# 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:** 2024-11-01 09:42

**Series Name:** M1009

## 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

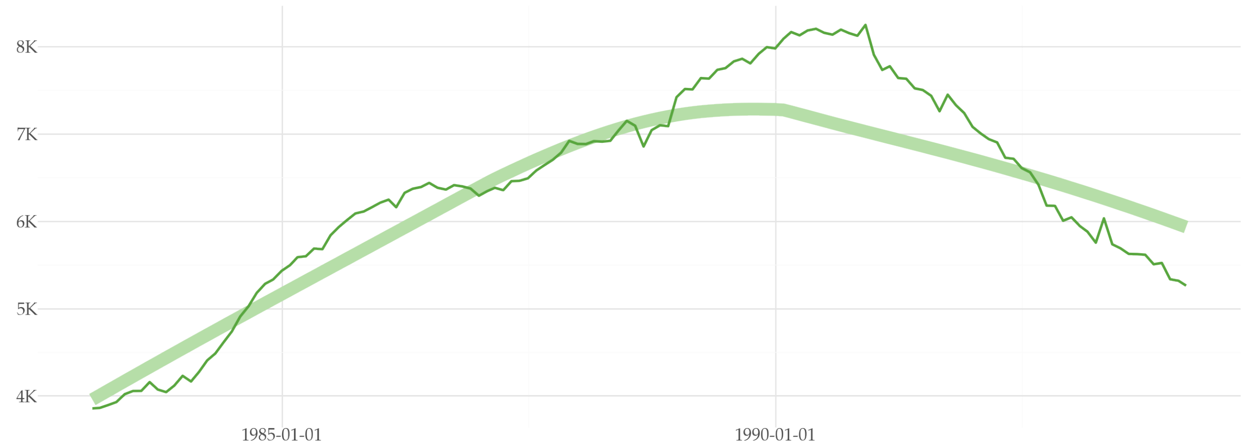


Figure 1: Time series line plot.

- A total of 134 month observations which span from January 1983 to February 1994.
- The mean value of the series is 6369.75 (median equal to 6404.6), with a standard deviation of 1206.69. The data ranges from a minimum of 3855 to a maximum of 8246.8.

## Data Distribution

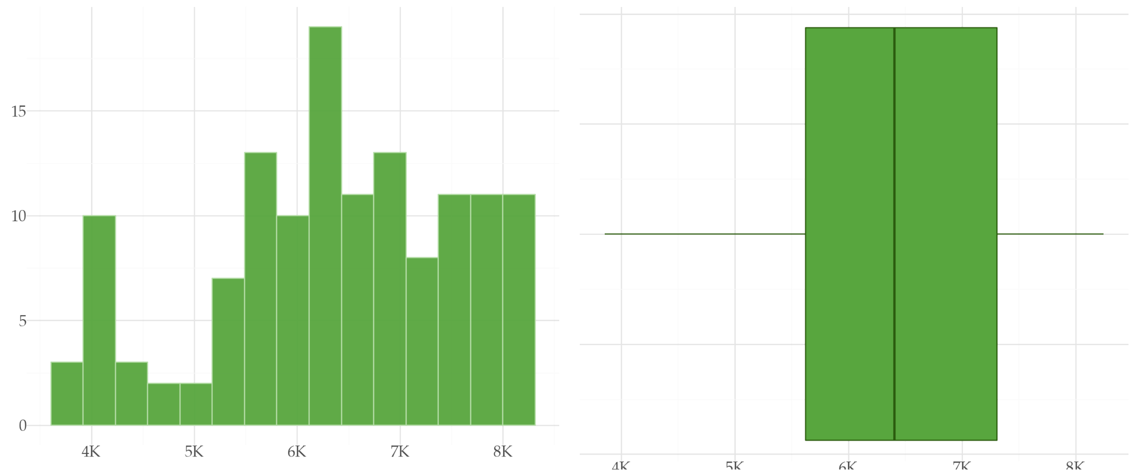


Figure 2: Distribution of the time series using an histogram (left) and a boxplot (right).

- The Kolmogorov-Smirnov test rejects the hypothesis that the series is distributed according to the following distributions: Cauchy, Exponential, and Pareto
- The distribution with largest p-value is Gaussian (p-value equal to 0.62). But, we cannot reject the hypothesis that the data follows the following distributions

(ordered decreasingly by p-value): Log-Normal, Exponentially Modified Gaussian distribution, Chi-squared, Logistic, Gamma, and Power-law.

