

COVID-19 VACCINE ANALAYSIS PROJECT

ABSTRACT :

- *This dataset provides a comprehensive record of COVID-19 vaccine administrations in from December 2020 to early January 2021. It includes information on the location, date, vaccine type, and the total number of vaccinations administered. Notable observations include a consistent increase in the total number of vaccinations over time, with "Sputnik V" being the most administered vaccine. The dataset offers valuable insights for monitoring and optimizing vaccination campaigns, data quality maintenance, and strategic planning.*

OBJECTIVES :

- Certainly, here's the meaning of each column in the provided data:
- Location: This column specifies the geographical location where the vaccine administrations took place, which is Argentina in this case.
- Date: This column indicates the date on which the vaccinations occurred, in the format YYYY-MM-DD.
- Vaccine: This column specifies the type of vaccine that was administered, such as Moderna, Oxford/AstraZeneca, Sinopharm/Beijing, or Sputnik V.
- Total Vaccinations: This column represents the total number of vaccinations administered for the specific vaccine type on the given date in Argentina.

DATA SOURCE :

Data set link:(<https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>)

	location	date	vaccine	total_vaccinations			
1	Argentina	29/12/20	Moderna	2			
2	Argentina	29/12/20	Oxford/As	3			
3	Argentina	29/12/20	Sinopharm	1			
4	Argentina	29/12/20	Sputnik V	20481			
5	Argentina	30/12/20	Moderna	2			
6	Argentina	30/12/20	Oxford/As	3			
7	Argentina	30/12/20	Sinopharm	1			
8	Argentina	30/12/20	Sputnik V	40583			
9	Argentina	31/12/20	Moderna	2			
10	Argentina	31/12/20	Oxford/As	3			
11	Argentina	31/12/20	Sinopharm	1			
12	Argentina	31/12/20	Sputnik V	43388			
13	Argentina	01/01/21	Moderna	2			
14	Argentina	01/01/21	Oxford/As	5			
15	Argentina	01/01/21	Sinopharm	1			
16	Argentina	01/01/21	Sputnik V	43513			
17	Argentina	02/01/21	Moderna	2			
18	Argentina	02/01/21	Oxford/As	6			
19	Argentina	02/01/21	Sinopharm	1			
20	Argentina	02/01/21	Sputnik V	46824			
21	Argentina	03/01/21	Moderna	2			
22	Argentina	03/01/21	Oxford/As	6			
23	Argentina	03/01/21	Sinopharm	1			
24	Argentina	03/01/21	Sputnik V	47266			
25	Argentina	04/01/21	Moderna	2			
26	Argentina	04/01/21	Oxford/As	6			
27	Argentina	04/01/21	Sinopharm	1			
28	Argentina	04/01/21	Sputnik V	57726			
29	Argentina	05/01/21	Moderna	2			
30	Argentina	05/01/21	Oxford/As	6			
31	Argentina	05/01/21	Sinopharm	5			
32	Argentina	05/01/21	Sputnik V	68445			
33	Argentina	06/01/21	Moderna	2			
34	Argentina	06/01/21	Oxford/As	6			
35	Argentina	06/01/21	Sinopharm	8			
36	Argentina	06/01/21	Sputnik V	78551			
37	Argentina	07/01/21	Moderna	2			

DESIGN, THINKING & INNOVATION :

- Data exploration & understanding
- Data reprocessing
- Exploratory data analysis(EDA)
- Statistical analysis
- Virtualization
- Insights and recommendation
- Conclusion

Data exploration and understanding:

- Exploratory Data Analysis (EDA) is an important step to understand your data better. While I can't create visualizations in this text-based format, I can suggest some steps and analyses you can perform for EDA on your dataset:
- Summary Statistics: Calculate basic statistics for numerical columns like "Total Vaccinations" to understand the distribution, mean, median, and range of vaccinations.
- Data Distribution: Plot histograms to visualize the distribution of "Total Vaccinations" for each vaccine type.
- Time Series Analysis: Since you have a date column, you can plot the trends over time to see how the vaccination numbers for different vaccines have evolved.
- Compare Vaccine Types: Create bar plots or pie charts to compare the usage of different vaccine types over time or across different locations.
- Correlation Analysis: Check if there is any correlation between variables, for example, between the number of Moderna vaccinations and the number of Oxford/AstraZeneca vaccinations.

