

# wrangle\_act

January 30, 2019

## 1 Import Libraries

```
In [2]: import pandas as pd
import numpy as np
from numpy import nan
import tweepy
import requests
import json
import os
import re
import matplotlib.pyplot as plt
import datetime
import sys
```

## 2 Gather

```
In [3]: archive = pd.read_csv("./twitter-archive-enhanced.csv")
```

```
In [4]: ImageURL = "https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-pred
```

```
response = requests.get(ImageURL)
```

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

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

```
In [5]: # Personal API keys, secrets, and tokens have been replaced with placeholders
```

```
consumer_key = 'xvWlbHKWDXDHUYwFJiW4EpKQoe'
```

```
consumer_secret = 'M7xWLk2p2lemGtIsksz1NFau9MfRZXbQJkEVratJ5NMU1xeIrD'
```

```
access_token = '182197942-XPars3NGL01BI9BbMtJqoa65piK2atTmTa62LB87'
```

```
access_secret = 'MiXD0wIf5yOLNP1deDX72ew9JoqL69P1Z92HV5gmoEtFk'
```

```
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
```

```
auth.set_access_token(access_token, access_secret)
```

```

api = tweepy.API(auth_handler=auth, wait_on_rate_limit=True, wait_on_rate_limit_notify=True)

In [6]: id_errors = []
def get_all_tweets(screen_name):
    new_tweets = api.user_timeline(screen_name = screen_name, count=200)
    alltweets = []
    alltweets.extend(new_tweets)

    oldest = alltweets[-1].id - 1

    while len(new_tweets) > 0:
        try:
            new_tweets = api.user_timeline(screen_name = screen_name, count=200, max_id=oldest)

            alltweets.extend(new_tweets)

            oldest = alltweets[-1].id - 1

            print("%s completed" % (len(alltweets)))

        except Exception as e:
            print(new_tweets.id_str + ": " + str(e))
            id_errors.append(new_tweets)
            continue

    outtweets = [[tweet.id_str, tweet.retweet_count, tweet.favorite_count] for tweet in alltweets]
    df = pd.DataFrame.from_dict(outtweets)

    df.to_csv('./tweet_json.txt', header=True, index=None, sep=' ', mode='w')
    print('True')

if __name__ == '__main__':
    get_all_tweets("dog_rates")

399 completed
599 completed
799 completed
999 completed
1199 completed
1399 completed
1599 completed
1799 completed
1999 completed
2157 completed
2157 completed
True

```

```
In [7]: tweet_json = pd.read_csv('./tweet_json.txt', sep=' ')
        tweet_json.columns = ["tweet_id", "retweet_count", "favorite_count"]
```

### 3 Access

```
In [8]: archive
```

```
Out[8]:
```

|      | tweet_id           | in_reply_to_status_id | in_reply_to_user_id \ |
|------|--------------------|-----------------------|-----------------------|
| 0    | 892420643555336193 | NaN                   | NaN                   |
| 1    | 892177421306343426 | NaN                   | NaN                   |
| 2    | 891815181378084864 | NaN                   | NaN                   |
| 3    | 891689557279858688 | NaN                   | NaN                   |
| 4    | 891327558926688256 | NaN                   | NaN                   |
| 5    | 891087950875897856 | NaN                   | NaN                   |
| 6    | 890971913173991426 | NaN                   | NaN                   |
| 7    | 890729181411237888 | NaN                   | NaN                   |
| 8    | 890609185150312448 | NaN                   | NaN                   |
| 9    | 890240255349198849 | NaN                   | NaN                   |
| 10   | 890006608113172480 | NaN                   | NaN                   |
| 11   | 889880896479866881 | NaN                   | NaN                   |
| 12   | 889665388333682689 | NaN                   | NaN                   |
| 13   | 889638837579907072 | NaN                   | NaN                   |
| 14   | 889531135344209921 | NaN                   | NaN                   |
| 15   | 889278841981685760 | NaN                   | NaN                   |
| 16   | 888917238123831296 | NaN                   | NaN                   |
| 17   | 888804989199671297 | NaN                   | NaN                   |
| 18   | 888554962724278272 | NaN                   | NaN                   |
| 19   | 888202515573088257 | NaN                   | NaN                   |
| 20   | 888078434458587136 | NaN                   | NaN                   |
| 21   | 887705289381826560 | NaN                   | NaN                   |
| 22   | 887517139158093824 | NaN                   | NaN                   |
| 23   | 887473957103951883 | NaN                   | NaN                   |
| 24   | 887343217045368832 | NaN                   | NaN                   |
| 25   | 887101392804085760 | NaN                   | NaN                   |
| 26   | 886983233522544640 | NaN                   | NaN                   |
| 27   | 886736880519319552 | NaN                   | NaN                   |
| 28   | 886680336477933568 | NaN                   | NaN                   |
| 29   | 886366144734445568 | NaN                   | NaN                   |
| ...  | ...                | ...                   | ...                   |
| 2326 | 666411507551481857 | NaN                   | NaN                   |
| 2327 | 666407126856765440 | NaN                   | NaN                   |
| 2328 | 666396247373291520 | NaN                   | NaN                   |
| 2329 | 666373753744588802 | NaN                   | NaN                   |
| 2330 | 666362758909284353 | NaN                   | NaN                   |
| 2331 | 666353288456101888 | NaN                   | NaN                   |
| 2332 | 666345417576210432 | NaN                   | NaN                   |
| 2333 | 666337882303524864 | NaN                   | NaN                   |

|      |                    |     |     |
|------|--------------------|-----|-----|
| 2334 | 666293911632134144 | NaN | NaN |
| 2335 | 666287406224695296 | NaN | NaN |
| 2336 | 666273097616637952 | NaN | NaN |
| 2337 | 666268910803644416 | NaN | NaN |
| 2338 | 666104133288665088 | NaN | NaN |
| 2339 | 666102155909144576 | NaN | NaN |
| 2340 | 666099513787052032 | NaN | NaN |
| 2341 | 666094000022159362 | NaN | NaN |
| 2342 | 666082916733198337 | NaN | NaN |
| 2343 | 666073100786774016 | NaN | NaN |
| 2344 | 666071193221509120 | NaN | NaN |
| 2345 | 666063827256086533 | NaN | NaN |
| 2346 | 666058600524156928 | NaN | NaN |
| 2347 | 666057090499244032 | NaN | NaN |
| 2348 | 666055525042405380 | NaN | NaN |
| 2349 | 666051853826850816 | NaN | NaN |
| 2350 | 666050758794694657 | NaN | NaN |
| 2351 | 666049248165822465 | NaN | NaN |
| 2352 | 666044226329800704 | NaN | NaN |
| 2353 | 666033412701032449 | NaN | NaN |
| 2354 | 666029285002620928 | NaN | NaN |
| 2355 | 666020888022790149 | NaN | NaN |

|    | timestamp \               |
|----|---------------------------|
| 0  | 2017-08-01 16:23:56 +0000 |
| 1  | 2017-08-01 00:17:27 +0000 |
| 2  | 2017-07-31 00:18:03 +0000 |
| 3  | 2017-07-30 15:58:51 +0000 |
| 4  | 2017-07-29 16:00:24 +0000 |
| 5  | 2017-07-29 00:08:17 +0000 |
| 6  | 2017-07-28 16:27:12 +0000 |
| 7  | 2017-07-28 00:22:40 +0000 |
| 8  | 2017-07-27 16:25:51 +0000 |
| 9  | 2017-07-26 15:59:51 +0000 |
| 10 | 2017-07-26 00:31:25 +0000 |
| 11 | 2017-07-25 16:11:53 +0000 |
| 12 | 2017-07-25 01:55:32 +0000 |
| 13 | 2017-07-25 00:10:02 +0000 |
| 14 | 2017-07-24 17:02:04 +0000 |
| 15 | 2017-07-24 00:19:32 +0000 |
| 16 | 2017-07-23 00:22:39 +0000 |
| 17 | 2017-07-22 16:56:37 +0000 |
| 18 | 2017-07-22 00:23:06 +0000 |
| 19 | 2017-07-21 01:02:36 +0000 |
| 20 | 2017-07-20 16:49:33 +0000 |
| 21 | 2017-07-19 16:06:48 +0000 |
| 22 | 2017-07-19 03:39:09 +0000 |
| 23 | 2017-07-19 00:47:34 +0000 |

```

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

```

```

                                source \
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...

```

[illegible]

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 Ralpus. 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/ws2AMaNPW     | 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   | <a href="https://twitter.com/dog_rates/status/889665388...">https://twitter.com/dog_rates/status/889665388...</a>                   | 13  |
| 13   | <a href="https://twitter.com/dog_rates/status/889638837...">https://twitter.com/dog_rates/status/889638837...</a>                   | 12  |
| 14   | <a href="https://twitter.com/dog_rates/status/889531135...">https://twitter.com/dog_rates/status/889531135...</a>                   | 13  |
| 15   | <a href="https://twitter.com/dog_rates/status/889278841...">https://twitter.com/dog_rates/status/889278841...</a>                   | 13  |
| 16   | <a href="https://twitter.com/dog_rates/status/888917238...">https://twitter.com/dog_rates/status/888917238...</a>                   | 12  |
| 17   | <a href="https://twitter.com/dog_rates/status/888804989...">https://twitter.com/dog_rates/status/888804989...</a>                   | 13  |
| 18   | <a href="https://twitter.com/dog_rates/status/888554962...">https://twitter.com/dog_rates/status/888554962...</a>                   | 13  |
| 19   | <a href="https://twitter.com/dog_rates/status/887473957...">https://twitter.com/dog_rates/status/887473957...</a>                   | 13  |
| 20   | <a href="https://twitter.com/dog_rates/status/888078434...">https://twitter.com/dog_rates/status/888078434...</a>                   | 12  |
| 21   | <a href="https://twitter.com/dog_rates/status/887705289...">https://twitter.com/dog_rates/status/887705289...</a>                   | 13  |
| 22   | <a href="https://twitter.com/dog_rates/status/887517139...">https://twitter.com/dog_rates/status/887517139...</a>                   | 14  |
| 23   | <a href="https://twitter.com/dog_rates/status/887473957...">https://twitter.com/dog_rates/status/887473957...</a>                   | 13  |
| 24   | <a href="https://twitter.com/dog_rates/status/887343217...">https://twitter.com/dog_rates/status/887343217...</a>                   | 13  |
| 25   | <a href="https://twitter.com/dog_rates/status/887101392...">https://twitter.com/dog_rates/status/887101392...</a>                   | 12  |
| 26   | <a href="https://twitter.com/dog_rates/status/886983233...">https://twitter.com/dog_rates/status/886983233...</a>                   | 13  |
| 27   | <a href="https://www.gofundme.com/mingusneedsus">https://www.gofundme.com/mingusneedsus</a> , <a href="https://...">https://...</a> | 13  |
| 28   | <a href="https://twitter.com/dog_rates/status/886680336...">https://twitter.com/dog_rates/status/886680336...</a>                   | 13  |
| 29   | <a href="https://twitter.com/dog_rates/status/886366144...">https://twitter.com/dog_rates/status/886366144...</a>                   | 12  |
| ...  | ...   | ... |
| 2326 | <a href="https://twitter.com/dog_rates/status/666411507...">https://twitter.com/dog_rates/status/666411507...</a>                   | 2   |
| 2327 | <a href="https://twitter.com/dog_rates/status/666407126...">https://twitter.com/dog_rates/status/666407126...</a>                   | 7   |
| 2328 | <a href="https://twitter.com/dog_rates/status/666396247...">https://twitter.com/dog_rates/status/666396247...</a>                   | 9   |
| 2329 | <a href="https://twitter.com/dog_rates/status/666373753...">https://twitter.com/dog_rates/status/666373753...</a>                   | 11  |
| 2330 | <a href="https://twitter.com/dog_rates/status/666362758...">https://twitter.com/dog_rates/status/666362758...</a>                   | 6   |
| 2331 | <a href="https://twitter.com/dog_rates/status/666353288...">https://twitter.com/dog_rates/status/666353288...</a>                   | 8   |
| 2332 | <a href="https://twitter.com/dog_rates/status/666345417...">https://twitter.com/dog_rates/status/666345417...</a>                   | 10  |
| 2333 | <a href="https://twitter.com/dog_rates/status/666337882...">https://twitter.com/dog_rates/status/666337882...</a>                   | 9   |
| 2334 | <a href="https://twitter.com/dog_rates/status/666293911...">https://twitter.com/dog_rates/status/666293911...</a>                   | 3   |
| 2335 | <a href="https://twitter.com/dog_rates/status/666287406...">https://twitter.com/dog_rates/status/666287406...</a>                   | 1   |
| 2336 | <a href="https://twitter.com/dog_rates/status/666273097...">https://twitter.com/dog_rates/status/666273097...</a>                   | 11  |
| 2337 | <a href="https://twitter.com/dog_rates/status/666268910...">https://twitter.com/dog_rates/status/666268910...</a>                   | 10  |
| 2338 | <a href="https://twitter.com/dog_rates/status/666104133...">https://twitter.com/dog_rates/status/666104133...</a>                   | 1   |
| 2339 | <a href="https://twitter.com/dog_rates/status/666102155...">https://twitter.com/dog_rates/status/666102155...</a>                   | 11  |
| 2340 | <a href="https://twitter.com/dog_rates/status/666099513...">https://twitter.com/dog_rates/status/666099513...</a>                   | 8   |
| 2341 | <a href="https://twitter.com/dog_rates/status/666094000...">https://twitter.com/dog_rates/status/666094000...</a>                   | 9   |
| 2342 | <a href="https://twitter.com/dog_rates/status/666082916...">https://twitter.com/dog_rates/status/666082916...</a>                   | 6   |
| 2343 | <a href="https://twitter.com/dog_rates/status/666073100...">https://twitter.com/dog_rates/status/666073100...</a>                   | 10  |
| 2344 | <a href="https://twitter.com/dog_rates/status/666071193...">https://twitter.com/dog_rates/status/666071193...</a>                   | 9   |
| 2345 | <a href="https://twitter.com/dog_rates/status/666063827...">https://twitter.com/dog_rates/status/666063827...</a>                   | 10  |
| 2346 | <a href="https://twitter.com/dog_rates/status/666058600...">https://twitter.com/dog_rates/status/666058600...</a>                   | 8   |
| 2347 | <a href="https://twitter.com/dog_rates/status/666057090...">https://twitter.com/dog_rates/status/666057090...</a>                   | 9   |
| 2348 | <a href="https://twitter.com/dog_rates/status/666055525...">https://twitter.com/dog_rates/status/666055525...</a>                   | 10  |
| 2349 | <a href="https://twitter.com/dog_rates/status/666051853...">https://twitter.com/dog_rates/status/666051853...</a>                   | 2   |
| 2350 | <a href="https://twitter.com/dog_rates/status/666050758...">https://twitter.com/dog_rates/status/666050758...</a>                   | 10  |
| 2351 | <a href="https://twitter.com/dog_rates/status/666049248...">https://twitter.com/dog_rates/status/666049248...</a>                   | 5   |
| 2352 | <a href="https://twitter.com/dog_rates/status/666044226...">https://twitter.com/dog_rates/status/666044226...</a>                   | 6   |
| 2353 | <a href="https://twitter.com/dog_rates/status/666033412...">https://twitter.com/dog_rates/status/666033412...</a>                   | 9   |
| 2354 | <a href="https://twitter.com/dog_rates/status/666029285...">https://twitter.com/dog_rates/status/666029285...</a>                   | 7   |

|      | 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]

archive table has 2356 rows and 17 columns . 1. These dog stage columns are too redundant (doggo, floofer, pupper, and puppo, ...) which result into a tidyness issue. --> merge them into one dog\_stage column. 2. in\_reply\_to\_status\_id,in\_reply\_to\_user\_id,retweeted\_status\_id,retweeted\_status\_user\_id and retweeted\_status\_timestamp have a lot of missing value --> examine more in programmatical portion. 3. Timestamp variable has the suffix '+0000'. 4. Source column's value has "<a href=" and ".r" around text. some rows have the same source url. 5. Some naame variable has 'a', 'an' and 'the'. 6. The value inside 'text' is incomplete.

