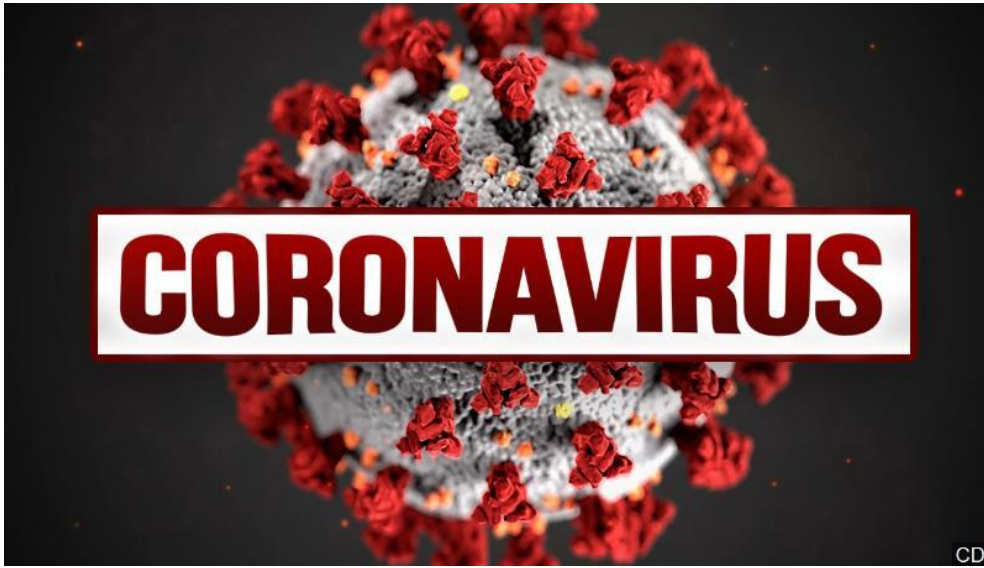
Analyzing people's opinion toward coronavirus

Chenchen, Qinwei, Jialiang, Jiaxin, Yang, Keyang, Yikun

Why we chose this topic?

1. Covid-19 has drawn public attention since Jan and there are many different opinions toward this virus. So we can gather enough data for this project
2. Covid-19 affects everyone in America so most people are interested in this topic



Goal

1. We want to analyze how location affect people's opinion toward covid-19
2. As time goes by, we want to know if people change their opinions.

Data Collection

Keywords: coronavirus, covid-19, outbreak, and pandemic etc

Location: NYC, Denver, Seattle, LA, Chicago and so on(12 cities in total)

Time: we collected data from 02/01/2020 to 04/19/2020

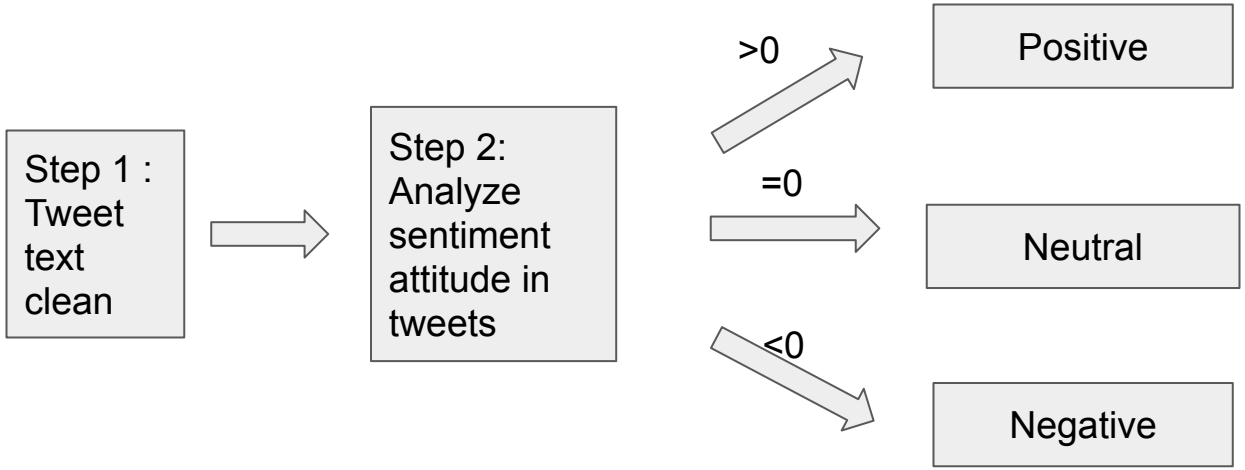
How ?

