

wrangle_act

November 24, 2019

1 Investigating WeRateDogs's Tweeter account

by modhi alhabradi

1.1 Installing needed packages and tweepy

```
In [55]: import tweepy
import pandas as pd
import numpy as np
import requests
import json
import time
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: consumer_key = 'YOUR CONSUMER KEY'
consumer_secret = 'YOUR CONSUMER SECRET'
access_token = 'YOUR ACCESS TOKEN'
access_secret = 'YOUR ACCESS SECRET'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)

api = tweepy.API(auth)
```

1.2 1. Data Gathering

```
In [3]: # Supplied file
tw_archive = pd.read_csv('twitter-archive-enhanced.csv', sep=',')
tw_archive.head()
```

```
Out[3]:
```

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

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

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

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

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

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

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

In [4]: # Requesting tweet image predictions

```
with open('image_predictions.tsv' , 'wb') as file:
    image_predictions = requests.get('https://d17h27t6h515a5.cloudfront.net/topher/2017/
    file.write(image_predictions.content)
```

```
In [5]: # Reading image predictions
```

```
image_predictions = pd.read_csv('image-predictions.tsv', sep='\t')
image_predictions.head()
```

```
Out[5]:
```

	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-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 [62]: # collecting data using API
```

```
collected = []
not_collected = [] # to avoid error/exception messages

with open('tweet_json.txt', 'w') as file:
    for tweet_id in list(tw_archive['tweet_id']):
        try:
            tweet_status = api.get_status(tweet_id, tweet_mode='extended')
            json.dump(tweet_status._json, file)
            file.write('\n')
            collected.append(tweet_id)
        except Exception as e:
            not_collected.append(tweet_id)
```

```
In [7]: #Reading JSON content as pandas dataframe
```

```
tweet_status = pd.read_json('tweet-json.txt', encoding='utf-8', lines = True)
```

```
In [8]: tweet_status.sample(6)
```

```
Out[8]:
```

	contributors	coordinates	created_at	display_text_range	\
692	NaN	NaN	2016-10-14 16:13:10	[0, 110]	
2322	NaN	NaN	2015-11-17 01:02:40	[0, 52]	
572	NaN	NaN	2016-11-22 20:58:07	[0, 56]	

2070	NaN	NaN	2015-11-29 23:30:32	[0, 135]
2010	NaN	NaN	2015-12-03 02:21:48	[0, 74]
1943	NaN	NaN	2015-12-07 03:34:14	[0, 116]

	entities \
692	{'hashtags': [], 'symbols': [], 'user_mentions...
2322	{'hashtags': [], 'symbols': [], 'user_mentions...
572	{'hashtags': [], 'symbols': [], 'user_mentions...
2070	{'hashtags': [], 'symbols': [], 'user_mentions...
2010	{'hashtags': [], 'symbols': [], 'user_mentions...
1943	{'hashtags': [], 'symbols': [], 'user_mentions...

	extended_entities	favorite_count \
692	{'media': [{ 'id': 786963058530906112, 'id_str'...	29725
2322	{'media': [{ 'id': 666421153020444674, 'id_str'...	327
572	{'media': [{ 'id': 801167887901724672, 'id_str'...	27386
2070	{'media': [{ 'id': 671109008019742721, 'id_str'...	1225
2010	{'media': [{ 'id': 672239273467359232, 'id_str'...	953
1943	{'media': [{ 'id': 673707042180341760, 'id_str'...	1177

	favorited	full_text	geo \
692	False	This is Rory. He's got an interview in a few m...	NaN
2322	False	*internally screaming* 12/10 https://t.co/YMcr...	NaN
572	False	This is Milo. I would do terrible things for M...	NaN
2070	False	This is Toby. He asked for chocolate cake for ...	NaN
2010	False	This is Ridley. He doesn't know how to couch. ...	NaN
1943	False	This is Raymond. He's absolutely terrified of ...	NaN

	...	\
692	...	
2322	...	
572	...	
2070	...	
2010	...	
1943	...	

	possibly_sensitive_appealable	quoted_status	quoted_status_id \
692	0.0	NaN	NaN
2322	0.0	NaN	NaN
572	0.0	NaN	NaN
2070	0.0	NaN	NaN
2010	0.0	NaN	NaN
1943	0.0	NaN	NaN

	quoted_status_id_str	retweet_count	retweeted	retweeted_status \
692	NaN	9327	False	NaN
2322	NaN	118	False	NaN
572	NaN	6961	False	NaN

2070	NaN	478	False	NaN
2010	NaN	347	False	NaN
1943	NaN	434	False	NaN

		source truncated	\
692	<a href="http://twitter.com/download/iphone" r...	False	
2322	<a href="http://twitter.com/download/iphone" r...	False	
572	<a href="http://twitter.com/download/iphone" r...	False	
2070	<a href="http://twitter.com/download/iphone" r...	False	
2010	<a href="http://twitter.com/download/iphone" r...	False	
1943	<a href="http://twitter.com/download/iphone" r...	False	

	user
692	{'id': 4196983835, 'id_str': '4196983835', 'na...
2322	{'id': 4196983835, 'id_str': '4196983835', 'na...
572	{'id': 4196983835, 'id_str': '4196983835', 'na...
2070	{'id': 4196983835, 'id_str': '4196983835', 'na...
2010	{'id': 4196983835, 'id_str': '4196983835', 'na...
1943	{'id': 4196983835, 'id_str': '4196983835', 'na...

[6 rows x 31 columns]

1.3 Data Assessment

1.3.1 Visual assessment

In [15]: tw_archive

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

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

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

5	2017-07-29	00:08:17	+0000
6	2017-07-28	16:27:12	+0000
7	2017-07-28	00:22:40	+0000
8	2017-07-27	16:25:51	+0000
9	2017-07-26	15:59:51	+0000
10	2017-07-26	00:31:25	+0000
11	2017-07-25	16:11:53	+0000
12	2017-07-25	01:55:32	+0000
13	2017-07-25	00:10:02	+0000
14	2017-07-24	17:02:04	+0000
15	2017-07-24	00:19:32	+0000
16	2017-07-23	00:22:39	+0000
17	2017-07-22	16:56:37	+0000
18	2017-07-22	00:23:06	+0000
19	2017-07-21	01:02:36	+0000
20	2017-07-20	16:49:33	+0000
21	2017-07-19	16:06:48	+0000
22	2017-07-19	03:39:09	+0000
23	2017-07-19	00:47:34	+0000
24	2017-07-18	16:08:03	+0000
25	2017-07-18	00:07:08	+0000
26	2017-07-17	16:17:36	+0000
27	2017-07-16	23:58:41	+0000
28	2017-07-16	20:14:00	+0000
29	2017-07-15	23:25:31	+0000
...			...
2326	2015-11-17	00:24:19	+0000
2327	2015-11-17	00:06:54	+0000
2328	2015-11-16	23:23:41	+0000
2329	2015-11-16	21:54:18	+0000
2330	2015-11-16	21:10:36	+0000
2331	2015-11-16	20:32:58	+0000
2332	2015-11-16	20:01:42	+0000
2333	2015-11-16	19:31:45	+0000
2334	2015-11-16	16:37:02	+0000
2335	2015-11-16	16:11:11	+0000
2336	2015-11-16	15:14:19	+0000
2337	2015-11-16	14:57:41	+0000
2338	2015-11-16	04:02:55	+0000
2339	2015-11-16	03:55:04	+0000
2340	2015-11-16	03:44:34	+0000
2341	2015-11-16	03:22:39	+0000
2342	2015-11-16	02:38:37	+0000
2343	2015-11-16	01:59:36	+0000
2344	2015-11-16	01:52:02	+0000
2345	2015-11-16	01:22:45	+0000
2346	2015-11-16	01:01:59	+0000
2347	2015-11-16	00:55:59	+0000

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