In [9]: imagePrediction

```
Out[9]:
```

|    | tweet_id           | jpg_url \   |
|----|--------------------|---|
| 0  | 666020888022790149 | <a href="https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg">https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg</a>   |
| 1  | 666029285002620928 | <a href="https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg">https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg</a>   |
| 2  | 666033412701032449 | <a href="https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg">https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg</a>   |
| 3  | 666044226329800704 | <a href="https://pbs.twimg.com/media/CT5Dr8HUEAA-1Eu.jpg">https://pbs.twimg.com/media/CT5Dr8HUEAA-1Eu.jpg</a>   |
| 4  | 666049248165822465 | <a href="https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg">https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg</a>   |
| 5  | 666050758794694657 | <a href="https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg">https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg</a>   |
| 6  | 666051853826850816 | <a href="https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg">https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg</a>   |
| 7  | 666055525042405380 | <a href="https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg">https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg</a>   |
| 8  | 666057090499244032 | <a href="https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg">https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg</a>   |
| 9  | 666058600524156928 | <a href="https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg">https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg</a>   |
| 10 | 666063827256086533 | <a href="https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg">https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg</a>   |
| 11 | 666071193221509120 | <a href="https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg">https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg</a>   |
| 12 | 666073100786774016 | <a href="https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg">https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg</a>   |
| 13 | 666082916733198337 | <a href="https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg">https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg</a> |
| 14 | 666094000022159362 | <a href="https://pbs.twimg.com/media/CT5w9gUW4AAAsBNN.jpg">https://pbs.twimg.com/media/CT5w9gUW4AAAsBNN.jpg</a> |
| 15 | 666099513787052032 | <a href="https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg">https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg</a>   |

|      |                    |   |
|------|--------------------|---|
| 16   | 666102155909144576 | <a href="https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg">https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg</a>     |
| 17   | 666104133288665088 | <a href="https://pbs.twimg.com/media/CT56LSZWoAA1Jj2.jpg">https://pbs.twimg.com/media/CT56LSZWoAA1Jj2.jpg</a>     |
| 18   | 666268910803644416 | <a href="https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg">https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg</a>     |
| 19   | 666273097616637952 | <a href="https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg">https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg</a>     |
| 20   | 666287406224695296 | <a href="https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg">https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg</a>     |
| 21   | 666293911632134144 | <a href="https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg">https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg</a>     |
| 22   | 666337882303524864 | <a href="https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg">https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg</a>     |
| 23   | 666345417576210432 | <a href="https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg">https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg</a>     |
| 24   | 666353288456101888 | <a href="https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg">https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg</a>     |
| 25   | 666362758909284353 | <a href="https://pbs.twimg.com/media/CT91XGsUcAAyUft.jpg">https://pbs.twimg.com/media/CT91XGsUcAAyUft.jpg</a>     |
| 26   | 666373753744588802 | <a href="https://pbs.twimg.com/media/CT9vZEYWUAA1Z05.jpg">https://pbs.twimg.com/media/CT9vZEYWUAA1Z05.jpg</a>     |
| 27   | 666396247373291520 | <a href="https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg">https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg</a>     |
| 28   | 666407126856765440 | <a href="https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg">https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg</a>     |
| 29   | 666411507551481857 | <a href="https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg">https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg</a>     |
| ...  | ...                | ...   |
| 2045 | 886366144734445568 | <a href="https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg">https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg</a>     |
| 2046 | 886680336477933568 | <a href="https://pbs.twimg.com/media/DE4fEDzWAAAYHMM.jpg">https://pbs.twimg.com/media/DE4fEDzWAAAYHMM.jpg</a>     |
| 2047 | 886736880519319552 | <a href="https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg">https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg</a>     |
| 2048 | 886983233522544640 | <a href="https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.jpg">https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.jpg</a>     |
| 2049 | 887101392804085760 | <a href="https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg">https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg</a>     |
| 2050 | 887343217045368832 | <a href="https://pbs.twimg.com/ext_tw_video_thumb/88734...">https://pbs.twimg.com/ext_tw_video_thumb/88734...</a> |
| 2051 | 887473957103951883 | <a href="https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg">https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg</a>     |
| 2052 | 887517139158093824 | <a href="https://pbs.twimg.com/ext_tw_video_thumb/88751...">https://pbs.twimg.com/ext_tw_video_thumb/88751...</a> |
| 2053 | 887705289381826560 | <a href="https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg">https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg</a>     |
| 2054 | 888078434458587136 | <a href="https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg">https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg</a>     |
| 2055 | 888202515573088257 | <a href="https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg">https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg</a>     |
| 2056 | 888554962724278272 | <a href="https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg">https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg</a>     |
| 2057 | 888804989199671297 | <a href="https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg">https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg</a>     |
| 2058 | 888917238123831296 | <a href="https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg">https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg</a>     |
| 2059 | 889278841981685760 | <a href="https://pbs.twimg.com/ext_tw_video_thumb/88927...">https://pbs.twimg.com/ext_tw_video_thumb/88927...</a> |
| 2060 | 889531135344209921 | <a href="https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg">https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg</a>     |
| 2061 | 889638837579907072 | <a href="https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg">https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg</a>     |
| 2062 | 889665388333682689 | <a href="https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg">https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg</a>     |
| 2063 | 889880896479866881 | <a href="https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg">https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg</a>     |
| 2064 | 890006608113172480 | <a href="https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg">https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg</a>     |
| 2065 | 890240255349198849 | <a href="https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg">https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg</a>     |
| 2066 | 890609185150312448 | <a href="https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg">https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg</a>     |
| 2067 | 890729181411237888 | <a href="https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg">https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg</a>     |
| 2068 | 890971913173991426 | <a href="https://pbs.twimg.com/media/DF1e0mZXUAAALUcq.jpg">https://pbs.twimg.com/media/DF1e0mZXUAAALUcq.jpg</a>   |
| 2069 | 891087950875897856 | <a href="https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg">https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg</a>     |
| 2070 | 891327558926688256 | <a href="https://pbs.twimg.com/media/DF6hr6BUMAAZzGt.jpg">https://pbs.twimg.com/media/DF6hr6BUMAAZzGt.jpg</a>     |
| 2071 | 891689557279858688 | <a href="https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg">https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg</a>     |
| 2072 | 891815181378084864 | <a href="https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg">https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg</a>     |
| 2073 | 892177421306343426 | <a href="https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg">https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg</a>     |
| 2074 | 892420643555336193 | <a href="https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg">https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg</a>     |

|   | img_num |                        | p1       | p1_conf | p1_dog | \ |
|---|---------|------------------------|----------|---------|--------|---|
| 0 | 1       | Welsh_springer_spaniel | 0.465074 |         | True   |   |

|      |     |                             |          |       |
|------|-----|-----------------------------|----------|-------|
| 1    | 1   | redbone                     | 0.506826 | True  |
| 2    | 1   | German_shepherd             | 0.596461 | True  |
| 3    | 1   | Rhodesian_ridgeback         | 0.408143 | True  |
| 4    | 1   | miniature_pinscher          | 0.560311 | True  |
| 5    | 1   | Bernese_mountain_dog        | 0.651137 | True  |
| 6    | 1   | box_turtle                  | 0.933012 | False |
| 7    | 1   | chow                        | 0.692517 | True  |
| 8    | 1   | shopping_cart               | 0.962465 | False |
| 9    | 1   | miniature_poodle            | 0.201493 | True  |
| 10   | 1   | golden_retriever            | 0.775930 | True  |
| 11   | 1   | Gordon_setter               | 0.503672 | True  |
| 12   | 1   | Walker_hound                | 0.260857 | True  |
| 13   | 1   | pug                         | 0.489814 | True  |
| 14   | 1   | bloodhound                  | 0.195217 | True  |
| 15   | 1   | Lhasa                       | 0.582330 | True  |
| 16   | 1   | English_setter              | 0.298617 | True  |
| 17   | 1   | hen                         | 0.965932 | False |
| 18   | 1   | desktop_computer            | 0.086502 | False |
| 19   | 1   | Italian_greyhound           | 0.176053 | True  |
| 20   | 1   | Maltese_dog                 | 0.857531 | True  |
| 21   | 1   | three-toed_sloth            | 0.914671 | False |
| 22   | 1   | ox                          | 0.416669 | False |
| 23   | 1   | golden_retriever            | 0.858744 | True  |
| 24   | 1   | malamute                    | 0.336874 | True  |
| 25   | 1   | guinea_pig                  | 0.996496 | False |
| 26   | 1   | soft-coated_wheaten_terrier | 0.326467 | True  |
| 27   | 1   | Chihuahua                   | 0.978108 | True  |
| 28   | 1   | black-and-tan_coonhound     | 0.529139 | True  |
| 29   | 1   | coho                        | 0.404640 | False |
| ...  | ... | ...                         | ...      | ...   |
| 2045 | 1   | French_bulldog              | 0.999201 | True  |
| 2046 | 1   | convertible                 | 0.738995 | False |
| 2047 | 1   | kuvasz                      | 0.309706 | True  |
| 2048 | 2   | Chihuahua                   | 0.793469 | True  |
| 2049 | 1   | Samoyed                     | 0.733942 | True  |
| 2050 | 1   | Mexican_hairless            | 0.330741 | True  |
| 2051 | 2   | Pembroke                    | 0.809197 | True  |
| 2052 | 1   | limousine                   | 0.130432 | False |
| 2053 | 1   | basset                      | 0.821664 | True  |
| 2054 | 1   | French_bulldog              | 0.995026 | True  |
| 2055 | 2   | Pembroke                    | 0.809197 | True  |
| 2056 | 3   | Siberian_husky              | 0.700377 | True  |
| 2057 | 1   | golden_retriever            | 0.469760 | True  |
| 2058 | 1   | golden_retriever            | 0.714719 | True  |
| 2059 | 1   | whippet                     | 0.626152 | True  |
| 2060 | 1   | golden_retriever            | 0.953442 | True  |
| 2061 | 1   | French_bulldog              | 0.991650 | True  |
| 2062 | 1   | Pembroke                    | 0.966327 | True  |

|      |   |                          |          |       |
|------|---|--------------------------|----------|-------|
| 2063 | 1 | French_bulldog           | 0.377417 | True  |
| 2064 | 1 | Samoyed                  | 0.957979 | True  |
| 2065 | 1 | Pembroke                 | 0.511319 | True  |
| 2066 | 1 | Irish_terrier            | 0.487574 | True  |
| 2067 | 2 | Pomeranian               | 0.566142 | True  |
| 2068 | 1 | Appenzeller              | 0.341703 | True  |
| 2069 | 1 | Chesapeake_Bay_retriever | 0.425595 | True  |
| 2070 | 2 | basset                   | 0.555712 | True  |
| 2071 | 1 | paper_towel              | 0.170278 | False |
| 2072 | 1 | Chihuahua                | 0.716012 | True  |
| 2073 | 1 | Chihuahua                | 0.323581 | True  |
| 2074 | 1 | orange                   | 0.097049 | False |

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

|      |                     |          |       |                             |
|------|---------------------|----------|-------|-----------------------------|
| 2048 | toy_terrier         | 0.143528 | True  | can_opener                  |
| 2049 | Eskimo_dog          | 0.035029 | True  | Staffordshire_bullterrier   |
| 2050 | sea_lion            | 0.275645 | False | Weimaraner                  |
| 2051 | Rhodesian_ridgeback | 0.054950 | True  | beagle                      |
| 2052 | tow_truck           | 0.029175 | False | shopping_cart               |
| 2053 | redbone             | 0.087582 | True  | Weimaraner                  |
| 2054 | pug                 | 0.000932 | True  | bull_mastiff                |
| 2055 | Rhodesian_ridgeback | 0.054950 | True  | beagle                      |
| 2056 | Eskimo_dog          | 0.166511 | True  | malamute                    |
| 2057 | Labrador_retriever  | 0.184172 | True  | English_setter              |
| 2058 | Tibetan_mastiff     | 0.120184 | True  | Labrador_retriever          |
| 2059 | borzoi              | 0.194742 | True  | Saluki                      |
| 2060 | Labrador_retriever  | 0.013834 | True  | redbone                     |
| 2061 | boxer               | 0.002129 | True  | Staffordshire_bullterrier   |
| 2062 | Cardigan            | 0.027356 | True  | basenji                     |
| 2063 | Labrador_retriever  | 0.151317 | True  | muzzle                      |
| 2064 | Pomeranian          | 0.013884 | True  | chow                        |
| 2065 | Cardigan            | 0.451038 | True  | Chihuahua                   |
| 2066 | Irish_setter        | 0.193054 | True  | Chesapeake_Bay_retriever    |
| 2067 | Eskimo_dog          | 0.178406 | True  | Pembroke                    |
| 2068 | Border_collie       | 0.199287 | True  | ice_lolly                   |
| 2069 | Irish_terrier       | 0.116317 | True  | Indian_elephant             |
| 2070 | English_springer    | 0.225770 | True  | German_short-haired_pointer |
| 2071 | Labrador_retriever  | 0.168086 | True  | spatula                     |
| 2072 | malamute            | 0.078253 | True  | kelpie                      |
| 2073 | Pekinese            | 0.090647 | True  | papillon                    |
| 2074 | bagel               | 0.085851 | False | banana                      |

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



|      |          |       |
|------|----------|-------|
| 19   | 0.111152 | True  |
| 20   | 0.025581 | True  |
| 21   | 0.013207 | False |
| 22   | 0.102643 | True  |
| 23   | 0.014241 | True  |
| 24   | 0.093412 | True  |
| 25   | 0.000461 | False |
| 26   | 0.206803 | True  |
| 27   | 0.004577 | True  |
| 28   | 0.173810 | True  |
| 29   | 0.189945 | False |
| ...  | ...      | ...   |
| 2045 | 0.000076 | True  |
| 2046 | 0.044173 | False |
| 2047 | 0.086346 | True  |
| 2048 | 0.032253 | False |
| 2049 | 0.029705 | True  |
| 2050 | 0.134203 | True  |
| 2051 | 0.038915 | True  |
| 2052 | 0.026321 | False |
| 2053 | 0.026236 | True  |
| 2054 | 0.000903 | True  |
| 2055 | 0.038915 | True  |
| 2056 | 0.111411 | True  |
| 2057 | 0.073482 | True  |
| 2058 | 0.105506 | True  |
| 2059 | 0.027351 | True  |
| 2060 | 0.007958 | True  |
| 2061 | 0.001498 | True  |
| 2062 | 0.004633 | True  |
| 2063 | 0.082981 | False |
| 2064 | 0.008167 | True  |
| 2065 | 0.029248 | True  |
| 2066 | 0.118184 | True  |
| 2067 | 0.076507 | True  |
| 2068 | 0.193548 | False |
| 2069 | 0.076902 | False |
| 2070 | 0.175219 | True  |
| 2071 | 0.040836 | False |
| 2072 | 0.031379 | True  |
| 2073 | 0.068957 | True  |
| 2074 | 0.076110 | False |

[2075 rows x 12 columns]

imagePrediction table has 2075 rows and 12 columns. 1. In p1, p2, p3 columns (breeds of dogs)  
 2. Three confidence have 6 decimals. 3. 'Newfoundland','Lhasa' in p2 and p1 are not a breed of dog, but indicator = TRUE

```
In [10]: tweet_json
```

```
Out[10]:
```

|      | tweet_id            | retweet_count | favorite_count |
|------|---------------------|---------------|----------------|
| 0    | 1090301319825436672 | 2             | 305            |
| 1    | 1090300780270149632 | 11            | 1226           |
| 2    | 1090300624657252352 | 6989          | 79691          |
| 3    | 1090060665903927296 | 1             | 143            |
| 4    | 1090060504788197377 | 1             | 116            |
| 5    | 1090058668987404289 | 58            | 1282           |
| 6    | 1090055878013644800 | 3             | 1352           |
| 7    | 1090054426788691968 | 10442         | 120078         |
| 8    | 1090051186432958464 | 0             | 11             |
| 9    | 1090006105990750209 | 0             | 102            |
| 10   | 1089995552228487168 | 0             | 9              |
| 11   | 1089984884951592963 | 0             | 8              |
| 12   | 1089960608730992643 | 0             | 11             |
| 13   | 1089933184110743552 | 3391          | 0              |
| 14   | 1089608184073879552 | 169           | 6401           |
| 15   | 1089589687897939969 | 0             | 135            |
| 16   | 1089261994136416256 | 0             | 836            |
| 17   | 1089259355805896706 | 11893         | 122218         |
| 18   | 1089220307246665728 | 79            | 2550           |
| 19   | 1089217280049729536 | 4205          | 0              |
| 20   | 1089217166509916161 | 1             | 22             |
| 21   | 1088950936670633985 | 11            | 725            |
| 22   | 1088950484650479616 | 184           | 9256           |
| 23   | 1088883158982905856 | 165033        | 0              |
| 24   | 1088879155855294464 | 0             | 172            |
| 25   | 1088868189495189504 | 13            | 753            |
| 26   | 1088860190663790592 | 6569          | 50536          |
| 27   | 1088552419842084864 | 0             | 33             |
| 28   | 1088550291694223360 | 6265          | 0              |
| 29   | 1088513633804419072 | 17            | 3152           |
| ...  | ...                 | ...           | ...            |
| 2127 | 1028067313101156352 | 0             | 0              |
| 2128 | 1028067310592917504 | 0             | 0              |
| 2129 | 1028067009894871040 | 0             | 0              |
| 2130 | 1028066978173317120 | 0             | 0              |
| 2131 | 1028066930538700800 | 0             | 0              |
| 2132 | 1028066727190355968 | 0             | 0              |
| 2133 | 1028066622202818560 | 0             | 0              |
| 2134 | 1028066560189952000 | 0             | 0              |
| 2135 | 1028066488630968320 | 0             | 0              |
| 2136 | 1028066169834459137 | 0             | 0              |
| 2137 | 1028065866615676928 | 0             | 0              |
| 2138 | 1028065821619175425 | 0             | 0              |
| 2139 | 1028063180747165696 | 1             | 5              |
| 2140 | 1028051315493023746 | 0             | 0              |

|      |                     |       |        |
|------|---------------------|-------|--------|
| 2141 | 1028048323624591360 | 0     | 0      |
| 2142 | 1028040708266110976 | 0     | 3      |
| 2143 | 1028030559723765761 | 0     | 1      |
| 2144 | 1028029530957664268 | 0     | 5      |
| 2145 | 1028026872041492480 | 31134 | 154387 |
| 2146 | 1028024120414334976 | 0     | 1      |
| 2147 | 1028022323465646081 | 0     | 4      |
| 2148 | 1028020490080641029 | 0     | 0      |
| 2149 | 1027988812171898880 | 0     | 0      |
| 2150 | 1027987538554679296 | 0     | 0      |
| 2151 | 1027985302554595329 | 0     | 1      |
| 2152 | 1027980978910687233 | 0     | 0      |
| 2153 | 1027976875811819520 | 0     | 0      |
| 2154 | 1027970381766963202 | 0     | 2      |
| 2155 | 1027967697684815873 | 0     | 1      |
| 2156 | 1027963039029452800 | 0     | 0      |

[2157 rows x 3 columns]

tweet\_json table has 3225 rows and 5 columns.

