CIS 593 – BIG DATA

LAB ASSIGNMENT – 3

Name: Pooja A. Khatri

CSU ID: 2783752

Extra credit work:

automatic process for data collection, and storage and retrieval using mongoDB

1. Platforms

Language used: Python

Python IDE: PyCharm

Database design tool: mongoDB

2. Files included

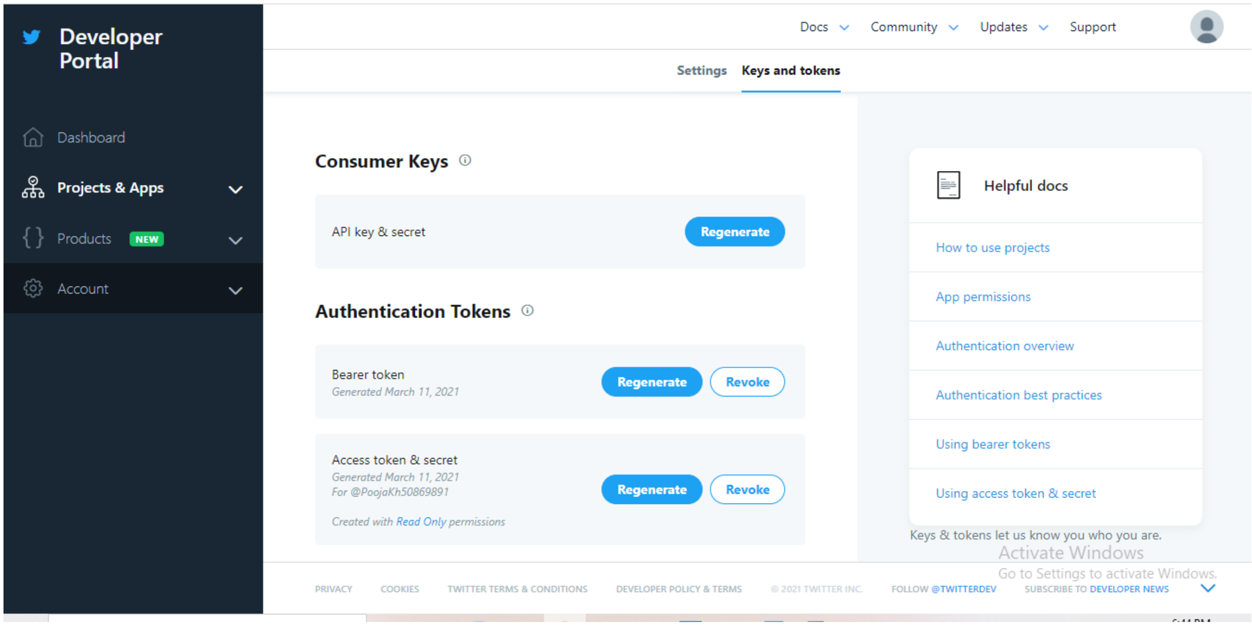
Main.py, QueryResults.xlsx showing query results for all queries in different sheets

3. Steps performed for lab

- Sign-In to Twitter developer account and generate API tokens
- Fetch JSON from tweeter using RestAPI
- Parse all data in python and make mongoDB connection
- Import all data into mongoDB Database
- Retrieve data from mongoDB dynamically

4. Screenshots

Twitter developer account obtained successfully



Consumer Keys

API key & secret

Authentication Tokens

Bearer token
Generated March 11, 2021

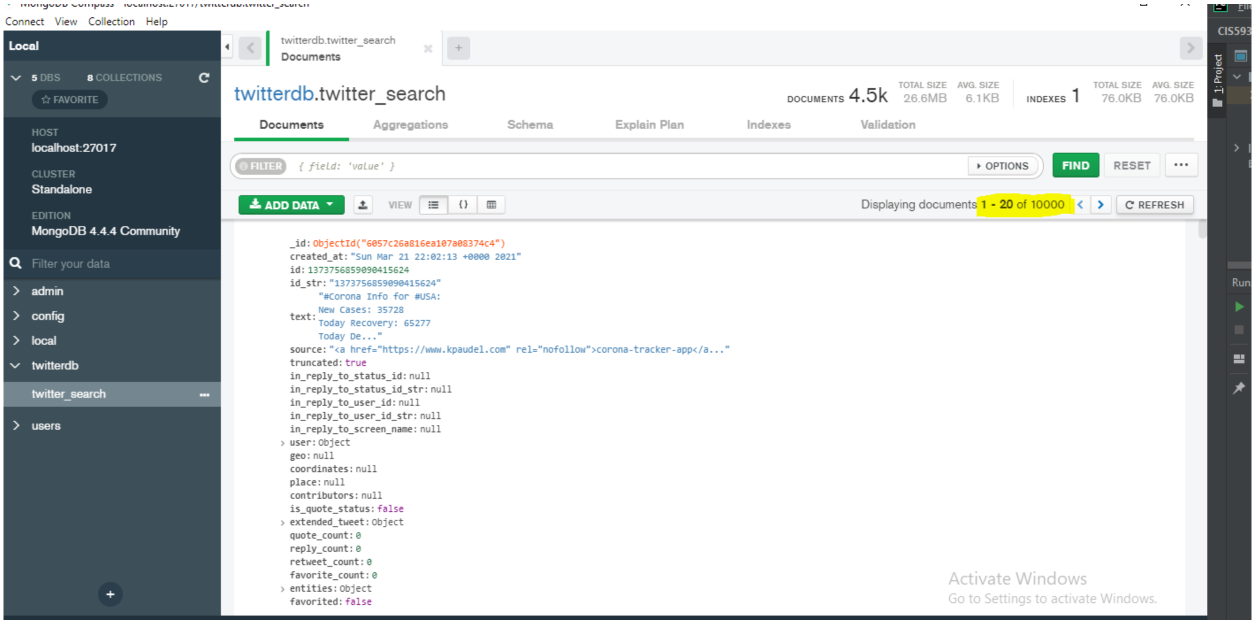
Access token & secret
Generated March 11, 2021
For @PoojaKh50869891

Created with Read Only permissions

Keys & tokens let us know you who you are.

