

wrangle_act

November 22, 2018

1 Project: We Rate Dogs

Data Wrangling and Analyzing

Presented By: Ali Hasan

Date: 11/22/2018

2 Introduction

Real-world data rarely comes clean. Using Python and its libraries, you will gather data from a variety of sources regarding one of Twitter accounts (We Rate Dogs) and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling. That's the first part. Second part would be to merge all datasets in one master dataset then conduct different types of analyses and visualization. Third part will be consisted of two PDFs. One for the Data Wrangling efforts and the second one for Analyses and Visualizations

2.1 First Part - Data Wrangling

- Gathering Data - Download CSV file - Create a TSV file from URL - Download Twitter API and JSON ##### - Assessing Data - Quality - Tidiness ##### - Cleaning Data ## Second Part - Storing, Analyzing, and Visualizing Wrangled Data ## Third Part - Reporting on: - Wrangling Efforts - Data Analyses and Visualizations

```
In [133]: # Import all required libraries
import pandas as pd
import json
import requests
import numpy as np
import tweepy
import matplotlib.pyplot as plt
import random
import matplotlib
%matplotlib inline
import seaborn as sns
import time
```

2.2 First Part - Data Wrangling

2.2.1 Gathering Data

Download CSV file `twitter_archive.csv`

```
In [2]: # Read downloaded twitter archive CSV
```

```
twitter_archive = pd.read_csv('twitter-archive-enhanced.csv')
twitter_archive.head()
```

```
Out[2]:
```

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

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

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

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

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

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

```

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

```

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

Create a TSV file from URL

```

In [3]: # download url
url = "https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-prediction-2017-08-17/599fd2ad_image-prediction-2017-08-17.jpg"
response = requests.get(url)

# Open tsv and write contents
with open('image.tsv', mode='wb') as file:
    file.write(response.content)

# Read TSV file
image = pd.read_csv('image.tsv', sep='\t')
image.head()

```

```

Out[3]:
      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

```

Download Twitter API and JSON

```

In [4]: # Tweeter API Developer Credentials
import tweepy

consumer_key = 'R4l3JLU7i3jiEyvzmji6DLAj0'
consumer_secret = 'ecp6uYvKa7ankFTUNTTBUXf2fb4XNODQzmkt6Vc7wRo0jQecu3'
access_token = '1061800277378981888-lgTwRZJyDDnQ2VegCRlRVA2lW8x9MN'
access_secret = 'wR7fMnT1wCYut3b7P1QZHQ37pY8lQUkUdtuersUT0mpeY'

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

api = tweepy.API(auth)

In [5]: #Twitter Query using tweet_id information from the twitter_archive CSV file.

tweet_ids = list(twitter_archive.tweet_id)

tweet_data = {}
error_list = []
for tweet in tweet_ids:
    try:
        tweet_status = api.get_status(tweet,
                                       wait_on_rate_limit=True,
                                       wait_on_rate_limit_notify=True)
        tweet_data[str(tweet)] = tweet_status._json
    except:
        print("Error for tweet_id: " + str(tweet))
        error_list.append(tweet)

Error for tweet_id: 888202515573088257
Error for tweet_id: 873697596434513921
Error for tweet_id: 872668790621863937
Error for tweet_id: 869988702071779329
Error for tweet_id: 866816280283807744
Error for tweet_id: 861769973181624320
Error for tweet_id: 845459076796616705
Error for tweet_id: 842892208864923648
Error for tweet_id: 837012587749474308
Error for tweet_id: 827228250799742977
Error for tweet_id: 802247111496568832
Error for tweet_id: 775096608509886464
Error for tweet_id: 770743923962707968
Rate limit reached. Sleeping for: 733
Error for tweet_id: 754011816964026368
Rate limit reached. Sleeping for: 734

In [9]: # size of dataframe

```

```

print("The number of tweed_ids is", len(tweet_data.keys()))
# The number of the errors
print("The number of tweet_ids with error is", len(error_list))

```

The number of tweed_ids is 2342
The number of tweet_ids with error is 14

```

In [10]: # Write JSON in a text file
with open('tweet_json.txt', 'w') as outfile:
    json.dump(tweet_data, outfile, sort_keys = True, indent=4, ensure_ascii = False)
# Read the new JSON file
twitter_api = pd.read_json('tweet_json.txt', orient='index')
twitter_api.head()

```

```

Out[10]:

```

	contributors	coordinates	created_at	\
1991-02-08 13:48:08.022790149	NaN	NaN	2015-11-15 22:32:08	
1991-02-08 16:08:05.002620928	NaN	NaN	2015-11-15 23:05:30	
1991-02-08 17:16:52.701032449	NaN	NaN	2015-11-15 23:21:54	
1991-02-08 20:17:06.329800704	NaN	NaN	2015-11-16 00:04:52	
1991-02-08 21:40:48.165822465	NaN	NaN	2015-11-16 00:24:50	

	entities	\
1991-02-08 13:48:08.022790149	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 16:08:05.002620928	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 17:16:52.701032449	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 20:17:06.329800704	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 21:40:48.165822465	{'hashtags': [], 'media': [{'display_url': 'pi...	

	extended_entities	\
1991-02-08 13:48:08.022790149	{'media': [{'display_url': 'pic.twitter.com/BL...	
1991-02-08 16:08:05.002620928	{'media': [{'display_url': 'pic.twitter.com/r7...	
1991-02-08 17:16:52.701032449	{'media': [{'display_url': 'pic.twitter.com/y6...	
1991-02-08 20:17:06.329800704	{'media': [{'display_url': 'pic.twitter.com/DW...	
1991-02-08 21:40:48.165822465	{'media': [{'display_url': 'pic.twitter.com/4B...	

	favorite_count	favorited	geo	\
1991-02-08 13:48:08.022790149	2544	0	NaN	
1991-02-08 16:08:05.002620928	129	0	NaN	
1991-02-08 17:16:52.701032449	125	0	NaN	
1991-02-08 20:17:06.329800704	296	0	NaN	
1991-02-08 21:40:48.165822465	107	0	NaN	

	id	id_str	\
1991-02-08 13:48:08.022790149	666020888022790144	666020888022790144	
1991-02-08 16:08:05.002620928	666029285002620928	666029285002620928	
1991-02-08 17:16:52.701032449	666033412701032448	666033412701032448	
1991-02-08 20:17:06.329800704	666044226329800704	666044226329800704	

1991-02-08 21:40:48.165822465	666049248165822464	666049248165822464
-------------------------------	--------------------	--------------------

	...	\
1991-02-08 13:48:08.022790149	...	
1991-02-08 16:08:05.002620928	...	
1991-02-08 17:16:52.701032449	...	
1991-02-08 20:17:06.329800704	...	
1991-02-08 21:40:48.165822465	...	

	quoted_status	quoted_status_id	\
1991-02-08 13:48:08.022790149	NaN	NaN	
1991-02-08 16:08:05.002620928	NaN	NaN	
1991-02-08 17:16:52.701032449	NaN	NaN	
1991-02-08 20:17:06.329800704	NaN	NaN	
1991-02-08 21:40:48.165822465	NaN	NaN	

	quoted_status_id_str	retweet_count	retweeted	\
1991-02-08 13:48:08.022790149	NaN	508	0	
1991-02-08 16:08:05.002620928	NaN	47	0	
1991-02-08 17:16:52.701032449	NaN	44	0	
1991-02-08 20:17:06.329800704	NaN	139	0	
1991-02-08 21:40:48.165822465	NaN	40	0	

	retweeted_status	\
1991-02-08 13:48:08.022790149	NaN	
1991-02-08 16:08:05.002620928	NaN	
1991-02-08 17:16:52.701032449	NaN	
1991-02-08 20:17:06.329800704	NaN	
1991-02-08 21:40:48.165822465	NaN	

	source	\
1991-02-08 13:48:08.022790149	<a href="http://twitter.com/download/iphone" r...	
1991-02-08 16:08:05.002620928	<a href="http://twitter.com/download/iphone" r...	
1991-02-08 17:16:52.701032449	<a href="http://twitter.com/download/iphone" r...	
1991-02-08 20:17:06.329800704	<a href="http://twitter.com/download/iphone" r...	
1991-02-08 21:40:48.165822465	<a href="http://twitter.com/download/iphone" r...	

	text	\
1991-02-08 13:48:08.022790149	Here we have a Japanese Irish Setter. Lost eye...	
1991-02-08 16:08:05.002620928	This is a western brown Mitsubishi terrier. Up...	
1991-02-08 17:16:52.701032449	Here is a very happy pup. Big fan of well-main...	
1991-02-08 20:17:06.329800704	This is a purebred Piers Morgan. Loves to Netf...	
1991-02-08 21:40:48.165822465	Here we have a 1949 1st generation vulpix. Enj...	

	truncated	\
1991-02-08 13:48:08.022790149	0	
1991-02-08 16:08:05.002620928	0	
1991-02-08 17:16:52.701032449	0	

```
1991-02-08 20:17:06.329800704      0
1991-02-08 21:40:48.165822465      0
```

```

                                user
1991-02-08 13:48:08.022790149 {'contributors_enabled': False, 'created_at': ...
1991-02-08 16:08:05.002620928 {'contributors_enabled': False, 'created_at': ...
1991-02-08 17:16:52.701032449 {'contributors_enabled': False, 'created_at': ...
1991-02-08 20:17:06.329800704 {'contributors_enabled': False, 'created_at': ...
1991-02-08 21:40:48.165822465 {'contributors_enabled': False, 'created_at': ...
```

```
[5 rows x 30 columns]
```

2.2.2 Assessing Data

Initial assessing for the three file visual and programatic:

```
In [11]: twitter_archive
```

```
Out[11]:
```

	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

[illegible]

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/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 [12]: # Get Data info
```

```
twitter_archive.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 2356 entries, 0 to 2355
```

```
Data columns (total 17 columns):
```

```
tweet_id                2356 non-null int64
in_reply_to_status_id   78 non-null float64
in_reply_to_user_id     78 non-null float64
timestamp               2356 non-null object
source                  2356 non-null object
text                    2356 non-null object
retweeted_status_id     181 non-null float64
retweeted_status_user_id 181 non-null float64
retweeted_status_timestamp 181 non-null object
expanded_urls           2297 non-null object
rating_numerator        2356 non-null int64
rating_denominator      2356 non-null int64
name                    2356 non-null object
```



```

doggo                2356 non-null object
floofer              2356 non-null object
pupper               2356 non-null object
puppo                2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB

```

```

In [13]: # Get summary of our quantative data
         twitter_archive.describe()

```

```

Out[13]:

```

	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 [14]: image

```

```

Out[14]:

```

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

5	666050758794694657	https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6	666051853826850816	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7	666055525042405380	https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8	666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9	666058600524156928	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10	666063827256086533	https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11	666071193221509120	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg
12	666073100786774016	https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13	666082916733198337	https://pbs.twimg.com/media/CT5m4VGWEAAAtKc8.jpg
14	666094000022159362	https://pbs.twimg.com/media/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/CT8g3BpUEAAUFjg.jpg
21	666293911632134144	https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22	666337882303524864	https://pbs.twimg.com/media/CT90wFIWEAMuRje.jpg
23	666345417576210432	https://pbs.twimg.com/media/CT9Vn7PWAAA_ZCM.jpg
24	666353288456101888	https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25	666362758909284353	https://pbs.twimg.com/media/CT9lXGsUcAAyUft.jpg
26	666373753744588802	https://pbs.twimg.com/media/CT9vZEYUAAALZ05.jpg
27	666396247373291520	https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28	666407126856765440	https://pbs.twimg.com/media/CT-NvwmW4AAAugGZ.jpg
29	666411507551481857	https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...
2045	886366144734445568	https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046	886680336477933568	https://pbs.twimg.com/media/DE4fEDzWAAAYHMM.jpg
2047	886736880519319552	https://pbs.twimg.com/media/DE5Se8FXcAAJFfx4.jpg
2048	886983233522544640	https://pbs.twimg.com/media/DE8yicJW0AAAABJ.jpg
2049	887101392804085760	https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050	887343217045368832	https://pbs.twimg.com/ext_tw_video_thumb/88734...
2051	887473957103951883	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2052	887517139158093824	https://pbs.twimg.com/ext_tw_video_thumb/88751...
2053	887705289381826560	https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054	888078434458587136	https://pbs.twimg.com/media/DFMWn56WsAAKA7B.jpg
2055	888202515573088257	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056	888554962724278272	https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg
2057	888804989199671297	https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058	888917238123831296	https://pbs.twimg.com/media/DFYRgsOUQAARGh0.jpg
2059	889278841981685760	https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060	889531135344209921	https://pbs.twimg.com/media/DFg_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 [15]: # Get Data info
         image.info()
```

