

# wrangle\_act

December 26, 2018

## 1 Report

### 1.1 Gather

- Neste ponto foram feitos esforços para adquirir os dados necessários para o projeto. Primeiramente, foi gerado um arquivo no formato TOML que contém os tokens usados na API do Twitter, este arquivo é aberto como um dicionário pelo python utilizando a extensão toml. As variáveis constantes (em maiúsculo) são criadas com os valores desse token facilitando a manutenção do código e o objeto da API é criado. São utilizados `wait_on_rate_limit` e `wait_on_rate_limit_notify` para que, quando o tempo limite for excedido, a api espere para continuar e avise dessa pausa;
- Utilizando a url fornecida para o projeto 2, o arquivo `image-predictions.tsv` é carregado como um DataFrame diretamente, aproveitando os recursos das últimas versões do pandas que tem a capacidade de fazer download diretamente ao abrir o url;
- O arquivo `twitter-archive-enhanced.csv` é carregado como o dataframe `twitter_archive`;
- Utilizando os identificadores do dataframe `image_prediccion` e a API do twitter, os tweets são baixados no formato json e gravados no arquivo `tweet_json.txt` linha por linha usando a forma de gravação `append` de arquivos do python. Também foi adicionado o comando mágico do ipython `%%time` para calcular o tempo de execução;
- Utilizando o arquivo `tweet_json.txt` e acessar cada linha deste arquivo pegando as informações úteis e adicionando-as a um dataframe vazio chamado `df_tweets`.

### 1.2 Assess

- Acessa informações úteis de cada dataframe para entender o conteúdo de seus dados, algumas das informações que são acessadas são: `sample(10)` para enxergar 10 amostras aleatórias; `info()` para verificar valores faltantes e os tipos de dados de cada coluna; `duplicated()` para verificar se valores estão duplicados. Ainda foram feitas outras visualizações que ajudaram a esclarecer o que poderia ser melhorado, como retweets existentes e replys de tweets originais;
- Cada link de `extended_urls` foi acessado para descobrir se existiam tweets inexistentes e assim facilitar suas remoções;
- Uma lista de tarefas para melhorar a qualidade e arrumação dos dataframes foi criada para servir como guia na limpeza.

### 1.3 Clean

- Primeiramente, para cada item da lista foi criada uma definição, ou seja, uma informação do que foi realizado;
- Essa ordem de limpeza foi definida por tabela, onde toda a limpeza de uma tabela era feita antes de prosseguir para a seguinte, a menos que a limpeza não fosse possível;
- A baixo de cada definição, foi realizada a limpeza do código e um ou mais testes para verificar a corretitude da tabela depois da limpeza;
- Ao terminar as limpezas de qualidade se seguiu para as limpezas de arrumação onde foi feita a junção das tabelas com as colunas necessárias e ideais;
- O novo dataframe gerado pela junção ainda teve de receber alguns procedimentos de limpeza que serviram para melhorar a qualidade antes de ser salvo como `twitter_archive_master.csv`

### 1.4 Analyze

- Algumas pequenas análises descritivas foram realizadas, com a geração de alguns gráficos, e foram escritos alguns insights dos resultados.

## 2 Wrangle Act

### 2.1 Gather

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

%matplotlib inline

In [2]: # Load twitter tokens
parse_tokens = toml.load('tokens.toml')

# Set tweepy tokens
CONSUMER_KEY = parse_tokens['tokens']['consumer_key']
CONSUMER_SECRET = parse_tokens['tokens']['consumer_secret']
ACCESS_TOKEN = parse_tokens['tokens']['access_token']
ACCESS_SECRET = parse_tokens['tokens']['access_secret']

auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
auth.set_access_token(ACCESS_TOKEN, ACCESS_SECRET)

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

```

In [ ]: r = requests.get(url)
        with open('/path/I/want/to/save/file/to/file_name.pdf', 'wb') as f:
            f.write(r.content)

In [91]: # Download image_predictions.tsv
url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions.tsv'
r = requests.get(url)
with open('image_predictions.tsv', 'wb') as f:
    f.write(r.content)

image_predictions = pd.read_table('image_predictions.tsv')

In [4]: # Load tweets
twitter_archive = pd.read_csv('twitter-archive-enhanced.csv')

In [5]: %%time
        # Take tweet json by tweet_id:
        # *AVISO*: Execução demorada
id_404_erros = []
# tweets = twitter_archive[(twitter_archive.retweeted_status_id.isnull()) & (twitter_archive.status_id.isnull())]
tweets = image_predictions
for tweet_id in tweets.tweet_id:
    try:
        tweet = api.get_status(tweet_id, tweet_mode='extended')
        if tweet.user.id == 4196983835:
            with open('tweet_json.txt', 'a') as outfile:
                json.dump(tweet._json, outfile)
                outfile.write('\n')
    except tweepy.TweepError:
        # if tweepy.TweepError.message[0]['code'] == 404:
        id_404_erros.append(tweet_id)
        continue

Rate limit reached. Sleeping for: 547
Rate limit reached. Sleeping for: 578

CPU times: user 1min 6s, sys: 2.54 s, total: 1min 9s
Wall time: 31min 56s

In [6]: # Gera DataFrame de tweets puxados por id
df_tweets = pd.DataFrame(columns=[
    'tweet_id', 'text', 'retweet_count', 'favorite_count', 'jpg_url'

with open('tweet_json.txt', 'r') as file:
    for line in file:
        try:
            line_json = json.loads(line)

```

```

        if 'media' not in line_json['entities']:
            continue
        line_dict = {
            'tweet_id': line_json['id_str'],
            'text': line_json['full_text'],
            'retweet_count': line_json['retweet_count'],
            'favorite_count': line_json['favorite_count'],
            'jpg_url': line_json['entities']['media'][0]['media_url_https'],
            'url': line_json['entities']['media'][0]['url'],
            'expanded_urls': line_json['entities']['media'][0]['expanded_url']
        }

        df_tweets = df_tweets.append(
            pd.Series(line_dict).to_frame().transpose(), ignore_index=True) # Tran

    except Exception:
        raise Exception(line)
df_tweets.head()

```

```

Out[6]:
      tweet_id      text \
0  666020888022790149  Here we have a Japanese Irish Setter. Lost eye...
1  666029285002620928  This is a western brown Mitsubishi terrier. Up...
2  666033412701032449  Here is a very happy pup. Big fan of well-main...
3  666044226329800704  This is a purebred Piers Morgan. Loves to Netf...
4  666049248165822465  Here we have a 1949 1st generation vulpix. Enj...

      retweet_count  favorite_count \
0              504             2533
1              47              128
2              43              124
3             139              294
4              41              107

      jpg_url      url \
0  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg  https://t.co/BLDqew2Ijj
1  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg  https://t.co/r7m0b2m0UI
2  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg  https://t.co/y671yMhoiR
3  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg  https://t.co/DWnyCj2mx
4  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg  https://t.co/4B7c0c1EDq

      expanded_urls
0  https://twitter.com/dog_rates/status/666020888...
1  https://twitter.com/dog_rates/status/666029285...
2  https://twitter.com/dog_rates/status/666033412...
3  https://twitter.com/dog_rates/status/666044226...
4  https://twitter.com/dog_rates/status/666049248...

```

## 2.2 Assess

In [7]: *# Show dataframe loaded head.*

```
image_predictions.sample(10)
```

Out [7]:

	tweet_id	jpg_url \
909	700518061187723268	<a href="https://pbs.twimg.com/media/Cbi9dI_UYAAgkyC.jpg">https://pbs.twimg.com/media/Cbi9dI_UYAAgkyC.jpg</a>
1355	760290219849637889	<a href="https://pbs.twimg.com/ext_tw_video_thumb/76028...">https://pbs.twimg.com/ext_tw_video_thumb/76028...</a>
1326	757354760399941633	<a href="https://pbs.twimg.com/media/CoKqIndWgAAattd.jpg">https://pbs.twimg.com/media/CoKqIndWgAAattd.jpg</a>
2009	878281511006478336	<a href="https://pbs.twimg.com/media/DDBIX9QVYAAohGa.jpg">https://pbs.twimg.com/media/DDBIX9QVYAAohGa.jpg</a>
735	687096057537363968	<a href="https://pbs.twimg.com/media/CYkON6CVAAAPXAc.jpg">https://pbs.twimg.com/media/CYkON6CVAAAPXAc.jpg</a>
1903	851953902622658560	<a href="https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg">https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg</a>
1994	874057562936811520	<a href="https://pbs.twimg.com/media/DCFGtdoXkAEsqIw.jpg">https://pbs.twimg.com/media/DCFGtdoXkAEsqIw.jpg</a>
1019	710140971284037632	<a href="https://pbs.twimg.com/media/Cdrtcr-W4AAqi5H.jpg">https://pbs.twimg.com/media/Cdrtcr-W4AAqi5H.jpg</a>
983	707411934438625280	<a href="https://pbs.twimg.com/media/CdE7ZktXIAEiWLj.jpg">https://pbs.twimg.com/media/CdE7ZktXIAEiWLj.jpg</a>
2017	880095782870896641	<a href="https://pbs.twimg.com/media/DDa6ckbXgAAM1vV.jpg">https://pbs.twimg.com/media/DDa6ckbXgAAM1vV.jpg</a>

	img_num	p1	p1_conf	p1_dog \
909	1	American_Staffordshire_terrier	0.569501	True
1355	1	Old_English_sheepdog	0.302200	True
1326	1	Italian_greyhound	0.914667	True
2009	1	basset	0.320420	True
735	1	Labrador_retriever	0.417107	True
1903	1	Staffordshire_bullterrier	0.757547	True
1994	1	flat-coated_retriever	0.832177	True
1019	1	Pekinese	0.953170	True
983	1	Lakeland_terrier	0.738277	True
2017	1	miniature_pinscher	0.120298	True

	p2	p2_conf	p2_dog \
909	Staffordshire_bullterrier	0.211308	True
1355	Lhasa	0.258803	True
1326	whippet	0.047774	True
2009	collie	0.215975	True
735	Chesapeake_Bay_retriever	0.341730	True
1903	American_Staffordshire_terrier	0.149950	True
1994	black-and-tan_coonhound	0.040437	True
1019	papillon	0.019517	True
983	Airedale	0.028515	True
2017	Rhodesian_ridgeback	0.106395	True

	p3	p3_conf	p3_dog
909	Chihuahua	0.121839	True
1355	briard	0.179200	True
1326	ice_lolly	0.015477	False
2009	Appenzeller	0.128507	True
735	German_short-haired_pointer	0.177702	True
1903	Chesapeake_Bay_retriever	0.047523	True
1994	Newfoundland	0.028228	True

1019	Japanese_spaniel	0.005821	True
983	giant_schnauzer	0.024876	True
2017	beagle	0.106073	True

In [8]: image\_predictions.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 [9]: image\_predictions.loc[image\_predictions.jpg\_url.isnull(), :]

Out[9]: Empty DataFrame

Columns: [tweet\_id, jpg\_url, img\_num, p1, p1\_conf, p1\_dog, p2, p2\_conf, p2\_dog, p3, p3\_conf, p3\_dog]  
Index: []

In [10]: twitter\_archive.sample(10)

```
Out[10]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
1396	699788877217865730	NaN	NaN	
680	788908386943430656	NaN	NaN	
2214	668544745690562560	NaN	NaN	
1284	708479650088034305	NaN	NaN	
726	782598640137187329	NaN	NaN	
802	772193107915964416	NaN	NaN	
1024	746507379341139972	NaN	NaN	
29	886366144734445568	NaN	NaN	
930	754449512966619136	NaN	NaN	
1799	677187300187611136	NaN	NaN	

	timestamp	\
1396	2016-02-17 02:54:04 +0000	
680	2016-10-20 01:03:11 +0000	
2214	2015-11-22 21:41:02 +0000	

1284	2016-03-12 02:28:06 +0000
726	2016-10-02 15:10:30 +0000
802	2016-09-03 22:02:38 +0000
1024	2016-06-25 00:56:43 +0000
29	2017-07-15 23:25:31 +0000
930	2016-07-16 22:55:55 +0000
1799	2015-12-16 18:03:28 +0000

	source \
1396	<a href="http://twitter.com/download/iphone" r...
680	<a href="http://twitter.com/download/iphone" r...
2214	<a href="http://twitter.com/download/iphone" r...
1284	<a href="http://twitter.com/download/iphone" r...
726	<a href="http://twitter.com/download/iphone" r...
802	<a href="http://twitter.com/download/iphone" r...
1024	<a href="http://twitter.com/download/iphone" r...
29	<a href="http://twitter.com/download/iphone" r...
930	<a href="http://twitter.com/download/iphone" r...
1799	<a href="http://twitter.com/download/iphone" r...

	text	retweeted_status_id \
1396	Say hello to Nala. She's a Freckled High Brus...	NaN
680	This is Lucy. She destroyed not one, but two r...	NaN
2214	It is an honor to rate this pup. He is a Snork...	NaN
1284	Meet Bear. He's a Beneboop Cumberclap. Extreme...	NaN
726	This is Timmy. He's quite large. According to ...	NaN
802	This is Willem. He's a Penn State pupper. Thin...	NaN
1024	This is Shiloh. She did not pass the soft mout...	NaN
29	This is Roscoe. Another pupper fallen victim t...	NaN
930	This is Dex. He can see into your past and fut...	NaN
1799	Here we see a Byzantine Rigatoni. Very aerodyn...	NaN

	retweeted_status_user_id	retweeted_status_timestamp \
1396	NaN	NaN
680	NaN	NaN
2214	NaN	NaN
1284	NaN	NaN
726	NaN	NaN
802	NaN	NaN
1024	NaN	NaN
29	NaN	NaN
930	NaN	NaN
1799	NaN	NaN

	expanded_urls	rating_numerator \
1396	https://twitter.com/dog_rates/status/699788877...	12
680	https://twitter.com/dog_rates/status/788908386...	11
2214	https://twitter.com/dog_rates/status/668544745...	10

1284	https://twitter.com/dog_rates/status/708479650...	13
726	https://twitter.com/dog_rates/status/782598640...	11
802	https://twitter.com/dog_rates/status/772193107...	12
1024	https://twitter.com/dog_rates/status/746507379...	10
29	https://twitter.com/dog_rates/status/886366144...	12
930	https://twitter.com/dog_rates/status/754449512...	11
1799	https://twitter.com/dog_rates/status/677187300...	9

	rating_denominator	name	doggo	floofer	pupper	puppo
1396	10	Nala	None	None	None	None
680	10	Lucy	None	None	None	None
2214	10	None	None	None	None	None
1284	10	Bear	None	None	None	None
726	10	Timmy	None	None	None	None
802	10	Willem	None	None	pupper	None
1024	10	Shiloh	None	None	None	None
29	10	Roscoe	None	None	pupper	None
930	10	Dex	None	None	None	None
1799	10	None	None	None	None	None

```
In [11]: # Retweets e replies
         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 [12]: # Formato do texto
         twitter_archive.loc[0, 'text']
```



```
Out[12]: "This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/1
```

```
In [13]: # Retweet externo existente
twitter_archive.loc[546, :]
```

```
Out[13]: tweet_id                        805823200554876929
in_reply_to_status_id                    NaN
in_reply_to_user_id                     NaN
timestamp                               2016-12-05 17:16:37 +0000
source                                  <a href="http://twitter.com/download/iphone" r...
text                                    RT @dog_rates: This is Reginald. He's one magi...
retweeted_status_id                      7.84183e+17
retweeted_status_user_id                 4.19698e+09
retweeted_status_timestamp                2016-10-07 00:06:50 +0000
expanded_urls                           https://vine.co/v/5ghHLBMMdlV,https://vine.co/...
rating_numerator                         12
rating_denominator                       10
name                                     Reginald
doggo                                    None
floofer                                  None
pupper                                   None
puppo                                    puppo
Name: 546, dtype: object
```

```
In [14]: # Reply
twitter_archive.loc[twitter_archive.in_reply_to_status_id.notnull(), :]
twitter_archive.loc[149, :]
```

```
Out[14]: tweet_id                        863079547188785154
in_reply_to_status_id                    6.67152e+17
in_reply_to_user_id                     4.19698e+09
timestamp                               2017-05-12 17:12:53 +0000
source                                  <a href="http://twitter.com/download/iphone" r...
text                                    Ladies and gentlemen... I found Pipsy. He may ...
retweeted_status_id                      NaN
retweeted_status_user_id                 NaN
retweeted_status_timestamp                NaN
expanded_urls                           https://twitter.com/dog_rates/status/863079547...
rating_numerator                         14
rating_denominator                       10
name                                     None
doggo                                    None
floofer                                  None
pupper                                   None
puppo                                    None
Name: 149, dtype: object
```

```
In [15]: # Lista de replies
twitter_archive[twitter_archive.in_reply_to_status_id.notnull()]
```

