

wrangle_act1

January 12, 2021

1 Wranlge data WeRateDogs

1. Gathering
2. Assessing
3. Cleaning
4. Storing, Anlyzing
5. Visualdze data
6. Reporting

2 Importing libraries and data

```
In [373]: import pandas as pd
import numpy as np
import requests
import tweepy
import re
import json
import os
from os import path
import time
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
import seaborn as sns
from scipy import stats
from IPython.display import HTML, display
from PIL import Image
from wordcloud import wordcloud, STOPWORDS, ImageColorGenerator
```

importing data and reading csv file

```
In [374]: twitter_arch=pd.read_csv('twitter-archive-enhanced.csv')
twitter_arch.head()
```

```
Out[374]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	NaN	NaN	
1	892177421306343426	NaN	NaN	

2	891815181378084864	NaN	NaN
3	891689557279858688	NaN	NaN
4	891327558926688256	NaN	NaN

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

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

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

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

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

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

importing grammatically tsv using the file and Requests image library predictions and the downloaded following pro-URL:

https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv

```
In [375]: url='https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv'
r=requests.get(url)
with open('image-predictions.tsv',mode='wb') as f:
    f.write(r.content)
img_pred=pd.read_csv('image-predictions.tsv',sep='\t')
```

```
In [376]: img_pred.head()
```

```
Out[376]:
```

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

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

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

3 JSON file and getting data from Twitter API

```
In [377]: '''
import tweepy
from tweepy import OAuthHandler
import json
from timeit import default_timer as timer

# Query Twitter API for each tweet in the Twitter archive and save JSON in a text file
# These are hidden to comply with Twitter's API terms and conditions
consumer_key = 'HIDDEN'
consumer_secret = 'HIDDEN'
access_token = 'HIDDEN'
access_secret = 'HIDDEN'

auth = OAuthHandler(consumer_key, consumer_secret)
```

```

auth.set_access_token(access_token, access_secret)

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

# NOTE TO STUDENT WITH MOBILE VERIFICATION ISSUES:
# df_1 is a DataFrame with the twitter_archive_enhanced.csv file. You may have to
# change line 17 to match the name of your DataFrame with twitter_archive_enhanced.csv
# NOTE TO REVIEWER: this student had mobile verification issues so the following
# Twitter API code was sent to this student from a Udacity instructor
# Tweet IDs for which to gather additional data via Twitter's API
tweet_ids = df_1.tweet_id.values
len(tweet_ids)

# Query Twitter's API for JSON data for each tweet ID in the Twitter archive
count = 0
fails_dict = {}
start = timer()
# Save each tweet's returned JSON as a new line in a .txt file
with open('tweet_json.txt', 'w') as outfile:
    # This loop will likely take 20-30 minutes to run because of Twitter's rate limit
    for tweet_id in tweet_ids:
        count += 1
        print(str(count) + ": " + str(tweet_id))
        try:
            tweet = api.get_status(tweet_id, tweet_mode='extended')
            print("Success")
            json.dump(tweet._json, outfile)
            outfile.write('\n')
        except tweepy.TweepError as e:
            print("Fail")
            fails_dict[tweet_id] = e
            pass
    end = timer()
    print(end - start)
    print(fails_dict)
'''

```

Out[377]: '\nimport tweepy\nfrom tweepy import OAuthHandler\nimport json\nfrom timeit import def

Data Frame creation txt file

```

In [378]: tweets_list=[]
          with open('tweet-json.txt') as file:
              for line in file:
                  try:
                      tweet=json.loads(line)
                      tweets_list.append(tweet)
                  except:

```

```

        continue
df_tweets=pd.DataFrame(tweets_list, columns=list(tweets_list[0].keys()))

In [379]: df_tweets.head()

Out[379]:
```

		created_at		id	id_str \
0	Tue Aug 01 16:23:56 +0000 2017	892420643555336193	892420643555336193		
1	Tue Aug 01 00:17:27 +0000 2017	892177421306343426	892177421306343426		
2	Mon Jul 31 00:18:03 +0000 2017	891815181378084864	891815181378084864		
3	Sun Jul 30 15:58:51 +0000 2017	891689557279858688	891689557279858688		
4	Sat Jul 29 16:00:24 +0000 2017	891327558926688256	891327558926688256		

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

	display_text_range	entities \
0	[0, 85]	{'hashtags': [], 'symbols': [], 'user_mentions...
1	[0, 138]	{'hashtags': [], 'symbols': [], 'user_mentions...
2	[0, 121]	{'hashtags': [], 'symbols': [], 'user_mentions...
3	[0, 79]	{'hashtags': [], 'symbols': [], 'user_mentions...
4	[0, 138]	{'hashtags': [{'text': 'BarkWeek', 'indices': ...

	extended_entities \
0	{'media': [{'id': 892420639486877696, 'id_str'...
1	{'media': [{'id': 892177413194625024, 'id_str'...
2	{'media': [{'id': 891815175371796480, 'id_str'...
3	{'media': [{'id': 891689552724799489, 'id_str'...
4	{'media': [{'id': 891327551943041024, 'id_str'...

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

	...	place	contributors	is_quote_status	retweet_count	favorite_count \
0	...	None	None	False	8853	39467
1	...	None	None	False	6514	33819
2	...	None	None	False	4328	25461
3	...	None	None	False	8964	42908
4	...	None	None	False	9774	41048

	favorited	retweeted	possibly_sensitive	possibly_sensitive_appealable	lang
--	-----------	-----------	--------------------	-------------------------------	------

0	False	False	False	False	en
1	False	False	False	False	en
2	False	False	False	False	en
3	False	False	False	False	en
4	False	False	False	False	en

[5 rows x 27 columns]

```
In [380]: df_twitter= df_tweets[['id','retweet_count','favorite_count']]
```

```
In [381]: df_twitter.head()
```

```
Out[381]:
```

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

4 Assessing

inspecting twitter data

```
In [382]: twitter_arch
```

```
Out[382]:
```

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

22	887517139158093824	NaN	NaN
23	887473957103951883	NaN	NaN
24	887343217045368832	NaN	NaN
25	887101392804085760	NaN	NaN
26	886983233522544640	NaN	NaN
27	886736880519319552	NaN	NaN
28	886680336477933568	NaN	NaN
29	886366144734445568	NaN	NaN
...
2326	666411507551481857	NaN	NaN
2327	666407126856765440	NaN	NaN
2328	666396247373291520	NaN	NaN
2329	666373753744588802	NaN	NaN
2330	666362758909284353	NaN	NaN
2331	666353288456101888	NaN	NaN
2332	666345417576210432	NaN	NaN
2333	666337882303524864	NaN	NaN
2334	666293911632134144	NaN	NaN
2335	666287406224695296	NaN	NaN
2336	666273097616637952	NaN	NaN
2337	666268910803644416	NaN	NaN
2338	666104133288665088	NaN	NaN
2339	666102155909144576	NaN	NaN
2340	666099513787052032	NaN	NaN
2341	666094000022159362	NaN	NaN
2342	666082916733198337	NaN	NaN
2343	666073100786774016	NaN	NaN
2344	666071193221509120	NaN	NaN
2345	666063827256086533	NaN	NaN
2346	666058600524156928	NaN	NaN
2347	666057090499244032	NaN	NaN
2348	666055525042405380	NaN	NaN
2349	666051853826850816	NaN	NaN
2350	666050758794694657	NaN	NaN
2351	666049248165822465	NaN	NaN
2352	666044226329800704	NaN	NaN
2353	666033412701032449	NaN	NaN
2354	666029285002620928	NaN	NaN
2355	666020888022790149	NaN	NaN

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

7	2017-07-28	00:22:40	+0000
8	2017-07-27	16:25:51	+0000
9	2017-07-26	15:59:51	+0000
10	2017-07-26	00:31:25	+0000
11	2017-07-25	16:11:53	+0000
12	2017-07-25	01:55:32	+0000
13	2017-07-25	00:10:02	+0000
14	2017-07-24	17:02:04	+0000
15	2017-07-24	00:19:32	+0000
16	2017-07-23	00:22:39	+0000
17	2017-07-22	16:56:37	+0000
18	2017-07-22	00:23:06	+0000
19	2017-07-21	01:02:36	+0000
20	2017-07-20	16:49:33	+0000
21	2017-07-19	16:06:48	+0000
22	2017-07-19	03:39:09	+0000
23	2017-07-19	00:47:34	+0000
24	2017-07-18	16:08:03	+0000
25	2017-07-18	00:07:08	+0000
26	2017-07-17	16:17:36	+0000
27	2017-07-16	23:58:41	+0000
28	2017-07-16	20:14:00	+0000
29	2017-07-15	23:25:31	+0000
...			...
2326	2015-11-17	00:24:19	+0000
2327	2015-11-17	00:06:54	+0000
2328	2015-11-16	23:23:41	+0000
2329	2015-11-16	21:54:18	+0000
2330	2015-11-16	21:10:36	+0000
2331	2015-11-16	20:32:58	+0000
2332	2015-11-16	20:01:42	+0000
2333	2015-11-16	19:31:45	+0000
2334	2015-11-16	16:37:02	+0000
2335	2015-11-16	16:11:11	+0000
2336	2015-11-16	15:14:19	+0000
2337	2015-11-16	14:57:41	+0000
2338	2015-11-16	04:02:55	+0000
2339	2015-11-16	03:55:04	+0000
2340	2015-11-16	03:44:34	+0000
2341	2015-11-16	03:22:39	+0000
2342	2015-11-16	02:38:37	+0000
2343	2015-11-16	01:59:36	+0000
2344	2015-11-16	01:52:02	+0000
2345	2015-11-16	01:22:45	+0000
2346	2015-11-16	01:01:59	+0000
2347	2015-11-16	00:55:59	+0000
2348	2015-11-16	00:49:46	+0000
2349	2015-11-16	00:35:11	+0000

2350 2015-11-16 00:30:50 +0000
2351 2015-11-16 00:24:50 +0000
2352 2015-11-16 00:04:52 +0000
2353 2015-11-15 23:21:54 +0000
2354 2015-11-15 23:05:30 +0000
2355 2015-11-15 22:32:08 +0000