## 2. Programatical assessment

### 2.1 archive table

In [11]: archive.info()

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

memory usage: 313.0+ KB

1. timestamp variable's type should be time format rather than string.
2. Expanded\_urls,in\_reply\_to\_status\_id,in\_reply\_to\_user\_id only has 78 non-null value.
3. retweeted\_status\_id,retweeted\_status\_user\_id and retweeted\_status\_timestamp have 2175 missing value
4. Tweet\_ID should be string rather than numeric value

In [12]: archive.tail()

```
Out[12]:
```

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

  

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

  

|      | source  | \ |
|------|---|---|
| 2351 | <a href="http://twitter.com/download/iphone" r... |   |
| 2352 | <a href="http://twitter.com/download/iphone" r... |   |
| 2353 | <a href="http://twitter.com/download/iphone" r... |   |
| 2354 | <a href="http://twitter.com/download/iphone" r... |   |
| 2355 | <a href="http://twitter.com/download/iphone" r... |   |

  

|      | text  | retweeted_status_id | \ |
|------|---|---------------------|---|
| 2351 | Here we have a 1949 1st generation vulpix. Enj... | NaN                 |   |
| 2352 | This is a purebred Piers Morgan. Loves to Netf... | NaN                 |   |
| 2353 | Here is a very happy pup. Big fan of well-main... | NaN                 |   |
| 2354 | This is a western brown Mitsubishi terrier. Up... | NaN                 |   |
| 2355 | Here we have a Japanese Irish Setter. Lost eye... | NaN                 |   |

  

|      | retweeted_status_user_id | retweeted_status_timestamp | \ |
|------|--------------------------|----------------------------|---|
| 2351 | NaN                      | NaN                        |   |
| 2352 | NaN                      | NaN                        |   |
| 2353 | NaN                      | NaN                        |   |
| 2354 | NaN                      | NaN                        |   |
| 2355 | NaN                      | NaN                        |   |

  

|      | expanded_urls                                     | rating_numerator | \ |
|------|---|------------------|---|
| 2351 | https://twitter.com/dog_rates/status/666049248... | 5                |   |

|      |   |   |
|------|---|---|
| 2352 | <a href="https://twitter.com/dog_rates/status/666044226...">https://twitter.com/dog_rates/status/666044226...</a> | 6 |
| 2353 | <a href="https://twitter.com/dog_rates/status/666033412...">https://twitter.com/dog_rates/status/666033412...</a> | 9 |
| 2354 | <a href="https://twitter.com/dog_rates/status/666029285...">https://twitter.com/dog_rates/status/666029285...</a> | 7 |
| 2355 | <a href="https://twitter.com/dog_rates/status/666020888...">https://twitter.com/dog_rates/status/666020888...</a> | 8 |

|      | rating_denominator | name | doggo | floofer | pupper | puppo |
|------|--------------------|------|-------|---------|--------|-------|
| 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  |

```
In [13]: pd.set_option('display.max_colwidth', -1)
```

```
In [14]: archive['name'].value_counts()
```

```
Out[14]: None      745
a          55
Charlie    12
Lucy       11
Oliver     11
Cooper     11
Tucker     10
Penny      10
Lola       10
Bo         9
Winston    9
the        8
Sadie      8
Toby       7
Bailey     7
an         7
Buddy      7
Daisy      7
Bella      6
Leo        6
Koda       6
Jax        6
Scout      6
Oscar      6
Jack       6
Milo       6
Rusty     6
Dave       6
Stanley    6
Bentley    5
..
Karl       1
```

|            |   |
|------------|---|
| Katie      | 1 |
| Jiminus    | 1 |
| Chesterson | 1 |
| Goose      | 1 |
| Jessifer   | 1 |
| Tobi       | 1 |
| Rueben     | 1 |
| Sojourner  | 1 |
| Karma      | 1 |
| Asher      | 1 |
| Rooney     | 1 |
| Shelby     | 1 |
| Wishes     | 1 |
| Rizzo      | 1 |
| all        | 1 |
| Spanky     | 1 |
| Kawhi      | 1 |
| Willie     | 1 |
| Pepper     | 1 |
| Brian      | 1 |
| Carter     | 1 |
| Sephie     | 1 |
| Chef       | 1 |
| Grizzwald  | 1 |
| Kobe       | 1 |
| Eriq       | 1 |
| Brandonald | 1 |
| Lizzie     | 1 |
| space      | 1 |

Name: name, Length: 957, dtype: int64

The name variable has 745 'None', 55 'a', 7 'an' name and 8 'the'.

In [15]: `archive.rating_denominator.value_counts()`

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

```

130    1
150    1
170    1
7      1
0      1
Name: rating_denominator, dtype: int64

```

There is 1 record with denominator = 0, which is invalid rating.

```
In [16]: archive[archive['rating_denominator'] ==0]
```

```

Out[16]:
          tweet_id  in_reply_to_status_id  in_reply_to_user_id \
313  835246439529840640  8.352460e+17  26259576.0

          timestamp \
313  2017-02-24 21:54:03 +0000

source
313  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

313  @jonnyusun @Lin_Manuel ok jomny I know you're excited but 960/00 isn't a valid rati

          retweeted_status_id  retweeted_status_user_id  retweeted_status_timestamp \
313  NaN  NaN  NaN

          expanded_urls  rating_numerator  rating_denominator  name doggo floofer \
313  NaN  960  0  None  None  None

          pupper puppo
313  None  None

```

```
In [17]: archive[archive['rating_denominator'] ==2]
```

```

Out[17]:
          tweet_id  in_reply_to_status_id  in_reply_to_user_id \
2335  666287406224695296  NaN  NaN

          timestamp \
2335  2015-11-16 16:11:11 +0000

source
2335  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

2335  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood floo

          retweeted_status_id  retweeted_status_user_id \
2335  NaN  NaN

```

```

retweeted_status_timestamp \
2335 NaN

expanded_urls \
2335 https://twitter.com/dog_rates/status/666287406224695296/photo/1

rating_numerator rating_denominator name doggo floofer pupper puppo
2335 1 2 an None None None None

In [18]: archive[archive['rating_denominator'] ==7]

Out[18]:
tweet_id in_reply_to_status_id in_reply_to_user_id \
516 810984652412424192 NaN NaN

timestamp \
516 2016-12-19 23:06:23 +0000

source
516 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

516 Meet Sam. She smiles 24/7 & secretly aspires to be a reindeer. \nKeep Sam smile

retweeted_status_id retweeted_status_user_id retweeted_status_timestamp \
516 NaN NaN NaN

516 https://www.gofundme.com/sams-smile,https://twitter.com/dog_rates/status/810984652

rating_numerator rating_denominator name doggo floofer pupper puppo
516 24 7 Sam None None None None

In [19]: archive[archive['rating_denominator'] ==11]

Out[19]:
tweet_id in_reply_to_status_id in_reply_to_user_id \
784 775096608509886464 NaN NaN
1068 740373189193256964 NaN NaN
1662 682962037429899265 NaN NaN

timestamp \
784 2016-09-11 22:20:06 +0000
1068 2016-06-08 02:41:38 +0000
1662 2016-01-01 16:30:13 +0000

source
784 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
1068 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
1662 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

```



```

784 RT @dog_rates: After so many requests, this is Bretagne. She was the last survivi
1068 After so many requests, this is Bretagne. She was the last surviving 9/11 search
1662 This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was ju

```

```

retweeted_status_id retweeted_status_user_id \
784 7.403732e+17 4.196984e+09
1068 NaN NaN
1662 NaN NaN

```

```

retweeted_status_timestamp \
784 2016-06-08 02:41:38 +0000
1068 NaN
1662 NaN

```

```

784 https://twitter.com/dog_rates/status/740373189193256964/photo/1,https://twitter.c
1068 https://twitter.com/dog_rates/status/740373189193256964/photo/1,https://twitter.c
1662 https://twitter.com/dog_rates/status/682962037429899265/photo/1

```

```

rating_numerator rating_denominator name doggo floofer pupper puppo
784 9 11 None None None None None
1068 9 11 None None None None None
1662 7 11 Darrel None None None None

```

```
In [20]: archive.rating_numerator.value_counts()
```

```

Out[20]: 12 558
11 464
10 461
13 351
9 158
8 102
7 55
14 54
5 37
6 32
3 19
4 17
1 9
2 9
420 2
0 2
15 2
75 2
80 1
20 1
24 1

```

```

26      1
44      1
50      1
60      1
165     1
84      1
88      1
144     1
182     1
143     1
666     1
960     1
1776    1
17      1
27      1
45      1
99      1
121     1
204     1
Name: rating_numerator, dtype: int64

```

```
In [21]: archive[archive['rating_numerator'] == 1776]
```

```

Out[21]:
          tweet_id  in_reply_to_status_id  in_reply_to_user_id \
979  749981277374128128 NaN
          timestamp \
979  2016-07-04 15:00:45 +0000
          source
979  <a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
          text \
979  This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkh
          retweeted_status_id  retweeted_status_user_id  retweeted_status_timestamp \
979 NaN NaN NaN
          expanded_urls \
979  https://twitter.com/dog_rates/status/749981277374128128/photo/1
          rating_numerator  rating_denominator  name  doggo  floofer  pupper  puppo
979  1776 10 Atticus None None None None

```

```
In [22]: archive[archive['rating_numerator'] ==0]
```

```

Out[22]:
          tweet_id  in_reply_to_status_id  in_reply_to_user_id \
315  835152434251116546 NaN
1016  746906459439529985  7.468859e+17  4.196984e+09

```

```

                                timestamp \
315  2017-02-24 15:40:31 +0000
1016 2016-06-26 03:22:31 +0000

                                source
315  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
1016 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

315  When you're so blinded by your systematic plagiarism that you forget what day it
1016 PUPDATE: can't see any. Even if I could, I couldn't reach them to pet. 0/10 much

    retweeted_status_id retweeted_status_user_id \
315  NaN                NaN
1016 NaN                NaN

    retweeted_status_timestamp \
315  NaN
1016 NaN

315  https://twitter.com/dog_rates/status/835152434251116546/photo/1,https://twitter.c
1016 https://twitter.com/dog_rates/status/746906459439529985/photo/1

    rating_numerator rating_denominator name doggo floofer pupper puppo
315  0                10                None None None      None      None
1016 0                10                None None None      None      None

```

```
In [23]: archive[archive['rating_numerator'] == 5]
```

```

Out[23]:
    tweet_id in_reply_to_status_id in_reply_to_user_id \
45  883482846933004288 NaN                NaN
730 781661882474196992 NaN                NaN
956 751583847268179968 NaN                NaN
1399 699691744225525762 NaN               NaN
1461 694925794720792577 NaN               NaN
1508 691483041324204033 NaN               NaN
1583 687102708889812993 NaN               NaN
1618 684969860808454144 6.849598e+17      4.196984e+09
1619 684959798585110529 NaN               NaN
1624 684880619965411328 NaN               NaN
1645 683849932751646720 NaN               NaN
1680 682003177596559360 NaN               NaN
1689 681340665377193984 6.813394e+17      4.196984e+09
1727 679877062409191424 NaN               NaN
1796 677301033169788928 NaN               NaN
1808 676897532954456065 NaN               NaN

```

|      |                    |     |     |
|------|--------------------|-----|-----|
| 1820 | 676588346097852417 | NaN | NaN |
| 1861 | 675483430902214656 | NaN | NaN |
| 1874 | 675135153782571009 | NaN | NaN |
| 1901 | 674646392044941312 | NaN | NaN |
| 1904 | 674632714662858753 | NaN | NaN |
| 1925 | 674063288070742018 | NaN | NaN |
| 1979 | 672980819271634944 | NaN | NaN |
| 2013 | 672231046314901505 | NaN | NaN |
| 2026 | 671879137494245376 | NaN | NaN |
| 2063 | 671159727754231808 | NaN | NaN |
| 2092 | 670782429121134593 | NaN | NaN |
| 2109 | 670449342516494336 | NaN | NaN |
| 2134 | 670069087419133954 | NaN | NaN |
| 2139 | 670037189829525505 | NaN | NaN |
| 2153 | 669661792646373376 | NaN | NaN |
| 2181 | 668994913074286592 | NaN | NaN |
| 2206 | 668631377374486528 | NaN | NaN |
| 2242 | 667911425562669056 | NaN | NaN |
| 2260 | 667550882905632768 | NaN | NaN |
| 2312 | 666776908487630848 | NaN | NaN |
| 2351 | 666049248165822465 | NaN | NaN |

|      | timestamp \               |
|------|---------------------------|
| 45   | 2017-07-08 00:28:19 +0000 |
| 730  | 2016-09-30 01:08:10 +0000 |
| 956  | 2016-07-09 01:08:47 +0000 |
| 1399 | 2016-02-16 20:28:06 +0000 |
| 1461 | 2016-02-03 16:49:55 +0000 |
| 1508 | 2016-01-25 04:49:38 +0000 |
| 1583 | 2016-01-13 02:43:46 +0000 |
| 1618 | 2016-01-07 05:28:35 +0000 |
| 1619 | 2016-01-07 04:48:36 +0000 |
| 1624 | 2016-01-06 23:33:58 +0000 |
| 1645 | 2016-01-04 03:18:23 +0000 |
| 1680 | 2015-12-30 01:00:03 +0000 |
| 1689 | 2015-12-28 05:07:27 +0000 |
| 1727 | 2015-12-24 04:11:37 +0000 |
| 1796 | 2015-12-17 01:35:24 +0000 |
| 1808 | 2015-12-15 22:52:02 +0000 |
| 1820 | 2015-12-15 02:23:26 +0000 |
| 1861 | 2015-12-12 01:12:54 +0000 |
| 1874 | 2015-12-11 02:08:58 +0000 |
| 1901 | 2015-12-09 17:46:48 +0000 |
| 1904 | 2015-12-09 16:52:27 +0000 |
| 1925 | 2015-12-08 03:09:46 +0000 |
| 1979 | 2015-12-05 03:28:25 +0000 |
| 2013 | 2015-12-03 01:49:05 +0000 |
| 2026 | 2015-12-02 02:30:43 +0000 |

2063 2015-11-30 02:52:03 +0000  
 2092 2015-11-29 01:52:48 +0000  
 2109 2015-11-28 03:49:14 +0000  
 2134 2015-11-27 02:38:14 +0000  
 2139 2015-11-27 00:31:29 +0000  
 2153 2015-11-25 23:39:47 +0000  
 2181 2015-11-24 03:29:51 +0000  
 2206 2015-11-23 03:25:17 +0000  
 2242 2015-11-21 03:44:27 +0000  
 2260 2015-11-20 03:51:47 +0000  
 2312 2015-11-18 00:36:17 +0000  
 2351 2015-11-16 00:24:50 +0000

source

45 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 730 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 956 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1399 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1461 <a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>  
 1508 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1583 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1618 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1619 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1624 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1645 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1680 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1689 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1727 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1796 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1808 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1820 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1861 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1874 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1901 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1904 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1925 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 1979 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2013 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2026 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2063 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2092 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2109 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2134 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2139 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2153 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2181 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2206 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 2242 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

2260 <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>  
2312 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
2351 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

45 This is Bella. She hopes her smile made you smile. If not, she is also offering y  
730 Who keeps sending in pictures without dogs in them? This needs to stop. 5/10 for  
956 Please stop sending it pictures that don't even have a doggo or pupper in them. C  
1399 This is Dave. He's a tropical pup. Short lil legs (dachshund mix?) Excels underwa  
1461 Please only send in dogs. This t-rex is very scary. 5/10 ...might still pet (vid  
1508 When bae says they can't go out but you see them with someone else that same nigh  
1583 Army of water dogs here. None of them know where they're going. Have no real purp  
1618 For those who claim this is a goat, u are wrong. It is not the Greatest Of All Ti  
1619 This is Jerry. He's a neat dog. No legs (tragic). Has more horns than a dog usual  
1624 Here we have a basking dino pupper. Looks powerful. Occasionally shits eggs. Does  
1645 This is Jiminy. He's not the brightest dog. Needs to lay off the kibble. 5/10 sti  
1680 Unique dog here. Wrinkly as hell. Weird segmented neck. Finger on fire. Doesn't s  
1689 I've been told there's a slight possibility he's checking his mirror. We'll bump  
1727 Meet Penelope. She's a bacon frise. Total babe (lol get it like the movie). Doesn  
1796 This is Juckson. He's totally on his way to a nascar race. 5/10 for Juckson https  
1808 Exotic handheld dog here. Appears unathletic. Feet look deadly. Can be thrown a g  
1820 This is Bubbles. He kinda resembles a fish. Always makes eye contact with u no ma  
1861 Rare shielded battle dog here. Very happy about abundance of lettuce. Painfully s  
1874 This is Steven. He got locked outside. Damn it Steven. 5/10 nice grill tho https:  
1901 Two gorgeous dogs here. Little waddling dog is a rebel. Refuses to look at camera  
1904 Rare submerged pup here. Holds breath for a long time. Frowning because that spoo  
1925 This is Earl. Earl is lost. Someone help Earl. He has no tags. Just trying to get  
1979 Extraordinary dog here. Looks large. Just a head. No body. Rather intrusive. 5/10  
2013 Exotic underwater dog here. Very shy. Wont return tennis balls I toss him. Never  
2026 This is Brad. He's a chubby lil pup. Doesn't really need the food he's trying to  
2063 This is Anthony. He just finished up his masters at Harvard. Unprofessional tatto  
2092 This dude slaps your girl's ass what do you do?\n5/10 https://t.co/6dioUL6gcP  
2109 Vibrant dog here. Fabulous tail. Only 2 legs tho. Has wings but can barely fly (l  
2134 This is Randall. He's from Chernobyl. Built playground himself. Has been stuck up  
2139 Awesome dog here. Not sure where it is tho. Spectacular camouflage. Enjoys leaves  
2153 This is a brave dog. Excellent free climber. Trying to get closer to God. Not ver  
2181 Two gorgeous pups here. Both have cute fake horns(adorable). Barn in the back loc  
2206 Meet Zeek. He is a grey Cumulonimbus. Zeek is hungry. Someone should feed Zeek as  
2242 Wow. Armored dog here. Ready for battle. Face looks dangerous. Not very loyal. Li  
2260 RT @dogratingrating: Unoriginal idea. Blatant plagiarism. Curious grammar. -5/10  
2312 This is Josep. He is a Rye Manganese mix. Can drive w eyes closed. Very irrespons  
2351 Here we have a 1949 1st generation vulpix. Enjoys sweat tea and Fox News. Cannot

