Air Passengers Time Series Forecasting using ARIMA

Manish Shukla
Plano, TX, USA
Independent researcher
manishshukla.ms18@gmail.com

August 27, 2025

Abstract

The number of people who fly on commercial aircraft has grown markedly since the middle of the twentieth century. Reliable forecasts of air passenger demand help airlines and regulators plan fleets, schedules and infrastructure. The *AirPassengers* data set — monthly counts of international airline passengers from 1949 to 1960 — is a classic benchmark for time series analysis because it exhibits both trend and seasonality [2]. In this report I revisit these data from a practical point of view. Inspired by a publicly available repository of code and plots8, I outline a reproducible workflow that involves exploratory graphics, variance-stabilising transformations, stationarity testing, decomposition of trend and seasonality, autocorrelation analysis and the fitting of a seasonal ARIMA model. I split the series into training and test samples, compare competing models using information criteria, and evaluate out-of-sample forecasts using root mean square error and mean absolute percentage error. The emphasis throughout is on clarity and brevity rather than exhaustiveness. The accompanying scripts and figures make it straightforward for readers to replicate the analysis and adapt the approach to other seasonal time series.

Keywords: Time series forecasting; Seasonal ARIMA (SARIMA); AirPassengers dataset; Box–Jenkins methodology; Stationarity testing; Model selection (AIC/BIC).

1 Introduction

Airline companies, regulators and travellers all rely on credible forecasts of passenger demand to inform decisions about capacity, scheduling and investment. When demand is underestimated, resources may be insufficient; when it is overestimated, assets lie idle. I approach the forecasting problem using a small but instructive data set: the monthly totals of international air travellers from 1949 to 1960 known as the *AirPassengers* data. Because the series is short and exhibits both a steadily increasing trend and pronounced annual seasonality, it provides a simple yet rich case study for seasonal time series models.

My aim in this report is to reproduce and explain a straightforward analytical pipeline for these data. Following the outline provided in a public repository8, I begin by inspecting the raw counts to understand their structure, apply a logarithmic transformation to stabilise the variance, and test for stationarity using the augmented Dickey–Fuller procedure. I then remove trend and seasonality through differencing, study the autocorrelation and partial autocorrelation functions, and fit a seasonal ARIMA model. The process is carried out in `Python` using open-source libraries. By splitting the series into training and test segments, I evaluate how well the selected model forecasts future months. Each step is explained in plain language and accompanied by figures, so that readers without specialised statistical backgrounds can follow along.

While many sophisticated forecasting methods exist, my goal here is to demonstrate that a classical approach suffices to capture the main features of the data and to produce reasonable predictions. I therefore focus on parsimony and transparency rather than on exhaustive comparisons. The remainder of the report is structured as follows. Section 3 summarises the data. Section 4 outlines the transformation, differencing and modeling techniques used. Section 6 presents the fitted model and its forecasting performance. Section **??** discusses the broader implications of the findings, and Section 7 concludes.

2 Literature Review

The analysis of time series data has its roots in the early twentieth century, when statisticians sought to understand economic cycles and physical phenomena. Among the earliest systematic treatments is the work of Slutsky (1937), who explored random summation of independent variables and its role in generating cycles. Another milestone is Yule's (1927) investigation of the autoregressive process to explain periodicities in sunspot numbers. These early works laid the foundation for the autoregressive moving average (ARMA) models that would be popularized by Box and Jenkins (1970).

Box and Jenkins introduced a systematic methodology for modeling time series. Their approach emphasised iterative steps: identification, estimation and diagnostic checking. In the identification stage, the analyst uses plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF) to hypothesize appropriate model orders for autoregressive (AR), moving average (MA) and integration (I) components. For non-stationary series, differencing is applied until stationarity is achieved. In the estimation stage, parameters are estimated via maximum likelihood or least squares. Diagnostic checking involves examining residuals for independence and homoscedasticity, often through Ljung–Box tests and further ACF/PACF plots. The Box–Jenkins method remains a cornerstone of time series analysis and underlies many software implementations.

When it comes to forecasting air passenger demand, various methods have been proposed in the literature. Classical approaches such as exponential smoothing (Brown, 1956; Holt, 1957; Winters, 1960) model trend and seasonality explicitly through smoothing parameters. These methods are easy to implement and often perform well for short-term forecasting. Holt–Winters triple exponential smoothing, for example, decomposes a series into level, trend and seasonal components and updates them recursively. In the presence of multiplicative seasonality—as in the AirPassengers data—multiplicative Holt–Winters smoothing can produce robust forecasts.

