
silky_terrier	1
clumber	1
standard_schnauzer	1
Scotch_terrier	1
Japanese_spaniel	1
Bouvier_des_Flandres	1
Irish_wolfhound	1

Name: breed, Length: 114, dtype: int64


```
In [31]: #Dropping Unidentifiable breeds
df_master = df_master[df_master.breed != 'Unidentifiable']
```

Test

```
In [35]: df_master.breed.value_counts()
```

```
Out[35]: golden_retriever      173
Labrador_retriever           113
Pembroke                     95
Chihuahua                    93
pug                          65
toy_poodle                   52
chow                         51
Samoyed                      46
Pomeranian                   42
malamute                     34
cocker_spaniel               33
Chesapeake_Bay_retriever     31
French_bulldog               30
miniature_pinscher           26
Cardigan                     23
Eskimo_dog                   22
Staffordshire_bullterrier     22
German_shepherd              21
beagle                       21
Shih-Tzu                     20
Siberian_husky               20
Shetland_sheepdog            19
kuvasz                       19
Lakeland_terrier             19
Maltese_dog                  19
Rottweiler                   19
basset                       17
Italian_greyhound            17
West_Highland_white_terrier   16
American_Staffordshire_terrier 16
...
Afghan_hound                  4
Weimaraner                    4
Scottish_deerhound            4
giant_schnauzer               4
Tibetan_terrier               4
bluetick                      4
keeshond                      4
Greater_Swiss_Mountain_dog     3
Leonberg                     3
curly-coated_retriever         3
```

toy_terrier	3
komondor	3
briard	3
Brabancon_griffon	3
cairn	3
Irish_water_spaniel	3
groenendael	2
Sussex_spaniel	2
black-and-tan_coonhound	2
Appenzeller	2
wire-haired_fox_terrier	2
Australian_terrier	2
Irish_wolfhound	1
EntleBucher	1
clumber	1
Japanese_spaniel	1
silky_terrier	1
Bouvier_des_Flandres	1
Scotch_terrier	1
standard_schnauzer	1

Name: breed, Length: 113, dtype: int64

Task 4 : Handle tweets issues & drop unnecessary columns

Define We will handle the tweet issues such as removing duplicate tweets and deleting columns which we do not need for our data analysis.

Code

```
In [32]: #Checking and analyze the column names which I want to delete
df_master.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1745 entries, 1 to 2067
Data columns (total 23 columns):
tweet_id                1745 non-null int64
in_reply_to_status_id   20 non-null float64
in_reply_to_user_id     20 non-null float64
timestamp               1745 non-null object
source                  1745 non-null object
text                    1745 non-null object
retweeted_status_id     60 non-null float64
retweeted_status_user_id 60 non-null float64
retweeted_status_timestamp 60 non-null object
expanded_urls           1745 non-null object
rating_numerator         1745 non-null int64
rating_denominator      1745 non-null int64
name                    1745 non-null object
```

```

favorites          1745 non-null int64
retweets           1745 non-null int64
user_followers     1745 non-null int64
user_favourites    1745 non-null int64
date_time          1745 non-null object
jpg_url            1745 non-null object
img_num            1745 non-null int64
dog_stages         1745 non-null object
breed              1745 non-null object
confidence         1745 non-null float64
dtypes: float64(5), int64(8), object(10)
memory usage: 327.2+ KB

```

Code

```

In [33]: # Delete the retweets
         df_master = df_master[pd.isnull(df_master.retweeted_status_id)]

         # Delete duplicated tweet_id
         df_master = df_master.drop_duplicates()

         # Delete tweets with no pictures
         df_master = df_master.dropna(subset = ['jpg_url'])

         # Test
         len(df_master)

```

Out[33]: 1685

Removing the columns which are not required in analysis

```

In [34]: df_master.drop(['in_reply_to_status_id', 'in_reply_to_user_id', 'retweeted_status_id', 're

```

Test

```

In [196]: df_master

```

```

Out[196]:
   tweet_id  timestamp \
1  892177421306343426  2017-08-01 00:17:27 +0000
2  891815181378084864  2017-07-31 00:18:03 +0000
3  891689557279858688  2017-07-30 15:58:51 +0000
4  891327558926688256  2017-07-29 16:00:24 +0000
5  891087950875897856  2017-07-29 00:08:17 +0000
6  890971913173991426  2017-07-28 16:27:12 +0000
7  890729181411237888  2017-07-28 00:22:40 +0000
8  890609185150312448  2017-07-27 16:25:51 +0000
9  890240255349198849  2017-07-26 15:59:51 +0000
10 890006608113172480  2017-07-26 00:31:25 +0000

```

11	889880896479866881	2017-07-25	16:11:53	+0000
12	889665388333682689	2017-07-25	01:55:32	+0000
13	889638837579907072	2017-07-25	00:10:02	+0000
14	889531135344209921	2017-07-24	17:02:04	+0000
15	889278841981685760	2017-07-24	00:19:32	+0000
16	888917238123831296	2017-07-23	00:22:39	+0000
17	888804989199671297	2017-07-22	16:56:37	+0000
18	888554962724278272	2017-07-22	00:23:06	+0000
19	888078434458587136	2017-07-20	16:49:33	+0000
20	887705289381826560	2017-07-19	16:06:48	+0000
22	887473957103951883	2017-07-19	00:47:34	+0000
23	887343217045368832	2017-07-18	16:08:03	+0000
24	887101392804085760	2017-07-18	00:07:08	+0000
25	886983233522544640	2017-07-17	16:17:36	+0000
26	886736880519319552	2017-07-16	23:58:41	+0000
28	886366144734445568	2017-07-15	23:25:31	+0000
29	886258384151887873	2017-07-15	16:17:19	+0000
30	885984800019947520	2017-07-14	22:10:11	+0000
31	885528943205470208	2017-07-13	15:58:47	+0000
33	885167619883638784	2017-07-12	16:03:00	+0000
...
2032	666437273139982337	2015-11-17	02:06:42	+0000
2033	666435652385423360	2015-11-17	02:00:15	+0000
2034	666430724426358785	2015-11-17	01:40:41	+0000
2035	666428276349472768	2015-11-17	01:30:57	+0000
2036	666421158376562688	2015-11-17	01:02:40	+0000
2037	666418789513326592	2015-11-17	00:53:15	+0000
2039	666407126856765440	2015-11-17	00:06:54	+0000
2040	666396247373291520	2015-11-16	23:23:41	+0000
2041	666373753744588802	2015-11-16	21:54:18	+0000
2043	666353288456101888	2015-11-16	20:32:58	+0000
2044	666345417576210432	2015-11-16	20:01:42	+0000
2045	666337882303524864	2015-11-16	19:31:45	+0000
2047	666287406224695296	2015-11-16	16:11:11	+0000
2048	666273097616637952	2015-11-16	15:14:19	+0000
2051	666102155909144576	2015-11-16	03:55:04	+0000
2052	666099513787052032	2015-11-16	03:44:34	+0000
2053	666094000022159362	2015-11-16	03:22:39	+0000
2054	666082916733198337	2015-11-16	02:38:37	+0000
2055	666073100786774016	2015-11-16	01:59:36	+0000
2056	666071193221509120	2015-11-16	01:52:02	+0000
2057	666063827256086533	2015-11-16	01:22:45	+0000
2058	666058600524156928	2015-11-16	01:01:59	+0000
2059	666057090499244032	2015-11-16	00:55:59	+0000
2060	666055525042405380	2015-11-16	00:49:46	+0000
2062	666050758794694657	2015-11-16	00:30:50	+0000
2063	666049248165822465	2015-11-16	00:24:50	+0000
2064	666044226329800704	2015-11-16	00:04:52	+0000

2065 666033412701032449 2015-11-15 23:21:54 +0000
2066 666029285002620928 2015-11-15 23:05:30 +0000
2067 666020888022790149 2015-11-15 22:32:08 +0000

```

source \
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...
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...
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...
31     <a href="http://twitter.com/download/iphone" r...
33     <a href="http://twitter.com/download/iphone" r...
...
2032  <a href="http://twitter.com/download/iphone" r...
2033  <a href="http://twitter.com/download/iphone" r...
2034  <a href="http://twitter.com/download/iphone" r...
2035  <a href="http://twitter.com/download/iphone" r...
2036  <a href="http://twitter.com/download/iphone" r...
2037  <a href="http://twitter.com/download/iphone" r...
2039  <a href="http://twitter.com/download/iphone" r...
2040  <a href="http://twitter.com/download/iphone" r...
2041  <a href="http://twitter.com/download/iphone" r...
2043  <a href="http://twitter.com/download/iphone" r...
2044  <a href="http://twitter.com/download/iphone" r...
2045  <a href="http://twitter.com/download/iphone" r...
```

2047 <a href="http://twitter.com/download/iphone" r...
 2048 <a href="http://twitter.com/download/iphone" r...
 2051 <a href="http://twitter.com/download/iphone" r...
 2052 <a href="http://twitter.com/download/iphone" r...
 2053 <a href="http://twitter.com/download/iphone" r...
 2054 <a href="http://twitter.com/download/iphone" r...
 2055 <a href="http://twitter.com/download/iphone" r...
 2056 <a href="http://twitter.com/download/iphone" r...
 2057 <a href="http://twitter.com/download/iphone" r...
 2058 <a href="http://twitter.com/download/iphone" r...
 2059 <a href="http://twitter.com/download/iphone" r...
 2060 <a href="http://twitter.com/download/iphone" r...
 2062 <a href="http://twitter.com/download/iphone" r...
 2063 <a href="http://twitter.com/download/iphone" r...
 2064 <a href="http://twitter.com/download/iphone" r...
 2065 <a href="http://twitter.com/download/iphone" r...
 2066 <a href="http://twitter.com/download/iphone" r...
 2067 <a href="http://twitter.com/download/iphone" r...

