Modeling and Forecasting of COVID-19 Cases in Kenya using ARIMA Model

Research Proposal Presented By:

Benson Mwangi, Christopher Njimu, Collins Kipkirui, Faith Chepkoech

Introduction

- Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus, whose family of viruses also caused SARS(2002) and MERS(2012).
- The first case of human infection was reported in Wuhan, China (Dec,2019).
- 11th March 2020, declared a pandemic by the WHO.
- 13th March 2020, first confirmed case in Kenya.
- 17 months into the pandemic, Kenya has had over 220,000 confirmed cases, 4 300 attributable deaths and currently cases are surging into the 4th wave.
- The pandemic has also had serious implications on widening social inequalities, loss of livelihoods, reduced economic productivity, stretched the health sector, disrupted the school calendar.
- With still a long way to go in terms of increasing percentage of vaccinated population, and the virus property of constantly mutating into new variants, there's still need to reduce the spread by informing decisions using updated data, in order to cushion livelihoods while saving lives.

Problem statement

- Covid-19 pandemic has disrupted the economic and health sectors of the country. It has caused deaths, reduced productivity, business closures, trade disruptions, school calendar disruption and decimation of the tourism industry.
- Community transmission is happening within the country, making it hard to trace sources of exposure.
- Need to predict possible future surges in order to tighten measures in good time to avoid loss of lives and interrupt predicted surges.
- Need for regular updating of models to factor in new variants and vaccination progress patterns.
- Government needs COVID-19 forecasts to inform its decisions.
- Our study focused on developing a suitable ARIMA model that was used to forecast the future trend of COVID-19 cases in Kenya.

Justification

- Through modeling and forecasting of the covid-19 cases, it will establish the likely Impact of the pandemic and help inform national and local government planning, budgeting, readiness, response and monitoring.

- Help the government in adjusting existing measures accordingly, by striking a fine balance in between saving lives and maintaining livelihoods.

- Help the health sector to proper plan through anticipating future possible surges to avoid a similar situation as that seen in India.

General and specific objectives

General objective

To model and forecast COVID-19 cases in Kenya using ARIMA model.

Specific objectives.

- To identify the time series components at work
- To analyze the stationarity of Covid-19 cases in Kenya
- To develop an ARIMA model for Covid-19 cases in Kenya
- To forecast Covid-19 cases for the next twelve weeks

Research Questions and Scope

Research Questions

- What time series components are at work in the Covid-19 cases Kenyan data?
- Is the Covid-19 cases data in Kenya stationary?
- What is the most suitable ARIMA model for the Covid-19 cases?
- What are the expected number of cases in the near future?

Scope of the study

This study analyzed covid-19 cases in Kenya from March 2020 to October 2021, fitted a suitable ARIMA model, and projected future covid-19 cases.

Literature Review

Previous related studies:

- India, (Gupta, 2020) conducted an analysis of COVID-19 cases using ARIMA model.
- Globally, (Chaurisia and Pal, 2020) applied ML in a time series analysis, comparing various models for prediction of the Covid-19 pandemic. ARIMA was the second best after Naïve method.
- Saudi Arabia,(Alzahrani et al.,2020) used the ARIMA model to forecast the spread of Covid-19 pandemic in Saudi Arabia considering the measures put in place.
- China, (Wang, 2018) modelled and compared the ARIMA and Grey Models for Hepatitis B monthly cases. ARIMA showing better performance than Grey Models.
- Kenya, (Sam el al., 2021) developed Otio-NARIMA model for forecasting seasonality of Covid-19 waves in Kenya.

Following this researched studies, we arrived at the conclusion that since the ARIMA model has been used before to model other susceptible diseases, including COVID-19 in other nations, the ARIMA model might be suitable to model COVID-19 cases in Kenya and to forecast the expected near-future tendencies.

Methodology

1.Research Design - Analytical research design.

2. Data

Weekly reported cases for the period *March 2020 – October 2021*.

R Statistical software, will be used for analysis of data.

3. Sources of data

Ministry of Health Kenya and Worldometer.