ARIMA models extend the ARMA framework by allowing for integration of non-stationary series. Seasonal ARIMA (SARIMA) models incorporate seasonal differencing and seasonal AR and MA terms to handle periodic behaviour. The ARIMA$(0,1,1)(0,1,1)_{12}$ model is known to fit the AirPassengers data reasonably well; Box and Jenkins (1970) used it as an illustrative example. Estimating seasonal ARIMA models requires careful selection of orders (p, d, q) for the non-seasonal part and (P, D, Q, s) for the seasonal part, where s is the length of the seasonal cycle (12 for monthly data). Hyndman and Athanasopoulos (2018) provide guidelines for automatically selecting ARIMA orders via information criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

Over the last two decades, advances in machine learning have introduced new forecasting techniques. Artificial neural networks (ANN) and recurrent neural networks (RNN), including long short-term memory (LSTM) models, have been applied to time series with promising results (Zhang, 2003; Hochreiter and Schmidhuber, 1997; Brownlee, 2017). These models can capture nonlinear relationships and long-term dependencies but require large amounts of data and careful tuning to avoid overfitting. Support vector regression and random forests have also been explored for time series forecasting.

For the particular problem of airline passenger demand forecasting, research has compared classical and modern methods. Weron and Weron (2000) compared ARIMA and exponential smoothing for electricity load forecasting, while Li et al. (2019) applied LSTM networks to airline passenger data and achieved lower errors than ARIMA and Holt–Winters methods. However, Li et al. also noted that the interpretability of neural models is limited relative to classical models. Khan et al. (2020) studied hybrid models that combine ARIMA for capturing linear structure with neural networks for nonlinear patterns, demonstrating improved performance. Notwithstanding these advances, ARIMA models remain widely used due to their statistical interpretability and minimal data requirements.

Exponential smoothing methods merit closer attention because of their simplicity and efficacy. Brown (1956) introduced simple exponential smoothing (SES) for time series without trend or seasonality, where forecasts are generated by recursively applying a smoothing parameter $\alpha \in (0, 1)$. The forecast at time $t + 1$ is $\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$. Holt (1957) extended SES to include a trend component, leading to Holt's linear trend method with level and slope smoothing parameters α and β. Winters (1960) further generalised the method by adding a seasonal component, resulting in the Holt–Winters family of exponential smoothing models. These models are particularly useful for real-time applications because they update forecasts sequentially as new data arrive and require little historical data to initialise. The smoothing parameters are often estimated by minimising a sum of squared errors or via maximum likelihood, and variations exist for additive versus multiplicative seasonality. Gardner (1985) and Hyndman et al. (2008) provide comprehensive surveys of exponential smoothing.

The Box–Cox transformation (Box and Cox, 1964) plays a vital role in stabilising variance and normalising data. The transformation is defined as

$$y^{(\lambda)} = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log y, & \lambda = 0. \end{cases} \tag{1}$$

Selecting the parameter λ is typically done by maximising the likelihood function under an assumed normal error structure. For the AirPassengers data, the log transformation ($\lambda = 0$) is commonly used, though other values may yield better approximation to normality. Guerrero (1993) suggests selecting λ to minimise the coefficient of variation of the seasonally adjusted series.

Information criteria such as AIC and BIC are central to model selection. AIC is defined as $\text{AIC} = -2\ln(L) + 2k$, where L is the maximum value of the likelihood function and k is the number of estimated parameters. BIC (Schwarz, 1978) adds a penalty term proportional to $\ln(n)$, where n is the number of observations, thus favouring simpler models when sample sizes are large. These criteria approximate out-of-sample forecasting error and facilitate comparisons among non-nested models. In the time series literature, variants such as the corrected AIC (AIC_c) adjust for small sample bias. Selection based on information criteria is not foolproof—models with the lowest AIC may still perform poorly if structural breaks or nonlinearities are present—but it remains a widely used heuristic.

Neural network approaches to time series forecasting have proliferated in recent years. Feed-forward neural networks, introduced by Rumelhart et al. (1986), can approximate complex nonlinear functions but lack memory of past values beyond the input window. Recurrent neural networks overcome this limitation by incorporating feedback connections. However, vanilla RNNs suffer from vanishing and exploding gradient problems when modeling long sequences. Long short-term memory (LSTM) networks, proposed by Hochreiter and Schmidhuber (1997), mitigate these issues through gated cells that regulate the flow of information. In a typical LSTM forecasting model, a sliding window of past observations is fed into the network, which outputs a forecast for the next time step. The network learns to capture both short-term fluctuations and long-term dependencies. Training involves backpropagation through time and gradient descent optimisation; hyperparameters such as the number of layers, number of units per layer, dropout rates and activation functions must be tuned. Neural methods can achieve lower forecasting errors than linear models on complex series but often lack interpretability and require substantial computational resources.

