# Effectiveness of China's measures during the COVID-19 epidemic from the perspective of public opinion research

# Abstract

In this paper, through web crawler technology, I get the epidemic data, Baidu hot search, Sina Weibo hot search comments, video comments, CCTV News data, online shopping data, stock data, etc. Using Python's data processing library and visualization library to visualize the data. In this paper, I use the method of emotion analysis based on dictionary to analyze the crawled public opinion data. After constructing emotion dictionary, negative word dictionary, degree adverb dictionary and stopping word dictionary, I use Hidden Markov Model (HMM) to segment words, TF-IDF algorithm and TextRank algorithm to extract feature words, and use self-defined way to extract text emotion according to the dictionary propensity calculation. Combined with the time development and government measures, this paper analyzes the impact of the epidemic situation and the measures of the Chinese government from multiple perspectives.



**Keywords:** Internet Public Sentiment Web Crawler Feature Extraction Sentiment Analysis Visualization Analysis COVID-19

# Catalog

# 1 Introduction

## 1.1 Research background and significance

In the outbreak of COVID-19, the changes in the epidemic situation always attracts our attention. The Chinese government took measures to deal with it. With the development of time, the epidemic was under control in China. The resumption of work and rehabilitation began in succession. Meanwhile, the outbreak of the epidemic was abroad, and the multinational government should be adversely affected. The Chinese government continued to strictly control and prevent the epidemic in China once again.

The Chinese government's control of the epidemic twice is inseparable from the cooperation of policies and measures with the people. Public opinion effectively reflects the implementation of policies and measures. In the current era of Internet popularization, the analysis of network public opinion data can best reflect the real views and emotions of the masses, and the real effect of epidemic prevention measures can be seen. Considering the factors of time and space, the analysis of public opinion data can summarize the experience and carry out the work of fighting against the epidemic situation better.

## 1.2 Research content and research status at home and abroad

### 1.2.1 Web crawler technology

Web crawler is a program or script file that can automatically get the specified content from a specified web page. It is the basic part of a search engine.

The engineering principle of web crawler technology is: Based on the principle of multi thread concurrent search, visit each node in the Internet, set a certain period interval to grab web page information, at the same time extract and analyze other web page information through the link relationship between web pages; the searcher is used to sort the captured information on the page, and index the database; then, network crawls Starting from one or a group of Web sites, the insect constantly extracts web page information from the network, searches for the required web page information as soon as possible, and avoids dead links and invalid links by means of regular updates.

At present, the mainstream web crawler technology includes two types: the first is the general web crawler technology, that is, the information crawling from a group of source addresses to the entire web page; the second is the topic based web crawler technology, which selects the web pages related to the topic for information crawling. Among them, the topic based web crawler technology is a hot research topic in the field of network information capture. The technology takes the topic information as the starting point, and network users determine and obtain information through the selected web page topic.

### 1.2.2 Data visualization

The form of data visualization is result visualization, which can clearly display the mined information through visual symbols, and its core is to express complex data in a way that users understand. Visualization technology has developed rapidly in recent years. In the current big data era, visualization technology is regarded as an important data mining method in various fields. In the big data environment, it is very important to analyze and process the data visually and to mine its value. During the COVID-19 epidemic, the epidemic map, epidemic development and change map, personnel flow map and so on frequently appeared in the public view, and data visualization played a great role in the analysis of the epidemic situation.

In this paper, through the visual analysis of epidemic data and public opinion data, the complex data is expressed in a way that is easy to understand, and the epidemic map, epidemic development map, text word cloud map, etc. are made to make the epidemic public opinion data more intuitive.

### 1.2.3 Text sentiment analysis

Sentiment analysis is a process of obtaining the emotional tendency of the text after information mining with personal subjective consciousness. According to the length of the text to be processed, it can be divided into two categories: one is short text sentiment analysis, such as hot search, comment, news, etc. the other is long text sentiment analysis, such as novels and essays. At present, there are many research methods of text sentiment analysis. In this paper, we use the method based on emotion dictionary to match and extract the emotional words in the text by constructing the emotion dictionary as complete as possible. At the same time, we should fully consider the influence of other texts, such as negative words, degree adverbs and punctuation marks, to calculate the sentiment of the text through the two indicators of emotional polarity and polarity intensity perceptual value is used to judge the emotional tendency of the text. Experimental data show that the analysis results and cross domain text processing ability based on sentiment dictionary analysis method are better.

# 2 Related theory and technology

## 2.1 Web crawler

### 2.1.1 The concept of web crawler

Web crawler is a kind of program or script, which can automatically grab the Internet information according to certain rules. It is widely used in the field of Internet search engine. It can automatically collect and obtain the accessible web content, and then provide it to the search engine for further processing, so that users can retrieve the information they need more quickly.

### 2.1.2 Working principle

The working process of web crawler can be divided into three steps: data collection, processing and storage. Data collection generally starts from one or more initial web page addresses, then crawls the contents of web pages by matching or searching, and extracts new web addresses from the current page, and puts them into the URL queue to be fetched in a certain order. The whole process is executed circularly until the conditions are met, and then the captured data is sorted and stored to display the corresponding data.

### 2.1.3 Crawling strategy

Generally, it can be divided into depth first strategy, best priority strategy and breadth first strategy. Depth first strategy is a strategy to search leaf nodes from the root node. In a web page, select a hyperlink to search the web page in depth first to form a search chain until there are no other hyperlinks; the best priority strategy confirms the order by the similarity between the description text of the hyperlink and the target page; the breadth first strategy starts from the root node, searches and traverses the current level, and then completes the search for the next level search traversal, we can use breadth first strategy when we need to capture a lot of Web information.

### 2.1.4 Three kinds of filtering technology of web page data

(1) Regular expression is a kind of logical formula, which can operate on strings, form regular strings with specific characters, and filter out key data information.

(2) XPath path language can be used to obtain the element content of specified attributes in XML document. In this paper, we use Python lxml library to realize the function of extracting tag content of XML document.

(3) Beautiful Soup can automatically convert the content of the input document into Unicode code. After completing the data grabbing and filtering operations specified by the user, the output document is automatically converted to UTF-8

encoding, which is a kind of HTML / XML parser.

### 2.1.5 Component

(1) Main control module: carry out initialization work, generate initial URL and put it into URL queue to be crawled, download, parse and extract web pages, control and coordinate the workflow of each module;

(2) Web page download module: download the web page. For the web pages that can be accessed anonymously, the pages that need to be authenticated can be downloaded after the simulated users log in. For those who need digital signature or digital certificate to access the website, the corresponding certificate should be loaded into the program, and then the web page can be downloaded after verification;

(3) Web page analysis module: extract the information meeting the specified requirements from the web page and transmit it to the data cleaning module. The extracted URL address is transmitted to the URL scheduling module, and the data meeting the specific requirements is extracted and transferred to the data cleaning module;

(4) The first mock exam module of URL: receiving the URL address sent by the previous module and comparing it with the address that has been crawled before, and dropping the existing address and putting the URL address that is not present in the corresponding location.

(5) Data cleaning module: receiving the data from step 3, cleaning, sorting and putting it into the database;

(6) Data display module: process crawled data, display data by text or text, or store data.

## 2.2 Data analysis

### 2.2.1 Visual analysis

Visual analysis can explain COVID-19's development and related influence, and provide relevant clues and guidelines for epidemic prevention and control strategy.

As an object-oriented interpretive programming language, Python has many visual libraries, such as bokeh, plotly, Matplotlib, Seaborn, pandas, wordcloud and pyechart. In this paper, we mainly use pandas, wordcloud and pyechart.

Python's pandas library has a powerful ability of data statistical analysis, which can process various types of data such as TXT, CSV, dataframe, etc. it also has the function of data visualization. The underlying foundation of visualization is Matplotlib. Python's wordcloud library is a word cloud visualization library with powerful customization function. Some of the word cloud images in this paper are generated by wordcloud library, as shown in Figure 2.1. Echarts is an open source visualization library implemented by JavaScript, which is compatible with most of the current browsers. Pyecharts is a python class library that can generate echarts charts.

```
# 绘制词云图
# 设置词云参数
wc = WordCloud(font_path="SimHei.ttf", #字体路径（SimHei.ttf黑体字）
               background_color="white", #词云图背景颜色
               max_words=2000,   #显示词的最大数量
               max_font_size=200,  #最大字体
               random_state=42  #为每个单词返回一个PIL颜色
                                #PIL模块是Python中处理图像的标准库
              )
wc2 = wc.fit_words(word_count) #根据词频生成词云
# 绘制词云图
plt.figure(figsize=(16,8)) #绘制
plt.imshow(wc2)            #对图像进行处理
plt.axis("off")            #关闭坐标刻度
plt.show()                 #显示处理后的结果
wc.to_file("wordcloud.png") #输出到文件
```

Figure 2.1 Making of word cloud

2.2.2 Text sentiment analysis

Text sentiment analysis is a sub domain of natural language processing. Emotional information in text data is extracted by data mining through computer language. The current research directions include subjective identification, emotional tendency calculation, emotional information classification, etc. the research levels are divided into word level, sentence level and text level according to the granularity of the text to be processed.

This paper uses the emotion analysis method based on dictionary, the steps are as follows:

(1) Dictionary construction

Sentiment analysis methods based on dictionaries usually regard the text as a collection of words, mining keywords such as emotional words, negative adverbs, degree adverbs and so on in the text, and through the establishment of certain rules, the emotional score of the text is calculated as the emotional tendency of the text.

Dictionary preparation:

① Emotion Dictionary: This paper combines the commonly used emotion dictionaries in the field of Chinese sentiment analysis to form a new emotion dictionary, and adds words of corresponding fields according to the specific analysis data. For example, epidemic domain words can be added when analyzing epidemic data, and "outbreak" is defined as negative words.

