Rochester Institute of Technology

RIT Digital Institutional Repository

5-2022

Covid-19 Vaccine Hesitancy and Governments Challenges

Laila Abdin
laa4801@rit.edu

Follow this and additional works at: https://repository.rit.edu/theses

Covid-19 Vaccine Hesitancy and Governments Challenges

by

Laila Abdin

A Capstone Submitted in Partial Fulfilment of the Requirements for

the Degree of Master of Science in Professional Studies:

Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology
May 2022

RIT

Master of Science in Professional Studies:
Data Analytics

Graduate Capstone Approval

Student Name**:** Laila Abdin

Graduate Capstone Title**:**

Covid-19 Vaccine Hesitancy and Governments Challenges

Graduate Capstone Committee:

Name: Dr. Sanjay Modak   Date:
              Chair of committee

Name: Dr.Ioannis Karamitsos   Date:
              Member of committee

Acknowledgments

Abstract

Governments face the challenge of increasing the number of people who have been infected. Millions of daily examinations are being conducted around the world to detect the disease. The pandemic is growing increasingly challenging as the number of confirmed cases increases. Governments encounter a variety of obstacles in approving the benefits of Covid-19 vaccinations. Indeed, the idea of getting vaccinated is still fuzzy, and does vaccination protect against infectious disease? There are a lot of questions floating around. The lack of clear information about the Covid-19 vaccine will stymie the growing number of people who have been immunized. Therefore, governments are being pushed to demonstrate the consequences of delaying or rejecting the Covid-19 vaccine. In this study, I examined the number of people tested, infected, and vaccinated and found out what are the factors that affected taking the vaccine specially in the United Arab Emirates. The reasons why we need to estimate the number of people who have been vaccinated. How to increase the number of people vaccinated. Also identified who the 20 countries with highest number of cases are, what motivating factors have influenced the increase in the number of people vaccinated, and what will happen after the pandemic. Increased testing of the Covid-19 virus will lead to the discovery of more infected people and encourage people to get vaccinated because infected people are suffering. More vaccination will be given to test the hypothesis that more Covid-19 testing will occur. For this study, I used a dataset collected from health ministries around the world on the number of cases, deaths, age, male/female, some diseases, and whether the infected person is a smoker or not. The world's lockdown variables, population demographics, mortality rates, infection rates, and health were collected, focusing on the 20 countries with the highest numbers of infected persons, deaths , and vaccinated persons.

Infection transmission and mortality data were presented in a linear stepwise fashion using ridge and lasso regression. Factor analysis was used to investigate the overall structure of the Covid-19 data, resulting in a theoretical model that was tested using latent variable path analysis. The mathematical model can help determine when these intervention methods are most effective for disease control and how they may impact disease dynamics. As a result, in this project, I developed a mathematical model that predicts the relationship between new cases of Covid-19 and those who have been fully vaccinated. I began my research on the pandemic environment by looking at a map of cases around the world, then looking at the top 20 countries by total number of cases and other factors, and finally doing some studies on the United Arab Emirates. During the last two years, I discovered that there have been four waves of the epidemic.
It was important to conduct regular testing for citizens so that sick people could be easily identified and helped. I also discovered that some countries had more cases than others, which would allow other researchers to investigate the causes behind this and take preventative steps in similar occurrences in the future. I used linear regression to predict the new number of people who have been vaccinated, which will help stakeholders make decisions about reopening, especially in areas with large populations, such as schools, universities, ministries, and other areas.

Keywords: Covid-19, Covid-19 vaccine, Active cases, Death cases, Pandemic

Table of Contents

List of Figures:

List of Tables

Chapter 1

1.1 Introduction

From the end of 2019 to the present, we have all been dealing with the problem of the massive spread of Covid-19. The entire world has begun to look for a vaccination to protect people, because of the large number of individuals who have died, prompting pharmaceutical corporations to begin looking for a vaccine and subsequently medical therapy.

Governments began to get a variety of responses; some accepted the vaccination and registered to receive it, while others refused. Governments and stakeholders started to look for new approaches to raise the number of people who have been vaccinated and to eliminate any factors that create hostility to the vaccine.

It is critical to know who had the vaccine and who did not, as this information will aid in the opening of various vitality departments such as schools, construction projects, industries, airports, and other projects. Because the isolation of countries has had a negative impact on many industries, vaccination is becoming increasingly crucial.

Counting the number of people who have been vaccinated and encouraging others to do so becomes a global aim. As a result, I concentrated my research on boosting the number of people who have been vaccinated. The increase in the number of Covid-19 tests and the number of fatality cases can be used to predict vaccine willingness. Then, using data mining methods, continue to predict the circumstance after taking the vaccination after the project is completed.

1.2   Project goals

I recognized the main motivations and problems for using the Covid-19 vaccination in this study. The reasons why we need to estimate the number of persons who have been vaccinated. How did they increase the number of persons that were vaccinated? Apart from identifying who the targeted people in society are, what were the incentive elements that influenced the increase in the number of people who were vaccinated, and what will happen after the pandemic.

1.3   Aims and Objectives

My main goal in this research is to figure out world data to study the similarities and variations in vaccination use, active cases, and mortality rates throughout the world, with a focus on the top 20 nations with high Covid-19 cases, and then particularly the UAE.

The main goal is to get a complete image in order to reopen and resume regular living. Furthermore, worldwide organizations may have thoughts regarding nations and regions where greater vaccination knowledge is needed, as well as what is happening in the world around us and how to reach out to them in comparable situations.

1.4   Research Questions:

R1: How many people got the vaccine.

R2: What are the methods that helped governments to increase the rate of vaccinated people

R3: What help the stakeholders to take the right decision for the reopening.

1.5   Research Methodology

In my project, I used the CRISP DM model prosses to illustrate the framework over six stages. CRISP DM stands for Cross Industry Standard Prosses for Data Mining. The below phases of CRISP DM are the lifecycle for data mining, analytics, and machine learning prosses.

This diagram shows the integrated steps for Crisp DM methodology.

Figure 1: CRISP DM Methodology

1. **Business Understanding**: The Business Understanding is the first step in the process. I describe the problem I want to solve or which questions I want to answer at that phase, which was what factors affected completely vaccinated people, I mean was the testing, increasing number of infected people or the dearth cases. At the end of the project, I was able to predict the number of vaccinated people by linear regression method.

2. **Data Understanding**: The second phase is to understand the data that will be used to help solve the business challenge. My data was useful to reach my goal, however it will be better for digging deeper on the problem if there are more data about the number of people on the ICUs, if they are male or female or their ages. I had a look on the attributes and try to find relations among them and how to reach my aim which is how to increase the number of vaccinated people.

3. **Data Preparation**: In this stage I started to clean my data that to raise the quality of data analysis such it impacts directly on the results. I used heat map visualization method to understand the relation among the attributes then accordingly, I deleted many of attributes. Then I study the Null values and deleted the empty attributes, I used method which is called LOCF for the missing values. I also studied the outliers' values for more insights about my data.

4. **Modelling**: I tried different ML models the I considered that the linear regression is the best fitting actual and prediction.

5. **Evaluation:** Previous assessment phases looked at things like the model's accuracy and generality. During this stage, I examined how well the model fulfils my objectives and try to figure out whether there's a business reason for its shortcomings. If time and resources allow, another alternative is to test the model on test applications in the real application. I analysed the data mining findings I've created throughout the assessment process. My model that is required to be connected to the original business objectives are included in data mining results, as are all other discoveries that not need to be related to the original business objectives but may reveal new issues, information, or clues for future.

