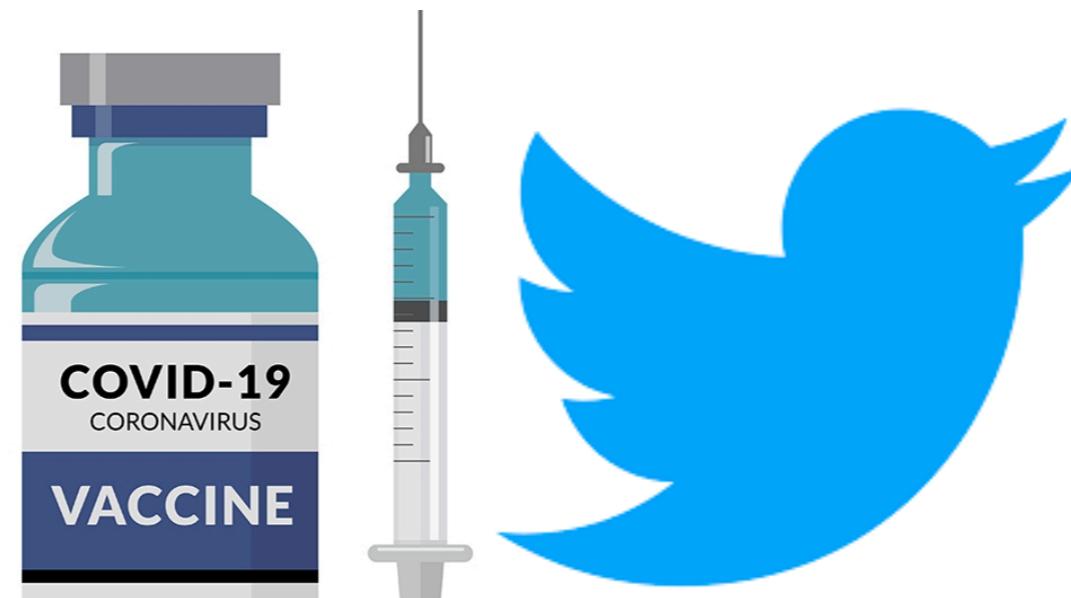


# Identifying conversation themes regarding covid-19 vaccines on Twitter



Ignasi Sols

# Introduction

Official global covid-19 deaths



5.1m

The  
Economist

November 11th, 2021



**Business impact:** Provide a better understanding of the current conversation on Twitter regarding covid-19 vaccines.

## **Data Science Goal:**

**Machine learning unsupervised model that find topics and clusters for covid-19 vaccine tweets.**

**Data:** Most recent 10,000 tweets for the query ‘Covid vaccine’ (all tweets are From November 7th, in English), scrapped.

**Tools:** Pandas, snscreape, spacy, sklearn.

## Data Science path:

- Pre-process the corpus.
- Matrix vectorization (TfidfVectorizer) —> 1000 features (terms)
- Topic modeling (NMF) —> 5 topics
- K-means clustering —> 7 clusters

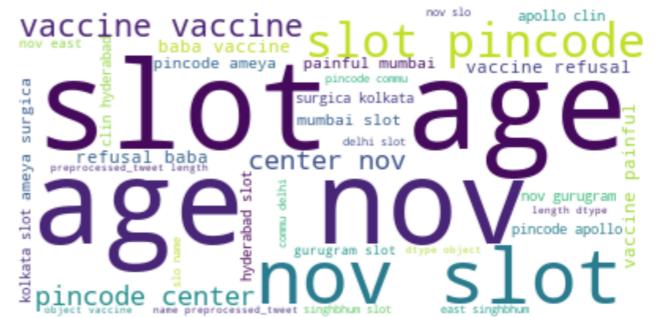
## **Assumptions/Risks**

- All Tweets are from the same day - our algorithm could be too influenced by events that happened this day.
- All Tweets are in English, limits the scope.

