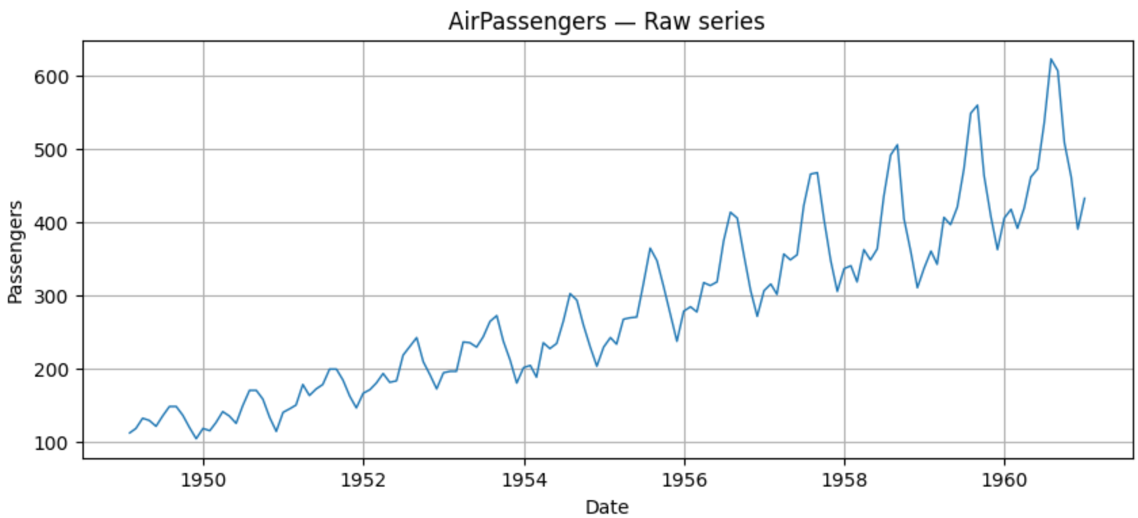
Exploratory Data Analysis (EDA) of Time Series – AirPassengers

This document accompanies the notebook '01_eda_time_series.ipynb' and summarizes the key concepts, visual analyses and statistical diagnostics used in the exploratory data analysis (EDA) of the AirPassengers time series. The figures referenced in each section should be inserted accordingly.

1. Dataset Overview

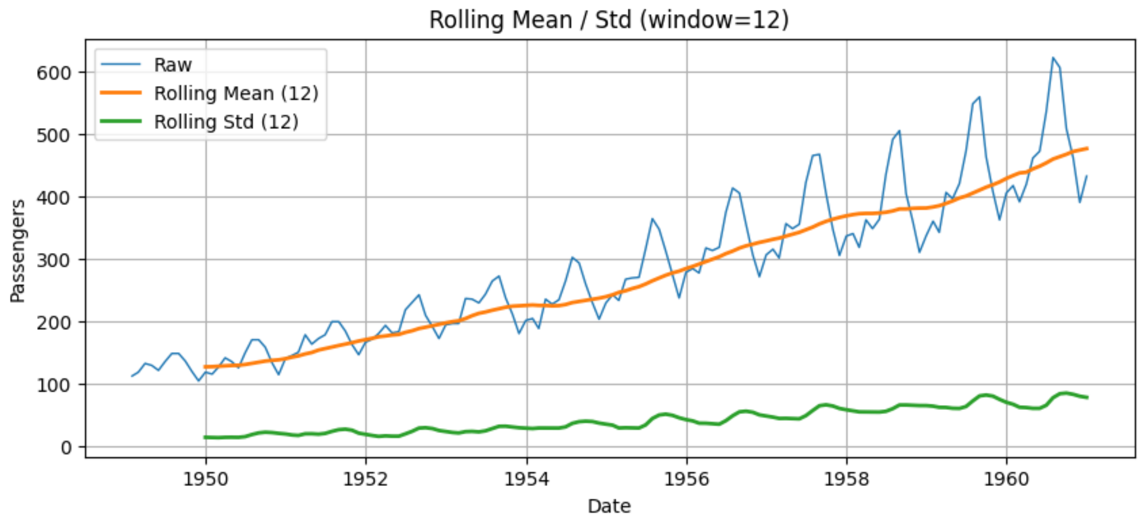
The AirPassengers dataset contains monthly totals of international airline passengers from 1949 to 1960. It is a classical benchmark dataset in time series analysis due to the presence of trend, seasonality, and increasing variance over time.



