Forecasting Analysis of International Tourist Arrivals to Hyderabad, India, Using ARIMA Model

Narige Kiran[1*], D. Mallikarjuna Reddy[2]

[1]*Department of Mathematics, Anurag University, Venkatapur, Hyderabad-500 088, India.*

Email: narigekiran2004@gmail.com

[2]*Department of Mathematics, Gitam University, Rudraram, Hyderabad-502 329, India.*

Email:drmallikreddyd@gmail.com

Article Info

Page Number: 3801 - 3812

Publication Issue:

Vol 71 No. 4 (2022)

Abstract

This study aims to forecast the arrivals of tourists from various geographical regions across the world by employing time series and forecasting models. As we know, Tourism has been one of the rapidly growing and vital economic sectors worldwide. Similarly, Hyderabad is one of the most popular cities historically. Therefore, every year, progressively more people visit Hyderabad, which increases the tourism business in the country. The country gets a lot of money from tourism. It is imperative to start planning for tourists soon. There were a lot of international visitors to Hyderabad, India, every month from January 2014 to June 2019. This model is based on the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model. People who make ARIMA models use the Bayesian Information Criterion to choose the best one. Fitted ARIMA (9, 1, 4) can predict international tourist visits from July 2019 to December 2019 with a high level of accuracy, about 79.6%. A prediction of what will happen can deal with future problems. This kind of analysis is beneficial in lines of quality-of-service guarantee to the tourism industry to grow even more.

Article History

Article Received*: 25 March 2022*

Revised*: 30 April 2022*

Accepted*: 15 June 2022*

Publication*: 19 August 2022*

Keywords: International tourism, ARIMA, Forecasting methods, Bayesian Information Criterion.

I. Introduction.

Tourism is a significant one of the world's largest and fastest developing service businesses, contributing to about 10 per cent of global GDP. As a consequence of its rich historical and cultural past, A lot of people visit South India, and one of the best places to go in Telangana. Hyderabad, often renowned as the "City of Pearls," is one of India's more developed cities and a modern hub for informatics, ITES, and biotechnology. It is often recognized as "The City of Pearls." Hyderabad is acclaimed for its great heritage, environment, and architecture, representing the city's distinctive identity as a meeting platform for people from India and its multilingual culture and populace. Hyderabad was named the second greatest destination in the world to visit in 2015 by the annual guide of National Geographic's 'Traveler' magazine, which was released in December (2015).

The report placed India 34th out of 140 nations. India increased six places from last year's survey, the most among the top 25% rated. The survey ranks India's tourist industry 13th in terms of price competitiveness. It covers India's acceptable land and port infrastructure and its reasonable air transport system (ranked 33rd) (ranked 28th). It also ranks well in naturalresources, cultural resources, and business travel (ranked eighth). However, some components of its tourist infrastructure are still lacking. Compared to other countries, it has numerous hotel rooms per capita and few ATMs. According to the World Tourism Organization, India ranked 16th globally and 7th among Asian and Pacific nations(2012).

When it comes to Gross National Happiness (GNH), Hyderabad is an extraordinary place to visit because it has specific tourism policies, incredible natural attractions, growing and expanding private sectors with qualified people who know how to market it, and marketing outlets and relationships that are well-established It has grown into a highly profitable company with many possibilities for expansion. As a result, we developed a model and forecasted the number of individuals who would go to Hyderabad.

Lim and McAleer (2002) made their projections for Australian travel outside the country using the Box-Jenkins methodology. They used an ARIMA forecasting model. Malaysia, Singapore, and Hong Kong used data from 1975 to 1989. They demonstrate how to choose the most appropriate model and use it for forecasting. Before beginning construction on the model, the authors conducted both a unit root test and a seasonal unit root test to examine the data. This test treats the data as stable even though the vast majority of it relates to tourist seasons. Evaluations are performed at each stage of the modelling process in order to improve the precision of the projections provided by the models. Hong Kong has ARIMA (3, 1, 1), Malaysia has ARIMA (3, 1, 2), and Singapore has ARIMA (4, 1, 0). The ARIMA model had a higher degree of precision than the second and third models (3, 1, 1).