|      | retweeted_status_id | retweeted_status_user_id \ |
|------|---------------------|----------------------------|
| 45   | NaN                 | NaN                        |
| 730  | NaN                 | NaN                        |
| 956  | NaN                 | NaN                        |
| 1399 | NaN                 | NaN                        |

|      |              |              |
|------|--------------|--------------|
| 1461 | NaN          | NaN          |
| 1508 | NaN          | NaN          |
| 1583 | NaN          | NaN          |
| 1618 | NaN          | NaN          |
| 1619 | NaN          | NaN          |
| 1624 | NaN          | NaN          |
| 1645 | NaN          | NaN          |
| 1680 | NaN          | NaN          |
| 1689 | NaN          | NaN          |
| 1727 | NaN          | NaN          |
| 1796 | NaN          | NaN          |
| 1808 | NaN          | NaN          |
| 1820 | NaN          | NaN          |
| 1861 | NaN          | NaN          |
| 1874 | NaN          | NaN          |
| 1901 | NaN          | NaN          |
| 1904 | NaN          | NaN          |
| 1925 | NaN          | NaN          |
| 1979 | NaN          | NaN          |
| 2013 | NaN          | NaN          |
| 2026 | NaN          | NaN          |
| 2063 | NaN          | NaN          |
| 2092 | NaN          | NaN          |
| 2109 | NaN          | NaN          |
| 2134 | NaN          | NaN          |
| 2139 | NaN          | NaN          |
| 2153 | NaN          | NaN          |
| 2181 | NaN          | NaN          |
| 2206 | NaN          | NaN          |
| 2242 | NaN          | NaN          |
| 2260 | 6.675484e+17 | 4.296832e+09 |
| 2312 | NaN          | NaN          |
| 2351 | NaN          | NaN          |

|      | retweeted_status_timestamp \ |
|------|------------------------------|
| 45   | NaN                          |
| 730  | NaN                          |
| 956  | NaN                          |
| 1399 | NaN                          |
| 1461 | NaN                          |
| 1508 | NaN                          |
| 1583 | NaN                          |
| 1618 | NaN                          |
| 1619 | NaN                          |
| 1624 | NaN                          |
| 1645 | NaN                          |
| 1680 | NaN                          |
| 1689 | NaN                          |

|      |                           |
|------|---------------------------|
| 1727 | NaN                       |
| 1796 | NaN                       |
| 1808 | NaN                       |
| 1820 | NaN                       |
| 1861 | NaN                       |
| 1874 | NaN                       |
| 1901 | NaN                       |
| 1904 | NaN                       |
| 1925 | NaN                       |
| 1979 | NaN                       |
| 2013 | NaN                       |
| 2026 | NaN                       |
| 2063 | NaN                       |
| 2092 | NaN                       |
| 2109 | NaN                       |
| 2134 | NaN                       |
| 2139 | NaN                       |
| 2153 | NaN                       |
| 2181 | NaN                       |
| 2206 | NaN                       |
| 2242 | NaN                       |
| 2260 | 2015-11-20 03:41:59 +0000 |
| 2312 | NaN                       |
| 2351 | NaN                       |

|      |   |
|------|---|
| 45   | <a href="https://twitter.com/dog_rates/status/883482846933004288/photo/1">https://twitter.com/dog_rates/status/883482846933004288/photo/1</a> , <a href="https://twitter.com/dog_rates/status/883482846933004288/photo/1">https://twitter.com/dog_rates/status/883482846933004288/photo/1</a> |
| 730  | <a href="https://twitter.com/dog_rates/status/781661882474196992/photo/1">https://twitter.com/dog_rates/status/781661882474196992/photo/1</a>   |
| 956  | <a href="https://twitter.com/dog_rates/status/751583847268179968/photo/1">https://twitter.com/dog_rates/status/751583847268179968/photo/1</a>   |
| 1399 | <a href="https://twitter.com/dog_rates/status/699691744225525762/photo/1">https://twitter.com/dog_rates/status/699691744225525762/photo/1</a>   |
| 1461 | <a href="https://vine.co/v/iJvUqWQ166L">https://vine.co/v/iJvUqWQ166L</a>   |
| 1508 | <a href="https://twitter.com/dog_rates/status/691483041324204033/photo/1">https://twitter.com/dog_rates/status/691483041324204033/photo/1</a> , <a href="https://twitter.com/dog_rates/status/691483041324204033/photo/1">https://twitter.com/dog_rates/status/691483041324204033/photo/1</a> |
| 1583 | <a href="https://twitter.com/dog_rates/status/687102708889812993/photo/1">https://twitter.com/dog_rates/status/687102708889812993/photo/1</a>   |
| 1618 | NaN   |
| 1619 | <a href="https://twitter.com/dog_rates/status/684959798585110529/photo/1">https://twitter.com/dog_rates/status/684959798585110529/photo/1</a>   |
| 1624 | <a href="https://twitter.com/dog_rates/status/684880619965411328/photo/1">https://twitter.com/dog_rates/status/684880619965411328/photo/1</a>   |
| 1645 | <a href="https://twitter.com/dog_rates/status/683849932751646720/photo/1">https://twitter.com/dog_rates/status/683849932751646720/photo/1</a>   |
| 1680 | <a href="https://twitter.com/dog_rates/status/682003177596559360/photo/1">https://twitter.com/dog_rates/status/682003177596559360/photo/1</a>   |
| 1689 | NaN   |
| 1727 | <a href="https://twitter.com/dog_rates/status/679877062409191424/photo/1">https://twitter.com/dog_rates/status/679877062409191424/photo/1</a>   |
| 1796 | <a href="https://twitter.com/dog_rates/status/677301033169788928/photo/1">https://twitter.com/dog_rates/status/677301033169788928/photo/1</a>   |
| 1808 | <a href="https://twitter.com/dog_rates/status/676897532954456065/photo/1">https://twitter.com/dog_rates/status/676897532954456065/photo/1</a>   |
| 1820 | <a href="https://twitter.com/dog_rates/status/676588346097852417/photo/1">https://twitter.com/dog_rates/status/676588346097852417/photo/1</a>   |
| 1861 | <a href="https://twitter.com/dog_rates/status/675483430902214656/photo/1">https://twitter.com/dog_rates/status/675483430902214656/photo/1</a>   |
| 1874 | <a href="https://twitter.com/dog_rates/status/675135153782571009/photo/1">https://twitter.com/dog_rates/status/675135153782571009/photo/1</a>   |
| 1901 | <a href="https://twitter.com/dog_rates/status/674646392044941312/photo/1">https://twitter.com/dog_rates/status/674646392044941312/photo/1</a>   |
| 1904 | <a href="https://twitter.com/dog_rates/status/674632714662858753/photo/1">https://twitter.com/dog_rates/status/674632714662858753/photo/1</a>   |
| 1925 | <a href="https://twitter.com/dog_rates/status/674063288070742018/photo/1">https://twitter.com/dog_rates/status/674063288070742018/photo/1</a>   |



1979 [https://twitter.com/dog\\_rates/status/672980819271634944/photo/1](https://twitter.com/dog_rates/status/672980819271634944/photo/1)  
 2013 [https://twitter.com/dog\\_rates/status/672231046314901505/photo/1](https://twitter.com/dog_rates/status/672231046314901505/photo/1)  
 2026 [https://twitter.com/dog\\_rates/status/671879137494245376/photo/1](https://twitter.com/dog_rates/status/671879137494245376/photo/1)  
 2063 [https://twitter.com/dog\\_rates/status/671159727754231808/photo/1](https://twitter.com/dog_rates/status/671159727754231808/photo/1)  
 2092 [https://twitter.com/dog\\_rates/status/670782429121134593/photo/1](https://twitter.com/dog_rates/status/670782429121134593/photo/1)  
 2109 [https://twitter.com/dog\\_rates/status/670449342516494336/photo/1](https://twitter.com/dog_rates/status/670449342516494336/photo/1)  
 2134 [https://twitter.com/dog\\_rates/status/670069087419133954/photo/1](https://twitter.com/dog_rates/status/670069087419133954/photo/1)  
 2139 [https://twitter.com/dog\\_rates/status/670037189829525505/photo/1](https://twitter.com/dog_rates/status/670037189829525505/photo/1)  
 2153 [https://twitter.com/dog\\_rates/status/669661792646373376/photo/1](https://twitter.com/dog_rates/status/669661792646373376/photo/1)  
 2181 [https://twitter.com/dog\\_rates/status/668994913074286592/photo/1](https://twitter.com/dog_rates/status/668994913074286592/photo/1)  
 2206 [https://twitter.com/dog\\_rates/status/668631377374486528/photo/1](https://twitter.com/dog_rates/status/668631377374486528/photo/1)  
 2242 [https://twitter.com/dog\\_rates/status/667911425562669056/photo/1](https://twitter.com/dog_rates/status/667911425562669056/photo/1)  
 2260 <https://twitter.com/dogratingrating/status/667548415174144001/photo/1>,<https://twitter.com/dogratingrating/status/667548415174144001/photo/1>  
 2312 [https://twitter.com/dog\\_rates/status/666776908487630848/photo/1](https://twitter.com/dog_rates/status/666776908487630848/photo/1)  
 2351 [https://twitter.com/dog\\_rates/status/666049248165822465/photo/1](https://twitter.com/dog_rates/status/666049248165822465/photo/1)

|      | rating_numerator | rating_denominator | name     | doggo | floofer | pupper | \ |
|------|------------------|--------------------|----------|-------|---------|--------|---|
| 45   | 5                | 10                 | Bella    | None  | None    | None   |   |
| 730  | 5                | 10                 | None     | None  | None    | None   |   |
| 956  | 5                | 10                 | None     | doggo | None    | pupper |   |
| 1399 | 5                | 10                 | Dave     | None  | None    | None   |   |
| 1461 | 5                | 10                 | None     | None  | None    | None   |   |
| 1508 | 5                | 10                 | None     | None  | None    | None   |   |
| 1583 | 5                | 10                 | None     | None  | None    | None   |   |
| 1618 | 5                | 10                 | None     | None  | None    | None   |   |
| 1619 | 5                | 10                 | Jerry    | None  | None    | None   |   |
| 1624 | 5                | 10                 | None     | None  | None    | pupper |   |
| 1645 | 5                | 10                 | Jiminy   | None  | None    | None   |   |
| 1680 | 5                | 10                 | None     | None  | None    | None   |   |
| 1689 | 5                | 10                 | None     | None  | None    | None   |   |
| 1727 | 5                | 10                 | Penelope | None  | None    | None   |   |
| 1796 | 5                | 10                 | Jackson  | None  | None    | None   |   |
| 1808 | 5                | 10                 | None     | None  | None    | None   |   |
| 1820 | 5                | 10                 | Bubbles  | None  | None    | None   |   |
| 1861 | 5                | 10                 | None     | None  | None    | None   |   |
| 1874 | 5                | 10                 | Steven   | None  | None    | None   |   |
| 1901 | 5                | 10                 | None     | None  | None    | None   |   |
| 1904 | 5                | 10                 | None     | None  | None    | None   |   |
| 1925 | 5                | 10                 | Earl     | None  | None    | None   |   |
| 1979 | 5                | 10                 | None     | None  | None    | None   |   |
| 2013 | 5                | 10                 | None     | None  | None    | None   |   |
| 2026 | 5                | 10                 | Brad     | None  | None    | None   |   |
| 2063 | 5                | 10                 | Anthony  | None  | None    | None   |   |
| 2092 | 5                | 10                 | None     | None  | None    | None   |   |
| 2109 | 5                | 10                 | None     | None  | None    | None   |   |
| 2134 | 5                | 10                 | Randall  | None  | None    | None   |   |
| 2139 | 5                | 10                 | None     | None  | None    | None   |   |
| 2153 | 5                | 10                 | a        | None  | None    | None   |   |

|      |   |    |       |      |      |      |
|------|---|----|-------|------|------|------|
| 2181 | 5 | 10 | None  | None | None | None |
| 2206 | 5 | 10 | Zeek  | None | None | None |
| 2242 | 5 | 10 | None  | None | None | None |
| 2260 | 5 | 10 | None  | None | None | None |
| 2312 | 5 | 10 | Josep | None | None | None |
| 2351 | 5 | 10 | None  | None | None | None |

puppo

|      |      |
|------|------|
| 45   | None |
| 730  | None |
| 956  | None |
| 1399 | None |
| 1461 | None |
| 1508 | None |
| 1583 | None |
| 1618 | None |
| 1619 | None |
| 1624 | None |
| 1645 | None |
| 1680 | None |
| 1689 | None |
| 1727 | None |
| 1796 | None |
| 1808 | None |
| 1820 | None |
| 1861 | None |
| 1874 | None |
| 1901 | None |
| 1904 | None |
| 1925 | None |
| 1979 | None |
| 2013 | None |
| 2026 | None |
| 2063 | None |
| 2092 | None |
| 2109 | None |
| 2134 | None |
| 2139 | None |
| 2153 | None |
| 2181 | None |
| 2206 | None |
| 2242 | None |
| 2260 | None |
| 2312 | None |
| 2351 | None |

In [24]: archive.timestamp.sort\_values()

Out[24]: 2355      2015-11-15 22:32:08 +0000

|      |            |          |       |
|------|------------|----------|-------|
| 2354 | 2015-11-15 | 23:05:30 | +0000 |
| 2353 | 2015-11-15 | 23:21:54 | +0000 |
| 2352 | 2015-11-16 | 00:04:52 | +0000 |
| 2351 | 2015-11-16 | 00:24:50 | +0000 |
| 2350 | 2015-11-16 | 00:30:50 | +0000 |
| 2349 | 2015-11-16 | 00:35:11 | +0000 |
| 2348 | 2015-11-16 | 00:49:46 | +0000 |
| 2347 | 2015-11-16 | 00:55:59 | +0000 |
| 2346 | 2015-11-16 | 01:01:59 | +0000 |
| 2345 | 2015-11-16 | 01:22:45 | +0000 |
| 2344 | 2015-11-16 | 01:52:02 | +0000 |
| 2343 | 2015-11-16 | 01:59:36 | +0000 |
| 2342 | 2015-11-16 | 02:38:37 | +0000 |
| 2341 | 2015-11-16 | 03:22:39 | +0000 |
| 2340 | 2015-11-16 | 03:44:34 | +0000 |
| 2339 | 2015-11-16 | 03:55:04 | +0000 |
| 2338 | 2015-11-16 | 04:02:55 | +0000 |
| 2337 | 2015-11-16 | 14:57:41 | +0000 |
| 2336 | 2015-11-16 | 15:14:19 | +0000 |
| 2335 | 2015-11-16 | 16:11:11 | +0000 |
| 2334 | 2015-11-16 | 16:37:02 | +0000 |
| 2333 | 2015-11-16 | 19:31:45 | +0000 |
| 2332 | 2015-11-16 | 20:01:42 | +0000 |
| 2331 | 2015-11-16 | 20:32:58 | +0000 |
| 2330 | 2015-11-16 | 21:10:36 | +0000 |
| 2329 | 2015-11-16 | 21:54:18 | +0000 |
| 2328 | 2015-11-16 | 23:23:41 | +0000 |
| 2327 | 2015-11-17 | 00:06:54 | +0000 |
| 2326 | 2015-11-17 | 00:24:19 | +0000 |
|      |            | ...      |       |
| 29   | 2017-07-15 | 23:25:31 | +0000 |
| 28   | 2017-07-16 | 20:14:00 | +0000 |
| 27   | 2017-07-16 | 23:58:41 | +0000 |
| 26   | 2017-07-17 | 16:17:36 | +0000 |
| 25   | 2017-07-18 | 00:07:08 | +0000 |
| 24   | 2017-07-18 | 16:08:03 | +0000 |
| 23   | 2017-07-19 | 00:47:34 | +0000 |
| 22   | 2017-07-19 | 03:39:09 | +0000 |
| 21   | 2017-07-19 | 16:06:48 | +0000 |
| 20   | 2017-07-20 | 16:49:33 | +0000 |
| 19   | 2017-07-21 | 01:02:36 | +0000 |
| 18   | 2017-07-22 | 00:23:06 | +0000 |
| 17   | 2017-07-22 | 16:56:37 | +0000 |
| 16   | 2017-07-23 | 00:22:39 | +0000 |
| 15   | 2017-07-24 | 00:19:32 | +0000 |
| 14   | 2017-07-24 | 17:02:04 | +0000 |
| 13   | 2017-07-25 | 00:10:02 | +0000 |
| 12   | 2017-07-25 | 01:55:32 | +0000 |

```

11      2017-07-25 16:11:53 +0000
10      2017-07-26 00:31:25 +0000
9       2017-07-26 15:59:51 +0000
8       2017-07-27 16:25:51 +0000
7       2017-07-28 00:22:40 +0000
6       2017-07-28 16:27:12 +0000
5       2017-07-29 00:08:17 +0000
4       2017-07-29 16:00:24 +0000
3       2017-07-30 15:58:51 +0000
2       2017-07-31 00:18:03 +0000
1       2017-08-01 00:17:27 +0000
0       2017-08-01 16:23:56 +0000
Name: timestamp, Length: 2356, dtype: object

