COVID-19 Vaccination Time Series Analysis

Table of Content

1. Team Members

2. Problem Statement

3. Objectives

4. Data Sources

5. Content of Data

6. Plan of Attack

7. Methodology

References

1. Team Members

1) Jayarani Emekar
2) Viraj Sonavane
3) Monika Gadage

2. Problem Statement

The COVID-19 pandemic is the most crucial health disaster that has surrounded the world for the past year. Predicting the COVID-19 vaccination trend has become a challenging issue. Many health professionals, statisticians, researchers, and programmers have been tracking the spread of the virus in different regions of the world using various approaches.

The rise in various vaccines developed by talented scientists spurred curiosity about learning more about ongoing vaccine programs and a keen interest in finding meaningful insights from data drove us to work on this particular endeavor.

3. Objectives

The project aims to convey the analysis of different ongoing vaccination programs around the globe by using the inferences discovered from the scraped data from the internet. Time series analysis is helpful to determine the progress of vaccination.

This Exploratory Data Analysis and Time series analysis strives to highlight the battle, forefronted by science and the presence of vaccines, between the virus and humanity. With many vaccination programs having started in late 2020, day by day, the world is slowly becoming more and more protected.

4. **Data Sources**

We will be using data from Our World in Data GitHub repository for covid-19, where the data is collected daily and uploaded monthly on the repository.

Data file will be in csv format country vaccination data.

5. **Content of Data**

The data contains following information:

- **Country-** this is the country for which the vaccination information is provided;
- **Country ISO Code -** ISO code for the country;
- **Date -** date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;
- **Total number of vaccinations -** this is the absolute number of total immunizations in the country;
- **Total number of people vaccinated -** a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccination might be larger than the number of people;
- **Total number of people fully vaccinated -** this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme;
- **Daily vaccinations (raw) -** for a certain data entry, the number of vaccination for that date/country;
- **Daily vaccinations -** for a certain data entry, the number of vaccination for that date/country;
- **Total vaccinations per hundred -** ratio (in percent) between vaccination number and total population up to the date in the country;
- **Total number of people vaccinated per hundred** - ratio (in percent) between population immunized and total population up to the date in the country;
- **Total number of people fully vaccinated per hundred** - ratio (in percent) between population fully immunized and total population up to the date in the country;
- **Number of vaccinations per day** - number of daily vaccination for that day and country;
- **Daily vaccinations per million** - ratio (in ppm) between vaccination number and total population for the current date in the country;
- **Vaccines used in the country** - total number of vaccines used in the country (up to date);
- Source name - source of the information (national authority, international organization, local organization etc.);
- **Source website -** website of the source of information;

6. **Plan of Attack**

1) Data preparation and cleaning
2) Find relation between the different data columns and finalize the same
3) Exploratory data analysis (EDA) with visualization on histograms and geo plots
    a) Vaccinated people Information analysis and how Vaccination Process changed over time
    b) Where are more people vaccinated per day? But in terms of percent from the entire population ?
    c) In which country the vaccination programme is more advanced?
    d) In which month has the most number of people vaccinated?

    e) Which vaccination schemes (combination of vaccines) are used and in which countries?
    f) Which country vaccinates/immunizes a larger percent from its population?
4) Analysis predictive Auto Regressive Integrated Moving Average (ARIMA) model parameters and work on the following point:
    a) Trace the daily vaccinations dynamic
    b) Time series transformations to make it stationary (as data is univariate)
    c) Predict the values based on stationary time series
    d) Create predictions for the upcoming weeks.
    e) When will the word achieve Herd immunity?

7. **Methodology**

1) We are planning to use AutoRegressive Integrated Moving Average (ARIMA). It is a class of model that captures a suite of different standard temporal structures in time series data
2) It combines both Autoregression (AR) and Moving Average (MA) models as well as a differencing pre-processing step of the sequence to make the sequence stationary, called integration (I) .
3) The method is suitable for univariate time series with trend and without seasonal components.

References

1) https://ourworldindata.org/covid-vaccinations
2) https://github.com/owid/covid-19-data
3) https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average
4) https://towardsdatascience.com/an-overview-of-time-series-forecasting-models