```

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

2335 <a href="http://twitter.com/download/iphone" r...
 2336 <a href="http://twitter.com/download/iphone" r...
 2337 <a href="http://twitter.com/download/iphone" r...
 2338 <a href="http://twitter.com/download/iphone" r...
 2339 <a href="http://twitter.com/download/iphone" r...
 2340 <a href="http://twitter.com/download/iphone" r...
 2341 <a href="http://twitter.com/download/iphone" r...
 2342 <a href="http://twitter.com/download/iphone" r...
 2343 <a href="http://twitter.com/download/iphone" r...
 2344 <a href="http://twitter.com/download/iphone" r...
 2345 <a href="http://twitter.com/download/iphone" r...
 2346 <a href="http://twitter.com/download/iphone" r...
 2347 <a href="http://twitter.com/download/iphone" r...
 2348 <a href="http://twitter.com/download/iphone" r...
 2349 <a href="http://twitter.com/download/iphone" r...
 2350 <a href="http://twitter.com/download/iphone" r...
 2351 <a href="http://twitter.com/download/iphone" r...
 2352 <a href="http://twitter.com/download/iphone" r...
 2353 <a href="http://twitter.com/download/iphone" r...
 2354 <a href="http://twitter.com/download/iphone" r...
 2355 <a href="http://twitter.com/download/iphone" r...

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

25	This... is a Jubilant Antarctic House Bear. We...	NaN
26	This is Maya. She's very shy. Rarely leaves he...	NaN
27	This is Mingus. He's a wonderful father to his...	NaN
28	This is Derek. He's late for a dog meeting. 13...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
...
2326	This is quite the dog. Gets really excited whe...	NaN
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN
2328	Oh goodness. A super rare northeast Qdoba kang...	NaN
2329	Those are sunglasses and a jean jacket. 11/10 ...	NaN
2330	Unique dog here. Very small. Lives in containe...	NaN
2331	Here we have a mixed Asiago from the Galápagos...	NaN
2332	Look at this jokester thinking seat belt laws ...	NaN
2333	This is an extremely rare horned Parthenon. No...	NaN
2334	This is a funny dog. Weird toes. Won't come do...	NaN
2335	This is an Albanian 3 1/2 legged Episcopalian...	NaN
2336	Can take selfies 11/10 https://t.co/ws2AMaWpW	NaN
2337	Very concerned about fellow dog trapped in com...	NaN
2338	Not familiar with this breed. No tail (weird)...	NaN
2339	Oh my. Here you are seeing an Adobe Setter giv...	NaN
2340	Can stand on stump for what seems like a while...	NaN
2341	This appears to be a Mongolian Presbyterian mi...	NaN
2342	Here we have a well-established sunblockerspan...	NaN
2343	Let's hope this flight isn't Malaysian (lol). ...	NaN
2344	Here we have a northern speckled Rhododendron...	NaN
2345	This is the happiest dog you will ever see. Ve...	NaN
2346	Here is the Rand Paul of retrievers folks! He'...	NaN
2347	My oh my. This is a rare blond Canadian terrie...	NaN
2348	Here is a Siberian heavily armored polar bear ...	NaN
2349	This is an odd dog. Hard on the outside but lo...	NaN
2350	This is a truly beautiful English Wilson Staff...	NaN
2351	Here we have a 1949 1st generation vulpix. Enj...	NaN
2352	This is a purebred Piers Morgan. Loves to Netf...	NaN
2353	Here is a very happy pup. Big fan of well-main...	NaN
2354	This is a western brown Mitsubishi terrier. Up...	NaN
2355	Here we have a Japanese Irish Setter. Lost eye...	NaN

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

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

2353	NaN	NaN
2354	NaN	NaN
2355	NaN	NaN

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

2338	https://twitter.com/dog_rates/status/666104133...	1
2339	https://twitter.com/dog_rates/status/666102155...	11
2340	https://twitter.com/dog_rates/status/666099513...	8
2341	https://twitter.com/dog_rates/status/666094000...	9
2342	https://twitter.com/dog_rates/status/666082916...	6
2343	https://twitter.com/dog_rates/status/666073100...	10
2344	https://twitter.com/dog_rates/status/666071193...	9
2345	https://twitter.com/dog_rates/status/666063827...	10
2346	https://twitter.com/dog_rates/status/666058600...	8
2347	https://twitter.com/dog_rates/status/666057090...	9
2348	https://twitter.com/dog_rates/status/666055525...	10
2349	https://twitter.com/dog_rates/status/666051853...	2
2350	https://twitter.com/dog_rates/status/666050758...	10
2351	https://twitter.com/dog_rates/status/666049248...	5
2352	https://twitter.com/dog_rates/status/666044226...	6
2353	https://twitter.com/dog_rates/status/666033412...	9
2354	https://twitter.com/dog_rates/status/666029285...	7
2355	https://twitter.com/dog_rates/status/666020888...	8

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

28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None
2343	10	None	None	None	None	None
2344	10	None	None	None	None	None
2345	10	the	None	None	None	None
2346	10	the	None	None	None	None
2347	10	a	None	None	None	None
2348	10	a	None	None	None	None
2349	10	an	None	None	None	None
2350	10	a	None	None	None	None
2351	10	None	None	None	None	None
2352	10	a	None	None	None	None
2353	10	a	None	None	None	None
2354	10	a	None	None	None	None
2355	10	None	None	None	None	None

[2356 rows x 17 columns]

In [383]: 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 [384]: twitter_arch.describe()
```

```

Out[384]:
      tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
count  2.356000e+03                7.800000e+01          7.800000e+01
mean    7.427716e+17                7.455079e+17          2.014171e+16
std     6.856705e+16                7.582492e+16          1.252797e+17
min     6.660209e+17                6.658147e+17          1.185634e+07
25%    6.783989e+17                6.757419e+17          3.086374e+08
50%    7.196279e+17                7.038708e+17          4.196984e+09
75%    7.993373e+17                8.257804e+17          4.196984e+09
max     8.924206e+17                8.862664e+17          8.405479e+17

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

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

```

```
In [385]: twitter_arch['source'].value_counts()
```

```

Out[385]:
<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>

```



```
<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
<a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
Name: source, dtype: int64
```

```
In [386]: twitter_arch['rating_numerator'].value_counts()
```

```
Out[386]: 12      558
          11      464
          10      461
          13      351
           9      158
           8      102
           7       55
          14       54
           5       37
           6       32
           3       19
           4       17
           1        9
           2        9
          420        2
           0        2
          15        2
          75        2
          80        1
          20        1
          24        1
          26        1
          44        1
          50        1
          60        1
          165        1
          84        1
          88        1
          144        1
          182        1
          143        1
          666        1
          960        1
          1776       1
           17        1
          27        1
          45        1
          99        1
          121        1
          204        1
          Name: rating_numerator, dtype: int64
```

```
In [387]: twitter_arch['rating_denominator'].value_counts()
```

```

Out[387]: 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 [388]: twitter_arch.name

```

