Enhancing Educational and Tourism Applications through Predictive Modeling of Cultural Heritage Site Visitation: use of Arima and autoregressive models

Pablo Rosser

pablo.rosser@unir.net

Universidad Internacional De La Rioja

Seila Soler

Universidad Isabel I de Castilla

Additional Declarations: No competing interests reported.

Abstract

This study focuses on the use of ARIMA and Autoregressive (AR) models to predict visitor flow to Civil War shelters in Alicante, highlighting seasonal patterns and differences among various visitor groups, with an enriching approach towards educational and tourism applications. Through a retrospective longitudinal design covering from August 2023 to January 2024, it analyzes the time series of visits, differentiating between the general public and school groups, as well as examining geographical demand. The research emphasizes the effectiveness and simplicity of the ARIMA(0, 0, 0) model with Logarithmic Transformation in modeling time series, while the AR(6) model proves indispensable for capturing short-term temporal dependencies. Despite the usefulness of these forecasts for future planning, the existence of uncertainties highlights the importance of adopting flexible management approaches and incorporating additional variables to refine predictions. This approach not only improves the management of visitor flows but also significantly contributes to the creation of more effective educational and tourism strategies, promoting the sustainability and appreciation of cultural heritage.

1. Introduction

In scientific terms, autocorrelation refers to the quantification of the linear interdependence between successive observations of a time sequence, comparing current instances with their counterparts in previous intervals, known as lags. A high magnitude of autocorrelation suggests that historical observations exert a significant influence on contemporary values. As the temporal distance between compared observations increases — that is, the number of lags —, the magnitude of autocorrelation tends to decrease, though these values remain statistically significant ($p < 0.001$). This characteristic is common in time series analysis, indicating that, even at extended temporal distances, there is a non-random linear correlation between successive observations (Azad et al., 2022; Bottomley et al., 2023; Cicuéndez et al., 2023; Gao et al., 2019).

The ARIMA (AutoRegressive Integrated Moving Average) model is a statistical model used to analyze and predict time series. It is capable of capturing a range of standard patterns in temporal data for future projections. ARIMA combines three basic components: autoregressive models (AR), integrated differentiation (I), and moving averages (MA) (Faujdar & Joshi, 1 C.E.; Lin, 2023).

The ARIMA model is an essential analytical tool for forecasting and analyzing time series, consisting of three key elements: Autoregressive (AR) p, Integrated (I) d, and Moving Average (MA) q. The AR p component captures the relationship between an observation and its past values, basing the prediction on the historical behavior of the series. "p" represents the number of past observations considered. The I d component focuses on the differentiation of the time series, essential for achieving stationarity, a condition where the statistical properties of the series, such as mean and variance, remain constant. "d" indicates the degree of differentiation needed. The MA q component models the error in the prediction from a combination of past errors, allowing for the capture of unexpected temporal variations. "q" refers to the number of error terms incorporated into the model.

Determining the optimal parameters for p and q is performed through the analysis of autocorrelation (ACF) and partial autocorrelation (PACF) graphs. These graphs are fundamental for identifying the appropriate structure of the ARIMA model, allowing the determination of the necessary number of AR and MA terms by observing significant correlation bars outside the confidence zones, thereby adjusting the model to accurately reflect the analyzed time series dynamics.

1. *RESEARCH OBJECTIVES:*

General objective of the research: Analyze and predict the influx of visitors to memorial cultural heritage through the application of ARIMA and Autoregressive statistical models, to better understand and manage the impact of visits to the Civil War shelters in Alicante.

Specific objectives

1. Investigate the variability in attendance at the shelters between August 2023 and January 2024 to identify seasonal patterns and peaks of visits.

2. Examine differences in visit frequency between the general public and school groups, highlighting distinct preferences and typologies.

3. Analyze the geographical distribution of interest in the shelters to locate areas of high demand at various levels.

4. Apply ARIMA and AR models to represent the dynamics and temporal dependencies of visits.

5. Use statistical models to project future visitor influxes, assessing accuracy and associated uncertainty.

6. Implement unit root tests, such as Dickey-Fuller, to verify stationarity and determine the need for time series differentiation.

7. Compare ARIMA and AR models based on their simplicity, relevance of coefficients, and predictive efficacy to choose the most suitable one.

8. Study autocorrelations and partials to understand the influence of past values on future ones and adjust the relevant model.

9. Future Forecast Evaluation: Project short-term visitor influx using the selected model, as a tool for planning and decision-making.

10. Contrast the models in terms of accuracy and reliability of forecasts through statistical indicators and confidence intervals, for their validation.

11. Suggest the incorporation of new data and variables, as well as the exploration of other models, to enrich the analysis and improve the accuracy of predictions.

2. Methodology

This quantitative study examines the influence of visits to the Civil War shelters in Alicante on memorial cultural heritage. Using a retrospective longitudinal design, it analyzes data from August 2023 to January 2024, focusing on two groups of visitors: the general public and school groups. Participants total 406 individuals, both local and international, who have made reservations through electronic and telephone means.

The data collection methodology was based on a visitation record, which includes information about the date, type, and origin of visitors, and their knowledge of the shelters. This information was systematized by the managing company, providing a representative sample throughout different seasons, though not throughout the entire year.

Data have been analyzed quantitatively with SPSS Statistics v29.0.1.0, R Studio, and Python, applying statistical models including ARIMA, autoregressions, and autocorrelations, to detect patterns and predict trends. Models were selected based on their complexity and significance.

The study complies with ethical principles, ensuring informed consent and data confidentiality. However, it suffers from temporal limitations by not covering a full year, which may affect the generalization of the results.

3. Results

First, descriptive and statistical analyses were performed to gain as comprehensive a view as possible of the niche of work where the study was to be conducted, which for lack of space we do not include in depth in this publication.

Subsequently, in the analyses, three statistical models were used indiscriminately: ARIMA(1, 0, 1), ARIMA(0, 0, 0) with Logarithmic Transformation, and AR(6), evaluating their capacity to model and predict trends in visits to the shelters. The selection of the optimal model was based on simplicity, significance of coefficients, and the capacity to capture the temporal dynamics of visits. The ARIMA(0, 0, 0) model with Logarithmic Transformation was identified as the most suitable for the analyzed time series, providing an optimal balance between simplicity and forecast accuracy. However, the utility of the AR(6) model to capture significant short-term dependencies is recognized. The choice of model considered the specific nature of the data and the objectives of the analysis.

1.1.   PREDICTION OF VISITOR INFLUX THROUGH ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE) AND AUTOREGRESSIVE MODELS.

1.1.1. PREDICTION BY DATE OF VISIT THROUGH THE ARIMA MODEL

We conducted an analysis of the autocorrelations of the time series corresponding to the "Date of the visit". Autocorrelations measure the linear relationship between current values of a time series and its previous values (lags). High autocorrelation indicates that past values have a strong influence on current values.

The time series exhibits significant autocorrelations for the first 16 lags, indicating strong temporal persistence; that is, visits are highly dependent on their previous values. The values of autocorrelation decrease as the number of lags increases but remain significant ($p < 0.001$), which is common in time series data.

Indeed, the autocorrelation values are high for the initial lags, starting at 0.737 for the first lag and gradually decreasing to 0.631 by the sixteenth lag. This suggests that visits on a given day are strongly influenced by visits on the preceding days.

Table 1. Autocorrelations

Series: Date of the visit

Lag	Autocorrelation	Standard Error[a]	Ljung-Box Statistics		
			Value	gl	Sig.[b]
1	,737	,050	222,925	1	<,001
2	,729	,072	441,251	2	<,001
3	,722	,088	655,875	3	<,001
4	,714	,101	866,207	4	<,001
5	,706	,113	1072,479	5	<,001
6	,699	,123	1275,550	6	<,001
7	,693	,133	1475,332	7	,000
8	,686	,141	1671,845	8	,000
9	,680	,149	1865,066	9	,000
10	,672	,157	2054,393	10	,000
11	,664	,164	2239,816	11	,000
12	,656	,170	2421,144	12	,000
13	,648	,176	2598,367	13	,000
14	,641	,182	2772,341	14	,000
15	,636	,188	2943,981	15	,000
16	,631	,193	3113,288	16	,000

a. The underlying process assumed is MA with the order equal to the number of lag minus one. Bartlett's approximation is used. b. Based on the asymptotic chi-square approximation

b. The underlying process assumed is MA with the order equal to the number of lag minus one. Bartlett's approximation is used. b. Based on the asymptotic chi-square approximation

The fact that autocorrelation values decrease slowly and remain positive for all considered lags indicates persistent time series behavior. This means visits tend to follow a "memory" of their past behavior over time.

