



# COVID-19

# TWEETS

Topic Modeling  
&  
Sentiment Analysis

YINGQING QIU  
METIS DATA SCIENCE BOOTCAMP

# INTRODUCTION



## GOAL:

Provide data analysis support for **social scientists** and **psychologists** to study how people express emotions and what they are most caring about when catastrophic events happen.

## DESIGN:

- Topic Modeling
- Dynamic Topic Modeling
- Sentiment Analysis

# **DATA** - Kaggle Data Source

**~40k**  
**TWEETS**

**March~April 2020**  
**TIME RANGE**

**March~April 2020**  
**COVID CASES**

**44%**  
**POSITIVE**

**37%**  
**NEGATIVE**

**19%**  
**NEUTRAL**

# TOPIC MODELING & SENTIMENT ANALYSIS WORKFLOW

## Preprocessing

NLTK  
LancasterStemmer  
NLTK stopwords

## Vectorizer

CountVectorizer  
TF-IDF Vectorizer

## Topic Modeling

NMF  
LSA

## Dynamic Topic Modeling

gensim  
ldaseqmodel

## Sentiment Analysis

VADER  
SentimentIntensityAnalyzer

# KEYWORDS & TOPICS

## NMF

## LSA

*CountVectorizer*

'need, demand, stock, panic, food',  
'retail, going, like, grocery, store',  
'oil, shopping, pandemic, consumer, covid',  
'time, like, local, staff, supermarket',  
'like, panic, need, shopping, people'

'covid, food, store, grocery, people',  
'store, grocery, going, like, retail',  
'food, people, panic, stock, need',  
'supermarket, people, shopping, like, home',  
'people, consumer, shopping, behavior, help'

*TfidfVectorizer*

'demand, need, panic, stock, food',  
'like, go, work, groceri, store',  
'impact, pandem, shop, consum, covid',  
'time, work, shop, peopl, supermarket',  
'roll, peopl, hand, toilet, paper'

'store, groceri, covid, food, peopl',  
'store, groceri, go, thank, line',  
'consum, covid, pandem, impact, behavior',  
'supermarket, shop, consum, work, covid',  
'paper, toilet, consum, hand, roll'



