

# wrangle\_act

June 28, 2022

## 1 Project: Wrangling and Analyze Data

### 1.1 Data Gathering

In the cell below, gather **all** three pieces of data for this project and load them in the notebook. **Note:** the methods required to gather each data are different. 1. Directly download the WeRate-Dogs Twitter archive data (twitter\_archive\_enhanced.csv)

```
In [40]: #importing the required packages
import pandas as pd
import numpy as np
import requests
import tweepy
import json
import os
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
In [41]: import pandas as pd
# Reading twitter _archive_enhance data
df_arch=pd.read_csv('twitter-archive-enhanced.csv')
df_arch.head()
```

```
Out[41]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	NaN	NaN	
1	892177421306343426	NaN	NaN	
2	891815181378084864	NaN	NaN	
3	891689557279858688	NaN	NaN	
4	891327558926688256	NaN	NaN	

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

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

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

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

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

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

## 2. Use the Requests library to download the tweet image prediction (image\_predictions.tsv)

```
In [42]: #downloading tweet_image_prediction
import requests as req
with req.get("https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-p
    with open("image_prediction.tsv",'wb') as file:
        file.write(rq.content)

In [43]: #Reading the downloaded image_prediction
image=pd.read_csv('image_prediction.tsv', sep='\t')
image.head()

Out[43]:
```

	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 [44]: # importing the Api data using the data given in the udacity

```

import json
df_list=[]
with open('tweet-json.txt','r') as file:
    for line in file:
        tweets=json.loads(line)
        tweet_id=tweets['id']
        retweet_count=tweets['retweet_count']
        favorite_count=tweets['favorite_count']
        df_list.append({'tweet_id':tweet_id,'retweet_count':retweet_count,'favorite_count':favorite_count})
tweet_count=pd.DataFrame(df_list,columns=['tweet_id','retweet_count','favorite_count'])

```

In [45]: tweet\_count

```

Out[45]:
   tweet_id  retweet_count  favorite_count
0  892420643555336193      8853         39467
1  892177421306343426      6514         33819
2  891815181378084864      4328         25461
3  891689557279858688      8964         42908
4  891327558926688256      9774         41048
5  891087950875897856      3261         20562
6  890971913173991426      2158         12041
7  890729181411237888     16716         56848
8  890609185150312448      4429         28226
9  890240255349198849      7711         32467
10 890006608113172480      7624         31166
11 889880896479866881      5156         28268
12 889665388333682689      8538         38818

```

13	889638837579907072	4735	27672
14	889531135344209921	2321	15359
15	889278841981685760	5637	25652
16	888917238123831296	4709	29611
17	888804989199671297	4559	26080
18	888554962724278272	3732	20290
19	888078434458587136	3653	22201
20	887705289381826560	5609	30779
21	887517139158093824	12082	46959
22	887473957103951883	18781	69871
23	887343217045368832	10737	34222
24	887101392804085760	6167	31061
25	886983233522544640	8084	35859
26	886736880519319552	3443	12306
27	886680336477933568	4610	22798
28	886366144734445568	3316	21524
29	886267009285017600	4	117
...	...	...	...
2324	666411507551481857	339	459
2325	666407126856765440	44	113
2326	666396247373291520	92	172
2327	666373753744588802	100	194
2328	666362758909284353	595	804
2329	666353288456101888	77	229
2330	666345417576210432	146	307
2331	666337882303524864	96	204
2332	666293911632134144	368	522
2333	666287406224695296	71	152
2334	666273097616637952	82	184
2335	666268910803644416	37	108
2336	666104133288665088	6871	14765
2337	666102155909144576	16	81
2338	666099513787052032	73	164
2339	666094000022159362	79	169
2340	666082916733198337	47	121
2341	666073100786774016	174	335
2342	666071193221509120	67	154
2343	666063827256086533	232	496
2344	666058600524156928	61	115
2345	666057090499244032	146	304
2346	666055525042405380	261	448
2347	666051853826850816	879	1253
2348	666050758794694657	60	136
2349	666049248165822465	41	111
2350	666044226329800704	147	311
2351	666033412701032449	47	128
2352	666029285002620928	48	132
2353	666020888022790149	532	2535

```
[2354 rows x 3 columns]
```

```
In [7]: df_list[0].keys()
```

```
Out[7]: dict_keys(['tweet_id', 'retweet_count', 'favorite_count'])
```

```
In [13]:
```

## 1.2 Assessing Data

In this section, detect and document at least **eight (8) quality issues and two (2) tidiness issue**. You must use **both** visual assessment programmatic assessement to assess the data.

**Note:** pay attention to the following key points when you access the data.

- You only want original ratings (no retweets) that have images. Though there are 5000+ tweets in the dataset, not all are dog ratings and some are retweets.
- Assessing and cleaning the entire dataset completely would require a lot of time, and is not necessary to practice and demonstrate your skills in data wrangling. Therefore, the requirements of this project are only to assess and clean at least 8 quality issues and at least 2 tidiness issues in this dataset.
- The fact that the rating numerators are greater than the denominators does not need to be cleaned. This [unique rating system](#) is a big part of the popularity of WeRateDogs.
- You do not need to gather the tweets beyond August 1st, 2017. You can, but note that you won't be able to gather the image predictions for these tweets since you don't have access to the algorithm used.

```
In [8]: df_arch
```

```
Out[8]:
```

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

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

	timestamp \
0	2017-08-01 16:23:56 +0000
1	2017-08-01 00:17:27 +0000
2	2017-07-31 00:18:03 +0000
3	2017-07-30 15:58:51 +0000

4	2017-07-29	16:00:24	+0000
5	2017-07-29	00:08:17	+0000
6	2017-07-28	16:27:12	+0000
7	2017-07-28	00:22:40	+0000
8	2017-07-27	16:25:51	+0000
9	2017-07-26	15:59:51	+0000
10	2017-07-26	00:31:25	+0000
11	2017-07-25	16:11:53	+0000
12	2017-07-25	01:55:32	+0000
13	2017-07-25	00:10:02	+0000
14	2017-07-24	17:02:04	+0000
15	2017-07-24	00:19:32	+0000
16	2017-07-23	00:22:39	+0000
17	2017-07-22	16:56:37	+0000
18	2017-07-22	00:23:06	+0000
19	2017-07-21	01:02:36	+0000
20	2017-07-20	16:49:33	+0000
21	2017-07-19	16:06:48	+0000
22	2017-07-19	03:39:09	+0000
23	2017-07-19	00:47:34	+0000
24	2017-07-18	16:08:03	+0000
25	2017-07-18	00:07:08	+0000
26	2017-07-17	16:17:36	+0000
27	2017-07-16	23:58:41	+0000
28	2017-07-16	20:14:00	+0000
29	2017-07-15	23:25:31	+0000
...			...
2326	2015-11-17	00:24:19	+0000
2327	2015-11-17	00:06:54	+0000
2328	2015-11-16	23:23:41	+0000
2329	2015-11-16	21:54:18	+0000
2330	2015-11-16	21:10:36	+0000
2331	2015-11-16	20:32:58	+0000
2332	2015-11-16	20:01:42	+0000
2333	2015-11-16	19:31:45	+0000
2334	2015-11-16	16:37:02	+0000
2335	2015-11-16	16:11:11	+0000
2336	2015-11-16	15:14:19	+0000
2337	2015-11-16	14:57:41	+0000
2338	2015-11-16	04:02:55	+0000
2339	2015-11-16	03:55:04	+0000
2340	2015-11-16	03:44:34	+0000
2341	2015-11-16	03:22:39	+0000
2342	2015-11-16	02:38:37	+0000
2343	2015-11-16	01:59:36	+0000
2344	2015-11-16	01:52:02	+0000
2345	2015-11-16	01:22:45	+0000
2346	2015-11-16	01:01:59	+0000

2347 2015-11-16 00:55:59 +0000  
2348 2015-11-16 00:49:46 +0000  
2349 2015-11-16 00:35:11 +0000  
2350 2015-11-16 00:30:50 +0000  
2351 2015-11-16 00:24:50 +0000  
2352 2015-11-16 00:04:52 +0000  
2353 2015-11-15 23:21:54 +0000  
2354 2015-11-15 23:05:30 +0000  
2355 2015-11-15 22:32:08 +0000

