

# Tourists' Arrival Prediction Using Regression Techniques

John C. Amar

D. Mgt., Associate Professor V  
College of Computer Studies  
University of Antique Tario-Lim Memorial Campus  
Tibiao, Antique

**Abstract:** *This study was conducted to predict the tourists' arrival in the Island of Boracay, Malay, Aklan, Philippines. It adopted the developmental type of research, for it is aimed to come up with a system to predict the arrival of tourist per year and determine the factors affecting such visit. It made use of Linear Regression technique in order to predict the arrival of tourists per year. It was found out to be precise and reliable. Also, it utilized the Multiple Linear Regression Technique in order to know the factors affecting the tourists' visit in the island. These mathematical tools are accorded with Linear Probability Plot, Normal Probability Plot, and Probability Plot. These were used in order to test the reliability of the variables used in the study, especially the 4 'As' of tourism which include, amenities, accommodation, accessibility, and attraction. It was found out that the use of this technique was found to be significant. Residual Plot Analysis was used in the study and found out to be effective and accurate. As proven, the results of the study would validate that when plots are scattered in the horizontal axis would mean that the analysis is found to be precise. Line Fit Plot Analysis was also used in the study, which proved that if there are increasing values of scattered points, it shows significant correlation between the independent and dependent variables. Normal Probability Plot was further utilized where it was proven that if the data are plotted in a theoretical normal and typical distribution in the sample percentile with regards to the number of tourists' arrivals form an approximate or estimated straight line. The general objective of the study was to develop an online application that will predict the tourists' arrival using regression techniques. **Specific Objectives:** 1. Apply linear regression to predict the tourist arrival per year. 2. Implement multiple linear regression to determine the factors affecting tourists' arrival. 3. Integrate these regression techniques to the developed online application. **Findings:** Linear Regression Analysis was used to predict the arrivals of tourists and was found out to be precise and reliable. Multiple Linear Regression Analysis was also utilized to determine the factors affecting tourists' arrivals. This was used in order to test the reliability of the variables used in the study, especially the 4 'As' of tourism which include, amenities, accommodation, accessibility, and attraction. The results proved that the use of this technique was found to be reliable and significant.*

**Keywords:** Tourist, Prediction, Data Mining, Linear Regression, Multiple Linear Regression

## 1. Introduction

Precise prediction of international and local tourists that will visit tourism spots in the future from its major country sources is a conclusive obligatory function for effective tourism development (Cheng Li et al., 2020). For a known country, like the Philippines, which has all the potentials to become a force to surmise in international tourism, particularly in our neighbors in Asia, effectual planning and decision making are really needed (Nafte, 2017). The administration on its part is not in a slipshod in coming up with tourism and other related plans and strategies over the years, yet, the country, has been padding in this field over other countries in the region – as one of the world's fastest mounting tourism destinations (Rufino, 2018).

According to the data from the Department of Tourism, there is an increase of 1, 092.446 of tourist arrival in the country from year 2018 to 2019. The newest merit is when El Nido Palawan found in the southwest part of Manila, Philippines is considered to be the best island in world for year 2020 by the world-renowned travel publication Travel + Leisure (Clemence, S.2020). Also, in the study of (Crisostomo, A. and Gustillo, R 2019) titled "Tourism: Asian Country Ranking Using Analytic Hierarchy Process" the Philippines ranked number five as favorite tourist destination among the Asian countries.

The number of tourists has significant impacts on our economy. These impacts can be considered as both negative

and positive. The most significant outcome of tourism on economy can be documented as number of changes on supply and required chain in the destination which is the host of tourists. Tourists' demand or simply tourist expenditure contributes to GDP, increasing the employment rate, making new source of revenue for local people, private and public sectors and destination's government and so on (Faladeobalade and Dubey, 2019).

Moreover, the tourism business is one of the most critical sectors for a flourishing economy as it reflects for a big area of some countries' Growth Domestic Product (GDP) and service figures (OECD, 2018). Tourism is regarded as by great differences in numbers on a yearly basis, and, as a result, predicting upcoming arrivals is a very complicated job. Predictions of tourists' influx are essential for setting up, strategy making, and budgeting purposes (Dezan Shira and Associates, 2019). Furthermore, tourism has become a major sector for economic development and provides a good implications on public revenue as tourist expenditures are evenly distributed in various sectors (Rizal, A. and Hartati, S.2016). (Xin, Y. et al.2016) considered tourism as one of the main source of foreign exchange earnings and that the growth in the tourism section will promote an increase in the country's overall income (Kan, M. et al.2016). The extreme expectation for tourism to promote improvement on the country's economic strength requires accurate prediction and planning to ensure sustainable demand of the needs of the tourism industry (Jun, Y. et al.2018).

Additionally, no study to our knowledge has attempted to predict future tourists' arrivals in Boracay, which is one of the most well-liked tourist destinations worldwide. This notion prompted to conduct such in order to help improve the services offered by concerned government agencies.

The system would predict the number of tourists' arrival in Boracay. The system would also provide reports which are needed by the Local Tourism Office especially in the protection and preservation of the island.

This study, adopted the Data Mining techniques of the tourists' arrival in Boracay, Malay, Aklan, Philippines using Regression techniques. It is said that, regression analysis is one of the most commonly used tools in research as it allows analyzation of relationships between dependent and independent variables and provides beneficial information such as significant relationship between these variables and most importantly prediction (Amy Gallo, 2016).

## 2. Methodology

### 2.1 Research Design

The researcher adopted the developmental type of research to achieve the purpose of the study. This method is a fact-finding study that encompasses sufficient and precise interpretation of data and their findings. This approach is appropriate wherever the object of a class varies among themselves and one is interested in knowing the extent to which different conditions obtain among objects. The data from a descriptive survey was used as basis for influence that may aid in solving practical complications.

### 2.2 Participants of the Study

This study was conducted at Boracay Island, Malay, Aklan, Philippines. The researcher prepared questionnaires for 50 participants which included foreign and local tourists as well as OFWs in order to know the reasons why they chose to visit the island of Boracay based on the 4 As of tourism. The said participants were considered as part of the computation using multiple regressions. On the other hand, historical data of tourists' arrival from 2010 to 2019 were utilized in order to predict future arrival using linear regression.

### 2.3 Data Gathering Instruments and Techniques

The major instrument that was used in gathering the data was the survey questionnaire. This was the main instrument used in soliciting responses from the subject of this study. The questionnaire is consists of a sequence of questions especially on some personal details of the tourists as well as their reasons of visiting the island. The researcher also conducted an interview with the proper authorities especially at the Local Government Unit of Malay, Aklan. The purposes of the survey and interview were to determine the factors affecting the visit of tourists as well as to gather pertinent information regarding the operation of Boracay Island.

### 2.4 Preparation of Instruments

Data gathering is the most important part in conducting the study. The researcher prepared the structured questions on the instruments especially on the aspect of the 4 'As' of tourism which include: attraction, accommodation, amenities, and accessibility to be distributed to the respective participants to be able to come up with the idea and data needed in the study.

