wrangle_act

February 10, 2021

1 Gathering the data

In [1]:
```
# Import essential libraries
import pandas as pd
import requests
import tweepy
import json
import time
import matplotlib.pyplot as plt
import seaborn as sns
```

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

In [3]:
```
r = requests.get(url)
with  open('image_predictions.tsv', 'wb') as f:
    f.write(r.content)
```

In [4]:
```
twitter_data= pd.read_csv('twitter-archive-enhanced.csv')
prediction_data= pd.read_csv('image_predictions.tsv',sep='\t')
```

In [5]:
```
twitter_data
```

Out[5]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	NaN	NaN	
1	892177421306343426	NaN	NaN	
2	891815181378084864	NaN	NaN	
3	891689557279858688	NaN	NaN	
4	891327558926688256	NaN	NaN	
5	891087950875897856	NaN	NaN	
6	890971913173991426	NaN	NaN	
7	890729181411237888	NaN	NaN	
8	890609185150312448	NaN	NaN	
9	890240255349198849	NaN	NaN	
10	890006608113172480	NaN	NaN	
11	889880896479866881	NaN	NaN	
12	889665388333682689	NaN	NaN	
13	889638837579907072	NaN	NaN	
14	889531135344209921	NaN	NaN	

```
15    889278841981685760                NaN                NaN
16    888917238123831296                NaN                NaN
17    888804989199671297                NaN                NaN
18    888554962724278272                NaN                NaN
19    888202515573088257                NaN                NaN
20    888078434458587136                NaN                NaN
21    887705289381826560                NaN                NaN
22    887517139158093824                NaN                NaN
23    887473957103951883                NaN                NaN
24    887343217045368832                NaN                NaN
25    887101392804085760                NaN                NaN
26    886983233522544640                NaN                NaN
27    886736880519319552                NaN                NaN
28    886680336477933568                NaN                NaN
29    886366144734445568                NaN                NaN
...                  ...                ...                ...
2326  666411507551481857                NaN                NaN
2327  666407126856765440                NaN                NaN
2328  666396247373291520                NaN                NaN
2329  666373753744588802                NaN                NaN
2330  666362758909284353                NaN                NaN
2331  666353288456101888                NaN                NaN
2332  666345417576210432                NaN                NaN
2333  666337882303524864                NaN                NaN
2334  666293911632134144                NaN                NaN
2335  666287406224695296                NaN                NaN
2336  666273097616637952                NaN                NaN
2337  666268910803644416                NaN                NaN
2338  666104133288665088                NaN                NaN
2339  666102155909144576                NaN                NaN
2340  666099513787052032                NaN                NaN
2341  666094000022159362                NaN                NaN
2342  666082916733198337                NaN                NaN
2343  666073100786774016                NaN                NaN
2344  666071193221509120                NaN                NaN
2345  666063827256086533                NaN                NaN
2346  666058600524156928                NaN                NaN
2347  666057090499244032                NaN                NaN
2348  666055525042405380                NaN                NaN
2349  666051853826850816                NaN                NaN
2350  666050758794694657                NaN                NaN
2351  666049248165822465                NaN                NaN
2352  666044226329800704                NaN                NaN
2353  666033412701032449                NaN                NaN
2354  666029285002620928                NaN                NaN
2355  666020888022790149                NaN                NaN

                    timestamp  \
```

```
0     2017-08-01 16:23:56 +0000
1     2017-08-01 00:17:27 +0000
2     2017-07-31 00:18:03 +0000
3     2017-07-30 15:58:51 +0000
4     2017-07-29 16:00:24 +0000
5     2017-07-29 00:08:17 +0000
6     2017-07-28 16:27:12 +0000
7     2017-07-28 00:22:40 +0000
8     2017-07-27 16:25:51 +0000
9     2017-07-26 15:59:51 +0000
10    2017-07-26 00:31:25 +0000
11    2017-07-25 16:11:53 +0000
12    2017-07-25 01:55:32 +0000
13    2017-07-25 00:10:02 +0000
14    2017-07-24 17:02:04 +0000
15    2017-07-24 00:19:32 +0000
16    2017-07-23 00:22:39 +0000
17    2017-07-22 16:56:37 +0000
18    2017-07-22 00:23:06 +0000
19    2017-07-21 01:02:36 +0000
20    2017-07-20 16:49:33 +0000
21    2017-07-19 16:06:48 +0000
22    2017-07-19 03:39:09 +0000
23    2017-07-19 00:47:34 +0000
24    2017-07-18 16:08:03 +0000
25    2017-07-18 00:07:08 +0000
26    2017-07-17 16:17:36 +0000
27    2017-07-16 23:58:41 +0000
28    2017-07-16 20:14:00 +0000
29    2017-07-15 23:25:31 +0000
...                          ...
2326  2015-11-17 00:24:19 +0000
2327  2015-11-17 00:06:54 +0000
2328  2015-11-16 23:23:41 +0000
2329  2015-11-16 21:54:18 +0000
2330  2015-11-16 21:10:36 +0000
2331  2015-11-16 20:32:58 +0000
2332  2015-11-16 20:01:42 +0000
2333  2015-11-16 19:31:45 +0000
2334  2015-11-16 16:37:02 +0000
2335  2015-11-16 16:11:11 +0000
2336  2015-11-16 15:14:19 +0000
2337  2015-11-16 14:57:41 +0000
2338  2015-11-16 04:02:55 +0000
2339  2015-11-16 03:55:04 +0000
2340  2015-11-16 03:44:34 +0000
2341  2015-11-16 03:22:39 +0000
2342  2015-11-16 02:38:37 +0000
```

```
2343  2015-11-16 01:59:36 +0000
2344  2015-11-16 01:52:02 +0000
2345  2015-11-16 01:22:45 +0000
2346  2015-11-16 01:01:59 +0000
2347  2015-11-16 00:55:59 +0000
2348  2015-11-16 00:49:46 +0000
2349  2015-11-16 00:35:11 +0000
2350  2015-11-16 00:30:50 +0000
2351  2015-11-16 00:24:50 +0000
2352  2015-11-16 00:04:52 +0000
2353  2015-11-15 23:21:54 +0000
2354  2015-11-15 23:05:30 +0000
2355  2015-11-15 22:32:08 +0000

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

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

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

```
18    This is Ralphus. He's powering up. Attempting ...                NaN
19    RT @dog_rates: This is Canela. She attempted s...       8.874740e+17
20    This is Gerald. He was just told he didn't get...                NaN
21    This is Jeffrey. He has a monopoly on the pool...                NaN
22    I've yet to rate a Venezuelan Hover Wiener. Th...                NaN
23    This is Canela. She attempted some fancy porch...                NaN
24    You may not have known you needed to see this ...                NaN
25    This... is a Jubilant Antarctic House Bear. We...                NaN
26    This is Maya. She's very shy. Rarely leaves he...                NaN
27    This is Mingus. He's a wonderful father to his...                NaN
28    This is Derek. He's late for a dog meeting. 13...                NaN
29    This is Roscoe. Another pupper fallen victim t...                NaN
...                                                 ...                ...
2326  This is quite the dog. Gets really excited whe...                NaN
2327  This is a southern Vesuvius bumblegruff. Can d...                NaN
2328  Oh goodness. A super rare northeast Qdoba kang...                NaN
2329  Those are sunglasses and a jean jacket. 11/10 ...                NaN
2330  Unique dog here. Very small. Lives in containe...                NaN
2331  Here we have a mixed Asiago from the Galápagos...                NaN
2332  Look at this jokester thinking seat belt laws ...                NaN
2333  This is an extremely rare horned Parthenon. No...                NaN
2334  This is a funny dog. Weird toes. Won't come do...                NaN
2335  This is an Albanian 3 1/2 legged  Episcopalian...                NaN
2336     Can take selfies 11/10 https://t.co/ws2AMaNwPW                NaN
2337  Very concerned about fellow dog trapped in com...                NaN
2338  Not familiar with this breed. No tail (weird)...                NaN
2339  Oh my. Here you are seeing an Adobe Setter giv...                NaN
2340  Can stand on stump for what seems like a while...                NaN
2341  This appears to be a Mongolian Presbyterian mi...                NaN
2342  Here we have a well-established sunblockerspan...                NaN
2343  Let's hope this flight isn't Malaysian (lol). ...                NaN
2344  Here we have a northern speckled Rhododendron...                 NaN
2345  This is the happiest dog you will ever see. Ve...                NaN
2346  Here is the Rand Paul of retrievers folks! He'...                NaN
2347  My oh my. This is a rare blond Canadian terrie...                NaN
2348  Here is a Siberian heavily armored polar bear ...                NaN
2349  This is an odd dog. Hard on the outside but lo...                NaN
2350  This is a truly beautiful English Wilson Staff...                NaN
2351  Here we have a 1949 1st generation vulpix. Enj...                NaN
2352  This is a purebred Piers Morgan. Loves to Netf...                NaN
2353  Here is a very happy pup. Big fan of well-main...                NaN
2354  This is a western brown Mitsubishi terrier. Up...                NaN
2355  Here we have a Japanese Irish Setter. Lost eye...                NaN

      retweeted_status_user_id retweeted_status_timestamp  \
0                          NaN                         NaN
1                          NaN                         NaN
2                          NaN                         NaN
```

3	NaN	NaN
4	NaN	NaN
5	NaN	NaN
6	NaN	NaN
7	NaN	NaN
8	NaN	NaN
9	NaN	NaN
10	NaN	NaN
11	NaN	NaN
12	NaN	NaN
13	NaN	NaN
14	NaN	NaN
15	NaN	NaN
16	NaN	NaN
17	NaN	NaN
18	NaN	NaN
19	4.196984e+09	2017-07-19 00:47:34 +0000
20	NaN	NaN
21	NaN	NaN
22	NaN	NaN
23	NaN	NaN
24	NaN	NaN
25	NaN	NaN
26	NaN	NaN
27	NaN	NaN
28	NaN	NaN
29	NaN	NaN
...	...	...
2326	NaN	NaN
2327	NaN	NaN
2328	NaN	NaN
2329	NaN	NaN
2330	NaN	NaN
2331	NaN	NaN
2332	NaN	NaN
2333	NaN	NaN
2334	NaN	NaN
2335	NaN	NaN
2336	NaN	NaN
2337	NaN	NaN
2338	NaN	NaN
2339	NaN	NaN
2340	NaN	NaN
2341	NaN	NaN
2342	NaN	NaN
2343	NaN	NaN
2344	NaN	NaN
2345	NaN	NaN

