

TOPIC MODELLING OF COVID-19 MEDIA COVERAGE IN UGANDA

By Julian Park, Joyce Yanru Jiang & Anne Whitehead
@ Whitehead Communications 2021

INTRODUCTION TO THE RESEARCH

As the COVID-19 pandemic carries on into 2021, despite vaccine development, countries like Uganda continue to see their fight against the pandemic reflected in news media coverage. In 2020, Whitehead Communications applied a machine learning technique to analyse the trends in Ugandan online news coverage of the pandemic. We have now updated our study using LDA Topic Modelling to identify dominant topics of Ugandan COVID-19 coverage from March 2020 up to the end of June 2021. This analysis shows an evolution in the pertinent themes of COVID-19 coverage in Uganda as well as a changing volume of overall coverage over time.

Whitehead Communications gathered 32,212 English-language news articles that include the keywords “COVID” and/or “Corona” from a selection of Uganda’s leading news websites. These websites include Uganda’s major print publications – Daily Monitor, Red Pepper and The Observer – as well as major online-only Ugandan media houses The Independent, Chimp Reports, Nile Post, PML Daily, Softpower, The Tower Post, Eagle News and Trumpet News. (Unfortunately, Uganda’s New Vision data was unavailable at the time of this analysis.) This research was done to broaden the chronological scope and topic diversity of previous exploratory research. As a result, the methodology of this research is largely consistent with said previous research. We used the machine learning method called Latent Dirichlet Allocation (LDA), which is a type of Natural Language Processing (NLP) that groups content into categories based on identification of keywords. As with the previous study, this research is a demonstration of qualitative methodology. It provides insights into the prevalence of the story over time and indicates key trends in how the subject of Coronavirus was covered in Uganda, though it does not provide comprehensive data on all aspects of media coverage.

PURPOSE

This research aims to expand our exploration of how COVID-19 has been covered by Ugandan online news media from the beginning of 2020 to June 2021 using machine learning. We identified and scraped articles from top Ugandan news sites publishing in English and processed them through a LDA Topic Modelling computational technique to determine which topics are covered related to the pandemic. Our research is intended to deliver further insights to parties with an interest in Uganda’s media industry, or its COVID-19 response, or the machine learning methodology as it applies to media and communications research. This update to our previous research is part of a wider research project by Whitehead Communications exploring the application of multiple new research methods to see the bigger picture, draw correlations and build stronger research-based foundations on which to develop communications strategy.

OVERVIEW OF RESULTS

When compared to the results of our first study, the number of articles more than doubled since July 2020 (from 13,000+ to 32,000+). However, the articles per month were less than in the first few months of the pandemic, indicating a drop in interest in late 2020 and early 2021. When the first Ugandan COVID-19 case was confirmed on March 21, 2020, alongside the declaration of Uganda's first nation-wide lockdown, there was a spike in the number of articles including mentions of COVID/Corona. April 2020 marked the peak in number of articles published through online Ugandan media in one month of 2020, which was followed by a steady decline in the following months. However, since the lowest point in January 2021, we saw a significant increase in the number of Coronavirus related articles alongside a resurgence in Ugandan COVID-19 cases and new lockdown in June 2021, setting a new all-time high for COVID-19 coverage in one month.

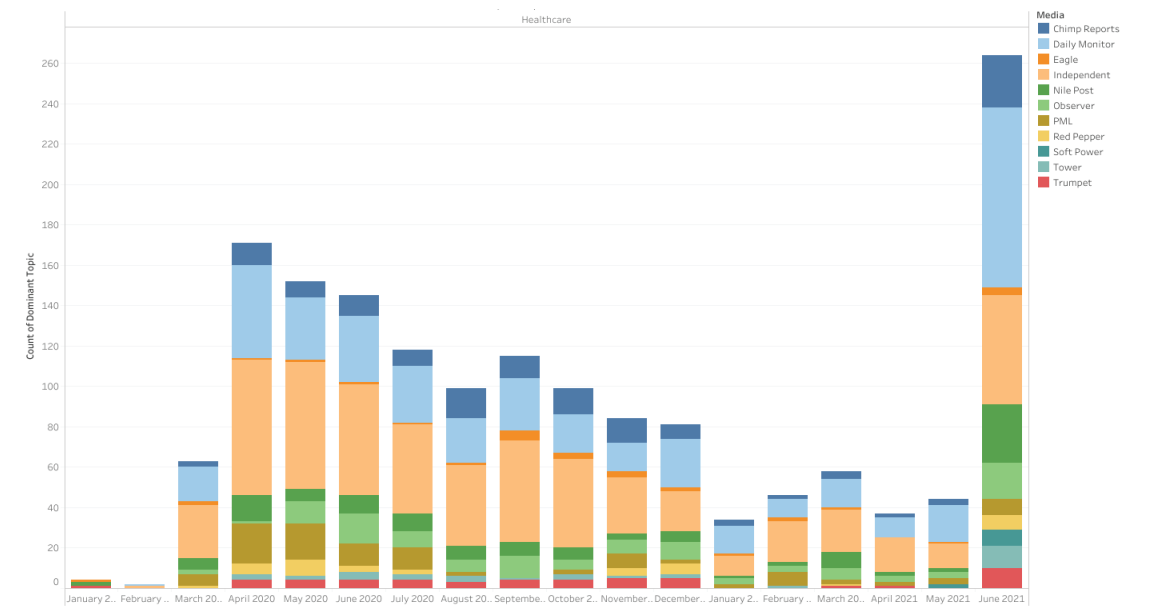
It is worth noting that Uganda had an election in January 2021, which was accompanied by a total shutdown of the internet for about a week, during which time online media was not able to publish as usual. This focus on a political season, along with a normalisation of the pandemic and slow progression of cases in the country, may help explain why coverage decreased between May 2020 and May 2021.