From 1997-2007, how many nights were spent in New Zealand? Lim et al. (2009) employed Box Jenkins autoregressive, moving average, and SARMA models. Model A was utilised. Model A uses Holt-Winter triple exponential smoothing.There were two pairs of things then that were compared to each other The Box-Jenkins method was used to make the ARMA and SARMA models. Those were the best ones because they were the same volume and had the exact contours (0, 1, 1). Model A forecasts that visitor nights will go down in the research, while Model B anticipates going up. Research on predicting foreign visitor arrivals using time series and regression methods has been done. They say that regression models cannot be used in this study because guest nights could be international and domestic. This means that they cannot use them. It is not good because it does not have enough demand characteristics.

For Australia's economy, China is a crucial contributor since it is a primary source of tourists. In 2015, research utilizing the SARIMA model was presented to estimate Chinese tourist visits to Australia, according to Emily Ma et al. (2016). Their data came from the Australian Bureau of Statistics monthly to construct their computations, and they employed STATA 13 Zero (zero), which is the ARIMA model (0, 1, 1). Using the best-fitting model, we estimated Chinese visitors' arrivals three periods in advance. According to the study's conclusions, visitors from China to Australia are expected to remain seasonal for the next two years. We will do a time series analysis using the available data. Taking into account a huge range of models, a large number of scholars have contributed their time and effort to this project in order to investigate the features of forecasting the number of visitors from other countries. (Nanthakumar et al. (2010); Box et al. (2015); Singh(2013); Makridakis et al.(2008); Manohar(2021), Peiris, H. (2016), Chu, F. L. (2009), Petrevska, B. (2017), Cho, V. (2003), Chang, Y. W., & Liao, M. Y. (2010), Baldigara, T., &Mamula, M. (2015), Turner, L., Kulendran, N., & Fernando, H. (1997)).

2. Methodology

2.1 Datacollection& Study Area

This study uses secondary data collected and maintained by the state of Telangana (data.telangana.gov.in). Foreign tourist arrivals in Hyderabad are included in the data set from January 2014 to June 2019, and information on the data's accessibility and relevance to the country's tourism demand.

2.2 Statisticalanalysis

For this study, the Box-Jenkins (2015) time series model is employed to evaluate the arrival of international visitors in Hyderabad. The following are the steps to putting the Box-Jenkins model into action: Begin by gathering and classifying all of the various design options. This involves giving the AR, MA, ARMA, and other model orders and additional model orders. The Autocorrelation function (ACF) and Partial autocorrelation function (PACF) are used to identify the series of interest in this process. The next step is to calculate the model's coefficients. AR model coefficients may be calculated using least squares regression. Although estimating a parameter often involves more complex iteration processes, computer programming generates it.

In the last phase, the model is tested. This is sometimes referred to as diagnostic testing or verification. The residuals must be white noise, and the computed parameters must be statistically significant. The Bayesian Information Criterion (BIC) is used to pick the best ARIMA model from a group of competing models that may be acceptable for the series. Forecasts are made based on the chosen model.

2.3 Auto-Regressive Moving Average Model (ARIMA)

In the event that the process does not remain stationary, we will have to incorporate a differencing term known as $(1-B)^d$ into the process in order to account for this. According to (Choden&Unhahpipet (2018)), a model is said to be an ARIMA (p, d, q) model when it is used to a time series which has been differenced d times and then the model ARMA (p, q) is used to analyse the data. As a consequence of this, it is possible to describe the generic ARIMA (p, d, q) using the backshift operator as shown below:

$$\phi_p(B)(1-B)^d = \theta_0 + \theta_q(B)a_t \tag{1}$$

Where $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots\cdots - \phi_p B^p$ is the stationary AR operator,

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is the invertible MA operator and

$\theta_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$ is deterministic trend term: (Choden&Unhahpipet(2018)).

2.4 Ljung-BoxTest

TheLjung–Boxtestmaybedefinedinterms ofhypothesisas:

H_0: Thedataareindependentlydistributed(absenceofserialcorrelation)

H_1: Thedataarenotindependentlydistributed(presenceof serialcorrelation)

The test statistics is $Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$ (2)

Where n=sample size, $\hat{\rho}_k$ =sample autocorrelation at lag k, and h=number of lags being tested.

