wrangle_act

February 22, 2021

1 1. Data Wrangling

1.1 1.1. Gathering Data

1.1.1 1.1.1. Define:

- import csv file (twitter-archive-enhanced.csv) into a DataFrame (df_twitter_raw)
- Download & import tsv file (image-predictions.tsv) into a DataFrame (df_images_raw)
- import json file (tweet-json.txt) into a DataFrame (df_tweets_raw)

1.1.2 1.1.2. Code:

```
In [1]: import pandas as pd
        import json
        import numpy as np
        import tweepy as tw
        import os
        import requests as req
        import re
        import matplotlib.pyplot as plt
        %matplotlib inline
        import seaborn as sb
```

1.1.2.1. import csv file (twitter-archive-enhanced.csv) into a DataFrame (df_twitter_raw)

```
In [2]: df_twitter_raw = pd.read_csv('twitter-archive-enhanced.csv')
```

1.1.2.2. Download & import tsv file (image-predictions.tsv) into a DataFrame (df_images_raw)

```
In [3]: url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictio
        file = url.split('/')[-1]
        r = req.get(url)
        if not os.path.isfile(file):
            with open(file, 'wb') as f:
                f.write(r.content)
```

```
In [4]: df_images_raw = pd.read_csv('image-predictions.tsv', sep='\t')
```

1.1.2.3. import json file (tweet-json.txt) into a DataFrame (df_tweets_raw)

```
In [5]: tweets = []
        with open('tweet-json.txt') as f:
            for line in f:
                tweets.append(json.loads(line))
        dic_tweets = tweets[0].keys()
        df_tweets_raw = pd.DataFrame(tweets, columns = dic_tweets)

In [6]: df_tweets_raw.to_csv('tweets_raw.csv', index= False)
```

1.2 1.2. Assessment

1.2.1 1.2.1. Define:

- asses the 3 dataframes visually and programatically

1.2.2 1.2.2. Code:

1.2.2.1. df_twitter_raw

```
In [7]: pd.set_option('display.max_colwidth', -1)

In [8]: df_twitter_raw
```

Out[8]:

	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" rel="nofollow">Twitter for iPhone</a>
1     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
3     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
4     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
5     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
6     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
7     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
8     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
9     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
10    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
11    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
12    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
13    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
14    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
15    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
16    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
17    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
18    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
19    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
20    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
21    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
22    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
23    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
24    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
25    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
26    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
27    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
28    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
29    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
...                                                                                               ...
2326  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2327  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2328  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2329  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2330  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2331  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2332  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2333  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2334  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2335  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2336  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
```

```
2337  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2338  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2339  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2340  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2341  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2342  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2343  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2344  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2345  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2346  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2347  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2348  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2349  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2350  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2351  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2352  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2353  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2354  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2355  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
```

```
0     This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13
1     This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she'
2     This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall grass. Yo
3     This is Darla. She commenced a snooze mid meal. 13/10 happens to the best of us ht
4     This is Franklin. He would like you to stop calling him "cute." He is a very fierc
5     Here we have a majestic great white breaching off South Africa's coast. Absolutely
6     Meet Jax. He enjoys ice cream so much he gets nervous around it. 13/10 help Jax en
7     When you watch your owner call another dog a good boy but then they turn back to y
8     This is Zoey. She doesn't want to be one of the scary sharks. Just wants to be a s
9     This is Cassie. She is a college pup. Studying international doggo communication a
10    This is Koda. He is a South Australian deckshark. Deceptively deadly. Frighteningl
11    This is Bruno. He is a service shark. Only gets out of the water to assist you. 13
12    Here's a puppo that seems to be on the fence about something haha no but seriously
13    This is Ted. He does his best. Sometimes that's not enough. But it's ok. 12/10 wou
14    This is Stuart. He's sporting his favorite fanny pack. Secretly filled with bones
15    This is Oliver. You're witnessing one of his many brutal attacks. Seems to be play
16    This is Jim. He found a fren. Taught him how to sit like the good boys. 12/10 for
17    This is Zeke. He has a new stick. Very proud of it. Would like you to throw it for
18    This is Ralphus. He's powering up. Attempting maximum borkdrive. 13/10 inspiration
19    RT @dog_rates: This is Canela. She attempted some fancy porch pics. They were unsu
20    This is Gerald. He was just told he didn't get the job he interviewed for. A h*cki
21    This is Jeffrey. He has a monopoly on the pool noodles. Currently running a 'boop
22    I've yet to rate a Venezuelan Hover Wiener. This is such an honor. 14/10 paw-inspi
23    This is Canela. She attempted some fancy porch pics. They were unsuccessful. 13/10
24    You may not have known you needed to see this today. 13/10 please enjoy (IG: emmyl
25    This... is a Jubilant Antarctic House Bear. We only rate dogs. Please only send do
26    This is Maya. She's very shy. Rarely leaves her cup. 13/10 would find her an envir
```

```
27    This is Mingus. He's a wonderful father to his smol pup. Confirmed 13/10, but he n
28    This is Derek. He's late for a dog meeting. 13/10 pet...al to the metal https://t.
29    This is Roscoe. Another pupper fallen victim to spontaneous tongue ejections. Get
...
2326  This is quite the dog. Gets really excited when not in water. Not very soft tho. B
2327  This is a southern Vesuvius bumblegruff. Can drive a truck (wow). Made friends wit
2328  Oh goodness. A super rare northeast Qdoba kangaroo mix. Massive feet. No pouch (di
2329  Those are sunglasses and a jean jacket. 11/10 dog cool af https://t.co/uHXrPkUEyl
2330  Unique dog here. Very small. Lives in container of Frosted Flakes (?). Short legs.
2331  Here we have a mixed Asiago from the Galápagos Islands. Only one ear working. Big
2332  Look at this jokester thinking seat belt laws don't apply to him. Great tongue tho
2333  This is an extremely rare horned Parthenon. Not amused. Wears shoes. Overall very
2334  This is a funny dog. Weird toes. Won't come down. Loves branch. Refuses to eat his
2335  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood floor
2336  Can take selfies 11/10 https://t.co/ws2AMaNwPW
2337  Very concerned about fellow dog trapped in computer. 10/10 https://t.co/OyxApIikpk
2338  Not familiar with this breed. No tail (weird). Only 2 legs. Doesn't bark. Surprisi
2339  Oh my. Here you are seeing an Adobe Setter giving birth to twins!!! The world is a
2340  Can stand on stump for what seems like a while. Built that birdhouse? Impressive.
2341  This appears to be a Mongolian Presbyterian mix. Very tired. Tongue slip confirmed
2342  Here we have a well-established sunblockerspaniel. Lost his other flip-flop. 6/10
2343  Let's hope this flight isn't Malaysian (lol). What a dog! Almost completely camouf
2344  Here we have a northern speckled Rhododendron. Much sass. Gives 0 fucks. Good tong
2345  This is the happiest dog you will ever see. Very committed owner. Nice couch. 10/1
2346  Here is the Rand Paul of retrievers folks! He's probably good at poker. Can drink
2347  My oh my. This is a rare blond Canadian terrier on wheels. Only $8.98. Rather doci
2348  Here is a Siberian heavily armored polar bear mix. Strong owner. 10/10 I would do
2349  This is an odd dog. Hard on the outside but loving on the inside. Petting still fu
2350  This is a truly beautiful English Wilson Staff retriever. Has a nice phone. Privil
2351  Here we have a 1949 1st generation vulpix. Enjoys sweat tea and Fox News. Cannot b
2352  This is a purebred Piers Morgan. Loves to Netflix and chill. Always looks like he
2353  Here is a very happy pup. Big fan of well-maintained decks. Just look at that tong
2354  This is a western brown Mitsubishi terrier. Upset about leaf. Actually 2 dogs here
2355  Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big fan of relaxing

      retweeted_status_id  retweeted_status_user_id  \
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    8.874740e+17         4.196984e+09
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

     retweeted_status_timestamp  \
0     NaN
1     NaN
2     NaN
3     NaN
4     NaN
5     NaN
6     NaN
7     NaN
8     NaN
9     NaN
10    NaN
11    NaN
12    NaN
13    NaN
14    NaN
15    NaN
16    NaN
17    NaN
18    NaN
19    2017-07-19 00:47:34 +0000
20    NaN
21    NaN
22    NaN
23    NaN
24    NaN
25    NaN
26    NaN
27    NaN
28    NaN
29    NaN
...   ...
2326  NaN
2327  NaN
2328  NaN
2329  NaN
2330  NaN
2331  NaN
2332  NaN
2333  NaN
2334  NaN
2335  NaN
2336  NaN
2337  NaN
2338  NaN
2339  NaN
```

```
2340  NaN
2341  NaN
2342  NaN
2343  NaN
2344  NaN
2345  NaN
2346  NaN
2347  NaN
2348  NaN
2349  NaN
2350  NaN
2351  NaN
2352  NaN
2353  NaN
2354  NaN
2355  NaN
```

```
0     https://twitter.com/dog_rates/status/892420643555336193/photo/1
1     https://twitter.com/dog_rates/status/892177421306343426/photo/1
2     https://twitter.com/dog_rates/status/891815181378084864/photo/1
3     https://twitter.com/dog_rates/status/891689557279858688/photo/1
4     https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://twitter.co
5     https://twitter.com/dog_rates/status/891087950875897856/photo/1
6     https://gofundme.com/ydvmve-surgery-for-jax,https://twitter.com/dog_rates/status/8
7     https://twitter.com/dog_rates/status/890729181411237888/photo/1,https://twitter.co
8     https://twitter.com/dog_rates/status/890609185150312448/photo/1
9     https://twitter.com/dog_rates/status/890240255349198849/photo/1
10    https://twitter.com/dog_rates/status/890006608113172480/photo/1,https://twitter.co
11    https://twitter.com/dog_rates/status/889880896479866881/photo/1
12    https://twitter.com/dog_rates/status/889665388333682689/photo/1
13    https://twitter.com/dog_rates/status/889638837579907072/photo/1,https://twitter.co
14    https://twitter.com/dog_rates/status/889531135344209921/photo/1
15    https://twitter.com/dog_rates/status/889278841981685760/video/1
16    https://twitter.com/dog_rates/status/888917238123831296/photo/1
17    https://twitter.com/dog_rates/status/888804989199671297/photo/1,https://twitter.co
18    https://twitter.com/dog_rates/status/888554962724278272/photo/1,https://twitter.co
19    https://twitter.com/dog_rates/status/887473957103951883/photo/1,https://twitter.co
20    https://twitter.com/dog_rates/status/888078434458587136/photo/1,https://twitter.co
21    https://twitter.com/dog_rates/status/887705289381826560/photo/1
22    https://twitter.com/dog_rates/status/887517139158093824/video/1
23    https://twitter.com/dog_rates/status/887473957103951883/photo/1,https://twitter.co
24    https://twitter.com/dog_rates/status/887343217045368832/video/1
25    https://twitter.com/dog_rates/status/887101392804085760/photo/1
26    https://twitter.com/dog_rates/status/886983233522544640/photo/1,https://twitter.co
27    https://www.gofundme.com/mingusneedsus,https://twitter.com/dog_rates/status/886736
28    https://twitter.com/dog_rates/status/886680336477933568/photo/1
29    https://twitter.com/dog_rates/status/886366144734445568/photo/1,https://twitter.co
```

```
...
2326  https://twitter.com/dog_rates/status/666411507551481857/photo/1
2327  https://twitter.com/dog_rates/status/666407126856765440/photo/1
2328  https://twitter.com/dog_rates/status/666396247373291520/photo/1
2329  https://twitter.com/dog_rates/status/666373753744588802/photo/1
2330  https://twitter.com/dog_rates/status/666362758909284353/photo/1
2331  https://twitter.com/dog_rates/status/666353288456101888/photo/1
2332  https://twitter.com/dog_rates/status/666345417576210432/photo/1
2333  https://twitter.com/dog_rates/status/666337882303524864/photo/1
2334  https://twitter.com/dog_rates/status/666293911632134144/photo/1
2335  https://twitter.com/dog_rates/status/666287406224695296/photo/1
2336  https://twitter.com/dog_rates/status/666273097616637952/photo/1
2337  https://twitter.com/dog_rates/status/666268910803644416/photo/1
2338  https://twitter.com/dog_rates/status/666104133288665088/photo/1
2339  https://twitter.com/dog_rates/status/666102155909144576/photo/1
2340  https://twitter.com/dog_rates/status/666099513787052032/photo/1
2341  https://twitter.com/dog_rates/status/666094000022159362/photo/1
2342  https://twitter.com/dog_rates/status/666082916733198337/photo/1
2343  https://twitter.com/dog_rates/status/666073100786774016/photo/1
2344  https://twitter.com/dog_rates/status/666071193221509120/photo/1
2345  https://twitter.com/dog_rates/status/666063827256086533/photo/1
2346  https://twitter.com/dog_rates/status/666058600524156928/photo/1
2347  https://twitter.com/dog_rates/status/666057090499244032/photo/1
2348  https://twitter.com/dog_rates/status/666055525042405380/photo/1
2349  https://twitter.com/dog_rates/status/666051853826850816/photo/1
2350  https://twitter.com/dog_rates/status/666050758794694657/photo/1
2351  https://twitter.com/dog_rates/status/666049248165822465/photo/1
2352  https://twitter.com/dog_rates/status/666044226329800704/photo/1
2353  https://twitter.com/dog_rates/status/666033412701032449/photo/1
2354  https://twitter.com/dog_rates/status/666029285002620928/photo/1
2355  https://twitter.com/dog_rates/status/666020888022790149/photo/1

      rating_numerator  rating_denominator      name  doggo floofer  pupper  \
0     13                10                   Phineas   None   None    None
1     13                10                   Tilly     None   None    None
2     12                10                   Archie    None   None    None
3     13                10                   Darla     None   None    None
4     12                10                   Franklin  None   None    None
5     13                10                   None      None   None    None
6     13                10                   Jax       None   None    None
7     13                10                   None      None   None    None
8     13                10                   Zoey      None   None    None
9     14                10                   Cassie    doggo  None    None
10    13                10                   Koda      None   None    None
11    13                10                   Bruno     None   None    None
12    13                10                   None      None   None    None
13    12                10                   Ted       None   None    None
14    13                10                   Stuart    None   None    None
```

15	13	10	Oliver	None	None	None
16	12	10	Jim	None	None	None
17	13	10	Zeke	None	None	None
18	13	10	Ralphus	None	None	None
19	13	10	Canela	None	None	None
20	12	10	Gerald	None	None	None
21	13	10	Jeffrey	None	None	None
22	14	10	such	None	None	None
23	13	10	Canela	None	None	None
24	13	10	None	None	None	None
25	12	10	None	None	None	None
26	13	10	Maya	None	None	None
27	13	10	Mingus	None	None	None
28	13	10	Derek	None	None	None
29	12	10	Roscoe	None	None	pupper
...	..	..	...	...	...	...
2326	2	10	quite	None	None	None
2327	7	10	a	None	None	None
2328	9	10	None	None	None	None
2329	11	10	None	None	None	None
2330	6	10	None	None	None	None
2331	8	10	None	None	None	None
2332	10	10	None	None	None	None
2333	9	10	an	None	None	None
2334	3	10	a	None	None	None
2335	1	2	an	None	None	None
2336	11	10	None	None	None	None
2337	10	10	None	None	None	None
2338	1	10	None	None	None	None
2339	11	10	None	None	None	None
2340	8	10	None	None	None	None
2341	9	10	None	None	None	None
2342	6	10	None	None	None	None
2343	10	10	None	None	None	None
2344	9	10	None	None	None	None
2345	10	10	the	None	None	None
2346	8	10	the	None	None	None
2347	9	10	a	None	None	None
2348	10	10	a	None	None	None
2349	2	10	an	None	None	None
2350	10	10	a	None	None	None
2351	5	10	None	None	None	None
2352	6	10	a	None	None	None
2353	9	10	a	None	None	None
2354	7	10	a	None	None	None
2355	8	10	None	None	None	None

puppo

```
0     None
1     None
2     None
3     None
4     None
5     None
6     None
7     None
8     None
9     None
10    None
11    None
12    puppo
13    None
14    puppo
15    None
16    None
17    None
18    None
19    None
20    None
21    None
22    None
23    None
24    None
25    None
26    None
27    None
28    None
29    None
...    ...
2326  None
2327  None
2328  None
2329  None
2330  None
2331  None
2332  None
2333  None
2334  None
2335  None
2336  None
2337  None
2338  None
2339  None
2340  None
2341  None
2342  None
```

```
2343  None
2344  None
2345  None
2346  None
2347  None
2348  None
2349  None
2350  None
2351  None
2352  None
2353  None
2354  None
2355  None

[2356 rows x 17 columns]
```

```
In [9]: df_twitter_raw.info()
```

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

```
In [10]: df_twitter_raw.describe()
```

```
Out[10]:            tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         count  2.356000e+03  7.800000e+01           7.800000e+01
         mean   7.427716e+17  7.455079e+17           2.014171e+16
         std    6.856705e+16  7.582492e+16           1.252797e+17
         min    6.660209e+17  6.658147e+17           1.185634e+07
```

```
25%    6.783989e+17  6.757419e+17          3.086374e+08
50%    7.196279e+17  7.038708e+17          4.196984e+09
75%    7.993373e+17  8.257804e+17          4.196984e+09
max    8.924206e+17  8.862664e+17          8.405479e+17

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

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

In [11]:
```
df_twitter_raw['rating_denominator'].value_counts()
```

Out[11]:
```
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 [12]:
```
df_twitter_raw['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]: sum(df_twitter_raw.duplicated())
```

```
Out[13]: 0
```

- some tweets are replies and retweets
- timestamp column dtype not correct

- dog stages columns contain 'None' as values
- rating_numerator contains wrong inputs
- rating_denominator contains wrong inputs
- tweet_id column dtype not correct
- Some tweets are not with expanded_urls (no Images)
- source of tweeting included in the source url
- columns not required in master sheet (in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp, expanded_urls)
- Set column (tweet_id) as index when merging dataframes

1.2.2.2. df_images_raw

```
In [14]: df_images_raw
```

```
Out[14]:                 tweet_id  \
         0     666020888022790149
         1     666029285002620928
         2     666033412701032449
         3     666044226329800704
         4     666049248165822465
         5     666050758794694657
         6     666051853826850816
         7     666055525042405380
         8     666057090499244032
         9     666058600524156928
         10    666063827256086533
         11    666071193221509120
         12    666073100786774016
         13    666082916733198337
         14    666094000022159362
         15    666099513787052032
         16    666102155909144576
         17    666104133288665088
         18    666268910803644416
         19    666273097616637952
         20    666287406224695296
         21    666293911632134144
         22    666337882303524864
         23    666345417576210432
         24    666353288456101888
         25    666362758909284353
         26    666373753744588802
         27    666396247373291520
         28    666407126856765440
         29    666411507551481857
         ...                  ...
         2045  886366144734445568
```

2046	886680336477933568
2047	886736880519319552
2048	886983233522544640
2049	887101392804085760
2050	887343217045368832
2051	887473957103951883
2052	887517139158093824
2053	887705289381826560
2054	888078434458587136
2055	888202515573088257
2056	888554962724278272
2057	888804989199671297
2058	888917238123831296
2059	889278841981685760
2060	889531135344209921
2061	889638837579907072
2062	889665388333682689
2063	889880896479866881
2064	890006608113172480
2065	890240255349198849
2066	890609185150312448
2067	890729181411237888
2068	890971913173991426
2069	891087950875897856
2070	891327558926688256
2071	891689557279858688
2072	891815181378084864
2073	892177421306343426
2074	892420643555336193

0	https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg
1	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
5	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8	https://pbs.twimg.com/media/CT5PY9OWoAAQGLo.jpg
9	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11	https://pbs.twimg.com/media/CT5cN_3WEAAlOoZ.jpg
12	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13	https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.jpg
14	https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg
15	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16	https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg

```
17    https://pbs.twimg.com/media/CT56LSZWoAAlJj2.jpg
18    https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
19    https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
20    https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
21    https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22    https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg
23    https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg
24    https://pbs.twimg.com/media/CT9cxOtUEAAhNN_.jpg
25    https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
26    https://pbs.twimg.com/media/CT9vZEYWUAAlZO5.jpg
27    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
29    https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...                                                ...
2045  https://pbs.twimg.com/media/DEOBTnQUwAApKEH.jpg
2046  https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2047  https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
2048  https://pbs.twimg.com/media/DE8yicJWOAAAvBJ.jpg
2049  https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050  https://pbs.twimg.com/ext_tw_video_thumb/887343120832229379/pu/img/6HSuFrW1lzI_9M
2051  https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2052  https://pbs.twimg.com/ext_tw_video_thumb/887517108413886465/pu/img/WanJKwssZj4VJv
2053  https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054  https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
2055  https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056  https://pbs.twimg.com/media/DFTH_O-UQAACu20.jpg
2057  https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg
2059  https://pbs.twimg.com/ext_tw_video_thumb/889278779352338437/pu/img/VlbFB3v8H8VwzV
2060  https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg
2061  https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062  https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063  https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg
2064  https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065  https://pbs.twimg.com/media/DFrEyVuWOAAO3t9.jpg
2066  https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067  https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068  https://pbs.twimg.com/media/DF1eOmZXUAALUcq.jpg
2069  https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074  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]
```

```
In [15]: df_images_raw.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
```

```
Data columns (total 12 columns):
tweet_id    2075 non-null int64
jpg_url     2075 non-null object
img_num     2075 non-null int64
p1          2075 non-null object
p1_conf     2075 non-null float64
p1_dog      2075 non-null bool
p2          2075 non-null object
p2_conf     2075 non-null float64
p2_dog      2075 non-null bool
p3          2075 non-null object
p3_conf     2075 non-null float64
p3_dog      2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

```
In [16]: df_images_raw.describe()
```

Out[16]:

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

```
In [17]: df_images_raw['img_num'].value_counts()
```

```
Out[17]: 1    1780
         2    198
         3    66
         4    31
         Name: img_num, dtype: int64
```

```
In [18]: sum(df_images_raw.duplicated())
```

```
Out[18]: 0
```

- There are entries in df_twitter_raw without images data
- tweet_id column dtype not correct
- columns labels not expressive
- images breed predection and if it's dog (distributed over 9 columns)
- identify the Non-dog images for tweet ids
- Set column (tweet_id) as index when merging dataframes

1.2.2.1. df_tweets_raw

```
In [19]: df_tweets_raw
```

Out[19]:

	created_at	id	id_str
0	Tue Aug 01 16:23:56 +0000 2017	892420643555336193	892420643555336193
1	Tue Aug 01 00:17:27 +0000 2017	892177421306343426	892177421306343426
2	Mon Jul 31 00:18:03 +0000 2017	891815181378084864	891815181378084864
3	Sun Jul 30 15:58:51 +0000 2017	891689557279858688	891689557279858688
4	Sat Jul 29 16:00:24 +0000 2017	891327558926688256	891327558926688256
5	Sat Jul 29 00:08:17 +0000 2017	891087950875897856	891087950875897856
6	Fri Jul 28 16:27:12 +0000 2017	890971913173991426	890971913173991426
7	Fri Jul 28 00:22:40 +0000 2017	890729181411237888	890729181411237888
8	Thu Jul 27 16:25:51 +0000 2017	890609185150312448	890609185150312448
9	Wed Jul 26 15:59:51 +0000 2017	890240255349198849	890240255349198849
10	Wed Jul 26 00:31:25 +0000 2017	890006608113172480	890006608113172480
11	Tue Jul 25 16:11:53 +0000 2017	889880896479866881	889880896479866881
12	Tue Jul 25 01:55:32 +0000 2017	889665388333682689	889665388333682689
13	Tue Jul 25 00:10:02 +0000 2017	889638837579907072	889638837579907072
14	Mon Jul 24 17:02:04 +0000 2017	889531135344209921	889531135344209921
15	Mon Jul 24 00:19:32 +0000 2017	889278841981685760	889278841981685760
16	Sun Jul 23 00:22:39 +0000 2017	888917238123831296	888917238123831296
17	Sat Jul 22 16:56:37 +0000 2017	888804989199671297	888804989199671297
18	Sat Jul 22 00:23:06 +0000 2017	888554962724278272	888554962724278272
19	Thu Jul 20 16:49:33 +0000 2017	888078434458587136	888078434458587136
20	Wed Jul 19 16:06:48 +0000 2017	887705289381826560	887705289381826560
21	Wed Jul 19 03:39:09 +0000 2017	887517139158093824	887517139158093824
22	Wed Jul 19 00:47:34 +0000 2017	887473957103951883	887473957103951883
23	Tue Jul 18 16:08:03 +0000 2017	887343217045368832	887343217045368832
24	Tue Jul 18 00:07:08 +0000 2017	887101392804085760	887101392804085760
25	Mon Jul 17 16:17:36 +0000 2017	886983233522544640	886983233522544640
26	Sun Jul 16 23:58:41 +0000 2017	886736880519319552	886736880519319552
27	Sun Jul 16 20:14:00 +0000 2017	886680336477933568	886680336477933568
28	Sat Jul 15 23:25:31 +0000 2017	886366144734445568	886366144734445568
29	Sat Jul 15 16:51:35 +0000 2017	886267009285017600	886267009285017600
...	...	...	...
2324	Tue Nov 17 00:24:19 +0000 2015	666411507551481857	666411507551481857
2325	Tue Nov 17 00:06:54 +0000 2015	666407126856765440	666407126856765440
2326	Mon Nov 16 23:23:41 +0000 2015	666396247373291520	666396247373291520
2327	Mon Nov 16 21:54:18 +0000 2015	666373753744588802	666373753744588802
2328	Mon Nov 16 21:10:36 +0000 2015	666362758909284353	666362758909284353
2329	Mon Nov 16 20:32:58 +0000 2015	666353288456101888	666353288456101888
2330	Mon Nov 16 20:01:42 +0000 2015	666345417576210432	666345417576210432
2331	Mon Nov 16 19:31:45 +0000 2015	666337882303524864	666337882303524864
2332	Mon Nov 16 16:37:02 +0000 2015	666293911632134144	666293911632134144
2333	Mon Nov 16 16:11:11 +0000 2015	666287406224695296	666287406224695296
2334	Mon Nov 16 15:14:19 +0000 2015	666273097616637952	666273097616637952
2335	Mon Nov 16 14:57:41 +0000 2015	666268910803644416	666268910803644416
2336	Mon Nov 16 04:02:55 +0000 2015	666104133288665088	666104133288665088