```

Out[388]: 0      Phineas
          1      Tilly
          2      Archie
          3      Darla
          4      Franklin
          5      None
          6      Jax
          7      None
          8      Zoey
          9      Cassie
         10      Koda
         11      Bruno
         12      None
         13      Ted
         14      Stuart
         15      Oliver
         16      Jim
         17      Zeke
         18      Ralphus
         19      Canela
         20      Gerald
         21      Jeffrey
         22      such
         23      Canela
         24      None
         25      None

```

```

26      Maya
27      Mingus
28      Derek
29      Roscoe
...
2326    quite
2327      a
2328    None
2329    None
2330    None
2331    None
2332    None
2333      an
2334      a
2335      an
2336    None
2337    None
2338    None
2339    None
2340    None
2341    None
2342    None
2343    None
2344    None
2345      the
2346      the
2347      a
2348      a
2349      an
2350      a
2351    None
2352      a
2353      a
2354      a
2355    None
Name: name, Length: 2356, dtype: object

```

```
In [389]: len(twitter_arch[twitter_arch.retweeted_status_id.isnull() == False])
```

```
Out[389]: 181
```

```
In [390]: twitter_arch.head()
```

```
Out[390]:
```

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

	timestamp \		source \		text	retweeted_status_id \
0	2017-08-01 16:23:56 +0000		<a href="http://twitter.com/download/iphone" r...		This is Phineas. He's a mystical boy. Only eve...	NaN
1	2017-08-01 00:17:27 +0000		<a href="http://twitter.com/download/iphone" r...		This is Tilly. She's just checking pup on you...	NaN
2	2017-07-31 00:18:03 +0000		<a href="http://twitter.com/download/iphone" r...		This is Archie. He is a rare Norwegian Pouncin...	NaN
3	2017-07-30 15:58:51 +0000		<a href="http://twitter.com/download/iphone" r...		This is Darla. She commenced a snooze mid meal...	NaN
4	2017-07-29 16:00:24 +0000		<a href="http://twitter.com/download/iphone" r...		This is Franklin. He would like you to stop ca...	NaN

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

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

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

Issues of Quality: 1. in twitter_arch.info() ...The variables are float but it needs to be numeric
=> in_reply_to & retweeted_status 2. retweets presented in data 3. dog names are not correct 4.
there are useless columns 5. source are <a href formate

tidiness 1. too much information in source and expanded urls 2. if you look at twit-
ter_arch.head() you will found separate dogs names for different stages in columns

In [391]: img_pred

```
Out[391]:
```

	tweet_id	jpg_url \
0	666020888022790149	https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
5	666050758794694657	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6	666051853826850816	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7	666055525042405380	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8	666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9	666058600524156928	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10	666063827256086533	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11	666071193221509120	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
12	666073100786774016	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13	666082916733198337	https://pbs.twimg.com/media/CT5m4VGWEAAATKc8.jpg
14	666094000022159362	https://pbs.twimg.com/media/CT5w9gUW4AAAsBNN.jpg
15	666099513787052032	https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16	666102155909144576	https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg
17	666104133288665088	https://pbs.twimg.com/media/CT56LSZWAAALJj2.jpg
18	666268910803644416	https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
19	666273097616637952	https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
20	666287406224695296	https://pbs.twimg.com/media/CT8g3BpUEAAAFjg.jpg
21	666293911632134144	https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22	666337882303524864	https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg
23	666345417576210432	https://pbs.twimg.com/media/CT9Vn7PWAAA_ZCM.jpg
24	666353288456101888	https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25	666362758909284353	https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg
26	666373753744588802	https://pbs.twimg.com/media/CT9vZEYUAAALZ05.jpg
27	666396247373291520	https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28	666407126856765440	https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.jpg
29	666411507551481857	https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...
2045	886366144734445568	https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046	886680336477933568	https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2047	886736880519319552	https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
2048	886983233522544640	https://pbs.twimg.com/media/DE8yicJW0AAAABJ.jpg
2049	887101392804085760	https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050	887343217045368832	https://pbs.twimg.com/ext_tw_video_thumb/88734...
2051	887473957103951883	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2052	887517139158093824	https://pbs.twimg.com/ext_tw_video_thumb/88751...
2053	887705289381826560	https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054	888078434458587136	https://pbs.twimg.com/media/DFMWn56WsAAKA7B.jpg
2055	888202515573088257	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056	888554962724278272	https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg
2057	888804989199671297	https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058	888917238123831296	https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg

2059	889278841981685760	https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060	889531135344209921	https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg
2061	889638837579907072	https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062	889665388333682689	https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063	889880896479866881	https://pbs.twimg.com/media/DF199B1WsAITKsg.jpg
2064	890006608113172480	https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065	890240255349198849	https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg
2066	890609185150312448	https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067	890729181411237888	https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068	890971913173991426	https://pbs.twimg.com/media/DF1eOmZXUAAALUcq.jpg
2069	891087950875897856	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070	891327558926688256	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071	891689557279858688	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072	891815181378084864	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073	892177421306343426	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074	892420643555336193	https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

	img_num	p1	p1_conf	p1_dog	\
0	1	Welsh_springer_spaniel	0.465074	True	
1	1	redbone	0.506826	True	
2	1	German_shepherd	0.596461	True	
3	1	Rhodesian_ridgeback	0.408143	True	
4	1	miniature_pinscher	0.560311	True	
5	1	Bernese_mountain_dog	0.651137	True	
6	1	box_turtle	0.933012	False	
7	1	chow	0.692517	True	
8	1	shopping_cart	0.962465	False	
9	1	miniature_poodle	0.201493	True	
10	1	golden_retriever	0.775930	True	
11	1	Gordon_setter	0.503672	True	
12	1	Walker_hound	0.260857	True	
13	1	pug	0.489814	True	
14	1	bloodhound	0.195217	True	
15	1	Lhasa	0.582330	True	
16	1	English_setter	0.298617	True	
17	1	hen	0.965932	False	
18	1	desktop_computer	0.086502	False	
19	1	Italian_greyhound	0.176053	True	
20	1	Maltese_dog	0.857531	True	
21	1	three-toed_sloth	0.914671	False	
22	1	ox	0.416669	False	
23	1	golden_retriever	0.858744	True	
24	1	malamute	0.336874	True	
25	1	guinea_pig	0.996496	False	
26	1	soft-coated_wheaten_terrier	0.326467	True	
27	1	Chihuahua	0.978108	True	
28	1	black-and-tan_coonhound	0.529139	True	
29	1	coho	0.404640	False	

...
2045	1	French_bulldog	0.999201	True
2046	1	convertible	0.738995	False
2047	1	kuvasz	0.309706	True
2048	2	Chihuahua	0.793469	True
2049	1	Samoyed	0.733942	True
2050	1	Mexican_hairless	0.330741	True
2051	2	Pembroke	0.809197	True
2052	1	limousine	0.130432	False
2053	1	basset	0.821664	True
2054	1	French_bulldog	0.995026	True
2055	2	Pembroke	0.809197	True
2056	3	Siberian_husky	0.700377	True
2057	1	golden_retriever	0.469760	True
2058	1	golden_retriever	0.714719	True
2059	1	whippet	0.626152	True
2060	1	golden_retriever	0.953442	True
2061	1	French_bulldog	0.991650	True
2062	1	Pembroke	0.966327	True
2063	1	French_bulldog	0.377417	True
2064	1	Samoyed	0.957979	True
2065	1	Pembroke	0.511319	True
2066	1	Irish_terrier	0.487574	True
2067	2	Pomeranian	0.566142	True
2068	1	Appenzeller	0.341703	True
2069	1	Chesapeake_Bay_retriever	0.425595	True
2070	2	basset	0.555712	True
2071	1	paper_towel	0.170278	False
2072	1	Chihuahua	0.716012	True
2073	1	Chihuahua	0.323581	True
2074	1	orange	0.097049	False

	p2	p2_conf	p2_dog	p3 \
0	collie	0.156665	True	Shetland_sheepdog
1	miniature_pinscher	0.074192	True	Rhodesian_ridgeback
2	malinois	0.138584	True	bloodhound
3	redbone	0.360687	True	miniature_pinscher
4	Rottweiler	0.243682	True	Doberman
5	English_springer	0.263788	True	Greater_Swiss_Mountain_dog
6	mud_turtle	0.045885	False	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 [392]: `img_pred.head()`

```
Out[392]:
```

	tweet_id	jpg_url	\
0	666020888022790149	https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg	
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg	
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg	
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg	

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

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

In [393]: `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 [394]: twitter_arch