② Dictionary of negative words: when negative words appear in a sentence, such as no, none, nothing etc., the emotional polarity of the sentence will change.

③ Dictionary of Degree Adverbs: the absolute value of the score indicates the emotional strength of the text to be tested. Therefore, if the degree word is modified, the basic score of the word should be multiplied by different coefficients according to its degree.

④ Stop word dictionary: stop words usually refer to a kind of words with high frequency but lack of emotional analysis meaning. If stop word filtering is not carried

out, the number of words in the segmentation set will be increased, resulting in the increase of calculation amount and waste of computing resources. There are 852 stop lists in this paper after reorganizing and recombining according to the existing resources on the Internet.

(2) Text preprocessing

The preprocessing process of text data includes data cleaning, word segmentation, stop word processing, word frequency statistical sorting and so on.

In this graduation project, we use Python pandas library and numpy library to clean the text data to be analyzed, use stuttering Chinese word segmentation to segment the text, use TF-IDF Algorithm and TexRrank Algorithm to extract keywords, and then use pandas library for word frequency statistics and sorting.

(3) Affective tendency calculation

Text sentiment classification is an important part of text sentiment analysis. The main research is to identify and classify the emotional information in text. This paper classifies text sentiment by calculating emotional tendency.

## 2.3 Web application development of Python

Python has a variety of web development frameworks, this paper uses Python Flask for graduation design.

### 2.3.1 Introduction of Python Flask

Flask is a Python based web development framework, with three-tier architecture design, including control layer, business layer, data layer, suitable for the development of lightweight web applications. The structure of flask is extensible, and it is convenient to add other functions.

### 2.3.2 Three layer architecture design of Python Flask

(1) Control layer: it is responsible for receiving the URL, parameters and request method of the mobile terminal, and forwarding them to the business layer. Specifically, it refers to the business logic on the routing function of calling app.

(2) Business layer: it is responsible for parsing the incoming data, and the operations related to the database will be handled by the data layer, and the data returned from the database will be returned to the mobile terminal after certain processing. The functions in the business layer will first parse the incoming HTTP request type (usually get and post), and then execute the corresponding business logic according to the HTTP request type.

(3) Data layer: responsible for interaction with database. Flask Sqlalchemy is a flask extension, which has the basic functions of relational database framework Sqlalchemy. It uses Flask Sqlalchemy to complete data layer related operations.

### 2.3.3 Python Flask project file design

By default, the flask project includes static (static files, such as JS, CSS),

templates (template files, such as HTML), and app.py Script file, only 7 lines of code can complete the construction of the flask project, as shown in Figure 2.2.

```
1  from flask import Flask # 从flask库导入Flask模块
2  app = Flask(__name__)   # 定义flask app
3  @app.route('/')     #控制层路由
4  def index():        #业务层函数名
5      return 'index' #业务层逻辑实现
6  if __name__ == '__main__': #函数主入口
7      app.run()       #启动flask app
```

Figure 2.2 Basic structure of flask project

# 3 Design and Implementation

## 3.1 Overall design

This paper is divided into six parts: epidemic map, epidemic development, public opinion analysis, study resumption analysis, online shopping analysis and economic analysis. The data of the six parts are analyzed visually. Combined with the development of time and the implementation of measures, the data changes are compared and analyzed to test the effectiveness of policy measures to deal with the epidemic situation.

## 3.2 Data acquisition

Using web crawler technology, crawling data from official website or API interface, and processing data to obtain target data. The method of data acquisition is legal, and the data source is official and scientific.

## 3.3 Public opinion analysis

### 3.3.1 Baidu public opinion

As a Chinese Internet search portal, Baidu covers more than 95% of domestic netizens. To a certain extent, Baidu hot search has become an important reference to reflect the most real and objective concerns of our people.

Crawl Baidu national hot search and Baidu area hot search, according to the hot search content to analyze the issues that people are most concerned about, and look at the changes of people's concerns according to the development of time.

To crawl Baidu hot search "Today's Hot spot" as an example for analysis and processing, as shown in figure 3.1 and figure 3.2.

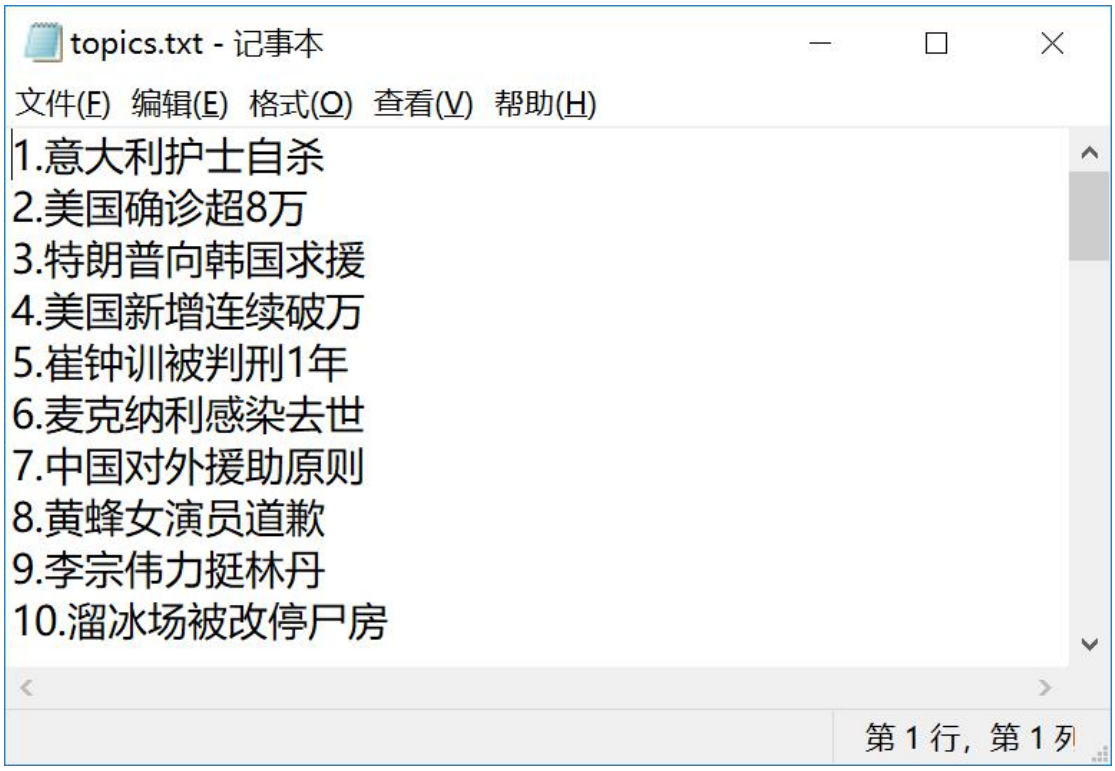


Figure 3.1 crawling Baidu hot search

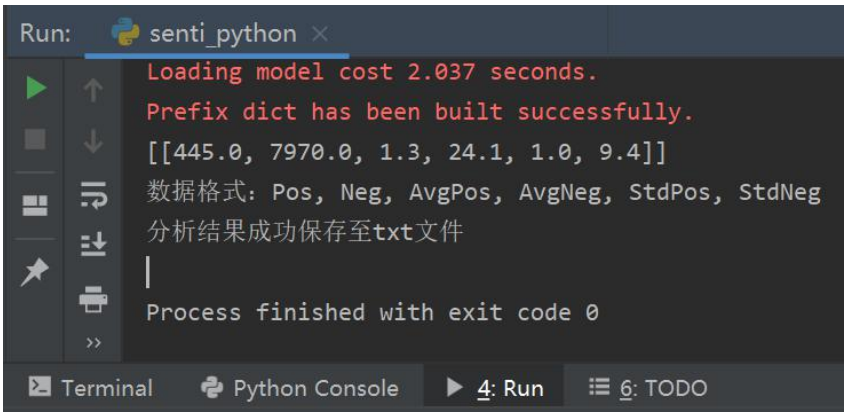


Figure 3.2 positive and negative emotional analysis of short text

Based on the analysis of Baidu hot search data and regional hot search data from January 2020 to April 2020, we can see that the hot search data are basically the same all over the country, and people in Hubei Province pay more attention to the epidemic.

When the epidemic is serious, some people are pessimistic, and on the whole, the masses generally hold a positive attitude towards the epidemic. With the development of time and the implementation of measures, the attitude of the masses on the epidemic is becoming more and more optimistic and positive.

### 3.3.2 Sina Weibo public opinion

Sina Weibo is similar to twitter, which is widely used by Chinese people. It has a

wide influence. In the process of fighting against epidemic situation, Sina Weibo resolutely implements the important directive spirit of President Xi on epidemic prevention and control work. It sticks to the correct political direction, public opinion orientation and value orientation in epidemic prevention propaganda work, and over 2,6000 government official micro-blog participated in the epidemic information release. More than 550000 related Weibo have been published, with more than 11.4 billion Weibo read.

Compared with news, blogs, forums and other types of users, Weibo users can get the latest information more easily. Hot search on Weibo can see current hot topics, and using Weibo has a strong sense of interaction and participation. Can express and respond in a timely manner, which fully reflects the user's right to speak.

Therefore, the analysis of Weibo public opinion will help to timely judge and guide the factors that form group problems and effectively control social public opinion.

Using web crawler technology, crawl Weibo under the specified hot topic of Weibo within a specified period of time, analyze the emotion of Weibo under this topic, and understand people' mentality to a topic.

Take the topic "The opening of Xiaotangshan Model Hospital in Wuhan" as an example, as shown in figures 3.3.

