

# wrangle\_act

February 5, 2018

## 1 Introduction

Real-world data rarely comes clean. This project on 'Data Wrangling' involves 3 main steps namely:- 1. Gathering 2. Assessing 3. Cleaning

Throughout this project several python libraries will be used to gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it.

### The Dataset

The dataset that i will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog\_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

### 1.1 1. Gathering

This step involves gathering data from 3 different sources as shown below :-

**1. Download the file "twitter\_archive\_enhanced.csv" manually from the given link and read it. This file is from the WeRateDogs Twitter archive.**

```
In [1]: import pandas as pd
```

```
twitter_archive = pd.read_csv("twitter-archive-enhanced.csv")
twitter_archive.head()
```

```
Out[1]:
```

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

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

```

3 2017-07-30 15:58:51 +0000
4 2017-07-29 16:00:24 +0000

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

**2. Programmatically download “tweet image predictions” using Requests library** The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file (image\_predictions.tsv) is hosted on Udacity’s servers and should be downloaded programmatically using the Requests library and the following URL: [https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad\\_image-predictions/image-predictions.tsv](https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv)

```

In [2]: import requests
        import os

```

```

folder_name = 'Data Wrangling Project Files'
if not os.path.exists(folder_name):
    os.makedirs(folder_name)

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

with open(os.path.join(folder_name, url.split('/')[-1]), mode = 'wb') as file:
    file.write(response.content)

In [3]: os.listdir(folder_name)

Out[3]: ['image-predictions.tsv']

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

```

**3. Use Tweepy library to query Twitter's API for additional data** — Code used to query the twitter API to gather tweet\_json.txt

```

Create API object to access the needed information
import tweepy
consumer_key = 'xxxxxxxxxxxxxxxxxxxxxxxxxxxx'
consumer_secret = 'xxxxxxxxxxxxxxxxxxxxxxxxxxxx'
access_token = 'xxxxxxxxxxxxxxxxxxxxxxxxxxxx'
access_secret = 'xxxxxxxxxxxxxxxxxxxxxxxxxxxx'
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)
api = tweepy.API(auth)
Query and write the collected data to a text file
import pandas as pd
twitter_archive = pd.read_csv("twitter-archive-enhanced.csv")
tweet_id = twitter_archive.tweet_id
file = open('tweet_json.txt', 'w')
for id in tweet_id:

try:
    tweet = api.get_status(id)
    a= tweet.text
    b= tweet.favorite_count
    c= tweet.retweet_count
    file.write(a + '\n')
    file.write( str(b) + '\n')
    file.write(str(c) + '\n')

except:
    file.write('NA'+ '\n')
    file.write('NA'+ '\n')
    file.write('NA'+ '\n')

```

```
file.close()
```

```
In [5]: # counting total number of lines in txt file
        with open('tweet_json.txt') as f:
            print sum(1 for _ in f)
```

7068

```
In [6]: df_list = []
        with open('tweet_json.txt') as file:
            # 7068/3 = 2356, so use a range from 1 to 2357 to get all rows from "tweet_json.txt"
            for x in range(1, 2357):
                text = file.readline()[:-1]
                favorite_count = file.readline()[:-1]
                retweet_count = file.readline()[:-1]
                df_list.append({'text': text,
                               'favorite_count': favorite_count,
                               'retweet_count': retweet_count})
                x= x+1
        import pandas as pd
        df = pd.DataFrame(df_list, columns=['text', 'favorite_count', 'retweet_count'])
```

```
In [7]: df
```

```
Out[7]:
```

	text	favorite_count \
0	This is Phineas. He's a mystical boy. Only eve...	39164
1	NA	NA
2	NA	NA
3	This is Darla. She commenced a snooze mid meal...	42534
4	NA	NA
5	NA	NA
6	NA	NA
7	NA	NA
8	NA	NA
9	NA	NA
10	NA	NA
11	This is Bruno. He is a service shark. Only get...	28037
12	Here's a puppo that seems to be on the fence a...	48329
13	This is Ted. He does his best. Sometimes that'...	27427
14	NA	NA
15	NA	NA
16	This is Jim. He found a fren. Taught him how t...	29345
17	NA	NA
18	This is Ralphus. He's powering up. Attempting ...	20115
19	NA	NA
20	NA	NA
21	NA	NA
22	I've yet to rate a Venezuelan Hover Wiener. Th...	46621

23	This is Canela. She attempted some fancy porch...	69808
24	You may not have known you needed to see this ...	33977
25	NA	NA
26	This is Maya. She's very shy. Rarely leaves he...	35517
27	NA	NA
28	This is Derek. He's late for a dog meeting. 13...	22631
29	NA	NA
...	...	...
2326	This is quite the dog. Gets really excited whe...	453
2327	This is a southern Vesuvius bumblegruff. Can d...	112
2328	Oh goodness. A super rare northeast Qdoba kang...	170
2329	Those are sunglasses and a jean jacket. 11/10 ...	193
2330	Unique dog here. Very small. Lives in containe...	790
2331	NA	NA
2332	Look at this jokester thinking seat belt laws ...	305
2333	This is an extremely rare horned Parthenon. No...	202
2334	This is a funny dog. Weird toes. Won't come do...	513
2335	This is an Albanian 3 1/2 legged Episcopalian...	151
2336	Can take selfies 11/10 <a href="https://t.co/ws2AMaNwPW">https://t.co/ws2AMaNwPW</a>	179
2337	Very concerned about fellow dog trapped in com...	105
2338	Not familiar with this breed. No tail (weird)...	14574
2339	Oh my. Here you are seeing an Adobe Setter giv...	80
2340	Can stand on stump for what seems like a while...	159
2341	This appears to be a Mongolian Presbyterian mi...	166
2342	Here we have a well-established sunblockerspan...	120
2343	Let's hope this flight isn't Malaysian (lol). ...	327
2344	Here we have a northern speckled Rhododendron...	153
2345	This is the happiest dog you will ever see. Ve...	486
2346	Here is the Rand Paul of retrievers folks! He'...	115
2347	My oh my. This is a rare blond Canadian terrie...	303
2348	Here is a Siberian heavily armored polar bear ...	446
2349	This is an odd dog. Hard on the outside but lo...	1240
2350	This is a truly beautiful English Wilson Staff...	135
2351	Here we have a 1949 1st generation vulpix. Enj...	109
2352	This is a purebred Piers Morgan. Loves to Netf...	305
2353	Here is a very happy pup. Big fan of well-main...	126
2354	This is a western brown Mitsubishi terrier. Up...	131
2355	Here we have a Japanese Irish Setter. Lost eye...	2528

	retweet_count
0	8719
1	NA
2	NA
3	8814
4	NA
5	NA
6	NA
7	NA

8	NA
9	NA
10	NA
11	5067
12	10264
13	4646
14	NA
15	NA
16	4607
17	NA
18	3673
19	NA
20	NA
21	NA
22	11913
23	18658
24	10612
25	NA
26	7940
27	NA
28	4561
29	NA
...	...
2326	335
2327	42
2328	90
2329	95
2330	583
2331	NA
2332	144
2333	95
2334	364
2335	69
2336	79
2337	36
2338	6760
2339	14
2340	72
2341	76
2342	46
2343	172
2344	65
2345	226
2346	58
2347	144
2348	259
2349	867
2350	59

2351	40
2352	143
2353	46
2354	47
2355	52

[2356 rows x 3 columns]

## 1.2 2. Assessing

This step involves assessing the 3 datasets collected above (from 3 different sources) both visually and programmatically and find out quality and tidiness issues.

### Visual Assessment

In [8]: twitter\_archive

```
Out[8]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	NaN	NaN	
1	892177421306343426	NaN	NaN	
2	891815181378084864	NaN	NaN	
3	891689557279858688	NaN	NaN	
4	891327558926688256	NaN	NaN	
5	891087950875897856	NaN	NaN	
6	890971913173991426	NaN	NaN	
7	890729181411237888	NaN	NaN	
8	890609185150312448	NaN	NaN	
9	890240255349198849	NaN	NaN	
10	890006608113172480	NaN	NaN	
11	889880896479866881	NaN	NaN	
12	889665388333682689	NaN	NaN	
13	889638837579907072	NaN	NaN	
14	889531135344209921	NaN	NaN	
15	889278841981685760	NaN	NaN	
16	888917238123831296	NaN	NaN	
17	888804989199671297	NaN	NaN	
18	888554962724278272	NaN	NaN	
19	888202515573088257	NaN	NaN	
20	888078434458587136	NaN	NaN	
21	887705289381826560	NaN	NaN	
22	887517139158093824	NaN	NaN	
23	887473957103951883	NaN	NaN	
24	887343217045368832	NaN	NaN	
25	887101392804085760	NaN	NaN	
26	886983233522544640	NaN	NaN	
27	886736880519319552	NaN	NaN	
28	886680336477933568	NaN	NaN	
29	886366144734445568	NaN	NaN	

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

	timestamp \
0	2017-08-01 16:23:56 +0000
1	2017-08-01 00:17:27 +0000
2	2017-07-31 00:18:03 +0000
3	2017-07-30 15:58:51 +0000
4	2017-07-29 16:00:24 +0000
5	2017-07-29 00:08:17 +0000
6	2017-07-28 16:27:12 +0000
7	2017-07-28 00:22:40 +0000
8	2017-07-27 16:25:51 +0000
9	2017-07-26 15:59:51 +0000
10	2017-07-26 00:31:25 +0000
11	2017-07-25 16:11:53 +0000
12	2017-07-25 01:55:32 +0000
13	2017-07-25 00:10:02 +0000
14	2017-07-24 17:02:04 +0000