```

Out[394]:
      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...

```

[illegible]

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

	text	retweeted_status_id \
0	This is Phineas. He's a mystical boy. Only eve...	NaN
1	This is Tilly. She's just checking pup on you...	NaN
2	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN
5	Here we have a majestic great white breaching ...	NaN
6	Meet Jax. He enjoys ice cream so much he gets ...	NaN
7	When you watch your owner call another dog a g...	NaN
8	This is Zoey. She doesn't want to be one of th...	NaN
9	This is Cassie. She is a college pup. Studying...	NaN
10	This is Koda. He is a South Australian decksha...	NaN
11	This is Bruno. He is a service shark. Only get...	NaN
12	Here's a puppo that seems to be on the fence a...	NaN
13	This is Ted. He does his best. Sometimes that'...	NaN
14	This is Stuart. He's sporting his favorite fan...	NaN
15	This is Oliver. You're witnessing one of his m...	NaN
16	This is Jim. He found a fren. Taught him how t...	NaN
17	This is Zeke. He has a new stick. Very proud o...	NaN
18	This is Ralphus. He's powering up. Attempting ...	NaN
19	RT @dog_rates: This is Canela. She attempted s...	8.874740e+17
20	This is Gerald. He was just told he didn't get...	NaN
21	This is Jeffrey. He has a monopoly on the pool...	NaN
22	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN
23	This is Canela. She attempted some fancy porch...	NaN
24	You may not have known you needed to see this ...	NaN
25	This... is a Jubilant Antarctic House Bear. We...	NaN
26	This is Maya. She's very shy. Rarely leaves he...	NaN
27	This is Mingus. He's a wonderful father to his...	NaN
28	This is Derek. He's late for a dog meeting. 13...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
...
2326	This is quite the dog. Gets really excited whe...	NaN
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN
2328	Oh goodness. A super rare northeast Qdoba kang...	NaN
2329	Those are sunglasses and a jean jacket. 11/10 ...	NaN
2330	Unique dog here. Very small. Lives in containe...	NaN
2331	Here we have a mixed Asiago from the Galápagos...	NaN

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

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

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

	expanded_urls	rating_numerator	\
0	https://twitter.com/dog_rates/status/892420643...	13	
1	https://twitter.com/dog_rates/status/892177421...	13	
2	https://twitter.com/dog_rates/status/891815181...	12	
3	https://twitter.com/dog_rates/status/891689557...	13	
4	https://twitter.com/dog_rates/status/891327558...	12	
5	https://twitter.com/dog_rates/status/891087950...	13	
6	https://gofundme.com/ydvmve-surgery-for-jax,ht...	13	

7	https://twitter.com/dog_rates/status/890729181...	13
8	https://twitter.com/dog_rates/status/890609185...	13
9	https://twitter.com/dog_rates/status/890240255...	14
10	https://twitter.com/dog_rates/status/890006608...	13
11	https://twitter.com/dog_rates/status/889880896...	13
12	https://twitter.com/dog_rates/status/889665388...	13
13	https://twitter.com/dog_rates/status/889638837...	12
14	https://twitter.com/dog_rates/status/889531135...	13
15	https://twitter.com/dog_rates/status/889278841...	13
16	https://twitter.com/dog_rates/status/888917238...	12
17	https://twitter.com/dog_rates/status/888804989...	13
18	https://twitter.com/dog_rates/status/888554962...	13
19	https://twitter.com/dog_rates/status/887473957...	13
20	https://twitter.com/dog_rates/status/888078434...	12
21	https://twitter.com/dog_rates/status/887705289...	13
22	https://twitter.com/dog_rates/status/887517139...	14
23	https://twitter.com/dog_rates/status/887473957...	13
24	https://twitter.com/dog_rates/status/887343217...	13
25	https://twitter.com/dog_rates/status/887101392...	12
26	https://twitter.com/dog_rates/status/886983233...	13
27	https://www.gofundme.com/mingusneedsus , https://...	13
28	https://twitter.com/dog_rates/status/886680336...	13
29	https://twitter.com/dog_rates/status/886366144...	12
...
2326	https://twitter.com/dog_rates/status/666411507...	2
2327	https://twitter.com/dog_rates/status/666407126...	7
2328	https://twitter.com/dog_rates/status/666396247...	9
2329	https://twitter.com/dog_rates/status/666373753...	11
2330	https://twitter.com/dog_rates/status/666362758...	6
2331	https://twitter.com/dog_rates/status/666353288...	8
2332	https://twitter.com/dog_rates/status/666345417...	10
2333	https://twitter.com/dog_rates/status/666337882...	9
2334	https://twitter.com/dog_rates/status/666293911...	3
2335	https://twitter.com/dog_rates/status/666287406...	1
2336	https://twitter.com/dog_rates/status/666273097...	11
2337	https://twitter.com/dog_rates/status/666268910...	10
2338	https://twitter.com/dog_rates/status/666104133...	1
2339	https://twitter.com/dog_rates/status/666102155...	11
2340	https://twitter.com/dog_rates/status/666099513...	8
2341	https://twitter.com/dog_rates/status/666094000...	9
2342	https://twitter.com/dog_rates/status/666082916...	6
2343	https://twitter.com/dog_rates/status/666073100...	10
2344	https://twitter.com/dog_rates/status/666071193...	9
2345	https://twitter.com/dog_rates/status/666063827...	10
2346	https://twitter.com/dog_rates/status/666058600...	8
2347	https://twitter.com/dog_rates/status/666057090...	9
2348	https://twitter.com/dog_rates/status/666055525...	10
2349	https://twitter.com/dog_rates/status/666051853...	2

2350	https://twitter.com/dog_rates/status/666050758...	10
2351	https://twitter.com/dog_rates/status/666049248...	5
2352	https://twitter.com/dog_rates/status/666044226...	6
2353	https://twitter.com/dog_rates/status/666033412...	9
2354	https://twitter.com/dog_rates/status/666029285...	7
2355	https://twitter.com/dog_rates/status/666020888...	8

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

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

[2356 rows x 17 columns]

In [395]: 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 [396]: img_pred.describe()
```

```
Out[396]:
```

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

5 Quality observation

1. confusing columns names to understand
2. several underscore in p1, p2, p3
3. datatype wrong numeric tweetid

```
In [397]: df_twitter
```

```
Out[397]:
```

	id	retweet_count	favorite_count
0	892420643555336193	8853	39467
1	892177421306343426	6514	33819
2	891815181378084864	4328	25461
3	891689557279858688	8964	42908
4	891327558926688256	9774	41048
5	891087950875897856	3261	20562
6	890971913173991426	2158	12041
7	890729181411237888	16716	56848
8	890609185150312448	4429	28226
9	890240255349198849	7711	32467
10	890006608113172480	7624	31166
11	889880896479866881	5156	28268
12	889665388333682689	8538	38818
13	889638837579907072	4735	27672
14	889531135344209921	2321	15359
15	889278841981685760	5637	25652
16	888917238123831296	4709	29611
17	888804989199671297	4559	26080
18	888554962724278272	3732	20290
19	888078434458587136	3653	22201
20	887705289381826560	5609	30779
21	887517139158093824	12082	46959
22	887473957103951883	18781	69871
23	887343217045368832	10737	34222
24	887101392804085760	6167	31061
25	886983233522544640	8084	35859
26	886736880519319552	3443	12306

27	886680336477933568	4610	22798
28	886366144734445568	3316	21524
29	886267009285017600	4	117
...
2324	666411507551481857	339	459
2325	666407126856765440	44	113
2326	666396247373291520	92	172
2327	666373753744588802	100	194
2328	666362758909284353	595	804
2329	666353288456101888	77	229
2330	666345417576210432	146	307
2331	666337882303524864	96	204
2332	666293911632134144	368	522
2333	666287406224695296	71	152
2334	666273097616637952	82	184
2335	666268910803644416	37	108
2336	666104133288665088	6871	14765
2337	666102155909144576	16	81
2338	666099513787052032	73	164
2339	666094000022159362	79	169
2340	666082916733198337	47	121
2341	666073100786774016	174	335
2342	666071193221509120	67	154
2343	666063827256086533	232	496
2344	666058600524156928	61	115
2345	666057090499244032	146	304
2346	666055525042405380	261	448
2347	666051853826850816	879	1253
2348	666050758794694657	60	136
2349	666049248165822465	41	111
2350	666044226329800704	147	311
2351	666033412701032449	47	128
2352	666029285002620928	48	132
2353	666020888022790149	532	2535