```

                                source \
0      <a href="http://twitter.com/download/iphone" 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 <a href="https://t.co/ws2AMaWpPW">https://t.co/ws2AMaWpPW</a>	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	<a href="https://twitter.com/dog_rates/status/666287406...">https://twitter.com/dog_rates/status/666287406...</a>	1
2336	<a href="https://twitter.com/dog_rates/status/666273097...">https://twitter.com/dog_rates/status/666273097...</a>	11
2337	<a href="https://twitter.com/dog_rates/status/666268910...">https://twitter.com/dog_rates/status/666268910...</a>	10
2338	<a href="https://twitter.com/dog_rates/status/666104133...">https://twitter.com/dog_rates/status/666104133...</a>	1
2339	<a href="https://twitter.com/dog_rates/status/666102155...">https://twitter.com/dog_rates/status/666102155...</a>	11
2340	<a href="https://twitter.com/dog_rates/status/666099513...">https://twitter.com/dog_rates/status/666099513...</a>	8
2341	<a href="https://twitter.com/dog_rates/status/666094000...">https://twitter.com/dog_rates/status/666094000...</a>	9
2342	<a href="https://twitter.com/dog_rates/status/666082916...">https://twitter.com/dog_rates/status/666082916...</a>	6
2343	<a href="https://twitter.com/dog_rates/status/666073100...">https://twitter.com/dog_rates/status/666073100...</a>	10
2344	<a href="https://twitter.com/dog_rates/status/666071193...">https://twitter.com/dog_rates/status/666071193...</a>	9
2345	<a href="https://twitter.com/dog_rates/status/666063827...">https://twitter.com/dog_rates/status/666063827...</a>	10
2346	<a href="https://twitter.com/dog_rates/status/666058600...">https://twitter.com/dog_rates/status/666058600...</a>	8
2347	<a href="https://twitter.com/dog_rates/status/666057090...">https://twitter.com/dog_rates/status/666057090...</a>	9
2348	<a href="https://twitter.com/dog_rates/status/666055525...">https://twitter.com/dog_rates/status/666055525...</a>	10
2349	<a href="https://twitter.com/dog_rates/status/666051853...">https://twitter.com/dog_rates/status/666051853...</a>	2
2350	<a href="https://twitter.com/dog_rates/status/666050758...">https://twitter.com/dog_rates/status/666050758...</a>	10
2351	<a href="https://twitter.com/dog_rates/status/666049248...">https://twitter.com/dog_rates/status/666049248...</a>	5
2352	<a href="https://twitter.com/dog_rates/status/666044226...">https://twitter.com/dog_rates/status/666044226...</a>	6
2353	<a href="https://twitter.com/dog_rates/status/666033412...">https://twitter.com/dog_rates/status/666033412...</a>	9
2354	<a href="https://twitter.com/dog_rates/status/666029285...">https://twitter.com/dog_rates/status/666029285...</a>	7
2355	<a href="https://twitter.com/dog_rates/status/666020888...">https://twitter.com/dog_rates/status/666020888...</a>	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 [9]: df\_arch.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                2356 non-null int64
in_reply_to_status_id   78 non-null float64
in_reply_to_user_id     78 non-null float64
timestamp               2356 non-null object
```

```

source                2356 non-null object
text                  2356 non-null object
retweeted_status_id    181 non-null float64
retweeted_status_user_id 181 non-null float64
retweeted_status_timestamp 181 non-null object
expanded_urls          2297 non-null object
rating_numerator        2356 non-null int64
rating_denominator      2356 non-null int64
name                   2356 non-null object
doggo                  2356 non-null object
floofer                2356 non-null object
pupper                 2356 non-null object
puppo                  2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB

```

In [10]: image

