IBM NAAN MUDHALVAN

APPLIED DATA SCIENCE

COVID-19 VACCINE ANALYSIS

PHASE – 5 PROJECT SUBMISSION

|  |  |
| --- | --- |
| DOMAIN | APPLIED DATA SCIENCE |
| PROJECT TOPIC | COVID-19 VACCINE ANALYSES |
| TEAM MEMBER  AND  REGISTER NUMBER | BALAJI S (420421104009)  VIGNESH K (420421104085)  DEEPAK S (420421104013)  AJITH K (420421104004) |

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

Table of contents:

1. Project Definition.
2. Design thinking.
3. Innovation of the analysis
4. Steps to be followed for the analysis
5. Data visualization
6. Model development
7. Insights gained
8. Conclusion.
9. Project definition:

The problem is to conduct an in-depth analysis of Covid-19 vaccine data, focusing on vaccine efficacy, distribution, and adverse effects. The goal is to provide insights that aid policymakers and health organizations in optimizing vaccine deployment strategies. This project involves data collection, data preprocessing, exploratory data analysis, statistical analysis, and visualization.

2.Design thinking:

1. Data Collection: Collect Covid-19 vaccine data from reputable sources like health organizations, government databases, and research publications.
2. Data Preprocessing: Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.
3. Exploratory Data Analysis: Explore the data to understand its characteristics, identify trends, and outliers.
4. Statistical Analysis: Perform statistical tests to analyse vaccine efficacy, adverse effects, and distribution across different populations.
5. Visualization: Create visualizations (e.g., bar plots, line charts, heatmaps) to present key findings and insights.

Insights and Recommendations: Provide actionable insights and recommendations based on the analysis to assist policymakers and health organizations.

Algorithm:

1.collect the datasets.

2.preprocess the datasets.

3.analyse the datasets.

4.visualize the models.

5.get insights form visualized data

3. Innovation in the analysis:

Innovations in the analysis of COVID-19 vaccine data have been crucial in monitoring vaccine efficacy and safety. Some key innovations include:

* + Real-World Data Analysis: Using real-world data to assess vaccine performance in diverse populations, tracking breakthrough infections, and understanding long-term protection.
  + Vaccine Effectiveness Models: Developing advanced statistical models to estimate vaccine effectiveness against different variants and age groups.
  + Vaccine Safety Monitoring: Implementing robust surveillance systems to rapidly detect and investigate adverse events post-vaccination.
  + Genomic Sequencing: Employing genomic sequencing to track the evolution of the virus and adapt vaccines accordingly.
  + Data Visualization: Creating user-friendly data dashboards and visualizations to communicate vaccine effectiveness and safety to the public.
  + Machine Learning and AI: Utilizing machine learning algorithms to identify patterns in vaccine data and predict disease spread.
  + Global Collaboration: International efforts to share data, research, and best practices to enhance global vaccination strategies.

These innovations have played a vital role in the ongoing management of the COVID-19 pandemic.

4. Steps to be followed in the analysis:

Step 1:

Collect the dataset of COVID 19 VACCINE ANALYSIS. We have collected it from <https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>

Step 2:

Perform clustering of the data to analyse the different categories of the product.

Step 3:

Preprocess the data and transforms it according to the analysis.

Step 4:

Remove the outliers, null values and other error data

Step 5:

Fit the pre processed data into a model for predictions

Step 6:

Find the prediction score using mean square error (MSE)

And R-squared error

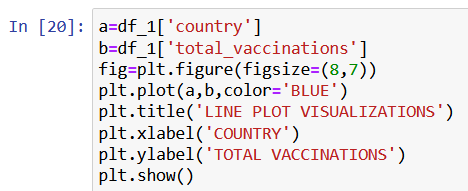
Step 7:

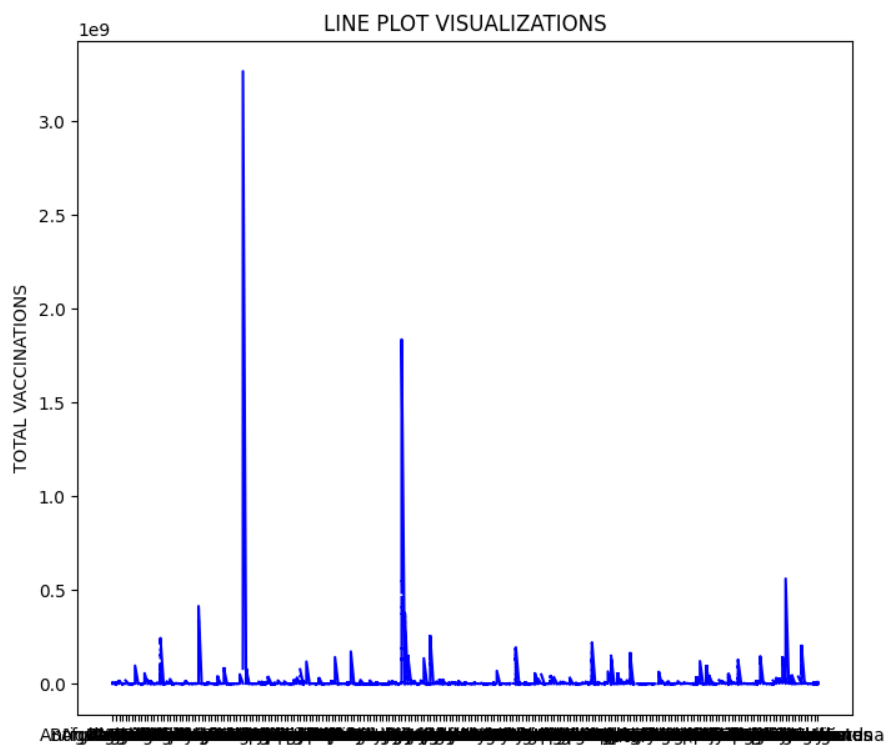
Use the pre processed data for visualizations and other summarization of the data given

