# Comprehensive Analysis and Forecasting of Monkeypox (MPOX) Cases Using Time Series Models

## Table of Contents:

# Abstract

In the context of escalating global health threats and the emergence of infectious diseases, effective monitoring and forecasting of outbreaks are crucial for public health response. Traditional epidemiological approaches often face challenges, including delays in reporting and data accessibility.

This study explores the utility of time series forecasting models, specifically ARIMA and SARIMA, in analysing and predicting Monkeypox (MPOX) cases during the 2022-2024 outbreak. Utilising comprehensive datasets obtained from global health databases, we employed a systematic approach to examine historical case data and identify trends and patterns in transmission dynamics.

The methodology involved time series decomposition, rigorous data preprocessing, and the application of ARIMA and SARIMA models to forecast future case incidence. Key analyses included lagged correlations between new cases and total cases, as well as seasonal variations in case trends.

The findings indicate that while the SARIMA model demonstrated superior performance in capturing seasonal fluctuations, both models provided valuable insights into the trajectory of the outbreak. Limitations in forecasting accuracy were observed, necessitating further refinement of the models to enhance predictive reliability.

This study underscores the importance of continuous monitoring and the integration of advanced forecasting techniques in the strategic planning of public health interventions to mitigate the impact of future outbreaks.
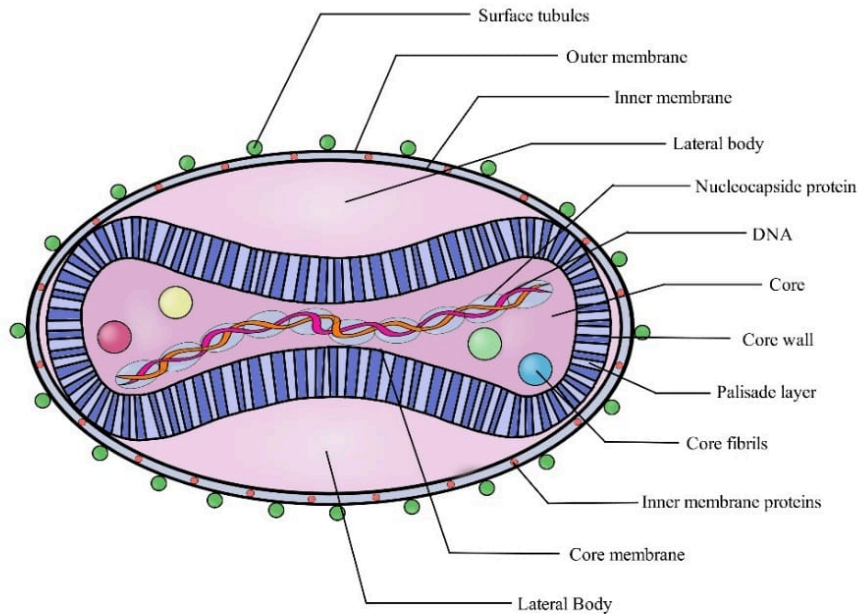
# 1. Executive Summary

- This report presents an in-depth analysis of global Monkeypox (MPOX) cases and uses advanced time series forecasting techniques to predict future trends. By leveraging historical data, the report aims to offer actionable insights for health authorities to better manage the outbreak.

- Using models like <u>ARIMA</u> and <u>SARIMA</u>, the analysis reveals key trends: cases surged dramatically in 2022 but have shown signs of stabilisation due to effective public health interventions. Forecasts indicate a potential decline in new cases over the next year, though vigilance is still required in hotspot regions.

- The report also highlights the regions most impacted and identifies factors contributing to varying case fatality rates. These findings serve as a critical resource for policy makers, public health experts, and researchers focusing on epidemic management and containment.

# 2. Introduction

## 2.1 Background

Monkeypox (MPOX), an emerging <u>zoonotic virus</u>, has transitioned from its endemic status in certain African regions to a global public health concern. With multiple outbreaks reported worldwide, the virus has stressed the importance of timely surveillance, effective response measures, and predictive modelling.

Surface tubules
Outer membrane
Inner membrane
Lateral body
Nucleocapside protein
DNA
Core
Core wall
Palisade layer
Core fibrils
Inner membrane proteins
Core membrane
Lateral Body

## 2.2 Objectives

The primary objective is to use historical data and time series models to analyse and forecast the global trends of Monkeypox, assisting in resource allocation and the management of outbreaks.

## 2.3 Importance of the Study

Time series forecasting is a crucial tool in epidemiology, enabling authorities to anticipate future case numbers, resource needs, and the possible trajectory of outbreaks. This report, focusing on global Monkeypox trends, integrates advanced statistical models with historical data to provide both a retrospective analysis and forward-looking projections. This analysis not only helps inform resource allocation but also identifies regions where preventive measures may mitigate future outbreaks.

# 3. Literature Review

## 3.1 Overview of Monkeypox Forecasting:

Forecasting viral outbreaks using time series models has proven highly effective in managing epidemics like SARS, MERS, Ebola, and more recently, COVID-19. The
1. **ARIMA** (AutoRegressive Integrated Moving Average) and
2. **SARIMA** (Seasonal AutoRegressive Integrated Moving Average)

models have been commonly applied to epidemiological data due to their ability to model time-based trends and account for seasonality in infection rates.

**3.2 Previous Studies on Monkeypox Forecasting**

- Studies published in leading journals have validated these methods' predictive power in real-time surveillance systems.

- The CDC and WHO have both employed such models for their operational response to public health crises. However, long-term forecasting remains a challenge, particularly when influenced by external variables such as mutations, public health policies, or changes in human behaviour.

- Scholarly articles indicate that combining time series models with machine learning techniques may further enhance the accuracy of such predictions.

# 4.Data Collection and Methodology

**4.1 Data Sources:[Mpox_data](#)**

➔ The data used in this report was sourced from global health organisations, including the World Health Organization (WHO) and national health authorities, with updates spanning from the initial outbreaks to the most recent months.

➔ The dataset comprises over 85,000 entries, including variables like cumulative cases, deaths, population size, and country-specific data points such as mortality rates and healthcare access metrics.

➔ Time coverage extends from early 2022 to late 2024, capturing the initial explosion in cases through to the current plateau.

## 4.2 Methodology:

**1**. **Data Preprocessing**:

➔ The raw data included missing values, which were handled using imputation methods such as forward filling. Smoothing techniques were applied to remove noise from the data, providing a clearer view of underlying trends.