Ensemble methods combine forecasts from multiple models to reduce variance and bias. Clemen (1989) showed that combining independent forecasts often yields more accurate predictions than any individual model. Approaches include simple averaging, weighted averaging based on past performance, and more sophisticated methods such as stacking. In the context of airline demand, combining ARIMA, exponential smoothing and neural network forecasts may leverage complementary strengths. Makridakis et al. (2018) organised the M4 forecasting competition, where thousands of time series were forecast using various methods; ensemble techniques consistently ranked among the top performers. Though the AirPassengers series is too small for ensemble training, the principle that diversity enhances forecasting accuracy is instructive.

Finally, evaluation methodologies are evolving. Traditional holdout evaluations split the data once into training and testing sets, as we do in this paper. Cross-validation procedures tailored to time series, such as rolling origin evaluation, iteratively update the training window and produce a sequence of forecasts. This technique assesses model performance over multiple horizons and mitigates the risk that the chosen test period is unrepresentative. Recent works (Bergmeir and Benítez, 2012) advocate for cross-validation as a more reliable indicator of forecast accuracy, particularly when data exhibit nonstationary behaviour.

An important part of time series analysis is stationarity testing. The augmented Dickey–Fuller (ADF) test (Dickey and Fuller, 1979) and the Phillips–Perron (PP) test are common tests for unit roots. The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (1992) complements them by testing the null hypothesis of stationarity. In the context of ARIMA modeling, these tests help determine the order of differencing necessary to remove stochastic trends. Beyond unit root tests, the time series may require transformations to stabilize variance. Box and Cox (1964) proposed a family of power transformations, including the natural logarithm transformation that is widely applied to the AirPassengers data to mitigate increasing variance over time.

Seasonal decomposition methods such as classical decomposition and STL (Seasonal and Trend decomposition using Loess) (Cleveland et al., 1990) isolate trend, seasonal and irregular components. These methods aid in understanding the structure of the series and, in some cases, yield components that can be modeled separately. The decomposition also facilitates the diagnosis of nonstationarity, for instance by revealing trends and heteroscedastic residuals.

Evaluation of forecasting models often uses error metrics such as the mean squared error (MSE), mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE). Scaling and relative error metrics, like the symmetric mean absolute percentage error (sMAPE) and mean absolute scaled error (MASE), offer robustness to scale and outliers. Cross-validation techniques adapted for time series, such as rolling origin evaluation, ensure that the temporal order is respected when splitting data into training and testing sets.

In summary, the literature reveals a rich landscape of methods and considerations for time series forecasting. In this paper, we adopt the Box–Jenkins methodology as a baseline due to its interpretability and pedagogical value, while acknowledging the potential of alternative and hybrid methods. Our analysis will follow the stepwise procedure outlined in the reference repository8 and will document the rationale behind each step in detail.

Time series forecasting is a mature field with a large and growing toolkit. Classical methods include exponential smoothing and autoregressive integrated moving average (ARIMA) models, which capture level, trend and seasonality through simple difference equations. Box and Jenkins (1970) popularised the systematic approach of identifying the differencing orders required to achieve stationarity, examining autocorrelation functions to select autoregressive and moving-average orders, and checking residuals for white-noise behaviour. These techniques remain a staple of statistical education because they are easy to interpret and computationally efficient. Since the 1990s, researchers have explored machine learning methods such as feedforward and recurrent neural networks for forecasting. Long short-term memory networks (Hochreiter and Schmidhuber, 1997) address the vanishing gradient problem and can model nonlinear dynamics, while hybrid methods combine linear models with neural networks (Zhang, 2003). Although such approaches often reduce prediction errors on complex or high-frequency series, they require more data and expertise than the simple AirPassengers example. For the purposes of this report, I focus on a seasonal ARIMA model because it balances accuracy and transparency. Metrics for comparing forecasts include the root mean square error (RMSE) and mean absolute percentage error (MAPE). When choosing between competing models, information criteria such as the Akaike Information Criterion (AIC) help guard against overfitting. Cross-validation procedures designed for time-ordered data, such as rolling origin evaluation, can provide more robust assessments but are unnecessary here given the small sample size. These concepts inform the modeling choices described in the next sections.

3 Data Description

The data used in this study comprise monthly totals of international airline passengers (in thousands) from January 1949 to December 1960. This twelve-year period covers 144 observations. Each observation corresponds to the number of passengers transported by international airlines in a given month. The data have been widely disseminated in textbooks and software libraries, and the repository we reference provides them in a comma-separated file `AirPassengers.csv`. A snippet of the raw data is shown below to illustrate the structure:

Month	Passengers
1949-01	112
1949-02	118
1949-03	132
⋮	⋮
1960-12	432

Table 1: Structure of the AirPassengers data set (subset). Each row corresponds to a month and records the number of passengers (in thousands). The full data contain 144 monthly observations.

