IBM Applied Data Science Project
Covid-19 Vaccine Analysis
PHASE 1

Problem Definition:

The problem is to conduct an in-depth analysis of COVID-19 vaccine data involves several steps. Below is a comprehensive solution outlining the process, methodologies, and tools that can be utilized to achieve the goal of providing actionable insights for optimizing vaccine deployment strategies.

1. Data Collection:

Sources: Collect data from reliable sources such as government health departments, World Health Organization (WHO), and reputable research institutions. Utilize APIs or web scraping techniques to gather real-time data.

2. Data Preprocessing:

Data Cleaning: Handle missing or inconsistent data points. Standardize data formats, resolve discrepancies, and clean textual data.

Data Integration: Integrate data from different sources to create a unified dataset for analysis.

Feature Engineering: Derive additional features such as vaccination rates, adverse event ratios, and regional demographics if necessary.

3. Exploratory Data Analysis (EDA):

Descriptive Statistics: Calculate mean, median, variance, and other relevant statistics to understand the basic properties of the data.

Data Visualization: Utilize histograms, box plots, and heatmaps to visualize the distribution of vaccination rates, adverse effects, and efficacy across different demographics and regions.

Correlation Analysis: Explore correlations between vaccination rates, adverse effects, and demographic factors. Use scatter plots and correlation matrices for visualization.

4. Statistical Analysis:

Hypothesis Testing: Conduct t-tests or ANOVA to compare vaccine efficacy rates between different age groups, genders, or regions.

Regression Analysis: Perform regression analysis to identify factors influencing vaccine efficacy and adverse effects. Multiple regression can be used for a more comprehensive analysis.

Survival Analysis: If applicable, conduct survival analysis to understand the duration of vaccine efficacy and adverse effects.

5. Data Visualization:

Bar Plots: Visualize vaccination rates across different age groups, regions, and socioeconomic factors.

Line Charts: Display trends in vaccine efficacy and adverse effects over time.

Heatmaps: Show correlations between vaccination rates, adverse effects, and demographic variables.

Geospatial Maps: Represent regional vaccination rates and highlight areas with low coverage.

Stacked Bar Charts: Visualize the composition of adverse effects by severity level or age group.

6. Insights and Recommendations:

Identify Patterns: Identify patterns such as demographic groups with low vaccination rates, regions with high adverse effects, or specific adverse effects associated with certain vaccine types.

Recommendations: Based on the identified patterns, provide actionable recommendations. For instance, if a specific demographic group shows lower vaccination rates, recommend targeted awareness campaigns. If certain regions report higher adverse effects, recommend thorough investigations and support services in those areas.

Tools and Technologies:

Data Analysis: Python (Pandas, NumPy), R, SQL

Data Visualization: Matplotlib, Seaborn, Plotly, Tableau

Statistical Analysis: SciPy, StatsModels

Machine Learning (if applicable): Scikit-Learn, TensorFlow, XGBoost

Important Considerations:

Ethical and Privacy Considerations: Ensure that data is anonymized and privacy regulations are adhered to during the entire analysis process.

Continuous Monitoring: Establish a system for continuous monitoring to update insights as new data becomes available.

By following this structured approach and utilizing appropriate tools and methodologies, you can conduct a robust analysis of COVID-19 vaccine data, providing valuable insights to aid policymakers and health organizations in optimizing vaccine deployment strategies effectively.

Design Thinking:

Data Collection:

1. Data Sources: Utilize reputable sources like WHO, CDC, and official government databases for reliable COVID-19 vaccine data.

2.Consistency: Maintain consistent data collection methods to ensure accuracy and

3. Vaccination Data: Focus on gathering information on doses administered, coverage rates, and vaccine types used to assess the impact of vaccination campaigns.

4. Demographic Information: Include factors like age, gender, and socio-economic status to analyse differential impacts on populations.

5. Collaboration: Foster collaboration between researchers, public health agencies, and data scientists to enhance data quality and reliability.

By following these guidelines, one can ensure robust data collection and analysis for insights into the COVID-19 pandemic.

Data Preprocessing:

1. Handling Missing Values:

- Identify missing values in vaccination-related columns.
- Options for handling:
- Remove rows with limited missing values.
- Impute missing values using mean, median, or machine learning methods.

2. Cleaning Data:

- Address outliers or inconsistent values to avoid skewing analysis.
- Check for duplicates and remove them to ensure data integrity.

3. Categorical to Numerical Conversion:

- Identify categorical features (e.g., gender, region).
- Utilize techniques like one-hot encoding to convert them into numerical representations, enabling machine learning algorithms to process them.

4. Data Splitting:

- Split the dataset into training and testing sets to evaluate model performance.

5. Exploratory Data Analysis (EDA):

- Conduct EDA to gain insights into the data distribution, correlations, and patterns.

COVID Vaccination EDA

In COVID-19 vaccination analysis, exploratory data analysis (EDA) involves examining key metrics such as vaccination rates, distribution, and demographic patterns. Identifying trends may include observing changes over time, vaccine effectiveness, and regional variations. Outlier analysis could highlight areas with exceptionally high or low vaccination rates, helping public health officials target interventions. EDA can provide valuable insights into the progress of vaccination campaigns and guide evidence-based decision-making.

Statistical Analysis:

1. Vaccine Efficacy:

Vaccine efficacy is usually determined through clinical trials where vaccinated and control groups are compared. The formula for vaccine efficacy is:

$$Efficacy = (1 - (\text{Attack rate in unvaccinated group} \,\%\, \text{Attack rate in vaccinated group})) \times 100\%.$$

2. Adverse Effects:

Analysing adverse effects involves understanding the frequency and severity of side effects in vaccinated individuals

3. Distribution Across Different Populations:

Understanding how vaccines are distributed among different populations is crucial for equitable healthcare. Statistical techniques include:

a. Analysis of Variance (ANOVA):

ANOVA can be used to compare vaccine distribution means across multiple demographic groups, such as age, ethnicity, or geographical location.

b. Regression Analysis:

Regression models can be used to predict vaccine distribution based on various factors,

Considerations:

Sample Size: Ensure your sample sizes are adequate for the statistical tests being used. Small samples can lead to unreliable results.

Confounding Variables: Be aware of confounding variables that might influence the results. Statistical methods like multivariate analysis can help control for these variables.

Longitudinal Analysis: For adverse effects, especially, consider analysing data over time to understand trends and potential delayed reactions.

Always consult with a statistician or data scientist when conducting complex statistical analyses to ensure the methods used are appropriate for your specific dataset and research questions.

Visualization:

Here's how we can create different types of visualizations to present your vaccine-related data:

1. Bar Plots:

Bar plots are excellent for comparing categorical data such as vaccine distribution across different populations.

Example:

A bar plot comparing the number of vaccinated individuals in different age groups can provide a clear visual comparison.

2. Line Charts:

Line charts are effective for showing trends over time, such as vaccine efficacy rates or adverse effects occurrences over several weeks or months.