2.5 BayesianInformationCriterion(BIC)

Bayesianinformationcriterion(BIC)isacriterionformodelselectionamongafinitesetofdeveloped models.The lowerthe BICvalue, the betterthemodelis.

Mathematically BIC can be written as

BIC= $\ln(n)k - 2\ln(\hat{L})$; (3)

Where, $\hat{L}$ =maximum value of likelihood function of the model, n=number of data points, k=number of free parameters to be estimated.

The analysis was executed using SPSS Version 26.

Numerical Results& Discussions

ARIMA (p, d, q) models were used to anticipate the arrival of foreign tourists in Hyderabad in the future. Determining the three parameters p, d, and q is crucial to the Box-Jenkins technique. ACFvalues and PACFvalues are used to determine the appropriate values of p and q parameters.

As seen in fig.1, the first stage in establishing a model is creating visuals for the data series.

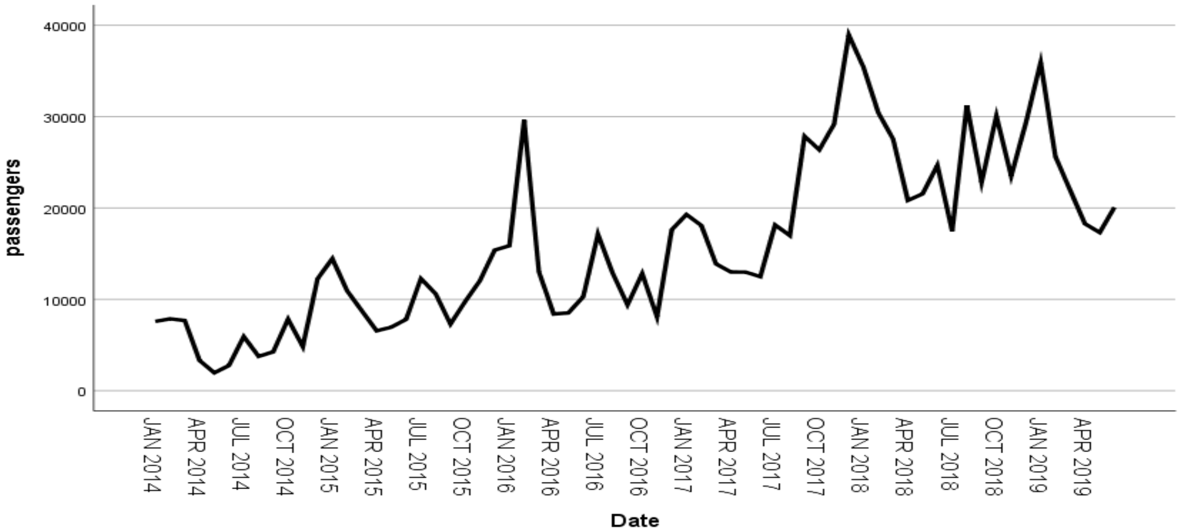

Fig.1 Trends of International Tourists

As seen in fig.1, the data series must be differentiated because of the trend component. The ACF and PACF of tourist arrival data are crucial for clearly confirming the data series's stationary and auto-regressive components. Figs.2 and 3 shows the autocorrelation function and partial autocorrelation functions, respectively