```

```

In [25]: archive[archive.duplicated() == True]
         archive[archive.tweet_id.duplicated() == True]

```

```

Out[25]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text]
         Index: []

```

```

In [26]: archive.source.value_counts()

```

```

Out[26]: <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         <a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>
         <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
         <a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
         Name: source, dtype: int64

```

```

In [27]: sum(archive.expanded_urls.isnull())

```

```

Out[27]: 59

```

```

In [28]: sum(archive.in_reply_to_user_id.isnull())

```

```

Out[28]: 2278

```

```

In [29]: sum(archive.retweeted_status_user_id.isnull())

```

```

Out[29]: 2175

```

## 2.2 imagePrediction table

```

In [30]: imagePrediction.head()

```

```

Out[30]:
   tweet_id                                     jpg_url \
0  666020888022790149  https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg
1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2  666033412701032449  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3  666044226329800704  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg

```

```
4 666049248165822465 https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
```

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

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

```
In [31]: imagePrediction.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 [32]: imagePrediction.describe()
```

```
Out[32]:
```

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

```
In [33]: imagePrediction['p1_conf'].describe().apply(lambda x: format(x, 'f'))
```

```
Out[33]: count      2075.000000
         mean       0.594548
         std        0.271174
         min        0.044333
         25%        0.364412
         50%        0.588230
         75%        0.843855
         max        1.000000
         Name: p1_conf, dtype: object
```

```
In [34]: imagePrediction['p2_conf'].describe().apply(lambda x: format(x, 'f'))
```

```
Out[34]: count      2075.000000
         mean       0.134589
         std        0.100666
         min        0.000000
         25%        0.053886
         50%        0.118181
         75%        0.195566
         max        0.488014
         Name: p2_conf, dtype: object
```

```
In [35]: imagePrediction['p3_conf'].describe().apply(lambda x: format(x, 'f'))
```

```
Out[35]: count      2075.000000
         mean       0.060324
         std        0.050906
         min        0.000000
         25%        0.016222
         50%        0.049444
         75%        0.091808
         max        0.273419
         Name: p3_conf, dtype: object
```

```
In [36]: imagePrediction.p1.value_counts()
```

```
Out[36]: golden_retriever      150
         Labrador_retriever    100
         Pembroke              89
         Chihuahua             83
         pug                   57
         chow                  44
         Samoyed               43
         toy_poodle            39
         Pomeranian            38
         malamute              30
         cocker_spaniel        30
```

|                           |    |
|---------------------------|----|
| French_bulldog            | 26 |
| miniature_pinscher        | 23 |
| Chesapeake_Bay_retriever  | 23 |
| seat_belt                 | 22 |
| Siberian_husky            | 20 |
| German_shepherd           | 20 |
| Staffordshire_bullterrier | 20 |
| Cardigan                  | 19 |
| web_site                  | 19 |
| beagle                    | 18 |
| Eskimo_dog                | 18 |
| Shetland_sheepdog         | 18 |
| Maltese_dog               | 18 |
| teddy                     | 18 |
| Lakeland_terrier          | 17 |
| Rottweiler                | 17 |
| Shih-Tzu                  | 17 |
| kuvasz                    | 16 |
| Italian_greyhound         | 16 |
| ..                        |    |
| platypus                  | 1  |
| four-poster               | 1  |
| orange                    | 1  |
| grey_fox                  | 1  |
| tricycle                  | 1  |
| leaf_beetle               | 1  |
| polecat                   | 1  |
| conch                     | 1  |
| stove                     | 1  |
| Japanese_spaniel          | 1  |
| bison                     | 1  |
| water_bottle              | 1  |
| leopard                   | 1  |
| mud_turtle                | 1  |
| maze                      | 1  |
| mortarboard               | 1  |
| pencil_box                | 1  |
| pillow                    | 1  |
| nail                      | 1  |
| Madagascar_cat            | 1  |
| ibex                      | 1  |
| lawn_mower                | 1  |
| coil                      | 1  |
| swab                      | 1  |
| wild_boar                 | 1  |
| hummingbird               | 1  |
| studio_couch              | 1  |
| giant_panda               | 1  |

```

terrapiin          1
envelope           1
Name: p1, Length: 378, dtype: int64

```

```
In [37]: imagePrediction.p2.value_counts()
```

```

Out[37]: Labrador_retriever          104
golden_retriever                    92
Cardigan                           73
Chihuahua                          44
Pomeranian                         42
Chesapeake_Bay_retriever           41
French_bulldog                     41
toy_poodle                         37
cocker_spaniel                     34
Siberian_husky                     33
miniature_poodle                   33
beagle                             28
Pembroke                           27
Eskimo_dog                         27
collie                             27
kuvasz                             26
Italian_greyhound                  22
Pekinese                           21
American_Staffordshire_terrier     21
chow                               20
miniature_pinscher                 20
toy_terrier                        20
malinois                           20
Samoyed                            20
Norwegian_elkhound                 19
Boston_bull                        19
Staffordshire_bullterrier           18
pug                                17
Irish_terrier                       17
kelpie                             16
..
bearskin                           1
spotlight                         1
otter                             1
soccer_ball                        1
knee_pad                           1
snail                              1
giant_panda                        1
dumbbell                           1
waffle_iron                        1
tray                               1
can_opener                         1

```



|                                     |   |
|-------------------------------------|---|
| cowboy_hat                          | 1 |
| Madagascar_cat                      | 1 |
| boathouse                           | 1 |
| television                          | 1 |
| computer_keyboard                   | 1 |
| china_cabinet                       | 1 |
| patio                               | 1 |
| timber_wolf                         | 1 |
| trench_coat                         | 1 |
| coral_reef                          | 1 |
| shower_cap                          | 1 |
| lesser_panda                        | 1 |
| white_wolf                          | 1 |
| rule                                | 1 |
| lighter                             | 1 |
| lawn_mower                          | 1 |
| pelican                             | 1 |
| handkerchief                        | 1 |
| mosquito_net                        | 1 |
| Name: p2, Length: 405, dtype: int64 |   |

In [38]: imagePrediction.p3.value\_counts()

|                                |    |
|--------------------------------|----|
| Labrador_retriever             | 79 |
| Chihuahua                      | 58 |
| golden_retriever               | 48 |
| Eskimo_dog                     | 38 |
| kelpie                         | 35 |
| kuvasz                         | 34 |
| Staffordshire_bullterrier      | 32 |
| chow                           | 32 |
| cocker_spaniel                 | 31 |
| beagle                         | 31 |
| toy_poodle                     | 29 |
| Pekinese                       | 29 |
| Pomeranian                     | 29 |
| Pembroke                       | 27 |
| Chesapeake_Bay_retriever       | 27 |
| Great_Pyrenees                 | 27 |
| malamute                       | 26 |
| French_bulldog                 | 26 |
| American_Staffordshire_terrier | 24 |
| Cardigan                       | 23 |
| pug                            | 23 |
| basenji                        | 21 |
| bull_mastiff                   | 20 |
| toy_terrier                    | 20 |
| Siberian_husky                 | 19 |

|                   |    |
|-------------------|----|
| Boston_bull       | 17 |
| Shetland_sheepdog | 17 |
| Lakeland_terrier  | 16 |
| boxer             | 16 |
| doormat           | 16 |
| ..                |    |
| chickadee         | 1  |
| moped             | 1  |
| quill             | 1  |
| bullfrog          | 1  |
| gorilla           | 1  |
| golfcart          | 1  |
| toyshop           | 1  |
| pot               | 1  |
| notebook          | 1  |
| wallet            | 1  |
| barbell           | 1  |
| wombat            | 1  |
| pool_table        | 1  |
| green_lizard      | 1  |
| Band_Aid          | 1  |
| screen            | 1  |
| cliff             | 1  |
| French_horn       | 1  |
| red_wolf          | 1  |
| chest             | 1  |
| wing              | 1  |
| guillotine        | 1  |
| power_drill       | 1  |
| joystick          | 1  |
| sea_cucumber      | 1  |
| maze              | 1  |
| cup               | 1  |
| crossword_puzzle  | 1  |
| pickup            | 1  |
| vacuum            | 1  |

Name: p3, Length: 408, dtype: int64

In [39]: imagePrediction.p1\_dog.value\_counts()

Out[39]: True        1532  
False        543  
Name: p1\_dog, dtype: int64

In [40]: imagePrediction.p2\_dog.value\_counts()

Out[40]: True        1553  
False        522  
Name: p2\_dog, dtype: int64

```
In [41]: imagePrediction.p3_dog.value_counts()
```

```
Out[41]: True      1499
         False     576
         Name: p3_dog, dtype: int64
```

## 2.3 imagePrediction table

```
In [42]: tweet_json.head()
```

```
Out[42]:
```

|   | tweet_id            | retweet_count | favorite_count |
|---|---------------------|---------------|----------------|
| 0 | 1090301319825436672 | 2             | 305            |
| 1 | 1090300780270149632 | 11            | 1226           |
| 2 | 1090300624657252352 | 6989          | 79691          |
| 3 | 1090060665903927296 | 1             | 143            |
| 4 | 1090060504788197377 | 1             | 116            |

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2157 entries, 0 to 2156
Data columns (total 3 columns):
tweet_id      2157 non-null int64
retweet_count 2157 non-null int64
favorite_count 2157 non-null int64
dtypes: int64(3)
memory usage: 50.6 KB
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2157 entries, 0 to 2156
Data columns (total 3 columns):
tweet_id      2157 non-null int64
retweet_count 2157 non-null int64
favorite_count 2157 non-null int64
dtypes: int64(3)
memory usage: 50.6 KB
```

```
In [45]: tweet_json.describe()
```

```
Out[45]:
```

|       | tweet_id     | retweet_count | favorite_count |
|-------|--------------|---------------|----------------|
| count | 2.157000e+03 | 2157.000000   | 2157.000000    |
| mean  | 1.043777e+18 | 2811.560037   | 10408.850255   |
| std   | 1.998247e+16 | 19050.964888  | 40845.247559   |
| min   | 1.027963e+18 | 0.000000      | 0.000000       |
| 25%   | 1.028138e+18 | 0.000000      | 0.000000       |
| 50%   | 1.029407e+18 | 0.000000      | 1.000000       |
| 75%   | 1.060962e+18 | 4.000000      | 357.000000     |
| max   | 1.090301e+18 | 466025.000000 | 690028.000000  |

## 2.4 Merge the table names a

```
In [46]: all_columns= pd.Series(list/archive) +list(imagePrediction)+ list(tweet_json))
all_columns[all_columns.duplicated()]
```

```
Out[46]: 17    tweet_id
          29    tweet_id
dtype: object
```

## 4 Clean

### 4.0.1 1. Tidiness

copy all data frames before making any cleaning process using copy() function to save them in case you need to return back with the original

```
In [47]: d1 = archive.copy()
```

```
In [48]: Archive_clean = archive.copy()
imagePrediction_clean = imagePrediction.copy()
```

#### Tidiness

**doggo, floofer, pupper, and puppo column in archive table are all describing dog age stage.**

**Define** The doggo, floofer, pupper, and puppo column should be merged into one column called dog\_stage to reduce redundancy.

#### Code

```
In [49]: Archive_clean.replace('None', np.nan, inplace=True)
Archive_clean['dog_stage'] = Archive_clean[['doggo', 'floofer', 'pupper', 'puppo']].fil
#Drop redundant columns
Archive_clean = Archive_clean.drop(['doggo', 'floofer', 'pupper', 'puppo'], axis=1)
```

#### Test

```
In [50]: Archive_clean['dog_stage'].value_counts()
```

```
Out[50]:
pupper      245
doggo        83
puppo        29
doggopupper  12
floofer       9
doggopuppo   1
doggofloofer  1
Name: dog_stage, dtype: int64
```

There are 1 doggo and puppo value and 1 doggo and floofer value, and 12 doggo pupper value. This may happen when text has both word. I will need to check these rows one by one to verify if it is right.

```
In [51]: Archive_clean[Archive_clean['dog_stage'] == 'doggopuppo'] #It's clearly a puppo
```

```
Out[51]:
```

|     | tweet_id            | in_reply_to_status_id    | in_reply_to_user_id        | \          | timestamp   | \ | source   |
|-----|---------------------|--------------------------|----------------------------|------------|---|---|--|
| 191 | 855851453814013952  | NaN                      | NaN                        |            | 2017-04-22 18:31:02 +0000                                       |   | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 191 |                     |                          |                            |            |   |   | Here's a puppo participating in the #ScienceMarch. Cleverly disguising her own dog |
|     | retweeted_status_id | retweeted_status_user_id | retweeted_status_timestamp | \          | expanded_urls   | \ |  |
| 191 | NaN                 | NaN                      | NaN                        |            | https://twitter.com/dog_rates/status/855851453814013952/photo/1 |   |  |
|     | rating_numerator    | rating_denominator       | name                       | dog_stage  |   |   |  |
| 191 | 13                  | 10                       | NaN                        | doggopuppo |   |   |  |

```
In [52]: Archive_clean.loc[Archive_clean['dog_stage'] == 'doggopuppo', 'dog_stage'] = 'puppo'
```

```
In [53]: Archive_clean[Archive_clean['dog_stage'] == 'doggofloofer'] #this should be a floofer
```

```
Out[53]:
```

|     | tweet_id            | in_reply_to_status_id    | in_reply_to_user_id        | \ | timestamp                 | \ | source   |
|-----|---------------------|--------------------------|----------------------------|---|---------------------------|---|--|
| 200 | 854010172552949760  | NaN                      | NaN                        |   | 2017-04-17 16:34:26 +0000 |   | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 200 |                     |                          |                            |   |                           |   | At first I thought this was a shy doggo, but it's actually a Rare Canadian Floofer |
|     | retweeted_status_id | retweeted_status_user_id | retweeted_status_timestamp | \ |                           |   |  |
| 200 | NaN                 | NaN                      | NaN                        |   |                           |   |  |
| 200 |                     |                          |                            |   |                           |   | https://twitter.com/dog_rates/status/854010172552949760/photo/1,https://twitter.co |

|     | rating_numerator | rating_denominator | name | dog_stage    |
|-----|------------------|--------------------|------|--------------|
| 200 | 11               | 10                 | NaN  | doggofloofer |

In [54]: Archive\_clean.loc[Archive\_clean['dog\_stage'] == 'doggofloofer', 'dog\_stage'] = 'floofer'

In [55]: Archive\_clean[Archive\_clean['dog\_stage'] == 'doggopupper']  
*#Some tweets have both doggo and pupper for two dogs, this will be kept and used as one*

Out[55]:

|      | tweet_id           | in_reply_to_status_id | in_reply_to_user_id | \ |
|------|--------------------|-----------------------|---------------------|---|
| 460  | 817777686764523521 | NaN                   | NaN                 |   |
| 531  | 808106460588765185 | NaN                   | NaN                 |   |
| 565  | 802265048156610565 | 7.331095e+17          | 4.196984e+09        |   |
| 575  | 801115127852503040 | NaN                   | NaN                 |   |
| 705  | 785639753186217984 | NaN                   | NaN                 |   |
| 733  | 781308096455073793 | NaN                   | NaN                 |   |
| 778  | 775898661951791106 | NaN                   | NaN                 |   |
| 822  | 77009376776997377  | NaN                   | NaN                 |   |
| 889  | 759793422261743616 | NaN                   | NaN                 |   |
| 956  | 751583847268179968 | NaN                   | NaN                 |   |
| 1063 | 741067306818797568 | NaN                   | NaN                 |   |
| 1113 | 733109485275860992 | NaN                   | NaN                 |   |

|      | timestamp                 | \ |
|------|---------------------------|---|
| 460  | 2017-01-07 16:59:28 +0000 |   |
| 531  | 2016-12-12 00:29:28 +0000 |   |
| 565  | 2016-11-25 21:37:47 +0000 |   |
| 575  | 2016-11-22 17:28:25 +0000 |   |
| 705  | 2016-10-11 00:34:48 +0000 |   |
| 733  | 2016-09-29 01:42:20 +0000 |   |
| 778  | 2016-09-14 03:27:11 +0000 |   |
| 822  | 2016-08-29 03:00:36 +0000 |   |
| 889  | 2016-07-31 16:50:42 +0000 |   |
| 956  | 2016-07-09 01:08:47 +0000 |   |
| 1063 | 2016-06-10 00:39:48 +0000 |   |
| 1113 | 2016-05-19 01:38:16 +0000 |   |

|      | source   |
|------|--|
| 460  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 531  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 565  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 575  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 705  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 733  | <a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>                    |
| 778  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 822  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 889  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 956  | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |
| 1063 | <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> |

1113 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

460 This is Dido. She's playing the lead role in "Pupper Stops to Catch Snow Before R  
531 Here we have Burke (pupper) and Dexter (doggo). Pupper wants to be exactly like d  
565 Like doggo, like pupper version 2. Both 11/10 https://t.co/9IxWAXFqze  
575 This is Bones. He's being haunted by another doggo of roughly the same size. 12/1  
705 This is Pinot. He's a sophisticated doggo. You can tell by the hat. Also pointier  
733 Pupper butt 1, Doggo 0. Both 12/10 https://t.co/WQvcPEpH2u  
778 RT @dog\_rates: Like father (doggo), like son (pupper). Both 12/10 https://t.co/pG  
822 RT @dog\_rates: This is just downright precious af. 12/10 for both pupper and dogg  
889 Meet Maggie & Lila. Maggie is the doggo, Lila is the pupper. They are sisters  
956 Please stop sending it pictures that don't even have a doggo or pupper in them. C  
1063 This is just downright precious af. 12/10 for both pupper and doggo https://t.co/  
1113 Like father (doggo), like son (pupper). Both 12/10 https://t.co/pG2inLa0da

|      | retweeted_status_id | retweeted_status_user_id \ |
|------|---------------------|----------------------------|
| 460  | NaN                 | NaN                        |
| 531  | NaN                 | NaN                        |
| 565  | NaN                 | NaN                        |
| 575  | NaN                 | NaN                        |
| 705  | NaN                 | NaN                        |
| 733  | NaN                 | NaN                        |
| 778  | 7.331095e+17        | 4.196984e+09               |
| 822  | 7.410673e+17        | 4.196984e+09               |
| 889  | NaN                 | NaN                        |
| 956  | NaN                 | NaN                        |
| 1063 | NaN                 | NaN                        |
| 1113 | NaN                 | NaN                        |

|      | retweeted_status_timestamp \ |
|------|------------------------------|
| 460  | NaN                          |
| 531  | NaN                          |
| 565  | NaN                          |
| 575  | NaN                          |
| 705  | NaN                          |
| 733  | NaN                          |
| 778  | 2016-05-19 01:38:16 +0000    |
| 822  | 2016-06-10 00:39:48 +0000    |
| 889  | NaN                          |
| 956  | NaN                          |
| 1063 | NaN                          |
| 1113 | NaN                          |

460 https://twitter.com/dog\_rates/status/817777686764523521/video/1  
531 https://twitter.com/dog\_rates/status/808106460588765185/photo/1  
565 https://twitter.com/dog\_rates/status/802265048156610565/photo/1

