

wrangle_act

November 25, 2020

1 Data Wrangling

1.1 Introduction

- The goal of this project is to wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations.

1.1.1 The Data We Have

- Enhanced Twitter Archive
- Additional Data via the Twitter API
- Image Predictions File ### Steps
- Data wrangling, which consists of:
 - Gathering data
 - Assessing data
 - Cleaning data
- Storing, analyzing, and visualizing our wrangled data
- Reporting on 1) our data wrangling efforts and 2) our data analyses and visualizations

```
In [1]: import pandas as pd
import numpy as np
import requests
import tweepy
import os
import json
import time
import re
import warnings
import matplotlib
import matplotlib.pyplot as plt
#import seaborn as sns
%matplotlib inline
matplotlib.style.use('ggplot')
```

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

```

Out[2]:
      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 [3]: 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 [4]: url="https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions
        response = requests.get(url)
        with open('image_predictions.tsv', 'wb') as file:
            file.write(response.content)
```

```
        image_predictions = pd.read_csv('image_predictions.tsv', sep='\t')
```

```
In [5]: 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
```

Use Twitter API and Python's Tweepy library and store each tweet's set of JSON data in a file.

```
In [6]: # Create a dataframe with tweet ID, retweet count, and favorite count
df_list = []
with open('tweet-json.txt') as file:
    for line in file:
        tweet = json.loads(line)

        df_list.append({
            'tweet_id': tweet['id'],
            'retweet_count': tweet['retweet_count'],
            'favorite_count': tweet['favorite_count']
        })
df_tweet = pd.DataFrame(df_list, columns = ['tweet_id', 'retweet_count', 'favorite_count'])

df_tweet.head()
```

```
Out[6]:
```

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

1.1.2 Assess Data

- Inspecting the dataset for two things: data quality issues (i.e. content issues) and lack of tidiness (i.e. structural issues).

Twitter archive

- Visual Inspection of twitter archive-

```
In [7]: twitter_archive
```

```
Out[7]:
```

	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/ws2AMaWpW	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 [8]: twitter_archive.head()

```
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	

	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 [9]: twitter_archive.tail()

```
Out[9]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
2351	666049248165822465	NaN	NaN	
2352	666044226329800704	NaN	NaN	
2353	666033412701032449	NaN	NaN	
2354	666029285002620928	NaN	NaN	
2355	666020888022790149	NaN	NaN	

	timestamp	\
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	\
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	\
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	\
2351	NaN	NaN	
2352	NaN	NaN	
2353	NaN	NaN	
2354	NaN	NaN	
2355	NaN	NaN	

	expanded_urls	rating_numerator	\
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
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

Programmatic Inspection of twitter archive

```
In [10]: 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 [11]: twitter_archive.describe()
```

```
Out[11]:
```

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

	retweeted_status_id	retweeted_status_user_id	rating_numerator	\
--	---------------------	--------------------------	------------------	---

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

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

```
In [12]: sum(twitter_archive.duplicated())
```

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

```
In [13]: twitter_archive['source'].value_counts()
```

```
Out[13]: <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 [14]: twitter_archive['rating_denominator'].value_counts()
```

```
Out[14]: 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 [15]: # Check if there are any records in twitter_arc which are retweets
        len(twitter_archive[twitter_archive.retweeted_status_id.isnull() == False])
```

```
Out[15]: 181
```

```
In [16]: # Sort by rating_denominator values
        twitter_archive.rating_denominator.value_counts().sort_index()
```

```
Out[16]: 0          1
         2          1
         7          1
        10        2333
        11          3
        15          1
        16          1
        20          2
        40          1
        50          3
        70          1
        80          2
        90          1
       110          1
       120          1
       130          1
       150          1
       170          1
Name: rating_denominator, dtype: int64
```

```
In [17]: # Sort by rating_numerator values
        twitter_archive.rating_numerator.value_counts().sort_index()
```

```
Out[17]: 0          2
         1          9
         2          9
         3         19
         4         17
         5         37
         6         32
         7         55
         8        102
         9        158
        10        461
        11        464
        12        558
        13        351
```

14	54
15	2
17	1
20	1
24	1
26	1
27	1
44	1
45	1
50	1
60	1
75	2
80	1
84	1
88	1
99	1
121	1
143	1
144	1
165	1
182	1
204	1
420	2
666	1
960	1
1776	1

Name: rating_numerator, dtype: int64

1.2 OBSERVATIONS For Twitter archive

1.2.1 Quality

- * ID variables are sometimes integers or floats (numeric)
- * "in_reply_to..." and "retweeted_status..." variables are numeric
- * retweets are present in the data
- * some of the column names are not meaningful
- * "timestamp" and "retweeted_status_timestamp" are not a datetime variable
- * "source" values are formatted as <a> href=url <a/>
- * rating_numerators are not always correctly accounting for decimals
- * the dog names are not standardized

1.2.2 tidyness

- * more than one stage is filled for a particular dog
- * "source" and "expanded_urls" have several informations inside them
- * columns "doggo", "floofer", "pupper" and "puppo" refer to the same measurement unit, i.e, dog

Image Predictions Visual Inspection of image predictions-

```
In [18]: image_predictions
```

```
Out[18]:
```

	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/CT5w9gUW4AAsBNN.jpg
15	666099513787052032	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16	666102155909144576	https://pbs.twimg.com/media/CT54YGiwUAEZnoK.jpg
17	666104133288665088	https://pbs.twimg.com/media/CT56LSZWAAALJj2.jpg
18	666268910803644416	https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
19	666273097616637952	https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
20	666287406224695296	https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
21	666293911632134144	https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22	666337882303524864	https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg
23	666345417576210432	https://pbs.twimg.com/media/CT9Vn7PWAAA_ZCM.jpg
24	666353288456101888	https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25	666362758909284353	https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg
26	666373753744588802	https://pbs.twimg.com/media/CT9vZEYUAAALZ05.jpg
27	666396247373291520	https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28	666407126856765440	https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.jpg
29	666411507551481857	https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...
2045	886366144734445568	https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046	886680336477933568	https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2047	886736880519319552	https://pbs.twimg.com/media/DE5Se8FXcAAJFfx4.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-UQAACu20.jpg
2057	888804989199671297	https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058	888917238123831296	https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg

2059	889278841981685760	https://pbs.twimg.com/ext_tw_video_thumb/88927...
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/DFrEyVuWAAA03t9.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		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 [19]: image_predictions.head()

```
Out[19]:
```

	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 [20]: image_predictions.tail()

```
Out[20]:
```

	tweet_id	jpg_url	\
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	p2	p2_conf	\
--	---------	----	---------	--------	----	---------	---

2070	2	basset	0.555712	True	English_springer	0.225770
2071	1	paper_towel	0.170278	False	Labrador_retriever	0.168086
2072	1	Chihuahua	0.716012	True	malamute	0.078253
2073	1	Chihuahua	0.323581	True	Pekinese	0.090647
2074	1	orange	0.097049	False	bagel	0.085851

	p2_dog		p3	p3_conf	p3_dog
2070	True	German_short-haired_pointer	0.175219	True	True
2071	True		spatula	0.040836	False
2072	True		kelpie	0.031379	True
2073	True		papillon	0.068957	True
2074	False		banana	0.076110	False

Programmatic assesment of image predictions:

```
In [21]: image_predictions.shape
```

```
Out[21]: (2075, 12)
```

```
In [22]: 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 [23]: sum(image_predictions.duplicated())
```

```
Out[23]: 0
```

```
In [24]: image_predictions.img_num.unique()
```

```
Out[24]: array([1, 4, 2, 3])
```

```
In [25]: sum(image_predictions.jpg_url.duplicated()==True)
```

```
Out[25]: 66
```

```
In [26]: image_predictions['img_num'].value_counts()
```

```
Out[26]: 1    1780
         2     198
         3      66
         4      31
         Name: img_num, dtype: int64
```

```
In [27]: image_predictions.describe()
```

```
Out[27]:
```

	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

1.3 OBSERVATIONS

1.3.1 Quality

- * "tweet_id" and "tweet_id" are numeric and not categorical (string)
- * 2075 tweet ids present (while the archive dataset has a total of 2356 ids, so 281 IDs are missing)
- * column names can be improved
- * p1, p2, and p3 contain underscores instead of spaces in the labels

JSON file Visual Inspection of json file-

```
In [28]: df_tweet
```

```
Out[28]:
```

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

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

[2354 rows x 3 columns]

```
In [29]: df_tweet.head()
```

```
Out[29]:
```

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

```
In [30]: df_tweet.tail()
```

```
Out[30]:
```

	tweet_id	retweet_count	favorite_count
2349	666049248165822465	41	111
2350	666044226329800704	147	311
2351	666033412701032449	47	128
2352	666029285002620928	48	132
2353	666020888022790149	532	2535

Programmatic assesment of json file :

```
In [31]: df_tweet.info()
```

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

```
In [32]: df_tweet.describe()
```

```
Out[32]:
```

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

```
In [33]: sum(df_tweet.duplicated()==True)
```

```
Out[33]: 0
```

```
In [34]: #Number of missing IDs
2356-2338
```

```
Out[34]: 18
```

1.4 OBSERVATIONS FOR API TWITTETR

1.4.1 Quality

* Number of missing IDs = 2356-2338 = 18

(Let's assume they have been deleted, the errors during API data download)

2 Clean

Trying to resolve the issues

```
In [35]: # let's prepare a working copy of each separate dataset
         twitter_archive_copy = twitter_archive.copy()
         image_prediction1 = image_predictions.copy()
         df_tweet1 = df_tweet.copy()
```

3 Twitter Archive

3.1 Define

- Replace all faulty names to none

```
In [36]: twitter_archive_copy[twitter_archive_copy.name.str.islower()==True]['name'].unique()
```

```
Out[36]: array(['such', 'a', 'quite', 'not', 'one', 'incredibly', 'mad', 'an',
               'very', 'just', 'my', 'his', 'actually', 'getting', 'this',
               'unacceptable', 'all', 'old', 'infuriating', 'the', 'by',
               'officially', 'life', 'light', 'space'], dtype=object)
```

3.2 Code

```
In [37]: faulty = ['such', 'a', 'quite', 'not', 'one', 'incredibly', 'mad', 'an',
                  'very', 'just', 'my', 'his', 'actually', 'getting', 'this',
                  'unacceptable', 'all', 'old', 'infuriating', 'the', 'by',
                  'officially', 'life', 'light', 'space']
```

```
In [38]: for x in faulty:
         twitter_archive_copy.name.replace(x, 'None', inplace=True)
```

3.3 Test

```
In [39]: twitter_archive.name.unique()
```

```
Out[39]: array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', 'None', 'Jax',
               'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver', 'Jim',
               'Zeke', 'Ralphus', 'Canela', 'Gerald', 'Jeffrey', 'such', 'Maya',
               'Mingus', 'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Lilly',
               'Earl', 'Lola', 'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald',
               'Rusty', 'Gus', 'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'a',
```