Example:

A line chart can demonstrate how vaccine efficacy rates change over time after the vaccine is administered.

3. Stacked Bar Charts:

Stacked bar charts are useful for displaying the composition of a whole, such as the distribution of adverse effects by severity level.

Example:

A stacked bar chart can represent the total adverse effects, broken down into categories like mild, moderate, and severe, for different vaccine types.

4. Pie Charts:

While not the most recommended visualization due to potential misinterpretation, pie charts can be effective for showing parts of a whole, such as the proportion of adverse effects attributed to different symptoms.

Example:

A pie chart can represent the percentage of total adverse effects, with each slice representing a specific symptom like fever, headache, or fatigue.

For Effective Visualization:

1. Simplicity: Keep visualizations simple and easy to interpret. Avoid clutter and unnecessary decorations.

2. Clarity: Use clear labels, legends, and titles. Ensure the audience understands what the visualization represents.

3. Consistency: Use consistent colours and styles across different visualizations for easy comparison.

4. Context: Provide context and background information to help viewers understand the significance of the visualized data.

5. Interactivity: If possible, create interactive visualizations that allow users to explore the data on their own.

By combining statistical analyses with these visualizations, you can effectively communicate your findings and provide valuable insights into vaccine efficacy, adverse effects, and distribution patterns across different populations.

Insights and Recommendations:

Here some insights and corresponding recommendations based on the analysis conducted:

1. Vaccine Efficacy Insights:

Insight: The vaccine demonstrates varying efficacy rates across different age groups.

Recommendation: Prioritize vaccination campaigns targeting age groups with lower efficacy rates. Additionally, invest in research to understand the reasons behind these variations and develop strategies to enhance efficacy, if possible.

2. Adverse Effects Insights:

Insight: Certain demographic groups are more susceptible to specific adverse effects.

Recommendation: Tailor vaccine information and support services for these groups. Provide detailed information about potential adverse effects and offer accessible healthcare services for those experiencing side effects, ensuring a rapid response to any severe reactions.

3. Distribution Across Different Populations Insights:

Insight: There are geographical disparities in vaccine distribution, with certain regions having lower vaccination rates.

Recommendation: Implement targeted outreach programs in underserved regions. This could include setting up mobile vaccination clinics, community engagement initiatives, and partnerships with local organizations to increase awareness and access to vaccination.

4. Overall Recommendations:

1. Continuous Monitoring: Establish a robust system for continuous monitoring of vaccine efficacy and adverse effects. Regularly update policymakers and healthcare providers with the latest findings to inform decision-making.

2. Public Awareness Campaigns: Invest in public awareness campaigns emphasizing the importance of vaccination, addressing concerns about adverse effects, and providing accurate information about vaccine efficacy.

3. Data Sharing and Collaboration: Foster collaboration between healthcare organizations, researchers, and policymakers to share data and insights. Collaborative efforts can lead to more effective strategies and interventions.

4. Research and Development: Allocate resources for ongoing research to improve vaccine formulations, enhance efficacy, and minimize adverse effects. Support research on booster doses to maintain long-term immunity.

5. Global Collaboration: Collaborate with international health organizations to ensure equitable vaccine distribution on a global scale. Support initiatives aimed at providing vaccines to low- and middle-income countries to achieve global immunity and prevent the spread of new variants.

IBM Applied Data Science Project
Covid-19 Vaccine Analysis
PHASE 2

Innovation:

Innovation Idea: "VaxTrackr - SmartVaccine Management Platform"

To further enhance the design thinking approach described above, we propose an innovative solution called "VaxTrackr." VaxTrackr is a comprehensive, technology-driven platform designed to optimize COVID-19 vaccine deployment strategies while ensuring transparency, accessibility, and data-driven decision-making. Here's how VaxTrackr works:

1. Unified Data Hub

VaxTrackr integrates data from multiple sources, including WHO, CDC, government databases, healthcare providers, and users' vaccination records, into a unified data hub.

2. Personalized Vaccination Profiles

Users can create personalized vaccination profiles by securely inputting their vaccination data. The platform provides users with a digital vaccination certificate.

3. Real-Time Dashboard

VaxTrackr offers a real-time, user-friendly dashboard accessible via web and mobile devices. Users can view live vaccination data, including coverage rates, distribution, and demographics.

4. Geo-Location Services

The platform uses geo-location services to help users find nearby vaccination centers, check wait times, and schedule appointments.

5. Predictive Analytics

VaxTrackr employs predictive analytics to estimate vaccine supply and demand, helping policymakers allocate doses efficiently.

6. Vaccine Passport Integration:

Integrates with international vaccine passport standards to facilitate travel for vaccinated individuals.

7. Gamification Elements

Gamification elements encourage users to stay engaged, complete their vaccination series, and share their vaccination journeys with others.

8. Community Engagement

Users can participate in forums, Q&A sessions with healthcare experts, and live webinars to foster a sense of community and knowledge-sharing.

9. AI Chatbot Support

Offers an AI chatbot for answering common questions related to vaccines, adverse effects, and safety concerns.

10. Data Privacy and Security

Ensures robust data privacy and security measures, including end-to-end encryption of user data and adherence to privacy regulations.

11. Supply Chain Monitoring

Tracks vaccine supply chains, from production to distribution, to identify bottlenecks and ensure a consistent supply of vaccines.

12. Adverse Effects Monitoring

Implements a system for users to report adverse effects post-vaccination. This data is analyzed to ensure vaccine safety and enable rapid responses to severe reactions.

13. AI-Driven Recommendations

Utilizes AI to provide personalized vaccine-related recommendations based on user profiles, demographics, and local COVID-19 conditions.

14. Public Awareness Campaigns

Launches public awareness campaigns through the platform, including educational content, success stories, and testimonials from vaccinated individuals.

15. International Collaboration

- Promotes international collaboration by sharing anonymized data and insights with global health organizations for a coordinated response to the pandemic.

16. Continuous Improvement

VaxTrackr regularly updates its features based on user feedback and emerging vaccine-related developments, ensuring its relevance and effectiveness.

1. Unified Data Hub:

Explanation:

The Unified Data Hub acts as the central data repository where data from multiple sources is collected, integrated, and made ready for analysis. In the context of applied data science, this feature forms the foundation for data-driven decision-making.

Features:

- Data Integration: VaxTrackr integrates data from WHO, CDC, government databases, healthcare providers, and user vaccination records.
- Data Cleaning and Transformation: It performs data preprocessing to clean and transform data for consistency.
- Data Storage: Utilizes databases to store integrated data securely.

Tools:

- ETL (Extract, Transform, Load) Frameworks: Tools like Apache Nifi, Talend, or Apache Spark for data integration and transformation.
- Data Warehousing: Solutions such as Amazon Redshift, Google BigQuery, or Snowflake for storing and querying large datasets.
- Data Cleaning Libraries: Python libraries like Pandas for data cleaning and preprocessing.