2337	Mon Nov 16 03:55:04 +0000 2015	666102155909144576	666102155909144576
2338	Mon Nov 16 03:44:34 +0000 2015	666099513787052032	666099513787052032
2339	Mon Nov 16 03:22:39 +0000 2015	666094000022159362	666094000022159362
2340	Mon Nov 16 02:38:37 +0000 2015	666082916733198337	666082916733198337
2341	Mon Nov 16 01:59:36 +0000 2015	666073100786774016	666073100786774016
2342	Mon Nov 16 01:52:02 +0000 2015	666071193221509120	666071193221509120
2343	Mon Nov 16 01:22:45 +0000 2015	666063827256086533	666063827256086533
2344	Mon Nov 16 01:01:59 +0000 2015	666058600524156928	666058600524156928
2345	Mon Nov 16 00:55:59 +0000 2015	666057090499244032	666057090499244032
2346	Mon Nov 16 00:49:46 +0000 2015	666055525042405380	666055525042405380
2347	Mon Nov 16 00:35:11 +0000 2015	666051853826850816	666051853826850816
2348	Mon Nov 16 00:30:50 +0000 2015	666050758794694657	666050758794694657
2349	Mon Nov 16 00:24:50 +0000 2015	666049248165822465	666049248165822465
2350	Mon Nov 16 00:04:52 +0000 2015	666044226329800704	666044226329800704
2351	Sun Nov 15 23:21:54 +0000 2015	666033412701032449	666033412701032449
2352	Sun Nov 15 23:05:30 +0000 2015	666029285002620928	666029285002620928
2353	Sun Nov 15 22:32:08 +0000 2015	666020888022790149	666020888022790149

0	This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 1
1	This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she
2	This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall grass. Y
3	This is Darla. She commenced a snooze mid meal. 13/10 happens to the best of us h
4	This is Franklin. He would like you to stop calling him "cute." He is a very fier
5	Here we have a majestic great white breaching off South Africa's coast. Absolutel
6	Meet Jax. He enjoys ice cream so much he gets nervous around it. 13/10 help Jax e
7	When you watch your owner call another dog a good boy but then they turn back to
8	This is Zoey. She doesn't want to be one of the scary sharks. Just wants to be a
9	This is Cassie. She is a college pup. Studying international doggo communication
10	This is Koda. He is a South Australian deckshark. Deceptively deadly. Frightening
11	This is Bruno. He is a service shark. Only gets out of the water to assist you. 1
12	Here's a puppo that seems to be on the fence about something haha no but seriousl
13	This is Ted. He does his best. Sometimes that's not enough. But it's ok. 12/10 wo
14	This is Stuart. He's sporting his favorite fanny pack. Secretly filled with bones
15	This is Oliver. You're witnessing one of his many brutal attacks. Seems to be pla
16	This is Jim. He found a fren. Taught him how to sit like the good boys. 12/10 for
17	This is Zeke. He has a new stick. Very proud of it. Would like you to throw it fo
18	This is Ralphus. He's powering up. Attempting maximum borkdrive. 13/10 inspiratio
19	This is Gerald. He was just told he didn't get the job he interviewed for. A h*ck
20	This is Jeffrey. He has a monopoly on the pool noodles. Currently running a 'boop
21	I've yet to rate a Venezuelan Hover Wiener. This is such an honor. 14/10 paw-insp
22	This is Canela. She attempted some fancy porch pics. They were unsuccessful. 13/1
23	You may not have known you needed to see this today. 13/10 please enjoy (IG: emmy
24	This... is a Jubilant Antarctic House Bear. We only rate dogs. Please only send d
25	This is Maya. She's very shy. Rarely leaves her cup. 13/10 would find her an envi
26	This is Mingus. He's a wonderful father to his smol pup. Confirmed 13/10, but he
27	This is Derek. He's late for a dog meeting. 13/10 pet...al to the metal https://t
28	This is Roscoe. Another pupper fallen victim to spontaneous tongue ejections. Get

```
29    @NonWhiteHat @MayhewMayhem omg hello tanner you are a scary good boy 12/10 would
...
2324  This is quite the dog. Gets really excited when not in water. Not very soft tho.
2325  This is a southern Vesuvius bumblegruff. Can drive a truck (wow). Made friends wi
2326  Oh goodness. A super rare northeast Qdoba kangaroo mix. Massive feet. No pouch (d
2327  Those are sunglasses and a jean jacket. 11/10 dog cool af https://t.co/uHXrPkUEyl
2328  Unique dog here. Very small. Lives in container of Frosted Flakes (?). Short legs
2329  Here we have a mixed Asiago from the Galápagos Islands. Only one ear working. Big
2330  Look at this jokester thinking seat belt laws don't apply to him. Great tongue th
2331  This is an extremely rare horned Parthenon. Not amused. Wears shoes. Overall very
2332  This is a funny dog. Weird toes. Won't come down. Loves branch. Refuses to eat hi
2333  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood floo
2334  Can take selfies 11/10 https://t.co/ws2AMaNwPW
2335  Very concerned about fellow dog trapped in computer. 10/10 https://t.co/OyxApIikp
2336  Not familiar with this breed. No tail (weird). Only 2 legs. Doesn't bark. Surpris
2337  Oh my. Here you are seeing an Adobe Setter giving birth to twins!!! The world is
2338  Can stand on stump for what seems like a while. Built that birdhouse? Impressive.
2339  This appears to be a Mongolian Presbyterian mix. Very tired. Tongue slip confirme
2340  Here we have a well-established sunblockerspaniel. Lost his other flip-flop. 6/10
2341  Let's hope this flight isn't Malaysian (lol). What a dog! Almost completely camou
2342  Here we have a northern speckled Rhododendron. Much sass. Gives 0 fucks. Good ton
2343  This is the happiest dog you will ever see. Very committed owner. Nice couch. 10/
2344  Here is the Rand Paul of retrievers folks! He's probably good at poker. Can drink
2345  My oh my. This is a rare blond Canadian terrier on wheels. Only $8.98. Rather doc
2346  Here is a Siberian heavily armored polar bear mix. Strong owner. 10/10 I would do
2347  This is an odd dog. Hard on the outside but loving on the inside. Petting still f
2348  This is a truly beautiful English Wilson Staff retriever. Has a nice phone. Privi
2349  Here we have a 1949 1st generation vulpix. Enjoys sweat tea and Fox News. Cannot
2350  This is a purebred Piers Morgan. Loves to Netflix and chill. Always looks like he
2351  Here is a very happy pup. Big fan of well-maintained decks. Just look at that ton
2352  This is a western brown Mitsubishi terrier. Upset about leaf. Actually 2 dogs her
2353  Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big fan of relaxin

      truncated display_text_range  \
0     False      [0, 85]
1     False      [0, 138]
2     False      [0, 121]
3     False      [0, 79]
4     False      [0, 138]
5     False      [0, 138]
6     False      [0, 140]
7     False      [0, 118]
8     False      [0, 122]
9     False      [0, 133]
10    False      [0, 130]
11    False      [0, 107]
12    False      [0, 106]
13    False      [0, 91]
```

```
14    False      [0, 118]
15    False      [0, 138]
16    False      [0, 86]
17    False      [0, 128]
18    False      [0, 87]
19    False      [0, 127]
20    False      [0, 127]
21    False      [0, 108]
22    False      [0, 99]
23    False      [0, 88]
24    False      [0, 129]
25    False      [0, 101]
26    False      [0, 121]
27    False      [0, 71]
28    False      [0, 131]
29    False      [27, 105]
...     ...            ...
2324  False      [0, 140]
2325  False      [0, 139]
2326  False      [0, 137]
2327  False      [0, 81]
2328  False      [0, 140]
2329  False      [0, 135]
2330  False      [0, 112]
2331  False      [0, 139]
2332  False      [0, 138]
2333  False      [0, 136]
2334  False      [0, 46]
2335  False      [0, 82]
2336  False      [0, 134]
2337  False      [0, 128]
2338  False      [0, 140]
2339  False      [0, 132]
2340  False      [0, 125]
2341  False      [0, 137]
2342  False      [0, 137]
2343  False      [0, 107]
2344  False      [0, 135]
2345  False      [0, 124]
2346  False      [0, 140]
2347  False      [0, 138]
2348  False      [0, 140]
2349  False      [0, 120]
2350  False      [0, 137]
2351  False      [0, 130]
2352  False      [0, 139]
2353  False      [0, 131]
```

```
0     {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
1     {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2     {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
3     {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
4     {'hashtags': [{'text': 'BarkWeek', 'indices': [129, 138]}], 'symbols': [], 'user_
5     {'hashtags': [{'text': 'BarkWeek', 'indices': [129, 138]}], 'symbols': [], 'user_
6     {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [{'url': 'https://t.
7     {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
8     {'hashtags': [{'text': 'BarkWeek', 'indices': [113, 122]}], 'symbols': [], 'user_
9     {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
10    {'hashtags': [{'text': 'BarkWeek', 'indices': [121, 130]}], 'symbols': [], 'user_
11    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
12    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
13    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
14    {'hashtags': [{'text': 'BarkWeek', 'indices': [109, 118]}], 'symbols': [], 'user_
15    {'hashtags': [{'text': 'BarkWeek', 'indices': [129, 138]}], 'symbols': [], 'user_
16    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
17    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
18    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
19    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
20    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
21    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
22    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
23    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
24    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
25    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
26    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [{'url': 'https://t.
27    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
28    {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
29    {'hashtags': [], 'symbols': [], 'user_mentions': [{'screen_name': 'NonWhiteHat',
...
2324  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2325  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2326  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2327  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2328  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2329  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2330  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2331  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2332  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2333  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2334  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2335  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2336  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2337  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2338  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2339  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
```

```
2340  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2341  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2342  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2343  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2344  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2345  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2346  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2347  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2348  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2349  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2350  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2351  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2352  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':
2353  {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': [], 'media': [{'id':


0     {'media': [{'id': 892420639486877696, 'id_str': '892420639486877696', 'indices':
1     {'media': [{'id': 892177413194625024, 'id_str': '892177413194625024', 'indices':
2     {'media': [{'id': 891815175371796480, 'id_str': '891815175371796480', 'indices':
3     {'media': [{'id': 891689552724799489, 'id_str': '891689552724799489', 'indices':
4     {'media': [{'id': 891327551943041024, 'id_str': '891327551943041024', 'indices':
5     {'media': [{'id': 891087942176911360, 'id_str': '891087942176911360', 'indices':
6     {'media': [{'id': 890971906207338496, 'id_str': '890971906207338496', 'indices':
7     {'media': [{'id': 890729118844600320, 'id_str': '890729118844600320', 'indices':
8     {'media': [{'id': 890609177319665665, 'id_str': '890609177319665665', 'indices':
9     {'media': [{'id': 890240245463175168, 'id_str': '890240245463175168', 'indices':
10    {'media': [{'id': 890006600089468928, 'id_str': '890006600089468928', 'indices':
11    {'media': [{'id': 889880888800096258, 'id_str': '889880888800096258', 'indices':
12    {'media': [{'id': 889665366129029120, 'id_str': '889665366129029120', 'indices':
13    {'media': [{'id': 889638825424826374, 'id_str': '889638825424826374', 'indices':
14    {'media': [{'id': 889531127467266049, 'id_str': '889531127467266049', 'indices':
15    {'media': [{'id': 889278779352338437, 'id_str': '889278779352338437', 'indices':
16    {'media': [{'id': 888917229776945152, 'id_str': '888917229776945152', 'indices':
17    {'media': [{'id': 888804981515575296, 'id_str': '888804981515575296', 'indices':
18    {'media': [{'id': 888554915546542081, 'id_str': '888554915546542081', 'indices':
19    {'media': [{'id': 888078426338406400, 'id_str': '888078426338406400', 'indices':
20    {'media': [{'id': 887705281597243393, 'id_str': '887705281597243393', 'indices':
21    {'media': [{'id': 887517108413886465, 'id_str': '887517108413886465', 'indices':
22    {'media': [{'id': 887473949361045505, 'id_str': '887473949361045505', 'indices':
23    {'media': [{'id': 887343120832229379, 'id_str': '887343120832229379', 'indices':
24    {'media': [{'id': 887101385971384320, 'id_str': '887101385971384320', 'indices':
25    {'media': [{'id': 886983218871902208, 'id_str': '886983218871902208', 'indices':
26    {'media': [{'id': 886736868116754432, 'id_str': '886736868116754432', 'indices':
27    {'media': [{'id': 886680331239161856, 'id_str': '886680331239161856', 'indices':
28    {'media': [{'id': 886366138128449536, 'id_str': '886366138128449536', 'indices':
29    NaN
...   ...
2324  {'media': [{'id': 666411498068123649, 'id_str': '666411498068123649', 'indices':
```

```
2325  {'media': [{'id': 666407121513275392, 'id_str': '666407121513275392', 'indices':
2326  {'media': [{'id': 666396240351993856, 'id_str': '666396240351993856', 'indices':
2327  {'media': [{'id': 666373746337402880, 'id_str': '666373746337402880', 'indices':
2328  {'media': [{'id': 666362717482020864, 'id_str': '666362717482020864', 'indices':
2329  {'media': [{'id': 666353280906170368, 'id_str': '666353280906170368', 'indices':
2330  {'media': [{'id': 666345414279471104, 'id_str': '666345414279471104', 'indices':
2331  {'media': [{'id': 666337857791987715, 'id_str': '666337857791987715', 'indices':
2332  {'media': [{'id': 666293909010702337, 'id_str': '666293909010702337', 'indices':
2333  {'media': [{'id': 666287399580733440, 'id_str': '666287399580733440', 'indices':
2334  {'media': [{'id': 666273081518768128, 'id_str': '666273081518768128', 'indices':
2335  {'media': [{'id': 666268904428277760, 'id_str': '666268904428277760', 'indices':
2336  {'media': [{'id': 666104129232740352, 'id_str': '666104129232740352', 'indices':
2337  {'media': [{'id': 666102150364286977, 'id_str': '666102150364286977', 'indices':
2338  {'media': [{'id': 666099505364733952, 'id_str': '666099505364733952', 'indices':
2339  {'media': [{'id': 666093996847063040, 'id_str': '666093996847063040', 'indices':
2340  {'media': [{'id': 666082912819875840, 'id_str': '666082912819875840', 'indices':
2341  {'media': [{'id': 666073098362486784, 'id_str': '666073098362486784', 'indices':
2342  {'media': [{'id': 666071190449033216, 'id_str': '666071190449033216', 'indices':
2343  {'media': [{'id': 666063820255862784, 'id_str': '666063820255862784', 'indices':
2344  {'media': [{'id': 666058597072306176, 'id_str': '666058597072306176', 'indices':
2345  {'media': [{'id': 666057085227016192, 'id_str': '666057085227016192', 'indices':
2346  {'media': [{'id': 666055517517848576, 'id_str': '666055517517848576', 'indices':
2347  {'media': [{'id': 666051848592334848, 'id_str': '666051848592334848', 'indices':
2348  {'media': [{'id': 666050754986266625, 'id_str': '666050754986266625', 'indices':
2349  {'media': [{'id': 666049244999131136, 'id_str': '666049244999131136', 'indices':
2350  {'media': [{'id': 666044217047650304, 'id_str': '666044217047650304', 'indices':
2351  {'media': [{'id': 666033409081393153, 'id_str': '666033409081393153', 'indices':
2352  {'media': [{'id': 666029276303482880, 'id_str': '666029276303482880', 'indices':
2353  {'media': [{'id': 666020881337073664, 'id_str': '666020881337073664', 'indices':

                                                                                sourc
0     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
1     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
3     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
6     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
7     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
8     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
9     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
10    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
11    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
12    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
13    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
14    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
15    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
16    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
```

```
17    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
18    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
19    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
20    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
21    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
22    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
23    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
24    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
25    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
26    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
27    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
28    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
29    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
...                                                                                  ..
2324  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2325  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2326  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2327  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2328  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2329  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2330  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2331  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2332  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2333  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2334  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2335  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2336  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2337  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2338  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2339  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2340  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2341  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2342  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2343  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2344  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2345  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2346  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2347  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2348  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2349  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2350  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2351  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2352  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
2353  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a

      in_reply_to_status_id  ...  place  contributors is_quote_status  \
0    NaN                     ...   None  None          False
1    NaN                     ...   None  None          False
```

```
2     NaN                    ...   None  None          False
3     NaN                    ...   None  None          False
4     NaN                    ...   None  None          False
5     NaN                    ...   None  None          False
6     NaN                    ...   None  None          False
7     NaN                    ...   None  None          False
8     NaN                    ...   None  None          False
9     NaN                    ...   None  None          False
10    NaN                    ...   None  None          False
11    NaN                    ...   None  None          False
12    NaN                    ...   None  None          False
13    NaN                    ...   None  None          False
14    NaN                    ...   None  None          False
15    NaN                    ...   None  None          False
16    NaN                    ...   None  None          False
17    NaN                    ...   None  None          False
18    NaN                    ...   None  None          False
19    NaN                    ...   None  None          False
20    NaN                    ...   None  None          False
21    NaN                    ...   None  None          False
22    NaN                    ...   None  None          False
23    NaN                    ...   None  None          False
24    NaN                    ...   None  None          False
25    NaN                    ...   None  None          False
26    NaN                    ...   None  None          False
27    NaN                    ...   None  None          False
28    NaN                    ...   None  None          False
29     8.862664e+17          ...   None  None          False
...             ...           ...    ...   ...            ...
2324  NaN                    ...   None  None          False
2325  NaN                    ...   None  None          False
2326  NaN                    ...   None  None          False
2327  NaN                    ...   None  None          False
2328  NaN                    ...   None  None          False
2329  NaN                    ...   None  None          False
2330  NaN                    ...   None  None          False
2331  NaN                    ...   None  None          False
2332  NaN                    ...   None  None          False
2333  NaN                    ...   None  None          False
2334  NaN                    ...   None  None          False
2335  NaN                    ...   None  None          False
2336  NaN                    ...   None  None          False
2337  NaN                    ...   None  None          False
2338  NaN                    ...   None  None          False
2339  NaN                    ...   None  None          False
2340  NaN                    ...   None  None          False
2341  NaN                    ...   None  None          False
2342  NaN                    ...   None  None          False
```

```
2343 NaN                   ...   None  None          False
2344 NaN                   ...   None  None          False
2345 NaN                   ...   None  None          False
2346 NaN                   ...   None  None          False
2347 NaN                   ...   None  None          False
2348 NaN                   ...   None  None          False
2349 NaN                   ...   None  None          False
2350 NaN                   ...   None  None          False
2351 NaN                   ...   None  None          False
2352 NaN                   ...   None  None          False
2353 NaN                   ...   None  None          False

     retweet_count favorite_count favorited retweeted possibly_sensitive  \
0     8853          39467          False     False     False
1     6514          33819          False     False     False
2     4328          25461          False     False     False
3     8964          42908          False     False     False
4     9774          41048          False     False     False
5     3261          20562          False     False     False
6     2158          12041          False     False     False
7     16716         56848          False     False     False
8     4429          28226          False     False     False
9     7711          32467          False     False     False
10    7624          31166          False     False     False
11    5156          28268          False     False     False
12    8538          38818          False     False     False
13    4735          27672          False     False     False
14    2321          15359          False     False     False
15    5637          25652          False     False     False
16    4709          29611          False     False     False
17    4559          26080          False     False     False
18    3732          20290          False     False     False
19    3653          22201          False     False     False
20    5609          30779          False     False     False
21    12082         46959          True      False     False
22    18781         69871          False     False     False
23    10737         34222          False     False     False
24    6167          31061          False     False     False
25    8084          35859          False     False     False
26    3443          12306          False     False     False
27    4610          22798          False     False     False
28    3316          21524          False     False     False
29    4             117            False     False     NaN
...   ..            ...              ...       ...     ...
2324  339           459            False     False     False
2325  44            113            False     False     False
2326  92            172            False     False     False
2327  100           194            False     False     False
```

```
2328  595           804           False     False     False
2329  77            229           False     False     False
2330  146           307           False     False     False
2331  96            204           False     False     False
2332  368           522           False     False     False
2333  71            152           False     False     False
2334  82            184           False     False     False
2335  37            108           False     False     False
2336  6871          14765         False     False     False
2337  16            81            False     False     False
2338  73            164           False     False     False
2339  79            169           False     False     False
2340  47            121           False     False     False
2341  174           335           False     False     False
2342  67            154           False     False     False
2343  232           496           False     False     False
2344  61            115           False     False     False
2345  146           304           False     False     False
2346  261           448           False     False     False
2347  879           1253          False     False     False
2348  60            136           False     False     False
2349  41            111           False     False     False
2350  147           311           False     False     False
2351  47            128           False     False     False
2352  48            132           False     False     False
2353  532           2535          False     False     False

     possibly_sensitive_appealable  lang
0     False                         en
1     False                         en
2     False                         en
3     False                         en
4     False                         en
5     False                         en
6     False                         en
7     False                         en
8     False                         en
9     False                         en
10    False                         en
11    False                         en
12    False                         en
13    False                         en
14    False                         en
15    False                         en
16    False                         en
17    False                         en
18    False                         en
19    False                         en
```

```
20    False                         en
21    False                         en
22    False                         en
23    False                         en
24    False                         en
25    False                         en
26    False                         en
27    False                         en
28    False                         en
29    NaN                           en
...   ...                           ..
2324  False                         en
2325  False                         en
2326  False                         en
2327  False                         en
2328  False                         en
2329  False                         en
2330  False                         en
2331  False                         en
2332  False                         en
2333  False                         en
2334  False                         en
2335  False                         en
2336  False                         en
2337  False                         en
2338  False                         en
2339  False                         en
2340  False                         en
2341  False                         en
2342  False                         en
2343  False                         en
2344  False                         en
2345  False                         en
2346  False                         en
2347  False                         en
2348  False                         en
2349  False                         en
2350  False                         en
2351  False                         en
2352  False                         en
2353  False                         en

[2354 rows x 27 columns]
```

```
In [20]: df_tweets_raw.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
```

```
Data columns (total 27 columns):
created_at                          2354 non-null object
id                                  2354 non-null int64
id_str                              2354 non-null object
full_text                           2354 non-null object
truncated                           2354 non-null bool
display_text_range                  2354 non-null object
entities                            2354 non-null object
extended_entities                   2073 non-null object
source                              2354 non-null object
in_reply_to_status_id               78 non-null float64
in_reply_to_status_id_str           78 non-null object
in_reply_to_user_id                 78 non-null float64
in_reply_to_user_id_str             78 non-null object
in_reply_to_screen_name             78 non-null object
user                                2354 non-null object
geo                                 0 non-null object
coordinates                         0 non-null object
place                               1 non-null object
contributors                        0 non-null object
is_quote_status                     2354 non-null bool
retweet_count                       2354 non-null int64
favorite_count                      2354 non-null int64
favorited                           2354 non-null bool
retweeted                           2354 non-null bool
possibly_sensitive                  2211 non-null object
possibly_sensitive_appealable       2211 non-null object
lang                                2354 non-null object
dtypes: bool(4), float64(2), int64(3), object(18)
memory usage: 432.3+ KB
```

```
In [21]: sum(df_tweets_raw['id'].duplicated())
```

```
Out[21]: 0
```

- change the column name from id to tweet_id
- Set column (tweet_id) as index when merging dataframes
- keep only the required columns ('id','retweet_count','favorite_count')
- two entries less than those in df_twitter_raw

1.2.3 1.2.3. Findings:

1.2.3.1. Quality Issues

1.2.3.1.1. df_twitter_raw

- some tweets are replies and retweets
- Some tweets are not with expanded_urls (no Image)

- timestamp column dtype not correct
- tweet_id column dtype not correct
- columns not required in master sheet (in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp, expanded_urls)
- dog stages columns contain 'None' as values
- rating_numerator contains wrong inputs
- rating_denominator contains wrong inputs
- source of tweeting included in the source url

1.2.3.1.2. df_images_raw

- There are entries in df_twitter_raw without images data
- columns labels not expressive
- tweet_id column dtype not correct
- images breed predection and if it's dog (distributed over 9 columns)
- identify the Non-dog images for tweet ids
- Set column (tweet_id) as index when merging dataframes

1.2.3.1.3. df_tweets_raw

- change the column name from id to tweet_id
- Set column (tweet_id) as index when merging dataframes
- keep only the required columns ('id','retweet_count','favorite_count')
- two entries less than those in df_twitter_raw

1.2.3.2. Tidiness Issues

- columns not required in master sheet (in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp, expanded_urls)
- Set column (tweet_id) as index when merging dataframes
- arrange dataframe by timestamp

1.2.4 1.3. Cleaning Data:

1.2.5 1.3.1. Quality Issues

1.3.1.1. df_twitter_raw

- Drop rows of replies and retweets

```
In [22]: df_twitter_clean = df_twitter_raw[df_twitter_raw['in_reply_to_status_id'].isnull()]
```

```
In [23]: df_twitter_clean = df_twitter_clean[df_twitter_clean['retweeted_status_id'].isnull()]
```

```
In [24]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2097 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                      2097 non-null int64
in_reply_to_status_id         0 non-null float64
in_reply_to_user_id           0 non-null float64
timestamp                     2097 non-null object
source                        2097 non-null object
text                          2097 non-null object
retweeted_status_id           0 non-null float64
retweeted_status_user_id      0 non-null float64
retweeted_status_timestamp    0 non-null object
expanded_urls                 2094 non-null object
rating_numerator              2097 non-null int64
rating_denominator            2097 non-null int64
name                          2097 non-null object
doggo                         2097 non-null object
floofer                       2097 non-null object
pupper                        2097 non-null object
puppo                         2097 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 294.9+ KB
```

