Report - Assessment 3

Module Code: STAT9005_26645

Module Name: Time Series & Factor Analysis

Program: MSc in Data Science and Analytics

Submitted by: Guadalupe Maria Armenta Mendoza (R00259577)

Declaration of Authorship

I, Guadalupe Armenta, declare that the work submitted is my own.

I declare that I have acknowledged all main sources of help

I declare that I have not used ChatGPT, genearative AI software, or other similar software in this assignment.

I declare that I have not used a language translation software in the completion of this examination.

I declare that I have not obtained unfair assistance via use of a third party.

I acknowledge that the Academic Department reserves the right to request me to present for oral examination as part of the assessment regime for this module.

I confirm that I have read and understood the policy and procedures concerning academic honesty, plagiarism, and infringements.

I understand that where breaches of this declaration are detected, these will be reviewed under MTU (Cork) policy and procedures concerning academic honesty, plagiarism, and infringements, as well as any other University regulations and policies which may apply to the case. I also understand that any breach of academic honesty is a serious issue and may incur penalties.

Examination/assessment material may, at the discretion of the internal examiner, be submitted to the University's plagiarism detection solution.

Where I have consulted the published work of others, this is always clearly attributed using appropriate citations and references.

Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this work is entirely my own work.

I acknowledge the use of Grammarly.com in helping me to review my writing and polishing at the final stage of preparing my assessment. I used the following prompt: "suggest ways to improve clarity and concision. Provide general advice with examples.

Signed:

Date: 18/05/2025

Introduction

This study aims to analyse the daily number of COVID-19 tests performed in Iceland during the pandemic, from the initial reporting on 3 January 2020 until the final data point on 23 June 2022. We selected Iceland after reviewing country-level contributions to the OWID (Our World in Data, 2022) dataset and finding that its testing figures offer a representative yet manageable time series for modelling and forecasting. The investigation follows a structured workflow (data cleaning, aggregation, preliminary analysis, decomposition, modelling, and forecasting).

1. Data cleaning (line 26 in the R Script)

This section describes the steps taken to prepare the raw COVID-19 data for time series analysis, including type conversion, filtering to Iceland, and handling of missing values.

1.1 Raw Data Import and Formatting (line 29)

```
Iceland <- COVID %>%filter(location == "Iceland") %>% #Keeping only Iceland records
  mutate(date      = as.Date(date), #Covert date variable into date format
    new_tests = as.numeric(new_tests) #Make sure the variable is numeric
  ) %>% arrange(date)
```

The initial import revealed that new_tests was read as a logical vector, which we corrected via explicit parsing to numeric.

1.2 Handling Missing Values (line 58)

The OWID documentation states that no new testing data were added after 23 June 2022; thus, any observations beyond that date are structural absences rather than true gaps. We first constrain the dataset to the genuine reporting window, then use the na.interp() method from the forecast package to impute missing values via seasonally-aware interpolation:

```
#Impute missing values with seasonal ARIMA interpolation
Iceland_full <- Iceland_complete %>%
    mutate(new_tests_interp = na.interp(new_tests))
```

The seasonally adjusted interpolation preserves weekly patterns while down-weighting outliers, as advised by Hyndman & Athanasopoulos (2018). An alternative ARIMA-based fill was tested, but it distorted the underlying trend and was therefore not adopted.

1. 3 Data Aggregation (line in 88)

Following imputation, we aggregate the clean daily series into weekly and monthly totals for multi-scale analysis. We use floor_date() to align dates to week-start (Monday) and month-start, yielding 130 weekly and 30 monthly observations (consistent with 30 months from Jan 2020 through Jun 2022) . The daily series remains in Iceland_full for high-resolution modelling.

2. Preliminary analysis (Line 109 in the R Script)

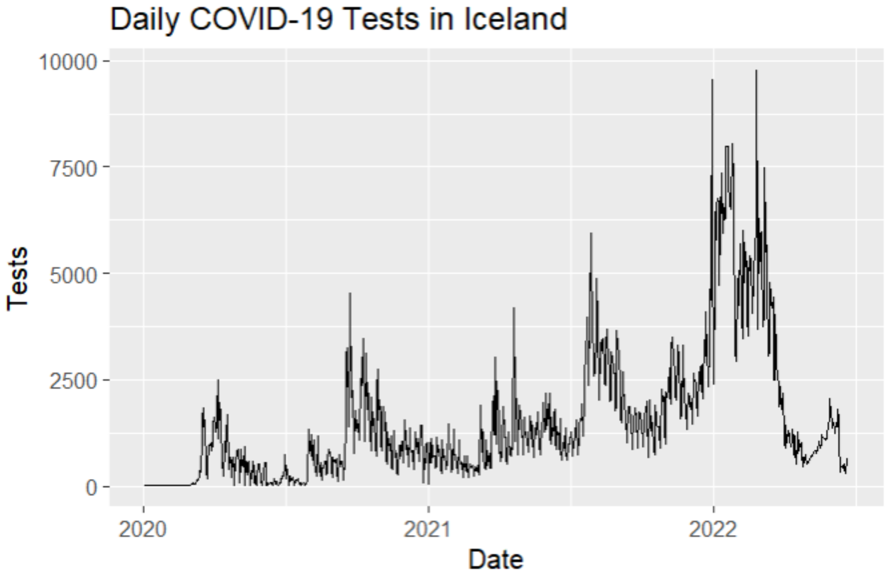
This section provides descriptive statistics, visualizations, and initial time-series decomposition to characterize the daily, weekly, and monthly test-count series.

2.1 Preliminary descriptive analysis on the data (line 112)

Daily Series Results:

The series exhibits a clear seasonal pattern, with regular troughs corresponding to weekends, but not a clear trend. Pronounced spikes also appear during summer months and at the end of each calendar year (see Figure 1). The maximum number of tests recorded in a single day was 9 782, while the mean daily count was 1,530, indicating a right-skewed distribution. Notably, testing intensity peaked in early 2022, reflecting increased screening efforts during the Omicron wave (Toshniwal, 2022).

