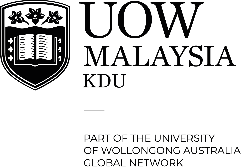
A close up of text on a white background

Description automatically generated

Assignment Coversheet **Please complete all details required clearly.** For softcopy submissions,please ensure this cover sheet is included at the start of your document or in the file folder.

**Assignment & Course Details:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Subject Code:** *(e.g. XCAT1234)*  XBDS 2014 | | **Subject Name** *(e.g. Fundamentals of Computing*)**:**  Introduction to Data Science | |
| **Course** *(e.g. Bachelor in Computing)* :  Bachelor of Computer Science | | | |
| **Lecturer Name:**  Dr. Law Foong Li | | | |
| **Assessment Due Date:** *(dd/mm/yy)* | 09/08/2021 | **Assessment Title:** | Group Assignment |

FOUNDATION DIPLOMA  DEGREE  MASTER



*Please tick* ✓ *or click if using MS WORD*

**Marks Breakdown**

Signature: **Total**

**For office use only** – Lecturer comments (if applicable)

*I/We declare that:*

* *This assignment is my/our own original work, except where I/we have appropriately cited the original source.*
* *This assignment or parts of it has not previously been submitted for assessment in this or any other subject.*
* *I/We allow the assessor of this assignment to test any work submitted by me/us, using text comparison software for plagiarism.****(For more information, Please read the Academic Integrity Guidelines)***

**Name :** Thean Jun Chao **Student ID:** 0127122 **Email :** 0127122@kdu-online.com **Mobile No:** 0173479830 **Signature:  
Date:** 08/08/2021

SCCM/CSForm/Rev. Ver 2.4

**Name :** Taashwin Reddy A/L Ramamoorthy **Student ID:** 0128664 **Email :** 0128664@kdu-online.com **Mobile No:** 0172632461 **Signature:  
Date:** 08/08/2021

**Name :  
Student ID:  
Email :  
Mobile No: Signature:  
Date:**

**Name :  
Student ID:  
Email :  
Mobile No: Signature:  
Date:**

**Name :  
Student ID:  
Email :  
Mobile No: Signature:  
Date:**

**Name :** Bernard Lee Hanlin **Student ID:** 0131545 **Email :** 0131545 @kdu-online.com **Mobile No:** 0163477019 **Signature:  
Date:** 08/08/2021

**Table of Contents**

[1 Introduction 3](#_Toc79326356)

[1.1 Problem Statement 3](#_Toc79326357)

[1.2 Limitations of Existing Solutions 3](#_Toc79326358)

[1.3 Proposed Solution and Approach 4](#_Toc79326359)

[1.4 Hypothesis & Research Question 4](#_Toc79326360)

[2 Methodology 5](#_Toc79326361)

[2.1 Data Collection 5](#_Toc79326362)

[2.2 Programming Language 5](#_Toc79326363)

[2.3 Data Understanding 5](#_Toc79326364)

[2.4 Data Preparation & Feature Engineering 7](#_Toc79326365)

[2.5 Modelling 11](#_Toc79326366)

[2.6 Platform 12](#_Toc79326367)

[3 Results 12](#_Toc79326368)

[3.1 Evaluation 12](#_Toc79326369)

[3.2 Samples Screenshot of the Application 19](#_Toc79326370)

[4 Discussion 30](#_Toc79326371)

[4.1 Changes in Tools and Methods Used 30](#_Toc79326372)

[4.2 Answer to the Research Question 31](#_Toc79326373)

[4.2.1 Research Question 1 (RQ1) 31](#_Toc79326374)

[4.2.2 Research Question 2 (RQ2) 32](#_Toc79326375)

[5 Conclusion 33](#_Toc79326376)

[5.1 Implications of the Findings to Machine Learning Area 33](#_Toc79326377)

[5.2 Limitations & Future Enhancement 33](#_Toc79326378)

[6 References 34](#_Toc79326379)

# Introduction

## Problem Statement

The covid-19 pandemic has adversely affected people’s lives and livelihoods around the world. Economies worldwide are severely affected by the pandemic as companies have gone bankrupt or have to fire some employees to survive. Atradius, a Dutch insurance company predicted a 26% increase in bankruptcies globally. LegalJobs reported that as of September 2020, 470 companies had gone bankrupt, including NPC International Inc; the largest franchisee of PizzaHut restaurants. Besides, the healthcare sector is suffering tremendously as front liners are getting infected or killed. A shortage of oxygen tanks in some countries, causing more distraught among the people as people are dying at a much faster rate since newer mutated variants of the virus are more contagious and dangerous. Frontliners are overworked for more than a year to standby to aid infected patients. Ensuring vaccinations progress at a steady increasing rate worldwide is necessary as vaccines are the main solution to overcoming this pandemic. However, some people are hesitant to get vaccinated mainly due to the questions regarding the vaccine safety. A study conducted on the acceptance of the covid-19 vaccination in China found that despite having a 91.3% acceptance rate among 2058 participants from various provinces, only 52.2% wanted to get vaccinated as soon as possible while the other 47.8% would delay until the safety of the vaccines is confirmed (Wang et al., 2020). To persuade and educate the public to get vaccinated as soon as possible, strong evidence of the effectiveness of vaccination must be displayed to the. Therefore, scientific research shall be carried out on the effectiveness of the vaccination in reducing the positive cases and death cases due to covid-19, as well as when a nation can reach herd immunity.

## Limitations of Existing Solutions

Several observational studies have been taken to assess the vaccine effectiveness by the Centers for Disease Control and Prevention (CDC, 2021).

1. Case-control studies

Two groups of participants are asked if they have been vaccinated. The two groups are the case group, referring to people infected with covid-19 and the control, referring to people who have not been infected with covid-19 (CDC, 2021).

1. Cohort studies

Observing groups of vaccinated and unvaccinated people for some time to see if they get infected by the covid-19 virus later (CDC, 2021).

1. Screening method assessments

Collecting the vaccination status among a group of people infected with the covid-19 virus (CDC, 2021).

1. Ecologic analysis assessments

Collecting data from people from different locations or at different times to find a relationship between those vaccinated and those infected with covid-19 (CDC, 2021).

One of the limitations of observational studies is the long observational time and high cost is needed to conduct the study (WHO, 2021). The different situations in different locations and rapid vaccination rollout makes it harder to compare vaccination rates between vaccinated and unprotected locations (WHO, 2021). Recall bias can occur as the participants may have forgotten the vaccine brand and the dates of their vaccination (WHO, 2021). Also, the study can be further enhanced by predicting the future number of new cases, new deaths, and when herd immunity can be attained.

## Proposed Solution and Approach

To convince more people to get vaccinated as soon as possible, we will use machine learning to train trusted and publicly available data to identify the trend of the vaccination against the new cases and new deaths. A decreasing trend is expected as the number of people who received their vaccination increases. Furthermore, we will also compare the effectiveness of partial vaccination (1 dose) against (2 doses) in terms of reducing the new cases and new deaths. Lastly, we would also allow the user to predict roughly when the United States can attain herd immunity.

## Hypothesis & Research Question

This paper aims to study the following research questions:

* RQ1: How does one dose of vaccination compare with two doses of vaccination in the effectiveness of reducing the number of positive Covid-19 cases and the number of deaths due to the Covid-19 in the USA?
  + Research hypothesis: Two doses of vaccination will be more effective in reducing the number of positive Covid-19 cases and the number of deaths due to the Covid-19 in the USA.
* RQ2: When will the US attain herd immunity (70% vaccinated)?
  + Research hypothesis: The USA can attain herd immunity by winter 2021 (D’souza and Dowdy, 2021)

# Methodology

## Data Collection

The dataset used in this study was collected from a public source in Github. The data contained within the dataset is sourced from a variety of legitimate sources, including the official data collated by Our World in Data; the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University; a peer study from Franciscco Arroyo-Marioli, Francisco Bullano, Simas Kucinskas, and Carlos Rondon-Moreno; Oxford COVID-19 Government Response Tracker, and International organizations including but not limited to UN, World Bank, OECD (Organisation for Economic Co-operation and Development) and IMHE (Institute for Health Metrics and Evaluation). The dataset includes 60 columns of data, each representing one variable of the dataset. The date range is between 2020-01-22 and 2021-07-31.

## Programming Language

The study is conducted by implementing a project using python programming, a general-purpose language that is easy to learn as the high-level syntax that can be understood easily.

## Data Understanding

The initial step taken to understand the dataset is by performing the exploratory data analysis (EDA). The EDA is conducted to determine the significance between variables; several libraries are imported into python. These libraries are:

* Pandas: Library for data manipulation and analysis
* Numpy: Support library for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
* Matlplotlib: Plotting library that provides an object-oriented API for embedding plots into applications.
* Seaborn: Data visualization library built on top of matplotlib and closely integrated with pandas data structure in Python. Visualisation is the central part of seaborn which helps in the exploration and understanding of data.

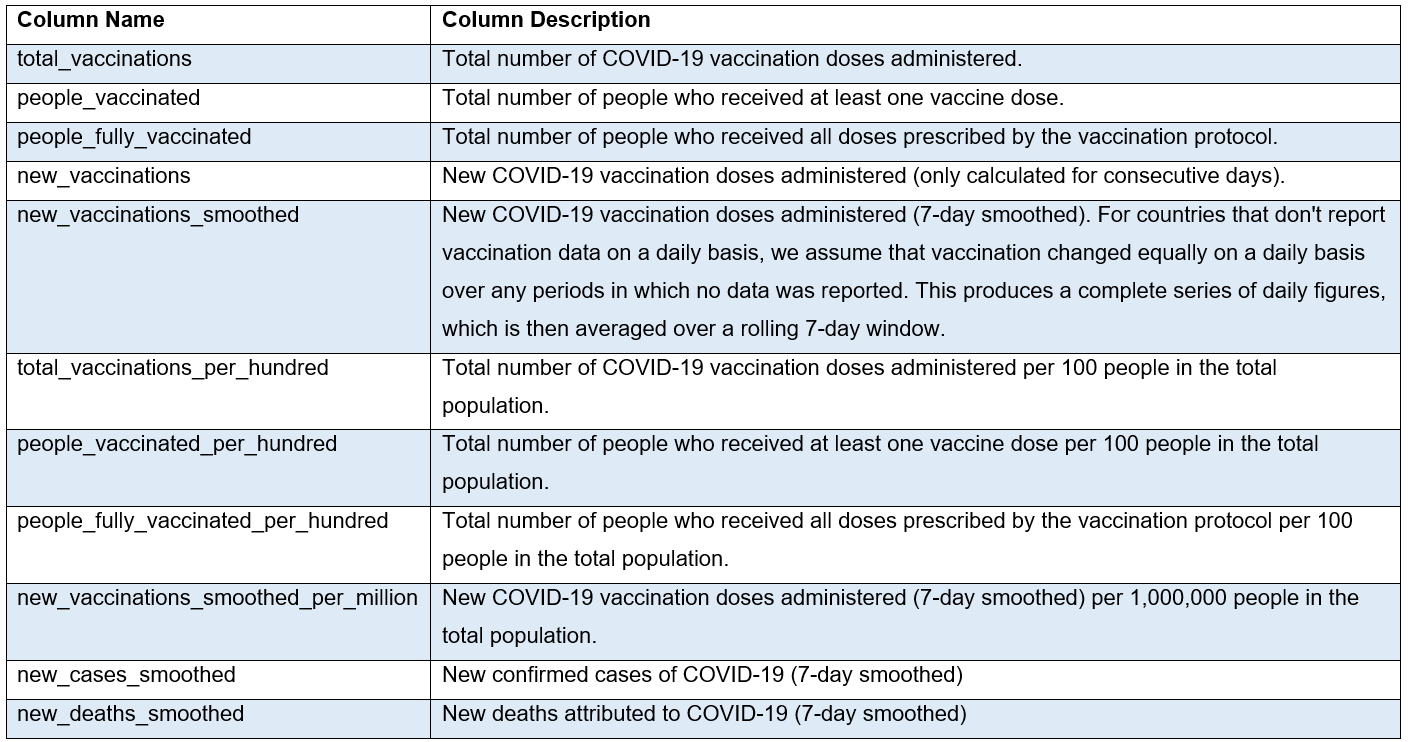
As shown in figure 1, the dataset contains 96,654 rows and 60 columns.



Figure 1: The shape of the dataset

Eleven columns were selected to perform the initial EDA process. These columns are listed in table 1 below.

Table 1: Columns used to perform the initial EDA process.



Two python functions namely generate\_heat\_map and plot\_scatter\_plot is defined to perform the EDA. The scatter plots help to identify the trends of the data. Heatmap on the other hand displays the correlation of the target variable against the other variables to identify the relevant dependent variables for the independent variable. Figure 2 below is the heatmap generated.

