

# Project: Wrangle and Analyze Data (WeRateDogs)

This project is based on tackling with data wrangling efforts.

## Table of Contents

- [Introduction](#)
- [Part I - Gathering Data](#)
- [Part II - Assessing Data](#)
- [Part III - Cleaning Data](#)
- [Part IV - Analyzing and visualizing Wrangled data](#)
- [Part V - Conclusion](#)

## Introduction

Wrangling and analyzing data is a demanding process for data analyst and its work. In this project, I will use the data wrangling skills to gather data from Twitter, clean them, and do some analysis. Then, I will get the original Twitter data from Twitter user @dog\_rates, along with a image prediction dataset. I will document my wrangling efforts in a Jupyter Notebook, and showcase them through analyses and visualizations using Python (and its libraries).

As WeRateDogs is a popular Twitter hash tag, people rate dogs with a denominator of 10 and the numerator is usually higher than 10 to show how lovely the dog is. Furthermore, WeRateDogs has over 4 million followers and has received international media coverage.

## Part I - Gathering Data

Data will be gathered from three resources:

1. The WeRateDogs Twitter archive. The twitter\_archive\_enhanced.csv file was provided to me.
2. The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file was provided to me.
3. Twitter API and Python's Tweepy library to gather each tweet's retweet count and favorite ("like") count at minimum, and any additional data I find interesting.

In [22]:

```
# Because tweepy module cannot be encompassed, it should be install as Anaconda Command.  
# !pip install tweepy
```

In [1]:

```
# import libraries I use to build my project  
import numpy as np  
import pandas as pd  
import requests  
import json  
import datetime  
import tweepy  
import sys  
import time  
import seaborn as sns  
import matplotlib.pyplot as plt  
from IPython.display import Image  
%matplotlib inline
```

1. Gathering the WeRateDogs Twitter archive Whose filename is twitter\_archive\_enhanced.csv

In [98]:

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

1. Downloading the image predictions file from the Udacity server and writing all content to image\_predictions.tsv file

In [99]:

```
url="https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv"
response = requests.get(url)
with open('image_predictions.tsv', 'wb') as file:
    file.write(response.content)

image_predictions = pd.read_csv('image_predictions.tsv', sep='\t')
```

1. Querying Twitter's API for JSON data - match to each Tweet ID from the archive

In [41]:

```
consumer_key = "MY_CUSTOMER_KEY"
consumer_secret = "MY_CONSUMER_SECRET"
access_token = "ACCESS_TOKEN"
access_secret = "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)

if(api):
    print("Connected Twitter via API")

# As, it has a long time process fo loop part, I would like to calculate how much time is it done.
start = time.time()

errors = []
tweet_ids = twitter_archive["tweet_id"]
with open('tweet_json.txt', 'w') as tweet_json_file:
    for tweet_id in tweet_ids:
        try:
            tweet = api.get_status(tweet_id, tweet_mode='extended', wait_on_rate_limit = True,
                                   wait_on_rate_limit_notify = True)
            json.dump(tweet._json, tweet_json_file)
            tweet_json_file.write('\n')
            print("Tweet ID : {}".format(tweet_id), end="\r")
        except tweepy.TweepError as e:
            errors.append(tweet_id)
            print("Error : {}".format(e), end="\r")
        pass

# this stops the timer
end = time.time()

minutes_time = int( (end - start) / 60)
seconds_time = int( (end - start) % 60)
print("Time the process completed : {} minutes {} seconds ".format(minutes_time, seconds_time))
```

```
Connected Twitter via API
Tweet ID : 892420643555336193
Tweet ID : 892177421306343426
Tweet ID : 891815181378084864
Tweet ID : 891689557279858688
Tweet ID : 891327558926688256
Tweet ID : 891087950875897856
Tweet ID : 890971913173991426
Tweet ID : 890729181411237888
Tweet ID : 890609185150312448
Tweet ID : 890240255349198849
Tweet ID : 890006608113172480
Tweet ID : 889880896479866881
Tweet ID : 889665388333682689
Tweet ID : 889638837579907072
Tweet ID : 889531135344209921
Tweet ID : 8892788411881685760
```

Tweet ID : 889278841981685760  
Tweet ID : 888917238123831296  
Tweet ID : 888804989199671297  
Tweet ID : 888554962724278272  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 888078434458587136  
Tweet ID : 887705289381826560  
Tweet ID : 887517139158093824  
Tweet ID : 887473957103951883  
Tweet ID : 887343217045368832  
Tweet ID : 887101392804085760  
Tweet ID : 886983233522544640  
Tweet ID : 886736880519319552  
Tweet ID : 886680336477933568  
Tweet ID : 886366144734445568  
Tweet ID : 886267009285017600  
Tweet ID : 886258384151887873  
Tweet ID : 886054160059072513  
Tweet ID : 885984800019947520  
Tweet ID : 885528943205470208  
Tweet ID : 885518971528720385  
Tweet ID : 885311592912609280  
Tweet ID : 885167619883638784  
Tweet ID : 884925521741709313  
Tweet ID : 884876753390489601  
Tweet ID : 884562892145688576  
Tweet ID : 884441805382717440  
Tweet ID : 884247878851493888  
Tweet ID : 884162670584377345  
Tweet ID : 883838122936631299  
Tweet ID : 883482846933004288  
Tweet ID : 883360690899218434  
Tweet ID : 883117836046086144  
Tweet ID : 882992080364220416  
Tweet ID : 882762694511734784  
Tweet ID : 882627270321602560  
Tweet ID : 882268110199369728  
Tweet ID : 882045870035918850  
Tweet ID : 881906580714921986  
Tweet ID : 881666595344535552  
Tweet ID : 881633300179243008  
Tweet ID : 881536004380872706  
Tweet ID : 881268444196462592  
Tweet ID : 880935762899988482  
Tweet ID : 880872448815771648  
Tweet ID : 880465832366813184  
Tweet ID : 880221127280381952  
Tweet ID : 880095782870896641  
Tweet ID : 879862464715927552  
Tweet ID : 879674319642796034  
Tweet ID : 879492040517615616  
Tweet ID : 879415818425184262  
Tweet ID : 879376492567855104  
Tweet ID : 879130579576475649  
Tweet ID : 879050749262655488  
Tweet ID : 879008229531029506  
Tweet ID : 878776093423087618  
Tweet ID : 878604707211726852  
Tweet ID : 878404777348136964  
Tweet ID : 878316110768087041  
Tweet ID : 878281511006478336  
Tweet ID : 878057613040115712  
Tweet ID : 877736472329191424  
Tweet ID : 877611172832227328  
Tweet ID : 877556246731214848  
Tweet ID : 877316821321428993  
Tweet ID : 877201837425926144  
Tweet ID : 876838120628539392  
Tweet ID : 876537666061221889  
Tweet ID : 876484053909872640  
Tweet ID : 876120275196170240  
Tweet ID : 875747767867523072  
Tweet ID : 875144289856114688  
Tweet ID : 875097192612077568  
Tweet ID : 875021211251597312  
Tweet ID : 874680097055178752  
Tweet ID : 874434818259525634  
Tweet ID : 874006800500600000

Tweet ID : 874296783580663808  
Tweet ID : 874057562936811520  
Tweet ID : 874012996292530176  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 873580283840344065  
Tweet ID : 873337748698140672  
Tweet ID : 873213775632977920  
Tweet ID : 872967104147763200  
Tweet ID : 872820683541237760  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 872620804844003328  
Tweet ID : 872486979161796608  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 872122724285648897  
Tweet ID : 871879754684805121  
Tweet ID : 871762521631449091  
Tweet ID : 871515927908634625  
Tweet ID : 871166179821445120  
Tweet ID : 871102520638267392  
Tweet ID : 871032628920680449  
Tweet ID : 870804317367881728  
Tweet ID : 870726314365509632  
Tweet ID : 870656317836468226  
Tweet ID : 870374049280663552  
Tweet ID : 870308999962521604  
Tweet ID : 870063196459192321  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 869772420881756160  
Tweet ID : 869702957897576449  
Tweet ID : 869596645499047938  
Tweet ID : 869227993411051520  
Tweet ID : 868880397819494401  
Tweet ID : 868639477480148993  
Tweet ID : 868622495443632128  
Tweet ID : 868552278524837888  
Tweet ID : 867900495410671616  
Tweet ID : 867774946302451713  
Tweet ID : 867421006826221569  
Tweet ID : 867072653475098625  
Tweet ID : 867051520902168576  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 866720684873056260  
Tweet ID : 866686824827068416  
Tweet ID : 866450705531457537  
Tweet ID : 866334964761202691  
Tweet ID : 866094527597207552  
Tweet ID : 865718153858494464  
Tweet ID : 865359393868664832  
Tweet ID : 865006731092295680  
Tweet ID : 864873206498414592  
Tweet ID : 864279568663928832  
Tweet ID : 864197398364647424  
Tweet ID : 863907417377173506  
Tweet ID : 863553081350529029  
Tweet ID : 863471782782697472  
Tweet ID : 863432100342583297  
Tweet ID : 863427515083354112  
Tweet ID : 863079547188785154  
Tweet ID : 863062471531167744  
Tweet ID : 862831371563274240  
Tweet ID : 862722525377298433  
Tweet ID : 862457590147678208  
Tweet ID : 862096992088072192  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 861383897657036800  
Tweet ID : 861288531465048066  
Tweet ID : 861005113778896900  
Tweet ID : 860981674716409858  
Tweet ID : 860924035999428608  
Tweet ID : 860563773140209665  
Tweet ID : 860524505164394496  
Tweet ID : 860276583193509888  
Tweet ID : 860184849394610176  
Tweet ID : 860177593139703809  
Tweet ID : 859924526012018688  
Tweet ID : 859851578198683649  
Tweet ID : 859607811541651456  
- - - - -

Tweet ID : 859196978902773760  
Tweet ID : 859074603037188101  
Tweet ID : 858860390427611136  
Tweet ID : 858843525470990336  
Tweet ID : 858471635011153920  
Tweet ID : 858107933456039936  
Tweet ID : 857989990357356544  
Tweet ID : 857746408056729600  
Tweet ID : 857393404942143489  
Tweet ID : 857263160327368704  
Tweet ID : 857214891891077121  
Tweet ID : 857062103051644929  
Tweet ID : 857029823797047296  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 856543823941562368  
Tweet ID : 856526610513747968  
Tweet ID : 856330835276025856  
Tweet ID : 856288084350160898  
Tweet ID : 856282028240666624  
Tweet ID : 855862651834028034  
Tweet ID : 855860136149123072  
Tweet ID : 855857698524602368  
Tweet ID : 855851453814013952  
Tweet ID : 855818117272018944  
Tweet ID : 855459453768019968  
Tweet ID : 855245323840757760  
Tweet ID : 855138241867124737  
Tweet ID : 854732716440526848  
Tweet ID : 854482394044301312  
Tweet ID : 854365224396361728  
Tweet ID : 854120357044912130  
Tweet ID : 854010172552949760  
Tweet ID : 853760880890318849  
Tweet ID : 853639147608842240  
Tweet ID : 853299958564483072  
Tweet ID : 852936405516943360  
Tweet ID : 852912242202992640  
Tweet ID : 852672615818899456  
Tweet ID : 852553447878664193  
Tweet ID : 852311364735569921  
Tweet ID : 852226086759018497  
Tweet ID : 852189679701164033  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 851861385021730816  
Tweet ID : 851591660324737024  
Tweet ID : 851464819735769094  
Tweet ID : 851224888060895234  
Tweet ID : 850753642995093505  
Tweet ID : 850380195714523136  
Tweet ID : 850333567704068097  
Tweet ID : 850145622816686080  
Tweet ID : 850019790995546112  
Tweet ID : 849776966551130114  
Tweet ID : 849668094696017920  
Tweet ID : 849412302885593088  
Tweet ID : 849336543269576704  
Tweet ID : 849051919805034497  
Tweet ID : 848690551926992896  
Tweet ID : 848324959059550208  
Tweet ID : 848213670039564288  
Tweet ID : 848212111729840128  
Tweet ID : 847978865427394560  
Tweet ID : 847971574464610304  
Tweet ID : 847962785489326080  
Tweet ID : 847842811428974592  
Tweet ID : 847617282490613760  
Tweet ID : 847606175596138505  
Tweet ID : 847251039262605312  
Tweet ID : 847157206088847362  
Tweet ID : 847116187444137987  
Tweet ID : 846874817362120707  
Tweet ID : 846514051647705089  
Tweet ID : 846505985330044928  
Tweet ID : 846153765933735936  
Tweet ID : 846139713627017216  
Tweet ID : 846042936437604353  
Tweet ID : 845812042753855489  
-----

Tweet ID : 845677943972139009  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 845397057150107648  
Tweet ID : 845306882940190720  
Tweet ID : 845098359547420673  
Tweet ID : 844979544864018432  
Tweet ID : 844973813909606400  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 844580511645339650  
Tweet ID : 844223788422217728  
Tweet ID : 843981021012017153  
Tweet ID : 843856843873095681  
Tweet ID : 843604394117681152  
Tweet ID : 843235543001513987  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 842846295480000512  
Tweet ID : 842765311967449089  
Tweet ID : 842535590457499648  
Tweet ID : 842163532590374912  
Tweet ID : 842115215311396866  
Tweet ID : 841833993020538882  
Tweet ID : 841680585030541313  
Tweet ID : 841439858740625411  
Tweet ID : 841320156043304961  
Tweet ID : 841314665196081154  
Tweet ID : 841077006473256960  
Tweet ID : 840761248237133825  
Tweet ID : 840728873075638272  
Tweet ID : 840698636975636481  
Tweet ID : 840696689258311684  
Tweet ID : 840632337062862849  
Tweet ID : 840370681858686976  
Tweet ID : 840268004936019968  
Tweet ID : 839990271299457024  
Tweet ID : 839549326359670784  
Tweet ID : 839290600511926273  
Tweet ID : 839239871831150596  
Tweet ID : 838952994649550848  
Tweet ID : 838921590096166913  
Tweet ID : 838916489579200512  
Tweet ID : 838831947270979586  
Tweet ID : 838561493054533637  
Tweet ID : 838476387338051585  
Tweet ID : 838201503651401729  
Tweet ID : 838150277551247360  
Tweet ID : 838085839343206401  
Tweet ID : 838083903487373313  
Tweet ID : 837820167694528512  
Tweet ID : 837482249356513284  
Tweet ID : 837471256429613056  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 837110210464448512  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 836989968035819520  
Tweet ID : 836753516572119041  
Tweet ID : 836677758902222849  
Tweet ID : 836648853927522308  
Tweet ID : 836397794269200385  
Tweet ID : 836380477523124226  
Tweet ID : 836260088725786625  
Tweet ID : 836001077879255040  
Tweet ID : 835685285446955009  
Tweet ID : 835574547218894849  
Tweet ID : 835536468978302976  
Tweet ID : 835309094223372289  
Tweet ID : 835297930240217089  
Tweet ID : 835264098648616962  
Tweet ID : 835246439529840640  
Tweet ID : 835172783151792128  
Tweet ID : 835152434251116546  
Tweet ID : 834931633769889797  
Tweet ID : 834786237630337024  
Tweet ID : 834574053763584002  
Tweet ID : 834477809192075265  
Tweet ID : 834458053273591808  
Tweet ID : 834209720923721728  
Tweet ID : 834167344700198914

Tweet ID : 834089966724603904  
Tweet ID : 834086379323871233  
Tweet ID : 833863086058651648  
Tweet ID : 833826103416520705  
Tweet ID : 833732339549220864  
Tweet ID : 833722901757046785  
Tweet ID : 833479644947025920  
Tweet ID : 833124694597443584  
Tweet ID : 832998151111966721  
Tweet ID : 832769181346996225  
Tweet ID : 832757312314028032  
Tweet ID : 832682457690300417  
Tweet ID : 832645525019123713  
Tweet ID : 832636094638288896  
Tweet ID : 832397543355072512  
Tweet ID : 832369877331693569  
Tweet ID : 832273440279240704  
Tweet ID : 832215909146226688  
Tweet ID : 832215726631055365  
Tweet ID : 832088576586297345  
Tweet ID : 832040443403784192  
Tweet ID : 832032802820481025  
Tweet ID : 831939777352105988  
Tweet ID : 831926988323639298  
Tweet ID : 831911600680497154  
Tweet ID : 831670449226514432  
Tweet ID : 831650051525054464  
Tweet ID : 831552930092285952  
Tweet ID : 831322785565769729  
Tweet ID : 831315979191906304  
Tweet ID : 831309418084069378  
Tweet ID : 831262627380748289  
Tweet ID : 830956169170665475  
Tweet ID : 830583320585068544  
Tweet ID : 830173239259324417  
Tweet ID : 830097400375152640  
Tweet ID : 829878982036299777  
Tweet ID : 829861396166877184  
Tweet ID : 829501995190984704  
Tweet ID : 829449946868879360  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 829141528400556032  
Tweet ID : 829011960981237760  
Tweet ID : 828801551087042563  
Tweet ID : 828770345708580865  
Tweet ID : 828708714936930305  
Tweet ID : 828650029636317184  
Tweet ID : 828409743546925057  
Tweet ID : 828408677031882754  
Tweet ID : 828381636999917570  
Tweet ID : 828376505180889089  
Tweet ID : 828372645993398273  
Tweet ID : 828361771580813312  
Tweet ID : 828046555563323392  
Tweet ID : 828011680017821696  
Tweet ID : 827933404142436356  
Tweet ID : 827653905312006145  
Tweet ID : 827600520311402496  
Tweet ID : 827324948884643840  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 827199976799354881  
Tweet ID : 826958653328592898  
Tweet ID : 826848821049180160  
Tweet ID : 826615380357632002  
Tweet ID : 826598799820865537  
Tweet ID : 826598365270007810  
Tweet ID : 826476773533745153  
Tweet ID : 826240494070030336  
Tweet ID : 826204788643753985  
Tweet ID : 826115272272650244  
Tweet ID : 825876512159186944  
Tweet ID : 825829644528148480  
Tweet ID : 825535076884762624  
Tweet ID : 825147591692263424  
Tweet ID : 825120256414846976  
Tweet ID : 825026590719483904  
Tweet ID : 824796380199809024

Tweet ID : 824775126675836928  
Tweet ID : 824663926340194305  
Tweet ID : 824325613288833024  
Tweet ID : 824297048279236611  
Tweet ID : 824025158776213504  
Tweet ID : 823939628516474880  
Tweet ID : 823719002937630720  
Tweet ID : 823699002998870016  
Tweet ID : 823581115634085888  
Tweet ID : 823333489516937216  
Tweet ID : 823322678127919110  
Tweet ID : 823269594223824897  
Tweet ID : 822975315408461824  
Tweet ID : 822872901745569793  
Tweet ID : 822859134160621569  
Tweet ID : 822647212903690241  
Tweet ID : 822610361945911296  
Tweet ID : 822489057087389700  
Tweet ID : 822462944365645825  
Tweet ID : 822244816520155136  
Tweet ID : 822163064745328640  
Tweet ID : 821886076407029760  
Tweet ID : 821813639212650496  
Tweet ID : 821765923262631936  
Tweet ID : 821522889702862852  
Tweet ID : 821421320206483457  
Tweet ID : 821407182352777218  
Tweet ID : 821153421864615936  
Tweet ID : 821149554670182400  
Tweet ID : 821107785811234820  
Tweet ID : 821044531881721856  
Tweet ID : 820837357901512704  
Tweet ID : 820749716845686786  
Tweet ID : 820690176645140481  
Tweet ID : 820494788566847489  
Tweet ID : 820446719150292993  
Tweet ID : 820314633777061888  
Tweet ID : 820078625395449857  
Tweet ID : 820013781606658049  
Tweet ID : 819952236453363712  
Tweet ID : 819924195358416896  
Tweet ID : 819711362133872643  
Tweet ID : 819588359383371776  
Tweet ID : 819347104292290561  
Tweet ID : 819238181065359361  
Tweet ID : 819227688460238848  
Tweet ID : 819015337530290176  
Tweet ID : 819015331746349057  
Tweet ID : 819006400881917954  
Tweet ID : 819004803107983360  
Tweet ID : 818646164899774465  
Tweet ID : 818627210458333184  
Tweet ID : 818614493328580609  
Tweet ID : 818588835076603904  
Tweet ID : 818536468981415936  
Tweet ID : 818307523543449600  
Tweet ID : 818259473185828864  
Tweet ID : 818145370475810820  
Tweet ID : 817908911860748288  
Tweet ID : 817827839487737858  
Tweet ID : 817777686764523521  
Tweet ID : 817536400337801217  
Tweet ID : 817502432452313088  
Tweet ID : 817423860136083457  
Tweet ID : 817415592588222464  
Tweet ID : 817181837579653120  
Tweet ID : 817171292965273600  
Tweet ID : 817120970343411712  
Tweet ID : 817056546584727552  
Tweet ID : 816829038950027264  
Tweet ID : 816816676327063552  
Tweet ID : 816697700272001025  
Tweet ID : 816450570814898180  
Tweet ID : 816336735214911488  
Tweet ID : 816091915477250048  
Tweet ID : 816062466425819140  
Tweet ID : 816014286006976512



Tweet ID : 815990720817401858  
Tweet ID : 815966073409433600  
Tweet ID : 815745968457060357  
Tweet ID : 815736392542261248  
Tweet ID : 815639385530101762  
Tweet ID : 815390420867969024  
Tweet ID : 814986499976527872  
Tweet ID : 814638523311648768  
Tweet ID : 814578408554463233  
Tweet ID : 814530161257443328  
Tweet ID : 814153002265309185  
Tweet ID : 813944609378369540  
Tweet ID : 813910438903693312  
Tweet ID : 813812741911748608  
Tweet ID : 813800681631023104  
Tweet ID : 813217897535406080  
Tweet ID : 813202720496779264  
Tweet ID : 813187593374461952  
Tweet ID : 813172488309972993  
Tweet ID : 813157409116065792  
Tweet ID : 813142292504645637  
Tweet ID : 813130366689148928  
Tweet ID : 813127251579564032  
Tweet ID : 813112105746448384  
Tweet ID : 813096984823349248  
Tweet ID : 813081950185472002  
Tweet ID : 813066809284972545  
Tweet ID : 813051746834595840  
Tweet ID : 812781120811126785  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 812709060537683968  
Tweet ID : 812503143955202048  
Tweet ID : 812466873996607488  
Tweet ID : 812372279581671427  
Tweet ID : 811985624773361665  
Tweet ID : 811744202451197953  
Tweet ID : 811647686436880384  
Tweet ID : 811627233043480576  
Tweet ID : 811386762094317568  
Tweet ID : 810984652412424192  
Tweet ID : 810896069567610880  
Tweet ID : 810657578271330305  
Tweet ID : 810284430598270976  
Tweet ID : 810254108431155201  
Tweet ID : 809920764300447744  
Tweet ID : 809808892968534016  
Tweet ID : 809448704142938112  
Tweet ID : 809220051211603969  
Tweet ID : 809084759137812480  
Tweet ID : 808838249661788160  
Tweet ID : 808733504066486276  
Tweet ID : 808501579447930884  
Tweet ID : 808344865868283904  
Tweet ID : 808134635716833280  
Tweet ID : 808106460588765185  
Tweet ID : 808001312164028416  
Tweet ID : 807621403335917568  
Tweet ID : 807106840509214720  
Tweet ID : 807059379405148160  
Tweet ID : 807010152071229440  
Tweet ID : 806629075125202948  
Tweet ID : 806620845233815552  
Tweet ID : 806576416489959424  
Tweet ID : 806542213899489280  
Tweet ID : 806242860592926720  
Tweet ID : 806219024703037440  
Tweet ID : 805958939288408065  
Tweet ID : 805932879469572096  
Tweet ID : 805826884734976000  
Tweet ID : 805823200554876929  
Tweet ID : 805520635690676224  
Tweet ID : 805487436403003392  
Tweet ID : 805207613751304193  
Tweet ID : 804738756058218496  
Tweet ID : 804475857670639616  
Tweet ID : 804413760345620481  
Tweet ID : 804026241225523202

