

Project Write Up

In this project, I want to first find out the current situation of COVID in the U.S. Then, I look up the data of the percentage of the frequency of mask used by county to answer what the real situation of people wearing masks in real life is. Next, through the COVID vaccine tracker in the U.S., I can know the real situation of COVID vaccination in the U.S. and see how far for the U.S. to reach a rate of COVID vaccination that people expect. Moreover, I will check the top articles on the homepage of New York Times website and look up how many of them are related to COVID; go through the most viewed articles in 7 days on New York Times website and find out how many of them are related to COVID; and dig through the most shared articles in 7 days on New York Times website and figure out how many of them are related to COVID. Through all these processes, we should be able to have a brief understanding of the situation of COVID in the U.S. and people's attitude toward COVID in the U.S.

In the first section, I import the data of cumulative cases and deaths of COVID in the U.S. I sort the data on Cases descending first then Date ascending to find out the total number of cases and deaths in the very recent dates. It shows that on 12.08.2021, the number of cases is 49505304 and the number of deaths is 791933, which are very sad to see. Then, by using the query method, I find out the date when the cases of COVID exceed 10000000 is 11.08.2020, and the date when the deaths of COVID exceed 100000 is 05.27.2020. That tells us the cases and deaths of COVID both reached to a sorrow number in less than a year when COVID enters the U.S. In the end, I plot two line-charts to look up the trend of the cases increased by date and the trend of the deaths increased by date. These two plots indicates that although there are some days when the number of new cases and deaths dropped, in general the number of new cases and deaths increase almost every day. The data indicates the situation is urgent and dangerous to everyone in the U.S.

In the second section, I import the data of the percentage of the frequency of mask used by county. By using the query method, I find out the number of counties where the percentage of people always wearing mask is larger than 0.5, which is 1544 and about half of the number of the counties. And I am shocked by the number of counties where the percentage of people always wearing mask is less than and equal to 0.3, which is 248. In addition, I calculate the average percentage of each frequency of people wearing masks and a kind of relieve that at least the average of people always wearing mask is 0.5. It is still not a lot but better than I expected before I worked on this data. Finally, I included a bar chart to show the average percentage of each frequency of people wearing masks. It visually tells us that the average percentage of people always wearing masks is significantly higher than other average percentage of frequencies. This data shows that a noticeable number of people in the U.S. does care about themselves and other who around them. However, it could be better.

In the third section, I imported the data of COVID vaccine tracker in the U.S. I first use .groupby to obtain the completed vaccination for each state. Then, I need the row labels to be datapoints so that I am able to calculate the percentage of completed vaccination for each state. Eventually, I calculate the percentage of completed vaccination for each state. The reason why I am going through all these steps is to find out the real situation of COVID completed vaccination for each state. I calculate the average percentage of completed vaccination for each state, which is about 0.6. This percentage of completed vaccination is good considering that the vaccine in fact came out less than a year. It shows that the majority of people in the U.S. has confidence to the vaccine. It also implies that the majority of people in the U.S. does strive to fight COVID. I include a bar chart at the end to just visually show the percentage of completed vaccination for each state.

In the fourth section, I use New York Times API to check the top articles on the homepage of New York Times. I use `request.get` to a dictionary that contains the information of the top stories on the homepage of New York Times. I use two for loops to get the information of articles about COVID and not about COVID. Then, I plot a bar chart to compare the number of articles about COVID and not about COVID. When we first look at the bar chart, we might think that the number of articles about COVID on the homepage is way less than that of other articles on the homepage. However, as we are reading through the description keywords of other articles on the homepage, we should realize that the topics varied. So, even though we cannot be sure about people's interest on the topic of COVID from this data. We could at least have a guess that the website is posting enough number of articles about COVID on its homepage. And it might be due to two reasons. First, the media encourages people to gather more information about COVID. Second, people give these articles enough attention and push these articles on Homepage.

In the fifth section, I again use New York Times API to check the most viewed articles for 7 days on New York Times. I use `request.get` to a dictionary that contains the information of the most viewed articles in 7 days of New York Times. I use two for loops to get the information of articles about COVID and not about COVID. Then, I plot a bar chart to compare the number of articles about COVID and not about COVID. As we seen, the number of most viewed articles about COVID is more than half of the number of other most viewed articles. And by reading through the description keywords of other most viewed articles, we notice that the topics varied. Therefore, we conclude that among the most viewed articles, the most common topic of the articles is COVID.

In the sixth, I use New York Times API to check the most shared articles for 7 days on New York Times. I use `request.get` to a dictionary that contains the information of the most

shared articles in 7 days of New York Times. I use two for loops to get the information of articles about COVID and not about COVID. Then, I plot a bar chart to compare the number of articles about COVID and not about COVID. As we seen, the number of most shared articles about COVID is only a little bit less than the number of other most shared articles. And by reading through the description keywords of other most shared articles, we notice that the topics varied. Therefore, we conclude that among the most shared articles, the most common topic of the articles is COVID. Moreover, people are more willing to share information about COVID to people around them.

Through the project, no wonder people will pay a significant amount of attention to the information about COVID. And it might be time to act more aggressively to stop this tragedy.