Well-being and happiness among the elderly

What have Canadian societies discussed concerning the context of the Elderly and the Pandemic?

Ayushi Singh	100359100
Carlos Becerra	100357358
Erich Silva	100370053
Lucas Felix	100365550

April 17, 2022

Langara College

DANA-4850-001 Capstone Project

Instructor Quynh **Monica** Nguyen

ABSTRACT

To capture and assess the community's perception of elderly well-being in Canada, methods of data scrapping were utilized within three distinct sources: 1 – Twitter, as a representation of social media platforms, 2 – News bulletins of online journals, and 3 – Academic articles of Canadian institutions. All website and platform-use policies were verified and adhered to during the process. Data scraping techniques varied according to the source. For Twitter, the research team was granted access to the Twitter Academic Research API v2.0, which enabled the use of a more comprehensive query through the research study window that ranged from January 2018 to February 2022. Following their collection, datasets were cleaned and analyzed using Natural Language Processing methods and tools, including TextBlob [7], NLTK [1], Gensim [5] Python libraries, and the Hugging Face Transformer[11] library. Finally, the techniques were applied to grouped and ungrouped data within the three distinct sources to produce a dynamic dashboard that amplified the analysis and interpretation of trends and topics present in the period.

1 Foundational knowledge

The foundational knowledge of the project revolved around understanding how well-being is defined in the context of elders. According to the literature [9], well-being can be defined as:

> *"Well-being has been defined as the combination of feeling good and functioning well; the experience of positive emotions such as happiness and contentment as well as the development of one's potential, having some control over one's life, having a sense of purpose, and experiencing positive relationships." (Ruggeri et al., 2020).* [9]

This term of well-being is linked directly to positive mental health. In addition, it is related to one's success in life's professional, personal, and interpersonal aspects.

> *"Well-being goes beyond hedonism and the pursuit of happiness or pleasurable experience, and beyond a global evaluation (life satisfaction): it encompasses how well people are functioning, known as eudaimonic, or psychological well-being. The features encompass both hedonic and eudaimonic aspects of well-being:* ***competence, emotional stability, engagement, meaning, optimism, positive emotion, positive relationships, resilience, self-esteem, and vitality****." (Ruggeri et al., 2020).* [9]

After understanding what to look for, a dictionary of keywords was defined to be used with elderly-related terms and help with the search for data. This list was revisited multiple times and improved with feedback from two external consultants: Colleen Fraser and Bonnie Milne.

The list, Appendix A, was then expanded with synonyms and antonyms, resulting in 407 words used to compare the results of the topics found during our research.

2 Data scraping and cleaning

The scraping and cleaning processes were performed to compose our corpora from three proposed sources.

1. Twitter
2. Academic Articles
3. News Bulletins

Each data had to be scraped and cleaned in a particular way, and every file has its related code, listed in the following sections.

2.1 Twitter

Twitter data was scraped using Twitter API v2.0 – Academic research access [12]. The query was based on set parameters, including:

- Start Date: 01/01/2018.
- End Date: 25/02/2022.
- Geographical scope: All tweets sent from Canada.
- Only tweets in the English language.

The query arguments were defined after trying multiple combination of keywords. The goal was to maximise the results of relevant, on-topic tweets, while avoiding losing important information. The final query argument is shown below.

Query argument: (elders OR elder OR elderly OR "senior citizen" OR "senior citizens" OR "senior men" OR "senior women" OR "senior adult" OR "senior adults" OR seniority OR seniors OR "older adults" OR "older adult" OR "old men" OR "old women" OR "old ladies" OR "older ladies" OR "aging population" OR "aging people" OR "aging men" OR "aging women" OR "old age" OR OAP).

The response was stored in JSON format and converted to python dictionaries for cleaning and preprocessing. The following fields of each tweet were kept for later use during the analysis phase:

- Date_time
- Tweet_ID
- Author_ID
- Place_ID
- Cleaned_Tweet_Text

The cleaning process consisted of removing Unicode format emotes, symbols (#, @, $, …), and punctuations. The text was also lowercased. Hashtags were kept but with the pound (#) symbol removed.

Along with a generalized query, a selective one was also performed to target Canadian news profiles using the same query argument, including @CBCNews, @CdnPressNews, @CP24, @CTVNews, @globeandmail, and @nationalpost. They were kept in their dedicated CSV files.

Table 1 describes the academic paper scraping and cleaning process files.

Table 1. Twitter data structure

Twitter				
Original data	**Cleaned Data [.csv]**	**Cleaning code [.jpynb]**	**Count**	**File Size [kb]**
elderly_tweets	cleaned_elderly_tweets	Twitter_Scraping	57,138	25,591
CBCNews_tweets	cleaned_CBCNews_tweets	Twitter_Scraping	260	96
CdnPressNews_tweets	cleaned_CdnPressNews_tweets	Twitter_Scraping	15	5
CP24_tweets	cleaned_CP24_tweets	Twitter_Scraping	240	67
CTVNews_tweets	cleaned_CTVNews_tweets	Twitter_Scraping	482	131
globeandmail_tweets	cleaned_globeandmail_tweets	Twitter_Scraping	371	117
nationalpost_tweets	cleaned_nationalpost_tweets	Twitter_Scraping	210	59
		Total	**58,716**	**26,066**

It is important to observe that the tweets resulting from scrapping selected news accounts are not necessarily present in our general query results. This is because the latter considered only tweets with geo-localization tags.

2.2 Academic articles

Academic data was scraped from the website cognit.ca. The following filters were applied:

- Years: 2018, 2019, 2020 (this means period from 2018-2019, 2019-2020 and 2020-2021).
- Geographical scope: All Canada was included in this search.
- Keywords: "elder or seniors and well-being."

The original dataset consists of academic research documents based on its author, university, and year. For the data cleaning, only the unique articles were kept, but the fields of author and university were consolidated to be accurate with attributions.

Figure 1 and Figure 2 show the top ten publications by university and author over the research period, respectively.

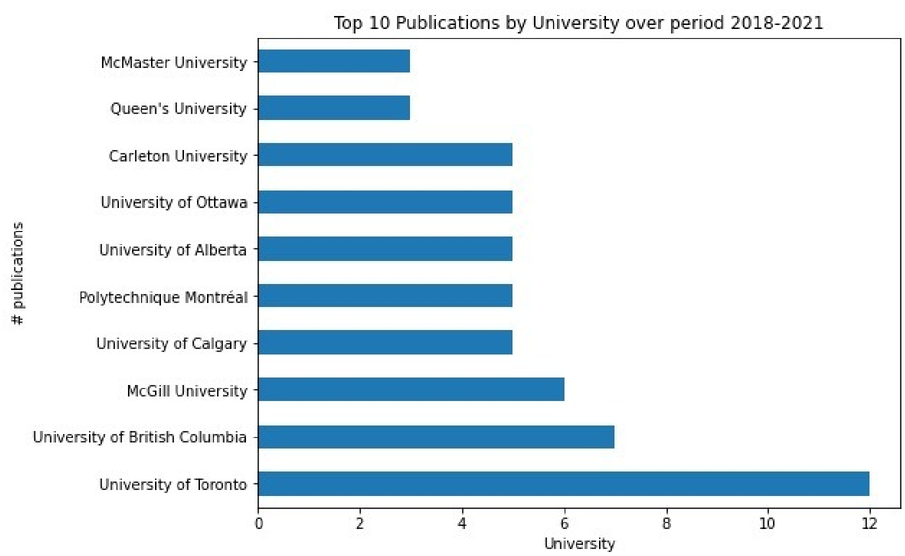

Figure 1. Top 10 publications by university

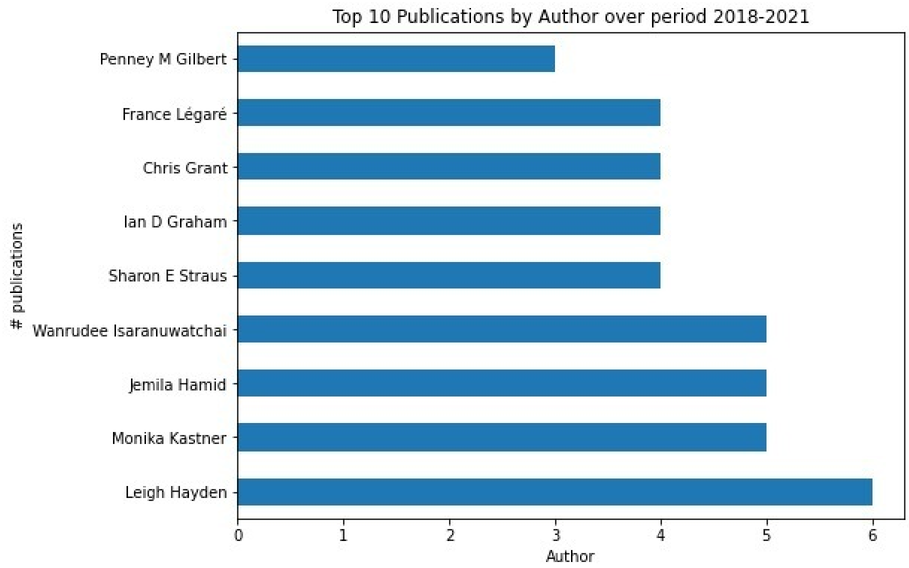

Figure 2. Top 10 publications by author

Table 2 describes the files related to the Academic paper scraping and cleaning process.

Table 2. Academic articles' data structure

Academia					
Original data	**Cleaned Data [.csv]**	**Cleaning code [.jpynb]**		**Count**	**File Size [kb]**
cognit_papers	cognit_papers_unique	CognitPages_Scraping		149	465
			Total	**149**	**465**

2.3 News websites

The following list shows the news websites scrapped. They were selected amongst the most relevant Canadian news bulletins that showed no restriction against web scraping in their use policy. Also, some other important sources that produce content relevant for the study, such as Seniors Advocate BC were included.

- https://www.seniorsadvocatebc.ca/reports/
- https://www.seniorsadvocatebc.ca/news-media/
- http://seniorsfirstbc.ca/news/
- https://www.vancouverislandfreedaily.com/tag/seniors/
- https://www.nanaimobulletin.com/tag/seniors/
- https://www.mcmasteroptimalaging.org/
- https://news.bchousing.org/?q=seniors&search.x=0&search.y=0
- https://helpagecanada.ca/news/
- https://www.aarpethel.com/

After scraping the data, the group analyzed the most frequent bigrams, a combination of two words. By doing so, we were able to identify and remove unwanted comments that might affect the result of the models. In addition, for some analyses such as word cloud and sentiment analysis, punctuations, words containing numbers, brackets, and square brackets were removed.

These steps were conducted before the preprocessing of the text data to assure that our data are free of words unrelated to the topic, such as: Follow us on Facebook, Twitter, PressLike, email addresses, and others. Table 3 describes the file structure, the original dataset, the cleaned version, and which jupyter notebook file contains the code for the cleaning.