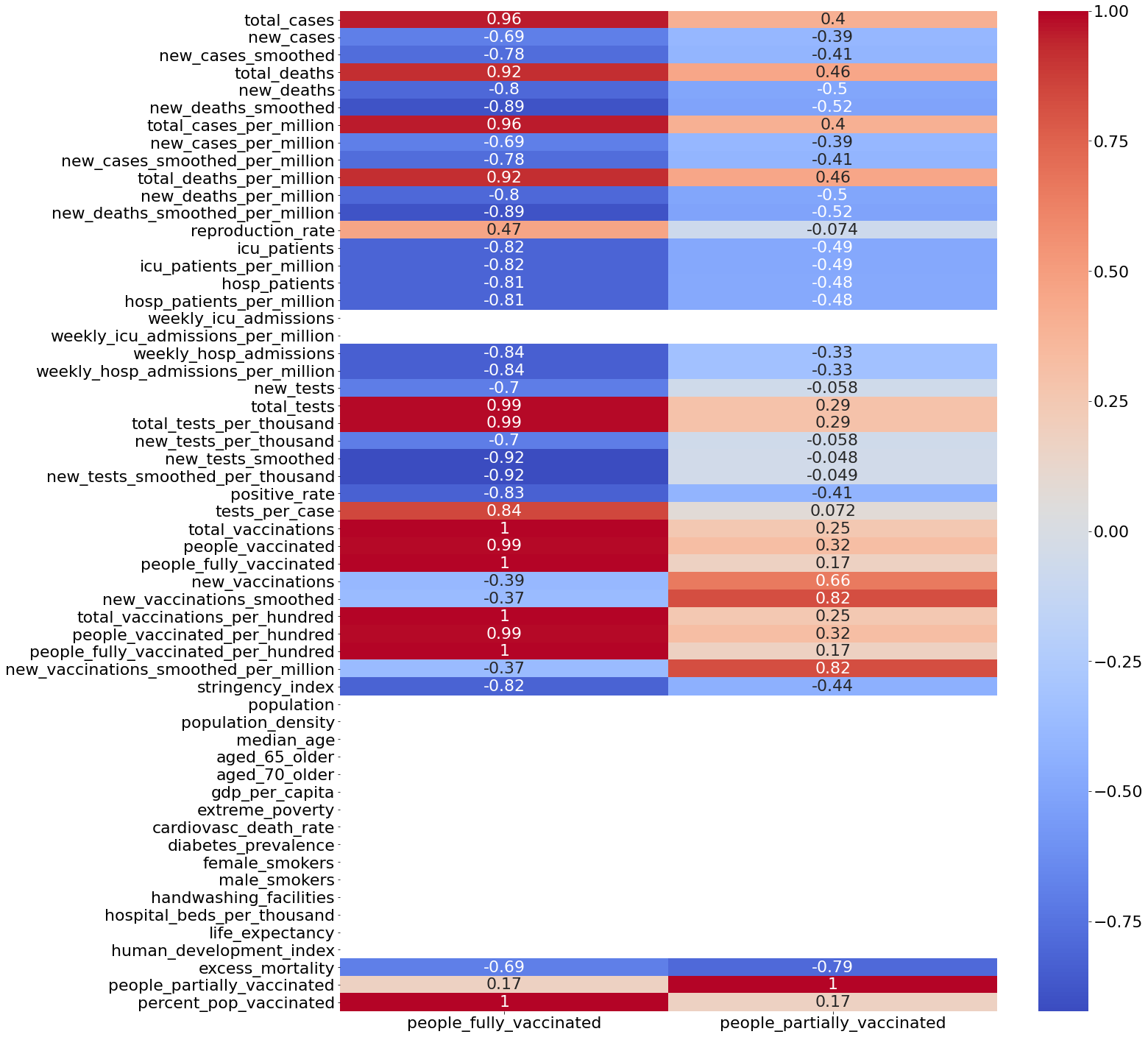
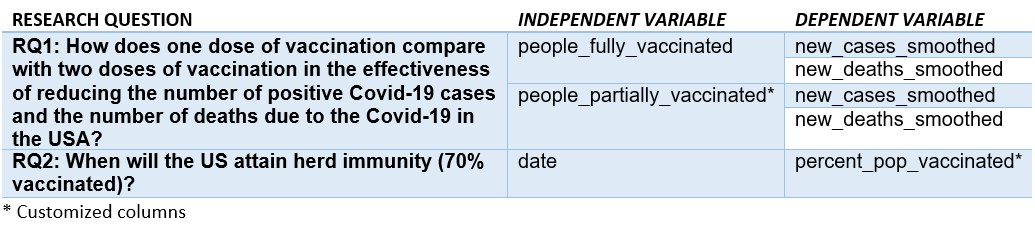


Figure 2: Heatmap

## Data Preparation & Feature Engineering

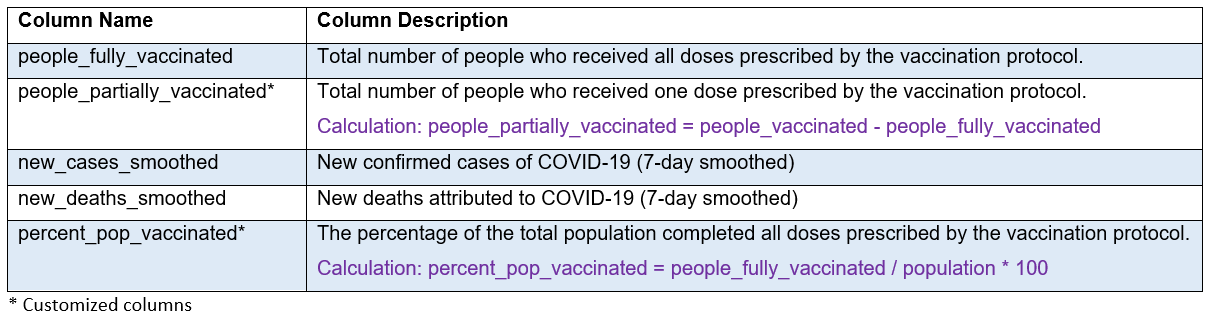
The data preparation process involves cleaning and transforming the raw data into the desired format, and performing feature engineering for processing. Through EDA, a number of columns shown in table 2 below are to be used in this study.

Table 2: Independent & dependent variables (columns)



Two customized columns have been created to satisfy the needs of our research questions namely people\_partially\_vaccinated and percent\_pop\_vaccinated. The calculations are included in the column description.

Table 3: Descriptions of the columns used



Furthermore, new\_cases\_smoothed and new\_deaths\_smoothed columns are used instead of new\_cases and new\_deaths. They are already present in the dataset. The smoothed data refers to the value obtained by averaging the values from the past 7 days. Smoothed data is used to eliminate the noises in the dataset so that the trends can be easily identified (Irizarry, 2021). The diagram below shows how the data becomes cleaner after removing the noises by smoothing.

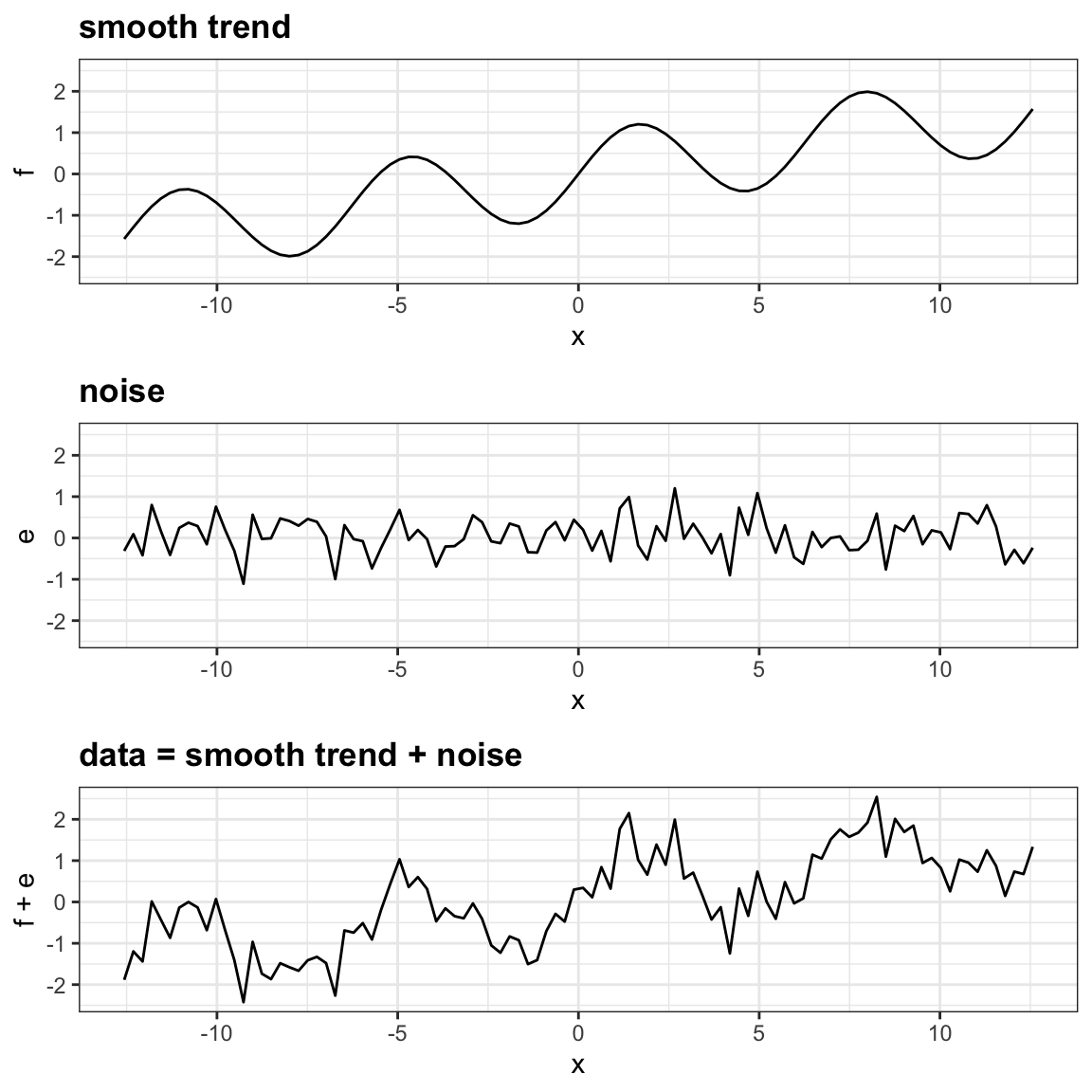


Figure 3: Data smoothing

The study will focus on the data of the United States of America as it is inappropriate to use every country in the world due to different population sizes and different situations in different countries. Figure 4 shows that the number of records available is 189 and 6 columns/features are kept for modelling. Figure 5 below shows a scatter plot of many trend lines due to the different population sizes of each country.

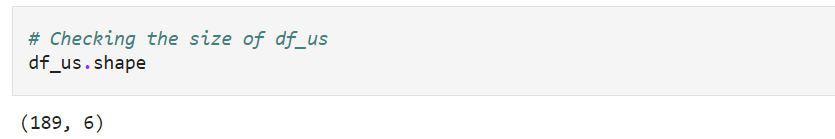


Figure 4: The shape of df\_us

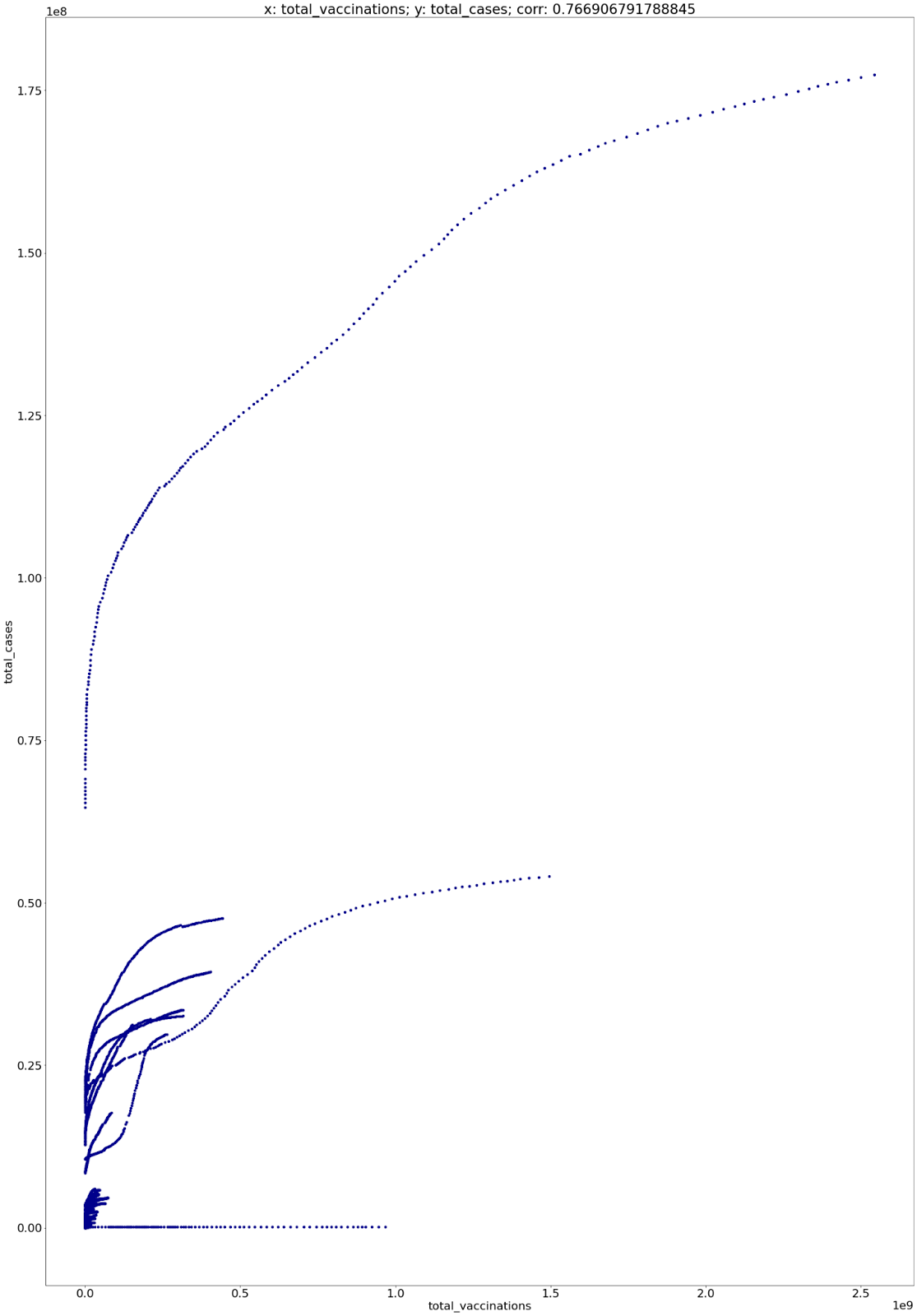


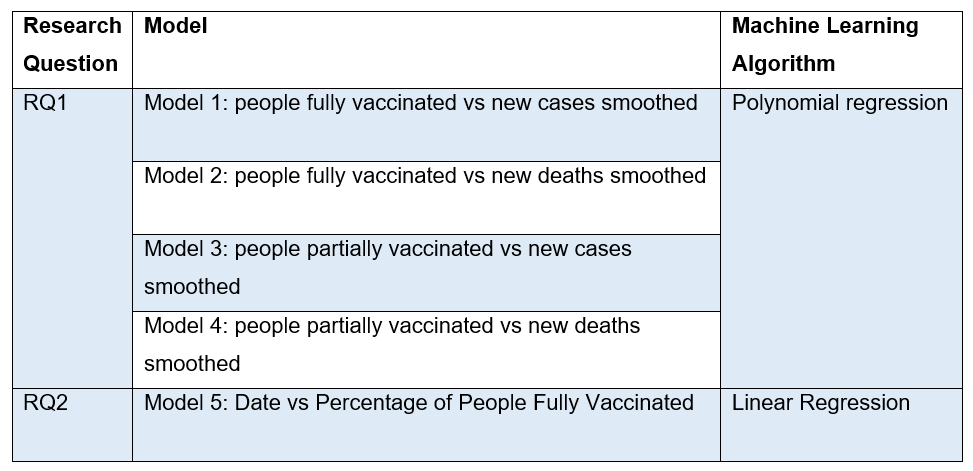
Figure 5: Sample scatter plot

Two functions are defined to perform the data cleaning process, namely “display\_missing\_val\_dates” and “group\_missing\_val\_by\_month”. These are used to discover missing values hidden in the data. From the discovery, the missing values of the variables "people\_fully\_vaccinated", "people\_partially\_vaccinated", and "percent\_pop\_vaccinated" are from the beginning of January 2020 until 18 January 2021, totalling up to 365 each. It indicates that the vaccination only starts taking place in early 2021. Therefore, all data before the date 2021-01-18 have to be discarded. Besides, four other missing values were found after the date 2021-01-18. . Since there are only four missing data, case deletion to handle the missing data whereby those columns are eliminated from the dataset as it is the easiest way and deleting a few data will not affect the overall performance of the model. Both "new\_cases\_smoothed" and "new\_deaths\_smoothed" only has 6 missing values, which is the first six days of the record that have been deleted earlier. Therefore, the actual data range used is between 2021-01-18 and 2021-07-31.