We use GetOldTweets3 to search all the tweets that are related to those keywords.
Overall, We collected about 70,000 tweets

Mining Sentiment Information in Tweets

Sentiment Attitude: Positive, Negative, Neutral

Implementation:



TextBlob

We used default training and validation data of TextBlob to build a classifier model to give the input tweet a sentimental attitude. TextBlob uses NaiveBayesClassifier as classifier model.

TextBlob will generate a list of words with sentiment. By comparing this words' negative points to positive points, classifier will finally gain a total points of the whole sentence (tweet) and return the final points.

Pros and Cons of TextBlob

Pros:

Since, it is built on the shoulders of NLTK and Pattern, therefore making it simple for beginners by providing an intuitive interface to NLTK.

Cons:

It is little slower in the comparison to spacy but faster than NLTK. (Spacy > TextBlob > NLTK)

It does not provide features like dependency parsing, word vectors etc. which is provided by spacy.

NLP tasks using TextBlob

1. Tokenization
2. Part-of-Speech Tagging & Filter Words by Tags
3. Words Inflection and Lemmatization
4. Generate N-grams Lists
5. Sentiment Analysis
6. Spelling Correction

1. Tokenization

```
from textblob import TextBlob

blob = TextBlob("I have a dream, and she has dream.\
                \nFrom now on, this place is called lbw square.\
                \nI bought a Taylor guitar in New York last year")
```

2. Part-of-Speech Tagging & Filter Words by Tags

```
#Part-of-Speech Tagging
for words, tag in blob.tags:
    print(words,tag)
```

```
I PRP
have VBP
a DT
dream NN
and CC
she PRP
has VBZ
dream NN
From IN
now RB
on IN
this DT
place NN
is VBZ
```

3 Words Inflection and Lemmatization

In [45]:
```
## lemmatization
blob = TextBlob("I have a dream, and she has dream. \
                \nFrom now on, this place is called lbw square. \
                \nI have 2 guitars now")
print(blob.sentences[2].words[3].singularize())

from textblob import Word
w = Word('running')
w.lemmatize("v")
```

```
guitar
```

Out[45]:
```
'run'
```

In [16]:
```
w = Word('bought')
w.lemmatize("v")
```

Out[16]:
```
'buy'
```

4. Generate N-grams Lists

```
# N-grams
blob = TextBlob("I love deep learning")
for ngram in blob.ngrams(2): # Bi-grams
    print(ngram)
for ngram in blob.ngrams(3): # Tri-grams
    print(ngram)
```

```
['I', 'love']
['love', 'deep']
['deep', 'learning']
['I', 'love', 'deep']
['love', 'deep', 'learning']
```

5. Sentiment Analysis

```
print (blob)
blob.sentiment
```

```
I have a dream, and she has dream.
 From now on, this place is called lbw square.
 I bought a Taylor guitar in New York last year
```

```
Sentiment(polarity=0.06818181818181818, subjectivity=0.2606060606060606)
```

```
blob = TextBlob("The coronavirus COVID-19 is \
                affecting 210 countries and territories around\
                the world and 2 international conveyances.\n\
                Most people infected with the COVID-19 virus will\
                experience mild to moderate respiratory illness and\
                recover without requiring special treatment.\
                Older people, and those with underlying medical \
                problems like cardiovascular disease, diabetes, \
                chronic respiratory disease, and cancer are more \
                likely to develop serious illness.\n\
                The best way to prevent and slow down transmission \
                is be well informed about the COVID-19 virus, the disease \
                it causes and how it spreads. Protect yourself and others \
                from infection by washing your hands or using an alcohol \
                based rub frequently and not touching your face. ")
blob.sentiment
```

```
Sentiment(polarity=0.1278835978835979, subjectivity=0.43735449735449733)
```

