

Group Coursework Submission Form

Specialist Masters Programme

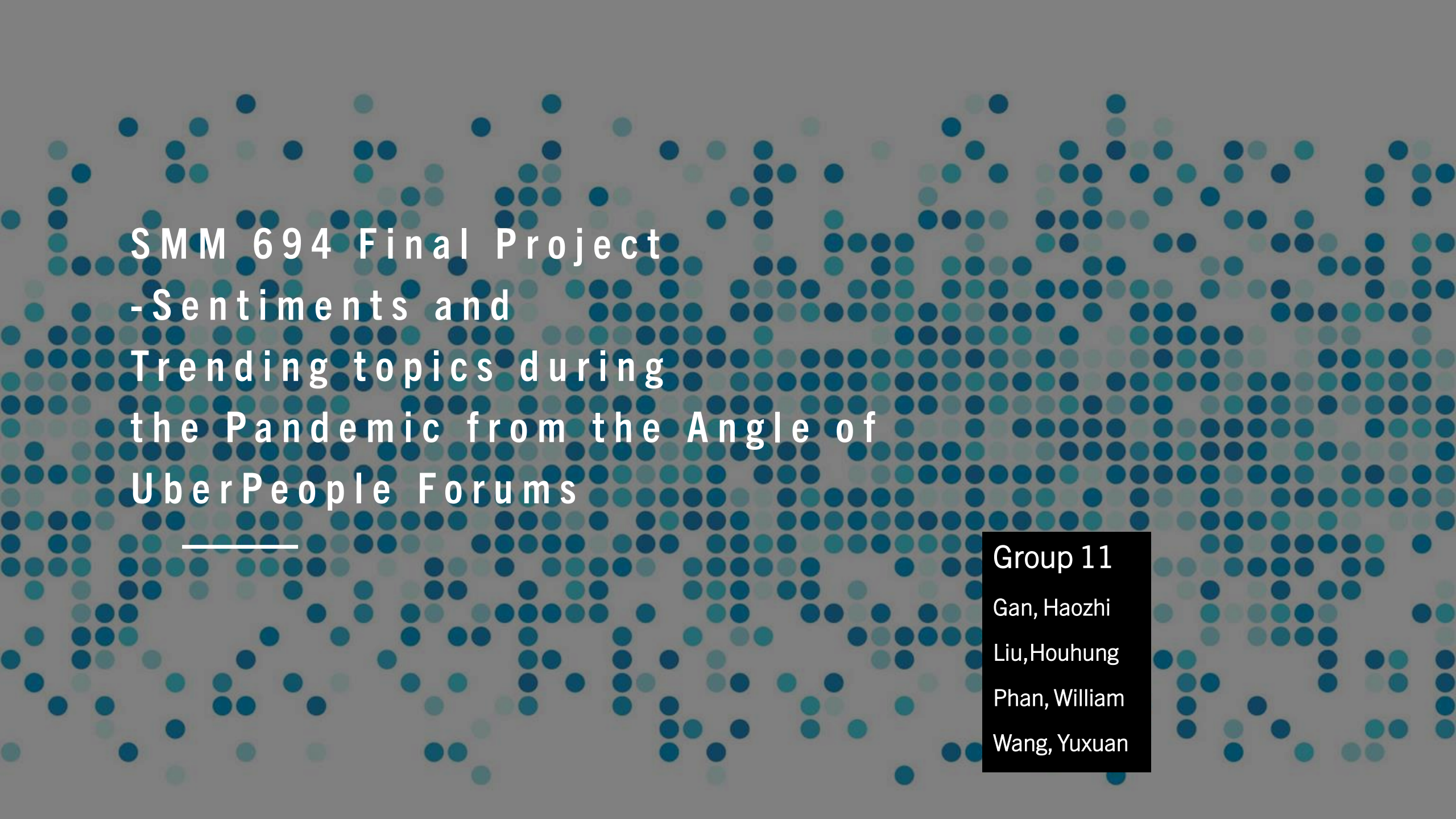
Please list all names of group members: (Surname, first name) 1. Gan, Haozhi 2. Liu, Houhung 3. Phan, William 4. Wang, Yuxuan		GROUP NUMBER	11
MSc in: Business Analytics			
Module Code: SMM694			
Module Title: Applied Natural Language Processing (NLP)			
Lecturer: Simone Santoni		Submission Date: 16/7/2021	
Declaration: <p>By submitting this work, we declare that this work is entirely our own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work we acknowledge that we have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. We also acknowledge that this work will be subject to a variety of checks for academic misconduct.</p> <p>We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.</p>			
Marker's Comments (if not being marked on-line):			

Deduction for Late Submission:

--

Final Mark:

□ %



SMM 694 Final Project - Sentiments and Trending topics during the Pandemic from the Angle of UberPeople Forums

Group 11

Gan, Haozhi

Liu, Houhung

Phan, William

Wang, Yuxuan

TABLE OF CONTENTS

- Purpose and Steps of the Projects
- Sentiment Analysis
- Topic Modelling
- Dynamic Topic Modelling

PURPOSE AND STEPS OF THE PROJECTS

Background:

Concerns raised from the start of 2020 as COVID-19 started its rapid spread. WHO Issues Global Health Emergency on 31st January 2020. Many countries soon took approaches: Trump Declares COVID-19 a National Emergency on March 13; UK went into lockdown on 23 March.

Ride hailing is one of the most affected industries during the pandemic. Uber being the number one ride-sharing company in the US has been experiencing hard times with its business ever since then.

