What is the problem you want to solve?

​ The terrorist organization ISIS (“Islamic State of Iraq and the Levant”) is referred to by many names such as ISIL, IS and Da’esh. These acronyms refer to different stages of the same XXX terrorist which are occupying and controlling several countries in the Middle East. ISIS’ the terrorist organization extremelymodern attract new sis d aloneOriginally, was as it is difficult for a corporation to decide between hate crime and free speech. Additionally, it is difficult to distinguish between actual ISIS accounts and conversations about the organization. To add to the array of issues, the perpetrators switched opened new account faster than Twitter could identify them and shut them down. However, due to public pressure Twitter ramped up the company’s efforts and has360,000 accounts

The primary goal my project is to analyze the 17,000 twitter messages made available on Kaggle.com as part of an open-source XXX to prevent, counter-message and limit ISIS’s ability to utilize Twitter as its’ broadcast and recruiting tool.

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

​National Security Agency, Central Intelligence Agency, Center for Terrorist Control, Twitter, Google Think Tan, (story), Anonymous (hacker group).

Client can be a counter-terrorist government agencies, social media companies which may be negatively affected by terrorist's utilization of their network, non-profit “Think Tanks” which work on improving human living conditions, world's aid organization's such as Red Cross which responds to crisis worldwide, and lastly hacker groups such as “Anonymous” which have taken upon themselves to be the vigilantes on Twitter by launching a campaign called “#OpsISIS” on Twitter to identify, report and ultimately shut down pro-ISIS accounts. Additionally, the counter-messaging campaign has made it a lot more difficult to connect with ISIS recruiters due to the sheer volume of “tweets” generated by this effort.

Lastly, I would like to mention that Google's Think Tank "Jigsaw" was not only deeply interested in understanding pro-ISIS campaigns, they took it a step further by starting a project called “The Redirect Method” which “is focused on reaching those who are actively looking for extremist content and connections. Rather than create new content and counter-narratives, our approach tries to divert young people off the path to extremism using pre-existing YouTube content and targeted advertising.”

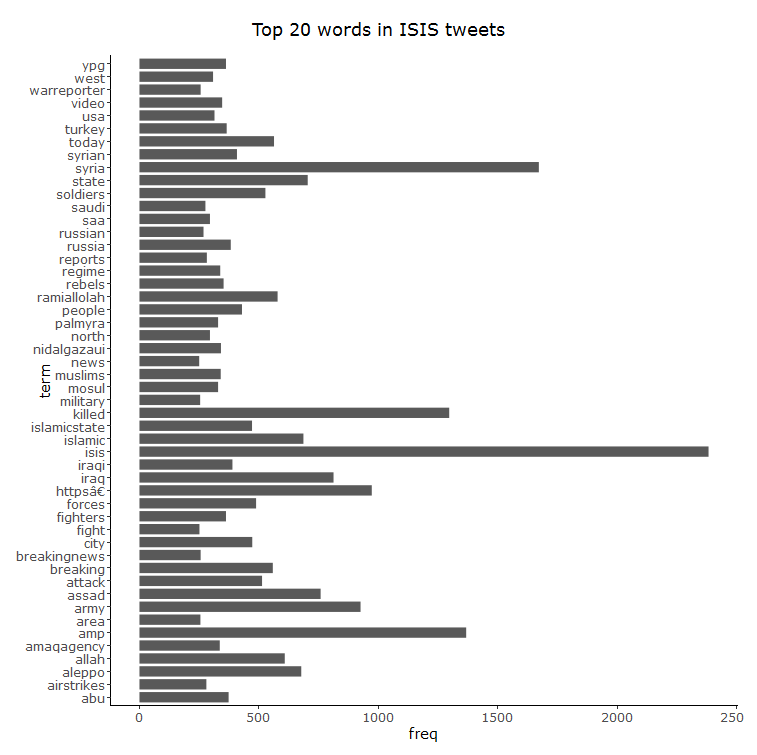
A deeper dive into the data set: What important fields and information does the data set have?

