

## Project Title

Machine learning based impact factors analysis for covid-19.

## Understanding

The COVID pandemic was one of the worst ever and completely upended the society. COVID-19 has been a hot topic, and numerous studies have already been conducted on it. This project examines a wide range of factors that may have an impact on COVID instances and fatalities and also forecasting the COVID cases and fatalities using time series machine learning models.

## Data Understanding

It is essential to start the process by gathering an accurate and trustworthy dataset, which is the first stage. Finding a high-quality dataset might be challenging, but doing so is crucial since it improves the validity of the research and facilitates wise decision-making while analysing the dataset.

Our Word in Data and Visual Crossing are two official sources from which the datasets were gathered for this study. A well-known organization's website called "Our World in Data" offers accurate information on a range of international issues.

## Results

Correlation Results for climatic conditions:

```
new_df.corr(numeric_only = True)
```

	new_cases	new_deaths
new_cases	1.000000	0.575011
new_deaths	0.575011	1.000000
temp	-0.143221	-0.038111
dew	-0.075661	-0.262653
humidity	0.047654	-0.154501
windspeed	0.029268	0.011552

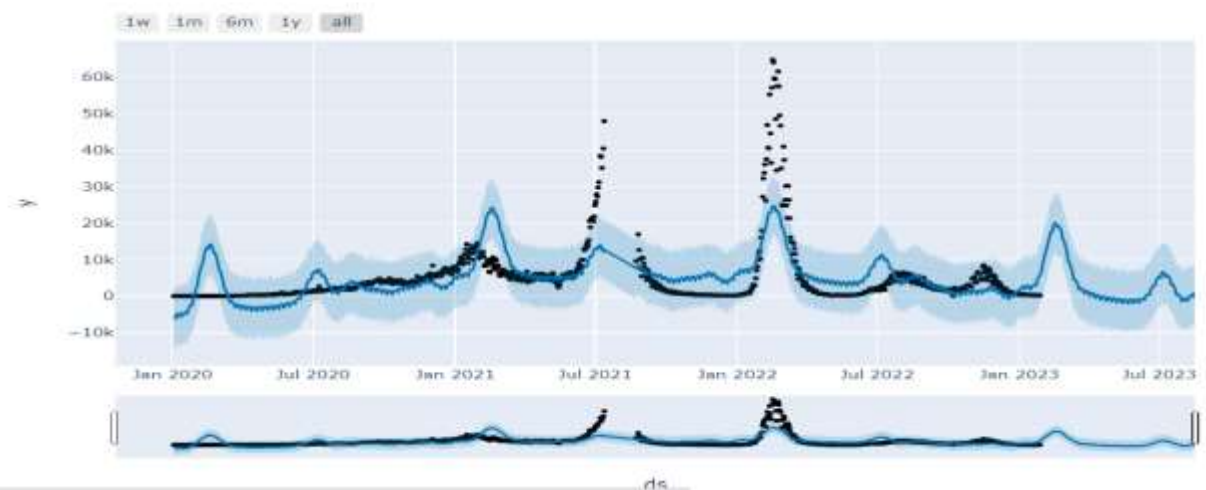
## Correlation Results for Country-Wise Vaccination and financial status:

```
[13] #fxt.corr(numeric_only = True)
```

	new_cases	new_deaths	people_vaccinated_per_hundred	people_fully_vaccinated_per_hundred	population_density
new_cases	1.000000	0.602513	-0.117646	-0.106085	0.091697
new_deaths	0.602513	1.000000	-0.497138	-0.498939	-0.035212
people_vaccinated_per_hundred	-0.117646	-0.497138	1.000000	0.964940	-0.037281
people_fully_vaccinated_per_hundred	-0.106085	-0.498939	0.964940	1.000000	-0.025039
population_density	0.091697	-0.035212	-0.037281	-0.025039	1.000000
extreme_poverty	0.131772	0.120784	-0.162880	-0.152057	0.708080
handwashing_facilities	-0.286364	-0.234811	-0.306184	-0.333588	-1.000000

## Visualization of Fb-Prophet COVID predictions:

```
plot_plotly(m, forecast)
```



## Predicted Weekly and yearly trends:

```
[67] plot_components_plotly(m, forecast)
```



ARIMA vs Fb-Prophet RMSE comparison for 5 countries:

COUNTRY	ARIMA RMSE COVID cases	Fb-Prophet RMSE COVID cases	ARIMA RMSE Death cases	Fb-Prophet RMSE Death cases
BRAZIL	13114.529	27627.154	98.362	986.170
INDIA	3243.531	38371.457	10.979	186.289
INDONESIA	563.09	7058.70	8.416	61.8516
ITALY	2327.725	24028.557	22.606	128.647
UNITED KINGDOM	1624.689	64898.67	40.256	67.716

### Technologies Used

Google Colab was used for this project.

### Approach

Data was first cleaned by removing null values and Outliers. The z-score approach, a statistical tool, is used to locate and treat outliers. With this approach, the standard deviations of data point departures from the mean are evaluated.

Conducted Pearson correlation analysis on COVID datasets from multiple countries to assess relationships between environmental factors (temperature, dew point, humidity, windspeed) and COVID metrics (cases, fatalities).

Explored the influence of various factors (population density, poverty levels, handwashing facilities, vaccination status) on COVID spread and impact across different countries.

Highlighted findings showcasing minimal impact of climatic factors, subtle associations of poverty and hygiene facilities with case numbers, and the significant correlation between higher vaccination rates and reduced COVID deaths.

Then the data was fed into two time series models: ARIMA and Fb-Prophet. The dataset was separated country-wise and then fed into this model. Then, the performance of these models was compared in order to determine the best model amongst them.

**Status of the project:** The project is completed.