There were several recurring topics through multiple iterations of our modelling process. Topics such as "Education", which was primarily centred around school closures and the impact of COVID-19 on students, is an example of this. Other Topics such as "Healthcare", "Sports" and "Police Action" also appeared in almost every result in both years despite changing various parameters of the modelling process such as the number of topics. The articles that are most associated with these topics are highly distinct in vocabulary usage from the subject matter of other articles. For example, the aforementioned "Police Action" topic is associated with words such as "detain" or "arrested". These words have a much lower chance of appearing in non-police related articles, which makes it easy for the algorithm to pick up on this pattern regardless of parameters. In addition to repeating topics, there were topics that evolved due to the additional articles used in our update.

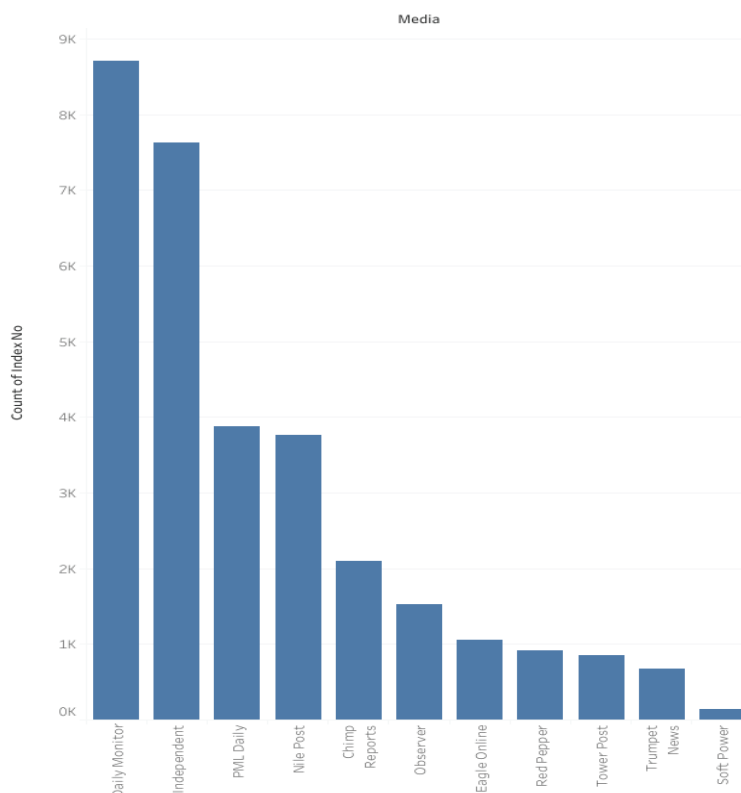
There were several topic changes between 2020 and 2021, which roughly reflected the global shift in perception about the pandemic – from an unprecedented emergency in the spring of 2020 to a continuous concern as new treatments and vaccines began to be introduced and the public and private sectors adjusted to getting on with things despite the continuing threats of the pandemic. New emerging topics in 2021 included "Travel & Tourism", "Local Donations" and "Innovation Response To Domestic Outbreak".

Timeseries of News Coverage

The following visualisation shows how the overall number of articles produced within our sample of Ugandan online media decreased between May 2020 and May 2021, then increased dramatically in June of 2021.



News Coverage by Media House



The number of articles we scraped across different media houses is represented in the chart below. Daily Monitor – one of Uganda’s leading daily newspapers – produced the most online coverage among those in our sample, followed by The Independent. Other major online media PML Daily and Nile Post came in at a distant 3rd and 4th, followed by Chimp Reports, The Observer and other media houses.