Main Purpose of Project:

We'd like to know how COVID affect the Uber markets and Uber users globally and to investigate the most trending topics and the sentiments inside the forum.

Since the awareness of COVID was raised by the UK and the US government in March 2020, we collect data starting from March 2020 to the early July of 2021.

We select "UberPeople Advice" community as the source of web scraping. From this forum, we can understand what kinds of advices were asked and hence the concerns that might arise during the lockdown.



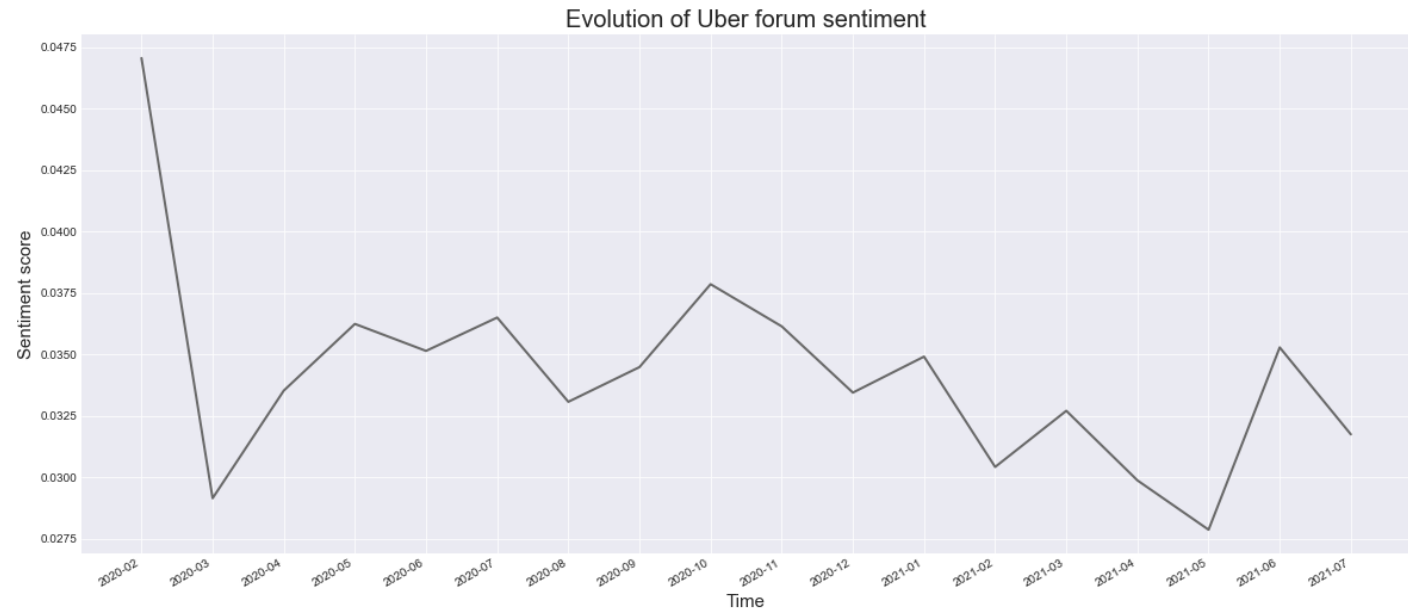
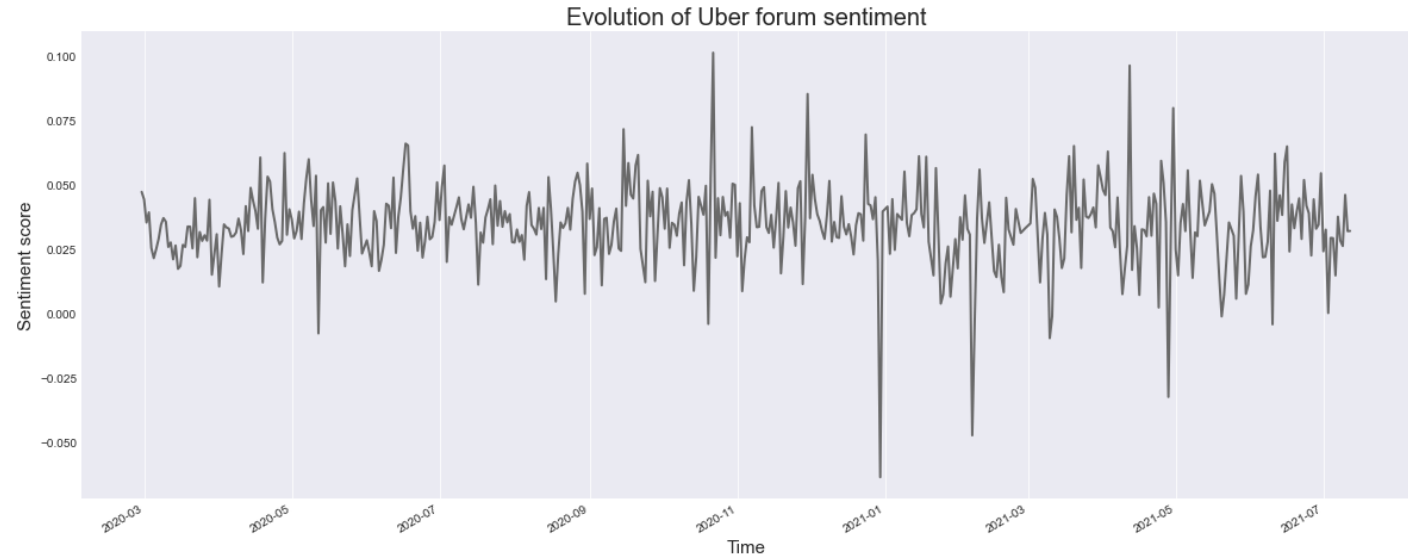
SENTIMENT ANALYSIS