text \

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

31 This is Maisey. She fell asleep mid-excavation...
 33 Here we have a corgi undercover as a malamute...
 ...
 2032 Here we see a lone northeastern Cumberbatch. H...
 2033 "Can you behave? You're ruining my wedding day...
 2034 Oh boy what a pup! Sunglasses take this one to...
 2035 Here we have an Austrian Pulitzer. Collectors ...
 2036 *internally screaming* 12/10 <https://t.co/YMcr...>
 2037 This is Walter. He is an Alaskan Terrapin. Lov...
 2039 This is a southern Vesuvius bumblegruff. Can d...
 2040 Oh goodness. A super rare northeast Qdoba kang...
 2041 Those are sunglasses and a jean jacket. 11/10 ...
 2043 Here we have a mixed Asiago from the Galápagos...
 2044 Look at this jokester thinking seat belt laws ...
 2045 This is an extremely rare horned Parthenon. No...
 2047 This is an Albanian 3 1/2 legged Episcopalian...
 2048 Can take selfies 11/10 <https://t.co/ws2AMaWpW>
 2051 Oh my. Here you are seeing an Adobe Setter giv...
 2052 Can stand on stump for what seems like a while...
 2053 This appears to be a Mongolian Presbyterian mi...
 2054 Here we have a well-established sunblockerspan...
 2055 Let's hope this flight isn't Malaysian (lol). ...
 2056 Here we have a northern speckled Rhododendron...
 2057 This is the happiest dog you will ever see. Ve...
 2058 Here is the Rand Paul of retrievers folks! He'...
 2059 My oh my. This is a rare blond Canadian terrie...
 2060 Here is a Siberian heavily armored polar bear ...
 2062 This is a truly beautiful English Wilson Staff...
 2063 Here we have a 1949 1st generation vulpix. Enj...
 2064 This is a purebred Piers Morgan. Loves to Netf...
 2065 Here is a very happy pup. Big fan of well-main...
 2066 This is a western brown Mitsubishi terrier. Up...
 2067 Here we have a Japanese Irish Setter. Lost eye...

	expanded_urls	rating_numerator \
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/888078434...	12
20	https://twitter.com/dog_rates/status/887705289...	13
22	https://twitter.com/dog_rates/status/887473957...	13
23	https://twitter.com/dog_rates/status/887343217...	13
24	https://twitter.com/dog_rates/status/887101392...	12
25	https://twitter.com/dog_rates/status/886983233...	13
26	https://www.gofundme.com/mingusneedsus , https://...	13
28	https://twitter.com/dog_rates/status/886366144...	12
29	https://twitter.com/dog_rates/status/886258384...	13
30	https://twitter.com/dog_rates/status/885984800...	12
31	https://twitter.com/dog_rates/status/885528943...	13
33	https://twitter.com/dog_rates/status/885167619...	13
...
2032	https://twitter.com/dog_rates/status/666437273...	7
2033	https://twitter.com/dog_rates/status/666435652...	10
2034	https://twitter.com/dog_rates/status/666430724...	6
2035	https://twitter.com/dog_rates/status/666428276...	7
2036	https://twitter.com/dog_rates/status/666421158...	12
2037	https://twitter.com/dog_rates/status/666418789...	10
2039	https://twitter.com/dog_rates/status/666407126...	7
2040	https://twitter.com/dog_rates/status/666396247...	9
2041	https://twitter.com/dog_rates/status/666373753...	11
2043	https://twitter.com/dog_rates/status/666353288...	8
2044	https://twitter.com/dog_rates/status/666345417...	10
2045	https://twitter.com/dog_rates/status/666337882...	9
2047	https://twitter.com/dog_rates/status/666287406...	1
2048	https://twitter.com/dog_rates/status/666273097...	11
2051	https://twitter.com/dog_rates/status/666102155...	11
2052	https://twitter.com/dog_rates/status/666099513...	8
2053	https://twitter.com/dog_rates/status/666094000...	9
2054	https://twitter.com/dog_rates/status/666082916...	6
2055	https://twitter.com/dog_rates/status/666073100...	10
2056	https://twitter.com/dog_rates/status/666071193...	9
2057	https://twitter.com/dog_rates/status/666063827...	10
2058	https://twitter.com/dog_rates/status/666058600...	8
2059	https://twitter.com/dog_rates/status/666057090...	9
2060	https://twitter.com/dog_rates/status/666055525...	10
2062	https://twitter.com/dog_rates/status/666050758...	10
2063	https://twitter.com/dog_rates/status/666049248...	5
2064	https://twitter.com/dog_rates/status/666044226...	6
2065	https://twitter.com/dog_rates/status/666033412...	9
2066	https://twitter.com/dog_rates/status/666029285...	7
2067	https://twitter.com/dog_rates/status/666020888...	8

	rating_denominator	name	favorites	retweets	user_followers	\
1	10	Tilly	32772	6183	7427787	
2	10	Archie	24683	4091	7427787	
3	10	Darla	41553	8514	7427787	
4	10	Franklin	39739	9220	7427787	
5	10	None	19946	3068	7427787	
6	10	Jax	11673	2034	7427787	
7	10	None	64521	18586	7427787	
8	10	Zoey	27418	4207	7427787	
9	10	Cassie	31464	7275	7427787	
10	10	Koda	30238	7222	7427788	
11	10	Bruno	27395	4900	7427788	
12	10	None	47447	9906	7427788	
13	10	Ted	26790	4478	7427788	
14	10	Stuart	14891	2208	7427788	
15	10	Oliver	24902	5301	7427788	
16	10	Jim	28679	4427	7427788	
17	10	Zeke	25207	4234	7427788	
18	10	Ralphus	19560	3498	7427788	
19	10	Gerald	21456	3445	7427788	
20	10	Jeffrey	29752	5309	7427788	
22	10	Canela	68123	17915	7427787	
23	10	None	33210	10249	7427787	
24	10	None	30140	5876	7427787	
25	10	Maya	34676	7658	7427787	
26	10	Mingus	11877	3232	7427787	
28	10	Roscoe	20891	3152	7427787	
29	10	Waffles	27599	6190	7427787	
30	10	Jimbo	32213	6698	7427787	
31	10	Maisey	35481	6334	7427787	
33	10	None	21622	4320	7427787	
...	
2032	10	None	124	48	7427810	
2033	10	None	165	53	7427810	
2034	10	None	316	196	7427810	
2035	10	None	165	83	7427810	
2036	10	None	315	111	7427810	
2037	10	Walter	123	45	7427810	
2039	10	a	106	40	7428467	
2040	10	None	167	84	7427810	
2041	10	None	186	92	7427810	
2043	10	None	218	71	7427810	
2044	10	None	290	133	7427810	
2045	10	an	197	90	7427810	
2047	2	an	146	64	7427810	
2048	10	None	173	76	7427810	
2051	10	None	80	12	7427810	

2052	10	None	153	67	7427810
2053	10	None	163	72	7427810
2054	10	None	116	44	7427810
2055	10	None	318	162	7427810
2056	10	None	144	59	7427810
2057	10	the	471	215	7427810
2058	10	the	111	57	7427810
2059	10	a	293	141	7427810
2060	10	a	431	245	7427810
2062	10	a	131	59	7427810
2063	10	None	107	40	7427810
2064	10	a	295	139	7427810
2065	10	a	125	44	7427810
2066	10	a	129	47	7427810
2067	10	None	2543	508	7427810

	user_favourites	jpg_url \
1	139530	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2	139530	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
3	139530	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
4	139530	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
5	139530	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
6	139530	https://pbs.twimg.com/media/DF1e0mZXUAAALUc.jpg
7	139530	https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
8	139530	https://pbs.twimg.com/media/DFwUU_XcAEpyXI.jpg
9	139530	https://pbs.twimg.com/media/DFrEyVuW0AAO3t9.jpg
10	139530	https://pbs.twimg.com/media/DFnwsY4WAAAMliS.jpg
11	139530	https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg
12	139530	https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
13	139530	https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
14	139530	https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg
15	139530	https://pbs.twimg.com/ext_tw_video_thumb/88927...
16	139530	https://pbs.twimg.com/media/DFYRgsOUQAARgh0.jpg
17	139530	https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
18	139530	https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg
19	139530	https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
20	139530	https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
22	139530	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
23	139530	https://pbs.twimg.com/ext_tw_video_thumb/88734...
24	139530	https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
25	139530	https://pbs.twimg.com/media/DE8yicJW0AAAavBJ.jpg
26	139530	https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
28	139530	https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
29	139530	https://pbs.twimg.com/media/DEyfTG4UMAE4aE9.jpg
30	139530	https://pbs.twimg.com/media/DEumeWWVOAA-Z61.jpg
31	139530	https://pbs.twimg.com/media/DEoH3yvXgAAzQtS.jpg
33	139530	https://pbs.twimg.com/media/DEi_N9qXYAAgEEw.jpg
...

2032	139530	https://pbs.twimg.com/media/CT-pKmRWIAAxUWj.jpg
2033	139530	https://pbs.twimg.com/media/CT-nsTQWEAEkyDn.jpg
2034	139530	https://pbs.twimg.com/media/CT-jNYqW4AAPi2M.jpg
2035	139530	https://pbs.twimg.com/media/CT-g-ODUwAEQdSn.jpg
2036	139530	https://pbs.twimg.com/media/CT-aggCXAAIMfT3.jpg
2037	139530	https://pbs.twimg.com/media/CT-YWb7U8AA7QnN.jpg
2039	139531	https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
2040	139530	https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
2041	139530	https://pbs.twimg.com/media/CT9vZEYUAA1Z05.jpg
2043	139530	https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
2044	139530	https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg
2045	139530	https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg
2047	139530	https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
2048	139530	https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
2051	139530	https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg
2052	139530	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
2053	139530	https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg
2054	139530	https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg
2055	139530	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
2056	139530	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
2057	139530	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
2058	139530	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
2059	139530	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
2060	139530	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
2062	139530	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
2063	139530	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
2064	139530	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
2065	139530	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
2066	139530	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2067	139530	https://pbs.twimg.com/media/CT4udn0WwAAOaMy.jpg

	img_num	dog_stages	breed	confidence
1	1	None	Chihuahua	0.323581
2	1	None	Chihuahua	0.716012
3	1	None	Labrador_retriever	0.168086
4	2	None	basset	0.555712
5	1	None	Chesapeake_Bay_retriever	0.425595
6	1	None	Appenzeller	0.341703
7	2	None	Pomeranian	0.566142
8	1	None	Irish_terrier	0.487574
9	1	doggo	Pembroke	0.511319
10	1	None	Samoyed	0.957979
11	1	None	French_bulldog	0.377417
12	1	puppo	Pembroke	0.966327
13	1	None	French_bulldog	0.991650
14	1	puppo	golden_retriever	0.953442
15	1	None	whippet	0.626152
16	1	None	golden_retriever	0.714719