```

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

```


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

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

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

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

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

2351	NaN	NaN
2352	NaN	NaN
2353	NaN	NaN
2354	NaN	NaN
2355	NaN	NaN

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

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

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

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

[2356 rows x 17 columns]

In [16]: image_predictions

Out[16]:	tweet_id	jpg_url \
0	666020888022790149	https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
5	666050758794694657	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6	666051853826850816	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg

7	666055525042405380	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8	666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9	666058600524156928	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10	666063827256086533	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11	666071193221509120	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
12	666073100786774016	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13	666082916733198337	https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg
14	666094000022159362	https://pbs.twimg.com/media/CT5w9gUW4AAAsBNN.jpg
15	666099513787052032	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16	666102155909144576	https://pbs.twimg.com/media/CT54YGiWUAENZnoK.jpg
17	666104133288665088	https://pbs.twimg.com/media/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/CT9Vn7PWAA_ZCM.jpg
24	666353288456101888	https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25	666362758909284353	https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg
26	666373753744588802	https://pbs.twimg.com/media/CT9vZEYWUAA1Z05.jpg
27	666396247373291520	https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28	666407126856765440	https://pbs.twimg.com/media/CT-NvwmW4AAUGGZ.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/DE5Se8FXcAAAJFx4.jpg
2048	886983233522544640	https://pbs.twimg.com/media/DE8yicJW0AAAABJ.jpg
2049	887101392804085760	https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050	887343217045368832	https://pbs.twimg.com/ext_tw_video_thumb/88734...
2051	887473957103951883	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2052	887517139158093824	https://pbs.twimg.com/ext_tw_video_thumb/88751...
2053	887705289381826560	https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054	888078434458587136	https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
2055	888202515573088257	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056	888554962724278272	https://pbs.twimg.com/media/DFTH_0-UQAAACu20.jpg
2057	888804989199671297	https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058	888917238123831296	https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg
2059	889278841981685760	https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060	889531135344209921	https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg
2061	889638837579907072	https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062	889665388333682689	https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063	889880896479866881	https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg
2064	890006608113172480	https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065	890240255349198849	https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg
2066	890609185150312448	https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067	890729181411237888	https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068	890971913173991426	https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg

2069	891087950875897856	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070	891327558926688256	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071	891689557279858688	https://pbs.twimg.com/media/DF_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 [17]: tweet_status
```

```

Out[17]:      contributors  coordinates      created_at  display_text_range \
0          NaN          NaN  2017-08-01 16:23:56      [0, 85]
1          NaN          NaN  2017-08-01 00:17:27      [0, 138]
2          NaN          NaN  2017-07-31 00:18:03      [0, 121]
3          NaN          NaN  2017-07-30 15:58:51      [0, 79]
4          NaN          NaN  2017-07-29 16:00:24      [0, 138]
5          NaN          NaN  2017-07-29 00:08:17      [0, 138]
6          NaN          NaN  2017-07-28 16:27:12      [0, 140]
7          NaN          NaN  2017-07-28 00:22:40      [0, 118]
8          NaN          NaN  2017-07-27 16:25:51      [0, 122]
9          NaN          NaN  2017-07-26 15:59:51      [0, 133]
10         NaN          NaN  2017-07-26 00:31:25      [0, 130]
11         NaN          NaN  2017-07-25 16:11:53      [0, 107]
12         NaN          NaN  2017-07-25 01:55:32      [0, 106]
13         NaN          NaN  2017-07-25 00:10:02      [0, 91]
14         NaN          NaN  2017-07-24 17:02:04      [0, 118]
15         NaN          NaN  2017-07-24 00:19:32      [0, 138]
16         NaN          NaN  2017-07-23 00:22:39      [0, 86]
17         NaN          NaN  2017-07-22 16:56:37      [0, 128]
18         NaN          NaN  2017-07-22 00:23:06      [0, 87]
19         NaN          NaN  2017-07-20 16:49:33      [0, 127]
20         NaN          NaN  2017-07-19 16:06:48      [0, 127]
21         NaN          NaN  2017-07-19 03:39:09      [0, 108]
22         NaN          NaN  2017-07-19 00:47:34      [0, 99]
23         NaN          NaN  2017-07-18 16:08:03      [0, 88]
24         NaN          NaN  2017-07-18 00:07:08      [0, 129]
25         NaN          NaN  2017-07-17 16:17:36      [0, 101]
26         NaN          NaN  2017-07-16 23:58:41      [0, 121]
27         NaN          NaN  2017-07-16 20:14:00      [0, 71]
28         NaN          NaN  2017-07-15 23:25:31      [0, 131]
29         NaN          NaN  2017-07-15 16:51:35      [27, 105]
...         ...         ...         ...         ...
2324        NaN          NaN  2015-11-17 00:24:19      [0, 140]
2325        NaN          NaN  2015-11-17 00:06:54      [0, 139]
2326        NaN          NaN  2015-11-16 23:23:41      [0, 137]
2327        NaN          NaN  2015-11-16 21:54:18      [0, 81]
2328        NaN          NaN  2015-11-16 21:10:36      [0, 140]
2329        NaN          NaN  2015-11-16 20:32:58      [0, 135]
2330        NaN          NaN  2015-11-16 20:01:42      [0, 112]
2331        NaN          NaN  2015-11-16 19:31:45      [0, 139]

```

2332	NaN	NaN	2015-11-16 16:37:02	[0, 138]
2333	NaN	NaN	2015-11-16 16:11:11	[0, 136]
2334	NaN	NaN	2015-11-16 15:14:19	[0, 46]
2335	NaN	NaN	2015-11-16 14:57:41	[0, 82]
2336	NaN	NaN	2015-11-16 04:02:55	[0, 134]
2337	NaN	NaN	2015-11-16 03:55:04	[0, 128]
2338	NaN	NaN	2015-11-16 03:44:34	[0, 140]
2339	NaN	NaN	2015-11-16 03:22:39	[0, 132]
2340	NaN	NaN	2015-11-16 02:38:37	[0, 125]
2341	NaN	NaN	2015-11-16 01:59:36	[0, 137]
2342	NaN	NaN	2015-11-16 01:52:02	[0, 137]
2343	NaN	NaN	2015-11-16 01:22:45	[0, 107]
2344	NaN	NaN	2015-11-16 01:01:59	[0, 135]
2345	NaN	NaN	2015-11-16 00:55:59	[0, 124]
2346	NaN	NaN	2015-11-16 00:49:46	[0, 140]
2347	NaN	NaN	2015-11-16 00:35:11	[0, 138]
2348	NaN	NaN	2015-11-16 00:30:50	[0, 140]
2349	NaN	NaN	2015-11-16 00:24:50	[0, 120]
2350	NaN	NaN	2015-11-16 00:04:52	[0, 137]
2351	NaN	NaN	2015-11-15 23:21:54	[0, 130]
2352	NaN	NaN	2015-11-15 23:05:30	[0, 139]
2353	NaN	NaN	2015-11-15 22:32:08	[0, 131]

	entities \
0	{'hashtags': [], 'symbols': [], 'user_mentions...
1	{'hashtags': [], 'symbols': [], 'user_mentions...
2	{'hashtags': [], 'symbols': [], 'user_mentions...
3	{'hashtags': [], 'symbols': [], 'user_mentions...
4	{'hashtags': [{'text': 'BarkWeek', 'indices': ...
5	{'hashtags': [{'text': 'BarkWeek', 'indices': ...
6	{'hashtags': [], 'symbols': [], 'user_mentions...
7	{'hashtags': [], 'symbols': [], 'user_mentions...
8	{'hashtags': [{'text': 'BarkWeek', 'indices': ...
9	{'hashtags': [], 'symbols': [], 'user_mentions...
10	{'hashtags': [{'text': 'BarkWeek', 'indices': ...
11	{'hashtags': [], 'symbols': [], 'user_mentions...
12	{'hashtags': [], 'symbols': [], 'user_mentions...
13	{'hashtags': [], 'symbols': [], 'user_mentions...
14	{'hashtags': [{'text': 'BarkWeek', 'indices': ...
15	{'hashtags': [{'text': 'BarkWeek', 'indices': ...
16	{'hashtags': [], 'symbols': [], 'user_mentions...
17	{'hashtags': [], 'symbols': [], 'user_mentions...
18	{'hashtags': [], 'symbols': [], 'user_mentions...
19	{'hashtags': [], 'symbols': [], 'user_mentions...
20	{'hashtags': [], 'symbols': [], 'user_mentions...
21	{'hashtags': [], 'symbols': [], 'user_mentions...
22	{'hashtags': [], 'symbols': [], 'user_mentions...
23	{'hashtags': [], 'symbols': [], 'user_mentions...

```

24   {'hashtags': [], 'symbols': [], 'user_mentions...
25   {'hashtags': [], 'symbols': [], 'user_mentions...
26   {'hashtags': [], 'symbols': [], 'user_mentions...
27   {'hashtags': [], 'symbols': [], 'user_mentions...
28   {'hashtags': [], 'symbols': [], 'user_mentions...
29   {'hashtags': [], 'symbols': [], 'user_mentions...
...
2324 {'hashtags': [], 'symbols': [], 'user_mentions...
2325 {'hashtags': [], 'symbols': [], 'user_mentions...
2326 {'hashtags': [], 'symbols': [], 'user_mentions...
2327 {'hashtags': [], 'symbols': [], 'user_mentions...
2328 {'hashtags': [], 'symbols': [], 'user_mentions...
2329 {'hashtags': [], 'symbols': [], 'user_mentions...
2330 {'hashtags': [], 'symbols': [], 'user_mentions...
2331 {'hashtags': [], 'symbols': [], 'user_mentions...
2332 {'hashtags': [], 'symbols': [], 'user_mentions...
2333 {'hashtags': [], 'symbols': [], 'user_mentions...
2334 {'hashtags': [], 'symbols': [], 'user_mentions...
2335 {'hashtags': [], 'symbols': [], 'user_mentions...
2336 {'hashtags': [], 'symbols': [], 'user_mentions...
2337 {'hashtags': [], 'symbols': [], 'user_mentions...
2338 {'hashtags': [], 'symbols': [], 'user_mentions...
2339 {'hashtags': [], 'symbols': [], 'user_mentions...
2340 {'hashtags': [], 'symbols': [], 'user_mentions...
2341 {'hashtags': [], 'symbols': [], 'user_mentions...
2342 {'hashtags': [], 'symbols': [], 'user_mentions...
2343 {'hashtags': [], 'symbols': [], 'user_mentions...
2344 {'hashtags': [], 'symbols': [], 'user_mentions...
2345 {'hashtags': [], 'symbols': [], 'user_mentions...
2346 {'hashtags': [], 'symbols': [], 'user_mentions...
2347 {'hashtags': [], 'symbols': [], 'user_mentions...
2348 {'hashtags': [], 'symbols': [], 'user_mentions...
2349 {'hashtags': [], 'symbols': [], 'user_mentions...
2350 {'hashtags': [], 'symbols': [], 'user_mentions...
2351 {'hashtags': [], 'symbols': [], 'user_mentions...
2352 {'hashtags': [], 'symbols': [], 'user_mentions...
2353 {'hashtags': [], 'symbols': [], 'user_mentions...

```

		extended_entities	favorite_count	\
0	{'media': [{ 'id': 892420639486877696, 'id_str'...		39467	
1	{'media': [{ 'id': 892177413194625024, 'id_str'...		33819	
2	{'media': [{ 'id': 891815175371796480, 'id_str'...		25461	
3	{'media': [{ 'id': 891689552724799489, 'id_str'...		42908	
4	{'media': [{ 'id': 891327551943041024, 'id_str'...		41048	
5	{'media': [{ 'id': 891087942176911360, 'id_str'...		20562	
6	{'media': [{ 'id': 890971906207338496, 'id_str'...		12041	
7	{'media': [{ 'id': 890729118844600320, 'id_str'...		56848	
8	{'media': [{ 'id': 890609177319665665, 'id_str'...		28226	

