A Comparison Study on Fuzzy Time Series and Holt-Winter Model in Forecasting Tourist Arrival in Langkawi, Kedah

Nur Fatihah Fauzi[1]*, Nurul Shahiera Ahmadi[2], Nor Hayati Shafii[3], Huda Zuhrah Ab. Halim[4]
[1][1,2,3,4] Faculty of Computer and Mathematical Sciences ,
Universiti Teknologi MARA, Perlis Branch, Arau Campus, 02600 Arau, Perlis

ABSTRACT

The tourism industry in Malaysia has been growing significantly over the years. Tourism has been one of the major donors to Malaysia's economy. Based on the report from the Department of Statistics, a total of domestic visitors in Malaysia were recorded at about 221.3 million in 2018 with an increase of 7.7% alongside a higher record in visitor arrivals and tourism expenditure. This study aims to make a comparison between two methods, which are Fuzzy Time Series and Holt-Winter in forecasting the number of tourist arrival in Langkawi based on the monthly tourist arrival data from January 2015 to December 2019. Both models were generated using Microsoft Excel in obtaining the forecast value. The Mean Square Error (MSE) has been calculated in this study to get the best model by looking at the lowest value. The result found that Holt-Winter has the lowest value that is 713524285 compared to the Fuzzy Time Series with a value of 2625517469. Thus, the Holt-Winter model is the best method and has been used to forecast the tourist arrival for the next 2 years. The forecast value for the years 2020 and 2021 are displayed by month.

INTRODUCTION

The domestic tourism spent a total of MYR 92 561 million compared to MYR 83 103 million in the preceding year. The tourism-led growth theory recommended and suggested that arriving tourists in Malaysia can be varied to produce economic growth (Tang and Tan, 2015). Many advanced countries around the world depend on tourists' investments which contributes to the Gross Domestic Product (GDP) of those countries. Langkawi is enclosed by a bundle of small islands and is one of Malaysia's most popular beach destinations perfect for tourist attraction, providing a peaceful and divine experience for a fantastic visit for both business and holiday purposes. Langkawi has been a potential tourist destination for a long time, although the early improvement was witnessed in 1948 and the momentum gained in the mid-1980. The island has grown rapidly over the last four decades with the transition from a beautiful tropical island that was a place for trawling and agricultural activities to a modern island for tourism (Irwana Omar et al., 2014). Thus, this study is conducted to seek out what is the best forecasting of the number of tourist arrival by comparing two models. The main objective of this study is to determine the best forecasting numbers of tourist arrival in Langkawi, by comparing the Fuzzy Time Series and Holt-Winter models. The data is were obtained from Langkawi's development agency (LADA) that consists of the number of tourists arriving in Langkawi from January 2015 until December 2019. Hence, this study aims to forecast the number of tourists arriving in Langkawi for the next 2 years by using the best model.

The achievement of the tourism field in Malaysia is based on the intensity of biodiversity and natural surroundings. Malaysian society has a wealth of traditions and culture as well as pleasant diplomatic affairs with other countries by being involved in international events, giving a big impact on tourism (Pillai, 2018). The research that has been done by Wahid et al. (2016) repeated that visitors in Langkawi possess higher value than first-time visitors with a value of 61.2% and 38.8, respectively. Repeated visitors in certain places make them popular ones. An increase in the number of visitors can be concluded as an attractive and higher value in the destination image. Most visitors visited Langkawi for leisure and preferred to stay at hotels or rented houses.

The method that has been suggested by Khairina et al. (2019) used the single exponential smoothing (SES) method to forecast groundwater tax revenue. This study was about groundwater tax revenue of Kutai Kartaegara Regional Office, Indonesia from January 2013 to January 2017. The forecasting was calculated based on Mean Average Percent Error (MAPE) and the best value is 45.868 with the value of alpha 0.1. Furthermore, the result shows that forecasting of groundwater tax for the year 2018 is Rp 443.906.600,7192. Other than that, recent studies by Alsharif et al. (2019) have proposed the method of ARIMA and SARIMA models in predicting daily and monthly average global solar radiation. The case study was conducted in Seoul, South Korea with the objective to develop both models in predicting daily and monthly solar radiation based on hourly solar radiation data that has been acquired from Korean Meteorological Administration from 1981 until 2017. The result shows that ARIMA (1,2,2) is used to delegate daily solar radiation and monthly, it is represented by SARIMA (4,1,1) with the average monthly use from 176 to 377 Wh/m2.