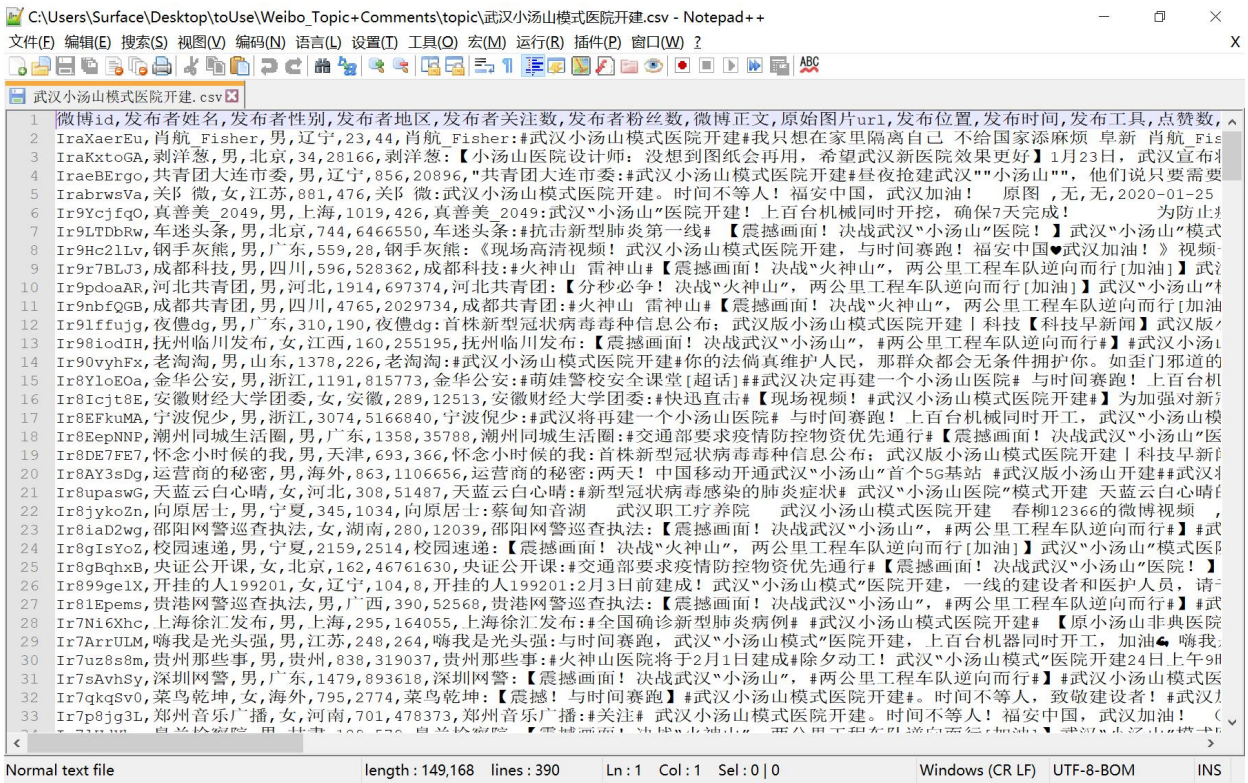


Figure 3.3 Get weibo data

Using web crawler technology, crawl the comments of the designated popular Weibo, analyze the emotion of the comment data, and understand the people' state of mind to a certain content. Take People's Daily's Weibo comments under the topic

"Construction of Wuhan Xiaotangshan Model Hospital" as an example, as shown in figures 3.4 .

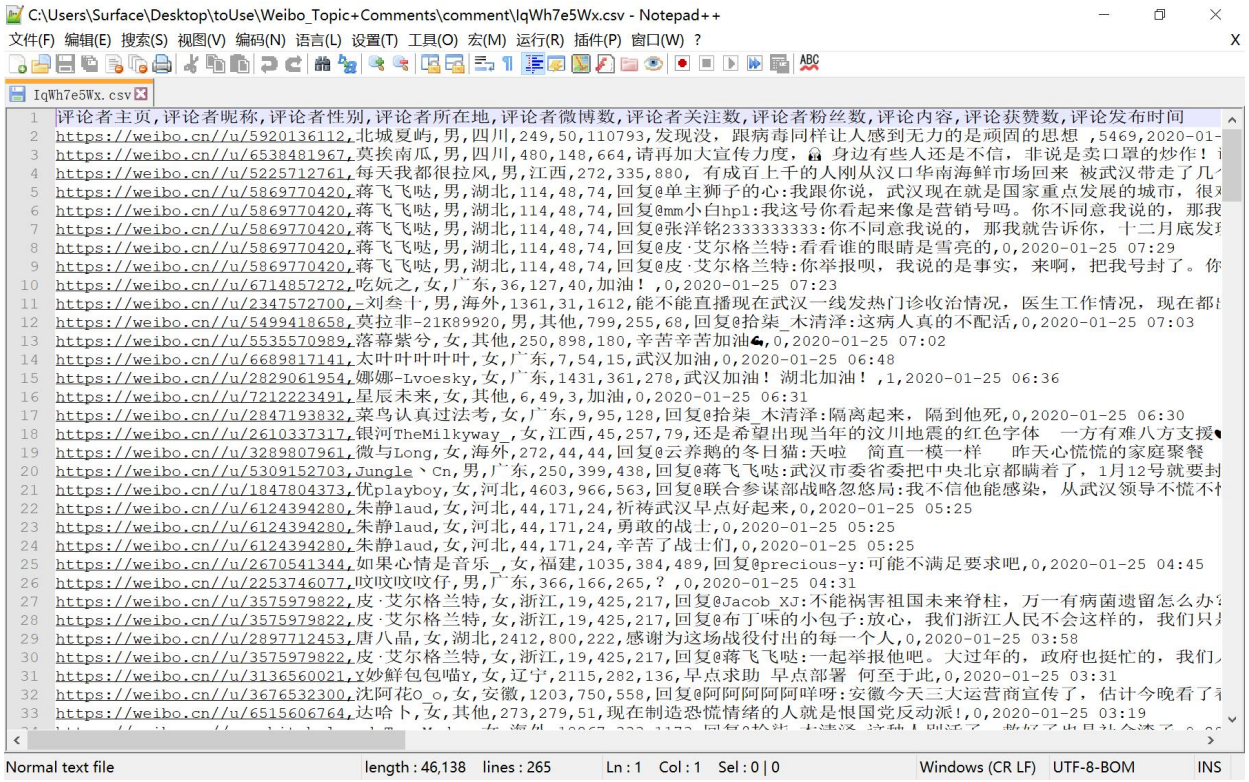


Figure 3.4 Get comments data

Through the analysis, we can see that in the early stage of the epidemic, people reacted strongly to the opening of the Xiaotangshan Model Hospital in Wuhan, optimistic about the role of the hospital, and felt sad for the victims of the epidemic, while being afraid of the unknown epidemic.

This paper analyzed many popular searches on Weibo, such as "The construction of Wuhan Xiaotangshan Model Hospital", "COVID-19" and "Health release", as well as the comment areas of the popular Weibo. We can see that during this epidemic, people paid high attention to the epidemic. With the continuous release of policies and measures, all parties actively cooperate, the mentality of people from the initial fear and hope, to hope to overcome the epidemic.

Thus it can be seen that the state media have implemented the spirit of President Xi's instructions, released authoritative information in a timely manner, disseminated scientific methods, and responded to hot issues, and played an important role in guiding public opinion. It can be seen that the policy measures are more effective at the same time.

### 3.3.3 Video comment on-screen of public opinion

Video comment on-screen comment is one of the main ways for people to express their views in recent years. On-screen comment is a real-time interactive comment way to break through the limitations of real time and space. The analysis of

video comment is one of the important ways to study the opinions of the masses.

This paper takes the two most representative video platforms, Tencent Video and Bilibili, as examples to analyze and study the public opinion on the video.

（1） Bilibili APP

Bilibili, as one of the largest original video platforms in China, has high user activity. Bilibili's data analysis can reflect the mentality of the masses on the epidemic to a certain extent.

The keyword "new-type coronavirus" is used as an example, as shown in figures 3.5.

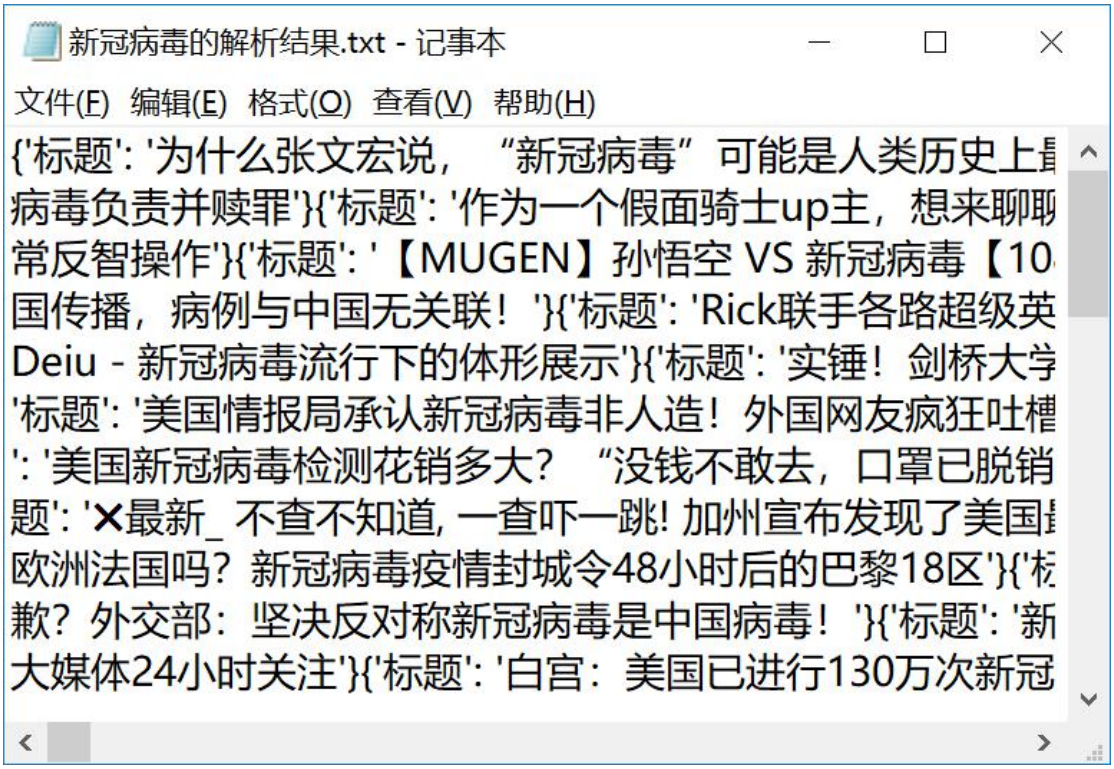


Figure 3.5 Get all the related video titles with keywords and co-exist as a TXT file