9	{'media': [{'id': 890240245463175168, 'id_str'...	32467
10	{'media': [{'id': 890006600089468928, 'id_str'...	31166
11	{'media': [{'id': 889880888800096258, 'id_str'...	28268
12	{'media': [{'id': 889665366129029120, 'id_str'...	38818
13	{'media': [{'id': 889638825424826374, 'id_str'...	27672
14	{'media': [{'id': 889531127467266049, 'id_str'...	15359
15	{'media': [{'id': 889278779352338437, 'id_str'...	25652
16	{'media': [{'id': 888917229776945152, 'id_str'...	29611
17	{'media': [{'id': 888804981515575296, 'id_str'...	26080
18	{'media': [{'id': 888554915546542081, 'id_str'...	20290
19	{'media': [{'id': 888078426338406400, 'id_str'...	22201
20	{'media': [{'id': 887705281597243393, 'id_str'...	30779
21	{'media': [{'id': 887517108413886465, 'id_str'...	46959
22	{'media': [{'id': 887473949361045505, 'id_str'...	69871
23	{'media': [{'id': 887343120832229379, 'id_str'...	34222
24	{'media': [{'id': 887101385971384320, 'id_str'...	31061
25	{'media': [{'id': 886983218871902208, 'id_str'...	35859
26	{'media': [{'id': 886736868116754432, 'id_str'...	12306
27	{'media': [{'id': 886680331239161856, 'id_str'...	22798
28	{'media': [{'id': 886366138128449536, 'id_str'...	21524
29	NaN	117
...
2324	{'media': [{'id': 666411498068123649, 'id_str'...	459
2325	{'media': [{'id': 666407121513275392, 'id_str'...	113
2326	{'media': [{'id': 666396240351993856, 'id_str'...	172
2327	{'media': [{'id': 666373746337402880, 'id_str'...	194
2328	{'media': [{'id': 666362717482020864, 'id_str'...	804
2329	{'media': [{'id': 666353280906170368, 'id_str'...	229
2330	{'media': [{'id': 666345414279471104, 'id_str'...	307
2331	{'media': [{'id': 666337857791987715, 'id_str'...	204
2332	{'media': [{'id': 666293909010702337, 'id_str'...	522
2333	{'media': [{'id': 666287399580733440, 'id_str'...	152
2334	{'media': [{'id': 666273081518768128, 'id_str'...	184
2335	{'media': [{'id': 666268904428277760, 'id_str'...	108
2336	{'media': [{'id': 666104129232740352, 'id_str'...	14765
2337	{'media': [{'id': 666102150364286977, 'id_str'...	81
2338	{'media': [{'id': 666099505364733952, 'id_str'...	164
2339	{'media': [{'id': 666093996847063040, 'id_str'...	169
2340	{'media': [{'id': 666082912819875840, 'id_str'...	121
2341	{'media': [{'id': 666073098362486784, 'id_str'...	335
2342	{'media': [{'id': 666071190449033216, 'id_str'...	154
2343	{'media': [{'id': 666063820255862784, 'id_str'...	496
2344	{'media': [{'id': 666058597072306176, 'id_str'...	115
2345	{'media': [{'id': 666057085227016192, 'id_str'...	304
2346	{'media': [{'id': 666055517517848576, 'id_str'...	448
2347	{'media': [{'id': 666051848592334848, 'id_str'...	1253
2348	{'media': [{'id': 666050754986266625, 'id_str'...	136
2349	{'media': [{'id': 666049244999131136, 'id_str'...	111

2350	{'media': [{'id': 666044217047650304, 'id_str'...	311
2351	{'media': [{'id': 666033409081393153, 'id_str'...	128
2352	{'media': [{'id': 666029276303482880, 'id_str'...	132
2353	{'media': [{'id': 666020881337073664, 'id_str'...	2535

	favorited	full_text	geo \
0	False	This is Phineas. He's a mystical boy. Only eve...	NaN
1	False	This is Tilly. She's just checking pup on you...	NaN
2	False	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	False	This is Darla. She commenced a snooze mid meal...	NaN
4	False	This is Franklin. He would like you to stop ca...	NaN
5	False	Here we have a majestic great white breaching ...	NaN
6	False	Meet Jax. He enjoys ice cream so much he gets ...	NaN
7	False	When you watch your owner call another dog a g...	NaN
8	False	This is Zoey. She doesn't want to be one of th...	NaN
9	False	This is Cassie. She is a college pup. Studying...	NaN
10	False	This is Koda. He is a South Australian decksha...	NaN
11	False	This is Bruno. He is a service shark. Only get...	NaN
12	False	Here's a puppo that seems to be on the fence a...	NaN
13	False	This is Ted. He does his best. Sometimes that'...	NaN
14	False	This is Stuart. He's sporting his favorite fan...	NaN
15	False	This is Oliver. You're witnessing one of his m...	NaN
16	False	This is Jim. He found a fren. Taught him how t...	NaN
17	False	This is Zeke. He has a new stick. Very proud o...	NaN
18	False	This is Ralphus. He's powering up. Attempting ...	NaN
19	False	This is Gerald. He was just told he didn't get...	NaN
20	False	This is Jeffrey. He has a monopoly on the pool...	NaN
21	True	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN
22	False	This is Canela. She attempted some fancy porch...	NaN
23	False	You may not have known you needed to see this ...	NaN
24	False	This... is a Jubilant Antarctic House Bear. We...	NaN
25	False	This is Maya. She's very shy. Rarely leaves he...	NaN
26	False	This is Mingus. He's a wonderful father to his...	NaN
27	False	This is Derek. He's late for a dog meeting. 13...	NaN
28	False	This is Roscoe. Another pupper fallen victim t...	NaN
29	False	@NonWhiteHat @MayhewMayhem omg hello tanner yo...	NaN
...
2324	False	This is quite the dog. Gets really excited whe...	NaN
2325	False	This is a southern Vesuvius bumblegruff. Can d...	NaN
2326	False	Oh goodness. A super rare northeast Qdoba kang...	NaN
2327	False	Those are sunglasses and a jean jacket. 11/10 ...	NaN
2328	False	Unique dog here. Very small. Lives in containe...	NaN
2329	False	Here we have a mixed Asiago from the Galápagos...	NaN
2330	False	Look at this jokester thinking seat belt laws ...	NaN
2331	False	This is an extremely rare horned Parthenon. No...	NaN
2332	False	This is a funny dog. Weird toes. Won't come do...	NaN
2333	False	This is an Albanian 3 1/2 legged Episcopalian...	NaN
2334	False	Can take selfies 11/10 https://t.co/ws2AMaWpW	NaN

2335	False	Very concerned about fellow dog trapped in com...	NaN
2336	False	Not familiar with this breed. No tail (weird)...	NaN
2337	False	Oh my. Here you are seeing an Adobe Setter giv...	NaN
2338	False	Can stand on stump for what seems like a while...	NaN
2339	False	This appears to be a Mongolian Presbyterian mi...	NaN
2340	False	Here we have a well-established sunblockerspan...	NaN
2341	False	Let's hope this flight isn't Malaysian (lol). ...	NaN
2342	False	Here we have a northern speckled Rhododendron...	NaN
2343	False	This is the happiest dog you will ever see. Ve...	NaN
2344	False	Here is the Rand Paul of retrievers folks! He'...	NaN
2345	False	My oh my. This is a rare blond Canadian terrie...	NaN
2346	False	Here is a Siberian heavily armored polar bear ...	NaN
2347	False	This is an odd dog. Hard on the outside but lo...	NaN
2348	False	This is a truly beautiful English Wilson Staff...	NaN
2349	False	Here we have a 1949 1st generation vulpix. Enj...	NaN
2350	False	This is a purebred Piers Morgan. Loves to Netf...	NaN
2351	False	Here is a very happy pup. Big fan of well-main...	NaN
2352	False	This is a western brown Mitsubishi terrier. Up...	NaN
2353	False	Here we have a Japanese Irish Setter. Lost eye...	NaN

	...	\
0	...	
1	...	
2	...	
3	...	
4	...	
5	...	
6	...	
7	...	
8	...	
9	...	
10	...	
11	...	
12	...	
13	...	
14	...	
15	...	
16	...	
17	...	
18	...	
19	...	
20	...	
21	...	
22	...	
23	...	
24	...	
25	...	
26	...	

27	...
28	...
29	...
...	...
2324	...
2325	...
2326	...
2327	...
2328	...
2329	...
2330	...
2331	...
2332	...
2333	...
2334	...
2335	...
2336	...
2337	...
2338	...
2339	...
2340	...
2341	...
2342	...
2343	...
2344	...
2345	...
2346	...
2347	...
2348	...
2349	...
2350	...
2351	...
2352	...
2353	...

	possibly_sensitive_appealable	quoted_status	quoted_status_id \
0	0.0	NaN	NaN
1	0.0	NaN	NaN
2	0.0	NaN	NaN
3	0.0	NaN	NaN
4	0.0	NaN	NaN
5	0.0	NaN	NaN
6	0.0	NaN	NaN
7	0.0	NaN	NaN
8	0.0	NaN	NaN
9	0.0	NaN	NaN
10	0.0	NaN	NaN
11	0.0	NaN	NaN

12	0.0	NaN	NaN
13	0.0	NaN	NaN
14	0.0	NaN	NaN
15	0.0	NaN	NaN
16	0.0	NaN	NaN
17	0.0	NaN	NaN
18	0.0	NaN	NaN
19	0.0	NaN	NaN
20	0.0	NaN	NaN
21	0.0	NaN	NaN
22	0.0	NaN	NaN
23	0.0	NaN	NaN
24	0.0	NaN	NaN
25	0.0	NaN	NaN
26	0.0	NaN	NaN
27	0.0	NaN	NaN
28	0.0	NaN	NaN
29	NaN	NaN	NaN
...
2324	0.0	NaN	NaN
2325	0.0	NaN	NaN
2326	0.0	NaN	NaN
2327	0.0	NaN	NaN
2328	0.0	NaN	NaN
2329	0.0	NaN	NaN
2330	0.0	NaN	NaN
2331	0.0	NaN	NaN
2332	0.0	NaN	NaN
2333	0.0	NaN	NaN
2334	0.0	NaN	NaN
2335	0.0	NaN	NaN
2336	0.0	NaN	NaN
2337	0.0	NaN	NaN
2338	0.0	NaN	NaN
2339	0.0	NaN	NaN
2340	0.0	NaN	NaN
2341	0.0	NaN	NaN
2342	0.0	NaN	NaN
2343	0.0	NaN	NaN
2344	0.0	NaN	NaN
2345	0.0	NaN	NaN
2346	0.0	NaN	NaN
2347	0.0	NaN	NaN
2348	0.0	NaN	NaN
2349	0.0	NaN	NaN
2350	0.0	NaN	NaN
2351	0.0	NaN	NaN
2352	0.0	NaN	NaN

