Ziad Sakr

COVID-19

Math Modelling Final project

**COVID-19**

1. **Description of the Problem:**

COVID-19 is a short for corona virus disease 2019, which is the official name that is given by the World Health Organization to the coronavirus disease. The number of the people who get infected by the COVID-19 are changing rapidly which has an effect on most if not all the countries: economically, politically, and financially. It’s been declared as a pandemic by the WHO (World Health Organization) and that because of its rapidly spread. The virus is targeting people with weaker immune system and more especially people older than 65. “According to a report published in the CDC's Morbidity and Mortality Weekly Report (MMWR) in late March, nearly 40% of people hospitalized for COVID-19 between mid-February and mid-March were between the ages of 20 and 54. Drilling further down by age, MMWR reported that 20% of hospitalized patients and 12% of COVID-19 patients in ICUs were between the ages of 20 and 44” 1.

The goal in our research: Create a model that will be automatically be updated by connecting our data to a reference of WHO data using web scrapping in python. Web Scrapping is a tool in python that the purpose of it is to extract a specific data from a website and update your model automatically. That will allow us to predict what it will be in the future from our data. I will be tackling that by referring the targeted website using pandas to read the targeted csv file in this website. I will be focusing on the United States of America. Because it’s one of the highest COVID-19 cases in the world and it changes rapidly every day. We will be doing analysis on couple states; NY, CT, UT, TX but specifically we will be focusing on NY.

* **Questions to be answered for our model:**

How serious is COVID-19?

* The answer here could really depend on what we are looking for. in our case, we will focus on the total number of cases from modelling data by focusing on the cases curve from the beginning of the pandemic till today.

How lockdown effect on reducing the spread of the virus?

* The lockdown is one of the social districting or isolation restriction. Meaning that a person shall not leave their space for a specific period of time and the study shows that this period is 14 days. It is also very important to prevent the COVID-19 pandemic and over the past couple month it showed that the spread of the virus can be reduced by this strategy.

Is the COVID-19 Test helps in reducing cases?

* We will answer this question in the end of our analysis by taking NY state as an example. We will see that COVID test helps in reducing the positive cases. As the higher the rate of the COVID test the faster you will be able to maintain its speediness either by the lockdown as we talked about it earlier or by Isolation or by going to the hospital and take the prescribed medicines for it.

**2- Assumptions and Model formulation:**

1. Covid-19 total test result and Lockdown:

The focus of this analysis is that most of the Countries announced that it will go for a full lockdown while others only have a curfew for specific hours. In our model we will doing our analysis on the US and we will be looking for each state individually but focusing on NY state specifically.

* How these states are dealing with COVID-19 differently than the others. In our model we will be looking at how these different countries/states are dealing with the virus and the result of the virus behavior by looking at the daily test results.
  + I will take New York, USA as a state to focus on and compare it with other states for multiple reasons: first, NY was the highest state that recorded COVID-19 cases between March to May. However, increasing the testing for the virus allowed the state to know how many cases and who is having it and from here they will be able to control these cases from spreading by putting them under quarantine. We will look at this in our analysis by comparing the daily test result vs daily positive cases in NY with other states as well.
* How long this country been lockdown and how large is the COVID test results in this country?

By knowing that, we will get a clear mind of the risk that this country is facing from this pandemic and how long until it can control it. As according to the WHO

* “Any person who is in close contact (within 2 meters) with a COVID-19 case experiencing respiratory symptoms, even if mild, (e.g., sneezing, coughing) is at risk of being exposed to potentially infective respiratory droplets”2
* “AGMP have been associated with an increased risk of transmission of coronavirus by inhalation.”2
* The reason of this analysis is to prove that testing and quarantine could be the key for reducing the daily cases which what we will see in our analysis models.