Fuzzy time series (FTS) is a modeling technique based on fuzzy logic, which has the same interest. FTS can be treated as another side of statistical time series analysis using a fuzzy set. The research done by Güler Dincer and Akkuş (2018) formulated the fuzzy time series model to forecast air pollution. This model was based on the use of a robust fuzzy clustering algorithm called Fuzzy K-meloid in the fuzzification step of the FTS model building algorithm. The proposed model (FKMF), FTS models based on fuzzy c-means (FCMF) and Gustafson-Kessel (GKF) clustering algorithms were utilized to the time series to calculate the forecasting achievement of the proposed model and analyse it with other fuzzy clustering-based FTS models. The time series of weekly SO2 concentrations and forecasting results than have been compared between RMSE and MAPE goodness of fit measures. The result showed the observation that proved that the method has provided successful forecasting results specifically in time series which are using various outliers.

The Holt-Winter method is suitable when there is a presence of trend and seasonality patterns. This method is built for a time series that shows the presence of a linear trend. There are two types of Holt-Winter, which are additive effect and multiplicative effect. The additive is used when seasonal variation remains unchanged, while the multiplicative effect is used when there is an increase in seasonal variation (Tirkes et al., 2017). This method has also been presented by Rahman and Ahmar (2017) in predicting primary energy consumption. The data consisted of the total primary energy consumption of several sectors in the United States including the residential sector, commercial sector, industrial sector, transportation sector, electric power sector and energy consumption balancing item. The data collected were from January 1973 to December 2016. This study aimed to compare two methods: ARIMA model and Holt-Winter model, based on error measurements MAE, RSS, MSE and RMS. The result found that Holt-Winter Additive type has the best value of MSE 258350.1

The main objective of this study is to determine the best accuracy in forecasting numbers of tourist arrival in Langkawi, by comparing the Fuzzy Time Series and Holt-Winter models and to forecast the number of tourist arrival in Langkawi for the next 2 years by using the best method. This study is conducted to seek out what is the best forecasting of the number of tourist arrival by comparing two methods, which are the Fuzzy time series and Holt-Winter. Langkawi Island has been chosen as the location for conducting this

study. Langkawi is an island located in Kedah, Malaysia, famous among tourists ranging from locals or those from all over the world. This study uses secondary data that were collected monthly. The data consists of the number of tourist arrival in Langkawi from January 2011 until December 2019. Besides, the data is also obtained from Langkawi's development agency (LADA).

This study will primarily benefit the authority particularly in the advertising activities and related tourism industry to conduct strategic planning to attract more tourists to visit Langkawi in the upcoming years. By expanding this sector, the employment aspect will grow, and the nation can get the advantage directly and indirectly. Forecasting at high precision becomes valuable since it may guarantee the development and the willingness of all tourism agencies such as hotels, transportation, food, and services industries. Furthermore, it also can help ensure these complementary services are available when and where future visitors need them, which will rebound to the benefit of their hotels or other individual tourism facilities. Also, it will give advantages to Langkawi's development agency to improve their infrastructure, to fund conservation efforts and promote more interesting tourism advertising.

RESEARCH METHODOLOGY

Fuzzy Time Series
There are 7 steps involved as follows:

Step 1: By using Microsoft excel, all data has been analysed and changed into percentage from year to year. The equation of changes is calculated as follows:

where,
X_n = total of tourist arrival in a month,
X_{n-1} = total of tourist arrival in the month before.

Step 2: The minimum and maximum values need to be identified from the value of percentage changes. Then, by using a rule where is the Universe of discourse that needs to be identified. $D1$ and $D2$ need to be positive numbers that have been assigned in U.

Step 3: The length of interval needs to be constructed at this step and the value needs to be added in order to make a fuzzification of interval where fuzzy sets U_i and i is equal to 1 until 7. The frequency distribution of each interval also needs to be produced at this step and the calculation of length of interval be follows:

Length of interval $= \dfrac{\left(D_{\min} + D_2\right) - \left(D_{\min} - D_1\right)}{7}$.