17	1	None	golden_retriever	0.469760
18	3	None	Siberian_husky	0.700377
19	1	None	French_bulldog	0.995026
20	1	None	basset	0.821664
22	2	None	Pembroke	0.809197
23	1	None	Mexican_hairless	0.330741
24	1	floof	Samoyed	0.733942
25	2	None	Chihuahua	0.793469
26	1	None	kuvasz	0.309706
28	1	pupper	French_bulldog	0.999201
29	1	None	pug	0.943575
30	1	None	Blenheim_spaniel	0.972494
31	1	None	pug	0.369275
33	4	None	malamute	0.812482
...
2032	1	None	Chihuahua	0.671853
2033	1	None	Chesapeake_Bay_retriever	0.184130
2034	1	None	Irish_terrier	0.104109
2035	1	None	Pembroke	0.371361
2036	1	None	Blenheim_spaniel	0.906777
2037	1	None	toy_terrier	0.149680
2039	1	None	black-and-tan_coonhound	0.529139
2040	1	None	Chihuahua	0.978108
2041	1	None	soft-coated_wheaten_terrier	0.326467
2043	1	None	malamute	0.336874
2044	1	None	golden_retriever	0.858744
2045	1	None	Newfoundland	0.278407
2047	1	None	Maltese_dog	0.857531
2048	1	None	Italian_greyhound	0.176053
2051	1	None	English_setter	0.298617
2052	1	None	Lhasa	0.582330
2053	1	None	bloodhound	0.195217
2054	1	None	pug	0.489814
2055	1	None	Walker_hound	0.260857
2056	1	None	Gordon_setter	0.503672
2057	1	None	golden_retriever	0.775930
2058	1	None	miniature_poodle	0.201493
2059	1	None	golden_retriever	0.007959
2060	1	None	chow	0.692517
2062	1	None	Bernese_mountain_dog	0.651137
2063	1	None	miniature_pinscher	0.560311
2064	1	None	Rhodesian_ridgeback	0.408143
2065	1	None	German_shepherd	0.596461
2066	1	None	redbone	0.506826
2067	1	None	Welsh_springer_spaniel	0.465074

[1685 rows x 17 columns]

Task5: Converting variables to correct datatype

Define For proper analysis it is very important to set correct datatypes for all the variables included. Here I will set the correct datatypes.

Code

```
In [35]: df_master['tweet_id'] = df_master['tweet_id'].astype(str)
df_master['timestamp'] = pd.to_datetime(df_master.timestamp)
df_master['source'] = df_master['source'].astype('category')
df_master['img_num'] = df_master['img_num'].astype('category')
df_master['dog_stages'] = df_master['dog_stages'].astype('category')
df_master['rating_numerator'] = df_master['rating_numerator'].astype('float64')
df_master['rating_denominator'] = df_master['rating_denominator'].astype('float64')
```

Test

```
In [36]: df_master
```

```
Out[36]:
```

	tweet_id	timestamp \
1	892177421306343426	2017-08-01 00:17:27
2	891815181378084864	2017-07-31 00:18:03
3	891689557279858688	2017-07-30 15:58:51
4	891327558926688256	2017-07-29 16:00:24
5	891087950875897856	2017-07-29 00:08:17
6	890971913173991426	2017-07-28 16:27:12
7	890729181411237888	2017-07-28 00:22:40
8	890609185150312448	2017-07-27 16:25:51
9	890240255349198849	2017-07-26 15:59:51
10	890006608113172480	2017-07-26 00:31:25
11	889880896479866881	2017-07-25 16:11:53
12	889665388333682689	2017-07-25 01:55:32
13	889638837579907072	2017-07-25 00:10:02
14	889531135344209921	2017-07-24 17:02:04
15	889278841981685760	2017-07-24 00:19:32
16	888917238123831296	2017-07-23 00:22:39
17	888804989199671297	2017-07-22 16:56:37
18	888554962724278272	2017-07-22 00:23:06
19	888078434458587136	2017-07-20 16:49:33
20	887705289381826560	2017-07-19 16:06:48
22	887473957103951883	2017-07-19 00:47:34
23	887343217045368832	2017-07-18 16:08:03
24	887101392804085760	2017-07-18 00:07:08
25	886983233522544640	2017-07-17 16:17:36
26	886736880519319552	2017-07-16 23:58:41
28	886366144734445568	2017-07-15 23:25:31
29	886258384151887873	2017-07-15 16:17:19
30	885984800019947520	2017-07-14 22:10:11

```

31      885528943205470208 2017-07-13 15:58:47
33      885167619883638784 2017-07-12 16:03:00
...
2032    666437273139982337 2015-11-17 02:06:42
2033    666435652385423360 2015-11-17 02:00:15
2034    666430724426358785 2015-11-17 01:40:41
2035    666428276349472768 2015-11-17 01:30:57
2036    666421158376562688 2015-11-17 01:02:40
2037    666418789513326592 2015-11-17 00:53:15
2039    666407126856765440 2015-11-17 00:06:54
2040    666396247373291520 2015-11-16 23:23:41
2041    666373753744588802 2015-11-16 21:54:18
2043    666353288456101888 2015-11-16 20:32:58
2044    666345417576210432 2015-11-16 20:01:42
2045    666337882303524864 2015-11-16 19:31:45
2047    666287406224695296 2015-11-16 16:11:11
2048    666273097616637952 2015-11-16 15:14:19
2051    666102155909144576 2015-11-16 03:55:04
2052    666099513787052032 2015-11-16 03:44:34
2053    666094000022159362 2015-11-16 03:22:39
2054    666082916733198337 2015-11-16 02:38:37
2055    666073100786774016 2015-11-16 01:59:36
2056    666071193221509120 2015-11-16 01:52:02
2057    666063827256086533 2015-11-16 01:22:45
2058    666058600524156928 2015-11-16 01:01:59
2059    666057090499244032 2015-11-16 00:55:59
2060    666055525042405380 2015-11-16 00:49:46
2062    666050758794694657 2015-11-16 00:30:50
2063    666049248165822465 2015-11-16 00:24:50
2064    666044226329800704 2015-11-16 00:04:52
2065    666033412701032449 2015-11-15 23:21:54
2066    666029285002620928 2015-11-15 23:05:30
2067    666020888022790149 2015-11-15 22:32:08

```

```

source \
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...

```

[illegible]

text \

1 This is Tilly. She's just checking pup on you...
 2 This is Archie. He is a rare Norwegian Pouncin...
 3 This is Darla. She commenced a snooze mid meal...
 4 This is Franklin. He would like you to stop ca...
 5 Here we have a majestic great white breaching ...
 6 Meet Jax. He enjoys ice cream so much he gets ...
 7 When you watch your owner call another dog a g...
 8 This is Zoey. She doesn't want to be one of th...
 9 This is Cassie. She is a college pup. Studying...
 10 This is Koda. He is a South Australian decksha...
 11 This is Bruno. He is a service shark. Only get...
 12 Here's a puppo that seems to be on the fence a...
 13 This is Ted. He does his best. Sometimes that'...
 14 This is Stuart. He's sporting his favorite fan...
 15 This is Oliver. You're witnessing one of his m...
 16 This is Jim. He found a fren. Taught him how t...
 17 This is Zeke. He has a new stick. Very proud o...
 18 This is Ralphus. He's powering up. Attempting ...
 19 This is Gerald. He was just told he didn't get...
 20 This is Jeffrey. He has a monopoly on the pool...
 22 This is Canela. She attempted some fancy porch...
 23 You may not have known you needed to see this ...
 24 This... is a Jubilant Antarctic House Bear. We...
 25 This is Maya. She's very shy. Rarely leaves he...
 26 This is Mingus. He's a wonderful father to his...
 28 This is Roscoe. Another pupper fallen victim t...
 29 This is Waffles. His doggles are pupside down...
 30 Viewer discretion advised. This is Jimbo. He w...
 31 This is Maisey. She fell asleep mid-excavation...
 33 Here we have a corgi undercover as a malamute...
 ...
 2032 Here we see a lone northeastern Cumberbatch. H...
 2033 "Can you behave? You're ruining my wedding day...
 2034 Oh boy what a pup! Sunglasses take this one to...
 2035 Here we have an Austrian Pulitzer. Collectors ...
 2036 *internally screaming* 12/10 <https://t.co/YMcr...>
 2037 This is Walter. He is an Alaskan Terrapin. Lov...
 2039 This is a southern Vesuvius bumblegruff. Can d...
 2040 Oh goodness. A super rare northeast Qdoba kang...
 2041 Those are sunglasses and a jean jacket. 11/10 ...
 2043 Here we have a mixed Asiago from the Galápagos...
 2044 Look at this jokester thinking seat belt laws ...
 2045 This is an extremely rare horned Parthenon. No...
 2047 This is an Albanian 3 1/2 legged Episcopalian...
 2048 Can take selfies 11/10 <https://t.co/ws2AMaNPW>
 2051 Oh my. Here you are seeing an Adobe Setter giv...

2052 Can stand on stump for what seems like a while...
 2053 This appears to be a Mongolian Presbyterian mi...
 2054 Here we have a well-established sunblockerspan...
 2055 Let's hope this flight isn't Malaysian (lol). ...
 2056 Here we have a northern speckled Rhododendron...
 2057 This is the happiest dog you will ever see. Ve...
 2058 Here is the Rand Paul of retrievers folks! He'...
 2059 My oh my. This is a rare blond Canadian terrie...
 2060 Here is a Siberian heavily armored polar bear ...
 2062 This is a truly beautiful English Wilson Staff...
 2063 Here we have a 1949 1st generation vulpix. Enj...
 2064 This is a purebred Piers Morgan. Loves to Netf...
 2065 Here is a very happy pup. Big fan of well-main...
 2066 This is a western brown Mitsubishi terrier. Up...
 2067 Here we have a Japanese Irish Setter. Lost eye...

	expanded_urls	rating_numerator \
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
19	https://twitter.com/dog_rates/status/888078434...	12.0
20	https://twitter.com/dog_rates/status/887705289...	13.0
22	https://twitter.com/dog_rates/status/887473957...	13.0
23	https://twitter.com/dog_rates/status/887343217...	13.0
24	https://twitter.com/dog_rates/status/887101392...	12.0
25	https://twitter.com/dog_rates/status/886983233...	13.0
26	https://www.gofundme.com/mingusneedsus,https:/...	13.0
28	https://twitter.com/dog_rates/status/886366144...	12.0
29	https://twitter.com/dog_rates/status/886258384...	13.0
30	https://twitter.com/dog_rates/status/885984800...	12.0
31	https://twitter.com/dog_rates/status/885528943...	13.0
33	https://twitter.com/dog_rates/status/885167619...	13.0
...