## Modelling

Machine learning is used in modelling. The data have been randomly split into a train set and a test set in the 7:3 ratio. The machine learning algorithms used for modelling are polynomial regression and linear regression. The polynomial regression of power 4 (quartic) is used in RQ1 as the graphs have non-linear trends while linear regression is used in RQ2 as the graph presents a positive linear relationship between the date and the percentage of people vaccinated. Table 4 below summarizes the models used to investigate each research question and the machine learning algorithm used.

Table 4: Machine learning algorithm used



The following equations are used in modelling:

RQ1:

RQ2:

## Platform

Jupyter Notebook is used from the data understanding process right up to the model evaluation process. It’s a free open-source platform that supports Python and helps document the entire data science process. Jupyter allows for codes to be run block by block which makes it easier to highlight possible bugs or errors that might be present.

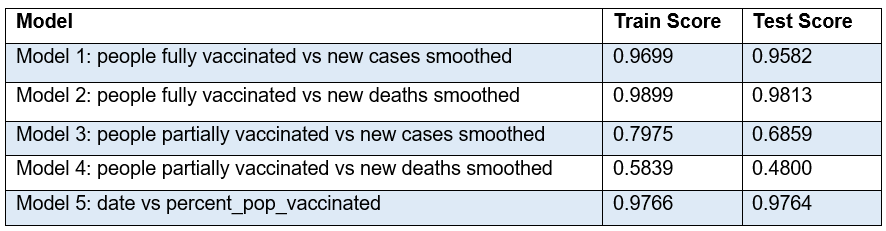
The spyder platform is used to develop the user interface as it supports the Streamlit library. Streamlit is an open-source application framework made to develop machine learning and data science applications using Python. It allows for a seamless transition of the inputted script and converts it into shareable and visually appealing web applications, making it easier for those who are not accustomed to data science to view the data patterns or trends through visualizations to better understand the data and make better decisions.

# Results

## Evaluation

Model 1 graphs the number of people fully vaccinated against the new positive cases. It has obtained an accuracy of 0.9699 and 0.9582 on the training set and testing set respectively. Model 2 graphs the number of people fully vaccinated against the new death cases. It has obtained an accuracy of 0.9899 and 0.9813 on the training set and testing set respectively. Model 3 graphs the number of people partially vaccinated against the new positive cases. It has obtained an accuracy of 0.7975 and 0.6859 on the training set and testing set respectively. Model 4 graphs the number of people partially vaccinated against the new deaths cases. It has obtained an accuracy of 0.5839 and 0.4800 on the training set and testing set respectively. Model 5 graphs the date against the percentage of the population who received 2 doses of vaccinations. It has obtained an accuracy of 0.9766 and 0.9764 on the training set and testing set respectively. Table 5 below summarizes the accuracy scores for all models.

Table 5: Model accuracy scores



The figures below show models 1 to 5. The scatter plots are the actual values while the red predicted line is the predicted regression line.

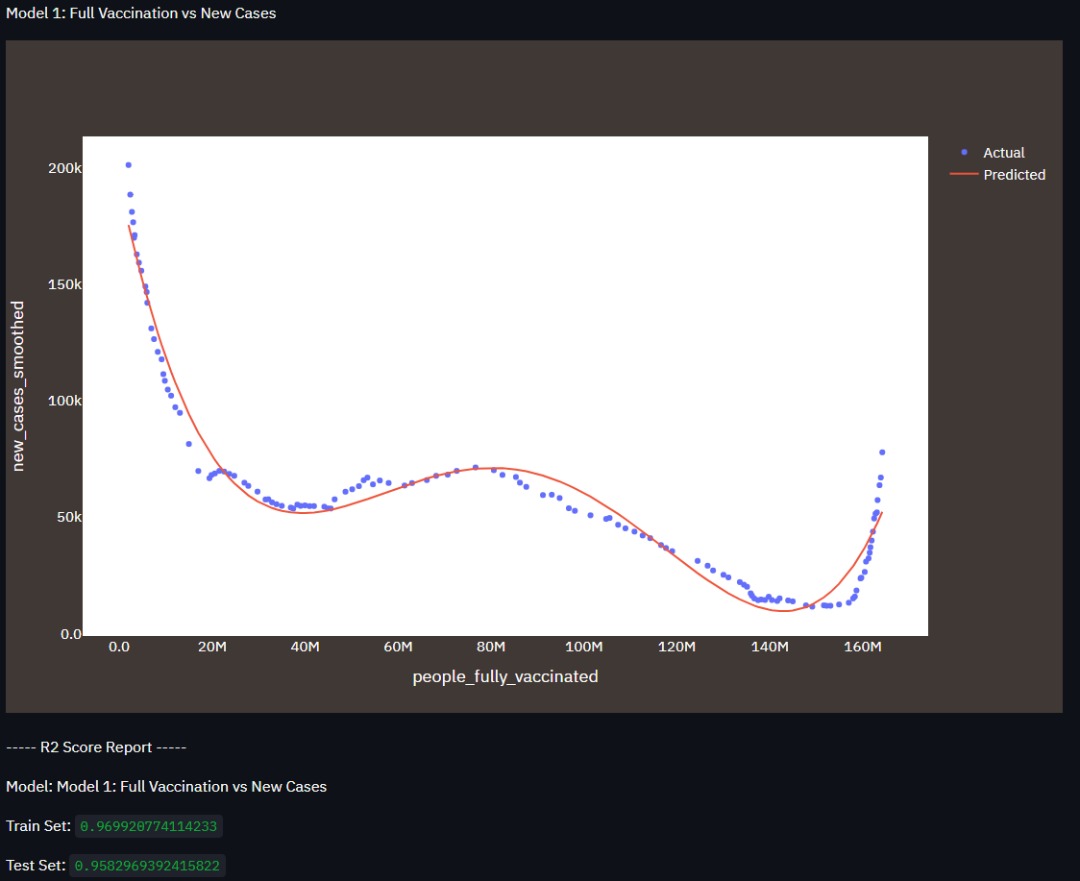


Figure 6: Model 1 scatter plot and regression curve

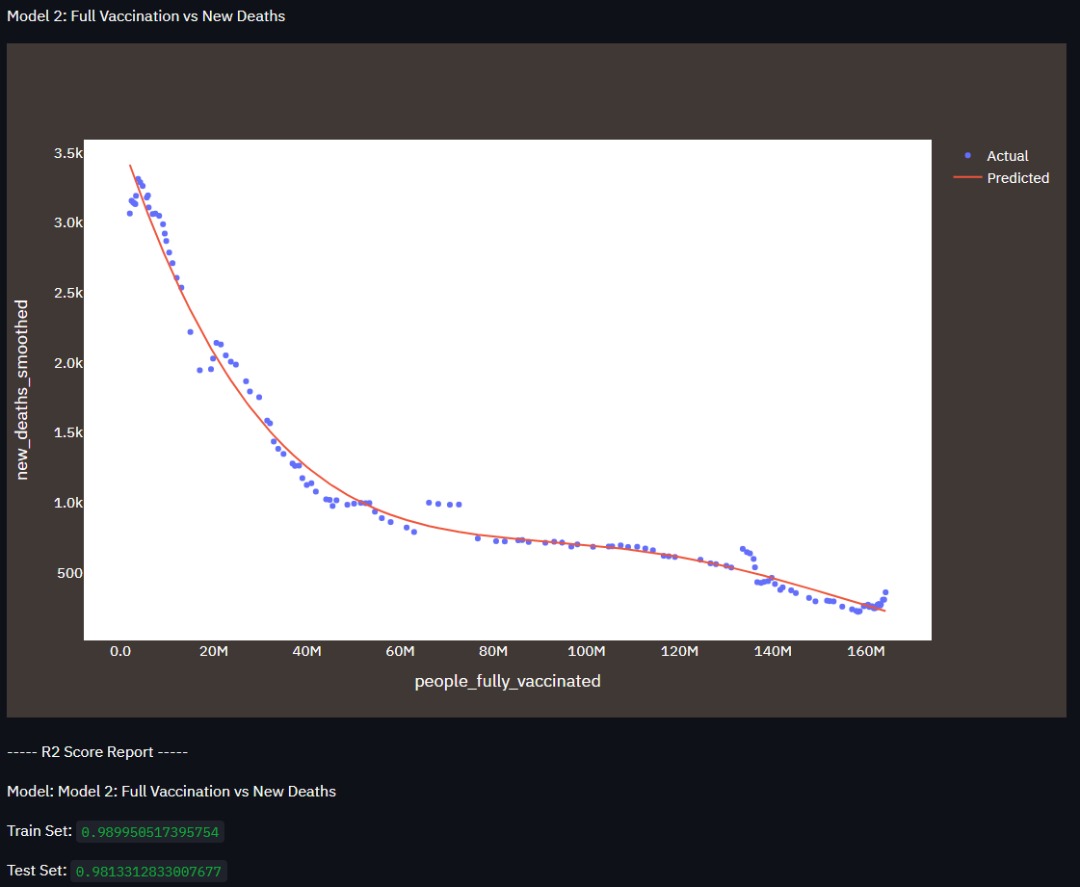


Figure 7: Model 2 scatter plot and regression curve



Figure 8: Model 3 scatter plot and regression curve

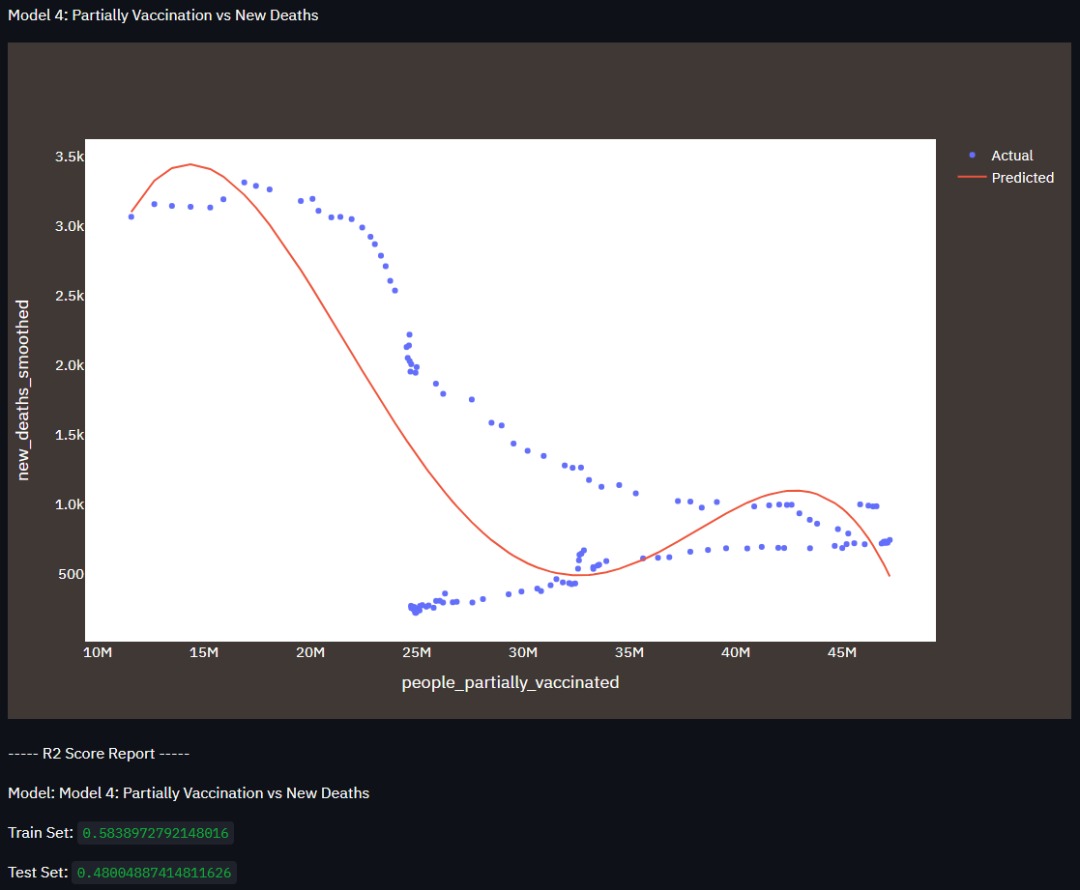


Figure 9: Model 4 scatter plot and regression curve

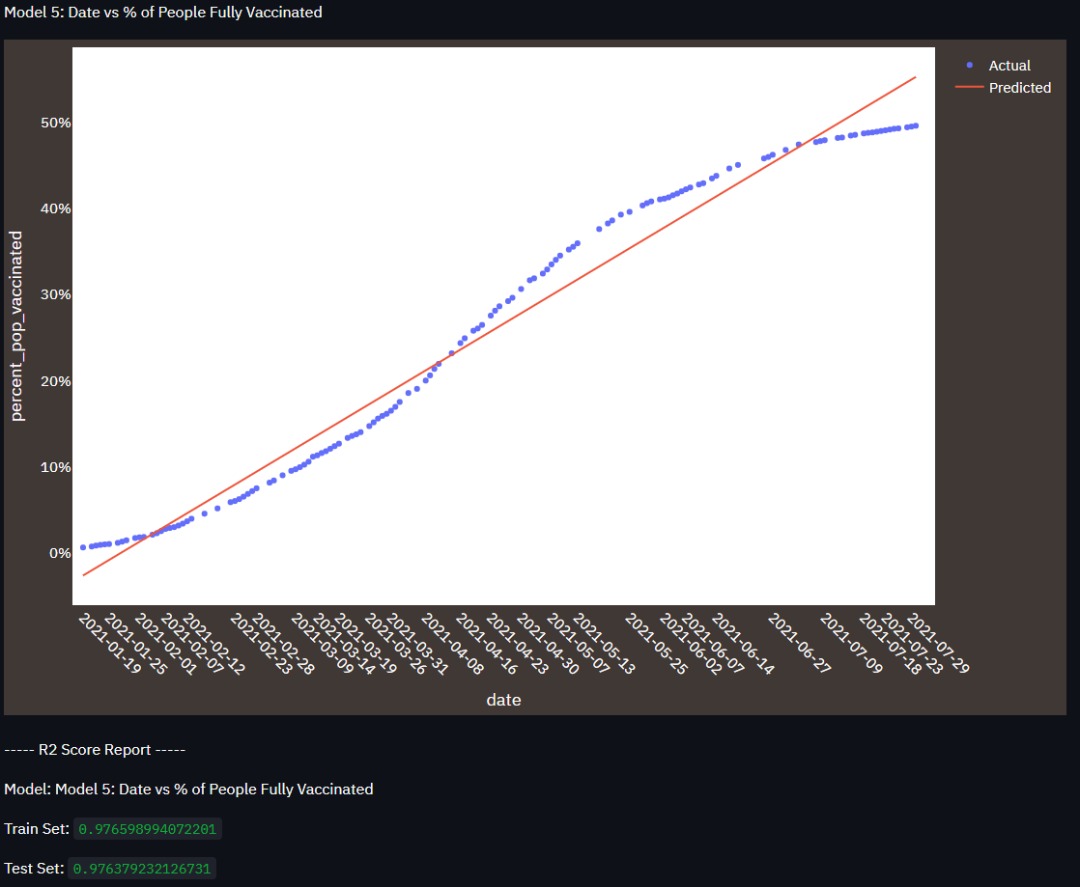


Figure 10: Model 5 scatter plot and regression curve

The figures below show comparisons between models 1 and 3 as well as models 2 and 4.