15	2017-07-24	00:19:32	+0000
16	2017-07-23	00:22:39	+0000
17	2017-07-22	16:56:37	+0000
18	2017-07-22	00:23:06	+0000
19	2017-07-21	01:02:36	+0000
20	2017-07-20	16:49:33	+0000
21	2017-07-19	16:06:48	+0000
22	2017-07-19	03:39:09	+0000
23	2017-07-19	00:47:34	+0000
24	2017-07-18	16:08:03	+0000
25	2017-07-18	00:07:08	+0000
26	2017-07-17	16:17:36	+0000
27	2017-07-16	23:58:41	+0000
28	2017-07-16	20:14:00	+0000
29	2017-07-15	23:25:31	+0000
...			...
2326	2015-11-17	00:24:19	+0000
2327	2015-11-17	00:06:54	+0000
2328	2015-11-16	23:23:41	+0000
2329	2015-11-16	21:54:18	+0000
2330	2015-11-16	21:10:36	+0000
2331	2015-11-16	20:32:58	+0000
2332	2015-11-16	20:01:42	+0000
2333	2015-11-16	19:31:45	+0000
2334	2015-11-16	16:37:02	+0000
2335	2015-11-16	16:11:11	+0000
2336	2015-11-16	15:14:19	+0000
2337	2015-11-16	14:57:41	+0000
2338	2015-11-16	04:02:55	+0000
2339	2015-11-16	03:55:04	+0000
2340	2015-11-16	03:44:34	+0000
2341	2015-11-16	03:22:39	+0000
2342	2015-11-16	02:38:37	+0000
2343	2015-11-16	01:59:36	+0000
2344	2015-11-16	01:52:02	+0000
2345	2015-11-16	01:22:45	+0000
2346	2015-11-16	01:01:59	+0000
2347	2015-11-16	00:55:59	+0000
2348	2015-11-16	00:49:46	+0000
2349	2015-11-16	00:35:11	+0000
2350	2015-11-16	00:30:50	+0000
2351	2015-11-16	00:24:50	+0000
2352	2015-11-16	00:04:52	+0000
2353	2015-11-15	23:21:54	+0000
2354	2015-11-15	23:05:30	+0000
2355	2015-11-15	22:32:08	+0000

source \

[illegible]

2343 <a href="http://twitter.com/download/iphone" r...  
 2344 <a href="http://twitter.com/download/iphone" r...  
 2345 <a href="http://twitter.com/download/iphone" r...  
 2346 <a href="http://twitter.com/download/iphone" r...  
 2347 <a href="http://twitter.com/download/iphone" r...  
 2348 <a href="http://twitter.com/download/iphone" r...  
 2349 <a href="http://twitter.com/download/iphone" r...  
 2350 <a href="http://twitter.com/download/iphone" r...  
 2351 <a href="http://twitter.com/download/iphone" r...  
 2352 <a href="http://twitter.com/download/iphone" r...  
 2353 <a href="http://twitter.com/download/iphone" r...  
 2354 <a href="http://twitter.com/download/iphone" r...  
 2355 <a href="http://twitter.com/download/iphone" r...

	text	retweeted_status_id \
0	This is Phineas. He's a mystical boy. Only eve...	NaN
1	This is Tilly. She's just checking pup on you...	NaN
2	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN
5	Here we have a majestic great white breaching ...	NaN
6	Meet Jax. He enjoys ice cream so much he gets ...	NaN
7	When you watch your owner call another dog a g...	NaN
8	This is Zoey. She doesn't want to be one of th...	NaN
9	This is Cassie. She is a college pup. Studying...	NaN
10	This is Koda. He is a South Australian decksha...	NaN
11	This is Bruno. He is a service shark. Only get...	NaN
12	Here's a puppo that seems to be on the fence a...	NaN
13	This is Ted. He does his best. Sometimes that'...	NaN
14	This is Stuart. He's sporting his favorite fan...	NaN
15	This is Oliver. You're witnessing one of his m...	NaN
16	This is Jim. He found a fren. Taught him how t...	NaN
17	This is Zeke. He has a new stick. Very proud o...	NaN
18	This is Ralphus. He's powering up. Attempting ...	NaN
19	RT @dog_rates: This is Canela. She attempted s...	8.874740e+17
20	This is Gerald. He was just told he didn't get...	NaN
21	This is Jeffrey. He has a monopoly on the pool...	NaN
22	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN
23	This is Canela. She attempted some fancy porch...	NaN
24	You may not have known you needed to see this ...	NaN
25	This... is a Jubilant Antarctic House Bear. We...	NaN
26	This is Maya. She's very shy. Rarely leaves he...	NaN
27	This is Mingus. He's a wonderful father to his...	NaN
28	This is Derek. He's late for a dog meeting. 13...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
...	...	...
2326	This is quite the dog. Gets really excited whe...	NaN
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN

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

	retweeted_status_user_id	retweeted_status_timestamp \
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
5	NaN	NaN
6	NaN	NaN
7	NaN	NaN
8	NaN	NaN
9	NaN	NaN
10	NaN	NaN
11	NaN	NaN
12	NaN	NaN
13	NaN	NaN
14	NaN	NaN
15	NaN	NaN
16	NaN	NaN
17	NaN	NaN

18	NaN	NaN
19	4.196984e+09	2017-07-19 00:47:34 +0000
20	NaN	NaN
21	NaN	NaN
22	NaN	NaN
23	NaN	NaN
24	NaN	NaN
25	NaN	NaN
26	NaN	NaN
27	NaN	NaN
28	NaN	NaN
29	NaN	NaN
...	...	...
2326	NaN	NaN
2327	NaN	NaN
2328	NaN	NaN
2329	NaN	NaN
2330	NaN	NaN
2331	NaN	NaN
2332	NaN	NaN
2333	NaN	NaN
2334	NaN	NaN
2335	NaN	NaN
2336	NaN	NaN
2337	NaN	NaN
2338	NaN	NaN
2339	NaN	NaN
2340	NaN	NaN
2341	NaN	NaN
2342	NaN	NaN
2343	NaN	NaN
2344	NaN	NaN
2345	NaN	NaN
2346	NaN	NaN
2347	NaN	NaN
2348	NaN	NaN
2349	NaN	NaN
2350	NaN	NaN
2351	NaN	NaN
2352	NaN	NaN
2353	NaN	NaN
2354	NaN	NaN
2355	NaN	NaN

	expanded_urls	rating_numerator \
0	<a href="https://twitter.com/dog_rates/status/892420643...">https://twitter.com/dog_rates/status/892420643...</a>	13
1	<a href="https://twitter.com/dog_rates/status/892177421...">https://twitter.com/dog_rates/status/892177421...</a>	13
2	<a href="https://twitter.com/dog_rates/status/891815181...">https://twitter.com/dog_rates/status/891815181...</a>	12

3	<a href="https://twitter.com/dog_rates/status/891689557...">https://twitter.com/dog_rates/status/891689557...</a>	13
4	<a href="https://twitter.com/dog_rates/status/891327558...">https://twitter.com/dog_rates/status/891327558...</a>	12
5	<a href="https://twitter.com/dog_rates/status/891087950...">https://twitter.com/dog_rates/status/891087950...</a>	13
6	<a href="https://gofundme.com/ydvmve-surgery-for-jax,ht...">https://gofundme.com/ydvmve-surgery-for-jax,ht...</a>	13
7	<a href="https://twitter.com/dog_rates/status/890729181...">https://twitter.com/dog_rates/status/890729181...</a>	13
8	<a href="https://twitter.com/dog_rates/status/890609185...">https://twitter.com/dog_rates/status/890609185...</a>	13
9	<a href="https://twitter.com/dog_rates/status/890240255...">https://twitter.com/dog_rates/status/890240255...</a>	14
10	<a href="https://twitter.com/dog_rates/status/890006608...">https://twitter.com/dog_rates/status/890006608...</a>	13
11	<a href="https://twitter.com/dog_rates/status/889880896...">https://twitter.com/dog_rates/status/889880896...</a>	13
12	<a href="https://twitter.com/dog_rates/status/889665388...">https://twitter.com/dog_rates/status/889665388...</a>	13
13	<a href="https://twitter.com/dog_rates/status/889638837...">https://twitter.com/dog_rates/status/889638837...</a>	12
14	<a href="https://twitter.com/dog_rates/status/889531135...">https://twitter.com/dog_rates/status/889531135...</a>	13
15	<a href="https://twitter.com/dog_rates/status/889278841...">https://twitter.com/dog_rates/status/889278841...</a>	13
16	<a href="https://twitter.com/dog_rates/status/888917238...">https://twitter.com/dog_rates/status/888917238...</a>	12
17	<a href="https://twitter.com/dog_rates/status/888804989...">https://twitter.com/dog_rates/status/888804989...</a>	13
18	<a href="https://twitter.com/dog_rates/status/888554962...">https://twitter.com/dog_rates/status/888554962...</a>	13
19	<a href="https://twitter.com/dog_rates/status/887473957...">https://twitter.com/dog_rates/status/887473957...</a>	13
20	<a href="https://twitter.com/dog_rates/status/888078434...">https://twitter.com/dog_rates/status/888078434...</a>	12
21	<a href="https://twitter.com/dog_rates/status/887705289...">https://twitter.com/dog_rates/status/887705289...</a>	13
22	<a href="https://twitter.com/dog_rates/status/887517139...">https://twitter.com/dog_rates/status/887517139...</a>	14
23	<a href="https://twitter.com/dog_rates/status/887473957...">https://twitter.com/dog_rates/status/887473957...</a>	13
24	<a href="https://twitter.com/dog_rates/status/887343217...">https://twitter.com/dog_rates/status/887343217...</a>	13
25	<a href="https://twitter.com/dog_rates/status/887101392...">https://twitter.com/dog_rates/status/887101392...</a>	12
26	<a href="https://twitter.com/dog_rates/status/886983233...">https://twitter.com/dog_rates/status/886983233...</a>	13
27	<a href="https://www.gofundme.com/mingusneedsus,https:/...">https://www.gofundme.com/mingusneedsus,https:/...</a>	13
28	<a href="https://twitter.com/dog_rates/status/886680336...">https://twitter.com/dog_rates/status/886680336...</a>	13
29	<a href="https://twitter.com/dog_rates/status/886366144...">https://twitter.com/dog_rates/status/886366144...</a>	12
...	...	...
2326	<a href="https://twitter.com/dog_rates/status/666411507...">https://twitter.com/dog_rates/status/666411507...</a>	2
2327	<a href="https://twitter.com/dog_rates/status/666407126...">https://twitter.com/dog_rates/status/666407126...</a>	7
2328	<a href="https://twitter.com/dog_rates/status/666396247...">https://twitter.com/dog_rates/status/666396247...</a>	9
2329	<a href="https://twitter.com/dog_rates/status/666373753...">https://twitter.com/dog_rates/status/666373753...</a>	11
2330	<a href="https://twitter.com/dog_rates/status/666362758...">https://twitter.com/dog_rates/status/666362758...</a>	6
2331	<a href="https://twitter.com/dog_rates/status/666353288...">https://twitter.com/dog_rates/status/666353288...</a>	8
2332	<a href="https://twitter.com/dog_rates/status/666345417...">https://twitter.com/dog_rates/status/666345417...</a>	10
2333	<a href="https://twitter.com/dog_rates/status/666337882...">https://twitter.com/dog_rates/status/666337882...</a>	9
2334	<a href="https://twitter.com/dog_rates/status/666293911...">https://twitter.com/dog_rates/status/666293911...</a>	3
2335	<a href="https://twitter.com/dog_rates/status/666287406...">https://twitter.com/dog_rates/status/666287406...</a>	1
2336	<a href="https://twitter.com/dog_rates/status/666273097...">https://twitter.com/dog_rates/status/666273097...</a>	11
2337	<a href="https://twitter.com/dog_rates/status/666268910...">https://twitter.com/dog_rates/status/666268910...</a>	10
2338	<a href="https://twitter.com/dog_rates/status/666104133...">https://twitter.com/dog_rates/status/666104133...</a>	1
2339	<a href="https://twitter.com/dog_rates/status/666102155...">https://twitter.com/dog_rates/status/666102155...</a>	11
2340	<a href="https://twitter.com/dog_rates/status/666099513...">https://twitter.com/dog_rates/status/666099513...</a>	8
2341	<a href="https://twitter.com/dog_rates/status/666094000...">https://twitter.com/dog_rates/status/666094000...</a>	9
2342	<a href="https://twitter.com/dog_rates/status/666082916...">https://twitter.com/dog_rates/status/666082916...</a>	6
2343	<a href="https://twitter.com/dog_rates/status/666073100...">https://twitter.com/dog_rates/status/666073100...</a>	10
2344	<a href="https://twitter.com/dog_rates/status/666071193...">https://twitter.com/dog_rates/status/666071193...</a>	9
2345	<a href="https://twitter.com/dog_rates/status/666063827...">https://twitter.com/dog_rates/status/666063827...</a>	10