- After loading General Inquirer categories of words from Harvard, we built sentiment axis based on below formula :
"Positive + Power + Pleasure – Negative – Pain – Submit"
- If the sentiment score is positive, the reviews are more positive and confident.
- Conversely, if the sentiment score is minus, the reviewers are in low mood and feel depressed.
- To sum up, the average sentimental score in UberPeople forum is positive, meaning the users has positive feeling during pandemic.

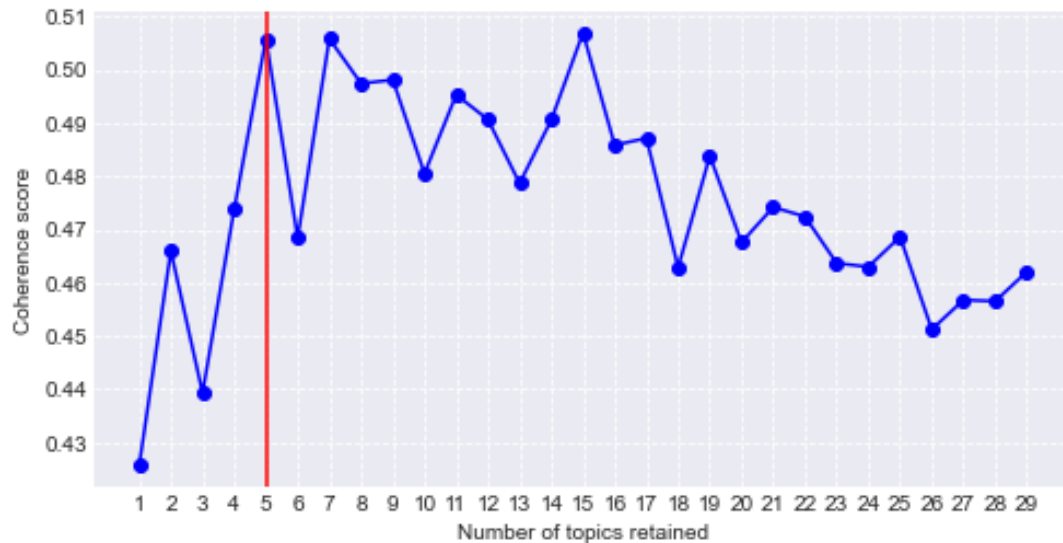
sentiment_score	
count	58785.000000
mean	0.033309
std	0.050093
min	-0.300198
25%	0.011632
50%	0.037911
75%	0.064457
max	0.292673

EVOLUTION OF UBER FORUM SENTIMENT

- Both line graphs indicate the fluctuation of sentiment score from 2020/2 to 2021/7; however, the first graph is on a daily basis while the second one is on a monthly basis.
- Overall, the peak of score arrives around February of 2020, and the bottommost score is around May of 2021.
- It is reasonable that sentiment score dramatically decreases from March 2020 due to severe outbreak of COVID and oncoming lockdown.
- In May 2021, Uber was facing historic driver shortages, leading to steeper fares, longer waits and the highest wages in years leads sentiment score reach the lowest point.



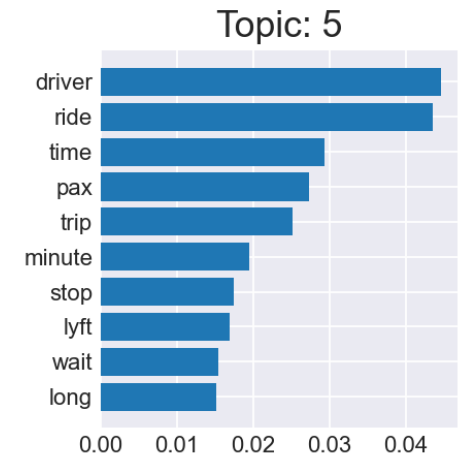
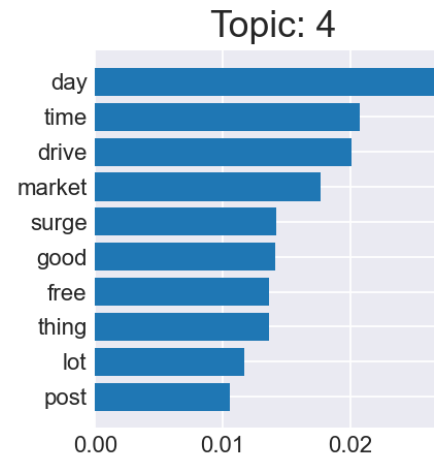
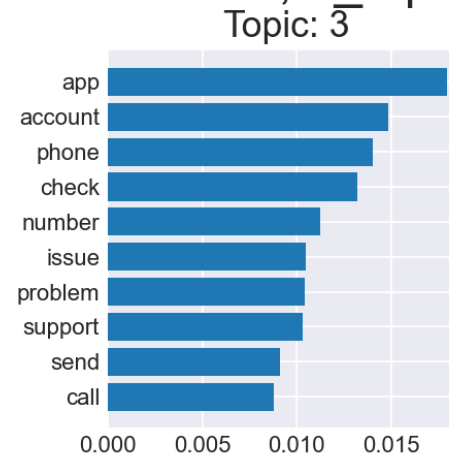
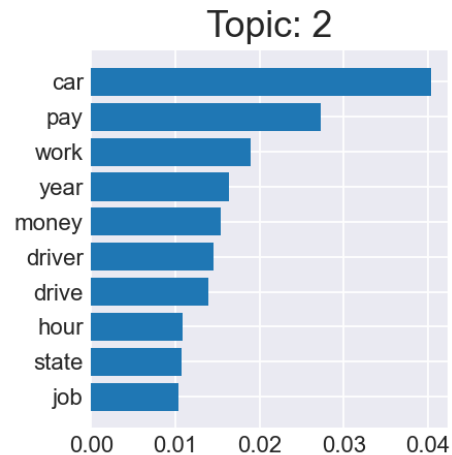
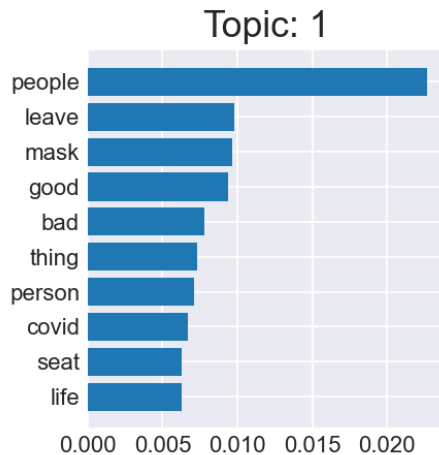
TOPIC MODELLING