The Ljung-Box statistic values are very high and significant (less than 0.001), indicating the autocorrelations for each of the lags are not zero and, therefore, there is a significant temporal dependence structure in the series of visits.

Partial autocorrelations, which show the correlation between two time points with the influence of the intervening points removed, decrease more rapidly than simple autocorrelations. This may evidence that

the direct effect of previous values fades more quickly than the total effect.

Indeed, the partial autocorrelation is high for the first lag and then decreases rapidly, becoming insignificant from the thirteenth lag onward. This is characteristic of an autoregressive (AR) process, where previous values affect future values up to a certain point, and then the influence stabilizes.

The high autocorrelation in the initial lags means that recent values in the series have a significant influence on future values.

The significance of partial autocorrelations for the first lag implies that an AR(1) model could be appropriate for the data. However, the partial autocorrelations do not fade completely to zero at higher lags, indicating a more complex AR model or the presence of other dynamics in the series.

Table 2. Partial Autocorrelations

Series: Date of the visit		
Lag	Partial Autocorrelation	Standard Error
1	,737	,050
2	,406	,050
3	,271	,050
4	,192	,050
5	,142	,050
6	,111	,050
7	,087	,050
8	,069	,050
9	,054	,050
10	,039	,050
11	,028	,050
12	,018	,050
13	,010	,050
14	,010	,050
15	,014	,050
16	,015	,050

In the following figures, autocorrelation and partial autocorrelation are visually represented at different lags, illustrating the strong temporal persistence and the direct effects of past values on future values of the time series data.

Figure 1. Autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs based on the provided values.

Based on the provided data from the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), some possible conclusions can be drawn:

- The ACF value at lag 1 is very high (0.737), indicating a strong positive correlation between each observation and the next in the series. This suggests the data may be non-stationary, as there is persistence of values over time.

- ACF values show a gradual decrease as lags increase but remain significant up to lag 16 (all significant at p < 0.001). This slow decay evidences a possible long-memory process or that the series could be integrated of a certain order (I(d)), meaning it may be necessary to differentiate the data d times to achieve stationarity.

- The PACF is also significant at lag 1 and decreases afterward. This may indicate an AR(1) process, where the current value is significantly influenced by the immediate previous value.

- The PACF after lag 1 shows non-significant values (except for lags 15 and 16, which have small but significant autocorrelations), meaning it may not be necessary to include additional AR terms.

- Patterns in ACF and PACF are often used for model identification in time series analysis. In this case, ACF and PACF suggest an ARIMA model could be appropriate. Specifically, given the slow decay of ACF, an ARIMA model with differentiation might be needed to account for non-stationarity.

In summary, the data likely require some form of differentiation to achieve stationarity, and an AR(1) process could be a good starting point for modeling. However, further analysis, such as unit root tests (e.g., augmented Dickey-Fuller test), is needed to confirm non-stationarity and the required order of differentiation, and model fitting with diagnostic checking would be needed to select the final model.

Given the prior analysis of autocorrelations, a reasonable starting point for the ARIMA model could be $p = 1, d = 1$, and $q = 1$, adjusting as necessary based on model diagnostics.

We will use an augmented Dickey-Fuller test to assess the time series' stationarity and determine if it's necessary to differentiate the series to make it stationary.

Table 3. Dickey-Fuller Test

Augmented Dickey-Fuller test for the time series of the number of visitors

Test Statistic	-7.687
p-value	1.45×10^{-11}
Used Lag	0
Number of Observations Used	95
Critical Values	Para un nivel de confianza del 1%, -3.501; para el 5%, -2.892; y para el 10%, -2.583.

Since the p-value is significantly less than 0.05, and the test statistic is lower than the critical values for common confidence levels (1%, 5%, and 10%), we reject the null hypothesis that the series has a unit root and conclude the series is stationary. This means it's not necessary to differentiate the series to achieve stationarity, suggesting $d = 0$ might be appropriate for the ARIMA model.

Now, we will construct an ARIMA model with the initial parameters suggested by the prior autocorrelation analysis, adjusting $d = 0$ based on the outcome of the Dickey-Fuller test.

Table 4. The adjusted ARIMA(1, 0, 1) model for the time series of the number of visitors

Coefficients	Comments
const: 4.2312 (p-value < 0.001),	indicating the constant term is significantly different from zero.
ar.L1: 0.2777 (p-value = 0.585)	showing the autoregressive coefficient is not significantly different from zero at the standard confidence level.
ma.L1: -0.0638 (p-value = 0.911)	indicating the moving average coefficient is also not significantly different from zero.
The Log Likelihood es -250.581	Additional information.
The information criteria AIC and BIC are 509.162 and 519.420, respectively	Additional information.
The Ljung-Box test for autocorrelated residuals shows a p-value of 0.99	indicating there is no evidence of autocorrelated residuals in the model..
The Jarque-Bera test yields a p-value of 0.00	suggesting the residuals do not follow a normal distribution, which could be an area for investigation and improvement.

The evaluation of an ARIMA(1, 0, 1) model applied to time series data shows that, although the AR and MA terms are not statistically significant, the absence of autocorrelated residuals indicates the temporal structure is well represented. However, the lack of normality in the residuals could signal omissions in capturing the data's dynamics. The implementation of logarithmic transformations improved the log-likelihood and information criteria (AIC and BIC), though the AR and MA terms remain non-significant. This suggests a more simplified model might be appropriate.

The improvement in the normality of the residuals following transformation is evidenced in the Jarque-Bera test, with skewness and kurtosis closer to those of a normal distribution, and a p-value that does not reject the hypothesis of normality. Despite the improvement in the distribution of residuals, the insignificance of the AR and MA coefficients points towards considering a more simple model, like an ARIMA(0, 0, 0) with a constant, which would assume a constant mean in the series.

Before adjusting a simpler model, the seasonality in the transformed series should be examined through visual analysis. If no seasonal patterns are detected, a simplified non-seasonal model could proceed.

Preliminary visual analysis reveals no evident seasonality, which could justify the use of a simplified non-seasonal model.

Figure 2. Graph on a time series transformed via logarithm

Given the lack of clear visual evidence of seasonality and the nonsignificant coefficients in the ARIMA(1, 0, 1) model for the transformed series, exploring a simpler model seems reasonable. We will adjust an ARIMA(0, 0, 0) model with a constant to the transformed series as a starting point for comparing model complexity. This approach will allow us to assess whether a model that assumes a constant mean is sufficient to capture the dynamics of the time series, based on criteria such as the AIC, BIC, and the significance of the model parameters, as well as the normality of the residuals. The ARIMA(0, 0, 0) model fitted to the time series transformed through logarithm (essentially modeling the series as a constant mean) provides the following results: the Log Likelihood is -109.429, with an AIC of 222.858 and a BIC of 227.987, which is slightly worse than the ARIMA(1, 0, 1) model in terms of AIC and BIC, but the difference is minimal. The coefficient for the constant is significant (p-value < 0.000), indicating that a constant mean is a significant component of the model. The Jarque-Bera Test for Residuals: Skewness (Skew): 0.069, indicating a symmetric distribution of residuals. Kurtosis: -0.708, suggesting a less pronounced distribution of residuals compared to a normal distribution. The Jarque-Bera test now shows a p-value of 0.33, indicating that there is not enough evidence to reject the hypothesis of normality in the residuals. The fit of the ARIMA(0, 0, 0) model to the transformed series suggests that a simple model, which considers a constant mean, is sufficiently good for modeling this time series, based on the model selection criteria and the normality of the residuals. This suggests that additional complexity through AR or MA terms may not be necessary for this particular time series. The absence of significant seasonal terms, both in visual inspection and in the fit of the simple model, suggests that there is no clear seasonality in the data that requires a SARIMA model. We have used the ARIMA(0, 0, 0) Model with a Logarithmic Transformation for the next 5 periods, adjusted to the time series transformed through logarithm, and they are constants, with an approximate value of 3.199 on the logarithmic scale. When transforming these values back to the original scale, this suggests an expected number of visitors that reflects the constant mean modeled by the ARIMA(0, 0, 0).

Table 5. Visitors according to the ARIMA(0, 0, 0) Model with a Logarithmic Transformation

	Forecast	Lower CI	Upper CI
96	3,198985	0,726242	14,09105
97	3,198985	0,726242	14,09105
98	3,198985	0,726242	14,09105
99	3,198985	0,726242	14,09105
100	3,198985	0,726242	14,09105

The confidence intervals for these forecasts on the original scale range from approximately 0.726 to 14.091 visitors. These wide intervals reflect the inherent uncertainty in future forecasting based on a model of constant mean.

Figure 3. Graph on confidence intervals based on a model of constant mean

These forecasts and their confidence intervals provide an estimate of the number of visitors expected in the next 5 periods, based on the historical trend modeled and assuming a constant mean. The wide range of the confidence interval underscores the potential variability in the forecasts and highlights the importance of considering this uncertainty when planning or making decisions based on these forecasts.