## **SELECTED TOPICS:** **TFIDFVECTORIZER + NMF**

### **PANIC**

demand, need, panic, stock,  
food, hoard, bank

### **GROCERY**

like, go, work, groceri, store,  
local, today, day, shop

### **RETAIL**

impact, pandem, shop,  
consum, covid, protect, retail,  
oil, outbreak

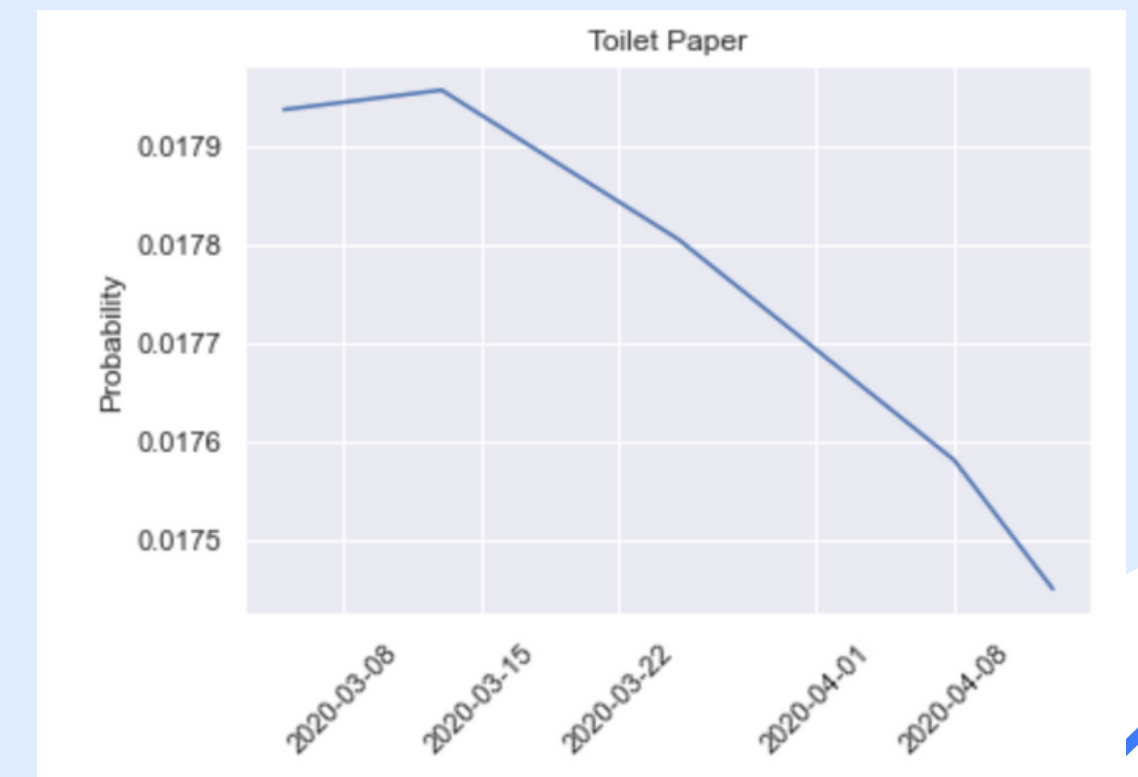
