



Coronavirus Tweets Analysis

-Xufei Li



Goal:

- ▶ To know what people are talking about during the pandemic

Data:

- ▶ From Kaggle: 41,000+ tweets
- ▶ Twitter API: 5,000 new tweets

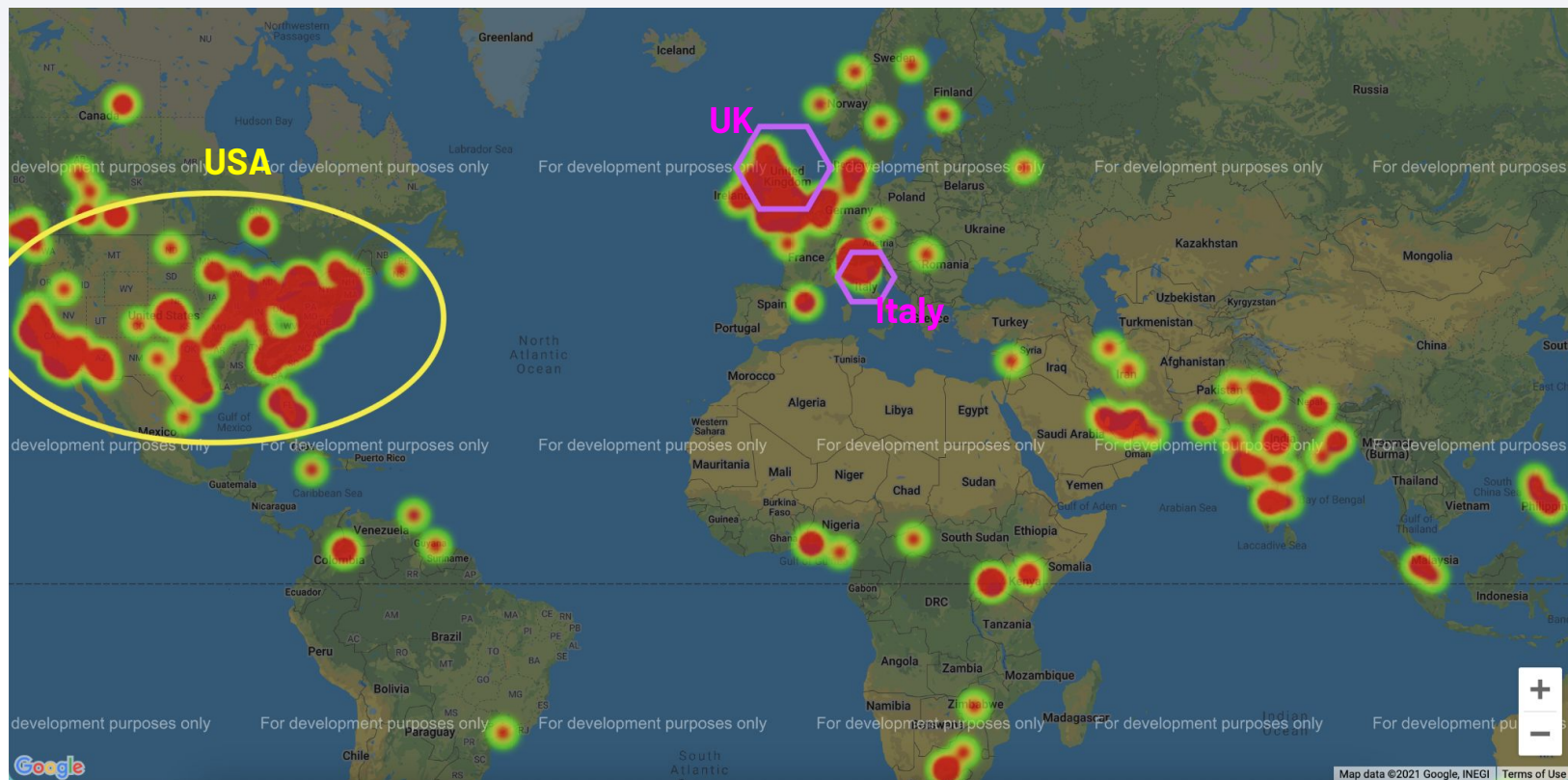
Solution path:

- ▶ Top topics people are discussing on March 2020(beginning of pandemic)
- ▶ Compare by countries
- ▶ Group tweets to different group for further analysis
- ▶ Top topics for May 2021(current)



Data Cleaning & Heatmap

Filter & Clean Kaggle Data - March, 2020 & English Tweets only - 21,000+ data points