6. **Deployment:** At the end of the project, my model can predict the vaccinated number of people for the future. This is very important for the reopening in critical places like schools and specially the primary schools.

1.6 Limitations of the Study

The attributes in the dataset were the cause of the limitations of my study. For an extensive analysis is required more information about the location of active and death cases in the UAE, as well as more information about the age of those infected and whether they worked online and when they moved to their offices. In addition it was a race against time.

Chapter 2 – Literature Review

2.1   Introduction

The title of our life during the Covid-19 pandemic was "No one is safe unless everyone is safe." The Covid-19 vaccination became in high demand. The restrictions and way of life have been affected. This pandemic had an impact on all aspect of our lives. For the general population, new products and behaviors become commonplace. Many studies and research into all the changes caused by the pandemic have begun, some of which are related to the vaccination and others to related habits such as shopping, studying online, and others. Governments around the world began researching vaccination apprehension in different people based on their sex, age, level of education, health status, religious commitments, and financial level. Governments devised strategies for dealing with various obstacles that may arise because of vaccination. Many groups sought to stop the spread of negative information regarding the vaccination.

2.2 Covid-19 Challenges

By the end of the year 2019, the Covid-19 virus considered as a pandemic by the WHO (World Health Organisation), the head of the Global Alliance for Vaccines and Immunization (GAVI) has warned, "Nobody is safe unless everybody is safe," from the new coronavirus (Enitan et al. 2020, p 1-2). The disease started to spread from one to another rapidly. A huge number of infected and dead people over the world. The expected loss for those two years about 9 trillion American Dollar (Murphy et al., 2020). Demand for the Covid-19 vaccine becomes a must. The World Bank and the Coalition for Epidemic Preparedness Innovations (CEPI), which funds the development of epidemic vaccines, co-hosted a global consultation that led to the launch of a Covid-19 Vaccine Development Taskforce, with the mission to finance and manufacture Covid-19 vaccines for global access (Naiankara et al., 2020). The vaccine is "Vaccines are biologics that provide active

adaptive immunity against specific diseases" (Kashte et al. 2021, p.2). Scientifically, producing antibodies to kill specific microorganisms is the simplest definition of the vaccine in general.

At the beginning of the pandemic, a lot of restrictions were set, and several procedures were taken to reduce the spread of the disease. Countries locked down completely or partially, closed their schools and universities, airports, many governments, and private establishments. People started working online. Lifestyle has been changed with this pandemic, wearing face masks, having social distances between people, alcohol sterilization, and decreased the families' activities. In this situation some countries sought to achieve herd immunity, however, most countries approached lockdown waiting for the vaccine to be available (Harper et al. ,2020). Locking down the countries affected all domains, education, health, construction, and traveling. Accordingly, working on finding the vaccine and competition among the companies started to find the solution.

Policymakers, focus on the statistics models about the situation of Covid-19 to take their decisions (Lee et al. 2021). For example, studied the impact of social distance and wearing face masks to reduce the spread of the virus (Lee et al., 2021). Accordingly, most of the countries launched Apps like AlHoson App in the UAE, to show the case of the person if he is infected (active case), not infected, or contacted with sick people. Physical isolation was the first reaction taken for the infected people (active cases) (Oliver et al., 2020). That later followed by wearing an electronic charm to alert indicated health organization if the person left his isolated place. This data about moving the infected people from one place to another which are taken from their mobiles will be valued for decision-makers to analyze the situation and study if the decisions were taken were correct or not (Oliver et al., 2020).

In the case of the Covid-19 vaccine, many candidates to approve the vaccine were produced then developed to be available for public use. In addition to the researchers who are finding the vaccine, there are other researchers started to study the dramatic of daily cases, dead cases, the behaviour of this disease, willingness to tacking the vaccine,

vaccinated people, and the impact of this vaccine, despite the lack of information about the disease. For example, "researcher has developed a mathematical model to identify the number of days students could attend school to allow them a better learning experience while mitigating infections of Covid-19" (Mohammad et al. 2021). One study was conducted about the acceptance of the Covid-19 vaccine showed that the highest acceptance rate was in Ecuador and the least was in Kuwait (Sallam, 2021). However, in the United States by a study on acceptance, the Covid-19 revealed that 20% intended to delay the Covid-19 vaccine due to safety and novelty (Naiankara et al. 2020).

Similarly, the survey was conducted in Saudi Arabia, their finding revealed 64.72% of Saudis show interest in accepting the vaccine from different society categories based on age, married status, Saudi or not, and degree level (Naiankara et al. 2020). Besides, in the same previous study, people who will be forced to take the vaccine by their work was high percentage also was observed in this study older people accepting the vaccine more than the younger (Lazarus et al. 2020).

In Indonesia, old people rejected the idea of the Covid-19 vaccine because they set most of their time in their houses and not traveling like younger people, so they are less perceived at risk of being infected with the Covid-19 virus (Harapan et al. 2020). Black and Hispanic Americans who are lower academic achievers also rejected the vaccine (Whitehead and Perry,2020). Religious people in America refused the vaccine because they distrust science or scientists, also linked to libertarian populist attitudes that incline Americans to ignore the recommendations of scientists and the mainstream media regarding infectious diseases (Whitehead and Perry,2020).

There are different factors such as the cognitive, psychologic, socio-demographic, and cultural effects of the hesitancy of taking the vaccine (Sallam, 2021). An earlier study showed that the high-income people refused the safety of the Covid-19 vaccine comparing with the low income over different countries (Sallam, 2021). Based on a study, women and old people are more willing than others of the sample of that study (Lazarus et al., 2020). Hesitancy in taking the Covid-19 vaccine or refuse it is the situation for all

nations. Hence, the government needs to find the factors that affect the acceptance of the Covid-19 vaccine. Different studies were conducted over the world to know the essential acceptance of the Covid-19 vaccine. The hesitancy of tacking the vaccine stands as an obstacle for the governments. Many reasons are behind refusing the vaccine, such as the education level, gender, age, lack of knowledge and religious believes and fears from the disease also will affect negatively on spreading the virus and accepting the vaccine (Harper et al.,2020). Hence, countries supposed to find the factors which are impacted directly for wide acceptancy for the Covid-19 vaccine. Factors are limited by complacency, convenience, and confidence. Governments are likely to optimize vaccine uptake in different communities by (i) establishing trust, (ii) providing both the risks and benefits of vaccination, (iii) giving the facts before addressing the myths, (iv) using visual aids, and finally (v) testing the communication material before launching (Naiankara et al.,2020). Moreover, the genetic mutation on the Covid-19 virus was a big reason to reduce the acceptance of the vaccine (Sallam, 2021). People must know the benefits of the vaccine for them and who are most closely identify (Murphy et al., 2020).

Consequently, having the vaccine, then count the vaccinated people will help predictors to predict how long the institutions are opening back and return to their life (for future work on my project). That all to reduce the material and immaterial losses. Specific moral foundations plus the commitment to the government restrictions and rules will reduce the acceleration of spreading the disease and finish this endemic faster (Harper et al.,2020).

It is expected that all governments will provide the vaccine for all equally and fairly and concentrated on the targeted people for this vaccine who are the old people, who work as healthcare workers and who have some diseases (Enitan et al., 2020), which will increase the acceptance of this vaccine to achieve community immunity. Some anti-vaccination activists are working in many countries against the Covid-19 vaccine (Lazarus et al., 2020). However, governments and public health officials by some websites and by national and international advertisements work to encourage people to take the vaccine. Encourage people to take the vaccine than build vaccine literacy, and solve some issues

related to the philosophy or religion (Lazarus et al., 2020). Besides, governments appealed to some trust organizations like Red Cross to increase the acceptance of the vaccine (Lazarus et al. 2020). Government supposes to launch campaigns to explain more about the vaccine, how it works, how it is developed, the needed time to vaccine all people achieving community immunity, and to encourage people to cooperate with Covid-19 control restrictions (Lazarus et al., 2020).

