

# wrangle\_act

February 18, 2021

## Wrangle Analyze Data Project

### 0.0.1 Importing Needed Packages

```
In [1]: # importing useful packages
import pandas as pd
import numpy as np
import requests
import tweepy
import os
import json
import time
import re
import matplotlib.pyplot as plt
import warnings
import IPython
import re
import seaborn as sns
import datetime
import PIL
import glob

%matplotlib inline
```

### 0.1 Gathering Files And Creating Dataframes

```
In [2]: # Reading twitter archive file and assign a df as df_twitter_arch
df_twitter_arch = pd.read_csv('twitter-archive-enhanced.csv')
```

```
In [3]: # Downloading Image Predictions file
url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions-2017-08-01/image-predictions-2017-08-01.ts'
response = requests.get(url)
with open(os.path.join(url.split('/')[-1]), 'wb') as file:
    file.write(response.content)
```

```
In [4]: # Reading Image Predictions file and assign a df as df_img_pred
df_img_pred = pd.read_csv('image-predictions.tsv', sep = '\t')
```

```
In [5]: # As I can not get my Twitter developer credentials till now, That's why I can not get a
# Opening tweet-json file and assign a df as df_api
with open('tweet-json.txt') as file:
    df_api = pd.read_json(file, lines=True, encoding = 'utf-8')
```

## 0.2 Assessing dataframes.

### 0.2.1 Visual Assessment

```
In [6]: # Display df_twitter_arch
df_twitter_arch
```

```
Out[6]:
```

|      | 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...



2349 <a href="http://twitter.com/download/iphone" r...  
 2350 <a href="http://twitter.com/download/iphone" r...  
 2351 <a href="http://twitter.com/download/iphone" r...  
 2352 <a href="http://twitter.com/download/iphone" r...  
 2353 <a href="http://twitter.com/download/iphone" r...  
 2354 <a href="http://twitter.com/download/iphone" r...  
 2355 <a href="http://twitter.com/download/iphone" r...

|      | text  | retweeted_status_id \ |
|------|---|-----------------------|
| 0    | This is Phineas. He's a mystical boy. Only eve... | NaN                   |
| 1    | This is Tilly. She's just checking pup on you...  | NaN                   |
| 2    | This is Archie. He is a rare Norwegian Pouncin... | NaN                   |
| 3    | This is Darla. She commenced a snooze mid meal... | NaN                   |
| 4    | This is Franklin. He would like you to stop ca... | NaN                   |
| 5    | Here we have a majestic great white breaching ... | NaN                   |
| 6    | Meet Jax. He enjoys ice cream so much he gets ... | NaN                   |
| 7    | When you watch your owner call another dog a g... | NaN                   |
| 8    | This is Zoey. She doesn't want to be one of th... | NaN                   |
| 9    | This is Cassie. She is a college pup. Studying... | NaN                   |
| 10   | This is Koda. He is a South Australian decksha... | NaN                   |
| 11   | This is Bruno. He is a service shark. Only get... | NaN                   |
| 12   | Here's a puppo that seems to be on the fence a... | NaN                   |
| 13   | This is Ted. He does his best. Sometimes that'... | NaN                   |
| 14   | This is Stuart. He's sporting his favorite fan... | NaN                   |
| 15   | This is Oliver. You're witnessing one of his m... | NaN                   |
| 16   | This is Jim. He found a fren. Taught him how t... | NaN                   |
| 17   | This is Zeke. He has a new stick. Very proud o... | NaN                   |
| 18   | This is Ralphus. He's powering up. Attempting ... | NaN                   |
| 19   | RT @dog_rates: This is Canela. She attempted s... | 8.874740e+17          |
| 20   | This is Gerald. He was just told he didn't get... | NaN                   |
| 21   | This is Jeffrey. He has a monopoly on the pool... | NaN                   |
| 22   | I've yet to rate a Venezuelan Hover Wiener. Th... | NaN                   |
| 23   | This is Canela. She attempted some fancy porch... | NaN                   |
| 24   | You may not have known you needed to see this ... | NaN                   |
| 25   | This... is a Jubilant Antarctic House Bear. We... | NaN                   |
| 26   | This is Maya. She's very shy. Rarely leaves he... | NaN                   |
| 27   | This is Mingus. He's a wonderful father to his... | NaN                   |
| 28   | This is Derek. He's late for a dog meeting. 13... | NaN                   |
| 29   | This is Roscoe. Another pupper fallen victim t... | NaN                   |
| ...  | ...   | ...                   |
| 2326 | This is quite the dog. Gets really excited whe... | NaN                   |
| 2327 | This is a southern Vesuvius bumblegruff. Can d... | NaN                   |
| 2328 | Oh goodness. A super rare northeast Qdoba kang... | NaN                   |
| 2329 | Those are sunglasses and a jean jacket. 11/10 ... | NaN                   |
| 2330 | Unique dog here. Very small. Lives in containe... | NaN                   |
| 2331 | Here we have a mixed Asiago from the Galápagos... | NaN                   |
| 2332 | Look at this jokester thinking seat belt laws ... | NaN                   |
| 2333 | This is an extremely rare horned Parthenon. No... | NaN                   |

|      |  |     |
|------|--|-----|
| 2334 | This is a funny dog. Weird toes. Won't come do...                                    | NaN |
| 2335 | This is an Albanian 3 1/2 legged Episcopalian...                                     | NaN |
| 2336 | Can take selfies 11/10 <a href="https://t.co/ws2AMaWwPW">https://t.co/ws2AMaWwPW</a> | NaN |
| 2337 | Very concerned about fellow dog trapped in com...                                    | NaN |
| 2338 | Not familiar with this breed. No tail (weird)...                                     | NaN |
| 2339 | Oh my. Here you are seeing an Adobe Setter giv...                                    | NaN |
| 2340 | Can stand on stump for what seems like a while...                                    | NaN |
| 2341 | This appears to be a Mongolian Presbyterian mi...                                    | NaN |
| 2342 | Here we have a well-established sunblockerspan...                                    | NaN |
| 2343 | Let's hope this flight isn't Malaysian (lol). ...                                    | NaN |
| 2344 | Here we have a northern speckled Rhododendron...                                     | NaN |
| 2345 | This is the happiest dog you will ever see. Ve...                                    | NaN |
| 2346 | Here is the Rand Paul of retrievers folks! He'...                                    | NaN |
| 2347 | My oh my. This is a rare blond Canadian terrie...                                    | NaN |
| 2348 | Here is a Siberian heavily armored polar bear ...                                    | NaN |
| 2349 | This is an odd dog. Hard on the outside but lo...                                    | NaN |
| 2350 | This is a truly beautiful English Wilson Staff...                                    | NaN |
| 2351 | Here we have a 1949 1st generation vulpix. Enj...                                    | NaN |
| 2352 | This is a purebred Piers Morgan. Loves to Netf...                                    | NaN |
| 2353 | Here is a very happy pup. Big fan of well-main...                                    | NaN |
| 2354 | This is a western brown Mitsubishi terrier. Up...                                    | NaN |
| 2355 | Here we have a Japanese Irish Setter. Lost eye...                                    | NaN |

|    | retweeted_status_user_id | retweeted_status_timestamp | \ |
|----|--------------------------|----------------------------|---|
| 0  | NaN                      | NaN                        |   |
| 1  | NaN                      | NaN                        |   |
| 2  | NaN                      | NaN                        |   |
| 3  | NaN                      | NaN                        |   |
| 4  | NaN                      | NaN                        |   |
| 5  | NaN                      | NaN                        |   |
| 6  | NaN                      | NaN                        |   |
| 7  | NaN                      | NaN                        |   |
| 8  | NaN                      | NaN                        |   |
| 9  | NaN                      | NaN                        |   |
| 10 | NaN                      | NaN                        |   |
| 11 | NaN                      | NaN                        |   |
| 12 | NaN                      | NaN                        |   |
| 13 | NaN                      | NaN                        |   |
| 14 | NaN                      | NaN                        |   |
| 15 | NaN                      | NaN                        |   |
| 16 | NaN                      | NaN                        |   |
| 17 | NaN                      | NaN                        |   |
| 18 | NaN                      | NaN                        |   |
| 19 | 4.196984e+09             | 2017-07-19 00:47:34 +0000  |   |
| 20 | NaN                      | NaN                        |   |
| 21 | NaN                      | NaN                        |   |
| 22 | NaN                      | NaN                        |   |
| 23 | NaN                      | NaN                        |   |

|      |     |     |
|------|-----|-----|
| 24   | NaN | NaN |
| 25   | NaN | NaN |
| 26   | NaN | NaN |
| 27   | NaN | NaN |
| 28   | NaN | NaN |
| 29   | NaN | NaN |
| ...  | ... | ... |
| 2326 | NaN | NaN |
| 2327 | NaN | NaN |
| 2328 | NaN | NaN |
| 2329 | NaN | NaN |
| 2330 | NaN | NaN |
| 2331 | NaN | NaN |
| 2332 | NaN | NaN |
| 2333 | NaN | NaN |
| 2334 | NaN | NaN |
| 2335 | NaN | NaN |
| 2336 | NaN | NaN |
| 2337 | NaN | NaN |
| 2338 | NaN | NaN |
| 2339 | NaN | NaN |
| 2340 | NaN | NaN |
| 2341 | NaN | NaN |
| 2342 | NaN | NaN |
| 2343 | NaN | NaN |
| 2344 | NaN | NaN |
| 2345 | NaN | NaN |
| 2346 | NaN | NaN |
| 2347 | NaN | NaN |
| 2348 | NaN | NaN |
| 2349 | NaN | NaN |
| 2350 | NaN | NaN |
| 2351 | NaN | NaN |
| 2352 | NaN | NaN |
| 2353 | NaN | NaN |
| 2354 | NaN | NaN |
| 2355 | NaN | NaN |

|   | expanded_urls   | rating_numerator \ |
|---|---|--------------------|
| 0 | <a href="https://twitter.com/dog_rates/status/892420643...">https://twitter.com/dog_rates/status/892420643...</a> | 13                 |
| 1 | <a href="https://twitter.com/dog_rates/status/892177421...">https://twitter.com/dog_rates/status/892177421...</a> | 13                 |
| 2 | <a href="https://twitter.com/dog_rates/status/891815181...">https://twitter.com/dog_rates/status/891815181...</a> | 12                 |
| 3 | <a href="https://twitter.com/dog_rates/status/891689557...">https://twitter.com/dog_rates/status/891689557...</a> | 13                 |
| 4 | <a href="https://twitter.com/dog_rates/status/891327558...">https://twitter.com/dog_rates/status/891327558...</a> | 12                 |
| 5 | <a href="https://twitter.com/dog_rates/status/891087950...">https://twitter.com/dog_rates/status/891087950...</a> | 13                 |
| 6 | <a href="https://gofundme.com/ydvmve-surgery-for-jax,ht...">https://gofundme.com/ydvmve-surgery-for-jax,ht...</a> | 13                 |
| 7 | <a href="https://twitter.com/dog_rates/status/890729181...">https://twitter.com/dog_rates/status/890729181...</a> | 13                 |
| 8 | <a href="https://twitter.com/dog_rates/status/890609185...">https://twitter.com/dog_rates/status/890609185...</a> | 13                 |



|      |   |     |
|------|---|-----|
| 9    | <a href="https://twitter.com/dog_rates/status/890240255...">https://twitter.com/dog_rates/status/890240255...</a>                   | 14  |
| 10   | <a href="https://twitter.com/dog_rates/status/890006608...">https://twitter.com/dog_rates/status/890006608...</a>                   | 13  |
| 11   | <a href="https://twitter.com/dog_rates/status/889880896...">https://twitter.com/dog_rates/status/889880896...</a>                   | 13  |
| 12   | <a href="https://twitter.com/dog_rates/status/889665388...">https://twitter.com/dog_rates/status/889665388...</a>                   | 13  |
| 13   | <a href="https://twitter.com/dog_rates/status/889638837...">https://twitter.com/dog_rates/status/889638837...</a>                   | 12  |
| 14   | <a href="https://twitter.com/dog_rates/status/889531135...">https://twitter.com/dog_rates/status/889531135...</a>                   | 13  |
| 15   | <a href="https://twitter.com/dog_rates/status/889278841...">https://twitter.com/dog_rates/status/889278841...</a>                   | 13  |
| 16   | <a href="https://twitter.com/dog_rates/status/888917238...">https://twitter.com/dog_rates/status/888917238...</a>                   | 12  |
| 17   | <a href="https://twitter.com/dog_rates/status/888804989...">https://twitter.com/dog_rates/status/888804989...</a>                   | 13  |
| 18   | <a href="https://twitter.com/dog_rates/status/888554962...">https://twitter.com/dog_rates/status/888554962...</a>                   | 13  |
| 19   | <a href="https://twitter.com/dog_rates/status/887473957...">https://twitter.com/dog_rates/status/887473957...</a>                   | 13  |
| 20   | <a href="https://twitter.com/dog_rates/status/888078434...">https://twitter.com/dog_rates/status/888078434...</a>                   | 12  |
| 21   | <a href="https://twitter.com/dog_rates/status/887705289...">https://twitter.com/dog_rates/status/887705289...</a>                   | 13  |
| 22   | <a href="https://twitter.com/dog_rates/status/887517139...">https://twitter.com/dog_rates/status/887517139...</a>                   | 14  |
| 23   | <a href="https://twitter.com/dog_rates/status/887473957...">https://twitter.com/dog_rates/status/887473957...</a>                   | 13  |
| 24   | <a href="https://twitter.com/dog_rates/status/887343217...">https://twitter.com/dog_rates/status/887343217...</a>                   | 13  |
| 25   | <a href="https://twitter.com/dog_rates/status/887101392...">https://twitter.com/dog_rates/status/887101392...</a>                   | 12  |
| 26   | <a href="https://twitter.com/dog_rates/status/886983233...">https://twitter.com/dog_rates/status/886983233...</a>                   | 13  |
| 27   | <a href="https://www.gofundme.com/mingusneedsus">https://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 |
| 2355 | <a href="https://twitter.com/dog_rates/status/666020888...">https://twitter.com/dog_rates/status/666020888...</a> | 8 |

|      | rating_denominator | name     | doggo | floofer | pupper | puppo |
|------|--------------------|----------|-------|---------|--------|-------|
| 0    | 10                 | Phineas  | None  | None    | None   | None  |
| 1    | 10                 | Tilly    | None  | None    | None   | None  |
| 2    | 10                 | Archie   | None  | None    | None   | None  |
| 3    | 10                 | Darla    | None  | None    | None   | None  |
| 4    | 10                 | Franklin | None  | None    | None   | None  |
| 5    | 10                 | None     | None  | None    | None   | None  |
| 6    | 10                 | Jax      | None  | None    | None   | None  |
| 7    | 10                 | None     | None  | None    | None   | None  |
| 8    | 10                 | Zoey     | None  | None    | None   | None  |
| 9    | 10                 | Cassie   | doggo | None    | None   | None  |
| 10   | 10                 | Koda     | None  | None    | None   | None  |
| 11   | 10                 | Bruno    | None  | None    | None   | None  |
| 12   | 10                 | None     | None  | None    | None   | puppo |
| 13   | 10                 | Ted      | None  | None    | None   | None  |
| 14   | 10                 | Stuart   | None  | None    | None   | puppo |
| 15   | 10                 | Oliver   | None  | None    | None   | None  |
| 16   | 10                 | Jim      | None  | None    | None   | None  |
| 17   | 10                 | Zeke     | None  | None    | None   | None  |
| 18   | 10                 | Ralphus  | None  | None    | None   | None  |
| 19   | 10                 | Canela   | None  | None    | None   | None  |
| 20   | 10                 | Gerald   | None  | None    | None   | None  |
| 21   | 10                 | Jeffrey  | None  | None    | None   | None  |
| 22   | 10                 | such     | None  | None    | None   | None  |
| 23   | 10                 | Canela   | None  | None    | None   | None  |
| 24   | 10                 | None     | None  | None    | None   | None  |
| 25   | 10                 | None     | None  | None    | None   | None  |
| 26   | 10                 | Maya     | None  | None    | None   | None  |
| 27   | 10                 | Mingus   | None  | None    | None   | None  |
| 28   | 10                 | Derek    | None  | None    | None   | None  |
| 29   | 10                 | Roscoe   | None  | None    | pupper | None  |
| ...  | ...                | ...      | ...   | ...     | ...    | ...   |
| 2326 | 10                 | quite    | None  | None    | None   | None  |
| 2327 | 10                 | a        | None  | None    | None   | None  |
| 2328 | 10                 | None     | None  | None    | None   | None  |
| 2329 | 10                 | None     | None  | None    | None   | None  |
| 2330 | 10                 | None     | None  | None    | None   | None  |
| 2331 | 10                 | None     | None  | None    | None   | None  |
| 2332 | 10                 | None     | None  | None    | None   | None  |
| 2333 | 10                 | an       | None  | None    | None   | None  |
| 2334 | 10                 | a        | None  | None    | None   | None  |
| 2335 | 2                  | an       | None  | None    | None   | None  |
| 2336 | 10                 | None     | None  | None    | None   | None  |

|      |    |      |      |      |      |      |
|------|----|------|------|------|------|------|
| 2337 | 10 | None | None | None | None | None |
| 2338 | 10 | None | None | None | None | None |
| 2339 | 10 | None | None | None | None | None |
| 2340 | 10 | None | None | None | None | None |
| 2341 | 10 | None | None | None | None | None |
| 2342 | 10 | None | None | None | None | None |
| 2343 | 10 | None | None | None | None | None |
| 2344 | 10 | None | None | None | None | None |
| 2345 | 10 | the  | None | None | None | None |
| 2346 | 10 | the  | None | None | None | None |
| 2347 | 10 | a    | None | None | None | None |
| 2348 | 10 | a    | None | None | None | None |
| 2349 | 10 | an   | None | None | None | None |
| 2350 | 10 | a    | None | None | None | None |
| 2351 | 10 | None | None | None | None | None |
| 2352 | 10 | a    | None | None | None | None |
| 2353 | 10 | a    | None | None | None | None |
| 2354 | 10 | a    | None | None | None | None |
| 2355 | 10 | None | None | None | None | None |

