

# Sathish\_Kumar\_Rajendiran\_Week8\_8.2\_Twitter\_Activity

August 19, 2020

Name: Sathish Kumar Rajendiran Task: Week8: 8.2 MongoDB & Tweets Collection Date: 8/19/2020

```
[ ]: #pip install pymongo library
!pip install pymongo
!pip install tweepy
```

```
[1]: #import libraries

# standard library
import os
import sys
from datetime import datetime
import time
import re

# csv, xls, pandas & json
import pandas as pd
import json
import csv
import xlrd

#twitter libraries
import tweepy
from tweepy import StreamListener
from tweepy import Stream
import preprocessor as p
# from tweet-preprocessor import clean,tokenize,parse

#MongoDB libraries
import pymongo
from pymongo import MongoClient

#visualization
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS
import seaborn as sns
%matplotlib inline
```

```
import timeit
# import random

os.getcwd()
```

```
[1]: '/Users/sathishrajendiran/ist652-python'
```

```
[12]: # Connecting to the database
# Connection to Mongo DB
try:
    client = MongoClient('localhost', 27017)
    print('Authentication OK - You're now connected to the MongoDB.\n')
    # use database named usgs or create it if not there already
    db = client.bball
    # create collection named earthquakes or create it if not there already
    coll = db.bbcoll

    # print('MongoDB: Tweets db with tweets collection has been created')
    print('MongoDB database: ' + str(db))
    print('MongoDB collection: ' + str(coll))

except pymongo.errors.ConnectionFailure as e:
    print('Could not connect to MongoDB: %s' % e )
```

Authentication OK - You're now connected to the MongoDB.

MongoDB database: Database(MongoClient(host=['localhost:27017'], document\_class=dict, tz\_aware=False, connect=True), 'bball')

MongoDB collection: Collection(Database(MongoClient(host=['localhost:27017'], document\_class=dict, tz\_aware=False, connect=True), 'bball'), 'bbcoll')

```
[3]: #list database defined
client.list_database_names
```

```
[3]: <bound method MongoClient.list_database_names of
MongoClient(host=['localhost:27017'], document_class=dict, tz_aware=False,
connect=True)>
```

```
[13]: #search the first item from the collection
coll.find_one()
```

```
[13]: {'_id': ObjectId('58d992de64a4f3e56d2db3c7'),
'user': {'profile_background_tile': True,
'friends_count': 217,
'profile_sidebar_fill_color': 'EFEFEF',
'id_str': '513196438',
```

```

'is_translation_enabled': False,
'profile_link_color': 'F70808',
'followers_count': 256,
'location': '',
'protected': False,
'default_profile_image': False,
'contributors_enabled': False,
'favourites_count': 2187,
'profile_background_color': 'BFD0D9',
'statuses_count': 1104,
'id': 513196438,
'profile_banner_url':
'https://pbs.twimg.com/profile_banners/513196438/1437359097',
'created_at': 'Sat Mar 03 13:54:37 +0000 2012',
'profile_image_url_https':
'https://pbs.twimg.com/profile_images/732514428051181568/0k0Va8Ia_normal.jpg',
'time_zone': 'Eastern Time (US & Canada)',
'follow_request_sent': None,
'listed_count': 1,
'utc_offset': -14400,
'lang': 'en',
'is_translator': False,
'name': 'Will',
'description': "Do not worry about tomorrow, for tomorrow will worry about
itself.-Matthew 6:34 HSC'20",
'profile_use_background_image': True,
'verified': False,
'geo_enabled': True,
'profile_text_color': '333333',
'profile_image_url':
'http://pbs.twimg.com/profile_images/732514428051181568/0k0Va8Ia_normal.jpg',
'entities': {'description': {'urls': []}},
'notifications': None,
'url': None,
'translator_type': 'none',
'has_extended_profile': True,
'default_profile': False,
'screen_name': 'frenchythe1st',
'following': None,
'profile_background_image_url': 'http://pbs.twimg.com/profile_background_image
s/850934042/26e0a43c1f821ac098571fb3de80944c.jpeg',
'profile_sidebar_border_color': 'FFFFFF',
'profile_background_image_url_https': 'https://pbs.twimg.com/profile_backgroun
d_images/850934042/26e0a43c1f821ac098571fb3de80944c.jpeg',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'favorited': False,
'in_reply_to_user_id': None,

```

```

'text': 'RT @HowardWKYT: The final seconds of the Kentucky-North Carolina game
was an emotional roller coaster for Big Blue fans. #marchmadness #WKY...',
'retweet_count': 9739,
'id': 8.464896333870899e+17,
'in_reply_to_screen_name': None,
'created_at': 'Mon Mar 27 22:30:30 +0000 2017',
'is_quote_status': False,
'retweeted_status': {'user': {'profile_background_tile': False,
'friends_count': 573,
'profile_sidebar_fill_color': 'C0DFEC',
'id_str': '27059989',
'is_translation_enabled': False,
'profile_link_color': '0084B4',
'followers_count': 3330,
'location': 'Lexington, KY',
'protected': False,
'default_profile_image': False,
'contributors_enabled': False,
'favourites_count': 1231,
'profile_background_color': '000000',
'statuses_count': 9463,
'id': 27059989,
'profile_banner_url':
'https://pbs.twimg.com/profile_banners/27059989/1431999643',
'created_at': 'Fri Mar 27 18:09:16 +0000 2009',
'profile_image_url_https':
'https://pbs.twimg.com/profile_images/575376840304386049/atONJG3G_normal.jpeg',
'time_zone': 'Eastern Time (US & Canada)',
'follow_request_sent': None,
'listed_count': 77,
'utc_offset': -14400,
'lang': 'en',
'is_translator': False,
'name': 'Lee K. Howard',
'description': 'Sports Anchor/Reporter for CBS/FOX in Lexington Kentucky,
providing sports news and my random rants and chants!',
'profile_use_background_image': True,
'verified': True,
'geo_enabled': True,
'profile_text_color': '333333',
'profile_image_url':
'http://pbs.twimg.com/profile_images/575376840304386049/atONJG3G_normal.jpeg',
'entities': {'description': {'urls': []},
'url': {'urls': [{'expanded_url':
'http://www.facebook.com/profile.php?id=100003241678454',
'display_url': 'facebook.com/profile.php?id...',
"indices': [0, 22],