```

Out[15]:
      tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
30    886267009285017600          8.862664e+17      2.281182e+09
55    881633300179243008          8.816070e+17      4.738443e+07
64    879674319642796034          8.795538e+17      3.105441e+09
113   870726314365509632          8.707262e+17      1.648776e+07
148   863427515083354112          8.634256e+17      7.759620e+07
149   863079547188785154          6.671522e+17      4.196984e+09
179   857214891891077121          8.571567e+17      1.806710e+08
184   856526610513747968          8.558181e+17      4.196984e+09
186   856288084350160898          8.562860e+17      2.792810e+08
188   855862651834028034          8.558616e+17      1.943518e+08
189   855860136149123072          8.558585e+17      1.361572e+07
218   850333567704068097          8.503288e+17      2.195506e+07
228   848213670039564288          8.482121e+17      4.196984e+09
234   847617282490613760          8.476062e+17      4.196984e+09
251   844979544864018432          7.590995e+17      4.196984e+09
274   840698636975636481          8.406983e+17      8.405479e+17
290   838150277551247360          8.381455e+17      2.195506e+07
291   838085839343206401          8.380855e+17      2.894131e+09
313   835246439529840640          8.352460e+17      2.625958e+07
342   832088576586297345          8.320875e+17      3.058208e+07
346   831926988323639298          8.319030e+17      2.068372e+07
387   826598799820865537          8.265984e+17      4.196984e+09
409   823333489516937216          8.233264e+17      1.582854e+09
427   821153421864615936          8.211526e+17      1.132119e+08
498   813130366689148928          8.131273e+17      4.196984e+09
513   811647686436880384          8.116272e+17      4.196984e+09
565   802265048156610565          7.331095e+17      4.196984e+09
570   801854953262350336          8.018543e+17      1.185634e+07
576   800859414831898624          8.008580e+17      2.918590e+08
611   797165961484890113          7.971238e+17      2.916630e+07
...
1479  693582294167244802          6.935722e+17      1.198989e+09
1497  692423280028966913          6.924173e+17      4.196984e+09
1501  692142790915014657          6.920419e+17      4.196984e+09
1523  690607260360429569          6.903413e+17      4.670367e+08
1598  686035780142297088          6.860340e+17      4.196984e+09
1605  685681090388975616          6.855479e+17      4.196984e+09
1618  684969860808454144          6.849598e+17      4.196984e+09
1630  684538444857667585          6.844811e+17      4.196984e+09
1634  684225744407494656          6.842229e+17      4.196984e+09
1663  682808988178739200          6.827884e+17      4.196984e+09
1689  681340665377193984          6.813394e+17      4.196984e+09
1774  678023323247357953          6.780211e+17      4.196984e+09
1819  676590572941893632          6.765883e+17      4.196984e+09
1842  675870721063669760          6.757073e+17      4.196984e+09
1844  675849018447167488          6.758457e+17      4.196984e+09
1852  675707330206547968          6.754971e+17      4.196984e+09

```

1866	675349384339542016	6.749998e+17	4.196984e+09
1882	674999807681908736	6.747934e+17	4.196984e+09
1885	674793399141146624	6.717299e+17	4.196984e+09
1892	674754018082705410	6.747522e+17	4.196984e+09
1895	674742531037511680	6.747400e+17	4.196984e+09
1905	674606911342424069	6.744689e+17	4.196984e+09
1914	674330906434379776	6.658147e+17	1.637468e+07
1940	673716320723169284	6.737159e+17	4.196984e+09
2036	671729906628341761	6.715610e+17	4.196984e+09
2038	671550332464455680	6.715449e+17	4.196984e+09
2149	669684865554620416	6.693544e+17	4.196984e+09
2169	669353438988365824	6.678065e+17	4.196984e+09
2189	668967877119254528	6.689207e+17	2.143566e+07
2298	667070482143944705	6.670655e+17	4.196984e+09

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

1497 2016-01-27 19:05:49 +0000  
 1501 2016-01-27 00:31:15 +0000  
 1523 2016-01-22 18:49:36 +0000  
 1598 2016-01-10 04:04:10 +0000  
 1605 2016-01-09 04:34:45 +0000  
 1618 2016-01-07 05:28:35 +0000  
 1630 2016-01-06 00:54:18 +0000  
 1634 2016-01-05 04:11:44 +0000  
 1663 2016-01-01 06:22:03 +0000  
 1689 2015-12-28 05:07:27 +0000  
 1774 2015-12-19 01:25:31 +0000  
 1819 2015-12-15 02:32:17 +0000  
 1842 2015-12-13 02:51:51 +0000  
 1844 2015-12-13 01:25:37 +0000  
 1852 2015-12-12 16:02:36 +0000  
 1866 2015-12-11 16:20:15 +0000  
 1882 2015-12-10 17:11:09 +0000  
 1885 2015-12-10 03:30:58 +0000  
 1892 2015-12-10 00:54:28 +0000  
 1895 2015-12-10 00:08:50 +0000  
 1905 2015-12-09 15:09:55 +0000  
 1914 2015-12-08 20:53:11 +0000  
 1940 2015-12-07 04:11:02 +0000  
 2036 2015-12-01 16:37:44 +0000  
 2038 2015-12-01 04:44:10 +0000  
 2149 2015-11-26 01:11:28 +0000  
 2169 2015-11-25 03:14:30 +0000  
 2189 2015-11-24 01:42:25 +0000  
 2298 2015-11-18 20:02:51 +0000

source \
 30 <a href="http://twitter.com/download/iphone" r...  
 55 <a href="http://twitter.com/download/iphone" r...  
 64 <a href="http://twitter.com/download/iphone" r...  
 113 <a href="http://twitter.com/download/iphone" r...  
 148 <a href="http://twitter.com/download/iphone" r...  
 149 <a href="http://twitter.com/download/iphone" r...  
 179 <a href="http://twitter.com/download/iphone" r...  
 184 <a href="http://twitter.com/download/iphone" r...  
 186 <a href="http://twitter.com/download/iphone" r...  
 188 <a href="http://twitter.com/download/iphone" r...  
 189 <a href="http://twitter.com/download/iphone" r...  
 218 <a href="http://twitter.com/download/iphone" r...  
 228 <a href="http://twitter.com/download/iphone" r...  
 234 <a href="http://twitter.com/download/iphone" r...  
 251 <a href="http://twitter.com/download/iphone" r...  
 274 <a href="http://twitter.com/download/iphone" r...  
 290 <a href="http://twitter.com/download/iphone" r...

291 <a href="http://twitter.com/download/iphone" r...  
 313 <a href="http://twitter.com/download/iphone" r...  
 342 <a href="http://twitter.com/download/iphone" r...  
 346 <a href="http://twitter.com/download/iphone" r...  
 387 <a href="http://twitter.com/download/iphone" r...  
 409 <a href="http://twitter.com/download/iphone" r...  
 427 <a href="http://twitter.com/download/iphone" r...  
 498 <a href="http://twitter.com/download/iphone" r...  
 513 <a href="http://twitter.com/download/iphone" r...  
 565 <a href="http://twitter.com/download/iphone" r...  
 570 <a href="http://twitter.com/download/iphone" r...  
 576 <a href="http://twitter.com/download/iphone" r...  
 611 <a href="http://twitter.com/download/iphone" r...  
 ...  
 1479 <a href="http://twitter.com/download/iphone" r...  
 1497 <a href="http://twitter.com/download/iphone" r...  
 1501 <a href="http://twitter.com/download/iphone" r...  
 1523 <a href="http://twitter.com/download/iphone" r...  
 1598 <a href="http://twitter.com/download/iphone" r...  
 1605 <a href="http://twitter.com/download/iphone" r...  
 1618 <a href="http://twitter.com/download/iphone" r...  
 1630 <a href="http://twitter.com/download/iphone" r...  
 1634 <a href="http://twitter.com/download/iphone" r...  
 1663 <a href="http://twitter.com/download/iphone" r...  
 1689 <a href="http://twitter.com/download/iphone" r...  
 1774 <a href="http://twitter.com/download/iphone" r...  
 1819 <a href="http://twitter.com/download/iphone" r...  
 1842 <a href="http://twitter.com/download/iphone" r...  
 1844 <a href="http://twitter.com/download/iphone" r...  
 1852 <a href="http://twitter.com/download/iphone" r...  
 1866 <a href="http://twitter.com/download/iphone" r...  
 1882 <a href="http://twitter.com/download/iphone" r...  
 1885 <a href="http://twitter.com/download/iphone" r...  
 1892 <a href="http://twitter.com/download/iphone" r...  
 1895 <a href="http://twitter.com/download/iphone" r...  
 1905 <a href="http://twitter.com/download/iphone" r...  
 1914 <a href="http://twitter.com/download/iphone" r...  
 1940 <a href="http://twitter.com/download/iphone" r...  
 2036 <a href="http://twitter.com/download/iphone" r...  
 2038 <a href="http://twitter.com/download/iphone" r...  
 2149 <a href="http://twitter.com/download/iphone" r...  
 2169 <a href="http://twitter.com/download/iphone" r...  
 2189 <a href="http://twitter.com/download/iphone" r...  
 2298 <a href="http://twitter.com/download/iphone" r...

		text	retweeted_status_id \
30	@NonWhiteHat	@MayhewMayhem omg hello tanner yo...	NaN
55	@rroushfenway	These are good dogs but 17/10 is ...	NaN

64	@RealKentMurphy 14/10 confirmed	NaN
113	@ComplicitOwl @ShopWeRateDogs >10/10 is res...	NaN
148	@Jack_Septic_Eye I'd need a few more pics to p...	NaN
149	Ladies and gentlemen... I found Pipsy. He may ...	NaN
179	@Marc_IRL pixelated af 12/10	NaN
184	THIS IS CHARLIE, MARK. HE DID JUST WANT TO SAY...	NaN
186	@xianmcguire @Jenna_Marbles Kardashians wouldn...	NaN
188	@dhmontgomery We also gave snoop dogg a 420/10...	NaN
189	@s8n You tried very hard to portray this good ...	NaN
218	@markhoppus MARK THAT DOG HAS SEEN AND EXPERIE...	NaN
228	Jerry just apuppologized to me. He said there ...	NaN
234	.@breaannanicolee PUPDATE: Cannon has a heart ...	NaN
251	PUPDATE: I'm proud to announce that Toby is 23...	NaN
274	@O_kelvin_0 >10/10 is reserved for puppos s...	NaN
290	@markhoppus 182/10	NaN
291	@bragg6of8 @Andy_Pace_ we are still looking fo...	NaN
313	@jonnyusun @Lin_Manuel ok jomny I know you're e...	NaN
342	@docmisterio account started on 11/15/15	NaN
346	@UNC can confirm 12/10	NaN
387	I was going to do 007/10, but the joke wasn't ...	NaN
409	@HistoryInPics 13/10	NaN
427	@imgur for a polar bear tho I'd say 13/10 is a...	NaN
498	I've been informed by multiple sources that th...	NaN
513	PUPDATE: I've been informed that Augie was act...	NaN
565	Like doggo, like pupper version 2. Both 11/10 ...	NaN
570	.@NBCSports OMG THE TINY HAT I'M GOING TO HAVE...	NaN
576	@SkyWilliams doggo simply protecting you from ...	NaN
611	@JODYHIGHROLLER it may be an 11/10 but what do...	NaN
...	...	...
1479	Personally I'd give him an 11/10. Not sure why...	NaN
1497	PUPDATE: just noticed this dog has some extra ...	NaN
1501	These are some pictures of Teddy that further ...	NaN
1523	12/10 @LightningHoltt	NaN
1598	Yes I do realize a rating of 4/20 would've bee...	NaN
1605	Jack deserves another round of applause. If yo...	NaN
1618	For those who claim this is a goat, u are wron...	NaN
1630	After watching this video, we've determined th...	NaN
1634	Two sneaky puppers were not initially seen, mo...	NaN
1663	I'm aware that I could've said 20/16, but here...	NaN
1689	I've been told there's a slight possibility he...	NaN
1774	After getting lost in Reese's eyes for several...	NaN
1819	After some outrage from the crowd. Bubbles is ...	NaN
1842	& this is Yoshi. Another world record cont...	NaN
1844	This dog is being demoted to a 9/10 for not we...	NaN
1852	We've got ourselves a battle here. Watch out R...	NaN
1866	Yea I lied. Here's more. All 13/10 https://t.c...	NaN
1882	Ok last one of these. I may try to make some m...	NaN
1885	I have found another. 13/10 https://t.co/HwroP...	NaN

1892	Just received another perfect photo of dogs an...	NaN
1895	Some clarification is required. The dog is sin...	NaN
1905	The 13/10 also takes into account this impecca...	NaN
1914	13/10\n@ABC7	NaN
1940	The millennials have spoken and we've decided ...	NaN
2036	I'm just going to leave this one here as well...	NaN
2038	After 22 minutes of careful deliberation this ...	NaN
2149	After countless hours of research and hundreds...	NaN
2169	This is Tessa. She is also very pleased after ...	NaN
2189	12/10 good shit Bubka\n@wane15	NaN
2298	After much debate this dog is being upgraded t...	NaN

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

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

	expanded_urls	rating_numerator \
30	NaN	12
55	NaN	17
64	NaN	14
113	NaN	10
148	NaN	12
149	https://twitter.com/dog_rates/status/863079547...	14
179	NaN	12
184	https://twitter.com/dog_rates/status/856526610...	14
186	NaN	14
188	NaN	420
189	NaN	666
218	NaN	13
228	NaN	11
234	NaN	13
251	https://twitter.com/dog_rates/status/844979544...	13
274	NaN	10
290	NaN	182
291	NaN	15
313	NaN	960
342	NaN	11



346		NaN	12
387		NaN	7
409		NaN	13
427		NaN	13
498		NaN	12
513		NaN	11
565	<a href="https://twitter.com/dog_rates/status/802265048...">https://twitter.com/dog_rates/status/802265048...</a>		11
570		NaN	11
576		NaN	11
611		NaN	11
...		...	...
1479		NaN	11
1497		NaN	9
1501	<a href="https://twitter.com/dog_rates/status/692142790...">https://twitter.com/dog_rates/status/692142790...</a>		13
1523		NaN	12
1598		NaN	4
1605		NaN	14
1618		NaN	5
1630	<a href="https://twitter.com/dog_rates/status/684538444...">https://twitter.com/dog_rates/status/684538444...</a>		12
1634	<a href="https://twitter.com/dog_rates/status/684225744...">https://twitter.com/dog_rates/status/684225744...</a>		143
1663		NaN	20
1689		NaN	5
1774		NaN	13
1819		NaN	7
1842	<a href="https://twitter.com/dog_rates/status/675870721...">https://twitter.com/dog_rates/status/675870721...</a>		11
1844		NaN	9
1852	<a href="https://twitter.com/dog_rates/status/675707330...">https://twitter.com/dog_rates/status/675707330...</a>		11
1866	<a href="https://twitter.com/dog_rates/status/675349384...">https://twitter.com/dog_rates/status/675349384...</a>		13
1882	<a href="https://twitter.com/dog_rates/status/674999807...">https://twitter.com/dog_rates/status/674999807...</a>		13
1885	<a href="https://twitter.com/dog_rates/status/674793399...">https://twitter.com/dog_rates/status/674793399...</a>		13
1892	<a href="https://twitter.com/dog_rates/status/674754018...">https://twitter.com/dog_rates/status/674754018...</a>		12
1895		NaN	11
1905		NaN	13
1914		NaN	13
1940		NaN	1
2036	<a href="https://twitter.com/dog_rates/status/671729906...">https://twitter.com/dog_rates/status/671729906...</a>		13
2038		NaN	1
2149		NaN	11
2169	<a href="https://twitter.com/dog_rates/status/669353438...">https://twitter.com/dog_rates/status/669353438...</a>		10
2189		NaN	12
2298		NaN	10

	rating_denominator	name	doggo	floofer	pupper	puppo
30	10	None	None	None	None	None
55	10	None	None	None	None	None
64	10	None	None	None	None	None
113	10	None	None	None	None	None
148	10	None	None	None	None	None

149	10	None	None	None	None	None
179	10	None	None	None	None	None
184	10	None	None	None	None	None
186	10	None	None	None	None	None
188	10	None	None	None	None	None
189	10	None	None	None	None	None
218	10	None	None	None	None	None
228	10	None	None	None	None	None
234	10	None	None	None	None	None
251	10	None	None	None	None	None
274	10	None	None	None	None	None
290	10	None	None	None	None	None
291	10	None	None	None	None	None
313	0	None	None	None	None	None
342	15	None	None	None	None	None
346	10	None	None	None	None	None
387	10	None	None	None	None	None
409	10	None	None	None	None	None
427	10	None	None	None	None	None
498	10	None	None	None	None	None
513	10	None	None	None	None	None
565	10	None	doggo	None	pupper	None
570	10	None	None	None	None	None
576	10	None	doggo	None	None	None
611	10	None	None	None	None	None
...	...	...	...	...	...	...
1479	10	None	None	None	None	None
1497	10	None	None	None	None	None
1501	10	None	None	None	None	None
1523	10	None	None	None	None	None
1598	20	None	None	None	None	None
1605	10	None	None	None	None	None
1618	10	None	None	None	None	None
1630	10	None	None	None	None	None
1634	130	None	None	None	None	None
1663	16	None	None	None	None	None
1689	10	None	None	None	None	None
1774	10	None	None	None	None	None
1819	10	None	None	None	None	None
1842	10	None	None	None	None	None
1844	10	None	None	None	None	None
1852	10	None	None	None	None	None
1866	10	None	None	None	None	None
1882	10	None	None	None	None	None
1885	10	None	None	None	None	None
1892	10	None	None	None	None	None
1895	10	None	None	None	None	None
1905	10	None	None	None	None	None

1914	10	None	None	None	None	None
1940	10	None	None	None	None	None
2036	10	None	None	None	None	None
2038	10	None	None	None	None	None
2149	10	None	None	None	None	None
2169	10	Tessa	None	None	None	None
2189	10	None	None	None	None	None
2298	10	None	None	None	None	None

[78 rows x 17 columns]

In [16]: # *Lista de retweets*

twitter\_archive[twitter\_archive.retweeted\_status\_id.notnull()]

Out[16]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
19	888202515573088257	NaN	NaN	
32	886054160059072513	NaN	NaN	
36	885311592912609280	NaN	NaN	
68	879130579576475649	NaN	NaN	
73	878404777348136964	NaN	NaN	
74	878316110768087041	NaN	NaN	
78	877611172832227328	NaN	NaN	
91	874434818259525634	NaN	NaN	
95	873697596434513921	NaN	NaN	
97	873337748698140672	NaN	NaN	
101	872668790621863937	NaN	NaN	
109	871166179821445120	NaN	NaN	
118	869988702071779329	NaN	NaN	
124	868639477480148993	NaN	NaN	
130	867072653475098625	NaN	NaN	
132	866816280283807744	NaN	NaN	
137	866094527597207552	NaN	NaN	
146	863471782782697472	NaN	NaN	
155	861769973181624320	NaN	NaN	
159	860981674716409858	NaN	NaN	
160	860924035999428608	NaN	NaN	
165	860177593139703809	NaN	NaN	
171	858860390427611136	NaN	NaN	
180	857062103051644929	NaN	NaN	
182	856602993587888130	NaN	NaN	
185	856330835276025856	NaN	NaN	
194	855245323840757760	NaN	NaN	
195	855138241867124737	NaN	NaN	
204	852936405516943360	NaN	NaN	
211	851953902622658560	NaN	NaN	
...	...	...	...	
784	775096608509886464	NaN	NaN	
794	773336787167145985	NaN	NaN	

800	772615324260794368	NaN	NaN
811	771171053431250945	NaN	NaN
815	771004394259247104	NaN	NaN
818	770743923962707968	NaN	NaN
822	770093767776997377	NaN	NaN
826	769335591808995329	NaN	NaN
829	768909767477751808	NaN	NaN
833	768554158521745409	NaN	NaN
841	766864461642756096	NaN	NaN
847	766078092750233600	NaN	NaN
860	763167063695355904	NaN	NaN
868	761750502866649088	NaN	NaN
872	761371037149827077	NaN	NaN
885	760153949710192640	NaN	NaN
890	759566828574212096	NaN	NaN
895	759159934323924993	NaN	NaN
908	757729163776290825	NaN	NaN
911	757597904299253760	NaN	NaN
926	754874841593970688	NaN	NaN
937	753298634498793472	NaN	NaN
943	752701944171524096	NaN	NaN
949	752309394570878976	NaN	NaN
1012	747242308580548608	NaN	NaN
1023	746521445350707200	NaN	NaN
1043	743835915802583040	NaN	NaN
1242	711998809858043904	NaN	NaN
2259	667550904950915073	NaN	NaN
2260	667550882905632768	NaN	NaN

	timestamp \
19	2017-07-21 01:02:36 +0000
32	2017-07-15 02:45:48 +0000
36	2017-07-13 01:35:06 +0000
68	2017-06-26 00:13:58 +0000
73	2017-06-24 00:09:53 +0000
74	2017-06-23 18:17:33 +0000
78	2017-06-21 19:36:23 +0000
91	2017-06-13 01:14:41 +0000
95	2017-06-11 00:25:14 +0000
97	2017-06-10 00:35:19 +0000
101	2017-06-08 04:17:07 +0000
109	2017-06-04 00:46:17 +0000
118	2017-05-31 18:47:24 +0000