Now there are some factors we cannot ignore them like social media which have a big effect on Covid-19 acceptance, governments must be oriented to the social media platforms and spread more information about the vaccine and talk to people from different aspects religious, philosophy and scientifically (Enitan et al. 2020). A study in the UK said that people who are not likely to consume traditional resources like TV, newspapers, radio, and government agencies and were more likely to take news from social media, which is an anti-vaccine platform, refuse the vaccine compared with the others (Murphy et al., 2020). Also, they need to be aware of whatever is circulated on those platforms such as fake news (Enitan et al., 2020). Some rumors about Bill Gates which are related to conspiracy theory said that he created the Covid-19 virus to monitor and control people through an injected microchip or quantum-dot spy software (Enitan et al. 2020). Fully supervision of pharmaceutical companies to avoid any bad information spread among people, fair distribution among countries, free of political, religious, or racial bias must be controlled by governments (Enitan et al. 2020). Also, people need to know nothing is called conspiracy theories are related to Covid-19, it is a global epidemic, we need to thwart all of us to be safe. Experts must illustrate the benefits outwit the side effects for the long term that dispute myths and falsehoods about vaccine safety and efficacy are essential (Murphy et al., 2020).

There are some challenges that poor governments faced having enough vaccine shots for all and how to save this vaccine at low temperatures (Enitan et al., 2020). One of the challenges to accept the vaccine was the way to take it, some people prefer to take it by oral route, subcutaneous or intravenous infusion (Enitan et al 2020). The price of the

Covid-19 vaccine is 20 to 200 EUR that is another challenge for poor countries like Indonesia (Harapan et al., 2020).

Therefore, governments must reach hesitant people through different ways and everywhere like hospital care, prisons, refugee centers. In general, accepting the vaccine is required to maintain herd immunity. Regarding what is published on the social media platforms, Dr. James Madara, the chief executive officer of the American Medical Association, said to the technology companies to ensure leaders make sure about the accurate information on the safety and efficacy of vaccinations on their platforms (Puri et al. 2020), also to reduce the negativity around this vaccine, social media platforms must disable any related advertisements and comments, for example in the US in Twitter, the only hashtag could have appeared is vaccine.gov (Puri et al.,2020). The core of the problem is, most of the time people are sitting at home during the pandemic that means they spend their time on social media, and there they are educated and fed by vaccines as opposed. The disadvantage of the platforms of social media is, people can access and spread whatever they want and remain anonymous and non-human accounts. In addition to some features in social media like suggesting related subjects to their interests which is very bad in spreading some negative ideas. Unfortunately, Twitter reported, hashtag, related to Covid-19 became very fast which is about every 45 milliseconds. Accordingly, health providers, governments, and health agencies must improve their presence on social media platforms to be able to contact directly with different levels of people and explain the right information related to any medical issue (Puri et al., 2020). By the history, taking the vaccine for sure will reduce the disease spread and death rate (Harrison and Wu, 2020). In my opinion, countries can give perks to vaccinated people to encourage others to take the vaccine. In the UAE they offer people free Covid tests to let people evaluate the size of the pandemic than to decide to take the vaccine.

Chapter 3- Project Description

3.1 Introduction

The current pandemic, named Covid-19, is an infectious disease caused by the virus, severe acute respiratory syndrome Covid-19. Covid-19 disease was first diagnosed in late December 2019 in Wuhan, the capital of Hubei Province in China, triggering the first pandemic of the century.

Covid-19, an infectious disease that began spreading internationally since the beginning of the year 2020, has resulted in approximately three million positive cases to date (till 27th April, 2020 there with about one million closed cases having a 20 percent death rate). Covid-19, as we all know, has become a pandemic for a variety of reasons. Some of these include (i) lack of appropriate vaccination and accurate treatment, and (ii) high disease transmission rate. Covid-19 has affected up to 210 countries and territories around the world.

Most government entities have adopted the imposition of curfews to maintain social isolation as the main measure. This method is a good way to prevent the spread of the disease. However, from an economic perspective, complete lockdown could lead to a major financial disaster in the near future.

Lockdown in high-density countries, in particular, may minimize disease transmission rates, while complete control may not be possible. As a result, in order to keep a country's economic position alive, a complete lockdown for an indefinite duration is not ideal under any circumstances. Consequently, a reasonable balance should be struck between the two different aspects of government policies: complete lockdown and healthy free conditions. Thus, an appropriate mathematical model will be able to examine the relationship between PCR testing, an increase in infection cases, and deaths as the

number of vaccinated people increases. It may be able to predict the disease's behavioral feature in the near future.

Despite the fact that the fundamental mathematical models on theoretical epidemiology appear to be very simple, they are the major works on mathematical epidemiology from a fresh perspective. We can now create a complicated mathematical model and study it completely in a scientific manner because to advances in computer tools and software.

Stakeholders can now use the results of my project and similar projects to plan for the reopening of government and private entities. We can also expand this in order to achieve economic improvement. Furthermore, universities have the opportunity to change how they admit students, for example, by admitting more students in fields such as medicine, pharmacology, medical analysis, nursing, and genetics.

Chapter 4- Data Analysis

4.1 Dataset Analysis

My dataset   https://ourworldindata.org/coronavirus/country/united-arab-emirates
I found this data on the Our World in Data website and contains 60 columns and 65535(from 29 Jan 2020 to 14 Jan2022) records in a CSV file. A group of researchers started to collect this data from the beginning of the pandemic from government health agencies around the world and keep updating it to the present.

Table 1: First Five Rows of the Dataset

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	female_smokers
0	AFG	Asia	Afghanistan	2020-02-24	5.0000	5.0000	NaN	NaN	NaN	NaN	...	NaN
1	AFG	Asia	Afghanistan	2020-02-25	5.0000	0.0000	NaN	NaN	NaN	NaN	...	NaN
2	AFG	Asia	Afghanistan	2020-02-26	5.0000	0.0000	NaN	NaN	NaN	NaN	...	NaN
3	AFG	Asia	Afghanistan	2020-02-27	5.0000	0.0000	NaN	NaN	NaN	NaN	...	NaN
4	AFG	Asia	Afghanistan	2020-02-28	5.0000	0.0000	NaN	NaN	NaN	NaN	...	NaN

5 rows × 67 columns

4.2 Understanding my Data

Table 2: Types of Columns

```
 #   Column                               Non-Null Count    Dtype
---  ------                               --------------    -----
 0   iso_code                             154561 non-null   object
 1   continent                            145255 non-null   object
 2   location                             154561 non-null   object
 3   date                                 154561 non-null   object
 4   total_cases                          151890 non-null   float64
 5   new_cases                            151844 non-null   float64
 6   new_cases_smoothed                   150693 non-null   float64
 7   total_deaths                         134569 non-null   float64
 8   new_deaths                           134742 non-null   float64
 9   new_deaths_smoothed                  134610 non-null   float64
10   total_cases_per_million              151182 non-null   float64
11   new_cases_per_million                151136 non-null   float64
12   new_cases_smoothed_per_million       149990 non-null   float64
13   total_deaths_per_million             133874 non-null   float64
14   new_deaths_per_million               134047 non-null   float64
15   new_deaths_smoothed_per_million      133915 non-null   float64
16   reproduction_rate                    116450 non-null   float64
17   icu_patients                         20395 non-null    float64
18   icu_patients_per_million             20395 non-null    float64
19   hosp_patients                        21563 non-null    float64
20   hosp_patients_per_million            21563 non-null    float64
21   weekly_icu_admissions                4286 non-null     float64
22   weekly_icu_admissions_per_million    4286 non-null     float64
23   weekly_hosp_admissions               8687 non-null     float64
24   weekly_hosp_admissions_per_million   8687 non-null     float64
25   new_tests                            61107 non-null    float64
26   total_tests                          62746 non-null    float64
27   total_tests_per_thousand             62746 non-null    float64
```

