

Pandemic State of Mind

Using Natural Language Processing to
Understand Communication by New York
Leadership During Covid-19

Taiwanese Data Professionals Meetup
Angeline Protacio
July 23, 2020

Outline

1. Motivation
2. Methods
3. Data Collection
4. Results - Topic Modeling
5. Results - Parts-of-Speech Analysis
6. Results - Sentiment Analysis
7. Conclusions and Future Work



How a friendly, airtight relationship between the Democratic heavyweights turned ugly. Is it beyond repair?

De Blasio Used Last-Minute Text to Tell Cuomo Schools Would Stay Shut

The governor and mayor, who have long had a combative relationship, can't seem to make peace during the biggest crisis of their careers.



After Mayor Bill de Blasio said that the city's public schools would stay closed through June, Gov. Andrew M. Cuomo questioned the mayor's authority to make that decision. Dave Sanders for The New York Times

Methods

Data Sources (January 1, 2020 - May 22, 2020)

- Mayor de Blasio's Speech/Interview Transcripts (n=211)
- Governor Cuomo's Speech/Interview Transcripts (n=184)

Libraries:

- requests/BeautifulSoup/Selenium
- pandas/NumPy/re/matplotlib/seaborn
- Scikit-learn
- spaCy
- VADER



Web Scraping

- Good for collecting data that is otherwise not easily obtained
- Code is likely to break if site updates frequently
- Some sites don't like scrapers
- Data quality is highly variable

The screenshot displays the official website of the City of New York, specifically the Office of the Mayor. At the top, a blue banner contains links for 'Information on coronavirus', 'Agency service suspensions/reductions', and 'Apply to be an Open Restaurant'. Below this, the site's navigation bar includes 'NYC Resources', 'NYC311', 'Office of the Mayor' (selected), 'Events', 'Connect', 'Jobs', and a search bar. A secondary navigation bar features buttons for 'Mayor', 'First Lady', 'News' (highlighted), and 'Officials'. The main content area is titled 'The latest news from the Office of the Mayor of New York City' and shows a list of news stories. On the left, there are filters for 'CATEGORY', 'DATE' (with 'Select a Date' buttons), and 'TYPE'. The news stories listed include: 'Mayor de Blasio Announces 'Take Care Initiative' to Help New Yorkers Safely Separate' (May 27, 2020), 'Transcript: Mayor de Blasio Holds Media Availability' (May 27, 2020), 'Mayor de Blasio Appoints Annabel Palma as Test & Trace Corps' Lead Equity Officer' (May 26, 2020), 'Mayor de Blasio Signs COVID-19 Relief Package Into Law' (May 26, 2020), 'Mayor de Blasio, First Lady McCray Join with DJ David Guetta to Support COVID-19 Relief with Virtual Charity Event' (May 26, 2020), and 'Transcript: Mayor de Blasio Holds Hearing and Signs Intros 1898-A, 1908-B, 1914-A, 1916-A, 1932-A, 1936-A, 1940-A'.

Transcript Data

“the total hospitalization rate is down again in the state of new york. were down to 16000. if you look at the numbers we were at 18000 people hospitalized for a period of time. it flattened there for a while it paused there. then it went down to 17000 but this is a low from our high point of 18000.”

- Governor Cuomo, 4/19/2020

“we have obviously had it a harder hit than almost any place but what we have to be really smart about is understanding this disease which has ravaged the world. you know its ravaged big cities and small towns in every kind of country. you let it in the door its going to cause you endless trouble.”

- Mayor de Blasio, 4/19/2020

Text Pre-Processing

- Lemmatization - converts words to base form
 - studies = study
 - studying = study
- Stopwords - commonly used words without meaning
 - “How”, “to”, “and”, “the”
 - Names of interviewers ('lehrer', 'brian', 'darden', 'moderator')

Topic Modeling Overview

- Document: Speech transcript
- Term: word

	Term1	Term2	Term3	Term4	Term5
Document1	0	0	1	0	0
Document2	1	1	3	0	0
Document3	0	0	0	0	1

Document-term Matrix

=

	Term1	Term2
Topic1	?	?
Topic2	?	?
Topic3	?	?