2346	<a href="https://twitter.com/dog_rates/status/666058600...">https://twitter.com/dog_rates/status/666058600...</a>	8
2347	<a href="https://twitter.com/dog_rates/status/666057090...">https://twitter.com/dog_rates/status/666057090...</a>	9
2348	<a href="https://twitter.com/dog_rates/status/666055525...">https://twitter.com/dog_rates/status/666055525...</a>	10
2349	<a href="https://twitter.com/dog_rates/status/666051853...">https://twitter.com/dog_rates/status/666051853...</a>	2
2350	<a href="https://twitter.com/dog_rates/status/666050758...">https://twitter.com/dog_rates/status/666050758...</a>	10
2351	<a href="https://twitter.com/dog_rates/status/666049248...">https://twitter.com/dog_rates/status/666049248...</a>	5
2352	<a href="https://twitter.com/dog_rates/status/666044226...">https://twitter.com/dog_rates/status/666044226...</a>	6
2353	<a href="https://twitter.com/dog_rates/status/666033412...">https://twitter.com/dog_rates/status/666033412...</a>	9
2354	<a href="https://twitter.com/dog_rates/status/666029285...">https://twitter.com/dog_rates/status/666029285...</a>	7
2355	<a href="https://twitter.com/dog_rates/status/666020888...">https://twitter.com/dog_rates/status/666020888...</a>	8

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10	Phineas	None	None	None	None
1	10	Tilly	None	None	None	None
2	10	Archie	None	None	None	None
3	10	Darla	None	None	None	None
4	10	Franklin	None	None	None	None
5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None
8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
19	10	Canela	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
...	...	...	...	...	...	...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None

2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None
2343	10	None	None	None	None	None
2344	10	None	None	None	None	None
2345	10	the	None	None	None	None
2346	10	the	None	None	None	None
2347	10	a	None	None	None	None
2348	10	a	None	None	None	None
2349	10	an	None	None	None	None
2350	10	a	None	None	None	None
2351	10	None	None	None	None	None
2352	10	a	None	None	None	None
2353	10	a	None	None	None	None
2354	10	a	None	None	None	None
2355	10	None	None	None	None	None

[2356 rows x 17 columns]

In [9]: image\_pred

Out[9]:	tweet_id	jpg_url \
0	666020888022790149	<a href="https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg">https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg</a>
1	666029285002620928	<a href="https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg">https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg</a>
2	666033412701032449	<a href="https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg">https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg</a>
3	666044226329800704	<a href="https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg">https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg</a>
4	666049248165822465	<a href="https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg">https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg</a>
5	666050758794694657	<a href="https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg">https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg</a>
6	666051853826850816	<a href="https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg">https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg</a>
7	666055525042405380	<a href="https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg">https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg</a>
8	666057090499244032	<a href="https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg">https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg</a>
9	666058600524156928	<a href="https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg">https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg</a>
10	666063827256086533	<a href="https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg">https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg</a>
11	666071193221509120	<a href="https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg">https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg</a>
12	666073100786774016	<a href="https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg">https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg</a>
13	666082916733198337	<a href="https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg">https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg</a>
14	666094000022159362	<a href="https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg">https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg</a>
15	666099513787052032	<a href="https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg">https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg</a>
16	666102155909144576	<a href="https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg">https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg</a>



17	666104133288665088	<a href="https://pbs.twimg.com/media/CT56LSZW0AALJj2.jpg">https://pbs.twimg.com/media/CT56LSZW0AALJj2.jpg</a>
18	666268910803644416	<a href="https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg">https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg</a>
19	666273097616637952	<a href="https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg">https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg</a>
20	666287406224695296	<a href="https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg">https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg</a>
21	666293911632134144	<a href="https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg">https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg</a>
22	666337882303524864	<a href="https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg">https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg</a>
23	666345417576210432	<a href="https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg">https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg</a>
24	666353288456101888	<a href="https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg">https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg</a>
25	666362758909284353	<a href="https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg">https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg</a>
26	666373753744588802	<a href="https://pbs.twimg.com/media/CT9vZEYUWAA1Z05.jpg">https://pbs.twimg.com/media/CT9vZEYUWAA1Z05.jpg</a>
27	666396247373291520	<a href="https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg">https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg</a>
28	666407126856765440	<a href="https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg">https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg</a>
29	666411507551481857	<a href="https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg">https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg</a>
...	...	...
2045	886366144734445568	<a href="https://pbs.twimg.com/media/DE0BTnQUwAapKEH.jpg">https://pbs.twimg.com/media/DE0BTnQUwAapKEH.jpg</a>
2046	886680336477933568	<a href="https://pbs.twimg.com/media/DE4fEDzWAAAYHMM.jpg">https://pbs.twimg.com/media/DE4fEDzWAAAYHMM.jpg</a>
2047	886736880519319552	<a href="https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg">https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg</a>
2048	886983233522544640	<a href="https://pbs.twimg.com/media/DE8yicJW0AAAavBJ.jpg">https://pbs.twimg.com/media/DE8yicJW0AAAavBJ.jpg</a>
2049	887101392804085760	<a href="https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg">https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg</a>
2050	887343217045368832	<a href="https://pbs.twimg.com/ext_tw_video_thumb/88734...">https://pbs.twimg.com/ext_tw_video_thumb/88734...</a>
2051	887473957103951883	<a href="https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg">https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg</a>
2052	887517139158093824	<a href="https://pbs.twimg.com/ext_tw_video_thumb/88751...">https://pbs.twimg.com/ext_tw_video_thumb/88751...</a>
2053	887705289381826560	<a href="https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg">https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg</a>
2054	888078434458587136	<a href="https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg">https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg</a>
2055	888202515573088257	<a href="https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg">https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg</a>
2056	888554962724278272	<a href="https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg">https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg</a>
2057	888804989199671297	<a href="https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg">https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg</a>
2058	888917238123831296	<a href="https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg">https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg</a>
2059	889278841981685760	<a href="https://pbs.twimg.com/ext_tw_video_thumb/88927...">https://pbs.twimg.com/ext_tw_video_thumb/88927...</a>
2060	889531135344209921	<a href="https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg">https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg</a>
2061	889638837579907072	<a href="https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg">https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg</a>
2062	889665388333682689	<a href="https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg">https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg</a>
2063	889880896479866881	<a href="https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg">https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg</a>
2064	890006608113172480	<a href="https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg">https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg</a>
2065	890240255349198849	<a href="https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg">https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg</a>
2066	890609185150312448	<a href="https://pbs.twimg.com/media/DFwUU_XcAEpyXI.jpg">https://pbs.twimg.com/media/DFwUU_XcAEpyXI.jpg</a>
2067	890729181411237888	<a href="https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg">https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg</a>
2068	890971913173991426	<a href="https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg">https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg</a>
2069	891087950875897856	<a href="https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg">https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg</a>
2070	891327558926688256	<a href="https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg">https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg</a>
2071	891689557279858688	<a href="https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg">https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg</a>
2072	891815181378084864	<a href="https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg">https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg</a>
2073	892177421306343426	<a href="https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg">https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg</a>
2074	892420643555336193	<a href="https://pbs.twimg.com/media/DGKD1-bX0AAIAUK.jpg">https://pbs.twimg.com/media/DGKD1-bX0AAIAUK.jpg</a>

	img_num		p1	p1_conf	p1_dog	\
0	1	Welsh_springer_spaniel	0.465074	True		
1	1	redbone	0.506826	True		

2	1	German_shepherd	0.596461	True
3	1	Rhodesian_ridgeback	0.408143	True
4	1	miniature_pinscher	0.560311	True
5	1	Bernese_mountain_dog	0.651137	True
6	1	box_turtle	0.933012	False
7	1	chow	0.692517	True
8	1	shopping_cart	0.962465	False
9	1	miniature_poodle	0.201493	True
10	1	golden_retriever	0.775930	True
11	1	Gordon_setter	0.503672	True
12	1	Walker_hound	0.260857	True
13	1	pug	0.489814	True
14	1	bloodhound	0.195217	True
15	1	Lhasa	0.582330	True
16	1	English_setter	0.298617	True
17	1	hen	0.965932	False
18	1	desktop_computer	0.086502	False
19	1	Italian_greyhound	0.176053	True
20	1	Maltese_dog	0.857531	True
21	1	three-toed_sloth	0.914671	False
22	1	ox	0.416669	False
23	1	golden_retriever	0.858744	True
24	1	malamute	0.336874	True
25	1	guinea_pig	0.996496	False
26	1	soft-coated_wheaten_terrier	0.326467	True
27	1	Chihuahua	0.978108	True
28	1	black-and-tan_coonhound	0.529139	True
29	1	coho	0.404640	False
...	...	...	...	...
2045	1	French_bulldog	0.999201	True
2046	1	convertible	0.738995	False
2047	1	kuvasz	0.309706	True
2048	2	Chihuahua	0.793469	True
2049	1	Samoyed	0.733942	True
2050	1	Mexican_hairless	0.330741	True
2051	2	Pembroke	0.809197	True
2052	1	limousine	0.130432	False
2053	1	basset	0.821664	True
2054	1	French_bulldog	0.995026	True
2055	2	Pembroke	0.809197	True
2056	3	Siberian_husky	0.700377	True
2057	1	golden_retriever	0.469760	True
2058	1	golden_retriever	0.714719	True
2059	1	whippet	0.626152	True
2060	1	golden_retriever	0.953442	True
2061	1	French_bulldog	0.991650	True
2062	1	Pembroke	0.966327	True
2063	1	French_bulldog	0.377417	True

