

Data Analysis for Involved

Ninad Tungare (ntungare@bu.edu), Harshad Reddy Nalla (harshad@bu.edu)
Caleb McDermott (caleb@involved.vote)

The project to analyze the social media data based political issue and political representatives in different constituents. Also to provide more insights on the Boston 311 data.

The Idea

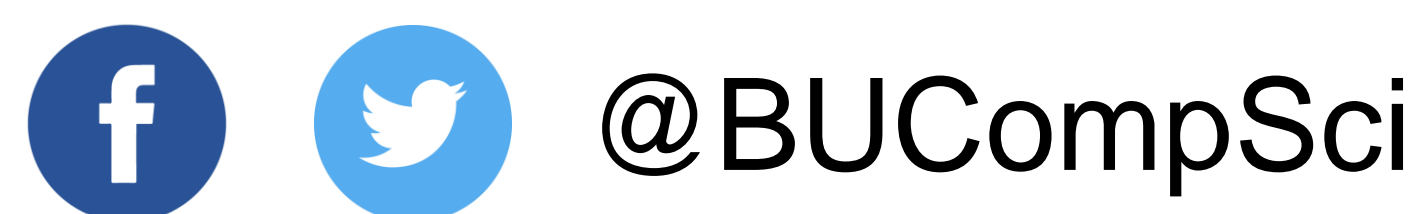
We provided analyses on the most popular topics being discussed upon a specific representative on Twitter and most requested Boston 311 Issues. We generated the dataset from social media and Boston 311, and used different analyses techniques to generate required results, such as finding the most common issues and topics that were discussed by the communities in Massachusetts.

Questions that were Answered

1. With reference to a representative, what topics are being discussed most frequently on Twitter?
2. Using Boston 311 data, which departments are getting the most number of issues in a given time period?
3. Using Boston 311 data, in a selected department what are the most prominent issues?

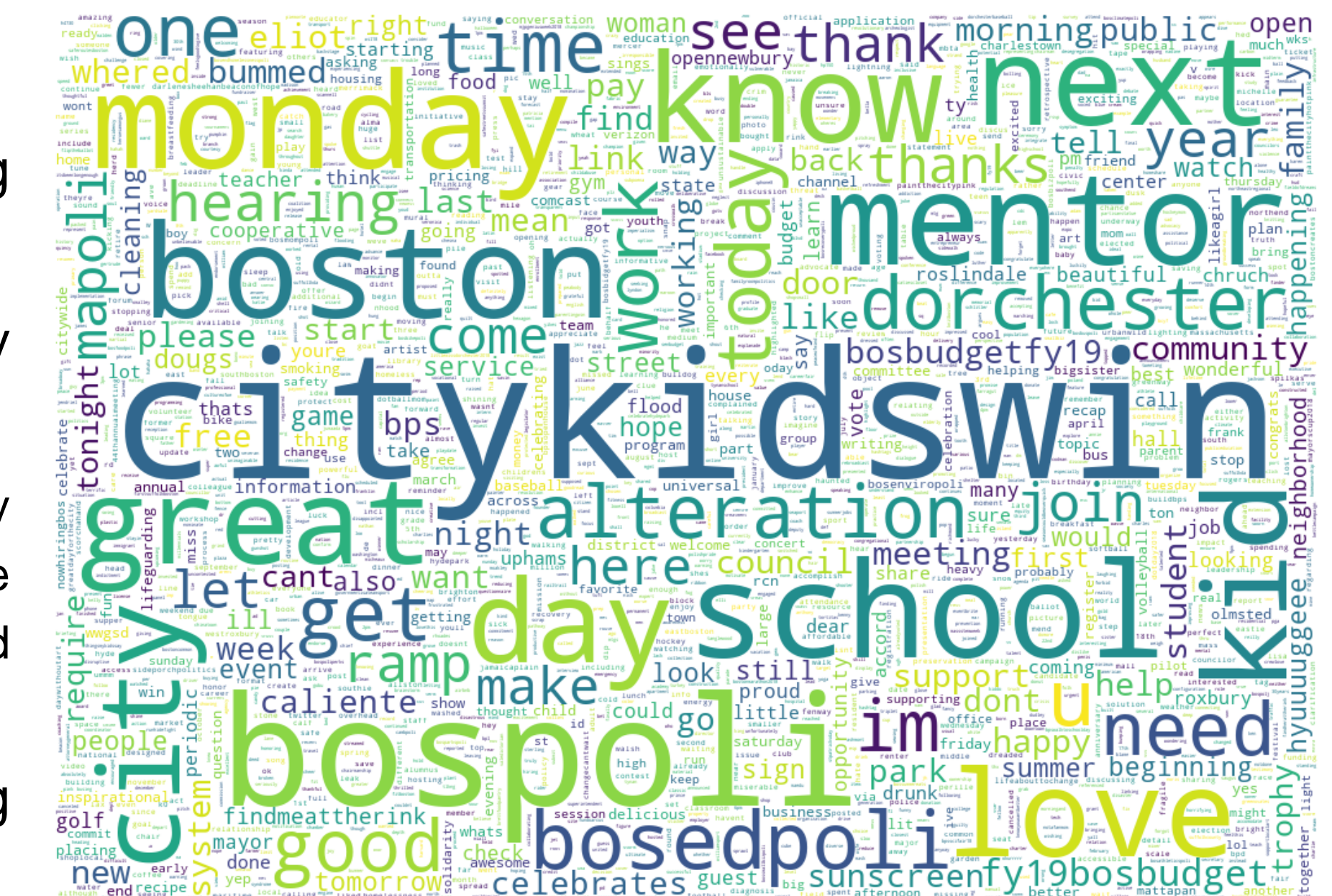
Source of Data

1. Twitter: Given a Twitter handle/name we collect 4 types of tweets per handle:
 1. Tweets sent by the handle
 2. Tweets/comments on the tweets sent by the handle
 3. Tweets which mention the account
 4. Tweets which are comments and mentions the account
2. Boston 311: Boston 311 data is a data source which is constantly updated with the latest Boston 311 issues and cases and with any updates on the old ones.