2032	https://twitter.com/dog_rates/status/666437273...	7.0
2033	https://twitter.com/dog_rates/status/666435652...	10.0
2034	https://twitter.com/dog_rates/status/666430724...	6.0
2035	https://twitter.com/dog_rates/status/666428276...	7.0
2036	https://twitter.com/dog_rates/status/666421158...	12.0
2037	https://twitter.com/dog_rates/status/666418789...	10.0
2039	https://twitter.com/dog_rates/status/666407126...	7.0
2040	https://twitter.com/dog_rates/status/666396247...	9.0
2041	https://twitter.com/dog_rates/status/666373753...	11.0
2043	https://twitter.com/dog_rates/status/666353288...	8.0
2044	https://twitter.com/dog_rates/status/666345417...	10.0
2045	https://twitter.com/dog_rates/status/666337882...	9.0
2047	https://twitter.com/dog_rates/status/666287406...	1.0
2048	https://twitter.com/dog_rates/status/666273097...	11.0
2051	https://twitter.com/dog_rates/status/666102155...	11.0
2052	https://twitter.com/dog_rates/status/666099513...	8.0
2053	https://twitter.com/dog_rates/status/666094000...	9.0
2054	https://twitter.com/dog_rates/status/666082916...	6.0
2055	https://twitter.com/dog_rates/status/666073100...	10.0
2056	https://twitter.com/dog_rates/status/666071193...	9.0
2057	https://twitter.com/dog_rates/status/666063827...	10.0
2058	https://twitter.com/dog_rates/status/666058600...	8.0
2059	https://twitter.com/dog_rates/status/666057090...	9.0
2060	https://twitter.com/dog_rates/status/666055525...	10.0
2062	https://twitter.com/dog_rates/status/666050758...	10.0
2063	https://twitter.com/dog_rates/status/666049248...	5.0
2064	https://twitter.com/dog_rates/status/666044226...	6.0
2065	https://twitter.com/dog_rates/status/666033412...	9.0
2066	https://twitter.com/dog_rates/status/666029285...	7.0
2067	https://twitter.com/dog_rates/status/666020888...	8.0

	rating_denominator	name	favorites	retweets	user_followers \
1	10.0	Tilly	32767	6182	7431629
2	10.0	Archie	24674	4091	7431629
3	10.0	Darla	41551	8512	7431630
4	10.0	Franklin	39737	9218	7431630
5	10.0	None	19946	3065	7431630
6	10.0	Jax	11670	2034	7431630
7	10.0	None	64506	18583	7431630
8	10.0	Zoey	27415	4207	7431630
9	10.0	Cassie	31455	7275	7431630
10	10.0	Koda	30242	7221	7431630
11	10.0	Bruno	27397	4900	7431630
12	10.0	None	47439	9904	7431630
13	10.0	Ted	26784	4477	7431630
14	10.0	Stuart	14888	2208	7431630
15	10.0	Oliver	24898	5300	7431630
16	10.0	Jim	28676	4427	7431630

17	10.0	Zeke	25204	4234	7431630
18	10.0	Ralphus	19556	3498	7431630
19	10.0	Gerald	21450	3445	7431630
20	10.0	Jeffrey	29745	5308	7431630
22	10.0	Canela	68114	17913	7431630
23	10.0	None	33208	10247	7431630
24	10.0	None	30132	5875	7431630
25	10.0	Maya	34671	7657	7431630
26	10.0	Mingus	11879	3232	7431630
28	10.0	Roscoe	20885	3152	7431630
29	10.0	Waffles	27600	6189	7431630
30	10.0	Jimbo	32212	6696	7431630
31	10.0	Maisey	35476	6333	7431630
33	10.0	None	21619	4320	7431630
...
2032	10.0	None	124	48	7431653
2033	10.0	None	165	53	7431653
2034	10.0	None	316	196	7431653
2035	10.0	None	165	83	7431653
2036	10.0	None	315	111	7431653
2037	10.0	Walter	123	45	7431653
2039	10.0	a	106	40	7431653
2040	10.0	None	167	84	7431653
2041	10.0	None	186	92	7431653
2043	10.0	None	218	71	7431653
2044	10.0	None	290	133	7431653
2045	10.0	an	197	90	7431653
2047	2.0	an	146	64	7431653
2048	10.0	None	173	76	7431653
2051	10.0	None	80	12	7431653
2052	10.0	None	153	67	7431653
2053	10.0	None	163	72	7431653
2054	10.0	None	116	44	7431653
2055	10.0	None	318	162	7431653
2056	10.0	None	144	59	7431653
2057	10.0	the	471	215	7431653
2058	10.0	the	111	57	7431653
2059	10.0	a	293	141	7431653
2060	10.0	a	431	244	7431653
2062	10.0	a	131	59	7431653
2063	10.0	None	107	40	7431653
2064	10.0	a	295	139	7431653
2065	10.0	a	125	44	7431653
2066	10.0	a	129	47	7431653
2067	10.0	None	2543	508	7431653

	user_favourites	jpg_url \
1	139602	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg

2	139602	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
3	139602	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
4	139602	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
5	139602	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
6	139602	https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg
7	139602	https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
8	139602	https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
9	139602	https://pbs.twimg.com/media/DFrEyVuW0AAO3t9.jpg
10	139602	https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
11	139602	https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg
12	139602	https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
13	139602	https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
14	139602	https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg
15	139602	https://pbs.twimg.com/ext_tw_video_thumb/88927...
16	139602	https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg
17	139602	https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
18	139602	https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg
19	139602	https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
20	139602	https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
22	139602	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
23	139602	https://pbs.twimg.com/ext_tw_video_thumb/88734...
24	139602	https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
25	139602	https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.jpg
26	139602	https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
28	139602	https://pbs.twimg.com/media/DE0BTnQUwAAPKEH.jpg
29	139602	https://pbs.twimg.com/media/DEyfTG4UMAE4aE9.jpg
30	139602	https://pbs.twimg.com/media/DEumeWWV0AA-Z61.jpg
31	139602	https://pbs.twimg.com/media/DEoH3yvXgAAzQtS.jpg
33	139602	https://pbs.twimg.com/media/DEi_N9qXYAAgEEw.jpg
...
2032	139602	https://pbs.twimg.com/media/CT-pKmRWIAAxUWj.jpg
2033	139602	https://pbs.twimg.com/media/CT-nsTQWEAEkyDn.jpg
2034	139602	https://pbs.twimg.com/media/CT-jNYqW4AAPi2M.jpg
2035	139602	https://pbs.twimg.com/media/CT-g-ODUwAEQdSn.jpg
2036	139602	https://pbs.twimg.com/media/CT-aggCXAAlMfT3.jpg
2037	139602	https://pbs.twimg.com/media/CT-YWb7U8AA7QnN.jpg
2039	139602	https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
2040	139602	https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
2041	139602	https://pbs.twimg.com/media/CT9vZEYWUAA1ZO5.jpg
2043	139602	https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
2044	139602	https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg
2045	139602	https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg
2047	139602	https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
2048	139602	https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
2051	139602	https://pbs.twimg.com/media/CT54YGiwUAEEZnoK.jpg
2052	139602	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
2053	139602	https://pbs.twimg.com/media/CT5w9gUW4AAAsBNN.jpg
2054	139602	https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg

2055	139602	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
2056	139602	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
2057	139602	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
2058	139602	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
2059	139602	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
2060	139602	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
2062	139602	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
2063	139602	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
2064	139602	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
2065	139602	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
2066	139602	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2067	139602	https://pbs.twimg.com/media/CT4udn0WwAAOaMy.jpg

	img_num	dog_stages	breed	confidence
1	1	None	Chihuahua	0.323581
2	1	None	Chihuahua	0.716012
3	1	None	Labrador_retriever	0.168086
4	2	None	basset	0.555712
5	1	None	Chesapeake_Bay_retriever	0.425595
6	1	None	Appenzeller	0.341703
7	2	None	Pomeranian	0.566142
8	1	None	Irish_terrier	0.487574
9	1	doggo	Pembroke	0.511319
10	1	None	Samoyed	0.957979
11	1	None	French_bulldog	0.377417
12	1	puppo	Pembroke	0.966327
13	1	None	French_bulldog	0.991650
14	1	puppo	golden_retriever	0.953442
15	1	None	whippet	0.626152
16	1	None	golden_retriever	0.714719
17	1	None	golden_retriever	0.469760
18	3	None	Siberian_husky	0.700377
19	1	None	French_bulldog	0.995026
20	1	None	basset	0.821664
22	2	None	Pembroke	0.809197
23	1	None	Mexican_hairless	0.330741
24	1	floof	Samoyed	0.733942
25	2	None	Chihuahua	0.793469
26	1	None	kuvasz	0.309706
28	1	pupper	French_bulldog	0.999201
29	1	None	pug	0.943575
30	1	None	Blenheim_spaniel	0.972494
31	1	None	pug	0.369275
33	4	None	malamute	0.812482
...
2032	1	None	Chihuahua	0.671853
2033	1	None	Chesapeake_Bay_retriever	0.184130
2034	1	None	Irish_terrier	0.104109