Activate Windows
Go to Settings to activate Windows.

mongoDB showing 10000 documents



HOST
localhost:27017

CLUSTER
Standalone

EDITION
MongoDB 4.4.4 Community

twitterdb.twitter_search

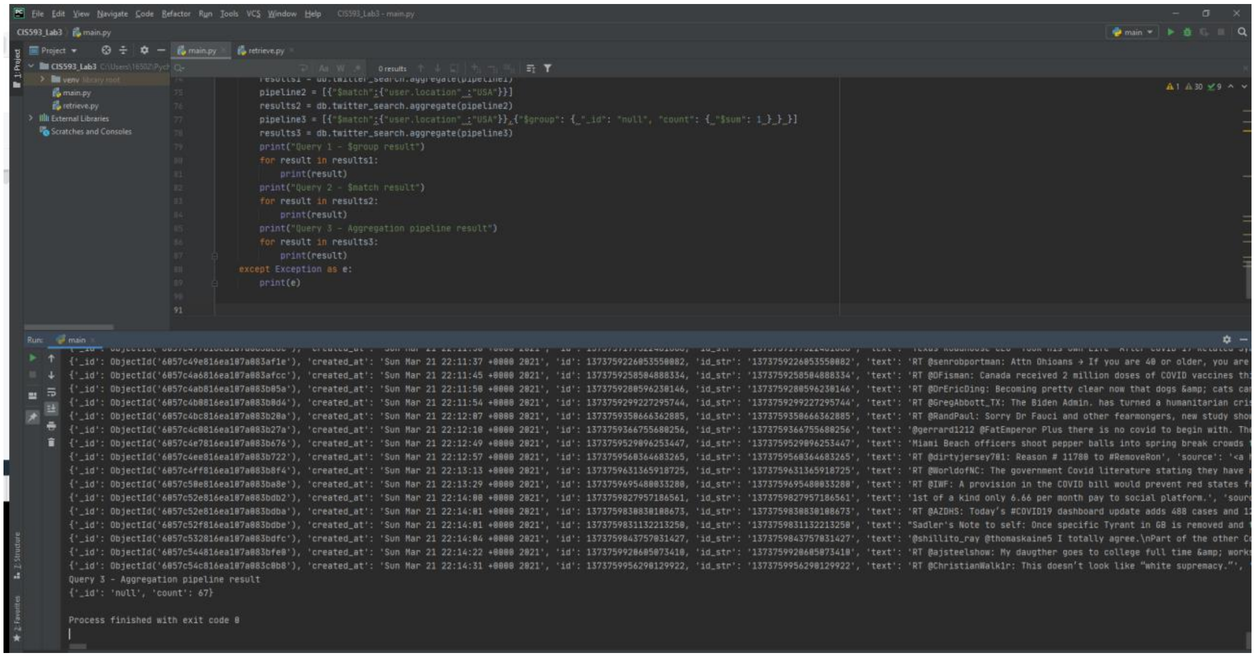
DOCUMENTS 4.5k TOTAL SIZE 26.6MB AVG. SIZE 6.1KB INDEXES 1 TOTAL SIZE 76.0KB AVG. SIZE 76.0KB

{ field: 'value' }

Displaying documents 1 - 20 of 10000

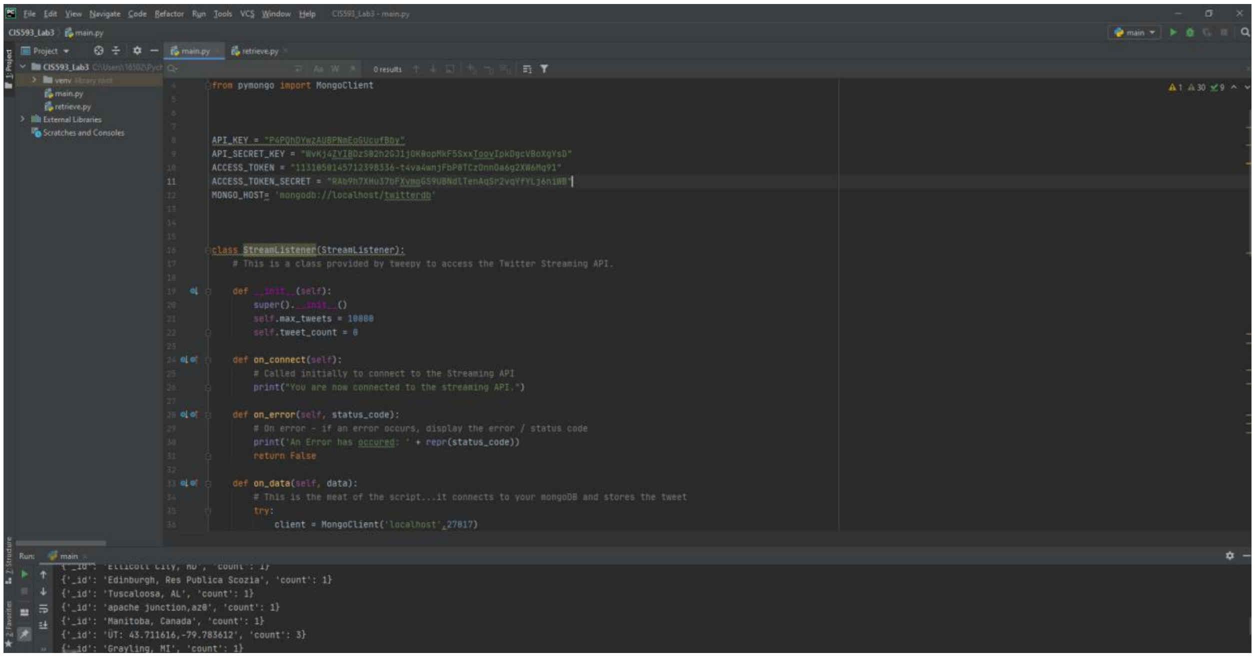
```
_id: ObjectId("6057c26a816ea107a08374c4")
created_at: "Sun Mar 21 22:02:13 +0000 2021"
id: 1373756859090415624
id_str: "1373756859090415624"
      "#Corona Info for #USA:
text: New Cases: 35728
      Today Recovery: 65277
      Today De..."
source: "<a href="https://www.kpaudel.com" rel="nofollow">corona-tracker-app</a..."
truncated: true
in_reply_to_status_id: null
in_reply_to_status_id_str: null
in_reply_to_user_id: null
in_reply_to_user_id_str: null
in_reply_to_screen_name: null
> user: Object
geo: null
coordinates: null
place: null
contributors: null
is_quote_status: false
> extended_tweet: Object
quote_count: 0
reply_count: 0
retweet_count: 0
favorite_count: 0
> entities: Object
favorited: false
```