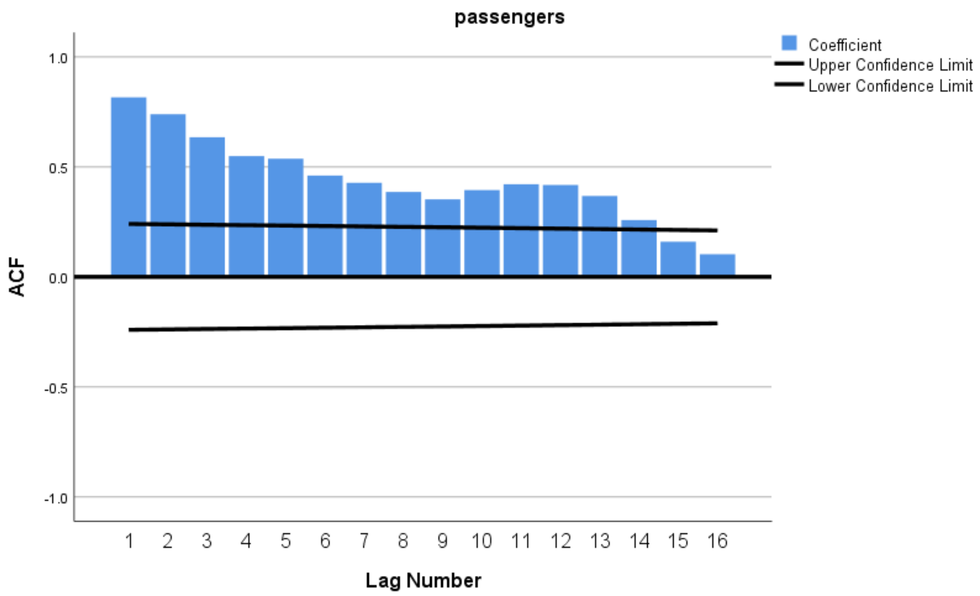

Fig.2: Autocorrelations function.

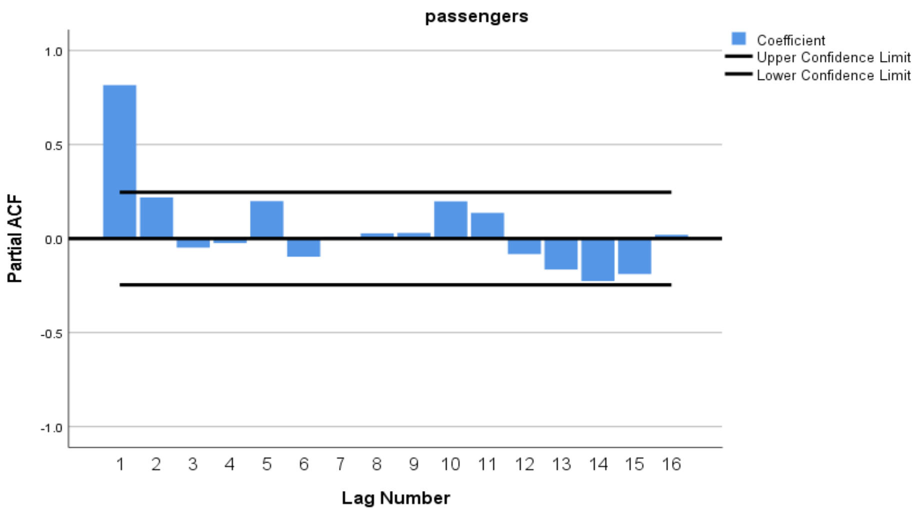

Fig.3: Partial Autocorrelations function.

Figs. 2 and 3 demonstrate strong correlations. Those who become statistically significant even at greater delays, and their significance gradually declines with increasing time. This is clear evidence of non-stationary series. There is statistical significance in the Ljung-Box data even for non-stationary series. Next, the researchers want to determine whether the tourists' arrivals are random strolls or not. It is possible to see this in fig.4, which shows a stationary time series as the initial discrepancy. Stable series may be seen in the plot, with the mean and variance remaining relatively stable throughout time.

