**Project Title: COVID-19 Vaccines Analysis**

**Problem Statement**

The primary goal of this project is to analyze the effectiveness and impact of various COVID-19 vaccines. The analysis involves exploring the efficacy, distribution, side effects, and overall performance of different vaccines in combating the spread of the virus. The project aims to provide insights to aid in better vaccine deployment strategies, understand the real-world effectiveness, and address concerns related to vaccination programs.

**Design Thinking Process**

**1. Understanding:** Initial research and understanding of available data, vaccine types, and their manufacturers.

**2. Ideation:** Identifying key metrics for analysis such as vaccination rates, adverse effects, efficacy, and distribution.

**3. Prototyping:** Developing data models, performing exploratory analysis, and assessing the feasibility of different statistical and machine learning techniques.

**4. Testing:** Validating the results, refining models, and ensuring the accuracy of findings.

**5. Implementation:** Presenting findings and providing actionable insights and recommendations.

**Phases of Development**

**1. Data Collection:** Gathering data from reliable sources including health organizations, government reports, and vaccine manufacturers.

**2. Data Preprocessing:** Cleaning the data, handling missing values, standardizing formats, and merging datasets if required.

**3. Exploratory Data Analysis (EDA):** Understanding the data through visualizations, identifying trends, correlations, and potential areas for deeper analysis.

**4. Statistical Analysis and Modeling:** Applying statistical methods and machine learning algorithms to extract insights and predict trends.

**5. Evaluation and Interpretation:** Assessing the models' performance and interpreting the results in the context of the problem statement.

**6. Reporting and Visualization:** Presenting findings through reports, visualizations, and actionable recommendations.

**Dataset Used**

The dataset encompasses information on vaccine types, manufacturers, administration rates, adverse reactions, efficacy rates, geographical distribution, and demographic factors. Sources include official health department reports, peer-reviewed studies, and public repositories such as CDC, WHO, and scientific journals.

**Data Preprocessing Steps**

- Handling missing or inconsistent data by imputation or removal.

- Standardizing and normalizing data for consistency.

- Cleaning outliers and irrelevant information.

- Merging and aggregating data from various sources for comprehensive analysis.

**Analysis Techniques Applied**

- Descriptive statistics for an initial overview of the data.

- Correlation and regression analysis to identify relationships between variables.

- Time series analysis to understand trends in vaccination rates.

- Machine learning algorithms for predictive modeling and clustering analysis to identify patterns.

**Key Findings and Insights**

**1.Efficacy Variation:** Different vaccines show varying efficacy rates against different strains and demographic groups.

**2. Adverse Effects Analysis:** Understanding and quantifying the adverse effects associated with specific vaccines.

**3. Geographical Disparities:** Discrepancies in vaccine distribution and administration rates across regions.

**4. Temporal Trends**: Analysis of the evolution of vaccination rates and their impact on infection rates.

**Recommendations**

**1. Optimized Distribution Strategies:** Target areas with lower vaccination rates for increased coverage.

**2. Surveillance and Monitoring:** Continual monitoring of adverse effects and efficacy rates for timely interventions.

**3. Public Awareness Campaigns:** Educate the population about vaccine benefits and dispel misinformation.

**PROGRAM:**

# Import necessary libraries import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns

# Load your dataset (replace 'data.csv' with your actual data file) df = pd.read\_csv('country\_vaccinations.csv')

# Display the first few rows of the dataset\ print("\n\nPerforming exploratory data analysis") print(df.head())

# Summary statistics print("\n\nStatistical analysis") summary\_stats = df.describe() print(summary\_stats)

# Check for missing values missing\_values = df.isnull().sum() print(missing\_values)

# Data visualization

# Example: Create a histogram of a numerical column print("\n\nVisualization") plt.figure(figsize=(8, 6)) sns.histplot(data=df, x='country', bins=20, kde=True) plt.title('Covid Analysis') plt.xlabel('Values') plt.ylabel('Frequency') plt.show()

# Example: Create a pair plot to visualize relationships between numerical variables sns.pairplot(data=df) plt.show()

# Example: Create a correlation heatmap correlation\_matrix = df.corr() plt.figure(figsize=(10, 8))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm')

plt.title('Correlation Heatmap') plt.show()

**OUTPUT:**

# Performing exploratory data analysis

country iso\_code date total\_vaccinations people\_vaccinated \ 0 Afghanistan AFG 2021-02-22 0.0 0.0

1. Afghanistan AFG 2021-02-23 NaN NaN
2. Afghanistan AFG 2021-02-24 NaN NaN
3. Afghanistan AFG 2021-02-25 NaN NaN 4 Afghanistan AFG 2021-02-26 NaN NaN

people\_fully\_vaccinated daily\_vaccinations\_raw daily\_vaccinations \ 0 NaN NaN NaN

1. NaN NaN 1367.0
2. NaN NaN 1367.0
3. NaN NaN 1367.0 4 NaN NaN 1367.0 total\_vaccinations\_per\_hundred people\_vaccinated\_per\_hundred \ 0 0.0 0.0
4. NaN NaN
5. NaN NaN
6. NaN NaN 4 NaN NaN

people\_fully\_vaccinated\_per\_hundred daily\_vaccinations\_per\_million \ 0 NaN NaN

1. NaN 34.0
2. NaN 34.0
3. NaN 34.0 4 NaN 34.0

vaccines \ 0 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...

1. Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
2. Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
3. Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
4. Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...

source\_name source\_website 0 World Health Organization <https://covid19.who.int/>

1. World Health Organization <https://covid19.who.int/>
2. World Health Organization <https://covid19.who.int/>
3. World Health Organization <https://covid19.who.int/>
4. World Health Organization <https://covid19.who.int/>

# Statistical analysis

total\_vaccinations people\_vaccinated people\_fully\_vaccinated \ count 4.360700e+04 4.129400e+04 3.880200e+04 mean 4.592964e+07 1.770508e+07 1.413830e+07 std 2.246004e+08 7.078731e+07 5.713920e+07 min 0.000000e+00 0.000000e+00 1.000000e+00 25% 5.264100e+05 3.494642e+05 2.439622e+05

50% 3.590096e+06 2.187310e+06 1.722140e+06 75% 1.701230e+07 9.152520e+06 7.559870e+06 max 3.263129e+09 1.275541e+09 1.240777e+09

daily\_vaccinations\_raw daily\_vaccinations \ count 3.536200e+04 8.621300e+04 mean 2.705996e+05 1.313055e+05 std 1.212427e+06 7.682388e+05 min 0.000000e+00 0.000000e+00 25% 4.668000e+03 9.000000e+02

50% 2.530900e+04 7.343000e+03 75% 1.234925e+05 4.409800e+04 max 2.474100e+07 2.242429e+07

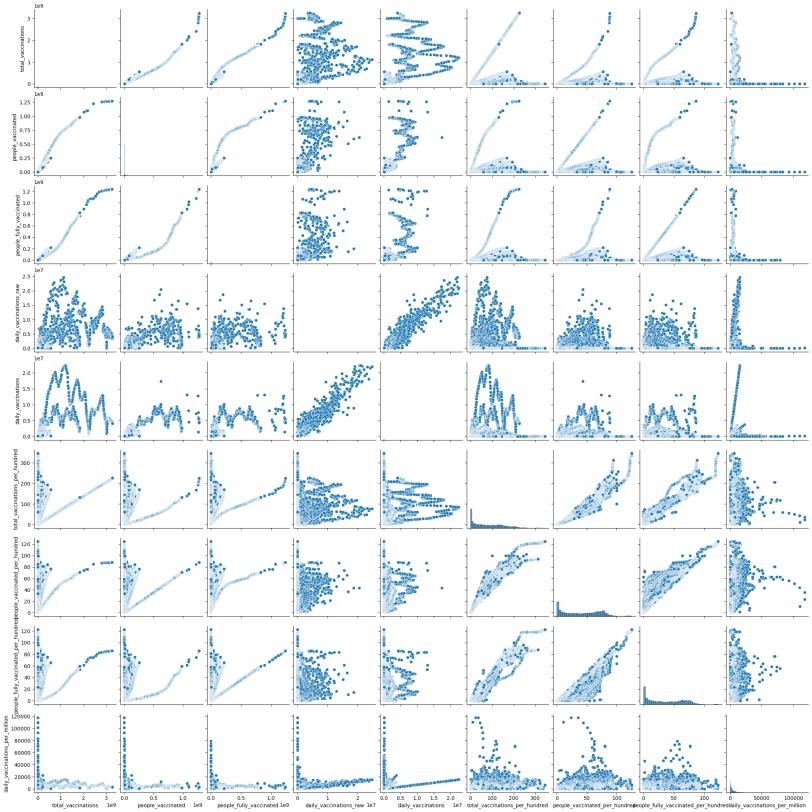
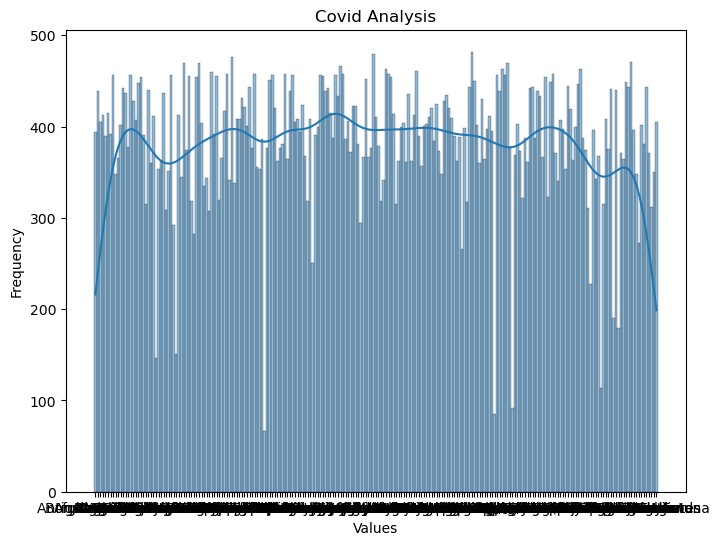
total\_vaccinations\_per\_hundred people\_vaccinated\_per\_hundred \ count 43607.000000 41294.000000 mean 80.188543 40.927317 std 67.913577 29.290759 min 0.000000 0.000000 25% 16.050000 11.370000