1.1.2. PREDICTION THROUGH THE AUTOREGRESSIVE MODEL (AR) OF ORDER 6

Based on the information from partial autocorrelations that we have obtained, it seems that an autoregressive model (AR) of order 6 might also be suitable for modeling the time series of visits to the civil war shelters in Alicante. The significant partial autocorrelation up to lag 6 suggests that past values up to 6 periods back have a significant influence on the current value of the series.

An AR(6) model will attempt to predict the current value of the series as a linear combination of the previous six values. This type of model is suitable when the time series shows a clear dependency on its past values up to a certain point, as indicated by the significant partial autocorrelations up to lag 6.

Table 6. Coefficients according to the autoregressive model (AR) of order 6

	Coefficient
0	4,113207
1	0,237784
2	0,037871
3	-0,17917
4	0,061418
5	-0,16254
6	0,040825

We will proceed to construct and estimate an AR(6) model for the time series. This model will be adjusted to the data to better understand the temporal dynamics of visits to the shelters and to make future forecasts based on historical information.

Table 7. Coefficients according to the model (AR) 6

Coefficients	Comments
const: 4.1132 (p-value < 0.001)	indicating a baseline level of visitors.
ar.L1: 0.2378 (p-value = 0.024)	indicating that the most recent value has a positive influence on the current value of the series.
The other lags (L2, L3, L4, L5, L6) have coefficients that are not significantly different from zero at the standard confidence level	although L3 and L5 show some level of negative influence.
The Log Likelihood is -234.416.	Additional information
The AIC and BIC information criteria are 484.833 and 504.831, respectively	which can help compare this model with other potential models
The S.D. of the innovations is 3.273	providing a measure of the variability of the model's errors

The AR(6) model indicates there is a significant dependency on the most recent value of the time series to predict the current value. However, the significance of the other lags is limited, suggesting that the influence of the previous values beyond the immediate one may not be as strong as initially expected.

This graph (Figure 4) will show the magnitude and significance of each lag coefficient (L1 to L6), which will help visualize how each past value contributes to the current value of the time series.

Figure 4. Graph to visualize the coefficients of the autoregressive model (AR) of order 6 obtained in the last analysis.

Based on the partial autocorrelation, an AR model could be suitable for this time series. The last significant lag based on partial autocorrelation is lag 6, suggesting that an autoregressive (AR) model of order 6 might be suitable for the time series of visits to Civil War shelters in Alicante.

Table 8. Coefficients for the (AR) 6 model; time series of visits

	Real	Imaginary	Modulus	Frequency
AR.1	1.2714	-0.0000j	1.2714	-0.0000
AR.2	0.3843	-1.5299j	1.5775	-0.2112
AR.3	0.3843	+1.5299j	1.5775	0.2112
AR.4	-1.0748	-0.9190j	1.4141	-0.3874
AR.5	-1.0748	+0.9190j	1.4141	0.3874
AR.6	3.8718	-0.0000j	.8718	-0.5000

The unit root test (ADF Test) indicates that the time series of visits is stationary, as the p-value is less than 0.05 (p-value = 1.45e-11), allowing us to reject the null hypothesis of a unit root.

The model selection process based on the AIC criterion suggests that the best autoregressive model for the data on visits to the Civil War shelters in Alicante is also an AR(6), which is the same as initially identified with the partial autocorrelations.

The constant (intercept) and the first lag (L1) are significantly different from zero, indicating that they have a statistically significant influence on the model. The other lags are not statistically significant at the 95% confidence level, although lag 3 is close to the significance threshold (p-value = 0.095).

The intercept of 4.1132 suggests that, in the absence of previous visits (i.e., when all lags are zero), the model predicts a base number of approximately 4 visitors. This value is statistically significant, as indicated by the p-value less than 0.05.

Table 9. Coefficients for the (AR) 6 model

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	4.1132	1.064	3.865	0.000	2.027	6.199
Number of Visitors.L1	0.2378	0.106	2.249	0.024	0.031	0.445
Number of Visitors.L2	0.0379	0.107	0.352	0.725	-0.173	0.249
Number of Visitors.L3	-0.1792	0.107	-1.668	0.095	-0.390	0.031
Number of Visitors.L4	0.0614	0.107	0.573	0.567	-0.149	0.271
Number of Visitors.L5	-0.1625	0.107	-1.514	0.130	-0.373	0.048
Number of Visitors.L6	0.0408	0.107	0.383	0.702	-0.168	0.250

The Coefficients of the Lags are as follows:

- Number of Visitors.L1: The coefficient of 0.2378 for the first lag indicates that visits on the previous day have a positive relationship with the current visits. For each additional visitor the previous day, we would expect to see an increase of approximately 0.2378 visitors on the current day. This effect is statistically significant.

- Number of Visitors.L2 to L6: The coefficients for lags 2 to 6 vary in magnitude and direction, but except for lag 3, none of them are statistically significant at the 95% level. This suggests that the influence of previous visits on current visits decreases or becomes less predictable after the first day.

The standard deviation of innovations (model errors) is 3.273, which gives us an idea of how much the actual observations vary around the model's predictions.

The AIC (4.822) and BIC (5.048) values are criteria used to compare models. In this context, they only provide an internal reference, as we are not comparing this model with others.

The roots of the characteristic polynomial indicate the stability of the model and the temporal dynamics of the series. All roots are real or pairs of complex conjugates with moduli greater than 1, suggesting that the model is stable. The pairs of complex conjugates imply oscillations in the time series, but the presence of dominant real roots suggests that these oscillations are not the main component of the series dynamics.

The AR(6) model suggests that visits to the Civil War shelters in Alicante are significantly influenced by the previous day's visits, but visits from more distant previous days have a lesser or uncertain effect on current visits. This might imply that promotional campaigns or special events would have a more immediate impact on visits that would dissipate relatively quickly.

However, the limited significance of lags beyond the first suggests that other factors not captured by this model might be influencing visits, warranting further investigation that could include external variables or considering different types of models.

1.1.3.   FUTURE FORECASTS

The AR(6) model can be used to make future forecasts, taking into account the significant influence of the first lag and the model's stability. Making forecasts will involve using the last 6 observed values to predict future values, iterating the process for each step forward in time that we wish to forecast.

The results seem to be consistent and provide a solid basis for decision-making and planning based on the forecasts generated by the model. However, as always, it is prudent to consider the inclusion of additional data or the exploration of other models to validate these findings and improve the accuracy of the forecasts.

To make future forecasts using the AR(6) model, we will specify the number of future periods we wish to forecast. Let's assume, for example, that we want to make forecasts for the next 5 future periods. We

will use the AR(6) model we have previously adjusted to generate these forecasts.

These files include the historical data of visits to the Civil War shelters in Alicante and the future forecasts generated by the AR(6) model.

Table 10. Date of the visits and number of visitors at each of them.

Date of Visit	Number of Visitors
2023-08-03	5
2023-08-18	3
2023-08-19	6
2023-08-24	4
2023-08-25	4
2023-08-26	2
2023-08-27	4
2023-08-28	2
2023-08-30	3
2023-09-02	8
2023-09-03	4
2023-09-05	3
2023-09-06	1
2023-09-09	9
2023-09-10	4
2023-09-13	3
2023-09-16	10
2023-09-17	7
2023-09-21	9
2023-09-26	3
2023-09-27	3
2023-10-03	2
2023-10-04	3
2023-10-06	2
2023-10-07	3
2023-10-08	4
2023-10-11	3
2023-10-13	1

2023-10-15	6
2023-10-20	8
2023-10-21	4
2023-10-24	3
2023-10-25	1
2023-10-25	1
2023-10-26	2
2023-10-27	14
2023-10-29	4
2023-11-01	3
2023-11-02	1
2023-11-02	3
2023-11-03	1
2023-11-04	14
2023-11-07	5
2023-11-09	4
2023-11-10	2
2023-11-11	9
2023-11-12	5
2023-11-15	4
2023-11-17	1
2023-11-18	3
2023-11-19	2
2023-11-21	3
2023-11-22	1
2023-11-24	2
2023-11-25	10
2023-11-26	1
2023-11-27	2

2023-11-28	6
2023-11-29	6
2023-11-30	5
2023-12-01	5
2023-12-02	8
2023-12-03	5
2023-12-04	2
2023-12-05	6
2023-12-06	3
2023-12-07	10
2023-12-08	8
2023-12-09	18
2023-12-10	3
2023-12-12	1
2023-12-12	2
2023-12-13	2
2023-12-15	4
2023-12-16	2
2023-12-17	6
2023-12-18	1
2023-12-19	2
2023-12-20	1
2023-12-21	3
2023-12-22	1
2023-12-26	7
2023-12-27	9
2023-12-28	14
2023-12-30	10
2024-01-03	5

2024-01-20	3
2024-01-21	1
2024-01-22	1
2024-01-24	2
2024-01-25	2
2024-01-26	2
2024-01-27	2
2024-01-28	3
2024-01-29	1
2024-01-30	1

Let's proceed to make and visualize these forecasts:

Table 11. Forecasts by Date and Number of Visitors

	0
2024-01-31	3,730753
2024-02-01	4,799849
2024-02-02	4,872107
2024-02-03	4,806394
2024-02-04	4,688019

The chart (Figure 5) displays future forecasts for visits to the Civil War shelters in Alicante, using the autoregressive model of order 6 (AR(6)). Historical data are presented in green, while future forecasts are shown in red.

