



Predictive analytics of COVID-19 cases and tourist arrivals in ASEAN based on covid-19 cases

Shubashini Rathina Velu¹ · Vinayakumar Ravi² · Kayalvily Tabianan³

Received: 23 May 2022 / Accepted: 22 September 2022

© The Author(s) under exclusive licence to International Union for Physical and Engineering Sciences in Medicine (IUPESM) 2022

Abstract

Purpose Research into predictive analytics, which helps predict future values using historical data, is crucial. In order to foresee future instances of COVID-19, a method based on the Seasonal ARIMA (SARIMA) model is proposed here. Additionally, the suggested model is able to predict tourist arrivals in the tourism business by factoring in COVID-19 during the pandemic. In this paper, we present a model that uses time-series analysis to predict the impact of a pandemic event, in this case the spread of the Coronavirus pandemic (Covid-19).

Methods The proposed approach outperformed the Autoregressive Integrated Moving Average (ARIMA) and Holt Winters models in all experiments for forecasting future values using COVID-19 and tourism datasets, with the lowest mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), and root mean squared error (RMSE). The SARIMA model predicts COVID-19 and tourist arrivals with and without the COVID-19 pandemic with less than 5% MAPE error.

Results The suggested method provides a dashboard that shows COVID-19 and tourism-related information to end users. The suggested tool can be deployed in the healthcare, tourism, and government sectors to monitor the number of COVID-19 cases and determine the correlation between COVID-19 cases and tourism.

Conclusion Management in the tourism industries and stakeholders are expected to benefit from this study in making decisions about whether or not to keep funding a given tourism business. The datasets, codes, and all the experiments are available for further research, and details are included in the appendix.

Keywords COVID-19 · Tourism · Predictive analytics · Regression models

1 Introduction

In the past, many researchers and analysts have performed their own forecasting in their own research area, including the tourism industry by using big data which is made up of an enormous volume of structured data, unstructured data, and semi-structured data. However, sometimes, the forecast might not be similar to the actual results when the future has exactly come. There is a big difference between the prediction and actual when it comes to reality could be due to the occurrence of unexpected factors which could not be predicted by using the historical statistic. For instance, the Coronavirus pandemic caused chaos worldwide starting from the year 2019 (Covid-19) until today in the year 2021 [1, 2]. There is no doubt that the sudden outbreak of coronavirus is unpredictable by any human being, hence, the relevant business predictions done in the past might have become inaccurate due to the new factors: the spread of the Covid-19 virus and restriction actions applied by governments.

This article is part of the COVID-19 Health Technology: Design, Regulation, Management, Assessment.

✉ Vinayakumar Ravi
vravi@pmu.edu.sa

Shubashini Rathina Velu
svelu@pmu.edu.sa

Kayalvily Tabianan
kayalvily.tabianan@newinti.edu.my

¹ Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

² Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

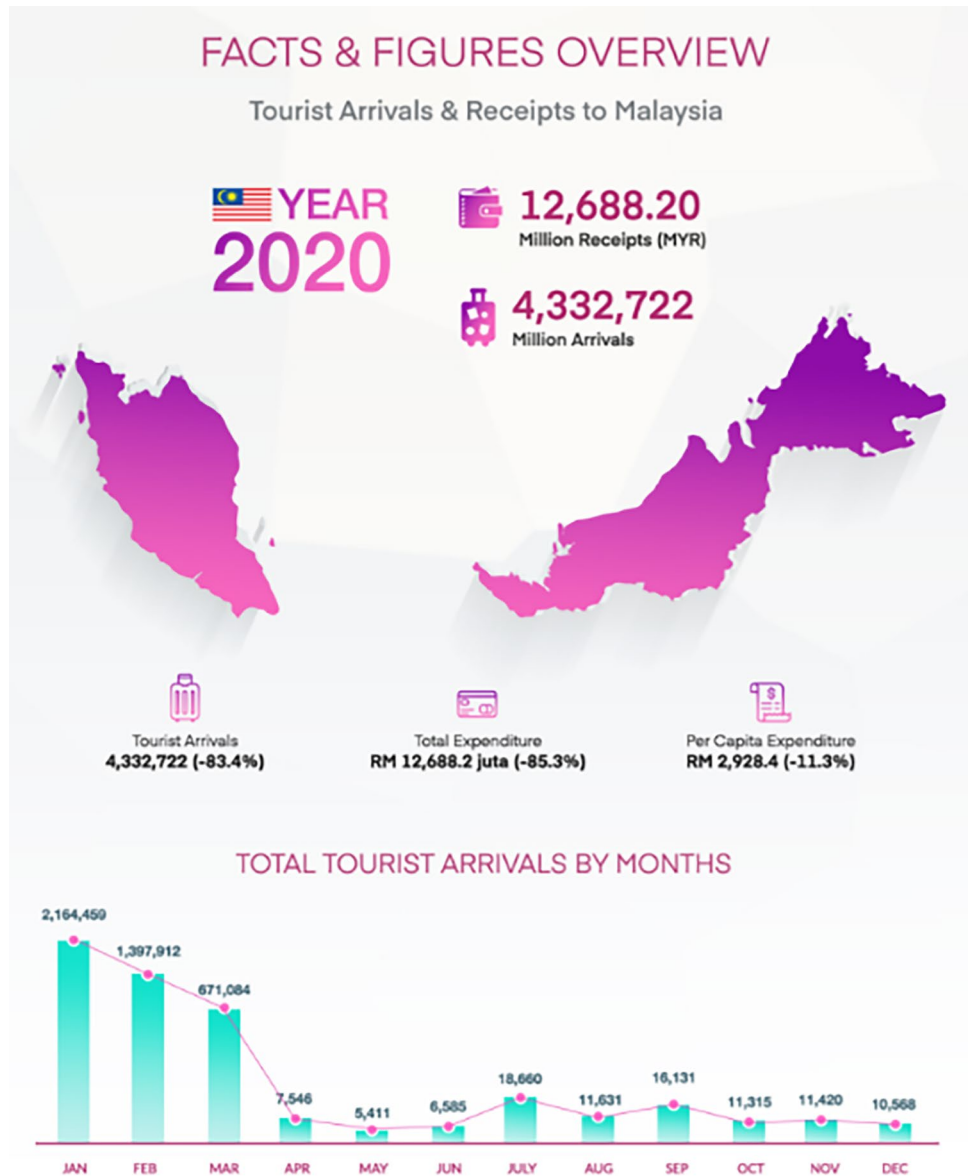
³ Faculty of Information Technology, Inti International University, Persiaran Perdana BBN Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

The worldwide tourism industry has been negatively affected by the Covid-19 pandemic. To handle the number of confirmed cases to prevent it from further increasing, most countries have started to implement the movement control order (MCO) action in order to reduce the mobility of crossing the countries or states within their own countries. Due to this restriction and the fear among the public, the public has to cancel their plans and outdoor activities to protect themselves and the people surrounding them. As a result, the number of tourist arrivals has drastically decreased compared to the one before the pandemic. For instance, in Malaysia, the number of tourist arrivals decreased from 671084 in March 2020 to 7546 in April 2020 as shown in statistics in Fig. 1, which is published on Tourism Malaysia’s official website.

However, there are still a few travel activities being carried out under the MCO, the only difference is the services provided and revenue will be much lower compared to before the pandemic unless the government applies the lock-down action where all entertainment sectors are prohibited to run business. Therefore, the action of travel cancellations has caused profit loss in tourism sectors since the tourism agencies are also being requested for refunds whenever the public wants to cancel their bookings during the pandemic period were made.

In addition, the MCO has decreased the productivity of labor [3]. The situation become worse since the tourism sector is facing the financial issue. Therefore, the adjustment of budget and resource usage, as well as the marketing strategies are needed in order to ensure the

Fig. 1 Statistics of Tourist Arrivals in Malaysia 2020 (Tourism Malaysia Corporate Site, n.d.)



organization is still able to operate with a much lower revenue during the pandemic. Furthermore, a new appropriate strategy enables the tourism sectors to prepare themselves to provide a better and more attractive service for their customers after the end of the outbreak. However, the exact end of the pandemic is difficult to be predicted, causing the tourism sectors will be difficult to determine until when the currently occupied resources are able to support them. Therefore, forecasting tourism activities during the Covid-19 pandemic can be done to gain insight if the sectors need to assign more resources during a specific segment of time.

In past research, the researchers have implemented multiple popular methodologies in forecasting, including the tourist arrivals prediction before and during the pandemic period however, the covid-19 cases have not been included as one of the predictor factors. These models include but are not limited to the linear regression model, Artificial Neural Network Model (ANN) model, Autoregressive Integrated Moving Average (ARIMA) models, Seasonal ARIMA (SARIMA) model, Holt-Winters model, even with hybrid models made up of several models combined by the researchers. In this work, the number of tourist arrivals in ASEAN countries in near future will be predicted by using several time-series predictive models. However, due to the availability of the dataset, only eight out of ten ASEAN countries will be involved in this forecasting work. The involved countries are Cambodia, Indonesia, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam.

Finally, in the past research, the researchers have made a conclusion on the model that has the best performance in time-series forecasting. However, the human behavior that does not follow the restriction rules has caused a fluctuation in the confirmed Covid-19 cases, as a result, the government has further strengthened the restriction level of mobility, eventually making the model lost its accuracy in tourist arrival forecasting. In addition, the countries that have experienced a health pandemic before, for instance, the SARS pandemic, will require less time for tourism recovery since the government and residents know how to handle the pandemic. Therefore, the Covid-19 cases will be included as one of the predictors in the chosen models to be investigated, following the validation, and comparing the model performances, as well as selecting the most accurate model for tourist arrivals forecasting will be done in this work. The main objectives of the proposed work are given below.

- 1) To study the current predictive models used for time-series prediction: The purpose of this research is to study the current predictive models that have been used for time-series predictions. From the models identified in past five to six years of literature works, a minimum of three models will be chosen to carry out the tourism forecasting regarding the scope of this work.
- 2) To study the tourist arrivals trend in ASEAN countries two years before and during the Covid-19 pandemic to obtain future forecast: Besides, this work will investigate the trend of tourist arrivals in eight ASEAN countries in the years 2018 and 2019 (before the pandemic), and year 2020 (during a pandemic). While the future prediction is expected to be obtained from the result of observation.
- 3) To investigate and justify if the ARIMA model is the best model for time-series prediction: Another purpose of this work is to investigate and justify the performance of the ARIMA model if it is able to perform better time-series prediction compared to other chosen models. The comparison will be made by comparing the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) of each model.
- 4) To provide data visualization on tourist arrivals prediction in near future using the most appropriate predictive model: Furthermore, after the model validation, this work is expected to provide data visualization of new tourist arrivals prediction in near future using the predictive model that has the best performance in forecasting tourist arrivals.

The rest of the sections are organized as follows. Section 2 includes the literature survey; proposed work details are included in Sect. 5. Dataset information is included in Sect. 5.3. Results, discussions, and limitations are included in Sect. 1. Finally, concludes the paper with the conclusion in Sect. 2.1.

2 Literature review

In this section, the contents are mainly the reviews and evaluation of the literature papers of past analysis and forecasting work. Firstly, the impact of the Covid-19 pandemic is being justified and followed by the general papers of time-series forecasting in the past. Next, the papers on forecasting in the more specific area of industry, which is the tourism industry before and during the pandemic are studied. The contents that were being looked through are the models used by the researchers and how the models performed compared with each other. Furthermore, the equations and components of three models, which are ARIMA, SARIMA, and Holt-Winters are being studied since these three models are mostly used in time-series forecasting works. Lastly, the performance of the models used in the past and the idea of adding Covid-19 cases as predictors are discussed.

2.1 Impact of covid-19 on tourism and economy

The Tourism industry is one of the assets that help in the country and world development, especially in terms of economy. There are multiple factors that are controlling the tourism demand and growth in each country. These factors can either decide whether the tourism industry will develop in a positive direction or negative direction. Eventually, deciding if a country is able to move forward or backward in the new era, as well as determine if a country is compatible with others. Khan et al. [4] have listed out these factors which include tourist number, quarantine, income level, as well as the fare of the hotel. These factors are able to display the level of development of the country in worldwide. Besides, the tourism industry is also will be affected by unexpected events such as the outbreak of disease. For instance, the Covid-19 pandemic. A detailed survey on COVID-19 and its impact in healthcare domain is studied and in addition, authors studied the worldwide market implementation including the security and privacy issues [5].

In fact, Covid-19 is not the first health pandemic faced by the worldwide. However, it is causing the most serious losses to worldwide tourism and the economy [6, 7]. According to the research, the time taken for tourism industry recovery due to Covid-19 is much longer than the historical health pandemic i.e. SARS, H1N1 since Covid-19 outbreak brings greater destruction and revenue losses [8]. Moreover, most of the countries have less experience in handling the health pandemic in the past, such as Thailand and Malaysia, while Taiwan and Hong Kong are the countries that experienced the SARS epidemic which allowed the governments to have a faster response in taking the more effective precautional actions, resulting these two countries have a much smaller impact in the tourism industry in terms of travel restrictions [9].

Apart from that, Alwi et al. [10] justified the sudden decrease of tourist arrivals is due to the travel cancellations that are triggered by the movement control orders that the government of some countries implemented starting from the earlier phase of Covid-19 outbreak. The cancellation of the travel accompanied by the action of requesting a refund by the travellers has further increased the financial and the economy losses in the tourism industry.

At the same time, through research done by Yang et al. [3], the researchers justified there is a relationship between the Covid-19 cases, government travel restriction orders, tourism activity, and employee's health status and productivity. Travelling has caused a larger area of spreading the Covid-19 virus, causing the government to enforce travel restrictions, affecting the health of employees at the same time lowering the tourism activities and services due to lower productivity from the employees. Apart from that, the tourism demand has decreased due to this pandemic since

the traveling frequency of residents is being controlled by the lockdown actions and hence resulting in a decrease in demand price. This statement is supported by the study done by Bakar and Rosbi [11].

From the economic perspective, most of the organizations are facing the issue of maintaining the operations either externally or internally. This problem is more obvious in a small organization. For instance, the small standard hoteliers are within a situation where they are unable to receive their loan to purchase the properties and assets for business operations during the pandemic that did not show symptoms of getting better [12]. Therefore, to ensure the business to continue to run as usual during the pandemic, the businesses should think about how to preserve the current assets until can be used until and after the outbreak instead of attracting more customers. This statement is supported by Abbas et al. [12] that the researchers indicated the tourism sectors should provide better, more comfortable, and affordable services.

2.2 Time-series analysis

Time-series analysis also done using multiple popular models in past research in different industries. For instance, Casini and Rocchetti [13] have investigated the relationship between number of tourists and the spread of Covid-19 virus in Italian during summer through a cross regional analysis. The analysis is done by using three models which are simple linear model, negative binomial regression model, and the Artificial Neural Network (ANN) cognitive model. The result of the analysis has shown that the reason of number of Covid-19 cases increased is highly affected by the tourism activity. While the ANN model given the most capability in prediction compared to the other two models. Besides, there is another researched done by Singh et al. [14], the researchers has implement ARIMA model in forecasting the Covid-19 cases in Malaysia. The research shows that ARIMA (0,1,0) model is the most accurate model for prediction the number of confirmed cases with the lowest MAPE compared to other ARIMA models.

On the other hand, more hybrid models were tested and obtained the results saying the hybrid model is more accurate than the plain models in forecasting works in multiple industries. In financial markets forecasting, time-series analysis also being implemented. Khashei et al. [15] carried out a comparison of four types of ARIMA models which are ARIMA, Fuzzy Auto-Regressive Integrated Moving Average (FARIMA), Fuzzy ANN (FANN), and Hybrid Fuzzy Auto-Regressive Integrated Moving Average (FARIMAH) in financial markets forecasting. From the research, ARIMA model is considered having the lowest capabilities in forecasting, while FANN having the best performance among the used models.

Apart from that, another hybrid models also implemented in a video traffic forecasting work [16], the researchers tried on improving the prediction by using FARIMA-based models. Throughout the research, the authors successfully improve the accuracy and stated that hybrid FARIMA/GARCH-MLP model is further better than plain FARIMA model. A detailed analysis of explainable deep learning approach is done for COVID-19 cases prediction across different states in India [17].

2.3 Tourism industry and forecasting

Forecasting is important in each industry including tourism industry. Either it is during the period of before or during the pandemic, a good prediction tool helps the management in a better decision making and construct the new policy and marketing strategies for allowing the organization could be last long. As mentioned by Abbas et al. [12], during his pandemic environment, the tourism sectors should not aim for increasing the number of customers but should aim on providing to their customers a better, comfortable, and affordable services. The forecast results still also allowed the organization to prepare the precautionary actions to protect the customers from the coronavirus. In this section, the tourism related forecasting research before the outbreak of Covid-19 will be evaluated, followed by the research done during the pandemic.

2.3.1 Before covid-19

Earlier in 2019, the ARIMA model and Holt-Winter model are used by Purwanto et al. [18] in a study of tourist arrival forecasting in Indonesia. At the same time, a linear trend model and a hybrid model (combination of ARIMA model and linear trend model) were used by the researchers in the prediction of tourist arrivals. The comparison of four models was done and showed that the hybrid model has the lowest Root Mean Square Error (RMSE) value compared to the other three models.

Furthermore, the SARIMA–GARCH hybrid model was developed in tourism demand forecasting in Taiwan [19]. The researcher compared the prediction results SARIMA–GARCH model with the results predicted by the other four methodologies. These models are the regression analysis, exponential smoothing, Holt-winter exponential smoothing, as well as the back-propagation neural network with a genetic algorithm approach. While the results have shown that the SARIMA–GARCH hybrid model has the best prediction performance.

Despite this, the hybrid model is not always the better methodology compared to its plain model in time-series forecasting. The hybrid model, ARIMA-GARCH also has been used by Chhorn and Chaiboonsri in tourist arrivals

forecasting in Cambodia (2017). The plain ARIMA and GARCH models were also applied to do the accuracy comparison. Based on the RMSE measurements of the prediction results, the ARIMA and ARIMA-GARCH models are the best two models among the five models implemented. However, in this research, the plain ARIMA model has slightly higher accuracy compared to the hybrid ARIMA-GARCH model, with a difference of 0.0003 RMSE value. This comparison is different from the time-series works that applying and comparing the hybrid model and its plain model where the results show the hybrid models have better prediction performance.

Besides, Lip et al. [20] carried out a prediction on tourist arrivals from three countries in Malaysia from the year 2013 to the year 2017 by using the ARIMA model, more specifically, the Box-Jenkins SARIMA models and Holt-Winters model to investigate which model is better in prediction performance. The results of the investigation proved that the Holt-Winters model is suitable for time-series prediction of tourist arrivals. However, it is only suitable for forecasting the tourist arrivals from United States and Korea. While the SARIMA model is suitable for forecasting the tourist arrivals from Korea. The researchers claimed that unpredicted factors such as political and natural events are one of the reasons that increase the difficulty to obtain the most accurate value of prediction. Besides, the researchers suggested testing the tourist arrival forecasting with the implementation of other ARIMA models such as the Fuzzy Seasonal ARIMA model (FSARIMA) in the future study.

Time-series analysis is also being used in forecasting tourism demand. In the previous research, both the plain model and hybrid model have been implemented in different works since the past. Firstly, the seasonal ARIMA model also known as the SARIMA model is being applied in tourism demand forecasting in Malaysia [20]. The result of the research showed that the ARIMA model without the seasonal effect is the most suitable model for predicting tourist arrivals from ASEAN.

Moreover, [21] whom one of the researchers applied the ARIMA model in tourism demand forecasting in Macedonia, says that even though the ARIMA model is considered as good, valid, and reliable, the accuracy is not high enough in the study. The researcher clarified this could be due to the structural breaks in data as the result of unexpected events such as tourist behaviors, as well as economic shocks.

Last but not least, in another study of United States' tourist arrivals forecasting [22], three models known as paired Neural Network with HP filter (pNN-HP), paired Neural network with Wavelet Transformation (pNN-WT) and with Moving Average (pNN-MA), which under the family of pair of Neural Network (pNN) are constructed and compared with the other three traditional models which are ARIMA, SARIMA, and

ARFIMA. This study has shown that the pNN models have higher accuracy compared to the traditional models.

2.3.2 During covid-19

Qiu et al. [23] conducted tourist arrival prediction of 20 countries by using 11 single predictive models and 26 stacking models. The results of the study clearly highlighted that the best three time series analysis and predictive models with the lowest MASE values are the SARIMA model, ETS model, and STL model among the single models. Certainly, stacking models, the combination of multiple methods as the input of second-level algorithms as the new models, will have higher accuracy compared to single models. However, the researchers justified that the optimal number of combinations for the stacking model is five single models in the study.

Apart from that, a forecasting using rolling-window approach has been used in tourism recovery prediction work for 16 countries. The researchers applied the Generalised Additive Model (GAM) linear and auto-regressive model, and the Long Short-Term Memory (LSTM) neural network [24]. The researchers indicate that there are limitations in their prediction work. The first limitation is the researchers did not include the data of tourism revenue that was contributed by local travellers, while another limitation is the model did not include the data of Covid-19 cases after the actions taken by the government to minimize the spread of the Covid-19 virus. Therefore, the research was having the problem of lack of data. Hence, other information such as the number of Covid-19 cases could be included to obtain a better prediction, which is also mentioned by the authors in the research paper, along with other consideration factors, for instance, the hospitalizations, the number of deaths, and the number of populations received the vaccinations.

Christidis and Christodoulou [25] carried out a prediction on the potential spread of Covid-19 from China connected with the infected travellers. As one of the parts of the research, they implemented the ARIMA model to predict the number of travellers who will leave Wuhan, China in January 2020. The authors indicated that the model is considered to have a high accuracy, however, based on the provided graph of estimations and real data, there are some estimation lines that did not have a similar pattern to the real data. This situation happened could be due to the lack of data, one of the factors mentioned also by the authors. Other factors stated by the authors include consisting of the same traveller travelling to different destinations. This means there are multiple duplicate data of travelling records made by the same person.

2.4 Performance of models in past research

Based on the observation of the past time-series works, the most frequently models used by the researchers in general time-series analysis are the ARIMA, SARIMA, Holt-Winters, and neural network models. At the same time, some of the authors proposed the hybrid models which are the combination of multiple plain models. As the results, most of the research indicated that the hybrid models are having higher accuracy compared to the plain model. However, some of the researchers obtained results saying the plain model is better than the hybrid model. There are different voices from the research results supporting the respective models. On the other hand, the time-series analysis in tourism forecasting works mostly applied the ARIMA model and SARIMA model, as well as the hybrid models. Similarly, there are researchers saying that the hybrid model has the best performance, while there is another saying ARIMA is better. This situation could be due to several considerations on model application. For instance, the type of model chosen by the researcher in the studied area where either the seasonal model or non-seasonal model should be used, and the number of relevant and significant predictors included in the model. Moreover, there is less relevant past tourist arrivals forecasting is done during the Covid-19 pandemic by including the Covid-19 cases as one of the predictors. Therefore, this work is going to develop a model with Covid-19 cases and justify if there will be an improvement occurred.

In conclusion, the impact of the Covid-19 pandemic on the tourism industry and economy has been looked through where the main negative effect is the loss of revenue of the business and unable to maintain operations as the outbreak is continuing for a long period. The travellers are cancelling their travel plans and requesting for a refund, as well as the decreased productivity of labour also causing the tourism sectors must reduce their services to ensure the health of people during the pandemic. Furthermore, several general time-series analysis in the past has been reviewed as well as further reviewed the models used in multiple industries. There are some models that are popular used by researchers in multiple industries including tourist arrivals forecasting in the tourism industry, where including but not limited to the ARIMA, SARIMA, Holt-Winters, and ANN. Most of the models are validated by identifying and comparing the RMSE of the models, where the lower the RMSE, the more accurate the model performed. Different researchers have come out with different results where some plain models are better in some analysis work, while some hybrid models performed better in other works. Therefore, the outcome and performance of the prediction are affected by the number and the significance of the predictors, as well as affected by external unexpected

events, such as economic shocks, the outbreak such as the current Covid-19 pandemic, and more. On the other hand, currently, there is fewer tourist arrivals work that include the Covid-19 cases in the prediction, which will be done in this work using the ARIMA, SARIMA, and Holt-Winters models.

3 Methodology: Time-series analysis models

3.1 ARIMA

The Auto-Regressive Integrated Moving Average (ARIMA) model is also known as Box-Jenkins methodology can be represented as the equation below (Khashei, Montazeri, & Bijari, 2015) [15]:

$$y_t = \vartheta_0 + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_y - \vartheta_1 \varepsilon_{t-1} - \dots - \vartheta_q \varepsilon_{t-q} \tag{1}$$

Khashei et al. [15] indicated the y_t is the actual value and ε_t represents as the random error at a time period, t . While θ and φ are the parameters of the mode that, and p and q are representing the autoregressive and the moving average orders of the model [20].

Khashei et al. [15] also explained there are three processes in the methodology. The first process is model identification which identifies if there are theoretical autocorrelation properties that can indicate if the time-series is developed by using ARIMA. The researcher also mentioned that Box and Jenkins, the developers of the methodology proposed two tools which are the autocorrelation function (ACF) and the partial autocorrelation function (PACF). These tools will help to identify the order of the model. The next process is parameter estimation which is done to minimize the measurement errors by estimating the parameters. While the last process is the diagnostic checking to check if the model is adequate. If the model has low adequacy, the processes should be repeated to identify another model.

3.2 SARIMA

Seasonal ARIMA (SARIMA) model is similar to ARIMA model but includes the seasonality component in time-series forecasting [26]. Velos et al. [26] explained that seasonality refers to the repeated pattern of the value changes in specific time intervals for instance the tourist arrivals in several countries that show the seasonal patterns (Baldigara and Mamula 2015), cited in Velos et al. [26].

The equations below representing the seasonal component $(P,D,Q)_m$ is directly added onto the ARIMA model by the multiplication, where (p,d,q) is the non-seasonal component and m refers to the number of observations in a period or cycle [27].

$$ARIMA(p,d,q)(P,D,Q)_m \tag{2}$$

The SARIMA model needs the backshift operator which is represented as B as shown in the equation below, which is the way of writing $ARIMA(p,d,q)(P,D,Q)_4$ without the constant [27].

$$(1 - \phi_1 B)(1 - \phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 + \vartheta_1 B)(1 + \Theta_1 B^4)\varepsilon_t \tag{3}$$

3.3 Holt-winters

Holt-Winters methodology is suitable for seasonal time-series forecasting [28]. It involves four equations, one is the forecast equation and the other three are smoothing equations for level, trend, and seasonal components three smoothing components α , β , and γ where the values of smoothing components are only between 0 and 1 [28]. Lip et al. [28] also did mention the Holt-Winters model introduces two types of methods which are the additive method and multiplicative method for each component. Furthermore, based on the research paper by Kurniasih et al. [29] two methods apply different equations as shown as below.

The equations for Holt-Winters Addictive Methods:
Forecast:

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \tag{4}$$

Level:

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \tag{5}$$

Trend:

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \tag{6}$$

Seasonal:

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s} \tag{7}$$

The equations for Holt-Winters Multiplicative Methods:
Forecast:

$$F_{t+m} = (L_t - b_t m)S_{t-s+m} \tag{8}$$

Level:

$$L_t = \alpha\left(\frac{Y_t}{S_t}\right) + (1 - \alpha)(L_{t-1} + b_{t-1}) \tag{9}$$

Trend:

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \tag{10}$$

Seasonal:

$$S_t = \gamma\left(\frac{Y_t}{L_t}\right) + (1 - \gamma)S_{t-s} \tag{11}$$

4 Description of datasets

The use of secondary sources will be the data collection method in this study. Two categories of the dataset will be used throughout this study, which are the dataset on the daily Coronavirus cases number and tourist arrivals in chosen ASEAN countries. Both datasets are obtained from the Internet. Coronavirus cases dataset is the open-source dataset published in GitHub which is prepared by the Center for Systems Science Engineering (CSSE) of Johns Hopkins University [30] recording the daily number of cases confirmed, deaths, recovery, and other variables. While the tourist arrivals dataset is obtained from the Trading Economics website which needs to purchase for study usage. Both datasets support the quantitative analysis in this study which can be easily processed for mathematical and statistical modelling. Figure 2 shows the webpage of one of the data sources since the author needs data for eight countries.

Based on the user requirement, the interviewee would like to know who their target market is, in order to prepare services based on their culture, in order to protect the reputation of the company. The data can be obtained from Google's Destination Insights under the Origins of Demand section. A sample data visualization is shown in Figs. 3 and 4.

The data understanding process is a stage of analysts clarifying the attributes within original dataset provided by the data publisher. This is an essential process whereas analysts will have to decide which attribute would be used for

analysis and remove the attributes which are not during data pre-processing for a more effective analysis.

The tourist arrivals dataset which will be used in time-series forecasting consists of the monthly tourist arrivals records for years, whereas there is total of 1831 records from a combination of eight countries' CSV files and a total of 7 attributes in each file which are Country, Category, DateTime, Value, Frequency, HistoricalDataSymbol, LastUpdate. A sample dataset is shown in Figs. 5, 6, and 7.

While target market dataset is made up of eight CSV files as well, with a total of 50 rows of data and 2 attributes which are Location and Rank in each. While The Covid-19 repository consists of 578 CSV files and the data is collected from 22 January 2020 to 21 August 2020. Each file consists of 3963 rows of data with 14 attributes which are FIPS, Admin2, Province_State, Country_Region, Last_Update, Latitude, Longitude, Confirmed, Deaths, Recovered, Active, Combined_Key, Incidence_Rate, Case-Fatality_Ratio. These two datasets will be processed and used in dashboard development for descriptive analysis.

The process of predictive model development is an iterative stage that includes the process of trial and error. To test the validation of the model, the usage of train and test datasets is needed to test the model repeatedly before the actual forecasting. The train dataset will be used to train the model for prediction and compare the results with the test dataset which acts as actual data.

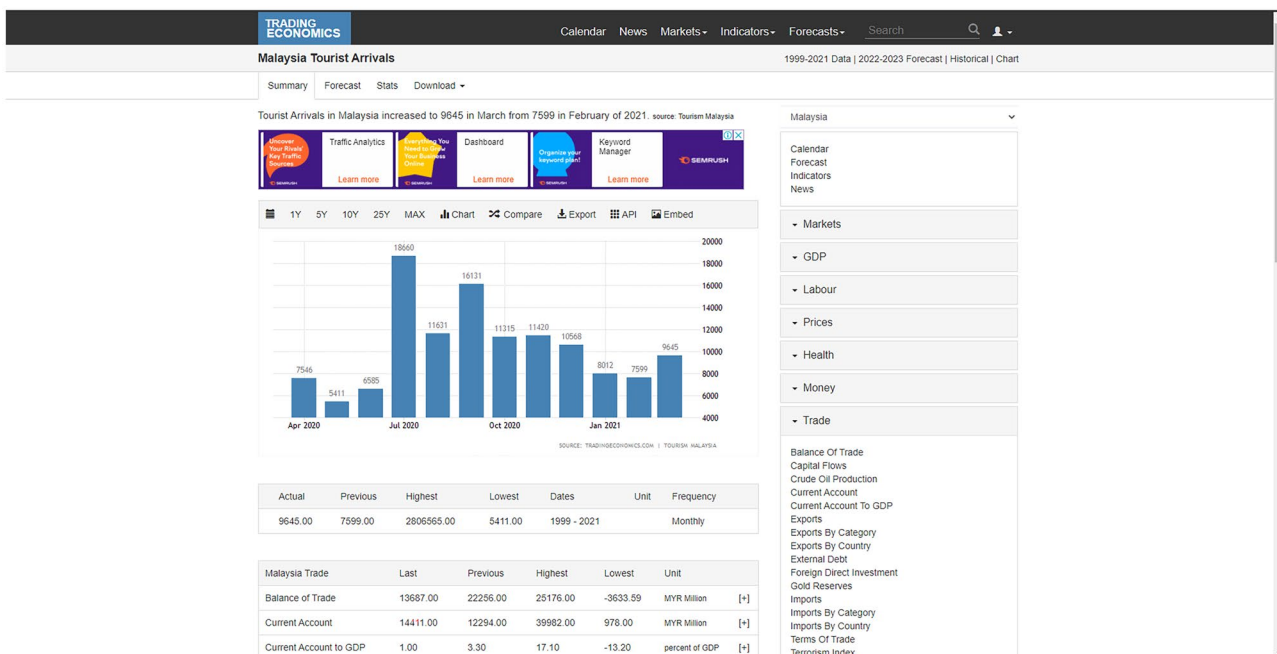


Fig. 2 Trading Economics Dataset Extraction Sample Interface

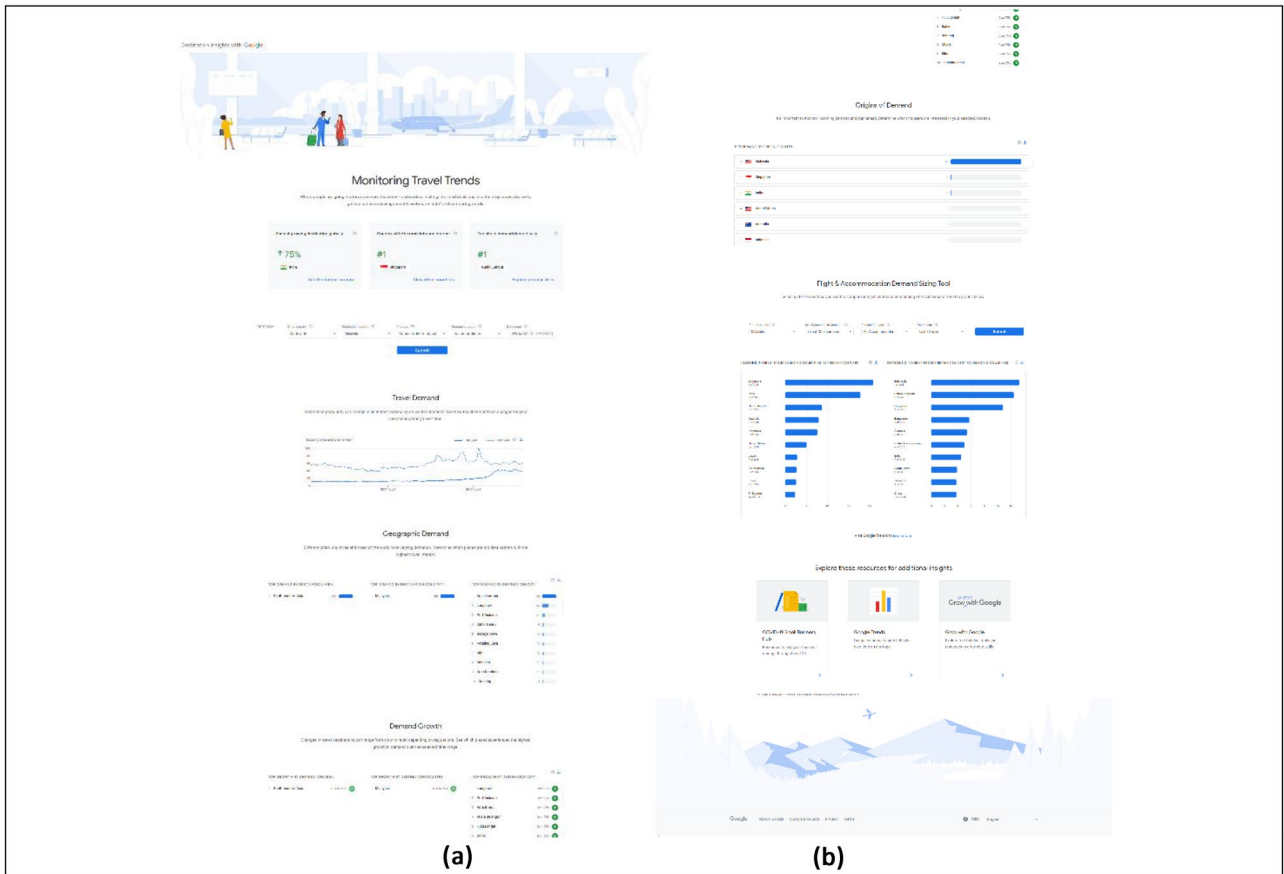


Fig. 3 Destination Insights with Google Interface

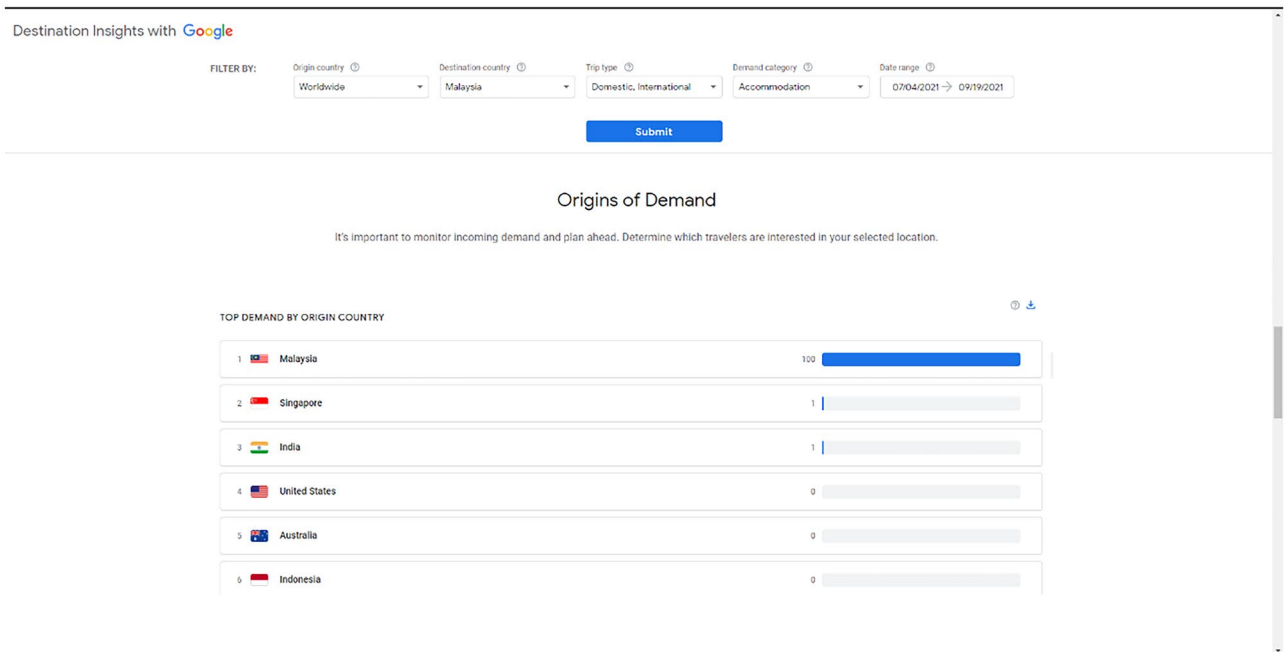


Fig. 4 Required Dataset Section from Google's Destination Insights

AutoSave Off historical_country_Cambodia_indicator_Tourist_Arrivals.csv											
File Home Insert Draw Page Layout Formulas Data Review View Help Data Minin											
A1 Country											
	A	B	C	D	E	F	G	H	I	J	K
1	Country	Category	DateTime	Value	Frequenc	Historical	LastUpdate				
2	Cambodia	Tourist Ar	2012-12-31	382326	Monthly	CAMBODI	2015-05-19T16:17:00				
3	Cambodia	Tourist Ar	2013-06-30	301412	Monthly	CAMBODI	2015-05-19T16:17:00				
4	Cambodia	Tourist Ar	2013-07-31	338761	Monthly	CAMBODI	2015-05-19T16:17:00				
5	Cambodia	Tourist Ar	2013-08-31	342064	Monthly	CAMBODI	2015-05-19T16:17:00				
6	Cambodia	Tourist Ar	2013-09-30	283787	Monthly	CAMBODI	2015-05-19T16:17:00				
7	Cambodia	Tourist Ar	2013-10-31	334410	Monthly	CAMBODI	2015-05-19T16:17:00				
8	Cambodia	Tourist Ar	2013-11-30	386737	Monthly	CAMBODI	2015-05-19T16:17:00				
9	Cambodia	Tourist Ar	2013-12-31	431807	Monthly	CAMBODI	2015-05-19T16:17:00				
10	Cambodia	Tourist Ar	2014-01-31	442045	Monthly	CAMBODI	2015-05-19T16:17:00				
11	Cambodia	Tourist Ar	2014-02-29	425801	Monthly	CAMBODI	2015-05-19T16:17:00				
12	Cambodia	Tourist Ar	2014-03-31	400076	Monthly	CAMBODI	2015-05-19T16:17:00				
13	Cambodia	Tourist Ar	2014-04-30	332690	Monthly	CAMBODI	2015-05-19T16:17:00				
14	Cambodia	Tourist Ar	2014-05-31	300302	Monthly	CAMBODI	2015-05-19T16:17:00				
15	Cambodia	Tourist Ar	2014-06-30	300454	Monthly	CAMBODI	2015-05-19T16:17:00				
16	Cambodia	Tourist Ar	2014-07-31	340091	Monthly	CAMBODI	2015-05-19T16:17:00				
17	Cambodia	Tourist Ar	2014-08-31	347211	Monthly	CAMBODI	2015-05-19T16:17:00				
18	Cambodia	Tourist Ar	2014-09-30	311388	Monthly	CAMBODI	2015-05-19T16:17:00				
19	Cambodia	Tourist Ar	2014-10-31	390637	Monthly	CAMBODI	2015-05-19T16:17:00				
20	Cambodia	Tourist Ar	2014-11-30	411501	Monthly	CAMBODI	2015-05-19T16:17:00				
21	Cambodia	Tourist Ar	2014-12-31	500579	Monthly	CAMBODI	2015-05-19T16:17:00				
22	Cambodia	Tourist Ar	2015-01-31	460577	Monthly	CAMBODI	2015-06-05T10:11:00				
23	Cambodia	Tourist Ar	2015-02-29	430207	Monthly	CAMBODI	2015-06-05T10:11:00				
24	Cambodia	Tourist Ar	2015-03-31	417052	Monthly	CAMBODI	2015-07-16T16:47:00				
25	Cambodia	Tourist Ar	2015-04-30	361139	Monthly	CAMBODI	2015-08-10T18:26:00				
26	Cambodia	Tourist Ar	2015-05-31	314748	Monthly	CAMBODI	2015-09-04T18:55:00				
27	Cambodia	Tourist Ar	2015-06-30	318267	Monthly	CAMBODI	2015-10-27T19:09:00				
28	Cambodia	Tourist Ar	2015-07-31	364325	Monthly	CAMBODI	2015-11-25T07:16:00				
29	Cambodia	Tourist Ar	2015-08-31	366096	Monthly	CAMBODI	2016-01-27T18:27:00				
30	Cambodia	Tourist Ar	2015-09-30	314459	Monthly	CAMBODI	2016-02-03T03:58:00				
31	Cambodia	Tourist Ar	2015-10-31	408922	Monthly	CAMBODI	2016-02-17T09:19:00				
32	Cambodia	Tourist Ar	2015-11-30	444640	Monthly	CAMBODI	2016-03-24T03:55:00				
33	Cambodia	Tourist Ar	2015-12-31	574799	Monthly	CAMBODI	2016-05-08T12:14:00				
34	Cambodia	Tourist Ar	2016-01-31	466086	Monthly	CAMBODI	2016-05-08T12:31:00				
35	Cambodia	Tourist Ar	2016-02-29	448468	Monthly	CAMBODI	2016-06-07T11:55:00				
36	Cambodia	Tourist Ar	2016-03-31	427923	Monthly	CAMBODI	2016-07-20T16:27:00				
37	Cambodia	Tourist Ar	2016-04-30	367684	Monthly	CAMBODI	2016-07-20T16:27:00				
38	Cambodia	Tourist Ar	2016-05-31	320601	Monthly	CAMBODI	2016-08-05T17:51:00				
39	Cambodia	Tourist Ar	2016-06-30	330170	Monthly	CAMBODI	2016-12-22T13:15:00				
40	Cambodia	Tourist Ar	2016-07-31	395761	Monthly	CAMBODI	2016-12-22T13:14:00				
41	Cambodia	Tourist Ar	2016-08-31	406214	Monthly	CAMBODI	2016-12-22T13:14:00				
42	Cambodia	Tourist Ar	2016-09-30	345508	Monthly	CAMBODI	2016-12-22T13:14:00				
43	Cambodia	Tourist Ar	2016-10-31	414077	Monthly	CAMBODI	2017-03-22T16:08:00				

Fig. 5 Cambodia's Tourist Arrival Dataset

	A	B	C	D	E	F	G	H	I	J	K	L
1	location	rank										
2	Indonesia	100										
3	Australia	7										
4	Singapore	2										
5	India	2										
6	United States	1										
7	United Kingdom	1										
8	Malaysia	0										
9	Germany	0										
10	Netherlands	0										
11	France	0										
12	Japan	0										
13	New Zealand	0										
14	Canada	0										
15	South Africa	0										
16	Philippines	0										
17	Hong Kong	0										
18	Belgium	0										
19	Thailand	0										
20	Spain	0										
21	Taiwan	0										
22	Russia	0										
23	Saudi Arabia	0										
24	Ukraine	0										
25	Italy	0										
26	Denmark	0										
27	United Arab Emirates	0										
28	South Korea	0										
29	Switzerland	0										
30	Sweden	0										
31	Vietnam	0										
32	Turkey	0										
33	Poland	0										
34	Austria	0										
35	Portugal	0										
36	Romania	0										
37	Brazil	0										
38	Egypt	0										
39	Bangladesh	0										
40	Iran	0										
41	Sri Lanka	0										
42	Norway	0										
43	Ireland	0										

Fig. 6 Indonesia’s Origin of Demand Dataset

Fig. 7 Sample Covid-19 Dataset

In this work, the dataset is split into the ratio of 80% as the train dataset and 20% as the test dataset. Apart from that, since the number of records for different countries is different in size, the dataset is split by multiplying 80% by the size

of data extracted based on the investigated country to obtain the size of train data. The train data is then assigned with records from 0 index to the index with the value of train size, while the rest will be the test dataset.

Model Type	ARIMA				SARIMA				Holt-winters				Best Choice
	MAE	MAPE	MSE	RMSE	MAE	MAPE	MSE	RMSE	MAE	MAPE	MSE	RMSE	
Cambodia	(13,1,0)				(0,1,1)x(0,1,12)12								SARIMA
Order	45559.631	7.52772	3.194E+09	56516.031	13837.852	2.4459957	408374403	20208.276	41019.892	7.2482455	3.296E+09	57411.363	
Indonesia	(30,2,1)				(0,1,1)x(0,1,12)12								SARIMA
Order	126877.9	9.2436778	2.576E+10	160489.88	68520.486	5.1351503	6.247E+09	79040.416	156296.79	11.084716	3.975E+10	199382.07	
Malaysia	(11,1,6)				(0,1,1)x(1,1,12)12								SARIMA
Order	118036.08	5.2746107	2.146E+10	146505.59	82821.35	3.8157419	1.269E+10	112662.38	227931.86	10.382257	8.747E+10	295753.12	
Myanmar	(30,2,6)				(0,1,1)x(0,1,12)12								SARIMA
Order	25439.139	7.340931	933565577	30554.305	13993.622	4.2428587	309447144	17591.11	43137.312	12.286383	2.403E+09	49024.887	
Philippines	(13,0,12)				(1,1,1)x(0,1,12)12								SARIMA
Order	139876.14	23.76326	2.916E+10	170764.5	19863.946	3.6565234	702623578	26507.048	68602.401	12.394492	7.876E+09	88745	
Singapore	(12,1,10)				(0,1,1)x(1,1,12)12								SARIMA
Order	90296.592	6.4327621	1.268E+10	112602.32	40624.825	3.0826359	2.566E+09	50659.127	137320.29	9.4504525	2.971E+10	172358.4	
Thailand	(12,1,12)				(0,0,1)x(0,1,12)12								SARIMA
Order	332311.36	10.660164	1.603E+11	400421.62	167431.3	5.8005088	4.331E+10	208109.17	370439.23	11.944513	2.034E+11	450988.83	
Vietnam	(16,1,8)				(1,1,1)x(1,1,0,12)12								SARIMA
Order	3207.9641	216.08908	14074775	3751.6364	70.581572	4.7071667	9186.0505	95.843886	93.290867	6.2417158	13703.649	117.06258	

Fig. 8 MAE, MAPE, MSE, and RMSE Values of Each Model by Country Using Data Before Pandemic

Model Type		SARIMA			
Country		MAE	MAPE	MSE	RMSE
Cambodia	Order	(0,1,1)x(0,1,1,12)12			
		611243.01	14.524923	4.354E+11	659821.36
Indonesia	Order	(0,1,1,x(0,1,1,12)12			
		2470943	9.1906631	8.117E+12	2849040.6
Malaysia	Order	(0,1,1)x(1,1,1,12)12			
		368863.59	17.755955	2.219E+11	471029.24
Myanmar	Order	(0,1,1)x(0,1,1,12)12			
		7839354.2	39.779689	2.398E+14	15484608
Philippines	Order	(1,1,1)x(0,1,1,12)12			
		3444726.5	568.87831	2.862E+13	5349402.1
Singapore	Order	(0,1,1)x(1,1,1,12)12			
		1022146.7	29.872001	1.345E+12	1159846.2
Thailand	Order	(0,0,1)x(0,1,1,12)12			
		3841006.9	486.7521	3.645E+13	6037296.1
Vietnam	Order	(1,1,1)x(1,1,0,12)12			
		2073.6865	85.328413	5218692.4	2284.4457

Fig. 9 MAE, MAPE, MSE, and RMSE Values of SARIMA Model with Same Order Values Using Data During Covid-19 Pandemic

5 Results, discussions, limitations, and future works

The statistical models are implemented using statsmodels in python programming. As for data visualization for the historical and final forecasting results after processing data using Python in PyCharm, Tableau is selected as the tool to perform the last stage of presenting data to end users.

In this section, the author discusses on the accuracy of the models in terms of MAE, MAPE, MSE, and RMSE values, which are calculated from the difference between prediction values using developed models and actual values from test dataset. It is crucial to validate the accuracy of the model in order to provide answers that have reliability as high as possible, so that clients can achieve their business goals based on the results given by models. While MAE, MAPE, MSE, and RMSE are part of the popular metrics that are being used to evaluate the performance of the model. Results tabulated in Fig. 8 below are the metric values of each model by the country that were calculated from the results generated during the model training stage.

From the figure above, the author notices that the SARIMA model performed better than ARIMA and Holt-Winters models using the tourist arrivals dataset before the Covid-19 pandemic for each country. Four metric values of the SARIMA model are obviously lower than ARIMA and Holt-Winters models. Therefore, the SARIMA model is selected as the best model in tourist arrivals forecasting.

Next, while applying the SARIMA model with the same order for respective countries using a dataset with a time index starting from 2018, most of the metrics show higher results as shown in Figs. 9, 10, and 11.

The author then reruns the codes for identifying the order values based on the lowest AIC values and assigns the identified order values. The results are shown in Fig. 12.

Based on the results as shown in Fig. 12, most of the sub-models have a better performance compared to which using same order value to forecast tourist arrivals assuming there is no pandemic. The results of forecasting during pandemic period using new orders are as shown in Figs. 13 and 14.

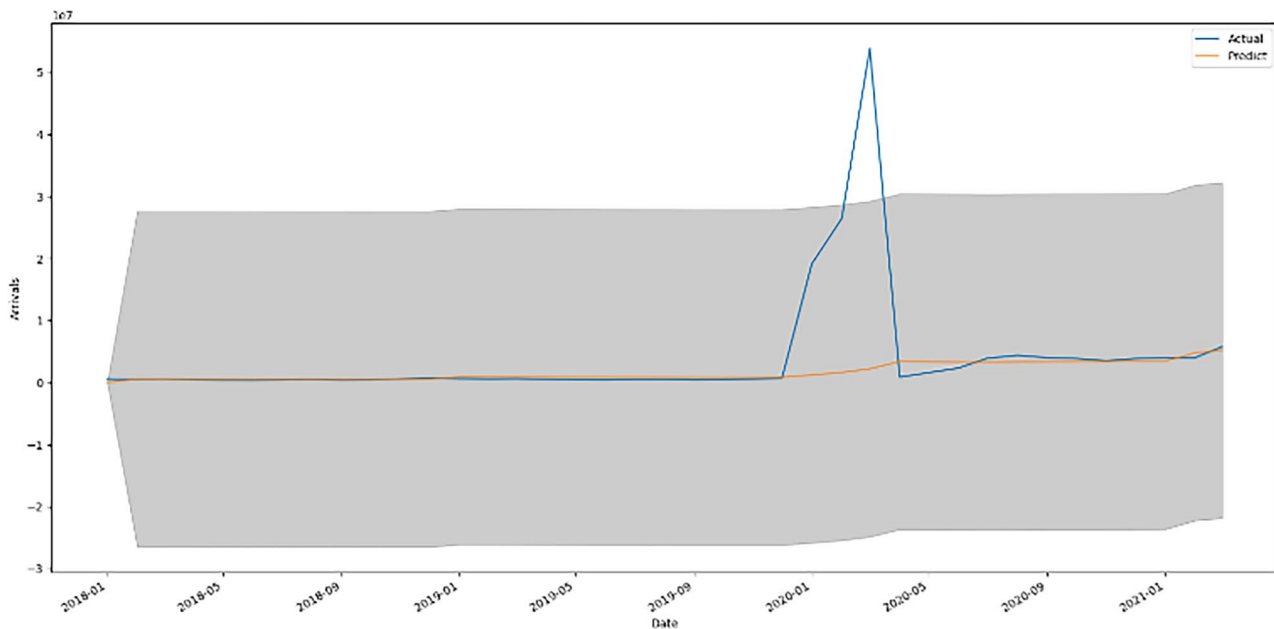


Fig. 10 Prediction and Actual Values using dataset year 2018-latest (Cambodia) – Same SARIMA Order Values

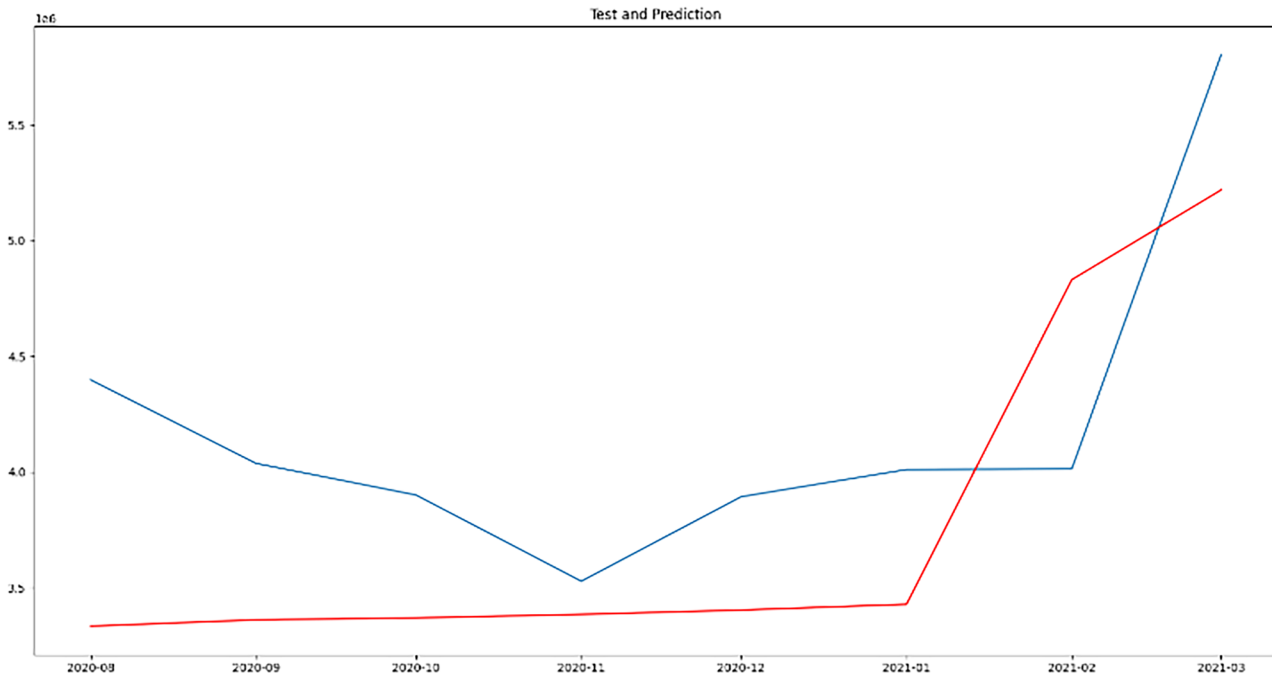


Fig. 11 Prediction and Actual Values (Aug 2020 – Mar 2021) using dataset year 2018-latest (Cambodia) – Same SARIMA Order Values

5.1 Data visualization

After developing the predictive model and forecasting future tourist arrivals value using the SARIMA model, the author moves to the last stage which is data visualization. Even though data visualization can be done using Python in PyCharm IDE, the outcome of this work is to display the forecasting results to any parties in the tourism industry, which means the target audience could lack of programming knowledge. Therefore, data visualization tools such are needed to act as intervals for end users to communicate with the datasets. Therefore, Tableau’s applications are the tools that are being used to present the forecasting and historical data as a dashboard to the target audience of this work.

With the assistance of Tableau Desktop Professional Edition, the dashboard can be created easily using the drag and drop method. The dashboard created in Tableau Desktop can be accessed by anyone by saving the workbook to the Desktop Public website.

5.1.1 Removing potential outliers

Referring to Figs. 15, 16, 17, and 18, there are forecasting values for the Covid-19 condition that lies below zero or much higher than the values predicted assuming no pandemic, which can be the outliers of the data. Therefore, the author eliminates some of the values by selecting and excluding the data point without making much of changes to the forecasting pattern.

5.2 Dashboard

In the section are the interfaces of the current dashboard created to be published to Tableau Public so that everyone can access the dashboard. The URL address of the dashboard on Tableau Public is <https://public.tableau.com/app/profile/tan.zhi.xuan/viz/TouristArrivalsForecastingDashboardBasedonCovid-19/Main>. The workbook also can be downloaded from Tableau

Model Type		SARIMA			
Country	Order	MAE	MAPE	MSE	RMSE
Cambodia	Order	(0,1,0)x(1,1,1,12)12			
		361020.9	8.8253005	1.907E+11	436721.26
Indonesia	Order	(0,1,1)x(0,1,1,12)12			
		2470943	9.1906631	8.117E+12	2849040.6
Malaysia	Order	(0,1,1)x(0,1,1,12)12			
		469505.09	23.969392	3.476E+11	589557.99
Myanmar	Order	(0,1,1)x(0,1,1,12)12			
		7839354.2	39.779689	2.398E+14	15484608
Philippines	Order	(0,1,1)x(0,1,1,12)12			
		1004018	103.48999	1.378E+12	1174039.2
Singapore	Order	(1,1,1)x(0,1,1,12)12			
		844656.67	24.626715	1.01E+12	1005036.9
Thailand	Order	(0,1,1)x(0,1,1,12)12			
		298471.32	23.402638	1.186E+11	344394.47
Vietnam	Order	(0,1,1)x(0,1,1,12)12			
		10298.741	387.21047	107080927	10347.991

Fig. 12 MAE, MAPE, MSE, and RMSE Values of SARIMA Model with new Order Values Using Data During Covid-19 Pandemic

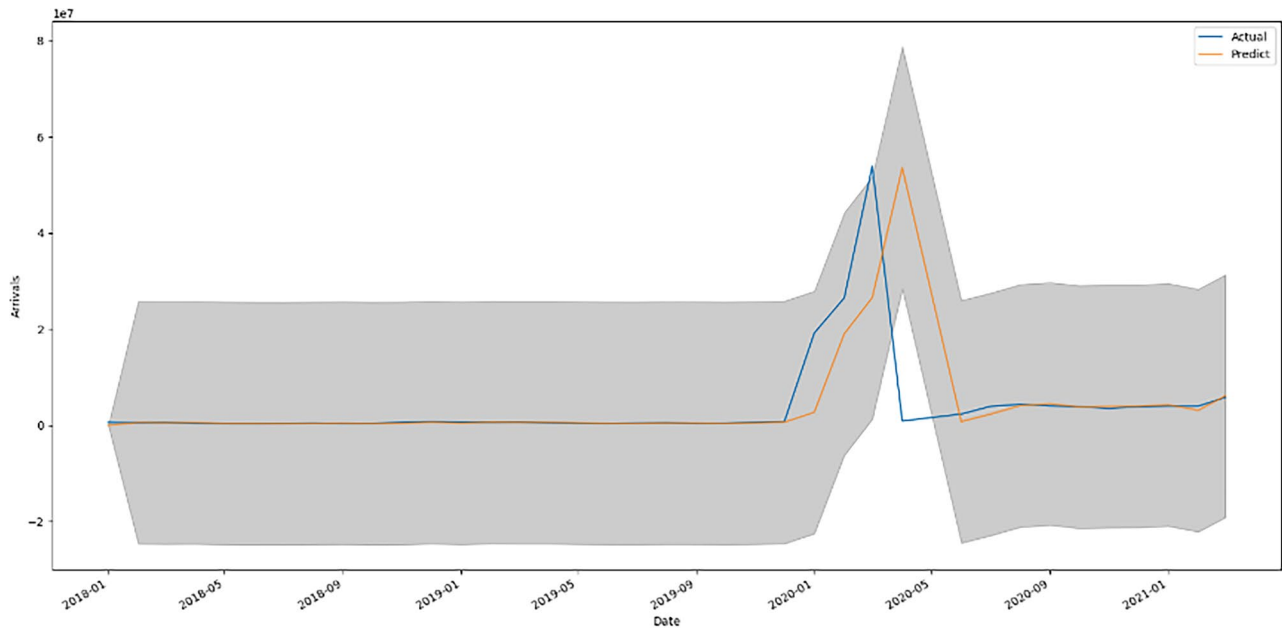


Fig. 13 Prediction and Actual Values using dataset year 2018-latest (Cambodia) – New SARIMA Order Values

Public without the issue of reading the icon images. The dashboard is shown in Figs. 19, 20, 21, 22, 23, and 24.

5.3 Limitations and future works

In this study, the data set that would be used does not involve the data of all the ASEAN countries. Tourist arrivals data of eight out of ten ASEAN countries are only provided on the

Internet and mostly can be obtained on Trading Economics website. This is due to data availability and data accessibility issues. Data availability issue refers to the dataset of certain countries that are not provided on any websites. While data accessibility issue refers to the tourist arrivals information of the country only being published in visualized statistic graph and the download option of the raw dataset is not provided by government official tourism websites.

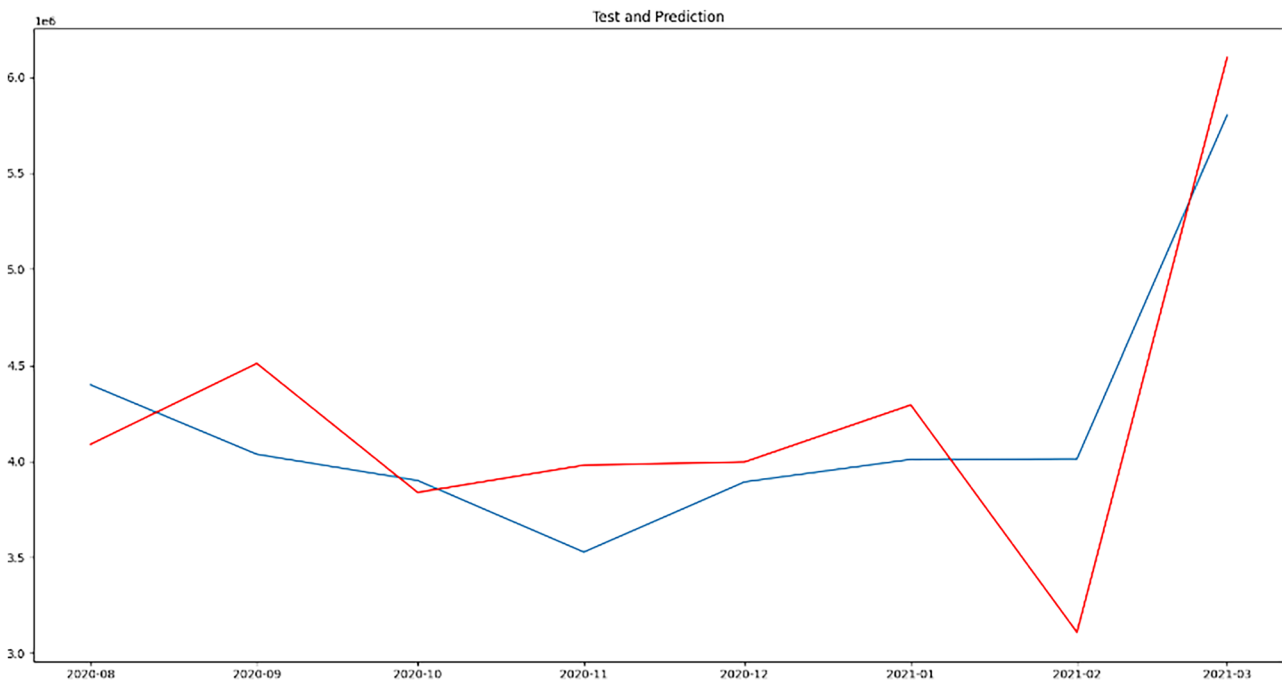


Fig. 14 Prediction and Actual Values (Aug 2020 – Mar 2021) using dataset year 2018-latest (Cambodia) – New SARIMA Order Values

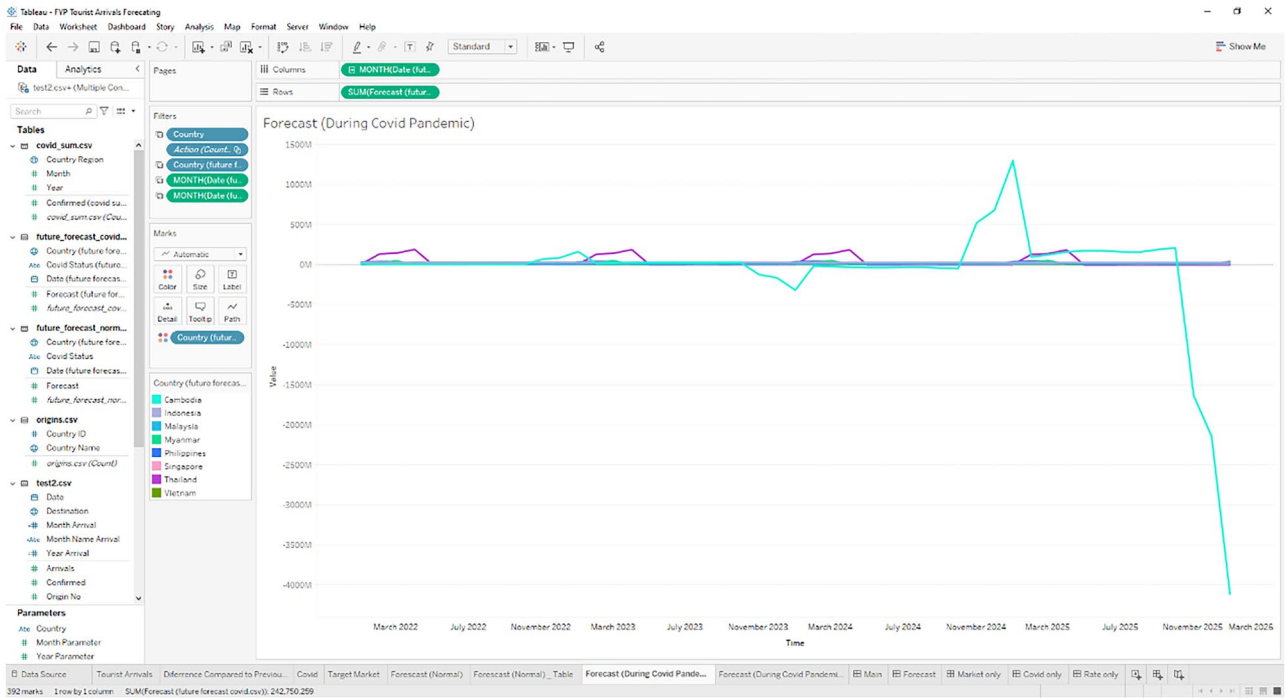


Fig. 15 Before Excluding Potential Forecasted Outliers

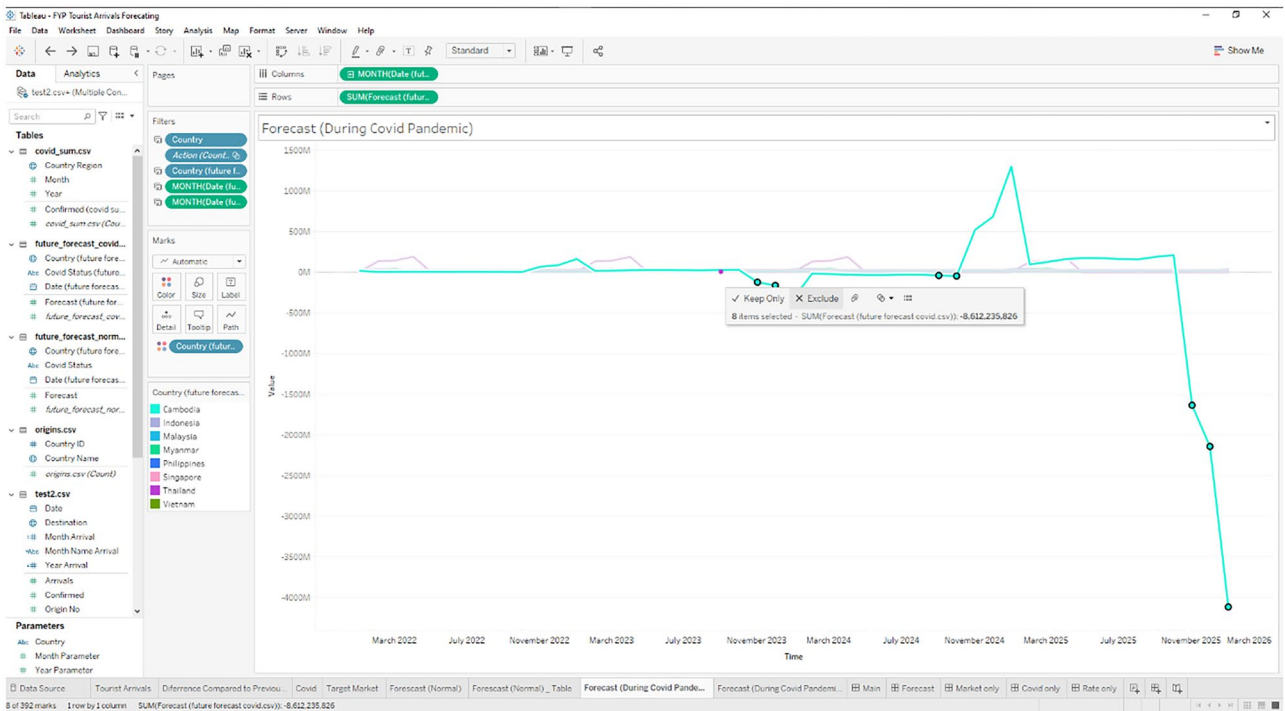


Fig. 16 Excluding Parts of Negative Forecasted Outliers

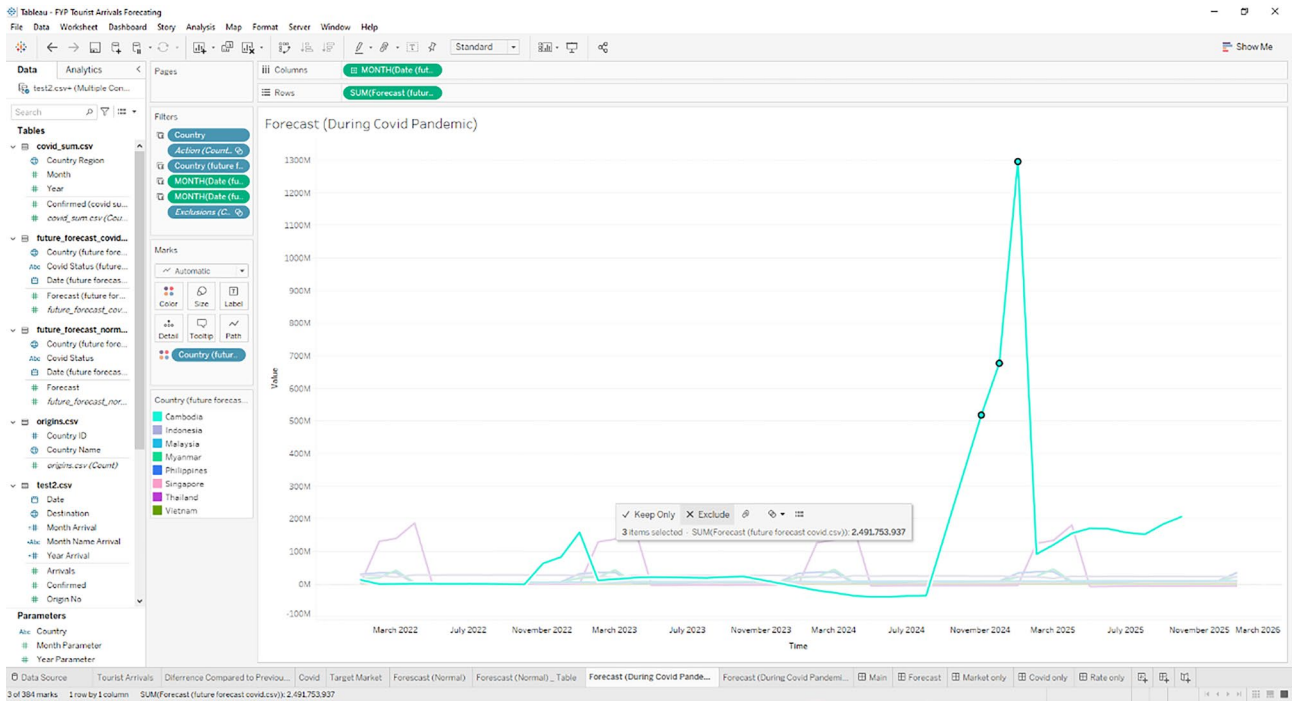


Fig. 17 Excluding Parts of Potential Positive Forecasted Outliers

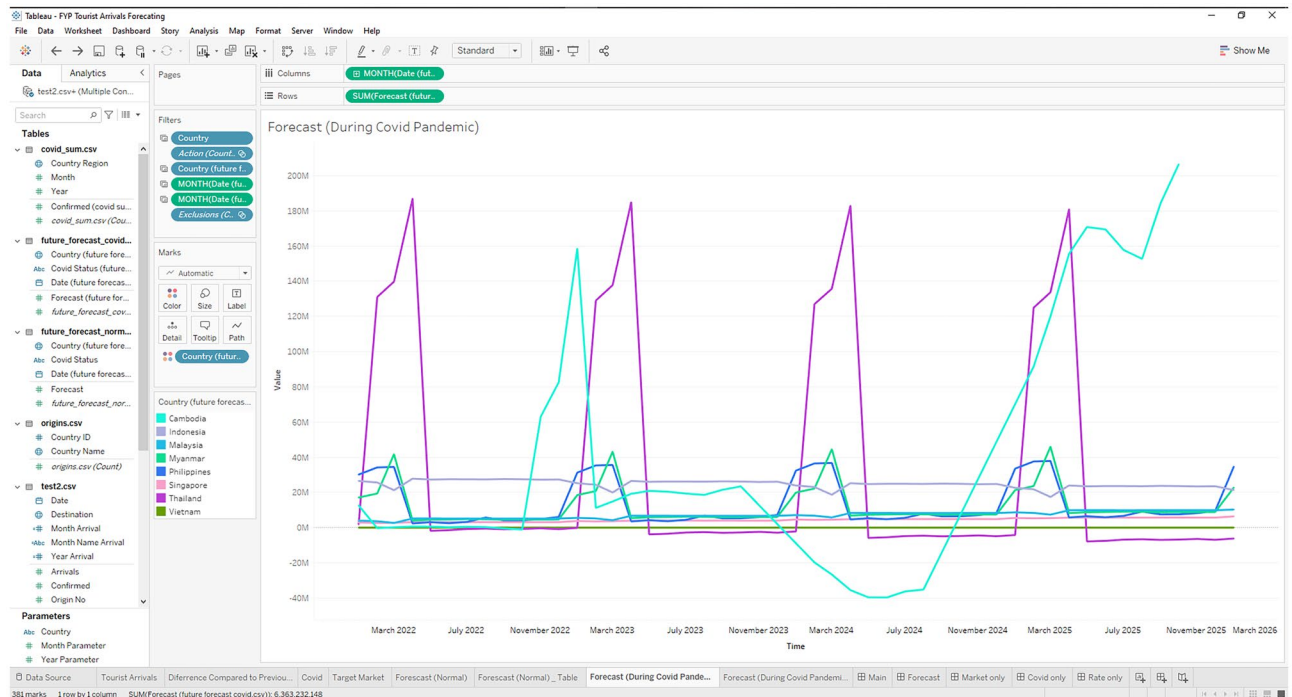


Fig. 18 Graph of Forecasted Tourist Arrivals Values (Covid) After Excluding Potential Outliers

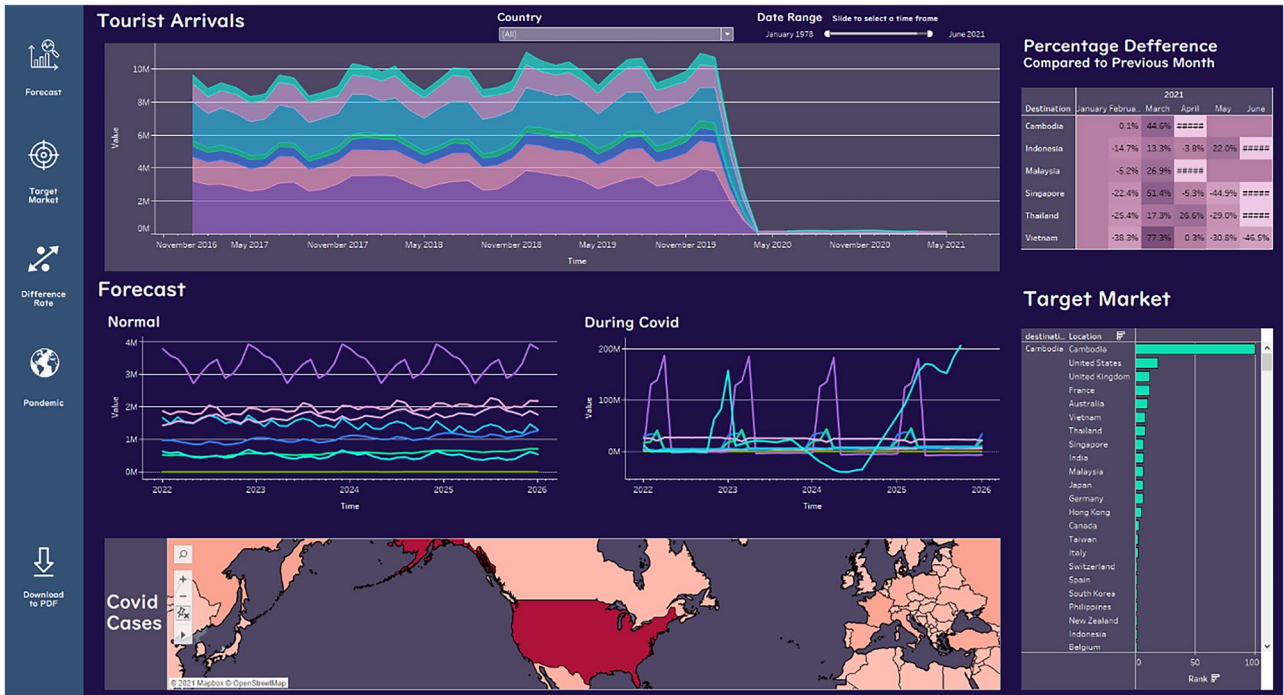


Fig. 19 Main Page

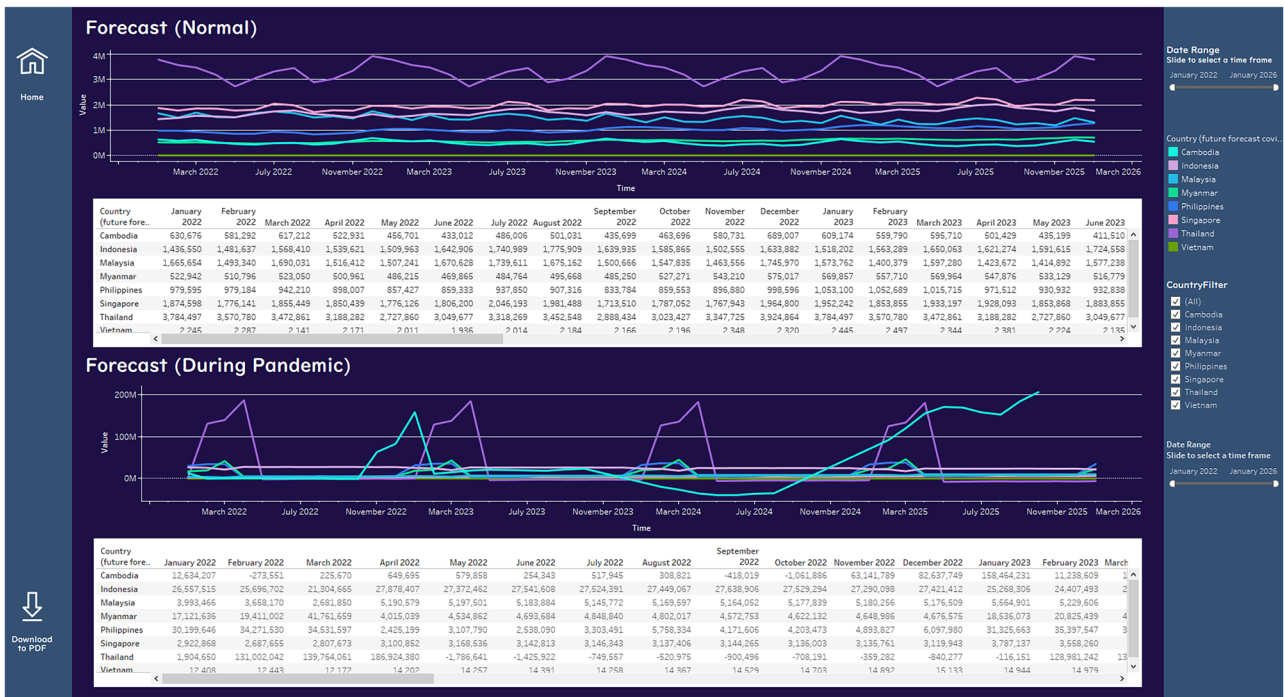


Fig. 20 Forecasted Value Page

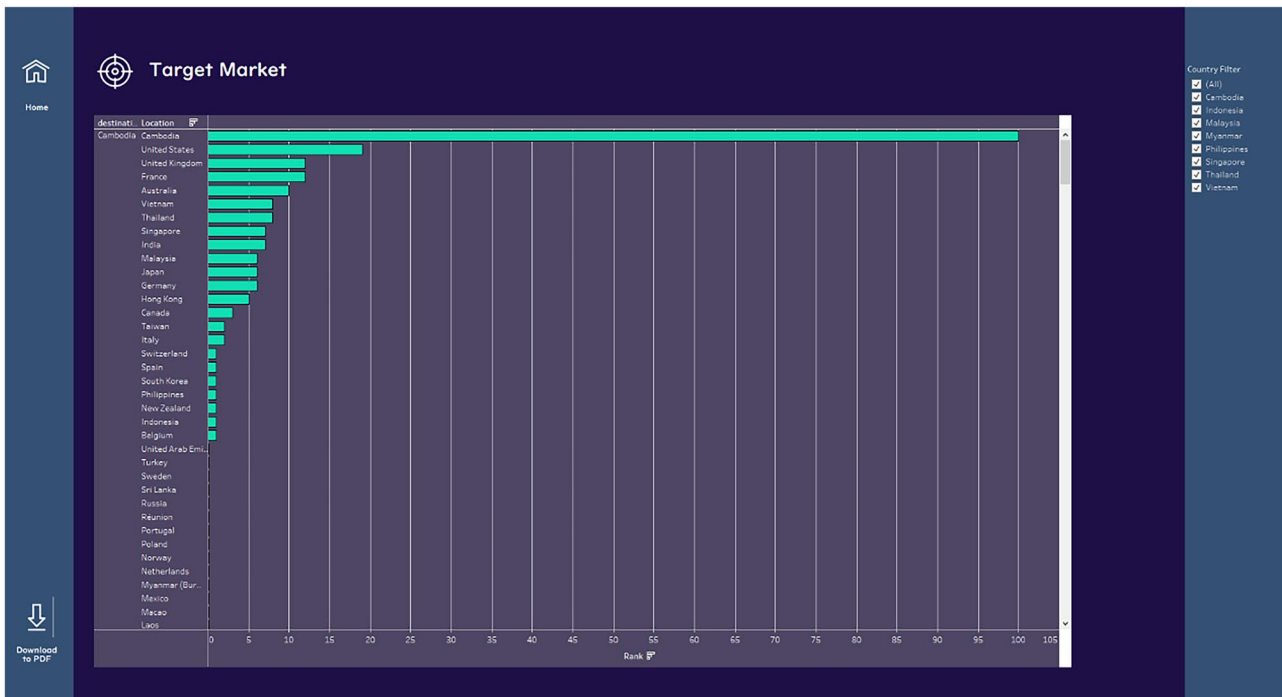


Fig. 21 Target Market Page



Fig. 22 Rate Difference Page

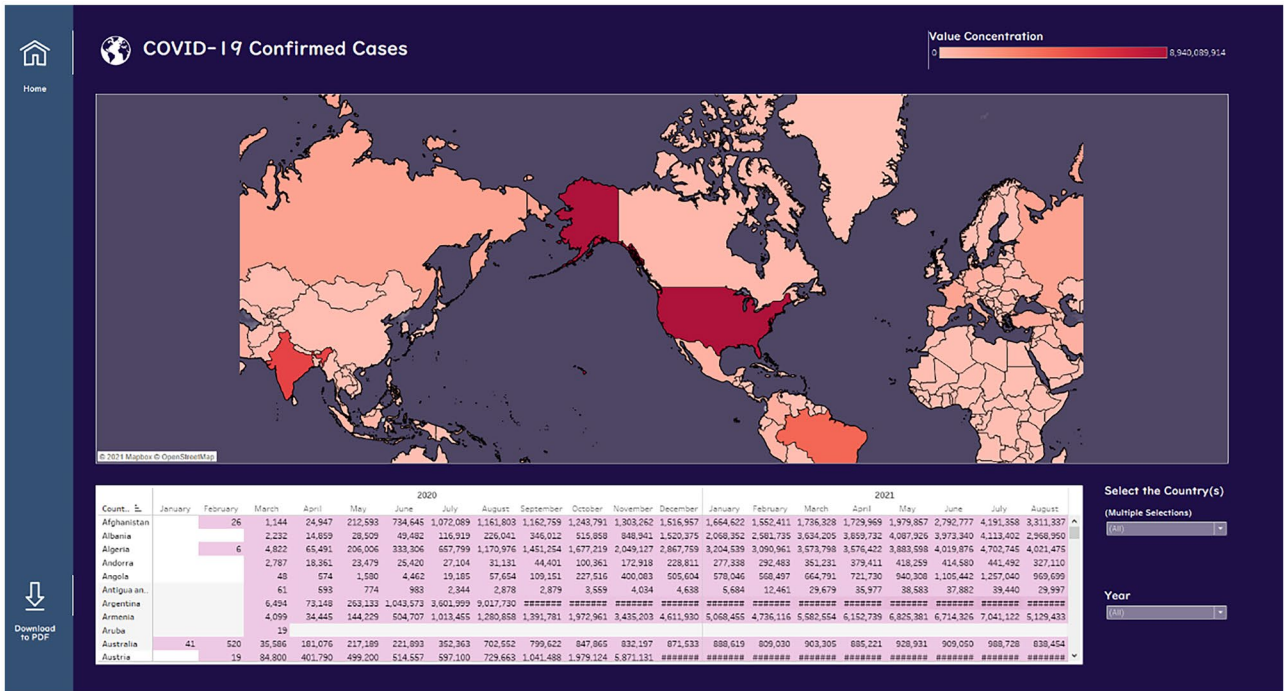


Fig. 23 Covid Cases Proportions Page

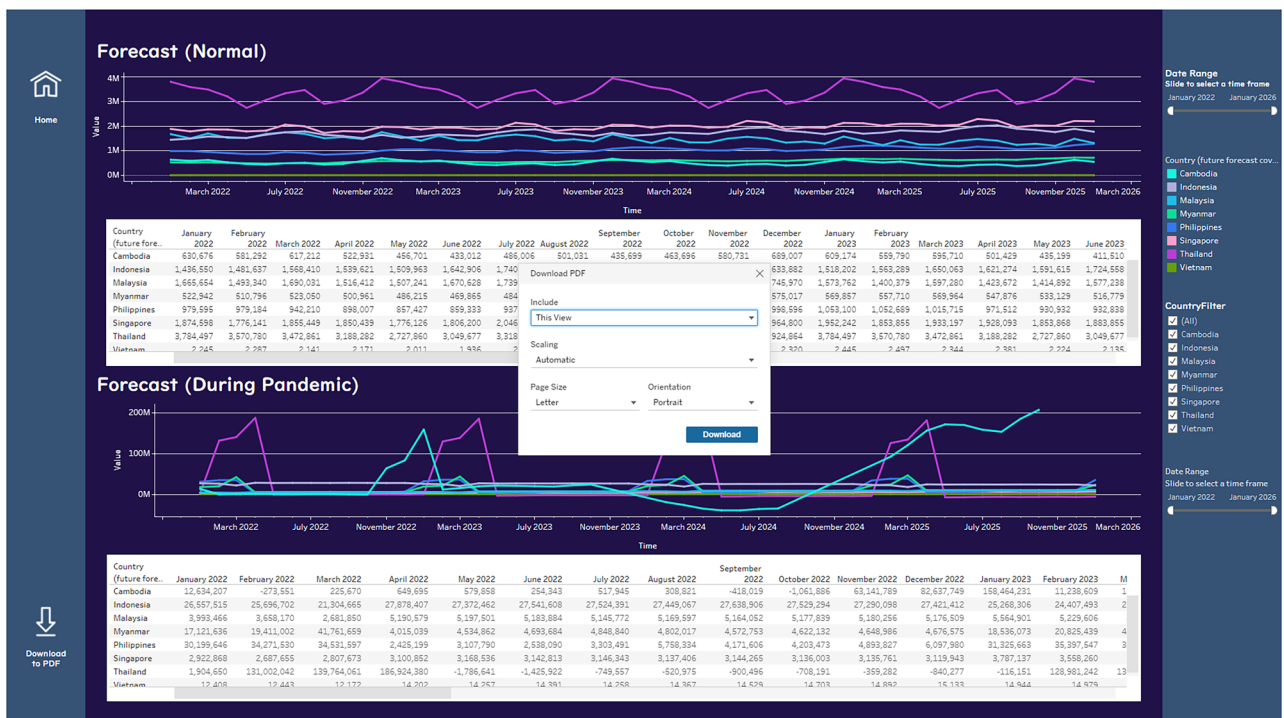


Fig. 24 Download Dashboard as PDF

Furthermore, the time frame of the tourist arrivals data will be much smaller if the analysis is only done on one country. Therefore, the tourist arrivals data as well as the Covid-19 cases number of totals of eight countries will be retrieved and used in this study, to increase the size of raw data for a more accurate prediction.

According to the four metrics for evaluating the performance of the selected predictive model out of three, the implementation of the SARIMA model has not performed well to forecast the tourist arrivals numbers in a few ASEAN countries assuming Covid-19 pandemic is still going on, even though it performed more accurate while assuming there is no Covid-19 from the beginning, some predictive even has negative values. Even though it can be eliminated by excluding these values, it could be a sign that there is still modification that can be done to improve the model for prediction.

This could be due to the lack of tourist arrivals records of respective countries with lower predictive performance occurred. For instance, Cambodia and Malaysia datasets have records until March 2021, Indonesia, Singapore, and Thailand datasets have records until May 2021, while Vietnam's records until June 2021, but Myanmar and the Philippines have only records until 2020. Therefore, the comparison of the SARIMA model's accuracy is barely can be done by comparing the results among eight countries' forecasted values. However, the datasets are still being updated on data source website, the Trading Economics. Therefore, the model can be trained and perform the forecasting for the same period to investigate if there will be a higher performance of tourist arrivals forecasting during the pandemic done by the SARIMA model in each country, using the updated datasets purchased from the website.

6 Conclusion

This work has discussed the current problems in tourism that are affected by the Covid-19 pandemic, as well as the challenges could be faced by tourism sectors with the tourism arrivals number that could be fluctuated due to the number of Covid-19 cases that decreased and increased inconsistently due to the restriction level of MCO and human behaviours. Therefore, time-series tourist arrivals that include the number of Covid-19 cases number as a prediction indicator is going to be done, since the pandemic started less than the past two years, and fewer time-series research based on Covid-19 cases in the tourism sector has been done. Besides, this work also aims to determine if the ARIMA model still has the best performance compared to Holt-Winters model and the SARIMA model in tourist arrivals forecasting based on Covid-19 cases. Another objective of this work is to develop a dashboard for data visualization on tourist arrivals forecasting using the

model with the highest accuracy in prediction. Therefore, after doing a literature review of past papers regarding to time-series analysis, the forecasting will be done in eight ASEAN countries due to the limitation of data availability and data accessibility. Finally, this research is expected to assist the management in tourism sectors such as entertainment, food and beverage, hospitality, management, and others for decision making, as well as the stakeholders, to decide if desire to continue supporting a particular tourism business. Recent literature on machine learning shows that the models are not robust and generalizable in an adversarial environment. Though this work has done a detailed investigation and analysis of the prediction of cases of Covid-19 and tourism, the model robustness and generalizability are not shown. This type of work will be considered as one of the significant directions towards future works.

Appendices

Appendix 1 Data Source

Tourist Arrivals

Country	URL
Cambodia	https://tradingeconomics.com/cambodia/tourist-arrivals
Indonesia	https://tradingeconomics.com/indonesia/tourist-arrivals
Malaysia	https://tradingeconomics.com/malaysia/tourist-arrivals
Myanmar	https://tradingeconomics.com/myanmar/tourist-arrivals
Philippines	https://tradingeconomics.com/philippines/tourist-arrivals
Singapore	https://tradingeconomics.com/singapore/tourist-arrivals
Thailand	https://tradingeconomics.com/thailand/tourist-arrivals
Vietnam	https://tradingeconomics.com/vietnam/tourist-arrivals

Covid-19

<https://github.com/CSSEGISandData/COVID-19>

Destination Insights with Google

https://destinationinsights.withgoogle.com/intl/en_ALL/

The small-sized or purchased raw datasets, preprocessed, integrated datasets as well as the forecasting results in CSV files can be accessed via the OneDrive folder given URL address below.

OneDrive folder:

https://studentnewintiedumy-my.sharepoint.com/:f:/g/personal/i19017476_student_newinti_edu_my/EiVwyBOSqLJAg4NkI8oGHMsB1MVsBMoGeQUmUUjQYB60dg?e=2xYO6X

Note: The size of Covid-19 raw data is large but also is open source and free to be downloaded by anyone. Therefore, it is not saved in the author's OneDrive folder.

Author contributions All authors have equal contributions.

Data availability The data that support the findings of this study are available from the first author upon reasonable request.

Code availability The code is available from the first author upon reasonable request.

Declarations

Informed consent None.

Conflicts of interest The authors declare no conflict of interest.

References

- Arunachalam A, Ravi V, Acharya V, Pham TD. Toward data-model-agnostic autonomous machine-generated data labeling and annotation platform: COVID-19 autoannotation use case. *IEEE Transact Eng Manag.* 2021.
- Ravi V, Narasimhan H, Chakraborty C, Pham TD. Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images. *Multimedia Syst.* 2022;28(4):1401–15.
- Yang Y, Zhang H, Chen X. Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak. *Ann Tour Res.* 2020;83: 102913.
- Khan N, Hassan AU, Fahad S, Naushad M. Factors affecting tourism industry and its impacts on global economy of the world. 2020. Available at SSRN 3559353.
- Shakeel T, Habib S, Boulila W, Koubaa A, Javed AR, Rizwan M, Sufiyan M. A survey on COVID-19 impact in the healthcare domain: worldwide market implementation, applications, security and privacy issues, challenges and future prospects. *Complex Intel Syst.* 2022;1–32.
- Manoj M, Srivastava G, Somayaji SRK, Gadekallu TR, Maddikunta PKR, Bhattacharya S. An incentive based approach for COVID-19 planning using blockchain technology. In *2020 IEEE Globecom Workshops GC Wkshps.* IEEE. 2020;1–6.
- Saleem K, Saleem M, Zeeshan R, Javed AR, Alazab M, Gadekallu TR, Suleman A. Situation-aware BDI reasoning to detect early symptoms of covid 19 using smartwatch. *IEEE Sens J.* 2022.
- Škare M, Soriano DR, Porada-Rochoń M. Impact of COVID-19 on the travel and tourism industry. *Technol Forecast Soc Chang.* 2021;163: 120469.
- Tran BL, Chen CC, Tseng WC, Liao SY. Tourism under the early phase of COVID-19 in four APEC economies: An estimation with special focus on SARS experiences. *Int J Environ Res Public Health.* 2020;17(20):7543.
- Alwi MK, Patwary AK, Ramly NI. Impact of covid-19 towards tourism industry: an evidence from Malaysia. *PalArch's Journal of Archaeology of Egypt/Egyptology.* 2020;17(8):354–82.
- Bakar NA, Rosbi S. Effect of Coronavirus disease (COVID-19) to tourism industry. *Int J Adv Eng Res Sci.* 2020;7(4):189–93.
- Abbas J, Mubeen R, Iorember PT, Raza S, Mamirkulova G. Exploring the impact of COVID-19 on tourism: transformational potential and implications for a sustainable recovery of the travel and leisure industry. *Curr Res Behav Sci.* 2021;2: 100033.
- Casini L, Roccetti M. A cross-regional analysis of the COVID-19 spread during the 2020 Italian vacation period: Results from three computational models are compared. *Sensors.* 2020;20(24):7319.
- Singh S, Sundram BM, Rajendran K, Law KB, Aris T, Ibrahim H, Dass SC, Gill BS. Forecasting daily confirmed COVID-19 cases in Malaysia using ARIMA models. *J Infect Develop Countries.* 2020;14(09):971–6.
- Khashei M, Montazeri MA, Bijari M. Comparison of four interval ARIMA-base time series methods for exchange rate forecasting. Published Online July. 2015.
- Katris C, Daskalaki S. Dynamic bandwidth allocation for video traffic using FARIMA-based forecasting models. *J Netw Syst Manage.* 2019;27(1):39–65.
- Pandianchery MS, Ravi V. Explainable AI Framework for COVID-19 Prediction in Different Provinces of India. 2022. <https://arxiv.org/abs/2201.06997>.
- Purwanto, Sunardi, Julfia, FT, Paramananda A. Hybrid model of ARIMA-linear trend model for tourist arrivals prediction model in Surakarta City, Indonesia. In *AIP Conf Proc.* 2019;2114(1), 060010. AIP Publishing LLC.
- Liang YH. Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. *Comput Ind Eng.* 2014;74:111–9.
- Nanthakumar L, Subramaniam T, Kogid M. Is'malaysia truly asia'? Forecasting tourism demand from asean using sarima approach. *Tourismos.* 2012;7(1).
- Petrevska B. Predicting tourism demand by ARIMA models. *Economic research-Ekonomska istraživanja.* 2017;30(1):939–50.
- Yao Y, Cao Y, Ding X, Zhai J, Liu J, Luo Y, Ma S, Zou K. A paired neural network model for tourist arrival forecasting. *Expert Syst Appl.* 2018;114:588–614.
- Qiu RT, Wu DC, Drotsky V, Petit S, Pratt S, Ohe Y. Visitor arrivals forecasts amid COVID-19: A perspective from the Asia and Pacific team. *Ann Tour Res.* 2021;88:103155. <https://doi.org/10.1016/j.annals.2021.103155>.
- Polyzos S, Fotiadis A, Samitas A. COVID-19 Tourism Recovery in the ASEAN and East Asia Region: Asymmetric Patterns and Implications. *ERIA Discussion Paper Series, Paper,* 2021;(379).
- Christidis P, Christodoulou A. The predictive capacity of air travel patterns during the global spread of the COVID-19 pandemic: risk, uncertainty and randomness. *Int J Environ Res Public Health.* 2020;17(10):3356.
- Velos SP, Go MB, Bate GP, Joyohoy EB. A Seasonal Autoregressive Integrated Moving Average (SARIMA) Model to Forecasting Tourist Arrival in the Philippines: A Case Study in Moalboal, Cebu (Philippines). *Recoletos Multidisciplin Res J.* 2020;8(1):67–78.
- Hyndman RJ, Athanasopoulos G. *Forecasting: principles and practice*, 3rd edition. [e-book] Melbourne: OTexts. 2021. Available at: <https://otexts.com/fpp3/>. [Accessed 5 July 2021].
- Lip NM, Jumery NS, Termizi FAA, Mulyadi NA, Anuar N, Ithnin H. Forecasting international tourist arrivals in malaysia using sarima and holt-winters model. *Management.* 2019;5(18):41–51.
- Kurniasih N, Ahmar AS, Hidayat DR, Agustin H, Rizal E. Forecasting infant mortality rate for China: a comparison between α -Sutte Indicator, ARIMA, and Holt-winters. In *J. Phys.: Conf. Series* 2018; 1028(1):012195. IOP Publishing.
- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020;20(5):533–4.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.