[2356 rows x 17 columns]

In [7]: # Display df\_img\_pred  
df\_img\_pred

| Out [7]: | tweet_id           | jpg_url \   |
|----------|--------------------|---|
| 0        | 666020888022790149 | <a href="https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg">https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg</a>   |
| 1        | 666029285002620928 | <a href="https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg">https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg</a>   |
| 2        | 666033412701032449 | <a href="https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg">https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg</a>   |
| 3        | 666044226329800704 | <a href="https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg">https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg</a>   |
| 4        | 666049248165822465 | <a href="https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg">https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg</a>   |
| 5        | 666050758794694657 | <a href="https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg">https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg</a>   |
| 6        | 666051853826850816 | <a href="https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg">https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg</a>   |
| 7        | 666055525042405380 | <a href="https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg">https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg</a>   |
| 8        | 666057090499244032 | <a href="https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg">https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg</a>   |
| 9        | 666058600524156928 | <a href="https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg">https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg</a>   |
| 10       | 666063827256086533 | <a href="https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg">https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg</a>   |
| 11       | 666071193221509120 | <a href="https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg">https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg</a>   |
| 12       | 666073100786774016 | <a href="https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg">https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg</a>   |
| 13       | 666082916733198337 | <a href="https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg">https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg</a> |
| 14       | 666094000022159362 | <a href="https://pbs.twimg.com/media/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/CT56LSZW0AA1Jj2.jpg">https://pbs.twimg.com/media/CT56LSZW0AA1Jj2.jpg</a>   |
| 18       | 666268910803644416 | <a href="https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg">https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg</a>   |
| 19       | 666273097616637952 | <a href="https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg">https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg</a>   |
| 20       | 666287406224695296 | <a href="https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg">https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg</a>   |
| 21       | 666293911632134144 | <a href="https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg">https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg</a>   |

|      |                    |   |
|------|--------------------|---|
| 22   | 666337882303524864 | <a href="https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg">https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg</a>     |
| 23   | 666345417576210432 | <a href="https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg">https://pbs.twimg.com/media/CT9Vn7PW0AA_ZCM.jpg</a>     |
| 24   | 666353288456101888 | <a href="https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg">https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg</a>     |
| 25   | 666362758909284353 | <a href="https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg">https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg</a>     |
| 26   | 666373753744588802 | <a href="https://pbs.twimg.com/media/CT9vZEYWUAAI205.jpg">https://pbs.twimg.com/media/CT9vZEYWUAAI205.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-NvwmW4AAAugGZ.jpg">https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.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/DE5Se8FXcAAJF4.jpg">https://pbs.twimg.com/media/DE5Se8FXcAAJF4.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/DF1eOmZXUAAALUcq.jpg">https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg</a>   |
| 2069 | 891087950875897856 | <a href="https://pbs.twimg.com/media/DF3HwyEWsAAABqE6.jpg">https://pbs.twimg.com/media/DF3HwyEWsAAABqE6.jpg</a>   |
| 2070 | 891327558926688256 | <a href="https://pbs.twimg.com/media/DF6hr6BUMAAZgT.jpg">https://pbs.twimg.com/media/DF6hr6BUMAAZgT.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  | terrapien                   |
| 7    | Tibetan_mastiff          |        | 0.058279 | True   | fur_coat                    |
| 8    | shopping_basket          |        | 0.014594 | False  | golden_retriever            |
| 9    | komondor                 |        | 0.192305 | True   | soft-coated_wheaten_terrier |
| 10   | Tibetan_mastiff          |        | 0.093718 | True   | Labrador_retriever          |
| 11   | Yorkshire_terrier        |        | 0.174201 | True   | Pekinese                    |
| 12   | English_foxhound         |        | 0.175382 | True   | Ibizan_hound                |
| 13   | bull_mastiff             |        | 0.404722 | True   | French_bulldog              |
| 14   | German_shepherd          |        | 0.078260 | True   | malinois                    |
| 15   | Shih-Tzu                 |        | 0.166192 | True   | Dandie_Dinmont              |
| 16   | Newfoundland             |        | 0.149842 | True   | borzoi                      |
| 17   | cock                     |        | 0.033919 | False  | partridge                   |
| 18   | desk                     |        | 0.085547 | False  | bookcase                    |
| 19   | toy_terrier              |        | 0.111884 | True   | basenji                     |
| 20   | toy_poodle               |        | 0.063064 | True   | miniature_poodle            |
| 21   | otter                    |        | 0.015250 | False  | great_grey_owl              |
| 22   | Newfoundland             |        | 0.278407 | True   | groenendael                 |
| 23   | Chesapeake_Bay_retriever |        | 0.054787 | True   | Labrador_retriever          |
| 24   | Siberian_husky           |        | 0.147655 | True   | Eskimo_dog                  |
| 25   | skunk                    |        | 0.002402 | False  | hamster                     |
| 26   | Afghan_hound             |        | 0.259551 | True   | briard                      |
| 27   | toy_terrier              |        | 0.009397 | True   | papillon                    |
| 28   | bloodhound               |        | 0.244220 | True   | flat-coated_retriever       |
| 29   | barracouta               |        | 0.271485 | False  | gar                         |
| ...  | ...                      |        | ...      | ...    | ...                         |
| 2045 | Chihuahua                |        | 0.000361 | True   | Boston_bull                 |
| 2046 | sports_car               |        | 0.139952 | False  | car_wheel                   |
| 2047 | Great_Pyrenees           |        | 0.186136 | True   | Dandie_Dinmont              |
| 2048 | toy_terrier              |        | 0.143528 | True   | can_opener                  |
| 2049 | Eskimo_dog               |        | 0.035029 | True   | Staffordshire_bullterrier   |
| 2050 | sea_lion                 |        | 0.275645 | False  | Weimaraner                  |
| 2051 | Rhodesian_ridgeback      |        | 0.054950 | True   | beagle                      |
| 2052 | tow_truck                |        | 0.029175 | False  | shopping_cart               |
| 2053 | redbone                  |        | 0.087582 | True   | Weimaraner                  |

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

|    | p3_conf  | p3_dog |
|----|----------|--------|
| 0  | 0.061428 | True   |
| 1  | 0.072010 | True   |
| 2  | 0.116197 | True   |
| 3  | 0.222752 | True   |
| 4  | 0.154629 | True   |
| 5  | 0.016199 | True   |
| 6  | 0.017885 | False  |
| 7  | 0.054449 | False  |
| 8  | 0.007959 | True   |
| 9  | 0.082086 | True   |
| 10 | 0.072427 | True   |
| 11 | 0.109454 | True   |
| 12 | 0.097471 | True   |
| 13 | 0.048960 | True   |
| 14 | 0.075628 | True   |
| 15 | 0.089688 | True   |
| 16 | 0.133649 | True   |
| 17 | 0.000052 | False  |
| 18 | 0.079480 | False  |
| 19 | 0.111152 | True   |
| 20 | 0.025581 | True   |
| 21 | 0.013207 | False  |
| 22 | 0.102643 | True   |
| 23 | 0.014241 | True   |
| 24 | 0.093412 | True   |

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

[2075 rows x 12 columns]

```
In [8]: # Display df_api
df_api.head()
```

```
Out[8]:
```

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



```

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

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

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

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

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

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

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

```

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

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

```
[5 rows x 31 columns]
```

## 0.2.2 Programmatic Assessment

Programmatic Assessment for df\_twitter\_arch

```
In [9]: # Displaying df_twitter_arch information
df_twitter_arch.info()
```

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

```
In [10]: # Cheacking if there any duplicated tweet_id
df_twitter_arch.tweet_id.duplicated().sum()
```

```
Out[10]: 0
```

```
In [11]: # Checking if there invalid denominator lager than normal = 10
df_twitter_arch.rating_denominator.value_counts()
```

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

```
In [12]: # Checking if there invalid denominator and numerator and the max number by .describe()
df_twitter_arch[['rating_numerator', 'rating_denominator']].describe()
```

```
Out[12]:
```

|       | rating_numerator | rating_denominator |
|-------|------------------|--------------------|
| count | 2356.000000      | 2356.000000        |
| mean  | 13.126486        | 10.455433          |
| std   | 45.876648        | 6.745237           |
| min   | 0.000000         | 0.000000           |
| 25%   | 10.000000        | 10.000000          |
| 50%   | 11.000000        | 10.000000          |
| 75%   | 12.000000        | 10.000000          |
| max   | 1776.000000      | 170.000000         |

```
In [13]: # Checking if there invalid names by istitle() and unique().
df_twitter_arch[~df_twitter_arch.name.str.istitle()].name.unique()
```

```
Out[13]: array(['such', 'a', 'quite', 'not', 'one', 'incredibly', 'BeBe', 'mad',
               'an', 'very', 'just', 'DonDon', 'my', 'his', 'actually', 'getting',
               'this', 'unacceptable', 'all', 'old', 'infuriating', 'CeCe', 'the',
               'by', 'officially', 'life', 'light', 'space', 'JD', 'DayZ'], dtype=object)
```

```
In [14]: # Checking if there retweets in text column should be removed.
# Creating a variable to search for text starts with "RT" as a retweet text.
search = 'RT'
# Searching for texts starts with "RT" in text column.
df_twitter_arch[df_twitter_arch['text'].str.startswith(search)]
```

```
Out[14]:
```

|    | tweet_id           | in_reply_to_status_id | in_reply_to_user_id \ |
|----|--------------------|-----------------------|-----------------------|
| 19 | 888202515573088257 | NaN                   | NaN                   |

|     |                    |     |     |
|-----|--------------------|-----|-----|
| 32  | 886054160059072513 | NaN | NaN |
| 36  | 885311592912609280 | NaN | NaN |
| 68  | 879130579576475649 | NaN | NaN |
| 73  | 878404777348136964 | NaN | NaN |
| 74  | 878316110768087041 | NaN | NaN |
| 78  | 877611172832227328 | NaN | NaN |
| 91  | 874434818259525634 | NaN | NaN |
| 95  | 873697596434513921 | NaN | NaN |
| 97  | 873337748698140672 | NaN | NaN |
| 101 | 872668790621863937 | NaN | NaN |
| 109 | 871166179821445120 | NaN | NaN |
| 118 | 869988702071779329 | NaN | NaN |
| 124 | 868639477480148993 | NaN | NaN |
| 130 | 867072653475098625 | NaN | NaN |
| 132 | 866816280283807744 | NaN | NaN |
| 137 | 866094527597207552 | NaN | NaN |
| 146 | 863471782782697472 | NaN | NaN |
| 155 | 861769973181624320 | NaN | NaN |
| 159 | 860981674716409858 | NaN | NaN |
| 160 | 860924035999428608 | NaN | NaN |
| 165 | 860177593139703809 | NaN | NaN |
| 171 | 858860390427611136 | NaN | NaN |
| 180 | 857062103051644929 | NaN | NaN |
| 182 | 856602993587888130 | NaN | NaN |
| 185 | 856330835276025856 | NaN | NaN |
| 194 | 855245323840757760 | NaN | NaN |
| 195 | 855138241867124737 | NaN | NaN |
| 204 | 852936405516943360 | NaN | NaN |
| 211 | 851953902622658560 | NaN | NaN |
| ... | ...                | ... | ... |
| 800 | 772615324260794368 | NaN | NaN |
| 811 | 771171053431250945 | NaN | NaN |
| 815 | 771004394259247104 | NaN | NaN |
| 818 | 770743923962707968 | NaN | NaN |
| 822 | 770093767776997377 | NaN | NaN |
| 826 | 769335591808995329 | NaN | NaN |
| 829 | 768909767477751808 | NaN | NaN |
| 833 | 768554158521745409 | NaN | NaN |
| 841 | 766864461642756096 | NaN | NaN |
| 847 | 766078092750233600 | NaN | NaN |
| 860 | 763167063695355904 | NaN | NaN |
| 868 | 761750502866649088 | NaN | NaN |
| 872 | 761371037149827077 | NaN | NaN |
| 885 | 760153949710192640 | NaN | NaN |
| 890 | 759566828574212096 | NaN | NaN |
| 895 | 759159934323924993 | NaN | NaN |
| 908 | 757729163776290825 | NaN | NaN |
| 911 | 757597904299253760 | NaN | NaN |

|      |                    |     |     |
|------|--------------------|-----|-----|
| 926  | 754874841593970688 | NaN | NaN |
| 937  | 753298634498793472 | NaN | NaN |
| 943  | 752701944171524096 | NaN | NaN |
| 949  | 752309394570878976 | NaN | NaN |
| 1012 | 747242308580548608 | NaN | NaN |
| 1023 | 746521445350707200 | NaN | NaN |
| 1043 | 743835915802583040 | NaN | NaN |
| 1242 | 711998809858043904 | NaN | NaN |
| 1286 | 708400866336894977 | NaN | NaN |
| 1860 | 675489971617296384 | NaN | NaN |
| 2259 | 667550904950915073 | NaN | NaN |
| 2260 | 667550882905632768 | NaN | NaN |

|     | timestamp \               |
|-----|---------------------------|
| 19  | 2017-07-21 01:02:36 +0000 |
| 32  | 2017-07-15 02:45:48 +0000 |
| 36  | 2017-07-13 01:35:06 +0000 |
| 68  | 2017-06-26 00:13:58 +0000 |
| 73  | 2017-06-24 00:09:53 +0000 |
| 74  | 2017-06-23 18:17:33 +0000 |
| 78  | 2017-06-21 19:36:23 +0000 |
| 91  | 2017-06-13 01:14:41 +0000 |
| 95  | 2017-06-11 00:25:14 +0000 |
| 97  | 2017-06-10 00:35:19 +0000 |
| 101 | 2017-06-08 04:17:07 +0000 |
| 109 | 2017-06-04 00:46:17 +0000 |
| 118 | 2017-05-31 18:47:24 +0000 |
| 124 | 2017-05-28 01:26:04 +0000 |
| 130 | 2017-05-23 17:40:04 +0000 |
| 132 | 2017-05-23 00:41:20 +0000 |
| 137 | 2017-05-21 00:53:21 +0000 |
| 146 | 2017-05-13 19:11:30 +0000 |
| 155 | 2017-05-09 02:29:07 +0000 |
| 159 | 2017-05-06 22:16:42 +0000 |
| 160 | 2017-05-06 18:27:40 +0000 |
| 165 | 2017-05-04 17:01:34 +0000 |
| 171 | 2017-05-01 01:47:28 +0000 |
| 180 | 2017-04-26 02:41:43 +0000 |
| 182 | 2017-04-24 20:17:23 +0000 |
| 185 | 2017-04-24 02:15:55 +0000 |
| 194 | 2017-04-21 02:22:29 +0000 |
| 195 | 2017-04-20 19:16:59 +0000 |
| 204 | 2017-04-14 17:27:40 +0000 |
| 211 | 2017-04-12 00:23:33 +0000 |
| ... | ...                       |
| 800 | 2016-09-05 02:00:22 +0000 |
| 811 | 2016-09-01 02:21:21 +0000 |
| 815 | 2016-08-31 15:19:06 +0000 |

818 2016-08-30 22:04:05 +0000  
 822 2016-08-29 03:00:36 +0000  
 826 2016-08-27 00:47:53 +0000  
 829 2016-08-25 20:35:48 +0000  
 833 2016-08-24 21:02:45 +0000  
 841 2016-08-20 05:08:29 +0000  
 847 2016-08-18 01:03:45 +0000  
 860 2016-08-10 00:16:21 +0000  
 868 2016-08-06 02:27:27 +0000  
 872 2016-08-05 01:19:35 +0000  
 885 2016-08-01 16:43:19 +0000  
 890 2016-07-31 01:50:18 +0000  
 895 2016-07-29 22:53:27 +0000  
 908 2016-07-26 00:08:05 +0000  
 911 2016-07-25 15:26:30 +0000  
 926 2016-07-18 03:06:01 +0000  
 937 2016-07-13 18:42:44 +0000  
 943 2016-07-12 03:11:42 +0000  
 949 2016-07-11 01:11:51 +0000  
 1012 2016-06-27 01:37:04 +0000  
 1023 2016-06-25 01:52:36 +0000  
 1043 2016-06-17 16:01:16 +0000  
 1242 2016-03-21 19:31:59 +0000  
 1286 2016-03-11 21:15:02 +0000  
 1860 2015-12-12 01:38:53 +0000  
 2259 2015-11-20 03:51:52 +0000  
 2260 2015-11-20 03:51:47 +0000

source \

19 <a href="http://twitter.com/download/iphone" r...  
 32 <a href="http://twitter.com/download/iphone" r...  
 36 <a href="http://twitter.com/download/iphone" r...  
 68 <a href="http://twitter.com/download/iphone" r...  
 73 <a href="http://twitter.com/download/iphone" r...  
 74 <a href="http://twitter.com/download/iphone" r...  
 78 <a href="http://twitter.com/download/iphone" r...  
 91 <a href="http://twitter.com/download/iphone" r...  
 95 <a href="http://twitter.com/download/iphone" r...  
 97 <a href="http://twitter.com/download/iphone" r...  
 101 <a href="http://twitter.com/download/iphone" r...  
 109 <a href="http://twitter.com/download/iphone" r...  
 118 <a href="http://twitter.com/download/iphone" r...  
 124 <a href="http://twitter.com/download/iphone" r...  
 130 <a href="http://twitter.com/download/iphone" r...  
 132 <a href="http://twitter.com/download/iphone" r...  
 137 <a href="http://twitter.com/download/iphone" r...  
 146 <a href="http://twitter.com/download/iphone" r...  
 155 <a href="http://twitter.com/download/iphone" r...

