

COVID Vaccines Analysis

DAC_Phase4 (Development Part 2)

Title:

The Covid-19 vaccine analysis by collecting and conducting the analysis by performing the exploratory data analysis, statistical data analysis and visualization of the given data.

Abstract:

The rapid distribution and administration of vaccines are of paramount importance in the global effort to combat infectious diseases, and the comprehensive analysis of vaccine distribution data can provide valuable insights for optimizing public health strategies. With the ongoing challenges posed by the COVID-19 pandemic and the necessity to manage vaccine supply chains effectively, the utilization of advanced machine learning techniques has emerged as a key innovation.

This document focuses on the systematic collection and analysis of data related to COVID-19 vaccines. The study leverages diverse datasets, including clinical trial results, real-world vaccination data, adverse event reports, and epidemiological statistics, to provide a comprehensive analysis of the vaccines' performance. The goal is to gain insights into vaccine efficacy, safety, and the impact of different vaccination strategies.

Exploratory data analysis:

Exploratory Data Analysis (EDA) is a crucial step in understanding and gaining insights from data. When conducting an EDA of COVID-19 vaccine data, you'll typically have access to various datasets that can provide information on vaccine distribution, effectiveness, side effects, and more.

Dataset link:

<https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>

Understand Data Structure:

EDA helps you understand the basic structure of your dataset, such as the number of variables (features) and data points (observations). It also helps in recognizing the types of data (numerical, categorical, etc.).

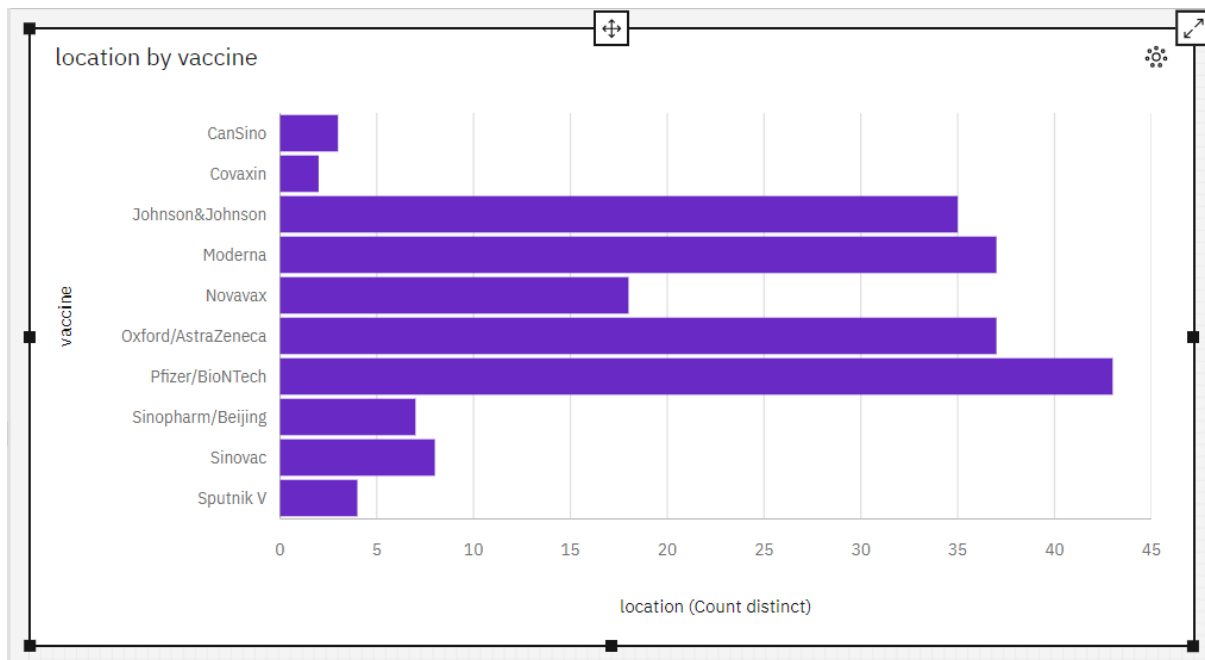
Data Visualization:

One of the most important aspects of EDA is data visualization. Visualization techniques include histograms, bar charts, scatter plots, box plots, and more. These visualizations can reveal patterns, relationships, and potential outliers in the data.