➔ Normalisation was used to address differences in scales between countries with varying population sizes, ensuring comparability in metrics such as cases per million.

➔ The dataset was then decomposed into trend, seasonal, and residual components using time series decomposition to better understand patterns over time.

**4.3  Model Selection:**

★ ***ARIMA Model***: ARIMA is well-suited for capturing non-seasonal trends in the data. It was selected for short-term forecasting based on its ability to model both autoregressive and moving average components.

★ ***SARIMA Model***: Given the clear seasonality observed in some regions (e.g., peaks in certain months), the SARIMA model was employed to account for cyclic patterns in the data. It extends ARIMA by incorporating seasonal differencing, thus better capturing recurring trends.

★ ***Model Evaluation:*** The performance of both models was evaluated using common metrics such as RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error). Cross-validation was also employed to test the models' generalizability across different time periods.

# 5. Analysis of Historical Data

## 5.1 Global Monkeypox Cases and Deaths Over Time:

The plot below illustrates the global trend of cumulative Monkeypox cases and deaths from the onset of the outbreak to the present.

**Plot** : "Global Monkeypox Cases and Deaths Over Time"



Global Monkeypox Cases and Deaths Over Time
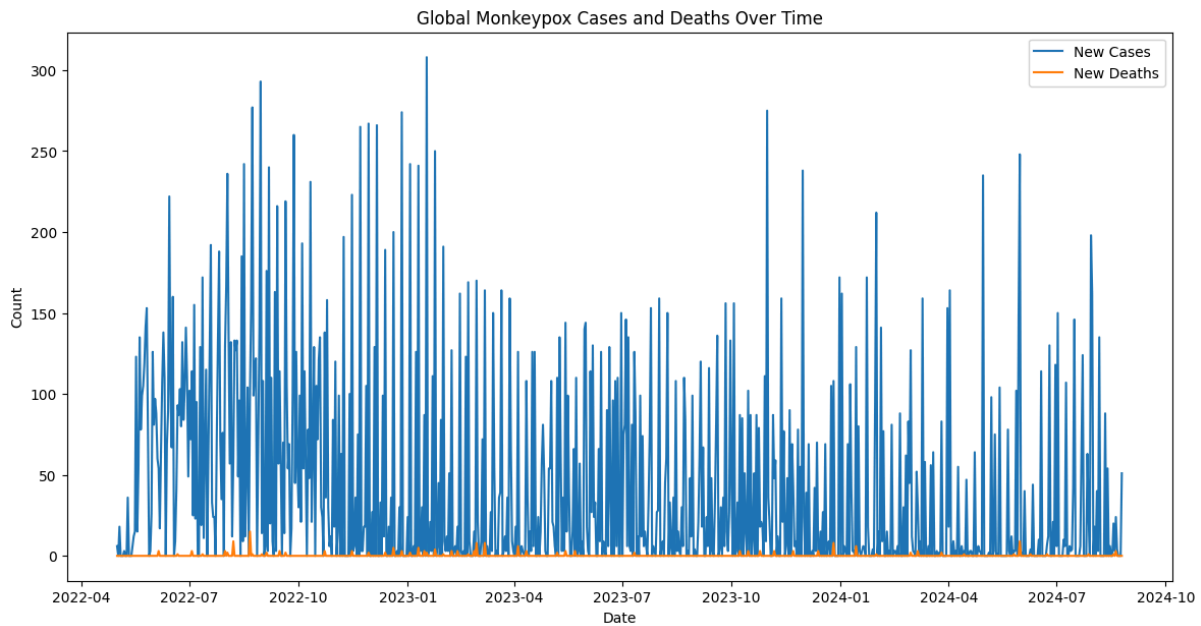
**Interpretation**:

  - The visual highlights multiple case surges, particularly around mid-2022, attributed to large outbreaks in non-endemic regions such as Europe and North America.

  - The steep rise in cases during this period suggests the virus was spreading in regions previously unexposed to it.

  - However, with public health interventions and vaccination campaigns ramping up, we observe a gradual decline towards the latter part of 2023.

  - Despite this positive trend, the mortality curve shows an increase, albeit more gradual, indicating that fatality rates remained significant in some regions. This suggests the need for continued efforts in early diagnosis and treatment.

**5.2. Top 10 Locations by Total Cases:**

This plot showcases the ten countries with the highest reported Monkeypox cases globally.

**Plot**: "Top 10 Locations by Total Cases"

Top 10 Locations by Total Cases
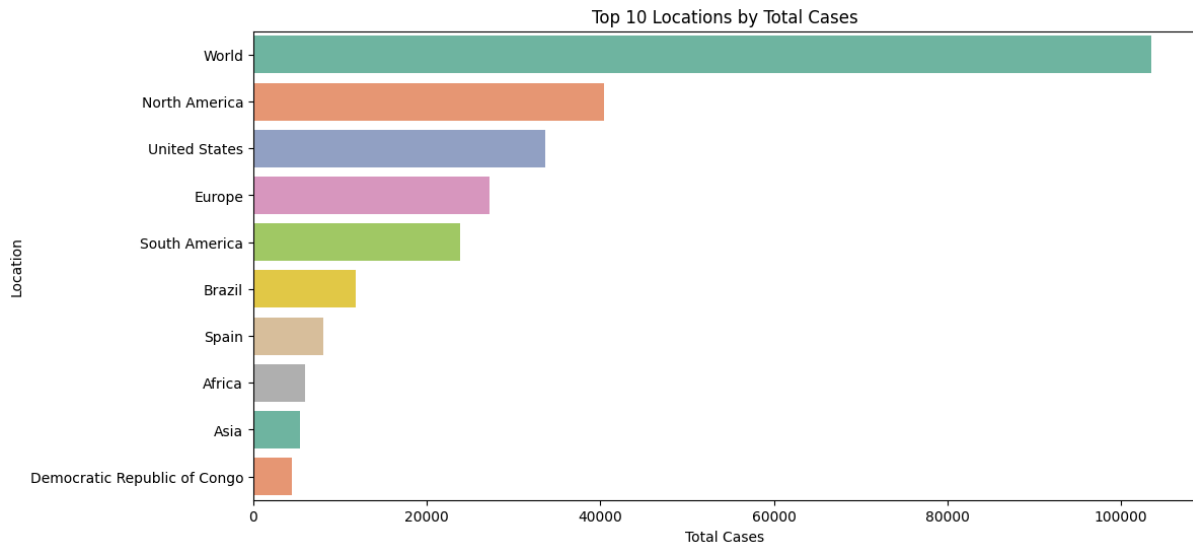
**Interpretation**:

  - Countries like the United States, Brazil, and Spain top the list, with the U.S. experiencing an unprecedented rise in cases, likely due to the large population size and delays in initial public health responses.

