Available online at
http://journale-academiauitmt.uitm.edu.my

e-Academia Journal

e-Academia Journal of UiTM Cawangan Terengganu 14(1) 73-86, June 2025

Forecasting Tourist Arrivals in Langkawi Island: A Time Series Analysis

Nur Hawa Hana Humaira Mohd Shukry [1], *Isnewati Ab Malek [2] & Haslinda Ab Malek [3]

[1, 2, 3]*Faculty of Computer and Mathematical Sciences,*
Universiti Teknologi MARA Negeri Sembilan Branch,
Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

*Corresponding author's email: isnewati@uitm.edu.my

Submission date: 7 March 2025 Accepted date: 27 May 2025 Published date: 29 June 2025

To cite this article: Shukry, N. H. H. H. M., Malek, I. A., Malek, H. A. (2025). Forecasting Tourist Arrivals in Langkawi Island: A Time Series Analysis. e-Academia Journal of UiTM Cawangan Terengganu, 14 (1) 73-86, June 2025

ABSTRACT

Langkawi, a renowned island in Malaysia, is one of the country's main tourist attractions celebrated for its natural beauty and unique features. The Covid-19 pandemic significantly affected global travel, causing major disruptions and a sharp decline in tourist arrivals. To support economic recovery and strategic planning, the state government of Kedah has launched the *Visiting Kedah 2025* programme to encourage tourism, especially to Langkawi. Given this context, it is reasonable to forecast an increase in tourist arrivals to Langkawi by 2025. Hence, this study aims to analyse the tourist arrival pattern, identify the best forecasting model, and provide monthly forecast value for tourist arrivals in Langkawi for the next 13 months. Therefore, the number of tourist arrivals is modelled and forecasted using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to have a better understanding towards the trend pattern. This study used the data set of the number of tourist arrivals for both domestic and international arrivals from January 2010 until June 2024, which was obtained from the official website of the Langkawi Development Authority (LADA). The best model was chosen based on the lowest value of Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). The SARIMA $(0,1,1)\,(0,1,1)_{12}$ model was chosen to be the most suitable for forecasting the tourist arrivals in Langkawi from July 2024 to July 2025. Based on the findings, the forecast value for the number of tourist arrivals is projected to rise from 239067.7 in July 2024 to 246778.0 in July 2025.

Keywords: *Forecasting, Tourism, Langkawi, ARIMA, Box-Jenkins*

1.0 INTRODUCTION

Tourism can best be described as the use of commercial services for leisure, pleasure, and relaxation while spending time away from home, often to explore unfamiliar places (UNWTO, 2025). In simple terms, it refers to people's short-term movement from their homes to temporary destinations for various activities. These experiences are generally believed to enhance personal growth and expand one’s understanding of different cultures. According to the United Nations World Tourism Organization (UNWTO, 2025), international tourism includes any activity where individuals travel outside their home region and stay for more than 24 hours. The impact of tourism activities is evident across various sectors of a national economy, as they involve both the consumption and purchase of products and services by travellers.

Located approximately 30 km off the northwest coast of Malaysia, Langkawi, in the state of Kedah, is an archipelago of 99 duty-free islands. It is one of the most popular islands in Malaysia for tourists to explore its beauty that is full of nature. According to the Domestic Tourism Survey, Langkawi Island was selected as the top travel destination for both domestic tourists and international tourists visiting Kedah in 2020 respectively. Parallel to this progress, Langkawi Island is actively marketing itself to become a globally recognized tourist destination. Moreover, Langkawi Island is well-known for being a prominent venue for well-known sporting events such as the Langkawi International Dialogues (LID), Le Tour de Langkawi, Langkawi International Regatta and Langkawi International Maritime and Aerospace (LIMA) (Shariff & Abidin, 2020). A popular tourist destination in Langkawi is the geopark that was designated as the United Nations Educational, Scientific and Cultural Organisation (UNESCO) Geopark, making it the first geopark in Malaysia and Southeast Asia. It generates revenue for the local community while also aiding in the preservation of the region’s cultural legacy. Because of its about 500-million-year-old geological history, UNESCO awarded it as a Global Geopark in 2007 (Harun et al., 2022). In conjunction, the consistent increase in international tourism exposes nations and communities to effects that extend beyond the economy to include sociocultural and environmental factors (Al-sakkaf et al., 2022). Hence, the reasons why the business owners and operations work as both guarantors of the economy’s stability, particularly on the island to boost the local communities’ economic development. Thus, the growth of tourism has benefited the local community.

Tourism is one of Malaysia’s largest economic contributors, and Langkawi remains one of the top destinations for both local and international visitors. Effective planning and resource allocation are essential, and forecasting tourist arrivals allows businesses and governments to adapt to demand fluctuations (Zheng & Zhang, 2023). However, unpredictable seasonal factors such as weather, economic shifts, and school holidays can affect the accuracy of these forecasts. Although univariate time series models are often used for forecasting, not all approaches are sufficiently robust in capturing both trend and seasonal components (Alioglu, 2022). The Box-Jenkins methodology, known for its wide application in time series forecasting, was adopted in this study. Therefore, this study findings are based on three objectives which are to study the tourist arrival pattern in Langkawi, to identify the best model in forecasting the number of tourist arrivals in Langkawi and to derive monthly forecast value regarding the number of tourist arrivals in Langkawi for the next 13 months.

2.0 LITERATURE REVIEW

2.1 Tourism Industry in Langkawi

According to Irwana Omar et al. (2014), Langkawi Island has undergone significant transformation over the years—from a quiet tropical island home to farmers and fishermen to a major tourist destination and the island’s main source of income. It is reasonable to conclude that Langkawi has developed a strong and sustainable tourism economy. Apart from that, Shafikullah and Nayan (2021) has studied the impact of tourism activities on the economy of the community in the coastal area of Langkawi Island, Malaysia and based on the study, it is stated that the development of Langkawi Island has progressed extreme quickly since it was designated as a duty-free port on January 1, 1987 and recognized by UNESCO as the first Geopark in Malaysia and Southeast Asia on June 1, 2007. Due to the recognition also, Langkawi has gradually attracted more tourists, scientist and environmental enthusiasts (Shariff & Abidin, 2020).