```

```

    'url': 'http://t.co/C2UFfLkjB1'}}}},
    'notifications': None,
    'url': 'http://t.co/C2UFfLkjB1',
    'translator_type': 'none',
    'has_extended_profile': False,
    'default_profile': False,
    'screen_name': 'HowardWKYT',
    'following': None,
    'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/596008542/6qfqg0jndpp3su53ay5.jpeg',
    'profile_sidebar_border_color': 'A8C7F7',
    'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/596008542/6qfqg0jndpp3su53ay5.jpeg',
    'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
    'favorited': False,
    'in_reply_to_user_id': None,
    'text': 'The final seconds of the Kentucky-North Carolina game was an emotional roller coaster for Big Blue fans.... https://t.co/TPZ6PuXHxH',
    'possibly_sensitive': False,
    'retweet_count': 9739,
    'id': 8.462142300261786e+17,
    'in_reply_to_screen_name': None,
    'created_at': 'Mon Mar 27 04:16:08 +0000 2017',
    'is_quote_status': False,
    'favorite_count': 12590,
    'contributors': None,
    'lang': 'en',
    'in_reply_to_status_id_str': None,
    'source': '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>',
    'in_reply_to_status_id': None,
    'place': {'country': 'United States',
    'contained_within': [],
    'id': '6ffcf3b0b904bbcb',
    'country_code': 'US',
    'bounding_box': {'type': 'Polygon',
    'coordinates': [[[-89.57151, 36.497129],
    [-81.964971, 36.497129],
    [-81.964971, 39.147359],
    [-89.57151, 39.147359]]]},
    'place_type': 'admin',
    'full_name': 'Kentucky, USA',
    'url': 'https://api.twitter.com/1.1/geo/id/6ffcf3b0b904bbcb.json',
    'name': 'Kentucky',
    'attributes': {}},
    'entities': {'hashtags': [],
    'urls': [{'expanded_url':

```

```
'https://twitter.com/i/web/status/846214230026178564',
  'display_url': 'twitter.com/i/web/status/8...',
  'indices': [106, 129],
  'url': 'https://t.co/TPZ6PuXHxH'}],
  'symbols': [],
  'user_mentions': [],
  'geo': None,
  'truncated': True,
  'coordinates': None,
  'in_reply_to_user_id_str': None,
  'retweeted': False,
  'id_str': '846214230026178564'},
  'favorite_count': 0,
  'contributors': None,
  'lang': 'en',
  'in_reply_to_status_id_str': None,
  'source': '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter
for iPhone</a>',
  'in_reply_to_status_id': None,
  'place': None,
  'entities': {'hashtags': [{'indices': [121, 134], 'text': 'marchmadness'}]},
  'urls': [],
  'symbols': [],
  'user_mentions': [{'indices': [3, 14],
    'name': 'Lee K. Howard',
    'id': 27059989,
    'id_str': '27059989',
    'screen_name': 'HowardWKYT'}]},
  'geo': None,
  'truncated': False,
  'coordinates': None,
  'in_reply_to_user_id_str': None,
  'retweeted': False,
  'id_str': '846489633387089920'}
```

```
[14]: docs = coll.find()
print('docs cursor ready!')
```

docs cursor ready!

```
[15]: doclist = [tweet for tweet in docs]
print('Number of docs: ',len(doclist))
```

Number of docs: 2000

```
[17]: def print_tweet_data(tweets):
    for tweet in tweets:
```

```
print('\nDate:', tweet['created_at'])
print('From:', tweet['user']['name'])
print('Screen Name:', tweet['user']['screen_name'])
print('Message', tweet['text'])
```

```
[18]: print_tweet_data(doclist[:5])
```

Date: Mon Mar 27 22:30:30 +0000 2017

From: Will

Screen Name: frenchythe1st

Message RT @HowardWKYT: The final seconds of the Kentucky-North Carolina game was an emotional roller coaster for Big Blue fans. #marchmadness #WKY...

Date: Mon Mar 27 22:30:22 +0000 2017

From: Kahlen Donatell

Screen Name: KahDon

Message RT @WhistleSports: When you perfectly time the #UNC buzzer beater #MarchMadness

(via:@SamuelGrubbs1 ) <https://t.co/0l2ibpZjB4>

Date: Mon Mar 27 22:30:21 +0000 2017

From: Jesús

Screen Name: \_jonwilliams

Message RT @HowardWKYT: The final seconds of the Kentucky-North Carolina game was an emotional roller coaster for Big Blue fans. #marchmadness #WKY...

Date: Mon Mar 27 22:30:14 +0000 2017

From: plug,

Screen Name: 21shawtyy

Message RT @BleacherReport: And then there were four... #MarchMadness  
<https://t.co/0MbxpgAuUC>

Date: Mon Mar 27 22:30:11 +0000 2017

From: Joni Dickerson

Screen Name: jonidickerson

Message RT @mycarolinastdnt: RT if you'll be cheering on @GamecockWBB and @dawnstaley tonight. Let's Go Gamecocks! #MarchMadness