4. **The model: ARIMA model Procedural Steps:**

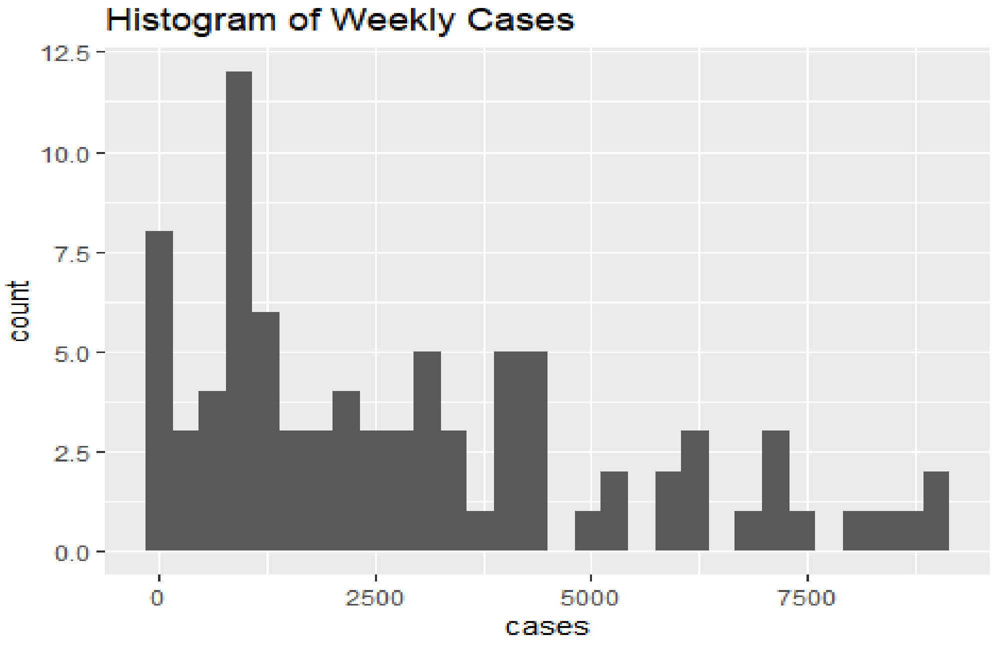
i. Data validation.
ii. Testing the stationarity of the data.
iii. Estimation of Parameters and order selection.
iv. Model validation.
v. Forecasting.

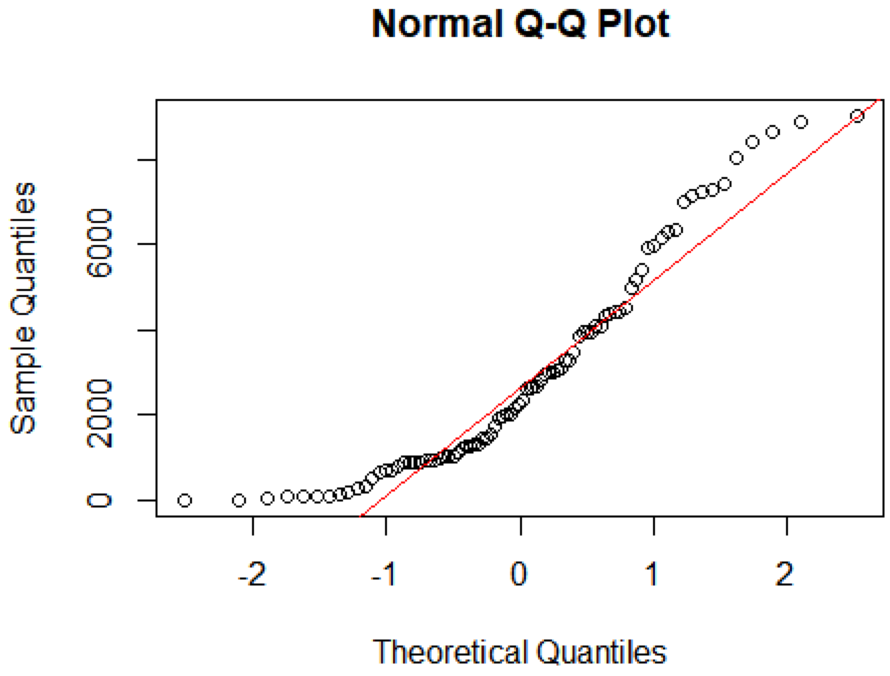
RESULTS AND DISCUSSION

Summary Statistics	
nobs	86
NAs	0
Minimum	7
Maximum	9014
1. Quartile	945.75
3. Quartile	4335.25
Mean	2945.465116
Median	2305
Sum	253310
SE Mean	267.02209
LCL Mean	2414.553665
UCL Mean	3476.376568
Variance	6131868.487
Stdev	2476.260989
Skewness	0.823212
Kurtosis	-0.361027