### 2.5 Validation of Instruments

An instrument is valid if it measures what it is intended to measure and accurately achieves the purpose for which it was designed. Declaration of information on the research must be reliable and certain and in order to meet this standard. The initial set of the instrument prepared by the researcher was checked and validated by the adviser and some professionals who are capable in doing the task as trial participants. After the final draft was finalized the researcher distributed the questionnaire to the selected participants.

### 2.6 Data Gathering Procedure

In gathering the data, the researcher distributed personally the analyzed and evaluated questionnaire to the participants. First, the researcher gave a request letter to the Municipal Mayor asking permission to conduct the study. Second, was the distribution of the questionnaires to the participants of the study. Lastly, was the retrieval of the questionnaire so that the researcher could start the interpretation of their responses on the Prediction of Tourists' Arrival. Through this procedure, he was able to get one hundred percent retrieval of the instrument.

### 2.7 Statistical Tools

When the questionnaires were retrieved, a table was prepared in accordance with the Prediction of Tourist's arrival using Regression techniques utilizing Multiple and Linear Regression Analyses to organize, analyze, and summarize the data needed. Ranking was also used in this study, to rank the factors affecting the tourists' arrival such as accommodation, amenities, accessibility, and attraction known as the 4 "As" of tourism.

#### 2.7.1 Weighted Mean

The weighted mean for each item was obtained by multiplying the scale value of responses by the total number of responses indicating it to get the total weighted points and dividing them by the total number or responses. The mean is the measure of central tendency. It points to where the majority of the participants' answers to a question cluster.

$$\bar{X} = \frac{\sum fx}{n}$$

Where:

**X**= Weighted Mean

**F**=Frequency

**X** =Scores

**n** = Total number of participants

$\Sigma$  = Summation symbol

**2.7.2 Likert Scale:** In the interpretation of the Weighted Mean (WM), Likert’s Scale method was used by the researcher using the following intervals and verbal interpretations. This 5-point scale was also used in order to determine the rank or the adjectival description of the weighted mean of the responses for the proposed Prediction Model. The fields represent the rating, range, and the adjectival description for each rating. These ratings are:

**Table 1**

| Unit Weight | Verbal Interpretation |
|-------------|-----------------------|
| 5           | Excellent             |
| 4           | Great                 |
| 3           | Good                  |
| 2           | Fair                  |
| 1           | Poor                  |

**Likert Scale with Verbal Interpretation**

The table above shows the Likert Scale of the weighted mean. The weighted mean is categorized in five weighted points. The verbal interpretation for the 5-point scale is shown in the table.

**2.7.3 Ranking**

This was used to get the rank average for each answers choice and determine which ever is the highest and lowest rank based on the results.

**2.8 Context Diagram**

This diagram was used to define the boundaries of the different processes of taking a visit to the Island of Boracay. It is a higher view of Tourists’ Arrival Prediction.

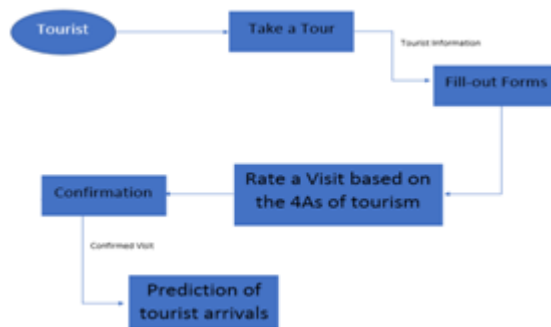


**Figure 2**

**Context Diagram of Predicting Tourist’s Arrival**

**2.9 Data Flow Diagram**

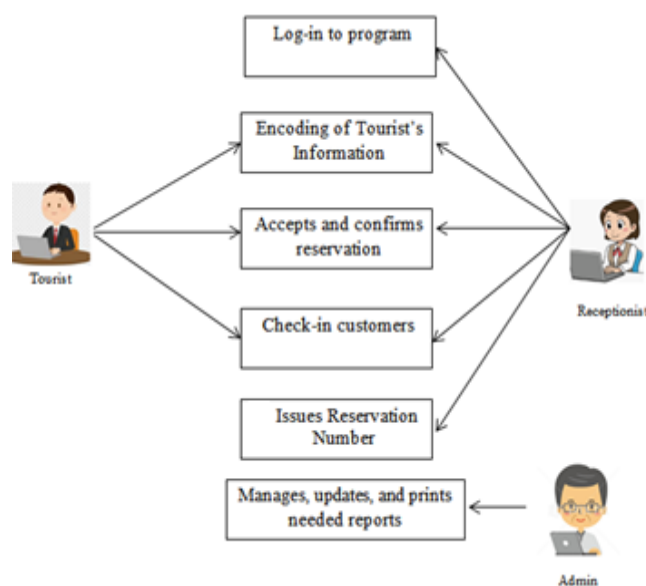
The graphical representation of the flow of data that was used to predict the tourists’ arrival in the Island of Boracay.



**Figure 3: Data Flow Diagram**

**2.10 Use Case Diagram**

This diagram was used to show the relationships between tourists and use cases within the software. The process will measure the number of tourists both foreign and local who visited the island of Boracay.



**Figure 4**

**Use Case Diagram of the Prediction Model**

**Tourist:** Encodes customer’s information, accepts as well as confirms reservation, and allows check-in.

**Receptionist:** Logs in to the program, encodes customer’s information, accepts as well as confirms reservation, checks-in customers, and issues reservation number.

**Admin:** Manages, updates data, and prints needed reports.

**2.11 Software Model**

This section gives a description of the methods used in developing the proposed system. The Prediction Model was developed using the Rapid Application Development (RAD) model.

The researcher used this model because the application’s requirement is very well documented, fixed, and clear.

The figure below shows the process of Rapid Application Development; (1) Requirements Planning, (2) User Design, (3) Construction, and (4) Cutover.

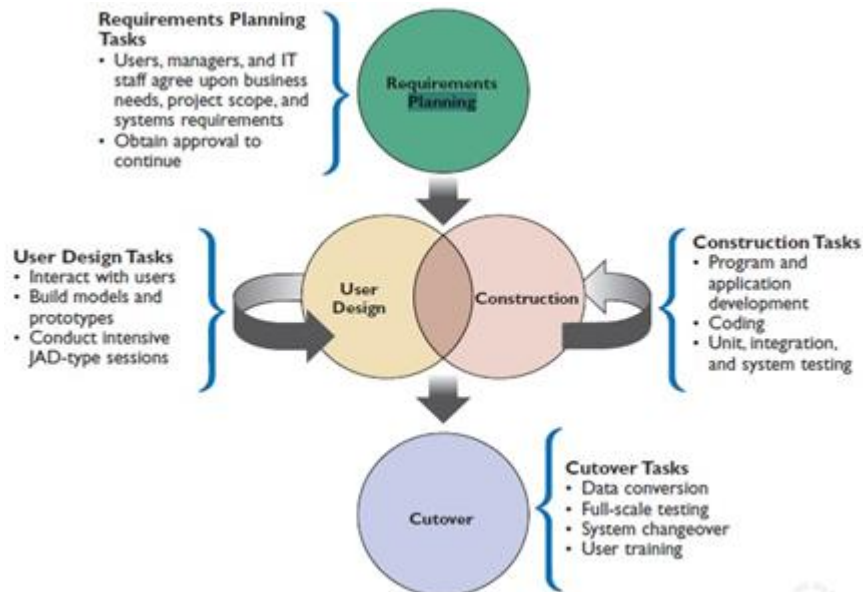


Figure 5

### Rapid Application Development Model

Rapid application development (RAD) is a team-based technique that speeds up information systems development and produces a functioning information system. Like JAD, RAD uses a group approach, but goes much further. While the end product of JAD is a requirements model, the end product of RAD is the new information system. RAD is a complete methodology, with a four-phase life cycle that parallels the traditional SDLC phases. Companies use RAD to reduce cost and development time, and increase the probability of success.