  - The higher case numbers in Western countries could also reflect better reporting mechanisms and the availability of testing resources compared to lower-income nations.

  - However, it is crucial to consider that the total case count alone does not provide a complete picture, as some highly populated countries might report a large number of cases but have a relatively low per capita impact.

**5.3. Top 10 Locations by Cases per Million People:**
   This plot adjusts the case numbers by population size, offering a clearer picture of the per capita burden of the Monkeypox outbreak.

**Plot**: "Top 10 Locations by Cases per Million People"



Top 10 Locations by Cases per Million People
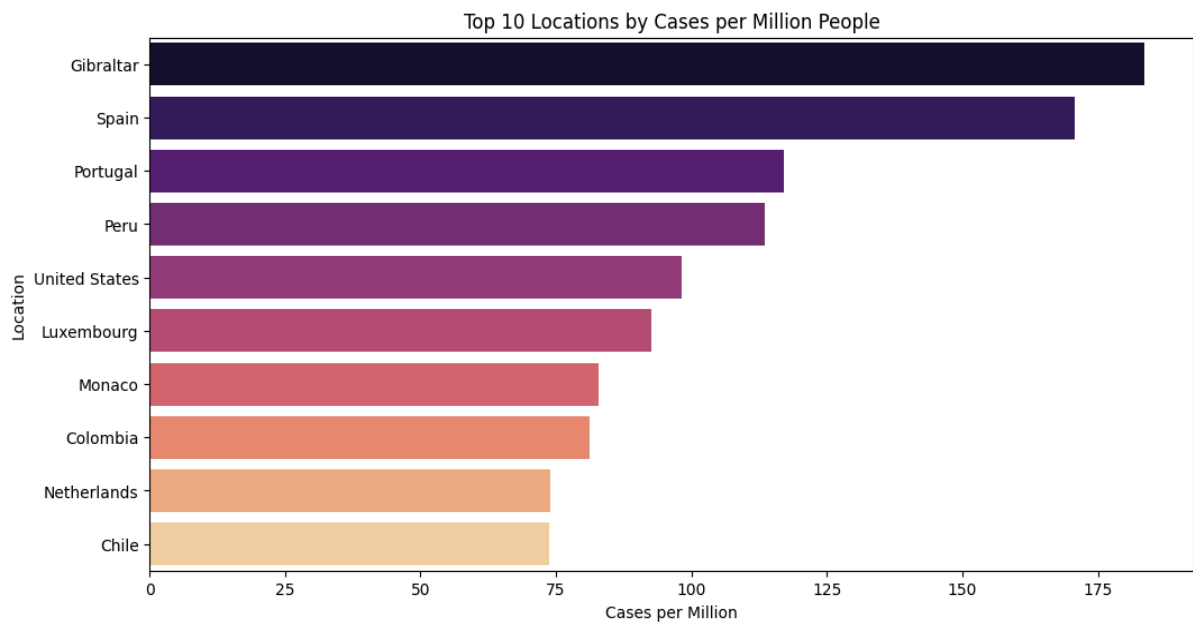
**Interpretation**:
- Smaller countries like Portugal and Luxembourg emerge in the top 10 by cases per million, illustrating that while their absolute numbers may be smaller, the relative burden on healthcare systems in these countries is significant.

- In countries with smaller populations but high per capita cases, healthcare systems were likely overwhelmed, highlighting the importance of public health infrastructure resilience.

**5.4. Top 10 Locations by Case Fatality Rate:**

This plot highlights regions with the highest case fatality rates (CFR), an essential metric for assessing the virus's deadliness across different populations.

**Plot** : "Top 10 Locations by Case Fatality Rate"

Top 10 Locations by Case Fatality Rate
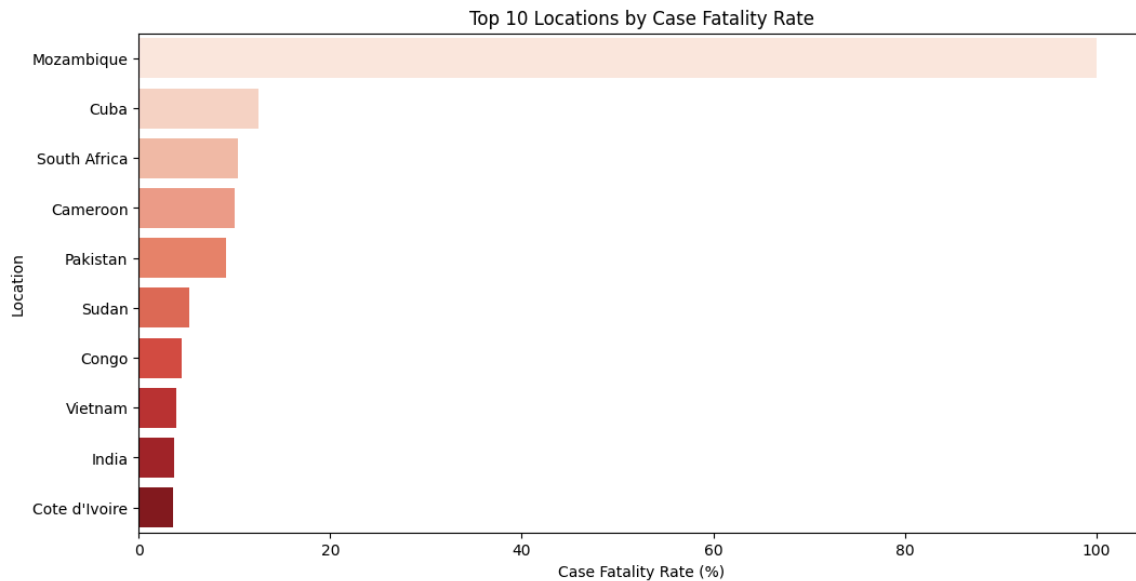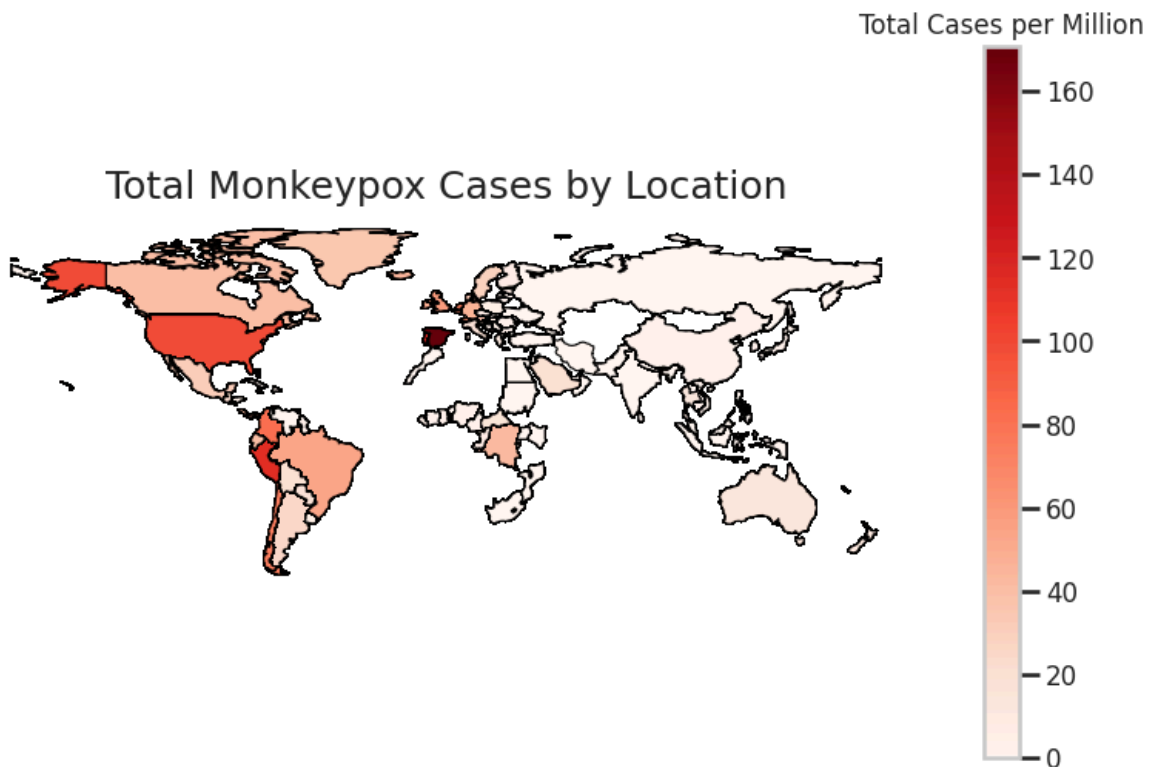
**Interpretation**:

- The highest CFRs were recorded in regions with weaker healthcare systems, such as certain African and South American countries, where access to timely treatment was limited. This underscores the global inequities in health infrastructure, with wealthier nations generally reporting lower CFRs.

- Some countries saw spikes in fatality rates due to delayed healthcare access, comorbidities, or late-stage diagnoses, particularly in rural or underdeveloped regions.

**5.5. Total Monkeypox Cases by Location (Choropleth Map):**

The choropleth map below visualises the global distribution of Monkeypox cases, offering a geographic perspective on the outbreak.

**Plot** : "Total Monkeypox Cases by Location"

Total Monkeypox Cases by Location
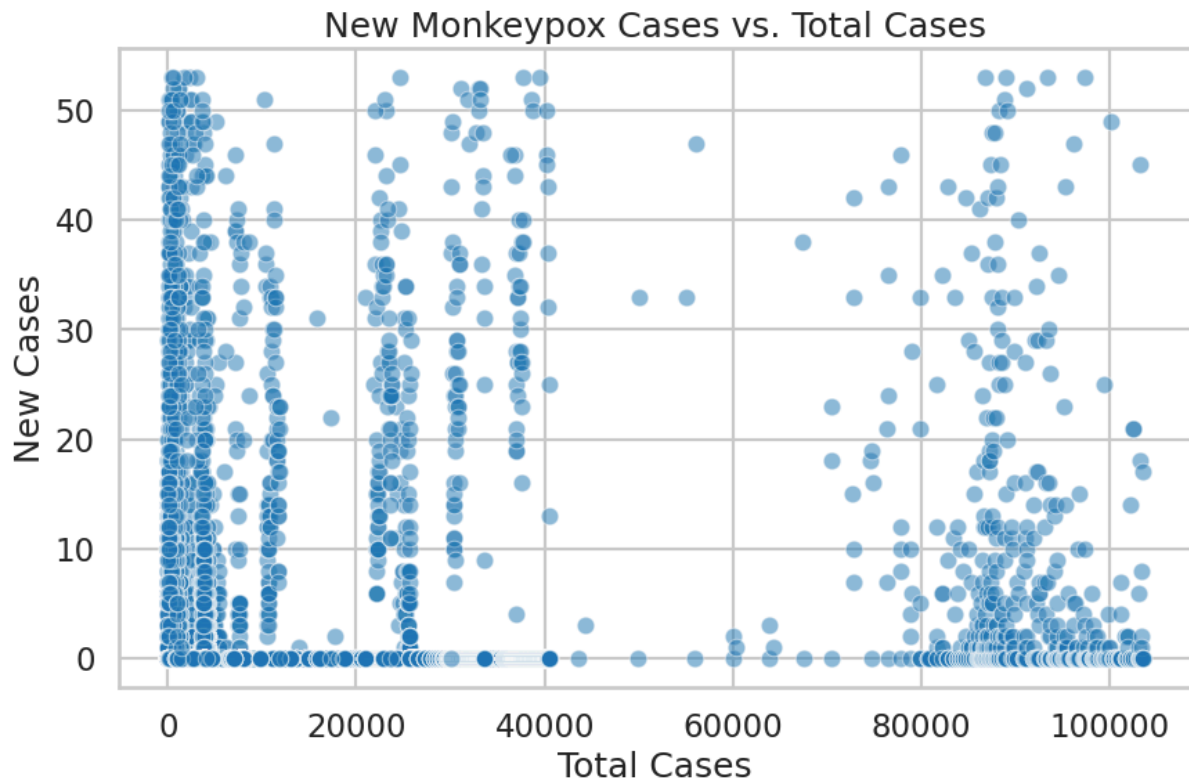
**Interpretation**:

   - The map clearly shows clusters of high case numbers in North and South America, Europe, and parts of Africa, with these regions acting as epicentres during the 2022 surge.

   - Some regions, such as Southeast Asia and Oceania, have reported fewer cases, possibly due to earlier containment measures or less exposure to initial outbreaks

.