TOPICS

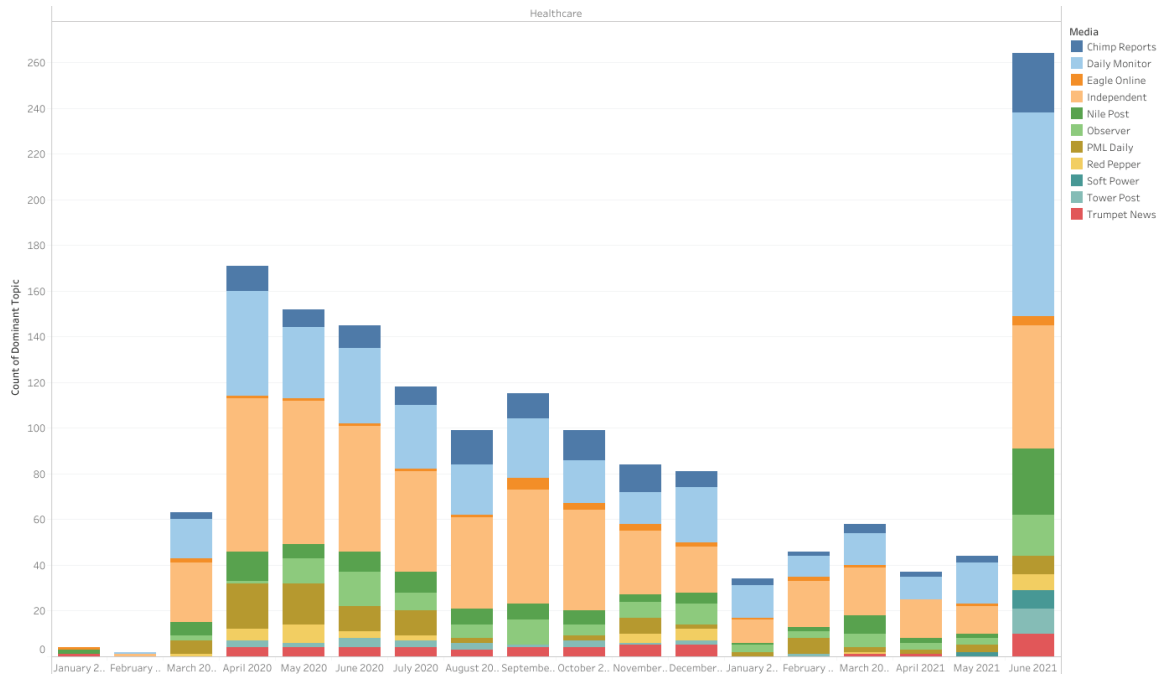
We tested out several parameters to determine the most effective new LDA Topic Modelling categorisation of articles. The following topics emerged in Ugandan articles about COVID-19 from the beginning (January 2020) up to the end of June 2021.

1. Healthcare
2. Vaccines
3. Economy & Finance
4. Travel & Tourism
5. Cases & Testing
6. Editorials & Personal Stories
7. President Museveni & Regional Politics
8. Electoral Politics
9. Police Action
10. Children & Women
11. Innovation Response To Domestic Outbreak
12. Sports
13. Education
14. Local Donations
15. Global Outbreak
16. Parliament, Courts & Justice

These topics indicate the main types of COVID-19 stories being published in Uganda related to coverage of the pandemic and were identified using an unsupervised machine learning algorithm, which was run multiple times with fine-tuned parameters to determine the optimal results.

Most of these topics are carried over from the previous 2020 clustering process with some modifications. However, there were some differences from last year's assessment. One main difference was the development and distribution of COVID-19 vaccines and other treatments. Other differences indicate changes in various industries as Uganda begins to accept that life with Coronavirus will not be over as quickly as expected. This is shown in Cluster #11 where various Ugandan organisations are responding to the need for increased e-commerce and other COVID-19 caused needs. Furthermore, articles focused on travel shifted from focusing on restrictions in 2020 to integrating more stories about the re-emergence of the tourism industry in 2021 and how it has been faring with the effects of COVID-19. More details on our methodology and an examination of key topics with representative articles are shared below.

