Univariate Time Series Data and Model Card

Generated by Cardtale - Automated Model and Data Card Generator

This report provides an automated, comprehensive analysis of univariate time series data. Generated by Cardtale, it explores basic aspects and potential challenges in your data to support informed decision-making and modeling choices.

Generated: 2024-10-29 22:08

Series Name: series_0

Table of Contents

Other aspects were explored but omitted from the final report:

Data Overview

This section examines the core characteristics and statistical properties of the time series. Understanding these attributes is important for assessing data quality and gaining a preliminary context. We explore the temporal structure, summary statistics, and distribution patterns to create a baseline understanding of your data.

Time Series Plot

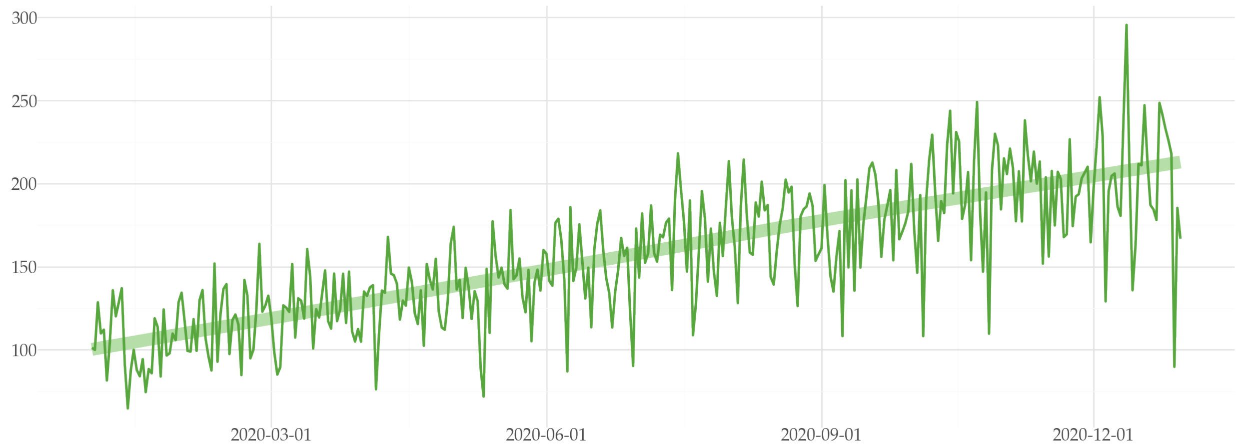

Figure 1: Time series line plot.

- A total of 365 daily observations which span from 2020-01-01 to 2020-12-30.
- The mean value of the series is 155.95 (median equal to 151.88), with a standard deviation of 40.72. The data ranges from a minimum of 64.67 to a maximum of 295.54.

Data Distribution

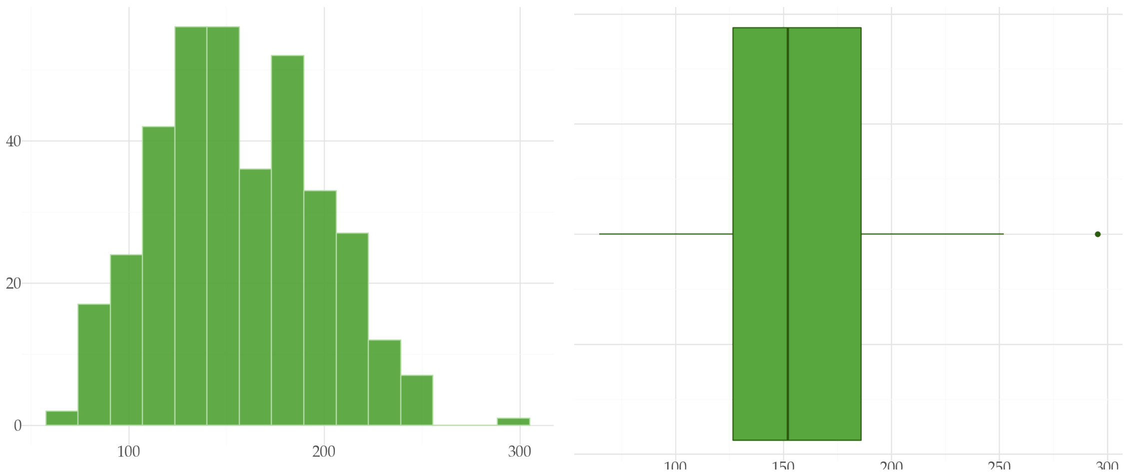

Figure 2: Distribution of the time series using an histogram (left) and a boxplot (right).