Using coherence score to tune the optimal number of topic. In this case, the optimal number of topics is 5.

- **Topic 1** is relevant to pandemic, such as Covid, and mask.
- **Topic 2** shows the topic for jobs such as pay, job, working time and money.
- **Topic 3** seems to be the technical topic such as app, phone issues.
- **Topic 4** is related to the state of feelings, such as good and bad.
- **Topic 5** is about driving time, waiting time, or stop time.

Uber forum LDA, $n_topics = 5$

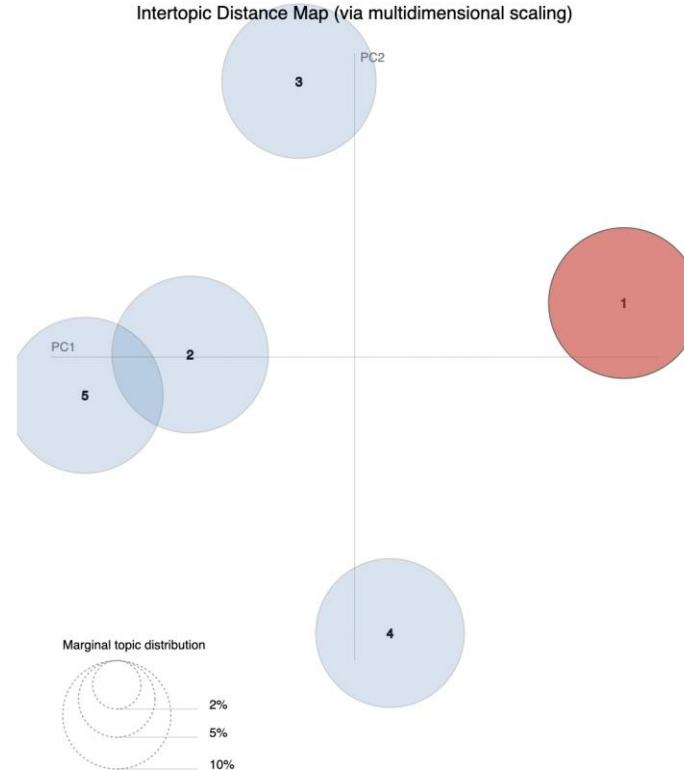


INTERACTIVE VISUALIZATION OF TOPIC MODELLING

- The **Topic 2** and **Topic 5** are close to each other, meaning both topics are similar. As in the previous slide, both topics are related to time (working time and waiting time).
- To be more specific, **Topic 1** covered lots of terms relating to policies of Covid such as, people have to wear mask in the vehicle, hand-washing, and open window.

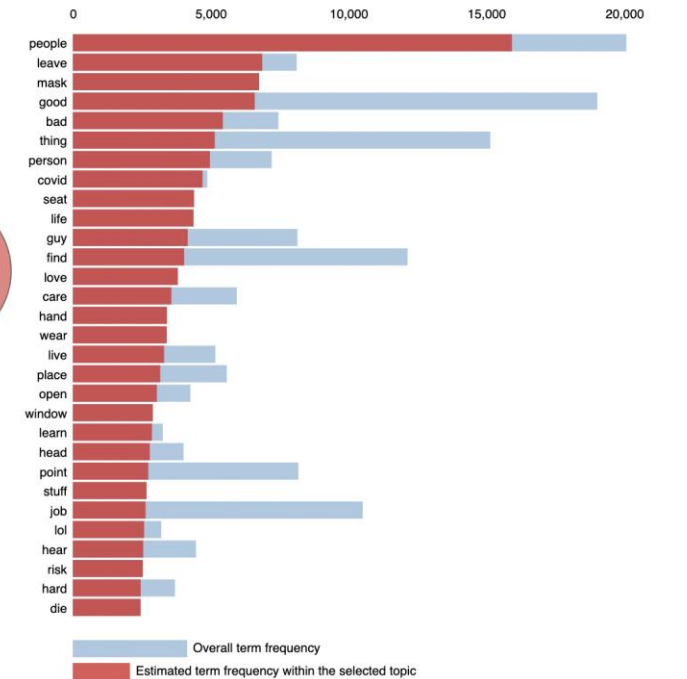
Selected Topic:

Intertopic Distance Map (via multidimensional scaling)



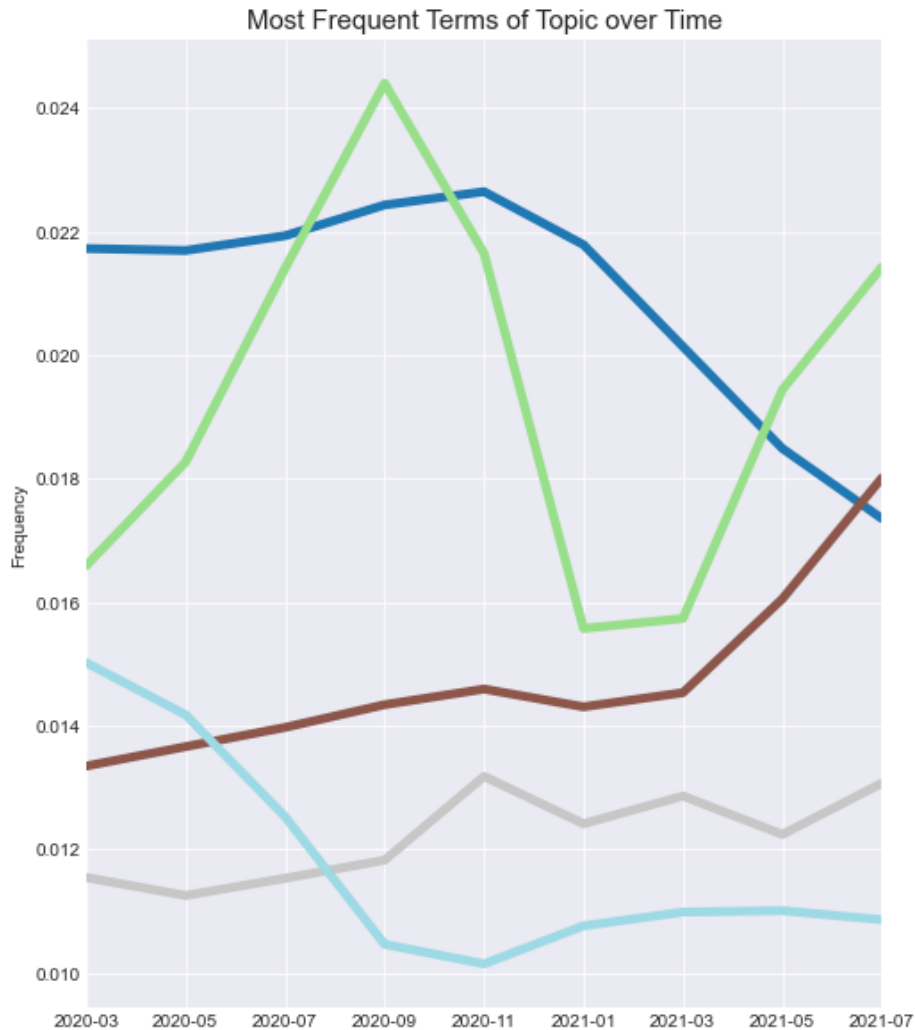
Slide to adjust relevance metric:(2)
 $\lambda = 1$ 0.0 0.2 0.4 0.6 0.8 1

Top-30 Most Relevant Terms for Topic 1 (19.3% of tokens)



1. $saliency(term\ w) = frequency(w) * [\sum_t p(t|w) * \log(p(t|w)/p(t))]$ for topics t ; see Chuang et. al (2012)
2. $relevance(term\ w\ i\ topic\ t) = \lambda * p(w\ i\ t) + (1 - \lambda) * p(w\ i\ t)/p(w)$; see Sievert & Shirley (2014)

