Mini Project | Real Time Data Analytics

Dear Participant,  
  
Please find below the Mini Project for Real-Time Data Analytics. Kindly submit your assignment before the deadline.

**Sentiment Analysis on Twitter in Real Time -**

**Problem Summary:** On the real time tweets streams data apply Sentiment Analysis to get the opinion & trends on different topics in real time. Also find influential people or POI (person of Interest) from the twitter data.

Analysis to be done on the Twitter streaming data is:  (**Take time interval to be 10 seconds.**)

Steps to be followed :

1. **Find Influential people in twitter:**
   1. **Hint** - For simplicity assume the algorithm to find influential person is directly proportional to followers.
   2. Find top 10 Influential personalities from the twitter across the globe. (In an interval of 10 seconds).

1. **Get trending Topics in twitter right now:**
   1. Find the word from the tweet which is occurring most of the times in the whole tweets corpus.
   2. Top 10 trending topic (In an interval of 10 seconds).

1. **Sentiment Analysis:**
   1. Filter tweets & take tweets which has mention of “iphone”.
   2. Get sentiments about iphone. (ie. In an interval of 10 seconds) - [positive, negative or neutral]

**Optional (For practice) :**

1. **Geo based analysis:**
   1. Find the country based tweets count.
2. **Find Top tweeting user:**
   1. Find User who is tweeting a lot.
   2. Find top 50 across the world.
3. **Language based analysis:**
   1. Find counts of tweets in different languages.
   2. Get top 7 languages in terms of volume.
4. **Most popular tweets:**
   1. Most popular tweets means here is the tweet which has been re-tweeted maximum number of times.
   2. Get top 100 most re-tweeted tweets in last 1 hour related to “iphone”.
5. **Plot** all above analysis (in bar chart or for Geo analysis heat map). It totally depends on your creativity to get some interesting insights & visualization plots from all above analysis.

**Hints:**

1. Attaching library to do sentiment analysis: (optional)
   1. from textblob import TextBlob
   2. sentiment\_details = TextBlob(tweet\_text)
   3. sentiment\_polarity = sentiment\_details.sentiment.polarity
2. For plot the graphs & charts, you can use matplotlib.
3. You can create server in your local machine(cloud lab).
4. For any doubt, please refer to the code which I had shared in the classroom while showing Live Twitter Demo.
5. In case of any doubt, please reach out to us.

**Twitter Fields of json object:**

Can access from:  [https://dev.twitter.com/overview/api/tweets (Links to an external site.)Links to an external site. (Links to an external site.)Links to an external site.](https://dev.twitter.com/overview/api/tweets)

If you are not able to find anything, Please find below a sample tweet json:

{

            "text": "RT @PostGradProblem: In preparation for the NFL lockout, I will be spending twice as much time analyzing my fantasy baseball team during ...",

            "truncated": true,

            "in\_reply\_to\_user\_id": null,

            "in\_reply\_to\_status\_id": null,

            "favorited": false,

            "source": "<a href=\"http://twitter.com/\" rel=\"nofollow\">Twitter for iPhone</a>",

            "in\_reply\_to\_screen\_name": null,

            "in\_reply\_to\_status\_id\_str": null,

            "id\_str": "54691802283900928",

            "entities": {

            "user\_mentions": [

                        {

                        "indices": [

                                    3,

                                    19

                        ],

                        "screen\_name": "PostGradProblem",

                            "id\_str": "271572434",

                        "name": "PostGradProblems",

                        "id": 271572434

                        }

            ],

            "urls": [ ],

            "hashtags": [ ]

            },

            "contributors": null,

            "retweeted": false,

            "in\_reply\_to\_user\_id\_str": null,

            "place": null,

            "retweet\_count": 4,

            "created\_at": "Sun Apr 03 23:48:36 +0000 2011",

            "retweeted\_status": {

            "text": "In preparation for the NFL lockout, I will be spending twice as much time analyzing my fantasy baseball team during company time. #PGP",

            "truncated": false,

            "in\_reply\_to\_user\_id": null,

            "in\_reply\_to\_status\_id": null,

            "favorited": false,

            "source": "<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">HootSuite</a>",