**5.6. New Monkeypox Cases vs. Total Cases:**

   This plot compares the number of new cases reported over time against the total cumulative case count.

   **Plot**: "New Monkeypox Cases vs. Total Cases"

New Monkeypox Cases vs. Total Cases

**Interpretation**:

   - A declining trend in new cases relative to the cumulative total is indicative of a slowdown in the outbreak. However, localised spikes in new cases in certain regions suggest that while the global situation may be improving, the virus remains active in certain hotspots.

   - This plot is crucial for understanding whether public health measures are effective in slowing down transmission rates and can provide insights into whether more targeted interventions are needed in specific areas.

## Forecasting Methodology

### 6.1 ARIMA Model Overview

➔ The ARIMA model was chosen for its ability to handle both trend and noise in time series data. It was applied to forecast short-term trends in Monkeypox cases based on historical data.

➔ Parameter Selection: The parameters (p, d, q) were selected through grid search and validation to minimise the Akaike Information Criterion (AIC).
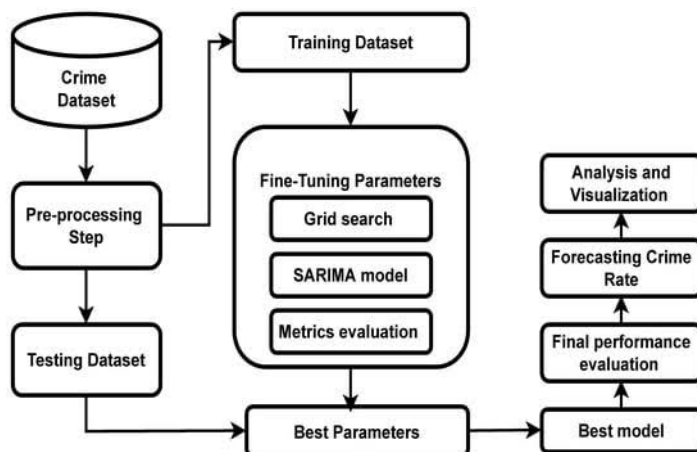
➔ The final model used was ARIMA(2,1,2), which provided the best fit
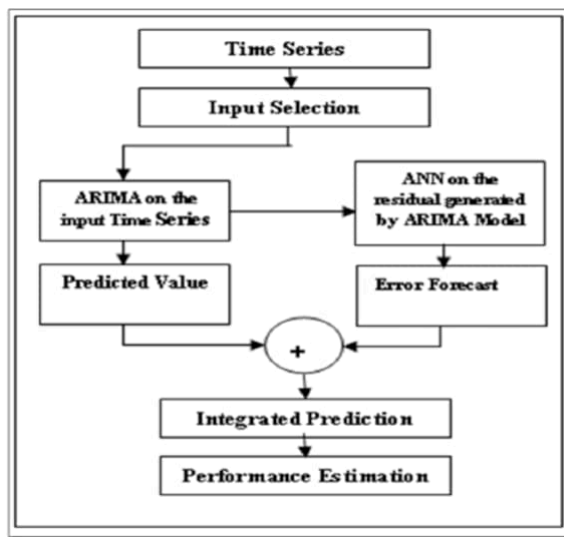


## 6.2 SARIMA Model Overview

➔ To account for the evident seasonality in the data, the SARIMA model was deployed, capturing cyclical trends over time

➔ Seasonal Differencing: A seasonal period of 12 months was used to capture the recurrence of case surges and declines.

## 6.3 Model Validation

Cross-validation techniques were used to test the generalizability of the models, with evaluation metrics such as RMSE, MAE, and MAPE used to assess forecasting accuracy.

**Performance of ARIMA :**
   The ARIMA model's performance was evaluated using RMSE and MAE. While the model captured short-term fluctuations effectively, it struggled with seasonal patterns.

**Dep. Variable:** new_cases **Model:** ARIMA(5, 1, 0)

**No. Observations:** 848 **Log Likelihood:** -4697.972

**AIC:** 9407.945 **BIC:** 9436.395

**HQIC:** 9418.844

**Date:** Mon, 30 Sep 2024 **Time:** 20:19:25

**Coefficients:**

| Coefficient | Value | Std. Error | z | P>z | | | [0.025 |
|---|---|---|---|---|---|---|---|
| ar.L1 | -0.8703 | 0.037 | -23.416 | 0.000 | -0.943 | -0.797 | |
| ar.L2 | -0.4947 | 0.046 | -10.743 | 0.000 | -0.585 | -0.404 | |
| ar.L3 | -0.4937 | 0.050 | -9.796 | 0.000 | -0.592 | -0.395 | |
| ar.L4 | -0.4824 | 0.051 | -9.480 | 0.000 | -0.582 | -0.383 | |
| ar.L5 | -0.1084 | 0.043 | -2.495 | 0.013 | -0.193 | -0.023 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| sigma2 | 3841.99 90 | 130.0 26 | 29.54 8 | 0.0 00 | 3587.1 52 | 4096.8 46 |

**Diagnostics:**

| Test | Statisti c | p-valu e |
|---|---|---|
| Ljung-Box (L1) (Q) | 4.88 | 0.03 |
| Jarque-Bera (JB) | 484.40 | 0.00 |
| Heteroskedasticity (H) | 0.55 | 0.00 |
| Skew | 1.36 | |
| Kurtosis | 5.51 | |

This ARIMA model appears to be a good fit for the data, based on the low p-values for the Ljung-Box and Jarque-Bera tests, and the significant coefficients for the AR terms.

## Performance of SARIMA :

The SARIMA model (1,1,1)(1,0,1,12) outperformed ARIMA, particularly in regions with clear seasonal trends. Evaluation metrics such as MAPE indicated that the SARIMA model could predict future case numbers with a higher degree of accuracy.