1. **Analysis:**

In the analysis part I will focus on doing three different kind of analysis. That will help us to look at the data from different perspectives and will help us to predict the future of the pandemic in the US. First analysis is by doing the best line fit using the concepts of Polyfit that we covered in class on the new COVID-19 cases daily. This will help us to see the coefficients matrix of the coefficients p. As Polyfit issues a Rank Warning when the least squares fit is badly conditioned which implies that the best fit is not well defined due to numerical errors. I am improving here my fit by having three different degrees in a list and run a for loop on this list and for there we will be able to see the best fit. This will avoid us the issue of fitting the polynomial coefficients that could led to badly conditions when the degree of the polynomial is large, or the interval of the dataset is badly centered. That could lead to overfitting. Therefore, the quality of the fit should always be checked in these cases. In this case I tried different polynomials degrees, but the best line fit is the exponential function. The reason I choose exponential function for this curve fitting as we can see there is a peak that happened in the daily positive cases and then because of an increase in our testing the positive cases has reduced. Which is will being our claim that we will be using for our analysis for our modelling prediction.

Our chart here is needed for our end analysis. We are aiming for the trend of comparison between the daily positive cases vs the daily test results. In the curve fitting below our **x axis** **is daily test results** and **y axis** **is positive daily cases**. We want to see if having many tests per day will result in reducing and controlling the positive cases. Our fit here is telling is showing us this trend and I will explain how. Our function is a exponential function. By looking at our function we assume here that with having more tests will result in future reducing of positive cases. We are hoping that our curve will stay steady and/or go down in the future. This why I think that having a exponential function will be a great use in our analysis. We will talk closely about these numbers in the end of our analysis. This curve fitting, we will be used as a support for a start of the analysis process. **All of our analysis are done in NY state and focusing on the daily cases.**

Chart

Description automatically generated

Secondly, I will continue my analysis by showing the death peak happened on the new cases using the bar graph. The reason I choose to do that is to answer our question that we had in the beginning regarding the seriousness of the virus. Also, to support my point on increasing testing and quarantine will help in reducing cases and/ death.

Chart, histogram

Description automatically generated

The data below help us to see how the daily positive cases vs negative daily cases. are doing over the time. We can see that our positive cases have passed the peak and it is going in a fast reduce and then on a steady number. However, for the negative cases we can see the increase in the negative cases, which is basically supporting our claim here. From here, we will continue on showing analysis on the test results vs the positive cases (daily) to be able to predict if the peak of this pandemic happened or there will be another one by doing my analysis it will show us if it happened already or not.

Chart, histogram

Description automatically generated

Thirdly, for the last analysis I am trying to model the total cases vs new tests. As a way to show if the new tests help to reduce the total cases. I am supporting my analysis by creating two vertical lines in my data. The first line will determine the the start of the restriction laws that the USA made and that will help us to see the impact of this decision on the growth of the total cases. The second vertical line is to determine the affection of this restrictions on the new cases after 14 days of the isolation. I am starting this analysis by creating two variables: tcases which will hold the total cases and test which will hold the new tests. Then I am creating two lists y and test growth which will be holding the data for the total current cases / new tests and test growth to hold the total new tests.

Creating a variable x to hold the total days since the start of the pandemic in the US which is detected on the 3/7/2020. Our focus will be on a specific day and on specific cases that will help us to look at the data more efficiently.

A screenshot of a cell phone

Description automatically generated

As we can see that the isolation didn’t make as of a big impact on reducing the total cases of COVID-19 in the US as the decision came late. The decision came after there was already cases in the USA. Our model show that the virus spread really fast and my prediction that if this decision came early, we could be able to manage It more. As from the decision day on the 17th till 14 days after the shutdown decision. There is more than 30% daily growth. My claim here is that this daily growth rate could be very overwhelmed long before we reached the summer. As there are no reports stated that the virus will slow down in the summer. However, our data and my prediction doesn’t expect around a 30% daily growth rate to continue, but we will see another peak happening if the government reopen the social events, travels and the economy.

In my final phases analysis, I wanted to show the total result of the tests in each state. I am tackling that by creating a function that has a list as its parameter and store this list in an array. I run the array inside a loop and doing some arithmetic on it using python built in function so it can be able to give the desire plot. The comparison happened to compare daily test results at NY state with other states as well that been hit with the virus. The reason I am showing that is to support my point on increasing the tests/day will result in reducing and controlling the positive cases and increasing the negative cases. I choose NY here as our main factor because it is the most state that been hit hardly with the virus. I compare it with other states as well that been hit similarly or less than NY. We will not get deep here in the analysis of the other states, but the purpose of this plot is to show the total tests/day each state has taken to reduce the cases. NY is in number one here in this list.