Figure 5. Future Forecast Chart for Visits, Using the Autoregressive Model of Order 6 (AR(6))

The AR(6) model projects future visits by focusing on immediate autocorrelation and model stability. Useful for resource management, it allows for demand forecasting and facilitates decision-making. However, inherent uncertainty and confidence limits should be considered, and forecasts could be refined with more data or alternative models.

1.2.   *CHOOSING A MODEL: COMPARISONS:*

As verified, we have used various models to analyze future forecasts, the forecasts on visits to the shelters in the city of Alicante.

The ARIMA(1, 0, 1) aims to capture autoregression and moving average in the time series, indicating complex temporal relationships. However, the non-significance of its AR and MA coefficients questions its necessity. Alternatively, the ARIMA(0, 0, 0) with logarithmic transformation represents the series through a constant mean, suitable for stationary series as confirmed by the Dickey-Fuller test, and seeks to stabilize the variance. This simple model does not show significant residual autocorrelations, suggesting an effective capture of the temporal dependency structure without additional components. It provides forecasts based on the stability of the mean, albeit with some uncertainty reflected in the confidence intervals.

The choice between these two models depends on the desired balance between simplicity and the ability to capture complexities in the data. The ARIMA(0, 0, 0) with logarithmic transformation appears to be sufficient and more parsimonious for modeling the given time series, reflecting the "less is more" philosophy in statistical modeling.

The significance of the constant in the ARIMA(0, 0, 0) model suggests that, for this particular time series, the additional complexity of AR or MA terms may not be necessary.

The adequacy of the model should be evaluated not only in terms of statistical fit but also in its ability to produce accurate and useful forecasts. The simplicity of the ARIMA(0, 0, 0) model, along with the normality of residuals, makes it preferable for interpretation and practical application in this case.

The ARIMA(1, 0, 1), ARIMA(0, 0, 0) with Logarithmic Transformation, and AR(6) models differ in complexity and approach for analyzing visits to the Civil War shelters in Alicante. ARIMA(1, 0, 1) seeks to capture short-term dependencies, while ARIMA(0, 0, 0) simplifies the series to a constant mean, improving the normality of residuals with logarithmic transformation. AR(6), an autoregressive model, predicts current values using information up to six periods prior, confirmed its stationarity with the unit root test. The significant constant in ARIMA(0, 0, 0) suggests that a constant mean adequately models the series. In AR(6), the importance of the first lag emphasizes the impact of the most recent value. Finally, both ARIMA models indicate a good fit by not presenting significant residual autocorrelations, while AR(6) might better capture temporal dynamics by focusing on autoregression.

Figure 6. Comparison of ARIMA and AR(6) Models

The selection between ARIMA(0, 0, 0) with logarithmic transformation and AR(6) is dictated by the complexity of the series and the purpose of the analysis. ARIMA(0, 0, 0) simplifies modeling to a constant mean, while AR(6) leverages recent temporal dependencies. The ARIMA(1, 0, 1) is found inadequate due to the insignificance of its coefficients. AR(6) is preferable for recognizing short-term autoregressive patterns, though the simplicity of ARIMA(0, 0, 0) may be beneficial where ease of

interpretation is a priority. The decision is based on the balance between simplicity and accuracy, adjusting to the specificity of the time series.

1.2.1. EVALUATING FORECAST RESULTS BETWEEN THE ARIMA(0, 0, 0) MODEL WITH LOGARITHMIC TRANSFORMATION AND THE AR(6) MODEL

The comparison between the ARIMA(0, 0, 0) model with Logarithmic Transformation and AR(6) reveals that the former, due to its simplicity and assumption of a constant mean, is preferable for stable series, although its utility is limited by uncertainty reflected in wide confidence intervals. On the other hand, AR(6), leveraging data from six previous periods, is superior in capturing dynamics and temporal trends, ideal for series with significant recent autocorrelations and variations. The choice between the two depends on the nature of the time series and the balance between simplicity and predictive accuracy.

Figure 7. Chart evaluating and comparing, through a simulation, the forecast results between the ARIMA(0, 0, 0) model with Logarithmic Transformation and the AR(6) model

Figure 8. Diagram offering a comparative view of the ARIMA(0, 0, 0) model with Logarithmic Transformation and AR(6) in terms of forecast accuracy and reliability

The selection of forecasting models, AR(6) or ARIMA(0, 0, 0) with logarithmic transformation, depends on the dynamics of the time series and the analytical purpose. AR(6) is optimal for short-term dependencies, while ARIMA(0, 0, 0) is better for stable series. Predictive accuracy and uncertainty should be evaluated through confidence intervals and analysis of historical variability.

Figure 9. Flowcharts illustrating the evaluation of forecast results between the ARIMA(0, 0, 0) model with Logarithmic Transformation and the AR(6) model

4. Discussion

Significant analyses of time series autocorrelations and related topics have been conducted. There is a study on the disturbances of a time series and the removal of trend through regressions over time or as a function of time. Various simulation processes using the Monte Carlo procedure have been carried out to study the presence of autocorrelations in the analysis of residuals and the most important statistics (Núñez, 1986). It is also used in medicine, for example, for causes of poor metabolism (Agámez Pertuz et al., 2006; Ertuğrul et al., 2022).

The analysis of time series autocorrelations has also been used in the context of tourism. Thus, cultural tourism as an emerging product is addressed, analyzing changes in Western society and its impact on heritage and culture. Reflections on new concepts of heritage are offered, and changes in the definition of cultural tourism, as well as strategies for transforming heritage resources into tourist products, are analyzed (Ibarra, 2023). Another study provides an overview of the main characteristics of people engaging in rural tourism in Spain, based on the results of a study conducted throughout the year 1994 (García, 2023). This work discusses the need for reliable and coherent statistics on tourism and its

interdependence with other economic and social sectors, highlighting the importance of designing a statistical system that represents the reality of tourism (Amorim, 2016).

The ARIMA (AutoRegressive Integrated Moving Average) model is a statistical model used to analyze and predict time series. It has been used, for example, to predict the flow of the Amprong River, emphasizing the importance of accurate prediction for water management in agriculture (Rahayu et al., 2020). To predict the prices of rice grain, demonstrating its utility in the agricultural sector (Ramadhani et al., 2020). For forecasting time series (Faujdar & Joshi, 1 C.E.). The ARIMA model has also been used to project the yield of Chinese potatoes, showing its applicability in the prediction of agricultural yields (Cheng-Zhi et al., 2016). It was used to predict the trajectory of the COVID-19 pandemic in the 15 most affected countries (Singh et al., 2020).

Regarding the use of the ARIMA model in the tourism sector, one study selected five different data sets on airlines, hotels, car rentals, and travel agencies in the U.S. tourism industry and used ETS and ARIMA models to predict data from 2000 to 2020 (Lin, 2023). Another focuses on the demand for whale watching tourism in Ulsan, using the seasonal ARIMA model for forecasts (신승훈 et al., 2022). A study applies the ARIMA model to predict the demand for health tourism in Turkey (Yilmaz, 2022). Another investigates how rural tourism can promote high-quality development in Guangshan County region, using the ARIMA model (Ma, 2022). Based on time series data of the number of domestic tourists in Hunan Province from 2000 to 2019, this study constructs an ARIMA model to predict the number of tourists in the next four years (Qin, 2021). The ARIMA model has also been used in the context of tourism in Spain. Thus, one study improves the forecasting of tourist flows to Spain using Google search indexes related to travel to Spain. Two models are compared for Germany, the United Kingdom, and France: a conventional ARIMA model and a model augmented with the Google index (Artola et al., 2015). Furthermore, the possibility of improving the predictive capacity of a tourism demand model with meteorological variables has been investigated, using as a case study the monthly British tourist demand towards the Balearic Islands (Spain). The results are compared with those obtained by non-causal methods such as an ARIMA model ((Álvarez-Díaz M & Rosselló-Nadal, 2008; Álvarez-Díaz & Rosselló-Nadal, 2010)