Table 3. News websites' data structure

News				
Original data	**Cleaned Data [.csv]**	**Cleaning code [.jpynb]**	**Count**	**File Size [kb]**
seniors_advocate_bc_reports	seniors_advocate_bc_reports_cleaned	News_Cleaning_Seniors_Advocate_BC_Reports	19	21
seniors_advocate_bc_news	seniors_advocate_bc_news_cleaned	News_Cleaning_Seniors_Advocate_BC	58	110
seniors_first_bc_news	seniors_first_bc_news_cleaned	News_Cleaning_Seniors_First_BC	39	34
seniors_vancouverisland_seniors	seniors_vancouverisland_seniors_cleaned	News_Cleaning_VancouverIsland_Seniors	173	247
seniors_nanaimo_news	seniors_nanaimo_news_cleaned	News_Cleaning_Nanaimo_Seniors	152	226
mcmaster_blog_posts	mcmaster_blog_posts_cleaned	News_Cleaning_Mcmaster_Blog_Posts	379	617
bchousing_news	bchousing_news_cleaned	News_Cleaning_bchousing	280	2368
helpagecanada_news	helpage_news_cleaned	News_Cleaning_helpage	70	249
AARPEthel_articles	AARPEthel_articles_cleaned	Cleaning_AARP	172	893
		Total	**1,315**	**4,765**

The details on the steps used for removing the unwanted words can be found on the jupyter notebooks aforementioned.

3 Gathered data description

In total, we have:

- 58,716 tweets.
- 149 Academic articles.
- 1315 News publications.

However, after some detailed analysis, it was possible to identify duplicated article's bodies and News Publications' content thus they were removed, and the data were reduced to the following numbers:

- 58,716 tweets.
- 101 Academic articles
- 868 News publications.

4 Sentiment Analysis

The website lexalytics.com2 [10], states that *Sentiment Analysis is the process of determining whether a piece of writing is positive, negative, or neutral*. This process combines different machine learning techniques and Natural Language Processing (NLP) to assign scores to the topics, themes, and categories in a phrase or text.

Ronen Feldman3 defines sentiment analysis as: “the task of finding the opinions of authors about specific entities.” These entities can be documents, sentences, social media activity, reviews, etc. In addition, the author explains that these sentiments are valuable to companies to keep track of their products and services and how the users are reacting to them, in a positive or negative way, and take action accordingly.

For this project, the sentiments were extracted using the TextBlob python library and sentiment function. The function receives a string of text as input and returns the polarity score and the subjectivity.

According to TextBlob documentation [7]:

> *“The polarity score is a float within the range [-1.0, 1.0]. The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective.”*

The process was conducted with the different sources separately, as described in the following paragraphs.

4.1 Twitter

On the Twitter scraped data, a combination of TextBlob and Transformers pipeline functions was used to calculate each tweet's sentiment score and classify them as with positive or negative connotations.

Since tweets have a limited number of characters and usually are smaller texts compared to the other 2 sources of data (News and Academic Articles), the presence of strong polarized words, either positive or negative, has a significant impact on the overall sentiment of the analyzed text.

Another investigation conducted was regarding the essence of Twitter social media. Twitter is known for being significantly related to the personal opinion of individuals complaining and creating trending topics to demonstrate dissatisfaction or, sometimes, praising a good service or product. Figure 3 shows the polarity distribution of the 57.139 tweets gathered to clarify the previous assumption.

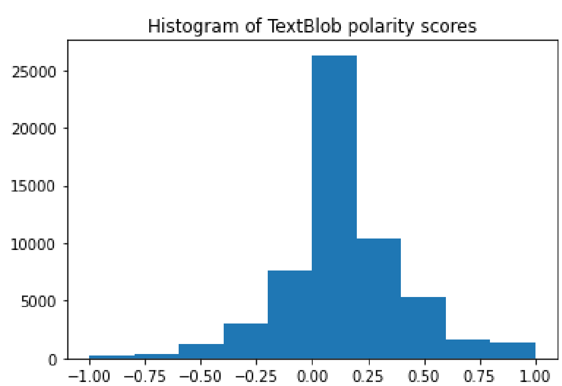

Figure 3. Tweets' TextBlob polarity histogram

Contrary to expectations, it is possible to conclude that most tweets are close to 0, representing a neutral sentiment. The average polarity presented was 0.1261.

To improve the results and interpretability, we applied a transformers-based sentiment analysis. The Hugging Face Transformer library [11] provides a built-in pipeline for the task. Figure 4 shows the result of the alternative, improved sentiment analysis. It is now possible to observe well-defined groups of positive and negative tweets.

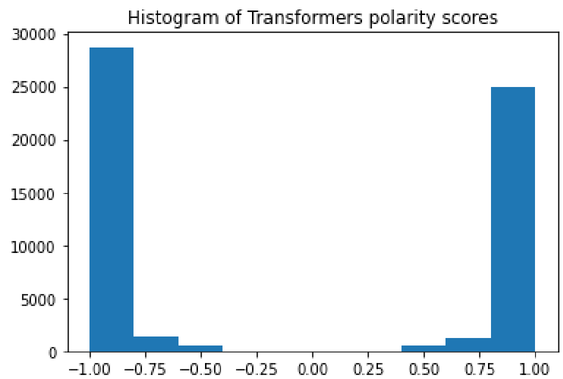

Figure 4. Tweets' Transformers polarity histogram

4.2 Academic articles

Academic Articles had their sentiment scores calculated and assigned alongside each observation. Then a detailed analysis of the negative sentiment articles was conducted. Their titles and scores are described in Table 4.

Table 4. Negative sentiment articles

Title	Polarity	Length
COVID-19 Physical Distancing and Post-Traumatic Stress Injury: Utilization of Digital Health and Remote Mental Health Services for Military, Veterans, and Public Safety Personnel (2020/21)	-0.0177	1976
Extending the validation and reliability testing of SCREEN a nutrition risk screening tool for community-living seniors. (2020/21)	-0.0148	1974
A collaborative evaluation of the Iiyuu Ahtaawin Miyupimaatisiiun Planning (IAMP) Initiative for Cree Health (2015/16-2018/19)	-0.0119	3424
Canadian Frailty Network - Technology Evaluation in the Elderly Network (TVN) (2017/18-2020/21)	-0.0095	1602

Looking into the titles above, we can clearly identify words with negative polarity, such as Traumatic Stress Injury and nutrition risk. In addition, the length of the text is not short enough for us to conclude that the presence of some negative words would reduce the overall score. Thus, it means that, on average, the whole text had a negative connotation of words.

Although the negative values, the polarity scores were close to the 0-mark, neutral speech, which is the expected tone of speech on academic documents. The presented polarity average score for all academic articles was 0.1159.

4.3 News websites

For the scraped News, the polarity of the body contents was appended to the dictionary to its respective News texts. Additionally, the score is plotted on a timeline to verify how it spread among all the News gathered. The result is illustrated in Figure 5.

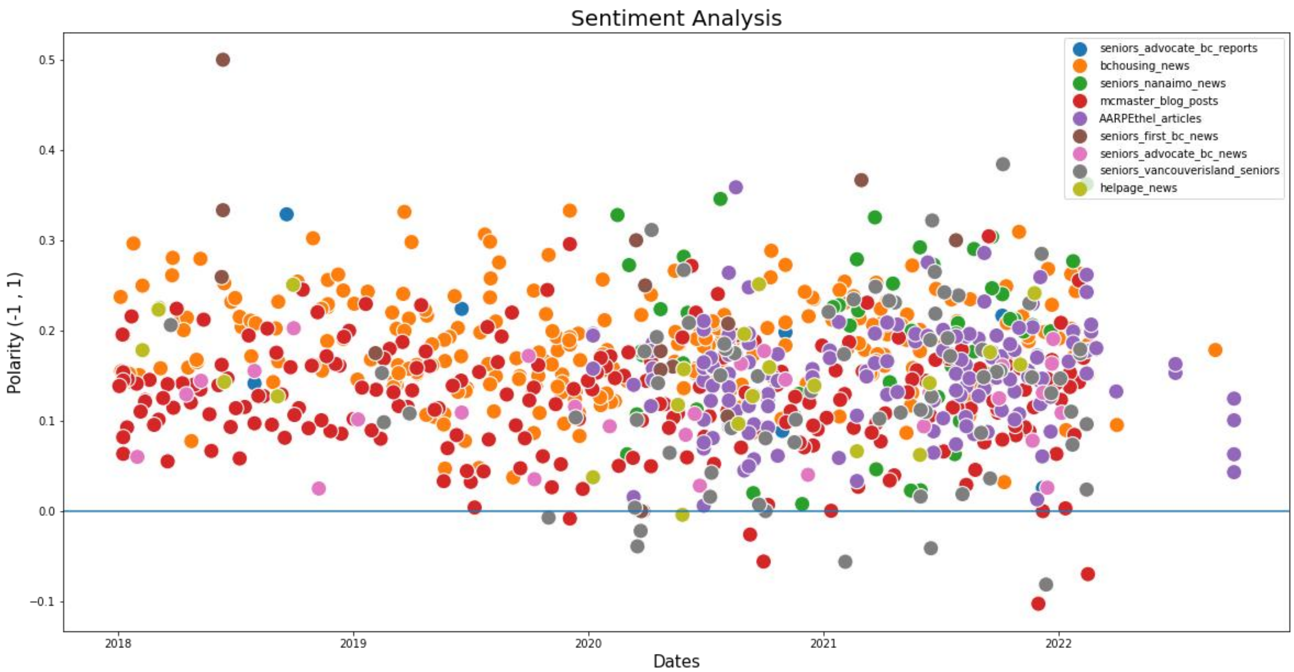

Figure 5. News Sentiment Analysis over years

The average polarity was 0.1507, and after 2019 some negative values started to appear. The graph highlighted outliers and negative polarity News that were further investigated. The chart also revealed an increasing number of reports with spreader polarity in between over the years.

The News with the highest polarity contained the following content:

"June 15 is World Elder Abuse Awareness Day (WEAAD). Over 50 WEAAD events will take place across BC – Find Out More WEAR PURPLE to show your support!"

This outlier highlighted the possibility of having no detailed data regarding Elder-Wellbeing. However, after checking the news with the shortest texts (less than 300 words) the group decided to keep all of them since they were placeholders for reports containing important information for Elders, described in Table 5.