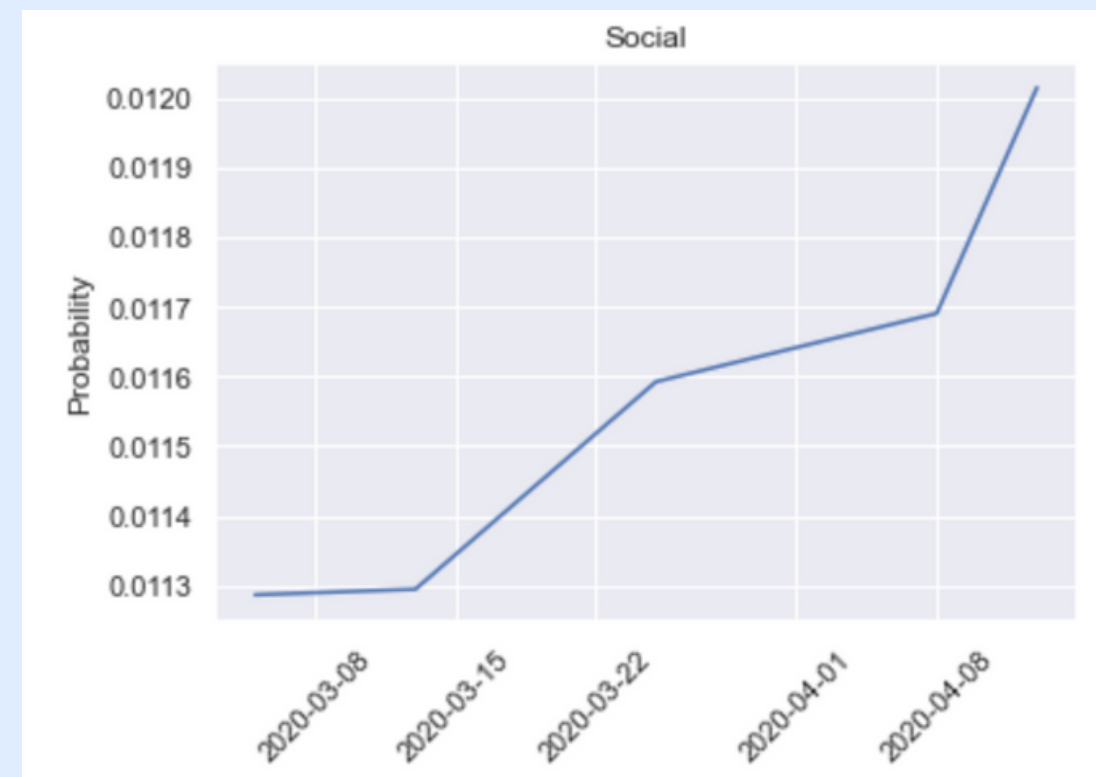
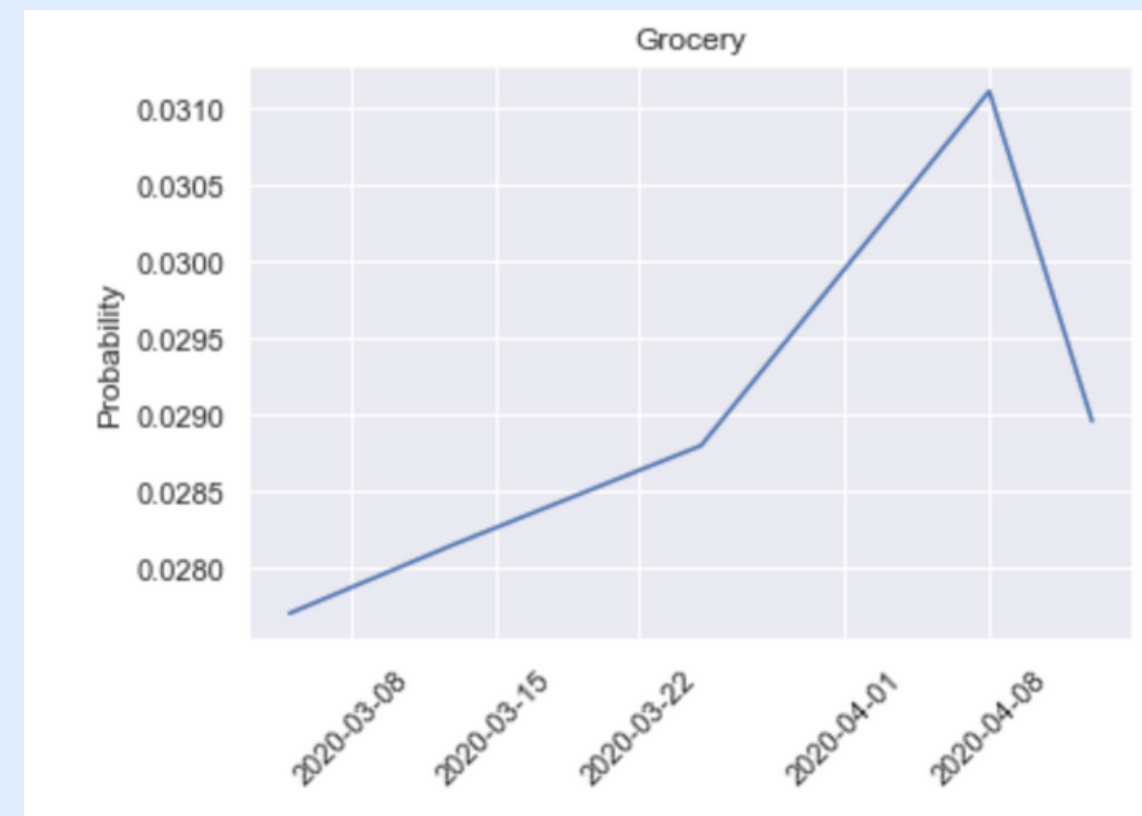
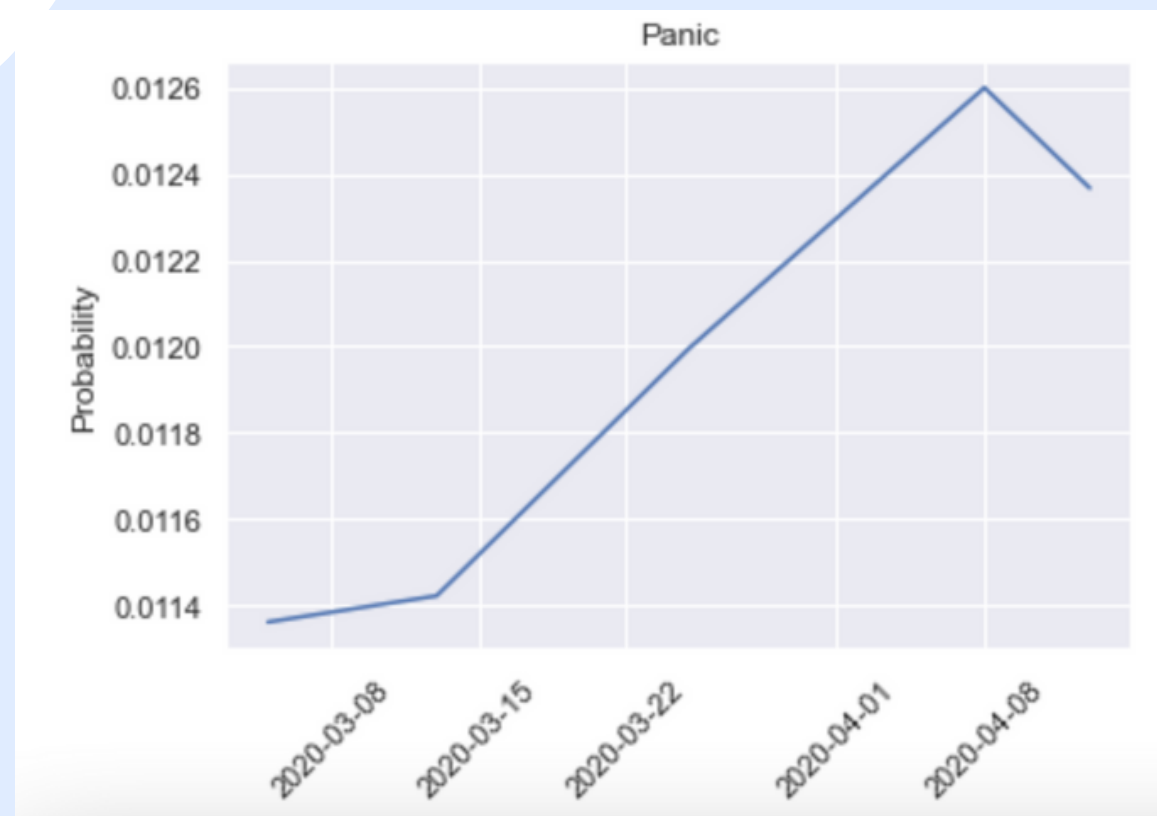
### **SOCIAL**