159 <a href="http://twitter.com/download/iphone" r...  
 160 <a href="http://twitter.com/download/iphone" r...  
 165 <a href="http://twitter.com/download/iphone" r...  
 171 <a href="http://twitter.com/download/iphone" r...  
 180 <a href="http://twitter.com/download/iphone" r...  
 182 <a href="http://twitter.com/download/iphone" r...  
 185 <a href="http://twitter.com/download/iphone" r...  
 194 <a href="http://twitter.com/download/iphone" r...  
 195 <a href="http://twitter.com/download/iphone" r...  
 204 <a href="http://twitter.com/download/iphone" r...  
 211 <a href="http://twitter.com/download/iphone" r...  
 ... ..  
 800 <a href="http://twitter.com/download/iphone" r...  
 811 <a href="http://twitter.com/download/iphone" r...  
 815 <a href="http://twitter.com/download/iphone" r...  
 818 <a href="http://twitter.com/download/iphone" r...  
 822 <a href="http://twitter.com/download/iphone" r...  
 826 <a href="http://twitter.com/download/iphone" r...  
 829 <a href="http://twitter.com/download/iphone" r...  
 833 <a href="http://twitter.com/download/iphone" r...  
 841 <a href="http://twitter.com/download/iphone" r...  
 847 <a href="http://twitter.com/download/iphone" r...  
 860 <a href="http://twitter.com/download/iphone" r...  
 868 <a href="http://twitter.com/download/iphone" r...  
 872 <a href="http://twitter.com/download/iphone" r...  
 885 <a href="http://twitter.com/download/iphone" r...  
 890 <a href="http://twitter.com/download/iphone" r...  
 895 <a href="http://twitter.com/download/iphone" r...  
 908 <a href="http://twitter.com/download/iphone" r...  
 911 <a href="http://twitter.com/download/iphone" r...  
 926 <a href="http://twitter.com/download/iphone" r...  
 937 <a href="http://twitter.com/download/iphone" r...  
 943 <a href="http://twitter.com/download/iphone" r...  
 949 <a href="http://twitter.com/download/iphone" r...  
 1012 <a href="http://twitter.com/download/iphone" r...  
 1023 <a href="http://twitter.com/download/iphone" r...  
 1043 <a href="http://twitter.com/download/iphone" r...  
 1242 <a href="http://twitter.com/download/iphone" r...  
 1286 <a href="http://vine.co" rel="nofollow">Vine -...  
 1860 <a href="http://twitter.com/download/iphone" r...  
 2259 <a href="http://twitter.com" rel="nofollow">Tw...  
 2260 <a href="http://twitter.com" rel="nofollow">Tw...

|    | text  | retweeted_status_id \ |
|----|---|-----------------------|
| 19 | RT @dog_rates: This is Canela. She attempted s... | 8.874740e+17          |
| 32 | RT @Athletics: 12/10 #BATP https://t.co/WxwJmv... | 8.860537e+17          |
| 36 | RT @dog_rates: This is Lilly. She just paralle... | 8.305833e+17          |
| 68 | RT @dog_rates: This is Emmy. She was adopted t... | 8.780576e+17          |

|     |  |              |
|-----|--|--------------|
| 73  | RT @dog_rates: Meet Shadow. In an attempt to r...  | 8.782815e+17 |
| 74  | RT @dog_rates: Meet Terrance. He's being yelle...  | 6.690004e+17 |
| 78  | RT @rachel2195: @dog_rates the boyfriend and h...  | 8.768508e+17 |
| 91  | RT @dog_rates: This is Coco. At first I though...  | 8.663350e+17 |
| 95  | RT @dog_rates: This is Walter. He won't start ...  | 8.688804e+17 |
| 97  | RT @dog_rates: This is Sierra. She's one preci...  | 8.732138e+17 |
| 101 | RT @loganamnosis: Penelope here is doing me qu...  | 8.726576e+17 |
| 109 | RT @dog_rates: This is Dawn. She's just checki...  | 8.410770e+17 |
| 118 | RT @dog_rates: We only rate dogs. This is quit...  | 8.591970e+17 |
| 124 | RT @dog_rates: Say hello to Cooper. His expres...  | 8.685523e+17 |
| 130 | RT @rachaeleasler: these @dog_rates hats are 1...  | 8.650134e+17 |
| 132 | RT @dog_rates: This is Jamesy. He gives a kiss...  | 8.664507e+17 |
| 137 | RT @dog_rates: Here's a pupper before and afte...  | 8.378202e+17 |
| 146 | RT @dog_rates: Say hello to Quinn. She's quite...  | 8.630625e+17 |
| 155 | RT @dog_rates: "Good afternoon class today we'...  | 8.066291e+17 |
| 159 | RT @dog_rates: Meet Lorenzo. He's an avid nift...  | 8.605638e+17 |
| 160 | RT @tallylott: h*ckin adorable promposal. 13/1...  | 8.609145e+17 |
| 165 | RT @dog_rates: Ohboyohboyohboyohboyohboyohboy...   | 7.616730e+17 |
| 171 | RT @dog_rates: Meet Winston. He knows he's a l...  | 8.395493e+17 |
| 180 | RT @AaronChewning: First time wearing my @dog_...  | 8.570611e+17 |
| 182 | RT @dog_rates: This is Luna. It's her first ti...  | 8.447048e+17 |
| 185 | RT @Jenna_Marbles: @dog_rates Thanks for ratin...  | 8.563302e+17 |
| 194 | RT @dog_rates: Meet George. He looks slightly ...  | 8.421635e+17 |
| 195 | RT @frasercampbell_: oh my... what's that... b...  | 8.551225e+17 |
| 204 | RT @dog_rates: I usually only share these on F...  | 8.316501e+17 |
| 211 | RT @dog_rates: This is Astrid. She's a guide d...  | 8.293743e+17 |
| ... | ...  | ...          |
| 800 | RT @dog_rates: This is Gromit. He's pupset bec...  | 7.652221e+17 |
| 811 | RT @dog_rates: This is Frankie. He's wearing b...  | 6.733201e+17 |
| 815 | RT @katieornah: @dog_rates learning a lot at c...  | 7.710021e+17 |
| 818 | RT @dog_rates: Here's a doggo blowing bubbles...   | 7.392382e+17 |
| 822 | RT @dog_rates: This is just downright precious...  | 7.410673e+17 |
| 826 | RT @dog_rates: Ever seen a dog pet another dog...  | 7.069045e+17 |
| 829 | RT @dog_rates: When it's Janet from accounting...  | 7.001438e+17 |
| 833 | RT @dog_rates: This is Nollie. She's waving at...  | 7.399792e+17 |
| 841 | RT @dog_rates: We only rate dogs... this is a ...  | 7.599238e+17 |
| 847 | RT @dog_rates: This is Colby. He's currently r...  | 7.258423e+17 |
| 860 | RT @dog_rates: Meet Eve. She's a raging alcoho...  | 6.732953e+17 |
| 868 | RT @dog_rates: "Tristan do not speak to me wit...  | 6.853251e+17 |
| 872 | RT @dog_rates: Oh. My. God. 13/10 magical af h...  | 7.116948e+17 |
| 885 | RT @hownottodraw: The story/person behind @dog...  | 7.601538e+17 |
| 890 | RT @dog_rates: This... is a Tyrannosaurus rex...   | 7.395441e+17 |
| 895 | RT @dog_rates: AT DAWN...\nWE RIDE\n\n11/10 ht...  | 6.703191e+17 |
| 908 | RT @dog_rates: This is Chompsky. He lives up t...  | 6.790626e+17 |
| 911 | RT @jon_hill1987: @dog_rates There is a cunning... | 7.575971e+17 |
| 926 | RT @dog_rates: This is Rubio. He has too much ...  | 6.791584e+17 |
| 937 | RT @dog_rates: This is Carly. She's actually 2...  | 6.815232e+17 |
| 943 | RT @dog_rates: HEY PUP WHAT'S THE PART OF THE ...  | 6.835159e+17 |



|      |   |              |
|------|---|--------------|
| 949  | RT @dog_rates: Everyone needs to watch this. 1... | 6.753544e+17 |
| 1012 | RT @dog_rates: This pupper killed this great w... | 7.047611e+17 |
| 1023 | RT @dog_rates: This is Shaggy. He knows exactl... | 6.678667e+17 |
| 1043 | RT @dog_rates: Extremely intelligent dog here...  | 6.671383e+17 |
| 1242 | RT @twitter: @dog_rates Awesome Tweet! 12/10. ... | 7.119983e+17 |
| 1286 | RT if you are as ready for summer as this pup ... | NaN          |
| 1860 | RT until we find this dog. Clearly a cool dog ... | NaN          |
| 2259 | RT @dogratingrating: Exceptional talent. Origi... | 6.675487e+17 |
| 2260 | RT @dogratingrating: Unoriginal idea. Blatant ... | 6.675484e+17 |

|     | retweeted_status_user_id | retweeted_status_timestamp | \ |
|-----|--------------------------|----------------------------|---|
| 19  | 4.196984e+09             | 2017-07-19 00:47:34 +0000  |   |
| 32  | 1.960740e+07             | 2017-07-15 02:44:07 +0000  |   |
| 36  | 4.196984e+09             | 2017-02-12 01:04:29 +0000  |   |
| 68  | 4.196984e+09             | 2017-06-23 01:10:23 +0000  |   |
| 73  | 4.196984e+09             | 2017-06-23 16:00:04 +0000  |   |
| 74  | 4.196984e+09             | 2015-11-24 03:51:38 +0000  |   |
| 78  | 5.128045e+08             | 2017-06-19 17:14:49 +0000  |   |
| 91  | 4.196984e+09             | 2017-05-21 16:48:45 +0000  |   |
| 95  | 4.196984e+09             | 2017-05-28 17:23:24 +0000  |   |
| 97  | 4.196984e+09             | 2017-06-09 16:22:42 +0000  |   |
| 101 | 1.547674e+08             | 2017-06-08 03:32:35 +0000  |   |
| 109 | 4.196984e+09             | 2017-03-13 00:02:39 +0000  |   |
| 118 | 4.196984e+09             | 2017-05-02 00:04:57 +0000  |   |
| 124 | 4.196984e+09             | 2017-05-27 19:39:34 +0000  |   |
| 130 | 7.874618e+17             | 2017-05-18 01:17:25 +0000  |   |
| 132 | 4.196984e+09             | 2017-05-22 00:28:40 +0000  |   |
| 137 | 4.196984e+09             | 2017-03-04 00:21:08 +0000  |   |
| 146 | 4.196984e+09             | 2017-05-12 16:05:02 +0000  |   |
| 155 | 4.196984e+09             | 2016-12-07 22:38:52 +0000  |   |
| 159 | 4.196984e+09             | 2017-05-05 18:36:06 +0000  |   |
| 160 | 3.638908e+08             | 2017-05-06 17:49:42 +0000  |   |
| 165 | 4.196984e+09             | 2016-08-05 21:19:27 +0000  |   |
| 171 | 4.196984e+09             | 2017-03-08 18:52:12 +0000  |   |
| 180 | 5.870972e+07             | 2017-04-26 02:37:47 +0000  |   |
| 182 | 4.196984e+09             | 2017-03-23 00:18:10 +0000  |   |
| 185 | 6.669901e+07             | 2017-04-24 02:13:14 +0000  |   |
| 194 | 4.196984e+09             | 2017-03-16 00:00:07 +0000  |   |
| 195 | 7.475543e+17             | 2017-04-20 18:14:33 +0000  |   |
| 204 | 4.196984e+09             | 2017-02-14 23:43:18 +0000  |   |
| 211 | 4.196984e+09             | 2017-02-08 17:00:26 +0000  |   |
| ... | ...                      | ...                        |   |
| 800 | 4.196984e+09             | 2016-08-15 16:22:20 +0000  |   |
| 811 | 4.196984e+09             | 2015-12-06 01:56:44 +0000  |   |
| 815 | 1.732729e+09             | 2016-08-31 15:10:07 +0000  |   |
| 818 | 4.196984e+09             | 2016-06-04 23:31:25 +0000  |   |
| 822 | 4.196984e+09             | 2016-06-10 00:39:48 +0000  |   |
| 826 | 4.196984e+09             | 2016-03-07 18:09:06 +0000  |   |

|      |              |                           |
|------|--------------|---------------------------|
| 829  | 4.196984e+09 | 2016-02-18 02:24:13 +0000 |
| 833  | 4.196984e+09 | 2016-06-07 00:36:02 +0000 |
| 841  | 4.196984e+09 | 2016-08-01 01:28:46 +0000 |
| 847  | 4.196984e+09 | 2016-04-29 00:21:01 +0000 |
| 860  | 4.196984e+09 | 2015-12-06 00:17:55 +0000 |
| 868  | 4.196984e+09 | 2016-01-08 05:00:14 +0000 |
| 872  | 4.196984e+09 | 2016-03-20 23:23:54 +0000 |
| 885  | 1.950368e+08 | 2016-08-01 16:42:51 +0000 |
| 890  | 4.196984e+09 | 2016-06-05 19:47:03 +0000 |
| 895  | 4.196984e+09 | 2015-11-27 19:11:49 +0000 |
| 908  | 4.196984e+09 | 2015-12-21 22:15:18 +0000 |
| 911  | 2.804798e+08 | 2016-07-25 15:23:28 +0000 |
| 926  | 4.196984e+09 | 2015-12-22 04:35:49 +0000 |
| 937  | 4.196984e+09 | 2015-12-28 17:12:42 +0000 |
| 943  | 4.196984e+09 | 2016-01-03 05:11:12 +0000 |
| 949  | 4.196984e+09 | 2015-12-11 16:40:19 +0000 |
| 1012 | 4.196984e+09 | 2016-03-01 20:11:59 +0000 |
| 1023 | 4.196984e+09 | 2015-11-21 00:46:50 +0000 |
| 1043 | 4.196984e+09 | 2015-11-19 00:32:12 +0000 |
| 1242 | 7.832140e+05 | 2016-03-21 19:29:52 +0000 |
| 1286 | NaN          | NaN                       |
| 1860 | NaN          | NaN                       |
| 2259 | 4.296832e+09 | 2015-11-20 03:43:06 +0000 |
| 2260 | 4.296832e+09 | 2015-11-20 03:41:59 +0000 |

|     | expanded_urls                                     | rating_numerator \ |
|-----|---|--------------------|
| 19  | https://twitter.com/dog_rates/status/887473957... | 13                 |
| 32  | https://twitter.com/dog_rates/status/886053434... | 12                 |
| 36  | https://twitter.com/dog_rates/status/830583320... | 13                 |
| 68  | https://twitter.com/dog_rates/status/878057613... | 14                 |
| 73  | https://www.gofundme.com/3yd6y1c,https://twitt... | 13                 |
| 74  | https://twitter.com/dog_rates/status/669000397... | 11                 |
| 78  | https://twitter.com/rachel2195/status/87685077... | 14                 |
| 91  | https://twitter.com/dog_rates/status/866334964... | 12                 |
| 95  | https://twitter.com/dog_rates/status/868880397... | 14                 |
| 97  | https://www.gofundme.com/help-my-baby-sierra-g... | 12                 |
| 101 | https://twitter.com/loganamnosis/status/872657... | 14                 |
| 109 | https://twitter.com/dog_rates/status/841077006... | 12                 |
| 118 | https://twitter.com/dog_rates/status/859196978... | 12                 |
| 124 | https://www.gofundme.com/3ti3nps,https://twitt... | 12                 |
| 130 | https://twitter.com/rachaeleasler/status/86501... | 13                 |
| 132 | https://twitter.com/dog_rates/status/866450705... | 13                 |
| 137 | https://twitter.com/dog_rates/status/837820167... | 12                 |
| 146 | https://www.gofundme.com/helpquinny,https://tw... | 13                 |
| 155 | https://twitter.com/dog_rates/status/806629075... | 13                 |
| 159 | https://www.gofundme.com/help-lorenzo-beat-can... | 13                 |
| 160 | https://twitter.com/tallylott/status/860914485... | 13                 |
| 165 | https://twitter.com/dog_rates/status/761672994... | 10                 |