5. Conclusion

The research on predicting visitor influx to the memorial cultural heritage of Alicante using ARIMA and Autoregressive models has revealed how fluctuations in the number of visitors can affect the management and conservation of Civil War shelters, identifying seasonal patterns and preferences among different types of visitors. This analysis has underscored the importance of resource planning tailored to temporal variations and the need to customize communication and education strategies for different visitor groups. Additionally, the geographically diverse interest in the shelters suggests potentials for promotion and collaboration at various levels. Among the statistical models evaluated, the ARIMA(0, 0, 0) with Logarithmic Transformation was identified as the most effective for modeling the time series of visits, though the AR(6) proved to be crucial for capturing short-term dynamics,

highlighting the complexity of the time series. Despite the utility of the forecasts generated for future planning, the presence of uncertainty in predictions underscores the need for flexible management approaches. Therefore, the inclusion of more variables and the exploration of new models to refine the understanding and accuracy of predictions are recommended. The findings reinforce the relevance of adopting data-based strategies in cultural heritage management, pointing towards adaptability and anticipation as keys to its effective preservation. Thus, this study contributes to the field of cultural heritage management, offering a replicable methodological approach and highlighting the critical importance of academic research in the promotion and conservation of cultural heritage.

Declarations

Conflict of Interest Statement

The authors declare no conflicts of interest.

Ethics Statement

Surveys conducted in this study were anonymous, and informed consent was obtained from all participants prior to their participation.

Funding Statement

Author Contribution

Author Contribution StatementDr. Pablo Rosser:- Conceptualization: Developed the original concept and research framework for the study on predicting visitor flow to Civil War shelters in Alicante using ARIMA and autoregressive models.- Methodology: Designed the study methodology, including the selection and application of ARIMA and AR models, and oversaw the data collection process.- Data Analysis: Conducted the primary data analysis using statistical tools such as SPSS Statistics v29.0.1.0, R Studio, and Python.- Supervision: Supervised the overall research project, ensuring the integrity and accuracy of the research findings.- Writing - Original Draft: Wrote the initial draft of the manuscript, including the introduction, methodology, results, and discussion sections.- Visualization: Created the graphs and charts presented in the manuscript to illustrate the statistical findings.Dr. Seila Soler:- Literature Review: Conducted a comprehensive literature review to support the study's theoretical framework and context.- Data Collection: Assisted in the organization and systematization of the visitation records used in the study.- Statistical Analysis: Collaborated in the statistical analysis and interpretation of the results, particularly in the application of unit root tests and the evaluation of model parameters.- Writing - Review & Editing: Contributed to the critical revision of the manuscript, enhancing the clarity and quality of the

final document.- Project Administration: Managed the logistical aspects of the research project, including coordination between institutions and communication with stakeholders.- Funding Acquisition: Secured funding and resources necessary for the completion of the research project.Both authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Data Availability

Sequence data that support the findings of this study have been deposited in zenodo with the primary accession code 10.5281/zenodo.11419481

References

1. Agámez Pertuz, Y. Y., Oviedo Aguiar, L. A., Uribe, U. N., Centeno, M. A., & Odriozola, J. A. (2006). Análisis de la microporosidad de catalizadores de fcc. *Academia Colombiana de Ciencias Exactas, Fisicas y Naturales. Revista*, *30*(115), 271–278. https://doi.org/10.18257/raccefyn.30(115).2006.2248

2. Álvarez-Díaz M & Rosselló-Nadal. (2008). Forecasting British Tourist Arrivals to Balearic Islands Using Meteorological Variables and Artificial Neural Networks. *Centre de Recerca Econòmica*, *2*.

3. Álvarez-Díaz, M., & Rosselló-Nadal, J. (2010). Forecasting British Tourist Arrivals in the Balearic Islands Using Meteorological Variables. *Tourism Economics*, *16*(1), 153–168. https://doi.org/10.5367/000000010790872079

4. Amorim, A. (2016). *Uso de indicadores químicos na avaliação da qualidade do Argissolo vermelho amarelo distrocoeso em um sistema de cultivo em aleias*.

5. Artola, C., Pinto, F., & Pablo, de P. G. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower*, *36*(1), 103–116. https://doi.org/10.1108/IJM-12-2014-0259

6. Azad, A. S., Sokkalingam, R., Daud, H., Adhikary, S. K., Khurshid, H., Mazlan, S. N. A., & Rabbani, M. B. A. (2022). Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study. *Sustainability: Science Practice and Policy*, *14*(3), 1843. https://doi.org/10.3390/su14031843

7. Bottomley, C., Ooko, M., Gasparrini, A., & Keogh, R. H. (2023). In praise of Prais-Winsten: An evaluation of methods used to account for autocorrelation in interrupted time series. *Statistics in Medicine*, *42*(8), 1277–1288. https://doi.org/10.1002/sim.9669

8. Cheng-Zhi, C., Hong-Lan, M., & Ying, L. (2016). Chinese potato yield projected on ARIMA (Autoregressive Integrated Moving Average) model basis. *Research on Crops*, *17*(4), 769. https://doi.org/10.5958/2348-7542.2016.00130.3

9. Cicuéndez, V., Litago, J., Sánchez-Girón, V., Román-Cascón, C., Recuero, L., Saénz, C., Yagüe, C., & Palacios-Orueta, A. (2023). Dynamic relationships between gross primary production and energy

partitioning in three different ecosystems based on eddy covariance time series analysis. *Frontiers in Forests and Global Change*, 6. https://doi.org/10.3389/ffgc.2023.1017365

10. Ertuğrul, A., Anıl Yağcıoğlu, A. E., Ağaoğlu, E., Karakaşlı, A. A., Ak, S., Yazıcı, M. K., & Leon, J. de. (2022). Valproate, obesity and other causes of clozapine poor metabolism in the context of rapid titration may explain clozapine-induced myocarditis: A re-analysis of a Turkish case series. *Revista de Psiquiatria y Salud Mental*, *15*(4), 281–286. https://doi.org/10.1016/j.rpsm.2021.10.003

11. Faujdar, N., & Joshi, A. (1 C.E.). *Time Series Analysis for Crime Forecasting Using ARIMA (Autoregressive Integrated Moving Average) Model*. IGI Global. https://doi.org/10.4018/978-1-7998-2795-5.ch007

12. Gao, Y., Cheng, J., Meng, H., & Liu, Y. (2019). Measuring spatio-temporal autocorrelation in time series data of collective human mobility. *Geo-Spatial Information Science = Diqui Kongjian Xinxi Kexue Xuebao / Edited by Editorial Board of Geomatics and Information Science of Wuhan University*, *22*(3), 166–173. https://doi.org/10.1080/10095020.2019.1643609

13. García, R. F. (2023). Análisis de las principales características de la demanda de turismo rural en España. *Revista de Estudios Turísticos*. https://doi.org/10.61520/et.1271995.723

14. Ibarra, J. (2023). Análisis de la oferta de turismo cultural en España. *Revista de Estudios Turísticos*. https://doi.org/10.61520/et.1502001.877

15. Lin, S. (2023). Forecasting the trend of tourism industry in the United States: using ARIMA model and ETS model. *Highlights in Business, Economics and Management*, *10*, 111–121. https://doi.org/10.54097/hbem.v10i.7964

16. Ma, J. (2022). Research on rural tourism promoting regional high quality development based on ARIMA model: A case study of guangshan county, Xinyang city. *Proceedings of the 4th International Conference on Economic Management and Model Engineering*. The International Conference on Economic Management and Model Engineering, Nanjing, China. https://doi.org/10.5220/0012023400003620

17. Núñez, J. M. (1986). *Ausencia de estacionaridad en las perturbaciones de una serie temporal y su influencia en la regresión sobre el tiempo*. https://dialnet.unirioja.es/servlet/tesis?codigo=12218

18. Qin, Q. (2021). Forecasting tourism market demand in hunan province using arima model. *Delta: Jurnal Ilmiah Pendidikan Matematika*, *9*(2), 211–220. https://doi.org/10.31941/delta.v9i2.1410

19. Rahayu, W. S., Juwono, P. T., & Soetopo, W. (2020). Discharge prediction of Amprong river using the ARIMA (autoregressive integrated moving average) model. *IOP Conference Series: Earth and Environmental Science*, *437*(1), 012032. https://doi.org/10.1088/1755-1315/437/1/012032

20. Ramadhani, F., Sukiyono, K., & Suryanty, M. (2020). Forecasting of Paddy Grain and Rice's Price: An ARIMA (Autoregressive Integrated Moving Average) Model Application. *SOCA: Jurnal Sosial Ekonomi Pertanian*, *14*(2), 224–239. https://doi.org/10.24843/SOCA.2020.v14.i02.p04