124	2017-05-28 01:26:04 +0000
130	2017-05-23 17:40:04 +0000
132	2017-05-23 00:41:20 +0000
137	2017-05-21 00:53:21 +0000
146	2017-05-13 19:11:30 +0000

155 2017-05-09 02:29:07 +0000  
 159 2017-05-06 22:16:42 +0000  
 160 2017-05-06 18:27:40 +0000  
 165 2017-05-04 17:01:34 +0000  
 171 2017-05-01 01:47:28 +0000  
 180 2017-04-26 02:41:43 +0000  
 182 2017-04-24 20:17:23 +0000  
 185 2017-04-24 02:15:55 +0000  
 194 2017-04-21 02:22:29 +0000  
 195 2017-04-20 19:16:59 +0000  
 204 2017-04-14 17:27:40 +0000  
 211 2017-04-12 00:23:33 +0000  
 ... ..  
 784 2016-09-11 22:20:06 +0000  
 794 2016-09-07 01:47:12 +0000  
 800 2016-09-05 02:00:22 +0000  
 811 2016-09-01 02:21:21 +0000  
 815 2016-08-31 15:19:06 +0000  
 818 2016-08-30 22:04:05 +0000  
 822 2016-08-29 03:00:36 +0000  
 826 2016-08-27 00:47:53 +0000  
 829 2016-08-25 20:35:48 +0000  
 833 2016-08-24 21:02:45 +0000  
 841 2016-08-20 05:08:29 +0000  
 847 2016-08-18 01:03:45 +0000  
 860 2016-08-10 00:16:21 +0000  
 868 2016-08-06 02:27:27 +0000  
 872 2016-08-05 01:19:35 +0000  
 885 2016-08-01 16:43:19 +0000  
 890 2016-07-31 01:50:18 +0000  
 895 2016-07-29 22:53:27 +0000  
 908 2016-07-26 00:08:05 +0000  
 911 2016-07-25 15:26:30 +0000  
 926 2016-07-18 03:06:01 +0000  
 937 2016-07-13 18:42:44 +0000  
 943 2016-07-12 03:11:42 +0000  
 949 2016-07-11 01:11:51 +0000  
 1012 2016-06-27 01:37:04 +0000  
 1023 2016-06-25 01:52:36 +0000  
 1043 2016-06-17 16:01:16 +0000  
 1242 2016-03-21 19:31:59 +0000  
 2259 2015-11-20 03:51:52 +0000  
 2260 2015-11-20 03:51:47 +0000

source \  
 19 <a href="http://twitter.com/download/iphone" r...  
 32 <a href="http://twitter.com/download/iphone" r...  
 36 <a href="http://twitter.com/download/iphone" r...

[illegible]

926 <a href="http://twitter.com/download/iphone" r...  
 937 <a href="http://twitter.com/download/iphone" r...  
 943 <a href="http://twitter.com/download/iphone" r...  
 949 <a href="http://twitter.com/download/iphone" r...  
 1012 <a href="http://twitter.com/download/iphone" r...  
 1023 <a href="http://twitter.com/download/iphone" r...  
 1043 <a href="http://twitter.com/download/iphone" r...  
 1242 <a href="http://twitter.com/download/iphone" r...  
 2259 <a href="http://twitter.com" rel="nofollow">Tw...  
 2260 <a href="http://twitter.com" rel="nofollow">Tw...

	text	retweeted_status_id \
19	RT @dog_rates: This is Canela. She attempted s...	8.874740e+17
32	RT @Athletics: 12/10 #BATP https://t.co/WxwJmv...	8.860537e+17
36	RT @dog_rates: This is Lilly. She just paralle...	8.305833e+17
68	RT @dog_rates: This is Emmy. She was adopted t...	8.780576e+17
73	RT @dog_rates: Meet Shadow. In an attempt to r...	8.782815e+17
74	RT @dog_rates: Meet Terrance. He's being yelle...	6.690004e+17
78	RT @rachel2195: @dog_rates the boyfriend and h...	8.768508e+17
91	RT @dog_rates: This is Coco. At first I though...	8.663350e+17
95	RT @dog_rates: This is Walter. He won't start ...	8.688804e+17
97	RT @dog_rates: This is Sierra. She's one preci...	8.732138e+17
101	RT @loganamnosis: Penelope here is doing me qu...	8.726576e+17
109	RT @dog_rates: This is Dawn. She's just checki...	8.410770e+17
118	RT @dog_rates: We only rate dogs. This is quit...	8.591970e+17
124	RT @dog_rates: Say hello to Cooper. His expres...	8.685523e+17
130	RT @rachaeleasler: these @dog_rates hats are 1...	8.650134e+17
132	RT @dog_rates: This is Jamesy. He gives a kiss...	8.664507e+17
137	RT @dog_rates: Here's a pupper before and afte...	8.378202e+17
146	RT @dog_rates: Say hello to Quinn. She's quite...	8.630625e+17
155	RT @dog_rates: "Good afternoon class today we'...	8.066291e+17
159	RT @dog_rates: Meet Lorenzo. He's an avid nift...	8.605638e+17
160	RT @tallylott: h*ckin adorable promposal. 13/1...	8.609145e+17
165	RT @dog_rates: Ohboyohboyohboyohboyohboyohboy...	7.616730e+17
171	RT @dog_rates: Meet Winston. He knows he's a l...	8.395493e+17
180	RT @AaronChewning: First time wearing my @dog_...	8.570611e+17
182	RT @dog_rates: This is Luna. It's her first ti...	8.447048e+17
185	RT @Jenna_Marbles: @dog_rates Thanks for ratin...	8.563302e+17
194	RT @dog_rates: Meet George. He looks slightly ...	8.421635e+17
195	RT @frasercampbell_: oh my... what's that... b...	8.551225e+17
204	RT @dog_rates: I usually only share these on F...	8.316501e+17
211	RT @dog_rates: This is Astrid. She's a guide d...	8.293743e+17
...	...	...
784	RT @dog_rates: After so many requests, this is...	7.403732e+17
794	RT @dog_rates: Meet Fizz. She thinks love is a...	7.713808e+17
800	RT @dog_rates: This is Gromit. He's pupset bec...	7.652221e+17
811	RT @dog_rates: This is Frankie. He's wearing b...	6.733201e+17
815	RT @katieornah: @dog_rates learning a lot at c...	7.710021e+17

818	RT @dog_rates: Here's a doggo blowing bubbles...	7.392382e+17
822	RT @dog_rates: This is just downright precious...	7.410673e+17
826	RT @dog_rates: Ever seen a dog pet another dog...	7.069045e+17
829	RT @dog_rates: When it's Janet from accounting...	7.001438e+17
833	RT @dog_rates: This is Nollie. She's waving at...	7.399792e+17
841	RT @dog_rates: We only rate dogs... this is a ...	7.599238e+17
847	RT @dog_rates: This is Colby. He's currently r...	7.258423e+17
860	RT @dog_rates: Meet Eve. She's a raging alcoho...	6.732953e+17
868	RT @dog_rates: "Tristan do not speak to me wit...	6.853251e+17
872	RT @dog_rates: Oh. My. God. 13/10 magical af h...	7.116948e+17
885	RT @hownottodraw: The story/person behind @dog...	7.601538e+17
890	RT @dog_rates: This... is a Tyrannosaurus rex...	7.395441e+17
895	RT @dog_rates: AT DAWN...\nWE RIDE\n\n11/10 ht...	6.703191e+17
908	RT @dog_rates: This is Chompsky. He lives up t...	6.790626e+17
911	RT @jon_hill987: @dog_rates There is a cunning...	7.575971e+17
926	RT @dog_rates: This is Rubio. He has too much ...	6.791584e+17
937	RT @dog_rates: This is Carly. She's actually 2...	6.815232e+17
943	RT @dog_rates: HEY PUP WHAT'S THE PART OF THE ...	6.835159e+17
949	RT @dog_rates: Everyone needs to watch this. 1...	6.753544e+17
1012	RT @dog_rates: This pupper killed this great w...	7.047611e+17
1023	RT @dog_rates: This is Shaggy. He knows exactl...	6.678667e+17
1043	RT @dog_rates: Extremely intelligent dog here...	6.671383e+17
1242	RT @twitter: @dog_rates Awesome Tweet! 12/10. ...	7.119983e+17
2259	RT @dogratingrating: Exceptional talent. Origi...	6.675487e+17
2260	RT @dogratingrating: Unoriginal idea. Blatant ...	6.675484e+17

	retweeted_status_user_id	retweeted_status_timestamp	\
19	4.196984e+09	2017-07-19 00:47:34 +0000	
32	1.960740e+07	2017-07-15 02:44:07 +0000	
36	4.196984e+09	2017-02-12 01:04:29 +0000	
68	4.196984e+09	2017-06-23 01:10:23 +0000	
73	4.196984e+09	2017-06-23 16:00:04 +0000	
74	4.196984e+09	2015-11-24 03:51:38 +0000	
78	5.128045e+08	2017-06-19 17:14:49 +0000	
91	4.196984e+09	2017-05-21 16:48:45 +0000	
95	4.196984e+09	2017-05-28 17:23:24 +0000	
97	4.196984e+09	2017-06-09 16:22:42 +0000	
101	1.547674e+08	2017-06-08 03:32:35 +0000	
109	4.196984e+09	2017-03-13 00:02:39 +0000	
118	4.196984e+09	2017-05-02 00:04:57 +0000	
124	4.196984e+09	2017-05-27 19:39:34 +0000	
130	7.874618e+17	2017-05-18 01:17:25 +0000	
132	4.196984e+09	2017-05-22 00:28:40 +0000	
137	4.196984e+09	2017-03-04 00:21:08 +0000	
146	4.196984e+09	2017-05-12 16:05:02 +0000	
155	4.196984e+09	2016-12-07 22:38:52 +0000	
159	4.196984e+09	2017-05-05 18:36:06 +0000	
160	3.638908e+08	2017-05-06 17:49:42 +0000	



165	4.196984e+09	2016-08-05 21:19:27	+0000
171	4.196984e+09	2017-03-08 18:52:12	+0000
180	5.870972e+07	2017-04-26 02:37:47	+0000
182	4.196984e+09	2017-03-23 00:18:10	+0000
185	6.669901e+07	2017-04-24 02:13:14	+0000
194	4.196984e+09	2017-03-16 00:00:07	+0000
195	7.475543e+17	2017-04-20 18:14:33	+0000
204	4.196984e+09	2017-02-14 23:43:18	+0000
211	4.196984e+09	2017-02-08 17:00:26	+0000
...	...	...	...
784	4.196984e+09	2016-06-08 02:41:38	+0000
794	4.196984e+09	2016-09-01 16:14:48	+0000
800	4.196984e+09	2016-08-15 16:22:20	+0000
811	4.196984e+09	2015-12-06 01:56:44	+0000
815	1.732729e+09	2016-08-31 15:10:07	+0000
818	4.196984e+09	2016-06-04 23:31:25	+0000
822	4.196984e+09	2016-06-10 00:39:48	+0000
826	4.196984e+09	2016-03-07 18:09:06	+0000
829	4.196984e+09	2016-02-18 02:24:13	+0000
833	4.196984e+09	2016-06-07 00:36:02	+0000
841	4.196984e+09	2016-08-01 01:28:46	+0000
847	4.196984e+09	2016-04-29 00:21:01	+0000
860	4.196984e+09	2015-12-06 00:17:55	+0000
868	4.196984e+09	2016-01-08 05:00:14	+0000
872	4.196984e+09	2016-03-20 23:23:54	+0000
885	1.950368e+08	2016-08-01 16:42:51	+0000
890	4.196984e+09	2016-06-05 19:47:03	+0000
895	4.196984e+09	2015-11-27 19:11:49	+0000
908	4.196984e+09	2015-12-21 22:15:18	+0000
911	2.804798e+08	2016-07-25 15:23:28	+0000
926	4.196984e+09	2015-12-22 04:35:49	+0000
937	4.196984e+09	2015-12-28 17:12:42	+0000
943	4.196984e+09	2016-01-03 05:11:12	+0000
949	4.196984e+09	2015-12-11 16:40:19	+0000
1012	4.196984e+09	2016-03-01 20:11:59	+0000
1023	4.196984e+09	2015-11-21 00:46:50	+0000
1043	4.196984e+09	2015-11-19 00:32:12	+0000
1242	7.832140e+05	2016-03-21 19:29:52	+0000
2259	4.296832e+09	2015-11-20 03:43:06	+0000
2260	4.296832e+09	2015-11-20 03:41:59	+0000

	expanded_urls	rating_numerator	\
19	<a href="https://twitter.com/dog_rates/status/887473957...">https://twitter.com/dog_rates/status/887473957...</a>	13	
32	<a href="https://twitter.com/dog_rates/status/886053434...">https://twitter.com/dog_rates/status/886053434...</a>	12	
36	<a href="https://twitter.com/dog_rates/status/830583320...">https://twitter.com/dog_rates/status/830583320...</a>	13	
68	<a href="https://twitter.com/dog_rates/status/878057613...">https://twitter.com/dog_rates/status/878057613...</a>	14	
73	<a href="https://www.gofundme.com/3yd6y1c">https://www.gofundme.com/3yd6y1c</a> , <a href="https://twitt...">https://twitt...</a>	13	
74	<a href="https://twitter.com/dog_rates/status/669000397...">https://twitter.com/dog_rates/status/669000397...</a>	11	

78	<a href="https://twitter.com/rachel2195/status/87685077...">https://twitter.com/rachel2195/status/87685077...</a>	14
91	<a href="https://twitter.com/dog_rates/status/866334964...">https://twitter.com/dog_rates/status/866334964...</a>	12
95	<a href="https://twitter.com/dog_rates/status/868880397...">https://twitter.com/dog_rates/status/868880397...</a>	14
97	<a href="https://www.gofundme.com/help-my-baby-sierra-g...">https://www.gofundme.com/help-my-baby-sierra-g...</a>	12
101	<a href="https://twitter.com/loganamnosis/status/872657...">https://twitter.com/loganamnosis/status/872657...</a>	14
109	<a href="https://twitter.com/dog_rates/status/841077006...">https://twitter.com/dog_rates/status/841077006...</a>	12
118	<a href="https://twitter.com/dog_rates/status/859196978...">https://twitter.com/dog_rates/status/859196978...</a>	12
124	<a href="https://www.gofundme.com/3ti3nps">https://www.gofundme.com/3ti3nps</a> , <a href="https://twitt...">https://twitt...</a>	12
130	<a href="https://twitter.com/rachaeleasler/status/86501...">https://twitter.com/rachaeleasler/status/86501...</a>	13
132	<a href="https://twitter.com/dog_rates/status/866450705...">https://twitter.com/dog_rates/status/866450705...</a>	13
137	<a href="https://twitter.com/dog_rates/status/837820167...">https://twitter.com/dog_rates/status/837820167...</a>	12
146	<a href="https://www.gofundme.com/helpquinny">https://www.gofundme.com/helpquinny</a> , <a href="https://tw...">https://tw...</a>	13
155	<a href="https://twitter.com/dog_rates/status/806629075...">https://twitter.com/dog_rates/status/806629075...</a>	13
159	<a href="https://www.gofundme.com/help-lorenzo-beat-can...">https://www.gofundme.com/help-lorenzo-beat-can...</a>	13
160	<a href="https://twitter.com/tallylott/status/860914485...">https://twitter.com/tallylott/status/860914485...</a>	13
165	<a href="https://twitter.com/dog_rates/status/761672994...">https://twitter.com/dog_rates/status/761672994...</a>	10
171	<a href="https://twitter.com/dog_rates/status/839549326...">https://twitter.com/dog_rates/status/839549326...</a>	12
180	<a href="https://twitter.com/AaronChewning/status/85706...">https://twitter.com/AaronChewning/status/85706...</a>	13
182	<a href="https://twitter.com/dog_rates/status/844704788...">https://twitter.com/dog_rates/status/844704788...</a>	13
185	NaN	14
194	<a href="https://twitter.com/dog_rates/status/842163532...">https://twitter.com/dog_rates/status/842163532...</a>	12
195	<a href="https://twitter.com/frasercampbell_/status/855...">https://twitter.com/frasercampbell_/status/855...</a>	14
204	<a href="http://www.gofundme.com/bluethewhitehusky">http://www.gofundme.com/bluethewhitehusky</a> , <a href="http...">http...</a>	13
211	<a href="https://twitter.com/dog_rates/status/829374341...">https://twitter.com/dog_rates/status/829374341...</a>	13
...	...	...
784	<a href="https://twitter.com/dog_rates/status/740373189...">https://twitter.com/dog_rates/status/740373189...</a>	9
794	<a href="https://twitter.com/dog_rates/status/771380798...">https://twitter.com/dog_rates/status/771380798...</a>	11
800	<a href="https://twitter.com/dog_rates/status/765222098...">https://twitter.com/dog_rates/status/765222098...</a>	10
811	<a href="https://twitter.com/dog_rates/status/673320132...">https://twitter.com/dog_rates/status/673320132...</a>	11
815	<a href="https://twitter.com/katieornah/status/77100213...">https://twitter.com/katieornah/status/77100213...</a>	12
818	<a href="https://twitter.com/dog_rates/status/739238157...">https://twitter.com/dog_rates/status/739238157...</a>	13
822	<a href="https://twitter.com/dog_rates/status/741067306...">https://twitter.com/dog_rates/status/741067306...</a>	12
826	<a href="https://vine.co/v/iXQAm5Lgrh">https://vine.co/v/iXQAm5Lgrh</a> , <a href="https://vine.co/...">https://vine.co/...</a>	13
829	<a href="https://twitter.com/dog_rates/status/700143752...">https://twitter.com/dog_rates/status/700143752...</a>	10
833	<a href="https://twitter.com/dog_rates/status/739979191...">https://twitter.com/dog_rates/status/739979191...</a>	12
841	<a href="https://twitter.com/dog_rates/status/759923798...">https://twitter.com/dog_rates/status/759923798...</a>	10
847	<a href="https://twitter.com/dog_rates/status/725842289...">https://twitter.com/dog_rates/status/725842289...</a>	12
860	<a href="https://twitter.com/dog_rates/status/673295268...">https://twitter.com/dog_rates/status/673295268...</a>	8
868	<a href="https://twitter.com/dog_rates/status/685325112...">https://twitter.com/dog_rates/status/685325112...</a>	10
872	<a href="https://twitter.com/dog_rates/status/711694788...">https://twitter.com/dog_rates/status/711694788...</a>	13
885	<a href="https://weratedogs.com/pages/about-us">https://weratedogs.com/pages/about-us</a> , <a href="https://...">https://...</a>	11
890	<a href="https://twitter.com/dog_rates/status/739544079...">https://twitter.com/dog_rates/status/739544079...</a>	10
895	<a href="https://twitter.com/dog_rates/status/670319130...">https://twitter.com/dog_rates/status/670319130...</a>	11
908	<a href="https://twitter.com/dog_rates/status/679062614...">https://twitter.com/dog_rates/status/679062614...</a>	11
911	<a href="https://twitter.com/jon_hill987/status/7575971...">https://twitter.com/jon_hill987/status/7575971...</a>	11
926	<a href="https://twitter.com/dog_rates/status/679158373...">https://twitter.com/dog_rates/status/679158373...</a>	11
937	<a href="https://twitter.com/dog_rates/status/681523177...">https://twitter.com/dog_rates/status/681523177...</a>	12
943	<a href="https://vine.co/v/ibvnzrauFuV">https://vine.co/v/ibvnzrauFuV</a> , <a href="https://vine.co/...">https://vine.co/...</a>	11

949	<a href="https://twitter.com/dog_rates/status/675354435...">https://twitter.com/dog_rates/status/675354435...</a>	13
1012	<a href="https://twitter.com/dog_rates/status/704761120...">https://twitter.com/dog_rates/status/704761120...</a>	13
1023	<a href="https://twitter.com/dog_rates/status/667866724...">https://twitter.com/dog_rates/status/667866724...</a>	10
1043	<a href="https://twitter.com/dog_rates/status/667138269...">https://twitter.com/dog_rates/status/667138269...</a>	10
1242	<a href="https://twitter.com/twitter/status/71199827977...">https://twitter.com/twitter/status/71199827977...</a>	12
2259	<a href="https://twitter.com/dogratingrating/status/667...">https://twitter.com/dogratingrating/status/667...</a>	12
2260	<a href="https://twitter.com/dogratingrating/status/667...">https://twitter.com/dogratingrating/status/667...</a>	5

	rating_denominator	name	doggo	floofer	pupper	puppo
19	10	Canela	None	None	None	None
32	10	None	None	None	None	None
36	10	Lilly	None	None	None	None
68	10	Emmy	None	None	None	None
73	10	Shadow	None	None	None	None
74	10	Terrance	None	None	None	None
78	10	None	None	None	pupper	None
91	10	Coco	None	None	None	None
95	10	Walter	None	None	None	None
97	10	Sierra	None	None	pupper	None
101	10	None	None	None	None	None
109	10	Dawn	None	None	None	None
118	10	quite	None	None	None	None
124	10	Cooper	None	None	None	None
130	10	None	None	None	None	None
132	10	Jamesy	None	None	pupper	None
137	10	None	None	None	pupper	None
146	10	Quinn	None	None	None	None
155	10	None	None	None	None	None
159	10	Lorenzo	None	None	None	None
160	10	None	None	None	None	None
165	10	None	None	None	None	None
171	10	Winston	None	None	None	None
180	10	None	None	None	None	None
182	10	Luna	None	None	None	None
185	10	None	None	None	None	None
194	10	George	None	None	None	None
195	10	None	None	None	None	None
204	10	None	None	None	None	None
211	10	Astrid	doggo	None	None	None
...	...	...	...	...	...	...
784	11	None	None	None	None	None
794	10	Fizz	None	None	None	None
800	10	Gromit	None	None	None	None
811	10	Frankie	None	None	None	None
815	10	None	None	None	pupper	None
818	10	None	doggo	None	None	None
822	10	just	doggo	None	pupper	None
826	10	None	None	None	None	None

829	10	None	None	None	pupper	None
833	10	Nollie	None	None	None	None
841	10	None	None	None	None	None
847	10	Colby	None	None	None	None
860	10	Eve	None	None	pupper	None
868	10	None	None	None	None	None
872	10	None	None	None	None	None
885	10	None	None	None	None	None
890	10	None	None	None	None	None
895	10	None	None	None	None	None
908	10	Chompsky	None	None	None	None
911	10	None	None	None	pupper	None
926	10	Rubio	None	None	None	None
937	10	Carly	None	None	None	None
943	10	None	None	None	None	None
949	10	None	None	None	None	None
1012	10	None	None	None	pupper	None
1023	10	Shaggy	None	None	None	None
1043	10	None	None	None	None	None
1242	10	None	None	None	None	None
2259	10	None	None	None	None	None
2260	10	None	None	None	None	None

[181 rows x 17 columns]