Table 5. Short texts analysis

Body	**Polarity**	**Length**
June 15 is World Elder Abuse Awareness Day (WEAAD) Over 50 WEAAD events will take place across BC –Find Out More WEAR PURPLE to show your support!	0.500	147
Seniors' poverty in British Columbia is the highest rate in the country, based on latest data from Statistics Canada (2015). Read more online –B.C. Seniors' Poverty Report Card	0.333	177
Seniors First BC recently released its Annual Report for 2020-2021. The report provides updates and statistics about its programs during this period. To read the report, visit here.	0.125	180
Featuring: Legal Staff – Seniors First BC Victim Services Worker – Family Services of Greater Vancouver Public Guardian and Trustee of BC Designated Agency – Fraser Health Watch online	0.175	186
Our Seniors First BC COVID-19 resource list gives all the information for risks, the number of cases, symptoms, precautions and where to get help during this crisis. COVID-19 RESOURCES UPDATED May 15, 2020.	0.250	207
Find all the resources announced by the Province of BC regarding COVID-19 situation, it gives information on how to protect yourself, your family and your community and the steps to take if you suspect to have the virus. COVID-19 Provincial Support and Information.	0	264
Vaccine call centres for seniors open March 8, 2021. You can book a vaccine appointment for yourself or your spouse. It's easy and safe to book over the phone. Our SAIL staff can assist as well. Call us at Call: 604-437-1940 or Toll Free: 1-866-437-1940 Click here for more details:	0.367	282

The News with negative polarity was also further analyzed. Those content were related to grief, kill, death, scam, failure, illness, etc. Words with a strong negative sentiment reduced the overall score.

The websites were also analyzed separately to identify speech patterns and polarity of sentiment. However, on average, every news media presented a positive polarity.

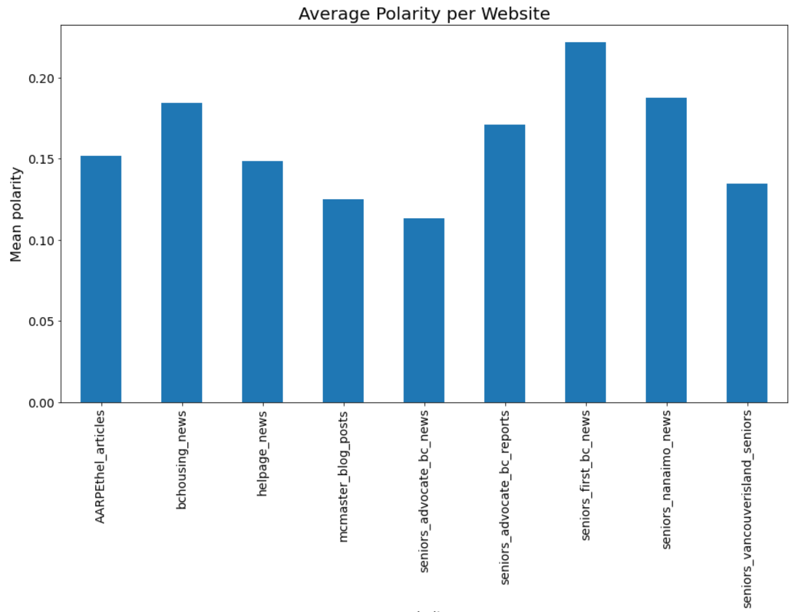

Figure 6. Average polarity per News website

Furthermore, the websites were plotted in different timeline charts. For example, McMaster and Vancouver Island Seniors were the ones that presented news with negative sentiment, as illustrated in Figure 7 and Figure 8.

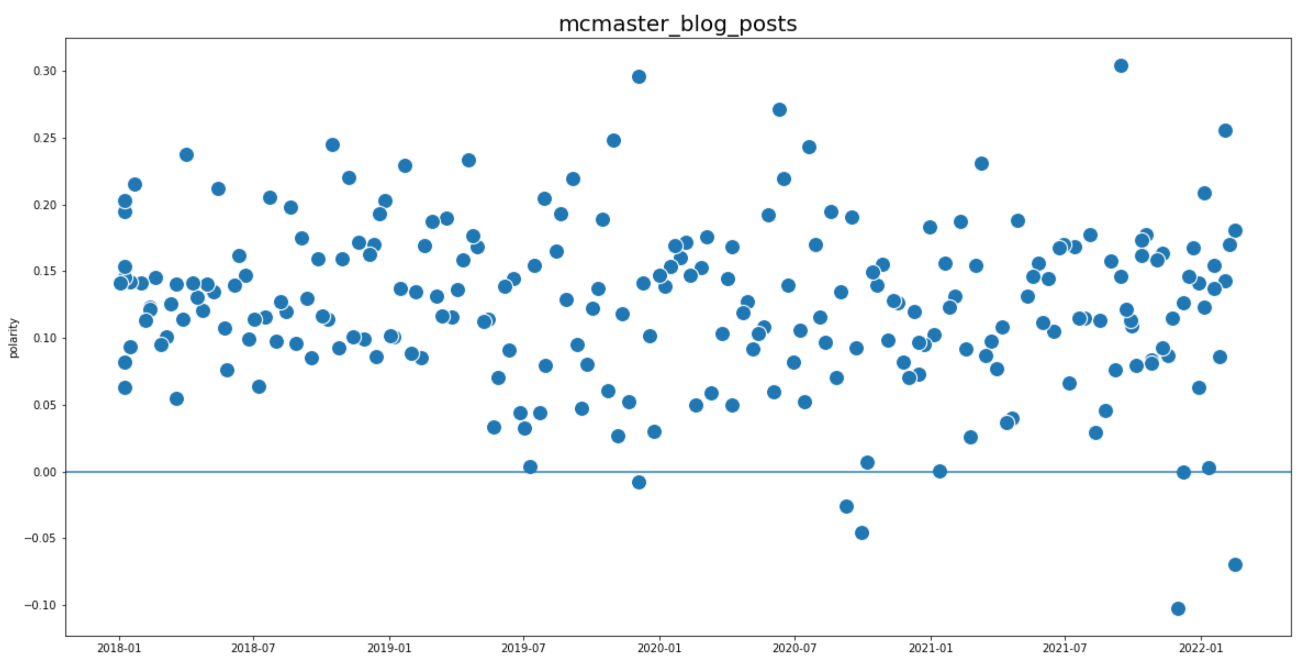

Figure 7. McMaster Blog sentiment score timeline

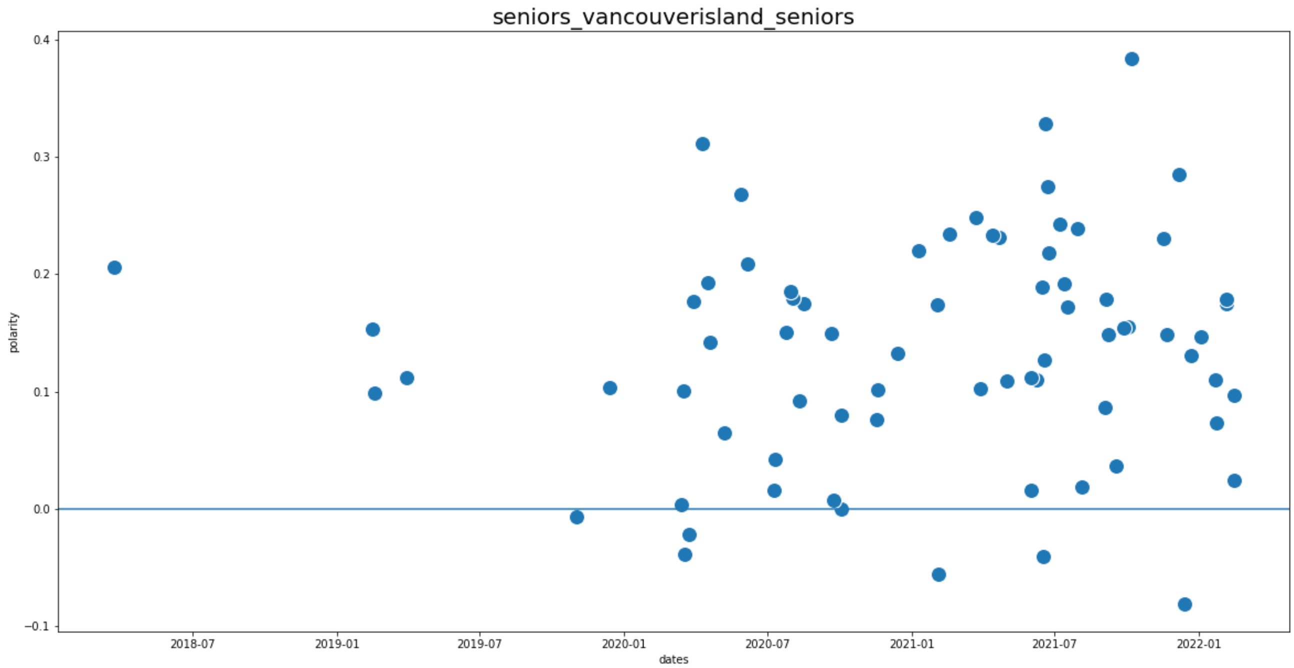

Figure 8. Vancouver Island Seniors sentiment score timeline

Again, it is possible to identify the increasing trend in the amount of news as the years passed.

5 Word cloud

A word cloud is a visual representation of word frequency. The more commonly the term appears within the text, the larger the word appears in the image generated. Word clouds are increasingly being employed as a simple tool to identify the focus of written material. Hence, we wanted to understand what the focus of our cleaned text looked like for the different sources.

Figure 9. Word Cloud for all Academic articles

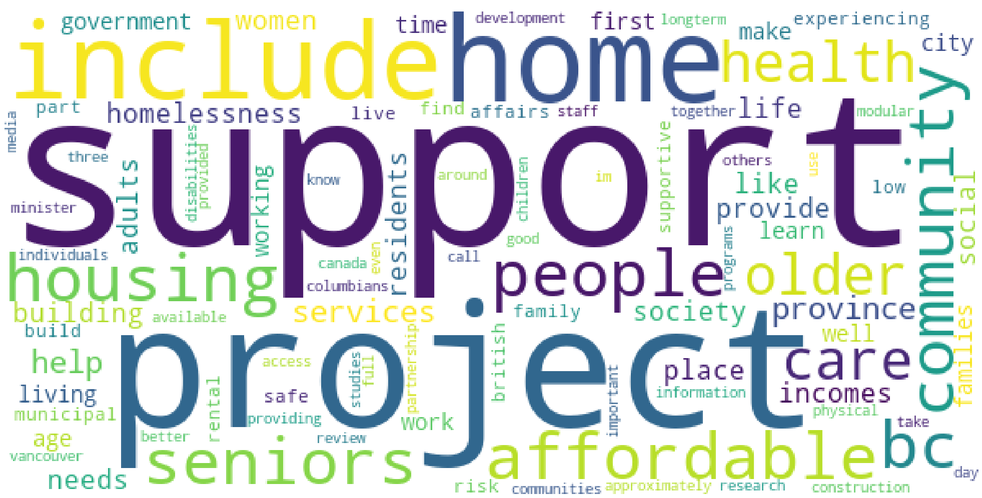

Figure 10. Word Cloud for all News combined

Figure 9 and Figure 10 are the word clouds output we obtained from the clean contents of NEWS and Academic articles. It is possible to observe that the content is majorly focused on support for the elderly, Projects related to the elderly, inclusion, home, affordability, housing aspects, health, community, etc. The size of the words signifies the frequency in which the topic was discussed. (Colors in the picture are just for beautification. The color does not indicate anything.)

By combining the word cloud from multiple sources with Topic Modelling and Sentiment analysis, we created a dashboard that clearly illustrated the concerns for elders per period, described in detail in section 7.

