When will Every Canadian Be Fully Vaccinated?

Preface: This report highlights my effort in science communication for a layman during my internship at Science for All Audience (SCIFAA). My goal is to translate the technical language and communicate it to every non-technical person. In this context, I will explain in layman language what data science is and how its power can be applied to see a trend of Canadian vaccinations and forecast when Canada will be fully vaccinated.

Table of Contents

Introduction

Coronavirus, short for COVID-19, has been a hot topic since its spread. Like regular germs, viruses are tiny particles. When they get inside a body, be it an animal, plant, or human, they cause sicknesses, such as flu, colds, chickenpox, and other diseases. The COVID-19 virus caused pneumonia-like infections, and the first person infected by the virus was in Wuhan, China, in late October 2020. COVID-19 has a transmission capability greater than common flu; this led the outbreak to spread to many countries, and hotspots began to appear globally by the end of January 2020. The outbreak was declared an International Health Emergency, then a pandemic on 11 March 2020, and lockdown began in most parts of the world. According to the John Hopkins University, as of 25 July 2021, up to 194 million COVID cases have been registered, and it has taken 4.14 million deaths [1]. Everyone is tired and stressed about the lockdown and wants to get back to life and meet people freely. Therefore, every country has been putting in tremendous effort to combat this virus. Being one of the leading countries, Canada is taking every possible action to ensure its citizens' safety from COVID-19. One of the measures is the Health Canada COVID-19 vaccination program aiming to vaccinate every Canadian by early September, at least with one dose. The question here is how Health Canada foresaw the time where every Canadian have gotten at least one or both doses of the COVID-19 vaccine. And will Health Canada meet this target or not? This report answers these research questions, and the research is conducted by leveraging the power of data science.

The Power of Data Science

Data Science used a bit of mathematics and computing power to understand loads of information. Here's how it is done.

Data Collection and Preparation

It all starts with collecting information, which is called data, in the language of data science. When information/data is collected, it is then cleaned to ensure if the information/data is entered correctly and fixed by replacing the missing values.

Data Analysis

The next step is to understand the data or taking a bird's eye view of it. This step is called data analysis and provides insights in the form of numbers. In this step, information/data is processed by applying mathematical operations and technological knowledge to find every insight possible. The insights/numbers are knowns as 'Descriptive Statistics'. This raises the question of why we need descriptive statics to describe data? This is simply because these insights give the idea of the overall shape of the data, which can be depicted on a chart like a bar chart, histogram, or dot plot for the data visualization step.

Data analysis for descriptive statistics has been performed using R programming in the R studio. It is one of the leading software environments to work with data and its statical modeling and analysis. The major reason to use R for this project is its packages for creating visually appealing plots.

Software packages and functions are like a library, and each book in that library is a function. For example, a book (function) from a mathematics library (library) will be needed to add two numbers.

Data Visualization

Many times, some of the simple questions or judgments are taken from information/data by visually looking at its numbers/statistics. This is done by putting statistics in charts or graphs. Moreover, it is easier to read graphs than lots of data. This work is called data visualization.

Data Modeling

The insights/numbers, along with charts or graphs, are then used to support any opinions. A common example is looking at the weather information/data and forecasting if the weather will be sunny or rainy.

Research for Data and Insights

For the data science project at the Science for All Audience (SCIFAA), I am pursuing this opportunity to interpret the COVID-19 vaccination data of Canada. The ultimate goal is to forecast the possible time when every Canadian will get the first dose or both of the vaccines.

Data is taken from the database ‘Our World in Data,’ and the population (people) that have received the COVID vaccine till July is considered [2]. Moreover, the reference for this project is the estimation of the vaccination effort in the United States of America [3].

It has been assumed that the population provided includes people from all groups, including adults and kids of 16 years of age or younger. Moreover, the population has either received one dose of their COVID-19 vaccine or both doses. It has also been assumed that a certain percentage of the population has not received even the first dose of vaccines, such as pregnant women, mothers, people who are afraid of needle injection for vaccination, and newly-born kids.

The data index presented in Fig. 1. is its raw form and contains four columns. The first column is for Day number, which started from 1, the second and third columns mentioned the country name and its code, respectively. The fourth column shows vaccination progress and its titles as people vaccinated per hundred. Each row in this column presents the vaccinated people in a group of 100 people in a day. For example, the first row of the column is zero, and this means the first day, marked as 22 February 2021, had zero people vaccinated in a group of 100 people. Moreover, as you go down this column, the vaccinations begin to increase with time.

Number	Country	Code	Date	people_vaccinated_per_hundred
1	Afghanistan	AFG	2021-02-22	0.00
2	Afghanistan	AFG	2021-02-28	0.02
3	Afghanistan	AFG	2021-03-16	0.14
4	Afghanistan	AFG	2021-04-07	0.31
5	Afghanistan	AFG	2021-04-22	0.62
6	Afghanistan	AFG	2021-05-11	1.15
7	Afghanistan	AFG	2021-05-20	1.21
8	Afghanistan	AFG	2021-05-24	1.22
9	Afghanistan	AFG	2021-05-26	1.23
10	Afghanistan	AFG	2021-05-27	1.23
11	Afghanistan	AFG	2021-05-30	1.23
12	Afghanistan	AFG	2021-06-02	1.24
13	Afghanistan	AFG	2021-06-03	1.24
14	Afghanistan	AFG	2021-06-08	1.24
15	Afghanistan	AFG	2021-06-14	1.25
16	Afghanistan	AFG	2021-06-22	1.50
17	Afghanistan	AFG	2021-06-27	1.67
18	Afghanistan	AFG	2021-06-30	1.80
19	Afghanistan	AFG	2021-07-05	1.87

Fig. 1. Share of people who received at least one dose of COVID-19 vaccine. (n.d.). Our World in Data.

The Story Data Tells

The data [2] was huge, and the ultimate goal was to find out any trend in it. A trend in any data can find out by checking if there is any relationship between two or more columns. And then it can be observed if the trend is increasing or decreasing. In order to find a trend in the worldwide COVID-19 vaccination data, each country in the data index is sliced to the day ‘2021-07-14’, and the respective people vaccinated per hundred are taken into consideration for data analysis as depicted in Fig. 2.

Number	Country	Code	Date	people_vaccinated_per_hundred
1	Afghanistan	AFG	2021-07-14	1.98
2	Africa	AF	2021-07-14	2.95
3	Albania	ALB	2021-07-14	20.66
4	Argentina	ARG	2021-07-14	45.59
5	Aruba	ABW	2021-07-14	66.27
6	Asia	AS	2021-07-14	25.97
7	Australia	AUS	2021-07-14	27.38
8	Austria	AUT	2021-07-14	56.45
9	Azerbaijan	AZE	2021-07-14	24.78
10	Bahrain	BHR	2021-07-14	64.16
11	Barbados	BRB	2021-07-14	33.75
12	Belgium	BEL	2021-07-14	66.33
13	Brazil	BRA	2021-07-14	42.42
14	Bulgaria	BGR	2021-07-14	14.71
15	Cambodia	KHM	2021-07-14	31.34
16	Canada	CAN	2021-07-14	69.81
17	Cayman Islands	CYM	2021-07-14	74.63
18	Chile	CHL	2021-07-14	70.28
19	Colombia	COL	2021-07-14	28.15

Fig. 2. People vaccinated per hundred till 14/07/2021 in each country

This is followed by estimating descriptive statistics on the data, finding the top 5 countries with the highest and lowest vaccinated people share per hundred as bar charts in Fig. 3 and Fig 4. The horizontal axis is labeled with the name of countries, and the vertical axis is labeled with the number of people vaccinated per hundred. The bar chart in Fig. 3 shows that Gibraltar has the highest number of people vaccinated, which is 116.7 in a group of 100 people. Moreover, the bar chart in Fig. 4 shows that Zambia has the lowest number of people immunized, that 0.9 people vaccinated per hundred.

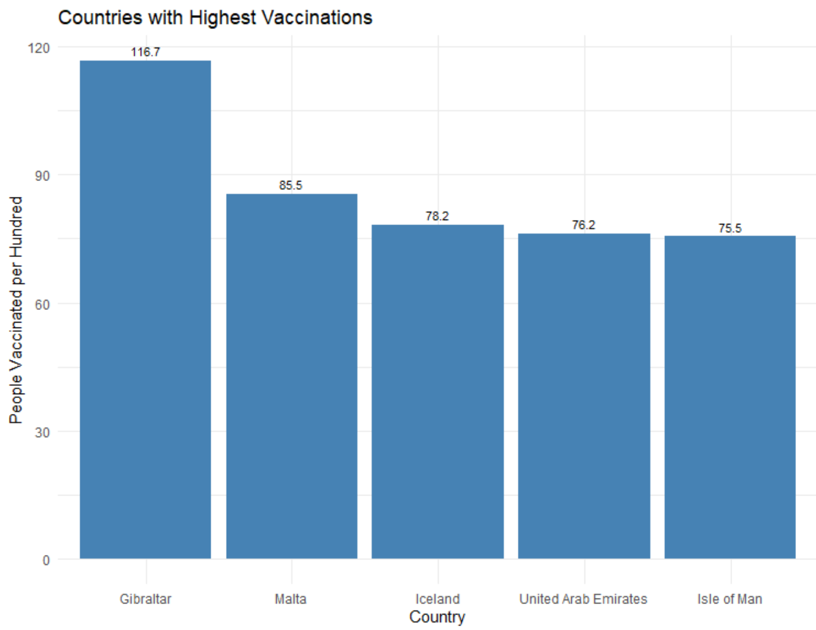

Fig. 3. Top 5 countries with most vaccinated people till 14/07/2021

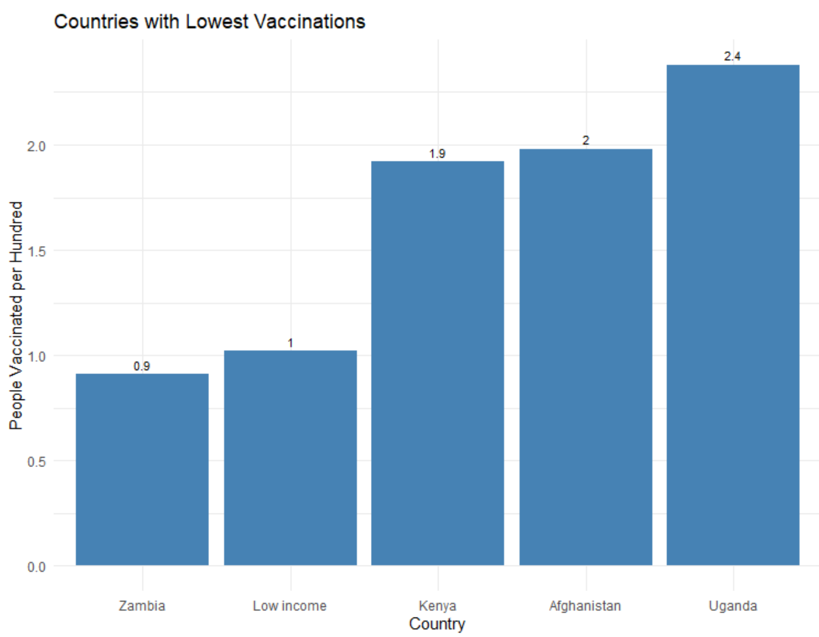

Fig. 4. 5 countries with lowest vaccinated people till 14/07/2021

The data [2] is then indexed for Canada and, as depicted in Fig. 5., showing the COVID-19 vaccination progress in Canada from 15/12/2020 to 15/07/2021.

Number	Country	Code	Date	people_vaccinated_per_hundred
1	Canada	CAN	2020-12-15	0.00
2	Canada	CAN	2020-12-16	0.01
3	Canada	CAN	2020-12-17	0.02
4	Canada	CAN	2020-12-18	0.03
5	Canada	CAN	2020-12-19	0.03
6	Canada	CAN	2020-12-20	0.03
7	Canada	CAN	2020-12-21	0.06
8	Canada	CAN	2020-12-22	0.07
9	Canada	CAN	2020-12-23	0.09
10	Canada	CAN	2020-12-24	0.12
11	Canada	CAN	2020-12-25	0.12
12	Canada	CAN	2020-12-26	0.13
13	Canada	CAN	2020-12-27	0.14
14	Canada	CAN	2020-12-28	0.16
15	Canada	CAN	2020-12-29	0.19
16	Canada	CAN	2020-12-30	0.22
17	Canada	CAN	2020-12-31	0.26
18	Canada	CAN	2021-01-01	0.27
19	Canada	CAN	2021-01-02	0.29

Fig. 5. Canadians vaccinated per hundred till 14/07/2021

According to the eligibility criteria to receive the vaccine for Canadians by Health Canada, it has been assumed that the share of the population includes adults and kids of 16 years of age or younger who have received one dose of their COVID-19 vaccine or both doses. Moreover, the population, including newly-born children, pregnant women, expecting mothers, and people who aren't comfortable with the vaccination, has been omitted from the data.

An important descriptive statistic, known as correlation co-efficient, was estimated between people vaccinated per hundred and the number of days for the vaccinations equal to 0.95. This positive correlation co-efficient hints that the nature of the data is linear, or as one increases, the other is also increasing, and that too, in the same direction.

The data in the column 'people_vaccinated_per_hundred' is then visualized using ggplot() library and geom_point function, Fig. 6. Its horizontal axis shows the time as in day and month, and the vertical axis shows people vaccinated per hundred. A trend was observed in December of 2020 where people weren't opened to the idea of being vaccinated. As a result, hardly 1 to 5 people were getting vaccination each day. However, the vaccination progress slightly increases during January and February. Then from March to May, more people got the vaccinations, which can be clearly observed with the steepness of the curve in Fig. 6. And after this, more Canadians have been getting vaccinations, but the increase is slow. In a nutshell, the plot of data in Fig. 6. shows that people who are being vaccinated are increasing with time, and this is another evidence of data being linear, which was proved by the correlation coefficient.

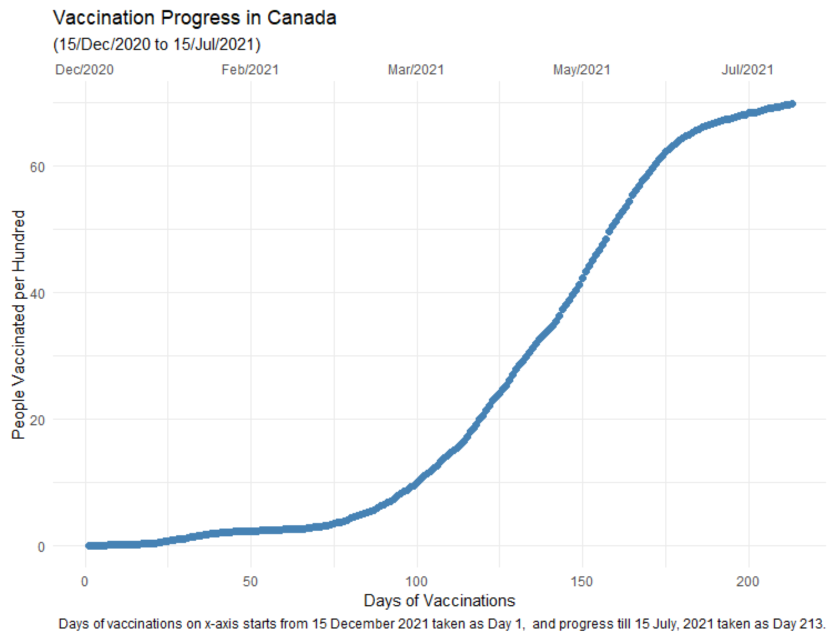

Fig. 6. Vaccination Progress in Canada as scatter plot

Having seen this linear trend, a model based on linear regression has been developed and then fitted on the Fig. 7 graph as a regression line in red color. A regression line is presented by the equation of a straight line:

$$y = m*x + c$$

where m = slope of a straight line

c = y-intercept of a straight line

x = input (people vaccinated per Hundred)

y = output (Time)

R programming has a built-in function ‘lm()’ used to calculate the slope and y-intercept of the linear regression line. This linear regression model can greatly help predict the time when every Canadian will be vaccinated.

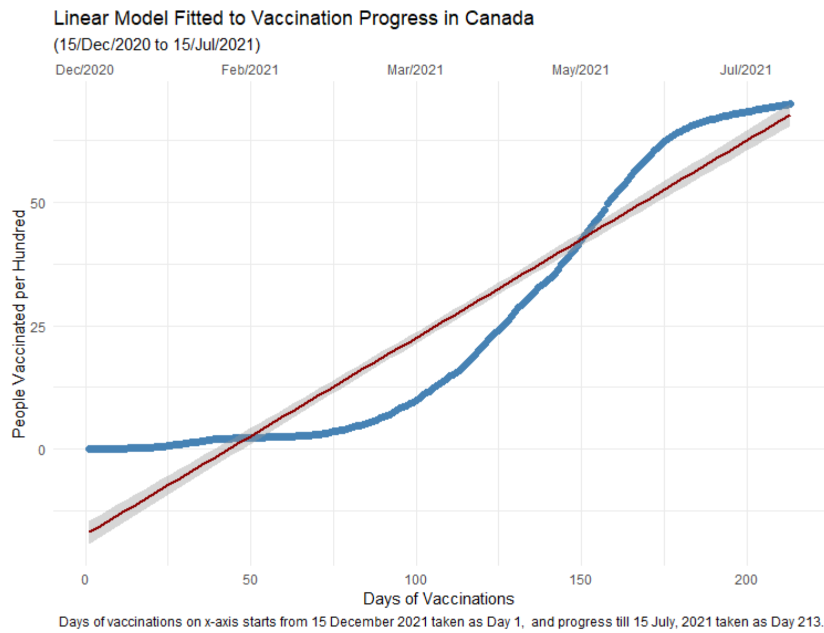

Fig. 7. Vaccination Progress in Canada as scatter plot (blue); linear regression of the data (dark red)

Prediction – When will Canada be Vaccinated Completely?

The accuracy of the linear regression model was estimated using the statistics calculated through the model of the regression line to predict, which was approximately 80 percent. With accuracy, the model was tested to predict different instances of time in the future with respect to the number of people in a group of 100 and presented the results as the scatter plot with black points, Fig. 8. The accuracy can be improved by using more data.

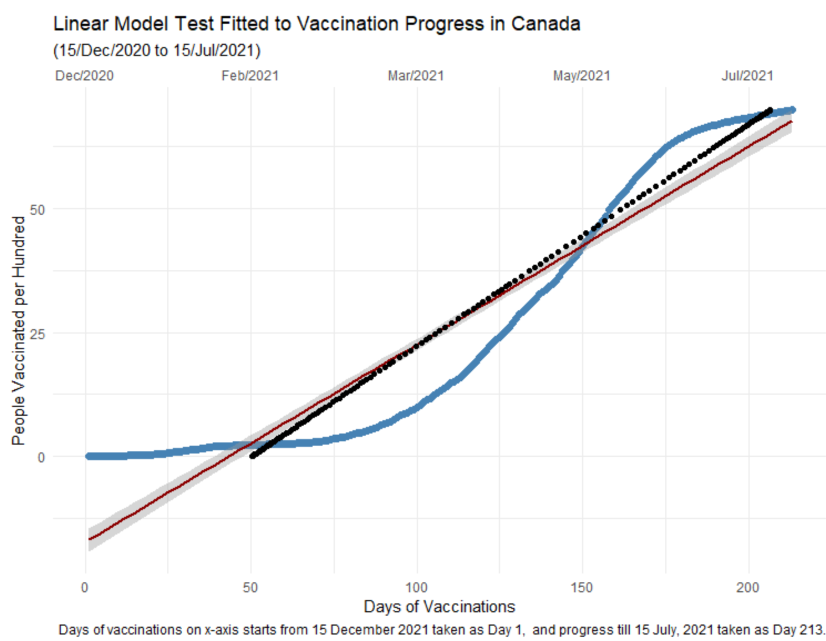

Fig. 8. The testing model with prediction and estimated results presented as a scatter plot (black)

Another prediction test on the model was conducted to find the time when 100 percent of Canadians will be completely vaccinated with one COVID-19 dose, and the model estimated the date of **14 September 2021**.

Conclusion

Based on the research conducted on the COVID-19 vaccination data in Canada, a linear regression model was developed to predict the time that would have every Canadian vaccinated with at least one dose. With respect to the results estimated by the model, hopefully, in September 2021, every Canadian will be completely vaccinated by one dosage of COVID-19 vaccines. Knowing this information, a possible time in future can be gathered when everyone will go back to pre-COVID regular life, when all the shops could be open with full capacity and when we could visit friends at their home without any restrictions. These results from the model also strengthen the Health Canada target and their approach to get the entire Canada vaccinated by September 2021.

This project has space for improvement. Using the statistics estimated by the linear regression model, the time can be predicted for the second dose of vaccine. Moreover, a trend of complete 2021 year can be created visually. Furthermore, by incorporating more of the original data, the model's accuracy can also be improved.

References

[1] COVID-19 Map. (n.d.). Johns Hopkins Coronavirus Resource Center. Retrieved 12 August, 2021, from https://coronavirus.jhu.edu/map.html

[2] Share of people who received at least one dose of COVID-19 vaccine. (n.d.). Our World in Data. Retrieved 12 July, 2021, from https://ourworldindata.org/grapher/share-people-vaccinated-covid

[3] Gerrard, B. (2021, 30 March). How I Built a Forecast to Estimate When the U.S. Will Be Vaccinated. Medium. https://towardsdatascience.com/when-will-the-us-be-vaccinated-1b24890a8c38