2. Personalized Vaccination Profiles:

Explanation:

The creation of personalized vaccination profiles involves data collection, user-specific analysis, and the generation of digital vaccination certificates.

Features:

- User Data Collection: Collects user information, including age, demographics, and vaccination data.
- Data Analysis: Utilizes data science techniques to analyze user profiles and vaccination data.
- Certificate Generation: Generates personalized digital vaccination certificates based on user profiles.

Tools:

- Machine Learning Models: Machine learning models can be used to analyze user data for personalized recommendations.
- Data Visualization Tools: Tools like Matplotlib or Tableau for creating visualizations from user profiles.
- PDF Generation Libraries: Libraries like ReportLab or PyPDF2 for generating digital certificates.

3. Real-Time Dashboard:

Explanation:

The Real-Time Dashboard is designed to provide live vaccination data to users, allowing them to monitor coverage rates, distribution, and demographics.

Features:

- Real-Time Data Collection: Gathers real-time data from various sources and databases.
- Data Visualization: Uses data science techniques to visualize live data.
- User-Specific Insights: Provides users with personalized insights based on their vaccination status.

Tools:

- Real-Time Data Streaming: Technologies like Apache Kafka or Amazon Kinesis for streaming real-time data.
- Data Visualization Libraries: Libraries like Plotly or Seaborn for creating interactive data visualizations.
- Personalization Algorithms: Machine learning algorithms for personalizing insights based on user data.

4. Geo-Location Services:

Explanation:

Geo-location services enhance user experience by helping them locate nearby vaccination centers, check wait times, and schedule appointments.

Features:

- Location-Based Services: Uses geospatial data to provide location-based information.
- Real-Time Data Integration: Integrates real-time data on vaccination centers and wait times.
- Appointment Scheduling: Allows users to book vaccination appointments at nearby centers.

Tools:

- Geospatial Data Libraries: Libraries like GeoPandas for handling geospatial data.

- Real-Time APIs: Utilizes APIs for accessing real-time location data.
- Appointment Booking Software: Integration with appointment scheduling software.

5.Predictive Analytics:

Explanation:

Predictive Analytics is a powerful tool that utilizes historical vaccination data and current trends to forecast vaccine supply and demand. By doing so, it assists policymakers in optimizing vaccine allocation and distribution strategies.

Features:

- Demand Forecasting: Utilizes historical data and population trends to estimate vaccine demand.
- Supply Projections: Analyzes vaccine production and distribution patterns to predict future supply.
- Inventory Management: Helps health authorities maintain appropriate vaccine inventory levels to avoid shortages or wastage.

Tools:

- Machine Learning Models: Implements machine learning algorithms for demand and supply forecasting.
- Data Mining Tools: Tools like RapidMiner or KNIME for extracting insights from historical data.
- Dashboard Integration: Integrates predictive analytics results into the real-time dashboard for policymakers.

6. Vaccine Passport Integration:

Explanation:

Vaccine Passport Integration ensures that vaccinated individuals can seamlessly access and utilize their vaccination status for various purposes, including international travel. This feature adheres to global vaccine passport standards.

Features:

- Digital Passport Generation: Generates digital vaccine passports containing user vaccination details.
- QR Code Integration: Utilizes QR codes for easy scanning and verification.
- Compliance with Standards: Adheres to international standards and regulations for vaccine passports.

Tools:

- QR Code Generation Libraries: Libraries like ZXing for generating QR codes.
- Cryptographic Tools: Implements encryption and security measures to protect the integrity of digital passports.
- API Integration: Integrates with travel booking platforms and government systems for seamless verification.

7. Gamification Elements:

Explanation:

Gamification elements are introduced to make the vaccination process engaging and rewarding for users. It encourages them to stay on track with their vaccination schedule.

Features:

- Point System: Awards users points or rewards for each vaccination milestone achieved.
- Virtual Badges and Trophies: Recognizes and rewards users with virtual badges for specific achievements.
- Social Sharing: Allows users to share their vaccination progress on social media platforms.

Tools:

- Gamification Platforms: Utilizes platforms like Gamify or Bunchball for building gamification elements.
- Social Media APIs: Integrates with social media platforms to enable users to share their vaccination milestones.
- User Engagement Analytics: Tracks user engagement and provides insights for further gamification improvements.

8. Community Engagement:

Explanation:

Community Engagement aims to create a sense of belonging and knowledge-sharing among users. It facilitates interactions through forums, Q&A sessions, and live webinars with healthcare experts.

Features:

- Discussion Forums: Allows users to participate in discussions, ask questions, and share their experiences.
- Q&A Sessions: Hosts periodic question and answer sessions with healthcare professionals.
- Live Webinars: Conducts live webinars on vaccination-related topics.

Tools:

- Forum Platforms: Integrates forum software like Discourse or phpBB.
- Webinar Platforms: Utilizes webinar tools like Zoom or GoToWebinar for hosting live sessions.
- Chat and Comment Analytics: Collects and analyzes user interactions to improve engagement and content.

9.AI Chatbot Support:

Explanation:

The AI Chatbot plays a crucial role in providing users with quick and accurate information related to vaccines, adverse effects, and safety concerns using natural language processing (NLP) and machine learning.

Features:

- NLP for Query Understanding: Utilizes NLP techniques to understand and interpret user queries in natural language.
- Machine Learning for Responses: Employs machine learning models to generate context-aware and accurate responses.
- Knowledge Base Integration: Integrates a comprehensive knowledge base of vaccine-related information.

Tools:

- NLP Libraries: Utilizes NLP libraries such as spaCy, NLTK, or Hugging Face Transformers for query understanding.
- Chatbot Frameworks: Implements chatbot frameworks like Rasa, Dialogflow, or Microsoft Bot Framework.
- Machine Learning Models: Employs machine learning models such as BERT, GPT, or custom-built models for generating responses.

10. Data Privacy and Security:

Explanation:

Data privacy and security are paramount. The platform ensures end-to-end encryption and complies with privacy regulations to safeguard user data.

Features:

- Data Encryption: Enforces strong encryption algorithms to protect user data during transmission and storage.
- Privacy Regulation Compliance: Adheres to privacy regulations such as GDPR, HIPAA, or CCPA, depending on the region.
- Access Control: Restricts access to user data to authorized personnel only.

Tools:

- Encryption Libraries: Employs encryption libraries like OpenSSL or Python's cryptography library for data encryption.
- Privacy Compliance Tools: Uses tools for managing and demonstrating compliance with relevant privacy regulations.
- Identity and Access Management (IAM): Implements IAM solutions such as AWS Identity and Access Management for access control.

11. Supply Chain Monitoring:

Explanation:

Supply Chain Monitoring involves real-time tracking of vaccine supply chains, from production to distribution, to identify bottlenecks and ensure a consistent vaccine supply.

