Assignment #2

Sanchit Singhal

Fall 2018, October 12, 2018

MIS 381N – User Generated Content Analytics w/ Dr. Anitesh Barua

McCombs School of Business, University of Texas at Austin

Task A

Approximately 10,000 tweets were collected via a Twitter API by providing the keywords of 'Senate' and 'Texas'.

Task B

After calculating the word frequencies, 4 key issues that were mentioned were found:

- Kavanaugh – 636 mentions
- Energy – 478 mentions
- Women – 406 mentions
- Jobs – 135 mentions

All variations of Cruz and Beto were replaced to ensure all relevant phrases are captured accurately.

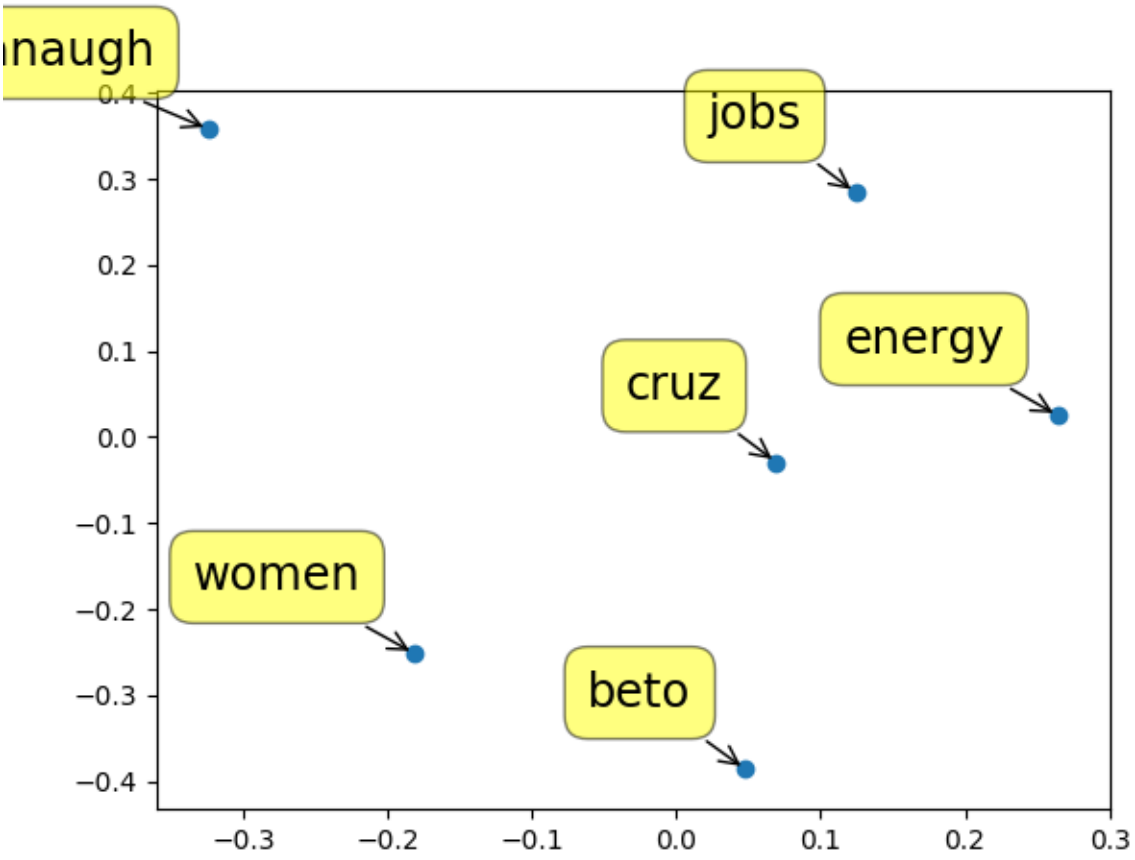
Task C

Lift and Sentiment scores were calculated for all pairs of Candidates and Key Issues:

	Lift	**Sentiment Score (if lift > 1)**
Cruz & Kavanaugh	0.046855	
Cruz & Energy	1.562649	0.462212 - Positive
Cruz & Women	0.143255	
Cruz & Jobs	1.575808	0.540307 - Positive
Beto & Kavanaugh	0.033002	
Beto & Energy	1.223895	0.555142 - Positive
Beto & Women	1.155435	0.281926 - Positive
Beto & Jobs	0.015506	