Topic-term matrix

*

	Topic1	Topic2
Document1	?	?
Document2	?	?
Document3	?	?

Document-topic matrix

Count Vectorization

	bit	bite	bitta	bitter	bizarre	black	blackout	blame	blank	blanket	blare	blasio	bleach	blend
Document0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Document1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Document2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Document3	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Document4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Document5	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Document6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Document7	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Document8	1	0	0	0	0	0	0	4	0	0	0	0	0	0
Document9	0	0	0	0	0	0	0	1	0	0	0	0	0	0

“ive spoken to mayor de **blasio**. its going to require a lot of assistance from the nypd its going to require a lot of assistance from different city agencies...i think we have the mayor who is on the telephone or some electronic means. there he is mayor bill de **blasio**.”

Topic-Term Matrix

	discussion	disdainful	disease	disgust	disgusting	dish	disinfect
topic0	0.233	0.004	5.981	0.000	0.019	0.000	0.000
topic1	0.180	0.000	0.063	0.015	0.007	0.011	0.483
topic2	0.088	0.000	0.512	0.014	0.001	0.000	0.000
topic3	0.177	0.000	3.375	0.000	0.000	0.000	0.000
topic4	0.311	0.000	0.000	0.019	0.010	0.000	0.000
topic5	0.116	0.000	0.000	0.007	0.000	0.000	0.000
topic6	0.157	0.000	0.000	0.000	0.000	0.034	0.000
topic7	0.058	0.000	0.000	0.008	0.076	0.000	0.015
topic8	0.142	0.025	2.780	0.002	0.000	0.000	0.059
topic9	0.000	0.000	0.000	0.000	0.006	0.000	0.000
topic10	0.029	0.003	0.000	0.000	0.001	0.000	0.000
topic11	0.224	0.000	1.023	0.000	0.004	0.000	0.000