```

Out[10]:
      tweet_id  jpg_url \
0    666020888022790149  https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg
1    666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2    666033412701032449  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3    666044226329800704  https://pbs.twimg.com/media/CT5Dr8HUEAA-1Eu.jpg
4    666049248165822465  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
5    666050758794694657  https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6    666051853826850816  https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7    666055525042405380  https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8    666057090499244032  https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9    666058600524156928  https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10   666063827256086533  https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11   666071193221509120  https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
12   666073100786774016  https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13   666082916733198337  https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg
14   666094000022159362  https://pbs.twimg.com/media/CT5w9gUW4AAaBNN.jpg
15   666099513787052032  https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16   666102155909144576  https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg
17   666104133288665088  https://pbs.twimg.com/media/CT56LSZWAAALJj2.jpg
18   666268910803644416  https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
19   666273097616637952  https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
20   666287406224695296  https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
21   666293911632134144  https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22   666337882303524864  https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg
23   666345417576210432  https://pbs.twimg.com/media/CT9Vn7PWAAA_ZCM.jpg
24   666353288456101888  https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25   666362758909284353  https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
26   666373753744588802  https://pbs.twimg.com/media/CT9vZEYWUAA1Z05.jpg
27   666396247373291520  https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg

```

28 666407126856765440 <https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.jpg>  
29 666411507551481857 <https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg>  
... ..  
2045 886366144734445568 <https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg>  
2046 886680336477933568 <https://pbs.twimg.com/media/DE4fEDzWAAAYHMM.jpg>  
2047 886736880519319552 <https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg>  
2048 886983233522544640 <https://pbs.twimg.com/media/DE8yicJW0AAAABJ.jpg>  
2049 887101392804085760 <https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg>  
2050 887343217045368832 [https://pbs.twimg.com/ext\\_tw\\_video\\_thumb/88734...](https://pbs.twimg.com/ext_tw_video_thumb/88734...)  
2051 887473957103951883 <https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg>  
2052 887517139158093824 [https://pbs.twimg.com/ext\\_tw\\_video\\_thumb/88751...](https://pbs.twimg.com/ext_tw_video_thumb/88751...)  
2053 887705289381826560 <https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg>  
2054 888078434458587136 <https://pbs.twimg.com/media/DFMWn56WsAAKA7B.jpg>  
2055 888202515573088257 <https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg>  
2056 888554962724278272 [https://pbs.twimg.com/media/DFTH\\_0-UQAACu20.jpg](https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg)  
2057 888804989199671297 <https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg>  
2058 888917238123831296 <https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg>  
2059 889278841981685760 [https://pbs.twimg.com/ext\\_tw\\_video\\_thumb/88927...](https://pbs.twimg.com/ext_tw_video_thumb/88927...)  
2060 889531135344209921 [https://pbs.twimg.com/media/DFg\\_2PVW0AEHN3p.jpg](https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg)  
2061 889638837579907072 <https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg>  
2062 889665388333682689 <https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg>  
2063 889880896479866881 <https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg>  
2064 890006608113172480 <https://pbs.twimg.com/media/DFnwsY4WAAAMliS.jpg>  
2065 890240255349198849 <https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg>  
2066 890609185150312448 [https://pbs.twimg.com/media/DFwUU\\_-XcAEpyXI.jpg](https://pbs.twimg.com/media/DFwUU_-XcAEpyXI.jpg)  
2067 890729181411237888 <https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg>  
2068 890971913173991426 <https://pbs.twimg.com/media/DF1e0mZXUAAALUc.jpg>  
2069 891087950875897856 <https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg>  
2070 891327558926688256 <https://pbs.twimg.com/media/DF6hr6BUMAAZgT.jpg>  
2071 891689557279858688 [https://pbs.twimg.com/media/DF\\_q7IAWsAEuuN8.jpg](https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg)  
2072 891815181378084864 <https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg>  
2073 892177421306343426 <https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg>  
2074 892420643555336193 <https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg>

	img_num		p1	p1_conf	p1_dog \
0	1	Welsh_springer_spaniel	0.465074	True	
1	1	redbone	0.506826	True	
2	1	German_shepherd	0.596461	True	
3	1	Rhodesian_ridgeback	0.408143	True	
4	1	miniature_pinscher	0.560311	True	
5	1	Bernese_mountain_dog	0.651137	True	
6	1	box_turtle	0.933012	False	
7	1	chow	0.692517	True	
8	1	shopping_cart	0.962465	False	
9	1	miniature_poodle	0.201493	True	
10	1	golden_retriever	0.775930	True	
11	1	Gordon_setter	0.503672	True	
12	1	Walker_hound	0.260857	True	



13	1	pug	0.489814	True
14	1	bloodhound	0.195217	True
15	1	Lhasa	0.582330	True
16	1	English_setter	0.298617	True
17	1	hen	0.965932	False
18	1	desktop_computer	0.086502	False
19	1	Italian_greyhound	0.176053	True
20	1	Maltese_dog	0.857531	True
21	1	three-toed_sloth	0.914671	False
22	1	ox	0.416669	False
23	1	golden_retriever	0.858744	True
24	1	malamute	0.336874	True
25	1	guinea_pig	0.996496	False
26	1	soft-coated_wheaten_terrier	0.326467	True
27	1	Chihuahua	0.978108	True
28	1	black-and-tan_coonhound	0.529139	True
29	1	coho	0.404640	False
...	...	...	...	...
2045	1	French_bulldog	0.999201	True
2046	1	convertible	0.738995	False
2047	1	kuvasz	0.309706	True
2048	2	Chihuahua	0.793469	True
2049	1	Samoyed	0.733942	True
2050	1	Mexican_hairless	0.330741	True
2051	2	Pembroke	0.809197	True
2052	1	limousine	0.130432	False
2053	1	basset	0.821664	True
2054	1	French_bulldog	0.995026	True
2055	2	Pembroke	0.809197	True
2056	3	Siberian_husky	0.700377	True
2057	1	golden_retriever	0.469760	True
2058	1	golden_retriever	0.714719	True
2059	1	whippet	0.626152	True
2060	1	golden_retriever	0.953442	True
2061	1	French_bulldog	0.991650	True
2062	1	Pembroke	0.966327	True
2063	1	French_bulldog	0.377417	True
2064	1	Samoyed	0.957979	True
2065	1	Pembroke	0.511319	True
2066	1	Irish_terrier	0.487574	True
2067	2	Pomeranian	0.566142	True
2068	1	Appenzeller	0.341703	True
2069	1	Chesapeake_Bay_retriever	0.425595	True
2070	2	basset	0.555712	True
2071	1	paper_towel	0.170278	False
2072	1	Chihuahua	0.716012	True
2073	1	Chihuahua	0.323581	True
2074	1	orange	0.097049	False

	p2	p2_conf	p2_dog	p3 \
0	collie	0.156665	True	Shetland_sheepdog
1	miniature_pinscher	0.074192	True	Rhodesian_ridgeback
2	malinois	0.138584	True	bloodhound
3	redbone	0.360687	True	miniature_pinscher
4	Rottweiler	0.243682	True	Doberman
5	English_springer	0.263788	True	Greater_Swiss_Mountain_dog
6	mud_turtle	0.045885	False	terrapien
7	Tibetan_mastiff	0.058279	True	fur_coat
8	shopping_basket	0.014594	False	golden_retriever
9	komondor	0.192305	True	soft-coated_wheaten_terrier
10	Tibetan_mastiff	0.093718	True	Labrador_retriever
11	Yorkshire_terrier	0.174201	True	Pekinese
12	English_foxhound	0.175382	True	Ibizan_hound
13	bull_mastiff	0.404722	True	French_bulldog
14	German_shepherd	0.078260	True	malinois
15	Shih-Tzu	0.166192	True	Dandie_Dinmont
16	Newfoundland	0.149842	True	borzoi
17	cock	0.033919	False	partridge
18	desk	0.085547	False	bookcase
19	toy_terrier	0.111884	True	basenji
20	toy_poodle	0.063064	True	miniature_poodle
21	otter	0.015250	False	great_grey_owl
22	Newfoundland	0.278407	True	groenendael
23	Chesapeake_Bay_retriever	0.054787	True	Labrador_retriever
24	Siberian_husky	0.147655	True	Eskimo_dog
25	skunk	0.002402	False	hamster
26	Afghan_hound	0.259551	True	briard
27	toy_terrier	0.009397	True	papillon
28	bloodhound	0.244220	True	flat-coated_retriever
29	barracouta	0.271485	False	gar
...	...	...	...	...
2045	Chihuahua	0.000361	True	Boston_bull
2046	sports_car	0.139952	False	car_wheel
2047	Great_Pyrenees	0.186136	True	Dandie_Dinmont
2048	toy_terrier	0.143528	True	can_opener
2049	Eskimo_dog	0.035029	True	Staffordshire_bullterrier
2050	sea_lion	0.275645	False	Weimaraner
2051	Rhodesian_ridgeback	0.054950	True	beagle
2052	tow_truck	0.029175	False	shopping_cart
2053	redbone	0.087582	True	Weimaraner
2054	pug	0.000932	True	bull_mastiff
2055	Rhodesian_ridgeback	0.054950	True	beagle
2056	Eskimo_dog	0.166511	True	malamute
2057	Labrador_retriever	0.184172	True	English_setter
2058	Tibetan_mastiff	0.120184	True	Labrador_retriever
2059	borzoi	0.194742	True	Saluki

2060	Labrador_retriever	0.013834	True	redbone
2061	boxer	0.002129	True	Staffordshire_bullterrier
2062	Cardigan	0.027356	True	basenji
2063	Labrador_retriever	0.151317	True	muzzle
2064	Pomeranian	0.013884	True	chow
2065	Cardigan	0.451038	True	Chihuahua
2066	Irish_setter	0.193054	True	Chesapeake_Bay_retriever
2067	Eskimo_dog	0.178406	True	Pembroke
2068	Border_collie	0.199287	True	ice_lolly
2069	Irish_terrier	0.116317	True	Indian_elephant
2070	English_springer	0.225770	True	German_short-haired_pointer
2071	Labrador_retriever	0.168086	True	spatula
2072	malamute	0.078253	True	kelpie
2073	Pekinese	0.090647	True	papillon
2074	bagel	0.085851	False	banana

	p3_conf	p3_dog
0	0.061428	True
1	0.072010	True
2	0.116197	True
3	0.222752	True
4	0.154629	True
5	0.016199	True
6	0.017885	False
7	0.054449	False
8	0.007959	True
9	0.082086	True
10	0.072427	True
11	0.109454	True
12	0.097471	True
13	0.048960	True
14	0.075628	True
15	0.089688	True
16	0.133649	True
17	0.000052	False
18	0.079480	False
19	0.111152	True
20	0.025581	True
21	0.013207	False
22	0.102643	True
23	0.014241	True
24	0.093412	True
25	0.000461	False
26	0.206803	True
27	0.004577	True
28	0.173810	True
29	0.189945	False
...	...	...

2045	0.000076	True
2046	0.044173	False
2047	0.086346	True
2048	0.032253	False
2049	0.029705	True
2050	0.134203	True
2051	0.038915	True
2052	0.026321	False
2053	0.026236	True
2054	0.000903	True
2055	0.038915	True
2056	0.111411	True
2057	0.073482	True
2058	0.105506	True
2059	0.027351	True
2060	0.007958	True
2061	0.001498	True
2062	0.004633	True
2063	0.082981	False
2064	0.008167	True
2065	0.029248	True
2066	0.118184	True
2067	0.076507	True
2068	0.193548	False
2069	0.076902	False
2070	0.175219	True
2071	0.040836	False
2072	0.031379	True
2073	0.068957	True
2074	0.076110	False

[2075 rows x 12 columns]

In [11]: df.describe()

```
Out[11]:
```

	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 [12]: image.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id      2075 non-null int64
jpg_url       2075 non-null object
img_num       2075 non-null int64
p1            2075 non-null object
p1_conf       2075 non-null float64
p1_dog        2075 non-null bool
p2            2075 non-null object
p2_conf       2075 non-null float64
p2_dog        2075 non-null bool
p3            2075 non-null object
p3_conf       2075 non-null float64
p3_dog        2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

In [11]: image.describe()

Out[11]:	tweet_id	img_num	p1_conf	p2_conf	p3_conf
count	2.075000e+03	2075.000000	2075.000000	2.075000e+03	2.075000e+03
mean	7.384514e+17	1.203855	0.594548	1.345886e-01	6.032417e-02
std	6.785203e+16	0.561875	0.271174	1.006657e-01	5.090593e-02
min	6.660209e+17	1.000000	0.044333	1.011300e-08	1.740170e-10
25%	6.764835e+17	1.000000	0.364412	5.388625e-02	1.622240e-02
50%	7.119988e+17	1.000000	0.588230	1.181810e-01	4.944380e-02
75%	7.932034e+17	1.000000	0.843855	1.955655e-01	9.180755e-02
max	8.924206e+17	4.000000	1.000000	4.880140e-01	2.734190e-01

In [14]: image

