wrangle_act

July 17, 2019

0.1 Project Details

Your tasks in this project are as follows:

- Data wrangling, which consists of:
- Gathering data (downloadable file in the Resources tab in the left most panel of your classroom and linked in step 1 below).
- Assessing data
- Cleaning data
- Storing, analyzing, and visualizing your wrangled data
- Reporting on 1) your data wrangling efforts and 2) your data analyses and visualizations

0.1.1 Gathering Data for this Project

Gather each of the three pieces of data as described below in a Jupyter Notebook titled wrangle_act.ipynb:

1. The WeRateDogs Twitter archive. I am giving this file to you, so imagine it as a file on hand. Download this file manually by clicking the following link: twitter_archive_enhanced.csv

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

3. Each tweet's retweet count and favorite ("like") count at minimum, and any additional data you find interesting. Using the tweet IDs in the WeRateDogs Twitter archive, query the Twitter API for each tweet's JSON data using Python's Tweepy library and store each tweet's entire set of JSON data in a file called tweet_json.txt file. Each tweet's JSON data should be written to its own line. Then read this .txt file line by line into a pandas DataFrame with (at minimum) tweet ID, retweet count, and favorite count. Note: do not include your Twitter API keys, secrets, and tokens in your project submission.

0.1.2 Assessing Data for this Project

After gathering each of the above pieces of data, assess them visually and programmatically for quality and tidiness issues. Detect and document at least **eight (8) quality issues** and **two (2) tidiness issues** in your wrangle_act.ipynb Jupyter Notebook. To meet specifications, the issues that satisfy the Project Motivation (see the Key Points header on the previous page) must be assessed.

0.1.3 Cleaning Data for this Project

Clean each of the issues you documented while assessing. Perform this cleaning in wrangle_act.ipynb as well. The result should be a high quality and tidy master pandas DataFrame (or DataFrames, if appropriate). Again, the issues that satisfy the Project Motivation must be cleaned.

0.1.4 Storing, Analyzing, and Visualizing Data for this Project

Store the clean DataFrame(s) in a CSV file with the main one named twitter_archive_master.csv. If additional files exist because multiple tables are required for tidiness, name these files appropriately. Additionally, you may store the cleaned data in a SQLite database (which is to be submitted as well if you do).

0.1.5 Deliverables

Analyze and visualize your wrangled data in your wrangle_act.ipynb Jupyter Notebook. At least **three (3) insights and one (1) visualization** must be produced.

Create a **300-600 word written report** called wrangle_report.pdf or wrangle_report.html that briefly describes your wrangling efforts. This is to be framed as an internal document.

Create a **250-word-minimum written report** called act_report.pdf or act_report.html that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

```
In [1]: import pandas as pd
        import numpy as np
        import requests as req
        import tweepy
        import json
        import matplotlib.pyplot as plt
        import seaborn as sns

        % matplotlib inline
```

0.1.6 Gathering Data

1. Gather data from twitter_archive_enhanced.csv

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

```
In [3]: df_twi_enhan.head()
```

```
Out[3]:              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
        0  892420643555336193                    NaN                  NaN   
        1  892177421306343426                    NaN                  NaN   
        2  891815181378084864                    NaN                  NaN   
        3  891689557279858688                    NaN                  NaN   
        4  891327558926688256                    NaN                  NaN   

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

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

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

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

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

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

2. Gather data from image_predictions.tsv

```
In [4]: request = req.get('https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_ima
```

```
In [5]: tweet_image = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-p
```

```
In [6]: df_breed = pd.read_csv(tweet_image, sep = '\t')
```

```
In [7]: df_breed.head()
```

```
Out[7]:              tweet_id                                          jpg_url  \
        0  666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
        1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
        2  666033412701032449  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
        3  666044226329800704  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
        4  666049248165822465  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

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

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

3. df_tweets Dataset

```
In [8]: consumer_key = "ZXva16vi0adGQTBT0cTTHiWG0"
        consumer_secret = "RrKjXya8UZfsYXKuzxL4kHPkRsrFItpzjNdcoigxSqH3tqA6wQ"
        access_token = '1144286188435628034-S0sHyd2tv0fetHUt0D6ZCS9KyazM1S'
        access_secret = '3Ks3EWUXHXyzlPt0ZiRKGtTUGSTs6USd3Mmmp0AaetJYc'

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

        api = tweepy.API(auth, wait_on_rate_limit=True)
```

```
In [9]: id_list = df_twi_enhan['tweet_id']
```

```
In [10]: id_list
```

```
Out[10]: 0       892420643555336193
         1       892177421306343426
```

2	891815181378084864
3	891689557279858688
4	891327558926688256
5	891087950875897856
6	890971913173991426
7	890729181411237888
8	890609185150312448
9	890240255349198849
10	890006608113172480
11	889880896479866881
12	889665388333682689
13	889638837579907072
14	889531135344209921
15	889278841981685760
16	888917238123831296
17	888804989199671297
18	888554962724278272
19	888202515573088257
20	888078434458587136
21	887705289381826560
22	887517139158093824
23	887473957103951883
24	887343217045368832
25	887101392804085760
26	886983233522544640
27	886736880519319552
28	886680336477933568
29	886366144734445568
	...
2326	666411507551481857
2327	666407126856765440
2328	666396247373291520
2329	666373753744588802
2330	666362758909284353
2331	666353288456101888
2332	666345417576210432
2333	666337882303524864
2334	666293911632134144
2335	666287406224695296
2336	666273097616637952
2337	666268910803644416
2338	666104133288665088
2339	666102155909144576
2340	666099513787052032
2341	666094000022159362
2342	666082916733198337
2343	666073100786774016
2344	666071193221509120

```
2345    666063827256086533
2346    666058600524156928
2347    666057090499244032
2348    666055525042405380
2349    666051853826850816
2350    666050758794694657
2351    666049248165822465
2352    666044226329800704
2353    666033412701032449
2354    666029285002620928
2355    666020888022790149
Name: tweet_id, Length: 2356, dtype: int64
```

In [11]:

```
# creating a list for the exceptions
exceptions_list = []

# opening the file to write
with open('tweet_json.txt', 'w', encoding = 'utf-8') as f:
    for id_tweet in id_list:
        try:
            tweet = api.get_status(id_tweet, tweet_mode= 'extended')
            json.dump(tweet._json, f)
# writing the content witt new paragraphs
            f.write("\n")
        except Exception as e:
            exceptions_list.append(id_tweet)
# printing out the exception messages
            print(str(e))
```

```
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 179, 'message': 'Sorry, you are not authorized to see this status.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
```

```
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
[{'code': 144, 'message': 'No status found with that ID.'}]
```

```
In [12]: df_json_tweets = pd.read_json('tweet_json.txt', orient='records', lines = True)

In [13]: df_tweets = df_json_tweets[['id','favorite_count', 'retweet_count', 'retweeted']]

In [14]: df_tweets.head()

Out[14]:                    id  favorite_count  retweet_count  retweeted
         0  892420643555336193           37353           8016      False
         1  892177421306343426           32098           5945      False
         2  891815181378084864           24197           3932      False
         3  891689557279858688           40642           8175      False
         4  891327558926688256           38875           8853      False
```

0.1.7 Assessing Data

1. twitter_archive_enhanced Dataset

```
In [15]: df_twi_enhan

Out[15]:                 tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         0     892420643555336193                    NaN                  NaN
         1     892177421306343426                    NaN                  NaN
         2     891815181378084864                    NaN                  NaN
         3     891689557279858688                    NaN                  NaN
         4     891327558926688256                    NaN                  NaN
         5     891087950875897856                    NaN                  NaN
         6     890971913173991426                    NaN                  NaN
         7     890729181411237888                    NaN                  NaN
         8     890609185150312448                    NaN                  NaN
         9     890240255349198849                    NaN                  NaN
         10    890006608113172480                    NaN                  NaN
         11    889880896479866881                    NaN                  NaN
         12    889665388333682689                    NaN                  NaN
         13    889638837579907072                    NaN                  NaN
         14    889531135344209921                    NaN                  NaN
         15    889278841981685760                    NaN                  NaN
         16    888917238123831296                    NaN                  NaN
         17    888804989199671297                    NaN                  NaN
         18    888554962724278272                    NaN                  NaN
         19    888202515573088257                    NaN                  NaN
         20    888078434458587136                    NaN                  NaN
         21    887705289381826560                    NaN                  NaN
         22    887517139158093824                    NaN                  NaN
```

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

                      timestamp  \
0     2017-08-01 16:23:56 +0000
1     2017-08-01 00:17:27 +0000
2     2017-07-31 00:18:03 +0000
3     2017-07-30 15:58:51 +0000
4     2017-07-29 16:00:24 +0000
5     2017-07-29 00:08:17 +0000
6     2017-07-28 16:27:12 +0000
7     2017-07-28 00:22:40 +0000
```

```
8     2017-07-27 16:25:51 +0000
9     2017-07-26 15:59:51 +0000
10    2017-07-26 00:31:25 +0000
11    2017-07-25 16:11:53 +0000
12    2017-07-25 01:55:32 +0000
13    2017-07-25 00:10:02 +0000
14    2017-07-24 17:02:04 +0000
15    2017-07-24 00:19:32 +0000
16    2017-07-23 00:22:39 +0000
17    2017-07-22 16:56:37 +0000
18    2017-07-22 00:23:06 +0000
19    2017-07-21 01:02:36 +0000
20    2017-07-20 16:49:33 +0000
21    2017-07-19 16:06:48 +0000
22    2017-07-19 03:39:09 +0000
23    2017-07-19 00:47:34 +0000
24    2017-07-18 16:08:03 +0000
25    2017-07-18 00:07:08 +0000
26    2017-07-17 16:17:36 +0000
27    2017-07-16 23:58:41 +0000
28    2017-07-16 20:14:00 +0000
29    2017-07-15 23:25:31 +0000
...                         ...
2326  2015-11-17 00:24:19 +0000
2327  2015-11-17 00:06:54 +0000
2328  2015-11-16 23:23:41 +0000
2329  2015-11-16 21:54:18 +0000
2330  2015-11-16 21:10:36 +0000
2331  2015-11-16 20:32:58 +0000
2332  2015-11-16 20:01:42 +0000
2333  2015-11-16 19:31:45 +0000
2334  2015-11-16 16:37:02 +0000
2335  2015-11-16 16:11:11 +0000
2336  2015-11-16 15:14:19 +0000
2337  2015-11-16 14:57:41 +0000
2338  2015-11-16 04:02:55 +0000
2339  2015-11-16 03:55:04 +0000
2340  2015-11-16 03:44:34 +0000
2341  2015-11-16 03:22:39 +0000
2342  2015-11-16 02:38:37 +0000
2343  2015-11-16 01:59:36 +0000
2344  2015-11-16 01:52:02 +0000
2345  2015-11-16 01:22:45 +0000
2346  2015-11-16 01:01:59 +0000
2347  2015-11-16 00:55:59 +0000
2348  2015-11-16 00:49:46 +0000
2349  2015-11-16 00:35:11 +0000
2350  2015-11-16 00:30:50 +0000
```

```
2351  2015-11-16 00:24:50 +0000
2352  2015-11-16 00:04:52 +0000
2353  2015-11-15 23:21:54 +0000
2354  2015-11-15 23:05:30 +0000
2355  2015-11-15 22:32:08 +0000