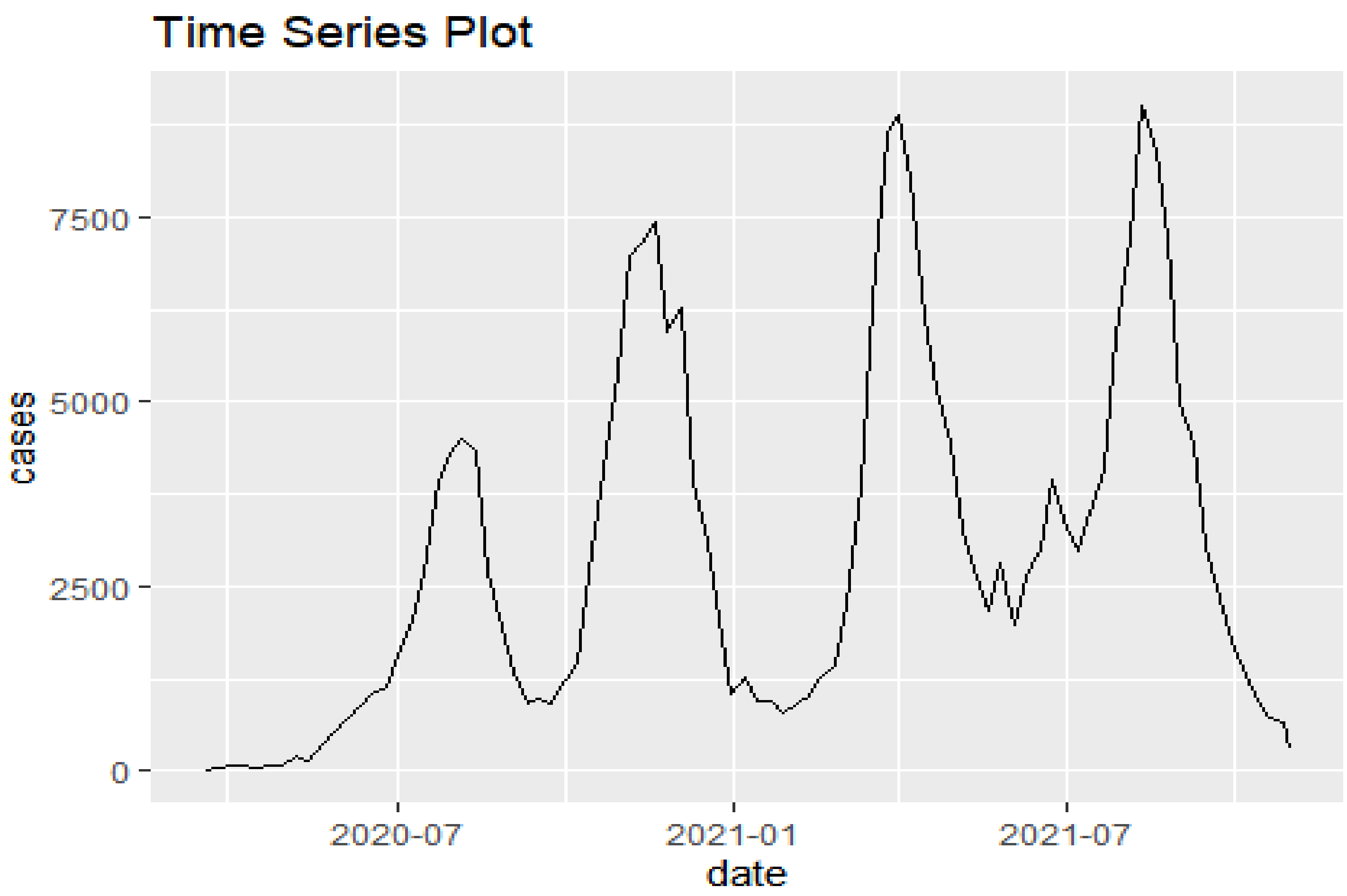
RESULTS AND DISCUSSION

Data Exploration

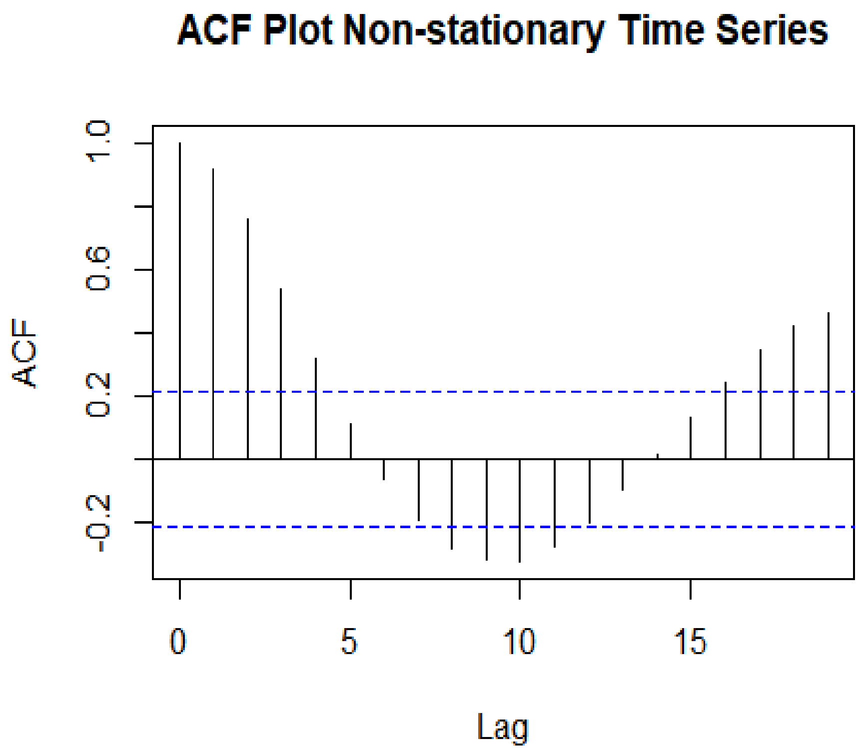


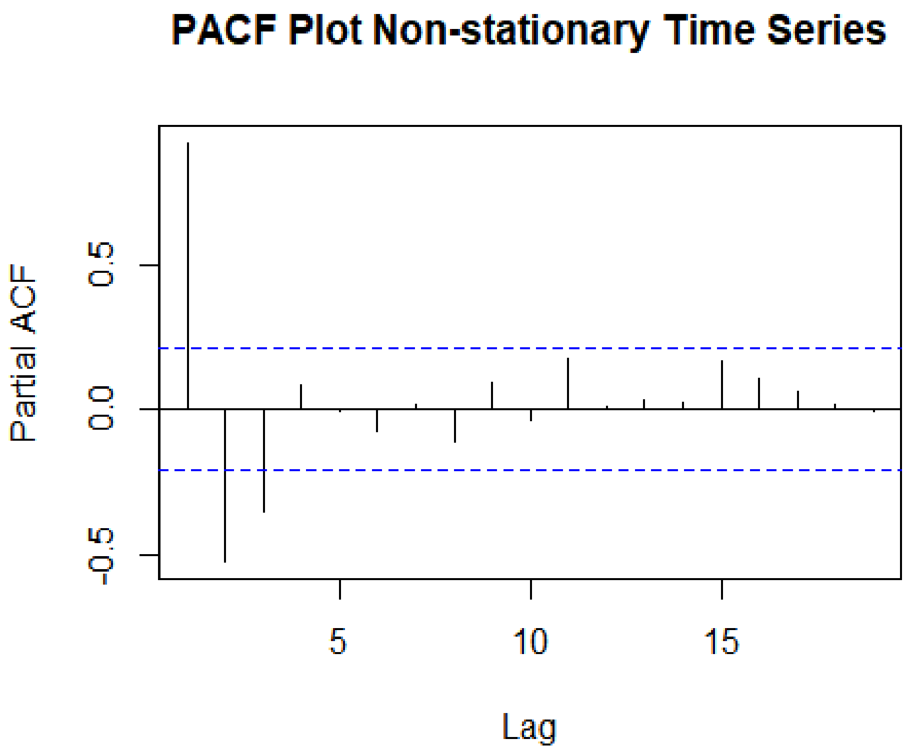


RESULTS AND DISCUSSION



RESULTS AND DISCUSSION





RESULTS AND DISCUSSION

Test for Stationarity

Augmented Dickey-Fuller (ADF) test at 5% level of significance.