2. Visual Inspection and Rolling Statistics

The raw time series plot reveals a clear upward trend and strong yearly seasonality. In addition, the amplitude of seasonal fluctuations increases as the level of the series grows, suggesting heteroscedastic behavior.

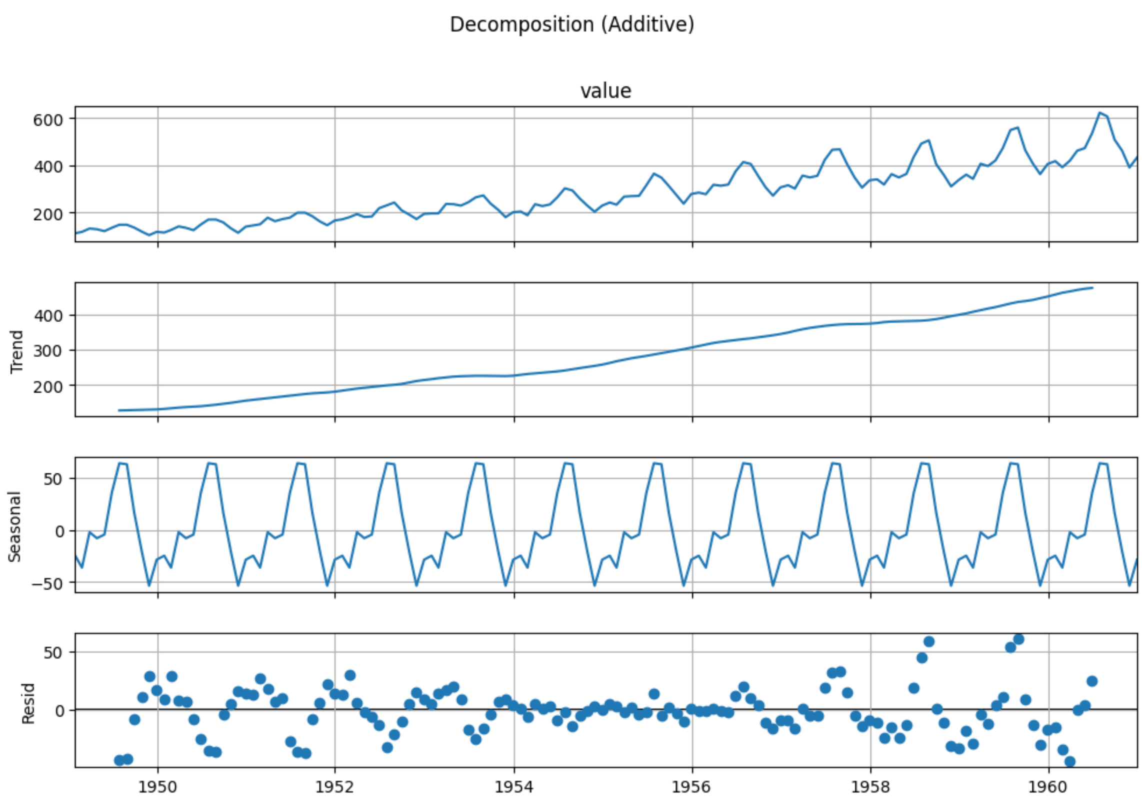
Rolling mean and rolling standard deviation (using a 12-month window) help visualize non-stationarity. Changes in the rolling mean indicate trend, while changes in rolling standard deviation indicate non-constant variance.



