Zili&Jiabo_6210_hw3

April 8, 2020

1 Video Game and Social Media Database

Group Member: Zili Huang, Jiabo Cheng Github : https://github.com/ZiliHuang1

1.1 Abstract

We are working on a video game database with game informatin gathered from giantbomb.com API and social media data collected from twitter API. After collecting the data, we will build a noSQL database using MongoDB Atlas and insert the dataset. The NoSQL database extend the SQL database and would be able to answer the questions of this assignment.

1.2 Data

Data sources: Giantbomb.comhttps://www.giantbomb.com/forums/api-developers-3017/ Data consists of: Detailed video game company information and video game information. Twitter: Data consists of: user and their friends, streaming tweets about related topics.

1.3 Step One: Converting SQL database to NoSQL database.

1.3.1 Get information about video games and their companies.

[7]:

```
import urllib
import requests
import json
import csv
import pymongo
import pandas as pd

client = pymongo.MongoClient('mongodb+srv://m220student:
 ↪m220password@mflix-6auna.mongodb.net/test?retryWrites=true&w=majority')

hw03 = client.hw03
game = hw03.game
```

[ ]:

```
user_agent = 'Mozilla/4.0 (compatible; MSIE 5.5; Windows NT)'
headers = {'User-Agent':user_agent}

with open('companylist.csv')as comp:
    c_csv = csv.reader(comp)
    mongodb = []
    for c in c_csv:
        url ='http://www.giantbomb.com/api/company/'+ c[1] + '/?
 ↪api_key=75a93f4b071b6002e1f328293149fb3cd166b91a&format=json'
        #print(url)
        req_c = urllib.request.Request(url=url, headers=headers)
        json_obj_c = urllib.request.urlopen(req_c)
        data_c = json.load(json_obj_c)
        res = data_c['results']
        for i in res['published_games'][1:10]:
            url_g = i['api_detail_url'] + '?
 ↪api_key=75a93f4b071b6002e1f328293149fb3cd166b91a&format=json'
            print(i['name'])
            req_g = urllib.request.Request(url=url_g, headers=headers)
            json_obj_g = urllib.request.urlopen(req_g)
            data_g = json.load(json_obj_g)
            res_g = data_g['results']
            print(res_g)
            mongodb.append(res_g)

game.insert_many(mongodb)
```

Sample Data:

[ ]:

```
```

[8]:

```
pd.DataFrame(list(game.find()))
```

[8]:

```
                          _id                       aliases  \
0    5e8a97b31fb82067c2eb87e5  Star Wars - Jedi: Fallen Order
1    5e8a97b31fb82067c2eb87e6                      NFS Heat
2    5e8a97b31fb82067c2eb87e7                          None
3    5e8a97b31fb82067c2eb87e8                          None
4    5e8a97b31fb82067c2eb87e9                          None
..                          …                             …
157  5e8a97b31fb82067c2eb8882                          None
158  5e8a97b31fb82067c2eb8883                          None
159  5e8a97b31fb82067c2eb8884                          None
160  5e8a97b31fb82067c2eb8885                          None
161  5e8a97b31fb82067c2eb8886                          None

                                   api_detail_url           date_added  \
0    https://www.giantbomb.com/api/game/3030-66000/  2018-01-30 15:12:26
1    https://www.giantbomb.com/api/game/3030-73586/  2019-05-29 12:23:03
2    https://www.giantbomb.com/api/game/3030-75048/  2019-09-04 10:33:43
3    https://www.giantbomb.com/api/game/3030-73601/  2019-05-29 18:25:19
4    https://www.giantbomb.com/api/game/3030-73977/  2019-06-19 18:46:59
```

```
..                                               …                  …
157  https://www.giantbomb.com/api/game/3030-76007/  2019-11-15 02:40:50
158  https://www.giantbomb.com/api/game/3030-76006/  2019-11-15 02:31:18
159  https://www.giantbomb.com/api/game/3030-72779/  2019-04-07 00:33:48
160  https://www.giantbomb.com/api/game/3030-77744/  2020-03-27 08:43:03
161  https://www.giantbomb.com/api/game/3030-74176/  2019-07-05 23:53:17

       date_last_updated                                               deck  \
0    2019-11-19 12:28:24  An action-adventure game in the Star Wars seri…
1    2020-02-04 01:34:44  Street racers drive around a Miami-inspired ci…
2    2019-10-20 10:38:03  The third title in the Garden Warfare spin-off…
3    2019-10-15 19:28:32  EA Sports' storied soccer/football franchise a…
4    2019-10-13 10:40:24  EA Sports' hockey franchise returns for the 20…
..                   …                                                 …
157  2020-02-21 11:54:20  A second helping of 11 Namco games for the Eve…
158  2020-02-21 11:53:48      A compilation of 11 Namco games for Evercade.
159  2020-02-17 14:31:40  Sword Art Online: Alicization Lycoris is a Jap…
160  2020-03-27 08:47:05                 Famista returns for the year 2020.
161  2020-01-21 21:21:39      One Piece returns to the Musō series in 2020.

                                           description  expected_release_day  \
0    <h2>Overview</h2><p>Star Wars Jedi: Fallen Ord…                   NaN
1    <h2>Overview</h2><p>…</p><h2>Development</h2…                   NaN
2    <h2>Overview</h2><p>Plants vs. Zombies: Battle…                   NaN
3                                                 None                   NaN
4    <h2>Overview</h2><p>While the official release…                   NaN
..                                                 …                   …
157  <p>Included in this collection:</p><ul><li><a …                  22.0
158  <p>Included in this collection</p><ul><li><a h…                  22.0
159                                               None                   NaN
160                                               None                   NaN
161  <h2>Overview</h2><p><strong>One Piece: Pirate …                   NaN

     expected_release_month  expected_release_quarter  …  killed_characters  \
0                       NaN                       NaN  …               None
1                       NaN                       NaN  …               None
2                       NaN                       NaN  …               None
3                       NaN                       NaN  …               None
4                       NaN                       NaN  …               None
..                      …                       …  …                …
157                     5.0                       NaN  …               None
158                     5.0                       NaN  …               None
159                     NaN                       2.0  …               None
160                     NaN                       NaN  …               None
161                     NaN                       NaN  …               None

                                             locations  \
0    [{'api_detail_url': 'https://www.giantbomb.com…
1    [{'api_detail_url': 'https://www.giantbomb.com…
2                                                 None
3    [{'api_detail_url': 'https://www.giantbomb.com…
4                                                 None
..                                                 …
157                                               None
158  [{'api_detail_url': 'https://www.giantbomb.com…
159                                               None
160                                               None
161                                               None

                                               objects  \
0    [{'api_detail_url': 'https://www.giantbomb.com…
1                                                 None
2                                                 None
3                                                 None
4    [{'api_detail_url': 'https://www.giantbomb.com…
..                                                 …
157  [{'api_detail_url': 'https://www.giantbomb.com…
158  [{'api_detail_url': 'https://www.giantbomb.com…
159                                               None
160  [{'api_detail_url': 'https://www.giantbomb.com…
161  [{'api_detail_url': 'https://www.giantbomb.com…

                                                people  \
0    [{'api_detail_url': 'https://www.giantbomb.com…
1                                                 None
2                                                 None
3                                                 None
4    [{'api_detail_url': 'https://www.giantbomb.com…
..                                                 …
157                                               None
158                                               None
159                                               None
160                                               None
161                                               None

                                            publishers  \
0    [{'api_detail_url': 'https://www.giantbomb.com…
1    [{'api_detail_url': 'https://www.giantbomb.com…
2    [{'api_detail_url': 'https://www.giantbomb.com…
3    [{'api_detail_url': 'https://www.giantbomb.com…
4    [{'api_detail_url': 'https://www.giantbomb.com…
..                                                 …
157  [{'api_detail_url': 'https://www.giantbomb.com…
158  [{'api_detail_url': 'https://www.giantbomb.com…
```

```
159  [{'api_detail_url': 'https://www.giantbomb.com…
160  [{'api_detail_url': 'https://www.giantbomb.com…
161  [{'api_detail_url': 'https://www.giantbomb.com…

                                              releases  \
0    [{'api_detail_url': 'https://www.giantbomb.com…
1    [{'api_detail_url': 'https://www.giantbomb.com…
2    [{'api_detail_url': 'https://www.giantbomb.com…
3    [{'api_detail_url': 'https://www.giantbomb.com…
4    [{'api_detail_url': 'https://www.giantbomb.com…
..                                                 …
157  [{'api_detail_url': 'https://www.giantbomb.com…
158  [{'api_detail_url': 'https://www.giantbomb.com…
159                                                NaN
160                                                NaN
161  [{'api_detail_url': 'https://www.giantbomb.com…

                                               reviews  \
0    [{'api_detail_url': 'https://www.giantbomb.com…
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN
..                                                 …
157                                                NaN
158                                                NaN
159                                                NaN
160                                                NaN
161                                                NaN

                                         similar_games  \
0                                                 None
1                                                 None
2                                                 None
3    [{'api_detail_url': 'https://www.giantbomb.com…
4                                                 None
..                                                 …
157                                               None
158                                               None
159                                               None
160  [{'api_detail_url': 'https://www.giantbomb.com…
161                                               None

                                                themes dlcs
0    [{'api_detail_url': 'https://www.giantbomb.com…  NaN
1    [{'api_detail_url': 'https://www.giantbomb.com…  NaN
2    [{'api_detail_url': 'https://www.giantbomb.com…  NaN
```

```
3                                                  NaN  NaN
4                                                  NaN  NaN
..                                                 …  …
157                                                NaN  NaN
158                                                NaN  NaN
159  [{'api_detail_url': 'https://www.giantbomb.com…  NaN
160                                                NaN  NaN
161  [{'api_detail_url': 'https://www.giantbomb.com…  NaN

[162 rows x 43 columns]
```