Hypothesis

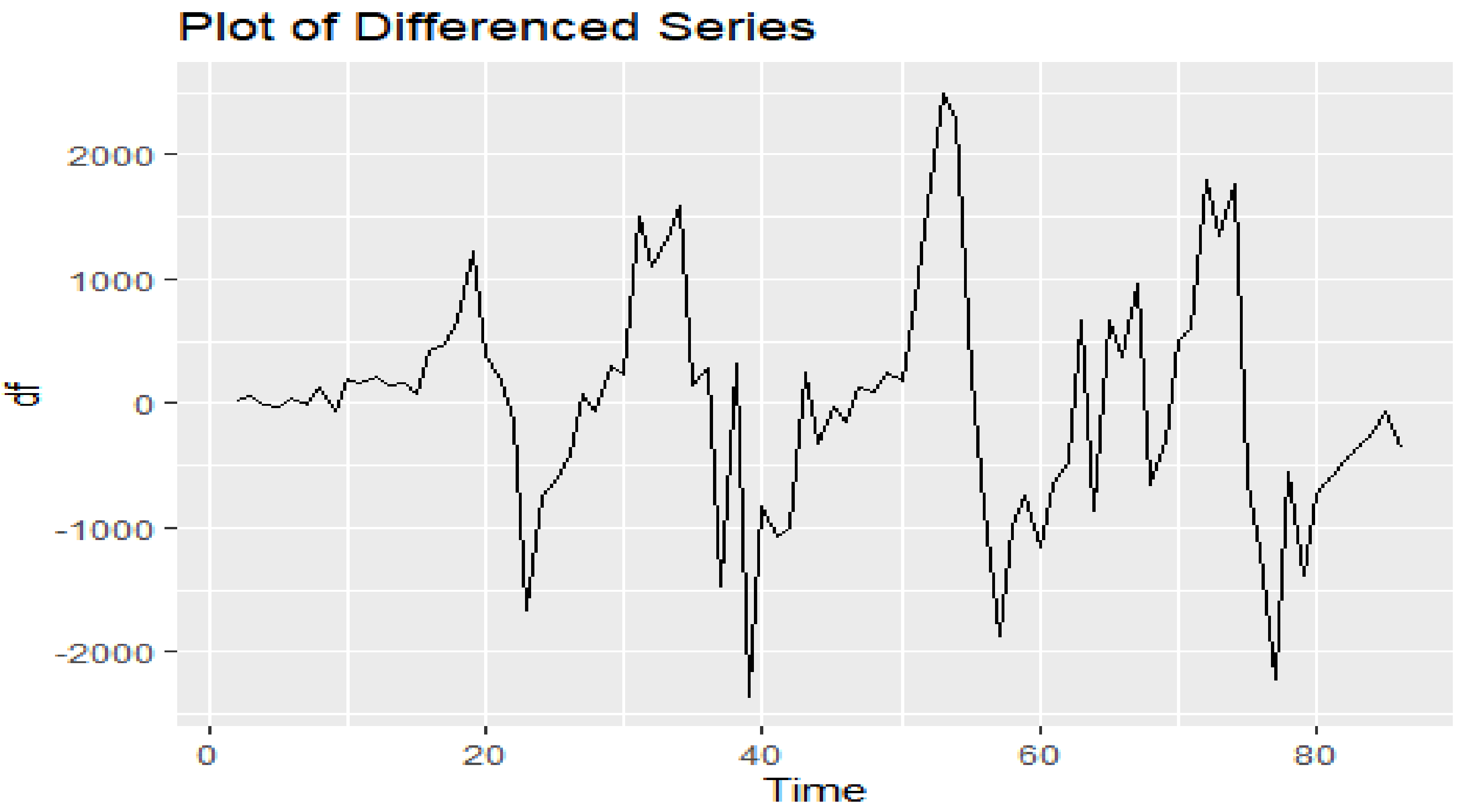
$\boldsymbol{H_0}$: The time series is non-stationary vs. $\boldsymbol{H_1}$: The time series is stationary.