6. Spelling Correction

```
blob = TextBlob("The coronavirus COVID-19 is \
                bffecting 210 countiies and territories around\
                the world and 2 international conveyances.")
blob.correct()
```

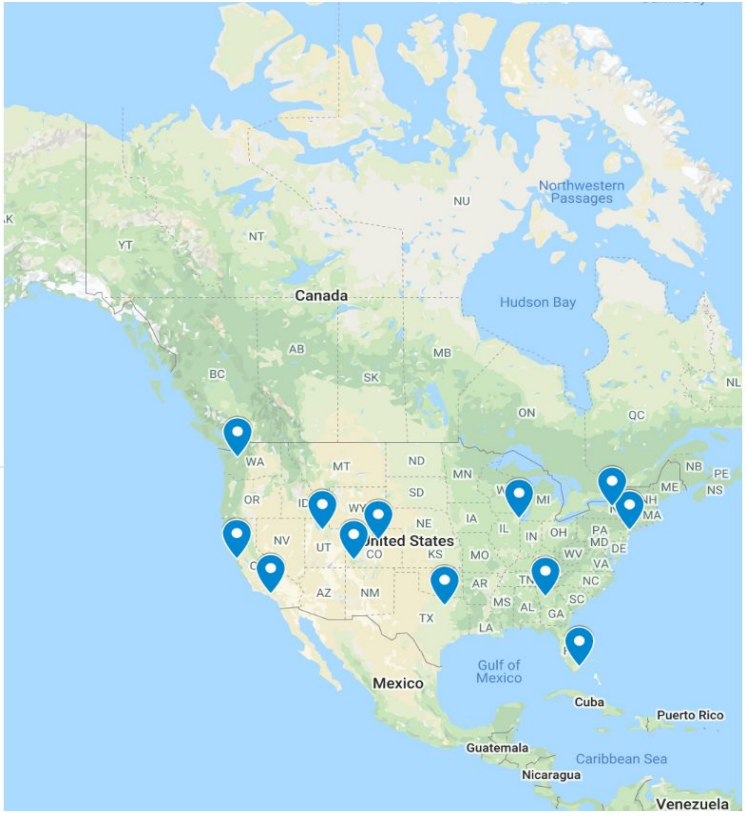
```
TextBlob("The coronavirus COVID-19 is affecting 210 countries and territories a
round the world and 2 international conveyances.")
```

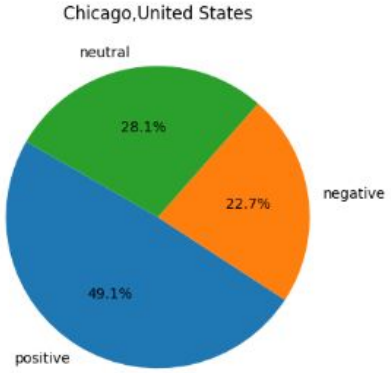
Location analysis

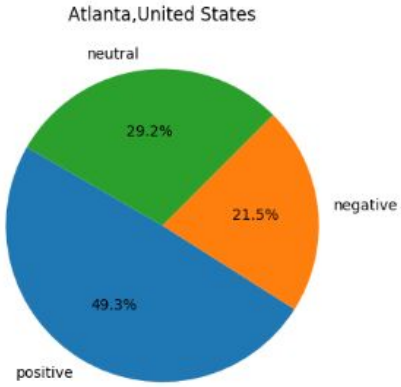
Chosen location

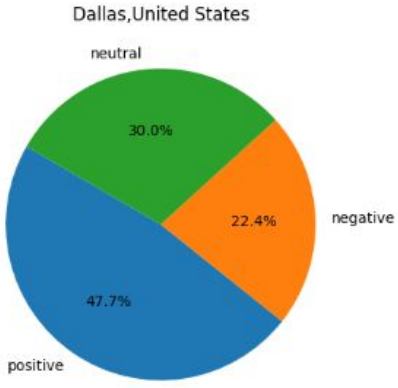
Untitled layer

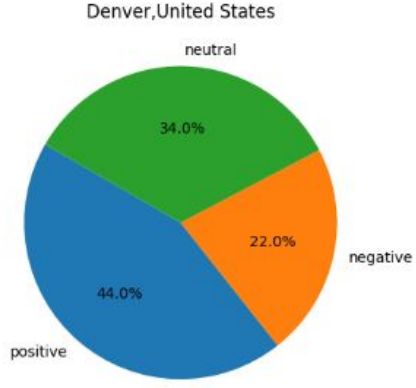
- Los Angeles
- New York
- Seattle
- San Francisco
- Salt Lake City
- Denver
- Miami
- Dallas
- Chicago
- Atlanta
- Syracuse
- Rico