Tweet ID : 803773340896923648  
Tweet ID : 803692223237865472  
Tweet ID : 803638050916102144  
Tweet ID : 803380650405482500  
Tweet ID : 803321560782307329  
Tweet ID : 803276597545603072  
Tweet ID : 802952499103731712  
Tweet ID : 802624713319034886  
Tweet ID : 802600418706604034  
Tweet ID : 802572683846291456  
Tweet ID : 802323869084381190  
Tweet ID : 802265048156610565  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 802239329049477120  
Tweet ID : 802185808107208704  
Tweet ID : 801958328846974976  
Tweet ID : 801854953262350336  
Tweet ID : 801538201127157760  
Tweet ID : 801285448605831168  
Tweet ID : 801167903437357056  
Tweet ID : 801127390143516673  
Tweet ID : 801115127852503040  
Tweet ID : 800859414831898624  
Tweet ID : 800855607700029440  
Tweet ID : 800751577355128832  
Tweet ID : 800513324630806528  
Tweet ID : 800459316964663297  
Tweet ID : 800443802682937345  
Tweet ID : 800388270626521089  
Tweet ID : 800188575492947969  
Tweet ID : 800141422401830912  
Tweet ID : 800018252395122689  
Tweet ID : 799774291445383169  
Tweet ID : 799757965289017345  
Tweet ID : 799422933579902976  
Tweet ID : 799308762079035393  
Tweet ID : 799297110730567681  
Tweet ID : 799063482566066176  
Tweet ID : 798933969379225600  
Tweet ID : 798925684722855936  
Tweet ID : 798705661114773508  
Tweet ID : 798701998996647937  
Tweet ID : 798697898615730177  
Tweet ID : 798694562394996736  
Tweet ID : 798686750113755136  
Tweet ID : 798682547630837760  
Tweet ID : 798673117451325440  
Tweet ID : 798665375516884993  
Tweet ID : 798644042770751489  
Tweet ID : 798628517273620480  
Tweet ID : 798585098161549313  
Tweet ID : 798576900688019456  
Tweet ID : 798340744599797760  
Tweet ID : 798209839306514432  
Tweet ID : 797971864723324932  
Tweet ID : 797545162159308800  
Tweet ID : 797236660651966464  
Tweet ID : 797165961484890113  
Tweet ID : 796904159865868288  
Tweet ID : 796865951799083009  
Tweet ID : 796759840936919040  
Tweet ID : 796563435802726400  
Tweet ID : 796484825502875648  
Tweet ID : 796387464403357696  
Tweet ID : 796177847564038144  
Tweet ID : 796149749086875649  
Tweet ID : 796125600683540480  
Tweet ID : 796116448414461957  
Tweet ID : 796080075804475393  
Tweet ID : 796031486298386433  
Tweet ID : 795464331001561088  
Tweet ID : 795400264262053889  
Tweet ID : 795076730285391872  
Tweet ID : 794983741416415232  
Tweet ID : 794926597468000259  
Tweet ID : 794355576146903043  
Tweet ID : 794332329137291264

Tweet ID : 794205286408003585  
Tweet ID : 793962221541933056  
Tweet ID : 793845145112371200  
Tweet ID : 793614319594401792  
Tweet ID : 793601777308463104  
Tweet ID : 793500921481273345  
Tweet ID : 793286476301799424  
Tweet ID : 793271401113350145  
Tweet ID : 793256262322548741  
Tweet ID : 793241302385262592  
Tweet ID : 793226087023144960  
Tweet ID : 793210959003287553  
Tweet ID : 793195938047070209  
Tweet ID : 793180763617361921  
Tweet ID : 793165685325201412  
Tweet ID : 793150605191548928  
Tweet ID : 793135492858580992  
Tweet ID : 793120401413079041  
Tweet ID : 792913359805018113  
Tweet ID : 792883833364439040  
Tweet ID : 792773781206999040  
Tweet ID : 792394556390137856  
Tweet ID : 792050063153438720  
Tweet ID : 791821351946420224  
Tweet ID : 791784077045166082  
Tweet ID : 791780927877898241  
Tweet ID : 791774931465953280  
Tweet ID : 791672322847637504  
Tweet ID : 791406955684368384  
Tweet ID : 791312159183634433  
Tweet ID : 791026214425268224  
Tweet ID : 790987426131050500  
Tweet ID : 790946055508652032  
Tweet ID : 790723298204217344  
Tweet ID : 790698755171364864  
Tweet ID : 790581949425475584  
Tweet ID : 790337589677002753  
Tweet ID : 790277117346975746  
Tweet ID : 790227638568808452  
Tweet ID : 789986466051088384  
Tweet ID : 789960241177853952  
Tweet ID : 789903600034189313  
Tweet ID : 789628658055020548  
Tweet ID : 789599242079838210  
Tweet ID : 789530877013393408  
Tweet ID : 789314372632018944  
Tweet ID : 789280767834746880  
Tweet ID : 789268448748703744  
Tweet ID : 789137962068021249  
Tweet ID : 788908386943430656  
Tweet ID : 788765914992902144  
Tweet ID : 788552643979468800  
Tweet ID : 788412144018661376  
Tweet ID : 788178268662984705  
Tweet ID : 788150585577050112  
Tweet ID : 788070120937619456  
Tweet ID : 788039637453406209  
Tweet ID : 787810552592695296  
Tweet ID : 787717603741622272  
Tweet ID : 787397959788929025  
Tweet ID : 787322443945877504  
Tweet ID : 787111942498508800  
Tweet ID : 786963064373534720

Rate limit reached. Sleeping for: 414

Tweet ID : 786729988674449408  
Tweet ID : 786709082849828864  
Tweet ID : 786664955043049472  
Tweet ID : 786595970293370880  
Tweet ID : 786363235746385920  
Tweet ID : 786286427768250368  
Tweet ID : 786233965241827333  
Tweet ID : 786051337297522688  
Tweet ID : 786036967502913536  
Tweet ID : 785927819176054784

Tweet ID : 785872687017132033  
Tweet ID : 785639753186217984  
Tweet ID : 785533386513321988  
Tweet ID : 785515384317313025  
Tweet ID : 785264754247995392  
Tweet ID : 785170936622350336  
Tweet ID : 784826020293709826  
Tweet ID : 784517518371221505  
Tweet ID : 784431430411685888  
Tweet ID : 784183165795655680  
Tweet ID : 784057939640352768  
Tweet ID : 783839966405230592  
Tweet ID : 783821107061198850  
Tweet ID : 783695101801398276  
Tweet ID : 783466772167098368  
Tweet ID : 783391753726550016  
Tweet ID : 783347506784731136  
Tweet ID : 783334639985389568  
Tweet ID : 783085703974514689  
Tweet ID : 782969140009107456  
Tweet ID : 782747134529531904  
Tweet ID : 782722598790725632  
Tweet ID : 782598640137187329  
Tweet ID : 782305867769217024  
Tweet ID : 782021823840026624  
Tweet ID : 781955203444699136  
Tweet ID : 781661882474196992  
Tweet ID : 781655249211752448  
Tweet ID : 781524693396357120  
Tweet ID : 781308096455073793  
Tweet ID : 781251288990355457  
Tweet ID : 781163403222056960  
Tweet ID : 780931614150983680  
Tweet ID : 780858289093574656  
Tweet ID : 780800785462489090  
Tweet ID : 780601303617732608  
Tweet ID : 780543529827336192  
Tweet ID : 780496263422808064  
Tweet ID : 780476555013349377  
Tweet ID : 780459368902959104  
Tweet ID : 780192070812196864  
Tweet ID : 780092040432480260  
Tweet ID : 780074436359819264  
Tweet ID : 779834332596887552  
Tweet ID : 779377524342161408  
Tweet ID : 779124354206535695  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 779056095788752897  
Tweet ID : 778990705243029504  
Tweet ID : 778774459159379968  
Tweet ID : 778764940568104960  
Tweet ID : 778748913645780993  
Tweet ID : 778650543019483137  
Tweet ID : 778624900596654080  
Tweet ID : 778408200802557953  
Tweet ID : 778396591732486144  
Tweet ID : 778383385161035776  
Tweet ID : 778286810187399168  
Tweet ID : 778039087836069888  
Tweet ID : 778027034220126208  
Tweet ID : 777953400541634568  
Tweet ID : 777885040357281792  
Tweet ID : 777684233540206592  
Tweet ID : 777641927919427584  
Tweet ID : 777621514455814149  
Tweet ID : 777189768882946048  
Tweet ID : 776819012571455488  
Tweet ID : 776813020089548800  
Tweet ID : 776477788987613185  
Tweet ID : 776249906839351296  
Tweet ID : 776218204058357768  
Tweet ID : 776201521193218049  
Tweet ID : 776113305656188928  
Tweet ID : 776088319444877312  
Tweet ID : 775898661951791106  
Tweet ID : 775842724423557120  
Tweet ID : 775733305207554048

Tweet ID : 775755505207554040  
Tweet ID : 775729183532220416  
Tweet ID : 775364825476165632  
Tweet ID : 775350846108426240  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 775085132600442880  
Tweet ID : 774757898236878852  
Tweet ID : 774639387460112384  
Tweet ID : 774314403806253056  
Tweet ID : 773985732834758656  
Tweet ID : 773922284943896577  
Tweet ID : 773704687002451968  
Tweet ID : 773670353721753600  
Tweet ID : 773547596996571136  
Tweet ID : 773336787167145985  
Tweet ID : 773308824254029826  
Tweet ID : 773247561583001600  
Tweet ID : 773191612633579521  
Tweet ID : 772877495989305348  
Tweet ID : 772826264096874500  
Tweet ID : 772615324260794368  
Tweet ID : 772581559778025472  
Tweet ID : 772193107915964416  
Tweet ID : 772152991789019136  
Tweet ID : 772117678702071809  
Tweet ID : 772114945936949249  
Tweet ID : 772102971039580160  
Tweet ID : 771908950375665664  
Tweet ID : 771770456517009408  
Tweet ID : 771500966810099713  
Tweet ID : 771380798096281600  
Tweet ID : 771171053431250945  
Tweet ID : 771136648247640064  
Tweet ID : 771102124360998913  
Tweet ID : 771014301343748096  
Error : [{'code': 179, 'message': 'Sorry, you are not authorized to see this status.'}]  
Tweet ID : 770787852854652928  
Tweet ID : 770772759874076672  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 770655142660169732  
Tweet ID : 770414278348247044  
Tweet ID : 770293558247038976  
Tweet ID : 770093767776997377  
Tweet ID : 770069151037685760  
Tweet ID : 769940425801170949  
Tweet ID : 769695466921623552  
Tweet ID : 769335591808995329  
Tweet ID : 769212283578875904  
Tweet ID : 768970937022709760  
Tweet ID : 768909767477751808  
Tweet ID : 768855141948723200  
Tweet ID : 768609597686943744  
Tweet ID : 768596291618299904  
Tweet ID : 768554158521745409  
Tweet ID : 768473857036525572  
Tweet ID : 768193404517830656  
Tweet ID : 767884188863397888  
Tweet ID : 767754930266464257  
Tweet ID : 767500508068192258  
Tweet ID : 767191397493538821  
Tweet ID : 767122157629476866  
Tweet ID : 766864461642756096  
Tweet ID : 766793450729734144  
Tweet ID : 766714921925144576  
Tweet ID : 766693177336135680  
Tweet ID : 766423258543644672  
Tweet ID : 766313316352462849  
Tweet ID : 766078092750233600  
Tweet ID : 766069199026450432  
Tweet ID : 766008592277377025  
Tweet ID : 765719909049503744  
Tweet ID : 765669560888528897  
Tweet ID : 765395769549590528  
Tweet ID : 765371061932261376  
Tweet ID : 765222098633691136  
Tweet ID : 764857477905154048  
Tweet ID : 764259802650378240  
Tweet ID : 763856072077010045

Tweet ID : 763936972077010943  
Tweet ID : 763837565564780549  
Tweet ID : 763183847194451968  
Tweet ID : 763167063695355904  
Tweet ID : 763103485927849985  
Tweet ID : 762699858130116608  
Tweet ID : 762471784394268675  
Tweet ID : 762464539388485633  
Tweet ID : 762316489655476224  
Tweet ID : 762035686371364864  
Tweet ID : 761976711479193600  
Tweet ID : 761750502866649088  
Tweet ID : 761745352076779520  
Tweet ID : 761672994376806400  
Tweet ID : 761599872357261312  
Tweet ID : 761371037149827077  
Tweet ID : 761334018830917632  
Tweet ID : 761292947749015552  
Tweet ID : 761227390836215808  
Tweet ID : 761004547850530816  
Tweet ID : 760893934457552897  
Tweet ID : 760656994973933572  
Tweet ID : 760641137271070720  
Tweet ID : 760539183865880579  
Tweet ID : 760521673607086080  
Tweet ID : 760290219849637889  
Tweet ID : 760252756032651264  
Tweet ID : 760190180481531904  
Tweet ID : 760153949710192640  
Tweet ID : 759943073749200896  
Tweet ID : 759923798737051648  
Tweet ID : 759846353224826880  
Tweet ID : 759793422261743616  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 759557299618865152  
Tweet ID : 759447681597108224  
Tweet ID : 759446261539934208  
Tweet ID : 759197388317847553  
Tweet ID : 759159934323924993  
Tweet ID : 759099523532779520  
Tweet ID : 759047813560868866  
Tweet ID : 758854675097526272  
Tweet ID : 758828659922702336  
Tweet ID : 758740312047005698  
Tweet ID : 758474966123810816  
Tweet ID : 758467244762497024  
Tweet ID : 758405701903519748  
Tweet ID : 758355060040593408  
Tweet ID : 758099635764359168  
Tweet ID : 758041019896193024  
Tweet ID : 757741869644341248  
Tweet ID : 757729163776290825  
Tweet ID : 757725642876129280  
Tweet ID : 757611664640446465  
Tweet ID : 757597904299253760  
Tweet ID : 757596066325864448  
Tweet ID : 757400162377592832  
Tweet ID : 757393109802180609  
Tweet ID : 757354760399941633  
Tweet ID : 756998049151549440  
Tweet ID : 756939218950160384  
Tweet ID : 756651752796094464  
Tweet ID : 756526248105566208  
Tweet ID : 756303284449767430  
Tweet ID : 756288534030475264  
Tweet ID : 756275833623502848  
Tweet ID : 755955933503782912  
Tweet ID : 755206590534418437  
Tweet ID : 755110668769038337  
Tweet ID : 754874841593970688  
Tweet ID : 754856583969079297  
Tweet ID : 754747087846248448  
Tweet ID : 754482103782404096  
Tweet ID : 754449512966619136  
Tweet ID : 754120377874386944  
Error : [{'code': 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 753655901052166144  
Tweet ID : 753400000000000000

Tweet ID : 753420520834629632  
Tweet ID : 753398408988139520  
Tweet ID : 753375668877008896  
Tweet ID : 753298634498793472  
Tweet ID : 753294487569522689  
Tweet ID : 753039830821511168  
Tweet ID : 753026973505581056  
Tweet ID : 752932432744185856  
Tweet ID : 752917284578922496  
Tweet ID : 752701944171524096  
Tweet ID : 752682090207055872  
Tweet ID : 752660715232722944  
Tweet ID : 752568224206688256  
Tweet ID : 752519690950500352  
Tweet ID : 752334515931054080  
Tweet ID : 752309394570878976  
Tweet ID : 752173152931807232  
Tweet ID : 751950017322246144  
Tweet ID : 751937170840121344  
Tweet ID : 751830394383790080  
Tweet ID : 751793661361422336  
Tweet ID : 751598357617971201  
Tweet ID : 751583847268179968  
Tweet ID : 751538714308972544  
Tweet ID : 751456908746354688  
Tweet ID : 751251247299190784  
Tweet ID : 751205363882532864  
Tweet ID : 751132876104687617  
Tweet ID : 750868782890057730  
Tweet ID : 750719632563142656  
Tweet ID : 750506206503038976  
Tweet ID : 750429297815552001  
Tweet ID : 750383411068534784  
Tweet ID : 750381685133418496  
Tweet ID : 750147208377409536  
Tweet ID : 750132105863102464  
Tweet ID : 750117059602808832  
Tweet ID : 750101899009982464  
Tweet ID : 750086836815486976  
Tweet ID : 750071704093859840  
Tweet ID : 750056684286914561  
Tweet ID : 750041628174217216  
Tweet ID : 750026558547456000  
Tweet ID : 750011400160841729  
Tweet ID : 749996283729883136  
Tweet ID : 749981277374128128  
Tweet ID : 749774190421639168  
Tweet ID : 749417653287129088  
Tweet ID : 749403093750648834  
Tweet ID : 749395845976588288  
Tweet ID : 749317047558017024  
Tweet ID : 749075273010798592  
Tweet ID : 749064354620928000  
Tweet ID : 749036806121881602  
Tweet ID : 748977405889503236  
Tweet ID : 748932637671223296  
Tweet ID : 748705597323898880  
Tweet ID : 748699167502000129  
Tweet ID : 748692773788876800  
Tweet ID : 748575535303884801  
Tweet ID : 748568946752774144  
Tweet ID : 748346686624440324  
Tweet ID : 748337862848962560  
Tweet ID : 748324050481647620  
Tweet ID : 748307329658011649  
Tweet ID : 748220828303695873  
Tweet ID : 747963614829678593  
Tweet ID : 747933425676525569  
Tweet ID : 747885874273214464  
Tweet ID : 747844099428986880  
Tweet ID : 747816857231626240  
Tweet ID : 747651430853525504  
Tweet ID : 747648653817413632  
Tweet ID : 747600769478692864  
Tweet ID : 747594051852075008  
Tweet ID : 747512671126323200  
Tweet ID : 747461612269887489  
Tweet ID : 747433450710506400

Tweet ID : 74743945012596480  
 Tweet ID : 747242308580548608  
 Tweet ID : 747219827526344708  
 Tweet ID : 747204161125646336  
 Tweet ID : 747103485104099331  
 Tweet ID : 746906459439529985  
 Tweet ID : 746872823977771008  
 Tweet ID : 746818907684614144  
 Tweet ID : 746790600704425984  
 Tweet ID : 746757706116112384  
 Tweet ID : 746726898085036033  
 Tweet ID : 746542875501609625  
 Tweet ID : 746521445350707200  
 Tweet ID : 74650737934113997  
 Tweet ID : 746369468511756288  
 Tweet ID : 746131877086527488  
 Tweet ID : 746056683365994496  
 Tweet ID : 745789745784041472  
 Tweet ID : 745712589599014916  
 Tweet ID : 745433870967832576  
 Tweet ID : 745422732645535745  
 Tweet ID : 745314880350101504  
 Tweet ID : 745074613265149952  
 Tweet ID : 745057283344719872  
 Tweet ID : 744995568523612160  
 Tweet ID : 744971049620602880  
 Tweet ID : 744709971296780288  
 Tweet ID : 744334592493166593  
 Tweet ID : 744234799360020481  
 Tweet ID : 744223424764059648  
 Tweet ID : 743980027717509120  
 Tweet ID : 743895849529389061  
 Tweet ID : 743835915802583040  
 Tweet ID : 743609206067040256  
 Tweet ID : 743595368194129920  
 Tweet ID : 743545585370791937  
 Tweet ID : 743510151680958465  
 Tweet ID : 743253157753532416  
 Tweet ID : 7432225934720234624  
 Tweet ID : 743210557239623680  
 Tweet ID : 742534281772302336  
 Tweet ID : 742528092657332225  
 Tweet ID : 742465774154047488  
 Tweet ID : 742423170473463808  
 Tweet ID : 742385895052087300  
 Tweet ID : 742161199639494656  
 Tweet ID : 742150290887731712  
 Tweet ID : 741793263812808706  
 Tweet ID : 741743634094141440  
 Tweet ID : 741438259667034112  
 Tweet ID : 741303864243200000  
 Tweet ID : 741099773336379392  
 Tweet ID : 741067306818797568  
 Tweet ID : 740995100987665933  
 Tweet ID : 740711788199743490  
 Tweet ID : 740699697422163968  
 Tweet ID : 740676976021798912  
 Tweet ID : 740373189193256964  
 Tweet ID : 740365076218183684  
 Tweet ID : 740359016048689152  
 Tweet ID : 740214038584557568  
 Tweet ID : 739979191639244800  
 Tweet ID : 739932936087216128  
 Tweet ID : 739844404073074688  
 Tweet ID : 739623569819336705  
 Tweet ID : 739606147276148736  
 Tweet ID : 739544079319588864  
 Tweet ID : 739485634323156992  
 Tweet ID : 739238157791694849  
 Tweet ID : 738891149612572673  
 Tweet ID : 738885046782832640  
 Tweet ID : 738883359779196928  
 Tweet ID : 738537504001953792  
 Tweet ID : 738402415918125056  
 Tweet ID : 738184450748633089  
 Tweet ID : 738166403467907072  
 Tweet ID : 738156290900254722



Tweet ID : 737826014890496000  
Tweet ID : 737800304142471168  
Tweet ID : 737678689543020544  
Tweet ID : 737445876994609152  
Tweet ID : 737322739594330112  
Tweet ID : 737310737551491075  
Tweet ID : 736736130620620800  
Tweet ID : 736392552031657984  
Tweet ID : 736365877722001409  
Tweet ID : 736225175608430592  
Tweet ID : 736010884653420544  
Tweet ID : 735991953473572864  
Tweet ID : 735648611367784448  
Tweet ID : 735635087207878657  
Tweet ID : 735274964362878976  
Tweet ID : 735256018284875776  
Tweet ID : 735137028879360001  
Tweet ID : 734912297295085568  
Tweet ID : 734787690684657664  
Tweet ID : 734776360183431168  
Tweet ID : 734559631394082816  
Tweet ID : 733828123016450049  
Tweet ID : 733822306246479872  
Tweet ID : 733482008106668032  
Tweet ID : 733460102733135873  
Tweet ID : 733109485275860992  
Tweet ID : 732732193018155009  
Tweet ID : 732726085725589504  
Tweet ID : 732585889486888962  
Tweet ID : 732375214819057664  
Tweet ID : 732005617171337216  
Tweet ID : 731285275100512256  
Tweet ID : 731156023742988288  
Tweet ID : 730924654643314689  
Tweet ID : 730573383004487680  
Tweet ID : 730427201120833536  
Tweet ID : 730211855403241472  
Tweet ID : 730196704625098752  
Tweet ID : 729854734790754305  
Tweet ID : 729838605770891264  
Tweet ID : 729823566028484608  
Tweet ID : 729463711119904772  
Tweet ID : 729113531270991872  
Tweet ID : 728986383096946689  
Tweet ID : 728760639972315136  
Tweet ID : 728751179681943552  
Tweet ID : 728653952833728512  
Tweet ID : 728409960103686147  
Tweet ID : 728387165835677696  
Tweet ID : 728046963732717569  
Tweet ID : 728035342121635841  
Tweet ID : 728015554473250816  
Tweet ID : 727685679342333952  
Tweet ID : 727644517743104000  
Tweet ID : 727524757080539137  
Tweet ID : 727314416056803329  
Tweet ID : 727286334147182592  
Tweet ID : 727175381690781696  
Tweet ID : 727155742655025152  
Tweet ID : 726935089318363137  
Tweet ID : 726887082820554753  
Tweet ID : 726828223124897792  
Tweet ID : 726224900189511680  
Tweet ID : 725842289046749185  
Tweet ID : 725786712245440512  
Tweet ID : 725729321944506368  
Tweet ID : 725458796924002305  
Tweet ID : 724983749226668032  
Tweet ID : 724771698126512129  
Tweet ID : 724405726123311104  
Tweet ID : 724049859469295616  
Tweet ID : 724046343203856385  
Tweet ID : 724004602748780546  
Tweet ID : 723912936180330496  
Tweet ID : 723688335806480385  
Tweet ID : 723673163800948736  
Tweet ID : 723179728551723008