                                                 source  \
0     <a href="http://twitter.com/download/iphone" r...
1     <a href="http://twitter.com/download/iphone" r...
2     <a href="http://twitter.com/download/iphone" r...
3     <a href="http://twitter.com/download/iphone" r...
4     <a href="http://twitter.com/download/iphone" r...
5     <a href="http://twitter.com/download/iphone" r...
6     <a href="http://twitter.com/download/iphone" r...
7     <a href="http://twitter.com/download/iphone" r...
8     <a href="http://twitter.com/download/iphone" r...
9     <a href="http://twitter.com/download/iphone" r...
10    <a href="http://twitter.com/download/iphone" r...
11    <a href="http://twitter.com/download/iphone" r...
12    <a href="http://twitter.com/download/iphone" r...
13    <a href="http://twitter.com/download/iphone" r...
14    <a href="http://twitter.com/download/iphone" r...
15    <a href="http://twitter.com/download/iphone" r...
16    <a href="http://twitter.com/download/iphone" r...
17    <a href="http://twitter.com/download/iphone" r...
18    <a href="http://twitter.com/download/iphone" r...
19    <a href="http://twitter.com/download/iphone" r...
20    <a href="http://twitter.com/download/iphone" r...
21    <a href="http://twitter.com/download/iphone" r...
22    <a href="http://twitter.com/download/iphone" r...
23    <a href="http://twitter.com/download/iphone" r...
24    <a href="http://twitter.com/download/iphone" r...
25    <a href="http://twitter.com/download/iphone" r...
26    <a href="http://twitter.com/download/iphone" r...
27    <a href="http://twitter.com/download/iphone" r...
28    <a href="http://twitter.com/download/iphone" r...
29    <a href="http://twitter.com/download/iphone" r...
...                                                 ...
2326  <a href="http://twitter.com/download/iphone" r...
2327  <a href="http://twitter.com/download/iphone" r...
2328  <a href="http://twitter.com/download/iphone" r...
2329  <a href="http://twitter.com/download/iphone" r...
2330  <a href="http://twitter.com/download/iphone" r...
2331  <a href="http://twitter.com/download/iphone" r...
2332  <a href="http://twitter.com/download/iphone" r...
2333  <a href="http://twitter.com/download/iphone" r...
2334  <a href="http://twitter.com/download/iphone" r...
2335  <a href="http://twitter.com/download/iphone" r...
```

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

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

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

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

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

```
2354                       NaN                        NaN
2355                       NaN                        NaN

                                          expanded_urls  rating_numerator  \
0     https://twitter.com/dog_rates/status/892420643...                13
1     https://twitter.com/dog_rates/status/892177421...                13
2     https://twitter.com/dog_rates/status/891815181...                12
3     https://twitter.com/dog_rates/status/891689557...                13
4     https://twitter.com/dog_rates/status/891327558...                12
5     https://twitter.com/dog_rates/status/891087950...                13
6     https://gofundme.com/ydvmve-surgery-for-jax,ht...                13
7     https://twitter.com/dog_rates/status/890729181...                13
8     https://twitter.com/dog_rates/status/890609185...                13
9     https://twitter.com/dog_rates/status/890240255...                14
10    https://twitter.com/dog_rates/status/890006608...                13
11    https://twitter.com/dog_rates/status/889880896...                13
12    https://twitter.com/dog_rates/status/889665388...                13
13    https://twitter.com/dog_rates/status/889638837...                12
14    https://twitter.com/dog_rates/status/889531135...                13
15    https://twitter.com/dog_rates/status/889278841...                13
16    https://twitter.com/dog_rates/status/888917238...                12
17    https://twitter.com/dog_rates/status/888804989...                13
18    https://twitter.com/dog_rates/status/888554962...                13
19    https://twitter.com/dog_rates/status/887473957...                13
20    https://twitter.com/dog_rates/status/888078434...                12
21    https://twitter.com/dog_rates/status/887705289...                13
22    https://twitter.com/dog_rates/status/887517139...                14
23    https://twitter.com/dog_rates/status/887473957...                13
24    https://twitter.com/dog_rates/status/887343217...                13
25    https://twitter.com/dog_rates/status/887101392...                12
26    https://twitter.com/dog_rates/status/886983233...                13
27    https://www.gofundme.com/mingusneedsus,https:/...                13
28    https://twitter.com/dog_rates/status/886680336...                13
29    https://twitter.com/dog_rates/status/886366144...                12
...                                                 ...               ...
2326  https://twitter.com/dog_rates/status/666411507...                 2
2327  https://twitter.com/dog_rates/status/666407126...                 7
2328  https://twitter.com/dog_rates/status/666396247...                 9
2329  https://twitter.com/dog_rates/status/666373753...                11
2330  https://twitter.com/dog_rates/status/666362758...                 6
2331  https://twitter.com/dog_rates/status/666353288...                 8
2332  https://twitter.com/dog_rates/status/666345417...                10
2333  https://twitter.com/dog_rates/status/666337882...                 9
2334  https://twitter.com/dog_rates/status/666293911...                 3
2335  https://twitter.com/dog_rates/status/666287406...                 1
2336  https://twitter.com/dog_rates/status/666273097...                11
2337  https://twitter.com/dog_rates/status/666268910...                10
2338  https://twitter.com/dog_rates/status/666104133...                 1
```

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

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

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

[2356 rows x 17 columns]
```

```
In [16]: df_twi_enhan.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 [17]: df_twi_enhan.shape
```

```
Out[17]: (2356, 17)
```

1.1. Check if there are any duplicated id.

```
In [18]: df_twi_enhan[df_twi_enhan['tweet_id'].duplicated(keep=False)]
```

```
Out[18]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text
         Index: []
```

1.2 Check if there are any misinterpreted info in the dataset. Results show that name start with lowercase letter are all invaid name.

```
In [19]: df_twi_enhan.groupby('name')['name'].size()
```

```
Out[19]: name
         Abby                2
         Ace                 1
         Acro                1
         Adele               1
         Aiden               1
         Aja                 1
         Akumi               1
         Al                  1
         Albert              2
         Albus               2
         Aldrick             1
         Alejandro           1
         Alexander           1
         Alexanderson        1
         Alf                 1
         Alfie               5
         Alfy                1
         Alice               2
```

```
Amber            1
Ambrose          1
Amy              1
Amélie           1
Anakin           2
Andru            1
Andy             1
Angel            1
Anna             1
Anthony          1
Antony           1
Apollo           1
                ..
Ziva             1
Zoe              1
Zoey             3
Zooey            1
Zuzu             1
a               55
actually         2
all              1
an               7
by               1
getting          2
his              1
incredibly       1
infuriating      1
just             4
life             1
light            1
mad              2
my               1
not              2
officially       1
old              1
one              4
quite            4
space            1
such             1
the              8
this             1
unacceptable     1
very             5
Name: name, Length: 957, dtype: int64
```

2. image_predictions Dataset

```
In [20]: df_breed
```

```
Out[20]:            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_3WEAAlOoZ.jpg
12    666073100786774016    https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13    666082916733198337    https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.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/CT56LSZWoAAlJj2.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/CT9OwFIWEAMuRje.jpg
23    666345417576210432    https://pbs.twimg.com/media/CT9Vn7PWoAA_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/CT9vZEYWUAAlZ05.jpg
27    666396247373291520    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28    666407126856765440    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
29    666411507551481857    https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...                  ...                                                ...
2045  886366144734445568    https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046  886680336477933568    https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2047  886736880519319552    https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
2048  886983233522544640    https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.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_O-UQAACu20.jpg
2057  888804989199671297    https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  888917238123831296    https://pbs.twimg.com/media/DFYRgsOUQAARGhO.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/DFl99B1WsAITKsg.jpg
2064  890006608113172480    https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065  890240255349198849    https://pbs.twimg.com/media/DFrEyVuW0AAO3t9.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/DF1eOmZXUAALUcq.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 [21]: df_breed.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 [22]: df_breed.shape
```

```
Out[22]: (2075, 12)
```

2.1. Check if there are any null values.

```
In [23]: df_breed.isnull().sum()
```

```
Out[23]: tweet_id    0
         jpg_url     0
         img_num     0
         p1          0
         p1_conf     0
         p1_dog      0
```

```
p2          0
p2_conf     0
p2_dog      0
p3          0
p3_conf     0
p3_dog      0
dtype: int64
```

3. df_tweets Dataset

```
In [24]: df_tweets.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2333 entries, 0 to 2332
Data columns (total 4 columns):
id                  2333 non-null int64
favorite_count      2333 non-null int64
retweet_count       2333 non-null int64
retweeted           2333 non-null bool
dtypes: bool(1), int64(3)
memory usage: 57.0 KB
```

```
In [25]: df_tweets.shape
```

```
Out[25]: (2333, 4)
```

```
In [26]: df_tweets
```

```
Out[26]:                        id  favorite_count  retweet_count  retweeted
         0     892420643555336193           37353           8016      False
         1     892177421306343426           32098           5945      False
         2     891815181378084864           24197           3932      False
         3     891689557279858688           40642           8175      False
         4     891327558926688256           38875           8853      False
         5     891087950875897856           19541           2948      False
         6     890971913173991426           11407           1949      False
         7     890729181411237888           62957          17861      False
         8     890609185150312448           26877           4048      False
         9     890240255349198849           30807           6974      False
         10    890006608113172480           29623           6939      False
         11    889880896479866881           26873           4719      False
         12    889665388333682689           46431           9496      False
         13    889638837579907072           26155           4266      False
         14    889531135344209921           14603           2134      False
         15    889278841981685760           24356           5075      False
         16    888917238123831296           28124           4258      False
         17    888804989199671297           24685           4033      False
         18    888554962724278272           19123           3333      False
```

```
19    888078434458587136           21005          3287      False
20    887705289381826560           29130          5099      False
21    887517139158093824           44701         11099      False
22    887473957103951883           66459         17104      False
23    887343217045368832           32489          9878      False
24    887101392804085760           29521          5648      False
25    886983233522544640           33859          7283      False
26    886736880519319552           11596          3071      False
27    886680336477933568           21653          4212      False
28    886366144734445568           20423          3017      False
29    886267009285017600             116             4      False
...                  ...             ...           ...        ...
2303  666411507551481857             425           316      False
2304  666407126856765440             103            36      False
2305  666396247373291520             161            80      False
2306  666373753744588802             179            86      False
2307  666362758909284353             749           542      False
2308  666353288456101888             211            71      False
2309  666345417576210432             281           130      False
2310  666337882303524864             188            87      False
2311  666293911632134144             482           335      False
2312  666287406224695296             141            63      False
2313  666273097616637952             166            74      False
2314  666268910803644416              99            32      False
2315  666104133288665088           13845          6219      False
2316  666102155909144576              75            11      False
2317  666099513787052032             147            63      False
2318  666094000022159362             158            69      False
2319  666082916733198337             109            42      False
2320  666073100786774016             310           152      False
2321  666071193221509120             140            55      False
2322  666063827256086533             458           205      False
2323  666058600524156928             108            56      False
2324  666057090499244032             279           132      False
2325  666055525042405380             419           228      False
2326  666051853826850816            1177           811      False
2327  666050758794694657             128            56      False
2328  666049248165822465             102            41      False
2329  666044226329800704             285           135      False
2330  666033412701032449             120            43      False
2331  666029285002620928             125            46      False
2332  666020888022790149            2486           486      False

[2333 rows x 4 columns]
```

3.1. Check if there are any retweeted values.

```
In [27]: len(df_tweets.query('retweeted == True'))
```

```
Out[27]: 0
```

3.2. Check if there are any duplicated id.

```
In [28]: df_tweets[df_tweets['id'].duplicated(keep=False)]

Out[28]: Empty DataFrame
         Columns: [id, favorite_count, retweet_count, retweeted]
         Index: []
```

Conclusion **Quality:** 1. df_twi_enhan contains retweets, where retweeted_status_id has a number instead of NaN. 2. "Timestamp is a string" is a wrong datatype. 3. Dogs'name start with lowercase letter which are invaid infomation, ex. "a", "actually", "all", etc. 4. doggo, floofer, pupper, and puppo are using the string "None" instead of NaN 5. tweet_id, in_reply_to_status_id, in_reply_to_user_id should be changed to string. 6. Rating scale is not accurate. 7. df_breed dataset contains data not related to dogs. 8. The p1, p2, and p3 contents are not consistent, some are capitalized, some contain underscores.

Tidiness: 9. Some info are not useful, such as source column in df_twi_enhan table and img_num in df_breed table. 10. There are three seperate data sources instead of one gaint table, since they all talking about the same tweet. 11. doggo, floofer, pupper, and puppo are all talking about dogs characteristics, can combine them into one column, and drop them afterall.