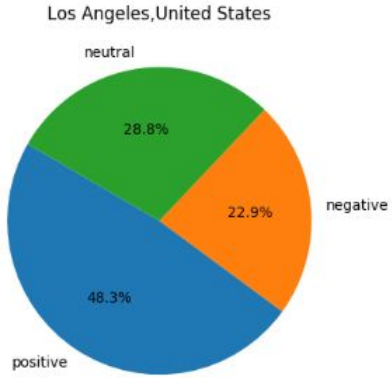


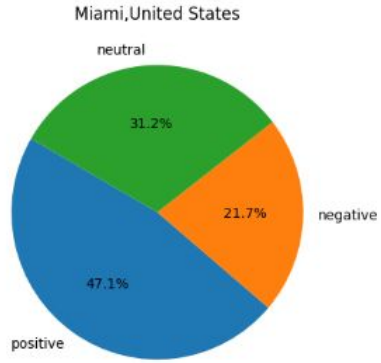

Chicago,United States
neutral
28.1%
negative
22.7%
49.1%
positive

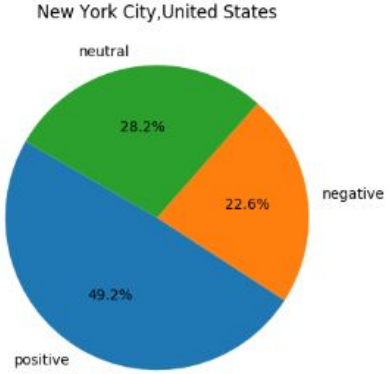

Atlanta,United States
neutral
29.2%
negative
21.5%
49.3%
positive

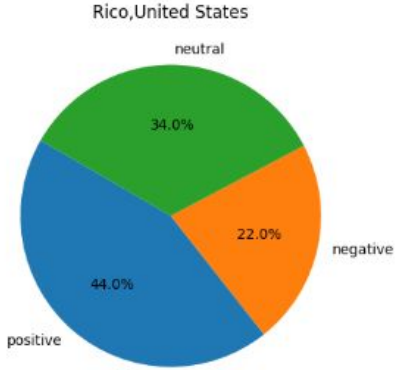

Dallas,United States
neutral
30.0%
negative
22.4%
47.7%
positive

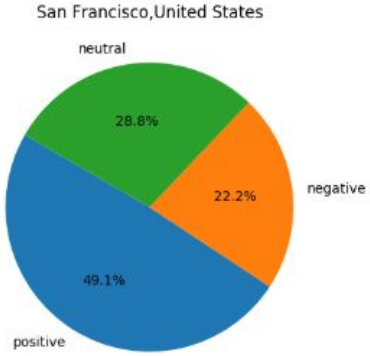

Denver,United States
neutral
34.0%
negative
22.0%
44.0%
positive

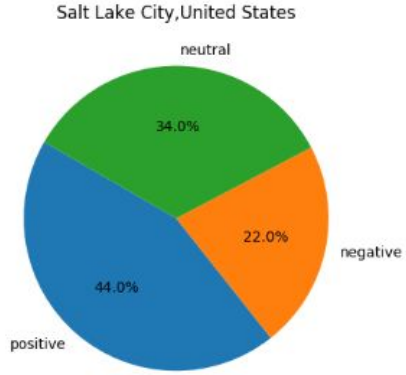

Los Angeles,United States