'Elliot', 'Louis', 'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack',
 'Emmy', 'Steven', 'Beau', 'Snoopy', 'Shadow', 'Terrance', 'Aja',
 'Penny', 'Dante', 'Nelly', 'Ginger', 'Benedict', 'Venti', 'Goose',
 'Nugget', 'Cash', 'Coco', 'Jed', 'Sebastian', 'Walter', 'Sierra',
 'Monkey', 'Harry', 'Kody', 'Lassie', 'Rover', 'Napolean', 'Dawn',
 'Boomer', 'Cody', 'Rumble', 'Clifford', 'quite', 'Dewey', 'Scout',
 'Gizmo', 'Cooper', 'Harold', 'Shikha', 'Jamesy', 'Lili', 'Sammy',
 'Meatball', 'Paisley', 'Albus', 'Neptune', 'Quinn', 'Belle',
 'Zooey', 'Dave', 'Jersey', 'Hobbes', 'Burt', 'Lorenzo', 'Carl',
 'Jordy', 'Milky', 'Trooper', 'Winston', 'Sophie', 'Wyatt', 'Rosie',
 'Thor', 'Oscar', 'Luna', 'Callie', 'Cermet', 'George', 'Marlee',
 'Arya', 'Einstein', 'Alice', 'Rumpole', 'Benny', 'Aspen', 'Jarod',
 'Wiggles', 'General', 'Sailor', 'Astrid', 'Iggy', 'Snoop', 'Kyle',
 'Leo', 'Riley', 'Gidget', 'Noosh', 'Odin', 'Jerry', 'Charlie',
 'Georgie', 'Rontu', 'Cannon', 'Furzey', 'Daisy', 'Tuck', 'Barney',
 'Vixen', 'Jarvis', 'Mimosa', 'Pickles', 'Bungalo', 'Brady', 'Margo',
 'Sadie', 'Hank', 'Tycho', 'Stephan', 'Indie', 'Winnie', 'Bentley',
 'Ken', 'Max', 'Maddie', 'Pipsy', 'Monty', 'Sojourner', 'Odie',
 'Arlo', 'Sunny', 'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera',
 'Buddy', 'Ava', 'Rory', 'Eli', 'Ash', 'Tucker', 'Tobi', 'Chester',
 'Wilson', 'Sunshine', 'Lipton', 'Gabby', 'Bronte', 'Poppy', 'Rhino',
 'Willow', 'not', 'Orion', 'Eevee', 'Smiley', 'Logan', 'Moreton',
 'Klein', 'Miguel', 'Emanuel', 'Kuyu', 'Dutch', 'Pete', 'Scooter',
 'Reggie', 'Kyro', 'Samson', 'Loki', 'Mia', 'Malcolm', 'Dexter',
 'Alfie', 'Fiona', 'one', 'Mutt', 'Bear', 'Doobert', 'Beebop',
 'Alexander', 'Sailer', 'Brutus', 'Kona', 'Boots', 'Ralphie', 'Phil',
 'Cupid', 'Pawnd', 'Pilot', 'Ike', 'Mo', 'Toby', 'Sweet', 'Pablo',
 'Nala', 'Balto', 'Crawford', 'Gabe', 'Mattie', 'Jimison',
 'Hercules', 'Duchess', 'Harlso', 'Sampson', 'Sundance', 'Luca',
 'Flash', 'Finn', 'Peaches', 'Howie', 'Jazzy', 'Anna', 'Bo',
 'Seamus', 'Wafer', 'Chelsea', 'Tom', 'Moose', 'Florence', 'Autumn',
 'Dido', 'Eugene', 'Herschel', 'Strudel', 'Tebow', 'Chloe', 'Betty',
 'Timber', 'Binky', 'Dudley', 'Comet', 'Larry', 'Levi', 'Akumi',
 'Titan', 'Olivia', 'Alf', 'Oshie', 'Bruce', 'Chubbs', 'Sky',
 'Atlas', 'Eleanor', 'Layla', 'Rocky', 'Baron', 'Tyr', 'Bauer',
 'Swagger', 'Brandi', 'Mary', 'Moe', 'Halo', 'Augie', 'Craig', 'Sam',
 'Hunter', 'Pavlov', 'Maximus', 'Wallace', 'Ito', 'Milo', 'Ollie',
 'Cali', 'Lennon', 'incredibly', 'Major', 'Duke', 'Reginald',
 'Sansa', 'Shooter', 'Django', 'Diogi', 'Sonny', 'Philbert',
 'Marley', 'Severus', 'Ronnie', 'Anakin', 'Bones', 'Mauve', 'Chef',
 'Doc', 'Sobe', 'Longfellow', 'Mister', 'Iroh', 'Baloo', 'Stubert',
 'Paull', 'Timison', 'Davey', 'Pancake', 'Tyrone', 'Snicku', 'Ruby',
 'Brody', 'Rizzy', 'Mack', 'Butter', 'Nimbus', 'Laika', 'Dobby',
 'Juno', 'Maude', 'Lily', 'Newt', 'Benji', 'Nida', 'Robin',
 'Monster', 'BeBe', 'Remus', 'Mabel', 'Misty', 'Happy', 'Mosby',
 'Maggie', 'Leela', 'Ralphie', 'Brownie', 'Meyer', 'Stella', 'mad',
 'Frank', 'Tonks', 'Lincoln', 'Oakley', 'Dale', 'Rizzo', 'Arnie',
 'Pinot', 'Dallas', 'Hero', 'Frankie', 'Stormy', 'Mairi', 'Loomis',

'Godi', 'Kenny', 'Deacon', 'Timmy', 'Harper', 'Chipson', 'Combo',
 'Dash', 'Bell', 'Hurley', 'Jay', 'Mya', 'Strider', 'an', 'Wesley',
 'Solomon', 'Huck', 'very', 'O', 'Blue', 'Finley', 'Sprinkles',
 'Heinrich', 'Shakespeare', 'Fizz', 'Chip', 'Grey', 'Roosevelt',
 'Gromit', 'Willem', 'Dakota', 'Dixie', 'Al', 'Jackson', 'just',
 'Carbon', 'DonDon', 'Kirby', 'Lou', 'Nollie', 'Chevy', 'Tito',
 'Louie', 'Rupert', 'Rufus', 'Brudge', 'Shadoe', 'Colby', 'Angel',
 'Brat', 'Tove', 'my', 'Aubie', 'Kota', 'Eve', 'Glenn', 'Shelby',
 'Sephie', 'Bonaparte', 'Albert', 'Wishes', 'Rose', 'Theo', 'Rocco',
 'Fido', 'Emma', 'Spencer', 'Lilli', 'Boston', 'Brandonald', 'Corey',
 'Leonard', 'Chompsky', 'Beckham', 'Devón', 'Gert', 'Watson',
 'Rubio', 'Keith', 'Dex', 'Carly', 'Ace', 'Tayzie', 'Grizzie',
 'Fred', 'Gilbert', 'Zoe', 'Stewie', 'Calvin', 'Lilah', 'Spanky',
 'Jameson', 'Piper', 'Atticus', 'Blu', 'Dietrich', 'Divine', 'Tripp',
 'his', 'Cora', 'Huxley', 'Keurig', 'Bookstore', 'Linus', 'Abby',
 'Shaggy', 'Shiloh', 'Gustav', 'Arlen', 'Percy', 'Lenox', 'Sugar',
 'Harvey', 'Blanket', 'actually', 'Geno', 'Stark', 'Beya', 'Kilo',
 'Kayla', 'Maxaroni', 'Doug', 'Edmund', 'Aqua', 'Theodore', 'Chase',
 'getting', 'Rorie', 'Simba', 'Charles', 'Bayley', 'Axel',
 'Storkson', 'Remy', 'Chadrick', 'Kellogg', 'Buckley', 'Livvie',
 'Terry', 'Hermione', 'Ralpher', 'Aldrick', 'this', 'unacceptable',
 'Rooney', 'Crystal', 'Ziva', 'Stefan', 'Pupcasso', 'Puff',
 'Flurpson', 'Coleman', 'Enchilada', 'Raymond', 'all', 'Rueben',
 'Cilantro', 'Karll', 'Sprout', 'Blitz', 'Bloop', 'Lillie',
 'Ashleigh', 'Kreggory', 'Sarge', 'Luther', 'Ivar', 'Jangle',
 'Schnitzel', 'Panda', 'Berkeley', 'Ralphé', 'Charleson', 'Clyde',
 'Harnold', 'Sid', 'Pippa', 'Otis', 'Carper', 'Bowie',
 'Alexanderson', 'Suki', 'Barclay', 'Skittle', 'Ebby', 'Flávio',
 'Smokey', 'Link', 'Jennifur', 'Ozzy', 'Bluebert', 'Stephanus',
 'Bubbles', 'old', 'Zeus', 'Bertson', 'Nico', 'Michelangelo',
 'Siba', 'Calbert', 'Curtis', 'Travis', 'Thumas', 'Kanu', 'Lance',
 'Opie', 'Kane', 'Olive', 'Chuckles', 'Staniel', 'Sora', 'Beemo',
 'Gunner', 'infuriating', 'Lacy', 'Tater', 'Olaf', 'Cecil', 'Vince',
 'Karma', 'Billy', 'Walker', 'Rodney', 'Klevin', 'Malikai', 'Bobble',
 'River', 'Jebberson', 'Remington', 'Farfle', 'Jiminus', 'Clarkus',
 'Finnegus', 'Cupcake', 'Kathmandu', 'Ellie', 'Katie', 'Kara',
 'Adele', 'Zara', 'Ambrose', 'Jimothy', 'Bode', 'Terrenth', 'Reese',
 'Chesterson', 'Lucia', 'Bisquick', 'Ralphson', 'Socks', 'Rambo',
 'Rudy', 'Fiji', 'Rilo', 'Bilbo', 'Coopson', 'Yoda', 'Millie',
 'Chet', 'Crouton', 'Daniel', 'Kaia', 'Murphy', 'Dotsy', 'Eazy',
 'Coops', 'Fillup', 'Miley', 'Charl', 'Reagan', 'Yukon', 'CeCe',
 'Cuddles', 'Claude', 'Jessiga', 'Carter', 'Ole', 'Pherb', 'Blipson',
 'Reptar', 'Trevith', 'Berb', 'Bob', 'Colin', 'Brian', 'Oliviér',
 'Grady', 'Kobe', 'Freddery', 'Bodie', 'Dunkin', 'Wally', 'Tupawc',
 'Amber', 'Edgar', 'Teddy', 'Kingsley', 'Brockly', 'Richie', 'Molly',
 'Vinscent', 'Cedrick', 'Hazel', 'Lolo', 'Eriq', 'Phred', 'the',
 'Oddie', 'Maxwell', 'Geoff', 'Covach', 'Durg', 'Fynn', 'Ricky',
 'Herald', 'Lucky', 'Ferg', 'Trip', 'Clarence', 'Hamrick', 'Brad',


```
'Pubert', 'Frönq', 'Derby', 'Lizzie', 'Ember', 'Blakely', 'Opal',
'Marq', 'Kramer', 'Barry', 'Gordon', 'Baxter', 'Mona', 'Horace',
'Crimson', 'Birf', 'Hammond', 'Lorelei', 'Marty', 'Brooks',
'Petrick', 'Hubertson', 'Gerbald', 'Oreo', 'Bruiser', 'Perry',
'Bobby', 'Jeph', 'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe', 'Tiger',
'Jiminy', 'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Darrel', 'Taco',
'Joey', 'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses', 'Toffee',
'Apollo', 'Asher', 'Glacier', 'Chuck', 'Champ', 'Ozzie', 'Griswold',
'Cheesy', 'Moofasa', 'Hector', 'Goliath', 'Kawhi', 'by', 'Emmie',
'Penelope', 'Willie', 'Rinna', 'Mike', 'William', 'Dwight', 'Evy',
'officially', 'Rascal', 'Linda', 'Tug', 'Tango', 'Grizz', 'Jerome',
'Crumpet', 'Jessifer', 'Izzy', 'Ralph', 'Sandy', 'Humphrey',
'Tassy', 'Juckson', 'Chuq', 'Tyrus', 'Karl', 'Godzilla', 'Vinnie',
'Kenneth', 'Herm', 'Bert', 'Striker', 'Donny', 'Pepper', 'Bernie',
'Buddah', 'Lenny', 'Arnold', 'Zuzu', 'Mollie', 'Laela', 'Tedders',
'Superpup', 'Rufio', 'Jeb', 'Rodman', 'Jonah', 'Chesney', 'life',
'Henry', 'Bobbay', 'Mitch', 'Kaiya', 'Acro', 'Aiden', 'Obie', 'Dot',
'Shnuggles', 'Kendall', 'Jeffri', 'Steve', 'Mac', 'Fletcher',
'Kenzie', 'Pumpkin', 'Schnozz', 'Gustaf', 'Cheryl', 'Ed',
'Leonidas', 'Norman', 'Caryl', 'Scott', 'Taz', 'Darby', 'Jackie',
'light', 'Jazz', 'Franq', 'Pippin', 'Rolf', 'Snickers', 'Ridley',
'Cal', 'Bradley', 'Bubba', 'Tuco', 'Patch', 'Mojo', 'Batdog',
'Dylan', 'space', 'Mark', 'JD', 'Alejandro', 'Scruffers', 'Pip',
'Julius', 'Tanner', 'Sparky', 'Anthony', 'Holly', 'Jett', 'Amy',
'Sage', 'Andy', 'Mason', 'Trigger', 'Antony', 'Creg', 'Traviss',
'Gin', 'Jeffrie', 'Danny', 'Ester', 'Pluto', 'Bloo', 'Edd', 'Willy',
'Herb', 'Damon', 'Peanut', 'Nigel', 'Butters', 'Sandra', 'Fabio',
'Randall', 'Liam', 'Tommy', 'Ben', 'Raphael', 'Julio', 'Andru',
'Kloey', 'Shawwn', 'Skye', 'Kollin', 'Ronduh', 'Billl', 'Saydee',
'Dug', 'Tessa', 'Sully', 'Kirk', 'Ralf', 'Clarq', 'Jaspers',
'Samsom', 'Harrison', 'Chaz', 'Jeremy', 'Jaycob', 'Lambeau',
'Ruffles', 'Amélie', 'Bobb', 'Banditt', 'Kevon', 'Winifred', 'Hanz',
'Churlie', 'Zeek', 'Timofy', 'Maks', 'Jomathan', 'Kallie', 'Marvin',
'Spark', 'Gòrdón', 'Jo', 'DayZ', 'Jareld', 'Torque', 'Ron',
'Skittles', 'Cleopatra', 'Erik', 'Stu', 'Tedrick', 'Filup',
'Kial', 'Naphaniel', 'Dook', 'Hall', 'Philippe', 'Biden', 'Fwed',
'Genevieve', 'Joshwa', 'Bradlay', 'Clybe', 'Keet', 'Carll',
'Jockson', 'Josep', 'Lugan', 'Christoper'], dtype=object)
```

```
In [40]: twitter_archive_copy.name.unique()
```

```
Out[40]: array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', 'None', 'Jax',
'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver', 'Jim',
'Zeke', 'Ralphus', 'Canela', 'Gerald', 'Jeffrey', 'Maya', 'Mingus',
'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Lilly', 'Earl',
'Lola', 'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald', 'Rusty',
'Gus', 'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'Elliot', 'Louis',
'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack', 'Emmy', 'Steven',
```