            "in\_reply\_to\_screen\_name": null,

            "in\_reply\_to\_status\_id\_str": null,

            "id\_str": "54640519019642881",

            "entities": {

                        "user\_mentions": [ ],

                        "urls": [ ],

                        "hashtags": [

                        {

                                    "text": "PGP",

                                "indices": [

                                    130,

                                    134

                                    ]

                        }

                        ]

            },

            "contributors": null,

            "retweeted": false,

            "in\_reply\_to\_user\_id\_str": null,

            "place": null,

            "retweet\_count": 4,

            "created\_at": "Sun Apr 03 20:24:49 +0000 2011",

            "user": {

                        "notifications": null,

                        "profile\_use\_background\_image": true,

                        "statuses\_count": 31,

                        "profile\_background\_color": "C0DEED",

                        "followers\_count": 3066,

                        "profile\_image\_url": "http://a2.twimg.com/profile\_images/1285770264/PGP\_normal.jpg",

                        "listed\_count": 6,

                        "profile\_background\_image\_url": "http://a3.twimg.com/a/1301071706/images/themes/theme1/bg.png",

                        "description": "",

                     "screen\_name": "PostGradProblem",

                        "default\_profile": true,

                        "verified": false,

                        "time\_zone": null,

                        "profile\_text\_color": "333333",

                        "is\_translator": false,

                        "profile\_sidebar\_fill\_color": "DDEEF6",

                        "location": "",

                        "id\_str": "271572434",

                        "default\_profile\_image": false,

                        "profile\_background\_tile": false,

                        "lang": "en",

                        "friends\_count": 21,

                        "protected": false,

                        "favourites\_count": 0,

                        "created\_at": "Thu Mar 24 19:45:44 +0000 2011",

                        "profile\_link\_color": "0084B4",

                        "name": "PostGradProblems",

                        "show\_all\_inline\_media": false,

                        "follow\_request\_sent": null,

                        "geo\_enabled": false,

                        "profile\_sidebar\_border\_color": "C0DEED",

                        "url": null,

                        "id": 271572434,

                        "contributors\_enabled": false,

                        "following": null,

                        "utc\_offset": null

            },

            "id": 54640519019642880,

            "coordinates": null,

            "geo": null

            },

            "user": {

            "notifications": null,

            "profile\_use\_background\_image": true,

            "statuses\_count": 351,

            "profile\_background\_color": "C0DEED",

            "followers\_count": 48,

            "profile\_image\_url": "http://a1.twimg.com/profile\_images/455128973/gCsVUnofNqqyd6tdOGevROvko1\_500\_normal.jpg",

            "listed\_count": 0,

            "profile\_background\_image\_url": "http://a3.twimg.com/a/1300479984/images/themes/theme1/bg.png",

            "description": "watcha doin in my waters?",

            "screen\_name": "OldGREG85",

            "default\_profile": true,

             "verified": false,

            "time\_zone": "Hawaii",

            "profile\_text\_color": "333333",

            "is\_translator": false,

            "profile\_sidebar\_fill\_color": "DDEEF6",

            "location": "Texas",

            "id\_str": "80177619",

            "default\_profile\_image": false,

            "profile\_background\_tile": false,

            "lang": "en",

            "friends\_count": 81,

            "protected": false,

            "favourites\_count": 0,

            "created\_at": "Tue Oct 06 01:13:17 +0000 2009",

            "profile\_link\_color": "0084B4",

            "name": "GG",

            "show\_all\_inline\_media": false,

            "follow\_request\_sent": null,

            "geo\_enabled": false,

            "profile\_sidebar\_border\_color": "C0DEED",

            "url": null,

            "id": 80177619,

            "contributors\_enabled": false,

            "following": null,

            "utc\_offset": -36000

            },

            "id": 54691802283900930,

            "coordinates": null,

            "geo": null

}

**Please Note:**

* **You are only required to answer questions 1-3 for evaluation, the remaining questions are optional and for practice**
* Total marks allotted for Mini Project is 50.
* Please submit working code and output(of each step) along with it in pdf,html and ipynb format.
* Please add necessary comments also in your solution file.
* Name format of files should always be **Group Number\_RTD\_MiniPro.extension**

Marks Distribution is as follows --

1. Find Influential people in twitter: **17.5 points**
2. Get trending Topics in twitter right now: **17.5 points**
3. Sentiment Analysis: **25 points**

Regards,  
Program Office