Listed below are the four phases for RAD model:

**Requirements Planning:** The **requirements planning phase** combines elements of the systems planning and systems analysis phases of the SDLC. This phase requires intense involvement from Users, managers, and IT staff members to discuss and agree on business needs, project scope, constraints, and system requirements. The requirement planning phase focus always remains on reaching the goals and end when the team agrees on the key issues and obtains management authorization to continue.

**User Design:** During the **user design phase**, users interact with systems analysts and develop models and prototypes that represent all system processes, outputs, and inputs. The RAD group or subgroups typically use a combination of JAD techniques and CASE tools to translate user needs into working models. User design phase is a continuous, interactive process that allows users to understand, modify, and eventually approve a working model of the system that meets their needs.

**Construction:** The **construction phase** focuses on program and application development tasks similar to the SDLC. In RAD, however, users continue to participate and still can suggest changes or improvements as actual screens or reports are developed.

**Cutover:** The **cutover phase** resembles the final tasks in the SDLC implementation phase, including data conversion, testing, changeover to the new system, and user training. Compared with traditional methods, the entire process is compressed. As a result, the new system is built in precise manner, delivered, and placed in operation much sooner.

### 2.12 Model Description

Regression analysis techniques utilized multiple linear regressions as shown in equation 1 wherein the independent variables to be used are attraction, accommodation, amenities, and accessibility.

$$Y = mx_1 + mx_2 + mx_3 + b$$

Where:

Y= the dependent variable of the regression

M= slope of the regression

X<sub>1</sub>=first independent variable of the regression

The x<sub>2</sub>=second independent variable of the regression

The x<sub>3</sub>=third independent variable of the regression

B= constant

Linear regression as reflected in equation 2 was used to predict the tourist arrival per year wherein the independent variable to be used is the number of tourist per year.

$$Y = a + bX$$

Where:

N = number of observations, or years

X = year index

Y = population size for given census years

### 3. Results and Discussion

The architecture and design of the application were based on the researchers' concepts, vision, and judgment of the members of the panel and suggestion as well as the demands of the participants.

The model used by the researcher for the development of the system that requires a highly interactive or complex user interface was the Rapid Application Development (RAD), because it is a complete methodology that relies on prototyping and involvement of the user in every phase of the development that helps to reduce cost and development time, and increase the probability of success. While the development was still on-going there was a constant communication between the developer and the office of the

Tourism Officer of Malay, Aklan for the improvement of the study.

Regression analysis was used in this study. Multiple and Linear Regression Analyses were utilized in order to come up with accurate results especially in the prediction of tourists' arrival.

The development of a Prediction model was utilized in order to predict the tourists' arrival in Boracay.

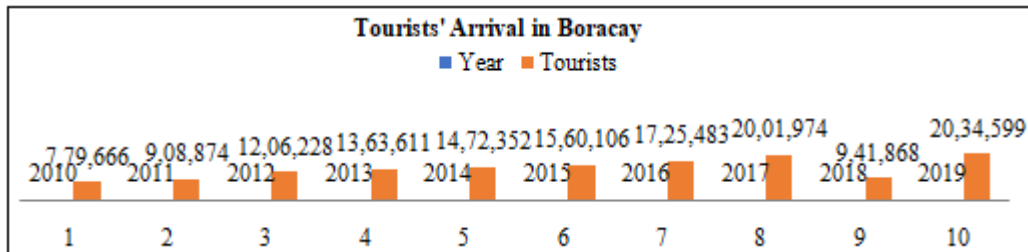


Figure 6

**Distribution of Tourists' Arrival from 2010-2019**

The figure above presents the tourists' arrival in Boracay Island from 2010 to 2019. It shows the tourists' arrival from 2010 to 2017 was increasing in number. As computed the average rate of arrival from the said periods of time is increasing. This further shows that the island of Boracay has still the power of captivating the hearts of tourists. It was in

2018 where there was a drop of tourists who visited the island because of its closure due to rehabilitation which was considered as government intervention. The arrival gained its momentum in 2019, where tourists' arrival had a significant increase compared to the previous years.

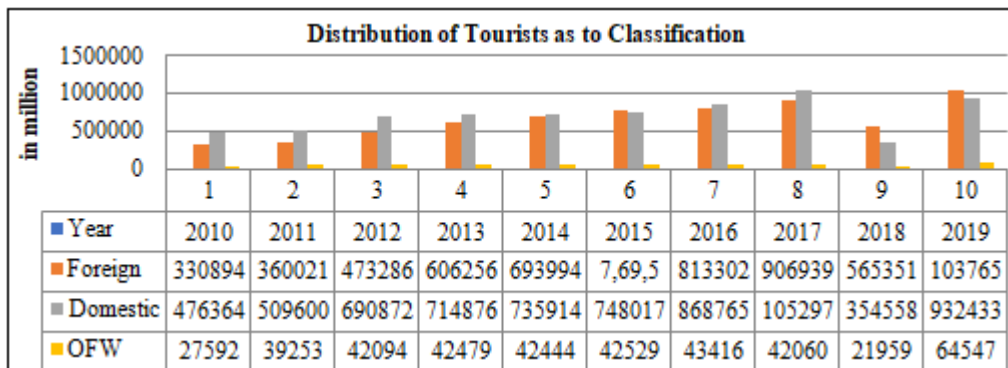


Figure 7

**Distribution of Tourists' Arrival as to Classification from 2010-2019**

The figure above shows that domestic tourists dominated the arrival in 2010 to 2014 and 2016 to 2019. While in 2015, foreign tourists surpassed the arrival of tourists over

domestic and Overseas Filipino Workers. On the other hand, OFWs maintained its position to always be on the last spot compared to Foreign and Domestic tourists.