[ ]:

```

```

1.4 Step Two: Extend Database by Getting Social Media Data.

1.4.1 Get Streaming Tweets Using Name of Games be the Related Hashtags.

[6]:

```
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream
import time
import pymongo
import json
```

[ ]:

```
class TwitterStreamer():

    def stream_tweets(self, timeout, stream,hash_tag_list):

        listener = StdOutListener(timeout,stream,hash_tag_list)
        auth = OAuthHandler('LwG8GPbodB36m4Bb6U5BKiggD',␣
 ↪"3EiAJXuRDci8nnFmdLAhRgIqtAbezYAVEJpN7eVbIwKpmvB2am")
        auth.
 ↪set_access_token('1174542277630468096-3Wn6VBSMLOgV1Kcq0glxTBjirzQzG5',␣
 ↪'pmXdaGMos8LbWytFREMzOLNagAvNiEn38yTCWTJEBRPRx')
        stream = Stream(auth, listener)
        stream.filter(track=hash_tag_list)

class StdOutListener(StreamListener):

    def __init__(self, timeout,stream,hash_tag_list):
        self.timeout= timeout
        self.fetched_tweets_filename=stream
        self.hash_tag_list = hash_tag_list

    def on_data(self, data):
        try:
            print(data)
            data = json.loads(data)
            self.fetched_tweets_filename.insert_one(data)

            # Limiting the duration of collecting data.
            if time.time() > timeout:
                print('Times up!')
                return False
            else:

                return True
        except BaseException as e:
            print("Error on_data %s" % str(e))
        return True

    def on_error(self, status):
        print(status)


if __name__ == '__main__':

    client = pymongo.MongoClient(
        'mongodb+srv://m220student:m220password@mflix-6auna.mongodb.net/test?
 ↪retryWrites=true&w=majority')
    hw03 = client.hw03
    game = hw03.game
    stream = hw03.stream
    game =game.distinct('name')

    timeout = time.time() + 60 * 10
    hash_tag_list = []
    for i in game:

        hash_tag_list.append(i)
        print(i)

    twitter_streamer = TwitterStreamer()
    twitter_streamer.stream_tweets(timeout,stream,hash_tag_list)
```

1.4.2 Sample Data:

[9]:

```
hw03 = client.hw03
str = hw03.stream
pd.DataFrame(list(str.find()))
```