```
In [17]: # Valores que não foram verificados com machine learning
temp = pd.merge(twitter_archive, image_predictions,
                on=['tweet_id'], how='outer')
temp[temp.p1.isnull()]
```

```
Out[17]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
30	886267009285017600	8.862664e+17	2.281182e+09	
32	886054160059072513	NaN	NaN	
35	885518971528720385	NaN	NaN	
42	884247878851493888	NaN	NaN	
55	881633300179243008	8.816070e+17	4.738443e+07	
64	879674319642796034	8.795538e+17	3.105441e+09	
68	879130579576475649	NaN	NaN	
72	878604707211726852	NaN	NaN	
73	878404777348136964	NaN	NaN	
74	878316110768087041	NaN	NaN	
83	876537666061221889	NaN	NaN	
88	875097192612077568	NaN	NaN	
91	874434818259525634	NaN	NaN	
97	873337748698140672	NaN	NaN	
101	872668790621863937	NaN	NaN	
109	871166179821445120	NaN	NaN	
110	871102520638267392	NaN	NaN	

113	870726314365509632	8.707262e+17	1.648776e+07
118	869988702071779329	NaN	NaN
124	868639477480148993	NaN	NaN
132	866816280283807744	NaN	NaN
133	866720684873056260	NaN	NaN
137	866094527597207552	NaN	NaN
146	863471782782697472	NaN	NaN
148	863427515083354112	8.634256e+17	7.759620e+07
159	860981674716409858	NaN	NaN
165	860177593139703809	NaN	NaN
171	858860390427611136	NaN	NaN
179	857214891891077121	8.571567e+17	1.806710e+08
180	857062103051644929	NaN	NaN
...	...	...	...
1625	684830982659280897	NaN	NaN
1628	684588130326986752	NaN	NaN
1640	684147889187209216	NaN	NaN
1650	683515932363329536	NaN	NaN
1663	682808988178739200	6.827884e+17	4.196984e+09
1676	682088079302213632	NaN	NaN
1689	681340665377193984	6.813394e+17	4.196984e+09
1706	680805554198020098	NaN	NaN
1728	679872969355714560	NaN	NaN
1743	679405845277462528	NaN	NaN
1750	679001094530465792	NaN	NaN
1760	678708137298427904	NaN	NaN
1774	678023323247357953	6.780211e+17	4.196984e+09
1776	677961670166224897	NaN	NaN
1791	677335745548390400	NaN	NaN
1807	676916996760600576	NaN	NaN
1818	676593408224403456	NaN	NaN
1819	676590572941893632	6.765883e+17	4.196984e+09
1834	676121918416756736	NaN	NaN
1844	675849018447167488	6.758457e+17	4.196984e+09
1895	674742531037511680	6.747400e+17	4.196984e+09
1905	674606911342424069	6.744689e+17	4.196984e+09
1914	674330906434379776	6.658147e+17	1.637468e+07
1916	674307341513269249	NaN	NaN
1940	673716320723169284	6.737159e+17	4.196984e+09
2038	671550332464455680	6.715449e+17	4.196984e+09
2149	669684865554620416	6.693544e+17	4.196984e+09
2189	668967877119254528	6.689207e+17	2.143566e+07
2212	668587383441514497	NaN	NaN
2298	667070482143944705	6.670655e+17	4.196984e+09

	timestamp \
30	2017-07-15 16:51:35 +0000
32	2017-07-15 02:45:48 +0000

35	2017-07-13	15:19:09	+0000
42	2017-07-10	03:08:17	+0000
55	2017-07-02	21:58:53	+0000
64	2017-06-27	12:14:36	+0000
68	2017-06-26	00:13:58	+0000
72	2017-06-24	13:24:20	+0000
73	2017-06-24	00:09:53	+0000
74	2017-06-23	18:17:33	+0000
83	2017-06-18	20:30:39	+0000
88	2017-06-14	21:06:43	+0000
91	2017-06-13	01:14:41	+0000
97	2017-06-10	00:35:19	+0000
101	2017-06-08	04:17:07	+0000
109	2017-06-04	00:46:17	+0000
110	2017-06-03	20:33:19	+0000
113	2017-06-02	19:38:25	+0000
118	2017-05-31	18:47:24	+0000
124	2017-05-28	01:26:04	+0000
132	2017-05-23	00:41:20	+0000
133	2017-05-22	18:21:28	+0000
137	2017-05-21	00:53:21	+0000
146	2017-05-13	19:11:30	+0000
148	2017-05-13	16:15:35	+0000
159	2017-05-06	22:16:42	+0000
165	2017-05-04	17:01:34	+0000
171	2017-05-01	01:47:28	+0000
179	2017-04-26	12:48:51	+0000
180	2017-04-26	02:41:43	+0000
...			...
1625	2016-01-06	20:16:44	+0000
1628	2016-01-06	04:11:43	+0000
1640	2016-01-04	23:02:22	+0000
1650	2016-01-03	05:11:12	+0000
1663	2016-01-01	06:22:03	+0000
1676	2015-12-30	06:37:25	+0000
1689	2015-12-28	05:07:27	+0000
1706	2015-12-26	17:41:07	+0000
1728	2015-12-24	03:55:21	+0000
1743	2015-12-22	20:59:10	+0000
1750	2015-12-21	18:10:50	+0000
1760	2015-12-20	22:46:44	+0000
1774	2015-12-19	01:25:31	+0000
1776	2015-12-18	21:20:32	+0000
1791	2015-12-17	03:53:20	+0000
1807	2015-12-16	00:09:23	+0000
1818	2015-12-15	02:43:33	+0000
1819	2015-12-15	02:32:17	+0000
1834	2015-12-13	19:30:01	+0000

1844 2015-12-13 01:25:37 +0000  
 1895 2015-12-10 00:08:50 +0000  
 1905 2015-12-09 15:09:55 +0000  
 1914 2015-12-08 20:53:11 +0000  
 1916 2015-12-08 19:19:32 +0000  
 1940 2015-12-07 04:11:02 +0000  
 2038 2015-12-01 04:44:10 +0000  
 2149 2015-11-26 01:11:28 +0000  
 2189 2015-11-24 01:42:25 +0000  
 2212 2015-11-23 00:30:28 +0000  
 2298 2015-11-18 20:02:51 +0000

```

                                     source \
30  <a href="http://twitter.com/download/iphone" r...
32  <a href="http://twitter.com/download/iphone" r...
35  <a href="http://twitter.com/download/iphone" r...
42  <a href="http://twitter.com/download/iphone" r...
55  <a href="http://twitter.com/download/iphone" r...
64  <a href="http://twitter.com/download/iphone" r...
68  <a href="http://twitter.com/download/iphone" r...
72  <a href="http://twitter.com/download/iphone" r...
73  <a href="http://twitter.com/download/iphone" r...
74  <a href="http://twitter.com/download/iphone" r...
83  <a href="http://twitter.com/download/iphone" r...
88  <a href="http://twitter.com/download/iphone" r...
91  <a href="http://twitter.com/download/iphone" r...
97  <a href="http://twitter.com/download/iphone" r...
101 <a href="http://twitter.com/download/iphone" r...
109 <a href="http://twitter.com/download/iphone" r...
110 <a href="http://twitter.com/download/iphone" r...
113 <a href="http://twitter.com/download/iphone" r...
118 <a href="http://twitter.com/download/iphone" r...
124 <a href="http://twitter.com/download/iphone" r...
132 <a href="http://twitter.com/download/iphone" r...
133 <a href="http://twitter.com/download/iphone" r...
137 <a href="http://twitter.com/download/iphone" r...
146 <a href="http://twitter.com/download/iphone" r...
148 <a href="http://twitter.com/download/iphone" r...
159 <a href="http://twitter.com/download/iphone" r...
165 <a href="http://twitter.com/download/iphone" r...
171 <a href="http://twitter.com/download/iphone" r...
179 <a href="http://twitter.com/download/iphone" r...
180 <a href="http://twitter.com/download/iphone" r...
...
1625 <a href="http://vine.co" rel="nofollow">Vine -...
1628 <a href="http://vine.co" rel="nofollow">Vine -...
1640 <a href="http://vine.co" rel="nofollow">Vine -...
1650 <a href="http://vine.co" rel="nofollow">Vine -...
  
```

1663 <a href="http://twitter.com/download/iphone" r...  
 1676 <a href="http://vine.co" rel="nofollow">Vine -...  
 1689 <a href="http://twitter.com/download/iphone" r...  
 1706 <a href="http://vine.co" rel="nofollow">Vine -...  
 1728 <a href="http://vine.co" rel="nofollow">Vine -...  
 1743 <a href="http://vine.co" rel="nofollow">Vine -...  
 1750 <a href="http://vine.co" rel="nofollow">Vine -...  
 1760 <a href="http://vine.co" rel="nofollow">Vine -...  
 1774 <a href="http://twitter.com/download/iphone" r...  
 1776 <a href="http://vine.co" rel="nofollow">Vine -...  
 1791 <a href="http://vine.co" rel="nofollow">Vine -...  
 1807 <a href="http://vine.co" rel="nofollow">Vine -...  
 1818 <a href="http://vine.co" rel="nofollow">Vine -...  
 1819 <a href="http://twitter.com/download/iphone" r...  
 1834 <a href="http://vine.co" rel="nofollow">Vine -...  
 1844 <a href="http://twitter.com/download/iphone" r...  
 1895 <a href="http://twitter.com/download/iphone" r...  
 1905 <a href="http://twitter.com/download/iphone" r...  
 1914 <a href="http://twitter.com/download/iphone" r...  
 1916 <a href="http://vine.co" rel="nofollow">Vine -...  
 1940 <a href="http://twitter.com/download/iphone" r...  
 2038 <a href="http://twitter.com/download/iphone" r...  
 2149 <a href="http://twitter.com/download/iphone" r...  
 2189 <a href="http://twitter.com/download/iphone" r...  
 2212 <a href="http://vine.co" rel="nofollow">Vine -...  
 2298 <a href="http://twitter.com/download/iphone" r...

	text	retweeted_status_id \
30	@NonWhiteHat @MayhewMayhem omg hello tanner yo...	NaN
32	RT @Athletics: 12/10 #BATP https://t.co/WxwJmv...	8.860537e+17
35	I have a new hero and his name is Howard. 14/1...	NaN
42	OMG HE DIDN'T MEAN TO HE WAS JUST TRYING A LIT...	NaN
55	@roushfenway These are good dogs but 17/10 is ...	NaN
64	@RealKentMurphy 14/10 confirmed	NaN
68	RT @dog_rates: This is Emmy. She was adopted t...	8.780576e+17
72	Martha is stunning how h*ckin dare you. 13/10 ...	NaN
73	RT @dog_rates: Meet Shadow. In an attempt to r...	8.782815e+17
74	RT @dog_rates: Meet Terrance. He's being yelle...	6.690004e+17
83	I can say with the pupmost confidence that the...	NaN
88	You'll get your package when that precious man...	NaN
91	RT @dog_rates: This is Coco. At first I though...	8.663350e+17
97	RT @dog_rates: This is Sierra. She's one preci...	8.732138e+17
101	RT @loganamnosis: Penelope here is doing me qu...	8.726576e+17
109	RT @dog_rates: This is Dawn. She's just checki...	8.410770e+17
110	Never doubt a doggo 14/10 https://t.co/AbBLh2FZCH	NaN
113	@ComplicitOwl @ShopWeRateDogs &gt;10/10 is res...	NaN
118	RT @dog_rates: We only rate dogs. This is quit...	8.591970e+17
124	RT @dog_rates: Say hello to Cooper. His expres...	8.685523e+17



132	RT @dog_rates: This is Jamesy. He gives a kiss...	8.664507e+17
133	He was providing for his family 13/10 how dare...	NaN
137	RT @dog_rates: Here's a pupper before and afte...	8.378202e+17
146	RT @dog_rates: Say hello to Quinn. She's quite...	8.630625e+17
148	@Jack_Septic_Eye I'd need a few more pics to p...	NaN
159	RT @dog_rates: Meet Lorenzo. He's an avid nift...	8.605638e+17
165	RT @dog_rates: Ohboyohboyohboyohboyohboyohboy...	7.616730e+17
171	RT @dog_rates: Meet Winston. He knows he's a l...	8.395493e+17
179	@Marc_IRL pixelated af 12/10	NaN
180	RT @AaronChewning: First time wearing my @dog_...	8.570611e+17
...	...	...
1625	This little fella really hates stairs. Prefers...	NaN
1628	This pupper just got his first kiss. 12/10 he'...	NaN
1640	This is Sweets the English Bulldog. Waves back...	NaN
1650	HEY PUP WHAT'S THE PART OF THE HUMAN BODY THAT...	NaN
1663	I'm aware that I could've said 20/16, but here...	NaN
1676	I'm not sure what this dog is doing but it's p...	NaN
1689	I've been told there's a slight possibility he...	NaN
1706	This guy's dog broke. So sad. 9/10 would still...	NaN
1728	This is Rocco. He's in a very intense game of ...	NaN
1743	Crazy unseen footage from Jurassic Park. 10/10...	NaN
1750	This is Rascal. He's paddling an imaginary can...	NaN
1760	Here we are witnessing a wild field pupper. Lo...	NaN
1774	After getting lost in Reese's eyes for several...	NaN
1776	This is Izzy. She's showing off the dance move...	NaN
1791	Downright inspiring 12/10 <a href="https://t.co/vSLtYBWHcQ">https://t.co/vSLtYBWHcQ</a>	NaN
1807	Super speedy pupper. Does not go gentle into t...	NaN
1818	This pupper loves leaves. 11/10 for committed ...	NaN
1819	After some outrage from the crowd. Bubbles is ...	NaN
1834	Here we are witnessing a very excited dog. Cle...	NaN
1844	This dog is being demoted to a 9/10 for not we...	NaN
1895	Some clarification is required. The dog is sin...	NaN
1905	The 13/10 also takes into account this impecca...	NaN
1914	13/10\n@ABC7	NaN
1916	This is life-changing. 12/10 <a href="https://t.co/SroT...">https://t.co/SroT...</a>	NaN
1940	The millennials have spoken and we've decided ...	NaN
2038	After 22 minutes of careful deliberation this ...	NaN
2149	After countless hours of research and hundreds...	NaN
2189	12/10 good shit Bubka\n@wane15	NaN
2212	Never forget this vine. You will not stop watc...	NaN
2298	After much debate this dog is being upgraded t...	NaN

	retweeted_status_user_id	retweeted_status_timestamp	\
30	NaN	NaN	
32	1.960740e+07	2017-07-15 02:44:07 +0000	
35	NaN	NaN	
42	NaN	NaN	
55	NaN	NaN	