Text Preprocessing

Preprocessing & Tokenization:

- ▶ 're' library
- ▶ Stop words
- ▶ TfidfVectorizer



Topic Modeling - NMF



Topic 1	Panic buying & stocking food
Topic 2	Grocery & retail
Topic 3	Online shopping & delivery
Topic 4	Gas price decline
Topic 5	Toilet paper hoarding
Topic 6	Food shortage (empty shelves)
Topic 7	Mask & Hand sanitizer shortage
Topic 8	Thanks ALL front-line workers
Topic 9	Covid impact on business & economy
Topic 10	Stay at home

A fun tweet:

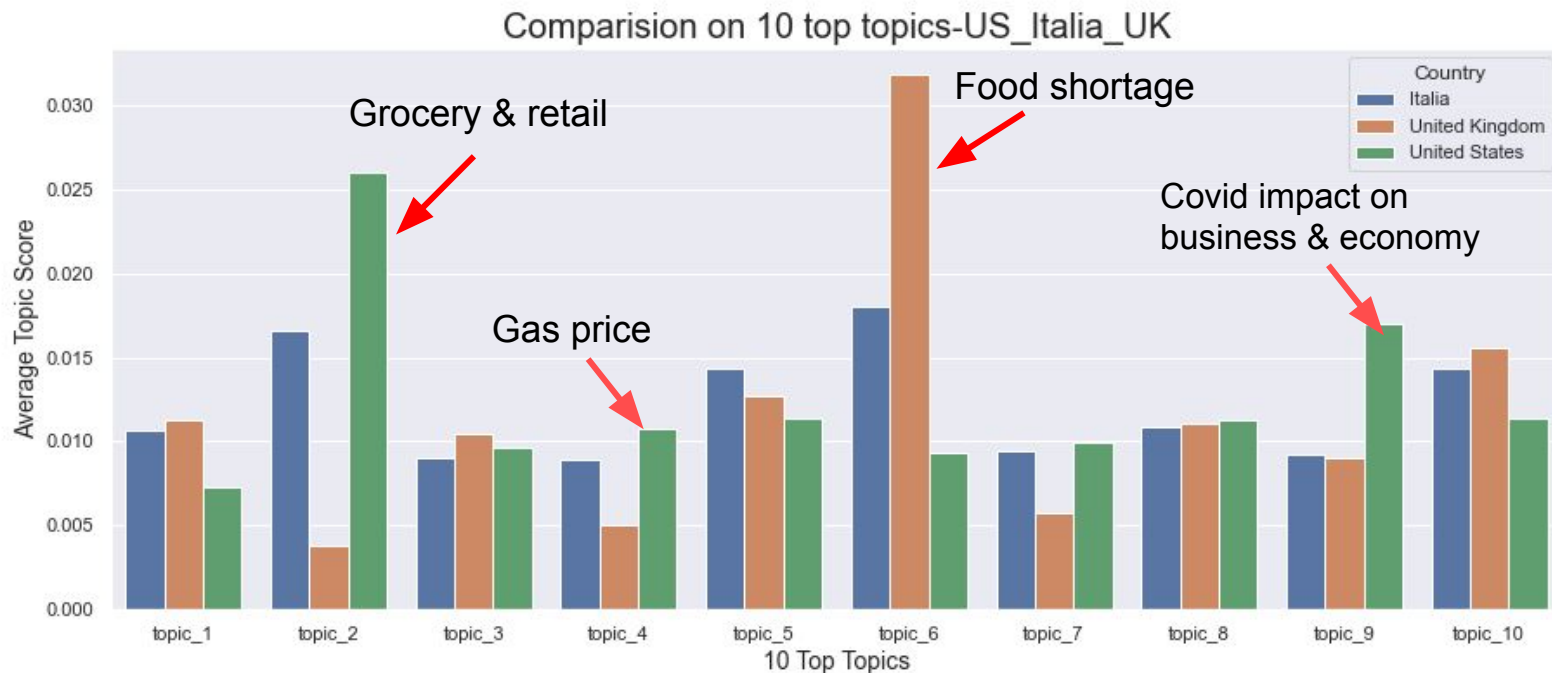
[Toilet paper hoarding is sh**.](#)

- 4 cases toilet paper from Costco
- Family of 4
- Quarantine for 14 days

= 182 sh**/day

So, calm down!