- The Kolmogorov-Smirnov test rejects the hypothesis that the series is distributed according to the following distributions: Cauchy, Power-law, Exponential, Pareto, and Chi-squared
- The distribution with largest p-value is Exponentially Modified Gaussian distribution (p-value equal to 0.37). But, we cannot reject the hypothesis that the

data follows the following distributions (ordered decreasingly by p-value): Log-Normal, Gamma, Gaussian, and Logistic.

- There are 1 outliers in the data, all of which are upper outliers. The outliers represent 0.27% of the complete data set.
- The excess kurtosis is equal to -0.46. This indicates that the data has a light tailed distribution.
- The skewness is equal to 0.22, which is close to zero. This indicates a symmetric distribution, though there is a slight right skewness.

Trend and Seasonality

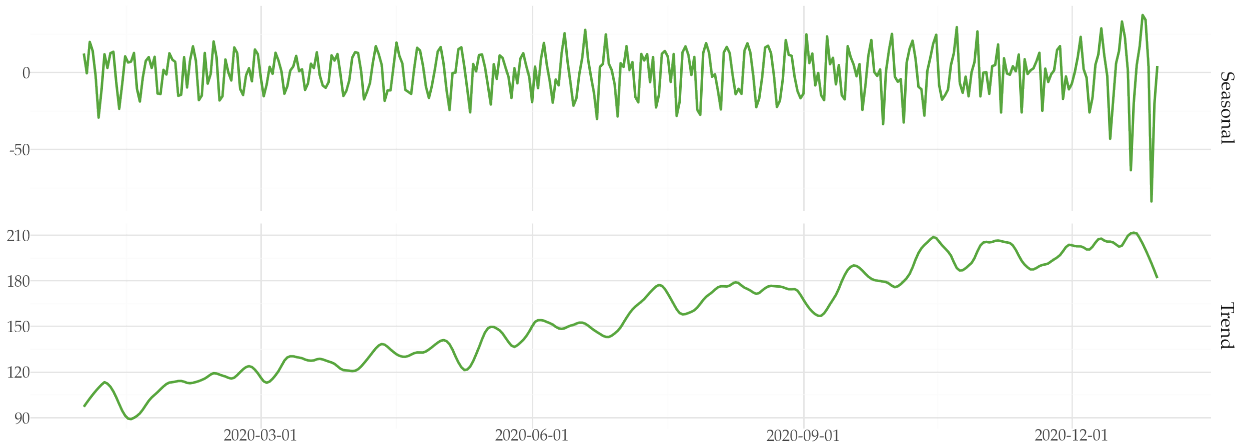

Figure 3: Seasonal and Trend components after decomposition using the STL (Season-Trend decomposition using LOESS) method.

- All hypothesis tests carried out (KPSS, Augmented Dickey-Fuller, and Philips-Perron) indicate that the time series is stationary in trend/level.
- The following tests indicate that the time series is non-stationary in seasonality for the specified period: OCSB. On the other hand, other tests (Wang-Smith-Hyndman) fail to reject the hypothesis that the data is stationary