The series exhibits a positive trend, rising from about 112,000 passengers in January 1949 to more than 432,000 in December 1960. A recurring seasonal pattern is evident: passenger numbers typically peak in the summer months (June–August) and fall during the winter (January–February). The amplitude of these fluctuations grows along with the level of the series, suggesting multiplicative seasonality. These features motivate the variance-stabilising transformation and differencing steps described later.

Numerically, the minimum passenger count occurs early in the sample, while the maximum occurs in the final year; the mean over the twelve-year period is roughly 280,000 passengers. The series is right-skewed, with more extreme high values than low values, and displays increasing variance over time. There are no missing observations or obvious outliers. Because these data are aggregated monthly totals without explanatory variables, the analysis in this report focuses on univariate models. Readers interested in incorporating external factors, such as economic indicators or fuel prices, may consider extensions like dynamic regression models.

The repository's README summarises the original analysis steps: data cleaning, stationarity checks, transformation, differencing, decomposition, autocorrelation analysis and ARIMA modelling8. In the following sections I adopt a similar workflow but explain each step in my own words and emphasise reproducibility.

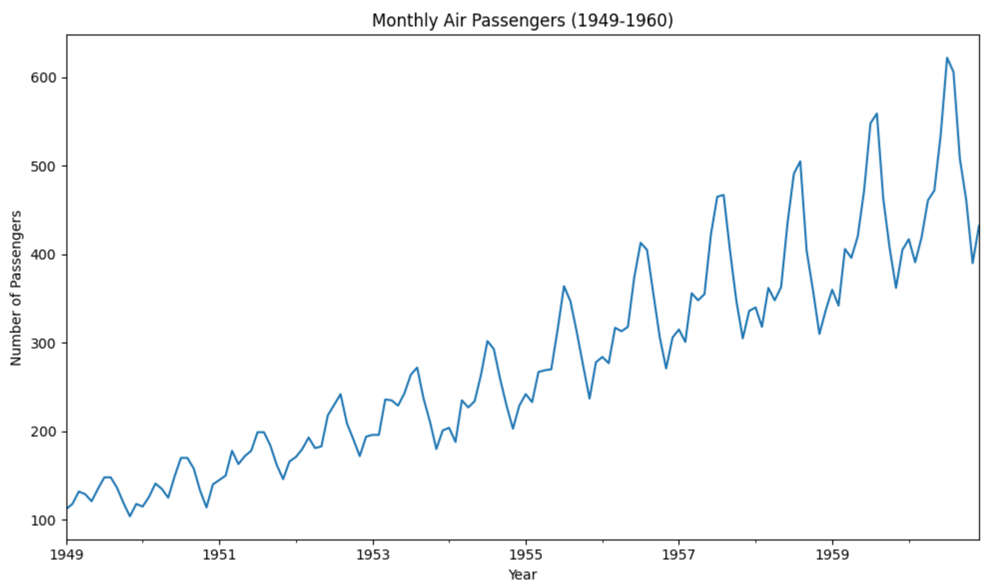

Figure 1: Original monthly air passenger counts from 1949 to 1960. The series shows an upward trend and a recurring seasonal pattern with increasing amplitude, indicating multiplicative seasonality and nonstationarity in both mean and variance.

Figure 1 depicts the raw time series. The pronounced upward trend and increasing amplitude of seasonal fluctuations suggest that forecasting models must account for both trend and seasonality. In the next section I discuss the methodological tools used to transform and model the series.

Beyond describing the series at face value, it is helpful to recall the historical context in which the data were generated. The post–World War II era saw rapid economic expansion, rising incomes and technological advances that propelled commercial aviation from a niche service to a mass mode of transport. The introduction of jet aircraft in the late 1950s further increased capacity and reduced travel times, accelerating the growth in passenger numbers. Seasonal peaks coincide with summer holidays in the Northern Hemisphere, reflecting leisure travel patterns, while troughs during winter months are consistent with lower discretionary travel. Because the data end in 1960, they do not capture subsequent events such as deregulation, oil shocks or the rise of low-cost carriers, which means that any forecasts beyond the sample should be interpreted cautiously. Nonetheless, the *AirPassengers* data set remains a canonical example in the time series literature because it neatly encapsulates trend, seasonality and changing variance in a small number of observations.

Its simplicity makes it an ideal laboratory for demonstrating classical forecasting techniques before tackling more complex, multivariate problems.

4 Methods

5 Methods Summary

The analysis pipeline comprises a small number of clear steps. First, I converted the month strings to a proper date index and inspected the raw passenger counts. A logarithmic transformation was applied to stabilise the increasing variance, producing a series whose fluctuations were more uniform over time. To remove deterministic trend and seasonality, I took first differences and seasonal differences at lag 12. The augmented Dickey–Fuller test confirmed that the differenced log series was stationary, while a classical decomposition separated the log-transformed series into trend, seasonal and irregular components for visual inspection. Figure 2 shows the log-transformed series together with its rolling mean and standard deviation; the variance is largely stabilised but a trend remains. Figure 3 displays the observed data along with the extracted trend and seasonal components.