Chart, line chart

Description automatically generated

From my point of view the rise of the daily total test result shown that the states that most effected by the virus has the highest daily tests. Which in result we will see in the next graph that is facing a low cases per day. Recall our first curve fitting: we can see that our x-axis of total test daily results and y-axis of positive daily cases has passed the peak of risen in positive cases. Our exponential function here is supporting our claim on if the NY state continued on increasing the test result per day as they are doing so far that will keep the crurve steady and not increase. Which by that they will be able to put individuals under quarantine or isolation if needed to prevent the virus from spreading. **(refer to second page on information about virus spreading).** However, if the tests/day has reduced this could put the state under dangerous of a second peak.

Chart

Description automatically generated

This analysis is based on the data provided and not having in consideration natural causes e.g sickness, weather change, finding vaccines, etc...

Chart, scatter chart

Description automatically generated

1. **Discussion and Conclusions:**

In conclusion, recall our analysis from the daily total positive case per day and compare this to the analysis of case before and after quarantine we can see that cases after quarantine has been decreasing down and that the peak already happened around March 20th- May 07th. The peak happened couple of factors; lack of knowledge about the virus, not doing the right tests to be able to recognize who has COVID-19, not maintain quarantine or social distancing. On the other hand, we can see reducing in the peak and the decreasing in the daily positive cases which happened because of many factors; Increasing the daily tests per state which will help to Capturing the virus in the early stage will help in reducing its speediness, social distancing (6 feet min) between two persons, if person feeling symptoms being in two-week self-quarantine will low the effectiveness of the virus in the body and in the community. From our analysis in Total test result for the different states, we can see that NY which has the highest Total tests per day has path through the peak and the cases are going down in our last cases and this is based on our exponential function curve fitting. We proved our main point in the assumption section that testing and quarantine are be the key for reducing the daily cases.

1. **Future improvements:**

Time sometimes is the key for improvements and accuracy. My analysis was effective in supporting my assumptions and answering the question however it is not the most efficient way but because lack of time and the minimal data resources that we have on our topic. COVID-19 is a very sensitive topic; politically, economically, and personally. We only see the numbers that the government wants us to see. I will not say this is case for all the countries however in some countries they got exposed that their numbers were not true. This why we don’t get the right data because of political issues which will affect the economy and individuals as well. The reason I stopped where I am now in my analysis because of the deadline for this project and again the limited access to the External resources regarding COVID-19. I believe that I could do better analysis if I had true and better data resources. I claim here that If I got more time, I would improve my analysis using Machine Learning. However, Machine Learning is a weapon with two heads. It could be very efficient, or it could destroy your analysis if you tackled it the wrong way. In our case we will assume the best case. Developing a Machine Learning model to predict our future data cases based on our assumptions in this project will be very useful by allowing the model to get trained on the past trends that we have in our data and be able to predict the future behavior by connecting the dots of the past data.

References:

1Harvard Medical school Health Publishing.

<https://www.health.harvard.edu/diseases-and-conditions/covid-19-basics>

2Governmennt of Canada

<https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/assumptions.html>

3WHO. Q&A on coronaviruses. [Online] 9 March 2020 [Accessed at:

<https://www.who.int/news-room/q-a-detail/q-a-coronaviruses>].

4Zhang, W., Du, R. Li, B. et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. Journal of Emerging Microbes & Infections. [Accessed at: <https://www.tandfonline.com/doi/full/10.1080/22221751.2020.1729071>]

US COVID-19 Data Citation:

<https://covid.ourworldindata.org/data/owid-covid-data.csv>

Data Citation:

<http://covidtracking.com/api/states/daily.csv> (Getting first data)

<https://covid.ourworldindata.org/data/owid-covid-data.csv> (Getting second dataset for lockdown analysis)

<https://github.com/altair-viz/altair> (Documnetation of altair library)