0.1.8 Cleaning Data

```
In [29]: df_twi_enhan_clean = df_twi_enhan.copy()
         df_breed_clean = df_breed.copy()
         df_tweets_clean = df_tweets.copy()
```

1. Drop duplicated data and retweets data

```
In [30]: df_twi_enhan.drop_duplicates()
         df_breed.drop_duplicates()
         df_tweets.drop_duplicates()

Out[30]:                     id  favorite_count  retweet_count  retweeted
         0   892420643555336193           37353           8016      False
         1   892177421306343426           32098           5945      False
         2   891815181378084864           24197           3932      False
         3   891689557279858688           40642           8175      False
         4   891327558926688256           38875           8853      False
         5   891087950875897856           19541           2948      False
         6   890971913173991426           11407           1949      False
         7   890729181411237888           62957          17861      False
         8   890609185150312448           26877           4048      False
         9   890240255349198849           30807           6974      False
         10  890006608113172480           29623           6939      False
         11  889880896479866881           26873           4719      False
         12  889665388333682689           46431           9496      False
```

```
13    889638837579907072          26155          4266        False
14    889531135344209921          14603          2134        False
15    889278841981685760          24356          5075        False
16    888917238123831296          28124          4258        False
17    888804989199671297          24685          4033        False
18    888554962724278272          19123          3333        False
19    888078434458587136          21005          3287        False
20    887705289381826560          29130          5099        False
21    887517139158093824          44701         11099        False
22    887473957103951883          66459         17104        False
23    887343217045368832          32489          9878        False
24    887101392804085760          29521          5648        False
25    886983233522544640          33859          7283        False
26    886736880519319552          11596          3071        False
27    886680336477933568          21653          4212        False
28    886366144734445568          20423          3017        False
29    886267009285017600            116             4        False
...                  ...            ...           ...          ...
2303  666411507551481857            425           316        False
2304  666407126856765440            103            36        False
2305  666396247373291520            161            80        False
2306  666373753744588802            179            86        False
2307  666362758909284353            749           542        False
2308  666353288456101888            211            71        False
2309  666345417576210432            281           130        False
2310  666337882303524864            188            87        False
2311  666293911632134144            482           335        False
2312  666287406224695296            141            63        False
2313  666273097616637952            166            74        False
2314  666268910803644416             99            32        False
2315  666104133288665088          13845          6219        False
2316  666102155909144576             75            11        False
2317  666099513787052032            147            63        False
2318  666094000022159362            158            69        False
2319  666082916733198337            109            42        False
2320  666073100786774016            310           152        False
2321  666071193221509120            140            55        False
2322  666063827256086533            458           205        False
2323  666058600524156928            108            56        False
2324  666057090499244032            279           132        False
2325  666055525042405380            419           228        False
2326  666051853826850816           1177           811        False
2327  666050758794694657            128            56        False
2328  666049248165822465            102            41        False
2329  666044226329800704            285           135        False
2330  666033412701032449            120            43        False
2331  666029285002620928            125            46        False
2332  666020888022790149           2486           486        False
```

```
[2333 rows x 4 columns]
```

```
In [31]: df_twi_enhan.drop(['retweeted_status_user_id','retweeted_status_id','retweeted_status_t
```

```
In [32]: df_twi_enhan.head()
```

```
Out[32]:             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  \
         0  This is Phineas. He's a mystical boy. Only eve...
         1  This is Tilly. She's just checking pup on you...
         2  This is Archie. He is a rare Norwegian Pouncin...
         3  This is Darla. She commenced a snooze mid meal...
         4  This is Franklin. He would like you to stop ca...

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

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

2. Datatype - Timestamp

```
In [33]: # Remove the time zone information from 'timestamp' column
         df_twi_enhan['timestamp'] = df_twi_enhan['timestamp'].str.slice(start=0, stop=-6)
```

```
In [34]: df_twi_enhan['timestamp'] = pd.to_datetime(df_twi_enhan['timestamp'], format = "%Y-%m-%
```

```
In [35]: df_twi_enhan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 14 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 datetime64[ns]
source                   2356 non-null object
text                     2356 non-null object
expanded_urls            2297 non-null object
rating_numerator         2356 non-null int64
rating_denominator       2356 non-null int64
name                     2356 non-null object
doggo                    2356 non-null object
floofer                  2356 non-null object
pupper                   2356 non-null object
puppo                    2356 non-null object
dtypes: datetime64[ns](1), float64(2), int64(3), object(8)
memory usage: 257.8+ KB
```

```
In [36]: df_twi_enhan.head(2)
```

```
Out[36]:              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         0  892420643555336193                    NaN                  NaN   
         1  892177421306343426                    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...   

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

                                                 expanded_urls  rating_numerator  \
         0  https://twitter.com/dog_rates/status/892420643...                13   
         1  https://twitter.com/dog_rates/status/892177421...                13   

            rating_denominator     name doggo floofer pupper puppo  
         0                  10  Phineas  None    None   None  None  
         1                  10    Tilly  None    None   None  None  
```

3. Inaccurate name (ex. dogs'name)

```
In [37]: lowercase_names = []
         for row in df_twi_enhan['name']:
             if row[0].islower() and row not in lowercase_names:
                 lowercase_names.append(row)
         print(lowercase_names)
```

```
['such', 'a', 'quite', 'not', 'one', 'incredibly', 'mad', 'an', 'very', 'just', 'my', 'his', 'ac
```

```
In [38]: df_twi_enhan['name'].replace(lowercase_names,
                                       np.NaN,
                                      inplace = True)
```

```
In [39]: df_twi_enhan['name'].replace('None',
                                       np.NaN,
                                      inplace = True)
```

```
In [40]: df_twi_enhan['name'].value_counts()
```

```
Out[40]: Charlie      12
         Cooper       11
         Lucy         11
         Oliver       11
         Penny        10
         Lola         10
         Tucker       10
         Winston       9
         Bo            9
         Sadie         8
         Bailey        7
         Buddy         7
         Toby          7
         Daisy         7
         Rusty         6
         Leo           6
         Jax           6
         Scout         6
         Jack          6
         Stanley       6
         Oscar         6
         Milo          6
         Dave          6
         Bella         6
         Koda          6
         Chester       5
         Louis         5
         Oakley        5
```

```
Sunny          5
Larry          5
              ..
Willem         1
Tuco           1
Clifford       1
Smiley         1
Lupe           1
Bode           1
Beya           1
Holly          1
General        1
Snoopy         1
Jaspers        1
Jarod          1
Jaycob         1
Major          1
Sprinkles      1
Gunner         1
Snoop          1
Horace         1
Wesley         1
Skye           1
Millie         1
Brutus         1
Tiger          1
Ronduh         1
Goliath        1
Jordy          1
Sojourner      1
Clarkus        1
Tobi           1
Pluto          1
Name: name, Length: 931, dtype: int64
```

```
In [ ]:
```

4&11. Dogs' characteristics

```
In [41]: df_twi_enhan['dog_charac'] = df_twi_enhan['text'].str.extract('(doggo|floofer|pupper|pu
```

```
In [42]: df_twi_enhan[['dog_charac','doggo', 'floofer', 'pupper', 'puppo']]
```

```
Out[42]:      dog_charac  doggo floofer  pupper  puppo
         0           NaN   None    None    None   None
         1           NaN   None    None    None   None
         2           NaN   None    None    None   None
         3           NaN   None    None    None   None
```

```
4           NaN   None    None    None   None
5           NaN   None    None    None   None
6           NaN   None    None    None   None
7           NaN   None    None    None   None
8           NaN   None    None    None   None
9         doggo  doggo    None    None   None
10          NaN   None    None    None   None
11          NaN   None    None    None   None
12        puppo   None    None    None  puppo
13          NaN   None    None    None   None
14        puppo   None    None    None  puppo
15          NaN   None    None    None   None
16          NaN   None    None    None   None
17          NaN   None    None    None   None
18          NaN   None    None    None   None
19          NaN   None    None    None   None
20          NaN   None    None    None   None
21          NaN   None    None    None   None
22          NaN   None    None    None   None
23          NaN   None    None    None   None
24          NaN   None    None    None   None
25          NaN   None    None    None   None
26          NaN   None    None    None   None
27          NaN   None    None    None   None
28          NaN   None    None    None   None
29       pupper   None    None  pupper   None
...         ...    ...     ...     ...    ...
2326        NaN   None    None    None   None
2327        NaN   None    None    None   None
2328        NaN   None    None    None   None
2329        NaN   None    None    None   None
2330        NaN   None    None    None   None
2331        NaN   None    None    None   None
2332        NaN   None    None    None   None
2333        NaN   None    None    None   None
2334        NaN   None    None    None   None
2335        NaN   None    None    None   None
2336        NaN   None    None    None   None
2337        NaN   None    None    None   None
2338        NaN   None    None    None   None
2339        NaN   None    None    None   None
2340        NaN   None    None    None   None
2341        NaN   None    None    None   None
2342        NaN   None    None    None   None
2343        NaN   None    None    None   None
2344        NaN   None    None    None   None
2345        NaN   None    None    None   None
2346        NaN   None    None    None   None
```

```
2347        NaN   None    None    None   None
2348        NaN   None    None    None   None
2349        NaN   None    None    None   None
2350        NaN   None    None    None   None
2351        NaN   None    None    None   None
2352        NaN   None    None    None   None
2353        NaN   None    None    None   None
2354        NaN   None    None    None   None
2355        NaN   None    None    None   None

[2356 rows x 5 columns]
```

```
In [43]: df_twi_enhan = df_twi_enhan.drop(['doggo', 'floofer', 'pupper', 'puppo'], axis=1)
```

```
In [44]: df_twi_enhan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 11 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 datetime64[ns]
source                   2356 non-null object
text                     2356 non-null object
expanded_urls            2297 non-null object
rating_numerator         2356 non-null int64
rating_denominator       2356 non-null int64
name                     1502 non-null object
dog_charac               399 non-null object
dtypes: datetime64[ns](1), float64(2), int64(3), object(5)
memory usage: 202.5+ KB
```

```
In [45]: df_twi_enhan.shape
```

```
Out[45]: (2356, 11)
```

5. Datatype - in_reply_to_status_id / in_reply_to_user_id / tweet_id

```
In [46]: df_twi_enhan['in_reply_to_status_id'] = df_twi_enhan['in_reply_to_status_id'].astype(st
```

```
In [47]: df_twi_enhan['in_reply_to_user_id'] = df_twi_enhan['in_reply_to_user_id'].astype(str)
```

```
In [48]: df_twi_enhan['tweet_id'] = df_twi_enhan['tweet_id'].astype(str)
```

```
In [49]: df_twi_enhan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 11 columns):
tweet_id                 2356 non-null object
in_reply_to_status_id    2356 non-null object
in_reply_to_user_id      2356 non-null object
timestamp                2356 non-null datetime64[ns]
source                   2356 non-null object
text                     2356 non-null object
expanded_urls            2297 non-null object
rating_numerator         2356 non-null int64
rating_denominator       2356 non-null int64
name                     1502 non-null object
dog_charac               399 non-null object
dtypes: datetime64[ns](1), int64(2), object(8)
memory usage: 202.5+ KB
```

```
In [50]: df_tweets['id'] = df_tweets['id'].astype(str)
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  """Entry point for launching an IPython kernel.
```

```
In [51]: df_tweets.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2333 entries, 0 to 2332
Data columns (total 4 columns):
id                2333 non-null object
favorite_count    2333 non-null int64
retweet_count     2333 non-null int64
retweeted         2333 non-null bool
dtypes: bool(1), int64(2), object(1)
memory usage: 57.0+ KB
```

6. Rating scale

```
In [52]: df_twi_enhan.head()
```

```
Out[52]:              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  \
0  This is Phineas. He's a mystical boy. Only eve...
1  This is Tilly. She's just checking pup on you...
2  This is Archie. He is a rare Norwegian Pouncin...
3  This is Darla. She commenced a snooze mid meal...
4  This is Franklin. He would like you to stop ca...