[2354 rows x 3 columns]

which include id retweet_count favorite_count

In [398]: df_tweets.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 27 columns):
created_at    2354 non-null object
id            2354 non-null int64
id_str        2354 non-null object
full_text     2354 non-null object
```

```

truncated                2354 non-null bool
display_text_range       2354 non-null object
entities                 2354 non-null object
extended_entities        2073 non-null object
source                   2354 non-null object
in_reply_to_status_id    78 non-null float64
in_reply_to_status_id_str 78 non-null object
in_reply_to_user_id      78 non-null float64
in_reply_to_user_id_str  78 non-null object
in_reply_to_screen_name  78 non-null object
user                     2354 non-null object
geo                      0 non-null object
coordinates              0 non-null object
place                    1 non-null object
contributors             0 non-null object
is_quote_status          2354 non-null bool
retweet_count            2354 non-null int64
favorite_count           2354 non-null int64
favorited                2354 non-null bool
retweeted                2354 non-null bool
possibly_sensitive       2211 non-null object
possibly_sensitive_appealable 2211 non-null object
lang                     2354 non-null object
dtypes: bool(4), float64(2), int64(3), object(18)
memory usage: 432.3+ KB

```

```
In [399]: df_tweets.describe()
```

```

Out[399]:
```

	id	in_reply_to_status_id	in_reply_to_user_id \
count	2.354000e+03	7.800000e+01	7.800000e+01
mean	7.426978e+17	7.455079e+17	2.014171e+16
std	6.852812e+16	7.582492e+16	1.252797e+17
min	6.660209e+17	6.658147e+17	1.185634e+07
25%	6.783975e+17	6.757419e+17	3.086374e+08
50%	7.194596e+17	7.038708e+17	4.196984e+09
75%	7.993058e+17	8.257804e+17	4.196984e+09
max	8.924206e+17	8.862664e+17	8.405479e+17

	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

we have 2 missing IDs from visual observation `twitter_arch.info()`=>RangeIndex: 2356 entries
&`df_tweets.info()`=>RangeIndex: 2354 entries

6 Clean

```
In [400]: twitter_arch_clean=twitter_arch.copy()
          img_pred_clean=img_pred.copy()
          df_twitter_clean=df_twitter.copy()
```

remove wrong names

```
In [401]: wrong_names=twitter_arch_clean[~twitter_arch_clean.name.str.istitle()].name.unique()
```

```
In [402]: for k in wrong_names:
          twitter_arch_clean.name.replace(k, 'None', inplace=True)
```