According to the analysis, Bilibili up masters' works on the topic of "new-type coronavirus" generally show a positive attitude and spread the positive energy of the network.

The following is an analysis of people' attitudes towards related videos through the comments area. Bilibili video as an example, with the keyword "new-type coronavirus" in the work No. 93727719, it is titled "what happened in the past two months when new-type coronavirus broke out?" , obtain video-related information, user name and comment content, and analyze the data, as shown in figures 3.6 and 3.7 .

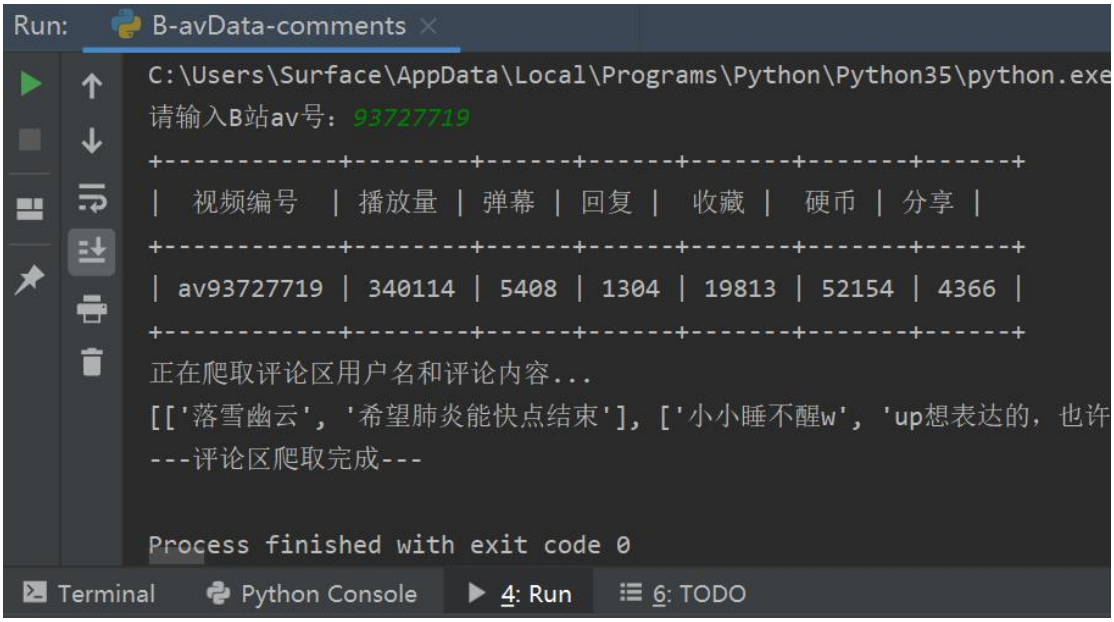


Figure 3.6 input AV number of station B video to obtain relevant information

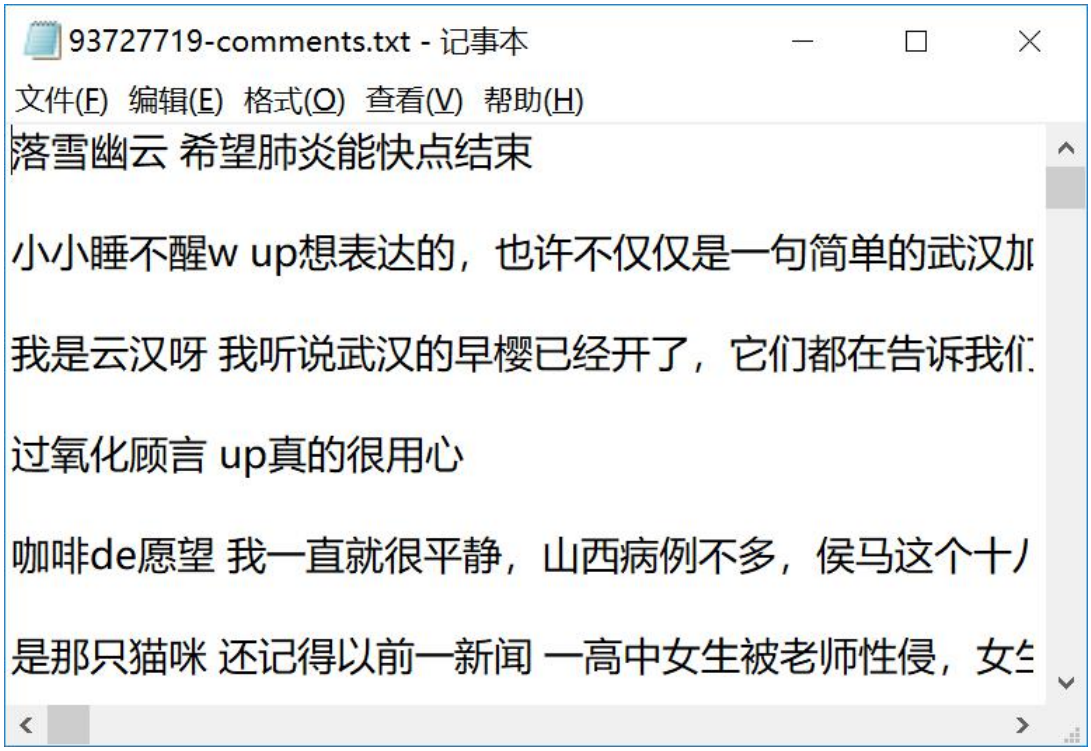


Figure 3.7 getting the comment area information

According to the analysis results, people' attitude towards the video is generally positive, and they are full of confidence in overcoming the epidemic situation.

This paper analyzes the video title short text corresponding to several keywords such as "Wuhan epidemic situation", "Hubei Province epidemic situation" and "fighting pneumonia", as well as the related video comment area containing keywords. It can be seen that during the epidemic period, users of Bilibili spontaneously produced videos of the introduction of new coronavirus, introduction of personal

epidemic prevention measures, and development status of epidemic situation in various places Positive message interaction, mutual encouragement and positive energy transmission.

（2） Tencent Video

Tencent Video is one of the largest video playing platforms in China. It can crawl through the video barrage to understand the mentality of people. According to the crawling comments, we use word segmentation technology and stop word dictionary to screen out valuable words. After counting the word frequency, we draw the word cloud, as shown in figures 3.8 .


Figure 3.8 Tencent video bullet screen word cloud

According to the analysis of the barrage of epidemic related videos broadcast by multiple Tencent video platforms, it can be seen that people are generally optimistic about the epidemic situation and agree with government policies.

（3） Public opinion analysis of video barrage comments

During the epidemic period, people understood the epidemic situation by watching relevant videos, and participated in the interaction through comments. The relevant videos generally spread positive content and people' mentality was generally optimistic.

### 3.3.4 News data analysis

During the epidemic period, news is our main source of information. Among the news media, CCTV has a wide audience and is authoritative as an official media. Therefore, the news data analysis refers to CCTV news articles.

Use web crawler technology to crawl all CCTV news releases from December 2019 to April 2020 for data analysis. Use TextRank Algorithm to extract keywords from the news release, record the global word segmentation and count the word frequency. Select the keywords that appear more frequently to make a cloud chart to reflect the main news content of the month.