time, work, shop, peopl,  
supermarket, social, home,  
staff

### **TOILET PAPER**

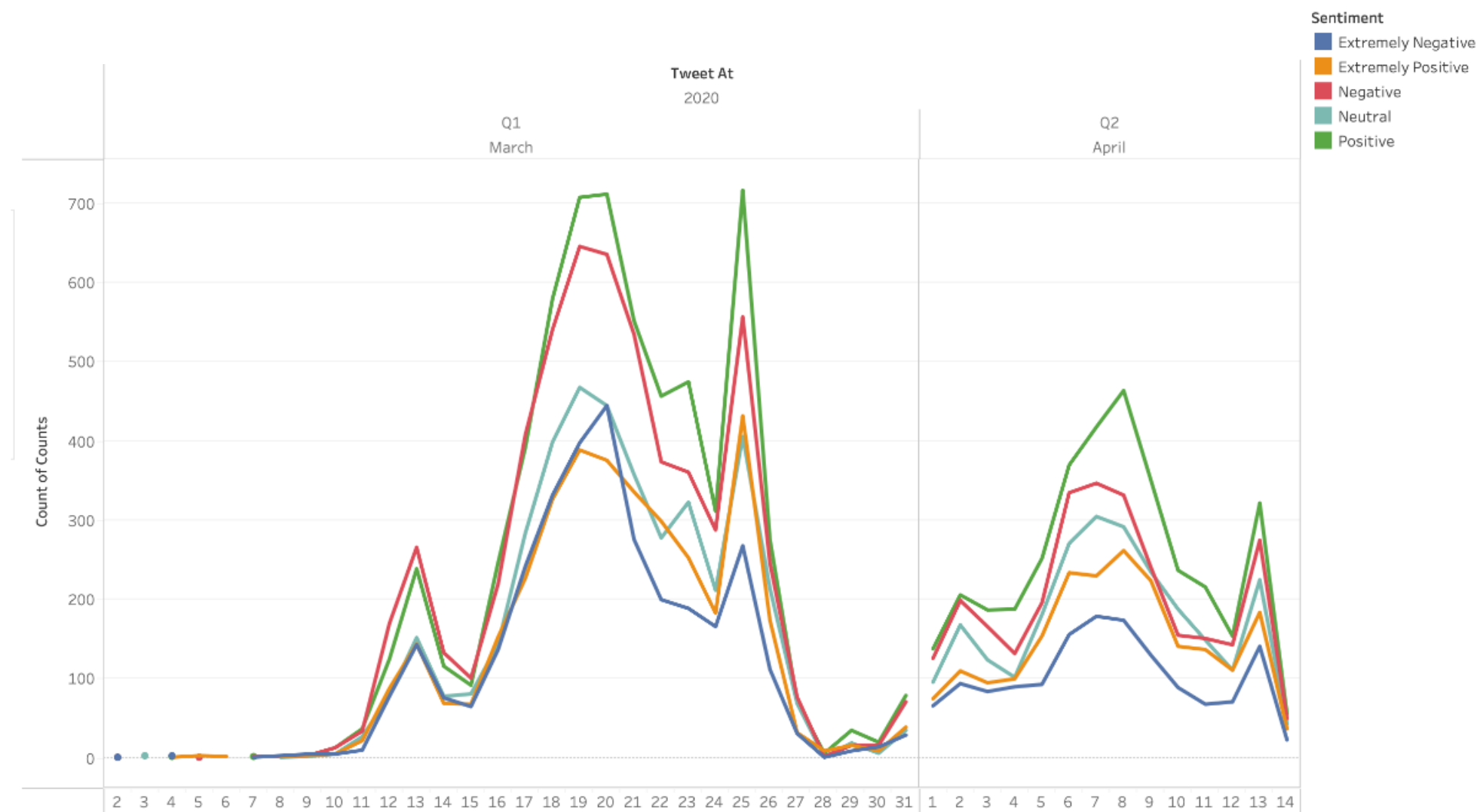
roll, peopl, hand, toilet,  
paper, buy, shortag, water,  
hoard