3. Time Series Decomposition

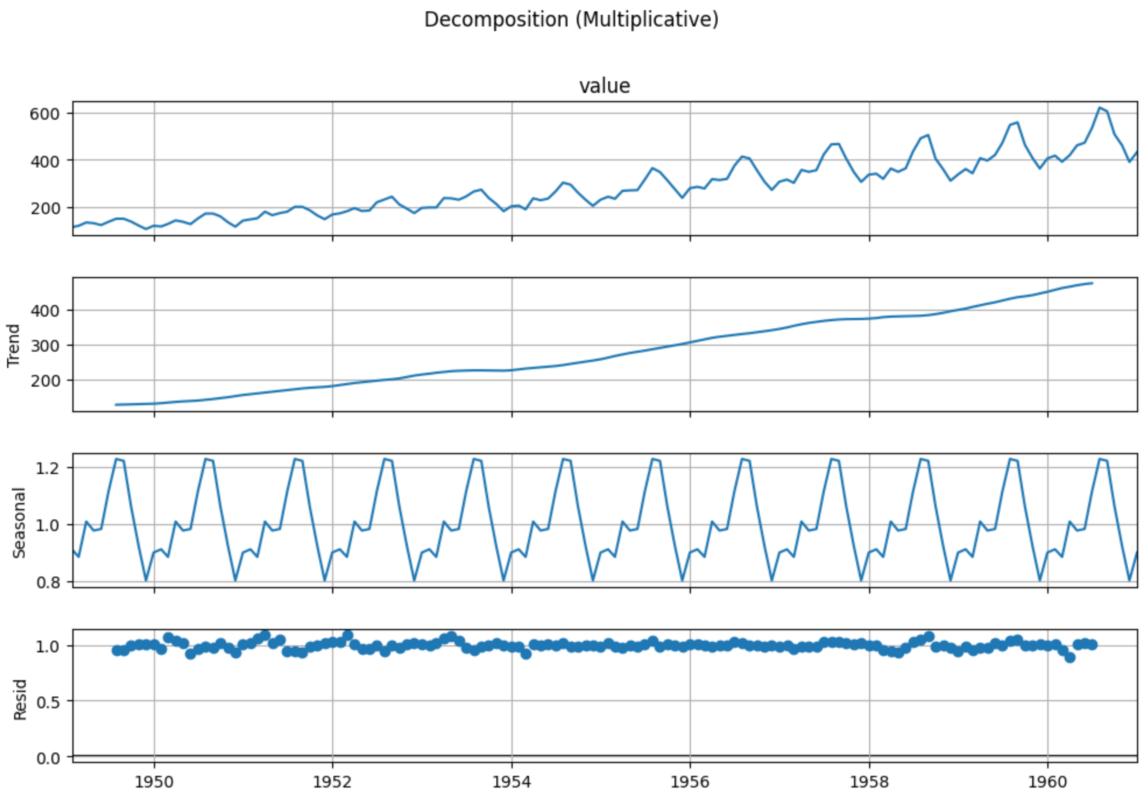
Time series decomposition separates the observed series into trend, seasonal and residual components. Two formulations are commonly used: additive and multiplicative.

Additive decomposition assumes that components add linearly. While it captures the overall trend, it forces seasonal effects to have constant magnitude, which is not consistent with the observed data.



Multiplicative decomposition assumes that seasonal effects scale with the level of the series.