# 5 topics

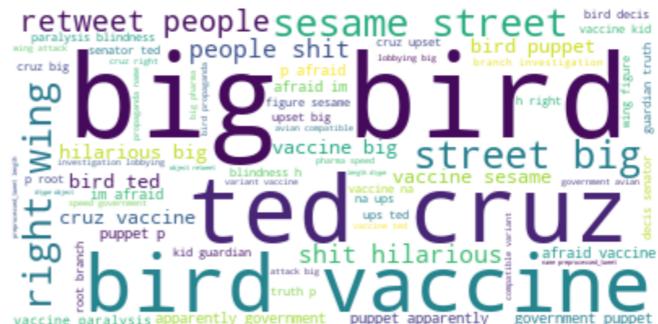
## India:

slot, pincode, nov, center, age, **mumbai**, ameya, nh, surgical, general, pmc, pune, multispeci, ho, **kolkata**, hyderabad, site, gurugram, nagar, singhbhum, laxmi, uphc, hospital, east, community, **delhi**, kapur, dr, burari, nurse



# Sesame Street Bird gets vaccinated

bird, big, **cruz**, **ted**, sesame, street, **propaganda**, character, right, **government**, wing, guardian, attack, icon, delight, child, **gop**, **republican**, old, kid, sen, winger, elmo, cnn, anti, pharma, smartnews, bigbird, today, new



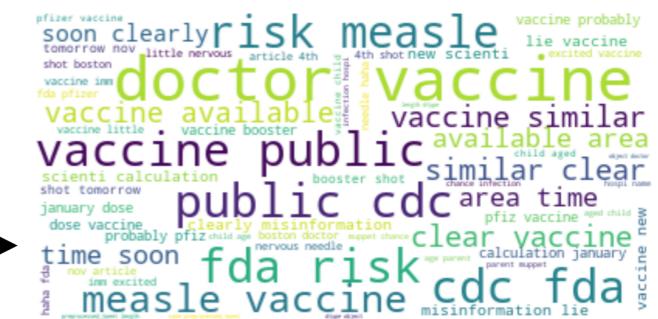
# Vaccine mandates:

**mandate, biden**, court, federal, business, administration, appeal, rule, worker, surgeon, general, admin, employee, ruling, test, new, temporarily, private, joe, amp, large, news, company, governor, president, prepared, staff, testing, mask, white



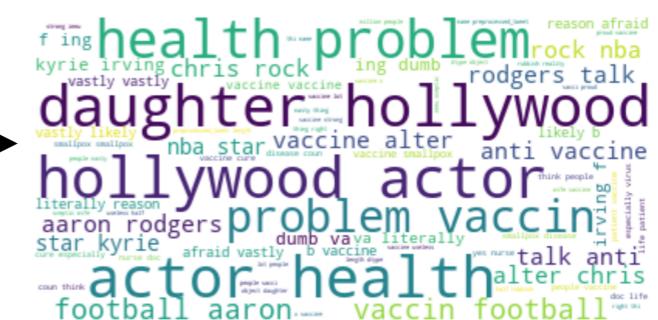
# **Shots for children**

**Kid, child, shot, school, pfizer, booster, risk, age, flu, old, parent, young, dose, time, like, amp, today, cdc, safe, polio, day, clinic, week, family, trial, effect, long, available, health, month**