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

   rating_denominator      name dog_charac
0                  10   Phineas        NaN
1                  10     Tilly        NaN
2                  10    Archie        NaN
3                  10     Darla        NaN
4                  10  Franklin        NaN
```

```
In [53]: df_twi_enhan = df_twi_enhan.drop('rating_denominator', axis=1)
```

```
In [54]: df_twi_enhan = df_twi_enhan.rename(index=str, columns={"rating_numerator": "rating_scal
```

```
In [55]: df_twi_enhan['rating_scale_10'] = df_twi_enhan['rating_scale_10'].astype('float')
```

```
In [56]: df_twi_enhan['rating_scale_10'].value_counts()
```

```
Out[56]: 12.0      558
         11.0      464
         10.0      461
         13.0      351
         9.0       158
         8.0       102
         7.0        55
         14.0       54
         5.0        37
         6.0        32
```

```
3.0        19
4.0        17
1.0         9
2.0         9
75.0        2
15.0        2
420.0       2
0.0         2
144.0       1
666.0       1
121.0       1
182.0       1
165.0       1
17.0        1
45.0        1
204.0       1
960.0       1
1776.0      1
84.0        1
24.0        1
27.0        1
88.0        1
99.0        1
50.0        1
80.0        1
60.0        1
44.0        1
20.0        1
26.0        1
143.0       1
Name: rating_scale_10, dtype: int64
```

```
In [57]: df_wrong_numerator = df_twi_enhan[df_twi_enhan.text.str.contains(r"(\d+\.\d*\/\d+)")]
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: UserWarning: This pattern has ma
  """Entry point for launching an IPython kernel.
```

```
In [58]: numerator = []
         for number in  df_wrong_numerator['text']:
             seperated = number.split('/')
             numerator.append(seperated[0].split()[-1])
         print(numerator)
```

```
['13.5', '9.75', '9.75', '11.27', '9.5', '11.26']
```

```
In [59]: df_twi_enhan_id_list = df_wrong_numerator['tweet_id'].tolist()
```

```
        for i in range(len(df_twi_enhan_id_list)):
            df_twi_enhan.loc[(df_twi_enhan['tweet_id'] == df_twi_enhan_id_list[i]), ['rating_sc
            print(df_twi_enhan.loc[(df_twi_enhan['tweet_id'] ==  df_twi_enhan_id_list[i]), ['ra
```

```
   rating_scale_10
45            13.5
    rating_scale_10
340            9.75
    rating_scale_10
695            9.75
    rating_scale_10
763           11.27
     rating_scale_10
1689             9.5
     rating_scale_10
1712           11.26
```

```
In [60]: df_twi_enhan.head(46)
```

```
Out[60]:               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
30  886267009285017600     8.86266357075e+17        2281181600.0
31  886258384151887873                   nan                 nan
32  886054160059072513                   nan                 nan
33  885984800019947520                   nan                 nan
34  885528943205470208                   nan                 nan
35  885518971528720385                   nan                 nan
36  885311592912609280                   nan                 nan
37  885167619883638784                   nan                 nan
38  884925521741709313                   nan                 nan
39  884876753390489601                   nan                 nan
40  884562892145688576                   nan                 nan
41  884441805382717440                   nan                 nan
42  884247878851493888                   nan                 nan
43  884162670584377345                   nan                 nan
44  883838122936631299                   nan                 nan
45  883482846933004288                   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...
5  2017-07-29 00:08:17  <a href="http://twitter.com/download/iphone" r...
6  2017-07-28 16:27:12  <a href="http://twitter.com/download/iphone" r...
7  2017-07-28 00:22:40  <a href="http://twitter.com/download/iphone" r...
8  2017-07-27 16:25:51  <a href="http://twitter.com/download/iphone" r...
9  2017-07-26 15:59:51  <a href="http://twitter.com/download/iphone" r...
10 2017-07-26 00:31:25  <a href="http://twitter.com/download/iphone" r...
11 2017-07-25 16:11:53  <a href="http://twitter.com/download/iphone" r...
12 2017-07-25 01:55:32  <a href="http://twitter.com/download/iphone" r...
13 2017-07-25 00:10:02  <a href="http://twitter.com/download/iphone" r...
14 2017-07-24 17:02:04  <a href="http://twitter.com/download/iphone" r...
15 2017-07-24 00:19:32  <a href="http://twitter.com/download/iphone" r...
16 2017-07-23 00:22:39  <a href="http://twitter.com/download/iphone" r...
17 2017-07-22 16:56:37  <a href="http://twitter.com/download/iphone" r...
18 2017-07-22 00:23:06  <a href="http://twitter.com/download/iphone" r...
19 2017-07-21 01:02:36  <a href="http://twitter.com/download/iphone" r...
20 2017-07-20 16:49:33  <a href="http://twitter.com/download/iphone" r...
21 2017-07-19 16:06:48  <a href="http://twitter.com/download/iphone" r...
22 2017-07-19 03:39:09  <a href="http://twitter.com/download/iphone" r...
23 2017-07-19 00:47:34  <a href="http://twitter.com/download/iphone" r...
24 2017-07-18 16:08:03  <a href="http://twitter.com/download/iphone" r...
25 2017-07-18 00:07:08  <a href="http://twitter.com/download/iphone" r...
26 2017-07-17 16:17:36  <a href="http://twitter.com/download/iphone" r...
```

```
27 2017-07-16 23:58:41  <a href="http://twitter.com/download/iphone" r...
28 2017-07-16 20:14:00  <a href="http://twitter.com/download/iphone" r...
29 2017-07-15 23:25:31  <a href="http://twitter.com/download/iphone" r...
30 2017-07-15 16:51:35  <a href="http://twitter.com/download/iphone" r...
31 2017-07-15 16:17:19  <a href="http://twitter.com/download/iphone" r...
32 2017-07-15 02:45:48  <a href="http://twitter.com/download/iphone" r...
33 2017-07-14 22:10:11  <a href="http://twitter.com/download/iphone" r...
34 2017-07-13 15:58:47  <a href="http://twitter.com/download/iphone" r...
35 2017-07-13 15:19:09  <a href="http://twitter.com/download/iphone" r...
36 2017-07-13 01:35:06  <a href="http://twitter.com/download/iphone" r...
37 2017-07-12 16:03:00  <a href="http://twitter.com/download/iphone" r...
38 2017-07-12 00:01:00  <a href="http://twitter.com/download/iphone" r...
39 2017-07-11 20:47:12  <a href="http://twitter.com/download/iphone" r...
40 2017-07-11 00:00:02  <a href="http://twitter.com/download/iphone" r...
41 2017-07-10 15:58:53  <a href="http://twitter.com/download/iphone" r...
42 2017-07-10 03:08:17  <a href="http://twitter.com/download/iphone" r...
43 2017-07-09 21:29:42  <a href="http://twitter.com/download/iphone" r...
44 2017-07-09 00:00:04  <a href="http://twitter.com/download/iphone" r...
45 2017-07-08 00:28:19  <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 ...
19  RT @dog_rates: This is Canela. She attempted s...
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...
30  @NonWhiteHat @MayhewMayhem omg hello tanner yo...
31  This is Waffles. His doggles are pupside down...
32  RT @Athletics: 12/10 #BATP https://t.co/WxwJmv...
33  Viewer discretion advised. This is Jimbo. He w...
34  This is Maisey. She fell asleep mid-excavation...
35  I have a new hero and his name is Howard. 14/1...
36  RT @dog_rates: This is Lilly. She just paralle...
37  Here we have a corgi undercover as a malamute...
38  This is Earl. He found a hat. Nervous about wh...
39  This is Lola. It's her first time outside. Mus...
40  This is Kevin. He's just so happy. 13/10 what ...
41  I present to you, Pup in Hat. Pup in Hat is gr...
42  OMG HE DIDN'T MEAN TO HE WAS JUST TRYING A LIT...
43  Meet Yogi. He doesn't have any important dog m...
44  This is Noah. He can't believe someone made th...
45  This is Bella. She hopes her smile made you sm...

                                        expanded_urls rating_scale_10  \
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
30                                                NaN              12
31  https://twitter.com/dog_rates/status/886258384...              13
32  https://twitter.com/dog_rates/status/886053434...              12
33  https://twitter.com/dog_rates/status/885984800...              12
34  https://twitter.com/dog_rates/status/885528943...              13
35  https://twitter.com/4bonds2carbon/status/88551...              14
36  https://twitter.com/dog_rates/status/830583320...              13
37  https://twitter.com/dog_rates/status/885167619...              13
38  https://twitter.com/dog_rates/status/884925521...              12
39  https://twitter.com/dog_rates/status/884876753...              13
40  https://twitter.com/dog_rates/status/884562892...              13
41  https://twitter.com/dog_rates/status/884441805...              14
42  https://twitter.com/kaijohnson_19/status/88396...              13
43  https://twitter.com/dog_rates/status/884162670...              12
44  https://twitter.com/dog_rates/status/883838122...              12
45  https://twitter.com/dog_rates/status/883482846...            13.5

        name dog_charac
0    Phineas        NaN
1      Tilly        NaN
2     Archie        NaN
3      Darla        NaN
4   Franklin        NaN
5        NaN        NaN
6        Jax        NaN
7        NaN        NaN
8       Zoey        NaN
9     Cassie      doggo
10      Koda        NaN
11     Bruno        NaN
12       NaN      puppo
13       Ted        NaN
14    Stuart      puppo
15    Oliver        NaN
16       Jim        NaN
17      Zeke        NaN
18   Ralphus        NaN
19    Canela        NaN
20    Gerald        NaN
21   Jeffrey        NaN
22       NaN        NaN
23    Canela        NaN
24       NaN        NaN
25       NaN        NaN
26      Maya        NaN
```

```
         27    Mingus        NaN
         28     Derek        NaN
         29    Roscoe     pupper
         30       NaN        NaN
         31   Waffles        NaN
         32       NaN        NaN
         33     Jimbo        NaN
         34    Maisey        NaN
         35       NaN        NaN
         36     Lilly        NaN
         37       NaN        NaN
         38      Earl        NaN
         39      Lola        NaN
         40     Kevin        NaN
         41       NaN        NaN
         42       NaN        NaN
         43      Yogi      doggo
         44      Noah        NaN
         45     Bella        NaN
```

7. df_breed dataset contains data not related to dogs.