```
2346                       NaN                       NaN
2347                       NaN                       NaN
2348                       NaN                       NaN
2349                       NaN                       NaN
2350                       NaN                       NaN
2351                       NaN                       NaN
2352                       NaN                       NaN
2353                       NaN                       NaN
2354                       NaN                       NaN
2355                       NaN                       NaN

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

```
2331  https://twitter.com/dog_rates/status/666353288...                8
2332  https://twitter.com/dog_rates/status/666345417...               10
2333  https://twitter.com/dog_rates/status/666337882...                9
2334  https://twitter.com/dog_rates/status/666293911...                3
2335  https://twitter.com/dog_rates/status/666287406...                1
2336  https://twitter.com/dog_rates/status/666273097...               11
2337  https://twitter.com/dog_rates/status/666268910...               10
2338  https://twitter.com/dog_rates/status/666104133...                1
2339  https://twitter.com/dog_rates/status/666102155...               11
2340  https://twitter.com/dog_rates/status/666099513...                8
2341  https://twitter.com/dog_rates/status/666094000...                9
2342  https://twitter.com/dog_rates/status/666082916...                6
2343  https://twitter.com/dog_rates/status/666073100...               10
2344  https://twitter.com/dog_rates/status/666071193...                9
2345  https://twitter.com/dog_rates/status/666063827...               10
2346  https://twitter.com/dog_rates/status/666058600...                8
2347  https://twitter.com/dog_rates/status/666057090...                9
2348  https://twitter.com/dog_rates/status/666055525...               10
2349  https://twitter.com/dog_rates/status/666051853...                2
2350  https://twitter.com/dog_rates/status/666050758...               10
2351  https://twitter.com/dog_rates/status/666049248...                5
2352  https://twitter.com/dog_rates/status/666044226...                6
2353  https://twitter.com/dog_rates/status/666033412...                9
2354  https://twitter.com/dog_rates/status/666029285...                7
2355  https://twitter.com/dog_rates/status/666020888...                8
```

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

```
21                   10   Jeffrey   None    None    None   None
22                   10      such   None    None    None   None
23                   10    Canela   None    None    None   None
24                   10      None   None    None    None   None
25                   10      None   None    None    None   None
26                   10      Maya   None    None    None   None
27                   10    Mingus   None    None    None   None
28                   10     Derek   None    None    None   None
29                   10    Roscoe   None    None  pupper   None
...                 ...       ...    ...     ...     ...    ...
2326                 10     quite   None    None    None   None
2327                 10         a   None    None    None   None
2328                 10      None   None    None    None   None
2329                 10      None   None    None    None   None
2330                 10      None   None    None    None   None
2331                 10      None   None    None    None   None
2332                 10      None   None    None    None   None
2333                 10        an   None    None    None   None
2334                 10         a   None    None    None   None
2335                  2        an   None    None    None   None
2336                 10      None   None    None    None   None
2337                 10      None   None    None    None   None
2338                 10      None   None    None    None   None
2339                 10      None   None    None    None   None
2340                 10      None   None    None    None   None
2341                 10      None   None    None    None   None
2342                 10      None   None    None    None   None
2343                 10      None   None    None    None   None
2344                 10      None   None    None    None   None
2345                 10       the   None    None    None   None
2346                 10       the   None    None    None   None
2347                 10         a   None    None    None   None
2348                 10         a   None    None    None   None
2349                 10        an   None    None    None   None
2350                 10         a   None    None    None   None
2351                 10      None   None    None    None   None
2352                 10         a   None    None    None   None
2353                 10         a   None    None    None   None
2354                 10         a   None    None    None   None
2355                 10      None   None    None    None   None

[2356 rows x 17 columns]
```

```
In [6]: prediction_data.head(1)
```

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

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

   p2_dog                 p3   p3_conf  p3_dog
0    True  Shetland_sheepdog  0.061428    True
```

In [7]:
```
consumer_key = '##################'
consumer_secret = '######################'
access_token = '###############-#########'
access_secret = '############################################'
```

In [8]:
```
'''auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)
api = tweepy.API(auth,
                 wait_on_rate_limit = True,
                 wait_on_rate_limit_notify = True)
# store tweets IDs that's exist and not in diffrent formats
# the json for the exists tweets and list for f

cant_find_tweets = []

with open('tweet_json.txt', 'w') as f:
    start = time.time()

    for tweet_id in twitter_data['tweet_id']:
        try:

            tweet = api.get_status(tweet_id,tweet_mode='extended')
            json.dump(tweet._json, f)
            f.write('\n')
        except Exception as e:
            cant_find_tweets.append(tweet_id)
    end= time.time()
print('time needed : ',end-start)
print ('tweets doesn't have data in API :',len(cant_find_tweets))'''
```

Out[8]:
```
"auth = tweepy.OAuthHandler(consumer_key, consumer_secret)\nauth.set_access_token(access
```

In [9]:
```
tweet_json = pd.read_json('tweet_json.txt',lines=True)
```

In [ ]:

In [ ]:

2 Assessing the data

In [ ]:

In [10]:
```
twitter_data
```

Out[10]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	NaN	NaN	
1	892177421306343426	NaN	NaN	
2	891815181378084864	NaN	NaN	
3	891689557279858688	NaN	NaN	
4	891327558926688256	NaN	NaN	
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	887978434458587136	NaN	NaN	
21	887705289381826560	NaN	NaN	
22	887517139158093824	NaN	NaN	
23	887473957103951883	NaN	NaN	
24	887343217045368832	NaN	NaN	
25	887101392804085760	NaN	NaN	
26	886983233522544640	NaN	NaN	
27	886736880519319552	NaN	NaN	
28	886680336477933568	NaN	NaN	
29	886366144734445568	NaN	NaN	
...	...	...	...	
2326	666411507551481857	NaN	NaN	
2327	666407126856765440	NaN	NaN	
2328	666396247373291520	NaN	NaN	
2329	666373753744588802	NaN	NaN	
2330	666362758909284353	NaN	NaN	
2331	666353288456101888	NaN	NaN	
2332	666345417576210432	NaN	NaN	
2333	666337882303524864	NaN	NaN	
2334	666293911632134144	NaN	NaN	
2335	666287406224695296	NaN	NaN	
2336	666273097616637952	NaN	NaN	
2337	666268910803644416	NaN	NaN	
2338	666104133288665088	NaN	NaN	
2339	666102155909144576	NaN	NaN	
2340	666099513787052032	NaN	NaN	
2341	666094000022159362	NaN	NaN	

```
2342  666082916733198337                    NaN                    NaN
2343  666073100786774016                    NaN                    NaN
2344  666071193221509120                    NaN                    NaN
2345  666063827256086533                    NaN                    NaN
2346  666058600524156928                    NaN                    NaN
2347  666057090499244032                    NaN                    NaN
2348  666055525042405380                    NaN                    NaN
2349  666051853826850816                    NaN                    NaN
2350  666050758794694657                    NaN                    NaN
2351  666049248165822465                    NaN                    NaN
2352  666044226329800704                    NaN                    NaN
2353  666033412701032449                    NaN                    NaN
2354  666029285002620928                    NaN                    NaN
2355  666020888022790149                    NaN                    NaN

                      timestamp  \
0     2017-08-01 16:23:56 +0000
1     2017-08-01 00:17:27 +0000
2     2017-07-31 00:18:03 +0000
3     2017-07-30 15:58:51 +0000
4     2017-07-29 16:00:24 +0000
5     2017-07-29 00:08:17 +0000
6     2017-07-28 16:27:12 +0000
7     2017-07-28 00:22:40 +0000
8     2017-07-27 16:25:51 +0000
9     2017-07-26 15:59:51 +0000
10    2017-07-26 00:31:25 +0000
11    2017-07-25 16:11:53 +0000
12    2017-07-25 01:55:32 +0000
13    2017-07-25 00:10:02 +0000
14    2017-07-24 17:02:04 +0000
15    2017-07-24 00:19:32 +0000
16    2017-07-23 00:22:39 +0000
17    2017-07-22 16:56:37 +0000
18    2017-07-22 00:23:06 +0000
19    2017-07-21 01:02:36 +0000
20    2017-07-20 16:49:33 +0000
21    2017-07-19 16:06:48 +0000
22    2017-07-19 03:39:09 +0000
23    2017-07-19 00:47:34 +0000
24    2017-07-18 16:08:03 +0000
25    2017-07-18 00:07:08 +0000
26    2017-07-17 16:17:36 +0000
27    2017-07-16 23:58:41 +0000
28    2017-07-16 20:14:00 +0000
29    2017-07-15 23:25:31 +0000
...                         ...
2326  2015-11-17 00:24:19 +0000
```

```
2327  2015-11-17 00:06:54 +0000
2328  2015-11-16 23:23:41 +0000
2329  2015-11-16 21:54:18 +0000
2330  2015-11-16 21:10:36 +0000
2331  2015-11-16 20:32:58 +0000
2332  2015-11-16 20:01:42 +0000
2333  2015-11-16 19:31:45 +0000
2334  2015-11-16 16:37:02 +0000
2335  2015-11-16 16:11:11 +0000
2336  2015-11-16 15:14:19 +0000
2337  2015-11-16 14:57:41 +0000
2338  2015-11-16 04:02:55 +0000
2339  2015-11-16 03:55:04 +0000
2340  2015-11-16 03:44:34 +0000
2341  2015-11-16 03:22:39 +0000
2342  2015-11-16 02:38:37 +0000
2343  2015-11-16 01:59:36 +0000
2344  2015-11-16 01:52:02 +0000
2345  2015-11-16 01:22:45 +0000
2346  2015-11-16 01:01:59 +0000
2347  2015-11-16 00:55:59 +0000
2348  2015-11-16 00:49:46 +0000
2349  2015-11-16 00:35:11 +0000
2350  2015-11-16 00:30:50 +0000
2351  2015-11-16 00:24:50 +0000
2352  2015-11-16 00:04:52 +0000
2353  2015-11-15 23:21:54 +0000
2354  2015-11-15 23:05:30 +0000
2355  2015-11-15 22:32:08 +0000