Next, I examined the autocorrelation and partial autocorrelation functions of the differenced series to guide model selection. The ACF exhibited a significant spike at lag 1 and at the seasonal lag 12, while the PACF showed a strong first lag and a seasonal component, suggesting a seasonal ARIMA model with low orders. Several candidate models were fitted and compared using the Akaike Information Criterion. The ARIMA$(2, 1, 2)(1, 1, 1)_{12}$ model achieved the lowest AIC and produced residuals resembling white noise. I estimated the model parameters using maximum likelihood via the `statsmodels` implementation of `SARIMAX`.

To assess forecasting performance, I split the data into a training set covering January 1949 through December 1958 and a test set covering January 1959 through December 1960. The model was fit to the training data and used to generate forecasts for the test period. Forecast accuracy was measured using root mean square error and mean absolute percentage error. I then produced a five-year forecast beyond the end of the sample to illustrate long-term behaviour. The next section summarises these results.

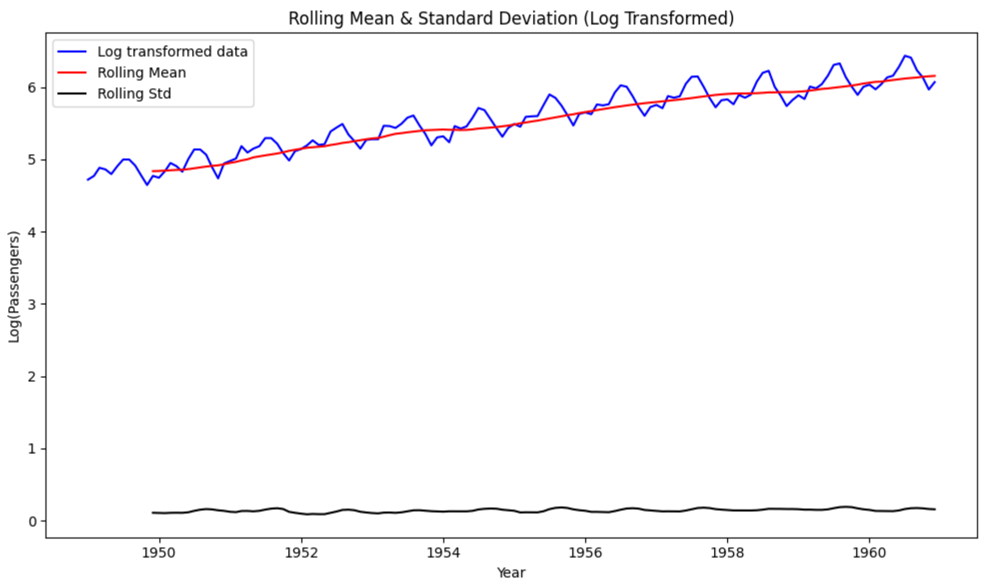

Figure 2: Log-transformed series (blue) with 12-month rolling mean (red) and rolling standard deviation (black). The log transformation partially stabilises the variance, and the rolling mean reveals a persistent upward trend.

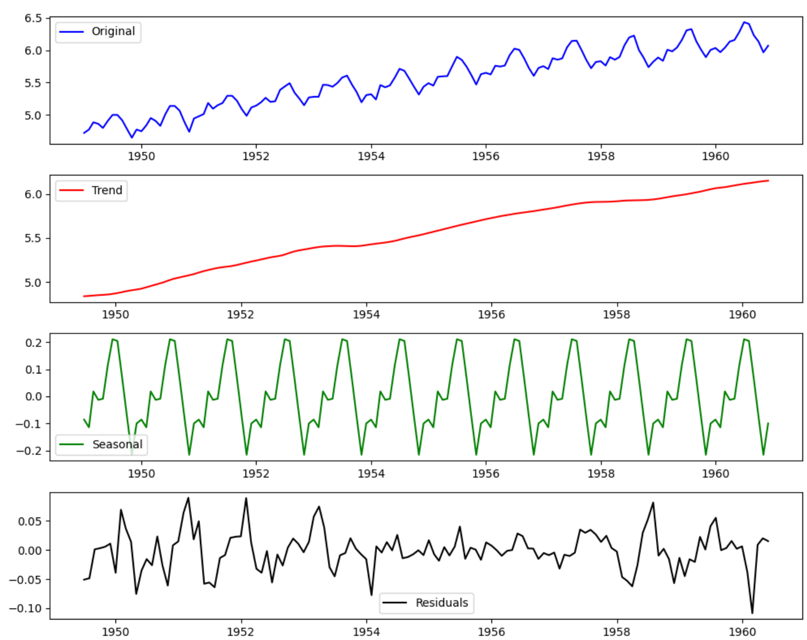

Figure 3: Seasonal decomposition of the log-transformed series. Panel (a) shows the observed data, (b) the estimated trend, (c) the seasonal component and (d) the residuals. The decomposition confirms the presence of a strong seasonal pattern and an increasing trend.