R Code Output

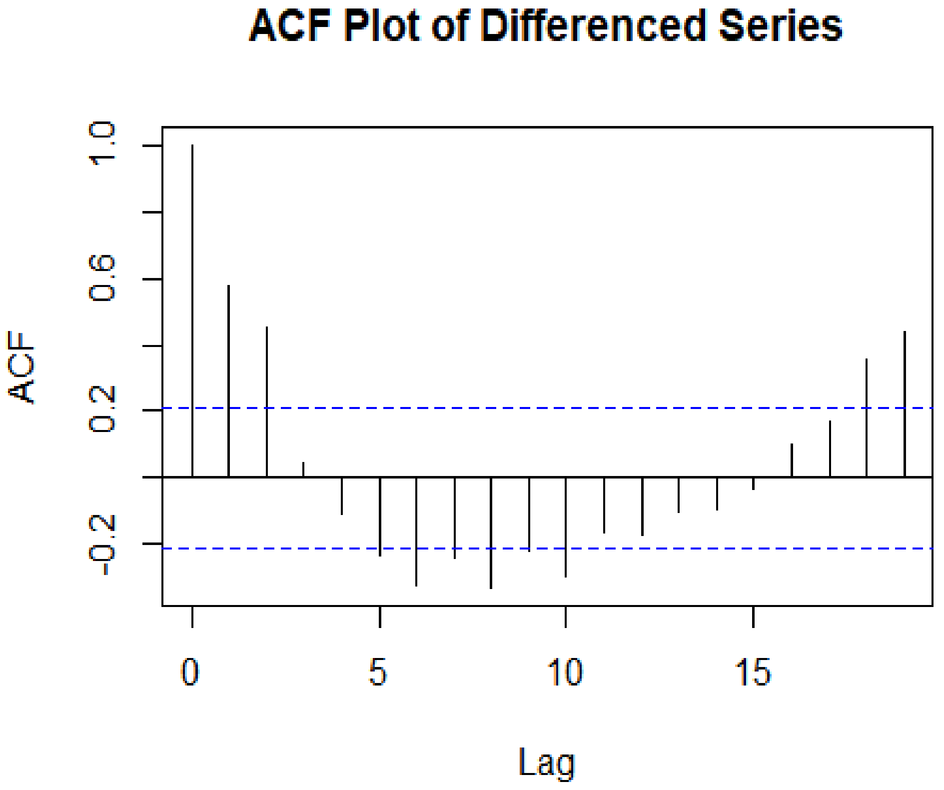
Augmented Dickey-Fuller Test

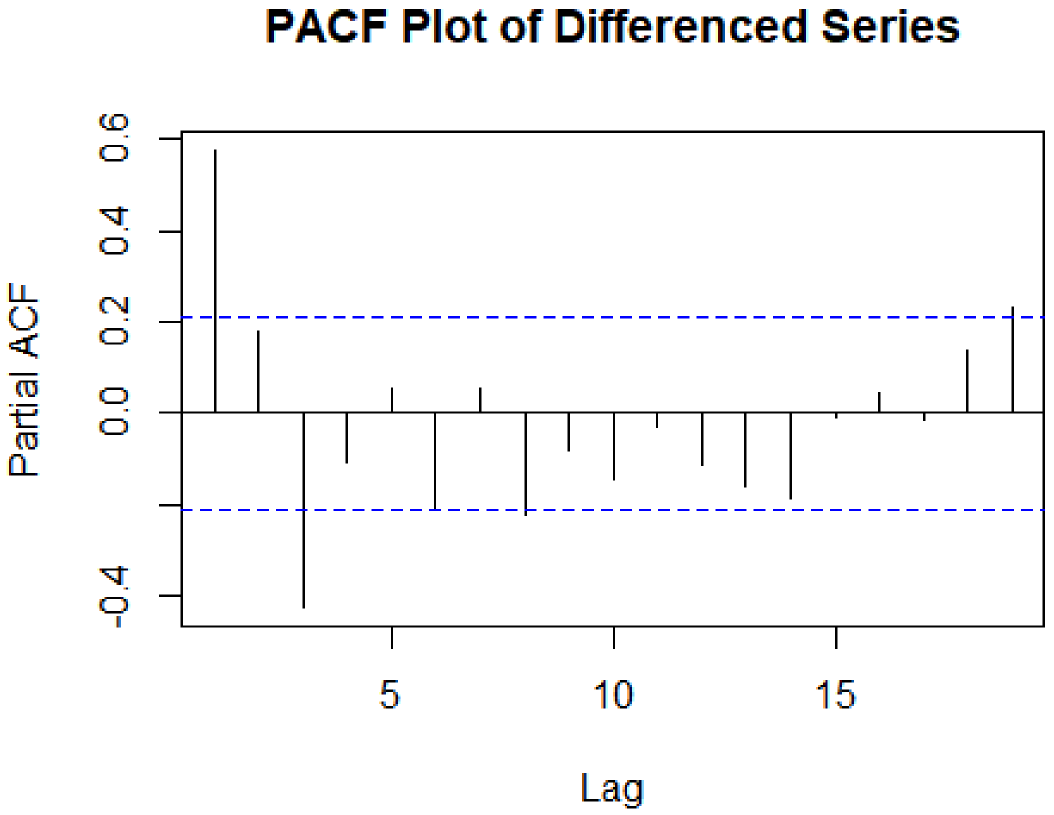
Dickey-Fuller = -3.1565, Lag order = 4, p-value = 0.101
alternative hypothesis: stationary

RESULTS AND DISCUSSION



Model Identification





Model Identification

ARIMA MODEL		AICc
ARIMA(0,1,4)		277.8872
ARIMA(0,1,5)		278.3766
ARIMA(1,1,2)		277.8917
ARIMA(1,1,4)		278.6327
ARIMA(2,1,2)		278.6227
ARIMA(3,1,0)		277.7939
ARIMA(3,1,1)		276.2863
ARIMA(3,1,1)	with drift	278.2367
ARIMA(3,1,1)(0,0,1)[12]		278.5256
ARIMA(3,1,1)(1,0,0)[12]		278.5355
ARIMA(3,1,2)		278.525
ARIMA(4,1,0)		276.3725
ARIMA(4,1,0)	with drift	278.5001
ARIMA(4,1,0)(0,0,1)[12]		278.5226
ARIMA(4,1,0)(1,0,0)[12]		278.5417
ARIMA(4,1,1)		276.6969
ARIMA(4,1,1)	with drift	278.9212
ARIMA(5,1,0)		278.6834