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

```
17    <a href="http://twitter.com/download/iphone" r...
18    <a href="http://twitter.com/download/iphone" r...
19    <a href="http://twitter.com/download/iphone" r...
20    <a href="http://twitter.com/download/iphone" r...
21    <a href="http://twitter.com/download/iphone" r...
22    <a href="http://twitter.com/download/iphone" r...
23    <a href="http://twitter.com/download/iphone" r...
24    <a href="http://twitter.com/download/iphone" r...
25    <a href="http://twitter.com/download/iphone" r...
26    <a href="http://twitter.com/download/iphone" r...
27    <a href="http://twitter.com/download/iphone" r...
28    <a href="http://twitter.com/download/iphone" r...
29    <a href="http://twitter.com/download/iphone" r...
...                                                 ...
2326  <a href="http://twitter.com/download/iphone" r...
2327  <a href="http://twitter.com/download/iphone" r...
2328  <a href="http://twitter.com/download/iphone" r...
2329  <a href="http://twitter.com/download/iphone" r...
2330  <a href="http://twitter.com/download/iphone" r...
2331  <a href="http://twitter.com/download/iphone" r...
2332  <a href="http://twitter.com/download/iphone" r...
2333  <a href="http://twitter.com/download/iphone" r...
2334  <a href="http://twitter.com/download/iphone" r...
2335  <a href="http://twitter.com/download/iphone" r...
2336  <a href="http://twitter.com/download/iphone" r...
2337  <a href="http://twitter.com/download/iphone" r...
2338  <a href="http://twitter.com/download/iphone" r...
2339  <a href="http://twitter.com/download/iphone" r...
2340  <a href="http://twitter.com/download/iphone" r...
2341  <a href="http://twitter.com/download/iphone" r...
2342  <a href="http://twitter.com/download/iphone" r...
2343  <a href="http://twitter.com/download/iphone" r...
2344  <a href="http://twitter.com/download/iphone" r...
2345  <a href="http://twitter.com/download/iphone" r...
2346  <a href="http://twitter.com/download/iphone" r...
2347  <a href="http://twitter.com/download/iphone" r...
2348  <a href="http://twitter.com/download/iphone" r...
2349  <a href="http://twitter.com/download/iphone" r...
2350  <a href="http://twitter.com/download/iphone" r...
2351  <a href="http://twitter.com/download/iphone" r...
2352  <a href="http://twitter.com/download/iphone" r...
2353  <a href="http://twitter.com/download/iphone" r...
2354  <a href="http://twitter.com/download/iphone" r...
2355  <a href="http://twitter.com/download/iphone" r...

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

2	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN
5	Here we have a majestic great white breaching ...	NaN
6	Meet Jax. He enjoys ice cream so much he gets ...	NaN
7	When you watch your owner call another dog a g...	NaN
8	This is Zoey. She doesn't want to be one of th...	NaN
9	This is Cassie. She is a college pup. Studying...	NaN
10	This is Koda. He is a South Australian decksha...	NaN
11	This is Bruno. He is a service shark. Only get...	NaN
12	Here's a puppo that seems to be on the fence a...	NaN
13	This is Ted. He does his best. Sometimes that'...	NaN
14	This is Stuart. He's sporting his favorite fan...	NaN
15	This is Oliver. You're witnessing one of his m...	NaN
16	This is Jim. He found a fren. Taught him how t...	NaN
17	This is Zeke. He has a new stick. Very proud o...	NaN
18	This is Ralphus. He's powering up. Attempting ...	NaN
19	RT @dog_rates: This is Canela. She attempted s...	8.874740e+17
20	This is Gerald. He was just told he didn't get...	NaN
21	This is Jeffrey. He has a monopoly on the pool...	NaN
22	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN
23	This is Canela. She attempted some fancy porch...	NaN
24	You may not have known you needed to see this ...	NaN
25	This... is a Jubilant Antarctic House Bear. We...	NaN
26	This is Maya. She's very shy. Rarely leaves he...	NaN
27	This is Mingus. He's a wonderful father to his...	NaN
28	This is Derek. He's late for a dog meeting. 13...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
...	...	...
2326	This is quite the dog. Gets really excited whe...	NaN
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN
2328	Oh goodness. A super rare northeast Qdoba kang...	NaN
2329	Those are sunglasses and a jean jacket. 11/10 ...	NaN
2330	Unique dog here. Very small. Lives in containe...	NaN
2331	Here we have a mixed Asiago from the Galápagos...	NaN
2332	Look at this jokester thinking seat belt laws ...	NaN
2333	This is an extremely rare horned Parthenon. No...	NaN
2334	This is a funny dog. Weird toes. Won't come do...	NaN
2335	This is an Albanian 3 1/2 legged Episcopalian...	NaN
2336	Can take selfies 11/10 https://t.co/ws2AMaNwPW	NaN
2337	Very concerned about fellow dog trapped in com...	NaN
2338	Not familiar with this breed. No tail (weird)...	NaN
2339	Oh my. Here you are seeing an Adobe Setter giv...	NaN
2340	Can stand on stump for what seems like a while...	NaN
2341	This appears to be a Mongolian Presbyterian mi...	NaN
2342	Here we have a well-established sunblockerspan...	NaN
2343	Let's hope this flight isn't Malaysian (lol). ...	NaN
2344	Here we have a northern speckled Rhododendron...	NaN

```
2345  This is the happiest dog you will ever see. Ve...                NaN
2346  Here is the Rand Paul of retrievers folks! He'...                NaN
2347  My oh my. This is a rare blond Canadian terrie...                NaN
2348  Here is a Siberian heavily armored polar bear ...                NaN
2349  This is an odd dog. Hard on the outside but lo...                NaN
2350  This is a truly beautiful English Wilson Staff...                NaN
2351  Here we have a 1949 1st generation vulpix. Enj...                NaN
2352  This is a purebred Piers Morgan. Loves to Netf...                NaN
2353  Here is a very happy pup. Big fan of well-main...                NaN
2354  This is a western brown Mitsubishi terrier. Up...                NaN
2355  Here we have a Japanese Irish Setter. Lost eye...                NaN

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

```
2330                        NaN                         NaN
2331                        NaN                         NaN
2332                        NaN                         NaN
2333                        NaN                         NaN
2334                        NaN                         NaN
2335                        NaN                         NaN
2336                        NaN                         NaN
2337                        NaN                         NaN
2338                        NaN                         NaN
2339                        NaN                         NaN
2340                        NaN                         NaN
2341                        NaN                         NaN
2342                        NaN                         NaN
2343                        NaN                         NaN
2344                        NaN                         NaN
2345                        NaN                         NaN
2346                        NaN                         NaN
2347                        NaN                         NaN
2348                        NaN                         NaN
2349                        NaN                         NaN
2350                        NaN                         NaN
2351                        NaN                         NaN
2352                        NaN                         NaN
2353                        NaN                         NaN
2354                        NaN                         NaN
2355                        NaN                         NaN

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

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

      rating_denominator      name  doggo floofer  pupper  puppo
0                     10   Phineas   None    None    None   None
1                     10     Tilly   None    None    None   None
2                     10    Archie   None    None    None   None
3                     10     Darla   None    None    None   None
4                     10  Franklin   None    None    None   None
```

5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None
8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
19	10	Canela	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
...	...	...	...	...	...	...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None
2343	10	None	None	None	None	None
2344	10	None	None	None	None	None
2345	10	the	None	None	None	None
2346	10	the	None	None	None	None
2347	10	a	None	None	None	None

```
2348                  10          a   None    None    None   None
2349                  10         an   None    None    None   None
2350                  10          a   None    None    None   None
2351                  10       None   None    None    None   None
2352                  10          a   None    None    None   None
2353                  10          a   None    None    None   None
2354                  10          a   None    None    None   None
2355                  10       None   None    None    None   None

[2356 rows x 17 columns]
```

```
In [11]: twitter_data.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 [12]: twitter_data.duplicated().sum()
```

```
Out[12]: 0
```

```
In [13]: twitter_data.name.value_counts()
```

```
Out[13]: None         745
         a             55
         Charlie       12
         Oliver        11
         Lucy          11
         Cooper        11
```

```
Lola          10
Tucker        10
Penny         10
Bo             9
Winston        9
the            8
Sadie          8
an             7
Bailey         7
Daisy          7
Buddy          7
Toby           7
Rusty          6
Milo           6
Leo            6
Koda           6
Stanley        6
Scout          6
Jack           6
Bella          6
Jax            6
Oscar          6
Dave           6
Phil           5
              ...
Spark          1
Koko           1
Reptar         1
Margo          1
Dietrich       1
by             1
Godi           1
Katie          1
Fletcher       1
Lizzie         1
Duchess        1
Biden          1
Shooter        1
Clifford       1
Dido           1
Sora           1
Jonah          1
Maxwell        1
Hermione       1
Todo           1
Buddah         1
Sprinkles      1
Birf           1
```

```
Remy           1
Carbon         1
Jennifur       1
Barclay        1
General        1
Halo           1
Clyde          1
Name: name, Length: 957, dtype: int64
```

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

Out[14]:

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

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

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

In [15]:
```
prediction_data.head()
```

Out[15]:

	tweet_id	jpg_url
0	666020888022790149	https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