|      |  |     |
|------|--|-----|
| 171  | <a href="https://twitter.com/dog_rates/status/839549326...">https://twitter.com/dog_rates/status/839549326...</a>                    | 12  |
| 180  | <a href="https://twitter.com/AaronChewning/status/85706...">https://twitter.com/AaronChewning/status/85706...</a>                    | 13  |
| 182  | <a href="https://twitter.com/dog_rates/status/844704788...">https://twitter.com/dog_rates/status/844704788...</a>                    | 13  |
| 185  | NaN  | 14  |
| 194  | <a href="https://twitter.com/dog_rates/status/842163532...">https://twitter.com/dog_rates/status/842163532...</a>                    | 12  |
| 195  | <a href="https://twitter.com/frasercampbell_/status/855...">https://twitter.com/frasercampbell_/status/855...</a>                    | 14  |
| 204  | <a href="http://www.gofundme.com/bluethewhitehusky">http://www.gofundme.com/bluethewhitehusky</a> , <a href="http://...">http...</a> | 13  |
| 211  | <a href="https://twitter.com/dog_rates/status/829374341...">https://twitter.com/dog_rates/status/829374341...</a>                    | 13  |
| ...  | ...  | ... |
| 800  | <a href="https://twitter.com/dog_rates/status/765222098...">https://twitter.com/dog_rates/status/765222098...</a>                    | 10  |
| 811  | <a href="https://twitter.com/dog_rates/status/673320132...">https://twitter.com/dog_rates/status/673320132...</a>                    | 11  |
| 815  | <a href="https://twitter.com/katieornah/status/77100213...">https://twitter.com/katieornah/status/77100213...</a>                    | 12  |
| 818  | <a href="https://twitter.com/dog_rates/status/739238157...">https://twitter.com/dog_rates/status/739238157...</a>                    | 13  |
| 822  | <a href="https://twitter.com/dog_rates/status/741067306...">https://twitter.com/dog_rates/status/741067306...</a>                    | 12  |
| 826  | <a href="https://vine.co/v/iXQAm5Lrgrh">https://vine.co/v/iXQAm5Lrgrh</a> , <a href="https://vine.co/...">https://vine.co/...</a>    | 13  |
| 829  | <a href="https://twitter.com/dog_rates/status/700143752...">https://twitter.com/dog_rates/status/700143752...</a>                    | 10  |
| 833  | <a href="https://twitter.com/dog_rates/status/739979191...">https://twitter.com/dog_rates/status/739979191...</a>                    | 12  |
| 841  | <a href="https://twitter.com/dog_rates/status/759923798...">https://twitter.com/dog_rates/status/759923798...</a>                    | 10  |
| 847  | <a href="https://twitter.com/dog_rates/status/725842289...">https://twitter.com/dog_rates/status/725842289...</a>                    | 12  |
| 860  | <a href="https://twitter.com/dog_rates/status/673295268...">https://twitter.com/dog_rates/status/673295268...</a>                    | 8   |
| 868  | <a href="https://twitter.com/dog_rates/status/685325112...">https://twitter.com/dog_rates/status/685325112...</a>                    | 10  |
| 872  | <a href="https://twitter.com/dog_rates/status/711694788...">https://twitter.com/dog_rates/status/711694788...</a>                    | 13  |
| 885  | <a href="https://weratedogs.com/pages/about-us">https://weratedogs.com/pages/about-us</a> , <a href="https://...">https://...</a>    | 11  |
| 890  | <a href="https://twitter.com/dog_rates/status/739544079...">https://twitter.com/dog_rates/status/739544079...</a>                    | 10  |
| 895  | <a href="https://twitter.com/dog_rates/status/670319130...">https://twitter.com/dog_rates/status/670319130...</a>                    | 11  |
| 908  | <a href="https://twitter.com/dog_rates/status/679062614...">https://twitter.com/dog_rates/status/679062614...</a>                    | 11  |
| 911  | <a href="https://twitter.com/jon_hill1987/status/7575971...">https://twitter.com/jon_hill1987/status/7575971...</a>                  | 11  |
| 926  | <a href="https://twitter.com/dog_rates/status/679158373...">https://twitter.com/dog_rates/status/679158373...</a>                    | 11  |
| 937  | <a href="https://twitter.com/dog_rates/status/681523177...">https://twitter.com/dog_rates/status/681523177...</a>                    | 12  |
| 943  | <a href="https://vine.co/v/ibvnzrauFuV">https://vine.co/v/ibvnzrauFuV</a> , <a href="https://vine.co/...">https://vine.co/...</a>    | 11  |
| 949  | <a href="https://twitter.com/dog_rates/status/675354435...">https://twitter.com/dog_rates/status/675354435...</a>                    | 13  |
| 1012 | <a href="https://twitter.com/dog_rates/status/704761120...">https://twitter.com/dog_rates/status/704761120...</a>                    | 13  |
| 1023 | <a href="https://twitter.com/dog_rates/status/667866724...">https://twitter.com/dog_rates/status/667866724...</a>                    | 10  |
| 1043 | <a href="https://twitter.com/dog_rates/status/667138269...">https://twitter.com/dog_rates/status/667138269...</a>                    | 10  |
| 1242 | <a href="https://twitter.com/twitter/status/71199827977...">https://twitter.com/twitter/status/71199827977...</a>                    | 12  |
| 1286 | <a href="https://vine.co/v/iHFqnjKVbIQ">https://vine.co/v/iHFqnjKVbIQ</a>  | 12  |
| 1860 | <a href="https://twitter.com/dog_rates/status/675489971...">https://twitter.com/dog_rates/status/675489971...</a>                    | 10  |
| 2259 | <a href="https://twitter.com/dogratingrating/status/667...">https://twitter.com/dogratingrating/status/667...</a>                    | 12  |
| 2260 | <a href="https://twitter.com/dogratingrating/status/667...">https://twitter.com/dogratingrating/status/667...</a>                    | 5   |

|    | rating_denominator | name     | doggo | floofer | pupper | puppo |
|----|--------------------|----------|-------|---------|--------|-------|
| 19 | 10                 | Canela   | None  | None    | None   | None  |
| 32 | 10                 | None     | None  | None    | None   | None  |
| 36 | 10                 | Lilly    | None  | None    | None   | None  |
| 68 | 10                 | Emmy     | None  | None    | None   | None  |
| 73 | 10                 | Shadow   | None  | None    | None   | None  |
| 74 | 10                 | Terrance | None  | None    | None   | None  |
| 78 | 10                 | None     | None  | None    | pupper | None  |

|      |     |          |       |      |        |      |
|------|-----|----------|-------|------|--------|------|
| 91   | 10  | Coco     | None  | None | None   | None |
| 95   | 10  | Walter   | None  | None | None   | None |
| 97   | 10  | Sierra   | None  | None | pupper | None |
| 101  | 10  | None     | None  | None | None   | None |
| 109  | 10  | Dawn     | None  | None | None   | None |
| 118  | 10  | quite    | None  | None | None   | None |
| 124  | 10  | Cooper   | None  | None | None   | None |
| 130  | 10  | None     | None  | None | None   | None |
| 132  | 10  | Jamesy   | None  | None | pupper | None |
| 137  | 10  | None     | None  | None | pupper | None |
| 146  | 10  | Quinn    | None  | None | None   | None |
| 155  | 10  | None     | None  | None | None   | None |
| 159  | 10  | Lorenzo  | None  | None | None   | None |
| 160  | 10  | None     | None  | None | None   | None |
| 165  | 10  | None     | None  | None | None   | None |
| 171  | 10  | Winston  | None  | None | None   | None |
| 180  | 10  | None     | None  | None | None   | None |
| 182  | 10  | Luna     | None  | None | None   | None |
| 185  | 10  | None     | None  | None | None   | None |
| 194  | 10  | George   | None  | None | None   | None |
| 195  | 10  | None     | None  | None | None   | None |
| 204  | 10  | None     | None  | None | None   | None |
| 211  | 10  | Astrid   | doggo | None | None   | None |
| ...  | ... | ...      | ...   | ...  | ...    | ...  |
| 800  | 10  | Gromit   | None  | None | None   | None |
| 811  | 10  | Frankie  | None  | None | None   | None |
| 815  | 10  | None     | None  | None | pupper | None |
| 818  | 10  | None     | doggo | None | None   | None |
| 822  | 10  | just     | doggo | None | pupper | None |
| 826  | 10  | None     | None  | None | None   | None |
| 829  | 10  | None     | None  | None | pupper | None |
| 833  | 10  | Nollie   | None  | None | None   | None |
| 841  | 10  | None     | None  | None | None   | None |
| 847  | 10  | Colby    | None  | None | None   | None |
| 860  | 10  | Eve      | None  | None | pupper | None |
| 868  | 10  | None     | None  | None | None   | None |
| 872  | 10  | None     | None  | None | None   | None |
| 885  | 10  | None     | None  | None | None   | None |
| 890  | 10  | None     | None  | None | None   | None |
| 895  | 10  | None     | None  | None | None   | None |
| 908  | 10  | Chompsky | None  | None | None   | None |
| 911  | 10  | None     | None  | None | pupper | None |
| 926  | 10  | Rubio    | None  | None | None   | None |
| 937  | 10  | Carly    | None  | None | None   | None |
| 943  | 10  | None     | None  | None | None   | None |
| 949  | 10  | None     | None  | None | None   | None |
| 1012 | 10  | None     | None  | None | pupper | None |
| 1023 | 10  | Shaggy   | None  | None | None   | None |

|      |    |      |      |      |      |      |
|------|----|------|------|------|------|------|
| 1043 | 10 | None | None | None | None | None |
| 1242 | 10 | None | None | None | None | None |
| 1286 | 10 | None | None | None | None | None |
| 1860 | 10 | None | None | None | None | None |
| 2259 | 10 | None | None | None | None | None |
| 2260 | 10 | None | None | None | None | None |

[183 rows x 17 columns]

### Programmatic Assessment for df\_img\_pred

```
In [15]: # Displaying df_img_pred information
df_img_pred.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id    2075 non-null int64
jpg_url     2075 non-null object
img_num     2075 non-null int64
p1          2075 non-null object
p1_conf     2075 non-null float64
p1_dog      2075 non-null bool
p2          2075 non-null object
p2_conf     2075 non-null float64
p2_dog      2075 non-null bool
p3          2075 non-null object
p3_conf     2075 non-null float64
p3_dog      2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

```
In [16]: # Cheacking if there any duplicated tweet_id
df_img_pred.tweet_id.duplicated().sum()
```

```
Out[16]: 0
```

```
In [17]: # Cheacking if there any duplicated jpg_url
df_img_pred.jpg_url.duplicated().sum()
```

```
Out[17]: 66
```

```
In [18]: # Checking img_num Column is it usless or not
df_img_pred.img_num.value_counts()
```

```
Out[18]: 1    1780
         2     198
         3      66
         4      31
         Name: img_num, dtype: int64
```

```
In [19]: # Checking img_num column has value =4 to be sure that it's useless
df_img_pred.query('img_num == 4')
```