```

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

```

```
In [16]: # Get summary of our quantative data
         image.describe()
```

```

Out[16]:

```

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

```
In [17]: twitter_api
```

```

Out[17]:

```

	contributors	coordinates	created_at \
1991-02-08 13:48:08.022790149	NaN	NaN	2015-11-15 22:32:08
1991-02-08 16:08:05.002620928	NaN	NaN	2015-11-15 23:05:30
1991-02-08 17:16:52.701032449	NaN	NaN	2015-11-15 23:21:54

1991-02-08 20:17:06.329800704	NaN	NaN 2015-11-16 00:04:52
1991-02-08 21:40:48.165822465	NaN	NaN 2015-11-16 00:24:50
1991-02-08 22:05:58.794694657	NaN	NaN 2015-11-16 00:30:50
1991-02-08 22:24:13.826850816	NaN	NaN 2015-11-16 00:35:11
1991-02-08 23:25:25.042405380	NaN	NaN 2015-11-16 00:49:46
1991-02-08 23:51:30.499244032	NaN	NaN 2015-11-16 00:55:59
1991-02-09 00:16:40.524156928	NaN	NaN 2015-11-16 01:01:59
1991-02-09 01:43:47.256086533	NaN	NaN 2015-11-16 01:22:45
1991-02-09 03:46:33.221509120	NaN	NaN 2015-11-16 01:52:02
1991-02-09 04:18:20.786774016	NaN	NaN 2015-11-16 01:59:36
1991-02-09 07:01:56.733198337	NaN	NaN 2015-11-16 02:38:37
1991-02-09 10:06:40.022159362	NaN	NaN 2015-11-16 03:22:39
1991-02-09 11:38:33.787052032	NaN	NaN 2015-11-16 03:44:34
1991-02-09 12:22:35.909144576	NaN	NaN 2015-11-16 03:55:04
1991-02-09 12:55:33.288665088	NaN	NaN 2015-11-16 04:02:55
1991-02-11 10:41:50.803644416	NaN	NaN 2015-11-16 14:57:41
1991-02-11 11:51:37.616637952	NaN	NaN 2015-11-16 15:14:19
1991-02-11 15:50:06.224695296	NaN	NaN 2015-11-16 16:11:11
1991-02-11 17:38:31.632134144	NaN	NaN 2015-11-16 16:37:02
1991-02-12 05:51:22.303524864	NaN	NaN 2015-11-16 19:31:45
1991-02-12 07:56:57.576210432	NaN	NaN 2015-11-16 20:01:42
1991-02-12 10:08:08.456101888	NaN	NaN 2015-11-16 20:32:58
1991-02-12 12:45:58.909284353	NaN	NaN 2015-11-16 21:10:36
1991-02-12 15:49:13.744588802	NaN	NaN 2015-11-16 21:54:18
1991-02-12 22:04:07.373291520	NaN	NaN 2015-11-16 23:23:41
1991-02-13 01:05:26.856765440	NaN	NaN 2015-11-17 00:06:54
1991-02-13 02:18:27.551481857	NaN	NaN 2015-11-17 00:24:19
...
1998-01-31 17:16:49.285017600	NaN	NaN 2017-07-15 16:51:35
1998-02-01 20:49:04.734445568	NaN	NaN 2017-07-15 23:25:31
1998-02-05 12:05:36.477933568	NaN	NaN 2017-07-16 20:14:00
1998-02-06 03:48:00.519319552	NaN	NaN 2017-07-16 23:58:41
1998-02-09 00:13:53.522544640	NaN	NaN 2017-07-17 16:17:36
1998-02-10 09:03:12.804085760	NaN	NaN 2017-07-18 00:07:08
1998-02-13 04:13:37.045368832	NaN	NaN 2017-07-18 16:08:03
1998-02-14 16:32:37.103951883	NaN	NaN 2017-07-19 00:47:34
1998-02-15 04:32:19.158093824	NaN	NaN 2017-07-19 03:39:09
1998-02-17 08:48:09.381826560	NaN	NaN 2017-07-19 16:06:48
1998-02-21 16:27:14.458587136	NaN	NaN 2017-07-20 16:49:33
1998-02-27 04:49:22.724278272	NaN	NaN 2017-07-22 00:23:06
1998-03-02 02:16:29.199671297	NaN	NaN 2017-07-22 16:56:37
1998-03-03 09:27:18.123831296	NaN	NaN 2017-07-23 00:22:39
1998-03-07 13:54:01.981685760	NaN	NaN 2017-07-24 00:19:32
1998-03-10 11:58:55.344209921	NaN	NaN 2017-07-24 17:02:04
1998-03-11 17:53:57.579907072	NaN	NaN 2017-07-25 00:10:02
1998-03-12 01:16:28.333682689	NaN	NaN 2017-07-25 01:55:32
1998-03-14 13:08:16.479866881	NaN	NaN 2017-07-25 16:11:53
1998-03-16 00:03:28.113172480	NaN	NaN 2017-07-26 00:31:25

1998-03-18 16:57:35.349198849	NaN	NaN 2017-07-26 15:59:51
1998-03-22 23:26:25.150312448	NaN	NaN 2017-07-27 16:25:51
1998-03-24 08:46:21.411237888	NaN	NaN 2017-07-28 00:22:40
1998-03-27 04:11:53.173991426	NaN	NaN 2017-07-28 16:27:12
1998-03-28 12:25:50.875897856	NaN	NaN 2017-07-29 00:08:17
1998-03-31 06:59:18.926688256	NaN	NaN 2017-07-29 16:00:24
1998-04-04 11:32:37.279858688	NaN	NaN 2017-07-30 15:58:51
1998-04-05 22:26:21.378084864	NaN	NaN 2017-07-31 00:18:03
1998-04-10 03:03:41.306343426	NaN	NaN 2017-08-01 00:17:27
1998-04-12 22:37:23.555336193	NaN	NaN 2017-08-01 16:23:56

		entities \
1991-02-08 13:48:08.022790149	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 16:08:05.002620928	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 17:16:52.701032449	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 20:17:06.329800704	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 21:40:48.165822465	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 22:05:58.794694657	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 22:24:13.826850816	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 23:25:25.042405380	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-08 23:51:30.499244032	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 00:16:40.524156928	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 01:43:47.256086533	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 03:46:33.221509120	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 04:18:20.786774016	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 07:01:56.733198337	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 10:06:40.022159362	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 11:38:33.787052032	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 12:22:35.909144576	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-09 12:55:33.288665088	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-11 10:41:50.803644416	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-11 11:51:37.616637952	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-11 15:50:06.224695296	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-11 17:38:31.632134144	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-12 05:51:22.303524864	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-12 07:56:57.576210432	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-12 10:08:08.456101888	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-12 12:45:58.909284353	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-12 15:49:13.744588802	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-12 22:04:07.373291520	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-13 01:05:26.856765440	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-13 02:18:27.551481857	{'hashtags': [], 'media': [{'display_url': 'pi...	
...	...	
1998-01-31 17:16:49.285017600	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1998-02-01 20:49:04.734445568	{'hashtags': [], 'symbols': [], 'urls': [{'dis...	
1998-02-05 12:05:36.477933568	{'hashtags': [], 'media': [{'display_url': 'pi...	
1998-02-06 03:48:00.519319552	{'hashtags': [], 'symbols': [], 'urls': [{'dis...	
1998-02-09 00:13:53.522544640	{'hashtags': [], 'media': [{'display_url': 'pi...	

1998-02-10 09:03:12.804085760	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-02-13 04:13:37.045368832	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-02-14 16:32:37.103951883	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-02-15 04:32:19.158093824	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-02-17 08:48:09.381826560	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-02-21 16:27:14.458587136	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-02-27 04:49:22.724278272	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-03-02 02:16:29.199671297	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-03 09:27:18.123831296	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-03-07 13:54:01.981685760	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-10 11:58:55.344209921	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-11 17:53:57.579907072	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-03-12 01:16:28.333682689	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-03-14 13:08:16.479866881	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-03-16 00:03:28.113172480	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-18 16:57:35.349198849	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-22 23:26:25.150312448	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-24 08:46:21.411237888	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-27 04:11:53.173991426	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-28 12:25:50.875897856	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-03-31 06:59:18.926688256	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-04-04 11:32:37.279858688	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-04-05 22:26:21.378084864	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-04-10 03:03:41.306343426	{'hashtags': [], 'symbols': [], 'urls': [{'dis...
1998-04-12 22:37:23.555336193	{'hashtags': [], 'media': [{'display_url': 'pi...

extended_entities \

1991-02-08 13:48:08.022790149	{'media': [{'display_url': 'pic.twitter.com/BL...
1991-02-08 16:08:05.002620928	{'media': [{'display_url': 'pic.twitter.com/r7...
1991-02-08 17:16:52.701032449	{'media': [{'display_url': 'pic.twitter.com/y6...
1991-02-08 20:17:06.329800704	{'media': [{'display_url': 'pic.twitter.com/DW...
1991-02-08 21:40:48.165822465	{'media': [{'display_url': 'pic.twitter.com/4B...
1991-02-08 22:05:58.794694657	{'media': [{'display_url': 'pic.twitter.com/fv...
1991-02-08 22:24:13.826850816	{'media': [{'display_url': 'pic.twitter.com/v5...
1991-02-08 23:25:25.042405380	{'media': [{'display_url': 'pic.twitter.com/rd...
1991-02-08 23:51:30.499244032	{'media': [{'display_url': 'pic.twitter.com/yW...
1991-02-09 00:16:40.524156928	{'media': [{'display_url': 'pic.twitter.com/pY...
1991-02-09 01:43:47.256086533	{'media': [{'display_url': 'pic.twitter.com/Rh...
1991-02-09 03:46:33.221509120	{'media': [{'display_url': 'pic.twitter.com/Zo...
1991-02-09 04:18:20.786774016	{'media': [{'display_url': 'pic.twitter.com/Yk...
1991-02-09 07:01:56.733198337	{'media': [{'display_url': 'pic.twitter.com/3R...
1991-02-09 10:06:40.022159362	{'media': [{'display_url': 'pic.twitter.com/mn...
1991-02-09 11:38:33.787052032	{'media': [{'display_url': 'pic.twitter.com/Ri...
1991-02-09 12:22:35.909144576	{'media': [{'display_url': 'pic.twitter.com/11...
1991-02-09 12:55:33.288665088	{'media': [{'display_url': 'pic.twitter.com/As...
1991-02-11 10:41:50.803644416	{'media': [{'display_url': 'pic.twitter.com/Oy...
1991-02-11 11:51:37.616637952	{'media': [{'display_url': 'pic.twitter.com/ws...
1991-02-11 15:50:06.224695296	{'media': [{'display_url': 'pic.twitter.com/d9...

1991-02-11 17:38:31.632134144	{'media': [{'display_url': 'pic.twitter.com/II...	
1991-02-12 05:51:22.303524864	{'media': [{'display_url': 'pic.twitter.com/Qp...	
1991-02-12 07:56:57.576210432	{'media': [{'display_url': 'pic.twitter.com/VF...	
1991-02-12 10:08:08.456101888	{'media': [{'display_url': 'pic.twitter.com/tl...	
1991-02-12 12:45:58.909284353	{'media': [{'display_url': 'pic.twitter.com/XM...	
1991-02-12 15:49:13.744588802	{'media': [{'display_url': 'pic.twitter.com/uH...	
1991-02-12 22:04:07.373291520	{'media': [{'display_url': 'pic.twitter.com/Dc...	
1991-02-13 01:05:26.856765440	{'media': [{'display_url': 'pic.twitter.com/Lo...	
1991-02-13 02:18:27.551481857	{'media': [{'display_url': 'pic.twitter.com/aM...	
...		...
1998-01-31 17:16:49.285017600		NaN
1998-02-01 20:49:04.734445568		NaN
1998-02-05 12:05:36.477933568	{'media': [{'display_url': 'pic.twitter.com/BC...	
1998-02-06 03:48:00.519319552		NaN
1998-02-09 00:13:53.522544640	{'media': [{'display_url': 'pic.twitter.com/I6...	
1998-02-10 09:03:12.804085760		NaN
1998-02-13 04:13:37.045368832	{'media': [{'additional_media_info': {'monetiz...	
1998-02-14 16:32:37.103951883	{'media': [{'display_url': 'pic.twitter.com/cL...	
1998-02-15 04:32:19.158093824	{'media': [{'additional_media_info': {'monetiz...	
1998-02-17 08:48:09.381826560		NaN
1998-02-21 16:27:14.458587136		NaN
1998-02-27 04:49:22.724278272	{'media': [{'display_url': 'pic.twitter.com/Yn...	
1998-03-02 02:16:29.199671297		NaN
1998-03-03 09:27:18.123831296	{'media': [{'display_url': 'pic.twitter.com/ch...	
1998-03-07 13:54:01.981685760		NaN
1998-03-10 11:58:55.344209921		NaN
1998-03-11 17:53:57.579907072	{'media': [{'display_url': 'pic.twitter.com/f8...	
1998-03-12 01:16:28.333682689	{'media': [{'display_url': 'pic.twitter.com/Bx...	
1998-03-14 13:08:16.479866881	{'media': [{'display_url': 'pic.twitter.com/u1...	
1998-03-16 00:03:28.113172480		NaN
1998-03-18 16:57:35.349198849		NaN
1998-03-22 23:26:25.150312448		NaN
1998-03-24 08:46:21.411237888		NaN
1998-03-27 04:11:53.173991426		NaN
1998-03-28 12:25:50.875897856		NaN
1998-03-31 06:59:18.926688256		NaN
1998-04-04 11:32:37.279858688	{'media': [{'display_url': 'pic.twitter.com/tD...	
1998-04-05 22:26:21.378084864		NaN
1998-04-10 03:03:41.306343426		NaN
1998-04-12 22:37:23.555336193	{'media': [{'display_url': 'pic.twitter.com/Mg...	

	favorite_count	favorited	geo	\
1991-02-08 13:48:08.022790149	2544	0	NaN	
1991-02-08 16:08:05.002620928	129	0	NaN	
1991-02-08 17:16:52.701032449	125	0	NaN	
1991-02-08 20:17:06.329800704	296	0	NaN	
1991-02-08 21:40:48.165822465	107	0	NaN	
1991-02-08 22:05:58.794694657	132	0	NaN	

1991-02-08 22:24:13.826850816	1209	0	NaN
1991-02-08 23:25:25.042405380	432	0	NaN
1991-02-08 23:51:30.499244032	294	0	NaN
1991-02-09 00:16:40.524156928	111	0	NaN
1991-02-09 01:43:47.256086533	471	0	NaN
1991-02-09 03:46:33.221509120	145	0	NaN
1991-02-09 04:18:20.786774016	318	0	NaN
1991-02-09 07:01:56.733198337	116	0	NaN
1991-02-09 10:06:40.022159362	163	0	NaN
1991-02-09 11:38:33.787052032	153	0	NaN
1991-02-09 12:22:35.909144576	80	0	NaN
1991-02-09 12:55:33.288665088	14161	0	NaN
1991-02-11 10:41:50.803644416	102	0	NaN
1991-02-11 11:51:37.616637952	174	0	NaN
1991-02-11 15:50:06.224695296	146	0	NaN
1991-02-11 17:38:31.632134144	494	0	NaN
1991-02-12 05:51:22.303524864	196	0	NaN
1991-02-12 07:56:57.576210432	290	0	NaN
1991-02-12 10:08:08.456101888	217	0	NaN
1991-02-12 12:45:58.909284353	770	0	NaN
1991-02-12 15:49:13.744588802	186	0	NaN
1991-02-12 22:04:07.373291520	167	0	NaN
1991-02-13 01:05:26.856765440	107	0	NaN
1991-02-13 02:18:27.551481857	440	0	NaN
...
1998-01-31 17:16:49.285017600	116	0	NaN
1998-02-01 20:49:04.734445568	20885	0	NaN
1998-02-05 12:05:36.477933568	22114	0	NaN
1998-02-06 03:48:00.519319552	11873	0	NaN
1998-02-09 00:13:53.522544640	34661	0	NaN
1998-02-10 09:03:12.804085760	30131	0	NaN
1998-02-13 04:13:37.045368832	33204	0	NaN
1998-02-14 16:32:37.103951883	68096	0	NaN
1998-02-15 04:32:19.158093824	45602	0	NaN
1998-02-17 08:48:09.381826560	29744	0	NaN
1998-02-21 16:27:14.458587136	21445	0	NaN
1998-02-27 04:49:22.724278272	19547	0	NaN
1998-03-02 02:16:29.199671297	25200	0	NaN
1998-03-03 09:27:18.123831296	28673	0	NaN
1998-03-07 13:54:01.981685760	24888	0	NaN
1998-03-10 11:58:55.344209921	14886	0	NaN
1998-03-11 17:53:57.579907072	26773	0	NaN
1998-03-12 01:16:28.333682689	47431	0	NaN
1998-03-14 13:08:16.479866881	27396	0	NaN
1998-03-16 00:03:28.113172480	30239	0	NaN
1998-03-18 16:57:35.349198849	31452	0	NaN
1998-03-22 23:26:25.150312448	27402	0	NaN
1998-03-24 08:46:21.411237888	64480	0	NaN

1998-03-27 04:11:53.173991426	11670	0	NaN
1998-03-28 12:25:50.875897856	19938	0	NaN
1998-03-31 06:59:18.926688256	39725	0	NaN
1998-04-04 11:32:37.279858688	41532	0	NaN
1998-04-05 22:26:21.378084864	24670	0	NaN
1998-04-10 03:03:41.306343426	32763	0	NaN
1998-04-12 22:37:23.555336193	38210	0	NaN

	id	id_str \
1991-02-08 13:48:08.022790149	666020888022790144	666020888022790144
1991-02-08 16:08:05.002620928	666029285002620928	666029285002620928
1991-02-08 17:16:52.701032449	666033412701032448	666033412701032448
1991-02-08 20:17:06.329800704	666044226329800704	666044226329800704
1991-02-08 21:40:48.165822465	666049248165822464	666049248165822464
1991-02-08 22:05:58.794694657	666050758794694656	666050758794694656
1991-02-08 22:24:13.826850816	666051853826850816	666051853826850816
1991-02-08 23:25:25.042405380	666055525042405376	666055525042405376
1991-02-08 23:51:30.499244032	666057090499244032	666057090499244032
1991-02-09 00:16:40.524156928	666058600524156928	666058600524156928
1991-02-09 01:43:47.256086533	666063827256086528	666063827256086528
1991-02-09 03:46:33.221509120	666071193221509120	666071193221509120
1991-02-09 04:18:20.786774016	666073100786774016	666073100786774016
1991-02-09 07:01:56.733198337	666082916733198336	666082916733198336
1991-02-09 10:06:40.022159362	666094000022159360	666094000022159360
1991-02-09 11:38:33.787052032	666099513787052032	666099513787052032
1991-02-09 12:22:35.909144576	666102155909144576	666102155909144576
1991-02-09 12:55:33.288665088	666104133288665088	666104133288665088
1991-02-11 10:41:50.803644416	666268910803644416	666268910803644416
1991-02-11 11:51:37.616637952	666273097616637952	666273097616637952
1991-02-11 15:50:06.224695296	666287406224695296	666287406224695296
1991-02-11 17:38:31.632134144	666293911632134144	666293911632134144
1991-02-12 05:51:22.303524864	666337882303524864	666337882303524864
1991-02-12 07:56:57.576210432	666345417576210432	666345417576210432
1991-02-12 10:08:08.456101888	666353288456101888	666353288456101888
1991-02-12 12:45:58.909284353	666362758909284352	666362758909284352
1991-02-12 15:49:13.744588802	666373753744588800	666373753744588800
1991-02-12 22:04:07.373291520	666396247373291520	666396247373291520
1991-02-13 01:05:26.856765440	666407126856765440	666407126856765440
1991-02-13 02:18:27.551481857	666411507551481856	666411507551481856
...
1998-01-31 17:16:49.285017600	886267009285017600	886267009285017600
1998-02-01 20:49:04.734445568	886366144734445568	886366144734445568
1998-02-05 12:05:36.477933568	886680336477933568	886680336477933568
1998-02-06 03:48:00.519319552	886736880519319552	886736880519319552
1998-02-09 00:13:53.522544640	886983233522544640	886983233522544640
1998-02-10 09:03:12.804085760	887101392804085760	887101392804085760
1998-02-13 04:13:37.045368832	887343217045368832	887343217045368832
1998-02-14 16:32:37.103951883	887473957103951872	887473957103951872

1998-02-15	04:32:19.158093824	887517139158093824	887517139158093824
1998-02-17	08:48:09.381826560	887705289381826560	887705289381826560
1998-02-21	16:27:14.458587136	888078434458587136	888078434458587136
1998-02-27	04:49:22.724278272	888554962724278272	888554962724278272
1998-03-02	02:16:29.199671297	888804989199671296	888804989199671296
1998-03-03	09:27:18.123831296	888917238123831296	888917238123831296
1998-03-07	13:54:01.981685760	889278841981685760	889278841981685760
1998-03-10	11:58:55.344209921	889531135344209920	889531135344209920
1998-03-11	17:53:57.579907072	889638837579907072	889638837579907072
1998-03-12	01:16:28.333682689	889665388333682688	889665388333682688
1998-03-14	13:08:16.479866881	889880896479866880	889880896479866880
1998-03-16	00:03:28.113172480	890006608113172480	890006608113172480
1998-03-18	16:57:35.349198849	890240255349198848	890240255349198848
1998-03-22	23:26:25.150312448	890609185150312448	890609185150312448
1998-03-24	08:46:21.411237888	890729181411237888	890729181411237888
1998-03-27	04:11:53.173991426	890971913173991424	890971913173991424
1998-03-28	12:25:50.875897856	891087950875897856	891087950875897856
1998-03-31	06:59:18.926688256	891327558926688256	891327558926688256
1998-04-04	11:32:37.279858688	891689557279858688	891689557279858688
1998-04-05	22:26:21.378084864	891815181378084864	891815181378084864
1998-04-10	03:03:41.306343426	892177421306343424	892177421306343424
1998-04-12	22:37:23.555336193	892420643555336192	892420643555336192

1991-02-08	13:48:08.022790149	...
1991-02-08	16:08:05.002620928	...
1991-02-08	17:16:52.701032449	...
1991-02-08	20:17:06.329800704	...
1991-02-08	21:40:48.165822465	...
1991-02-08	22:05:58.794694657	...
1991-02-08	22:24:13.826850816	...
1991-02-08	23:25:25.042405380	...
1991-02-08	23:51:30.499244032	...
1991-02-09	00:16:40.524156928	...
1991-02-09	01:43:47.256086533	...
1991-02-09	03:46:33.221509120	...
1991-02-09	04:18:20.786774016	...
1991-02-09	07:01:56.733198337	...
1991-02-09	10:06:40.022159362	...
1991-02-09	11:38:33.787052032	...
1991-02-09	12:22:35.909144576	...
1991-02-09	12:55:33.288665088	...
1991-02-11	10:41:50.803644416	...
1991-02-11	11:51:37.616637952	...
1991-02-11	15:50:06.224695296	...
1991-02-11	17:38:31.632134144	...
1991-02-12	05:51:22.303524864	...
1991-02-12	07:56:57.576210432	...

1991-02-12 10:08:08.456101888	...
1991-02-12 12:45:58.909284353	...
1991-02-12 15:49:13.744588802	...
1991-02-12 22:04:07.373291520	...
1991-02-13 01:05:26.856765440	...
1991-02-13 02:18:27.551481857	...
...	...
1998-01-31 17:16:49.285017600	...
1998-02-01 20:49:04.734445568	...
1998-02-05 12:05:36.477933568	...
1998-02-06 03:48:00.519319552	...
1998-02-09 00:13:53.522544640	...
1998-02-10 09:03:12.804085760	...
1998-02-13 04:13:37.045368832	...
1998-02-14 16:32:37.103951883	...
1998-02-15 04:32:19.158093824	...
1998-02-17 08:48:09.381826560	...
1998-02-21 16:27:14.458587136	...
1998-02-27 04:49:22.724278272	...
1998-03-02 02:16:29.199671297	...
1998-03-03 09:27:18.123831296	...
1998-03-07 13:54:01.981685760	...
1998-03-10 11:58:55.344209921	...
1998-03-11 17:53:57.579907072	...
1998-03-12 01:16:28.333682689	...
1998-03-14 13:08:16.479866881	...
1998-03-16 00:03:28.113172480	...
1998-03-18 16:57:35.349198849	...
1998-03-22 23:26:25.150312448	...
1998-03-24 08:46:21.411237888	...
1998-03-27 04:11:53.173991426	...
1998-03-28 12:25:50.875897856	...
1998-03-31 06:59:18.926688256	...
1998-04-04 11:32:37.279858688	...
1998-04-05 22:26:21.378084864	...
1998-04-10 03:03:41.306343426	...
1998-04-12 22:37:23.555336193	...

	quoted_status	quoted_status_id \
1991-02-08 13:48:08.022790149	NaN	NaN
1991-02-08 16:08:05.002620928	NaN	NaN
1991-02-08 17:16:52.701032449	NaN	NaN
1991-02-08 20:17:06.329800704	NaN	NaN
1991-02-08 21:40:48.165822465	NaN	NaN
1991-02-08 22:05:58.794694657	NaN	NaN
1991-02-08 22:24:13.826850816	NaN	NaN
1991-02-08 23:25:25.042405380	NaN	NaN
1991-02-08 23:51:30.499244032	NaN	NaN

1991-02-09 00:16:40.524156928	NaN	NaN
1991-02-09 01:43:47.256086533	NaN	NaN
1991-02-09 03:46:33.221509120	NaN	NaN
1991-02-09 04:18:20.786774016	NaN	NaN
1991-02-09 07:01:56.733198337	NaN	NaN
1991-02-09 10:06:40.022159362	NaN	NaN
1991-02-09 11:38:33.787052032	NaN	NaN
1991-02-09 12:22:35.909144576	NaN	NaN
1991-02-09 12:55:33.288665088	NaN	NaN
1991-02-11 10:41:50.803644416	NaN	NaN
1991-02-11 11:51:37.616637952	NaN	NaN
1991-02-11 15:50:06.224695296	NaN	NaN
1991-02-11 17:38:31.632134144	NaN	NaN
1991-02-12 05:51:22.303524864	NaN	NaN
1991-02-12 07:56:57.576210432	NaN	NaN
1991-02-12 10:08:08.456101888	NaN	NaN
1991-02-12 12:45:58.909284353	NaN	NaN
1991-02-12 15:49:13.744588802	NaN	NaN
1991-02-12 22:04:07.373291520	NaN	NaN
1991-02-13 01:05:26.856765440	NaN	NaN
1991-02-13 02:18:27.551481857	NaN	NaN
...
1998-01-31 17:16:49.285017600	NaN	NaN
1998-02-01 20:49:04.734445568	NaN	NaN
1998-02-05 12:05:36.477933568	NaN	NaN
1998-02-06 03:48:00.519319552	NaN	NaN
1998-02-09 00:13:53.522544640	NaN	NaN
1998-02-10 09:03:12.804085760	NaN	NaN
1998-02-13 04:13:37.045368832	NaN	NaN
1998-02-14 16:32:37.103951883	NaN	NaN
1998-02-15 04:32:19.158093824	NaN	NaN
1998-02-17 08:48:09.381826560	NaN	NaN
1998-02-21 16:27:14.458587136	NaN	NaN
1998-02-27 04:49:22.724278272	NaN	NaN
1998-03-02 02:16:29.199671297	NaN	NaN
1998-03-03 09:27:18.123831296	NaN	NaN
1998-03-07 13:54:01.981685760	NaN	NaN
1998-03-10 11:58:55.344209921	NaN	NaN
1998-03-11 17:53:57.579907072	NaN	NaN
1998-03-12 01:16:28.333682689	NaN	NaN
1998-03-14 13:08:16.479866881	NaN	NaN
1998-03-16 00:03:28.113172480	NaN	NaN
1998-03-18 16:57:35.349198849	NaN	NaN
1998-03-22 23:26:25.150312448	NaN	NaN
1998-03-24 08:46:21.411237888	NaN	NaN
1998-03-27 04:11:53.173991426	NaN	NaN
1998-03-28 12:25:50.875897856	NaN	NaN
1998-03-31 06:59:18.926688256	NaN	NaN

1998-04-04 11:32:37.279858688	NaN	NaN
1998-04-05 22:26:21.378084864	NaN	NaN
1998-04-10 03:03:41.306343426	NaN	NaN
1998-04-12 22:37:23.555336193	NaN	NaN

	quoted_status_id_str	retweet_count	retweeted	\
1991-02-08 13:48:08.022790149	NaN	508	0	
1991-02-08 16:08:05.002620928	NaN	47	0	
1991-02-08 17:16:52.701032449	NaN	44	0	
1991-02-08 20:17:06.329800704	NaN	139	0	
1991-02-08 21:40:48.165822465	NaN	40	0	
1991-02-08 22:05:58.794694657	NaN	59	0	
1991-02-08 22:24:13.826850816	NaN	841	0	
1991-02-08 23:25:25.042405380	NaN	244	0	
1991-02-08 23:51:30.499244032	NaN	141	0	
1991-02-09 00:16:40.524156928	NaN	57	0	
1991-02-09 01:43:47.256086533	NaN	215	0	
1991-02-09 03:46:33.221509120	NaN	59	0	
1991-02-09 04:18:20.786774016	NaN	162	0	
1991-02-09 07:01:56.733198337	NaN	44	0	
1991-02-09 10:06:40.022159362	NaN	72	0	
1991-02-09 11:38:33.787052032	NaN	67	0	
1991-02-09 12:22:35.909144576	NaN	12	0	
1991-02-09 12:55:33.288665088	NaN	6482	0	
1991-02-11 10:41:50.803644416	NaN	35	0	
1991-02-11 11:51:37.616637952	NaN	76	0	
1991-02-11 15:50:06.224695296	NaN	64	0	
1991-02-11 17:38:31.632134144	NaN	352	0	
1991-02-12 05:51:22.303524864	NaN	90	0	
1991-02-12 07:56:57.576210432	NaN	133	0	
1991-02-12 10:08:08.456101888	NaN	71	0	
1991-02-12 12:45:58.909284353	NaN	564	0	
1991-02-12 15:49:13.744588802	NaN	92	0	
1991-02-12 22:04:07.373291520	NaN	84	0	
1991-02-13 01:05:26.856765440	NaN	40	0	
1991-02-13 02:18:27.551481857	NaN	323	0	
...	
1998-01-31 17:16:49.285017600	NaN	4	0	
1998-02-01 20:49:04.734445568	NaN	3152	0	
1998-02-05 12:05:36.477933568	NaN	4396	0	
1998-02-06 03:48:00.519319552	NaN	3229	0	
1998-02-09 00:13:53.522544640	NaN	7654	0	
1998-02-10 09:03:12.804085760	NaN	5873	0	
1998-02-13 04:13:37.045368832	NaN	10244	0	
1998-02-14 16:32:37.103951883	NaN	17907	0	
1998-02-15 04:32:19.158093824	NaN	11487	0	
1998-02-17 08:48:09.381826560	NaN	5304	0	
1998-02-21 16:27:14.458587136	NaN	3444	0	

1998-02-27 04:49:22.724278272	NaN	3493	0
1998-03-02 02:16:29.199671297	NaN	4225	0
1998-03-03 09:27:18.123831296	NaN	4423	0
1998-03-07 13:54:01.981685760	NaN	5290	0
1998-03-10 11:58:55.344209921	NaN	2208	0
1998-03-11 17:53:57.579907072	NaN	4473	0
1998-03-12 01:16:28.333682689	NaN	9898	0
1998-03-14 13:08:16.479866881	NaN	4898	0
1998-03-16 00:03:28.113172480	NaN	7218	0
1998-03-18 16:57:35.349198849	NaN	7271	0
1998-03-22 23:26:25.150312448	NaN	4206	0
1998-03-24 08:46:21.411237888	NaN	18570	0
1998-03-27 04:11:53.173991426	NaN	2032	0
1998-03-28 12:25:50.875897856	NaN	3065	0
1998-03-31 06:59:18.926688256	NaN	9213	0
1998-04-04 11:32:37.279858688	NaN	8505	0
1998-04-05 22:26:21.378084864	NaN	4090	0
1998-04-10 03:03:41.306343426	NaN	6179	0
1998-04-12 22:37:23.555336193	NaN	8370	0

	retweeted_status \
1991-02-08 13:48:08.022790149	NaN
1991-02-08 16:08:05.002620928	NaN
1991-02-08 17:16:52.701032449	NaN
1991-02-08 20:17:06.329800704	NaN
1991-02-08 21:40:48.165822465	NaN
1991-02-08 22:05:58.794694657	NaN
1991-02-08 22:24:13.826850816	NaN
1991-02-08 23:25:25.042405380	NaN
1991-02-08 23:51:30.499244032	NaN
1991-02-09 00:16:40.524156928	NaN
1991-02-09 01:43:47.256086533	NaN
1991-02-09 03:46:33.221509120	NaN
1991-02-09 04:18:20.786774016	NaN
1991-02-09 07:01:56.733198337	NaN
1991-02-09 10:06:40.022159362	NaN
1991-02-09 11:38:33.787052032	NaN
1991-02-09 12:22:35.909144576	NaN
1991-02-09 12:55:33.288665088	NaN
1991-02-11 10:41:50.803644416	NaN
1991-02-11 11:51:37.616637952	NaN
1991-02-11 15:50:06.224695296	NaN
1991-02-11 17:38:31.632134144	NaN
1991-02-12 05:51:22.303524864	NaN
1991-02-12 07:56:57.576210432	NaN
1991-02-12 10:08:08.456101888	NaN
1991-02-12 12:45:58.909284353	NaN
1991-02-12 15:49:13.744588802	NaN

1991-02-12 22:04:07.373291520	NaN
1991-02-13 01:05:26.856765440	NaN
1991-02-13 02:18:27.551481857	NaN
...	...
1998-01-31 17:16:49.285017600	NaN
1998-02-01 20:49:04.734445568	NaN
1998-02-05 12:05:36.477933568	NaN
1998-02-06 03:48:00.519319552	NaN
1998-02-09 00:13:53.522544640	NaN
1998-02-10 09:03:12.804085760	NaN
1998-02-13 04:13:37.045368832	NaN
1998-02-14 16:32:37.103951883	NaN
1998-02-15 04:32:19.158093824	NaN
1998-02-17 08:48:09.381826560	NaN
1998-02-21 16:27:14.458587136	NaN
1998-02-27 04:49:22.724278272	NaN
1998-03-02 02:16:29.199671297	NaN
1998-03-03 09:27:18.123831296	NaN
1998-03-07 13:54:01.981685760	NaN
1998-03-10 11:58:55.344209921	NaN
1998-03-11 17:53:57.579907072	NaN
1998-03-12 01:16:28.333682689	NaN
1998-03-14 13:08:16.479866881	NaN
1998-03-16 00:03:28.113172480	NaN
1998-03-18 16:57:35.349198849	NaN
1998-03-22 23:26:25.150312448	NaN
1998-03-24 08:46:21.411237888	NaN
1998-03-27 04:11:53.173991426	NaN
1998-03-28 12:25:50.875897856	NaN
1998-03-31 06:59:18.926688256	NaN
1998-04-04 11:32:37.279858688	NaN
1998-04-05 22:26:21.378084864	NaN
1998-04-10 03:03:41.306343426	NaN
1998-04-12 22:37:23.555336193	NaN

	source \
1991-02-08 13:48:08.022790149	<a href="http://twitter.com/download/iphone" r...
1991-02-08 16:08:05.002620928	<a href="http://twitter.com/download/iphone" r...
1991-02-08 17:16:52.701032449	<a href="http://twitter.com/download/iphone" r...
1991-02-08 20:17:06.329800704	<a href="http://twitter.com/download/iphone" r...
1991-02-08 21:40:48.165822465	<a href="http://twitter.com/download/iphone" r...
1991-02-08 22:05:58.794694657	<a href="http://twitter.com/download/iphone" r...
1991-02-08 22:24:13.826850816	<a href="http://twitter.com/download/iphone" r...
1991-02-08 23:25:25.042405380	<a href="http://twitter.com/download/iphone" r...
1991-02-08 23:51:30.499244032	<a href="http://twitter.com/download/iphone" r...
1991-02-09 00:16:40.524156928	<a href="http://twitter.com/download/iphone" r...
1991-02-09 01:43:47.256086533	<a href="http://twitter.com/download/iphone" r...
1991-02-09 03:46:33.221509120	<a href="http://twitter.com/download/iphone" r...

1991-02-09 04:18:20.786774016	<a href="http://twitter.com/download/iphone" r...
1991-02-09 07:01:56.733198337	<a href="http://twitter.com/download/iphone" r...
1991-02-09 10:06:40.022159362	<a href="http://twitter.com/download/iphone" r...
1991-02-09 11:38:33.787052032	<a href="http://twitter.com/download/iphone" r...
1991-02-09 12:22:35.909144576	<a href="http://twitter.com/download/iphone" r...
1991-02-09 12:55:33.288665088	<a href="http://twitter.com/download/iphone" r...
1991-02-11 10:41:50.803644416	<a href="http://twitter.com/download/iphone" r...
1991-02-11 11:51:37.616637952	<a href="http://twitter.com/download/iphone" r...
1991-02-11 15:50:06.224695296	<a href="http://twitter.com/download/iphone" r...
1991-02-11 17:38:31.632134144	<a href="http://twitter.com/download/iphone" r...
1991-02-12 05:51:22.303524864	<a href="http://twitter.com/download/iphone" r...
1991-02-12 07:56:57.576210432	<a href="http://twitter.com/download/iphone" r...
1991-02-12 10:08:08.456101888	<a href="http://twitter.com/download/iphone" r...
1991-02-12 12:45:58.909284353	<a href="http://twitter.com/download/iphone" r...
1991-02-12 15:49:13.744588802	<a href="http://twitter.com/download/iphone" r...
1991-02-12 22:04:07.373291520	<a href="http://twitter.com/download/iphone" r...
1991-02-13 01:05:26.856765440	<a href="http://twitter.com/download/iphone" r...
1991-02-13 02:18:27.551481857	<a href="http://twitter.com/download/iphone" r...
...	...
1998-01-31 17:16:49.285017600	<a href="http://twitter.com/download/iphone" r...
1998-02-01 20:49:04.734445568	<a href="http://twitter.com/download/iphone" r...
1998-02-05 12:05:36.477933568	<a href="http://twitter.com/download/iphone" r...
1998-02-06 03:48:00.519319552	<a href="http://twitter.com/download/iphone" r...
1998-02-09 00:13:53.522544640	<a href="http://twitter.com/download/iphone" r...
1998-02-10 09:03:12.804085760	<a href="http://twitter.com/download/iphone" r...
1998-02-13 04:13:37.045368832	<a href="http://twitter.com/download/iphone" r...
1998-02-14 16:32:37.103951883	<a href="http://twitter.com/download/iphone" r...
1998-02-15 04:32:19.158093824	<a href="http://twitter.com/download/iphone" r...
1998-02-17 08:48:09.381826560	<a href="http://twitter.com/download/iphone" r...
1998-02-21 16:27:14.458587136	<a href="http://twitter.com/download/iphone" r...
1998-02-27 04:49:22.724278272	<a href="http://twitter.com/download/iphone" r...
1998-03-02 02:16:29.199671297	<a href="http://twitter.com/download/iphone" r...
1998-03-03 09:27:18.123831296	<a href="http://twitter.com/download/iphone" r...
1998-03-07 13:54:01.981685760	<a href="http://twitter.com/download/iphone" r...
1998-03-10 11:58:55.344209921	<a href="http://twitter.com/download/iphone" r...
1998-03-11 17:53:57.579907072	<a href="http://twitter.com/download/iphone" r...
1998-03-12 01:16:28.333682689	<a href="http://twitter.com/download/iphone" r...
1998-03-14 13:08:16.479866881	<a href="http://twitter.com/download/iphone" r...
1998-03-16 00:03:28.113172480	<a href="http://twitter.com/download/iphone" r...
1998-03-18 16:57:35.349198849	<a href="http://twitter.com/download/iphone" r...
1998-03-22 23:26:25.150312448	<a href="http://twitter.com/download/iphone" r...
1998-03-24 08:46:21.411237888	<a href="http://twitter.com/download/iphone" r...
1998-03-27 04:11:53.173991426	<a href="http://twitter.com/download/iphone" r...
1998-03-28 12:25:50.875897856	<a href="http://twitter.com/download/iphone" r...
1998-03-31 06:59:18.926688256	<a href="http://twitter.com/download/iphone" r...
1998-04-04 11:32:37.279858688	<a href="http://twitter.com/download/iphone" r...
1998-04-05 22:26:21.378084864	<a href="http://twitter.com/download/iphone" r...
1998-04-10 03:03:41.306343426	<a href="http://twitter.com/download/iphone" r...

1998-04-12 22:37:23.555336193 <a href="http://twitter.com/download/iphone" r...

text \

1991-02-08 13:48:08.022790149 Here we have a Japanese Irish Setter. Lost eye...

1991-02-08 16:08:05.002620928 This is a western brown Mitsubishi terrier. Up...

1991-02-08 17:16:52.701032449 Here is a very happy pup. Big fan of well-main...

1991-02-08 20:17:06.329800704 This is a purebred Piers Morgan. Loves to Netf...

1991-02-08 21:40:48.165822465 Here we have a 1949 1st generation vulpix. Enj...

1991-02-08 22:05:58.794694657 This is a truly beautiful English Wilson Staff...

1991-02-08 22:24:13.826850816 This is an odd dog. Hard on the outside but lo...

1991-02-08 23:25:25.042405380 Here is a Siberian heavily armored polar bear ...

1991-02-08 23:51:30.499244032 My oh my. This is a rare blond Canadian terrie...

1991-02-09 00:16:40.524156928 Here is the Rand Paul of retrievers folks! He'...

1991-02-09 01:43:47.256086533 This is the happiest dog you will ever see. Ve...

1991-02-09 03:46:33.221509120 Here we have a northern speckled Rhododendron...

1991-02-09 04:18:20.786774016 Let's hope this flight isn't Malaysian (lol). ...

1991-02-09 07:01:56.733198337 Here we have a well-established sunblockerspan...

1991-02-09 10:06:40.022159362 This appears to be a Mongolian Presbyterian mi...

1991-02-09 11:38:33.787052032 Can stand on stump for what seems like a while...

1991-02-09 12:22:35.909144576 Oh my. Here you are seeing an Adobe Setter giv...

1991-02-09 12:55:33.288665088 Not familiar with this breed. No tail (weird)...

1991-02-11 10:41:50.803644416 Very concerned about fellow dog trapped in com...

1991-02-11 11:51:37.616637952 Can take selfies 11/10 <https://t.co/ws2AMaWpW>

1991-02-11 15:50:06.224695296 This is an Albanian 3 1/2 legged Episcopalian...

1991-02-11 17:38:31.632134144 This is a funny dog. Weird toes. Won't come do...

1991-02-12 05:51:22.303524864 This is an extremely rare horned Parthenon. No...

1991-02-12 07:56:57.576210432 Look at this jokester thinking seat belt laws ...

1991-02-12 10:08:08.456101888 Here we have a mixed Asiago from the Galápagos...

1991-02-12 12:45:58.909284353 Unique dog here. Very small. Lives in containe...

1991-02-12 15:49:13.744588802 Those are sunglasses and a jean jacket. 11/10 ...

1991-02-12 22:04:07.373291520 Oh goodness. A super rare northeast Qdoba kang...

1991-02-13 01:05:26.856765440 This is a southern Vesuvius bumblegruff. Can d...

1991-02-13 02:18:27.551481857 This is quite the dog. Gets really excited whe...

...

1998-01-31 17:16:49.285017600 @NonWhiteHat @MayhewMayhem omg hello tanner yo...

1998-02-01 20:49:04.734445568 This is Roscoe. Another pupper fallen victim t...

1998-02-05 12:05:36.477933568 This is Derek. He's late for a dog meeting. 13...

1998-02-06 03:48:00.519319552 This is Mingus. He's a wonderful father to his...

1998-02-09 00:13:53.522544640 This is Maya. She's very shy. Rarely leaves he...

1998-02-10 09:03:12.804085760 This... is a Jubilant Antarctic House Bear. We...

1998-02-13 04:13:37.045368832 You may not have known you needed to see this ...

1998-02-14 16:32:37.103951883 This is Canela. She attempted some fancy porch...

1998-02-15 04:32:19.158093824 I've yet to rate a Venezuelan Hover Wiener. Th...

1998-02-17 08:48:09.381826560 This is Jeffrey. He has a monopoly on the pool...

1998-02-21 16:27:14.458587136 This is Gerald. He was just told he didn't get...

1998-02-27 04:49:22.724278272 This is Ralphus. He's powering up. Attempting ...

1998-03-02 02:16:29.199671297 This is Zeke. He has a new stick. Very proud o...

1998-03-03 09:27:18.123831296 This is Jim. He found a fren. Taught him how t...

1998-03-07 13:54:01.981685760	This is Oliver. You're witnessing one of his m...
1998-03-10 11:58:55.344209921	This is Stuart. He's sporting his favorite fan...
1998-03-11 17:53:57.579907072	This is Ted. He does his best. Sometimes that'...
1998-03-12 01:16:28.333682689	Here's a puppo that seems to be on the fence a...
1998-03-14 13:08:16.479866881	This is Bruno. He is a service shark. Only get...
1998-03-16 00:03:28.113172480	This is Koda. He is a South Australian decksha...
1998-03-18 16:57:35.349198849	This is Cassie. She is a college pup. Studying...
1998-03-22 23:26:25.150312448	This is Zoey. She doesn't want to be one of th...
1998-03-24 08:46:21.411237888	When you watch your owner call another dog a g...
1998-03-27 04:11:53.173991426	Meet Jax. He enjoys ice cream so much he gets ...
1998-03-28 12:25:50.875897856	Here we have a majestic great white breaching ...
1998-03-31 06:59:18.926688256	This is Franklin. He would like you to stop ca...
1998-04-04 11:32:37.279858688	This is Darla. She commenced a snooze mid meal...
1998-04-05 22:26:21.378084864	This is Archie. He is a rare Norwegian Pouncin...
1998-04-10 03:03:41.306343426	This is Tilly. She's just checking pup on you...
1998-04-12 22:37:23.555336193	This is Phineas. He's a mystical boy. Only eve...

	truncated \
1991-02-08 13:48:08.022790149	0
1991-02-08 16:08:05.002620928	0
1991-02-08 17:16:52.701032449	0
1991-02-08 20:17:06.329800704	0
1991-02-08 21:40:48.165822465	0
1991-02-08 22:05:58.794694657	0
1991-02-08 22:24:13.826850816	0
1991-02-08 23:25:25.042405380	0
1991-02-08 23:51:30.499244032	0
1991-02-09 00:16:40.524156928	0
1991-02-09 01:43:47.256086533	0
1991-02-09 03:46:33.221509120	0
1991-02-09 04:18:20.786774016	0
1991-02-09 07:01:56.733198337	0
1991-02-09 10:06:40.022159362	0
1991-02-09 11:38:33.787052032	0
1991-02-09 12:22:35.909144576	0
1991-02-09 12:55:33.288665088	0
1991-02-11 10:41:50.803644416	0
1991-02-11 11:51:37.616637952	0
1991-02-11 15:50:06.224695296	0
1991-02-11 17:38:31.632134144	0
1991-02-12 05:51:22.303524864	0
1991-02-12 07:56:57.576210432	0
1991-02-12 10:08:08.456101888	0
1991-02-12 12:45:58.909284353	0
1991-02-12 15:49:13.744588802	0
1991-02-12 22:04:07.373291520	0
1991-02-13 01:05:26.856765440	0
1991-02-13 02:18:27.551481857	0