[9]:

```
                           _id                      created_at  \
0     5e8ba01550126ee36a48ff7c  Mon Apr 06 21:32:50 +0000 2020
1     5e8ba11322d419e75f047f5f  Mon Apr 06 21:37:18 +0000 2020
```

```
2     5e8bc7270af538741e89b68d  Tue Apr 07 00:19:45 +0000 2020
3     5e8bc7300af538741e89b68e  Tue Apr 07 00:19:55 +0000 2020
4     5e8bc76b0af538741e89b68f  Tue Apr 07 00:20:54 +0000 2020
…                        …                              …
3604  5e8d52380df45260e0722508  Wed Apr 08 04:24:53 +0000 2020
3605  5e8d52380df45260e0722509  Wed Apr 08 04:24:53 +0000 2020
3606  5e8d52380df45260e072250a  Wed Apr 08 04:24:53 +0000 2020
3607  5e8d52380df45260e072250b  Wed Apr 08 04:24:53 +0000 2020
3608  5e8d52380df45260e072250c  Wed Apr 08 04:24:53 +0000 2020

                       id               id_str  \
0     1247276101979844611  1247276101979844611
1     1247277226405625856  1247277226405625856
2     1247318108722925574  1247318108722925574
3     1247318150200295426  1247318150200295426
4     1247318395181305857  1247318395181305857
…                     …                    …
3604  1247742183588777984  1247742183588777984
3605  1247742184268292103  1247742184268292103
3606  1247742184901627906  1247742184901627906
3607  1247742185371365376  1247742185371365376
3608  1247742185975549952  1247742185975549952

                                                   text  \
0     RT @JuanjoSaezF1: Pues bueno, un SORTEO para h…
1     RT @InsideDlpCastle: Bonjour à toutes et à tou…
2     @Codemasters  Please, get an update to F1 2019…
3     @ana_olivm2 Prefiro o mais fácil, porque não t…
4     RT @JuanjoSaezF1: Pues bueno, un SORTEO para h…
…                                                   …
3604  RT @Ruff_3D: "THE MERC WITH A MOUTH AND THE EX…
3605  PlayStation 4          #PS4live (DiRT …
3606  Porsche 911 GT3 RS Race - NEED FOR SPEED HEAT …
3607  Check out my broadcast from my PlayStation 4! …
3608  RT @BORRACHITO: Muchos no lo entenderán por se…

                                                 source  truncated  \
0     <a href="http://twitter.com/#!/download/ipad" …      False
1     <a href="http://twitter.com/download/iphone" r…      False
2     <a href="https://mobile.twitter.com" rel="nofo…       True
3     <a href="http://twitter.com/download/android" …      False
4     <a href="https://mobile.twitter.com" rel="nofo…      False
…                                                   …        …
3604  <a href="https://mobile.twitter.com" rel="nofo…      False
3605  <a href="http://www.playstation.com" rel="nofo…      False
3606  <a href="http://twitter.com" rel="nofollow">Tw…      False
3607  <a href="http://www.playstation.com" rel="nofo…      False
```

```
3608  <a href="http://twitter.com/download/iphone" r…      False

      in_reply_to_status_id in_reply_to_status_id_str  in_reply_to_user_id  \
0                       NaN                      None                  NaN
1                       NaN                      None                  NaN
2                       NaN                      None         2.005574e+07
3              1.247182e+18       1247181552263614470         1.219428e+18
4                       NaN                      None                  NaN
…                     …                       …                  …
3604                    NaN                      None                  NaN
3605                    NaN                      None                  NaN
3606                    NaN                      None                  NaN
3607                    NaN                      None                  NaN
3608                    NaN                      None                  NaN

      … lang   timestamp_ms  \
0     …   es  1586208770638
1     …   fr  1586209038722
2     …   en  1586218785826
3     …   pt  1586218795715
4     …   es  1586218854123
…   …  …            …
3604  …   en  1586319893153
3605  …   ja  1586319893315
3606  …   en  1586319893466
3607  …   en  1586319893578
3608  …   es  1586319893722

                                         extended_tweet display_text_range  \
0                                                   NaN                NaN
1                                                   NaN                NaN
2     {'full_text': '@Codemasters  Please, get an up…               NaN
3                                                   NaN           [12, 84]
4                                                   NaN                NaN
…                                                 …                …
3604                                                NaN                NaN
3605                                                NaN                NaN
3606                                                NaN                NaN
3607                                                NaN                NaN
3608                                                NaN                NaN

     possibly_sensitive quoted_status_id quoted_status_id_str quoted_status  \
0                   NaN              NaN                  NaN           NaN
1                   NaN              NaN                  NaN           NaN
2                   NaN              NaN                  NaN           NaN
3                   NaN              NaN                  NaN           NaN
4                   NaN              NaN                  NaN           NaN
```

```
…                  …              …                  …           …
3604                NaN              NaN                  NaN           NaN
3605              False              NaN                  NaN           NaN
3606              False              NaN                  NaN           NaN
3607              False              NaN                  NaN           NaN
3608              False              NaN                  NaN           NaN

      quoted_status_permalink  \
0                         NaN
1                         NaN
2                         NaN
3                         NaN
4                         NaN
…                       …
3604                      NaN
3605                      NaN
3606                      NaN
3607                      NaN
3608                      NaN

                                      extended_entities
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   NaN
…                                                 …
3604                                                NaN
3605                                                NaN
3606                                                NaN
3607                                                NaN
3608  {'media': [{'id': 1247607432513126400, 'id_str…

[3609 rows x 37 columns]
```