Auto-Correlation

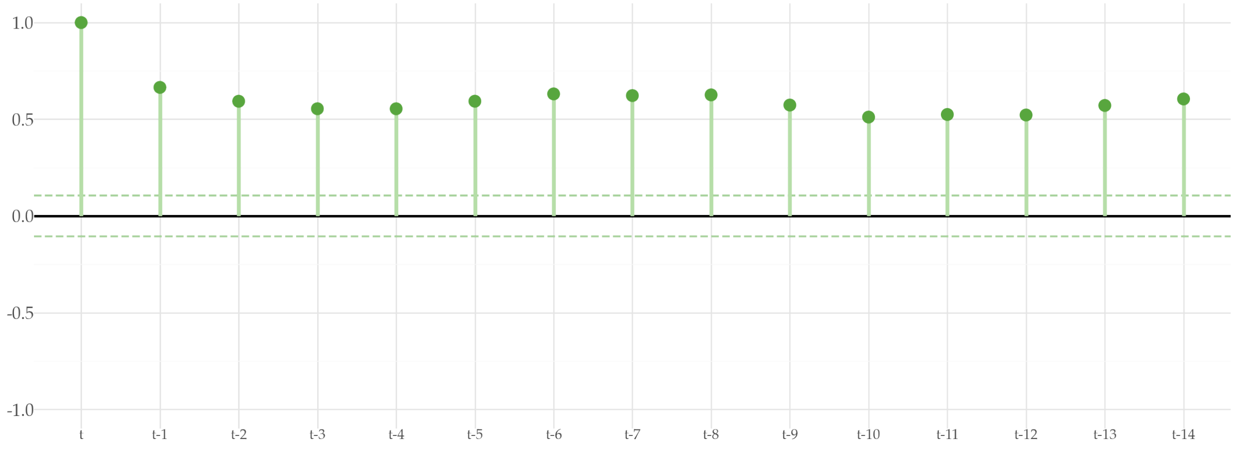

Figure 4: Auto-correlation plot up to 14 lags.

- The following lags show significant autocorrelation: t-1, t-2, t-3, t-4, t-5, t-6, t-7, t-8, t-9, t-10, t-11, t-12, t-13, and t-14. The autocorrelation is positive for all lags with a significant value.
- All lags relative to the seasonal period (t-7 and t-14) show a significant autocorrelation.

Partial Auto-Correlation

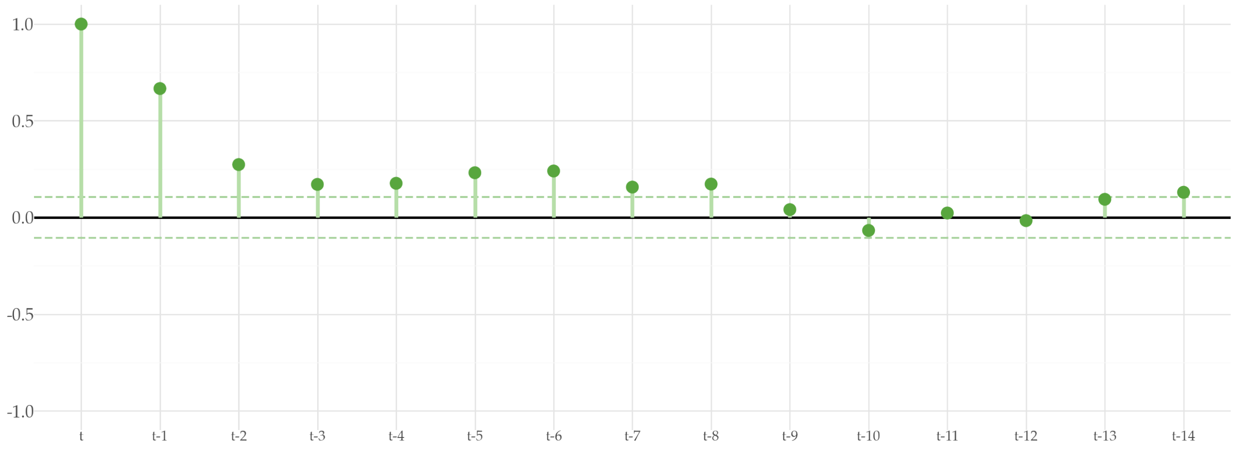

Figure 5: Partial Auto-correlation plot up to 14 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-1, t-2, t-3, t-4, t-5, t-6, t-7, t-8, and t-14.
- All lags relative to the seasonal period (t-7 and t-14) show a significant partial autocorrelation.

Trend

Trend refers to the long-term change in the mean level of a time series. It reflects systematic and gradual changes in the data over time. Understanding the trend is important for identifying long-term growth or decline, structural changes, and making informed modeling decisions. This section examines the characteristics of the trend of the time series.