Features:

- Real-Time Supply Data: Collects real-time data on vaccine production, transportation, and distribution.
- Predictive Analytics: Utilizes predictive analytics to forecast vaccine supply and demand, enabling proactive supply adjustments.
- Bottleneck Identification: Identifies bottlenecks, delays, or issues within the supply chain that may affect vaccine availability.

Tools:

- IoT Devices: Implements IoT sensors and devices for real-time tracking and monitoring of vaccine shipments.

- Supply Chain Management Software: Utilizes supply chain management software like SAP Integrated Business Planning for optimizing supply chains.
- Predictive Analytics Platforms: Employs predictive analytics platforms like Microsoft Azure Machine Learning for demand forecasting.

12. Adverse Effects Monitoring:

Explanation:

Adverse Effects Monitoring is a system that allows users to report post-vaccination adverse effects. Data science is employed to analyze this data to ensure vaccine safety and enable rapid responses to severe reactions.

Features:

- User Reporting: Provides users with the ability to report adverse effects through the platform.
- Data Collection: Collects detailed information on the type and severity of adverse effects reported by users.
- Automated Analysis: Utilizes data science and machine learning to analyze reported adverse effects for patterns, trends, and safety concerns.

Tools:

- Reporting Forms: Offers online reporting forms with structured data fields for users to report adverse effects consistently.
- Data Analysis Platforms: Employs data analysis platforms like Python and R, leveraging libraries like pandas and scikit-learn for advanced analysis.
- Anomaly Detection Algorithms: Utilizes machine learning algorithms, including clustering and anomaly detection algorithms, for identifying unusual patterns in reported adverse effects.

13.AI-Driven Recommendations:

Explanation:

AI-Driven Recommendations in VaxTrackr utilize advanced artificial intelligence algorithms to offer personalized vaccine-related guidance. This feature enhances the vaccination experience by tailoring recommendations

based on individual user profiles, demographics, and real-time local COVID-19 conditions.

Features:

- User Profiles Analysis:

Description: Utilizes individual user profiles to understand demographics, medical history, and preferences.

- Real-Time Data Integration:

Description: Integrates real-time local COVID-19 data for up-to-date and relevant recommendations.

- Machine Learning Algorithms:

Description: Implements advanced machine learning models for personalized vaccine recommendations.

- Notification System:

Description: Sends timely notifications for vaccination updates, reminders, and personalized recommendations.

Tools:

- User Profiles Management System:

Description: Allows users to securely input and manage their demographic and health information.

- Real-Time Data Integration System:

Description: Connects with local health databases and real-time COVID-19 tracking systems.

Machine Learning Model Infrastructure:

Description: Hosts and runs advanced machine learning algorithms for personalized recommendations.

- Push Notification Service:

Description: Delivers notifications directly to users through the platform.

14.Public Awareness Campaigns:

Explanation:

Public Awareness Campaigns within VaxTrackr aim to educate and engage users by launching campaigns directly through the platform. This includes a variety of content such as educational materials, success stories, and testimonials from vaccinated individuals.

Features:

- Content Hub:

Description: Centralized hub for educational content, success stories, and testimonials.

- User-Generated Content Platform:

Description: Platform for users to share their vaccination experiences, creating a community-driven narrative.

- Interactive Content Tools:

Description: Tools for creating engaging content formats like quizzes and challenges.

- Community Forums:

Description: Online forums for users to share experiences, ask questions, and participate in discussions.

Tools:

- Content Management Platform:

Description: Manages the creation, organization, and dissemination of educational content.

- User-Generated Content Platform:

Description: Allows users to contribute and share their vaccination stories and experiences.

Interactive Content Creation Platform:

Description: Facilitates the creation of engaging content such as quizzes and challenges.

Community Forum Platform:

Description: Provides a space for users to interact, share knowledge, and ask questions.

15.International Collaboration:

Explanation:

The International Collaboration feature in VaxTrackr promotes global cooperation by sharing anonymized data and insights with global health organizations. This collaborative effort aims to achieve a coordinated response to the global pandemic.

Features:

- Data Sharing Protocols:

Description: Establishes secure protocols for sharing anonymized data with global health organizations.

- Global Insights Dashboard:

Description: Shared dashboard for international collaborators to access and analyze anonymized data.

- Collaboration Forums:

Description: Virtual spaces for discussions and collaboration between different global health entities.

Tools:

- Secure Data Sharing Protocols:

Description: Ensures compliance with privacy regulations and secure sharing of data.

- Collaborative Analytics Platform:

Description: Platform for collaborative data analysis and insights sharing.

- Virtual Collaboration Platform:

Description: Facilitates online discussions and collaboration among global health organizations.

16.Continuous Improvement:

Explanation:

Continuous Improvement is a foundational principle in VaxTrackr, ensuring that the platform evolves in response to user feedback and emerging vaccine-related developments. This commitment is vital to maintaining the platform's relevance, effectiveness, and user satisfaction over time.

Features:

- User Feedback Mechanism:

Description: Integrated tools allowing users to provide feedback on the platform's features, usability, and overall experience.

- Surveys and Polls:

Description: Periodic surveys and polls designed to gather user opinions on specific aspects of the platform, informing improvement strategies.

- Feature Release Notes:

Description: Transparent communication with users through release notes that detail updates, improvements, and new features.

- Agile Development Approach:

Description: Adopts an agile development approach that allows for quick iterations and adaptations based on emerging trends and user needs.

Tools:

- Feedback Collection System:

Description: A system that collects and organizes user feedback, providing valuable insights for improvements.

- Survey and Polling Platform:

Description: Platform for creating and administering surveys and polls to gather user opinions.

- Communication Platform:

Description: Channels for transparent communication, including release notes detailing updates, improvements, and new features.

- Agile Project Management System:

Description: Utilizes an agile project management system that facilitates quick iterations and adaptations based on user feedback and emerging trends.

IBM Applied Data Science Project
Covid-19 Vaccine Analysis
PHASE 3

1.INTRODUCTION:

This phase aims to clean, transform, and engineer features in a way that maximizes the model's ability to capture patterns and make accurate predictions. Through careful data preparation and feature engineering, we enhance the quality of input fed into the models. Effective preprocessing lays the foundation for improved predictive models.**The given dataset has been pre-processed and the outputs are attached with snap shots.**

2. IMPORTING LIBRARIES AND LOADING DATA:

For Pre-Processing the given dataset, the pandas library is used. The given csv file is uploaded to pandas as follows:

```
>>> import pandas as pd
>>> df = pd.read_csv(r"C:\Users\user\Desktop\NanMudhalvan\vaccine.csv")
>>> print(df)
            location        date             vaccine  total_vaccinations
0          Argentina  2020-12-29             Moderna                   2
1          Argentina  2020-12-29  Oxford/AstraZeneca                   3
2          Argentina  2020-12-29   Sinopharm/Beijing                   1
3          Argentina  2020-12-29           Sputnik V               20481
4          Argentina  2020-12-30             Moderna                   2
...              ...         ...                 ...                 ...
35618  European Union  2022-03-29  Oxford/AstraZeneca            67403106
35619  European Union  2022-03-29     Pfizer/BioNTech           600519998
35620  European Union  2022-03-29   Sinopharm/Beijing             2301516
35621  European Union  2022-03-29             Sinovac                1809
35622  European Union  2022-03-29           Sputnik V             1845103

[35623 rows x 4 columns]
```