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

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

```
In [16]: prediction_data.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 [17]: #bring only this columns from the json file id ,retweet_count,favorite_count
         tweet_json_modified= tweet_json.loc[:,['id','retweet_count','favorite_count']]
         tweet_json_modified= tweet_json_modified.rename(columns={"id": "tweet_id"})
```

```
In [18]: tweet_json_modified.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2331 entries, 0 to 2330
Data columns (total 3 columns):
tweet_id          2331 non-null int64
retweet_count     2331 non-null int64
favorite_count    2331 non-null int64
dtypes: int64(3)
```

```
memory usage: 54.7 KB
```

```
In [19]: tweet_json_modified.describe()
```

```
Out[19]:            tweet_id  retweet_count  favorite_count
         count  2.331000e+03     2331.00000     2331.000000
         mean   7.419079e+17     2603.55813     7336.684256
         std    6.823170e+16     4404.48562    11394.554122
         min    6.660209e+17        1.00000        0.000000
         25%    6.782670e+17      528.50000     1275.000000
         50%    7.182469e+17     1215.00000     3185.000000
         75%    7.986692e+17     3020.50000     8973.000000
         max    8.924206e+17    74767.00000   151101.000000
```

```
In [20]: prediction_data.duplicated().sum()
```

```
Out[20]: 0
```

```
In [21]: prediction_data.p3_dog.value_counts()
```

```
Out[21]: True     1499
         False     576
         Name: p3_dog, dtype: int64
```

```
In [ ]:
```

```
In [22]: print('twitter_data rows = ',twitter_data.shape[0])
         print('prediction_data rows = ',prediction_data.shape[0])
```

```
twitter_data rows =  2356
prediction_data rows =  2075
```

```
In [23]: twitter_data.text[0]
```

```
Out[23]: "This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/10
```

2.0.1 Tidiness Issues

- prediction data should be in the same file in the archived Twitter data file
- columns (tweet_id, retweet_count,favorite_count) in the JSON file should be in the same file with twitter data and prediction data -- ### Quality Issues in twitter archived data
- there are retweets in the data
- timestamp not in a datetime format
- the most name in the "name" column is 'a'
- there some outliers in (rating_numerator, rating_denominator)

Quality Issues in Twitter prediction data

- the prediction for the dog breed it's not actually a dog all the time

Quality Issues in JSON file

- there are some uninformative columns
- some tweets in the original file doesn't have a match in the JSON file
- there a min values that don't make sense in (retweet_count,favorite_count)

3 Cleaning the data

3.0.1 solving the tidness issues

```
In [24]: # Merge all dataframes in one place

         df_clean = pd.merge(twitter_data, prediction_data,on='tweet_id', how='inner') #1
         df_clean = pd.merge(df_clean, tweet_json_modified,on='tweet_id', how='inner') #2

In [25]: df_clean.columns
         df_clean
```

Out[25]:

	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	890066081131724 80	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	888078434458587136	NaN	NaN	
20	887705289381826560	NaN	NaN	
21	887517139158093824	NaN	NaN	
22	887473957103951883	NaN	NaN	
23	887343217045368832	NaN	NaN	
24	887101392804085760	NaN	NaN	
25	886983233522544640	NaN	NaN	

```
26    886736880519319552                  NaN                  NaN
27    886680336477933568                  NaN                  NaN
28    886366144734445568                  NaN                  NaN
29    886258384151887873                  NaN                  NaN
...                  ...                  ...                  ...
2029  666411507551481857                  NaN                  NaN
2030  666407126856765440                  NaN                  NaN
2031  666396247373291520                  NaN                  NaN
2032  666373753744588802                  NaN                  NaN
2033  666362758909284353                  NaN                  NaN
2034  666353288456101888                  NaN                  NaN
2035  666345417576210432                  NaN                  NaN
2036  666337882303524864                  NaN                  NaN
2037  666293911632134144                  NaN                  NaN
2038  666287406224695296                  NaN                  NaN
2039  666273097616637952                  NaN                  NaN
2040  666268910803644416                  NaN                  NaN
2041  666104133288665088                  NaN                  NaN
2042  666102155909144576                  NaN                  NaN
2043  666099513787052032                  NaN                  NaN
2044  666094000022159362                  NaN                  NaN
2045  666082916733198337                  NaN                  NaN
2046  666073100786774016                  NaN                  NaN
2047  666071193221509120                  NaN                  NaN
2048  666063827256086533                  NaN                  NaN
2049  666058600524156928                  NaN                  NaN
2050  666057090499244032                  NaN                  NaN
2051  666055525042405380                  NaN                  NaN
2052  666051853826850816                  NaN                  NaN
2053  666050758794694657                  NaN                  NaN
2054  666049248165822465                  NaN                  NaN
2055  666044226329800704                  NaN                  NaN
2056  666033412701032449                  NaN                  NaN
2057  666029285002620928                  NaN                  NaN
2058  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    2017-07-20 16:49:33 +0000
20    2017-07-19 16:06:48 +0000
21    2017-07-19 03:39:09 +0000
22    2017-07-19 00:47:34 +0000
23    2017-07-18 16:08:03 +0000
24    2017-07-18 00:07:08 +0000
25    2017-07-17 16:17:36 +0000
26    2017-07-16 23:58:41 +0000
27    2017-07-16 20:14:00 +0000
28    2017-07-15 23:25:31 +0000
29    2017-07-15 16:17:19 +0000
...                          ...
2029  2015-11-17 00:24:19 +0000
2030  2015-11-17 00:06:54 +0000
2031  2015-11-16 23:23:41 +0000
2032  2015-11-16 21:54:18 +0000
2033  2015-11-16 21:10:36 +0000
2034  2015-11-16 20:32:58 +0000
2035  2015-11-16 20:01:42 +0000
2036  2015-11-16 19:31:45 +0000
2037  2015-11-16 16:37:02 +0000
2038  2015-11-16 16:11:11 +0000
2039  2015-11-16 15:14:19 +0000
2040  2015-11-16 14:57:41 +0000
2041  2015-11-16 04:02:55 +0000
2042  2015-11-16 03:55:04 +0000
2043  2015-11-16 03:44:34 +0000
2044  2015-11-16 03:22:39 +0000
2045  2015-11-16 02:38:37 +0000
2046  2015-11-16 01:59:36 +0000
2047  2015-11-16 01:52:02 +0000
2048  2015-11-16 01:22:45 +0000
2049  2015-11-16 01:01:59 +0000
2050  2015-11-16 00:55:59 +0000
2051  2015-11-16 00:49:46 +0000
2052  2015-11-16 00:35:11 +0000
2053  2015-11-16 00:30:50 +0000
2054  2015-11-16 00:24:50 +0000
2055  2015-11-16 00:04:52 +0000
2056  2015-11-15 23:21:54 +0000
```

```
2057  2015-11-15 23:05:30 +0000
2058  2015-11-15 22:32:08 +0000

                                                  source  \
0     <a href="http://twitter.com/download/iphone" r...
1     <a href="http://twitter.com/download/iphone" r...
2     <a href="http://twitter.com/download/iphone" r...
3     <a href="http://twitter.com/download/iphone" r...
4     <a href="http://twitter.com/download/iphone" r...
5     <a href="http://twitter.com/download/iphone" r...
6     <a href="http://twitter.com/download/iphone" r...
7     <a href="http://twitter.com/download/iphone" r...
8     <a href="http://twitter.com/download/iphone" r...
9     <a href="http://twitter.com/download/iphone" r...
10    <a href="http://twitter.com/download/iphone" r...
11    <a href="http://twitter.com/download/iphone" r...
12    <a href="http://twitter.com/download/iphone" r...
13    <a href="http://twitter.com/download/iphone" r...
14    <a href="http://twitter.com/download/iphone" r...
15    <a href="http://twitter.com/download/iphone" r...
16    <a href="http://twitter.com/download/iphone" r...
17    <a href="http://twitter.com/download/iphone" r...
18    <a href="http://twitter.com/download/iphone" r...
19    <a href="http://twitter.com/download/iphone" r...
20    <a href="http://twitter.com/download/iphone" r...
21    <a href="http://twitter.com/download/iphone" r...
22    <a href="http://twitter.com/download/iphone" r...
23    <a href="http://twitter.com/download/iphone" r...
24    <a href="http://twitter.com/download/iphone" r...
25    <a href="http://twitter.com/download/iphone" r...
26    <a href="http://twitter.com/download/iphone" r...
27    <a href="http://twitter.com/download/iphone" r...
28    <a href="http://twitter.com/download/iphone" r...
29    <a href="http://twitter.com/download/iphone" r...
...                                                  ...
2029  <a href="http://twitter.com/download/iphone" r...
2030  <a href="http://twitter.com/download/iphone" r...
2031  <a href="http://twitter.com/download/iphone" r...
2032  <a href="http://twitter.com/download/iphone" r...
2033  <a href="http://twitter.com/download/iphone" r...
2034  <a href="http://twitter.com/download/iphone" r...
2035  <a href="http://twitter.com/download/iphone" r...
2036  <a href="http://twitter.com/download/iphone" r...
2037  <a href="http://twitter.com/download/iphone" r...
2038  <a href="http://twitter.com/download/iphone" r...
2039  <a href="http://twitter.com/download/iphone" r...
2040  <a href="http://twitter.com/download/iphone" r...
2041  <a href="http://twitter.com/download/iphone" r...
```

```
2042  <a href="http://twitter.com/download/iphone" r...
2043  <a href="http://twitter.com/download/iphone" r...
2044  <a href="http://twitter.com/download/iphone" r...
2045  <a href="http://twitter.com/download/iphone" r...
2046  <a href="http://twitter.com/download/iphone" r...
2047  <a href="http://twitter.com/download/iphone" r...
2048  <a href="http://twitter.com/download/iphone" r...
2049  <a href="http://twitter.com/download/iphone" r...
2050  <a href="http://twitter.com/download/iphone" r...
2051  <a href="http://twitter.com/download/iphone" r...
2052  <a href="http://twitter.com/download/iphone" r...
2053  <a href="http://twitter.com/download/iphone" r...
2054  <a href="http://twitter.com/download/iphone" r...
2055  <a href="http://twitter.com/download/iphone" r...
2056  <a href="http://twitter.com/download/iphone" r...
2057  <a href="http://twitter.com/download/iphone" r...
2058  <a href="http://twitter.com/download/iphone" r...

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