2064	1	Samoyed	0.957979	True
2065	1	Pembroke	0.511319	True
2066	1	Irish_terrier	0.487574	True
2067	2	Pomeranian	0.566142	True
2068	1	Appenzeller	0.341703	True
2069	1	Chesapeake_Bay_retriever	0.425595	True
2070	2	basset	0.555712	True
2071	1	paper_towel	0.170278	False
2072	1	Chihuahua	0.716012	True
2073	1	Chihuahua	0.323581	True
2074	1	orange	0.097049	False

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

2049	Eskimo_dog	0.035029	True	Staffordshire_bullterrier
2050	sea_lion	0.275645	False	Weimaraner
2051	Rhodesian_ridgeback	0.054950	True	beagle
2052	tow_truck	0.029175	False	shopping_cart
2053	redbone	0.087582	True	Weimaraner
2054	pug	0.000932	True	bull_mastiff
2055	Rhodesian_ridgeback	0.054950	True	beagle
2056	Eskimo_dog	0.166511	True	malamute
2057	Labrador_retriever	0.184172	True	English_setter
2058	Tibetan_mastiff	0.120184	True	Labrador_retriever
2059	borzoi	0.194742	True	Saluki
2060	Labrador_retriever	0.013834	True	redbone
2061	boxer	0.002129	True	Staffordshire_bullterrier
2062	Cardigan	0.027356	True	basenji
2063	Labrador_retriever	0.151317	True	muzzle
2064	Pomeranian	0.013884	True	chow
2065	Cardigan	0.451038	True	Chihuahua
2066	Irish_setter	0.193054	True	Chesapeake_Bay_retriever
2067	Eskimo_dog	0.178406	True	Pembroke
2068	Border_collie	0.199287	True	ice_lolly
2069	Irish_terrier	0.116317	True	Indian_elephant
2070	English_springer	0.225770	True	German_short-haired_pointer
2071	Labrador_retriever	0.168086	True	spatula
2072	malamute	0.078253	True	kelpie
2073	Pekinese	0.090647	True	papillon
2074	bagel	0.085851	False	banana

	p3_conf	p3_dog
0	0.061428	True
1	0.072010	True
2	0.116197	True
3	0.222752	True
4	0.154629	True
5	0.016199	True
6	0.017885	False
7	0.054449	False
8	0.007959	True
9	0.082086	True
10	0.072427	True
11	0.109454	True
12	0.097471	True
13	0.048960	True
14	0.075628	True
15	0.089688	True
16	0.133649	True
17	0.000052	False
18	0.079480	False
19	0.111152	True

20	0.025581	True
21	0.013207	False
22	0.102643	True
23	0.014241	True
24	0.093412	True
25	0.000461	False
26	0.206803	True
27	0.004577	True
28	0.173810	True
29	0.189945	False
...	...	...
2045	0.000076	True
2046	0.044173	False
2047	0.086346	True
2048	0.032253	False
2049	0.029705	True
2050	0.134203	True
2051	0.038915	True
2052	0.026321	False
2053	0.026236	True
2054	0.000903	True
2055	0.038915	True
2056	0.111411	True
2057	0.073482	True
2058	0.105506	True
2059	0.027351	True
2060	0.007958	True
2061	0.001498	True
2062	0.004633	True
2063	0.082981	False
2064	0.008167	True
2065	0.029248	True
2066	0.118184	True
2067	0.076507	True
2068	0.193548	False
2069	0.076902	False
2070	0.175219	True
2071	0.040836	False
2072	0.031379	True
2073	0.068957	True
2074	0.076110	False

[2075 rows x 12 columns]

```
In [10]: tweet_json = df
         tweet_json
```

```
Out[10]:
```

	text	favorite_count	\
0	This is Phineas. He's a mystical boy. Only eve...	39164	

1		NA	NA
2		NA	NA
3	This is Darla. She commenced a snooze mid meal...		42534
4		NA	NA
5		NA	NA
6		NA	NA
7		NA	NA
8		NA	NA
9		NA	NA
10		NA	NA
11	This is Bruno. He is a service shark. Only get...		28037
12	Here's a puppo that seems to be on the fence a...		48329
13	This is Ted. He does his best. Sometimes that'...		27427
14		NA	NA
15		NA	NA
16	This is Jim. He found a fren. Taught him how t...		29345
17		NA	NA
18	This is Ralphus. He's powering up. Attempting ...		20115
19		NA	NA
20		NA	NA
21		NA	NA
22	I've yet to rate a Venezuelan Hover Wiener. Th...		46621
23	This is Canela. She attempted some fancy porch...		69808
24	You may not have known you needed to see this ...		33977
25		NA	NA
26	This is Maya. She's very shy. Rarely leaves he...		35517
27		NA	NA
28	This is Derek. He's late for a dog meeting. 13...		22631
29		NA	NA
...		...	...
2326	This is quite the dog. Gets really excited whe...		453
2327	This is a southern Vesuvius bumblegruff. Can d...		112
2328	Oh goodness. A super rare northeast Qdoba kang...		170
2329	Those are sunglasses and a jean jacket. 11/10 ...		193
2330	Unique dog here. Very small. Lives in containe...		790
2331		NA	NA
2332	Look at this jokester thinking seat belt laws ...		305
2333	This is an extremely rare horned Parthenon. No...		202
2334	This is a funny dog. Weird toes. Won't come do...		513
2335	This is an Albanian 3 1/2 legged Episcopalian...		151
2336	Can take selfies 11/10 <a href="https://t.co/ws2AMaWpPW">https://t.co/ws2AMaWpPW</a>		179
2337	Very concerned about fellow dog trapped in com...		105
2338	Not familiar with this breed. No tail (weird)...		14574
2339	Oh my. Here you are seeing an Adobe Setter giv...		80
2340	Can stand on stump for what seems like a while...		159
2341	This appears to be a Mongolian Presbyterian mi...		166
2342	Here we have a well-established sunblockerspan...		120
2343	Let's hope this flight isn't Malaysian (lol). ...		327

2344	Here we have a northern speckled Rhododendron...	153
2345	This is the happiest dog you will ever see. Ve...	486
2346	Here is the Rand Paul of retrievers folks! He'...	115
2347	My oh my. This is a rare blond Canadian terrie...	303
2348	Here is a Siberian heavily armored polar bear ...	446
2349	This is an odd dog. Hard on the outside but lo...	1240
2350	This is a truly beautiful English Wilson Staff...	135
2351	Here we have a 1949 1st generation vulpix. Enj...	109
2352	This is a purebred Piers Morgan. Loves to Netf...	305
2353	Here is a very happy pup. Big fan of well-main...	126
2354	This is a western brown Mitsubishi terrier. Up...	131
2355	Here we have a Japanese Irish Setter. Lost eye...	2528

	retweet_count
0	8719
1	NA
2	NA
3	8814
4	NA
5	NA
6	NA
7	NA
8	NA
9	NA
10	NA
11	5067
12	10264
13	4646
14	NA
15	NA
16	4607
17	NA
18	3673
19	NA
20	NA
21	NA
22	11913
23	18658
24	10612
25	NA
26	7940
27	NA
28	4561
29	NA
...	...
2326	335
2327	42
2328	90

2329	95
2330	583
2331	NA
2332	144
2333	95
2334	364
2335	69
2336	79
2337	36
2338	6760
2339	14
2340	72
2341	76
2342	46
2343	172
2344	65
2345	226
2346	58
2347	144
2348	259
2349	867
2350	59
2351	40
2352	143
2353	46
2354	47
2355	52

[2356 rows x 3 columns]

## Programmatic Assessment

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

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



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

```
In [12]: image_pred.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: 152.0+ KB
```

```
In [13]: tweet_json.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 3 columns):
text          2356 non-null object
favorite_count 2356 non-null object
retweet_count 2356 non-null object
dtypes: object(3)
memory usage: 55.3+ KB
```

```
In [14]: twitter_archive.doggo.value_counts()
```

```
Out[14]: None      2259
         doggo      97
         Name: doggo, dtype: int64
```

```
In [15]: twitter_archive.floofer.value_counts()
```

```
Out[15]: None          2346
         floofer       10
         Name: floofer, dtype: int64
```

```
In [16]: twitter_archive.pupper.value_counts()
```

```
Out[16]: None          2099
         pupper        257
         Name: pupper, dtype: int64
```

```
In [17]: twitter_archive.puppo.value_counts()
```

```
Out[17]: None          2326
         puppo         30
         Name: puppo, dtype: int64
```

```
In [18]: image_pred.p1.unique()
```

```
Out[18]: array(['Welsh_springer_spaniel', 'redbone', 'German_shepherd',
                'Rhodesian_ridgeback', 'miniature_pinscher', 'Bernese_mountain_dog',
                'box_turtle', 'chow', 'shopping_cart', 'miniature_poodle',
                'golden_retriever', 'Gordon_setter', 'Walker_hound', 'pug',
                'bloodhound', 'Lhasa', 'English_setter', 'hen', 'desktop_computer',
                'Italian_greyhound', 'Maltese_dog', 'three-toed_sloth', 'ox',
                'malamute', 'guinea_pig', 'soft-coated_wheaten_terrier',
                'Chihuahua', 'black-and-tan_coonhound', 'coho', 'toy_terrier',
                'Blenheim_spaniel', 'Pembroke', 'llama', 'Chesapeake_Bay_retriever',
                'curly-coated_retriever', 'dalmatian', 'Ibizan_hound',
                'Border_collie', 'Labrador_retriever', 'seat_belt', 'snail',
                'miniature_schnauzer', 'Airedale', 'triceratops', 'swab', 'hay',
                'hyena', 'jigsaw_puzzle', 'West_Highland_white_terrier',
                'toy_poodle', 'giant_schnauzer', 'vizsla', 'vacuum', 'Rottweiler',
                'Siberian_husky', 'teddy', 'papillon', 'Saint_Bernard', 'porcupine',
                'goose', 'Tibetan_terrier', 'borzoi', 'beagle', 'hare',
                'Yorkshire_terrier', 'Pomeranian', 'electric_fan', 'web_site',
                'ibex', 'kuvasz', 'fire_engine', 'lorikeet',
                'flat-coated_retriever', 'toyshop', 'common_iguana',
                'Norwegian_elkhound', 'frilled_lizard', 'leatherback_turtle',
                'hamster', 'Angora', 'Arctic_fox', 'trombone', 'canoe',
                'king_penguin', 'shopping_basket', 'standard_poodle',
                'Staffordshire_bullterrier', 'basenji', 'Lakeland_terrier',
                'American_Staffordshire_terrier', 'bearskin', 'Shih-Tzu', 'bustard',
                'crash_helmet', 'French_bulldog', 'Pekinese', 'komondor',
                'ski_mask', 'malinois', 'kelpie', 'Brittany_spaniel',
                'cocker_spaniel', 'shower_curtain', 'basset', 'jellyfish',
                'doormat', 'Arabian_camel', 'lynx', 'hog', 'comic_book', 'minivan',
                'seashore', 'cuirass', 'Brabancon_griffon', 'candle', 'Eskimo_dog',
                'weasel', 'Christmas_stocking', 'washbasin', 'car_mirror',
                'piggy_bank', 'pot', 'boathouse', 'mud_turtle',
```