Tweet ID : 722974582966214656  
Tweet ID : 722613351520608256  
Tweet ID : 721503162398597120  
Tweet ID : 721001180231503872  
Tweet ID : 720785406564900865  
Tweet ID : 720775346191278080  
Tweet ID : 720415127506415616  
Tweet ID : 720389942216527872  
Tweet ID : 720340705894408192  
Tweet ID : 720059472081784833  
Tweet ID : 720043174954147842  
Tweet ID : 719991154352222208  
Tweet ID : 719704490224398336  
Tweet ID : 719551379208073216  
Tweet ID : 719367763014393856  
Tweet ID : 719339463458033665  
Tweet ID : 719332531645071360  
Tweet ID : 718971898235854848  
Tweet ID : 718939241951195136  
Tweet ID : 718631497683582976  
Tweet ID : 718613305783398402  
Tweet ID : 718540630683709445  
Tweet ID : 718460005985447936  
Tweet ID : 718454725339934721  
Tweet ID : 718246886998687744  
Tweet ID : 718234618122661888  
Tweet ID : 717841801130979328  
Tweet ID : 717790033953034240  
Tweet ID : 717537687239008257  
Tweet ID : 717428917016076293  
Tweet ID : 717421804990701568  
Tweet ID : 717047459982213120  
Tweet ID : 717009362452090881  
Tweet ID : 716802964044845056  
Tweet ID : 716791146589110272  
Tweet ID : 716730379797970944  
Tweet ID : 716447146686459905  
Tweet ID : 716439118184652801  
Tweet ID : 716285507865542656  
Tweet ID : 716080869887381504  
Tweet ID : 715928423106027520  
Tweet ID : 715758151270801409  
Tweet ID : 715733265223708672  
Tweet ID : 715704790270025728  
Tweet ID : 715696743237730304  
Tweet ID : 715680795826982913  
Tweet ID : 715360349751484417  
Tweet ID : 715342466308784130  
Tweet ID : 715220193576927233  
Tweet ID : 715200624753819648  
Tweet ID : 715009755312439296  
Tweet ID : 714982300363173890  
Tweet ID : 714962719905021952  
Tweet ID : 714957620017307648  
Tweet ID : 714631576617938945  
Tweet ID : 714606013974974464  
Tweet ID : 714485234495041536  
Tweet ID : 714258258790387713  
Tweet ID : 714251586676113411  
Tweet ID : 714214115368108032  
Tweet ID : 714141408463036416  
Tweet ID : 713919462244790272  
Tweet ID : 713909862279876608  
Tweet ID : 713900603437621249  
Tweet ID : 713761197720473600  
Tweet ID : 713411074226274305  
Tweet ID : 713177543487135744  
Tweet ID : 713175907180089344  
Tweet ID : 712809025985978368  
Tweet ID : 712717840512598017  
Tweet ID : 712668654853337088  
Tweet ID : 712438159032893441  
Tweet ID : 712309440758808576  
Tweet ID : 712097430750289920  
Tweet ID : 712092745624633345  
Tweet ID : 712085617388212225  
Tweet ID : 712065007010385924

Tweet ID : 711998809858043904  
Tweet ID : 711968124745228288  
Tweet ID : 711743778164514816  
Tweet ID : 711732680602345472  
Tweet ID : 711694788429553666  
Tweet ID : 711652651650457602  
Tweet ID : 711363825979756544  
Tweet ID : 711306686208872448  
Tweet ID : 711008018775851008  
Tweet ID : 710997087345876993  
Tweet ID : 710844581445812225  
Tweet ID : 710833117892898816  
Tweet ID : 710658690886586372  
Tweet ID : 710609963652087808  
Tweet ID : 710588934686908417  
Tweet ID : 710296729921429505  
Tweet ID : 710283270106132480  
Tweet ID : 710272297844797440  
Tweet ID : 710269109699739648  
Tweet ID : 710153181850935296  
Tweet ID : 710140971284037632  
Tweet ID : 710117014656950272  
Tweet ID : 709918798883774466  
Tweet ID : 709901256215666688  
Tweet ID : 709852847387627521  
Tweet ID : 709566166965075968  
Tweet ID : 709556954897764353  
Tweet ID : 709519240576036864  
Tweet ID : 709449600415961088  
Tweet ID : 709409458133323776  
Tweet ID : 709225125749587968  
Tweet ID : 709207347839836162  
Tweet ID : 709198395643068416  
Tweet ID : 709179584944730112  
Tweet ID : 709158332880297985  
Tweet ID : 709042156699303936  
Tweet ID : 708853462201716736  
Tweet ID : 708845821941387268  
Tweet ID : 708834316713893888  
Tweet ID : 708810915978854401  
Tweet ID : 708738143638450176  
Tweet ID : 708711088997666817  
Tweet ID : 708479650088034305  
Tweet ID : 708469915515297792  
Tweet ID : 708400866336894977  
Tweet ID : 708356463048204288  
Tweet ID : 708349470027751425  
Tweet ID : 708149363256774660  
Tweet ID : 708130923141795840  
Tweet ID : 708119489313951744  
Tweet ID : 708109389455101952  
Tweet ID : 708026248782585858  
Tweet ID : 707995814724026368  
Tweet ID : 707983188426153984  
Tweet ID : 707969809498152960  
Tweet ID : 707776935007539200  
Tweet ID : 707741517457260545  
Tweet ID : 707738799544082433  
Tweet ID : 707693576495472641  
Tweet ID : 707629649552134146  
Tweet ID : 707610948723478529  
Tweet ID : 707420581654872064  
Tweet ID : 707411934438625280  
Tweet ID : 707387676719185920  
Tweet ID : 707377100785885184  
Tweet ID : 707315916783140866  
Tweet ID : 707297311098011648  
Tweet ID : 707059547140169728  
Tweet ID : 707038192327901184  
Tweet ID : 707021089608753152  
Tweet ID : 707014260413456384  
Tweet ID : 706904523814649856  
Tweet ID : 706901761596989440  
Tweet ID : 706681918348251136  
Tweet ID : 706644897839910912  
Tweet ID : 706593038911545345  
Tweet ID : 706538006853918722

Tweet ID : 706516534877929472  
Tweet ID : 706346369204748288  
Tweet ID : 706310011488698368  
Tweet ID : 706291001778950144  
Tweet ID : 706265994973601792  
Tweet ID : 706169069255446529  
Tweet ID : 706166467411222528  
Tweet ID : 706153300320784384  
Tweet ID : 705975130514706432  
Tweet ID : 705970349788291072  
Tweet ID : 705898680587526145  
Tweet ID : 705786532653883392  
Tweet ID : 705591895322394625  
Tweet ID : 705475953783398401  
Tweet ID : 705442520700944385  
Tweet ID : 705428427625635840  
Tweet ID : 705239209544720384  
Tweet ID : 705223444686888960  
Tweet ID : 705102439679201280  
Tweet ID : 705066031337840642  
Tweet ID : 704871453724954624  
Tweet ID : 704859558691414016  
Tweet ID : 704847917308362754  
Tweet ID : 704819833553219584  
Tweet ID : 704761120771465216  
Tweet ID : 704499785726889984  
Tweet ID : 704491224099647488  
Tweet ID : 704480331685040129  
Tweet ID : 704364645503647744  
Tweet ID : 704347321748819968  
Tweet ID : 704134088924532736  
Tweet ID : 704113298707505153  
Tweet ID : 704054845121142784  
Tweet ID : 703774238772166656  
Tweet ID : 703769065844768768  
Tweet ID : 703631701117943808  
Tweet ID : 703611486317502464  
Tweet ID : 703425003149250560  
Tweet ID : 703407252292673536  
Tweet ID : 703382836347330562  
Tweet ID : 703356393781329922  
Tweet ID : 703268521220972544  
Tweet ID : 703079050210877440  
Tweet ID : 703041949650034688  
Tweet ID : 702932127499816960  
Tweet ID : 702899151802126337  
Tweet ID : 702684942141153280  
Tweet ID : 702671118226825216  
Tweet ID : 702598099714314240  
Tweet ID : 702539513671897089  
Tweet ID : 702332542343577600  
Tweet ID : 702321140488925184  
Tweet ID : 702276748847800320  
Tweet ID : 702217446468493312  
Tweet ID : 701981390485725185  
Tweet ID : 701952816642965504  
Tweet ID : 701889187134500865  
Tweet ID : 701805642395348998  
Tweet ID : 701601587219795968  
Tweet ID : 701570477911896070  
Tweet ID : 701545186879471618  
Tweet ID : 701214700881756160  
Tweet ID : 700890391244103680  
Tweet ID : 700864154249383937  
Tweet ID : 700847567345688576  
Tweet ID : 700796979434098688  
Tweet ID : 700747788515020802  
Tweet ID : 700518061187723268  
Tweet ID : 700505138482569216  
Tweet ID : 700462010979500032  
Tweet ID : 700167517596164096  
Tweet ID : 700151421916807169  
Tweet ID : 700143752053182464  
Tweet ID : 700062718104104960  
Tweet ID : 700029284593901568  
Tweet ID : 700002074055016451  
Tweet ID : 699801817392291840

Tweet ID : 699788877217865730  
Tweet ID : 699779630832685056  
Tweet ID : 699775878809702401  
Tweet ID : 699691744225525762  
Tweet ID : 699446877801091073  
Tweet ID : 699434518667751424  
Tweet ID : 699423671849451520  
Tweet ID : 699413908797464576  
Tweet ID : 699370870310113280  
Tweet ID : 699323444782047232  
Tweet ID : 699088579889332224  
Tweet ID : 699079609774645248  
Tweet ID : 699072405256409088  
Tweet ID : 699060279947165696  
Tweet ID : 699036661657767936  
Tweet ID : 698989035503689728  
Tweet ID : 698953797952008193  
Tweet ID : 698907974262222848  
Tweet ID : 698710712454139905  
Tweet ID : 698703483621523456  
Tweet ID : 698635131305795584  
Tweet ID : 698549713696649216  
Tweet ID : 698355670425473025  
Tweet ID : 698342080612007937  
Tweet ID : 698262614669991936  
Tweet ID : 698195409219559425  
Tweet ID : 698178924120031232  
Tweet ID : 697995514407682048  
Tweet ID : 697990423684476929  
Tweet ID : 697943111201378304  
Tweet ID : 697881462549430272  
Tweet ID : 697630435728322560  
Tweet ID : 697616773278015490  
Tweet ID : 697596423848730625  
Tweet ID : 697575480820686848  
Tweet ID : 697516214579523584  
Tweet ID : 697482927769255936  
Tweet ID : 697463031882764288  
Tweet ID : 697270446429966336  
Tweet ID : 697259378236399616  
Tweet ID : 697255105972801536  
Tweet ID : 697242256848379904  
Tweet ID : 696900204696625153  
Tweet ID : 696894894812565505  
Tweet ID : 696886256886657024  
Tweet ID : 696877980375769088  
Tweet ID : 696754882863349760  
Tweet ID : 696744641916489729  
Tweet ID : 696713835009417216  
Tweet ID : 696518437233913856  
Tweet ID : 696490539101908992  
Tweet ID : 696488710901260288  
Tweet ID : 696405997980676096  
Tweet ID : 696100768806522880  
Tweet ID : 695816827381944320  
Tweet ID : 695794761660297217  
Tweet ID : 695767669421768709  
Tweet ID : 695629776980148225  
Tweet ID : 695446424020918272  
Tweet ID : 695409464418041856  
Tweet ID : 695314793360662529  
Tweet ID : 695095422348574720  
Tweet ID : 695074328191332352  
Tweet ID : 695064344191721472  
Tweet ID : 695051054296211456  
Tweet ID : 694925794720792577  
Tweet ID : 694905863685980160  
Tweet ID : 694669722378485760  
Tweet ID : 694356675654983680  
Tweet ID : 694352839993344000  
Tweet ID : 694342028726001664  
Tweet ID : 694329668942569472  
Tweet ID : 694206574471057408  
Tweet ID : 694183373896572928  
Tweet ID : 694001791655137281  
Tweet ID : 693993230313091072  
Tweet ID : 693942351086120961

Tweet ID : 693647888581312512  
Tweet ID : 693644216740769793  
Tweet ID : 693642232151285760  
Tweet ID : 693629975228977152  
Tweet ID : 693622659251335168  
Tweet ID : 693590843962331137  
Tweet ID : 693582294167244802  
Tweet ID : 693486665285931008  
Tweet ID : 693280720173801472  
Tweet ID : 693267061318012928  
Tweet ID : 693262851218264065  
Tweet ID : 693231807727280129  
Tweet ID : 693155686491000832  
Tweet ID : 693109034023534592  
Tweet ID : 693095443459342336  
Tweet ID : 692919143163629568  
Tweet ID : 692905862751522816  
Tweet ID : 692901601640583168  
Tweet ID : 692894228850999298  
Tweet ID : 692828166163931137  
Tweet ID : 692752401762250755  
Tweet ID : 692568918515392513  
Tweet ID : 692535307825213440  
Tweet ID : 692530551048294401  
Tweet ID : 692423280028966913  
Tweet ID : 692417313023332352  
Tweet ID : 692187005137076224  
Tweet ID : 692158366030913536  
Tweet ID : 692142790915014657  
Tweet ID : 692041934689402880  
Tweet ID : 692017291282812928  
Tweet ID : 691820333922455552  
Tweet ID : 691793053716221953  
Tweet ID : 691756958957883396  
Tweet ID : 691675652215414786  
Tweet ID : 691483041324204033  
Tweet ID : 691459709405118465  
Tweet ID : 691444869282295808  
Tweet ID : 691416866452082688  
Tweet ID : 691321916024623104  
Tweet ID : 691096613310316544  
Tweet ID : 691090071332753408  
Tweet ID : 690989312272396288  
Tweet ID : 690959652130045952  
Tweet ID : 690938899477221376  
Tweet ID : 690932576555528194  
Tweet ID : 690735892932222976  
Tweet ID : 690728923253055490  
Tweet ID : 690690673629138944  
Tweet ID : 690649993829576704  
Tweet ID : 690607260360429569  
Tweet ID : 690597161306841088  
Tweet ID : 690400367696297985  
Tweet ID : 690374419777196032  
Tweet ID : 690360449368465409  
Tweet ID : 690348396616552449  
Tweet ID : 690248561355657216  
Tweet ID : 690021994562220032  
Tweet ID : 690015576308211712  
Tweet ID : 690005060500217858  
Tweet ID : 689999384604450816  
Tweet ID : 689993469801164801  
Tweet ID : 68997755533848577  
Tweet ID : 689905486972461056  
Tweet ID : 689877686181715968  
Tweet ID : 689835978131935233  
Tweet ID : 689661964914655233  
Tweet ID : 689659372465688576  
Tweet ID : 689623661272240129  
Tweet ID : 689599056876867584  
Tweet ID : 689557536375177216  
Tweet ID : 689517482558820352  
Tweet ID : 689289219123089408  
Tweet ID : 689283819090870273  
Tweet ID : 689280876073582592  
Tweet ID : 689275259254616065  
Tweet ID : 689255633275777024

Tweet ID : 689154315265683456  
Tweet ID : 689143371370250240  
Tweet ID : 688916208532455424  
Tweet ID : 688908934925697024  
Tweet ID : 688898160958271489  
Tweet ID : 688894073864884227  
Tweet ID : 688828561667567616  
Tweet ID : 688804835492233216  
Tweet ID : 688789766343622656  
Tweet ID : 688547210804498433  
Tweet ID : 688519176466644993  
Tweet ID : 688385280030670848  
Tweet ID : 688211956440801280  
Tweet ID : 688179443353796608  
Tweet ID : 688116655151435777  
Tweet ID : 688064179421470721  
Tweet ID : 687841446767013888  
Tweet ID : 687826841265172480  
Tweet ID : 687818504314159109  
Tweet ID : 687807801670897665  
Tweet ID : 687732144991551489  
Tweet ID : 687704180304273409  
Tweet ID : 687664829264453632  
Tweet ID : 687494652870668288  
Tweet ID : 687480748861947905  
Tweet ID : 687476254459715584  
Tweet ID : 687460506001633280  
Tweet ID : 687399393394311168  
Tweet ID : 687317306314240000  
Tweet ID : 687312378585812992  
Tweet ID : 687127927494963200  
Tweet ID : 687124485711986689  
Tweet ID : 687109925361856513  
Tweet ID : 687102708889812993  
Tweet ID : 687096057537363968  
Tweet ID : 686947101016735744  
Tweet ID : 686760001961103360  
Tweet ID : 686749460672679938  
Tweet ID : 686730991906516992  
Tweet ID : 686683045143953408  
Tweet ID : 686618349602762752  
Tweet ID : 686606069955735556  
Tweet ID : 686394059078897668  
Tweet ID : 686386521809772549

Rate limit reached. Sleeping for: 509

Tweet ID : 686377065986265092  
Tweet ID : 686358356425093120  
Tweet ID : 686286779679375361  
Tweet ID : 686050296934563840  
Tweet ID : 686035780142297088  
Tweet ID : 686034024800862208  
Tweet ID : 686007916130873345  
Tweet ID : 686003207160610816  
Tweet ID : 685973236358713344  
Tweet ID : 685943807276412928  
Tweet ID : 685906723014619143  
Tweet ID : 685681090388975616  
Tweet ID : 685667379192414208  
Tweet ID : 685663452032069632  
Tweet ID : 685641971164143616  
Tweet ID : 685547936038666240  
Tweet ID : 685532292383666176  
Tweet ID : 685325112850124800  
Tweet ID : 685321586178670592  
Tweet ID : 685315239903100929  
Tweet ID : 685307451701334016  
Tweet ID : 685268753634967552  
Tweet ID : 685198997565345792  
Tweet ID : 685169283572338688  
Tweet ID : 684969860808454144  
Tweet ID : 684959798585110529  
Tweet ID : 684940049151070208  
Tweet ID : 684926975086034944  
Tweet ID : 684914660081053696

Tweet ID : 684902183876321280  
Tweet ID : 684880619965411328  
Tweet ID : 684830982659280897  
Tweet ID : 684800227459624960  
Tweet ID : 684594889858887680  
Tweet ID : 684588130326986752  
Tweet ID : 684567543613382656  
Tweet ID : 684538444857667585  
Tweet ID : 684481074559381504  
Tweet ID : 684460069371654144  
Tweet ID : 684241637099323392  
Tweet ID : 684225744407494656  
Tweet ID : 684222868335505415  
Tweet ID : 684200372118904832  
Tweet ID : 684195085588783105  
Tweet ID : 684188786104872960  
Tweet ID : 684177701129875456  
Tweet ID : 684147889187209216  
Tweet ID : 684122891630342144  
Tweet ID : 684097758874210310  
Tweet ID : 683857920510050305  
Tweet ID : 683852578183077888  
Tweet ID : 683849932751646720  
Tweet ID : 683834909291606017  
Tweet ID : 683828599284170753  
Tweet ID : 683773439333797890  
Tweet ID : 683742671509258241  
Tweet ID : 683515932363329536  
Tweet ID : 683498322573824003  
Tweet ID : 683481228088049664  
Tweet ID : 683462770029932544  
Tweet ID : 683449695444799489  
Tweet ID : 683391852557561860  
Tweet ID : 683357973142474752  
Tweet ID : 683142553609318400  
Tweet ID : 683111407806746624  
Tweet ID : 683098815881154561  
Tweet ID : 683078886620553216  
Tweet ID : 683030066213818368  
Tweet ID : 682962037429899265  
Tweet ID : 682808988178739200  
Tweet ID : 682788441537560576  
Tweet ID : 682750546109968385  
Tweet ID : 682697186228989953  
Tweet ID : 682662431982772225  
Tweet ID : 682638830361513985  
Tweet ID : 682429480204398592  
Tweet ID : 682406705142087680  
Tweet ID : 682393905736888321  
Tweet ID : 682389078323662849  
Tweet ID : 682303737705140231  
Tweet ID : 682259524040966145  
Tweet ID : 682242692827447297  
Tweet ID : 682088079302213632  
Tweet ID : 682059653698686977  
Tweet ID : 682047327939461121  
Tweet ID : 682032003584274432  
Tweet ID : 682003177596559360  
Tweet ID : 681981167097122816  
Tweet ID : 681891461017812993  
Tweet ID : 681694085539872773  
Tweet ID : 681679526984871937  
Tweet ID : 681654059175129088  
Tweet ID : 681610798867845120  
Tweet ID : 681579835668455424  
Tweet ID : 681523177663676416  
Tweet ID : 681340665377193984  
Tweet ID : 681339448655802368  
Tweet ID : 681320187870711809  
Tweet ID : 681302363064414209  
Tweet ID : 681297372102656000  
Tweet ID : 681281657291280384  
Tweet ID : 681261549936340994  
Tweet ID : 681242418453299201  
Tweet ID : 681231109724700672  
Tweet ID : 681193455364796417  
Tweet ID : 680970795137544192



Tweet ID : 680959110691590145  
Tweet ID : 680940246314430465  
Tweet ID : 680934982542561280  
Tweet ID : 680913438424612864  
Tweet ID : 680889648562991104  
Tweet ID : 680836378243002368  
Tweet ID : 680805554198020098  
Tweet ID : 680801747103793152  
Tweet ID : 680798457301471234  
Tweet ID : 680609293079592961  
Tweet ID : 680583894916304897  
Tweet ID : 680497766108381184  
Tweet ID : 680494726643068929  
Tweet ID : 680473011644985345  
Tweet ID : 680440374763077632  
Tweet ID : 680221482581123072  
Tweet ID : 680206703334408192  
Tweet ID : 680191257256136705  
Tweet ID : 680176173301628928  
Tweet ID : 680161097740095489  
Tweet ID : 680145970311643136  
Tweet ID : 680130881361686529  
Tweet ID : 680115823365742593  
Tweet ID : 680100725817409536  
Tweet ID : 680085611152338944  
Tweet ID : 680070545539371008  
Error : [{"code": 144, 'message': 'No status found with that ID.'}]  
Tweet ID : 679877062409191424  
Tweet ID : 679872969355714560  
Tweet ID : 679862121895714818  
Tweet ID : 679854723806179328  
Tweet ID : 679844490799091713  
Tweet ID : 679828447187857408  
Tweet ID : 679777920601223168  
Tweet ID : 679736210798047232  
Tweet ID : 679729593985699840  
Tweet ID : 679722016581222400  
Tweet ID : 679530280114372609  
Tweet ID : 679527802031484928  
Tweet ID : 679511351870550016  
Tweet ID : 679503373272485890  
Tweet ID : 679475951516934144  
Tweet ID : 679462823135686656  
Tweet ID : 679405845277462528  
Tweet ID : 679158373988876288  
Tweet ID : 679148763231985668  
Tweet ID : 679132435750195208  
Tweet ID : 679111216690831360  
Tweet ID : 679062614270468097  
Tweet ID : 679047485189439488  
Tweet ID : 679001094530465792  
Tweet ID : 678991772295516161  
Tweet ID : 678969228704284672  
Tweet ID : 678800283649069056  
Tweet ID : 678798276842360832  
Tweet ID : 678774928607469569  
Tweet ID : 678767140346941444  
Tweet ID : 678764513869611008  
Tweet ID : 678755239630127104  
Tweet ID : 678740035362037760  
Tweet ID : 678708137298427904  
Tweet ID : 678675843183484930  
Tweet ID : 678643457146150913  
Tweet ID : 678446151570427904  
Tweet ID : 678424312106393600  
Tweet ID : 678410210315247616  
Tweet ID : 678399652199309312  
Tweet ID : 678396796259975168  
Tweet ID : 678389028614488064  
Tweet ID : 678380236862578688  
Tweet ID : 678341075375947776  
Tweet ID : 678334497360859136  
Tweet ID : 678278586130948096  
Tweet ID : 678255464182861824  
Tweet ID : 678023323247357953  
Tweet ID : 678021115718029313  
Tweet ID : 677861670166224887

tweet ID : 677918531514703872  
Tweet ID : 677895101218201600  
Tweet ID : 677716515794329600  
Tweet ID : 677700003327029250  
Tweet ID : 677698403548192770  
Tweet ID : 677687604918272002  
Tweet ID : 677673981332312066  
Tweet ID : 677662372920729601  
Tweet ID : 677644091929329666  
Tweet ID : 677573743309385728  
Tweet ID : 677565715327688705  
Tweet ID : 677557565589463040  
Tweet ID : 677547928504967168  
Tweet ID : 677530072887205888  
Tweet ID : 677335745548390400  
Tweet ID : 677334615166730240  
Tweet ID : 677331501395156992  
Tweet ID : 677328882937298944  
Tweet ID : 677314812125323265  
Tweet ID : 677301033169788928  
Tweet ID : 677269281705472000  
Tweet ID : 677228873407442944  
Tweet ID : 677187300187611136  
Tweet ID : 676975532580409345  
Tweet ID : 676957860086095872  
Tweet ID : 676949632774234114  
Tweet ID : 676948236477857792  
Tweet ID : 676946864479084545  
Tweet ID : 676942428000112642  
Tweet ID : 676936541936185344  
Tweet ID : 676916996760600576  
Tweet ID : 676897532954456065  
Tweet ID : 676864501615042560  
Tweet ID : 676821958043033607  
Tweet ID : 676819651066732545  
Tweet ID : 676811746707918848  
Tweet ID : 676776431406465024  
Tweet ID : 676617503762681856  
Tweet ID : 676613908052996102  
Tweet ID : 676606785097199616  
Tweet ID : 676603393314578432  
Tweet ID : 676593408224403456  
Tweet ID : 676590572941893632  
Tweet ID : 676588346097852417  
Tweet ID : 676582956622721024  
Tweet ID : 676575501977128964  
Tweet ID : 676533798876651520  
Tweet ID : 676496375194980353  
Tweet ID : 676470639084101634  
Tweet ID : 676440007570247681  
Tweet ID : 676430933382295552  
Tweet ID : 676263575653122048  
Tweet ID : 676237365392908289  
Tweet ID : 676219687039057920  
Tweet ID : 676215927814406144  
Tweet ID : 676191832485810177  
Tweet ID : 676146341966438401  
Tweet ID : 676121918416756736  
Tweet ID : 676101918813499392  
Tweet ID : 676098748976615425  
Tweet ID : 676089483918516224  
Tweet ID : 675898130735476737  
Tweet ID : 675891555769696257  
Tweet ID : 675888385639251968  
Tweet ID : 675878199931371520  
Tweet ID : 675870721063669760  
Tweet ID : 675853064436391936  
Tweet ID : 675849018447167488  
Tweet ID : 675845657354215424  
Tweet ID : 675822767435051008  
Tweet ID : 675820929667219457  
Tweet ID : 675798442703122432  
Tweet ID : 675781562965868544  
Tweet ID : 675740360753160193  
Tweet ID : 675710890956750848  
Tweet ID : 675707330206547968  
Tweet ID : 675706620471360020