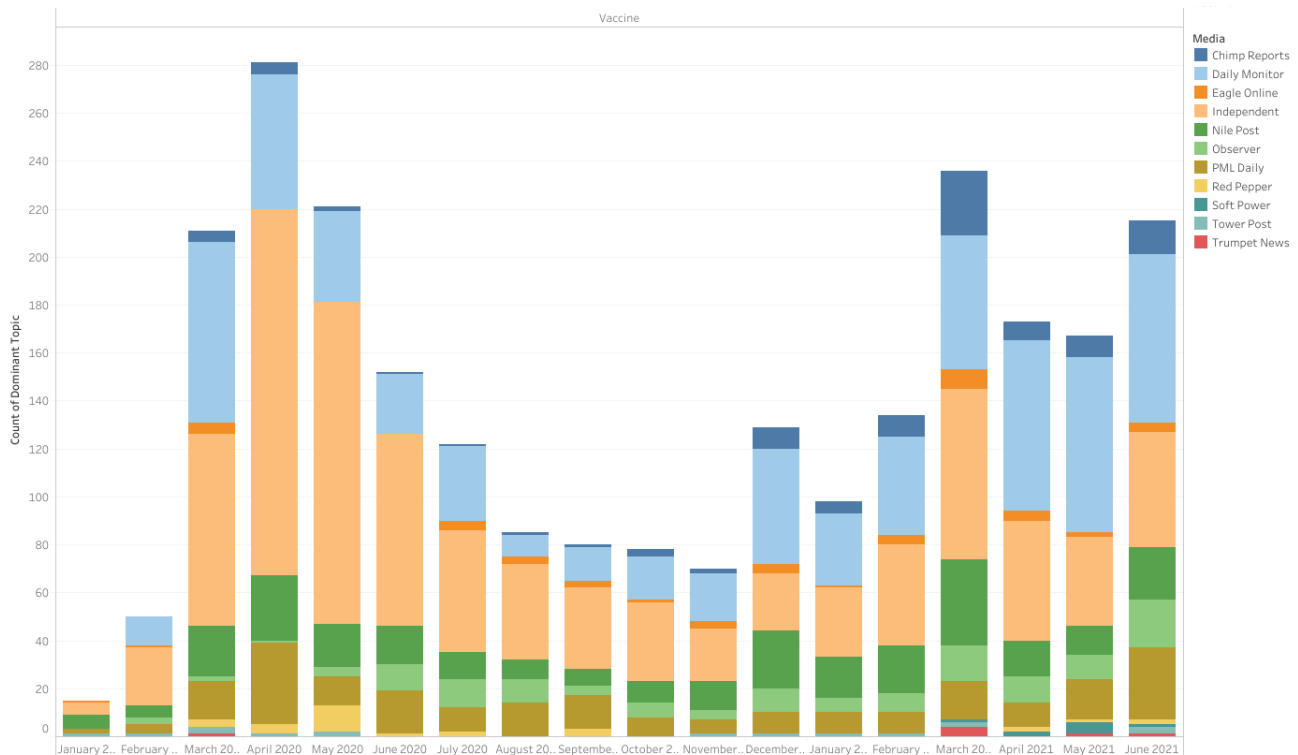
Healthcare



Healthcare-related articles reflected the wider trend in pandemic coverage through the year, with a major spike in the month of June 2021. As shown by the sample headlines, this was due to the pandemic spreading in Uganda. ICUs and other medical facilities were overwhelmed with news coverage closely following the situation. This topic also picked up government regulation is local treatments like COVIDEX. Sample headlines paint a stark picture as Uganda's medical infrastructure struggles to manage the pandemic with limited resources and equipment.

Topic	Sample Headlines	
Healthcare	18/6/2020	Namboole stadium to hold 40,000 beds
	/11/5/2021	Lack of trained staff leaves ICUs in regional referral hospitals idle
	18/6/2021	Uganda needs 8,000 cylinders of oxygen daily – Dr. Aceng
	27/5/2021	COVID-19: Entebbe hospital runs out of beds
	9/6/2021	Kabale hospital runs out of space for Covid-19 patients
	14/6/2021	COVIDEX herbal Medicine is Illegal, says Government
	23/6/2021	Gulu Hospital runs out of Covid drugs