```
In [61]: df_breed
```

```
Out[61]:                 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_3WEAAlOoZ.jpg
         12     666073100786774016    https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
         13     666082916733198337    https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.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/CT56LSZWoAAlJj2.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/CT9OwFIWEAMuRje.jpg
```

```
23    666345417576210432    https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg
24    666353288456101888    https://pbs.twimg.com/media/CT9cxOtUEAAhNN_.jpg
25    666362758909284353    https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
26    666373753744588802    https://pbs.twimg.com/media/CT9vZEYWUAAlZ05.jpg
27    666396247373291520    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28    666407126856765440    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
29    666411507551481857    https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...                  ...                                                ...
2045  886366144734445568    https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046  886680336477933568    https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2047  886736880519319552    https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
2048  886983233522544640    https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.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_O-UQAACu20.jpg
2057  888804989199671297    https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  888917238123831296    https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg
2059  889278841981685760  https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060  889531135344209921    https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg
2061  889638837579907072    https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062  889665388333682689    https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063  889880896479866881    https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg
2064  890006608113172480    https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065  890240255349198849    https://pbs.twimg.com/media/DFrEyVuW0AAO3t9.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/DF1eOmZXUAALUcq.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 [62]: df_breed.query('p1_dog == False and p2_dog == False and p3_dog == False')
```

```
Out[62]:                 tweet_id                                          jpg_url  \
         6     666051853826850816  https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
         17    666104133288665088  https://pbs.twimg.com/media/CT56LSZWoAAlJj2.jpg
         18    666268910803644416  https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
         21    666293911632134144  https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
         25    666362758909284353  https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
         29    666411507551481857  https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
         45    666786068205871104  https://pbs.twimg.com/media/CUDmZIkWcAAIPPe.jpg
```

```
50    666837028449972224    https://pbs.twimg.com/media/CUEUva1WsAA2jPb.jpg
51    666983947667116034    https://pbs.twimg.com/media/CUGaXDhW4AY9JUH.jpg
53    667012601033924608    https://pbs.twimg.com/media/CUG0bC0U8AAw2su.jpg
56    667065535570550784    https://pbs.twimg.com/media/CUHkkJpXIAA2w3n.jpg
69    667188689915760640    https://pbs.twimg.com/media/CUJUk2iWUAAVtOv.jpg
73    667369227918143488    https://pbs.twimg.com/media/CUL4xR9UkAEdlJ6.jpg
77    667437278097252352    https://pbs.twimg.com/media/CUM2qWaWoAUZ06L.jpg
78    667443425659232256    https://pbs.twimg.com/media/CUM8QZwW4AAVsBl.jpg
93    667549055577362432    https://pbs.twimg.com/media/CUOcVCwWsAERUKY.jpg
94    667550882905632768    https://pbs.twimg.com/media/CUObvUJVEAAnYPF.jpg
96    667724302356258817    https://pbs.twimg.com/media/CUQ7tv3W4AA3KlI.jpg
98    667766675769573376    https://pbs.twimg.com/media/CURiQMnUAAAPT2M.jpg
100   667782464991965184    https://pbs.twimg.com/media/CURwm3cUkAARcO6.jpg
106   667866724293877760    https://pbs.twimg.com/media/CUS9PlUWwAANeAD.jpg
107   667873844930215936    https://pbs.twimg.com/media/CUTDtyGXIAARxus.jpg
112   667911425562669056    https://pbs.twimg.com/media/CUTl5m1WUAAabZG.jpg
115   667937095915278337    https://pbs.twimg.com/media/CUT9PuQWwAABQv7.jpg
117   668142349051129856    https://pbs.twimg.com/media/CUW37BzWsAAlJlN.jpg
118   668154635664932864    https://pbs.twimg.com/media/CUXDGR2WcAAUQKz.jpg
123   668226093875376128    https://pbs.twimg.com/media/CUYEFlQXAAUkPGm.jpg
130   668291999406125056    https://pbs.twimg.com/media/CUZABzGW4AE5F0k.jpg
132   668466899341221888    https://pbs.twimg.com/media/CUbfGbbWoAApZth.jpg
140   668544745690562560    https://pbs.twimg.com/media/CUcl5jeWsAA6ufS.jpg
...                  ...                                                ...
1839  837482249356513284    https://pbs.twimg.com/media/C59VqMUXEAAzldG.jpg
1844  838916489579200512    https://pbs.twimg.com/media/C6RkiQZUsAAM4R4.jpg
1847  839290600511926273    https://pbs.twimg.com/media/C6XBt9XXEAEEW9U.jpg
1851  840370681858686976    https://pbs.twimg.com/media/C6mYrK0UwAANhep.jpg
1853  840696689258311684    https://pbs.twimg.com/media/C6rBLenU0AAr8MN.jpg
1869  844580511645339650    https://pbs.twimg.com/media/C7iNfq1W0AAcbsR.jpg
1886  847962785489326080    https://pbs.twimg.com/media/C8SRpHNUIAARB3j.jpg
1887  847971574464610304    https://pbs.twimg.com/media/C8SZH1EWAAAIRRF.jpg
1891  849051919805034497    https://pbs.twimg.com/media/C8hwNxbXYAAwyVG.jpg
1892  849336543269576704    https://pbs.twimg.com/media/C8lzFC4XcAAQxB4.jpg
1900  851464819735769094    https://pbs.twimg.com/media/C9ECujZXsAAPCSM.jpg
1902  851861385021730816    https://pbs.twimg.com/media/C8W6sY_W0AEmttW.jpg
1905  852226086759018497  https://pbs.twimg.com/ext_tw_video_thumb/85222...
1906  852311364735569921    https://pbs.twimg.com/media/C9QEqZ7XYAIR7fS.jpg
1910  853299958564483072    https://pbs.twimg.com/media/C9eHyF7XgAAOxPM.jpg
1931  859074603037188101    https://pbs.twimg.com/media/C-wLyufW0AA546I.jpg
1936  860184849394610176    https://pbs.twimg.com/media/C-_9jWWUwAAnwkd.jpg
1937  860276583193509888    https://pbs.twimg.com/media/C_BQ_NlVwAAgYGD.jpg
1940  860924035999428608    https://pbs.twimg.com/media/C_KVJjDXsAEUCWn.jpg
1946  862457590147678208    https://pbs.twimg.com/media/C_gQmaTUMAAPYSS.jpg
1953  863907417377173506    https://pbs.twimg.com/media/C_O3NPeUQAAgrMl.jpg
1956  864873206498414592    https://pbs.twimg.com/media/DAClmHkXcAA1kSv.jpg
1975  870063196459192321    https://pbs.twimg.com/media/DBMV3NnXUAAm0Pp.jpg
1979  870804317367881728    https://pbs.twimg.com/media/DBW35ZsVoAEWZUU.jpg
```

```
2012  879050749262655488   https://pbs.twimg.com/media/DDMD_phXoAQ1qf0.jpg
2021  880935762899988482   https://pbs.twimg.com/media/DDm2Z5aXUAEDS2u.jpg
2022  881268444196462592   https://pbs.twimg.com/media/DDrk-f9WAAI-WQv.jpg
2046  886680336477933568   https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2052  887517139158093824  https://pbs.twimg.com/ext_tw_video_thumb/88751...
2074  892420643555336193   https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

      img_num                p1   p1_conf  p1_dog                   p2  \
6           1        box_turtle  0.933012   False           mud_turtle
17          1               hen  0.965932   False                 cock
18          1  desktop_computer  0.086502   False                 desk
21          1  three-toed_sloth  0.914671   False                otter
25          1        guinea_pig  0.996496   False                skunk
29          1              coho  0.404640   False           barracouta
45          1             snail  0.999888   False                 slug
50          1       triceratops  0.442113   False            armadillo
51          1              swab  0.589446   False            chain_saw
53          1             hyena  0.987230   False  African_hunting_dog
56          1     jigsaw_puzzle  0.560001   False              doormat
69          1            vacuum  0.335830   False                 swab
73          1             teddy  0.709545   False           bath_towel
77          1         porcupine  0.989154   False           bath_towel
78          1             goose  0.980815   False                drake
93          1      electric_fan  0.984377   False            spotlight
94          1          web_site  0.998258   False           dishwasher
96          1              ibex  0.619098   False              bighorn
98          1       fire_engine  0.883493   False            tow_truck
100         1          lorikeet  0.466149   False          hummingbird
106         1     jigsaw_puzzle  1.000000   False           prayer_rug
107         1     common_iguana  0.999647   False       frilled_lizard
112         1    frilled_lizard  0.257695   False                   ox
115         1           hamster  0.172078   False           guinea_pig
117         1            Angora  0.918834   False                  hen
118         1        Arctic_fox  0.473584   False              wallaby
123         1          trombone  0.390339   False               cornet
130         1          web_site  0.995535   False                skunk
132         1   shopping_basket  0.398361   False               hamper
140         1          bearskin  0.427870   False                  bow
...       ...               ...       ...     ...                  ...
1839        2         birdhouse  0.541196   False           can_opener
1844        2          web_site  0.993651   False              monitor
1847        1          web_site  0.670892   False              monitor
1851        1            teapot  0.981819   False                  cup
1853        1          web_site  0.841768   False                 rule
1869        1            washer  0.903064   False           dishwasher
1886        1          sea_lion  0.882654   False                 mink
1887        1        coffee_mug  0.633652   False                  cup
1891        1          fountain  0.997509   False  American_black_bear
```

```
1892         1           patio  0.521788   False              prison
1900         2        web_site  0.919649   False                menu
1902         1      pencil_box  0.662183   False               purse
1905         1          prison  0.352793   False          dishwasher
1906         1         barbell  0.971581   False            dumbbell
1910         1          grille  0.652280   False         beach_wagon
1931         1        revolver  0.190292   False          projectile
1936         1      chimpanzee  0.267612   False             gorilla
1937         1        lakeside  0.312299   False                dock
1940         2        envelope  0.933016   False        oscilloscope
1946         1    home_theater  0.496348   False        studio_couch
1953         1          marmot  0.358828   False             meerkat
1956         2            pole  0.478616   False            lakeside
1975         1      comic_book  0.534409   False            envelope
1979         1    home_theater  0.168290   False             sandbar
2012         1           tabby  0.311861   False       window_screen
2021         1     street_sign  0.251801   False            umbrella
2022         1          tusker  0.473303   False      Indian_elephant
2046         1     convertible  0.738995   False          sports_car
2052         1       limousine  0.130432   False           tow_truck
2074         1          orange  0.097049   False               bagel

          p2_conf  p2_dog                 p3       p3_conf  p3_dog
6    4.588540e-02   False           terrapin  1.788530e-02   False
17   3.391940e-02   False          partridge  5.206580e-05   False
18   8.554740e-02   False           bookcase  7.947970e-02   False
21   1.525000e-02   False     great_grey_owl  1.320720e-02   False
25   2.402450e-03   False            hamster  4.608630e-04   False
29   2.714850e-01   False                gar  1.899450e-01   False
45   5.514170e-05   False              acorn  2.625800e-05   False
50   1.140710e-01   False      common_iguana  4.325530e-02   False
51   1.901420e-01   False                wig  3.450970e-02   False
53   1.260080e-02   False             coyote  5.735010e-05   False
56   1.032590e-01   False       space_heater  4.256800e-02   False
69   2.652780e-01   False      toilet_tissue  1.407030e-01   False
73   1.272850e-01   False  Christmas_stocking  2.856750e-02   False
77   6.300490e-03   False             badger  9.663400e-04   False
78   6.917770e-03   False                hen  5.255170e-03   False
93   7.736710e-03   False          lampshade  1.901230e-03   False
94   2.010840e-04   False       oscilloscope  1.417360e-04   False
96   1.251190e-01   False                ram  7.467320e-02   False
98   7.473390e-02   False               jeep  1.277260e-02   False
100  8.301100e-02   False       African_grey  5.424740e-02   False
106  1.011300e-08   False            doormat  1.740170e-10   False
107  1.811500e-04   False   African_chameleon  1.283570e-04   False
112  2.351600e-01   False        triceratops  8.531690e-02   False
115  9.492420e-02   False           Band_Aid  5.999520e-02   False
117  3.779340e-02   False        wood_rabbit  1.101490e-02   False
```

```
118   2.614110e-01   False          white_wolf  8.094780e-02   False
123   3.141490e-01   False         French_horn  2.551820e-01   False
130   1.363490e-03   False              badger  6.856500e-04   False
132   3.632220e-01   False            bassinet  8.417350e-02   False
140   2.588580e-01   False             panpipe  2.156260e-02   False
...            ...     ...                 ...           ...     ...
1839  1.210940e-01   False              carton  5.613670e-02   False
1844  1.405900e-03   False            envelope  1.093090e-03   False
1847  1.015650e-01   False              screen  7.530610e-02   False
1851  1.402580e-02   False           coffeepot  2.420540e-03   False
1853  7.087310e-03   False            envelope  6.820300e-03   False
1869  3.248900e-02   False             printer  1.645620e-02   False
1886  6.688020e-02   False               otter  2.567870e-02   False
1887  2.733920e-01   False       toilet_tissue  6.665580e-02   False
1891  1.413120e-03   False             sundial  6.811150e-04   False
1892  1.495440e-01   False          restaurant  2.715260e-02   False
1900  2.630610e-02   False    crossword_puzzle  3.481510e-03   False
1902  6.650550e-02   False              pillow  4.472530e-02   False
1905  1.107230e-01   False                file  9.411200e-02   False
1906  2.841790e-02   False             go-kart  5.595040e-07   False
1910  1.128460e-01   False         convertible  8.625230e-02   False
1931  1.490640e-01   False            fountain  6.604660e-02   False
1936  1.042930e-01   False           orangutan  5.990750e-02   False
1937  1.598420e-01   False               canoe  7.079450e-02   False
1940  1.259140e-02   False         paper_towel  1.117850e-02   False
1946  1.672560e-01   False        barber_chair  5.262500e-02   False
1953  1.747030e-01   False              weasel  1.234850e-01   False
1956  1.141820e-01   False               wreck  5.592650e-02   False
1975  2.807220e-01   False         book_jacket  4.378550e-02   False
1979  9.804040e-02   False          television  7.972940e-02   False
2012  1.691230e-01   False        Egyptian_cat  1.329320e-01   False
2021  1.151230e-01   False       traffic_light  6.953380e-02   False
2022  2.456460e-01   False                ibex  5.566070e-02   False
2046  1.399520e-01   False           car_wheel  4.417270e-02   False
2052  2.917540e-02   False       shopping_cart  2.632080e-02   False
2074  8.585110e-02   False              banana  7.611000e-02   False

