Assignment 1

Topics covered:

- Docker containers and Docker-compose
- NoSQL

Marking Scheme (Total 100 points)

- Create keys to Collect Data from Twitter (10 points)
- Dockerize the app to collect data from Twitter (20 points)
- Publish the docker image of your app in docker hub in your account (20 points)
- Change the app to store data in mongoDb. Run the app with mongoDB (using images from docker-hub) using docker-compose (50 points)

Date of submission: 26th of April
Mode of submission: Email
Format of submission:
1) Code of your application with twitter API keys
2) DockerFile for your app you used to dockerize the app.
3) Link of your published docker image from docker hub. (make sure you make your docker image public)
4) Final docker compose file to run either the published docker image or from local docker image.
5) Full Name and Student ID:

Instructions

1) Collecting Data from Twitter (10 points)
Generate an account in twitter and create an app to get credentials for getting data

https://developer.twitter.com/en/apps

Sample of how to create an account:
https://towardsdatascience.com/access-data-from-twitter-api-using-r-and-or-python-b8ac342d3efe

This might be old so look up latest instructions or just follow getting started in Twitter API web page

As a Data scientist you would be spending time to collect data from different source. Take this is an exercise on learning how to get data from public sources based on documentation.

If you cannot finish this step send me an email and I will share my API keys with you for this step. You will not get points in that case.

2) Dockerize the app to collect data from Twitter :

Get the sample code from github link (https://github.com/rohit-nlp/BigDataCourseUB) from folder assisgnment_1 to read data from Twitter. You can use your own code if you want. I am using Twython library so make sure to install in your python env.

Now dockerize your application.
Hint: https://runnable.com/docker/python/dockerize-your-python-application

3) Publish the docker image of your app in docker hub in your account:

Once you test your docker image locally publish the docker image in your docker hub account. (We already saw this in class or you can check internet for steps)
Make sure to make your image.
You would need to créate an account in docker hub(https://hub.docker.com/) and make sure to login to docker using command
$ docker login

4) Change the app to store data in mongoDb. Run the app with mongoDB (using images from docker-hub) using docker-compose.

The sample code gets data from Twitter and saves it a json. Change it to save the data in mongoDB.
Use mongoDB docker image to run it locally and then connect to that docker using pymongo library.

Create a docker-compose file to run your updated application and mongoDB together.
Submit the docker-compose file.