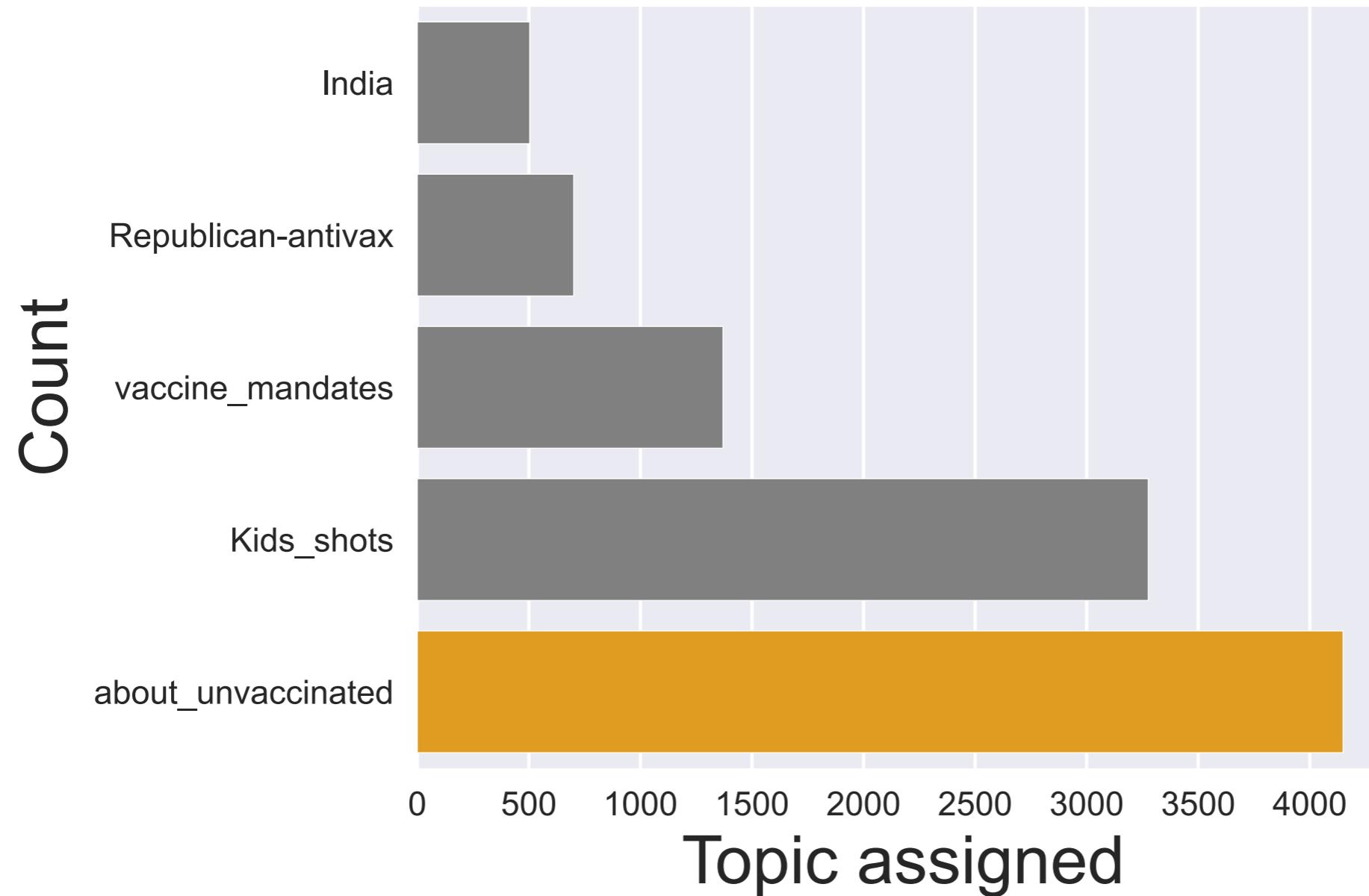


# Discussion about the unvaccinated

people, death, like, amp, virus, immunity, risk, effect, unvaccinated, case, time, million, long, hospital, effective, good, way, booster, thing, bad, flu, safe, anti, right, reaction, dose, natural, rate, likely, fact



## Count of tweets assigned to each topic



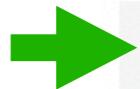
## K-means clustering: interpretation

| Cluster | India   | about_unvaccinated | Sesame_street_bird | Kids_shots | vaccine_mandates |
|---------|---------|--------------------|--------------------|------------|------------------|
| 0       | 0.00000 | 0.00000            | 0.00000            | 0.000000   | 0.126655         |

“White House Chief of Staff Ron Klain on Sunday said he is confident **President Biden**’s decision to **mandate** the COVID-19 **vaccine** or weekly testing at large businesses will hold up in court over”

## K-means clustering: interpretation

| Cluster | India   | about_unvaccinated | Sesame_street_bird | Kids_shots | vaccine_mandates |
|---------|---------|--------------------|--------------------|------------|------------------|
| 0       | 0.00000 | 0.00000            | 0.00000            | 0.000000   | 0.126655         |
| 1       | 0.00000 | 0.00319            | 0.00000            | 0.065240   | 0.000200         |



“Pfizer’s COVID-19 vaccine for **kids** five to 11 is now in Alabama. At least they will get the anti-coagulant to hopefully stave off deadly blood clots and heart inflammation, unlike adults”

“Ivermectin won a Nobel prize for something totally unrelated to covid. There’s **no evidence that the vaccine is deadly to kids**. Y’all **just see shit and repeat** it without any fact checking at all, huh?”

## K-means clustering: interpretation

| Cluster | India   | about_unvaccinated | Sesame_street_bird | Kids_shots | vaccine_mandates |
|---------|---------|--------------------|--------------------|------------|------------------|
| 0       | 0.00000 | 0.00000            | 0.00000            | 0.000000   | 0.126655         |
| 1       | 0.00000 | 0.00319            | 0.00000            | 0.065240   | 0.000200         |
| 2       | 0.23419 | 0.00000            | 0.00000            | 0.000000   | 0.000000         |



PUNE: 154570 slots found for age 18+!