* Name
* Username
* Description
* Location
* Number of followers at the time the tweet was downloaded
* Number of statuses by the user when the tweet was downloaded
* Date and timestamp of the tweet
* The tweet itself

Thus, we cannot analyze long periods of tweet history and its effect on the user’s influence. It would be also beneficial if the edidentify ly createddestroyed

“”

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Frequency of words | | |  |  | | #aleppo | 276 | | #amaqagency | 314 | | #breaking | 290 | | #breakingnews | 250 | | #iraq | 409 | | #is | 652 | | #isis | 1489 | | #islamicstate | 431 | | #syria | 1059 | | @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 | | Word Cloud C:\Users\ababen\AppData\Local\Microsoft\Windows\INetCacheContent.Word\Rplot.png |

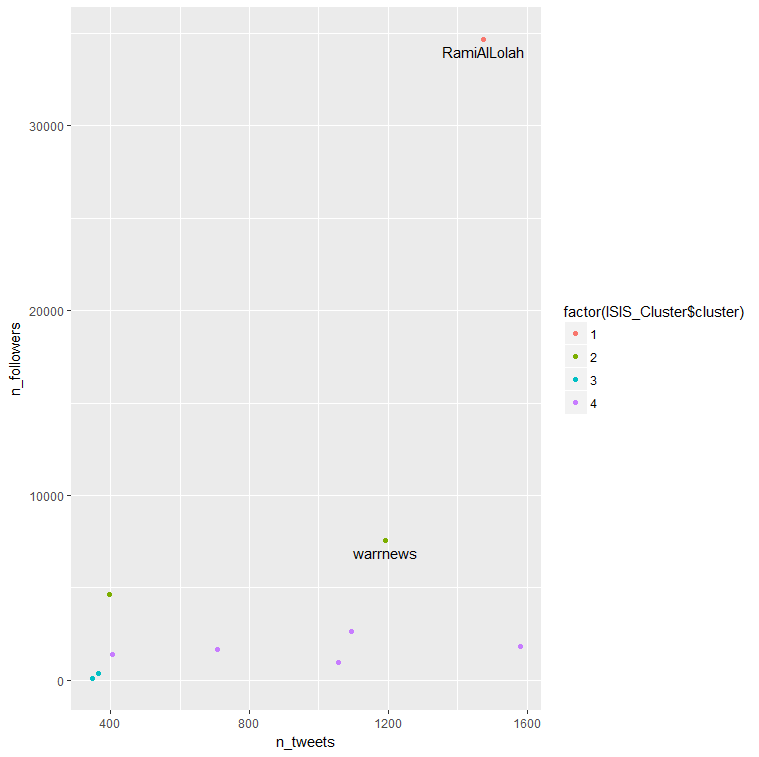




​



A handful of users do the majority of posting(s).



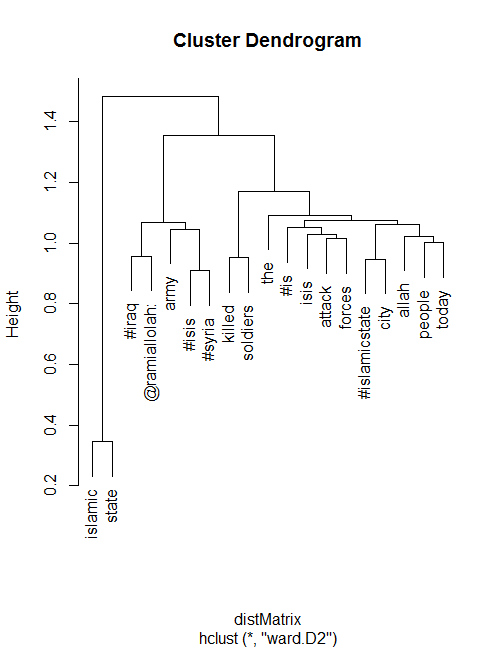
Based on these findings, what approach are you going to take? How has your approach changed from what you initially proposed, if applicable?

I would like to find how many of the original users posting are still posting today?

Is there any link between attacks and tweets?

**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/



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.

TO DO: Build a time series around time of events (i.e. attacks, elections, etc.)