Previous work:
1. Data preprocessing pipeline
    - Trained on Vaccine/Travel data separately
2. Models:
    - Linear SVM (full data):
        - Vaccine: accuracy: 0.80, f1 score:0.68
        - Travel: accuracy:0.68, f1 score:0.63
    - BERT (10 days data): accuracy: 0.77, f1 score:0.76
    - CNN: fine tuning

Will do this week:
1. Create new linguistic features: NER
2. Explore new tags by unsupervised learning: LDA, k-means and explore new methods by reading papers.

What we need this week:
1. More verified data (Got)
    - Only recent 10 days data on vaccine
    - More domains such as travel
2. Hierarchy on other domains

Previous work:

1. **Tuning models on full data set:**
   - **Linear SVM -- faster**:
     - Vaccine: accuracy: 0.80, f1 score:0.68
     - Travel: accuracy:0.68, f1 score:0.63
     - Benefits: accuracy:0.71, f1 score:0.53
     - Tax: accuracy:0.73, f1 score:0.47
   - **BERT -- more accurate**:
     - Vaccine: accuracy: 0.86, f1 score:0.86
     - Travel: accuracy:0.72, f1 score:0.72
     - Benefits: accuracy:0.75, f1 score:0.74
     - Tax: accuracy:0.76, f1 score:0.75

2. **Linguistic Features:**
   - Name-entity Recognition (NER)
   - Sentiment Analysis
   - Dependency Structure
   - Semantic Roles

3. **New tagging methods**

Comment:

"Critical missing info:

Fed Govt needs to make it mandatory that Ontario provide detailedinfo

1) Cumulative percent of adult population who have received at least one dose of a COVID-19 vaccine

2) Distribution by PH Unit geographic catchment

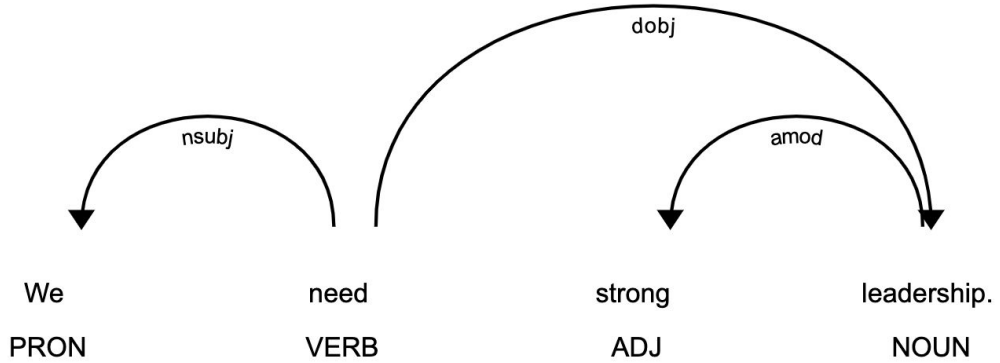3) Naming of # distributed to each health care institution"

NER:
```
[('Fed Govt', 'ORGANIZATION'), ('Ontario', 'LOCATION'), ('COVID-19', 'MISC'), ('PH Unit',
'ORGANIZATION')]
```

Noun Phrase chunks:
```
['info', 'Fed Govt', 'it mandatory', 'Ontario', 'Cumulative percent', 'adult population',
'dose', 'a COVID-19 vaccine', 'Distribution', 'PH Unit geographic catchment', 'Naming',
'each health care institution']
```

```
We need strong leadership.
((We, 0), (need, 1), 'nsubj')
((need, 1), (need, 1), 'ROOT')
((strong, 2), (leadership, 3), 'amod')
((leadership, 3), (need, 1), 'dobj')
((., 4), (need, 1), 'punct')
```

Critical missing info: Fed Govt needs to make it mandatory that Ontario provide detailedinfo 1) Cumulative percent of adult population who have received at least one dose of a COVID-19 vaccine 2) Distribution by PH Unit geographic catchment 3) Naming of # distributed to each health care institution

Sentiment: (1.3, [('missing', -1.2), ('mandatory', 0.3), ('care', 2.2)])

How reliable the shipment is ?? Spending on vaccine is not the solution. Why Canada can not make vaccine ?? Aren't we a develop country ??? No technology at all !! Aren't this TRUDO GOVT. realize that Canada is to vulnerable when is come to safety ?? We need strong leadership. This guy NO GOOD