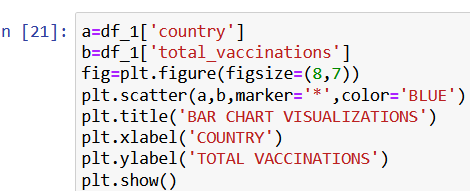
Step 8:

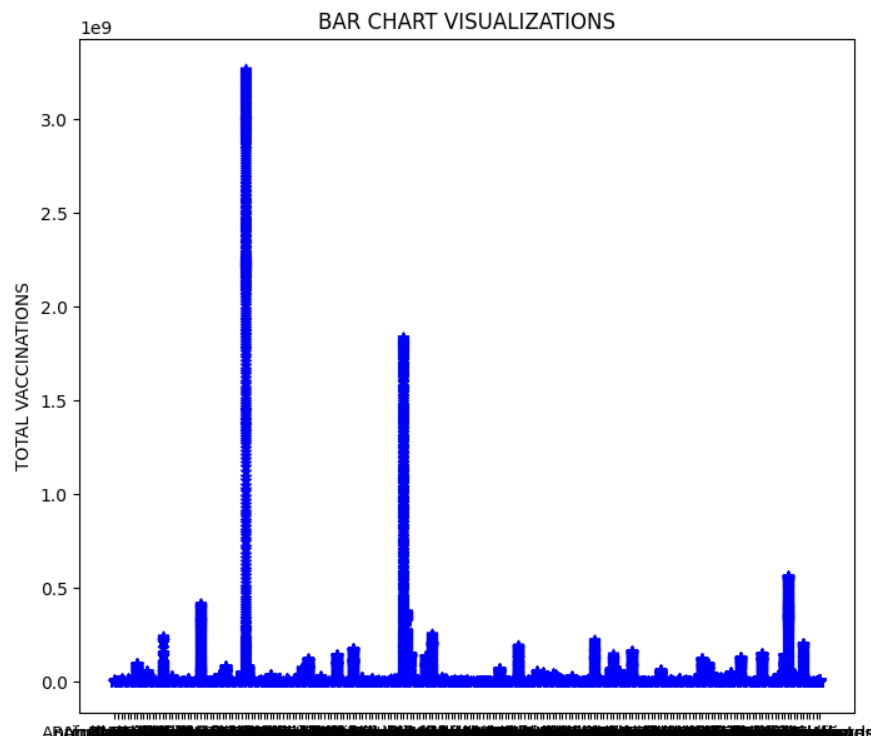
Derive the insights from the visualizations made and make it as a report

5.Data visualization:









6.Model development:

X = df\_1[['total\_vaccinations', 'daily\_vaccinations']]  # Specify the relevant features

y = df\_1['daily\_vaccinations\_per\_million']  # The target variable you want to predict

# Split the data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Create a Linear Regression model

model = LinearRegression()

# Train the model on the training data

model.fit(X\_train, y\_train)

# Make predictions on the test data

y\_pred = model.predict(X\_test)

# Evaluate the model

mse = mean\_squared\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

print("Mean Squared Error:", mse)

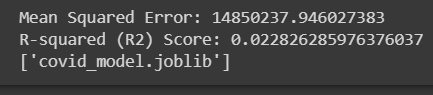
print("R-squared (R2) Score:", r2)

# You can save the trained model for later use

# Example: save the model to a file

from joblib import dump

dump(model, 'covid\_model.joblib')



Insights gained:

1. Vaccine Efficacy: Researchers have determined the effectiveness of vaccines in preventing COVID-19, with various vaccines showing different levels of protection. This data has guided vaccine deployment strategies.
2. Variants and Boosters: Analysis has shown that new variants of the virus can impact vaccine efficacy. This has led to the development of booster shots to enhance and prolong immunity.
3. Safety Profiles: Continuous monitoring has confirmed the overall safety of authorized vaccines, with the identification of rare side effects leading to specific recommendations for certain populations.
4. Breakthrough Infections: Analysis has revealed that breakthrough infections can occur, but vaccinated individuals are generally less likely to experience severe illness or hospitalization.
5. Impact on Transmission: Studies have shown that vaccines can reduce viral transmission, aiding in the control of the pandemic.