2353		0.0	NaN	NaN	
	quoted_status_id_str	retweet_count	retweeted	retweeted_status	\
0	NaN	8853	False	NaN	
1	NaN	6514	False	NaN	
2	NaN	4328	False	NaN	
3	NaN	8964	False	NaN	
4	NaN	9774	False	NaN	
5	NaN	3261	False	NaN	
6	NaN	2158	False	NaN	
7	NaN	16716	False	NaN	
8	NaN	4429	False	NaN	
9	NaN	7711	False	NaN	
10	NaN	7624	False	NaN	
11	NaN	5156	False	NaN	
12	NaN	8538	False	NaN	
13	NaN	4735	False	NaN	
14	NaN	2321	False	NaN	
15	NaN	5637	False	NaN	
16	NaN	4709	False	NaN	
17	NaN	4559	False	NaN	
18	NaN	3732	False	NaN	
19	NaN	3653	False	NaN	
20	NaN	5609	False	NaN	
21	NaN	12082	False	NaN	
22	NaN	18781	False	NaN	
23	NaN	10737	False	NaN	
24	NaN	6167	False	NaN	
25	NaN	8084	False	NaN	
26	NaN	3443	False	NaN	
27	NaN	4610	False	NaN	
28	NaN	3316	False	NaN	
29	NaN	4	False	NaN	
...	
2324	NaN	339	False	NaN	
2325	NaN	44	False	NaN	
2326	NaN	92	False	NaN	
2327	NaN	100	False	NaN	
2328	NaN	595	False	NaN	
2329	NaN	77	False	NaN	
2330	NaN	146	False	NaN	
2331	NaN	96	False	NaN	
2332	NaN	368	False	NaN	
2333	NaN	71	False	NaN	
2334	NaN	82	False	NaN	
2335	NaN	37	False	NaN	
2336	NaN	6871	False	NaN	
2337	NaN	16	False	NaN	

2338	NaN	73	False	NaN
2339	NaN	79	False	NaN
2340	NaN	47	False	NaN
2341	NaN	174	False	NaN
2342	NaN	67	False	NaN
2343	NaN	232	False	NaN
2344	NaN	61	False	NaN
2345	NaN	146	False	NaN
2346	NaN	261	False	NaN
2347	NaN	879	False	NaN
2348	NaN	60	False	NaN
2349	NaN	41	False	NaN
2350	NaN	147	False	NaN
2351	NaN	47	False	NaN
2352	NaN	48	False	NaN
2353	NaN	532	False	NaN

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

...		False
2324	<a href="http://twitter.com/download/iphone"	False
2325	<a href="http://twitter.com/download/iphone"	False
2326	<a href="http://twitter.com/download/iphone"	False
2327	<a href="http://twitter.com/download/iphone"	False
2328	<a href="http://twitter.com/download/iphone"	False
2329	<a href="http://twitter.com/download/iphone"	False
2330	<a href="http://twitter.com/download/iphone"	False
2331	<a href="http://twitter.com/download/iphone"	False
2332	<a href="http://twitter.com/download/iphone"	False
2333	<a href="http://twitter.com/download/iphone"	False
2334	<a href="http://twitter.com/download/iphone"	False
2335	<a href="http://twitter.com/download/iphone"	False
2336	<a href="http://twitter.com/download/iphone"	False
2337	<a href="http://twitter.com/download/iphone"	False
2338	<a href="http://twitter.com/download/iphone"	False
2339	<a href="http://twitter.com/download/iphone"	False
2340	<a href="http://twitter.com/download/iphone"	False
2341	<a href="http://twitter.com/download/iphone"	False
2342	<a href="http://twitter.com/download/iphone"	False
2343	<a href="http://twitter.com/download/iphone"	False
2344	<a href="http://twitter.com/download/iphone"	False
2345	<a href="http://twitter.com/download/iphone"	False
2346	<a href="http://twitter.com/download/iphone"	False
2347	<a href="http://twitter.com/download/iphone"	False
2348	<a href="http://twitter.com/download/iphone"	False
2349	<a href="http://twitter.com/download/iphone"	False
2350	<a href="http://twitter.com/download/iphone"	False
2351	<a href="http://twitter.com/download/iphone"	False
2352	<a href="http://twitter.com/download/iphone"	False
2353	<a href="http://twitter.com/download/iphone"	False

				user
0	{	'id': 4196983835,	'id_str': '4196983835',	'na...
1	{	'id': 4196983835,	'id_str': '4196983835',	'na...
2	{	'id': 4196983835,	'id_str': '4196983835',	'na...
3	{	'id': 4196983835,	'id_str': '4196983835',	'na...
4	{	'id': 4196983835,	'id_str': '4196983835',	'na...
5	{	'id': 4196983835,	'id_str': '4196983835',	'na...
6	{	'id': 4196983835,	'id_str': '4196983835',	'na...
7	{	'id': 4196983835,	'id_str': '4196983835',	'na...
8	{	'id': 4196983835,	'id_str': '4196983835',	'na...
9	{	'id': 4196983835,	'id_str': '4196983835',	'na...
10	{	'id': 4196983835,	'id_str': '4196983835',	'na...
11	{	'id': 4196983835,	'id_str': '4196983835',	'na...
12	{	'id': 4196983835,	'id_str': '4196983835',	'na...
13	{	'id': 4196983835,	'id_str': '4196983835',	'na...
14	{	'id': 4196983835,	'id_str': '4196983835',	'na...

[illegible]

```
[2354 rows x 31 columns]
```

1.3.2 Programmatic assessment

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

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

```
In [19]: tw_archive.isnull().sum()
```

```
Out[19]: tweet_id                0
         in_reply_to_status_id    2278
         in_reply_to_user_id      2278
         timestamp                0
         source                   0
         text                     0
         retweeted_status_id      2175
         retweeted_status_user_id 2175
         retweeted_status_timestamp 2175
         expanded_urls            59
         rating_numerator         0
         rating_denominator       0
         name                     0
         doggo                    0
         floofer                  0
         pupper                   0
         puppo                    0
         dtype: int64
```

```
In [20]: tw_archive.describe()
```



```

Out[20]:
      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 [21]: tw_archive.rating_denominator.value_counts()
```

```

Out[21]:
10      2333
11         3
50         3
80         2
20         2
2          1
16         1
40         1
70         1
15         1
90         1
110        1
120        1
130        1
150        1
170        1

```

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

```
In [22]: tw_archive.rating_numerator.value_counts()
```

```
Out[22]: 12      558
        11      464
        10      461
        13      351
         9      158
         8      102
         7       55
        14       54
         5       37
         6       32
         3       19
         4       17
         1        9
         2        9
        420        2
         0        2
        15        2
        75        2
        80        1
        20        1
        24        1
        26        1
        44        1
        50        1
        60        1
       165        1
        84        1
        88        1
       144        1
       182        1
       143        1
       666        1
       960        1
      1776        1
        17        1
        27        1
        45        1
        99        1
       121        1
       204        1
Name: rating_numerator, dtype: int64
```

```
In [23]: tw_archive.source.value_counts()
```

```
Out[23]: <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>
<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
<a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
Name: source, dtype: int64
```

```
In [24]: image_predictions.head()
```

```
Out[24]:
```

	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-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 [25]: image_predictions.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 [26]: image_predictions.describe()
```

```
Out[26]:
```

	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 [27]: tweet_status.head()
```

```
Out[27]:
```

	contributors	coordinates	created_at	display_text_range	\
0	NaN	NaN	2017-08-01 16:23:56	[0, 85]	
1	NaN	NaN	2017-08-01 00:17:27	[0, 138]	
2	NaN	NaN	2017-07-31 00:18:03	[0, 121]	
3	NaN	NaN	2017-07-30 15:58:51	[0, 79]	
4	NaN	NaN	2017-07-29 16:00:24	[0, 138]	

	entities	\
0	{'hashtags': [], 'symbols': [], 'user_mentions': ...}	
1	{'hashtags': [], 'symbols': [], 'user_mentions': ...}	
2	{'hashtags': [], 'symbols': [], 'user_mentions': ...}	
3	{'hashtags': [], 'symbols': [], 'user_mentions': ...}	
4	{'hashtags': [{'text': 'BarkWeek', 'indices': ...}	

	extended_entities	favorite_count	\
0	{'media': [{'id': 892420639486877696, 'id_str': ...}	39467	
1	{'media': [{'id': 892177413194625024, 'id_str': ...}	33819	
2	{'media': [{'id': 891815175371796480, 'id_str': ...}	25461	
3	{'media': [{'id': 891689552724799489, 'id_str': ...}	42908	
4	{'media': [{'id': 891327551943041024, 'id_str': ...}	41048	

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

	...	\
0	...	
1	...	
2	...	
3	...	
4	...	

	possibly_sensitive_appealable	quoted_status	quoted_status_id	\
0	0.0	NaN	NaN	
1	0.0	NaN	NaN	
2	0.0	NaN	NaN	
3	0.0	NaN	NaN	
4	0.0	NaN	NaN	

	quoted_status_id_str	retweet_count	retweeted	retweeted_status	\
0	NaN	8853	False	NaN	
1	NaN	6514	False	NaN	
2	NaN	4328	False	NaN	
3	NaN	8964	False	NaN	
4	NaN	9774	False	NaN	

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

	user
0	{'id': 4196983835, 'id_str': '4196983835', 'na...
1	{'id': 4196983835, 'id_str': '4196983835', 'na...
2	{'id': 4196983835, 'id_str': '4196983835', 'na...
3	{'id': 4196983835, 'id_str': '4196983835', 'na...
4	{'id': 4196983835, 'id_str': '4196983835', 'na...