Sentiment: (-6.8, [('??', -2), ('??', -2), ('???', -2), ('!!', -2), ('vulnerable', -0.9), ('safety', 1.8), ('??', -2), ('strong', 2.3)])

Linguistic feature - Semantic Role Labelling (SRL)

John	met	him	privately	in Ottawa	,	May 3
ARG0	V	ARG1	ARGM-MNR	ARGM-LOC		ARGM-TMP

INPUT: a comment

Critical missing info: Fed Govt needs to make it mandatory that Ontario provide detailedinfo 1) Cumulative percent of adult population who have received at least one dose of a COVID-19 vaccine 2) Distribution by PH Unit geographic catchment 3) Naming of # distributed to each health care institution

OUTPUT: a table that answers "**who did what to whom in which way**"

Agent	Action	Item	Extra
N/A	missing	info	: Fed Govt needs to make it mandatory that Ontario provide detailedinfo 1) Cumulative percent of adult population who have received at least one dose of a COVID-19 vaccine 2) Distribution by PH Unit geographic catchment 3) Naming of # distributed to each health care institution
Fed Govt	needs	make it mandatory that Ontario provide detailedinfo 1) Cumulative percent of adult population who have received at least one dose of a COVID-19 vaccine 2) Distribution by PH Unit geographic catchment 3) Naming of # distributed to each health care institution.	N/A

Tags Generation - Topic Model

Vaccine Tags:

Tags	Count
Getting vaccinated - When / Where / What	7605
Getting vaccinated - Registration / notifications	7282
Vaccine safety (health issues / ingredients / side effects / research)	4732
Data and tracking vaccines	3558
Vaccine effectiveness / delayed dosage	3185
Vaccine strategy: Authorization / Eligibility / Priorities	1475
Vaccines - Other	1264
Life after vaccination	191

Generated Topics:

Topic	Count	Explain
-1_the_to_vaccine_is	17812	Getting vaccinated - When / Where / What
53_vaccin_de_le_je	2605	French comments
61_old_when_years_for	1869	Vaccine for aged people
130_no_johnson_not_vaccine	1163	No information about johnson vaccine
102_vaccine_covid_covid vaccine_for	1091	Register for vaccine
87_vaccination_vaccinated_to_get vaccinated	896	Getting vaccinated - When / Where / What
107_side_effects_side effects_vaccine	888	Vaccine safety (health issues / ingredients / side effects / research)
30_second_dose_2nd_second dose	843	Second dose
55_19_covid 19_covid19_covid	720	Covid-19
60_in_ontario_when_vaccine	704	Vaccine in Ontario
97_book_appointment_to book_an	701	Book appointment

Tags Generation - Topic Model

Travel Tags:

Tags	Count
Restrictions or Requirements	2678
Hotels	2613
Can I enter Canada	2146
Quarantine	1618
Testing	1274
Provincial restrictions	927
Other (Contact / Travel outside / ArriveCan / Mad people)	765
Design / content	693

Generated Topics:

Topic	Count	Explain
-1_to_the_in_and	6213	Can I enter Canada
7_not_no_the_is	1263	No information
0_de_je_pour_la	863	French comments
24_vaccinated_have_if_covid	794	Do I need to quarantine if vaccinated
3_hotel_the_quarantine_days	756	Quarantine days in hotel
35_vaccinated_canada_if_have	598	Enter Canada if vaccinated
4_my_to_and_in	564	Relatives travel
46_my_canadian_to_and	488	Family visit
30_in_to_hotel_calgary	426	Quarantine hotel in different destination
41_canada_to_travel_province	407	Provincial restrictions
31_to_flight_from_the	353	Flight and airport

Will do this week:
1. Add linguistic features to machine learning models.
2. Keep exploring unsupervised learning models to generate new tags.

What we need this week:
1. Do we need to make prediction to domains from comments?
2. Will we name new tags or will you provide a tag pool?