64	NaN	NaN
68	4.196984e+09	2017-06-23 01:10:23 +0000
72	NaN	NaN
73	4.196984e+09	2017-06-23 16:00:04 +0000
74	4.196984e+09	2015-11-24 03:51:38 +0000
83	NaN	NaN
88	NaN	NaN
91	4.196984e+09	2017-05-21 16:48:45 +0000
97	4.196984e+09	2017-06-09 16:22:42 +0000
101	1.547674e+08	2017-06-08 03:32:35 +0000
109	4.196984e+09	2017-03-13 00:02:39 +0000
110	NaN	NaN
113	NaN	NaN
118	4.196984e+09	2017-05-02 00:04:57 +0000
124	4.196984e+09	2017-05-27 19:39:34 +0000
132	4.196984e+09	2017-05-22 00:28:40 +0000
133	NaN	NaN
137	4.196984e+09	2017-03-04 00:21:08 +0000
146	4.196984e+09	2017-05-12 16:05:02 +0000
148	NaN	NaN
159	4.196984e+09	2017-05-05 18:36:06 +0000
165	4.196984e+09	2016-08-05 21:19:27 +0000
171	4.196984e+09	2017-03-08 18:52:12 +0000
179	NaN	NaN
180	5.870972e+07	2017-04-26 02:37:47 +0000
...	...	...
1625	NaN	NaN
1628	NaN	NaN
1640	NaN	NaN
1650	NaN	NaN
1663	NaN	NaN
1676	NaN	NaN
1689	NaN	NaN
1706	NaN	NaN
1728	NaN	NaN
1743	NaN	NaN
1750	NaN	NaN
1760	NaN	NaN
1774	NaN	NaN
1776	NaN	NaN
1791	NaN	NaN
1807	NaN	NaN
1818	NaN	NaN
1819	NaN	NaN
1834	NaN	NaN
1844	NaN	NaN
1895	NaN	NaN
1905	NaN	NaN

1914	NaN	NaN
1916	NaN	NaN
1940	NaN	NaN
2038	NaN	NaN
2149	NaN	NaN
2189	NaN	NaN
2212	NaN	NaN
2298	NaN	NaN

	expanded_urls	...	img_num	p1	\
30	NaN	...	NaN	NaN	
32	https://twitter.com/dog_rates/status/886053434...	...	NaN	NaN	
35	https://twitter.com/4bonds2carbon/status/88551...	...	NaN	NaN	
42	https://twitter.com/kaijohnson_19/status/88396...	...	NaN	NaN	
55	NaN	...	NaN	NaN	
64	NaN	...	NaN	NaN	
68	https://twitter.com/dog_rates/status/878057613...	...	NaN	NaN	
72	https://twitter.com/bbcworld/status/8785998685...	...	NaN	NaN	
73	https://www.gofundme.com/3yd6y1c,https://twitt...	...	NaN	NaN	
74	https://twitter.com/dog_rates/status/669000397...	...	NaN	NaN	
83	https://twitter.com/mpstowerham/status/8761629...	...	NaN	NaN	
88	https://twitter.com/drboondoc/status/874413398...	...	NaN	NaN	
91	https://twitter.com/dog_rates/status/866334964...	...	NaN	NaN	
97	https://www.gofundme.com/help-my-baby-sierra-g...	...	NaN	NaN	
101	https://twitter.com/loganamnosis/status/872657...	...	NaN	NaN	
109	https://twitter.com/dog_rates/status/841077006...	...	NaN	NaN	
110	https://twitter.com/animalcog/status/871075758...	...	NaN	NaN	
113	NaN	...	NaN	NaN	
118	https://twitter.com/dog_rates/status/859196978...	...	NaN	NaN	
124	https://www.gofundme.com/3ti3nps,https://twitt...	...	NaN	NaN	
132	https://twitter.com/dog_rates/status/866450705...	...	NaN	NaN	
133	https://twitter.com/nbcnews/status/86645871888...	...	NaN	NaN	
137	https://twitter.com/dog_rates/status/837820167...	...	NaN	NaN	
146	https://www.gofundme.com/helpquinny,https://tw...	...	NaN	NaN	
148	NaN	...	NaN	NaN	
159	https://www.gofundme.com/help-lorenzo-beat-can...	...	NaN	NaN	
165	https://twitter.com/dog_rates/status/761672994...	...	NaN	NaN	
171	https://twitter.com/dog_rates/status/839549326...	...	NaN	NaN	
179	NaN	...	NaN	NaN	
180	https://twitter.com/AaronChewning/status/85706...	...	NaN	NaN	
...	...	...	...	...	
1625	https://vine.co/v/eEZXZI1rqxX	...	NaN	NaN	
1628	https://vine.co/v/ihWlxntjt07	...	NaN	NaN	
1640	https://vine.co/v/ib2nT0EuuOI	...	NaN	NaN	
1650	https://vine.co/v/ibvnzrauFuV	...	NaN	NaN	
1663	NaN	...	NaN	NaN	
1676	https://vine.co/v/iqMjlxULzbn	...	NaN	NaN	
1689	NaN	...	NaN	NaN	

1706	<a href="https://vine.co/v/iAPOUgzi2P0">https://vine.co/v/iAPOUgzi2P0</a>	...	NaN	NaN
1728	<a href="https://vine.co/v/iAAxTbj1UAM">https://vine.co/v/iAAxTbj1UAM</a>	...	NaN	NaN
1743	<a href="https://vine.co/v/iKVFEigMLxP">https://vine.co/v/iKVFEigMLxP</a>	...	NaN	NaN
1750	<a href="https://vine.co/v/iKIwAzEatd6">https://vine.co/v/iKIwAzEatd6</a>	...	NaN	NaN
1760	<a href="https://vine.co/v/eQjxxYaQ60K">https://vine.co/v/eQjxxYaQ60K</a>	...	NaN	NaN
1774	NaN	...	NaN	NaN
1776	<a href="https://vine.co/v/iKuMDuYV0aZ">https://vine.co/v/iKuMDuYV0aZ</a>	...	NaN	NaN
1791	<a href="https://vine.co/v/hbLbH77Ar67">https://vine.co/v/hbLbH77Ar67</a>	...	NaN	NaN
1807	<a href="https://vine.co/v/imJ0BdZ0JTw">https://vine.co/v/imJ0BdZ0JTw</a>	...	NaN	NaN
1818	<a href="https://vine.co/v/eEQQaPFbg0Y">https://vine.co/v/eEQQaPFbg0Y</a>	...	NaN	NaN
1819	NaN	...	NaN	NaN
1834	<a href="https://vine.co/v/iZXg7VpeDAv">https://vine.co/v/iZXg7VpeDAv</a>	...	NaN	NaN
1844	NaN	...	NaN	NaN
1895	NaN	...	NaN	NaN
1905	NaN	...	NaN	NaN
1914	NaN	...	NaN	NaN
1916	<a href="https://vine.co/v/i7nWzrenw5h">https://vine.co/v/i7nWzrenw5h</a>	...	NaN	NaN
1940	NaN	...	NaN	NaN
2038	NaN	...	NaN	NaN
2149	NaN	...	NaN	NaN
2189	NaN	...	NaN	NaN
2212	<a href="https://vine.co/v/ea00wvPTx9l">https://vine.co/v/ea00wvPTx9l</a>	...	NaN	NaN
2298	NaN	...	NaN	NaN

	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog
30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
32	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
35	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
42	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
55	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
64	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
68	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
72	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
74	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
83	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
88	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
91	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
97	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
101	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
109	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
110	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
113	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
118	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
124	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
132	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
133	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
137	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

146	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
148	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
159	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
165	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
171	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
179	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
180	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
1625	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1628	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1640	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1650	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1663	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1676	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1689	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1706	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1728	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1743	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1750	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1760	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1774	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1776	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1791	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1807	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1818	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1819	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1834	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1844	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1895	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1905	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1914	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1916	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1940	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2038	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2149	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2189	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2212	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2298	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

[281 rows x 28 columns]

```
In [18]: # Verificando existencia de duplicados
         twitter_archive.tweet_id.duplicated().sum()
```

Out[18]: 0

```
In [19]: image_predictions.tweet_id.duplicated().sum()
```

Out[19]: 0

```
In [20]: twitter_archive[twitter_archive.rating_numerator == 1776].text.values
```

```
Out[20]: array(["This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkl",
               dtype=object)
```

```
In [22]: %%time
         # Verificando links de tweets inexistentes
         # AVISO: Execução demorada.
         count = 0
         id_404 = []
         for index, row in twitter_archive[twitter_archive.expanded_urls.notna()].iterrows():
             if requests.get(row.expanded_urls.split(',')[0]).status_code == 404:
                 print(row.expanded_urls.split(',')[0])
                 count += 1
                 id_404.append(row.tweet_id)
         count
```

```
https://twitter.com/4bonds2carbon/status/885517367337512960
https://twitter.com/loganamnosis/status/872657584259551233/photo/1
https://twitter.com/perfy/status/855857318168150016
https://www.petfinder.com/petdetail/37334596
https://twitter.com/brianstack153/status/796796054100471809
https://twitter.com/KennyFromDaBlok/status/837011344666812416/photo/1
https://twitter.com/stevekopack/status/834086676934836224
https://www.petfinder.com/petdetail/34918210
https://twitter.com/dog_rates/status/680055455951884288/photo/1
https://twitter.com/littlewiewel/status/800852955880628224
https://www.patreon.com/WeRateDogs
https://weratedogs.com/pages/about-us
https://twitter.com/dog_rates/status/754011816964026368/photo/1
https://twitter.com/dog_rates/status/680055455951884288/photo/1
CPU times: user 1min 26s, sys: 4.06 s, total: 1min 30s
Wall time: 33min 50s
```

```
In [23]: # Mostra linhas que url retorna erro 404
         twitter_archive[twitter_archive.tweet_id.isin(id_404)]
```

```
Out[23]:
```

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
35	885518971528720385	NaN	NaN	
101	872668790621863937	NaN	NaN	
190	855857698524602368	NaN	NaN	
237	847157206088847362	NaN	NaN	
256	843981021012017153	NaN	NaN	
298	837012587749474308	NaN	NaN	
323	834089966724603904	NaN	NaN	
344	832032802820481025	NaN	NaN	
506	812747805718642688	NaN	NaN	
577	800855607700029440	NaN	NaN	

742	780476555013349377	NaN	NaN
885	760153949710192640	NaN	NaN
932	754011816964026368	NaN	NaN
1726	680055455951884288	NaN	NaN

	timestamp \
35	2017-07-13 15:19:09 +0000
101	2017-06-08 04:17:07 +0000
190	2017-04-22 18:55:51 +0000
237	2017-03-29 18:43:12 +0000
256	2017-03-21 00:22:10 +0000
298	2017-03-01 18:52:06 +0000
323	2017-02-21 17:18:39 +0000
344	2017-02-16 01:04:13 +0000
506	2016-12-24 19:52:31 +0000
577	2016-11-22 00:17:10 +0000
742	2016-09-26 18:38:05 +0000
885	2016-08-01 16:43:19 +0000
932	2016-07-15 17:56:40 +0000
1726	2015-12-24 16:00:30 +0000

	source \
35	<a href="http://twitter.com/download/iphone" r...
101	<a href="http://twitter.com/download/iphone" r...
190	<a href="http://twitter.com/download/iphone" r...
237	<a href="http://twitter.com/download/iphone" r...
256	<a href="http://twitter.com/download/iphone" r...
298	<a href="http://twitter.com/download/iphone" r...
323	<a href="http://twitter.com/download/iphone" r...
344	<a href="http://twitter.com/download/iphone" r...
506	<a href="http://twitter.com/download/iphone" r...
577	<a href="http://twitter.com/download/iphone" r...
742	<a href="http://twitter.com/download/iphone" r...
885	<a href="http://twitter.com/download/iphone" r...
932	<a href="http://twitter.com/download/iphone" r...
1726	<a href="https://about.twitter.com/products/tw...

	text	retweeted_status_id \
35	I have a new hero and his name is Howard. 14/1...	NaN
101	RT @loganamnosis: Penelope here is doing me qu...	8.726576e+17
190	HE'S LIKE "WAIT A MINUTE I'M AN ANIMAL THIS IS...	NaN
237	Meet Daisy. She's been pup for adoption for mo...	NaN
256	HE WAS DOING A SNOOZE NO SHAME IN A SNOOZE 13/...	NaN
298	RT @KennyFromDaBlok: 14/10 h*ckin good hats. w...	8.370113e+17
323	DOGGO ON THE LOOSE I REPEAT DOGGO ON THE LOOSE...	NaN
344	This is Miguel. He was the only remaining dogg...	NaN
506	RT @dog_rates: Meet Sammy. At first I was like...	6.800555e+17
577	RT @Lin_Manuel: 11/10 would recommend. https:/...	8.008540e+17

742	RT @Patreon: Well. @dog_rates is on Patreon. \...	7.804657e+17
885	RT @hownottodraw: The story/person behind @dog...	7.601538e+17
932	This is Charlie. He pouts until he gets to go ...	NaN
1726	Meet Sammy. At first I was like "that's a snow...	NaN

	retweeted_status_user_id	retweeted_status_timestamp	\
35	NaN	NaN	
101	1.547674e+08	2017-06-08 03:32:35 +0000	
190	NaN	NaN	
237	NaN	NaN	
256	NaN	NaN	
298	7.266347e+08	2017-03-01 18:47:10 +0000	
323	NaN	NaN	
344	NaN	NaN	
506	4.196984e+09	2015-12-24 16:00:30 +0000	
577	7.992370e+07	2016-11-22 00:10:52 +0000	
742	1.228326e+09	2016-09-26 17:55:00 +0000	
885	1.950368e+08	2016-08-01 16:42:51 +0000	
932	NaN	NaN	
1726	NaN	NaN	

	expanded_urls	rating_numerator	\
35	https://twitter.com/4bonds2carbon/status/88551...	14	
101	https://twitter.com/loganamnosis/status/872657...	14	
190	https://twitter.com/perfy/status/8558573181681...	13	
237	https://www.petfinder.com/petdetail/37334596,h...	11	
256	https://twitter.com/brianstack153/status/79679...	13	
298	https://twitter.com/KennyFromDaBlok/status/837...	14	
323	https://twitter.com/stevekopack/status/8340866...	10	
344	https://www.petfinder.com/petdetail/34918210,h...	12	
506	https://twitter.com/dog_rates/status/680055455...	10	
577	https://twitter.com/littlewiewel/status/800852...	11	
742	https://www.patreon.com/WeRateDogs,https://twi...	12	
885	https://weratedogs.com/pages/about-us,https://...	11	
932	https://twitter.com/dog_rates/status/754011816...	12	
1726	https://twitter.com/dog_rates/status/680055455...	10	

	rating_denominator	name	doggo	floofer	pupper	puppo
35	10	None	None	None	None	None
101	10	None	None	None	None	None
190	10	None	None	None	None	None
237	10	Daisy	None	None	None	None
256	10	None	None	None	None	None
298	10	None	None	None	None	None
323	10	None	doggo	None	None	None
344	10	Miguel	doggo	None	None	None
506	10	Sammy	None	None	None	None
577	10	None	None	None	None	None



742	10	None	None	None	None	None
885	10	None	None	None	None	None
932	10	Charlie	None	None	None	None
1726	10	Sammy	None	None	None	None

In [24]: df\_tweets.sample(10)

```
Out[24]:
```

	tweet_id	text \
745	687664829264453632	Meet Opal. He's a Belgian Dijon Poofster. Upse...
2037	886366144734445568	This is Roscoe. Another pupper fallen victim t...
785	690374419777196032	This is Phred. He's an Albanian Flepperkush. T...
24	666353288456101888	Here we have a mixed Asiago from the Galápagos...
315	671735591348891648	This is Oscar. He's ready for Christmas. 11/10...
853	696713835009417216	This is Trevith. He's a Swiss Mountain Roadwoo...
1638	807621403335917568	This is Ollie Vue. He was a 3 legged pupper on...
1265	749403093750648834	Duuun dun... duuun dun... dunn dun. dunn dun...
2008	879862464715927552	This is Romeo. He would like to do an entrance...
313	671561002136281088	This is the best thing I've ever seen so sprea...

	retweet_count	favorite_count \
745	546	2039
2037	3139	20836
785	919	3432
24	71	216
315	778	1476
853	689	2494
1638	4043	15614
1265	588	2774
2008	3441	21962
313	7442	13052

	jpg_url \
745	<a href="https://pbs.twimg.com/media/CYsTg1XUsAEPjxE.jpg">https://pbs.twimg.com/media/CYsTg1XUsAEPjxE.jpg</a>
2037	<a href="https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg">https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg</a>
785	<a href="https://pbs.twimg.com/media/CZSsz3vWXEAACElU.jpg">https://pbs.twimg.com/media/CZSsz3vWXEAACElU.jpg</a>
24	<a href="https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg">https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg</a>
315	<a href="https://pbs.twimg.com/media/CVJ79MxWEAEIWVi.jpg">https://pbs.twimg.com/media/CVJ79MxWEAEIWVi.jpg</a>
853	<a href="https://pbs.twimg.com/media/Cas5h-wWcAA3nAc.jpg">https://pbs.twimg.com/media/Cas5h-wWcAA3nAc.jpg</a>
1638	<a href="https://pbs.twimg.com/media/CzU_YVGUoAA1Bla.jpg">https://pbs.twimg.com/media/CzU_YVGUoAA1Bla.jpg</a>
1265	<a href="https://pbs.twimg.com/media/CmZqIslWIAQFique.jpg">https://pbs.twimg.com/media/CmZqIslWIAQFique.jpg</a>
2008	<a href="https://pbs.twimg.com/media/DDXmPreXUAA3QGJ.jpg">https://pbs.twimg.com/media/DDXmPreXUAA3QGJ.jpg</a>
313	<a href="https://pbs.twimg.com/media/CVHdK-7WwAAsuyc.jpg">https://pbs.twimg.com/media/CVHdK-7WwAAsuyc.jpg</a>

	url \
745	<a href="https://t.co/IAt3jRZ5ez">https://t.co/IAt3jRZ5ez</a>
2037	<a href="https://t.co/RGE08MIJox">https://t.co/RGE08MIJox</a>
785	<a href="https://t.co/VpfFCKE28C">https://t.co/VpfFCKE28C</a>
24	<a href="https://t.co/tltQ5w9aU0">https://t.co/tltQ5w9aU0</a>

```

315 https://t.co/TON0Irzgwr
853 https://t.co/6J8Ibwy1X6
1638 https://t.co/qTRY2qX9y4
1265 https://t.co/9qdJ2Q1Cwx
2008 https://t.co/Qh5aEkRQm9
313 https://t.co/q6Rsu0VYwU

                                expanded_urls
745 https://twitter.com/dog_rates/status/687664829...
2037 https://twitter.com/dog_rates/status/886366144...
785 https://twitter.com/dog_rates/status/690374419...
24 https://twitter.com/dog_rates/status/666353288...
315 https://twitter.com/dog_rates/status/671735591...
853 https://twitter.com/dog_rates/status/696713835...
1638 https://twitter.com/dog_rates/status/807621403...
1265 https://twitter.com/dog_rates/status/749403093...
2008 https://twitter.com/dog_rates/status/879862464...
313 https://twitter.com/dog_rates/status/671561002...