2035	1	None	Pembroke	0.371361
2036	1	None	Blenheim_spaniel	0.906777
2037	1	None	toy_terrier	0.149680
2039	1	None	black-and-tan_coonhound	0.529139
2040	1	None	Chihuahua	0.978108
2041	1	None	soft-coated_wheaten_terrier	0.326467
2043	1	None	malamute	0.336874
2044	1	None	golden_retriever	0.858744
2045	1	None	Newfoundland	0.278407
2047	1	None	Maltese_dog	0.857531
2048	1	None	Italian_greyhound	0.176053
2051	1	None	English_setter	0.298617
2052	1	None	Lhasa	0.582330
2053	1	None	bloodhound	0.195217
2054	1	None	pug	0.489814
2055	1	None	Walker_hound	0.260857
2056	1	None	Gordon_setter	0.503672
2057	1	None	golden_retriever	0.775930
2058	1	None	miniature_poodle	0.201493
2059	1	None	golden_retriever	0.007959
2060	1	None	chow	0.692517
2062	1	None	Bernese_mountain_dog	0.651137
2063	1	None	miniature_pinscher	0.560311
2064	1	None	Rhodesian_ridgeback	0.408143
2065	1	None	German_shepherd	0.596461
2066	1	None	redbone	0.506826
2067	1	None	Welsh_springer_spaniel	0.465074

[1685 rows x 17 columns]

In [37]: df_master.describe()

```
Out[37]:
```

	rating_numerator	rating_denominator	favorites	retweets \
count	1685.000000	1685.000000	1685.000000	1685.000000
mean	11.473591	10.534718	9110.648071	2720.528783
std	8.109189	6.947941	13326.513716	4919.970189
min	0.000000	2.000000	80.000000	12.000000
25%	10.000000	10.000000	2073.000000	610.000000
50%	11.000000	10.000000	4272.000000	1359.000000
75%	12.000000	10.000000	11246.000000	3080.000000
max	165.000000	150.000000	164966.000000	84148.000000

	user_followers	user_favourites	confidence
count	1.685000e+03	1685.000000	1685.000000
mean	7.431644e+06	139602.005341	0.549905
std	4.897948e+01	0.072910	0.299357
min	7.431629e+06	139602.000000	0.000010
25%	7.431633e+06	139602.000000	0.302685

50%	7.431640e+06	139602.000000	0.550506
75%	7.431644e+06	139602.000000	0.822048
max	7.432312e+06	139603.000000	0.999956

Task 6 : Fixing the dog names

Define I noticed while accessing the data that there are some invalid dog names in 'names' column. Let's check dog incorrect dog names usually starting from lower case letters

Code

```
In [38]: # Using FOR loop to get all the names starting from lowercase
lowercase_names = []
for row in df_master['name']:
    if row[0].islower() and row not in lowercase_names:
        lowercase_names.append(row)
print(lowercase_names)
```

```
['a', 'quite', 'one', 'incredibly', 'my', 'not', 'an', 'very', 'just', 'getting', 'unacceptable']
```

Code

```
In [39]: # Replace all names that start with a lowercase letter with a NaN
df_master['name'].replace(lowercase_names,
                          np.nan,
                          inplace = True)

# Replace all 'None's with a NaN
df_master['name'].replace('None',
                          np.nan,
                          inplace = True)

# Replace the name '0' with "0'Malley"
df_master['name'].replace('0',
                          "0'Malley",
                          inplace = True)
```

Test

```
In [40]: df_master['name'].value_counts()
```

```
Out[40]: Lucy          10
         Cooper        10
         Tucker         9
         Charlie        9
         Oliver         9
         Penny          8
```

Daisy	7
Winston	7
Sadie	7
Koda	6
Lola	6
Jax	6
Toby	6
Bella	5
Bo	5
Rusty	5
Stanley	5
Leo	5
Oscar	5
Louis	4
Brody	4
Finn	4
Dave	4
Scout	4
Cassie	4
Sophie	4
Bentley	4
Oakley	4
Duke	4
Milo	4
..	
Finnegus	1
Schnozz	1
Cilantro	1
Kramer	1
Wishes	1
Berkeley	1
Brutus	1
Biden	1
Aldrick	1
Leonard	1
Fwed	1
Arnold	1
Willem	1
Eleanor	1
Fizz	1
Brad	1
Flash	1
Naphaniel	1
Remy	1
Tonks	1
Eve	1
Jangle	1
Burt	1


```

Jimison      1
Bronte       1
Jennifur     1
Mollie       1
Maxaroni     1
Hector       1
Layla        1
Name: name, Length: 833, dtype: int64

```

Task 7 : To create a new column 'rating'

Define Creating a single column named 'rating' based on Numerator and Denominator values

Code

```
In [41]: rating_denominator_idx = df_master.columns.get_loc('rating_denominator')
```

```
column_name = 'rating'
```

```
rating_idx = rating_denominator_idx + 1
```

```
df_master = df_master.query('rating_denominator == 10')
```

```
df_master.insert(rating_idx, column_name, round(np.divide(df_master['rating_numerator']
df_master['rating_denominator']),2))
```

```
# Test
```

```
df_master.head()
```

```

Out[41]:
      tweet_id      timestamp \
1  892177421306343426  2017-08-01 00:17:27
2  891815181378084864  2017-07-31 00:18:03
3  891689557279858688  2017-07-30 15:58:51
4  891327558926688256  2017-07-29 16:00:24
5  891087950875897856  2017-07-29 00:08:17

      source \
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...

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

```

5 Here we have a majestic great white breaching ...

		expanded_urls	rating_numerator	\
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	

	rating_denominator	rating	name	favorites	retweets	user_followers	\
1	10.0	1.3	Tilly	32767	6182	7431629	
2	10.0	1.2	Archie	24674	4091	7431629	
3	10.0	1.3	Darla	41551	8512	7431630	
4	10.0	1.2	Franklin	39737	9218	7431630	
5	10.0	1.3	NaN	19946	3065	7431630	

	user_favourites	jpg_url	img_num	\
1	139602	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg	1	
2	139602	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg	1	
3	139602	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg	1	
4	139602	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg	2	
5	139602	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg	1	

	dog_stages	breed	confidence
1	None	Chihuahua	0.323581
2	None	Chihuahua	0.716012
3	None	Labrador_retriever	0.168086
4	None	basset	0.555712
5	None	Chesapeake_Bay_retriever	0.425595

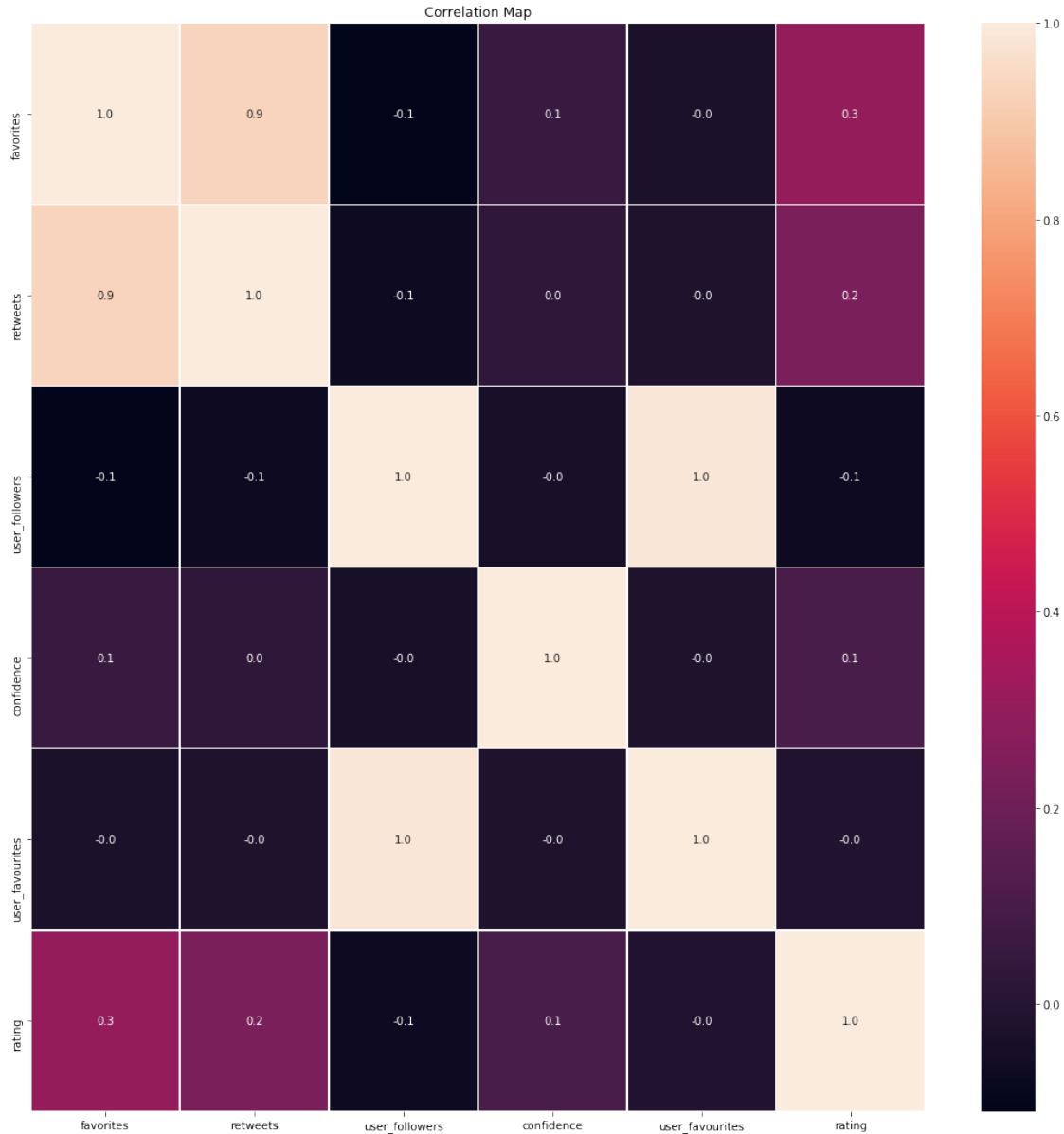
0.2.4 Analyze and Visualize

I will create correlation matrix for having quick idea for important variable correlations

1. What is the correlation of various variable in the given data

Code

```
In [42]: f,ax = plt.subplots(figsize=(18, 18))
sns.heatmap(df_master[['source', 'favorites', 'retweets',
                        'user_followers', 'confidence',
                        'user_favourites', 'rating']].corr(), annot=True, linewidths=.5, f
plt.title('Correlation Map')
plt.savefig('Correlation.png')
```