- Drop rows of tweets with no expanded_urls (no Image)

```
In [25]: df_twitter_clean = df_twitter_clean[df_twitter_clean['expanded_urls'].notnull()]
```

```
In [26]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2094 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                      2094 non-null int64
in_reply_to_status_id         0 non-null float64
in_reply_to_user_id           0 non-null float64
timestamp                     2094 non-null object
source                        2094 non-null object
text                          2094 non-null object
retweeted_status_id           0 non-null float64
retweeted_status_user_id      0 non-null float64
retweeted_status_timestamp    0 non-null object
expanded_urls                 2094 non-null object
rating_numerator              2094 non-null int64
rating_denominator            2094 non-null int64
name                          2094 non-null object
doggo                         2094 non-null object
floofer                       2094 non-null object
pupper                        2094 non-null object
```

```
puppo                         2094 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 294.5+ KB
```

- Modify timestamp column dtype
- Modify tweet_id column dtype

```
In [27]: df_twitter_clean['timestamp']= pd.to_datetime(df_twitter_clean['timestamp'])
         df_twitter_clean['tweet_id'] = df_twitter_clean['tweet_id'].astype(str)

In [28]: df_twitter_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2094 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                      2094 non-null object
in_reply_to_status_id         0 non-null float64
in_reply_to_user_id           0 non-null float64
timestamp                     2094 non-null datetime64[ns]
source                        2094 non-null object
text                          2094 non-null object
retweeted_status_id           0 non-null float64
retweeted_status_user_id      0 non-null float64
retweeted_status_timestamp    0 non-null object
expanded_urls                 2094 non-null object
rating_numerator              2094 non-null int64
rating_denominator            2094 non-null int64
name                          2094 non-null object
doggo                         2094 non-null object
floofer                       2094 non-null object
pupper                        2094 non-null object
puppo                         2094 non-null object
dtypes: datetime64[ns](1), float64(4), int64(2), object(10)
memory usage: 294.5+ KB
```

- columns not required in master sheet (in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp, expanded_urls)

```
In [29]: df_twitter_clean = df_twitter_clean.drop(columns = ['in_reply_to_status_id', 'in_reply_

In [30]: df_twitter_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2094 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id              2094 non-null object
```

```
timestamp             2094 non-null datetime64[ns]
source                2094 non-null object
text                  2094 non-null object
expanded_urls         2094 non-null object
rating_numerator      2094 non-null int64
rating_denominator    2094 non-null int64
name                  2094 non-null object
doggo                 2094 non-null object
floofer               2094 non-null object
pupper                2094 non-null object
puppo                 2094 non-null object
dtypes: datetime64[ns](1), int64(2), object(9)
memory usage: 212.7+ KB
```

```
In [31]: df_twitter_clean['doggo'].value_counts()

Out[31]: None     2011
         doggo    83
         Name: doggo, dtype: int64

In [32]: df_twitter_clean['floofer'].value_counts()

Out[32]: None       2084
         floofer    10
         Name: floofer, dtype: int64

In [33]: df_twitter_clean['pupper'].value_counts()

Out[33]: None      1865
         pupper    229
         Name: pupper, dtype: int64

In [34]: df_twitter_clean['puppo'].value_counts()

Out[34]: None     2070
         puppo    24
         Name: puppo, dtype: int64
```

- convert 'None' in dog stages columns to Nan
- combine the 4 dog stages colums into a 'dog_stages' column

```
In [35]: df_twitter_clean.iloc[:, -4:]= df_twitter_clean.iloc[:,-4:].replace('None', '')

In [36]: df_twitter_clean['dog_stage']= df_twitter_clean['doggo'] + df_twitter_clean['floofer']

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2094 entries, 0 to 2355
Data columns (total 13 columns):
tweet_id              2094 non-null object
timestamp             2094 non-null datetime64[ns]
source                2094 non-null object
text                  2094 non-null object
expanded_urls         2094 non-null object
rating_numerator      2094 non-null int64
rating_denominator    2094 non-null int64
name                  2094 non-null object
doggo                 2094 non-null object
floofer               2094 non-null object
pupper                2094 non-null object
puppo                 2094 non-null object
dog_stage             2094 non-null object
dtypes: datetime64[ns](1), int64(2), object(10)
memory usage: 229.0+ KB
```

```
In [38]: df_twitter_clean['dog_stage'].value_counts()

Out[38]:                1759
         pupper          220
         doggo            72
         puppo            23
         doggopupper       9
         floofer           9
         doggopuppo        1
         doggofloofer      1
         Name: dog_stage, dtype: int64

In [39]: df_twitter_clean['dog_stage'].replace({'doggopupper':'doggo-pupper','doggopuppo':'doggo

In [40]: df_twitter_clean['dog_stage'].value_counts()

Out[40]:                 1759
         pupper           220
         doggo             72
         puppo             23
         floofer            9
         doggo-pupper       9
         doggo-puppo        1
         doggo-floofer      1
         Name: dog_stage, dtype: int64

In [41]: df_twitter_clean.loc[df_twitter_clean.dog_stage == '', 'dog_stage'] = np.nan

In [42]: df_twitter_clean['dog_stage'].value_counts()
```

```
Out[42]: pupper           220
         doggo            72
         puppo            23
         floofer          9
         doggo-pupper     9
         doggo-puppo      1
         doggo-floofer    1
         Name: dog_stage, dtype: int64

In [43]: df_twitter_clean = df_twitter_clean.drop(columns = ['doggo', 'floofer', 'pupper','puppo

In [44]: df_twitter_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2094 entries, 0 to 2355
Data columns (total 9 columns):
tweet_id              2094 non-null object
timestamp             2094 non-null datetime64[ns]
source                2094 non-null object
text                  2094 non-null object
expanded_urls         2094 non-null object
rating_numerator      2094 non-null int64
rating_denominator    2094 non-null int64
name                  2094 non-null object
dog_stage             335 non-null object
dtypes: datetime64[ns](1), int64(2), object(6)
memory usage: 163.6+ KB
```

- extract the tweets source from column 'source'

```
In [45]: df_twitter_clean.source[:3]

Out[45]: 0    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         1    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         2    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         Name: source, dtype: object

In [46]: df_twitter_clean['source']=df_twitter_clean.source.apply(lambda x:re.findall (r'>(.*)<'

In [47]: df_twitter_clean['source'].value_counts()

Out[47]: Twitter for iPhone     1962
         Vine - Make a Scene    91
         Twitter Web Client     30
         TweetDeck              11
         Name: source, dtype: int64
```

- correct rating_denominator values by reviewing text

- correct rating_numerator values by reviewing text

```
In [48]: df_twitter_clean.describe()
```

```
Out[48]:        rating_numerator  rating_denominator
         count  2094.000000       2094.000000
         mean   12.191500         10.449379
         std    40.393858         6.649800
         min    0.000000          2.000000
         25%    10.000000         10.000000
         50%    11.000000         10.000000
         75%    12.000000         10.000000
         max    1776.000000       170.000000
```

```
In [49]: df_twitter_clean['rating_denominator'].value_counts()
```

```
Out[49]: 10     2077
         50     3
         11     2
         80     2
         7      1
         170    1
         150    1
         120    1
         110    1
         90     1
         70     1
         40     1
         20     1
         2      1
         Name: rating_denominator, dtype: int64
```

```
In [50]: df_twitter_clean['rating_numerator'].value_counts()
```

```
Out[50]: 12      485
         10      434
         11      413
         13      287
         9       153
         8       98
         7       52
         14      38
         5       34
         6       32
         3       19
         4       16
         2       9
         1       5
         26      1
```

```
44      1
165     1
24      1
60      1
50      1
144     1
80      1
84      1
88      1
121     1
204     1
420     1
1776    1
27      1
45      1
75      1
99      1
0       1
Name: rating_numerator, dtype: int64
```

In [51]:
```
df_twitter_clean[df_twitter_clean['rating_numerator'] > 20][['text','rating_numerator',
```

Out[51]:

```
433   The floofs have been released I repeat the floofs have been released. 84/70 https
516   Meet Sam. She smiles 24/7 &amp;amp; secretly aspires to be a reindeer. \nKeep Sam smi
695   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
763   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
902   Why does this never happen at my front door... 165/150 https://t.co/HmwrdfEfUE
979   This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkh
1120  Say hello to this unbelievably well behaved squad of doggos. 204/170 would try to
1202  This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. Amaz
1228  Happy Saturday here's 9 puppers on a bench. 99/90 good work everybody https://t.c
1254  Here's a brigade of puppers. All look very prepared for whatever happens next. 80
1274  From left to right:\nCletus, Jerome, Alejandro, Burp, &amp;amp; Titson\nNone know whe
1351  Here is a whole flock of puppers.  60/50 I'll take the lot https://t.co/9dpcw6MdW
1433  Happy Wednesday here's a bucket of pups. 44/40 would pet all at once https://t.co
1635  Someone help the girl is being mugged. Several are distracting her while two stea
1712  Here we have uncovered an entire battalion of holiday puppers. Average of 11.26/1
1779  IT'S PUPPERGEDDON. Total of 144/120 ...I think https://t.co/ZanVtAtvIq
1843  Here we have an entire platoon of puppers. Total score: 88/80 would pet all at on
2074  After so many requests... here you go.\n\nGood dogg. 420/10 https://t.co/yfAAo1gd
```

	rating_numerator	rating_denominator
433	84	70
516	24	7
695	75	10
763	27	10
902	165	150

```
979   1776              10
1120  204               170
1202  50                50
1228  99                90
1254  80                80
1274  45                50
1351  60                50
1433  44                40
1635  121               110
1712  26                10
1779  144               120
1843  88                80
2074  420               10
```

In [52]:
```
df_twitter_clean['rating_numerator'] = df_twitter_clean.text.str.extract('(\d+\.?\d?\d?
```

In [53]:
```
df_twitter_clean[df_twitter_clean['rating_numerator'] > 20][['text','rating_numerator',
```

Out[53]:

```
433   The floofs have been released I repeat the floofs have been released. 84/70 https
516   Meet Sam. She smiles 24/7 &amp;amp; secretly aspires to be a reindeer. \nKeep Sam smi
902   Why does this never happen at my front door... 165/150 https://t.co/HmwrdfEfUE
979   This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkh
1120  Say hello to this unbelievably well behaved squad of doggos. 204/170 would try to
1202  This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. Amaz
1228  Happy Saturday here's 9 puppers on a bench. 99/90 good work everybody https://t.c
1254  Here's a brigade of puppers. All look very prepared for whatever happens next. 80
1274  From left to right:\nCletus, Jerome, Alejandro, Burp, &amp;amp; Titson\nNone know whe
1351  Here is a whole flock of puppers.  60/50 I'll take the lot https://t.co/9dpcw6MdW
1433  Happy Wednesday here's a bucket of pups. 44/40 would pet all at once https://t.co
1635  Someone help the girl is being mugged. Several are distracting her while two stea
1779  IT'S PUPPERGEDDON. Total of 144/120 ...I think https://t.co/ZanVtAtvIq
1843  Here we have an entire platoon of puppers. Total score: 88/80 would pet all at on
2074  After so many requests... here you go.\n\nGood dogg. 420/10 https://t.co/yfAAo1gd

      rating_numerator  rating_denominator
433   84.0              70
516   24.0              7
902   165.0             150
979   1776.0            10
1120  204.0             170
1202  50.0              50
1228  99.0              90
1254  80.0              80
1274  45.0              50
1351  60.0              50
1433  44.0              40
1635  121.0             110
```

```
1779  144.0             120
1843  88.0              80
2074  420.0             10
```

```
In [54]: df_twitter_clean[(df_twitter_clean['rating_denominator'] != 10) & (df_twitter_clean['ra
```

Out[54]:

```
433   The floofs have been released I repeat the floofs have been released. 84/70 https
902   Why does this never happen at my front door... 165/150 https://t.co/HmwrdfEfUE
1120  Say hello to this unbelievably well behaved squad of doggos. 204/170 would try to
1165  Happy 4/20 from the squad! 13/10 for all https://t.co/eV1diwds8a
1202  This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. Amaz
1228  Happy Saturday here's 9 puppers on a bench. 99/90 good work everybody https://t.c
1254  Here's a brigade of puppers. All look very prepared for whatever happens next. 80
1274  From left to right:\nCletus, Jerome, Alejandro, Burp, &amp;amp; Titson\nNone know whe
1351  Here is a whole flock of puppers.  60/50 I'll take the lot https://t.co/9dpcw6MdW
1433  Happy Wednesday here's a bucket of pups. 44/40 would pet all at once https://t.co
1635  Someone help the girl is being mugged. Several are distracting her while two stea
1779  IT'S PUPPERGEDDON. Total of 144/120 ...I think https://t.co/ZanVtAtvIq
1843  Here we have an entire platoon of puppers. Total score: 88/80 would pet all at on

      rating_numerator  rating_denominator
433   84.0              70
902   165.0             150
1120  204.0             170
1165  4.0               20
1202  50.0              50
1228  99.0              90
1254  80.0              80
1274  45.0              50
1351  60.0              50
1433  44.0              40
1635  121.0             110
1779  144.0             120
1843  88.0              80
```

```
In [55]: df_twitter_clean.loc[df_twitter_clean.index == 1202, ['rating_numerator','rating_denomi
```

```
In [56]: df_twitter_clean.loc[df_twitter_clean.index == 1165, ['rating_numerator','rating_denomi
```

```
In [57]: df_twitter_clean[(df_twitter_clean['rating_denominator'] != 10) & (df_twitter_clean['ra
```

Out[57]:

```
                tweet_id           timestamp              source  \
433   820690176645140481 2017-01-15 17:52:40  Twitter for iPhone
902   758467244762497024 2016-07-28 01:00:57  Twitter for iPhone
1120  731156023742988288 2016-05-13 16:15:54  Twitter for iPhone
1228  713900603437621249 2016-03-27 01:29:02  Twitter for iPhone
1254  710658690886586372 2016-03-18 02:46:49  Twitter for iPhone
1274  709198395643068416 2016-03-14 02:04:08  Twitter for iPhone
```

```
1351  704054845121142784 2016-02-28 21:25:30  Twitter for iPhone
1433  697463031882764288 2016-02-10 16:51:59  Twitter for iPhone
1635  684222868335505415 2016-01-05 04:00:18  Twitter for iPhone
1779  677716515794329600 2015-12-18 05:06:23  Twitter for iPhone
1843  675853064436391936 2015-12-13 01:41:41  Twitter for iPhone


433   The floofs have been released I repeat the floofs have been released. 84/70 https
902   Why does this never happen at my front door... 165/150 https://t.co/HmwrdfEfUE
1120  Say hello to this unbelievably well behaved squad of doggos. 204/170 would try to
1228  Happy Saturday here's 9 puppers on a bench. 99/90 good work everybody https://t.c
1254  Here's a brigade of puppers. All look very prepared for whatever happens next. 80
1274  From left to right:\nCletus, Jerome, Alejandro, Burp, &amp;amp; Titson\nNone know whe
1351  Here is a whole flock of puppers.  60/50 I'll take the lot https://t.co/9dpcw6MdW
1433  Happy Wednesday here's a bucket of pups. 44/40 would pet all at once https://t.co
1635  Someone help the girl is being mugged. Several are distracting her while two stea
1779  IT'S PUPPERGEDDON. Total of 144/120 ...I think https://t.co/ZanVtAtvIq
1843  Here we have an entire platoon of puppers. Total score: 88/80 would pet all at on


433   https://twitter.com/dog_rates/status/820690176645140481/photo/1,https://twitter.c
902   https://twitter.com/dog_rates/status/758467244762497024/video/1
1120  https://twitter.com/dog_rates/status/731156023742988288/photo/1
1228  https://twitter.com/dog_rates/status/713900603437621249/photo/1
1254  https://twitter.com/dog_rates/status/710658690886586372/photo/1
1274  https://twitter.com/dog_rates/status/709198395643068416/photo/1
1351  https://twitter.com/dog_rates/status/704054845121142784/photo/1
1433  https://twitter.com/dog_rates/status/697463031882764288/photo/1
1635  https://twitter.com/dog_rates/status/684222868335505415/photo/1
1779  https://twitter.com/dog_rates/status/677716515794329600/photo/1
1843  https://twitter.com/dog_rates/status/675853064436391936/photo/1,https://twitter.c

      rating_numerator  rating_denominator  name dog_stage
433   84.0              70                  None  NaN
902   165.0             150                 None  NaN
1120  204.0             170                 this  NaN
1228  99.0              90                  None  NaN
1254  80.0              80                  None  NaN
1274  45.0              50                  None  NaN
1351  60.0              50                  a     NaN
1433  44.0              40                  None  NaN
1635  121.0             110                 None  NaN
1779  144.0             120                 None  NaN
1843  88.0              80                  None  NaN
```

In [58]:
```
dogs_count = df_twitter_clean.rating_denominator[df_twitter_clean['rating_denominator']
```

In [59]:
```
dogs_count.value_counts()
```

```
Out[59]: 1.0     2079
         5.0     2
         8.0     2
         12.0    1
         11.0    1
         4.0     1
         9.0     1
         17.0    1
         15.0    1
         7.0     1
         Name: rating_denominator, dtype: int64

In [60]: df_twitter_clean.loc[df_twitter_clean['rating_denominator'] %10 ==0, ['rating_numerator

In [61]: df_twitter_clean[df_twitter_clean['rating_denominator'] != 10]

Out[61]:                tweet_id           timestamp              source  \
         516   810984652412424192 2016-12-19 23:06:23  Twitter for iPhone
         1068  740373189193256964 2016-06-08 02:41:38  Twitter for iPhone
         1662  682962037429899265 2016-01-01 16:30:13  Twitter for iPhone
         2335  666287406224695296 2015-11-16 16:11:11  Twitter for iPhone


         516   Meet Sam. She smiles 24/7 &amp;amp; secretly aspires to be a reindeer. \nKeep Sam smi
         1068  After so many requests, this is Bretagne. She was the last surviving 9/11 search
         1662  This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was ju
         2335  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood floo


         516   https://www.gofundme.com/sams-smile,https://twitter.com/dog_rates/status/8109846
         1068  https://twitter.com/dog_rates/status/740373189193256964/photo/1,https://twitter.c
         1662  https://twitter.com/dog_rates/status/682962037429899265/photo/1
         2335  https://twitter.com/dog_rates/status/666287406224695296/photo/1

               rating_numerator  rating_denominator    name dog_stage
         516   24.0              7                   Sam     NaN
         1068  9.0               11                  None    NaN
         1662  7.0               11                  Darrel  NaN
         2335  1.0               2                   an      NaN

In [62]: df_twitter_clean.loc[df_twitter_clean.index == 516, ['rating_numerator','rating_denomin

In [63]: df_twitter_clean[df_twitter_clean['rating_denominator'] != 10]

Out[63]:                tweet_id           timestamp              source  \
         516   810984652412424192 2016-12-19 23:06:23  Twitter for iPhone
         1068  740373189193256964 2016-06-08 02:41:38  Twitter for iPhone
         1662  682962037429899265 2016-01-01 16:30:13  Twitter for iPhone
         2335  666287406224695296 2015-11-16 16:11:11  Twitter for iPhone
```

```
516   Meet Sam. She smiles 24/7 &amp;amp; secretly aspires to be a reindeer. \nKeep Sam smi
1068  After so many requests, this is Bretagne. She was the last surviving 9/11 search
1662  This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was ju
2335  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood floo


516   https://www.gofundme.com/sams-smile,https://twitter.com/dog_rates/status/81098465
1068  https://twitter.com/dog_rates/status/740373189193256964/photo/1,https://twitter.c
1662  https://twitter.com/dog_rates/status/682962037429899265/photo/1
2335  https://twitter.com/dog_rates/status/666287406224695296/photo/1

      rating_numerator  rating_denominator    name dog_stage
516   NaN               NaN                 Sam     NaN
1068   9.0               11.0               None    NaN
1662   7.0               11.0               Darrel  NaN
2335   1.0               2.0                an      NaN
```

```
In [64]: df_twitter_clean.loc[[1165,1068,2335], ['rating_numerator','rating_denominator']] = [[1
         df_twitter_clean.loc[[1165,1068,2335], ['rating_numerator','rating_denominator']]
```

```
Out[64]:      rating_numerator  rating_denominator
         1165  14.0              10.0
         1068  10.0              10.0
         2335  9.0               10.0
```

- save Outcome

```
In [65]: df_twitter_clean.to_csv('twitter_clean.csv', index=False)
```

1.3.1.2. df_images_raw

- Change tweet_id column dtype to str

```
In [66]: df_images_clean = df_images_raw.copy()
```

```
In [67]: df_images_clean['tweet_id'] = df_images_clean['tweet_id'].astype(str)
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id    2075 non-null object
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(1), object(5)
memory usage: 152.1+ KB
```

- Change the columns labels to be expressive

```
In [69]: cols = ['tweet_id', 'jpg_url', 'img_num',
                 'prediction_1', 'confidence_1', 'breed_1',
                 'prediction_2', 'confidence_2', 'breed_2',
                 'prediction_3', 'confidence_3', 'breed_3']
         df_images_clean.columns = cols
```

- identify the most propable breed for tweet ids
- identify the Non-dog images for tweet ids

```
In [70]: df_images_clean[df_images_clean['breed_1'] == False].prediction_1.value_counts()
```

```
Out[70]: seat_belt              22
         web_site               19
         teddy                  18
         tennis_ball            9
         dingo                  9
         doormat                8
         hamster                7
         Siamese_cat            7
         bath_towel             7
         swing                  7
         tub                    7
         ice_bear               6
         car_mirror             6
         home_theater           6
         llama                  6
         porcupine              5
         minivan                5
         hippopotamus           5
         shopping_cart          5
         ox                     5
         bathtub                4
         bow_tie                4
         jigsaw_puzzle          4
         Arctic_fox             4
         patio                  4
```

```
hog                    4
barrow                 4
guinea_pig             4
brown_bear             4
goose                  4
                      ..
canoe                  1
stove                  1
desktop_computer       1
panpipe                1
otter                  1
syringe                1
radio_telescope        1
grille                 1
rain_barrel            1
lawn_mower             1
mortarboard            1
bookshop               1
sunglasses             1
coffee_mug             1
teapot                 1
harp                   1
dhole                  1
studio_couch           1
grey_fox               1
African_hunting_dog    1
maillot                1
tiger_shark            1
bookcase               1
candle                 1
platypus               1
shopping_basket        1
revolver               1
cougar                 1
African_grey           1
espresso               1
Name: prediction_1, Length: 267, dtype: int64
```

```
In [71]: df_images_clean[df_images_clean['breed_2'] == False].prediction_1.value_counts()
```

```
Out[71]: Chihuahua                18
         web_site                 14
         chow                     12
         teddy                    11
         Labrador_retriever       11
         pug                      10
         Samoyed                   8
         tub                       7
```

```
Pembroke                    6
home_theater                6
toy_poodle                  6
minivan                     5
doormat                     5
bath_towel                  5
porcupine                   5
Pomeranian                  4
kelpie                      4
hamster                     4
hippopotamus                4
Chesapeake_Bay_retriever    4
guinea_pig                  4
jigsaw_puzzle               4
patio                       4
llama                       4
car_mirror                  4
bow_tie                     4
Maltese_dog                 3
wood_rabbit                 3
washbasin                   3
seat_belt                   3
                           ..
walking_stick               1
Madagascar_cat              1
alp                         1
lawn_mower                  1
hammer                      1
mortarboard                 1
water_bottle                1
padlock                     1
sunglasses                  1
coffee_mug                  1
teapot                      1
harp                        1
paper_towel                 1
studio_couch                1
African_hunting_dog         1
maillot                     1
tiger_shark                 1
panpipe                     1
bookcase                    1
candle                      1
otter                       1
toyshop                     1
coil                        1
shopping_basket             1
revolver                    1
```

```
African_grey                1
grey_fox                    1
Eskimo_dog                  1
syringe                     1
tabby                       1
Name: prediction_1, Length: 284, dtype: int64
```

```
In [72]: df_images_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id        2075 non-null object
jpg_url         2075 non-null object
img_num         2075 non-null int64
prediction_1    2075 non-null object
confidence_1    2075 non-null float64
breed_1         2075 non-null bool
prediction_2    2075 non-null object
confidence_2    2075 non-null float64
breed_2         2075 non-null bool
prediction_3    2075 non-null object
confidence_3    2075 non-null float64
breed_3         2075 non-null bool
dtypes: bool(3), float64(3), int64(1), object(5)
memory usage: 152.1+ KB
```

```
In [73]: df_images_clean['dog_breed'] = np.nan
         for i, row in df_images_clean.iterrows():
             if row['breed_1'] == True:
                 df_images_clean.loc[i, ['dog_breed']]= row['prediction_1']
             elif row['breed_2'] == True:
                 df_images_clean.loc[i, ['dog_breed']]= row['prediction_2']
             elif row['breed_3'] == True:
                 df_images_clean.loc[i, ['dog_breed']]= row['prediction_3']
             else:
                 df_images_clean.loc[i, ['dog_breed']]= 'Not Dog'
```

```
In [74]: df_images_clean
```

```
Out[74]:                tweet_id  \
         0     666020888022790149
         1     666029285002620928
         2     666033412701032449
         3     666044226329800704
         4     666049248165822465
         5     666050758794694657
         6     666051853826850816
```