[324 rows x 12 columns]
```

```
In [63]: df_breed = df_breed.query('p1_dog == True or p2_dog == True or p3_dog == True')
```

```
In [64]: df_breed.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1751 entries, 0 to 2073
Data columns (total 12 columns):
tweet_id    1751 non-null int64
jpg_url     1751 non-null object
```

```
img_num     1751 non-null int64
p1          1751 non-null object
p1_conf     1751 non-null float64
p1_dog      1751 non-null bool
p2          1751 non-null object
p2_conf     1751 non-null float64
p2_dog      1751 non-null bool
p3          1751 non-null object
p3_conf     1751 non-null float64
p3_dog      1751 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 141.9+ KB
```

```
In [65]: df_breed.query('p1_dog == False and p2_dog == False and p3_dog == False')
```

```
Out[65]: Empty DataFrame
         Columns: [tweet_id, jpg_url, img_num, p1, p1_conf, p1_dog, p2, p2_conf, p2_dog, p3, p3_
         Index: []
```

8. p1, p2, and p3 data consistency.

```
In [66]: df_breed['p1'] = df_breed.p1.str.capitalize()
         df_breed['p2'] = df_breed.p1.str.capitalize()
         df_breed['p3'] = df_breed.p1.str.capitalize()
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  """Entry point for launching an IPython kernel.
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#

/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  This is separate from the ipykernel package so we can avoid doing imports until
```

```
In [67]: df_breed['p1'].replace(regex=True,inplace=True,to_replace="_",value=r' ')
         df_breed['p2'].replace(regex=True,inplace=True,to_replace="_",value=r' ')
         df_breed['p3'].replace(regex=True,inplace=True,to_replace="_",value=r' ')
```

```
/opt/conda/lib/python3.6/site-packages/pandas/core/generic.py:5890: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  self._update_inplace(new_data)
```

```
In [68]: df_breed
```

```
Out[68]:                 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
         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_3WEAAlOoZ.jpg
         12    666073100786774016    https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
         13    666082916733198337    https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.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
         19    666273097616637952    https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
         20    666287406224695296    https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
         22    666337882303524864    https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg
         23    666345417576210432    https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg
         24    666353288456101888    https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
         26    666373753744588802    https://pbs.twimg.com/media/CT9vZEYWUAAlZ05.jpg
         27    666396247373291520    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
         28    666407126856765440    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
         30    666418789513326592    https://pbs.twimg.com/media/CT-YWb7U8AA7QnN.jpg
         31    666421158376562688    https://pbs.twimg.com/media/CT-aggCXAAIMfT3.jpg
         32    666428276349472768    https://pbs.twimg.com/media/CT-g-0DUwAEQdSn.jpg
         33    666430724426358785    https://pbs.twimg.com/media/CT-jNYqW4AAPi2M.jpg
         34    666435652385423360    https://pbs.twimg.com/media/CT-nsTQWEAEkyDn.jpg
         35    666437273139982337    https://pbs.twimg.com/media/CT-pKmRWIAAxUWj.jpg
         ...                  ...                                               ...
         2042  885528943205470208    https://pbs.twimg.com/media/DEoH3yvXgAAzQtS.jpg
         2043  885984800019947520    https://pbs.twimg.com/media/DEumeWWV0AA-Z61.jpg
         2044  886258384151887873    https://pbs.twimg.com/media/DEyfTG4UMAE4aE9.jpg
         2045  886366144734445568    https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
         2047  886736880519319552    https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
         2048  886983233522544640    https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.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
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_O-UQAACu2O.jpg
2057  888804989199671297    https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  888917238123831296    https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg
2059  889278841981685760  https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060  889531135344209921    https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg
2061  889638837579907072    https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062  889665388333682689    https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063  889880896479866881    https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg
2064  890006608113172480    https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065  890240255349198849    https://pbs.twimg.com/media/DFrEyVuW0AAO3t9.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/DF1eOmZXUAALUcq.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

      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
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
19          1          Italian greyhound  0.176053    True
20          1                Maltese dog  0.857531    True
22          1                         Ox  0.416669   False
23          1          Golden retriever  0.858744    True
24          1                   Malamute  0.336874    True
26          1  Soft-coated wheaten terrier  0.326467    True
27          1                  Chihuahua  0.978108    True
```

```
28          1      Black-and-tan coonhound  0.529139    True
30          1                  Toy terrier  0.149680    True
31          1             Blenheim spaniel  0.906777    True
32          1                     Pembroke  0.371361    True
33          1                        Llama  0.505184   False
34          1     Chesapeake bay retriever  0.184130    True
35          1                    Chihuahua  0.671853    True
...       ...                          ...       ...     ...
2042        1                          Pug  0.369275    True
2043        1             Blenheim spaniel  0.972494    True
2044        1                          Pug  0.943575    True
2045        1               French bulldog  0.999201    True
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
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

                               p2   p2_conf  p2_dog  \
0          Welsh springer spaniel  0.156665    True
1                         Redbone  0.074192    True
2                  German shepherd  0.138584    True
3             Rhodesian ridgeback  0.360687    True
4              Miniature pinscher  0.243682    True
5            Bernese mountain dog  0.263788    True
7                            Chow  0.058279    True
8                   Shopping cart  0.014594   False
```

```
9                 Miniature poodle  0.192305    True
10                Golden retriever  0.093718    True
11                   Gordon setter  0.174201    True
12                    Walker hound  0.175382    True
13                             Pug  0.404722    True
14                      Bloodhound  0.078260    True
15                           Lhasa  0.166192    True
16                  English setter  0.149842    True
19               Italian greyhound  0.111884    True
20                     Maltese dog  0.063064    True
22                              Ox  0.278407    True
23                Golden retriever  0.054787    True
24                        Malamute  0.147655    True
26     Soft-coated wheaten terrier  0.259551    True
27                       Chihuahua  0.009397    True
28         Black-and-tan coonhound  0.244220    True
30                     Toy terrier  0.148258    True
31                Blenheim spaniel  0.090346    True
32                        Pembroke  0.249394    True
33                           Llama  0.104109    True
34        Chesapeake bay retriever  0.056775   False
35                       Chihuahua  0.124680    True
...                            ...       ...     ...
2042                           Pug  0.265835    True
2043              Blenheim spaniel  0.006630    True
2044                           Pug  0.025286   False
2045                French bulldog  0.000361    True
2047                        Kuvasz  0.186136    True
2048                     Chihuahua  0.143528    True
2049                       Samoyed  0.035029    True
2050              Mexican hairless  0.275645   False
2051                      Pembroke  0.054950    True
2053                        Basset  0.087582    True
2054                French bulldog  0.000932    True
2055                      Pembroke  0.054950    True
2056                Siberian husky  0.166511    True
2057              Golden retriever  0.184172    True
2058              Golden retriever  0.120184    True
2059                       Whippet  0.194742    True
2060              Golden retriever  0.013834    True
2061                French bulldog  0.002129    True
2062                      Pembroke  0.027356    True
2063                French bulldog  0.151317    True
2064                       Samoyed  0.013884    True
2065                      Pembroke  0.451038    True
2066                 Irish terrier  0.193054    True
2067                    Pomeranian  0.178406    True
2068                   Appenzeller  0.199287    True
```

```
2069     Chesapeake bay retriever  0.116317    True
2070                       Basset  0.225770    True
2071                  Paper towel  0.168086    True
2072                    Chihuahua  0.078253    True
2073                    Chihuahua  0.090647    True

                               p3   p3_conf  p3_dog
0          Welsh springer spaniel  0.061428    True
1                         Redbone  0.072010    True
2                 German shepherd  0.116197    True
3             Rhodesian ridgeback  0.222752    True
4              Miniature pinscher  0.154629    True
5            Bernese mountain dog  0.016199    True
7                            Chow  0.054449   False
8                   Shopping cart  0.007959    True
9                Miniature poodle  0.082086    True
10               Golden retriever  0.072427    True
11                  Gordon setter  0.109454    True
12                   Walker hound  0.097471    True
13                            Pug  0.048960    True
14                     Bloodhound  0.075628    True
15                          Lhasa  0.089688    True
16                 English setter  0.133649    True
19              Italian greyhound  0.111152    True
20                    Maltese dog  0.025581    True
22                             Ox  0.102643    True
23               Golden retriever  0.014241    True
24                       Malamute  0.093412    True
26    Soft-coated wheaten terrier  0.206803    True
27                      Chihuahua  0.004577    True
28        Black-and-tan coonhound  0.173810    True
30                    Toy terrier  0.142860    True
31               Blenheim spaniel  0.001117    True
32                       Pembroke  0.241878    True
33                          Llama  0.062071   False
34       Chesapeake bay retriever  0.036763   False
35                      Chihuahua  0.044094    True
...                           ...       ...     ...
2042                          Pug  0.134697    True
2043             Blenheim spaniel  0.006239    True
2044                          Pug  0.002849   False
2045               French bulldog  0.000076    True
2047                       Kuvasz  0.086346    True
2048                    Chihuahua  0.032253   False
2049                      Samoyed  0.029705    True
2050             Mexican hairless  0.134203    True
2051                     Pembroke  0.038915    True
2053                       Basset  0.026236    True
```

```
2054              French bulldog  0.000903   True
2055                    Pembroke  0.038915   True
2056              Siberian husky  0.111411   True
2057            Golden retriever  0.073482   True
2058            Golden retriever  0.105506   True
2059                     Whippet  0.027351   True
2060            Golden retriever  0.007958   True
2061              French bulldog  0.001498   True
2062                    Pembroke  0.004633   True
2063              French bulldog  0.082981  False
2064                     Samoyed  0.008167   True
2065                    Pembroke  0.029248   True
2066               Irish terrier  0.118184   True
2067                  Pomeranian  0.076507   True
2068                 Appenzeller  0.193548  False
2069    Chesapeake bay retriever  0.076902  False
2070                      Basset  0.175219   True
2071                 Paper towel  0.040836  False
2072                   Chihuahua  0.031379   True
2073                   Chihuahua  0.068957   True

[1751 rows x 12 columns]
```

9 Drop not useful columns (ex. source / img_num)

```
In [69]: df_twi_enhan.drop(['source'], axis=1, inplace=True)
```

```
In [70]: df_breed.drop(['img_num'], axis=1, inplace=True)
```

```
/opt/conda/lib/python3.6/site-packages/pandas/core/frame.py:3697: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  errors=errors)
```

```
In [71]: df_breed.head(2)
```

```
Out[71]:              tweet_id                                          jpg_url  \
         0  666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
         1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg

                               p1   p1_conf  p1_dog                      p2   p2_conf  \
         0  Welsh springer spaniel  0.465074    True  Welsh springer spaniel  0.156665
         1                 Redbone  0.506826    True                 Redbone  0.074192

            p2_dog                      p3   p3_conf  p3_dog
         0    True  Welsh springer spaniel  0.061428    True
         1    True                 Redbone  0.072010    True
```

```
In [72]: df_twi_enhan.head(2)
```

```
Out[72]:              tweet_id in_reply_to_status_id in_reply_to_user_id  \
         0  892420643555336193                   nan                 nan
         1  892177421306343426                   nan                 nan

                     timestamp                                               text  \
         0 2017-08-01 16:23:56  This is Phineas. He's a mystical boy. Only eve...
         1 2017-08-01 00:17:27  This is Tilly. She's just checking pup on you...