- There is strong correlation between retweets and favourites. The dogs having more retweets, are more favourites which is as expected.
- We can see positive correlation of dog rating with favourites, retweets & confidence.
- User followers and favourites doesn't have much effect on each other

2. How retweets and favourites are correlated for Popular breeds

```
In [43]: #Grouping the top 8 breeds by their favourites score and assigning to dog_favorites var
dog_favorites = df_master.groupby('breed')['favorites'].sum().sort_values(ascending=False)
dog_favorites
```

```
Out[43]: breed
golden_retriever      1858893
Labrador_retriever    1255254
Pembroke               1021658
Chihuahua              719039
French_bulldog         591775
Samoyed                562894
chow                   438498
cocker_spaniel         400368
Name: favorites, dtype: int64
```

```
In [44]: #Grouping the top 8 breeds by their retweets score and assigning to variable dog_retweets
dog_retweets = df_master.groupby('breed')['retweets'].sum().sort_values(ascending=False)
dog_retweets
```

```
Out[44]: breed
golden_retriever      547477
Labrador_retriever    389524
Pembroke              279688
Chihuahua             232193
Samoyed               192295
French_bulldog        155960
cocker_spaniel        140822
chow                  126062
Name: retweets, dtype: int64
```

```
In [45]: fig = plt.figure() # Creates matplotlib figure

ax = fig.add_subplot(111) # Creates matplotlib axes
ax2 = ax.twinx() # Creates another axes that shares the same x-axis as ax.

width = 0.4

dog_favorites.plot(figsize = (10,6), kind='bar', color='#1C86EE', ax=ax, width=width, p
dog_retweets.plot(figsize = (10,6), kind='bar', color='#008B8B', ax=ax2, width=width, p

# Turns off grid on the left Axis.
ax.grid(False)

# Turns off grid on the secondary (right) Axis.
ax2.grid(False)

ax.set_ylabel('Likes')
ax2.set_ylabel('Retweets')

ax.set_xticklabels(dog_favorites.index, rotation=30)

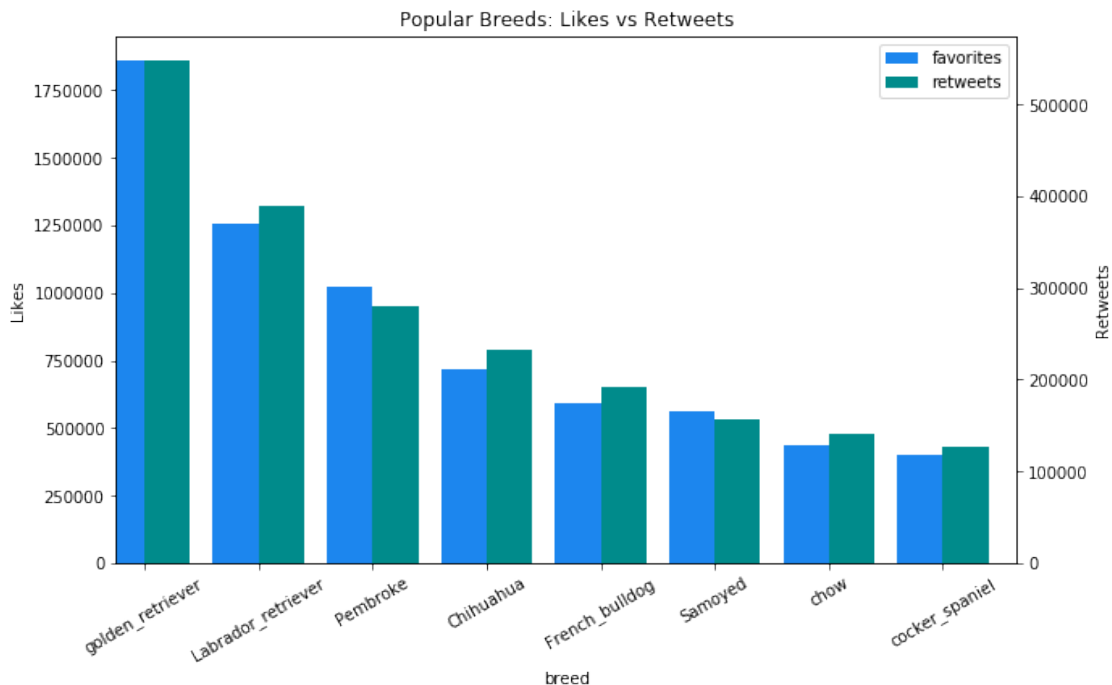
h1, l1 = ax.get_legend_handles_labels()
```

```

h2, l2 = ax2.get_legend_handles_labels()

plt.legend(h1+h2, l1+l2, loc=1)
plt.show()
plt.savefig('Popular_Breeds.png')

```



<matplotlib.figure.Figure at 0x7ff2e577d6d8>

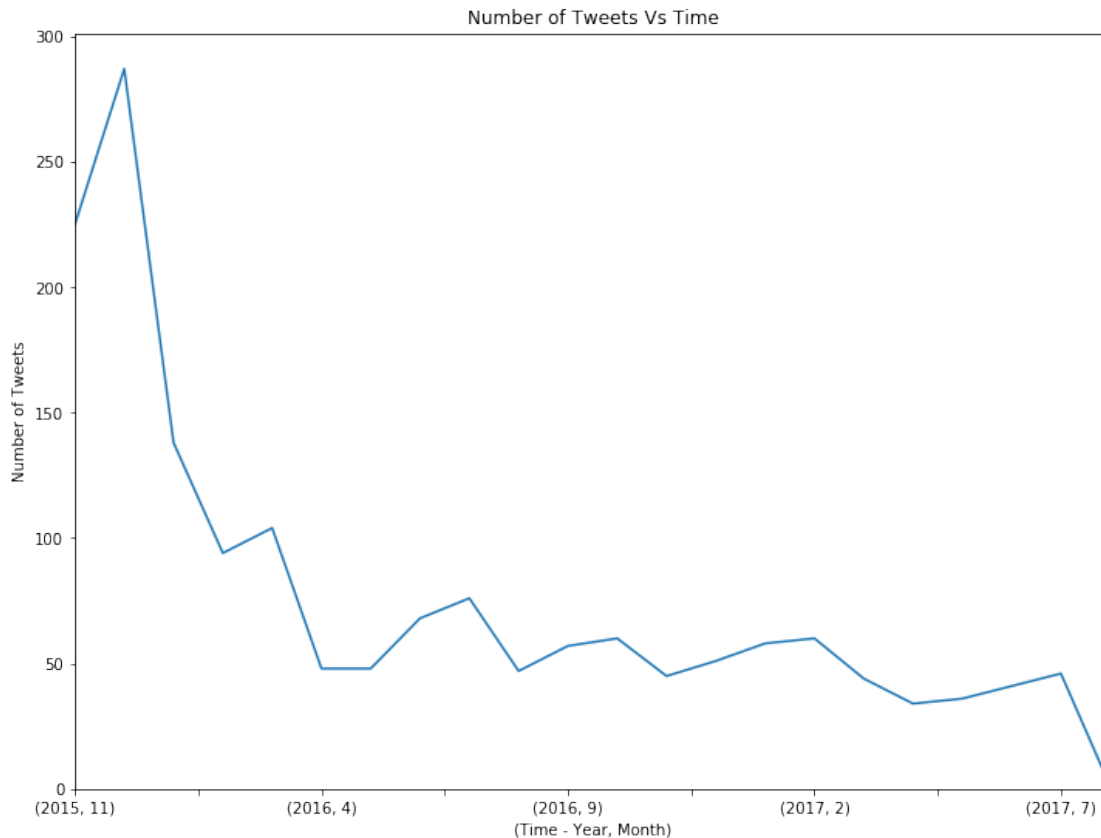
We can see from above plot that breed golden retriever is most popular with highest likes and retweets.

3. How all the tweets are distributed in dataframe over the period of time

```

In [46]: plt.rcParams["figure.figsize"] = [12, 9]
data = df_master.tweet_id.groupby([df_master["timestamp"].dt.year, df_master["timestamp"]
ax = data.plot('line', title='Number of Tweets Vs Time')
ax.set_xlabel("(Time - Year, Month)")
ax.set_ylabel("Number of Tweets")
ax.set_ylim(0, )
plt.savefig('tweet_trends')

```



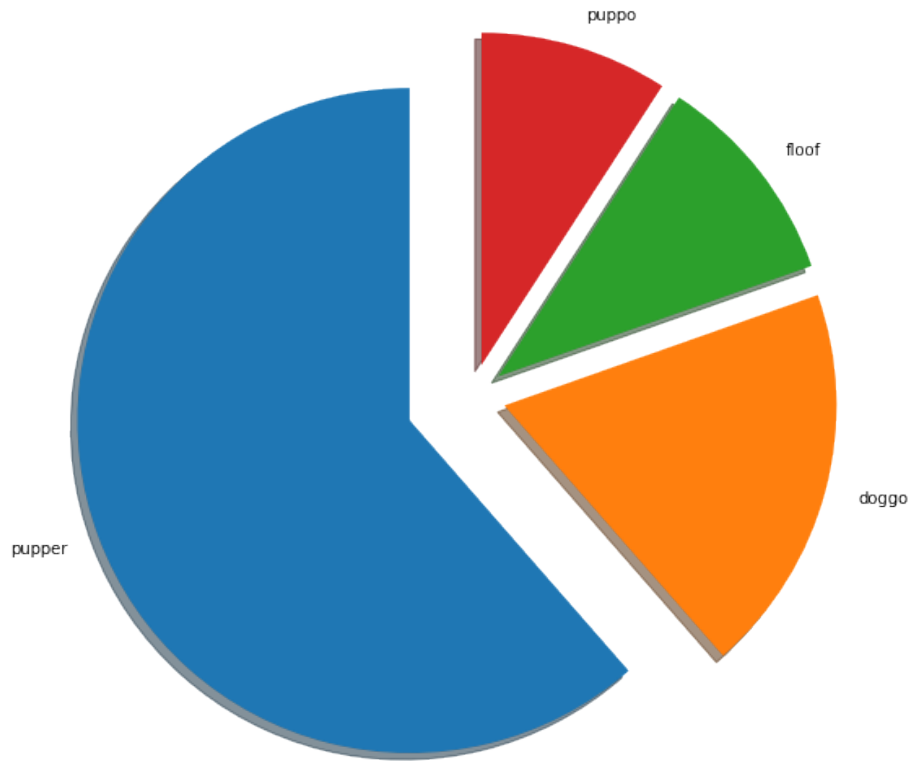
It clearly shows that the tweets have been substantially decreased over the period of time. We can see fast increase of tweets when the #WeRateDog account just started but after that it is downfall in number of tweets as the time passes.