7	666055525042405380
8	666057090499244032
9	666058600524156928
10	666063827256086533
11	666071193221509120
12	666073100786774016
13	666082916733198337
14	666094000022159362
15	666099513787052032
16	666102155909144576
17	666104133288665088
18	666268910803644416
19	666273097616637952
20	666287406224695296
21	666293911632134144
22	666337882303524864
23	666345417576210432
24	666353288456101888
25	666362758909284353
26	666373753744588802
27	666396247373291520
28	666407126856765440
29	666411507551481857
...	...
2045	886366144734445568
2046	886680336477933568
2047	886736880519319552
2048	886983233522544640
2049	887101392804085760
2050	887343217045368832
2051	887473957103951883
2052	887517139158093824
2053	887705289381826560
2054	888078434458587136
2055	888202515573088257
2056	888554962724278272
2057	888804989199671297
2058	888917238123831296
2059	889278841981685760
2060	889531135344209921
2061	889638837579907072
2062	889665388333682689
2063	889880896479866881
2064	890006608113172480
2065	890240255349198849
2066	890609185150312448
2067	890729181411237888
2068	890971913173991426

2069	891087950875897856
2070	891327558926688256
2071	891689557279858688
2072	891815181378084864
2073	892177421306343426
2074	892420643555336193

		j
0	https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg	
1	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	
2	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg	
3	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg	
4	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg	
5	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg	
6	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg	
7	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg	
8	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg	
9	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg	
10	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg	
11	https://pbs.twimg.com/media/CT5cN_3WEAAlOoZ.jpg	
12	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg	
13	https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.jpg	
14	https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg	
15	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg	
16	https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg	
17	https://pbs.twimg.com/media/CT56LSZWoAAlJj2.jpg	
18	https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg	
19	https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg	
20	https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg	
21	https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg	
22	https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg	
23	https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg	
24	https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg	
25	https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg	
26	https://pbs.twimg.com/media/CT9vZEYWUAAlZO5.jpg	
27	https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg	
28	https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg	
29	https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg	
...	...	
2045	https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg	
2046	https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg	
2047	https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg	
2048	https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.jpg	
2049	https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg	
2050	https://pbs.twimg.com/ext_tw_video_thumb/887343120832229379/pu/img/6HSuFrW1lzI_9M	
2051	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg	
2052	https://pbs.twimg.com/ext_tw_video_thumb/887517108413886465/pu/img/WanJKwssZj4VJv	
2053	https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg	

```
2054  https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
2055  https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056  https://pbs.twimg.com/media/DFTH_O-UQAACu2O.jpg
2057  https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg
2059  https://pbs.twimg.com/ext_tw_video_thumb/889278779352338437/pu/img/VlbFB3v8H8VwzV
2060  https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg
2061  https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062  https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063  https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg
2064  https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065  https://pbs.twimg.com/media/DFrEyVuWOAAO3t9.jpg
2066  https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067  https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068  https://pbs.twimg.com/media/DF1eOmZXUAALUcq.jpg
2069  https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

      img_num                 prediction_1  confidence_1  breed_1  \
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

                 prediction_2  confidence_2  breed_2  \
0     collie                   0.156665      True
1     miniature_pinscher       0.074192      True
2     malinois                 0.138584      True
3     redbone                  0.360687      True
4     Rottweiler               0.243682      True
5     English_springer         0.263788      True
6     mud_turtle               0.045885      False
7     Tibetan_mastiff          0.058279      True
8     shopping_basket          0.014594      False
9     komondor                 0.192305      True
```

10	Tibetan_mastiff	0.093718	True
11	Yorkshire_terrier	0.174201	True
12	English_foxhound	0.175382	True
13	bull_mastiff	0.404722	True
14	German_shepherd	0.078260	True
15	Shih-Tzu	0.166192	True
16	Newfoundland	0.149842	True
17	cock	0.033919	False
18	desk	0.085547	False
19	toy_terrier	0.111884	True
20	toy_poodle	0.063064	True
21	otter	0.015250	False
22	Newfoundland	0.278407	True
23	Chesapeake_Bay_retriever	0.054787	True
24	Siberian_husky	0.147655	True
25	skunk	0.002402	False
26	Afghan_hound	0.259551	True
27	toy_terrier	0.009397	True
28	bloodhound	0.244220	True
29	barracouta	0.271485	False
...	...	...	...
2045	Chihuahua	0.000361	True
2046	sports_car	0.139952	False
2047	Great_Pyrenees	0.186136	True
2048	toy_terrier	0.143528	True
2049	Eskimo_dog	0.035029	True
2050	sea_lion	0.275645	False
2051	Rhodesian_ridgeback	0.054950	True
2052	tow_truck	0.029175	False
2053	redbone	0.087582	True
2054	pug	0.000932	True
2055	Rhodesian_ridgeback	0.054950	True
2056	Eskimo_dog	0.166511	True
2057	Labrador_retriever	0.184172	True
2058	Tibetan_mastiff	0.120184	True
2059	borzoi	0.194742	True
2060	Labrador_retriever	0.013834	True
2061	boxer	0.002129	True
2062	Cardigan	0.027356	True
2063	Labrador_retriever	0.151317	True
2064	Pomeranian	0.013884	True
2065	Cardigan	0.451038	True
2066	Irish_setter	0.193054	True
2067	Eskimo_dog	0.178406	True
2068	Border_collie	0.199287	True
2069	Irish_terrier	0.116317	True
2070	English_springer	0.225770	True
2071	Labrador_retriever	0.168086	True

```
2072  malamute                  0.078253      True
2073  Pekinese                  0.090647      True
2074  bagel                     0.085851      False

                    prediction_3  confidence_3  breed_3  \
0     Shetland_sheepdog             0.061428      True
1     Rhodesian_ridgeback           0.072010      True
2     bloodhound                    0.116197      True
3     miniature_pinscher            0.222752      True
4     Doberman                      0.154629      True
5     Greater_Swiss_Mountain_dog    0.016199      True
6     terrapin                      0.017885      False
7     fur_coat                      0.054449      False
8     golden_retriever              0.007959      True
9     soft-coated_wheaten_terrier   0.082086      True
10    Labrador_retriever            0.072427      True
11    Pekinese                      0.109454      True
12    Ibizan_hound                  0.097471      True
13    French_bulldog                0.048960      True
14    malinois                      0.075628      True
15    Dandie_Dinmont                0.089688      True
16    borzoi                        0.133649      True
17    partridge                     0.000052      False
18    bookcase                      0.079480      False
19    basenji                       0.111152      True
20    miniature_poodle              0.025581      True
21    great_grey_owl                0.013207      False
22    groenendael                   0.102643      True
23    Labrador_retriever            0.014241      True
24    Eskimo_dog                    0.093412      True
25    hamster                       0.000461      False
26    briard                        0.206803      True
27    papillon                      0.004577      True
28    flat-coated_retriever         0.173810      True
29    gar                           0.189945      False
...   ...                                ...        ...
2045  Boston_bull                   0.000076      True
2046  car_wheel                     0.044173      False
2047  Dandie_Dinmont                0.086346      True
2048  can_opener                    0.032253      False
2049  Staffordshire_bullterrier     0.029705      True
2050  Weimaraner                    0.134203      True
2051  beagle                        0.038915      True
2052  shopping_cart                 0.026321      False
2053  Weimaraner                    0.026236      True
2054  bull_mastiff                  0.000903      True
2055  beagle                        0.038915      True
2056  malamute                      0.111411      True
```

2057	English_setter	0.073482	True
2058	Labrador_retriever	0.105506	True
2059	Saluki	0.027351	True
2060	redbone	0.007958	True
2061	Staffordshire_bullterrier	0.001498	True
2062	basenji	0.004633	True
2063	muzzle	0.082981	False
2064	chow	0.008167	True
2065	Chihuahua	0.029248	True
2066	Chesapeake_Bay_retriever	0.118184	True
2067	Pembroke	0.076507	True
2068	ice_lolly	0.193548	False
2069	Indian_elephant	0.076902	False
2070	German_short-haired_pointer	0.175219	True
2071	spatula	0.040836	False
2072	kelpie	0.031379	True
2073	papillon	0.068957	True
2074	banana	0.076110	False

	dog_breed
0	Welsh_springer_spaniel
1	redbone
2	German_shepherd
3	Rhodesian_ridgeback
4	miniature_pinscher
5	Bernese_mountain_dog
6	Not Dog
7	chow
8	golden_retriever
9	miniature_poodle
10	golden_retriever
11	Gordon_setter
12	Walker_hound
13	pug
14	bloodhound
15	Lhasa
16	English_setter
17	Not Dog
18	Not Dog
19	Italian_greyhound
20	Maltese_dog
21	Not Dog
22	Newfoundland
23	golden_retriever
24	malamute
25	Not Dog
26	soft-coated_wheaten_terrier
27	Chihuahua

```
28    black-and-tan_coonhound
29    Not Dog
...       ...
2045  French_bulldog
2046  Not Dog
2047  kuvasz
2048  Chihuahua
2049  Samoyed
2050  Mexican_hairless
2051  Pembroke
2052  Not Dog
2053  basset
2054  French_bulldog
2055  Pembroke
2056  Siberian_husky
2057  golden_retriever
2058  golden_retriever
2059  whippet
2060  golden_retriever
2061  French_bulldog
2062  Pembroke
2063  French_bulldog
2064  Samoyed
2065  Pembroke
2066  Irish_terrier
2067  Pomeranian
2068  Appenzeller
2069  Chesapeake_Bay_retriever
2070  basset
2071  Labrador_retriever
2072  Chihuahua
2073  Chihuahua
2074  Not Dog

[2075 rows x 13 columns]
```

```
In [75]: df_images_clean['dog_breed']=df_images_clean['dog_breed'].str.title()
```

```
In [76]: df_images_clean['dog'] = np.nan
         for i, row in df_images_clean.iterrows():
             if row['dog_breed'] == 'Not Dog':
                 df_images_clean.loc[i, ['dog']]= False
             else:
                 df_images_clean.loc[i, ['dog']]= True
```

- save Outcome

```
In [77]: df_images_clean.to_csv('images_clean.csv', index=False)
```

1.3.1.3. df_tweets_raw

- keep only the required columns ('id','retweet_count','favorite_count')

```
In [78]: df_tweets_clean = df_tweets_raw[['id','created_at','retweet_count','favorite_count']].c
```

- change the column name from id to tweet_id

```
In [79]: df_tweets_clean = df_tweets_clean.rename(columns = {'id':'tweet_id'})
```

- Change tweet_id column dtype to str

```
In [80]: df_tweets_clean['tweet_id'] = df_tweets_clean['tweet_id'].astype(str)
```

```
In [81]: df_tweets_clean['created_at']= pd.to_datetime(df_tweets_clean['created_at'])
```

- save Outcome

```
In [82]: df_tweets_clean.to_csv('tweets_clean.csv', index=False)
```

1.2.6 1.3.2. Tidiness Issues

- Set column (tweet_id) as index before merging

```
In [83]: df_twitter_clean.set_index('tweet_id', inplace =True)
         df_twitter_clean.head(2)
```

```
Out[83]:                             timestamp              source  \
         tweet_id
         892420643555336193 2017-08-01 16:23:56  Twitter for iPhone
         891774213063434 26 2017-08-01 00:17:27  Twitter for iPhone


         tweet_id
         892420643555336193  This is Phineas. He's a mystical boy. Only ever appears in the hole
         891774213063434 26  This is Tilly. She's just checking pup on you. Hopes you're doing o

                                                                                expanded_urls  \
         tweet_id
         892420643555336193  https://twitter.com/dog_rates/status/892420643555336193/photo/1
         891774213063434 26  https://twitter.com/dog_rates/status/891774213063434 26/photo/1

                             rating_numerator  rating_denominator     name dog_stage
         tweet_id
         892420643555336193  13.0              10.0                Phineas  NaN
         891774213063434 26  13.0              10.0                Tilly    NaN
```

```
In [84]: df_images_clean.set_index('tweet_id', inplace=True)
         df_images_clean.head(2)
```

```
Out[84]:                                                                jpg_url  img_num  \
         tweet_id
         666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg  1
         666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg  1

                                    prediction_1  confidence_1  breed_1  \
         tweet_id
         666020888022790149  Welsh_springer_spaniel  0.465074      True
         666029285002620928  redbone                 0.506826      True

                                   prediction_2  confidence_2  breed_2  \
         tweet_id
         666020888022790149  collie              0.156665      True
         666029285002620928  miniature_pinscher  0.074192      True

                                    prediction_3  confidence_3  breed_3  \
         tweet_id
         666020888022790149  Shetland_sheepdog    0.061428      True
         666029285002620928  Rhodesian_ridgeback  0.072010      True

                                       dog_breed   dog
         tweet_id
         666020888022790149  Welsh_Springer_Spaniel  True
         666029285002620928  Redbone                 True

In [85]: df_tweets_clean.set_index('tweet_id', inplace= True)
         df_tweets_clean.head(2)

Out[85]:                             created_at  retweet_count  favorite_count
         tweet_id
         892420643555336193 2017-08-01 16:23:56  8853           39467
         892177421306343426 2017-08-01 00:17:27  6514           33819
```

- combine a master dataframe by concatinating the 3 clean Dataframes with index (tweet_id)

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

<class 'pandas.core.frame.DataFrame'>
Index: 2094 entries, 892420643555336193 to 666020888022790149
Data columns (total 8 columns):
timestamp             2094 non-null datetime64[ns]
source                2094 non-null object
text                  2094 non-null object
expanded_urls         2094 non-null object
rating_numerator      2093 non-null float64
rating_denominator    2093 non-null float64
name                  2094 non-null object
dog_stage             335 non-null object
dtypes: datetime64[ns](1), float64(2), object(5)
```

```
memory usage: 147.2+ KB
```

```
In [87]: df_images_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2075 entries, 666020888022790149 to 892420643555336193
Data columns (total 13 columns):
jpg_url         2075 non-null object
img_num         2075 non-null int64
prediction_1    2075 non-null object
confidence_1    2075 non-null float64
breed_1         2075 non-null bool
prediction_2    2075 non-null object
confidence_2    2075 non-null float64
breed_2         2075 non-null bool
prediction_3    2075 non-null object
confidence_3    2075 non-null float64
breed_3         2075 non-null bool
dog_breed       2075 non-null object
dog             2075 non-null bool
dtypes: bool(4), float64(3), int64(1), object(5)
memory usage: 170.2+ KB
```

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

```
<class 'pandas.core.frame.DataFrame'>
Index: 2354 entries, 892420643555336193 to 666020888022790149
Data columns (total 3 columns):
created_at        2354 non-null datetime64[ns]
retweet_count     2354 non-null int64
favorite_count    2354 non-null int64
dtypes: datetime64[ns](1), int64(2)
memory usage: 73.6+ KB
```

```
In [89]: df_twitter_tweets = pd.concat([df_twitter_clean, df_tweets_clean], axis = 1, sort=True)
```

```
In [90]: df_twitter_master = pd.concat([df_twitter_tweets, df_images_clean], axis = 1, sort=True
```

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

```
<class 'pandas.core.frame.DataFrame'>
Index: 2356 entries, 666020888022790149 to 892420643555336193
Data columns (total 24 columns):
timestamp             2094 non-null datetime64[ns]
source                2094 non-null object
text                  2094 non-null object
```

```
expanded_urls         2094 non-null object
rating_numerator      2093 non-null float64
rating_denominator    2093 non-null float64
name                  2094 non-null object
dog_stage             335 non-null object
created_at            2354 non-null datetime64[ns]
retweet_count         2354 non-null float64
favorite_count        2354 non-null float64
jpg_url               2075 non-null object
img_num               2075 non-null float64
prediction_1          2075 non-null object
confidence_1          2075 non-null float64
breed_1               2075 non-null object
prediction_2          2075 non-null object
confidence_2          2075 non-null float64
breed_2               2075 non-null object
prediction_3          2075 non-null object
confidence_3          2075 non-null float64
breed_3               2075 non-null object
dog_breed             2075 non-null object
dog                   2075 non-null object
dtypes: datetime64[ns](2), float64(8), object(14)
memory usage: 460.2+ KB
```

- keep only tweets with images predections

```
In [92]: df_twitter_master.dropna(subset = ['jpg_url'], axis= 0,  inplace = True)
```

```
In [93]: df_twitter_master.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2075 entries, 666020888022790149 to 892420643555336193
Data columns (total 24 columns):
timestamp             1971 non-null datetime64[ns]
source                1971 non-null object
text                  1971 non-null object
expanded_urls         1971 non-null object
rating_numerator      1970 non-null float64
rating_denominator    1970 non-null float64
name                  1971 non-null object
dog_stage             303 non-null object
created_at            2073 non-null datetime64[ns]
retweet_count         2073 non-null float64
favorite_count        2073 non-null float64
jpg_url               2075 non-null object
img_num               2075 non-null float64
prediction_1          2075 non-null object
confidence_1          2075 non-null float64
```

```
breed_1               2075 non-null object
prediction_2          2075 non-null object
confidence_2          2075 non-null float64
breed_2               2075 non-null object
prediction_3          2075 non-null object
confidence_3          2075 non-null float64
breed_3               2075 non-null object
dog_breed             2075 non-null object
dog                   2075 non-null object
dtypes: datetime64[ns](2), float64(8), object(14)
memory usage: 405.3+ KB
```

```
In [94]: df_twitter_master.dropna(subset = ['timestamp'], axis= 0,  inplace = True)
```

```
In [95]: df_twitter_master.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1971 entries, 666020888022790149 to 892420643555336193
Data columns (total 24 columns):
timestamp             1971 non-null datetime64[ns]
source                1971 non-null object
text                  1971 non-null object
expanded_urls         1971 non-null object
rating_numerator      1970 non-null float64
rating_denominator    1970 non-null float64
name                  1971 non-null object
dog_stage             303 non-null object
created_at            1971 non-null datetime64[ns]
retweet_count         1971 non-null float64
favorite_count        1971 non-null float64
jpg_url               1971 non-null object
img_num               1971 non-null float64
prediction_1          1971 non-null object
confidence_1          1971 non-null float64
breed_1               1971 non-null object
prediction_2          1971 non-null object
confidence_2          1971 non-null float64
breed_2               1971 non-null object
prediction_3          1971 non-null object
confidence_3          1971 non-null float64
breed_3               1971 non-null object
dog_breed             1971 non-null object
dog                   1971 non-null object
dtypes: datetime64[ns](2), float64(8), object(14)
memory usage: 385.0+ KB
```

- verify the time stamp and created_at from tweets & twitter dataframes are the same

```
In [96]: df_twitter_master['timestamp_verify'] = df_twitter_master['timestamp'] == df_twitter_ma

In [97]: df_twitter_master['timestamp_verify'].value_counts()

Out[97]: True    1971
         Name: timestamp_verify, dtype: int64

In [98]: df_twitter_master.drop(columns = ['timestamp_verify', 'created_at'], inplace=True)

In [99]: df_twitter_master.info()

<class 'pandas.core.frame.DataFrame'>
Index: 1971 entries, 666020888022790149 to 892420643555336193
Data columns (total 23 columns):
timestamp             1971 non-null datetime64[ns]
source                1971 non-null object
text                  1971 non-null object
expanded_urls         1971 non-null object
rating_numerator      1970 non-null float64
rating_denominator    1970 non-null float64
name                  1971 non-null object
dog_stage             303 non-null object
retweet_count         1971 non-null float64
favorite_count        1971 non-null float64
jpg_url               1971 non-null object
img_num               1971 non-null float64
prediction_1          1971 non-null object
confidence_1          1971 non-null float64
breed_1               1971 non-null object
prediction_2          1971 non-null object
confidence_2          1971 non-null float64
breed_2               1971 non-null object
prediction_3          1971 non-null object
confidence_3          1971 non-null float64
breed_3               1971 non-null object
dog_breed             1971 non-null object
dog                   1971 non-null object
dtypes: datetime64[ns](1), float64(8), object(14)
memory usage: 449.6+ KB
```

- arrange dataframe by timestamp

```
In [100]: df_twitter_master.sort_values(by='timestamp')

Out[100]:                              timestamp              source  \
          666020888022790149 2015-11-15 22:32:08  Twitter for iPhone
          666029285002620928 2015-11-15 23:05:30  Twitter for iPhone
          666033412701032449 2015-11-15 23:21:54  Twitter for iPhone
```

```
666044226329800704 2015-11-16 00:04:52  Twitter for iPhone
666049248165822465 2015-11-16 00:24:50  Twitter for iPhone
666050758794694657 2015-11-16 00:30:50  Twitter for iPhone
666051853826850816 2015-11-16 00:35:11  Twitter for iPhone
666055525042405380 2015-11-16 00:49:46  Twitter for iPhone
666057090499244032 2015-11-16 00:55:59  Twitter for iPhone
666058600524156928 2015-11-16 01:01:59  Twitter for iPhone
666063827256086533 2015-11-16 01:22:45  Twitter for iPhone
666071193221509120 2015-11-16 01:52:02  Twitter for iPhone
666073100786774016 2015-11-16 01:59:36  Twitter for iPhone
666082916733198337 2015-11-16 02:38:37  Twitter for iPhone
666094000022159362 2015-11-16 03:22:39  Twitter for iPhone
666099513787052032 2015-11-16 03:44:34  Twitter for iPhone
666102155909144576 2015-11-16 03:55:04  Twitter for iPhone
666104133288665088 2015-11-16 04:02:55  Twitter for iPhone
666268910803644416 2015-11-16 14:57:41  Twitter for iPhone
666273097616637952 2015-11-16 15:14:19  Twitter for iPhone
666287406224695296 2015-11-16 16:11:11  Twitter for iPhone
666293911632134144 2015-11-16 16:37:02  Twitter for iPhone
666337882303524864 2015-11-16 19:31:45  Twitter for iPhone
666345417576210432 2015-11-16 20:01:42  Twitter for iPhone
666353288456101888 2015-11-16 20:32:58  Twitter for iPhone
666362758909284353 2015-11-16 21:10:36  Twitter for iPhone
666373753744588802 2015-11-16 21:54:18  Twitter for iPhone
666396247373291520 2015-11-16 23:23:41  Twitter for iPhone
666407126856765440 2015-11-17 00:06:54  Twitter for iPhone
666411507551481857 2015-11-17 00:24:19  Twitter for iPhone
...                                ...                 ...
886258384151887873 2017-07-15 16:17:19  Twitter for iPhone
886366144734445568 2017-07-15 23:25:31  Twitter for iPhone
886680336477933568 2017-07-16 20:14:00  Twitter for iPhone
886736880519319552 2017-07-16 23:58:41  Twitter for iPhone
886983233522544640 2017-07-17 16:17:36  Twitter for iPhone
887101392804085760 2017-07-18 00:07:08  Twitter for iPhone
887343217045368832 2017-07-18 16:08:03  Twitter for iPhone
887473957103951883 2017-07-19 00:47:34  Twitter for iPhone
887517139158093824 2017-07-19 03:39:09  Twitter for iPhone
887705289381826560 2017-07-19 16:06:48  Twitter for iPhone
888078434458587136 2017-07-20 16:49:33  Twitter for iPhone
888554962724278272 2017-07-22 00:23:06  Twitter for iPhone
888804989199671297 2017-07-22 16:56:37  Twitter for iPhone
889017238123831296 2017-07-23 00:22:39  Twitter for iPhone
889278841981685760 2017-07-24 00:19:32  Twitter for iPhone
889531135344209921 2017-07-24 17:02:04  Twitter for iPhone
889638837579907072 2017-07-25 00:10:02  Twitter for iPhone
889665388333682689 2017-07-25 01:55:32  Twitter for iPhone
889880896479866881 2017-07-25 16:11:53  Twitter for iPhone
890006608113172480 2017-07-26 00:31:25  Twitter for iPhone
```

```
890240255349198849 2017-07-26 15:59:51  Twitter for iPhone
890609185150312448 2017-07-27 16:25:51  Twitter for iPhone
890729181411237888 2017-07-28 00:22:40  Twitter for iPhone
890971913173991426 2017-07-28 16:27:12  Twitter for iPhone
891087950875897856 2017-07-29 00:08:17  Twitter for iPhone
891327558926688256 2017-07-29 16:00:24  Twitter for iPhone
891689557279858688 2017-07-30 15:58:51  Twitter for iPhone
891815181378084864 2017-07-31 00:18:03  Twitter for iPhone
892177421306343426 2017-08-01 00:17:27  Twitter for iPhone
892420643555336193 2017-08-01 16:23:56  Twitter for iPhone
```