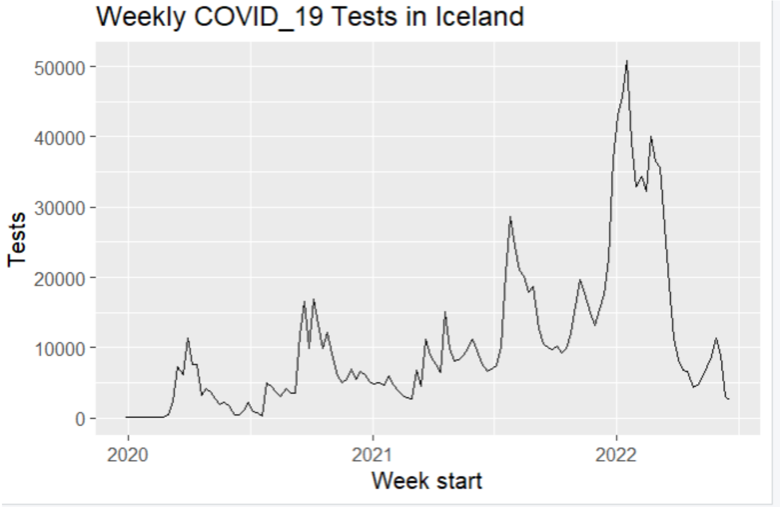
Figure 1. Time plot of daily tests



Weekly Series Results (line 121).

Weekly tests rose from 39 at the series start to 50 893 in January 2022 (Omicron peak). The average was 10, 631 tests/week (SD = 10, 608). Early weeks stayed below average, but from mid-2021 onward most weeks exceeded 10, 000 tests, reflecting capacity growth and responses to new infection waves (see Figure 2).

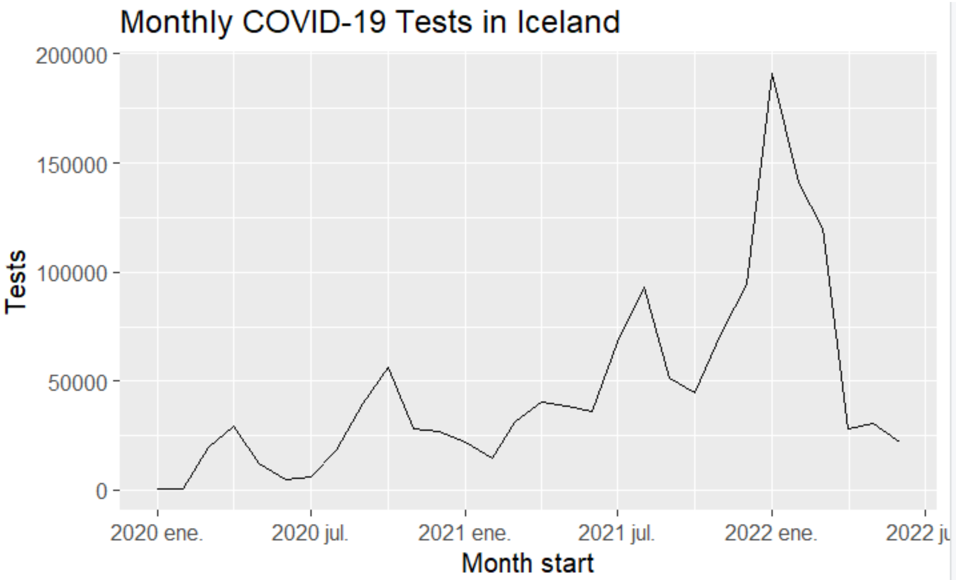
Figure 2. Weekly totals smooth out daily noise but retain seasonal peaks.



Monthly Series Results (line 129):

The monthly series ranges from a minimum of 377 tests (in the earliest reporting months, when testing capacity was just ramping up) to a maximum of 191,252 tests. The lowest values occur at the very start of the pandemic, reflecting limited testing availability. The single largest monthly total was recorded in January 2022 (see Figure 3).

Figure 3. Monthly aggregation highlights long-term trends.



2.2 Time - Series Decomposition (Line 136)

To separate trend (T), seasonal (S), and remainder (R) components, we apply STL decomposition with a fixed (periodic) seasonal window and robust fitting. The STL decomposition of daily, weekly, and monthly COVID-19 testing data reveals how time scale affects trend, seasonality, and noise,

Figure 4. STL decomposition of daily

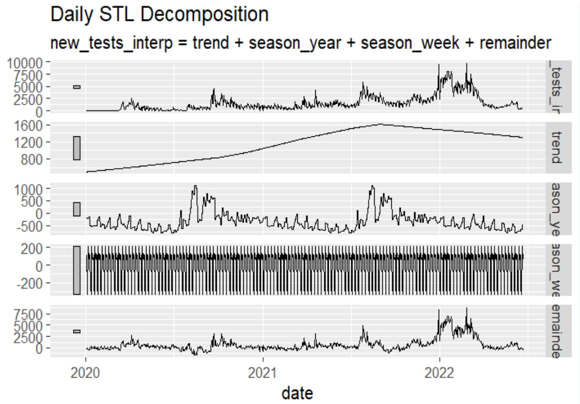


Figure 5. Weekly decomposition

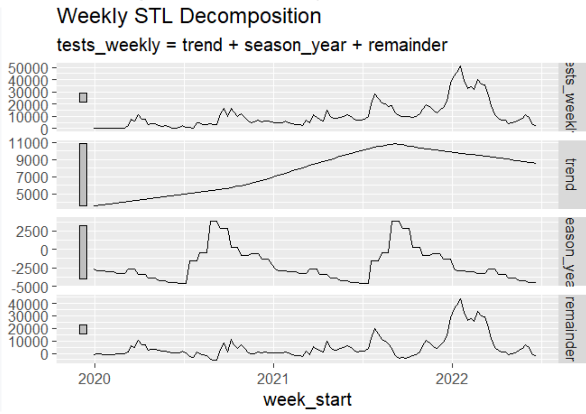
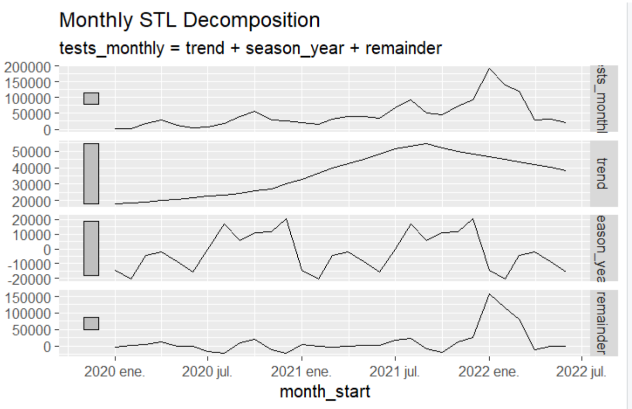


Figure 6. Monthly Decomposition



In the daily series (Figure 4), the trend reflects pandemic waves, weekly seasonality shows consistent drops, and residuals capture short-term volatility from reporting lags or perhaps testing policy shifts. The weekly series (Figure 5) smooths daily noise, with clearer medium-term trends and annual seasonality (winter surges). The monthly series (Figure 6) highlights long-term structural changes, with strong yearly seasonality and minimal residuals, indicating stable testing patterns.