4. Which dog_stage is having the a high proportion in the provided table

```
In [47]: # Plot the data partitioned by dog stage
dog_stage_count = list(df_master[df_master['dog_stages'] != 'None']['dog_stages'].value_counts())
dog_stages = df_master[df_master['dog_stages'] != 'None']['dog_stages'].value_counts().index
explode = (0.2, 0.1, 0.1, 0.1)

fig1, ax1 = plt.subplots()
ax1.pie(dog_stage_count, explode = explode, labels = dog_stages, shadow = True, startangle=90)
ax1.axis('equal')
```

```
Out[47]: (-1.3152196293355578,
          1.2304343477800763,
          -1.1892043309943232,
          1.2047108916820504)
```



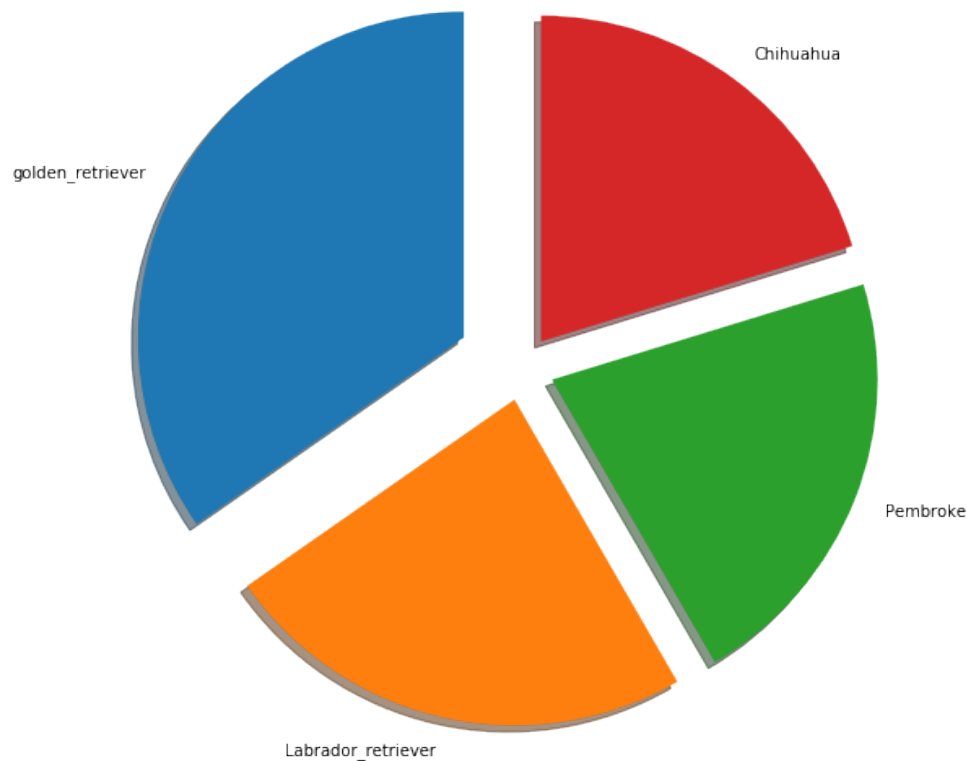
Ok, so from above pie chart it is clear that more than half of the data is related to dog stage pupper. It would be interesting to see how the breed is divided in provided data with the given context of dog stage above.

5. Which dog breed is having the highest distribution in provided data Let's create another pie chart displaying the dog breed proportion

```
In [48]: # Plot the data partitioned by dog stage
dog_stage_count = list(df_master[df_master['breed'] != 'Unidentifiable']['breed'].value_counts())
dog_stages = df_master[df_master['breed'] != 'Unidentifiable']['breed'].value_counts().index
explode = (0.2, 0.1, 0.1, 0.1)

fig1, ax1 = plt.subplots()
ax1.pie(dog_stage_count, explode = explode, labels = dog_stages, shadow = True, startangle=90)
ax1.axis('equal')
```

```
Out[48]: (-1.3020110348114706,
1.2305555614256674,
-1.2391766942374782,
1.2035938901078009)
```



Hmmm, golden retriever breed is more in the provided data which understandably reflects as more favourite dog breed with high likes and retweets. Let's check first 10 breeds highest in numbers in provided data

```
In [49]: breed_count = df_master['breed'].value_counts().head(10)
breed_count
```

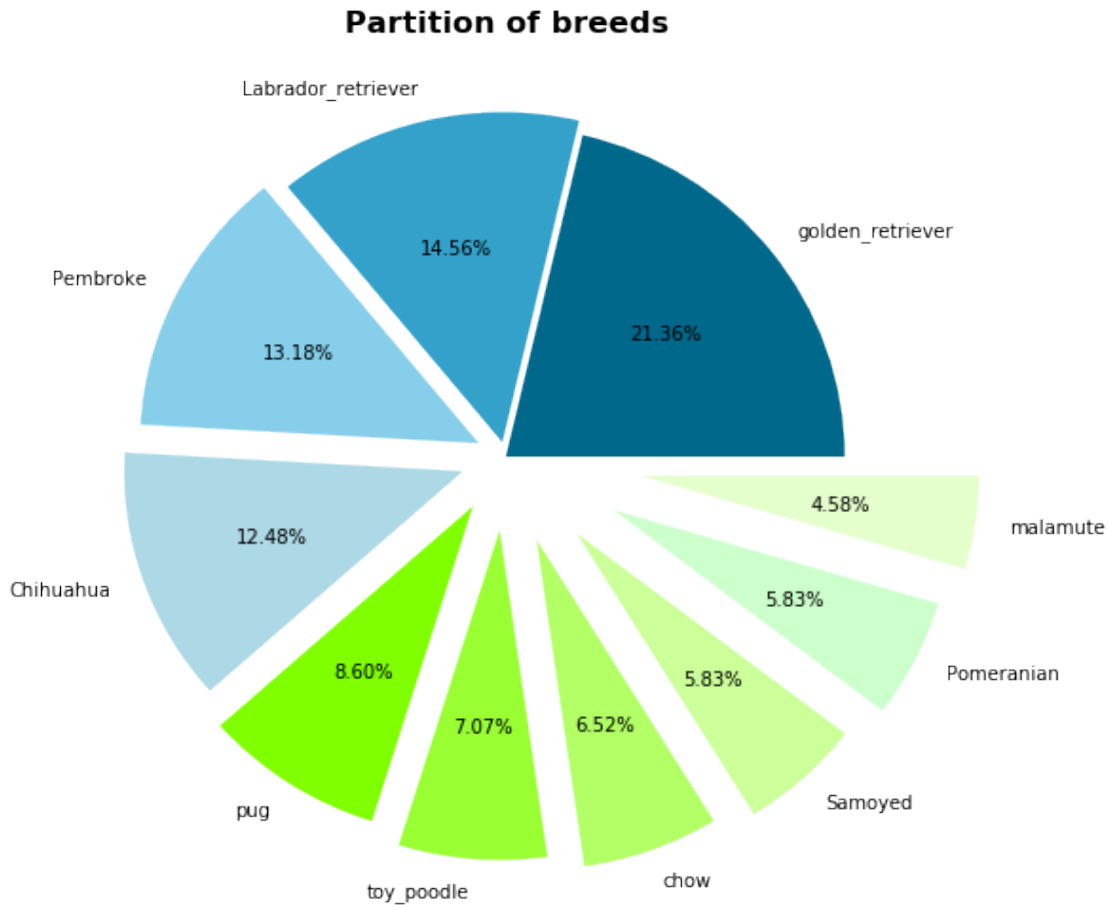
```
Out[49]: golden_retriever    154
Labrador_retriever    105
Pembroke              95
Chihuahua             90
pug                   62
toy_poodle            51
chow                  47
Pomeranian            42
Samoyed               42
malamute              33
Name: breed, dtype: int64
```

Ok, we got the numbers but how these 10 breeds reflects in proportion in large picture. We will create pie chart this time with their proportion percentage