Trend Line Plot

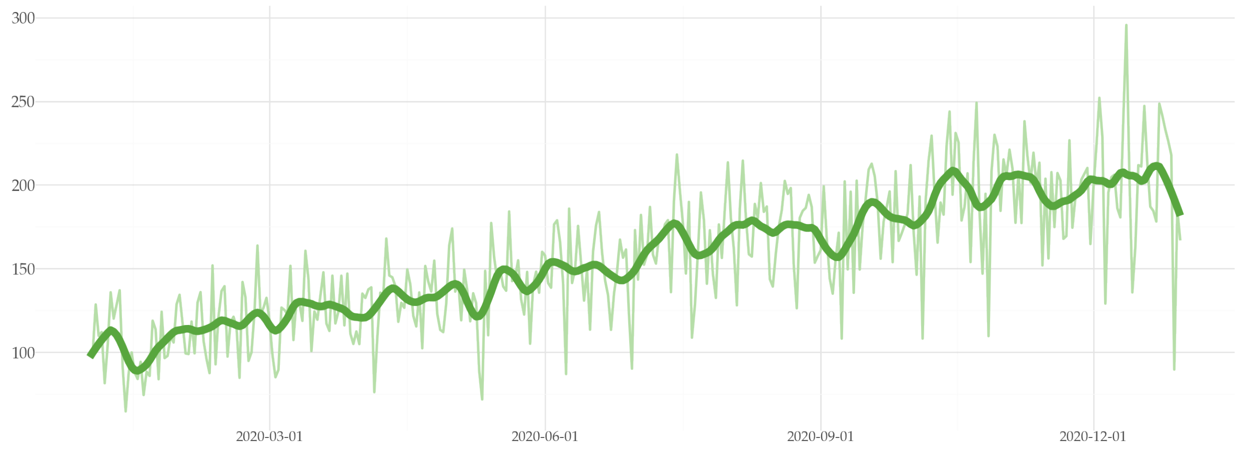

Figure 6: Time series trend plot.

- The time series has non-stationary trend according to the statistical test(s): . There is a strong upward trend.
- On the other hand, the methods KPSS, Augmented Dickey-Fuller, and Philips-Perron did not find evidence for the presence of trend.
- The same tests were applied to analyse whether the time series is stationary around a constant level. The method(s) KPSS and Augmented Dickey-Fuller reject this hypothesis. But, the test(s) Philips-Perron fail to reject.
- Including a trend explanatory variable which denotes the position of each observation does not improve forecasting performance.

Distribution of Differences

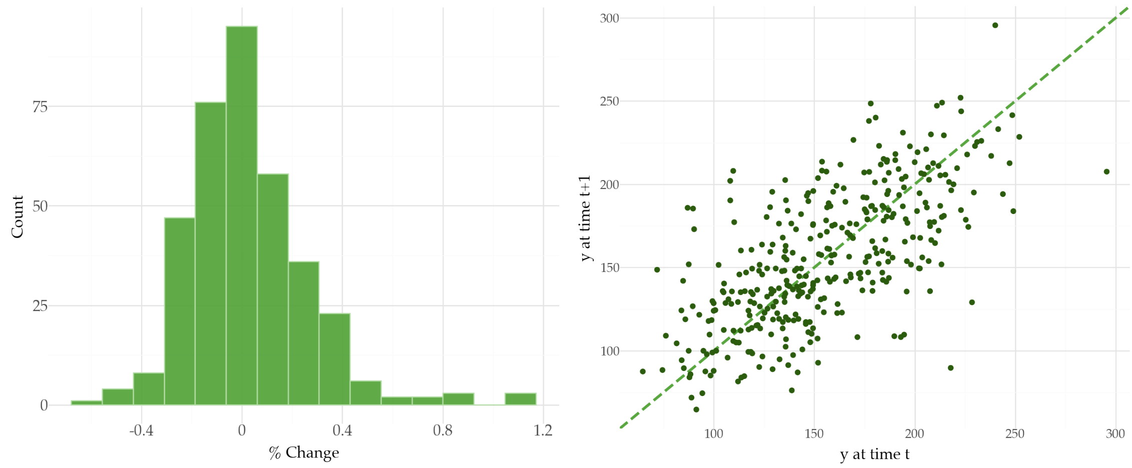

Figure 7: Distribution of percentage changes (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram shown the distribution of these changes. The lag-plot depicts the randomness in the data. The time series shows greater randomness as the points deviate from the dotted line.

- The Kolmogorov-Smirnov test rejects the hypothesis that the differenced series is distributed according to the following distributions: Cauchy, Power-law, Exponential, Pareto, and Chi-squared.
- The distribution with largest p-value is Logistic (p-value equal to 0.98). But, we cannot reject the hypothesis that the differenced series follows the following distributions (ordered decreasingly by p-value): Exponentially Modified Gaussian distribution, Log-Normal, Gamma, and Gaussian.
- The excess kurtosis of the differenced series is equal to 0.76. This indicates that the data has a heavy tailed distribution.
- The skewness of the differenced series is equal to 0.0, which is close to zero. This indicates a symmetric distribution, though there is a slight right skewness.
- **Forecasting experiments**: Taking first differences does not improve forecasting performance.

