**Collecting tweets related to stock market**

**Objectives**

* Gain experience of collecting data from Twitter using Twitter API
* Gain experience of using database to store the data and query
* Gain experience of collecting real-time data

**Important Notes**

* Work in groups of two students
* All reports must be submitted as a PDF, with source code as an archive (e.g. zip, tar.gz)
* Save the submission as “group name.zip”
* Push all the codes and report in to your repository in Git Hub. It is recommended that you become familiar with version control systems, in particular Git as it will be very beneficial for your collaborative work. The following links will help you to become familiar with git:

<https://help.github.com/articles/set-up-git/#platform-windows>

<https://try.github.io/levels/1/challenges/1>

* If you have any issue contact the TA with your submission at [somayyeh.aghababaei@uoit.ca](mailto:somayyeh.aghababaei@uoit.ca)

**Assignment Details**

This assignment consists of three parts:

1. **Collecting data:** In this assignment you need to collect data related to stock market from Twitter for one week. In Twitter, ticker symbols like [#gold]($APPL) are used for stocks and companies. You are requested to collect the tweets with some specific keywords and store them in a database. The following keywords should be used:
   1. **Altcoin**
   2. **Bitcoin**
   3. **Coindesk**
   4. **Cryptocurrency**
   5. **Gold**
   6. **APPL**
   7. **GOOG**
   8. **YHOO**

Each tweet is a json file with the following format:

{"created\_at":”……….”,

"id":”………..”,

"text":" Time to buy some ether!\n#ethereum #investing #cryptocurrency”

“user\_id”:”………..”

…

}

1. **Saving data:** You need to save the requested data in one database with 8 collections where data related to each keyword is saved. The tweets related to each keyword are saved separately to one collection. As a result you need to have a database (named “tweets”) and 8 different collections (Altcoin, Bicoin,…).
2. **Visualizing data:** You need to present the daily number of tweets for each keyword as well as the daily number of users.

**Recourses**

# It is recommended for this assignment to use Python for programming with Tweepy and PyMongo that are very useful for Twitter APIs and MongoDB.

# The following resources are helpful.

# <https://dev.twitter.com/overview/documentation>

# <https://www.mongodb.com/>

# <https://www.python.org/doc/>