[5 rows x 31 columns]

In [28]: tweet_status.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 31 columns):
contributors      0 non-null float64
coordinates       0 non-null float64
created_at        2354 non-null datetime64[ns]
display_text_range 2354 non-null object
entities          2354 non-null object
extended_entities 2073 non-null object
favorite_count    2354 non-null int64
favorited         2354 non-null bool
full_text         2354 non-null object
geo              0 non-null float64
id               2354 non-null int64
id_str           2354 non-null int64
```

```

in_reply_to_screen_name      78 non-null object
in_reply_to_status_id        78 non-null float64
in_reply_to_status_id_str    78 non-null float64
in_reply_to_user_id          78 non-null float64
in_reply_to_user_id_str      78 non-null float64
is_quote_status              2354 non-null bool
lang                          2354 non-null object
place                         1 non-null object
possibly_sensitive            2211 non-null float64
possibly_sensitive_appealable 2211 non-null float64
quoted_status                28 non-null object
quoted_status_id             29 non-null float64
quoted_status_id_str         29 non-null float64
retweet_count                 2354 non-null int64
retweeted                     2354 non-null bool
retweeted_status              179 non-null object
source                        2354 non-null object
truncated                     2354 non-null bool
user                          2354 non-null object
dtypes: bool(4), datetime64[ns](1), float64(11), int64(4), object(11)
memory usage: 505.8+ KB

```

In [29]: tweet_status.describe()

```

Out[29]:
      contributors  coordinates  favorite_count  geo  id  \
count           0.0           0.0      2354.000000  0.0  2.354000e+03
mean           NaN           NaN       8080.968564  NaN  7.426978e+17
std            NaN           NaN      11814.771334  NaN  6.852812e+16
min            NaN           NaN           0.000000  NaN  6.660209e+17
25%            NaN           NaN       1415.000000  NaN  6.783975e+17
50%            NaN           NaN       3603.500000  NaN  7.194596e+17
75%            NaN           NaN      10122.250000  NaN  7.993058e+17
max            NaN           NaN     132810.000000  NaN  8.924206e+17

      id_str  in_reply_to_status_id  in_reply_to_status_id_str  \
count  2.354000e+03      7.800000e+01      7.800000e+01
mean   7.426978e+17      7.455079e+17      7.455079e+17
std    6.852812e+16      7.582492e+16      7.582492e+16
min    6.660209e+17      6.658147e+17      6.658147e+17
25%    6.783975e+17      6.757419e+17      6.757419e+17
50%    7.194596e+17      7.038708e+17      7.038708e+17
75%    7.993058e+17      8.257804e+17      8.257804e+17
max    8.924206e+17      8.862664e+17      8.862664e+17

      in_reply_to_user_id  in_reply_to_user_id_str  possibly_sensitive  \
count      7.800000e+01      7.800000e+01      2211.0
mean       2.014171e+16      2.014171e+16      0.0

```

std	1.252797e+17	1.252797e+17	0.0
min	1.185634e+07	1.185634e+07	0.0
25%	3.086374e+08	3.086374e+08	0.0
50%	4.196984e+09	4.196984e+09	0.0
75%	4.196984e+09	4.196984e+09	0.0
max	8.405479e+17	8.405479e+17	0.0

	possibly_sensitive_appealable	quoted_status_id	quoted_status_id_str \
count	2211.0	2.900000e+01	2.900000e+01
mean	0.0	8.162686e+17	8.162686e+17
std	0.0	6.164161e+16	6.164161e+16
min	0.0	6.721083e+17	6.721083e+17
25%	0.0	7.888183e+17	7.888183e+17
50%	0.0	8.340867e+17	8.340867e+17
75%	0.0	8.664587e+17	8.664587e+17
max	0.0	8.860534e+17	8.860534e+17

	retweet_count
count	2354.000000
mean	3164.797366
std	5284.770364
min	0.000000
25%	624.500000
50%	1473.500000
75%	3652.000000
max	79515.000000

1.3.3 Assessments for quality:

1.3.4 Tw_archive

1- Some columns have huge amount of missing values, for example, "in_reply_to_status_id", "in_reply_to_user_id", "retweeted_status_id", "in_reply_to_user_id", "retweeted_status_id", "retweeted_status_user_id", "retweeted_status_timestamp". Since I don't need in_reply and retweet data in this project, I prefer to delete those columns directly. 2-The variable "expanded_urls" also has few missing values, which means some records had no images. Any ratings without images should not be taking into account. 3-Rename columns with more appropriate names: "timestamp" to "tweet_timestamp", "text" to "tweet_text", "name" to "dog_name". 4- Optimize the source contents for human reading habit; change the long url links to certain words. 5- Multiple cases where numerator of rating > 10. These entries will be assessed and removed if not authentic. 6-Multiple cases of where the numerator of rating < 10. These entries will be removed.

1.3.5 image_predictions

1-The "p1" and "p1_conf" columns will be renamed with more explanatory titles. 2-The column "jpg_url" will be removed since url data is already contained in the twitter archive data.

1.3.6 tweet_status

1- Drop unneeded columns

1.4 Assessments for tidiness:

Tw_archive

1-The columns (doggo, floofer, pupper and puppo) do not need to be separated. Each dog will be classified as one of these classifications. It is better to create one column for dog classification that contains the values (doggo, floofer, pupper and puppo). 2-Not all columns are needed for further analysis

1.4.1 tweet_status

1-Rename the column id to be tweet_id to facilitate merging

1.4.2 Merge all documents using tweet_id

2 Data Cleaning

```
In [2]: # Reading gathered files
```

```
tw_archive = pd.read_csv('twitter-archive-enhanced.csv')
image_predictions = pd.read_csv('image_predictions.tsv', sep='\t')
tweet_status = pd.read_json('tweet-json.txt', lines = True, encoding='utf-8')
```

```
In [3]: # create a cleaned dataframe from the archive
```

```
cleaned_archive = tw_archive.copy()
```

```
In [4]: # test
```

```
cleaned_archive.info()
```

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

```
RangeIndex: 2356 entries, 0 to 2355
```

```
Data columns (total 17 columns):
```

tweet_id	2356 non-null int64
in_reply_to_status_id	78 non-null float64
in_reply_to_user_id	78 non-null float64
timestamp	2356 non-null object
source	2356 non-null object
text	2356 non-null object
retweeted_status_id	181 non-null float64
retweeted_status_user_id	181 non-null float64
retweeted_status_timestamp	181 non-null object
expanded_urls	2297 non-null object
rating_numerator	2356 non-null int64
rating_denominator	2356 non-null int64
name	2356 non-null object
doggo	2356 non-null object
floofer	2356 non-null object


```
pupper                2356 non-null object
puppo                 2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

```
In [5]: # Define: Dropping unneded (missing too much values) columns using drop function.
```

```
    # code
    cleaned_archive = cleaned_archive.drop(['in_reply_to_status_id', 'in_reply_to_user_id',
```

```
In [6]: # Test
```

```
    cleaned_archive.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id            2356 non-null int64
timestamp           2356 non-null object
source              2356 non-null object
text                2356 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: int64(3), object(9)
memory usage: 221.0+ KB
```

```
In [7]: #Define
```

```
    # drop the 59 records without images
    cleaned_archive= cleaned_archive.dropna(subset = ['expanded_urls'])
```

```
In [8]: # Define
```

```
    # Rename columns with more appropriate names: "timestamp" to "tweet_timestamp",
    #      "text" to "tweet_text", "name" to "dog_name",
```

```
    # Code
```

```
    new_names = {'timestamp': 'tweet_timestamp', 'text': 'tweet_text',\
                  'name': 'dog_name'}
```

```
    cleaned_archive.rename(columns=new_names, inplace=True)
```

```
In [9]: cleaned_archive.head()
```

```

Out[9]:
      tweet_id      tweet_timestamp \
0  892420643555336193  2017-08-01 16:23:56 +0000
1  892177421306343426  2017-08-01 00:17:27 +0000
2  891815181378084864  2017-07-31 00:18:03 +0000
3  891689557279858688  2017-07-30 15:58:51 +0000
4  891327558926688256  2017-07-29 16:00:24 +0000

      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...

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

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

```

```
In [10]: # simplify the source contexts
```

```

cleaned_archive['source'] = cleaned_archive['source'].replace('<a href="http://twitter.'
                                                                'Twitter for
cleaned_archive['source'] = cleaned_archive['source'].replace('<a href="http://vine.co
                                                                'Vine - Make
cleaned_archive['source'] = cleaned_archive['source'].replace('<a href="http://twitter
                                                                'Twitter Web
cleaned_archive['source'] = cleaned_archive['source'].replace('<a href="https://about.
                                                                'TweetDeck')

# change the type of source to category for preparing later visualization
cleaned_archive['source'] = cleaned_archive['source'].astype('category')

```

```
In [11]: cleaned_archive.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2297 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id          2297 non-null int64
tweet_timestamp   2297 non-null object
source            2297 non-null category
tweet_text        2297 non-null object
expanded_urls     2297 non-null object
rating_numerator  2297 non-null int64
rating_denominator 2297 non-null int64
dog_name          2297 non-null object
doggo             2297 non-null object
floofer           2297 non-null object
pupper           2297 non-null object
puppo             2297 non-null object
dtypes: category(1), int64(3), object(8)
memory usage: 217.8+ KB
```

```
In [12]: cleaned_archive.source.value_counts()
```

```
Out[12]: Twitter for iphone    2163
         Vine - Make a Scene    91
         Twitter Web Client    32
         TweetDeck             11
         Name: source, dtype: int64
```

```
In [13]: cleaned_archive.rating_denominator.value_counts()
```

```
Out[13]: 10      2278
         11       3
         50       3
         80       2
          7       1
        170       1
        150       1
        130       1
        120       1
        110       1
         90       1
         70       1
         40       1
         20       1
          2       1
         Name: rating_denominator, dtype: int64
```

```
In [14]: cleaned_archive.rating_numerator.value_counts()
```

```

Out[14]: 12      547
        10      455
        11      455
        13      339
        9       156
        8       102
        7        53
        14       50
        5        35
        6        32
        3        19
        4        16
        2         9
        1         6
        0         2
        75        2
        60         1
        24         1
        26         1
        44         1
        50         1
        165        1
        80         1
        84         1
        88         1
        144        1
        143        1
        420        1
        1776       1
        15         1
        27         1
        45         1
        99         1
        121        1
        204        1
        Name: rating_numerator, dtype: int64

In [15]: # filter the wrong rating_denominator values

df1 = cleaned_archive[cleaned_archive['rating_denominator'] != 10]

df1[['tweet_id', 'tweet_text', 'rating_numerator', 'rating_denominator']].values

Out[15]: array([[820690176645140481,
        'The floofs have been released I repeat the floofs have been released. 84/70 ht
        84, 70],
        [810984652412424192,
        'Meet Sam. She smiles 24/7 & secretly aspires to be a reindeer. \nKeep Sam