```
666020888022790149  Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big
666029285002620928  This is a western brown Mitsubishi terrier. Upset about leaf. Actu
666033412701032449  Here is a very happy pup. Big fan of well-maintained decks. Just l
666044226329800704  This is a purebred Piers Morgan. Loves to Netflix and chill. Alway
666049248165822465  Here we have a 1949 1st generation vulpix. Enjoys sweat tea and Fo
666050758794694657  This is a truly beautiful English Wilson Staff retriever. Has a ni
666051853826850816  This is an odd dog. Hard on the outside but loving on the inside.
666055525042405380  Here is a Siberian heavily armored polar bear mix. Strong owner. 1
666057090499244032  My oh my. This is a rare blond Canadian terrier on wheels. Only $8
666058600524156928  Here is the Rand Paul of retrievers folks! He's probably good at p
666063827256086533  This is the happiest dog you will ever see. Very committed owner.
666071193221509120  Here we have a northern speckled Rhododendron. Much sass. Gives 0
666073100786774016  Let's hope this flight isn't Malaysian (lol). What a dog! Almost c
666082916733198337  Here we have a well-established sunblockerspaniel. Lost his other
666094000022159362  This appears to be a Mongolian Presbyterian mix. Very tired. Tongu
666099513787052032  Can stand on stump for what seems like a while. Built that birdhou
666102155909144576  Oh my. Here you are seeing an Adobe Setter giving birth to twins!!
666104133288665088  Not familiar with this breed. No tail (weird). Only 2 legs. Doesn'
666268910803644416  Very concerned about fellow dog trapped in computer. 10/10 https:/
666273097616637952  Can take selfies 11/10 https://t.co/ws2AMaNwPW
666287406224695296  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polishe
666293911632134144  This is a funny dog. Weird toes. Won't come down. Loves branch. Re
666337882303524864  This is an extremely rare horned Parthenon. Not amused. Wears shoe
666345417576210432  Look at this jokester thinking seat belt laws don't apply to him.
666353288456101888  Here we have a mixed Asiago from the Galápagos Islands. Only one e
666362758909284353  Unique dog here. Very small. Lives in container of Frosted Flakes
666373753744588802  Those are sunglasses and a jean jacket. 11/10 dog cool af https://
666396247373291520  Oh goodness. A super rare northeast Qdoba kangaroo mix. Massive fe
666407126856765440  This is a southern Vesuvius bumblegruff. Can drive a truck (wow).
666411507551481857  This is quite the dog. Gets really excited when not in water. Not
...
886258384151887873  This is Waffles. His doggles are pupside down. Unsure how to fix.
886366144734445568  This is Roscoe. Another pupper fallen victim to spontaneous tongue
886680336477933568  This is Derek. He's late for a dog meeting. 13/10 pet...al to the
886736880519319552  This is Mingus. He's a wonderful father to his smol pup. Confirmed
886983233522544640  This is Maya. She's very shy. Rarely leaves her cup. 13/10 would f
```

887101392804085760	This... is a Jubilant Antarctic House Bear. We only rate dogs. Ple
887343217045368832	You may not have known you needed to see this today. 13/10 please
887473957103951883	This is Canela. She attempted some fancy porch pics. They were uns
887517139158093824	I've yet to rate a Venezuelan Hover Wiener. This is such an honor.
887705289381826560	This is Jeffrey. He has a monopoly on the pool noodles. Currently
888078434458587136	This is Gerald. He was just told he didn't get the job he intervie
888554962724278272	This is Ralphus. He's powering up. Attempting maximum borkdrive. 1
888804989199671297	This is Zeke. He has a new stick. Very proud of it. Would like you
888917238123831296	This is Jim. He found a fren. Taught him how to sit like the good
889278841981685760	This is Oliver. You're witnessing one of his many brutal attacks.
889531135344209921	This is Stuart. He's sporting his favorite fanny pack. Secretly fi
889638837579907072	This is Ted. He does his best. Sometimes that's not enough. But it
889665388333682689	Here's a puppo that seems to be on the fence about something haha
889880896479866881	This is Bruno. He is a service shark. Only gets out of the water t
890006608113172480	This is Koda. He is a South Australian deckshark. Deceptively dead
890240255349198849	This is Cassie. She is a college pup. Studying international doggo
890609185150312448	This is Zoey. She doesn't want to be one of the scary sharks. Just
890729181411237888	When you watch your owner call another dog a good boy but then the
890971913173991426	Meet Jax. He enjoys ice cream so much he gets nervous around it. 1
891087950875897856	Here we have a majestic great white breaching off South Africa's c
891327558926688256	This is Franklin. He would like you to stop calling him "cute." He
891689557279858688	This is Darla. She commenced a snooze mid meal. 13/10 happens to t
891815181378084864	This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in th
892177421306343426	This is Tilly. She's just checking pup on you. Hopes you're doing
892420643555336193	This is Phineas. He's a mystical boy. Only ever appears in the hol

666020888022790149	https://twitter.com/dog_rates/status/666020888022790149/photo/1
666029285002620928	https://twitter.com/dog_rates/status/666029285002620928/photo/1
666033412701032449	https://twitter.com/dog_rates/status/666033412701032449/photo/1
666044226329800704	https://twitter.com/dog_rates/status/666044226329800704/photo/1
666049248165822465	https://twitter.com/dog_rates/status/666049248165822465/photo/1
666050758794694657	https://twitter.com/dog_rates/status/666050758794694657/photo/1
666051853826850816	https://twitter.com/dog_rates/status/666051853826850816/photo/1
666055525042405380	https://twitter.com/dog_rates/status/666055525042405380/photo/1
666057090499244032	https://twitter.com/dog_rates/status/666057090499244032/photo/1
666058600524156928	https://twitter.com/dog_rates/status/666058600524156928/photo/1
666063827256086533	https://twitter.com/dog_rates/status/666063827256086533/photo/1
666071193221509120	https://twitter.com/dog_rates/status/666071193221509120/photo/1
666073100786774016	https://twitter.com/dog_rates/status/666073100786774016/photo/1
666082916733198337	https://twitter.com/dog_rates/status/666082916733198337/photo/1
666094000022159362	https://twitter.com/dog_rates/status/666094000022159362/photo/1
666099513787052032	https://twitter.com/dog_rates/status/666099513787052032/photo/1
666102155909144576	https://twitter.com/dog_rates/status/666102155909144576/photo/1
666104133288665088	https://twitter.com/dog_rates/status/666104133288665088/photo/1
666268910803644416	https://twitter.com/dog_rates/status/666268910803644416/photo/1
666273097616637952	https://twitter.com/dog_rates/status/666273097616637952/photo/1
666287406224695296	https://twitter.com/dog_rates/status/666287406224695296/photo/1

```
666293911632134144  https://twitter.com/dog_rates/status/666293911632134144/photo/1
666337882303524864  https://twitter.com/dog_rates/status/666337882303524864/photo/1
666345417576210432  https://twitter.com/dog_rates/status/666345417576210432/photo/1
666353288456101888  https://twitter.com/dog_rates/status/666353288456101888/photo/1
666362758909284353  https://twitter.com/dog_rates/status/666362758909284353/photo/1
666373753744588802  https://twitter.com/dog_rates/status/666373753744588802/photo/1
666396247373291520  https://twitter.com/dog_rates/status/666396247373291520/photo/1
666407126856765440  https://twitter.com/dog_rates/status/666407126856765440/photo/1
666411507551481857  https://twitter.com/dog_rates/status/666411507551481857/photo/1
...                                                                            ...
886258384151887873  https://twitter.com/dog_rates/status/886258384151887873/photo/1
886366144734445568  https://twitter.com/dog_rates/status/886366144734445568/photo/1,ht
886680336477933568  https://twitter.com/dog_rates/status/886680336477933568/photo/1
886736880519319552  https://www.gofundme.com/mingusneedsus,https://twitter.com/dog_rat
886983233522544640  https://twitter.com/dog_rates/status/886983233522544640/photo/1,ht
887101392804085760  https://twitter.com/dog_rates/status/887101392804085760/photo/1
887343217045368832  https://twitter.com/dog_rates/status/887343217045368832/video/1
887473957103951883  https://twitter.com/dog_rates/status/887473957103951883/photo/1,ht
887517139158093824  https://twitter.com/dog_rates/status/887517139158093824/video/1
887705289381826560  https://twitter.com/dog_rates/status/887705289381826560/photo/1
888078434458587136  https://twitter.com/dog_rates/status/888078434458587136/photo/1,ht
888554962724278272  https://twitter.com/dog_rates/status/888554962724278272/photo/1,ht
888804989199671297  https://twitter.com/dog_rates/status/888804989199671297/photo/1,ht
889017238123831296  https://twitter.com/dog_rates/status/889017238123831296/photo/1
889278841981685760  https://twitter.com/dog_rates/status/889278841981685760/video/1
889531135344209921  https://twitter.com/dog_rates/status/889531135344209921/photo/1
889638837579907072  https://twitter.com/dog_rates/status/889638837579907072/photo/1,ht
889665388333682689  https://twitter.com/dog_rates/status/889665388333682689/photo/1
889880896479866881  https://twitter.com/dog_rates/status/889880896479866881/photo/1
890006608113172480  https://twitter.com/dog_rates/status/890006608113172480/photo/1,ht
890240255349198849  https://twitter.com/dog_rates/status/890240255349198849/photo/1
890609185150312448  https://twitter.com/dog_rates/status/890609185150312448/photo/1
890729181411237888  https://twitter.com/dog_rates/status/890729181411237888/photo/1,ht
890971913173991426  https://gofundme.com/ydvmve-surgery-for-jax,https://twitter.com/do
891087950875897856  https://twitter.com/dog_rates/status/891087950875897856/photo/1
891327558926688256  https://twitter.com/dog_rates/status/891327558926688256/photo/1,ht
891689557279858688  https://twitter.com/dog_rates/status/891689557279858688/photo/1
891815181378084864  https://twitter.com/dog_rates/status/891815181378084864/photo/1
892177421306343426  https://twitter.com/dog_rates/status/892177421306343426/photo/1
892420643555336193  https://twitter.com/dog_rates/status/892420643555336193/photo/1

                    rating_numerator  rating_denominator      name dog_stage  \
666020888022790149  8.0               10.0                 None      NaN
666029285002620928  7.0               10.0                 a         NaN
666033412701032449  9.0               10.0                 a         NaN
666044226329800704  6.0               10.0                 a         NaN
666049248165822465  5.0               10.0                 None      NaN
666050758794694657  10.0              10.0                 a         NaN
```

666051853826850816	2.0	10.0	an	NaN
666055525042405380	10.0	10.0	a	NaN
666057090499244032	9.0	10.0	a	NaN
666058600524156928	8.0	10.0	the	NaN
666063827256086533	10.0	10.0	the	NaN
666071193221509120	9.0	10.0	None	NaN
666073100786774016	10.0	10.0	None	NaN
666082916733198337	6.0	10.0	None	NaN
666094000022159362	9.0	10.0	None	NaN
666099513787052032	8.0	10.0	None	NaN
666102155909144576	11.0	10.0	None	NaN
666104133288665088	1.0	10.0	None	NaN
666268910803644416	10.0	10.0	None	NaN
666273097616637952	11.0	10.0	None	NaN
666287406224695296	9.0	10.0	an	NaN
666293911632134144	3.0	10.0	a	NaN
666337882303524864	9.0	10.0	an	NaN
666345417576210432	10.0	10.0	None	NaN
666353288456101888	8.0	10.0	None	NaN
666362758909284353	6.0	10.0	None	NaN
666373753744588802	11.0	10.0	None	NaN
666396247373291520	9.0	10.0	None	NaN
666407126856765440	7.0	10.0	a	NaN
666411507551481857	2.0	10.0	quite	NaN
...	...	...	...	...
886258384151887873	13.0	10.0	Waffles	NaN
886366144734445568	12.0	10.0	Roscoe	pupper
886680336477933568	13.0	10.0	Derek	NaN
886736880519319552	13.0	10.0	Mingus	NaN
886983233522544640	13.0	10.0	Maya	NaN
887101392804085760	12.0	10.0	None	NaN
887343217045368832	13.0	10.0	None	NaN
887473957103951883	13.0	10.0	Canela	NaN
887517139158093824	14.0	10.0	such	NaN
887705289381826560	13.0	10.0	Jeffrey	NaN
888078434458587136	12.0	10.0	Gerald	NaN
888554962724278272	13.0	10.0	Ralphus	NaN
888804989199671297	13.0	10.0	Zeke	NaN
888917238123831296	12.0	10.0	Jim	NaN
889278841981685760	13.0	10.0	Oliver	NaN
889531135344209921	13.0	10.0	Stuart	puppo
889638837579907072	12.0	10.0	Ted	NaN
889665388333682689	13.0	10.0	None	puppo
889880896479866881	13.0	10.0	Bruno	NaN
890006608113172480	13.0	10.0	Koda	NaN
890240255349198849	14.0	10.0	Cassie	doggo
890609185150312448	13.0	10.0	Zoey	NaN
890729181411237888	13.0	10.0	None	NaN

```
890971913173991426  13.0              10.0               Jax       NaN
891087950875897856  13.0              10.0               None      NaN
891327558926688256  12.0              10.0               Franklin  NaN
891689557279858688  13.0              10.0               Darla     NaN
891815181378084864  12.0              10.0               Archie    NaN
892177421306343426  13.0              10.0               Tilly     NaN
892420643555336193  13.0              10.0               Phineas   NaN

                    retweet_count  favorite_count  ...   confidence_1  \
666020888022790149  532.0          2535.0          ...    0.465074
666029285002620928  48.0           132.0           ...    0.506826
666033412701032449  47.0           128.0           ...    0.596461
666044226329800704  147.0          311.0           ...    0.408143
666049248165822465  41.0           111.0           ...    0.560311
666050758794694657  60.0           136.0           ...    0.651137
666051853826850816  879.0          1253.0          ...    0.933012
666055525042405380  261.0          448.0           ...    0.692517
666057090499244032  146.0          304.0           ...    0.962465
666058600524156928  61.0           115.0           ...    0.201493
666063827256086533  232.0          496.0           ...    0.775930
666071193221509120  67.0           154.0           ...    0.503672
666073100786774016  174.0          335.0           ...    0.260857
666082916733198337  47.0           121.0           ...    0.489814
666094000022159362  79.0           169.0           ...    0.195217
666099513787052032  73.0           164.0           ...    0.582330
666102155909144576  16.0           81.0            ...    0.298617
666104133288665088  6871.0         14765.0         ...    0.965932
666268910803644416  37.0           108.0           ...    0.086502
666273097616637952  82.0           184.0           ...    0.176053
666287406224695296  71.0           152.0           ...    0.857531
666293911632134144  368.0          522.0           ...    0.914671
666337882303524864  96.0           204.0           ...    0.416669
666345417576210432  146.0          307.0           ...    0.858744
666353288456101888  77.0           229.0           ...    0.336874
666362758909284353  595.0          804.0           ...    0.996496
666373753744588802  100.0          194.0           ...    0.326467
666396247373291520  92.0           172.0           ...    0.978108
666407126856765440  44.0           113.0           ...    0.529139
666411507551481857  339.0          459.0           ...    0.404640
...                   ...            ...           ...         ...
886258384151887873  6523.0         28469.0         ...    0.943575
886366144734445568  3316.0         21524.0         ...    0.999201
886680336477933568  4610.0         22798.0         ...    0.738995
886736880519319552  3443.0         12306.0         ...    0.309706
886983233522544640  8084.0         35859.0         ...    0.793469
887101392804085760  6167.0         31061.0         ...    0.733942
887343217045368832  10737.0        34222.0         ...    0.330741
887473957103951883  18781.0        69871.0         ...    0.809197
```

```
887517139158093824  12082.0        46959.0         ...    0.130432
887705289381826560  5609.0         30779.0         ...    0.821664
888078434458587136  3653.0         22201.0         ...    0.995026
888554962724278272  3732.0         20290.0         ...    0.700377
888804989199671297  4559.0         26080.0         ...    0.469760
888917238123831296  4709.0         29611.0         ...    0.714719
889278841981685760  5637.0         25652.0         ...    0.626152
889531135344209921  2321.0         15359.0         ...    0.953442
889638837579907072  4735.0         27672.0         ...    0.991650
889665388333682689  8538.0         38818.0         ...    0.966327
889880896479866881  5156.0         28268.0         ...    0.377417
890006608113172480  7624.0         31166.0         ...    0.957979
890240255349198849  7711.0         32467.0         ...    0.511319
890609185150312448  4429.0         28226.0         ...    0.487574
890729181411237888  16716.0        56848.0         ...    0.566142
890971913173991426  2158.0         12041.0         ...    0.341703
891087950875897856  3261.0         20562.0         ...    0.425595
891327558926688256  9774.0         41048.0         ...    0.555712
891689557279858688  8964.0         42908.0         ...    0.170278
891815181378084864  4328.0         25461.0         ...    0.716012
892177421306343426  6514.0         33819.0         ...    0.323581
892420643555336193  8853.0         39467.0         ...    0.097049

                    breed_1              prediction_2  confidence_2 breed_2  \
666020888022790149  True     collie                    0.156665      True
666029285002620928  True     miniature_pinscher        0.074192      True
666033412701032449  True     malinois                  0.138584      True
666044226329800704  True     redbone                   0.360687      True
666049248165822465  True     Rottweiler                0.243682      True
666050758794694657  True     English_springer          0.263788      True
666051853826850816  False    mud_turtle                0.045885      False
666055525042405380  True     Tibetan_mastiff           0.058279      True
666057090499244032  False    shopping_basket           0.014594      False
666058600524156928  True     komondor                  0.192305      True
666063827256086533  True     Tibetan_mastiff           0.093718      True
666071193221509120  True     Yorkshire_terrier         0.174201      True
666073100786774016  True     English_foxhound          0.175382      True
666082916733198337  True     bull_mastiff              0.404722      True
666094000022159362  True     German_shepherd           0.078260      True
666099513787052032  True     Shih-Tzu                  0.166192      True
666102155909144576  True     Newfoundland              0.149842      True
666104133288665088  False    cock                      0.033919      False
666268910803644416  False    desk                      0.085547      False
666273097616637952  True     toy_terrier               0.111884      True
666287406224695296  True     toy_poodle                0.063064      True
666293911632134144  False    otter                     0.015250      False
666337882303524864  False    Newfoundland              0.278407      True
666345417576210432  True     Chesapeake_Bay_retriever  0.054787      True
```

666353288456101888	True	Siberian_husky	0.147655	True
666362758909284353	False	skunk	0.002402	False
666373753744588802	True	Afghan_hound	0.259551	True
666396247373291520	True	toy_terrier	0.009397	True
666407126856765440	True	bloodhound	0.244220	True
666411507551481857	False	barracouta	0.271485	False
...	...	...	...	...
886258384151887873	True	shower_cap	0.025286	False
886366144734445568	True	Chihuahua	0.000361	True
886680336477933568	False	sports_car	0.139952	False
886736880519319552	True	Great_Pyrenees	0.186136	True
886983233522544640	True	toy_terrier	0.143528	True
887101392804085760	True	Eskimo_dog	0.035029	True
887343217045368832	True	sea_lion	0.275645	False
887473957103951883	True	Rhodesian_ridgeback	0.054950	True
887517139158093824	False	tow_truck	0.029175	False
887705289381826560	True	redbone	0.087582	True
888078434458587136	True	pug	0.000932	True
888554962724278272	True	Eskimo_dog	0.166511	True
888804989199671297	True	Labrador_retriever	0.184172	True
888917238123831296	True	Tibetan_mastiff	0.120184	True
889278841981685760	True	borzoi	0.194742	True
889531135344209921	True	Labrador_retriever	0.013834	True
889638837579907072	True	boxer	0.002129	True
889665388333682689	True	Cardigan	0.027356	True
889880896479866881	True	Labrador_retriever	0.151317	True
890006608113172480	True	Pomeranian	0.013884	True
890240255349198849	True	Cardigan	0.451038	True
890609185150312448	True	Irish_setter	0.193054	True
890729181411237888	True	Eskimo_dog	0.178406	True
890971913173991426	True	Border_collie	0.199287	True
891087950875897856	True	Irish_terrier	0.116317	True
891327558926688256	True	English_springer	0.225770	True
891689557279858688	False	Labrador_retriever	0.168086	True
891815181378084864	True	malamute	0.078253	True
892177421306343426	True	Pekinese	0.090647	True
892420643555336193	False	bagel	0.085851	False

	prediction_3	confidence_3	breed_3	\
666020888022790149	Shetland_sheepdog	0.061428	True	
666029285002620928	Rhodesian_ridgeback	0.072010	True	
666033412701032449	bloodhound	0.116197	True	
666044226329800704	miniature_pinscher	0.222752	True	
666049248165822465	Doberman	0.154629	True	
666050758794694657	Greater_Swiss_Mountain_dog	0.016199	True	
666051853826850816	terrapin	0.017885	False	
666055525042405380	fur_coat	0.054449	False	
666057090499244032	golden_retriever	0.007959	True	

666058600524156928	soft-coated_wheaten_terrier	0.082086	True
666063827256086533	Labrador_retriever	0.072427	True
666071193221509120	Pekinese	0.109454	True
666073100786774016	Ibizan_hound	0.097471	True
666082916733198337	French_bulldog	0.048960	True
666094000022159362	malinois	0.075628	True
666099513787052032	Dandie_Dinmont	0.089688	True
666102155909144576	borzoi	0.133649	True
666104133288665088	partridge	0.000052	False
666268910803644416	bookcase	0.079480	False
666273097616637952	basenji	0.111152	True
666287406224695296	miniature_poodle	0.025581	True
666293911632134144	great_grey_owl	0.013207	False
666337882303524864	groenendael	0.102643	True
666345417576210432	Labrador_retriever	0.014241	True
666353288456101888	Eskimo_dog	0.093412	True
666362758909284353	hamster	0.000461	False
666373753744588802	briard	0.206803	True
666396247373291520	papillon	0.004577	True
666407126856765440	flat-coated_retriever	0.173810	True
666411507551481857	gar	0.189945	False
...	...	...	...
886258384151887873	Siamese_cat	0.002849	False
886366144734445568	Boston_bull	0.000076	True
886680336477933568	car_wheel	0.044173	False
886736880519319552	Dandie_Dinmont	0.086346	True
886983233522544640	can_opener	0.032253	False
887101392804085760	Staffordshire_bullterrier	0.029705	True
887343217045368832	Weimaraner	0.134203	True
887473957103951883	beagle	0.038915	True
887517139158093824	shopping_cart	0.026321	False
887705289381826560	Weimaraner	0.026236	True
888078434458587136	bull_mastiff	0.000903	True
888554962724278272	malamute	0.111411	True
888804989199671297	English_setter	0.073482	True
888917238123831296	Labrador_retriever	0.105506	True
889278841981685760	Saluki	0.027351	True
889531135344209921	redbone	0.007958	True
889638837579907072	Staffordshire_bullterrier	0.001498	True
889665388333682689	basenji	0.004633	True
889880896479866881	muzzle	0.082981	False
890006608113172480	chow	0.008167	True
890240255349198849	Chihuahua	0.029248	True
890609185150312448	Chesapeake_Bay_retriever	0.118184	True
890729181411237888	Pembroke	0.076507	True
890971913173991426	ice_lolly	0.193548	False
891087950875897856	Indian_elephant	0.076902	False
891327558926688256	German_short-haired_pointer	0.175219	True

```
891689557279858688  spatula                      0.040836      False
891815181378084864  kelpie                       0.031379      True
892177421306343426  papillon                     0.068957      True
892420643555336193  banana                       0.076110      False

                                    dog_breed    dog
666020888022790149  Welsh_Springer_Spaniel       True
666029285002620928  Redbone                      True
666033412701032449  German_Shepherd              True
666044226329800704  Rhodesian_Ridgeback          True
666049248165822465  Miniature_Pinscher           True
666050758794694657  Bernese_Mountain_Dog         True
666051853826850816  Not Dog                      False
666055525042405380  Chow                         True
666057090499244032  Golden_Retriever             True
666058600524156928  Miniature_Poodle             True
666063827256086533  Golden_Retriever             True
666071193221509120  Gordon_Setter                True
666073100786774016  Walker_Hound                 True
666082916733198337  Pug                          True
666094000022159362  Bloodhound                   True
666099513787052032  Lhasa                        True
666102155909144576  English_Setter               True
666104133288665088  Not Dog                      False
666268910803644416  Not Dog                      False
666273097616637952  Italian_Greyhound            True
666287406224695296  Maltese_Dog                  True
666293911632134144  Not Dog                      False
666337882303524864  Newfoundland                 True
666345417576210432  Golden_Retriever             True
666353288456101888  Malamute                     True
666362758909284353  Not Dog                      False
666373753744588802  Soft-Coated_Wheaten_Terrier  True
666396247373291520  Chihuahua                    True
666407126856765440  Black-And-Tan_Coonhound       True
666411507551481857  Not Dog                      False
...                     ...                        ...
886258384151887873  Pug                          True
886366144734445568  French_Bulldog               True
886680336477933568  Not Dog                      False
886736880519319552  Kuvasz                       True
886983233522544640  Chihuahua                    True
887101392804085760  Samoyed                      True
887343217045368832  Mexican_Hairless             True
887473957103951883  Pembroke                     True
887517139158093824  Not Dog                      False
887705289381826560  Basset                       True
888078434458587136  French_Bulldog               True
```

```
888554962724278272  Siberian_Husky                True
888804989199671297  Golden_Retriever              True
888917238123831296  Golden_Retriever              True
889278841981685760  Whippet                       True
889531135344209921  Golden_Retriever              True
889638837579907072  French_Bulldog                True
889665388333682689  Pembroke                      True
889880896479866881  French_Bulldog                True
890006608113172480  Samoyed                       True
890240255349198849  Pembroke                      True
890609185150312448  Irish_Terrier                 True
890729181411237888  Pomeranian                    True
890971913173991426  Appenzeller                   True
891087950875897856  Chesapeake_Bay_Retriever      True
891327558926688256  Basset                        True
891689557279858688  Labrador_Retriever            True
891815181378084864  Chihuahua                     True
892177421306343426  Chihuahua                     True
892420643555336193  Not Dog                       False

[1971 rows x 23 columns]
```

- Save result

```
In [101]: df_twitter_master.to_csv('twitter_archive_master.csv')
```

- Data now ready to act on

2 2. Data Visualization and Insights

2.1 2.1. exploratory chart

2.2 2.2. Profile over time

2.2.1 2.2.1. Interaction over time

2.2.2 2.2.2. Average tweets count over time

2.2.3 2.2.3. correlation matrix between retweets and favorite

2.2.4 2.2.3. Interactions of images of dogs vs not dog images

2.2.5 2.2.4. Average rating over time

2.3 2.3. Dog Breeds

2.3.1 2.3.1. Tweets counts for dog breeds

2.3.2 2.3.2. Retweet & favorite counts for dog breeds

2.3.3 2.3.3. Retweet & favorite average for dog breeds

2.3.4 2.3.4. Average Rating for dog breeds

2.4 2.4. Dog Stages

2.4.1 2.4.1. Tweets Counts for dog stages

2.4.2 2.4.2. Retweet & favorite for dog stages

2.4.3 2.4.3. Average Rating for dog stages