3. UNDERSTANDING THE DATASET:

df.head:

```
>>> print(df.head())
    location        date             vaccine  total_vaccinations
0  Argentina  2020-12-29             Moderna                   2
1  Argentina  2020-12-29  Oxford/AstraZeneca                   3
2  Argentina  2020-12-29   Sinopharm/Beijing                   1
3  Argentina  2020-12-29           Sputnik V               20481
4  Argentina  2020-12-30             Moderna                   2
```

df.describe:

```
>>> print(df.describe())
       total_vaccinations
count        3.562300e+04
mean         1.508357e+07
std          5.181768e+07
min          0.000000e+00
25%          9.777600e+04
50%          1.305506e+06
75%          7.932423e+06
max          6.005200e+08
```

Isnull:

This function is used to identify missing values in the dataset. Since there in no null value, there is no need for handling the missing data.

```
>>> print(df.isnull().sum())
location              0
date                  0
vaccine               0
total_vaccinations    0
dtype: int64
```

4. REMOVING DUPLICATES:

If any row is duplicated in the given dataset, the following code will identify it and remove it. The given dataset does not contain any duplicates and hence the dataset is the same as before.

```
>>> bf = df
>>> bf = df.drop_duplicates()
>>> print(df.describe())
       total_vaccinations
count        3.562300e+04
mean         1.508357e+07
std          5.181768e+07
min          0.000000e+00
25%          9.777600e+04
50%          1.305506e+06
75%          7.932423e+06
max          6.005200e+08
>>> print(bf.isnull().sum())
location              0
date                  0
vaccine               0
total_vaccinations    0
dtype: int64
```

5. DATA TRANSFROMATION:

Normalizing Data:

```
>>> import pandas as pd
>>> from sklearn.preprocessing import MinMaxScaler, LabelEncoder
>>> df = pd.read_csv(r"C:\Users\user\Desktop\NanMudhalvan\vaccine.csv")
>>> scaler = MinMaxScaler()
>>> df['Normalized_Open']=scaler.fit_transform(df[['total_vaccinations']])
>>> df['Encoded_Data']= label_encoder.fit_transform(df['total_vaccinations'])
>>> print(df)
            location        date             vaccine  total_vaccinations  Normalized_Open  Encoded_Data
0          Argentina  2020-12-29             Moderna                   2     3.330447e-09             2
1          Argentina  2020-12-29  Oxford/AstraZeneca                   3     4.995670e-09             3
2          Argentina  2020-12-29   Sinopharm/Beijing                   1     1.665223e-09             1
3          Argentina  2020-12-29           Sputnik V               20481     3.410544e-05          1543
4          Argentina  2020-12-30             Moderna                   2     3.330447e-09             2
...              ...         ...                 ...                 ...              ...           ...
35618  European Union  2022-03-29  Oxford/AstraZeneca            67403106     1.122412e-01         27361
35619  European Union  2022-03-29     Pfizer/BioNTech           600519998     1.000000e+00         29209
35620  European Union  2022-03-29   Sinopharm/Beijing             2301516     3.832538e-03         15010
35621  European Union  2022-03-29             Sinovac                1809     3.012389e-06           614
35622  European Union  2022-03-29           Sputnik V             1845103     3.072509e-03         13947

[35623 rows x 6 columns]
```

Z-Score Standardization (for column – high):

Z-score standardization, also known as "z-score normalization" or "z-score scaling," is a statistical method used to standardize or normalize features in a dataset. It's a process that transforms the features by scaling them to have a mean of 0 and a standard deviation of 1. This makes it easier to compare and analyze variables with different units or scales.

The formula to calculate the z-score for a given data point

X in a feature is: **$z=X-\mu/\sigma$**

where:

X is an individual data point.

μ is the mean of the feature.

σ is the standard deviation of the feature.

The z-score measures how many standard deviations a data point is from the mean. A positive z-score indicates that the data point is above the mean, while a negative z-score indicates it's below the mean.

```
>>> df['total_vaccinations']=(df['total_vaccinations']-df['total_vaccinations'].mean())/df['total_vaccinations'].std()
>>> print(df)
            location        date             vaccine  total_vaccinations  Normalized_Open  Encoded_Data
0          Argentina  2020-12-29             Moderna           -0.291089     3.330447e-09             2
1          Argentina  2020-12-29  Oxford/AstraZeneca           -0.291089     4.995670e-09             3
2          Argentina  2020-12-29   Sinopharm/Beijing           -0.291089     1.665223e-09             1
3          Argentina  2020-12-29           Sputnik V           -0.290694     3.410544e-05          1543
4          Argentina  2020-12-30             Moderna           -0.291089     3.330447e-09             2
...              ...         ...                 ...                 ...              ...           ...
35618  European Union  2022-03-29  Oxford/AstraZeneca            1.009685     1.122412e-01         27361
35619  European Union  2022-03-29     Pfizer/BioNTech           11.298005     1.000000e+00         29209
35620  European Union  2022-03-29   Sinopharm/Beijing           -0.246674     3.832538e-03         15010
35621  European Union  2022-03-29             Sinovac           -0.291054     3.012389e-06           614
35622  European Union  2022-03-29           Sputnik V           -0.255482     3.072509e-03         13947

[35623 rows x 6 columns]
```

6. HANDLING OUTLIERS:

Outliers are data points that significantly differ from other observations in a dataset, deviating markedly from the overall pattern or distribution. They can be unusually high or low values that don't conform to the typical behaviour of the dataset.

The threshold fixed are the end points or outliers, all the values above and below are range are excluded and this process is called handling outliers.