```

24, 7],
 [775096608509886464,
 'RT @dog_rates: After so many requests, this is Bretagne. She was the last surviving 9, 11],
 [758467244762497024,
 'Why does this never happen at my front door... 165/150 <https://t.co/HmwrdfEfUE>
 165, 150],
 [740373189193256964,
 'After so many requests, this is Bretagne. She was the last surviving 9/11 search 9, 11],
 [731156023742988288,
 'Say hello to this unbelievably well behaved squad of doggos. 204/170 would try 204, 170],
 [722974582966214656,
 'Happy 4/20 from the squad! 13/10 for all <https://t.co/eVldiws8a>',
 4, 20],
 [716439118184652801,
 'This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. A 50, 50],
 [713900603437621249,
 "Happy Saturday here's 9 puppies on a bench. 99/90 good work everybody <https://t.co/9dpcw6>
 99, 90],
 [710658690886586372,
 "Here's a brigade of puppies. All look very prepared for whatever happens next. 80, 80],
 [709198395643068416,
 'From left to right:\nCletus, Jerome, Alejandro, Burp, & Titson\nNone know 45, 50],
 [704054845121142784,
 "Here is a whole flock of puppies. 60/50 I'll take the lot <https://t.co/9dpcw6>
 60, 50],
 [697463031882764288,
 "Happy Wednesday here's a bucket of pups. 44/40 would pet all at once <https://t.co/9dpcw6>
 44, 40],
 [684225744407494656,
 'Two sneaky puppies were not initially seen, moving the rating to 143/130. Please 143, 130],
 [684222868335505415,
 'Someone help the girl is being mugged. Several are distracting her while two search 121, 110],
 [682962037429899265,
 'This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was 7, 11],
 [677716515794329600,
 "IT'S PUPPERGEDDON. Total of 144/120 ...I think <https://t.co/ZanVtAtvIq>",
 144, 120],
 [675853064436391936,
 'Here we have an entire platoon of puppies. Total score: 88/80 would pet all at

```

88, 80],
[666287406224695296,
'This is an Albanian 3 1/2 legged Episcopalian. Loves well-polished hardwood f
1, 2]], dtype=object)

```

```

In [16]: # tweet_id : 740373189193256964,722974582966214656,716439118184652801,68296203742989926

```

```

cleaned_archive.loc[cleaned_archive.tweet_id == 740373189193256964, ['rating_denominator',
cleaned_archive.loc[cleaned_archive.tweet_id == 722974582966214656, ['rating_denominator',
cleaned_archive.loc[cleaned_archive.tweet_id == 716439118184652801, ['rating_denominator',
cleaned_archive.loc[cleaned_archive.tweet_id == 682962037429899265, ['rating_denominator',
cleaned_archive.loc[cleaned_archive.tweet_id == 666287406224695296, ['rating_denominator',

```

```

In [17]: # tweet_id : 740373189193256964,722974582966214656,716439118184652801,68296203742989926

```

```

cleaned_archive.loc[cleaned_archive.tweet_id == 740373189193256964, ['rating_numerator',
cleaned_archive.loc[cleaned_archive.tweet_id == 722974582966214656, ['rating_numerator',
cleaned_archive.loc[cleaned_archive.tweet_id == 716439118184652801, ['rating_numerator',
cleaned_archive.loc[cleaned_archive.tweet_id == 682962037429899265, ['rating_numerator',
cleaned_archive.loc[cleaned_archive.tweet_id == 666287406224695296, ['rating_numerator',

```

```

In [18]: #test

```

```

cleaned_archive.loc[cleaned_archive.tweet_id == 740373189193256964]

```

```

Out[18]:
          tweet_id      tweet_timestamp      source \
1068  740373189193256964  2016-06-08 02:41:38 +0000  Twitter for iphone

          tweet_text \
1068  After so many requests, this is Bretagne. She ...

          expanded_urls  rating_numerator \
1068  https://twitter.com/dog_rates/status/740373189...      14

          rating_denominator  dog_name  doggo  floofer  pupper  puppo
1068              10      None  None      None  None  None

```

```

In [ ]:

```

```

In [19]: # Define: exclude all extreme and zero values from the numertor and denominator

```

```

#Code
cleaned_archive = cleaned_archive[cleaned_archive['rating_numerator'] != 0 ]
cleaned_archive = cleaned_archive[cleaned_archive['rating_denominator'] == 10 ]
cleaned_archive = cleaned_archive[cleaned_archive['rating_numerator'] <= 20 ]
#cleaned_archive = cleaned_archive[cleaned_archive['rating_denominator'] < 20 ]

```

```

In [20]: # test

```

```

len(cleaned_archive[cleaned_archive['rating_numerator'] > 20 ])

```

```

Out[20]: 0

```

```
In [21]: cleaned_archive.rating_numerator.value_counts()
```

```
Out[21]: 12    547
          11    456
          10    456
          13    340
           9    155
           8    102
           7     52
          14     51
           5     35
           6     32
           3     19
           4     15
           2      9
           1      5
          15      1
          Name: rating_numerator, dtype: int64
```

```
In [22]: cleaned_archive.rating_denominator.value_counts()
```

```
Out[22]: 10    2275
          Name: rating_denominator, dtype: int64
```

2.0.1 image_predictions

```
In [45]: image_predictions.head()
```

```
Out[45]:
```

	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-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 [23]: cleaned_images = image_predictions.copy()

In [24]: #Define
         # change the column names
         cleaned_images.rename(columns={'p1': 'first_prediction', 'p1_conf': 'first_confidence',
                                         'p2': 'second_prediction', 'p2_conf': 'second_confidence',
                                         'p3': 'third_prediction', 'p3_conf': 'third_confidence'})

In [25]: #test
         cleaned_images.columns

Out[25]: Index(['tweet_id', 'jpg_url', 'img_num', 'first_prediction',
                'first_confidence', 'first_dog', 'second_prediction',
                'second_confidence', 'second_dog', 'third_prediction',
                'third_confidence', 'third_dog'],
                dtype='object')

In [26]: cleaned_images.drop('jpg_url', inplace=True, axis=1)

In [27]: cleaned_images.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 11 columns):
tweet_id          2075 non-null int64
img_num           2075 non-null int64
first_prediction   2075 non-null object
first_confidence   2075 non-null float64
first_dog          2075 non-null bool
second_prediction  2075 non-null object
second_confidence  2075 non-null float64
second_dog         2075 non-null bool
third_prediction   2075 non-null object
third_confidence   2075 non-null float64
third_dog          2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(3)
memory usage: 135.8+ KB

```

2.0.2 tweet_status

```

In [28]: cleaned_status = tweet_status.copy()

In [29]: cleaned_status.head()

Out[29]:   contributors  coordinates  created_at  display_text_range \
0          NaN          NaN  2017-08-01 16:23:56          [0, 85]
1          NaN          NaN  2017-08-01 00:17:27          [0, 138]
2          NaN          NaN  2017-07-31 00:18:03          [0, 121]

```


3	NaN	NaN 2017-07-30 15:58:51	[0, 79]
4	NaN	NaN 2017-07-29 16:00:24	[0, 138]

	entities \
0	{'hashtags': [], 'symbols': [], 'user_mentions...
1	{'hashtags': [], 'symbols': [], 'user_mentions...
2	{'hashtags': [], 'symbols': [], 'user_mentions...
3	{'hashtags': [], 'symbols': [], 'user_mentions...
4	{'hashtags': [{'text': 'BarkWeek', 'indices': ...

	extended_entities	favorite_count \
0	{'media': [{'id': 892420639486877696, 'id_str'...	39467
1	{'media': [{'id': 892177413194625024, 'id_str'...	33819
2	{'media': [{'id': 891815175371796480, 'id_str'...	25461
3	{'media': [{'id': 891689552724799489, 'id_str'...	42908
4	{'media': [{'id': 891327551943041024, 'id_str'...	41048

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

	...	\
0	...	
1	...	
2	...	
3	...	
4	...	

	possibly_sensitive_appealable	quoted_status	quoted_status_id \
0	0.0	NaN	NaN
1	0.0	NaN	NaN
2	0.0	NaN	NaN
3	0.0	NaN	NaN
4	0.0	NaN	NaN

	quoted_status_id_str	retweet_count	retweeted	retweeted_status \
0	NaN	8853	False	NaN
1	NaN	6514	False	NaN
2	NaN	4328	False	NaN
3	NaN	8964	False	NaN
4	NaN	9774	False	NaN

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

```

2 <a href="http://twitter.com/download/iphone" r... False
3 <a href="http://twitter.com/download/iphone" r... False
4 <a href="http://twitter.com/download/iphone" r... False

```

```

                                user
0 {'id': 4196983835, 'id_str': '4196983835', 'na...
1 {'id': 4196983835, 'id_str': '4196983835', 'na...
2 {'id': 4196983835, 'id_str': '4196983835', 'na...
3 {'id': 4196983835, 'id_str': '4196983835', 'na...
4 {'id': 4196983835, 'id_str': '4196983835', 'na...

```

[5 rows x 31 columns]

```

In [30]: # Define: filter/select needed columns
# Code
cleaned_status = cleaned_status.filter(['id', 'favorite_count', 'retweet_count'] )

```

```

In [33]: # test
cleaned_status.head()

```

```

Out[33]:
      id  favorite_count  retweet_count
0  892420643555336193      39467         8853
1  892177421306343426      33819         6514
2  891815181378084864      25461         4328
3  891689557279858688      42908         8964
4  891327558926688256      41048         9774

```

2.0.3 Assessments for tidiness:

```

In [34]: # Define: use loc to add a new column dog_class = doggo, floofer, pupper or poppo. NaN

# Code
cleaned_archive.loc[cleaned_archive['doggo'] == 'doggo', 'dog_class'] = 'doggo'
cleaned_archive.loc[cleaned_archive['floofer'] == 'floofer', 'dog_class'] = 'floofer'
cleaned_archive.loc[cleaned_archive['pupper'] == 'pupper', 'dog_class'] = 'pupper'
cleaned_archive.loc[cleaned_archive['puppo'] == 'puppo', 'dog_class'] = 'puppo'

```