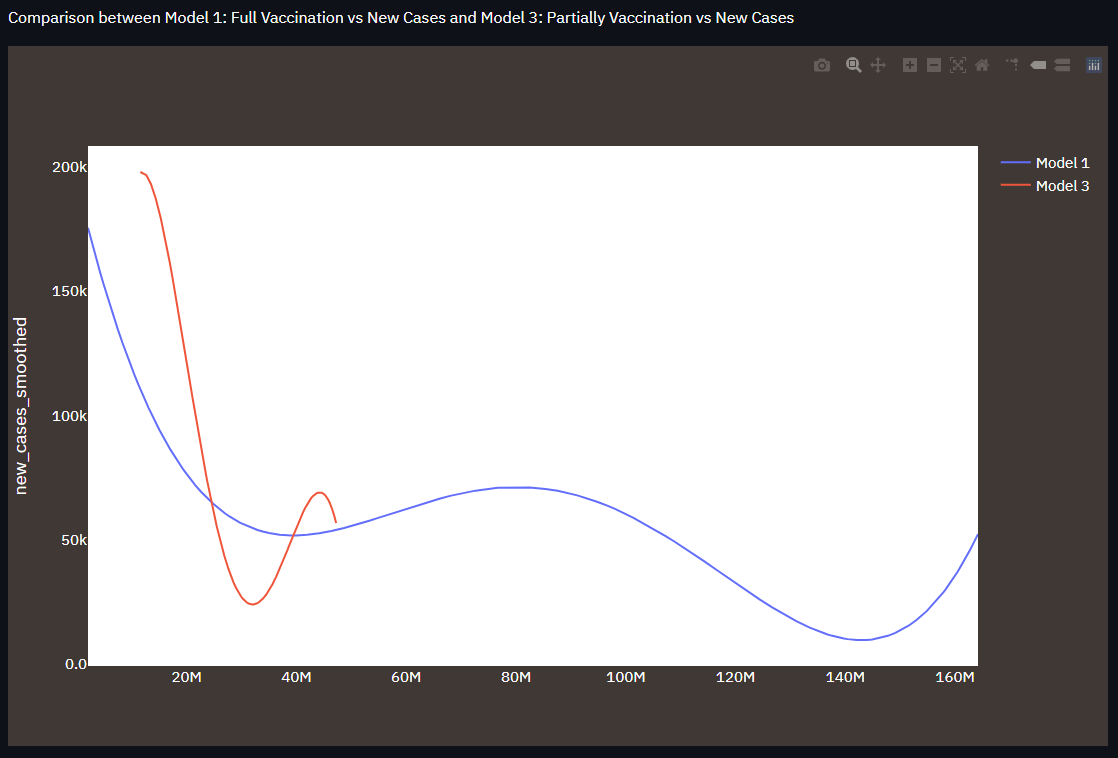


Figure 11: Comparison graph between model 1 and model 3

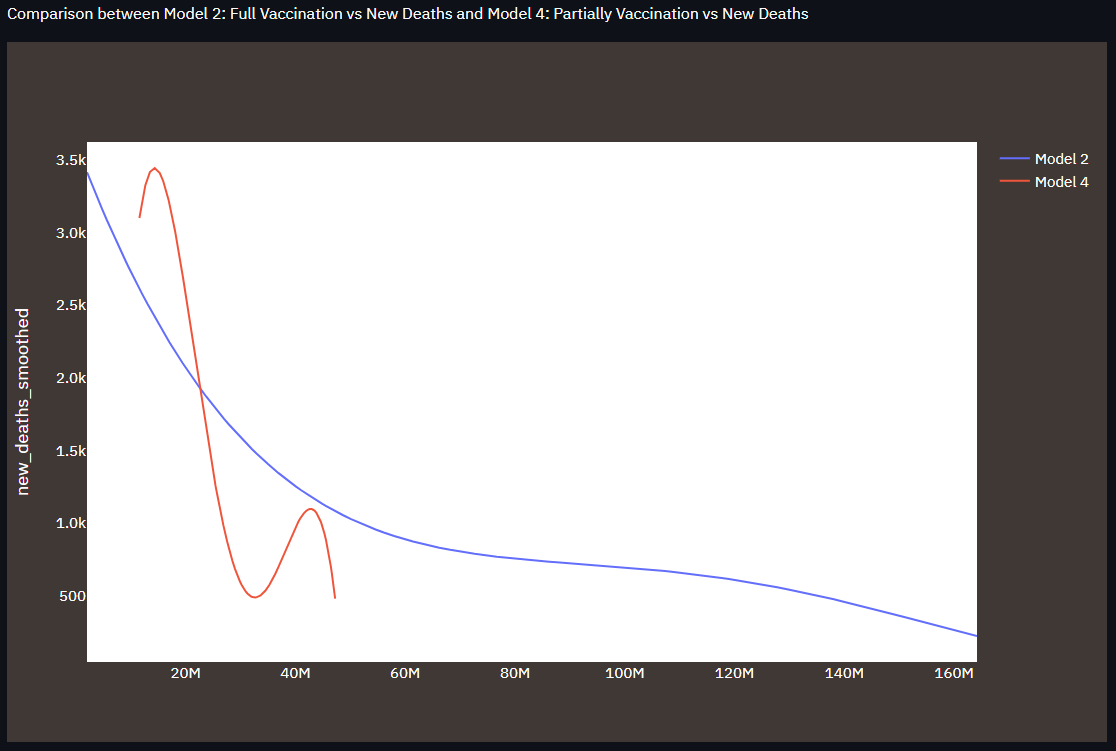


Figure 12: Comparison graph between model 2 and model 4

## Samples Screenshot of the Application

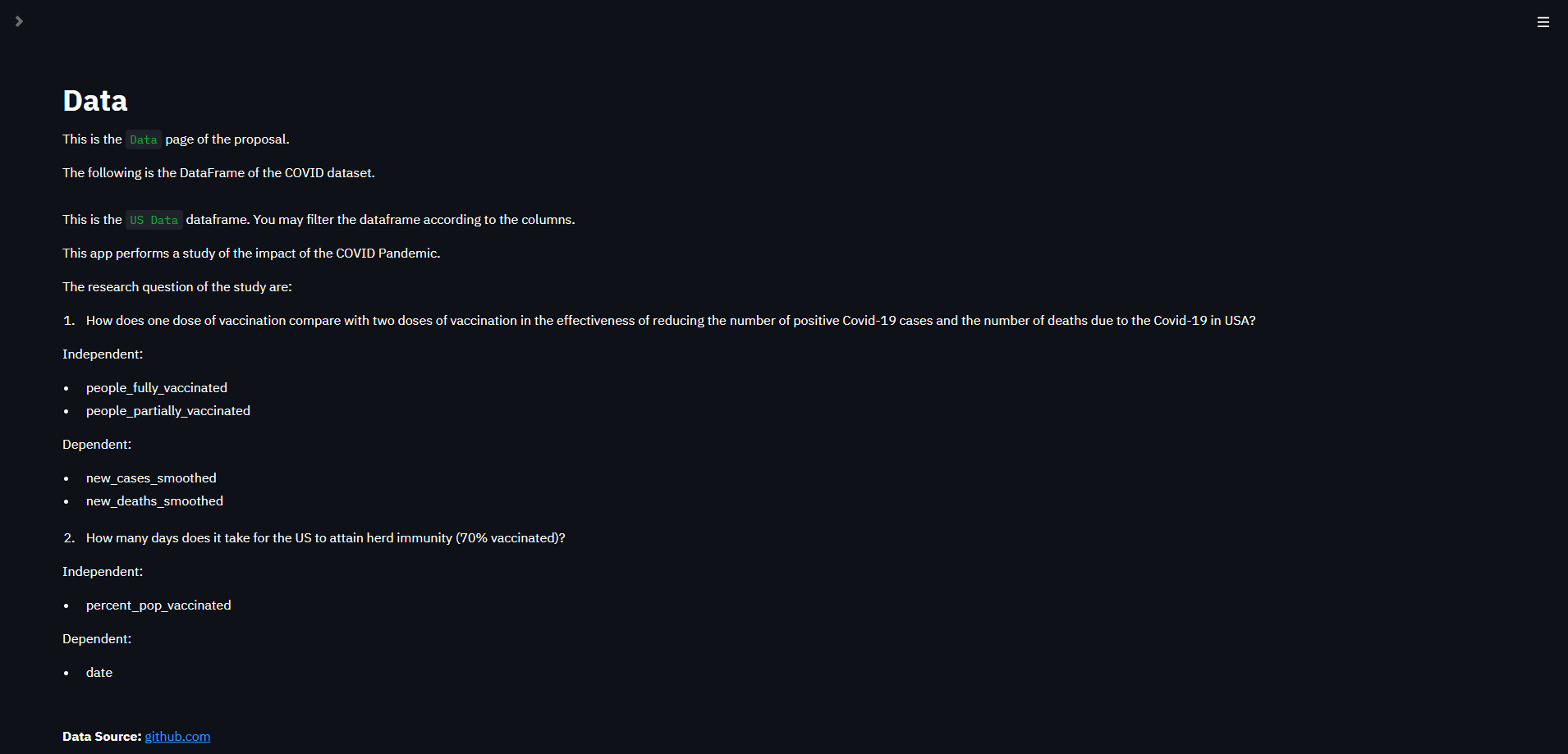


Figure 13: The introduction Data page of the Streamlit app

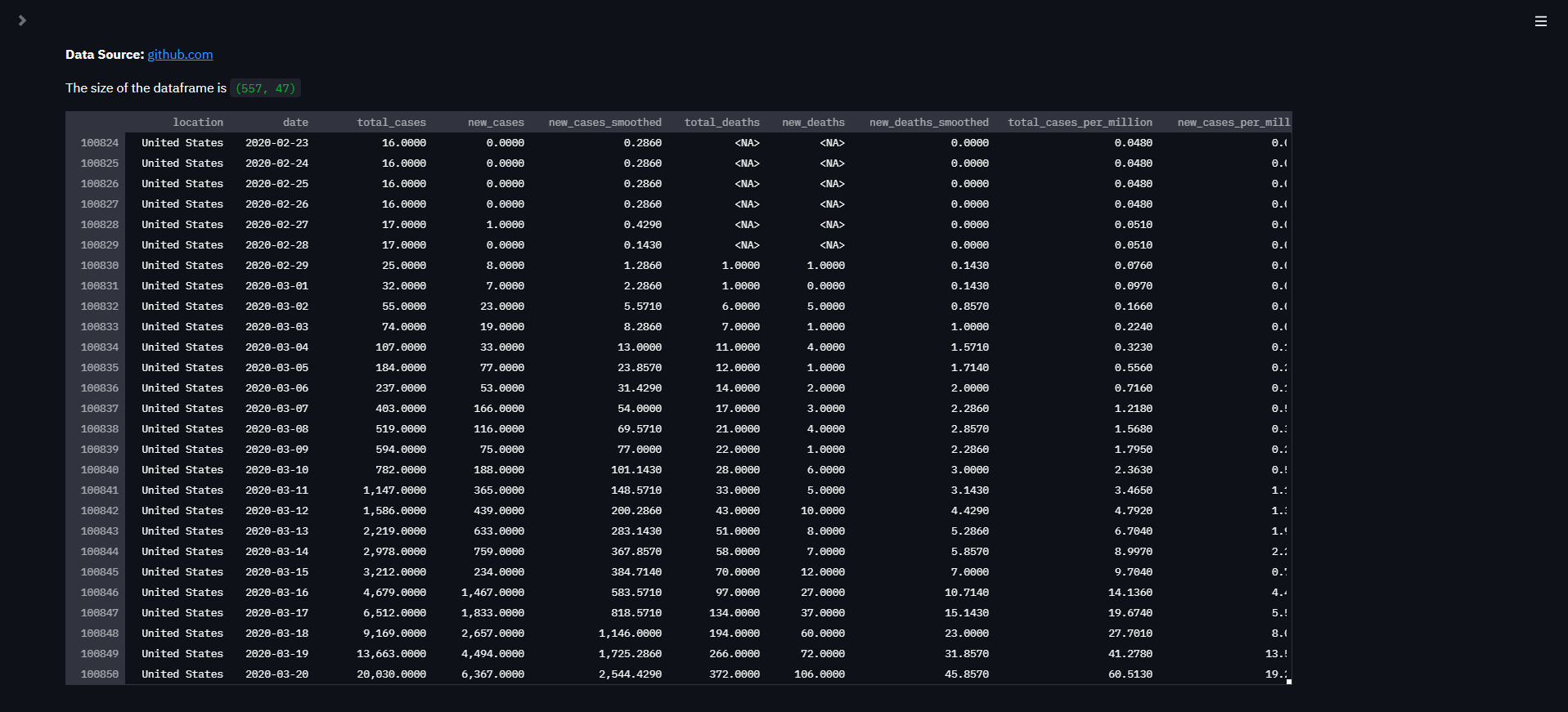


Figure 14: Dataframe displaying our dataset

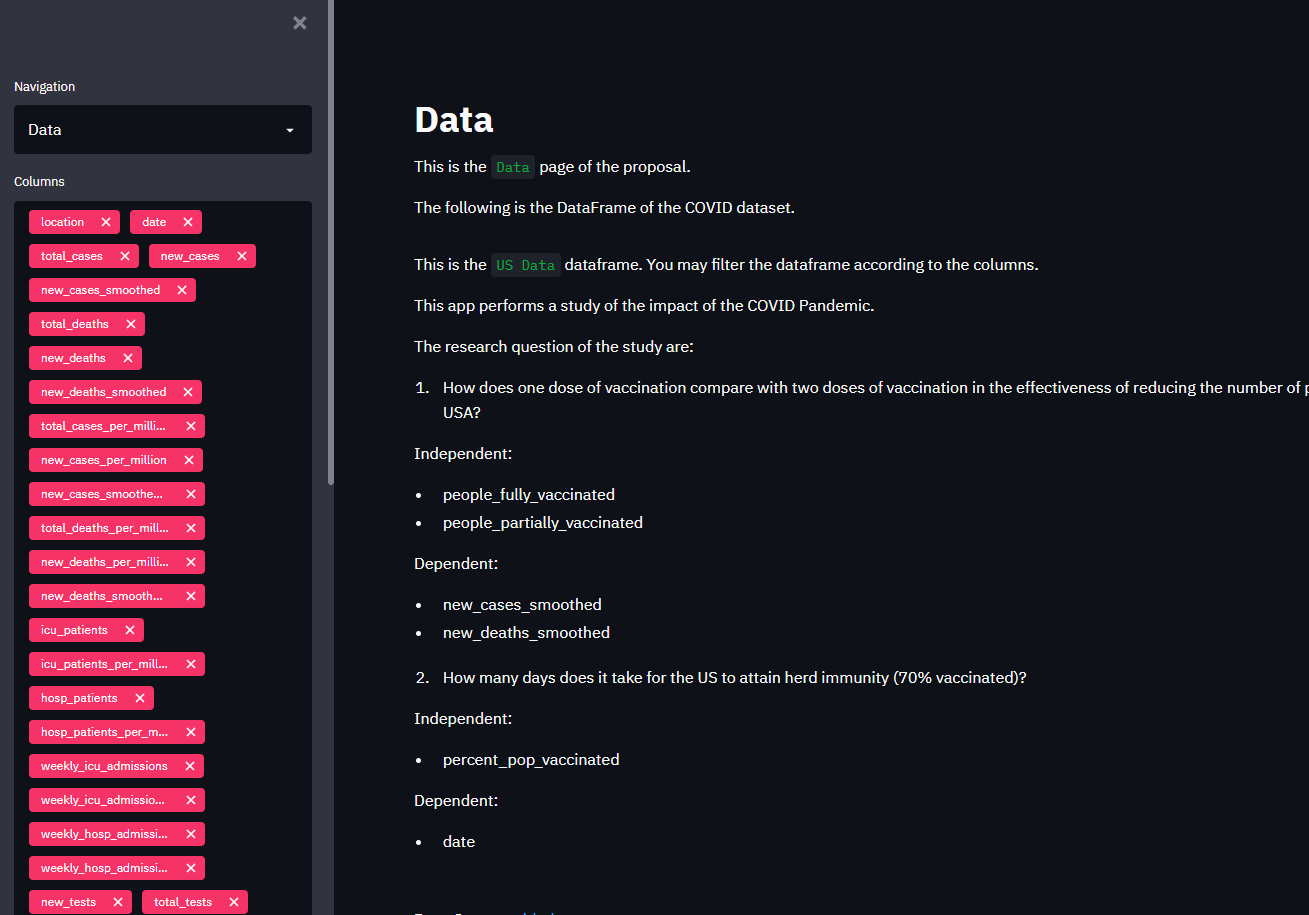


Figure 15: The sidebar to change page and choose columns in the dataframe