Pycharm showing successful execution



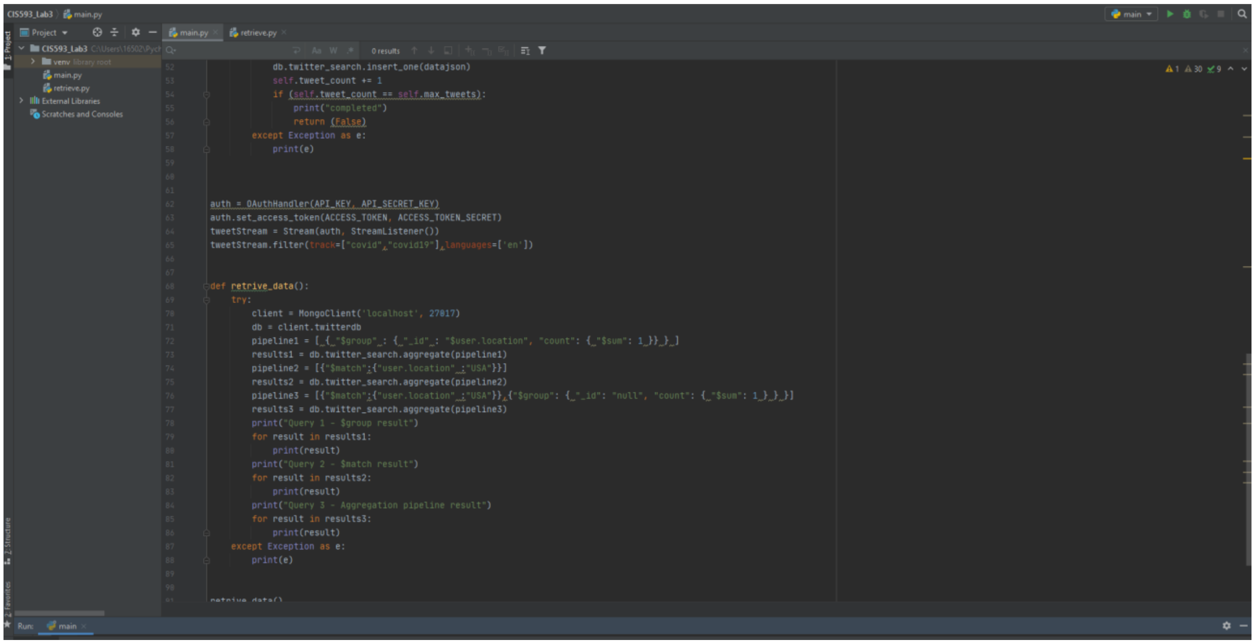
```
        pipeline2 = [{"$match":{"user.location" :"USA"}}]
        results2 = db.twitter_search.aggregate(pipeline2)
        pipeline3 = [{"$match":{"user.location" :"USA"}},{"$group": {"_id": "null", "count": {"$sum": 1}}}]
        results3 = db.twitter_search.aggregate(pipeline3)
        print("Query 1 - $group result")
        for result in results1:
            print(result)
        print("Query 2 - $match result")
        for result in results2:
            print(result)
        print("Query 3 - Aggregation pipeline result")
        for result in results3:
            print(result)
    except Exception as e:
        print(e)
```

```
{'_id': ObjectId('6057c49e816ea107a083af1e'), 'created_at': 'Sun Mar 21 22:11:37 +0000 2021', 'id': 1373759226053550082, 'id_str': '1373759226053550082', 'text': 'RT @senrobportman: Attn Ohioans → If you are 40 or older, you are
{'_id': ObjectId('6057c4a6816ea107a083afcc'), 'created_at': 'Sun Mar 21 22:11:45 +0000 2021', 'id': 1373759258504888334, 'id_str': '1373759258504888334', 'text': 'RT @DFisman: Canada received 2 million doses of COVID vaccines thi
{'_id': ObjectId('6057c4ab816ea107a083b05a'), 'created_at': 'Sun Mar 21 22:11:50 +0000 2021', 'id': 1373759280596230146, 'id_str': '1373759280596230146', 'text': 'RT @DrEricDing: Becoming pretty clear now that dogs &amp; cats car
{'_id': ObjectId('6057c4b0816ea107a083b0d4'), 'created_at': 'Sun Mar 21 22:11:54 +0000 2021', 'id': 1373759299227295744, 'id_str': '1373759299227295744', 'text': 'RT @GregAbbott_TX: The Biden Admin. has turned a humanitarian cris
{'_id': ObjectId('6057c4bc816ea107a083b20a'), 'created_at': 'Sun Mar 21 22:12:07 +0000 2021', 'id': 1373759350666362885, 'id_str': '1373759350666362885', 'text': 'RT @RandPaul: Sorry Dr Fauci and other fearmongers, new study show
{'_id': ObjectId('6057c4c0816ea107a083b27a'), 'created_at': 'Sun Mar 21 22:12:10 +0000 2021', 'id': 1373759366755680256, 'id_str': '1373759366755680256', 'text': '@gerrard1212 @FatEmperor Plus there is no covid to begin with. The
{'_id': ObjectId('6057c4e7816ea107a083b676'), 'created_at': 'Sun Mar 21 22:12:49 +0000 2021', 'id': 1373759529096253447, 'id_str': '1373759529096253447', 'text': 'Miami Beach officers shoot pepper balls into spring break crowds t
{'_id': ObjectId('6057c4ee816ea107a083b722'), 'created_at': 'Sun Mar 21 22:12:57 +0000 2021', 'id': 1373759560364683265, 'id_str': '1373759560364683265', 'text': 'RT @dirtyjersey701: Reason # 11780 to #RemoveRon', 'source': '<a h
{'_id': ObjectId('6057c4ff816ea107a083b8f4'), 'created_at': 'Sun Mar 21 22:13:13 +0000 2021', 'id': 1373759631365918725, 'id_str': '1373759631365918725', 'text': 'RT @WorldofNC: The government Covid literature stating they have m
{'_id': ObjectId('6057c50e816ea107a083ba8e'), 'created_at': 'Sun Mar 21 22:13:29 +0000 2021', 'id': 1373759695480033280, 'id_str': '1373759695480033280', 'text': 'RT @IWF: A provision in the COVID bill would prevent red states fr
{'_id': ObjectId('6057c52e816ea107a083bdb2'), 'created_at': 'Sun Mar 21 22:14:00 +0000 2021', 'id': 1373759827957186561, 'id_str': '1373759827957186561', 'text': '1st of a kind only 6.66 per month pay to social platform.', 'sourc
{'_id': ObjectId('6057c52e816ea107a083bdba'), 'created_at': 'Sun Mar 21 22:14:01 +0000 2021', 'id': 1373759830830108673, 'id_str': '1373759830830108673', 'text': 'RT @AZDHS: Today's #COVID19 dashboard update adds 488 cases and 12
{'_id': ObjectId('6057c52f816ea107a083bdbe'), 'created_at': 'Sun Mar 21 22:14:01 +0000 2021', 'id': 1373759831132213250, 'id_str': '1373759831132213250', 'text': "Sadler's Note to self: Once specific Tyrant in GB is removed and t
{'_id': ObjectId('6057c532816ea107a083bdfc'), 'created_at': 'Sun Mar 21 22:14:04 +0000 2021', 'id': 1373759843757031427, 'id_str': '1373759843757031427', 'text': '@shillito_ray @thomaskaine5 I totally agree.\nPart of the other Co
{'_id': ObjectId('6057c544816ea107a083bfe0'), 'created_at': 'Sun Mar 21 22:14:22 +0000 2021', 'id': 1373759920605073410, 'id_str': '1373759920605073410', 'text': 'RT @ajsteelshow: My daugther goes to college full time &amp; works
{'_id': ObjectId('6057c54c816ea107a083c0b8'), 'created_at': 'Sun Mar 21 22:14:31 +0000 2021', 'id': 1373759956290129922, 'id_str': '1373759956290129922', 'text': 'RT @ChristianWalk1r: This doesn't look like "white supremacy."',
Query 3 - Aggregation pipeline result
{'_id': 'null', 'count': 67}

Process finished with exit code 0
```