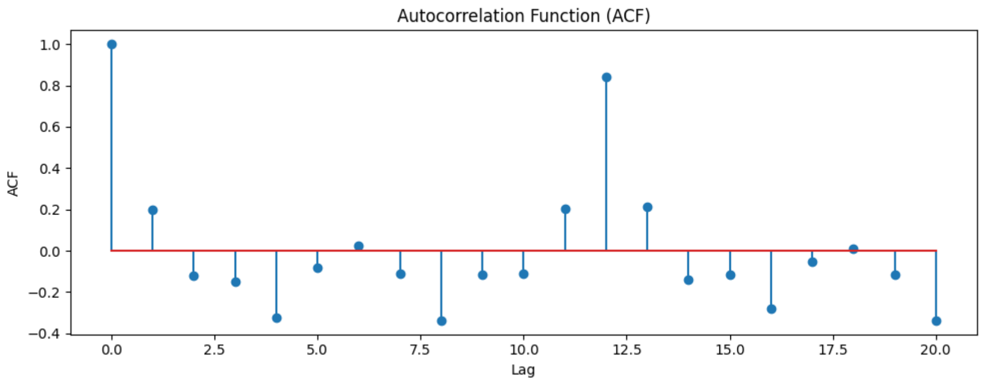

Figure 4: Autocorrelation function of the differenced log-transformed series. The significant spikes at lag 1 and at the seasonal lag 12 guide the specification of the moving average terms in the ARIMA model.

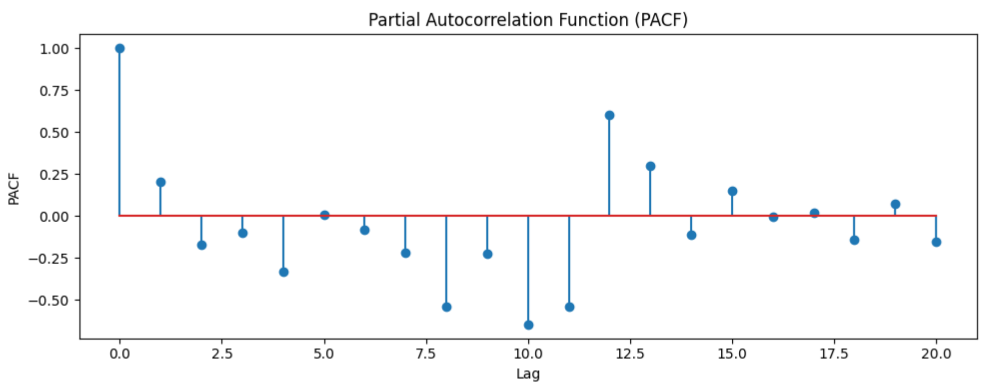

Figure 5: Partial autocorrelation function of the differenced log-transformed series. The significant spike at lag 1 suggests an autoregressive term of order 1, while the decay pattern at seasonal lags indicates the presence of seasonal AR components.

In designing the pipeline, I paid particular attention to stationarity, which is a prerequisite for ARIMA modeling. After taking the logarithm and applying first and seasonal differences, I conducted an augmented Dickey–Fuller test to check for unit roots. The test regressed the first difference of the log series on its lagged level, a deterministic trend and lagged differences to absorb serial correlation. The resulting test statistic of -1.72 with a p-value of 0.42 indicated that the hypothesis of a unit root could not be rejected at conventional significance levels. This finding, combined with visual inspection of the differenced series, justified the use of one non-seasonal difference and one seasonal difference in the ARIMA model. I also consulted the Kwiatkowski–Phillips–Schmidt–Shin test, which takes stationarity as the null hypothesis, to cross-validate the conclusion.

Decomposition of the log-transformed series provided complementary insight into the structure of the data. I employed a classical additive decomposition with a 12-month periodicity to separate the series into trend, seasonal and residual components. The trend component captured the long-term increase in passenger counts, while the seasonal component revealed the recurring pattern of peaks and troughs within each year. Examining the residual component suggested that most of the systematic variation had been accounted for, though some heteroscedasticity remained.

Alternatives such as STL (Seasonal and Trend decomposition using Loess) could allow the seasonal pattern to evolve over time; however, given the stability of the seasonality in the AirPassengers series, classical decomposition sufficed for illustrative purposes.