# DYNAMIC TOPIC MODELING



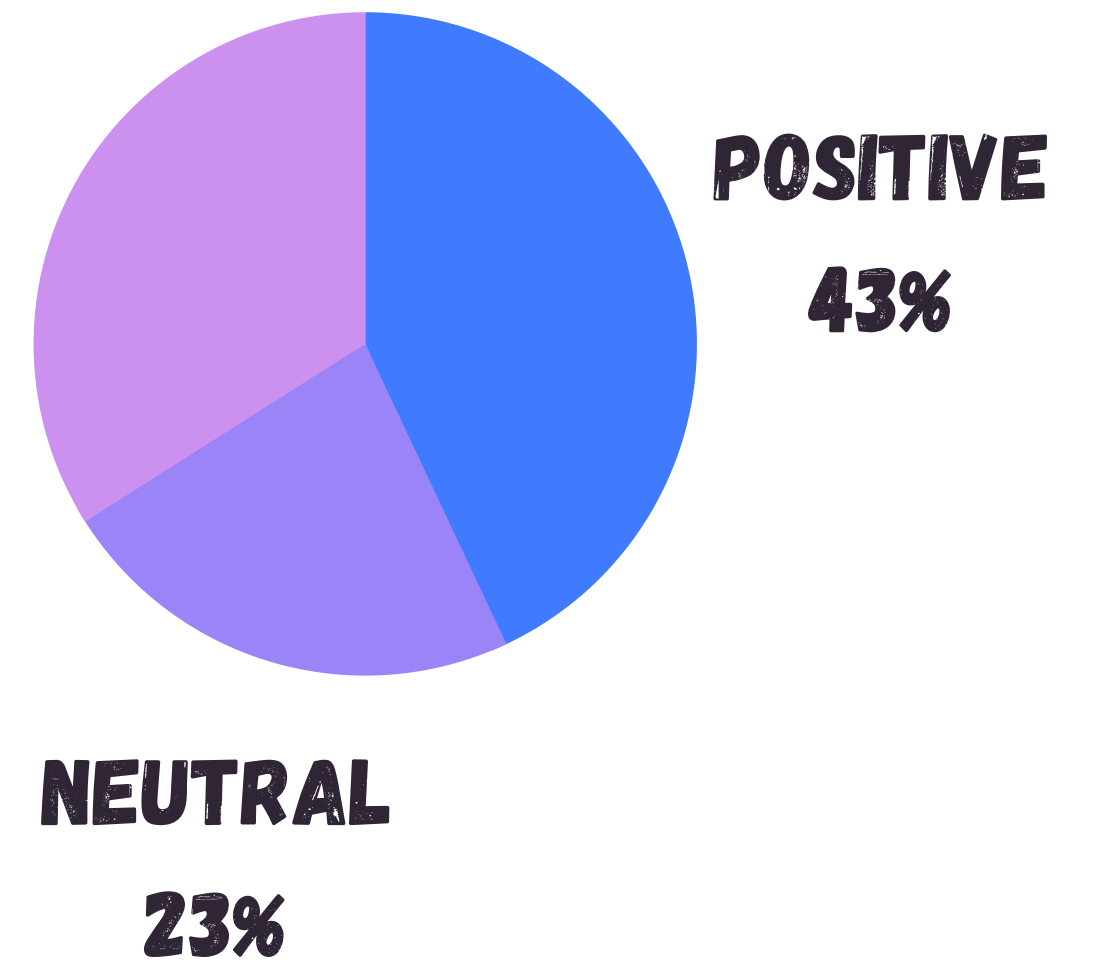
# SENTIMENT ANALYSIS

Given sentiments in March and April 2020



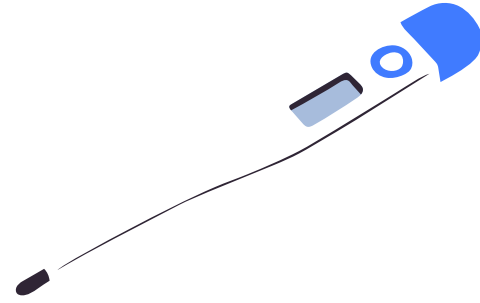
Predicted by **VADER**: -0.05~0.05

**NEGATIVE**  
**34%**

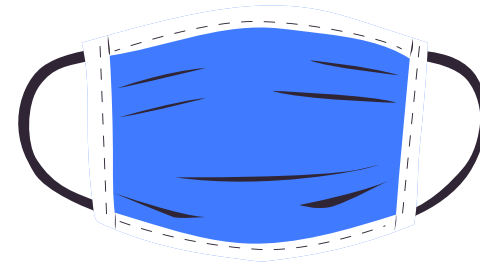




# CONCLUSIONS



- PANIC, GROCERY, RETAIL, SOCIAL, and TOILET PAPER are the five top topics.



- RETAIL has the most obvious rising trends among the top topics with the increasing of COVID cases.



- The sentiment does not obviously vary with time, while more people remain at a positive stage.

# NEXT STEPS



- Including Geographic effects.
- Consider government mask requirements



**THANK YOU!**  
**ANY QUESTIONS?**