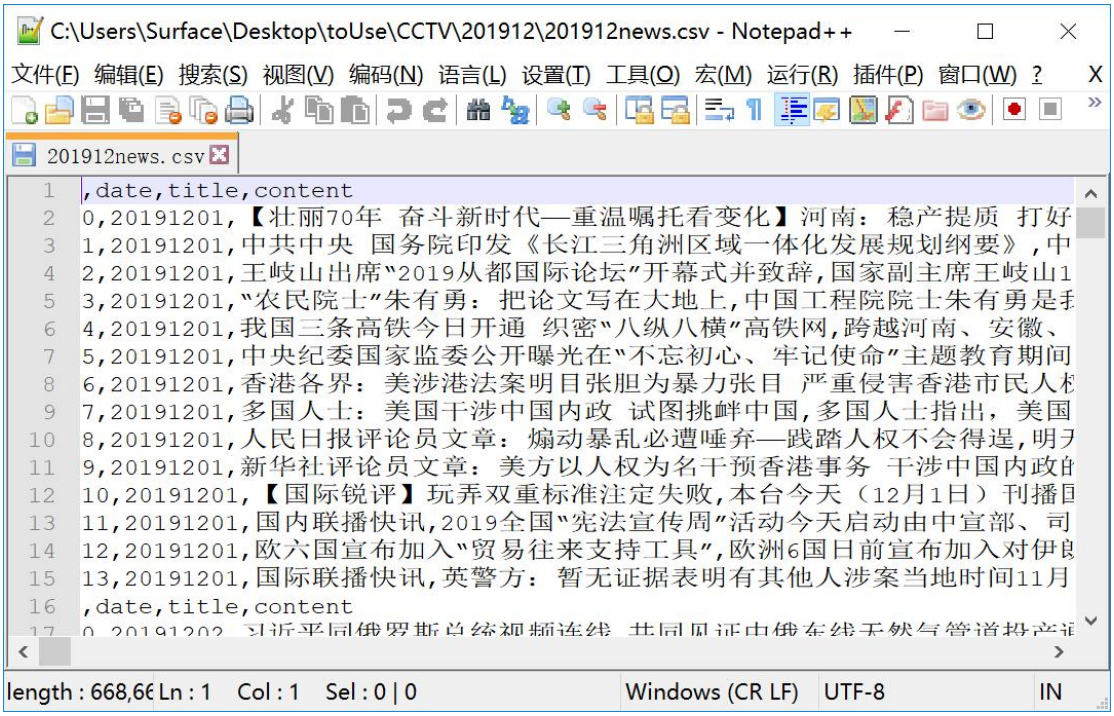


Figure 3.9 All CCTV press releases in December 2019


Figure 3.10 Select the keywords that appear frequently in the press release in December to make a cloud chart

According to the data analysis, it can be seen that:

In December 2019, Wuhan Health Commission reported that there were patients with pneumonia of unknown cause. According to the news analysis of that month, we can see that it did not cause high attention. In January 2020, at the beginning of January, Wuhan continuously reported patients with unexplained pneumonia, which attracted the attention of all parties. On January 15, the National Center for Disease Control and prevention launched level I emergency response, and the news will release the latest news of pneumonia every day. In February 2020, with the continuous development of the epidemic situation in China, the first level emergency response has been launched one after another. Due to the shortage of medical materials, Hubei Province is in a difficult situation, and medical teams have been sent to support them. The resumption of work and school needs to be postponed. In March 2020, with the policy intervention and the people's cooperation, the epidemic situation reached its peak and began to decline gradually. The resumption of work and production was carried out in an orderly manner according to the epidemic situation in various regions. In April 2020, the epidemic situation in foreign countries worsened. China sent medical teams to carry medical materials to help countries and donated medical materials and equipment to help countries. In China, the resumption of work and production was carried out in an orderly manner.

#### 3.3.5 Summary

After the COVID-19 epidemic appeared in China, people had panic in the face of this unknown infectious disease. The Chinese government promptly took a series of measures, such as isolation, city closure and testing. The public responded positively and had the confidence to overcome the epidemic. With the long-term cooperation of the government and the people, the epidemic situation was gradually under control, and the people were full of hope for overcoming the epidemic situation.

## 3.4 Analysis of returning to school

#### 3.4.1 University distribution

The college entrance examination website uses web crawler to crawl the college information, and the crawling content is saved as a college_ data.csv Visual analysis of the data was performed using pyecharts, as shown in figures 3.11 to 3.13 .

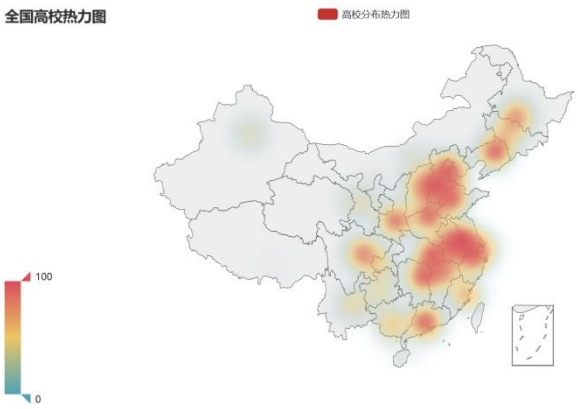


Figure 3.11 thermal diagram of National Universities

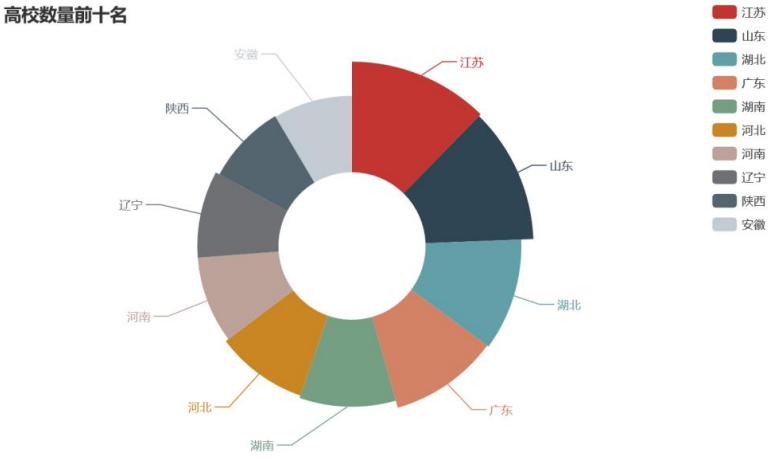


Figure 3.12 top ten universities

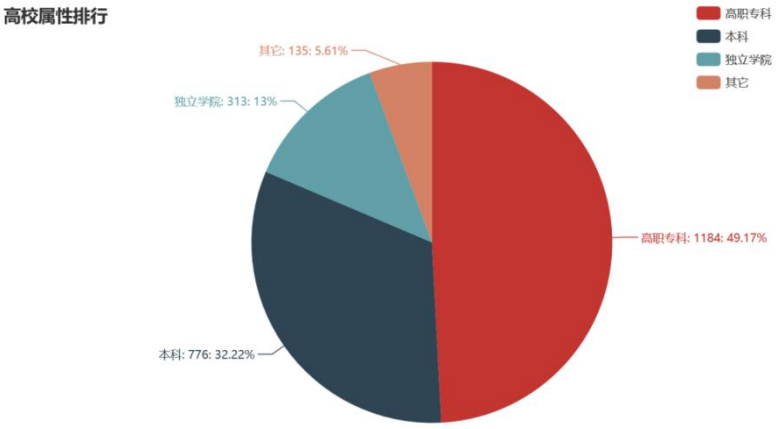


Figure 3.13 ranking of University attributes

From the distribution of the number of colleges and universities in China, the

number of colleges and universities in the eastern developed areas is more. In terms of the quality distribution of colleges and universities in China, the top three universities in 211 are Beijing, Jiangsu Province and Shanghai, and the top three universities in 985 are Beijing, Shanghai Province and Shandong Province. Among them, the number of 211 universities in Hubei Province ranks the fifth, and the number of 985 universities ranks the 11th.

3.4.2 Online course software

During the epidemic period, the online teaching method of "suspension of classes without suspension" and "cloud classroom" has become the mainstream in special period. There are many apps for online teaching, and the quality of software varies. Among the first batch of registered educational applications released by the Ministry of education, nailing was selected. Since then, nailing has become the online course platform choice of most schools. Using web crawler technology to crawl Apple's app store's nail rating and comments, and analyze the status quo of the user experience of the nail.

According to the 500 pieces of data in the first 10 pages, the number of people corresponding to one star, two stars, three stars, four stars and five stars are 320, 14, 10, 21 and 135 respectively. According to the content of the comment area, the proportion of comments with keywords is counted, as shown in Figure 3.14.

```
--在爬取的前10页500条评分数据中
--钉钉1-5评分分值出现的次数：
(array([1, 2, 3, 4, 5]), array([320,  14,  10,  21, 135], dtype=int64))
--统计含主要关键词的评论出现的比例
----五星出现的比例为：  7.60%
----一星出现的比例为：  8.80%
----好评出现的比例为：  9.40%
----差评出现的比例为：  2.80%
----老师出现的比例为：  2.80%
----学生出现的比例为：  6.60%
----分期出现的比例为：  5.80%
----下架出现的比例为：  5.40%
----垃圾出现的比例为：  5.40%
----加油出现的比例为：  1.40%

Process finished with exit code 0
```

Figure 3.14 proportion of scoring keywords

According to the crawling comments, the word segmentation technology and the stop word dictionary are used to screen out the valuable words. After counting the word frequency, the word cloud is drawn, as shown in Figure 3.15.


Figure 3.15 word cloud

According to the analysis, nail software quality is higher, user experience is better, some student users are dissatisfied with the compulsory download of the school, they vent their emotions on the software score, resulting in low software score. Furthermore, it can be seen that some students have resistance to online classes, and the normalization process of online teaching needs to be further explored.

3.4.3 Analysis

The types of colleges and Universities Participating in the statistics include undergraduate, independent college, higher vocational college, etc. The research shows that local students are more likely to go to local school. The key universities represented by 211 and 985 have a wide range of students, with a large number of foreign students. Kindergartens, primary schools, junior middle schools, senior high schools and technical secondary schools have many local residents. Therefore, the student structure can refer to the situation of local non-key universities.

In view of the differences in student structure, local epidemic situation, physical quality and age, and enrollment status, the regions meeting the three conditions given by the Ministry of education can consider resuming school. The third year of junior high school and the third year of senior high school will return to school first, and the students and teachers with important scientific research tasks will return to school first, and other grades will be determined. The time of resumption of 985 and 211 universities will be determined It should be discussed carefully in light of the specific situation.