```
In [102]: df_tw_plt = df_twitter_master.set_index('timestamp')

In [103]: df_tw_plt.drop(columns = ['img_num', 'jpg_url','text','expanded_urls', 'prediction_1',

In [104]: df_tw_plt.info()

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1971 entries, 2015-11-15 22:32:08 to 2017-08-01 16:23:56
Data columns (total 9 columns):
source                1971 non-null object
rating_numerator      1970 non-null float64
rating_denominator    1970 non-null float64
name                  1971 non-null object
dog_stage             303 non-null object
retweet_count         1971 non-null float64
favorite_count        1971 non-null float64
dog_breed             1971 non-null object
dog                   1971 non-null object
dtypes: float64(4), object(5)
memory usage: 154.0+ KB
```

```
In [105]: df_tw_plt.loc[df_tw_plt.rating_numerator>20]
```

```
Out[105]:                                source  rating_numerator  rating_denominator  \
          timestamp
          2015-11-29 05:52:33  Twitter for iPhone  420.0             10.0
          2016-07-04 15:00:45  TweetDeck           1776.0            10.0

                                  name dog_stage  retweet_count  favorite_count  \
          timestamp
          2015-11-29 05:52:33  None     NaN       4324.0         7989.0
          2016-07-04 15:00:45  Atticus  NaN       2772.0         5569.0

                               dog_breed    dog
          timestamp
          2015-11-29 05:52:33  Not Dog   False
          2016-07-04 15:00:45  Not Dog   False
```

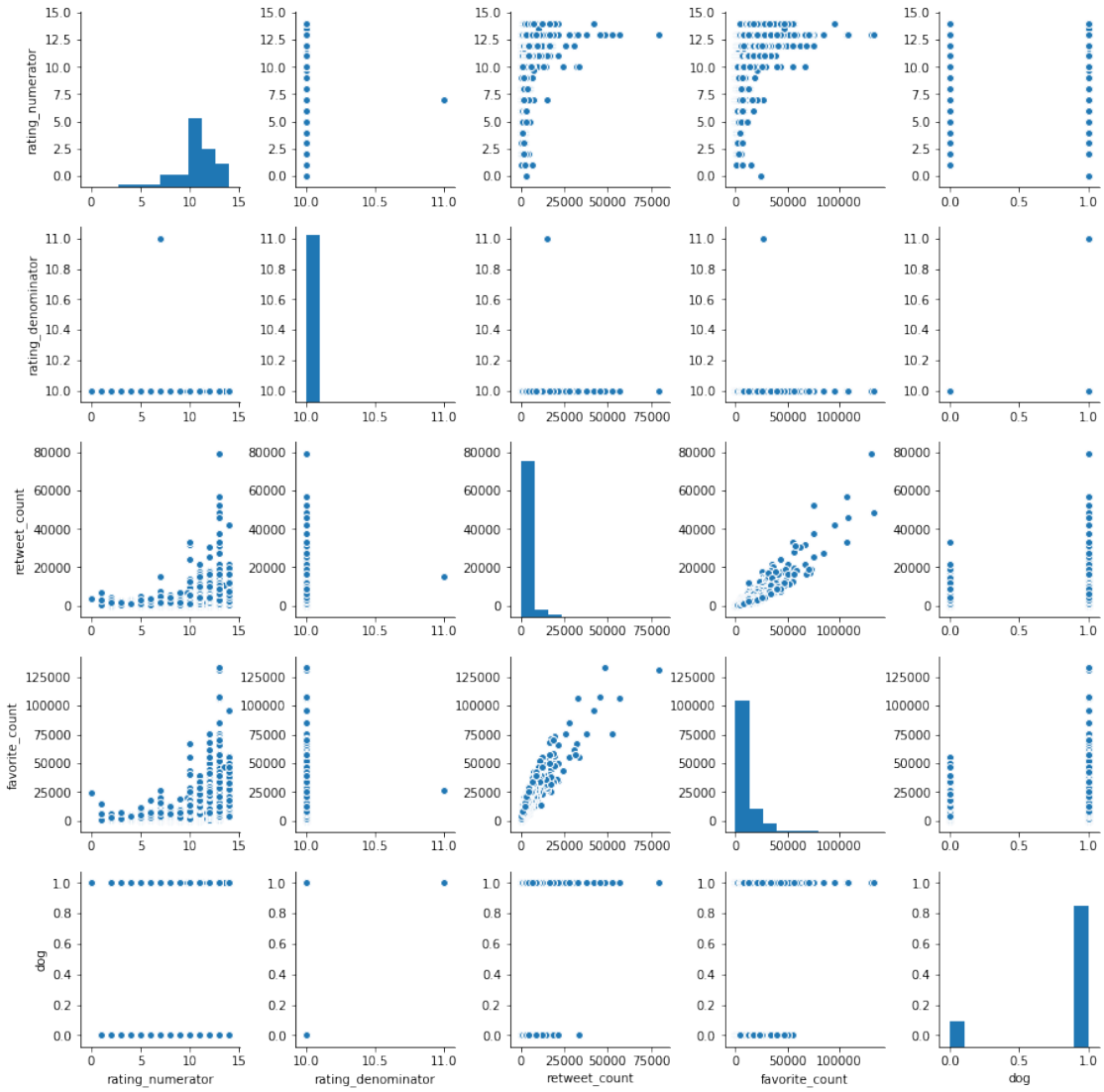
```
In [106]: df_tw = df_tw_plt.loc[df_tw_plt['rating_numerator'] < 20]
```

```
In [107]: df_tw.info()
```

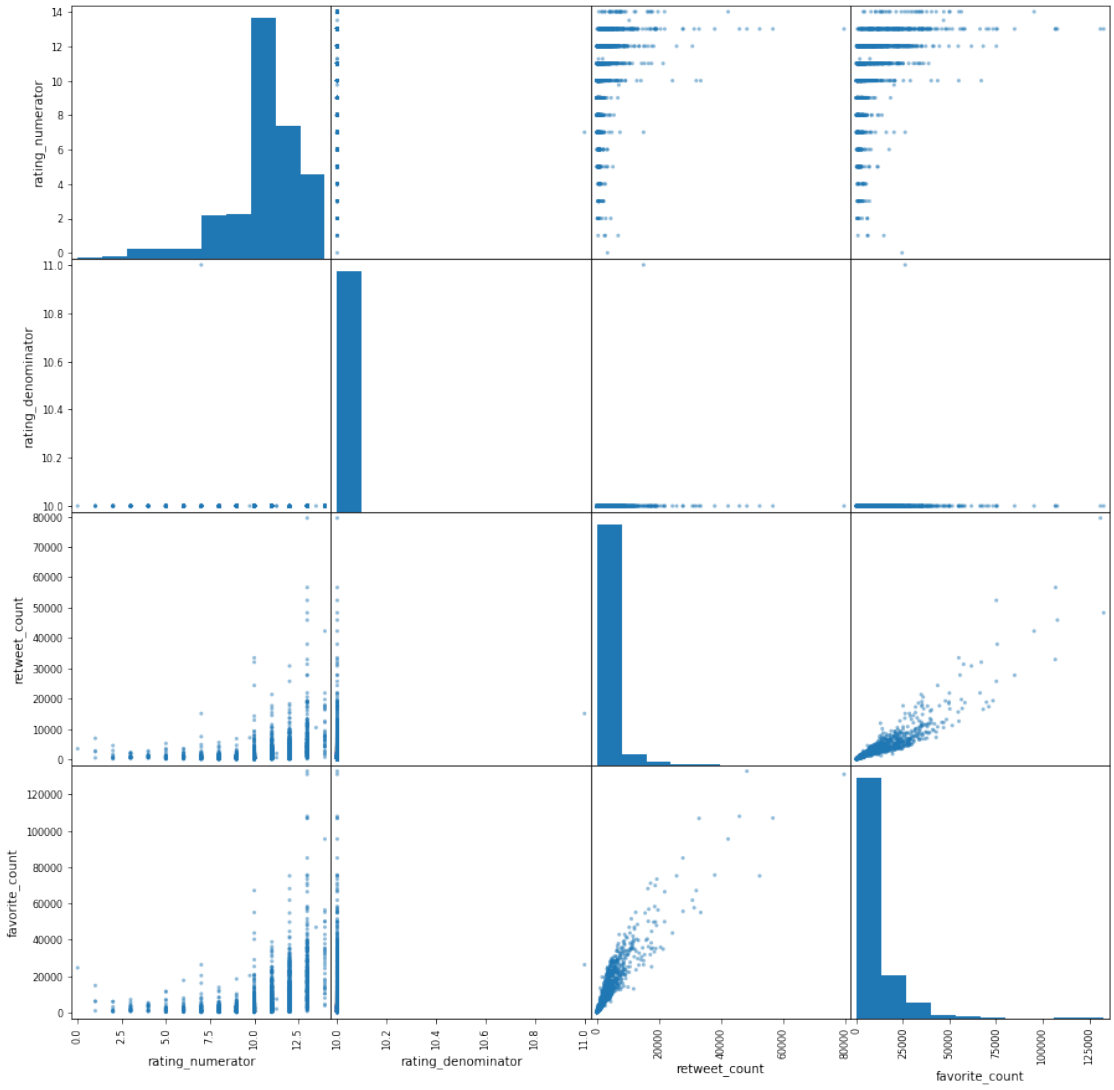
```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1968 entries, 2015-11-15 22:32:08 to 2017-08-01 16:23:56
Data columns (total 9 columns):
source                1968 non-null object
rating_numerator      1968 non-null float64
rating_denominator    1968 non-null float64
name                  1968 non-null object
dog_stage             303 non-null object
retweet_count         1968 non-null float64
favorite_count        1968 non-null float64
dog_breed             1968 non-null object
dog                   1968 non-null object
dtypes: float64(4), object(5)
memory usage: 153.8+ KB
```

2.5 2.1. exploratory chart

```
In [108]: sb.pairplot(df_tw)
          plt.savefig('figures/exploring.png')
```



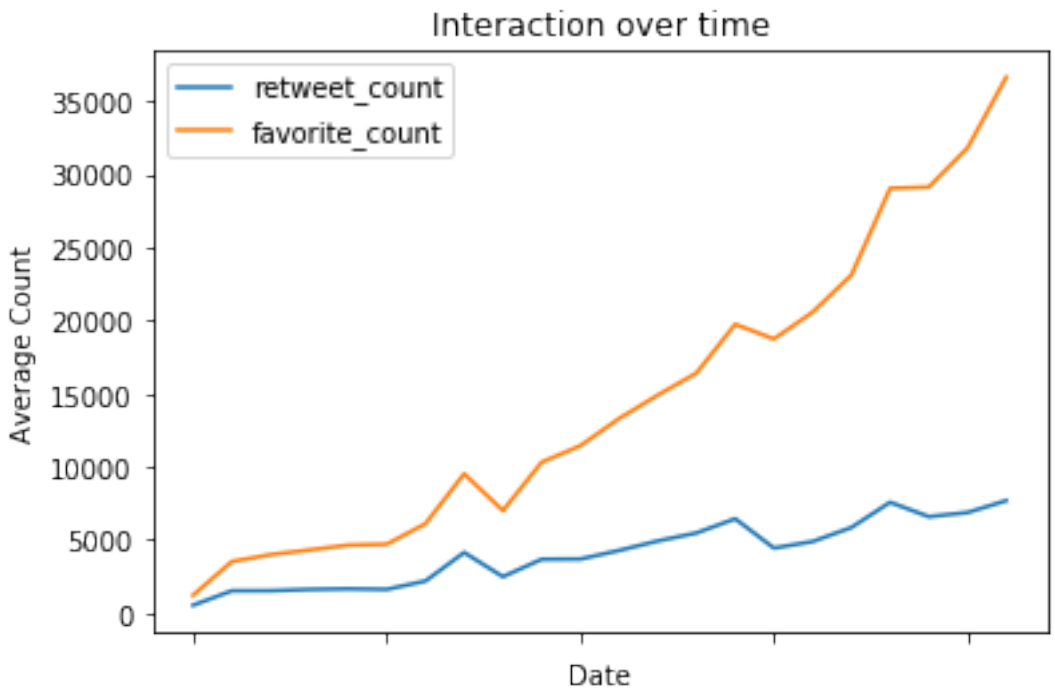

```
In [109]: pd.plotting.scatter_matrix(df_tw, figsize=(15,15))
          plt.savefig('figures/exploring2.png');
```



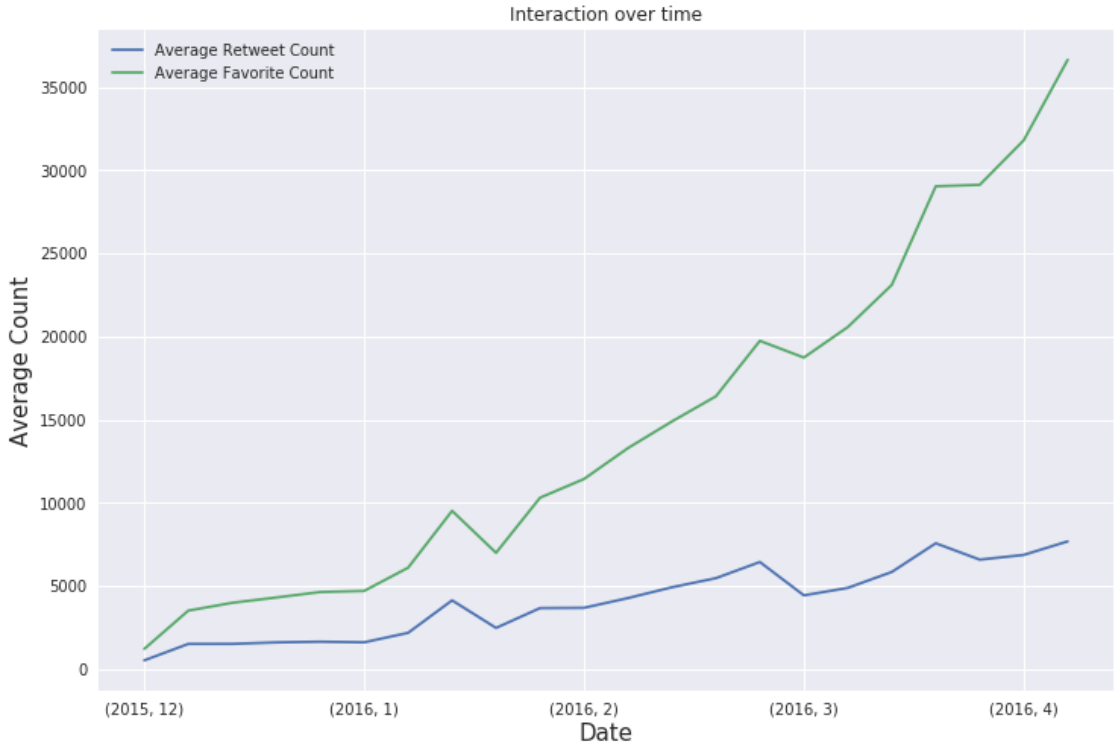
2.6 2.2. Profile over time

2.6.1 2.2.1. Interaction over time

```
In [110]: av_tw_month = df_tw.groupby([(df_tw.index.year),(df_tw.index.month)]).retweet_count.me
          av_fv_month = df_tw.groupby([(df_tw.index.year),(df_tw.index.month)]).favorite_count.m

          av_tw_month.plot(kind = 'line', title = 'Interaction over time', legend = True)
          av_fv_month.plot(kind = 'line', title = 'Interaction over time', legend = True)
          plt.xlabel('Date')
          plt.ylabel('Average Count')
          plt.xticks();
          plt.savefig('figures/Interaction over time.png')
```

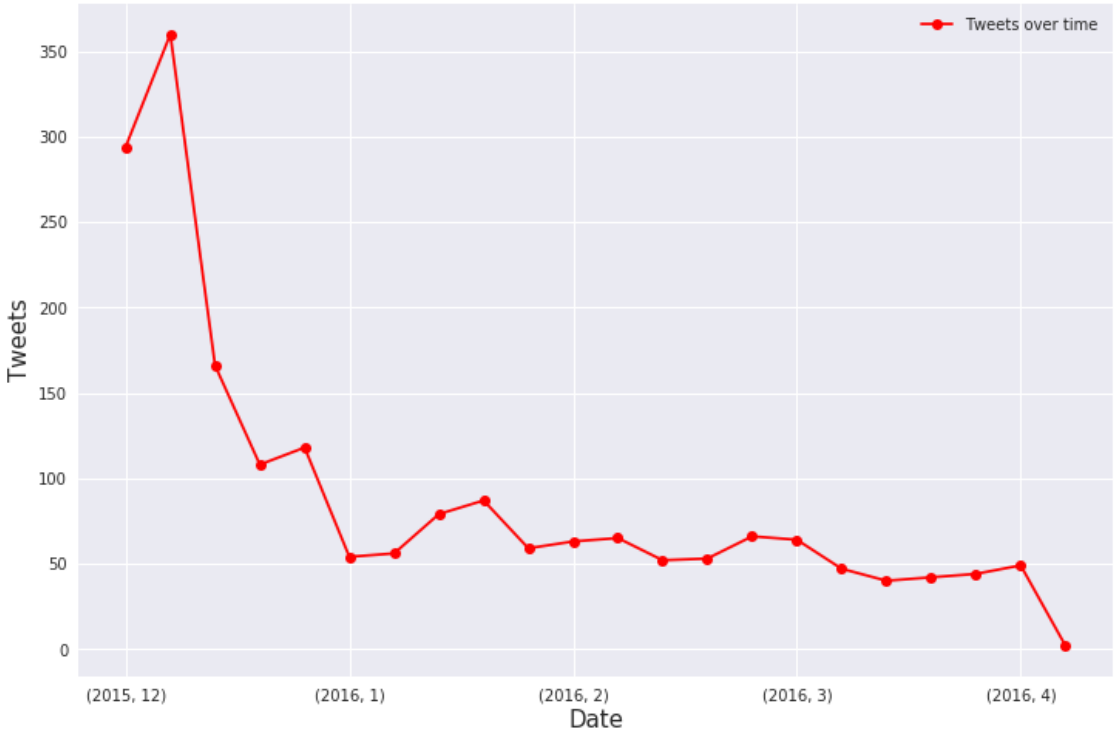


In [111]:
```
av_tw_month = df_tw.groupby([(df_tw.index.year),(df_tw.index.month)]).retweet_count.me
av_fv_month = df_tw.groupby([(df_tw.index.year),(df_tw.index.month)]).favorite_count.m


sb.set_context('notebook')
sb.set(rc={'figure.figsize':(12,8)})
fig, ax = plt.subplots()
ax = av_tw_month.plot(kind = 'line', label = 'Average Retweet Count', legend = True)
ax = av_fv_month.plot(kind = 'line', label = 'Average Favorite Count', legend = True)
plt.xlabel('Date', fontsize=15)
plt.ylabel('Average Count', fontsize=15)
plt.legend()
plt.title('Interaction over time')
ax.set_xticklabels(av_tw_month.index);
plt.savefig('figures/Interaction over time2.png')
```



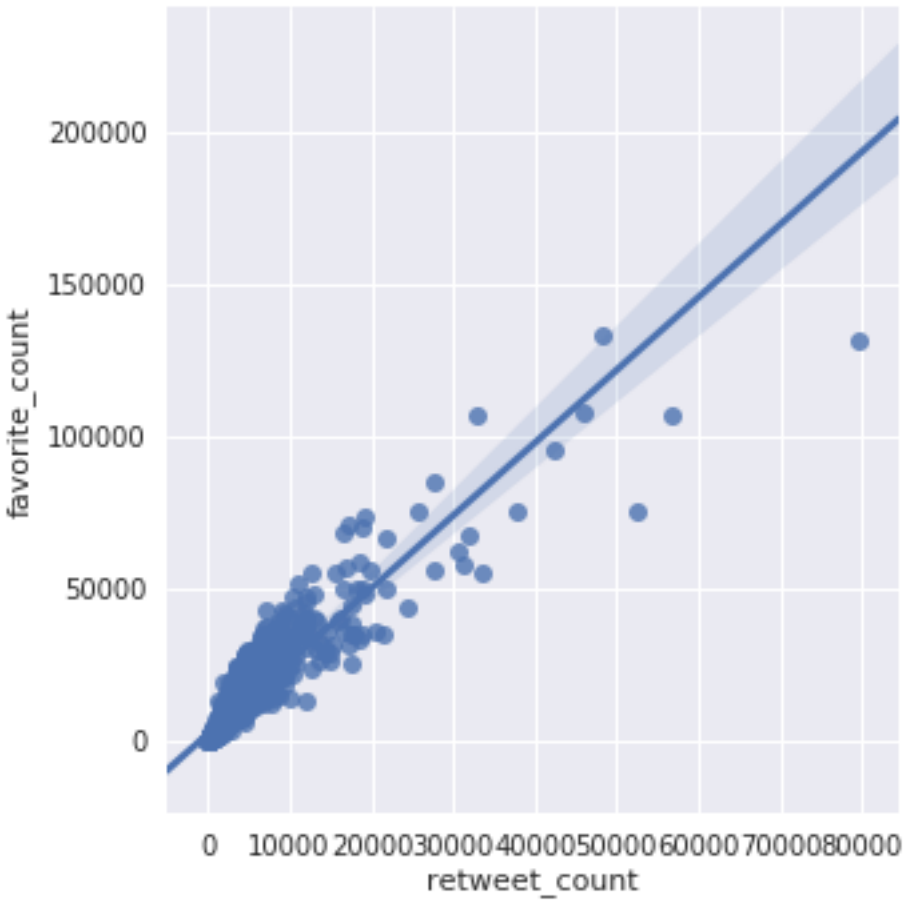
2.6.2 2.2.2. Average tweets count over time

```
In [112]: twt_month = df_tw.groupby([(df_tw.index.year),(df_tw.index.month)]).rating_denominator

          fig, ax = plt.subplots()
          twt_month.plot(style = '-ro', figsize = (12,8), label = 'Tweets over time')
          plt.xlabel('Date', fontsize=15)
          plt.ylabel('Tweets', fontsize=15)
          plt.legend()
          ax.set_xticklabels(twt_month.index);
          plt.savefig('figures/Average tweets count over time.png')
```

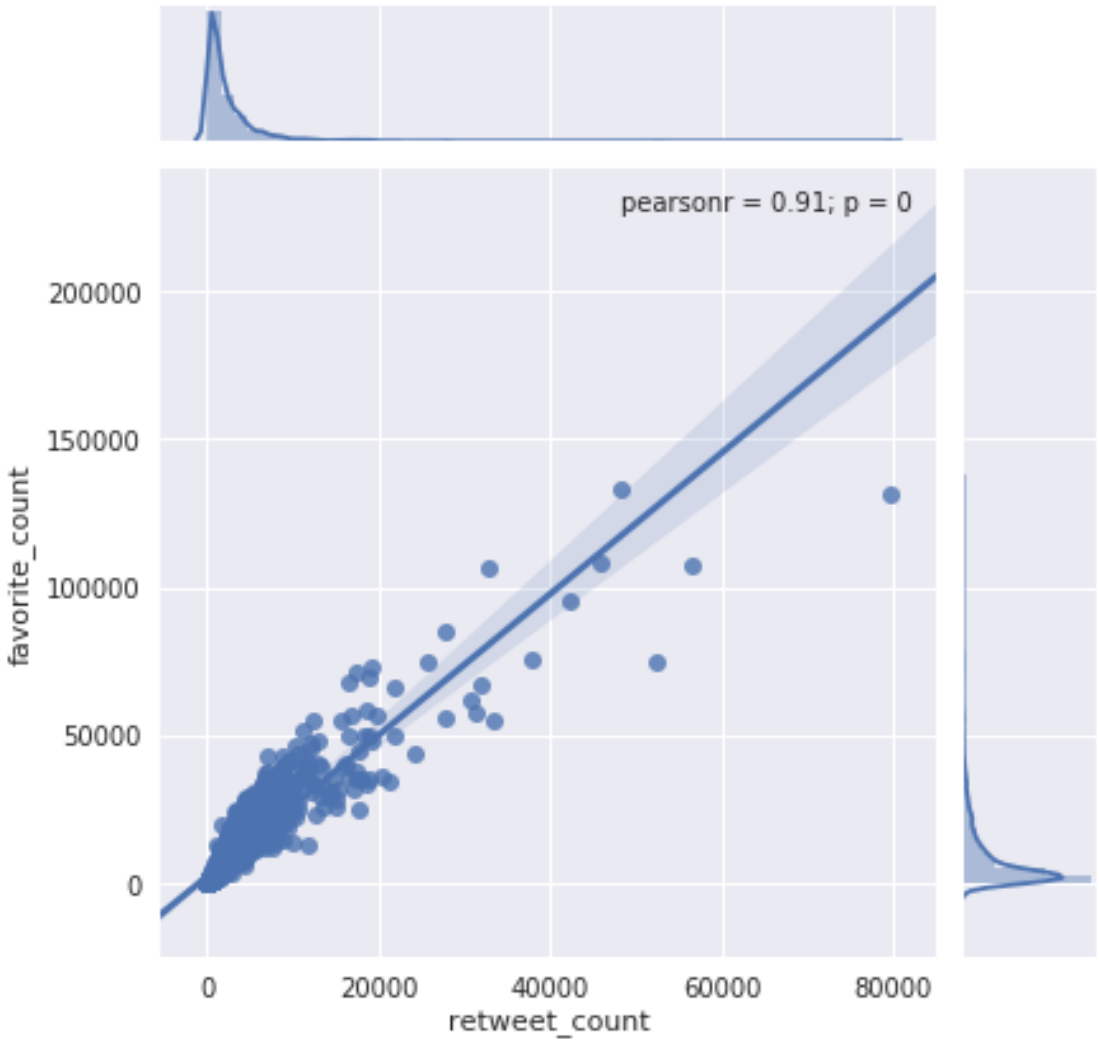


2.6.3 2.2.3. correlation matrix between retweets and favorite