```
29    This is Waffles. His doggles are pupside down...                  NaN
...                                                    ...                ...
2029  This is quite the dog. Gets really excited whe...                 NaN
2030  This is a southern Vesuvius bumblegruff. Can d...                 NaN
2031  Oh goodness. A super rare northeast Qdoba kang...                 NaN
2032  Those are sunglasses and a jean jacket. 11/10 ...                 NaN
2033  Unique dog here. Very small. Lives in containe...                 NaN
2034  Here we have a mixed Asiago from the Galápagos...                 NaN
2035  Look at this jokester thinking seat belt laws ...                 NaN
2036  This is an extremely rare horned Parthenon. No...                 NaN
2037  This is a funny dog. Weird toes. Won't come do...                 NaN
2038  This is an Albanian 3 1/2 legged  Episcopalian...                 NaN
2039     Can take selfies 11/10 https://t.co/ws2AMaNwPW                 NaN
2040  Very concerned about fellow dog trapped in com...                 NaN
2041  Not familiar with this breed. No tail (weird)...                  NaN
2042  Oh my. Here you are seeing an Adobe Setter giv...                 NaN
2043  Can stand on stump for what seems like a while...                 NaN
2044  This appears to be a Mongolian Presbyterian mi...                 NaN
2045  Here we have a well-established sunblockerspan...                 NaN
2046  Let's hope this flight isn't Malaysian (lol). ...                 NaN
2047  Here we have a northern speckled Rhododendron...                  NaN
2048  This is the happiest dog you will ever see. Ve...                 NaN
2049  Here is the Rand Paul of retrievers folks! He'...                 NaN
2050  My oh my. This is a rare blond Canadian terrie...                 NaN
2051  Here is a Siberian heavily armored polar bear ...                 NaN
2052  This is an odd dog. Hard on the outside but lo...                 NaN
2053  This is a truly beautiful English Wilson Staff...                 NaN
2054  Here we have a 1949 1st generation vulpix. Enj...                 NaN
2055  This is a purebred Piers Morgan. Loves to Netf...                 NaN
2056  Here is a very happy pup. Big fan of well-main...                 NaN
2057  This is a western brown Mitsubishi terrier. Up...                 NaN
2058  Here we have a Japanese Irish Setter. Lost eye...                 NaN

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

14	NaN	NaN
15	NaN	NaN
16	NaN	NaN
17	NaN	NaN
18	NaN	NaN
19	NaN	NaN
20	NaN	NaN
21	NaN	NaN
22	NaN	NaN
23	NaN	NaN
24	NaN	NaN
25	NaN	NaN
26	NaN	NaN
27	NaN	NaN
28	NaN	NaN
29	NaN	NaN
...	...	...
2029	NaN	NaN
2030	NaN	NaN
2031	NaN	NaN
2032	NaN	NaN
2033	NaN	NaN
2034	NaN	NaN
2035	NaN	NaN
2036	NaN	NaN
2037	NaN	NaN
2038	NaN	NaN
2039	NaN	NaN
2040	NaN	NaN
2041	NaN	NaN
2042	NaN	NaN
2043	NaN	NaN
2044	NaN	NaN
2045	NaN	NaN
2046	NaN	NaN
2047	NaN	NaN
2048	NaN	NaN
2049	NaN	NaN
2050	NaN	NaN
2051	NaN	NaN
2052	NaN	NaN
2053	NaN	NaN
2054	NaN	NaN
2055	NaN	NaN
2056	NaN	NaN
2057	NaN	NaN
2058	NaN	NaN

	expanded_urls	...	\
0	https://twitter.com/dog_rates/status/892420643...	...	
1	https://twitter.com/dog_rates/status/892177421...	...	
2	https://twitter.com/dog_rates/status/891815181...	...	
3	https://twitter.com/dog_rates/status/891689557...	...	
4	https://twitter.com/dog_rates/status/891327558...	...	
5	https://twitter.com/dog_rates/status/891087950...	...	
6	https://gofundme.com/ydvmve-surgery-for-jax,ht...	...	
7	https://twitter.com/dog_rates/status/890729181...	...	
8	https://twitter.com/dog_rates/status/890609185...	...	
9	https://twitter.com/dog_rates/status/890240255...	...	
10	https://twitter.com/dog_rates/status/890006608...	...	
11	https://twitter.com/dog_rates/status/889880896...	...	
12	https://twitter.com/dog_rates/status/889665388...	...	
13	https://twitter.com/dog_rates/status/889638837...	...	
14	https://twitter.com/dog_rates/status/889531135...	...	
15	https://twitter.com/dog_rates/status/889278841...	...	
16	https://twitter.com/dog_rates/status/888917238...	...	
17	https://twitter.com/dog_rates/status/888804989...	...	
18	https://twitter.com/dog_rates/status/888554962...	...	
19	https://twitter.com/dog_rates/status/888078434...	...	
20	https://twitter.com/dog_rates/status/887705289...	...	
21	https://twitter.com/dog_rates/status/887517139...	...	
22	https://twitter.com/dog_rates/status/887473957...	...	
23	https://twitter.com/dog_rates/status/887343217...	...	
24	https://twitter.com/dog_rates/status/887101392...	...	
25	https://twitter.com/dog_rates/status/886983233...	...	
26	https://www.gofundme.com/mingusneedsus,https:/...	...	
27	https://twitter.com/dog_rates/status/886680336...	...	
28	https://twitter.com/dog_rates/status/886366144...	...	
29	https://twitter.com/dog_rates/status/886258384...	...	
...	...	...	
2029	https://twitter.com/dog_rates/status/666411507...	...	
2030	https://twitter.com/dog_rates/status/666407126...	...	
2031	https://twitter.com/dog_rates/status/666396247...	...	
2032	https://twitter.com/dog_rates/status/666373753...	...	
2033	https://twitter.com/dog_rates/status/666362758...	...	
2034	https://twitter.com/dog_rates/status/666353288...	...	
2035	https://twitter.com/dog_rates/status/666345417...	...	
2036	https://twitter.com/dog_rates/status/666337882...	...	
2037	https://twitter.com/dog_rates/status/666293911...	...	
2038	https://twitter.com/dog_rates/status/666287406...	...	
2039	https://twitter.com/dog_rates/status/666273097...	...	
2040	https://twitter.com/dog_rates/status/666268910...	...	
2041	https://twitter.com/dog_rates/status/666104133...	...	
2042	https://twitter.com/dog_rates/status/666102155...	...	
2043	https://twitter.com/dog_rates/status/666099513...	...	
2044	https://twitter.com/dog_rates/status/666094000...	...	

```
2045  https://twitter.com/dog_rates/status/666082916...      ...
2046  https://twitter.com/dog_rates/status/666073100...      ...
2047  https://twitter.com/dog_rates/status/666071193...      ...
2048  https://twitter.com/dog_rates/status/666063827...      ...
2049  https://twitter.com/dog_rates/status/666058600...      ...
2050  https://twitter.com/dog_rates/status/666057090...      ...
2051  https://twitter.com/dog_rates/status/666055525...      ...
2052  https://twitter.com/dog_rates/status/666051853...      ...
2053  https://twitter.com/dog_rates/status/666050758...      ...
2054  https://twitter.com/dog_rates/status/666049248...      ...
2055  https://twitter.com/dog_rates/status/666044226...      ...
2056  https://twitter.com/dog_rates/status/666033412...      ...
2057  https://twitter.com/dog_rates/status/666029285...      ...
2058  https://twitter.com/dog_rates/status/666020888...      ...

       p1_conf  p1_dog                     p2   p2_conf p2_dog  \
0     0.097049   False                  bagel  0.085851  False
1     0.323581    True               Pekinese  0.090647   True
2     0.716012    True               malamute  0.078253   True
3     0.170278   False     Labrador_retriever  0.168086   True
4     0.555712    True       English_springer  0.225770   True
5     0.425595    True          Irish_terrier  0.116317   True
6     0.341703    True          Border_collie  0.199287   True
7     0.566142    True             Eskimo_dog  0.178406   True
8     0.487574    True           Irish_setter  0.193054   True
9     0.511319    True               Cardigan  0.451038   True
10    0.957979    True             Pomeranian  0.013884   True
11    0.377417    True     Labrador_retriever  0.151317   True
12    0.966327    True               Cardigan  0.027356   True
13    0.991650    True                  boxer  0.002129   True
14    0.953442    True     Labrador_retriever  0.013834   True
15    0.626152    True                 borzoi  0.194742   True
16    0.714719    True        Tibetan_mastiff  0.120184   True
17    0.469760    True     Labrador_retriever  0.184172   True
18    0.700377    True             Eskimo_dog  0.166511   True
19    0.995026    True                    pug  0.000932   True
20    0.821664    True               redbone  0.087582   True
21    0.130432   False             tow_truck  0.029175  False
22    0.809197    True    Rhodesian_ridgeback  0.054950   True
23    0.330741    True               sea_lion  0.275645  False
24    0.733942    True             Eskimo_dog  0.035029   True
25    0.793469    True           toy_terrier  0.143528   True
26    0.309706    True         Great_Pyrenees  0.186136   True
27    0.738995   False             sports_car  0.139952  False
28    0.999201    True              Chihuahua  0.000361   True
29    0.943575    True             shower_cap  0.025286  False
...        ...     ...                    ...       ...    ...
2029  0.404640   False             barracouta  0.271485  False
```

```
2030  0.529139    True                bloodhound  0.244220   True
2031  0.978108    True              toy_terrier  0.009397   True
2032  0.326467    True             Afghan_hound  0.259551   True
2033  0.996496   False                    skunk  0.002402  False
2034  0.336874    True           Siberian_husky  0.147655   True
2035  0.858744    True  Chesapeake_Bay_retriever  0.054787   True
2036  0.416669   False             Newfoundland  0.278407   True
2037  0.914671   False                    otter  0.015250  False
2038  0.857531    True              toy_poodle  0.063064   True
2039  0.176053    True              toy_terrier  0.111884   True
2040  0.086502   False                     desk  0.085547  False
2041  0.965932   False                     cock  0.033919  False
2042  0.298617    True             Newfoundland  0.149842   True
2043  0.582330    True                 Shih-Tzu  0.166192   True
2044  0.195217    True          German_shepherd  0.078260   True
2045  0.489814    True             bull_mastiff  0.404722   True
2046  0.260857    True         English_foxhound  0.175382   True
2047  0.503672    True        Yorkshire_terrier  0.174201   True
2048  0.775930    True          Tibetan_mastiff  0.093718   True
2049  0.201493    True                 komondor  0.192305   True
2050  0.962465   False          shopping_basket  0.014594  False
2051  0.692517    True          Tibetan_mastiff  0.058279   True
2052  0.933012   False               mud_turtle  0.045885  False
2053  0.651137    True         English_springer  0.263788   True
2054  0.560311    True               Rottweiler  0.243682   True
2055  0.408143    True                  redbone  0.360687   True
2056  0.596461    True                malinois  0.138584   True
2057  0.506826    True       miniature_pinscher  0.074192   True
2058  0.465074    True                   collie  0.156665   True

                            p3   p3_conf p3_dog  retweet_count  \
0                       banana  0.076110  False           7408
1                     papillon  0.068957   True           5515
2                       kelpie  0.031379   True           3638
3                      spatula  0.040836  False           7587
4     German_short-haired_pointer  0.175219   True           8159
5                Indian_elephant  0.076902  False           2739
6                     ice_lolly  0.193548  False           1772
7                      Pembroke  0.076507   True          16576
8        Chesapeake_Bay_retriever  0.118184   True           3788
9                     Chihuahua  0.029248   True           6425
10                        chow  0.008167   True           6448
11                      muzzle  0.082981  False           4375
12                      basenji  0.004633   True           8787
13     Staffordshire_bullterrier  0.001498   True           3929
14                      redbone  0.007958   True           1982
15                       Saluki  0.027351   True           4676
16            Labrador_retriever  0.105506   True           3937
```

```
17                English_setter  0.073482   True           3723
18                     malamute  0.111411   True           3040
19                bull_mastiff  0.000903   True           3048
20                   Weimaraner  0.026236   True           4742
21                shopping_cart  0.026321  False          10339
22                       beagle  0.038915   True          15790
23                   Weimaraner  0.134203   True           9256
24     Staffordshire_bullterrier  0.029705   True           5239
25                   can_opener  0.032253  False           6711
26                Dandie_Dinmont  0.086346   True           2792
27                    car_wheel  0.044173  False           3933
28                  Boston_bull  0.000076   True           2782
29                  Siamese_cat  0.002849  False           5558
...                         ...       ...    ...            ...
2029                        gar  0.189945  False            285
2030       flat-coated_retriever  0.173810   True             31
2031                    papillon  0.004577   True             73
2032                      briard  0.206803   True             76
2033                     hamster  0.000461  False            499
2034                  Eskimo_dog  0.093412   True             63
2035          Labrador_retriever  0.014241   True            128
2036                groenendael  0.102643   True             81
2037             great_grey_owl  0.013207  False            308
2038           miniature_poodle  0.025581   True             57
2039                     basenji  0.111152   True             70
2040                    bookcase  0.079480  False             32
2041                   partridge  0.000052  False           5763
2042                      borzoi  0.133649   True             11
2043             Dandie_Dinmont  0.089688   True             57
2044                    malinois  0.075628   True             66
2045             French_bulldog  0.048960   True             41
2046               Ibizan_hound  0.097471   True            140
2047                    Pekinese  0.109454   True             52
2048          Labrador_retriever  0.072427   True            190
2049  soft-coated_wheaten_terrier  0.082086   True             50
2050            golden_retriever  0.007959   True            118
2051                    fur_coat  0.054449  False            211
2052                    terrapin  0.017885  False            745
2053   Greater_Swiss_Mountain_dog  0.016199   True             51
2054                    Doberman  0.154629   True             38
2055          miniature_pinscher  0.222752   True            122
2056                  bloodhound  0.116197   True             39
2057         Rhodesian_ridgeback  0.072010   True             41
2058           Shetland_sheepdog  0.061428   True            444

     favorite_count
0             35141
1             30404
```