```
Out[19]:
```

|      | tweet_id           | jpg_url \   |
|------|--------------------|---|
| 144  | 668623201287675904 | <a href="https://pbs.twimg.com/media/CUdtP1xUYAIeBnE.jpg">https://pbs.twimg.com/media/CUdtP1xUYAIeBnE.jpg</a> |
| 779  | 689905486972461056 | <a href="https://pbs.twimg.com/media/CZMJYCRVAAE35Wk.jpg">https://pbs.twimg.com/media/CZMJYCRVAAE35Wk.jpg</a> |
| 1024 | 710588934686908417 | <a href="https://pbs.twimg.com/media/CdyE2x1W8AAeOTG.jpg">https://pbs.twimg.com/media/CdyE2x1W8AAeOTG.jpg</a> |
| 1161 | 734787690684657664 | <a href="https://pbs.twimg.com/media/CjJ9gQ1WgAAxQtJ.jpg">https://pbs.twimg.com/media/CjJ9gQ1WgAAxQtJ.jpg</a> |
| 1286 | 750868782890057730 | <a href="https://pbs.twimg.com/media/CmufLLsXYAAsUOr.jpg">https://pbs.twimg.com/media/CmufLLsXYAAsUOr.jpg</a> |
| 1325 | 756998049151549440 | <a href="https://pbs.twimg.com/media/CoFlsGAWgAA2YeV.jpg">https://pbs.twimg.com/media/CoFlsGAWgAA2YeV.jpg</a> |
| 1337 | 758405701903519748 | <a href="https://pbs.twimg.com/media/CoZl9fXWgAMoxOn.jpg">https://pbs.twimg.com/media/CoZl9fXWgAMoxOn.jpg</a> |
| 1342 | 758854675097526272 | <a href="https://pbs.twimg.com/media/Cof-SuqVYAAs4kZ.jpg">https://pbs.twimg.com/media/Cof-SuqVYAAs4kZ.jpg</a> |
| 1372 | 762464539388485633 | <a href="https://pbs.twimg.com/media/CpTRc4DUEAAyTq6.jpg">https://pbs.twimg.com/media/CpTRc4DUEAAyTq6.jpg</a> |
| 1437 | 773985732834758656 | <a href="https://pbs.twimg.com/media/Cr2_6R8WAAAUmtc.jpg">https://pbs.twimg.com/media/Cr2_6R8WAAAUmtc.jpg</a> |
| 1496 | 783391753726550016 | <a href="https://pbs.twimg.com/media/Ct8qn8EWIAAk9zP.jpg">https://pbs.twimg.com/media/Ct8qn8EWIAAk9zP.jpg</a> |
| 1542 | 791312159183634433 | <a href="https://pbs.twimg.com/media/CvtONV4WAAAQ3Rn.jpg">https://pbs.twimg.com/media/CvtONV4WAAAQ3Rn.jpg</a> |
| 1543 | 791406955684368384 | <a href="https://pbs.twimg.com/media/CvukbEkWAAAV-69.jpg">https://pbs.twimg.com/media/CvukbEkWAAAV-69.jpg</a> |
| 1549 | 792913359805018113 | <a href="https://pbs.twimg.com/media/CwD-eCLWIAA6vOB.jpg">https://pbs.twimg.com/media/CwD-eCLWIAA6vOB.jpg</a> |
| 1617 | 802323869084381190 | <a href="https://pbs.twimg.com/media/CyJtSmDUAAA2F9x.jpg">https://pbs.twimg.com/media/CyJtSmDUAAA2F9x.jpg</a> |
| 1713 | 818614493328580609 | <a href="https://pbs.twimg.com/media/C1xNgraVIAA3EVb.jpg">https://pbs.twimg.com/media/C1xNgraVIAA3EVb.jpg</a> |
| 1716 | 819006400881917954 | <a href="https://pbs.twimg.com/media/C12x-JTVIAAZdf1.jpg">https://pbs.twimg.com/media/C12x-JTVIAAZdf1.jpg</a> |
| 1717 | 819015331746349057 | <a href="https://pbs.twimg.com/media/C12x-JTVIAAZdf1.jpg">https://pbs.twimg.com/media/C12x-JTVIAAZdf1.jpg</a> |
| 1766 | 826848821049180160 | <a href="https://pbs.twimg.com/media/C3mOnZ_XUAAjr2V.jpg">https://pbs.twimg.com/media/C3mOnZ_XUAAjr2V.jpg</a> |
| 1768 | 827199976799354881 | <a href="https://pbs.twimg.com/media/C3rN-lcWEAA9CmR.jpg">https://pbs.twimg.com/media/C3rN-lcWEAA9CmR.jpg</a> |
| 1790 | 830097400375152640 | <a href="https://pbs.twimg.com/media/C4UZLZLWYAAOdcs.jpg">https://pbs.twimg.com/media/C4UZLZLWYAAOdcs.jpg</a> |
| 1795 | 831315979191906304 | <a href="https://pbs.twimg.com/media/C41st0bXAAE6MP8.jpg">https://pbs.twimg.com/media/C41st0bXAAE6MP8.jpg</a> |
| 1800 | 831911600680497154 | <a href="https://pbs.twimg.com/media/C4uLLGuUoAAkIHm.jpg">https://pbs.twimg.com/media/C4uLLGuUoAAkIHm.jpg</a> |
| 1829 | 836001077879255040 | <a href="https://pbs.twimg.com/media/C5oSiskUOAE8sJ_.jpg">https://pbs.twimg.com/media/C5oSiskUOAE8sJ_.jpg</a> |
| 1893 | 849412302885593088 | <a href="https://pbs.twimg.com/media/C8m3-iQVoAAETnF.jpg">https://pbs.twimg.com/media/C8m3-iQVoAAETnF.jpg</a> |
| 1914 | 854120357044912130 | <a href="https://pbs.twimg.com/media/C9px7jyVwAAnmWn.jpg">https://pbs.twimg.com/media/C9px7jyVwAAnmWn.jpg</a> |
| 1920 | 856282028240666624 | <a href="https://pbs.twimg.com/media/C-If9ZwXoAAfDX2.jpg">https://pbs.twimg.com/media/C-If9ZwXoAAfDX2.jpg</a> |
| 1934 | 859851578198683649 | <a href="https://pbs.twimg.com/media/C-70cfyXsAAsqzU.jpg">https://pbs.twimg.com/media/C-70cfyXsAAsqzU.jpg</a> |
| 1954 | 864197398364647424 | <a href="https://pbs.twimg.com/media/C_4-8iPVOAA1Twg.jpg">https://pbs.twimg.com/media/C_4-8iPVOAA1Twg.jpg</a> |
| 1978 | 870656317836468226 | <a href="https://pbs.twimg.com/media/DBUxSSTXsAA-Jn1.jpg">https://pbs.twimg.com/media/DBUxSSTXsAA-Jn1.jpg</a> |
| 2040 | 885167619883638784 | <a href="https://pbs.twimg.com/media/DEi_N9qXYAAgEEw.jpg">https://pbs.twimg.com/media/DEi_N9qXYAAgEEw.jpg</a> |

|      | img_num | p1                       | p1_conf  | p1_dog \ |
|------|---------|--------------------------|----------|----------|
| 144  | 4       | Chihuahua                | 0.708163 | True     |
| 779  | 4       | Pomeranian               | 0.943331 | True     |
| 1024 | 4       | Pembroke                 | 0.982004 | True     |
| 1161 | 4       | golden_retriever         | 0.883991 | True     |
| 1286 | 4       | toy_poodle               | 0.912648 | True     |
| 1325 | 4       | golden_retriever         | 0.678555 | True     |
| 1337 | 4       | Chesapeake_Bay_retriever | 0.702954 | True     |
| 1342 | 4       | barrow                   | 0.974047 | False    |
| 1372 | 4       | chow                     | 0.999953 | True     |
| 1437 | 4       | giant_panda              | 0.451149 | False    |
| 1496 | 4       | Norwegian_elkhound       | 0.877130 | True     |

|      |   |                         |          |       |
|------|---|-------------------------|----------|-------|
| 1542 | 4 | miniature_pinscher      | 0.892925 | True  |
| 1543 | 4 | Pembroke                | 0.972629 | True  |
| 1549 | 4 | web_site                | 0.226716 | False |
| 1617 | 4 | home_theater            | 0.765069 | False |
| 1713 | 4 | Chihuahua               | 0.450722 | True  |
| 1716 | 4 | prison                  | 0.907083 | False |
| 1717 | 4 | prison                  | 0.907083 | False |
| 1766 | 4 | Great_Pyrenees          | 0.858764 | True  |
| 1768 | 4 | Great_Dane              | 0.869681 | True  |
| 1790 | 4 | toy_poodle              | 0.442713 | True  |
| 1795 | 4 | briard                  | 0.982755 | True  |
| 1800 | 4 | bloodhound              | 0.777562 | True  |
| 1829 | 4 | Samoyed                 | 0.963558 | True  |
| 1893 | 4 | schipperke              | 0.907559 | True  |
| 1914 | 4 | black-and-tan_coonhound | 0.854861 | True  |
| 1920 | 4 | Chihuahua               | 0.876543 | True  |
| 1934 | 4 | Labrador_retriever      | 0.899086 | True  |
| 1954 | 4 | golden_retriever        | 0.945905 | True  |
| 1978 | 4 | Pembroke                | 0.945495 | True  |
| 2040 | 4 | malamute                | 0.812482 | True  |

|      |                                | p2                   | p2_conf  | p2_dog | p3 \                 |
|------|--------------------------------|----------------------|----------|--------|----------------------|
| 144  |                                | Pomeranian           | 0.091372 | True   | titi                 |
| 779  |                                | Shetland_sheepdog    | 0.023675 | True   | chow                 |
| 1024 |                                | Cardigan             | 0.008943 | True   | malamute             |
| 1161 |                                | chow                 | 0.023542 | True   | Labrador_retriever   |
| 1286 |                                | miniature_poodle     | 0.035059 | True   | seat_belt            |
| 1325 |                                | Labrador_retriever   | 0.072632 | True   | Border_terrier       |
| 1337 |                                | laptop               | 0.092277 | False  | notebook             |
| 1342 |                                | Old_English_sheepdog | 0.023791 | True   | komondor             |
| 1372 |                                | Tibetan_mastiff      | 0.000023 | True   | dhole                |
| 1437 |                                | fur_coat             | 0.148001 | False  | pug                  |
| 1496 |                                | cairn                | 0.086241 | True   | keeshond             |
| 1542 |                                | toy_terrier          | 0.095524 | True   | Doberman             |
| 1543 |                                | Cardigan             | 0.027026 | True   | basenji              |
| 1549 |                                | lighter              | 0.081941 | False  | switch               |
| 1617 |                                | television           | 0.203578 | False  | entertainment_center |
| 1713 |                                | Border_terrier       | 0.204177 | True   | beagle               |
| 1716 |                                | palace               | 0.020089 | False  | umbrella             |
| 1717 |                                | palace               | 0.020089 | False  | umbrella             |
| 1766 |                                | golden_retriever     | 0.023526 | True   | Pekinese             |
| 1768 | American_Staffordshire_terrier |                      | 0.026658 | True   | boxer                |
| 1790 |                                | Pomeranian           | 0.142073 | True   | Pekinese             |
| 1795 | soft-coated_wheaten_terrier    |                      | 0.009084 | True   | Bouvier_des_Flandres |
| 1800 |                                | Great_Dane           | 0.047418 | True   | Leonberg             |
| 1829 |                                | white_wolf           | 0.019848 | False  | malamute             |
| 1893 |                                | crossword_puzzle     | 0.017934 | False  | Chihuahua            |
| 1914 |                                | Doberman             | 0.050792 | True   | bluetick             |

|      |                    |          |      |                 |
|------|--------------------|----------|------|-----------------|
| 1920 | Italian_greyhound  | 0.032962 | True | Cardigan        |
| 1934 | golden_retriever   | 0.047091 | True | kuvasz          |
| 1954 | Labrador_retriever | 0.021264 | True | Tibetan_mastiff |
| 1978 | Cardigan           | 0.045875 | True | beagle          |
| 2040 | Siberian_husky     | 0.071712 | True | Eskimo_dog      |

|      | p3_conf  | p3_dog |
|------|----------|--------|
| 144  | 0.067325 | False  |
| 779  | 0.007165 | True   |
| 1024 | 0.007550 | True   |
| 1161 | 0.016056 | True   |
| 1286 | 0.026376 | False  |
| 1325 | 0.049033 | True   |
| 1337 | 0.032727 | False  |
| 1342 | 0.001246 | True   |
| 1372 | 0.000003 | False  |
| 1437 | 0.109570 | True   |
| 1496 | 0.011019 | True   |
| 1542 | 0.003544 | True   |
| 1543 | 0.000153 | True   |
| 1549 | 0.039009 | False  |
| 1617 | 0.018644 | False  |
| 1713 | 0.092774 | True   |
| 1716 | 0.007850 | False  |
| 1717 | 0.007850 | False  |
| 1766 | 0.017104 | True   |
| 1768 | 0.019866 | True   |
| 1790 | 0.125745 | True   |
| 1795 | 0.004693 | True   |
| 1800 | 0.017943 | True   |
| 1829 | 0.005904 | True   |
| 1893 | 0.016191 | True   |
| 1914 | 0.021762 | True   |
| 1920 | 0.020776 | True   |
| 1934 | 0.023206 | True   |
| 1954 | 0.020493 | True   |
| 1978 | 0.004329 | True   |
| 2040 | 0.055770 | True   |

### Programmatic Assessment for df\_api

```
In [20]: # Displaying df_api information
df_api.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 31 columns):
contributors                                0 non-null float64
```



```

coordinates          0 non-null float64
created_at           2354 non-null datetime64[ns]
display_text_range   2354 non-null object
entities             2354 non-null object
extended_entities    2073 non-null object
favorite_count       2354 non-null int64
favorited            2354 non-null bool
full_text            2354 non-null object
geo                  0 non-null float64
id                   2354 non-null int64
id_str               2354 non-null int64
in_reply_to_screen_name 78 non-null object
in_reply_to_status_id 78 non-null float64
in_reply_to_status_id_str 78 non-null float64
in_reply_to_user_id  78 non-null float64
in_reply_to_user_id_str 78 non-null float64
is_quote_status      2354 non-null bool
lang                 2354 non-null object
place                1 non-null object
possibly_sensitive    2211 non-null float64
possibly_sensitive_appealable 2211 non-null float64
quoted_status        28 non-null object
quoted_status_id     29 non-null float64
quoted_status_id_str 29 non-null float64
retweet_count        2354 non-null int64
retweeted             2354 non-null bool
retweeted_status      179 non-null object
source               2354 non-null object
truncated            2354 non-null bool
user                 2354 non-null object
dtypes: bool(4), datetime64[ns](1), float64(11), int64(4), object(11)
memory usage: 505.8+ KB

```

```

In [21]: # Cheacking if there any duplicated id
         df_api.id.duplicated().sum()

```

```

Out[21]: 0

```

```

In [22]: # Checking the max of retweets and favorite Counts
         df_api[['retweet_count', 'favorite_count']].describe()

```

```

Out[22]:
         retweet_count  favorite_count
count      2354.000000      2354.000000
mean       3164.797366      8080.968564
std        5284.770364     11814.771334
min          0.000000          0.000000
25%         624.500000      1415.000000
50%        1473.500000      3603.500000

```

|     |              |               |
|-----|--------------|---------------|
| 75% | 3652.000000  | 10122.250000  |
| max | 79515.000000 | 132810.000000 |

### 0.3 Assessment Summary in points.

Completeness, Validity, Accuracy and consistency

df\_twitter\_arch

Quality 1- timestamp column is str instead of datetime 2- The rating\_numerator column should be of type float and also has invalid values it should be correctly extracted from text column. 3- There are retweets should be removed. 4- There are invalid names in name column should be removed. 5- Source has a link should be extracted to get only source and should be changed to category. 6- (doggo, floofer, puppe and puppo) Columns has none or missing values. Tidiness 1- doggo, floofer, puppe and puppo columns should be merged in one column dog\_types

df\_img\_pred

Quality 1- There is 66 duplicated url in jpg\_url should be removed. 2- img\_num is useless and should be removed. Tidiness 1- p1\_conf, p2\_conf and p3\_conf columns should be merged in one column confidence. 2- p1\_dog, p2\_dog and p3\_dog columns should be merged in one column dog\_type.

df\_api

Quality 1- id column should be renamed to tweet\_id 2- All columns are useless except (id, favorite\_count and retweet\_count) and should be removed.

### 0.4 Cleaning dataframes.

#### 0.4.1 Creating a copy from each dataframe to begin the cleaning process.

```
In [23]: # Creating a copy from each dataframe to begin the cleaning process.
# Creating a copy from Twitter Archive Dataframe to work on it.
df_twitter_arch_clean = df_twitter_arch.copy()
# Creating a copy from Image Predictions Dataframe to work on it.
df_img_pred_clean = df_img_pred.copy()
# Creating a copy from Twitter API Dataframe to work on it.
df_api_clean = df_api.copy()
```

#### 0.4.2 Cleaning Twitter Archive from (df\_twitter\_arch to df\_twitter\_arch\_clean)

Timestamp column in df\_twitter\_arch is str. And need to be changed to datetime.

Code

```
In [24]: # Changing the timestamp column type from str to datetime type.
df_twitter_arch_clean.timestamp = pd.to_datetime(df_twitter_arch_clean.timestamp)
```

Test

```
In [25]: df_twitter_arch_clean.timestamp.dtype
```

```
Out[25]: dtype('<M8[ns]')
```

### 0.4.3 expanded\_urls has Nan values should be replaced

Code

```
In [26]: # Replacing all NaN values in expanded_urls column with empty to avoid errors with wran
df_twitter_arch_clean.expanded_urls.fillna('', inplace=True)
```

Test

```
In [27]: df_twitter_arch_clean['expanded_urls'].isnull().sum()
```

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

### 0.4.4 Removing all rows has values in retweets

Code

```
In [28]: # Removing all rows that have values in retweeted_status_id column.
df_twitter_arch_clean = df_twitter_arch_clean[df_twitter_arch_clean.retweeted_status_id
# Removing all rows that have values in retweeted_status_user_id column.
df_twitter_arch_clean = df_twitter_arch_clean[df_twitter_arch_clean.retweeted_status_us
# Removing all rows that have values in retweeted_status_timestamp column.
df_twitter_arch_clean = df_twitter_arch_clean[df_twitter_arch_clean.retweeted_status_ti
```

Test

```
In [29]: df_twitter_arch_clean.retweeted_status_id.describe()
```

```
Out[29]: count      0.0
mean      NaN
std       NaN
min       NaN
25%       NaN
50%       NaN
75%       NaN
max       NaN
Name: retweeted_status_id, dtype: float64
```

Test

```
In [30]: df_twitter_arch_clean.retweeted_status_user_id.describe()
```

```
Out[30]: count      0.0
mean      NaN
std       NaN
min       NaN
25%       NaN
50%       NaN
75%       NaN
max       NaN
Name: retweeted_status_user_id, dtype: float64
```

Test

```
In [31]: df_twitter_arch_clean.retweeted_status_timestamp.describe()
```

```
Out[31]: count      0
         unique      0
         Name: retweeted_status_timestamp, dtype: int64
```

The rating\_numerator column should of type float and also has invalid values it should be correctly extracted from text column.

Code

```
In [32]: # Extracting the correct rating_numerator from text column by using RegEx.
         ext_rating = df_twitter_arch_clean[df_twitter_arch_clean.text.str.contains(r"(\d+\.\d*)")
         ext_rating
```

/opt/conda/lib/python3.6/site-packages/ipykernel\_launcher.py:2: UserWarning: This pattern has ma

```
Out[32]:      0
         45    13.5
         695    9.75
         763   11.27
        1689    9.5
        1712   11.26
```

```
In [33]: # Adding the correct values to rating_numerator column
         df_twitter_arch_clean.loc[ext_rating.index, 'rating_numerator'] = ext_rating.values
```

```
In [34]: # Changing the rating_numerator to float
         df_twitter_arch_clean.rating_numerator = df_twitter_arch_clean.rating_numerator.astype(float)
```

Test

```
In [35]: df_twitter_arch_clean.rating_numerator.dtype
```

```
Out[35]: dtype('float64')
```

#### 0.4.5 The rating\_denominator column has invalid values should be removed.

Code

```
In [36]: # Removing all values greater than the normal rating denominator > 10.
         df_twitter_arch_clean = df_twitter_arch_clean[df_twitter_arch_clean['rating_denominator'] < 10]
```

Test

```
In [37]: df_twitter_arch_clean.rating_denominator.describe()
```

```
Out[37]: count      2153.0
         mean        10.0
         std         0.0
         min         10.0
         25%         10.0
         50%         10.0
         75%         10.0
         max         10.0
         Name: rating_denominator, dtype: float64
```

#### 0.4.6 There are retweets should be removed with RT start.

Code

```
In [38]: # Removing all retweets rows from text column that starts with "RT"
         df_twitter_arch = df_twitter_arch[~df_twitter_arch['text'].str.startswith('RT')]
```

Test

```
In [39]: df_twitter_arch[df_twitter_arch['text'].str.startswith('RT')]
```

```
Out[39]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text]
         Index: []
```

#### 0.4.7 There are invalid names in name column should be removed.

Code

```
In [40]: # detecting the invalid names from the name column
         df_twitter_arch_clean[~df_twitter_arch_clean.name.str.istitle()].name.unique()
```

```
Out[40]: array(['such', 'a', 'quite', 'not', 'one', 'incredibly', 'BeBe', 'very',
               'DonDon', 'my', 'his', 'an', 'actually', 'just', 'getting', 'mad',
               'unacceptable', 'all', 'old', 'infuriating', 'CeCe', 'the', 'by',
               'officially', 'life', 'light', 'space', 'JD', 'DayZ'], dtype=object)
```

```
In [41]: # Replacing the typo names to be correct.
         df_twitter_arch_clean.name.replace(['DonDon', 'CeCe'], ['Dondon', 'Cece'], inplace=True)
```

```
In [42]: # detecting the invalid names from the name column after correcting some.
         df_twitter_arch_clean[~df_twitter_arch_clean.name.str.istitle()].name.unique()
```

```
Out[42]: array(['such', 'a', 'quite', 'not', 'one', 'incredibly', 'BeBe', 'very',
               'my', 'his', 'an', 'actually', 'just', 'getting', 'mad',
               'unacceptable', 'all', 'old', 'infuriating', 'the', 'by',
               'officially', 'life', 'light', 'space', 'JD', 'DayZ'], dtype=object)
```