Tweet ID : 67570663941188032  
Tweet ID : 675534494439489536  
Tweet ID : 675531475945709568  
Tweet ID : 675522403582218240  
Tweet ID : 675517828909424640  
Tweet ID : 675501075957489664  
Tweet ID : 675497103322386432  
Tweet ID : 675489971617296384  
Tweet ID : 675483430902214656  
Tweet ID : 675432746517426176  
Tweet ID : 675372240448454658  
Tweet ID : 675362609739206656  
Tweet ID : 675354435921575936  
Tweet ID : 675349384339542016  
Tweet ID : 675334060156301312  
Tweet ID : 675166823650848770  
Tweet ID : 675153376133427200  
Tweet ID : 675149409102012420  
Tweet ID : 675147105808306176  
Tweet ID : 675146535592706048  
Tweet ID : 675145476954566656  
Tweet ID : 675135153782571009  
Tweet ID : 675113801096802304  
Tweet ID : 675111688094527488  
Tweet ID : 675109292475830276  
Tweet ID : 675047298674663426  
Tweet ID : 675015141583413248  
Tweet ID : 675006312288268288  
Tweet ID : 675003128568291329  
Tweet ID : 674999807681908736  
Tweet ID : 674805413498527744  
Tweet ID : 674800520222154752  
Tweet ID : 674793399141146624  
Tweet ID : 674790488185167872  
Tweet ID : 674788554665512960  
Tweet ID : 674781762103414784  
Tweet ID : 674774481756377088  
Tweet ID : 674767892831932416  
Tweet ID : 674764817387900928  
Tweet ID : 674754018082705410  
Tweet ID : 674752233200820224  
Tweet ID : 674743008475090944  
Tweet ID : 674742531037511680  
Tweet ID : 674739953134403584  
Tweet ID : 674737130913071104  
Tweet ID : 674690135443775488  
Tweet ID : 674670581682434048  
Tweet ID : 674664755118911488  
Tweet ID : 674646392044941312  
Tweet ID : 674644256330530816  
Tweet ID : 674638615994089473  
Tweet ID : 674632714662858753  
Tweet ID : 674606911342424069  
Tweet ID : 674468880899788800  
Tweet ID : 674447403907457024  
Tweet ID : 674436901579923456  
Tweet ID : 674422304705744896  
Tweet ID : 674416750885273600  
Tweet ID : 674410619106390016  
Tweet ID : 674394782723014656  
Tweet ID : 674372068062928900  
Tweet ID : 674330906434379776  
Tweet ID : 674318007229923329  
Tweet ID : 674307341513269249  
Tweet ID : 674291837063053312  
Tweet ID : 674271431610523648  
Tweet ID : 674269164442398721  
Tweet ID : 674265582246694913  
Tweet ID : 674262580978937856  
Tweet ID : 674255168825880576  
Tweet ID : 674082852460433408  
Tweet ID : 674075285688614912  
Tweet ID : 674063288070742018  
Tweet ID : 674053186244734976  
Tweet ID : 674051556661161984  
Tweet ID : 674045139690631169  
Tweet ID : 674042553264685056  
Tweet ID : 674038822568732717

Tweet ID : 674038233588123111  
Tweet ID : 674036086168010753  
Tweet ID : 674024893172875264  
Tweet ID : 674019345211760640  
Tweet ID : 674014384960745472  
Tweet ID : 674008982932058114  
Tweet ID : 673956914389192708  
Tweet ID : 673919437611909120  
Tweet ID : 673906403526995968  
Tweet ID : 673887867907739649  
Tweet ID : 673716320723169284  
Tweet ID : 673715861853720576  
Tweet ID : 673711475735838725  
Tweet ID : 673709992831262724  
Tweet ID : 673708611235921920  
Tweet ID : 673707060090052608  
Tweet ID : 673705679337693185  
Tweet ID : 673700254269775872  
Tweet ID : 673697980713705472  
Tweet ID : 673689733134946305  
Tweet ID : 673688752737402881  
Tweet ID : 673686845050527744  
Tweet ID : 673680198160809984  
Tweet ID : 673662677122719744  
Tweet ID : 673656262056419329  
Tweet ID : 673636718965334016  
Tweet ID : 673612854080196609  
Tweet ID : 673583129559498752  
Tweet ID : 673580926094458881  
Tweet ID : 673576835670777856  
Tweet ID : 673363615379013632  
Tweet ID : 673359818736984064  
Tweet ID : 673355879178194945  
Tweet ID : 673352124999274496  
Tweet ID : 673350198937153538  
Tweet ID : 673345638550134785  
Tweet ID : 673343217010679808  
Tweet ID : 673342308415348736  
Tweet ID : 673320132811366400  
Tweet ID : 673317986296586240  
Tweet ID : 673295268553605120  
Tweet ID : 673270968295534593  
Tweet ID : 673240798075449344  
Tweet ID : 673213039743795200  
Tweet ID : 673148804208660480  
Tweet ID : 672997845381865473  
Tweet ID : 672995267319328768  
Tweet ID : 672988786805112832  
Tweet ID : 672984142909456390  
Tweet ID : 672980819271634944  
Tweet ID : 672975131468300288  
Tweet ID : 672970152493887488  
Tweet ID : 672968025906282496  
Tweet ID : 672964561327235073  
Tweet ID : 672902681409806336  
Tweet ID : 672898206762672129  
Tweet ID : 672884426393653248  
Tweet ID : 672877615439593473  
Tweet ID : 672834301050937345  
Tweet ID : 672828477930868736  
Tweet ID : 672640509974827008  
Tweet ID : 672622327801233409  
Tweet ID : 672614745925664768  
Tweet ID : 672609152938721280  
Tweet ID : 672604026190569472  
Tweet ID : 672594978741354496  
Tweet ID : 672591762242805761  
Tweet ID : 672591271085670400  
Tweet ID : 672538107540070400  
Tweet ID : 672523490734551040  
Tweet ID : 672488522314567680  
Tweet ID : 672482722825261057  
Tweet ID : 672481316919734272  
Tweet ID : 672475084225949696  
Tweet ID : 672466075045466113  
Tweet ID : 672272411274932228  
Tweet ID : 672267570918129665  
-- -- 55886155575881558861

Tweet ID : 672264251789176834  
Tweet ID : 672256522047614977  
Tweet ID : 672254177670729728  
Tweet ID : 672248013293752320  
Tweet ID : 672245253877968896  
Tweet ID : 672239279297454080  
Tweet ID : 672231046314901505  
Tweet ID : 672222792075620352  
Tweet ID : 672205392827572224  
Tweet ID : 672169685991993344  
Tweet ID : 672160042234327040  
Tweet ID : 672139350159835138  
Tweet ID : 672125275208069120  
Tweet ID : 672095186491711488  
Tweet ID : 672082170312290304  
Tweet ID : 672068090318987265  
Tweet ID : 671896809300709376  
Tweet ID : 671891728106971137  
Tweet ID : 671882082306625538  
Tweet ID : 671879137494245376  
Tweet ID : 671874878652489728  
Tweet ID : 671866342182637568  
Tweet ID : 671855973984772097  
Tweet ID : 671789708968640512  
Tweet ID : 671768281401958400  
Tweet ID : 671763349865160704  
Tweet ID : 671744970634719232  
Tweet ID : 671743150407421952  
Tweet ID : 671735591348891648  
Tweet ID : 671729906628341761  
Tweet ID : 671561002136281088  
Tweet ID : 671550332464455680  
Tweet ID : 671547767500775424  
Tweet ID : 671544874165002241  
Tweet ID : 671542985629241344  
Tweet ID : 671538301157904385  
Tweet ID : 671536543010570240  
Tweet ID : 671533943490011136  
Tweet ID : 671528761649688577  
Tweet ID : 671520732782923777  
Tweet ID : 671518598289059840  
Tweet ID : 671511350426865664  
Tweet ID : 671504605491109889  
Tweet ID : 671497587707535361  
Tweet ID : 671488513339211776  
Tweet ID : 671486386088865792  
Tweet ID : 671485057807351808  
Tweet ID : 671390180817915904  
Tweet ID : 671362598324076544  
Tweet ID : 671357843010908160  
Tweet ID : 671355857343524864  
Tweet ID : 671347597085433856  
Tweet ID : 671186162933985280  
Tweet ID : 671182547775299584  
Tweet ID : 671166507850801152  
Tweet ID : 671163268581498880  
Tweet ID : 671159727754231808  
Tweet ID : 671154572044468225  
Tweet ID : 671151324042559489  
Tweet ID : 671147085991960577  
Tweet ID : 671141549288370177  
Tweet ID : 671138694582165504  
Tweet ID : 671134062904504320  
Tweet ID : 671122204919246848  
Tweet ID : 671115716440031232  
Tweet ID : 671109016219725825  
Tweet ID : 670995969505435648  
Tweet ID : 670842764863651840  
Tweet ID : 670840546554966016  
Tweet ID : 670838202509447168  
Tweet ID : 670833812859932673  
Tweet ID : 670832455012716544  
Tweet ID : 670826280409919488  
Tweet ID : 670823764196741120  
Tweet ID : 670822709593571328  
Tweet ID : 670815497391357952  
Tweet ID : 670811965569282048

Tweet ID : 670807719151067136  
Tweet ID : 670804601705242624  
Tweet ID : 670803562457407488  
Tweet ID : 670797304698376195  
Tweet ID : 670792680469889025  
Tweet ID : 670789397210615808  
Tweet ID : 670786190031921152  
Tweet ID : 670783437142401025  
Tweet ID : 670782429121134593  
Tweet ID : 670780561024270336  
Tweet ID : 670778058496974848  
Tweet ID : 670764103623966721  
Tweet ID : 670755717859713024  
Tweet ID : 670733412878163972  
Tweet ID : 670727704916926465  
Tweet ID : 670717338665226240  
Tweet ID : 670704688707301377  
Tweet ID : 670691627984359425  
Tweet ID : 670679630144274432  
Tweet ID : 670676092097810432  
Tweet ID : 670668383499735048  
Tweet ID : 670474236058800128  
Tweet ID : 670468609693655041  
Tweet ID : 670465786746662913  
Tweet ID : 670452855871037440  
Tweet ID : 670449342516494336  
Tweet ID : 670444955656130560  
Tweet ID : 670442337873600512  
Tweet ID : 670435821946826752  
Tweet ID : 670434127938719744  
Tweet ID : 670433248821026816  
Tweet ID : 670428280563085312  
Tweet ID : 670427002554466305  
Tweet ID : 670421925039075328  
Tweet ID : 670420569653809152  
Tweet ID : 670417414769758208  
Tweet ID : 670411370698022913  
Tweet ID : 670408998013820928  
Tweet ID : 670403879788544000  
Tweet ID : 670385711116361728  
Tweet ID : 670374371102445568  
Tweet ID : 670361874861563904  
Tweet ID : 670338931251150849  
Tweet ID : 670319130621435904  
Tweet ID : 670303360680108032  
Tweet ID : 670290420111441920  
Tweet ID : 670093938074779648  
Tweet ID : 670086499208155136  
Tweet ID : 670079681849372674  
Tweet ID : 670073503555706880  
Tweet ID : 670069087419133954  
Tweet ID : 670061506722140161  
Tweet ID : 670055038660800512  
Tweet ID : 670046952931721218  
Tweet ID : 670040295598354432  
Tweet ID : 670037189829525505  
Tweet ID : 670003130994700288  
Tweet ID : 669993076832759809  
Tweet ID : 669972011175813120  
Tweet ID : 669970042633789440  
Tweet ID : 669942763794931712  
Tweet ID : 669926384437997569  
Tweet ID : 669923323644657664  
Tweet ID : 669753178989142016  
Tweet ID : 669749430875258880  
Tweet ID : 669684865554620416  
Tweet ID : 669683899023405056  
Tweet ID : 669682095984410625  
Tweet ID : 669680153564442624  
Tweet ID : 669661792646373376  
Tweet ID : 669625907762618368  
Tweet ID : 669603084620980224  
Tweet ID : 669597912108789760  
Tweet ID : 669583744538451968  
Tweet ID : 669573570759163904  
Tweet ID : 669571471778410496  
Tweet ID : 669567591774625800

Tweet ID : 669564461267722241  
Tweet ID : 669393256313184256  
Tweet ID : 669375718304980992  
Tweet ID : 669371483794317312  
Tweet ID : 669367896104181761  
Tweet ID : 669363888236994561  
Tweet ID : 669359674819481600  
Tweet ID : 669354382627049472  
Tweet ID : 669353438988365824  
Tweet ID : 669351434509529089  
Tweet ID : 669328503091937280  
Tweet ID : 669327207240699904  
Tweet ID : 669324657376567296  
Tweet ID : 669216679721873412  
Tweet ID : 669214165781868544  
Tweet ID : 669203728096960512  
Tweet ID : 669037058363662336  
Tweet ID : 669015743032369152  
Tweet ID : 669006782128353280  
Tweet ID : 669000397445533696  
Tweet ID : 668994913074286592  
Tweet ID : 668992363537309700  
Tweet ID : 668989615043424256  
Tweet ID : 668988183816871936  
Tweet ID : 668986018524233728  
Tweet ID : 668981893510119424  
Tweet ID : 668979806671884288  
Tweet ID : 668975677807423489  
Tweet ID : 668967877119254528  
Tweet ID : 668960084974809088  
Tweet ID : 668955713004314625  
Tweet ID : 668932921458302977  
Tweet ID : 668902994700836864  
Tweet ID : 668892474547511297  
Tweet ID : 668872652652679168  
Tweet ID : 668852170888998912  
Tweet ID : 668826086256599040  
Tweet ID : 668815180734689280  
Tweet ID : 668779399630725120  
Tweet ID : 668655139528511488  
Tweet ID : 668645506898350081  
Tweet ID : 668643542311546881  
Tweet ID : 668641109086707712  
Tweet ID : 668636665813057536  
Tweet ID : 668633411083464705  
Tweet ID : 668631377374486528  
Tweet ID : 668627278264475648  
Tweet ID : 668625577880875008  
Tweet ID : 668623201287675904  
Tweet ID : 668620235289837568  
Tweet ID : 668614819948453888  
Tweet ID : 668587383441514497  
Tweet ID : 668567822092664832  
Tweet ID : 668544745690562560  
Tweet ID : 668542336805281792  
Tweet ID : 668537837512433665  
Tweet ID : 668528771708952576  
Tweet ID : 668507509523615744  
Tweet ID : 668496999348633600  
Tweet ID : 668484198282485761  
Tweet ID : 668480044826800133  
Tweet ID : 668466899341221888  
Tweet ID : 668297328638447616  
Tweet ID : 668291999406125056  
Tweet ID : 668286279830867968  
Tweet ID : 668274247790391296  
Tweet ID : 668268907921326080  
Tweet ID : 668256321989451776  
Tweet ID : 668248472370458624  
Tweet ID : 668237644992782336  
Tweet ID : 668226093875376128  
Tweet ID : 668221241640230912  
Tweet ID : 668204964695683073  
Tweet ID : 668190681446379520  
Tweet ID : 668171859951755264  
Tweet ID : 668154635664932864  
Tweet ID : 668142349051129856

Tweet ID : 668113020489474048  
Tweet ID : 667937095915278337  
Tweet ID : 667924896115245057  
Tweet ID : 667915453470232577  
Tweet ID : 667911425562669056  
Tweet ID : 667902449697558528  
Tweet ID : 667886921285246976  
Tweet ID : 667885044254572545  
Tweet ID : 667878741721415682  
Tweet ID : 667873844930215936  
Tweet ID : 667866724293877760  
Tweet ID : 667861340749471744  
Tweet ID : 667832474953625600  
Tweet ID : 667806454573760512  
Tweet ID : 667801013445750784  
Tweet ID : 667793409583771648  
Tweet ID : 667782464991965184  
Tweet ID : 667773195014021121  
Tweet ID : 667766675769573376  
Tweet ID : 667728196545200128  
Tweet ID : 667724302356258817  
Tweet ID : 667550904950915073  
Tweet ID : 667550882905632768  
Tweet ID : 667549055577362432  
Tweet ID : 667546741521195010  
Tweet ID : 667544320556335104  
Tweet ID : 667538891197542400  
Tweet ID : 667534815156183040  
Tweet ID : 667530908589760512  
Tweet ID : 667524857454854144  
Tweet ID : 667517642048163840  
Tweet ID : 667509364010450944  
Tweet ID : 667502640335572993  
Tweet ID : 667495797102141441  
Tweet ID : 667491009379606528  
Tweet ID : 667470559035432960  
Tweet ID : 667455448082227200  
Tweet ID : 667453023279554560  
Tweet ID : 667443425659232256  
Tweet ID : 667437278097252352  
Tweet ID : 667435689202614272  
Tweet ID : 667405339315146752  
Tweet ID : 667393430834667520  
Tweet ID : 667369227918143488  
Tweet ID : 667211855547486208  
Tweet ID : 667200525029539841  
Tweet ID : 667192066997374976  
Tweet ID : 667188689915760640  
Tweet ID : 667182792070062081  
Tweet ID : 667177989038297088  
Tweet ID : 667176164155375616  
Tweet ID : 667174963120574464  
Tweet ID : 667171260800061440  
Tweet ID : 667165590075940865  
Tweet ID : 667160273090932737  
Tweet ID : 667152164079423490  
Tweet ID : 667138269671505920  
Tweet ID : 667119796878725120  
Tweet ID : 667090893657276420  
Tweet ID : 667073648344346624  
Tweet ID : 667070482143944705  
Tweet ID : 667065535570550784  
Tweet ID : 667062181243039745  
Tweet ID : 667044094246576128  
Tweet ID : 667012601033924608  
Tweet ID : 666996132027977728  
Tweet ID : 666983947667116034  
Tweet ID : 666837028449972224  
Tweet ID : 666835007768551424  
Tweet ID : 666826780179869698  
Tweet ID : 666817836334096384  
Tweet ID : 666804364988780544  
Tweet ID : 666786068205871104  
Tweet ID : 666781792255496192  
Tweet ID : 666776908487630848  
Tweet ID : 666739327293083650  
Tweet ID : 666701168228331520



```
Tweet ID : 666691418707132416
Tweet ID : 666649482315059201
Tweet ID : 666644823164719104
Tweet ID : 666454714377183233
Tweet ID : 666447344410484738
Tweet ID : 666437273139982337
Tweet ID : 666435652385423360
Tweet ID : 666430724426358785
Tweet ID : 666428276349472768
Tweet ID : 666421158376562688
Tweet ID : 666418789513326592
Tweet ID : 666411507551481857
Tweet ID : 666407126856765440
Tweet ID : 666396247373291520
Tweet ID : 666373753744588802
Tweet ID : 666362758909284353
Tweet ID : 666353288456101888
Tweet ID : 666345417576210432
Tweet ID : 666337882303524864
Tweet ID : 666293911632134144
Tweet ID : 666287406224695296
Tweet ID : 666273097616637952
Tweet ID : 666268910803644416
Tweet ID : 666104133288665088
Tweet ID : 666102155909144576
Tweet ID : 666099513787052032
Tweet ID : 666094000022159362
Tweet ID : 666082916733198337
Tweet ID : 666073100786774016
Tweet ID : 666071193221509120
Tweet ID : 666063827256086533
Tweet ID : 666058600524156928
Tweet ID : 666057090499244032
Tweet ID : 666055525042405380
Tweet ID : 666051853826850816
Tweet ID : 666050758794694657
Tweet ID : 666049248165822465
Tweet ID : 666044226329800704
Tweet ID : 666033412701032449
Tweet ID : 666029285002620928
Tweet ID : 666020888022790149
Time the process completed : 33 minutes 22 seconds
```

In [42]:

```
print(errors)
```

```
[888202515573088257, 873697596434513921, 872668790621863937, 872261713294495745,
869988702071779329, 866816280283807744, 861769973181624320, 856602993587888130,
851953902622658560, 845459076796616705, 844704788403113984, 842892208864923648,
837366284874571778, 837012587749474308, 829374341691346946, 827228250799742977,
812747805718642688, 802247111496568832, 779123168116150273, 775096608509886464,
771004394259247104, 770743923962707968, 759566828574212096, 754011816964026368,
680055455951884288]
```

1. Using the Python Tweepy library , Gathering retweet count and like of tweet count thanks to the usage of the Twitter API

Reference Link : <https://stackoverflow.com/questions/47925828/how-to-create-a-pandas-dataframe-using-tweepy>

In [100]:

```
extracted_tweet_data = []
with open('tweet_json.txt') as tweet_json_file: # tweet-json.txt
    for line in tweet_json_file:
        json_data = json.loads(line)
        tweet_id = json_data['id']
        retweet_count = json_data['retweet_count']
        favorite_count = json_data['favorite_count']
        extracted_tweet_data.append(
            {
                'tweet_id': tweet_id,
                'retweet_count': int(retweet_count),
                'favorite_count': int(favorite_count),
            }
        )
```

```
twitter_api_extracted_data = pd.DataFrame(extracted_tweet_data, columns = ['tweet_id',
'retweeted_count', 'favorite_count'])
```

## Part II - Assessing Data

In this part, I determine whether all dataset is convenient or not by defining quality issues and tidiness issues.

Here are the features of both quality issues and tidiness issues shown below.

The four features of quality issues:

1. Completeness: Does a dataset have missing data?
2. Validity: Does a dataset make sense?
3. Accuracy: Does a dataset contain inaccurate data?
4. Consistency: Does the data of each row has standardization?