neutral
28.8%
negative
22.9%
48.3%
positive

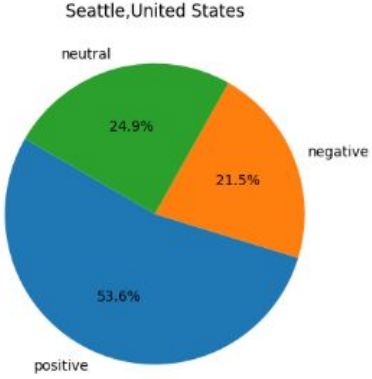

Miami,United States
neutral
31.2%
21.7%
negative
47.1%
positive

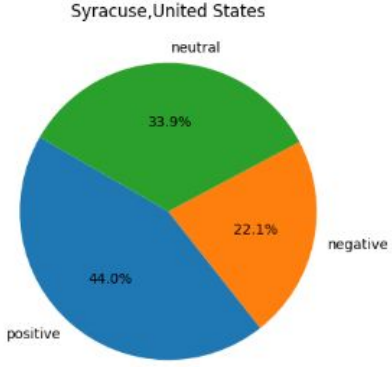

New York City,United States
neutral
28.2%
22.6%
negative
49.2%
positive


Rico,United States
neutral
34.0%
22.0%
negative
44.0%
positive

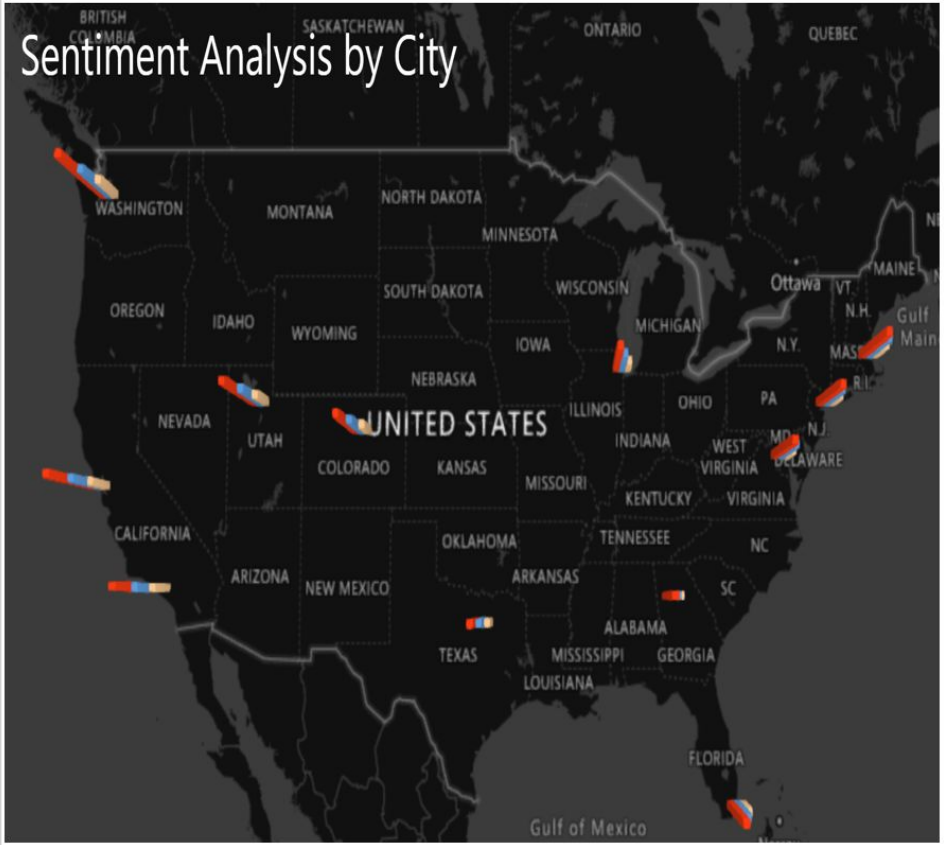

San Francisco,United States
neutral
28.8%
negative
22.2%
49.1%
positive

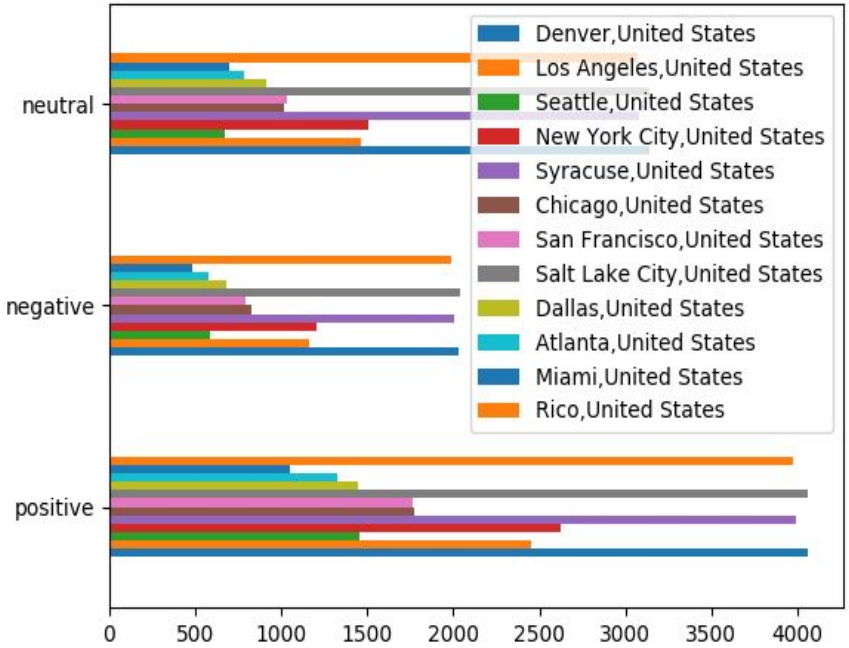

Salt Lake City,United States