```
In [43]: # Replacing all invalid names by np.nan.
df_twitter_arch_clean.name.replace(['such', 'a', 'quite', 'not', 'one', 'incredibly', '
    'my', 'his', 'an', 'actually', 'just', 'getting', 'mad',
    'unacceptable', 'all', 'old', 'infuriating', 'the', 'by',
    'officially', 'life', 'light', 'space', 'JD', 'DayZ', 'None'], np.nan, inplace=True)
```

Test

```
In [44]: df_twitter_arch_clean.name.unique()
```

```
Out[44]: array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', nan, 'Jax',
    'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver', 'Jim',
    'Zeke', 'Ralphus', 'Gerald', 'Jeffrey', 'Canela', 'Maya', 'Mingus',
    'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Earl', 'Lola',
    'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald', 'Rusty', 'Gus',
    'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'Elliot', 'Louis',
    'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack', 'Steven', 'Beau',
    'Snoopy', 'Shadow', 'Emmy', 'Aja', 'Penny', 'Dante', 'Nelly',
    'Ginger', 'Benedict', 'Venti', 'Goose', 'Nugget', 'Cash', 'Jed',
    'Sebastian', 'Sierra', 'Monkey', 'Harry', 'Kody', 'Lassie', 'Rover',
    'Napolean', 'Boomer', 'Cody', 'Rumble', 'Clifford', 'Dewey',
    'Scout', 'Gizmo', 'Walter', 'Cooper', 'Harold', 'Shikha', 'Lili',
    'Jamesy', 'Coco', 'Sammy', 'Meatball', 'Paisley', 'Albus',
    'Neptune', 'Belle', 'Quinn', 'Zooey', 'Dave', 'Jersey', 'Hobbes',
    'Burt', 'Lorenzo', 'Carl', 'Jordy', 'Milky', 'Trooper', 'Sophie',
    'Wyatt', 'Rosie', 'Thor', 'Oscar', 'Callie', 'Cermet', 'Marlee',
    'Arya', 'Einstein', 'Alice', 'Rumpole', 'Benny', 'Aspen', 'Jarod',
    'Wiggles', 'General', 'Sailor', 'Iggy', 'Snoop', 'Kyle', 'Leo',
    'Riley', 'Noosh', 'Odin', 'Jerry', 'Georgie', 'Rontu', 'Cannon',
    'Furzey', 'Daisy', 'Tuck', 'Barney', 'Vixen', 'Jarvis', 'Mimosa',
    'Pickles', 'Brady', 'Luna', 'Charlie', 'Margo', 'Sadie', 'Hank',
    'Tycho', 'Indie', 'Winnie', 'George', 'Bentley', 'Max', 'Dawn',
    'Maddie', 'Monty', 'Sojourner', 'Winston', 'Odie', 'Arlo',
    'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera', 'Ava', 'Eli', 'Ash',
    'Tucker', 'Tobi', 'Chester', 'Wilson', 'Sunshine', 'Lipton',
    'Bronte', 'Poppy', 'Gidget', 'Rhino', 'Willow', 'Orion', 'Eevee',
    'Smiley', 'Miguel', 'Emanuel', 'Kuyu', 'Dutch', 'Pete', 'Scooter',
    'Reggie', 'Lilly', 'Samson', 'Mia', 'Astrid', 'Malcolm', 'Dexter',
    'Alfie', 'Fiona', 'Mutt', 'Bear', 'Doobert', 'Beebop', 'Alexander',
    'Sailer', 'Brutus', 'Kona', 'Boots', 'Ralphie', 'Loki', 'Cupid',
    'Pawnd', 'Pilot', 'Ike', 'Mo', 'Toby', 'Sweet', 'Pablo', 'Nala',
    'Crawford', 'Gabe', 'Jimison', 'Duchess', 'Harlso', 'Sundance',
    'Luca', 'Flash', 'Sunny', 'Howie', 'Jazzy', 'Anna', 'Finn', 'Bo',
    'Wafer', 'Tom', 'Florence', 'Autumn', 'Buddy', 'Dido', 'Eugene',
    'Ken', 'Strudel', 'Tebow', 'Chloe', 'Timber', 'Binky', 'Moose',
    'Dudley', 'Comet', 'Akumi', 'Titan', 'Olivia', 'Alf', 'Oshie',
    'Chubbs', 'Sky', 'Atlas', 'Eleanor', 'Layla', 'Rocky', 'Baron',
    'Tyr', 'Bauer', 'Swagger', 'Brandi', 'Mary', 'Moe', 'Halo', 'Augie',
```

'Craig', 'Hunter', 'Pavlov', 'Phil', 'Kyro', 'Wallace', 'Ito',  
 'Seamus', 'Ollie', 'Stephan', 'Lennon', 'Major', 'Duke', 'Sansa',  
 'Shooter', 'Django', 'Diogi', 'Sonny', 'Marley', 'Severus',  
 'Ronnie', 'Milo', 'Bones', 'Mauve', 'Chef', 'Doc', 'Peaches',  
 'Sobe', 'Longfellow', 'Mister', 'Iroh', 'Pancake', 'Snicku', 'Ruby',  
 'Brody', 'Mack', 'Nimbus', 'Laika', 'Maximus', 'Dobby', 'Moreton',  
 'Juno', 'Maude', 'Lily', 'Newt', 'Benji', 'Nida', 'Robin',  
 'Monster', 'Remus', 'Levi', 'Mabel', 'Misty', 'Betty', 'Mosby',  
 'Maggie', 'Bruce', 'Happy', 'Ralphy', 'Brownie', 'Rizzy', 'Stella',  
 'Butter', 'Frank', 'Tonks', 'Lincoln', 'Rory', 'Logan', 'Dale',  
 'Rizzo', 'Arnie', 'Mattie', 'Pinot', 'Dallas', 'Hero', 'Frankie',  
 'Stormy', 'Reginald', 'Balto', 'Mairi', 'Loomis', 'Godi', 'Cali',  
 'Deacon', 'Timmy', 'Sampson', 'Chipson', 'Combo', 'Oakley', 'Dash',  
 'Hercules', 'Jay', 'Mya', 'Strider', 'Wesley', 'Solomon', 'Huck',  
 'O', 'Blue', 'Anakin', 'Finley', 'Sprinkles', 'Heinrich',  
 'Shakespeare', 'Chelsea', 'Bungalo', 'Chip', 'Grey', 'Roosevelt',  
 'Willem', 'Davey', 'Dakota', 'Fizz', 'Dixie', 'Al', 'Jackson',  
 'Carbon', 'Klein', 'Dondon', 'Kirby', 'Lou', 'Chevy', 'Tito',  
 'Philbert', 'Louie', 'Rupert', 'Rufus', 'Brudge', 'Shadoe', 'Angel',  
 'Brat', 'Tove', 'Gromit', 'Aubie', 'Kota', 'Leela', 'Glenn',  
 'Shelby', 'Sephie', 'Bonaparte', 'Albert', 'Wishes', 'Rose', 'Theo',  
 'Rocco', 'Fido', 'Emma', 'Spencer', 'Lilli', 'Boston', 'Brandonald',  
 'Corey', 'Leonard', 'Beckham', 'Devón', 'Gert', 'Watson', 'Keith',  
 'Dex', 'Ace', 'Tayzie', 'Grizzie', 'Fred', 'Gilbert', 'Meyer',  
 'Zoe', 'Stewie', 'Calvin', 'Lilah', 'Spanky', 'Jameson', 'Piper',  
 'Atticus', 'Blu', 'Dietrich', 'Divine', 'Tripp', 'Cora', 'Huxley',  
 'Keurig', 'Bookstore', 'Linus', 'Abby', 'Shiloh', 'Gustav', 'Arlen',  
 'Percy', 'Lenox', 'Sugar', 'Harvey', 'Blanket', 'Geno', 'Stark',  
 'Beya', 'Kilo', 'Kayla', 'Maxaroni', 'Bell', 'Doug', 'Edmund',  
 'Aqua', 'Theodore', 'Baloo', 'Chase', 'Nollie', 'Rorie', 'Simba',  
 'Charles', 'Bayley', 'Axel', 'Storkson', 'Remy', 'Chadrick',  
 'Kellogg', 'Buckley', 'Livvie', 'Terry', 'Hermione', 'Ralpher',  
 'Aldrick', 'Larry', 'Rooney', 'Crystal', 'Ziva', 'Stefan',  
 'Pupcasso', 'Puff', 'Flurpson', 'Coleman', 'Enchilada', 'Raymond',  
 'Rueben', 'Cilantro', 'Karll', 'Sprout', 'Blitz', 'Bloop', 'Colby',  
 'Lillie', 'Ashleigh', 'Kreggory', 'Sarge', 'Luther', 'Ivar',  
 'Jangle', 'Schnitzel', 'Panda', 'Berkeley', 'Ralphé', 'Charleson',  
 'Clyde', 'Harnold', 'Sid', 'Pippa', 'Otis', 'Carper', 'Bowie',  
 'Alexanderson', 'Suki', 'Barclay', 'Skittle', 'Ebby', 'Flávio',  
 'Smokey', 'Link', 'Jennifur', 'Ozzy', 'Stephanus', 'Bubbles',  
 'Zeus', 'Bertson', 'Nico', 'Michelangelo', 'Siba', 'Calbert',  
 'Curtis', 'Travis', 'Thumas', 'Kanu', 'Lance', 'Opie', 'Stubert',  
 'Kane', 'Olive', 'Chuckles', 'Staniel', 'Sora', 'Beemo', 'Gunner',  
 'Lacy', 'Tater', 'Olaf', 'Cecil', 'Vince', 'Karma', 'Billy',  
 'Walker', 'Rodney', 'Klevin', 'Malikai', 'Bobbie', 'River',  
 'Jebberson', 'Remington', 'Farfle', 'Jiminus', 'Harper', 'Clarkus',  
 'Finnegus', 'Cupcake', 'Kathmandu', 'Ellie', 'Katie', 'Kara',  
 'Adele', 'Zara', 'Ambrose', 'Jimothy', 'Bode', 'Terrenth', 'Reese',

'Chesterson', 'Lucia', 'Bisquick', 'Ralphson', 'Socks', 'Rambo',  
 'Rudy', 'Fiji', 'Rilo', 'Bilbo', 'Coopson', 'Yoda', 'Millie',  
 'Chet', 'Crouton', 'Daniel', 'Kaia', 'Murphy', 'Dotsy', 'Eazy',  
 'Coops', 'Fillup', 'Miley', 'Charl', 'Reagan', 'Yukon', 'Cece',  
 'Cuddles', 'Claude', 'Jessiga', 'Carter', 'Ole', 'Pherb', 'Blipson',  
 'Reptar', 'Trevith', 'Berb', 'Bob', 'Colin', 'Brian', 'Oliviér',  
 'Grady', 'Kobe', 'Freddery', 'Bodie', 'Dunkin', 'Wally', 'Tupawc',  
 'Amber', 'Herschel', 'Edgar', 'Teddy', 'Kingsley', 'Brockly',  
 'Richie', 'Molly', 'Vinscent', 'Cedrick', 'Hazel', 'Lolo', 'Eriq',  
 'Phred', 'Oddie', 'Maxwell', 'Geoff', 'Covach', 'Durg', 'Fynn',  
 'Ricky', 'Herald', 'Lucky', 'Ferg', 'Trip', 'Clarence', 'Hamrick',  
 'Brad', 'Pubert', 'Frönq', 'Derby', 'Lizzie', 'Ember', 'Blakely',  
 'Opal', 'Marq', 'Kramer', 'Barry', 'Tyrone', 'Gordon', 'Baxter',  
 'Mona', 'Horace', 'Crimson', 'Birf', 'Hammond', 'Lorelei', 'Marty',  
 'Brooks', 'Petrick', 'Hubertson', 'Gerbald', 'Oreo', 'Bruiser',  
 'Perry', 'Bobby', 'Jeph', 'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe',  
 'Tiger', 'Jiminy', 'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Taco',  
 'Joey', 'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses', 'Toffee',  
 'Apollo', 'Carly', 'Asher', 'Glacier', 'Chuck', 'Champ', 'Ozzie',  
 'Griswold', 'Cheesy', 'Moofasa', 'Hector', 'Goliath', 'Kawhi',  
 'Emmie', 'Penelope', 'Willie', 'Rinna', 'Mike', 'William', 'Dwight',  
 'Evy', 'Hurley', 'Rubio', 'Chompsky', 'Rascal', 'Linda', 'Tug',  
 'Tango', 'Grizz', 'Jerome', 'Crumpet', 'Jessifer', 'Izzy', 'Ralph',  
 'Sandy', 'Humphrey', 'Tassy', 'Juckson', 'Chuq', 'Tyrus', 'Karl',  
 'Godzilla', 'Vinnie', 'Kenneth', 'Herm', 'Bert', 'Striker', 'Donny',  
 'Pepper', 'Bernie', 'Buddah', 'Lenny', 'Arnold', 'Zuzu', 'Mollie',  
 'Laela', 'Tedders', 'Superpup', 'Rufio', 'Jeb', 'Rodman', 'Jonah',  
 'Chesney', 'Kenny', 'Henry', 'Bobbay', 'Mitch', 'Kaiya', 'Acro',  
 'Aiden', 'Obie', 'Dot', 'Shnuggles', 'Kendall', 'Jeffri', 'Steve',  
 'Eve', 'Mac', 'Fletcher', 'Kenzie', 'Pumpkin', 'Schnozz', 'Gustaf',  
 'Cheryl', 'Ed', 'Leonidas', 'Norman', 'Caryl', 'Scott', 'Taz',  
 'Darby', 'Jackie', 'Jazz', 'Franq', 'Pippin', 'Rolf', 'Snickers',  
 'Ridley', 'Cal', 'Bradley', 'Bubba', 'Tuco', 'Patch', 'Mojo',  
 'Batdog', 'Dylan', 'Mark', 'Alejandro', 'Scruffers', 'Pip',  
 'Julius', 'Tanner', 'Sparky', 'Anthony', 'Holly', 'Jett', 'Amy',  
 'Sage', 'Andy', 'Mason', 'Trigger', 'Antony', 'Creg', 'Traviss',  
 'Gin', 'Jeffrie', 'Danny', 'Ester', 'Pluto', 'Bloo', 'Edd', 'Paull',  
 'Sam', 'Willy', 'Herb', 'Damon', 'Peanut', 'Nigel', 'Butters',  
 'Sandra', 'Fabio', 'Randall', 'Liam', 'Tommy', 'Ben', 'Raphael',  
 'Julio', 'Andru', 'Kloey', 'Shawwn', 'Skye', 'Kollin', 'Ronduh',  
 'Billl', 'Saydee', 'Dug', 'Tessa', 'Sully', 'Kirk', 'Ralf', 'Clarq',  
 'Jaspers', 'Samsom', 'Terrance', 'Harrison', 'Chaz', 'Jeremy',  
 'Jaycob', 'Lambeau', 'Ruffles', 'Amélie', 'Bobb', 'Banditt',  
 'Kevon', 'Winifred', 'Hanz', 'Churlie', 'Zeek', 'Timofy', 'Maks',  
 'Jomathan', 'Kallie', 'Marvin', 'Spark', 'Gördón', 'Jo', 'Jareld',  
 'Torque', 'Ron', 'Skittles', 'Cleopatra', 'Erik', 'Stu',  
 'Tedrick', 'Shaggy', 'Filup', 'Kial', 'Naphaniel', 'Dook', 'Hall',  
 'Philippe', 'Biden', 'Fwed', 'Genevieve', 'Joshwa', 'Timison',



```
'Bradlay', 'Pipsy', 'Clybe', 'Keet', 'Carll', 'Jockson', 'Josep',
'Lugan', 'Christoper'], dtype=object)
```

#### 0.4.8 There are row has not data for dogs should be removed.

Code

```
In [45]: # There are row has not data for dogs should be removed.
df_twitter_arch_clean = df_twitter_arch_clean[df_twitter_arch_clean.text.isnull()]
df_twitter_arch_clean = df_twitter_arch_clean.loc[~df_twitter_arch_clean.text.str.match
```

Test

```
In [46]: df_twitter_arch_clean.loc[df_twitter_arch_clean.text.str.match('.*only rate dogs')]

Out[46]: Empty DataFrame
Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text]
Index: []
```

#### 0.4.9 Source has a link should be extracted to get only source and should be changed to category.