```

575 https://twitter.com/dog_rates/status/801115127852503040/photo/1,https://twitter.c
705 https://twitter.com/dog_rates/status/785639753186217984/photo/1,https://twitter.c
733 https://vine.co/v/5rgu2Law2ut
778 https://twitter.com/dog_rates/status/733109485275860992/photo/1,https://twitter.c
822 https://twitter.com/dog_rates/status/741067306818797568/photo/1,https://twitter.c
889 https://twitter.com/dog_rates/status/759793422261743616/photo/1,https://twitter.c
956 https://twitter.com/dog_rates/status/751583847268179968/photo/1
1063 https://twitter.com/dog_rates/status/741067306818797568/photo/1
1113 https://twitter.com/dog_rates/status/733109485275860992/photo/1

```

|      | rating_numerator | rating_denominator | name   | dog_stage   |
|------|------------------|--------------------|--------|-------------|
| 460  | 13               | 10                 | Dido   | doggopupper |
| 531  | 12               | 10                 | NaN    | doggopupper |
| 565  | 11               | 10                 | NaN    | doggopupper |
| 575  | 12               | 10                 | Bones  | doggopupper |
| 705  | 10               | 10                 | Pinot  | doggopupper |
| 733  | 12               | 10                 | NaN    | doggopupper |
| 778  | 12               | 10                 | NaN    | doggopupper |
| 822  | 12               | 10                 | just   | doggopupper |
| 889  | 12               | 10                 | Maggie | doggopupper |
| 956  | 5                | 10                 | NaN    | doggopupper |
| 1063 | 12               | 10                 | just   | doggopupper |
| 1113 | 12               | 10                 | NaN    | doggopupper |

```
In [56]: Archive_clean['dog_stage'].value_counts()
```

```

Out[56]:
pupper      1976
doggo       245
puppo       83
doggopupper 30
floofer     12
Name: dog_stage, dtype: int64

```

The result looks correct. Let's work on the second tidy data issue - The column names in imagePrediction are not specific and not easy to understand. ### Define

Rename column names -(p1,p1\_conf,p1\_dog,p2,p2\_conf,p2\_dog,p3,p3\_conf,p3\_dog) in imagePrediction to make them user friendly

#### 4.0.2 Code

```

In [57]: imagePrediction_clean.rename(index=str, columns={"p1" : "prediction1", "p1_conf": "predic
            "p2" : "prediction2", "p2_conf": "predict
            "p3" : "prediction3", "p3_conf": "predic
            "img_num" : "most_confident_img_num"}),

```

#### 4.0.3 Test

```
In [58]: imagePrediction_clean.head(2)
```



```

Out [58]:
      tweet_id      jpg_url \
0  666020888022790149  https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg
1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg

      most_confident_img_num      prediction1      prediction_confidence1 \
0  1      Welsh_springer_spaniel      0.465074
1  1      redbone      0.506826

      dogbreed_indicator1      prediction2      prediction_confidence2 \
0  True      collie      0.156665
1  True      miniature_pinscher      0.074192

      dogbreed_indicator2      prediction3      prediction_confidence3 \
0  True      Shetland_sheepdog      0.061428
1  True      Rhodesian_ridgeback      0.072010

      dogbreed_indicator3
0  True
1  True

```

The columns name in imagePrediction table is changed and look make sense.

#### 4.0.4 Define

The last step should be merging tweets\_json table with Twitter\_Archive Table on tweet\_id column.

#### 4.0.5 Code

```

In [59]: Archive_clean = pd.merge(Archive_clean, tweet_json,
                                on=['tweet_id'], how='left')

```

Here, I will keep Archive\_clean table and image prediction table both for tidiness of data.

#### 4.0.6 Test

```

In [60]: Archive_clean.head()

```

```

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

      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" rel="nofollow">Twitter for iPhone</a>
1 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
2 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
3 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
4 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
```

```
0 This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/1
1 This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she's
2 This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall grass. You
3 This is Darla. She commenced a snooze mid meal. 13/10 happens to the best of us http
4 This is Franklin. He would like you to stop calling him "cute." He is a very fierce
```

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

```
0 https://twitter.com/dog_rates/status/892420643555336193/photo/1
1 https://twitter.com/dog_rates/status/892177421306343426/photo/1
2 https://twitter.com/dog_rates/status/891815181378084864/photo/1
3 https://twitter.com/dog_rates/status/891689557279858688/photo/1
4 https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://twitter.com/
```

```
rating_numerator rating_denominator name dog_stage retweet_count \
0 13 10 Phineas NaN
1 13 10 Tilly NaN
2 12 10 Archie NaN
3 13 10 Darla NaN
4 12 10 Franklin NaN
```

```
favorite_count
0 NaN
1 NaN
2 NaN
3 NaN
4 NaN
```

Here, I will keep Archive\_clean table and image prediction table both for tidiness of data.

## 5 Quality issues

Missing data is part of data quality issue.

### 5.0.1 Define

expanded\_urls column has 59 url with missing value, I can fill them up by combining 'https://twitter.com/dog\_rates/status/' with their tweet id and connect to '/photo/1'

```
In [61]: Archive_clean['expanded_urls'] = Archive_clean['expanded_urls'].fillna(
        value = 'https://twitter.com/dog_rates/status/' + Archive_clean['tweet_id'].astype(st
```

### 5.0.2 Test

```
In [67]: sum(Archive_clean['expanded_urls'].isnull())
```

```
Out[67]: 0
```

From the result, we can see all the missing value in expanded\_urls column is filled with correct url

in\_reply\_to\_status\_id,in\_reply\_to\_user\_id all only have 78 non-null value, 2278 missing value, 97% field in these columns are missing. retweeted\_status\_id,retweeted\_status\_user\_id and retweeted\_status\_timestamp have 2175 missing value out of 2356, about 92% fields are missing.

## 6 Define

Since we don't use reweet data, I will delete the rows populated with values under these columns and remove these columns which have a large amount of missing value.

## 7 Code

```
In [63]: Archive_clean[Archive_clean["in_reply_to_status_id"].isnull() ==False]
```