'Beau', 'Snoopy', 'Shadow', 'Terrance', 'Aja', 'Penny', 'Dante',
'Nelly', 'Ginger', 'Benedict', 'Venti', 'Goose', 'Nugget', 'Cash',
'Coco', 'Jed', 'Sebastian', 'Walter', 'Sierra', 'Monkey', 'Harry',
'Kody', 'Lassie', 'Rover', 'Napolean', 'Dawn', 'Boomer', 'Cody',
'Rumble', 'Clifford', 'Dewey', 'Scout', 'Gizmo', 'Cooper', 'Harold',
'Shikha', 'Jamesy', 'Lili', 'Sammy', 'Meatball', 'Paisley', 'Albus',
'Neptune', 'Quinn', 'Belle', 'Zooey', 'Dave', 'Jersey', 'Hobbes',
'Burt', 'Lorenzo', 'Carl', 'Jordy', 'Milky', 'Trooper', 'Winston',
'Sophie', 'Wyatt', 'Rosie', 'Thor', 'Oscar', 'Luna', 'Callie',
'Cermet', 'George', 'Marlee', 'Arya', 'Einstein', 'Alice',
'Rumpole', 'Benny', 'Aspen', 'Jarod', 'Wiggles', 'General',
'Sailor', 'Astrid', 'Iggy', 'Snoop', 'Kyle', 'Leo', 'Riley',
'Gidget', 'Noosh', 'Odin', 'Jerry', 'Charlie', 'Georgie', 'Rontu',
'Cannon', 'Furzey', 'Daisy', 'Tuck', 'Barney', 'Vixen', 'Jarvis',
'Mimosa', 'Pickles', 'Bungalo', 'Brady', 'Margo', 'Sadie', 'Hank',
'Tycho', 'Stephan', 'Indie', 'Winnie', 'Bentley', 'Ken', 'Max',
'Maddie', 'Pipsy', 'Monty', 'Sojourner', 'Odie', 'Arlo', 'Sunny',
'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera', 'Buddy', 'Ava',
'Rory', 'Eli', 'Ash', 'Tucker', 'Tobi', 'Chester', 'Wilson',
'Sunshine', 'Lipton', 'Gabby', 'Bronte', 'Poppy', 'Rhino', 'Willow',
'Orion', 'Eevee', 'Smiley', 'Logan', 'Moreton', 'Klein', 'Miguel',
'Emanuel', 'Kuyu', 'Dutch', 'Pete', 'Scooter', 'Reggie', 'Kyro',
'Samson', 'Loki', 'Mia', 'Malcolm', 'Dexter', 'Alfie', 'Fiona',
'Mutt', 'Bear', 'Doobert', 'Beebop', 'Alexander', 'Sailer',
'Brutus', 'Kona', 'Boots', 'Ralphie', 'Phil', 'Cupid', 'Pawnd',
'Pilot', 'Ike', 'Mo', 'Toby', 'Sweet', 'Pablo', 'Nala', 'Balto',
'Crawford', 'Gabe', 'Mattie', 'Jimison', 'Hercules', 'Duchess',
'Harlso', 'Sampson', 'Sundance', 'Luca', 'Flash', 'Finn', 'Peaches',
'Howie', 'Jazzy', 'Anna', 'Bo', 'Seamus', 'Wafer', 'Chelsea', 'Tom',
'Moose', 'Florence', 'Autumn', 'Dido', 'Eugene', 'Herschel',
'Strudel', 'Tebow', 'Chloe', 'Betty', 'Timber', 'Binky', 'Dudley',
'Comet', 'Larry', 'Levi', 'Akumi', 'Titan', 'Olivia', 'Alf',
'Oshie', 'Bruce', 'Chubbs', 'Sky', 'Atlas', 'Eleanor', 'Layla',
'Rocky', 'Baron', 'Tyr', 'Bauer', 'Swagger', 'Brandi', 'Mary',
'Moe', 'Halo', 'Augie', 'Craig', 'Sam', 'Hunter', 'Pavlov',
'Maximus', 'Wallace', 'Ito', 'Milo', 'Ollie', 'Cali', 'Lennon',
'Major', 'Duke', 'Reginald', 'Sansa', 'Shooter', 'Django', 'Diogi',
'Sonny', 'Philbert', 'Marley', 'Severus', 'Ronnie', 'Anakin',
'Bones', 'Mauve', 'Chef', 'Doc', 'Sobe', 'Longfellow', 'Mister',
'Iroh', 'Baloo', 'Stubert', 'Paull', 'Timison', 'Davey', 'Pancake',
'Tyrone', 'Snicku', 'Ruby', 'Brody', 'Rizzy', 'Mack', 'Butter',
'Nimbus', 'Laika', 'Dobby', 'Juno', 'Maude', 'Lily', 'Newt',
'Benji', 'Nida', 'Robin', 'Monster', 'BeBe', 'Remus', 'Mabel',
'Misty', 'Happy', 'Mosby', 'Maggie', 'Leela', 'Ralphy', 'Brownie',
'Meyer', 'Stella', 'Frank', 'Tonks', 'Lincoln', 'Oakley', 'Dale',
'Rizzo', 'Arnie', 'Pinot', 'Dallas', 'Hero', 'Frankie', 'Stormy',
'Mairi', 'Loomis', 'Godi', 'Kenny', 'Deacon', 'Timmy', 'Harper',
'Chipson', 'Combo', 'Dash', 'Bell', 'Hurley', 'Jay', 'Mya',