I moved to the next stage; I dropped the following 25 columns:

'total_cases_per_million', 'new_cases_per_million', 'total_deaths_per_million',

'new_cases_smoothed_per_million', 'new_deaths_smoothed_per_million',

'new_deaths_per_million', 'icu_patients_per_million', 'hosp_patients_per_million',

'weekly_icu_admissions_per_million', 'weekly_hosp_admissions_per_million',

'total_tests_per_thousand', 'new_tests_per_thousand',

'new_tests_smoothed_per_thousand', 'total_vaccinations_per_hundred',

'people_vaccinated_per_hundred', 'people_fully_vaccinated_per_hundred',

'total_boosters_per_hundred', 'new_vaccinations_smoothed_per_million',

'new_people_vaccinated_smoothed',

'new_people_vaccinated_smoothed_per_hundred',

'stringency_index', 'handwashing_facilities', 'hospital_beds_per_thousand',

'life_expectancy',

'human_development_index','excess_mortality_cumulative_per_million'

I noticed connections between the columns and their effects on one another.
I used a heat map visualization, which is an excellent way to visualize complex
statistical data.
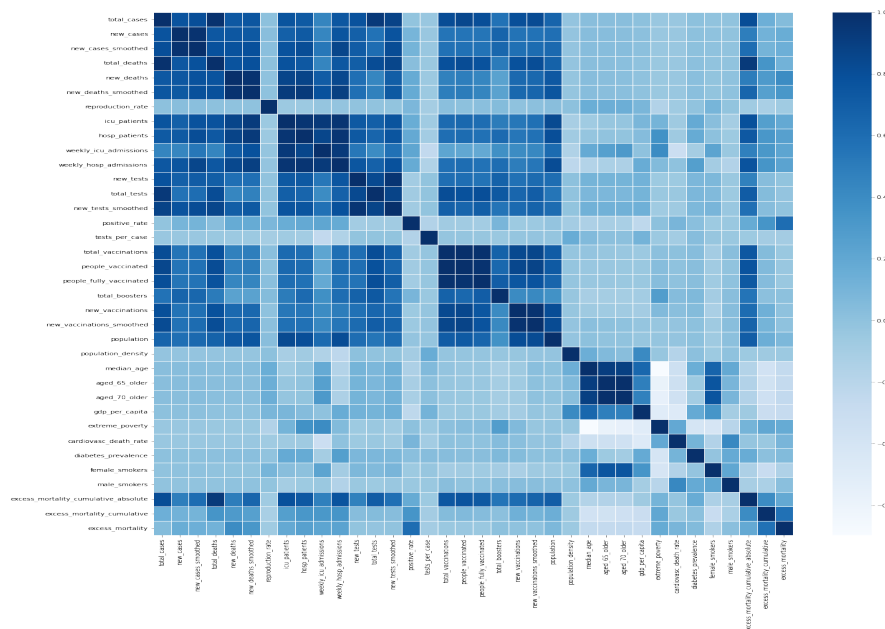
Figure 2: Heart Map for the Dataset

After this, the following columns were dropped:

'population_density','median_age', 'aged_65_older', 'aged_70_older', 'gdp_per_capita', 'extreme_poverty', 'cardiovasc_death_rate','diabetes_prevalence', 'female_smokers', 'male_smokers', 'excess_mortality_cumulative', 'excess_mortality', 'reproduction_rate', 'tests_per_case', 'positive_rate', 'new_cases_smoothed', 'new_deaths_smoothed', 'new_tests_smoothed'.

For more understanding another heat map for the remaining data is developed.
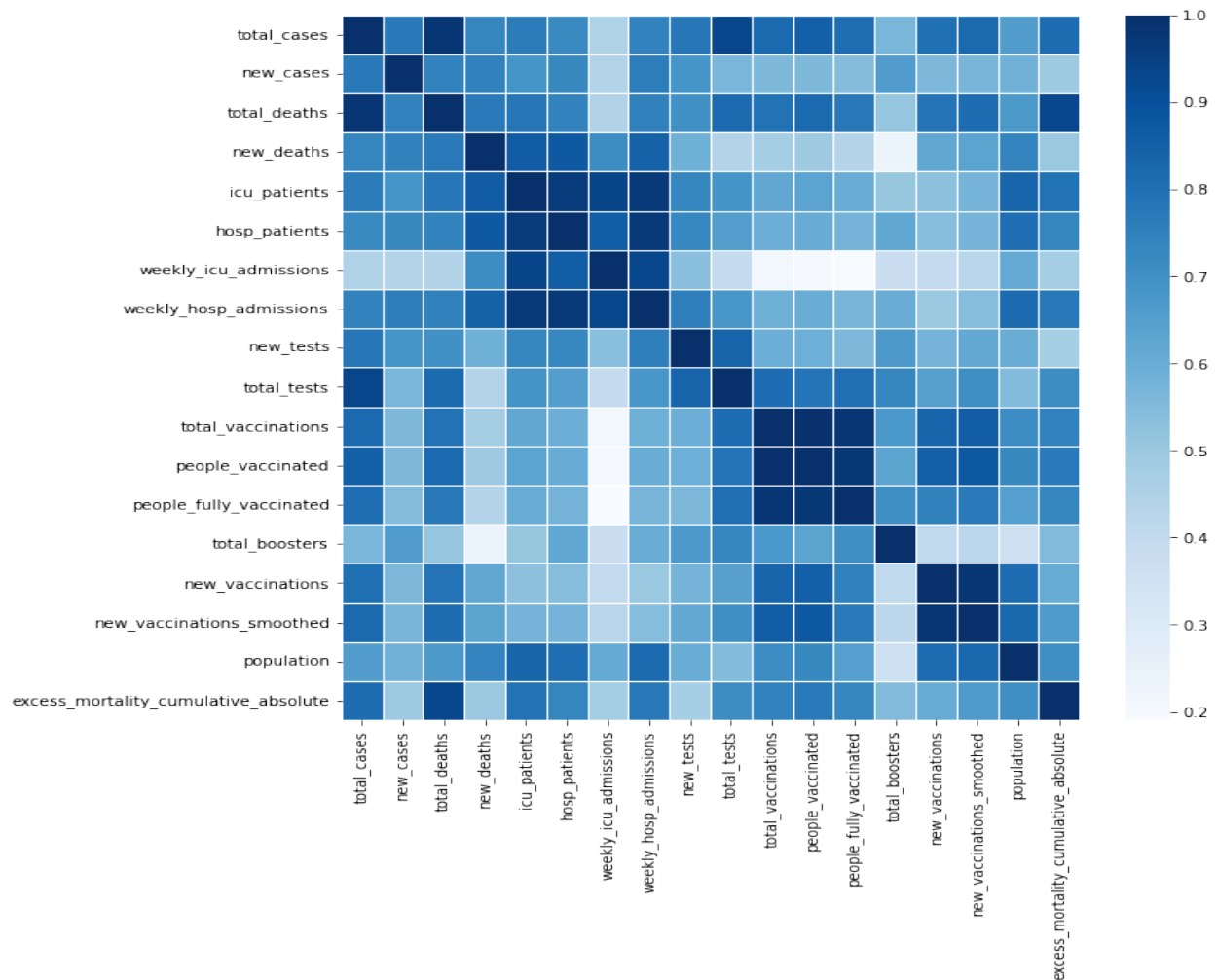
Figure 3: Heart Map for the Remaining Columns

The reader can see that the overall number of cases has a high impact on the total number of people who die, as well as the total number of tests, thus more tests will help find more cases and estimate the death rate.

