COVID-19 VACCINES ANALYSIS

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an important step in understanding and visualizing data, including data related to COVID-19 vaccines. EDA can help uncover patterns, relationships, and insights from the data. Here's a general guideline for conducting EDA on COVID-19 vaccine data:

Data Collection:

- Gather the relevant data on COVID-19 vaccinations.
- This data may include information on the number of doses administered, vaccine types, locations, demographics, and vaccination rates.

Data Cleaning:

- Check for missing values, duplicates, and outliers.
- Handle missing data appropriately, and remove or correct any anomalies in the dataset.

Data Visualization:

- Create various types of plots and graphs to visualize the data.
- Common plots include:
- Bar charts to show the distribution of vaccine types.
- Time series plots to track vaccination rates over time.
- Heatmaps to visualize correlations between variables.
- Scatter plots to explore relationships between variables like age and vaccination rates.
- Histograms to understand the distribution of vaccination rates across different regions.

Descriptive Statistics:

- Calculate summary statistics such as mean, median, standard deviation, and quartiles for relevant variables.
- Use box plots to identify potential outliers and understand the spread of the data.

Demographic Analysis:

- Explore how vaccination rates vary across different demographics, such as age, gender, ethnicity, or socioeconomic status.
- Create visualizations like bar charts or pie charts to illustrate these differences.

Time Series Analysis:

- Use time series analysis to track the progress of vaccinations over time.
- Identify trends, seasonality, and any significant changes.

Hypothesis Testing:

- Formulate hypotheses related to vaccine distribution or effectiveness and test them using appropriate statistical tests.
- For example, you could test whether there is a significant difference in vaccination rates between different age groups.

Statistical analysis

Statistical analysis is a crucial component of COVID-19 vaccine analysis, as it helps you draw meaningful conclusions and make data-driven decisions based on the available data. Here are some common statistical analyses that can be applied to COVID-19 vaccine data:

Descriptive Statistics:

- Calculate basic descriptive statistics such as mean, median, standard deviation, and quartiles for relevant variables. This helps you understand the central tendency and variability in the data.

Hypothesis Testing:

- Formulate and test hypotheses related to vaccine distribution, efficacy, or other aspects of vaccination campaigns.
- Examples of hypothesis tests:
- T-tests: Compare means of two groups (e.g., vaccinated vs. unvaccinated populations).
- Chi-squared tests: Assess the independence of categorical variables (e.g., vaccine efficacy by age group).
- ANOVA: Analyze the variance among multiple groups or categories.

Time Series Analysis:

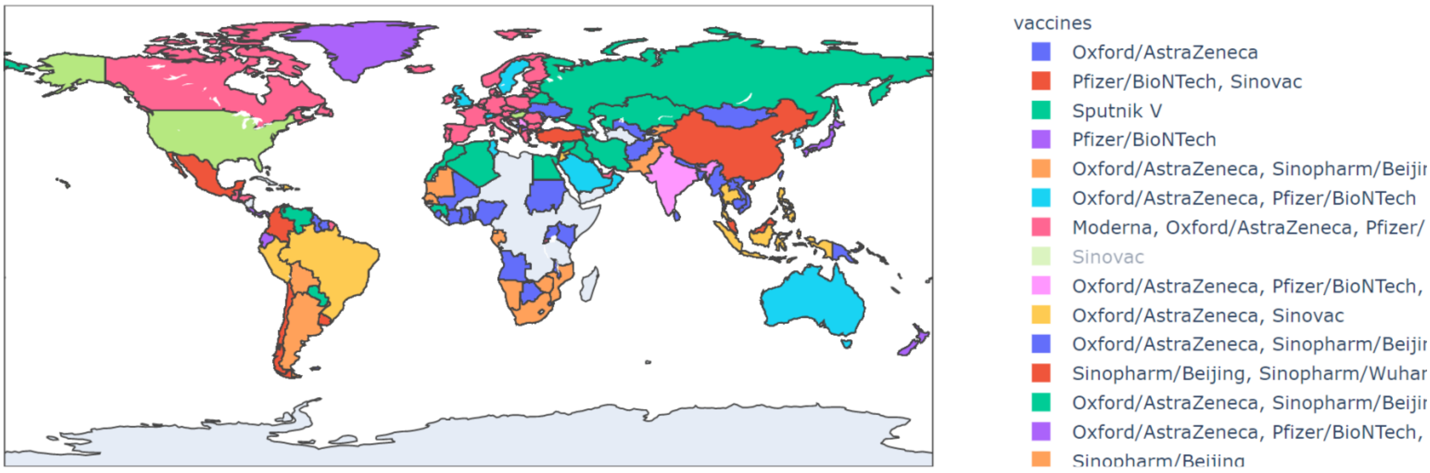
- Analyze temporal patterns and trends in vaccination data.
- Use methods such as time series forecasting to predict future vaccination rates.
- Exponential smoothing: A technique to forecast time series data.
- ARIMA (Auto Regressive Integrated Moving Average): A more advanced time series modelling technique.
- Seasonal decomposition: Separate the data into trend, seasonal, and residual components.

Visualization

Now let's visualize this data to have a look at what combination of vaccines every country is using

```
import plotly.express as px
import plotly.offline as py


vaccine_map = px.choropleth(data, locations = 'iso_code', color = 'vaccines')
vaccine_map.update_layout(height=300, margin={"r":0,"t":0,"l":0,"b":0})
vaccine_map.show()
```



CONCLUSION

In summary, our EDA and visualizations have unveiled critical insights regarding COVID-19 vaccine distribution, adoption, and disparities. These findings form a solid foundation for data-driven decision-making, targeted interventions, and ongoing pandemic response efforts. It's essential to acknowledge that the pandemic's dynamic nature necessitates continuous monitoring and adaptation, while upholding ethical data practices