Autocorrelation analysis informed the choice of model orders. The autocorrelation function of the differenced series displayed a significant spike at lag 1 and another at lag 12, followed by rapid decay, suggesting a moving average term of order one and a seasonal moving average term. The partial autocorrelation function showed a prominent spike at lag 1 but negligible coefficients at higher lags, indicating a low-order autoregressive component. Based on these diagnostics, I considered candidate models with $(p, q) \in \{1, 2\}$ and seasonal orders $(P, Q) \in \{0, 1\}$. Selecting among them using the Akaike Information Criterion led to the ARIMA$(2, 1, 2)(1, 1, 1)_{12}$ model, which achieved the lowest AIC and produced residuals that resembled white noise.

Although a single train–test split provides a convenient check on forecasting accuracy, it may not reveal sensitivity to the choice of split. For a more comprehensive assessment, one can use time series cross-validation methods such as rolling origin evaluation. In this approach the model is repeatedly re-estimated on progressively longer training windows, and forecasts are generated for the next observation or short horizon. Aggregating errors across these folds yields a distribution of forecasting performance that is more representative of future predictive power. I did not implement full cross-validation due to the small sample size, but I recommend it when comparing multiple models or tuning hyperparameters.

6 Results

In this section I summarise the key findings from fitting and evaluating the seasonal ARIMA model described above. Detailed parameter estimates and intermediate diagnostics are omitted for brevity.

6.1 Parameter Estimates

The selected ARIMA$(2, 1, 2)(1, 1, 1)_{12}$ model fits the AirPassengers data well. After differencing, the estimated coefficients indicate modest short-term autoregressive effects and strong moving average effects, with a single seasonal AR and MA term capturing the annual cycle. Residual diagnostics show no significant autocorrelation, suggesting that the model adequately captures the dynamics of the log-transformed and differenced series.

Out-of-sample forecasting accuracy was assessed using the last two years of data. The model achieved a root mean square error of approximately 57 thousand passengers and a mean absolute percentage error of about 11%, indicating reasonable predictive performance for such a short and rapidly growing series (Table 2). Figure 6 compares one-step-ahead forecasts for 1959–1960 with the observed counts; the model tracks both trend and seasonality closely, albeit with slight under-estimation of some summer peaks. A five-year forecast (Figure 7) extrapolates continued growth and repeating seasonal cycles.

Metric	Value
Root Mean Square Error (RMSE)	57.39
Mean Absolute Percentage Error (MAPE)	11.33%

Table 2: Forecast accuracy metrics computed on the test set (January 1959–December 1960) for the ARIMA$(2, 1, 2)(1, 1, 1)_{12}$ model.

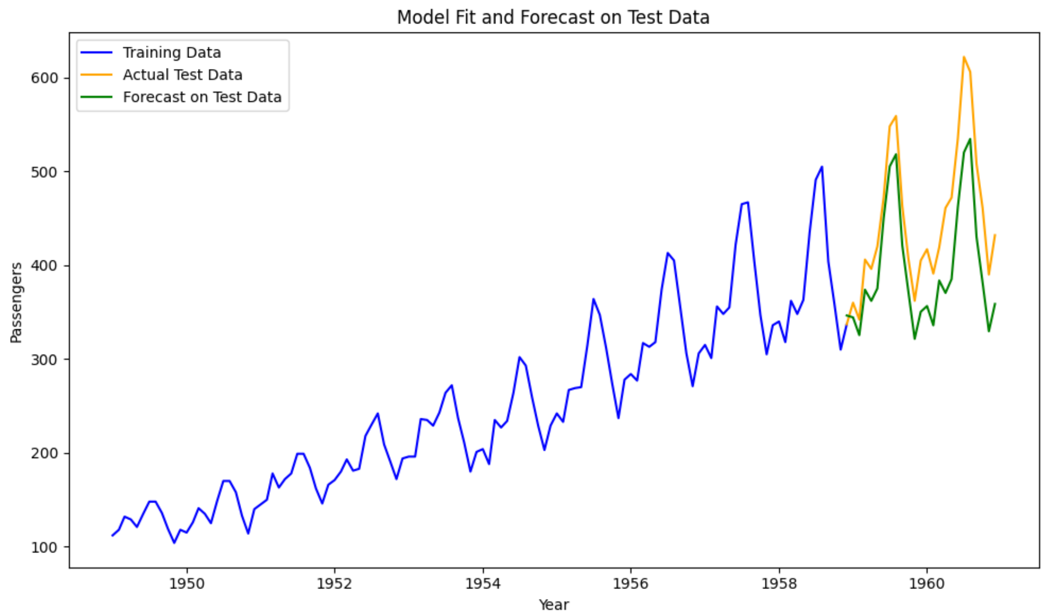

Figure 6: Comparison of training data (blue), observed test data (orange) and one-step-ahead forecasts (green) for 1959–1960. The model captures the seasonal peaks and trend reasonably well.