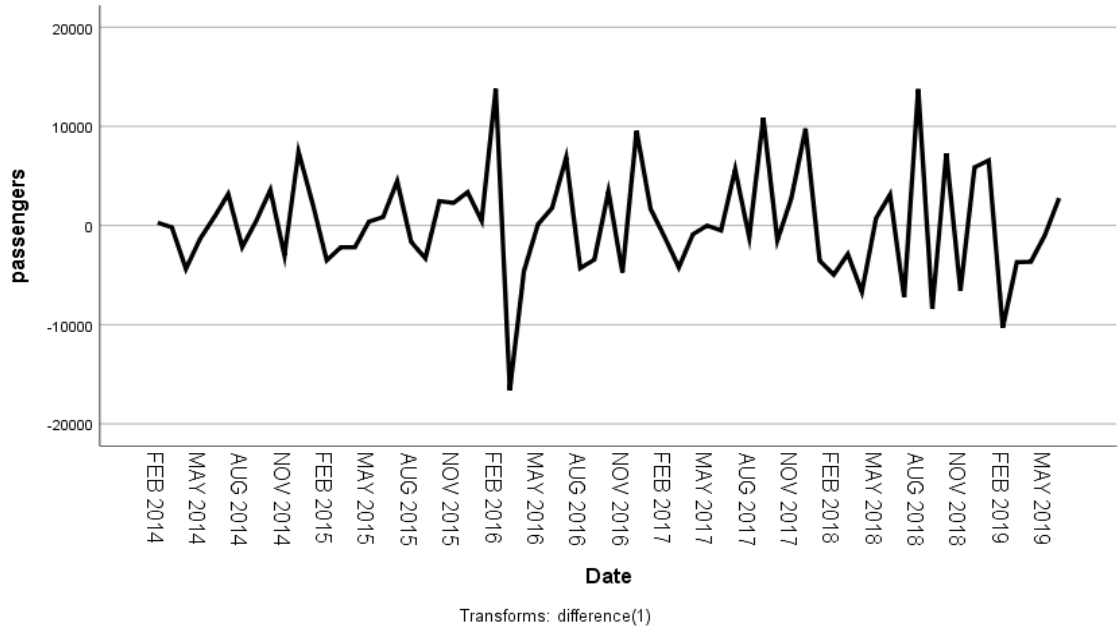

Fig.4 First-difference graph

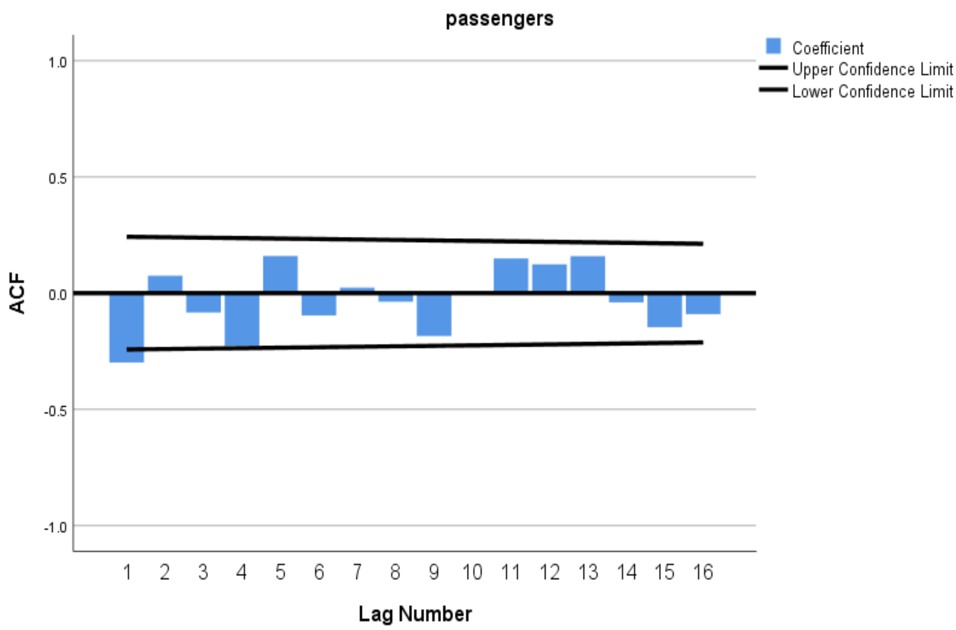

Fig.5 ACF, after the first differencing

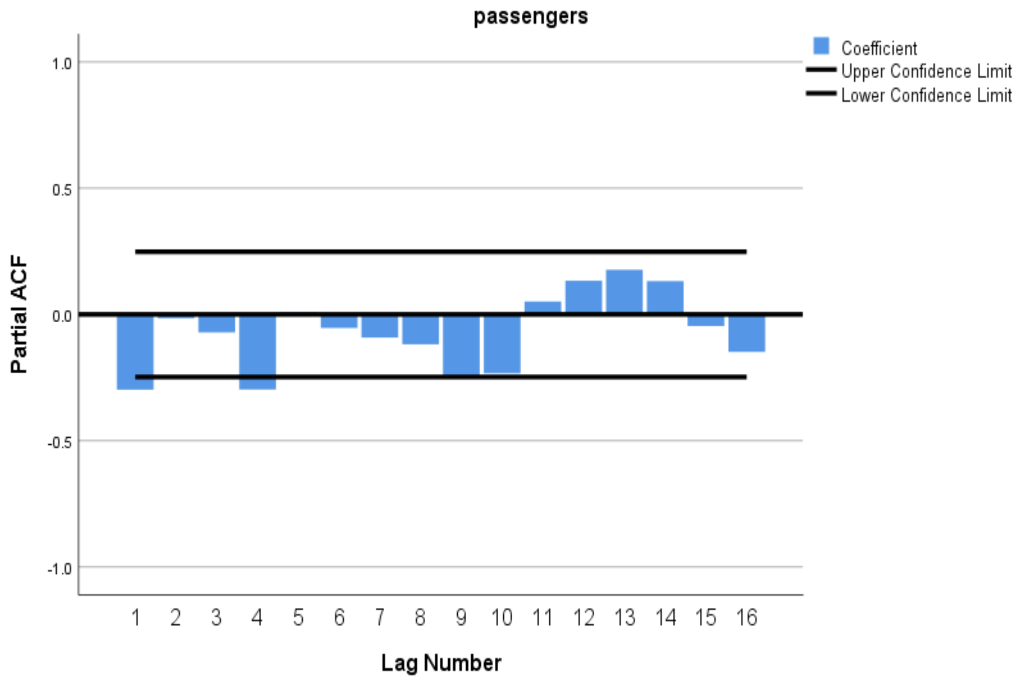

Fig.6PACF, after the first differencing

ACF and PACF samples are shown in Figs.5 and 6 with 95 % confidence limits and 16 lag values for comparison. The ACF and PACF spikes are illuminated. As a result, the confidence intervals indicate that the sample ACF spiked at lag 1 and 4, while the sample PACF spiked at lag 1, 4, and 9 with the initial differencing (outside of the confidence intervals). As a result, the preliminary ARIMA models are as follows: (1, 1, 1), (1, 1, 4), (4, 1, 1), (4, 1, 4),(9, 1, 1) and (9, 1, 4). Several factors are considered while picking the best-fitting model from among the available models: the BIC, the Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE), the Mean Average Error (MAE), and the R^2 coefficient. Table.1 contains the models that were developed. ARIMA (9, 1, 4) is found to have the lowest BIC, least root mean square error, least mean absolute deviation (MAD), least mean absolute deviation (MAPE), and greatest R^2 ,suggesting that it is the most appropriate model for forecasting monthly international tourist arrivals through Hyderabad between January 2014 and July 2019.