```

```

In [25]: # valores duplicados
df_tweets.tweet_id.value_counts().head(10)

```

```

Out[25]: 729463711119904772    1
          809808892968534016    1
          831309418084069378    1
          683498322573824003    1
          775898661951791106    1
          826476773533745153    1
          670433248821026816    1
          852226086759018497    1
          743210557239623680    1
          678969228704284672    1
          Name: tweet_id, dtype: int64

```

## 2.2.1 Quality

### image\_predictions table

- Tipo de dados errado em tweet\_id;
- Textos com diferentes formas minusculos e maiusculos nas colunas p1, p2 e p3;
- Separação dos textos das colunas p1, p2 e p3 usando \_ ao invés de espaço.

### twitter\_archive table

- in\_reply\_to\_status\_id, in\_reply\_to\_user\_id, retweeted\_status\_id, retweeted\_status\_user\_id com valores incorres no formato 6.67152e+17;

- Tipo de dados errado em `tweet_id`, `in_reply_to_status_id`, `in_reply_to_user_id`, `timestamp`, `retweeted_status_id`, `retweeted_status_user_id`, `retweeted_status_timestamp`;
- Valores nulos sendo representados como `nan` ao invés de `NaN` nas colunas `in_reply_to_user_id`, `retweeted_status_id`, `retweeted_status_id` e `retweeted_status_user_id`;
- Valores nulos sendo representados como `None` em `doggo`, `floofer`, `pupper` e `puppo`;
- Url repetidas na coluna `expanded_urls`;
- Alguns links em `expanded_urls` levam a páginas com error 404;
- Nem todos os ids tem predições.
- Urls vazias;
- Nomes de pets como `None` e a.

## 2.3 ##### `df_tweets` table

### `twitter_archive_master` table

- Limpar urls que não levam a `dog_rates`

#### 2.3.1 Tidiness

- Colunas `source` desnecessária em `twitter_archive`;
- Retweets existentes, colunas `retweeted_status_id`, `retweeted_status_user_id` e `retweeted_status_timestamp`.
- Replies existentes, colunas `in_reply_to_status_id`, `in_reply_to_user_id`;
- `text` contém (texto, nota, url) em `twitter_archive` e `df_tweets`;
- Quatro colunas em `twitter_archive` (`doggo`, `floofer`, `pupper` e `puppo`) ao invés de uma;
- Denominador se mostra desnecessário sendo sempre 10 em `twitter_archive` na coluna `rating_denominator`;
- Colunas `in_reply_to_status_id`, `in_reply_to_user_id`, `retweeted_status_id`, `retweeted_status_user_id` e `retweeted_status_timestamp`;
- Numero de imagens na coluna `img_num` se mostra desnecessário em `image_predictions`.
- Junção de `df_tweets`, `image_predictions` e `twitter_archive`;

## 2.4 Clean

```
In [26]: image_predictions_clean = image_predictions.copy()
         twitter_archive_clean = twitter_archive.copy()
         df_tweets_clean = df_tweets.copy()
```

### 2.4.1 Clean - image\_predictions table

#### Define - Tipo de dados errado em tweet\_id

- Modifica o tipos dos valores da coluna tweet\_id de inteiro para string usando astype

#### Code

```
In [27]: image_predictions_clean.tweet_id = image_predictions_clean.tweet_id.astype('str')
```

#### Test

```
In [28]: image_predictions_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id      2075 non-null object
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(1), object(5)
memory usage: 152.1+ KB
```

#### Define

- Textos com diferentes formas minusculos e maiusculos nas colunas p1, p2 e p3

#### Code

```
In [29]: image_predictions_clean.p1 = image_predictions_clean.p1.apply(
        lambda x: x.lower().replace('_', ' '))
image_predictions_clean.p2 = image_predictions_clean.p2.apply(
        lambda x: x.lower().replace('_', ' '))
image_predictions_clean.p3 = image_predictions_clean.p3.apply(
        lambda x: x.lower().replace('_', ' '))
```

## Test

```
In [30]: image_predictions_clean.p1.sample(10)
```

```
Out[30]: 479          borzoi
1685    miniature poodle
435      toilet tissue
259      chihuahua
1305      bluetick
1322      geyser
364      tusker
246      wool
486      vizsla
1488      briard
Name: p1, dtype: object
```

```
In [31]: image_predictions_clean.p2.sample(10)
```

```
Out[31]: 1498    golden retriever
264          beagle
1439          basenji
757    blenheim spaniel
85    labrador retriever
247          orange
78          drake
2053          redbone
335          crate
1097    lawn mower
Name: p2, dtype: object
```

```
In [32]: image_predictions_clean.p3.sample(10)
```

```
Out[32]: 1884          doormat
2036          chihuahua
854    giant schnauzer
1858          eskimo dog
992          kelpie
394          shovel
1905          file
2024    golden retriever
733          bagel
1720          doberman
Name: p3, dtype: object
```

### 2.4.2 Clean - twitter\_archive table

**Define - Valores nulos sendo representados como None em doggo, floofer, pupper e puppo**

- Modifica valores None para np.nan nas colunas doggo, floofer, pupper e puppo

## Code

```
In [33]: twitter_archive_clean.doggo = twitter_archive_clean.doggo.apply(
        lambda x: np.nan if x == 'None' else x)
twitter_archive_clean.floofer = twitter_archive_clean.floofer.apply(
        lambda x: np.nan if x == 'None' else x)
twitter_archive_clean.pupper = twitter_archive_clean.pupper.apply(
        lambda x: np.nan if x == 'None' else x)
twitter_archive_clean.puppo = twitter_archive_clean.puppo.apply(
        lambda x: np.nan if x == 'None' else x)
```

## Test

```
In [34]: twitter_archive_clean.doggo.sample(10)
```

```
Out[34]: 2338    NaN
        1349    NaN
        301    NaN
        2069    NaN
        1008    NaN
        327    NaN
        397    NaN
        594    NaN
        148    NaN
        610    NaN
        Name: doggo, dtype: object
```

```
In [35]: twitter_archive_clean.floofer.sample(10)
```

```
Out[35]: 1810    NaN
        1366    NaN
        241    NaN
        366    NaN
        568    NaN
        2345    NaN
        592    NaN
        1699    NaN
        2076    NaN
        1960    NaN
        Name: floofer, dtype: object
```

```
In [36]: twitter_archive_clean.pupper.sample(10)
```

```
Out[36]: 1316    NaN
        1558    NaN
        926    NaN
        966    pupper
        992    NaN
        38     NaN
```

```

89      NaN
390      NaN
1253     NaN
1620     NaN
Name: pupper, dtype: object

```

```
In [37]: twitter_archive_clean.puppo.sample(10)
```

```

Out[37]: 421      NaN
1762     NaN
336      NaN
1995     NaN
1788     NaN
649      NaN
1356     NaN
1538     NaN
676      NaN
2220     NaN
Name: puppo, dtype: object

```

**Define - Tipo de dados errado em tweet\_id, in\_reply\_to\_status\_id, in\_replay\_to\_user\_id, timestamp, retweeted\_status\_id, retweeted\_status\_user\_id, retweeted\_status\_timestamp**

- Converte int para string nas colunas tweet\_id, in\_reply\_to\_status\_id, in\_replay\_to\_user\_id, retweeted\_status\_id e 'retweeted\_status\_user\_id
- Converte string para datetime nas colunas timestamp e retweeted\_status\_timestamp ##### Code

```

In [38]: # Conversão para string
twitter_archive_clean.tweet_id = twitter_archive_clean.tweet_id.astype('str')
twitter_archive_clean.in_reply_to_status_id = twitter_archive_clean.in_reply_to_status_id.astype('str')
twitter_archive_clean.in_reply_to_user_id = twitter_archive_clean.in_reply_to_user_id.astype('str')
twitter_archive_clean.retweeted_status_id = twitter_archive_clean.retweeted_status_id.astype('str')
twitter_archive_clean.retweeted_status_user_id = twitter_archive_clean.retweeted_status_user_id.astype('str')

# Conversão para datetime
twitter_archive_clean.timestamp = pd.to_datetime(
    twitter_archive_clean.timestamp)
twitter_archive_clean.retweeted_status_timestamp = pd.to_datetime(
    twitter_archive_clean.retweeted_status_timestamp)

```

**Test**

```
In [39]: twitter_archive_clean.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                2356 non-null object
in_reply_to_status_id   2356 non-null object
in_reply_to_user_id     2356 non-null object
timestamp               2356 non-null datetime64[ns]
source                 2356 non-null object
text                   2356 non-null object
retweeted_status_id     2356 non-null object
retweeted_status_user_id 2356 non-null object
retweeted_status_timestamp 181 non-null datetime64[ns]
expanded_urls          2297 non-null object
rating_numerator       2356 non-null int64
rating_denominator     2356 non-null int64
name                   2356 non-null object
doggo                  97 non-null object
floofer               10 non-null object
pupper                257 non-null object
puppo                 30 non-null object
dtypes: datetime64[ns](2), int64(2), object(13)
memory usage: 313.0+ KB

```

```
In [40]: twitter_archive_clean.timestamp.sample(10)
```

```

Out[40]: 829      2016-08-25 20:35:48
         2260    2015-11-20 03:51:47
         2298    2015-11-18 20:02:51
         1171    2016-04-14 00:55:25
         751     2016-09-22 20:33:42
         195     2017-04-20 19:16:59
         1959    2015-12-06 18:56:46
         1134    2016-05-06 18:33:34
         1667    2015-12-31 20:39:41
         1734    2015-12-23 18:51:56
         Name: timestamp, dtype: datetime64[ns]

```

### Define - Url repetidas na coluna expanded\_urls

- Linhas onde a url se repete separados por , são divididas e apenas o primeiro valor permanece

### Code

```

In [41]: twitter_archive_clean.expanded_urls = twitter_archive_clean.expanded_urls.apply(
         lambda x: x if x is np.nan else x.split(',')[0])

```



## Test

```
In [42]: twitter_archive_clean.expanded_urls.sample(10).values
```

```
Out[42]: array(['https://twitter.com/dog_rates/status/703382836347330562/photo/1',  
               'https://twitter.com/dog_rates/status/762699858130116608/photo/1',  
               'https://twitter.com/dog_rates/status/802323869084381190/photo/1',  
               nan,  
               'https://twitter.com/dog_rates/status/670803562457407488/photo/1',  
               'https://twitter.com/dog_rates/status/670434127938719744/photo/1',  
               'https://twitter.com/dog_rates/status/696754882863349760/photo/1',  
               'https://twitter.com/dog_rates/status/667443425659232256/photo/1',  
               'https://twitter.com/dog_rates/status/796865951799083009/photo/1',  
               'https://twitter.com/dog_rates/status/669375718304980992/photo/1'],  
              dtype=object)
```

## Define - Alguns links em expanded\_urls levam a páginas com error 404

- Pela lista criada em id\_404 com identificadores que retornaram erro 404, seleciona as linhas e altera o valor para np.nan

## Code

```
In [43]: twitter_archive_clean.loc[twitter_archive_clean.tweet_id.isin(pd.Series(id_404).astype(  
               twitter_archive_clean.loc[twitter_archive_clean.tweet_id.isin(  
               pd.Series(id_404).astype('str')), 'expanded_urls'].apply(lambda x: np.nan)
```

## Test

```
In [44]: twitter_archive_clean.loc[twitter_archive_clean.tweet_id.isin(  
               pd.Series(id_404).astype('str')), 'expanded_urls']
```

```
Out[44]: 35      NaN  
        101      NaN  
        190      NaN  
        237      NaN  
        256      NaN  
        298      NaN  
        323      NaN  
        344      NaN  
        506      NaN  
        577      NaN  
        742      NaN  
        885      NaN  
        932      NaN  
        1726     NaN  
        Name: expanded_urls, dtype: object
```

## Define - Nem todos os ids tem predições

- Remove linhas desnecessárias que não possuem predição, selecionando os ids inexistentes na predição e removendo esses índices.

### Code

```
In [45]: tweets_no_predictions = twitter_archive_clean[~twitter_archive_clean.tweet_id.isin(
        image_predictions_clean.tweet_id)]
        twitter_archive_clean.drop(tweets_no_predictions.index, axis=0, inplace=True)
        # tweets_no_predictions
```

### Test

```
In [46]: twitter_archive_clean[~twitter_archive_clean.tweet_id.isin(
        image_predictions_clean.tweet_id)]
```

Out [46]: Empty DataFrame

Columns: [tweet\_id, in\_reply\_to\_status\_id, in\_reply\_to\_user\_id, timestamp, source, text]  
Index: []

Define - Valores nulos sendo representados como nan ao invés de NaN nas colunas in\_reply\_to\_status\_id, in\_reply\_to\_user\_id, retweeted\_status\_id e retweeted\_status\_user\_id

- Verifica as linhas em cada coluna e caso encontre nan troca por np.nan.

### Code

```
In [47]: twitter_archive_clean.in_reply_to_status_id = twitter_archive_clean.in_reply_to_status_id.
        lambda x: x if x != 'nan' else np.nan)
        twitter_archive_clean.in_reply_to_user_id = twitter_archive_clean.in_reply_to_user_id.
        lambda x: x if x != 'nan' else np.nan)
        twitter_archive_clean.retweeted_status_id = twitter_archive_clean.retweeted_status_id.
        lambda x: x if x != 'nan' else np.nan)
        twitter_archive_clean.retweeted_status_user_id = twitter_archive_clean.retweeted_status_user_id.
        lambda x: x if x != 'nan' else np.nan)
```

### Test

```
In [48]: twitter_archive_clean.sample(5)
```

Out [48]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
1693	681297372102656000		NaN	NaN
988	748977405889503236		NaN	NaN
8	890609185150312448		NaN	NaN
517	810896069567610880		NaN	NaN
1392	700062718104104960		NaN	NaN

	timestamp	source	\
1693	2015-12-28 02:15:26	<a href="http://twitter.com/download/iphone" r...	
988	2016-07-01 20:31:43	<a href="http://twitter.com/download/iphone" r...	
8	2017-07-27 16:25:51	<a href="http://twitter.com/download/iphone" r...	
517	2016-12-19 17:14:23	<a href="http://twitter.com/download/iphone" r...	
1392	2016-02-17 21:02:13	<a href="http://twitter.com/download/iphone" r...	

	text	retweeted_status_id	\
1693	This is actually a lion. We only rate dogs. Fo...	NaN	
988	What jokester sent in a pic without a dog in i...	NaN	
8	This is Zoey. She doesn't want to be one of th...	NaN	
517	This is Hunter. He just found out he needs bra...	NaN	
1392	This is Eazy-E. He's colorful af. Must be rare...	NaN	

	retweeted_status_user_id	retweeted_status_timestamp	\
1693	NaN	NaT	
988	NaN	NaT	
8	NaN	NaT	
517	NaN	NaT	
1392	NaN	NaT	

	expanded_urls	rating_numerator	\
1693	https://twitter.com/dog_rates/status/681297372...	12	
988	https://twitter.com/dog_rates/status/748977405...	10	
8	https://twitter.com/dog_rates/status/890609185...	13	
517	https://twitter.com/dog_rates/status/810896069...	11	
1392	https://twitter.com/dog_rates/status/700062718...	6	

	rating_denominator	name	doggo	floofer	pupper	puppo
1693	10	actually	NaN	NaN	NaN	NaN
988	10	not	NaN	NaN	NaN	NaN
8	10	Zoey	NaN	NaN	NaN	NaN
517	10	Hunter	NaN	NaN	NaN	NaN
1392	10	Eazy	NaN	NaN	NaN	NaN

## Define - Urls vazias.

- Selecciona as linhas onde a url se encontra como np.nan e remove essas linhas pelo id

## Code

```
In [49]: id_list = twitter_archive_clean[(twitter_archive_clean.expanded_urls.isna())]
twitter_archive_clean.drop(id_list.index, axis=0, inplace=True)
```

## Test

```
In [50]: twitter_archive_clean[(twitter_archive_clean.expanded_urls.isna())]
```

```
Out[50]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text]
         Index: []
```

**Define - Nomes de pets como None e a.**

- Selecciona nomes de pets que se encontram com valores None e a e os altera para np.nan

**Code**

```
In [51]: twitter_archive_clean.name = twitter_archive_clean.name.apply(
         lambda x: np.nan if x == 'None' or x == 'a' else x)
```

**Test**

```
In [52]: twitter_archive_clean.query('name == "None" or name == "a"')
```

```
Out[52]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text]
         Index: []
```

```
In [53]: twitter_archive_clean.name.sample(10)
```

```
Out[53]: 306      NaN
         803      NaN
         560    Marley
        1837      NaN
         219    Riley
        1192      NaN
        1176      NaN
         426    Sundance
         526      Ito
        1602    Lorelei
         Name: name, dtype: object
```