'German\_short-haired\_pointer', 'Shetland\_sheepdog', 'Irish\_terrier',  
'cairn', 'platypus', 'English\_springer', 'whippet',  
'ping-pong\_ball', 'sea\_urchin', 'bow\_tie', 'window\_shade',  
'jack-o'-lantern', 'sorrel', 'Sussex\_spaniel', 'peacock', 'axolotl',  
'wool', 'banana', 'Dandie\_Dinmont', 'Norwich\_terrier',  
'wood\_rabbit', 'dhole', 'keeshond', 'Norfolk\_terrier', 'lacewing',  
'dingo', 'brown\_bear', 'Old\_English\_sheepdog', 'scorpion',  
'flamingo', 'microphone', 'Samoyed', 'pitcher',  
'African\_hunting\_dog', 'refrigerator', 'picket\_fence', 'tub',  
'zebra', 'hermit\_crab', 'swing', 'Doberman', 'park\_bench',  
'feather\_boa', 'Loafer', 'stone\_wall', 'ice\_bear', 'prayer\_rug',  
'chimpanzee', 'china\_cabinet', 'bee\_eater', 'tennis\_ball', 'carton',  
'killer\_whale', 'ostrich', 'terrapin', 'Siamese\_cat', 'gondola',  
'Great\_Pyrenees', 'microwave', 'starfish', 'sandbar', 'tusk',  
'motor\_scooter', 'ram', 'leaf\_beetle', 'wombat', 'schipperke',  
'Newfoundland', 'bull\_mastiff', 'water\_bottle', 'suit',  
'toilet\_seat', 'collie', 'robin', 'Cardigan',  
'Greater\_Swiss\_Mountain\_dog', 'slug', 'toilet\_tissue',  
'acorn\_squash', 'soccer\_ball', 'African\_crocodile', 'tick',  
'ocarina', 'boxer', 'street\_sign', 'bow', 'stove', 'paper\_towel',  
'upright', 'dough', 'Scottish\_deerhound', 'bath\_towel',  
'standard\_schnauzer', 'walking\_stick', 'Irish\_water\_spaniel',  
'bubble', 'Boston\_bull', 'book\_jacket', 'rain\_barrel',  
'black-footed\_ferret', 'guenon', 'Japanese\_spaniel',  
'water\_buffalo', 'patio', 'cowboy\_hat', 'dogsled', 'maze', 'harp',  
'panpipe', 'cash\_machine', 'mailbox', 'wallaby', 'EntleBucher',  
'earthstar', 'pillow', 'bluetick', 'space\_heater', 'carousel',  
'Irish\_setter', 'birdhouse', 'snorkel', 'bald\_eagle', 'koala',  
'Leonberg', 'cheetah', 'minibus', 'Weimaraner', 'clog',  
'dishwasher', 'white\_wolf', 'sliding\_door', 'damsel\_fly',  
'Great\_Dane', 'Tibetan\_mastiff', 'cheeseburger', 'fiddler\_crab',  
'bannister', 'crane', 'Scotch\_terrier', 'snowmobile', 'badger',  
'bighorn', 'geyser', 'barrow', 'bison', 'Mexican\_hairless',  
'ice\_lolly', 'sea\_lion', 'dining\_table', 'groenendael',  
'Australian\_terrier', 'beaver', 'briard', 'Appenzeller', 'grey\_fox',  
'mousetrap', 'hippopotamus', 'Border\_terrier', 'hummingbird',  
'tailed\_frog', 'otter', 'Egyptian\_cat', 'four-poster', 'wild\_boar',  
'bathtub', 'agama', 'muzzle', 'hotdog', 'bib', 'espresso',  
'timber\_wolf', 'meerkat', 'nail', 'hammer', 'home\_theater', 'alp',  
'bonnet', 'handkerchief', 'hand\_blower', 'polecat', 'lakeside',  
'studio\_couch', 'cup', 'cliff', 'Bedlington\_terrier', 'lawn\_mower',  
'balloon', 'sunglasses', 'rapeseed', 'traffic\_light', 'coil',  
'binoculars', 'paddle', 'tiger\_shark', 'sulphur-crested\_cockatoo',  
'wire-haired\_fox\_terrier', 'Saluki', 'American\_black\_bear',  
'rotisserie', 'conch', 'skunk', 'bookshop', 'radio\_telescope',  
'cougar', 'African\_grey', 'coral\_reef', 'lion', 'maillot',  
'Madagascar\_cat', 'tabby', 'silky\_terrier', 'giant\_panda',  
'long-horned\_beetle', 'Afghan\_hound', 'clumber', 'sundial',

```
'padlock', 'pool_table', 'quilt', 'beach_wagon', 'remote_control',
'bakery', 'pedestal', 'gas_pump', 'bookcase', 'shield', 'loupe',
'restaurant', 'prison', 'school_bus', 'cowboy_boot', 'jersey',
'wooden_spoon', 'leopard', 'mortarboard', 'teapot',
'military_uniform', 'washer', 'coffee_mug', 'fountain',
'pencil_box', 'barbell', 'grille', 'revolver', 'envelope',
'syringe', 'marmot', 'pole', 'laptop', 'basketball', 'tricycle',
'convertible', 'limousine', 'orange'], dtype=object)
```

```
In [19]: image_pred.p2.unique()
```

```
Out[19]: array(['collie', 'miniature_pinscher', 'malinois', 'redbone', 'Rottweiler',
'English_springer', 'mud_turtle', 'Tibetan_mastiff',
'shopping_basket', 'komondor', 'Yorkshire_terrier',
'English_foxhound', 'bull_mastiff', 'German_shepherd', 'Shih-Tzu',
'Newfoundland', 'cock', 'desk', 'toy_terrier', 'toy_poodle',
'otter', 'Chesapeake_Bay_retriever', 'Siberian_husky', 'skunk',
'Afghan_hound', 'bloodhound', 'barracouta', 'papillon',
'cocker_spaniel', 'chow', 'Irish_terrier', 'chain_saw', 'beagle',
'giant_schnauzer', 'Labrador_retriever', 'Pembroke', 'Chihuahua',
'Weimaraner', 'slug', 'Brittany_spaniel', 'standard_schnauzer',
'teddy', 'armadillo', 'African_hunting_dog', 'vizsla', 'doormat',
'pug', 'Italian_greyhound', 'Samoyed', 'Pomeranian',
'miniature_poodle', 'Lakeland_terrier', 'Irish_setter', 'swab',
'malamute', 'bath_towel', 'Border_collie', 'Leonberg', 'drake',
'French_bulldog', 'ice_bear', 'Christmas_stocking',
'golden_retriever', 'standard_poodle', 'dhole', 'kuvasz',
'Cardigan', 'silky_terrier', 'spotlight', 'dishwasher', 'bighorn',
'tow_truck', 'hummingbird', 'English_setter', 'prayer_rug',
'frilled_lizard', 'Pekinese', 'ox', 'boxer', 'hog', 'guinea_pig',
'hen', 'wallaby', 'cowboy_boot', 'cornet', 'minivan', 'paddle',
'basset', 'hamper', 'Bedlington_terrier', 'Shetland_sheepdog',
'bow', 'Lhasa', 'pelican', 'toaster', 'groenendael',
'Australian_terrier', 'llama', 'knee_pad', 'pillow', 'Ibizan_hound',
'Old_English_sheepdog', 'Welsh_springer_spaniel', 'coral_reef',
'bison', 'waffle_iron', 'tabby', 'bib', 'police_van', 'Eskimo_dog',
'breastplate', 'German_short-haired_pointer', 'Norfolk_terrier',
'Blenheim_spaniel', 'pickup', 'miniature_schnauzer', 'lampshade',
'Tibetan_terrier', 'Siamese_cat', 'borzoi', 'studio_couch',
'toilet_seat', 'hamster', 'seat_belt', 'keeshond', 'koala',
'hair_spray', 'Saint_Bernard', 'tray',
'American_Staffordshire_terrier', 'birdhouse', 'terrapien',
'Staffordshire_bullterrier', 'briard',
'West_Highland_white_terrier', 'spotted_salamander', 'tennis_ball',
'porcupine', 'cardigan', 'corn', 'basenji', 'otterhound',
'European_gallinule', 'indri', 'tailed_frog', 'beach_wagon',
'siamang', 'orange', 'home_theater', 'cairn', 'hare',
'Norwegian_elkhound', 'Brabancon_griffon', 'American_black_bear',
```