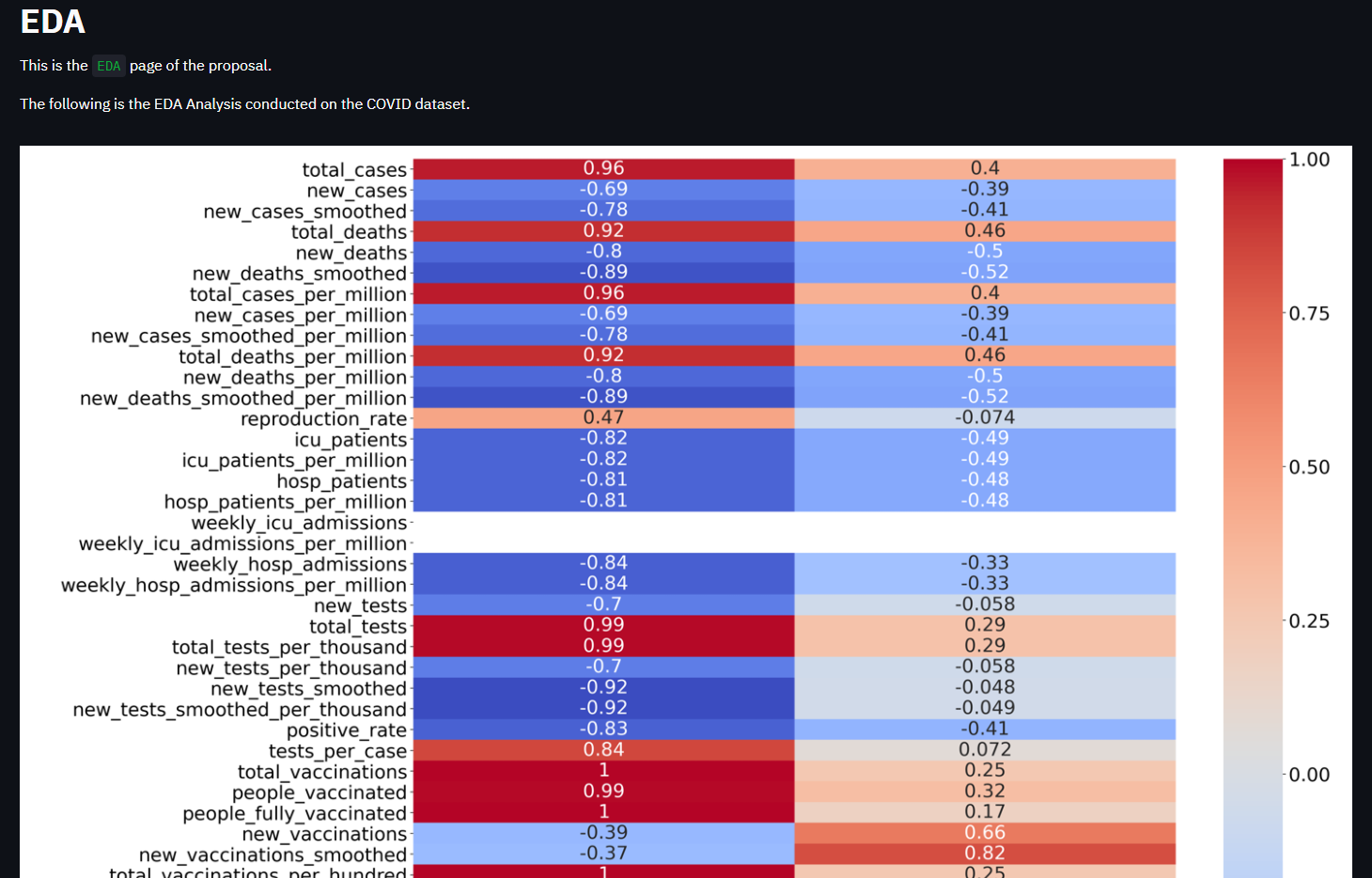


Figure 16: The EDA Page. The heatmap of correlation between independent variables and other variables.

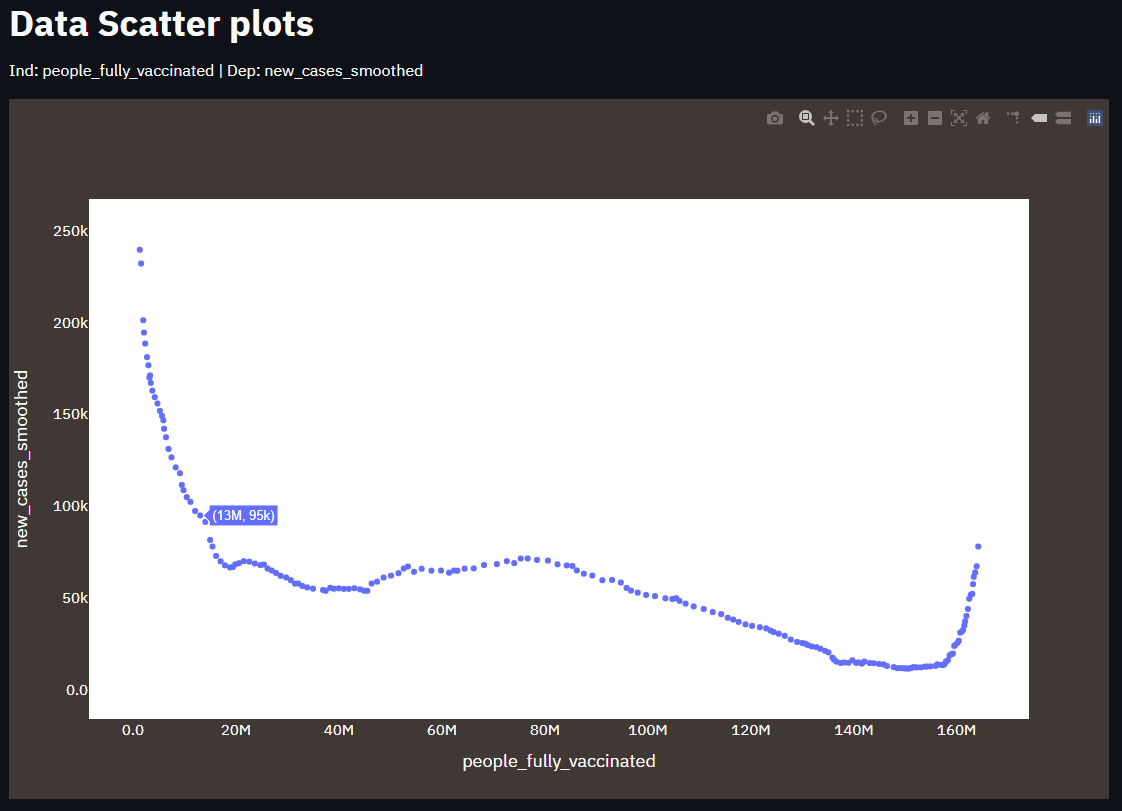


Figure 17: People\_fully\_vaccinated against new\_cases\_smoothed.

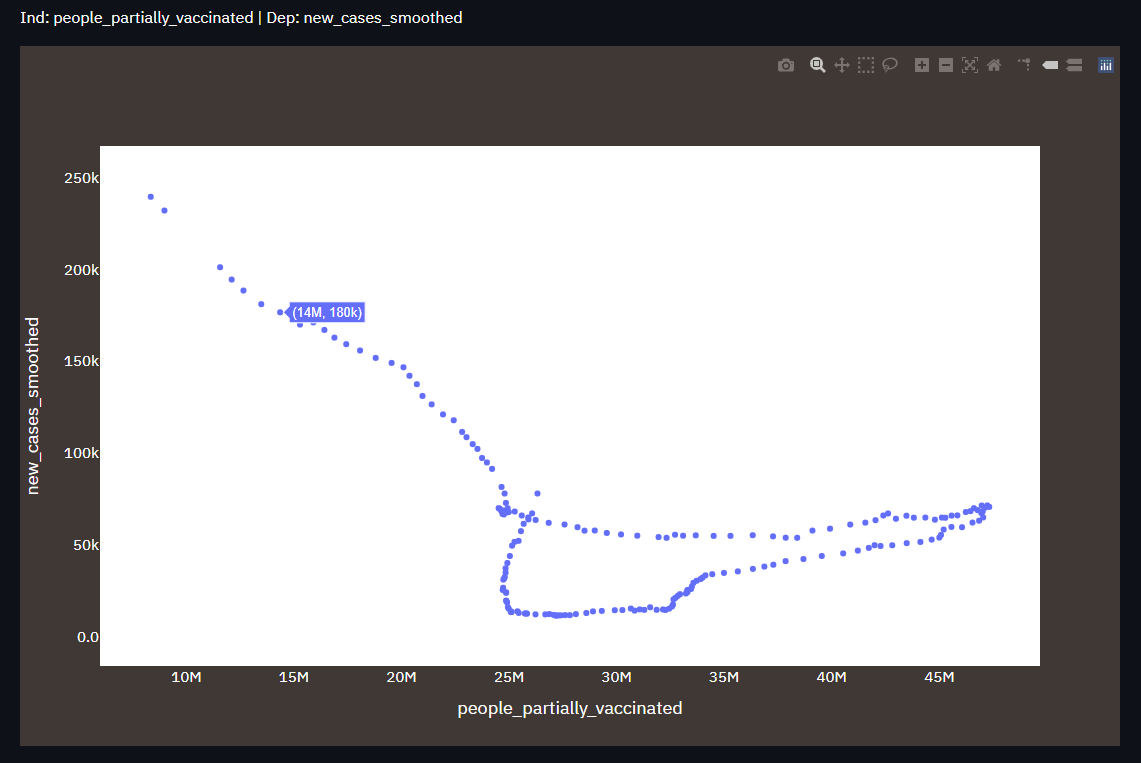


Figure 18: people\_partially vaccinated against new\_cases\_smoothed

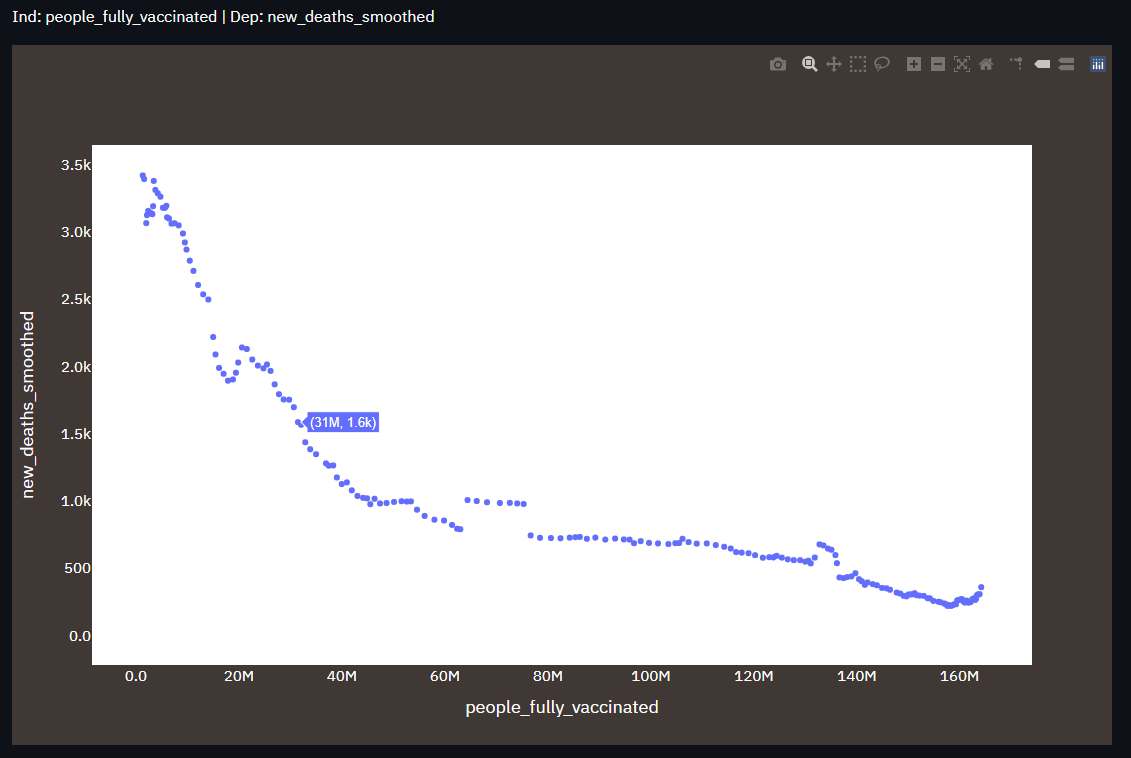


Figure 19: for people\_fully\_vaccinated against new\_deaths\_smoothed.

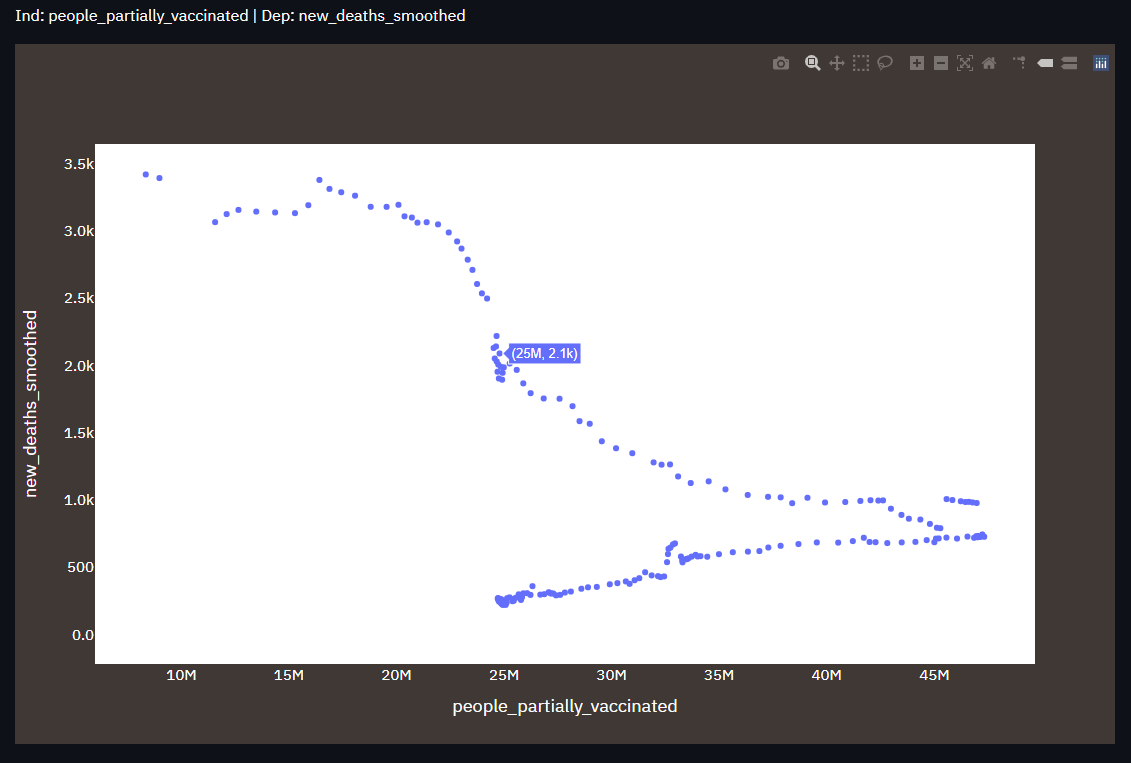


Figure 20: people\_partially\_vaccinated against new\_deaths\_smoothed.

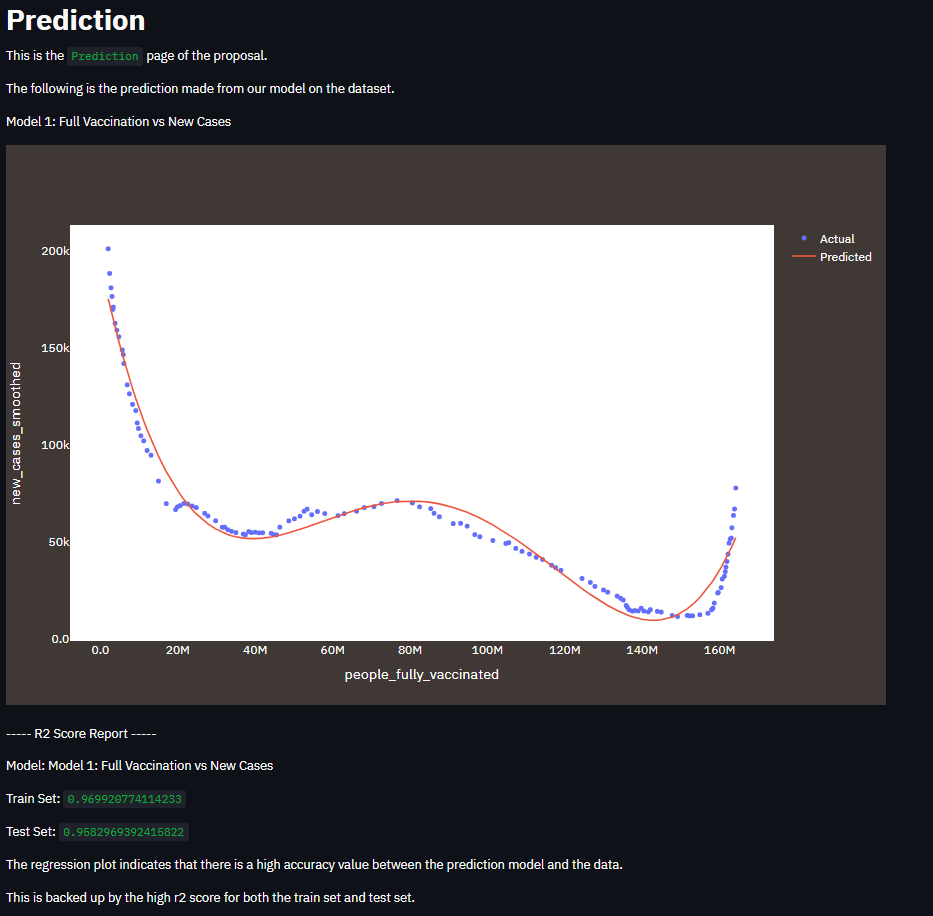


Figure 21: Prediction Page, Prediction model between people\_full\_vaccinated against new\_cases\_smoothed, R2 Score Report.

* The graph allows users to select charts in the legends.



Figure 22: User input prediction for new\_cases\_smoothed based on people\_fully\_vaccinated.



Figure 23: Prediction model between people\_full\_vaccinated against new\_deaths\_smoothed, R2 Score Report.

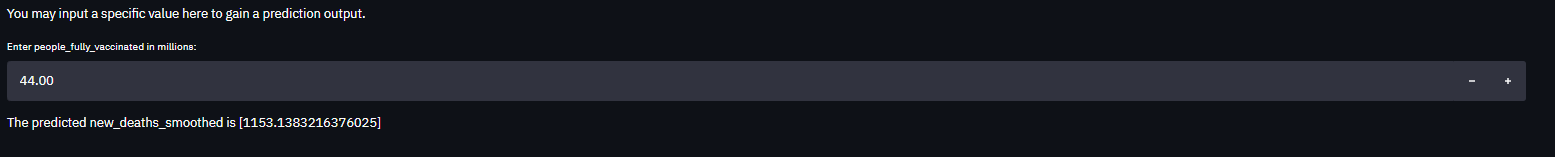


Figure 24: User input prediction for new\_deaths\_smoothed based on people\_fully\_vaccinated.

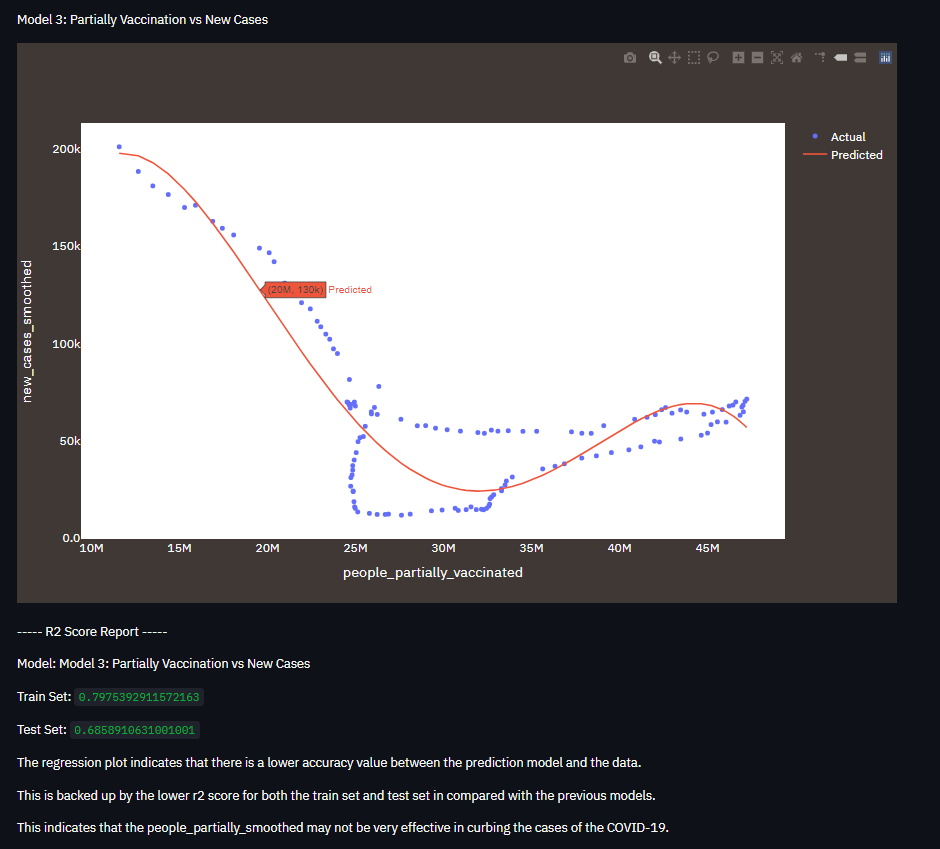


Figure 25: Prediction model between people\_partially\_vaccinated against new\_cases\_smoothed, R2 Score Report.

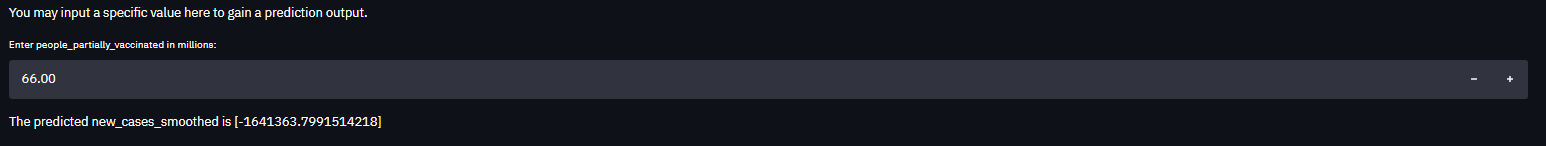


Figure 26: User input prediction for new\_cases\_smoothed based on people\_partially\_vaccinated.

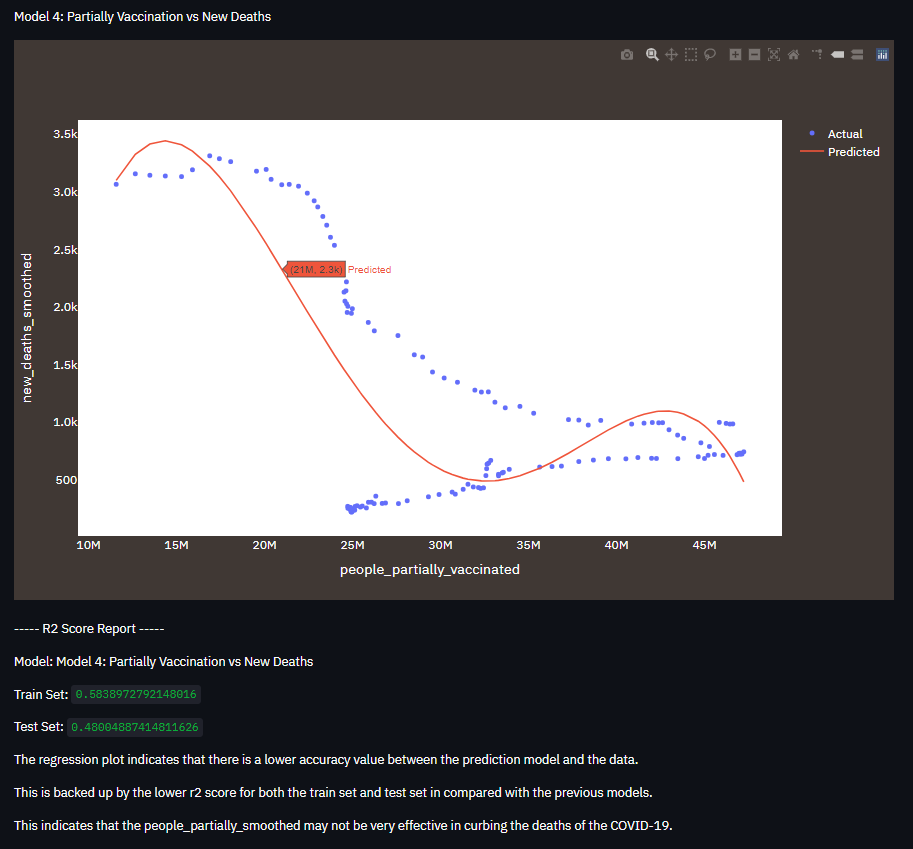


Figure 27: Prediction model between people\_partially\_vaccinated against new\_deaths\_smoothed, R2 Score Report.



Figure 28: User input prediction for new\_deaths\_smoothed based on people\_partially\_vaccinated.

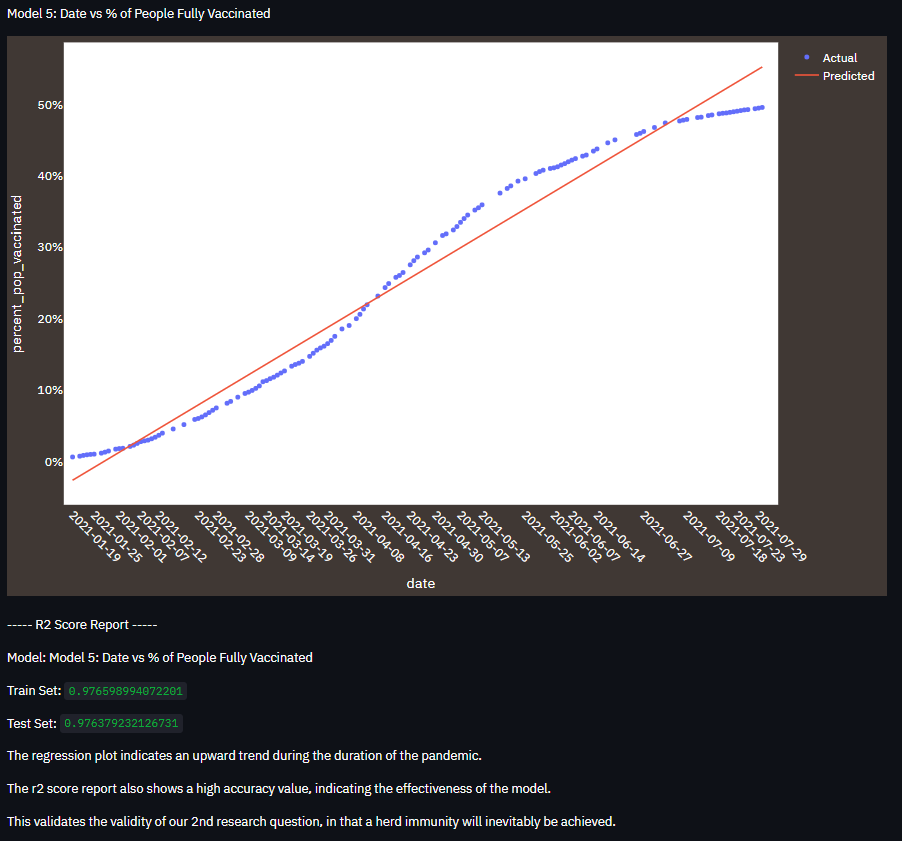


Figure 29: Prediction model between date against people\_pop\_vaccinated, R2 Score Report.

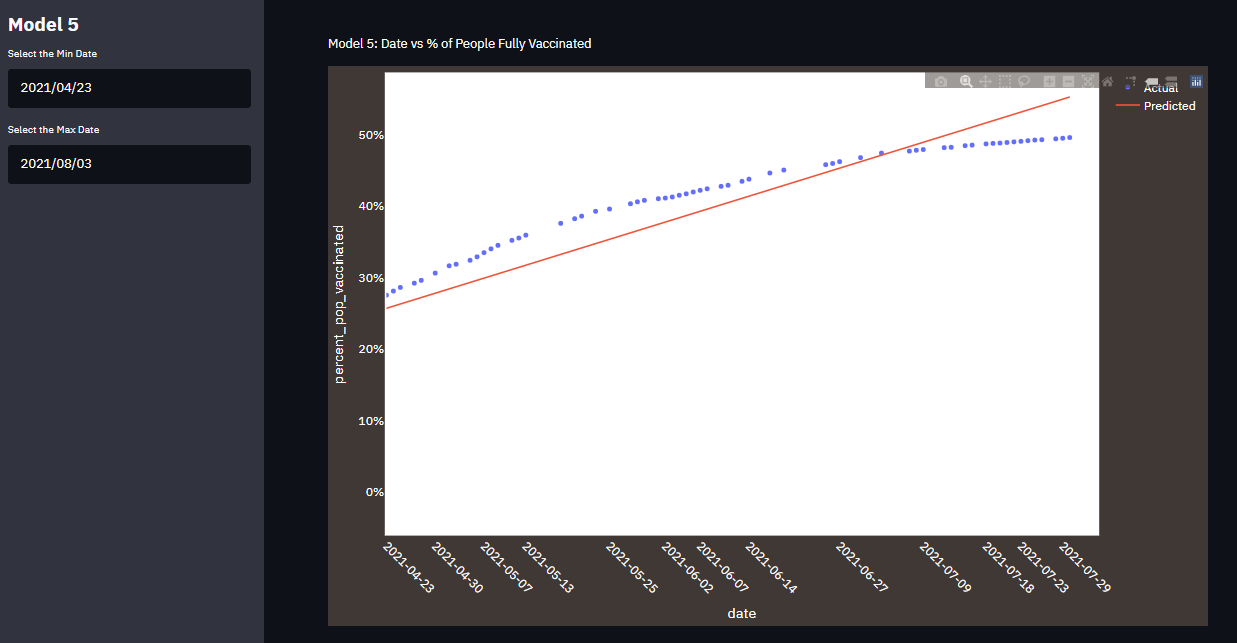


Figure 30: Date picker for model 5

* Date input is limited to the range of the dataset, 12-1-2021 to 3/8/2021.

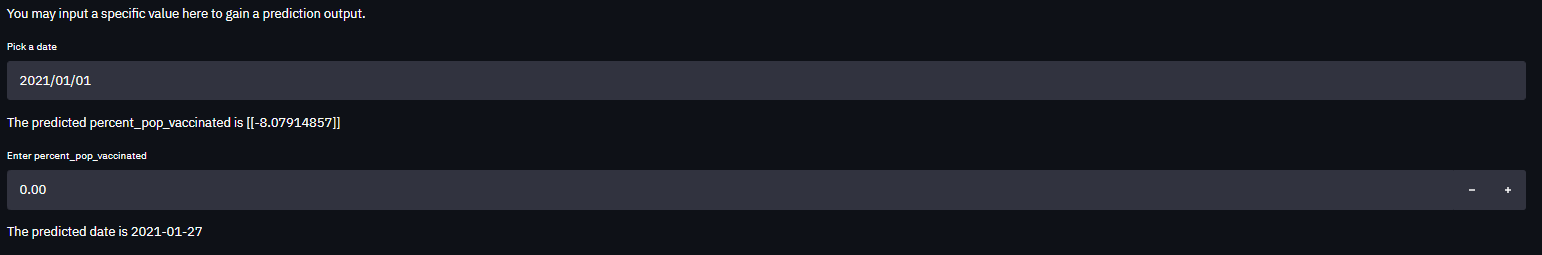


Figure 31: User input prediction for percent\_pop\_vaccinated based on date, and date prediction based on percent\_pop\_vaccinated.

* Date input is limited to 1-1-2021 to 1-1-2023

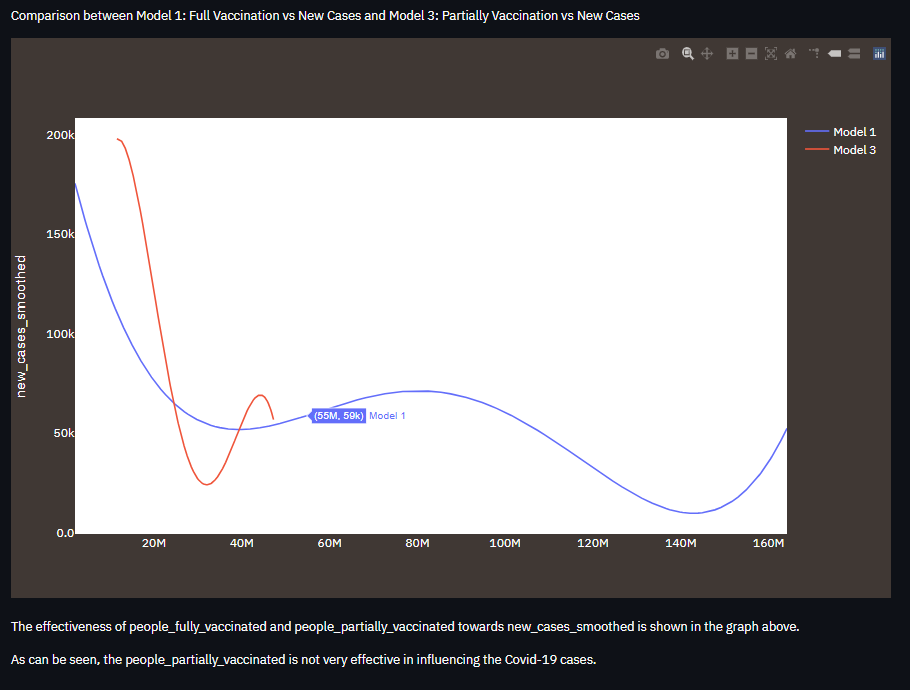


Figure 32: Regression line comparison between model 1 and model 3.

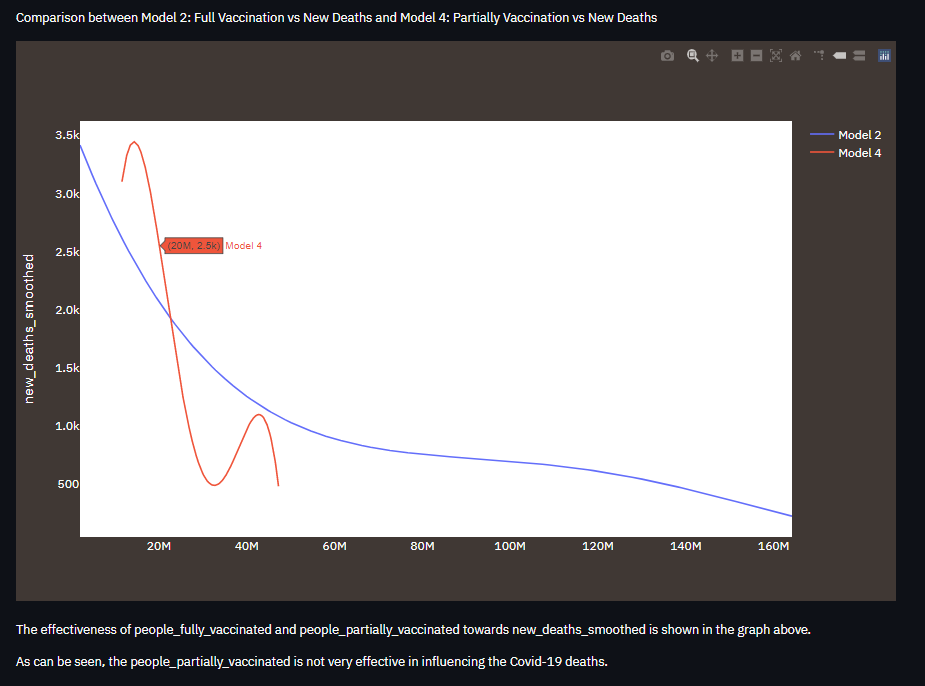


Figure 33: Regression line comparison between model 2 and model 4.

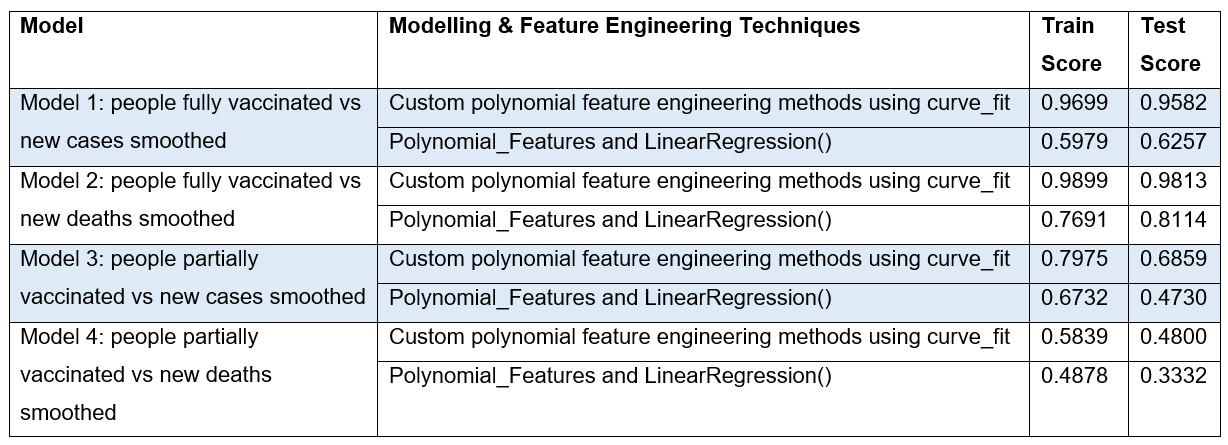
# Discussion

## Changes in Tools and Methods Used

A custom feature engineering technique has been used to apply polynomial regression to the models under the RQ1. In this case, a quartic function is defined and the scipy.optimize.curve\_fit function is used. The scipy.optimize.curve\_fit is a library that uses a nonlinear least square method to fit a function by returning a series of optimized coefficients for the curve. It is typically used to sketch best-fit curves for scatter plot graphs with a nonlinear trend. The curve\_fit parameter requires 3 variable inputs, including the polynomial function, the x-array, and the y-array. The generated optimized curve coefficients and the x\_train values will be used in the quartic function to train and predict the model.

The reason for using a customized feature engineering technique rather than using the conventional PolynomialFeatures together with the LinearRegression modules in scikit-learn is because the conventional way does not train our model well. Table 6 below compares the result of the train and test score using the conventional method and the custom method.

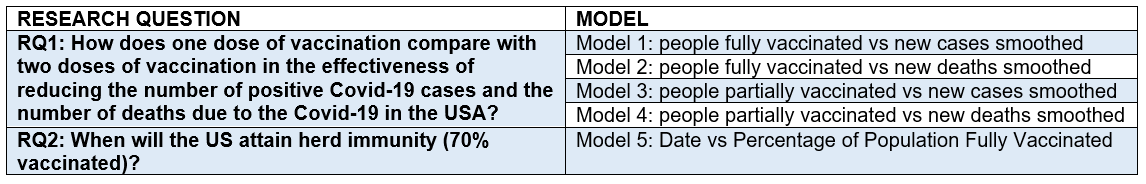
Table 6: PolynomialFeatures vs custom features



As observed above, our custom features performed better compared to the conventional polynomial regression model.

## Answer to the Research Question

Table 7: Models used in each research question.



### Research Question 1 (RQ1)

Models 1 and 2 show an exceptionally high train and test score that is over 95% accuracy. Model 3 shows a lower but acceptable accuracy train and test scores, 0.7975 and 0.6859 respectively. However, model 4 has a low accuracy train and test score that is below 0.5.

Models 3 and 4 faced serious data fluctuations to the point that they almost looked like a one-to-many relation (see figure 12). A weird trend is observed in figures 10 and 11 whereby those partially vaccinated regression lines in red show a greater decrease in the number of new cases and new deaths as compared to the fully vaccinated. Besides, the regression lines of models 3 and 4 came to a stop at around 50 million people as compared to 160 million people in models 1 and 2. This is because those who have received their first dose of vaccine will have their second dose appointment scheduled in approximately three weeks after their first dose, therefore the number of people partially vaccinated is lesser than the number of people fully vaccinated.

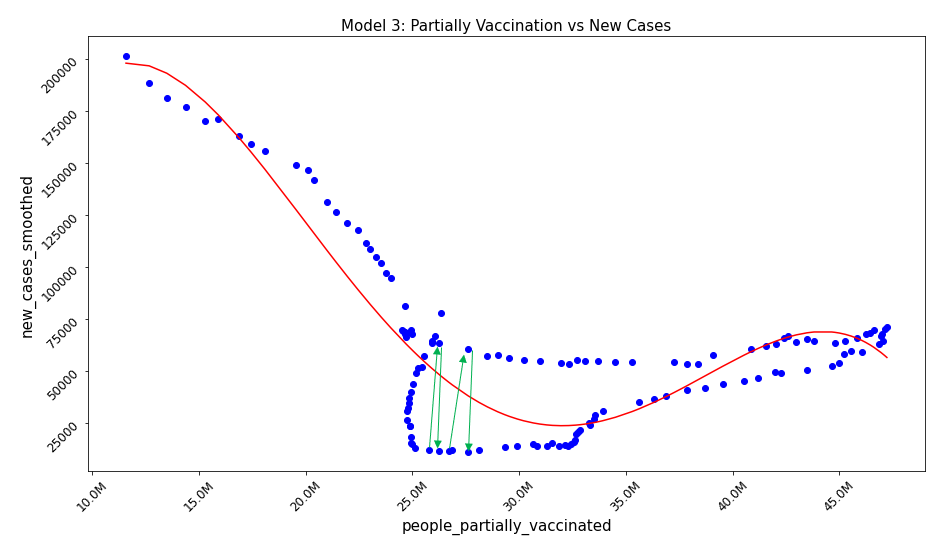


Figure 34: Data fluctuation

It is difficult to make a direct comparison between the number of people fully vaccinated (models 1 and 2) and the number of people partially vaccinated (models 3 and 4) in terms of their effectiveness in controlling the new cases and new deaths. However, we can safely say that the number of people fully vaccinated is certainly more reliable data to investigate the effectiveness of vaccination in controlling the new cases (model 1) and new deaths (model 2) as the data are more stable. Model 1 does fluctuate at some point but it is not as severe as compared to model 3. The fluctuation could be caused by some other extraneous variables that have some influence on the new cases, like mutated delta variant of the covid-19 virus, which is more contagious; and the loosen of lockdown restrictions. The death cases under model 1 show a constant decreasing trend without any fluctuation. It shows that although completing two doses of vaccination may not fully prevent one from getting infected by the covid-19 virus, however, it is effective in reducing the severity of the disease since it reduces the number of deaths. In conclusion, there is no strong evidence to claim that the number of people fully vaccinated is more effective than the number of people partially vaccinated in reducing the new cases and death cases; however, the number of people fully vaccinated is a more reliable data to investigate the effectiveness of vaccines.

### Research Question 2 (RQ2)

As for RQ2, model 5 has a very high train and test accuracy score of 0.9766 and 0.9764 respectively. Figure 35 below shows a prediction made by the system in which 70% of the population will be fully vaccinated by 2021-09-16. Therefore, according to our model, the hypothesis claiming that the USA can attain herd immunity by winter 2021 is valid as September is the fall season.

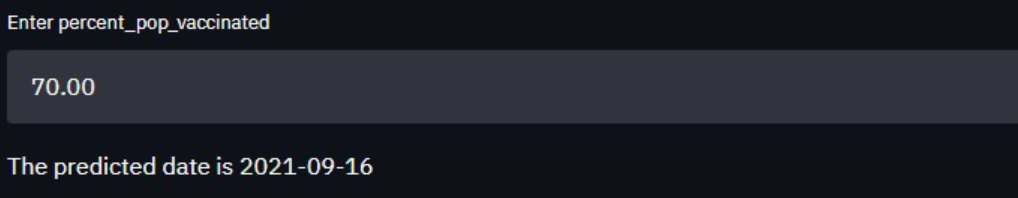


Figure 35: Model 5 herd immunity prediction

# Conclusion

## Implications of the Findings to Machine Learning Area

The use of the *scipy.optimize.curve\_fit* to get the best-fit curve coefficients and train the model by fitting the training dataset with the coefficients could be a new potential way of modelling polynomial regression when the traditional method of using *sklearn.preprocessing.PolynomialFeatures* and *sklearn.linear\_model.LinearRegression* does not produce the desired result. The curve\_fit package is normally used to sketch the best fit curve for non-linear data. To the best of our knowledge, we have not come across a researcher using this method to model a polynomial regression.

## Limitations & Future Enhancement

Our study focuses on the USA; therefore, the training data is limited. The initial 96645 records as shown in figure 1 reduced to 189 records as shown in figure 4 after data cleaning and narrowing down to the USA. The 189 records are further divided into train set and test set in the ratio of 7:3, leaving only 132 records to train the model. As discussed in section 2.4 and demonstrated in figure 5, focusing on one country was necessary due to different population sizes. Perhaps feature scaling on the population size of other countries as well as on the independent and dependent variables can be done to get more training data to train our data more accurately. However, accuracy cannot be guaranteed as different countries are in different situations. Lastly, this study does not support uncontrollable factors such as the emergence of new mutated variants of covid-19 viruses that could be potentially more contagious and fatal.

---Total word count: 3288 (including titles and picture captions)---

# References

COVID-19 Vaccination. (2021). Retrieved 8 August 2021, from <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness/how-they-work.html>

Dowdy, D., & D'Souza, G. (2021). What is Herd Immunity and How Can We Achieve It With COVID-19?. Retrieved 8 August 2021, from <https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html>

Evaluation of COVID-19 vaccine effectiveness. (2021). Retrieved 8 August 2021, from <https://www.who.int/publications/i/item/WHO-2019-nCoV-vaccine_effectiveness-measurement-2021.1>

Gerryn, C. (2021). Bankruptcies expected to increase 26% globally | Atradius. Retrieved 8 August 2021, from <https://group.atradius.com/press/press-releases/bankrupticies-expected-to-grow-twenty-six-percent-in-2021.html>

Irizarry, R. (2021). Chapter 28 Smoothing | Introduction to Data Science. Retrieved 8 August 2021, from <https://rafalab.github.io/dsbook/smoothing.html>

Kuadli, J. (2021). 11+ Mind-Blowing Bankruptcy Statistics for 2021. Retrieved 8 August 2021, from <https://legaljobs.io/blog/bankruptcy-statistics/>

Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., & Appel, C. et al. (2021). A global database of COVID-19 vaccinations. Nature Human Behaviour, 5(7), 947-953. doi: 10.1038/s41562-021-01122-8

Wang, J., Jing, R., Lai, X., Zhang, H., Lyu, Y., Knoll, M., & Fang, H. (2020). Acceptance of COVID-19 Vaccination during the COVID-19 Pandemic in China. Vaccines, 8(3), 482. doi: 10.3390/vaccines8030482

**ASSESSMENT RUBRIC**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **CRITERIA** | **MARKS** | | | | | | |
| **Methodology (30%)** | **16-20** | **13-15** | **10-12** | | **8-9** | **0-7** | **Comments** |
| Excellent in documenting the methodology.  • Generates complete, clear and unambiguous requirements specification.  • Identifies ambiguity in givens and states necessary assumptions.  • Uses appropriate diagrams to describe implementation clearly including the design decisions.  Overall contents comprehensively articulates all relevant and pertinent issues related to the overall solution.  All areas are at least good. May be outstanding is some areas and good in others and hence is on balance excellent. Good or above in all areas. Likely to contain minor errors, omissions or additions which prevent the methodology from being outstanding. Overall an excellent methodology. | Good in documenting the methodology.  • Generates requirements specification with minor residual ambiguity.  • Identifies ambiguity in givens however necessary assumptions are not fully stated.  • Uses appropriate diagrams to describe implementation however contain a small number of errors, omissions or additions.  It is possible that the methodology is outstanding or excellent in some areas and satisfactory in others but on balance is good. Satisfactory or above in all areas. Likely to contain a small number of errors, omissions or additions which prevent the methodology from being excellent. Overall a good methodology. | | Satisfactory in documenting the methodology.  • Generates requirements specification with some residual ambiguity.  • Omits ambiguity in givens and states necessary assumptions ambiguously.  • Uses appropriate diagrams to describe implementation however contain a number of errors, omissions or additions.  It is possible that the methodology is good or above in some areas and satisfactory in others. Weak in no more than two areas. Likely to contain a number of errors, omissions or additions which prevent the methodology from being good. Overall satisfactory. | Weak in documenting the methodology.  • Generates requirements specification with substantial ambiguity.  • Omits ambiguity in given and necessary assumptions.  • Uses inappropriate diagrams to describe implementation and contain large number of errors, omissions or additions.  It is possible that the methodology is satisfactory or above in some areas and unsatisfactory in others. Likely to be weak in more than three areas. It might be unsatisfactory in one area but no more. Likely to contain errors, omissions, additions, or misunderstandings which prevent the methodology from being satisfactory. Overall poor. | Unsatisfactory in documenting the methodology.  • Generates requirements specification with substantial ambiguity.  • Omits ambiguity in given and necessary assumptions.  • No diagrams to describe software architecture.  It is possible that the methodology is weak or above in some areas and unsatisfactory in others. Unsatisfactory in two or more areas. Likely to contain errors, omissions, additions, or misunderstandings which prevent the design model from being weak. May not be recognisable as a methodology, might have majors errors in content or a combination of the two. |  |
|  | **16-20** | **13-15** | | **10-12** | **8-9** | **0-7** |  |
| **Implementation (30%)** | Excellent in the implementation, comments, indentation, consistency and syntax. Correct following of object-oriented concepts. The work shows particular insight or originality in its approach. Excellent functionalities which identifies the underlying principles behind the problem. All areas are at least good. May be outstanding is some areas and good in others and hence is on balance excellent. Good or above in all areas. Likely to contain minor errors, omissions or additions which prevent the implementation from being outstanding. Overall an excellent implementation. | Good in the implementation, comments, indentation, consistency and syntax. It is possible that the application is outstanding or excellent in some areas and satisfactory in others but on balance is good. Satisfactory or above in all areas. Likely to contain a small number of errors, omissions or additions which prevent the implementation from being excellent. Overall a good implementation. | | Satisfactory in the implementation, comments, indentation, consistency and syntax. It is possible that the application is good or above in some areas and satisfactory in others. Likely to be weak in no more than two areas. Likely to contain a number of errors, omissions or additions which prevent the implementation from being good. Overall a satisfactory implementation. | Weak in the implementation, comments, indentation, consistency and syntax. It is possible that the application is satisfactory or above in some areas and unsatisfactory in others. Likely to be weak in more than three areas. It might be unsatisfactory in one area but no more. Likely to contain errors, omissions, additions, or misunderstandings which prevent the implementation from being satisfactory. Still recognisable as an object-oriented application of the problem in focus. Overall poor but satisfactory. | Unsatisfactory in the implementation, comments, indentation, consistency and syntax. It is possible that the application is weak or above in some areas and unsatisfactory in others. Unsatisfactory in two or more areas. Likely to contain errors, omissions, additions, or misunderstandings which prevent the implementation from being weak. May not be recognisable as an object-oriented application, might have majors errors in implementation or a combination of the two. |  |
| **Results (20%)** | **16-20** | **13-15** | | **10-12** | **8-9** | **0-7** |  |
|  | Excellent in the results discussion. The explanation and justification of how it meets specified requirements shows outstanding insight into the issues involved and alternatives available. The presentation clearly and concisely demonstrates a deep understanding of the project implementation. May be outstanding is some areas and good in others and hence is on balance excellent. Good or above in all areas. Likely to contain minor errors, omissions or additions which prevent the results discussion from being outstanding. Overall an excellent results discussion. | Good in the results discussion. The explanation and justification of how it meets specified requirements shows good understanding into the issues involved and alternatives available. The presentation demonstrates a good understanding of the project implementation. It is possible that the presentation is outstanding or excellent in some areas and satisfactory in others but on balance is good. Satisfactory or above in all areas. Likely to contain a small number of errors, omissions or additions which prevent the results discussion from being excellent. Overall a good results discussion. | | Satisfactory in the results discussion. The explanation and justification of how it meets specified requirements covers relevant aspects into the issues involved and alternatives available, but is not outstanding in any respect. It is possible that the presentation is good or above in some areas and satisfactory in others. Likely to be weak in no more than two areas. Likely to contain a number of errors, omissions or additions which prevent the results discussion from being good. Overall a satisfactory results discussion. | Weak in the areas of the results discussion. The explanation and justification of how it meets specified required is inadequate. It is possible that the presentation is satisfactory or above in some areas and unsatisfactory in others. Likely to be weak in more than three areas. It might be unsatisfactory in one area but no more. Likely to contain errors, omissions, additions, or misunderstandings which prevent the results discussion from being satisfactory. Overall poor but satisfactory. | Unsatisfactory in the areas of the results discussion. It conveys little understanding of solution of the problem or no useful explanation and justification of how it meets specified requirements. It is possible that the presentation is weak or above in some areas and unsatisfactory in others. Unsatisfactory in two or more areas. Likely to contain errors, omissions, additions, or misunderstandings which prevent the results discussion from being weak. |  |
| **Communication (15%)** | **16-20** | **13-15** | | **10-12** | **8-9** | **0-7** | **Comments** |
|  | Excellent, well-directed presentation, logically and coherently structured. It is free or almost free grammatical errors. The format is clear and consistent with appropriate use of headings and paragraphs. English usage is easily understandable. References and quotations are utilized appropriately to indicate sources. | Good presentation, logically structured. There are occasional spelling and grammatical errors, but the reader does not struggle to interpret the writer’s intended meaning. The writing would benefit from the use of organizational tools (e.g. headings, paragraphs) and more consistent use of references to sources. | | Satisfactory presentation, well structured. It contains number of spelling and grammatical errors, but the reader does not struggle to interpret the writer’s intended meaning. It is possible that the use of organizational tools (e.g. headings and paragraphs) are good or above in some areas and satisfactory in others. | Weak presentation and structure. Spelling and grammatical errors force the reader to struggle to determine the intended meaning. Organizational tools such as headings, paragraphs are used inconsistently. References are not used properly to indicate the sources of material. | Unsatisfactory presentation and structure. Numerous spelling and grammatical errors and a lack of clear consistent organization interfere with the writer’s ability to communicate to the key points. The reader frequently cannot determine the intended meaning. There are no references to indicate material taken from other sources. |  |

**Turnitin Report**