Three features for tidiness:

1. Each variable is defined only just in one column
2. Duplicate rows cannot be contained in dataset.
3. If dataset is combined with another one, the column of both all dataset applying join process should be same type.

First, I assess twitter\_archive dataset to find all kinds of issues defined above.

### twitter\_archive

- *tweet\_id*: the unique identifier of tweets
- *in\_reply\_to\_status\_id*: if the represented Tweet is a reply, this field will contain the integer representation of the original Tweet's ID
- *in\_reply\_to\_user\_id*: if the represented Tweet is a reply, this field will contain the integer representation of the original Tweet's author ID
- *timestamp*: timestamp creation of the Tweet
- *source*: device from where the Tweet was posted, as an HTML-formatted string. e.g. Twitter for Android, Twitter for iPhone, Twitter Web Client
- *text*: actual text of the status update
- *retweeted\_status\_id*: if the represented Tweet is a retweet, this field will contain the integer representation of the original Tweet's ID
- *retweeted\_status\_user\_id*: if the represented Tweet is a retweet, this field will contain the integer representation of the original Tweet's author ID
- *retweeted\_status\_timestamp*: timestamp of retweet
- *expanded\_urls*: tweet URL
- *rating\_numerator*: numerator of the rating of a dog
- *rating\_denominator*: denominator of the rating of a dog
- *name*: name of the dog
- *doggo*: respective dog stage
- *floofer*: respective dog stage
- *pupper*: respective dog stage
- *puppo*: respective dog stage

In [101]:

```
# Show first 5 data of twitter_archive dataframe
twitter_archive.head(5)
```

Out[101]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retwee
0	892420643555336193	NaN	NaN	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone"	This is Phineas. He's a mystical	

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retwe
1	892177421306343426	NaN	NaN	2017-08-01 00:17:27+0000	href="http://twitter.com/download/iphone"	This is Tilly. She's just checking pup on you....	
2	891815181378084864	NaN	NaN	2017-07-31 00:18:03+0000	href="http://twitter.com/download/iphone"	This is Archie. He is a rare Norwegian Pouncin...	
3	891689557279858688	NaN	NaN	2017-07-30 15:58:51+0000	href="http://twitter.com/download/iphone"	This is Darla. She commenced a snooze mid meal...	
4	891327558926688256	NaN	NaN	2017-07-29 16:00:24+0000	href="http://twitter.com/download/iphone"	This is Franklin. He would like you to stop ca...	

In [102]:

```
# Show last 5 data of twitter_archive dataframe
twitter_archive.tail(5)
```

Out[102]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retw
2351	666049248165822465	NaN	NaN	2015-11-16 00:24:50+0000	href="http://twitter.com/download/iphone"	Here we have a 1949 1st generation vulpix. Enj...	
2352	666044226329800704	NaN	NaN	2015-11-16 00:04:52+0000	href="http://twitter.com/download/iphone"	This is a purebred Piers Morgan. Loves to Netf...	
2353	666033412701032449	NaN	NaN	2015-11-15 23:21:54+0000	href="http://twitter.com/download/iphone"	Here is a very happy pup. Big fan of well-main...	
2354	666029285002620928	NaN	NaN	2015-11-15 23:05:30+0000	href="http://twitter.com/download/iphone"	This is a western brown Mitsubishi terrier. Up...	
2355	666020888022790149	NaN	NaN	2015-11-15 22:32:08+0000	href="http://twitter.com/download/iphone"	Here we have a Japanese Irish Setter. Lost eye...	

In [103]:

```
# Show how many variables there are in twitter_archive dataframe
twitter_archive.count()
```

Out[103]:

```
tweet_id      2356
in_reply_to_status_id  78
in_reply_to_user_id    78
timestamp      2356
source         2356
text           2356
```

```

retweeted_status_id      2356
retweeted_status_user_id  181
retweeted_status_timestamp 181
expanded_urls            2297
rating_numerator          2356
rating_denominator        2356
name                      2356
doggo                     2356
floofer                   2356
pupper                    2356
puppo                     2356
dtype: int64

```

In [104]:

```

# Show detailed information of each column about the twitter_archive dataframe
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 [105]:

```

# Show descriptive statistics of twitter_archive dataframe
twitter_archive.describe()

```

Out[105]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	retweeted_status_id	retweeted_status_user_id	rating_numerator	rating
count	2.356000e+03	7.800000e+01	7.800000e+01	1.810000e+02	1.810000e+02	2356.000000	
mean	7.427716e+17	7.455079e+17	2.014171e+16	7.720400e+17	1.241698e+16	13.126486	
std	6.856705e+16	7.582492e+16	1.252797e+17	6.236928e+16	9.599254e+16	45.876648	
min	6.660209e+17	6.658147e+17	1.185634e+07	6.661041e+17	7.832140e+05	0.000000	
25%	6.783989e+17	6.757419e+17	3.086374e+08	7.186315e+17	4.196984e+09	10.000000	
50%	7.196279e+17	7.038708e+17	4.196984e+09	7.804657e+17	4.196984e+09	11.000000	
75%	7.993373e+17	8.257804e+17	4.196984e+09	8.203146e+17	4.196984e+09	12.000000	
max	8.924206e+17	8.862664e+17	8.405479e+17	8.874740e+17	7.874618e+17	1776.000000	

In [106]:

```

# Deterime the number of duplicate values with respect too twitter_id column in twitter_archive da
taframe
sum(twitter_archive['tweet_id'].duplicated())

```

Out[106]:

0

In [107]:

```
# Show the number of numerators with counts of each data value in twitter_archive dataframe
twitter_archive["rating_numerator"].value_counts()
```

Out[107]:

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

Name: rating\_numerator, dtype: int64

In [108]:

```
# Show the number of names with counts of each data value in twitter_archive dataframe
twitter_archive['name'].value_counts()
```

Out[108]:

None	745
a	55
Charlie	12
Cooper	11
Oliver	11
Lucy	11
Penny	10
Tucker	10
Lola	10
Bo	9
Winston	9
the	8
Sadie	8
an	7
Daisy	7
Bailey	7
Toby	7
Ruddy	7

```

Buddy      /
Bella      6
Jax        6
Rusty      6
Jack       6
Scout      6
Koda       6
Stanley    6
Milo       6
Leo        6
Oscar      6
Dave       6
Chester    5
...
Bubba      1
Maude      1
Bauer      1
Alexanderson 1
Marq       1
Zuzu       1
Toffee     1
Vince      1
Ralphie    1
Rascal     1
Tedrick    1
Ivar       1
Ralphie    1
Liam       1
Jiminus    1
Marlee     1
Jackie     1
Laela      1
Robin      1
Sprout     1
Pinot      1
Cilantro   1
Wishes     1
Zeek       1
Gert       1
Aiden      1
Divine     1
O          1
Meera      1
William    1
Name: name, Length: 957, dtype: int64

```

In [109]:

```

# Show the number of rating denominators with counts of each data value in twitter_archive dataframe
twitter_archive["rating_denominator"].value_counts()

```

Out[109]:

```

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

```

In [110]:

```
# Show source value for the first value of twitter_archive dataframe
twitter_archive["source"][0]
```

Out[110]:

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

In [111]:

```
# Show None value of name for twitter_archive dataframe
twitter_archive[twitter_archive["name"] == 'None']
```

Out[111]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	
5	891087950875897856	NaN	NaN	2017-07-29 00:08:17 +0000	<a href="http://twitter.com/download/iphone" r...	Here we have a great white bre
7	890729181411237888	NaN	NaN	2017-07-28 00:22:40 +0000	<a href="http://twitter.com/download/iphone" r...	When you w owner call anot
12	889665388333682689	NaN	NaN	2017-07-25 01:55:32 +0000	<a href="http://twitter.com/download/iphone" r...	Here's a pt seems to be on t
24	887343217045368832	NaN	NaN	2017-07-18 16:08:03 +0000	<a href="http://twitter.com/download/iphone" r...	You may not hav you needed to s
25	887101392804085760	NaN	NaN	2017-07-18 00:07:08 +0000	<a href="http://twitter.com/download/iphone" r...	This... is a Antarctic Hou
30	886267009285017600	8.862664e+17	2.281182e+09	2017-07-15 16:51:35 +0000	<a href="http://twitter.com/download/iphone" r...	@Nonl @MayhewMayt hello tar
32	886054160059072513	NaN	NaN	2017-07-15 02:45:48 +0000	<a href="http://twitter.com/download/iphone" r...	RT @Athleti https://t.co/W
35	885518971528720385	NaN	NaN	2017-07-13 15:19:09 +0000	<a href="http://twitter.com/download/iphone" r...	I have a new hen name is Howar
37	885167619883638784	NaN	NaN	2017-07-12 16:03:00 +0000	<a href="http://twitter.com/download/iphone" r...	Here we hav underc mal
41	884441805382717440	NaN	NaN	2017-07-10 15:58:53 +0000	<a href="http://twitter.com/download/iphone" r...	I present to yo Hat. Pup in H
42	884247878851493888	NaN	NaN	2017-07-10 03:08:17 +0000	<a href="http://twitter.com/download/iphone" r...	OMG HE DIDN' TO HE W/ TRYINC
47	883117836046086144	NaN	NaN	2017-07-07 00:17:54 +0000	<a href="http://twitter.com/download/iphone" r...	Please only se We c mec
55	881633300179243008	8.816070e+17	4.738443e+07	2017-07-02 21:58:53 +0000	<a href="http://twitter.com/download/iphone" r...	@roushfenway T good dogs but 1
59	880872448815771648	NaN	NaN	2017-06-30 19:35:32 +0000	<a href="http://twitter.com/download/iphone" r...	Ugh not again. rate dogs. Pleas
62	880095782870896641	NaN	NaN	2017-06-28 16:00:00	<a href="http://twitter.com/download/iphone" r...	Please don' photos withou

tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text
64	879674319642796034	8.795538e+17	3.105441e+09	2017-06-27 12:14:36 +0000	href="http://twitter.com/download/iphone" r... @RealKentMurp c
72	878604707211726852	NaN	NaN	2017-06-24 13:24:20 +0000	href="http://twitter.com/download/iphone" r... Martha is stun h*ckin dare you.
78	877611172832227328	NaN	NaN	2017-06-21 19:36:23 +0000	href="http://twitter.com/download/iphone" r... RT @rac @dog_rates the l
83	876537666061221889	NaN	NaN	2017-06-18 20:30:39 +0000	href="http://twitter.com/download/iphone" r... I can say pupmost confide
88	875097192612077568	NaN	NaN	2017-06-14 21:06:43 +0000	href="http://twitter.com/download/iphone" r... You'll get your when that precio
89	875021211251597312	NaN	NaN	2017-06-14 16:04:48 +0000	href="http://twitter.com/download/iphone" r... Guys please stop pictures witho
93	874057562936811520	NaN	NaN	2017-06-12 00:15:36 +0000	href="http://twitter.com/download/iphone" r... I can't believe th happening.
96	873580283840344065	NaN	NaN	2017-06-10 16:39:04 +0000	href="http://twitter.com/download/iphone" r... We usually c Deck-bound Si
99	872967104147763200	NaN	NaN	2017-06-09 00:02:31 +0000	href="http://twitter.com/download/iphone" r... Here's a very le He has a date
100	872820683541237760	NaN	NaN	2017-06-08 14:20:41 +0000	href="http://twitter.com/download/iphone" r... Here are my #dogsatpollin
101	872668790621863937	NaN	NaN	2017-06-08 04:17:07 +0000	href="http://twitter.com/download/iphone" r... RT @logan: Penelope here
103	872486979161796608	NaN	NaN	2017-06-07 16:14:40 +0000	href="http://twitter.com/download/iphone" r... We. Only. Rate. I not send in o
110	871102520638267392	NaN	NaN	2017-06-03 20:33:19 +0000	href="http://twitter.com/download/iphone" r... Never doubt https://t.co/AbBL
112	870804317367881728	NaN	NaN	2017-06-03 00:48:22 +0000	href="http://twitter.com/download/iphone" r... Real funny guys. in a pic witho
113	870726314365509632	8.707262e+17	1.648776e+07	2017-06-02 19:38:25 +0000	href="http://twitter.com/download/iphone" r... @Corr @ShopWeR &gt;10/10
...	...	...	...	...	...
2294	667138269671505920	NaN	NaN	2015-11-19 00:32:12 +0000	href="http://twitter.com/download/iphone" r... Extremely intelli here. Has lea
2298	667070482143944705	6.670655e+17	4.196984e+09	2015-11-18 20:02:51 +0000	href="http://twitter.com/download/iphone" r... After much de dog is being upg
2299	667065535570550784	NaN	NaN	2015-11-18 19:43:11 +0000	href="http://twitter.com/download/iphone" r... Here w Hufflepuff. Lo E
2299	667065535570550784	NaN	NaN	2015-11-18	href="http://twitter.com/download/iphone" r... 12/10 qin



id	text	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	https://t.co/QZAn
2301	6670440942465/6128	NaN	NaN	18-11-17 18:17:59+0000	href="http://twitter.com/download/iphone"	
2305	666837028449972224	NaN	NaN	2015-11-18 04:35:11+0000	href="http://twitter.com/download/iphone" r...	My goodness. ' dog here. Large
2306	666835007768551424	NaN	NaN	2015-11-18 04:27:09+0000	href="http://twitter.com/download/iphone" r...	These are Feldspars. The
2307	666826780179869698	NaN	NaN	2015-11-18 03:54:28+0000	href="http://twitter.com/download/iphone" r...	12/10 simply bril https://t.co/V
2310	666786068205871104	NaN	NaN	2015-11-18 01:12:41+0000	href="http://twitter.com/download/iphone" r...	Unfamiliar with th Ears pointy
2316	666649482315059201	NaN	NaN	2015-11-17 16:09:56+0000	href="http://twitter.com/download/iphone" r...	Cool dog. Enjoy Low monotone ba
2320	666437273139982337	NaN	NaN	2015-11-17 02:06:42+0000	href="http://twitter.com/download/iphone" r...	Here we se nort Cumberb
2321	666435652385423360	NaN	NaN	2015-11-17 02:00:15+0000	href="http://twitter.com/download/iphone" r...	"Can you behave ruining my weddi
2322	666430724426358785	NaN	NaN	2015-11-17 01:40:41+0000	href="http://twitter.com/download/iphone" r...	Oh boy wh Sunglasses take
2323	666428276349472768	NaN	NaN	2015-11-17 01:30:57+0000	href="http://twitter.com/download/iphone" r...	Here we have an Pulitzer. Coll
2324	666421158376562688	NaN	NaN	2015-11-17 01:02:40+0000	href="http://twitter.com/download/iphone" r...	*internally sci 12/10 https://t.co
2328	666396247373291520	NaN	NaN	2015-11-16 23:23:41+0000	href="http://twitter.com/download/iphone" r...	Oh goodness rare northea
2329	666373753744588802	NaN	NaN	2015-11-16 21:54:18+0000	href="http://twitter.com/download/iphone" r...	Those are sungla: a jean jacket.
2330	666362758909284353	NaN	NaN	2015-11-16 21:10:36+0000	href="http://twitter.com/download/iphone" r...	Unique dog h small. Lives in c
2331	666353288456101888	NaN	NaN	2015-11-16 20:32:58+0000	href="http://twitter.com/download/iphone" r...	Here we have Asiago Gals
2332	666345417576210432	NaN	NaN	2015-11-16 20:01:42+0000	href="http://twitter.com/download/iphone" r...	Look at this thinking seat be
2336	666273097616637952	NaN	NaN	2015-11-16 15:14:19+0000	href="http://twitter.com/download/iphone" r...	Can take selfi https://t.co/ws2AN
2337	666268910803644416	NaN	NaN	2015-11-16 14:57:41+0000	href="http://twitter.com/download/iphone" r...	Very concern fellow dog tr
2338	666104133288665088	NaN	NaN	2015-11-16 04:02:55+0000	href="http://twitter.com/download/iphone" r...	Not familiar breed. No tail (
2339	666102155909144576	NaN	NaN	2015-11-16 03:55:04	href="http://twitter.com/download/iphone" r...	Oh my. Here seeing an Adol

tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text
2340	666099513787052032	NaN	2015-11-16 03:44:34 +0000	<a href="http://twitter.com/download/iphone" r...	Can stand on s what seems like
2341	666094000022159362	NaN	2015-11-16 03:22:39 +0000	<a href="http://twitter.com/download/iphone" r...	This appear Mongolian Pre
2342	666082916733198337	NaN	2015-11-16 02:38:37 +0000	<a href="http://twitter.com/download/iphone" r...	Here we hav esi sunblock
2343	666073100786774016	NaN	2015-11-16 01:59:36 +0000	<a href="http://twitter.com/download/iphone" r...	Let's hope this f Malaysia
2344	666071193221509120	NaN	2015-11-16 01:52:02 +0000	<a href="http://twitter.com/download/iphone" r...	Here we have a Rhodode
2351	666049248165822465	NaN	2015-11-16 00:24:50 +0000	<a href="http://twitter.com/download/iphone" r...	Here we have a generation vulp
2355	666020888022790149	NaN	2015-11-15 22:32:08 +0000	<a href="http://twitter.com/download/iphone" r...	Here we have a J Irish Setter. L

745 rows × 17 columns

In [112]:

```
# Show Null values of each column in twitter archive dataframe
twitter_archive.isnull().sum()
```

Out[112]:

```
tweet_id          0
in_reply_to_status_id    2278
in_reply_to_user_id    2278
timestamp          0
source             0
text              0
retweeted_status_id    2175
retweeted_status_user_id  2175
retweeted_status_timestamp  2175
expanded_urls        59
rating_numerator      0
rating_denominator    0
name                 0
doggo                0
floofer              0
pupper              0
puppo                0
dtype: int64
```

In [113]:

```
# Check whether duplicated values exist or not in twitter archive dataframe
twitter_archive.duplicated().sum()
```

Out[113]:

0

In [114]:

```
twitter_archive.groupby(["doggo", "floofer", "pupper",
"puppo"]).size().reset_index().rename(columns={0: "count"})
```

Out[114]:

	doggo	floofer	pupper	puppo	count
0	None	None	None	None	1976
1	None	None	None	puppo	29
2	None	None	pupper	None	245
3	None	floofer	None	None	9
4	doggo	None	None	None	83
5	doggo	None	None	puppo	1
6	doggo	None	pupper	None	12
7	doggo	floofer	None	None	1

After assessing first dataframe named for twitter\_archive, Here are the results derived from its assessment.

**twitter\_archive:**

### Quality Issues

- **Completeness:**
  - missing data in the following columns: in\_reply\_to\_status\_id, in\_reply\_to\_user\_id, retweeted\_status\_id, retweeted\_status\_user\_id, retweeted\_status\_timestamp, expanded\_urls
  - All tweet\_id is int type so that other tables' tweet id should be int type
- **Validity:**
  - Dog names: some dogs contain 'None', 'a', and 'an.'
  - Because this dataset includes retweets part and it leads to duplicated data, retweeted\_status\_id, retweeted\_status\_user\_id and retweeted\_status\_timestamp columns are empty.
- **Accuracy:**
  - The last value of rating\_numerator is 1776
  - timestamp is defined as an object instead of datetime
  - retweeted\_status\_timestamp is defined as an object instead of datetime
- **Consistency:**
  - Source column contains the HTML tags
  - Value of rating\_denominator column should be 10 as a standart, but there are an other values such as 11, 50 80

### Tidiness Issues

- The last four columns all relate to the same variable (dogoo, floofer, pupper, puppo) instead of defined only just one column named for dog\_stage as categorial approach.

Secondly, I assess image\_predictions dataset to find all kinds of issues defined above.

### image\_predictions

- *tweet\_id*: the unique identifier of tweet
- *jpg\_url*: dog's image URL
- *img\_num*: the image number that corresponded to the most confident prediction (numbered 1 to 4 since tweets can have up to four images)
- *p1*: algorithm's #1 prediction for the image in the tweet
- *p1\_conf*: how confident the algorithm is in its #1 prediction
- *p1\_dog*: whether or not the #1 prediction is a breed of dog
- *p2*: algorithm's #2 prediction for the image in the tweet
- *p2\_conf*: how confident the algorithm is in its #2 prediction
- *p2\_dog*: whether or not the #2 prediction is a breed of dog
- *p3*: algorithm's #3 prediction for the image in the tweet
- *p3\_conf*: how confident the algorithm is in its #3 prediction

- `p3_conf`: how confident the algorithm is in its #3 prediction
- `p3_dog`: whether or not the #3 prediction is a breed of dog!

In [115]:

```
# Get first 5 rows from image predictions dataframe
image_predictions.head()
```

Out[115]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	
0	666020888022790149	https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg	1	Welsh_springer_spaniel	0.465074	True	
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	1	redbone	0.506826	True	miniature
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg	1	German_shepherd	0.596461	True	
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-IEu.jpg	1	Rhodesian_ridgeback	0.408143	True	
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg	1	miniature_pinscher	0.560311	True	F

In [116]:

```
# Get last 5 rows from image predictions dataframe
image_predictions.tail()
```

Out[116]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p
2070	891327558926688256	https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg	2	basset	0.555712	True	English_springe
2071	891689557279858688	https://pbs.twimg.com/media/DF_q7IAWsaEuuN8.jpg	1	paper_towel	0.170278	False	Labrador_retrieve
2072	891815181378084864	https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg	1	Chihuahua	0.716012	True	malamute
2073	892177421306343426	https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg	1	Chihuahua	0.323581	True	Pekines
2074	892420643555336193	https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg	1	orange	0.097049	False	bag

In [117]:

```
# Show detailed information of each column about the image prediction dataframe
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 [118]:

```
# Show descriptive statistics of image predictions dataframe
image_predictions.describe()
```

Out[118]:

tweet_id	img_num	p1_conf	p2_conf	p3_conf
----------	---------	---------	---------	---------

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

In [120]:

```
# This is an image for tweet_id 666020888022790149
id_image = image_predictions[image_predictions["tweet_id"] == 666020888022790149]
["jpg_url"].values[0]
# show images Via Image Library
Image(url = str(id_image))
```

Out[120]:



In [121]:

```
# Check whether duplicated tweet_id values exist or not in image predictions dataframe
image_predictions[image_predictions["tweet_id"].duplicated()]
```

Out[121]:

tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog
----------	---------	---------	----	---------	--------	----	---------	--------	----	---------	--------

In [122]:

```
# Show the sum of duplicated jpg_url values of image_predictions dataframe
sum(image_predictions['jpg_url'].duplicated())
```

Out[122]:

66

In [123]:

```
# Get the exact duplicate records. The expected number is 132 because of 66*2
image_predictions_df_images_duplicates = image_predictions[image_predictions["jpg_url"].duplicated(
    keep=False)]
image_predictions_df_images_duplicates.sort_values('jpg_url')
```