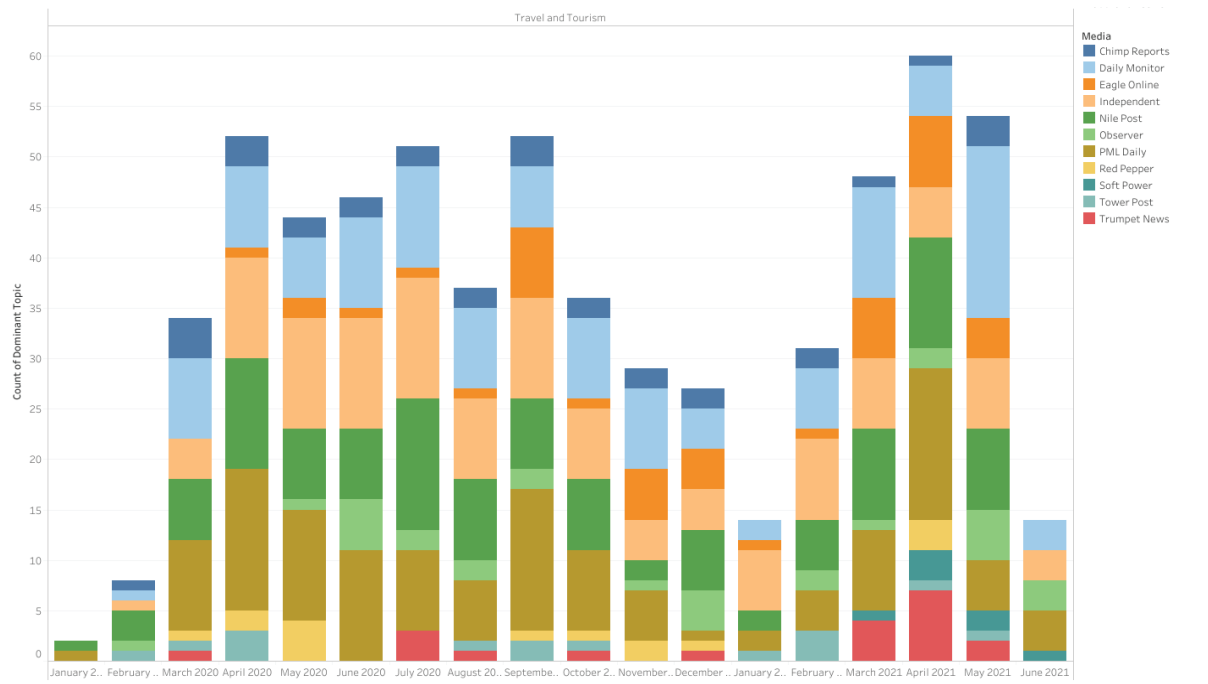
Vaccines



This topic did not appear in our 2020 model and demonstrates a significant development globally in the fight against COVID-19: the creation and distribution of COVID-19 vaccines and other treatments. This topic includes news about the testing and approval, distribution, shortages, and precautions associated with COVID vaccines.

Topic	Sample Headlines	
Vaccines	8/12/2020	Coronavirus declared a global health emergency by World Health Organization
	20/2/2021	Over 200 million vaccine doses administered globally
	2/3/2021	Uganda to receive first batch of Covid-19 vaccine on Friday
	25/3/2021	Uganda alters Covid-19 vaccine roll-out plan to cover more people
	16/3/2020	Low turn up for Covid-19 vaccines
	14/4/2021	Johnson and Johnson COVID-19 vaccine halted over blood clot fears
	7/5/2021	Ugandans to receive second Covid-19 vaccine after 12 weeks instead of 8
	1/6/2021	WHO requests for 20m doses of COVID-19
	6/12/2020	After Burundi president's death, African leaders response to Covid-19 comes under scrutiny
	29/6/2020	China refutes Pompeo's anti-China statements over China-Africa ties

Travel & Tourism

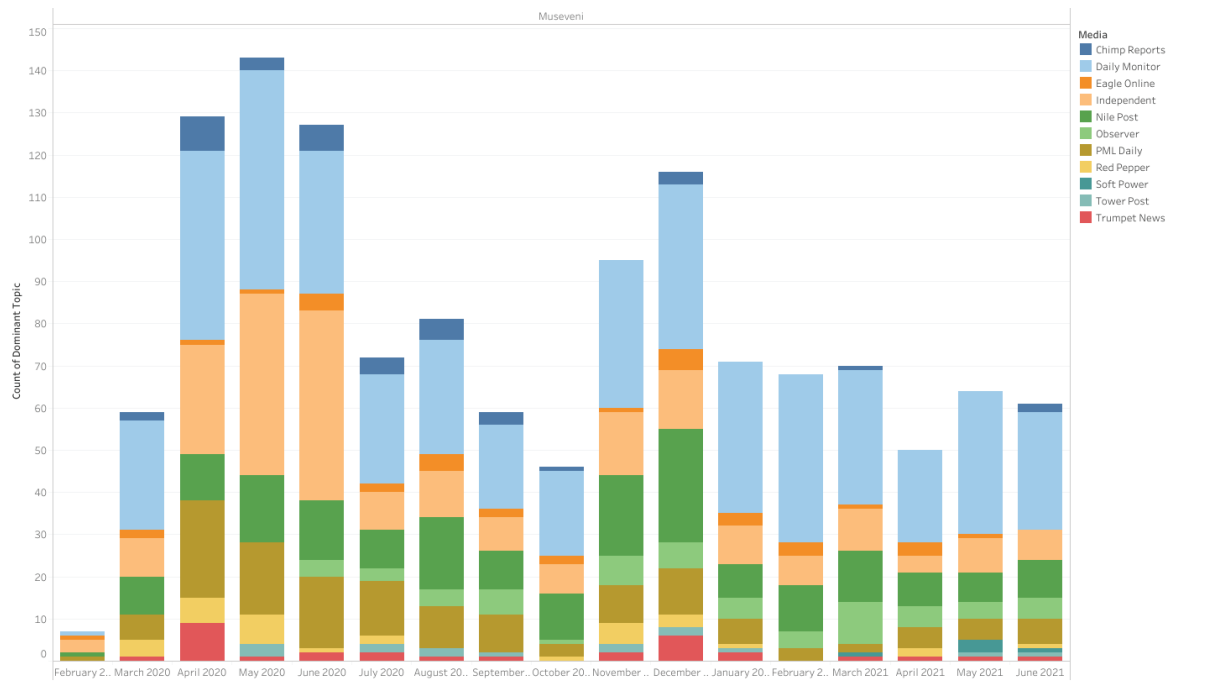


While “Travel Restrictions” appeared in our 2020 model, this year’s analysis revealed a development in stories about travel to include a strong narrative about the tourism industry. With travel restrictions gradually being lifted internationally, Uganda’s focus turned toward revitalizing its tourism industry. Stories around Travel & Tourism increased as plane routes were re-established and lockdowns around the world were lifted, allowing visitors into the country. Concentrated efforts were being made in these periods to promote Uganda as a travel destination, despite the ongoing pandemic.