Seasonality

Seasonality represents recurring patterns or cycles that appear at regular intervals in time series data. These are predictable fluctuations that reflect periodic influences such as monthly, quarterly, or yearly cycles. Understanding seasonal patterns is crucial for forecasting, trend analysis, and identifying anomalies. This section examines the presence, strength, and characteristics of seasonal components in the input time series.

Seasonal Line Plot (Dayly)

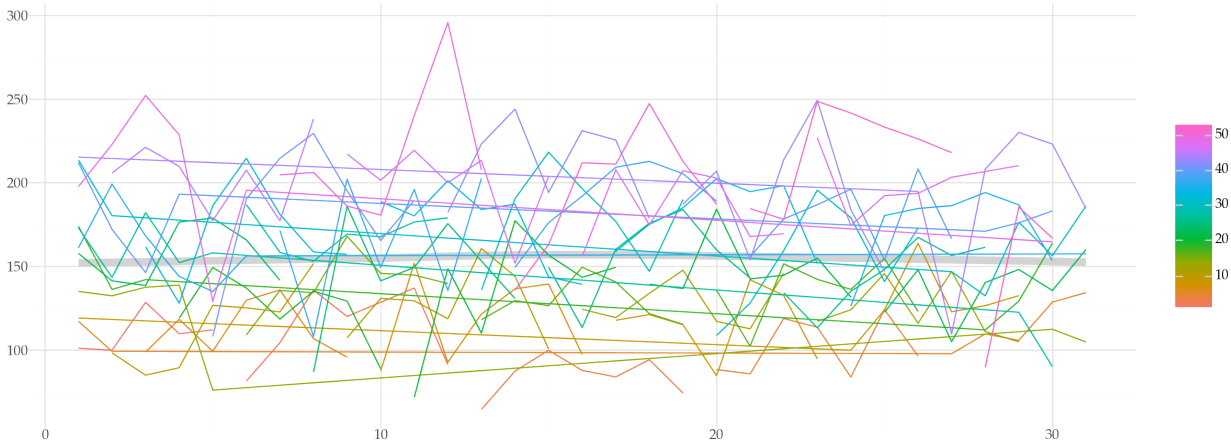

Figure 8: Seasonal plot of daily values grouped by week.

- The following tests indicate that the time series is non-stationary in seasonality for a weekly period: OCSB. On the other hand, other tests (Wang-Smith-Hyndman) fail to reject the stationary null hypothesis.
- **Forecasting experiments**: Including daily information in the predictive model decreases forecasting performance. This information was included as Fourier terms and repeating basis function terms in the explanatory variables.
- We also searched for Daily seasonal patterns. But, all statistical tests confirmed stationarity in those periods. Besides, including information about these periods in the forecasting model decreased its performance.
- Statistical tests were carried out to check for differences among means and variances across days. No significant differences were found.

Seasonal Sub-series Plot (Dayly)