```

In [35]: # test
cleaned_archive.sample(8)

```

```

Out[35]:
      tweet_id  tweet_timestamp  source \
2183  668989615043424256  2015-11-24 03:08:48 +0000  Twitter for iphone
966   750383411068534784  2016-07-05 17:38:41 +0000  Twitter for iphone
219   850145622816686080  2017-04-07 00:38:06 +0000  Twitter for iphone
58    880935762899988482  2017-06-30 23:47:07 +0000  Twitter for iphone
1997  672591271085670400  2015-12-04 01:40:29 +0000  Twitter for iphone
1898  674690135443775488  2015-12-09 20:40:38 +0000  Twitter for iphone
305   836260088725786625  2017-02-27 17:01:56 +0000  Twitter for iphone
2167  669359674819481600  2015-11-25 03:39:17 +0000  Twitter for iphone

```

	tweet_text \
2183	This is Bernie. He's taking his Halloween cost...
966	This is Zoe. She was trying to stealthily take...
219	This is Riley. He's making new friends. Jubila...
58	This is Louis. He's crossing. It's a big deal...
1997	Lots of pups here. All are Judea Hazelnuts. Ex...
1898	Meet Patrick. He's an exotic pup. Jumps great ...
305	This is Lucy. She spent all morning overseeing...
2167	Meet Saydee. She's a Rochester Ecclesiastical...

	expanded_urls	rating_numerator \
2183	https://twitter.com/dog_rates/status/668989615...	3
966	https://twitter.com/dog_rates/status/750383411...	9
219	https://twitter.com/dog_rates/status/850145622...	11
58	https://twitter.com/dog_rates/status/880935762...	13
1997	https://twitter.com/dog_rates/status/672591271...	8
1898	https://twitter.com/dog_rates/status/674690135...	3
305	https://twitter.com/dog_rates/status/836260088...	13
2167	https://twitter.com/dog_rates/status/669359674...	11

	rating_denominator	dog_name	doggo	floofer	pupper	puppo	dog_class
2183	10	Bernie	None	None	None	None	NaN
966	10	Zoe	None	None	pupper	None	pupper
219	10	Riley	None	None	None	None	NaN
58	10	Louis	None	None	None	None	NaN
1997	10	None	None	None	None	None	NaN
1898	10	Patrick	None	None	None	None	NaN
305	10	Lucy	None	None	None	None	NaN
2167	10	Saydee	None	None	None	None	NaN

In [36]: *# dropping unneded doggo, floofer, pupper or puppo columns*

```
cleaned_archive = cleaned_archive.drop(['doggo', 'floofer', 'pupper', 'puppo'], axis =
```

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2275 entries, 0 to 2355
Data columns (total 9 columns):
tweet_id          2275 non-null int64
tweet_timestamp   2275 non-null object
source            2275 non-null category
tweet_text        2275 non-null object
expanded_urls     2275 non-null object
rating_numerator  2275 non-null int64
rating_denominator 2275 non-null int64
dog_name          2275 non-null object
dog_class         373 non-null object
```

```
dtypes: category(1), int64(3), object(5)
memory usage: 162.4+ KB
```

2.0.4 tweet_status

```
In [38]: # Define: rename the column id
```

```
# Code
```

```
cleaned_status = cleaned_status.rename(columns={'id': 'tweet_id'})
```

```
In [39]: cleaned_status.head()
```

```
Out[39]:
```

	tweet_id	favorite_count	retweet_count
0	892420643555336193	39467	8853
1	892177421306343426	33819	6514
2	891815181378084864	25461	4328
3	891689557279858688	42908	8964
4	891327558926688256	41048	9774

2.1 Merging documents to form a working dataframe

```
In [40]: tweeter_df = pd.merge(cleaned_archive, cleaned_images, how='outer', on=['tweet_id'])
```

```
In [41]: tweeter_df = pd.merge(tweeter_df, cleaned_status, how = 'outer', on=['tweet_id'])
```

```
In [42]: tweeter_df.sample(5)
```

```
Out[42]:
```

	tweet_id	tweet_timestamp	source \
2037	670428280563085312	2015-11-28 02:25:32 +0000	Twitter for iphone
2094	669324657376567296	2015-11-25 01:20:08 +0000	Twitter for iphone
1754	676496375194980353	2015-12-14 20:17:59 +0000	Twitter for iphone
1582	683834909291606017	2016-01-04 02:18:42 +0000	Twitter for iphone
1923	672523490734551040	2015-12-03 21:11:09 +0000	Twitter for iphone

	tweet_text \
2037	This is Willy. He's millennial af. 11/10 https...
2094	Meet Ralf. He's a miniature Buick DiCaprio. Ca...
1754	Say hello to Griffin. He's upset because his c...
1582	Here we see a faulty pupper. Might need to rep...
1923	When she says she'll be ready in a minute but ...

	expanded_urls	rating_numerator \
2037	https://twitter.com/dog_rates/status/670428280...	11.0
2094	https://twitter.com/dog_rates/status/669324657...	11.0
1754	https://twitter.com/dog_rates/status/676496375...	9.0
1582	https://twitter.com/dog_rates/status/683834909...	9.0
1923	https://twitter.com/dog_rates/status/672523490...	10.0

	rating_denominator	dog_name	dog_class	img_num	...	\
2037	10.0	Willy	NaN	1.0	...	
2094	10.0	Ralf	NaN	1.0	...	
1754	10.0	Griffin	NaN	1.0	...	
1582	10.0	None	pupper	1.0	...	
1923	10.0	None	NaN	1.0	...	

	first_confidence	first_dog	second_prediction	second_confidence	\
2037	0.335269	True	golden_retriever	0.305850	
2094	0.201659	False	Cardigan	0.131544	
1754	0.985387	True	Norwegian_elkhound	0.004417	
1582	0.738449	True	toy_poodle	0.102992	
1923	0.565981	True	chow	0.081212	

	second_dog	third_prediction	third_confidence	third_dog	favorite_count	\
2037	True	Tibetan_mastiff	0.063325	True	1484.0	
2094	True	sandbar	0.101430	False	525.0	
1754	True	French_bulldog	0.003893	True	1641.0	
1582	True	Samoyed	0.023247	True	2880.0	
1923	True	Irish_terrier	0.061596	True	696.0	

	retweet_count
2037	694.0
2094	223.0
1754	602.0
1582	1265.0
1923	189.0

[5 rows x 21 columns]

```
In [43]: tweeter_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 21 columns):
tweet_id          2356 non-null int64
tweet_timestamp   2275 non-null object
source            2275 non-null category
tweet_text        2275 non-null object
expanded_urls     2275 non-null object
rating_numerator  2275 non-null float64
rating_denominator 2275 non-null float64
dog_name          2275 non-null object
dog_class         373 non-null object
img_num          2075 non-null float64
first_prediction  2075 non-null object
first_confidence  2075 non-null float64
first_dog         2075 non-null object
```

```

second_prediction      2075 non-null object
second_confidence      2075 non-null float64
second_dog             2075 non-null object
third_prediction        2075 non-null object
third_confidence       2075 non-null float64
third_dog              2075 non-null object
favorite_count         2354 non-null float64
retweet_count          2354 non-null float64
dtypes: category(1), float64(8), int64(1), object(11)
memory usage: 389.0+ KB

```

```

In [44]: # Saving df as csv
         tweeter_df.to_csv('twitter_archive_master.csv')

```

2.2 Analyze and Visualize

```

In [45]: tweeter_df.head()

```

```

Out[45]:
   tweet_id      tweet_timestamp      source \
0  892420643555336193  2017-08-01 16:23:56 +0000  Twitter for iphone
1  892177421306343426  2017-08-01 00:17:27 +0000  Twitter for iphone
2  891815181378084864  2017-07-31 00:18:03 +0000  Twitter for iphone
3  891689557279858688  2017-07-30 15:58:51 +0000  Twitter for iphone
4  891327558926688256  2017-07-29 16:00:24 +0000  Twitter for iphone

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

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

   rating_denominator  dog_name  dog_class  img_num  ... \
0          10.0    Phineas      NaN      1.0    ...
1          10.0     Tilly      NaN      1.0    ...
2          10.0    Archie      NaN      1.0    ...
3          10.0     Darla      NaN      1.0    ...
4          10.0   Franklin      NaN      2.0    ...

   first_confidence  first_dog  second_prediction  second_confidence \

```

0	0.097049	False	bagel	0.085851
1	0.323581	True	Pekinese	0.090647
2	0.716012	True	malamute	0.078253
3	0.170278	False	Labrador_retriever	0.168086
4	0.555712	True	English_springer	0.225770

	second_dog	third_prediction	third_confidence	third_dog \
0	False	banana	0.076110	False
1	True	papillon	0.068957	True
2	True	kelpie	0.031379	True
3	True	spatula	0.040836	False
4	True	German_short-haired_pointer	0.175219	True

	favorite_count	retweet_count
0	39467.0	8853.0
1	33819.0	6514.0
2	25461.0	4328.0
3	42908.0	8964.0
4	41048.0	9774.0

[5 rows x 21 columns]

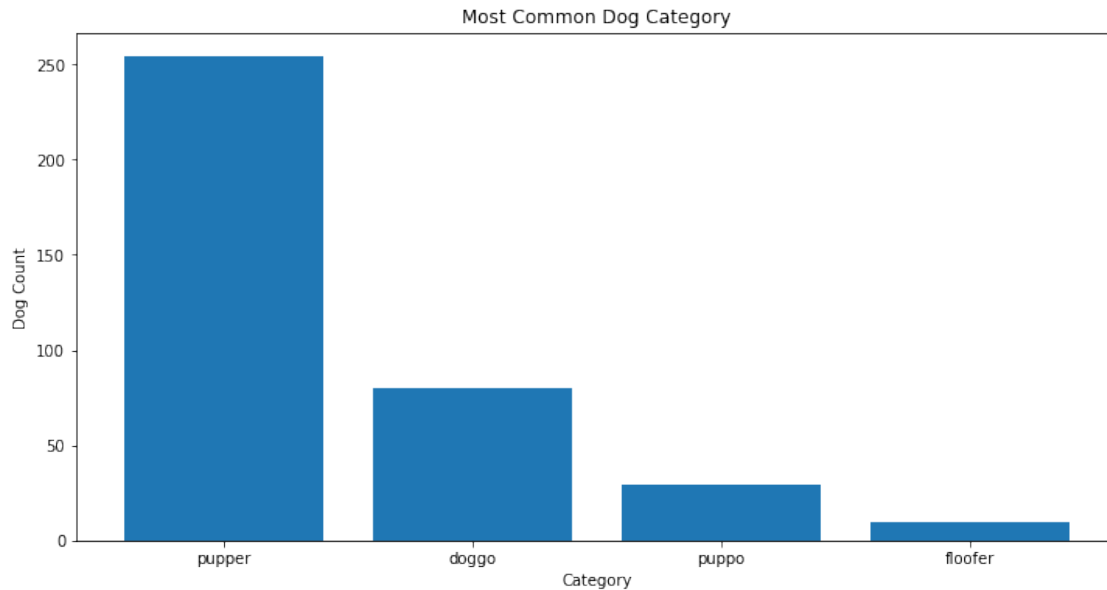
2.3 Most Common Dog Category

```
In [47]: tweeter_df.dog_class.value_counts()
```

```
Out[47]: pupper      254
         doggo       80
         puppo       29
         floofer     10
         Name: dog_class, dtype: int64
```

```
In [48]: dog_type = ['pupper', 'doggo', 'puppo', 'floofer']
         dog_counts = [254, 80, 29, 10]
```

```
fig,ax = plt.subplots(figsize = (12,6))
ax.bar(dog_type, dog_counts, width = 0.8)
ax.set_ylabel('Dog Count')
ax.set_xlabel('Category')
plt.title("Most Common Dog Category")
plt.show()
```