'sulphur\_butterfly', 'Sealyham\_terrier', 'Walker\_hound',  
'tarantula', 'Persian\_cat', 'coral\_fungus', 'accordion', 'kelpie',  
'Great\_Pyrenees', 'wood\_rabbit', 'black-and-tan\_coonhound',  
'sunglasses', 'plow', 'whippet', 'rain\_barrel', 'bathtub', 'tiger',  
'snail', 'tick', 'wire-haired\_fox\_terrier', 'water\_bottle', 'wig',  
'platypus', 'Irish\_wolfhound', 'ram', 'gorilla',  
'entertainment\_center', 'toucan', 'mask', 'shopping\_cart',  
'Sussex\_spaniel', 'crate', 'grey\_whale', 'badger', 'Airedale',  
'Arabian\_camel', 'cockroach', 'lifeboat', 'rotisserie', 'goldfish',  
'stingray', 'warthog', 'bobsled', 'rhinoceros\_beetle', 'beaver',  
'brown\_bear', 'Maltese\_dog', 'weasel', 'quill',  
'Rhodesian\_ridgeback', 'Arctic\_fox', 'ashcan', 'bow\_tie',  
'soft-coated\_wheaten\_terrier', 'schipperke', 'bearskin',  
'Kerry\_blue\_terrier', 'ice\_lolly', 'American\_alligator',  
'mosquito\_net', 'sea\_lion', 'Boston\_bull', 'nail',  
'black-footed\_ferret', 'promontory', 'sarong', 'Doberman',  
'space\_heater', 'Great\_Dane', 'mailbox', 'Saluki', 'bakery',  
'sandal', 'leafhopper', 'barrel', 'water\_buffalo', 'polecat',  
'macaque', 'Japanese\_spaniel', 'folding\_chair', 'trench\_coat',  
'Angora', 'junco', 'crib', 'dalmatian', 'snowmobile',  
'flat-coated\_retriever', 'streetcar', 'window\_screen', 'bannister',  
'hair\_slide', 'meerkat', 'cannon', 'jaguar', 'Border\_terrier',  
'Greater\_Swiss\_Mountain\_dog', 'hay', 'apron', 'cloak', 'radiator',  
'muzzle', 'feather\_boa', 'rifle', 'chimpanzee', 'loggerhead',  
'torch', 'Mexican\_hairless', 'spindle', 'triceratops',  
'Appenzeller', 'stove', 'dingo', 'oscilloscope', 'common\_newt',  
'hotdog', 'medicine\_chest', 'quail', 'horse\_cart', 'four-poster',  
'pier', 'red\_fox', 'affenpinscher', 'assault\_rifle',  
'mashed\_potato', 'moped', 'hyena', 'seashore', 'tub', 'sports\_car',  
'swing', 'mink', 'bluetick', 'neck\_brace', 'grey\_fox', 'mongoose',  
'fur\_coat', 'Scotch\_terrier', 'spatula', 'paper\_towel', 'shoji',  
'toyshop', 'banded\_gecko', 'peacock', 'crossword\_puzzle',  
'tree\_frog', 'wombat', 'turnstile', 'sleeping\_bag', 'quilt',  
'Gila\_monster', 'giant\_panda', 'Dandie\_Dinmont', 'handkerchief',  
'sombrero', 'Indian\_elephant', 'coffee\_mug', 'gibbon', 'carton',  
'screw', 'minibus', 'hatchet', 'window\_shade', 'lawn\_mower',  
'washbasin', 'sock', 'prison', 'patio', 'china\_cabinet',  
'chain\_mail', 'breakwater', 'computer\_keyboard', 'goose',  
'lakeside', 'solar\_dish', 'table\_lamp', 'Windsor\_tie',  
'punching\_bag', 'comic\_book', 'sunglass', 'great\_white\_shark',  
'timber\_wolf', 'fountain', 'dugong', 'marmot', 'barbershop',  
'shovel', 'curly-coated\_retriever', 'lesser\_panda', 'monitor',  
'crutch', 'cash\_machine', 'printer', 'volcano', 'wallet', 'laptop',  
'bathing\_cap', 'confectionery', 'dam', 'killer\_whale', 'canoe',  
'Madagascar\_cat', 'jean', 'boathouse', 'cliff', 'maillot', 'iPod',  
'hand-held\_computer', 'black\_widow', 'Norwich\_terrier', 'necklace',  
'dining\_table', 'binoculars', 'cradle', 'sea\_urchin', 'cougar',  
'EntleBucher', 'basketball', 'lighter', 'saltshaker', 'harvester',

```

'television', 'envelope', 'house_finch', 'web_site', 'palace',
'shower_curtain', 'cab', 'snorkel', 'jigsaw_puzzle', 'sweatshirt',
'white_wolf', 'sliding_door', 'academic_gown', 'cowboy_hat',
'can_opener', 'cup', 'rule', 'soccer_ball', 'bucket', 'racket',
'menu', 'purse', 'Bernese_mountain_dog', 'dumbbell', 'projectile',
'dock', 'oxygen_mask', 'sandbar', 'umbrella', 'shower_cap', 'bagel'], dtype=object)

```

```
In [20]: image_pred.p3.unique()
```

```

Out[20]: array(['Shetland_sheepdog', 'Rhodesian_ridgeback', 'bloodhound',
'miniature_pinscher', 'Doberman', 'Greater_Swiss_Mountain_dog',
'terrapin', 'fur_coat', 'golden_retriever',
'soft-coated_wheaten_terrier', 'Labrador_retriever', 'Pekinese',
'Ibizan_hound', 'French_bulldog', 'malinois', 'Dandie_Dinmont',
'borzoi', 'partridge', 'bookcase', 'basenji', 'miniature_poodle',
'great_grey_owl', 'groenendael', 'Eskimo_dog', 'hamster', 'briard',
'papillon', 'flat-coated_retriever', 'gar', 'Chihuahua', 'Shih-Tzu',
'Pomeranian', 'dingo', 'power_drill', 'Saluki', 'Great_Pyrenees',
'West_Highland_white_terrier', 'collie', 'toy_poodle', 'vizsla',
'acorn', 'giant_schnauzer', 'teddy', 'common_iguana', 'wig',
'water_buffalo', 'coyote', 'seat_belt', 'kelpie', 'space_heater',
'Brabancon_griffon', 'standard_poodle', 'beagle',
'Irish_water_spaniel', 'bluetick', 'Weimaraner',
'Chesapeake_Bay_retriever', 'toilet_tissue',
'black-and-tan_coonhound', 'kuvasz', 'Christmas_stocking', 'badger',
'hen', 'Staffordshire_bullterrier', 'Yorkshire_terrier',
'Lakeland_terrier', 'weasel', 'ski_mask', 'cocker_spaniel',
'Australian_terrier', 'lampshade', 'oscilloscope', 'ram', 'jeep',
'ice_bear', 'African_grey', 'Great_Dane', 'curly-coated_retriever',
'door_mat', 'African_chameleon', 'schipperke', 'muzzle',
'triceratops', 'Newfoundland', 'Band_Aid', 'wood_rabbit',
'white_wolf', 'giant_panda', 'Welsh_springer_spaniel',
'French_horn', 'toy_terrier', 'Pembroke', 'Cardigan', 'bassinet',
'pug', 'Afghan_hound', 'American_Staffordshire_terrier', 'whippet',
'English_setter', 'panpipe', 'crane', 'mouse', 'titi', 'Angora',
'Boston_bull', 'silky_terrier', 'Japanese_spaniel', 'sandbar',
'balance_beam', 'black-footed_ferret', 'miniature_schnauzer',
'Blenheim_spaniel', 'bathtub', 'Saint_Bernard', 'redbone',
'goldfish', 'Norfolk_terrier', 'llama', 'koala', 'pillow', 'jersey',
'chow', 'minibus', 'malamute', 'bulletproof_vest', 'beach_wagon',
'cairn', 'plunger', 'paper_towel', 'wing', 'English_foxhound',
'Brittany_spaniel', 'bolete', 'ashcan', 'box_turtle', 'guinea_pig',
'bison', 'bull_mastiff', 'racket', 'cardoon', 'Tibetan_mastiff',
>window_screen', 'Irish_terrier', 'agama', 'common_newt',
'car_wheel', 'gorilla', 'bagel', 'clumber', 'Egyptian_cat',
'television', 'boxer', 'brown_bear', 'leafhopper',
'German_shepherd', 'Border_collie', 'menu', 'wolf_spider',
'bathing_cap', 'stinkhorn', 'drumstick', 'mask',

```

'Scottish\_deerhound', 'shower\_curtain', 'Appenzeller',  
'plastic\_bag', 'swimming\_trunks', 'prairie\_chicken', 'red\_wolf',  
'Maltese\_dog', 'snail', 'gibbon', 'Gordon\_setter', 'black\_swan',  
'beacon', 'wool', 'cowboy\_boot', 'Rottweiler', 'poncho', 'swing',  
'Arctic\_fox', 'bib', 'Italian\_greyhound', 'steam\_locomotive',  
'fountain', 'chickadee', 'abaya', 'Border\_terrier', 'bubble',  
'chimpanzee', 'hammerhead', 'Norwegian\_elkhound', 'Norwich\_terrier',  
'Airedale', 'Siamese\_cat', 'sea\_cucumber', 'seashore', 'nipple',  
'moped', 'Arabian\_camel', 'crayfish', 'wallaby',  
'wire-haired\_fox\_terrier', 'toilet\_seat', 'Old\_English\_sheepdog',  
'pajama', 'Walker\_hound', 'shovel', 'bucket', 'Sealyham\_terrier',  
'Windsor\_tie', 'Siberian\_husky', 'quill', 'Persian\_cat',  
'European\_fire\_salamander', 'three-toed\_sloth', 'swab', 'echidna',  
'tennis\_ball', 'Lhasa', 'coral\_reef', 'keeshond', 'mink', 'screw',  
'basset', 'wreck', 'kimono', 'German\_short-haired\_pointer',  
'joystick', 'microwave', 'Tibetan\_terrier', 'Irish\_wolfhound',  
'Samoyed', 'loggerhead', 'French\_loaf', 'Irish\_setter', 'komondor',  
'purse', 'greenhouse', 'broccoli', 'shopping\_basket', 'macaque',  
'squirrel\_monkey', 'green\_lizard', 'parallel\_bars', 'cloak',  
'chest', 'sundial', 'mosquito\_net', 'bath\_towel', 'cuirass',  
'zebra', 'lumbermill', 'wallet', 'feather\_boa', 'English\_springer',  
'electric\_fan', 'hippopotamus', 'ox', 'quilt', 'assault\_rifle',  
'axolotl', 'pot', 'toyshop', 'pizza', 'scuba\_diver', 'beaver',  
'Mexican\_hairless', 'cliff', 'loupe', 'wild\_boar', 'jaguar', 'hog',  
'polecat', 'lion', 'EntleBucher', 'hand-held\_computer', 'washbasin',  
'whiptail', 'rock\_crab', 'hare', 'shoji', 'sombrero', 'bell\_cote',  
'rifle', 'goose', 'pickup', 'sunglasses', 'limousine', 'bow\_tie',  
'pretzel', 'marmot', 'ice\_lolly', 'vacuum', 'dalmatian', 'prison',  
'shower\_cap', 'sliding\_door', 'dugong', 'otterhound', 'eel',  
'binder', 'bullfrog', 'soap\_dispenser', 'sea\_lion', 'carton',  
'brass', 'mitten', 'golfcart', 'cougar', 'warthog', 'umbrella',  
'neck\_brace', 'cup', 'book\_jacket', 'padlock', 'cab', 'chime',  
'Leonberg', 'viaduct', 'American\_black\_bear', 'tub', 'hand\_blower',  
'king\_penguin', 'rotisserie', 'bannister', 'passenger\_car',  
'mongoose', 'dhole', 'consomme', 'valley', 'park\_bench', 'mushroom',  
'barrow', 'parachute', 'desktop\_computer', 'snorkel', 'wok',  
'affenpinscher', 'space\_shuttle', 'rain\_barrel', 'ballplayer',  
'mountain\_tent', 'oxcart', 'buckeye', 'sunglass', 'croquet\_ball',  
'refrigerator', 'snow\_leopard', 'tripod', 'rapeseed', 'tiger\_cat',  
'Bernese\_mountain\_dog', 'notebook', 'maraca', 'pool\_table',  
'lakeside', 'theater\_curtain', 'pier', 'cheetah', 'mousetrap',  
'pop\_bottle', 'soccer\_ball', 'wombat', 'rhinoceros\_beetle',  
'paddlewheel', 'paintbrush', 'maze', 'hatchet', 'chain',  
'jigsaw\_puzzle', 'switch', 'Kerry\_blue\_terrier', 'barbell',  
'convertible', 'entertainment\_center', 'file', 'guillotine', 'nail',  
'standard\_schnauzer', 'bow', 'grocery\_store', 'boathouse', 'conch',  
'Bouvier\_des\_Flandres', 'grey\_fox', 'shopping\_cart', 'meerkat',  
'grand\_piano', 'envelope', 'screen', 'coffeepot', 'printer',

```

'otter', 'restaurant', 'bonnet', 'crossword_puzzle', 'go-kart',
'Sussex_spaniel', 'orangutan', 'canoe', 'barber_chair',
'traffic_light', 'ibex', 'can_opener', 'Indian_elephant', 'spatula',
'banana'], dtype=object)