```
Out[14]:
```

	tweet_id	jpg_url \
0	666020888022790149	<a href="https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg">https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg</a>
1	666029285002620928	<a href="https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg">https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg</a>
2	666033412701032449	<a href="https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg">https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg</a>
3	666044226329800704	<a href="https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg">https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg</a>
4	666049248165822465	<a href="https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg">https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg</a>
5	666050758794694657	<a href="https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg">https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg</a>
6	666051853826850816	<a href="https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg">https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg</a>
7	666055525042405380	<a href="https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg">https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg</a>
8	666057090499244032	<a href="https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg">https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg</a>
9	666058600524156928	<a href="https://pbs.twimg.com/media/CT5Qw94XAAA-2dP.jpg">https://pbs.twimg.com/media/CT5Qw94XAAA-2dP.jpg</a>
10	666063827256086533	<a href="https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg">https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg</a>
11	666071193221509120	<a href="https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg">https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg</a>
12	666073100786774016	<a href="https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg">https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg</a>
13	666082916733198337	<a href="https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg">https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg</a>
14	666094000022159362	<a href="https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg">https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg</a>
15	666099513787052032	<a href="https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg">https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg</a>
16	666102155909144576	<a href="https://pbs.twimg.com/media/CT54YGiwUAEZnoK.jpg">https://pbs.twimg.com/media/CT54YGiwUAEZnoK.jpg</a>
17	666104133288665088	<a href="https://pbs.twimg.com/media/CT56LSZWAAALJj2.jpg">https://pbs.twimg.com/media/CT56LSZWAAALJj2.jpg</a>
18	666268910803644416	<a href="https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg">https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg</a>
19	666273097616637952	<a href="https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg">https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg</a>
20	666287406224695296	<a href="https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg">https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg</a>
21	666293911632134144	<a href="https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg">https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg</a>
22	666337882303524864	<a href="https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg">https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg</a>
23	666345417576210432	<a href="https://pbs.twimg.com/media/CT9Vn7PWAAA_ZCM.jpg">https://pbs.twimg.com/media/CT9Vn7PWAAA_ZCM.jpg</a>
24	666353288456101888	<a href="https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg">https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg</a>
25	666362758909284353	<a href="https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg">https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg</a>
26	666373753744588802	<a href="https://pbs.twimg.com/media/CT9vZEYUAAALZ05.jpg">https://pbs.twimg.com/media/CT9vZEYUAAALZ05.jpg</a>
27	666396247373291520	<a href="https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg">https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg</a>
28	666407126856765440	<a href="https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.jpg">https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.jpg</a>
29	666411507551481857	<a href="https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg">https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg</a>
...	...	...
2045	886366144734445568	<a href="https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg">https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg</a>
2046	886680336477933568	<a href="https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg">https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg</a>
2047	886736880519319552	<a href="https://pbs.twimg.com/media/DE5Se8FXcAAJFfx4.jpg">https://pbs.twimg.com/media/DE5Se8FXcAAJFfx4.jpg</a>
2048	886983233522544640	<a href="https://pbs.twimg.com/media/DE8yicJW0AAAABJ.jpg">https://pbs.twimg.com/media/DE8yicJW0AAAABJ.jpg</a>
2049	887101392804085760	<a href="https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg">https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg</a>
2050	887343217045368832	<a href="https://pbs.twimg.com/ext_tw_video_thumb/88734...">https://pbs.twimg.com/ext_tw_video_thumb/88734...</a>
2051	887473957103951883	<a href="https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg">https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg</a>
2052	887517139158093824	<a href="https://pbs.twimg.com/ext_tw_video_thumb/88751...">https://pbs.twimg.com/ext_tw_video_thumb/88751...</a>
2053	887705289381826560	<a href="https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg">https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg</a>
2054	888078434458587136	<a href="https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg">https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg</a>
2055	888202515573088257	<a href="https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg">https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg</a>
2056	888554962724278272	<a href="https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg">https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg</a>
2057	888804989199671297	<a href="https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg">https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg</a>
2058	888917238123831296	<a href="https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg">https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg</a>

2059	889278841981685760	<a href="https://pbs.twimg.com/ext_tw_video_thumb/88927...">https://pbs.twimg.com/ext_tw_video_thumb/88927...</a>
2060	889531135344209921	<a href="https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg">https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg</a>
2061	889638837579907072	<a href="https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg">https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg</a>
2062	889665388333682689	<a href="https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg">https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg</a>
2063	889880896479866881	<a href="https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg">https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg</a>
2064	890006608113172480	<a href="https://pbs.twimg.com/media/DFnwsY4WAAAMliS.jpg">https://pbs.twimg.com/media/DFnwsY4WAAAMliS.jpg</a>
2065	890240255349198849	<a href="https://pbs.twimg.com/media/DFrEyVuWAAA03t9.jpg">https://pbs.twimg.com/media/DFrEyVuWAAA03t9.jpg</a>
2066	890609185150312448	<a href="https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg">https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg</a>
2067	890729181411237888	<a href="https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg">https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg</a>
2068	890971913173991426	<a href="https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg">https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg</a>
2069	891087950875897856	<a href="https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg">https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg</a>
2070	891327558926688256	<a href="https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg">https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg</a>
2071	891689557279858688	<a href="https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg">https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg</a>
2072	891815181378084864	<a href="https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg">https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg</a>
2073	892177421306343426	<a href="https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg">https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg</a>
2074	892420643555336193	<a href="https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg">https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg</a>

	img_num		p1	p1_conf	p1_dog	\
0	1	Welsh_springer_spaniel	0.465074	True		
1	1	redbone	0.506826	True		
2	1	German_shepherd	0.596461	True		
3	1	Rhodesian_ridgeback	0.408143	True		
4	1	miniature_pinscher	0.560311	True		
5	1	Bernese_mountain_dog	0.651137	True		
6	1	box_turtle	0.933012	False		
7	1	chow	0.692517	True		
8	1	shopping_cart	0.962465	False		
9	1	miniature_poodle	0.201493	True		
10	1	golden_retriever	0.775930	True		
11	1	Gordon_setter	0.503672	True		
12	1	Walker_hound	0.260857	True		
13	1	pug	0.489814	True		
14	1	bloodhound	0.195217	True		
15	1	Lhasa	0.582330	True		
16	1	English_setter	0.298617	True		
17	1	hen	0.965932	False		
18	1	desktop_computer	0.086502	False		
19	1	Italian_greyhound	0.176053	True		
20	1	Maltese_dog	0.857531	True		
21	1	three-toed_sloth	0.914671	False		
22	1	ox	0.416669	False		
23	1	golden_retriever	0.858744	True		
24	1	malamute	0.336874	True		
25	1	guinea_pig	0.996496	False		
26	1	soft-coated_wheaten_terrier	0.326467	True		
27	1	Chihuahua	0.978108	True		
28	1	black-and-tan_coonhound	0.529139	True		
29	1	coho	0.404640	False		

...	...	...	...	...
2045	1	French_bulldog	0.999201	True
2046	1	convertible	0.738995	False
2047	1	kuvasz	0.309706	True
2048	2	Chihuahua	0.793469	True
2049	1	Samoyed	0.733942	True
2050	1	Mexican_hairless	0.330741	True
2051	2	Pembroke	0.809197	True
2052	1	limousine	0.130432	False
2053	1	basset	0.821664	True
2054	1	French_bulldog	0.995026	True
2055	2	Pembroke	0.809197	True
2056	3	Siberian_husky	0.700377	True
2057	1	golden_retriever	0.469760	True
2058	1	golden_retriever	0.714719	True
2059	1	whippet	0.626152	True
2060	1	golden_retriever	0.953442	True
2061	1	French_bulldog	0.991650	True
2062	1	Pembroke	0.966327	True
2063	1	French_bulldog	0.377417	True
2064	1	Samoyed	0.957979	True
2065	1	Pembroke	0.511319	True
2066	1	Irish_terrier	0.487574	True
2067	2	Pomeranian	0.566142	True
2068	1	Appenzeller	0.341703	True
2069	1	Chesapeake_Bay_retriever	0.425595	True
2070	2	basset	0.555712	True
2071	1	paper_towel	0.170278	False
2072	1	Chihuahua	0.716012	True
2073	1	Chihuahua	0.323581	True
2074	1	orange	0.097049	False

		p2	p2_conf	p2_dog		p3 \
0		collie	0.156665	True		Shetland_sheepdog
1	miniature_pinscher		0.074192	True		Rhodesian_ridgeback
2	malinois		0.138584	True		bloodhound
3	redbone		0.360687	True		miniature_pinscher
4	Rottweiler		0.243682	True		Doberman
5	English_springer		0.263788	True	Greater_Swiss_Mountain_dog	
6	mud_turtle		0.045885	False		terrapin
7	Tibetan_mastiff		0.058279	True		fur_coat
8	shopping_basket		0.014594	False		golden_retriever
9	komondor		0.192305	True	soft-coated_wheaten_terrier	
10	Tibetan_mastiff		0.093718	True	Labrador_retriever	
11	Yorkshire_terrier		0.174201	True		Pekinese
12	English_foxhound		0.175382	True		Ibizan_hound
13	bull_mastiff		0.404722	True		French_bulldog
14	German_shepherd		0.078260	True		malinois



15	Shih-Tzu	0.166192	True	Dandie_Dinmont
16	Newfoundland	0.149842	True	borzoi
17	cock	0.033919	False	partridge
18	desk	0.085547	False	bookcase
19	toy_terrier	0.111884	True	basenji
20	toy_poodle	0.063064	True	miniature_poodle
21	otter	0.015250	False	great_grey_owl
22	Newfoundland	0.278407	True	groenendael
23	Chesapeake_Bay_retriever	0.054787	True	Labrador_retriever
24	Siberian_husky	0.147655	True	Eskimo_dog
25	skunk	0.002402	False	hamster
26	Afghan_hound	0.259551	True	briard
27	toy_terrier	0.009397	True	papillon
28	bloodhound	0.244220	True	flat-coated_retriever
29	barracouta	0.271485	False	gar
...	...	...	...	...
2045	Chihuahua	0.000361	True	Boston_bull
2046	sports_car	0.139952	False	car_wheel
2047	Great_Pyrenees	0.186136	True	Dandie_Dinmont
2048	toy_terrier	0.143528	True	can_opener
2049	Eskimo_dog	0.035029	True	Staffordshire_bullterrier
2050	sea_lion	0.275645	False	Weimaraner
2051	Rhodesian_ridgeback	0.054950	True	beagle
2052	tow_truck	0.029175	False	shopping_cart
2053	redbone	0.087582	True	Weimaraner
2054	pug	0.000932	True	bull_mastiff
2055	Rhodesian_ridgeback	0.054950	True	beagle
2056	Eskimo_dog	0.166511	True	malamute
2057	Labrador_retriever	0.184172	True	English_setter
2058	Tibetan_mastiff	0.120184	True	Labrador_retriever
2059	borzoi	0.194742	True	Saluki
2060	Labrador_retriever	0.013834	True	redbone
2061	boxer	0.002129	True	Staffordshire_bullterrier
2062	Cardigan	0.027356	True	basenji
2063	Labrador_retriever	0.151317	True	muzzle
2064	Pomeranian	0.013884	True	chow
2065	Cardigan	0.451038	True	Chihuahua
2066	Irish_setter	0.193054	True	Chesapeake_Bay_retriever
2067	Eskimo_dog	0.178406	True	Pembroke
2068	Border_collie	0.199287	True	ice_lolly
2069	Irish_terrier	0.116317	True	Indian_elephant
2070	English_springer	0.225770	True	German_short-haired_pointer
2071	Labrador_retriever	0.168086	True	spatula
2072	malamute	0.078253	True	kelpie
2073	Pekinese	0.090647	True	papillon
2074	bagel	0.085851	False	banana

p3\_conf p3\_dog

0	0.061428	True
1	0.072010	True
2	0.116197	True
3	0.222752	True
4	0.154629	True
5	0.016199	True
6	0.017885	False
7	0.054449	False
8	0.007959	True
9	0.082086	True
10	0.072427	True
11	0.109454	True
12	0.097471	True
13	0.048960	True
14	0.075628	True
15	0.089688	True
16	0.133649	True
17	0.000052	False
18	0.079480	False
19	0.111152	True
20	0.025581	True
21	0.013207	False
22	0.102643	True
23	0.014241	True
24	0.093412	True
25	0.000461	False
26	0.206803	True
27	0.004577	True
28	0.173810	True
29	0.189945	False
...	...	...
2045	0.000076	True
2046	0.044173	False
2047	0.086346	True
2048	0.032253	False
2049	0.029705	True
2050	0.134203	True
2051	0.038915	True
2052	0.026321	False
2053	0.026236	True
2054	0.000903	True
2055	0.038915	True
2056	0.111411	True
2057	0.073482	True
2058	0.105506	True
2059	0.027351	True
2060	0.007958	True
2061	0.001498	True

2062	0.004633	True
2063	0.082981	False
2064	0.008167	True
2065	0.029248	True
2066	0.118184	True
2067	0.076507	True
2068	0.193548	False
2069	0.076902	False
2070	0.175219	True
2071	0.040836	False
2072	0.031379	True
2073	0.068957	True
2074	0.076110	False

[2075 rows x 12 columns]

In [12]: tweet\_count

Out[12]:	tweet_id	retweet_count	favorite_count
0	892420643555336193	8853	39467
1	892177421306343426	6514	33819
2	891815181378084864	4328	25461
3	891689557279858688	8964	42908
4	891327558926688256	9774	41048
5	891087950875897856	3261	20562
6	890971913173991426	2158	12041
7	890729181411237888	16716	56848
8	890609185150312448	4429	28226
9	890240255349198849	7711	32467
10	890006608113172480	7624	31166
11	889880896479866881	5156	28268
12	889665388333682689	8538	38818
13	889638837579907072	4735	27672
14	889531135344209921	2321	15359
15	889278841981685760	5637	25652
16	888917238123831296	4709	29611
17	888804989199671297	4559	26080
18	888554962724278272	3732	20290
19	888078434458587136	3653	22201
20	887705289381826560	5609	30779
21	887517139158093824	12082	46959
22	887473957103951883	18781	69871
23	887343217045368832	10737	34222
24	887101392804085760	6167	31061
25	886983233522544640	8084	35859
26	886736880519319552	3443	12306
27	886680336477933568	4610	22798
28	886366144734445568	3316	21524

29	886267009285017600	4	117
...	...	...	...
2324	666411507551481857	339	459
2325	666407126856765440	44	113
2326	666396247373291520	92	172
2327	666373753744588802	100	194
2328	666362758909284353	595	804
2329	666353288456101888	77	229
2330	666345417576210432	146	307
2331	666337882303524864	96	204
2332	666293911632134144	368	522
2333	666287406224695296	71	152
2334	666273097616637952	82	184
2335	666268910803644416	37	108
2336	666104133288665088	6871	14765
2337	666102155909144576	16	81
2338	666099513787052032	73	164
2339	666094000022159362	79	169
2340	666082916733198337	47	121
2341	666073100786774016	174	335
2342	666071193221509120	67	154
2343	666063827256086533	232	496
2344	666058600524156928	61	115
2345	666057090499244032	146	304
2346	666055525042405380	261	448
2347	666051853826850816	879	1253
2348	666050758794694657	60	136
2349	666049248165822465	41	111
2350	666044226329800704	147	311
2351	666033412701032449	47	128
2352	666029285002620928	48	132
2353	666020888022790149	532	2535

[2354 rows x 3 columns]

In [18]: tweet\_count.describe()