- There are no outliers in the data set according to the boxplot representation.
- The skewness is equal to -0.42, indicates that the left tail is long relative to the right tail.
- The excess kurtosis is equal to -0.61. This value is similar to that found from data following a Gaussian distribution

## Trend and Seasonality

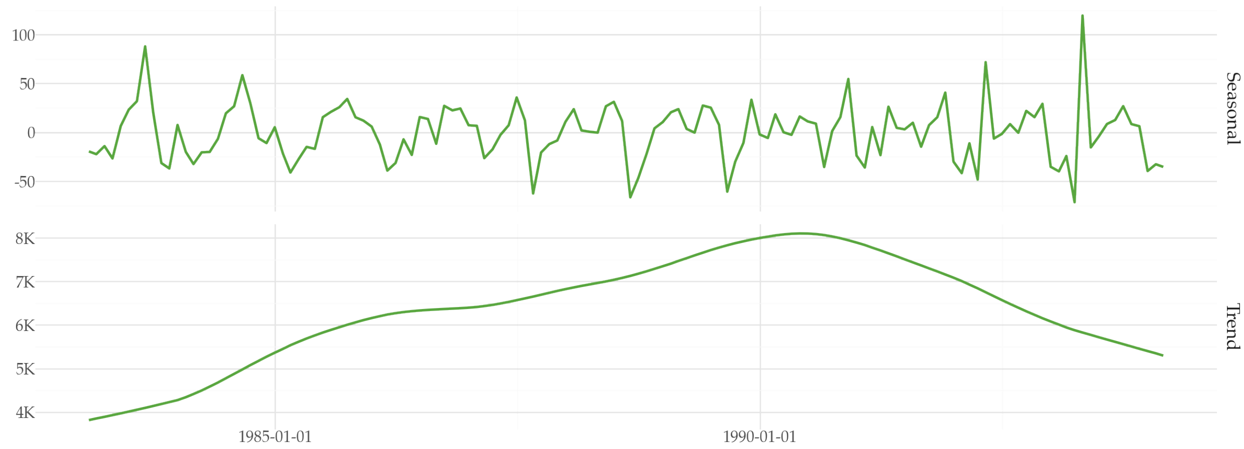


Figure 3: Seasonal and Trend components after decomposition using the STL (Season-Trend decomposition using LOESS) method.

- All hypothesis tests carried out (Augmented Dickey-Fuller and Philips-Perron) indicate that the time series is not stationary in trend/level.
- All hypothesis tests carried out (Wang-Smith-Hyndman and OCSB) indicate that the time series is not stationary in seasonality for the specified period.

## Auto-Correlation

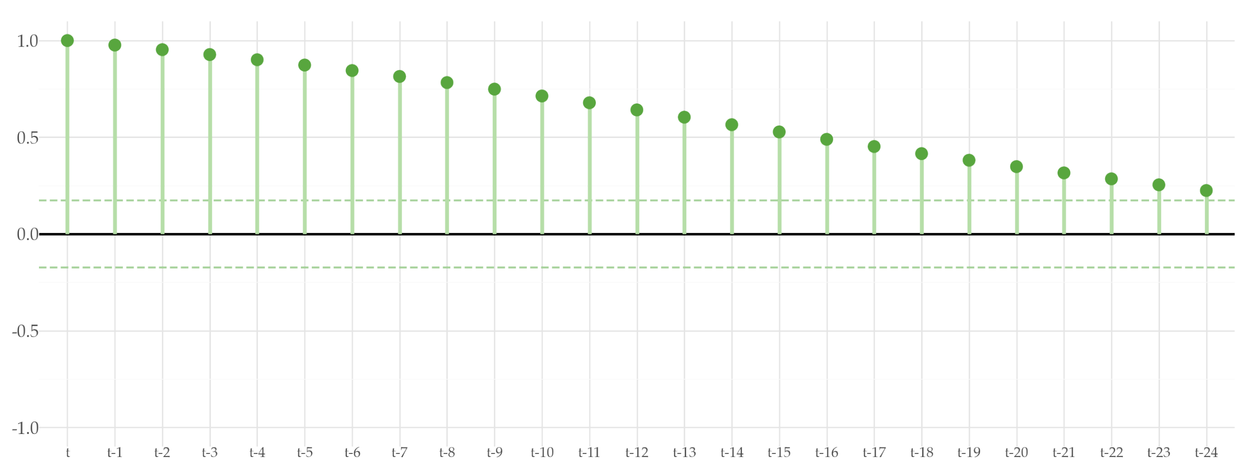

Figure 4: Auto-correlation plot up to 24 lags.