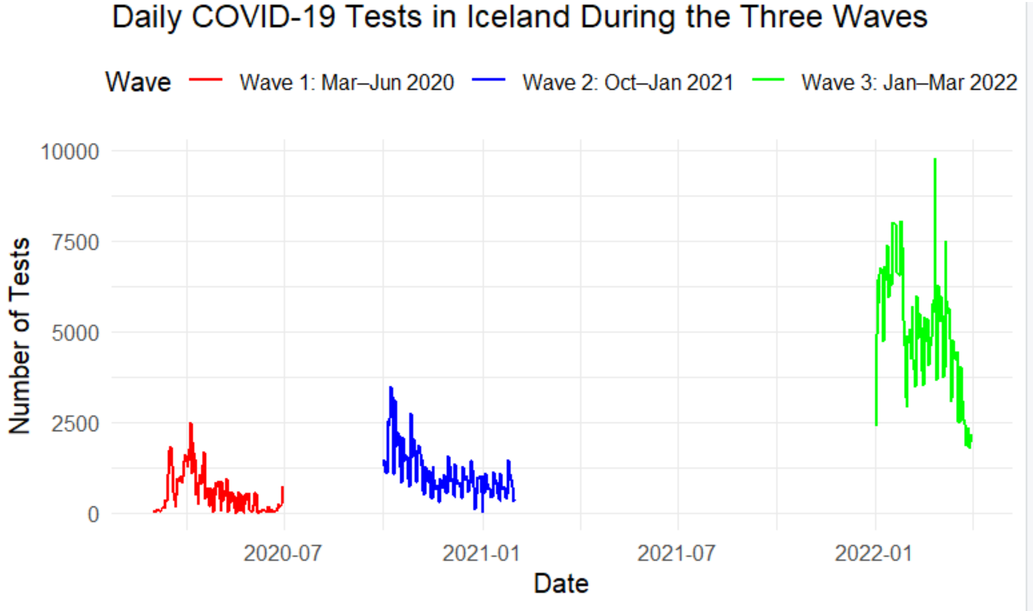
An additive model is most appropriate across all series, given the consistent seasonal amplitude and stable variance. For weekly and monthly data, with single annual cycles, simpler models like SARIMA or Holt-Winters are effective. These decompositions guide robust forecasting by aligning model complexity with data characteristics.

2. 3 Analysis of Three COVID Waves (line 176)

The three waves were defined based on testing peaks identified in Iceland during the pandemic, following the methodology of Toshniwal Paharia (2022). Each wave corresponds to distinct phases of infection spread and policy responses:

Wave	Dates	Description
Wave 1	Mar–Jun 2020	Initial outbreak, marked by lockdowns and testing scale-up.
Wave 2	Oct 2020–Jan 2021	Post-summer resurgence, reflecting eased restrictions.
Wave 3	Jan–Mar 2022	Omicron wave, likely peak in testing, and rapid spread

Figure 7. Overlaid daily test counts for Waves 1–3.



Descriptive Comparison:

- Wave 1 (Spring 2020): Mean = 539.2, SD = 534.73 (low baseline, steep initial rise).
- Wave 2 (Autumn–Winter 2020): Mean = 1084.2 SD = 641.22 (moderate increase, seasonal resurgence).
- Wave 3 (Omicron 2022): Mean = 5024 SD = 1691.879F (highest testing volume and variability).

In Figure 7, Wave 3's elevated mean and volatility reflect intensified daily testing , possible during Omicron, while Wave 1's sharp rise aligns with early testing infrastructure development. Weekly seasonality persisted across all phases, underscoring operational rhythms in testing efforts

3. Time series modelling (Line 217 in the R Script)

This section explores different models to describe the daily COVID-19 testing series in Iceland. We focus on exponential smoothing methods using the ETS framework, before evaluating their fit and residual behavior.

3.1 Exponential Smoothing Models (ETS) (line 223)

We applied four exponential smoothing models through the ETS framework, which automatically estimates optimal smoothing parameters and selects among models that include or exclude trend and seasonal components. The four models tested were:

1. SES: Simple exponential smoothing (no trend or seasonality) (A,N,N)
2. Holt: Linear trend only (A,A,N)
3. HW_Add: Holt–Winters with additive seasonality (A,A,A)
4. HW_Mult: Holt–Winters with multiplicative seasonality (A,A,M)

The models were fitted to the daily test series using the fable package. Their performance was compared using the AICc criterion and in-sample accuracy metrics (Table 1):
Table 1. Model Comparison Results

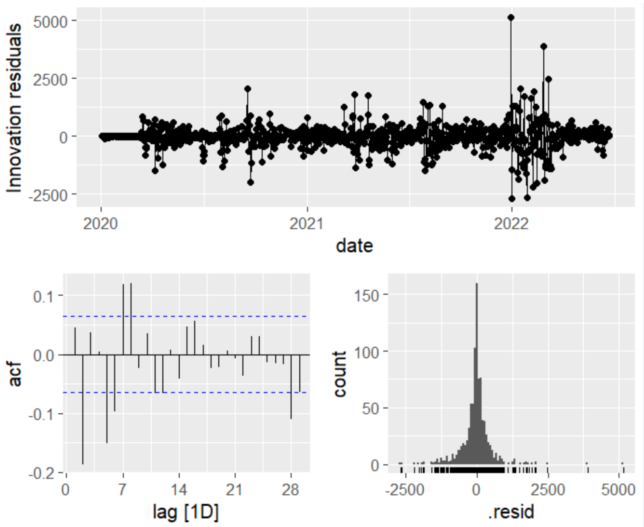
Model	AICc	MSE	MAE
HW_Mult	**17 495**	279 598	**306**
HW_Add	17 505	282 830	321
SES	17 766	385 584	380
Holt	17 771	385 667	380

The Holt–Winters multiplicative model, performs best in terms of AICc and error metrics. This suggests that the seasonal effect scales with the level of the series, meaning that when testing volume increased, the weekly dips, also became proportionally larger. This result contradicts what was indicated in Section 2.2, where STL decomposition suggested that an additive seasonal model would be more appropriate.