According to the Langkawi Development Authority (LADA), in 2023, the total number of tourist arrivals for both domestic and international tourists are 2,587,195 and 223,127 respectively. Majority of the tourist arrivals in Langkawi are contributed from domestic tourists and in 2023, it shows that around 92.1% which more than half of the total number of tourist arrivals are coming from the local people. Hence, it can be concluded that Langkawi has been chosen as a tourist main place to visit and travel as it shows a big number of tourist arrivals from both domestic and international arrival. Mrs. Siti Hajar Yunus, the Deputy Director of Tourism Malaysia Northern Region (TMWU) has stated that Penang and Langkawi continue to be well-liked destinations for Malaysian travelers.

2.2 Impact of Covid-19 on Tourism in Langkawi

The Covid-19 pandemic's widespread presence has had a major impact on how people perceive and handle travel danger. Not only that, but it has also impacted the paths and avenues of distribution that customers take. The global tourism sector has been impacted by the Covid-19 pandemic (Yuzaidi et al., 2022). Furthermore, a worldwide alert regarding the outbreak was released, recommending that elderly individuals and those with long-term health issues delay any unnecessary travel (World Health organization et al., 2020). Hence, 96% of the world's population was affected by an international travel restriction because of travel bans (Gössling et al., 2020). The Covid-19 global epidemic is negatively affecting the travel and tourist industry as well as the economics of the nations that significantly depend on it, claim Fernández et al. (2022). As shown in the statistics of monthly tourist arrivals in Langkawi (LADA), the number of tourist arrivals in Langkawi can be seen to have a significant drop starting from March 2020 which define for domestic to be from 196,580 in February 2020 to 135,449 in March 2020 and only 6372 in April 2020. Moreover, for international tourist arrivals, it is shown to have 58,852 in February 2020 and it began to drop to 17,651 in March 2020. Due to the lockdown in March 2020, there are only 47 of tourist arrivals in April 2020 and national borders continue to be closed immediately for international travel (Altahir et al., 2020).

After a while, Covid-19 pandemic can be well controlled with the latest innovations to prevent the pandemic from spreading faster from one to another and Malaysia has not taken a long time to take that proactive effort to protect the national health system (Tang, 2020). Each day goes by, the number of Covid-19 cases begin to show a downward trend, and most things slowly seem to be well-organized and structured with the proper implementation by the government. After the Malaysian government reopened it's border, Langkawi was become as a pilot destination for the country's tourism bubble initiative. There are around 60,504 visitors between September 16 and October 6, 2021, generating RM24.9 million in tourism revenue (The Star, 2021). This positive trend indicates a strong resurgence in domestic travel interest despite ongoing recovery efforts from the COVID-19 pandemic. The chief executive officer of Langkawi Development Authority (LADA), Mr. Nasaruddin Abdul Muttalib stated that up to September 30, the travel bubble's arrivals of tourists brought in RM15.97 million for Langkawi's tourism sector. This trend is enough to be said that people still have the interest to travel to Langkawi even after the pandemic of Covid-19 and this number of tourist arrivals will keep showing a trend in the future (Astro Awani, 2021). Tourism is a significant contributor to Langkawi's economy. Forecasting helps stakeholders anticipate revenue, employment needs, and business opportunities, thereby supporting local livelihoods and guiding investment decisions (Tang, 2020).

2.3 Forecasting The Number of Tourist Arrivals in Langkawi

Forecasting techniques can assist host nations in anticipating and meeting tourists' needs well in advance. Apart from that, the process of forecasting involves making precise and optimal forecasts. Forecasting is used by many corporate stakeholders for a variety of purposes such as budget planning, quantity projections and future cost estimation (Mishra et al., 2021). To plan and market tourism, modelling and predicting the number of tourist arrivals are essential. As a result, it is critical for making policy decisions that will promote sustainable tourism development (Msofe & Mbago, 2019). A tourism forecasting system can guide the preparation and organization of necessary things for tourists by the administration. The forecasting method will enable the government to implement the necessary modifications more quickly and effectively (González -Rivera et al., 2019).

After all, this study findings are based on the Box-Jenkins seasonal autoregressive integrated moving average (SARIMA) approach since monthly data on tourist arrivals are more likely assumed to have the presence of seasonality and is widely used in the time series data. A time-series data forecasting has made extensive use of the ARIMA model and its seasonal extension, SARIMA which can also handle the data's seasonal pattern and behaviors (Adineh et al., 2020). Not only that, although there are many types of forecasting models available to use for this kind of time series data, yet there are only few studies that cover Box-Jenkins seasonal autoregressive integrated moving average (SARIMA) model for forecasting the number of tourist arrivals in Langkawi. The Box-Jenkins approach is extensively utilized in various domains, specifically for predicting the cement production, vegetable productivity and life expectancy as used by Hamid and Abushaba (2020), Thapa et al. (2022) and Husin and Aziz (2021). Research from Mansor and Ishak (2015) has used the Box-Jenkins technique of forecasting, which comprised the autoregressive fractionally integrated model (ARFIMA) and the autoregressive integrated moving average (ARIMA) model. They also mentioned in the study that the ARFIMA model worked best on the modelling arrival of tourists in Langkawi. Moreover, Halim and Muda (2021) have done a study regarding the modelling and forecasting the tourism demand in Malaysia using Box-Jenkins and Singular Spectrum Analysis. They used secondary data of monthly tourism arrival in the year of 1990 to 2014 and forecasted for the next year which is 2015. After all, the study has concluded where Box-Jenkins time series method is the best model compared to the singular spectrum analysis.

To sum up, this research aims to clarify the trend patterns of Langkawi's tourist arrivals, to identify the best model of forecasting the number of tourist arrivals and lastly is to forecast the tourist arrivals in Langkawi from July 2024 until July 2025. The use of forecasting techniques to reliably estimate tourist arrivals is supported by a few studies and this is important for predicting how Langkawi's tourism sector would rebound after the state government has announced the program of 'Visiting Kedah 2025'. Research on forecasting the number of tourist arrivals in Langkawi from the year 2024 until 2025 are quite limited. To overcome the issue, this research applied the Box-Jenkins methodology to find the best model for forecasting the value of tourist arrivals in Langkawi. The following chapter goes into more detail about these techniques.

3.0 METHODOLOGY

All the data came from Langkawi's tourism official website, LADA. This data consists of monthly data from January 2010 until June 2024, totalling 174 samples. This study focuses on a single quantitative variable: the number of tourist arrivals, which covers both domestic and international visitors. Descriptive analysis is the most straightforward and basic method of describing the specific variable. A trend line graph including time series data was useful for this research in observing the monthly tourist arrivals in Langkawi. A line graph, which connects points with trend lines, illustrates how the number of tourist arrivals can make slight changes over time.