6 Topic Modeling

Before building a model based on the text scrapped from twitter, news and academic articles, a preprocessing task was conducted to eliminate "stop words" and apply lemmatization and stemming on the corpus.

When analyzing text with NLP processes, "stop words" are often eliminated. These words include articles, some connectors, and everyday words/expressions.

Stemming is a popular way to minimize the vocabulary size in NLP tasks as this process converts words with the same "stem" or root to a single word. Stemming can be based on rules or based on context. Due to the amount of data and the length of the text – especially in the case of Twitter – in this project, rules-based treatment was selected.

Lemmatization usually refers to doing things properly using a vocabulary and morphological analysis of words, usually aiming to remove inflectional endings only and return the base or dictionary form of a word, known as the "lemma" [8].

To perform the semantic similarity detection and calculate the most relevant topics in the project's corpus, several options were considered, but BERT was selected per the results obtained.

6.1 BERT topic modeling

An acronym for Bidirectional Encoder Representations from Transformers, *BERT is a language representation model. A pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications* [3].

In essence, using a pre-trained language model allows us to get closer to the message intent of each text body. The representation of each tweet, academic article, or news page can be embedded in a vector space that captures the semantics and sentence structures of the text.

The model created using BERT provided a range of distinct topics, relevant and irrelevant for our research.

Table 6. Top ten topics for Twitter corpus

Topic	Count	Name
-1	30633	-1_and_to_the_of
0	1032	0_mens_senior_championship_team
1	655	1_citizen_senior_seniors_citizens
2	405	2_housing_affordable_rental_units
3	388	3_pharmacare_drugs_prescriptions_drug
4	350	4_dog_cat_cats_dogs
5	344	5_trudeau_trudeaus_justin_harper
6	325	6_christmas_stockingstuffersforseniors_gifts_s...
7	303	7_music_album_songs_song
8	301	8_abuse_awareness_elder_elderabuse

Table 6 shows the top ten topics for Twitter retrieved from BERT built-in architecture in Huggin Face Transformers library [11]. Topic enumerated as -1 contains the observations not included in the other groups.

Table 7. Topics for Academic articles corpus

Topic	Count	Name
-1	49	-1_indigen_communiti_research_health
0	28	0_communiti_research_indigen_elder
1	24	1_care_patient_hospit_studi

Table 8. Top ten topics for News corpus

Topic	Count	Name
0	303	0_hous_home_afford_peopl
-1	208	-1_care_health_senior_older
1	151	1_year_like_hair_love
2	69	2_blood_diet_diabet_heart
3	62	3_senior_care_report_advoc

4	58	4_older_social_adult_care
5	39	5_vaccin_covid_outbreak_staff
6	35	6_walk_exercis_fall_speed
7	28	7_cancer_smoke_quit_screen
8	26	8_pain_exercis_pillow_treatment

Figure 11 illustrate the topics captured from our dataset. In the case of Twitter, topic 2 is built around the “housing”, “affordable”, “rental”, “units”, and “rent”, thus is easy to see this subset of data is about the housing conditions affecting Elders. On the Other hand, the topic 0 contains the words “men”, “senior”, “championship”, “team”, “game” which might mean that the subset of tweets is related to some sports competition, hence it is off-topic and is not related to our Project’s topic.

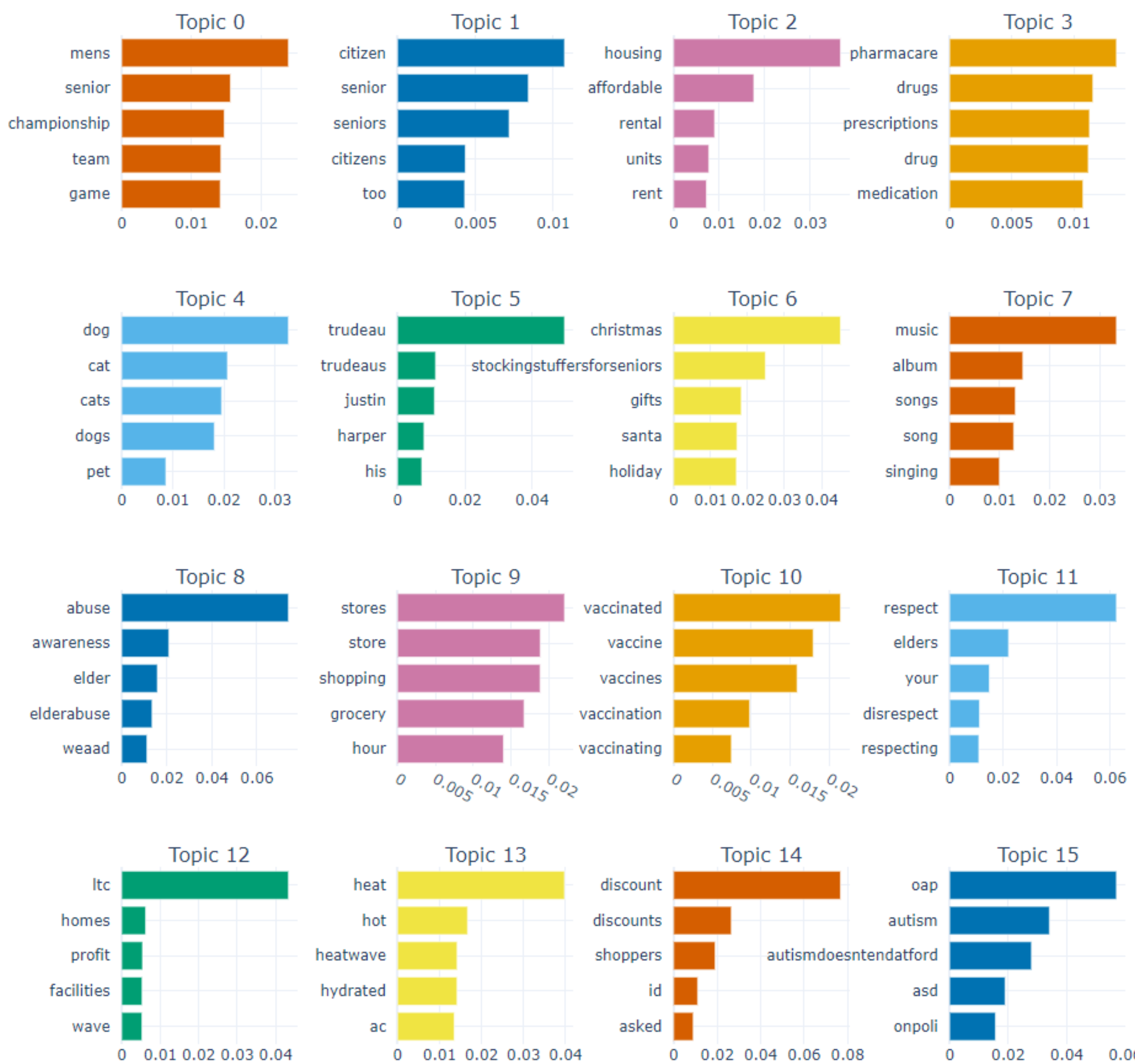

Figure 11. Tweets topic word scores

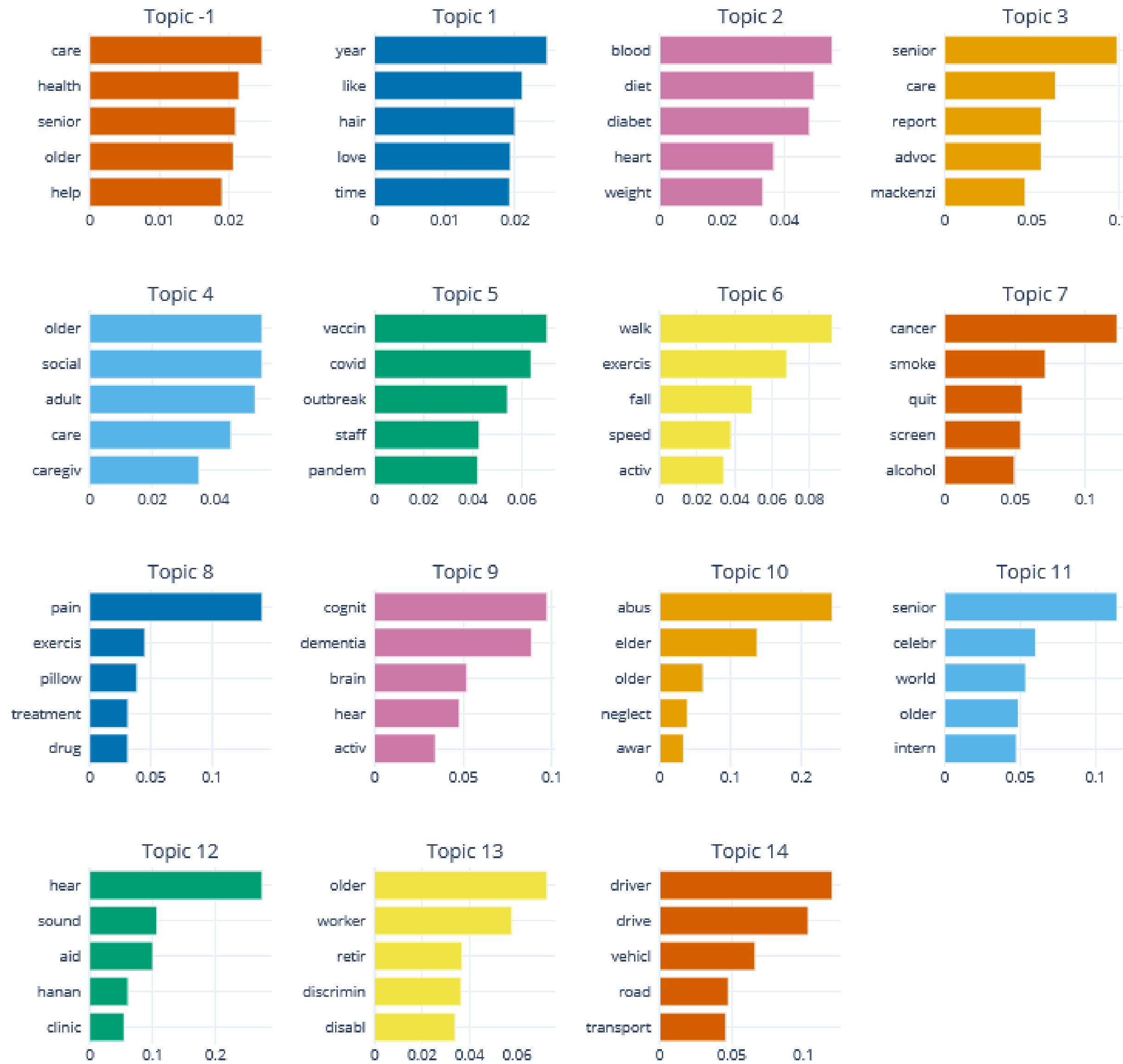

Figure 12. News topic word scores

In terms of the topics from the news media analysis, Figure 12 shows that, for instance, topic 0 is built around the concept of "care", "health", "senior", "older", and "help", making clear that those news articles are related to the health care.