Topic Modeling

Topic 0 - disease, phase, fight, support, information

Topic 1 - rate, virus, reopen, infection, business

Topic 2 - business, small, clinic, support, commissioner

Topic 3 - commissioner, information, **disease**, barbot, symptom

Topic 4 - medicaid, budget, percent, pay, billion

Topic 5 - supply, personnel, military, ventilator, katz

Topic 6 - kid, chancellor, parent, learn, educator

Topic 7 - applause, stop, young, save, kid

Topic 8 - street, nypd, **disease**, safe, homeless

Topic 9 - ventilator, bed, increase, curve, equipment

Topic 10 - attack, nation, law, progressive, build

Topic 11 - food, folk, basic, provide, safe

Topic Modeling

Covid Mechanics - disease, phase, fight, support, information

Reopening Metrics - rate, virus, reopen, infection, business

Neighborhood Impact - business, small, clinic, support, commissioner

DOH Communication - commissioner, information, disease, barbot, symptom

Healthcare - medicaid, budget, percent, pay, billion

Hospital Needs - supply, personnel, military, ventilator, katz

Education/School - kid, chancellor, parent, learn, educator

Neighborhood Resilience - applause, stop, young, save, kid

Homelessness - street, nypd, disease, safe, homeless

Hospital Status - ventilator, bed, increase, curve, equipment

Hate Crime - attack, nation, law, progressive, build

New Yorkers - food, folk, basic, provide, safe

Document-Topic Matrix

Topic8: Homelessness - street, nypd, disease, safe, homeless

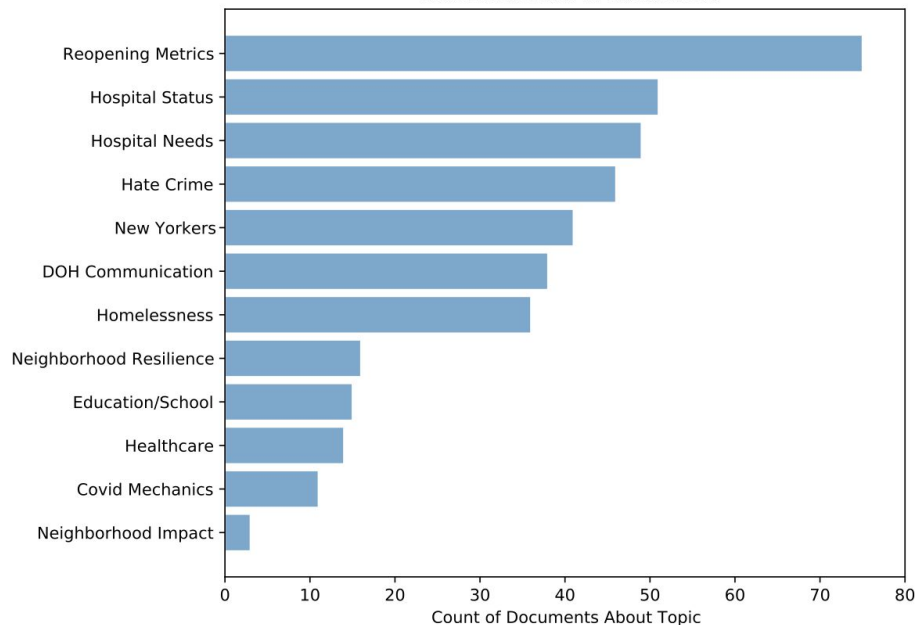
	topic0	topic1	topic2	topic3	topic4	topic5	topic6	topic7	topic8	topic9	topic10	topic11
Document15	0.03896	0.00000	0.0000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00940	0.36612	0.07127	0.00000
Document16	0.00000	0.25782	0.0000	0.29491	0.00000	0.05613	0.00000	0.00000	0.01892	0.18787	0.15989	0.00000
Document17	0.00000	1.63322	0.0000	0.13565	0.00000	0.00000	0.13608	0.00000	0.00000	0.01294	0.05510	0.00000
Document18	0.01188	0.09943	0.0017	0.08291	0.08216	0.00000	0.01563	0.00000	0.00000	0.00000	0.57768	0.00000
Document19	0.00000	0.79539	0.0000	0.00000	0.03797	0.00000	0.00000	0.18578	0.52266	0.00000	0.00000	0.00000
Document20	0.00000	0.20330	0.0000	0.14762	0.00532	0.00000	0.00000	0.00000	0.00000	0.12484	0.00000	0.00000
Document21	0.00000	0.22289	0.0000	0.00000	0.10890	0.00000	0.00000	0.00000	0.23922	0.00000	0.00000	0.02922
Document22	0.01528	0.05069	0.0000	0.00844	0.00000	0.00000	0.00000	0.01137	0.00000	0.02246	0.00000	0.00000
Document23	0.00000	0.27088	0.0000	0.00000	0.47841	0.00000	0.00000	0.23669	0.69773	0.00000	0.89267	0.00000
Document24	0.00921	0.19771	0.0000	0.12089	0.12153	0.04369	0.00000	0.00000	0.00000	0.00000	0.96057	0.00000

"the nypd do the security for the system so and to do something like this everybody has to work together. its never been done before. weve never closed down the subway system. the homeless have to be taken off the train to do this obviously. you cant disinfected the train if you have homeless people on it."