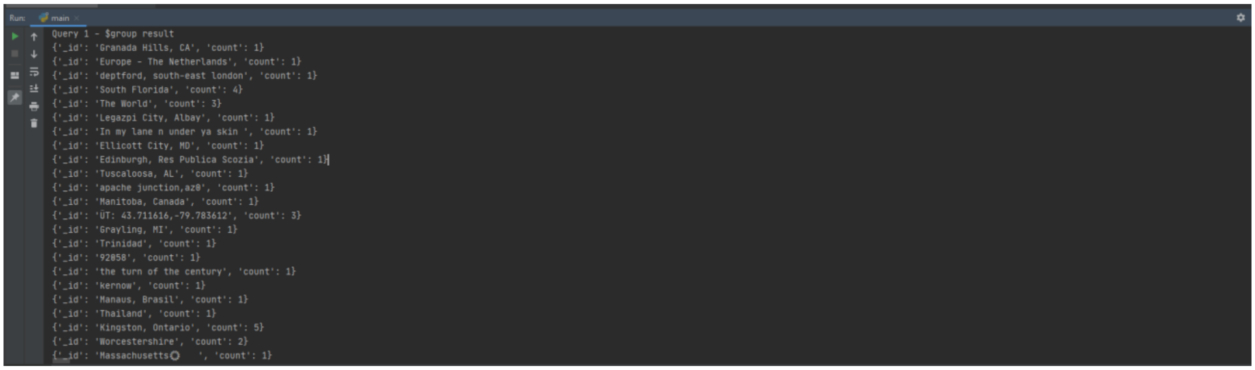
```
from pymongo import MongoClient



API_KEY = "P4PQhDYwzAUBPNmEoGUcufBby"
API_SECRET_KEY = "WvKj4ZYIBDzS02h2GJ1jOK8opMkF5SxxTooyIpkDgcVBoXgYsD"
ACCESS_TOKEN = "1131050145712398336-t4va4wnjFbP8TCzOnn0a6g2XW6Mq91"
ACCESS_TOKEN_SECRET = "RAb9h7XHu37bFXvmoGS9UBNdlTenAqSr2vqYfYLj6n1WB"
MONGO_HOST= 'mongodb://localhost/twitterdb'



class StreamListener(StreamListener):
    # This is a class provided by tweepy to access the Twitter Streaming API.

    def __init__(self):
        super().__init__()
        self.max_tweets = 10000
        self.tweet_count = 0

    def on_connect(self):
        # Called initially to connect to the Streaming API
        print("You are now connected to the streaming API.")

    def on_error(self, status_code):
        # On error - if an error occurs, display the error / status code
        print('An Error has occured: ' + repr(status_code))
        return False

    def on_data(self, data):
        # This is the meat of the script...it connects to your mongoDB and stores the tweet
        try:
            client = MongoClient('localhost',27017)
```

```
{'_id': 'Ellicott City, MD', 'count': 1}
{'_id': 'Edinburgh, Res Publica Scozia', 'count': 1}
{'_id': 'Tuscaloosa, AL', 'count': 1}
{'_id': 'apache junction,az0', 'count': 1}
{'_id': 'Manitoba, Canada', 'count': 1}
{'_id': 'UT: 43.711616,-79.783612', 'count': 3}
{'_id': 'Grayling, MI', 'count': 1}
```



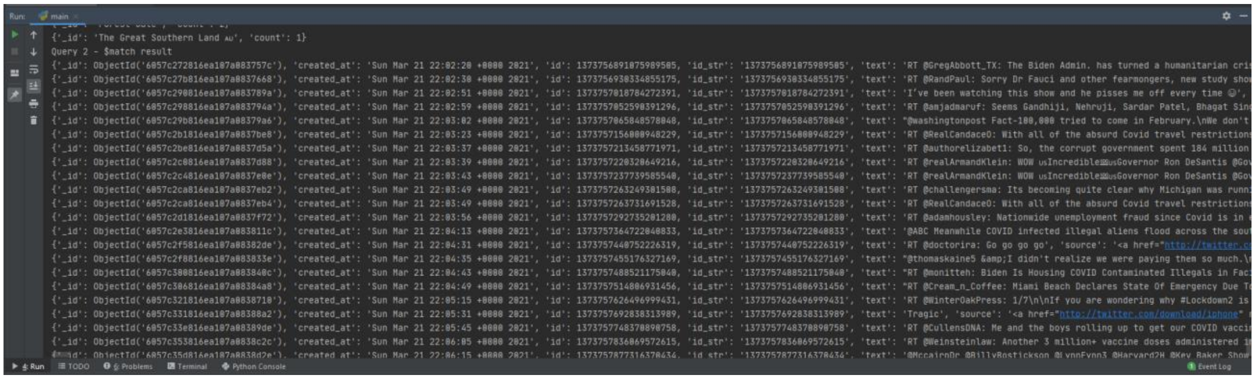
```
                db.twitter_search.insert_one(datajson)
                self.tweet_count += 1
                if (self.tweet_count == self.max_tweets):
                    print("completed")
                    return (False)
            except Exception as e:
                print(e)



auth = OAuthHandler(API_KEY, API_SECRET_KEY)
auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)
tweetStream = Stream(auth, StreamListener())
tweetStream.filter(track=["covid","covid19"],languages=['en'])


def retrive_data():
    try:
        client = MongoClient('localhost', 27017)
        db = client.twitterdb
        pipeline1 = [ { "$group" : { "_id" : "$user.location", "count": { "$sum": 1 }} } ]
        results1 = db.twitter_search.aggregate(pipeline1)
        pipeline2 = [{"$match":{"user.location" :"USA"}}]
        results2 = db.twitter_search.aggregate(pipeline2)
        pipeline3 = [{"$match":{"user.location" :"USA"}},{"$group": { "_id": "null", "count": { "$sum": 1 } } }]
        results3 = db.twitter_search.aggregate(pipeline3)
        print("Query 1 - $group result")
        for result in results1:
            print(result)
        print("Query 2 - $match result")
        for result in results2:
            print(result)
        print("Query 3 - Aggregation pipeline result")
        for result in results3:
            print(result)
    except Exception as e:
        print(e)


retrive_data()
```