```
In [403]: twitter_arch_clean.name.unique()
```

```
Out[403]: array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', 'None', 'Jax',
                  'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver', 'Jim',
                  'Zeke', 'Ralphus', 'Canela', 'Gerald', 'Jeffrey', 'Maya', 'Mingus',
                  'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Lilly', 'Earl',
                  'Lola', 'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald', 'Rusty',
                  'Gus', 'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'Elliot', 'Louis',
                  'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack', 'Emmy', 'Steven',
                  'Beau', 'Snoopy', 'Shadow', 'Terrance', 'Aja', 'Penny', 'Dante',
                  'Nelly', 'Ginger', 'Benedict', 'Venti', 'Goose', 'Nugget', 'Cash',
                  'Coco', 'Jed', 'Sebastian', 'Walter', 'Sierra', 'Monkey', 'Harry',
                  'Kody', 'Lassie', 'Rover', 'Napolean', 'Dawn', 'Boomer', 'Cody',
                  'Rumble', 'Clifford', 'Dewey', 'Scout', 'Gizmo', 'Cooper', 'Harold',
                  'Shikha', 'Jamesy', 'Lili', 'Sammy', 'Meatball', 'Paisley', 'Albus',
                  'Neptune', 'Quinn', 'Belle', 'Zooey', 'Dave', 'Jersey', 'Hobbes',
                  'Burt', 'Lorenzo', 'Carl', 'Jordy', 'Milky', 'Trooper', 'Winston',
                  'Sophie', 'Wyatt', 'Rosie', 'Thor', 'Oscar', 'Luna', 'Callie',
                  'Cermet', 'George', 'Marlee', 'Arya', 'Einstein', 'Alice',
                  'Rumpole', 'Benny', 'Aspen', 'Jarod', 'Wiggles', 'General',
                  'Sailor', 'Astrid', 'Iggy', 'Snoop', 'Kyle', 'Leo', 'Riley',
                  'Gidget', 'Noosh', 'Odin', 'Jerry', 'Charlie', 'Georgie', 'Rontu',
                  'Cannon', 'Furzey', 'Daisy', 'Tuck', 'Barney', 'Vixen', 'Jarvis',
                  'Mimosa', 'Pickles', 'Bungalo', 'Brady', 'Margo', 'Sadie', 'Hank',
                  'Tycho', 'Stephan', 'Indie', 'Winnie', 'Bentley', 'Ken', 'Max',
                  'Maddie', 'Pipsy', 'Monty', 'Sojourner', 'Odie', 'Arlo', 'Sunny',
                  'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera', 'Buddy', 'Ava',
                  'Rory', 'Eli', 'Ash', 'Tucker', 'Tobi', 'Chester', 'Wilson',
                  'Sunshine', 'Lipton', 'Gabby', 'Bronte', 'Poppy', 'Rhino', 'Willow',
                  'Orion', 'Eevee', 'Smiley', 'Logan', 'Moreton', 'Klein', 'Miguel',
                  'Emanuel', 'Kuyu', 'Dutch', 'Pete', 'Scooter', 'Reggie', 'Kyro',
                  'Samson', 'Loki', 'Mia', 'Malcolm', 'Dexter', 'Alfie', 'Fiona',
```


'Mutt', 'Bear', 'Doobert', 'Beebop', 'Alexander', 'Sailer',
'Brutus', 'Kona', 'Boots', 'Ralphie', 'Phil', 'Cupid', 'Pawnd',
'Pilot', 'Ike', 'Mo', 'Toby', 'Sweet', 'Pablo', 'Nala', 'Balto',
'Crawford', 'Gabe', 'Mattie', 'Jimison', 'Hercules', 'Duchess',
'Harlso', 'Sampson', 'Sundance', 'Luca', 'Flash', 'Finn', 'Peaches',
'Howie', 'Jazzy', 'Anna', 'Bo', 'Seamus', 'Wafer', 'Chelsea', 'Tom',
'Moose', 'Florence', 'Autumn', 'Dido', 'Eugene', 'Herschel',
'Strudel', 'Tebow', 'Chloe', 'Betty', 'Timber', 'Binky', 'Dudley',
'Comet', 'Larry', 'Levi', 'Akumi', 'Titan', 'Olivia', 'Alf',
'Oshie', 'Bruce', 'Chubbs', 'Sky', 'Atlas', 'Eleanor', 'Layla',
'Rocky', 'Baron', 'Tyr', 'Bauer', 'Swagger', 'Brandi', 'Mary',
'Moe', 'Halo', 'Augie', 'Craig', 'Sam', 'Hunter', 'Pavlov',
'Maximus', 'Wallace', 'Ito', 'Milo', 'Ollie', 'Cali', 'Lennon',
'Major', 'Duke', 'Reginald', 'Sansa', 'Shooter', 'Django', 'Diogi',
'Sonny', 'Philbert', 'Marley', 'Severus', 'Ronnie', 'Anakin',
'Bones', 'Mauve', 'Chef', 'Doc', 'Sobe', 'Longfellow', 'Mister',
'Iroh', 'Baloo', 'Stubert', 'Paull', 'Timison', 'Davey', 'Pancake',
'Tyrone', 'Snicku', 'Ruby', 'Brody', 'Rizzy', 'Mack', 'Butter',
'Nimbus', 'Laika', 'Dobby', 'Juno', 'Maude', 'Lily', 'Newt',
'Benji', 'Nida', 'Robin', 'Monster', 'Remus', 'Mabel', 'Misty',
'Happy', 'Mosby', 'Maggie', 'Leela', 'Ralphy', 'Brownie', 'Meyer',
'Stella', 'Frank', 'Tonks', 'Lincoln', 'Oakley', 'Dale', 'Rizzo',
'Arnie', 'Pinot', 'Dallas', 'Hero', 'Frankie', 'Stormy', 'Mairi',
'Loomis', 'Godi', 'Kenny', 'Deacon', 'Timmy', 'Harper', 'Chipson',
'Combo', 'Dash', 'Bell', 'Hurley', 'Jay', 'Mya', 'Strider',
'Wesley', 'Solomon', 'Huck', 'O', 'Blue', 'Finley', 'Sprinkles',
'Heinrich', 'Shakespeare', 'Fizz', 'Chip', 'Grey', 'Roosevelt',
'Gromit', 'Willem', 'Dakota', 'Dixie', 'Al', 'Jackson', 'Carbon',
'Kirby', 'Lou', 'Nollie', 'Chevy', 'Tito', 'Louie', 'Rupert',
'Rufus', 'Brudge', 'Shadoe', 'Colby', 'Angel', 'Brat', 'Tove',
'Aubie', 'Kota', 'Eve', 'Glenn', 'Shelby', 'Sephie', 'Bonaparte',
'Albert', 'Wishes', 'Rose', 'Theo', 'Rocco', 'Fido', 'Emma',
'Spencer', 'Lilli', 'Boston', 'Brandonald', 'Corey', 'Leonard',
'Chompsky', 'Beckham', 'Devón', 'Gert', 'Watson', 'Rubio', 'Keith',
'Dex', 'Carly', 'Ace', 'Tayzie', 'Grizzie', 'Fred', 'Gilbert',
'Zoe', 'Stewie', 'Calvin', 'Lilah', 'Spanky', 'Jameson', 'Piper',
'Atticus', 'Blu', 'Dietrich', 'Divine', 'Tripp', 'Cora', 'Huxley',
'Keurig', 'Bookstore', 'Linus', 'Abby', 'Shaggy', 'Shiloh',
'Gustav', 'Arlen', 'Percy', 'Lenox', 'Sugar', 'Harvey', 'Blanket',
'Geno', 'Stark', 'Beya', 'Kilo', 'Kayla', 'Maxaroni', 'Doug',
'Edmund', 'Aqua', 'Theodore', 'Chase', 'Rorie', 'Simba', 'Charles',
'Bayley', 'Axel', 'Storkson', 'Remy', 'Chadrick', 'Kellogg',
'Buckley', 'Livvie', 'Terry', 'Hermione', 'Ralpher', 'Aldrick',
'Rooney', 'Crystal', 'Ziva', 'Stefan', 'Pupcasso', 'Puff',
'Flurpson', 'Coleman', 'Enchilada', 'Raymond', 'Rueben', 'Cilantro',
'Karll', 'Sprout', 'Blitz', 'Bloop', 'Lillie', 'Ashleigh',
'Kreggory', 'Sarge', 'Luther', 'Ivar', 'Jangle', 'Schnitzel',
'Panda', 'Berkeley', 'Ralphé', 'Charleson', 'Clyde', 'Harnold',

'Sid', 'Pippa', 'Otis', 'Carper', 'Bowie', 'Alexanderson', 'Suki',
'Barclay', 'Skittle', 'Ebby', 'Flávio', 'Smokey', 'Link',
'Jennifur', 'Ozzy', 'Bluebert', 'Stephanus', 'Bubbles', 'Zeus',
'Bertson', 'Nico', 'Michelangelo', 'Siba', 'Calbert', 'Curtis',
'Travis', 'Thumas', 'Kanu', 'Lance', 'Opie', 'Kane', 'Olive',
'Chuckles', 'Staniel', 'Sora', 'Beemo', 'Gunner', 'Lacy', 'Tater',
'Olaf', 'Cecil', 'Vince', 'Karma', 'Billy', 'Walker', 'Rodney',
'Klevin', 'Malikai', 'Bobbie', 'River', 'Jebberson', 'Remington',
'Farfle', 'Jiminus', 'Clarkus', 'Finnegus', 'Cupcake', 'Kathmandu',
'Ellie', 'Katie', 'Kara', 'Adele', 'Zara', 'Ambrose', 'Jimothy',
'Bode', 'Terrenth', 'Reese', 'Chesterson', 'Lucia', 'Bisquick',
'Ralphson', 'Socks', 'Rambo', 'Rudy', 'Fiji', 'Rilo', 'Bilbo',
'Coopson', 'Yoda', 'Millie', 'Chet', 'Crouton', 'Daniel', 'Kaia',
'Murphy', 'Dotsy', 'Eazy', 'Coops', 'Fillup', 'Miley', 'Charl',
'Reagan', 'Yukon', 'Cuddles', 'Claude', 'Jessiga', 'Carter', 'Ole',
'Pherb', 'Blipson', 'Reptar', 'Trevith', 'Berb', 'Bob', 'Colin',
'Brian', 'Olivier', 'Grady', 'Kobe', 'Freddery', 'Bodie', 'Dunkin',
'Wally', 'Tupawc', 'Amber', 'Edgar', 'Teddy', 'Kingsley', 'Brockly',
'Richie', 'Molly', 'Vinscent', 'Cedrick', 'Hazel', 'Lolo', 'Eriq',
'Phred', 'Oddie', 'Maxwell', 'Geoff', 'Covach', 'Durg', 'Fynn',
'Ricky', 'Herald', 'Lucky', 'Ferg', 'Trip', 'Clarence', 'Hamrick',
'Brad', 'Pubert', 'Frönq', 'Derby', 'Lizzie', 'Ember', 'Blakely',
'Opal', 'Marq', 'Kramer', 'Barry', 'Gordon', 'Baxter', 'Mona',
'Horace', 'Crimson', 'Birf', 'Hammond', 'Lorelei', 'Marty',
'Brooks', 'Petrick', 'Hubertson', 'Gerbald', 'Oreo', 'Bruiser',
'Perry', 'Bobby', 'Jeph', 'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe',
'Tiger', 'Jiminy', 'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Darrel',
'Taco', 'Joey', 'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses',
'Toffee', 'Apollo', 'Asher', 'Glacier', 'Chuck', 'Champ', 'Ozzie',
'Griswold', 'Cheesy', 'Moofasa', 'Hector', 'Goliath', 'Kawhi',
'Emmie', 'Penelope', 'Willie', 'Rinna', 'Mike', 'William', 'Dwight',
'Evy', 'Rascal', 'Linda', 'Tug', 'Tango', 'Grizz', 'Jerome',
'Crumpet', 'Jessifer', 'Izzy', 'Ralph', 'Sandy', 'Humphrey',
'Tassy', 'Juckson', 'Chuq', 'Tyrus', 'Karl', 'Godzilla', 'Vinnie',
'Kenneth', 'Herm', 'Bert', 'Striker', 'Donny', 'Pepper', 'Bernie',
'Buddah', 'Lenny', 'Arnold', 'Zuzu', 'Mollie', 'Laela', 'Tedders',
'Superpup', 'Rufio', 'Jeb', 'Rodman', 'Jonah', 'Chesney', 'Henry',
'Bobbay', 'Mitch', 'Kaiya', 'Acro', 'Aiden', 'Obie', 'Dot',
'Shnuggles', 'Kendall', 'Jeffri', 'Steve', 'Mac', 'Fletcher',
'Kenzie', 'Pumpkin', 'Schnozz', 'Gustaf', 'Cheryl', 'Ed',
'Leonidas', 'Norman', 'Caryl', 'Scott', 'Taz', 'Darby', 'Jackie',
'Jazz', 'Franq', 'Pippin', 'Rolf', 'Snickers', 'Ridley', 'Cal',
'Bradley', 'Bubba', 'Tuco', 'Patch', 'Mojo', 'Batdog', 'Dylan',
'Mark', 'Alejandro', 'Scruffers', 'Pip', 'Julius', 'Tanner',
'Sparky', 'Anthony', 'Holly', 'Jett', 'Amy', 'Sage', 'Andy',
'Mason', 'Trigger', 'Antony', 'Creg', 'Traviss', 'Gin', 'Jeffrie',
'Danny', 'Ester', 'Pluto', 'Bloo', 'Edd', 'Willy', 'Herb', 'Damon',
'Peanut', 'Nigel', 'Butters', 'Sandra', 'Fabio', 'Randall', 'Liam',

```

'Tommy', 'Ben', 'Raphael', 'Julio', 'Andru', 'Kloey', 'Shawwn',
'Skye', 'Kollin', 'Ronduh', 'Billl', 'Saydee', 'Dug', 'Tessa',
'Sully', 'Kirk', 'Ralf', 'Clarq', 'Jaspers', 'Samsom', 'Harrison',
'Chaz', 'Jeremy', 'Jaycob', 'Lambeau', 'Ruffles', 'Amélie', 'Bobb',
'Banditt', 'Kevon', 'Winifred', 'Hanz', 'Churlie', 'Zeek', 'Timofy',
'Maks', 'Jomathan', 'Kallie', 'Marvin', 'Spark', 'Gòrdón', 'Jo',
'Jareld', 'Torque', 'Ron', 'Skittles', 'Cleopatrícia', 'Erik',
'Stu', 'Tedrick', 'Filup', 'Kial', 'Naphaniel', 'Dook', 'Hall',
'Philippe', 'Biden', 'Fwed', 'Genevieve', 'Joshwa', 'Bradlay',
'Clybe', 'Keet', 'Carll', 'Jockson', 'Josep', 'Lugan', 'Christoper'], dtype=obj

```