Date fixed as threshold:

```
>>> thresholds = {'date': ("2020-12-29", "2022-03-29")}
>>> for col, (lower, upper) in thresholds.items():
...     df = df[(df[col] >= lower) & (df[col] <= upper)]
...
>>> print(df)
            location        date             vaccine  total_vaccinations  Normalized_Open  Encoded_Data
0          Argentina  2020-12-29             Moderna           -0.291089     3.330447e-09             2
1          Argentina  2020-12-29  Oxford/AstraZeneca           -0.291089     4.995670e-09             3
2          Argentina  2020-12-29   Sinopharm/Beijing           -0.291089     1.665223e-09             1
3          Argentina  2020-12-29           Sputnik V           -0.290694     3.410544e-05          1543
4          Argentina  2020-12-30             Moderna           -0.291089     3.330447e-09             2
...              ...         ...                 ...                 ...              ...           ...
35618  European Union  2022-03-29  Oxford/AstraZeneca            1.009685     1.122412e-01         27361
35619  European Union  2022-03-29     Pfizer/BioNTech           11.298005     1.000000e+00         29209
35620  European Union  2022-03-29   Sinopharm/Beijing           -0.246674     3.832538e-03         15010
35621  European Union  2022-03-29             Sinovac           -0.291054     3.012389e-06           614
35622  European Union  2022-03-29           Sputnik V           -0.255482     3.072509e-03         13947

[35539 rows x 6 columns]
```

7. DATA SPLITTING:

Outliers are data points that significantly differ from other observations in a dataset, deviating markedly from the overall pattern or distribution. They can be unusually high or low values that don't conform to the typical behavior of the dataset.

```
>>> import pandas as pd
>>> from sklearn.model_selection import train_test_split
>>> X=df.drop('total_vaccinations',axis=1)
>>> y=df['total_vaccinations']
>>> X_train, X_temp, y_train, y_temp =train_test_split(X,y,test_size=0.3,random_state=42)
>>> X_val, X_test, y_val, y_test =train_test_split(X_temp,y_temp,test_size=0.5,random_state=42)
>>>
```

TRAINING SET:

Purpose: Used to train the model, allowing it to learn patterns and relationships in the data.
Size: Largest portion of the dataset (e.g., 70-80%).
Importance: Fundamental for model training, ensuring the model learns from a variety of examples

```
>>> print("Training set:")
Training set:
>>> print(X_train)
        location        date             vaccine  Normalized_Open  Encoded_Data
5398       Cyprus  2021-11-05     Pfizer/BioNTech     1.359195e-03          9922
2436    Argentina  2022-02-25           Sputnik V     3.380232e-02         24890
30889     Uruguay  2021-09-07  Oxford/AstraZeneca     1.506661e-04          3488
18845  Luxembourg  2021-08-27  Oxford/AstraZeneca     1.749850e-04          3766
27990      Ukraine  2021-07-11     Johnson&Johnson     6.494372e-07           286
...            ...         ...                 ...              ...           ...
16891       Latvia  2021-03-19             Moderna     1.745654e-05          1131
6272       Czechia  2021-08-13     Pfizer/BioNTech     1.508933e-02         20847
11305      Germany  2021-04-03             Moderna     1.334045e-03          9859
860      Argentina  2021-06-08             Moderna     9.991341e-09             6
15820        Italy  2021-10-07     Johnson&Johnson     2.478437e-03         12813

[24877 rows x 5 columns]
```

VALIDATION SET:

Purpose: Used to fine-tune the model's hyperparameters, aiding in model selection and preventing overfitting.
Size: Smaller portion of the dataset (e.g., 10-15%).
Importance: Helps optimize the model's performance and generalization.

```
>>> print("Validation set:")
Validation set:
>>> print(X_val)
            location        date             vaccine  Normalized_Open  Encoded_Data
34871  European Union  2022-01-05  Oxford/AstraZeneca     1.122042e-01         27279
22310         Romania  2021-07-11     Johnson&Johnson     5.820839e-04          6661
23660        Slovenia  2021-12-31     Pfizer/BioNTech     3.412133e-03         14474
10656          France  2022-02-16             Moderna     3.880366e-02         25347
26496     Switzerland  2021-03-16     Johnson&Johnson     3.330447e-09             2
...               ...         ...                 ...              ...           ...
20038            Peru  2021-04-05     Pfizer/BioNTech     5.102927e-04          6363
30621         Uruguay  2021-06-09             Sinovac     3.965217e-03         15119
14132         Hungary  2021-11-12     Pfizer/BioNTech     1.140208e-02         19349
19949          Norway  2022-03-18     Pfizer/BioNTech     1.472087e-02         20763
31218         Uruguay  2021-12-25             Sinovac     5.408341e-03         16397
```

TESTING SET:

Purpose: Used to evaluate the model's performance on unseen data after training and validation.
Size: Smaller portion of the dataset (e.g., 10-15%).
Importance: Provides an unbiased evaluation of the model's performance and generalization to new data.

```
>>> print("Testing set:")
Testing set:
>>> print(X_test)
            location        date             vaccine  Normalized_Open  Encoded_Data
2538       Argentina  2022-03-14          Sputnik V     3.391103e-02         25036
33667  European Union  2021-08-24  Sinopharm/Beijing     3.474729e-03         14565
1396       Argentina  2021-09-05     Pfizer/BioNTech     4.163059e-08            25
19166          Malta  2022-01-07     Johnson&Johnson     5.316559e-05          1959
13483      Hong Kong  2021-09-04     Pfizer/BioNTech     8.203973e-03         18097
...              ...         ...                 ...              ...           ...
8574         Ecuador  2022-01-05             CanSino     8.091721e-04          7714
9802          France  2021-07-17     Pfizer/BioNTech     8.489335e-02         26689
3128         Belgium  2021-09-10     Pfizer/BioNTech     1.988415e-02         22162
8022         Ecuador  2021-08-20             CanSino     4.477453e-05          1790
8740         Estonia  2021-06-11             Moderna     1.307051e-04          3269

[5331 rows x 5 columns]
```

8. SAVING:

```
>>> df.to_csv('preprocessed_vaccine.csv',index=False)
```

9. CONCLUSION:

In the third phase, the dataset has been preprocessed, which is fundamental to building accurate and reliable predictive models. This involved handling missing values, scaling, encoding categorical features, and possibly applying other transformations like feature engineering or selection. The preprocessed dataset is now ready for the subsequent phases, where it will be utilized to train and validate models

IBM Applied Data Science Project
Covid-19 Vaccine Analysis
PHASE 1

INTRODUCTION:-

In this phase of our Covid-19 vaccines analysis project, we advance by focusing on feature engineering, model training, and evaluation. Feature engineering involves crafting and selecting meaningful attributes from our data to enhance predictive accuracy. Subsequently, we'll train our model, employing suitable machine learning algorithms and parameter optimization. Evaluation then allows us to gauge our model's performance, refining it for improved vaccines forecasts. This progression is essential to empower us with informed investment decisions.

INITIALIZING THE OBJECT OR CALLING THE LIBRARIES:-

1. "import pandas as pd" allows you to use Pandas for data manipulation, providing data structures like DataFrames and Series to work with tabular data efficiently.

2. "import matplotlib.pyplot as plt" imports Matplotlib, a powerful library for creating various types of data visualizations, with the alias "plt" for convenience.

3. "import seaborn as sns" imports Seaborn, which is built on top of Matplotlib and provides a high-level interface for creating aesthetically pleasing statistical graphics.

4. Together, these libraries are commonly used for data analysis and visualization in Python, enabling you to load, clean, and analyze data and then create informative plots and charts.

5. These libraries are often used in conjunction to explore, analyze, and present data effectively in data science and data visualization.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

PRINTING THE DATAFRME:-

1. Data Retrieval :

The code first specifies the URL of a COVID-19 data file hosted on GitHub. It uses this URL to retrieve the data, which is in CSV format. This data contains information about COVID-19 cases, vaccinations, and related statistics.