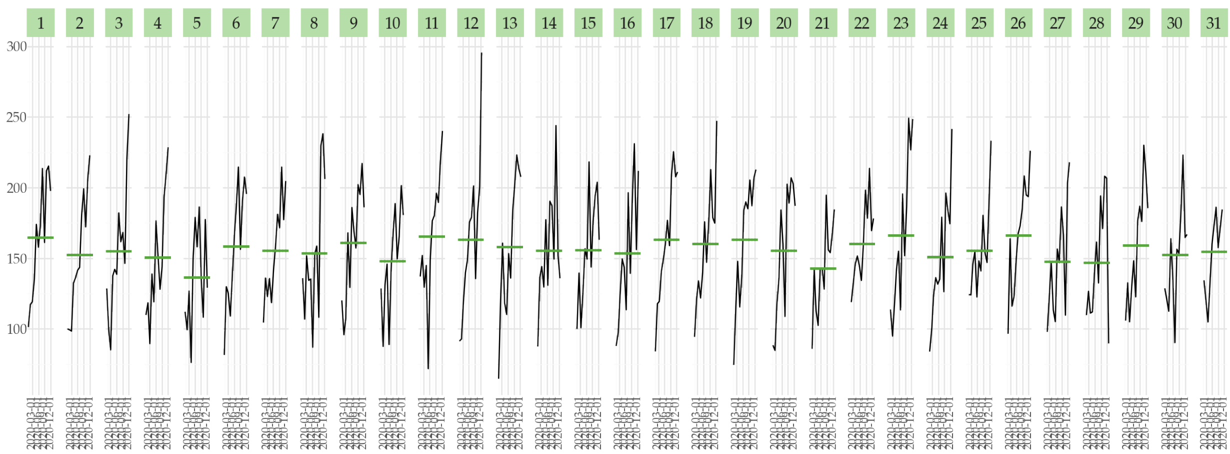

Figure 9: Dayly seasonal sub-series. This plot helps to understand how the data varies within and across dayly groups.

- Statistical tests were carried out to check for differences among means and variances across days. No significant differences were found.

- **Forecasting experiments**: There is evidence for a weekly seasonal pattern from statistical tests. Yet, including information about this period in the forecasting model decreased its performance.

Seasonal Sub-series Plot (Monthly)

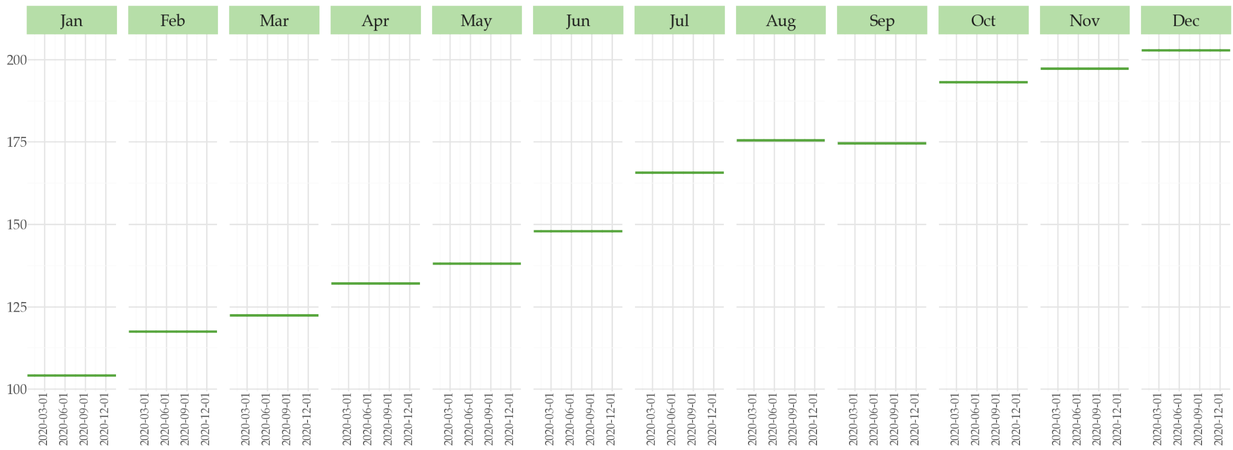

Figure 10: Monthly seasonal sub-series. This plot helps to understand how the data varies within and across monthly groups.

- The following tests indicate that the time series is non-stationary in seasonality for a monthly period: OCSB. On the other hand, other tests (Wang-Smith-Hyndman) fail to reject the stationary null hypothesis.

- Overall, there is a reasonable evidence that the time series is not stationary around a constant level. Within each group, there is also indication that the data is not constant around a level in 8% of the Months.

- **Forecasting experiments**: There is evidence for a monthly seasonal pattern from statistical tests. Besides, including information about this period in the forecasting model improved its performance.