```
Out[18]:
```

	tweet_id	retweet_count	favorite_count
count	2.354000e+03	2354.000000	2354.000000
mean	7.426978e+17	3164.797366	8080.968564
std	6.852812e+16	5284.770364	11814.771334
min	6.660209e+17	0.000000	0.000000
25%	6.783975e+17	624.500000	1415.000000
50%	7.194596e+17	1473.500000	3603.500000
75%	7.993058e+17	3652.000000	10122.250000
max	8.924206e+17	79515.000000	132810.000000

In [13]: tweet\_count.info()

```

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

```

### 1.2.1 Quality issues

1.Changing None in doggo,floofer,pupper,puppo to Nan

2.name column has many invalid value like a, the,None. 3.text has a link of tweets and rating at the end

4.In rating\_denominator some entry are more than and below 10 need to be 10

5.Rating\_numerator some entry are less than 10 and supposed to be more than 10

6.Timestamp column is string instead of datetime

7.Expanded url has Nan-values 8.In\_reply\_to\_user\_id,In\_reply\_to\_status\_id,retweet\_status\_id,retweet\_status\_id has missing values

### 1.2.2 Tidiness issues

1.Doggo,flooter,pupper and Puppo need to be one column in image\_prediction.csv

2.All the three dataset are related yet divided into three dataframe

## 1.3 Cleaning Data

In this section, clean **all** of the issues you documented while assessing.

**Note:** Make a copy of the original data before cleaning. Cleaning includes merging individual pieces of data according to the rules of [tidy data](#). The result should be a high-quality and tidy master pandas DataFrame (or DataFrames, if appropriate).