### 2.4.3 Tidiness

**Define - Colunas source desnecessária em twitter\_archive**

- Remove a coluna source utilizando drop

**Code**

```
In [54]: twitter_archive_clean.drop(columns='source', inplace=True)
```

## Test

```
In [55]: twitter_archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2069 entries, 0 to 2355
Data columns (total 16 columns):
tweet_id                2069 non-null object
in_reply_to_status_id   23 non-null object
in_reply_to_user_id     23 non-null object
timestamp               2069 non-null datetime64[ns]
text                   2069 non-null object
retweeted_status_id      79 non-null object
retweeted_status_user_id 79 non-null object
retweeted_status_timestamp 79 non-null datetime64[ns]
expanded_urls           2069 non-null object
rating_numerator         2069 non-null int64
rating_denominator       2069 non-null int64
name                    1438 non-null object
doggo                   79 non-null object
floofer                 8 non-null object
pupper                  222 non-null object
puppo                   24 non-null object
dtypes: datetime64[ns](2), int64(2), object(12)
memory usage: 274.8+ KB
```

**Define - Retweets existentes, colunas `retweeted_status_id`, `retweeted_status_user_id` e `retweeted_status_timestamp`**

- Seleciona linhas que possuam identificação de retweets e as remove utilizando drop

## Code

```
In [56]: # Se existe `retweeted_status_id` então existe `retweeted_status_user_id` e `retweeted_status_timestamp`
twitter_archive_clean.drop(
    twitter_archive_clean[twitter_archive_clean.retweeted_status_id.notna()].index, axis=0, inplace=True)
```

## Test

```
In [57]: twitter_archive_clean[twitter_archive_clean.retweeted_status_id.notna()]
```

```
Out[57]: Empty DataFrame
```

```
Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, text, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp, expanded_urls, rating_numerator, rating_denominator, name, doggo, floofer, pupper, puppo]
Index: []
```

**Define - Repls existentes, colunas `in_reply_to_status_id`, `in_reply_to_user_id`.**

- Seleciona linhas que possuam identificação de replys e as remove utilizando drop

## Code

```
In [58]: # Se existe `in_reply_to_status_id` então existe `in_reply_to_user_id`
        twitter_archive_clean.drop(
            twitter_archive_clean[twitter_archive_clean.in_reply_to_status_id.notna()].index,
```

## Test

```
In [59]: twitter_archive_clean[twitter_archive_clean.in_reply_to_status_id.notna()]
```

```
Out [59]: Empty DataFrame
```

```
Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, text, retw
Index: []
```

## Define - text contém (texto, nota, url).

- Limpa a coluna texto removendo a nota e a url utilizando expressão regular, mantendo apenas o texto do tweet

## Code

```
In [60]: twitter_archive_clean.text = twitter_archive_clean.text.str.replace(r'(\d+/\d+)', '')
        r'http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[*\(\),]|(?:%[0-9a-fA-F][0-9a-fA-F]))+'
        df_tweets_clean.text = df_tweets_clean.text.str.replace(r'(\d+/\d+)', '').str.replace(
            r'http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[*\(\),]|(?:%[0-9a-fA-F][0-9a-fA-F]))+'
            r'http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[*\(\),]|(?:%[0-9a-fA-F][0-9a-fA-F]))+')'
```

## Test

```
In [61]: twitter_archive_clean.text.sample(10).values
```

```
Out [61]: array(["This is an Iraqi Speed Kangaroo. It is not a dog. Please only send in dogs. I
               "This is Gromit. He's pupset because there's no need to beware of him. Just war
               'This is Churlie. AKA Fetty Woof. Lost eye saving a school bus full of toddlers
               "This is Bubbles. He's a Yorkshire Piccolo. would snug aggressively",
               "Say hello to Carbon. This is his first time swimming. He's having a h*ckin bla
               "This is DonDon. He's way up but doesn't feel blessed. Rather uncomfortable ac
               "These two pirates crashed their ship and don't know what to do now. Very irre
               "This is Buddy. He's photogenic af. Loves to sexily exit pond. Very striped. C
               'This is Spanky. He was a member of the 2002 USA Winter Olympic speed skating t
               "Meet Saydee. She's a Rochester Ecclesiastical. Jumped off cliff and caught s
dtype=object)
```

```
In [62]: df_tweets_clean.text.sample(10).values
```

```
Out [62]: array(["This is Carl. He just wants to make sure you're having a good day. just a sw
               "This is Beckham. He fell asleep at the wheel. Very churlish. Looks to have a l
               'This is Titan. His nose is quite chilly. Requests to return to the indoors. v
               "This is Chubbs. He dug a hole and now he's stuck in it. Dang h*ckin doggo. w
               'This is Hobbes. He was told he was going to the park. Ended up at the vet. H*
```

```

        "Say hello to Maggie. She's a Western Septic Downy. Pretends to be Mexican. Gr
        'This is Scooter. His lack of opposable thumbs is rendering his resistance to
        'This is Bungalo. She uses that face to get what she wants. It works unbeliev
        "This is Scruffers. He's being violated on multiple levels and is not happy ab
        "This is Louis. He's crossing. It's a big deal. h*ckin breathtaking"],
        dtype=object)

```

**Define - Quatro colunas em twitter\_archive (doggo, floofer, pupper e puppo) ao invés de uma.**

- Gera uma nova colunas pet\_class e adiciona os valores doggo, floofer, pupper e puppo, ou a junção desses em caso de mais de uma classificação.
- Remove as colunas doggo, floofer, pupper e puppo desnecessarias

## Code

```

In [63]: twitter_archive_clean['pet_class'] = twitter_archive_clean[[
        'doggo', 'floofer', 'pupper', 'puppo']].fillna('').sum(axis=1)
        twitter_archive_clean.pet_class = twitter_archive_clean.pet_class.apply(
            lambda x: np.nan if x is '' else x)
        twitter_archive_clean.drop(
            columns=['doggo', 'floofer', 'pupper', 'puppo'], inplace=True)

```

## Test

```

In [64]: twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1967 entries, 0 to 2355
Data columns (total 13 columns):
tweet_id                1967 non-null object
in_reply_to_status_id   0 non-null object
in_reply_to_user_id     0 non-null object
timestamp               1967 non-null datetime64[ns]
text                   1967 non-null object
retweeted_status_id     0 non-null object
retweeted_status_user_id 0 non-null object
retweeted_status_timestamp 0 non-null datetime64[ns]
expanded_urls          1967 non-null object
rating_numerator        1967 non-null int64
rating_denominator      1967 non-null int64
name                   1388 non-null object
pet_class              302 non-null object
dtypes: datetime64[ns](2), int64(2), object(9)
memory usage: 215.1+ KB

```

```

In [65]: twitter_archive_clean.sample(10)

```

```

Out [65]:
      tweet_id in_reply_to_status_id in_reply_to_user_id \
1541  689623661272240129                NaN                NaN
710   784826020293709826                NaN                NaN
1450  695816827381944320                NaN                NaN
1719  680161097740095489                NaN                NaN
1753  678800283649069056                NaN                NaN
452   818614493328580609                NaN                NaN
580   800459316964663297                NaN                NaN
1394  700002074055016451                NaN                NaN
2285  667188689915760640                NaN                NaN
1370  702321140488925184                NaN                NaN

      timestamp                                text \
1541 2016-01-20 01:41:08 This is Lucy. She's terrified of the stuffed b...
710  2016-10-08 18:41:19 This is Rusty. He's going D1 for sure. Insane ...
1450 2016-02-06 03:50:33 Here's a dog enjoying a sunset. would trade l...
1719 2015-12-24 23:00:17 Meet Goliath. He's an example of irony. Head i...
1753 2015-12-21 04:52:53 Here's a pupper with some mean tan lines. Snaz...
452  2017-01-10 00:24:38 This is Bear. He's a passionate believer of th...
580  2016-11-20 22:02:27 Here's a very sleepy pupper. Appears to be por...
1394 2016-02-17 17:01:14 This is Thumas. He covered himself in nanners ...
2285 2015-11-19 03:52:34 Quite an advanced dog here. Impressively dress...
1370 2016-02-24 02:36:23 Please enjoy this picture as much as I did.

      retweeted_status_id retweeted_status_user_id retweeted_status_timestamp \
1541                NaN                NaN                NaT
710                NaN                NaN                NaT
1450                NaN                NaN                NaT
1719                NaN                NaN                NaT
1753                NaN                NaN                NaT
452                NaN                NaN                NaT
580                NaN                NaN                NaT
1394                NaN                NaN                NaT
2285                NaN                NaN                NaT
1370                NaN                NaN                NaT

      expanded_urls rating_numerator \
1541 https://twitter.com/dog_rates/status/689623661... 10
710  https://twitter.com/dog_rates/status/784826020... 13
1450 https://twitter.com/dog_rates/status/695816827... 11
1719 https://twitter.com/dog_rates/status/680161097... 12
1753 https://twitter.com/dog_rates/status/678800283... 12
452  https://twitter.com/dog_rates/status/818614493... 12
580  https://twitter.com/dog_rates/status/800459316... 12
1394 https://twitter.com/dog_rates/status/700002074... 9
2285 https://twitter.com/dog_rates/status/667188689... 10
1370 https://twitter.com/dog_rates/status/702321140... 12

```



	rating_denominator	name	pet_class
1541	10	Lucy	pupper
710	10	Rusty	NaN
1450	10	NaN	NaN
1719	10	Goliath	NaN
1753	10	NaN	pupper
452	10	Bear	NaN
580	10	NaN	pupper
1394	10	Thumas	NaN
2285	10	NaN	NaN
1370	10	NaN	NaN

**Define - Denominador se mostra desnecessário sendo sempre 10 em twitter\_archive na coluna rating\_denominator - Colunas in\_reply\_to\_status\_id, in\_reply\_to\_user\_id, retweeted\_status\_id, retweeted\_status\_user\_id e retweeted\_status\_timestamp.**

- Remove as colunas rating\_denominator, in\_reply\_to\_status\_id, in\_reply\_to\_user\_id, retweeted\_status\_id, retweeted\_status\_user\_id e retweeted\_status\_timestamp utilizando drop

## Code

```
In [66]: twitter_archive_clean.drop(columns=['rating_denominator', 'in_reply_to_status_id', 'in_reply_to_user_id', 'retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timestamp'])
```

## Test

```
In [67]: twitter_archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1967 entries, 0 to 2355
Data columns (total 7 columns):
tweet_id      1967 non-null object
timestamp     1967 non-null datetime64[ns]
text          1967 non-null object
expanded_urls 1967 non-null object
rating_numerator 1967 non-null int64
name          1388 non-null object
pet_class     302 non-null object
dtypes: datetime64[ns](1), int64(1), object(5)
memory usage: 122.9+ KB
```

**Define - Numero de imagens na coluna img\_num se mostra desnecessário em image\_predictions.**

- Remove a coluna img\_num utilizando drop

## Code

```
In [68]: image_predictions_clean.drop(columns='img_num', axis=1, inplace=True)
```

## Test

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 11 columns):
tweet_id      2075 non-null object
jpg_url       2075 non-null object
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), object(5)
memory usage: 135.8+ KB
```

## Define - Junção de df\_tweets, image\_predictions e twitter\_archive.

- Gera um merge chamado twitter\_archive\_master com a junção das tabelas twitter\_archive e df\_tweets
- Junta twitter\_archive\_master com a tabela faltante image\_predictions

## Code

```
In [70]: # junta `twitter_archive_clean` com `df_tweets_clean`
twitter_archive_master = twitter_archive_clean.merge(
    df_tweets_clean, how='left', on=['tweet_id', 'text', 'expanded_urls'])

In [71]: # junta `twitter_archive_master` com `image_predictions_clean`
twitter_archive_master = twitter_archive_master.merge(
    image_predictions_clean, how='left', on=['tweet_id', 'jpg_url'])
```

## Test

```
In [72]: twitter_archive_master.head()
```

```
Out[72]:
```

	tweet_id	timestamp	\
0	892420643555336193	2017-08-01 16:23:56	
1	892177421306343426	2017-08-01 00:17:27	

```

2 891815181378084864 2017-07-31 00:18:03
3 891689557279858688 2017-07-30 15:58:51
4 891327558926688256 2017-07-29 16:00:24

```

```

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

```

```

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

```

```

        name pet_class retweet_count favorite_count \
0   Phineas      NaN          8332          38111
1    Tilly      NaN          6156          32699
2   Archie      NaN          4074          24611
3    Darla      NaN          8476          41468
4  Franklin      NaN          9169          39634

```

```

                                jpg_url  url \
0 https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg https://t.co/MgUWQ76dJU
1 https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg https://t.co/0Xxu71qeIV
2 https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg https://t.co/wUnZnhtVJB
3 https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg https://t.co/tD36da7qLQ
4 https://pbs.twimg.com/media/DF6hr6AVYAAZ8G8.jpg https://t.co/AtUZn91f7f

```

```

        p1  p1_conf p1_dog  p2  p2_conf p2_dog \
0   orange 0.097049 False   bagel 0.085851 False
1  chihuahua 0.323581 True   pekinese 0.090647 True
2  chihuahua 0.716012 True   malamute 0.078253 True
3 paper towel 0.170278 False  labrador retriever 0.168086 True
4      NaN      NaN      NaN      NaN      NaN      NaN

```

```

        p3  p3_conf p3_dog
0   banana 0.076110 False
1  papillon 0.068957 True
2   kelpie 0.031379 True
3  spatula 0.040836 False
4      NaN      NaN      NaN

```

```
In [73]: twitter_archive_master.info()
```

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

```

Int64Index: 1967 entries, 0 to 1966
Data columns (total 20 columns):
tweet_id      1967 non-null object
timestamp     1967 non-null datetime64[ns]
text          1967 non-null object
expanded_urls 1967 non-null object
rating_numerator 1967 non-null int64
name          1388 non-null object
pet_class     302 non-null object
retweet_count 1940 non-null object
favorite_count 1940 non-null object
jpg_url       1940 non-null object
url           1940 non-null object
p1            1672 non-null object
p1_conf       1672 non-null float64
p1_dog        1672 non-null object
p2            1672 non-null object
p2_conf       1672 non-null float64
p2_dog        1672 non-null object
p3            1672 non-null object
p3_conf       1672 non-null float64
p3_dog        1672 non-null object
dtypes: datetime64[ns](1), float64(3), int64(1), object(15)
memory usage: 322.7+ KB

```

## 2.4.4 Clean - twitter\_archive\_master table

### Define - Limpar urls que não levam a dog\_rates

- Seleciona urls que não contem `https://twitter.com/dog_rates` e remove as linhas utilizando os índices. ##### Code

```
In [74]: twitter_archive_master_clean = twitter_archive_master.copy()
```

```
In [75]: # Remove por tweets que não pertençam a https://twitter.com/dog_rates
no_tweets_index = twitter_archive_master_clean[~twitter_archive_master_clean.expanded_urls.str.contains(
    'https://twitter.com/dog_rates', regex=False)].index
twitter_archive_master_clean.drop(no_tweets_index, axis=0, inplace=True)
```

### Test

```
In [76]: twitter_archive_master_clean[~twitter_archive_master_clean.expanded_urls.str.contains(
    'https://twitter.com/dog_rates', regex=False)]
```

```
Out[76]: Empty DataFrame
```

```
Columns: [tweet_id, timestamp, text, expanded_urls, rating_numerator, name, pet_class]
Index: []
```

## 2.4.5 Save clean result

```
In [77]: twitter_archive_master_clean.to_csv('twitter_archive_master.csv', index=False)
```

## 2.5 Analyzes

```
In [78]: # carrega o dataframe
df_tweets_archive = pd.read_csv(
    'twitter_archive_master.csv', parse_dates=['timestamp'])
sns.set_style("whitegrid")
```

```
In [79]: df_tweets_archive.head()
```

```
Out[79]:
```

	tweet_id	timestamp	text	expanded_urls	rating_numerator	name	pet_class	retweet_count	favorite_count	jpg_url	url	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog
0	892420643555336193	2017-08-01 16:23:56	This is Phineas. He's a mystical boy. Only eve...	https://twitter.com/dog_rates/status/892420643...	13	Phineas	NaN	8332.0	38111.0	https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg	https://t.co/MgUWQ76dJU						
1	892177421306343426	2017-08-01 00:17:27	This is Tilly. She's just checking pup on you...	https://twitter.com/dog_rates/status/892177421...	13	Tilly	NaN	6156.0	32699.0	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg	https://t.co/0Xxu71qeIV						
2	891815181378084864	2017-07-31 00:18:03	This is Archie. He is a rare Norwegian Pouncin...	https://twitter.com/dog_rates/status/891815181...	12	Archie	NaN	4074.0	24611.0	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg	https://t.co/wUnZnhtVJB						
3	891689557279858688	2017-07-30 15:58:51	This is Darla. She commenced a snooze mid meal...	https://twitter.com/dog_rates/status/891689557...	13	Darla	NaN	8476.0	41468.0	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg	https://t.co/tD36da7qLQ						
4	891327558926688256	2017-07-29 16:00:24	This is Franklin. He would like you to stop ca...	https://twitter.com/dog_rates/status/891327558...	12	Franklin	NaN	9169.0	39634.0	https://pbs.twimg.com/media/DF6hr6AVYAAZ8G8.jpg	https://t.co/AtUZn91f7f						

0	orange	0.097049	False	bagel	0.085851	False
1	chihuahua	0.323581	True	pekinese	0.090647	True
2	chihuahua	0.716012	True	malamute	0.078253	True
3	paper towel	0.170278	False	labrador retriever	0.168086	True
4	NaN	NaN	NaN	NaN	NaN	NaN

	p3	p3_conf	p3_dog
0	banana	0.076110	False
1	papillon	0.068957	True
2	kelpie	0.031379	True
3	spatula	0.040836	False
4	NaN	NaN	NaN

```
In [80]: df_tweets_archive.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1941 entries, 0 to 1940
Data columns (total 20 columns):
tweet_id          1941 non-null int64
timestamp         1941 non-null datetime64[ns]
text              1941 non-null object
expanded_urls     1941 non-null object
rating_numerator  1941 non-null int64
name              1368 non-null object
pet_class         297 non-null object
retweet_count     1940 non-null float64
favorite_count    1940 non-null float64
jpg_url           1940 non-null object
url               1940 non-null object
p1                1672 non-null object
p1_conf           1672 non-null float64
p1_dog            1672 non-null object
p2                1672 non-null object
p2_conf           1672 non-null float64
p2_dog            1672 non-null object
p3                1672 non-null object
p3_conf           1672 non-null float64
p3_dog            1672 non-null object
dtypes: datetime64[ns](1), float64(5), int64(2), object(12)
memory usage: 303.4+ KB
```

```
In [81]: df_tweets_archive.describe().to_csv('csv/tab1.csv')
df_tweets_archive.describe()
```

```
Out[81]:
```

	tweet_id	rating_numerator	retweet_count	favorite_count	\
count	1.941000e+03	1941.000000	1940.000000	1940.000000	
mean	7.345651e+17	12.215353	2685.393299	8810.139175	
std	6.674540e+16	41.953730	4792.382675	12969.727538	

min	6.660209e+17	0.000000	12.000000	80.000000
25%	6.757109e+17	10.000000	589.500000	1864.500000
50%	7.079698e+17	11.000000	1271.000000	3880.500000
75%	7.844314e+17	12.000000	3065.250000	11009.500000
max	8.924206e+17	1776.000000	83828.000000	164571.000000

	p1_conf	p2_conf	p3_conf
count	1672.000000	1.672000e+03	1.672000e+03
mean	0.571605	1.405823e-01	6.269009e-02
std	0.274083	1.025783e-01	5.155378e-02
min	0.044333	1.011300e-08	1.740170e-10
25%	0.339591	5.925937e-02	1.716503e-02
50%	0.544931	1.245165e-01	5.454660e-02
75%	0.823429	2.024830e-01	9.476893e-02
max	1.000000	4.880140e-01	2.734190e-01

### 2.5.1 Qual a média das notas por classificação?