Pycharm showing execution result of $group query - [{ "$group" : { "_id" : "$user.location", "count": { "$sum": 1 }} }]



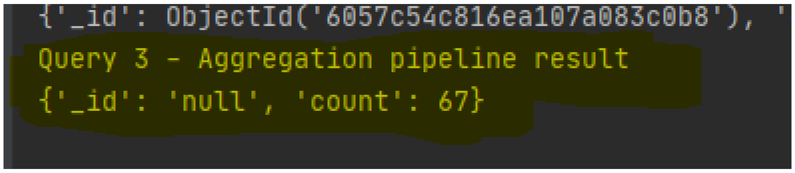
```
Query 1 - $group result
{'_id': 'Granada Hills, CA', 'count': 1}
{'_id': 'Europe - The Netherlands', 'count': 1}
{'_id': 'deptford, south-east london', 'count': 1}
{'_id': 'South Florida', 'count': 4}
{'_id': 'The World', 'count': 3}
{'_id': 'Legazpi City, Albay', 'count': 1}
{'_id': 'In my lane n under ya skin ', 'count': 1}
{'_id': 'Ellicott City, MD', 'count': 1}
{'_id': 'Edinburgh, Res Publica Scozia', 'count': 1}
{'_id': 'Tuscaloosa, AL', 'count': 1}
{'_id': 'apache junction,az0', 'count': 1}
{'_id': 'Manitoba, Canada', 'count': 1}
{'_id': 'ÜT: 43.711616,-79.783612', 'count': 3}
{'_id': 'Grayling, MI', 'count': 1}
{'_id': 'Trinidad', 'count': 1}
{'_id': '92058', 'count': 1}
{'_id': 'the turn of the century', 'count': 1}
{'_id': 'kernow', 'count': 1}
{'_id': 'Manaus, Brasil', 'count': 1}
{'_id': 'Thailand', 'count': 1}
{'_id': 'Kingston, Ontario', 'count': 5}
{'_id': 'Worcestershire', 'count': 2}
{'_id': 'Massachusetts○    ', 'count': 1}
```

Pycharm showing execution result of $match query – [{"$match":{"user.location" :"USA"}}]



