# Challenges Faced & Techniques Used

## 1. Introduction

Analysing COVID-19 data posed several challenges due to its irregularity, missing data, and fluctuating trends. In this section, we discuss the challenges encountered during the analysis and the techniques used to overcome them.

## 2. Challenges

### 2.1 Missing Data:

COVID-19 datasets often had gaps in reporting, leading to incomplete time series.

### Solution:

We used forward filling to impute missing values, ensuring continuous time series data for forecasting.

### 2.2 Outliers:

The data showed sharp fluctuations, especially during major events like lockdowns or new variant outbreaks.

### Solution:

We applied smoothing techniques like moving averages to reduce the impact of outliers and prevent them from skewing the forecasts.

**2.3 Seasonality and Trend Shifts:**

The pandemic exhibited strong seasonal fluctuations and unexpected shifts, complicating the modelling process.

**Solution:**

We utilized seasonal decomposition and SARIMA models to account for seasonal effects and trend shifts.

**2.4 Stationarity:**

The data was non-stationary, which meant we had to transform it to apply time series models effectively.

**Solution:**

We performed differencing and log transformations to make the data stationary.

# 3. Techniques Used

### 3.1 Differencing:

Applied to make the data stationary and suitable for ARIMA and SARIMA models.

### 3.2 Smoothing:

Used to handle volatility and reduce the effect of outliers.

### 3.3 Cross-validation:

Ensured model robustness by validating its generalization to unseen data.

## 4. Conclusion

By applying techniques like differencing, seasonal decomposition, and smoothing, we were able to tackle challenges such as missing data, seasonality, and non-stationarity. These techniques enabled the models to perform more effectively.