Topic	Sample Headlines	
Travel & Tourism	25/7/2020	Uganda tourism sector reopens after 4 months of lockdown
	29/9/2020	Uganda fronts endangered golden monkeys as new tourist attraction
	7/4/2021	UTB, Uganda Airlines recommit to joint promotion of destination Uganda
	26/4/2021	Why Private Sector Should Participate in Virtual Pearl of Africa Tourism Expo
	18/5/2021	Chinese tourists could be the future of African tourism
	27/5/2021	Covid-19 decimates Uganda tourism earnings, down by shs 4.1trillion in 2020



President Museveni & Regional Politics



Undeterred by COVID-19, political opposition, state violence and allegations of rigging, President Museveni was once again reelected to extend his 35-year rule in January 2021. The LDA Topic Model identified President Museveni as a news topic in itself, which tied together both domestic stories about his declarations on COVID-19 measures and electoral activities, and his role as a peacekeeper and strong man in the region.

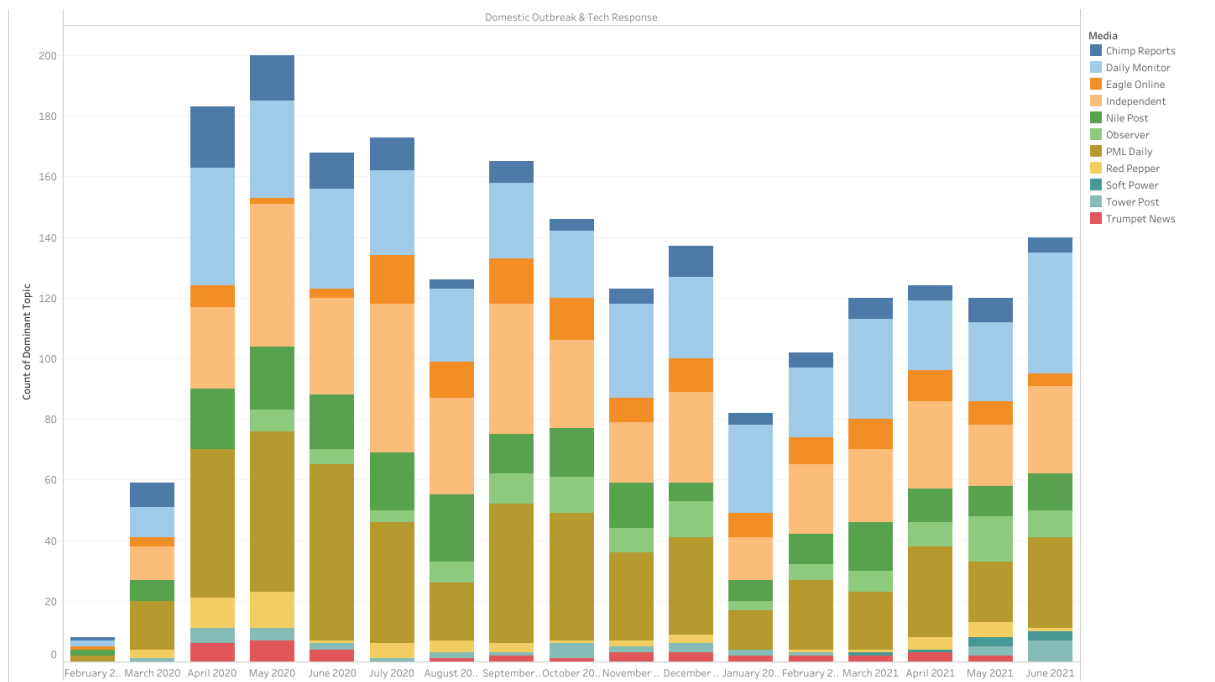
Sample Headlines

President Museveni & Regional Politics

22/4/2020	COVID-19 fight confirms Museveni leadership credentials
22/5/2020	Why the confusion in presidential addresses?
11/6/2020	Museveni's powerful Sudan fixer dies after contracting coronavirus
16/11/2020	TIGRAY CRISIS: Museveni Meets Ethiopian Deputy Premier, Calls for Negotiations
10/12/2020	Top US congressman calls for sanctions against President Museveni, bigwigs in security forces over grave human rights violation
8/3/2021	Museveni guards holding 50 missing persons
10/3/2021	Museveni: 24,000 soldiers deployed to defeat opposition plot
18/4/2021	"You are misinformed"— Uganda blasts US government over report on human rights abuse