Also, the total number of people who have been vaccinated is directly influenced by the total number of cases, total number of people who have been tested, and total number of people who have died.

Followed by preprocessing phase, the null values /cells were counted.

Table 3: Counting Null Values

```
iso_code                            0
date                                0
location                            0
population                       1025
total_cases                      2671
new_cases                        2717
continent                        9306
new_deaths                      19819
total_deaths                    19992
tests_units                     78616
new_vaccinations_smoothed       81570
total_tests                     91815
new_tests                       93454
total_vaccinations             114580
people_vaccinated              116306
people_fully_vaccinated        119137
new_vaccinations               121460
hosp_patients                  132998
icu_patients                   134166
dtype: int64
```

All numerical values were converted to integers, and statistical functions such as mean, median, maximum, minimum, and standard deviation were calculated. Then I saw that the number of cases was decreasing. For Covid-19, the situation is terrible, and they must work extremely hard to locate the vaccine while also encouraging people to use it.

Table 4: Descriptive Statistics

	total_cases	new_cases	total_deaths	new_deaths	icu_patients	hosp_patients	new_tests	total_tests
count	151890.0000	151844.0000	134569.0000	134742.0000	20395.0000	21563.0000	61107.0000	62746.0000
mean	2173589.9950	8959.1177	53377.7442	170.2777	921.0866	4052.6487	63440.2589	15084747.5770
std	12887619.4638	58421.7271	280507.7826	829.8944	2701.3774	11229.5510	209515.5484	57895106.1610
min	1.0000	-74347.0000	1.0000	-1918.0000	0.0000	0.0000	1.0000	0.0000
25%	1613.0000	1.0000	69.0000	0.0000	29.0000	115.0000	2340.0000	321338.2500
50%	21330.0000	71.0000	689.0000	2.0000	151.0000	604.0000	8550.0000	1646673.0000
75%	260292.0000	942.0000	6557.0000	19.0000	606.0000	2649.0000	34496.5000	7212181.2500
max	323567887.0000	3701643.0000	5529524.0000	18062.0000	28891.0000	149038.0000	3740296.0000	737405541.0000
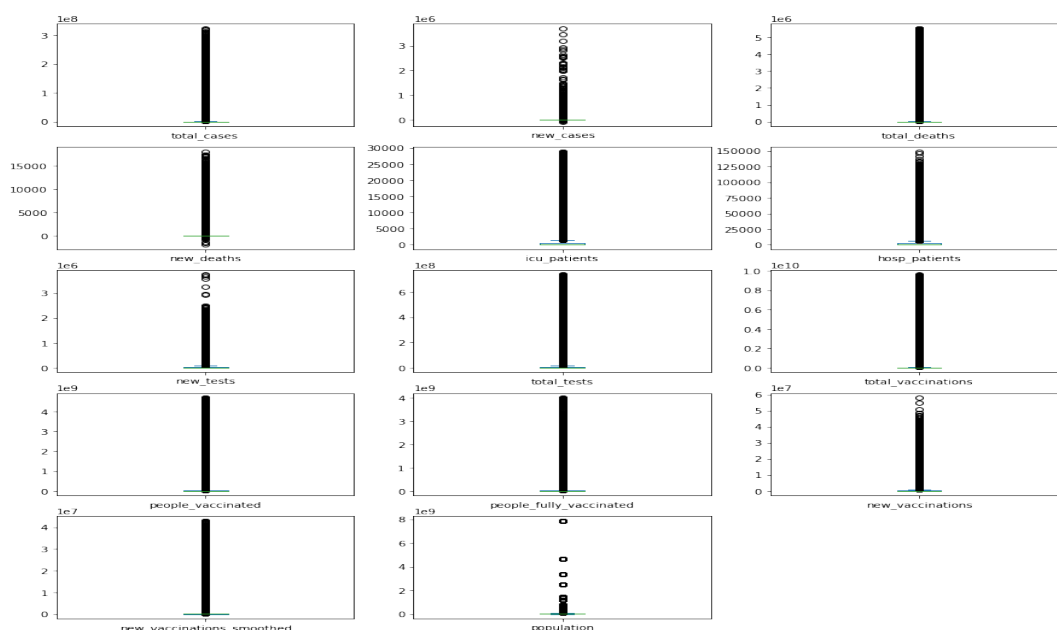
Figure 4: Outliers Analysis

I went on to investigate the missing values and outliers. I discovered using the boxplot that there are no outlier values on several attributes, such as the population, which is acceptable, and on new cases, vaccines, and testing, which are all acceptable. This is also appropriate because the population is directly proportionate. I noticed that the distribution of my data is left skewed, as the data is decreasing by time naturally.

To tackle the issue of missing values, I applied the method (LOCF), which is Last Observation carried forward look one row ahead instead of filling the missing value.

```
iso_code                          0
continent                         0
location                          0
date                              0
total_cases                       0
new_cases                         0
total_deaths                      0
new_deaths                        0
new_tests                         0
total_tests                       0
total_vaccinations                0
people_vaccinated                 0
people_fully_vaccinated           0
new_vaccinations                  0
new_vaccinations_smoothed         0
population                        0
dtype: int64
```

Figure 5: Missing Values

4.2 Visualization

I have created a data frame for the top 20 countries in the Covid-19 cases, and then used that dataset to investigate the global situation. This map displays the global distribution of Covid-19 cases; the deeper color indicates a higher density of cases compared to the general population.
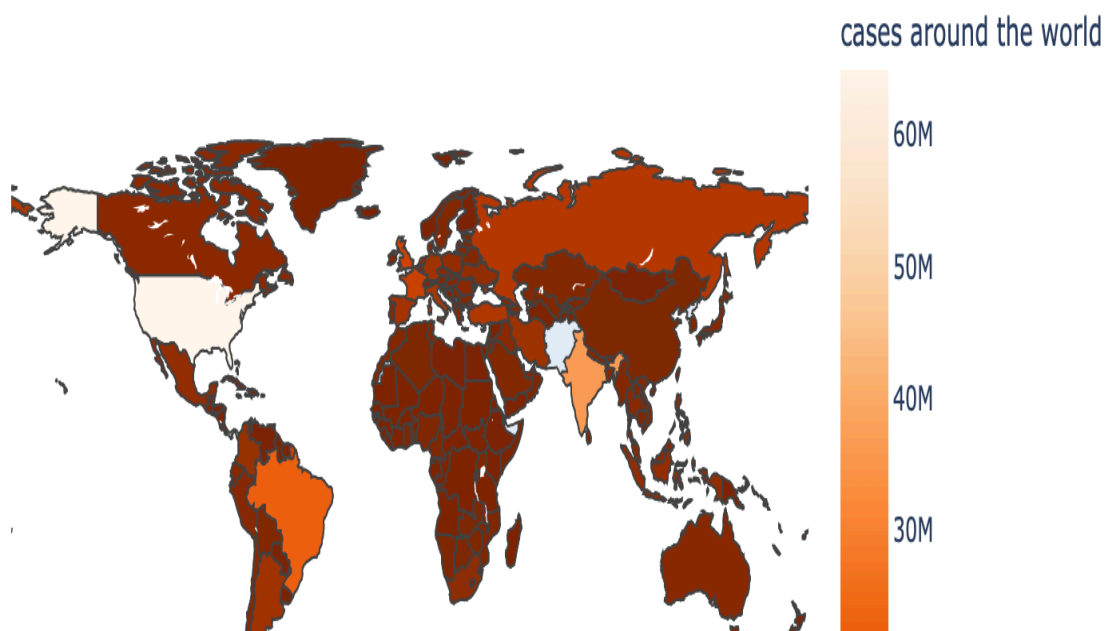