6.2 Topics refinement

To refine the results from topic modeling with BERT, we adopted an approach that could consider as input the aspects relevant to our research study. Therefore, the topics described in the previous section were aggregated by similarity to 19 parent topics: Covid, Health, House, Homecare, Money, Volunteer, Safety, Abuse, Physical Activity, Social Activity, First Nations, Schooling, Environment, Autonomy, Freedom, Climate, Temperature, Vaccine, and Grocery. The final results will be explored in the following sections.

7 Dashboard and results

By combining the techniques explained in the previous section, the group created a Tableau Dashboard capable of capturing the sentiment and the frequency of the topics on all the media sources studied in a monthly manner. However, since the academic articles presented only two major topics on its corpus, they were left aside from the dashboard analysis.

The interactive capabilities of the tool assisted in the discovery of patterns and in featuring the major concerns faced by the elderly society. Figure 13 shows the structure of the dashboard, and Figure 14 illustrates the final version.

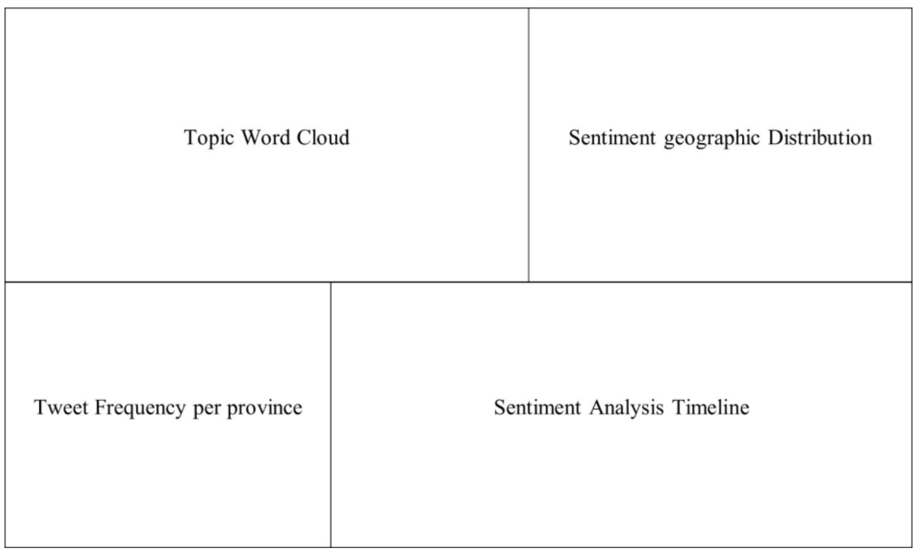

Figure 13. Dashboard structure

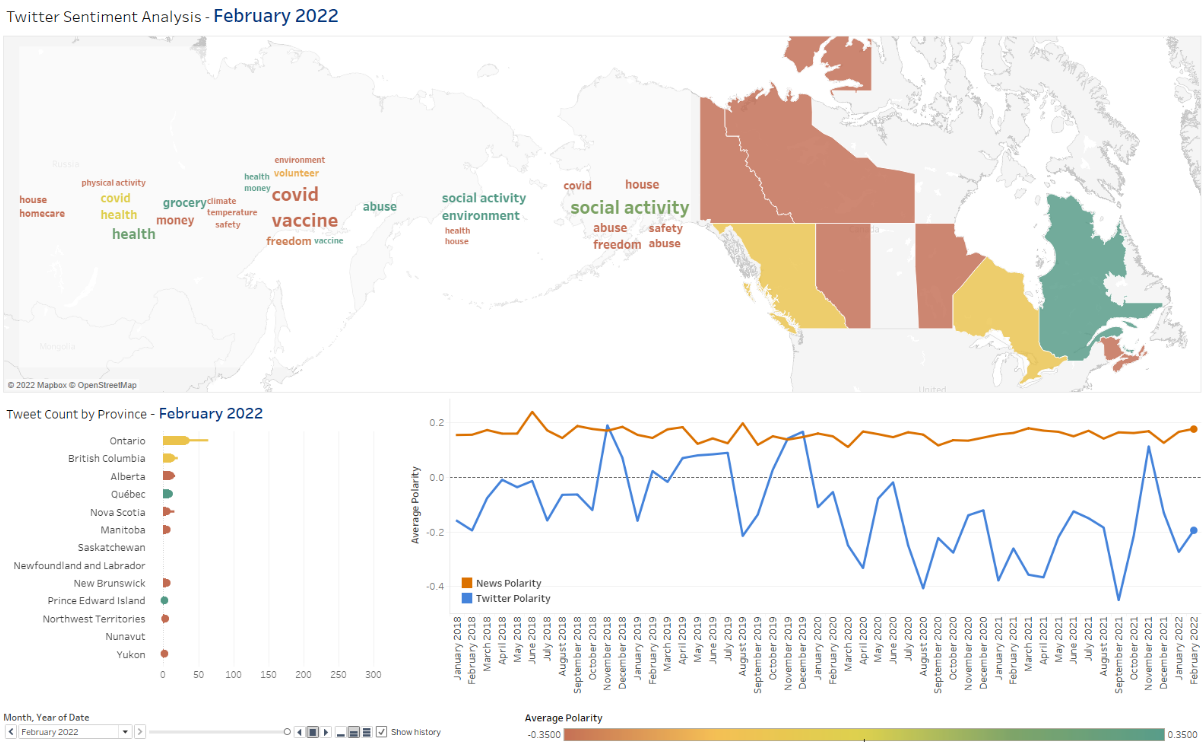

Figure 14. Tableau dashboard

In this case, the colors represent the polarity of the sentiments, where red means negative, and green, positive.

It is possible to observe the difference variation between tweets and news bulletins as expected News websites presented smaller variations in polarity, while Twitter showed great ups and downs during the period.

By analyzing that variation, it was possible to investigate what were the topics mentioned on those periods of major positivity and negativity and contextualize them with the world events. Some were described in the following images.

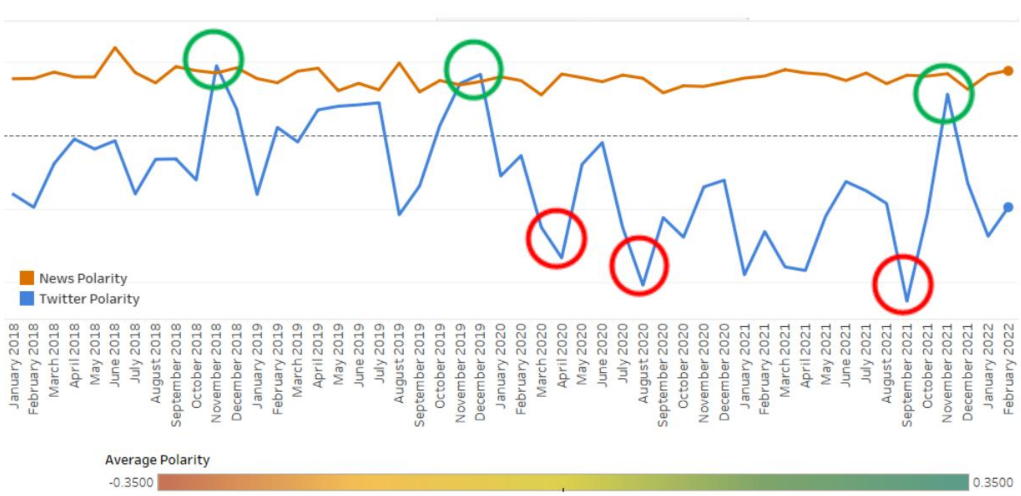

Figure 15. Periods with greater and lesser polarity

The next group of images represents the topics present in the Highest polarity periods.