```
In [404]: twitter_arch_clean.name.value_counts()
```

```

Out[404]: None      860
Charlie      12
Cooper       11
Oliver       11
Lucy         11
Lola         10
Tucker       10
Penny        10
Bo           9
Winston      9
Sadie        8
Toby         7
Daisy        7
Bailey       7
Buddy        7
Jax          6
Dave         6
Koda         6
Milo         6
Scout        6
Stanley      6
Bella        6
Leo          6
Oscar        6
Jack         6
Rusty        6
Sammy        5
Phil         5
Louis        5
Larry        5
...
Covach       1
Fabio        1
Julius       1
Timber       1

```

Winifred	1
Toffee	1
Sweets	1
Clarkus	1
Patch	1
Tanner	1
Josep	1
Dunkin	1
Huxley	1
Keet	1
Cupid	1
Barney	1
Tonks	1
Wafer	1
Nimbus	1
Jett	1
Skye	1
Napolean	1
Ralphson	1
Jersey	1
Lucia	1
Juckson	1
Kenzie	1
Bradlay	1
Shelby	1
Rontu	1

Name: name, Length: 927, dtype: int64

```
In [405]: twitter_arch_clean.rating_denominator=twitter_arch_clean['rating_denominator']=10
```

```
In [406]: twitter_arch_clean.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 [407]: twitter_arch_clean['timestamp']=pd.to_datetime(twitter_arch_clean['timestamp'])
```

```
In [408]: twitter_arch_clean['retweeted_status_timestamp']=pd.to_datetime(twitter_arch_clean['retweeted_status_timestamp'])
```

```
In [409]: twitter_arch_clean.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 datetime64[ns]
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 datetime64[ns]
expanded_urls           2297 non-null object
rating_numerator         2356 non-null int64
rating_denominator       2356 non-null int64
name                    2356 non-null object
doggo                   2356 non-null object
floofer                 2356 non-null object
pupper                  2356 non-null object
puppo                   2356 non-null object
dtypes: datetime64[ns](2), float64(4), int64(3), object(8)
memory usage: 313.0+ KB

```

```
In [410]: twitter_arch_clean.head()
```

```

Out[410]:
   tweet_id  in_reply_to_status_id  in_reply_to_user_id \
0  892420643555336193             NaN                 NaN
1  892177421306343426             NaN                 NaN
2  891815181378084864             NaN                 NaN
3  891689557279858688             NaN                 NaN
4  891327558926688256             NaN                 NaN

   timestamp                                     source \
0  2017-08-01 16:23:56  <a href="http://twitter.com/download/iphone" r...

```

```

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

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

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

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

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

```

you can see in `twitter_arch_clean.head()` timestamp that converted to normal date formatting
remove dog stages columns

```

In [411]: def dog_stages(row):
            stage=[]
            if row['doggo']=='doggo':
                stage.append('doggo')
            else:
                return ','.join(stage)
            if row['floofer']=='floofer':
                stage.append('floofer')
            else:
                return ','.join(stage)
            if row['pupper']=='pupper':
                stage.append('pupper')

```

```

        else:
            return ','.join(stage)
        if row['puppo']=='puppo':
            stage.append('puppo')
        else:
            return ','.join(stage)
        if not stage:
            return ','.join(stage)
twitter_arch_clean['dog_stages']=twitter_arch_clean.apply(lambda row:dog_stages(row),a

In [412]: twitter_arch_clean.drop(['doggo','floofer','pupper','puppo'],axis=1,inplace=True)

In [413]: twitter_arch_clean.head()

Out[413]:
```

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

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

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

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

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

	rating_denominator	name	dog_stages
0	10	Phineas	
1	10	Tilly	
2	10	Archie	
3	10	Darla	
4	10	Franklin	

cleared and combined into one column named dog_tages

```
In [414]: pd.value_counts(twitter_arch_clean['dog_stages'])
```

```
Out[414]:          2259
doggo          96
doggo,floofer    1
Name: dog_stages, dtype: int64
```

```
In [415]: twitter_arch_clean[['dog_stages']]
```

```
Out[415]:      dog_stages
0
1
2
3
4
5
6
7
8
9      doggo
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
```



```

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

```

```
[2356 rows x 1 columns]
```

7 img prediction

```
In [416]: img_pred.head()
```

```

Out[416]:
      tweet_id      jpg_url \
0  666020888022790149  https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg
1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2  666033412701032449  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3  666044226329800704  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4  666049248165822465  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

      img_num      p1      p1_conf      p1_dog      p2 \

```

0	1	Welsh_springer_spaniel	0.465074	True	collie
1	1	redbone	0.506826	True	miniature_pinscher
2	1	German_shepherd	0.596461	True	malinois
3	1	Rhodesian_ridgeback	0.408143	True	redbone
4	1	miniature_pinscher	0.560311	True	Rottweiler

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

```
In [417]: img_pred.p1=img_pred.p1.str.capitalize()
img_pred.p2=img_pred.p2.str.capitalize()
img_pred.p3=img_pred.p3.str.capitalize()
```

```
In [418]: img_pred.head()
```

```
Out[418]:
```

	tweet_id	jpg_url	\
0	666020888022790149	https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg	
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg	
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg	
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg	

	img_num	p1	p1_conf	p1_dog	p2	\
0	1	Welsh_springer_spaniel	0.465074	True	Collie	
1	1	Redbone	0.506826	True	Miniature_pinscher	
2	1	German_shepherd	0.596461	True	Malinois	
3	1	Rhodesian_ridgeback	0.408143	True	Redbone	
4	1	Miniature_pinscher	0.560311	True	Rottweiler	

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

by checking p1, p1 p3 you will find some underscore and dash to be replaced with spaces

```
In [419]: img_pred.p1=img_pred.p1.str.replace("-", " ")
img_pred.p2=img_pred.p2.str.replace("-", " ")
img_pred.p3=img_pred.p3.str.replace("-", " ")
```

```
In [420]: img_pred.head()
```

```
Out[420]:
```

	tweet_id	jpg_url	\
0	666020888022790149	https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg	

```

1 666029285002620928 https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2 666033412701032449 https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3 666044226329800704 https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4 666049248165822465 https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

```

	img_num	p1	p1_conf	p1_dog	p2 \
0	1	Welsh_springer_spaniel	0.465074	True	Collie
1	1	Redbone	0.506826	True	Miniature_pinscher
2	1	German_shepherd	0.596461	True	Malinois
3	1	Rhodesian_ridgeback	0.408143	True	Redbone
4	1	Miniature_pinscher	0.560311	True	Rottweiler

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

```

In [421]: img_pred.p1=img_pred.p1.str.replace("_"," ")
img_pred.p2=img_pred.p2.str.replace("_"," ")
img_pred.p3=img_pred.p3.str.replace("_"," ")

```

```

In [422]: img_pred.head()

```

```

Out [422]:          tweet_id          jpg_url \
0 666020888022790149 https://pbs.twimg.com/media/CT4udnOWwAA0aMy.jpg
1 666029285002620928 https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
2 666033412701032449 https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
3 666044226329800704 https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4 666049248165822465 https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

```

	img_num	p1	p1_conf	p1_dog	p2 \
0	1	Welsh springer spaniel	0.465074	True	Collie
1	1	Redbone	0.506826	True	Miniature pinscher
2	1	German shepherd	0.596461	True	Malinois
3	1	Rhodesian ridgeback	0.408143	True	Redbone
4	1	Miniature pinscher	0.560311	True	Rottweiler

	p2_conf	p2_dog	p3	p3_conf	p3_dog
0	0.156665	True	Shetland sheepdog	0.061428	True
1	0.074192	True	Rhodesian ridgeback	0.072010	True
2	0.138584	True	Bloodhound	0.116197	True
3	0.360687	True	Miniature pinscher	0.222752	True
4	0.243682	True	Doberman	0.154629	True