Nov 07, 2021: 930 slots found in pincode 411039 at 6 centers

Nov 07, 2021: 197 slots found in pincode 411042 at PMC P

Kamalnayan Multispeci Ho

# K-means clustering: interpretation

| Cluster | India   | about_unvaccinated | Sesame_street_bird | Kids_shots | vaccine_mandates |
|---------|---------|--------------------|--------------------|------------|------------------|
| 0       | 0.00000 | 0.00000            | 0.00000            | 0.000000   | 0.126655         |
| 1       | 0.00000 | 0.00319            | 0.00000            | 0.065240   | 0.000200         |
| 2       | 0.23419 | 0.00000            | 0.00000            | 0.000000   | 0.000000         |
| 3       | 0.00000 | 0.00000            | 0.19061            | 0.000000   | 0.000000         |



“The **Bird** knows”

“Is the COVID vaccine validated for **birds**? ”

"Cancun @tedcruz, why don't you pick on someone with a lower IQ level that is more in line with yours? I Ted Cruz leads right-wing attack on Big Bird for promoting 'propaganda' after the Sesame Street character got the COVID-19 vaccine"



## K-means clustering: interpretation

| Cluster | India   | about_unvaccinated | Sesame_street_bird | Kids_shots | vaccine_mandates |
|---------|---------|--------------------|--------------------|------------|------------------|
| 0       | 0.00000 | 0.00000            | 0.00000            | 0.000000   | 0.126655         |
| 1       | 0.00000 | 0.00319            | 0.00000            | 0.065240   | 0.000200         |
| 2       | 0.23419 | 0.00000            | 0.00000            | 0.000000   | 0.000000         |
| 3       | 0.00000 | 0.00000            | 0.19061            | 0.000000   | 0.000000         |
| 4       | 0.00000 | 0.00920            | 0.00000            | 0.007290   | 0.002590         |

"No, COVID-19 vaccines don't cause HIV, AIDS or cancer as **social media** posts claim\n\nIt is one of the most popular but completely untrue stories of the week, shared widely on social media. Here are the facts on that story and more:

**Other top samples suggest possible cluster related to social media?**

## K-means clustering: interpretation

| Cluster | India   | about_unvaccinated | Sesame_street_bird | Kids_shots | vaccine_mandates |
|---------|---------|--------------------|--------------------|------------|------------------|
| 0       | 0.00000 | 0.00000            | 0.00000            | 0.000000   | 0.126655         |
| 1       | 0.00000 | 0.00319            | 0.00000            | 0.065240   | 0.000200         |
| 2       | 0.23419 | 0.00000            | 0.00000            | 0.000000   | 0.000000         |
| 3       | 0.00000 | 0.00000            | 0.19061            | 0.000000   | 0.000000         |
| 4       | 0.00000 | 0.00920            | 0.00000            | 0.007290   | 0.002590         |
| 5       | 0.00000 | 0.05842            | 0.00000            | 0.000000   | 0.000000         |

“Vaccines do not protect anyone from getting covid. Vaccinated people are still getting it and dying. Vaccines only help reduce symptoms. Educate yourself before making an ignorant statement”

“People are afraid to get the COVID vaccine but if you turned your clocks back for daylight savings you are also a sheep. Nothing is real. Everything is made up”      ???

## K-means clustering: interpretation

| Cluster | India   | about_unvaccinated | Sesame_street_bird | Kids_shots | vaccine_mandates |
|---------|---------|--------------------|--------------------|------------|------------------|
| 0       | 0.00000 | 0.00000            | 0.00000            | 0.000000   | 0.126655         |
| 1       | 0.00000 | 0.00319            | 0.00000            | 0.065240   | 0.000200         |
| 2       | 0.23419 | 0.00000            | 0.00000            | 0.000000   | 0.000000         |
| 3       | 0.00000 | 0.00000            | 0.19061            | 0.000000   | 0.000000         |
| 4       | 0.00000 | 0.00920            | 0.00000            | 0.007290   | 0.002590         |
| 5       | 0.00000 | 0.05842            | 0.00000            | 0.000000   | 0.000000         |
| 6       | 0.00000 | 0.00000            | 0.00000            | 0.141925   | 0.000000         |

“Kids should not get the covid vaccine. The threat to children is less than the common flu.”