- Governor Cuomo, 4/30/2020

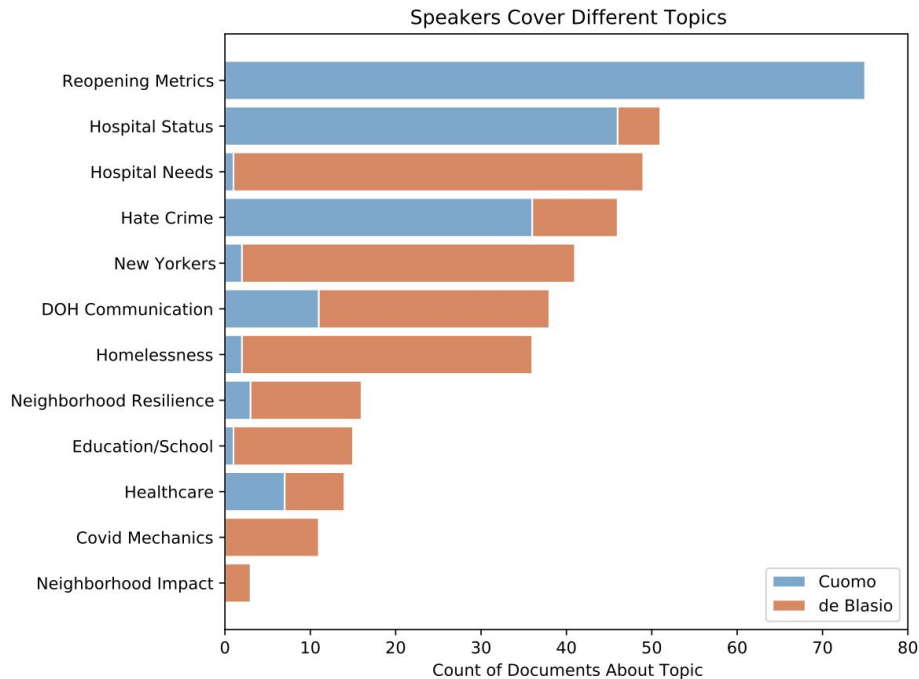
Topic Modeling

Reopening, Hospital Status, and Hospital Needs
Addressed Most in Documents



Reopening Metrics - rate, virus, reopen, infection, business, infect, hospitalization
Hospital Status - ventilator, bed, increase, percent, equipment, curve
Hospital Needs - supply, commissioner, ventilator, personnel, military

Topic Modeling



Reopening Metrics - rate, virus, reopen, infection, business, infect, hospitalization
Hospital Status - ventilator, bed, increase, percent, equipment, curve
Hospital Needs - supply, commissioner, ventilator, personnel, military

Parts-of-Speech using spaCy

look	mayor	de	blasio	the	mta	and	i
VERB	PROPN	PROPN	PROPN	DET	PROPN	CCONJ	PRON

are	all	together	on	this
AUX	ADV	ADV	ADP	DET

it	is	a	joint	effort
PRON	AUX	DET	ADJ	NOUN

Parts-of-Speech Analysis

Noun	Adjectives (Governor Cuomo)	Adjectives (Mayor de Blasio)
Testing	more, automated, diagnostic, own	more, widespread, diagnostic, automated

Parts-of-Speech Analysis

Noun	Adjectives (Governor Cuomo)	Adjectives (Mayor de Blasio)
Testing	more, automated, diagnostic, own	more, widespread, diagnostic, automated
Hospital	public, temporary	elmhurst, public

Parts-of-Speech Analysis

Noun	Adjectives (Governor Cuomo)	Adjectives (Mayor de Blasio)
Testing	more, automated, diagnostic, own	more, widespread, diagnostic, automated
Hospital	public, temporary	elmhurst, public
Yorkers	elderly, fewer	everyday, homeless, vulnerable, working

Parts-of-Speech Analysis

Noun	Adjectives (Governor Cuomo)	Adjectives (Mayor de Blasio)
Testing	more, automated, diagnostic, own	more, widespread, diagnostic, automated
Hospital	public, temporary	elmhurst, public
Yorkers	elderly, fewer	everyday, homeless, vulnerable, working
Government	best, active, competent, effective, proper	American, Chinese, Iranian