Mean and Standard Deviation Analysis (Monthly)

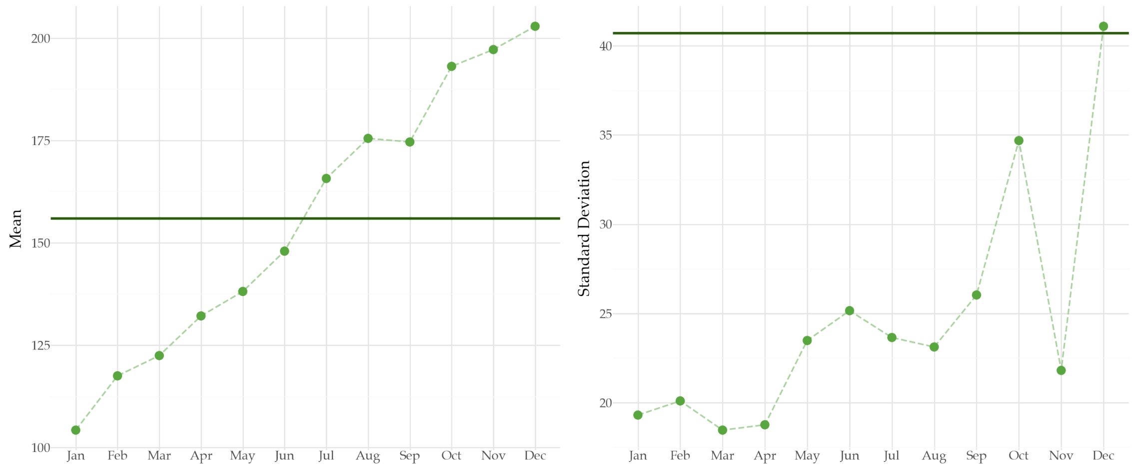

Figure 11: Mean plot (left) and standard deviation plot (right) of monthly values.

- The data shows significantly different mean and standard deviation when grouped by month.

Variance

Variance measures how data points spread around the average value in your time series. This section examines whether the variability remains stable (homoskedastic) or changes (heteroskedastic) over time. Understanding variance patterns is crucial for selecting appropriate modeling techniques, which can have a significant impact on forecasting accuracy.

Heteroskedasticity Testing

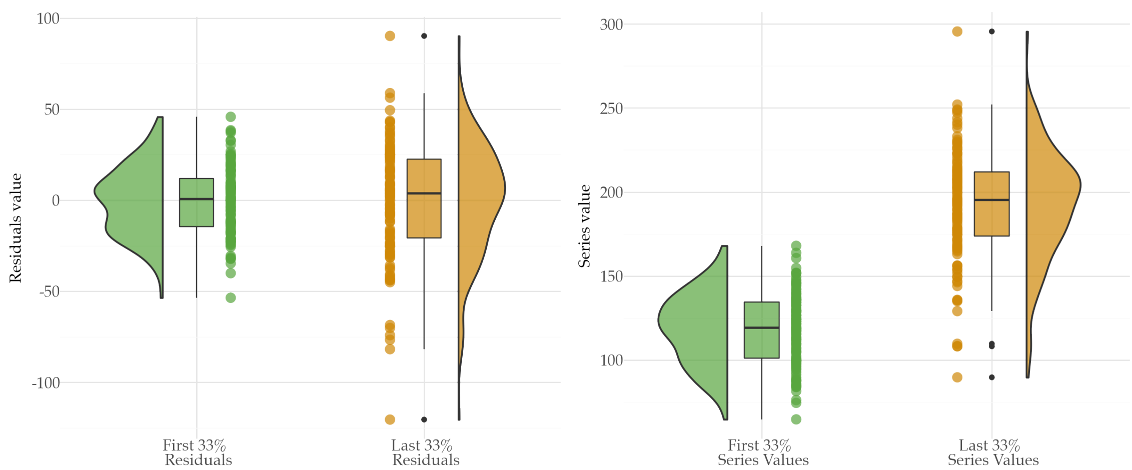

Figure 12: Time series residuals analysis. Difference in the distribution of the residuals (left) and the series (right) in the first and last thirds of the series, following a Goldfeld-Quand partition.

- As shown in the analysis of seasonality, there are significant differences in the dispersion of Monthly groups of observations.

- Compelling statistical evidence was found for the hypothesis that the time series is heteroskedastic, according to the White, Breusch-Pagan, and Goldfeld-Quandt tests.

- **Forecasting experiments**: Transforming the series with either the logarithm or the Box-Cox method did not improve forecasting performance.

- In the original scale, the Logarithm distribution is a reasonable fit to the data. But, after taking the Exponentially Modified Gaussian distribution transformation, a Logistic distribution was found to be a better fit.

Change Detection

Change points denote significant shifts in the underlying distribution of time series. These structural changes can manifest as sudden shifts in level, trend, variance, or seasonal patterns. Detecting and understanding these points is crucial as they often indicate important events or regime changes that affect modeling decisions. This section identifies potential change points and assesses their impact on the overall analysis strategy.