```

...
1998-01-31 17:16:49.285017600 0
1998-02-01 20:49:04.734445568 1
1998-02-05 12:05:36.477933568 0
1998-02-06 03:48:00.519319552 1
1998-02-09 00:13:53.522544640 0
1998-02-10 09:03:12.804085760 1
1998-02-13 04:13:37.045368832 0
1998-02-14 16:32:37.103951883 0
1998-02-15 04:32:19.158093824 0
1998-02-17 08:48:09.381826560 1
1998-02-21 16:27:14.458587136 1
1998-02-27 04:49:22.724278272 0
1998-03-02 02:16:29.199671297 1
1998-03-03 09:27:18.123831296 0
1998-03-07 13:54:01.981685760 1
1998-03-10 11:58:55.344209921 1
1998-03-11 17:53:57.579907072 0
1998-03-12 01:16:28.333682689 0
1998-03-14 13:08:16.479866881 0
1998-03-16 00:03:28.113172480 1
1998-03-18 16:57:35.349198849 1
1998-03-22 23:26:25.150312448 1
1998-03-24 08:46:21.411237888 1
1998-03-27 04:11:53.173991426 1
1998-03-28 12:25:50.875897856 1
1998-03-31 06:59:18.926688256 1
1998-04-04 11:32:37.279858688 0
1998-04-05 22:26:21.378084864 1
1998-04-10 03:03:41.306343426 1
1998-04-12 22:37:23.555336193 0

```

```

...
user
1991-02-08 13:48:08.022790149 {'contributors_enabled': False, 'created_at': ...
1991-02-08 16:08:05.002620928 {'contributors_enabled': False, 'created_at': ...
1991-02-08 17:16:52.701032449 {'contributors_enabled': False, 'created_at': ...
1991-02-08 20:17:06.329800704 {'contributors_enabled': False, 'created_at': ...
1991-02-08 21:40:48.165822465 {'contributors_enabled': False, 'created_at': ...
1991-02-08 22:05:58.794694657 {'contributors_enabled': False, 'created_at': ...
1991-02-08 22:24:13.826850816 {'contributors_enabled': False, 'created_at': ...
1991-02-08 23:25:25.042405380 {'contributors_enabled': False, 'created_at': ...
1991-02-08 23:51:30.499244032 {'contributors_enabled': False, 'created_at': ...
1991-02-09 00:16:40.524156928 {'contributors_enabled': False, 'created_at': ...
1991-02-09 01:43:47.256086533 {'contributors_enabled': False, 'created_at': ...
1991-02-09 03:46:33.221509120 {'contributors_enabled': False, 'created_at': ...
1991-02-09 04:18:20.786774016 {'contributors_enabled': False, 'created_at': ...
1991-02-09 07:01:56.733198337 {'contributors_enabled': False, 'created_at': ...
1991-02-09 10:06:40.022159362 {'contributors_enabled': False, 'created_at': ...

```

1991-02-09 11:38:33.787052032	{'contributors_enabled': False, 'created_at': ...
1991-02-09 12:22:35.909144576	{'contributors_enabled': False, 'created_at': ...
1991-02-09 12:55:33.288665088	{'contributors_enabled': False, 'created_at': ...
1991-02-11 10:41:50.803644416	{'contributors_enabled': False, 'created_at': ...
1991-02-11 11:51:37.616637952	{'contributors_enabled': False, 'created_at': ...
1991-02-11 15:50:06.224695296	{'contributors_enabled': False, 'created_at': ...
1991-02-11 17:38:31.632134144	{'contributors_enabled': False, 'created_at': ...
1991-02-12 05:51:22.303524864	{'contributors_enabled': False, 'created_at': ...
1991-02-12 07:56:57.576210432	{'contributors_enabled': False, 'created_at': ...
1991-02-12 10:08:08.456101888	{'contributors_enabled': False, 'created_at': ...
1991-02-12 12:45:58.909284353	{'contributors_enabled': False, 'created_at': ...
1991-02-12 15:49:13.744588802	{'contributors_enabled': False, 'created_at': ...
1991-02-12 22:04:07.373291520	{'contributors_enabled': False, 'created_at': ...
1991-02-13 01:05:26.856765440	{'contributors_enabled': False, 'created_at': ...
1991-02-13 02:18:27.551481857	{'contributors_enabled': False, 'created_at': ...
...	...
1998-01-31 17:16:49.285017600	{'contributors_enabled': False, 'created_at': ...
1998-02-01 20:49:04.734445568	{'contributors_enabled': False, 'created_at': ...
1998-02-05 12:05:36.477933568	{'contributors_enabled': False, 'created_at': ...
1998-02-06 03:48:00.519319552	{'contributors_enabled': False, 'created_at': ...
1998-02-09 00:13:53.522544640	{'contributors_enabled': False, 'created_at': ...
1998-02-10 09:03:12.804085760	{'contributors_enabled': False, 'created_at': ...
1998-02-13 04:13:37.045368832	{'contributors_enabled': False, 'created_at': ...
1998-02-14 16:32:37.103951883	{'contributors_enabled': False, 'created_at': ...
1998-02-15 04:32:19.158093824	{'contributors_enabled': False, 'created_at': ...
1998-02-17 08:48:09.381826560	{'contributors_enabled': False, 'created_at': ...
1998-02-21 16:27:14.458587136	{'contributors_enabled': False, 'created_at': ...
1998-02-27 04:49:22.724278272	{'contributors_enabled': False, 'created_at': ...
1998-03-02 02:16:29.199671297	{'contributors_enabled': False, 'created_at': ...
1998-03-03 09:27:18.123831296	{'contributors_enabled': False, 'created_at': ...
1998-03-07 13:54:01.981685760	{'contributors_enabled': False, 'created_at': ...
1998-03-10 11:58:55.344209921	{'contributors_enabled': False, 'created_at': ...
1998-03-11 17:53:57.579907072	{'contributors_enabled': False, 'created_at': ...
1998-03-12 01:16:28.333682689	{'contributors_enabled': False, 'created_at': ...
1998-03-14 13:08:16.479866881	{'contributors_enabled': False, 'created_at': ...
1998-03-16 00:03:28.113172480	{'contributors_enabled': False, 'created_at': ...
1998-03-18 16:57:35.349198849	{'contributors_enabled': False, 'created_at': ...
1998-03-22 23:26:25.150312448	{'contributors_enabled': False, 'created_at': ...
1998-03-24 08:46:21.411237888	{'contributors_enabled': False, 'created_at': ...
1998-03-27 04:11:53.173991426	{'contributors_enabled': False, 'created_at': ...
1998-03-28 12:25:50.875897856	{'contributors_enabled': False, 'created_at': ...
1998-03-31 06:59:18.926688256	{'contributors_enabled': False, 'created_at': ...
1998-04-04 11:32:37.279858688	{'contributors_enabled': False, 'created_at': ...
1998-04-05 22:26:21.378084864	{'contributors_enabled': False, 'created_at': ...
1998-04-10 03:03:41.306343426	{'contributors_enabled': False, 'created_at': ...
1998-04-12 22:37:23.555336193	{'contributors_enabled': False, 'created_at': ...

[2342 rows x 30 columns]


```
In [18]: # Get Data info
         twitter_api.info()

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 2342 entries, 1991-02-08 13:48:08.022790149 to 1998-04-12 22:37:23.555336193
Data columns (total 30 columns):
contributors          0 non-null float64
coordinates           0 non-null float64
created_at            2342 non-null datetime64[ns]
entities              2342 non-null object
extended_entities     1823 non-null object
favorite_count        2342 non-null int64
favorited             2342 non-null int64
geo                   0 non-null float64
id                    2342 non-null int64
id_str                2342 non-null int64
in_reply_to_screen_name 77 non-null object
in_reply_to_status_id  77 non-null float64
in_reply_to_status_id_str 77 non-null float64
in_reply_to_user_id    77 non-null float64
in_reply_to_user_id_str 77 non-null float64
is_quote_status       2342 non-null int64
lang                  2342 non-null object
place                 1 non-null object
possibly_sensitive     2206 non-null float64
possibly_sensitive_appealable 2206 non-null float64
quoted_status         24 non-null object
quoted_status_id       26 non-null float64
quoted_status_id_str   26 non-null float64
retweet_count          2342 non-null int64
retweeted              2342 non-null int64
retweeted_status       168 non-null object
source                2342 non-null object
text                  2342 non-null object
truncated             2342 non-null int64
user                  2342 non-null object
dtypes: datetime64[ns](1), float64(11), int64(8), object(10)
memory usage: 567.2+ KB
```

```
In [19]: # Get summary of our quantative data
         twitter_api.describe()
```

```
Out[19]:
```

	contributors	coordinates	favorite_count	favorited	geo	\
count	0.0	0.0	2342.000000	2342.0	0.0	
mean	NaN	NaN	8000.126388	0.0	NaN	
std	NaN	NaN	12382.293537	0.0	NaN	
min	NaN	NaN	0.000000	0.0	NaN	

25%	NaN	NaN	1382.750000	0.0	NaN
50%	NaN	NaN	3483.000000	0.0	NaN
75%	NaN	NaN	9802.750000	0.0	NaN
max	NaN	NaN	164933.000000	0.0	NaN

	id	id_str	in_reply_to_status_id \
count	2.342000e+03	2.342000e+03	7.700000e+01
mean	7.422212e+17	7.422212e+17	7.440692e+17
std	6.832408e+16	6.832408e+16	7.524295e+16
min	6.660209e+17	6.660209e+17	6.658147e+17
25%	6.783509e+17	6.783509e+17	6.757073e+17
50%	7.186224e+17	7.186224e+17	7.032559e+17
75%	7.986971e+17	7.986971e+17	8.233264e+17
max	8.924206e+17	8.924206e+17	8.862664e+17

	in_reply_to_status_id_str	in_reply_to_user_id \
count	7.700000e+01	7.700000e+01
mean	7.440692e+17	2.040329e+16
std	7.524295e+16	1.260797e+17
min	6.658147e+17	1.185634e+07
25%	6.757073e+17	3.589728e+08
50%	7.032559e+17	4.196984e+09
75%	8.233264e+17	4.196984e+09
max	8.862664e+17	8.405479e+17

	in_reply_to_user_id_str	is_quote_status	possibly_sensitive \
count	7.700000e+01	2342.000000	2206.0
mean	2.040329e+16	0.013237	0.0
std	1.260797e+17	0.114311	0.0
min	1.185634e+07	0.000000	0.0
25%	3.589728e+08	0.000000	0.0
50%	4.196984e+09	0.000000	0.0
75%	4.196984e+09	0.000000	0.0
max	8.405479e+17	1.000000	0.0

	possibly_sensitive_appealable	quoted_status_id	quoted_status_id_str \
count	2206.0	2.600000e+01	2.600000e+01
mean	0.0	8.113972e+17	8.113972e+17
std	0.0	6.295843e+16	6.295843e+16
min	0.0	6.721083e+17	6.721083e+17
25%	0.0	7.761338e+17	7.761338e+17
50%	0.0	8.281173e+17	8.281173e+17
75%	0.0	8.637581e+17	8.637581e+17
max	0.0	8.860534e+17	8.860534e+17

	retweet_count	retweeted	truncated
count	2342.000000	2342.0	2342.000000
mean	2950.495303	0.0	0.105465

std	4962.148408	0.0	0.307218
min	0.000000	0.0	0.000000
25%	592.500000	0.0	0.000000
50%	1377.000000	0.0	0.000000
75%	3442.250000	0.0	0.000000
max	84082.000000	0.0	1.000000

Quality twitter_archive

- As per project instructions, we need to exclude 181 retweets from our table
- Name column has got inconsistent value and NA need to be removed
- Rating numerator and denominator inconsistent values to be corrected
- Timestamp format to be datetime format. Then split it into three columns year, month, and day
- One thing more I found out I needed to do, make sure that "None" in dogs stages is programmable

image

- Drop records with duplicated jpg_url
- p1,p2 and p3 have inconsistent upper case
- Replace underscore by space in p1, p2, p3 values
- Create one column for image prediction and one column for confidence level

twitter_api

- Rename the id column to tweet_id
- 168 records have a retweeted_status which will need to be excluded

2.2.3 Tidiness

twitter_archive - Combine each dog stage column into a single column named "stage" - Convert Rating columns to float values - Remove columns that are not needed - All datasets

All datasets - All datasets Convert tweet_id column from an integer to a string value - Consolidate the 3 datasets into 1

2.2.4 Cleaning Data

twitter_archive dataset

```
In [20]: # Make copy for these 3 datasets
twitter_archive_clean = twitter_archive.copy()
```

Define Retweets are not original tweets and need to be removed

Code

```
In [21]: # drop retweets values so they don't count
twitter_archive_clean.drop(twitter_archive_clean[twitter_archive_clean.retweeted_status
```

Test

```
In [22]: twitter_archive_clean.info()
         twitter_archive_clean. shape
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                2175 non-null int64
in_reply_to_status_id   78 non-null float64
in_reply_to_user_id     78 non-null float64
timestamp               2175 non-null object
source                 2175 non-null object
text                   2175 non-null object
retweeted_status_id     0 non-null float64
retweeted_status_user_id 0 non-null float64
retweeted_status_timestamp 0 non-null object
expanded_urls          2117 non-null object
rating_numerator        2175 non-null int64
rating_denominator      2175 non-null int64
name                   2175 non-null object
doggo                  2175 non-null object
floofer                2175 non-null object
pupper                2175 non-null object
puppo                  2175 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 305.9+ KB
```

```
Out[22]: (2175, 17)
```

Define Remove all retweet related columns

Code

```
In [23]: # Remove retweets related columns
         twitter_archive_clean.drop(['retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timestamp'])
```

Test

```
In [24]: twitter_archive_clean.info()
         twitter_archive_clean. shape
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 14 columns):
tweet_id                2175 non-null int64
in_reply_to_status_id   78 non-null float64
in_reply_to_user_id     78 non-null float64
```

```

timestamp          2175 non-null object
source             2175 non-null object
text               2175 non-null object
expanded_urls      2117 non-null object
rating_numerator   2175 non-null int64
rating_denominator 2175 non-null int64
name               2175 non-null object
doggo              2175 non-null object
floofer            2175 non-null object
pupper            2175 non-null object
puppo              2175 non-null object
dtypes: float64(2), int64(3), object(9)
memory usage: 254.9+ KB

```

```
Out[24]: (2175, 14)
```

Define Name change to be verified

Code

```
In [25]: # Name column errors
```

```
        twitter_archive_clean.name.unique()
```

```
Out[25]: array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', 'None', 'Jax',
               'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver', 'Jim',
               'Zeke', 'Ralphus', 'Gerald', 'Jeffrey', 'such', 'Canela', 'Maya',
               'Mingus', 'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Earl',
               'Lola', 'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald', 'Rusty',
               'Gus', 'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'a', 'Elliot',
               'Louis', 'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack', 'Steven',
               'Beau', 'Snoopy', 'Shadow', 'Emmy', 'Aja', 'Penny', 'Dante',
               'Nelly', 'Ginger', 'Benedict', 'Venti', 'Goose', 'Nugget', 'Cash',
               'Jed', 'Sebastian', 'Sierra', 'Monkey', 'Harry', 'Kody', 'Lassie',
               'Rover', 'Napolean', 'Boomer', 'Cody', 'Rumble', 'Clifford',
               'Dewey', 'Scout', 'Gizmo', 'Walter', 'Cooper', 'Harold', 'Shikha',
               'Lili', 'Jamesy', 'Coco', 'Sammy', 'Meatball', 'Paisley', 'Albus',
               'Neptune', 'Belle', 'Quinn', 'Zooey', 'Dave', 'Jersey', 'Hobbes',
               'Burt', 'Lorenzo', 'Carl', 'Jordy', 'Milky', 'Trooper', 'quite',
               'Sophie', 'Wyatt', 'Rosie', 'Thor', 'Oscar', 'Callie', 'Cermet',
               'Marlee', 'Arya', 'Einstein', 'Alice', 'Rumpole', 'Benny', 'Aspen',
               'Jarod', 'Wiggles', 'General', 'Sailor', 'Iggy', 'Snoop', 'Kyle',
               'Leo', 'Riley', 'Noosh', 'Odin', 'Jerry', 'Georgie', 'Rontu',
               'Cannon', 'Furzey', 'Daisy', 'Tuck', 'Barney', 'Vixen', 'Jarvis',
               'Mimosa', 'Pickles', 'Brady', 'Luna', 'Charlie', 'Margo', 'Sadie',
               'Hank', 'Tycho', 'Indie', 'Winnie', 'George', 'Bentley', 'Max',
               'Dawn', 'Maddie', 'Monty', 'Sojourner', 'Winston', 'Odie', 'Arlo',
               'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera', 'Ava', 'Eli', 'Ash',
```

'Tucker', 'Tobi', 'Chester', 'Wilson', 'Sunshine', 'Lipton',
 'Bronte', 'Poppy', 'Gidget', 'Rhino', 'Willow', 'not', 'Orion',
 'Eevee', 'Smiley', 'Miguel', 'Emanuel', 'Kuyu', 'Dutch', 'Pete',
 'Scooter', 'Reggie', 'Lilly', 'Samson', 'Mia', 'Astrid', 'Malcolm',
 'Dexter', 'Alfie', 'Fiona', 'one', 'Mutt', 'Bear', 'Doobert',
 'Beebop', 'Alexander', 'Sailer', 'Brutus', 'Kona', 'Boots',
 'Ralphie', 'Loki', 'Cupid', 'Pawnd', 'Pilot', 'Ike', 'Mo', 'Toby',
 'Sweet', 'Pablo', 'Nala', 'Crawford', 'Gabe', 'Jimison', 'Duchess',
 'Harlso', 'Sundance', 'Luca', 'Flash', 'Sunny', 'Howie', 'Jazzy',
 'Anna', 'Finn', 'Bo', 'Wafer', 'Tom', 'Florence', 'Autumn', 'Buddy',
 'Dido', 'Eugene', 'Ken', 'Strudel', 'Tebow', 'Chloe', 'Timber',
 'Binky', 'Moose', 'Dudley', 'Comet', 'Akumi', 'Titan', 'Olivia',
 'Alf', 'Oshie', 'Chubbs', 'Sky', 'Atlas', 'Eleanor', 'Layla',
 'Rocky', 'Baron', 'Tyr', 'Bauer', 'Swagger', 'Brandi', 'Mary',
 'Moe', 'Halo', 'Augie', 'Craig', 'Sam', 'Hunter', 'Pavlov', 'Phil',
 'Kyro', 'Wallace', 'Ito', 'Seamus', 'Ollie', 'Stephan', 'Lennon',
 'incredibly', 'Major', 'Duke', 'Sansa', 'Shooter', 'Django',
 'Diogi', 'Sonny', 'Marley', 'Severus', 'Ronnie', 'Milo', 'Bones',
 'Mauve', 'Chef', 'Doc', 'Peaches', 'Sobe', 'Longfellow', 'Mister',
 'Iroh', 'Pancake', 'Snicku', 'Ruby', 'Brody', 'Mack', 'Nimbus',
 'Laika', 'Maximus', 'Dobby', 'Moreton', 'Juno', 'Maude', 'Lily',
 'Newt', 'Benji', 'Nida', 'Robin', 'Monster', 'BeBe', 'Remus',
 'Levi', 'Mabel', 'Misty', 'Betty', 'Mosby', 'Maggie', 'Bruce',
 'Happy', 'Ralphy', 'Brownie', 'Rizzy', 'Stella', 'Butter', 'Frank',
 'Tonks', 'Lincoln', 'Rory', 'Logan', 'Dale', 'Rizzo', 'Arnie',
 'Mattie', 'Pinot', 'Dallas', 'Hero', 'Frankie', 'Stormy',
 'Reginald', 'Balto', 'Mairi', 'Loomis', 'Godi', 'Cali', 'Deacon',
 'Timmy', 'Sampson', 'Chipson', 'Combo', 'Oakley', 'Dash',
 'Hercules', 'Jay', 'Mya', 'Strider', 'Wesley', 'Solomon', 'Huck',
 'O', 'Blue', 'Anakin', 'Finley', 'Sprinkles', 'Heinrich',
 'Shakespeare', 'Chelsea', 'Bungalo', 'Chip', 'Grey', 'Roosevelt',
 'Willem', 'Davey', 'Dakota', 'Fizz', 'Dixie', 'very', 'Al',
 'Jackson', 'Carbon', 'Klein', 'DonDon', 'Kirby', 'Lou', 'Chevy',
 'Tito', 'Philbert', 'Louie', 'Rupert', 'Rufus', 'Brudge', 'Shadoe',
 'Angel', 'Brat', 'Tove', 'my', 'Gromit', 'Aubie', 'Kota', 'Leela',
 'Glenn', 'Shelby', 'Sephie', 'Bonaparte', 'Albert', 'Wishes',
 'Rose', 'Theo', 'Rocco', 'Fido', 'Emma', 'Spencer', 'Lilli',
 'Boston', 'Brandonald', 'Corey', 'Leonard', 'Beckham', 'Devón',
 'Gert', 'Watson', 'Keith', 'Dex', 'Ace', 'Tayzie', 'Grizzie',
 'Fred', 'Gilbert', 'Meyer', 'Zoe', 'Stewie', 'Calvin', 'Lilah',
 'Spanky', 'Jameson', 'Piper', 'Atticus', 'Blu', 'Dietrich',
 'Divine', 'Tripp', 'his', 'Cora', 'Huxley', 'Keurig', 'Bookstore',
 'Linus', 'Abby', 'Shiloh', 'an', 'Gustav', 'Arlen', 'Percy',
 'Lenox', 'Sugar', 'Harvey', 'Blanket', 'actually', 'Geno', 'Stark',
 'Beya', 'Kilo', 'Kayla', 'Maxaroni', 'Bell', 'Doug', 'Edmund',
 'Aqua', 'Theodore', 'just', 'Baloo', 'Chase', 'getting', 'Nollie',
 'Rorie', 'Simba', 'Charles', 'Bayley', 'Axel', 'Storkson', 'Remy',
 'Chadrick', 'mad', 'Kellogg', 'Buckley', 'Livvie', 'Terry',

'Hermione', 'Ralpher', 'Aldrick', 'Larry', 'this', 'unacceptable',
 'Rooney', 'Crystal', 'Ziva', 'Stefan', 'Pupcasso', 'Puff',
 'Flurpson', 'Coleman', 'Enchilada', 'Raymond', 'all', 'Rueben',
 'Cilantro', 'Karll', 'Sprout', 'Blitz', 'Bloop', 'Colby', 'Lillie',
 'Ashleigh', 'Kreggory', 'Sarge', 'Luther', 'Ivar', 'Jangle',
 'Schnitzel', 'Panda', 'Berkeley', 'Ralphé', 'Charleson', 'Clyde',
 'Harnold', 'Sid', 'Pippa', 'Otis', 'Carper', 'Bowie',
 'Alexanderson', 'Suki', 'Barclay', 'Skittle', 'Ebby', 'Flávio',
 'Smokey', 'Link', 'Jennifur', 'Ozzy', 'Bluebert', 'Stephanus',
 'Bubbles', 'old', 'Zeus', 'Bertson', 'Nico', 'Michelangelo',
 'Siba', 'Calbert', 'Curtis', 'Travis', 'Thumas', 'Kanu', 'Lance',
 'Opie', 'Stubert', 'Kane', 'Olive', 'Chuckles', 'Staniel', 'Sora',
 'Beemo', 'Gunner', 'infuriating', 'Lacy', 'Tater', 'Olaf', 'Cecil',
 'Vince', 'Karma', 'Billy', 'Walker', 'Rodney', 'Klevin', 'Malikai',
 'Bobble', 'River', 'Jebberson', 'Remington', 'Farfle', 'Jiminus',
 'Harper', 'Clarkus', 'Finnegus', 'Cupcake', 'Kathmandu', 'Ellie',
 'Katie', 'Kara', 'Adele', 'Zara', 'Ambrose', 'Jimothy', 'Bode',
 'Terrenth', 'Reese', 'Chesterson', 'Lucia', 'Bisquick', 'Ralphson',
 'Socks', 'Rambo', 'Rudy', 'Fiji', 'Rilo', 'Bilbo', 'Coopson',
 'Yoda', 'Millie', 'Chet', 'Crouton', 'Daniel', 'Kaia', 'Murphy',
 'Dotsy', 'Eazy', 'Coops', 'Fillup', 'Miley', 'Charl', 'Reagan',
 'Yukon', 'CeCe', 'Cuddles', 'Claude', 'Jessiga', 'Carter', 'Ole',
 'Pherb', 'Blipson', 'Reptar', 'Trevith', 'Berb', 'Bob', 'Colin',
 'Brian', 'Olivier', 'Grady', 'Kobe', 'Freddery', 'Bodie', 'Dunkin',
 'Wally', 'Tupawc', 'Amber', 'Herschel', 'Edgar', 'Teddy',
 'Kingsley', 'Brockly', 'Richie', 'Molly', 'Vinscent', 'Cedrick',
 'Hazel', 'Lolo', 'Eriq', 'Phred', 'the', 'Oddie', 'Maxwell',
 'Geoff', 'Covach', 'Durg', 'Fynn', 'Ricky', 'Herald', 'Lucky',
 'Ferg', 'Trip', 'Clarence', 'Hamrick', 'Brad', 'Pubert', 'Frönq',
 'Derby', 'Lizzie', 'Ember', 'Blakely', 'Opal', 'Marq', 'Kramer',
 'Barry', 'Tyrone', 'Gordon', 'Baxter', 'Mona', 'Horace', 'Crimson',
 'Birf', 'Hammond', 'Lorelei', 'Marty', 'Brooks', 'Petrick',
 'Hubertson', 'Gerbald', 'Oreo', 'Bruiser', 'Perry', 'Bobby', 'Jeph',
 'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe', 'Tiger', 'Jiminy',
 'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Darrel', 'Taco', 'Joey',
 'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses', 'Toffee', 'Apollo',
 'Carly', 'Asher', 'Glacier', 'Chuck', 'Champ', 'Ozzie', 'Griswold',
 'Cheesy', 'Moofasa', 'Hector', 'Goliath', 'Kawhi', 'by', 'Emmie',
 'Penelope', 'Willie', 'Rinna', 'Mike', 'William', 'Dwight', 'Evy',
 'Hurley', 'Rubio', 'officially', 'Chompsky', 'Rascal', 'Linda',
 'Tug', 'Tango', 'Grizz', 'Jerome', 'Crumpet', 'Jessifer', 'Izzy',
 'Ralph', 'Sandy', 'Humphrey', 'Tassy', 'Juckson', 'Chuq', 'Tyrus',
 'Karl', 'Godzilla', 'Vinnie', 'Kenneth', 'Herm', 'Bert', 'Striker',
 'Donny', 'Pepper', 'Bernie', 'Buddah', 'Lenny', 'Arnold', 'Zuzu',
 'Mollie', 'Laela', 'Tedders', 'Superpup', 'Rufio', 'Jeb', 'Rodman',
 'Jonah', 'Chesney', 'life', 'Kenny', 'Henry', 'Bobbay', 'Mitch',
 'Kaiya', 'Acro', 'Aiden', 'Obie', 'Dot', 'Shnuggles', 'Kendall',
 'Jeffri', 'Steve', 'Eve', 'Mac', 'Fletcher', 'Kenzie', 'Pumpkin',