Parts-of-Speech Analysis

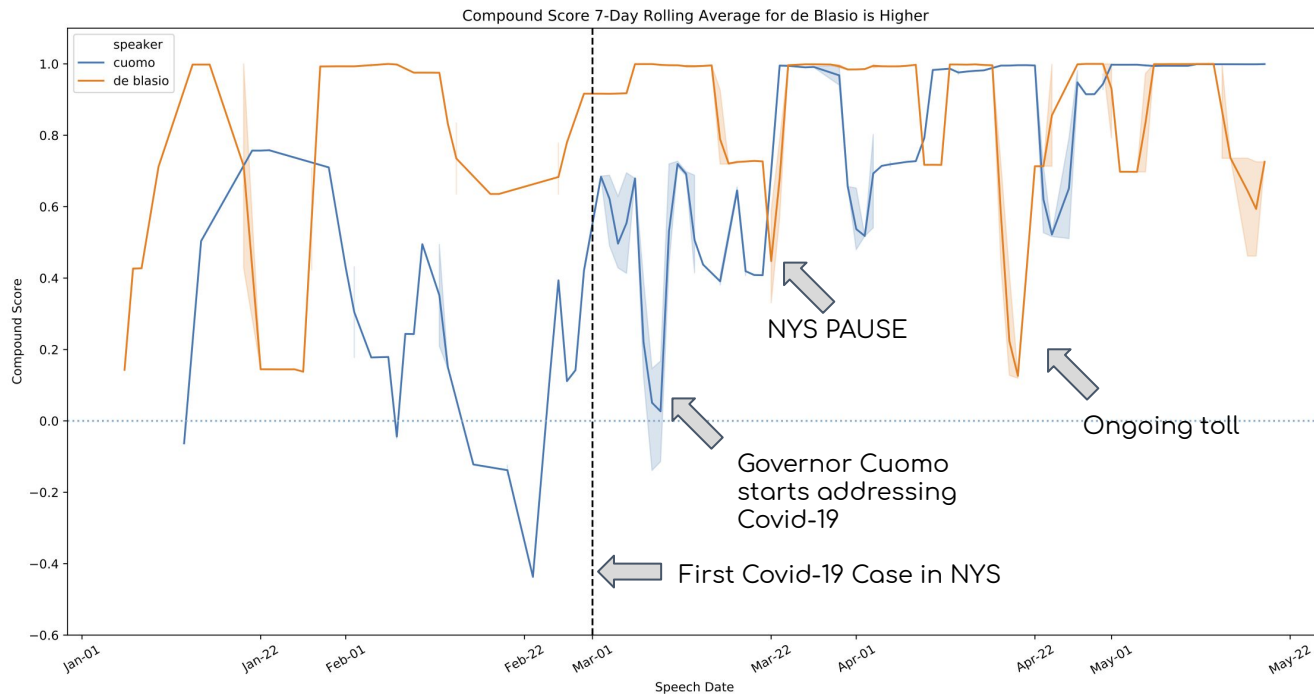
Noun	Adjectives (Governor Cuomo)	Adjectives (Mayor de Blasio)
Testing	more, automated, diagnostic, own	more, widespread, diagnostic, automated
Hospital	public, temporary	elmhurst, public
Yorkers	elderly, fewer	everyday, homeless, vulnerable, working
Government	best, active, competent, effective, proper	American, Chinese, Iranian
Care	appropriate, intensive	universal, urgent, available, public

Sentiment Analysis using VADER

Valence Aware Dictionary and sEntiment Reasoner

Input	Positive	Compound	Neutral	Negative
The book was good.	0.492	0.4404	0.508	0.0
At least it isn't a horrible book.	0.363	0.431	0.637	0.0
The book was only kind of good.	0.303	0.3832	0.697	0.0
The plot was good, but the characters are un compelling and the dialog is not great.	0.094	-0.7042	0.579	0.327

Sentiment Analysis



Sentiment Analysis (Negative)