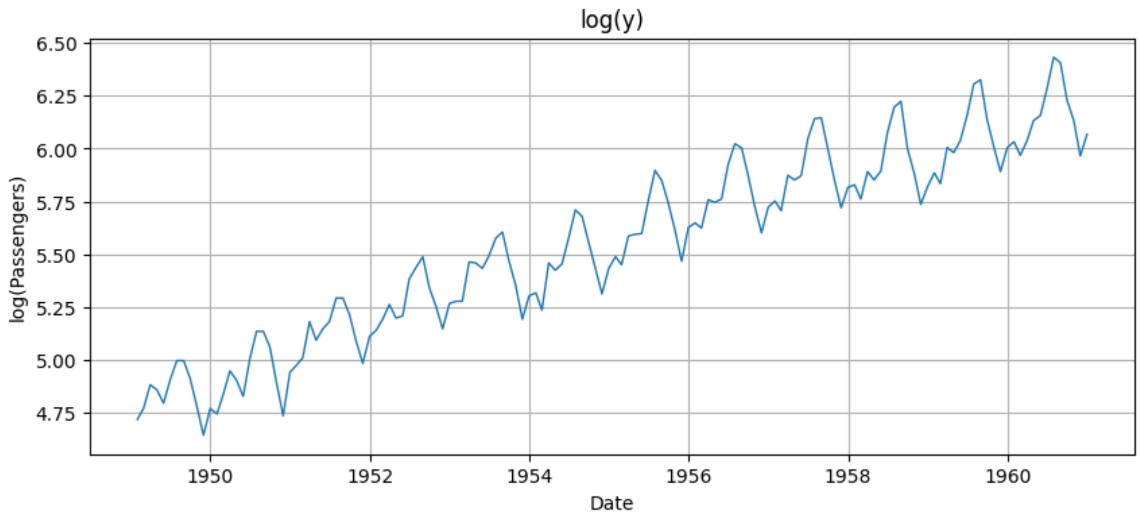
This formulation better captures the increasing amplitude of seasonal fluctuations observed in the data.