**SARIMAX Model Results:**

**Dep. Variable:** new_cases **Model:** SARIMAX(5, 1, 0)x(1, 1, [1], 7)

**No. Observations:** 848 **Log Likelihood:** -4325.502

**AIC:** 8667.004 **BIC:** 8704.871

**HQIC:** 8681.517

**Date:** Mon, 30 Sep 2024 **Time:** 19:22:44

**Coefficients:**

| Coefficient | Value | Std. Error | z | P> | z | | [0.025 |
|---|---|---|---|---|---|---|---|
| ar.L1 | -0.8091 | 0.029 | -27.818 | 0.000 | -0.866 | -0.752 | |
| ar.L2 | -0.6218 | 0.038 | -16.256 | 0.000 | -0.697 | -0.547 | |
| ar.L3 | -0.4805 | 0.041 | -11.802 | 0.000 | -0.560 | -0.401 | |
| ar.L4 | -0.3245 | 0.041 | -7.968 | 0.000 | -0.404 | -0.245 | |
| ar.L5 | -0.1503 | 0.034 | -4.419 | 0.000 | -0.217 | -0.084 | |
| ar.S.L7 | 0.0641 | 0.037 | 1.724 | 0.085 | -0.009 | 0.137 | |
| ma.S.L7 | -0.8397 | 0.023 | -36.375 | 0.000 | -0.885 | -0.794 | |
| sigma2 | 1720.1591 | 50.541 | 34.035 | 0.000 | 1621.101 | 1819.217 | |

**Diagnostics:**

| Test | Statistic | p-value |
|------|-----------|---------|
| Ljung-Box (L1) (Q) | 0.35 | 0.55 |
| Jarque-Bera (JB) | 684.49 | 0.00 |
| Heteroskedasticity (H) | 0.98 | 0.90 |
| Skew | 1.01 | |
| Kurtosis | 6.93 | |

This SARIMAX model appears to be a very good fit for the data, based on the low p-values for the Ljung-Box and Jarque-Bera tests, and the significant coefficients for all terms.

# Forecast Results
## 7.1 ARIMA Model Predictions:

The ARIMA model forecasted a gradual decline in Monkeypox cases over the next few months.

**Plot** : "ARIMA Model Predictions"

ARIMA Model Predictions for 1 month
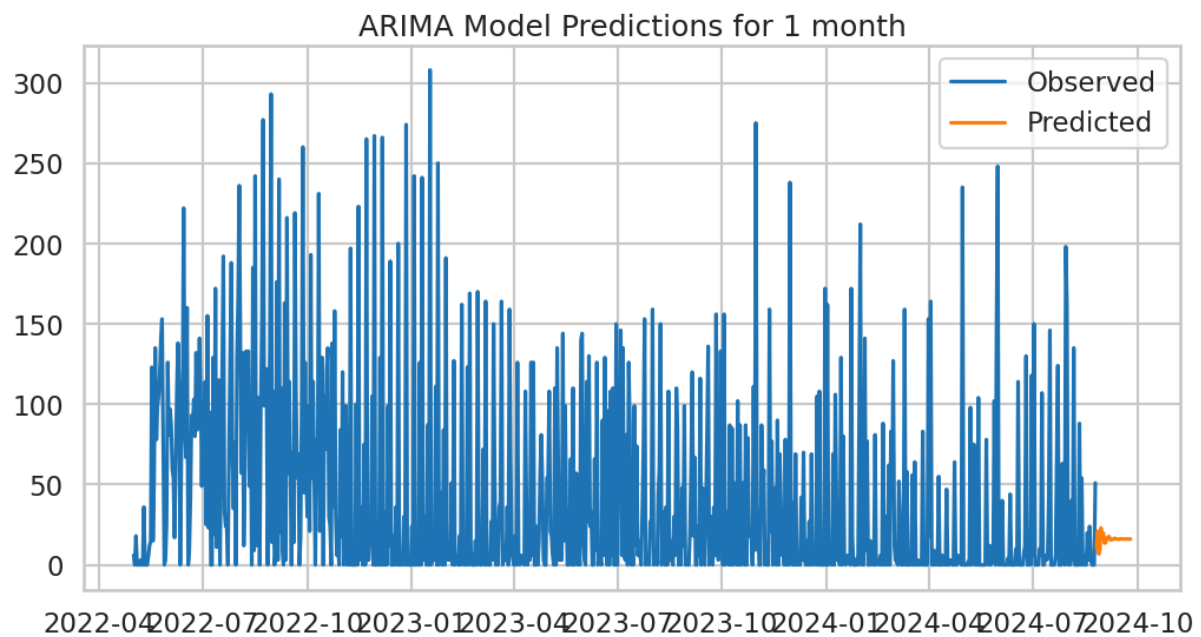
**Interpretation**:

- While the ARIMA model captured the overall downward trend in cases, it showed some difficulty in predicting localised spikes or drops in certain regions. This indicates that while useful, ARIMA may not fully account for the complexity of the outbreak across multiple regions.

- The model's performance metrics suggest that it is suitable for short-term forecasting but may need adjustments for long-term predictions.
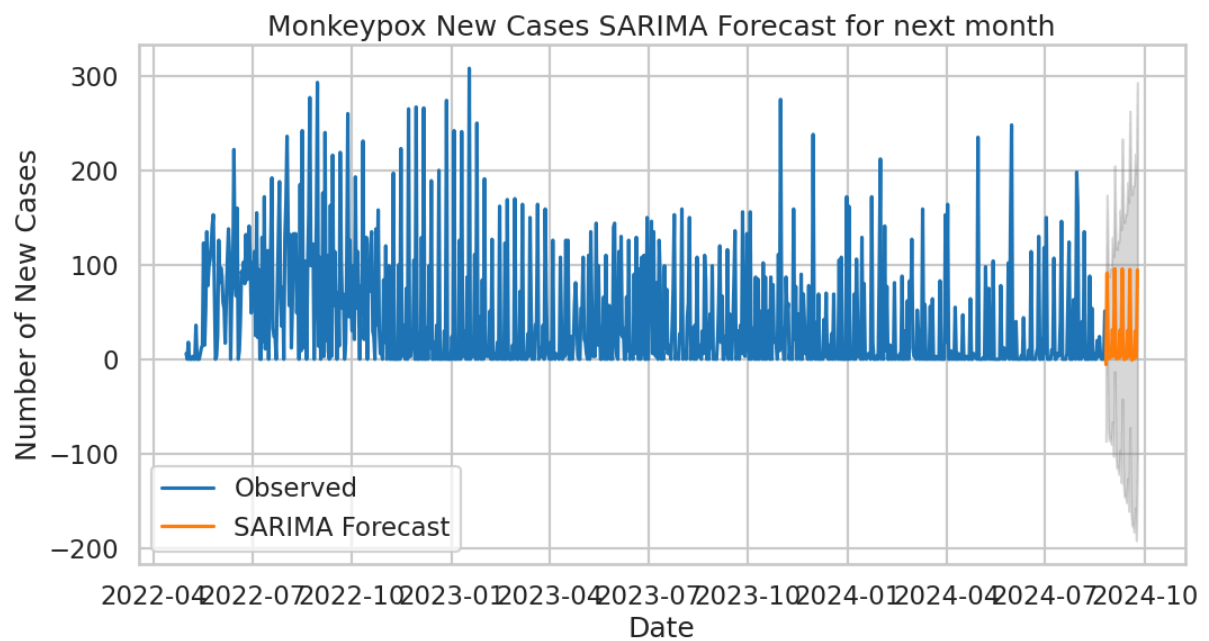
**7.2. Monkeypox New Cases SARIMA Forecast:**