“1,000+ Covid vaccines given at clinic for kids”

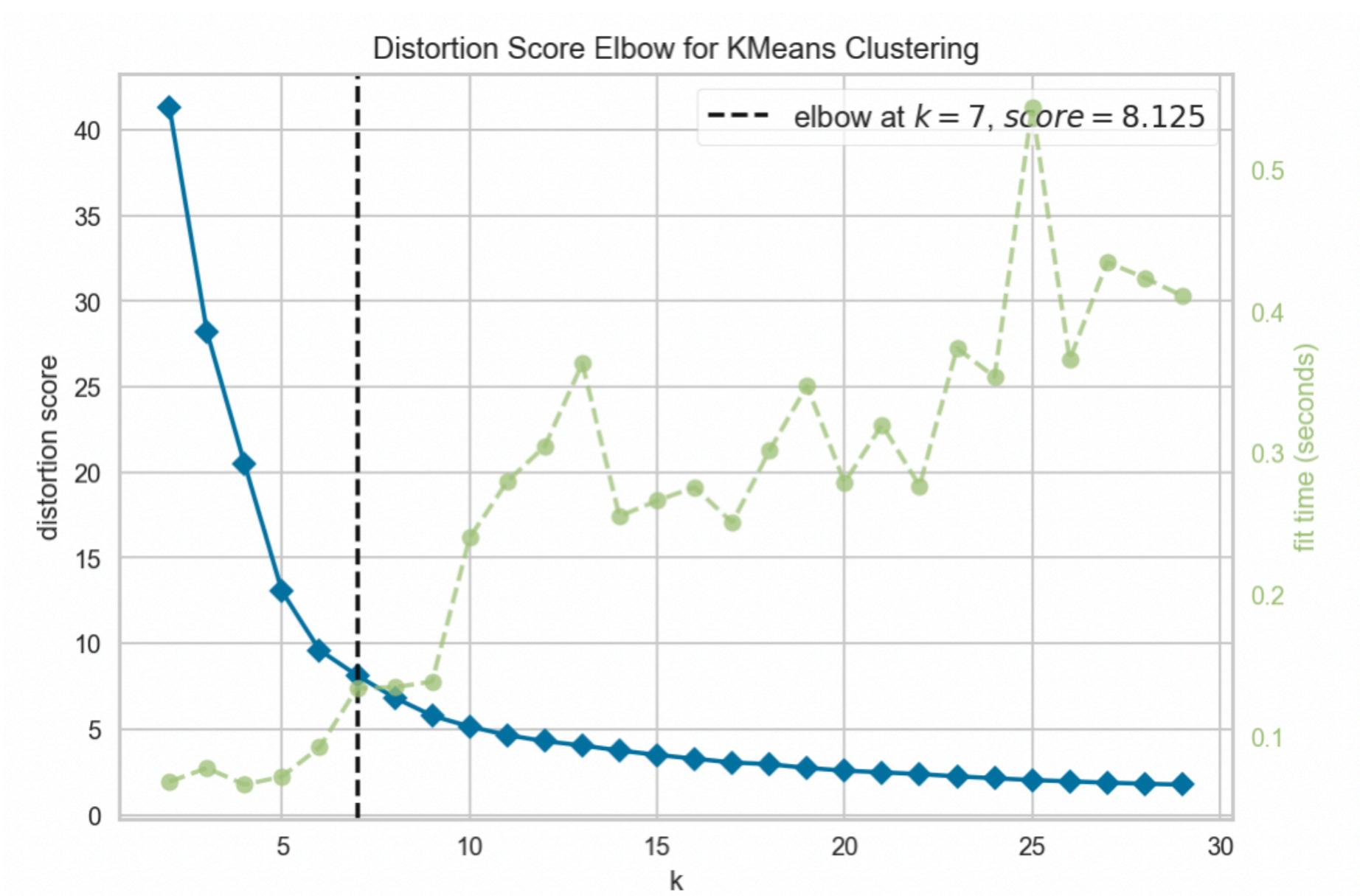
## **Future directions:**

- Work with CorEx, apply k-means clustering and DBSCAN.
- Increase the number of tweets.