```

```
In [21]: image_pred.p1_dog.value_counts()
```

```

Out[21]: True      1532
        False     543
        Name: p1_dog, dtype: int64

```

```
In [22]: image_pred.p2_dog.value_counts()
```

```

Out[22]: True      1553
        False     522
        Name: p2_dog, dtype: int64

```

```
In [23]: image_pred.p3_dog.value_counts()
```

```

Out[23]: True      1499
        False     576
        Name: p3_dog, dtype: int64

```

## Quality issues 'twitter\_archive' table

- tweet\_id column should be of object type, not integer.
- Missing values in 'in\_reply\_to\_status\_id', 'in\_reply\_to\_user\_id', 'retweeted\_status\_id', 'retweeted\_status\_user\_id', and 'retweeted\_status\_timestamp' columns can instead be left as null or filled with new values.
- 'timestamp' and 'retweeted\_status\_timestamp' are of object type instead of datetime.
- We only want to deal with non retweet tweets, hence the first rows where retweeted\_status\_id, retweeted\_status\_user\_id, and retweeted\_status\_timestamp are not empty, must be removed.
- The columns doggo, floofer, pupper and puppo can be converted to have true/false values instead of None values.

## 'tweet\_json' table

- Missing values in 'favorite\_count' and 'retweet\_count' columns. They can be left as null, or added with new values.
- 'favorite\_count' and 'retweet\_count' are object type and should be converted to numeric type.

## 'image\_pred' table

- Nondescriptive column headers p1, p2, p3, p1\_conf, p2\_conf, p3\_conf, p1\_dog, p2\_dog, p3\_dog.
- The columns p1, p2 and p3 are various breeds of dogs and hence must be categorical instead of object type.
- The columns p1\_dog, p2\_dog and p3\_dog must be of type bool.
- The columns 'p1', 'p2', 'p3' represent things which are not dogs. We can use False values in p1\_dog, p2\_dog and p3\_dog to remove these non dog entries.



### 1.2.1 Tidiness issues

- All the 3 tables namely 'twitter\_archive', 'image\_pred' and 'tweet\_json' have the same observational unit 'tweet\_id'.
- The rating\_numerator and rating\_denominator must be combined into single column Overall\_rating.

## 2 Cleaning

### 2.0.1 Define

- Combine all the 3 tables 'twitter\_archive', 'image\_pred' and 'tweet\_json' into a single table.

### 2.0.2 Code

```
In [24]: # Create a copy of the dataset before cleaning
df_clean = twitter_archive.copy()

# Add 2 new columns favorite_count and retweet_count to twitter_archive from tweet_json
df_clean['favorite_count'] = tweet_json.favorite_count
df_clean['retweet_count'] = tweet_json.retweet_count
# Now join all the columns of 'image_pred' to df_clean
df2_clean = pd.merge(df_clean,
                     image_pred,
                     left_on='tweet_id',
                     right_on='tweet_id',
                     how='right')
```

### 2.0.3 Test

```
In [25]: df2_clean.head()
```

```
Out[25]:
```

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

	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  ...  img_num  \
0 https://twitter.com/dog_rates/status/892420643...  ...        1
1 https://twitter.com/dog_rates/status/892177421...  ...        1
2 https://twitter.com/dog_rates/status/891815181...  ...        1
3 https://twitter.com/dog_rates/status/891689557...  ...        1
4 https://twitter.com/dog_rates/status/891327558...  ...        2

                                p1    p1_conf p1_dog                                p2    p2_conf p2_dog  \
0         orange  0.097049  False                                bagel  0.085851  False
1   Chihuahua  0.323581   True                                Pekinese  0.090647   True
2   Chihuahua  0.716012   True                                malamute  0.078253   True
3 paper_towel  0.170278  False  Labrador_retriever  0.168086   True
4      basset  0.555712   True    English_springer  0.225770   True

                                p3    p3_conf p3_dog
0                                banana  0.076110  False
1                                papillon  0.068957   True
2                                kelpie  0.031379   True
3                                spatula  0.040836  False
4 German_short-haired_pointer  0.175219   True

```

[5 rows x 30 columns]

## 2.0.4 Define

- Convert `tweet_id` column to object type as, we are not going to perform mathematical operations over them, those are labels or identifiers.
- Convert `'timestamp'` and `'retweeted_status_timestamp'` to datetime.
- Convert the columns `doggo`, `floofer`, `pupper` and `puppo` to bool type.

- Convert 'favorite\_count' and 'retweet\_count' to numeric type.
- Give descriptive column names to p1, p2, p3, p1\_conf, p2\_conf, p3\_conf, p1\_dog, p2\_dog, p3\_dog.

## 2.0.5 Code

```
In [26]: # Create another copy of the above cleaned dataset before further cleaning
df_clean_1 = df2_clean.copy()

# Convert tweet_id column to object type as, those are labels or identifiers.
df2_clean.tweet_id = df2_clean.tweet_id.astype('object')

# Convert 'timestamp' and 'retweeted_status_timestamp' to datetime
df2_clean['timestamp'] = pd.to_datetime(df2_clean['timestamp'])
df2_clean['retweeted_status_timestamp'] = pd.to_datetime(df2_clean['retweeted_status_timestamp'])

# Convert the columns doggo, floofer, pupper and puppo to true/false values
df2_clean['doggo'] = (df2_clean['doggo'] == 'doggo').astype(bool)
df2_clean['floofer'] = (df2_clean['floofer'] == 'floofer').astype(bool)
df2_clean['pupper'] = (df2_clean['pupper'] == 'pupper').astype(bool)
df2_clean['puppo'] = (df2_clean['puppo'] == 'puppo').astype(bool)

# Convert 'favorite_count' and 'retweet_count' to numeric type
df2_clean[['favorite_count', 'retweet_count']] = df2_clean[['favorite_count', 'retweet_count']].astype(int)

# Give descriptive column names to p1, p2, p3, p1_conf, p2_conf, p3_conf, p1_dog, p2_dog, p3_dog
df2_clean = df2_clean.rename(columns = {
    'p1': 'Possible_breed_1',
    'p2': 'Possible_breed_2',
    'p3': 'Possible_breed_3',
    'p1_conf': 'Breed_1_confidence',
    'p2_conf': 'Breed_2_confidence',
    'p3_conf': 'Breed_3_confidence',
    'p1_dog': 'Breed_1_present',
    'p2_dog': 'Breed_2_present',
    'p3_dog': 'Breed_3_present'})
```

## 2.0.6 Test

```
In [27]: df2_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2075 entries, 0 to 2074
Data columns (total 30 columns):
tweet_id                2075 non-null object
in_reply_to_status_id    23 non-null float64
```

```

in_reply_to_user_id      23 non-null float64
timestamp                2075 non-null datetime64[ns]
source                   2075 non-null object
text                     2075 non-null object
retweeted_status_id      81 non-null float64
retweeted_status_user_id 81 non-null float64
retweeted_status_timestamp 81 non-null datetime64[ns]
expanded_urls            2075 non-null object
rating_numerator         2075 non-null int64
rating_denominator       2075 non-null int64
name                     2075 non-null object
doggo                    2075 non-null bool
floofer                  2075 non-null bool
pupper                   2075 non-null bool
puppo                    2075 non-null bool
favorite_count           1794 non-null float64
retweet_count            1794 non-null float64
jpg_url                  2075 non-null object
img_num                  2075 non-null int64
Possible_breed_1         2075 non-null object
Breed_1_confidence       2075 non-null float64
Breed_1_present          2075 non-null bool
Possible_breed_2         2075 non-null object
Breed_2_confidence       2075 non-null float64
Breed_2_present          2075 non-null bool
Possible_breed_3         2075 non-null object
Breed_3_confidence       2075 non-null float64
Breed_3_present          2075 non-null bool
dtypes: bool(7), datetime64[ns](2), float64(9), int64(3), object(9)
memory usage: 403.2+ KB

```

### 2.0.7 Define

- Convert Possible\_breed\_1, Possible\_breed\_2, Possible\_breed\_3 to category type
- Convert Breed\_1\_present, Breed\_2\_present, Breed\_3\_present to the type bool.
- Remove the retweets so that we have only original tweets.
- Remove records of Possible\_breed\_1, Possible\_breed\_2, Possible\_breed\_3 which are not dogs i.e which have 'False' values in Breed\_1\_present, Breed\_2\_present and Breed\_3\_present.

### 2.0.8 Code