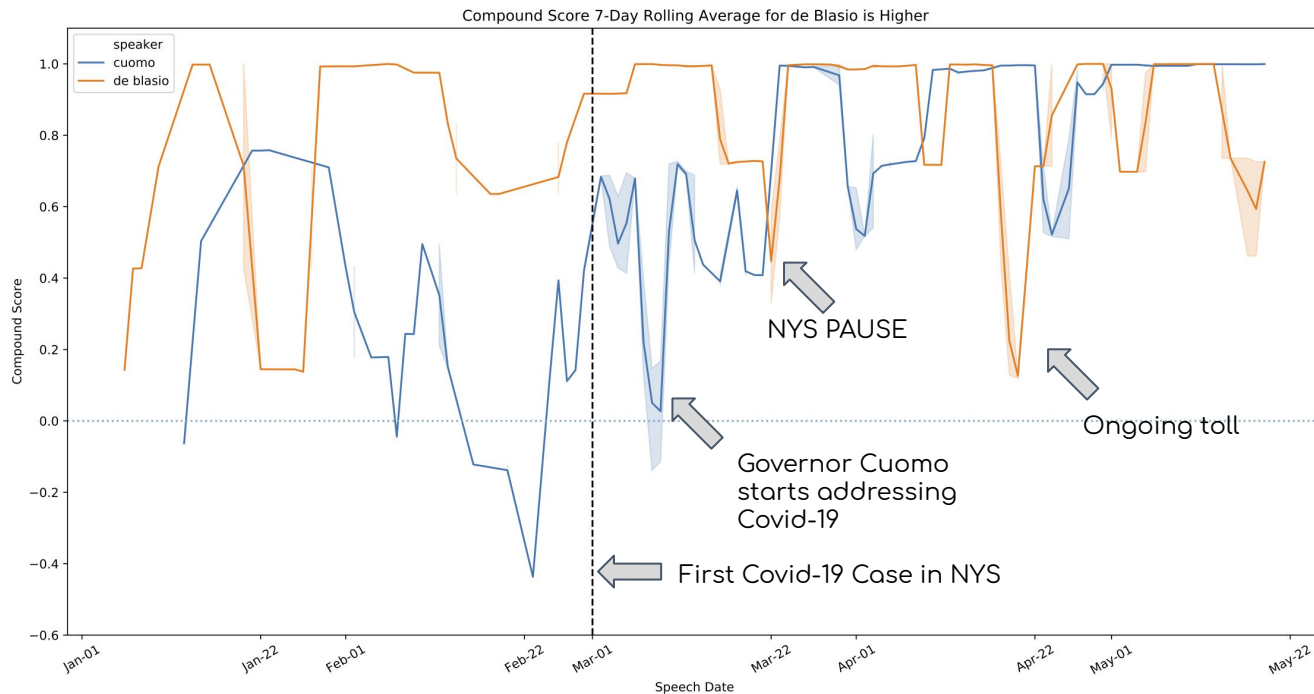
"The people who are in the vulnerable population - senior citizens, immune compromised, people with underlying illnesses. What am I worried about as Governor? People say to me, What are you worried about? I'm worried about undue fear and anxiety, hence why I'm constantly trying to communicate facts."