```
In [113]: sb.lmplot(data= df_tw, x= 'retweet_count', y='favorite_count');
          plt.savefig('figures/correlation matrix between retweets and favorite.png')
```

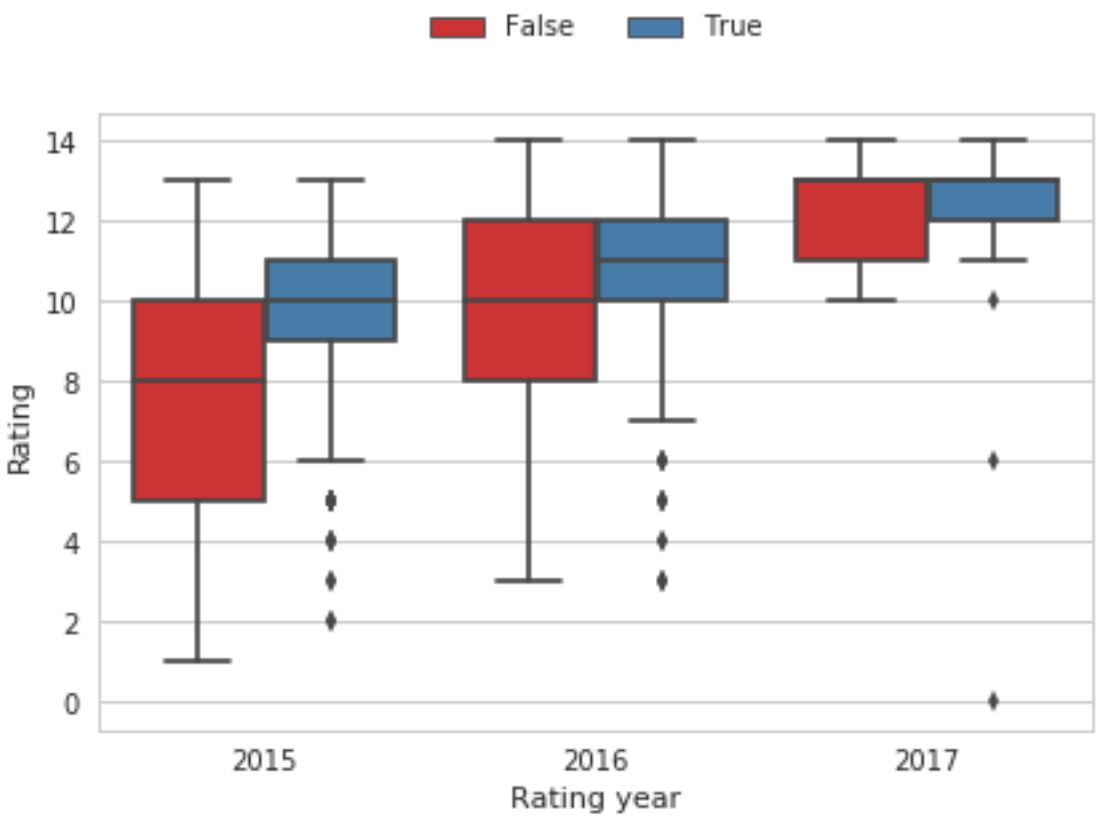



```
In [114]: sb.jointplot(data= df_tw, x= 'retweet_count', y='favorite_count', kind='reg');
          plt.savefig('figures/correlation matrix between retweets and favorite2.png')
```



2.6.4 2.2.3. Interactions of images of dogs vs not dog images

```
In [115]: sb.set_context('notebook')
          sb.set_style("whitegrid")
          plt.subplots(figsize=(6,4))
          sb.boxplot(df_tw.index.year, df_tw.rating_numerator , hue=df_tw.dog, palette= "Set1");
          plt.legend(loc=8)
          plt.legend(loc='upper center', bbox_to_anchor=(0.5,1.2), ncol=3, fancybox = True, shad
          plt.xlabel('Rating year');plt.ylabel('Rating');
          plt.tight_layout()
          plt.savefig('figures/interactions of images of dogs vs not dog images over time.png')
```

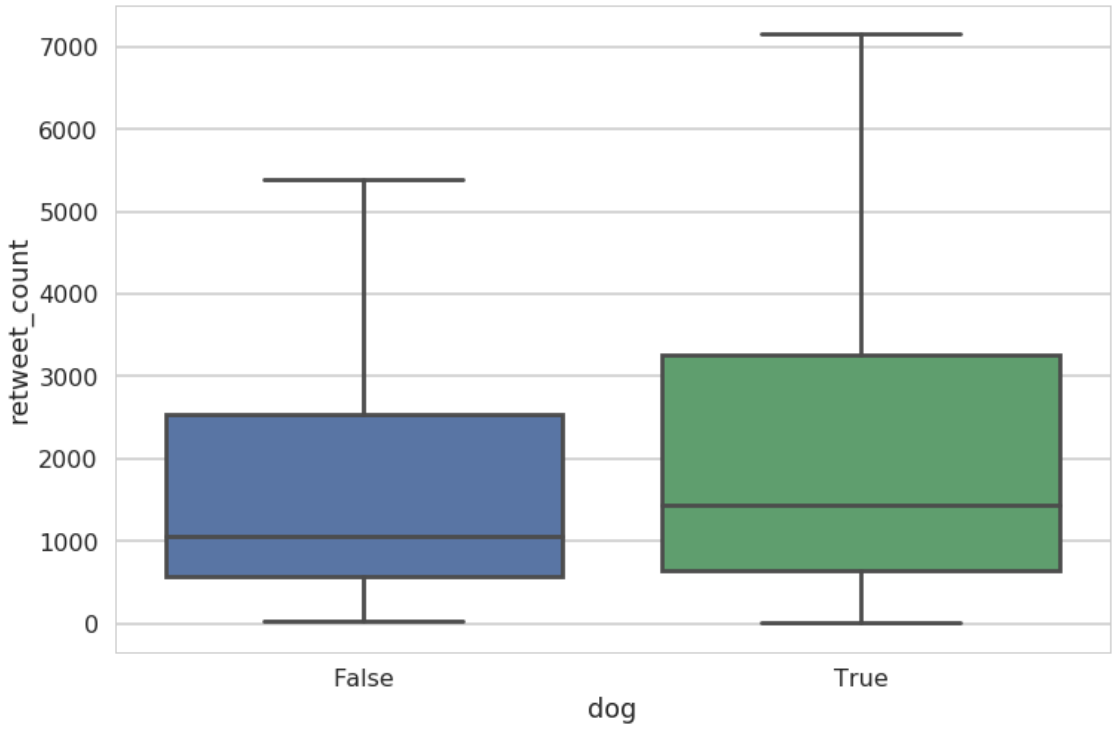



```
In [116]: df_tw.groupby(['dog']).retweet_count.describe()
```

Out[116]:

dog	count	mean	std	min	25%	50%	75%	max
False	303.0	2504.600660	3889.878411	34.0	573.0	1067.0	2541.5	33421.0
True	1665.0	2835.138138	4833.325952	16.0	650.0	1440.0	3261.0	79515.0

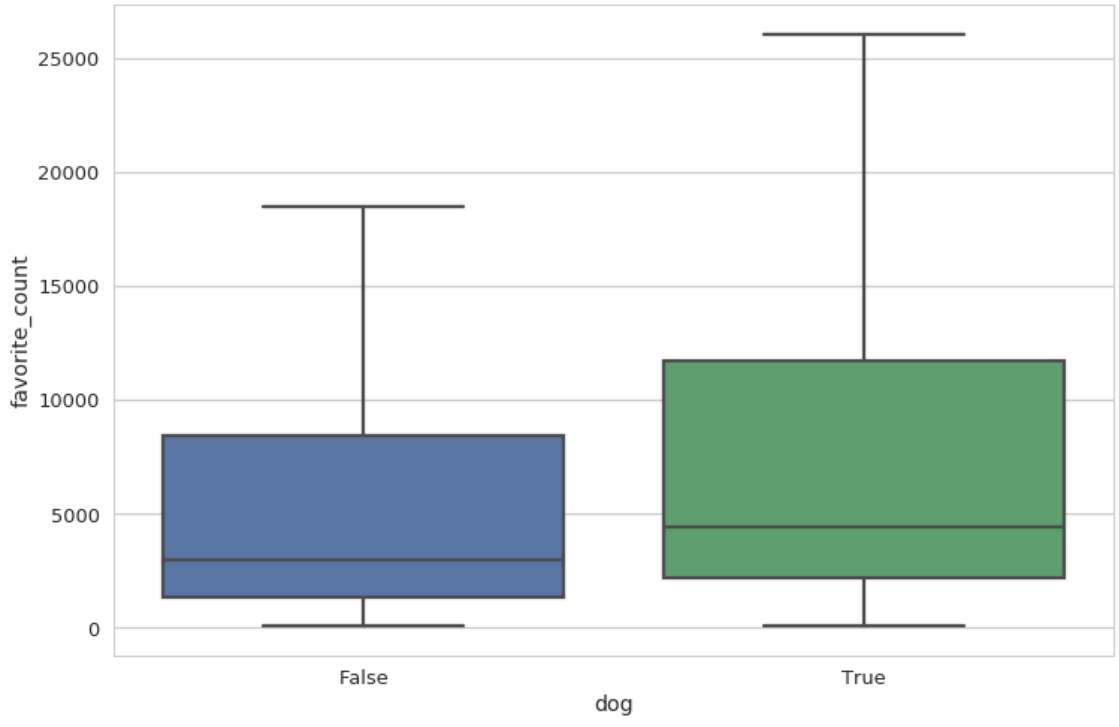
```
In [117]: sb.set_context('poster')
          ax=sb.boxplot(x='dog', y='retweet_count' , data=df_tw, showfliers= False);
          plt.savefig('figures/Retweets of images of dogs vs not dog images.png')
```




```
In [118]: sb.set_context('talk')
          ax=sb.boxplot(x='dog', y='favorite_count' , data=df_tw, showfliers= False)
          plt.setp(ax.get_xticklabels());
          plt.savefig('figures/Favorite of images of dogs vs not dog images.png')
```

```
  agg_filter: unknown
  alpha: float (0.0 transparent through 1.0 opaque)
  animated: [True | False]
  backgroundcolor: any matplotlib color
  bbox: FancyBboxPatch prop dict
  clip_box: a :class:`matplotlib.transforms.Bbox` instance
  clip_on: [True | False]
  clip_path: [ (:class:`~matplotlib.path.Path`, :class:`~matplotlib.transforms.Transform`) | :cl
  color: any matplotlib color
  contains: a callable function
  family or fontfamily or fontname or name: [FONTNAME | 'serif' | 'sans-serif' | 'cursive' | 'fa
  figure: a :class:`matplotlib.figure.Figure` instance
  fontproperties or font_properties: a :class:`matplotlib.font_manager.FontProperties` instance
  gid: an id string
  horizontalalignment or ha: [ 'center' | 'right' | 'left' ]
  label: string or anything printable with '%s' conversion.
  linespacing: float (multiple of font size)
  multialignment: ['left' | 'right' | 'center' ]
  path_effects: unknown
```

```
picker: [None|float|boolean|callable]
position: (x,y)
rasterized: [True | False | None]
rotation: [ angle in degrees | 'vertical' | 'horizontal' ]
rotation_mode: unknown
size or fontsize: [size in points | 'xx-small' | 'x-small' | 'small' | 'medium' | 'large' | 'x
sketch_params: unknown
snap: unknown
stretch or fontstretch: [a numeric value in range 0-1000 | 'ultra-condensed' | 'extra-condense
style or fontstyle: [ 'normal' | 'italic' | 'oblique']
text: string or anything printable with '%s' conversion.
transform: :class:`~matplotlib.transforms.Transform` instance
url: a url string
usetex: unknown
variant or fontvariant: [ 'normal' | 'small-caps' ]
verticalalignment or ma or va: [ 'center' | 'top' | 'bottom' | 'baseline' ]
visible: [True | False]
weight or fontweight: [a numeric value in range 0-1000 | 'ultralight' | 'light' | 'normal' | '
wrap: unknown
x: float
y: float
zorder: any number
```



2.6.5 2.2.4. Average rating over time

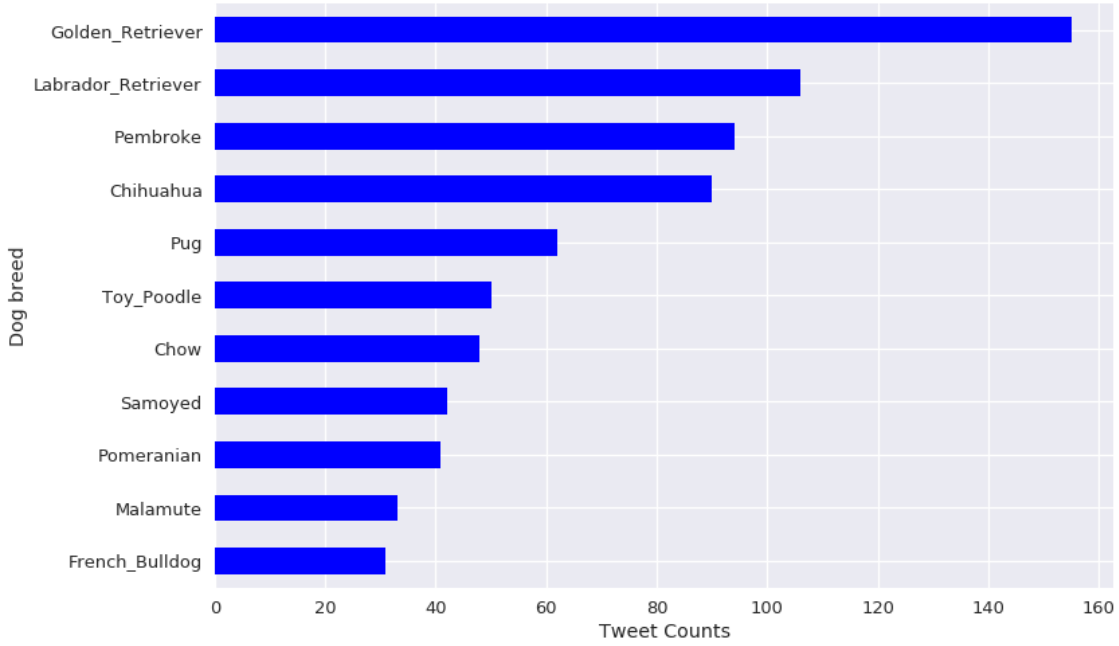
```
In [119]: tw_month_rating = df_tw.groupby([(df_tw.index.year),(df_tw.index.month)]).rating_numer

          sb.set_context('notebook')
          sb.set(rc={'figure.figsize':(12,8)})
          fig, ax = plt.subplots()

          tw_month_rating.plot(style = '-ro', figsize = (12,8), label = 'Dogs rating over time')
          plt.axhline(y=10.0, color='b', linestyle='--', label='Out of rating')

          plt.xlabel('Date', fontsize=15)
          plt.ylabel('Average Rating out of 10', fontsize=15)
          plt.legend()
          ax.set_xticklabels(tw_month_rating.index);
          plt.savefig('figures/Average rating over time.png')
```

2.7 2.3. Dog Breeds

2.7.1 2.3.1. Tweets counts for dog breeds

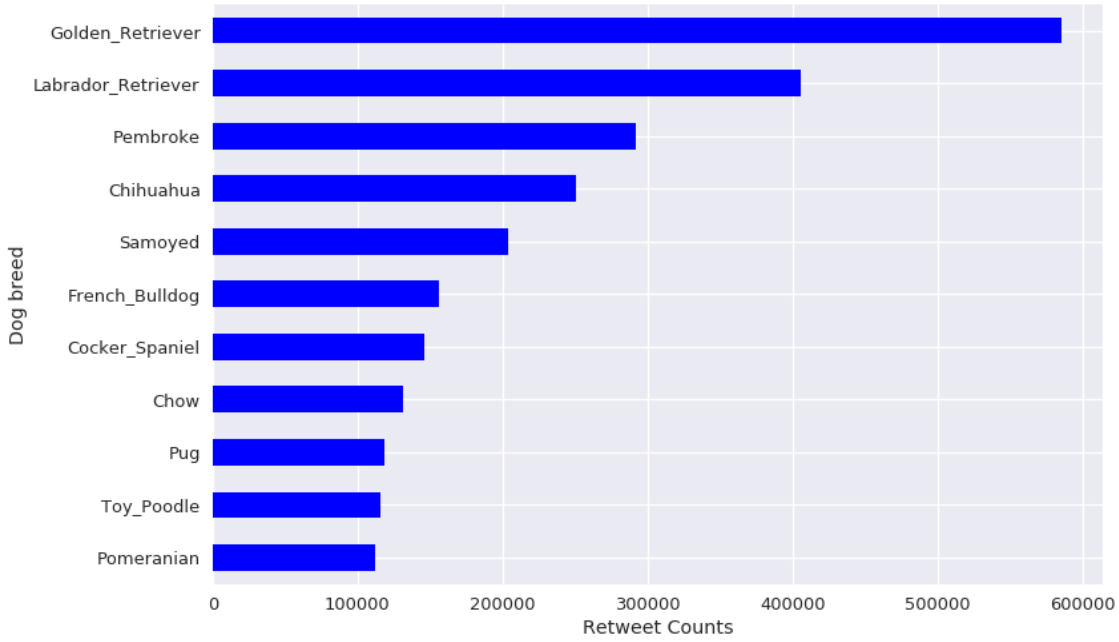
```
In [120]: sb.set_context('talk')
          df_tw[df_tw.dog == True].dog_breed.value_counts()[10: :-1].plot(kind='barh', color ='b
          plt.xlabel('Tweet Counts')
```

```
plt.ylabel('Dog breed');
plt.savefig('figures/Tweets counts for dog breeds.png')
```

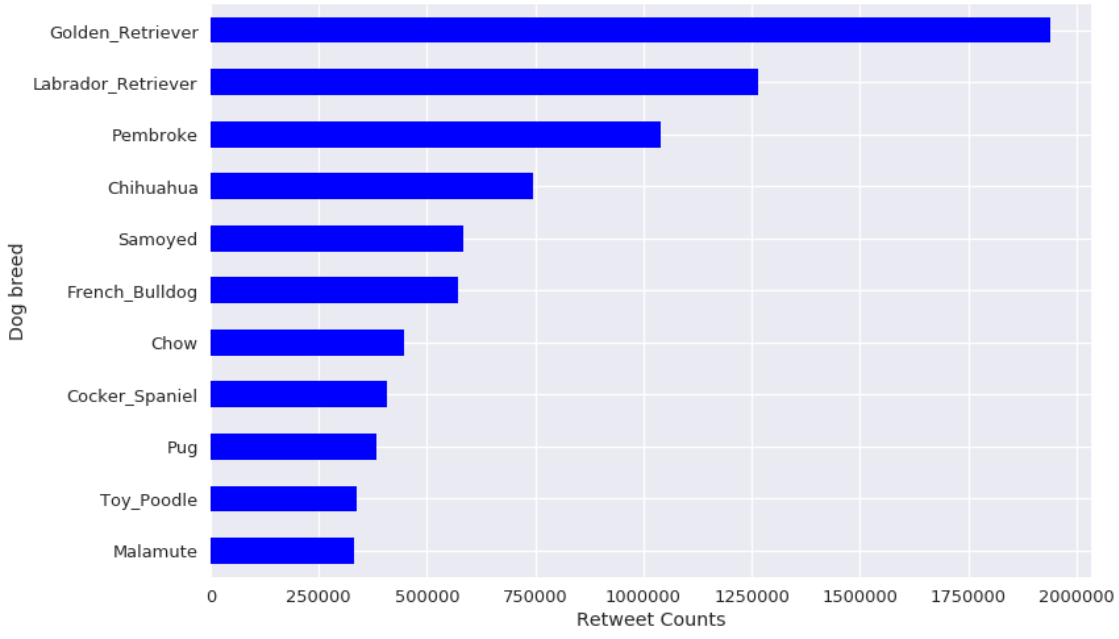


2.7.2 2.3.2. Retweet & favorite counts for dog breeds

```
In [121]: sb.set_context('talk')
          df_tw[df_tw.dog == True].groupby(['dog_breed']).retweet_count.sum().sort_values(ascend
          plt.xlabel('Retweet Counts')
          plt.ylabel('Dog breed');
          plt.savefig('figures/Retweet counts for dog breeds.png')
```

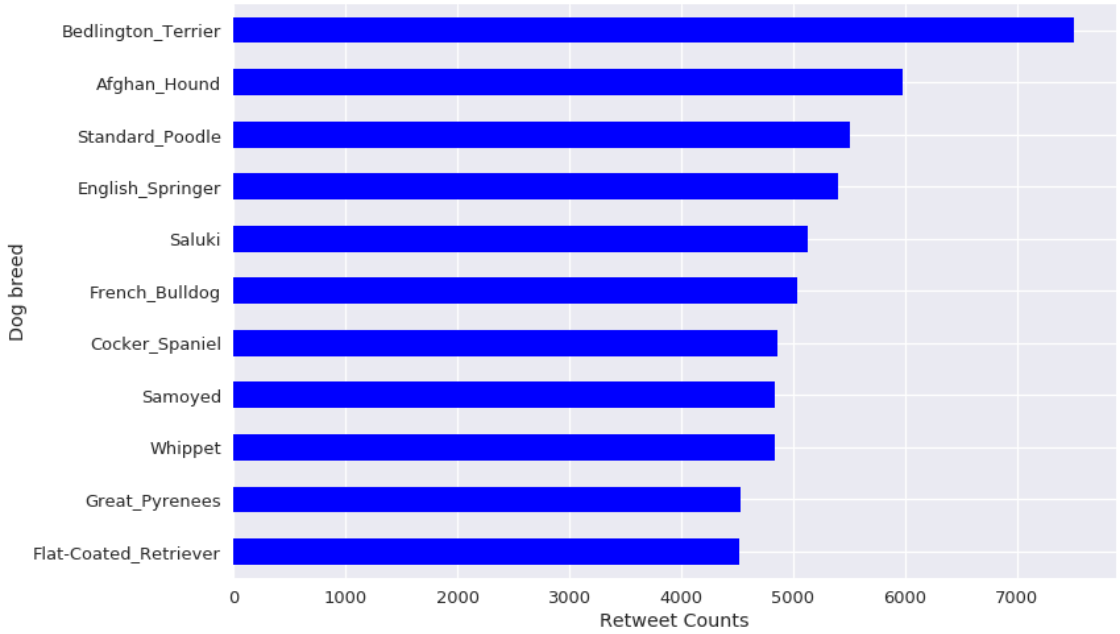



```
In [122]: sb.set_context('talk')
          df_tw[df_tw.dog == True].groupby(['dog_breed']).favorite_count.sum().sort_values(ascen
          plt.xlabel('Retweet Counts')
          plt.ylabel('Dog breed');
          plt.savefig('figures/Favorite counts for dog breeds.png')
```

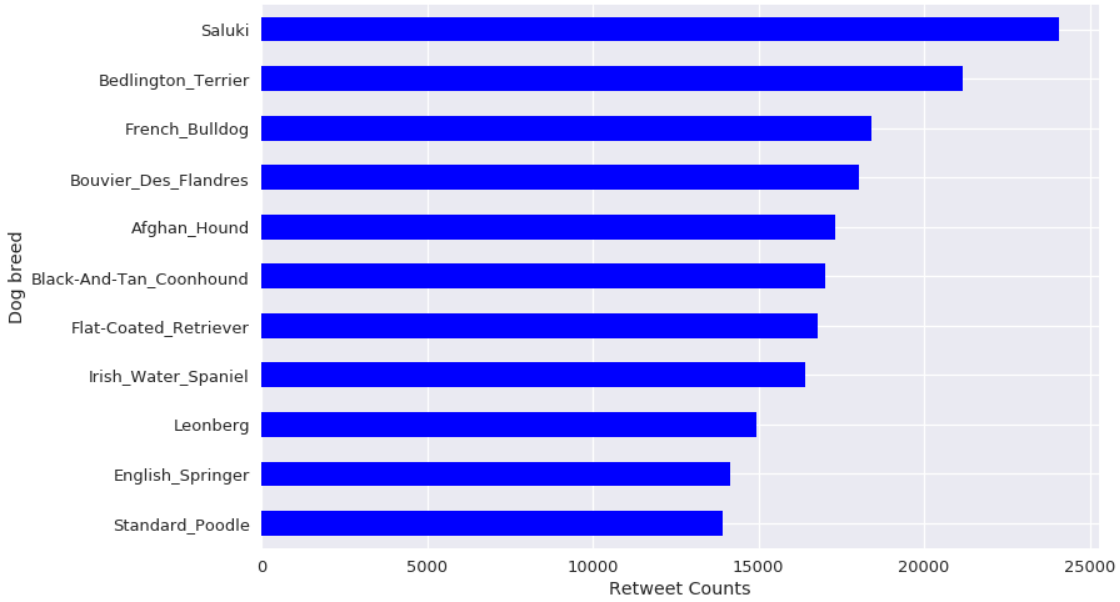


2.7.3 2.3.3. Retweet & favorite average for dog breeds