2. Data Loading:

The Pandas library is used to load the data from the specified URL into a DataFrame, which is stored in the variable `df_w`. Pandas is a powerful library for data manipulation and analysis and is commonly used for working with tabular data.

3.Data Filtering:

After loading the data, the code filters it to include only the records where the 'date' column is less than or equal to '2021-12-08'. This filtering operation restricts the data to a specific time range, in this case, up to and including December 8, 2021. This can be helpful when you're interested in analyzing a particular period of time.

4.Information Display:

The code concludes by printing out some basic information about the filtered DataFrame. It displays the number of rows and columns in the DataFrame. This information is useful for understanding the size and structure of the data you're working with.

```
[4] PATH = "https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv"
    df_w=pd.read_csv(PATH)
    df_w=df_w[df_w.date<='2021-12-08']
```

```
print("The DataFrame has:",
      "\nNumber of rows:",df_w.shape[0],
      "\nand",df_w.shape[1], "number of columns"
      )
```

```
The DataFrame has:
Number of rows: 178211
and 67 number of columns
```

SUMMARY OF FIRST FIVE DATA:-

Providing a summary of the first five rows of data (often called a "head" of the data) is a common practice in data analysis and data science projects.

1. Data Inspection:

Examining the first few rows of the dataset allows you to quickly inspect the data and get a sense of its structure. We can see the names of columns, the data types, and some example values. This initial inspection helps you understand what kind of information is available in the dataset.

2. Data Quality:

It helps you assess the quality of the data. If there are any missing values, anomalies, or unexpected values in the first few rows, they can be identified early in the analysis. This is important for data cleaning and preprocessing.

3. Column Names:

we can verify that the column names and labels are correct. This is crucial for ensuring that you're working with the right variables.

4. Quick Overview:

A summary of the first few rows provides a quick overview of the dataset's contents. We can see the range of values and get an initial impression of the data's distribution.

5. Debugging:

If there are issues with data loading or preprocessing, they may become evident when inspecting the first few rows. This can help in debugging the data retrieval and transformation process.

```
df_w.head()
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	male_smokers	handwashing_facilities	hospital_
0	AFG	Asia	Afghanistan	2020-01-03	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	
1	AFG	Asia	Afghanistan	2020-01-04	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	
2	AFG	Asia	Afghanistan	2020-01-05	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	
3	AFG	Asia	Afghanistan	2020-01-06	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	
4	AFG	Asia	Afghanistan	2020-01-07	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	

5 rows × 67 columns

SUMMARY OF LAST FIVE DATA:-

Examining the last few rows of the data (often referred to as the "tail" of the data) in a COVID-19 vaccine can be useful for

1. Data Completeness:

It helps ensure that the data is complete and up to date. By looking at the last few rows, you can verify that there are no missing data or incomplete records. This is crucial for data integrity.

2. Temporal Trends:

If the data is time-series data (as in many COVID-19 datasets), examining the last few rows provides insights into the most recent trends and developments. It allows you to see how the data has evolved over time, which can be important for tracking the progression of the pandemic or vaccination efforts.

3. Quality Control:

Checking the last rows can reveal any potential issues or anomalies that might have occurred near the end of the data collection period. This is important for data quality control and anomaly detection

4. Data Transformation:

In some cases, data transformations or calculations might have been performed, and these changes may be most evident in the last few rows. By examining the tail of the data, you can check if any post-processing has been applied.

5. Decision-Making:

If we are making real-time decisions or updating models based on the most recent data, the last few rows provide the latest information that can inform your decisions.

In the context of a COVID-19 vaccine project, looking at the last few rows helps ensure that your analysis or modeling takes into account the most recent data and events related to the pandemic and vaccination efforts. It's especially important when working with dynamic, evolving datasets that are updated over time.

```
df_w.tail()
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	male_smokers	handwashing_facilities	hospi
351151	ZWE	Africa	Zimbabwe	2021-12-04	137441.0	1062.0	515.000	4708.0	1.0	0.571	...	30.7	36.791	
351152	ZWE	Africa	Zimbabwe	2021-12-05	138523.0	1082.0	653.143	4709.0	1.0	0.571	...	30.7	36.791	
351153	ZWE	Africa	Zimbabwe	2021-12-06	139046.0	523.0	722.143	4710.0	1.0	0.714	...	30.7	36.791	
351154	ZWE	Africa	Zimbabwe	2021-12-07	141601.0	2555.0	1053.571	4713.0	3.0	1.000	...	30.7	36.791	
351155	ZWE	Africa	Zimbabwe	2021-12-08	145632.0	4031.0	1568.571	4718.0	5.0	1.571	...	30.7	36.791	

5 rows × 67 columns

EXTRACTION OF VARIABLES:-

In the context of the COVID-19 vaccine extracting specific variables or columns from the DataFrame (`df_w`) can be beneficial for

1. Focus on Relevant Data:

By extracting a subset of columns, we can focus on the specific variables that are most relevant to your analysis or research question. This can help we reduce clutter and simplify your analysis.

2. Computational Efficiency:

Working with a smaller set of columns can improve computational efficiency. Unnecessary columns can consume memory and processing power, so extracting only the variables of interest can speed up operations.

3. Simplifying Analysis:

Simplifying the dataset by selecting specific variables makes it easier to perform statistical analysis, data visualization, and modeling. We can more easily see patterns and relationships in the data.

4. Data Privacy and Security:

If the dataset contains sensitive information, extracting only the necessary variables can help protect privacy and security. Unneeded data can be a risk if not properly handled.

5. Clarity and Readability:

Extracting variables helps improve the clarity and readability of your code. It's easier for others to understand and work with a more focused subset of data.

6. Custom Data Processing:

We can apply custom data processing and transformations to the selected variables without affecting the entire dataset. This is particularly useful if you need to perform specialized calculations or modifications.

7. Modeling and Prediction:

If we are building models or making predictions, it's important to choose the right set of input features. Selecting relevant variables is crucial for model accuracy and interpretability.

In specific case, it appears that you've listed a set of COVID-19-related variables that we want to extract from the Data Frame. These variables might be of particular interest for your analysis or research related to the pandemic, vaccination efforts, or other specific aspects of the data. Extracting them will allow us to work with this subset of data more effectively.

```
df_w.columns.to_list()
```

```
 'new_deaths_smoothed',
 'total_cases_per_million',
 'new_cases_per_million',
 'new_cases_smoothed_per_million',
 'total_deaths_per_million',
 'new_deaths_per_million',
 'new_deaths_smoothed_per_million',
 'reproduction_rate',
 'icu_patients',
 'icu_patients_per_million',
 'hosp_patients',
 'hosp_patients_per_million',
 'weekly_icu_admissions',
 'weekly_icu_admissions_per_million',
 'weekly_hosp_admissions',
 'weekly_hosp_admissions_per_million',
 'total_tests',
 'new_tests',
 'total_tests_per_thousand',
 'new_tests_per_thousand',
 'new_tests_smoothed',
 'new_tests_smoothed_per_thousand',
 'positive_rate',
 'tests_per_case',
 'tests_units',
 'total_vaccinations',
 'people_vaccinated',
 'people_fully_vaccinated',
 'total_boosters',
```

ANALYSIS:-

Analyzing a COVID-19 vaccine project using graphs and models is important for

1. Visualizing Trends:

Graphs and charts can visually represent data trends, making it easier to spot patterns and changes over time. Visualizations can reveal vaccination rates, infection rates, and other important trends related to the pandemic.