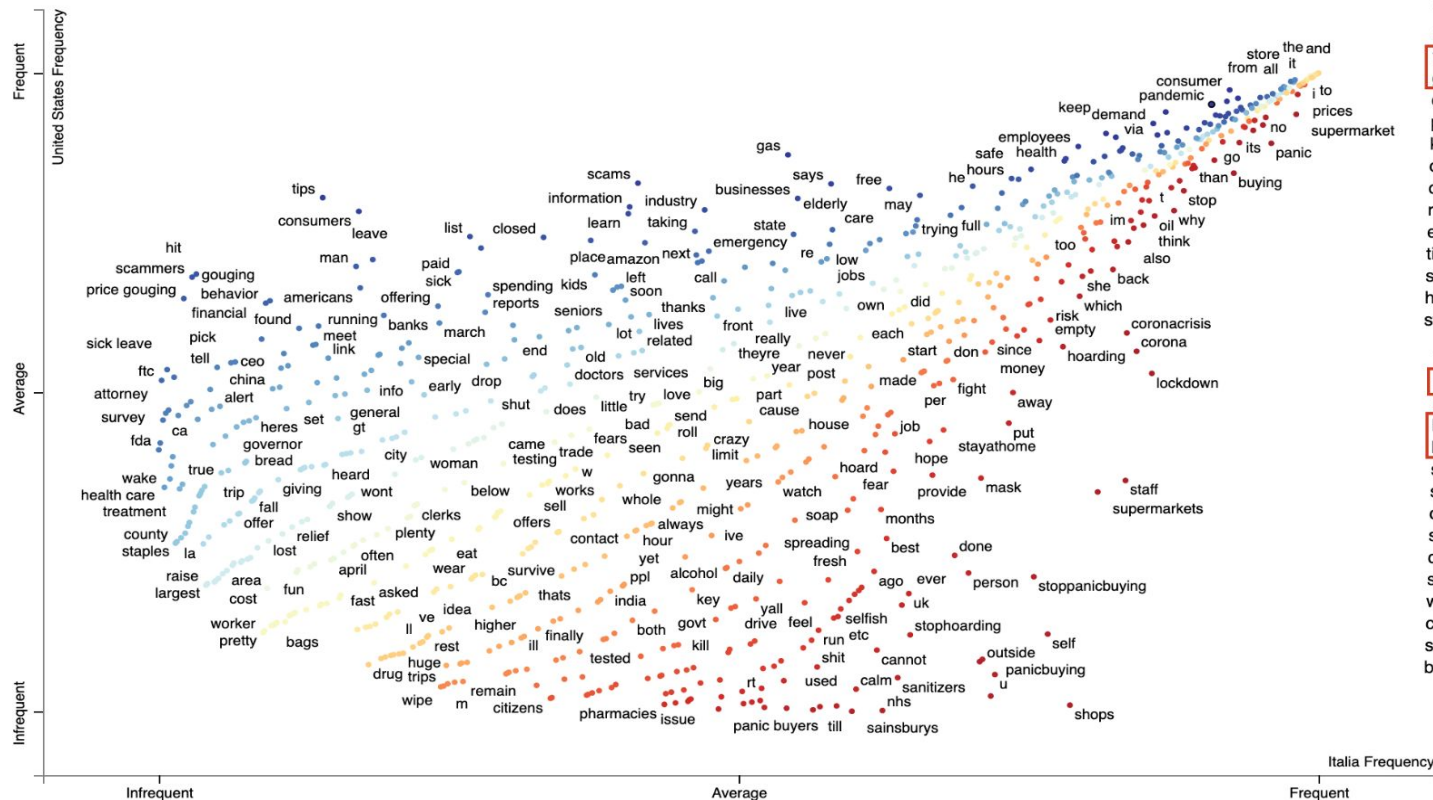
Visual - comparison among countries



Visual - Scattertext

United States vs. Italy (March, 2020)

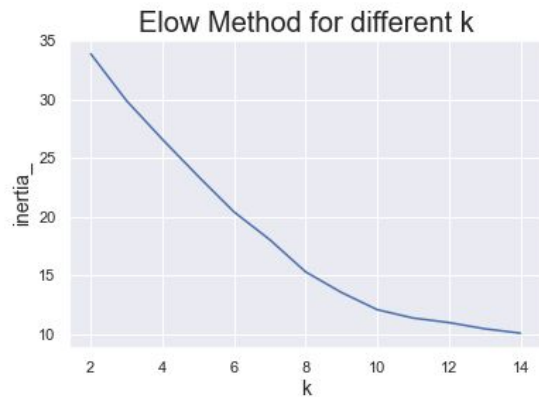
- USA vs. Italy
- 'panic'(39 vs. 99)
- 'toiletpaper'(71 vs. 36)
- sharing words: 'jobs', 'testing'