neutral
34.0%
negative
22.0%
44.0%
positive


Seattle,United States
neutral
24.9%
negative
21.5%
53.6%
positive

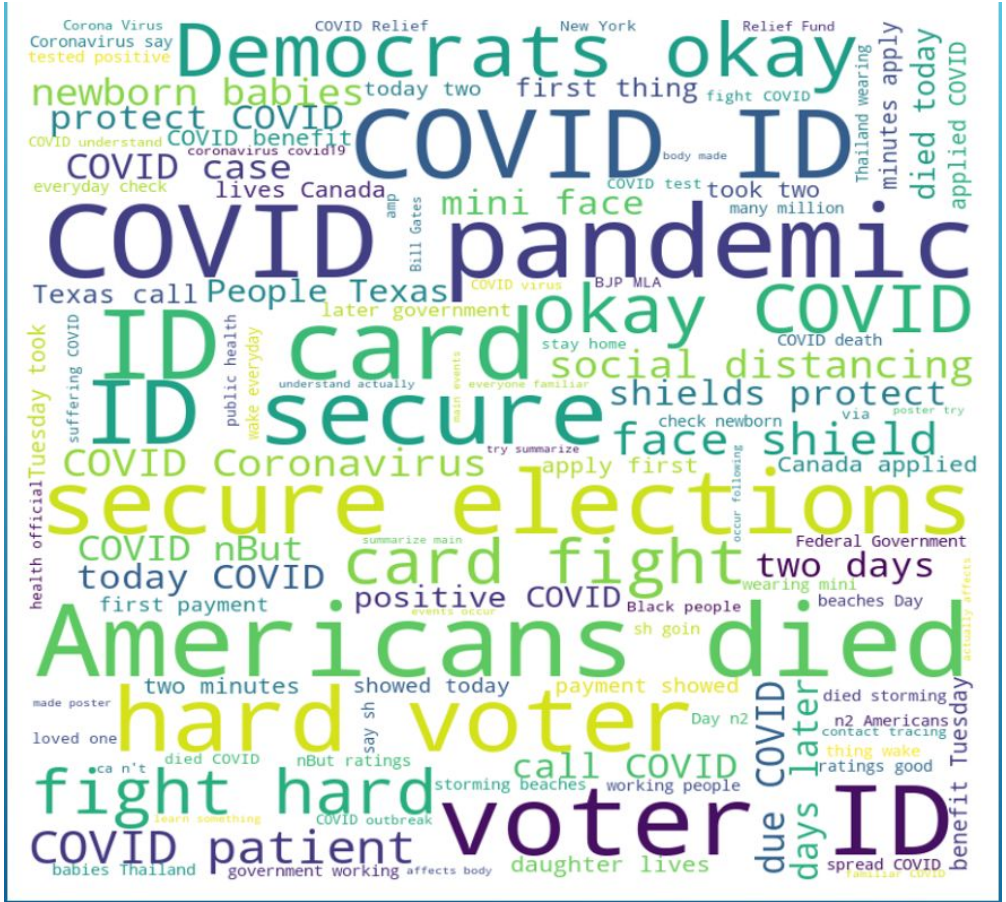

Syracuse,United States
neutral
33.9%
22.1%
negative
44.0%
positive

Result


Sentiment Analysis by City
BRITISH COLUMBIA
SASKATCHEWAN
ONTARIO
QUEBEC
WASHINGTON
MONTANA
NORTH DAKOTA
MINNESOTA
OREGON
IDAHO
WYOMING
SOUTH DAKOTA
WISCONSIN
MICHIGAN
Ottawa
MAINE
VT.
N.H.
N.Y.
IOWA
NEBRASKA
NEVADA
UTAH
UNITED STATES
ILLINOIS
OHIO
PA
INDIANA
WEST VIRGINIA
COLORADO
KANSAS
MISSOURI
KENTUCKY