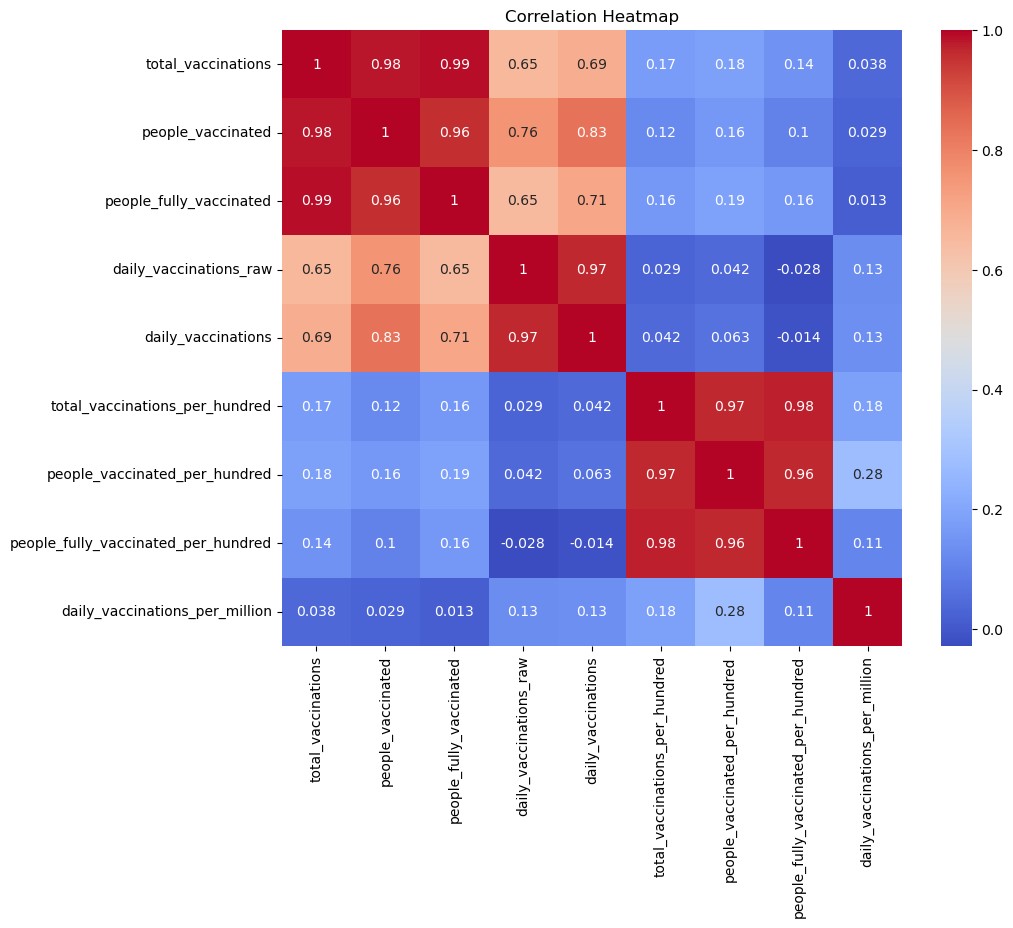
50% 67.520000 41.435000 75% 132.735000 67.910000 max 345.370000 124.760000

people\_fully\_vaccinated\_per\_hundred daily\_vaccinations\_per\_million count 38802.000000 86213.000000 mean 35.523243 3257.049157 std 28.376252 3934.312440 min 0.000000 0.000000 25% 7.020000 636.000000

50% 31.750000 2050.000000 75% 62.080000 4682.000000 max 122.370000 117497.000000 country 0 iso\_code 0 date 0 total\_vaccinations 42905 people\_vaccinated 45218 people\_fully\_vaccinated 47710 daily\_vaccinations\_raw 51150 daily\_vaccinations 299 total\_vaccinations\_per\_hundred 42905 people\_vaccinated\_per\_hundred 45218 people\_fully\_vaccinated\_per\_hundred 47710 daily\_vaccinations\_per\_million 299 vaccines 0 source\_name 0 source\_website 0 dtype: int64

**Visualization**





**Dataset link:**

[**https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress**](https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress)

**GitHub link:** [**https://github.com/sivasubramanianss40/NaanMudhalvan\_IBM.git**](https://github.com/sivasubramanianss40/NaanMudhalvan_IBM.git)

**Conclusion**

The COVID-19 vaccines analysis project provides a comprehensive understanding of vaccine effectiveness, distribution, and impact. The insights gained can aid policymakers, healthcare professionals, and communities in better understanding the dynamics of vaccination efforts and optimizing strategies to combat the pandemic.