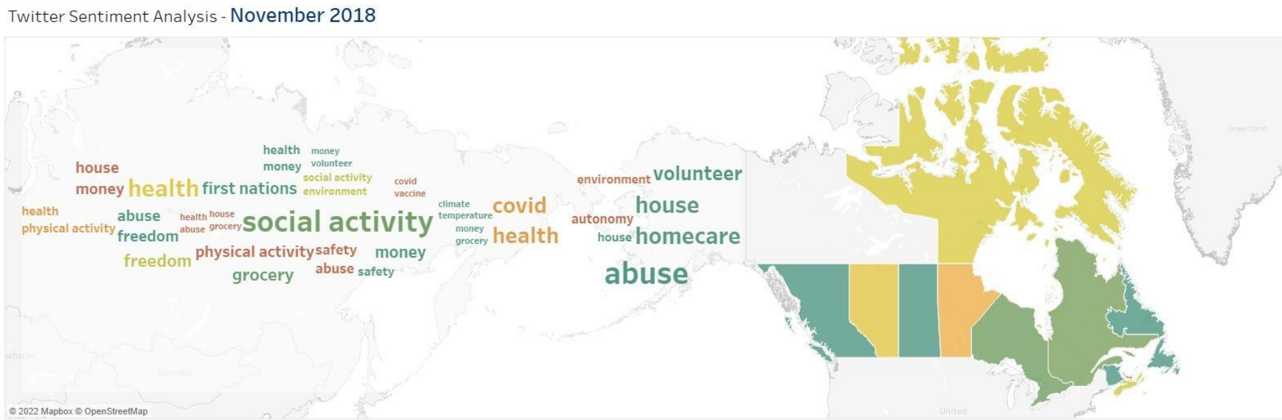

Figure 16. November 2018 Word cloud snapshot

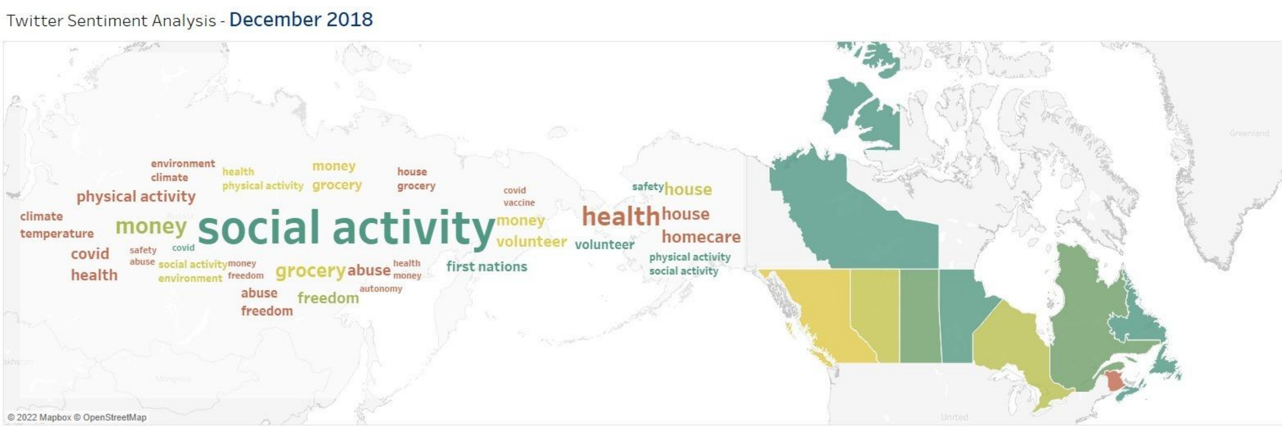

Figure 17. December 2018 Word cloud snapshot

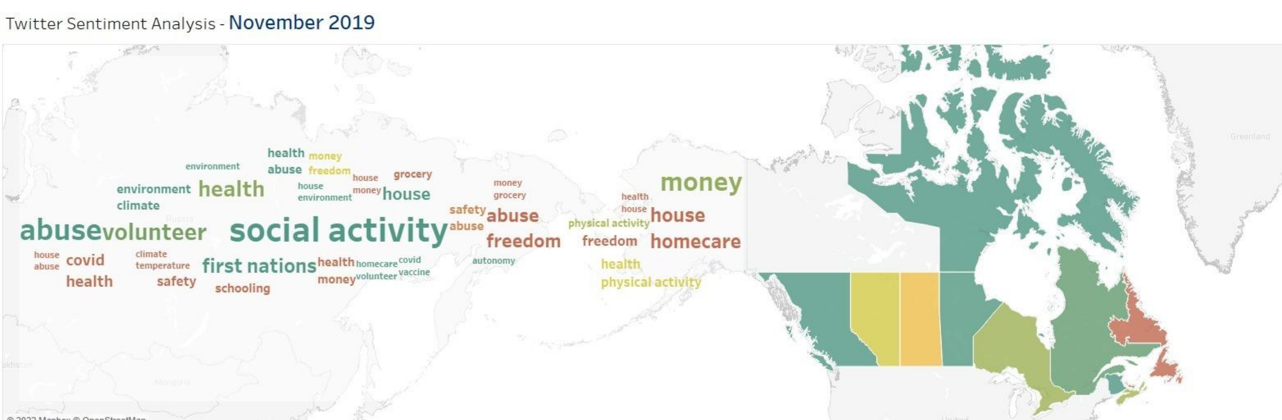

Figure 18. November 2019 Word cloud snapshot

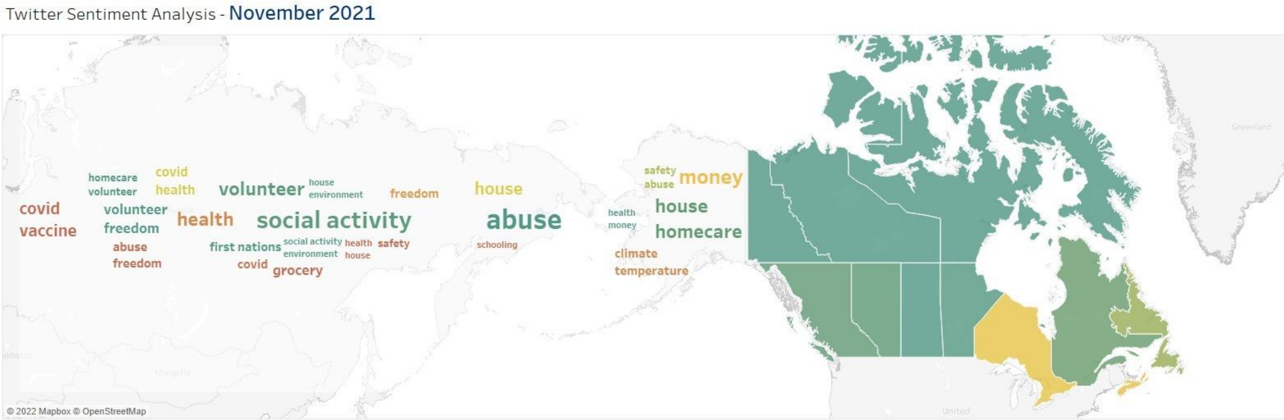

Figure 19. November 2021 Word cloud snapshot

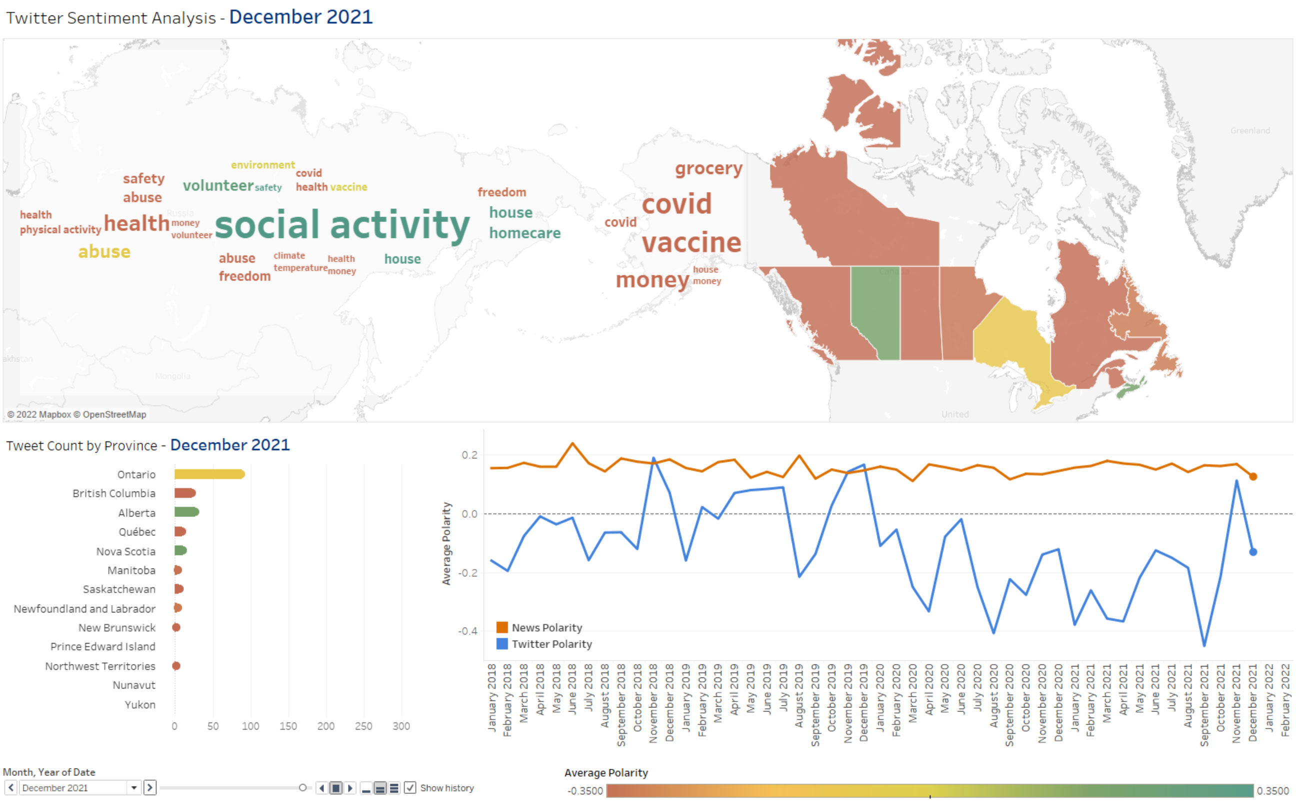

Figure 20. December 2021 snapshot

November and December had the social activity as the main topic with extreme positive polarity. This is due to the festivities and holidays, a period when there is a lot of social interaction when families gather around and share good times. When intergenerational relationships are at their peak, seniors and those who advocate for them say positive things.

The following images represent some of the most negative sentiments of the period.

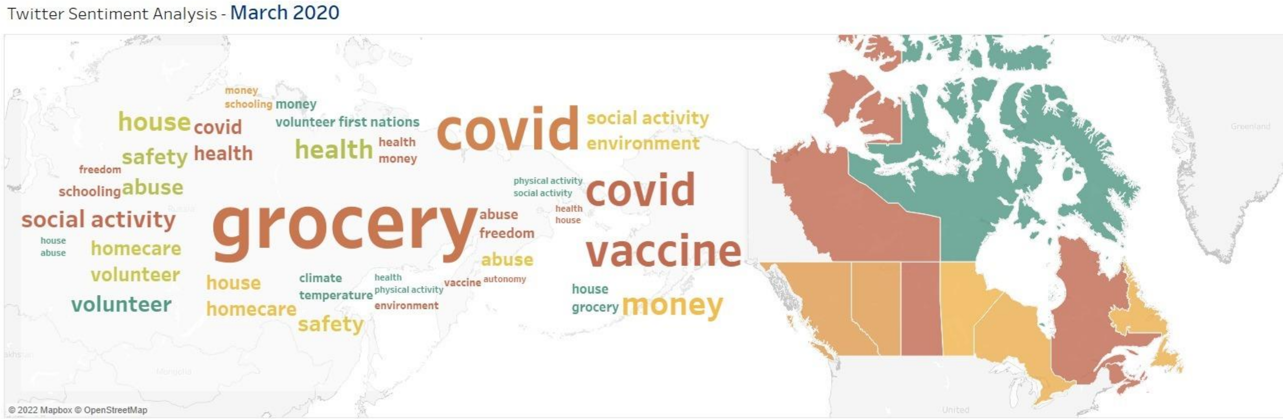

Figure 21. March 2020 Word cloud snapshot

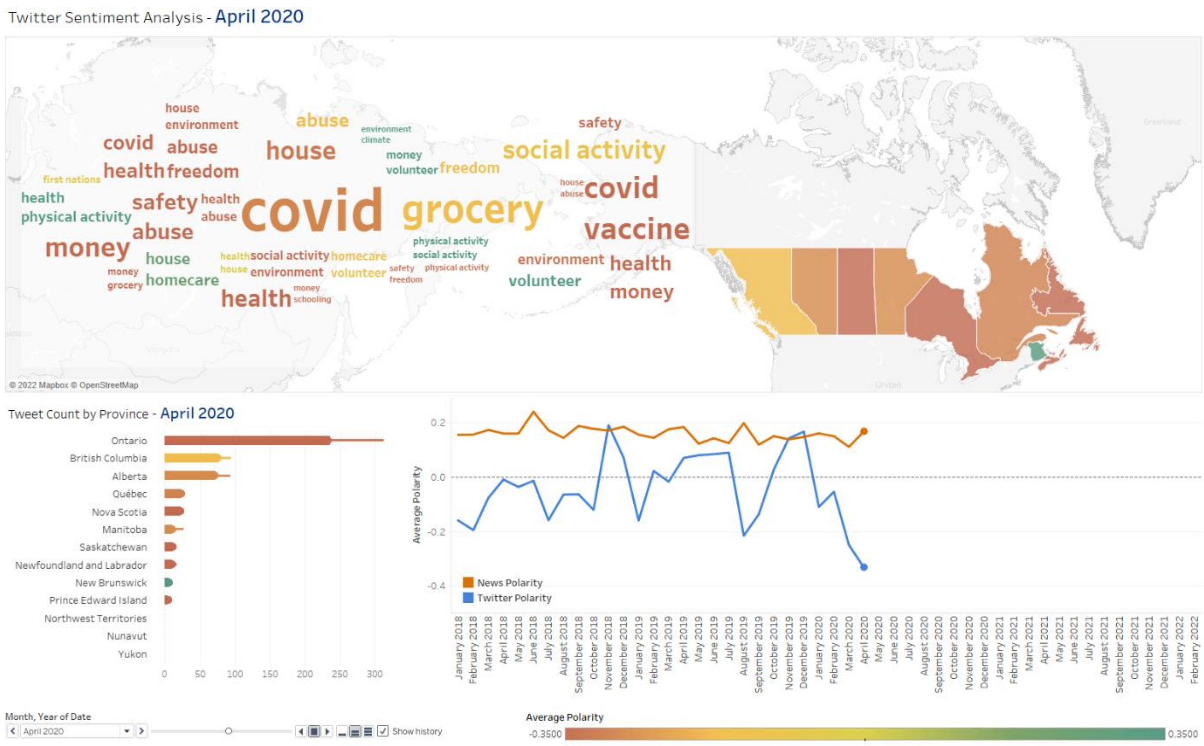

Figure 22. April 2020 snapshot

These months were when the first wave of COVID-19 happened. As we can see, the most mentioned topic on March 2020 was not the Pandemic itself but a concern related to groceries. Once the lockdown was in place, how elders would have access to food and essential products was a major discussion.