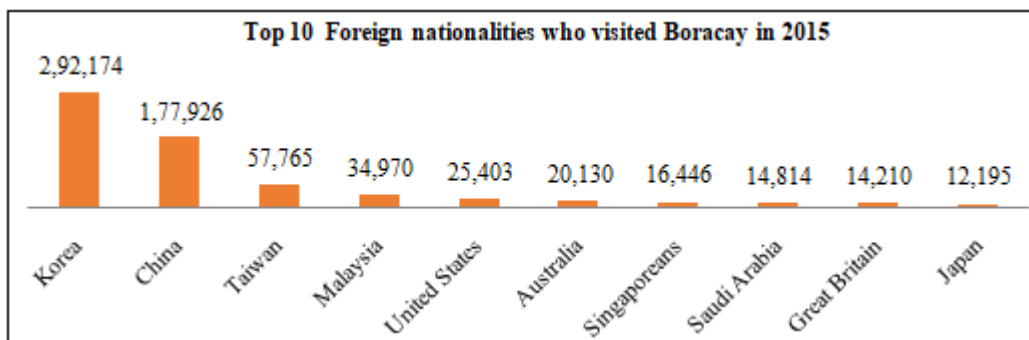


Figure 8

**Top 10 Foreign nationalities who visited Boracay in 2015**  
 Figure 8 shows that tourists coming from Korea made it on the top spot beating other nationalities. Almost 300 thousand Koreans visited the island, Chinese with more or less 180 thousand, Taiwanese with almost 58 thousand, while

Japanese came on the last spot with more or less 12, 200.completing the top ten were Americans, Australians, Saudi Arabians, and British who came in fourth, fifth, sixth, seventh, eighth, and ninth places respectively.

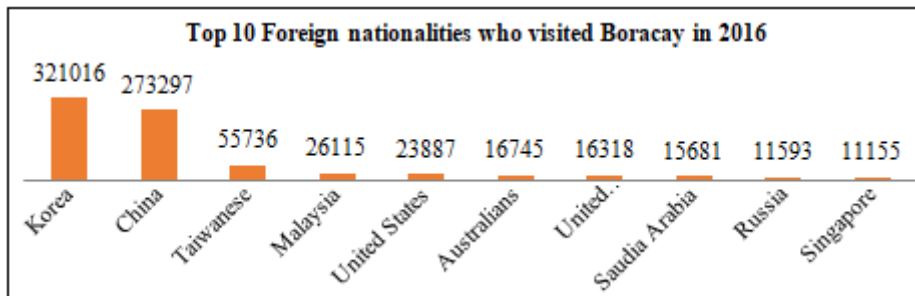


Figure 9

**Top 10 Foreign nationalities who visited Boracay in 2016**  
 Figure 9 shows that tourists coming from Korea made it on the top spot beating other nationalities. Almost 300 thousand Koreans visited the island, Chinese with more or less 274 thousand, Taiwanese with almost 58 thousand, while

Japanese came on the last spot with more or less 12, 200.completing the top ten were Americans, Australians, Saudi Arabians, and British who came in fourth, fifth, sixth, seventh, eighth, and ninth places respectively.

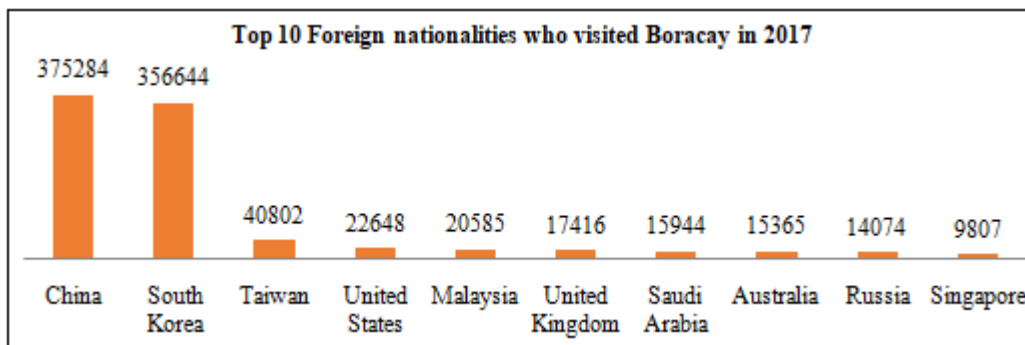


Figure 10

**Top 10 Foreign nationalities who visited Boracay in 2017**  
 The figure above shows that tourists coming from Korea made it on the top spot beating other nationalities. Almost 380 thousand Chinese visited the island, Koreans with more or less 180 thousand, Taiwanese with almost 58 thousand,

while Japanese came on the last spot with more or less 12, 200.completing the top ten were Americans, Australians, Saudi Arabians, and British who came in fourth, fifth, sixth, seventh, eighth, and ninth places respectively.

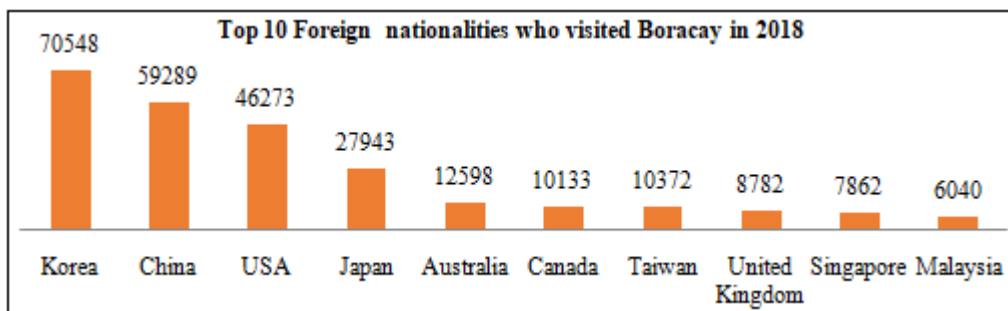


Figure 11

**Top 10 Foreign nationalities who visited Boracay in 2018**  
 Figure 11 shows that tourists coming from Korea made it on the top spot beating other nationalities. Koreans, a little more than 70 thousand visited the island, Chinese with more or less 6 thousand, Americans with almost 46 thousand, while Japanese with more or less 27 thousand. Completing

the top ten were Australians, Canadians, Taiwanese, British, Singaporeans, and Malaysians who came in fifth, sixth, seventh, eighth, ninth, and tenth places respectively.

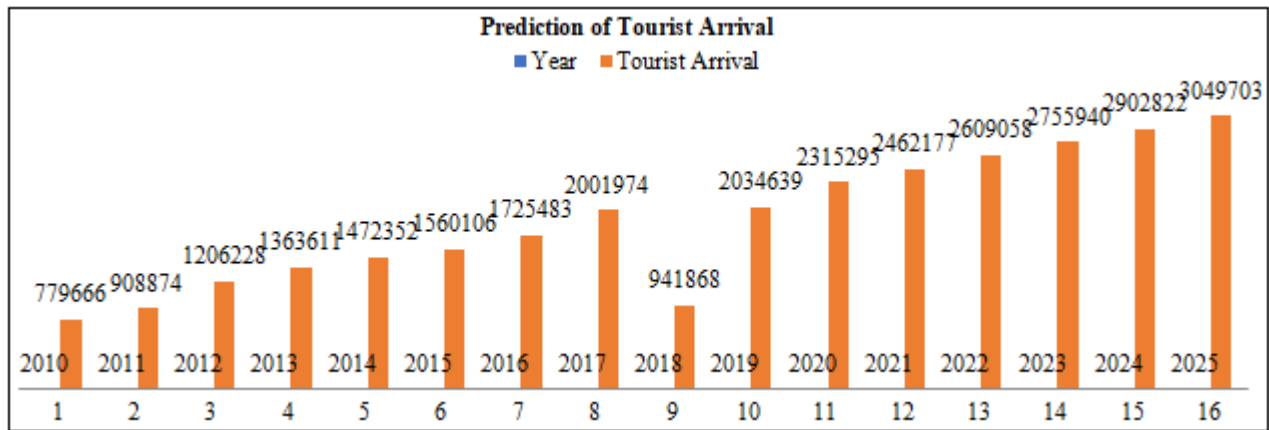


Figure 12

**Prediction of Tourists' Arrival**

The figure above shows the prediction of tourists' arrival using linear regression analysis. With the consideration of the data from 2010-2019, it is noted that there is an increase of tourists' arrival from 2020 to 2025. The projection is set to 5 – 10 years for more precise results. In this study, it was set to 6 years, until 2025.

**Table 2**

| Linear Regression Statistics |             |
|------------------------------|-------------|
| Multiple R                   | 0.981097509 |
| R Square                     | 0.962552321 |
| Adjusted R Square            | 0.957202653 |
| Standard Error               | 90885.21482 |
| Observations                 | 9           |

**Linear Regression Statistics**

The table above shows the regression of 0.98109 (98%) which shows that the independent variable and the

dependent variable are correlated with each other. The data in the table, considered 9 observations, from the year 2010-2019. The R square value of 0.96255 (96%) shows a very high prediction of tourists' arrival. This shows that the model is a very good predictive model in predicting tourists' arrival.

**Table 3**  
**ANOVA**

|            | df | SS          | MS       | F       | Significance F |
|------------|----|-------------|----------|---------|----------------|
| Regression | 1  | 1.48622E+12 | 1.49E+12 | 179.927 | 3.00155E-06    |
| Residual   | 7  | 57820855915 | 8.26E+09 |         |                |
| Total      | 8  | 1.54404E+12 |          |         |                |

**Test of Statistical Significance of the Model**

The table above shows the statistical significance of the forecasting model using linear regression analysis. The F-value of 3.00155E-06 is considered significant because it is below the normal value of 0.05. This therefore, shows that the model is statistically significant.

**Table 4**

|           | Coefficients | Standard Error | t Stat   | P-value     | Lower 95%    | Upper 95%    | Lower 95.0% | Upper 95.0%  |
|-----------|--------------|----------------|----------|-------------|--------------|--------------|-------------|--------------|
| Intercept | -294385533.4 | 22054767.46    | -13.3479 | 3.10305E-06 | -346536771.4 | -242234295.4 | -346536771  | -242234295.4 |
| year      | 146881.5984  | 10950.11419    | 13.4137  | 3.00155E-06 | 120988.6928  | 172774.504   | 120988.6928 | 172774.504   |

**Degree of Coefficients of the Independent and Dependent Variables**

Table 4 shows the intercept of the independent and dependent variables of -294385533.4 where the line of best fit is intersecting with the y-axis. It is also reflected that the p-value of the independent variable is, 3.00155E-06 which states that this value is lower than the standard value of 0.05, which further connotes that the said variable is really significant.

**Residual Plot of Forecasting Tourists' Arrival**

Figure 13 shows the residuals on the vertical axis and the independent variable on the horizontal axis. It is reflected that the points in the residual plot are randomly dispersed from the year 2018-2019 (i. e., the horizontal axis), which simply tells that the linear regression is appropriate for the data. This further connotes that the model is appropriate to be used and in defining and examining residual plots as well as in investigating potential deviations from the normality of assumptions.

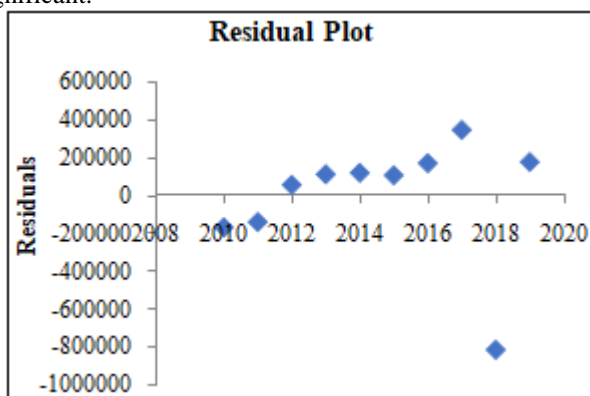


Figure 13

The results are in congruence with the findings of Law and Martin, 2015, that producing residual plots is a technique for checking model assumptions, and that a common use of such plots is to check for a linear trend and is highly advisable in applied work.

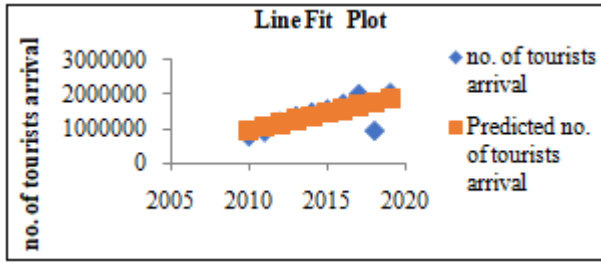


Figure 14

**Line Fit Plot of Predicting Tourists' Arrival**

The figure above shows the line of best fit through a scattered plot of data points that best expresses the relationship between these points. It is noted that the scattered points representing the number of tourist arrivals from 2010 to 2019 (not in horizontal slopes) are connected with a line of best fit showing powerful significant relationships.

This is contrast to the study made by Casey, 2016, wherein it was found out that the study with a horizontal slope is frequently indicative of a lack of linear association between two variables, as well as broader purposes for best fit lines (i. e., that fitting lines serve more than just descriptive purposes). While it is perhaps not sensible to fit a line to non-linear data, the inferences based on slope that can be made are powerful.

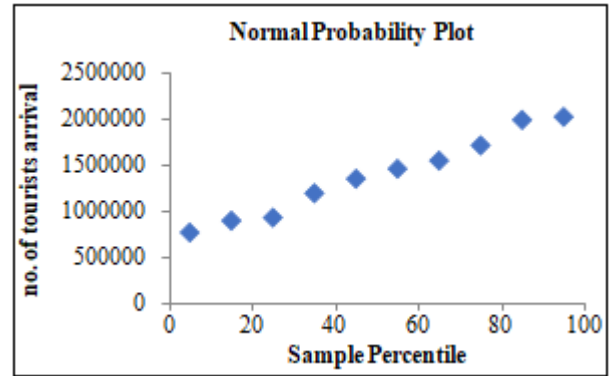


Figure 15

**Normal Probability Plot of Predicting Tourists' Arrival**

The figure above shows a normal probability plot which indicates that the data fit the normal distribution. The data are plotted in a theoretical normal distribution wherein the points from the sample percentile with regards to the number of tourists' arrivals form an approximate straight line.

The results of the study are similar with the findings of Chantarangsi, et. al., (2016), that a normal probability plot can be augmented by intervals for all the points if the population distribution is normal, then all the points should fall into the corresponding intervals simultaneously.

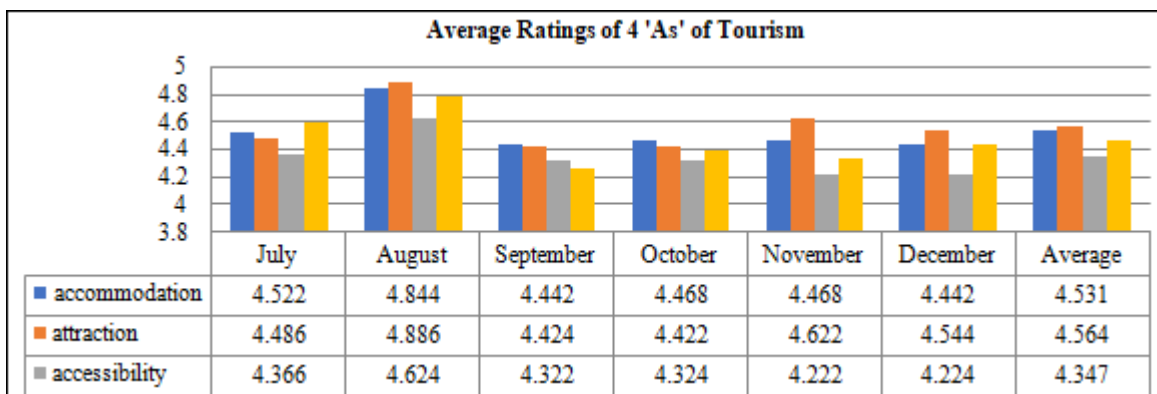


Figure 16

**Average Ratings of 4 'As' of Tourism**

The figure above shows the average ratings of 4 'As' of Tourism from July to December 2019. These include accommodation, attraction, accessibility, and amenities. It was found out that attraction ranked number 1 with an

average weighted mean of 4.564 followed by accommodation, amenities, and accessibility with average means of 4.531, 4.469, and 4.437 respectively. The results showed that accessibility ranked last maybe because of the mode of transportation going to Boracay.

Table 5

| <i>Multiple Linear Regression Statistics</i> |           |
|--|-----------|
| Multiple R                                   | 0.999999  |
| R Square                                     | 0.999979  |
| Adjusted R Square                            | 0.9999895 |
| Standard Error                               | 119.38484 |
| Observations                                 | 6         |

**Multiple Linear Regression Statistics**

The table above shows the regression of 0.99999895 (99.99%) which shows that the independent variable and the dependent variable are correlated with each other. The R

square value of.999997899 (96%) shows a very high prediction of tourists' arrival. This shows that the model is a very good predictive model in forecasting tourists' arrival. The data in the table considered 6 observations.



Table 6

| ANOVA      |    |            |             |             |                |
|------------|----|------------|-------------|-------------|----------------|
|            | df | SS         | MS          | F           | Significance F |
| Regression | 4  | 6785342362 | 1696335591  | 119018.2163 | 0.002173971    |
| Residual   | 1  | 14252.7391 | 14252.73914 |             |                |
| Total      | 5  | 6785356615 |             |             |                |

**Statistical Significance of the Model**

The table above shows the statistical significance of the forecasting model using multiple linear regression analysis. The F-value of 0.002173971, is considered significant

because it is below the normal value of .05. This therefore, shows that the forecasting model is statistically significant.

Table 7

|               | Coefficients | Standard Error | t Stat     | P-value   | Lower 95%  | Upper 95%  | Lower 95.0% | Upper 95.0% |
|---------------|--------------|----------------|------------|-----------|------------|------------|-------------|-------------|
| Intercept     | -1025249.6   | 2287.4306      | -448.21014 | 0.0014204 | -1054314.1 | -996185.03 | -1054314.1  | -996185     |
| accommodation | -223261.47   | 9906.7083      | -22.536392 | 0.02823   | -349138.13 | -97384.802 | -349138.13  | -97384.8    |
| attraction    | 157768.78    | 3639.837       | 43.345012  | 0.0146847 | 111520.26  | 204017.29  | 111520.26   | 204017.3    |
| accessibility | 251969.82    | 6095.0339      | 41.340184  | 0.0153965 | 174525.07  | 329414.57  | 174525.07   | 329414.6    |
| amenities     | 69730.629    | 996.31802      | 69.988325  | 0.0090955 | 57071.209  | 82390.05   | 57071.209   | 82390.05    |

**P-values of the Independent Variables**

Table 5 shows the intercept of the independent and dependent variables of -1025249.588 where the line of best fit is intersecting from x-axis with respect to the y-axis. It is also reflected that the p-values of the independent variable which are, accommodation, attraction, accessibility, and amenities, 0.028230004, 0.014684662, 0.015396536, and 0.009095466, respectively, state that these values are lower than the standard value of .15, which further connotes that the said variables are really significant.

plot are randomly dispersed from 4.4 to 4.9 (i. e., the horizontal axis), which simply tells that the linear regression is appropriate for the data. This further connotes that the model is appropriate to be used in defining and examining the residual plots.

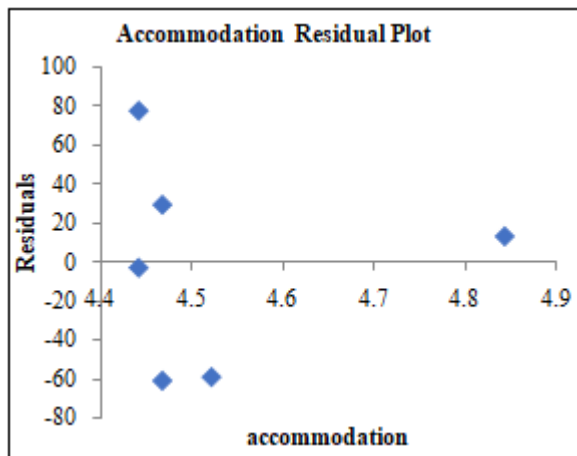


Figure 17

**Accommodation Residual Plot of 4 'As' of Tourism**

Figure 17 shows the residuals of the dependent variable on the vertical axis and the independent variables on the horizontal axis. It is reflected that the points in the residual

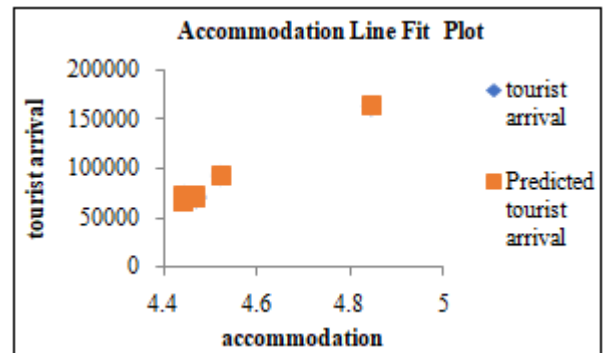


Figure 18

**Accommodation Line Fit Plot for Accommodation**

The figure above shows the line of best fit through a scattered plot of data points that best expresses the relationship between these points. It is noted that the scattered points representing the accessibility with mean averages of 4.4 to 5.0 are increasing, therefore, showing high correlation between the independent and dependent variables. As stated, the correlation is positive when the values increase.

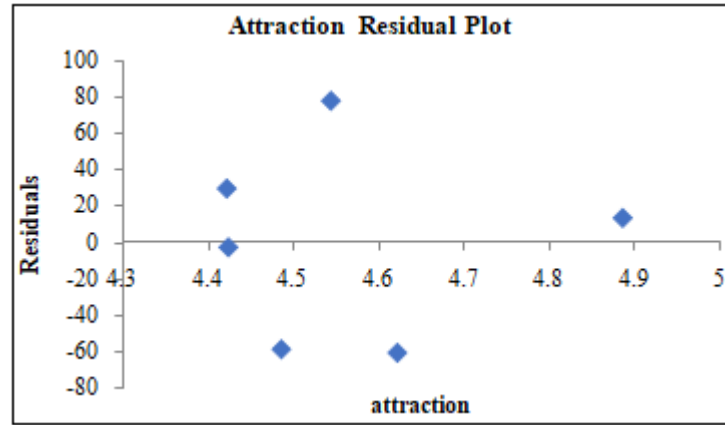


Figure 19

**Attraction Residual Plot**

Figure 19 shows the residuals on the vertical axis and the independent variable on the horizontal axis. It is reflected that the points in the residual plot are randomly dispersed from 4.4 to 4.85 (i. e., the horizontal axis), which simply tells that the linear regression is appropriate for the data. This further signifies that the model is appropriate to be used in defining and examining residual plots.

**Accessibility Residual Plot**

Figure 21 shows the residuals of the dependent variable on the vertical axis and the independent variables on the horizontal axis. It is reflected that the points in the residual plot (in the horizontal axis) are randomly dispersed from 4.4 to 4.7, which simply expresses that the multiple linear regression is appropriate for the data. This further connotes that the model is appropriate to be used and in defining and examining residual plots.

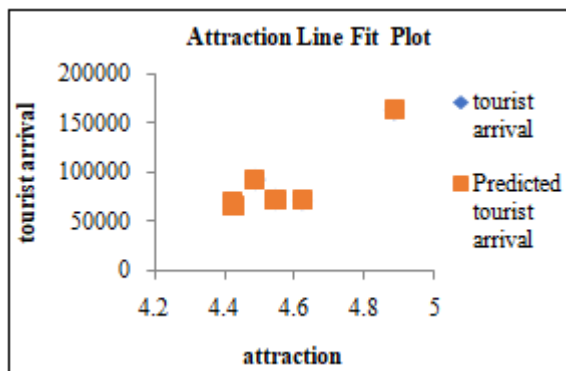


Figure 20

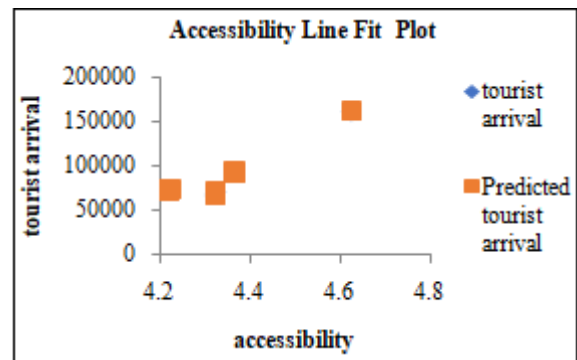


Figure 22

**Attraction Line Fit Plot**

The figure above shows the line of best fit through a scattered plot of data points that best expresses the relationship between these points. It is noted that the scattered points representing attraction with mean averages of 4.2 to 5.0 can still be associated because more scattered points are increasing, therefore, showing significant correlation between the independent and dependent variables.

**Accessibility Line Fit Plot**

The figure above shows the line of best fit through a scattered plot of data points that best expresses the relationship between points. It is noted that the scattered points representing the accessibility with mean averages of 4.2 to 4.8 can still be connected because more scattered points are increasing, therefore, showing significant correlation between the independent and dependent variables.

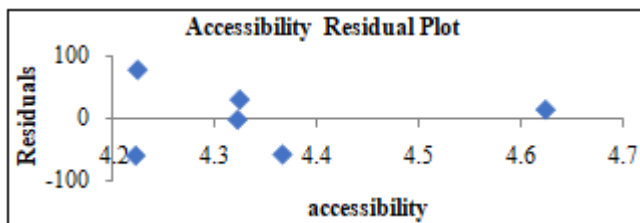


Figure 21

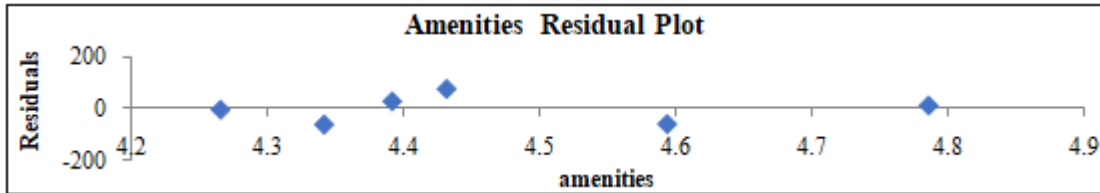


Figure 23

**Amenities Residual Plot**

Figure 23 shows the residuals on the vertical axis and the independent variable on the horizontal axis. It is reflected that the points in the residual plot are randomly dispersed from 4.4 to 4.85 (i. e., the horizontal axis), which simply

tells that the multiple linear regression is appropriate for the data. This further connotes that the model is appropriate to be used in defining and examining the residual plots.

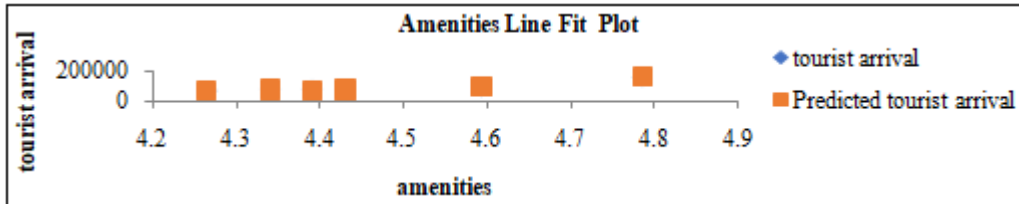


Figure 24

**Amenities Line Fit Plot**

Figure 24 shows the line of best fit through a scattered plot of data points that best expresses the relationship between points. It is noted that the scattered points representing amenities, with mean averages of 4.2 to 5.0 can still be

connected because more scattered points are increasing, therefore, showing significant correlation between the independent and dependent variables.

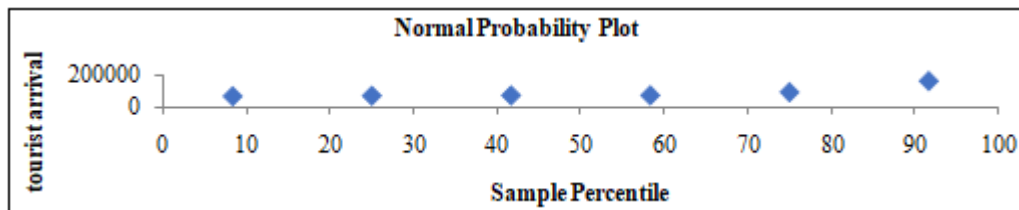


Figure 25

**Normal Probability Plot of the Components of 4 ‘As’ of Tourism**

Figure 25 shows a normal probability plot which indicates that the data fit the normal distribution. The data are plotted in a theoretical normal and typical distribution wherein the points from the sample percentile with regards to the number of tourists’ arrivals form an approximate or estimated straight line.

Tourism (DOT) and the Department of Environment and Natural Resources (DENR) through the Local Government Unit of Malay, Aklan should be provided with the necessary data to for them to oversee the operation of Boracay Island. Maintenance of the system should be done regularly so as to maintain its functionality especially that the tourists’ arrivals are expected to increase in the years to come.

**4. Recommendations**

It is recommended that other future researchers may use other algorithms in order to test and compare the results with the regression techniques that were used in this study. The office of the tourism officer of the Municipality of Malay, Aklan should have this kind of system in order to predict the arrivals of tourists in the island of Boracay so that it will be provided with real time data and would be properly aware of the predicted tourists’ arrival for preparedness purposes. The officials of the Local Government Unit of Malay, Aklan, as one of the decision-makers, should supervise and monitor the operations of the island by having accurate data generated by the system needed in the decision-making, especially in the management and control of the island as well as in the appropriation of funds. The Department of

**5. Conclusion**

Linear Regression Technique was used to predict the arrivals of tourists and was found out to be precise and reliable. Multiple Linear Regression Technique was also utilized to determine the factors affecting tourists’ arrivals. This was used in order to test the reliability of the variables used in the study, especially the 4 ‘As’ of tourism which include, amenities, accommodation, accessibility, and attraction. The results proved that the use of this technique was found to be reliable. Residual Plot Analysis was used in the study and found out that this analysis was found to be effective and accurate. As proven, the results of the study would validate that when plots are scattered in the horizontal axis would mean that the analysis is found to be precise. Line Fit Plot Analysis was also used in the study, which proved that if there are increasing values of scattered points,

it shows significant correlation between the independent and dependent variables. Normal Probability Plot was further utilized where it was proven that if the data are plotted in a theoretical *normal* and typical distribution in the sample percentile with regards to the number of tourists' arrivals form an approximate or estimated straight line. In relation to the foregoing conclusions, regression analysis techniques have significant contribution in the development of a tool in predicting tourists' arrival. The study further proved that tourist reservations are very important in the operation of the island and in order to observe its carrying capacity every day. Furthermore, regression analysis techniques are considered effective in prediction purposes.

## References

- [1] Andrew Saayman, Ilse Botha.2015. "Non-linear models for tourism demand forecasting. " *Tourism Economics* 23 (3): 594-613.
- [2] Artola, C., Pinto, F., & Garcia, P. D. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower*, 36, 103–116.
- [3] Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46, 454–464.
- [4] Baldigara, T., and M. Koić.2015. "Modelling the International Tourism Demand in Croatia Using a Polynomial Regression Analysis. " *The Business of Tourism* 15: 29–38
- [5] Baldigara, T., and M. Mamula.2015. "Modelling international Tourism Demand Using Seasonal ARIMA Models. " *Tourism and Hospitality Management* 21 (1): 19–31.
- [6] Claveria, O., Monte, E., & Torra, S. (2016).combination forecasts of tourism demand with machine learning models. *Applied Economics Letters*, 23, 428–431. Croce, V. (2017). Business confidence and international tourism Demand: Evidence from a global panel of experts. *Global Journal of Management and Business Research*, 16 (1), 29–42.
- [7] Chikobvu, Makoni Tendai and Delson. (2017). "Modelling international tourist arrivals and volatility to the Victoria Falls Rainforest, Zimbabwe: Application of the GARCH family of models. " *African Journal of Hospitality, Tourism and Leisure* 6 (4).
- [8] Chigora, F., and C. Vutete. (2015). "Comprehending Determinants of Demand: Zimbabwe Tourism Destination Scenario. " *International and Business* 1 (2): 17–19.
- [9] Clemence, S. (2020). "Travel + Leisure". 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) . 10.1109/HNICEM48295.2019.9072840
- [10] Clemence, S. (2020). "Travel + Leisure". 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) . 10.1109/HNICEM48295.2019.9072840
- [11] Jungmittag, A. (2016).combination of forecasts across estimation windows: An application to air travel demand. *Journal of Forecasting*, 35 (4), 373–380. Law, R., & Au, N. (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management*, 20, 89–97.
- [12] Katerina Volchek, School of Hotel and Tourism Management, The Hong Kong Polytechnic University, 17, Science Museum Road, TST, Kowloon, Hong Kong. Email: katerina.volchek[at]connect.polyu.hk, November 2018, DOI: 10.1177/1354816618811558
- [13] Li, X., Pan, B., Law, R., & Huang, X. K. (2017). Forecasting tourism demand with composite search index. *Tourism Management*, 59, 57–66. Li, Z. C., & Sheng, D. (2016). Forecasting passenger travel demand for air and high-speed rail integration service: A case study of Beijing-guangzhou corridor, China. *Transportation Research Part A: Policy and Practice*, 94 (1), 397–410.
- [14] Musundi, S. W., P. M. M'mukiira, and F. Mungai.2016. "Modelling and Forecasting Kenyan GDP using Autoregressive Integrated Moving Average (ARIMA) Models. " *Science Journal of Applied Mathematics and Statistics* 4 (2): 64–73. <https://doi.org/10.11648/j.sjams.20160402.18>.
- [15] Nafté, Anthony. "Up, up and way: ASEAN tourism" Published by Manila Bulletin. January 16, 2017.
- [16] Ndiege, B. O.2015. "An Analysis of the Performance of International Tourism Demand in Tanzania. " *ARA Journal* 5 (1): 53–62.
- [17] Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research*, 56 (7), 957–970.
- [18] Peng, G., Liu, Y., Wang, J., & Gu, J. (2017). Analysis of the prediction capability of web search data based on the HE-TDC method-prediction of the volume of daily tourism visitors. *Journal of Systems Science and Systems Engineering*, 26 (2), 163–182.
- [19] Saayman, A., and I. Botha.2015. "Evaluating Non-linear Approaches in Forecasting Tourist Arrivals. " *ERSA Working Paper* 492. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.667.741&rep=rep1&type=pdf>.
- [20] Sood S, Jain K (2017) Comparative Analysis of Techniques for Forecasting Tourists' Arrival. *J Tourism Hospit* 6: 285. doi: 10.4172/2167-0269.1000285