Figure 6: Distribution of COVID-19 Cases around the world

This chart shows the increase of Covid-19 infected cases from January 2020 to January 2022. In addition, for the past two years, I have been researching the increasing in the positive cases. In both statistics, I discovered that Covid-19 cases have been steadily growing, which raises numerous problems, such as what about the vaccine? As the vaccine dose is increased, the number of new cases increases. What is the most important component in locating new cases, and how can this be managed?
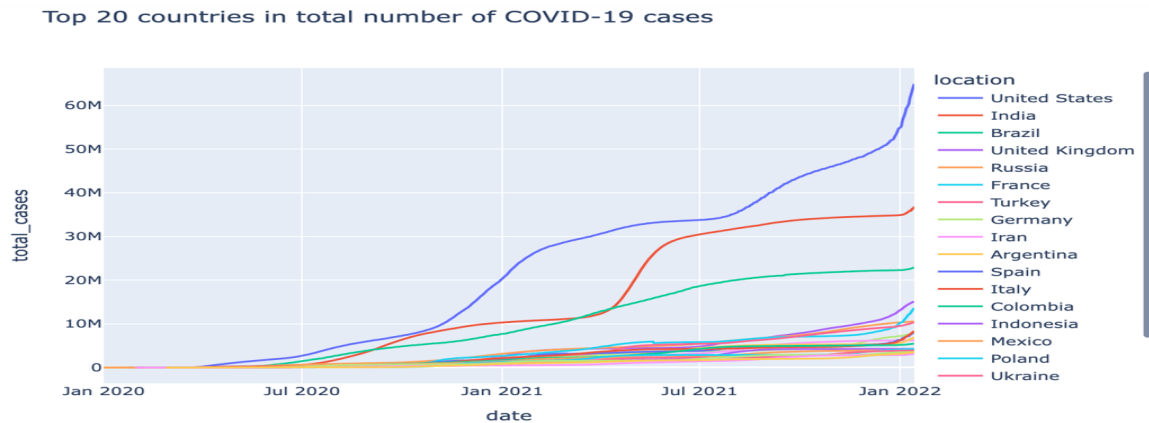
Figure 7: Top 20 Countries in Total Number of COVID-19 Cases

Then a deeper examination into the data and graphed the new cases over a two-year period is performed and it is discovered that there are four waves to this epidemic and reading that the instances spiked throughout the winter and fall seasons.
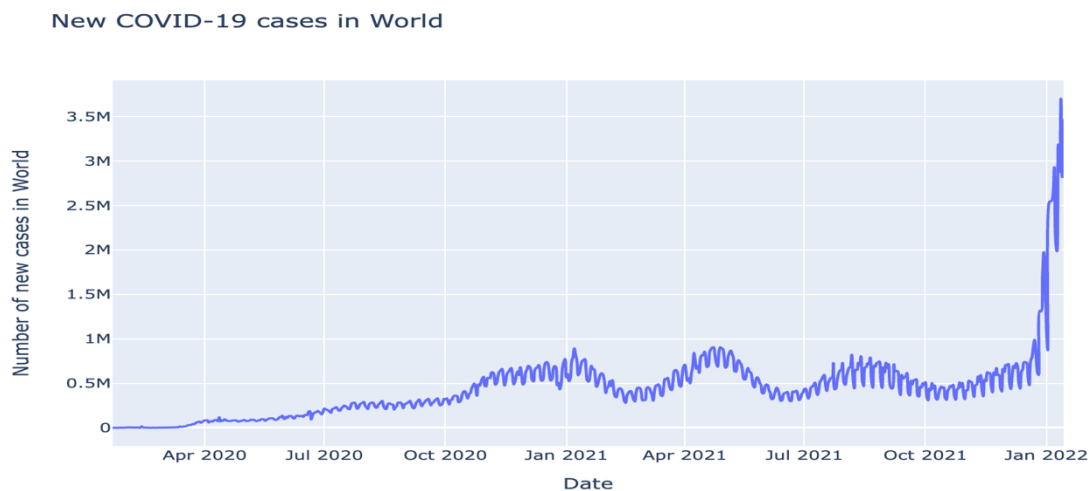

Figure 8: New COVID-19 Cases in World

I decided to look at the statistics for fully vaccinated people, and I discovered that the rate of fully vaccinated people was also increasing, accordingly that there may be a correlation between new cases and vaccinated people.
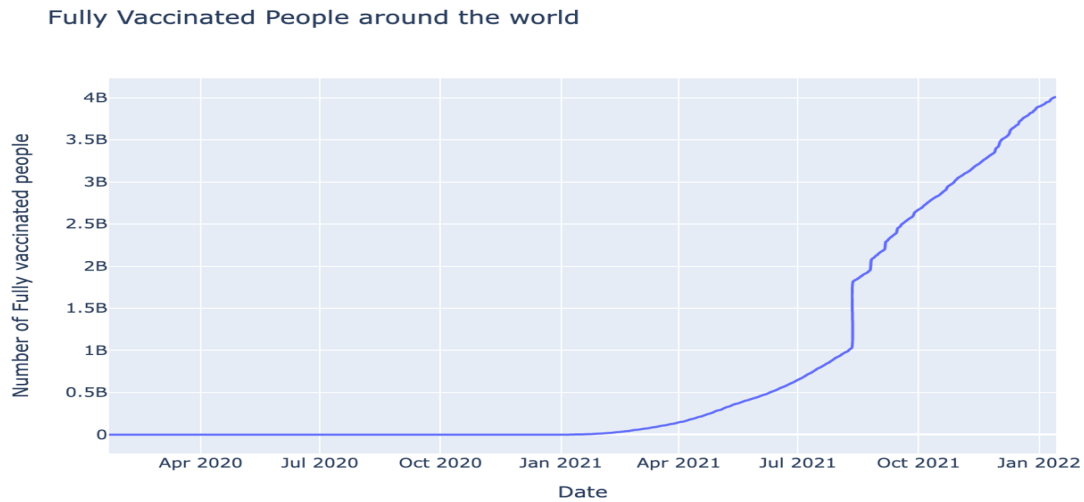
Figure 9: Fully Vaccinated People around the World