- Governor Cuomo, 3/5/2020, compound score: -0.97

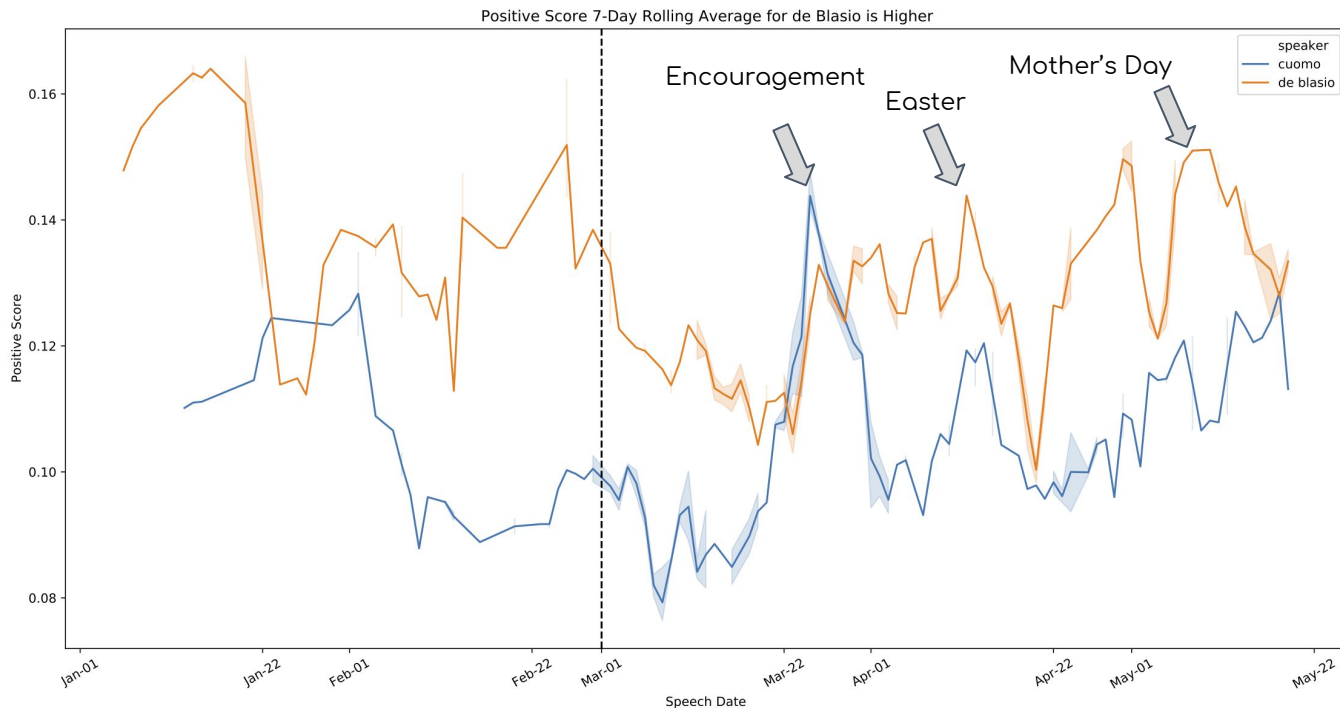
"A lot of people are going to die who don't have to die. That is the blunt truth. And it will start with dozens and it will go to hundreds and it will go farther than that when you talk about this country. The reality today, this minute, is if the president does not act in a matter of days, because look at the trajectory of this crisis."

- Mayor de Blasio, 3/20/2020, compound score: -0.89

Sentiment Analysis



Sentiment Analysis



Sentiment Analysis (Positive)

"New York loves everyone. That's why I love New York. It always has, it always will. And at the end of the day, my friends, even if it is a long day, and this is a long day, love wins. Always. And it will win again through this virus."

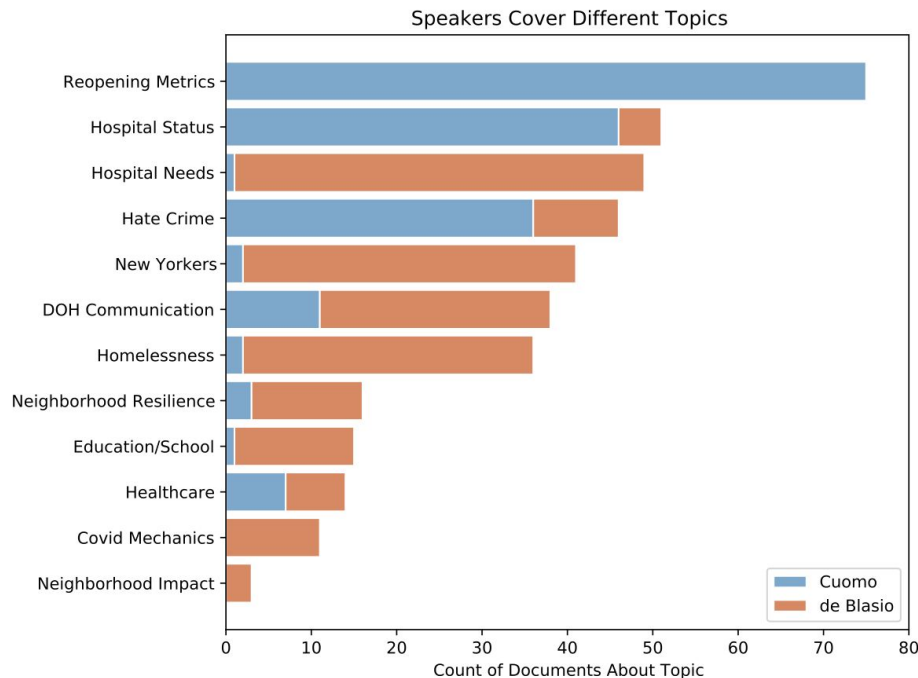
- Governor Cuomo, 3/24/2020, compound score: 0.9996