Previous work:

1. **Unsupervised learning on tags generation:**
   - Hierarchical Clustering
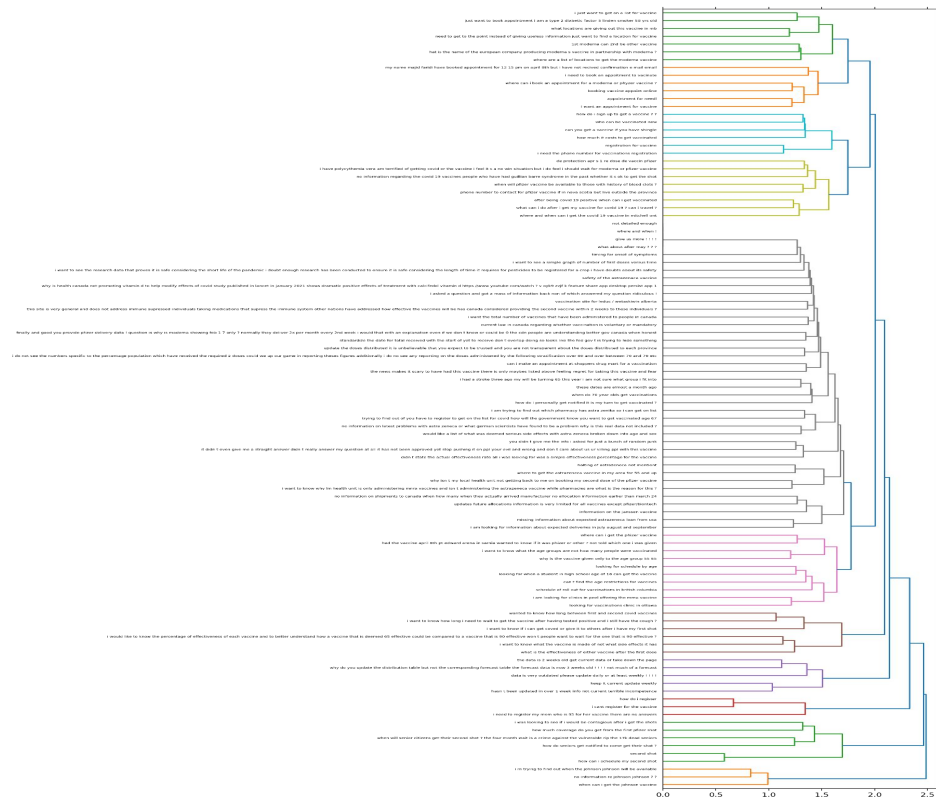   - K-Means Clustering
   - BERT Topic

2. **Supervised learning**
   - Built preprocessing pipeline from previous models
   - Built a classification model(Linear SVC) to predict the domains (Vaccine, Travel, Taxes, Benefits) from comments
   - 93% accuracy

3. **Linguistic features:**
   - Included the linguistic features in the classification models
     - No accuracy improvement with the NP chunks features
   - Extracted semantic roles using SRL BERT
   - Extracted predicates (Verbs) from each comment

Hierarchical Clustering

Main idea: Bottom up approach when forming groups

Pros: No need for a predefined number of clusters

Cons: Computationally expensive

Suggestions: Analyze a reasonably-sized sample instead of the full data set

Figure on the left is from a sample of 1000 rows (Vaccine data)

Hierarchical Clustering

C1 ['no information re johnson johnson ? ?', 'i m trying to find out when the johnson johnson will be available', 'when can i get the johnson vaccine', 'i want an appointment for vaccine', 'appointment for need1', 'booking vaccine appoint online', 'where can i book an appointment for a moderna or phyzer vaccine ?', 'i need to book an appoitment to vacinate', 'my name majid faridi have booked appointment for 12 15 pm on april 8th but i have not recived confirmation e mail email']

C2 ['how can i schedule my second shot', 'second shot', 'how do seniors get notified to come get their shot ?', 'when will senior citizens get their second shot ? the four month wait is a crime against the vulnerable rip the 17k dead seniors', 'how much coverage do you get from the first pfizer shot', 'i was looking to see if i would be contagious after i got the shots', 'hat is the name of the european company producing moderna s vaccine in partnership with moderna ?', '1st moderna can 2nd be other vaccine', 'where are a list of locations to get the moderna vaccine', 'need to get to the point instead of giving useless information just want to find a location for vaccine', 'what locations are giving out this vaccine in mb', 'just want to book appointment i am a type 2 diabetic factor 5 linden smoker 58 yrs old', 'i just want to get on a list for vaccine']

C3 ['i cant register for the vaccine', 'how do i register', 'i need to register my mom who is 95 for her vaccine there are no answers']

C4 ['hasn t been updated in over 1 week info not current terrible incompetence', 'keep it current update weekly', 'why do you update the distribution table but not the corresponding forecast table the forecast data is now 3 weeks old ! ! ! ! not much of a forecast', 'the data is 2 weeks old get current data or take down the page', 'data is very outdated please update daily or at least weekly ! ! ! !']