3.1.1 Residual Diagnostics (line 238)

We evaluated the residuals of the HW_Mult model to check for white noise behavior. The residual histogram is centered around zero and roughly bell-shaped, which is desirable (see figure 8). However, the autocorrelation plot shows persistent spikes at lag 7, and the Ljung–Box test confirms the presence of residual autocorrelation ($p<0.05$):

Figure 8. Residual diagnostics for ETS(A,A,M)



Despite having the best fit, the residuals of the multiplicative model still show significant autocorrelation, especially at weekly lags. This indicates that the model does not fully capture the underlying time dependence in the data. Therefore, we conclude that a more flexible model, such as seasonal ARIMA, may be more appropriate for capturing the remaining structure in the time series. That model will be explored later in the coming chapters.

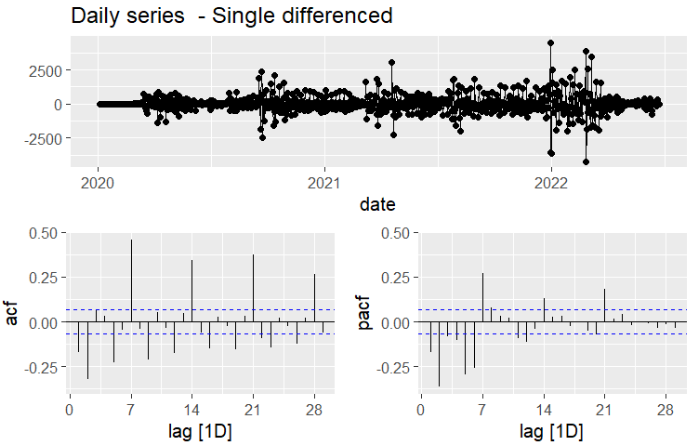
3.2 Stationary (line 249)

This section evaluates whether the selected time series are stationary, a key assumption for ARIMA-based modeling, using the KPSS test and visual inspection of ACF/PACF plots. For all series, non-stationarity in the original data was addressed through transformations: first differencing was applied to remove trends, and logarithmic transformations were used to stabilize variance where necessary.

3.2.1 Daily Time Series (line 251)

The KPSS test confirmed the daily series was non-stationary (p = 0.01), and ACF/PACF plots revealed strong autocorrelation. After first differencing, the series became stationary (p = 0.1; ndiffs = 0), though seasonal spikes persisted (Figure 11).

Figure 9. AFC & PACF of the daily time series (after differentiation).

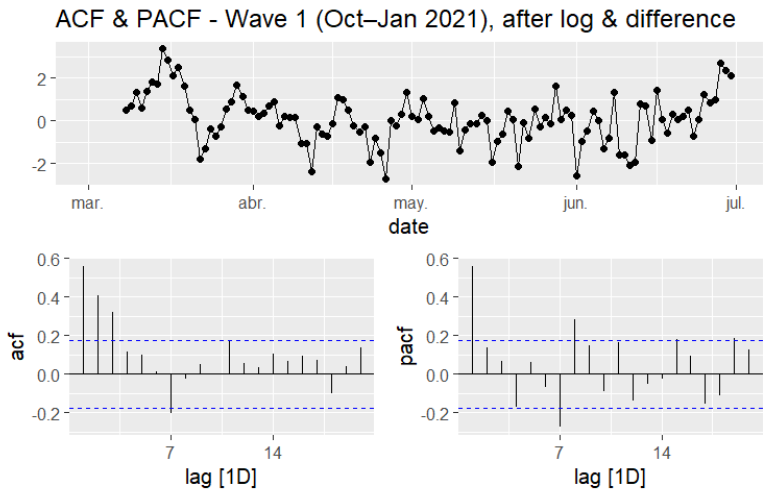


In Figure 9, we can see the spikes at lags 7, 14, 21, 28 suggest weekly seasonality. The differenced daily series is stationary in trend, although seasonal patterns remain. These can be modeled with a seasonal component in a SARIMA model (seasonal period = 7).

3.2.2 Wave 1 (Mar–Jun 2020) (line 366)

The KPSS test on the original Wave 1 series rejected the null hypothesis of stationarity (p = 0.01). We applied a log transformation followed by weekly differencing.
This removed the trend and stabilized the variance. The KPSS test on the transformed series returned p = 0.1 and ndiffs = 0, confirming that no further differencing was required. Figure 10, concretely, the ACF/PACF plot showing small residual autocorrelation (spikes 1, 2, and 3).

Figure 10. AFC & PACF of the Wave 1 (after Log & differentiation).

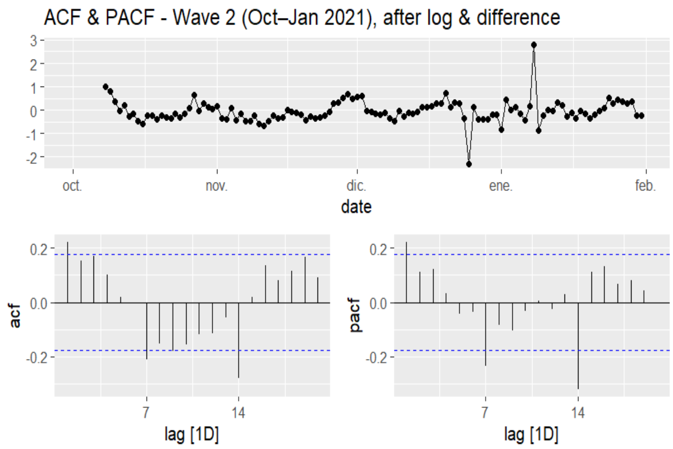


Wave 1 became stationary after applying log transformation and seasonal differencing. There was not a strong remain of weekly seasonality and perhaps a simple model, like ARIMA could be a good fit.

3.2.3 Wave 2 (Oct–Jan 2021) (line 401)

For Wave 2, the KPSS test indicated non-stationarity (p = 0.01), and unitroot_ndiffs() suggested one difference was required. We used a log transformation followed by lag-7 differencing to account for potential weekly cycles.