Out[123]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	
480	675354435921575936	https://pbs.twimg.com/ext_tw_video_thumb/67535...	1	upright	0.303415	False	
1297	752309394570878976	https://pbs.twimg.com/ext_tw_video_thumb/67535...	1	upright	0.303415	False	
1864	842892208864923648	https://pbs.twimg.com/ext_tw_video_thumb/80710...	1	Chihuahua	0.505370	True	
1641	807106840509214720	https://pbs.twimg.com/ext_tw_video_thumb/80710...	1	Chihuahua	0.505370	True	
1703	817181837579653120	https://pbs.twimg.com/ext_tw_video_thumb/81596...	1	Tibetan_mastiff	0.506312	True	
1691	815966073409433600	https://pbs.twimg.com/ext_tw_video_thumb/81596...	1	Tibetan_mastiff	0.506312	True	
1705	817423860136083457	https://pbs.twimg.com/ext_tw_video_thumb/81742...	1	ice_bear	0.336200	False	
1858	841833993020538882	https://pbs.twimg.com/ext_tw_video_thumb/81742...	1	ice_bear	0.336200	False	
1715	819004803107983360	https://pbs.twimg.com/media/C12whDoVEAALRxa.jpg	1	standard_poodle	0.351308	True	
1718	819015337530290176	https://pbs.twimg.com/media/C12whDoVEAALRxa.jpg	1	standard_poodle	0.351308	True	
1716	819006400881917954	https://pbs.twimg.com/media/C12x-JTVIAAzdf.jpg	4	prison	0.907083	False	
1717	819015331746349057	https://pbs.twimg.com/media/C12x-JTVIAAzdf.jpg	4	prison	0.907083	False	
1746	823269594223824897	https://pbs.twimg.com/media/C2kzTGxWEAEOpPL.jpg	1	Samoyed	0.585441	True	
1738	822244816520155136	https://pbs.twimg.com/media/C2kzTGxWEAEOpPL.jpg	1	Samoyed	0.585441	True	
1740	822489057087389700	https://pbs.twimg.com/media/C2oRbOuWEAAbVSI.jpg	1	Samoyed	0.416769	True	
1742	822647212903690241	https://pbs.twimg.com/media/C2oRbOuWEAAbVSI.jpg	1	Samoyed	0.416769	True	
1767	826958653328592898	https://pbs.twimg.com/media/C3nygbBWQAAjwcW.jpg	1	golden_retriever	0.617389	True	
1789	829878982036299777	https://pbs.twimg.com/media/C3nygbBWQAAjwcW.jpg	1	golden_retriever	0.617389	True	
1785	829374341691346946	https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg	1	Staffordshire_bullterrier	0.757547	True	Ameri
1903	851953902622658560	https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg	1	Staffordshire_bullterrier	0.757547	True	Ameri
2041	885311592912609280	https://pbs.twimg.com/media/C4bTH6nWMAAX_bJ.jpg	1	Labrador_retriever	0.908703	True	
1791	830583320585068544	https://pbs.twimg.com/media/C4bTH6nWMAAX_bJ.jpg	1	Labrador_retriever	0.908703	True	
224	670319130621435904	https://pbs.twimg.com/media/CU1zsMSUAAAS0qW.jpg	1	Irish_terrier	0.254856	True	
1345	759159934323924993	https://pbs.twimg.com/media/CU1zsMSUAAAS0qW.jpg	1	Irish_terrier	0.254856	True	
1590	798644042770751489	https://pbs.twimg.com/media/CU3mlTUWIAAfyQS.jpg	1	English_springer	0.403698	True	
241	670444955656130560	https://pbs.twimg.com/media/CU3mlTUWIAAfyQS.jpg	1	English_springer	0.403698	True	
1589	798628517273620480	https://pbs.twimg.com/media/CUN4Or5UAAAa5K4.jpg	1	beagle	0.636169	True	
85	667509364010450944	https://pbs.twimg.com/media/CUN4Or5UAAAa5K4.jpg	1	beagle	0.636169	True	
327	671896809300709376	https://pbs.twimg.com/media/CVMOIMiWwAA4Yxl.jpg	1	chow	0.243529	True	
1591	798665375516884993	https://pbs.twimg.com/media/CVMOIMiWwAA4Yxl.jpg	1	chow	0.243529	True	
...	...	...	...	...	...	...	
1571	794983741416415232	https://pbs.twimg.com/media/CvT6IV6WEAQhhV5.jpg	3	schipperke	0.363272	True	
1530	789530877013393408	https://pbs.twimg.com/media/CvT6IV6WEAQhhV5.jpg	3	schipperke	0.363272	True	
1538	790723298204217344	https://pbs.twimg.com/media/CvaYgDOWgAEfjls.jpg	1	tub	0.479477	False	
1533	789986466051088384	https://pbs.twimg.com/media/CvaYgDOWgAEfjls.jpg	1	tub	0.479477	False	
1683	813944609378369540	https://pbs.twimg.com/media/Cveg1-NXgAASaaT.jpg	1	Labrador_retriever	0.427742	True	
1584	788677447344887748	https://pbs.twimg.com/media/CvT6IV6WEAQhhV5.jpg	1	English_springer	0.403698	True	



1534	790271117346975746	https://pbs.twimg.com/media/Cveg1-NXgAASaaI.jpg	1	Labrador_retriever	0.427742	True
1539	790946055508652032	https://pbs.twimg.com/media/CvoBPWRWgAA4het.jpg	1	dishwasher	0.700466	False
1699	816829038950027264	https://pbs.twimg.com/media/CvoBPWRWgAA4het.jpg	1	dishwasher	0.700466	False
1544	791672322847637504	https://pbs.twimg.com/media/CvyVxQRWEAAAdSZS.jpg	1	golden_retriever	0.705092	True
1564	793614319594401792	https://pbs.twimg.com/media/CvyVxQRWEAAAdSZS.jpg	1	golden_retriever	0.705092	True
1804	832215726631055365	https://pbs.twimg.com/media/CwJR1okWIAA6XMp.jpg	1	Afghan_hound	0.274637	True
1561	793286476301799424	https://pbs.twimg.com/media/CwJR1okWIAA6XMp.jpg	1	Afghan_hound	0.274637	True
1566	793962221541933056	https://pbs.twimg.com/media/CwS4aqZXUAAe3IO.jpg	1	Labrador_retriever	0.861651	True
1652	809808892968534016	https://pbs.twimg.com/media/CwS4aqZXUAAe3IO.jpg	1	Labrador_retriever	0.861651	True
1572	795076730285391872	https://pbs.twimg.com/media/CwiuEJmW8AAZnit.jpg	2	gas_pump	0.676439	False
1755	824796380199809024	https://pbs.twimg.com/media/CwiuEJmW8AAZnit.jpg	2	gas_pump	0.676439	False
1579	796177847564038144	https://pbs.twimg.com/media/Cwx99pW8AMk_le.jpg	1	golden_retriever	0.600276	True
1578	796149749086875649	https://pbs.twimg.com/media/Cwx99pW8AMk_le.jpg	1	golden_retriever	0.600276	True
1610	801167903437357056	https://pbs.twimg.com/media/Cx5R8wPVEAALa9r.jpg	1	cocker_spaniel	0.740220	True
1645	808134635716833280	https://pbs.twimg.com/media/Cx5R8wPVEAALa9r.jpg	1	cocker_spaniel	0.740220	True
1603	800141422401830912	https://pbs.twimg.com/media/CxqsX-8XUAAEvjD.jpg	3	golden_retriever	0.938048	True
1727	820446719150292993	https://pbs.twimg.com/media/CxqsX-8XUAAEvjD.jpg	3	golden_retriever	0.938048	True
1944	861769973181624320	https://pbs.twimg.com/media/CzG425nWgAAAnP7P.jpg	2	Arabian_camel	0.366248	False
1638	806629075125202948	https://pbs.twimg.com/media/CzG425nWgAAAnP7P.jpg	2	Arabian_camel	0.366248	False
1970	868880397819494401	https://pbs.twimg.com/media/DA7iHL5U0AA1OQo.jpg	1	laptop	0.153718	False
1992	873697596434513921	https://pbs.twimg.com/media/DA7iHL5U0AA1OQo.jpg	1	laptop	0.153718	False
2051	887473957103951883	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg	2	Pembroke	0.809197	True
2055	888202515573088257	https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg	2	Pembroke	0.809197	True
1033	711694788429553666	https://pbs.twimg.com/tweet_video_thumb/CeBym7...	1	brown_bear	0.713293	False
1364	761371037149827077	https://pbs.twimg.com/tweet_video_thumb/CeBym7...	1	brown_bear	0.713293	False

132 rows × 7 columns



In [127]:

```
# Check whether duplicated values exist or not in image prediction dataframe
image_predictions.duplicated().sum()
```

Out[127]:

0

In [128]:

```
# Show Null values of each column in image prediction dataframe
image_predictions.isnull().sum()
```

Out[128]:

```
tweet_id      0
jpg_url       0
img_num       0
p1            0
p1_conf       0
p1_dog        0
p2           0
p2_conf       0
p2_dog        0
p3           0
p3_conf       0
p3_dog        0
dtype: int64
```

In [129]:

```
# Show what different kinds of values there are in p1,p2,p3 columns
```

```
print("p1 Column")
print(image_predictions['p1'].value_counts())
print("-----")
```

```
p1 Column
golden_retriever      150
Labrador_retriever    100
Pembroke               89
Chihuahua             83
pug                   57
chow                  44
Samoyed               43
toy_poodle            39
Pomeranian            38
cocker_spaniel        30
malamute              30
French_bulldog        26
Chesapeake_Bay_retriever 23
miniature_pinscher    23
seat_belt             22
German_shepherd       20
Staffordshire_bullterrier 20
Siberian_husky        20
web_site              19
Cardigan              19
Maltese_dog           18
beagle                18
Eskimo_dog            18
Shetland_sheepdog     18
teddy                 18
Lakeland_terrier      17
Rottweiler            17
Shih-Tzu              17
kuvasz                16
Italian_greyhound     16
...
slug                  1
pitcher               1
cliff                 1
candle                1
bee_eater             1
snowmobile           1
hummingbird           1
swab                  1
platypus              1
water_bottle          1
marmot                1
black-footed_ferret   1
military_uniform      1
hotdog                1
prayer_rug            1
damselfly             1
leopard               1
sliding_door          1
hand_blower           1
silky_terrier         1
rain_barrel           1
ocarina               1
grey_fox              1
guenon                1
shopping_basket       1
mortarboard           1
starfish              1
bison                 1
maze                  1
teapot                1
Name: p1, Length: 378, dtype: int64
-----
```

In [130]:

```
print("p2 Column")
print(image_predictions['p2'].value_counts())
print("-----")
```



```

p2 Column
Labrador_retriever      104
golden_retriever         92
Cardigan                 73
Chihuahua               44
Pomeranian              42
French_bulldog          41
Chesapeake_Bay_retriever 41
toy_poodle              37
cocker_spaniel          34
Siberian_husky          33
miniature_poodle        33
beagle                  28
Eskimo_dog              27
collie                  27
Pembroke                27
kuvasz                  26
Italian_greyhound       22
Pekinese                21
American_Staffordshire_terrier 21
malinois                20
chow                    20
toy_terrier             20
Samoyed                 20
miniature_pinscher      20
Boston_bull             19
Norwegian_elkhound      19
Staffordshire_bullterrier 18
Irish_terrier           17
pug                     17
Shih-Tzu                16

```

...

```

tree_frog                1
goose                    1
timber_wolf              1
spotted_salamander       1
saltshaker               1
wombat                   1
bearskin                 1
rifle                    1
assault_rifle            1
cornet                   1
sombbrero                1
quail                    1
cliff                    1
triceratops              1
iPod                     1
killer_whale             1
tarantula                1
siamang                  1
komondor                 1
junco                    1
oxygen_mask              1
necklace                 1
space_heater             1
bannister                1
cannon                   1
rule                     1
handkerchief             1
hotdog                   1
barracouta               1
indri                    1

```

Name: p2, Length: 405, dtype: int64

-----

In [63]:

```

print("p3 Column")
print(image_predictions['p3'].value_counts())
print("-----")

```

```

p3 Column
Labrador_retriever      79
Chihuahua               58
golden retriever        48

```

```

Eskimo_dog      38
kelpie          35
kuvasz          34
chow            32
Staffordshire_bullterrier  32
cocker_spaniel  31
beagle          31
Pekinese        29
Pomeranian      29
toy_poodle       29
Chesapeake_Bay_retriever  27
Pembroke        27
Great_Pyrenees  27
French_bulldog  26
malamute        26
American_Staffordshire_terrier  24
Cardigan        23
pug             23
basenji         21
bull_mastiff     20
toy_terrier      20
Siberian_husky  19
Boston_bull      17
Shetland_sheepdog  17
doormat         16
boxer           16
Lakeland_terrier  16
..
goose           1
jersey          1
dugong          1
desktop_computer  1
Kerry_blue_terrier  1
chimpanzee      1
go-kart         1
entertainment_center  1
coral_reef      1
cardoon         1
neck_brace      1
grocery_store   1
crossword_puzzle  1
rhinoceros_beetle  1
zebra           1
snorkel         1
boathouse       1
gar             1
bib             1
pool_table      1
pretzel         1
guillotine      1
partridge       1
goldfish        1
black_swan      1
assault_rifle   1
wok             1
hare            1
valley          1
loupe           1
Name: p3, Length: 408, dtype: int64
-----

```

After assessing first dataframe named for image\_predictions, Here are the results derived from its assessment.

#### image\_predictions:

##### Quality Issues

- **Completeness:**
  - No missing data in all columns
  - All tweet\_id is int type so that other tables' tweet id should be int type
- **Validity:**
  - p1, p2 and p3 columns have invalid data
  - Multiple duplicated jpg\_url (66 records)

- **Accuracy:**
  - No Accuracy Issue
- **Consistency:**
  - Values in columns p1, p2, p3 contain "-" and "\_" for values with more than one word
  - Values in columns p1, p2, p3 are in mixed cases(lower/upper)

### ***Tidiness Issues***

- Delete all duplicated jpg\_url tweets.
- Delete columns that cannot be used for analysis

### **twitter\_api\_extracted\_data**

- *tweet\_id*: the unique identifier of tweets
- *retweet\_count*: number of times this Tweet has been retweeted
- *favorite\_count*: indicates approximately how many times this Tweet has been liked by Twitter users!

In [131]:

```
# Get first 5 rows from twitter_api_extracted_data dataframe
twitter_api_extracted_data.head()
```

Out[131]:

	tweet_id	retweet_count	favorite_count
0	892420643555336193	7803	36633
1	892177421306343426	5780	31561
2	891815181378084864	3825	23774
3	891689557279858688	7970	39950
4	891327558926688256	8604	38153

In [132]:

```
# Get last 5 rows from twitter_api_extracted_data dataframe
twitter_api_extracted_data.tail()
```

Out[132]:

	tweet_id	retweet_count	favorite_count
2326	666049248165822465	41	99
2327	666044226329800704	133	273
2328	666033412701032449	41	115
2329	666029285002620928	43	121
2330	666020888022790149	470	2447

In [133]:

```
# Show detailed information of each column about the twitter_api_extracted_data dataframe
twitter_api_extracted_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2331 entries, 0 to 2330
Data columns (total 3 columns):
tweet_id          2331 non-null int64
retweet_count     2331 non-null int64
favorite_count    2331 non-null int64
dtypes: int64(3)
memory usage: 54.7 KB
```

In [134]:

```
# Show descriptive statistics of twitter_api_extracted_data dataframe
twitter_api_extracted_data.describe()
```

Out[134]:

	tweet_id	retweet_count	favorite_count
count	2.331000e+03	2331.000000	2331.000000
mean	7.419079e+17	2741.518662	7653.821536
std	6.823170e+16	4636.063127	11876.340787
min	6.660209e+17	1.000000	0.000000
25%	6.782670e+17	554.500000	1338.000000
50%	7.182469e+17	1287.000000	3333.000000
75%	7.986692e+17	3191.000000	9387.000000
max	8.924206e+17	78879.000000	158283.000000

In [135]:

```
# Show how many variables there are in twitter_api_extracted_data dataframe
twitter_api_extracted_data.count()
```

Out[135]:

```
tweet_id          2331
retweet_count      2331
favorite_count      2331
dtype: int64
```

In [136]:

```
# Check whether duplicated values exist or not in twitter_api_extracted_data dataframe
sum(twitter_api_extracted_data['tweet_id'].duplicated())
```

Out[136]:

```
0
```

In [137]:

```
# Show Null values of each column in twitter_api_extracted_data
twitter_api_extracted_data.isnull().sum()
```

Out[137]:

```
tweet_id          0
retweet_count      0
favorite_count      0
dtype: int64
```

In [138]:

```
# Show how many different kinds of variable there are in favorite_count of
twitter_api_extracted_data dataframe
twitter_api_extracted_data['favorite_count'].value_counts()
```

Out[138]:

```
0          163
1381         3
1541         3
1079         3
318          3
1617         3
931          3
3686         3
431          3
```

```

3722      3
2090      3
3446      3
1170      3
235       3
2666      2
4713      2
6774      2
21636     2
2253      2
2238      2
322       2
1081      2
734       2
201       2
3621      2
1085      2
301       2
3881      2
4446      2
1107      2
...
2602      1
2600      1
1358      1
549       1
4593      1
8863      1
12791     1
2552      1
4601      1
507       1
2558      1
8705      1
23044     1
517       1
10758     1
2570      1
523       1
33293     1
10766     1
18960     1
8721      1
8314      1
32583     1
539       1
12829     1
2360      1
2592      1
545       1
23076     1
11698     1
Name: favorite_count, Length: 1998, dtype: int64

```

In [15]:

```

# Show how many different kinds of variable there are in retweet_count of
twitter_api_extracted_data dataframe
twitter_api_extracted_data['retweet_count'].value_counts()

```

Out[15]:

```

231      5
227      5
230      5
561      5
609      5
54       4
630      4
3478     4
114      4
199      4
127      4
74       4
1107     4
210      4

```

```

040      4
1993     4
621      4
578      4
499      4
504      4
4923     4
1040     4
679      4
667      4
750      3
203      3
1381     3
1297     3
1150     3
1672     3
1088     3
..
6120     1
1120     1
2030     1
6132     1
862      1
14098    1
1       1
3       1
182     1
7       1
2004     1
2002     1
1940     1
1650     1
3991     1
1950     1
1207     1
1203     1
4003     1
4011     1
4015     1
4017     1
1153     1
6068     1
1976     1
3248     1
1988     1
1996     1
240      1
2047     1
Name: retweet_count, Length: 1681, dtype: int64

```

In [139]:

```

# Show the name of each column in twitter_archive dataframe
list(twitter_archive)

```

Out[139]:

```

['tweet_id',
 'in_reply_to_status_id',
 'in_reply_to_user_id',
 'timestamp',
 'source',
 'text',
 'retweeted_status_id',
 'retweeted_status_user_id',
 'retweeted_status_timestamp',
 'expanded_urls',
 'rating_numerator',
 'rating_denominator',
 'name',
 'doggo',
 'floofer',
 'pupper',
 'puppo']

```

In [140]:

```
# Show the name of each column in image_predictions dataframe
list(image_predictions)
```

Out[140]:

```
['tweet_id',
 'jpg_url',
 'img_num',
 'p1',
 'p1_conf',
 'p1_dog',
 'p2',
 'p2_conf',
 'p2_dog',
 'p3',
 'p3_conf',
 'p3_dog']
```

In [141]:

```
# Show the name of each column in twitter_api_extracted_data dataframe
list(twitter_api_extracted_data)
```

Out[141]:

```
['tweet_id', 'retweet_count', 'favorite_count']
```

In [142]:

```
# Check for duplicated columns in all these 3 datasets
duplicated_columns = pd.Series(list(twitter_archive) + list(image_predictions) + list(twitter_api_e
xtracted_data))
duplicated_columns[duplicated_columns.duplicated()]
```

Out[142]:

```
17    tweet_id
29    tweet_id
dtype: object
```

After assessing first dataframe named for twitter\_api\_extracted\_data, Here are the results derived from its assessment.

**twitter\_api\_extracted\_data:**

#### **Quality Issues**

- **Completeness:**
  - No missing data in all columns
- **Validity:**
  - No Validity Issues
- **Accuracy:**
  - No Accuracy Issue
- **Consistency:**
  - No Consistency Issues

#### **Tidiness Issues**

- No Tidiness Issues

## **Part III - Cleaning Data**

In this part, the process consists of the cleaning portion of the data wrangling process shown below as itemizing each one:

1. Define: convert our assessments into defined cleaning tasks. These definitions also serve as an instruction list so others (or yourself in the future) can look at your work and reproduce it.

2. Code: convert those definitions to code and run that code.
3. Test: test your dataset, visually or with code, to make sure your cleaning operations worked.

## Define

1. Remove Five unessential columns named for in\_reply\_to\_status\_id, in\_reply\_to\_user\_id, retweeted\_status\_id, retweeted\_status\_user\_id, and retweeted\_status\_timestamp in twitter\_archive dataframe.
2. Change the timestamp to correct datetime format in twitter\_archive dataframe.
3. Remove unessential html tags in source column of twitter\_archive dataframe.
4. Dog names: some dogs contain 'None', 'a', and 'an.'. All these names are changed as "None".
5. The four columns in twitter\_archive dataframe which are doggo, floof, pupper and puppo should be merged into one column named stage.
6. Drop duplicated jpg\_url and delete columns that won't be used for analysis in image\_predictions dataframe.
7. Change Non-descriptive column names named for p1, p1\_conf, p1\_dog, p2, p2\_conf, p2\_dog, p3, p3\_conf, p3\_dog into descriptive column names named for first\_prediction, first\_confidence, first\_dog, second\_prediction, second\_confidence, second\_dog, third\_prediction, third\_confidence, third\_dog consecutively in image\_predictions dataframe.
8. Merge all the 3 dataframe named for twitter\_archive, image\_predictions, twitter\_api\_extracted\_data to one dataframe based on tweet\_id.

In [143]:

```
# Make a copy of the all these 3 dataframes as table before doing the process of cleaning
twitter_archive_clean_df = twitter_archive.copy()
image_predictions_clean_df = image_predictions.copy()
twitter_api_extracted_data_clean_df = twitter_api_extracted_data.copy()
```

**Define 1:** Remove Five unessential columns named for in\_reply\_to\_status\_id, in\_reply\_to\_user\_id, retweeted\_status\_id, retweeted\_status\_user\_id, and retweeted\_status\_timestamp in twitter\_archive dataframe.

In [144]:

```
# Code
twitter_archive_clean_df.drop(['in_reply_to_status_id',
                              'in_reply_to_user_id',
                              'retweeted_status_id',
                              'retweeted_status_user_id',
                              'retweeted_status_timestamp'], axis = 1, inplace=True)
```

In [145]:

```
# Test
twitter_archive_clean_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id          2356 non-null int64
timestamp         2356 non-null object
source            2356 non-null object
text              2356 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: int64(3), object(9)
memory usage: 221.0+ KB
```

**Define 2 :** Change the timestamp to correct datetime format in twitter\_archive dataframe.

In [146]:

```
# Code
#Remove the time zone from the 'timestamp' column
twitter_archive_clean_df['timestamp'] = twitter_archive_clean_df['timestamp'].str.slice(start=0, st
```



```
op=-6)

# Change the 'timestamp' column to a datetime object
twitter_archive_clean_df['timestamp'] = pd.to_datetime(twitter_archive_clean_df['timestamp'], errors = 'coerce', format = "%Y-%m-%d %H:%M:%S")
```

In [147]:

```
# Test
twitter_archive_clean_df.loc[0]
```

Out [147]:

```
tweet_id      892420643555336193
timestamp     2017-08-01 16:23:56
source        <a href="http://twitter.com/download/iphone" r...
text          This is Phineas. He's a mystical boy. Only eve...
expanded_urls https://twitter.com/dog_rates/status/892420643...
rating_numerator      13
rating_denominator    10
name                  Phineas
doggo                 None
floofer              None
pupper              None
puppo               None
Name: 0, dtype: object
```

**Define 3 :** Remove unessential html tags in source column of twitter\_archive dataframe.

In [148]:

```
# Code
twitter_archive_clean_df["source"] = twitter_archive_clean_df["source"].str.replace(r'<(?:\a\b[^\>]*>|/a>)', '')

twitter_archive_clean_df["source"] = twitter_archive_clean_df["source"].astype('category')
```

In [149]:

```
# Test
twitter_archive_clean df["source"].value_counts()
```

Out [149]:

```
Twitter for iPhone      2221
Vine - Make a Scene    91
Twitter Web Client     33
TweetDeck              11
Name: source, dtype: int64
```

**Define 4 :** Dog names: some dogs contain 'None', 'a', and 'an.'. All these names are changed as "None".

In [150]:

```
# Code
# Create the array of small letter words (all are dictionary words) in the 'name' column.
non_names_dog = twitter_archive_clean_df["name"][(twitter_archive_clean_df["name"].str.islower() ==
True)].values

# Print the non_names_dog array
print(non_names_dog)
```

[ 'such' 'a' 'quite' 'quite' 'quite' 'not' 'one' 'incredibly' 'a' 'mad'  
'an' 'very' 'a' 'very' 'just' 'my' 'one' 'not' 'his' 'one' 'a' 'a' 'a'  
'an' 'very' 'actually' 'a' 'just' 'getting' 'mad' 'very' 'this'  
'unacceptable' 'all' 'a' 'old' 'a' 'infuriating' 'a' 'a' 'a' 'an' 'a' 'a'  
'very' 'getting' 'just' 'a' 'the' 'the' 'actually' 'by' 'a' 'officially'  
'a' 'the' 'the' 'a' 'a' 'a' 'a' 'life' 'a' 'one' 'a' 'a' 'a' 'light'  
'just' 'space' 'a' 'the' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'an' 'a'  
'the' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' ]

```
one a u u u u u u u u u quite a an a  
'an' 'the' 'the' 'a' 'a' 'an' 'a' 'a' 'a' 'a']
```

In [151]:

```
twitter_archive_clean_df.replace(to_replace = non_names_dog, value = np.NaN, inplace = True)
```

In [152]:

```
# Test  
  
# Re-create the non_names_dog array and it must be an empty array now.  
non_names_dog = twitter_archive_clean_df.name[(twitter_archive_clean_df.name.str.islower() == True)  
].values  
print(non_names_dog)
```

```
[]
```

**Define 5 :** The four columns in twitter\_archive dataframe which are doggo, floof, pupper and puppo should be merged into one column named stage.