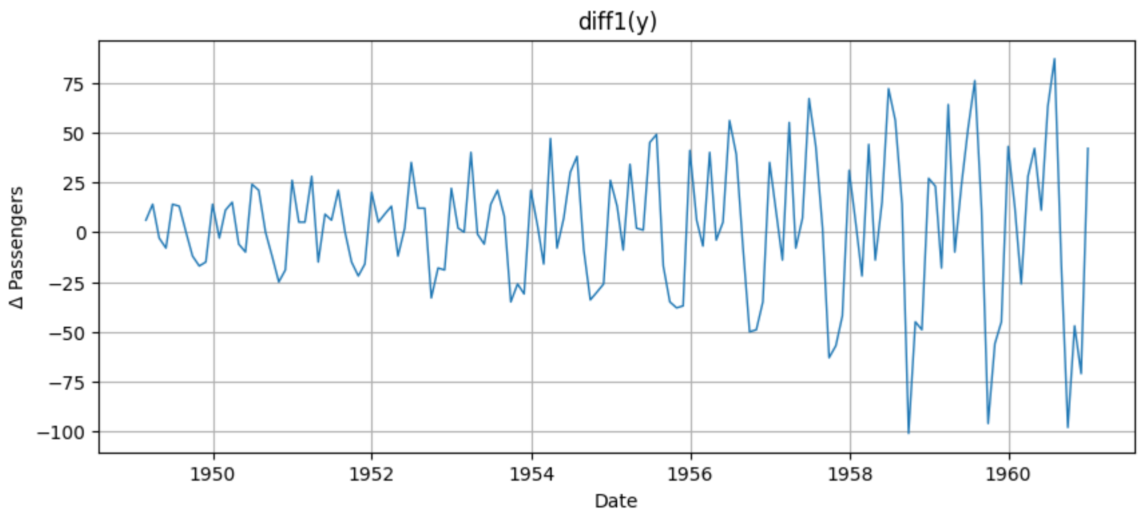


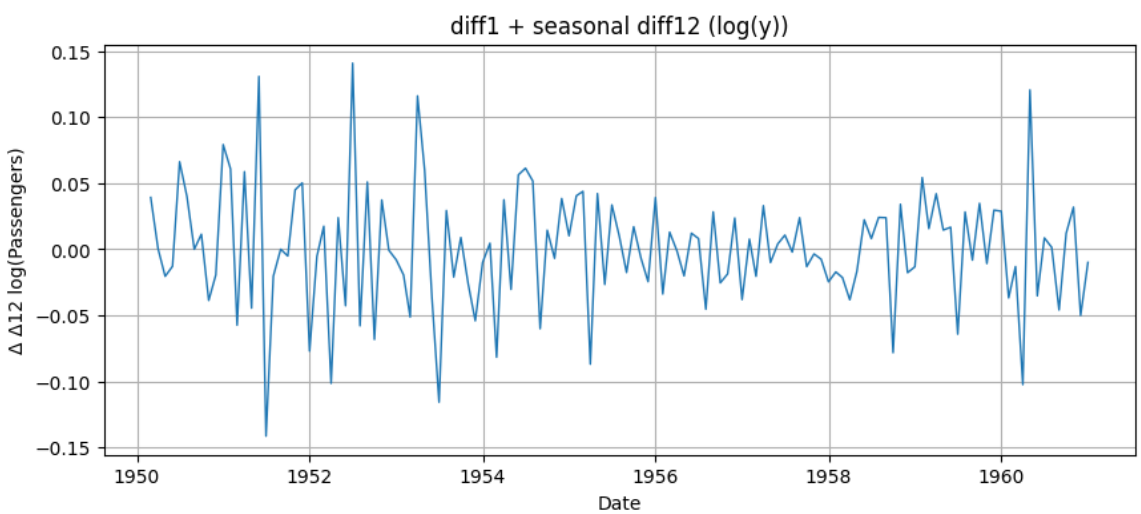
4. Transformations for Stationarity

To prepare the series for modeling, common transformations are applied. A logarithmic transformation stabilizes variance, while differencing removes trend and seasonality.

The combination of log transformation, first differencing, and seasonal differencing is commonly used as a preprocessing step for SARIMA models.