```
In [82]: df_tweets_archive.rating_numerator.describe().to_csv('csv/tab2.csv')
df_tweets_archive.rating_numerator.describe()
```

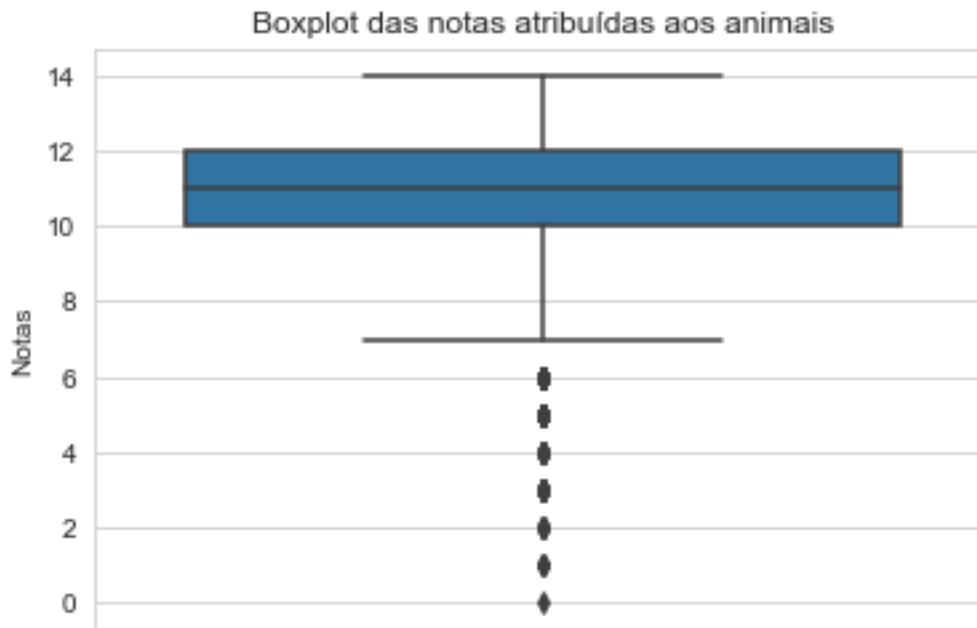
```
Out [82]: count      1941.000000
mean         12.215353
std          41.953730
min           0.000000
25%          10.000000
50%          11.000000
75%          12.000000
max          1776.000000
Name: rating_numerator, dtype: float64
```

```
In [83]: df_tweets_archive[df_tweets_archive.rating_numerator <=
          15].rating_numerator.describe().to_csv('csv/tab3.csv')
df_tweets_archive[df_tweets_archive.rating_numerator <=
          15].rating_numerator.describe()
```

```
Out [83]: count      1924.000000
mean         10.500000
std           2.192385
min           0.000000
25%          10.000000
50%          11.000000
75%          12.000000
max          14.000000
Name: rating_numerator, dtype: float64
```

```
In [84]: ax = sns.boxplot(
          data=df_tweets_archive[df_tweets_archive.rating_numerator <= 15].rating_numerator,
          ax.set_title('Boxplot das notas atribuídas aos animais'))
```

```
ax.set_ylabel('Notas')
ax.set_xticks([])
fig = ax.get_figure()
fig.savefig('imgs/fig1.png');
```



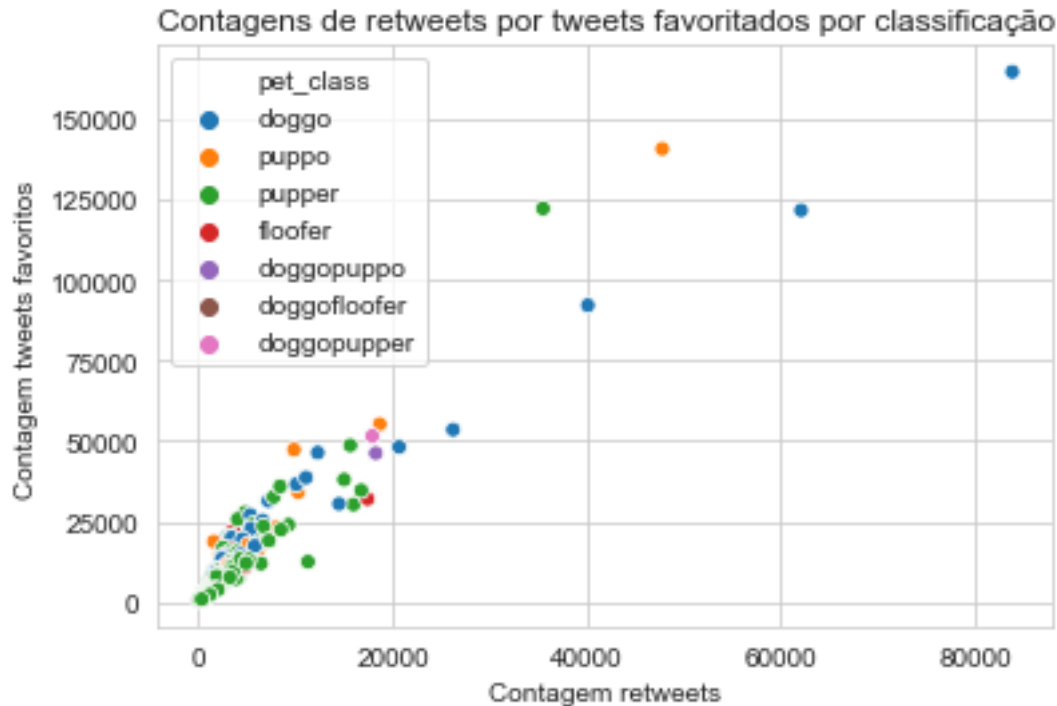
**Insight** Neste gráfico podemos verificar depois de remover as notas discrepantes, ou seja superiores a 15, assim removendo outliers (valores que podiam chegar a 1776, sendo que a nota deveria ser até 10 visto que o denominador é 10 mas os usuários como brincadeira classificam os animais com notas acima desse valor), e com média de 10.49 e desvio de 2.19. Observa-se que existe uma homogeneidade da variância no boxplot, visto que a diferença entre o terceiro quartil e a mediana assim como a diferença entre o primeiro quartil e a mediana parecem ser as mesmas.

## 2.5.2 Relação entre contagem de retweets e tweets favoritos

```
In [85]: sns.set_style("whitegrid")
```

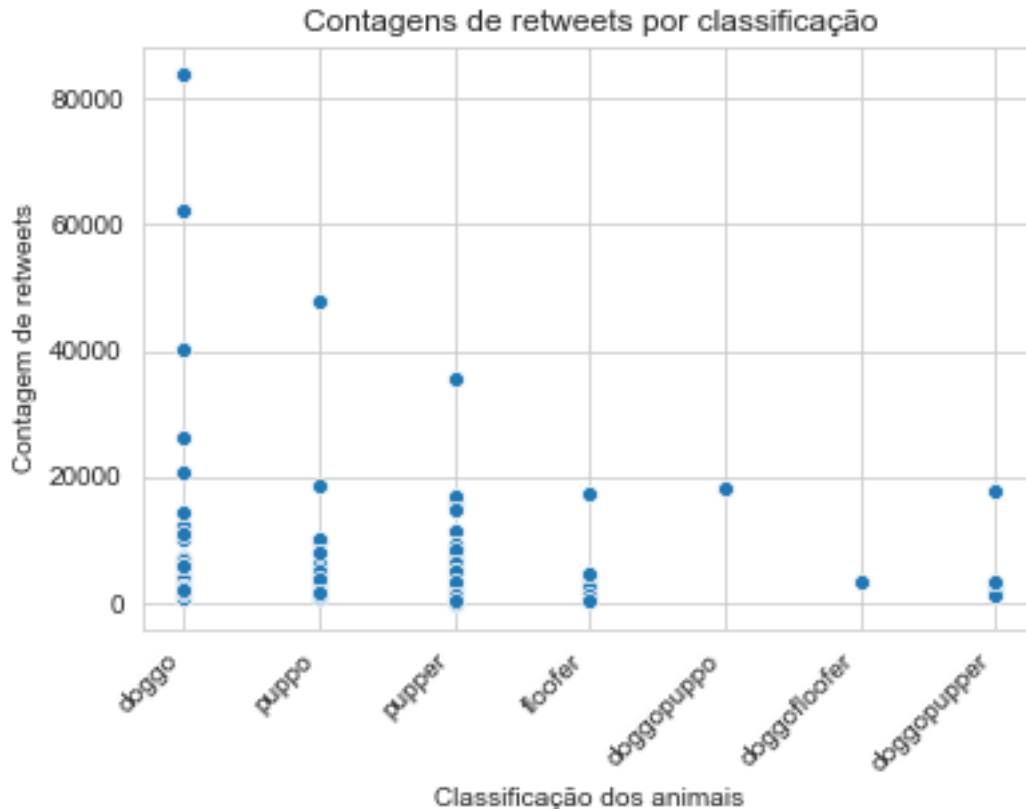
```
ax = sns.scatterplot(x="retweet_count", y="favorite_count",
                    hue='pet_class', data=df_tweets_archive)
ax.set_title('Contagens de retweets por tweets favoritos por classificação')
ax.set_xlabel('Contagem retweets')
ax.set_ylabel('Contagem tweets favoritos')
fig = ax.get_figure()
fig.savefig('imgs/fig2.png');
```





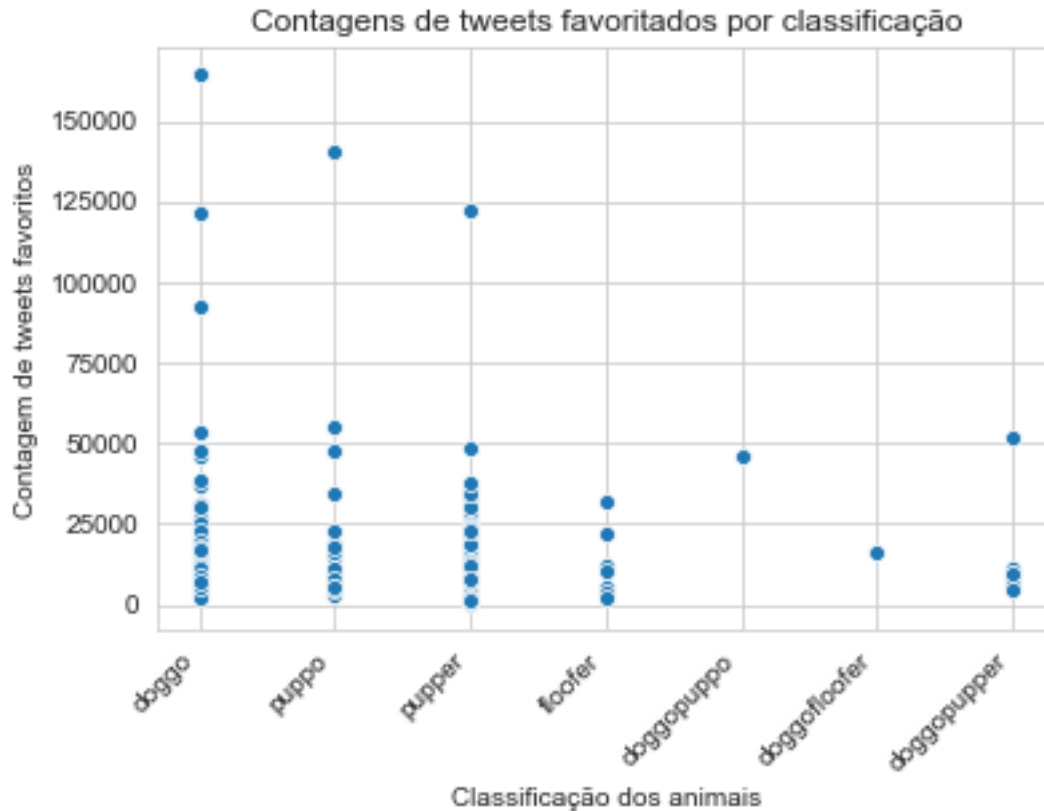
**Insight** Esse é um gráfico de dispersão, entre a contagem de retweets e a contagem de tweets favoritos para cada classificação. Observa-se que a medida que aumenta a contagem de retweets aumenta a contagem de tweets favoritos (o que pode ser devido a maior visibilidade do tweet) e a classe que possui a maior evidência é a classe pupper. Vale salientar que é possível observar que a classe doggo possui observações com maiores números de retweets, consequentemente, maior contagem de tweets favoritos.

```
In [86]: ax = sns.scatterplot(x="pet_class", y="retweet_count", data=df_tweets_archive)
labels = df_tweets_archive.pet_class.unique()[1:]
ax.set_xticklabels(rotation=45, ha='right', labels=labels)
ax.set_title('Contagens de retweets por classificação')
ax.set_xlabel('Classificação dos animais')
ax.set_ylabel('Contagem de retweets')
fig = ax.get_figure()
fig.savefig('imgs/fig3.png');
```



**Insight** Por este gráfico é possível verificar que realmente a classificação doggo é a que possui o tweet com maior número de retweets, porém também podemos ver que pupper tem um grande concentração de retweets até 20000. Cachorros que possuem duas classificações são poucos com relação aos demais.

```
In [87]: ax = sns.scatterplot(x="pet_class", y="favorite_count", data=df_tweets_archive)
labels = df_tweets_archive.pet_class.unique()[1:]
ax.set_xticklabels(rotation=45, ha='right', labels=labels)
ax.set_title('Contagens de tweets favoritados por classificação')
ax.set_xlabel('Classificação dos animais')
ax.set_ylabel('Contagem de tweets favoritos')
fig = ax.get_figure()
fig.savefig('imgs/fig4.png');
```



```
In [88]: df_tweets_archive[['retweet_count', 'favorite_count']].corr().to_csv('csv/tab4.csv')
df_tweets_archive[['retweet_count', 'favorite_count']].corr()
```

```
Out [88]:
```

	retweet_count	favorite_count
retweet_count	1.000000	0.929773
favorite_count	0.929773	1.000000

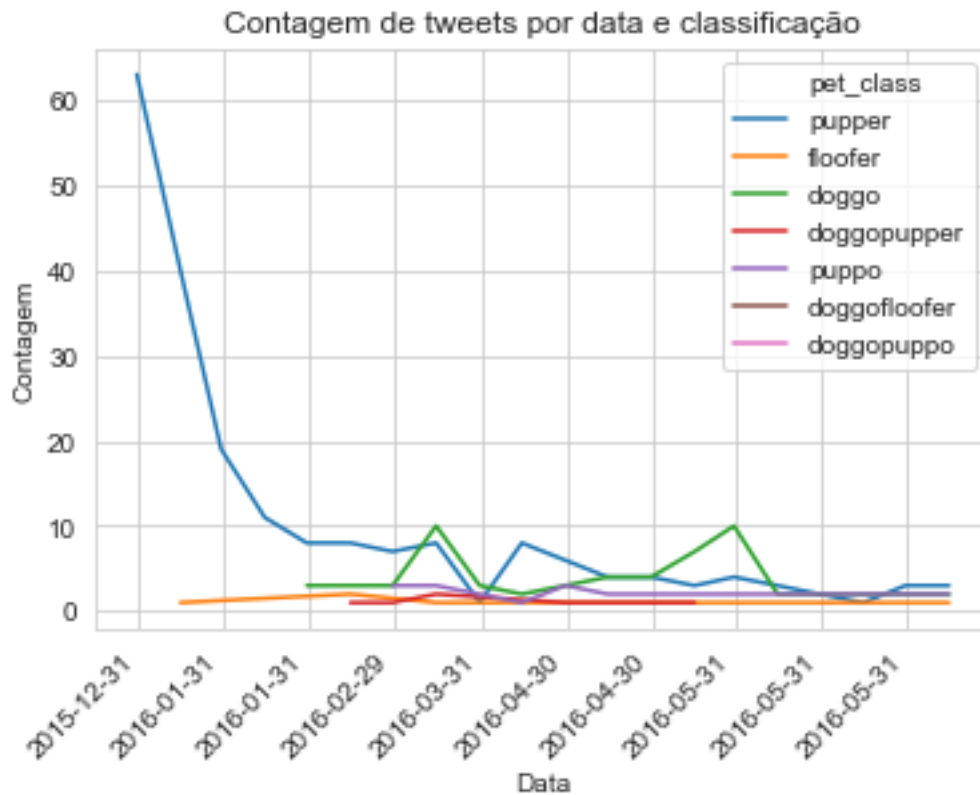
**Insight** Ao comparar a correlação entre `retweet_count` e `favorite_count` pode-se notar que existe uma forte correlação positiva entre as variáveis, isto é a medida que a contagem de retweets aumenta significa o quão favorito aquele tweet é com os usuários.

### 2.5.3 Contagem de tweets por data da post e classificação do animal

```
In [89]: graph = df_tweets_archive.set_index('timestamp')
graph = graph.groupby([pd.Grouper(freq='M'), 'pet_class']).count()
graph = graph.reset_index(level=['timestamp', 'pet_class'])

ax = sns.lineplot(x='timestamp', y='tweet_id', data=graph, hue='pet_class')
ax.set_xticklabels(rotation=45, ha='right', labels=graph.timestamp.dt.date)
ax.set_title('Contagem de tweets por data e classificação')
ax.set_xlabel('Data')
```

```
ax.set_ylabel('Contagem')
fig = ax.get_figure()
fig.savefig('imgs/fig5.png');
```



**Insight** Ao início é possível observar que a grande maioria dos tweets com classificação eram classificados unicamente como pupper, a medida que o tempo passa outras classificações são utilizadas e a contagem de tweets de cachorros denominados como pupper diminuiu. Observa-se que embora tenham ocorrido diferentes classificações, como doggo por exemplo, o número de tweets que possuem os animais denominados nessas classes diminuiu entre os anos.