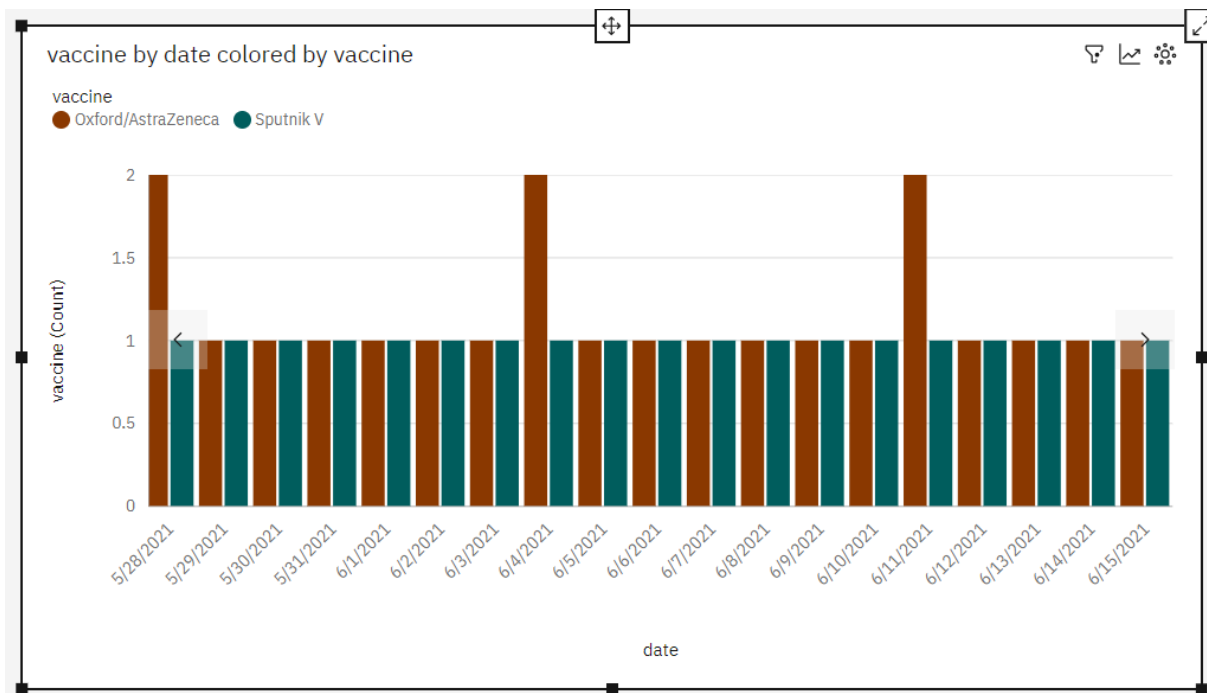
Data Preprocessing:

As you explore the data, you may discover the need for data cleaning and preprocessing steps, such as scaling, encoding categorical variables, or handling imbalanced datasets.

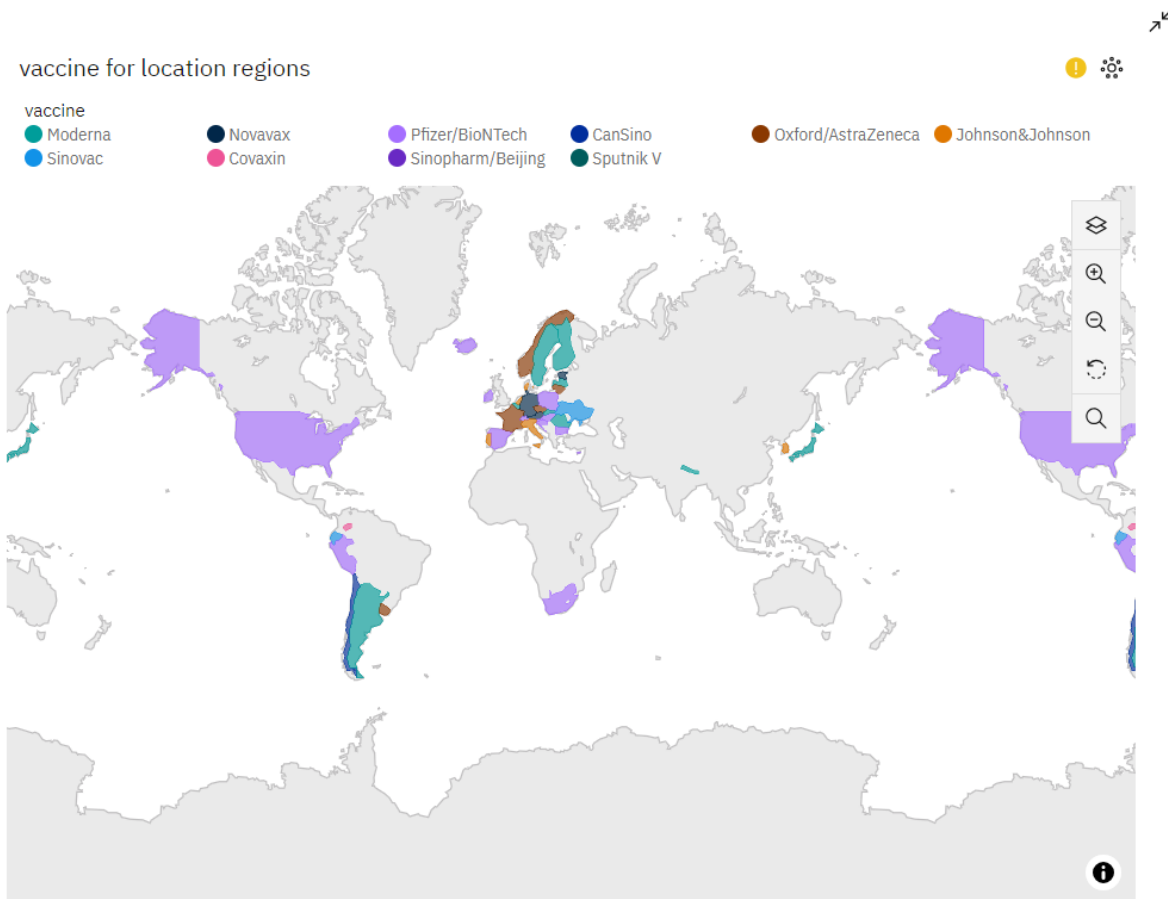
Statistical analysis and data Visualization using Cognos tool:



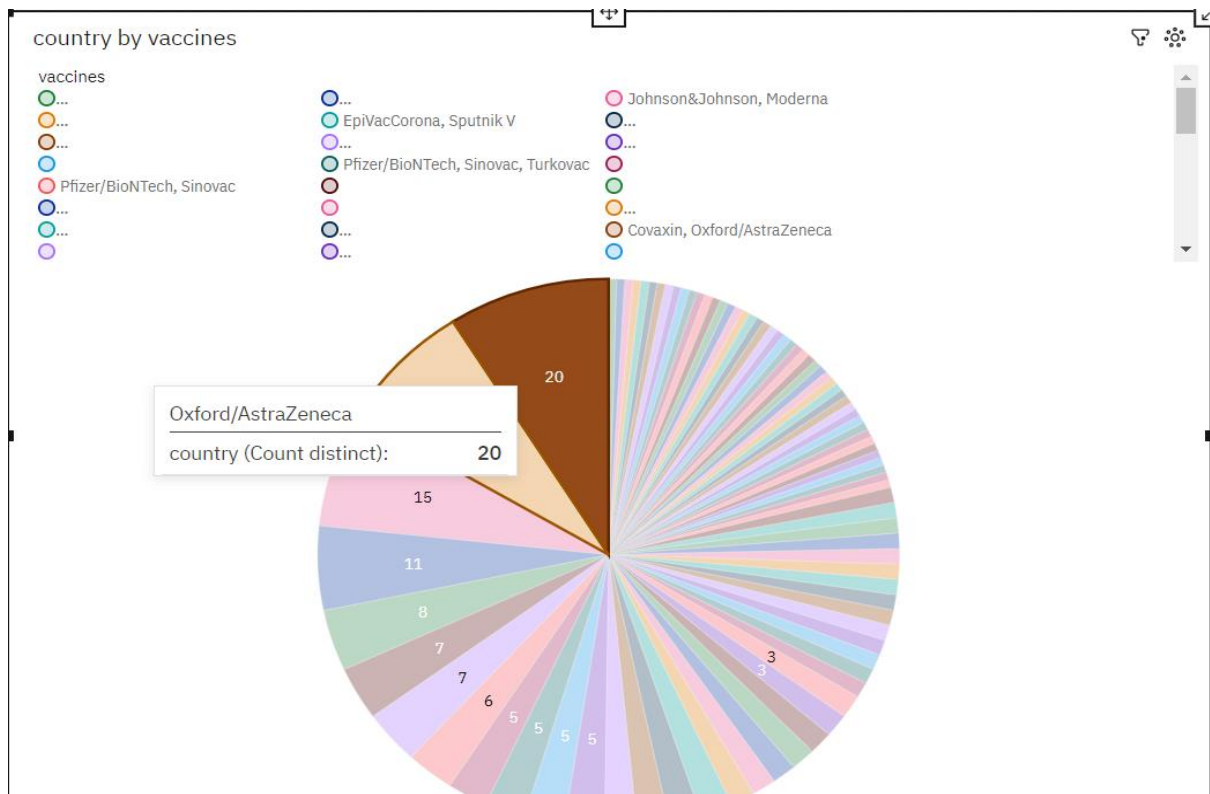
- Pfizer/BioNTech is the most frequently occurring category of vaccine with a count of 8888 items with location values (25 % of the total).
- The total number of results for location, across all vaccines, is almost 36 thousand.



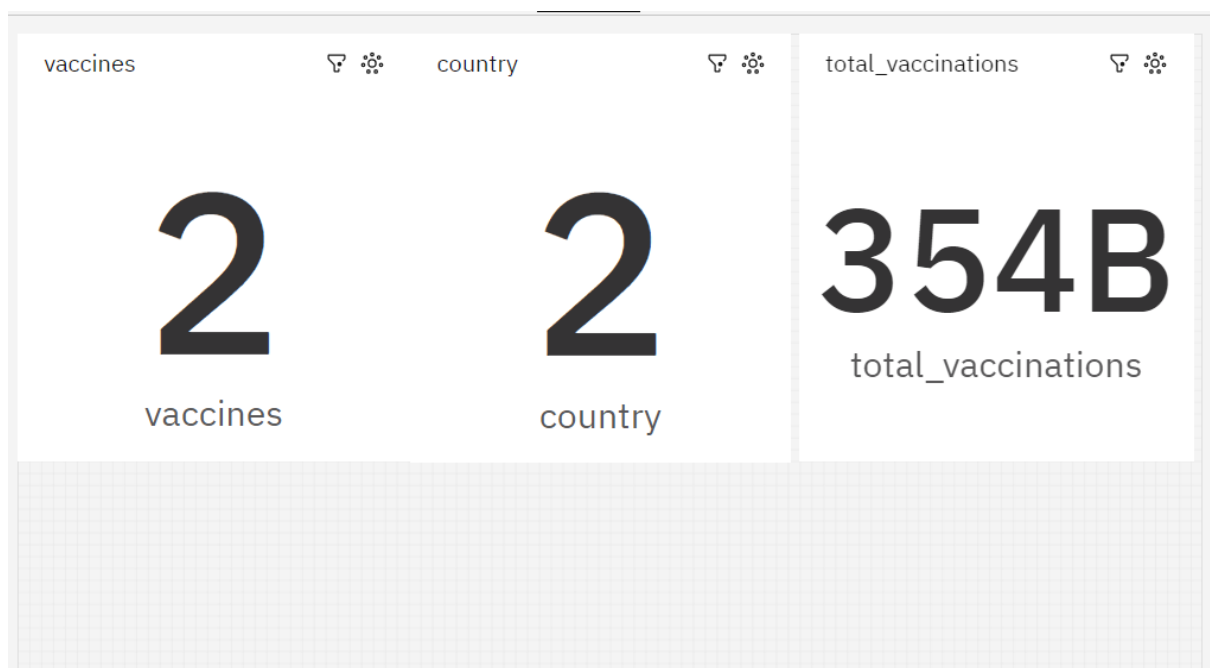
- It is projected that by 2022-06-29, Oxford/AstraZeneca will exceed Sputnik V in vaccine by 0.12.
- From 2021-02-04 to 2021-02-05, Oxford/AstraZeneca's vaccine increased by 100%.
- 2021-02-12 (0.3 %), 2021-12-17 (0.3 %), 2021-12-24 (0.3 %), 2021-02-05 (0.3 %), and 2021-12-31 (0.3 %) are the most frequently occurring categories of date with a combined count of 15 items with vaccine values (1.5 % of the total) .
- Oxford/AstraZeneca is the most frequently occurring category of vaccine with a count of 515 items with vaccine values (53 % of the total).
- The total number of results for vaccine, across all dates, is 971.



- European Union location accounted for 41% of Pfizer/BioNTech total_vaccinations compared to 24% for Moderna.
- location European Union has the highest total_vaccinations at approximately 193 billion, out of which vaccine Pfizer/BioNTech contributed the most at approximately 141 billion.
- vaccine Pfizer/BioNTech has the highest total total_vaccinations due to location European Union.
- The total number of results for vaccine, across all locations, is almost 36 thousand.



- Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech (23.3 %) and Moderna, Oxford/AstraZeneca, Pfizer/BioNTech (21.6 %) are the most frequently occurring categories of vaccines with a combined count of 1756 items with country values (44.9 % of the total).
- The total number of results for country, across all vaccines, is nearly four thousand.



- total_vaccinations has a weak weekly trend. The smallest values typically occur on Saturday.
- total_vaccinations has a strong upward trend.
- date 2021-01-15 has the lowest total total_vaccinations at 0.0, followed by 2021-01-16 at 191181.0.
- date 2021-02-14 has the highest total total_vaccinations at 0.0, followed by 2022-03-29 at 2.088846244E9.
- Based on the current forecasting, total_vaccinations may reach over 2.3 billion by date 2022-06-25.
- total_vaccinations has unusually low values for 9 time points, the most notable of which are 2021-08-12, 2021-07-18, 2021-07-31, 2021-08-15, and 2021-08-02.
- From 2021-08-15 to 2021-08-16, total_vaccinations increased by 502%.
- The overall number of results for total_vaccinations is 721.

Conclusion:

In the face of global health crises like the COVID-19 pandemic, advanced machine learning techniques offer innovative solutions to the challenges of vaccine distribution. This document explores the application of clustering and time series forecasting to uncover hidden patterns in vaccine distribution and adverse effects data. By doing so, we aim to contribute to more effective vaccine allocation, better planning, and informed decision-making in the realm of public health.