Step 4: Based on step 2, the interval of $v_1, v_2, \ldots, v_n$ needs to be constructed. The interval needs to be generated in the form of trapezoidal number and it can be represented as follows:

$$A_1 = \left[m_0, m_1, m_2, m_3\right],$$
$$A_2 = \left[m_1, m_2, m_3, m_4\right],$$
$$\vdots$$
$$A_n = \left[m_{n-1}, m_n, m_{n+1}, m_{n+2}\right].$$

Step 5: At this step, the fuzzified data need to be generated by classifying each data with the interval produced in step 4. To classify the interval, it needs to refer to the changes value. If the data shows the range of v_j, then the fuzzy number needs to be in A_j. After classification, the fuzzy logical relationship

needs to be produced and it can be referred to as $A_i \rightarrow A_j$, Where A_i is the present data and A_j is the future data.

Step 6: In this step, the fuzzy logical relationship rule needs to generate and put it based on the group. It can be arranged as follows:

$A_x \rightarrow A_b$
$A_x \rightarrow A_c$
$A_x \rightarrow A_d$
A_x can be grouped as $A_x \rightarrow A_b, A_c, A_d$.

Step 7: After that, each of the fuzzy relationship rules need to be classified between rule 1, rule 2 or rule 3 based on the calculation below. The example of the calculation is shown as follows.

Rule 1: When the fuzzy group A_j is empty, it means that the group has no relationship with others. It can be symbolized as $A_j \rightarrow A_\phi$ or $A_j \rightarrow A_j$. The forecast value for this rule is $F_{vt} = R\left[NSTFN\left(A_j\right) \right]$.

Rule 2: The fuzzy group is one to one where there is only one relationship rule that is related. For example, $A_j \rightarrow A_b$. The forecast value is $F_{vt} = R\left[NSTFN\left(A_b\right) \right]$.

Rule 3: The fuzzy group is one to many
$A_j \rightarrow A_b$
$A_j \rightarrow A_c$
$A_j \rightarrow A_d$
It would be written as $A_j \rightarrow A_b, A_c, A_d$ and the forecast can be calculated as

$$F_{vt} = R\left[\frac{NSTFN\left(A_b\right) + NSTFN\left(A_c\right) + NSTFN\left(A_d\right)}{n} \right]$$, where n is the number pairing of A_j .

Holt-Winter Model

The Holt–Winters method was originally developed by Winters and involves estimating three smoothing parameters associated with the level, trend, and seasonal factors.

<u>Multiplicative effect assumption</u>
On the assumption that the relationship of these components is multiplicative in nature, the equations are represented as follows:

Level component:
$$L_t = \alpha \frac{y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1})$$
Trend component:
$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$
Seasonality component:
$$S_t = \gamma \frac{y_t}{L_t} + (1-\gamma)S_{t-s}$$

The m-step-ahead forecast is calculated as:

$$F_{t+m} = \left(L_t + b_t \times m\right)S_{t-s+m}$$

Additive Effect Assumption

The additive effect assumption pertaining to the components of the models can also be made. The basic equations are

Level component:

$$L_t = \alpha\left(y_t - S_{t-s}\right) + \left(1 - \alpha\right)\left(L_{t-1} + b_{t-1}\right)$$

Trend component:

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

Seasonality component:

$$S_t = \gamma\left(y_t - L_t\right) + \left(1 - \gamma\right)S_{t-s}$$

The m-step-ahead forecast is calculated as:

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

where,

y_t: the actual values which include seasonality

L_t: the level component of the series, comprising the smoothened values but does not include the seasonality component.

b_t: the estimation of the trend component,

S_t: the estimation of seasonality component,

S: the length of seasonality,

α: the smoothing constant for level where $0 < \alpha < 1$,

β: the smoothing constant for the trend estimation where $0 < \beta < 1$,

γ: the smoothing constant for seasonality estimation where $0 < \gamma < 1$,

m: number of steps ahead to be forecast,

F_{t+m}: forecast for m-step-ahead.