```

In [46]: # Make copies of original pieces of data
         df_arch_clean=df_arch.copy()
         image_clean=image.copy()
         tweet_count_clean=tweet_count.copy()

```

### 1.3.1 1. All the three dataset are related yet divided into three dataframe

### 1.3.2 Define:Combining all the three dataset into one dataset

```

In [48]: #merging Enhanced archived data with tweet count data using tweet id
         df_clean1=pd.merge(df_arch_clean,tweet_count_clean, on='tweet_id', how='left')
         #merging again the formed data with the image prediction also using Tweet id
         df_clean=pd.merge(df_clean1,image_clean, on='tweet_id', how='left')

```

```

In [10]: df_clean.info()

```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 30 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
retweet_count           2354 non-null float64
favorite_count          2354 non-null float64
jpg_url                 2075 non-null object
img_num                 2075 non-null float64
p1                      2075 non-null object
p1_conf                 2075 non-null float64
p1_dog                  2075 non-null object
p2                      2075 non-null object
p2_conf                 2075 non-null float64
p2_dog                  2075 non-null object
p3                      2075 non-null object
p3_conf                 2075 non-null float64
p3_dog                  2075 non-null object
dtypes: float64(10), int64(3), object(17)
memory usage: 570.6+ KB

```

### 1.3.3 2: Changing None in doggo, floofer, pupper, puppo to Nan

### 1.3.4 : Define Removing all the None in the Doggo, floofer, pupper and puppo column to Nan

### 1.3.5 Code

```

In [49]: #Changing the word None on the dog levels columns to Nan
dog_level=['doggo','floofer','pupper','puppo']
for dog in dog_level:
    df_clean[dog]=df_clean[dog].replace('None',np.nan)

```

### 1.3.6 Test

```
In [10]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 30 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                   97 non-null object
floofer                 10 non-null object
pupper                  257 non-null object
puppo                   30 non-null object
retweet_count           2354 non-null float64
favorite_count          2354 non-null float64
jpg_url                 2075 non-null object
img_num                 2075 non-null float64
p1                      2075 non-null object
p1_conf                 2075 non-null float64
p1_dog                  2075 non-null object
p2                      2075 non-null object
p2_conf                 2075 non-null float64
p2_dog                  2075 non-null object
p3                      2075 non-null object
p3_conf                 2075 non-null float64
p3_dog                  2075 non-null object
dtypes: float64(10), int64(3), object(17)
memory usage: 570.6+ KB
```

### 1.3.7 3: doggo,floofer,pupper and puppo to be one column

### 1.3.8 Define: Combine the doggo,floofer,pupper and puppo to one column dog level

### 1.3.9 Code

```
In [50]: col_list=['doggo','floofer','pupper','puppo']
df_clean['dog_level']=df_clean[col_list].apply(lambda x: ', '.join(x.dropna()).astype(st
```

```
df_clean.dog_level=df_clean.dog_level.replace('',np.nan).astype('category')
df_clean=df_clean.drop(columns=col_list, axis=1)
```

### 1.3.10 Test

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 27 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
retweet_count           2354 non-null float64
favorite_count          2354 non-null float64
jpg_url                 2075 non-null object
img_num                 2075 non-null float64
p1                      2075 non-null object
p1_conf                 2075 non-null float64
p1_dog                  2075 non-null object
p2                      2075 non-null object
p2_conf                 2075 non-null float64
p2_dog                  2075 non-null object
p3                      2075 non-null object
p3_conf                 2075 non-null float64
p3_dog                  2075 non-null object
dog_level               380 non-null category
dtypes: category(1), float64(10), int64(3), object(13)
memory usage: 499.6+ KB
```

**1.3.11 4: name column has many invalid value like a, the, None.**

**1.3.12 Define: Remove all the invalid values from column name**

**1.3.13 Code**

```
In [51]: df_clean.name=df_clean.name.replace(regex=['^[a-z]+', 'None'], value=np.nan)
```



### 1.3.14 Test

```
In [12]: sum(df_clean.name.isnull())
```

```
Out[12]: 854
```

```
In [52]: # trying to extract Dog name from the text column
def function(text):
    txt_list=text.split()
    for word in txt_list:
        if word.lower() == 'named':
            name_index=txt_list.index(word) + 1
            return txt_list[name_index]
        else:
            pass
    return np.nan
```

```
In [53]: # Putting a condition for the extracted name only to be replace on Null spaces
df_clean.name=np.where(df_clean.name.isnull(), df_clean.text.apply(function),df_clean.name)
```

```
In [14]: # testing the function
sum(df_clean.name.isnull())
```

```
Out[14]: 830
```

It worked well the null values has reduced to 830 from 854

### 1.3.15 Issue 5: Text column has the link of tweets and ratings

#### 1.3.16 Define: remove tweet link and ratings from text column

#### 1.3.17 Code

```
In [54]: df_clean.text=df_clean.text.str.extract('(.(?=\s\d+/\d+\s))')
```

### 1.3.18 Test

```
In [68]: df_clean.text.sample(5)
```

```
Out[68]: 285          RT @KibaDva: I collected all the good dogs!!
1318    This is Oliver (pronounced "Ricardo"). He's a ...
1284    Meet Bear. He's a Beneboop Cumberclap. Extreme...
494     We only rate dogs. Please don't send in other ...
626     This is Bailey. She loves going down slides bu...
Name: text, dtype: object
```

### 1.3.19 6: In rating\_denominator some entry are more than and below 10 need to be 10

Define: Remove all values other than 10

### Code

```
In [55]: df_clean=df_clean[df_clean['rating_denominator'] == 10]
```

### Test

```
In [22]: df_clean['rating_denominator'].describe()
```

```
Out[22]: count      2333.0  
         mean        10.0  
         std         0.0  
         min        10.0  
         25%        10.0  
         50%        10.0  
         75%        10.0  
         max        10.0  
         Name: rating_denominator, dtype: float64
```

### 1.3.20 7:Rating\_numerators some entry are less than 10 and supposed to be more than 10

Define Remove all values on rating\_numerator that are less than 10

### Code

```
In [56]: df_clean=df_clean[~(df_clean['rating_numerator']<= 9)]
```

### Test

```
In [24]: df_clean['rating_numerator'].describe()
```

```
Out[24]: count      1899.000000  
         mean       13.392312  
         std       45.381156  
         min       10.000000  
         25%       11.000000  
         50%       12.000000  
         75%       12.000000  
         max      1776.000000  
         Name: rating_numerator, dtype: float64
```

### 1.3.21 8: Timestamp column is string instead of datetime

### 1.3.22 Define: Change timestamp column to datetime

### Code