6. What is the ratio of dog breeds in given data

In [50]: *# Shows the repartition of the breeds with a pie chart.*

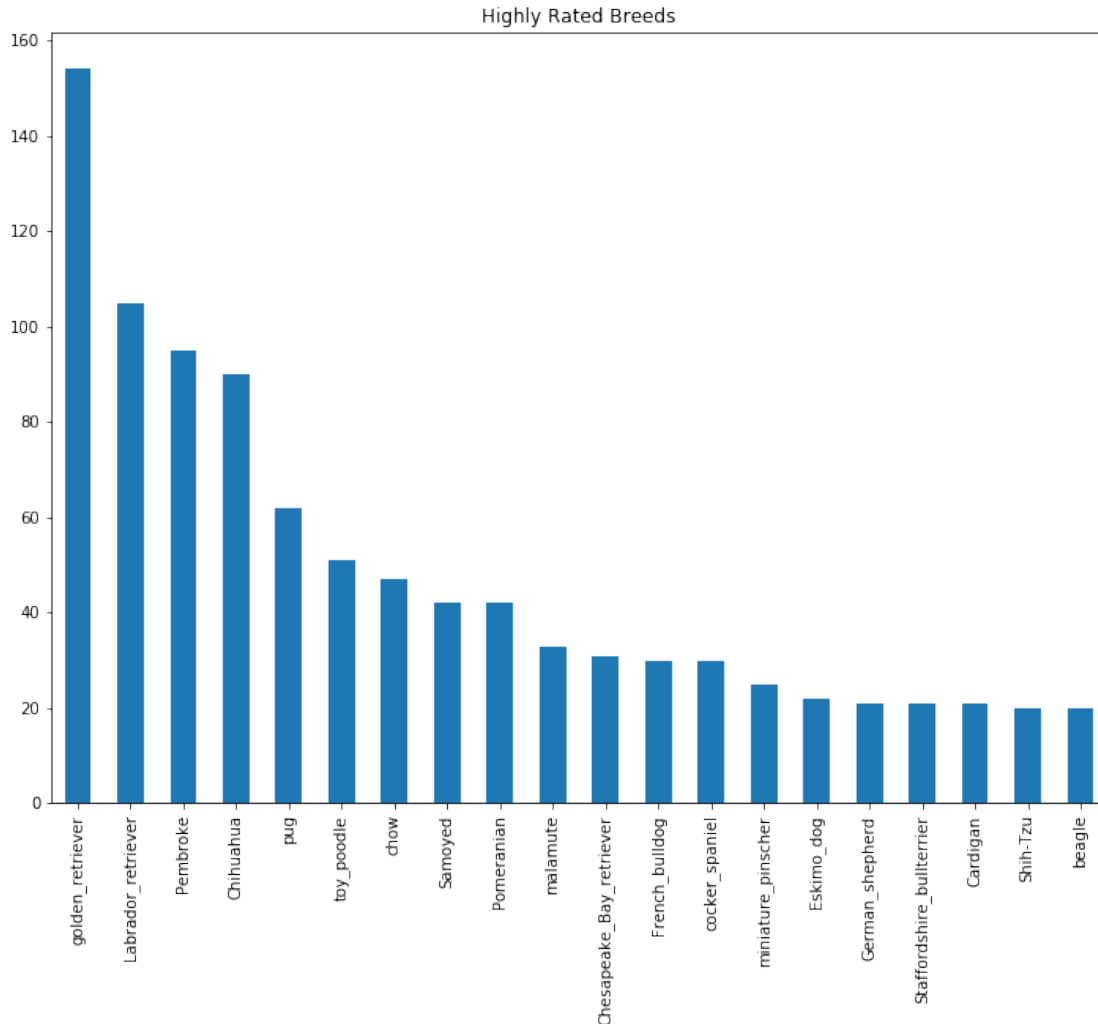
```
explode = np.linspace(0,0.4,10)
colors = ['#00688B', '#33A1C9', '#87CEEB', '#ADD8E6', '#80FF00', '#99FF33', '#B2FF66', '#90EE90', '#90EE90', '#90EE90']
breed_count.sort_values(ascending=False).plot.pie(legend=False, subplots=True, autopct=
plt.ylabel('')
plt.title('Partition of breeds', weight='bold', fontsize=16)
plt.savefig('Breeds_Partition.png')
```



Clearly shows that golden_retriever is highest in numbers as compared to other breeds. More than half of the data is covered by golden_retriever, Labrador_retriever, Pembroke and Chihuahua.

7. How much difference are in the ratings of high rated breeds. We can check it by taking the first 20 popular breeds to get wider picture for our analysis

```
In [51]: top=df_master.groupby('breed').filter(lambda x: len(x) >= 20)
top['breed'].value_counts().plot(kind = 'bar')
plt.title('Highly Rated Breeds');
```



golden retriever is almost 25% more popular than the second runner up Labrador retriever. May be this breed is more commonly found & more adored in the area from where this data is retrieved, can be many such factors which are responsible for their popularity. The next 6 breed dogs are having gradual decline and other breeds more or less are constant in ratings. It would be interesting to check top 8 dogs with the highest retweets. Let's check this :

8. Check top 8 most popular dogs and dog description

```
In [53]: # first, we will get the highest 10 values for the retweet count in our dataframe
top_8_retweets = df_master['retweets'].nlargest(8, keep='first')

# filter the dataframe to get only the records with the highest 8 retweets count in the
most_popular_dogs = df_master[df_master['retweets'].isin(top_8_retweets)]

# Sort the results descendingly by "retweet_count" field
most_popular_dogs = most_popular_dogs.sort_values(by=['retweets'], ascending=False)
```

most_popular_dogs

Out [53]:

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator	rating	name	favorites	retweets	user_followers
846	744234799360020481	2016-06-18 18:26:18	<a href="http://twitter.com/download/iphone" r...	Here's a doggo realizing you can stand in a po...	https://twitter.com/dog_rates/status/744234799...	13.0	10.0	1.3	NaN	164966	84148	7431639
881	739238157791694849	2016-06-04 23:31:25	<a href="http://twitter.com/download/iphone" r...	Here's a doggo blowing bubbles. It's downright...	https://twitter.com/dog_rates/status/739238157...	13.0	10.0	1.3	NaN	121961	62329	7431639
428	807106840509214720	2016-12-09 06:17:20	<a href="http://twitter.com/download/iphone" r...	This is Stephan. He just wants to help. 13/10 ...	https://twitter.com/dog_rates/status/807106840...	13.0	10.0	1.3	Stephan	127649	61334	7431633
325	822872901745569793	2017-01-21 18:26:02	<a href="http://twitter.com/download/iphone" r...	Here's a super supportive puppo participating ...	https://twitter.com/dog_rates/status/822872901...	13.0	10.0	1.3	NaN	141056	47999	7431633
59	879415818425184262	2017-06-26 19:07:24	<a href="http://twitter.com/download/iphone" r...	This is Duddles. He did an attempt. 13/10 some...	https://twitter.com/dog_rates/status/879415818...	13.0						
354	819004803107983360	2017-01-11 02:15:36	<a href="http://twitter.com/download/iphone" r...	This is Bo. He was a very good First Doggo. 14...	https://twitter.com/dog_rates/status/819004803...	14.0						
431	806629075125202948	2016-12-07 22:38:52	<a href="http://twitter.com/download/iphone" r...	"Good afternoon class today we're going to lea...	https://twitter.com/dog_rates/status/806629075...	13.0						
111	866450705531457537	2017-05-22 00:28:40	<a href="http://twitter.com/download/iphone" r...	This is Jamesy. He gives a kiss to every other...	https://twitter.com/dog_rates/status/866450705...	13.0						

59	10.0	1.3	Duddles	104488	43662	7431630
354	10.0	1.4	Bo	92344	40256	7431633
431	10.0	1.3	NaN	80773	38481	7431633
111	10.0	1.3	Jamesy	122489	35651	7431631

	user_favourites	jpg_url \
846	139602	https://pbs.twimg.com/ext_tw_video_thumb/74423...
881	139602	https://pbs.twimg.com/ext_tw_video_thumb/73923...
428	139602	https://pbs.twimg.com/ext_tw_video_thumb/80710...
325	139602	https://pbs.twimg.com/media/C2tugXLXgAARJ04.jpg
59	139602	https://pbs.twimg.com/ext_tw_video_thumb/87941...
354	139602	https://pbs.twimg.com/media/C12whDoVEAALRxa.jpg
431	139602	https://pbs.twimg.com/media/CzG425nWgAAnP7P.jpg
111	139602	https://pbs.twimg.com/media/DAZAUfBXcAAG_Nn.jpg

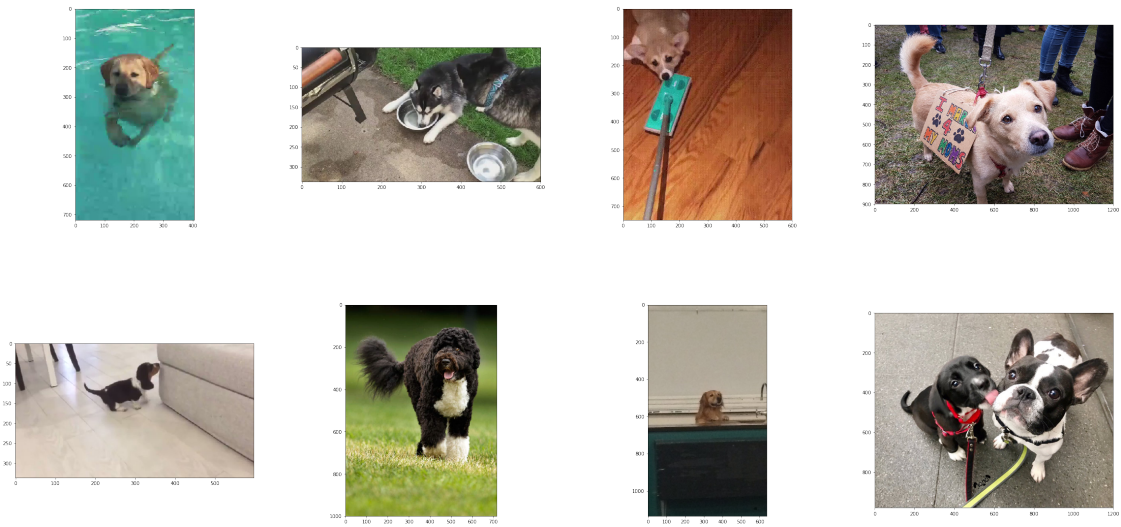
	img_num	dog_stages	breed	confidence
846	1	doggo	Labrador_retriever	0.825333
881	1	doggo	Eskimo_dog	0.503372
428	1	None	Chihuahua	0.505370
325	1	puppo	Lakeland_terrier	0.196015
59	1	None	English_springer	0.383404
354	1	doggo	standard_poodle	0.351308
431	2	None	cocker_spaniel	0.046403
111	2	pupper	French_bulldog	0.905334

Wow! here Labrador retriever breed got the first place in retweets and popularity. I am sure all of these 8 dogs must be cute. I would like to see how they look so let's check their picture in above order.

```
In [54]: fig = plt.figure(figsize=(40,30))
fig.suptitle("Most popular dogs having high retweets count", fontsize=24)
columns = 4
idx = 0
fig.subplots_adjust(hspace=0.4)

for i, row in most_popular_dogs.iterrows():
    img_url = row['jpg_url']
    image = io.imread(img_url)
    plt.subplot(most_popular_dogs.shape[0] / columns + 1, columns, idx + 1)
    plt.imshow(image)
    idx += 1
```

Most popular dogs having high retweets count



Aww! These are very cute dogs and no surprise that they are top 8 popular dogs.