In [153]:

```
# Code  
  
## Melt the doggo, floofer, pupper and puppo columns to dog and dogs_blood columns  
twitter_archive_clean_df = pd.melt(twitter_archive_clean_df,  
                                   id_vars=['tweet_id',  
                                           'timestamp',  
                                           'source',  
                                           'text',  
                                           'expanded_urls',  
                                           'rating_numerator',  
                                           'rating_denominator',  
                                           'name'],  
                                   var_name='dog', value_name='stage')  
  
## Drop dog column  
twitter_archive_clean_df = twitter_archive_clean_df.drop('dog', 1)  
  
## Sort by dogs_blood column and then drop duplicated based on tweet_id except the last occurrence  
twitter_archive_clean_df = twitter_archive_clean_df.sort_values('stage').drop_duplicates(subset='tweet_id', keep='last')
```

In [154]:

```
# Test  
twitter_archive_clean_df['stage'].value_counts()
```

Out[154]:

```
None          1976  
pupper         257  
doggo           83  
puppo           30  
floofer         10  
Name: stage, dtype: int64
```

**Define 6 :** Drop duplicated jpg\_url in image\_predictions dataframe.

In [155]:

```
# Code  
image_predictions_clean_df = image_predictions_clean_df.drop_duplicates(subset=['jpg_url'], keep='last')
```

In [156]:

```
# Test
```

```
sum(image_predictions_clean_df['jpg_url'].duplicated())
```

Out[156]:

0

**Define 7 :** Change Non-descriptive column names named for p1, p1\_conf, p1\_dog, p2, p2\_conf, p2\_dog, p3, p3\_conf, p3\_dog into descriptive column names named for first\_prediction, first\_confidence, first\_dog, second\_prediction, second\_confidence, second\_dog, third\_prediction, third\_confidence, third\_dog consecutively in image\_predictions dataframe

In [157]:

```
# Code
image_predictions_clean_df.rename(index = str, columns={"p1": "first_prediction",
                                                       "p1_conf": "first_confidence",
                                                       "p1_dog": "first_dog",
                                                       "p2": "second_prediction",
                                                       "p2_conf": "second_confidence",
                                                       "p2_dog": "second_dog",
                                                       "p3": "third_prediction",
                                                       "p3_conf": "third_confidence",
                                                       "p3_dog": "third_dog"}, inplace = True)
```

In [158]:

```
# Test
image_predictions_clean_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2009 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id          2009 non-null int64
jpg_url           2009 non-null object
img_num          2009 non-null int64
first_prediction  2009 non-null object
first_confidence  2009 non-null float64
first_dog         2009 non-null bool
second_prediction 2009 non-null object
second_confidence 2009 non-null float64
second_dog        2009 non-null bool
third_prediction  2009 non-null object
third_confidence  2009 non-null float64
third_dog         2009 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 162.8+ KB
```

**Define 8 :** Merge all the 3 dataframe named for twitter\_archive, image\_predictions, twitter\_api\_extracted\_data to one dataframe based on tweet\_id.

In [159]:

```
# Code
##Join all tables on Left Join using the Tweet_ID column
twitter_main_df = pd.merge(twitter_archive_clean_df, image_predictions_clean_df, how = 'left', on = ['tweet_id'])

# Only keeping rows that have pictures (jpg_url)
twitter_main_df = twitter_main_df[twitter_main_df['jpg_url'].notnull()]

##Create a new dataframe that merge full_twitter_df and tweet_json_df_clean
twitter_main_df = pd.merge(twitter_main_df, twitter_api_extracted_data, how = 'left', on = ['tweet_id'])
```

In [160]:

```
# Test
twitter_main_df.head()
```

Out[160]:

Out[155]:

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator
0	667550904950915073	2015-11-20 03:51:52	Twitter Web Client	RT @dogratingrating: Exceptional talent. Original...	https://twitter.com/dogratingrating/status/667...	12	
1	667550882905632768	2015-11-20 03:51:47	Twitter Web Client	RT @dogratingrating: Unoriginal idea. Blatant ...	https://twitter.com/dogratingrating/status/667...	5	
2	667549055577362432	2015-11-20 03:44:31	Twitter Web Client	Never seen dog like this. Breathes heavy. Tilt...	https://twitter.com/dog_rates/status/667549055...	1	
3	667546741521195010	2015-11-20 03:35:20	Twitter Web Client	Here is George. George took a selfie of his ne...	https://twitter.com/dog_rates/status/667546741...	9	
4	667544320556335104	2015-11-20 03:25:43	Twitter Web Client	This is Kial. Kial is either wearing a cape, w...	https://twitter.com/dog_rates/status/667544320...	10	

5 rows × 22 columns



In [161]:

```
twitter_main_df.tail()
```

Out[161]:

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator
2004	744995568523612160	2016-06-20 20:49:19	Twitter for iPhone	This is Abby. She got her face stuck in a glass...	https://twitter.com/dog_rates/status/744995568...	9	
2005	825535076884762624	2017-01-29 02:44:34	Twitter for iPhone	Here's a very loving and accepting puppo. Appe...	https://twitter.com/dog_rates/status/825535076...	14	
2006	825026590719483904	2017-01-27 17:04:02	Twitter for iPhone	Say hello to Pablo. He's one gorgeous puppo. A...	https://www.gofundme.com/my-puppys-double-cata...	12	
2007	803773340896923648	2016-11-30 01:31:12	Twitter for iPhone	This is Diogi. He fell in the pool as soon as ...	https://twitter.com/dog_rates/status/803773340...	12	
2008	859607811541651456	2017-05-03 03:17:27	Twitter for iPhone	Sorry for the lack of posts today. I came home...	https://twitter.com/dog_rates/status/859607811...	13	

5 rows × 22 columns



In [162]:

```
twitter_main_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2009 entries, 0 to 2008
Data columns (total 22 columns):
tweet_id          2009 non-null int64
timestamp         2009 non-null datetime64[ns]
source            2009 non-null category
```

```

text                2009 non-null object
expanded_urls       2009 non-null object
rating_numerator    2009 non-null int64
rating_denominator  2009 non-null int64
name                1911 non-null object
stage               2009 non-null object
jpg_url             2009 non-null object
img_num             2009 non-null float64
first_prediction    2009 non-null object
first_confidence    2009 non-null float64
first_dog           2009 non-null object
second_prediction   2009 non-null object
second_confidence   2009 non-null float64
second_dog          2009 non-null object
third_prediction    2009 non-null object
third_confidence    2009 non-null float64
third_dog           2009 non-null object
retweet_count       1994 non-null float64
favorite_count      1994 non-null float64
dtypes: category(1), datetime64[ns](1), float64(6), int64(3), object(11)
memory usage: 347.4+ KB

```

In [164]:

```

# Write dataframe to csv file named for
twitter_main_df.to_csv('twitter_main.csv', encoding='utf-8', index=False)

```

In [2]:

```

# Get all data from twitter_main.csv file
twitter_analyze_visualize_df = pd.read_csv('twitter_main.csv')

```

In [166]:

```
twitter_analyze_visualize_df.head()
```

Out[166]:

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_den
0	667550904950915073	2015-11-20 03:51:52	Twitter Web Client	RT @dogratingrating: Exceptional talent. Original...	https://twitter.com/dogratingrating/status/667...	12	
1	667550882905632768	2015-11-20 03:51:47	Twitter Web Client	RT @dogratingrating: Unoriginal idea. Blatant ...	https://twitter.com/dogratingrating/status/667...	5	
2	667549055577362432	2015-11-20 03:44:31	Twitter Web Client	Never seen dog like this. Breathes heavy. Tilt...	https://twitter.com/dog_rates/status/667549055...	1	
3	667546741521195010	2015-11-20 03:35:20	Twitter Web Client	Here is George. George took a selfie of his ne...	https://twitter.com/dog_rates/status/667546741...	9	
4	667544320556335104	2015-11-20 03:25:43	Twitter Web Client	This is Kial. Kial is either wearing a cape, w...	https://twitter.com/dog_rates/status/667544320...	10	

5 rows × 22 columns

## Part IV - Analyzing and Visualizing Wrangled data

In this part, there are some different kinds of graphs and images based on analyzing and visualizing wrangle data named for twitter\_analyze\_visualize\_df. All these are itemized in detail below.

1. Which sources are used for retweet, favorite tweet?
2. Which are the most 10 popular dog names?
3. Which is the most favorite and retweeted tweet?
4. What is the favorite and retweeted tweet count in terms of dog breeds based on first prediction second prediction third

- What is the average and retweeted tweet count in terms of dog breeds based on first prediction, second prediction, third prediction?
- What is the distribution of dogs' stage?
- How have the Retweet and Favorite Counts been changed over the time?

## 1. Which sources are used for retweet, favorite tweet?

In [167]:

```
# show Which sources are used for retweet, favorite tweet
twitter_analyze_visualize_df["source"].value_counts()
```

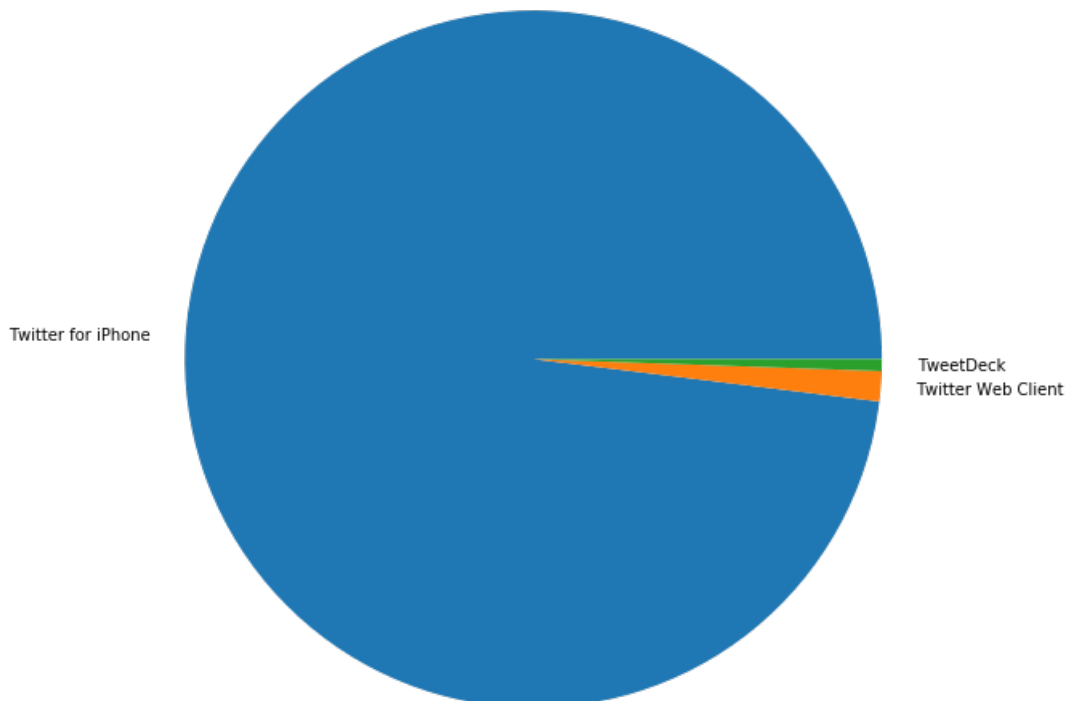
Out[167]:

```
Twitter for iPhone    1970
Twitter Web Client    28
TweetDeck             11
Name: source, dtype: int64
```

In [169]:

```
# show it via pie chart
twitter_analyze_visualize_df["source"].value_counts().plot(kind='pie',
                                                            label='',
                                                            title='Which sources are used for retweet,
favorite tweet');
plt.rcParams["figure.figsize"] = [15, 10]
```

Which sources are used for retweet, favorite tweet



## 2. Which are the most 10 popular dog names?

In [170]:

```
# show which the most 10 popular dog names are
twitter_analyze_visualize_df['name'].value_counts().head(10)
```

Out[170]:

```

None          561
Charlie        11
Lucy           10
Oliver         10
Cooper         10
Penny          9
Tucker         9
Winston        8
Sadie          8
Lola           7
Name: name, dtype: int64

```

```
In [171]:
```

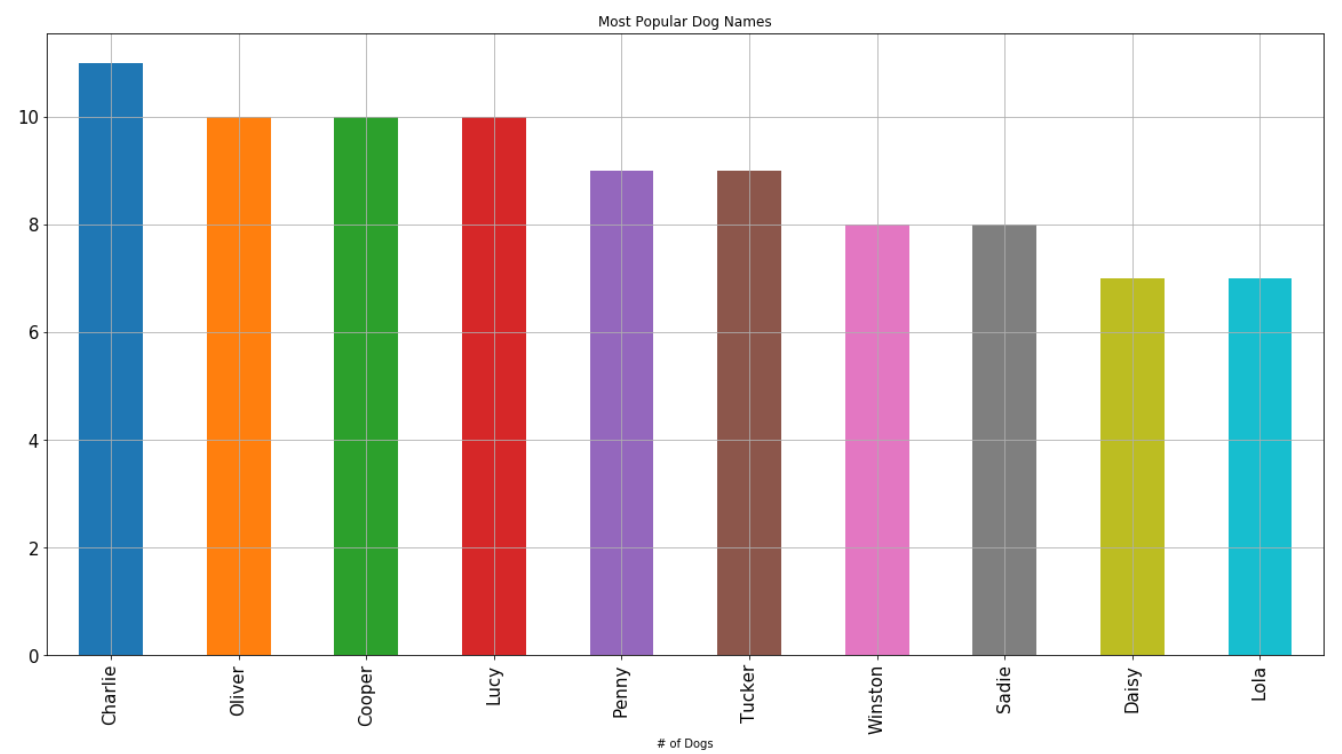
```

# By not taking into account "None" value, show the most 10 popular dog names.
dogs_names = twitter_analyze_visualize_df.query('name != "None"')
most_popular_dog_names_10 = dogs_names['name'].value_counts().head(10)

most_popular_dog_names_10.plot(kind='bar', figsize=(20,10),grid=True, fontsize =15,
                                title = 'Most Popular Dog Names').set_xlabel('# of Dogs')

plt.show()

```



### 3. Which is the most favorite and retweeted tweet?

```
In [172]:
```

```

most_favorite_and_retweeted_tweet = twitter_analyze_visualize_df.sort_values(['retweet_count', 'favorite_count'], ascending=[False, False]).head(1)
most_favorite_and_retweeted_tweet

```

```
Out[172]:
```

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator
1755	744234799360020481	2016-06-18 18:26:18	Twitter for iPhone	Here's a doggo realizing you can stand in a po...	https://twitter.com/dog_rates/status/744234799...	13	

1 rows × 22 columns

In [173]:

```
# This is an image for tweet_id 744234799360020481
id_image = most_favorite_and_retweeted_tweet[most_favorite_and_retweeted_tweet["tweet_id"] ==
744234799360020481]["jpg_url"].values[0]
# show images Via Image Library
Image(url = str(id_image))
```

Out[173]:



**4. What is the favorite and retweeted tweet count in terms of dog breeds based on first prediction second prediction, third prediction?**

In [7]:

```
# create a dataframe with only the rows based on favorite_count and retweet_count in terms of their "True" values in first prediction, second prediction, third prediction
retweet_favorite_df_prediction_df = twitter_analyze_visualize_df.query('first_dog == True and second_dog == True and third_dog == True')
retweet_favorite_df_prediction_1_df = retweet_favorite_df_prediction_df[['favorite_count', 'retweet_count', 'first_prediction']]
retweet_favorite_df_prediction_2_df = retweet_favorite_df_prediction_df[['favorite_count', 'retweet_count', 'second_prediction']]
retweet_favorite_df_prediction_3_df = retweet_favorite_df_prediction_df[['favorite_count', 'retweet_count', 'third_prediction']]
```

In [9]:

```
prediction_1_df = pd.DataFrame(retweet_favorite_df_prediction_1_df.groupby('first_prediction').sum
```



```
([['favorite_count', 'retweet_count']])
prediction_1_df.sort_values(by = "retweet_count", ascending = False, inplace = True)
prediction_1_df.reset_index()
prediction_1_df = prediction_1_df.head(10)
prediction_1_df
```

Out[9]:

	favorite_count	retweet_count
first_prediction		
golden_retriever	1259003.0	372208.0
Pembroke	830021.0	221510.0
Labrador_retriever	588405.0	160122.0
Samoyed	282711.0	105810.0
French_bulldog	355085.0	89442.0
Chihuahua	262604.0	84795.0
Eskimo_dog	220176.0	83265.0
chow	263752.0	73997.0
cocker_spaniel	202716.0	73694.0
toy_poodle	180169.0	68428.0

In [10]:

```
prediction_2_df = pd.DataFrame(retweet_favorite_df_prediction_2_df.groupby('second_prediction').sum()
[['favorite_count', 'retweet_count']])
prediction_2_df.sort_values(by = "retweet_count", ascending = False, inplace = True)
prediction_2_df.reset_index()
prediction_2_df = prediction_2_df.head(10)
prediction_2_df
```

Out[10]:

	favorite_count	retweet_count
second_prediction		
Labrador_retriever	1027108.0	290653.0
Cardigan	721122.0	196299.0
golden_retriever	445197.0	146262.0
Pomeranian	369101.0	124073.0
Siberian_husky	350098.0	115289.0
Boston_bull	336785.0	107438.0
toy_poodle	147984.0	76688.0
kuvasz	207722.0	69500.0
Chesapeake_Bay_retriever	228717.0	68988.0
Eskimo_dog	262109.0	68172.0

In [11]:

```
prediction_3_df = pd.DataFrame(retweet_favorite_df_prediction_3_df.groupby('third_prediction').sum()
[['favorite_count', 'retweet_count']])
prediction_3_df.sort_values(by = "retweet_count", ascending = False, inplace = True)
prediction_3_df.reset_index()
prediction_3_df = prediction_3_df.head(10)
prediction_3_df
```

Out[11]:

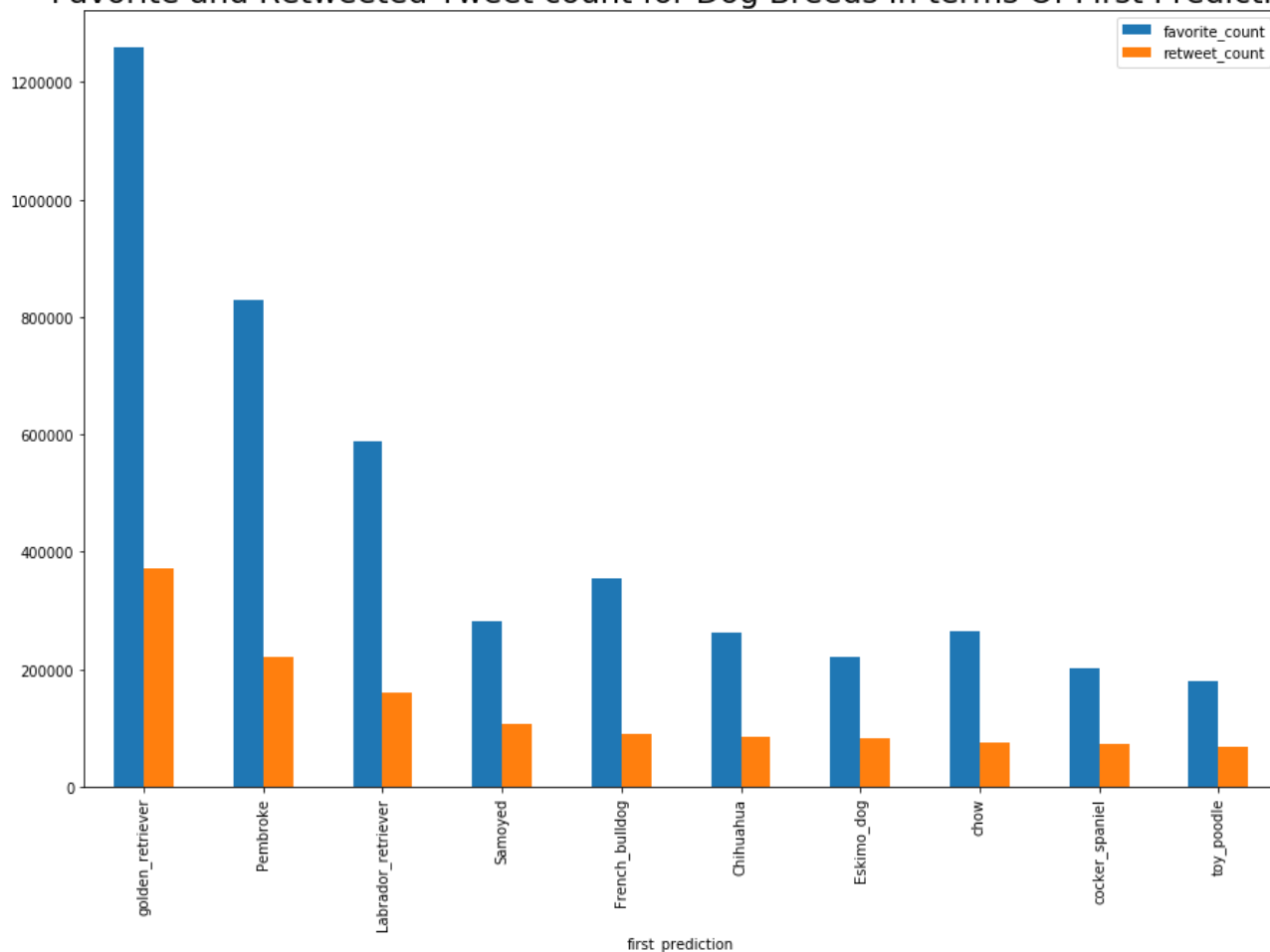
	favorite_count	retweet_count
third_prediction		
Labrador_retriever	554810.0	157677.0