C5 ['i want to know what the vaccine is made of not what side effects it has', 'i would like to know the percentage of effectiveness of each vaccine and to better understand how a vaccine that is deemed 65 effective could be compared to a vaccine that is 90 effective won t people want to wait for the one that is 90 effective ?', 'what is the effectiveness of either vaccine after the first dose', 'i want to know how long i need to wait to get the vaccine after having tested positive and i still have the cough ?', 'wanted to know how long between first and second covid vaccines', 'i want to know if i can get coved or give it to others after i have my first shot']

C6 ['looking for vaccinations clinic in ottawa', 'i am looking for clinics in peel offering the mrna vaccine', 'looking for when a student in high school age of 18 can get the vaccine', 'looking for schedule by age', 'can t find the age restrictions for vaccines', 'schedule of roll out for vaccinations in british columbia', 'why is the vaccine given only to the age group 55 65', 'i want to know what the age groups are not how many people were vaccinated', 'had the vaccine april 8th pt edward arena in sarnia wanted to know if it was phizer or other ? not told which one i was given', 'where can i get the phizer vaccine']

C7 ['i am looking for information about expected deliveries in july august and september', 'missing information about expected astrazeneca loan from usa', 'updates future allocations information is very limited for all vaccines except pfizer/biontech', 'no information on shipments to canada when how many when they actually arrived manufacturer no allocation information earlier than march 24', 'information on the janssen vaccine', 'i want to know why 1m health unit is only administering mnra vaccines and isn t administering the astrazeneca vaccine while pharmacies are what is the reason for this ?', 'why isn t my local health unit not getting back to me on booking my second dose of the pfizer vaccine', 'where to get the astrazeneca vaccine in my area for 55 and up', 'halting of astrazeneca not mentiont', 'it didn t even give me a straight answer didn t really answer my question at all it has not been approved yet stop pushing it on ppl your evil and wrong and don t care about us ur killing ppl with this vaccine', 'you didn t give me the info i asked for just a bunch of random junk', 'didn t state the actual effectiveness rate all i was looking for was a simple effectiveness percentage for the vaccine', 'would like a list of what was deemed serious side effects with astra zeneca broken down into age and sex', 'no information on latest problems with astra zeneca or what german scientists have found to be a problem why is this real data not included ?', 'trying to find out of you have to register to get on the list for covid how will the government know you want to get vaccinated age 67', 'i am trying to find out which pharmacy has astra zenika so i can get on list', 'these dates are almost a month ago', 'i had a stroke three ago my will be turning 65 this year i am not sure what group i fit into', 'when do 70 year olds get vaccinations', 'how do i presently get notified it is my turn to get vaccinated ?', 'the news makes it scary to have had this vaccine there is only maybes listed above feeling regret for taking this vaccine and fear', 'can i make an appointment at shoppers drug mart for a vaccination', 'i do not see the numbers specific to the percentage population which have received the required 2 doses could we up our game in reporting theses figures additionally i do no see any reporting on the doses administered by the following stratification over 80 and over between 70 and 79 etc', 'update the doses distributed it is unbelievable that you expect to be trusted and you are not transparent about the doses distributed to each province', 'standardize the date for total received with the start of yet to receive don t overlap doing so looks like the fed gov t is trying to hide something', 'finally and good you provide pfizer delivery data ! question is why is moderna showing feb 1 7 only ? normally they deliver 2x per month every 2nd week i would that with an explanation even if we don t know or could be 0 the cdn people are understanding better gov canada when honest', 'current law in canada regarding whether vaccination is voluntary or mandatory', 'i want the total number of vaccines that have been administered to people in canada', 'this site is very general and does not address immune supressed individuals taking medications that supress the immune system other nations have addressed how effective the vaccines will be has canada considered providing the second vaccine within 2 weeks to these individuals ?', 'vaccination site for leduc / wetaskiwin alberta', 'where and when !', 'not detailed enough', 'give us more ! ! ! !', 'what about after may ? ? ?', 'timing for onset of symptoms', 'i want to see a simple graph of number of first doses versus time', 'i want to see the research data that proves it is safe considering the short life of the pandemic i doubt enough research has been conducted to ensure it is safe considering the length of time it requires for pesticides to be registered for a crop i have doubts about its safety', 'safety of the astrazeneca vaccine', 'why is health canada not promoting vitamin d to help modify effects of covid study published in lancet in january 2021 shows dramatic positive effects of treatment with calcifedol vitamin d https //www youtube com/watch ? v oyk9 zvjf k feature share app desktop persist app 1', 'i asked a question and got a mass of information back non of which answered my question ridiculous !']