DYNAMIC TOPIC MODELLING - TOPIC 1

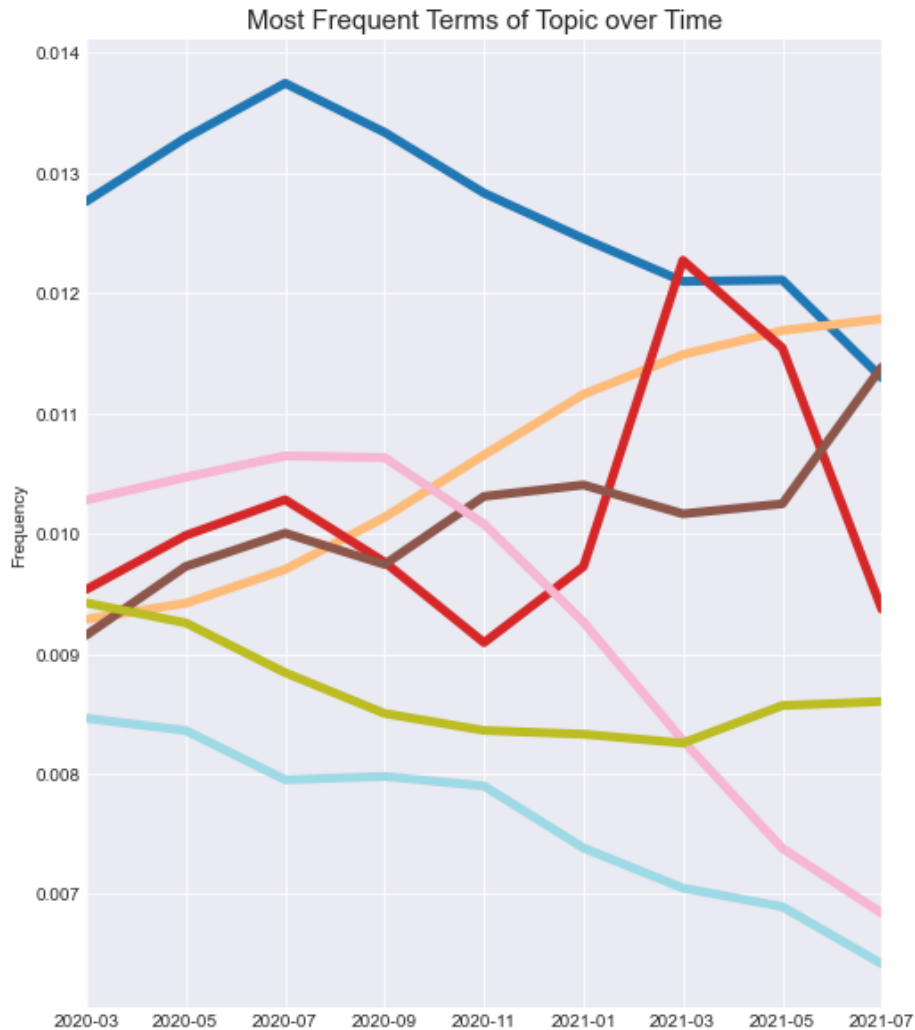


	2020/3	2020/5	2020/7	2020/9	2020/11	2021/1	2021/3	2021/5	2021/7
ride	ride	ride	ride	driver	ride	ride	ride	driver	driver
driver	driver	driver	driver	ride	driver	driver	driver	ride	time
time	drive	drive	time	time	time	time	time	time	ride
trip	time	time	drive	trip	trip	trip	trip	trip	trip
get	trip	trip	trip	hour	minute	drive	drive	drive	pax
day	get	get	get	drive	drive	minute	minute	pax	drive
pax	day	day	hour	get	get	get	pax	minute	minute
pay	pax	pax	pay	minute	pay	app	get	get	lyft
minute	pay	pay	day	pay	app	pax	app	lyft	pay
	minute	minute	minute	pax	pax	pay	lyft	pay	mile

Summary:

- It is evident that Topic 1 is about working environment in Uber such as working hour and salary. It is noticeable that from March to July in 2021 people started to pay more attention to Lyft, the second-largest ridesharing company in the United States after Uber meaning that the competitors of Uber start to grow during the Pandemic.
- As is shown in the figure, people are becoming more interested in 'time'. Besides, 'app' appears in the top terms from November 2020. It can be assumed that people are gradually changing their lifestyle like 'working from home'.