Data processing :

- Remove any duplicate entries: Check for and remove any duplicate rows in the dataset to ensure each record is unique.
- Handle missing values: Check for any missing values in the dataset and decide how to handle them, whether by filling them in or removing rows with missing data.
- Convert data types: Ensure that the "Date" column is in a proper date format and not a string.
- Aggregating data: If needed, you can aggregate the data based on your analysis goals. For example, you can calculate daily or cumulative vaccination numbers for each vaccine type.
- Check for outliers: Examine the data for outliers or erroneous values and decide how to handle them.

Exploratory Data Analysis :

- *Summary Statistics:* Calculate basic statistics for numerical columns like "Total Vaccinations" to understand the distribution, mean, median, and range of vaccinations.
- *Data Distribution:* Plot histograms to visualize the distribution of "Total Vaccinations" for each vaccine type.
- *Time Series Analysis:* Since you have a date column, you can plot the trends over time to see how the vaccination numbers for different vaccines have evolved.
- *Compare Vaccine Types:* Create bar plots or pie charts to compare the usage of different vaccine types over time or across different locations.
- *Correlation Analysis:* Check if there is any correlation between variables, for example, between the number of Moderna vaccinations and the number of Oxford/AstraZeneca vaccinations.
- *Outlier Detection:* Identify and analyze any outliers in the data that might need special attention.
- *Data by Location:* You can analyze how vaccination numbers vary by location within Argentina.

Statistical analysis :

- *Descriptive Statistics:* Calculate basic statistics like mean, median, standard deviation, minimum, and maximum for the "Total Vaccinations" column. This will provide an overview of the data's central tendency and variability.
- *Hypothesis Testing:* You can use hypothesis tests to determine if there are significant differences between vaccine types or between different dates. For example, you could perform a t-test or ANOVA to compare the means of "Total Vaccinations" for different vaccine types.
- *Correlation Analysis:* Calculate correlations between variables to understand relationships. For instance, you can determine if there is a correlation between the number of vaccinations of different vaccine types.
- *Time Series Analysis:* You can apply time series analysis techniques to understand how vaccination numbers have changed over time. This may involve decomposing the time series into trend, seasonality, and residuals.
- *Regression Analysis:* If you have additional data, such as demographic or geographic information, you can perform regression analysis to model and predict the number of vaccinations based on these factors.
- *Chi-squared Test:* If you want to test the independence of two categorical variables (e.g., vaccine type and location), you can use a chi-squared test.

Visualization :

- *To perform visualizations for your dataset, you can use various data visualization libraries in programming languages like Python. Here, I'll guide you on how to create some common types of visualizations using Python and the popular libraries Matplotlib and Seaborn. If you don't already have these libraries, you can install them using pip:*
 1. *Histogram*
 2. *Time series plot*
 3. *Bar plot for vaccine type*
 4. *Corelation heatmap*
 5. *Boxplot*

Insights & Recommendations :

- *Insights:*
- *Vaccination Trends: The dataset shows a time series of vaccine administrations in Argentina, indicating that vaccination efforts began in December 2020.*
- *Most Administered Vaccine: "Sputnik V" appears to be the most administered vaccine in the dataset.*
- *Recommendations:*
- *Monitor Vaccine Distribution: Continue to monitor the distribution and administration of different vaccines to ensure a balanced and effective vaccination campaign.*
- *Data Quality: Ensure data quality by regularly checking for and addressing duplicates, missing values, and anomalies in the dataset.*

Conclusion :

- *In conclusion, the dataset you provided offers valuable information about vaccine administrations in Argentina, particularly during the initial phases of the vaccination campaign. Here are the key takeaways:*
- *Vaccine Trends: The data shows that various vaccine types, including Moderna, Oxford/AstraZeneca, Sinopharm/Beijing, and Sputnik V, were administered in Argentina starting in December 2020. "Sputnik V" was the most administered vaccine type.*
- *Positive Trend: Over time, there is a noticeable increase in the total number of vaccinations, suggesting a positive trend in vaccination efforts.*
- *Daily Fluctuations: There are daily fluctuations in the number of vaccinations, which might be due to various operational and external factors.*

THANK YOU