**Twitter Text Capture and Analysis**

**Hima Bindu Sigili**

**Tulasi Seshadri**

**Samuel Buahin**

**Sunidhi Gupta**

**Jerin Roy**

**Abishek Surya Rathicka Sridhar**

**Objective:**

This activity deals with text analysis of data pulled from twitter.Using this data, we answer a few important questions that helps us to get better insight into the collected data and derive meaning from it.

**Procedure:**

1. Download the twitter script from the following link & look at the script.

<https://drive.google.com/open?id=0B3n4oQctrRw6cFJIM3N4Q3ZMZFE>

* 1. How does the program work?

**The program works by retrieving tweets from twitter based on a search string the user enters.**

* 1. How do you think you can use this code?

**This code can be used by entering a twitter ACCESS KEY, SECRET KEY, CONSUMER KEY AND CONSUMER SECRET into the main program.**

* 1. Can you think of different scenarios where this code could be used for data collection?

**It can used to pull tweets onto a feed on a website and can also be used to pull data about tweets pertaining to a particular user or topic.**

1. Add your twitter API key in the script. You can generate one using <https://apps.twitter.com/> .

ACCESS\_TOKEN = '67269536-lrcoKLBp6mWovGOPv4q29GwfFzsZpYjYjT6dd2Ih4'

ACCESS\_SECRET = '1zgufGUnudFwXNw9ULFQ4L3BYdPXi7CA0KSgIPnXBlJa9'

CONSUMER\_KEY = 'w1d7ZYSGKsITgSMOt3j9NEYvg'

CONSUMER\_SECRET = '8is1lsAZdsTOBnVEjsnQDKQcBWLyyXQiPva0r0jMgxKJaIqb4u

1. Run twitter.py , input the required details to generate the output file
2. The output file contains user name, time, tweet, retweet\_count, and the followers for the user.

**Challenge:**

1. Now write a program to read the file generated. Your program should get the value of n (the number of records to be displayed) from the user, and generate the following output and write them into separate files.
   1. The top n users who have tweeted the most for the entire timeline.
   2. The top n users who have tweeted the most for every hour.
   3. The top n users who have the maximum followers.
   4. The top n tweets which have the maximum retweet count.

**Guidelines:**

1. Attach a zip file containing text document, source code as well as the output generated for the above scenarios for top 10 users and include how to run your program in the word document.

**How to run a program:**

* 1. **Enter the path of the input file and output file.**
  2. **The program generates the output and writes them**

**to separate text files.**

1. Check in your code in github and specify the github repository link in the word document. You can refer the steps below to create a repository in github.

**Creating a repository in GitHub**

1. Go to<https://github.com/>
2. Select Start a project.
3. Specify a repository name in the text box. Click create repository.Use Git bash to move to the directory that contains the folder to be moved from local machine to Github remote repository.
   1. **To initialize git:** git init
   2. **To add the files to git:** git add README.md
   3. **Committing the files to repository:** git commit -m "first commit"
   4. **Add it to your github repository:** git remote add origin**“your github repository name”**
   5. **Push it to the repository:** git push -u origin master

**Future enhancements:**

1. Sentiment analysis based on some keywords that reflect the emotions related to the tweets.
2. Location based searching to identify local preferences and trends.