1.4.3 Get User Account Information and Friendship Relation.

```
[ ]: from random import sample

     auth = tweepy.OAuthHandler('LwG8GPbodB36m4Bb6U5BKiggD',␣
      ↪"3EiAJXuRDci8nnFmdLAhRgIqtAbezYAVEJpN7eVbIwKpmvB2am")
     auth.set_access_token('1174542277630468096-3Wn6VBSMLOgV1Kcq0glxTBjirzQzG5',␣
      ↪'pmXdaGMos8LbWytFREMzOLNagAvNiEn38yTCWTJEBRPRx')

     client = pymongo.MongoClient('mongodb+srv://m220student:
      ↪m220password@mflix-6auna.mongodb.net/test?retryWrites=true&w=majority')
```

```
hw03_1 = client.hw03
fri = hw03_1.friends_1

class TwitterClient():
    def __init__(self, twitter_user = None):
        self.auth = auth
        self.twitter_client = tweepy.API(self.auth)
        self.twitter_user = twitter_user

        def get_user_timeline_tweets(self,num_tweets):
        tweets = []
        for tweet in tweepy.Cursor(self.twitter_client.user_timeline, id = self.
 ↪twitter_user).items(num_tweets):
            tweets.append(tweet)
        return tweets

    def get_friend_list(self,num_friends):
        friend_list = []
        for friend in tweepy.Cursor(self.twitter_client.friends, id = self.
 ↪twitter_user).items(num_friends):
            friend_list.append(friend)
        return friend_list

if __name__ == '__main__':

    screen_name = 'RockstarGames' #Using company name to start.
    record = ['RockstarGames'] #Avoid back to searched account.
    for i in range(0,20):
        homearray = []
        friendarray = []

        twitter_client = TwitterClient(screen_name)

        for i in twitter_client.get_user_timeline_tweets(10):
            home = {}
            print(i.entities['hashtags'])
            home['id'] = i.id
            home['created_at'] = i.created_at
            home['hashtags'] = i.entities['hashtags']
            home['text'] = i.text
            home['favorite_count'] = i.favorite_count
            home['retweet_count'] = i.retweet_count
            home['lang'] = i.lang
            home['source'] = i.source
            home['user'] = i.user.screen_name
            homearray.append(home)
        twi.insert_many(homearray)

        max_followers = 0
        back_screen_name = []
        for user in twitter_client.get_friend_list(10):
            mongoDB = {}
            mongoDB['screen_name']= user.screen_name
            mongoDB['name'] = user.name
            mongoDB['description'] = user.description
            mongoDB['followers_count'] = user.followers_count
            mongoDB['friends_count'] = user.friends_count
            mongoDB['verified'] = user.verified
            friendarray.append(mongoDB)

            if user.followers_count > max_followers:
                screen_name = user.screen_name
                max_followers = user.followers_count

            if user.followers_count > 10000:
                back_screen_name.append(user.screen_name)

        fri.insert_many(friendarray) #Insert data into MongoDB

        #if the account is searched before, random select account with more␣
↪than 10000 followers:
        back_up_name = str(sample(back_screen_name, 1)[0])
        screen_name = back_up_name

        record.append(screen_name)
        print(screen_name)

        print(friendarray)

    print('end')
```

1.4.4 Sample Data

Account Information and Friendship Relation

[10]:

```
hw03 = client.hw03
fri = hw03.friends_1
pd.DataFrame(list(fri.find()))
```