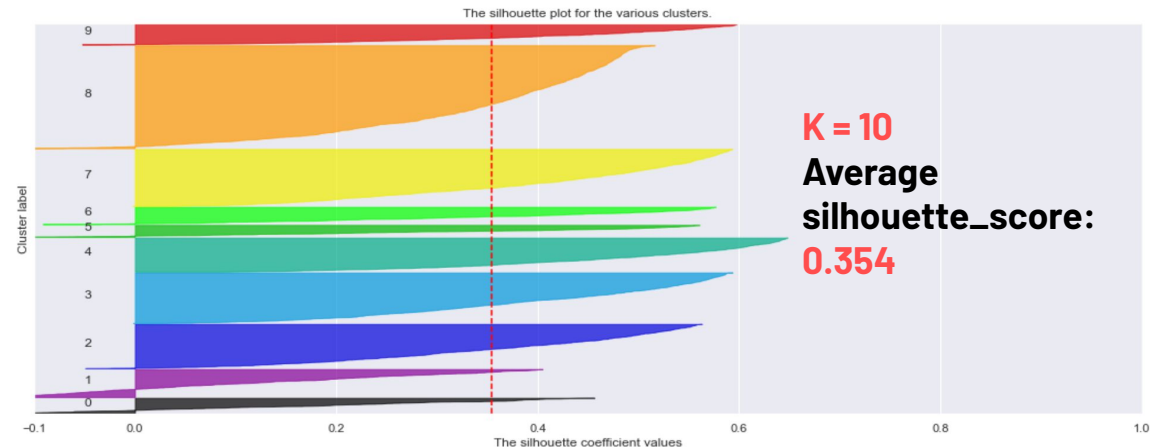
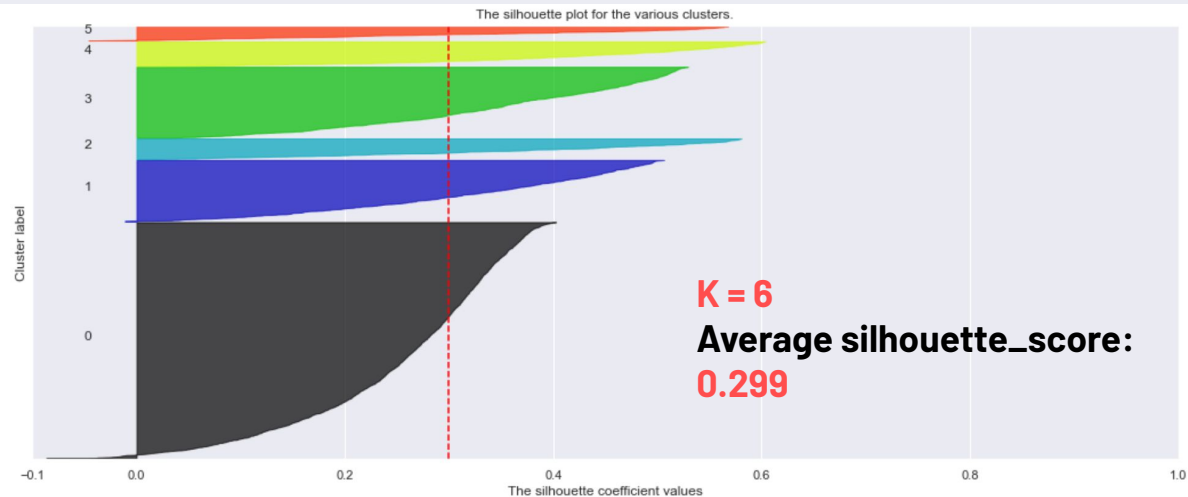
Top United States	Characteristic
consumer	coronavirus
toiletpaper	covid
gas	toiletpaper
demand	coronacrisis
pandemic	coronavirusoutbreak
keep	sanitizer
during	coronaviruspandemic
our	stoppanicbuying
retail	stayathome
employees	socialdistancing
tips	coronavirusupdate
scams	toiletpapercrisis
here	panicbuying
stores	toiletpaperpanic
	stophoarding
	stayhome
	handsanitizer
	realdonaldtrump
	coronaviruslockdown
	pricegouging
	washyourhands
	distancing
	toiletpaperapocalypse
	hoarding
	supermarket
	coronacrisis
	stock
	who
	covid
	self
	because