DYNAMIC TOPIC MODELLING - TOPIC 2

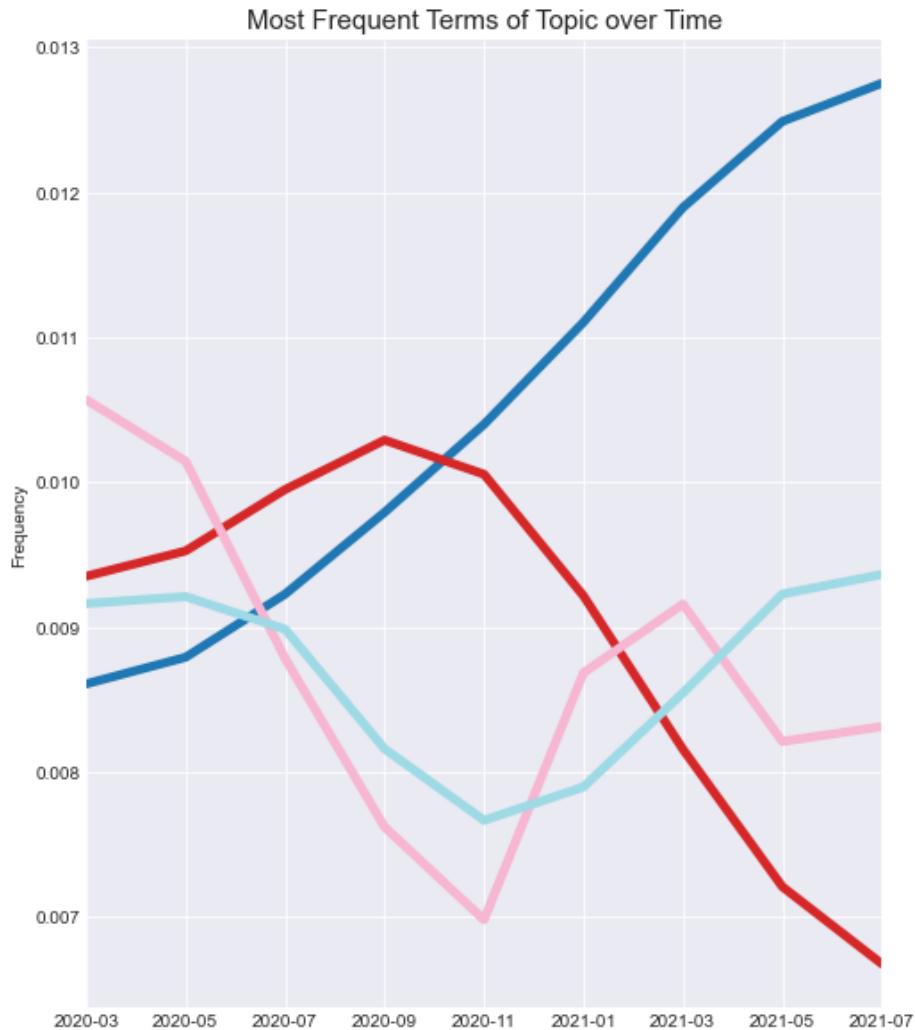


	2020/3	2020/5	2020/7	2020/9	2020/11	2021/1	2021/3	2021/5	2021/7
know	know	know	know	know	know	know	think	know	drive
people	people	people	people	people	car	car	know	car	car
time	time	time	time	time	get	get	car	think	get
think	think	think	think	think	time	think	get	get	know
get	go	get	get	get	people	time	drive	drive	think
car	car	car	car	car	think	drive	time	go	go
get	get	go	go	go	go	go	go	time	time
like	like	like	like	like	drive	people	like	like	ticket
drive	drive	drive	good	good	like	like	people	use	like
good	good	good	way	way	good	good	good	people	driver

Summary:

- This topic is related to the state of feelings. People felt discussing in the Uber forum is a good way to know people from the community during the pandemic.

DYNAMIC TOPIC MODELLING - TOPIC 3

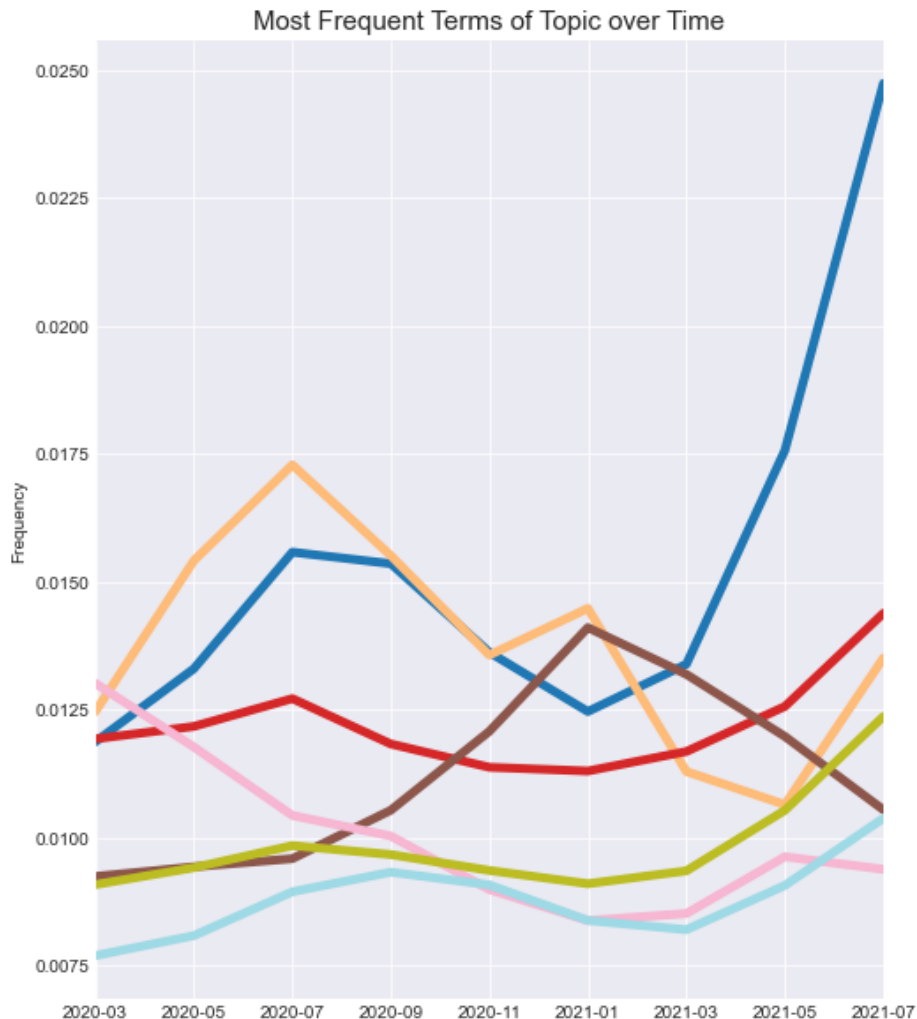


	2020/3	2020/5	2020/7	2020/9	2020/11	2021/1	2021/3	2021/5	2021/7
driver	pay	week	work	work	driver	driver	driver	driver	driver
work	week	pay	week	driver	work	account	pay	get	check
pay	work	work	driver	time	account	work	get	year	get
get	get	get	get	week	time	pay	account	check	year
	driver	driver	pay	get	get	time	work	pay	pay
	state	state	time	account	pay	get	year	account	account
	unemployment	time	state	state	app	know	time	work	background
	time	know	account	pay	year	year	know	know	know
	know	unemployment	money	money	know	tell	support	time	state
	money	account	know	job	week	support	state	state	work

Summary:

- According to the most frequent terms, Topic 3 is about employment, and people talk about payments and working time in Uber, which probably means more and more people are interested in working as Uber drivers and want to get further information from forum.
- After lockdown in the US and UK, people are increasingly interested in the topic about drivers; nevertheless, term of work also decreased since September 2020. Also, "support" appears since the start of 2021. It seems that the demand for drivers is growing and Uber is facing driver shortages. In fact, drivers are getting financial support from the government from January 2021.