2	22862
3	38416
4	36666
5	18488
6	10748
7	59149
8	25441
9	29052
10	28014
11	25462
12	43742
13	24588
14	13852
15	22946
16	26575
17	23317
18	17984
19	19879
20	27623
21	42275
22	62571
23	30698
24	27974
25	31684
26	10900
27	20520
28	19257
29	25517
...	...
2029	394
2030	98
2031	154
2032	169
2033	697
2034	194
2035	266
2036	174
2037	449
2038	131
2039	156
2040	93
2041	13196
2042	69
2043	140
2044	152
2045	101
2046	282
2047	134

```
2048            434
2049            103
2050            264
2051            397
2052           1088
2053            119
2054             94
2055            261
2056            107
2057            117
2058           2350

[2059 rows x 30 columns]
```

3.0.2 solving the quality issues

1- drop uninformative columns

```
In [26]: df_clean.drop(['in_reply_to_user_id',
                                'in_reply_to_status_id',
                                'retweeted_status_user_id',
                                'retweeted_status_timestamp'], axis=1,inplace = True)
```

2- remove retweets

```
In [27]: # drop the rows
         df_clean.drop(df_clean[df_clean.retweeted_status_id.notnull()== True].index, inplace=Tr

In [28]: #drop the columns
         df_clean.drop('retweeted_status_id',axis= 1,inplace=True)

In [29]: df_clean.columns

Out[29]: Index(['tweet_id', 'timestamp', 'source', 'text', 'expanded_urls',
                'rating_numerator', 'rating_denominator', 'name', 'doggo', 'floofer',
                'pupper', 'puppo', 'jpg_url', 'img_num', 'p1', 'p1_conf', 'p1_dog',
                'p2', 'p2_conf', 'p2_dog', 'p3', 'p3_conf', 'p3_dog', 'retweet_count',
                'favorite_count'],
               dtype='object')
```

3- convert object type to datetime in the timestamp column

```
In [30]: df_clean.timestamp=df_clean.timestamp.apply(pd.to_datetime)

In [31]: df_clean.timestamp.dtype

Out[31]: dtype('<M8[ns]')
```

4- replace the name 'a' with None value of the dog's name

```
In [32]: df_clean.name.value_counts()
```

```
Out[32]: None         546
         a             55
         Oliver        10
         Cooper        10
         Charlie       10
         Lucy           9
         Tucker         9
         Penny          9
         Winston        8
         Sadie          8
         Toby           7
         Daisy          7
         the            7
         Lola           7
         Jax            6
         Bella          6
         Koda           6
         Bo             6
         an             6
         Stanley        6
         Louis          5
         Scout          5
         Dave           5
         Chester        5
         Rusty          5
         Milo           5
         Bailey         5
         Leo            5
         Buddy          5
         Oscar          5
                       ...
         Mosby          1
         Banjo          1
         Aja            1
         Maks           1
         Barclay        1
         Remy           1
         Karl           1
         Birf           1
         Chadrick       1
         Stuart         1
         Mike           1
         Kawhi          1
         Godi           1
         Fletcher       1
```

```
Lizzie          1
Duchess         1
Biden           1
Shooter         1
Clifford        1
Dido            1
Sora            1
Jonah           1
Billy           1
Maxwell         1
Hermione        1
Todo            1
Buddah          1
Sprinkles       1
Paull           1
Clyde           1
Name: name, Length: 934, dtype: int64
```

```
In [33]: df_clean['name'] = df_clean['name'].replace(['a'],None)
```

```
In [34]: df_clean.name.value_counts()
```

```
Out[34]: None          565
         Oliver         12
         Charlie        10
         Cooper         10
         Penny           9
         Tucker          9
         Lucy            9
         the             9
         Daisy           8
         Winston         8
         an              8
         Sadie           8
         Lola            7
         Stanley         7
         Toby            7
         Koda            6
         Bo              6
         Jax             6
         Bella           6
         Chester         5
         Bailey          5
         Dave            5
         Gary            5
         Leo             5
         Oscar           5
         Scout           5
```

```
Rusty          5
Milo           5
Buddy          5
Louis          5
               ...
Shnuggles      1
Maks           1
Andru          1
Moofasa        1
Logan          1
Halo           1
Lilly          1
General        1
Jonah          1
Godi           1
Fletcher       1
Lizzie         1
Duchess        1
Biden          1
Shooter        1
Clifford       1
Dido           1
Sora           1
Billy          1
Barclay        1
Maxwell        1
Hermione       1
Todo           1
Buddah         1
Sprinkles      1
Paull          1
Birf           1
Remy           1
Carbon         1
Durg           1
Name: name, Length: 933, dtype: int64
```

```
In [ ]:
```

5-drop the outliers in (rating_numerator, rating_denominator)

```
In [35]: df_clean.rating_denominator.value_counts()
```

```
Out[35]: 10     1969
         50        3
         80        2
         11        2
         170       1
```

```
150       1
130       1
120       1
110       1
90        1
70        1
40        1
20        1
7         1
2         1
Name: rating_denominator, dtype: int64
```

In [36]:
```
df_clean.rating_numerator.value_counts()
```

Out[36]:
```
12      448
10      418
11      396
13      257
9       151
8        95
7        52
14       35
5        33
6        32
3        19
4        16
2         9
1         5
0         2
420       1
24        1
1776      1
27        1
44        1
45        1
50        1
60        1
75        1
80        1
84        1
88        1
99        1
121       1
143       1
144       1
165       1
204       1
26        1
Name: rating_numerator, dtype: int64
```

```
In [37]: print('value 170\n',df_clean[df_clean.rating_denominator == 170].text)
         print('value 150\n',df_clean[df_clean.rating_denominator == 150].text)
         print('value 130\n',df_clean[df_clean.rating_denominator == 130].text)
         print('value 120\n',df_clean[df_clean.rating_denominator == 120].text)
         print('value 110\n',df_clean[df_clean.rating_denominator == 110].text)

value 170
 911    Say hello to this unbelievably well behaved sq...
Name: text, dtype: object
value 150
 722    Why does this never happen at my front door...
Name: text, dtype: object
value 130
 1366    Two sneaky puppers were not initially seen, mo...
Name: text, dtype: object
value 120
 1498    IT'S PUPPERGEDDON. Total of 144/120 ...I think...
Name: text, dtype: object
value 110
 1367    Someone help the girl is being mugged. Several...
Name: text, dtype: object


In [38]: df_clean.drop(df_clean[df_clean['rating_numerator']>=400].index, inplace=True)

In [39]: df_clean.drop(df_clean[df_clean['rating_denominator']>=50].index, inplace=True)
```