C8 ['what can i do after i get my vaccine for covid 19 ? can i travel ?', 'after being covid 19 positive when can i get vaccinated', 'where and when can i get the covid 19 vaccine in mitchell ont', 'phone number to contact for pfizer vaccine if in nova scotia but live outside the province', 'when will pfizer vaccine be available to those with history of blood clots ?', 'i have polycythemia vera am terrified of getting covid or the vaccine i feel it s a no win situation but i do feel i should wait for moderna or pfizer vaccine', 'de protection apr s 1 re dose de vaccin pfizer', 'no information regarding the covid 19 vaccines people who have had guillian barre syndrome in the past whether it s ok to get the shot']

C9 ['i need the phone number for vaccinations registration', 'registration for vaccine', 'who can be vaccinated now', 'how do i sign up to get a vaccine ? ?', 'can you get a vaccine if you have shingle', 'how much it costs to get vaccinated']

C0 ['where can i book an appointment for a moderna or phyzer vaccine ?', 'how do i sign up to get a vaccine ? ?', 'how do seniors get notified to come get their shot ?']

K-Means Clustering

Main idea: Top down approach when forming groups

Pros: Way faster than Hierarchical Clustering

Cons: Need to know the number of clusters beforehand

Suggestions: Have a standard K value that tells you if a topic is relevant. For example, K=10 in the Vaccine data captures known topics such as vaccine registration, second dose, more info needed about vaccine, etc.

If a new topic comes up while K=10, this means that this new topic has become relevant and can be considered its own tag.

K-Means Clustering

Looking at individual words per cluster:

```
Top terms per cluster:

Cluster 0 words:{'bookings', 'sign', 'information', 'old', 'years', 'make'}

Cluster 1 words:{'receiving', 'second', 'want', 'register', 'years'}

Cluster 2 words:{'types', 'vaccine', 'different', 'know', 'used'}

Cluster 3 words:{'appointment', 'bookings', 'want'}

Cluster 4 words:{'age', 'vaccine', 'group'}

Cluster 5 words:{'dose', 'receiving', 'astrazeneca', 'blood', 'information'}

Cluster 6 words:{'effects', 'dose', 'second', 'pfizer'}

Cluster 7 words:{'register', 'covid', 'bookings', 'appointment'}

Cluster 8 words:{'johnson', 'information'}

Cluster 9 words:{'register', 'need', 'vaccine', 'want'}
```

K-Means Clustering

Looking at word pairs per cluster:

```
Top terms per cluster:

Cluster 0 words:{'book vaccin', 'year old', 'second shot', 'year old vaccin', 'age vaccin', 'appoint vaccin', 'sign vaccin', 'inform date', 'old vaccin', 'make appoint'}

Cluster 1 words:{'regist vaccin', 'year old', 'moderna vaccin', 'vaccin covid', 'second shot', 'want know', 'receiv vaccin', 'know vaccin', 'second vaccin', 'second dose'}

Cluster 2 words:{'use canada', 'differ vaccin use', 'type vaccin', 'type vaccin use', 'know vaccin use', 'vaccin use clinic', 'vaccin use site', 'use clinic', 'vaccin use canada', 'vaccin use
'}

Cluster 3 words:{'book appoint vaccin', 'tri book appoint', 'book appoint', 'want book appoint', 'appoint vaccin', 'appoint covid', 'book appoint covid', 'book appoint covid vaccin', 'want bo
ok', 'appoint covid vaccin'}

Cluster 4 words:{'vaccin age', 'look age group', 'vaccin age group', 'age group elig', 'age group', 'group elig', 'know age', 'group vaccin', 'know age group', 'age group vaccin'}

Cluster 5 words:{'dose astrazeneca vaccin', 'receiv astrazeneca vaccin', 'effect astrazeneca vaccin', 'dose astrazeneca', 'receiv astrazeneca', 'astrazeneca vaccin', 'effect astrazeneca', 'bl
ood clot', 'inform astrazeneca vaccin', 'inform astrazeneca'}

Cluster 6 words:{'book pfizer', 'second dose pfizer vaccin', 'effect pfizer vaccin', 'dose pfizer', 'book pfizer vaccin', 'pfizer vaccin', 'second dose pfizer', 'effect pfizer', 'dose pfizer
vaccin', 'second dose'}

Cluster 7 words:{'regist covid vaccin', 'receiv covid vaccin', 'covid vaccin', 'appoint covid', 'make appoint covid vaccin', 'receiv covid', 'regist covid', 'appoint covid vaccin', 'book covi
d vaccin', 'book covid'}

Cluster 8 words:{'johnson johnson', 'johnson vaccin', 'johnson johnson vaccin avail', 'inform johnson johnson', 'info johnson johnson', 'inform johnson', 'inform johnson johnson vaccin', 'joh
nson vaccin avail', 'johnson johnson vaccin', 'vaccin avail'}

Cluster 9 words:{'regist vaccin', 'vaccin covid', 'look regist vaccin', 'want regist', 'look regist', 'tri regist vaccin', 'regist vaccin covid', 'want regist vaccin', 'need regist', 'need re
gist vaccin'}
```