A month later, the COVID-19 took the lead on most discussed problems among the senior society, and grocery became the second most frequent topic, however, with a more neutral (yellow color) sentiment.

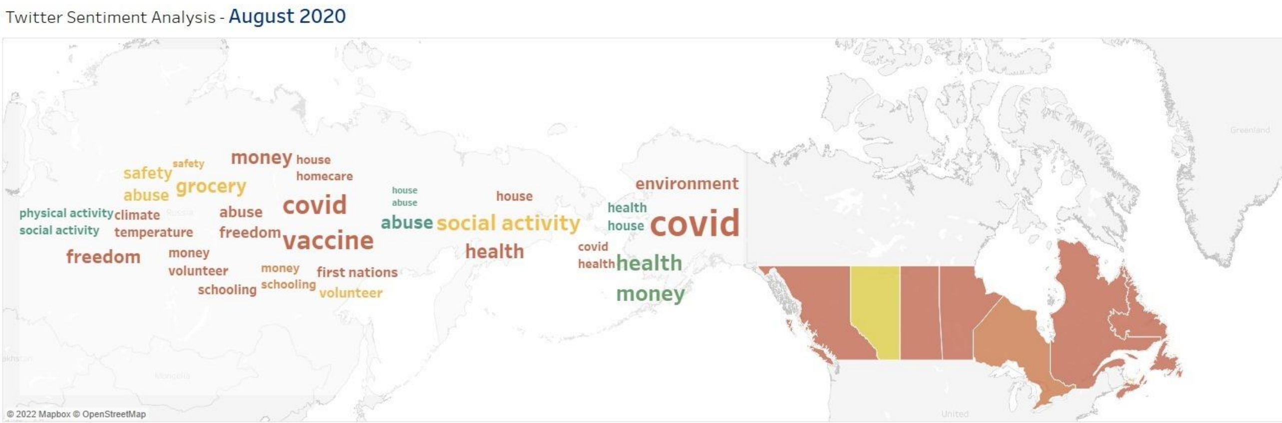

Figure 23. August 2020 Word cloud snapshot

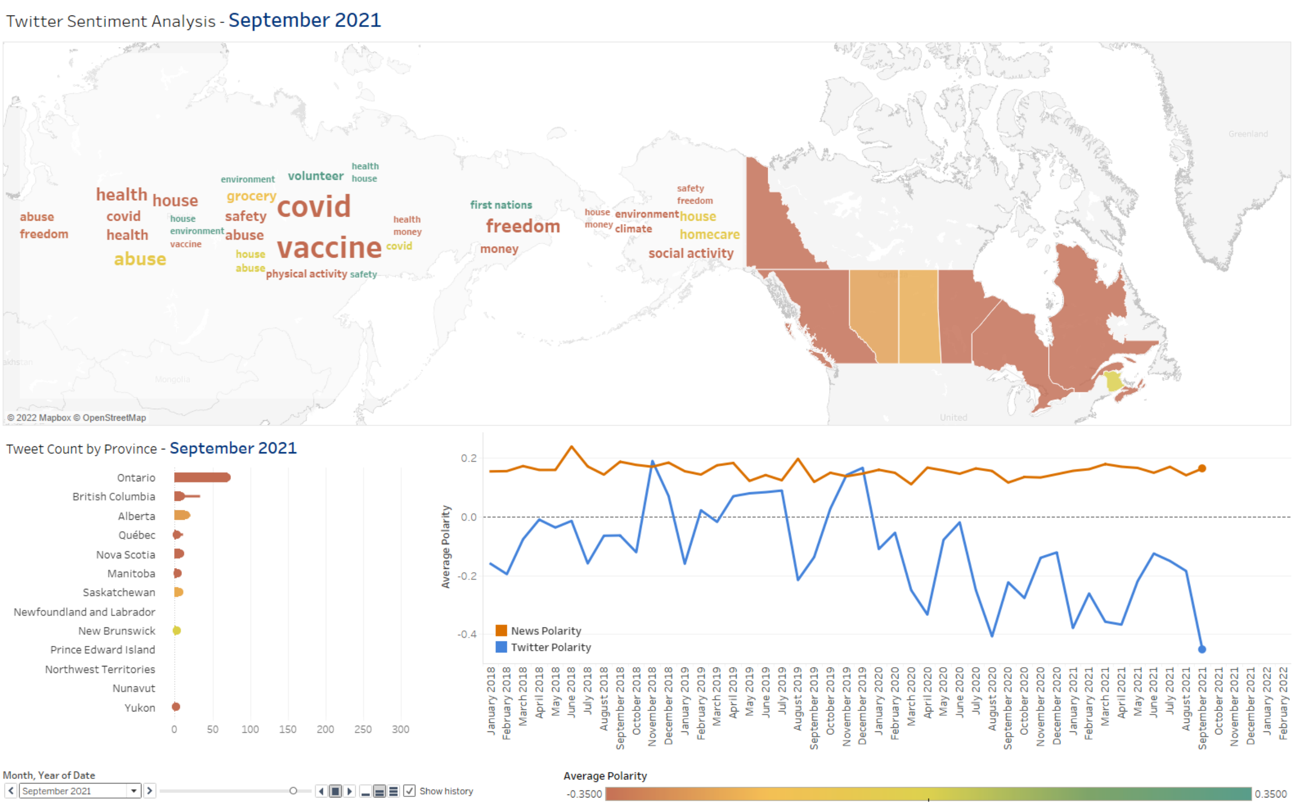

Figure 24. September 2021 snapshot

Both August 2020, and September 2021, also presented the Covid vaccine and COVID as a big concern, along with other topics. Expected results during the two pandemic years.

Another problem easy to identify using the dashboard was those related to climate change and temperature. For example, the following image exhibits the concern related to the temperatures during the summer of July 2020.

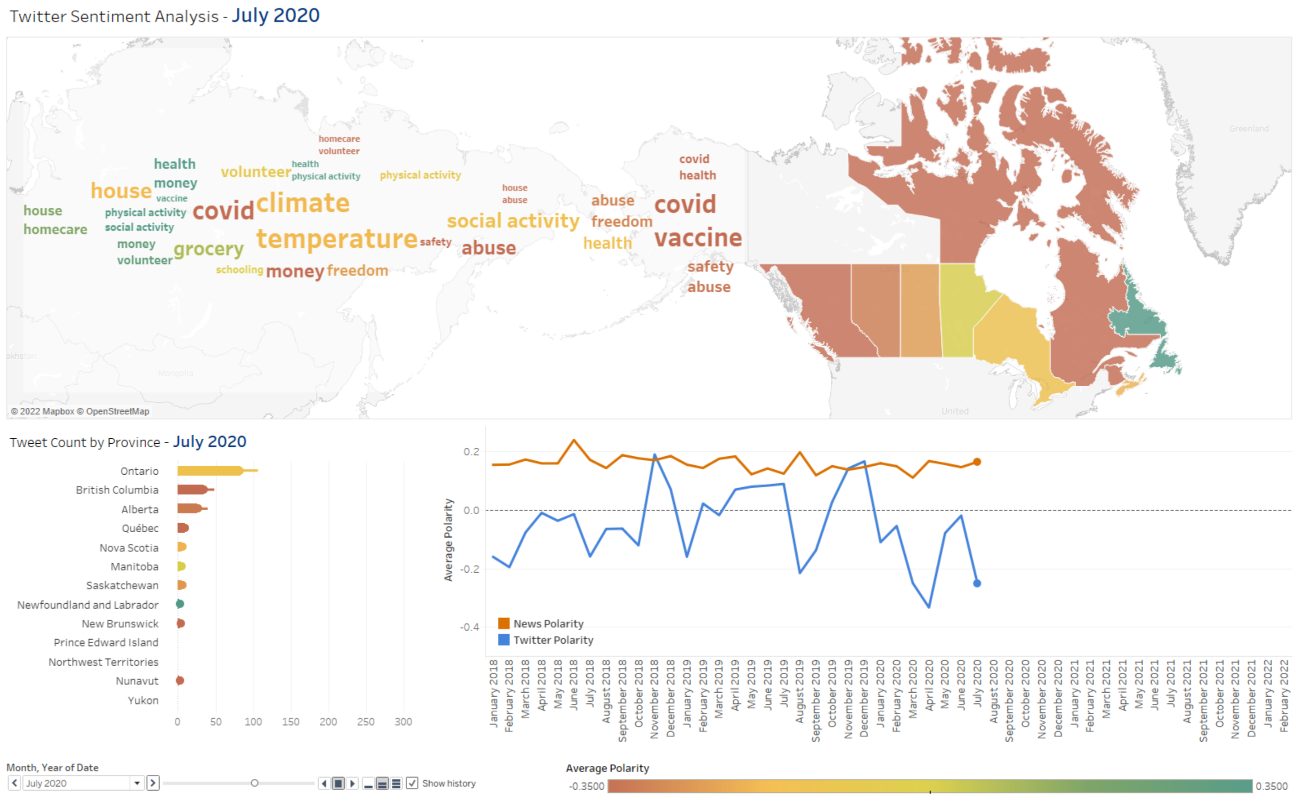

Figure 25. July 2020 snapshot

Although some data are easy to contextualize, others are not. For example, the following three images represent other negative polarity periods when the sentiment was spread around different topics.