6- the prediction for the dog breed it's not actully a dog all the time accroding the values of True and False of the prediction of 3 algorithms so we could consider this in our work

7- some tweets in the original file doesn't have match in json file this solved in merge method with inner join that is match only exist is on each file

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

Out[40]:            tweet_id  rating_numerator  rating_denominator      img_num  \
         count  1.973000e+03       1973.000000         1973.000000  1973.000000
         mean   7.357741e+17         10.604663           10.015712     1.203751
         std    6.752737e+16          2.802277            0.738163     0.562563
         min    6.660209e+17          0.000000            2.000000     1.000000
         25%    6.757816e+17         10.000000           10.000000     1.000000
         50%    7.081494e+17         11.000000           10.000000     1.000000
         75%    7.878106e+17         12.000000           10.000000     1.000000
         max    8.924206e+17         75.000000           40.000000     4.000000

                    p1_conf       p2_conf       p3_conf  retweet_count  favorite_count
         count  1973.000000  1.973000e+03  1.973000e+03    1973.000000     1973.000000
         mean      0.593553  1.346965e-01  6.029367e-02    2369.567157     8049.576280
```

std	0.272120	1.006635e-01	5.082766e-02	4243.295527	11855.727347
min	0.044333	1.011300e-08	1.740170e-10	11.000000	69.000000
25%	0.360428	5.413540e-02	1.619070e-02	529.000000	1703.000000
50%	0.587372	1.181810e-01	4.952370e-02	1144.000000	3632.000000
75%	0.845256	1.954050e-01	9.193000e-02	2704.000000	9994.000000
max	1.000000	4.880140e-01	2.710420e-01	74767.000000	151101.000000

8- there are a minimum values doesn't make sense in (retweet_count,favorite_count) droped with outlier in (rating_numerator , rating_denominator)

4 Store

```
In [41]: # store the data in csv file
         #df_clean.to_csv('twitter_archive_master.csv', encoding='utf-8', index=False)
```

4.0.1 Analyzing, and Visualizing Data for this Project

Through the project from gathering the data and assigning and cleaned and eventually the last step shows our work

```
In [42]: df = pd.read_csv('twitter_archive_master.csv')
```

the describes method give us intuition about the data here we have 1973 sample of the data after cleaning and dropping the outliers, also the mean of favorites of all tweets is 8049 and retweets is 2369

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

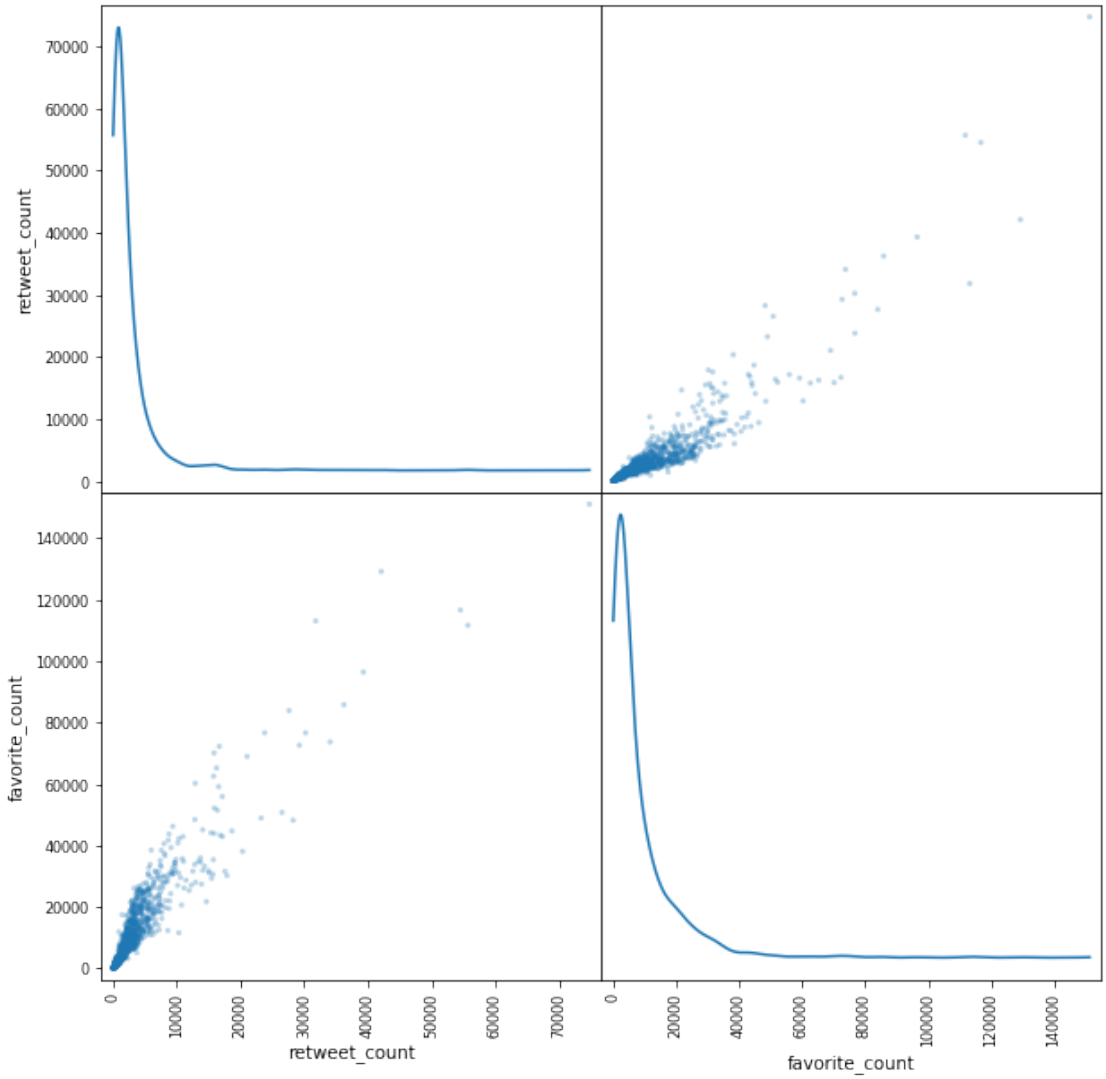
Out[48]:

	tweet_id	rating_numerator	rating_denominator	img_num	\
count	1.973000e+03	1973.000000	1973.000000	1973.000000	
mean	7.357741e+17	10.604663	10.015712	1.203751	
std	6.752737e+16	2.802277	0.738163	0.562563	
min	6.660209e+17	0.000000	2.000000	1.000000	
25%	6.757816e+17	10.000000	10.000000	1.000000	
50%	7.081494e+17	11.000000	10.000000	1.000000	
75%	7.878106e+17	12.000000	10.000000	1.000000	
max	8.924206e+17	75.000000	40.000000	4.000000	

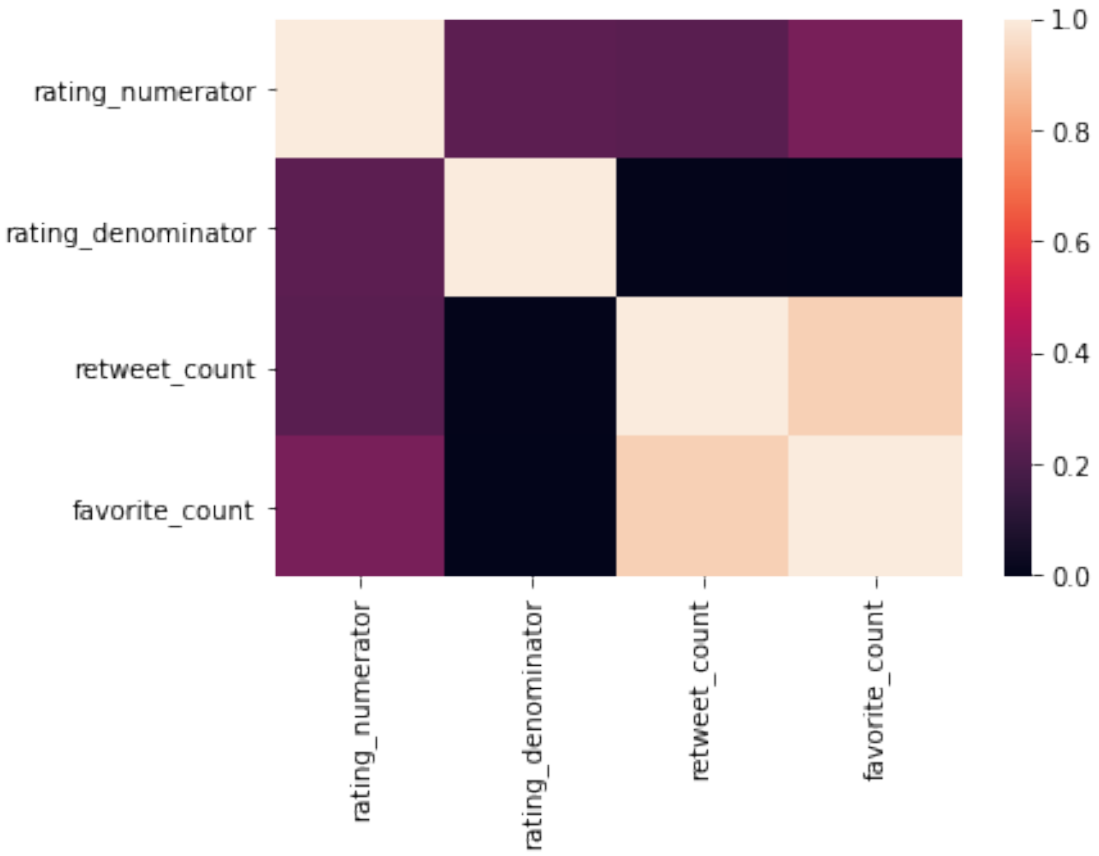
	p1_conf	p2_conf	p3_conf	retweet_count	favorite_count
count	1973.000000	1.973000e+03	1.973000e+03	1973.000000	1973.000000
mean	0.593553	1.346965e-01	6.029367e-02	2369.567157	8049.576280
std	0.272120	1.006635e-01	5.082766e-02	4243.295527	11855.727347
min	0.044333	1.011300e-08	1.740170e-10	11.000000	69.000000
25%	0.360428	5.413540e-02	1.619070e-02	529.000000	1703.000000
50%	0.587372	1.181810e-01	4.952370e-02	1144.000000	3632.000000
75%	0.845256	1.954050e-01	9.193000e-02	2704.000000	9994.000000
max	1.000000	4.880140e-01	2.710420e-01	74767.000000	151101.000000

In [43]:
```
# Produce a scatter matrix for each pair of features in the data
pd.scatter_matrix(df.loc[:,['retweet_count','favorite_count']], alpha = 0.3, figsize =
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:2: FutureWarning: pandas.scatter_ma
```



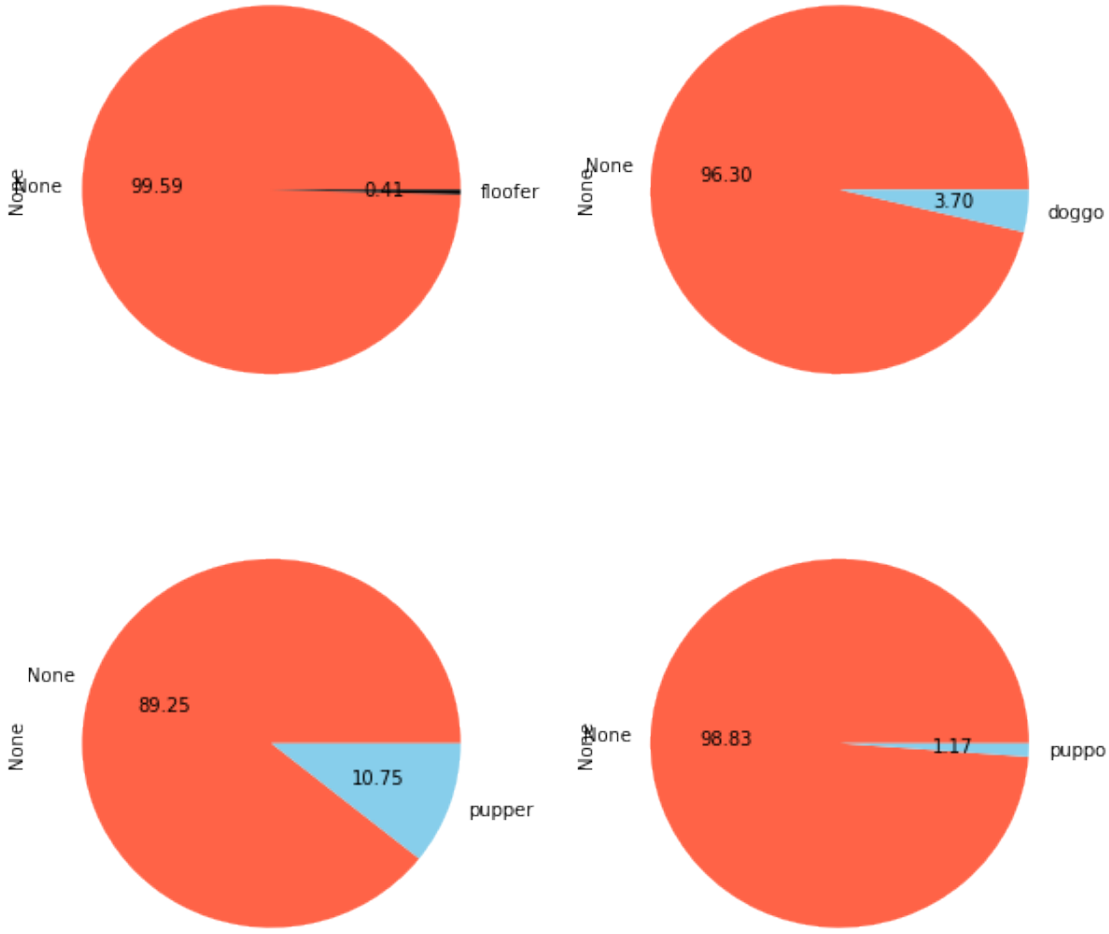
In [49]:
```
corr = df.loc[:,['rating_numerator','rating_denominator','retweet_count','favorite_coun
sns.heatmap(corr);
```



Those figures show several things:

a positive correlation between the favorites and retweets

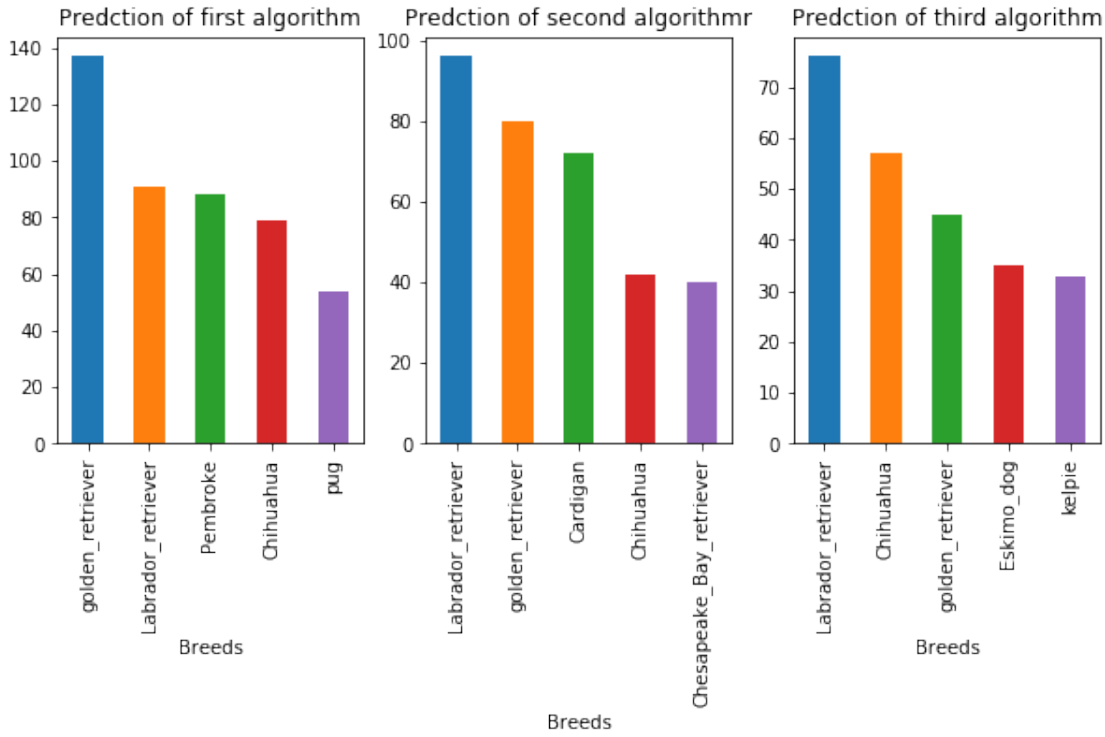
```
In [44]: fig, ((ax1, ax2),(ax3,ax4)) = plt.subplots(nrows=2,ncols=2, figsize=(10, 10))