- 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, t-14, t-15, t-16, t-17, t-18, t-19, t-20, t-21, t-22, t-23, and t-24. The autocorrelation is positive for all lags with a significant value.
- None of the lags relative to the seasonal period (t-12 and t-24) show any significant autocorrelation.

## Partial Auto-Correlation

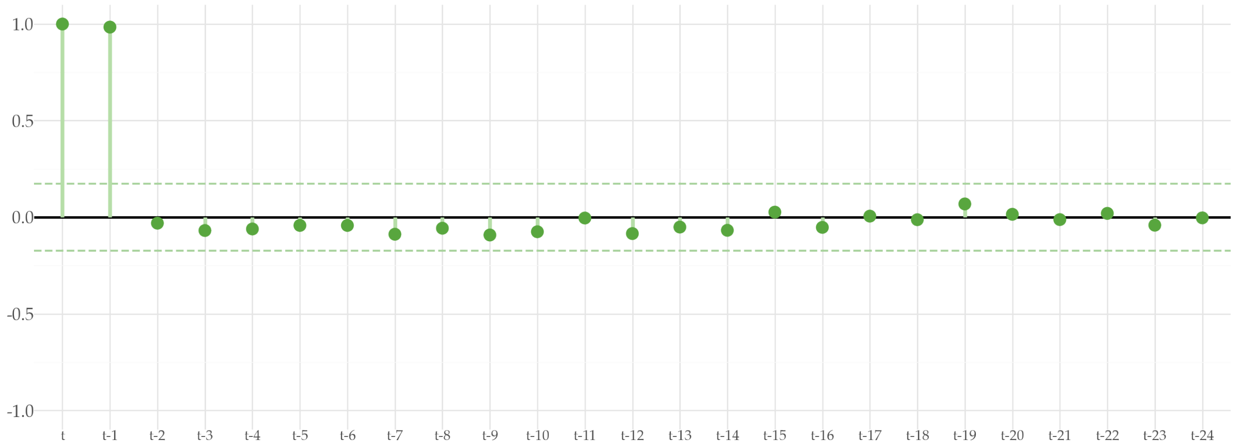


Figure 5: Partial Auto-correlation plot up to 24 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-1.
- None of the lags relative to the seasonal period (t-12 and t-24) show any 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

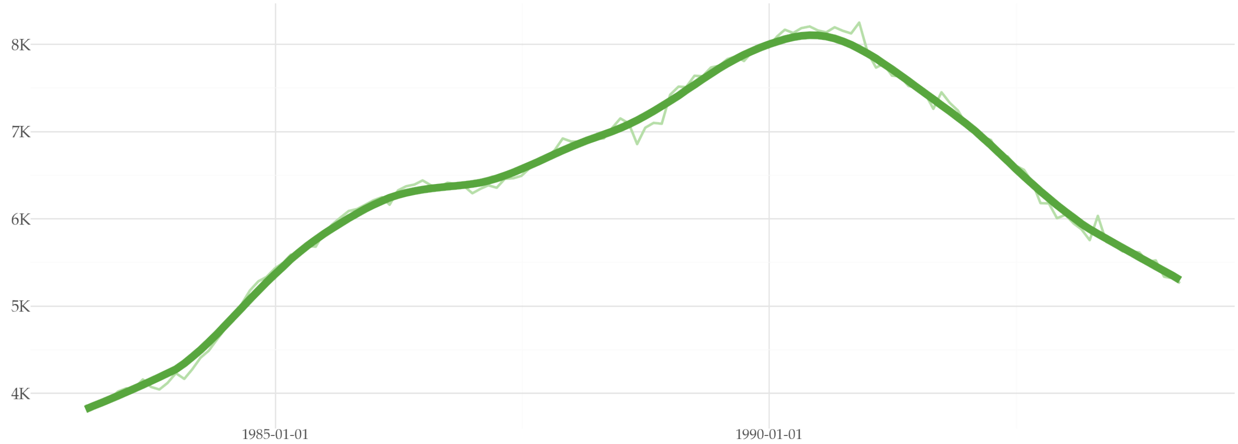


Figure 6: Time series trend plot.

- The time series has non-stationary trend according to the statistical test(s): Augmented Dickey-Fuller and Philips-Perron. There is a moderate upward trend.
- The same tests were applied to analyse whether the time series is stationary around a constant level. The method(s) Philips-Perron reject this hypothesis.
- Including a trend explanatory variable which denotes the position of each observation does not improve forecasting performance.