President Xi stressed at the Standing Committee meeting of the Political Bureau of the CPC Central Committee in May 14th that we should strengthen the anti rebound work in key areas and key places, and highlighted the epidemic prevention work in Hubei Province, Beijing, Heilongjiang Province and Jilin Province. The number of 985 and 211 colleges and universities in Beijing ranks first in China, which coincides with the two sessions. Beijing should be more careful in determining the time for returning to school than other regions. Wuhan, Hubei Province, as the source of domestic epidemic, should continue to strictly prevent and control. Up to the time of the meeting, the time for the resumption of universities in Heilongjiang Province has not yet been determined, and the grade of Jilin Province's reunification has been

changed to online teaching. Since June 6th, universities in Beijing have qualified colleges and universities to arrange graduating grades to return to school in succession. The time for the resumption of universities in Hubei Province has yet to be determined, and the above decision is in line with the spirit of President Xi's speech.

## 3.5 Online shopping data analysis

Affected by the epidemic situation, the demand for protective equipment has increased greatly, and there are obvious phenomena of bid up and selling fake products. It is rarely available in offline physical stores, and online purchase has become the main channel. According to users' habits, they generally choose Tmall APP's Medicine Division online shops to purchase. This data analysis takes the representative data of Mask online shopping as an example.

After being controlled by the relevant departments, the website crawler technology was used to crawl the comment data of mask from Tmall APP, word segmentation technology and stop word dictionary were used to screen out valuable words. After counting word frequency, word cloud was drawn, as shown in Figure 3.16 .


Figure 3.16 word cloud of mask comments

According to the results of data analysis, the quality and price of masks can be guaranteed and the user experience is good.

After analyzing a variety of epidemic related protective equipment, it can be seen that in the early stage of the epidemic, the quantity and quality of protective equipment are in short supply. The government timely supervises and factories work overtime. With the mutual cooperation of the government and enterprises, the situation has improved significantly. People can obtain qualified protective articles through online shopping in time.

## 3.6 Economic data analysis

The epidemic has had a great impact on the economy. This year's government work report did not put forward a specific target for economic growth. Premier Li Keqiang pointed out in the report that the reason for this is that the global epidemic situation and economic and trade forms are highly uncertain, and economic development is affected by some unpredictable factors. Therefore, it is necessary to analyze the economic data affected by the epidemic situation.

### 3.6.1 Relationship between GDP and epidemic situation

Gross Domestic Product (GDP) is often used to measure the economic situation of countries. This paper intends to explore the relationship between GDP and the epidemic situation. Countries with GDP over 0.1 trillion in 2019 are selected to make GDP map according to the amount, and epidemic map is drawn according to the epidemic data on March 25, 2020, as shown in figures 3.17 and 3.18 .

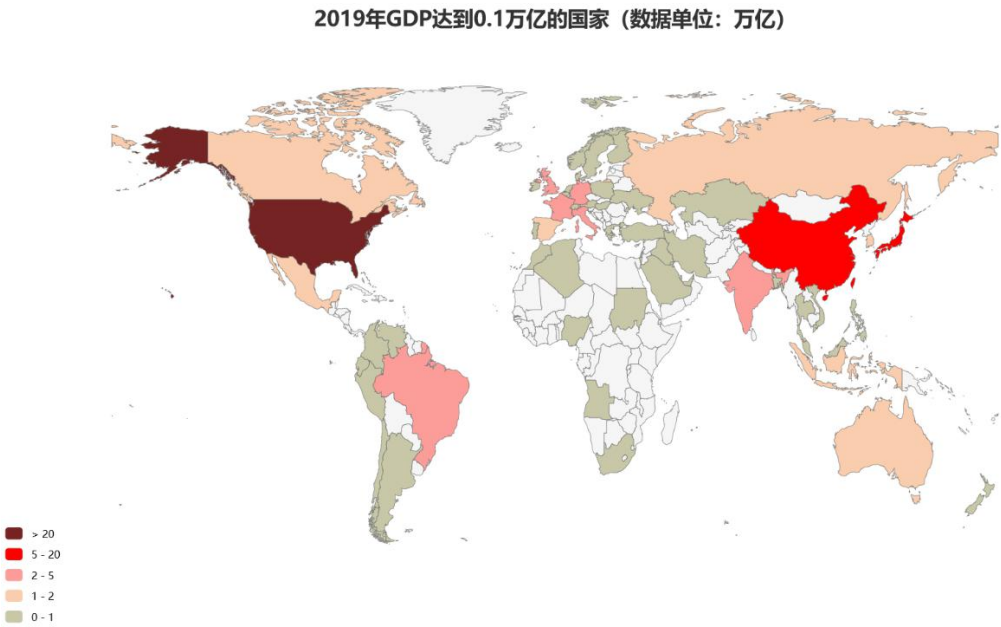


Figure 3.17 GDP map

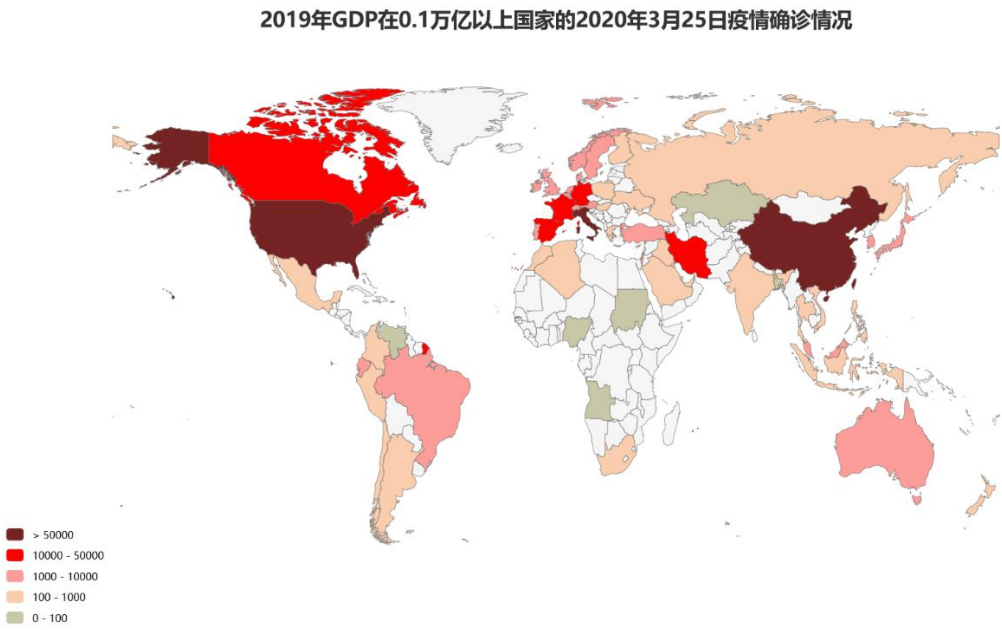


Figure 3.18 Epidemic map

It can be seen that there is a positive correlation between the GDP and the diagnosis of the epidemic situation. Considering the factors such as the total population, population density, medical level, traffic level and policy measures, countries with higher GDP generally have larger population and population density, generally better traffic conditions, and greater risk of infection among the masses. As the medical level is generally high, they have the ability to treat more people Therefore, the number of confirmed cases is high.

3.6.2 Stock trend analysis

The outbreak of COVID-19 in China has affected the financial market, and has seriously affected investor confidence. The stock market is under great pressure.

The epidemic situation of infectious diseases is closely related to the financial market turmoil. If the financial crisis is caused by the epidemic situation of infectious diseases, there will be a vicious circle between the epidemic situation of infectious diseases and the financial crisis. This is because the epidemic situation of infectious diseases and financial crisis will lead to macroeconomic recession, the income of various economic entities will decline, leading to the decline of the ability to deal with the epidemic situation of infectious diseases, accelerating the spread of infectious diseases and forming a more serious financial crisis. The financial crisis has led to a decline in economic growth, exacerbated economic inequality, and made the poor more likely to become potential communicators, making infectious diseases more prevalent .

In the overall economic downturn, the pharmaceutical industry is optimistic because of the epidemic situation. This paper makes a visual analysis on the representative medical stocks.

Moving average and daily trading volume are important indicators to judge the stock market. This paper uses these two indicators to visually analyze the stock

market data.

Moving Average (MA) is commonly used to observe the trend of securities price changes, and to count the securities prices within a certain period of time. After averaging the prices, the average value can be connected to form a MA line. MA5 is the five day moving average, MA10 is the ten day moving average, and MA20 is the twenty day moving average, which represent the weighted average prices of stocks on five days, ten days and twenty days respectively.

Stock trading volume refers to the specific number of transactions in a certain period of time, and the daily trading volume reflects the trading situation of the stock on that day.

To facilitate the analysis of the impact of the epidemic on the stock market, python flask is used to create a small stock trend query website. The data is updated in real time according to the date. After the user registers and logs in, he / she can input the Chinese name of the stock and query date range to query the moving average and trading volume, as shown in figures 3.19 to 3.24 .