3.1 Box-Jenkins Methodology

The Box-Jenkins method has been applied in this study to forecast the number of tourist arrivals in Langkawi. A methodical way to create and utilize ARIMA models for time series forecasting is the Box-Jenkins methodology. This model can do the prediction of future data patterns using historical data. In the forecasting of time series data, the Box-Jenkins method is directly related to ARIMA models since it is a quantitative approach that provides a rigorous approach to model estimation and validation.

3.1.1 Autoregressive Integrated Moving Average (ARIMA)

ARIMA modelling is a statistical technique that forecasts future trends using time series data, and it is helpful especially when the variable's stationarity assumption is not fully satisfied. It is necessary to differentiate the data series to attain stationarity. Autoregressive (AR) and Moving Average (MA) models are combined into a single framework by an ARIMA model. ARIMA (p,d,q) is a common representation of the model, where d is the number of times that the variable tourist arrival must be differenced to attain stationarity. A simple model case for ARIMA (1,1,1) model can be written as, $w_t = \mu + \phi_1 w_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t$, where $w_t = y_t - y_{t-1}$ serves as the data series' first difference and is stationary. The values of each parameter in the equation are $p = 1, d = 1, and\ q = 1$. The initial difference did not make the sequence stationary on certain occasions. Hence, if there is a need for second differencing, the second order series would be integrated, where the value of d is two. For example, in ARIMA (1,2,1), the AR and MA models were both first order, with p and q equal to one and d equal to two. If the data series has not yet achieved the stationarity, continue to apply differencing until the data achieve the stationarity.

The data series will become a Seasonal Autoregressive Integrated Moving Average (SARIMA) model if it has a seasonal component. The non-seasonal ARIMA model was known as ARIMA(p,d,q), where p stood for the autoregressive (AR) process order, d for the differencing order required for the time series to become stationery and q for the moving average (MA) order process. Meanwhile, the seasonal terms (P,D,Q) in the ARIMA model could be multiplied by the non-seasonal terms (p,d,q) to form a seasonal ARIMA (SARIMA) model. The notation applied was SARIMA $(p,d,q)(P,D,Q)_{[m]}$ where m is the data's frequency and the capital and lowercase letters denoted the seasonal and no-seasonal parameters of the model, respectively (Perone, 2022). For the SARIMA models with seasonal order terms (P,D,Q) and non-seasonal order terms (p,d,q), the approximated basic equation was as follows (Yin et al., 2023).

3.1.2 Stages in ARIMA Model Development

The Box-Jenkins modelling approach is based on three primary phases which consists of model identification, model estimation and validation and model application. Thus, all the three steps will be explained further in this section.

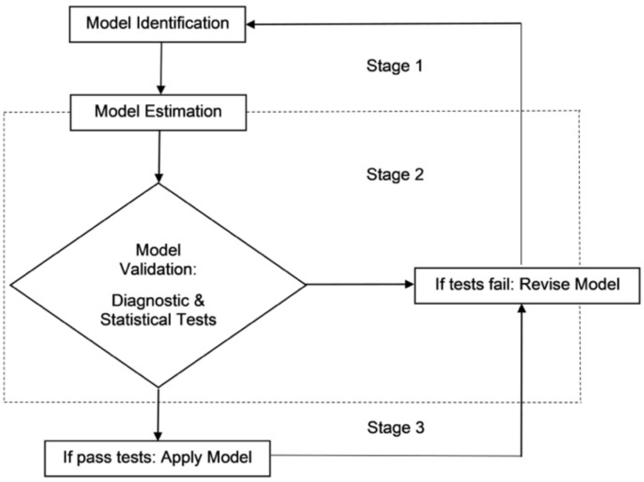

Figure 1: Stages in Box-Jenkins Methodology

Figure 1 illustrates the different stages of the Box-Jenkins Methodology. The first stage, known as model identification, involves determining the order of the Autoregressive (AR) and Moving Average (MA) components by examining the autocorrelation function (ACF) and partial autocorrelation function (PACF). The autocorrelation coefficients (ACs) measured the degree of relationship between the current value of time series and its past value at different lags (e.g., lag 1, 2, 3, etc.), while the partial autocorrelation coefficients (PACs) measured the degree of association between y_t and y_{t-k}. Once the time series has been transformed to achieve stationarity, the ACF and PACF plots are examined to identify any patterns. If PACF shows a sharp cutoff after a few significant spikes while the ACF gradually decays, the data are best modelled by an AR (p) process, where p represents the number of significant spikes in the PACF. Conversely, if the ACF display a sharp cutoff and the PACF decays gradually, an MA (q) model may be appropriate, where q denotes the number of significant spikes in the ACF.

Model validation and estimation are done in the second stage. The parameters of the AR and MA orders are determined by the model estimate, while the model's validity is evaluated through model validation. The validation procedure consists of three parts which are statistical validation of residual diagnostics, parameter validation and model validation. When the residuals are guaranteed to meet the requirements of a stationary univariate process, which is the situation when they are normally, randomly and independently distributed with mean zero and variance, σ^2. To compare the various models, this phase displays a table of the goodness of fit statistics using the Ljung-Box Test, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). To validate the ARIMA models, statistical test such as the Ljung-Box Statistic, Akaike's Information Criteria (AIC), and Bayesian Information Criterion (BIC) were used. The Ljung-Box test checks if the residuals from a time series model show no autocorrelation at lag (k) by computing chi-squared value. The hypothesis is, H_0: The errors are white noise against H_1: The errors are white noise. A white noise series has zero autocorrelation at all lags. If the Ljung-Box test fails to reject H_0 (p-value > 0.05), there is no significant serial correlation, meaning the model is adequate.

The AIC measure how well a model fits the data. The model with the lowest AIC is considered to have a superior fit. Nonetheless, the principle of parsimony continued to apply when selecting the best model, with the model with the least parameters being given the priority. Meanwhile, BIC selects the most accurate out-of-sample model by balancing complexity and goodness-of-fit. It is similar to AIC but penalizes additional parameters more strictly. The best model has the lowest BIC. AIC and BIC are calculated as:

$$AIC = e^{(\frac{2k}{T})} \frac{\sum_{t=1}^{T} e_t^2}{T} \qquad BIC = T^{\frac{k}{T}} \frac{\sum_{t=1}^{T} e_t^2}{T}$$

where k the number of estimated parameters (AR, MA and seasonal terms) and T is the number of observations. Since BIC and AIC were closely related and because one of their similarities is that the lower the BIC value, the better ARIMA model. Once a model meets validation criteria, it can be implemented. Otherwise, modifications are needed. The best-fitting Box-Jenkins model, selected based on the lowest AIC and BIC values, is used to forecast tourist arrivals in Langkawi. Forecasts are presented either as point estimates or within confidence intervals, with the model balancing accuracy and simplicity.

4.0 RESULT AND DISCUSSION

This chapter presents the findings to achieve the study's objectives. The analysis included descriptive analysis, and the Box-Jenkins methodology. The data on monthly tourist arrivals in Langkawi was divided into two, estimation part (January 2010 to August 2018) and evaluation part (September 2018 to June 2024). All outputs that were produced were examined to glean crucial information about the number of tourist arrivals in Langkawi.

4.1 Trend of The Number of Tourist Arrivals in Langkawi

This section presents the trend analysis of monthly tourist arrivals in Langkawi, using a time series line graph in Figure 2 to visually interpret the patterns. The graph spans from January 2010 to June 2024 and serves as the basis for understanding the overall dynamics of tourism on the island. The trend analysis reveals several notable patterns, including recurring seasonal fluctuations, long-term growth trends, and abrupt interruptions due to external events like the COVID-19 pandemic.

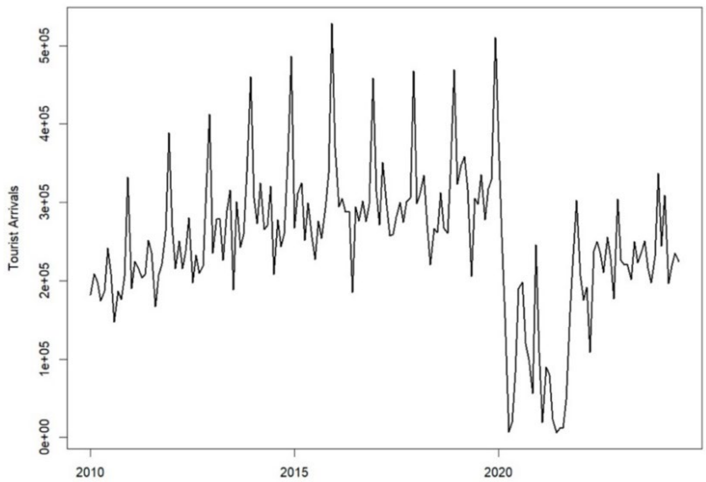

Figure 2: The Monthly Series of the Number of Tourist Arrivals in Langkawi

From 2010 until early 2020, the data exhibit a strong and consistent seasonal pattern, characterized by distinct peaks and troughs. These peaks typically occur during major holiday seasons, such as the year-end school holidays from mid-November to early January, and festive periods including Christmas and New Year celebrations. Another increase in tourist arrivals is commonly observed from March to September, which aligns with Malaysia's drier weather conditions. This period is especially appealing for beach tourism and outdoor activities, both of which are prominent features of Langkawi's tourism offerings. The recurring nature of these patterns highlights the seasonal nature of tourism demand on the island. In addition to seasonal variations, the trendline shows a gradual upward trajectory in the number of tourist arrivals over the years, suggesting that Langkawi has experienced steady growth in its tourism industry. This growth may be attributed to increased marketing efforts, improved infrastructure, and broader recognition of Langkawi as a premier tourist destination. The consistent rise in visitor numbers also reflects the island's ability to attract both domestic and international tourists due to its unique attractions, such as its UNESCO Global Geopark status, duty-free shopping, and a wide range of leisure activities. However, there is a noticeable decline in arrivals in 2020, most likely because of an unusual occurrence like the COVID-19 pandemic. This sharp downturn reflects the immediate impact of nationwide lockdowns, travel restrictions, and border closures. The usual seasonal peaks disappear during this period, and the trend becomes highly erratic, indicating an unstable recovery phase. While there are signs of gradual recovery post-2020, the seasonal patterns are less pronounced and more volatile, reflecting ongoing global uncertainties and changes in travel behaviour.

Despite these fluctuations, the post-pandemic trend indicates a slow yet steady rebound in tourist arrivals, especially after the implementation of the Langkawi tourism bubble in late 2021. This initiative helped restore tourist confidence and signalled a renewed interest in domestic travel. Descriptive analysis confirms the presence of seasonality, as the dataset comprises monthly time series data on tourist arrivals. Seasonality is a well-known characteristic of tourism trends and can be influenced by various factors such as rainfall, temperature, and wind patterns (Alshuqaiqi & Omar, 2019). For example, Jong et al. (2023) noted that April and May typically experience the hottest weather in Malaysia, resulting in lower tourist arrivals during those months. Conversely, January and December consistently record the highest arrivals due to school holidays and festive celebrations. These insights support the use of seasonal forecasting models, such as the SARIMA model, to project future tourist arrivals with a reasonable degree of accuracy.

In conclusion, the trend analysis of monthly tourist arrivals in Langkawi underscores both the resilience and sensitivity of the tourism sector. It highlights the importance of incorporating seasonal components into forecasting models, particularly in the context of post-pandemic recovery planning and future tourism development.

4.2 Box-Jenkins Methodology

The Box-Jenkins methodology is widely recognized for its effectiveness in modelling time series data, especially in contexts where data exhibit trend and seasonal variations. This section describes the application of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, which is a seasonal extension of the standard ARIMA model. In this study, SARIMA is employed to forecast monthly tourist arrivals in Langkawi, due to its ability to handle both trend and seasonal components found in tourism data. In general, the data series must be assumed to be stationary. Figure 3 and 4 shows the ACF and PACF of the data series.

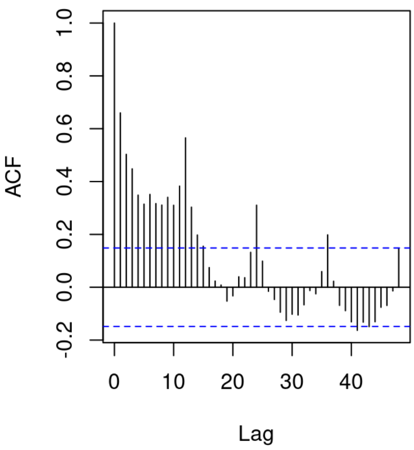

Figure 3: ACF of Tourist Arrivals

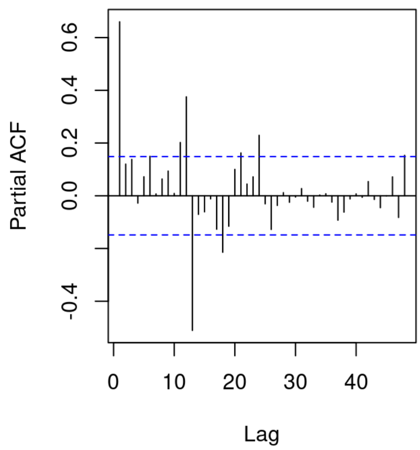

Figure 4: PACF of Tourist Arrivals

Figures 3 and 4 display an ACF and PACF series of the tourist arrivals in Langkawi with a gradual decay to zero and a PACF series with significant spikes at regular intervals, like lag 12 and lag 24. Consequently, it demonstrated the non-stationarity of the series. The correlation strength is indicated by the values on the y-axis, which reflect the correlation coefficient and range from 0.0 to 0.8. The series is said not to correlate if the lag value is 0.0, whereas a significant positive autocorrelation is indicated if the lag value is 0.8.

According to the ACF plot, significant spikes appear in the initial few lags, suggesting a strong association between adjacent observations. The ACF exhibits a slow decrease rather than a rapid drop to zero due to the existence of seasonality. Meanwhile, the PACF shows periodic spikes at lag 1 and happens at regular intervals like lag 12 and lag 24. Furthermore, a spike at lag 12 indicates a strong correlation between observations that are separated by 12 months, pointing to an annual pattern. On the other hand, a spike at lag 24 suggests that seasonality endures throughout several cycles. Hence, the PACF pattern points to an AR (1)-like structure but the ACF's non-stationary features may hint that differencing is necessary. Also, the p-value for the Augmented Dickey-Fuller (ADF) test was 0.1769, which was more than the alpha value which is $\alpha = 0.05$ and indicated that the series has a unit root. Overall, the series' non-stationarity is further demonstrated by the presence of a unit root. The data series of tourist arrivals in Langkawi must undergo differencing until the ACF and PACF plot exhibits no signs of decay to attain stationarity, since the data series meets all the requirements for non-stationary data. The data series may be used for forecasting once it is stationary.

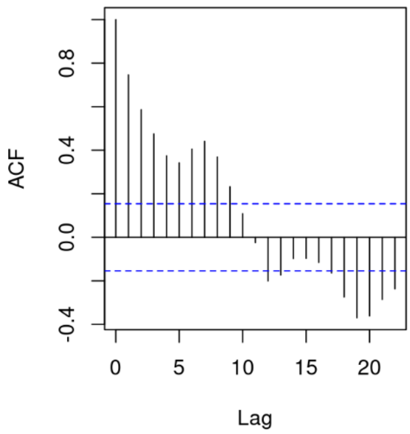

Figure 5: ACF of Tourist Arrivals After Seasonal Differencing

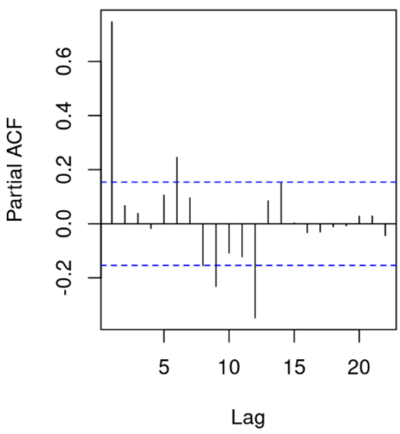

Figure 6: PACF of Tourist Arrivals After Seasonal Differencing

Figure 5 shows a gradual decay in the autocorrelations over the delays in the ACF plot. This slow decay implies that even after the seasonal component has been eliminated, the series may still exhibit non-stationarity in the time series. In the meantime, Figure 6 shows the PACF plot displays significant spikes at lag 1 and a few more minor spikes during the first few lags, which may imply some short-term autocorrelation structure. Additionally, the series exhibited a unit root, as indicated by the Augmented Dickey-Fuller (ADF) test p-value of 0.4343, which is greater than $\alpha = 0.05$. This indicates that a unit root still exists, suggesting the time series remains non-stationary. Therefore, non-seasonal differencing must be performed to eliminate the non-stationarity from the series. The observed behaviours of the ACF and PACF plots after seasonal differencing suggest that the series may still be integrated, implying that non-seasonal differencing could facilitate the achievement of stationarity.

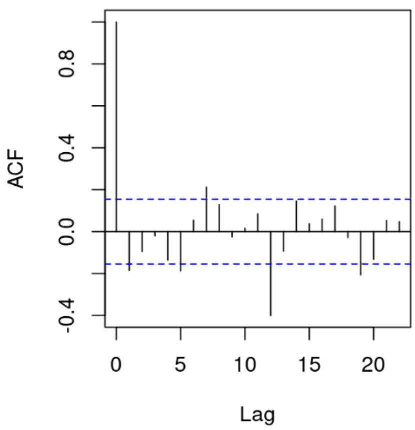

Figure 7: ACF of Tourist Arrivals After Non-Seasonal Differencing

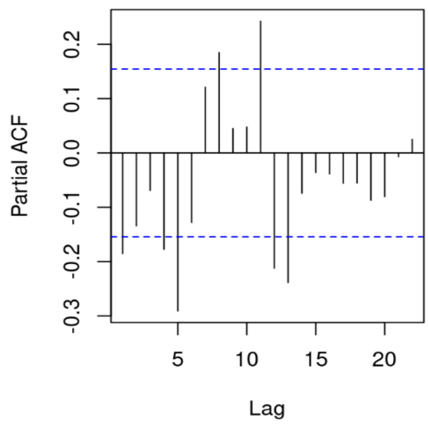

Figure 8: PACF of Tourist Arrivals After Non-Seasonal Differencing

Figure 7 and 8 shows that after the non-seasonal differencing, both ACF and PACF plots shows no signs of decay. Thus, the series is said to have reached stationarity because order differencing was performed where d = 1. Then, SARIMA $(p,d,q)(P,D,Q)_{[12]}$ could be used to carry out the model identification phase. To identify the proper SARIMA model, one must look at the lag that was greater than the standard error, SE band, in both the ACF and PACF plots in Figure 7 and 8. Significant lags at lag 1, lag 5 and lag 7 beyond the SE band are shown in the ACF plot in Figure 7. This suggests that there is an MA component of order 3, MA (3) in the non-seasonal part. Since d = 1, q = 3 and D = 1, the model identification would result in SARIMA $(p,1,3)(P,1,Q)_{[12]}$.

Meanwhile, the PACF plot in Figure 8 shows three significant spikes at lags 5, 8 and 11, which exceed the standard error bands. This demonstrates that there is an AR component of order 3, AR (3) in the time series. The seasonal SAR component of the model is also indicated by a rather significant spike at lag 12. Since the model would therefore have p = 3, SARIMA $(3,1,3)(1,1,1)_{[12]}$ was determined to be the proper SARIMA model.

However, despite these observations, the precision of the associated values *p* and *q* in the model cannot be guaranteed. To ensure that a well-defined model was not overlooked, several model formulations would be calculated. SARIMA $(3,1,3)(1,1,1)_{[12]}$, SARIMA $(2,1,2)(1,1,1)_{[12]}$, SARIMA $(1,1,2)(1,1,1)_{[12]}$, SARIMA $(1,1,1)(1,1,1)_{[12]}$, SARIMA $(0,1,1)(1,1,1)_{[12]}$, SARIMA $(1,1,1)(0,1,1)_{[12]}$ and SARIMA $(0,1,1)(0,1,1)_{[12]}$ were the seven models proposed. To find out if serial correlation occurred in a time series, the p-value was analyzed for each model of the Ljung-Box Test in Table 1.

Table 1: Summary of Ljung-Box Test

Model	P-value
SARIMA $(3,1,3)(1,1,1)_{[12]}$	0.7552
SARIMA $(2,1,2)(1,1,1)_{[12]}$	0.7951
SARIMA $(1,1,2)(1,1,1)_{[12]}$	0.6605
SARIMA $(1,1,1)(1,1,1)_{[12]}$	0.6509
SARIMA $(0,1,1)(1,1,1)_{[12]}$	0.4174
SARIMA $(1,1,1)(0,1,1)_{[12]}$	0.7867
SARIMA $(0,1,1)(0,1,1)_{[12]}$	0.5495

The Ljung-Box test, which assesses whether the errors in each SARIMA model satisfy the requirements for white noise, is summarized in Table 1. If a model's Ljung-Box test p-value is greater than 0.05, it is said to have no serial correlation. Since all seven SARIMA models have a p-value greater than $\alpha = 0.05$, these results led to the acceptance of the null hypothesis, and it can be concluded that all errors are white noise. Also, it was determined to satisfy the requirements for identifying the best model, since they were sufficient and well-defined.

4.3 Determining the Best SARIMA Model

After conducting model identification and ensuring stationarity through both seasonal and non-seasonal differencing, the next crucial step in the Box-Jenkins methodology is determining the best SARIMA model for forecasting the number of tourist arrivals in Langkawi.

Table 2: Summary of AIC and BIC value

Model	AIC	BIC
SARIMA $(3,1,3)(1,1,1)_{[12]}$	2166.1	2188.7
SARIMA $(2,1,2)(1,1,1)_{[12]}$	2164.0	2181.6
SARIMA $(1,1,2)(1,1,1)_{[12]}$	2163.0	2178.0
SARIMA $(1,1,1)(1,1,1)_{[12]}$	2161.1	2173.6
SARIMA $(0,1,1)(1,1,1)_{[12]}$	2161.6	2171.6
SARIMA $(1,1,1)(0,1,1)_{[12]}$	2159.5	2169.5
SARIMA $(0,1,1)(0,1,1)_{[12]}$	2159.3	2166.8

This involves comparing several candidate models based on their statistical adequacy and performance using model selection criteria. The Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used as standard model selection tools. In terms of the selection of the best model, the model that contains the lowest value of both AIC and BIC will be chosen as the best model. Table 2 shows that SARIMA $(0,1,1)(0,1,1)_{[12]}$ has AIC and BIC values of 2159.3 and 2166.8 which is the lowest among all. This indicates that SARIMA $(0,1,1)(0,1,1)_{[12]}$ is the simplest model that fits the data well. Hence, SARIMA $(0,1,1)(0,1,1)_{[12]}$ would be the best model to use in forecasting the number of tourist arrivals in Langkawi for the next 13 months since it has the best balance in terms of model fit and complexity.

Forecasting is widely used by corporate stakeholders for various purposes, including budget planning, quantity projections, and future cost estimation (Mishra et al., 2021). In the tourism sector, modeling and predicting the number of tourist arrivals is crucial for effective planning and marketing. A study by Shabri et al. (2020) focused on forecasting monthly tourist arrivals in Langkawi from 2002 to 2016 using ARIMA, ANN, DNN, and hybrid models. Their results indicated that the ARIMA $(1,0,2)(0,1,1)_{[12]}$ model was the most suitable, as it yielded the lowest Akaike Information Criterion (AIC) value. Similarly, Ahmad et al. (2022) found that among several SARIMA models, SARIMA $(1,1,0)(0,1,0)_{[12]}$ was the best for analyzing tourist arrival data at the MARDI Langkawi Agro Technology Park. The Box-Jenkins methodology has also been widely applied in other countries. For instance, Ibanez & Edradan (2024) identified SARIMA $(3,1,4)(0,1,2)_{[12]}$ as the best model for forecasting tourist arrivals in the Philippines. Likewise, Devi et al. (2024) reported that the same model was effective in forecasting tourist arrivals in India, based on the lowest AIC and Bayesian Information Criterion (BIC) values.

4.4 Forecast the Number of Tourist Arrivals in Langkawi

The forecast data and model based on SARIMA $(0,1,1)(0,1,1)_{[12]}$ were generated using the Box-Jenkins method. Thus, the SARIMA $(0,1,1)(0,1,1)_{[12]}$ model was used to forecast the number of tourist arrivals in Langkawi starting from July 2024 until July 2025 in monthly equivalents of 13 months since it was determined to be the best model for forecasting. The forecast values of the number of tourist arrivals in Langkawi for the upcoming 13 months, from July 2024 until July 2025, are displayed in Table 3 and were expressed in a graphical way in Figure 9.

Table 3: Summary of Forecast Values of The Number of Tourist Arrivals in Langkawi

Year	Month	Forecast	Low 95 CI	High 95 CI
2024	July	239067.7	142343.5	335792.0
2024	August	249404.2	136323.2	362485.2
2024	September	236340.0	108986.0	363694.0
2024	October	245553.5	105372.3	385734.6
2024	November	265678.3	113749.0	417607.5
2024	December	394230.0	231398.1	557061.8
2025	January	277843.2	104794.7	450891.7
2025	February	267657.3	84963.2	450351.4
2025	March	249982.7	58127.3	441838.1
2025	April	214520.1	13921.4	415118.9
2025	May	231411.7	22435.2	440388.3
2025	June	236585.6	19554.4	453616.8
2025	July	246778.0	15159.4	478396.5

Together with the lower and higher confidence interval (CI) at 95% intervals generated by applying the SARIMA $(0,1,1)(0,1,1)_{[12]}$ together with its parameters and the data in the dataset, Table 3 displays the forecasted number of tourist arrivals in Langkawi for the next 13 months. According to the forecasts, Langkawi's number of tourist arrivals is predicted to experience a steadily increasing trend over the next two years starting from July 2024 until July 2025. The interval for December 2024 is especially wide from 231398.1 to 557061.8 indicates that higher uncertainty during that month. Moreover, the lowest predicted value of the number of tourist arrivals at the 95% confidence level is 13921.4, while the maximum would be 557061.8. Hence, the predicted value of the number of tourist arrivals in Langkawi at 95% confidence interval is between 13921.4 and 557061.8.

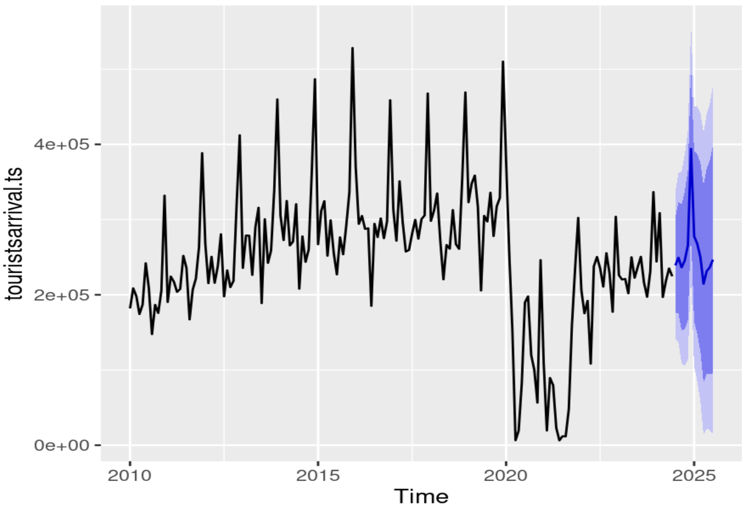

Figure 9: Forecast of The Number of Tourist Arrivals in Langkawi from July 2024 to July 2025

A line graph representing the time series of the actual number of tourist arrivals values for Langkawi from January 2010 to June 2024 is displayed in Figure 9. After that, the line intersects with the forecasted values for July 2024 until July 2025. It is forecasted that the number of tourist arrivals in Langkawi will be 239067.7 in July 2024. By July 2025, it should have steadily increased to 246778.0. As a result, the model's predictions have been steadily increasing.

5.0 CONCLUSION

Tourism is a vital component of Langkawi's economic landscape, contributing significantly to employment, income generation, and local development. Accurate forecasting of tourist arrivals is essential for strategic planning, resource management, and policy formulation, particularly in the context of economic recovery and tourism development initiatives such as the *Visiting Kedah 2025* programme. This study successfully applied the Box-Jenkins methodology, specifically the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, to forecast the number of tourist arrivals in Langkawi for the period from July 2024 to July 2025. The time series analysis of tourist arrival data from January 2010 to June 2024 revealed clear seasonal patterns, characterized by consistent peaks during school holidays and festive seasons, alongside a significant decline during the onset of the COVID-19 pandemic in 2020. These observations validated the presence of trend and seasonal components in the data, justifying the use of the SARIMA model for forecasting purposes. Among the seven SARIMA models tested, SARIMA (0,1,1) $(0,1,1)_{12}$ had the lowest AIC and BIC values, making it the most appropriate for forecasting. The model predicts a steady increase in tourist arrivals from July 2024 to July 2025, with forecasted values ranging between 13,921.4 and 557,061.8 at a 95% confidence interval. The number of arrivals is expected to grow from 239,067.7 in July 2024 to 246,778.9 in July 2025, indicating a positive trend. These findings suggest a promising outlook for Langkawi's tourism sector, reinforcing the SARIMA model's effectiveness in providing reliable forecasts. Future research could explore additional variables, such as economic factors or policy changes, to further enhance prediction accuracy. This study offers valuable insights for the Malaysian government, particularly the Ministry of Tourism, Arts and Culture (MOTAC) and the Langkawi Development Authority (LADA), by supporting accurate forecasting. These insights can assist in strategic planning, resource allocation, marketing initiatives, promotional efforts, and capacity management.

ACKNOWLEDGEMENTS

The authors are grateful to Universiti Teknologi MARA (UiTM), Seremban branch, and Langkawi Development Authority (LADA) for providing the data. They also extend their gratitude to other researchers for their valuable ideas and discussions.

REFERENCES

Adineh, A. H., Narimani, Z., & Satapathy, S. C. (2020). Importance of data preprocessing in time series prediction using sarima: A case study. *International Journal of Knowledge-based and Intelligent Engineering Systems*, *24*(4), 331–342. https://doi.org/10.3233/kes-200065

Ahmad, A. A., Nik Mohd Masdek, N. R., & Pono, Z. (2022). *Statistical modelling and forecasting of tourist arrivals in MARDI Langkawi Agro Technology Park using SARIMA*.

Alioglu, E. C. (2022). Tourism forecasting with time series: Comparison of univariate and multivariate approaches *[Doctoral dissertation, Tilburg University].*

Al-sakkaf, M. A., Mohaidin, Z., Iskan, Y. H. P., Murshid, M. A., & Amarneh, S. (2022). The influence of tourism impacts and destination social responsibility on support for tourism development: A conceptual model. *Journal of Tourism, Hospitality and Environment Management*, *7*(29), 76–93. https://doi.org/10.35631/jthem.729007

Alshuqaiqi, A., & Omar, S. I. (2019). Causes and implication of seasonality in tourism. *Journal of Advanced Research in Dynamical and Control Systems*, *11*(4), 1480-1486.

Altahir, A. A., Mathur, N., Thiruchelvam, L., Mustafa Abro, G. E., Radzi, S., Dass, S. C., Gill, B. S., Sebastian, P., Zulkifli, S. A., & Asirvadam, V. S. (2020). Modeling the impact of lock-down on covid-19 spread in malaysia. *bioRxiv*, 2020–07.

Astro Awani. (2021, October 2). *Langkawi tourism bubble project a great success - LADA*. https://international.astroawani.com/malaysia-news/langkawi-tourism-bubble-project-great-success---lada-323295

Devi, R., Agrawal, A., Dhar, J., & Misra, A. K. (2024). Forecasting of Indian tourism industry using modeling approach. *MethodsX*, 102723.

Fernández, J. A. S., Mart´ınez, J. M. G., & Mart´ın, J. M. M. (2022). An analysis of the competitiveness of the tourism industry in a context of economic recovery following the covid19 pandemic. *Technological Forecasting and Social Change*, *174*, 121301. https://doi.org/10.1016/j.techfore.2021.121301

González-Rivera, G., Loungani, P., & Sheng, X. S. (2019). Forecasting issues in developing economies. *International Journal of Forecasting*, *35*(3), 927 - 928. https://doi.org/10.1016/j.ijforecast.2019.04.005

Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: A rapid assessment of covid-19. *Journal of sustainable tourism*, *29*(1), 1–20. https://doi.org/10.1080/09669582.2020.1758708

Halim, S. S. A., & Muda, N. (2021). Modeling and forecasting of tourism demand in malaysia. *International Journal of Current Science Research and Review*, *03*(12). https://doi.org/10.47191/ijcsrr/v3-i12-06

Hamid, E. Y., & Abushaba, M. H. O. (2020). Use of exponential holt model and the box- jenkins methodology in predicting the time series of cement production in sudan. *Journal of Al-Qadisiyah for computer science and mathematics*, *12*(1), Page–42. https://doi.org/10.29304/jqcm.2020.12.1.666

Harun, M. K., Arifin, H. S., Anwar, S., Putri, E. I. K., & Tata, H. L. (2022). Agroforestry approaches in the restoration of peatland landscapes in central kalimantan, indonesia. In *Forest dynamics and conservation: Science, innovations and policies* (pp. 331–362). Springer. https://doi.org/10.1007/978-981-19-0071-6_16

Husin, W. Z. W., & Aziz, N. A. (2021). Box-jenkins model for forecasting malaysia life expectancy. *International Journal of Academic Research in Business and Social Sciences*. https://doi.org/10.6007/ijarbss/v11-i12/11980

Ibanez, D. A., & Edradan, M. Forecasting Tourist Arrivals of Philippines from August 2024 to December 2026 using Seasonal Autoregressive Integrated Moving Average.

Irwana Omar, S., Ghapar Othman, A., & Mohamed, B. (2014). The tourism life cycle: An overview of langkawi island, malaysia. *International Journal of Culture, Tourism and Hospitality Research*, *8*(3), 272–289. https://doi.org/10.1108/ijcthr-09-2013-0069

Jong, M. C., Puah, C. H., & Arip, M. A. (2023). assessing the impacts of tourism climate index and holiday climate index on tourism demand in Malaysia. *Asian Journal of Business Research Volume*, *13*(1), 67-81.

Mansor, K. A., & Ishak, W. I. (2015). Forecasting tourist arrivals to langkawi island malaysia. *Cross-sCultural Management Journal*, *17*(1), 69–76. https://EconPapers.repec.org/RePEc:cmj:journl:y:2015:i:7:p:69-76

Mishra, R. K., Urolagin, S., Jothi, J. A. A., Nawaz, N., & Ramkissoon, H. (2021). Machine learning based forecasting systems for worldwide international tourists' arrival. *International Journal of Advanced Computer Science and Applications*, *12*(11).

Mohd Shariff, N., & Zainol Abidin, A. (2020). Tourists'perceptions of langkawi island as a world tourism destination. *GeoJournal of Tourism and Geosites*, *32*(4), 1264–1269. https://doi.org/10.30892/gtg.32411-567

Msofe, Z. A., & Mbago, M. C. (2019). Forecasting international tourist arrivals in zanzibar using box-jenkins SARIMA model. *General Letters in Mathematics (GLM)*, *7*(2).

Shabri, A., Samsudin, R., & Yusoff, Y. (2020, May). Combining deep neural network and fourier series for tourist arrivals forecasting. In *IOP Conference Series: Materials Science and Engineering* (Vol. 864, No. 1, p. 012094). IOP Publishing.

Shariff, N. M., & Abidin, A. Z. (2020). TOURISTS'PERCEPTIONS OF LANGKAWI ISLAND AS A WORLD TOURISM DESTINATION. *Geo Journal of Tourism and Geosites*, *32*(4), 1264-1269.

Shafikhullah, S. N. A., & Nayan, N. (2021). Impact of tourism activities on the economy of the community in the coastal area of langkawi island, malaysia. *Forum Ilmu Sosial*, *48*(1), 11–24. https://doi.org/10.15294/fis.v48i1.30781

Tang, K. H. D. (2020). Movement control as an effective measure against covid-19 spread in malaysia: An overview. *Journal of Public Health*, 1–4. https://doi.org/10.1007/s10389-020-01316-w

Thapa, R., Devkota, S., Subedi, S., & Jamshidi, B. (2022). Forecasting area, production and productivity of vegetable crops in nepal using the box-jenkins arima model. *Turkish Journal of Agriculture-Food Science and Technology*, *10*(2), 174–181. https://doi.org/10.24925/turjaf.v10i2.174-181.4618

The Star. (2021, October 10). *Langkawi travel bubble generates income of RM24.9mil as of Wednesday (Oct 6)*. https://www.thestar.com.my/news/nation/2021/10/10/langkawi-travel-bubble-generates-income-of-rm249mil-as-of-wednesday-oct-6

Tourism, U. N. (2025). Glossary of tourism terms. *United Nations. Retrieved*, *3*. https://www.unwto.org/glossary-tourism-terms

World Health Organization, 2., et al. (2020). *Guidance on covid-19 for the care of older people and people living in long-term care facilities, other non-acute care facilities and home care* (tech. rep.). WHO Regional Office for the Western Pacific. https://iris.who.int/handle/10665/331913

Yin, C., Liu, K., Zhang, Q., Hu, K., Yang, Z., Yang, L., & Zhao, N. (2023). Sarima-based medium-and long-term load forecasting. *Strategic Planning for Energy and the Environment*, 283–306. https://doi.org/10.13052/spee1048-5236.4222

Yuzaidi, M. A., Saini, M. I. A. S., Zaini, N. S. H., Mazlan, N. S. B., & Safri, F. H. M. (2022). The tourist risk and perception from the effect covid 19 in langkawi, kedah.

Zheng, S., & Zhang, Z. (2023). Adaptive tourism forecasting using hybrid artificial intelligence model: A case study of xi'an international tourist arrivals. *PeerJ Computer Science*, *9*, e1573. https://doi.org/10.7717/peerj-cs.1573