DYNAMIC TOPIC MODELLING - TOPIC 4

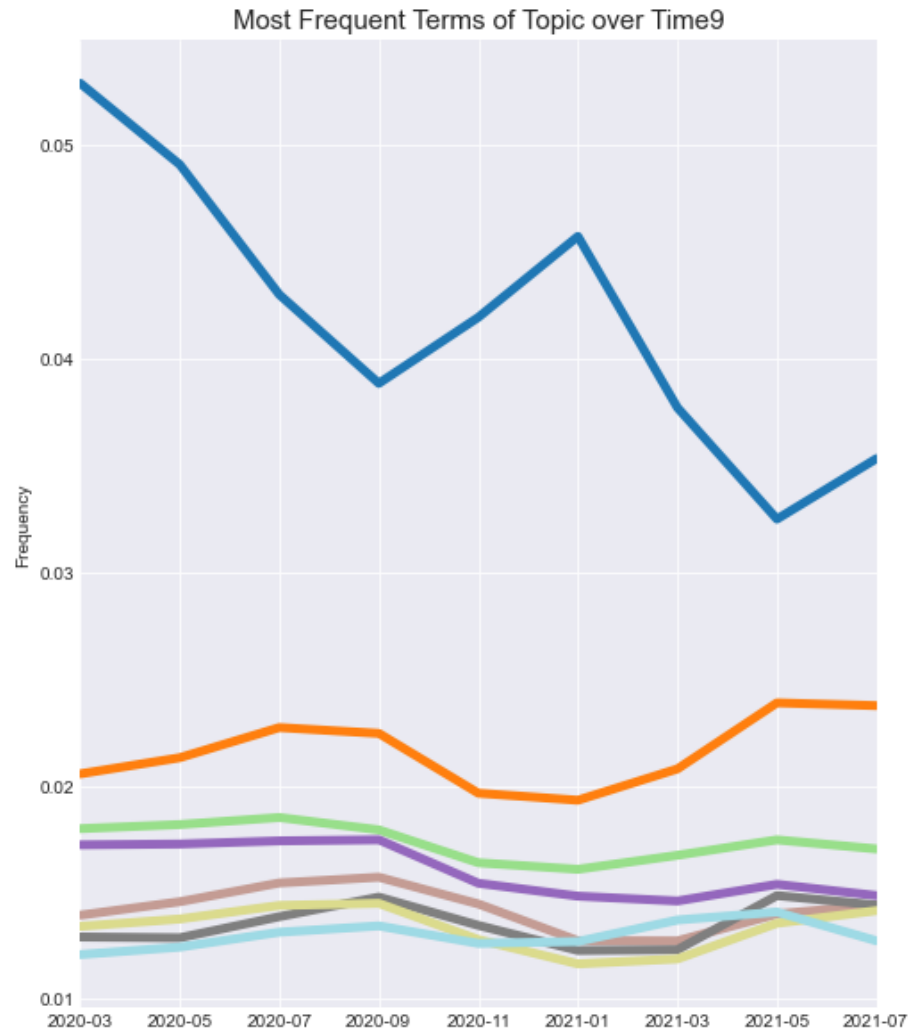


	2020/3	2020/5	2020/7	2020/9	2020/11	2021/1	2021/3	2021/5	2021/7
market	—	mask	mask	mask	market	mask	market	market	market
mask	people	—	market	market	mask	like	like	—	—
think	mask	market	think	think	like	market	Think	think	think
like	think	think	—	like	think	think	mask	like	mask
people	market	people	people	people	right	right	right	mask	right
right	virus	like	right	right	person	person	people	right	like
person	like	right	like	person	people	people	Pax	people	person
	right	virus	person	—	different	pax	person	person	people
	person	person	wear	different	pax	different	different	different	different
	wear	wear	different	attach	wear	wear	—	pax	wear

Summary:

- This topic is responsible for pandemic. People are talking about virus, about whether people should wear mask or whether the market was affected.
- The debt on 'mask' is increasing with the growing population vaccinated and the US and UK reopening. People discussed if wearing face masks is mandatory for Uber passengers and drivers.
- It is obvious that people have greater concerns in the market after March 2021 when the economy begins to resume.

DYNAMIC TOPIC MODELLING - TOPIC 5



2020/3	2020/5	2020/7	2020/9	2020/11	2021/1	2021/3	2021/5	2021/7
car	car	car	car	car	car	car	car	car
like	like	like	like	like	like	like	like	like
look	look	look	look	look	look	look	look	look
free	free	free	free	free	free	free	free	free
insurance	insurance	kid	kid	kid	Insurance	zone	seriously	kid
kid	kid	believe	seriously	seriously	kid	kid	zone	seriously
believe	believe	seriously	believe	believe	seriously	seriously	kid	believe
seriously	seriously	insurance	zone	zone	believe	believe	believe	drive
zone	zone	zone	insurance	insurance	zone	insurance	sarcasm	zone
vehicle	vehicle	sarcasm	sarcasm	sarcasm	drive	sarcasm	drive	sarcasm

Summary:

- This topic is responsible for car insurance and other social issue. Those are long-existing subjects and no matter outbreak of pandemic would not have significantly changes.