README:

This is a small project that I did for my course in databases.
It includes using the tweeps api for twitter data collection and folium api for map plotting.
Also you can refer this if you want to refer as to how to work with MongoDB and also dictionary and lists in Python.
It also includes a solution for sentiment analysis about tweets. So, that can also be helpful if you're trying to do some sentiment analysis or NLP.

Setup :

1) Install and check if Python 3 is working perfectly.
2) Install mongoDB and keep running it in background.
3) Create a tweep app and generate the api keys the are left blank in the code where we are actually collecting tweets.
4) Install the following libraries :
    1) Folium (Map plotting)
    2) TextBlob (for NLP)
    3) Emoji (for emoji detection and collection)
5) Copy the files in a depository and run pranav_pa3.py for collecting the tweets all around the world that contain a certain text in them. And if you want tweets region wise I have collected tweets from US in us_tweets.py
6) It will automatically create necessary databases.

File name : pranav_pa3.py
This file connects with the twitter api and scrapes the tweets, in real time, that contain WORDS given in the questions.
It the creates a new database named 'twitter_db' and stores it in its collection 'twitter_search' .

File name : pranav_part1.py
This file contains the answers as output to the questions in PA3_extra for 660 students. For the last question of sentiment analysis it shows the whole text or extended_tweet.full_text which if it exists and its sentiment, calculated on basis of the polarity, on the next line.

File name : us_tweets.py
This file connects with the twitter api and scrapes the tweets in real time that are tweeted in USA. this was done using the bounding box of USA.
I have added an extra condition the tweet should have country_code 'US' and the coordinates should be present, which will help us while plotting it on the map. So, it is

useful while mapping it to the Map. And it also helps to reduce the number of tweets encountered while creating dictionaries for the questions.
Having ‘country code’ condition does not capture tweets from other countries that are included in the bounding box of USA. (like Mexico)

File name : us_tweets_answers.py
This file has the python code which uses pymongo library and to run queries using mongodb query format.
Computation for every question is seperated by multiple hash tags.
The dictionary variable ‘state_per_emoji’ has following format {‘emoji’ :{‘state’: count}}
The dictionary variable ‘emoji_per_state’ has following format {‘state’ :{‘emoji’: count}}

File name : json_to_csv.py
This file converts the json data collected and stores it in ‘usa_tweets.csv’.

File name : map_gen.py
This file generates a ‘map.html’ using folium library for python on tweets. And maps every tweet on the map of usa.

File name : map.html
This is the html file generated by the map_gen.py. It is quite big so it takes some time to generate the map and therefore, you might have to wait for some time or run it a couple of times.
For reference I have included the Screenshot of the output.

Folder Extra_Credit:

File name: extra_credit.py
This file generates the the dictionary that contains state name as key and the top two emojis of the state as value.
And it also takes data of latitude and longitude from the zip_codes_states.csv file for every state in the USA.
Then it generates a map in the html format using folium that creates a pointer on that position included in the csv file, when clicked on that pointer it will show the top two emojis of that state.

File name: zip_codes_states.csv
This file has a state abbrevation and a random latitude longitude value in the bounding box of that state.

Here too, I have included a Screenshot of the output map.