21. Singh, R. K., Rani, M., Bhagavathula, A. S., Sah, R., Rodriguez-Morales, A. J., Kalita, H., Nanda, C., Sharma, S., Sharma, Y. D., Rabaan, A. A., Rahmani, J., & Kumar, P. (2020). Prediction of the COVID-19

Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model. *JMIR Public Health and Surveillance*, *6*(2), e19115. https://doi.org/10.2196/19115

22. Yilmaz, N. (2022). Turkey's health tourism demand forecast: The Arima model approach. *International Journal of Health Management and Tourism*. https://doi.org/10.31201/ijhmt.1065460

23. 김민경, 김현주, & 박종화. (2022). 계절ARIMA 모형을 이용한 국내 의료관광 수요예측에 관한 연구. 관광레저연구, *34*(9), 85–99. https://doi.org/10.31336/JTLR.2022.9.34.9.85
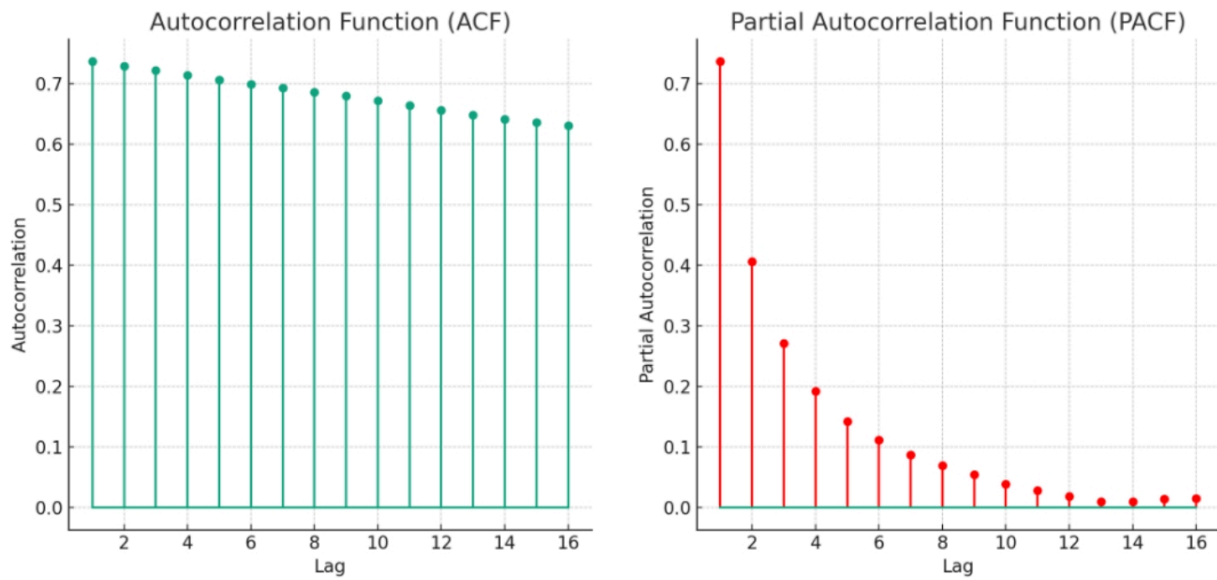
Figures

Figure 1

Autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs based on the provided values.
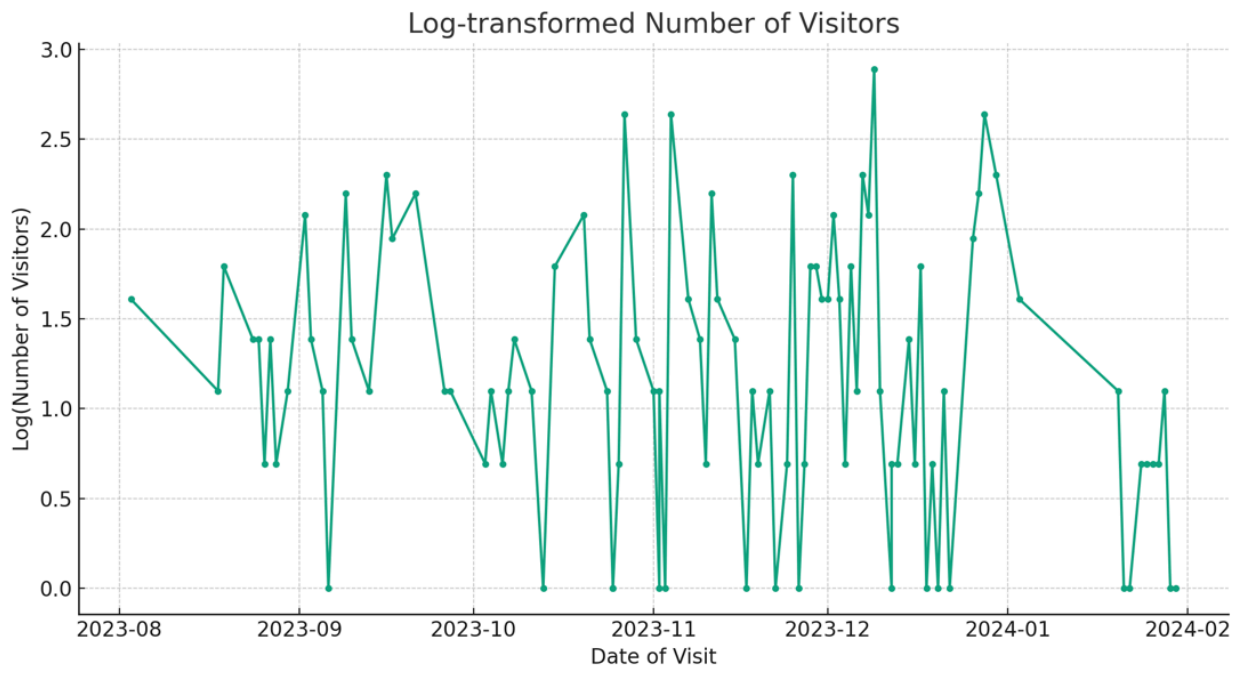
Figure 2