Throughout the process of this data collection, Microsoft Excel has been used to get the forecasting value for Fuzzy Time Series and Holt-Winter Model. For forecasting the next 2 years, Holt-Winter method has been used since it is the best model based on the lowest value of error measurement.

FINDINGS AND DISCUSSION

The data were analyzed and forecasted by using Microsoft Excel. Figures 1 and 2 show the relationship of actual data and forecasted value of tourist arrival for Fuzzy Time Series and Holt-Winter model. For Figure 2, the data needs to be analysed to determine whether it is additive or multiplicative by nature. Then, Figure 2 shows that the components are interacting in a multiplicative manner since the seasonal variation is increased when the time increases. The data that has been calculated was based on the Multiplicative assumption.
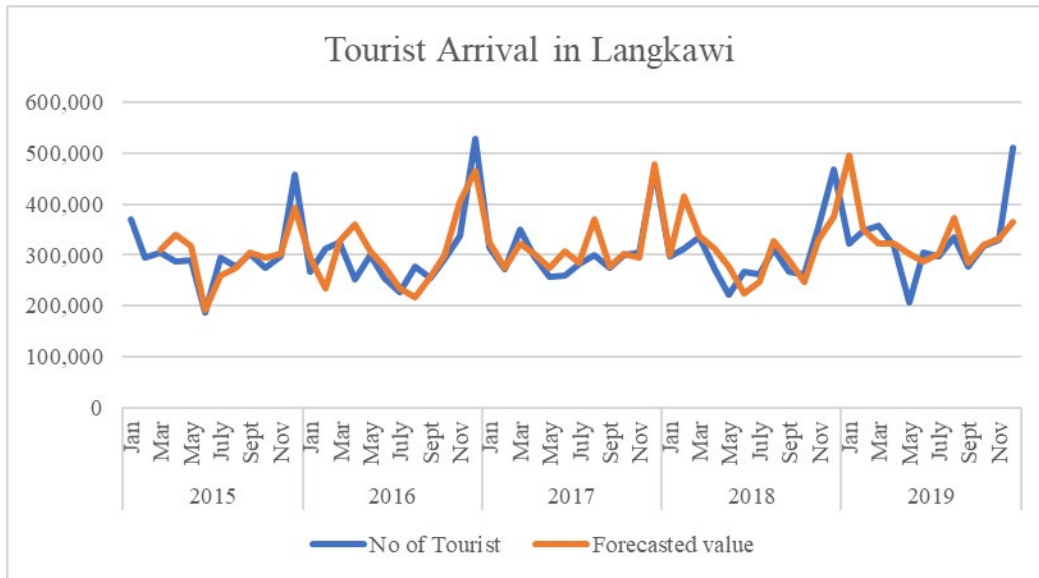
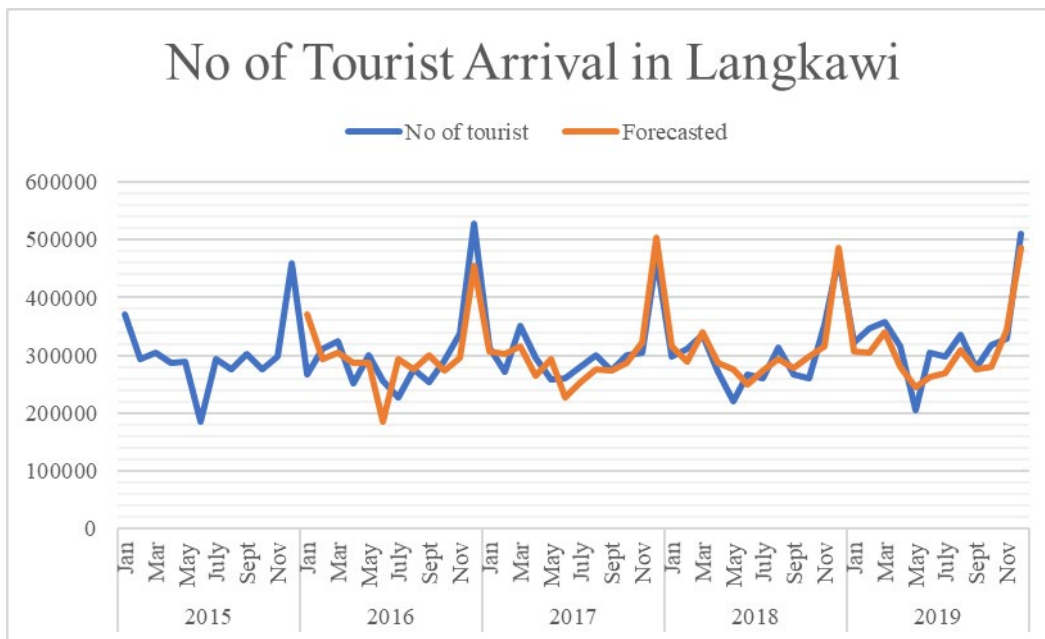Figure 1: Actual data and forecasted value for Fuzzy Time Series