```
Out[63]:
```

|     | tweet_id           | in_reply_to_status_id | in_reply_to_user_id \ |
|-----|--------------------|-----------------------|-----------------------|
| 30  | 886267009285017600 | 8.862664e+17          | 2.281182e+09          |
| 55  | 881633300179243008 | 8.816070e+17          | 4.738443e+07          |
| 64  | 879674319642796034 | 8.795538e+17          | 3.105441e+09          |
| 113 | 870726314365509632 | 8.707262e+17          | 1.648776e+07          |
| 148 | 863427515083354112 | 8.634256e+17          | 7.759620e+07          |
| 149 | 863079547188785154 | 6.671522e+17          | 4.196984e+09          |
| 179 | 857214891891077121 | 8.571567e+17          | 1.806710e+08          |
| 184 | 856526610513747968 | 8.558181e+17          | 4.196984e+09          |
| 186 | 856288084350160898 | 8.562860e+17          | 2.792810e+08          |
| 188 | 855862651834028034 | 8.558616e+17          | 1.943518e+08          |
| 189 | 855860136149123072 | 8.558585e+17          | 1.361572e+07          |
| 218 | 850333567704068097 | 8.503288e+17          | 2.195506e+07          |
| 228 | 848213670039564288 | 8.482121e+17          | 4.196984e+09          |
| 234 | 847617282490613760 | 8.476062e+17          | 4.196984e+09          |

|      |                    |              |              |
|------|--------------------|--------------|--------------|
| 251  | 844979544864018432 | 7.590995e+17 | 4.196984e+09 |
| 274  | 840698636975636481 | 8.406983e+17 | 8.405479e+17 |
| 290  | 838150277551247360 | 8.381455e+17 | 2.195506e+07 |
| 291  | 838085839343206401 | 8.380855e+17 | 2.894131e+09 |
| 313  | 835246439529840640 | 8.352460e+17 | 2.625958e+07 |
| 342  | 832088576586297345 | 8.320875e+17 | 3.058208e+07 |
| 346  | 831926988323639298 | 8.319030e+17 | 2.068372e+07 |
| 387  | 826598799820865537 | 8.265984e+17 | 4.196984e+09 |
| 409  | 823333489516937216 | 8.233264e+17 | 1.582854e+09 |
| 427  | 821153421864615936 | 8.211526e+17 | 1.132119e+08 |
| 498  | 813130366689148928 | 8.131273e+17 | 4.196984e+09 |
| 513  | 811647686436880384 | 8.116272e+17 | 4.196984e+09 |
| 565  | 802265048156610565 | 7.331095e+17 | 4.196984e+09 |
| 570  | 801854953262350336 | 8.018543e+17 | 1.185634e+07 |
| 576  | 800859414831898624 | 8.008580e+17 | 2.918590e+08 |
| 611  | 797165961484890113 | 7.971238e+17 | 2.916630e+07 |
| ...  | ...                | ...          | ...          |
| 1479 | 693582294167244802 | 6.935722e+17 | 1.198989e+09 |
| 1497 | 692423280028966913 | 6.924173e+17 | 4.196984e+09 |
| 1501 | 692142790915014657 | 6.920419e+17 | 4.196984e+09 |
| 1523 | 690607260360429569 | 6.903413e+17 | 4.670367e+08 |
| 1598 | 686035780142297088 | 6.860340e+17 | 4.196984e+09 |
| 1605 | 685681090388975616 | 6.855479e+17 | 4.196984e+09 |
| 1618 | 684969860808454144 | 6.849598e+17 | 4.196984e+09 |
| 1630 | 684538444857667585 | 6.844811e+17 | 4.196984e+09 |
| 1634 | 684225744407494656 | 6.842229e+17 | 4.196984e+09 |
| 1663 | 682808988178739200 | 6.827884e+17 | 4.196984e+09 |
| 1689 | 681340665377193984 | 6.813394e+17 | 4.196984e+09 |
| 1774 | 678023323247357953 | 6.780211e+17 | 4.196984e+09 |
| 1819 | 676590572941893632 | 6.765883e+17 | 4.196984e+09 |
| 1842 | 675870721063669760 | 6.757073e+17 | 4.196984e+09 |
| 1844 | 675849018447167488 | 6.758457e+17 | 4.196984e+09 |
| 1852 | 675707330206547968 | 6.754971e+17 | 4.196984e+09 |
| 1866 | 675349384339542016 | 6.749998e+17 | 4.196984e+09 |
| 1882 | 674999807681908736 | 6.747934e+17 | 4.196984e+09 |
| 1885 | 674793399141146624 | 6.717299e+17 | 4.196984e+09 |
| 1892 | 674754018082705410 | 6.747522e+17 | 4.196984e+09 |
| 1895 | 674742531037511680 | 6.747400e+17 | 4.196984e+09 |
| 1905 | 674606911342424069 | 6.744689e+17 | 4.196984e+09 |
| 1914 | 674330906434379776 | 6.658147e+17 | 1.637468e+07 |
| 1940 | 673716320723169284 | 6.737159e+17 | 4.196984e+09 |
| 2036 | 671729906628341761 | 6.715610e+17 | 4.196984e+09 |
| 2038 | 671550332464455680 | 6.715449e+17 | 4.196984e+09 |
| 2149 | 669684865554620416 | 6.693544e+17 | 4.196984e+09 |
| 2169 | 669353438988365824 | 6.678065e+17 | 4.196984e+09 |
| 2189 | 668967877119254528 | 6.689207e+17 | 2.143566e+07 |
| 2298 | 667070482143944705 | 6.670655e+17 | 4.196984e+09 |

|      | timestamp \               |
|------|---------------------------|
| 30   | 2017-07-15 16:51:35 +0000 |
| 55   | 2017-07-02 21:58:53 +0000 |
| 64   | 2017-06-27 12:14:36 +0000 |
| 113  | 2017-06-02 19:38:25 +0000 |
| 148  | 2017-05-13 16:15:35 +0000 |
| 149  | 2017-05-12 17:12:53 +0000 |
| 179  | 2017-04-26 12:48:51 +0000 |
| 184  | 2017-04-24 15:13:52 +0000 |
| 186  | 2017-04-23 23:26:03 +0000 |
| 188  | 2017-04-22 19:15:32 +0000 |
| 189  | 2017-04-22 19:05:32 +0000 |
| 218  | 2017-04-07 13:04:55 +0000 |
| 228  | 2017-04-01 16:41:12 +0000 |
| 234  | 2017-03-31 01:11:22 +0000 |
| 251  | 2017-03-23 18:29:57 +0000 |
| 274  | 2017-03-11 22:59:09 +0000 |
| 290  | 2017-03-04 22:12:52 +0000 |
| 291  | 2017-03-04 17:56:49 +0000 |
| 313  | 2017-02-24 21:54:03 +0000 |
| 342  | 2017-02-16 04:45:50 +0000 |
| 346  | 2017-02-15 18:03:45 +0000 |
| 387  | 2017-02-01 01:11:25 +0000 |
| 409  | 2017-01-23 00:56:15 +0000 |
| 427  | 2017-01-17 00:33:26 +0000 |
| 498  | 2016-12-25 21:12:41 +0000 |
| 513  | 2016-12-21 19:01:02 +0000 |
| 565  | 2016-11-25 21:37:47 +0000 |
| 570  | 2016-11-24 18:28:13 +0000 |
| 576  | 2016-11-22 00:32:18 +0000 |
| 611  | 2016-11-11 19:55:50 +0000 |
| ...  | ...                       |
| 1479 | 2016-01-30 23:51:19 +0000 |
| 1497 | 2016-01-27 19:05:49 +0000 |
| 1501 | 2016-01-27 00:31:15 +0000 |
| 1523 | 2016-01-22 18:49:36 +0000 |
| 1598 | 2016-01-10 04:04:10 +0000 |
| 1605 | 2016-01-09 04:34:45 +0000 |
| 1618 | 2016-01-07 05:28:35 +0000 |
| 1630 | 2016-01-06 00:54:18 +0000 |
| 1634 | 2016-01-05 04:11:44 +0000 |
| 1663 | 2016-01-01 06:22:03 +0000 |
| 1689 | 2015-12-28 05:07:27 +0000 |
| 1774 | 2015-12-19 01:25:31 +0000 |
| 1819 | 2015-12-15 02:32:17 +0000 |
| 1842 | 2015-12-13 02:51:51 +0000 |
| 1844 | 2015-12-13 01:25:37 +0000 |
| 1852 | 2015-12-12 16:02:36 +0000 |

1866 2015-12-11 16:20:15 +0000  
 1882 2015-12-10 17:11:09 +0000  
 1885 2015-12-10 03:30:58 +0000  
 1892 2015-12-10 00:54:28 +0000  
 1895 2015-12-10 00:08:50 +0000  
 1905 2015-12-09 15:09:55 +0000  
 1914 2015-12-08 20:53:11 +0000  
 1940 2015-12-07 04:11:02 +0000  
 2036 2015-12-01 16:37:44 +0000  
 2038 2015-12-01 04:44:10 +0000  
 2149 2015-11-26 01:11:28 +0000  
 2169 2015-11-25 03:14:30 +0000  
 2189 2015-11-24 01:42:25 +0000  
 2298 2015-11-18 20:02:51 +0000

30 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a> source  
 55 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 64 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 113 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 148 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 149 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 179 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 184 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 186 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 188 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 189 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 218 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 228 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 234 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 251 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 274 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 290 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 291 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 313 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 342 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 346 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 387 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 409 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 427 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 498 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 513 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 565 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 570 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 576 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 611 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
 ... ..  
 1479 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

1497 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1501 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1523 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1598 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1605 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1618 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1630 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1634 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1663 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1689 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1774 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1819 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1842 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1844 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1852 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1866 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1882 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1885 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1892 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1895 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1905 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1914 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
1940 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
2036 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
2038 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
2149 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
2169 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
2189 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>  
2298 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>

30 @NonWhiteHat @MayhewMayhem omg hello tanner you are a scary good boy 12/10 would  
55 @roushfenway These are good dogs but 17/10 is an emotional impulse rating. More 1  
64 @RealKentMurphy 14/10 confirmed  
113 @ComplicitOwl @ShopWeRateDogs &gt;10/10 is reserved for dogs  
148 @Jack\_Septic\_Eye I'd need a few more pics to polish a full analysis, but based on  
149 Ladies and gentlemen... I found Pipsy. He may have changed his name to Pablo, but  
179 @Marc\_IRL pixelated af 12/10  
184 THIS IS CHARLIE, MARK. HE DID JUST WANT TO SAY HI AFTER ALL. PUPGRADED TO A 14/10  
186 @xianmcguire @Jenna\_Marbles Kardashians wouldn't be famous if as a society we did  
188 @dhmontgomery We also gave snoop dogg a 420/10 but I think that predated your res  
189 @s8n You tried very hard to portray this good boy as not so good, but you have ul  
218 @markhoppus MARK THAT DOG HAS SEEN AND EXPERIENCED MANY THINGS. PROBABLY LOST OTH  
228 Jerry just apuppologized to me. He said there was no ill-intent to the slippage.  
234 .@breaannanicolee PUPDATE: Cannon has a heart on his nose. Pupgraded to a 13/10  
251 PUPDATE: I'm proud to announce that Toby is 236 days sober. Pupgraded to a 13/10.  
274 @O\_kelvin\_0 &gt;10/10 is reserved for puppos sorry Kevin  
290 @markhoppus 182/10

291 @bragg6of8 @Andy\_Pace\_ we are still looking for the first 15/10  
 313 @jonnyusun @Lin\_Manuel ok jomny I know you're excited but 960/00 isn't a valid rat  
 342 @docmisterio account started on 11/15/15  
 346 @UNC can confirm 12/10  
 387 I was going to do 007/10, but the joke wasn't worth the &lt;10 rating  
 409 @HistoryInPics 13/10  
 427 @imgur for a polar bear tho I'd say 13/10 is appropriate  
 498 I've been informed by multiple sources that this is actually a dog elf who's tire  
 513 PUPDATE: I've been informed that Augie was actually bringing his family these flo  
 565 Like doggo, like pupper version 2. Both 11/10 <https://t.co/9IxWAXFqze>  
 570 .@NBCSports OMG THE TINY HAT I'M GOING TO HAVE TO SAY 11/10 NBC  
 576 @SkyWilliams doggo simply protecting you from evil that which you cannot see. 11/  
 611 @JODYHiGROLLER it may be an 11/10 but what do I know  
 ...  
 1479 Personally I'd give him an 11/10. Not sure why you think you're qualified to rate  
 1497 PUPDATE: just noticed this dog has some extra legs. Very advanced. Revolutionary  
 1501 These are some pictures of Teddy that further justify his 13/10 rating. Please em  
 1523 12/10 @LightningHoltt  
 1598 Yes I do realize a rating of 4/20 would've been fitting. However, it would be unj  
 1605 Jack deserves another round of applause. If you missed this earlier today I stron  
 1618 For those who claim this is a goat, u are wrong. It is not the Greatest Of All Ti  
 1630 After watching this video, we've determined that Pippa will be upgraded to a 12/1  
 1634 Two sneaky puppies were not initially seen, moving the rating to 143/130. Please  
 1663 I'm aware that I could've said 20/16, but here at WeRateDogs we are very professi  
 1689 I've been told there's a slight possibility he's checking his mirror. We'll bump  
 1774 After getting lost in Reese's eyes for several minutes we're going to upgrade him  
 1819 After some outrage from the crowd. Bubbles is being upgraded to a 7/10. That's as  
 1842 & this is Yoshi. Another world record contender 11/10 (what the hell is happe  
 1844 This dog is being demoted to a 9/10 for not wearing a helmet while riding. Gotta  
 1852 We've got ourselves a battle here. Watch out Reggie. 11/10 <https://t.co/ALJvbtcfw>  
 1866 Yea I lied. Here's more. All 13/10 <https://t.co/ZQZf2U4xCP>  
 1882 Ok last one of these. I may try to make some myself. Anyway here ya go. 13/10 htt  
 1885 I have found another. 13/10 <https://t.co/HwroPYv8pY>  
 1892 Just received another perfect photo of dogs and the sunset. 12/10 <https://t.co/9Y>  
 1895 Some clarification is required. The dog is singing Cher and that is more than wor  
 1905 The 13/10 also takes into account this impeccable yard. Louis is great but the fu  
 1914 13/10\n@ABC7  
 1940 The millennials have spoken and we've decided to immediately demote to a 1/10. Th  
 2036 I'm just going to leave this one here as well. 13/10 <https://t.co/DaD5SyajWt>  
 2038 After 22 minutes of careful deliberation this dog is being demoted to a 1/10. The  
 2149 After countless hours of research and hundreds of formula alterations we have com  
 2169 This is Tessa. She is also very pleased after finally meeting her biological fath  
 2189 12/10 good shit Bubka\n@wane15  
 2298 After much debate this dog is being upgraded to 10/10. I repeat 10/10

|    | retweeted_status_id | retweeted_status_user_id | \ |
|----|---------------------|--------------------------|---|
| 30 | NaN                 | NaN                      |   |
| 55 | NaN                 | NaN                      |   |



|      |     |     |
|------|-----|-----|
| 64   | NaN | NaN |
| 113  | NaN | NaN |
| 148  | NaN | NaN |
| 149  | NaN | NaN |
| 179  | NaN | NaN |
| 184  | NaN | NaN |
| 186  | NaN | NaN |
| 188  | NaN | NaN |
| 189  | NaN | NaN |
| 218  | NaN | NaN |
| 228  | NaN | NaN |
| 234  | NaN | NaN |
| 251  | NaN | NaN |
| 274  | NaN | NaN |
| 290  | NaN | NaN |
| 291  | NaN | NaN |
| 313  | NaN | NaN |
| 342  | NaN | NaN |
| 346  | NaN | NaN |
| 387  | NaN | NaN |
| 409  | NaN | NaN |
| 427  | NaN | NaN |
| 498  | NaN | NaN |
| 513  | NaN | NaN |
| 565  | NaN | NaN |
| 570  | NaN | NaN |
| 576  | NaN | NaN |
| 611  | NaN | NaN |
| ...  | ..  | ..  |
| 1479 | NaN | NaN |
| 1497 | NaN | NaN |
| 1501 | NaN | NaN |
| 1523 | NaN | NaN |
| 1598 | NaN | NaN |
| 1605 | NaN | NaN |
| 1618 | NaN | NaN |
| 1630 | NaN | NaN |
| 1634 | NaN | NaN |
| 1663 | NaN | NaN |
| 1689 | NaN | NaN |
| 1774 | NaN | NaN |
| 1819 | NaN | NaN |
| 1842 | NaN | NaN |
| 1844 | NaN | NaN |
| 1852 | NaN | NaN |
| 1866 | NaN | NaN |
| 1882 | NaN | NaN |
| 1885 | NaN | NaN |

|      |     |     |
|------|-----|-----|
| 1892 | NaN | NaN |
| 1895 | NaN | NaN |
| 1905 | NaN | NaN |
| 1914 | NaN | NaN |
| 1940 | NaN | NaN |
| 2036 | NaN | NaN |
| 2038 | NaN | NaN |
| 2149 | NaN | NaN |
| 2169 | NaN | NaN |
| 2189 | NaN | NaN |
| 2298 | NaN | NaN |

|      | retweeted_status_timestamp \ |
|------|------------------------------|
| 30   | NaN                          |
| 55   | NaN                          |
| 64   | NaN                          |
| 113  | NaN                          |
| 148  | NaN                          |
| 149  | NaN                          |
| 179  | NaN                          |
| 184  | NaN                          |
| 186  | NaN                          |
| 188  | NaN                          |
| 189  | NaN                          |
| 218  | NaN                          |
| 228  | NaN                          |
| 234  | NaN                          |
| 251  | NaN                          |
| 274  | NaN                          |
| 290  | NaN                          |
| 291  | NaN                          |
| 313  | NaN                          |
| 342  | NaN                          |
| 346  | NaN                          |
| 387  | NaN                          |
| 409  | NaN                          |
| 427  | NaN                          |
| 498  | NaN                          |
| 513  | NaN                          |
| 565  | NaN                          |
| 570  | NaN                          |
| 576  | NaN                          |
| 611  | NaN                          |
| ...  | ...                          |
| 1479 | NaN                          |
| 1497 | NaN                          |
| 1501 | NaN                          |
| 1523 | NaN                          |

1598 NaN  
 1605 NaN  
 1618 NaN  
 1630 NaN  
 1634 NaN  
 1663 NaN  
 1689 NaN  
 1774 NaN  
 1819 NaN  
 1842 NaN  
 1844 NaN  
 1852 NaN  
 1866 NaN  
 1882 NaN  
 1885 NaN  
 1892 NaN  
 1895 NaN  
 1905 NaN  
 1914 NaN  
 1940 NaN  
 2036 NaN  
 2038 NaN  
 2149 NaN  
 2169 NaN  
 2189 NaN  
 2298 NaN

30 [https://twitter.com/dog\\_rates/status/886267009285017600/photo/1](https://twitter.com/dog_rates/status/886267009285017600/photo/1)  
 55 [https://twitter.com/dog\\_rates/status/881633300179243008/photo/1](https://twitter.com/dog_rates/status/881633300179243008/photo/1)  
 64 [https://twitter.com/dog\\_rates/status/879674319642796034/photo/1](https://twitter.com/dog_rates/status/879674319642796034/photo/1)  
 113 [https://twitter.com/dog\\_rates/status/870726314365509632/photo/1](https://twitter.com/dog_rates/status/870726314365509632/photo/1)  
 148 [https://twitter.com/dog\\_rates/status/863427515083354112/photo/1](https://twitter.com/dog_rates/status/863427515083354112/photo/1)  
 149 [https://twitter.com/dog\\_rates/status/863079547188785154/photo/1](https://twitter.com/dog_rates/status/863079547188785154/photo/1)  
 179 [https://twitter.com/dog\\_rates/status/857214891891077121/photo/1](https://twitter.com/dog_rates/status/857214891891077121/photo/1)  
 184 [https://twitter.com/dog\\_rates/status/856526610513747968/photo/1](https://twitter.com/dog_rates/status/856526610513747968/photo/1)  
 186 [https://twitter.com/dog\\_rates/status/856288084350160898/photo/1](https://twitter.com/dog_rates/status/856288084350160898/photo/1)  
 188 [https://twitter.com/dog\\_rates/status/855862651834028034/photo/1](https://twitter.com/dog_rates/status/855862651834028034/photo/1)  
 189 [https://twitter.com/dog\\_rates/status/855860136149123072/photo/1](https://twitter.com/dog_rates/status/855860136149123072/photo/1)  
 218 [https://twitter.com/dog\\_rates/status/850333567704068097/photo/1](https://twitter.com/dog_rates/status/850333567704068097/photo/1)  
 228 [https://twitter.com/dog\\_rates/status/848213670039564288/photo/1](https://twitter.com/dog_rates/status/848213670039564288/photo/1)  
 234 [https://twitter.com/dog\\_rates/status/847617282490613760/photo/1](https://twitter.com/dog_rates/status/847617282490613760/photo/1)  
 251 [https://twitter.com/dog\\_rates/status/844979544864018432/photo/1](https://twitter.com/dog_rates/status/844979544864018432/photo/1), [https://twitter.com/dog\\_rates/status/844979544864018432/photo/1](https://twitter.com/dog_rates/status/844979544864018432/photo/1)  
 274 [https://twitter.com/dog\\_rates/status/840698636975636481/photo/1](https://twitter.com/dog_rates/status/840698636975636481/photo/1)  
 290 [https://twitter.com/dog\\_rates/status/838150277551247360/photo/1](https://twitter.com/dog_rates/status/838150277551247360/photo/1)  
 291 [https://twitter.com/dog\\_rates/status/838085839343206401/photo/1](https://twitter.com/dog_rates/status/838085839343206401/photo/1)  
 313 [https://twitter.com/dog\\_rates/status/835246439529840640/photo/1](https://twitter.com/dog_rates/status/835246439529840640/photo/1)  
 342 [https://twitter.com/dog\\_rates/status/832088576586297345/photo/1](https://twitter.com/dog_rates/status/832088576586297345/photo/1)

346 [https://twitter.com/dog\\_rates/status/831926988323639298/photo/1](https://twitter.com/dog_rates/status/831926988323639298/photo/1)  
 387 [https://twitter.com/dog\\_rates/status/826598799820865537/photo/1](https://twitter.com/dog_rates/status/826598799820865537/photo/1)  
 409 [https://twitter.com/dog\\_rates/status/823333489516937216/photo/1](https://twitter.com/dog_rates/status/823333489516937216/photo/1)  
 427 [https://twitter.com/dog\\_rates/status/821153421864615936/photo/1](https://twitter.com/dog_rates/status/821153421864615936/photo/1)  
 498 [https://twitter.com/dog\\_rates/status/813130366689148928/photo/1](https://twitter.com/dog_rates/status/813130366689148928/photo/1)  
 513 [https://twitter.com/dog\\_rates/status/811647686436880384/photo/1](https://twitter.com/dog_rates/status/811647686436880384/photo/1)  
 565 [https://twitter.com/dog\\_rates/status/802265048156610565/photo/1](https://twitter.com/dog_rates/status/802265048156610565/photo/1)  
 570 [https://twitter.com/dog\\_rates/status/801854953262350336/photo/1](https://twitter.com/dog_rates/status/801854953262350336/photo/1)  
 576 [https://twitter.com/dog\\_rates/status/800859414831898624/photo/1](https://twitter.com/dog_rates/status/800859414831898624/photo/1)  
 611 [https://twitter.com/dog\\_rates/status/797165961484890113/photo/1](https://twitter.com/dog_rates/status/797165961484890113/photo/1)  
 ...  
 1479 [https://twitter.com/dog\\_rates/status/693582294167244802/photo/1](https://twitter.com/dog_rates/status/693582294167244802/photo/1)  
 1497 [https://twitter.com/dog\\_rates/status/692423280028966913/photo/1](https://twitter.com/dog_rates/status/692423280028966913/photo/1)  
 1501 [https://twitter.com/dog\\_rates/status/692142790915014657/photo/1](https://twitter.com/dog_rates/status/692142790915014657/photo/1),[https://twitter.com/dog\\_rates/status/692142790915014657/photo/1](https://twitter.com/dog_rates/status/692142790915014657/photo/1)  
 1523 [https://twitter.com/dog\\_rates/status/690607260360429569/photo/1](https://twitter.com/dog_rates/status/690607260360429569/photo/1)  
 1598 [https://twitter.com/dog\\_rates/status/686035780142297088/photo/1](https://twitter.com/dog_rates/status/686035780142297088/photo/1)  
 1605 [https://twitter.com/dog\\_rates/status/685681090388975616/photo/1](https://twitter.com/dog_rates/status/685681090388975616/photo/1)  
 1618 [https://twitter.com/dog\\_rates/status/684969860808454144/photo/1](https://twitter.com/dog_rates/status/684969860808454144/photo/1)  
 1630 [https://twitter.com/dog\\_rates/status/684538444857667585/video/1](https://twitter.com/dog_rates/status/684538444857667585/video/1)  
 1634 [https://twitter.com/dog\\_rates/status/684225744407494656/photo/1](https://twitter.com/dog_rates/status/684225744407494656/photo/1),[https://twitter.com/dog\\_rates/status/684225744407494656/photo/1](https://twitter.com/dog_rates/status/684225744407494656/photo/1)  
 1663 [https://twitter.com/dog\\_rates/status/682808988178739200/photo/1](https://twitter.com/dog_rates/status/682808988178739200/photo/1)  
 1689 [https://twitter.com/dog\\_rates/status/681340665377193984/photo/1](https://twitter.com/dog_rates/status/681340665377193984/photo/1)  
 1774 [https://twitter.com/dog\\_rates/status/678023323247357953/photo/1](https://twitter.com/dog_rates/status/678023323247357953/photo/1)  
 1819 [https://twitter.com/dog\\_rates/status/676590572941893632/photo/1](https://twitter.com/dog_rates/status/676590572941893632/photo/1)  
 1842 [https://twitter.com/dog\\_rates/status/675870721063669760/photo/1](https://twitter.com/dog_rates/status/675870721063669760/photo/1)  
 1844 [https://twitter.com/dog\\_rates/status/675849018447167488/photo/1](https://twitter.com/dog_rates/status/675849018447167488/photo/1)  
 1852 [https://twitter.com/dog\\_rates/status/675707330206547968/photo/1](https://twitter.com/dog_rates/status/675707330206547968/photo/1)  
 1866 [https://twitter.com/dog\\_rates/status/675349384339542016/photo/1](https://twitter.com/dog_rates/status/675349384339542016/photo/1),[https://twitter.com/dog\\_rates/status/675349384339542016/photo/1](https://twitter.com/dog_rates/status/675349384339542016/photo/1)  
 1882 [https://twitter.com/dog\\_rates/status/674999807681908736/photo/1](https://twitter.com/dog_rates/status/674999807681908736/photo/1)  
 1885 [https://twitter.com/dog\\_rates/status/674793399141146624/photo/1](https://twitter.com/dog_rates/status/674793399141146624/photo/1)  
 1892 [https://twitter.com/dog\\_rates/status/674754018082705410/photo/1](https://twitter.com/dog_rates/status/674754018082705410/photo/1)  
 1895 [https://twitter.com/dog\\_rates/status/674742531037511680/photo/1](https://twitter.com/dog_rates/status/674742531037511680/photo/1)  
 1905 [https://twitter.com/dog\\_rates/status/674606911342424069/photo/1](https://twitter.com/dog_rates/status/674606911342424069/photo/1)  
 1914 [https://twitter.com/dog\\_rates/status/674330906434379776/photo/1](https://twitter.com/dog_rates/status/674330906434379776/photo/1)  
 1940 [https://twitter.com/dog\\_rates/status/673716320723169284/photo/1](https://twitter.com/dog_rates/status/673716320723169284/photo/1)  
 2036 [https://twitter.com/dog\\_rates/status/671729906628341761/photo/1](https://twitter.com/dog_rates/status/671729906628341761/photo/1)  
 2038 [https://twitter.com/dog\\_rates/status/671550332464455680/photo/1](https://twitter.com/dog_rates/status/671550332464455680/photo/1)  
 2149 [https://twitter.com/dog\\_rates/status/669684865554620416/photo/1](https://twitter.com/dog_rates/status/669684865554620416/photo/1)  
 2169 [https://twitter.com/dog\\_rates/status/669353438988365824/photo/1](https://twitter.com/dog_rates/status/669353438988365824/photo/1)  
 2189 [https://twitter.com/dog\\_rates/status/668967877119254528/photo/1](https://twitter.com/dog_rates/status/668967877119254528/photo/1)  
 2298 [https://twitter.com/dog\\_rates/status/667070482143944705/photo/1](https://twitter.com/dog_rates/status/667070482143944705/photo/1)

|     | rating_numerator | rating_denominator | name | dog_stage | retweet_count \ |
|-----|------------------|--------------------|------|-----------|-----------------|
| 30  | 12               | 10                 | NaN  |           | NaN             |
| 55  | 17               | 10                 | NaN  |           | NaN             |
| 64  | 14               | 10                 | NaN  |           | NaN             |
| 113 | 10               | 10                 | NaN  |           | NaN             |
| 148 | 12               | 10                 | NaN  |           | NaN             |

|      |     |     |     |                 |
|------|-----|-----|-----|-----------------|
| 149  | 14  | 10  | NaN | NaN             |
| 179  | 12  | 10  | NaN | NaN             |
| 184  | 14  | 10  | NaN | NaN             |
| 186  | 14  | 10  | NaN | NaN             |
| 188  | 420 | 10  | NaN | NaN             |
| 189  | 666 | 10  | NaN | NaN             |
| 218  | 13  | 10  | NaN | NaN             |
| 228  | 11  | 10  | NaN | NaN             |
| 234  | 13  | 10  | NaN | NaN             |
| 251  | 13  | 10  | NaN | NaN             |
| 274  | 10  | 10  | NaN | NaN             |
| 290  | 182 | 10  | NaN | NaN             |
| 291  | 15  | 10  | NaN | NaN             |
| 313  | 960 | 0   | NaN | NaN             |
| 342  | 11  | 15  | NaN | NaN             |
| 346  | 12  | 10  | NaN | NaN             |
| 387  | 7   | 10  | NaN | NaN             |
| 409  | 13  | 10  | NaN | NaN             |
| 427  | 13  | 10  | NaN | NaN             |
| 498  | 12  | 10  | NaN | NaN             |
| 513  | 11  | 10  | NaN | NaN             |
| 565  | 11  | 10  | NaN | doggopupper NaN |
| 570  | 11  | 10  | NaN | NaN             |
| 576  | 11  | 10  | NaN | doggo NaN       |
| 611  | 11  | 10  | NaN | NaN             |
| ...  | ..  | ..  | ... | ..              |
| 1479 | 11  | 10  | NaN | NaN             |
| 1497 | 9   | 10  | NaN | NaN             |
| 1501 | 13  | 10  | NaN | NaN             |
| 1523 | 12  | 10  | NaN | NaN             |
| 1598 | 4   | 20  | NaN | NaN             |
| 1605 | 14  | 10  | NaN | NaN             |
| 1618 | 5   | 10  | NaN | NaN             |
| 1630 | 12  | 10  | NaN | NaN             |
| 1634 | 143 | 130 | NaN | NaN             |
| 1663 | 20  | 16  | NaN | NaN             |
| 1689 | 5   | 10  | NaN | NaN             |
| 1774 | 13  | 10  | NaN | NaN             |
| 1819 | 7   | 10  | NaN | NaN             |
| 1842 | 11  | 10  | NaN | NaN             |
| 1844 | 9   | 10  | NaN | NaN             |
| 1852 | 11  | 10  | NaN | NaN             |
| 1866 | 13  | 10  | NaN | NaN             |
| 1882 | 13  | 10  | NaN | NaN             |
| 1885 | 13  | 10  | NaN | NaN             |
| 1892 | 12  | 10  | NaN | NaN             |
| 1895 | 11  | 10  | NaN | NaN             |
| 1905 | 13  | 10  | NaN | NaN             |

|      |    |    |       |     |
|------|----|----|-------|-----|
| 1914 | 13 | 10 | NaN   | NaN |
| 1940 | 1  | 10 | NaN   | NaN |
| 2036 | 13 | 10 | NaN   | NaN |
| 2038 | 1  | 10 | NaN   | NaN |
| 2149 | 11 | 10 | NaN   | NaN |
| 2169 | 10 | 10 | Tessa | NaN |
| 2189 | 12 | 10 | NaN   | NaN |
| 2298 | 10 | 10 | NaN   | NaN |

favorite\_count

|      |     |
|------|-----|
| 30   | NaN |
| 55   | NaN |
| 64   | NaN |
| 113  | NaN |
| 148  | NaN |
| 149  | NaN |
| 179  | NaN |
| 184  | NaN |
| 186  | NaN |
| 188  | NaN |
| 189  | NaN |
| 218  | NaN |
| 228  | NaN |
| 234  | NaN |
| 251  | NaN |
| 274  | NaN |
| 290  | NaN |
| 291  | NaN |
| 313  | NaN |
| 342  | NaN |
| 346  | NaN |
| 387  | NaN |
| 409  | NaN |
| 427  | NaN |
| 498  | NaN |
| 513  | NaN |
| 565  | NaN |
| 570  | NaN |
| 576  | NaN |
| 611  | NaN |
| ...  | ..  |
| 1479 | NaN |
| 1497 | NaN |
| 1501 | NaN |
| 1523 | NaN |
| 1598 | NaN |
| 1605 | NaN |
| 1618 | NaN |