[10]:

```
                        _id      screen_name                 name  \
0  5e8b2f24528d22d1da03d179     IndieWorldNA          Indie World
1  5e8b2f24528d22d1da03d17a      Pocket_Camp          Pocket_camp
2  5e8b2f24528d22d1da03d17b  mariokarttourEN       Mario Kart Tour
3  5e8b2f24528d22d1da03d17c      FE_Heroes_EN    Fire Emblem Heroes
```

```
4     5e8b2f24528d22d1da03d17d  Drmarioworld_EN             Dr. Mario World
…                        …              …                         …
1042  5e8b7e58345d9e0a8ec66b38   AromaRevealBox  Aroma Reveal Self-Care Box
1043  5e8b7e58345d9e0a8ec66b39       Jess_Nneka    Makeup Scientist
1044  5e8b7e58345d9e0a8ec66b3a      Msgisselle1                 Ms.gisselle
1045  5e8b7e58345d9e0a8ec66b3b        Mirrissa3                  Mirrissa
1046  5e8b7e58345d9e0a8ec66b3c  h3h3productions                 Ethan Klein

                                            description  followers_count  \
0     The official #IndieWorld account of @NintendoA…            16415
1     This is the official channel for Animal Crossi…            76058
2     The official account for the smart phone app M…           157109
3     Official English account for the Fire Emblem H…            67212
4     The official account of the smart device app D…            14290
…                                                 …              …
1042  Subscription box of aromatherapy, skincare & b…               59
1043  N'neka is not a made up name…Jess (N-knee-ka…             2312
1044  Founder of Glamlite Cosmetics. Food Themed Mak…             1801
1045  hobby photographer\ninsta for my photography: …               82
1046  Host of @TheH3Podcast / Co-Founder of @TeddyFr…          2400486

      friends_count  verified
0                66      True
1                 4      True
2                 4      True
3                 1      True
4                 2      True
…             …       …
1042             65     False
1043            565     False
1044              4     False
1045            434     False
1046            436      True

[1047 rows x 7 columns]
```

Home Timeline Tweets of Each Account.

[11]:
```
hw03 = client.hw03
twi = hw03.twitter_1
pd.DataFrame(list(twi.find()))
```

[11]:
```
                           _id                   id           created_at  \
0     5e8b2f23528d22d1da03d16f  1246849300958556160 2020-04-05 17:16:53
1     5e8b2f23528d22d1da03d170  1246195773374492672 2020-04-03 22:00:00
2     5e8b2f23528d22d1da03d171  1246173124921622533 2020-04-03 20:30:00
3     5e8b2f23528d22d1da03d172  1246165575014830081 2020-04-03 20:00:00
```

```
4     5e8b2f23528d22d1da03d173  1246151089029709825 2020-04-03 19:02:26
…                        …                  …                 …
1005  5e8b7e59345d9e0a8ec66b42  1246955412588355584 2020-04-06 00:18:32
1006  5e8b7e59345d9e0a8ec66b43  1246955339196424193 2020-04-06 00:18:14
1007  5e8b7e59345d9e0a8ec66b44  1246955152973561856 2020-04-06 00:17:30
1008  5e8b7e59345d9e0a8ec66b45  1246953664658300929 2020-04-06 00:11:35
1009  5e8b7e59345d9e0a8ec66b46  1246953492691832832 2020-04-06 00:10:54

                                               hashtags  \
0             [{'text': 'ACNH', 'indices': [107, 112]}]
1     [{'text': 'AnimalCrossing', 'indices': [42, 57]}]
2                                                    []
3     [{'text': 'AnimalCrossing', 'indices': [58, 73]}]
4     [{'text': 'NintendoSwitchOnline', 'indices': […
…                                                 …
1005                                                 []
1006                                                 []
1007                                                 []
1008                                                 []
1009                                                 []

                                                   text  favorite_count  \
0     It's Sunday, and that means it's turnip time! …          18320
1     Kit &amp; Krysta give you a tour of their #Ani…           3158
2     Who's that riding around on Hornhead? It can't…           1380
3     Brighten up your spring with a video game gard…           4474
4     Interested in throwing your arms into the acti…           1350
…                                                 …             …
1005  @lil_h3_h3_h3 @lil_h3_h3 @lil_h3 @LilNasX @the…             16
1006  @lil_h3_h3 @lil_h3 @LilNasX @theh3podcast What…             28
1007                                  @niggamunch6969 K             117
1008              @HilaKleinH3 Theodore is a Skyrim npc            2559
1009                    @billieh3podcast @HilaKleinH3 K             194

      retweet_count lang                source            user
0              1134   en       Twitter Web App  NintendoAmerica
1               305   en  Twitter Media Studio  NintendoAmerica
2               141   en  Twitter Media Studio  NintendoAmerica
3               407   en  Twitter Media Studio  NintendoAmerica
4               204   en       Twitter Web App  NintendoAmerica
…             …  …                   …              …
1005              1   in   Twitter for Android  h3h3productions
1006              3   en   Twitter for Android  h3h3productions
1007              1  und   Twitter for Android  h3h3productions
1008             38   en   Twitter for Android  h3h3productions
1009              3  und   Twitter for Android  h3h3productions
```

```
[1010 rows x 10 columns]
```