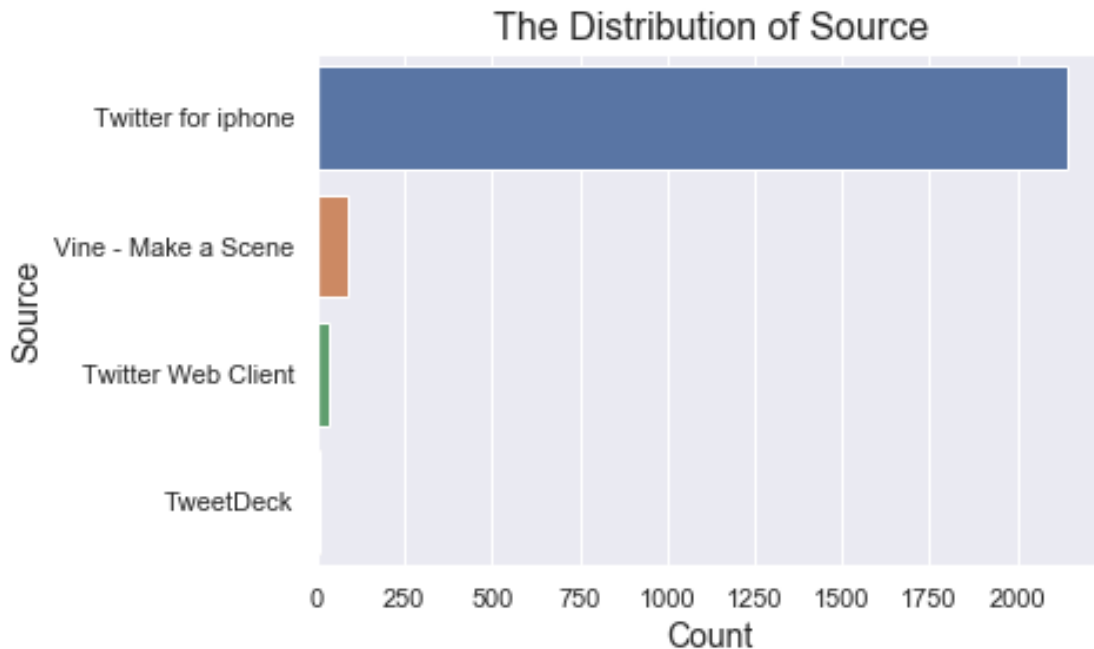
Similarly, I check the distribution of dog stages. It shows that ‘pupper’ is the most popular dog stage, followed by ‘doggo’ and ‘puppo’. It could be due to the young and unmatured dog is usually cuter than the adult dog. It should also be noticed that there’s huge amount missing data in dog stages, thus the distribution may not reflect the truth.

2.4 The Distribution of Source

```
In [56]: sorted_source = cleaned_archive['source'].value_counts().index
print(cleaned_archive['source'].value_counts())
sns.set(style="darkgrid")
sns.countplot(data = cleaned_archive, y = 'source', order = sorted_source)
plt.xticks(rotation = 360)
plt.xlabel('Count', fontsize=14)
plt.ylabel('Source', fontsize=14)
plt.title('The Distribution of Source', fontsize=16)
```

```
Twitter for iphone      2142
Vine - Make a Scene     91
Twitter Web Client      32
TweetDeck               10
Name: source, dtype: int64
```

```
Out[56]: Text(0.5, 1.0, 'The Distribution of Source')
```

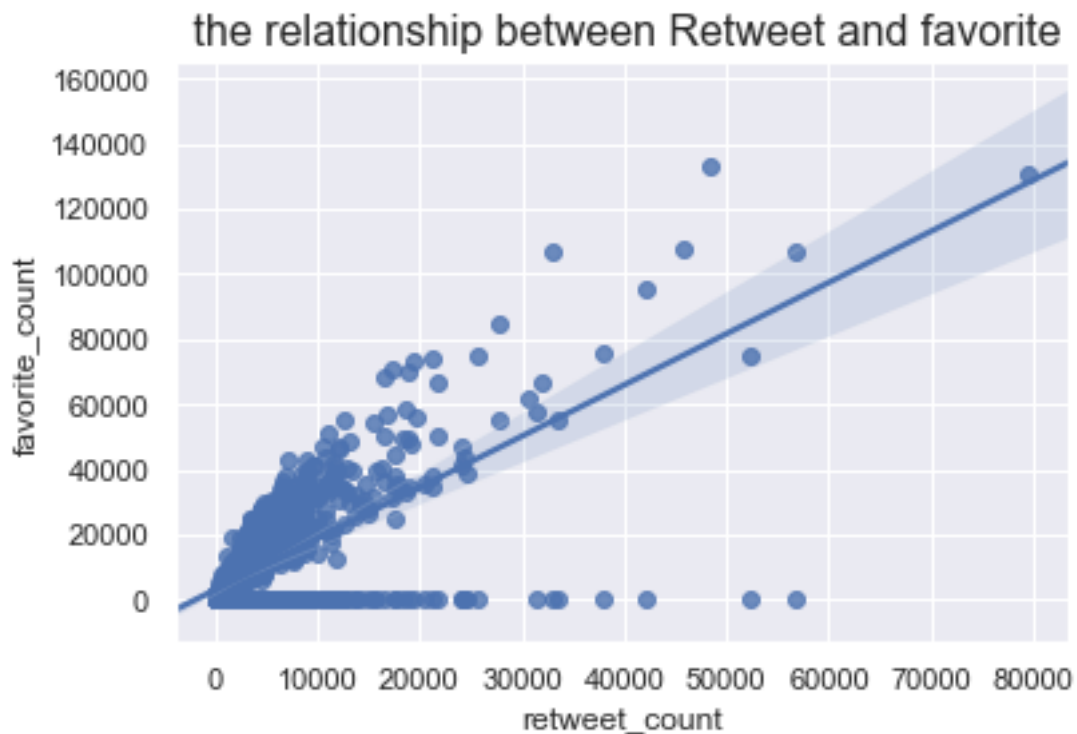



This plot above shows the distribution of source. We can see that the dominate source of tweets is from iPhone twitter app, which is 94% in the total. That means the twitter app is the main channel for people using to tweet, retweet, post, and others, while the TweetDesk is pretty rare (less than 1%).

2.4.1 Retweet_count and favorite_count

```
In [59]: sns.regplot(x="retweet_count", y="favorite_count", data=tweeter_df)
          plt.title(' the relationship between Retweet and favorite',fontsize=16)
```

```
Out[59]: Text(0.5, 1.0, ' the relationship between Retweet and favorite')
```



A reasonable hypothesis is that most popular tweets usually get a large number of retweets and favorite counts. I test the correlation between 'retweet_count' and 'favorite_count'. the plot showing a strong positive relationship between 'retweet_count' and 'favorite_count'. That confirms this hypothesis.

```
In [49]: tweeter_df.rating_numerator.describe()
```

```
Out[49]: count      2275.000000
         mean         10.712527
         std          2.124116
         min          1.000000
         25%         10.000000
         50%         11.000000
         75%         12.000000
         max          15.000000
         Name: rating_numerator, dtype: float64
```

```
In [65]: tweeter_df.first_prediction.value_counts()
```

```
Out[65]: golden_retriever      150
         Labrador_retriever    100
         Pembroke              89
         Chihuahua             83
         pug                   57
         chow                   44
```

Samoyed	43
toy_poodle	39
Pomeranian	38
cocker_spaniel	30
malamute	30
French_bulldog	26
miniature_pinscher	23
Chesapeake_Bay_retriever	23
seat_belt	22
Staffordshire_bullterrier	20
German_shepherd	20
Siberian_husky	20
web_site	19
Cardigan	19
Shetland_sheepdog	18
teddy	18
Maltese_dog	18
Eskimo_dog	18
beagle	18
Lakeland_terrier	17
Shih-Tzu	17
Rottweiler	17
kuvasz	16
Italian_greyhound	16
...	
clumber	1
ibex	1
otter	1
cowboy_boot	1
dining_table	1
bee_eater	1
restaurant	1
maillot	1
rapeseed	1
bannister	1
flamingo	1
minibus	1
barbell	1
polecat	1
hotdog	1
zebra	1
robin	1
carousel	1
Scotch_terrier	1
traffic_light	1
shopping_basket	1
bearskin	1
school_bus	1

```

African_grey          1
book_jacket           1
bakery                1
terrapin              1
wooden_spoon          1
standard_schnauzer    1
park_bench            1
Name: first_prediction, Length: 378, dtype: int64

```

```
In [63]: tweeter_df.retweet_count.describe()
```

```

Out[63]: count      2354.000000
         mean       3164.797366
         std        5284.770364
         min         0.000000
         25%        624.500000
         50%       1473.500000
         75%       3652.000000
         max       79515.000000
         Name: retweet_count, dtype: float64

```

```
In [64]: tweeter_df.favorite_count.describe()
```

```

Out[64]: count      2354.000000
         mean       8080.968564
         std       11814.771334
         min         0.000000
         25%       1415.000000
         50%       3603.500000
         75%      10122.250000
         max      132810.000000
         Name: favorite_count, dtype: float64

```

2.5 insight

The most popular dog type is a "pupper" The most popular dog, based on image predictions, is a Golden Retriever The median and mean ratings out of ten are 11 and 12.7 respectively The median and mean retweet count is 1337 and 2837 respectively The median and mean favorite count is 3744 and 8490 respectively

2.6 Reference

Twitter API Guide: <https://www.slickremix.com/docs/how-to-get-api-keys-and-tokens-for-twitter/>

Tweepy documentation: <https://media.readthedocs.org/pdf/tweepy/latest/tweepy.pdf>

WeRateDogs Twitter: https://twitter.com/dog_rates?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwg

```
In [ ]:
```