**Submit Your Data Story - Your Capstone Project Milestone Report**

Your Capstone project milestone report is an early draft of your final Capstone report. We encourage you and your mentor to plan multiple milestones if possible, since your story will evolve with exploration and analysis. This is a slightly longer (3-5 page) draft that should have the following:

* An introduction to the problem (based on your earlier Capstone submissions).

Unlike its predecessors, ISIS (Islamic State …) has been very successful in its propaganda campaign on Twitter. Furthermore, ISIS has been using Twitter as an effective recruiting tool.

* A deeper dive into the data set:
  + What important fields and information does the data set have?

Parse out ideas from here: <<< -http://blog.kaggle.com/2016/06/03/dataset-spotlight-how-isis-uses-twitter/ >>>

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Frequency of words | | |  |  | | #aleppo | 276 | | #amaqagency | 314 | | #breaking | 290 | | #breakingnews | 250 | | #iraq | 409 | | #is | 652 | | #isis | 1489 | | #islamicstate | 431 | | #syria | 1059 | | ... | 251 | | @nidalgazaui: | 302 | | @ramiallolah: | 505 | | abu | 348 | | allah | 489 | | army | 849 | | assad | 357 | | attack | 460 | | city | 357 | | fighters | 330 | | forces | 441 | | iraqi | 283 | | isis | 646 | | islamic | 643 | | killed | 1185 | | muslims | 251 | | north | 282 | | people | 376 | | rebels | 257 | | regime | 292 | | reports | 261 | | soldiers | 473 | | state | 597 | | syrian | 348 | | the | 968 | | this | 297 | | today | 383 | | west | 278 | | C:\Users\ababen\AppData\Local\Microsoft\Windows\INetCacheContent.Word\Rplot.png |

* + What are its limitations i.e. what are some questions that you cannot answer with this data set?

Twitter usernames of pro-ISIS fanboys are constantly changing because Twitter shuts them down.

Indicators of someone being pro-ISIS included: (a) keywords in the user’s name, description or tweets such as “Dawla” (which refers to the State), “Baqiyyah” (which denotes being part of the supposed ‘ever-expanding’ of the ISIS state), “Amaq” (the agency used by ISIS to issue official proclamations, “Wilayat” (used by ISIS to divide up the world into its provinces), etc. I also looked at imagery such as if a user had the ISIS flag or images of radical leaders like al-Baghdadi, Anwar Awlaki, etc. I also looked at who they were following and following them back. All of these factors helped me identify factors to collect the data.  
-http://blog.kaggle.com/2016/06/03/dataset-spotlight-how-isis-uses-twitter/

* + What kind of cleaning and wrangling did you need to do?

I had to use the “tm” and “stinger” package to remove stopwords (i.e. the, is, and, etc.), I removed links, retweets, hashtags, @people, punctuations, numbers, emojis and non-english characters.

* Any preliminary exploration you’ve performed and your initial findings.

For preliminary exploration, I looked at the dataset tweets to gather what the dataset looks like. Once I understood the format and structure, I wanted to know who are the most active twitter users in this dataset, what are the most frequent users, hashtags and words.

What can I learn from ISIS #hashtag campaigns?

* Based on these findings, what approach are you going to take? How has your approach changed from what you initially proposed, if applicable?

**What's the most interesting or insightful thing you've learned about the data?**

The most interesting insight I learned was how a handful of users are essentially thought leaders and influence much of the network. I thought it would be a bit more evenly spread across the network, but its clear that 3-5 people are generating the lion’s share of content and serve as connectors between content producers and content receivers.

-http://blog.kaggle.com/2016/06/03/dataset-spotlight-how-isis-uses-twitter/

Build a time series around time of events (i.e. attacks, elections, etc..

Basically, the milestone is an opportunity for you to practice your data story skills. We encourage you and your mentor to plan multiple milestones if possible.

Add your code and milestone report to the github repository. As before, once your mentor has approved your milestone document, please share the github repository URL on the community and ask the community for feedback.

Analyzed the network of words: <https://www.kaggle.com/ggospodinov/d/kzaman/how-isis-uses-twitter/tweet-analysis2>