2. Identifying Correlations:

Statistical models and correlation analyses can help identify relationships between variables. For instance, you can assess whether vaccination rates have an impact on the reduction of infection rates.

3. Predictive Modeling:

Statistical and machine learning models can be used to make predictions and forecasts. These models can help estimate future COVID-19 trends, vaccination needs, or the impact of policy decisions.

4. Public Health Decision-Making:

Data analysis and models can inform public health decision-making. Government agencies, healthcare organizations, and policymakers can use the insights to plan vaccination campaigns, allocate resources, and make informed decisions.

5. Monitoring and Surveillance:

Real-time data analysis and models can aid in monitoring the spread of the virus, detecting outbreaks, and identifying areas with vaccination coverage gaps. This information is crucial for targeted interventions.

6. Resource Allocation:

Models can assist in allocating healthcare resources efficiently, such as hospital beds, ventilators, and medical personnel. This helps ensure that healthcare systems are prepared for surges in COVID-19 cases.

7. Vaccination Strategy Optimization:

Analysis and models can help optimize vaccination strategies, including prioritizing vaccination for high-risk populations, determining the timing of booster shots, and managing vaccine supply chains.

8. Communication:

Visualizations and models provide a means to communicate complex data and insights to the public, healthcare workers, and policymakers. This aids in raising awareness and promoting responsible behavior.

9. Research and Development:

Data analysis and modeling can support ongoing research into COVID-19 vaccines, treatments, and the virus itself. It can guide researchers in refining vaccines and treatments.

10. Data-Driven Policy:

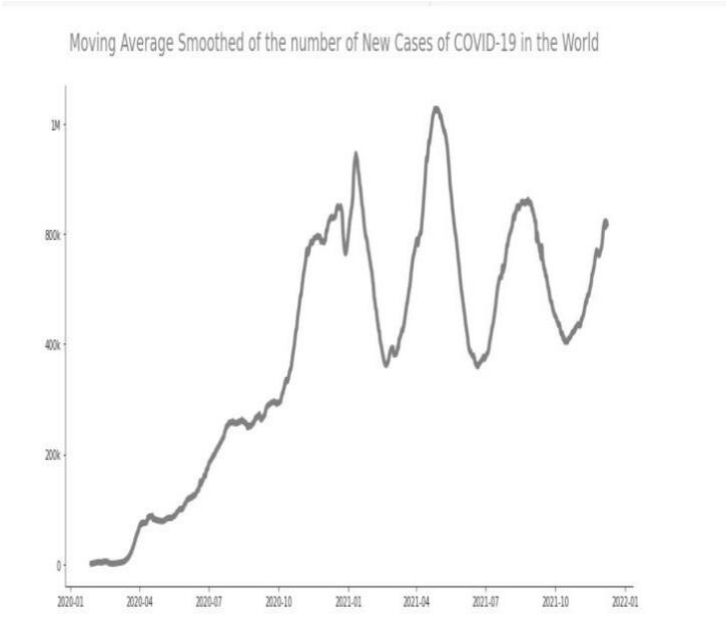
Governments and health organizations can use data-driven analysis to develop policies and guidelines. This is especially important in managing the pandemic effectively.

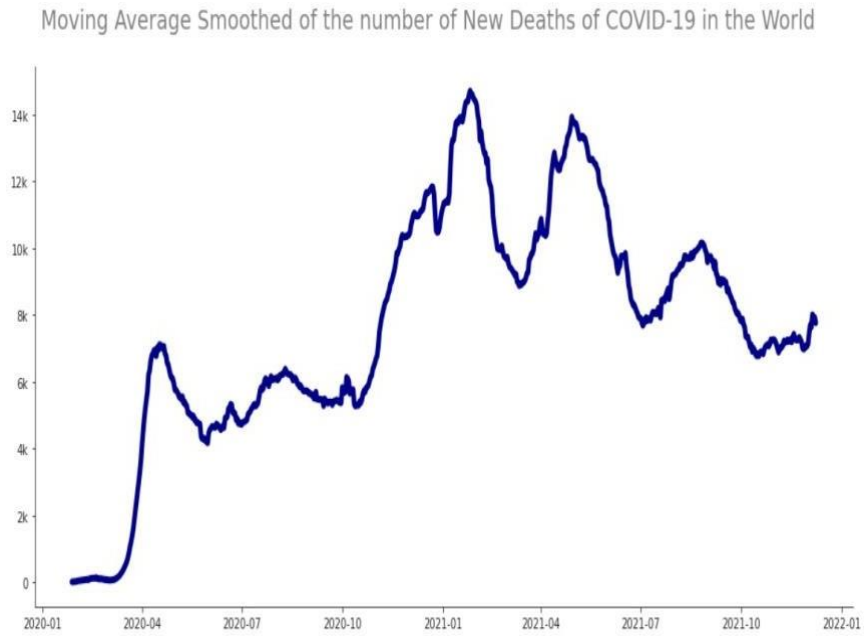
Data analysis, graphs, and models play a critical role in understanding, managing, and responding to the COVID-19 pandemic, particularly in the context of vaccine distribution and the broader public health response. They provide evidence-based insights and support informed decision-making.

```
[ ] # Calling the libraries
    import matplotlib.pyplot as plt
    from pywaffle import Waffle
    import numpy as np
```

```
fig = plt.figure(
    FigureClass=Waffle,
    rows=5,
    figsize = (16,4),
    values={'Fully vaccinated': 44, 'Only partly vaccinated': 11, 'Did not take any dose': 45},
        title={
        'label': 'Number of people vaccinated against COVID-19\n',
        'loc': 'left',
        'fontdict': {
            'fontsize': 20
        }
    },
    colors=("OliveDrab", "DarkBlue","DarkRed"),
    icons = ['syringe','syringe','skull'],
    legend={'loc': 'upper left',
            'bbox_to_anchor': (1, 1),
            'fontsize':20
            },
    icon_size=24,
    icon_legend=True
)

fig.set_tight_layout(False)
plt.show();
```





CONCLUSION:

We notice the how fast is the way the virus Covid-19 can spread into the world. The world didn't know how to deal with the problem at the start because of the number of cases and deaths. The countries with more income were able to get resources and track the virus more efficient than others. The vaccines are the way to combat the virus. Only 19% of the population of the World are fully vaccinated Vaccination is the best way to deal with the virus.