Table:1 International tourist arrivals to Hyderabad: R^2, RMSE, MAPE, MAE, BIC.

Model	R^2	RMSE	MAPE	MAE	BIC
ARIMA(1,1,1)	0.713	4973.242	29.446	3761.141	17.216

ARIMA(1,1,4)	0.734	4905.627	28.862	3636.976	17.382
ARIMA(4,1,1)	0.707	5146.605	28.641	3798.10	17.478
ARIMA(4,1,4)	0.741	4969.852	28.989	3665.87	17.600
ARIMA (9, 1, 1)	0.749	4975.723	26.124	3342.905	17.731
ARIMA (9, 1, 4)	**0.796**	**4625.054**	**25.541**	**3154083**	**17.77**

The goodnessoffit/diagnosticchecking

The selection of the best fit model in time series modeling is significantly linked to the quality of the analysis of the residuals analysis is conducted appropriately. The model ARIMA assumes that the residual should be white noise if the model is accurate.

Relative autocorrelation is shown in Fig.7 to be around 0, suggesting that the residuals are uncorrelated and do not reveal any patterns. The model makes the assumption that now the residuals have a zero mean and variance with one. Now that all lag orders have been considered, it is clear that the Ljung-Box statistic has a reasonable chance of being accurate (P=0.357). All model assumptions may be met because of the selected ARIMA models (9, 1, 4).