'Strider', 'Wesley', 'Solomon', 'Huck', 'O', 'Blue', 'Finley',
 'Sprinkles', 'Heinrich', 'Shakespeare', 'Fizz', 'Chip', 'Grey',
 'Roosevelt', 'Gromit', 'Willem', 'Dakota', 'Dixie', 'Al', 'Jackson',
 'Carbon', 'DonDon', 'Kirby', 'Lou', 'Nollie', 'Chevy', 'Tito',
 'Louie', 'Rupert', 'Rufus', 'Brudge', 'Shadoe', 'Colby', 'Angel',
 'Brat', 'Tove', 'Aubie', 'Kota', 'Eve', 'Glenn', 'Shelby', 'Sephie',
 'Bonaparte', 'Albert', 'Wishes', 'Rose', 'Theo', 'Rocco', 'Fido',
 'Emma', 'Spencer', 'Lilli', 'Boston', 'Brandonald', 'Corey',
 'Leonard', 'Chompsky', 'Beckham', 'Devón', 'Gert', 'Watson',
 'Rubio', 'Keith', 'Dex', 'Carly', 'Ace', 'Tayzie', 'Grizzie',
 'Fred', 'Gilbert', 'Zoe', 'Stewie', 'Calvin', 'Lilah', 'Spanky',
 'Jameson', 'Piper', 'Atticus', 'Blu', 'Dietrich', 'Divine', 'Tripp',
 'Cora', 'Huxley', 'Keurig', 'Bookstore', 'Linus', 'Abby', 'Shaggy',
 'Shiloh', 'Gustav', 'Arlen', 'Percy', 'Lenox', 'Sugar', 'Harvey',
 'Blanket', 'Geno', 'Stark', 'Beya', 'Kilo', 'Kayla', 'Maxaroni',
 'Doug', 'Edmund', 'Aqua', 'Theodore', 'Chase', 'Rorie', 'Simba',
 'Charles', 'Bayley', 'Axel', 'Storkson', 'Remy', 'Chadrick',
 'Kellogg', 'Buckley', 'Livvie', 'Terry', 'Hermione', 'Ralpher',
 'Aldrick', 'Rooney', 'Crystal', 'Ziva', 'Stefan', 'Pupcasso',
 'Puff', 'Flurpson', 'Coleman', 'Enchilada', 'Raymond', 'Rueben',
 'Cilantro', 'Karll', 'Sprout', 'Blitz', 'Bloop', 'Lillie',
 'Ashleigh', 'Kreggory', 'Sarge', 'Luther', 'Ivar', 'Jangle',
 'Schnitzel', 'Panda', 'Berkeley', 'Ralphé', 'Charleson', 'Clyde',
 'Harnold', 'Sid', 'Pippa', 'Otis', 'Carper', 'Bowie',
 'Alexanderson', 'Suki', 'Barclay', 'Skittle', 'Ebby', 'Flávio',
 'Smokey', 'Link', 'Jennifur', 'Ozzy', 'Bluebert', 'Stephanus',
 'Bubbles', 'Zeus', 'Bertson', 'Nico', 'Michelangelo', 'Siba',
 'Calbert', 'Curtis', 'Travis', 'Thumas', 'Kanu', 'Lance', 'Opie',
 'Kane', 'Olive', 'Chuckles', 'Staniel', 'Sora', 'Beemo', 'Gunner',
 'Lacy', 'Tater', 'Olaf', 'Cecil', 'Vince', 'Karma', 'Billy',
 'Walker', 'Rodney', 'Klevin', 'Malikai', 'Bobble', 'River',
 'Jebberson', 'Remington', 'Farfle', 'Jiminus', 'Clarkus',
 'Finnegus', 'Cupcake', 'Kathmandu', 'Ellie', 'Katie', 'Kara',
 'Adele', 'Zara', 'Ambrose', 'Jimothy', 'Bode', 'Terrenth', 'Reese',
 'Chesterson', 'Lucia', 'Bisquick', 'Ralphson', 'Socks', 'Rambo',
 'Rudy', 'Fiji', 'Rilo', 'Bilbo', 'Coopson', 'Yoda', 'Millie',
 'Chet', 'Crouton', 'Daniel', 'Kaia', 'Murphy', 'Dotsy', 'Eazy',
 'Coops', 'Fillup', 'Miley', 'Charl', 'Reagan', 'Yukon', 'CeCe',
 'Cuddles', 'Claude', 'Jessiga', 'Carter', 'Ole', 'Pherb', 'Blipson',
 'Reptar', 'Trevith', 'Berb', 'Bob', 'Colin', 'Brian', 'Olivier',
 'Grady', 'Kobe', 'Freddery', 'Bodie', 'Dunkin', 'Wally', 'Tupawc',
 'Amber', 'Edgar', 'Teddy', 'Kingsley', 'Brockly', 'Richie', 'Molly',
 'Vinscent', 'Cedrick', 'Hazel', 'Lolo', 'Eriq', 'Phred', 'Oddie',
 'Maxwell', 'Geoff', 'Covach', 'Durg', 'Fynn', 'Ricky', 'Herald',
 'Lucky', 'Ferg', 'Trip', 'Clarence', 'Hamrick', 'Brad', 'Pubert',
 'Frönq', 'Derby', 'Lizzie', 'Ember', 'Blakely', 'Opal', 'Marq',
 'Kramer', 'Barry', 'Gordon', 'Baxter', 'Mona', 'Horace', 'Crimson',
 'Birf', 'Hammond', 'Lorelei', 'Marty', 'Brooks', 'Petrick',

```
'Hubertson', 'Gerbald', 'Oreo', 'Bruiser', 'Perry', 'Bobby', 'Jeph',
'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe', 'Tiger', 'Jiminy',
'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Darrel', 'Taco', 'Joey',
'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses', 'Toffee', 'Apollo',
'Asher', 'Glacier', 'Chuck', 'Champ', 'Ozzie', 'Griswold', 'Cheesy',
'Moofasa', 'Hector', 'Goliath', 'Kawhi', 'Emmie', 'Penelope',
'Willie', 'Rinna', 'Mike', 'William', 'Dwight', 'Evy', 'Rascal',
'Linda', 'Tug', 'Tango', 'Grizz', 'Jerome', 'Crumpet', 'Jessifer',
'Izzy', 'Ralph', 'Sandy', 'Humphrey', 'Tassy', 'Juckson', 'Chuq',
'Tyrus', 'Karl', 'Godzilla', 'Vinnie', 'Kenneth', 'Herm', 'Bert',
'Striker', 'Donny', 'Pepper', 'Bernie', 'Buddah', 'Lenny', 'Arnold',
'Zuzu', 'Mollie', 'Laela', 'Tedders', 'Superpup', 'Rufio', 'Jeb',
'Rodman', 'Jonah', 'Chesney', 'Henry', 'Bobbay', 'Mitch', 'Kaiya',
'Acro', 'Aiden', 'Obie', 'Dot', 'Shnuggles', 'Kendall', 'Jeffri',
'Steve', 'Mac', 'Fletcher', 'Kenzie', 'Pumpkin', 'Schnozz',
'Gustaf', 'Cheryl', 'Ed', 'Leonidas', 'Norman', 'Caryl', 'Scott',
'Taz', 'Darby', 'Jackie', 'Jazz', 'Franq', 'Pippin', 'Rolf',
'Snickers', 'Ridley', 'Cal', 'Bradley', 'Bubba', 'Tuco', 'Patch',
'Mojo', 'Batdog', 'Dylan', 'Mark', 'JD', 'Alejandro', 'Scruffers',
'Pip', 'Julius', 'Tanner', 'Sparky', 'Anthony', 'Holly', 'Jett',
'Amy', 'Sage', 'Andy', 'Mason', 'Trigger', 'Antony', 'Creg',
'Traviss', 'Gin', 'Jeffrie', 'Danny', 'Ester', 'Pluto', 'Bloo',
'Edd', 'Willy', 'Herb', 'Damon', 'Peanut', 'Nigel', 'Butters',
'Sandra', 'Fabio', 'Randall', 'Liam', 'Tommy', 'Ben', 'Raphael',
'Julio', 'Andru', 'Kloey', 'Shawwn', 'Skye', 'Kollin', 'Ronduh',
'Billl', 'Saydee', 'Dug', 'Tessa', 'Sully', 'Kirk', 'Ralf', 'Clarq',
'Jaspers', 'Samsom', 'Harrison', 'Chaz', 'Jeremy', 'Jaycob',
'Lambeau', 'Ruffles', 'Amélie', 'Bobb', 'Banditt', 'Kevon',
'Winifred', 'Hanz', 'Churlie', 'Zeek', 'Timofy', 'Maks', 'Jomathan',
'Kallie', 'Marvin', 'Spark', 'Gördón', 'Jo', 'DayZ', 'Jareld',
'Torque', 'Ron', 'Skittles', 'Cleopatra', 'Erik', 'Stu',
'Tedrick', 'Filup', 'Kial', 'Naphaniel', 'Dook', 'Hall', 'Philippe',
'Biden', 'Fwed', 'Genevieve', 'Joshwa', 'Bradlay', 'Clybe', 'Keet',
'Carll', 'Jockson', 'Josep', 'Lugan', 'Christoper'], dtype=object)
```

3.4 Define

- Assign 10 to rating_denominator column value

3.4.1 Code

```
In [41]: twitter_archive_copy.rating_denominator= twitter_archive_copy['rating_denominator']=10
```

```
In [42]: twitter_archive_copy.rating_denominator.unique()
```

```
Out[42]: array([10])
```

3.5 Define

- Convert timestamp columns to date

3.6 Code

```
In [43]: twitter_archive_copy['timestamp'] = pd.to_datetime(twitter_archive_copy['timestamp'])
         twitter_archive_copy['retweeted_status_timestamp'] = pd.to_datetime(twitter_archive_copy['retweeted_status_timestamp'])
```

3.7 Test

```
In [44]: twitter_archive_copy.head()
```

```
Out[44]:
```

	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	source	\
0	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone" r...	
1	2017-08-01 00:17:27	<a href="http://twitter.com/download/iphone" r...	
2	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...	
3	2017-07-30 15:58:51	<a href="http://twitter.com/download/iphone" r...	
4	2017-07-29 16:00:24	<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	NaT	
1	NaN	NaT	
2	NaN	NaT	
3	NaN	NaT	
4	NaN	NaT	

	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

3.8 Define

- Create the "dog_stage" column by combining columns "doggo", "floofer", "pupper" and "puppo"

3.9 Code

```
In [45]: def dog_stage(row):
        stage = []
        if row['doggo'] == 'doggo':
            stage.append('doggo')
        if row['floofer'] == 'floofer':
            stage.append('floofer')
        if row['pupper'] == 'pupper':
            stage.append('pupper')
        if row['puppo'] == 'puppo':
            stage.append('puppo')

        if not stage:
            return "None"
        else:
            return ','.join(stage)

        twitter_archive_copy['dog_stage'] = twitter_archive_copy.apply(lambda row: dog_stage(row)

In [46]: # drop the columns
        twitter_archive_copy.drop(["doggo", "floofer", "pupper", "puppo"], axis=1, inplace=True)
```

3.10 Test

```
In [47]: twitter_archive_copy.head(4)
```

```
Out[47]:
```

	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	

	timestamp	source	\
0	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone" r...	
1	2017-08-01 00:17:27	<a href="http://twitter.com/download/iphone" r...	
2	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...	
3	2017-07-30 15:58:51	<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

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

	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

	rating_denominator	name	dog_stage
0	10	Phineas	None
1	10	Tilly	None
2	10	Archie	None
3	10	Darla	None

```
In [48]: pd.value_counts(twitter_archive_copy['dog_stage'])
```

```
Out[48]: None          1976
pupper          245
doggo           83
puppo           29
doggo,pupper    12
floofer         9
doggo,floofer   1
doggo,puppo     1
Name: dog_stage, dtype: int64
```

3.10.1 Define

- remove all rows that have values (not blank or non-null) in `retweeted_status_id`, `retweeted_status_user_id`, and `retweeted_status_timestamp` columns

3.10.2 Code

```
In [49]: twitter_archive_copy = twitter_archive_copy[twitter_archive_copy['in_reply_to_status_id']
twitter_archive_copy = twitter_archive_copy[twitter_archive_copy['retweeted_status_id']

columns_to_be_dropped = ["in_reply_to_status_id", "in_reply_to_user_id", "retweeted_sta
```

```
twitter_archive_copy.drop(columns_to_be_dropped, axis = 1, inplace=True)
columns_to_be_dropped = []
```

3.10.3 Test

```
In [50]: twitter_archive_copy
```

```
Out[50]:
```

	tweet_id	timestamp \
0	892420643555336193	2017-08-01 16:23:56
1	892177421306343426	2017-08-01 00:17:27
2	891815181378084864	2017-07-31 00:18:03
3	891689557279858688	2017-07-30 15:58:51
4	891327558926688256	2017-07-29 16:00:24
5	891087950875897856	2017-07-29 00:08:17
6	890971913173991426	2017-07-28 16:27:12
7	890729181411237888	2017-07-28 00:22:40
8	890609185150312448	2017-07-27 16:25:51
9	890240255349198849	2017-07-26 15:59:51
10	890006608113172480	2017-07-26 00:31:25
11	889880896479866881	2017-07-25 16:11:53
12	889665388333682689	2017-07-25 01:55:32
13	889638837579907072	2017-07-25 00:10:02
14	889531135344209921	2017-07-24 17:02:04
15	889278841981685760	2017-07-24 00:19:32
16	888917238123831296	2017-07-23 00:22:39
17	888804989199671297	2017-07-22 16:56:37
18	888554962724278272	2017-07-22 00:23:06
20	888078434458587136	2017-07-20 16:49:33
21	887705289381826560	2017-07-19 16:06:48
22	887517139158093824	2017-07-19 03:39:09
23	887473957103951883	2017-07-19 00:47:34
24	887343217045368832	2017-07-18 16:08:03
25	887101392804085760	2017-07-18 00:07:08
26	886983233522544640	2017-07-17 16:17:36
27	886736880519319552	2017-07-16 23:58:41
28	886680336477933568	2017-07-16 20:14:00
29	886366144734445568	2017-07-15 23:25:31
31	886258384151887873	2017-07-15 16:17:19
...
2326	666411507551481857	2015-11-17 00:24:19
2327	666407126856765440	2015-11-17 00:06:54
2328	666396247373291520	2015-11-16 23:23:41
2329	666373753744588802	2015-11-16 21:54:18
2330	666362758909284353	2015-11-16 21:10:36
2331	666353288456101888	2015-11-16 20:32:58
2332	666345417576210432	2015-11-16 20:01:42
2333	666337882303524864	2015-11-16 19:31:45
2334	666293911632134144	2015-11-16 16:37:02