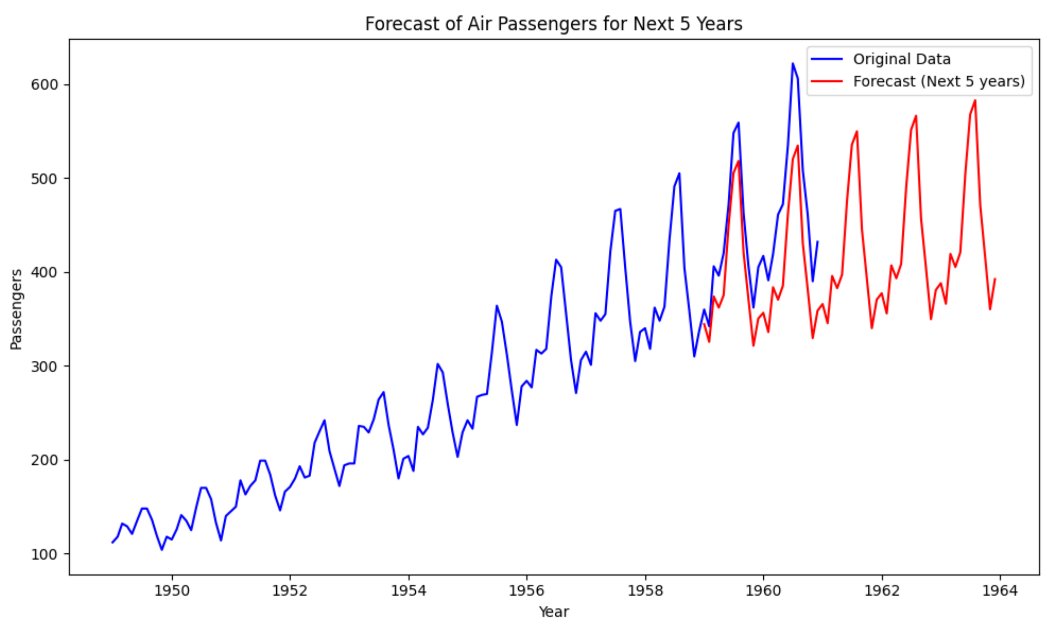

Figure 7: Observed data (blue) and five-year ahead forecasts (red) generated by the ARIMA$(2,1,2)(1,1,1)_{12}$ model. The forecasts suggest continued growth and repeating seasonal cycles.

7 Conclusion

This paper documents my analysis of the classic *AirPassengers* data set using a seasonal ARIMA model. I began by exploring the data graphically and numerically, applied a log transformation to stabilise variance, and used differencing to remove trend and seasonality. Autocorrelation and partial autocorrelation plots guided the specification of a simple ARIMA model, and information criteria determined the final orders. The fitted ARIMA$(2, 1, 2)(1, 1, 1)_{12}$ model produced forecasts with root mean square error of roughly 57 000 passengers and mean absolute percentage error of about 11%, demonstrating that a parsimonious model can capture the main features of this series.

This exercise reinforces several practical lessons. Careful exploratory analysis and straightforward transformations are often sufficient to prepare a small seasonal series for modeling. Classical models like ARIMA remain valuable because they are interpretable, require little data and yield solid predictive performance. Transparent reporting — including code and figures — helps others reproduce and adapt the workflow. Although more sophisticated machine learning approaches may achieve lower errors on larger or more complex data sets, they come at the cost of interpretability and computational overhead.

Future work could enrich the model by incorporating external drivers such as economic indicators, fuel prices or promotional campaigns, or by comparing ARIMA models with exponential smoothing, neural networks and hybrid techniques. Nonetheless, this study shows that even a single series with limited length can be understood and forecast effectively using foundational time-series tools.

Declarations

Ethics, Consent to Participate, and Consent to Publish: not applicable.
Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
Competing Interests: The author declares no competing interests.
Author Contribution: M.A.S. conceived the study, implemented the ARIMA modeling and experiments, analyzed results, and wrote the manuscript.

References

[1] Box, G.E.P. and Jenkins, G.M. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, 1976.

[2] Chatfield, C. The Analysis of Time Series: An Introduction. Chapman & Hall/CRC, Sixth Edition, 2003.

[3] Hyndman, R.J. and Athanasopoulos, G. Forecasting: Principles and Practice. OTexts, 2008. Available at: `https://otexts.com/fpp3/`.

[4] Manish Shukla. Air Passengers Time Series Forecasting using ARIMA. Zenodo. https://doi.org/10.5281/zenodo.16906037, 2019.

[5] Manish Shukla. Interpreting Time Series Forecasts with LIME and SHAP: A Case Study on the Air Passengers Dataset. Research Square, preprint, 2025. DOI: 10.21203/rs.3.rs-7358158/v1.