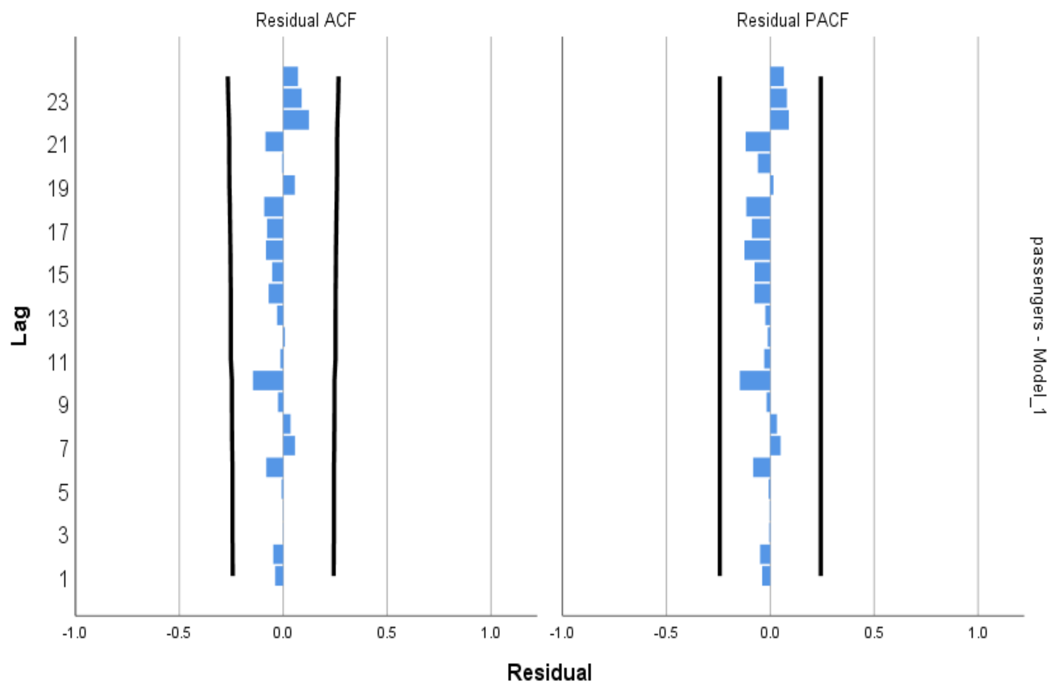

Fig.7 ACF and PACF samples of the residuals from ARIMA (9, 1, 4)

ForecastingusingARIMA(9,1,4)

Table.2 shows the ARIMA (9, 1, 4) model's anticipated values for monthly foreign visitor arrivals in Hyderabad, with standard error, lower and upper bounds, whereas Fig.8 shows the actual forecast values. Table.2 shows the ARIMA (9, 1, 4) model's anticipated values for monthly foreign visitor arrivals in Hyderabad, with standard error, lower and upper bounds, whereas Figure.6 shows the actual forecast values.

Table:2 International visitor arrivals during the first six months of 2019.

Date	Touristvisited	Forecasted	95% Confidence level	
			Lower	Upper
July-2019	27874	25091	16119	34063
Aug-2019	22755	29066	19091	39042
Sep-2019	34018	28992	17621	40362
Oct-2019	28706	28385	16451	40318
Nov-2019	32762	30508	17980	43037
Dec-2019	34084	29358	15633	43083