Task D

Using dissimilarity metric, the candidates and key issues were plotted on an MDS plot:



Task E

Using the location data from the tweets, Lift scores were calculated. Note – the top 10 cities by population were categorized as Large Cities and the rest as Small Towns (all from Texas):

	Beto	**Cruz**
Big Cities in Texas	1.273882	1.029067
Small Towns in Texas	1.095044	1.074642

Sub-sequentially, using the location data from the tweets sentiments scores were calculated. Note – again the top 10 cities by population were categorized as Large Cities and the rest as Small Towns (all from Texas):

	Beto	**Cruz**
Big Cities in Texas	0.331453- Positive	0.308256 - Positive
Small Towns in Texas	0.319262 - Positive	0.26806 - Positive

Task F

I would give the following recommendations to the candidates. It might be helpful for the concerned party but also to the counter-party as a competitive advantage:

- Both Beto and Cruz seem not have a high association with Kavanaugh at the time of tweet collection – hence it is furthest away from all others in the MDS. This is surprising because of the tremendous amount of media coverage around the topic – but this is neither a good thing or a bad thing for either candidate. I would advise the campaign teams to closely monitor the situation to best take advantage of a possible opportunity on a topic that is relatively neutral for both Beto and Cruz. I would be wary of jumping into the issue without much consideration though as one wrong move could lead to negative crowd sentiment as well.

- Both Beto and Cruz have a significant association with Energy – Cruz has a slightly higher lift value but Beto has a slightly higher positive sentiment. The role of Energy in the state of Texas is a dominant one and I would advise both candidates to maintain their current position on the subject.

- Women is an issue where the public perception with the two candidates seems to differ – while there is a high association with Beto, there hardly seems to be any with Cruz. Despite having a weak sentiment, Beto still maintains an overall positive sentiment on Women. I would advise Cruz's campaign to focus more on Women. After all, they make up almost 50% of voters and it would be unwise to let Beto take advantage of that.

- On the contrary, Cruz has a high association with Jobs as well as a relatively high positive sentiment – Beto seems to have almost no association. I would advise Cruz's campaign team to continue their efforts on this front as it is working well but Beto's advisors should be more aggressive in reaching the public around the issue.

- It is interesting to note that all four significant lift scores (associations above 1) have a positive sentiment. This tells me that both candidates, Beto and Cruz, and doing a good job at not maintaining negativity around their campaigns. The key issues they are associated with seem to be with a positive sentiments and the rest are not associated. Both campaigning strategies seem to be taking a very safe, conservative approach in this election.

- Cruz seems to be more in the middle of the MDS plot signifying that he is being associated to more issues than Beto.

- It's also interesting to note that jobs and energy are closely packed in a cluster on the MDS plot – which makes sense as the energy sector in Texas must create a lot of jobs for Texans.

- When analyzing location data, it seems like Beto has a slightly higher association in big cities while Cruz has a slightly higher association in small towns – although not by much. I would advise Beto's campaign to target the rural areas more aggressively. I would also advise Cruz's campaign to work more closely with the media in bigger cities.

- Sentiments for Beto and Cruz, in both big cities and small towns, seem to be positive. Again, I think the campaign's are doing a good job at maintaining non-negativity. It's interesting to note that bigger cities seem to have a more positive sentiment for both candidates than smaller towns – but only very slightly. This is again an indication that both parties are playing it very safe and avoiding too many controversial issues.

- I would advise both candidates to try to send more positivity in rural areas as it seems like they might be being ignored a little and that last push can help both candidates.

- When examining the number of positive, negative, or neutral sentiments it seems that most tweets are in fact neutral. This means that the crowd is generally refraining from expressing extreme emotions when discussing the elections on Twitter – again possibly due the campaign strategies.

- Overall, I think the Texas senate race is very tight. Both candidates stand fairly neutral on a lot of key issues. The two dividers being women and jobs from this analysis. Both campaigns are maintaining safe, positive approaches but could do a slightly better job at reaching the rural areas. At this point of the race, Cruz seems be holding the upper hand – but just slightly and there remain opportunities for either candidate to take the win by adapting their strategies using some of the recommendations provided above.

*Datasets and Analysis available upon request