**Thanks!**

# Appendix:

## Choosing k for k-means clustering

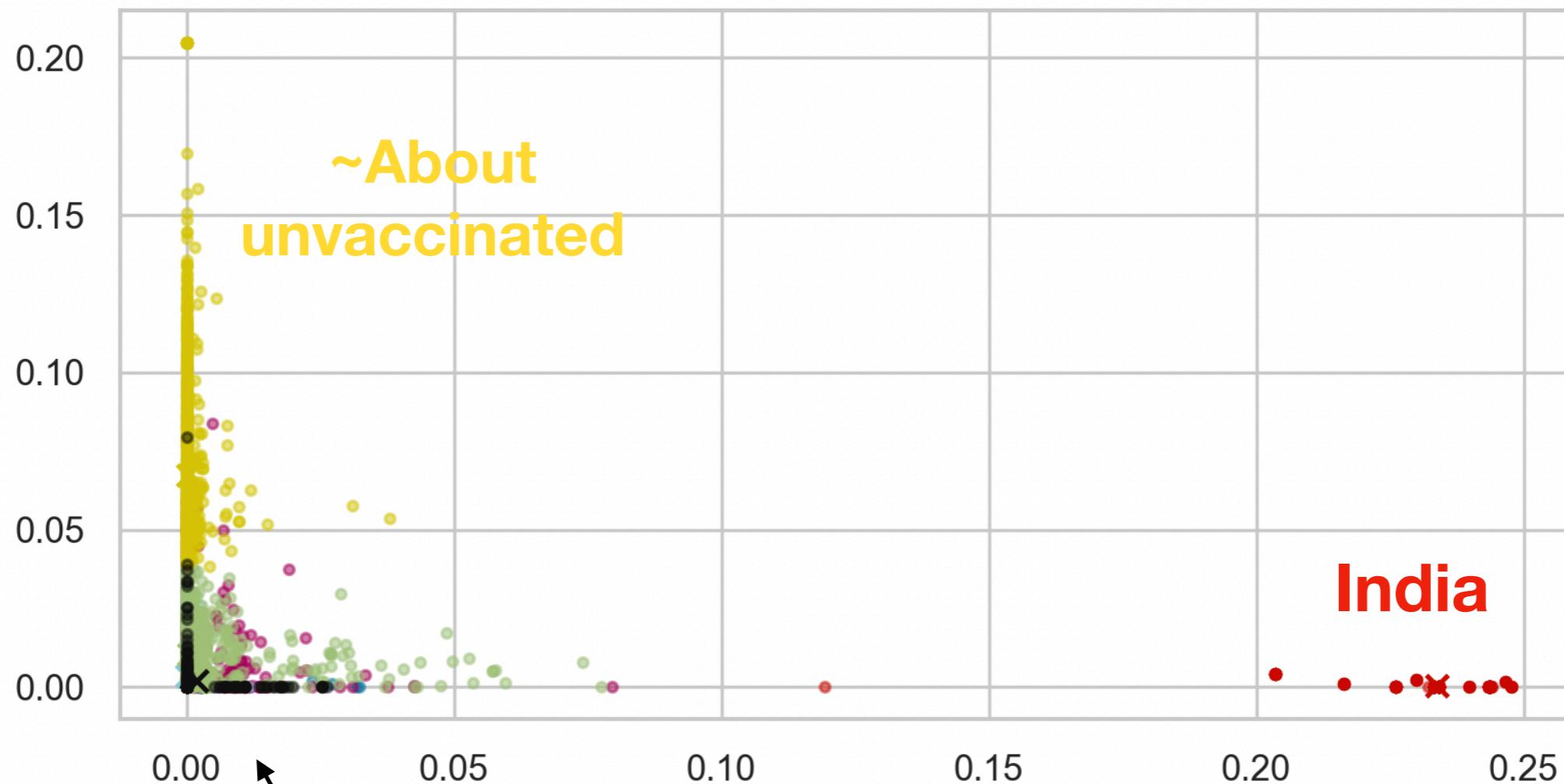


## Appendix:

### NMF: seven clusters identified

PUT a title. EXPLAIN IT.LEGEND,: who are the yellow, green, black, red? Overlap?

Tweet examples for each cluster.



**Kids  
shots**

# **Appendix:**

## **Corex - seven topics identified (with anchors)**

|                      |   |
|----------------------|---|
| <b>India:</b>        | slot, pincode, nov, center, age, mumbai, ameya, nh, surgical, pmc                               |
| <b>Ted Cruz:</b>     | bird, cruz, big, ted, sesame, street, propaganda, character, wing, icon                         |
| <b>Children</b>      | child, kid, elementary, classroom, school, young, educator, approval, key, student              |
| <b>Mandate</b>       | mandate, court, biden, appeal, federal, surgeon, business, pregnant, rule, mum                  |
| <b>Aaron Rodgers</b> | aaron, rodgers, snl, partnership, prevea, comment, bay, green, packers, davidson                |
| <b>Boosters</b>      | effectiveness, percent, dramatic, decline, study, month, efficacy, veteran, additional, booster |
| <b>Bmj/Israel</b>    | whistle, integrity, researcher, data, trial, bmj, pfizer, issue, whistleblower, journal         |