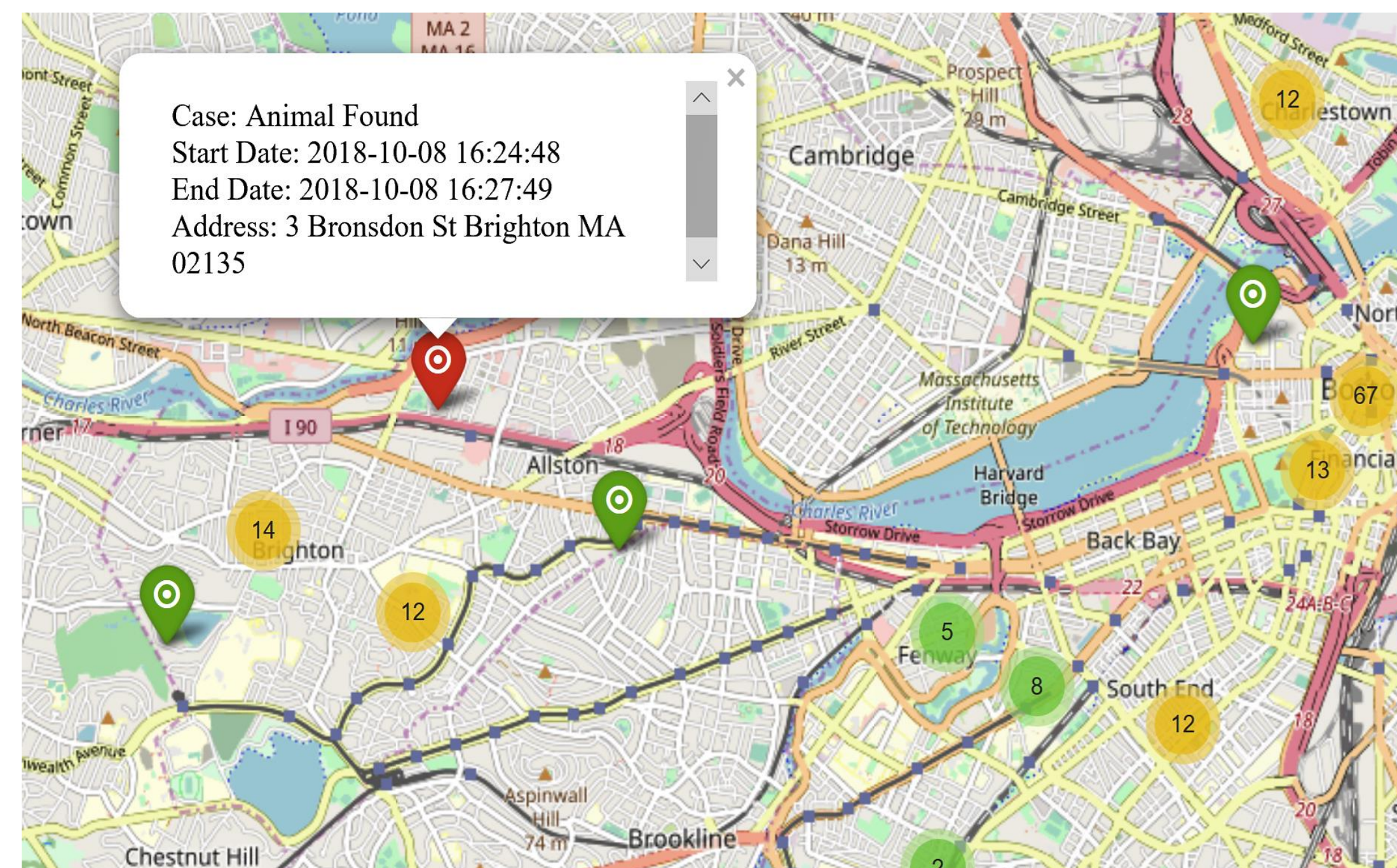
Twitter Analysis

- Tweets from the Massachusetts area were collected for the requested handles using Twitter API
- We developed an automated model which collects the data based on the filters, viz. they should be under the 4 types of tweets we are looking for.
 - Tweets were gathered using Tweepy and filtered out initially using the geocode and by whether they mention the handle we are looking for. We also extract all tweets in a random thread where someone is trying to bring this to the attention of the account and we do so by using the “in_reply_to_id” and iteratively extract all tweets and store all these tweets in mongo db.
- We ran TF-IDF on the collected data and created word weightage after removing unnecessary stop words.
- We then used these word weightages to generate this word cloud.



Boston 311 Map Visualization

- We developed an interactive map of the Boston 311 data, and we even integrated the following few filters:
 - Date-range: We can specify a start lookup date and end of lookup date.
 - Subject/Department: We can specify a department for the lookup.
 - Case: We can specify a case within a Subject/Department for the lookup.
- We also enabled clustering of data, such that an user can zoom into specific neighborhoods and see the data for only that area.



Boston 311 Analysis