```
In [57]: df_clean.timestamp=pd.to_datetime(df_clean.timestamp)
```

## Test

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1899 entries, 0 to 2350
Data columns (total 30 columns):
tweet_id                1899 non-null int64
in_reply_to_status_id   62 non-null float64
in_reply_to_user_id     62 non-null float64
timestamp               1899 non-null datetime64[ns]
source                  1899 non-null object
text                    1807 non-null object
retweeted_status_id     172 non-null float64
retweeted_status_user_id 172 non-null float64
retweeted_status_timestamp 172 non-null object
expanded_urls           1853 non-null object
rating_numerator        1899 non-null int64
rating_denominator      1899 non-null int64
name                    1241 non-null object
doggo                   92 non-null object
floofer                 10 non-null object
pupper                  214 non-null object
puppo                   29 non-null object
retweet_count           1897 non-null float64
favorite_count          1897 non-null float64
jpg_url                 1646 non-null object
img_num                 1646 non-null float64
p1                      1646 non-null object
p1_conf                 1646 non-null float64
p1_dog                  1646 non-null object
p2                      1646 non-null object
p2_conf                 1646 non-null float64
p2_dog                  1646 non-null object
p3                      1646 non-null object
p3_conf                 1646 non-null float64
p3_dog                  1646 non-null object
dtypes: datetime64[ns](1), float64(10), int64(3), object(16)
memory usage: 459.9+ KB
```

### 1.3.23 9: Expanded url has Misssing Values

### 1.3.24 Define: Drop all rows with missing values

### 1.3.25 code

```
In [59]: df_clean.dropna(subset=['expanded_urls'], inplace=True)
```

### 1.3.26 Test

```
In [28]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1853 entries, 0 to 2350
Data columns (total 30 columns):
tweet_id                1853 non-null int64
in_reply_to_status_id   20 non-null float64
in_reply_to_user_id     20 non-null float64
timestamp               1853 non-null datetime64[ns]
source                  1853 non-null object
text                    1789 non-null object
retweeted_status_id      171 non-null float64
retweeted_status_user_id 171 non-null float64
retweeted_status_timestamp 171 non-null object
expanded_urls           1853 non-null object
rating_numerator         1853 non-null int64
rating_denominator       1853 non-null int64
name                    1241 non-null object
doggo                   89 non-null object
floofer                 10 non-null object
pupper                 212 non-null object
puppo                   28 non-null object
retweet_count           1851 non-null float64
favorite_count          1851 non-null float64
jpg_url                 1646 non-null object
img_num                 1646 non-null float64
p1                      1646 non-null object
p1_conf                 1646 non-null float64
p1_dog                  1646 non-null object
p2                      1646 non-null object
p2_conf                 1646 non-null float64
p2_dog                  1646 non-null object
p3                      1646 non-null object
p3_conf                 1646 non-null float64
p3_dog                  1646 non-null object
dtypes: datetime64[ns](1), float64(10), int64(3), object(16)
memory usage: 448.8+ KB
```

### 1.3.27 10:Tweet id have wrong datatype

### 1.3.28 Define: Change twitter\_id int to string

### 1.3.29 Code

```
In [61]: df_clean.tweet_id=df_clean.tweet_id.astype(str)
```

### 1.3.30 Test

```
In [74]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1853 entries, 0 to 2350
Data columns (total 27 columns):
tweet_id                1853 non-null object
in_reply_to_status_id   20 non-null float64
in_reply_to_user_id     20 non-null float64
timestamp               1853 non-null datetime64[ns]
source                  1853 non-null object
text                    1789 non-null object
retweeted_status_id      171 non-null float64
retweeted_status_user_id 171 non-null float64
retweeted_status_timestamp 171 non-null object
expanded_urls           1853 non-null object
rating_numerator         1853 non-null int64
rating_denominator       1853 non-null int64
name                    1256 non-null object
retweet_count            1851 non-null float64
favorite_count           1851 non-null float64
jpg_url                 1646 non-null object
img_num                 1646 non-null float64
p1                      1646 non-null object
p1_conf                 1646 non-null float64
p1_dog                  1646 non-null object
p2                      1646 non-null object
p2_conf                 1646 non-null float64
p2_dog                  1646 non-null object
p3                      1646 non-null object
p3_conf                 1646 non-null float64
p3_dog                  1646 non-null object
dog_level                326 non-null category
dtypes: category(1), datetime64[ns](1), float64(10), int64(2), object(13)
memory usage: 393.0+ KB
```

```
In [78]: df_clean=df_clean.drop(columns=['retweeted_status_id','retweeted_status_user_id','retweeted_status_timestamp'])
```

```
In [79]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1853 entries, 0 to 2350
Data columns (total 24 columns):
tweet_id                1853 non-null object
in_reply_to_status_id   20 non-null float64
in_reply_to_user_id     20 non-null float64
timestamp               1853 non-null datetime64[ns]
```

```

source          1853 non-null object
text            1789 non-null object
expanded_urls   1853 non-null object
rating_numerator 1853 non-null int64
rating_denominator 1853 non-null int64
name            1256 non-null object
retweet_count   1851 non-null float64
favorite_count  1851 non-null float64
jpg_url         1646 non-null object
img_num         1646 non-null float64
p1              1646 non-null object
p1_conf         1646 non-null float64
p1_dog          1646 non-null object
p2              1646 non-null object
p2_conf         1646 non-null float64
p2_dog          1646 non-null object
p3              1646 non-null object
p3_conf         1646 non-null float64
p3_dog          1646 non-null object
dog_level       326 non-null category
dtypes: category(1), datetime64[ns](1), float64(8), int64(2), object(12)
memory usage: 349.6+ KB

```

## 1.4 Storing Data

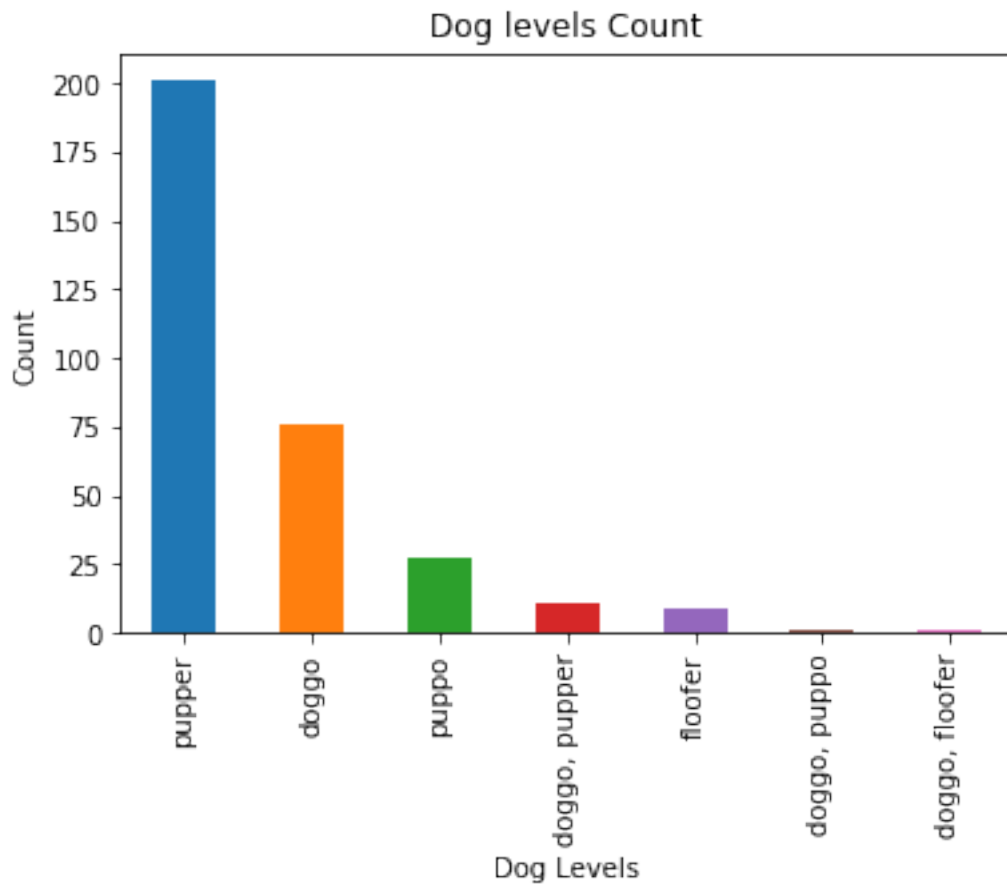
Save gathered, assessed, and cleaned master dataset to a CSV file named "twitter\_archive\_master.csv".

```
In [80]: df_clean.to_csv('twitter_archive_master.csv')
```