I mean, by encouraging people to do PCR tests or any other method to check for the Covid-19 and determine the cases around the world, people will be encouraged to take the vaccine.
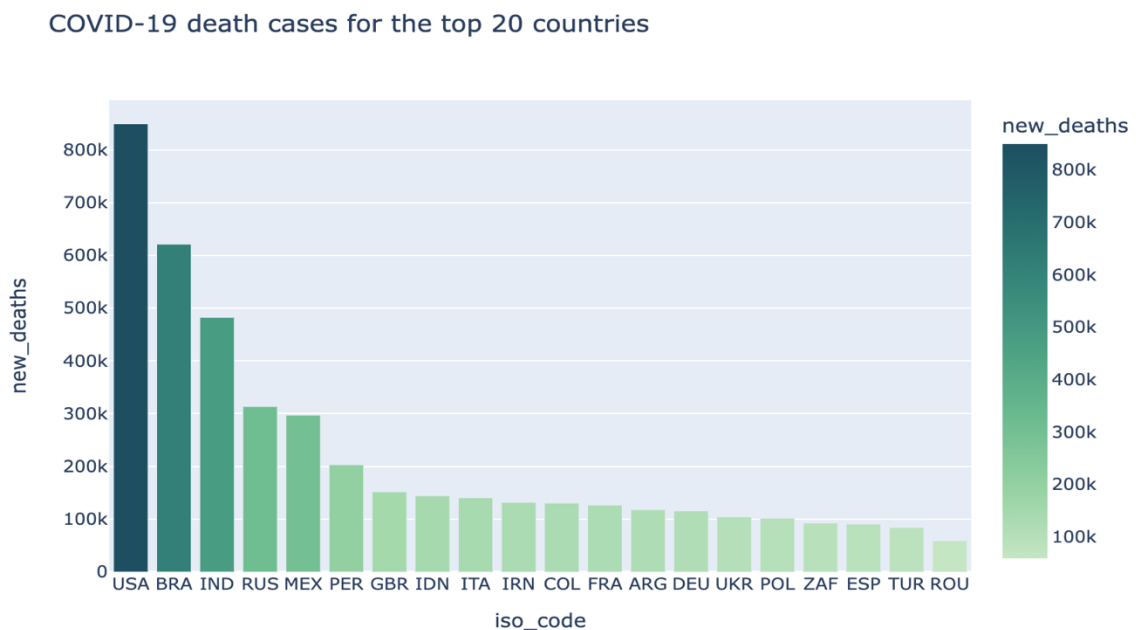
Figure 10: Death cases for the top 20 Countries
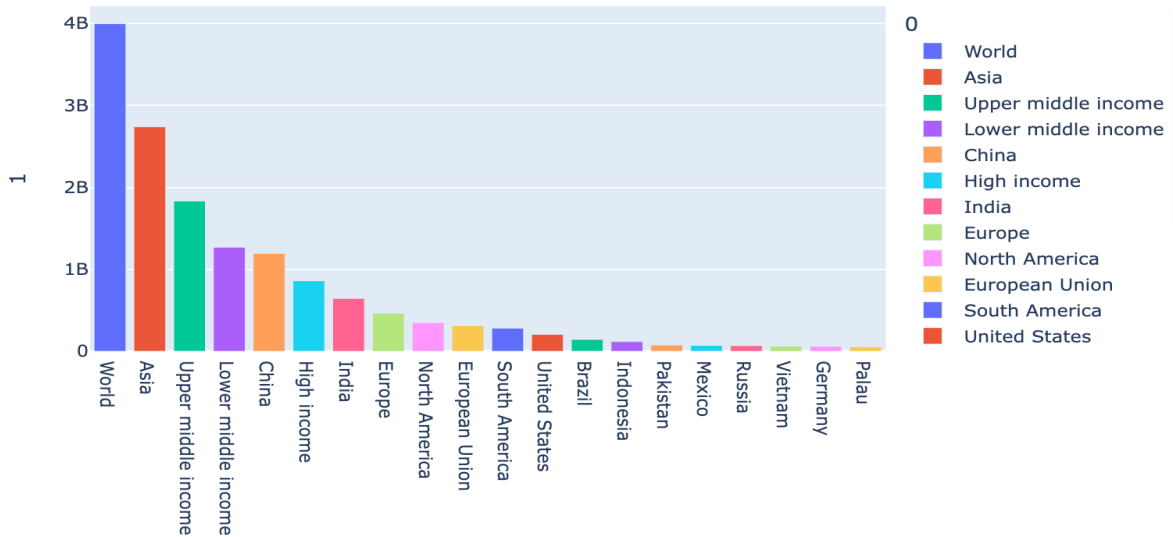
Numbers of fully vaccinated people

Figure 11: Vaccinated people for the top 20 Countries

It's crucial to figure out if there's anything else that could influence the number of people who have been vaccinated. When I looked at the death rate for the top 20 countries, I discovered that it was quite high, reaching about 1 million for the United States of America.

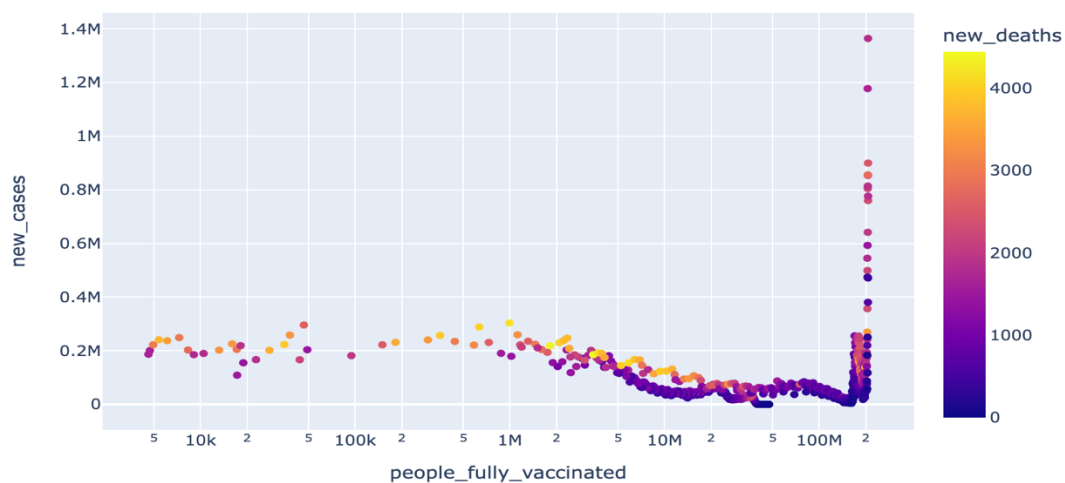Number of new cases relevance to number of vaccinated people -USA-

Figure 12: New cases relevance to the number of vaccinated people in USA

23

From this graph, I concluded that the number of fully vaccinated people increased throughout the interval, but the number of new cases remained stable and then decreased as the new death rate decreased. As a result, I came to the conclusion that more vaccination, fewer new cases, and a lower death rate all proportioned inversely.
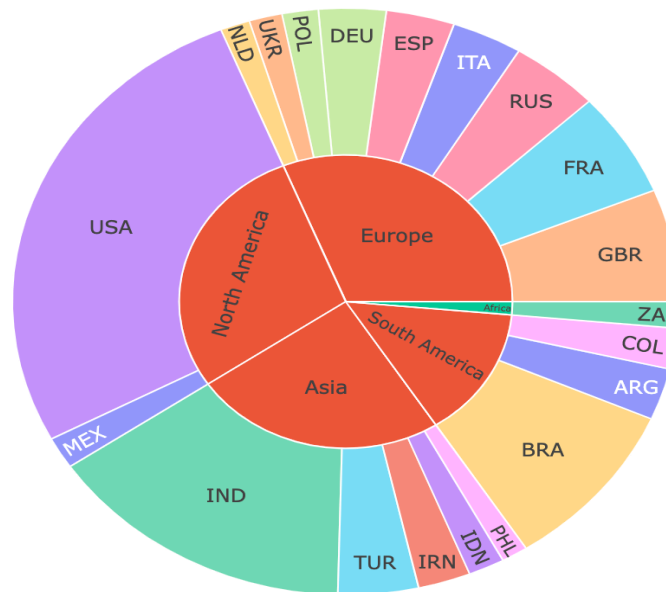
Figure 13: Vaccinated people per Continent

Making an examination at the percentage of death cases per content and then the percentage for the top 20 countries in this graph, we see that the problem is concentrated in North America, Europe and Asia, which means that the governments of these countries need to act to reduce the high death rate.