"Health care workers have been heroes. Our first responders have been heroes, but really everyday New Yorkers, you've been heroes too because you've really done an amazing job practicing social distancing, handling shelter in place, making it work. God bless you all."

- Mayor de Blasio, 4/9/2020, compound score: 1.000

Conclusions

- One speaker tends to dominate each topic



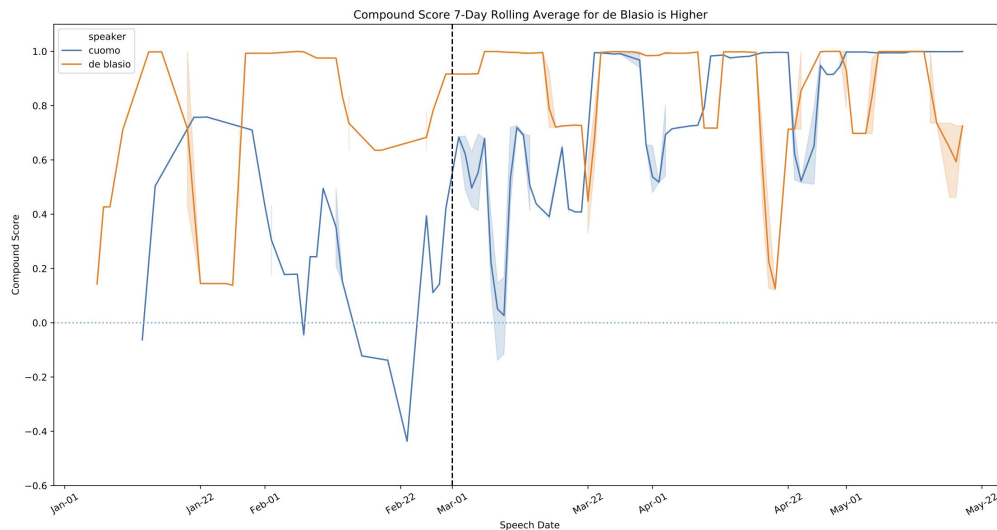
Conclusions

- No clear differences in how they describe frequent terms

Noun	Adjectives (Governor Cuomo)	Adjectives (Mayor de Blasio)
Testing	more, automated, diagnostic, own	more, widespread, diagnostic, automated
Hospital	public, temporary	elmhurst, public
Yorkers	elderly, fewer	everyday, homeless, vulnerable, working
Care	appropriate, intensive	universal, urgent, available, public

Conclusions

- Compound speech sentiment for both speakers reflects events and continued encouragement



Future Work

- Update speech transcript data
- More detailed reading of speeches to provide more context
- Application of word embedding

Thank you!

Email: email@angelineprotacio.com

Website: www.angelineprotacio.com

Github: <https://github.com/angelinepro>

LinkedIn: <https://www.linkedin.com/in/angeline-protacio/>