## 1.5 Analyzing and Visualizing Data

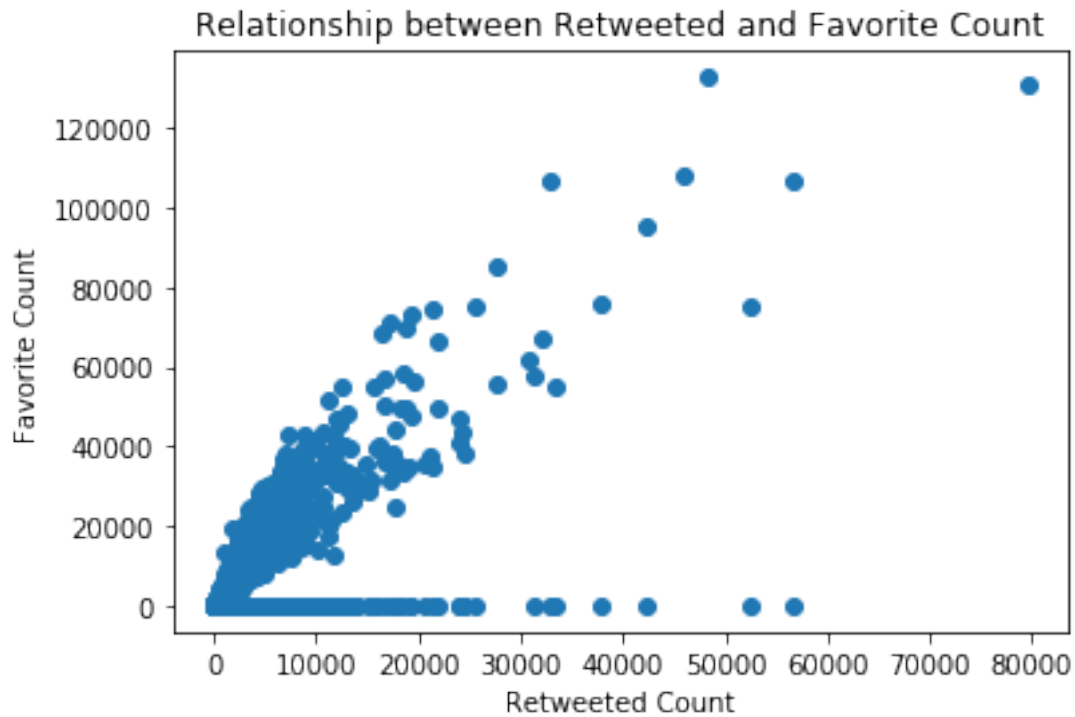
In this section, analyze and visualize your wrangled data. You must produce at least **three (3) insights and one (1) visualization**.

```
In [85]: df_clean['dog_level'].value_counts().plot(kind='bar')
plt.xlabel("Dog Levels")
plt.ylabel("Count")
plt.title("Dog levels Count");
```



Pupper is the Most common Dog

```
In [82]: #Relationship between retweet count and favourite count
plt.scatter(df_clean.retweet_count, df_clean.favorite_count)
plt.xlabel('Retweeted Count')
plt.ylabel('Favorite Count')
plt.title('Relationship between Retweeted and Favorite Count')
plt.show()
```



There is strong, positive relationship between Retweet count and favourite count and it's linear.

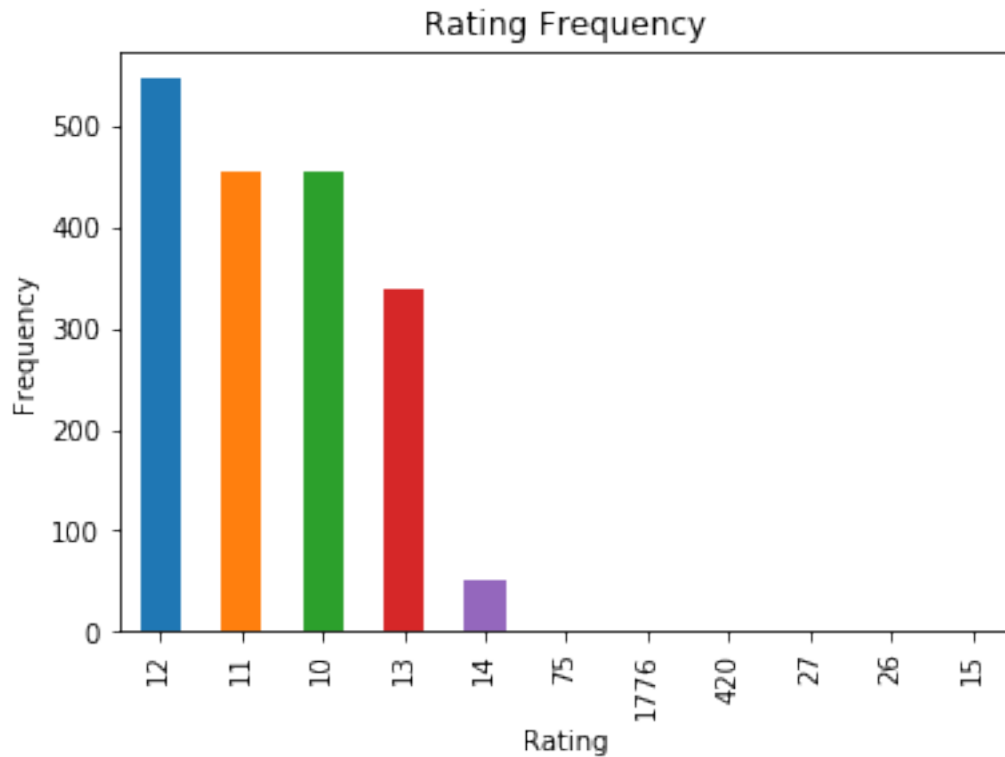
```
In [83]: df_clean.rating_numerator.value_counts()
```

```
Out[83]: 12      547
         11      455
         10      455
         13      339
         14       50
         75        2
        1776        1
         420        1
         27        1
         26        1
         15        1
        Name: rating_numerator, dtype: int64
```

Most dogs received rating of 12

```
In [28]: df_clean['rating_numerator'].value_counts().plot(kind='bar')
         plt.xlabel("Rating")
         plt.ylabel("Frequency")
         plt.title("Rating Frequency");
```





Most dogs receive rating rating of 12

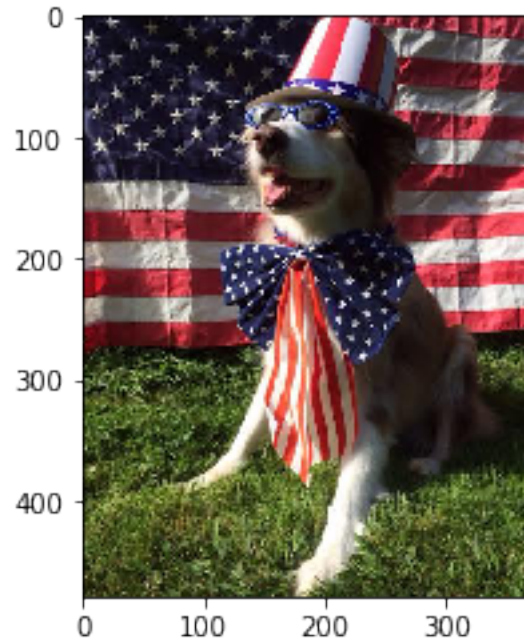
```
In [75]: highest_rating=df_clean[df_clean.rating_numerator==1776][['rating_numerator','name','jpg_url','text']]
```

```
In [76]: highest_rating
```

```
Out[76]:
```

	rating_numerator	name	jpg_url	text
979	1776	Atticus	https://pbs.twimg.com/media/CmgBZ7kWcAAlzFD.jpg	This is Atticus. He's quite simply America af.

```
In [72]: import requests, io
for index, row in highest_rating.iterrows():
    response = requests.get(row['jpg_url']).content
    img = plt.imread(io.BytesIO(response), format='JPG')
    plt.imshow(img)
plt.show()
```



Atticus was the most liked Dog has the highest number of rating (1776) even from the image its smart Dog

#### 1.5.1 Insights:

1. Pupper the most common dog level
2. Atticus dog with the highest rating 3.The retweeted count and favorite count are strongly related

#### 1.5.2 Visualization

The dog with highest rating. The relationship between Favorite count and retweeted count