log(y)
log(Passengers)
Date

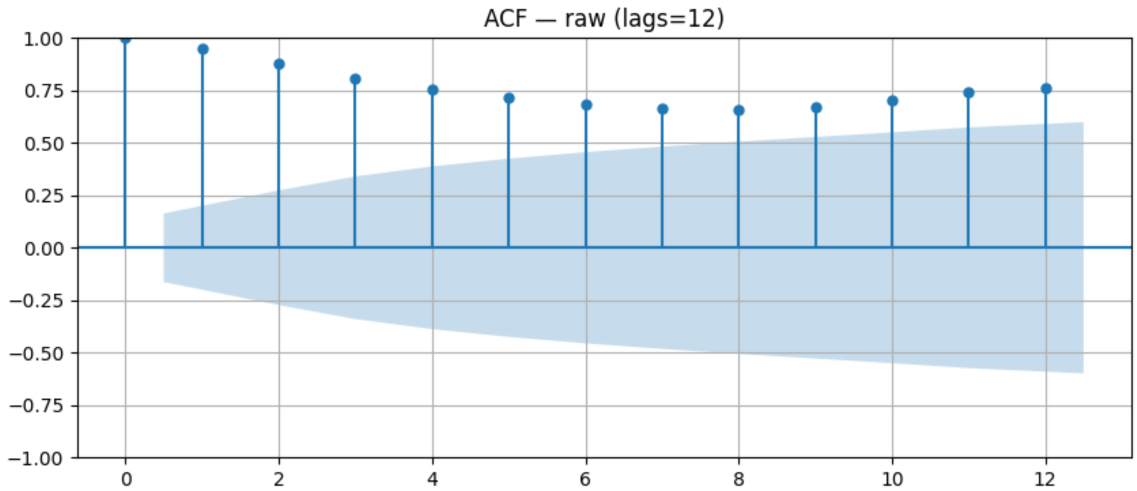

diff1(y)
Δ Passengers
Date

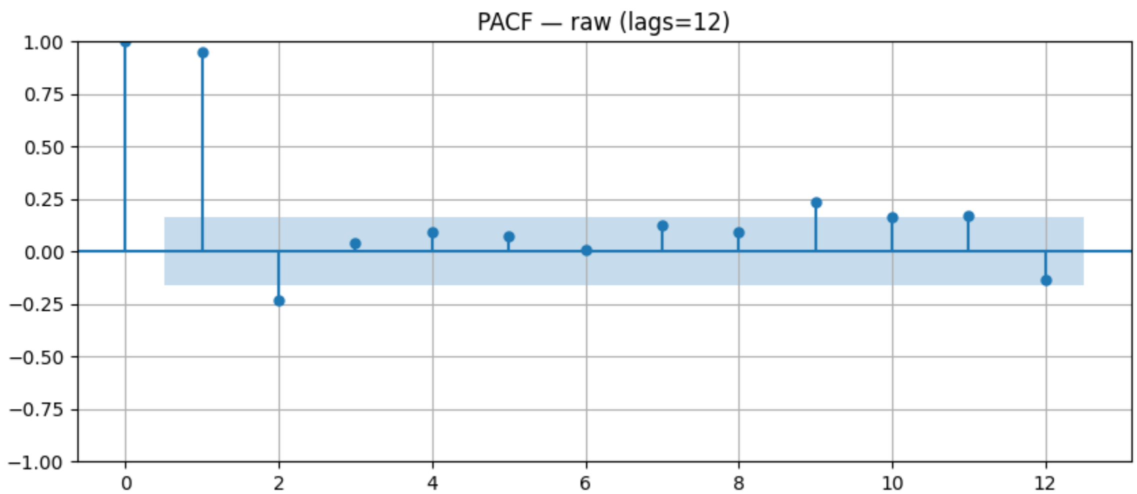

diff1 + seasonal diff12 (log(y))
Δ Δ12 log(Passengers)
Date

5. Autocorrelation Analysis (ACF and PACF)

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are used to analyze temporal dependence and guide model structure selection.

Comparing ACF and PACF before and after transformations helps assess whether the series is approaching stationarity.





6. Statistical Diagnostics

Formal statistical tests are used to complement visual inspection and provide quantitative evidence about the statistical properties of the time series.

Two complementary tests are applied to assess **stationarity**:

- **Augmented Dickey-Fuller (ADF)**
  Tests the presence of a unit root.
  - Null hypothesis (H0): the series is non-stationary
  - A small p-value ($p < 0.05$) suggests stationarity.
- **KPSS (Kwiatkowski–Phillips–Schmidt–Shin)**
  Tests stationarity directly under an opposite null hypothesis.
  - Null hypothesis (H0): the series is stationary
  - A small p-value ($p < 0.05$) suggests non-stationarity.

Using both tests together provides a more robust assessment of stationarity, reducing ambiguous conclusions.

In addition, the **Ljung-Box test** is applied to evaluate whether significant autocorrelation remains in the series up to specific lags:

- Lag 12 corresponds to one year (monthly data)
- Lag 24 corresponds to two years

A small p-value indicates that temporal dependence is still present.

The table below summarizes the diagnostics for selected transformations of the AirPassengers series.

series	n	ADF p-value	KPSS p-value	LB p@12	LB p@24	stationary?
raw	144	0.9919	0.0100	0.0000	0.0000	No
log(y)	144	0.4224	0.0100	0.0000	0.0000	No
log(y) diff1 + seasdiff12	131	0.0002	0.1000	0.0000	0.0000	Yes

The results indicate that:

- The **raw series** is clearly non-stationary.
- The **log-transformed series** stabilizes variance but remains non-stationary.
- The combination of **log transformation, first differencing and seasonal differencing** satisfies common stationarity criteria.

7. Key Takeaways

- The AirPassengers series exhibits trend, seasonality and increasing variance.
- A multiplicative structure better describes the data.
- Log and differencing transformations are required to approach stationarity.
- Proper EDA is essential before applying forecasting models.