```
'Schnozz', 'Gustaf', 'Cheryl', 'Ed', 'Leonidas', 'Norman', 'Caryl',
'Scott', 'Taz', 'Darby', 'Jackie', 'light', 'Jazz', 'Franq',
'Pippin', 'Rolf', 'Snickers', 'Ridley', 'Cal', 'Bradley', 'Bubba',
'Tuco', 'Patch', 'Mojo', 'Batdog', 'Dylan', 'space', 'Mark', 'JD',
'Alejandro', 'Scruffers', 'Pip', 'Julius', 'Tanner', 'Sparky',
'Anthony', 'Holly', 'Jett', 'Amy', 'Sage', 'Andy', 'Mason',
'Trigger', 'Antony', 'Creg', 'Traviss', 'Gin', 'Jeffrie', 'Danny',
'Ester', 'Pluto', 'Bloo', 'Edd', 'Paull', 'Willy', 'Herb', 'Damon',
'Peanut', 'Nigel', 'Butters', 'Sandra', 'Fabio', 'Randall', 'Liam',
'Tommy', 'Ben', 'Raphael', 'Julio', 'Andru', 'Kloey', 'Shawwn',
'Skye', 'Kollin', 'Ronduh', 'Billl', 'Saydee', 'Dug', 'Tessa',
'Sully', 'Kirk', 'Ralf', 'Clarq', 'Jaspers', 'Samsom', 'Terrance',
'Harrison', 'Chaz', 'Jeremy', 'Jaycob', 'Lambeau', 'Ruffles',
'Amélie', 'Bobb', 'Banditt', 'Kevon', 'Winifred', 'Hanz', 'Churlie',
'Zeek', 'Timofy', 'Maks', 'Jomathan', 'Kallie', 'Marvin', 'Spark',
'Gòrdón', 'Jo', 'DayZ', 'Jareld', 'Torque', 'Ron', 'Skittles',
'Cleopatricia', 'Erik', 'Stu', 'Tedrick', 'Shaggy', 'Filup', 'Kial',
'Naphaniel', 'Dook', 'Hall', 'Philippe', 'Biden', 'Fwed',
'Genevieve', 'Joshwa', 'Timison', 'Bradlay', 'Pipsy', 'Clybe',
'Keet', 'Carll', 'Jockson', 'Josep', 'Lugan', 'Christoper'], dtype=object)
```

```
In [26]: # Name consistency
for name in twitter_archive_clean.name:
    if name.islower():
        twitter_archive_clean.name.replace(name, 'None', inplace = True)
twitter_archive_clean.name = twitter_archive_clean.name.replace('0', 'None')
```

Test

```
In [27]: twitter_archive_clean.name.unique()
```

```
Out[27]: array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', 'None', 'Jax',
'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver', 'Jim',
'Zeke', 'Ralphus', 'Gerald', 'Jeffrey', 'Canela', 'Maya', 'Mingus',
'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Earl', 'Lola',
'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald', 'Rusty', 'Gus',
'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'Elliot', 'Louis',
'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack', 'Steven', 'Beau',
'Snoopy', 'Shadow', 'Emmy', 'Aja', 'Penny', 'Dante', 'Nelly',
'Ginger', 'Benedict', 'Venti', 'Goose', 'Nugget', 'Cash', 'Jed',
'Sebastian', 'Sierra', 'Monkey', 'Harry', 'Kody', 'Lassie', 'Rover',
'Napolean', 'Boomer', 'Cody', 'Rumble', 'Clifford', 'Dewey',
'Scout', 'Gizmo', 'Walter', 'Cooper', 'Harold', 'Shikha', 'Lili',
'Jamesy', 'Coco', 'Sammy', 'Meatball', 'Paisley', 'Albus',
'Neptune', 'Belle', 'Quinn', 'Zooey', 'Dave', 'Jersey', 'Hobbes',
'Burt', 'Lorenzo', 'Carl', 'Jordy', 'Milky', 'Trooper', 'Sophie',
'Wyatt', 'Rosie', 'Thor', 'Oscar', 'Callie', 'Cermet', 'Marlee',
'Arya', 'Einstein', 'Alice', 'Rumpole', 'Benny', 'Aspen', 'Jarod',
```


'Wiggles', 'General', 'Sailor', 'Iggy', 'Snoop', 'Kyle', 'Leo',
 'Riley', 'Noosh', 'Odin', 'Jerry', 'Georgie', 'Rontu', 'Cannon',
 'Furzey', 'Daisy', 'Tuck', 'Barney', 'Vixen', 'Jarvis', 'Mimosa',
 'Pickles', 'Brady', 'Luna', 'Charlie', 'Margo', 'Sadie', 'Hank',
 'Tycho', 'Indie', 'Winnie', 'George', 'Bentley', 'Max', 'Dawn',
 'Maddie', 'Monty', 'Sojourner', 'Winston', 'Odie', 'Arlo',
 'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera', 'Ava', 'Eli', 'Ash',
 'Tucker', 'Tobi', 'Chester', 'Wilson', 'Sunshine', 'Lipton',
 'Bronte', 'Poppy', 'Gidget', 'Rhino', 'Willow', 'Orion', 'Eevee',
 'Smiley', 'Miguel', 'Emanuel', 'Kuyu', 'Dutch', 'Pete', 'Scooter',
 'Reggie', 'Lilly', 'Samson', 'Mia', 'Astrid', 'Malcolm', 'Dexter',
 'Alfie', 'Fiona', 'Mutt', 'Bear', 'Doobert', 'Beebop', 'Alexander',
 'Sailer', 'Brutus', 'Kona', 'Boots', 'Ralphie', 'Loki', 'Cupid',
 'Pawnd', 'Pilot', 'Ike', 'Mo', 'Toby', 'Sweet', 'Pablo', 'Nala',
 'Crawford', 'Gabe', 'Jimison', 'Duchess', 'Harlso', 'Sundance',
 'Luca', 'Flash', 'Sunny', 'Howie', 'Jazzy', 'Anna', 'Finn', 'Bo',
 'Wafer', 'Tom', 'Florence', 'Autumn', 'Buddy', 'Dido', 'Eugene',
 'Ken', 'Strudel', 'Tebow', 'Chloe', 'Timber', 'Binky', 'Moose',
 'Dudley', 'Comet', 'Akumi', 'Titan', 'Olivia', 'Alf', 'Oshie',
 'Chubbs', 'Sky', 'Atlas', 'Eleanor', 'Layla', 'Rocky', 'Baron',
 'Tyr', 'Bauer', 'Swagger', 'Brandi', 'Mary', 'Moe', 'Halo', 'Augie',
 'Craig', 'Sam', 'Hunter', 'Pavlov', 'Phil', 'Kyro', 'Wallace',
 'Ito', 'Seamus', 'Ollie', 'Stephan', 'Lennon', 'Major', 'Duke',
 'Sansa', 'Shooter', 'Django', 'Diogi', 'Sonny', 'Marley', 'Severus',
 'Ronnie', 'Milo', 'Bones', 'Mauve', 'Chef', 'Doc', 'Peaches',
 'Sobe', 'Longfellow', 'Mister', 'Iroh', 'Pancake', 'Snicku', 'Ruby',
 'Brody', 'Mack', 'Nimbus', 'Laika', 'Maximus', 'Dobby', 'Moreton',
 'Juno', 'Maude', 'Lily', 'Newt', 'Benji', 'Nida', 'Robin',
 'Monster', 'BeBe', 'Remus', 'Levi', 'Mabel', 'Misty', 'Betty',
 'Mosby', 'Maggie', 'Bruce', 'Happy', 'Ralphy', 'Brownie', 'Rizzy',
 'Stella', 'Butter', 'Frank', 'Tonks', 'Lincoln', 'Rory', 'Logan',
 'Dale', 'Rizzo', 'Arnie', 'Mattie', 'Pinot', 'Dallas', 'Hero',
 'Frankie', 'Stormy', 'Reginald', 'Balto', 'Mairi', 'Loomis', 'Godi',
 'Cali', 'Deacon', 'Timmy', 'Sampson', 'Chipson', 'Combo', 'Oakley',
 'Dash', 'Hercules', 'Jay', 'Mya', 'Strider', 'Wesley', 'Solomon',
 'Huck', 'Blue', 'Anakin', 'Finley', 'Sprinkles', 'Heinrich',
 'Shakespeare', 'Chelsea', 'Bungalo', 'Chip', 'Grey', 'Roosevelt',
 'Willem', 'Davey', 'Dakota', 'Fizz', 'Dixie', 'Al', 'Jackson',
 'Carbon', 'Klein', 'DonDon', 'Kirby', 'Lou', 'Chevy', 'Tito',
 'Philbert', 'Louie', 'Rupert', 'Rufus', 'Brudge', 'Shadoe', 'Angel',
 'Brat', 'Tove', 'Gromit', 'Aubie', 'Kota', 'Leela', 'Glenn',
 'Shelby', 'Sephie', 'Bonaparte', 'Albert', 'Wishes', 'Rose', 'Theo',
 'Rocco', 'Fido', 'Emma', 'Spencer', 'Lilli', 'Boston', 'Brandonald',
 'Corey', 'Leonard', 'Beckham', 'Devón', 'Gert', 'Watson', 'Keith',
 'Dex', 'Ace', 'Tayzie', 'Grizzie', 'Fred', 'Gilbert', 'Meyer',
 'Zoe', 'Stewie', 'Calvin', 'Lilah', 'Spanky', 'Jameson', 'Piper',
 'Atticus', 'Blu', 'Dietrich', 'Divine', 'Tripp', 'Cora', 'Huxley',
 'Keurig', 'Bookstore', 'Linus', 'Abby', 'Shiloh', 'Gustav', 'Arlen',

'Percy', 'Lenox', 'Sugar', 'Harvey', 'Blanket', 'Geno', 'Stark',
 'Beya', 'Kilo', 'Kayla', 'Maxaroni', 'Bell', 'Doug', 'Edmund',
 'Aqua', 'Theodore', 'Baloo', 'Chase', 'Nollie', 'Rorie', 'Simba',
 'Charles', 'Bayley', 'Axel', 'Storkson', 'Remy', 'Chadrick',
 'Kellogg', 'Buckley', 'Livvie', 'Terry', 'Hermione', 'Ralpher',
 'Aldrick', 'Larry', 'Rooney', 'Crystal', 'Ziva', 'Stefan',
 'Pupcasso', 'Puff', 'Flurpson', 'Coleman', 'Enchilada', 'Raymond',
 'Rueben', 'Cilantro', 'Karll', 'Sprout', 'Blitz', 'Bloop', 'Colby',
 'Lillie', 'Ashleigh', 'Kreggory', 'Sarge', 'Luther', 'Ivar',
 'Jangle', 'Schnitzel', 'Panda', 'Berkeley', 'Ralphé', 'Charleson',
 'Clyde', 'Harnold', 'Sid', 'Pippa', 'Otis', 'Carper', 'Bowie',
 'Alexanderson', 'Suki', 'Barclay', 'Skittle', 'Ebby', 'Flávio',
 'Smokey', 'Link', 'Jennifur', 'Ozzy', 'Bluebert', 'Stephanus',
 'Bubbles', 'Zeus', 'Bertson', 'Nico', 'Michelangelo', 'Siba',
 'Calbert', 'Curtis', 'Travis', 'Thumas', 'Kanu', 'Lance', 'Opie',
 'Stubert', 'Kane', 'Olive', 'Chuckles', 'Staniel', 'Sora', 'Beemo',
 'Gunner', 'Lacy', 'Tater', 'Olaf', 'Cecil', 'Vince', 'Karma',
 'Billy', 'Walker', 'Rodney', 'Klevin', 'Malikai', 'Bobble', 'River',
 'Jebberson', 'Remington', 'Farfle', 'Jiminus', 'Harper', 'Clarkus',
 'Finnegus', 'Cupcake', 'Kathmandu', 'Ellie', 'Katie', 'Kara',
 'Adele', 'Zara', 'Ambrose', 'Jimothy', 'Bode', 'Terrenth', 'Reese',
 'Chesterson', 'Lucia', 'Bisquick', 'Ralphson', 'Socks', 'Rambo',
 'Rudy', 'Fiji', 'Rilo', 'Bilbo', 'Coopson', 'Yoda', 'Millie',
 'Chet', 'Crouton', 'Daniel', 'Kaia', 'Murphy', 'Dotsy', 'Eazy',
 'Coops', 'Fillup', 'Miley', 'Charl', 'Reagan', 'Yukon', 'CeCe',
 'Cuddles', 'Claude', 'Jessiga', 'Carter', 'Ole', 'Pherb', 'Blipson',
 'Reptar', 'Trevith', 'Berb', 'Bob', 'Colin', 'Brian', 'Olivier',
 'Grady', 'Kobe', 'Freddery', 'Bodie', 'Dunkin', 'Wally', 'Tupawc',
 'Amber', 'Herschel', 'Edgar', 'Teddy', 'Kingsley', 'Brockly',
 'Richie', 'Molly', 'Vinscent', 'Cedrick', 'Hazel', 'Lolo', 'Eriq',
 'Phred', 'Oddie', 'Maxwell', 'Geoff', 'Covach', 'Durg', 'Fynn',
 'Ricky', 'Herald', 'Lucky', 'Ferg', 'Trip', 'Clarence', 'Hamrick',
 'Brad', 'Pubert', 'Frönq', 'Derby', 'Lizzie', 'Ember', 'Blakely',
 'Opal', 'Marq', 'Kramer', 'Barry', 'Tyrone', 'Gordon', 'Baxter',
 'Mona', 'Horace', 'Crimson', 'Birf', 'Hammond', 'Lorelei', 'Marty',
 'Brooks', 'Petrick', 'Hubertson', 'Gerbald', 'Oreo', 'Bruiser',
 'Perry', 'Bobby', 'Jeph', 'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe',
 'Tiger', 'Jiminy', 'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Darrel',
 'Taco', 'Joey', 'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses',
 'Toffee', 'Apollo', 'Carly', 'Asher', 'Glacier', 'Chuck', 'Champ',
 'Ozzie', 'Griswold', 'Cheesy', 'Moofasa', 'Hector', 'Goliath',
 'Kawhi', 'Emmie', 'Penelope', 'Willie', 'Rinna', 'Mike', 'William',
 'Dwight', 'Evy', 'Hurley', 'Rubio', 'Chompsky', 'Rascal', 'Linda',
 'Tug', 'Tango', 'Grizz', 'Jerome', 'Crumpet', 'Jessifer', 'Izzy',
 'Ralph', 'Sandy', 'Humphrey', 'Tassy', 'Juckson', 'Chuq', 'Tyrus',
 'Karl', 'Godzilla', 'Vinnie', 'Kenneth', 'Herm', 'Bert', 'Striker',
 'Donny', 'Pepper', 'Bernie', 'Buddah', 'Lenny', 'Arnold', 'Zuzu',
 'Mollie', 'Laela', 'Tedders', 'Superpup', 'Rufio', 'Jeb', 'Rodman',

```
'Jonah', 'Chesney', 'Kenny', 'Henry', 'Bobbay', 'Mitch', 'Kaiya',
'Acro', 'Aiden', 'Obie', 'Dot', 'Shnuggles', 'Kendall', 'Jeffri',
'Steve', 'Eve', 'Mac', 'Fletcher', 'Kenzie', 'Pumpkin', 'Schnozz',
'Gustaf', 'Cheryl', 'Ed', 'Leonidas', 'Norman', 'Caryl', 'Scott',
'Taz', 'Darby', 'Jackie', 'Jazz', 'Franq', 'Pippin', 'Rolf',
'Snickers', 'Ridley', 'Cal', 'Bradley', 'Bubba', 'Tuco', 'Patch',
'Mojo', 'Batdog', 'Dylan', 'Mark', 'JD', 'Alejandro', 'Scrufflers',
'Pip', 'Julius', 'Tanner', 'Sparky', 'Anthony', 'Holly', 'Jett',
'Amy', 'Sage', 'Andy', 'Mason', 'Trigger', 'Antony', 'Creg',
'Traviss', 'Gin', 'Jeffrie', 'Danny', 'Ester', 'Pluto', 'Bloo',
'Edd', 'Paull', 'Willy', 'Herb', 'Damon', 'Peanut', 'Nigel',
'Butters', 'Sandra', 'Fabio', 'Randall', 'Liam', 'Tommy', 'Ben',
'Raphael', 'Julio', 'Andru', 'Kloey', 'Shawwn', 'Skye', 'Kollin',
'Ronduh', 'Billl', 'Saydee', 'Dug', 'Tessa', 'Sully', 'Kirk',
'Ralf', 'Clarq', 'Jaspers', 'Samsom', 'Terrance', 'Harrison',
'Chaz', 'Jeremy', 'Jaycob', 'Lambeau', 'Ruffles', 'Amélie', 'Bobb',
'Banditt', 'Kevon', 'Winifred', 'Hanz', 'Churlie', 'Zeek', 'Timofy',
'Maks', 'Jomathan', 'Kallie', 'Marvin', 'Spark', 'Gòrdón', 'Jo',
'DayZ', 'Jareld', 'Torque', 'Ron', 'Skittles', 'Cleopatricia',
'Erik', 'Stu', 'Tedrick', 'Shaggy', 'Filup', 'Kial', 'Naphaniel',
'Dook', 'Hall', 'Philippe', 'Biden', 'Fwed', 'Genevieve', 'Joshwa',
'Timison', 'Bradlay', 'Pipsy', 'Clybe', 'Keet', 'Carll', 'Jackson',
'Josep', 'Lugan', 'Christoper'], dtype=object)
```

Define Correct NA values in the name column

Code

```
In [28]: twitter_archive_clean.name = twitter_archive_clean.name.apply(lambda x: x if x != 'None')
```

Test

```
In [29]: twitter_archive_clean.name.unique()
```

```
Out[29]: array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', nan, 'Jax',
'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver', 'Jim',
'Zeke', 'Ralphus', 'Gerald', 'Jeffrey', 'Canela', 'Maya', 'Mingus',
'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Earl', 'Lola',
'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald', 'Rusty', 'Gus',
'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'Elliot', 'Louis',
'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack', 'Steven', 'Beau',
'Snoopy', 'Shadow', 'Emmy', 'Aja', 'Penny', 'Dante', 'Nelly',
'Ginger', 'Benedict', 'Venti', 'Goose', 'Nugget', 'Cash', 'Jed',
'Sebastian', 'Sierra', 'Monkey', 'Harry', 'Kody', 'Lassie', 'Rover',
'Napolean', 'Boomer', 'Cody', 'Rumble', 'Clifford', 'Dewey',
'Scout', 'Gizmo', 'Walter', 'Cooper', 'Harold', 'Shikha', 'Lili',
'Jamesy', 'Coco', 'Sammy', 'Meatball', 'Paisley', 'Albus',
'Neptune', 'Belle', 'Quinn', 'Zooey', 'Dave', 'Jersey', 'Hobbes',
```

'Burt', 'Lorenzo', 'Carl', 'Jordy', 'Milky', 'Trooper', 'Sophie',
'Wyatt', 'Rosie', 'Thor', 'Oscar', 'Callie', 'Cermet', 'Marlee',
'Arya', 'Einstein', 'Alice', 'Rumpole', 'Benny', 'Aspen', 'Jarod',
'Wiggles', 'General', 'Sailor', 'Iggy', 'Snoop', 'Kyle', 'Leo',
'Riley', 'Noosh', 'Odin', 'Jerry', 'Georgie', 'Rontu', 'Cannon',
'Furzey', 'Daisy', 'Tuck', 'Barney', 'Vixen', 'Jarvis', 'Mimosa',
'Pickles', 'Brady', 'Luna', 'Charlie', 'Margo', 'Sadie', 'Hank',
'Tycho', 'Indie', 'Winnie', 'George', 'Bentley', 'Max', 'Dawn',
'Maddie', 'Monty', 'Sojourner', 'Winston', 'Odie', 'Arlo',
'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera', 'Ava', 'Eli', 'Ash',
'Tucker', 'Tobi', 'Chester', 'Wilson', 'Sunshine', 'Lipton',
'Bronte', 'Poppy', 'Gidget', 'Rhino', 'Willow', 'Orion', 'Eevee',
'Smiley', 'Miguel', 'Emanuel', 'Kuyu', 'Dutch', 'Pete', 'Scooter',
'Reggie', 'Lilly', 'Samson', 'Mia', 'Astrid', 'Malcolm', 'Dexter',
'Alfie', 'Fiona', 'Mutt', 'Bear', 'Doobert', 'Beebop', 'Alexander',
'Sailer', 'Brutus', 'Kona', 'Boots', 'Ralphie', 'Loki', 'Cupid',
'Pawnd', 'Pilot', 'Ike', 'Mo', 'Toby', 'Sweet', 'Pablo', 'Nala',
'Crawford', 'Gabe', 'Jimison', 'Duchess', 'Harlso', 'Sundance',
'Luca', 'Flash', 'Sunny', 'Howie', 'Jazzy', 'Anna', 'Finn', 'Bo',
'Wafer', 'Tom', 'Florence', 'Autumn', 'Buddy', 'Dido', 'Eugene',
'Ken', 'Strudel', 'Tebow', 'Chloe', 'Timber', 'Binky', 'Moose',
'Dudley', 'Comet', 'Akumi', 'Titan', 'Olivia', 'Alf', 'Oshie',
'Chubbs', 'Sky', 'Atlas', 'Eleanor', 'Layla', 'Rocky', 'Baron',
'Tyr', 'Bauer', 'Swagger', 'Brandi', 'Mary', 'Moe', 'Halo', 'Augie',
'Craig', 'Sam', 'Hunter', 'Pavlov', 'Phil', 'Kyro', 'Wallace',
'Ito', 'Seamus', 'Ollie', 'Stephan', 'Lennon', 'Major', 'Duke',
'Sansa', 'Shooter', 'Django', 'Diogi', 'Sonny', 'Marley', 'Severus',
'Ronnie', 'Milo', 'Bones', 'Mauve', 'Chef', 'Doc', 'Peaches',
'Sobe', 'Longfellow', 'Mister', 'Iroh', 'Pancake', 'Snicku', 'Ruby',
'Brody', 'Mack', 'Nimbus', 'Laika', 'Maximus', 'Dobby', 'Moreton',
'Juno', 'Maude', 'Lily', 'Newt', 'Benji', 'Nida', 'Robin',
'Monster', 'BeBe', 'Remus', 'Levi', 'Mabel', 'Misty', 'Betty',
'Mosby', 'Maggie', 'Bruce', 'Happy', 'Ralphy', 'Brownie', 'Rizzy',
'Stella', 'Butter', 'Frank', 'Tonks', 'Lincoln', 'Rory', 'Logan',
'Dale', 'Rizzo', 'Arnie', 'Mattie', 'Pinot', 'Dallas', 'Hero',
'Frankie', 'Stormy', 'Reginald', 'Balto', 'Mairi', 'Loomis', 'Godi',
'Cali', 'Deacon', 'Timmy', 'Sampson', 'Chipson', 'Combo', 'Oakley',
'Dash', 'Hercules', 'Jay', 'Mya', 'Strider', 'Wesley', 'Solomon',
'Huck', 'Blue', 'Anakin', 'Finley', 'Sprinkles', 'Heinrich',
'Shakespeare', 'Chelsea', 'Bungalo', 'Chip', 'Grey', 'Roosevelt',
'Willem', 'Davey', 'Dakota', 'Fizz', 'Dixie', 'Al', 'Jackson',
'Carbon', 'Klein', 'DonDon', 'Kirby', 'Lou', 'Chevy', 'Tito',
'Philbert', 'Louie', 'Rupert', 'Rufus', 'Brudge', 'Shadoe', 'Angel',
'Brat', 'Tove', 'Gromit', 'Aubie', 'Kota', 'Leela', 'Glenn',
'Shelby', 'Sephie', 'Bonaparte', 'Albert', 'Wishes', 'Rose', 'Theo',
'Rocco', 'Fido', 'Emma', 'Spencer', 'Lilli', 'Boston', 'Brandonald',
'Corey', 'Leonard', 'Beckham', 'Devón', 'Gert', 'Watson', 'Keith',
'Dex', 'Ace', 'Tayzie', 'Grizzie', 'Fred', 'Gilbert', 'Meyer',

'Zoe', 'Stewie', 'Calvin', 'Lilah', 'Spanky', 'Jameson', 'Piper',
 'Atticus', 'Blu', 'Dietrich', 'Divine', 'Tripp', 'Cora', 'Huxley',
 'Keurig', 'Bookstore', 'Linus', 'Abby', 'Shiloh', 'Gustav', 'Arlen',
 'Percy', 'Lenox', 'Sugar', 'Harvey', 'Blanket', 'Geno', 'Stark',
 'Beya', 'Kilo', 'Kayla', 'Maxaroni', 'Bell', 'Doug', 'Edmund',
 'Aqua', 'Theodore', 'Baloo', 'Chase', 'Nollie', 'Rorie', 'Simba',
 'Charles', 'Bayley', 'Axel', 'Storkson', 'Remy', 'Chadrick',
 'Kellogg', 'Buckley', 'Livvie', 'Terry', 'Hermione', 'Ralpher',
 'Aldrick', 'Larry', 'Rooney', 'Crystal', 'Ziva', 'Stefan',
 'Pupcasso', 'Puff', 'Flurpson', 'Coleman', 'Enchilada', 'Raymond',
 'Rueben', 'Cilantro', 'Karll', 'Sprout', 'Blitz', 'Bloop', 'Colby',
 'Lillie', 'Ashleigh', 'Kreggory', 'Sarge', 'Luther', 'Ivar',
 'Jangle', 'Schnitzel', 'Panda', 'Berkeley', 'Ralphé', 'Charleson',
 'Clyde', 'Harnold', 'Sid', 'Pippa', 'Otis', 'Carper', 'Bowie',
 'Alexanderson', 'Suki', 'Barclay', 'Skittle', 'Ebby', 'Flávio',
 'Smokey', 'Link', 'Jennifur', 'Ozzy', 'Bluebert', 'Stephanus',
 'Bubbles', 'Zeus', 'Bertson', 'Nico', 'Michelangelo', 'Siba',
 'Calbert', 'Curtis', 'Travis', 'Thumas', 'Kanu', 'Lance', 'Opie',
 'Stubert', 'Kane', 'Olive', 'Chuckles', 'Staniel', 'Sora', 'Beemo',
 'Gunner', 'Lacy', 'Tater', 'Olaf', 'Cecil', 'Vince', 'Karma',
 'Billy', 'Walker', 'Rodney', 'Klevin', 'Malikai', 'Bobble', 'River',
 'Jebberson', 'Remington', 'Farfle', 'Jiminus', 'Harper', 'Clarkus',
 'Finnegus', 'Cupcake', 'Kathmandu', 'Ellie', 'Katie', 'Kara',
 'Adele', 'Zara', 'Ambrose', 'Jimothy', 'Bode', 'Terrenth', 'Reese',
 'Chesterson', 'Lucia', 'Bisquick', 'Ralphson', 'Socks', 'Rambo',
 'Rudy', 'Fiji', 'Rilo', 'Bilbo', 'Coopson', 'Yoda', 'Millie',
 'Chet', 'Crouton', 'Daniel', 'Kaia', 'Murphy', 'Dotsy', 'Eazy',
 'Coops', 'Fillup', 'Miley', 'Charl', 'Reagan', 'Yukon', 'CeCe',
 'Cuddles', 'Claude', 'Jessiga', 'Carter', 'Ole', 'Pherb', 'Blipson',
 'Reptar', 'Trevith', 'Berb', 'Bob', 'Colin', 'Brian', 'Oliviér',
 'Grady', 'Kobe', 'Freddery', 'Bodie', 'Dunkin', 'Wally', 'Tupawc',
 'Amber', 'Herschel', 'Edgar', 'Teddy', 'Kingsley', 'Brockly',
 'Richie', 'Molly', 'Vinscent', 'Cedrick', 'Hazel', 'Lolo', 'Eriq',
 'Phred', 'Oddie', 'Maxwell', 'Geoff', 'Covach', 'Durg', 'Fynn',
 'Ricky', 'Herald', 'Lucky', 'Ferg', 'Trip', 'Clarence', 'Hamrick',
 'Brad', 'Pubert', 'Frönq', 'Derby', 'Lizzie', 'Ember', 'Blakely',
 'Opal', 'Marq', 'Kramer', 'Barry', 'Tyrone', 'Gordon', 'Baxter',
 'Mona', 'Horace', 'Crimson', 'Birf', 'Hammond', 'Lorelei', 'Marty',
 'Brooks', 'Petrick', 'Hubertson', 'Gerbald', 'Oreo', 'Bruiser',
 'Perry', 'Bobby', 'Jeph', 'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe',
 'Tiger', 'Jiminy', 'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Darrel',
 'Taco', 'Joey', 'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses',
 'Toffee', 'Apollo', 'Carly', 'Asher', 'Glacier', 'Chuck', 'Champ',
 'Ozzie', 'Griswold', 'Cheesy', 'Moofasa', 'Hector', 'Goliath',
 'Kawhi', 'Emmie', 'Penelope', 'Willie', 'Rinna', 'Mike', 'William',
 'Dwight', 'Evy', 'Hurley', 'Rubio', 'Chompsky', 'Rascal', 'Linda',
 'Tug', 'Tango', 'Grizz', 'Jerome', 'Crumpet', 'Jessifer', 'Izzy',
 'Ralph', 'Sandy', 'Humphrey', 'Tassy', 'Juckson', 'Chuq', 'Tyrus',