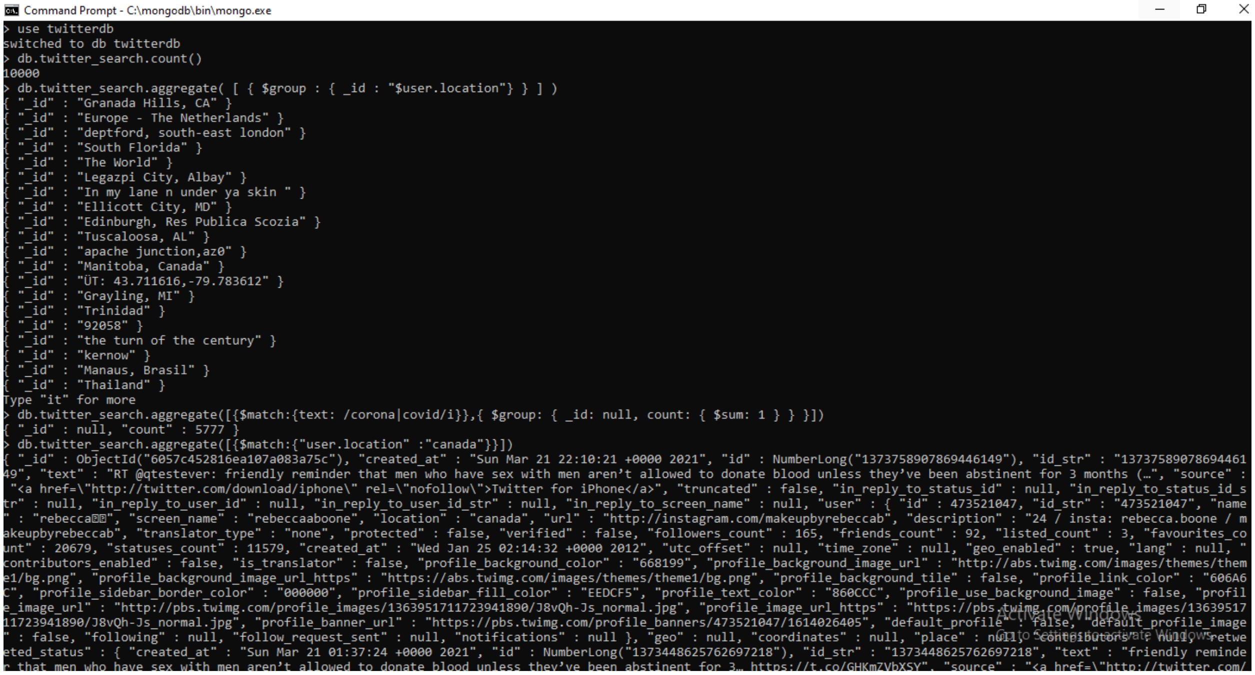
```
{'_id': 'The Great Southern Land AU', 'count': 1}
Query 2 - $match result
{'_id': ObjectId('6057c272816ea107a083757c'), 'created_at': 'Sun Mar 21 22:02:20 +0000 2021', 'id': 1373756891075989505, 'id_str': '1373756891075989505', 'text': 'RT @GregAbbott_TX: The Biden Admin. has turned a humanitarian cris
{'_id': ObjectId('6057c27b816ea107a0837668'), 'created_at': 'Sun Mar 21 22:02:30 +0000 2021', 'id': 1373756930334855175, 'id_str': '1373756930334855175', 'text': 'RT @RandPaul: Sorry Dr Fauci and other fearmongers, new study show
{'_id': ObjectId('6057c290816ea107a083789a'), 'created_at': 'Sun Mar 21 22:02:51 +0000 2021', 'id': 1373757018784272391, 'id_str': '1373757018784272391', 'text': 'I've been watching this show and he pisses me off every time 😡',
{'_id': ObjectId('6057c298816ea107a083794a'), 'created_at': 'Sun Mar 21 22:02:59 +0000 2021', 'id': 1373757052590391296, 'id_str': '1373757052590391296', 'text': 'RT @amjadmaruf: Seems Gandhiji, Nehruji, Sardar Patel, Bhagat Sing
{'_id': ObjectId('6057c29b816ea107a08379a6'), 'created_at': 'Sun Mar 21 22:03:02 +0000 2021', 'id': 1373757065848578048, 'id_str': '1373757065848578048', 'text': "@washingtonpost Fact-100,000 tried to come in February.\nWe don't
{'_id': ObjectId('6057c2b1816ea107a0837be8'), 'created_at': 'Sun Mar 21 22:03:23 +0000 2021', 'id': 1373757156808948229, 'id_str': '1373757156808948229', 'text': 'RT @RealCandaceO: With all of the absurd Covid travel restrictions
{'_id': ObjectId('6057c2be816ea107a0837d5a'), 'created_at': 'Sun Mar 21 22:03:37 +0000 2021', 'id': 1373757213458771971, 'id_str': '1373757213458771971', 'text': 'RT @authorelizabet1: So, the corrupt government spent 184 million
{'_id': ObjectId('6057c2c0816ea107a0837d88'), 'created_at': 'Sun Mar 21 22:03:39 +0000 2021', 'id': 1373757220320649216, 'id_str': '1373757220320649216', 'text': 'RT @realArmandKlein: WOW usIncredible▦usGovernor Ron DeSantis @Gov
{'_id': ObjectId('6057c2c4816ea107a0837e0e'), 'created_at': 'Sun Mar 21 22:03:43 +0000 2021', 'id': 1373757237739585540, 'id_str': '1373757237739585540', 'text': 'RT @realArmandKlein: WOW usIncredible▦usGovernor Ron DeSantis @Gov
{'_id': ObjectId('6057c2ca816ea107a0837eb2'), 'created_at': 'Sun Mar 21 22:03:49 +0000 2021', 'id': 1373757263249301508, 'id_str': '1373757263249301508', 'text': 'RT @challengersma: Its becoming quite clear why Michigan was runni
{'_id': ObjectId('6057c2ca816ea107a0837eb4'), 'created_at': 'Sun Mar 21 22:03:49 +0000 2021', 'id': 1373757263731691528, 'id_str': '1373757263731691528', 'text': 'RT @RealCandaceO: With all of the absurd Covid travel restrictions
{'_id': ObjectId('6057c2d1816ea107a0837f72'), 'created_at': 'Sun Mar 21 22:03:56 +0000 2021', 'id': 1373757292735201280, 'id_str': '1373757292735201280', 'text': 'RT @adamhousley: Nationwide unemployment fraud since Covid is in e
{'_id': ObjectId('6057c2e3816ea107a083811c'), 'created_at': 'Sun Mar 21 22:04:13 +0000 2021', 'id': 1373757364722040833, 'id_str': '1373757364722040833', 'text': '@ABC Meanwhile COVID infected illegal aliens flood across the sout
{'_id': ObjectId('6057c2f5816ea107a08382de'), 'created_at': 'Sun Mar 21 22:04:31 +0000 2021', 'id': 1373757440752226319, 'id_str': '1373757440752226319', 'text': 'RT @doctorira: Go go go go', 'source': '<a href="http://twitter.c
{'_id': ObjectId('6057c2f8816ea107a083833e'), 'created_at': 'Sun Mar 21 22:04:35 +0000 2021', 'id': 1373757455176327169, 'id_str': '1373757455176327169', 'text': "@thomaskaine5 &amp;I didn't realize we were paying them so much.\n
{'_id': ObjectId('6057c300816ea107a083840c'), 'created_at': 'Sun Mar 21 22:04:43 +0000 2021', 'id': 1373757488521175040, 'id_str': '1373757488521175040', 'text': "RT @monitteh: Biden Is Housing COVID Contaminated Illegals in Faci
{'_id': ObjectId('6057c306816ea107a08384a8'), 'created_at': 'Sun Mar 21 22:04:49 +0000 2021', 'id': 1373757514806931456, 'id_str': '1373757514806931456', 'text': "RT @Cream_n_Coffee: Miami Beach Declares State Of Emergency Due To
{'_id': ObjectId('6057c321816ea107a0838710'), 'created_at': 'Sun Mar 21 22:05:15 +0000 2021', 'id': 1373757626496999431, 'id_str': '1373757626496999431', 'text': 'RT @WinterOakPress: 1/7\n\nIf you are wondering why #Lockdown2 is
{'_id': ObjectId('6057c331816ea107a08388a2'), 'created_at': 'Sun Mar 21 22:05:31 +0000 2021', 'id': 1373757692838313989, 'id_str': '1373757692838313989', 'text': 'Tragic', 'source': '<a href="http://twitter.com/download/iphone" r
{'_id': ObjectId('6057c33e816ea107a08389de'), 'created_at': 'Sun Mar 21 22:05:45 +0000 2021', 'id': 1373757748370890758, 'id_str': '1373757748370890758', 'text': 'RT @CullensDNA: Me and the boys rolling up to get our COVID vaccin
{'_id': ObjectId('6057c353816ea107a0838c2c'), 'created_at': 'Sun Mar 21 22:06:05 +0000 2021', 'id': 1373757836069572615, 'id_str': '1373757836069572615', 'text': 'RT @Weinsteinlaw: Another 3 million+ vaccine doses administered in
{'_id': ObjectId('6057c35d816ea107a0838d2e'), 'created_at': 'Sun Mar 21 22:06:15 +0000 2021', 'id': 1373757877316370434, 'id_str': '1373757877316370434', 'text': '@MccairnDr @BillyBostickson @LynnFynn3 @Harvard2H @Kev_Baker Show
```

Pycharm showing execution result of Aggregation pipeline query - [{"$match":{"user.location" :"USA"}},{"$group": { "_id": "null", "count": { "$sum": 1 } } }]



```
{'_id': ObjectId('6057c54c816ea107a083c0b8'), '
Query 3 - Aggregation pipeline result
{'_id': 'null', 'count': 67}
```