```

In [28]: # Convert Possible_breed_1, Possible_breed_2, Possible_breed_3 to category type
df2_clean.Possible_breed_1 = df2_clean.Possible_breed_1.astype('category')
df2_clean.Possible_breed_2 = df2_clean.Possible_breed_2.astype('category')
df2_clean.Possible_breed_3 = df2_clean.Possible_breed_3.astype('category')

```

```

# Convert Breed_1_present, Breed_2_present, Breed_3_present to the type bool
df2_clean.Breed_1_present = df2_clean.Breed_1_present.astype('bool')
df2_clean.Breed_2_present = df2_clean.Breed_2_present.astype('bool')
df2_clean.Breed_3_present = df2_clean.Breed_3_present.astype('bool')

# Remove the retweets so that we have only original tweets i.e remove rows for which re
df2_clean = df2_clean[df2_clean.retweeted_status_id.isnull()]

In [29]: # Also we can delete the retweet columns as we don't need retweets
del df2_clean['retweeted_status_id']
del df2_clean['retweeted_status_user_id']
del df2_clean['retweeted_status_timestamp']

In [31]: # Create a copy called df_unclean which is the dataset having things which are not dogs
df_unclean = df2_clean.copy()

In [32]: # Remove records of Possible_breed_1, Possible_breed_2, Possible_breed_3 which are not
df2_clean = df2_clean[df2_clean['Breed_1_present'] | df2_clean['Breed_2_present'] | df2_c

```

## 2.0.9 Test

```

In [33]: df2_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1686 entries, 1 to 2074
Data columns (total 27 columns):
tweet_id                1686 non-null object
in_reply_to_status_id   20 non-null float64
in_reply_to_user_id     20 non-null float64
timestamp               1686 non-null datetime64[ns]
source                 1686 non-null object
text                   1686 non-null object
expanded_urls          1686 non-null object
rating_numerator        1686 non-null int64
rating_denominator      1686 non-null int64
name                   1686 non-null object
doggo                  1686 non-null bool
floofer                1686 non-null bool
pupper                1686 non-null bool
puppo                  1686 non-null bool
favorite_count          1442 non-null float64
retweet_count           1442 non-null float64
jpg_url                 1686 non-null object
img_num                 1686 non-null int64
Possible_breed_1        1686 non-null category
Breed_1_confidence      1686 non-null float64
Breed_1_present         1686 non-null bool
Possible_breed_2        1686 non-null category
Breed_2_confidence      1686 non-null float64

```

```

Breed_2_present      1686 non-null bool
Possible_breed_3     1686 non-null category
Breed_3_confidence   1686 non-null float64
Breed_3_present      1686 non-null bool
dtypes: bool(7), category(3), datetime64[ns](1), float64(7), int64(3), object(6)
memory usage: 267.8+ KB

```

## 2.0.10 Define

- Combine rating\_numerator and rating\_denominator into a single column Overall\_rating.

## 2.0.11 Code

```

In [34]: df2_clean['Overall_rating'] = (df2_clean.rating_numerator)/(df2_clean.rating_denominator)

In [35]: del df2_clean['rating_numerator']
         del df2_clean['rating_denominator']

```

## 2.0.12 Test

```

In [36]: df2_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1686 entries, 1 to 2074
Data columns (total 26 columns):
tweet_id      1686 non-null object
in_reply_to_status_id  20 non-null float64
in_reply_to_user_id  20 non-null float64
timestamp     1686 non-null datetime64[ns]
source        1686 non-null object
text          1686 non-null object
expanded_urls  1686 non-null object
name          1686 non-null object
doggo         1686 non-null bool
floofer       1686 non-null bool
pupper        1686 non-null bool
puppo         1686 non-null bool
favorite_count  1442 non-null float64
retweet_count  1442 non-null float64
jpg_url       1686 non-null object
img_num       1686 non-null int64
Possible_breed_1  1686 non-null category
Breed_1_confidence  1686 non-null float64
Breed_1_present  1686 non-null bool
Possible_breed_2  1686 non-null category
Breed_2_confidence  1686 non-null float64
Breed_2_present  1686 non-null bool
Possible_breed_3  1686 non-null category

```

```
Breed_3_confidence      1686 non-null float64
Breed_3_present         1686 non-null bool
Overall_rating          1686 non-null float64
dtypes: bool(7), category(3), datetime64[ns](1), float64(8), int64(1), object(6)
memory usage: 254.6+ KB
```

The missing values in 'in\_reply\_to\_status\_id', 'in\_reply\_to\_user\_id', 'retweeted\_status\_id', 'retweeted\_status\_id', 'retweeted\_status\_timestamp', 'favorite\_count' and 'retweet\_count' columns cannot be treated as we cannot collect and add those missing values.

### 3 Storing, Analyzing and Visualizing Data

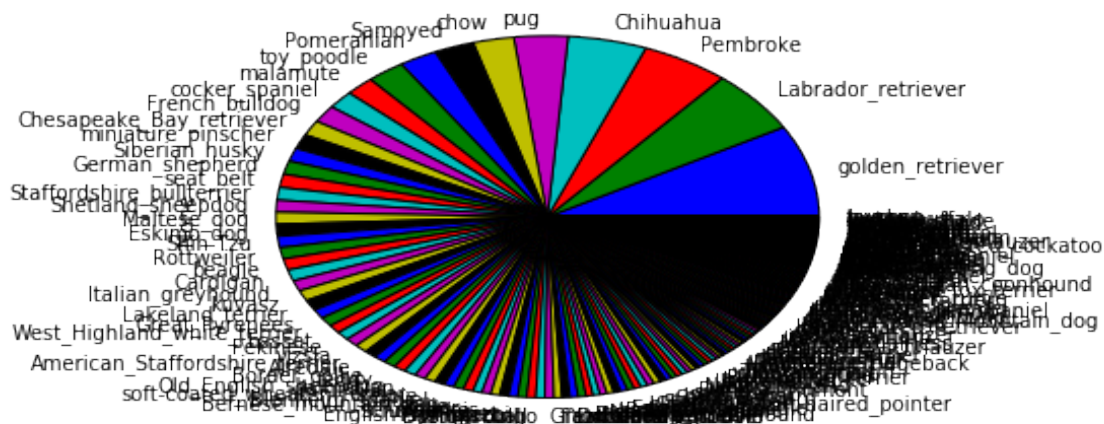
The final cleaned dataframe `df2_clean` can now be stored as CSV file named `twitter_archive_master.csv`.

```
In [37]: df2_clean.to_csv('twitter_archive_master.csv')
```

### Plot Piechart to find the difference in 'Possible\_breed\_1' column between clean dataset (having only dog entries) and unclean dataset

```
In [38]: import matplotlib.pyplot as plt
          %matplotlib inline
          # plot all possible dog breeds for df2_clean which has only dogs
          df2_clean.Possible_breed_1.value_counts().plot(kind='pie')
```

```
Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x929ae10>
```



```
In [39]: # plot all possible dog breeds for df_unclean which has things other than dogs
df_unclean.Possible_breed_1.value_counts().plot(kind='pie')
```

Out[39]: <matplotlib.axes.\_subplots.AxesSubplot at 0xab7cba8>



Analyze 'Overall\_rating' column to find number of entries with Overall\_rating >=1

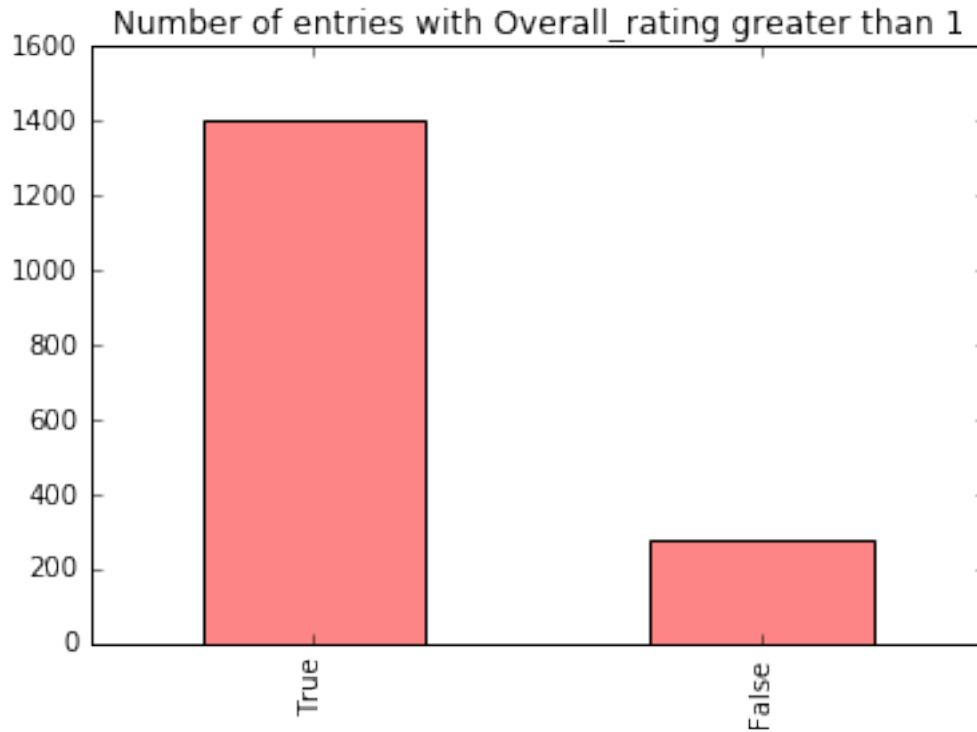
```
In [40]: # plotting number of entries having overall_rating >=1
df2_clean['rating_morethan_one'] = df2_clean.Overall_rating >=1
df2_clean.rating_morethan_one.value_counts()
```

```
Out[40]: True      1406
False       280
Name: rating_morethan_one, dtype: int64
```

```
In [41]: df2_clean.rating_morethan_one.value_counts().plot(kind='bar',
title='Number of entries with Overall_rating greater than 1',
color = '#fe8585')
```

Out[41]: <matplotlib.axes.\_subplots.AxesSubplot at 0xc9517f0>





### 3.0.1 Insights

- As seen from the two piecharts plotted above, the clean dataset i.e df2\_clean has higher proportion of dog breeds like golden\_retriever, Labrador\_retriever, Chihuahua, Pembroke etc , but there is comparatively lesser proportion of dogs in df\_unclean (the unclean dataset).
- Only 280 entries of dogs have Overall\_ratings less than 1 , and the rest 1406 entries have Overall\_rating greater than 1.
- The 5 most popular dog breeds from the above dataset are golden\_retriever, Labrador\_retriever, Pembroke, Chihuahua and pug.