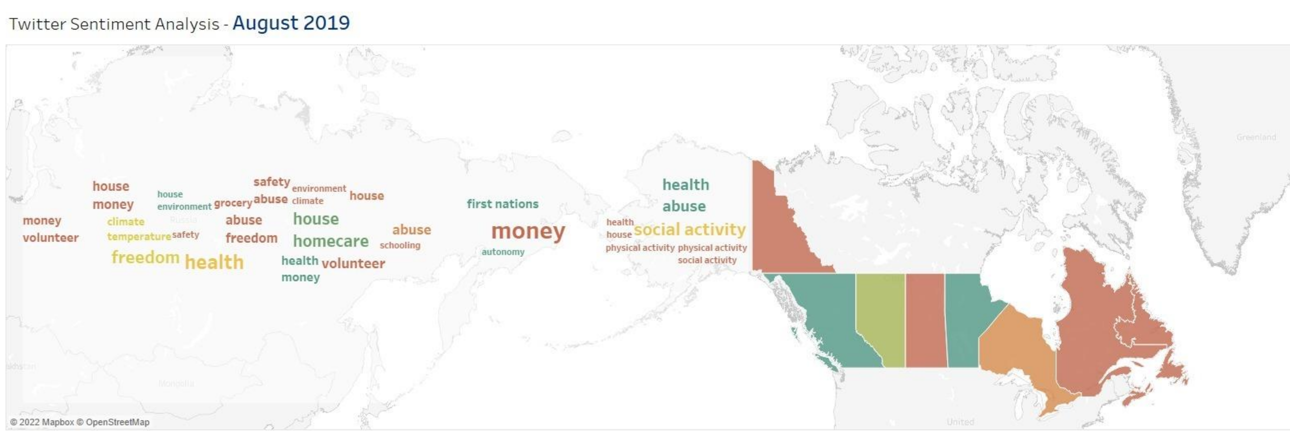

Figure 26. August 2019 Word cloud snapshot

In August 2019, the discussions were about money, health, social activity, and freedom, however, it's not clear what major event happened during that period.

There were other scenarios with low sentiments where the leading cause was not clearly defined. For those periods, a more detailed contextualization is required. Again, crosschecking News and other information sources would help identify what happened.

8 Recommendations and next steps

To continue this analysis, we recommend combining sentiment analysis, word cloud, and topic modeling techniques to further contextualize the topics highlighted in the dashboard. In addition, gathering more Academia material would be valuable for comparison with the other sources and to improve Academia's topic modeling results. Another possible approach is identifying government and other institutions' measures and responses to the featured problems and how they increased overall sentiment.

Furthermore, classify the Tweets according to related and non-related to elderly-wellbeing to even increase the perception among the target audience. For this specific process, we recommend some labeling techniques with clustering analysis over the 57 thousand tweets, using K-means to identify the relevant observations and propagate the label around the clusters defined. This is an effective, yet faster labeling method tested on a different dataset and resulted in meaningful accuracy increases over classification models. The details are provided in the paper Clustering for Semi-Supervised Learning on Disasters Tweets file in the submission folder.

9 Acknowledgement

The team would like to wholeheartedly thank Jessica Blaquiere and Karan Bhardwaj for their support and our instructor Quynh Monica Nguyen, for the opportunity to work on a transformative project.

This work received invaluable advice and domain-knowledge input from Colleen Fraser and Bonnie Milne.

REFERENCES

[1] Bird, Steven, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. *NLTK*. O'Reilly Media, 2016. https://www.nltk.org/book/.

[2] Cudmore, Cora-Lee. “2018 World Elder Abuse Awareness Day.” Seniors First BC, June 13, 2018. https://seniorsfirstbc.ca/news/2018-world-elder-abuse-awareness-day/.

[3] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” arXiv.org, May 24, 2019. https://arxiv.org/abs/1810.04805.

[4] Feldman, Ronen. “Techniques and Applications for Sentiment Analysis.” ACM, April 1, 2013. https://cacm.acm.org/magazines/2013/4/162501-techniques-and-applications-for-sentiment-analysis/fulltext.

[5] “Gensim: Topic Modelling for Humans.” Topics and Transformations - gensim, December 22, 2021. https://radimrehurek.com/gensim/auto_examples/core/run_topics_and_transformations.html#sphx-glr-auto-examples-core-run-topics-and-transformations-py.

[6] Grootendorst, Maarten P. “Bertopic.” BERTopic. Accessed April 18, 2022. https://maartengr.github.io/BERTopic/index.html.

[7] Loria, Steven. “Tutorial: Quickstart.” Tutorial: Quickstart - TextBlob 0.16.0 documentation, 2020. https://textblob.readthedocs.io/en/dev/quickstart.html.

[8] Manning, Christopher – Raghavan, Prabhakar – Schütze, Hinrich. “Introduction to Information Retrieval”, Cambridge University Press, 2008. https://nlp.stanford.edu/IR-book/html/htmledition/tf-idf-weighting-1.html

[9] Ruggeri, K., Garcia-Garzon, E., Maguire, Á., Matz, S. and Huppert, F.A. (2020), “Well-being is more than happiness and life satisfaction: A multidimensional analysis of 21 countries”, Health and Quality of Life Outcomes, BioMed Central Ltd., Vol. 18 No. 1, pp. 1–16.

[10] “Sentiment Analysis Explained”. 2022. Lexalytics. https://www.lexalytics.com/technology/sentiment-analysis

[11] “Transformers”. Accessed April 18, 2022. https://huggingface.co/docs/transformers/index.

[12] “Twitter API Documentation | Docs | Twitter Developer Platform.” Twitter. Twitter. Accessed April 18, 2022. https://developer.twitter.com/en/docs/twitter-api.

APPENDIX A

Primary Key Words	Synonyms - Secondary Key Words	Antonyms
Learning	Scholarly knowledge	Unlearn
Learning	Scholarship	Neglect
Learning	Erudition	Ignorance
Learning	Schooling	Lack of knowledge
Learning	Study	Noncognitive process
Learning	Acquisition	Loss
Learning	Information	Disinterest
Learning	Wisdom	Stupidity
Learning	Educated	Uneducated
Learning	Cognitive process	Learning disability
Social Support	Human society	Lonesomeness
Social Support	Communities	Isolations
Social Support	Societal	Outcast
Social Support	Sociable	Unsocial
Social Support	Ethnic	Introvert
Social Support	Ethnical	Lonely
Social Support	Cultural	Segregate
Social Support	Sociality	Noncultural
Social Support	Friendly	Social exclusion
Social Support	Interpersonal	Unfriendly
Social Support	Friendly companion	Antisocial
Respect	Admiration	Disapproval
Respect	Deference	Defiance
Respect	Respectfulness	Disrespect
Respect	Courtesy	Impoliteness
Respect	Obedience	Rudeness
Respect	Ovation	Misbehavior
Respect	Veneration	Detest
Respect	Worship	Despise
Respect	Obeisance	Inconsideration
Respect	Esteem	Bad behavior
Respect	Regard	Casual attitude
Respect	Value	Scorn
Respect	Prize	Disdain
Respect	Appreciation	Disgust
Respect	Honor	Hate
Respect	Consideration	Dislike
Respect	Curteous	Antipathy

Respect	Behavior	Misconduct
Freedom	From duty	Captivity
Freedom	Free	Limitation
Freedom	Liberty	Restraint
Freedom	Unrestraint	Restriction
Freedom	Opportunity	Suppression
Freedom	Power	Reserve
Freedom	Privilege	Dependence
Freedom	Right	Constraint
Freedom	Exemption	Permanence
Freedom	Flexibility	Immovableness
Freedom	Immunity	Vulnerability
Freedom	Mobility	Ban
Freedom	Being free	Disallowance
Helping Strangers	Amend	Discourage
Helping Strangers	Assist	Hurt
Helping Strangers	Better	Harm
Helping Strangers	Assistance	Frustrate
Helping Strangers	Improve	Discouragement
Helping Strangers	Facilitate	Prevent
Helping Strangers	Ameliorate	Complicate
Helping Strangers	Encourage	Delay
Helping Strangers	Boost	Block
Helping Strangers	Fulfillment	Inhibit
Volunteering	Without payment	Oblige
Volunteering	Unpaid	Paid labor
Volunteering	Offer	Gainful employment
Volunteering	Informal	Gainful occupation
Volunteering	Voluntarily	Wage employment
Life Satisfaction	Payment	Frustration
Life Satisfaction	Atonement	Discontent
Life Satisfaction	Change	Disappointment
Life Satisfaction	Compensation	Dissatisfaction
Life Satisfaction	Contentment	Displeasure
Life Satisfaction	Desire	Distaste
Positive mood	Affirmation	Negativity
Positive mood	Encouraging	Negativism
Positive mood	Positivism	Despair
Positive mood	Favorableness	Lack of hope
Wellbeing	Wellness	Unwell
Wellbeing	Healthy	Unhealthy

Wellbeing	Successfulness	Diseased
Wellbeing	Social Welfare	Detrimental
Elder	Senior	
Elder	Seniority	
Elder	Elderly	
Elder	Aged	
Elder	Aging	
Elder	Older	
Elder	Fourth Year	
Elder	Old	
Elder	Older Adults	
Elder	Aged	
Elder	Geriatric	
Elder	Aging	
Personal growth	Personal Development	Degradation of the individual
Personal growth	Development Of The Individual	Decay of personality
Personal growth	Self-Fulfilment	Personal decline
Personal growth	Individual Development	Degradation of person
Relatedness	Relationship	Separation
Relatedness	Kinship	Conflict
Relatedness	Affinity	Dissimilarity
Relatedness	Association	Isolation
Relatedness	Connection	Division
Relatedness	Connectivity	Solitude
Positive relations	Good Relations	Distant relationship
Positive relations	Beneficial Relations	Hostile relationship
Positive relations	Certain People	Unhealthy relationship
Positive relations	Constructive Interaction	Unfriendly relationship
Social Acceptance	Social Recognition	Disagreement
Social Acceptance	Public Acceptance	Self-isolation
Social Acceptance	Community	Social distancing
Social Acceptance	Collective Acceptance	
Social Acceptance	Community Acceptance	
Social Integration	Social Inclusion	
Social Integration	Social Insertion	
Social Integration	Social Inclusiveness	
Social Integration	Social Reinsertion	
Social Integration	Society Reintegration	
Self-efficacy	Self-Motivation	Hesitation
Self-efficacy	Belief In One’s Ability To Succeed	Fail
Self-efficacy	Personal Efficacy	Procrastination

Competence	Ability	Inability
Competence	Capability	Incapacity
Competence	Proficiency	Incompetence
Competence	Skill	Ineptitude
Competence	Expertise	Impotence
Competence	Functionality	Uselessness
Competence	Capacity	Inadequacy
Competence	Talent	Unfitness
Competence	Aptitude	Inefficiency
Autonomy	Freedom	Confinement
Autonomy	Independence	Dependency
Autonomy	Self-Sufficiency	Forbidding
Autonomy	Free Will	Prohibition
Autonomy	Empowerment	Denial
Autonomy	Independent	Rejection
Social contribution	Social Welfare Benefit	
Social contribution	Social Allowance	
Social contribution	Welfare Benefit	
Environmental mastery		
Perceived development of one's potential		
Money		
Happiness		
Family		
Content (feeling)		
Luxury		
Social Status		
Job		
Intergenerational contact		