                                                expanded_urls rating_scale_10     name  \
         0  https://twitter.com/dog_rates/status/892420643...              13  Phineas
         1  https://twitter.com/dog_rates/status/892177421...              13    Tilly

           dog_charac
         0        NaN
         1        NaN
```

```
In [ ]:
```

10. Merge three datasets into one.

```
In [73]: df_part = pd.merge(left=df_twi_enhan,right=df_tweets, left_on='tweet_id', right_on='id'
```

```
In [74]: df_part.head(1)
```

```
Out[74]:              tweet_id in_reply_to_status_id in_reply_to_user_id  \
         0  892420643555336193                   nan                 nan

                     timestamp                                               text  \
         0 2017-08-01 16:23:56  This is Phineas. He's a mystical boy. Only eve...

                                                expanded_urls rating_scale_10     name  \
         0  https://twitter.com/dog_rates/status/892420643...              13  Phineas

           dog_charac                  id  favorite_count  retweet_count retweeted
         0        NaN  892420643555336193         37353.0         8016.0     False
```

```
In [75]: df_part = df_part.drop(['id'], axis = 1)
```

```
In [76]: df_part.head(1)
```

```
Out[76]:              tweet_id in_reply_to_status_id in_reply_to_user_id  \
         0  892420643555336193                   nan                 nan

                     timestamp                                               text  \
         0 2017-08-01 16:23:56  This is Phineas. He's a mystical boy. Only eve...

                                                expanded_urls rating_scale_10     name  \
        0  https://twitter.com/dog_rates/status/892420643...              13  Phineas

          dog_charac  favorite_count  retweet_count retweeted
        0        NaN         37353.0         8016.0     False
```

```
In [77]: df_breed.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1751 entries, 0 to 2073
Data columns (total 11 columns):
tweet_id    1751 non-null int64
jpg_url     1751 non-null object
p1          1751 non-null object
p1_conf     1751 non-null float64
p1_dog      1751 non-null bool
p2          1751 non-null object
p2_conf     1751 non-null float64
p2_dog      1751 non-null bool
p3          1751 non-null object
p3_conf     1751 non-null float64
p3_dog      1751 non-null bool
dtypes: bool(3), float64(3), int64(1), object(4)
memory usage: 128.2+ KB
```

```
In [78]: df_breed['tweet_id'] = df_breed['tweet_id'].astype(str)
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  """Entry point for launching an IPython kernel.
```

```
In [79]: df_all = pd.merge(left=df_part,right=df_breed, left_on='tweet_id', right_on='tweet_id',
```

```
In [80]: df_all.columns
```

```
Out[80]: Index(['tweet_id', 'in_reply_to_status_id', 'in_reply_to_user_id', 'timestamp',
                'text', 'expanded_urls', 'rating_scale_10', 'name', 'dog_charac',
                'favorite_count', 'retweet_count', 'retweeted', 'jpg_url', 'p1',
                'p1_conf', 'p1_dog', 'p2', 'p2_conf', 'p2_dog', 'p3', 'p3_conf',
                'p3_dog'],
               dtype='object')
```

```
In [81]: df_part.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
```

```
Data columns (total 12 columns):
tweet_id                 2356 non-null object
in_reply_to_status_id    2356 non-null object
in_reply_to_user_id      2356 non-null object
timestamp                2356 non-null datetime64[ns]
text                     2356 non-null object
expanded_urls            2297 non-null object
rating_scale_10          2356 non-null object
name                     1502 non-null object
dog_charac               399 non-null object
favorite_count           2333 non-null float64
retweet_count            2333 non-null float64
retweeted                2333 non-null object
dtypes: datetime64[ns](1), float64(2), object(9)
memory usage: 239.3+ KB
```

```
In [82]: df_all.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 22 columns):
tweet_id                 2356 non-null object
in_reply_to_status_id    2356 non-null object
in_reply_to_user_id      2356 non-null object
timestamp                2356 non-null datetime64[ns]
text                     2356 non-null object
expanded_urls            2297 non-null object
rating_scale_10          2356 non-null object
name                     1502 non-null object
dog_charac               399 non-null object
favorite_count           2333 non-null float64
retweet_count            2333 non-null float64
retweeted                2333 non-null object
jpg_url                  1751 non-null object
p1                       1751 non-null object
p1_conf                  1751 non-null float64
p1_dog                   1751 non-null object
p2                       1751 non-null object
p2_conf                  1751 non-null float64
p2_dog                   1751 non-null object
p3                       1751 non-null object
p3_conf                  1751 non-null float64
p3_dog                   1751 non-null object
dtypes: datetime64[ns](1), float64(5), object(16)
memory usage: 423.3+ KB
```

0.1.9 Analyzing and Visualizing Data

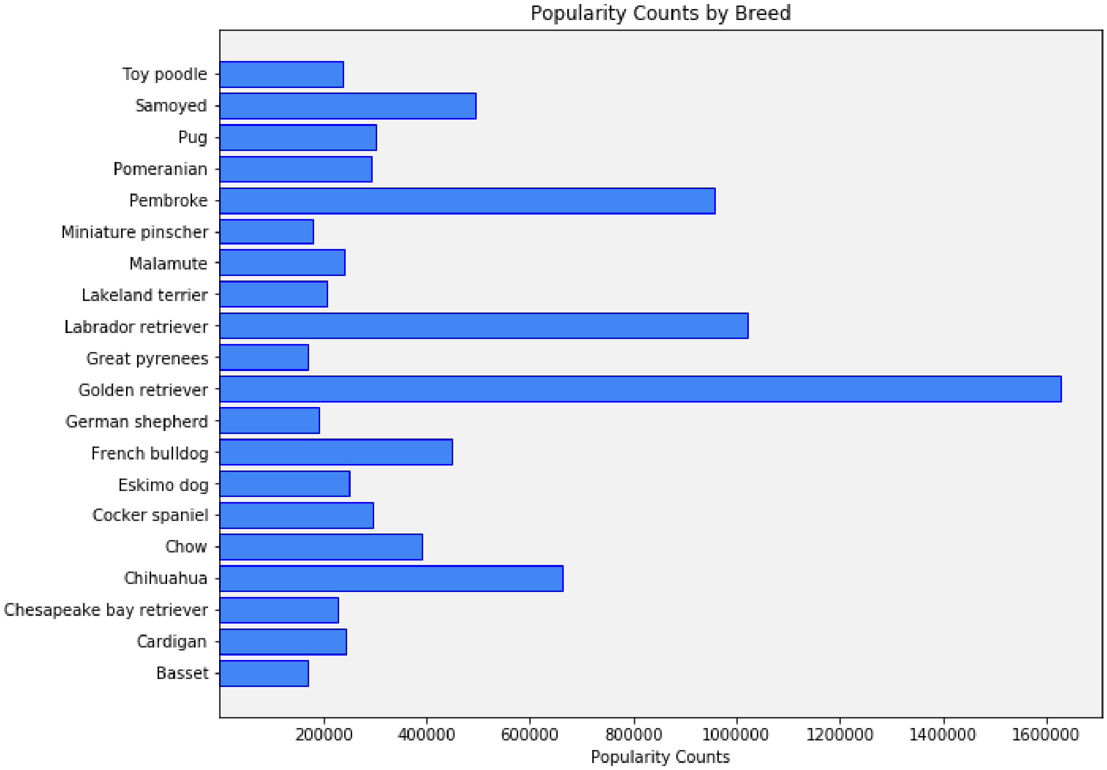
```
In [83]: df_all.head()
```

```
Out[83]:             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                                               text  \
         0 2017-08-01 16:23:56  This is Phineas. He's a mystical boy. Only eve...
         1 2017-08-01 00:17:27  This is Tilly. She's just checking pup on you...
         2 2017-07-31 00:18:03  This is Archie. He is a rare Norwegian Pouncin...
         3 2017-07-30 15:58:51  This is Darla. She commenced a snooze mid meal...
         4 2017-07-29 16:00:24  This is Franklin. He would like you to stop ca...

                                                expanded_urls rating_scale_10  \
         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

                name dog_charac  favorite_count  ...    \
         0   Phineas        NaN         37353.0  ...
         1     Tilly        NaN         32098.0  ...
         2    Archie        NaN         24197.0  ...
         3     Darla        NaN         40642.0  ...
         4  Franklin        NaN         38875.0  ...

                                                    jpg_url           p1   p1_conf  \
         0                                              NaN          NaN       NaN
         1  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg    Chihuahua  0.323581
         2  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg    Chihuahua  0.716012
         3  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg  Paper towel  0.170278
         4  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg       Basset  0.555712

          p1_dog           p2   p2_conf p2_dog           p3   p3_conf p3_dog
         0    NaN          NaN       NaN    NaN          NaN       NaN    NaN
         1   True    Chihuahua  0.090647   True    Chihuahua  0.068957   True
         2   True    Chihuahua  0.078253   True    Chihuahua  0.031379   True
         3  False  Paper towel  0.168086   True  Paper towel  0.040836  False
         4   True       Basset  0.225770   True       Basset  0.175219   True

         [5 rows x 22 columns]
```

```
In [84]: df_all.to_csv('twitter_archive_master.csv')
In [85]: df_all=pd.read_csv('twitter_archive_master.csv')

In [86]: df_all.shape

Out[86]: (2356, 23)

In [87]: df_all.describe()

Out[87]:         Unnamed: 0      tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         count  2356.000000  2.356000e+03           7.800000e+01         7.800000e+01
         mean   1177.500000  7.427716e+17           7.455079e+17         2.014171e+16
         std     680.262939  6.856705e+16           7.582492e+16         1.252797e+17
         min       0.000000  6.660209e+17           6.658147e+17         1.185634e+07
         25%     588.750000  6.783989e+17           6.757419e+17         3.086374e+08
         50%    1177.500000  7.196279e+17           7.038708e+17         4.196984e+09
         75%    1766.250000  7.993373e+17           8.257804e+17         4.196984e+09
         max    2355.000000  8.924206e+17           8.862664e+17         8.405479e+17

                rating_scale_10  favorite_count  retweet_count      p1_conf  \
         count      2356.000000     2333.000000     2333.00000  1751.000000
         mean         13.063680     7785.007715     2819.18817     0.604207
         std          45.839085    12077.971151     4769.51836     0.265911
         min           0.000000        0.000000        1.00000     0.044333
         25%          10.000000     1357.000000      567.00000     0.377079
         50%          11.000000     3389.000000     1317.00000     0.605304
         75%          12.000000     9549.000000     3287.00000     0.848720
         max        1776.000000   161254.000000    80960.00000     0.999984

                    p2_conf       p3_conf
         count  1751.000000  1.751000e+03
         mean      0.137715  6.161188e-02
         std       0.101297  5.192022e-02
         min       0.000010  2.160900e-07
         25%       0.055020  1.608055e-02
         50%       0.121811  5.000780e-02
         75%       0.199439  9.480810e-02
         max       0.467678  2.734190e-01