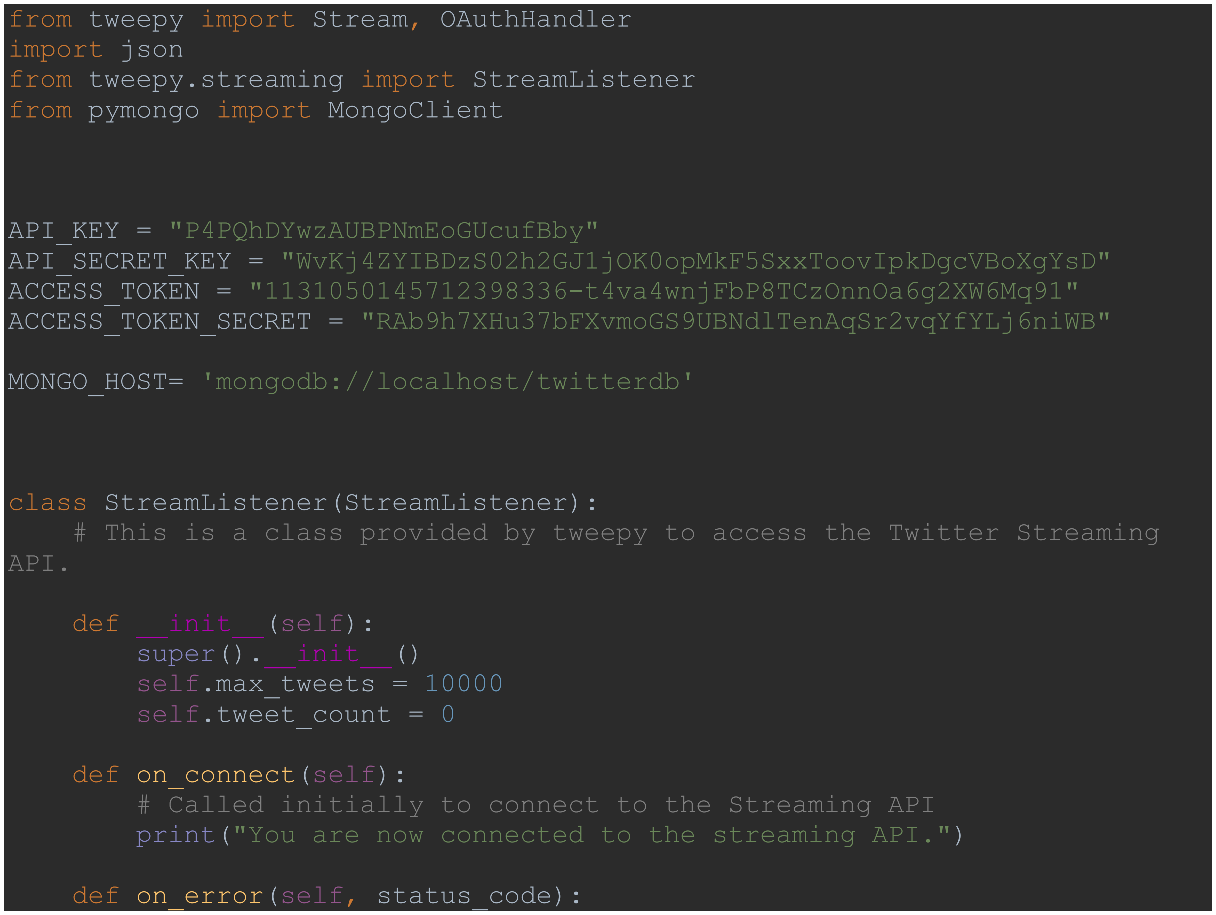
mongoDB Shell showing few query results



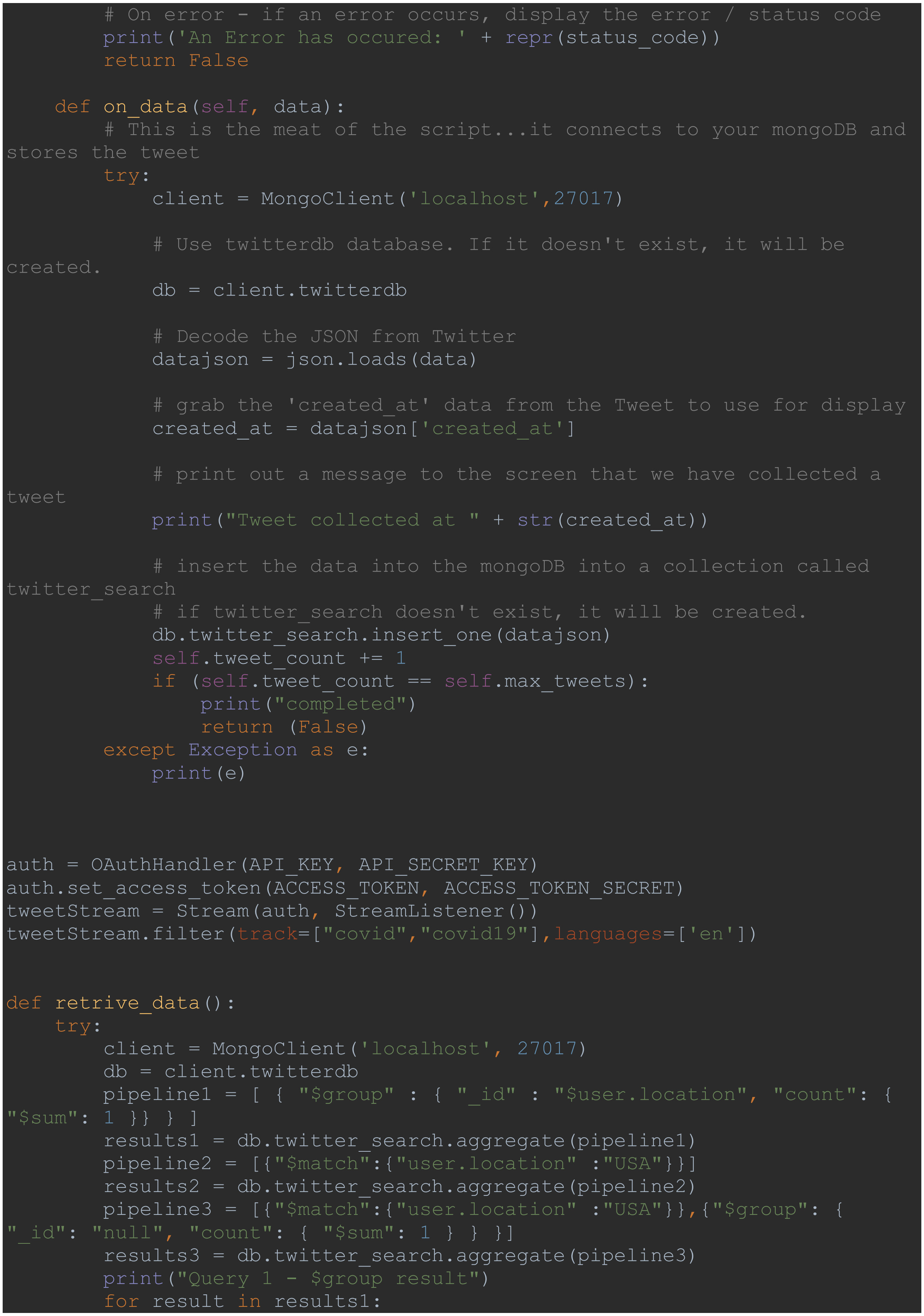
```
> use twitterdb
switched to db twitterdb
> db.twitter_search.count()
10000
> db.twitter_search.aggregate( [ { $group : { _id : "$user.location"} } ] )
{ "_id" : "Granada Hills, CA" }
{ "_id" : "Europe - The Netherlands" }
{ "_id" : "deptford, south-east london" }
{ "_id" : "South Florida" }
{ "_id" : "The World" }
{ "_id" : "Legazpi City, Albay" }
{ "_id" : "In my lane n under ya skin " }
{ "_id" : "Ellicott City, MD" }
{ "_id" : "Edinburgh, Res Publica Scozia" }
{ "_id" : "Tuscaloosa, AL" }
{ "_id" : "apache junction,az0" }
{ "_id" : "Manitoba, Canada" }
{ "_id" : "ÜT: 43.711616,-79.783612" }
{ "_id" : "Grayling, MI" }
{ "_id" : "Trinidad" }
{ "_id" : "92058" }
{ "_id" : "the turn of the century" }
{ "_id" : "kernow" }
{ "_id" : "Manaus, Brasil" }
{ "_id" : "Thailand" }
Type "it" for more
> db.twitter_search.aggregate([{$match:{text: /corona|covid/i}},{ $group: { _id: null, count: { $sum: 1 } } }])
{ "_id" : null, "count" : 5777 }
> db.twitter_search.aggregate([{$match:{"user.location" :"canada"}}])
{ "_id" : ObjectId("6057c452816ea107a083a75c"), "created_at" : "Sun Mar 21 22:10:21 +0000 2021", "id" : NumberLong("1373758907869446149"), "id_str" : "13737589078694461
49", "text" : "RT @qtestever: friendly reminder that men who have sex with men aren’t allowed to donate blood unless they’ve been abstinent for 3 months (…", "source" :
 "<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\">Twitter for iPhone</a>", "truncated" : false, "in_reply_to_status_id" : null, "in_reply_to_status_id_s
tr" : null, "in_reply_to_user_id" : null, "in_reply_to_user_id_str" : null, "in_reply_to_screen_name" : null, "user" : { "id" : 473521047, "id_str" : "473521047", "name
" : "rebecca��", "screen_name" : "rebeccaaboone", "location" : "canada", "url" : "http://instagram.com/makeupbyrebeccab", "description" : "24 / insta: rebecca.boone / m
akeupbyrebeccab", "translator_type" : "none", "protected" : false, "verified" : false, "followers_count" : 165, "friends_count" : 92, "listed_count" : 3, "favourites_co
unt" : 20679, "statuses_count" : 11579, "created_at" : "Wed Jan 25 02:14:32 +0000 2012", "utc_offset" : null, "time_zone" : null, "geo_enabled" : true, "lang" : null, "
contributors_enabled" : false, "is_translator" : false, "profile_background_color" : "668199", "profile_background_image_url" : "http://abs.twimg.com/images/themes/them
e1/bg.png", "profile_background_image_url_https" : "https://abs.twimg.com/images/themes/theme1/bg.png", "profile_background_tile" : false, "profile_link_color" : "606A6
C", "profile_sidebar_border_color" : "000000", "profile_sidebar_fill_color" : "EEDCF5", "profile_text_color" : "860CCC", "profile_use_background_image" : false, "profil
e_image_url" : "http://pbs.twimg.com/profile_images/1363951711723941890/J8vQh-Js_normal.jpg", "profile_image_url_https" : "https://pbs.twimg.com/profile_images/13639517
11723941890/J8vQh-Js_normal.jpg", "profile_banner_url" : "https://pbs.twimg.com/profile_banners/473521047/1614026405", "default_profile" : false, "default_profile_image
" : false, "following" : null, "follow_request_sent" : null, "notifications" : null }, "geo" : null, "coordinates" : null, "place" : null, "contributors" : null, "retwe
eted_status" : { "created_at" : "Sun Mar 21 01:37:24 +0000 2021", "id" : NumberLong("1373448625762697218"), "id_str" : "1373448625762697218", "text" : "friendly reminde
r that men who have sex with men aren’t allowed to donate blood unless they’ve been abstinent for 3… https://t.co/GHKmZVbXSY", "source" : "<a href=\"http://twitter.com/
```