```
'Karl', 'Godzilla', 'Vinnie', 'Kenneth', 'Herm', 'Bert', 'Striker',
'Donny', 'Pepper', 'Bernie', 'Buddah', 'Lenny', 'Arnold', 'Zuzu',
'Mollie', 'Laela', 'Tedders', 'Superpup', 'Rufio', 'Jeb', 'Rodman',
'Jonah', 'Chesney', 'Kenny', 'Henry', 'Bobbay', 'Mitch', 'Kaiya',
'Acro', 'Aiden', 'Obie', 'Dot', 'Shnuggles', 'Kendall', 'Jeffri',
'Steve', 'Eve', 'Mac', 'Fletcher', 'Kenzie', 'Pumpkin', 'Schnozz',
'Gustaf', 'Cheryl', 'Ed', 'Leonidas', 'Norman', 'Caryl', 'Scott',
'Taz', 'Darby', 'Jackie', 'Jazz', 'Franq', 'Pippin', 'Rolf',
'Snickers', 'Ridley', 'Cal', 'Bradley', 'Bubba', 'Tuco', 'Patch',
'Mojo', 'Batdog', 'Dylan', 'Mark', 'JD', 'Alejandro', 'Scruffers',
'Pip', 'Julius', 'Tanner', 'Sparky', 'Anthony', 'Holly', 'Jett',
'Amy', 'Sage', 'Andy', 'Mason', 'Trigger', 'Antony', 'Creg',
'Traviss', 'Gin', 'Jeffrie', 'Danny', 'Ester', 'Pluto', 'Bloo',
'Edd', 'Paull', 'Willy', 'Herb', 'Damon', 'Peanut', 'Nigel',
'Butters', 'Sandra', 'Fabio', 'Randall', 'Liam', 'Tommy', 'Ben',
'Raphael', 'Julio', 'Andru', 'Kloey', 'Shawwn', 'Skye', 'Kollin',
'Ronduh', 'Billl', 'Saydee', 'Dug', 'Tessa', 'Sully', 'Kirk',
'Ralf', 'Clarq', 'Jaspers', 'Samsom', 'Terrance', 'Harrison',
'Chaz', 'Jeremy', 'Jaycob', 'Lambeau', 'Ruffles', 'Amélie', 'Bobb',
'Banditt', 'Kevon', 'Winifred', 'Hanz', 'Churlie', 'Zeek', 'Timofy',
'Maks', 'Jomathan', 'Kallie', 'Marvin', 'Spark', 'Gòrdón', 'Jo',
'DayZ', 'Jareld', 'Torque', 'Ron', 'Skittles', 'Cleopatricia',
'Erik', 'Stu', 'Tedrick', 'Shaggy', 'Filup', 'Kial', 'Naphaniel',
'Dook', 'Hall', 'Philippe', 'Biden', 'Fwed', 'Genevieve', 'Joshwa',
'Timison', 'Bradlay', 'Pipsy', 'Clybe', 'Keet', 'Carll', 'Jockson',
'Josep', 'Lugan', 'Christopher'], dtype=object)
```

Define Rating numerator and denominator from integer to decimal

Code

```
In [30]: twitter_archive_clean['rating_numerator'] = twitter_archive_clean['rating_numerator'].a
        twitter_archive_clean['rating_denominator'] = twitter_archive_clean['rating_denominator']
```

Test

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 14 columns):
tweet_id                2175 non-null int64
in_reply_to_status_id    78 non-null float64
in_reply_to_user_id      78 non-null float64
timestamp                2175 non-null object
source                  2175 non-null object
text                    2175 non-null object
expanded_urls            2117 non-null object
```

```
rating_numerator      2175 non-null float64
rating_denominator     2175 non-null float64
name                   1390 non-null object
doggo                  2175 non-null object
floofer                2175 non-null object
pupper                 2175 non-null object
puppo                  2175 non-null object
dtypes: float64(4), int64(1), object(9)
memory usage: 254.9+ KB
```

Define Timestamp coversion to datetime format then split into three columns year, month, and day

Code

```
In [32]: # Change timestamp to datetime formart
         twitter_archive_clean['timestamp'] = pd.to_datetime(twitter_archive_clean['timestamp'])

         #extract year, month and day to new columns
         twitter_archive_clean['year'] = twitter_archive_clean['timestamp'].dt.year
         twitter_archive_clean['month'] = twitter_archive_clean['timestamp'].dt.month
         twitter_archive_clean['day'] = twitter_archive_clean['timestamp'].dt.day

         #Finally drop timestamp column
         twitter_archive_clean = twitter_archive_clean.drop('timestamp', 1)
```

Test

```
In [33]: list(twitter_archive_clean)
```

```
Out[33]: ['tweet_id',
          'in_reply_to_status_id',
          'in_reply_to_user_id',
          'source',
          'text',
          'expanded_urls',
          'rating_numerator',
          'rating_denominator',
          'name',
          'doggo',
          'floofer',
          'pupper',
          'puppo',
          'year',
          'month',
          'day']
```

image dataset

```
In [34]: image_clean = image.copy()
```

Define Drop duplicated records with same jpg_url

Code

```
In [35]: ## Duplicated jpg_url
         sum(image.jpg_url.duplicated())
```

```
Out[35]: 66
```

```
In [36]: image_clean = image_clean.drop_duplicates(subset=['jpg_url'], keep='last')
```

Test

```
In [37]: sum(image_clean['jpg_url'].duplicated())
```

```
Out[37]: 0
```

Define p1,p2 and p3 have inconsisitent upper case

```
In [38]: image_clean['p1'] = image_clean['p1'].str.lower()
         image_clean['p2'] = image_clean['p2'].str.lower()
         image_clean['p3'] = image_clean['p3'].str.lower()
```

Test

```
In [39]: image_clean.head(5)
```

```
Out[39]:
```

	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

Define Replace underscore by space in p1, p2, p3 values

Code

```
In [40]: image_clean['p1'] = image_clean['p1'].str.replace('_', ' ')
        image_clean['p2'] = image_clean['p2'].str.replace('_', ' ')
        image_clean['p3'] = image_clean['p3'].str.replace('_', ' ')
```

Test

```
In [41]: image_clean.head(5)
```

```
Out[41]:
```

	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

Define Create one column for image prediction and one column for confidence level

Code

```
In [42]: # The first true prediction (p1, p2 or p3) will be store in these lists
        dog_type = []
        confidence_list = []

        # create a function with nested if to capture the dog type and confidence level
        # from the first 'true' prediction
        def image(image_clean):
            if image_clean['p1_dog'] == True:
                dog_type.append(image_clean['p1'])
                confidence_list.append(image_clean['p1_conf'])
            elif image_clean['p2_dog'] == True:
```

```

        dog_type.append(image_clean['p2'])
        confidence_list.append(image_clean['p2_conf'])
    elif image_clean['p3_dog'] == True:
        dog_type.append(image_clean['p3'])
        confidence_list.append(image_clean['p3_conf'])
    else:
        dog_type.append('Error')
        confidence_list.append('Error')

# series objects having index the image_prediction_clean column.
image_clean.apply(image, axis=1)

# create new columns
image_clean['dog_type'] = dog_type
image_clean['confidence_list'] = confidence_list

In [43]: #drop rows that has prediction_list 'error'
        image_clean = image_clean[image_clean['dog_type'] != 'Error']

```

Test

```

In [44]: image_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1691 entries, 0 to 2073
Data columns (total 14 columns):
tweet_id      1691 non-null int64
jpg_url       1691 non-null object
img_num       1691 non-null int64
p1            1691 non-null object
p1_conf       1691 non-null float64
p1_dog        1691 non-null bool
p2            1691 non-null object
p2_conf       1691 non-null float64
p2_dog        1691 non-null bool
p3            1691 non-null object
p3_conf       1691 non-null float64
p3_dog        1691 non-null bool
dog_type      1691 non-null object
confidence_list 1691 non-null object
dtypes: bool(3), float64(3), int64(2), object(6)
memory usage: 163.5+ KB

```

twitter_api

```

In [45]: twitter_api_clean = twitter_api.copy()

```

Define Rename the id column to tweet_id

Code

```
In [46]: twitter_api_clean.rename(columns = {'id': 'tweet_id'}, inplace = True)
```

Test

```
In [47]: list(twitter_api_clean)
```

```
Out[47]: ['contributors',
          'coordinates',
          'created_at',
          'entities',
          'extended_entities',
          'favorite_count',
          'favorited',
          'geo',
          'tweet_id',
          'id_str',
          'in_reply_to_screen_name',
          'in_reply_to_status_id',
          'in_reply_to_status_id_str',
          'in_reply_to_user_id',
          'in_reply_to_user_id_str',
          'is_quote_status',
          'lang',
          'place',
          'possibly_sensitive',
          'possibly_sensitive_appealable',
          'quoted_status',
          'quoted_status_id',
          'quoted_status_id_str',
          'retweet_count',
          'retweeted',
          'retweeted_status',
          'source',
          'text',
          'truncated',
          'user']
```

```
In [48]: twitter_api_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
DatetimeIndex: 2342 entries, 1991-02-08 13:48:08.022790149 to 1998-04-12 22:37:23.555336193
```

```
Data columns (total 30 columns):
```

contributors	0 non-null float64
coordinates	0 non-null float64
created_at	2342 non-null datetime64[ns]
entities	2342 non-null object
extended_entities	1823 non-null object

```

favorite_count      2342 non-null int64
favorited           2342 non-null int64
geo                 0 non-null float64
tweet_id           2342 non-null int64
id_str             2342 non-null int64
in_reply_to_screen_name 77 non-null object
in_reply_to_status_id 77 non-null float64
in_reply_to_status_id_str 77 non-null float64
in_reply_to_user_id 77 non-null float64
in_reply_to_user_id_str 77 non-null float64
is_quote_status     2342 non-null int64
lang               2342 non-null object
place              1 non-null object
possibly_sensitive  2206 non-null float64
possibly_sensitive_appealable 2206 non-null float64
quoted_status       24 non-null object
quoted_status_id     26 non-null float64
quoted_status_id_str 26 non-null float64
retweet_count       2342 non-null int64
retweeted           2342 non-null int64
retweeted_status     168 non-null object
source             2342 non-null object
text               2342 non-null object
truncated          2342 non-null int64
user               2342 non-null object
dtypes: datetime64[ns](1), float64(11), int64(8), object(10)
memory usage: 567.2+ KB

```

Define 168 records have a retweeted_status which will need to be excluded

Code

```

In [49]: # find tweet_id with valid retweeted_status
         twitter_api_clean[twitter_api_clean['retweeted_status'].notnull()==True]

```

```

Out[49]:
          contributors  coordinates  created_at \
1991-02-26 06:48:02.905632768      NaN      NaN 2015-11-20 03:51:47
1991-02-26 06:48:24.950915073      NaN      NaN 2015-11-20 03:51:52
1992-07-24 17:26:49.858043904      NaN      NaN 2016-03-21 19:31:59
1993-07-28 05:05:15.802583040      NaN      NaN 2016-06-17 16:01:16
1993-08-28 07:04:05.350707200      NaN      NaN 2016-06-25 01:52:36
1993-09-05 15:18:28.580548608      NaN      NaN 2016-06-27 01:37:04
1993-11-03 06:49:54.570878976      NaN      NaN 2016-07-11 01:11:51
1993-11-07 19:52:24.171524096      NaN      NaN 2016-07-12 03:11:42
1993-11-14 17:37:14.498793472      NaN      NaN 2016-07-13 18:42:44
1993-12-02 23:27:21.593970688      NaN      NaN 2016-07-18 03:06:01
1994-01-03 11:51:44.299253760      NaN      NaN 2016-07-25 15:26:30

```

1994-01-05 00:19:23.776290825	NaN	NaN 2016-07-26 00:08:05
1994-01-21 13:45:34.323924993	NaN	NaN 2016-07-29 22:53:27
1994-01-26 06:47:08.574212096	NaN	NaN 2016-07-31 01:50:18
1994-02-02 01:52:29.710192640	NaN	NaN 2016-08-01 16:43:19
1994-02-16 03:57:17.149827077	NaN	NaN 2016-08-05 01:19:35
1994-02-20 13:21:42.866649088	NaN	NaN 2016-08-06 02:27:27
1994-03-08 22:51:03.695355904	NaN	NaN 2016-08-10 00:16:21
1994-04-11 15:28:12.750233600	NaN	NaN 2016-08-18 01:03:45
1994-04-20 17:54:21.642756096	NaN	NaN 2016-08-20 05:08:29
1994-05-10 07:15:58.521745409	NaN	NaN 2016-08-24 21:02:45
1994-05-14 10:02:47.477751808	NaN	NaN 2016-08-25 20:35:48
1994-05-19 08:19:51.808995329	NaN	NaN 2016-08-27 00:47:53
1994-05-28 02:56:07.776997377	NaN	NaN 2016-08-29 03:00:36
1994-06-07 15:53:14.259247104	NaN	NaN 2016-08-31 15:19:06
1994-06-09 14:10:53.431250945	NaN	NaN 2016-09-01 02:21:21
1994-06-26 07:22:04.260794368	NaN	NaN 2016-09-05 02:00:22
1994-07-04 15:46:27.167145985	NaN	NaN 2016-09-07 01:47:12
1994-08-03 07:24:21.951791106	NaN	NaN 2016-09-14 03:27:11
1994-08-07 08:58:26.839351296	NaN	NaN 2016-09-15 02:42:54
...
1996-09-04 10:46:33.020538882	NaN	NaN 2017-03-15 02:10:39
1996-10-12 05:32:39.547420673	NaN	NaN 2017-03-24 02:22:04
1996-11-14 11:39:34.464610304	NaN	NaN 2017-04-01 00:39:12
1996-11-14 13:41:05.427394560	NaN	NaN 2017-04-01 01:08:10
1996-12-04 02:54:54.696017920	NaN	NaN 2017-04-05 17:00:34
1996-12-29 12:09:45.021730816	NaN	NaN 2017-04-11 18:15:55
1996-12-30 13:51:42.622658560	NaN	NaN 2017-04-12 00:23:33
1997-01-10 22:46:45.516943360	NaN	NaN 2017-04-14 17:27:40
1997-02-05 10:24:01.867124737	NaN	NaN 2017-04-20 19:16:59
1997-02-06 16:08:43.840757760	NaN	NaN 2017-04-21 02:22:29
1997-02-19 05:40:35.276025856	NaN	NaN 2017-04-24 02:15:55
1997-02-22 09:16:33.587888130	NaN	NaN 2017-04-24 20:17:23
1997-02-27 16:48:23.051644929	NaN	NaN 2017-04-26 02:41:43
1997-03-20 12:19:50.427611136	NaN	NaN 2017-05-01 01:47:28
1997-04-04 18:13:13.139703809	NaN	NaN 2017-05-04 17:01:34
1997-04-13 09:33:55.999428608	NaN	NaN 2017-05-06 18:27:40
1997-04-14 01:34:34.716409858	NaN	NaN 2017-05-06 22:16:42
1997-05-12 21:16:22.782697472	NaN	NaN 2017-05-13 19:11:30
1997-06-12 05:48:47.597207552	NaN	NaN 2017-05-21 00:53:21
1997-06-23 13:30:53.475098625	NaN	NaN 2017-05-23 17:40:04
1997-07-11 16:44:37.480148993	NaN	NaN 2017-05-28 01:26:04
1997-08-09 22:36:19.821445120	NaN	NaN 2017-06-04 00:46:17
1997-09-04 01:49:08.698140672	NaN	NaN 2017-06-10 00:35:19
1997-09-16 18:33:38.259525634	NaN	NaN 2017-06-13 01:14:41
1997-10-23 12:52:52.832227328	NaN	NaN 2017-06-21 19:36:23
1997-10-31 16:41:50.768087041	NaN	NaN 2017-06-23 18:17:33
1997-11-01 17:19:37.348136964	NaN	NaN 2017-06-24 00:09:53
1997-11-10 02:56:19.576475649	NaN	NaN 2017-06-26 00:13:58

1998-01-20 15:53:12.912609280	NaN	NaN 2017-07-13 01:35:06
1998-01-29 06:09:20.059072513	NaN	NaN 2017-07-15 02:45:48

		entities \
1991-02-26 06:48:02.905632768	{'hashtags': [], 'media': [{'display_url': 'pi...	
1991-02-26 06:48:24.950915073	{'hashtags': [], 'media': [{'display_url': 'pi...	
1992-07-24 17:26:49.858043904	{'hashtags': [{'indices': [61, 73], 'text': 'L...	
1993-07-28 05:05:15.802583040	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1993-08-28 07:04:05.350707200	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1993-09-05 15:18:28.580548608	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1993-11-03 06:49:54.570878976	{'hashtags': [], 'media': [{'display_url': 'pi...	
1993-11-07 19:52:24.171524096	{'hashtags': [], 'symbols': [], 'urls': [{'dis...	
1993-11-14 17:37:14.498793472	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1993-12-02 23:27:21.593970688	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-01-03 11:51:44.299253760	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-01-05 00:19:23.776290825	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-01-21 13:45:34.323924993	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-01-26 06:47:08.574212096	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-02-02 01:52:29.710192640	{'hashtags': [], 'symbols': [], 'urls': [{'dis...	
1994-02-16 03:57:17.149827077	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-02-20 13:21:42.866649088	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-03-08 22:51:03.695355904	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1994-04-11 15:28:12.750233600	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-04-20 17:54:21.642756096	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1994-05-10 07:15:58.521745409	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1994-05-14 10:02:47.477751808	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1994-05-19 08:19:51.808995329	{'hashtags': [], 'symbols': [], 'urls': [{'dis...	
1994-05-28 02:56:07.776997377	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-06-07 15:53:14.259247104	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-06-09 14:10:53.431250945	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-06-26 07:22:04.260794368	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-07-04 15:46:27.167145985	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1994-08-03 07:24:21.951791106	{'hashtags': [], 'media': [{'display_url': 'pi...	
1994-08-07 08:58:26.839351296	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
...	...	
1996-09-04 10:46:33.020538882	{'hashtags': [], 'media': [{'display_url': 'pi...	
1996-10-12 05:32:39.547420673	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1996-11-14 11:39:34.464610304	{'hashtags': [], 'media': [{'display_url': 'pi...	
1996-11-14 13:41:05.427394560	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1996-12-04 02:54:54.696017920	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1996-12-29 12:09:45.021730816	{'hashtags': [], 'media': [{'display_url': 'pi...	
1996-12-30 13:51:42.622658560	{'hashtags': [], 'media': [{'display_url': 'pi...	
1997-01-10 22:46:45.516943360	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1997-02-05 10:24:01.867124737	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1997-02-06 16:08:43.840757760	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1997-02-19 05:40:35.276025856	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1997-02-22 09:16:33.587888130	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	
1997-02-27 16:48:23.051644929	{'hashtags': [], 'symbols': [], 'urls': [], 'u...	

1997-03-20 12:19:50.427611136	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-04-04 18:13:13.139703809	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-04-13 09:33:55.999428608	{'hashtags': [], 'media': [{'display_url': 'pi...
1997-04-14 01:34:34.716409858	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-05-12 21:16:22.782697472	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-06-12 05:48:47.597207552	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-06-23 13:30:53.475098625	{'hashtags': [], 'media': [{'display_url': 'pi...
1997-07-11 16:44:37.480148993	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-08-09 22:36:19.821445120	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-09-04 01:49:08.698140672	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-09-16 18:33:38.259525634	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-10-23 12:52:52.832227328	{'hashtags': [], 'media': [{'display_url': 'pi...
1997-10-31 16:41:50.768087041	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-11-01 17:19:37.348136964	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1997-11-10 02:56:19.576475649	{'hashtags': [], 'symbols': [], 'urls': [], 'u...
1998-01-20 15:53:12.912609280	{'hashtags': [], 'media': [{'display_url': 'pi...
1998-01-29 06:09:20.059072513	{'hashtags': [{'indices': [21, 26], 'text': 'B...

extended_entities \

1991-02-26 06:48:02.905632768	{'media': [{'display_url': 'pic.twitter.com/r7...
1991-02-26 06:48:24.950915073	{'media': [{'display_url': 'pic.twitter.com/ua...
1992-07-24 17:26:49.858043904	{'media': [{'display_url': 'pic.twitter.com/j6...
1993-07-28 05:05:15.802583040	NaN
1993-08-28 07:04:05.350707200	NaN
1993-09-05 15:18:28.580548608	NaN
1993-11-03 06:49:54.570878976	{'media': [{'additional_media_info': {'monetiz...
1993-11-07 19:52:24.171524096	NaN
1993-11-14 17:37:14.498793472	NaN
1993-12-02 23:27:21.593970688	{'media': [{'display_url': 'pic.twitter.com/NL...
1994-01-03 11:51:44.299253760	{'media': [{'display_url': 'pic.twitter.com/7b...
1994-01-05 00:19:23.776290825	{'media': [{'display_url': 'pic.twitter.com/Xl...
1994-01-21 13:45:34.323924993	{'media': [{'display_url': 'pic.twitter.com/Qn...
1994-01-26 06:47:08.574212096	{'media': [{'display_url': 'pic.twitter.com/zx...
1994-02-02 01:52:29.710192640	NaN
1994-02-16 03:57:17.149827077	{'media': [{'display_url': 'pic.twitter.com/Ez...
1994-02-20 13:21:42.866649088	{'media': [{'display_url': 'pic.twitter.com/VG...
1994-03-08 22:51:03.695355904	NaN
1994-04-11 15:28:12.750233600	{'media': [{'display_url': 'pic.twitter.com/vt...
1994-04-20 17:54:21.642756096	NaN
1994-05-10 07:15:58.521745409	NaN
1994-05-14 10:02:47.477751808	NaN
1994-05-19 08:19:51.808995329	NaN
1994-05-28 02:56:07.776997377	{'media': [{'display_url': 'pic.twitter.com/o5...
1994-06-07 15:53:14.259247104	{'media': [{'display_url': 'pic.twitter.com/nT...
1994-06-09 14:10:53.431250945	{'media': [{'display_url': 'pic.twitter.com/iJ...
1994-06-26 07:22:04.260794368	{'media': [{'display_url': 'pic.twitter.com/eS...
1994-07-04 15:46:27.167145985	NaN
1994-08-03 07:24:21.951791106	{'media': [{'display_url': 'pic.twitter.com/pG...