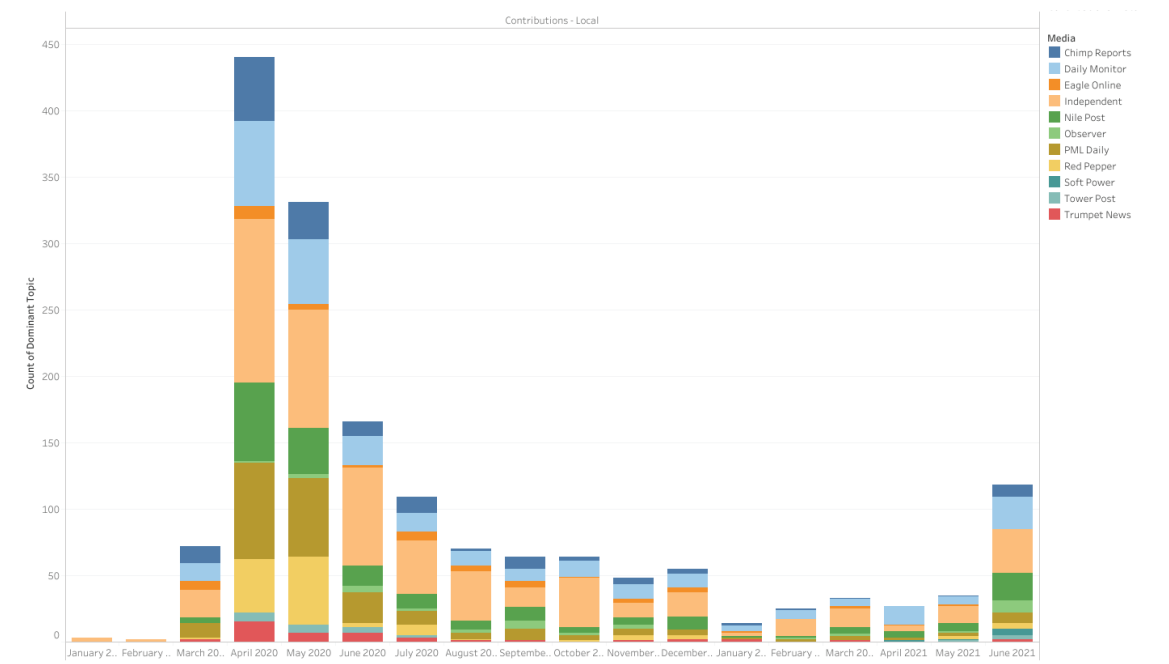
Innovation Response To Domestic Outbreak



With COVID-19 putting the brakes on Ugandan development, Uganda's leaders in the private sector, foundations and NGOs contributed to the recovery. This topic picked up on strategic contributions from different sectors of Ugandan society that were covered in Uganda's media for working to make progress with key initiatives in entrepreneurship, green energy, innovation and employment.

Topic	Sample Headlines	
Innovation Response To Domestic Outbreak	23/3/2020	Jumia announces actions to support governments' fight against coronavirus
	17/4/2020	COVID-19 CRISIS: FAO commits to safeguarding food security amid coronavirus lockdown
	27/5/2020	Advocate For Government and Private Sector Collaboration - Tony Elumelu
	2/9/2020	We must help Africa rebuild boldly, but smartly – AfDB President Adesina
	21/3/2021	UN Uganda and Equity Bank set new partnerships for acceleration of attainment of SDGs
	5/1/2021	Enhancing Uganda's innovative industry by commercializing intellectual property

Local Donations



Ugandan media coverage included a topic on local donations to those worst affected by the pandemic. While international aid flowing into Uganda emerged in our 2020 topic model, by 2021 the domestic efforts dominated support for vulnerable groups. This included donations by companies, NGOs, individuals and government relief funds, such as masks, cars and fuel, food, and relief funds distributed by the Prime Minister's Office.

Topic	Sample Headlines	
Local Donations	8/4/2020	Opposition gives out food to poor despite Museveni ban
	11/4/2020	Pastor Kayanja donates 5000Kgs of relief food to Karamoj
	15/5/2020	Covid-19: How residents of Kabarole donated and rescued vulnerable people
	20/5/2020	Kwagalana Group donates brand new double cabin 4*4 pick-up to COVID -19 National Task Force
	11/12/2020	Coca Cola donates hygiene kits to Lira
	11/3/2021	Government Facemasks finally arrive in Mbale, no facilitation to distribute them
	23/6/2021	Nile Breweries Donates 2500 Litres of Fuel to Buikwe, Jinja City COVID-19 Task Forces
	23/6/2021	Vulnerable people to get UGX.100,000 as govt moves to disburse UGX.53b Covid-19 cash relief
	23/6/2021	Nytil donates 30,000 masks to KCCA

METHODOLOGY

We used Latent Dirichlet Allocation (LDA) Topic Modelling, a machine learning technique, to categorise news based on Natural Language Processing automated analysis and assign topics to each article. LDA creates topics by first breaking up all the words in the data into a “bag of words”. There is no distinction between articles in LDA nor is there any weight given to the order that the words appear in. The algorithm then groups words based on their co-occurrence to generate distinct clusters. We can then use the frequency of these words to label the original articles under the most dominant topic. The following breaks down the main steps in our exploration of this methodology using 32,212 English-language news articles that include the stem keywords “COVID” and/or “Corona” from a selection of Uganda’s leading news websites between early 2020 and June 30, 2021.

