

PSOSM Assignment 2

Divay Prakash - 2014039

Ans (1)

Present in the form of 2 files titled 'delhipolice.json' and 'bostonpolice.json' present in the 'data' folder. The files contain all tweets (including retweets but excluding replies) made by the handles during 1st August 2017 - 7th September 2017

Ans (2) 1, 2, 4 and (3) 1

Parts 1, 2 and 4 of question 2 and part 1 of question 3 are present in the script logs below. Also time series graphs measuring handle activity using tweet frequency were created and are present in the 'graphs' folder as two SVG files titled 'delhi.svg' and 'boston.svg'.

'@DelhiPolice' logs

```
divay@divay-linux:~/psosm_a2$ python twitter.py
Collecting tweets...
Done!
```

```
No. of tweets by @DelhiPolice: 207
Frequency of tweets by @DelhiPolice: 5.44736842105 tweets per day
No. of ORIGINAL tweets by @DelhiPolice: 60
Frequency of ORIGINAL tweets by @DelhiPolice: 1.57894736842 tweets per day
```

```
IMP: Like/Retweet count is only for original tweets by handle and does not include
retweets as that taints data
```

```
Number of likes on these tweets: 8292
Number of retweets of these tweets: 4244
Engagement score for @DelhiPolice: 87.0
```

```
Number of tweets with text only: 42
Number of likes on these tweets: 6661
Number of retweets of these tweets: 3492
Average number of likes on tweets with text only: 158
Average number of retweets of tweets with text only: 83
```

```
Number of tweets with image: 18
Number of likes on these tweets: 1631
Number of retweets of these tweets: 752
Average number of likes on tweets with image: 90
Average number of retweets of tweets with image: 41
```

```
Number of tweets with video: 0
Number of likes on these tweets: 0
Number of retweets of these tweets: 0
```

```
@DelhiPolice has retweeted tweets from the following 10 user IDs the most -
DCPNEastDelhi 20
DCP_DelhiPolice 17
DCP_Shd 13
DCPSEastDelhi 13
adcp1_south 10
DCPEastDelhi 9
LtGovDelhi 8
surenderkr_ips 7
DcpNorthDelhi 6
CPDelhi 5
```

'@bostonpolice' logs

```
divay@divay-linux:~/psosm_a2$ python twitter.py
Collecting tweets...
Done!
```

```
No. of tweets by @bostonpolice: 177
Frequency of tweets by @bostonpolice: 4.65789473684 tweets per day
No. of ORIGINAL tweets by @bostonpolice: 175
Frequency of ORIGINAL tweets by @bostonpolice: 4.60526315789 tweets per day
```

IMP: Like/Retweet count is only for original tweets by handle and does not include retweets as that taints data

```
Number of likes on these tweets: 52337
Number of retweets of these tweets: 23850
Engagement score for @bostonpolice: 176.75
```

```
Number of tweets with text only: 150
Number of likes on these tweets: 49351
Number of retweets of these tweets: 23077
Average number of likes on tweets with text only: 329
Average number of retweets of tweets with text only: 153
```

```
Number of tweets with image: 25
Number of likes on these tweets: 2986
Number of retweets of these tweets: 773
Average number of likes on tweets with image: 119
Average number of retweets of tweets with image: 30
```

```
Number of tweets with video: 0
Number of likes on these tweets: 0
Number of retweets of these tweets: 0
```

```
@bostonpolice has retweeted tweets from the following 10 user IDs the most -
NBCNightlyNews 1
MichaelFlaher19 1
```

The function used to create 'Engagement Score' was-

```
#likeCount is number of likes on handle's ORIGINAL tweets
#retweetCount is number of retweets of handle's ORIGINAL tweets
likeAvg = likeCount/originalTweetCount
retweetAvg = retweetCount/originalTweetCount
likeScore = likeAvg * 0.25          #measure of tweet appreciation
retweetScore = retweetAvg * 0.75    #measure of tweet value
engagementScore = likeScore + retweetScore
```

I was not able to obtain replies to any tweets made by the police handles. The various methods I tried included -

- **Twitter Search API** using the 'to:' parameter. However, very less data was returned.
- **Deprecated and undocumented 'related_results' experimental API endpoint**
- **'conversation' API endpoint** which was found to be blocked to third-party clients and returned JSON: '{"errors":[{"message":"SSL is required","code":92}]}'
- **'mentions_timeline' API endpoint** which returns data which needs to be matched to the original posts 'id' using the 'in_reply_to_status_id' parameter, however data returned was of recent days only so did not represent an accurate picture of the analysis period tweets

Ans (2) 3

Ans (3) 2