Activate Windows
Go to Settings to activate Windows.

5. Code



```
from tweepy import Stream, OAuthHandler
import json
from tweepy.streaming import StreamListener
from pymongo import MongoClient



API_KEY = "P4PQhDYwzAUBPNmEoGUcufBby"
API_SECRET_KEY = "WvKj4ZYIBDzS02h2GJ1jOK0opMkF5SxxToovIpkDgcVBoXgYsD"
ACCESS_TOKEN = "1131050145712398336-t4va4wnjFbP8TCzOnnOa6g2XW6Mq91"
ACCESS_TOKEN_SECRET = "RAb9h7XHu37bFXvmoGS9UBNdlTenAqSr2vqYfYLj6niWB"

MONGO_HOST= 'mongodb://localhost/twitterdb'



class StreamListener(StreamListener):
    # This is a class provided by tweepy to access the Twitter Streaming
API.

    def __init__(self):
        super().__init__()
        self.max_tweets = 10000
        self.tweet_count = 0

    def on_connect(self):
        # Called initially to connect to the Streaming API
        print("You are now connected to the streaming API.")

    def on_error(self, status_code):
```



```
        # On error - if an error occurs, display the error / status code
        print('An Error has occured: ' + repr(status_code))
        return False

    def on_data(self, data):
        # This is the meat of the script...it connects to your mongoDB and
stores the tweet
        try:
            client = MongoClient('localhost',27017)

            # Use twitterdb database. If it doesn't exist, it will be
created.
            db = client.twitterdb

            # Decode the JSON from Twitter
            datajson = json.loads(data)

            # grab the 'created_at' data from the Tweet to use for display
            created_at = datajson['created_at']

            # print out a message to the screen that we have collected a
tweet
            print("Tweet collected at " + str(created_at))

            # insert the data into the mongoDB into a collection called
twitter_search
            # if twitter_search doesn't exist, it will be created.
            db.twitter_search.insert_one(datajson)
            self.tweet_count += 1
            if (self.tweet_count == self.max_tweets):
                print("completed")
                return (False)
        except Exception as e:
            print(e)



auth = OAuthHandler(API_KEY, API_SECRET_KEY)
auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)
tweetStream = Stream(auth, StreamListener())
tweetStream.filter(track=["covid","covid19"],languages=['en'])


def retrive_data():
    try:
        client = MongoClient('localhost', 27017)
        db = client.twitterdb
        pipeline1 = [ { "$group" : { "_id" : "$user.location", "count": {
"$sum": 1 }} } ]
        results1 = db.twitter_search.aggregate(pipeline1)
        pipeline2 = [{"$match":{"user.location" :"USA"}}]
        results2 = db.twitter_search.aggregate(pipeline2)
        pipeline3 = [{"$match":{"user.location" :"USA"}},{"$group": {
"_id": "null", "count": { "$sum": 1 } } }]
        results3 = db.twitter_search.aggregate(pipeline3)
        print("Query 1 - $group result")
        for result in results1:
            print(result)
        print("Query 2 - $match result")
        for result in results2:
            print(result)
        print("Query 3 - Aggregation pipeline result")
        for result in results3:
            print(result)
    except Exception as e:
        print(e)


retrive_data()
```