Change Points

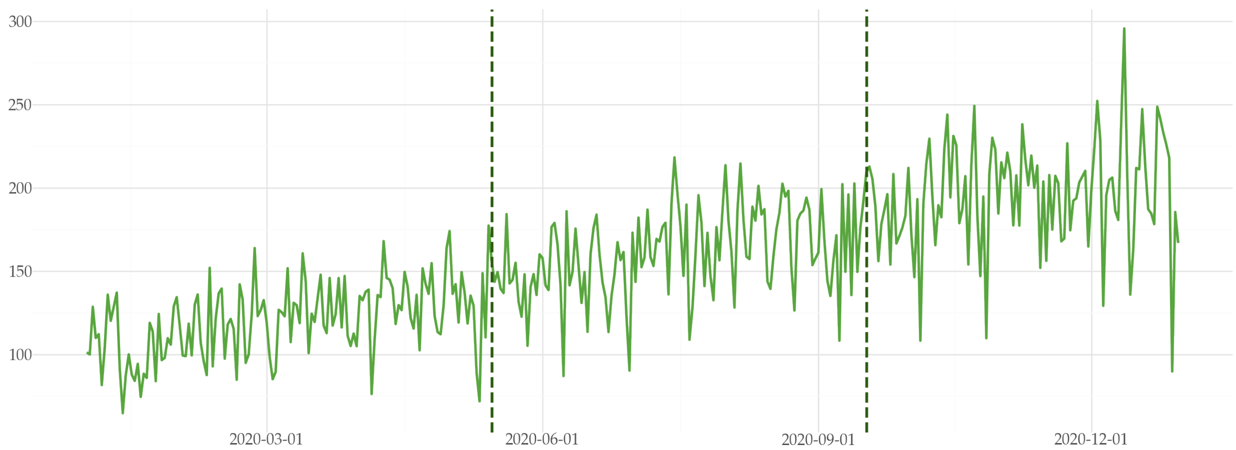

Figure 13: Time series plot with marked change points according to the PELT method.

- There are a total of 2 change points over the time series

- The first change point was found at 2020-05-15 where the time series shows an decreasing tendency.

Changes in Distribution

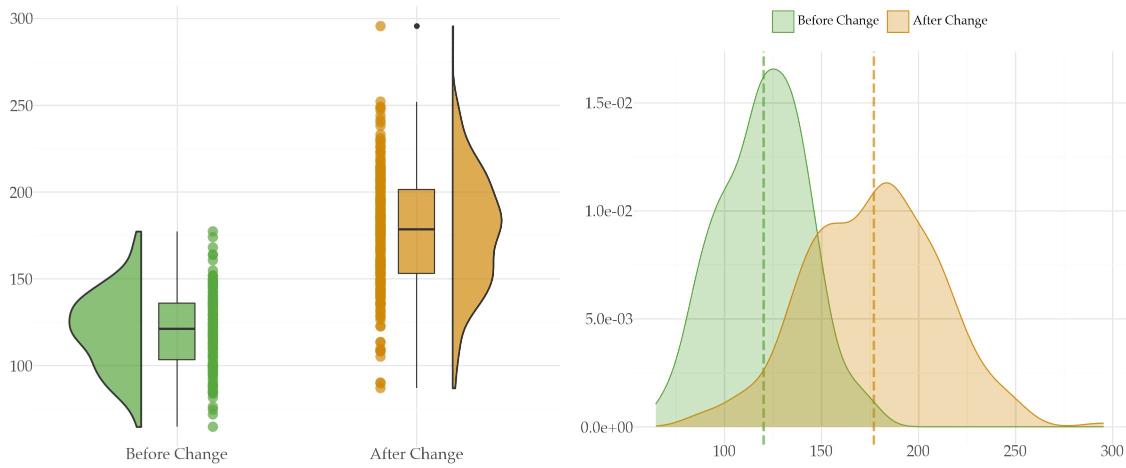

Figure 14: Time series analysis before and after the first detected change point occurs. Paired distributions before and after change (left), and overlapped density plot (right).

- The distribution before and after the first change point are significantly different.
- Before the change point, the data follows a Log-Normal distribution. But, after the first change point, a Gaussian distribution is a better fit.