VIRGINIA
CALIFORNIA
OKLAHOMA
TENNESSEE
NC
ARIZONA
NEW MEXICO
ARKANSAS
SC
ALABAMA
TEXAS
MISSISSIPPI
GEORGIA
LOUISIANA
FLORIDA
Gulf of Mexico

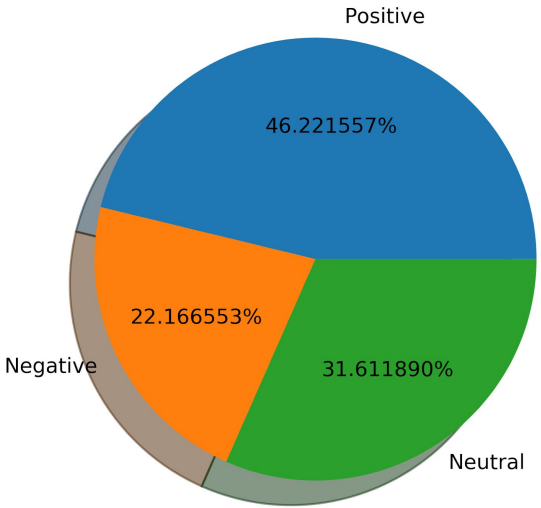

neutral
negative
positive
Denver,United States
Los Angeles,United States
Seattle,United States
New York City,United States
Syracuse,United States
Chicago,United States
San Francisco,United States
Salt Lake City,United States
Dallas,United States
Atlanta,United States
Miami,United States
Rico,United States
0
500
1000
1500
2000
2500
3000
3500
4000