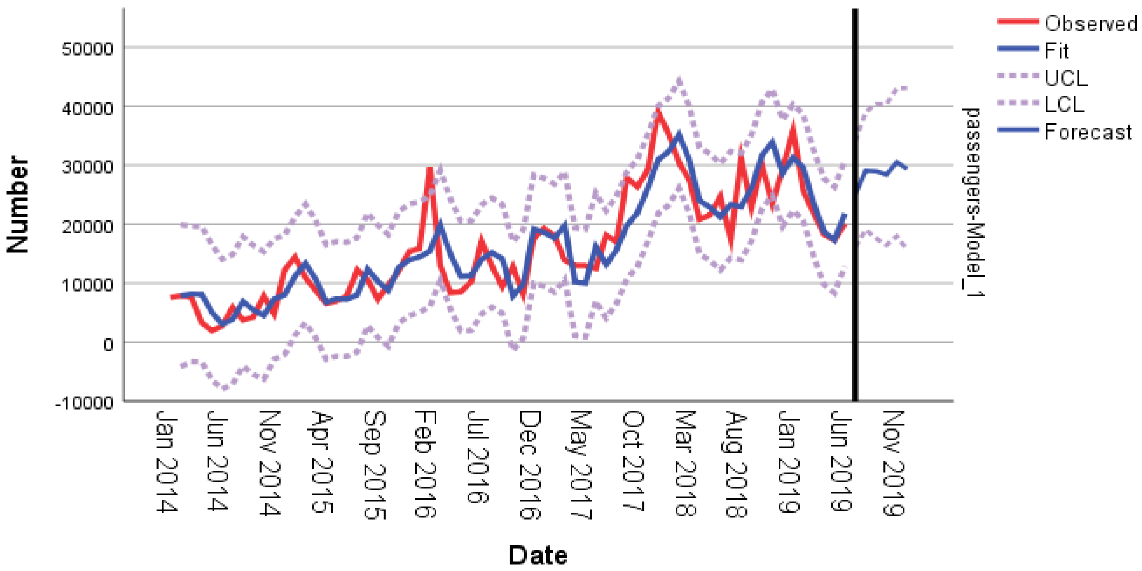

Fig.8 From January 2014 to June 2019, the ARIMA (9, 1, 4) values of monthly foreign visitor arrivals in Hyderabad were compared to the actual and forecasted values.

Conclusions

Some of the following inferences may be drawn from the model as mentioned above fits, result analysis, and findings: A univariate time series model was used for this investigation, based on data from the previous months' international visitor arrivals in Hyderabad. We calculated the number of foreign visitors expected between January 2014 and June 2019 using various ARIMA models, varying (p, d, q), and found that ARIMA (9, 1, 4) provided the most accurate results. We matched the best-predicted values with 79.6% accuracy from July 2019 to December 2019. This kind of research is quite helpful in ensuring that the tourist sector in Hyderabad and India continues to expand in terms of quality of service.

References:

[1] Tourism in India(2015) https://en.wikipedia.org/wiki/Tourism_in_India_by_state#Telangana

[2] Tourism in India(2012) https://en.wikipedia.org/wiki/Tourism_in_India.

[3] Lim, C., & McAleer, M. (2002). Time-series forecasts of international travel demand for Australia. *Tourism Management*, *23*(4), 389-396.

[4] Lim, C., Chang, C., & McAleer, M. (2009). Forecasting h (m) otel guest nights in New Zealand. *International journal of hospitality management*, *28*(2), 228-235.

[5] Ma, E., Liu, Y., Li, J., & Chen, S. (2016). Anticipating Chinese tourists arrivals in Australia: A time series analysis. *Tourism management perspectives*, *17*, 50-58.

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

[7] Box, G. E., Jenkins, G. M., Reinsel, G. C., &Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

[8] Singh, E. H. (2013). Forecasting tourist inflow in Bhutan using seasonal ARIMA. *International Journal of Science and Research*, *2*(9), 242-245.

[9] Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*. John wiley& sons.

[10] Dingari, M. (2021). Air Traffic Data Analysis Using Seasonal ARIMA. *Design Engineering*, 15876-15884.

[11] Choden., &Unhapipat, S. (2018, June). ARIMA model to forecast international tourist visit in Bumthang, Bhutan. In *Journal of Physics: Conference Series* (Vol. 1039, No. 1, p. 012023). IOP Publishing.

[12] Peiris, H. (2016). A Seasonal ARIMA Model of Tourism Forecasting: The Case of Sri Lanka. *Journal of Tourism, Hospitality and Sports*, *22*(1), 98-109.

[13] Chu, F. L. (2009). Forecasting tourism demand with ARMA-based methods. *Tourism Management*, *30*(5), 740-751.

[14] Petrevska, B. (2017). Predicting tourism demand by ARIMA models. *Economic research-Ekonomskaistraživanja*, *30*(1), 939-950.

[15] Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism management*, *24*(3), 323-330.

[16] Chang, Y. W., & Liao, M. Y. (2010). A seasonal ARIMA model of tourism forecasting: The case of Taiwan. *Asia Pacific journal of Tourism research*, *15*(2), 215-221.

[17] Baldigara, T., &Mamula, M. (2015). Modelling international tourism demand using seasonal ARIMA models. *Tourism and hospitality management*, *21*(1), 19-31.

[18] Turner, L., Kulendran, N., & Fernando, H. (1997). The use of composite national indicators for tourism forecasting. *Tourism Economics*, *3*(4), 309-317.