Figure 2: Actual data and forecasted value for Holt-Winter model

In order to evaluate the accuracy of forecast data by the two methods, we computed mean absolute percentage error (MAPE), mean absolute error (MAE), mean square error (MSE) and root mean square error (RMSE). For all three measures, the smaller the value, the better the fit of the model and calculate the forecasted value for the next 2 years. The value of MAPE, MAE, MSE and RMSE for both models can be seen in Table 4.9. However, the value of tourist arrival for January and February 2015 cannot be evaluated since there is no value before the data. These statistics are computed as follows:

e = actual data – forecasted,

$$\text{MAPE} = \frac{\sum \left| \frac{e}{\text{actual data}} \right|}{n} (100) \; ; \; y_t \neq 0,$$

$$\text{MAE} = \frac{\sum e}{n},$$

$$\text{MSE} = \frac{\sum e^2}{n},$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum e^2},$$

where n is number of data.

Table 1: The value of square error for both models

Year	Month	No. of Tourist Arrival	FTS $\left(e^2\right)$	HW $\left(e^2\right)$
2015	Jan	370,178		
	Feb	294,171		
	Mar	304,953	13194177	
	Apr	287,500	2638237238	
	May	288,496	828806521	
	Jun	185,110	45425443	
	July	294,506	1314717909	
	Aug	276,574	3239647	
	Sept	301,667	12674073	
	Oct	275,113	376534314	
	Nov	297,372	24772390	
	Dec	458,877	4405733179	
2016	Jan	267,075	725089928	10612090435
	Feb	311,547	6015955755	326176485
	Mar	324,634	4747183	425152003
	Apr	251,750	11877359853	1205485328
	May	299,160	78877490	147386486
	Jun	255,449	467986343	5104221958
	July	226,915	71931661	4395304170
	Aug	276,501	3537810206	2182695
	Sept	253,758	17786741	2124283269
	Oct	292,338	46130319	371916735
	Nov	336,736	4577220052	1741726284

	Dec	528,286	3811614230	5324073767
2017	Jan	313,710	142217321	42421764
	Feb	271,747	9574884	967724423
	Mar	351,091	781573849	1304558945
	Apr	296,483	43743148	971727829
	May	257,338	251891323	1361140567
	Jun	259,430	2168822256	1096896332
	July	281,138	7885622	722328552
	Aug	299,779	5082983613	525377227
	Sept	274,462	3512934	72107
	Oct	300,569	1133667	187457649
	Nov	305,591	147512813	306305196
	Dec	467,820	100041756	1330273050
2018	Jan	298152	2497266	277226937
	Feb	312000	10805988200	572842523
	Mar	334886	14128577	39021553
	Apr	271648	1589055357	264430285
	May	220407	3276923713	3138330115
	Jun	266305	1683215164	287348950
	July	261124	209447728	161565939
	Aug	312830	232446419	354186502
	Sept	267159	510043651	107889280
	Oct	260949	217916098	1377922835
	Nov	354291	567639507	1511107958
	Dec	469200	8877107954	306623388
2019	Jan	322916	30166576489	245093450
	Feb	347692	7845749	1873414169
	Mar	358647	1240723244	350679671
	Apr	316630	35249297	1310997982
	May	205568	9458910822	1594971524
	Jun	305062	333942117	1866324101
	July	297267	19707544	782905733
	Aug	335861	1415660100	715005515
	Sept	278003	54932968	6493237
	Oct	317597	3445654	1353130753
	Nov	328817	18855161	269132416
	Dec	510266	20991532713	551229018

Table 2 shows the value of MAPE, MAE, MSE and RMSE between Fuzzy Time Series with multiplicative and Holt-Winter. Thus, to make a forecast on tourist arrival in Langkawi, the Holt Winter model is the best model since the value of MAPE, MAE, MSE and RMSE is the lowest. Meanwhile, the forecasted value for the year 2020 and 2021 can be seen in Table 3.

Table 2: The value of MAPE, MAE, MSE and RMSE for both methods

Method	MAPE	RANK	MAE	RANK	MSE	RANK	RMSE	RANK
Fuzzy Time Series	10.3481	(2)	31981.3621	(2)	2437419693.6379	(1)	49370.2308	(2)
Holt-Winter	10.0651	(1)	29350.2447	(1)	1263041162.1040	(2)	35539.2904	(1)

Table 3: The forecasted value of tourist arrival by using Holt-Winter for next 2 year

Month	Forecasted value	
	Year 2020	Year 2021
Jan	324245	334788
Feb	338671	349653
Mar	360691	372356
Apr	310531	320547
May	228517	235868
Jun	296474	305985
July	295018	304458
Aug	335494	346200
Sept	286025	295128
Oct	312503	322422
Nov	346460	357428
Dec	517040	533365

The most important thing to note in choosing the best method for forecasting is by looking at the lowest errors measured, and freedom from some problems towards assumptions.

CONCLUSION AND RECOMMENDATION

The main objective of this study is to make a comparison between two models, the model which gives the lowest value of MSE is the Holt-Winter method with the value of 1248211606 representative of the best method in forecasting tourist arrival. Therefore, the main objective of this study has been accomplished by making a comparison between these two methods. Meanwhile, the sub-objective is to forecast the number of tourists arriving in Langkawi for the next 2 years by using the best model. The forecasted value is then generated and displayed by month for the years 2020 and 2021 by using the Holt-Winter model. There are many methods that can be used to make a comparison and determine which method is better to make a forecast. The forecasted value can be trained with many types of algorithm in order to receive a more accurate result. Lastly, the forecasted values for the years 2020 and 2021 may be slightly different since the pandemic took place around the world including Malaysia, termed as the COVID-19. Surely, the tourism industry is heavily impacted by the current events happening in the world.

Acknowledgments

REFERENCES

Alsharif, M. H., Younes, M. K., & Kim, J. (2019). Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea. Symmetry, 11(2), 1-17. https://doi.org/10.3390/sym11020240

Güler Dincer, N., & Akkuş, Ö. (2018). A new fuzzy time series model based on robust clustering for forecasting of air pollution. *Ecological Informatics*, 43, 157-164. https://doi.org/10.1016/j.ecoinf.2017.12.001

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

Khairina, D. M., Muaddam, A., Maharani, S., & Rahmania, H. (2019). Forecasting of Groundwater Tax Revenue Using Single Exponential Smoothing Method. *E3S Web of Conferences*, 125(2019), 1-5. https://doi.org/10.1051/e3sconf/201912523006

Pillai, A. (2018). " Research on Tourism in Malaysia " Retrived. February, 2018, from https://doi.org/10.13140/RG.2.2.19565.28640

Rahman, A., & Ahmar, A. S. (2017). Forecasting of primary energy consumption data in the United States: A comparison between ARIMA and Holter-Winters models. *3rd Electronic and Green Materials International Conference 2017*, 1885(1), 020163. https://doi.org/10.1063/1.5002357

Tang, C. F., & Tan, E. C. (2015). Does tourism effectively stimulate Malaysia's economic growth? *Tourism Management*, 46, 158-163. https://doi.org/10.1016/j.tourman.2014.06.020

Tirkeş, G., & Güray, C. (2017). *Tehnički vjesnik 24*. 24(2), 503–509. https://doi.org/http://dx.doi.org/10.17559/TV-20160615204011

Wahid, S. D. M., Aliman, N. K., Hashim, S. M. & Harudin S. (2016). First-time and repeat visitors to Langkawi Island, Malaysia. *Procedia Economics and Finance*, 35, 622-631. https://doi.org/10.1016/S2212-5671(16)00076-9