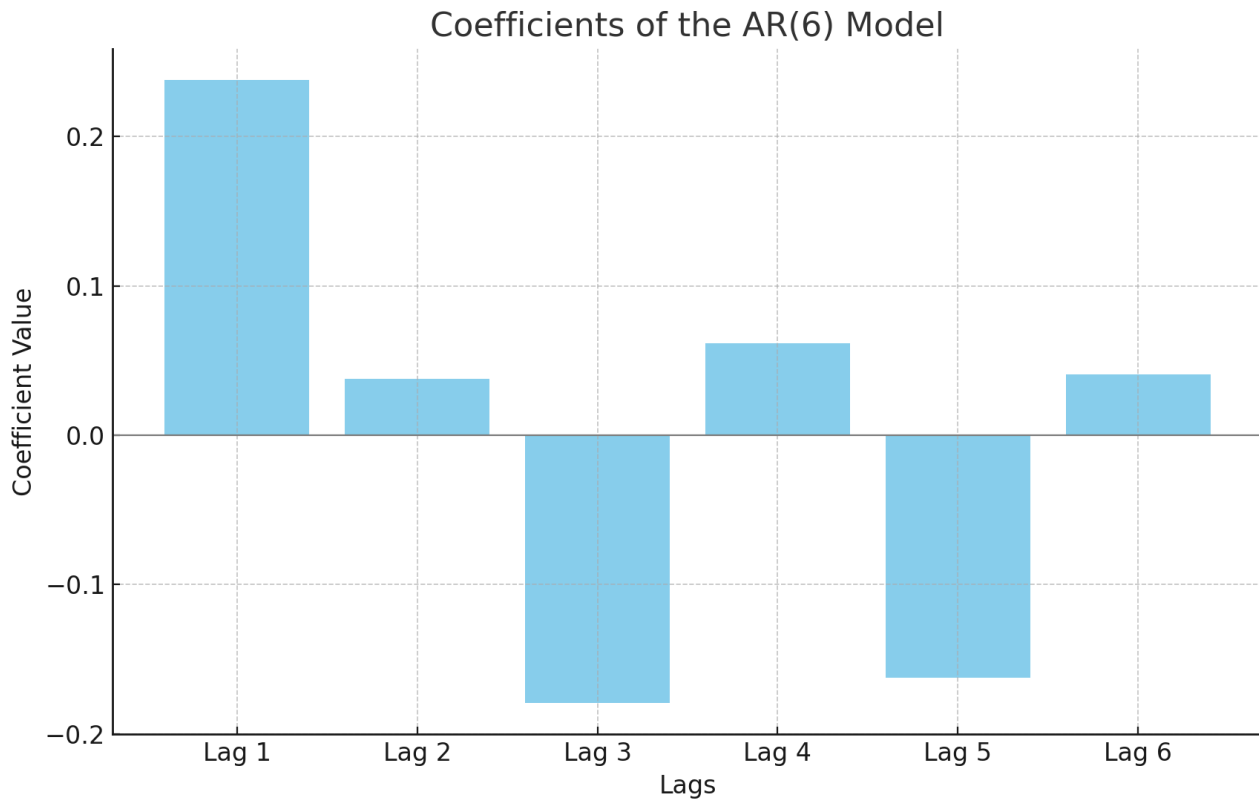Graph on a time series transformed via logarithm
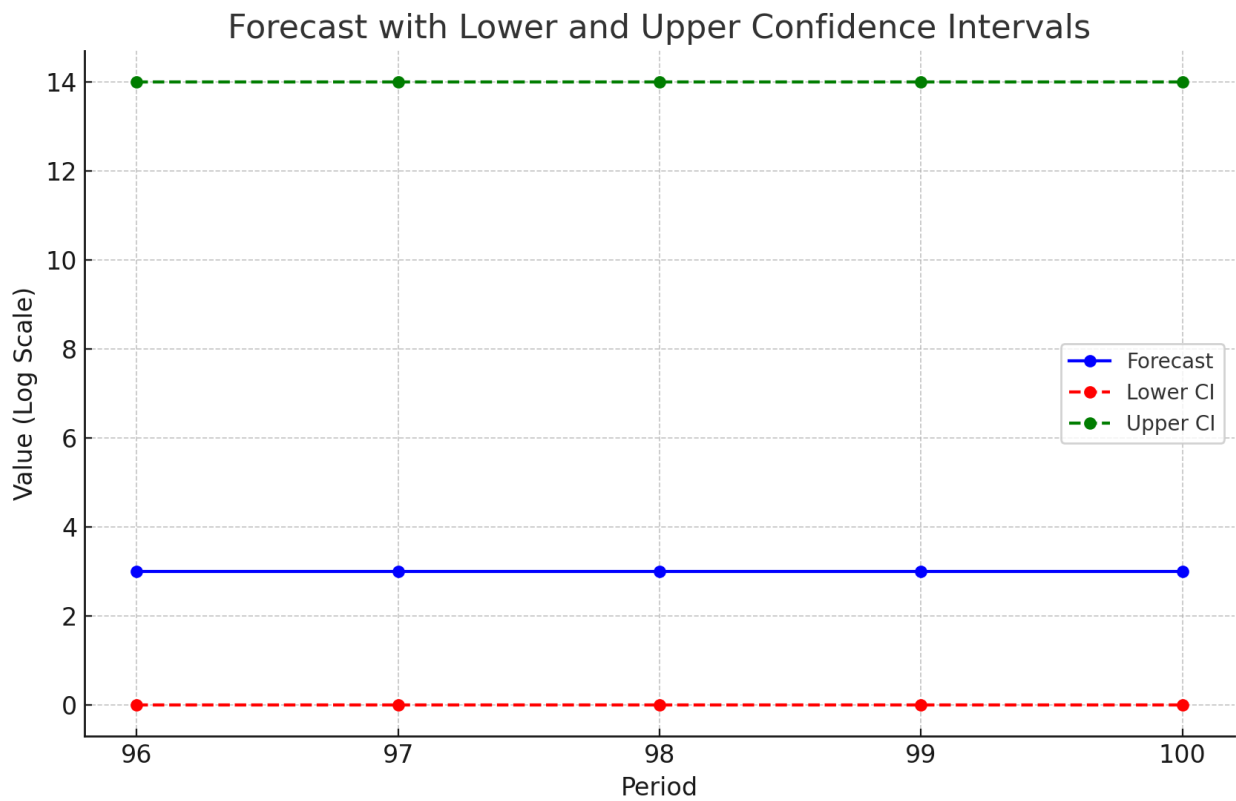
Figure 3

Graph on confidence intervals based on a model of constant mean

Figure 4

Graph to visualize the coefficients of the autoregressive model (AR) of order 6 obtained in the last analysis.
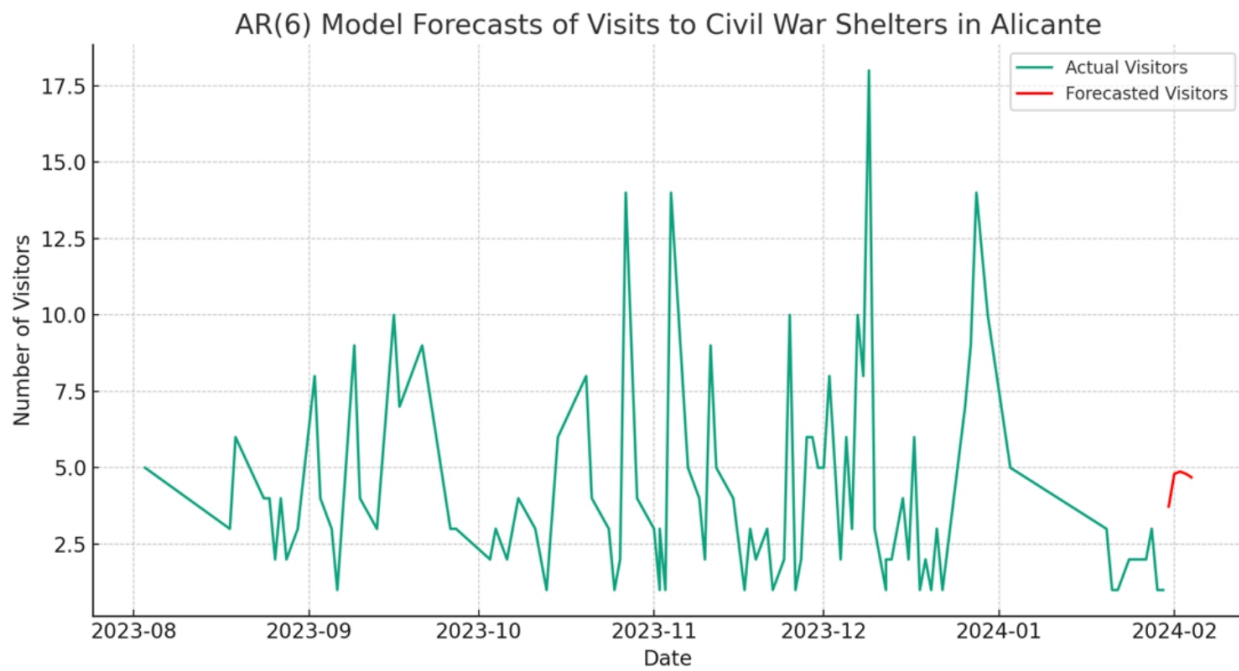
Figure 5
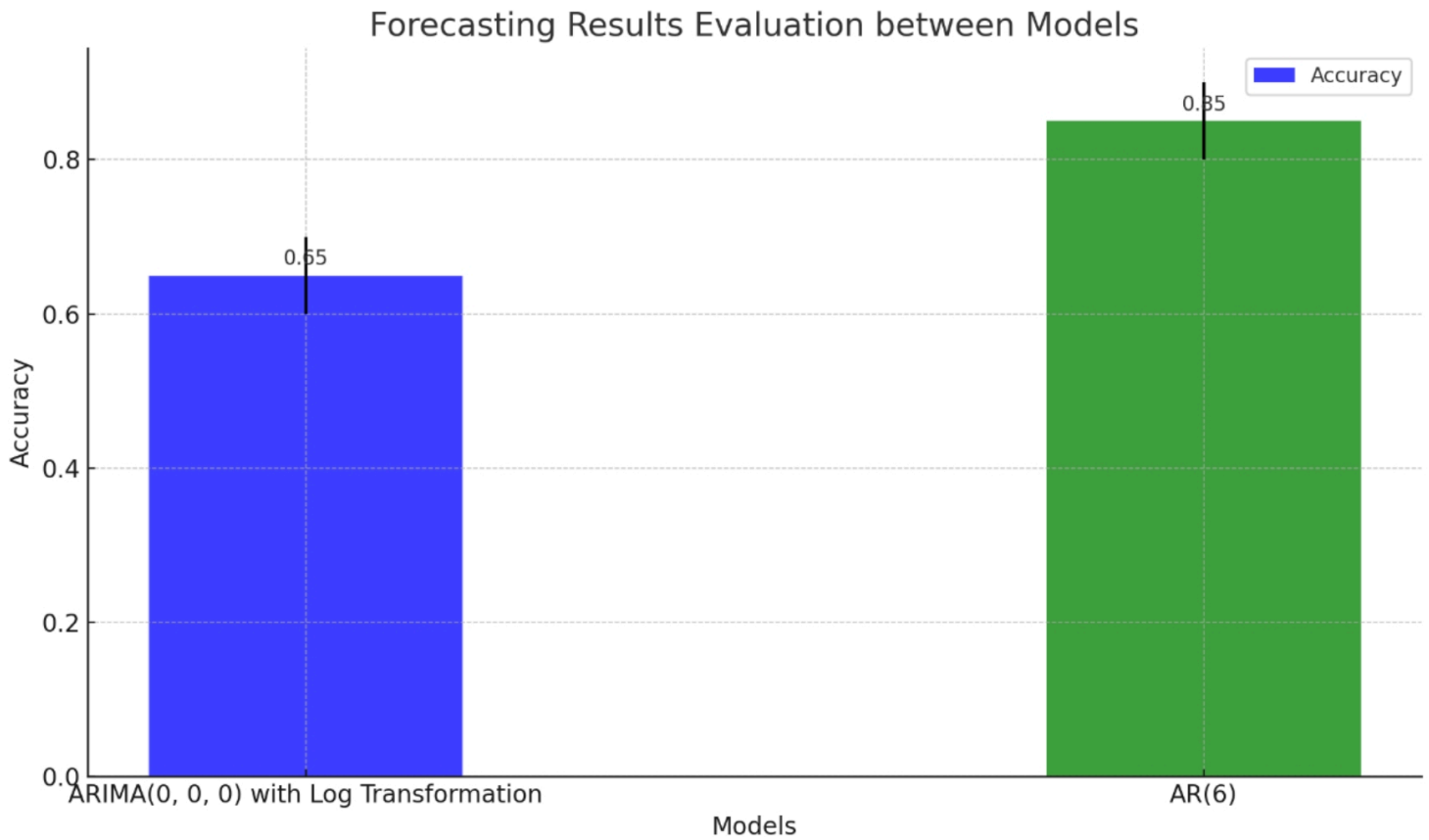
Future Forecast Chart for Visits, Using the Autoregressive Model of Order 6 (AR(6))