1994-08-07 08:58:26.839351296		NaN
...		...
1996-09-04 10:46:33.020538882	{'media': [{'additional_media_info': {'monetiz...	
1996-10-12 05:32:39.547420673		NaN
1996-11-14 11:39:34.464610304	{'media': [{'display_url': 'pic.twitter.com/Ob...	
1996-11-14 13:41:05.427394560		NaN
1996-12-04 02:54:54.696017920		NaN
1996-12-29 12:09:45.021730816	{'media': [{'display_url': 'pic.twitter.com/b0...	
1996-12-30 13:51:42.622658560	{'media': [{'display_url': 'pic.twitter.com/xo...	
1997-01-10 22:46:45.516943360		NaN
1997-02-05 10:24:01.867124737		NaN
1997-02-06 16:08:43.840757760		NaN
1997-02-19 05:40:35.276025856		NaN
1997-02-22 09:16:33.587888130		NaN
1997-02-27 16:48:23.051644929		NaN
1997-03-20 12:19:50.427611136		NaN
1997-04-04 18:13:13.139703809		NaN
1997-04-13 09:33:55.999428608	{'media': [{'display_url': 'pic.twitter.com/6n...	
1997-04-14 01:34:34.716409858		NaN
1997-05-12 21:16:22.782697472		NaN
1997-06-12 05:48:47.597207552		NaN
1997-06-23 13:30:53.475098625	{'media': [{'display_url': 'pic.twitter.com/nR...	
1997-07-11 16:44:37.480148993		NaN
1997-08-09 22:36:19.821445120		NaN
1997-09-04 01:49:08.698140672		NaN
1997-09-16 18:33:38.259525634		NaN
1997-10-23 12:52:52.832227328	{'media': [{'display_url': 'pic.twitter.com/dJ...	
1997-10-31 16:41:50.768087041		NaN
1997-11-01 17:19:37.348136964		NaN
1997-11-10 02:56:19.576475649		NaN
1998-01-20 15:53:12.912609280	{'media': [{'display_url': 'pic.twitter.com/SA...	
1998-01-29 06:09:20.059072513		NaN

	favorite_count	favorited	geo \
1991-02-26 06:48:02.905632768	0	0	NaN
1991-02-26 06:48:24.950915073	0	0	NaN
1992-07-24 17:26:49.858043904	0	0	NaN
1993-07-28 05:05:15.802583040	0	0	NaN
1993-08-28 07:04:05.350707200	0	0	NaN
1993-09-05 15:18:28.580548608	0	0	NaN
1993-11-03 06:49:54.570878976	0	0	NaN
1993-11-07 19:52:24.171524096	0	0	NaN
1993-11-14 17:37:14.498793472	0	0	NaN
1993-12-02 23:27:21.593970688	0	0	NaN
1994-01-03 11:51:44.299253760	0	0	NaN
1994-01-05 00:19:23.776290825	0	0	NaN
1994-01-21 13:45:34.323924993	0	0	NaN
1994-01-26 06:47:08.574212096	0	0	NaN

1994-02-02 01:52:29.710192640	0	0	NaN
1994-02-16 03:57:17.149827077	0	0	NaN
1994-02-20 13:21:42.866649088	0	0	NaN
1994-03-08 22:51:03.695355904	0	0	NaN
1994-04-11 15:28:12.750233600	0	0	NaN
1994-04-20 17:54:21.642756096	0	0	NaN
1994-05-10 07:15:58.521745409	0	0	NaN
1994-05-14 10:02:47.477751808	0	0	NaN
1994-05-19 08:19:51.808995329	0	0	NaN
1994-05-28 02:56:07.776997377	0	0	NaN
1994-06-07 15:53:14.259247104	0	0	NaN
1994-06-09 14:10:53.431250945	0	0	NaN
1994-06-26 07:22:04.260794368	0	0	NaN
1994-07-04 15:46:27.167145985	0	0	NaN
1994-08-03 07:24:21.951791106	0	0	NaN
1994-08-07 08:58:26.839351296	0	0	NaN
...
1996-09-04 10:46:33.020538882	0	0	NaN
1996-10-12 05:32:39.547420673	0	0	NaN
1996-11-14 11:39:34.464610304	0	0	NaN
1996-11-14 13:41:05.427394560	0	0	NaN
1996-12-04 02:54:54.696017920	0	0	NaN
1996-12-29 12:09:45.021730816	0	0	NaN
1996-12-30 13:51:42.622658560	0	0	NaN
1997-01-10 22:46:45.516943360	0	0	NaN
1997-02-05 10:24:01.867124737	0	0	NaN
1997-02-06 16:08:43.840757760	0	0	NaN
1997-02-19 05:40:35.276025856	0	0	NaN
1997-02-22 09:16:33.587888130	0	0	NaN
1997-02-27 16:48:23.051644929	0	0	NaN
1997-03-20 12:19:50.427611136	0	0	NaN
1997-04-04 18:13:13.139703809	0	0	NaN
1997-04-13 09:33:55.999428608	0	0	NaN
1997-04-14 01:34:34.716409858	0	0	NaN
1997-05-12 21:16:22.782697472	0	0	NaN
1997-06-12 05:48:47.597207552	0	0	NaN
1997-06-23 13:30:53.475098625	0	0	NaN
1997-07-11 16:44:37.480148993	0	0	NaN
1997-08-09 22:36:19.821445120	0	0	NaN
1997-09-04 01:49:08.698140672	0	0	NaN
1997-09-16 18:33:38.259525634	0	0	NaN
1997-10-23 12:52:52.832227328	0	0	NaN
1997-10-31 16:41:50.768087041	0	0	NaN
1997-11-01 17:19:37.348136964	0	0	NaN
1997-11-10 02:56:19.576475649	0	0	NaN
1998-01-20 15:53:12.912609280	0	0	NaN
1998-01-29 06:09:20.059072513	0	0	NaN

		tweet_id	id_str \
1991-02-26 06:48:02.905632768	667550882905632768	667550882905632768	
1991-02-26 06:48:24.950915073	667550904950915072	667550904950915072	
1992-07-24 17:26:49.858043904	711998809858043904	711998809858043904	
1993-07-28 05:05:15.802583040	743835915802583040	743835915802583040	
1993-08-28 07:04:05.350707200	746521445350707200	746521445350707200	
1993-09-05 15:18:28.580548608	747242308580548608	747242308580548608	
1993-11-03 06:49:54.570878976	752309394570878976	752309394570878976	
1993-11-07 19:52:24.171524096	752701944171524096	752701944171524096	
1993-11-14 17:37:14.498793472	753298634498793472	753298634498793472	
1993-12-02 23:27:21.593970688	754874841593970688	754874841593970688	
1994-01-03 11:51:44.299253760	757597904299253760	757597904299253760	
1994-01-05 00:19:23.776290825	757729163776290816	757729163776290816	
1994-01-21 13:45:34.323924993	759159934323924992	759159934323924992	
1994-01-26 06:47:08.574212096	759566828574212096	759566828574212096	
1994-02-02 01:52:29.710192640	760153949710192640	760153949710192640	
1994-02-16 03:57:17.149827077	761371037149827072	761371037149827072	
1994-02-20 13:21:42.866649088	761750502866649088	761750502866649088	
1994-03-08 22:51:03.695355904	763167063695355904	763167063695355904	
1994-04-11 15:28:12.750233600	766078092750233600	766078092750233600	
1994-04-20 17:54:21.642756096	766864461642756096	766864461642756096	
1994-05-10 07:15:58.521745409	768554158521745408	768554158521745408	
1994-05-14 10:02:47.477751808	768909767477751808	768909767477751808	
1994-05-19 08:19:51.808995329	769335591808995328	769335591808995328	
1994-05-28 02:56:07.776997377	770093767776997376	770093767776997376	
1994-06-07 15:53:14.259247104	771004394259247104	771004394259247104	
1994-06-09 14:10:53.431250945	771171053431250944	771171053431250944	
1994-06-26 07:22:04.260794368	772615324260794368	772615324260794368	
1994-07-04 15:46:27.167145985	773336787167145984	773336787167145984	
1994-08-03 07:24:21.951791106	775898661951791104	775898661951791104	
1994-08-07 08:58:26.839351296	776249906839351296	776249906839351296	
...	
1996-09-04 10:46:33.020538882	841833993020538880	841833993020538880	
1996-10-12 05:32:39.547420673	845098359547420672	845098359547420672	
1996-11-14 11:39:34.464610304	847971574464610304	847971574464610304	
1996-11-14 13:41:05.427394560	847978865427394560	847978865427394560	
1996-12-04 02:54:54.696017920	849668094696017920	849668094696017920	
1996-12-29 12:09:45.021730816	851861385021730816	851861385021730816	
1996-12-30 13:51:42.622658560	851953902622658560	851953902622658560	
1997-01-10 22:46:45.516943360	852936405516943360	852936405516943360	
1997-02-05 10:24:01.867124737	855138241867124736	855138241867124736	
1997-02-06 16:08:43.840757760	855245323840757760	855245323840757760	
1997-02-19 05:40:35.276025856	856330835276025856	856330835276025856	
1997-02-22 09:16:33.587888130	856602993587888128	856602993587888128	
1997-02-27 16:48:23.051644929	857062103051644928	857062103051644928	
1997-03-20 12:19:50.427611136	858860390427611136	858860390427611136	
1997-04-04 18:13:13.139703809	860177593139703808	860177593139703808	
1997-04-13 09:33:55.999428608	860924035999428608	860924035999428608	

1997-04-14	01:34:34.716409858	860981674716409856	860981674716409856
1997-05-12	21:16:22.782697472	863471782782697472	863471782782697472
1997-06-12	05:48:47.597207552	866094527597207552	866094527597207552
1997-06-23	13:30:53.475098625	867072653475098624	867072653475098624
1997-07-11	16:44:37.480148993	868639477480148992	868639477480148992
1997-08-09	22:36:19.821445120	871166179821445120	871166179821445120
1997-09-04	01:49:08.698140672	873337748698140672	873337748698140672
1997-09-16	18:33:38.259525634	874434818259525632	874434818259525632
1997-10-23	12:52:52.832227328	877611172832227328	877611172832227328
1997-10-31	16:41:50.768087041	878316110768087040	878316110768087040
1997-11-01	17:19:37.348136964	878404777348136960	878404777348136960
1997-11-10	02:56:19.576475649	879130579576475648	879130579576475648
1998-01-20	15:53:12.912609280	885311592912609280	885311592912609280
1998-01-29	06:09:20.059072513	886054160059072512	886054160059072512

1991-02-26	06:48:02.905632768	...
1991-02-26	06:48:24.950915073	...
1992-07-24	17:26:49.858043904	...
1993-07-28	05:05:15.802583040	...
1993-08-28	07:04:05.350707200	...
1993-09-05	15:18:28.580548608	...
1993-11-03	06:49:54.570878976	...
1993-11-07	19:52:24.171524096	...
1993-11-14	17:37:14.498793472	...
1993-12-02	23:27:21.593970688	...
1994-01-03	11:51:44.299253760	...
1994-01-05	00:19:23.776290825	...
1994-01-21	13:45:34.323924993	...
1994-01-26	06:47:08.574212096	...
1994-02-02	01:52:29.710192640	...
1994-02-16	03:57:17.149827077	...
1994-02-20	13:21:42.866649088	...
1994-03-08	22:51:03.695355904	...
1994-04-11	15:28:12.750233600	...
1994-04-20	17:54:21.642756096	...
1994-05-10	07:15:58.521745409	...
1994-05-14	10:02:47.477751808	...
1994-05-19	08:19:51.808995329	...
1994-05-28	02:56:07.776997377	...
1994-06-07	15:53:14.259247104	...
1994-06-09	14:10:53.431250945	...
1994-06-26	07:22:04.260794368	...
1994-07-04	15:46:27.167145985	...
1994-08-03	07:24:21.951791106	...
1994-08-07	08:58:26.839351296	...
...		...
1996-09-04	10:46:33.020538882	...

1996-10-12 05:32:39.547420673	...
1996-11-14 11:39:34.464610304	...
1996-11-14 13:41:05.427394560	...
1996-12-04 02:54:54.696017920	...
1996-12-29 12:09:45.021730816	...
1996-12-30 13:51:42.622658560	...
1997-01-10 22:46:45.516943360	...
1997-02-05 10:24:01.867124737	...
1997-02-06 16:08:43.840757760	...
1997-02-19 05:40:35.276025856	...
1997-02-22 09:16:33.587888130	...
1997-02-27 16:48:23.051644929	...
1997-03-20 12:19:50.427611136	...
1997-04-04 18:13:13.139703809	...
1997-04-13 09:33:55.999428608	...
1997-04-14 01:34:34.716409858	...
1997-05-12 21:16:22.782697472	...
1997-06-12 05:48:47.597207552	...
1997-06-23 13:30:53.475098625	...
1997-07-11 16:44:37.480148993	...
1997-08-09 22:36:19.821445120	...
1997-09-04 01:49:08.698140672	...
1997-09-16 18:33:38.259525634	...
1997-10-23 12:52:52.832227328	...
1997-10-31 16:41:50.768087041	...
1997-11-01 17:19:37.348136964	...
1997-11-10 02:56:19.576475649	...
1998-01-20 15:53:12.912609280	...
1998-01-29 06:09:20.059072513	...

	quoted_status	quoted_status_id \
1991-02-26 06:48:02.905632768	NaN	NaN
1991-02-26 06:48:24.950915073	NaN	NaN
1992-07-24 17:26:49.858043904	NaN	NaN
1993-07-28 05:05:15.802583040	NaN	NaN
1993-08-28 07:04:05.350707200	NaN	NaN
1993-09-05 15:18:28.580548608	NaN	NaN
1993-11-03 06:49:54.570878976	NaN	NaN
1993-11-07 19:52:24.171524096	NaN	NaN
1993-11-14 17:37:14.498793472	NaN	NaN
1993-12-02 23:27:21.593970688	NaN	NaN
1994-01-03 11:51:44.299253760	NaN	NaN
1994-01-05 00:19:23.776290825	NaN	NaN
1994-01-21 13:45:34.323924993	NaN	NaN
1994-01-26 06:47:08.574212096	NaN	NaN
1994-02-02 01:52:29.710192640	NaN	NaN
1994-02-16 03:57:17.149827077	NaN	NaN
1994-02-20 13:21:42.866649088	NaN	NaN

1994-03-08 22:51:03.695355904	NaN	NaN
1994-04-11 15:28:12.750233600	NaN	NaN
1994-04-20 17:54:21.642756096	NaN	NaN
1994-05-10 07:15:58.521745409	NaN	NaN
1994-05-14 10:02:47.477751808	NaN	NaN
1994-05-19 08:19:51.808995329	NaN	NaN
1994-05-28 02:56:07.776997377	NaN	NaN
1994-06-07 15:53:14.259247104	NaN	NaN
1994-06-09 14:10:53.431250945	NaN	NaN
1994-06-26 07:22:04.260794368	NaN	NaN
1994-07-04 15:46:27.167145985	NaN	NaN
1994-08-03 07:24:21.951791106	NaN	NaN
1994-08-07 08:58:26.839351296	NaN	NaN
...
1996-09-04 10:46:33.020538882	NaN	NaN
1996-10-12 05:32:39.547420673	NaN	NaN
1996-11-14 11:39:34.464610304	NaN	NaN
1996-11-14 13:41:05.427394560	NaN	NaN
1996-12-04 02:54:54.696017920	NaN	NaN
1996-12-29 12:09:45.021730816	NaN	NaN
1996-12-30 13:51:42.622658560	NaN	NaN
1997-01-10 22:46:45.516943360	NaN	NaN
1997-02-05 10:24:01.867124737	NaN	NaN
1997-02-06 16:08:43.840757760	NaN	NaN
1997-02-19 05:40:35.276025856	NaN	NaN
1997-02-22 09:16:33.587888130	NaN	NaN
1997-02-27 16:48:23.051644929	NaN	NaN
1997-03-20 12:19:50.427611136	NaN	NaN
1997-04-04 18:13:13.139703809	NaN	NaN
1997-04-13 09:33:55.999428608	NaN	NaN
1997-04-14 01:34:34.716409858	NaN	NaN
1997-05-12 21:16:22.782697472	NaN	NaN
1997-06-12 05:48:47.597207552	NaN	NaN
1997-06-23 13:30:53.475098625	NaN	NaN
1997-07-11 16:44:37.480148993	NaN	NaN
1997-08-09 22:36:19.821445120	NaN	NaN
1997-09-04 01:49:08.698140672	NaN	NaN
1997-09-16 18:33:38.259525634	NaN	NaN
1997-10-23 12:52:52.832227328	NaN	NaN
1997-10-31 16:41:50.768087041	NaN	NaN
1997-11-01 17:19:37.348136964	NaN	NaN
1997-11-10 02:56:19.576475649	NaN	NaN
1998-01-20 15:53:12.912609280	NaN	NaN
1998-01-29 06:09:20.059072513	NaN	8.860534e+17

	quoted_status_id_str	retweet_count	retweeted	\
1991-02-26 06:48:02.905632768	NaN	33	0	
1991-02-26 06:48:24.950915073	NaN	35	0	

1992-07-24 17:26:49.858043904	NaN	134	0
1993-07-28 05:05:15.802583040	NaN	2245	0
1993-08-28 07:04:05.350707200	NaN	1058	0
1993-09-05 15:18:28.580548608	NaN	3104	0
1993-11-03 06:49:54.570878976	NaN	18086	0
1993-11-07 19:52:24.171524096	NaN	3130	0
1993-11-14 17:37:14.498793472	NaN	6254	0
1993-12-02 23:27:21.593970688	NaN	8695	0
1994-01-03 11:51:44.299253760	NaN	317	0
1994-01-05 00:19:23.776290825	NaN	8796	0
1994-01-21 13:45:34.323924993	NaN	1267	0
1994-01-26 06:47:08.574212096	NaN	22970	0
1994-02-02 01:52:29.710192640	NaN	34	0
1994-02-16 03:57:17.149827077	NaN	19526	0
1994-02-20 13:21:42.866649088	NaN	4290	0
1994-03-08 22:51:03.695355904	NaN	3287	0
1994-04-11 15:28:12.750233600	NaN	2831	0
1994-04-20 17:54:21.642756096	NaN	6171	0
1994-05-10 07:15:58.521745409	NaN	6350	0
1994-05-14 10:02:47.477751808	NaN	2950	0
1994-05-19 08:19:51.808995329	NaN	8383	0
1994-05-28 02:56:07.776997377	NaN	3315	0
1994-06-07 15:53:14.259247104	NaN	240	0
1994-06-09 14:10:53.431250945	NaN	8261	0
1994-06-26 07:22:04.260794368	NaN	3690	0
1994-07-04 15:46:27.167145985	NaN	5569	0
1994-08-03 07:24:21.951791106	NaN	17999	0
1994-08-07 08:58:26.839351296	NaN	10038	0
...
1996-09-04 10:46:33.020538882	NaN	16436	0
1996-10-12 05:32:39.547420673	NaN	8173	0
1996-11-14 11:39:34.464610304	NaN	457	0
1996-11-14 13:41:05.427394560	NaN	3453	0
1996-12-04 02:54:54.696017920	NaN	5243	0
1996-12-29 12:09:45.021730816	NaN	23	0
1996-12-30 13:51:42.622658560	NaN	10138	0
1997-01-10 22:46:45.516943360	NaN	2105	0
1997-02-05 10:24:01.867124737	NaN	48	0
1997-02-06 16:08:43.840757760	NaN	6210	0
1997-02-19 05:40:35.276025856	NaN	702	0
1997-02-22 09:16:33.587888130	NaN	11047	0
1997-02-27 16:48:23.051644929	NaN	177	0
1997-03-20 12:19:50.427611136	NaN	8361	0
1997-04-04 18:13:13.139703809	NaN	31583	0
1997-04-13 09:33:55.999428608	NaN	840	0
1997-04-14 01:34:34.716409858	NaN	2204	0
1997-05-12 21:16:22.782697472	NaN	2498	0
1997-06-12 05:48:47.597207552	NaN	8471	0

1997-06-23	13:30:53.475098625	NaN	121	0
1997-07-11	16:44:37.480148993	NaN	2108	0
1997-08-09	22:36:19.821445120	NaN	5691	0
1997-09-04	01:49:08.698140672	NaN	1578	0
1997-09-16	18:33:38.259525634	NaN	14618	0
1997-10-23	12:52:52.832227328	NaN	80	0
1997-10-31	16:41:50.768087041	NaN	6576	0
1997-11-01	17:19:37.348136964	NaN	1273	0
1997-11-10	02:56:19.576475649	NaN	6741	0
1998-01-20	15:53:12.912609280	NaN	18288	0
1998-01-29	06:09:20.059072513	8.860534e+17	105	0

		retweeted_status \
1991-02-26	06:48:02.905632768	{'contributors': None, 'coordinates': None, 'c...
1991-02-26	06:48:24.950915073	{'contributors': None, 'coordinates': None, 'c...
1992-07-24	17:26:49.858043904	{'contributors': None, 'coordinates': None, 'c...
1993-07-28	05:05:15.802583040	{'contributors': None, 'coordinates': None, 'c...
1993-08-28	07:04:05.350707200	{'contributors': None, 'coordinates': None, 'c...
1993-09-05	15:18:28.580548608	{'contributors': None, 'coordinates': None, 'c...
1993-11-03	06:49:54.570878976	{'contributors': None, 'coordinates': None, 'c...
1993-11-07	19:52:24.171524096	{'contributors': None, 'coordinates': None, 'c...
1993-11-14	17:37:14.498793472	{'contributors': None, 'coordinates': None, 'c...
1993-12-02	23:27:21.593970688	{'contributors': None, 'coordinates': None, 'c...
1994-01-03	11:51:44.299253760	{'contributors': None, 'coordinates': None, 'c...
1994-01-05	00:19:23.776290825	{'contributors': None, 'coordinates': None, 'c...
1994-01-21	13:45:34.323924993	{'contributors': None, 'coordinates': None, 'c...
1994-01-26	06:47:08.574212096	{'contributors': None, 'coordinates': None, 'c...
1994-02-02	01:52:29.710192640	{'contributors': None, 'coordinates': None, 'c...
1994-02-16	03:57:17.149827077	{'contributors': None, 'coordinates': None, 'c...
1994-02-20	13:21:42.866649088	{'contributors': None, 'coordinates': None, 'c...
1994-03-08	22:51:03.695355904	{'contributors': None, 'coordinates': None, 'c...
1994-04-11	15:28:12.750233600	{'contributors': None, 'coordinates': None, 'c...
1994-04-20	17:54:21.642756096	{'contributors': None, 'coordinates': None, 'c...
1994-05-10	07:15:58.521745409	{'contributors': None, 'coordinates': None, 'c...
1994-05-14	10:02:47.477751808	{'contributors': None, 'coordinates': None, 'c...
1994-05-19	08:19:51.808995329	{'contributors': None, 'coordinates': None, 'c...
1994-05-28	02:56:07.776997377	{'contributors': None, 'coordinates': None, 'c...
1994-06-07	15:53:14.259247104	{'contributors': None, 'coordinates': None, 'c...
1994-06-09	14:10:53.431250945	{'contributors': None, 'coordinates': None, 'c...
1994-06-26	07:22:04.260794368	{'contributors': None, 'coordinates': None, 'c...
1994-07-04	15:46:27.167145985	{'contributors': None, 'coordinates': None, 'c...
1994-08-03	07:24:21.951791106	{'contributors': None, 'coordinates': None, 'c...
1994-08-07	08:58:26.839351296	{'contributors': None, 'coordinates': None, 'c...
...		...
1996-09-04	10:46:33.020538882	{'contributors': None, 'coordinates': None, 'c...
1996-10-12	05:32:39.547420673	{'contributors': None, 'coordinates': None, 'c...
1996-11-14	11:39:34.464610304	{'contributors': None, 'coordinates': None, 'c...
1996-11-14	13:41:05.427394560	{'contributors': None, 'coordinates': None, 'c...

1996-12-04 02:54:54.696017920	{'contributors': None, 'coordinates': None, 'c...
1996-12-29 12:09:45.021730816	{'contributors': None, 'coordinates': None, 'c...
1996-12-30 13:51:42.622658560	{'contributors': None, 'coordinates': None, 'c...
1997-01-10 22:46:45.516943360	{'contributors': None, 'coordinates': None, 'c...
1997-02-05 10:24:01.867124737	{'contributors': None, 'coordinates': None, 'c...
1997-02-06 16:08:43.840757760	{'contributors': None, 'coordinates': None, 'c...
1997-02-19 05:40:35.276025856	{'contributors': None, 'coordinates': None, 'c...
1997-02-22 09:16:33.587888130	{'contributors': None, 'coordinates': None, 'c...
1997-02-27 16:48:23.051644929	{'contributors': None, 'coordinates': None, 'c...
1997-03-20 12:19:50.427611136	{'contributors': None, 'coordinates': None, 'c...
1997-04-04 18:13:13.139703809	{'contributors': None, 'coordinates': None, 'c...
1997-04-13 09:33:55.999428608	{'contributors': None, 'coordinates': None, 'c...
1997-04-14 01:34:34.716409858	{'contributors': None, 'coordinates': None, 'c...
1997-05-12 21:16:22.782697472	{'contributors': None, 'coordinates': None, 'c...
1997-06-12 05:48:47.597207552	{'contributors': None, 'coordinates': None, 'c...
1997-06-23 13:30:53.475098625	{'contributors': None, 'coordinates': None, 'c...
1997-07-11 16:44:37.480148993	{'contributors': None, 'coordinates': None, 'c...
1997-08-09 22:36:19.821445120	{'contributors': None, 'coordinates': None, 'c...
1997-09-04 01:49:08.698140672	{'contributors': None, 'coordinates': None, 'c...
1997-09-16 18:33:38.259525634	{'contributors': None, 'coordinates': None, 'c...
1997-10-23 12:52:52.832227328	{'contributors': None, 'coordinates': None, 'c...
1997-10-31 16:41:50.768087041	{'contributors': None, 'coordinates': None, 'c...
1997-11-01 17:19:37.348136964	{'contributors': None, 'coordinates': None, 'c...
1997-11-10 02:56:19.576475649	{'contributors': None, 'coordinates': None, 'c...
1998-01-20 15:53:12.912609280	{'contributors': None, 'coordinates': None, 'c...
1998-01-29 06:09:20.059072513	{'contributors': None, 'coordinates': None, 'c...

source \

1991-02-26 06:48:02.905632768	Tw...
1991-02-26 06:48:24.950915073	Tw...
1992-07-24 17:26:49.858043904	<a href="http://twitter.com/download/iphone" r...
1993-07-28 05:05:15.802583040	<a href="http://twitter.com/download/iphone" r...
1993-08-28 07:04:05.350707200	<a href="http://twitter.com/download/iphone" r...
1993-09-05 15:18:28.580548608	<a href="http://twitter.com/download/iphone" r...
1993-11-03 06:49:54.570878976	<a href="http://twitter.com/download/iphone" r...
1993-11-07 19:52:24.171524096	<a href="http://twitter.com/download/iphone" r...
1993-11-14 17:37:14.498793472	<a href="http://twitter.com/download/iphone" r...
1993-12-02 23:27:21.593970688	<a href="http://twitter.com/download/iphone" r...
1994-01-03 11:51:44.299253760	<a href="http://twitter.com/download/iphone" r...
1994-01-05 00:19:23.776290825	<a href="http://twitter.com/download/iphone" r...
1994-01-21 13:45:34.323924993	<a href="http://twitter.com/download/iphone" r...
1994-01-26 06:47:08.574212096	<a href="http://twitter.com/download/iphone" r...
1994-02-02 01:52:29.710192640	<a href="http://twitter.com/download/iphone" r...
1994-02-16 03:57:17.149827077	<a href="http://twitter.com/download/iphone" r...
1994-02-20 13:21:42.866649088	<a href="http://twitter.com/download/iphone" r...
1994-03-08 22:51:03.695355904	<a href="http://twitter.com/download/iphone" r...
1994-04-11 15:28:12.750233600	<a href="http://twitter.com/download/iphone" r...
1994-04-20 17:54:21.642756096	<a href="http://twitter.com/download/iphone" r...

1994-05-10 07:15:58.521745409 <a href="http://twitter.com/download/iphone" r...
 1994-05-14 10:02:47.477751808 <a href="http://twitter.com/download/iphone" r...
 1994-05-19 08:19:51.808995329 <a href="http://twitter.com/download/iphone" r...
 1994-05-28 02:56:07.776997377 <a href="http://twitter.com/download/iphone" r...
 1994-06-07 15:53:14.259247104 <a href="http://twitter.com/download/iphone" r...
 1994-06-09 14:10:53.431250945 <a href="http://twitter.com/download/iphone" r...
 1994-06-26 07:22:04.260794368 <a href="http://twitter.com/download/iphone" r...
 1994-07-04 15:46:27.167145985 <a href="http://twitter.com/download/iphone" r...
 1994-08-03 07:24:21.951791106 <a href="http://twitter.com/download/iphone" r...
 1994-08-07 08:58:26.839351296 <a href="http://twitter.com/download/iphone" r...
 ...
 1996-09-04 10:46:33.020538882 <a href="http://twitter.com/download/iphone" r...
 1996-10-12 05:32:39.547420673 <a href="http://twitter.com/download/iphone" r...
 1996-11-14 11:39:34.464610304 <a href="http://twitter.com/download/iphone" r...
 1996-11-14 13:41:05.427394560 <a href="http://twitter.com/download/iphone" r...
 1996-12-04 02:54:54.696017920 <a href="http://twitter.com/download/iphone" r...
 1996-12-29 12:09:45.021730816 <a href="http://twitter.com/download/iphone" r...
 1996-12-30 13:51:42.622658560 <a href="http://twitter.com/download/iphone" r...
 1997-01-10 22:46:45.516943360 <a href="http://twitter.com/download/iphone" r...
 1997-02-05 10:24:01.867124737 <a href="http://twitter.com/download/iphone" r...
 1997-02-06 16:08:43.840757760 <a href="http://twitter.com/download/iphone" r...
 1997-02-19 05:40:35.276025856 <a href="http://twitter.com/download/iphone" r...
 1997-02-22 09:16:33.587888130 <a href="http://twitter.com/download/iphone" r...
 1997-02-27 16:48:23.051644929 <a href="http://twitter.com/download/iphone" r...
 1997-03-20 12:19:50.427611136 <a href="http://twitter.com/download/iphone" r...
 1997-04-04 18:13:13.139703809 <a href="http://twitter.com/download/iphone" r...
 1997-04-13 09:33:55.999428608 <a href="http://twitter.com/download/iphone" r...
 1997-04-14 01:34:34.716409858 <a href="http://twitter.com/download/iphone" r...
 1997-05-12 21:16:22.782697472 <a href="http://twitter.com/download/iphone" r...
 1997-06-12 05:48:47.597207552 <a href="http://twitter.com/download/iphone" r...
 1997-06-23 13:30:53.475098625 <a href="http://twitter.com/download/iphone" r...
 1997-07-11 16:44:37.480148993 <a href="http://twitter.com/download/iphone" r...
 1997-08-09 22:36:19.821445120 <a href="http://twitter.com/download/iphone" r...
 1997-09-04 01:49:08.698140672 <a href="http://twitter.com/download/iphone" r...
 1997-09-16 18:33:38.259525634 <a href="http://twitter.com/download/iphone" r...
 1997-10-23 12:52:52.832227328 <a href="http://twitter.com/download/iphone" r...
 1997-10-31 16:41:50.768087041 <a href="http://twitter.com/download/iphone" r...
 1997-11-01 17:19:37.348136964 <a href="http://twitter.com/download/iphone" r...
 1997-11-10 02:56:19.576475649 <a href="http://twitter.com/download/iphone" r...
 1998-01-20 15:53:12.912609280 <a href="http://twitter.com/download/iphone" r...
 1998-01-29 06:09:20.059072513 <a href="http://twitter.com/download/iphone" r...

text \

1991-02-26 06:48:02.905632768 RT @dogratingrating: Unoriginal idea. Blatant ...
 1991-02-26 06:48:24.950915073 RT @dogratingrating: Exceptional talent. Origina...
 1992-07-24 17:26:49.858043904 RT @twitter: @dog_rates Awesome Tweet! 12/10. ...
 1993-07-28 05:05:15.802583040 RT @dog_rates: Extremely intelligent dog here...
 1993-08-28 07:04:05.350707200 RT @dog_rates: This is Shaggy. He knows exactl...

