Aaron Daniels

CIS593

Lab 3

2430989

I first wrote a script to return my Mongo DB connection. This returns the client to the mongo db. Its named getdb.py

Next, I setup the twitter stream. My code is in twitterStream.py. I extended the tweepy StreamingClient with my class, IDPrinter. I rewrote the on\_data method that is called each time data is received from the twitter stream. At the end it calls the original super().on\_data method

def on\_data(self, raw\_data):

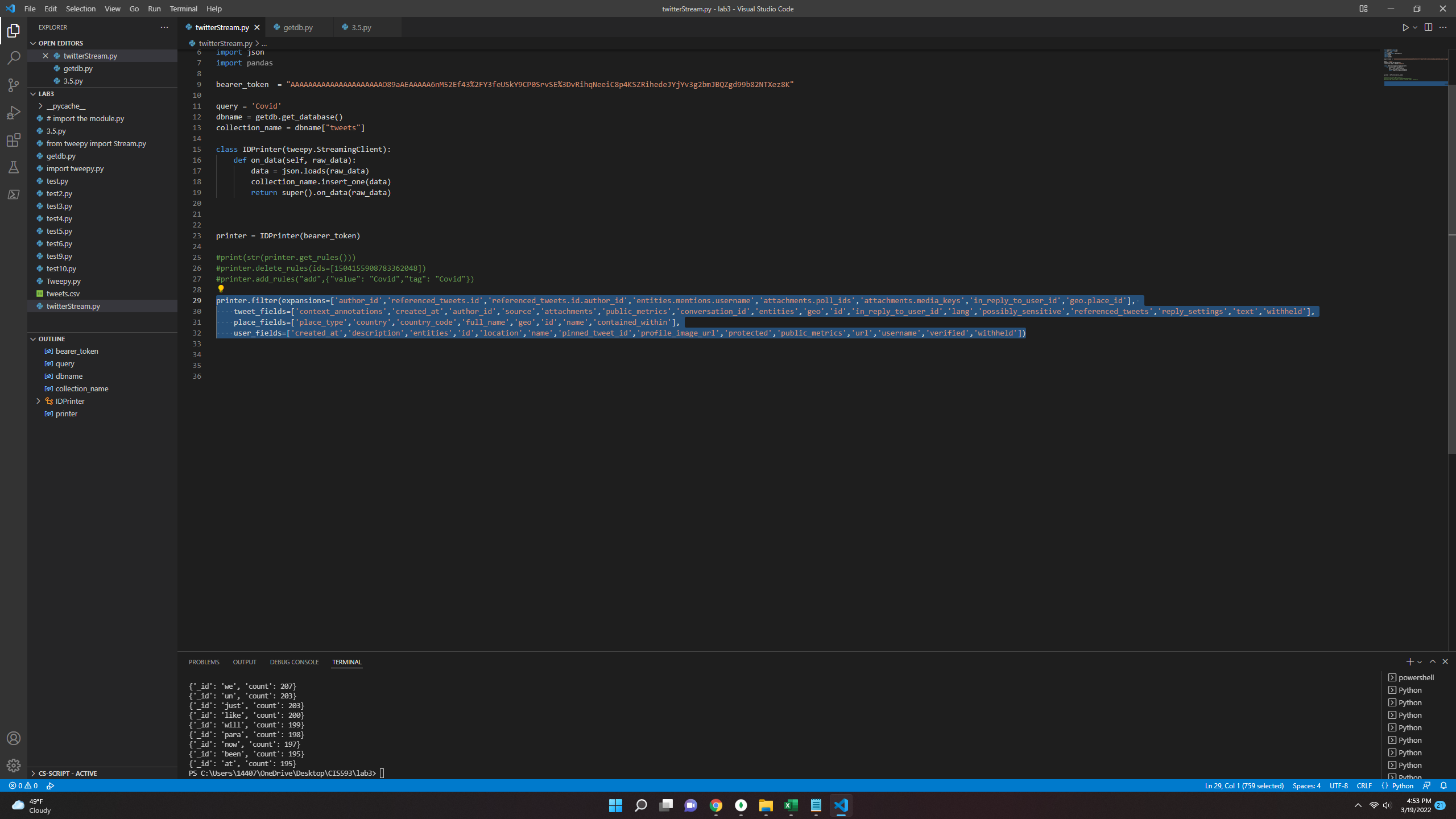
data = json.loads(raw\_data)

collection\_name.insert\_one(data)

return super().on\_data(raw\_data)

It adds a step to insert the json into my Mongo db. I added my twitter rule to filter for Covid. Then, I ran my script to start retrieving tweet data.

I wasn’t able to get any geo data for 3.4. I literally requested every piece of data that I could, and geo data just isn’t showing up. I did some google searches and eventually gave up. I think I needed to use twitter api v1.1 or not use tweepy.



I used all the expansions and optional data fields.

printer.filter(expansions=['author\_id','referenced\_tweets.id','referenced\_tweets.id.author\_id','entities.mentions.username','attachments.poll\_ids','attachments.media\_keys','in\_reply\_to\_user\_id','geo.place\_id'],

tweet\_fields=['context\_annotations','created\_at','author\_id','source','attachments','public\_metrics','conversation\_id','entities','geo','id','in\_reply\_to\_user\_id','lang','possibly\_sensitive','referenced\_tweets','reply\_settings','text','withheld'],

place\_fields=['place\_type','country','country\_code','full\_name','geo','id','name','contained\_within'],

user\_fields=['created\_at','description','entities','id','location','name','pinned\_tweet\_id','profile\_image\_url','protected','public\_metrics','url','username','verified','withheld'])

I used the Counter package to count the most common words