Figure 6

Comparison of ARIMA and AR(6) Models

Forecasting Results Evaluation between Models

Figure 7

Chart evaluating and comparing, through a simulation, the forecast results between the ARIMA(0, 0, 0) model with Logarithmic Transformation and the AR(6) model
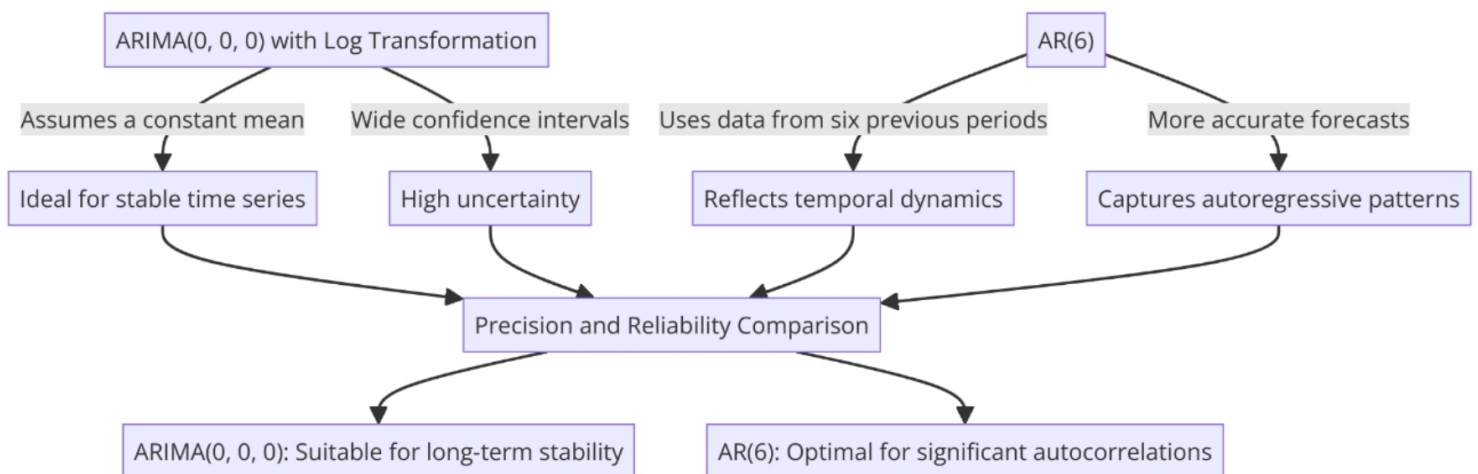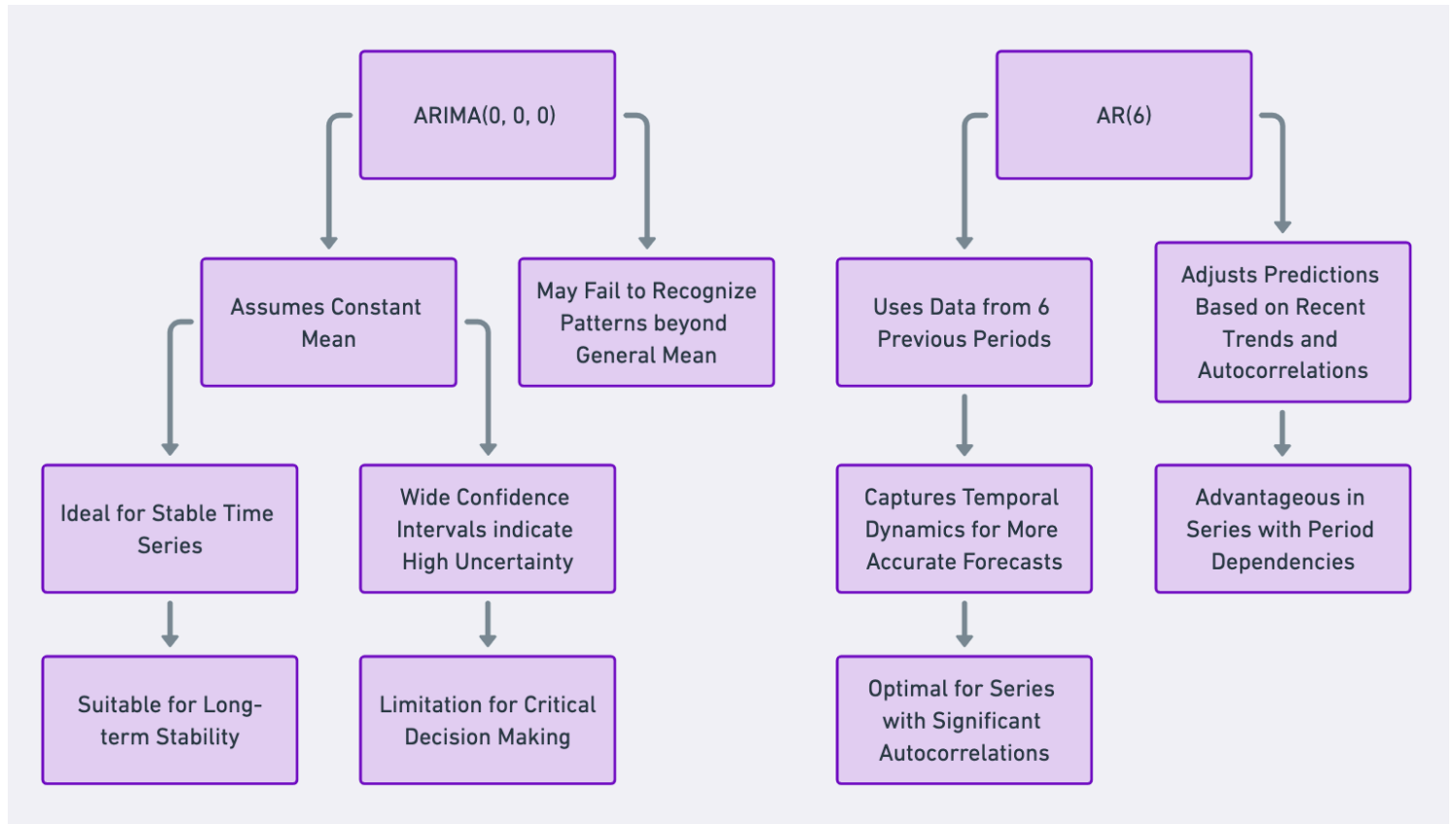

Figure 8

Diagram offering a comparative view of the ARIMA(0, 0, 0) model with Logarithmic Transformation and AR(6) in terms of forecast accuracy and reliability

Figure 9

Flowcharts illustrating the evaluation of forecast results between the ARIMA(0, 0, 0) model with Logarithmic Transformation and the AR(6) model