2335 666287406224695296 2015-11-16 16:11:11
 2336 666273097616637952 2015-11-16 15:14:19
 2337 666268910803644416 2015-11-16 14:57:41
 2338 666104133288665088 2015-11-16 04:02:55
 2339 666102155909144576 2015-11-16 03:55:04
 2340 666099513787052032 2015-11-16 03:44:34
 2341 666094000022159362 2015-11-16 03:22:39
 2342 666082916733198337 2015-11-16 02:38:37
 2343 666073100786774016 2015-11-16 01:59:36
 2344 666071193221509120 2015-11-16 01:52:02
 2345 666063827256086533 2015-11-16 01:22:45
 2346 666058600524156928 2015-11-16 01:01:59
 2347 666057090499244032 2015-11-16 00:55:59
 2348 666055525042405380 2015-11-16 00:49:46
 2349 666051853826850816 2015-11-16 00:35:11
 2350 666050758794694657 2015-11-16 00:30:50
 2351 666049248165822465 2015-11-16 00:24:50
 2352 666044226329800704 2015-11-16 00:04:52
 2353 666033412701032449 2015-11-15 23:21:54
 2354 666029285002620928 2015-11-15 23:05:30
 2355 666020888022790149 2015-11-15 22:32:08

```

                                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...
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...
31 <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 \

```

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...
5 Here we have a majestic great white breaching ...
6 Meet Jax. He enjoys ice cream so much he gets ...
7 When you watch your owner call another dog a g...
8 This is Zoey. She doesn't want to be one of th...
9 This is Cassie. She is a college pup. Studying...

```

10 This is Koda. He is a South Australian decksha...
 11 This is Bruno. He is a service shark. Only get...
 12 Here's a puppo that seems to be on the fence a...
 13 This is Ted. He does his best. Sometimes that'...
 14 This is Stuart. He's sporting his favorite fan...
 15 This is Oliver. You're witnessing one of his m...
 16 This is Jim. He found a fren. Taught him how t...
 17 This is Zeke. He has a new stick. Very proud o...
 18 This is Ralphus. He's powering up. Attempting ...
 20 This is Gerald. He was just told he didn't get...
 21 This is Jeffrey. He has a monopoly on the pool...
 22 I've yet to rate a Venezuelan Hover Wiener. Th...
 23 This is Canela. She attempted some fancy porch...
 24 You may not have known you needed to see this ...
 25 This... is a Jubilant Antarctic House Bear. We...
 26 This is Maya. She's very shy. Rarely leaves he...
 27 This is Mingus. He's a wonderful father to his...
 28 This is Derek. He's late for a dog meeting. 13...
 29 This is Roscoe. Another pupper fallen victim t...
 31 This is Waffles. His doggles are pupside down...
 ...
 2326 This is quite the dog. Gets really excited whe...
 2327 This is a southern Vesuvius bumblegruff. Can d...
 2328 Oh goodness. A super rare northeast Qdoba kang...
 2329 Those are sunglasses and a jean jacket. 11/10 ...
 2330 Unique dog here. Very small. Lives in containe...
 2331 Here we have a mixed Asiago from the Galápagos...
 2332 Look at this jokester thinking seat belt laws ...
 2333 This is an extremely rare horned Parthenon. No...
 2334 This is a funny dog. Weird toes. Won't come do...
 2335 This is an Albanian 3 1/2 legged Episcopalian...
 2336 Can take selfies 11/10 <https://t.co/ws2AMaNwPW>
 2337 Very concerned about fellow dog trapped in com...
 2338 Not familiar with this breed. No tail (weird)...
 2339 Oh my. Here you are seeing an Adobe Setter giv...
 2340 Can stand on stump for what seems like a while...
 2341 This appears to be a Mongolian Presbyterian mi...
 2342 Here we have a well-established sunblockerspan...
 2343 Let's hope this flight isn't Malaysian (lol). ...
 2344 Here we have a northern speckled Rhododendron...
 2345 This is the happiest dog you will ever see. Ve...
 2346 Here is the Rand Paul of retrievers folks! He'...
 2347 My oh my. This is a rare blond Canadian terrie...
 2348 Here is a Siberian heavily armored polar bear ...
 2349 This is an odd dog. Hard on the outside but lo...
 2350 This is a truly beautiful English Wilson Staff...
 2351 Here we have a 1949 1st generation vulpix. Enj...
 2352 This is a purebred Piers Morgan. Loves to Netf...

2353 Here is a very happy pup. Big fan of well-main...
 2354 This is a western brown Mitsubishi terrier. Up...
 2355 Here we have a Japanese Irish Setter. Lost eye...

	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
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
31	https://twitter.com/dog_rates/status/886258384...	13
...
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	dog_stage
0	10	Phineas	None
1	10	Tilly	None
2	10	Archie	None
3	10	Darla	None
4	10	Franklin	None
5	10	None	None
6	10	Jax	None
7	10	None	None
8	10	Zoey	None
9	10	Cassie	doggo
10	10	Koda	None
11	10	Bruno	None
12	10	None	puppo
13	10	Ted	None
14	10	Stuart	puppo
15	10	Oliver	None
16	10	Jim	None
17	10	Zeke	None
18	10	Ralphus	None
20	10	Gerald	None
21	10	Jeffrey	None
22	10	None	None
23	10	Canela	None
24	10	None	None
25	10	None	None
26	10	Maya	None
27	10	Mingus	None
28	10	Derek	None

29	10	Roscoe	pupper
31	10	Waffles	None
...
2326	10	None	None
2327	10	None	None
2328	10	None	None
2329	10	None	None
2330	10	None	None
2331	10	None	None
2332	10	None	None
2333	10	None	None
2334	10	None	None
2335	10	None	None
2336	10	None	None
2337	10	None	None
2338	10	None	None
2339	10	None	None
2340	10	None	None
2341	10	None	None
2342	10	None	None
2343	10	None	None
2344	10	None	None
2345	10	None	None
2346	10	None	None
2347	10	None	None
2348	10	None	None
2349	10	None	None
2350	10	None	None
2351	10	None	None
2352	10	None	None
2353	10	None	None
2354	10	None	None
2355	10	None	None

[2097 rows x 9 columns]

4 Image Predictions

4.1 Define

- Capitalize the first letter of names of p1,p2,p3 columns

4.2 Code

```
In [51]: image_prediction1.p1= image_prediction1.p1.str.capitalize()
         image_prediction1.p2= image_prediction1.p2.str.capitalize()
         image_prediction1.p3= image_prediction1.p3.str.capitalize()
```

```
In [52]: image_prediction1.head(3)
```

```
Out[52]:
```

	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	

	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

	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

```
In [53]: image_predictions.head(3)
```

```
Out[53]:
```

	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	

	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

	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

4.2.1 Define

- Replace _ and - with white space in p1,p2 and p3 columns

```
In [54]: image_prediction1.p1= image_prediction1.p1.str.replace('_', ' ')
image_prediction1.p2= image_prediction1.p2.str.replace('_', ' ')
image_prediction1.p3= image_prediction1.p3.str.replace('_', ' ')

image_prediction1.p1= image_prediction1.p1.str.replace('-', ' ')
image_prediction1.p2= image_prediction1.p2.str.replace('-', ' ')
image_prediction1.p3= image_prediction1.p3.str.replace('-', ' ')

```

4.3 Code

```
In [55]: image_prediction1.head(10)
```

```

Out [55]:          tweet_id          jpg_url \
0  666020888022790149  https://pbs.twimg.com/media/CT4udnOWwAAOaMy.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

          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
5           1  Bernese mountain dog  0.651137    True  English springer
6           1      Box turtle  0.933012   False    Mud turtle
7           1           Chow  0.692517    True  Tibetan mastiff
8           1  Shopping cart  0.962465   False  Shopping basket
9           1  Miniature poodle  0.201493    True    Komondor

          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
5  0.263788    True  Greater swiss mountain dog  0.016199    True
6  0.045885   False           Terrapin  0.017885   False
7  0.058279    True           Fur coat  0.054449   False
8  0.014594   False    Golden retriever  0.007959    True
9  0.192305    True  Soft coated wheaten terrier  0.082086    True

```

4.4 JSON file : Tweet Retweet and Favorite Counts ¶

4.4.1 Define

- set tweet_id to string with astype()

4.4.2 Code

```
In [56]: df_tweet1['tweet_id'] = df_tweet1['tweet_id'].astype('str')
```

4.5 Test

```
In [57]: df_tweet1.head()
```



```
Out [57]:
```

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

4.6 Define

- Setting the column in order

4.7 Code

```
In [58]: df_tweet1 = df_tweet1[['tweet_id', 'favorite_count', 'retweet_count']]
```

4.8 Test

```
In [59]: df_tweet1.head()
```

```
Out [59]:
```

	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

4.9 Define

- Renaming the column names. WHY? Because we have to merge two tables on ground of some key whose name must be common.

4.10 Code

```
In [60]: df_tweet1.rename(columns={'favorite_count': 'likes', 'retweet_count': 'retweets'}, inplace=True)
```

4.11 Test

```
In [61]: df_tweet1.head()
```

```
Out [61]:
```

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

```
In [62]: df_tweet1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 3 columns):
tweet_id    2354 non-null object
likes       2354 non-null int64
retweets    2354 non-null int64
dtypes: int64(2), object(1)
memory usage: 55.2+ KB
```

```
In [63]: df_tweet1.describe()
```

```
Out[63]:
```

	likes	retweets
count	2354.000000	2354.000000
mean	8080.968564	3164.797366
std	11814.771334	5284.770364
min	0.000000	0.000000
25%	1415.000000	624.500000
50%	3603.500000	1473.500000
75%	10122.250000	3652.000000
max	132810.000000	79515.000000

4.12 Define

- change tweet_id from number to string for all three data sets because to merge them we need consistency. Code:

4.13 Code

```
In [64]: # changing format type from number to string on all three dataframe

twitter_archive_copy['tweet_id'] = twitter_archive_copy['tweet_id'].astype('str')
image_prediction1['tweet_id'] = image_prediction1['tweet_id'].astype('str')
df_tweet1['tweet_id'] = df_tweet1['tweet_id'].astype('str')
```

4.14 Test

```
In [65]: # testing if tweet_id format changed
type(twitter_archive_copy['tweet_id'].iloc[0])
type(image_prediction1['tweet_id'].iloc[0])
type(df_tweet1['tweet_id'].iloc[0])
```

```
Out[65]: str
```

5 COMBINING DATASETS

5.1 Define

- Combine df_json1 and image_prediction1 to twitter_arc_copy using merge()

5.2 Code

```
In [66]: df_merge = pd.merge(twitter_archive_copy, image_prediction1, on=['tweet_id'], how='inner')
df_merge = pd.merge(df_merge, df_tweet1, on = 'tweet_id', how = 'inner' )
```

```
In [67]: df_merge.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1971 entries, 0 to 1970
Data columns (total 22 columns):
tweet_id          1971 non-null object
timestamp         1971 non-null datetime64[ns]
source            1971 non-null object
text              1971 non-null object
expanded_urls     1971 non-null object
rating_numerator  1971 non-null int64
rating_denominator 1971 non-null int64
name              1971 non-null object
dog_stage         1971 non-null object
jpg_url           1971 non-null object
img_num           1971 non-null int64
p1                1971 non-null object
p1_conf           1971 non-null float64
p1_dog            1971 non-null bool
p2                1971 non-null object
p2_conf           1971 non-null float64
p2_dog            1971 non-null bool
p3                1971 non-null object
p3_conf           1971 non-null float64
p3_dog            1971 non-null bool
likes             1971 non-null int64
retweets          1971 non-null int64
dtypes: bool(3), datetime64[ns](1), float64(3), int64(5), object(10)
memory usage: 313.7+ KB
```