1993-09-05 15:18:28.580548608	RT @dog_rates: This pupper killed this great w...
1993-11-03 06:49:54.570878976	RT @dog_rates: Everyone needs to watch this. 1...
1993-11-07 19:52:24.171524096	RT @dog_rates: HEY PUP WHAT'S THE PART OF THE ...
1993-11-14 17:37:14.498793472	RT @dog_rates: This is Carly. She's actually 2...
1993-12-02 23:27:21.593970688	RT @dog_rates: This is Rubio. He has too much ...
1994-01-03 11:51:44.299253760	RT @jon_hill987: @dog_rates There is a cunning...
1994-01-05 00:19:23.776290825	RT @dog_rates: This is Chompsky. He lives up t...
1994-01-21 13:45:34.323924993	RT @dog_rates: AT DAWN...\nWE RIDE\n\n11/10 ht...
1994-01-26 06:47:08.574212096	RT @dog_rates: This... is a Tyrannosaurus rex...
1994-02-02 01:52:29.710192640	RT @hownottodraw: The story/person behind @dog...
1994-02-16 03:57:17.149827077	RT @dog_rates: Oh. My. God. 13/10 magical af h...
1994-02-20 13:21:42.866649088	RT @dog_rates: "Tristan do not speak to me wit...
1994-03-08 22:51:03.695355904	RT @dog_rates: Meet Eve. She's a raging alcoho...
1994-04-11 15:28:12.750233600	RT @dog_rates: This is Colby. He's currently r...
1994-04-20 17:54:21.642756096	RT @dog_rates: We only rate dogs... this is a ...
1994-05-10 07:15:58.521745409	RT @dog_rates: This is Nollie. She's waving at...
1994-05-14 10:02:47.477751808	RT @dog_rates: When it's Janet from accounting...
1994-05-19 08:19:51.808995329	RT @dog_rates: Ever seen a dog pet another dog...
1994-05-28 02:56:07.776997377	RT @dog_rates: This is just downright precious...
1994-06-07 15:53:14.259247104	RT @katieornah: @dog_rates learning a lot at c...
1994-06-09 14:10:53.431250945	RT @dog_rates: This is Frankie. He's wearing b...
1994-06-26 07:22:04.260794368	RT @dog_rates: This is Gromit. He's pupset bec...
1994-07-04 15:46:27.167145985	RT @dog_rates: Meet Fizz. She thinks love is a...
1994-08-03 07:24:21.951791106	RT @dog_rates: Like father (doggo), like son (...)
1994-08-07 08:58:26.839351296	RT @dog_rates: We only rate dogs. Pls stop sen...
...	...
1996-09-04 10:46:33.020538882	RT @dog_rates: This is Ken. His cheeks are mag...
1996-10-12 05:32:39.547420673	RT @dog_rates: This is Bungalo. She uses that ...
1996-11-14 11:39:34.464610304	RT @basic_vacek_: I love my new mug easy 13/10...
1996-11-14 13:41:05.427394560	RT @dog_rates: This is Charlie. He fell asleep...
1996-12-04 02:54:54.696017920	RT @dog_rates: This is Gidget. She's a spy pup...
1996-12-29 12:09:45.021730816	RT @eddie_coe98: Thanks @dog_rates completed m...
1996-12-30 13:51:42.622658560	RT @dog_rates: This is Astrid. She's a guide d...
1997-01-10 22:46:45.516943360	RT @dog_rates: I usually only share these on F...
1997-02-05 10:24:01.867124737	RT @frasercampbell_: oh my... what's that... b...
1997-02-06 16:08:43.840757760	RT @dog_rates: Meet George. He looks slightly ...
1997-02-19 05:40:35.276025856	RT @Jenna_Marbles: @dog_rates Thanks for ratin...
1997-02-22 09:16:33.587888130	RT @dog_rates: This is Luna. It's her first ti...
1997-02-27 16:48:23.051644929	RT @AaronChewning: First time wearing my @dog_...
1997-03-20 12:19:50.427611136	RT @dog_rates: Meet Winston. He knows he's a l...
1997-04-04 18:13:13.139703809	RT @dog_rates: Ohboyohboyohboyohboyohboyohboy...
1997-04-13 09:33:55.999428608	RT @tallylott: h*ckin adorable promposal. 13/1...
1997-04-14 01:34:34.716409858	RT @dog_rates: Meet Lorenzo. He's an avid nift...
1997-05-12 21:16:22.782697472	RT @dog_rates: Say hello to Quinn. She's quite...
1997-06-12 05:48:47.597207552	RT @dog_rates: Here's a pupper before and afte...
1997-06-23 13:30:53.475098625	RT @rachaeleasler: these @dog_rates hats are 1...
1997-07-11 16:44:37.480148993	RT @dog_rates: Say hello to Cooper. His expres...
1997-08-09 22:36:19.821445120	RT @dog_rates: This is Dawn. She's just checki...

1997-09-04 01:49:08.698140672 RT @dog_rates: This is Sierra. She's one preci...
 1997-09-16 18:33:38.259525634 RT @dog_rates: This is Coco. At first I though...
 1997-10-23 12:52:52.832227328 RT @rachel2195: @dog_rates the boyfriend and h...
 1997-10-31 16:41:50.768087041 RT @dog_rates: Meet Terrance. He's being yelle...
 1997-11-01 17:19:37.348136964 RT @dog_rates: Meet Shadow. In an attempt to r...
 1997-11-10 02:56:19.576475649 RT @dog_rates: This is Emmy. She was adopted t...
 1998-01-20 15:53:12.912609280 RT @dog_rates: This is Lilly. She just paralle...
 1998-01-29 06:09:20.059072513 RT @Athletics: 12/10 #BATP <https://t.co/WxwJmv...>

truncated \

1991-02-26 06:48:02.905632768	0
1991-02-26 06:48:24.950915073	0
1992-07-24 17:26:49.858043904	0
1993-07-28 05:05:15.802583040	0
1993-08-28 07:04:05.350707200	0
1993-09-05 15:18:28.580548608	0
1993-11-03 06:49:54.570878976	0
1993-11-07 19:52:24.171524096	0
1993-11-14 17:37:14.498793472	0
1993-12-02 23:27:21.593970688	0
1994-01-03 11:51:44.299253760	0
1994-01-05 00:19:23.776290825	0
1994-01-21 13:45:34.323924993	0
1994-01-26 06:47:08.574212096	0
1994-02-02 01:52:29.710192640	0
1994-02-16 03:57:17.149827077	0
1994-02-20 13:21:42.866649088	0
1994-03-08 22:51:03.695355904	0
1994-04-11 15:28:12.750233600	0
1994-04-20 17:54:21.642756096	0
1994-05-10 07:15:58.521745409	0
1994-05-14 10:02:47.477751808	0
1994-05-19 08:19:51.808995329	0
1994-05-28 02:56:07.776997377	0
1994-06-07 15:53:14.259247104	0
1994-06-09 14:10:53.431250945	0
1994-06-26 07:22:04.260794368	0
1994-07-04 15:46:27.167145985	0
1994-08-03 07:24:21.951791106	0
1994-08-07 08:58:26.839351296	0
...	...
1996-09-04 10:46:33.020538882	0
1996-10-12 05:32:39.547420673	0
1996-11-14 11:39:34.464610304	0
1996-11-14 13:41:05.427394560	0
1996-12-04 02:54:54.696017920	0
1996-12-29 12:09:45.021730816	0
1996-12-30 13:51:42.622658560	0

1997-01-10	22:46:45.516943360	0
1997-02-05	10:24:01.867124737	0
1997-02-06	16:08:43.840757760	0
1997-02-19	05:40:35.276025856	0
1997-02-22	09:16:33.587888130	0
1997-02-27	16:48:23.051644929	0
1997-03-20	12:19:50.427611136	0
1997-04-04	18:13:13.139703809	0
1997-04-13	09:33:55.999428608	0
1997-04-14	01:34:34.716409858	0
1997-05-12	21:16:22.782697472	0
1997-06-12	05:48:47.597207552	0
1997-06-23	13:30:53.475098625	0
1997-07-11	16:44:37.480148993	0
1997-08-09	22:36:19.821445120	0
1997-09-04	01:49:08.698140672	0
1997-09-16	18:33:38.259525634	0
1997-10-23	12:52:52.832227328	0
1997-10-31	16:41:50.768087041	0
1997-11-01	17:19:37.348136964	0
1997-11-10	02:56:19.576475649	0
1998-01-20	15:53:12.912609280	0
1998-01-29	06:09:20.059072513	0

			user
1991-02-26	06:48:02.905632768	{'contributors_enabled': False, 'created_at': ...	
1991-02-26	06:48:24.950915073	{'contributors_enabled': False, 'created_at': ...	
1992-07-24	17:26:49.858043904	{'contributors_enabled': False, 'created_at': ...	
1993-07-28	05:05:15.802583040	{'contributors_enabled': False, 'created_at': ...	
1993-08-28	07:04:05.350707200	{'contributors_enabled': False, 'created_at': ...	
1993-09-05	15:18:28.580548608	{'contributors_enabled': False, 'created_at': ...	
1993-11-03	06:49:54.570878976	{'contributors_enabled': False, 'created_at': ...	
1993-11-07	19:52:24.171524096	{'contributors_enabled': False, 'created_at': ...	
1993-11-14	17:37:14.498793472	{'contributors_enabled': False, 'created_at': ...	
1993-12-02	23:27:21.593970688	{'contributors_enabled': False, 'created_at': ...	
1994-01-03	11:51:44.299253760	{'contributors_enabled': False, 'created_at': ...	
1994-01-05	00:19:23.776290825	{'contributors_enabled': False, 'created_at': ...	
1994-01-21	13:45:34.323924993	{'contributors_enabled': False, 'created_at': ...	
1994-01-26	06:47:08.574212096	{'contributors_enabled': False, 'created_at': ...	
1994-02-02	01:52:29.710192640	{'contributors_enabled': False, 'created_at': ...	
1994-02-16	03:57:17.149827077	{'contributors_enabled': False, 'created_at': ...	
1994-02-20	13:21:42.866649088	{'contributors_enabled': False, 'created_at': ...	
1994-03-08	22:51:03.695355904	{'contributors_enabled': False, 'created_at': ...	
1994-04-11	15:28:12.750233600	{'contributors_enabled': False, 'created_at': ...	
1994-04-20	17:54:21.642756096	{'contributors_enabled': False, 'created_at': ...	
1994-05-10	07:15:58.521745409	{'contributors_enabled': False, 'created_at': ...	
1994-05-14	10:02:47.477751808	{'contributors_enabled': False, 'created_at': ...	
1994-05-19	08:19:51.808995329	{'contributors_enabled': False, 'created_at': ...	

```

1994-05-28 02:56:07.776997377 {'contributors_enabled': False, 'created_at': ...
1994-06-07 15:53:14.259247104 {'contributors_enabled': False, 'created_at': ...
1994-06-09 14:10:53.431250945 {'contributors_enabled': False, 'created_at': ...
1994-06-26 07:22:04.260794368 {'contributors_enabled': False, 'created_at': ...
1994-07-04 15:46:27.167145985 {'contributors_enabled': False, 'created_at': ...
1994-08-03 07:24:21.951791106 {'contributors_enabled': False, 'created_at': ...
1994-08-07 08:58:26.839351296 {'contributors_enabled': False, 'created_at': ...
...
1996-09-04 10:46:33.020538882 {'contributors_enabled': False, 'created_at': ...
1996-10-12 05:32:39.547420673 {'contributors_enabled': False, 'created_at': ...
1996-11-14 11:39:34.464610304 {'contributors_enabled': False, 'created_at': ...
1996-11-14 13:41:05.427394560 {'contributors_enabled': False, 'created_at': ...
1996-12-04 02:54:54.696017920 {'contributors_enabled': False, 'created_at': ...
1996-12-29 12:09:45.021730816 {'contributors_enabled': False, 'created_at': ...
1996-12-30 13:51:42.622658560 {'contributors_enabled': False, 'created_at': ...
1997-01-10 22:46:45.516943360 {'contributors_enabled': False, 'created_at': ...
1997-02-05 10:24:01.867124737 {'contributors_enabled': False, 'created_at': ...
1997-02-06 16:08:43.840757760 {'contributors_enabled': False, 'created_at': ...
1997-02-19 05:40:35.276025856 {'contributors_enabled': False, 'created_at': ...
1997-02-22 09:16:33.587888130 {'contributors_enabled': False, 'created_at': ...
1997-02-27 16:48:23.051644929 {'contributors_enabled': False, 'created_at': ...
1997-03-20 12:19:50.427611136 {'contributors_enabled': False, 'created_at': ...
1997-04-04 18:13:13.139703809 {'contributors_enabled': False, 'created_at': ...
1997-04-13 09:33:55.999428608 {'contributors_enabled': False, 'created_at': ...
1997-04-14 01:34:34.716409858 {'contributors_enabled': False, 'created_at': ...
1997-05-12 21:16:22.782697472 {'contributors_enabled': False, 'created_at': ...
1997-06-12 05:48:47.597207552 {'contributors_enabled': False, 'created_at': ...
1997-06-23 13:30:53.475098625 {'contributors_enabled': False, 'created_at': ...
1997-07-11 16:44:37.480148993 {'contributors_enabled': False, 'created_at': ...
1997-08-09 22:36:19.821445120 {'contributors_enabled': False, 'created_at': ...
1997-09-04 01:49:08.698140672 {'contributors_enabled': False, 'created_at': ...
1997-09-16 18:33:38.259525634 {'contributors_enabled': False, 'created_at': ...
1997-10-23 12:52:52.832227328 {'contributors_enabled': False, 'created_at': ...
1997-10-31 16:41:50.768087041 {'contributors_enabled': False, 'created_at': ...
1997-11-01 17:19:37.348136964 {'contributors_enabled': False, 'created_at': ...
1997-11-10 02:56:19.576475649 {'contributors_enabled': False, 'created_at': ...
1998-01-20 15:53:12.912609280 {'contributors_enabled': False, 'created_at': ...
1998-01-29 06:09:20.059072513 {'contributors_enabled': False, 'created_at': ...

```

```
[168 rows x 30 columns]
```

```
In [50]: twitter_api_clean.drop(twitter_api_clean[twitter_api_clean['retweeted_status'].notnull()
```

Test

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

```
<class 'pandas.core.frame.DataFrame'>
```

```
DatetimeIndex: 2174 entries, 1991-02-08 13:48:08.022790149 to 1998-04-12 22:37:23.555336193
```

```

Data columns (total 30 columns):
contributors      0 non-null float64
coordinates       0 non-null float64
created_at        2174 non-null datetime64[ns]
entities          2174 non-null object
extended_entities 1748 non-null object
favorite_count    2174 non-null int64
favorited         2174 non-null int64
geo              0 non-null float64
tweet_id          2174 non-null int64
id_str            2174 non-null int64
in_reply_to_screen_name 77 non-null object
in_reply_to_status_id 77 non-null float64
in_reply_to_status_id_str 77 non-null float64
in_reply_to_user_id 77 non-null float64
in_reply_to_user_id_str 77 non-null float64
is_quote_status   2174 non-null int64
lang              2174 non-null object
place            1 non-null object
possibly_sensitive 2116 non-null float64
possibly_sensitive_appealable 2116 non-null float64
quoted_status     24 non-null object
quoted_status_id  25 non-null float64
quoted_status_id_str 25 non-null float64
retweet_count     2174 non-null int64
retweeted         2174 non-null int64
retweeted_status   0 non-null object
source            2174 non-null object
text              2174 non-null object
truncated         2174 non-null int64
user              2174 non-null object
dtypes: datetime64[ns](1), float64(11), int64(8), object(10)
memory usage: 526.5+ KB

```

2.2.5 Tidiness

Define Remove unwanted columns from all datasets

Code To do that, I need to pull all `datasets.info()` and verify what column to keep and what column to remove before merging all three datasets together

```

In [52]: twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 16 columns):
tweet_id          2175 non-null int64

```

```

in_reply_to_status_id    78 non-null float64
in_reply_to_user_id      78 non-null float64
source                   2175 non-null object
text                     2175 non-null object
expanded_urls            2117 non-null object
rating_numerator          2175 non-null float64
rating_denominator        2175 non-null float64
name                     1390 non-null object
doggo                    2175 non-null object
floofer                  2175 non-null object
pupper                   2175 non-null object
puppo                    2175 non-null object
year                     2175 non-null int64
month                    2175 non-null int64
day                      2175 non-null int64
dtypes: float64(4), int64(4), object(8)
memory usage: 288.9+ KB

```

```

In [53]: # Delete unwanted columns from twitter_archive_clean
         twitter_archive_clean = twitter_archive_clean.drop(['source',
                                                             'in_reply_to_status_id',
                                                             'in_reply_to_user_id',
                                                             'expanded_urls'], 1)

```

Test

```

In [54]: twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id          2175 non-null int64
text              2175 non-null object
rating_numerator  2175 non-null float64
rating_denominator 2175 non-null float64
name              1390 non-null object
doggo             2175 non-null object
floofer           2175 non-null object
pupper            2175 non-null object
puppo             2175 non-null object
year              2175 non-null int64
month             2175 non-null int64
day               2175 non-null int64
dtypes: float64(2), int64(4), object(6)
memory usage: 220.9+ KB

```

```

In [55]: # Remove unwanted columns in the image dataset
         image_clean.info()

```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1691 entries, 0 to 2073
Data columns (total 14 columns):
tweet_id          1691 non-null int64
jpg_url           1691 non-null object
img_num           1691 non-null int64
p1                1691 non-null object
p1_conf           1691 non-null float64
p1_dog            1691 non-null bool
p2                1691 non-null object
p2_conf           1691 non-null float64
p2_dog            1691 non-null bool
p3                1691 non-null object
p3_conf           1691 non-null float64
p3_dog            1691 non-null bool
dog_type          1691 non-null object
confidence_list    1691 non-null object
dtypes: bool(3), float64(3), int64(2), object(6)
memory usage: 163.5+ KB

```

Code

```

In [56]: #Delete columns
         image_clean = image_clean.drop(['img_num', 'p1',
                                         'p1_conf', 'p1_dog',
                                         'p2', 'p2_conf',
                                         'p2_dog', 'p3',
                                         'p3_conf',
                                         'p3_dog'], 1)

```

Test

```

In [57]: image_clean.info()

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

```

```

In [58]: twitter_api_clean.info()

```



```

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 2174 entries, 1991-02-08 13:48:08.022790149 to 1998-04-12 22:37:23.555336193
Data columns (total 30 columns):
contributors          0 non-null float64
coordinates           0 non-null float64
created_at            2174 non-null datetime64[ns]
entities              2174 non-null object
extended_entities      1748 non-null object
favorite_count        2174 non-null int64
favorited             2174 non-null int64
geo                   0 non-null float64
tweet_id              2174 non-null int64
id_str                2174 non-null int64
in_reply_to_screen_name 77 non-null object
in_reply_to_status_id  77 non-null float64
in_reply_to_status_id_str 77 non-null float64
in_reply_to_user_id    77 non-null float64
in_reply_to_user_id_str 77 non-null float64
is_quote_status        2174 non-null int64
lang                  2174 non-null object
place                 1 non-null object
possibly_sensitive     2116 non-null float64
possibly_sensitive_appealable 2116 non-null float64
quoted_status          24 non-null object
quoted_status_id       25 non-null float64
quoted_status_id_str   25 non-null float64
retweet_count          2174 non-null int64
retweeted              2174 non-null int64
retweeted_status       0 non-null object
source                2174 non-null object
text                  2174 non-null object
truncated             2174 non-null int64
user                  2174 non-null object
dtypes: datetime64[ns](1), float64(11), int64(8), object(10)
memory usage: 526.5+ KB

```

Code

```

In [59]: # Keep a few columns from twitter_api_clean and delete the rest
         twitter_api_clean.drop(['contributors',
                                'coordinates',
                                'created_at',
                                'entities',
                                'extended_entities',
                                'favorited',
                                'geo',
                                'id_str',

```

```

'in_reply_to_screen_name',
'in_reply_to_status_id',
'in_reply_to_status_id_str',
'in_reply_to_user_id',
'in_reply_to_user_id_str',
'is_quote_status',
'lang',
'place',
'possibly_sensitive',
'possibly_sensitive_appealable',
'quoted_status',
'quoted_status_id',
'quoted_status_id_str',
'retweeted',
'retweeted_status',
'source',
'truncated',
'user'], axis=1,inplace=True)

```

Test

In [60]: `twitter_api_clean.head()`

```

Out[60]:

```

	timestamp	favorite_count	tweet_id	retweet_count	text
1991-02-08	13:48:08.022790149	2544	666020888022790144	508	Here we have a Japanese Irish Setter. Lost eye...
1991-02-08	16:08:05.002620928	129	666029285002620928	47	This is a western brown Mitsubishi terrier. Up...
1991-02-08	17:16:52.701032449	125	666033412701032449	44	Here is a very happy pup. Big fan of well-main...
1991-02-08	20:17:06.329800704	296	666044226329800704	139	This is a purebred Piers Morgan. Loves to Netf...
1991-02-08	21:40:48.165822465	107	666049248165822464	40	Here we have a 1949 1st generation vulpix. Enj...

Define Twitter_archive - Combine each dog stage column into a single column named "stage"

code

```

In [61]: # Combine dogs & stage columns into one column 'stage' list
# append each stage list into a bigger list
stage = pd.melt(twitter_archive_clean, id_vars=['tweet_id'], value_vars=['doggo', 'floof', 'pupper', 'puppo'])
stage_list = []
for ids in stage.tweet_id.unique():
    stage_list.append(stage.query('tweet_id == @ids').value.tolist())

In [62]: # remove repeat value in each stage list
stage_list2 = []
for e in stage_list:
    stage_list2.append(list(set(e)))

In [63]: # remove the 'None' in stage list which contains any stage value, such as 'doggo', 'floof', 'pupper', 'puppo'
stage_list3 = []
for e in stage_list2:
    if len(e) > 1:
        e.remove('None')
        stage_list3.append(e)
    else:
        stage_list3.append(e)

In [64]: # remove '['' which wrapped outside the stage value
stage_list4 = []
for e in stage_list3:
    if len(e) > 1:
        stage_list4.append(",".join(e))
    else:
        stage_list4.append(e[0])

In [65]: twitter_archive_clean['stage'] = stage_list4

```

Test

```
In [66]: twitter_archive_clean.head()
```

```

Out[66]:
   tweet_id  text \
0  892420643555336193  This is Phineas. He's a mystical boy. Only eve...
1  892177421306343426  This is Tilly. She's just checking pup on you...
2  891815181378084864  This is Archie. He is a rare Norwegian Pouncin...
3  891689557279858688  This is Darla. She commenced a snooze mid meal...
4  891327558926688256  This is Franklin. He would like you to stop ca...

   rating_numerator  rating_denominator  name  doggo  floofer  pupper  puppo \
0                13.0                10.0  Phineas  None    None    None    None
1                13.0                10.0   Tilly  None    None    None    None
2                12.0                10.0  Archie  None    None    None    None
3                13.0                10.0   Darla  None    None    None    None
4                12.0                10.0 Franklin  None    None    None    None

```

	year	month	day	stage
0	2017	8	1	None
1	2017	8	1	None
2	2017	7	31	None
3	2017	7	30	None
4	2017	7	29	None

```
In [67]: twitter_archive_clean['stage'].value_counts()
```

```
Out[67]: None          1831
pupper          224
doggo           75
puppo           24
pupper,doggo    10
floofer         9
doggo,floofer   1
puppo,doggo     1
Name: stage, dtype: int64
```

Define There are some dogs or tweet_id that has two stages. We need to find them

Code

```
In [68]: # One floofer,doggo
twitter_archive_clean.query('stage == "floofer,doggo"')
```

```
Out[68]: Empty DataFrame
Columns: [tweet_id, text, rating_numerator, rating_denominator, name, doggo, floofer, pupper, puppo]
Index: []
```

```
In [69]: # One puppo,doggo
twitter_archive_clean.query('stage == "puppo,doggo"')
```

```
Out[69]:
```

	tweet_id	text	rating_numerator	rating_denominator	name	doggo	floofer	pupper	puppo
191	855851453814013952	Here's a puppo participating in the #ScienceMa...	13.0	10.0	NaN	doggo	None	None	puppo

	year	month	day	stage
191	2017	4	22	puppo,doggo

```
In [70]: # Ten doggo,pupper
twitter_archive_clean.query('stage == "doggo,pupper").head()
```

```
Out[70]: Empty DataFrame
Columns: [tweet_id, text, rating_numerator, rating_denominator, name, doggo, floofer, pupper, puppo]
Index: []
```

Define All datasets Convert tweet_id column from an integer to a string value

Code

```
In [71]: twitter_archive_clean['tweet_id'] = twitter_archive_clean['tweet_id'].astype('str')
         image_clean['tweet_id'] = image_clean['tweet_id'].astype('str')
         twitter_api_clean['tweet_id'] = twitter_api_clean['tweet_id'].astype('str')
```

Test

```
In [72]: type(twitter_archive_clean['tweet_id'].iloc[0]), type(image_clean['tweet_id'].iloc[0]),
         type(twitter_api_clean['tweet_id'].iloc[0])
```

```
Out[72]: (str, str, str)
```

Define Twitter_archive - Convert Rating columns to float values

Code

```
In [73]: # This step was done in the quality stage but I wanted to make sure it is float now
twitter_archive_clean['rating_numerator'] = twitter_archive_clean['rating_numerator'].astype(float)
twitter_archive_clean['rating_denominator'] = twitter_archive_clean['rating_denominator'].astype(float)
```

Test

```
In [74]: twitter_archive_clean['rating_numerator'].dtype, twitter_archive_clean['rating_denominator'].dtype
```

```
Out[74]: (dtype('float64'), dtype('float64'))
```

Define Consolidate the 3 datasets into 1

Code

[illegible]

Test

```
In [76]: twitter_archive_master.head()
```

```
Out[76]:
```

	tweet_id	text_x \
0	891815181378084864	This is Archie. He is a rare Norwegian Pouncin...
1	891689557279858688	This is Darla. She commenced a snooze mid meal...
2	891327558926688256	This is Franklin. He would like you to stop ca...
3	891087950875897856	Here we have a majestic great white breaching ...
4	890729181411237888	When you watch your owner call another dog a g...

	rating_numerator	rating_denominator	name	doggo	floofer	pupper	puppo \
0	12.0	10.0	Archie	None	None	None	None

1		13.0		10.0	Darla	None	None	None	None
2		12.0		10.0	Franklin	None	None	None	None
3		13.0		10.0	NaN	None	None	None	None
4		13.0		10.0	NaN	None	None	None	None

	year	month	day	stage	jpg_url	\
0	2017	7	31	None	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg	
1	2017	7	30	None	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg	
2	2017	7	29	None	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg	
3	2017	7	29	None	https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg	
4	2017	7	28	None	https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg	

	dog_type	confidence_list	favorite_count	retweet_count	\
0	chihuahua	0.716012	24670	4090	
1	labrador retriever	0.168086	41532	8505	
2	basset	0.555712	39725	9213	
3	chesapeake bay retriever	0.425595	19938	3065	
4	pomeranian	0.566142	64480	18570	

	text_y
0	This is Archie. He is a rare Norwegian Pouncin...
1	This is Darla. She commenced a snooze mid meal...
2	This is Franklin. He would like you to stop ca...
3	Here we have a majestic great white breaching ...
4	When you watch your owner call another dog a g...