         df.groupby('floofer').size().plot(kind='pie', autopct='%.2f',
                                           colors=['tomato', 'black'], ax=ax1)

         df.groupby('doggo').size().plot(kind='pie', autopct='%.2f',
                                           colors=['tomato', 'skyblue'], ax=ax2)

         df.groupby('pupper').size().plot(kind='pie', autopct='%.2f',
                                           colors=['tomato', 'skyblue'], ax=ax3)

         df.groupby('puppo').size().plot(kind='pie', autopct='%.2f',
                                           colors=['tomato', 'skyblue'], ax=ax4)

         plt.show()
```



these four pie charts showes the various stages of dog: doggo, pupper, puppo, and floofer and as showen the 'pupper' is the most domniated in the tweets

```
In [50]: fig, ((ax1), (ax2),(ax3)) = plt.subplots(nrows=1,ncols=3, figsize=(10, 4))

         df[df['p1_dog'] == True].p1.sort_values().value_counts().head(5).plot(kind='bar', ax=ax
         df[df['p2_dog'] == True].p2.sort_values().value_counts().head(5).plot(kind='bar', ax=ax
         df[df['p3_dog'] == True].p3.sort_values().value_counts().head(5).plot(kind='bar', ax=ax
         ax1.set_xlabel('Breeds')
         ax1.set_title('Predction of first algorithm')
         ax2.set_xlabel('Breeds')
         ax2.set_title('Predction of second algorithmr')
         ax3.set_xlabel('Breeds')
         ax3.set_title('Predction of third algorithm')
```

```
plt.show()
```



As it can be seen, the pair 'golden retriever / Labrador retriever' is the absolute leader in terms of top breeds which classified from the different three algorithms

```
In [71]: df.sort_values(by='favorite_count', ascending=False).iloc[0]
```

```
Out[71]: tweet_id                                            744234799360020481
         timestamp                                          2016-06-18 18:26:18
         source              <a href="http://twitter.com/download/iphone" r...
         text                Here's a doggo realizing you can stand in a po...
         expanded_urls       https://twitter.com/dog_rates/status/744234799...
         rating_numerator                                                   13
         rating_denominator                                                 10
         name                                                             None
         doggo                                                           doggo
         floofer                                                          None
         pupper                                                           None
         puppo                                                            None
         jpg_url             https://pbs.twimg.com/ext_tw_video_thumb/74423...
         img_num                                                             1
         p1                                                  Labrador_retriever
```

```
p1_conf                                                  0.825333
p1_dog                                                       True
p2                                                       ice_bear
p2_conf                                                 0.0446808
p2_dog                                                       False
p3                                                        whippet
p3_conf                                                 0.0184422
p3_dog                                                       True
retweet_count                                               74767
favorite_count                                             151101
Name: 766, dtype: object
```

4.0.2 case study

case study for the highest Tweet has retweets and favorites it's in 2016 and it's doggo and it's breed is 'Labrador retriever'

here the url

https://twitter.com/dog_rates/status/744234799360020481/video/1

In [ ]: