COVID-19 VACCINATION ANALYSIS

OBJECTIVES:

* Introduction
* Project objective
* Design thinking
* Dataset details
* Program code
* Correlation code
* Conclusion

INTRODUCTION TO COVID-19 VACCINE:

**Vaccines are the suspension of killed microbes which provides immunity towards a specific disease. Vaccine builds our immunity which helps to fight the microbial attack. The body creates antibodies to attack them when vaccines are ingested or injected into the patient's body.**

**PROJECT OBJECTIVES:**

1. **To quantify hospital-based outcomes and deaths, including in relation to sociodemographic characteristics and comorbidities as ascertained from hospital AND general practice data.**
2. **To estimate the strength of association between these outcomes and sociodemographic and health characteristics.**

Problem definition:**The problem is to conduct an in-depth analysis of covid-19 vaccine data,focusing on vaccine efficacy,distribution,and adverse effects.This project involves data collection,data processing,exploratory data analysis,statistical analysis,and Visualization.**

**Design thinking:**

**DATA COLLECTION:**

**Vaccines reduce risks of getting a disease by working with your body's natural defences to build protection.Vaccine effectiveness against COVID-19 mortality was 58.7% for a first dose, 88.6% for a second dose and 93.2% for a third dose. Protection increases with the increase of doses and is high for the third dose or booster, as has been shown in previous research.**

**DATA PREPROCESSING:**

**A preprocessing step is necessary to convert raw dirty data to trainable, understandable and analyzing data in the sentiment analysis algorithm.**

**EXPLORATORY DATA ANALYSIS:**

**we conducted exploratory data analysis on the existing data of COVID-19 vaccine. It reveals the types and quantities of vaccines currently in use, shows the comparison of vaccination data from different countries and global vaccination trends, made the comparison between China and India, and has a more in-depth and clear understanding of the world's fight against the epidemic situation.**

**STATISTICAL ANALYSIS:**

**we applied statistical methods to find factors that have contributed to the fast development of COVID-19Vaccine analysis.**

**VISUALIZATION:**

**The challenge in data visualization identified by the author is the visual scalability. In other word, large data must be compressed or reduced before it could be visualized.**

**DATASET DETAILS:**

**Dataset link:**

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

**Data Preprocessing:**

* **Data preprocessing is a crucial step within the statistics analysis and gadget gaining knowledge of pipeline.**
* **It includes a sequence of strategies and operations finished on uncooked statistics to clean, organize, and transform it right into a layout that is suitable for analysis or device mastering version schooling.**
* **Data preprocessing goals to enhance the first-class of the records, making it greater reliable and conducive to generating accurate consequences.**

**Here are some common tasks and techniques involved in data preprocessing:**

**Data Cleaning:**

* **Handling missing values: Deciding how to deal with missing data, whether by imputing values or removing incomplete records.**
* **Outlier detection and treatment: Identifying and handling data points that significantly deviate from the norm.**

**Noise reduction:**

* **Smoothing noisy data through techniques like filtering.**

**Data Transformation:**

* **Data normalization:** **Scaling numerical features to a standard range (e.g., between 0 and 1) to ensure that they have similar influence in the analysis.**
* **Encoding categorical variables: Converting categorical data into numerical format, such as one-hot encoding or label encoding.**
* **Feature engineering: Creating new features or modifying existing ones to capture more meaningful information from the data.**
* **Dimensionality reduction: Reducing the number of features while retaining essential information, using methods like Principal Component Analysis (PCA).**

**Data Integration:**

* **Merging or joining datasets: Combining data from multiple sources into a single dataset for analysis.**

**Aggregation:** **Summarizing data at a higher level of granularity, such as aggregating daily sales into monthly totals.**

**Data Reduction:**

* **Sampling:** **Reducing the size of a large dataset by randomly selecting a representative subset.**
* **Binning: Grouping continuous data into discrete bins to simplify analysis.**
* **Filtering: Selecting a subset of data based on specific criteria**.

**Data Standardization:**

* **Ensuring that data follows a consistent format and structure.**
* **Date and time format conversion: Converting date and time data into a uniform format.**
* **Currency conversion: Converting monetary values into a common currency.**

**Data Scaling:**

* **Scaling numerical data to a common range to prevent some features from dominating the analysis.**

**Data preprocessing is an iterative process that may involve several of these steps in various orders, depending on the specific dataset and the analysis goals. Proper data preprocessing is essential for improving the accuracy and effectiveness of machine learning models, as well as for making data more accessible for traditional statistical analysis.**

**PROGRAM FOR COVID-19:**

**import numpy as** **np**

**import pandas as** **pd**

**import matplotlib.pyplot as** **plt**

**import seaborn as** **sns**

**import** **os**

**for** **dirname, \_, filenames in os.walk('/kaggle/input'):**

**for** **filename in filenames:**

**print(os.path.join(dirname, filename))**

**data = pd.read\_csv(“http://www.kaggle.com/dataset/gpreda/covid-world-vaccination-progress”****)**

**data.head()**

**OUTPUT:**

| country | iso\_code | date | total\_vaccinations | people\_vaccinated | people\_fully\_vaccinated | daily\_vaccinations\_raw | daily\_vaccinations | total\_vaccinations\_per\_hundred | people\_vaccinated\_per\_hundred | people\_fully\_vaccinated\_per\_hundred | daily\_vaccinations\_per\_million | vaccines | source\_name | source\_website |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **Afghanistan** | **AFG** | **2021-02-22** | **0.0** | **0.0** | **NaN** | **NaN** | **NaN** | **0.0** | **0.0** | **NaN** | **NaN** | **Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...** | **World Health Organization** | **https://covid19.who.int/** |
| **1** | **Afghanistan** | **AFG** | **2021-02-23** | **NaN** | **NaN** | **NaN** | **NaN** | **1367.0** | **NaN** | **NaN** | **NaN** | **34.0** | **Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...** | **World Health Organization** | **https://covid19.who.int/** |
| **2** | **Afghanistan** | **AFG** | **2021-02-24** | **NaN** | **NaN** | **NaN** | **NaN** | **1367.0** | **NaN** | **NaN** | **NaN** | **34.0** | **Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...** | **World Health Organization** | **https://covid19.who.int/** |
| **3** | **Afghanistan** | **AFG** | **2021-02-25** | **NaN** | **NaN** | **NaN** | **NaN** | **1367.0** | **NaN** | **NaN** | **NaN** | **34.0** | **Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...** | **World Health Organization** | **https://covid19.who.int/** |
| **4** | **Afghanistan** | **AFG** | **2021-02-26** | **NaN** | **NaN** | **NaN** | **NaN** | **1367.0** | **NaN** | **NaN** | **NaN** | **34.0** | **Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...** | **World Health Organization** | **https://covid19.who.int/** |

**CORRELATION CODE:**

**Code:**

**import numpy as np**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**numeric\_dataset = dataset.select\_dtypes(include=['number'])**

**corr = numeric\_dataset.corr()**

**import matplotlib.pyplot as plt**

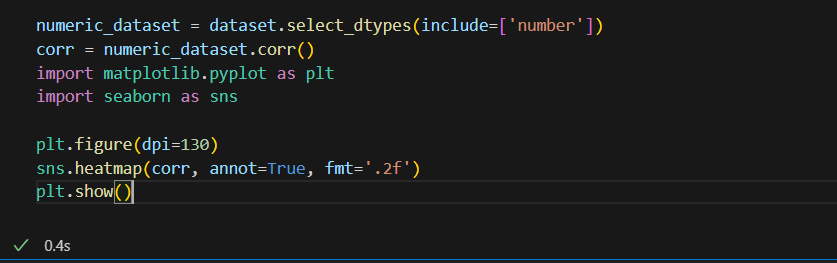
**import seaborn as sns**

**plt.figure(dpi=130)**

**sns.heatmap(corr, annot=True, fmt='.2f')**

**plt.show()**

**Output:**

****

****The data contains the following information:****

* ****Country** - this is the country for which the vaccination information is provided;**
* ****Country ISO Code** - ISO code for the country;**

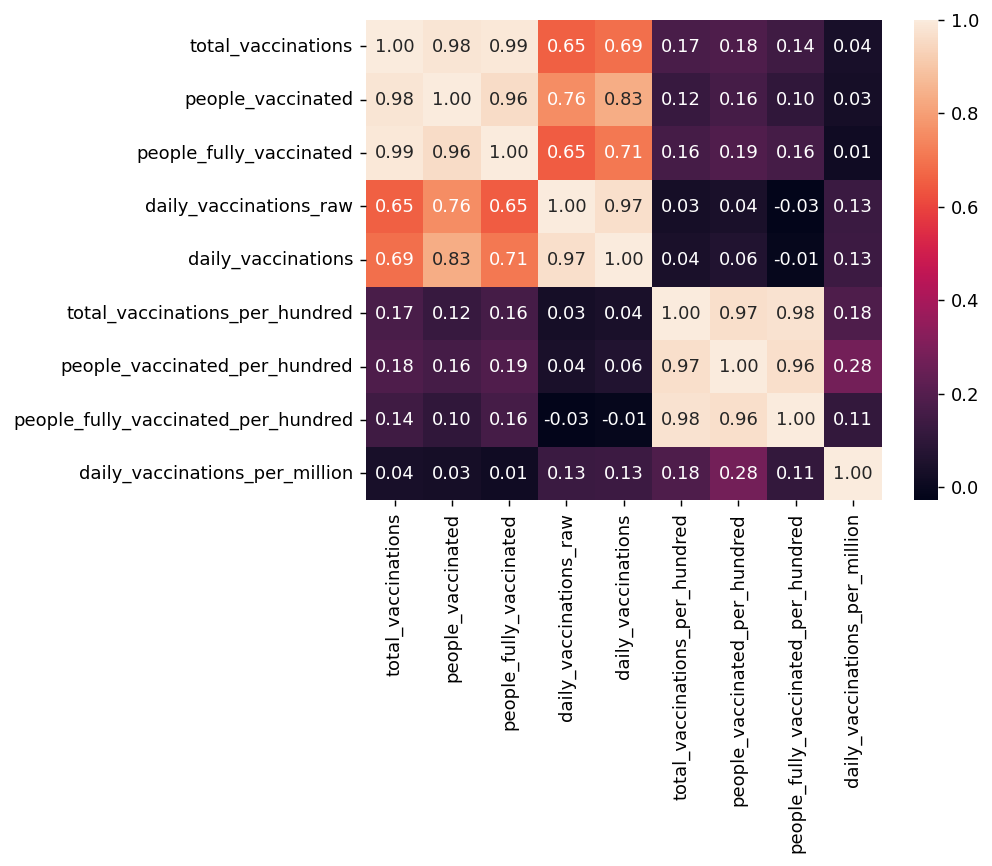
**ABOUT DATASET IMPLEMENTATION:**

****The data contains the 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 vaccine actions** - 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;**

**CONCLUSION:**

**CONCLUSION:**

**In conclusion, the development and distribution of COVID-19 vaccines have been pivotal in the global effort to combat the pandemic. These vaccines have demonstrated their effectiveness in reducing the spread of the virus and preventing severe illness. While challenges and questions have arisen, ongoing research and monitoring are essential to address them. The importance of equitable access to vaccines, public health measures, and international cooperation in the fight against COVID-19 cannot be overstated. It remains a dynamic situation, and continued vigilance and adaptation are key to our ongoing response to the pandemic.**