```
In [68]: df_merge.head(15)
```

```
Out[68]:
```

	tweet_id	timestamp	\
0	892420643555336193	2017-08-01 16:23:56	
1	892177421306343426	2017-08-01 00:17:27	
2	891815181378084864	2017-07-31 00:18:03	
3	891689557279858688	2017-07-30 15:58:51	
4	891327558926688256	2017-07-29 16:00:24	
5	891087950875897856	2017-07-29 00:08:17	
6	890971913173991426	2017-07-28 16:27:12	
7	890729181411237888	2017-07-28 00:22:40	
8	890609185150312448	2017-07-27 16:25:51	
9	890240255349198849	2017-07-26 15:59:51	

10 890006608113172480 2017-07-26 00:31:25
 11 889880896479866881 2017-07-25 16:11:53
 12 889665388333682689 2017-07-25 01:55:32
 13 889638837579907072 2017-07-25 00:10:02
 14 889531135344209921 2017-07-24 17:02:04

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...

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...
 5 Here we have a majestic great white breaching ...
 6 Meet Jax. He enjoys ice cream so much he gets ...
 7 When you watch your owner call another dog a g...
 8 This is Zoey. She doesn't want to be one of th...
 9 This is Cassie. She is a college pup. Studying...
 10 This is Koda. He is a South Australian decksha...
 11 This is Bruno. He is a service shark. Only get...
 12 Here's a puppo that seems to be on the fence a...
 13 This is Ted. He does his best. Sometimes that'...
 14 This is Stuart. He's sporting his favorite fan...

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

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

	jpg_url	...	p1_conf	\
0	https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg	...	0.097049	
1	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg	...	0.323581	
2	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg	...	0.716012	
3	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg	...	0.170278	
4	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg	...	0.555712	
5	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg	...	0.425595	
6	https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg	...	0.341703	
7	https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg	...	0.566142	
8	https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg	...	0.487574	
9	https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg	...	0.511319	
10	https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg	...	0.957979	
11	https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg	...	0.377417	
12	https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg	...	0.966327	
13	https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg	...	0.991650	
14	https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg	...	0.953442	

p1_dog	p2	p2_conf	p2_dog	p3	\
0	False	Bagel	0.085851	False	Banana
1	True	Pekinese	0.090647	True	Papillon
2	True	Malamute	0.078253	True	Kelpie
3	False	Labrador retriever	0.168086	True	Spatula

4	True	English springer	0.225770	True	German short haired pointer
5	True	Irish terrier	0.116317	True	Indian elephant
6	True	Border collie	0.199287	True	Ice lolly
7	True	Eskimo dog	0.178406	True	Pembroke
8	True	Irish setter	0.193054	True	Chesapeake bay retriever
9	True	Cardigan	0.451038	True	Chihuahua
10	True	Pomeranian	0.013884	True	Chow
11	True	Labrador retriever	0.151317	True	Muzzle
12	True	Cardigan	0.027356	True	Basenji
13	True	Boxer	0.002129	True	Staffordshire bullterrier
14	True	Labrador retriever	0.013834	True	Redbone

	p3_conf	p3_dog	likes	retweets
0	0.076110	False	39467	8853
1	0.068957	True	33819	6514
2	0.031379	True	25461	4328
3	0.040836	False	42908	8964
4	0.175219	True	41048	9774
5	0.076902	False	20562	3261
6	0.193548	False	12041	2158
7	0.076507	True	56848	16716
8	0.118184	True	28226	4429
9	0.029248	True	32467	7711
10	0.008167	True	31166	7624
11	0.082981	False	28268	5156
12	0.004633	True	38818	8538
13	0.001498	True	27672	4735
14	0.007958	True	15359	2321

[15 rows x 22 columns]

```
In [69]: # Calculate the value of 'rating'
df_merge['rating'] = df_merge['rating_numerator'] / df_merge['rating_denominator']
```

```
In [70]: # Check for presence of new column and correct calculation
df_merge.head(1)
```

```
Out[70]:
```

	tweet_id	timestamp	\	source	\	text	\	expanded_urls	rating_numerator	\
0	892420643555336193	2017-08-01 16:23:56		<a href="http://twitter.com/download/iphone" r...		This is Phineas. He's a mystical boy. Only eve...		https://twitter.com/dog_rates/status/892420643...	13	

```

      rating_denominator      name dog_stage \
0              10  Phineas      None

      jpg_url      ...      p1_dog      p2 \
0  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg      ...      False  Bagel

      p2_conf  p2_dog      p3  p3_conf  p3_dog  likes  retweets  rating
0  0.085851   False  Banana  0.07611   False  39467      8853      1.3

[1 rows x 23 columns]

```

```

In [72]: # Renaming some column names
df_merge.rename(columns={'rating_numerator': 'numerator',
                        'rating_denominator': 'denominator'}, inplace=True)

# Drop undesired columns to make dataframe easier to view
#df_merge.drop(['retweeted_status_id', 'retweeted_status_timestamp', 'retweeted_status_u

# Allow all columns and text to be shown
pd.set_option('display.max_columns', None)

```

```

In [73]: df_merge.head()

```

```

Out[73]:
      tweet_id      timestamp \
0  892420643555336193  2017-08-01 16:23:56
1  892177421306343426  2017-08-01 00:17:27
2  891815181378084864  2017-07-31 00:18:03
3  891689557279858688  2017-07-30 15:58:51
4  891327558926688256  2017-07-29 16:00:24

      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 \
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  numerator  denominator \
0  https://twitter.com/dog_rates/status/892420643...      13      10
1  https://twitter.com/dog_rates/status/892177421...      13      10

```

2	https://twitter.com/dog_rates/status/891815181...	12	10
3	https://twitter.com/dog_rates/status/891689557...	13	10
4	https://twitter.com/dog_rates/status/891327558...	12	10

	name	dog_stage	jpg_url \
0	Phineas	None	https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg
1	Tilly	None	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2	Archie	None	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
3	Darla	None	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
4	Franklin	None	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg

	img_num	p1	p1_conf	p1_dog	p2	p2_conf \
0	1	Orange	0.097049	False	Bagel	0.085851
1	1	Chihuahua	0.323581	True	Pekinese	0.090647
2	1	Chihuahua	0.716012	True	Malamute	0.078253
3	1	Paper towel	0.170278	False	Labrador retriever	0.168086
4	2	Basset	0.555712	True	English springer	0.225770

	p2_dog	p3	p3_conf	p3_dog	likes	retweets \
0	False	Banana	0.076110	False	39467	8853
1	True	Papillon	0.068957	True	33819	6514
2	True	Kelpie	0.031379	True	25461	4328
3	True	Spatula	0.040836	False	42908	8964
4	True	German short haired pointer	0.175219	True	41048	9774

	rating
0	1.3
1	1.3
2	1.2
3	1.3
4	1.2

In [74]: df_merge.tail()

```
Out[74]:
```

	tweet_id	timestamp \	source \
1966	666049248165822465	2015-11-16 00:24:50	<a href="http://twitter.com/download/iphone" r...
1967	666044226329800704	2015-11-16 00:04:52	<a href="http://twitter.com/download/iphone" r...
1968	666033412701032449	2015-11-15 23:21:54	<a href="http://twitter.com/download/iphone" r...
1969	666029285002620928	2015-11-15 23:05:30	<a href="http://twitter.com/download/iphone" r...
1970	666020888022790149	2015-11-15 22:32:08	<a href="http://twitter.com/download/iphone" r...

	text \
1966	Here we have a 1949 1st generation vulpix. Enj...
1967	This is a purebred Piers Morgan. Loves to Netf...
1968	Here is a very happy pup. Big fan of well-main...
1969	This is a western brown Mitsubishi terrier. Up...
1970	Here we have a Japanese Irish Setter. Lost eye...

	expanded_urls	numerator \
1966	https://twitter.com/dog_rates/status/666049248...	5
1967	https://twitter.com/dog_rates/status/666044226...	6
1968	https://twitter.com/dog_rates/status/666033412...	9
1969	https://twitter.com/dog_rates/status/666029285...	7
1970	https://twitter.com/dog_rates/status/666020888...	8

	denominator	name	dog_stage \
1966	10	None	None
1967	10	None	None
1968	10	None	None
1969	10	None	None
1970	10	None	None

	jpg_url	img_num \
1966	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg	1
1967	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg	1
1968	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg	1
1969	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	1
1970	https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg	1

	p1	p1_conf	p1_dog	p2	p2_conf \
1966	Miniature pinscher	0.560311	True	Rottweiler	0.243682
1967	Rhodesian ridgeback	0.408143	True	Redbone	0.360687
1968	German shepherd	0.596461	True	Malinois	0.138584
1969	Redbone	0.506826	True	Miniature pinscher	0.074192
1970	Welsh springer spaniel	0.465074	True	Collie	0.156665

	p2_dog	p3	p3_conf	p3_dog	likes	retweets	rating
1966	True	Doberman	0.154629	True	111	41	0.5
1967	True	Miniature pinscher	0.222752	True	311	147	0.6
1968	True	Bloodhound	0.116197	True	128	47	0.9
1969	True	Rhodesian ridgeback	0.072010	True	132	48	0.7
1970	True	Shetland sheepdog	0.061428	True	2535	532	0.8

5.3 STORING DATA

```
In [75]: df_merge.to_csv('twitter_archive_master_new.csv', encoding='utf-8', index=False)
          wrangled_df= pd.read_csv('twitter_archive_master_new.csv')
          wrangled_df.head()
```

```

Out[75]:
      tweet_id      timestamp \
0  892420643555336193  2017-08-01 16:23:56
1  892177421306343426  2017-08-01 00:17:27
2  891815181378084864  2017-07-31 00:18:03
3  891689557279858688  2017-07-30 15:58:51
4  891327558926688256  2017-07-29 16:00:24

      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 \
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  numerator  denominator \
0  https://twitter.com/dog_rates/status/892420643...      13      10
1  https://twitter.com/dog_rates/status/892177421...      13      10
2  https://twitter.com/dog_rates/status/891815181...      12      10
3  https://twitter.com/dog_rates/status/891689557...      13      10
4  https://twitter.com/dog_rates/status/891327558...      12      10

      name dog_stage      jpg_url \
0  Phineas      None  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg
1  Tilly      None  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2  Archie      None  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
3  Darla      None  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
4  Franklin      None  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg

      img_num      p1      p1_conf  p1_dog      p2      p2_conf \
0      1      Orange  0.097049  False      Bagel  0.085851
1      1      Chihuahua  0.323581  True      Pekinese  0.090647
2      1      Chihuahua  0.716012  True      Malamute  0.078253
3      1      Paper towel  0.170278  False  Labrador retriever  0.168086
4      2      Basset  0.555712  True      English springer  0.225770

      p2_dog      p3      p3_conf  p3_dog  likes  retweets \
0  False      Banana  0.076110  False  39467  8853
1  True      Papillon  0.068957  True  33819  6514
2  True      Kelpie  0.031379  True  25461  4328
3  True      Spatula  0.040836  False  42908  8964
4  True  German short haired pointer  0.175219  True  41048  9774

```

	rating
0	1.3
1	1.3
2	1.2
3	1.3
4	1.2