What we have got in the first 4 weeks:

1. **Supervised learning - Tags/Domains Classification**
   Based on what information(webs) Canada.ca already has:

   Examples:
   "*What is delivery schedule for July Pfizer moderne AstraZeneca and j&j*"
   → Domain: "Vaccine"
   → Tags: "Data and tracking vaccines"
   → Vaccines for COVID-19: Shipments and deliveries

   Models:
   SVM, CNN, BERT

   Results:
   93% accuracy on domains and 73% - 86% accuracy on tags

2.   **Unsupervised learning - Comment Clustering**
      Based on what information users need that Canada.ca may not have:

      Examples:
        - One group of 36 comments in **Travel**:
        - Key words: "dog", "pets", "walk with", "quarantine", etc
        - Example comment:
               "I was wondering if I can walk my dog off of my property during quarantine."
        - Tags: "Quarantine"
        - Link: Driving to Canada: COVID-19 testing for travellers (No information)

        - Another group of 324 comments in **Vaccine**:
        - Key words: "second", "2nd", "second vaccines", etc
        - Example comment:
               "What about the second dose of madorna vaccine"
        - Tags: "Vaccines - Other"
        - Link: Pfizer-BioNTech COVID-19 vaccine: What you should know (Incorrect website)

      Models: BERT topic, K-Means, Hierarchical Clustering

      Purpose: Make new tags from the clustering groups

What we have got in the first 4 weeks:

3. Linguistic features:

- Name-entity Recognition (NER)
- Sentiment Analysis
- Dependency Structure
- Semantic Roles

Purpose:
- Improve the performance of tags classification models
- Provide insights to new tags creating

What we have done in the first 4 weeks:

Will do this week:

1. Add linguistic features to machine learning models.

2. Keep exploring unsupervised learning models

Update:
1.   Error Analysis
2.   Clustering analysis
     a.   Good vs. need new label
     b.   New label pool
3.   Linguistic feature Embedding
4.   Topic Modelling

Will do this week:

1.   Finalize model

2.   Project report

Naive Bayes vs. BERT Error Analysis

One vs. Rest Multinomial Naive Bayes Classification

Pros: very fast!

Cons: not very accurate

Accuracy: 0.549

F1 score: 0.431

BERTforSequence Classification

Pros: very accurate!

Cons: data-hungry, slow to train

Accuracy: 0.717

F1 score: 0.716

Testing on Travel Data

Comment	True Tags	NB Predictions	BERT Predictions
is testing and quarantining required if fully vaccinated both doses coming back from the us via land crossing ?	Can I enter Canada	Restrictions or Requirements	Restrictions or Requirements
i would like to have you please address how fully documented vaccinated people property owners can travel into atlantic canada by car	Can I enter Canada	Can I enter Canada	Can I enter Canada
this is a concentration camp	Hotels	Can I enter Canada	Hotels
looking for quarantine regulations when arriving in canada no symptoms	Quarantine	Restrictions or Requirements	Quarantine
study permit for university of manitoba international students	Restrictions or Requirements	Provincial restrictions	Restrictions or Requirements
canadian citizen living in usa and am am vaccinated do i need to quarantine ?	Quarantine	Restrictions or Requirements	Quarantine

Topic Modelling

(share screen to show local deck)