```

In [423]: df_twitter_clean.id.apply(str)

```

```

Out [423]: 0      892420643555336193
1      892177421306343426

```

2	891815181378084864
3	891689557279858688
4	891327558926688256
5	891087950875897856
6	890971913173991426
7	890729181411237888
8	890609185150312448
9	890240255349198849
10	890006608113172480
11	889880896479866881
12	889665388333682689
13	889638837579907072
14	889531135344209921
15	889278841981685760
16	888917238123831296
17	888804989199671297
18	888554962724278272
19	888078434458587136
20	887705289381826560
21	887517139158093824
22	887473957103951883
23	887343217045368832
24	887101392804085760
25	886983233522544640
26	886736880519319552
27	886680336477933568
28	886366144734445568
29	886267009285017600
	...
2324	666411507551481857
2325	666407126856765440
2326	666396247373291520
2327	666373753744588802
2328	666362758909284353
2329	666353288456101888
2330	666345417576210432
2331	666337882303524864
2332	666293911632134144
2333	666287406224695296
2334	666273097616637952
2335	666268910803644416
2336	666104133288665088
2337	666102155909144576
2338	666099513787052032
2339	666094000022159362
2340	666082916733198337
2341	666073100786774016
2342	666071193221509120

```

2343    666063827256086533
2344    666058600524156928
2345    666057090499244032
2346    666055525042405380
2347    666051853826850816
2348    666050758794694657
2349    666049248165822465
2350    666044226329800704
2351    666033412701032449
2352    666029285002620928
2353    666020888022790149
Name: id, Length: 2354, dtype: object

```

```

In [424]: df_twitter_clean['id']=df_twitter['id'].astype('str')
          df_twitter_clean=df_twitter_clean[['id','favorite_count','retweet_count']]

```

```

In [425]: df_twitter_clean.head()

```

```

Out[425]:
           id  favorite_count  retweet_count
0  892420643555336193         39467         8853
1  892177421306343426         33819         6514
2  891815181378084864         25461         4328
3  891689557279858688         42908         8964
4  891327558926688256         41048         9774

```

```

In [426]: df_twitter_clean.info()

```

```

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

```

```

In [427]: df_twitter_clean.rename(columns={'id':'tweets_id'},inplace=True)

```

```

In [428]: df_twitter_clean.info()

```

```

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

```

```
In [429]: df_twitter_clean['tweets_id']=df_twitter_clean[str('tweets_id')]
```

```
In [430]: type(df_twitter_clean['tweets_id'].iloc[0])
```

```
Out[430]: str
```

```
In [431]: img_pred_clean.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 [432]: twitter_arch_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 14 columns):
tweet_id      2356 non-null int64
in_reply_to_status_id    78 non-null float64
in_reply_to_user_id      78 non-null float64
timestamp      2356 non-null datetime64[ns]
source         2356 non-null object
text           2356 non-null object
retweeted_status_id      181 non-null float64
retweeted_status_user_id  181 non-null float64
retweeted_status_timestamp 181 non-null datetime64[ns]
expanded_urls  2297 non-null object
rating_numerator  2356 non-null int64
rating_denominator  2356 non-null int64
name           2356 non-null object
dog_stages     2356 non-null object
dtypes: datetime64[ns](2), float64(4), int64(3), object(5)
memory usage: 257.8+ KB
```

```

In [433]: df_twitter_clean.info()

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

In [434]: df_twitter_clean.rename(columns={'tweets_id': 'tweet_id'}, inplace=True)

In [435]: df_twitter_clean.info()

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

In [436]: img_pred_clean['tweet_id']=img_pred_clean[str('tweet_id')]
          df_twitter_clean['tweet_id']=df_twitter_clean[str('tweet_id')]
          twitter_arch_clean['tweet_id']=twitter_arch_clean[str('tweet_id')]

In [437]: type(img_pred_clean['tweet_id'].iloc[0])
          type(df_twitter_clean['tweet_id'].iloc[0])
          type(twitter_arch_clean['tweet_id'].iloc[0])

Out[437]: numpy.int64

In [438]: img_pred_clean['tweet_id']=img_pred_clean['tweet_id'].astype('str')
          df_twitter_clean['tweet_id']=df_twitter_clean['tweet_id'].astype('str')
          twitter_arch_clean['tweet_id']=twitter_arch_clean['tweet_id'].astype('str')

In [439]: type(img_pred_clean['tweet_id'].iloc[0])
          type(df_twitter_clean['tweet_id'].iloc[0])
          type(twitter_arch_clean['tweet_id'].iloc[0])

Out[439]: str

In [440]: merge=pd.merge(twitter_arch_clean,img_pred_clean,on=['tweet_id'],how='inner')
          merge=pd.merge(merge,df_twitter_clean,on='tweet_id',how='inner')

In [441]: merge.info()

```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2073 entries, 0 to 2072
Data columns (total 27 columns):
tweet_id                2073 non-null object
in_reply_to_status_id   23 non-null float64
in_reply_to_user_id     23 non-null float64
timestamp               2073 non-null datetime64[ns]
source                  2073 non-null object
text                    2073 non-null object
retweeted_status_id     79 non-null float64
retweeted_status_user_id 79 non-null float64
retweeted_status_timestamp 79 non-null datetime64[ns]
expanded_urls           2073 non-null object
rating_numerator        2073 non-null int64
rating_denominator      2073 non-null int64
name                    2073 non-null object
dog_stages              2073 non-null object
jpg_url                 2073 non-null object
img_num                 2073 non-null int64
p1                      2073 non-null object
p1_conf                 2073 non-null float64
p1_dog                  2073 non-null bool
p2                      2073 non-null object
p2_conf                 2073 non-null float64
p2_dog                  2073 non-null bool
p3                      2073 non-null object
p3_conf                 2073 non-null float64
p3_dog                  2073 non-null bool
favorite_count          2073 non-null int64
retweet_count           2073 non-null int64
dtypes: bool(3), datetime64[ns](2), float64(7), int64(5), object(10)
memory usage: 411.0+ KB

```

merging was helpful using the inner index to join https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html

```
In [442]: merge.head()
```

```

Out[442]:
   tweet_id  in_reply_to_status_id  in_reply_to_user_id \
0  892420643555336193             NaN                 NaN
1  892177421306343426             NaN                 NaN
2  891815181378084864             NaN                 NaN
3  891689557279858688             NaN                 NaN
4  891327558926688256             NaN                 NaN

   timestamp                                     source \
0  2017-08-01 16:23:56  <a href="http://twitter.com/download/iphone" r...

```



```

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

```

```

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

```

```

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

```

```

                                expanded_urls  ...  p1_conf  \
0  https://twitter.com/dog_rates/status/892420643...  ...  0.097049
1  https://twitter.com/dog_rates/status/892177421...  ...  0.323581
2  https://twitter.com/dog_rates/status/891815181...  ...  0.716012
3  https://twitter.com/dog_rates/status/891689557...  ...  0.170278
4  https://twitter.com/dog_rates/status/891327558...  ...  0.555712

```

```

p1_dog      p2  p2_conf  p2_dog      p3  \
0  False      bagel  0.085851  False      banana
1  True       Pekinese  0.090647  True      papillon
2  True       malamute  0.078253  True      kelpie
3  False  Labrador_retriever  0.168086  True      spatula
4  True    English_springer  0.225770  True  German_short-haired_pointer

```

```

p3_conf  p3_dog  favorite_count  retweet_count
0  0.076110  False      39467      8853
1  0.068957  True      33819      6514
2  0.031379  True      25461      4328
3  0.040836  False      42908      8964
4  0.175219  True      41048      9774

```

```
[5 rows x 27 columns]
```

```
In [443]: merge.drop(['retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timesta
```

```
In [444]: merge.head()
```

```

Out[444]:
            tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
0  892420643555336193                  NaN                  NaN

```

1	892177421306343426	NaN	NaN
2	891815181378084864	NaN	NaN
3	891689557279858688	NaN	NaN
4	891327558926688256	NaN	NaN

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

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

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

	rating_denominator	name	...	p1_conf	p1_dog \
0	10	Phineas	...	0.097049	False
1	10	Tilly	...	0.323581	True
2	10	Archie	...	0.716012	True
3	10	Darla	...	0.170278	False
4	10	Franklin	...	0.555712	True

	p2	p2_conf	p2_dog	p3 \
0	bagel	0.085851	False	banana
1	Pekinese	0.090647	True	papillon
2	malamute	0.078253	True	kelpie
3	Labrador_retriever	0.168086	True	spatula
4	English_springer	0.225770	True	German_short-haired_pointer

	p3_conf	p3_dog	favorite_count	retweet_count
0	0.076110	False	39467	8853
1	0.068957	True	33819	6514
2	0.031379	True	25461	4328
3	0.040836	False	42908	8964
4	0.175219	True	41048	9774

[5 rows x 24 columns]