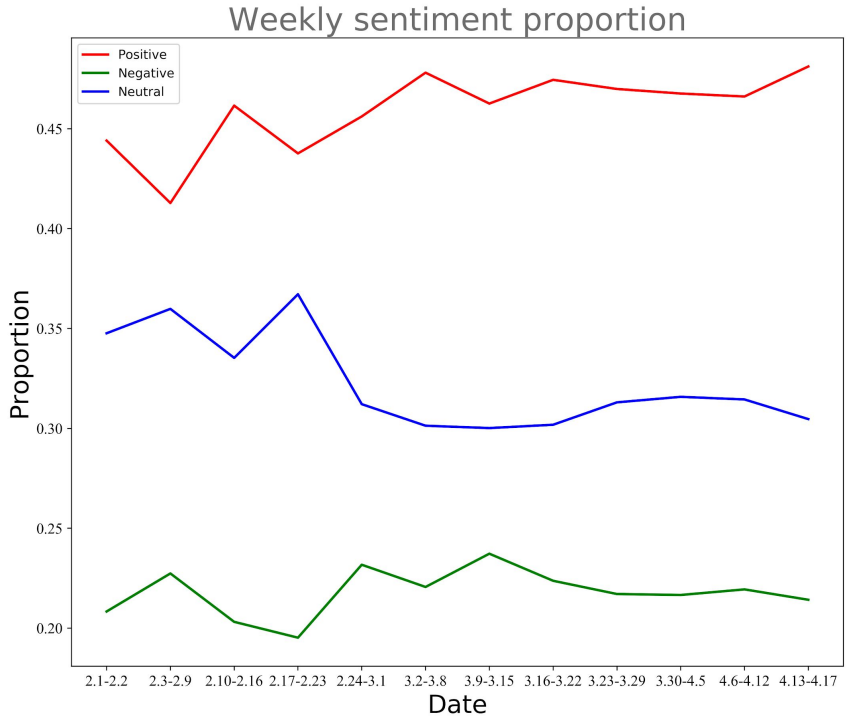
MOST FREQUENTLY MENTIONED TERMS



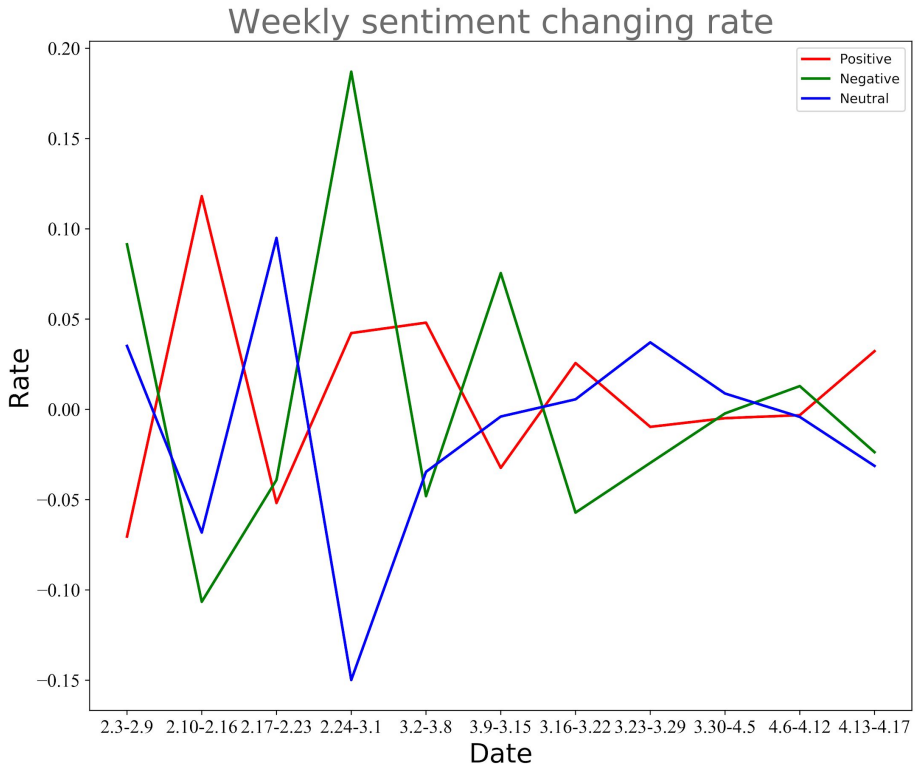
2.1-4.17 Sentiment analysis

Sentiment proportion 2.1-4.17





Weekly changing rate



News:

20 February

The United States confirmed one more case in California, bringing the total number to 16.

21 February

The United States confirmed 20 more cases, bringing the total number to 35. Furthermore, the Association of Public Health Laboratories (APHL) announced that only three states were capable of testing for the corona virus: California, Nebraska, and Illinois.

24 February

The United States confirmed 18 more cases including evacuated passengers from the cruise ship, bringing the total number to 53.

26 February

The United States confirmed three new cases - one being a domestic case in California with no travel history, and the other two being former passengers aboard *Diamond Princess* - bringing the total number to 60.

28 February

The United States confirmed 4 more cases, including 2 former passengers of *Diamond Princess*. Washington state authorities later confirmed two additional presumptive cases, bringing its total to 66. One had recently returned from South Korea, and the other case was unrelated and locally acquired.

29 February

The United States confirmed its first death, a man from Washington near the Seattle area.

Conclusion

Geography Distribution: There is no outstanding geographical difference in people’s attitudes towards COVID-19, which differs from our assumption.

Time Distribution: More and more people are holding neutral attitude towards this epidemic, while the number of people holding positive and negative attitudes are decreasing.