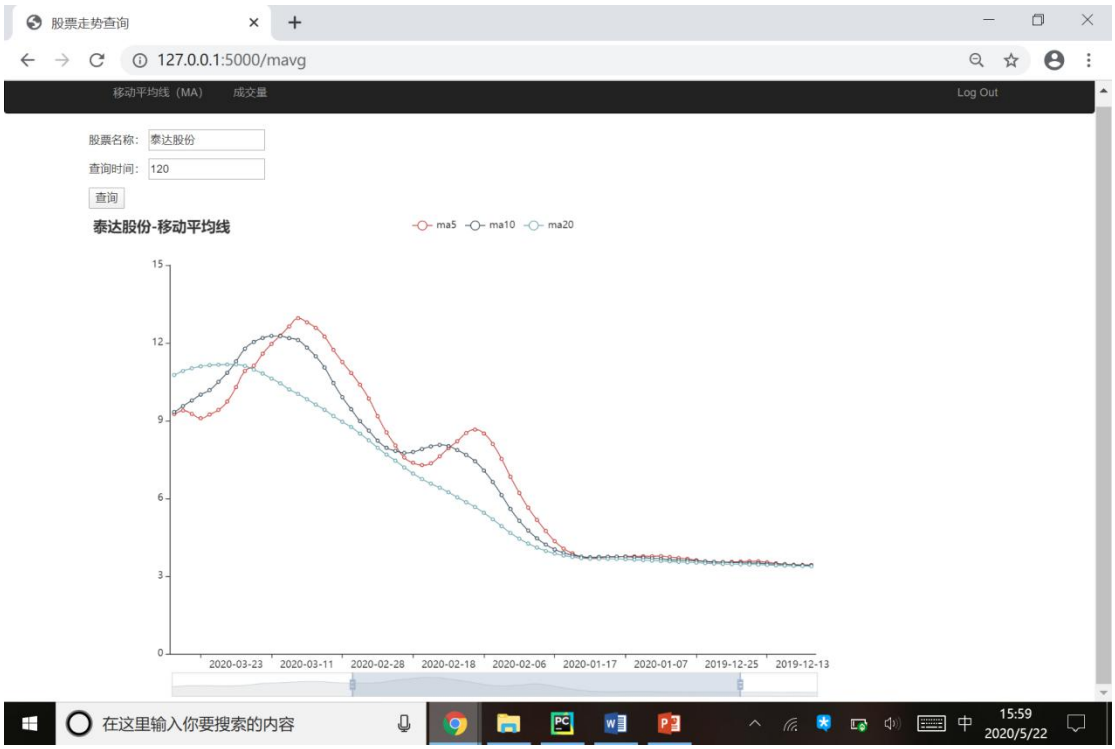


Figure 3.19 click "query" to display the data

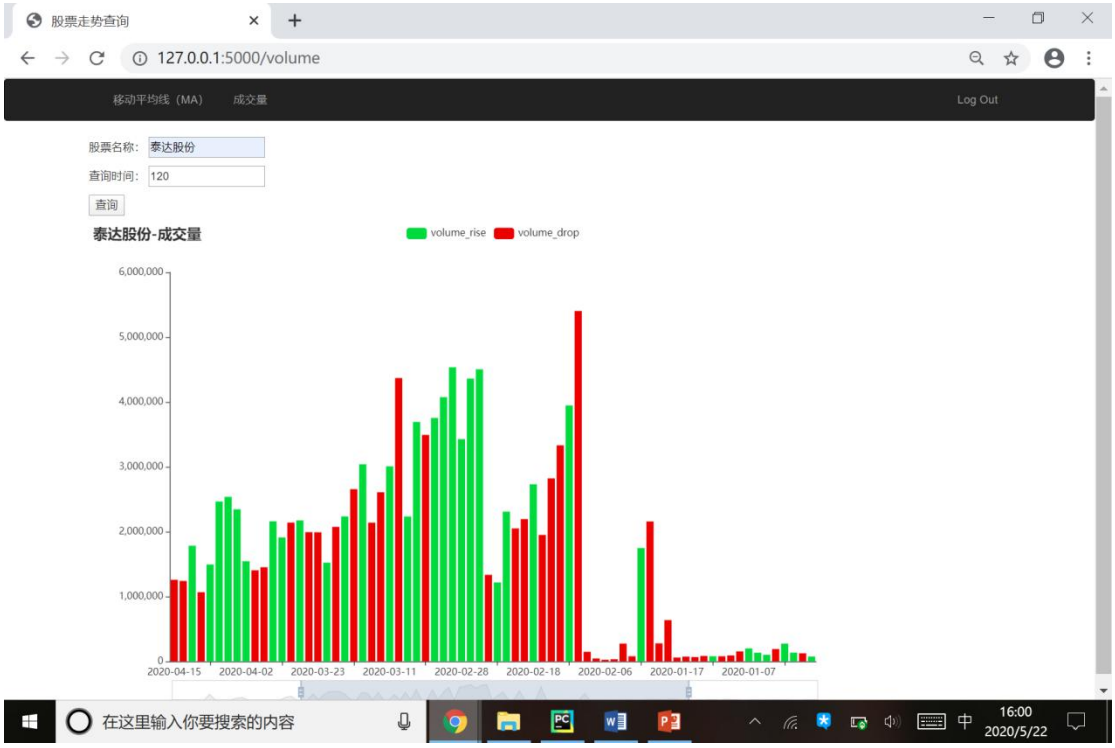


Figure 3.20 query volume data

TEDA shares is the largest manufacturer of mask filters in China. After the outbreak of the disease, the demand for respirators increased dramatically.

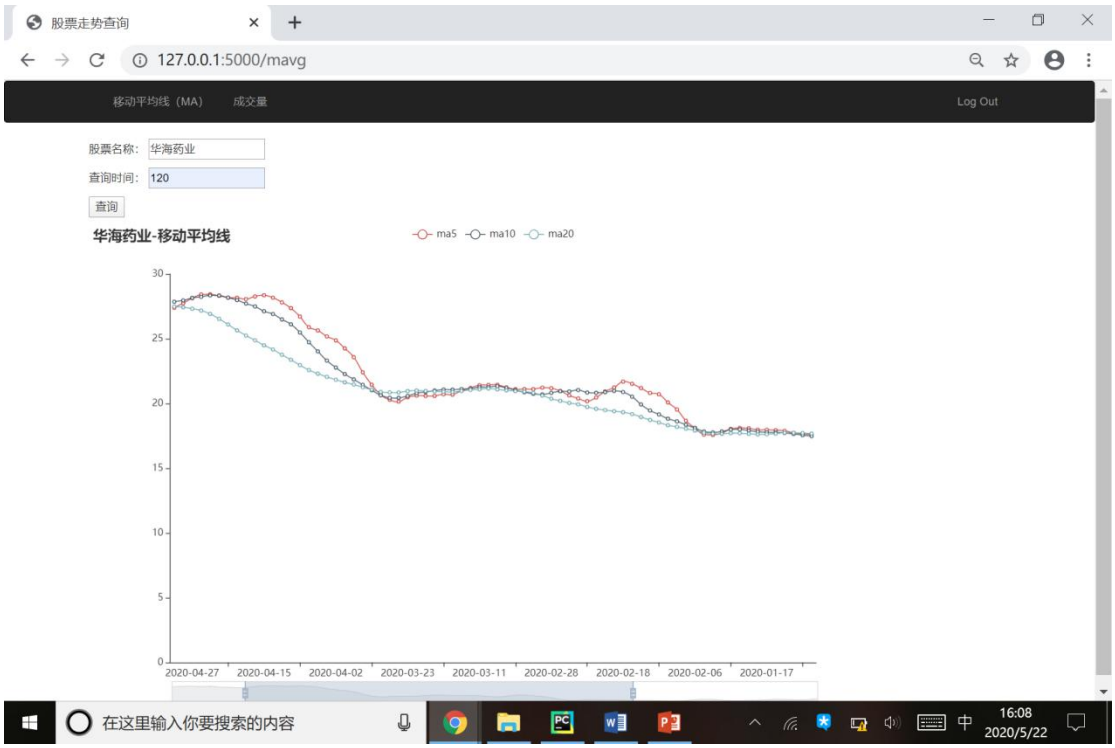


Figure 3.21 MA of Huahai pharmaceutical

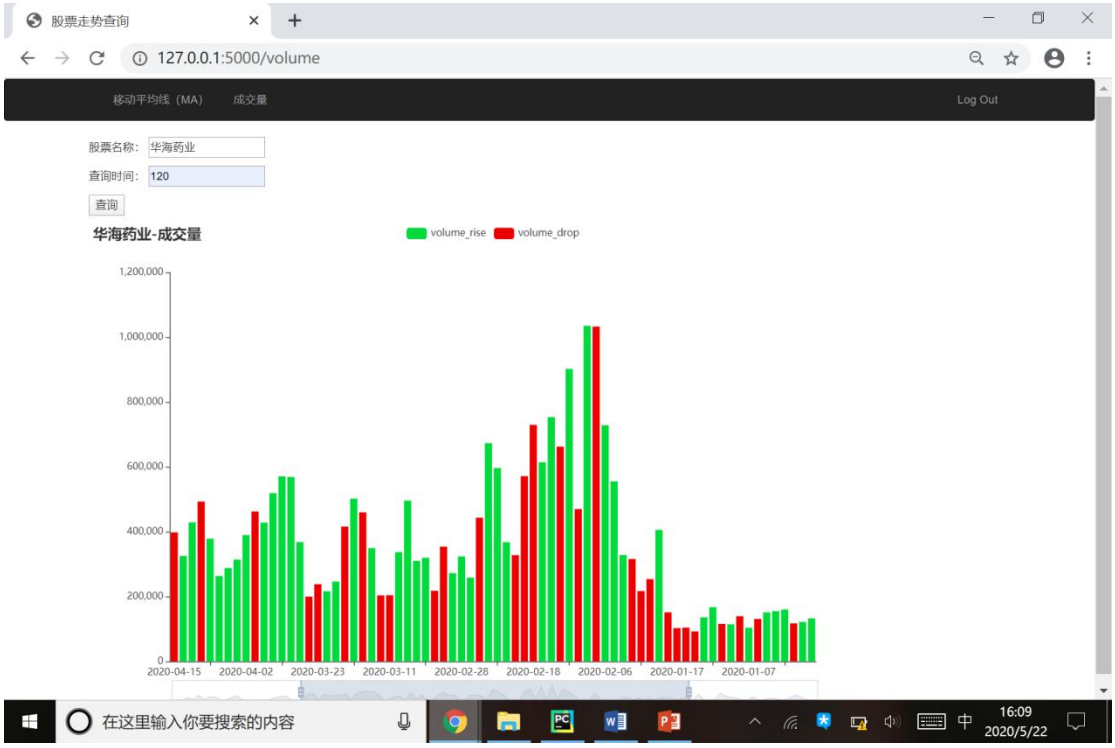


Figure 3.22 trading volume of Huahai pharmaceutical

Huahai Pharmaceutical Co., Ltd. imitated the new crown specific drug "redcivir" and put it into clinical trial, and its stock rose sharply.