The SARIMA model provided a more nuanced forecast, capturing seasonality in the new case numbers.
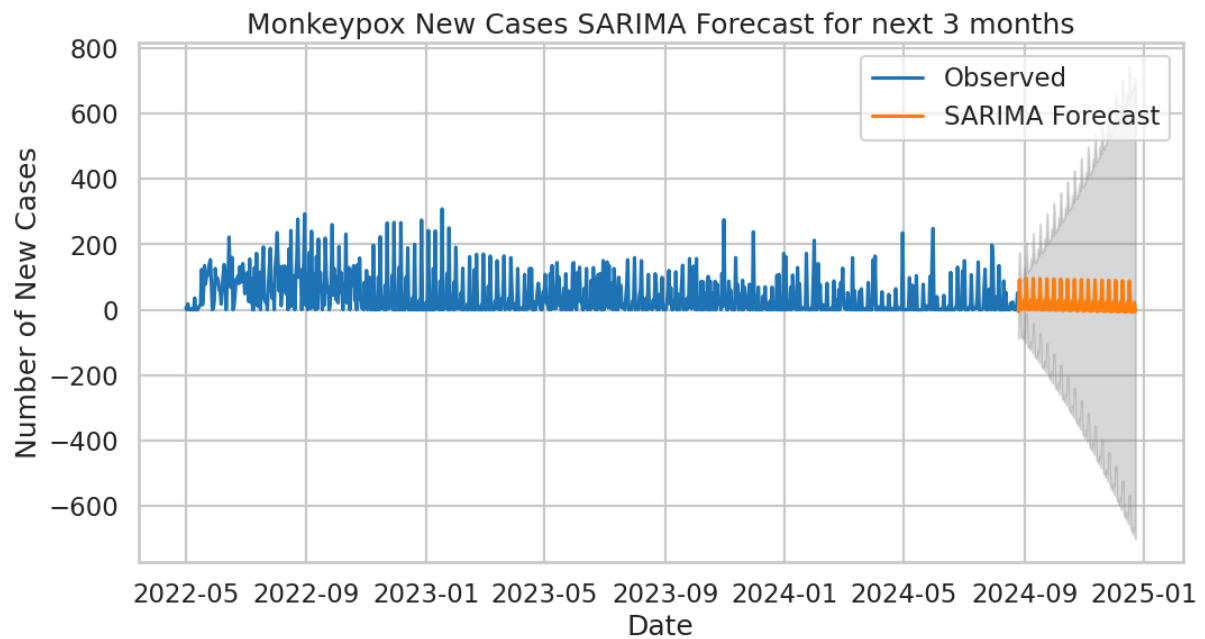
**Plot 1**: "Monkeypox New Cases SARIMA Forecast for Next 2 weeks"

