Case Studies in Data Analytics (Assignment 3)- Government Analytics
Report By: Girija Shingte(18230228) and Dhaval Salwala(18230845)

Introduction:

Mining pages on social networking websites like Facebook and Twitter is greatly useful in identifying sentiments of people, determining trending topics, identifying target customers for marketing, etc. The task we have considered for this assignment is determing the trending policy topics in Galway based on analysis of tweets from local media house pages like Galway Bay FM News, Galway City Council, Galway County Council, Connacht/City Tribune, RTE News, Galway 2020, Galway Drivers, St. Patrick's Fest, Bus Eireann and Galway Page. To perform this task, we have made use of Python (for collecting tweets and processing the tweets to identify trending policies) and R (for visualizing our output). Subsequent sections include information on how data was gathered and processed, results and findings.

Steps and Design Decisions:

- **Accessing tweets from Twitter:**
  The following table includes detailed information on how tweets were collected:

Module/Page	Description
Module: GetOldTweets3	This module in python allows us to access old tweets on twitter. This module was chosen as the functions available within this module are easy to access. To retrieve 6 month old tweets on the pages, we only needed to specify the names of the pages, the start date and end date of the time frame in which the tweets are posted.
Page: Galway Bay FM news	Posts include news on current affairs, traffic and sports from Galway Bay FM.
Page: Galway City Council	Posts come from local authority for Galway city.
Page: Galway County Council	Posts come from local authority for Galway county.
Page: Connacht/City Tribune	News, Videos and posts from Connacht Tribune.
Page: Galway Page	Live updates on current affairs, culture and lifestyle in Galway city and Galway county.
Page: St Patricks Fest	Live news and videos related to St. Patrick's Day
Page: RTE News	News channel of Ireland
Page: Galway 2020	News, information and posts about Galway International Art Festival
Page: Galway drivers	Traffic information related to Galway city
Page: Bus Eireann	Bus and coach transport information

- **Identifying relevant posts:**

  Step a: Filtering Tweets
  In this section, we first decided 6 domains to which the tweets would belong. These 6 domains include Transport, Government, Arts-Culture, Demographics, Health-Eduction, and Crime-EmergencyServices. To choose these domains we referred to http://www.dublindashboard.ie. Then, we selected some keywords related to these domains. For example: if the domain is demographics, keywords can be ["population", "male", "female", "men", "women"] to name a few. We then scanned through every tweet and checked if the word in that tweet belonged to either of the domains. If the word belongs to a particular domain, that particular tweet which contains that word is added to its respective domain.

  Step b: Topic Modelling
  In this section, we use the LDA technique to determine the 3 trending topics within each of the six domains.

- **Analysis of relevant posts:**
  From the trending topics that we found in the previous section, we calculated the number of tweets that belong to that particular trending topic. At the end of this step, we had created a csv in which the first column represents the trending topic number, second column represents the keywords related to the trending topic, and the third column represents the domain to which the topic belongs.

- **Visualization of Results:**
  In this step, we have found the word clouds of each trending topic in each domain as word clouds give us a better understanding about the trending topics than the tabular representation. Additionally, we have also plotted a bar chart that represents the number of tweets present in each of the trending topics.

Analysis and Visualization:

Figure 1 represents the total number of tweets in each topics and their respective domains. From Figure 1, we can see that the highest number of tweets from the pages we selected are for the health and education domain whereas the lowest number of tweets are for arts and culture domain. After health and education, the most talked domain is transport followed by crime and emergency.