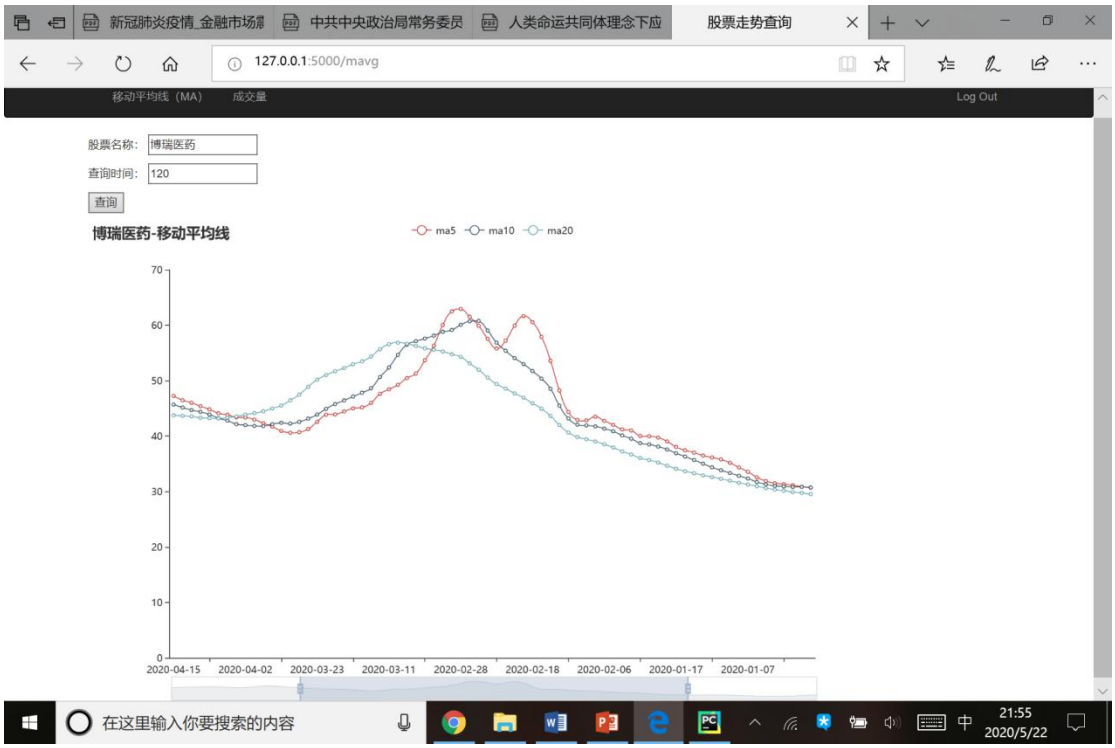


Figure 3.23 MA of Borui pharmaceutical

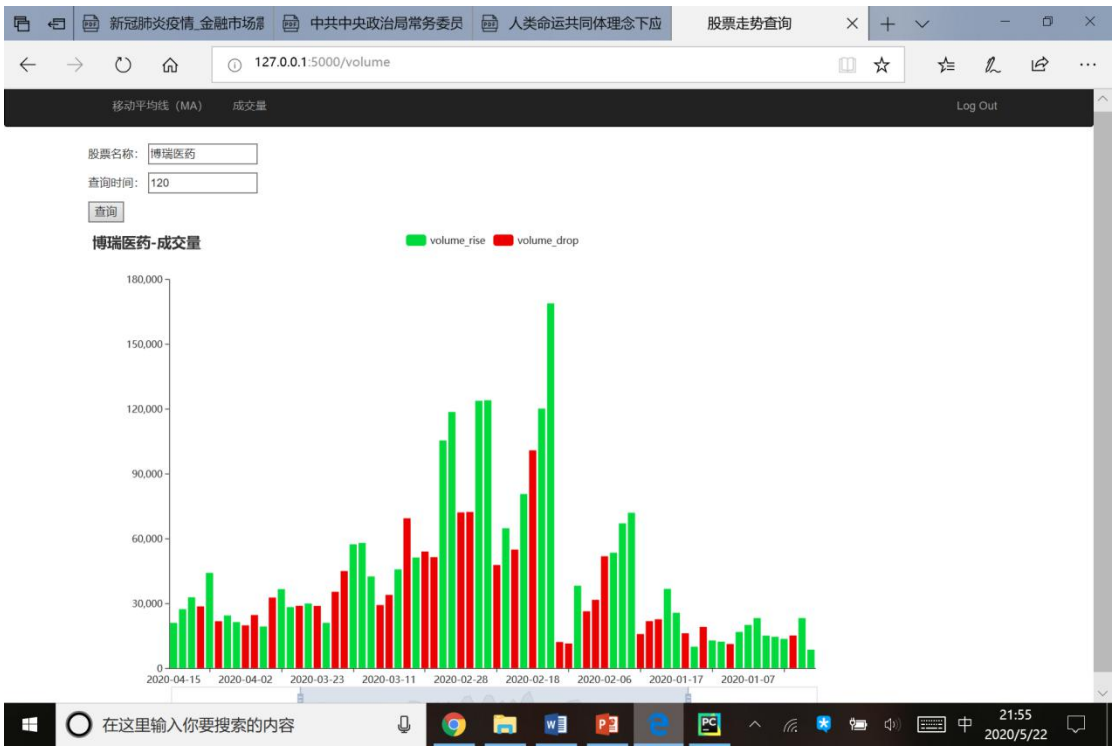


Figure 3.24 trading volume of Borui pharmaceutical

On February 11, Bray Pharmaceutical Co., Ltd. announced that it had successfully imitated redcivir, and its stock rose sharply. On February 14, the Shanghai Stock Exchange announced that it would focus on monitoring the concept stocks related to epidemic prevention and control. After that, Borui pharmaceutical entered a shock correction, but its cumulative increase was still more than 40%.

According to the above analysis, the biomedical industry was originally a weak cycle industry with long-term and stable growth. During the epidemic period, due to the demand relationship, it was chased by capital, and was also valued by the public. Some financial products began to launch "robust medical portfolio series", and investment in biomedical industry became a new trend of investment and financing.

During the epidemic period, from the rise of epidemic related stocks, especially domestic vaccine related stocks, we can see that the Chinese people are full of confidence in overcoming the epidemic situation.

# 4 Summary and Prospect

The COVID-19 epidemic will inevitably have a great impact on the economy and society. There are inherent contradictions between epidemic prevention and control and economic development. The Chinese government has always insisted on taking the people as the center and immediately implemented strict isolation and closure measures at the initial stage of the outbreak. When the spread of the epidemic has gradually been controlled, it has continued to strictly control the epidemic situation by adopting measures of classification and precise implementation. At the same time, various departments and governments at all levels have introduced a series of policies and measures to promote the economic order to return to normal, providing a set of Chinese experience and scheme for the countries affected by the epidemic.

The main work of this paper is summarized as follows:

(1) In the outbreak of COVID-19, in this graduation project, according to the epidemic situation data and the public opinion data and economic data related to the epidemic situation, according to the research background and significance, related theories and technology, feasibility analysis, demand analysis, design and implementation steps, we used the empirical research method, quantitative analysis method, case study method and interdisciplinary analysis method. A variety of methods, closely linked to current events, combined with time development, spatial transformation, policy measures, from the visual analysis, we can see that the epidemic situation is gradually controlled, public opinion is gradually turning to positive, and the economic situation is gradually becoming stable. Compared with the previous epidemic related papers, this paper's analysis is broader, the time span is larger, and more truly reflects the impact of the epidemic.

(2) In the process of data collection, this paper uses web crawler technology to crawl and collect various epidemic related data, such as epidemic data, Baidu hot search, Sina Weibo hot search and comments, video comment on-screen, CCTV News data, online shopping data, stock data, etc.

(3) In this paper, the crawled data is visually analyzed, and some classic visual libraries of Python are used to analyze and process the data, which makes the results more intuitive.

(4) This paper analyzes the text sentiment of the crawled public opinion data, and adopts the sentiment analysis method based on the dictionary. After constructing the sentiment dictionary, negative word dictionary, degree adverb dictionary and stop word dictionary, the Hidden Markov Model (HMM) is used for word segmentation, TF-IDF Algorithm and TextRank Algorithm are used for feature word extraction, and the emotional tendency of text is calculated according to the dictionary in a user-defined way.

(5) This paper uses Python Flask, combined with web crawler technology Flask Sqlalchemy database function to design and develop a small stock query website, which is convenient for stock data query according to time and stock name.

After being controlled in China, COVID-19 broke out again in Beijing, Jilin Province and Xinjiang Province, and was controlled again. Other countries can

control the epidemic situation only after adopting China's epidemic prevention experience. Up to now, China's measures are the only effective way to control the epidemic. In some countries, such as the United States and India, leaders pay little attention to the life and health of the people, resulting in the increasing number of COVID-19 patients and deaths. Although the appearance of COVID-19 is a natural disaster, the consequences of COVID-19 are actually man-made disasters.