```
In [123]: sb.set_context('talk')
          df_tw[df_tw.dog == True].groupby(['dog_breed']).retweet_count.mean().sort_values(ascen
          plt.xlabel('Retweet Counts')
          plt.ylabel('Dog breed');
          plt.savefig('figures/Retweet average for dog breeds.png')
```



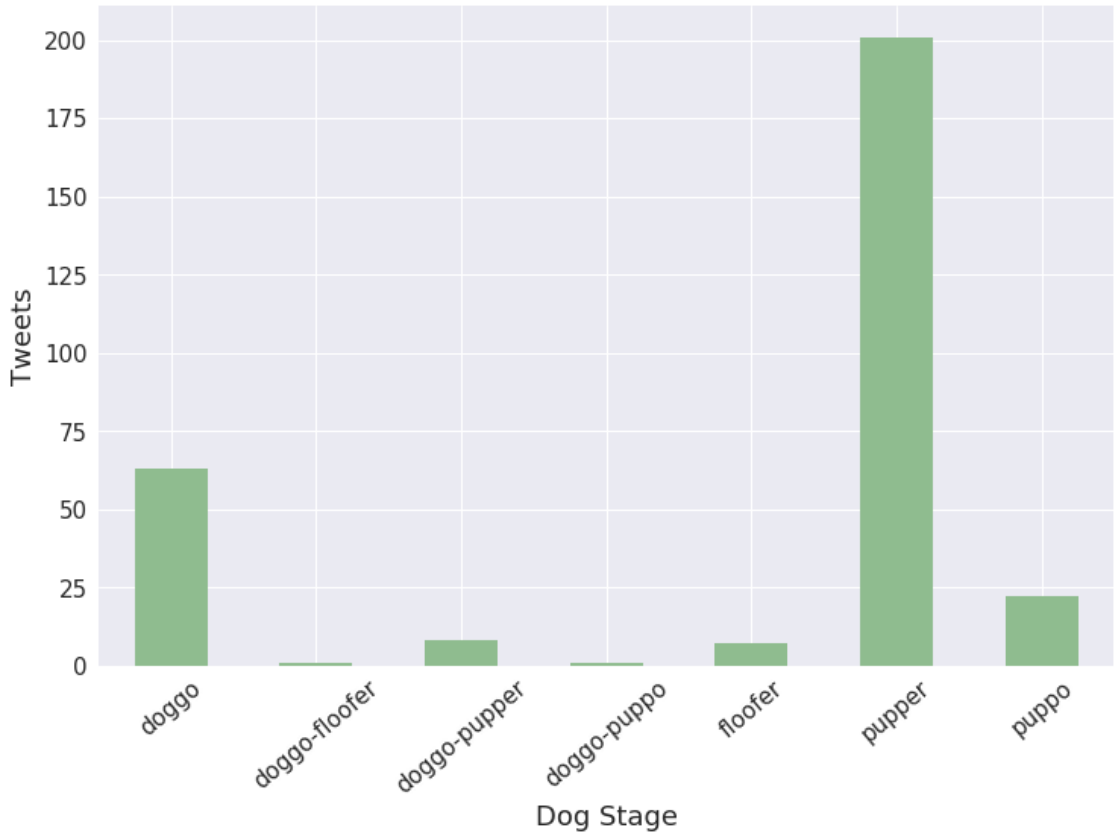

```
In [124]: sb.set_context('talk')
          df_tw[df_tw.dog == True].groupby(['dog_breed']).favorite_count.mean().sort_values(asce
          plt.xlabel('Retweet Counts')
          plt.ylabel('Dog breed');
          plt.savefig('figures/Favorite average for dog breeds.png')
```



2.8 2.4. Dog Stages

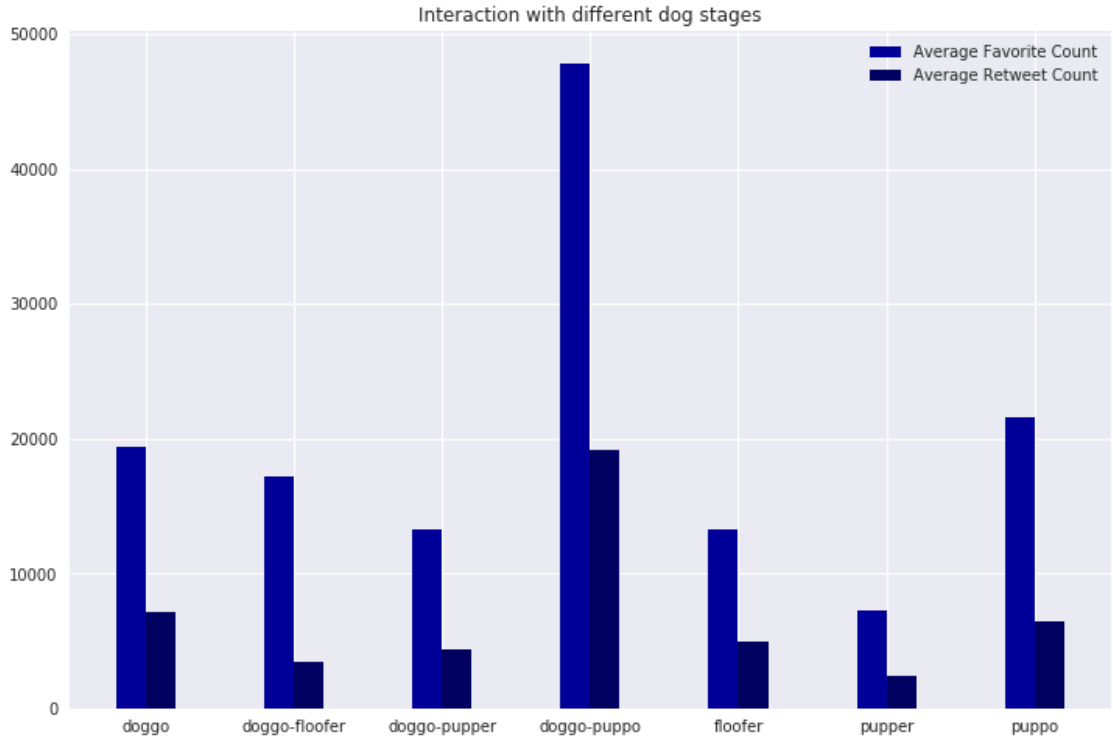
2.8.1 2.4.1. Tweets Counts for dog stages

```
In [125]: twt_month_stages = df_tw.groupby('dog_stage').rating_denominator.count()

          sb.set_context('talk')
          sb.set(rc={'figure.figsize':(12,8)})
          ind = np.arange(len(twt_month_stages))
          twt_month_stages.plot(kind= 'bar',fontsize=15, color= 'darkseagreen', rot=40)
          plt.xlabel('Dog Stage', fontsize=18)
          plt.ylabel('Tweets', fontsize=18);
          plt.savefig('figures/Tweets Counts for dog stages.png')
```



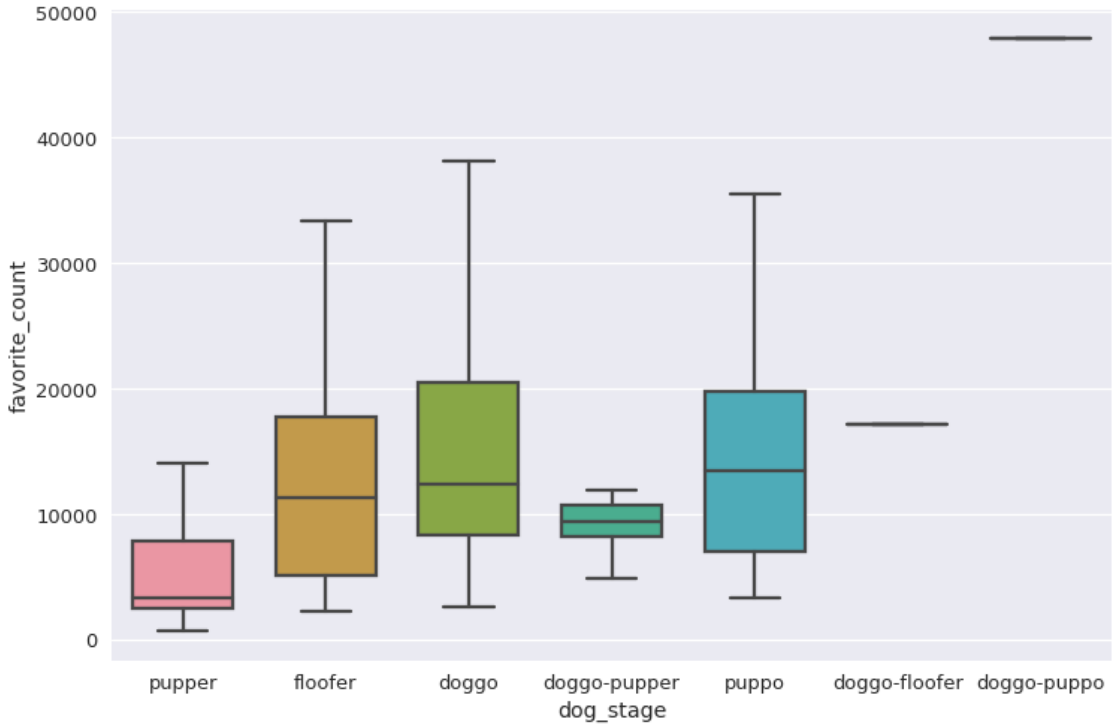
2.8.2 2.4.2. Retweet & favorite for dog stages

```
In [126]: avg_fav_stage = df_tw.groupby('dog_stage').favorite_count.mean()
          avg_retweet_stage = df_tw.groupby('dog_stage').retweet_count.mean()

          sb.set_context('notebook')
          sb.set(rc={'figure.figsize':(12,8)})
          width=0.2
          ind = np.arange(len(avg_fav_stage))
          locations = ind + width/2
          labels = avg_fav_stage.index

          plt.bar(ind, avg_fav_stage, width, color= '#000099', label= 'Average Favorite Count')
          plt.bar(ind+width, avg_retweet_stage, width, color= '#000060', label= 'Average Retweet
          plt.xticks(locations, labels)
          plt.legend()
          plt.title('Interaction with different dog stages');
          plt.savefig('figures/Retweet & favorite for dog stages.png')
```



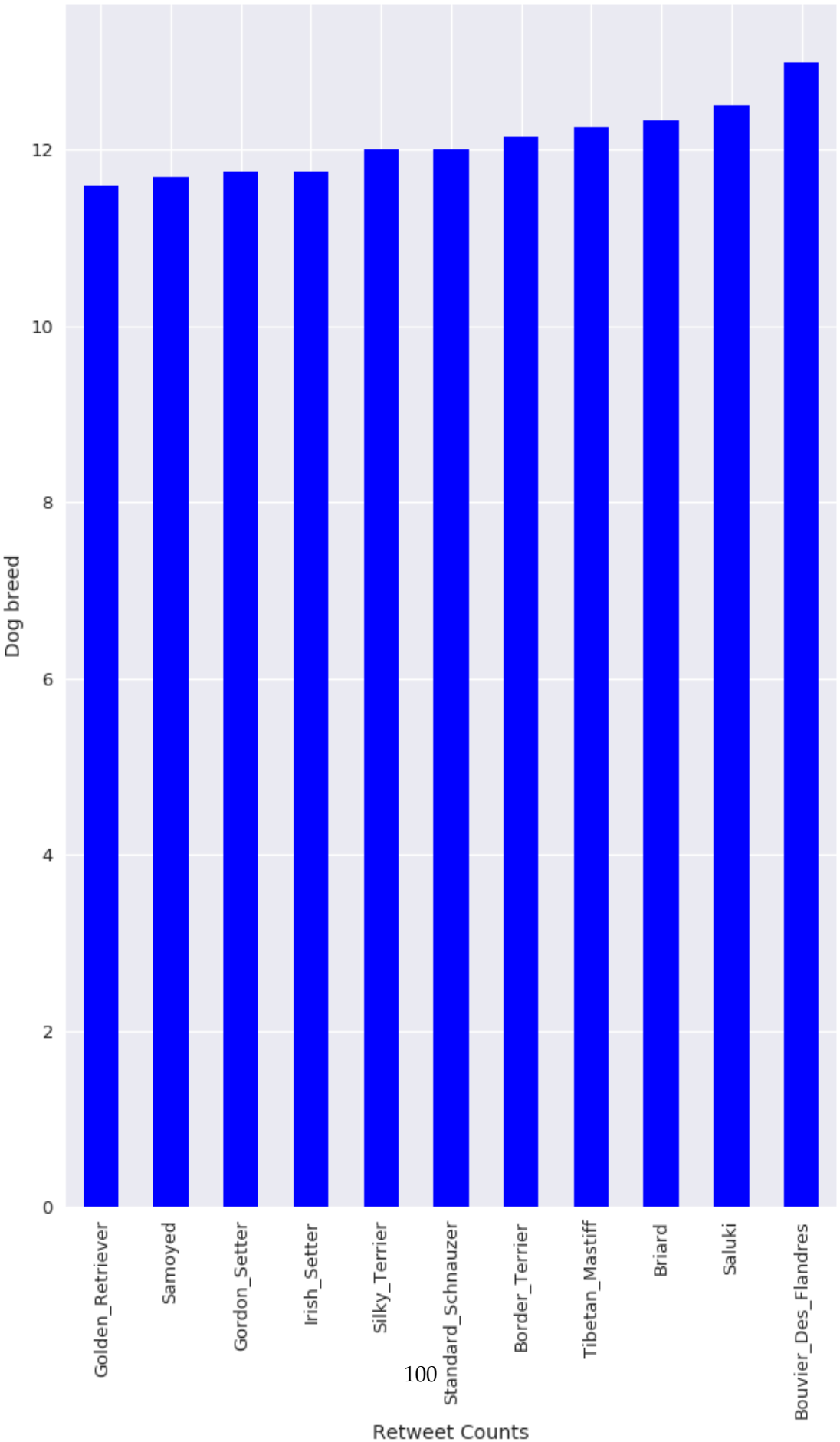
interacrtion against dog stages

```
In [127]: sb.set_context('talk')
          ax=sb.boxplot(x='dog_stage', y='favorite_count' , data=df_tw, showfliers= False, width
          sb.set(rc={'figure.figsize':(10,16)});
          plt.savefig('figures/Favorite for dog stages.png')
```

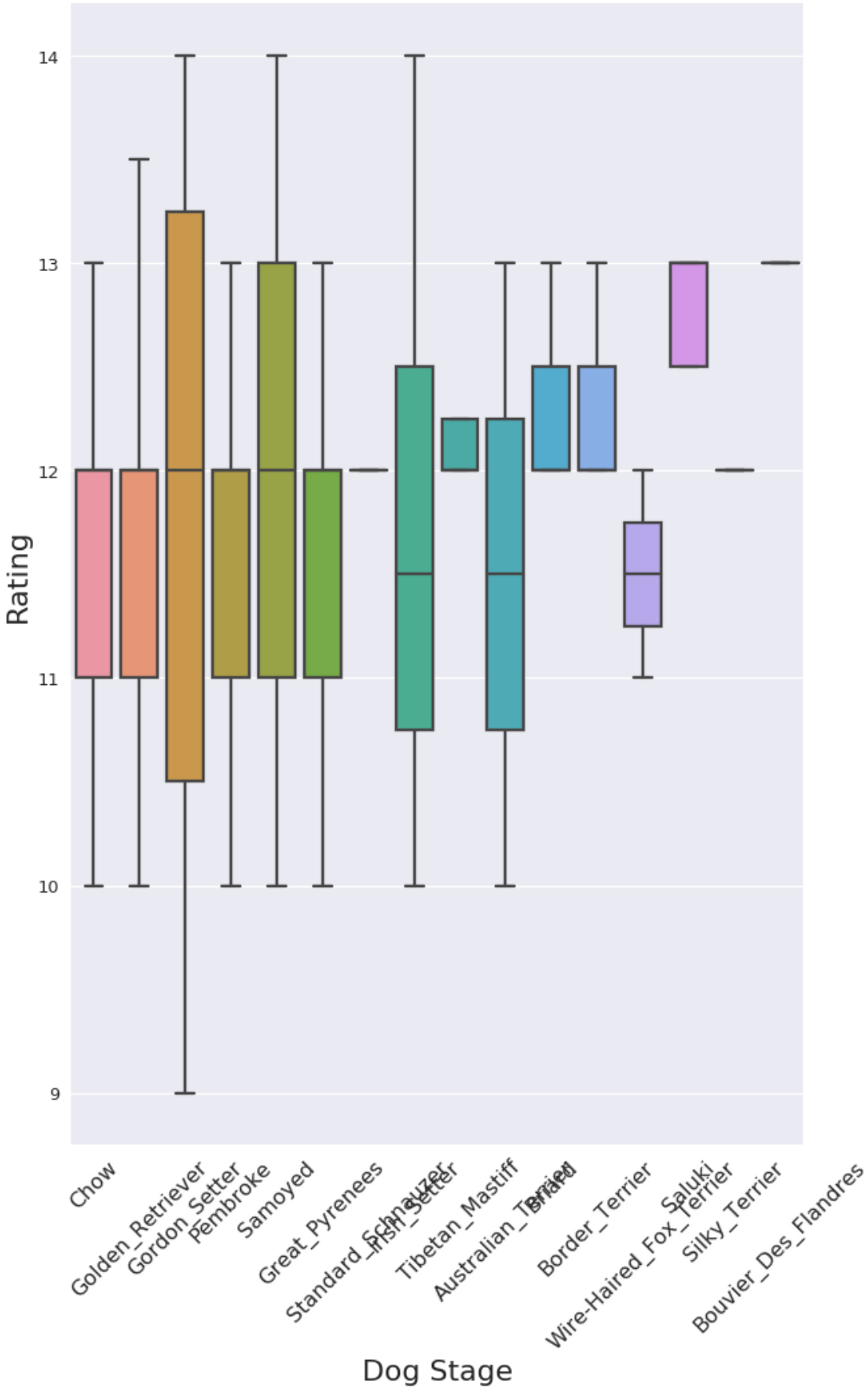


2.8.3 2.4.3. Average Rating for dog stages

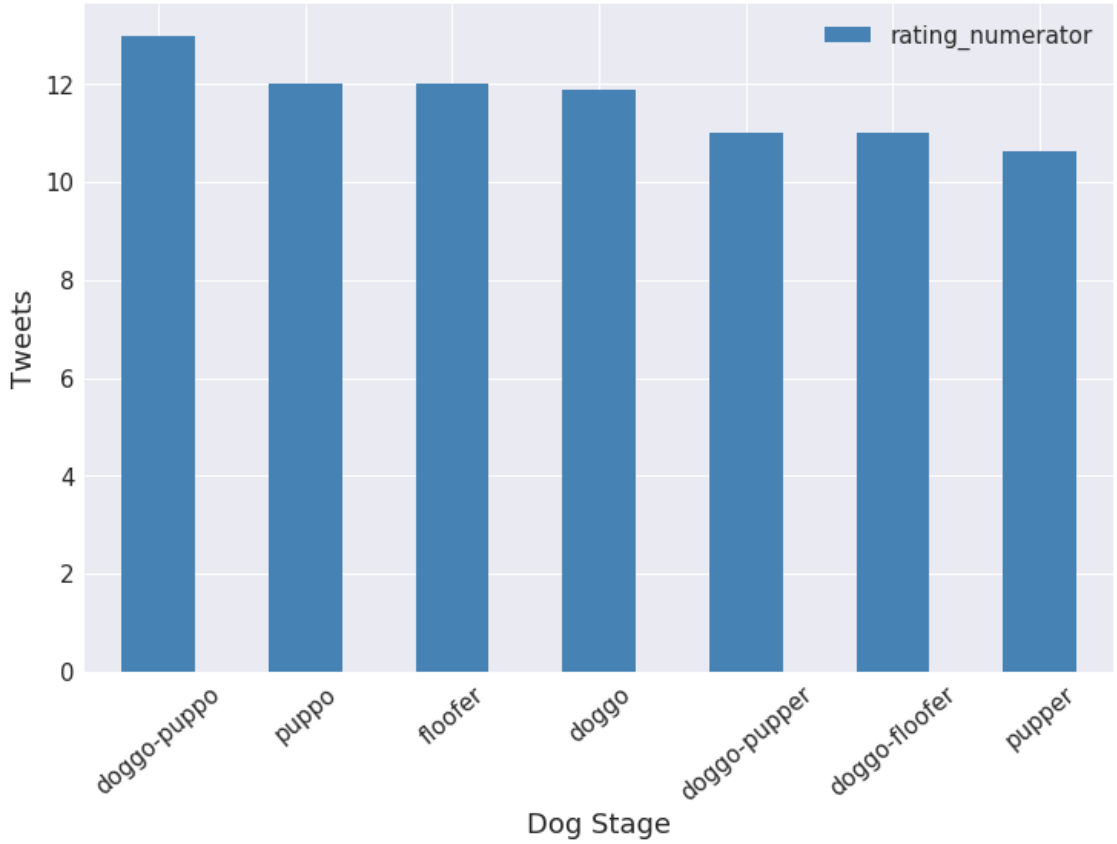
```
In [128]: sb.set_context('talk')
          df_tw[df_tw.dog == True].groupby(['dog_breed']).rating_numerator.mean().sort_values(as
          plt.xlabel('Retweet Counts')
          plt.ylabel('Dog breed');
          plt.savefig('figures/Average Rating for dog stages-bar.png')
```


Dog breed
0
2
4
6
8
10
12
Golden_Retriever
Samoyed
Gordon_Setter
Irish_Setter
Silky_Terrier
Standard_Schnauzer
Border_Terrier
Tibetan_Mastiff
Briard
Saluki
Bouvier_Des_Flandres
Retweet Counts

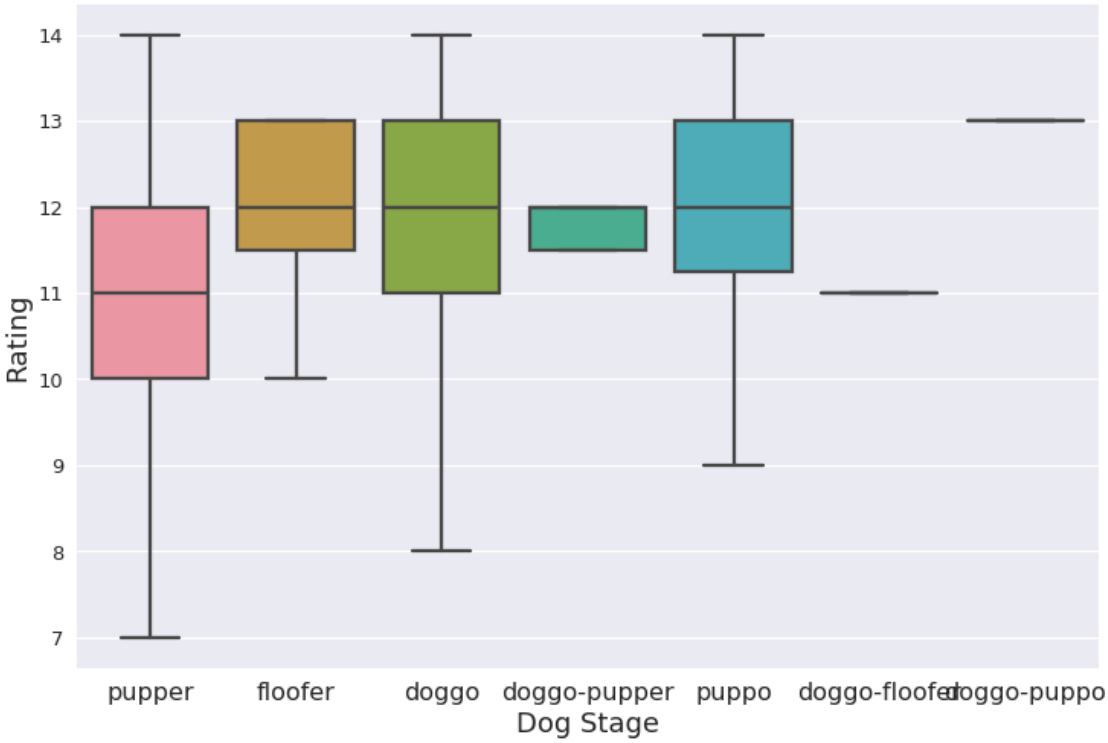
In [129]:
```
breed_filter = df_tw[df_tw.dog == True].groupby(['dog_breed']).rating_numerator.mean()


rating_breed = df_tw[df_tw.dog_breed.isin(breed_filter.index)]
sb.set_context('talk')
ax=sb.boxplot(x= 'dog_breed', y='rating_numerator' , data=rating_breed, showfliers= Fa
sb.set(rc={'figure.figsize':(10,16)})
plt.xlabel('Dog Stage', fontsize=22)
plt.ylabel('Rating', fontsize=22)
plt.setp(ax.get_xticklabels(), fontsize=16, rotation=45);
plt.savefig('figures/Average Rating for dog stages-box.png')
```


Rating
14
13
12
11
10
9
Chow
Golden_Retriever
Gordon_Setter
Pembroke
Samoyed
Great_Pyrenees
Tibetan_Mastiff
Border_Terrier
Saluki
Wire-Haired_Fox_Terrier
Silky_Terrier
Bouvier_Des_Flandres
Dog Stage

In [130]:
```
twt_rating_stages = df_tw.groupby('dog_stage').rating_numerator.mean().sort_values(asc

sb.set_context('notebook')
sb.set(rc={'figure.figsize':(12,8)})
ind = np.arange(len(twt_rating_stages))
twt_rating_stages.plot(kind= 'bar',fontsize=15, color= 'steelblue',sort_columns=True,
plt.xlabel('Dog Stage', fontsize=18)
plt.ylabel('Tweets', fontsize=18)
plt.legend(fontsize=15);
plt.savefig('figures/Average Rating for dog stages-bar2.png')
```



In [131]:
```
sb.set_context('talk')
ax=sb.boxplot(x='dog_stage', y='rating_numerator' , data=df_tw, showfliers= False,)
sb.set(rc={'figure.figsize':(18,16)})
plt.xlabel('Dog Stage', fontsize=18)
plt.ylabel('Rating', fontsize=18)
plt.setp(ax.get_xticklabels(), fontsize=16, rotation=0);
plt.savefig('figures/Average Rating for dog stages- box2.png')
```


14
13
12
11
10
9
8
7
Rating
pupper
floofer
doggo
doggo-pupper
puppo
doggo-floofer
doggo-puppo
Dog Stage