In [ ]:
```

1. Most popular breeds.

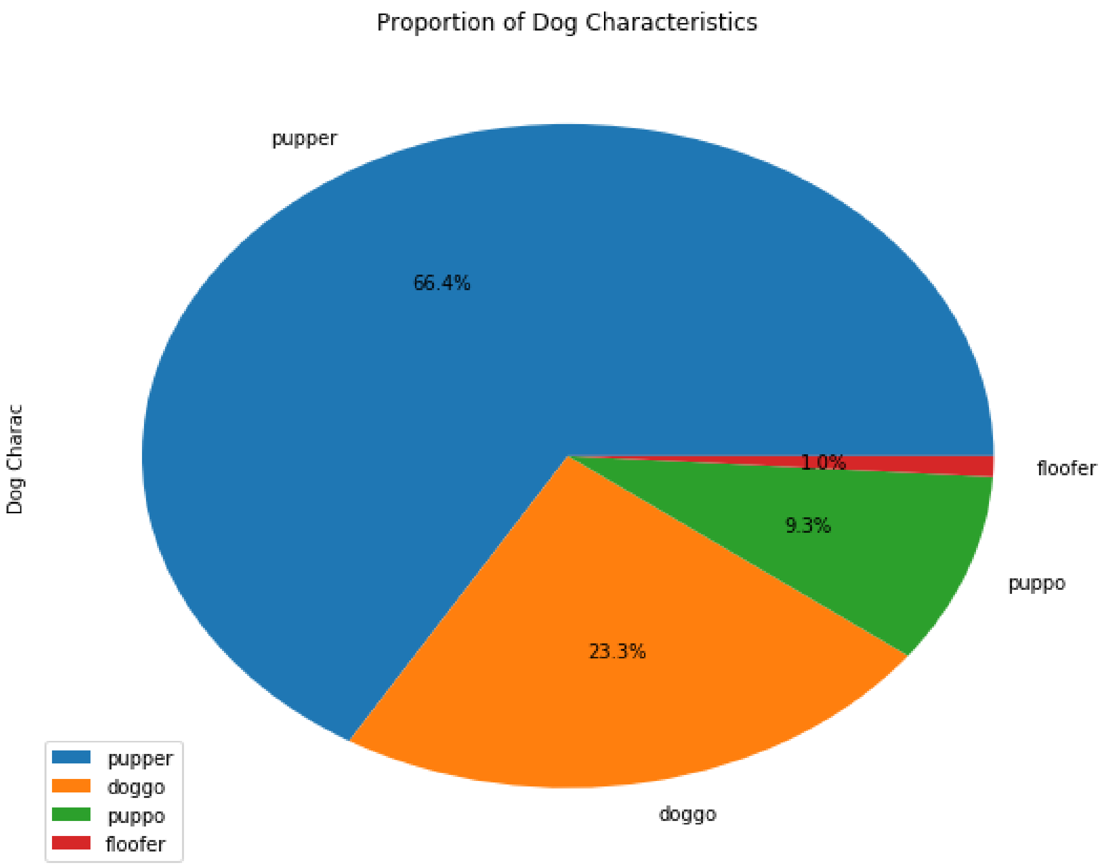
```
In [88]: df_popular = df_all.groupby('p1')['favorite_count'].sum().reset_index()
         df_sorted = df_popular.sort_values('favorite_count', ascending=False).head(20)
         ser_pop = df_sorted['favorite_count']
         ser_breed = df_sorted['p1']

         fig, ax = plt.subplots(figsize=(10,8))
         fav = plt.barh(ser_breed, ser_pop, color = "#4285F4", edgecolor = ['Blue']*len(ser_bree
        ax.set_facecolor('#F2F2F2')
        plt.xlabel('Popularity Counts')
        plt.title('Popularity Counts by Breed')
        plt.xticks(np.arange(200000, 1800000, 200000))
        plt.show();
```



Golden retriever is the most popular breed among the users with more than 1600000 likes, rank 2 breed is Labrador Retriever (about 1000000 likes) and Pembroke is rank 3 (slightly less than 1000000 likes).

2. Dog Characteristics

```
In [89]: df_charac = df_all[df_all['dog_charac'] != "None"]
         fig, ax = plt.subplots(figsize=(10,8))

         df_charac['dog_charac'].value_counts().plot(kind = 'pie', ax = ax, label = 'Dog Charac'
         plt.title('Proportion of Dog Characteristics')
         plt.legend();
```



The most frequent dog chrarcteristic is "pupper" with about 66%, the second one is "doggo" (about 23%). The rest of two "puppo and floofer" with about 10% altogether.

3. Rating

```
In [93]: df_all['rating_scale_10'].value_counts()
```

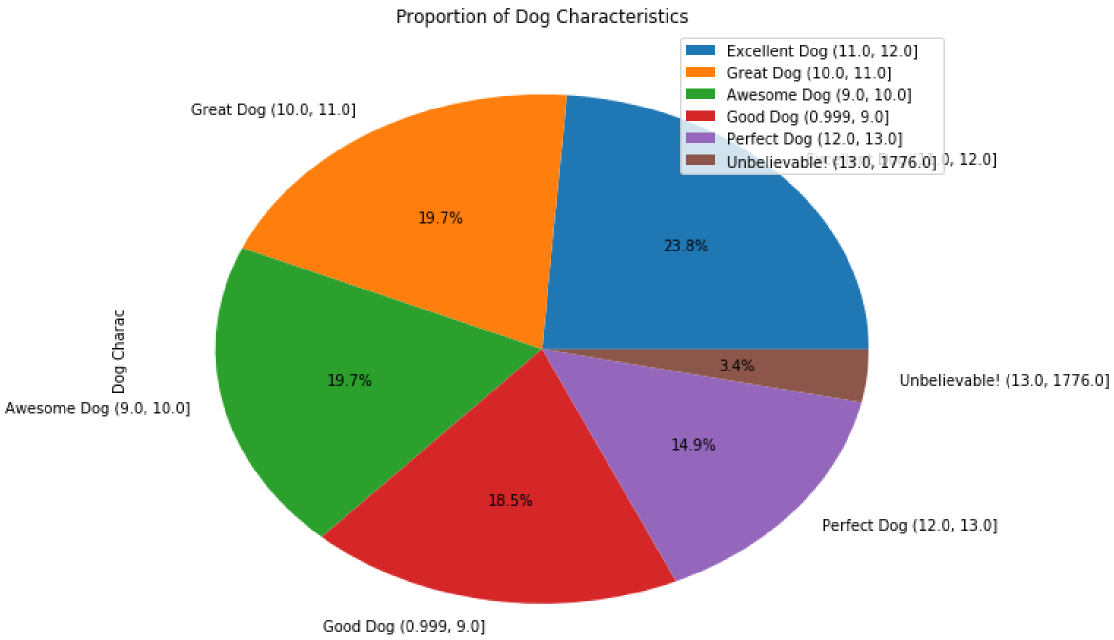
```
Out[93]: 12.00      558
         11.00      464
         10.00      461
         13.00      351
         9.00       158
         8.00       102
         7.00        55
         14.00       54
         5.00        35
         6.00        32
         3.00        19
         4.00        17
```

```
        2.00        9
        1.00        9
        0.00        2
        420.00      2
        9.75        2
        15.00       2
        960.00      1
        84.00       1
        24.00       1
        17.00       1
        13.50       1
        143.00      1
        121.00      1
        80.00       1
        182.00      1
        165.00      1
        45.00       1
        204.00      1
        1776.00     1
        666.00      1
        99.00       1
        11.27       1
        11.26       1
        88.00       1
        144.00      1
        9.50        1
        20.00       1
        44.00       1
        60.00       1
        50.00       1
        Name: rating_scale_10, dtype: int64
```

```
In [90]: df_rate = df_all[df_all['rating_scale_10'] > 0]
```

```
In [91]: rate_label = ['Good Dog (0.999, 9.0]', 'Awesome Dog (9.0, 10.0]', 'Great Dog (10.0, 11.

         rate_bins = pd.qcut(df_rate['rating_scale_10'], 6, labels = rate_label)
```

```
In [92]: fig, ax = plt.subplots(figsize=(10,8))
         rate_bins.value_counts().plot(kind = 'pie', ax = ax, label = 'Dog Charac', autopct='%1.
         plt.title('Proportion of Dog Characteristics')
         plt.legend();
```



I analyzed the rating scale, and divided the score into different buckets with assigned names.

Good Dog: 0.999<x<=9.0,

Awesome Dog: 9.0<x<=10.0,

Great Dog: 10.0<x<=11.0,

Excellent Dog 11.0<x<=12.0,

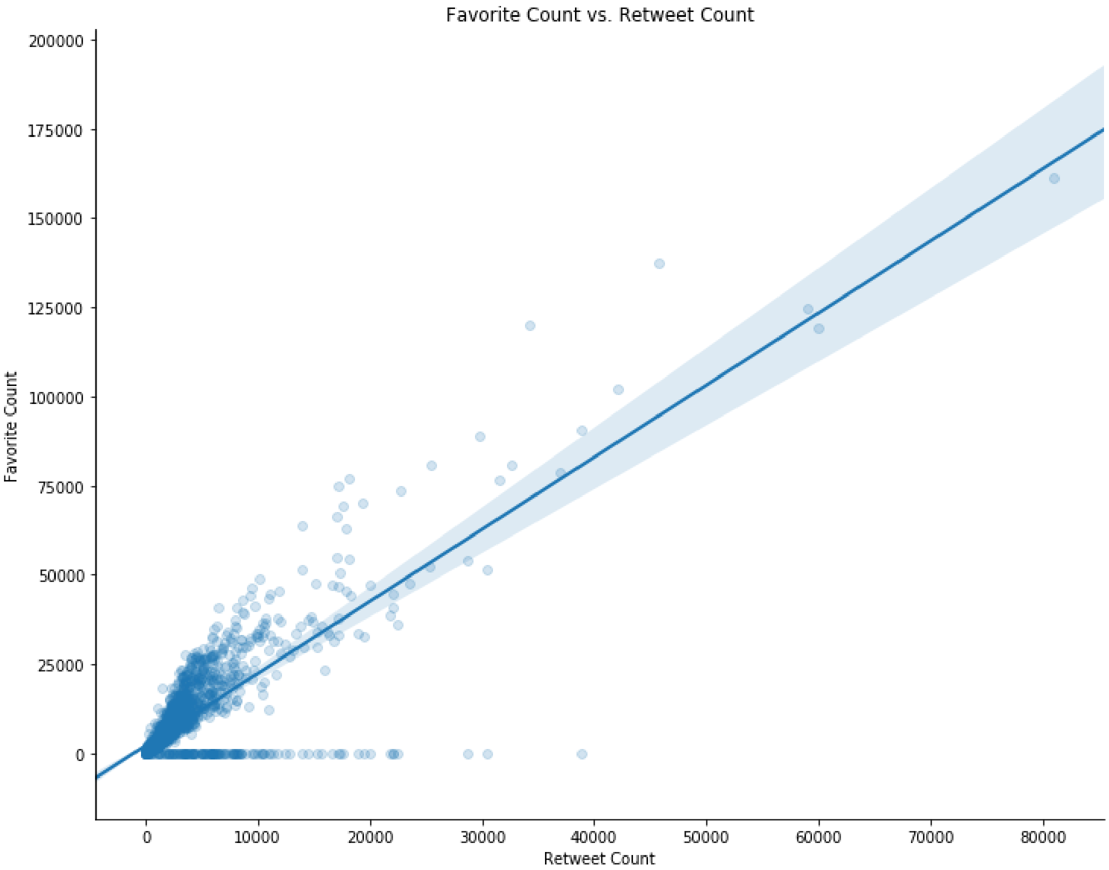
Perfect Dog: 12.0<x<=13.0,

Unbelievable!: 13.0<x<=1776.0

According to the results, most of the dogs are rated between 0 and 13, only 3.4% of the dogs are rated higher than 13. It has outliers, such as 1176, 960, but still more than half the dogs have a rating greater than 10. Most popular rating is between 11 and 12.

4. Favorite counts vs Retweet counts

```
In [99]: sns.lmplot(x="retweet_count",
                    y="favorite_count",
                    data=df_all,
                size = 8,
                aspect=1.3,
                scatter_kws={'alpha':1/5})
    plt.title('Favorite Count vs. Retweet Count')
    plt.xlabel('Retweet Count')
    plt.ylabel('Favorite Count');
```



Favorite counts and retweet counts have a positive correlation. The majority of the data falls below 35000 favorite counts and 10000 retweet counts, so for every 3-4 favourites, there is about 1 retweets.

```
In [ ]: from subprocess import call
        call(['python', '-m', 'nbconvert', 'wrangle_act.ipynb'])

In [ ]:
```