```
In [47]: # Source has a link should be extracted to get only source and should be changed to cat
df_twitter_arch_clean.source.unique()

Out[47]: array(['<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone',
'<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>',
'<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>',
'<a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck
```

Code

```
In [48]: def find_source(x):
frist= x.find('>') + 1
secound =x[frist:].find('<')
return x[frist:][:secound]

In [49]: df_twitter_arch_clean.source = df_twitter_arch_clean.source.apply(lambda x: find_source
```

Test

```
In [50]: df_twitter_arch_clean.source.unique()

Out[50]: [Twitter for iPhone, Twitter Web Client, Vine - Make a Scene, TweetDeck]
Categories (4, object): [Twitter for iPhone, Twitter Web Client, Vine - Make a Scene, T
```

#### 0.4.10 (doggo, floofer, puppe and puppo) Columns has none or missing values.

Code

```
In [51]: dogs_types = ['doggo', 'floofer', 'pupper', 'puppo']
         # creating a list to creat the loop
         for types in dogs_types:
             df_twitter_arch_clean[dogs_types] = df_twitter_arch_clean[dogs_types].replace('None', np.nan)

In [52]: # doggo, floofer, puppe and puppo columns should be merged in one column dog_types
         dogs_type = ['doggo', 'floofer', 'pupper', 'puppo']
         df_twitter_arch_clean['dog_stage'] = df_twitter_arch_clean[dogs_type].\
             apply(lambda x: ', '.join(x.dropna()).astype(str))
         df_twitter_arch_clean.dog_stage = df_twitter_arch_clean.dog_stage.replace(' ', np.nan)
         df_twitter_arch_clean = df_twitter_arch_clean.drop(columns = dogs_type, axis =1)
```

Test

```
In [53]: df_twitter_arch_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2098 entries, 0 to 2355
Data columns (total 14 columns):
tweet_id                2098 non-null int64
in_reply_to_status_id   72 non-null float64
in_reply_to_user_id     72 non-null float64
timestamp               2098 non-null datetime64[ns]
source                 2098 non-null category
text                   2098 non-null object
retweeted_status_id      0 non-null float64
retweeted_status_user_id 0 non-null float64
retweeted_status_timestamp 0 non-null object
expanded_urls           2098 non-null object
rating_numerator         2098 non-null float64
rating_denominator       2098 non-null int64
name                   1384 non-null object
dog_stage               344 non-null object
dtypes: category(1), datetime64[ns](1), float64(5), int64(2), object(5)
memory usage: 231.7+ KB
```

#### 0.4.11 Removing all retweet columns.

Code

```
In [54]: # Removing all retweet columns.
         dropped_columns = ['retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timestamp']
         df_twitter_arch_clean.drop(columns = dropped_columns, axis=1, inplace=True)
```

Test

```
In [55]: list(df_twitter_arch_clean.columns)
```

```
Out[55]: ['tweet_id',
          'timestamp',
          'source',
          'text',
          'expanded_urls',
          'rating_numerator',
          'rating_denominator',
          'name',
          'dog_stage']
```

#### 0.4.12 Cleaning Image Predictions from ( df\_img\_pred to df\_img\_pred\_clean )

#### 0.4.13 There is 66 duplicated url in jpg\_url should be removed.

```
In [56]: # There is 66 duplicated url in jpg_url should be removed.
df_img_pred_clean = df_img_pred_clean.drop_duplicates(subset=['jpg_url'], keep='last')
```

Test

```
In [57]: df_img_pred_clean.jpg_url.duplicated().sum()
```

```
Out[57]: 0
```

#### 0.4.14 p1\_conf, p2\_conf and p3\_conf columns should be merged in one column confidence and p1\_dog, p2\_dog and p3\_dog columns should be merged in one column dog\_type.

Code

```
In [58]: # p1_conf, p2_conf and p3_conf columns should be merged in one column confidence and p1_dog, p2_dog and p3_dog columns should be merged in one column dog_type.

dog_type = []
confidence_list = []
def image(df_img_pred_clean):
    if df_img_pred_clean['p1_dog'] == True:
        dog_type.append(df_img_pred_clean['p1'])
        confidence_list.append(df_img_pred_clean['p1_conf'])
    elif df_img_pred_clean['p2_dog'] == True:
        dog_type.append(df_img_pred_clean['p2'])
        confidence_list.append(df_img_pred_clean['p2_conf'])
    elif df_img_pred_clean['p3_dog'] == True:
        dog_type.append(df_img_pred_clean['p3'])
        confidence_list.append(df_img_pred_clean['p3_conf'])
    else:
        dog_type.append('Error')
        confidence_list.append('Error')

df_img_pred_clean.apply(image, axis=1)
```

```
df_img_pred_clean['dog_type'] = dog_type
df_img_pred_clean['confidence'] = confidence_list
```

```
In [59]: # Cleaning errors
```

```
df_img_pred_clean = df_img_pred_clean[df_img_pred_clean['dog_type'] != 'Error']
```

Test

```
In [60]: df_img_pred_clean.dog_type.unique()
```

```
Out[60]: array(['Welsh_springer_spaniel', 'redbone', 'German_shepherd',
                'Rhodesian_ridgeback', 'miniature_pinscher', 'Bernese_mountain_dog',
                'chow', 'golden_retriever', 'miniature_poodle', 'Gordon_setter',
                'Walker_hound', 'pug', 'bloodhound', 'Lhasa', 'English_setter',
                'Italian_greyhound', 'Maltese_dog', 'Newfoundland', 'malamute',
                'soft-coated_wheaten_terrier', 'Chihuahua',
                'black-and-tan_coonhound', 'toy_terrier', 'Blenheim_spaniel',
                'Pembroke', 'Irish_terrier', 'Chesapeake_Bay_retriever',
                'curly-coated_retriever', 'dalmatian', 'Ibizan_hound',
                'Border_collie', 'Labrador_retriever', 'miniature_schnauzer',
                'Airedale', 'Rottweiler', 'West_Highland_white_terrier',
                'toy_poodle', 'giant_schnauzer', 'vizsla', 'Siberian_husky',
                'papillon', 'Saint_Bernard', 'Tibetan_terrier', 'borzoi',
                'Yorkshire_terrier', 'Pomeranian', 'kuvasz',
                'flat-coated_retriever', 'Norwegian_elkhound', 'boxer',
                'Eskimo_dog', 'standard_poodle', 'Staffordshire_bullterrier',
                'basenji', 'Lakeland_terrier', 'American_Staffordshire_terrier',
                'Shih-Tzu', 'groenendael', 'French_bulldog', 'Pekinese', 'komondor',
                'beagle', 'malinois', 'kelpie', 'Brittany_spaniel',
                'cocker_spaniel', 'basset', 'English_springer', 'Cardigan',
                'Brabancon_griffon', 'German_short-haired_pointer',
                'Shetland_sheepdog', 'cairn', 'whippet', 'Sussex_spaniel',
                'Dandie_Dinmont', 'Norwich_terrier', 'keeshond', 'Norfolk_terrier',
                'Old_English_sheepdog', 'Samoyed', 'Scottish_deerhound', 'Doberman',
                'Irish_wolfhound', 'Great_Pyrenees', 'schipperke', 'bull_mastiff',
                'collie', 'Greater_Swiss_Mountain_dog', 'standard_schnauzer',
                'Irish_water_spaniel', 'Boston_bull', 'Japanese_spaniel',
                'Bedlington_terrier', 'EntleBucher', 'bluetick', 'Irish_setter',
                'Leonberg', 'Mexican_hairless', 'Weimaraner', 'Great_Dane',
                'Tibetan_mastiff', 'Scotch_terrier', 'Australian_terrier', 'briard',
                'Appenzeller', 'Border_terrier', 'wire-haired_fox_terrier',
                'Saluki', 'silky_terrier', 'Afghan_hound', 'clumber',
                'Bouvier_des_Flandres'], dtype=object)
```

```
In [61]: df_img_pred_clean.confidence.unique()
```

```
Out[61]: array([0.465074, 0.506826, 0.596461, ..., 0.168086, 0.716012, 0.323581], dtype=object)
```

#### 0.4.15 img\_num is useless and should be removed and all unnecessary columns.

Code

```
In [62]: # img_num is useless and should be removed and all unnecessary columns.
df_img_pred_clean = df_img_pred_clean.drop(['img_num',
      'p1',
      'p1_conf',
      'p1_dog',
      'p2',
      'p2_conf',
      'p2_dog',
      'p3',
      'p3_conf',
      'p3_dog'], 1)
```

Test

```
In [63]: list(df_img_pred_clean.columns)

Out[63]: ['tweet_id', 'jpg_url', 'dog_type', 'confidence']

In [64]: df_img_pred_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1691 entries, 0 to 2073
Data columns (total 4 columns):
tweet_id      1691 non-null int64
jpg_url       1691 non-null object
dog_type      1691 non-null object
confidence    1691 non-null object
dtypes: int64(1), object(3)
memory usage: 66.1+ KB
```

#### 0.4.16 Cleaning Twitter API from ( df\_api to df\_api\_clean )

#### 0.4.17 id column should be renamed to tweet\_it

Code

```
In [65]: # id column should be renamed to tweet_it
df_api_clean = df_api_clean.rename(columns={'id': 'tweet_id'})
```

Test

```
In [66]: list(df_api_clean.columns)
```

```
Out[66]: ['contributors',
          'coordinates',
          'created_at',
          'display_text_range',
          'entities',
          'extended_entities',
          'favorite_count',
          'favorited',
          'full_text',
          'geo',
          'tweet_id',
          'id_str',
          'in_reply_to_screen_name',
          'in_reply_to_status_id',
          'in_reply_to_status_id_str',
          'in_reply_to_user_id',
          'in_reply_to_user_id_str',
          'is_quote_status',
          'lang',
          'place',
          'possibly_sensitive',
          'possibly_sensitive_appealable',
          'quoted_status',
          'quoted_status_id',
          'quoted_status_id_str',
          'retweet_count',
          'retweeted',
          'retweeted_status',
          'source',
          'truncated',
          'user']
```

**0.4.18 All columns are useless except (id, favorite\_count and retweet\_count) and should be removed.**

Code

```
In [67]: # All columns are useless except (id, favorite_count and retweet_count) and should be removed
df_api_clean = df_api_clean.drop(['created_at',
                                   'id_str',
                                   'full_text',
                                   'truncated',
                                   'display_text_range',
                                   'entities',
                                   'extended_entities',
                                   'source',
                                   'in_reply_to_status_id',
                                   'in_reply_to_status_id_str',
```

```

'in_reply_to_user_id',
'in_reply_to_user_id_str',
'in_reply_to_screen_name',
'user',
'geo',
'coordinates',
'place',
'contributors',
'is_quote_status', 'favorited',
'retweeted',
'possibly_sensitive',
'possibly_sensitive_appealable',
'lang',
'retweeted_status',
'quoted_status_id',
'quoted_status_id_str',
'quoted_status'], 1)

```

Test

```
In [68]: list(df_api_clean.columns)
```

```
Out[68]: ['favorite_count', 'tweet_id', 'retweet_count']
```

```
In [69]: df_api_clean.info()
```

```

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

```

## 0.5 Storing and Acting on Wrangled Data

```

In [70]: # Merging all data frames in one data frame (df_twitter_arch_clean, df_img_pred_clean and
df_master = pd.merge(df_twitter_arch_clean, df_img_pred_clean, on='tweet_id')
df_master = pd.merge(df_master, df_api_clean, on='tweet_id')

```

```
In [71]: df_master.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1561 entries, 0 to 1560
Data columns (total 14 columns):
tweet_id          1561 non-null int64
timestamp         1561 non-null datetime64[ns]

```

```

source          1561 non-null category
text            1561 non-null object
expanded_urls   1561 non-null object
rating_numerator 1561 non-null float64
rating_denominator 1561 non-null int64
name            1136 non-null object
dog_stage       251 non-null object
jpg_url         1561 non-null object
dog_type        1561 non-null object
confidence      1561 non-null object
favorite_count  1561 non-null int64
retweet_count   1561 non-null int64
dtypes: category(1), datetime64[ns](1), float64(1), int64(4), object(7)
memory usage: 172.4+ KB

```

### 0.5.1 Save new dataframe to twitter\_archive\_master.csv

```

In [72]: df_master.to_csv('twitter_archive_master.csv',
                          index=False, encoding = 'utf-8')

```

```

In [73]: df_master.head()

```

```

Out[73]:
   tweet_id      timestamp      source \
0  892177421306343426  2017-08-01 00:17:27  Twitter for iPhone
1  891815181378084864  2017-07-31 00:18:03  Twitter for iPhone
2  891689557279858688  2017-07-30 15:58:51  Twitter for iPhone
3  891327558926688256  2017-07-29 16:00:24  Twitter for iPhone
4  891087950875897856  2017-07-29 00:08:17  Twitter for iPhone

   text \
0  This is Tilly. She's just checking pup on you...
1  This is Archie. He is a rare Norwegian Pouncin...
2  This is Darla. She commenced a snooze mid meal...
3  This is Franklin. He would like you to stop ca...
4  Here we have a majestic great white breaching ...

   expanded_urls  rating_numerator \
0  https://twitter.com/dog_rates/status/892177421...  13.0
1  https://twitter.com/dog_rates/status/891815181...  12.0
2  https://twitter.com/dog_rates/status/891689557...  13.0
3  https://twitter.com/dog_rates/status/891327558...  12.0
4  https://twitter.com/dog_rates/status/891087950...  13.0

   rating_denominator  name  dog_stage \
0                    10   Tilly      NaN
1                    10  Archie      NaN
2                    10  Darla      NaN

```



|   |    |          |     |
|---|----|----------|-----|
| 3 | 10 | Franklin | NaN |
| 4 | 10 | NaN      | NaN |

|   | jpg_url   | dog_type \               |
|---|---|--------------------------|
| 0 | https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg | Chihuahua                |
| 1 | https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg | Chihuahua                |
| 2 | https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg | Labrador_retriever       |
| 3 | https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg | basset                   |
| 4 | https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg | Chesapeake_Bay_retriever |

|   | confidence | favorite_count | retweet_count |
|---|------------|----------------|---------------|
| 0 | 0.323581   | 33819          | 6514          |
| 1 | 0.716012   | 25461          | 4328          |
| 2 | 0.168086   | 42908          | 8964          |
| 3 | 0.555712   | 41048          | 9774          |
| 4 | 0.425595   | 20562          | 3261          |

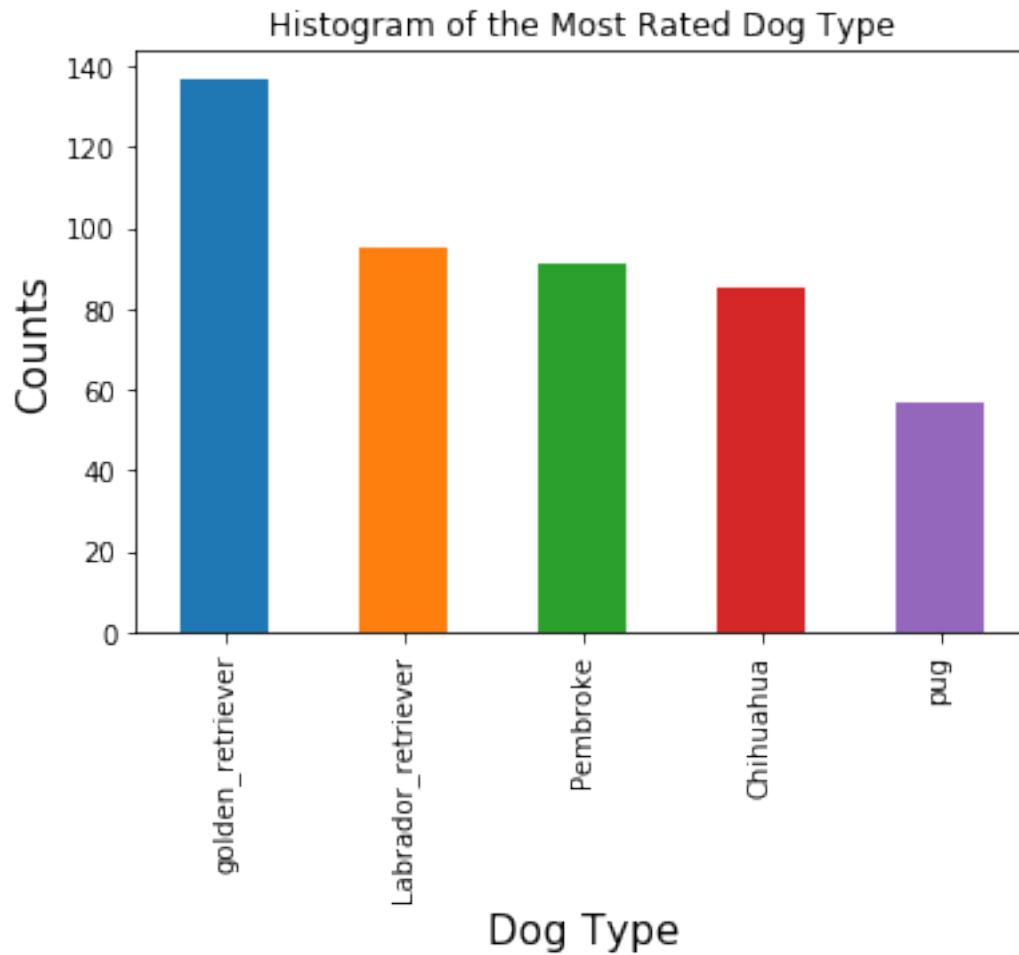
## 0.6 Visualizations

### 0.6.1 Top 5 Dog Types

```
In [74]: df_dog_type = df_master.groupby('dog_type').filter(lambda x: len(x) >= 55)
```

```
df_dog_type['dog_type'].value_counts().plot(kind = 'bar')
plt.title('Histogram of the Most Rated Dog Type')
plt.xlabel('Dog Type' , fontsize=15)
plt.ylabel('Counts', fontsize=15)