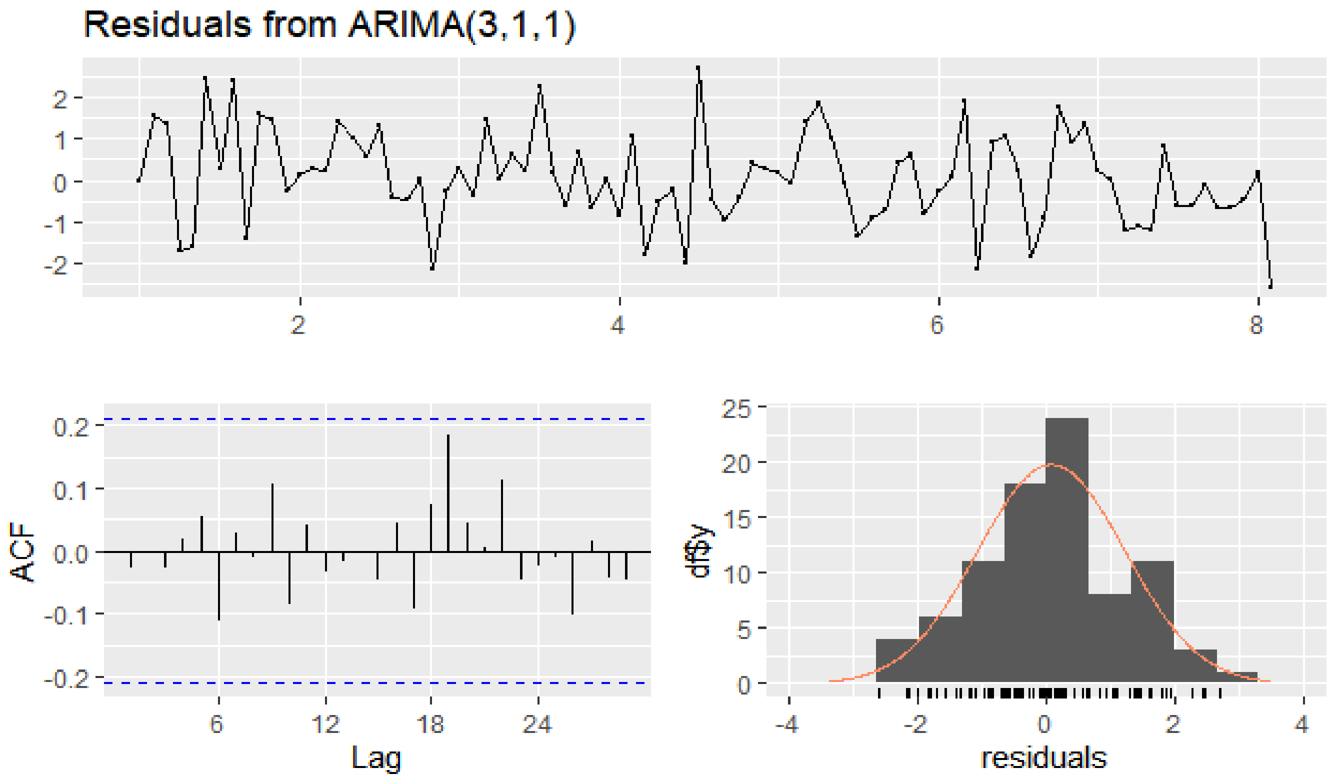
Best model: ARIMA(3,1,1)

Parameter Estimation

Using maximum likelihood estimation, the **ARIMA (3,1,1) model** with coefficients is given as:

$$У_t = 1.0106У_{t-1} + 0.3013У_{t-2} - 0.4906\, У_{t-3} + \varepsilon_t - 0.6849\varepsilon_{t-1}$$

Model Validation



The series of the residuals is stationary i.e. the statistical properties of the series are constant. The ACF plot shows there is no autocorrelation between the residuals. The histogram shows that the residuals are approximately normally distributed.

Model Validation

Ljung-Box Test for model adequacy

H_0: The model is a good fit

Vs

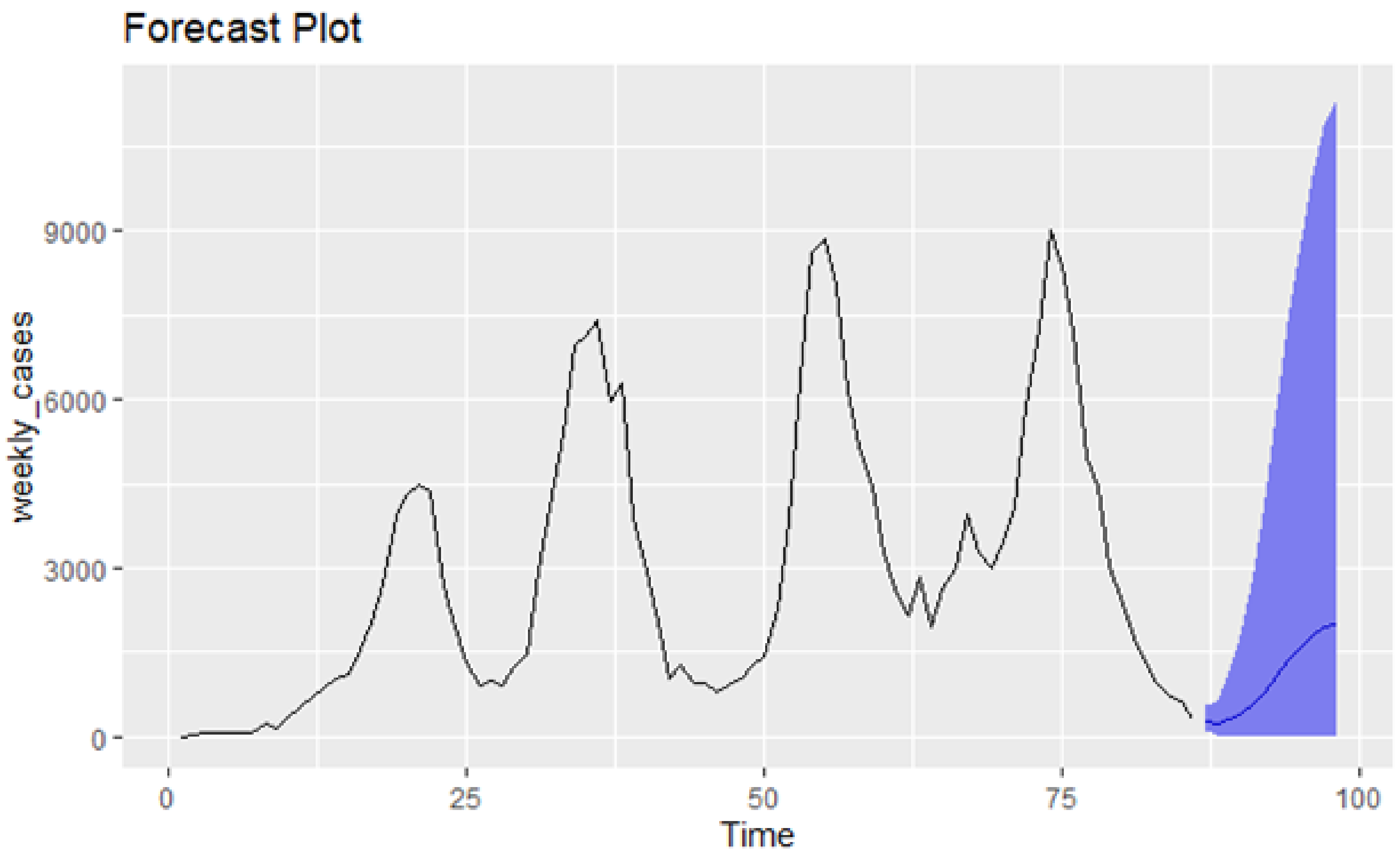
H_1: The model is not a good fit.

The Ljung-Box test for model adequacy is a good fit since the **p-value (0.967)** is greater than 5% level of significance hence fail to reject the null hypothesis.

Forecasts

Week	Point Forecast	Lo 95	Hi 95
87	271.9873	96.5857	542.3723
88	221.3478	22.5350	650.6111
89	297.0643	3.8604	1163.3529
90	389.2461	0.1509	1813.8225
91	574.7888	0.0000	2921.5264
92	791.2699	-0.0898	4227.9445
93	1065.8363	-0.3670	5812.2421
94	1335.7335	-0.7447	7364.7352
95	1593.986	-1.0573	8826.9551
96	1794.6639	-1.4610	9992.0684
97	1935.0146	-2.0410	10852.2218
98	2004.6242	-3.0865	11373.8692

Forecasts



Forecast Plot
9000
6000
3000
0
weekly_cases
0
25
50
75
100
Time

Conclusion and Recommendation

Conclusion

ARIMA model (3, 1, 1) resulted as the best model due to;

1. Low AICc
2. Low BIC
3. Low AIC
4. Goodness of fit tests also support the model

Recommendations

1. The government should tighten policies to reduce transmissions which are expected to peak from second week of December.
2. Continued vaccination of people along side other Covid-19 measures.

References

Alzahrani, S., Aljamaan, I., & Al-Fakih, E. (2020). Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using the ARIMA prediction model under current public health interventions. *Journal Of Infection And Public Health*, *13*(7), 914-919. https://doi.org/10.1016/j.jiph.2020.06.001

Chaurasia, V., & Pal, S. (2020). Application of machine learning time series analysis for prediction COVID-19 pandemic. *Research On Biomedical Engineering*. https://doi.org/10.1007/s42600-020-00105-4

Khan, F., & Gupta, R. (2020). ARIMA and NAR-based prediction model for time series analysis of COVID-19 cases in India. *Journal Of Safety Science And Resilience*, *1*(1), 12-18. https://doi.org/10.1016/j.jnlssr.2020.06.007

Sam, S. O., Pokhariyal, G. P., Rogo, K., & Ndhine, E. O. (2021). Otoi-NARIMA model for forecast seasonality of COVID-19 waves: Case of Kenya. International Journal of Statistics and Applied Mathematics 2021; 6(2): 48-58. https://doi.org/10.22271/maths.2021.v6.i2a.675

Wang, Y., Shen, Z., & Jiang, Y. (2018). Comparison of ARIMA and GM (1,1) models for prediction of hepatitis B in China. *PLOS ONE*, *13*(9), e0201987. https://doi.org/10.1371/journal.pone.0201987

THANK YOU-AHSANTE

Stay Safe