```
In [76]: wrangled_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1971 entries, 0 to 1970
Data columns (total 23 columns):
tweet_id      1971 non-null int64
timestamp     1971 non-null object
source        1971 non-null object
text          1971 non-null object
expanded_urls  1971 non-null object
numerator     1971 non-null int64
denominator   1971 non-null int64
name          1971 non-null object
dog_stage     1971 non-null object
jpg_url       1971 non-null object
img_num       1971 non-null int64
p1            1971 non-null object
p1_conf       1971 non-null float64
p1_dog        1971 non-null bool
p2            1971 non-null object
p2_conf       1971 non-null float64
p2_dog        1971 non-null bool
p3            1971 non-null object
p3_conf       1971 non-null float64
p3_dog        1971 non-null bool
likes         1971 non-null int64
retweets      1971 non-null int64
rating        1971 non-null float64
dtypes: bool(3), float64(4), int64(6), object(10)
memory usage: 313.8+ KB
```

5.4 Analyze and Visualize

```
In [77]: # this list of dogs for rating > 2
         wrangled_df.loc[wrangled_df['rating'] > 2]
```

```
Out[77]:
```

	tweet_id	timestamp	\
320	820690176645140481	2017-01-15 17:52:40	
382	810984652412424192	2016-12-19 23:06:23	
499	786709082849828864	2016-10-13 23:23:56	

549	778027034220126208	2016-09-20 00:24:34
658	758467244762497024	2016-07-28 01:00:57
722	749981277374128128	2016-07-04 15:00:45
842	731156023742988288	2016-05-13 16:15:54
918	716439118184652801	2016-04-03 01:36:11
939	713900603437621249	2016-03-27 01:29:02
963	710658690886586372	2016-03-18 02:46:49
981	709198395643068416	2016-03-14 02:04:08
1045	704054845121142784	2016-02-28 21:25:30
1120	697463031882764288	2016-02-10 16:51:59
1288	684222868335505415	2016-01-05 04:00:18
1359	680494726643068929	2015-12-25 21:06:00
1420	677716515794329600	2015-12-18 05:06:23
1478	675853064436391936	2015-12-13 01:41:41
1696	670842764863651840	2015-11-29 05:52:33

	source \
320	<a href="http://twitter.com/download/iphone" r...
382	<a href="http://twitter.com/download/iphone" r...
499	<a href="http://twitter.com/download/iphone" r...
549	<a href="http://twitter.com/download/iphone" r...
658	<a href="http://twitter.com/download/iphone" r...
722	<a href="https://about.twitter.com/products/tw...
842	<a href="http://twitter.com/download/iphone" r...
918	<a href="http://twitter.com/download/iphone" r...
939	<a href="http://twitter.com/download/iphone" r...
963	<a href="http://twitter.com/download/iphone" r...
981	<a href="http://twitter.com/download/iphone" r...
1045	<a href="http://twitter.com/download/iphone" r...
1120	<a href="http://twitter.com/download/iphone" r...
1288	<a href="http://twitter.com/download/iphone" r...
1359	<a href="http://twitter.com/download/iphone" r...
1420	<a href="http://twitter.com/download/iphone" r...
1478	<a href="http://twitter.com/download/iphone" r...
1696	<a href="http://twitter.com/download/iphone" r...

	text \
320	The floofs have been released I repeat the flo...
382	Meet Sam. She smiles 24/7 & secretly aspir...
499	This is Logan, the Chow who lived. He solemnly...
549	This is Sophie. She's a Jubilant Bush Pupper. ...
658	Why does this never happen at my front door...
722	This is Atticus. He's quite simply America af...
842	Say hello to this unbelievably well behaved sq...
918	This is Bluebert. He just saw that both #Final...
939	Happy Saturday here's 9 puppies on a bench. 99...
963	Here's a brigade of puppies. All look very pre...
981	From left to right:\nCletus, Jerome, Alejandro...

1045 Here is a whole flock of puppers. 60/50 I'll ...
 1120 Happy Wednesday here's a bucket of pups. 44/40...
 1288 Someone help the girl is being mugged. Several...
 1359 Here we have uncovered an entire battalion of ...
 1420 IT'S PUPPERGEDDON. Total of 144/120 ...I think...
 1478 Here we have an entire platoon of puppers. Tot...
 1696 After so many requests... here you go.\n\nGood...

	expanded_urls	numerator	\
320	https://twitter.com/dog_rates/status/820690176...	84	
382	https://www.gofundme.com/sams-smile , https://tw...	24	
499	https://twitter.com/dog_rates/status/786709082...	75	
549	https://twitter.com/dog_rates/status/778027034...	27	
658	https://twitter.com/dog_rates/status/758467244...	165	
722	https://twitter.com/dog_rates/status/749981277...	1776	
842	https://twitter.com/dog_rates/status/731156023...	204	
918	https://twitter.com/dog_rates/status/716439118...	50	
939	https://twitter.com/dog_rates/status/713900603...	99	
963	https://twitter.com/dog_rates/status/710658690...	80	
981	https://twitter.com/dog_rates/status/709198395...	45	
1045	https://twitter.com/dog_rates/status/704054845...	60	
1120	https://twitter.com/dog_rates/status/697463031...	44	
1288	https://twitter.com/dog_rates/status/684222868...	121	
1359	https://twitter.com/dog_rates/status/680494726...	26	
1420	https://twitter.com/dog_rates/status/677716515...	144	
1478	https://twitter.com/dog_rates/status/675853064...	88	
1696	https://twitter.com/dog_rates/status/670842764...	420	

	denominator	name	dog_stage	\
320	10	None	None	
382	10	Sam	None	
499	10	Logan	None	
549	10	Sophie	pupper	
658	10	None	None	
722	10	Atticus	None	
842	10	None	None	
918	10	Bluebert	None	
939	10	None	None	
963	10	None	None	
981	10	None	None	
1045	10	None	None	
1120	10	None	None	
1288	10	None	None	
1359	10	None	None	
1420	10	None	None	
1478	10	None	None	
1696	10	None	None	

	jpg_url	img_num	\
320	https://pbs.twimg.com/media/C20tWr0VQAEnS9r.jpg	2	
382	https://pbs.twimg.com/media/COEyPZbXAAAcSc.jpg	1	
499	https://pbs.twimg.com/media/CurzvFTXgAA2_AP.jpg	1	
549	https://pbs.twimg.com/media/Cswbc2yWcAAVsCJ.jpg	1	
658	https://pbs.twimg.com/ext_tw_video_thumb/75846...	1	
722	https://pbs.twimg.com/media/CmgBZ7kWcAAlzFD.jpg	1	
842	https://pbs.twimg.com/media/CiWWWhVNUYAAab_r.jpg	1	
918	https://pbs.twimg.com/media/CfFNk7cWAAA-hND.jpg	1	
939	https://pbs.twimg.com/media/CehIzzZWQAEyHH5.jpg	1	
963	https://pbs.twimg.com/media/CdzETn4W4AAVU5N.jpg	1	
981	https://pbs.twimg.com/media/CdeUKpcWoAAJAWJ.jpg	1	
1045	https://pbs.twimg.com/media/CcVOJEcXEAMOFHL.jpg	1	
1120	https://pbs.twimg.com/media/Ca3i7CzXIAMLhg8.jpg	1	
1288	https://pbs.twimg.com/media/CX7Y_ByWwAEJdUy.jpg	1	
1359	https://pbs.twimg.com/media/CXGaVxOWAADjhF.jpg	1	
1420	https://pbs.twimg.com/media/CWe7kw9W4AE8UJh.jpg	1	
1478	https://pbs.twimg.com/media/CWEcxqWVEAAHyGH.jpg	1	
1696	https://pbs.twimg.com/media/CU9P717W4AAOlKx.jpg	1	

	p1	p1_conf	p1_dog	p2	\
320	West highland white terrier	0.872064	True	Kuvasz	
382	Golden retriever	0.871342	True	Tibetan mastiff	
499	Pomeranian	0.467321	True	Persian cat	
549	Clumber	0.946718	True	Cocker spaniel	
658	Labrador retriever	0.436377	True	Chihuahua	
722	Bow tie	0.533941	False	Sunglasses	
842	Lakeside	0.501767	False	Breakwater	
918	Siberian husky	0.396495	True	Malamute	
939	Golden retriever	0.371816	True	Cocker spaniel	
963	Soft coated wheaten terrier	0.948617	True	Dandie dinmont	
981	Borzoi	0.490783	True	Wire haired fox terrier	
1045	Great pyrenees	0.667939	True	Kuvasz	
1120	Labrador retriever	0.999885	True	Golden retriever	
1288	Soft coated wheaten terrier	0.791182	True	Cocker spaniel	
1359	Kuvasz	0.438627	True	Samoyed	
1420	Teddy	0.662908	False	Crib	
1478	Labrador retriever	0.868367	True	Golden retriever	
1696	Microphone	0.096063	False	Accordion	

	p2_conf	p2_dog	p3	p3_conf	p3_dog	\
320	0.059526	True	Samoyed	0.037400	True	
382	0.036708	True	Labrador retriever	0.025823	True	
499	0.122978	False	Chow	0.102654	True	
549	0.015950	True	Lhasa	0.006519	True	
658	0.113956	True	American staffordshire terrier	0.099689	True	
722	0.080822	False	Sunglass	0.050776	False	
842	0.051351	False	King penguin	0.049444	False	

918	0.317053	True	Eskimo dog	0.273419	True
939	0.177413	True	Irish setter	0.092725	True
963	0.018664	True	Cairn	0.015943	True
981	0.083513	True	English setter	0.083184	True
1045	0.228764	True	Golden retriever	0.043885	True
1120	0.000098	True	Pug	0.000008	True
1288	0.072444	True	Teddy	0.071486	False
1359	0.111622	True	Great pyrenees	0.064061	True
1420	0.031891	False	Chow	0.025438	True
1478	0.043305	True	Vizsla	0.028207	True
1696	0.094075	False	Drumstick	0.061113	False

	likes	retweets	rating
320	13518	3716	8.4
382	5927	1655	2.4
499	20296	7069	7.5
549	7320	1885	2.7
658	5316	2539	16.5
722	5569	2772	177.6
842	4196	1434	20.4
918	2574	247	5.0
939	3062	829	9.9
963	2529	636	8.0
981	2634	721	4.5
1045	3201	1028	6.0
1120	3748	1552	4.4
1288	4225	1563	12.1
1359	1879	542	2.6
1420	3323	1104	14.4
1478	2927	1460	8.8
1696	7989	4324	42.0

5.5 DEFINE

- Create a function to remove links and apply it to get cleaner texts. Remove `http://` and followed links for better cloud words.

5.6 CODE

```
In [78]: def remove_link(x):
          http_pos = x.find("http")
          # If no link, retain row
          if http_pos == -1:
              x = x
          else:
              # Remove space before link to end
              x = x[:http_pos - 1]
```

```
        return x
wrangled_df.text = wrangled_df.text.apply(remove_link)
```

5.7 TEST

```
In [79]: for row in wrangled_df.text[:5]:
          print(row)
```

This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/10
This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she's available for
This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know what
This is Darla. She commenced a snooze mid meal. 13/10 happens to the best of us
This is Franklin. He would like you to stop calling him "cute." He is a very fierce shark and sh

```
In [80]: from PIL import Image
         from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
```

#Word Cloud of texts

```
text1 = " ".join(review for review in wrangled_df.text) #to collect all texts from column
# Create stopwords list:
stopwords = set(STOPWORDS)
stopwords.update(["puppy", "Bruno", "dog", "good", "boy", 'cool'])
```

```
# Generate a word cloud image
```

```
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generate(text1)
```

```
# Display the generated image:
```

the matplotlib way:

```
plt.imshow(wordcloud, interpolation='bilinear')
```

```
plt.axis("off")
```

```
plt.show()
```



5.7.1 Used stop words to give visualization of commonly occurring texts in reviews provided by WeRateDogs.

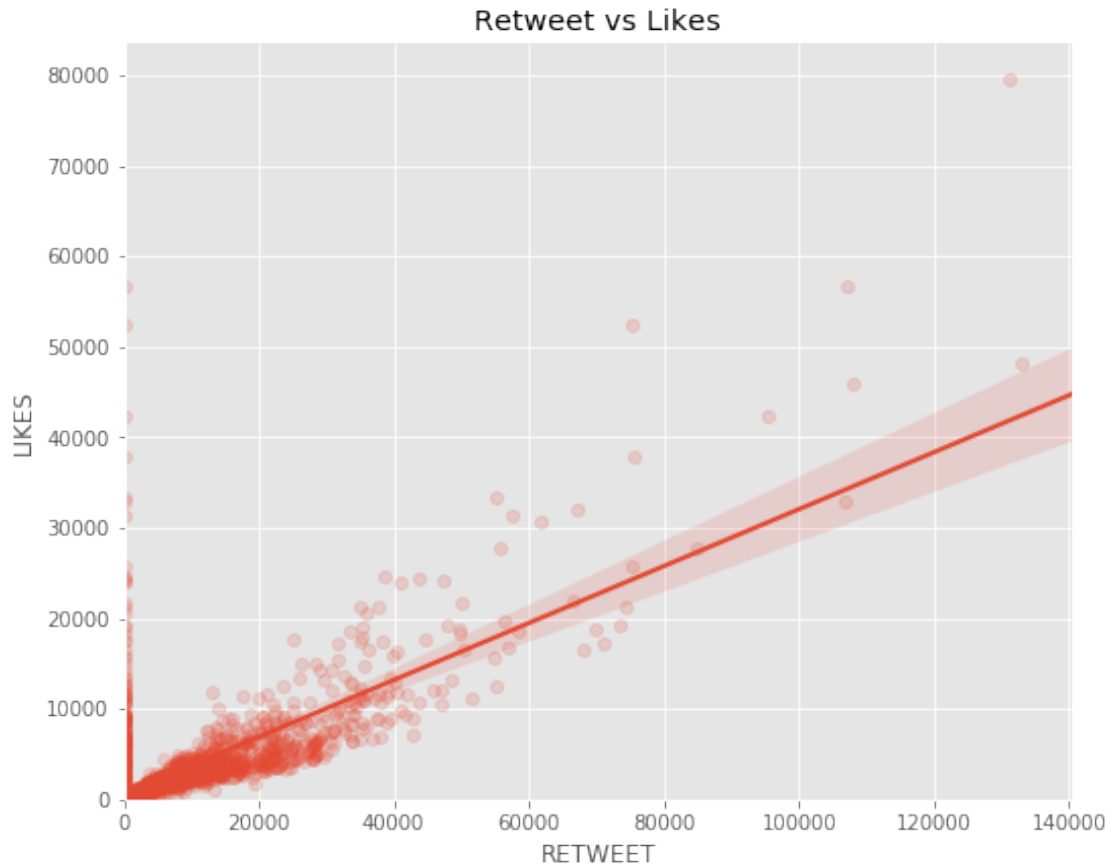
Stencil used as mask in the form of dog-paw.

Most of the dogs of denominator rating 10 have an incrementing trend here.

```
In [81]: # Plot scatterplot of retweet vs favorite count
import seaborn as sns
lm = sns.lmplot(x="likes",
                y="retweets",
                data=df_tweet1,
                size = 6,
                aspect=1.3,
                scatter_kws={'alpha':1/6})

axes = lm.axes
axes[0,0].set_xlim(0,)
axes[0,0].set_ylim(0,)
plt.title('Retweet vs Likes')
plt.xlabel('RETWEET')
plt.ylabel('LIKES')
```

```
Out[81]: Text(-8.7,0.5,'LIKES')
```

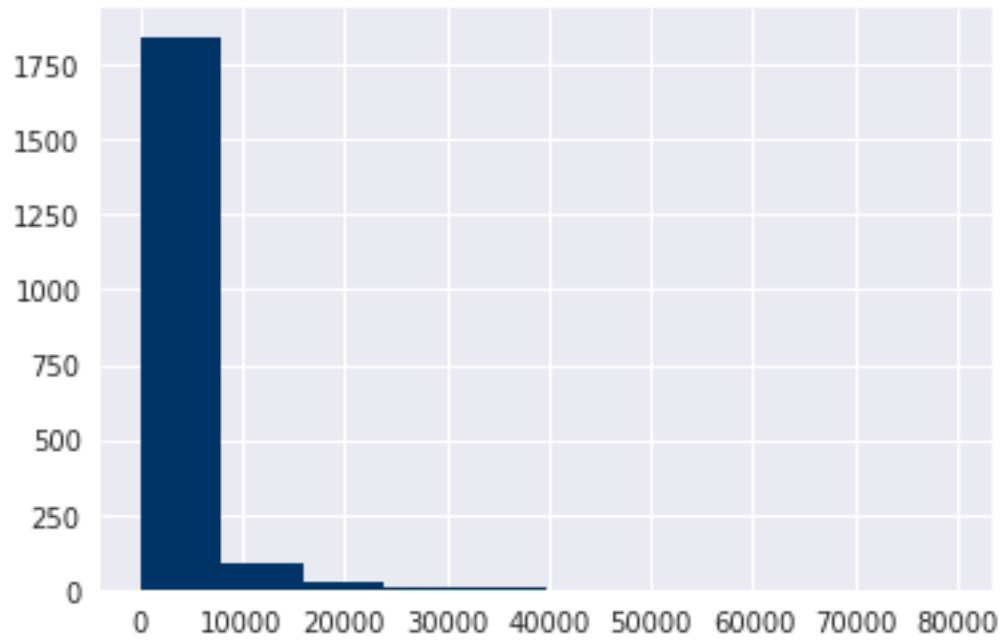


```
In [82]: #Retweets and Likes are linearly related , as the retweets increase,likes increase.
         wrangled_df.retweets.describe()
```

```
Out[82]: count      1971.000000
         mean       2784.449518
         std       4697.662893
         min        16.000000
         25%        628.500000
         50%       1367.000000
         75%       3239.000000
         max      79515.000000
         Name: retweets, dtype: float64
```

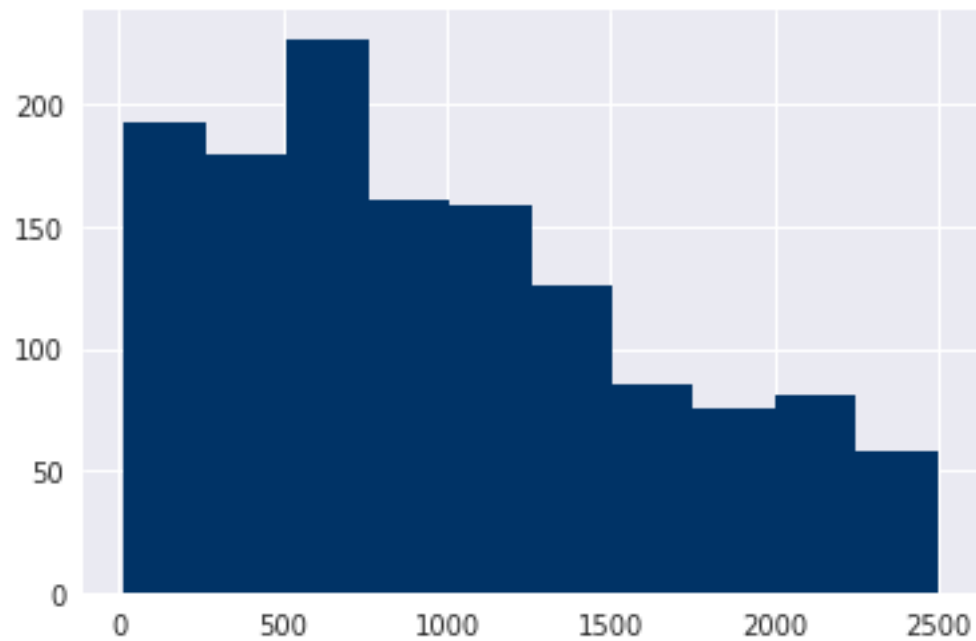
```
In [83]: def set_my_palette():
         sns.set()
         current_palette = sns.color_palette(my_palette)
         sns.set_palette(current_palette)

         my_palette = ['#003366']
         set_my_palette()
         wrangled_df.retweets.hist();
```



In [84]: *#Zooming it further*

```
wrangled_df[wrangled_df.retweets <= 2500].retweets.hist();
```



```

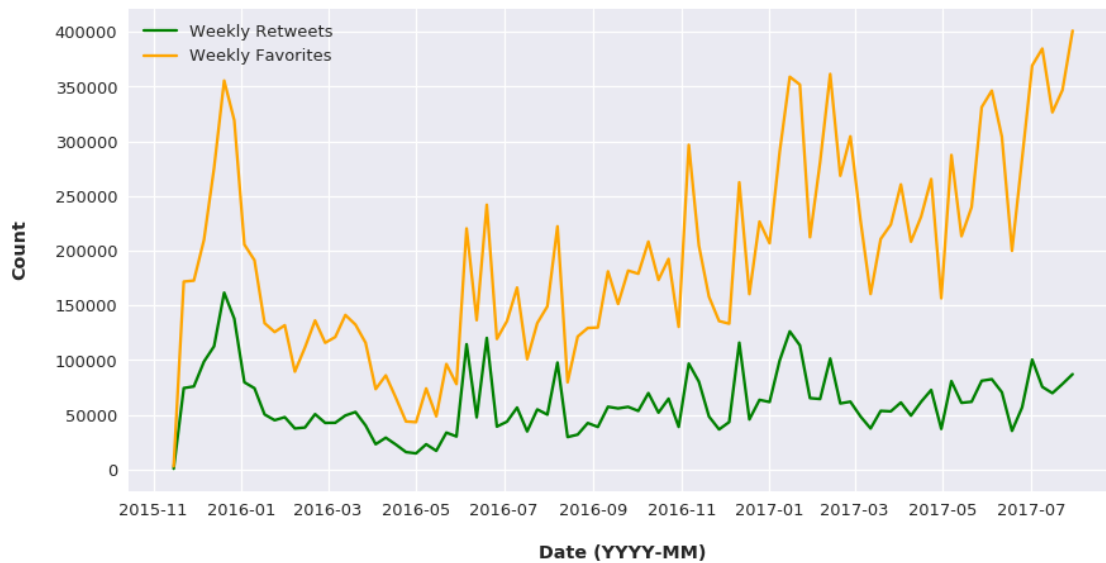
In [85]: wrangled_df['timestamp'] = pd.to_datetime(wrangled_df.timestamp)

weekly_retweet = wrangled_df.groupby(pd.Grouper(key='timestamp', freq='1w'))['retweets'].
    .reset_index().sort_values('timestamp')[: -1]

weekly_favorite = wrangled_df.groupby(pd.Grouper(key='timestamp', freq='1w'))['likes'].
    .reset_index().sort_values('timestamp')[: -1]

sns.set_context("talk")
plt.subplots(figsize=(14, 7))
plt.plot(weekly_retweet.timestamp, weekly_retweet.retweets, label="Weekly Retweets", co
plt.plot(weekly_favorite.timestamp, weekly_favorite.likes, label="Weekly Favorites", co
plt.xlabel('\nDate (YYYY-MM)', weight='bold')
plt.ylabel('Count\n', weight='bold')
plt.legend();

```



```

In [ ]:

```