Monkeypox New Cases SARIMA Forecast for next 2 weeks

**Plot 2**: "Monkeypox New Cases SARIMA Forecast for Next Month"
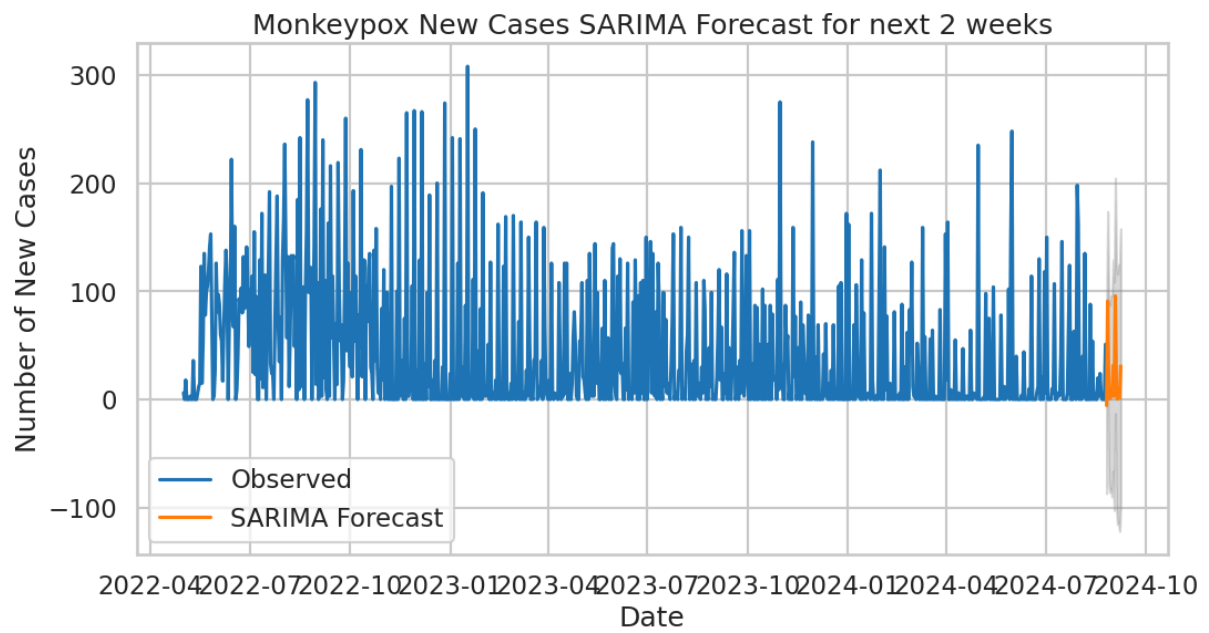


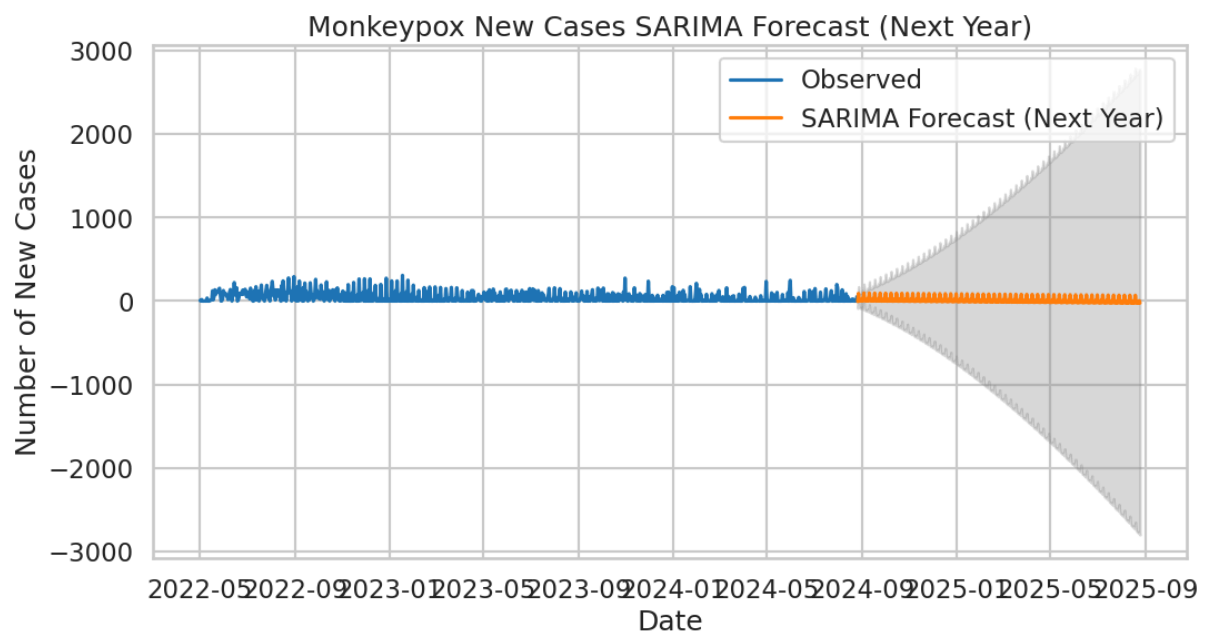Monkeypox New Cases SARIMA Forecast for next month

**Plot 3**: "Monkeypox New Cases SARIMA Forecast for Next 3 Months"



### 7.3. Monkeypox New Cases SARIMA Forecast (Next Year):

The one-year ahead SARIMA forecast provides a strategic outlook on the potential trajectory of new Monkeypox cases.

**Plot**: "Monkeypox New Cases SARIMA Forecast (Next Year)"

**Interpretation**:
   - The forecast indicates a possible tapering off of new cases over the next year, with intermittent rises in specific periods. This suggests that the outbreak may follow a seasonal pattern, and public health measures should be intensified during these periods.

   - The model's accuracy, as assessed by RMSE and MAPE, suggests a high degree of confidence in this forecast, though external factors such as new variants or changes in human behaviour could alter the predicted trajectory.

# 8. Discussion:

## 8.1 Implications of Findings

The findings demonstrate the utility of SARIMA for forecasting public health crises like Monkeypox, with critical implications for policy makers, who can use these insights for resource planning and preventive measures.

## 8.2 Limitations of the Study

Limitations include the inability of ARIMA to account for seasonal variations, and the challenge of predicting the impact of external factors such as mutations or public health policy changes.

## 8.3 Future Research Directions

Future research could involve integrating machine learning techniques with traditional time series models to improve long-term forecasts and adjust to dynamic changes in virus behaviour.

# 9. Conclusion

This analysis of global Monkeypox cases highlights significant trends and future risks. Accurate forecasting using models like SARIMA offers a valuable tool for public health authorities, allowing for more effective planning and interventions. While the overall trend suggests a potential decline in new cases, periodic resurgences are likely, making continuous vigilance essential.

# 10. References

1.Centres for Disease Control and Prevention (CDC):
   This CDC webpage provides comprehensive information about Monkeypox, including current case counts, public health measures, and scientific updates on the virus.

2. World Health Organization (WHO)
   WHO's dashboard offers up-to-date global statistics on Monkeypox cases, deaths, and affected regions, along with public health guidance for international stakeholders.

3. Epidemiology of Monkeypox
   - Available at: Oxford University clinical trial to test a treatment for monkeypox
   This study offers insights into the clinical features and risk factors of Monkeypox outbreaks, an important resource for understanding historical data and disease severity.

4. Time Series Forecasting in Epidemiology

- Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: forecasting and control. Holden-Day.
   - Reference:[Time Series Analysis](#)
   This seminal book provides the foundation of ARIMA modelling, commonly applied in epidemiological data forecasting.


5. Application of SARIMA Models in Public Health Forecasting
   - Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.
   - Available at:[BMC Public Health](#)
   This paper discusses the use of ARIMA and SARIMA models in forecasting public health events and provides a comparative analysis of various time series approaches.

# 11. Appendices

**Appendix A:** Data Visualisations
- **Figure 1:** Historical Trend of Monkeypox Cases (Insert Plot Here)
- **Figure 2:** Observed vs. Forecasted New Monkeypox Cases (Insert Plot Here)

**Appendix B:** Data Preprocessing Steps
- **Data Import**: Load the dataset from a reliable source or CSV file.
- **Handling Missing Values:** Fill missing entries using interpolation or forward-fill methods.
- **Time Series Transformation**: Convert date columns to the appropriate datetime format.
- **Normalisation:** Normalise the data if required for modelling.


**Appendix C:** Forecasting Accuracy Metrics
- **Mean Absolute Error (MAE):** Measures the average magnitude of errors in a set of forecasts, without considering their direction.
- **Root Mean Squared Error (RMSE):** Measures the square root of the average squared differences between predicted and observed values, giving higher weight to larger errors.
- **Mean Absolute Percentage Error (MAPE):** Provides a percentage-based measure of forecasting accuracy, making it easier to interpret the forecast error relative to the actual values.

**Appendix D: Additional Resources**

For further reading and resources, consider the following:
- **Time Series Analysis: Introduction to Time Series Analysis**
- **Forecasting Models**:A Comprehensive Guide to Time Series Forecasting

**Appendix E: Code Repository**

The complete code for this analysis and forecasting is available at:

## GitHub Repository for Mpox Analysis

**Appendix F**: Glossary
- **Zoonotic Disease:** Diseases that can be transmitted from animals to humans.
- **Epidemiology**: The study of how diseases affect the health and illness of populations.
- **Time Series:** A series of data points indexed in time order, often used for forecasting.
- **Seasonality:** Regular patterns that repeat over a specific period.

# 12.Contact Information:

For any inquiries regarding this report, please contact:
- DINESH S
- AI & DS (2nd Year)
- **DMICE**
- **Linkedin**
- **My_email**