```
In [77]: # I forgot to drop 'doggo', 'puppo', 'pupper', 'floofer' columns. I can implement it on
twitter_archive_master.drop(['doggo', 'puppo', 'pupper', 'floofer'], axis=1,inplace=True)
```

```
In [78]: # Order the table by tweet_id then reset index accordingly
twitter_archive_master.sort_values('tweet_id', inplace=True)
twitter_archive_master= twitter_archive_master.set_index('tweet_id')
```

```
In [79]: # add rating column
twitter_archive_master['rating'] = twitter_archive_master.rating_numerator/twitter_arch
```

```
In [80]: twitter_archive_master.head()
```

```
Out[80]:
```

	tweet_id	text_x	\
	666029285002620928	This is a western brown Mitsubishi terrier. Up...	
	666044226329800704	This is a purebred Piers Morgan. Loves to Netf...	
	666057090499244032	My oh my. This is a rare blond Canadian terrie...	
	666058600524156928	Here is the Rand Paul of retrievers folks! He'...	
	666071193221509120	Here we have a northern speckled Rhododendron...	

	rating_numerator	rating_denominator	name	year	month	\
tweet_id						
666029285002620928	7.0	10.0	NaN	2015	11	

666044226329800704	6.0	10.0	NaN	2015	11
666057090499244032	9.0	10.0	NaN	2015	11
666058600524156928	8.0	10.0	NaN	2015	11
666071193221509120	9.0	10.0	NaN	2015	11

tweet_id	day	stage	\
666029285002620928	15	None	
666044226329800704	16	None	
666057090499244032	16	None	
666058600524156928	16	None	
666071193221509120	16	None	

tweet_id	jpg_url	\
666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	
666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg	
666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg	
666058600524156928	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg	
666071193221509120	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg	

tweet_id	dog_type	confidence_list	favorite_count	\
666029285002620928	redbone	0.506826	129	
666044226329800704	rhodesian ridgeback	0.408143	296	
666057090499244032	golden retriever	0.00795896	294	
666058600524156928	miniature poodle	0.201493	111	
666071193221509120	gordon setter	0.503672	145	

tweet_id	retweet_count	\
666029285002620928	47	
666044226329800704	139	
666057090499244032	141	
666058600524156928	57	
666071193221509120	59	

tweet_id	text_y	rating
666029285002620928	This is a western brown Mitsubishi terrier. Up...	0.7
666044226329800704	This is a purebred Piers Morgan. Loves to Netf...	0.6
666057090499244032	My oh my. This is a rare blond Canadian terrie...	0.9
666058600524156928	Here is the Rand Paul of retrievers folks! He'...	0.8
666071193221509120	Here we have a northern speckled Rhododendron...	0.9

In [81]: twitter_archive_master.info()

<class 'pandas.core.frame.DataFrame'>

Index: 1056 entries, 666029285002620928 to 891815181378084864

```
Data columns (total 15 columns):
text_x                1056 non-null object
rating_numerator      1056 non-null float64
rating_denominator    1056 non-null float64
name                  730 non-null object
year                  1056 non-null int64
month                 1056 non-null int64
day                   1056 non-null int64
stage                 1056 non-null object
jpg_url               1056 non-null object
dog_type              1056 non-null object
confidence_list       1056 non-null object
favorite_count        1056 non-null int64
retweet_count         1056 non-null int64
text_y                1056 non-null object
rating                1056 non-null float64
dtypes: float64(3), int64(5), object(7)
memory usage: 132.0+ KB
```

```
In [82]: twitter_archive_master.describe()
```

```
Out[82]:
```

	rating_numerator	rating_denominator	year	month \
count	1056.000000	1056.000000	1056.000000	1056.000000
mean	11.379735	10.519886	2015.846591	7.145833
std	8.563492	7.219483	0.690592	4.077628
min	1.000000	2.000000	2015.000000	1.000000
25%	10.000000	10.000000	2015.000000	3.000000
50%	11.000000	10.000000	2016.000000	7.000000
75%	12.000000	10.000000	2016.000000	11.000000
max	165.000000	150.000000	2017.000000	12.000000

	day	favorite_count	retweet_count	rating
count	1056.000000	1056.000000	1056.000000	1056.000000
mean	16.110795	7911.772727	2255.431818	1.081182
std	8.759344	10653.075588	3281.921062	0.282588
min	1.000000	80.000000	12.000000	0.200000
25%	8.000000	1755.000000	560.750000	1.000000
50%	17.000000	3793.000000	1240.000000	1.100000
75%	24.000000	9637.500000	2715.000000	1.200000
max	31.000000	91088.000000	33954.000000	7.500000

```
In [83]: twitter_archive_master.shape
```

```
Out[83]: (1056, 15)
```

2.2.6 Late Quality issue correction

Define One thing more I found out I needed to do, make sure that "None" in dogs stages is programmable. Note this is a qauality issue

Code

```
In [84]: twitter_archive_master.stage = twitter_archive_master.stage.replace('None', np.nan)
```

Test

```
In [85]: twitter_archive_master.head()
```

```
Out[85]:
```

tweet_id	text_x \
666029285002620928	This is a western brown Mitsubishi terrier. Up...
666044226329800704	This is a purebred Piers Morgan. Loves to Netf...
666057090499244032	My oh my. This is a rare blond Canadian terrie...
666058600524156928	Here is the Rand Paul of retrievers folks! He'...
666071193221509120	Here we have a northern speckled Rhododendron...

tweet_id	rating_numerator	rating_denominator	name	year	month	\
666029285002620928	7.0	10.0	NaN	2015	11	
666044226329800704	6.0	10.0	NaN	2015	11	
666057090499244032	9.0	10.0	NaN	2015	11	
666058600524156928	8.0	10.0	NaN	2015	11	
666071193221509120	9.0	10.0	NaN	2015	11	

tweet_id	day	stage	\
666029285002620928	15	NaN	
666044226329800704	16	NaN	
666057090499244032	16	NaN	
666058600524156928	16	NaN	
666071193221509120	16	NaN	

tweet_id	jpg_url \
666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
666058600524156928	https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
666071193221509120	https://pbs.twimg.com/media/CT5cN_3WEAA10oZ.jpg

tweet_id	dog_type	confidence_list	favorite_count	\
666029285002620928	redbone	0.506826	129	
666044226329800704	rhodesian ridgeback	0.408143	296	
666057090499244032	golden retriever	0.00795896	294	
666058600524156928	miniature poodle	0.201493	111	
666071193221509120	gordon setter	0.503672	145	

tweet_id	retweet_count	\
----------	---------------	---

tweet_id	
666029285002620928	47
666044226329800704	139
666057090499244032	141
666058600524156928	57
666071193221509120	59

tweet_id	text_y	rating
666029285002620928	This is a western brown Mitsubishi terrier. Up...	0.7
666044226329800704	This is a purebred Piers Morgan. Loves to Netf...	0.6
666057090499244032	My oh my. This is a rare blond Canadian terrie...	0.9
666058600524156928	Here is the Rand Paul of retrievers folks! He'...	0.8
666071193221509120	Here we have a northern speckled Rhododendron...	0.9

```
In [89]: twitter_archive_master.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1056 entries, 666029285002620928 to 891815181378084864
Data columns (total 15 columns):
text_x                1056 non-null object
rating_numerator      1056 non-null float64
rating_denominator    1056 non-null float64
name                  730 non-null object
year                  1056 non-null int64
month                 1056 non-null int64
day                   1056 non-null int64
stage                 164 non-null object
jpg_url               1056 non-null object
dog_type              1056 non-null object
confidence_list       1056 non-null object
favorite_count        1056 non-null int64
retweet_count         1056 non-null int64
text_y                1056 non-null object
rating                1056 non-null float64
dtypes: float64(3), int64(5), object(7)
memory usage: 132.0+ KB
```

2.3 Second Part - Storing, Analyzing, and Visualizing Wrangled Data

2.3.1 Storing Wrangled Data

```
In [151]: twitter_archive_master.to_csv('twitter_archive_master.csv', index=False, encoding = 'u
```

2.3.2 Analyzing Wrangled Data

First Analysis and Visualizing - Dogs Counts

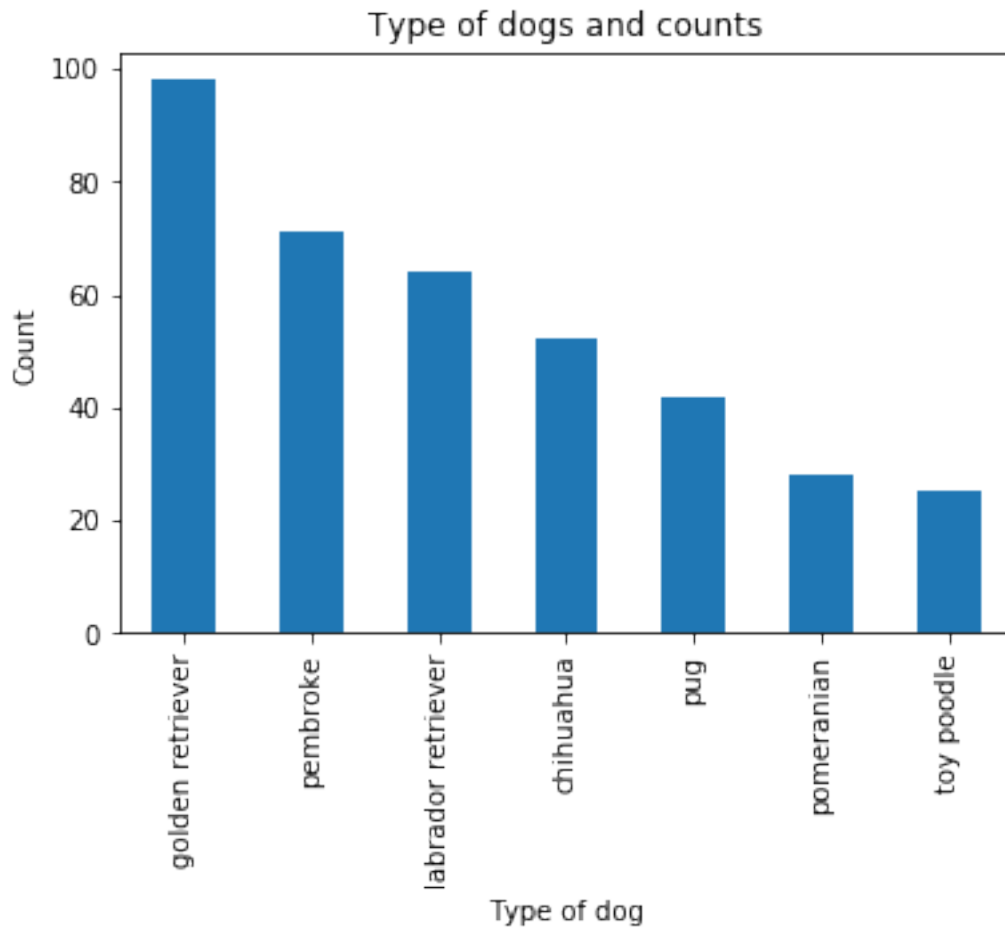
```
In [91]: # High counts of rated dogs are
        twitter_archive_master['dog_type'].value_counts().head(10)
```

```
Out[91]: golden retriever          98
         pembroke                  71
         labrador retriever        64
         chihuahua                 52
         pug                       42
         pomeranian                28
         toy poodle                25
         chow                      24
         chesapeake bay retriever  22
         cocker spaniel            21
         Name: dog_type, dtype: int64
```

```
In [92]: # Low counts of rated dogs are
        twitter_archive_master['dog_type'].value_counts().tail(15)
```

```
Out[92]: komondor                  2
         leonberg                  2
         brabancon griffon         2
         wire-haired fox terrier   2
         afghan hound              1
         standard schnauzer        1
         black-and-tan coonhound   1
         groenendael               1
         clumber                   1
         welsh springer spaniel    1
         bluetick                  1
         toy terrier               1
         irish wolfhound           1
         silky terrier             1
         japanese spaniel          1
         Name: dog_type, dtype: int64
```

```
In [93]: dog_type = twitter_archive_master.groupby('dog_type').filter(lambda x: len(x) >= 25)
        dog_type['dog_type'].value_counts().plot(kind = 'bar')
        plt.title('Type of dogs and counts')
        plt.xlabel('Type of dog')
        plt.ylabel('Count');
```



It seems like the Golden Retriever has the highest rating counts then pembroke the second. So it would make sense later when analyze the highly rated dogs to take the first 10 to be more accurate

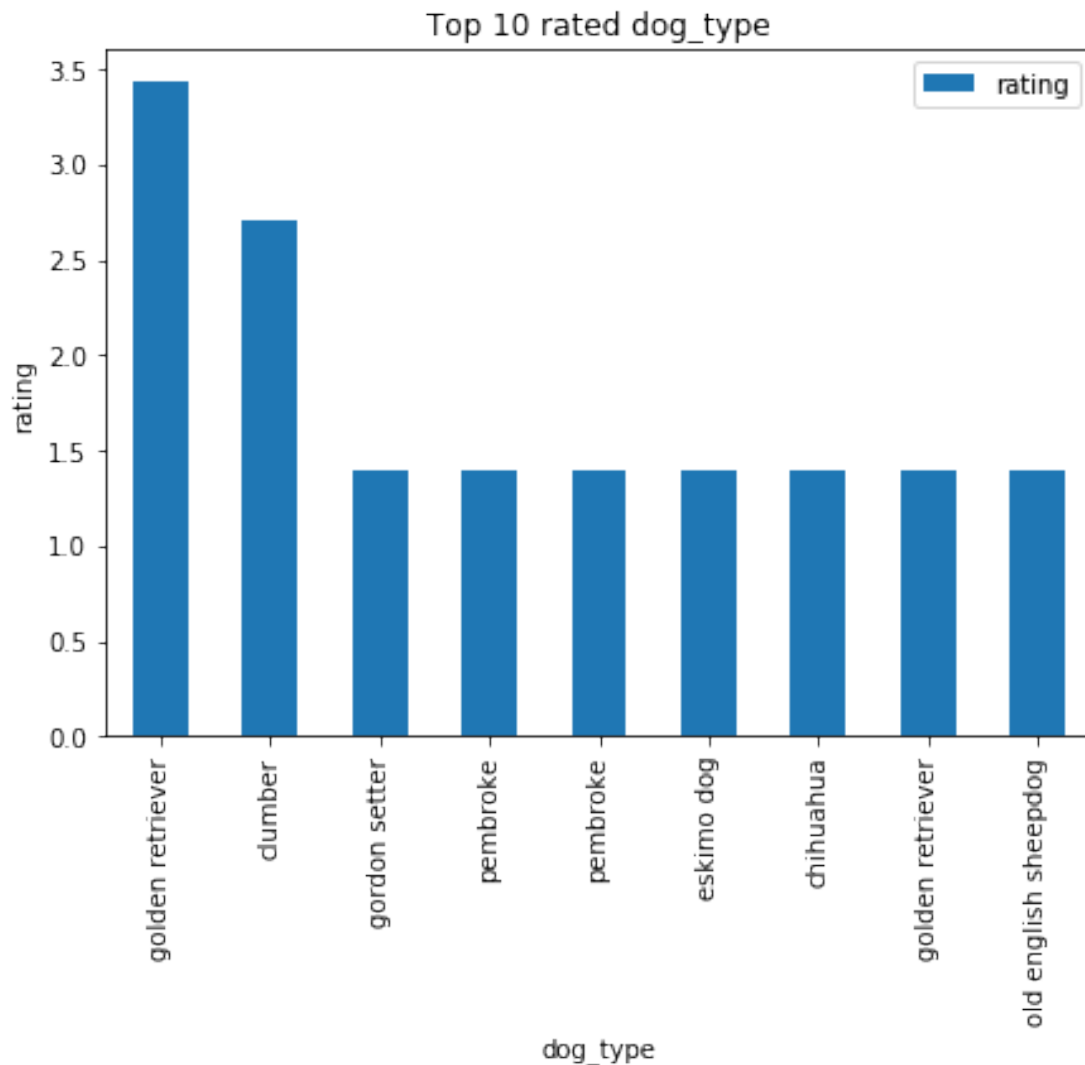
Dogs Rating

```
In [153]: top10_rating = twitter_archive_master.query('dog_type == dog_type').sort_values\
          ('rating', ascending = False)[1:10]
          top10_rating = top10_rating['dog_type']
          top10_rating
```

```
Out[153]: tweet_id
          810984652412424192    golden retriever
          778027034220126208          clumber
          820314633777061888    gordon setter
          832273440279240704    pembroke
          884441805382717440    pembroke
          774314403806253056    eskimo dog
          856282028240666624    chihuahua
          807621403335917568    golden retriever
```

```
856526610513747968    old english sheepdog
Name: dog_type, dtype: object
```

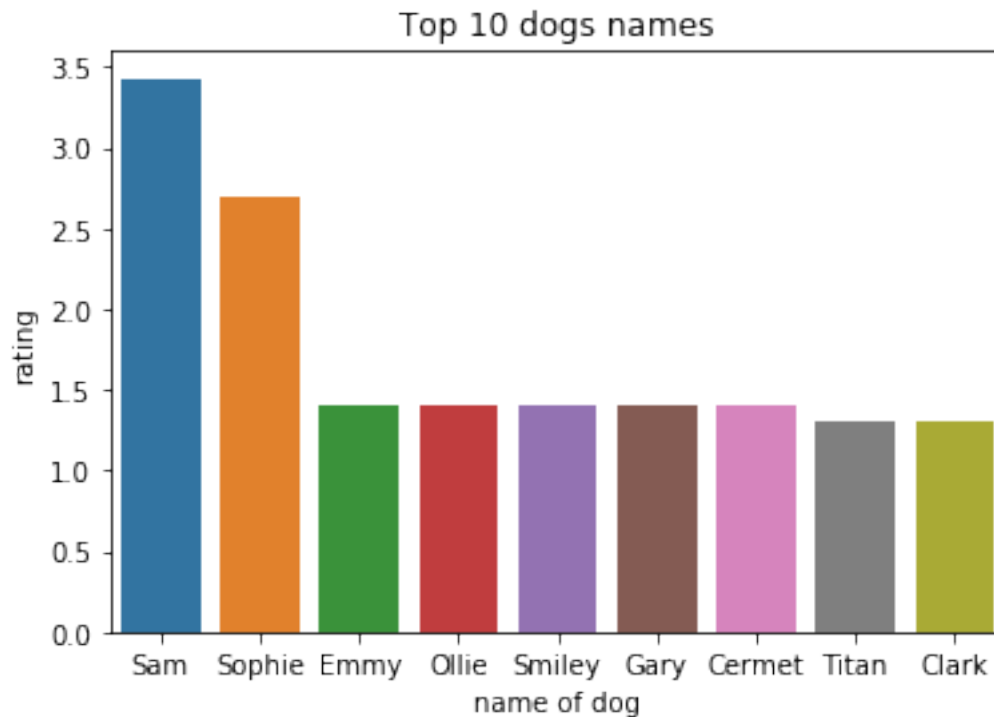
```
In [94]: # Top 10 rated dog_type
top10_rating.plot(x='dog_type', y='rating', kind='bar', figsize=(7,5))
plt.title('Top 10 rated dog_type')
plt.xlabel('dog_type')
plt.ylabel('rating');
```



Here I wanted to verify whether the dogs that were rated in terms of counts, are getting high ratings too. Our conclusion was Golden Retriever, Dumbles, and Golden Setter are three highest rate dogs while Golden Retriever, pembroke, and labrador retriever had the three highest counts

Second Analysis and Visualizing - Top 10 Rating Dogs' Names Out of curiosity I wanted to know the names of these dogs

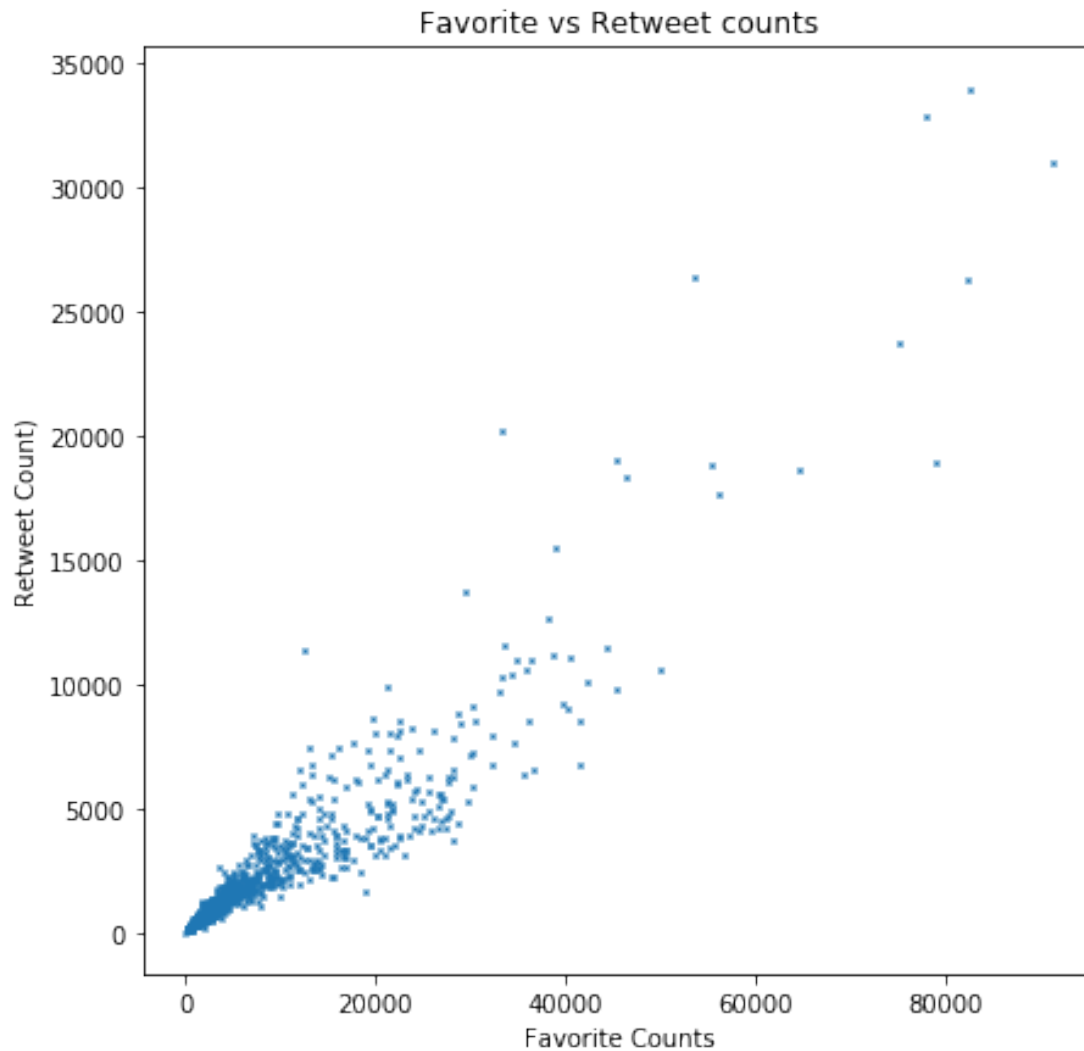
```
In [112]: # Top 10 dog names
top10_rating = twitter_archive_master.query('name == name').sort_values('rating', asce
sns.barplot(x="name", y="rating", data=top10_rating)
plt.title('Top 10 dogs names')
plt.xlabel('name of dog')
plt.ylabel('rating');
```



Sam, Sophie and Emmy are the names of these highly rated dogs

Third Analysis and Visualizing - Favorite vs Retweet Counts

```
In [113]: plt.figure(figsize = (7,7))
plt.plot(twitter_archive_master.favorite_count, twitter_archive_master.retweet_count,\
         marker='x', linestyle='', ms=2 );
plt.title('Favorite vs Retweet counts')
plt.xlabel('Favorite Counts')
plt.ylabel('Retweet Count')
plt.show()
```



Here I wanted to see the correlation between the counts of favorite and retweets. Seems like we have positive to strong positive correlation

Fourth Analysis and Visualizing - Averages of Dogs Stages

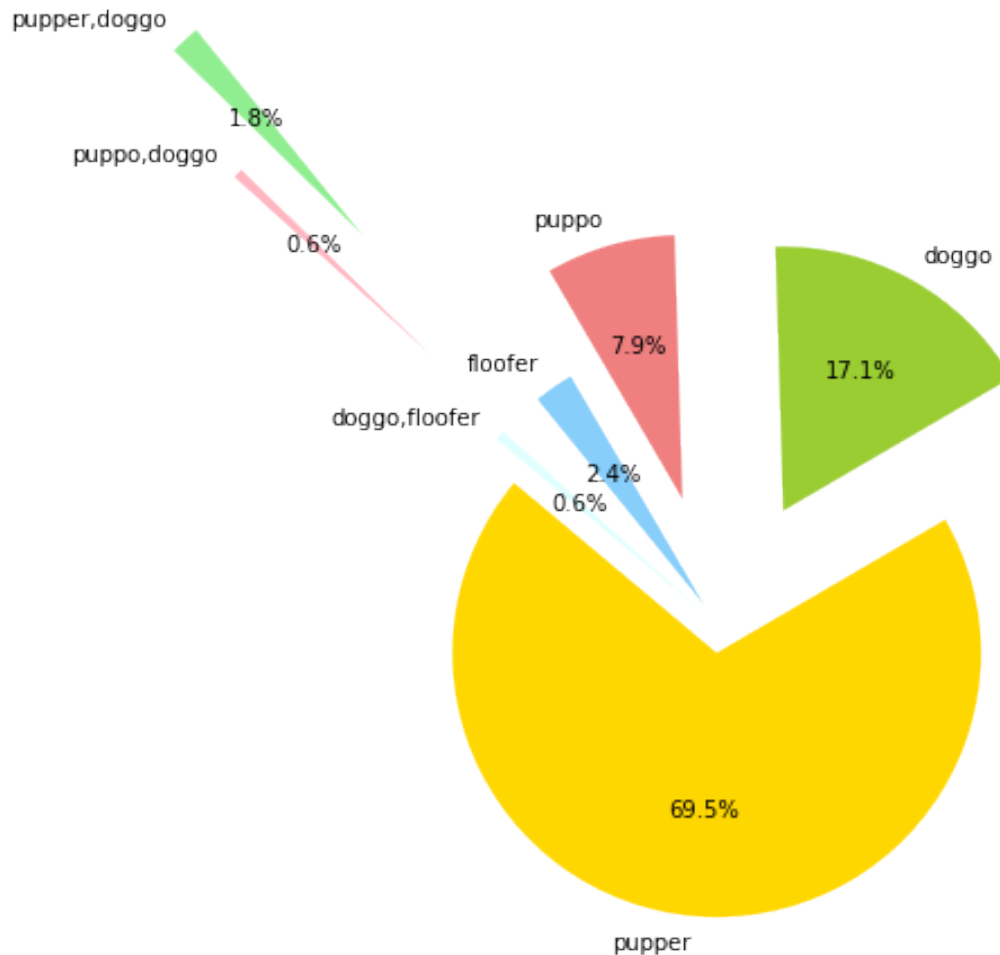
```
In [122]: twitter_archive_master.stage.value_counts().index.tolist()
```

```
Out[122]: pupper          114
          doggo           28
          puppo           13
          floofer          4
          pupper,doggo     3
          puppo,doggo      1
          doggo,floofer    1
          Name: stage, dtype: int64
```

```
In [150]: # exclude NA value first and then visualize the stage composition
labels = twitter_archive_master.stage.value_counts().index.tolist()
sizes = twitter_archive_master.stage.value_counts().tolist()

plt.figure(figsize=(8, 8))
plt.pie(sizes, labels=labels, explode=explode, colors=colors, autopct='%1.1f%%', startangle=90)

plt.axis('equal');
```



Puppies represent almost 70% then doggos represent 17%

2.4 Third Part - Reporting on:

- Wrangling Efforts
- Data Analyses and Visualizations

Both reports are attached in the submission file