After that focused on the United Arab Emirates, as I mentioned at the beginning of my project. I used the same dataset but focused on the UAE and came to the same conclusions.

4.3 Machine learning:

The linear regression model used in my project showed a linear relation between the independent variable "people_fully_vaccinated" and the independent variable "new_cases".

Linear regression is typically used for predictive analysis and the accuracy from my study is 99.7% , which proved to be the best model based on the data obtained using various machine learning methods.

The linear regression equation has the form Y= a+bX, such that Y is the independent variable that I want to predict I'm using to predict, X the dependent variable, b is the slope and finally the y-intercept is the value of a and it is equal to 0 in my model.

For the linear regression analysis is used OLS method. After that I used multi linear regression methods (Lasso and Ridge) then I found in the OLS result that the R-squared is equal to 1 which means how the independent variable makes change in the dependent variable.

I used LASSO regression to check the correlation, I found that it is similar to simple linear regression. That was clear on the fitting of actual data with the predicted data. We may say that a linear regression model can be used to explain completely vaccinated persons and new cases. Positive cases rise in lockstep with the number of persons who have been fully vaccinated. Regression models are frequently employed to provide statistical support for statements about everyday realities.

Overall, the results fulfilled our expectations by demonstrating the ability of several machine learning algorithms to predict the number of persons who have been vaccinated. The percentage of predictions for each algorithm varied slightly, but it successfully demonstrated that machine learning algorithms can extract useful information from large

amounts of data and predict fully vaccinated people. This will, in reality, play a significant part in anticipating the organizations' reopening.

Chapter 5 Conclusion

5.1 Conclusion

The main goal of this study was to look at several techniques for effecting the global increase of vaccinated persons in order to close the Covid-19 chapter.

Indeed, data science has overseen the creation of a Linear regression model that can determine the correlation between certain parameters and fully vaccinated persons. The goal of this project's thesis was to put a notion to the test by evaluating and testing the model's accuracy. The research findings have plainly defined the research theme.

Furthermore, the research's findings confirmed that the Linear Regression model is a good model for determining the number of persons who have been vaccinated, which can aid in making decisions and preventing the spread of disease. In fact, the study cycle has addressed nearly all of the research topics and achieved the primary purpose of being explored through discussion and examination.

5.2 Recommendations

After World War II, the world has never seen anything like the current circumstances brought on by the Covid-19 pandemic. Covid-19 has spread to 210 countries and territories since its discovery in Wuhan in December 2019, and this pandemic nature continues to grow.

Because these types of conditions are uncommon in modern society, several governments have implemented policies to limit the disease's devastating character as much as feasible.

Accordingly, the output model that I built can help the decision makers to assist people in receiving the vaccine in the instance of Covid-19. This will allow life to continue as usual while reducing losses.

I recommend that all governments prepare classified data for people in each city based on their ages, occupations, gender, and health status, as well as additional sub datasets to classify people by each key factor.

I also advise the owner of this data set to go a little deeper to find out who among the groups received the vaccine later than the others in order to raise awareness.

5.3 Future Work

For future work on this project, I recommend that the researchers collect data on companies and corporations to determine when they reopen and resume work from their offices, and what the relationship is between this and the regular PCR tests and infected people, as well as the most important factor, the dates of vaccination for employees.

They can also concentrate on the age and gender of the employees who received the vaccine first, which will aid in the addition of this as a factor in forecasting hiering employees. In addition, they can study the economic problems.

Bibliography

1. Kashte,Gulbake,El-Amin III and Gupta,Shivaji,Arvind,Saadiq and Ashim.(2021). Covid-19 vaccines: rapid development, implications, challenges and future prospects. *Human Cell*.

2. Mohammad, Trounev and Cattani. (2021). The dynamics of Covid-19 in the UAE based on fractional derivative modeling using Riesz wavelets simulation. UAE .Advances in Difference Equations.

3. Naiankara,Muqattash,Niankara and Traoret , Ibrahim,Riham, Aminata and Rachidatou . (2020). A Random Utility Model of Stated Vaccine Preferences Analysis: With application to Covid-19 Vaccine Demand in the UAE. UAE.AlAin University.

4. Oliver,Lepri,Sterly,Lambiotte,Deletaille,DeNadai ,Letouzé,Salah, Benjamins, Cattuto, Colizza, de Cordes, P. Fraiberger,Koebe, Lehmann, Murillo, Pentland, Pham, Pivetta, Saramäki, V. Scarpino, Tizzoni, Verhulst and Vinck , Nuria, Bruno, Harald, Renaud, Sébastien, Marco, Emmanuel, Albert, Richard, Ciro, Vittoria, Nicolas, Samuel , Till , Sune, Juan, Alex, Phuong, Frédéric , Jari, Samuel,. Scarpino, Michele, Stefaan, Patrick . (2020). Mobile phone data for informing public health actions across the Covid-19 pandemic life cycle.USA.

5. Harper, Satchell,Fido and Latzman . Craig A., Liam P., Dean and Robert D. (2020). Functional Fear Predicts Public Health Compliance in the Covid-19 Pandemic . UK. International Journal of Mental Health and Addiction.

6. Lee,Zabinsky,Wasserheit, Kofsky and Liu. Serin Zelda B. Judith N. Stephen M. and Shan.(2021). Covid-19 Pandemic Response Simulation in a Large City: Impact of Nonpharmaceutical Interventions on Reopening Society.USA. Medical Decision Making.

7. Sallam. Malek. (2021). Covid-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates. Jordan. MDPI.

8. Lazarus, Ratzan,Palayew,Gostin,Larson, Rabin, Kimball and El-Mohandes. Jeffery V.,Scott C., Adam,Lawrence O., Heidi,Kenneth,Spencer and Ayman. (2020). A global survey of potential acceptance of a Covid-19 vaccine. Spain. Naturemedicine.

9. Enitan, Oyekale , Akele , Olawuyi , Olabisi , Nwankiti , Adejumo , and Enitan. Seyi, Adesola, Richard, Kayode, Elisha, Amarachi, Esther, and Comfort. (2020). Assessment of Knowledge, Perception and Readiness of Nigerians to Participate in the Covid-19 Vaccine Trial. Naigeria. Sci Forschen.

10. Murphy,Bentall,Vallieres and Shevlin .Jamie,Richard,Frederique and Mark. (2020). Preparing for a Covid-19 vaccine: Identifying and psychologically profiling those who are vaccine hesitant or resistant in two general population samples. UK

11. Harapan, Wagner, Yufika, Winardi, Anwar, Gan, Setiawan, Rajamoorthy, Sofyan, and Mudatsir. Harapan, Abram L., Amanda,Wira, Samsul ,Alex, Abdul Malik, Yogambigai, Hizir, and Mudatsir . (2020). Acceptance of a Covid-19 Vaccine in Southeast Asia: A Cross-Sectional Study in Indonesia. Indonesia. PMC.

12. Whitehead and Perry. Andrew L. and Samuel L. (2020). How Culture Wars Delay Herd Immunity: Christian Nationalism and Anti-vaccine Attitudes. USA. SAGE Journal.

13. Puri,Coomes,Haghbayan and Gunaratune. Neha,Eric,Hourmazd and Keith. (2020). Social media and vaccine hesitancy: new updates for the era of Covid-19 and globalized infectious diseases. Canada. Human Vaccines & Immunotherapeutics

14. Harrison and Wu. Emily A. and Julia W. (2020). Vaccine confidence in the time of Covid-19. USA. European Journal of Epidemiology.