1.5 Step Three: Answering the questions.

What are tags are associated with a person, place or thing? We could query the dataset to get the info we needed. For example: Query the hashtags of tweets of ‘NintendoAmerica’(A game company name)

[21]:

```
from bson.json_util import dumps
pipeline =twi.aggregate(
    [
        {'$match':{'user':'NintendoAmerica'}},
        {'$unwind':'$hashtags'},
        {'$group':{'_id':'$hashtags.text',
                   'total':{'$sum':1}
        }},
        {'$sort':{'total':-1}}

    ]
)
dumps(pipeline)
```

[21]:

```
'[{"_id": "AnimalCrossing", "total": 4}, {"_id": "NintendoSwitchOnline",
"total": 4}, {"_id": "PokemonSwordShield", "total": 2}, {"_id": "ACNH", "total":
2}, {"_id": "NintendoPower", "total": 2}, {"_id": "FEHeroes", "total": 2},
{"_id": "MarioKart", "total": 2}]'
```

The hashtags most presents of NintendoAmerica is “#AnimalCrossing”,“# NintendoSwitchOnline” #### What social media users are like other social media users in your domain?

[23]:

```
pipeline =fri.aggregate(
    [
        {'$group':{'_id':'$name',
                   'Freq':{'$sum':1}
        }},
        {'$sort':{'Freq':-1}},
        {'$limit': 14}
    ]
)
dumps(pipeline)
```

[23]:

```
'[{"_id": "Fortnite", "Freq": 10}, {"_id": "Fortnite Status", "Freq": 7},
{"_id": "Fortnite Competitive", "Freq": 7}, {"_id": "Bert", "Freq": 5}, {"_id":
"Ernie", "Freq": 5}, {"_id": "Abby Cadabby", "Freq": 5}, {"_id": "Cookie
Monster", "Freq": 4}, {"_id": "Grover", "Freq": 4}, {"_id": "Big Bird", "Freq":
4}, {"_id": "Elmo", "Freq": 4}, {"_id": "Fortnite Creative", "Freq": 4}, {"_id":
"Count von Count", "Freq": 4}, {"_id": "Lady Gaga", "Freq": 4}, {"_id": "Battle
Breakers", "Freq": 3}]'
```

The result shows 10 people in the dataset all followed Fortnite, which means the Fortnite is the most likely user in my domain who would be followed. #### What people, places or things are popular in your domain? The most popular things in my domain: Query the hashtags and sort by the total number of their presents:

[46]:

```
pipeline=twi.aggregate(
    [
        {'$unwind':'$hashtags'},
        {'$group':{'_id':'$hashtags.text',
                   'total':{'$sum':1}
        }},
        {'$sort':{'total':-1}},
        {'$limit': 20}
    ]
)
dumps(pipeline)
```

[46]:

```
'[{"_id": "COVID19", "total": 17}, {"_id": "FortniteChampionSeries", "total":
10}, {"_id": "SteamDeals", "total": 8}, {"_id": "WrestleMania", "total": 6},
{"_id": "ThisIsMyStreet", "total": 6}, {"_id": "StayHome", "total": 5}, {"_id":
"AnimalCrossing", "total": 5}, {"_id": "Chromatica", "total": 5}, {"_id":
"SonicWatchParty", "total": 4}, {"_id": "NintendoSwitchOnline", "total": 4},
{"_id": "AskPhilSchoen", "total": 4}, {"_id": "JimmyKimmelLiveFromHisHouse",
"total": 4}, {"_id": "KatyKeene", "total": 4}, {"_id": "coronavirus", "total":
4}, {"_id": "ACNH", "total": 3}, {"_id": "Sesame50", "total": 3}, {"_id":
"AmericasFoodFund", "total": 3}, {"_id": "Wrestlemania36", "total": 2}, {"_id":
"ThirstyThursday", "total": 2}, {"_id": "CaringForEachOther", "total": 2}]'
```

The popular hashtags includes: COVID19, Fortnite, ThisIsMyStreet, StayHome, AnimalCrossing, SonicWatchParty, ect. #### What people, places or things are trending in your domain? The most trending things in my domain: Query the hashtags of stream tweets and sort by the total number of their presents:

[27]:

```
pipeline=str.aggregate(
    [
        {'$unwind':'$entities.hashtags'},
        {'$group':{'_id':'$entities.hashtags.text',                    'total':
 ↪{'$sum':1}
        }},
        {'$sort':{'total':-1}},
        {'$limit':30}
    ]
)
dumps(pipeline)
```

[27]:
```
'[{"_id": "PS4live", "total": 209}, {"_id": "Fortnite", "total": 144}, {"_id":
"LHHATL", "total": 58}, {"_id": "fortnite", "total": 44}, {"_id":
"\\u30d5\\u30a9\\u30fc\\u30c8\\u30ca\\u30a4\\u30c8", "total": 42}, {"_id":
"PS4", "total": 35}, {"_id": "FF7R", "total": 34}, {"_id": "NowPlaying",
"total": 22}, {"_id": "twitch", "total": 15}, {"_id": "XboxShare", "total": 14},
{"_id": "ApexLegends", "total": 12}, {"_id": "FortniteBattleRoyale", "total":
12}, {"_id": "F1", "total": 11}, {"_id":
"\\u3081\\u3056\\u307e\\u3057\\u30d7\\u30ec\\u30bc\\u30f3\\u30c8", "total": 11},
{"_id": "\\u0627\\u0644\\u0639\\u0628_\\u0648\\u0627\\u0646\\u062a_\\u0641\\u064
a_\\u0628\\u064a\\u062a\\u0643\\u0645", "total": 10}, {"_id":
"TomandanteBorolas", "total": 10}, {"_id": "clutchwin", "total": 10}, {"_id":
"LigaBBVAMX", "total": 10}, {"_id": "xboxclips", "total": 9}, {"_id": "Apex",
"total": 9}, {"_id": "streamer", "total": 8}, {"_id": "COVID19", "total": 8},
{"_id": "FortniteArt", "total": 8}, {"_id": "PS4share", "total": 8}, {"_id":
"\\u8056\\u52633ToM", "total": 7}, {"_id": "FF7\\u30ea\\u30e1\\u30a4\\u30af",
"total": 7}, {"_id": "gamer", "total": 7}, {"_id": "FortniteChapter2", "total":
7}, {"_id": "ad", "total": 7}, {"_id": "trapLAXradio", "total": 7}]'
```

The trending hashtags includes: PS4live, Fortnite, NowPlaying, FF7R, Twitch, ApexLegends,F1,PS4Share, ect.

The top 3 with highest number of favorite tweets is:

[42]:
```
pipeline=str.aggregate(
    [
        {'$sort':{'favorite_count':-1}},
        {'$limit':10},
        {'$project':{'_id':0,'user.name':1,'text':1}},
        {'$limit':3}
    ]
)
dumps(pipeline)
```

[42]:
```
'[{"text": "RT @JuanjoSaezF1: Pues bueno, un SORTEO para haceros m\\u00e1s
amenos estos d\\u00edas de confinamiento. CLAVE de F1 2019 para PC.
#F1\\n\\nSolo dos requis\\u2026", "user": {"name": "Pablo Vilafa\\u00f1e"}},
{"text": "RT @InsideDlpCastle: Bonjour \\u00e0 toutes et \\u00e0 tous!\\n\\nLes
06/07 Avril 2019, @DisneylandParis accueillait l\'\\u00e9curie @RenaultSportFR
et le champio\\u2026", "user": {"name": "Jace"}}, {"text": "@Codemasters
Please, get an update to F1 2019 with the cars and circuits of 2020 since it
does not look good that\\u2026 https://t.co/NIXLtAFTOP", "user": {"name":
"David"}}]'
```

1.6 Report

File used: companylist.csv Code used: MongoDB.py,stream.py, giantboob.py The data of video game information is gathered from giantbomb.com API, the social media data is collected from twitter. Next, insert data into NoSQL. Finally, using NoSQL query to extract the useful insight of the data, answering the questions.

1.7 Conclusion

Primary focus of this assignment is to learn how to get social media data from twitter, load the data into the NoSQL database, using aggregate method to query the useful data.

1.8 Contribution

We contributed by ourselves: 60% By External source: 40%

1.9 Citations

https://github.com/vprusso/youtube_tutorials/blob/master/twitter_python/part_1_streaming_tweets/tweepy_
https://www.youtube.com/watch?v=wlnx-7cm4Gg http://docs.tweepy.org/en/latest/api.html#timeline-methods https://www.youtube.com/watch?v=PrBSuxXURYs

```
[ ]:
```