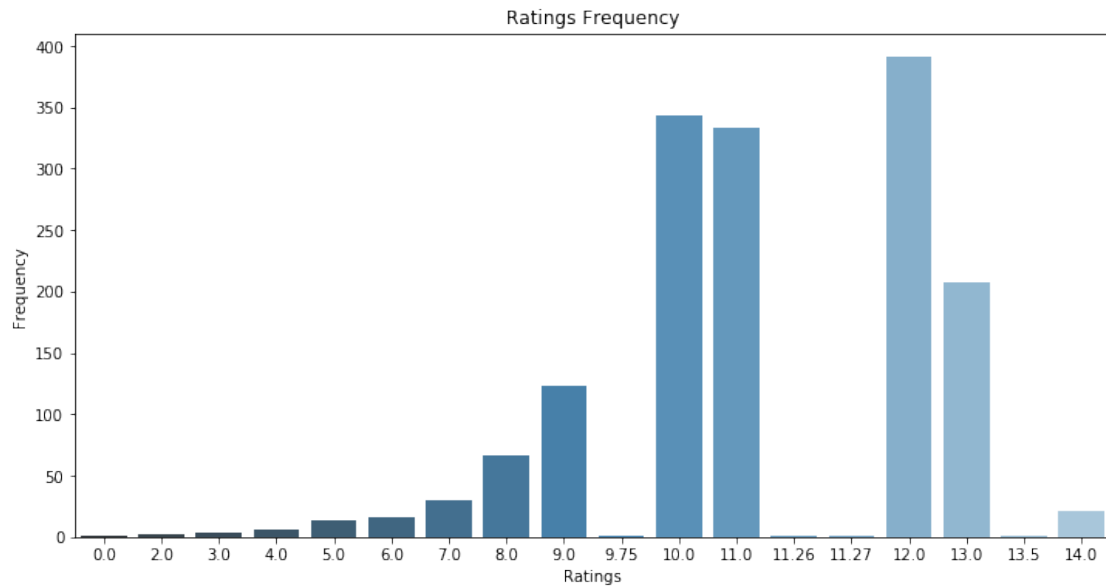
fig = plt.gcf()
fig.savefig('output.png',bbox_inches='tight');
```



## 0.6.2 Ratings Frequency

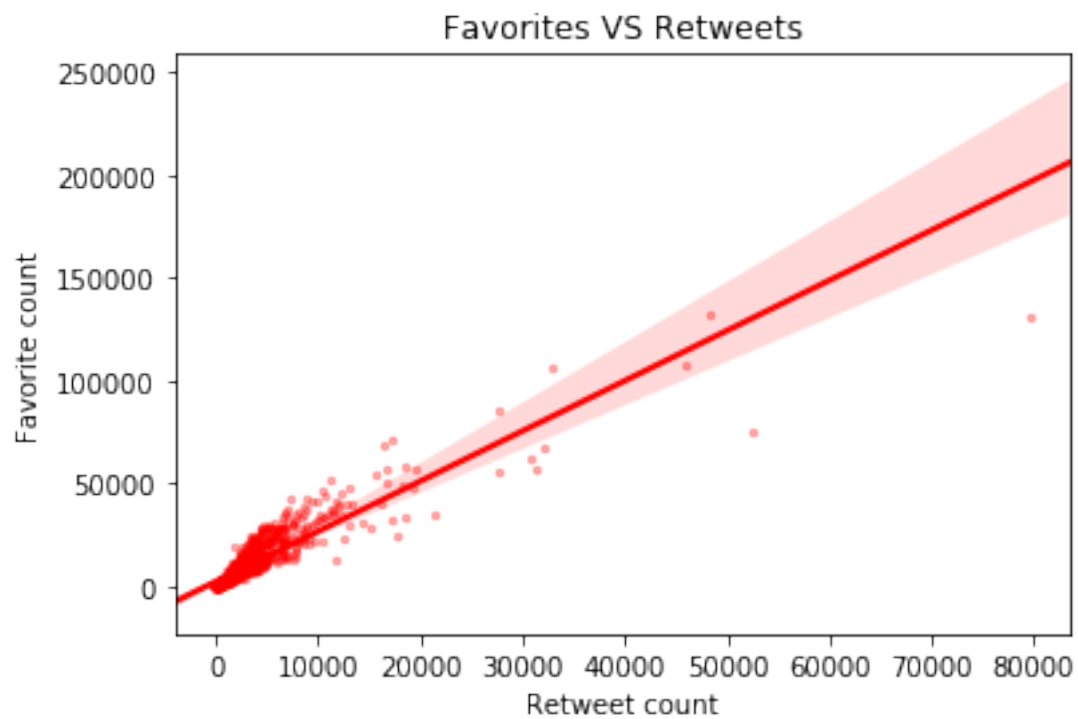
```
In [75]: data = df_master.rating_numerator.value_counts()
```

```
x = data.index  
y = data.values  
fig, mn = plt.subplots(figsize=(12, 6))  
g = sns.barplot(x, y, palette='Blues_d', ax=mn)  
mn.set(xlabel='Ratings', ylabel='Frequency', title='Ratings Frequency')  
plt.show()
```



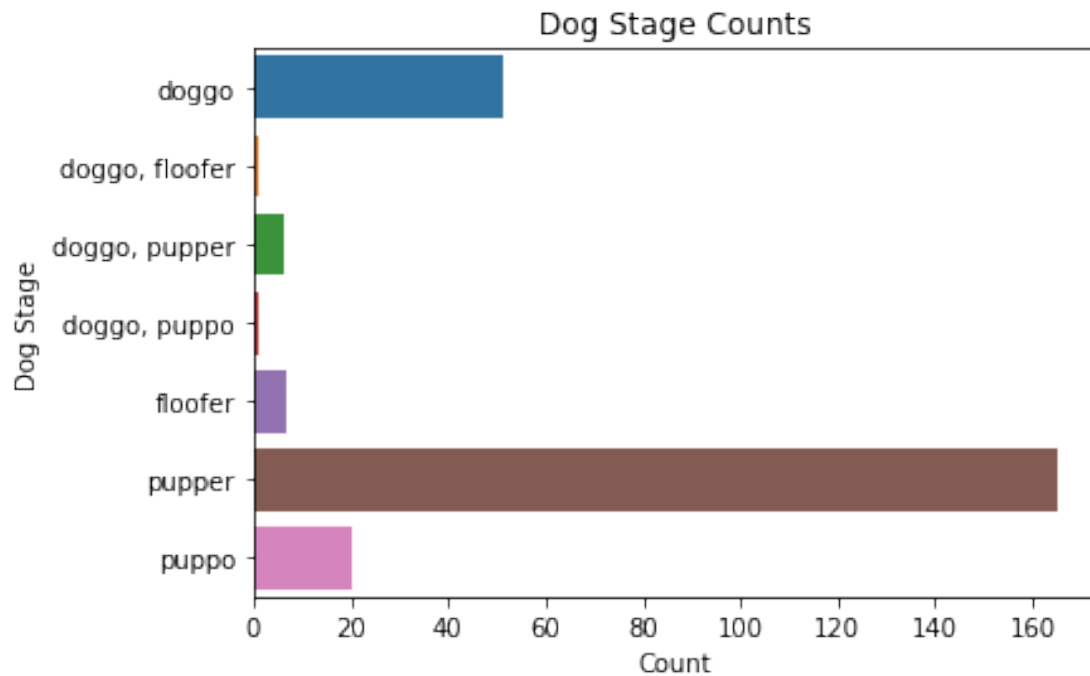
### 0.6.3 Chart Between Favorites Count and Retweet Count

```
In [76]: mn = sns.regplot(x='retweet_count', y='favorite_count', data=df_master, color='r', scatter_kws={'color': 'r'})
mn.set(xlabel='Retweet count', ylabel='Favorite count', title='Favorites VS Retweets')
plt.show()
```



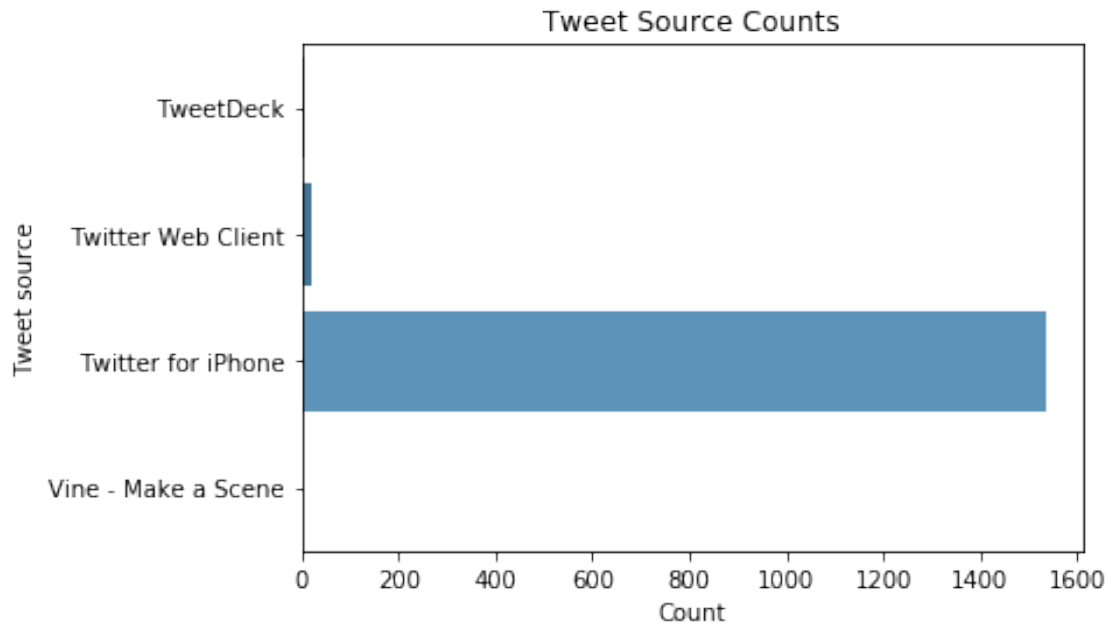
#### 0.6.4 Dogs Stage Count

```
In [77]: data = df_master.groupby('dog_stage').count()['tweet_id']
mn = sns.barplot(y=data.index, x=data.values)
mn.set(xlabel='Count', ylabel='Dog Stage', title='Dog Stage Counts')
plt.show()
```



#### 0.6.5 Chart to explain the most used source

```
In [78]: data = df_master.groupby('source').count()['tweet_id']
mn = sns.barplot(y=data.index, x=data.values, palette='Blues_d')
mn.set(xlabel='Count', ylabel='Tweet source', title='Tweet Source Counts')
plt.show()
```



### 0.6.6 Top 5 Frequent Names

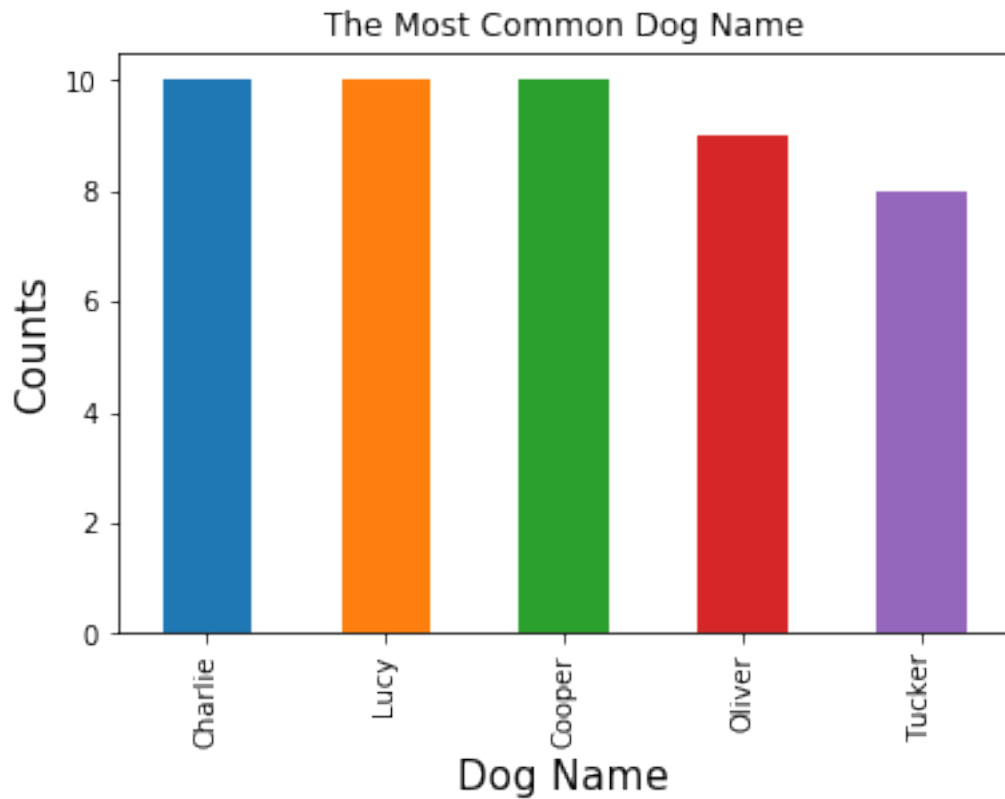
```
In [79]: n = 5
         top_five_names = df_master['name'].value_counts()[:n].index.tolist()
         top_five_names

Out[79]: ['Cooper', 'Charlie', 'Lucy', 'Oliver', 'Tucker']

In [80]: df_dog_name = df_master.groupby('name').filter(lambda x: len(x) >= 8)

         df_dog_name['name'].value_counts().plot(kind = 'bar')
         plt.title('The Most Common Dog Name')
         plt.xlabel('Dog Name' , fontsize=15)
         plt.ylabel('Counts', fontsize=15)

         fig = plt.gcf()
         fig.savefig('output.png',bbox_inches='tight');
```



### 0.6.7 Top Retweeted Dog

```
In [81]: # Getting the max retweet count
df_master.retweet_count.describe()
```

```
Out[81]: count      1561.000000
         mean       2502.057655
         std       4279.890452
         min        16.000000
         25%        611.000000
         50%       1328.000000
         75%       2915.000000
         max       79515.000000
         Name: retweet_count, dtype: float64
```

```
In [82]: # Getting the top retweeted index
df_master.query('retweet_count == 79515 ')
```

```
Out[82]:      tweet_id      timestamp      source \
626  744234799360020481  2016-06-18 18:26:18  Twitter for iPhone

         text \
```

```

626 Here's a doggo realizing you can stand in a po...

                                expanded_urls rating_numerator \
626 https://twitter.com/dog_rates/status/744234799...          13.0

    rating_denominator name dog_stage \
626                10 NaN      doggo

                                jpg_url      dog_type \
626 https://pbs.twimg.com/ext_tw_video_thumb/74423... Labrador_retriever

    confidence favorite_count retweet_count
626  0.825333      131075      79515

```

```

In [83]: # Displaying the top retweeted dog photo
ind = 626
url = df_master.iloc[ind, df_master.columns.get_loc('jpg_url')]
img = IPython.display.Image(url)
img

```

Out[83]:





```
In [84]: # Displaying the top retweeted dog tweet text
df_master.iloc[ind, df_master.columns.get_loc('text')]
```

```
Out[84]: "Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Co
```

## 0.6.8 Top Favorite Dog

```
In [85]: # Getting the max favorite count
df_master.favorite_count.describe()
```

```
Out[85]: count      1561.000000
mean        8427.795003
std         11555.136731
min           81.000000
25%         2040.000000
50%         4065.000000
75%        10512.000000
max        132810.000000
Name: favorite_count, dtype: float64
```

```
In [86]: # Getting the top favorite index
df_master.query('favorite_count == 132810 ')
```

```
Out[86]:
```

|     | tweet_id           | timestamp           | source             | text  | expanded_urls                                     | rating_numerator | rating_denominator | name | dog_stage | jpg_url   | dog_type         | confidence | favorite_count | retweet_count |
|-----|--------------------|---------------------|--------------------|---|---|------------------|--------------------|------|-----------|---|------------------|------------|----------------|---------------|
| 260 | 822872901745569793 | 2017-01-21 18:26:02 | Twitter for iPhone | Here's a super supportive puppo participating ... | https://twitter.com/dog_rates/status/822872901... | 13.0             | 10                 | NaN  | puppo     | https://pbs.twimg.com/media/C2tugXLXgAArJ04.jpg | Lakeland_terrier | 0.196015   | 132810         | 48265         |

```
In [87]: # Displaying the top favorite dog photo
ind=260
url = df_master.iloc[ind, df_master.columns.get_loc('jpg_url')]
img = IPython.display.Image(url)
img
```

Out [87] :



```
In [88]: # Displaying the top favorite dog tweet text
         df_master.iloc[ind, df_master.columns.get_loc('text')]
```

Out[88]: "Here's a super supportive puppo participating in the Toronto #WomensMarch today. 13/1

Thank You  
Created By:  
Mahmoud Hassan Khalil El-Tobgy

In [ ]: