COLLEGE OF BUSINESS & ECONOMICS

AFRICAN CENTER OF EXCELLENCE IN DATA SCIENCE

ACE-DS

# PREDICTION OF TOURISM REVENUES IN RWANDA:

# A CASE STUDY OF VOLCANO AND NYUNGWE NATIONAL PARK

Thesis Submitted in Partial fulfillment of Academic requirements for Award of the Master's degree in Data Science (ACEDS).

**By Alfred NTAGANDA**

**Reg. No: 219013920**

**Supervisor Name: Dr. Edouard MUSABANGANJI**

**Kigali, September 2020**

COLLEGE OF BUSINESS & ECONOMICS

AFRICAN CENTER OF EXCELLENCE IN DATA SCIENCE

ACE-DS

# PREDICTION OF TOURISM REVENUES IN RWANDA: A CASE STUDY OF VOLCANO AND NYUNGWE NATIONAL PARK

Thesis Submitted in Partial fulfillment of Academic requirements for the Award of the Master's degree in Data Science (ACEDS).

**By Alfred NTAGANDA**

**Reg. No: 219013920**

**Supervisor Name: Dr. Edouard MUSABANGANJI**

**Kigali, September 2020**

**DECLARATION**

I declare that this Dissertation contains my own work except where specifically acknowledged, and it has been passed through the anti-plagiarism system and found to be complaint and this is the approved final version of the Thesis:

Alfred NTAGANDA

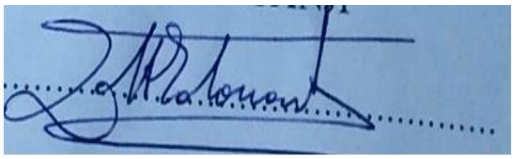
Reg. No: **219013920**

Signature


Date …31/12/2020…….

Dr. Edouard MUSABANGANJI

Signature



Date …31/12/2020…….

**ACKNOWLEDGEMENT**

Primarily, my sincere gratitude is to almighty **God** who helped me to accomplish this research sturdy.
I thank my supervisor **Dr. Edouard MUSABANGANJI** for his guidance and encouragement in the preparation and presentation of the research, His wisdom, patience and sacrifices throughout this study, greatly contributed to the accomplishment of this work.

I would like to extend my sincere appreciation to the government of Rwanda for allowing me to undertake this master's degree course and to the entire staff of the University of Rwanda especially those in the Africa Center of Excellence Data Science, especially in Data mining option for the knowledge they imparted in me.

My special thanks are addressed to the Rwanda Development Board (RDB) and Rwanda Online for the assistance given to me from the beginning of my research through provision of the data used in the study.

I cannot finish without thanking my **families, friends** and **colleagues** for any support given to me.

**SIGNATURE PAGE**

This is to certify that this Thesis has passed through the antiplagiarism system and found compliant and this is the approved final version of the Thesis:

_____

Name and signature of the Student

_____

The Thesis Committee for (Official Name of author).

_____

Name and Signature of Supervisor

_____

Name Signature of the Co-supervisor

Table of Contents

**LISTS OF TABLES**

**LIST OF FIGURES**

## LIST OF SYMBOLS AND ACCRONYMS

**ACF** Auto Correlation Function

**PACF** Partial Auto Correlation Function

**ADF** Augmented Dickey-Fuller

**AIC** Akaike Information Criterion

**SIC** Schwarz information criterion

**AR** Autoregressive

**ARIMA** Autoregressive Integrated Moving Average

**SARIMA** Seasonal Autoregressive Integrated Moving Average

**AU** Austria

**Be** Belgium

**CA** Canada

**CH** China

**DRC** Democratic republic of Congo

**FR** France

**GE** German

**IN** India

**IR** Ireland

**IT** Italy

**JP** Japan

**MA** Moving Average

**ME** Mexico

**MG** Mountain Gorillas

NE Nethlands

**NZ** Newzland

**PACF** Partial Auto Correlation Function

**RDB** Rwanda Development Board

**RW** Rwanda

**SA** South Africa

**SC** Schwarz Criterion

**SP** Spain

**SW** Switzland

**SZ** Swaziland

**UK** United Kingdom

**UNWTO** World Tourism Organization

**USA** United states of America

**USD** United States Dollars

**GARICH** generalized autoregressive conditional heteroscedastic

MARIMA multivariate ARMA

ARFIMA Autoregressive fractionally integrated moving average

## ABSTRACT

Rwanda's Tourism and Travel sector has been growing steadily at an annual rate of 9.6% between 1999 and 2018, with its significant contribution to GDP. The sector is indeed the single most contributor to the country's GDP. Its salience in national development creates the need to predict its future trend to guide policies and strategies for its promotion. Knowledge concerning future tourism revenues is crucial in guiding planning and policymaking. However, it was found that few studies have been done about tourism revenues forecasts for Rwanda. The purpose of this study was to analyze trends and patterns of tourism revenue from Volcano and Volcanoes National Parks for the years 2013-2019 and make future predictions for the next 18 months. This research used the Box–Jenkins technique to model and forecast tourism revenues. The results revealed that revenues exhibited a positive trend between 2013 and 2019, with some seasonal fluctuations. It is predicted that revenues will continue on a positive trajectory for the next 18 months until June 2021. The findings reiterate the importance of the tourism sector and the need for concerted efforts to maintain its high and growing performance. The proper management, funding and marketing of tourist sites like Nyungwe and Volcanoes National Parks will help owners of tourism-related businesses to allocate resources, prioritize marketing, advertising and devise overall strategies for the continued growth of tourism arrivals and revenues.



**KEYWORDS**

Tourism, tourism revenue, prediction, Box-Jenkins, Rwanda.

# CHAPTER ONE: INTRODUCTION

## 1.1 Background

Tourism plays a pivotal part in the development of both developing and advanced countries. Studies based on the tourism growth hypothesis designate that a positive association exists between the size of a country's tourism sector and its growth rate of Gross Domestic Product (GDP) (Brida et al., 2016). This growth role is played through inducing infrastructure improvements, job creation, domestic consumption and export diversity (Christie et al., 2014). It is not surprising therefore that the sector has received considerable policy attention in many countries, including developing countries that have shifted drastically from the Agriculture sector. Resultantly, global tourism grew substantially over the past ten years (Lenzen et al., 2018). This general trend however masked critical variations from region to region and country to country. In the case of Sub-Saharan Africa, the sector is providing opportunities for leapfrogging from agriculture directly into the service sector. Studies on the performance of the tourism sector employ varied measures ranging from tourist expenses in the destination country (Li et al., 2004), expenses on particular products like meals, exploration spending and spending (Au and Law, 2000). Additional tourism command variables revealed in this study incorporates tourism expenses. (Akal 2004), tourism occupation (Witt et al., 2004) as well as both imports and exports of tourism (Smeral, 2004).

Rwanda as a landlocked country with population of 12,926,997 million and known to be world's most reputable genocides may not attract you to be the excellent touristic visit for the next dream trip. However, tourism sector in Rwanda has shown tremendous progress over the past decade and is currently among the most favorite tourist destinations in Sub-Saharan Africa. The number of arrivals skyrocketed from 494,000 in 2006 to 932,000 in 2016, while tourism spending rose from USD 146 million to USD 377 million over the same period (World Bank, 2020). The international perception of the country has also changed for the better over the past two decades. Since 2010, Rwanda is regarded as one of the safest tourism destinations in East Africa. This reviving is related to the marketing of the country and in particular, the most popular tourist attraction – the mountain gorillas. The recovery of mountain gorilla tourism reveals that with an appropriate mix of effective policies, a post-conflict country can effectively boost its tourism sector. In addition, Rwanda has adopted a comprehensive and participatory approach to boosting the tourism sector by involving local communities (Nielsen and Spenceley, 2011).

The achievement of Rwanda tourism industry has remained prominent in the latest years. Rwanda has recognized the maximum attainments in the tourist arrivals; the total tourist arrivals enriched by 6.01% in 2013 to 7.18% in 2014. However, in the recent years like 2015, there were substantial declines in the number of arrivals and revenues. This prompted the government to take several measures to revive the sector, including a renewed focus on MICE (Meetings, Incentives, Conferences and Events) tourism as well as investment in various tourism-related infrastructure. The over-all monthly data of tourist arrivals and revenues from all over the world to Rwanda in January 2013 to December 2019 were obtained from Rwanda development board (RDB) and Rwanda online.

The original data of the total tourist revenues showed a positive trend starting from 2003. However, 2013 was selected as a starting point for analysis in this study because statistical records for the previous years are not so reliable. This is because the two national parks were not keeping proper records and did not have a good data capture template. Instead, many records were kept in hard copies and some revenues and arrivals were not accurately recorded. It is well known that Rwanda has been showing a great advancement in tourism to the globe. The total tourist arrivals to Rwanda have increased in present years. The overall tourist arrivals to Rwanda from all countries of the world in the years of 2018, 2019 and later 2020 increased due to investments made and partly as a result of signing contract with Arsenal Premier League team in first division and later with Paris Saint-Germain. These initiatives shows great achievement for Rwanda tourism industry (Jauhari, 2018).

Rwanda as hilly country has diverse touristic attractions, which are complemented by the beautiful environment and culture. The country has touristic opportunities that contribute to its growth. These potentials were recognized around 1960 to 1970 by American primatologist and conservationist called Dian Fossey. However, later in the years during and after 1994, some sites were ruined and some of the animals in the national parks were killed while others evacuated to the neighboring countries like Uganda and DRC due to Genocide against Tutsi.

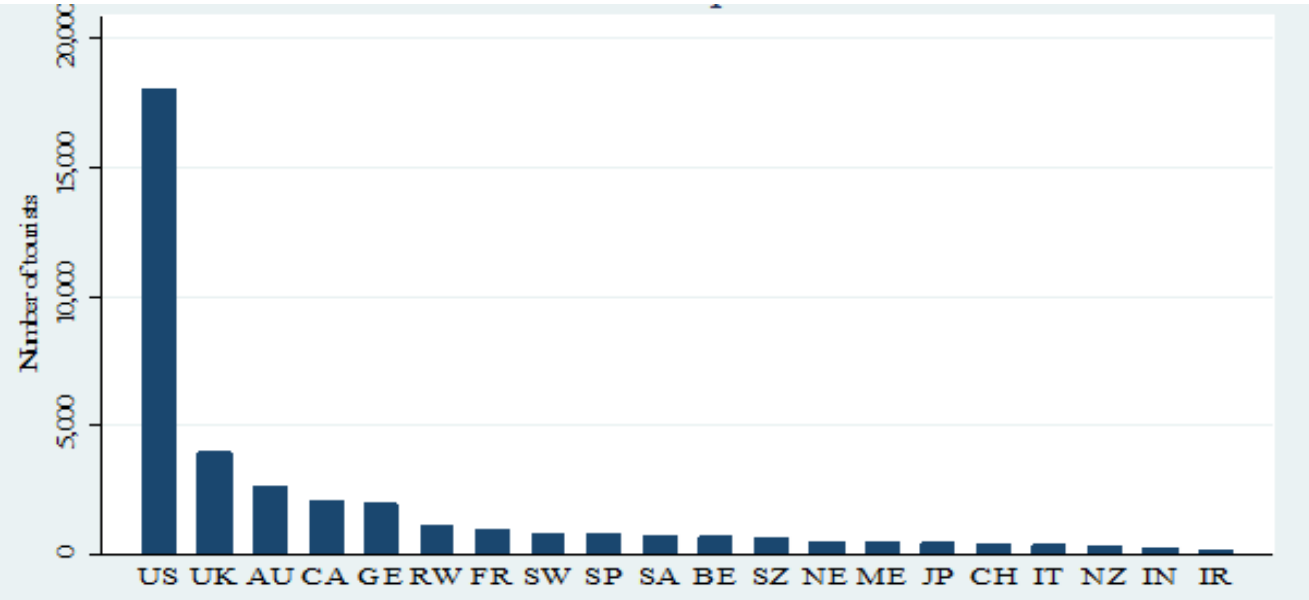
The mountains with beautiful and good-looking touristic sceneries are the main pull factor for foreign tourists to Rwanda. The approach of community tourism involving sharing of revenues with communities where tourist attractions are located has improved the standards of the people in the respective communities. The country has remained escalating considerably at regular rate of eight percent for the past two decades. However, due to the high population and limited land, Rwanda has encountered challenges in the tourism industry. There are also issues of limited physical and human capital, which limit investments, and proper management of the tourism sector. (Kambogo and Bizimana, 2016).

To embark on this study box Jenkins models (SARIMA) was though as the one to solve the highlighted issues and propose the best option to develop tourism revenues prediction in Rwanda based on the gaps that will be highlighted in the literatures bellow, some literatures tried to work on this but they have not completely shown how to get real tourism revenues and be able to predict for future, but with this research we will show the real revenues in the future based on the series of the previous years of our study where the trend will exhibit the same trajectories.

Rwanda has been reported to have more prospective sites to offer for the tourists and some were not thought of before. However, after the infamous 1994 Rwandan Tutsi genocide especially after 2000, Rwanda has heavily invested in the sector to maximize its resources form all tourist destinations in different locations of the country. The government prioritized the tourism industry, which then became the top income contributor the country (English et al., 2016). In 2017, tourism revenues raised to 438 million US Dollars from 390 million US Dollars.

Figure 1 presents the top countries of origin for the tourists who visited Nyungwe and Volcanoes National Parks in 2016. The figure reveals that US was the top source of tourists who visited the two parks, followed by UK, Austria, Canada and Germany, among others.

**FIGURE 1** : TOURISM ARRIVALS FROM TOP 10 COUNTRIES VISITING RWANDA IN 2016



## 1.2 Problem statement

Despite the clear importance of the tourism sector in Rwanda's development, there is scanty empirical work on tourism revenue forecasting based on the trends and patterns of tourist revenues. The limited empirical evidence is partly attributable to the lack of knowledge and skills in tourism data assessment, which leads to incompetency in classifying tourism fluctuations. Without proper classification and analysis, it is difficult to predict future revenues and arrivals .Currently, tourism statistics from national parks like Volcanoes and Nyungwe are captured using templates that are oriented for money generation with low usability for research purposes. Moreover, the lack of proper statistics limits the accurate prediction of future trends (Akal, 2004). Consequently, it is hard for the policy makers to make informed decisions for the future growth of the sector. The essence of this research is to solve a problem of how tourism data are not analyzed, leading to the failure to discover their trends and patterns. The lack of empirical evidence on the topic limits the flow of knowledge and skills needed for the better management of the tourism sector and specific tourist sites like Volcanoes and Nyungwe National Parks. Additionally, the improper record keeping and classification of tourist records not only limits the ability to explore patterns and trends but also hinders the accurate prediction of future revenue and arrivals.

Although the sector has realized some achievements after 1994 Rwanda Tutsi genocide, the sector has not yet realized its full potential. This is because of several challenges including but not limited to lack of properly planned investment in the sector, lack of infrastructure resources, which make some tourist attractions inaccessible, insecurity issues including poaching of critical animal species, lack of skilled guides and lack of marketing skills to promote tourism. Rwanda has been known to have richly attractive and beautiful sceneries including mountain gorillas, volcanic mountains, volcanoes and National Parks as well as historical places with features that reveal the kingdom weapons, their ways of fighting and bravely characters. In other words, Volcano and Nyungwe National Park have the majority of touristic structures that attract most of the tourists, which are unique to the world (Jauhari, 2018)

In measuring the performance of different models, preceding studies used diverse evaluation measures. The utmost commonly realistic measures machine learning model decision tree and random forest. In practice, when relating dissimilar models, it is hardly the case that one model dominate the other with respect to all assessment measures. The common way to solve the problem is to convey out the average numbers of some statistical measures and then compare the forecast models based on the parameter obtained. Some studies report that GARCH/ GARCH type models is the best model for forecasting time series.

Yet, due to few variables of the data set used, it was shown by metrics that Box Jenkins Models(SARIMA) are the best fit for this time series data of tourism revenues. Since SARIMA model of Box Jenkins will give the exact accuracy of the tourism revenues that will help decision makers and tourism industry to base on in correcting some issues in the industry itself.

### 1.3 Research questions

In order to respond to the above problem, this study attempts to answer two main research questions.

- What are the patterns of tourist revenues and arrivals for Nyungwe and Volcanoes national parks?
- How are tourist revenues and arrivals  expected to change between January 2020 and June 2021?

### 1.4 Research objectives

#### 1.4.1 General objectives

The general objective of this research is to analyze trends and patterns of tourism revenue in Volcano and Nyungwe National Park for the years 2013-2019 and make future predictions for the next 18 months.

#### 1.4.2. Specific Objectives

- To analyze patterns of tourism revenues and arrivals between 2013 and 2019.
- To predict future tourism revenues between January 2020 and June 2021.

# CHAPTER TWO: LITERATURE REVIEW

## 2.0 Literature review

## 2.1 Introduction

This section presents assessment of the theoretical literature and tourism revenues modelling and forecasting studies depend on secondary data. It reports the information found in the literature related to the study topic (Akal, 2004). This review will describe, summarize, evaluate and clarify the literature related to the study. It gives a theoretical and empirical base for the research and helps to find out the nature of the research topic. The section will also include definitions and clarification about tourism revenues and tourist arrivals including their patterns, trends and forecasts. Additionally, the section includes different authors' opinions about this topic.Eventhough the descriptive variables involved in the tourism revenues, models differ massively with study aims. The engagement of certain pointers the aspect of tourism revenues variables in modelling and forecasting tourism revenues that has been insignificant provocative (Witt and Song, 2000).

### 2.1.1 Economic dimensions of tourism

Nowadays tourism is the most prominent profitable activities on the global. Tourism revenues differ from country to country depending on the resources used and it is becoming a competitive feature for diverse countries. Many tourism researchers have presented different definitions. However, the most commonly used definitions are those of (McIntosh, 1990) defines tourism as *"A joint understanding of the interaction between the host government as tourism, business, and tourists"*by shite (2000); tourism is referred as the all ventures done in tempory motivations other than making a living but in search of refreshment, expedition and enjoyment (McIntosh, 1990).

It is very crucial to consider that tourism is an economic constituent and has played a substantial part in the world's development. Most countries have utilized tourism to achieve other policy objectives in an ultimate aim to enhance economic growth and development. International travels have shown significant increase after the World War II (World Tourism Organization, 1997).

This has improved revenues of developing countries. (Tosun, C., and Timothy, 2001). Foreign exchange revenues are greatly needed by third world countries for them to develop their economies.

The tourism sector or industry is made up of different activities, including entertainment, transportation activities with target of shopping, adventure and enjoying the topography of the land (Hillman, 2007).

### 2.1.2 The social-cultural dimensions of tourism and globalization

Many researchers discovered that tourism industry is social event. In other words, tourism activities mainly occur between two diverse groups of people. These include local visitors and foreign visitors. The tourism industry has played a big role in the improvement of relationships between societies with different cultures. This is one of the activities that allow collaboration among diverse customs, cultures and behaviors. Tourism has changed people's social structures, patterns, conducts and lifestyles. Primarily, tourism was seen as commercialized hospitality. Thus specifies that tourism was originally commercialized and ultimately developed and fixed traditional visitors or tourists and local host's relations. Tourism is also a leisure activity and involves a variety of pilgrimage travels. Furthermore, tourism is an ethic relation between tourists and the local residents or population in the host areas or countries. This is viewed in terms of invention of ethic arts for tourism market, and commercializing of ethnic characteristics for tourists.

Apart from tourism definitions from diverse insights, there are signals that tourism has sociological impacts. The real example is the relationship and partnership between tourists and local residents throughout their time of tour. The relationship is seen whenever the tourists are greeting, buying goods and services from their counterparts or when they are discussing or asking about the site to visit. This allows them to exchange their language, norms and as well as their cultural values. However, due to cultural varieties, tourist and local residents sometimes face challenges when they meet for the first time. Tourism by itself is a cultural event, since the majority of tourists are intentionally sightseeing to study others cultures. Thus, tourism is also considered as a cultural leisure activity. Tourism has critically affected culture in different ways since the growth of tourism in the world where by whenever tourists meet local residents they encounter culture variety including different religions, customs, traditions, dressing styles and behaviors or manners that differ from their own. These challenges lead to problems in communication between local citizens and foreign tourists (Mark et al., 2012).

National parks in Rwanda are frequently visited and became an international tourism activity where by many foreigners come each year to sightsee different products in the parks.

The relation between tourism and globalization has been argued after World War II (Weiss, T, 2004), since then, the tourist endures to move from one country to another. This prompted the arguments among scholars regarding the contribution of tourism in both tourist sender and receiver countries (Sugiyarto et al., 2003).

### 2.1.3 Tourism in the Third World

The tourism industry in the Third World countries is the most essential foundations of foreign exchange (Polland, 1976). However, statistics show clearly that the domestic tourism in many developing countries is still at a very low scale. In Rwanda, (Pankaj et al, 2017) found that the tendency of using Nyungwe National Park is generally higher among foreigners versus Rwandan nationals. International tourists are realized as the major source of the revenues in the developing countries (Ahmad, 1986). Most of the countries in West Africa have many tourist's attractions, which are quite pleasant for the international tourists. These include beaches, mass tourists which leads to the tourists coming looking for fun and sun. Scholars and tourists are mostly interested in studying the nature of the land and about people's culture. However, one critical issue with tourism in the developing countries is the absence of appropriate planning. (Robinson, 1972) illustrated that the developing countries require most studies on the tourism industry to change their mindset that tourism industry is for developed countries. There is slight investment and attention needed in tourism sector because the returns in terms of tourism revenues can be substantial (Mowforth and Munt, 2015).

### 2.1.4 Rural Tourism

In reality, tourism industry is a new philosophy. Tourism itself has revealed good improvement to the lives of the people as years go by. Particularly, rural tourism is progressing at a good rate compared to past years during the early 50's and 60's. The rise of rural tourism increased at substantial rates around the 70's, 80's and 90's. Tourism covers diverse activities including ecotourism, walking, climbing adventure, sports, healthy tourism, educational tourism, arts and heritage tourism, among others (Pandey, 1996; Cole and Sinclair, 2002).

### 2.1.5 Tourism typologies

Generally, many people have diverse desires and ambitions in life. Tourism has enormous economic significance with diverse types or categories that are valuable due to one's preferences. The classification of tourism services and activities is as follows:

- Pleasure or inclination tourism
- Exploration or adventure tourism
- Health tourism
- Spiritual or religious tourism
- Professional or business tourism
- Friends and family sightseeing tourism
- Other tourism.

Currently, tourism is one of the leisure activities people enjoy every year and spend a lot of money in doing research, sightseeing, adventures and meeting different people with diverse cultures. This allows tourists to acquire new styles of life including discovery of new foodstuffs, wearing styles and different cultural values. Globally, the desire for holiday tourism has rapidly increased and with time, the tourism needs have changed due to changes in lifestyles and general wellbeing of the people. It was recognized that most people aspire to visit old or traditional sceneries but nowadays people have a habit of visiting new places more than any other places (Isaac, 2008).

### 2.1.6 Tourism movements

Tourism traffic is an essential part of the tourism as it allows the transportation means for the tourists to reach their desired sites or destinations. Therefore, thinking of improving transportation to those areas is a very important factor to consider whenever investing in the tourism sector. Having all the transportations modes available and letting the tourists chose the best option for them is quite essential (Bunten and Graburn, 2018).

### 2.1.7 Seasons and climate changes

The tourism revenues are affected mostly by the seasons and climate atmospheres in the nations of origin and destination. When the climate and seasons are not favorable, it is hard for tourists to reach some remote sites due to the fact that some transportation means are affected by weather conditions. Again, time differences and cultural celebrations may affect tourist plans. For example, China in the period of their traditional festivals, spring festivals like in October and May, the number of tourists increases and afterwards, there is a decline (Ariano et al., 2010).

### 2.1.8 Analyzing tourism prompting factors

The tourism industry is the greatest economic activity that generate income to the nation through the expenditures made by the visitors. Analysis and forecasting of the tourist expenditures or revenues is useful for the growth of tourism and the progress of the nation at large. The main factors that affect tourism sector performance include tourism movement conditions, seasons and climate changes, tourism competitor status, among others (Zhu and Liao, 2014).

## 2.2 Modeling Tourism Revenues Patterns

Tourism revenues modelling and forecasting studies depend on secondary data. Even though the descriptive variables involved in the tourism revenues, models diverge greatly as far as objective of a certain research is concerned. (Witt and Song, 2000).

### 2.2.1 Predicting tourism demands

Due to advancement in computer technologies in the past years, there are many options to easily model tourism demand. The most commonly applied method in many literatures on time series are linear regression, exponential smoothing and autoregressive models (Giannopoulos et al., 2012).

However, artificial intelligence is also practical to these kinds of studies (Canestrelli and Costa, 1991). Recent research has shown that there is no single accurate model that is overall in as far as prediction is concerned (Li, Song and Witt, 2005). Actually, (Shen, Li and Song 2011) depicted that mixing dissimilar forecasting procedures can assist to increase the precision in tourism revenues. However, artificial intelligence methods are honestly novel, and they equally require more training data and energy to get precise forecasts (Rodríguez, 2017).

### 2.2.2 Sarima, Arfima, Marima Models

There are many studies on tourist revenues. (Goh and Law, 2011) and (Li, Song and Witt, 2005) have gathered a comprehensive evaluation of these researchers. Two main methods in this study concerning modeling and forecasting tourism revenues: the fundamental scientific method and the time series methods. The fundamental systematic approaches are positioned on the fundamental association concerning the request features and the tourists (Song and Li, 2008). The petition features are profits of different nations, reasonable amounts, the exchange taxes, transportation expenses and lodging charges.(Witt, 1995) and (Kalendran and King, 1997) presented models of univariate time series most of the time overtake the pivotal scientific models. Some researchers have been conducted on univariate time series models, like (Chu, 2008), (Lee, Song and Mjede, 2008), (Coshall, 2008) and Kulendran and (Witt, 2001), (Chu, 2008b) has establish that the Autoregressive Fractionally Integrated Moving Average model (ARFIMA) demonstrates the maximum predicting precision both tempory and in the permanent. However, the best performing model in the medium-run is shown to be the Seasonal Autoregressive Integrated Moving Average (SARIMA) is (Goh, 2011).

Preparation depend greatly on precise forecasts to decrease the errors in the decision making practice. That is to diminish the probabilities a decision will be unsuccessful to attain the foreseen aims. The essential of perfect prediction is consequently significant to regime and executives. For example, administration bodies use demand estimations to set marketing objectives, and discover prospective markets. Executives use mandate estimates to choose effective necessities for instance staffing, capacity and study mission opportunity including the financial possibility to build an innovative guesthouse. There occurs substantial study emphasizing tourism petition envisaging, but mostly emphases on obsolete time series techniques.

For example, "*use of Box-Jenkins' Autoregressive Integrated Moving Average (ARIMA) model to predicts tourist arrivals to Australia from Hong Kong, Malaysia and Singapore"*(Goh et al., 2011) shows usage of SARIMA and MARIMA time series prototypes with interferences in predicting tourism petitions using Hong Kong ten arrival series. Nevertheless, with our mindfulness, there is little energy engaged to examine the presentation of machine learning models in predicting tourism incomes apart from for neural network models. Neural networks remained primary practical in prediction of tourism returns by Law et al" *to predict Japanese demand for travel to Honk Kong*". However, most outcomes confirm that the basic structural method (BSM) remains a highly accurate method. Tentative outcomes indicates that neural network model predictions exceeds various regression, moving average, and exponent smoothing. Law spreads the suitability of neural networks in tourism returns predictions by using the multilayer perceptron of tourism revenues (Peng et al., 2014).

### 2.2.3 Gravity model

The Gravity model in comparison with other associated profession models is pretty prosperous in economics and sciences. This model is resulting from Newton's Law where the force upsurges by mass and drops by means of distance, which is referred as gravitation. This is the prominent Newton's Law of Universal Gravitation.

$$\boldsymbol{F_{ij} = \frac{GM_iM_j}{D^2{}_{ij}}}$$

Where:

F = Force of attractive

M = Mass

D = Distance

G = Gravitational constant

In the domain of tourism Gravity models, theoretical base of the model as mentioned by (Bergstrand, 1985) can help to delineate tourism revenues in an overseas nation. Travelling for sightseeing dedications has a significant sense behind it, in which that places are different and are extraordinary; therefore, purposes for visitors are different due to diverse missions (Durbarry, 2001).

The tourists are enforced to pay tax charges, transport expenses and money exchanges for them to access tourism facilities. The general gravity model for tourism is stated as follows:

$Y_{zjt} = \alpha_z + Y_j + \lambda_t + \beta_1 X1_{zjt} + \beta_2 X2_{zt} + \beta_3 X3_{zt} + \cdots .. + U_{zjt}$ Or

$Y_t = x + \beta_1 y_{(t-1)} + \beta_2 y_{(t-2)} + \cdots + \beta_p y_{(t-p)} + \varepsilon_t$

Where

$\boldsymbol{Y_{zjt}}$ is the capacity of tourism flow from country **z** to country **j** at time **t** as dependent variable $\boldsymbol{X1_{zjt}}$ are descriptive variables with differences in all three dimensions z, j and t, for example exchange rate, $\boldsymbol{X2_{zt}}$ are explanatory variables with inconsistency in measurement z and t. For instance, Gross Domestic Product (GDP). $\boldsymbol{X3_{zt}}$ is descriptive variable with discrepancy in dimensions j and t, α k is the origin country effect. $\boldsymbol{Y_j}$ Is the destination of country's effect, $\boldsymbol{\lambda_t}$ is the time effect and $\boldsymbol{u_{zjt}}$ is a white noise disruption term. The precise effects (**α, γ** and **λ)** can be preserved as random variables (Error Component Approach) or fixed parameters (fixed effect approach) when approximating these kind of models.

The majority of the future forecasts are established by the power of forecast. However, user skills are useful in accuracy of estimation using the original data. The expansion in technology has greatly helped in data collection and capturing the most complex methodologies used in different literatures. There are external factors that can steer to minimization of predictive power. There are the unidentified activities that might present irresistible errors to future forecasts. Although we expect some historical incidences to persist into the upcoming (Hyndman et al., 1998), the impending can convey unpredicted alterations, as it is not persistent over the period. This assists the plan of temporary forecast relatively to a permanent one. Majority of the researchers conclude that tourism revenues studies depend on secondary data, for the model to be created and evaluated for better observations. However, the descriptive variables intricate in tourism revenue models are diverse based on the research objectives and researchers' experiences (Witt and Song, 2000). Many models have been used but none is leading over others in as far as prediction is concerned (Veal, 2017).

Numerous researches on demonstrating univariate data using time series models have been applied in diverse areas, such as in businesses, tourism and transportation (Pelinescu et al., 2010); Koirala, 2012; Brojba, 2010). These variabilities of models have established a suitable gears in modeling univariate data, cheers to their accuracy for forecasting both in-sample and out-of-sample values.

The performance of ARIMA models on national incomes has improved in numerous years, with several writers decollating the use of time series models they have used. Pelinescu et al, (2010) inspected the Romanian local budget with the intention of helping policy makers to make a well-organized policy and achieve their income and spending by applying better strategic management tool. This came in due to the discovery by leaders that it is problematic to forecast future tourism revenues (Makananisa, 2015).

### 2.2.4 Garch Model

Modelling and forecasting of India's spices export data set, which shows seasonality characteristics, was done by the Box-Jenkins Autoregressive integrated moving average (ARIMA) method. However, in most cases generalized autoregressive conditional heteroscedastic (GARCH) nonlinear time-series model along with its approximation processes remain systematically considered. Lagrange multiplier test for occurrence of Autoregressive conditional heteroscedastic (ARCH) properties is also debated. The GARCH model is active for modelling and predicting different records. Relative research of the fitted ARIMA and GARCH models is performed from the perspective of dynamics in advance of forecast error variance alongside with Mean square prediction error (MSPE), Mean absolute prediction error (MAPE) and Relative mean absolute prediction error (RMAPE). The EViews, Ver. 10 software packages together with computer programs in R or python are applied for records exploration. Advantage of GARCH model above ARIMA method is attested by the data under consideration. Potential use of more precise forecasts attained by GARCH methodology vis-à-vis ARIMA methodology to ensure good accuracy in the forecast and having more real data for forecast is essential i.e having like 20 years data of tourism may help you to forecast in more than 5 years to come and this give an accurate forecast if other factors hold constant. (Paul and Himadri, 2009).

# CHAPTER THREE: METHODOLOGY

## 3.1 Data sources

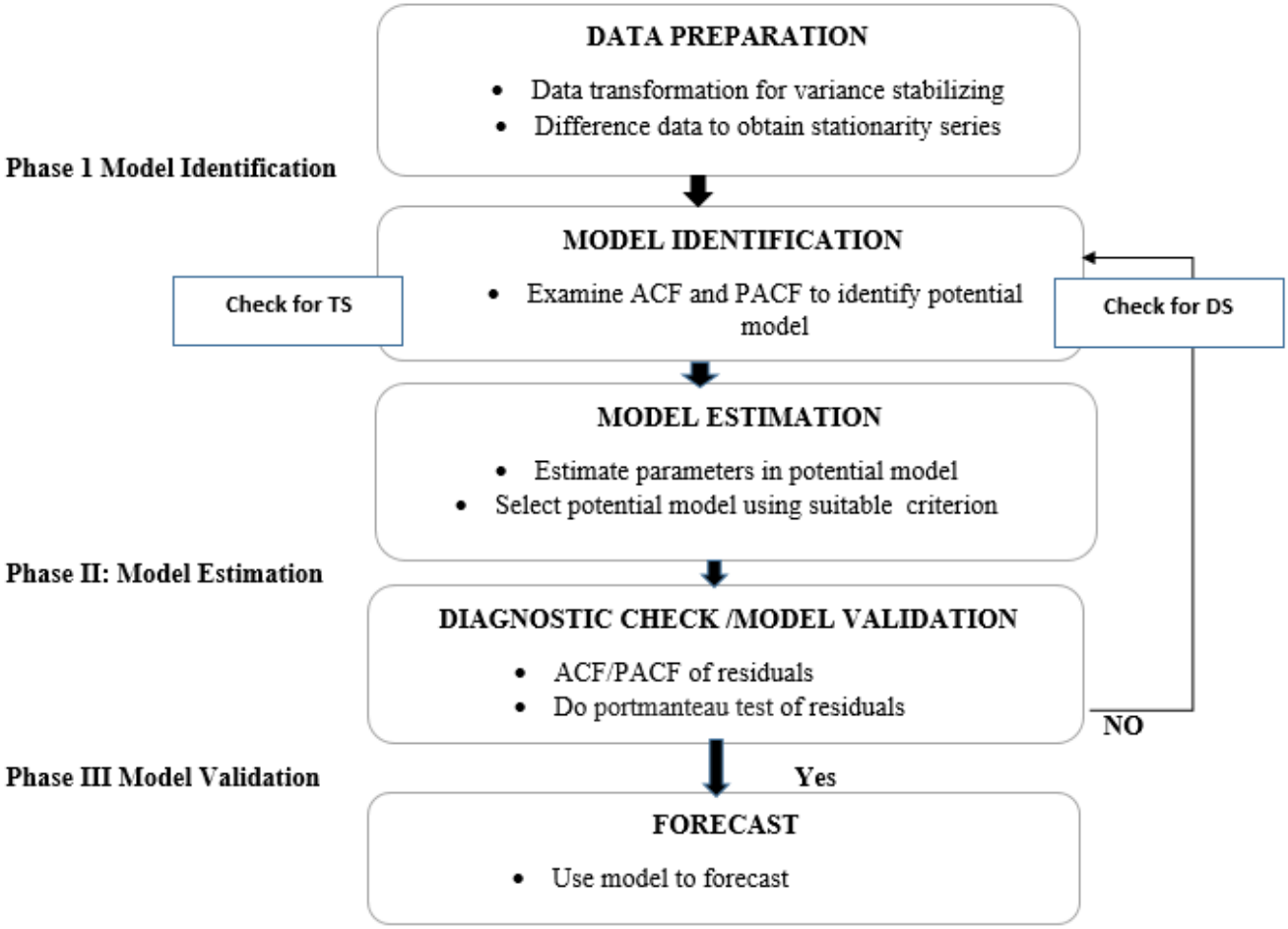
Tourism data was obtained from the Rwanda Development Board (RDB) and Rwanda online, based on records from Volcano and Nyungwe National Parks for the years between 2013 and 2019. The raw data contains daily records of tourist transaction in the two parks. This data was aggregated to monthly frequency by summing up the daily revenues in order to construct monthly time series of total revenue. Therefore, the main variables of interest in the data are the monthly number of tourist arrivals and revenues generated from the arrivals. The dataset contains 81 observations (12 months of 7 years minus three months of missing data at the beginning of 2013).

## 3.2 Analytical method: the Box-Jenkins technique

### 3.2.1 An overview on univariate Analysis

This section introduces the univariate analysis that will be done on the tourism data used in this study.one of the method chosen for this research is Box - Jenkins Analysis which refers to a formal technique of pinpoint, suitable , examining, and using seasonal integrated autoregressive, moving average (SARIMA) time series models. The method is suitable for time series of average length of minimum 50 observations.
This analysis will be done on the tourism revenues the series will be the month, years, revenues as well as arrivals since the more the arrivals the more tourism revenues generated and this will aid us to build the model that shows the seasonality and trends in the revenues in the years presented in the study up to the current situation, and this will help us to know when there is high and low tourism revenues in each year.

With univariate analysis, we will first check for the stationarity test form the raw series this will give us an idea of the presence of trends and seasonality and this will help us to know the realistic mode to be used. We will again check for the behavior of Autocorrelation function (ACF) and Partial autocorrelation function (PACF) and later analyses for seasonality of the stationary series. Lastly, this will lead us to the forecast stage of the process. During analysis of the tourism revenue series, we will be explaining more terms and concepts in later stages of the study. As it is illustrated from the bellow diagram, we are explaining the main phases of Box-Jenkins method.




### 3.2.2 Model identification

**a) Testing for the stationarity**

The first step in the identification of an appropriate model is the test for stationarity of the revenue series using the Augmented Dickey-Fuller (ADF) Test. The null hypothesis of the test is that revenue has a unit root, that is, it is non-stationary. The alternative hypothesis is that revenue has no unit root, which implies that it is a stationary series. The decision rule in the ADF test is that: if the ADF test statistic is more negative (bigger in absolute terms) than the critical values at the chosen level of significance, the null hypothesis is rejected, otherwise the null hypothesis is not rejected. An alternative statement is that: the null hypothesis is rejected whenever the P-value is smaller than the chosen level of significance. In this study, the five percent level is chosen to evaluate this decision rule.

If the null hypothesis is rejected an ARMA (p, q) model would be estimated based on the raw values of revenue, corresponding to an ARIMA (p, 0, q) model where d=0 represents integration order zero for a stationary series. On the other hand, rejection of the null hypothesis would mean that the series is non-stationary, leading to the estimation of a first-differenced model. This hence implies that the model is estimated as an autoregressive integrated moving average (ARIMA (p, d, q)) model, where the integration order d depends on how many times the differencing is made to make the series stationary.

In this chapter, we will test for the stationarity of the revenues series using the Augmented Dickey-Fuller (ADF) Test. The null hypothesis of the test is that revenue has a unit root, that is, it is non-stationary. The alternative hypothesis is that revenue has no unit root, which implies that it is a stationary series. As we are testing for the stationarity, we introduce "Unit Root Test as (Alleyne, D, 2006). " Unit root tests are tests used for stationarity in a time series. A time series has stationarity if a change in time does not cause an alteration in the form of the distribution; unit roots are one cause for non-stationarity. With root test, we are able to realize the occurrence of a non-stationarity. In addition, we will be capable to clearly understand the non-stationarity and be able to make a time series stationary.by use of the two important processes used Trend Stationary (TS) and Difference Stationary (DS) processes. We are going to differentiate these two important processes below.

### *I.* The Trend Stationary process (TS)

Trend stationary process: Demonstrates deterministic non-stationarity and is expressed in this form a trend Non-stationary process $\boldsymbol{R_t}$, as $\boldsymbol{R_t} = \boldsymbol{\theta_t} + \boldsymbol{\varepsilon_t}$

Where $\boldsymbol{\theta_t}$ is a deterministic mean trend and $\boldsymbol{\varepsilon_t}$ is a stationary stochastic process with mean zero.

For most applications, trend is the main concern. Time series decomposition methods emphasizes mostly on decaying $\boldsymbol{R_t}$ into different trend sources (e.g., secular trend component and seasonal component). You can decay series nonparametrically by means of filters (moving averages), or parametrically using regression methods. Assume an estimate $\boldsymbol{\theta_t}$, you can realize the residual series $\boldsymbol{R_t}$ - $\boldsymbol{\theta_t}$ for autocorrelation, and optionally model by means of a stationary stochastic process model.

The accurate technique of stationing this type of process is the ordinary least squares method.

***II.*** **The Difference Stationary process (DS)**

DS processes are processes that are made stationary by using filters. non-stationary time series continue to be differenced until stationarity is attained as Box-Jenkins modeling method approach recommends.by Clements, M. P., and Hendry, D. F. (2001).

. You can express a difference-stationary process, $\boldsymbol{R_t}$ **,** as $\Delta_{\boldsymbol{Rt}}^{\boldsymbol{P}} = \boldsymbol{\theta} + \boldsymbol{\delta}(\boldsymbol{L})\boldsymbol{\varepsilon_t}$

$\boldsymbol{\Delta^P} = (\mathbf{1} - \boldsymbol{L})^{\boldsymbol{P}}$ is $P^{th}$ degree differencing operator

$\boldsymbol{\delta}(\boldsymbol{L}) = (\mathbf{1} + \boldsymbol{\delta_1 L} + \boldsymbol{\delta_2 L^2} + \ldots.)$

As stated above DS is a process that brings the stationarity by use of filters on differences this implies that:

$(\mathbf{1} - \boldsymbol{F})^{\boldsymbol{d}}\, \boldsymbol{G_1} = \boldsymbol{\delta} + \boldsymbol{\varepsilon_t}$

Where $\boldsymbol{\varepsilon_t}$ is stationary process, $\boldsymbol{\delta}$ is a constant and **F** is the change operator and the order of the difference filter

The introduction of the constant $\boldsymbol{\delta}$ brings an attention to define two different processes:

$\boldsymbol{\delta} = \mathbf{0}$ **:** In DS is written without difference $\boldsymbol{G_1} = \boldsymbol{G_{t-1}} + \boldsymbol{\varepsilon_t}$.

$\boldsymbol{\varepsilon_t}$ is white noise

Amongst numerous kinds of stationarity test that formerly known, we chose to use Augmented Dickey–Fuller test (ADF).The hint to choose this test was because it will assist in identifying the presence of Unit root in the time series as well stationarity nature in the series as well as existence of correlation structure in the error term for which, according to Bourbonnais (2007), there is no motive to assume, a priori, their non-correlation and therefore their whiteness (white noise).

Time series are made stationary by differencing are known to be integrated processes. Precisely, when P differences are essential to create a series stationary, this kind of series are termed as integrated of order P, denoted I(P). Processes with P ≥ 1 are often said to have a unit root.by (Hamilton and Susmel, 1994).

According to the (Bourbonnais 2007, p. 231), under the alternative hypothesis **| Φ|<1**, the ADF test is based on the estimate of ordinary least squares of the following three models:

$\Delta G_t = \rho G_{t-1} - \sum_{j=2}^{p} \Phi\, \Delta G_{t-j+1} + \varepsilon_t$ **Model 1**

$\Delta G_t = \rho G_{t-1} - \sum_{j=2}^{p} \Phi\, \Delta G_{t-j+1} + c + \varepsilon_t$ **Model 2**

$\Delta G_t = \rho G_{t-1} - \sum_{j=2}^{p} \Phi\, \Delta G_{t-j+1} + c + bt + \varepsilon_t$ **Model 3**

**Where** $\varepsilon_t$ ~ independent and identically distributed (i.i.d.)

If the hypothesis H0 is tested, the chronicle $\boldsymbol{R_t}$ is not stationary irrespective of the model

retained.In the third model, if we accept H1: | **Φ**|<1 and if the coefficient b is significantly different from 0 then the process is a TS process that brings stationarity by calculating residuals versus trend estimated by ordinary least squares.

**b) Auto-correlation function and partial auto-correlation function**

The autocorrelation function (ACF) and partial autocorrelation function (PACF) are examined to choose between autoregressive (AR) and moving average (MA) components of the ARIMA model. As a rule of thumb governing the Box-Jenkins method, the AR model is selected if the ACF decays or persists and the PACF has insignificant spikes after lag order q. On the other hand, if the PACF persists and the ACF has insignificant spikes after autoregressive lag p, the preferred model is MA. The third possible scenario is that both ACF and PACF persist or have significant spikes, in which case a combination of AR and MA models is selected, leading to the estimation of an ARIMA (p, d, q) model with *p* autoregressive terms, *d* order of integration and *q* moving average terms. The ACF and PACF plots for the revenue series (or its first difference, in case of non-stationarity) will be examined in order to choose between AR and MA components, or a combination of the two.

There is a rule in using these two functions: Determines if there is clear trend in the data use differencing or drifting to "detrend" the data .normally one-lag differencing is used. In Box-Jenkins methodology both ACF and PACF autocorrelation function and partial autocorrelation function respectively are used to recognize the orders p,d,q,P,D,Q for SARIMA model(Box et al., 2008)

$$\boldsymbol{r^r = c_r / c_0} \qquad (1)$$

Where $c_r$ is auto covariance function

$$\boldsymbol{c_r = T^{-1} \sum_{t=r+1}^{T} (y_t - y^-)(y_{t-r} - y^-),\ r = 1,2,3 \ldots} \qquad (2)$$

and $\boldsymbol{c_0}$ is the derived variance from

$$c_0 = T^{-1} \sum_{t=1}^{T} (y_t - y^-).^2 \qquad (3)$$

Where T is the observation $\boldsymbol{y_t}$ with mean of $\boldsymbol{y^-}$ (Harvey, 1989).

**c) Model formulation**

The analytical method used in this study is the Autoregressive Integrated Moving Average (ARIMA) model, recognized by (Box- Jenkins 1970) as a one of the methods for time series forecasting. ARIMA models are used to designate autocorrelations that are within the data.

Forecasting is based on a linear grouping of the past observations that requires a stationary series without any precise trend in the data. The upcoming value of a variable in an ARIMA model expressed as a linear combination of the past values and past errors. In other words, the current value of the dependent variable depends on its past values and a moving average of its past errors. The terms p, d and q represent the autoregressive order, order of integration and moving average order, respectively. Therefore, the ARIMA model for tourist revenue is expressed as:

$$\Delta Revenue_t = \alpha + \boldsymbol{\beta_1}\ \Delta Revenue_{t-1} + \boldsymbol{\beta_2}\Delta Revenue_{t-2} + \cdots + \boldsymbol{\beta_p}\Delta Revenue_{t-p} + \Delta\boldsymbol{\varepsilon_t} + \rho_1\Delta\boldsymbol{\varepsilon_{t-1}} + \rho_2\Delta\boldsymbol{\varepsilon_{t-2}} + \ldots + \rho_q\Delta\boldsymbol{\varepsilon_{t-q}}$$

Where $\boldsymbol{\Delta y_t}$ is the change in tourist revenue (measured in United States dollars) realized in month *t*.

$\boldsymbol{\Delta y_{t-1}}$ is the change in tourist revenue for the previous period (month). The Coefficients ($\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, **….** $\boldsymbol{\beta_p}$) represent the extent to which the past values (lags 1, 2,…., p) affect the current revenues. The parameter $\alpha$ is a constant and $\boldsymbol{\varepsilon_t}$ is the error term, which is assumed to be "white-noise".

**3.2.3 Setting up the SARIMA (p, d, q) (P, D, Q) m model**

In case of seasonality in the revenue series, the standard ARIMA (p, d, q) model would not appropriately estimate tourist revenue. To account for this potential seasonality (monthly fluctuations in the trend of revenue between 2013 and 2019), the seasonal integrated moving average model (SARIMA) is estimated. In other words, a SARIMA (p, d, q) (P, D, Q)s model is defined for the revenue series, where the parameters p, d, and q are the same as specified in the ARIMA set up above. In addition to these terms, the SARIMA model adds seasonal lags (P), a seasonal difference order (D) and seasonal moving average terms (Q). Lastly, s represents the length of the seasonal cycle, which corresponds to 12 months in the case of the monthly revenue data used in this study. Following Brockwell and Davis (1991), the SARIMA model for a series $\boldsymbol{y_t}$ can be represented as:

$$\boldsymbol{\Phi}(\boldsymbol{L^s})\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{\Delta^d}\boldsymbol{\Delta^D}\boldsymbol{y_t} = \boldsymbol{\theta_0} + \boldsymbol{\Theta}(\boldsymbol{L^s})\boldsymbol{\theta}(\boldsymbol{L})\boldsymbol{\varepsilon_t}$$

**Note:** left part seasonal and non-seasonal AR multiply each other, again on the right side the seasonal and non-seasonal MA multiply each other.

Where L is the lag operator; $\boldsymbol{L^s}$ represents seasonal lags; $\boldsymbol{\Delta}$ is the difference operator. $\boldsymbol{\Delta^d}$ and $\boldsymbol{\Delta^D}$ represent the integration order of the series and the seasonal difference order, respectively. The polynomial structures for the autoregressive and moving average components are represent by the two equations below, respectively.

Non-Seasonal AR: $\boldsymbol{\Phi(L) = 1 - \Phi_1 L - \cdots - \Phi_p L^p}$

Non-Seasonal MA: $\boldsymbol{\theta(L) = 1 + \theta_1 L + \cdots - \theta_q L^q}$

Similarly, the seasonal polynomial structures are represented as:

Seasonal AR: $\boldsymbol{\Phi(L^s) = 1 - \Phi_1 L^s - \cdots - \Phi_P L^{Ps}}$

Seasonal MA: $\boldsymbol{\Theta(L^s) = 1 + \Theta_1 L^s + \cdots - \Theta_Q L^{Qs}}$

Finally, $\varepsilon_t$ is an error term that is assumed to be a Gaussian process with a zero mean and variance $\sigma^2$.

The model to be estimated for tourist revenue is then written as:

$$\boldsymbol{\Phi(L^s)\Phi(L)\Delta^d\Delta^D revenue_t = \theta_0 + \Theta(L^s)\theta(L)\varepsilon_t}$$

Where $s = 12$. The two difference operators $\boldsymbol{\Delta^d}$ and $\boldsymbol{\Delta^D}$ are used to transform the non-stationary series $\boldsymbol{revenue_t}$ into a difference-stationary series $\boldsymbol{drevenue_t}$.

### 3.2.4 Model estimation

To begin with before we describe about the model estimation let's talk on Maximum Likelihood Estimator Whenever a model is identified to explain the data.the first issue arise is to find the expected parameter estimates.In real perception, the best way to go is to use the Maximum Likelihood Estimator (MLE) which delivers an overall method for assessing a vector of unknown parameters in a probable multivariate distribution

To explain this let assume that a certain variable x is a random variable with probability density function f(x) which consist of unknown parameters $\theta = (\theta_1, \theta_2, \ldots . \theta_p)$ this is a random sample of G observations $(x_1, x_2, \ldots . x_G)$ is available and the likelihood L, is defined as the joint density of the observations; i.e.

**L=f**$\boldsymbol{(x_1, x_2, \ldots . x_G)}$ **= f** $\boldsymbol{(x_g; \theta)}$

**Note that** since the $x_G$ are arbitrarily nominated from the data, trial values can be measured to have been independently drawn, so the likelihood, which is the joint distribution of $(x_1, x_2, \ldots . x_G)$ is simply the artifact of the marginal densities.

The Maximum Likelihood Estimator (MLE), $(\theta_1, \theta_2, \ldots . \theta_p)$ is the value of the parameters that is most possible to have produced the observed sample of data. The attractiveness of MLE is that subject to irrelevant circumstances, it has very suitable possessions in bulky samples (i.e., asymptotically).

### 3.2.5 Model Validation/ Diagnostic checks for the selected model.

In this step, the different candidate models will be validated using either least squares or maximum likelihood methods. The Akaike Information Criterion (AIC) and Schwarz information criterion (SIC) will be used to select the most appropriate model, which is the one that best fits the data. In other words, a model with the lowest AIC and SIC respectively will be considered to be the one that gives the lowest estimation error and, most likely, the most accurate forecast. This step is the estimation of the ARIMA model identified in the previous step.

The diagnostic checks assess how best the estimated model fits the data. It involves checking either the normality of the residuals or patterns exhibited in the residuals. This is conducted using tests and/or residual plots after estimation. After estimating the selected model, it will then be tested for appropriateness using three alternative, yet complementary methods, namely, (i) the ACF and PACF plots of the associated residuals, (ii) the Ljung-Box test for serial correlation in the residuals and (iii) the Jarque-Bera test for the normality of residuals. If the estimated model passes all the three diagnostic checks, then forecasting based on the model is reliable and accurate.

#### a) ACF and PACF plots of the associated residuals

One way to examine the validity of the estimated model is to examine the ACF and PACF plots of the residuals is plotted, with an aim of examining the presence of any significant correlation in the residuals at different lag orders. Based on this diagnostic check, a reliable model should explain the biggest variation in the variable of interest and little or no significant information should be left unexplained. In other words, a reliable model is not expected to have residuals with significant spikes in either the ACF or PACF.

**b) Ljung-Box test for serial correlation in the residuals**

The Ljung-Box test is a test for autocorrelation in the residuals of the estimated model. The null hypothesis is that there no significant autocorrelation at the corresponding lag order. This is tested against the alternative hypothesis of significant serial correlation. Ideally, if the estimated model is reliable, it should not show signs of significant autocorrelation in the residuals. The decision rule for the Ljung-Box test is based on the P-values presented for the respective lag orders. If the P-value is smaller than the chosen level of significance (for example 0.05 for a 95% confidence level), the null hypothesis is rejected, implying that autocorrelation is present in the model's residuals at the specified lag order. On the other hand, a P-value that is bigger than the significance level implies no significant correlation in the residuals at the respective lag order.

**c) Jarque-Bera test for the normality of residuals**

The third diagnostic method is the test that residuals have a zero mean. This is conducted using the Jarque-Bera test for residual normality. The null hypothesis is that residuals are approximately normally distributed with a zero mean and variance $\sigma^2$, while the alternative hypothesis is that the residuals do not follow a normal distribution. The decision rule of the test is that: if the P-value of the test is bigger than the chosen level of significance (for example 0.05), the null hypothesis is rejected, otherwise it is not rejected.

### 3.2.6 Analytical Tool

Data was analyzed using EViews version 10. The EViews software is a commonly used analytical software which has powerful features and tools especially for time series analysis and forecasting. In addition, the EViews software was selected for analysis in this thesis because its flexible tools and user interface yet not compromising analytical performance.

# CHAPTER FOUR: RESULTS AND DISCUSSIONS

## 4.1. Descriptive analysis of the series

Figure 2 shows that the average monthly revenue is approximately 1.3 million United States dollars (USD. The lowest monthly revenue was slightly more than 59,000 USD and the maximum was over three million. The revenue values are highly diverse as indicated by a very large standard deviation relative to the mean (a standard deviation of 670,383 USD is over 48% of the mean value of 1,374,317 USD).

FIGURE 2: SUMMARY STATISTICS OF MONTHLY TOURIST REVENUE

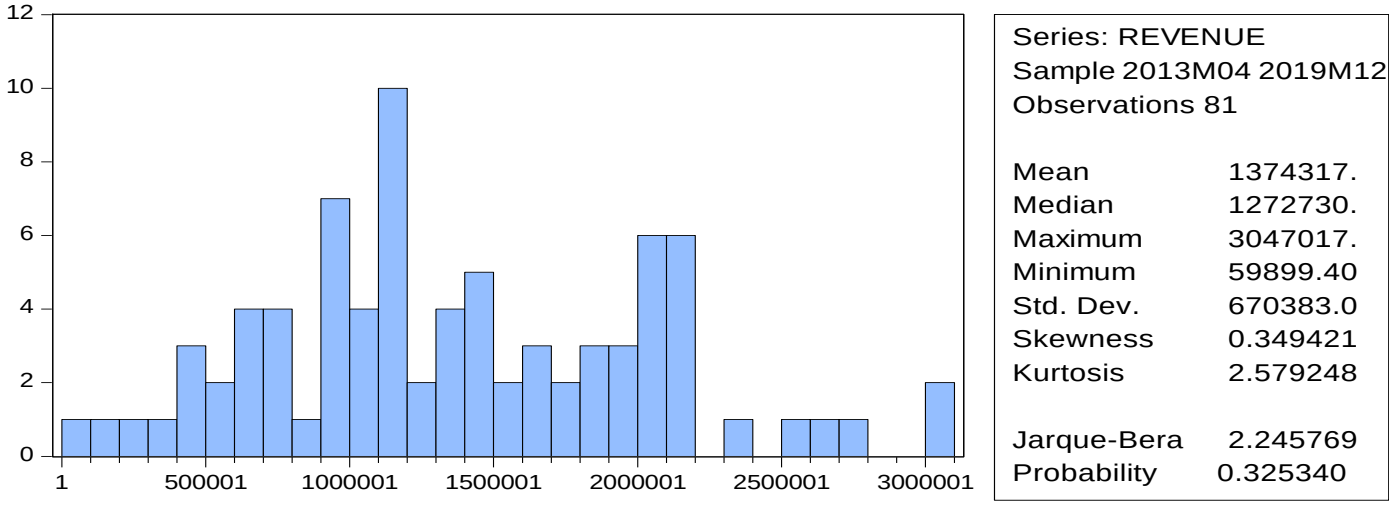
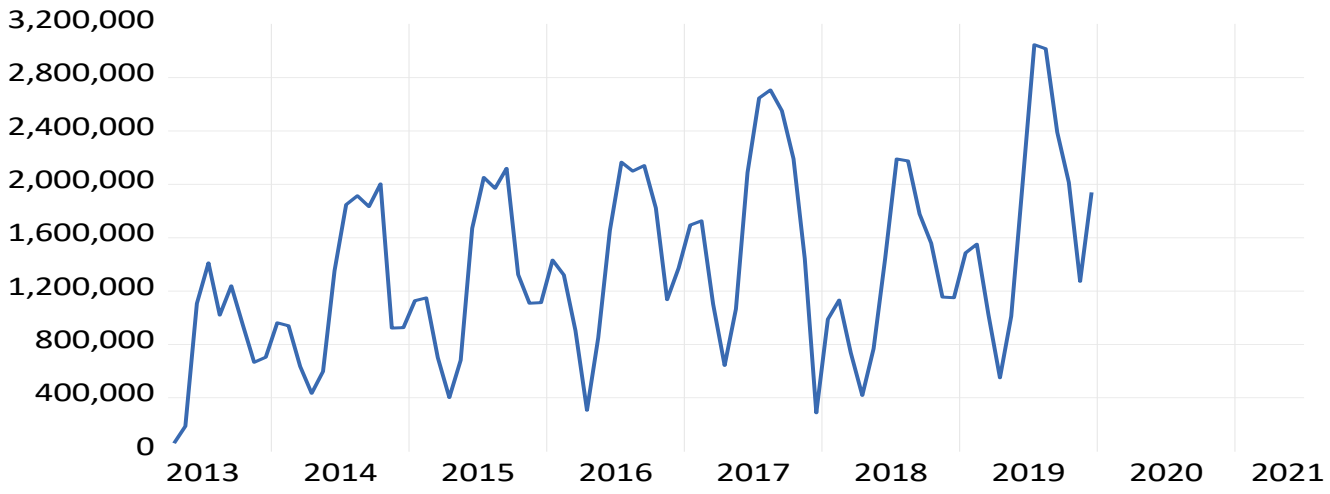



Figure 3 shows the monthly tourist revenue from the two study parks across the analysis period 2013-2019. Although the overall trend is generally positive and the revenue is much lower in 2013, it is important to note in that year, tourism statistics were not as accurately captured as they were in the latter years. In the year 2014, the low revenue could partly be attributed to the outbreak of Ebola in the Democratic Republic of Congo (DRC). Thereafter, revenue values exhibited a positive trend through 2019. There is also a seasonal dimension to the observed trend, where revenue is highest during the middle of the year in months of June-September. Therefore, the estimation results to be presented in the subsequent section will incorporate this seasonality by estimating a seasonal autoregressive moving average (SARIMA) model rather than a standard ARIMA model.

FIGURE 3:MONTHLY TOURISM REVENUES FOR VOLCANO AND NYUNGWE NATIONAL PARK: 2013-2019



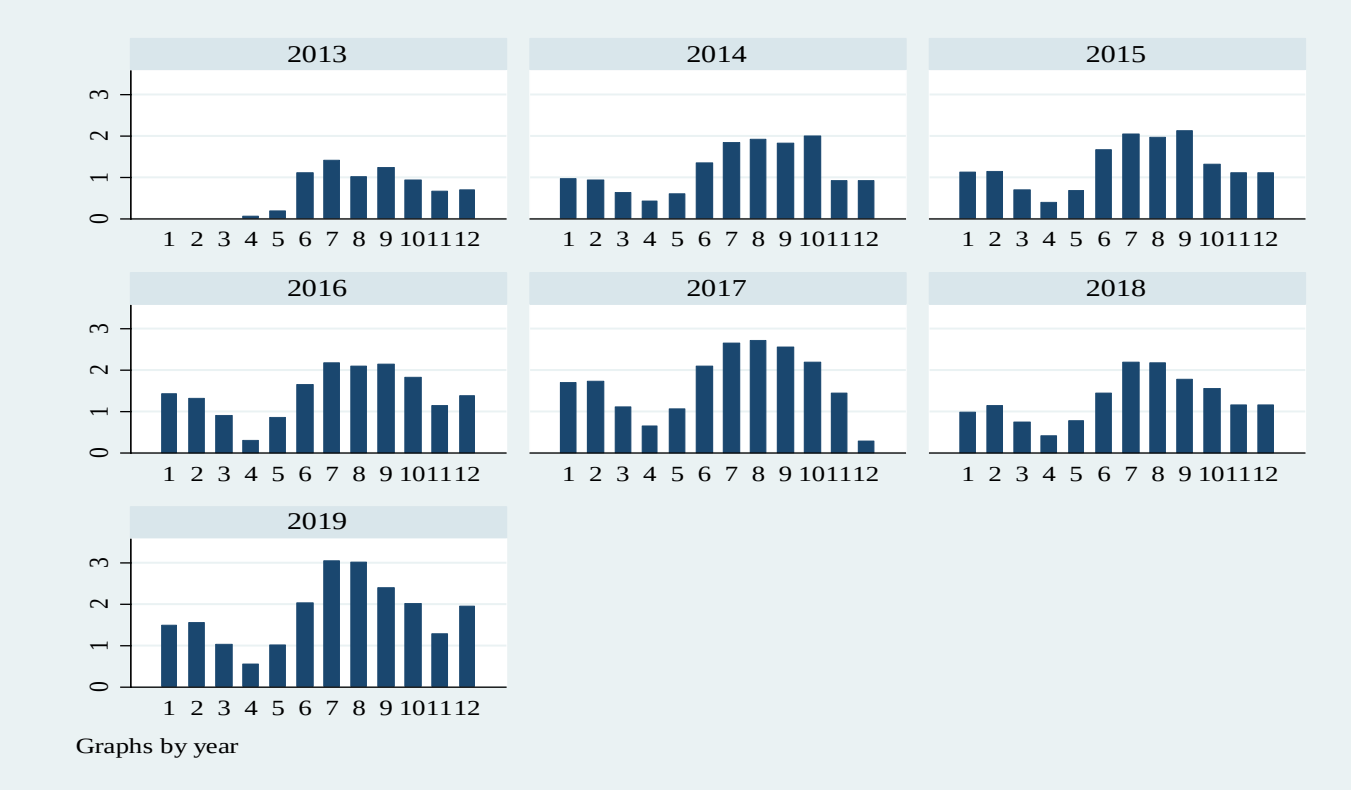

In addition to the tourist revenue, Table 1 shows the number of arrivals estimated in this thesis and compares it with the annual statistical records obtained from RDB. Although there are some differences in the estimated number of arrivals and the reports from RDB, these differences are quite small. It is important to note that during the process of data cleaning and processing, some observations with either missing or less informative information were not utilized for the study. This decision to eliminate some observations was made in consultation with the statistics team at RDB.

TABLE 1: NUMBER OF TOURIST ARRIVALS AND REVENUE ESTIMATED AND OBTAINED FROM RDB

| Year | Arrivals according to RDB Statistics | Number of confirmed arrivals | Tourism Revenues in million USD |
|---|---|---|---|
| 2013 | 32,101 | 25,896 | 7.340.114 |
| 2014 | 37,025 | 30,896 | 14.369.400 |
| 2015 | 35,928 | 32,760 | 15.420.308 |
| 2016 | 46,387 | 41,404 | 17.206.200 |
| 2017 | 49,982 | 41,888 | 20.150.700 |
| 2018 | 45,722 | 44,306 | 15.500.200 |
| 2019 | 52,751 | 57,572 | 21.332.000 |

Figure 4 presents tourist revenues disaggregated by year and month in order to clearly examine seasonality. The revenue figures exhibit seasonal cyclical behavior: they are highest between July and September, decline towards the end of the year and rise again in February. This provides evidence of seasonality in the series, which will be accounted for in the estimation procedure. In other words, the observed seasonality suggests the use of a seasonal ARIMA (i.e. SARIMA) model rather than a standard ARIMA model.

**FIGURE 4**: MONTHLY TOURIST REVENUE FOR NYUNGWE AND VOLCANOES NATIONAL PARKS: 2013-2019




### 4.2 Stationarity analysis

Based on the ADF test results in Table 2, the null hypothesis is not rejected because the P-value (0.5596) is bigger than any conventional level of significance. Therefore, the series revenue has a unit root i.e., it is non-stationary.

Therefore, revenue will be estimated in its first-differenced form rather than raw values, hence leading to the use of ARIMA rather than ARMA. The first difference of revenue is tested for stationary using the same test as above $\Delta revenue_t = revenue_t - revenue_{t-1}$ or $\Delta y_t = y_t - y_{t-1}$

**TABLE 2**: ADF TEST RESULTS FOR THE STATIONARITY OF TOURISM REVENUE

Null Hypothesis: REVENUE has a unit root
Exogenous: Constant
Lag Length: 10 (Automatic - based on SIC, maxlag=11)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -1.436530 | 0.5596 |
| Test critical values: | 1% level | -3.527045 | |
| | 5% level | -2.903566 | |
| | 10% level | -2.589227 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(REVENUE)
Method: Least Squares
Date: 09/17/20 Time: 21:15
Sample (adjusted): 2014M03 2019M12
Included observations: 70 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| REVENUE(-1) | -0.241900 | 0.168392 | -1.436530 | 0.1562 |
| D(REVENUE(-1)) | 0.087897 | 0.174226 | 0.504500 | 0.6158 |
| D(REVENUE(-2)) | -0.258738 | 0.169947 | -1.522465 | 0.1333 |
| D(REVENUE(-3)) | -0.357424 | 0.162924 | -2.193811 | 0.0323 |
| D(REVENUE(-4)) | -0.214217 | 0.155020 | -1.381865 | 0.1723 |
| D(REVENUE(-5)) | -0.294570 | 0.149948 | -1.964487 | 0.0543 |
| D(REVENUE(-6)) | -0.257427 | 0.133992 | -1.921204 | 0.0596 |
| D(REVENUE(-7)) | -0.361391 | 0.129703 | -2.786287 | 0.0072 |
| D(REVENUE(-8)) | -0.321936 | 0.121748 | -2.644268 | 0.0105 |
| D(REVENUE(-9)) | -0.365754 | 0.103472 | -3.534823 | 0.0008 |
| D(REVENUE(-10)) | -0.582561 | 0.112962 | -5.157126 | 0.0000 |
| C | 406637.3 | 233356.7 | 1.742557 | 0.0867 |

| | | | |
|---|---|---|---|
| R-squared | 0.692549 | Mean dependent var | 14279.26 |
| Adjusted R-squared | 0.634240 | S.D. dependent var | 509270.0 |
| S.E. of regression | 307997.0 | Akaike info criterion | 28.26837 |
| Sum squared resid | 5.50E+12 | Schwarz criterion | 28.65383 |
| Log likelihood | -977.3930 | Hannan-Quinn criter. | 28.42148 |
| F-statistic | 11.87710 | Durbin-Watson stat | 1.878756 |
| Prob(F-statistic) | 0.000000 | | |

In Table 3, The MacKinnon one-sided P-value of 0.0001 is smaller than any conventional level of significance. Similarly, the ADF test statistic (-9.971662) is more negative than the critical values at all levels of significance. Therefore, the null hypothesis that the first difference of revenue is non-stationary is rejected even at the one percent significance level. This therefore implies that the first difference of revenue is stationary and is the focus of the analysis that follows.

**TABLE 3**: ADF TEST RESULTS FOR THE STATIONARITY OF TOURISM REVENUE

Null Hypothesis: DREVENUE has a unit root
Exogenous: Constant
Lag Length: 9 (Automatic - based on SIC, maxlag=11)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -9.971662 | 0.0001 |
| Test critical values: | 1% level | -3.527045 | |
| | 5% level | -2.903566 | |
| | 10% level | -2.589227 | |

*MacKinnon (1996) one-sided p-values.

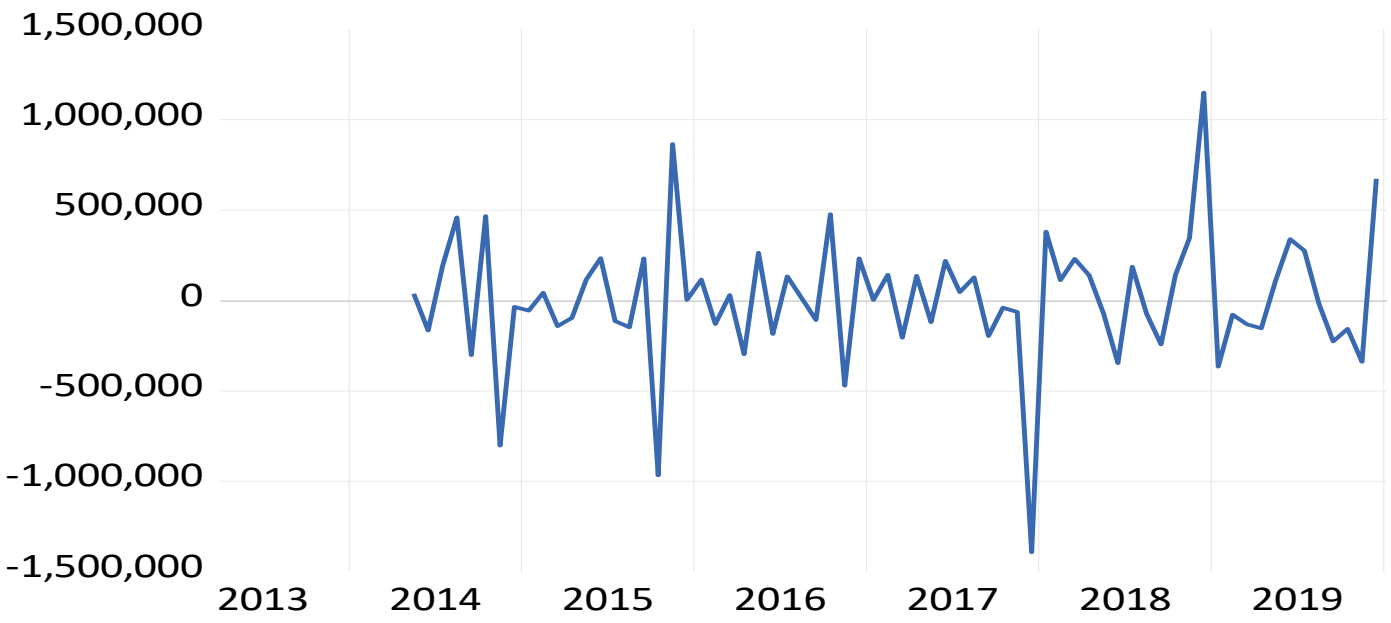
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DREVENUE)
Method: Least Squares
Date: 09/24/20 Time: 15:10
Sample (adjusted): 2014M03 2019M12
Included observations: 70 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| DREVENUE(-1) | -5.120167 | 0.513472 | -9.971662 | 0.0000 |
| D(DREVENUE(-1)) | 4.012550 | 0.445696 | 9.002887 | 0.0000 |
| D(DREVENUE(-2)) | 3.557021 | 0.425537 | 8.358892 | 0.0000 |
| D(DREVENUE(-3)) | 3.028708 | 0.381278 | 7.943563 | 0.0000 |
| D(DREVENUE(-4)) | 2.662927 | 0.334711 | 7.955903 | 0.0000 |
| D(DREVENUE(-5)) | 2.227490 | 0.285873 | 7.791884 | 0.0000 |
| D(DREVENUE(-6)) | 1.864617 | 0.239378 | 7.789436 | 0.0000 |
| D(DREVENUE(-7)) | 1.408676 | 0.187489 | 7.513366 | 0.0000 |
| D(DREVENUE(-8)) | 1.019001 | 0.135948 | 7.495532 | 0.0000 |
| D(DREVENUE(-9)) | 0.610234 | 0.112306 | 5.433659 | 0.0000 |
| C | 75769.88 | 37835.03 | 2.002638 | 0.0498 |

| | | | |
|---|---|---|---|
| R-squared | 0.757827 | Mean dependent var | 9852.141 |
| Adjusted R-squared | 0.716781 | S.D. dependent var | 583935.2 |
| S.E. of regression | 310760.8 | Akaike info criterion | 28.27476 |
| Sum squared resid | 5.70E+12 | Schwarz criterion | 28.62810 |
| Log likelihood | -978.6167 | Hannan-Quinn criter. | 28.41511 |
| F-statistic | 18.46273 | Durbin-Watson stat | 1.891081 |
| Prob(F-statistic) | 0.000000 | | |

Figure 5 plots the change in monthly tourist revenue. Clearly, the trend and seasonality in the variable has been removed after differencing.

**FIGURE 5**: LINE PLOT OF MONTHLY CHANGE IN TOURIST REVENUE: 2013-2019

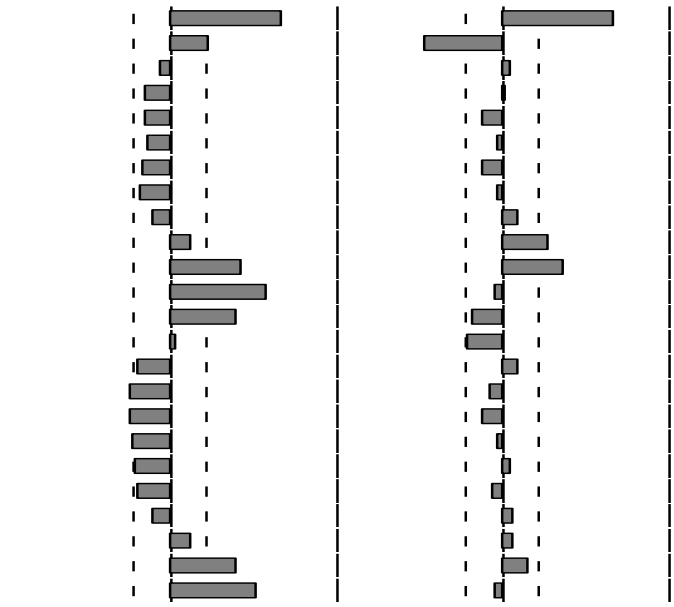



### 4.3 Analysis of ACF and PACF of revenue

Figure 6 shows the correlogram for the first-differenced revenue, which indicates two important observations. First, significant spikes of ACF and PACF indicate a mix of both AR and MA terms in the model to be estimated. Secondly, the appearance of significant spikes at fixed 12-period intervals indicates seasonality (with a 12-month frequency) of revenue. In other words, revenue changes seasonally where the season depends on the 12 months of each year covered in the data. Therefore, seasonality is taken into account in the estimation stage that follows.

**FIGURE 6**: ACF AND PACF FOR MONTHLY TOURIST REVENUE

Sample (adjusted): 2013M04 2019M12
Included observations: 81 after adjustments



| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.703 | 0.703 | 41.546 | 0.000 |
| | | 2 | 0.251 | -0.481 | 46.911 | 0.000 |
| | | 3 | -0.056 | 0.052 | 47.186 | 0.000 |
| | | 4 | -0.150 | 0.020 | 49.150 | 0.000 |
| | | 5 | -0.152 | -0.112 | 51.186 | 0.000 |
| | | 6 | -0.138 | -0.032 | 52.897 | 0.000 |
| | | 7 | -0.159 | -0.121 | 55.206 | 0.000 |
| | | 8 | -0.175 | -0.026 | 58.016 | 0.000 |
| | | 9 | -0.104 | 0.103 | 59.021 | 0.000 |
| | | 10 | 0.125 | 0.291 | 60.497 | 0.000 |
| | | 11 | 0.458 | 0.383 | 80.664 | 0.000 |
| | | 12 | 0.602 | -0.038 | 115.95 | 0.000 |
| | | 13 | 0.414 | -0.181 | 132.88 | 0.000 |
| | | 14 | 0.040 | -0.206 | 133.04 | 0.000 |
| | | 15 | -0.193 | 0.092 | 136.84 | 0.000 |
| | | 16 | -0.251 | -0.081 | 143.37 | 0.000 |
| | | 17 | -0.239 | -0.112 | 149.39 | 0.000 |
| | | 18 | -0.232 | -0.024 | 155.14 | 0.000 |
| | | 19 | -0.211 | 0.066 | 159.96 | 0.000 |
| | | 20 | -0.196 | -0.054 | 164.20 | 0.000 |
| | | 21 | -0.103 | 0.068 | 165.40 | 0.000 |
| | | 22 | 0.124 | 0.078 | 167.14 | 0.000 |
| | | 23 | 0.423 | 0.171 | 187.89 | 0.000 |
| | | 24 | 0.540 | -0.038 | 222.24 | 0.000 |

### 4.4. Model selection, estimation and validation

Table 4 below provides the AIC values for the different candidate models. The *SARIMA (1, 1, 1) (1, 1, 1) 12* model has the lowest AIC and SIC values among all potential models. Therefore, the estimation, forecasting and diagnostic checks presented in the subsequent sections will be based on this model.

**TABLE 4**: MODEL SELECTION BASED ON SCHWARZ INFORMATION CRITERION (SIC)

| Model specification | AIC | SIC | Decision |
|---|---|---|---|
| SARIMA (1, 1, 1, 1, 1, 1) 12 | 28.11118 | 28.32741 | Selected Model |
| SARIMA (2, 1, 2, 1, 1, 1) 12 | 28.46098 | 28.37848 | Rejected Model |
| SARIMA (3, 1, 3,1, 1, 1) 12 | 28.17999 | 28.44797 | Rejected Model |
| SARIMA (4, 1, 4, 1, 1, 1) 12 | 28.23832 | 28.58849 | Rejected Model |

Table 5 presents the estimation results from the SARIMA (1, 1, 1) (1, 1, 1) 12 model of change in tourism revenue.

**TABLE 5**: ESTIMATION RESULTS FOR SARIMA (1, 1, 1) (1, 1, 1) 12 MODEL

Dependent Variable: DREVENUE
Method: ARMA Maximum Likelihood (BFGS)
Date: 09/18/20 Time: 11:07
Sample: 2013M05 2019M12
Included observations: 80
Convergence achieved after 33 iterations
Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| AR(1) | 0.672366 | 0.116605 | 5.766168 | 0.0000 |
| SAR(12) | 0.992638 | 0.022039 | 45.04053 | 0.0000 |
| MA(1) | -0.970321 | 0.028162 | -34.45521 | 0.0000 |
| SMA(12) | -0.806878 | 0.259764 | -3.106200 | 0.0027 |
| SIGMASQ | 6.60E+10 | 1.41E+10 | 4.672880 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.726007 | Mean dependent var | 23497.43 |
| Adjusted R-squared | 0.711394 | S.D. dependent var | 493710.0 |
| S.E. of regression | 265231.2 | Akaike info criterion | 28.17853 |
| Sum squared resid | 5.28E+12 | Schwarz criterion | 28.32741 |
| Log likelihood | -1122.141 | Hannan-Quinn criter. | 28.23822 |
| Durbin-Watson stat | 1.824514 | | |

| | | | | |
|---|---|---|---|---|
| Inverted AR Roots | 1.00 | .87+.50i | .87-.50i | .67 |
| | .50+.87i | .50-.87i | .00+1.00i | -.00-1.00i |
| | -.50+.87i | -.50-.87i | -.87-.50i | -.87+.50i |
| | -1.00 | | | |
| Inverted MA Roots | .98 | .97 | .85+.49i | .85-.49i |
| | .49+.85i | .49-.85i | -.00-.98i | -.00+.98i |
| | -.49-.85i | -.49+.85i | -.85+.49i | -.85-.49i |
| | -.98 | | | |

Table 6 then summarizes the parameter coefficients and their standard errors for the estimated SARIMA (1, 1, 1) (1, 1, 1) 12 model.

**TABLE 6**: SUMMARY OF THE PARAMETERS OF THE ESTIMATED SARIMA (1, 1, 1) 1, 1, 1) 12 MODEL

| Parameter | Estimated coefficient | Standard error |
|---|---|---|
| Phi 1 ($\boldsymbol{\Phi_1 L}$) | 0.672366 | 0.116605 |
| Phi-season ($\boldsymbol{\Phi_1 L^s}$) | 0.992638 | 0.022039 |
| Theta 1 ($\boldsymbol{\theta_1 L}$) | -0.970321 | 0.028162 |
| Theta-season ($\boldsymbol{\Theta_1 L^s}$) | 0.806878 | 0.259764 |

### 4.4.1 Fitting the estimated model

Based on these estimation results, the final estimated model is then written as:

$$\Delta Revenue_t = 0.672366$$
$$\Delta Revenue_{t-1} + 0.992638\ SeasonAR_{t-12} - 0.970321\ \Delta\varepsilon_{t-1} - 0.806878\ SeasonMA_{t-12}$$

The coefficient of the first lag of the change in revenue is $0.672366$. This means that the change in revenue in the previous time period has a positive effect on the change in revenue in the current period. Specifically, when the change in revenue in the previous period was one dollar higher, the change in revenue in the current period is approximately 0.67 dollars higher, other factors held constant. The seasonal autoregressive and seasonal moving average terms are significant (P-value<0.05), implying that seasonality has a significant effect on the change in tourist revenue.

### 4.4.2 Testing the significance of coefficients

The significance of coefficients is tested individually using the student t-test. The null hypothesis that the respective coefficient is equal to zero is tested against the alternative hypothesis that the coefficient is statistically different from zero. For the three AR terms, the null hypothesis is rejected, given the P-values that are bigger than 0.05. This implies that the current value of change in revenue depends significantly on its past values. Additionally, the significance of seasonal AR and MA terms implies that monthly seasonality has a significant effect on changes in tourist revenue. It is also vital to note that the AIC for the model that accounts for seasonality is lower than the SIC for the model that ignores seasonality (Table 5). Therefore, accounting for seasonality in the SARIMA (1, 1, 1) (1, 1, 1) 12 model better fits the data than when seasonality is not accounted for.

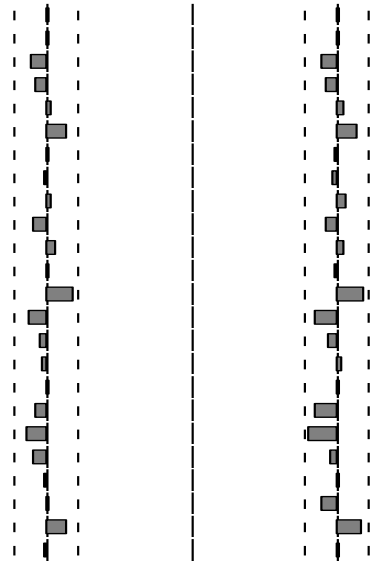
### 4.4.3 Diagnostic checks for SARIMA (1, 1, 1) (1, 1, 1) 12

#### 4.4.3.1 ACF and PACF plots of the associated residuals

Figure 7 shows the ACF and PACF of residuals from the estimated SARIMA (1, 1, 1) (1, 1, 1) 12 model which accounts for seasonality in tourist revenue. As indicated in the figure, there is no significant autocorrelation in the residuals of the underlying model. This clears the underlying model as appropriate for the estimation of tourist revenue. Figure 6 shows the ACF and PACF of residuals from the estimated SARIMA (1, 1, 1) (1, 1, 1) 12 model which accounts for seasonality in tourist revenue. As indicated in the figure, there is no significant autocorrelation in the residuals of the underlying model. This clears the underlying model as appropriate for the estimation of tourist revenue. Additionally, the absence of significant ACF and PACF in the residuals indicates that the model clarifies much of the variation in revenue and no substantial information remained unexplained (in the residuals). This is the model on which the forecast of future tourist revenues will be based. This is the model on which the forecast of future tourist revenues will be based.

**FIGURE 7**: CORRELOGRAM OF RESIDUALS AND LJUNG-BOX Q-STATISTICS FOR SARIMA (1 ,1, 1)(1,1,1)12

Date: 09/18/20 Time: 11:14
Sample (adjusted): 2013M05 2019M12
Q-statistic probabilities adjusted for 4 ARMA terms



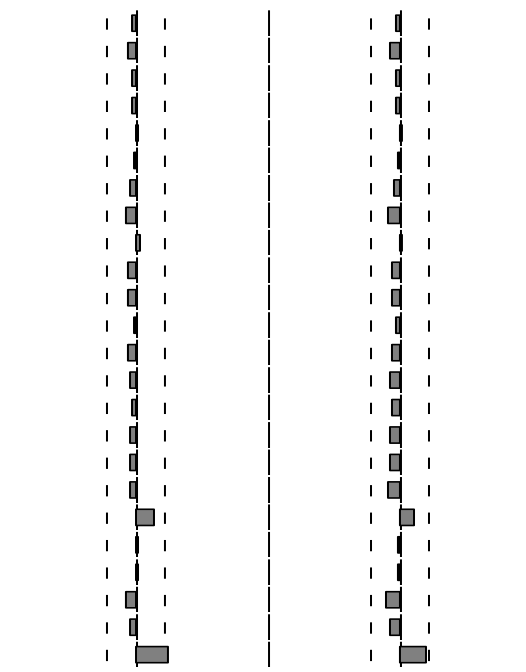
| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.015 | 0.015 | 0.0179 | |
| | | 2 | 0.016 | 0.016 | 0.0387 | |
| | | 3 | -0.102 | -0.102 | 0.9250 | |
| | | 4 | -0.080 | -0.078 | 1.4709 | |
| | | 5 | 0.049 | 0.055 | 1.6818 | 0.195 |
| | | 6 | 0.151 | 0.145 | 3.6920 | 0.158 |
| | | 7 | 0.005 | -0.016 | 3.6947 | 0.296 |
| | | 8 | -0.012 | -0.017 | 3.7076 | 0.447 |
| | | 9 | 0.040 | 0.080 | 3.8579 | 0.570 |
| | | 10 | -0.085 | -0.067 | 4.5298 | 0.605 |
| | | 11 | 0.073 | 0.055 | 5.0429 | 0.655 |
| | | 12 | 0.002 | -0.007 | 5.0432 | 0.753 |
| | | 13 | 0.198 | 0.203 | 8.8882 | 0.448 |
| | | 14 | -0.121 | -0.143 | 10.346 | 0.411 |
| | | 15 | -0.042 | -0.051 | 10.526 | 0.484 |
| | | 16 | -0.025 | 0.041 | 10.589 | 0.564 |
| | | 17 | 0.031 | 0.030 | 10.692 | 0.637 |
| | | 18 | -0.076 | -0.152 | 11.304 | 0.662 |
| | | 19 | -0.138 | -0.198 | 13.341 | 0.576 |
| | | 20 | -0.094 | -0.034 | 14.302 | 0.576 |
| | | 21 | -0.005 | 0.023 | 14.305 | 0.645 |
| | | 22 | 0.002 | -0.105 | 14.306 | 0.709 |
| | | 23 | 0.152 | 0.185 | 16.952 | 0.593 |
| | | 24 | -0.008 | 0.003 | 16.959 | 0.656 |

#### 4.4.3.2 Ljung-Box test for serial correlation in the residuals

The Ljung-Box Q-statistics for the estimated SARIMA (1, 1, 1) (1, 1, 1) 12 model are reported in Figure 8. The P-values for autocorrelation at all lags are bigger than 0.05 and hence, the null hypothesis of no serial correlation cannot be rejected. Therefore, the estimated SARIMA (1, 1, 1) (1, 1, 1) 12 model passes this diagnostic check as it does not suffer from a problem of autocorrelation in residuals.

**FIGURE 8**: CORRELOGRAM AND LJUNG-BOX Q-STATISTICS OF SQUARED RESIDUALS

Date: 09/18/20 Time: 11:15
Sample (adjusted): 2013M05 2019M12
Included observations: 80 after adjustments

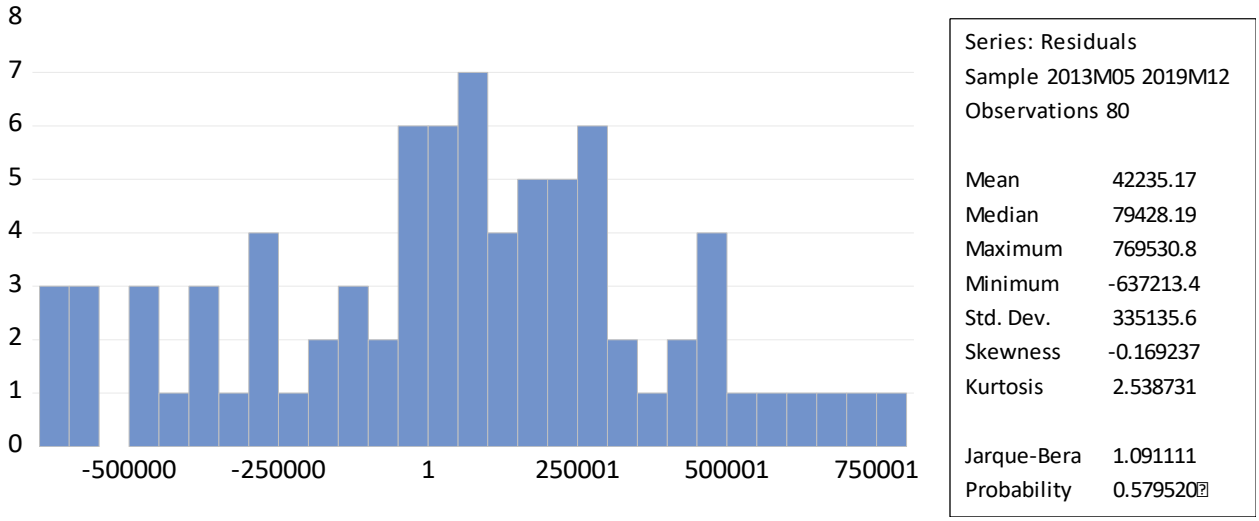



| | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|
| 1 | -0.025 | -0.025 | 0.0532 | 0.818 |
| 2 | -0.065 | -0.066 | 0.4076 | 0.816 |
| 3 | -0.028 | -0.031 | 0.4737 | 0.925 |
| 4 | -0.026 | -0.032 | 0.5310 | 0.970 |
| 5 | 0.014 | 0.009 | 0.5488 | 0.990 |
| 6 | -0.003 | -0.007 | 0.5498 | 0.997 |
| 7 | -0.046 | -0.047 | 0.7415 | 0.998 |
| 8 | -0.081 | -0.085 | 1.3348 | 0.995 |
| 9 | 0.033 | 0.022 | 1.4351 | 0.998 |
| 10 | -0.050 | -0.063 | 1.6653 | 0.998 |
| 11 | -0.053 | -0.062 | 1.9330 | 0.999 |
| 12 | -0.015 | -0.030 | 1.9545 | 0.999 |
| 13 | -0.050 | -0.063 | 2.1990 | 1.000 |
| 14 | -0.047 | -0.068 | 2.4188 | 1.000 |
| 15 | -0.025 | -0.052 | 2.4832 | 1.000 |
| 16 | -0.049 | -0.075 | 2.7295 | 1.000 |
| 17 | -0.041 | -0.067 | 2.9030 | 1.000 |
| 18 | -0.040 | -0.085 | 3.0764 | 1.000 |
| 19 | 0.140 | 0.109 | 5.1984 | 0.999 |
| 20 | 0.013 | -0.010 | 5.2164 | 1.000 |
| 21 | 0.002 | -0.012 | 5.2168 | 1.000 |
| 22 | -0.080 | -0.101 | 5.9366 | 1.000 |
| 23 | -0.048 | -0.075 | 6.2005 | 1.000 |
| 24 | 0.255 | 0.218 | 13.840 | 0.950 |

#### 4.4.3.3 Jarque-Bera test for the normality of residuals

Figure 9 shows the histogram and Jarque-Bera test for the normality of residuals. The figure reveals that the residuals are approximately normally distributed, given the relatively symmetric concentration of observations. Additionally, the Jarque-Bera test has a large P-value of 0.579520 which is bigger than 0.05. Therefore, the null hypothesis that the residuals are approximately normally dispersed with a zero mean and variance $\sigma^2$ cannot be rejected at the five percent significance level.

**FIGURE 9**: HISTOGRAM AND RESIDUAL NORMALITY TEST FOR SARIMA (1, 1, 1) (1, 1, 1) 12

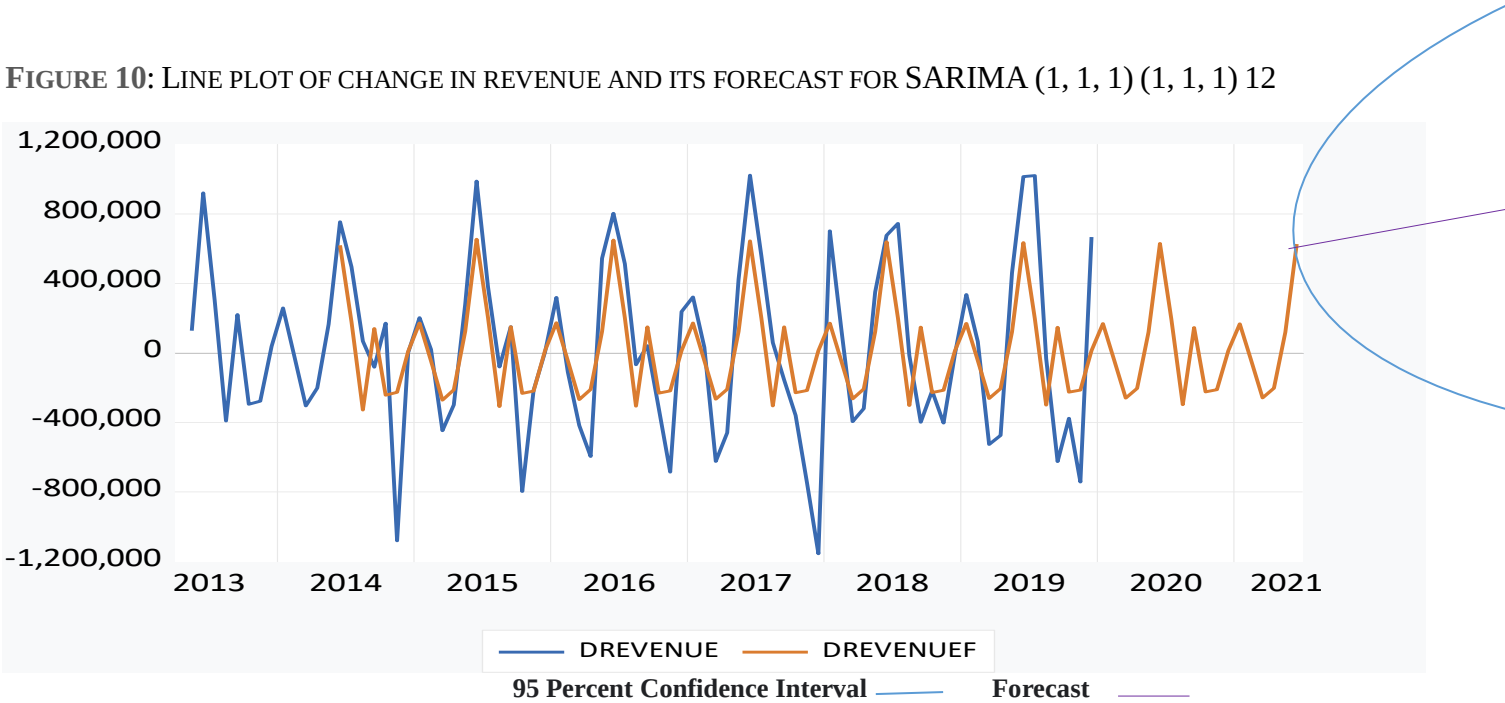



Therefore, based on all the three diagnostic checks, that is, (i) plots of ACF and PACF of residuals, (ii) Ljung Box Q-statistics and (iii) Jarque-Bera normality test, the estimated SARIMA (1, 1, 1) (1, 1, 1) 12 model is approved as a reliable model. Therefore, the final step of forecasting future changes in revenue is conducted based on the SARIMA (1, 1, 1) (1, 1, 1) 12 model.

### 4.5. Forecasting using the selected SARIMA (1, 1, 1) (1, 1, 1) 12 model

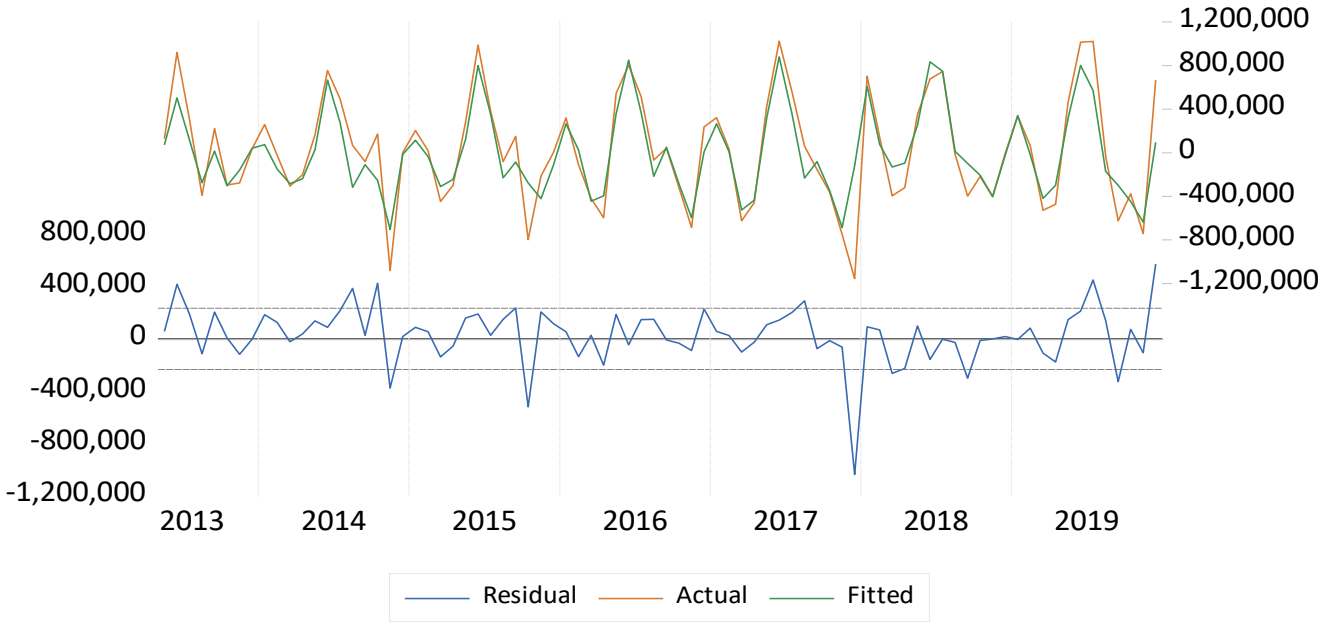
The future changes in tourist revenue are forecasted for the next 18 months between January 2020 and June 2021. This out of sample forecast is presented in Figure 10. The figure shows that the change in revenues will continue to exhibit the same trend for the next 18 months. Likewise, the series will continue to exhibit seasonality whereby months in the third quarter of the year are associated with higher revenue.

**FIGURE 10**: LINE PLOT OF CHANGE IN REVENUE AND ITS FORECAST FOR SARIMA (1, 1, 1) (1, 1, 1) 12




In order to check the quality/accuracy of the forecast, Figure 11 below presents within-sample forecasts and the actual and fitted values for the change in tourist revenue over the analysis period. The figure reveals that the fitted values of change in revenue are quite similar to the actual values. This further indicates that the forecast based on the SARIMA (1, 1, 1) (1, 1, 1) 12 model is a good forecast.

**FIGURE 11**: ACTUAL, FITTED AND RESIDUALS OF CHANGE IN REVENUE FOR SARIMA (1, 1, 1) (1, 1, 1) 12



Since the forecasts are for the change in revenue, negative values indicate that the value at a certain time period is expected to be lower than the value of revenue in the previous period (hence the negative values indicate a decline in revenue from the previous time period.

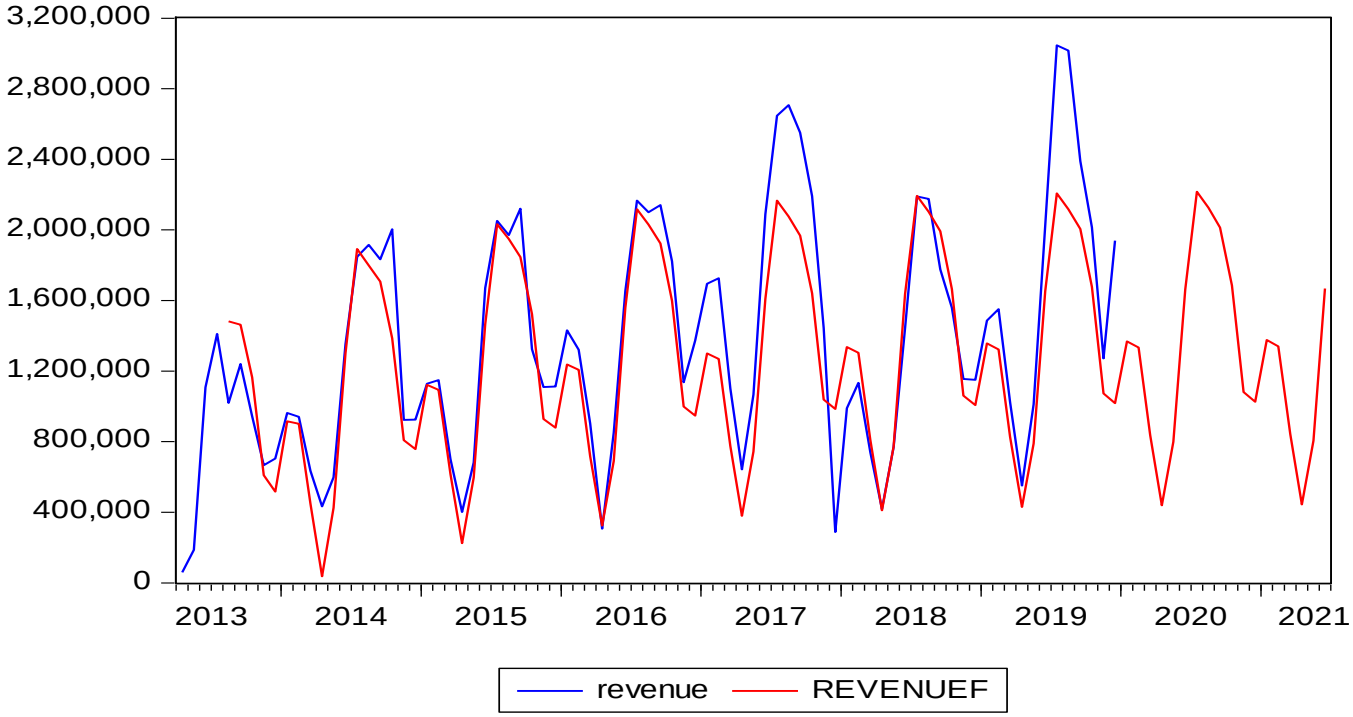
Table 7 below provides the values for the change in revenue and the forecast for the change in revenue for selected periods. Similar to the figure, the table also reveals that the change in tourist revenue will exhibit a similar pattern until June 2021.

**TABLE 7: ACTUAL AND FORECASTED REVENUE AND CHANGE IN REVENUE FOR SELECTED PERIODS: 2019-2021**

| PERIOD | REVENUE | REVENUEF | DREVENUE | DREVENUEF |
|---|---|---|---|---|
| 2019M01 | 1486301 | 1357462.428 | 336108 | 169570.4741 |
| 2019M02 | 1551553 | 1323179.393 | 65252 | -43877.31896 |
| 2019M03 | 1024681 | 826595.819 | -526872 | -262146.0407 |
| 2019M04 | 550452.2 | 429111.3442 | -474228.8 | -205596.4406 |
| 2019M05 | 1012193 | 789621.4749 | 461740.8 | 121160.849 |
| 2019M06 | 2026462 | 1654283.038 | 1014269 | 635208.3389 |
| 2019M07 | 3047017 | 2207854.596 | 1020555 | 197588.7529 |
| 2019M08 | 3015485 | 2117505.948 | -31532 | -299843.141 |
| 2019M09 | 2390498 | 2006086.063 | -624987 | 146709.475 |
| 2019M10 | 2014902 | 1677468.519 | -375596 | -226000.4131 |
| 2019M11 | 1272730 | 1073521.706 | -742172 | -213229.9711 |
| 2019M12 | 1939694 | 1018607.478 | 666964 | 12611.31997 |
| 2020M01 | | 1368939.892 | | 168322.0643 |
| 2020M02 | | 1334135.587 | | -43554.28585 |
| 2020M03 | | 837023.6042 | | -260216.0721 |
| 2020M04 | | 439064.7928 | | -204082.8009 |
| 2020M05 | | 799097.9462 | | 120268.84 |
| 2020M06 | | 1663326.355 | | 630531.8154 |
| 2020M07 | | 2216464.803 | | 196134.0672 |
| 2020M08 | | 2125721.582 | | -297635.6393 |
| 2020M09 | | 2013910.13 | | 145629.3722 |
| 2020M10 | | 1684933.109 | | -224336.5554 |
| 2020M11 | | 1080630.786 | | -211660.1318 |
| 2020M12 | | 1025389.033 | | 12518.47306 |
| 2021M01 | | 1375399.541 | | 167082.8455 |
| 2021M02 | | 1340297.207 | | -43233.63099 |
| 2021M03 | | 842893.0326 | | -258300.3123 |
| 2021M04 | | 444662.7637 | | -202580.3048 |
| 2021M05 | | 804431.0264 | | 119383.3981 |
| 2021M06 | | 1668412.561 | | 625889.7212 |

The figure below demonstrates the revenues and its forecast in the years of 2013 until June 2012

**Figure 12: Line plot of revenue and its forecast: 2013-2021**



# CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

## 5.1 Introduction

This study uses the Seasonal Autoregressive Integrated Moving Averages (SARIMA) to model monthly tourism revenues for Volcano and Nyungwe National Parks for the period 2013-2019. In addition, revenue is forecasted for the next 18 months between January 2020 and June 2021.The daily records of tourism revenue received from the parks were aggregated to form monthly values. This yielded 81 observations used in the analysis (12 months of seven years minus the first three months of 2013 when data was missing).

## 5.2 Summary of findings

This thesis uses ARIMA modelling by (Box-Jenkins ,1976) for the modelling of tourism revenues in Rwanda specifically in Volcano and Nyungwe National Parks using monthly information from the years 2013 to 2019. Due to the presence of seasonality in the data, seasonal adjustment was done, therefore updating the model to a Seasonal Autoregressive Integrated Moving Average (SARIMA) model. Using the Akaike Information Criterion (AIC) and Schwarz Information criterion (SIC), the SARIMA (1, 1, 1) (1, 1, 1) 12 model was selected as the most efficient model for the tourism revenues in Volcano and Nyungwe National Parks. In other words, the SARIMA (1, 1, 1) (1, 1, 1) 12 model had the lowest AIC value among all the candidate models estimated. This implies that it is the most reliable for both estimation and forecasting of tourism revenues.

The study makes three main findings. Firstly, there is a positive trend of the number of tourist arrivals and revenues for Nyungwe and Volcanoes national parks between 2013 and 2019. This finding reiterates the growing importance of the tourism sector to Rwanda's development. This trend is corroborated by global and national statistics that indicate that the country's tourism sector has been growing resiliently over the past decades (Rwanda Development Board, 2017). Secondly, the study presents tourism as one of the sectors with a high growth potential. This is corroborated by (English et al., 2016) who mentioned the tourism as the single most contributor to the country's GDP. Thirdly, the results of the study reveal fluctuations across some years. However, the greatest fluctuations are observed across months.

There is a high degree of seasonality in both arrivals and revenue over the analysis period. Specifically, both arrivals and revenues are highest in the months June-September, which coincides with the summer season in most countries of origin especially in Europe. In other words, it is mainly during the summer season that foreign tourists visit tropical countries like Rwanda partly because of the pleasant weather. Fourthly, tourist revenues are predicted to continue exhibiting a stable positive trend for the next 18 months until June 2021. This further reveals that the sector will perhaps continue maintaining is importance in national development. Lastly, analysis of the patterns of tourist arrivals indicate that international tourists account for the majority of visits. This is consistent with (Pankaj et al, 2017) who found that the tendency of using Nyungwe National Park to be higher among foreigners versus nationals. The study's findings have key policy implications. To begin with, the sector has potential for further expansion and contribution to the country's development agenda. Its continued contribution to GDP in the near future is seemingly robust. This requires policy orientation to fully harness the sector's potential. Specifically in the case of Nyungwe and Volcano national parks, more marketing could induce growth in both arrivals and revenue. Overall, the study's findings reiterate the need for concerted policy and regulatory measures including resource allocation to address gaps in the sector, promote it domestically and internationally with an ultimate goal of sustaining the observed positive trend and development impact. Since the prediction of tourist arrivals and revenues has often been challenging for policy makers, this study provides a partial remedy by conducting a comprehensive analysis of patterns, trends and predictions, which could guide informed decisions in the sector.

### 5.3 Conclusion and Recommendations

The importance of tourism in the development process cannot be under estimated. In Africa particularly, the sector is growing at high rates and is one of the single most contributors to GDP in countries like Rwanda. The growing importance of the sector also implies mounting scholarly and policy interest in further understanding the trends and patterns of the sector. This study contributes to this discourse by analyzing patterns and trends of tourist arrivals and revenues in Rwanda covering the period 2013-2019 and making predictions for the period 2020-2021. The analysis is based on tourist records from Nyungwe and Volcano national parks, which are used in this study as case studies. The analytical method used is the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, recognized by (Box and Jenkins, 1970) as a powerful tool for time series modelling and forecasting.

Seasonality was incorporated in the analysis using the SARIMA model. Among several alternative models, the SARIMA (1, 1, 1) (1, 1, 1) 12 model was chosen as one that best fits the underlying data based on the Akaike Information Criterion (AIC) and Schwarz Information criterion(SIC) .The model reveals a positive yet seasonal trend of tourist revenues for the two national parks across the study period. Forecasts based on the preferred model indicate that revenues are expected to maintain their past growth rates for the next 18 months until June 2021. The findings should however be interpreted with caution given some prevailing limitations. First, the combination of taxes and other incomes in the tourist statistics makes it hard to predict the former and their impact on the sector's performance. It is therefore recommended that data capture templates be modified to separate taxes from other incomes. Secondly, the revenue predictions were made before the corona outbreak. With the outbreak and lockdown in both the domestic and international markets, the predicted positive trend could be critically reversed. Further research is recommended in order to understand the pandemic's effect on the tourism sector and devise evidence-based policies and strategies to revive the sector in the post-COVID-19 period. Additionally, the study recommends further improvements in the way tourism statistics are recorded in order to make them more research-friendly and encourage further research on the topic as a way of steering evidence-based tourism policies and strategies.

### 5.4. Limitations/Challenges and further research

Nonetheless, this research shows some limitations that can pretense a breakthrough for prospect study in this topic. The first one is that my focuse was on volcano and Nyungwe national parks; thus the ability to generalize the conclusions is incomplete or limited. Another vital limitation is the outbreak of COVID-19 has not been taken into considerations during forecasting since my prediction based on the data and I hold other factors constant but corona virus outbreak will have great impact on the tourism industry both regional and globally. Further research is required in order to understand the pandemic's effect on the tourism sector and devise evidence-based policies and strategies to revive the sector in the post-COVID-19 period.

## References


Akal,M.(2004).Forecasting Turkey's tourism revenues by ARMAX model. *TourismManagement*, *25*(5), pp.565-580.

Ariano, R., Canonica, G.W. and Passalacqua, G. ( 2010). Possible role of climate changes in variations in pollen seasons and allergic sensitizations during 27 years. *Annals of Allergy, Asthma and Immunology*, *104*(3), pp.215-222.

Athanasopoulos,G.,Hyndman,R.J.,Song,H.andWu,D.C.(2011).Thetourismforecasting competition. *International Journal of Forecasting*, *27*(3), pp.822-844.

Belgiu, M. and Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote sensing of environment*, *204*, pp.509-523.

Blanco, L.F.L. and Hancco, R.W.M.(2020). Modeling and forecasting international tourism demand in Puno-Peru. *Revista Brasileira de Pesquisa em Turismo*, *14*(1), pp.34-55.

Brida, J.G., Cortes-Jimenez, I. and Pulina, M.(2016). Has the tourism-led growth hypothesis been validated? A literature review. *Current Issues in Tourism*, *19*(5), pp.394-430.

Brockwell, P.J., Davis, R.A. and Fienberg, S.E., 1991. *Time series: theory and methods: theory and methods*. Springer Science and Business Media.

Christie, I., Fernandes, E., Messerli, H. and Twining-Ward, L. (2014). *Tourism in Africa: Harnessing tourism for growth and improved livelihoods*. The World Bank.

Claveria, O. and Torra, S.(2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, *36*, pp.220-228.

English, P., McSharry, P. and Ggombe, K.(2016). Raising exports and attracting FDI in Rwanda. *IGC Policy Note*, *38402*.

Enright, M.J. and Newton, J.(2005). Determinants of tourism destination competitiveness in Asia Pacific: Comprehensiveness and universality. *Journal of travel research*, *43*(4), pp.339-350.

Fakriddinovich, E.R.( 2020). The Potentials of Tourism Industry in Uzbekistan and Effective Use of Them for Sustainable Development. *Indonesian Journal of Law and Economics Review*, *7*.

FaladeObalade, T.A. and Dubey, S.( 2014). Managing Tourism as a source of Revenue and Foreign direct investment inflow in a developing Country: The Jordanian Experience. *International journal of academic research in economics and management sciences*, *3*(3), pp.16-42.

Frechtling, D.( 2012). *Forecasting tourism demand*. Routledge. García Rodríguez, Ó. (2017).

Goh, C. and Law, R. ( 2011). The methodological progress of tourism demand forecasting: a review of related literature. *Journal of Travel and Tourism Marketing*, *28*(3), pp.296-317.

Gopal, B. and Biju, T., Relation between Residents' Perception on the Impact of Tourism and Their Support for Tourism Development.

Hillman, W.( 2007). Travel authenticated?: Postcards, tourist brochures, and travel photography. *Tourism Analysis*, *12*(3), pp.135-148.

HORVATH, Z. and Johnston, R. (2000). AR (1) Time series process. Econometrics 7590.

Bizimungu,J.(2019). In New Times. Retrieved from https https://www.busiweek.com/rwanda-gorilla-tourism-revenue-grows-by-25.

Kambogo, I. and Bizimana, J.P.(2016). Tourists' preferences for ecotourism planning and development around Nyungwe National Park, Rwanda. *Rwanda Journal*, *1*(1S).

Khan, A.A., Projections of Tourist Arrivals in the Eastern Black Sea Region of Turkey. *Coğrafi Bilimler Dergisi*, *18*(1), pp.85-98.

Lenzen, M., Sun, Y.Y., Faturay, F., Ting, Y.P., Geschke, A. and Malik, A.( 2018). The carbon footprint of global tourism. *Nature Climate Change*, *8*(6), pp.522-528.

Liu, X., Peng, H., Bai, Y., Zhu, Y. and Liao, L.( 2014). Tourism flows prediction based on an improved grey GM (1, 1) model. *Procedia-Social and Behavioral Sciences*, *138*(14), pp.767-775.

Loganathan, N. and Ibrahim, Y. (2010). Forecasting international tourism demand in Malaysia using Box Jenkins SARIMA application. *South Asian Journal of Tourism and Heritage*, *3*(2), pp.50-60.

Mak, A.H., Lumbers, M. and Eves, A.( 2012). Globalisation and food consumption in tourism. *Annals of tourism research*, *39*(1), pp.171-196.

Makananisa, M.P.( 2015). *Forecasting annual tax revenue of the South African taxes using time series Holt-Winters and ARIMA/SARIMA Models*. South Africa: University of South Africa.

Mbati, M. (2019). *Towards Improving the Skill of Seasonal Rainfall Prediction over Rwanda* (Doctoral dissertation, University of Nairobi).

Meo, M.S., Chowdhury, M.A.F., Shaikh, G.M., Ali, M. and Masood Sheikh, S., (2018). Asymmetric impact of oil prices, exchange rate, and inflation on tourism demand in Pakistan: new evidence from nonlinear ARDL. *Asia Pacific Journal of Tourism Research*, *23*(4), pp.408-422.

Mowforth, M. and Munt, I.( 2015). *Tourism and sustainability: Development, globalisation and new tourism in the third world*. routledge.

Peng, B., Song, H. and Crouch, G.I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, *45*, pp.181-193.

Petrevska, B.( 2012). Forecasting international tourism demand: The evidence of Macedonia. *UTMS Journal of Economics*, *3*(1), pp.45-55.

Rwanda Development Board. (2017). RDB Annual Report (2017). *Report, Rwanda Development Board*, Rwanda, December.

Statista, G. (2017). Travel and tourism industry–Statistics and facts.

Sugiyarto, G., Blake, A. and Sinclair, M.T. (2003). Tourism and globalization: Economic impact in Indonesia. *Annals of Tourism Research*, *30*(3), pp.683-701.

Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., Tan, H., Lin, D. and Wang, J. (2018). Trend analysis and forecast of PM2. 5 in Fuzhou, China using the ARIMA model. *Ecological indicators*, *95*, pp.702-710.