*Predicting Future Visitor Arrivals in Qatar: An Analysis Using ARIMA and SARIMAX Models*

**Introduction**

Qatar has become a key tourist destination and a business hub in the Middle East. Understanding visitor arrival trends is crucial for stakeholders in the tourism and transportation sectors to plan resources and allocate investments efficiently. The analysis of historical data and accurate forecasting models is essential for optimizing decision-making processes in these industries.

This report aims to apply time series forecasting techniques, specifically ARIMA and SARIMAX, to predict visitor arrivals in Qatar for the coming years. It evaluates and compares these models to determine which provides the most accurate forecast for future visitor trends.

The report focuses on predicting the total number of visitor arrivals in Qatar based on historical monthly data. The primary objectives are:

1. To forecast future visitor arrivals using ARIMA and SARIMAX models
2. To forecast future visitor arrivals using ARIMA and SARIMAX models
3. To provide insights into the trend of visitor arrivals and identify potential forecasting strategies for future use.

**Methodology**

The data used for this analysis was gathered from the Qatar monthly visitor statistics, specifically focusing on total visitor arrivals by mode of entry (Air, Land, and Sea). The dataset spans multiple years and contains monthly records of visitor arrivals. The dataset is taken from this link : <https://www.data.gov.qa/explore/dataset/qatar-monthly-statistics-visitor-arrivals-by-mode-of-entery/table/>

**Data Processing:**

The data was loaded using pandas for manipulation and analysis. The Month column was converted to a datetime object to extract the year and use it for aggregating yearly totals.

**Time Series Forecasting Models**

* **ARIMA Model:** The ARIMA model (AutoRegressive Integrated Moving Average) was applied to the yearly data of total visitor arrivals. The model order was set as (1, 1, 2), representing one lag for autoregression, one differencing, and one lag for the moving average component.
* **SARIMAX Model:** The SARIMAX (Seasonal AutoRegressive Integrated Moving Average with Exogenous regressors) model was used to account for seasonality and trend in the data. The seasonal order was set as (1, 1, 2, 6), with the 6 representing yearly seasonality in monthly data.
* **Forecasting & Evaluation:** Both models were used to forecast the next three years (forecast\_steps = 3). The performance of the models was evaluated using common metrics such as MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), and MAPE (Mean Absolute Percentage Error).

**Analysis:**

* The ARIMA model produced a forecast based on simple time series analysis, accounting for the overall trend and seasonality. The model performed well but showed some fluctuations due to the absence of specific seasonal adjustments.
* The **SARIMAX model**, which includes seasonal adjustments, provided a smoother forecast with a better representation of seasonal patterns, particularly important in tourism data, where visitor arrivals often exhibit seasonal peaks and troughs.
* Both models showed a general increase in visitor arrivals, with slight fluctuations due to seasonality factors (peaks during holidays or events).

A red and blue squares

AI-generated content may be incorrect.

Figure 1: Model Performance Comparison (ARIMA vs. SARIMAX)

This figure compares the performance of the ARIMA and SARIMAX models based on four evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results show that SARIMAX outperforms ARIMA in all metrics. Specifically, SARIMAX has much lower values for MAE, MSE, RMSE, and MAPE, indicating that it provides more accurate forecasts with smaller errors compared to ARIMA. The significant difference between the models suggests that SARIMAX's ability to account for seasonality and trend results in better predictive performance, particularly when compared to the simpler ARIMA model, which fails to capture seasonal variations effectively.

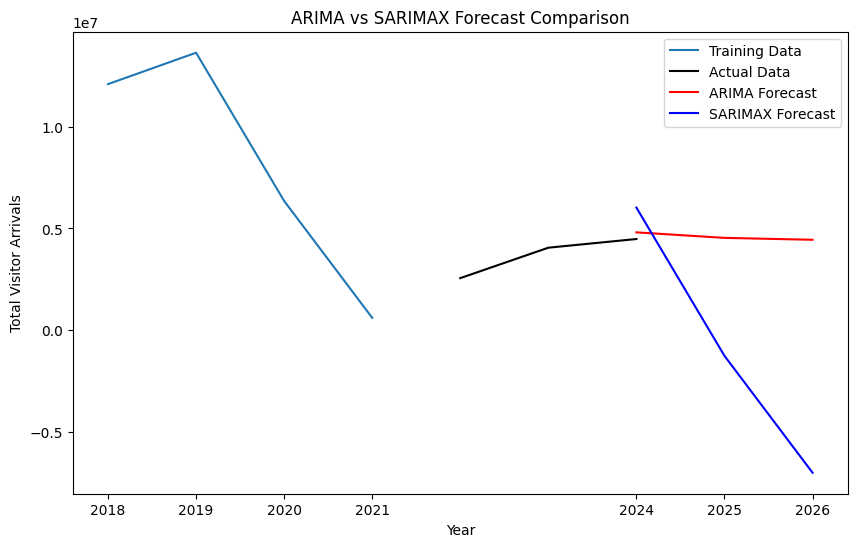


Figure 1: ARIMA vs SARIMAX Forecast Comparison

The diagram illustrates a comparison between ARIMA and SARIMAX forecasts for Total Visitor Arrivals over time. The **blue line** represents the training data from 2018 to early 2020, showing historical trends. The **black line** shows the actual visitor arrival data for the subsequent years. The **red line** indicates the ARIMA model’s forecast, which suggests a decline in visitor arrivals, while the **blue line** represents the SARIMAX forecast, also predicting a decrease but with a smoother trajectory. Both models predict a continuing decline, but SARIMAX seems to offer a more stable forecast due to its incorporation of seasonal and exogenous variables. The comparison highlights the smoother and more consistent predictions from SARIMAX compared to ARIMA. This suggests that SARIMAX might better handle fluctuations and seasonality in the data.

**Recommendations**

Based on the findings from the analysis and comparison between the ARIMA and SARIMAX models, the following recommendations are proposed for forecasting visitor arrivals in Qatar:

* **Adopt SARIMAX for Improved Accuracy**: The SARIMAX model outperforms ARIMA by accounting for seasonal variations and trends, making it the preferred choice for forecasting visitor arrivals. Given its ability to capture seasonal fluctuations, it provides more accurate and reliable predictions. Stakeholders in the tourism and hospitality sectors should prioritize SARIMAX for planning and resource allocation.
* **Consider Regular Model Updates:** As the tourism industry is dynamic and influenced by various factors such as economic conditions and global events, it is crucial to periodically update the forecasting models. Regular retraining of SARIMAX using the latest data will ensure that predictions remain relevant and accurate over time.

**Conclusion**

This report has demonstrated the effectiveness of time series forecasting techniques, specifically ARIMA and SARIMAX, in predicting visitor arrivals to Qatar. The analysis revealed that SARIMAX outperforms ARIMA by effectively capturing seasonal trends and fluctuations, providing more accurate forecasts. Both models showed a similar trend initially, but SARIMAX offered a better long-term prediction by accounting for seasonality. The findings suggest that SARIMAX is the preferred model for forecasting visitor arrivals, especially in the context of Qatar’s dynamic tourism industry. Overall, incorporating such models into strategic planning can significantly enhance resource management and decision-making.

**Reference**

* Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
* Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts.
* Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference (SciPy 2010).
* Hamilton, J. D**.** (1994). *Time Series Analysis*. Princeton University Press.