***Data collection:** First, using local media knowledge, we identified the Ugandan news websites likely to produce the highest volume of articles. Our goal was to gather the text of related articles and other associated information such as the title and date. We developed a scraping script in Python using the Beautiful Soup package for each news website to automate this process. On each website, we searched the capital-insensitive stem keywords “covid” and “corona” to gather articles related to COVID-19 news (this also included extended varieties of these keywords such as “coronavirus” and “COVID-19”). Daily Monitor required a manually intensive method of scraping using Chrome Driver due to the formatting of their query results. The data we scraped ranged in dates from January 2020 to June of 2021. This range significantly adds to the original period of the research.

***Data cleaning:** Prior to modelling our data, it was important to pre-process it. We took several steps to improve the consistency and coherence of our topic modelling.

- Removed duplicated articles
- Only kept articles whose headlines or article body text contained Covid-related keywords (ex. COVID, corona, lockdown, outbreak, curfew, reopen, facemask, respiratory, WHO, hospital, etc.). The keywords were determined in a discretionary way that may lead to some Type I and Type II errors. However, we believe our result is mathematically reliable.
- Removed articles that contain keyword “Toyota” (to exclude articles that were picked up in our data about the Toyota Corona car model) and “coronation”
- Removed articles that contain AFP (Agence France-Presse) and Xinhua (Chinese news agency) and do not have any Uganda-related keywords. Both AFP and Xinhua are international news agencies that sell news to Ugandan local media. Though the Ugandan audience is still exposed to this content, we chose to remove international articles unrelated to Uganda in order to focus on news produced locally.

***Challenges with specific media houses:**

**New Vision's website had a broken query system. This was due to a reformatting of their website during the period of research. We were not able to search for relevant articles much less scrape them. Although we were in communication with the team at New Vision, we were unable to rectify the challenge in time, and for this reason we have omitted them from the scraping process.

**The Brink News website had protections in place to prevent scraping. Our attempts to run Beautiful Soup through the query resulted in an error message: 403 Access Forbidden. Although we included this website in our first analysis and were in touch with the media house, we were unable to resolve the problem and thus excluded this website from our second dataset.

**Daily Monitor required a special scraping method because of how their query results were formatted. Rather than spitting out multiple pages of results, Daily Monitor's website required the user to scroll down and click "view more" to scrape more articles. Using Chrome Driver, we were able to automate the process of clicking and scrape all the relevant articles.

***LDA Topic Modelling:** We ran our algorithm in many variations, changing specifications of the number of topics between 14 to 18, and found that we received the most optimal results when we set the number of topic clusters to 16.

***LDA limitations:** We were able to determine both strengths and weaknesses of the LDA Topic Modelling method with our previous and recent experiences. The LDA algorithm was consistent at picking up certain types of articles such as on the topic of sports or education. This method was less successful at consistently discerning and categorizing topics such as judicial and parliamentary affairs. Considering that LDA measures the cooccurrence of words when fed an article, this makes sense. Topics such as sports have a clearly defined set of words that has very little crossover with other article topics. Words such as team, player, seasons, and league appear with very high frequency in sports articles while appearing very little in others. For the rest of our scraped COVID-19 related articles, topics relating to various branches of government, lockdowns, etc. have less of a distinct boundary when it comes to keywords. LDA has a much harder time giving consistent and coherent results for these topics.

***Topic analysis:** Our first trial for topic analysis was to check keywords for each cluster. We applied text analysis of article and headline contents using Python to develop word clouds/rankings for each cluster. Since some topics comprised of mixed subtopics, we were unable to tell the actual topic from analysing keywords themselves. We then reviewed the headlines alongside the topic percentage contribution per article and in some cases the body text of articles that returned high relevance scores. We used local knowledge to assess the makeup of each topic.

ACKNOWLEDGEMENTS

We must express our gratitude to everyone who offered their assistance to this project, from Whitehead Communications' own research team spread across Uganda, UK, USA and Hong Kong, to the editors and owners of various Ugandan media houses who offered feedback on our preliminary data and updated their websites to facilitate our research. We are also grateful to Ugandan PR practitioners and media researchers who offered their recommendations and moral support. Finally, we would like to recognize the AIEM research group at Boston University who took interest in our project when we realized that we were doing similar research and published our 2020 results at <https://covid19.philemerge.com/uganda>

We share this exploratory research in good faith, to encourage ethical use and contributions of African data to the global development of new communications research practices, and we remain open to learning and collaborating further with anyone interested. Please contact anne@whiteheadcommunications.com with any questions