Labrador_retriever	551848.0	157677.0
malamute	286012.0	104475.0
third_prediction		
Chihuahua	365030.0	102884.0
Pekinese	262206.0	93240.0
kelpie	286988.0	79871.0
Staffordshire_bulldog	256394.0	75785.0
Cardigan	235315.0	73235.0
kuvasz	242716.0	72308.0
basenji	219123.0	68504.0
Pembroke	182990.0	62868.0

In [32]:

```
ax1 = prediction_1_df.plot.bar()
ax1.set_title("Favorite and Retweeted Tweet count for Dog Breeds In terms Of First Prediction", fontsize=22)
plt.show()
```

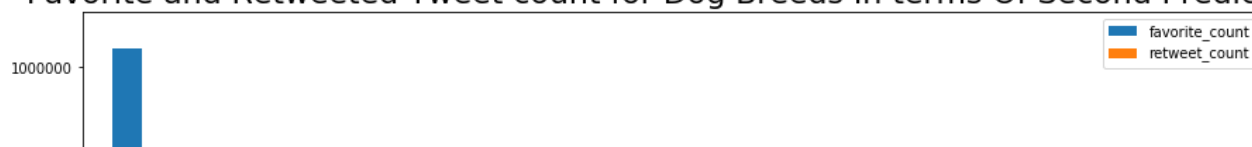
Favorite and Retweeted Tweet count for Dog Breeds In terms Of First Prediction

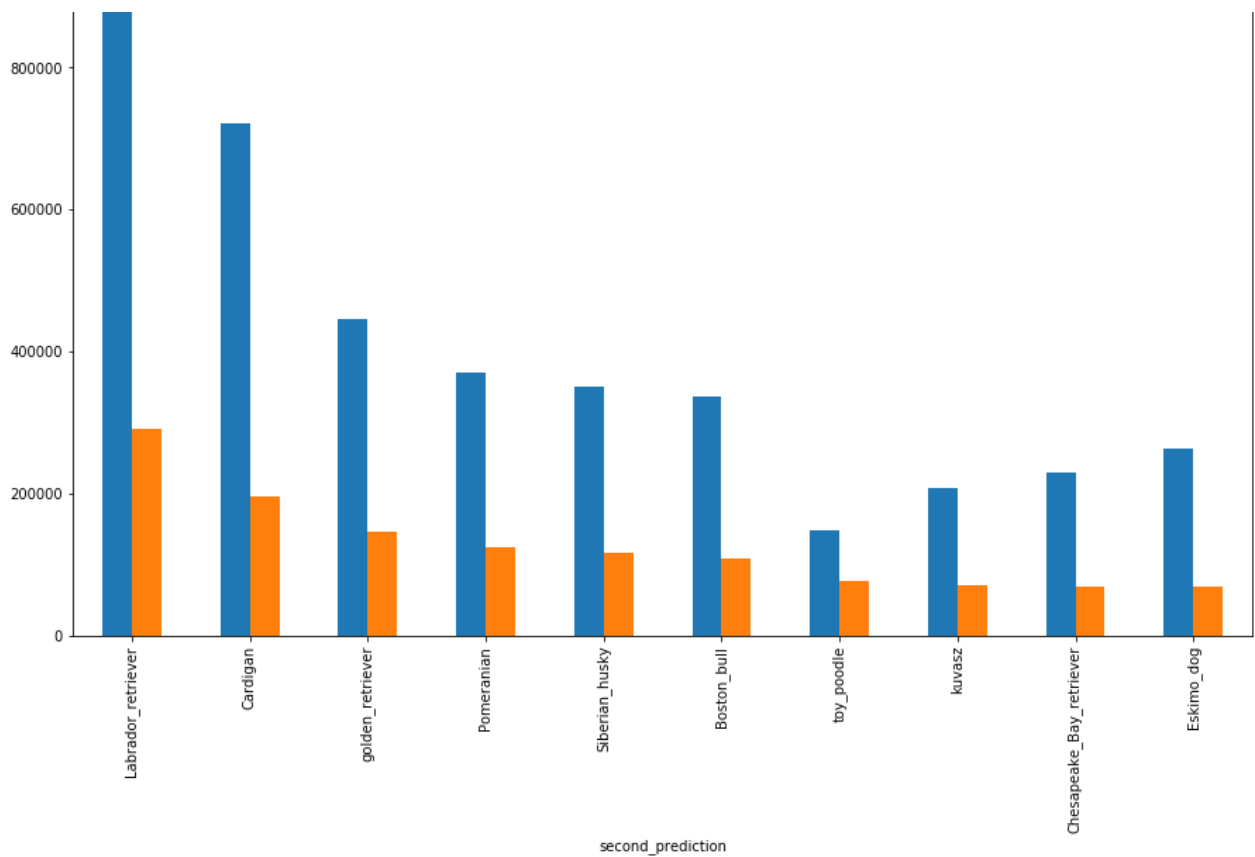


In [33]:

```
ax2 = prediction_2_df.plot.bar()
ax2.set_title("Favorite and Retweeted Tweet count for Dog Breeds In terms Of Second Prediction", fontsize=22)
plt.show()
```

Favorite and Retweeted Tweet count for Dog Breeds In terms Of Second Prediction

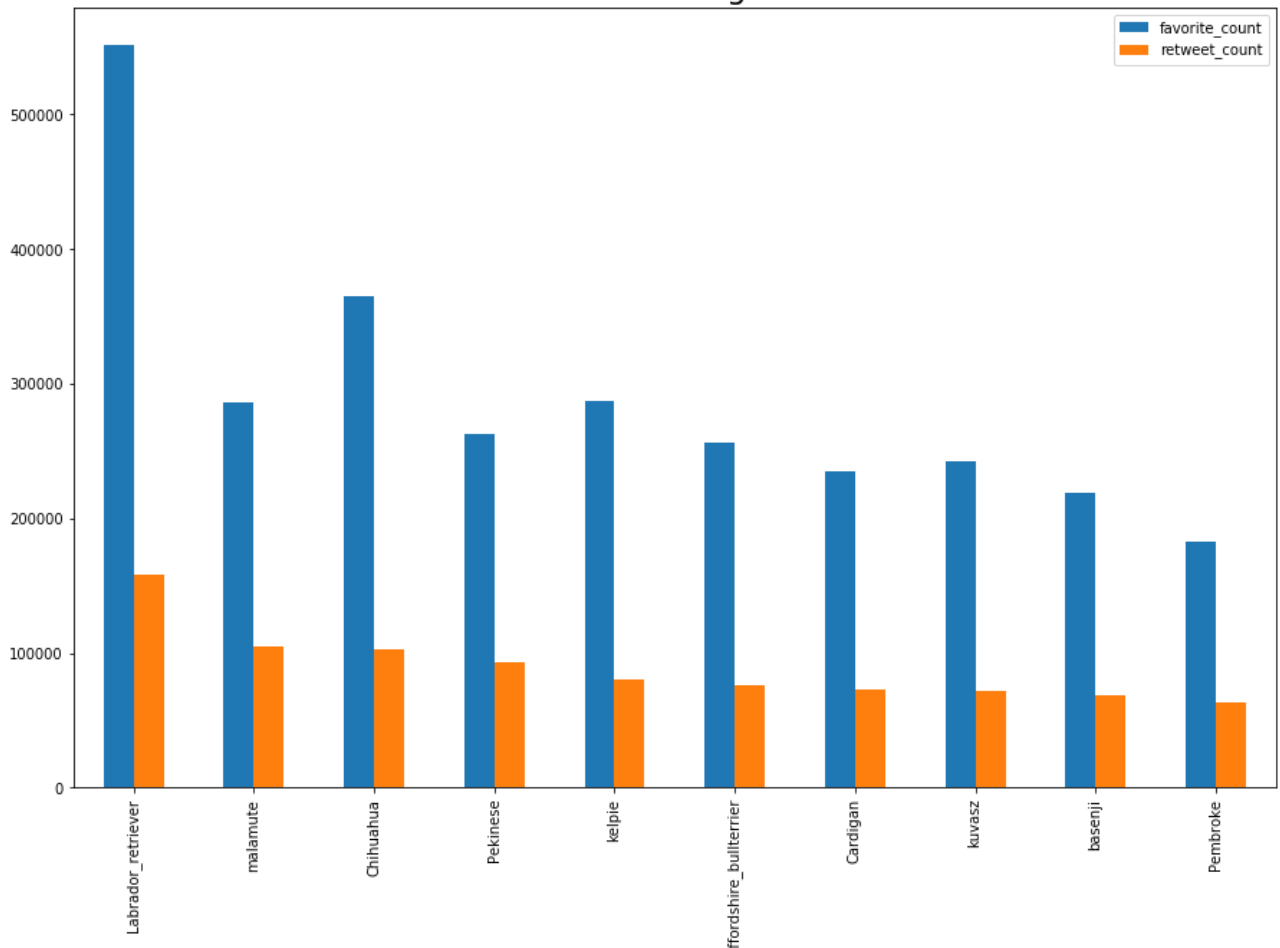




In [34]:

```
ax3 = prediction_3_df.plot.bar()
ax3.set_title("Favorite and Retweeted Tweet count for Dog Breeds In terms Of Third Prediction", fontsize=22)
plt.show()
```

Favorite and Retweeted Tweet count for Dog Breeds In terms Of Third Prediction



## 5. What is the distribution of dogs' stage?

In [174]:

```
# show the distribution of dogs' stage
twitter_analyze_visualize_df['stage'].value_counts()
```

Out[174]:

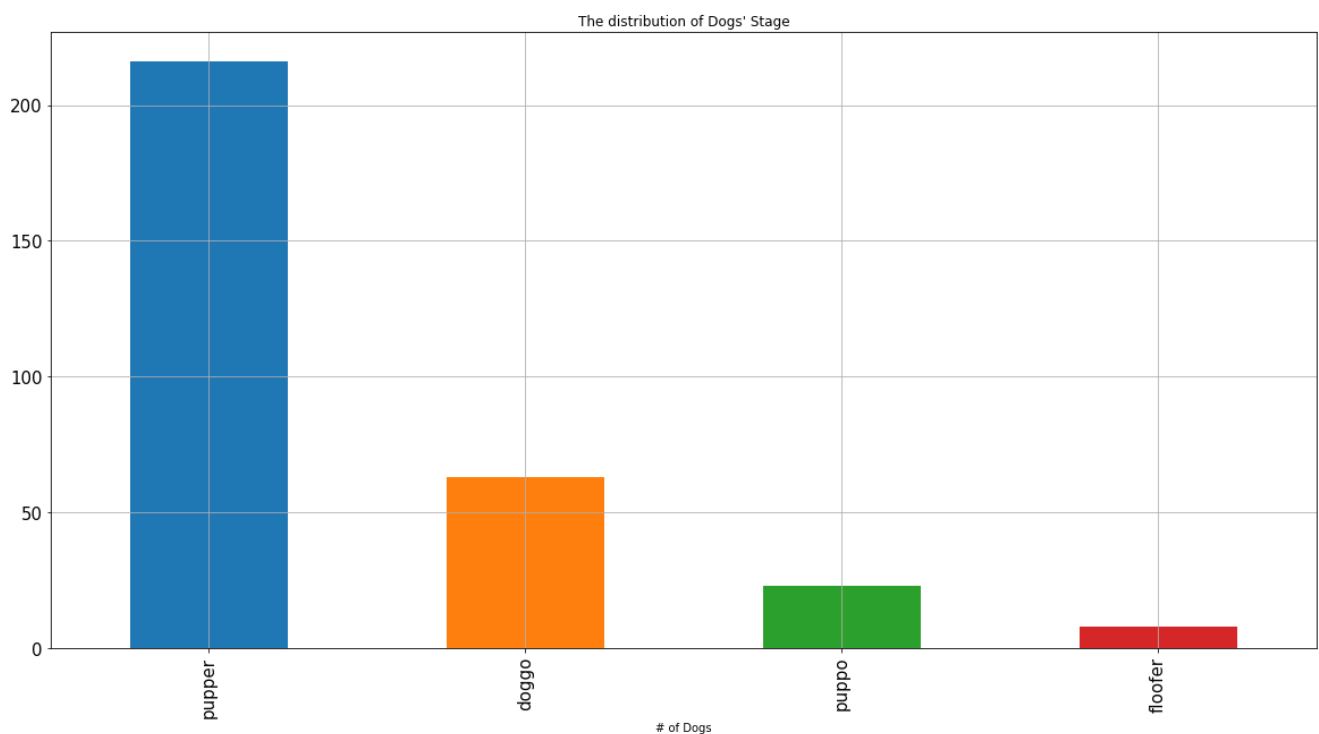
```
None      1699
pupper    216
doggo      63
puppo     23
floofer     8
Name: stage, dtype: int64
```

In [175]:

```
# By not taking into account "None" value, show the most 10 popular dog names.
dogs_names = twitter_analyze_visualize_df.query('stage != "None"')
most_popular_dog_names_10 = dogs_names['stage'].value_counts().head(10)

most_popular_dog_names_10.plot(kind='bar', figsize=(20,10), grid=True, fontsize=15,
                                title="The distribution of Dogs' Stage").set_xlabel('# of Dogs')

plt.show()
```



## 6. How have the Retweet and Favorite Counts been changed over the time?

In [4]:

```
twitter_analyze_visualize_df['timestamp'] =
pd.to_datetime(twitter_analyze_visualize_df['timestamp'], errors='coerce', format="%Y-%m-%d %H:%M:%S")
```

In [9]:

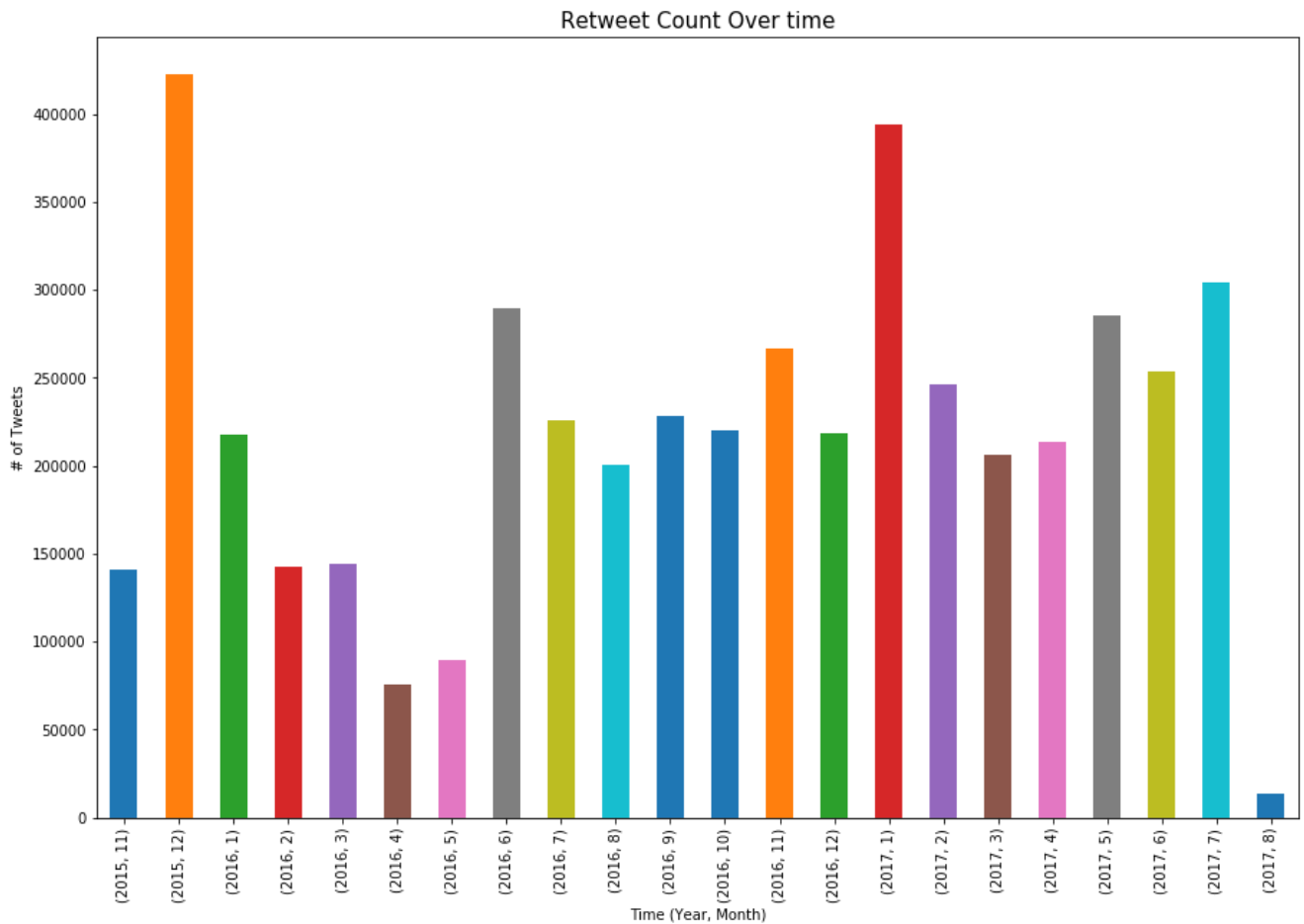
```
twitter_analyze_visualize_df["retweet_count"].groupby([
twitter_analyze_visualize_df["timestamp"]
```

```

t.year,
                                twitter_analyze_visualize_df["timestamp"]
t.month
                                ]
                                ).sum().plot(kind='bar')

plt.title('Retweet Count Over time', size =15)
plt.ylabel('# of Tweets')
plt.xlabel('Time (Year, Month)')
plt.show()
plt.rcParams["figure.figsize"] = [15, 10]

```



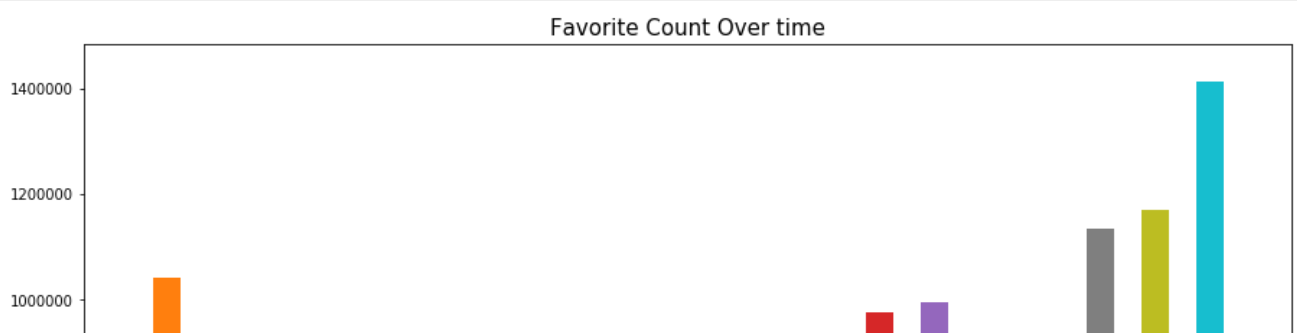
In [11]:

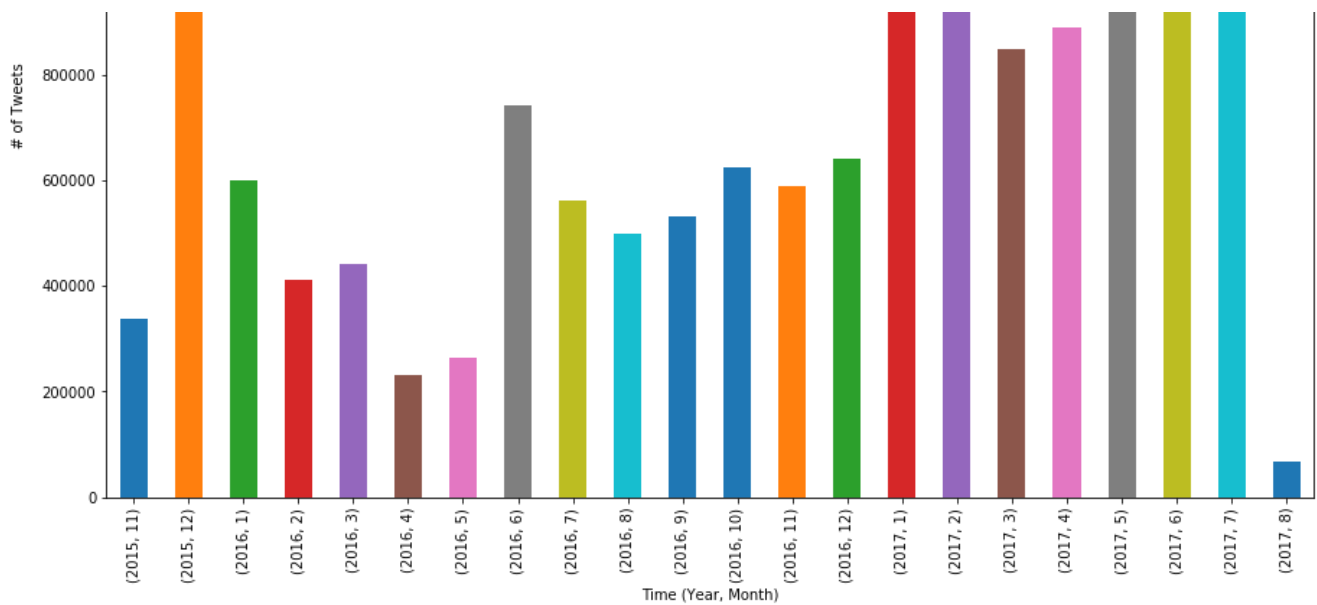
```

twitter_analyze_visualize_df["favorite_count"].groupby([
                                twitter_analyze_visualize_df["timestamp"]
dt.year,
                                twitter_analyze_visualize_df["timestamp"]
dt.month
                                ]
                                ).sum().plot(kind='bar')

plt.title('Favorite Count Over time', size =15)
plt.ylabel('# of Tweets')
plt.xlabel('Time (Year, Month)')
plt.show()
plt.rcParams["figure.figsize"] = [15, 10]

```





## Part V - Conclusion

To sum up, analyzing and wrangling data was a demanding process which gathers, assesses and lastly cleans the data to get a right dataframe that I can use in the project. Firstly, I could gather some data via different kinds of gathering methods like getting data from url link via request module, getting data from csv file and finally getting data from the usage of Twitter API. My milestone in this part was based on how to install tweepy in the Anaconda Jupyter because some command code snippets couldn't be worked. Then, I could pass to the part of assessment of data. In this part, I could check whether all dataframes from the part of gathering data have null value, duplicated data, how many values there were in each one as well as showing all types of data values in each one with its statistical description. After doing all these processes, I could determine the issues and tidiness of each dataframe. All these part was very exhausting for me because of determining each issue in terms of its kind in issue part in parallel with deciding what I did in tidiness part. After determining the issue and tidiness of each dataframe, I could clean all them with respect to their troubles. Then, I could combine all cleaning dataframes with only just one dataframe used for the visualization. Finally, in the visualization part, I could determine what kind of graph I drew thanks to defining 6 questions I formed then I drew all plots for each question. In this part, my limitation is based on how to show datetime in determining the sum of favorite and retweeted tweet in each time period because the period of time couldn't be appeared at the bottom of plot in 6th question. I could fix the problem to process the conversion of datetime. After all, I completed the project with my all effort and determination.