```
1630 NaN
1634 NaN
1663 NaN
1689 NaN
1774 NaN
1819 NaN
1842 NaN
1844 NaN
1852 NaN
1866 NaN
1882 NaN
1885 NaN
1892 NaN
1895 NaN
1905 NaN
1914 NaN
1940 NaN
2036 NaN
2038 NaN
2149 NaN
2169 NaN
2189 NaN
2298 NaN
```

```
[78 rows x 16 columns]
```

```
In [64]: Archive_clean = Archive_clean[Archive_clean["in_reply_to_status_id"].isnull()]
        Archive_clean = Archive_clean[Archive_clean["retweeted_status_id"].isnull()]
```

```
In [65]: Archive_clean.drop(['in_reply_to_status_id', 'in_reply_to_user_id', 'retweeted_status_id'])
```

## Test

```
In [66]: Archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2097 entries, 0 to 2355
Data columns (total 11 columns):
tweet_id          2097 non-null int64
timestamp         2097 non-null object
source           2097 non-null object
text             2097 non-null object
expanded_urls     2097 non-null object
rating_numerator  2097 non-null int64
rating_denominator 2097 non-null int64
name             1494 non-null object
dog_stage        2097 non-null object
retweet_count     0 non-null float64
```

```

favorite_count      0 non-null float64
dtypes: float64(2), int64(3), object(6)
memory usage: 196.6+ KB

```

Now the table only has 2097 rows and in\_reply\_to\_status\_id,in\_reply\_to\_user\_id, retweeted\_status\_id,retweeted\_status\_user\_id, retweeted\_status\_timestamp are removed. tweet\_id column should be of object type, not integer.

### 7.0.1 Define

tweet\_id columns in two will be both changed into object type.

```

In [68]: Archive_clean['tweet_id'] = Archive_clean['tweet_id'].astype(str)
         imagePrediction_clean['tweet_id'] = imagePrediction_clean['tweet_id'].astype(str)

In [69]: Archive_clean['tweet_id'].dtypes

Out[69]: dtype('O')

```

The type of tweet id is changed successfully.

Next step will be source columns. There are 4 different kinds of source -- 'Twitter for iPhone','Vine - Make a Scene','Twitter Web Client',and 'TweetDeck'.

### 7.0.2 Define

Source column should be converted into 4 category of source -'Twitter for iPhone','Vine - Make a Scene','Twitter Web Client',and 'TweetDeck'.

```

In [70]: sourcemap = {'<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for i
                  '<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>' : 'Vine - Make a Scen
                  '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>' : 'Twitter Web Clie
                  '<a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>':'

# Function to apply
def convert_source(Archive_clean):
    if Archive_clean['source'] in sourcemap.keys():
        sourceab = sourcemap[Archive_clean['source']]
        return sourceab
    else:
        return Archive_clean['source']

Archive_clean['source'] = Archive_clean.apply(convert_source, axis=1)

```



## Test

```
In [71]: Archive_clean['source'].value_counts()
```

```
Out[71]: Twitter for iPhone      1964
         Vine - Make a Scene     91
         Twitter Web Client      31
         TweetDeck               11
         Name: source, dtype: int64
```

The source column is converted successfully, but the type of column is int64, which should be changed into category.

**Erroneous datatypes (timestamp, source columns) Erroneous format in value of prediction1-3 dog breed , prediction\_confidence1-3 have 6 decimals and Name is extracted from the word after 'This is', a lot of text does not follow this rule, some texts just have 'a', 'an', 'the'.**

### 7.0.3 Define

timestamp should be removed '+0000', year and month variable are created and convert into datetime. p1, p2, p3 dog breed should be all lower case and p1\_conf, p2\_conf, p3\_conf should be trimmed into 4 decimal. Name is very hard to clean, I will replace 'None', 'a', 'an', 'the' with nan value.

### 7.0.4 Code

```
In [72]: # To category
         Archive_clean['source'] = Archive_clean['source'].astype('category')

         # Strip '+0000'
         Archive_clean.timestamp = Archive_clean.timestamp.str.strip('+0000')

         # Extract year and month variable
         Archive_clean['month'] = Archive_clean['timestamp'].str.split('-').str.get(0) + Archive_clean['timestamp'].str.get(1)
         Archive_clean['month'] = Archive_clean['month'].astype(str)

         Archive_clean['year'] = Archive_clean['timestamp'].str.split('-').str.get(2)

         # To datetime
         Archive_clean.timestamp = pd.to_datetime(Archive_clean.timestamp)

         # To lower case
         imagePrediction_clean['prediction1'] = imagePrediction_clean['prediction1'].str.lower()
         imagePrediction_clean['prediction2'] = imagePrediction_clean['prediction2'].str.lower()
         imagePrediction_clean['prediction3'] = imagePrediction_clean['prediction3'].str.lower()

         # to 4 decimal
         imagePrediction_clean.prediction_confidence1 = imagePrediction_clean.prediction_confidence1.round(4)
         imagePrediction_clean.prediction_confidence2 = imagePrediction_clean.prediction_confidence2.round(4)
         imagePrediction_clean.prediction_confidence3 = imagePrediction_clean.prediction_confidence3.round(4)
```

```

# replace 'None', 'a', 'an', 'the' with nan
Archive_clean.name = Archive_clean.name.replace(['a', 'an', 'None', 'the'], np.nan)

#To string
Archive_clean['name'] = Archive_clean['name'].apply(str)

```

## 7.0.5 Test

```
In [73]: Archive_clean.timestamp.head()
```

```

Out[73]: 0    2017-08-01 16:23:56
         1    2017-08-01 00:17:27
         2    2017-07-31 00:18:03
         3    2017-07-30 15:58:51
         4    2017-07-29 16:00:24
         Name: timestamp, dtype: datetime64[ns]

```

```
In [74]: imagePrediction_clean[['prediction1', 'prediction2', 'prediction3']].head()
```

```

Out[74]:
           prediction1      prediction2      prediction3
0  welsh_springer_spaniel    collie      shetland_sheepdog
1      redbone      miniature_pinscher  rhodesian_ridgeback
2  german_shepherd      malinois      bloodhound
3  rhodesian_ridgeback      redbone      miniature_pinscher
4  miniature_pinscher    rottweiler      doberman

```

```
In [75]: imagePrediction_clean[['prediction_confidence1', 'prediction_confidence2', 'prediction_confidence3']].head()
```

```

Out[75]:
           prediction_confidence1  prediction_confidence2  prediction_confidence3
0      0.4651      0.1567      0.0614
1      0.5068      0.0742      0.0720
2      0.5965      0.1386      0.1162
3      0.4081      0.3607      0.2228
4      0.5603      0.2437      0.1546

```

```
In [76]: Archive_clean.name.value_counts()
```

```

Out[76]: nan          672
         Charlie      11
         Lucy        11
         Oliver      10
         Cooper      10
         Penny       9
         Tucker      9
         Lola        8
         Sadie       8
         Winston     8
         Toby        7

```

|            |   |
|------------|---|
| Daisy      | 7 |
| Stanley    | 6 |
| Bella      | 6 |
| Jax        | 6 |
| Bailey     | 6 |
| Bo         | 6 |
| Koda       | 6 |
| Oscar      | 6 |
| Milo       | 5 |
| Leo        | 5 |
| Buddy      | 5 |
| Scout      | 5 |
| Rusty      | 5 |
| Louis      | 5 |
| Dave       | 5 |
| Chester    | 5 |
| Bentley    | 5 |
| Phil       | 4 |
| Boomer     | 4 |
| ..         |   |
| Eve        | 1 |
| Willie     | 1 |
| Brian      | 1 |
| Carter     | 1 |
| Sephie     | 1 |
| Chef       | 1 |
| Grizzwald  | 1 |
| Perry      | 1 |
| Buckley    | 1 |
| Patch      | 1 |
| Tom        | 1 |
| Brudge     | 1 |
| Donny      | 1 |
| Tanner     | 1 |
| Tater      | 1 |
| Grizz      | 1 |
| Kendall    | 1 |
| Molly      | 1 |
| incredibly | 1 |
| Grey       | 1 |
| Butter     | 1 |
| Hanz       | 1 |
| Keet       | 1 |
| Corey      | 1 |
| Snicku     | 1 |
| Jimbo      | 1 |
| Tiger      | 1 |
| Scott      | 1 |

```
In [77]: text1 = Archive_clean[Archive_clean['rating_numerator'] == 5]['text']
        for x in text1:
            print(re.findall("\d+\.\d+", x))
```

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

68

```

        ['rating_numerator', 'rating_denominator']] = 14,10
Archive_clean.loc[Archive_clean['tweet_id'] == 740373189193256964,
        ['rating_numerator', 'rating_denominator']] = 14,10
Archive_clean.loc[Archive_clean['tweet_id'] == 682962037429899265,
        ['rating_numerator', 'rating_denominator']] = 10,10

```

```

In [79]: imagePrediction_clean.loc[imagePrediction_clean['prediction1'].isin(['newfoundland', '1
imagePrediction_clean.loc[imagePrediction_clean['prediction2'].isin(['newfoundland', '1
imagePrediction_clean.loc[imagePrediction_clean['prediction3'].isin(['newfoundland', '1

```

## Test

```

In [80]: Archive_clean[Archive_clean['tweet_id']== '883482846933004288'][['text', 'rating_numerat

```

Out[80]:

```

45 This is Bella. She hopes her smile made you smile. If not, she is also offering you

        rating_numerator  rating_denominator
45 13.5                  10

```

```

In [81]: Archive_clean['rating_denominator'].value_counts()

```

```

Out[81]: 10      2081
         50       3
         11       2
         80       2
          7       1
        170       1
        150       1
        120       1
        110       1
         90       1
         70       1
         40       1
         20       1
Name: rating_denominator, dtype: int64

```

```

In [82]: imagePrediction_clean[imagePrediction_clean['prediction3'] == 'mitten']

```

```

Out[82]:          tweet_id                                     jpg_url \
950  704859558691414016  https://pbs.twimg.com/media/CcgqBNVW8AE76lv.jpg

        most_confident_img_num prediction1 prediction_confidence1 \
950  1                          pug          0.2844

        dogbreed_indicator1 prediction2 prediction_confidence2 \
950  True                  teddy          0.1563

        dogbreed_indicator2 prediction3 prediction_confidence3 \

```

```
950 False          mitten          0.1389
```

```
dogbreed_indicator3
```

```
950 False
```

```
In [83]: imagePrediction_clean[imagePrediction_clean['prediction2'] == 'newfoundland']['dogbreed
```

```
Out[83]: 16      False
```

```
22      False
```

```
732     False
```

```
1012    False
```

```
1846    False
```

```
1859    False
```

```
1890    False
```

```
Name: dogbreed_indicator2, dtype: bool
```

```
In [84]: imagePrediction_clean[imagePrediction_clean['prediction3'] == 'newfoundland']['dogbreed
```

```
Out[84]: 114     False
```

```
491     False
```

```
1147    False
```

```
1540    False
```

```
1842    False
```

```
1868    False
```

```
1875    False
```

```
1994    False
```

```
2032    False
```

```
Name: dogbreed_indicator3, dtype: bool
```

```
In [85]: Archive_clean['rating'] =Archive_clean['rating_numerator']/Archive_clean['rating_denomi
```

## 7.0.6 Test

```
In [86]: Archive_clean[['rating', 'rating_numerator', 'rating_denominator']].head()
```

```
Out[86]:
```

|   | rating | rating_numerator | rating_denominator |
|---|--------|------------------|--------------------|
| 0 | 1.3    | 13.0             | 10                 |
| 1 | 1.3    | 13.0             | 10                 |
| 2 | 1.2    | 12.0             | 10                 |
| 3 | 1.3    | 13.0             | 10                 |
| 4 | 1.2    | 12.0             | 10                 |

Archive\_clean['rating'] column is successfully created.

## Store Files

```
In [87]: Archive_clean.to_csv('./archive_master.csv', index=False)
         imagePrediction_clean.to_csv('./imagePrediction_master.csv', index=False)
```

```
In [88]: twitter_archive_master = pd.merge(Archive_clean, imagePrediction_clean, on='tweet_id')
```

```
In [89]: Archive_clean.to_csv('./master.csv', index=False)
```