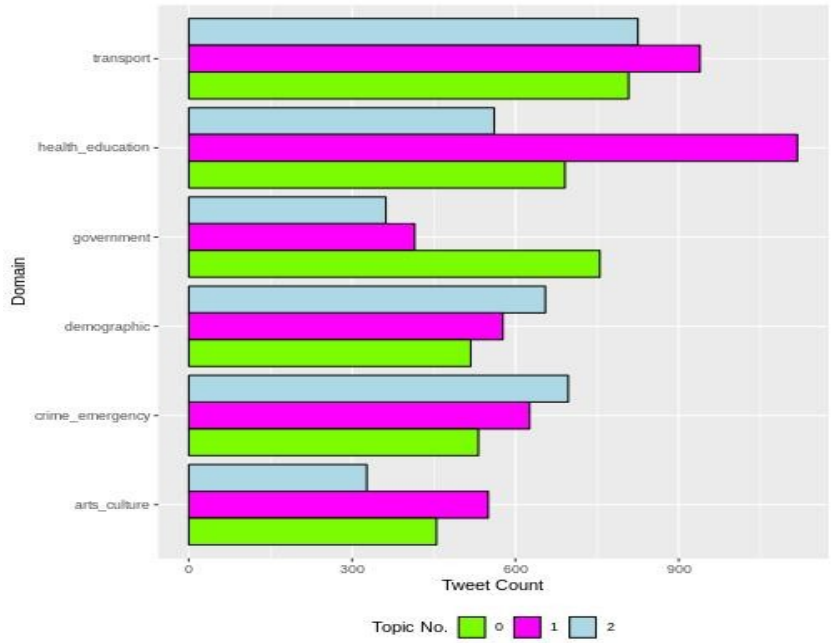

Figure 1: Tweet counts for respective domains and topics

The table below consists of the word clouds for each topic in each domain. With the help of these word clouds we have tried to analyse what the trending policy could be:

Topic and Domain	Analysis	Topic and Domain	Analysis
Domain: arts_culture -Topic No: 0 	From words like St Patrick's Festival, Parade, March, etc we can infer that the topic is about St Patrick's Day celebrations.	**Domain: arts_culture -Topic No: 1** 	From words like Galway, culture, project, volunteer, and launch we can infer that the topic is related to Galway International Art Festival 2020 as Galway is the host city. Galway is also the European Capital of Culture for the year 2020.
Domain: arts_culture -Topic No: 2 	In this topic, discussions about St Patrick's Day are not only restricted to Galway, but Dublin as well.	**Domain: crime_emergency -Topic No: 0** 	From the words like follow and safety, we can infer that the policies could be related to road safety.

Domain: crime_emergency -Topic No: 1 	From words in this cloud we can infer that the policies could be related to safety, policies against domestic violence, etc.	**Domain: crime_emergency -Topic No: 2** 	From words like safety, traffic and flood we can infer that the policies could be related to traffic management, road safety and flood control.
Domain: demographic -Topic No: 0 	From words like women, arrest, court, we can infer that the policies could be related to women rights.	**Domain: demographic -Topic No: 1** 	From words like women, city, library, we can infer that the ploicies could be related to basic public services offered by the government.
Domain: demographic -Topic No: 2 	From words like women, arrest, court, we can infer that the policies could be related to women rights.	**Domain: government -Topic No: 0** 	From words like brexit, british, vote, we can infer that the policies are related to Brexit.
Domain: government -Topic No: 1 	From words like social, house, develop , we can infer that the policies are related to city planning and development.	**Domain: government -Topic No: 2** 	From the words present in this cloud, we can infer that the policies are similar to the ones in topic 0 and 1.

Domain: health_education -Topic No: 0 	From words like special, contact, department, we can infer that the policies could be related to emergency services.	Domain: health_education -Topic No: 1 	From words like university, hospital, we can infer that the policies could be related to administration of hospitals and schools.
Domain: health_education -Topic No: 2 	From words like relevant, department, respond, we can infer that the policies could be related to how quickly departments should respond to an emergency situation.	Domain: transport -Topic No: 0 	From words like route, arrive, stop, road we can infer that the policies could be about road planning.
Domain: transport -Topic No: 1 	Due to presence of words like safety, advice, road, route we can infer that the policies could be related to road-safety.	Domain: transport -Topic No: 2 	From words like route, weather, eireann, delay. We can infer that the policies could be related to situations that arise out of weather conditions.

Conclusion:

Thus, after mining different pages on twitter that post news about Galway city/county and using different filtering criterias and topic modelling, we could identify different government policies related to different domains.

References:

Websites:

https://twitter.com/GalwayCityCo
https://twitter.com/GalwayCoCo
https://twitter.com/Galwaybayfmnews
https://twitter.com/CTribune
https://twitter.com/galwaypage
https://twitter.com/rtenews
https://twitter.com/Buseireann
https://twitter.com/GalwayDrivers
https://twitter.com/galway2020
https://twitter.com/stpatricksfest

Citations:

Reddick, C., Chatfield, A. and Ojo, A. (2017). A social media text analytics framework for double-loop learning for citizen-centric public services: A case study of a local government Facebook use. *Government Information Quarterly*, 34(1), pp.110-125.

Towards Data Science. (2019). *Topic Modelling in Python with NLTK and Gensim*. [online] Available at: https://towardsdatascience.com/topic-modelling-in-python-with-nltk-and-gensim-4ef03213cd21 [Accessed 19 Mar. 2019].

PyPI. (2019). *GetOldTweets3*. [online] Available at: https://pypi.org/project/GetOldTweets3 [Accessed 19 Mar. 2019].

Machine Learning Plus. (2019). *Topic Modeling in Python with Gensim*. [online] Available at: https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/ [Accessed 19 Mar. 2019].