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 espefcally 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.

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

What country we will focus on?

* We 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.

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 specifc 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-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 spreadness 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:

Most of the Countries announced that it will go for a full lockdown while others only for specific hours. In our model we will be looking for:

* 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
* “A person who is in direct physical contact (e.g., kissing, touching) with a person with COVID-19 is at risk of infection, via the transfer of the virus.”2.

1. Transmission:

“Symptomatic cases of COVID-19 are causing the majority of transmission;” 3 “however many people with covid-19 have only mild symptoms, especially at the early stage of the disease, and can still transmit to other people.”2

* Person to person transmission is mostly occurring via infectious
* “respiratory droplets Respiratory droplets and contact transmission are considered to be the most important routes of transmission of COVID-19 viruses, but do not fully account for the occurrence of all COVID-19 cases, previously known as novel coronavirus pneumonia (NCP), and the reasons for the rapid spread of this virus”4

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 Quartic. Because of it is function ax^{5}+bx^{4}+cx^{3}+dx^{2}+ex+f=0 the Quintic fit line will result in a huge increase in the data set and then it will eventually go down. Which is will being our prediction when modelling this pandemic.

A close up of a map

Description automatically generated

Secondly, I will continue my analysis on the new cases using the bar grapgh. The reason I choose to do that is to help us to see how the total cases are doing over the time and be able to predict when the peak of this pandemic happened/will happen and by doing my analysis it will show us if it happened already or not

A close up of a piece of paper

Description automatically generated

The data as shown is not normally disturbed and we can see that the curve of the new cases will go up. However, the first peak happened as we see but from our data and from my line fit function prediction there will be another second peak will happen.

A picture containing computer

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 pithon built in function so it can be able to give the desire plot below:

A close up of a map

Description automatically generated

From my point of view the rise of the 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

A close up of a map

Description automatically generated

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)