Top Italia
panic
supermarket
buying
lockdown
staff
shops
corona
supermarkets
coronacrisis
stoppanicbuying
person
stophoarding
ago
uk
selfish
run
etc
cannot
sanitizers
nhs
calm
used
panicbuyers
till
sainsburys
shops

Clustering



Choose K = 6



Result of clustering:

Top 3 topic for each cluster based on topic score



Cluster 0	'topic_5'	'topic_10'	'topic_8'
Cluster 1	'topic_2'	'topic_8'	'topic_10'
Cluster 2	'topic_4'	'topic_2'	'topic_10'
Cluster 3	'topic_7'	'topic_8'	'topic_10'
Cluster 4	'topic_6'	'topic_10'	'topic_8'
Cluster 5	'topic_9'	'topic_3'	'topic_1'



Topics for May_2021:

(Data is from twitter API, only for May_26_2021 night)



- Vaccine - 'appointment', 'second dose', 'side effects'
- News - 'LEAKED INTERNAL DOCS(fb)'
- News - 'Tennessee anti-vaxxer arrested'
- News - 'COVID OUTBREAK IN MELBOURNE, Australia'

Future work:

“
Ingesting whole month data for May 2021 then apply topic modeling to see the top discussions.



Appendix 1:

Top 6 words for each topic for March, 2020

```
[['people', 'stock', 'need', 'buying', 'panic', 'food'],  
 ['people', 'going', 'stores', 'retail', 'grocery', 'store'],  
 ['stay', 'amazon', 'delivery', 'home', 'online', 'shopping'],  
 ['amid', 'low', 'price', 'gas', 'oil', 'prices'],  
 ['toiletpaperpanic', 'rolls', 'roll', 'toiletpaper', 'paper', 'toilet'],  
 ['just', 'like', 'local', 'shelves', 'people', 'supermarket'],  
 ['use', 'masks', 'hands', 'make', 'hand', 'sanitizer'],  
 ['delivery', 'employees', 'staff', 'drivers', 'thank', 'workers'],  
 ['scams', 'business', 'behavior', 'impact', 'pandemic', 'consumer'],  
 ['support', 'social', 'stay', 'help', 'home', 'amp']]
```

Conclusion:

- Topic1: panic buying & stocking food
- Topic2: grocery & retail
- Topic3: online shopping & delivery
- Topic4: gas price decline
- Topic5: Toilet paper hoarding
- Topic6: supermarket shelves are empty - Food shortage
- Topic7: mask & Hand sanitizer shortage
- Topic8: Thanks ALL front-line workers
- Topic9: covid impact on business & economy
- Topic10: StayAtHomeSaveLives?



Appendix 2:

Example: detailed raw tweets for topic_1

```
: topic_check('topic_1') #topic: "panic buying & stocking food"
```

```
food      3.10264
panic     2.22801
buying     1.64335
need       0.84370
stock      0.79738
people     0.72541
Name: topic_1, dtype: float64
```

```
9651
#Covid_19 stop panic buying theres enough food etc for everyone https://t.co/FfABn13Upo
9283
virus we need to stop panic buying so the old and the people that are still working get food all this #corona
15452
Minister The country has enough food supplies there is no need for panic buying 19
1881 #Coronavirus seriously can people stop buying free from food cause they have already hoarded the other foo
d. People need the free from food due to medical issue. Please stop panic buy there is need.
5816 For all of you panic buy
ing. STOP DOING IT Vulnerable people can't get any food please think of others. #coronavirus
2448
STOP FUCKING PANIC BUYING FOOD, YOU IDIOTS!! #coronavirus
Name: Raw_Tweet, dtype: object
```

Appendix 3:

Example: detailed raw tweets for topic_4

```
topic_check('topic_4') #gas price decline
```

```
prices      2.18292
oil          0.86774
gas          0.55985
price        0.34274
low          0.31089
amid         0.25006
Name: topic_4, dtype: float64
```

```
3352                                Oil prices at $23.18\r\r\n#Oilprices #coronavirus
3353                                Oil prices at $22.90\r\r\n#Oilprices #coronavirus #CoronaVirusUpdate
4021                                Oil prices at 2002 already, are we back almost 20 years due to #coronavirus
19424                                I do like Covid 19 gas prices
6280    Toilet paper prices went up while gas prices going down. #covid_19\r\r\nhttps://t.co/jp5aYlra9R
3954                                Gas prices keep going down because of COVID-19. \r\r\nhttps://t.co/GCX7Mz3M21
Name: Raw_Tweet, dtype: object
```