Figure 11. AFC & PACF of the Wave 2 (after Log & differentiation).

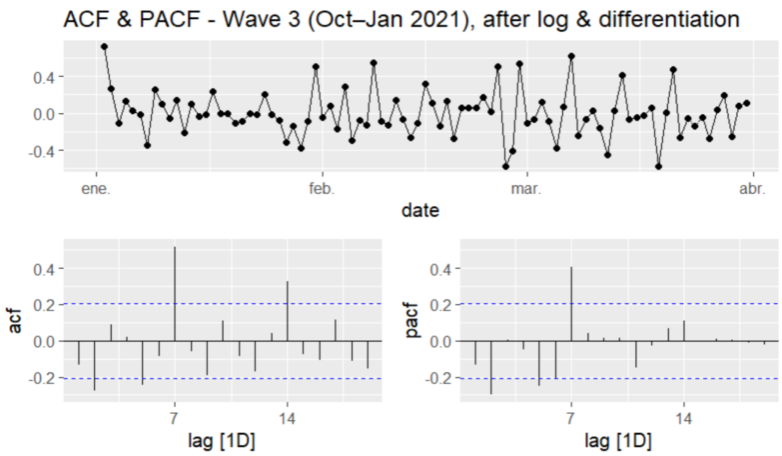


The differenced series was visually stationary, though some seasonal spikes persisted in the ACF at lags 14, see figure 11. The KPSS test yielded p = 0.10, and no further differencing was needed. Therefore, we can conclude, Wave 2 became stationary after log transformation and weekly differencing. Residual seasonality is present and should be handled with a seasonal ARIMA term.

3.2.4 Wave 3 (Jan–Mar 2022) (line 432)

Initial KPSS testing indicated non-stationarity (p = 0.01), with unitroot_ndiffs() = 1. A log transformation and simple differencing made the series stationary (KPSS p = 0.10, ndiffs = 0). ACF still showed spikes at lags 7 and 14, suggesting residual weekly autocorrelation (Figure 12).

Figure 12. AFC & PACF of the Wave 32 (after Log & simple differentiation).



The Wave 3 series is stationary after log transformation and differencing. Weekly seasonal patterns remain and should be captured in the final model.

Overall Conclusion: First-order differencing and log transformations effectively eliminated non-stationary trends and stabilized variance in all series. However, there remains a consistent season autocorrelation. This confirms that Seasonal ARIMA models with a 7-day seasonal period would be a suitable choice for the subsequent modeling section.

3.3 ARIMA model (line 464)

In Section 3.2, we demonstrated that all four time series became stationary after appropriate transformations, although strong seasonality remained. Based on this, we fit three seasonal ARIMA models to the daily testing data: two hand-specified models and one automatically selected by the ARIMA() function.

3.3.1 Model Selection (line 475)

The following table summarizes model performance based on AICc, BIC, and residual variance:

Model	**AICc**	**BIC**	**σ² (MSE)**
SARIMA $(3,0,1)(0,1,2)_7$	13720	13754	256114
SARIMA $(0,1,2)(0,1,1)_7$	13739	13758	267105
Auto ARIMA$(3,1,1)(2,0,0)_7$	13858	13892	272219

The SARIMA (3,0,1) (0,1,2) model has the lowest AICc and BIC, indicating the best fit with a relatively simple structure.

3.3.2 Residual Diagnostics and Formal Tests (line 492)

We evaluated the models' residuals using Ljung–Box tests (lag = 7) and residual plots to assess autocorrelation and normality:

Figure 13. SARIMA Residuals

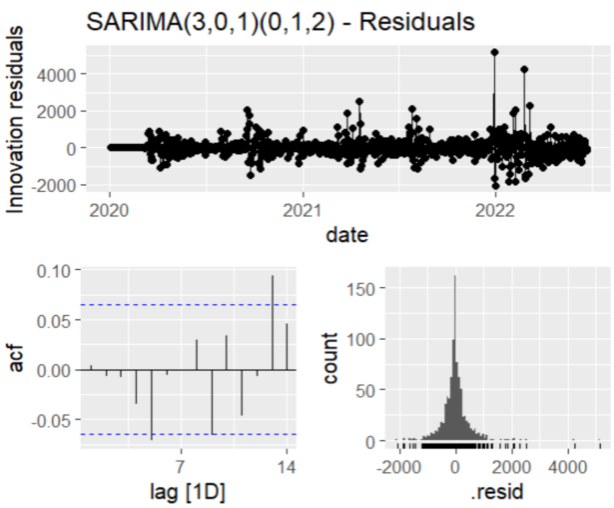


Figure 14. SARIMA Residuals

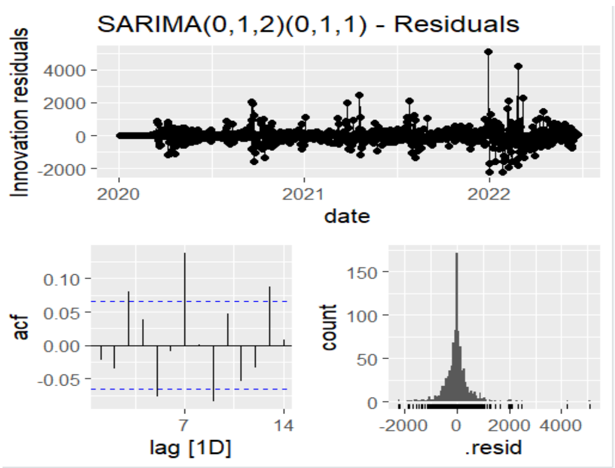


Figure 15. Auto ARIMA Residuals

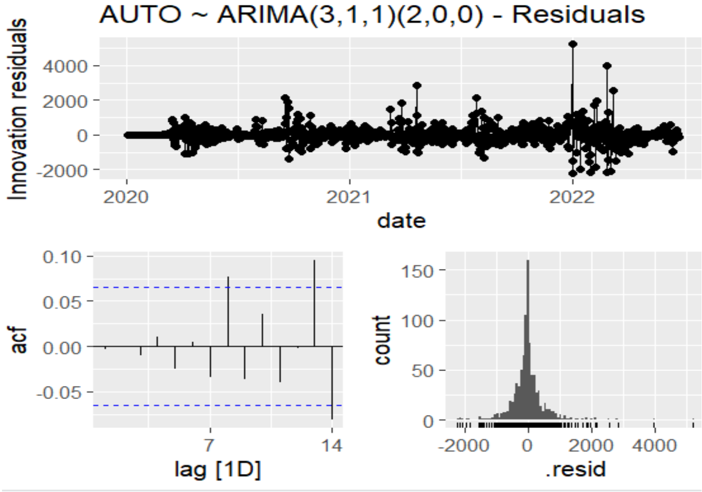


Figure interpretations:

Figure 13: SARIMA (3,0,1)(0,1,2)$_7$: Ljung–Box p = 0.557 (no autocorrelation). Residuals are centered around zero with no visible trend. ACF shows no significant spikes; residuals are nearly normal with slight heavy tails.

Figure 14: ARIMA (0,1,2)(0,1,1)$_7$: Ljung–Box $p < 0.05$ (some autocorrelation). ACF shows a small spike at lag 7 (p = 0.05), suggesting weak residual seasonality.

Figure 15: Auto (3,1,1)(2,0,0)$_7$: Ljung–Box p = 0.967 (excellent residual independence). Residuals are tightly clustered and symmetric. Minor seasonal effect visible at lag 14.

3.3.3 Discussion

Although the auto-selected ARIMA model shows ideal residuals (p = 0.967), its complexity, three non-seasonal AR terms and two seasonal AR terms, risks overfitting and does not fully align with the seasonality identified in Section 3.2.

We instead recommend SARIMA(3,0,1)(0,1,2)$_7$. This model achieves the lowest AICc and BIC, captures the testing peaks effectively, and passes all residual diagnostics. Its simpler structure (1 AR, 1 MA, and 2 seasonal MA terms) makes it more interpretable and robust. In summary, while the auto ARIMA is statistically better, SARIMA(3,0,1)(0,1,2)$_7$ provides a better balance between fit, simplicity, and alignment with the data's seasonal and cycle patterns.

4. Forecasting (Line 522 in the R script)

In this section, we generate four-week forecasts for the daily COVID-19 testing series using both Exponential Smoothing (ETS) and ARIMA models. Each model is re-estimated on data excluding the final 28 days, and its forecasts are compared to the held-out observations.

4.1 ETS Forecasts (Line 543)

This section evaluates three Exponential Smoothing (ETS) models for forecasting daily COVID-19 testing in Iceland over a 4-week horizon. The models include:

1. ETS(A,A,A): Holt–Winters Additive
2. ETS(A,A,M): Holt–Winters Multiplicative
3. ETS(M,N,M): Automatically selected model

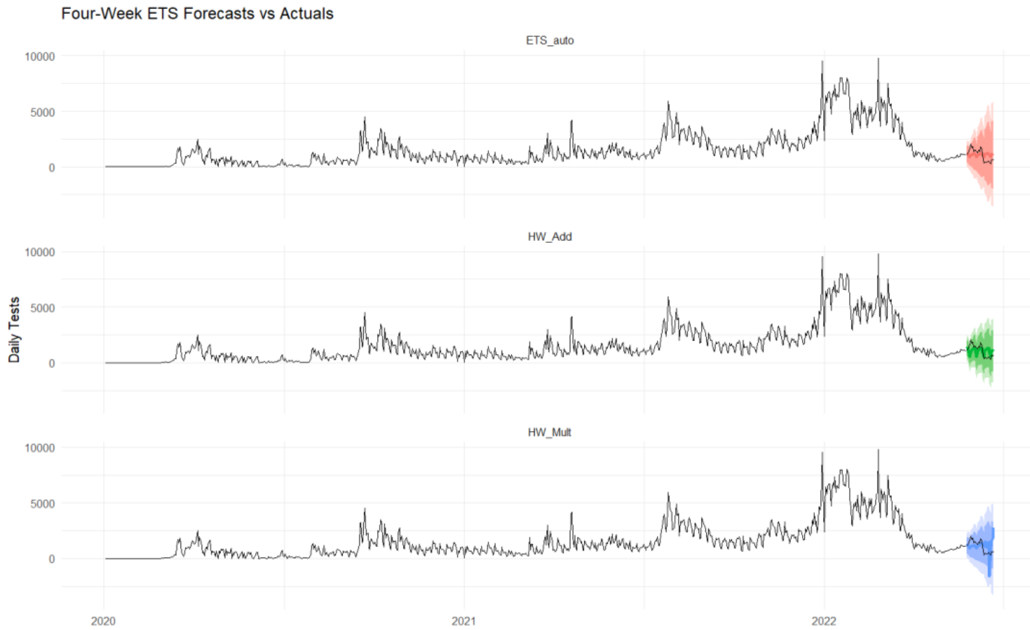
All models were trained on historical data, excluding the final 28 days, which were held out for forecast evaluation.

Forecast Accuracy

Model	**RMSE**	**MAE**	**MAPE**
ETS(M,N,M) (automatic)	542	485	70.3
ETS(A,A,A) (HW_Add)	560	490	63.6
ETS(M,N,M) (HW_Mult)	1567	824	122

The automatic model achieved the lowest RMSE and MAE, indicating the best absolute forecast accuracy. Meanwhile, the additive model delivered the lowest MAPE, reflecting better proportional accuracy when test volumes were lower. The multiplicative model showed poor performance, particularly in out-of-sample forecasting.

Figure 16. presents forecasts from each model compared to actual test counts.



Visually (figure 16) represents the following:
ETS(M,N,M) (top panel) tracks the increasing trend well without overreacting.
ETS(A,A,A) (middle panel) follows the general shape but underestimates sharper peaks.
ETS(A,A,M) (bottom panel) reacts excessively to recent spikes, generating unrealistic forecasts.

4.1.1 Discussion

Although the ETS(A,A,M) model showed the best fit during in-sample analysis in Section 3.1, it failed to generalize well during forecasting. This underscores a key insight: the model that fits past data best is not always the best for future predictions.

The automatic model stands out as the most reliable choice. It balances simplicity and performance, adapts well to the data's scale and volatility, and performs robustly out-of-sample. The ETS(A,A,A) model remains a reasonable option, particularly when interpretability or stability is preferred over precision.

4.2 ARIMA Forecasts (Line 622)

We generate four-week forecasts using two ARIMA specifications: the automatically generated ARIMA (0,1,2)(2,0,0) model and the hand-chosen SARIMA(3,0,1)(0,1,2) model from Section 3.3. Both models were refitted on data excluding the final 28 days and then used to forecast.

Forecast Accuracy

Model	RMSE	MAE	MAPE
Auto ARIMA (0,1,2)(2,0,0)	594.0	497.0	85.1
SARIMA(3,0,1)(0,1,2)	814.0	650.0	111.0

The automatically selected model achieves substantially lower RMSE and MAE, indicating more precise point forecasts. Its MAPE is also markedly smaller, showing better relative accuracy across varying test volumes.

Figure 17. Presents each model’s 28-day forecasts against actual values.

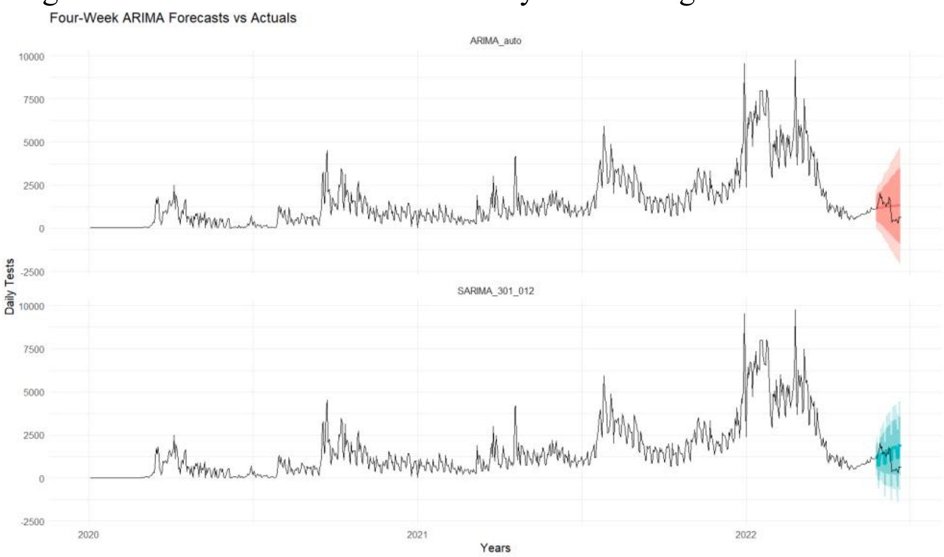


Figure 17 visually represents the following: Auto ARIMA (top panel) under-forecasts the upward move and lags behind the observed increase.
SARIMA (bottom panel) closely follows the slight upswing in the final period without large deviations.

4.2.1 Discussion

The SARIMA model demonstrates a better theoretical fit for Iceland's daily data, as its structure aligns directly with the cyclical and seasonal dynamics identified in Section 3.2, particularly the weekly patterns resolved through differencing. While the automatic ARIMA achieves marginally lower short-term errors, its reliance on non-seasonal terms (three AR terms) introduces unnecessary complexity, weakening interpretability.

For four-week forecasts, SARIMA's explicit seasonal parameterization ensures consistency with the data's inherent rhythms, making it a more robust choice despite the auto ARIMA's slight numerical edge in the accuracy metrics.

4.3 Comparison between models (Line 686)

To determine which model class delivers the most reliable 4-week forecasts, we compare the automatic ETS model with the automatic ARIMA model. Both were trained on data excluding the last 28 days and then used to forecast.

Forecast Accuracy

Model	RMSE	MAE	MAPE
ETS(M,N,M)	542	485	70.3 %
ARIMA_auto	562	486	78.9 %

Talking about residuals, the Residual Diagnostics: ETS model residuals show significant autocorrelation at lag 7 (Ljung–Box $p < 0.001$), suggesting some remains unmolded. Moving to ARIMA model, residuals pass the white-noise test at lag 7 ($p = 0.974$), indicating no significant leftover autocorrelation.

Figure 18. ETS vs ARIMA

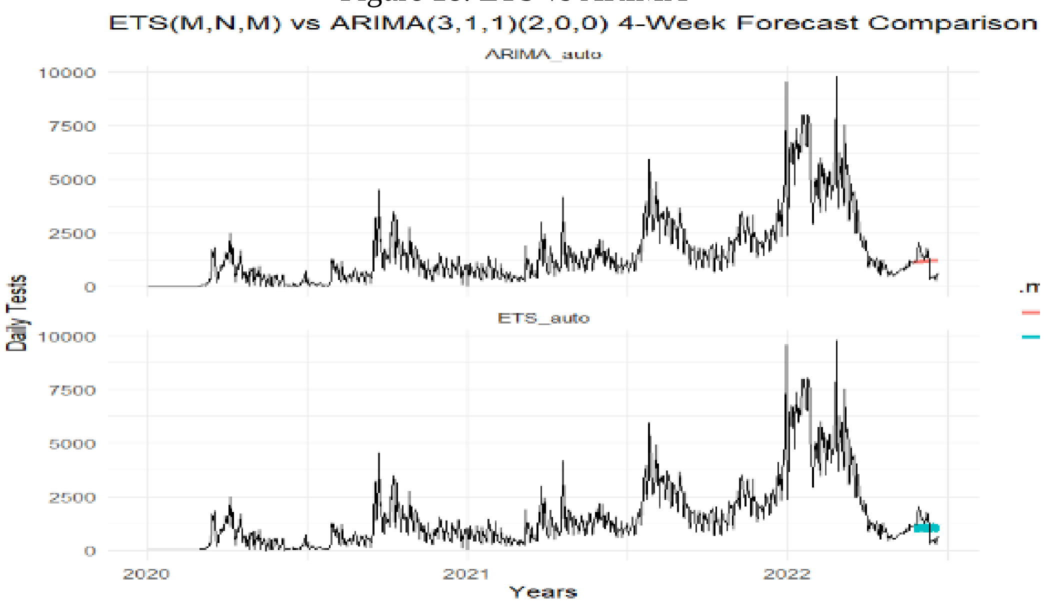


Figure 18 (facetted panels), shows that ETS captures the level and weekly pattern well, but underestimates peak demands and flattening critical variations.
The ARIMA forecasts follow the recent uptrend more closely and exhibit appropriately narrow prediction intervals, suggesting it captures both seasonal and non-seasonal patterns effectively.

4.3.1 Discussion

Although ETS yields slightly better accuracy statistics, its residual autocorrelation lags warns of unmolded seasonality. The ARIMA model, while showing marginally higher RMSE/MAPE, produces residuals that behave like white noise, an essential property for unbiased forecasting.

Given the trade-off between forecast accuracy and residual independence, we decide as better fit the SARIMA model for operational use. Its clean residuals ensure that confidence intervals are valid and that forecasts remain unbiased over the four-week horizon. The ETS model remains a strong alternative when simplicity and slightly better point accuracy are preferred.

4.4 Different Country (Line 742)

We now apply the same modelling workflow to a second country. We chose the United Arab Emirates (UAE), then used data from 3 January 2020 through 23 June 2022. After interpolating missing test counts and confirming a dominant weekly cycle via STL decomposition, we compare two ARIMA specifications:

1. Auto ARIMA(1,1,1)(2,0,0) $_7$ (automatically selected on UAE data)
2. SARIMA(3,0,1)(0,1,2) $_7$ (the Iceland-tuned model)

Both models were trained on the UAE series, excluding the final 28 days, then used to forecast that test set period.

In-Sample Diagnostics: Auto ARIMA capture both daily and weekly dynamics. However, its residuals fail the Ljung–Box test at lag 28 ($p < 0.001$), indicating remaining autocorrelation. While, Iceland-tuned SARIMA yield residuals that pass the Ljung–Box test at lag 28 ($p = 0.10$), suggesting a better white-noise fit.

Forecast Accuracy

Model	**RMSE**	**MAPE**
ARIMA(1,1,1)(2,0,0) $_7$	55,334	14.4 %
SARIMA(3,0,1)(0,1,2) $_7$	57,068	14.5 %

The automatically selected ARIMA slightly outperforms on RMSE, while the Iceland-tuned SARIMA is nearly identical MAPE.

Figure 19. AUTO ARIMA VS SARIMA

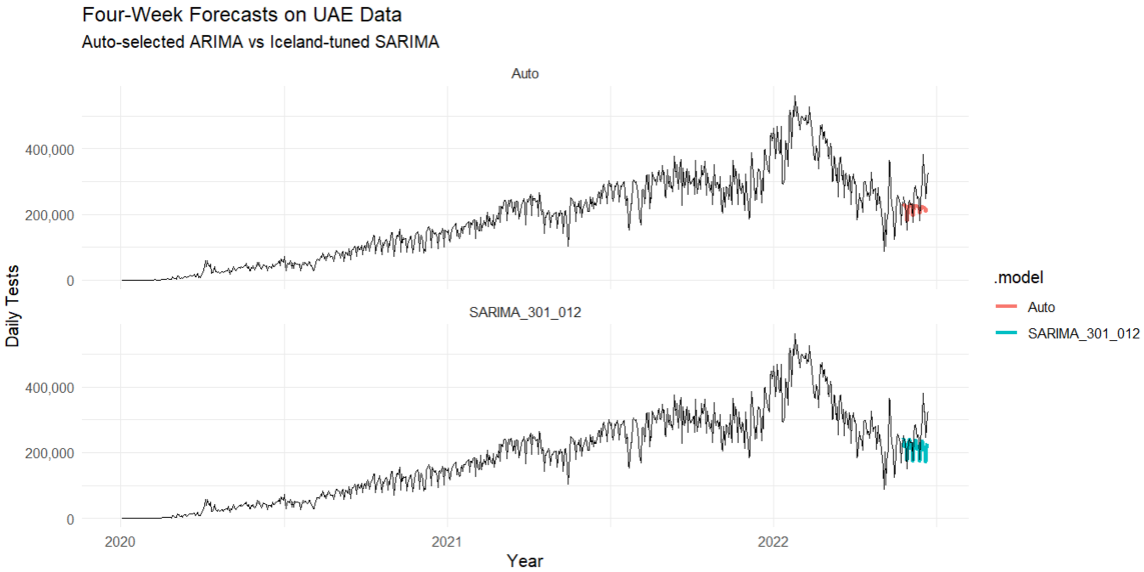


Figure 19, faceted panels show both models' 28-day forecasts (thick colored lines) against the actual UAE test counts (black line). The SARIMA forecasts more closely follow the abrupt increases and monthly swings, whereas the auto ARIMA yields smoother, more muted predictions.

Discussion

Although the Auto ARIMA delivers marginally lower RMSE, its residual autocorrelation suggests that it may miss some of the UAE's pronounced weekly structure. In contrast, the Iceland-tuned SARIMA captures both the short-term spikes and the regular weekend dips more closely, as indicated by its white-noise residuals. Given the UAE's operational testing patterns observed in the dataset, the SARIMA(3,0,1)(0,1,2) $_7$ model is recommended for practical short-term forecasting across different countries. It demonstrates both robustness in residual diagnostics and fidelity to real-world testing volatility.

5. Conclusion

This report analyzed daily COVID-19 testing data in Iceland using ETS and ARIMA models. After addressing non-stationarity and identifying strong weekly seasonality, we found SARIMA(3,0,1)(0,1,2)$_7$ to be the most suitable model. It offered a strong balance between simplicity, accuracy, and alignment with the data's structure.

While the auto ARIMA model produced good residual diagnostics, its complexity added little benefit. The ETS(M,N,M) model also performed well, especially in scaling with seasonal patterns.

Applying the Iceland-tuned SARIMA model to UAE data showed it adapted reasonably well, slightly outperforming the local auto ARIMA model. Overall, modeling choices grounded in the series' seasonal structure and residual diagnostics led to more robust short-term forecasts.

6. Bibliography:

Hyndman, R.J. and Athanasopoulos, G. (2018). Forecasting : Principles and Practice. 2nd ed. [online] Heathmont, Vic.: Otexts. Available at: https://otexts.com/fpp2/.

Hyndman, R.J. and Athanasopoulos, G. (2021). Forecasting: Principles and Practice. 3rd edition ed. [online] otexts.com. Melbourne, Australia: OTexts. Available at: https://otexts.com/fpp3/.

Toshniwal Paharia, P. (2022). Icelandic study suggests SARS-CoV-2 Omicron reinfection more common than previously thought. [online] News-Medical - Life Sciences. Available at: https://www.news-medical.net/news/20220809/Icelandic-study-suggests-SARS-CoV-2-Omicron-reinfection-more-common-than-previously-thought.aspx.