## Distribution of Differences

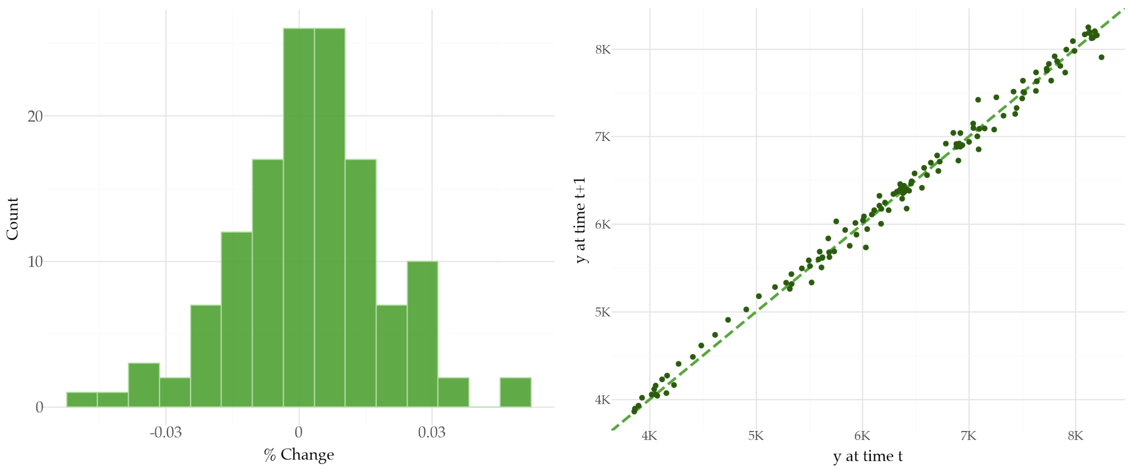


Figure 7: Distribution of percentage changes (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram shown the distribution of these changes. The lag-plot depicts the randomness in the data. The time series shows greater randomness 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: Power-law, Exponential, Pareto, Log-Normal, and Chi-squared.

- The distribution with largest p-value is Logistic (p-value equal to 0.9). But, we cannot reject the hypothesis that the differenced series follows the following distributions (ordered decreasingly by p-value): Exponentially Modified Gaussian distribution, Gaussian, Gamma, and Cauchy.
- The skewness is equal to -0.42, indicates that the left tail is long relative to the right tail.
- The excess kurtosis of the differenced series is equal to 1.16. This indicates that the data has a heavy tailed distribution.
- **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 (Monthly)

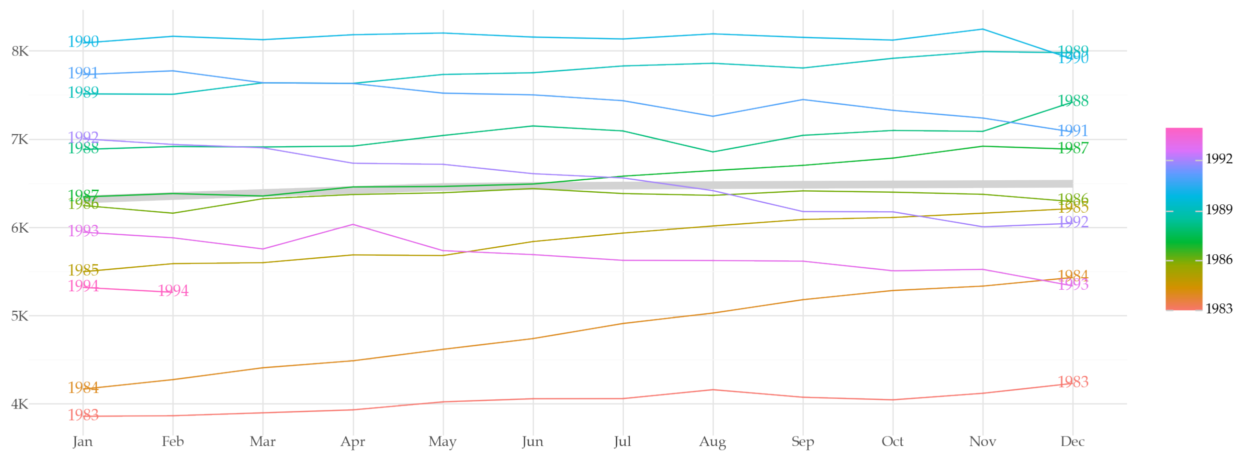


Figure 8: Seasonal plot of monthly values grouped by year.

- All hypothesis tests carried out (Wang-Smith-Hyndman and OCSB) indicate that the time series is stationary in seasonality for a yearly period.
- **Forecasting experiments**: Including monthly 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 Quarterly 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 months. No significant differences were found.

## 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

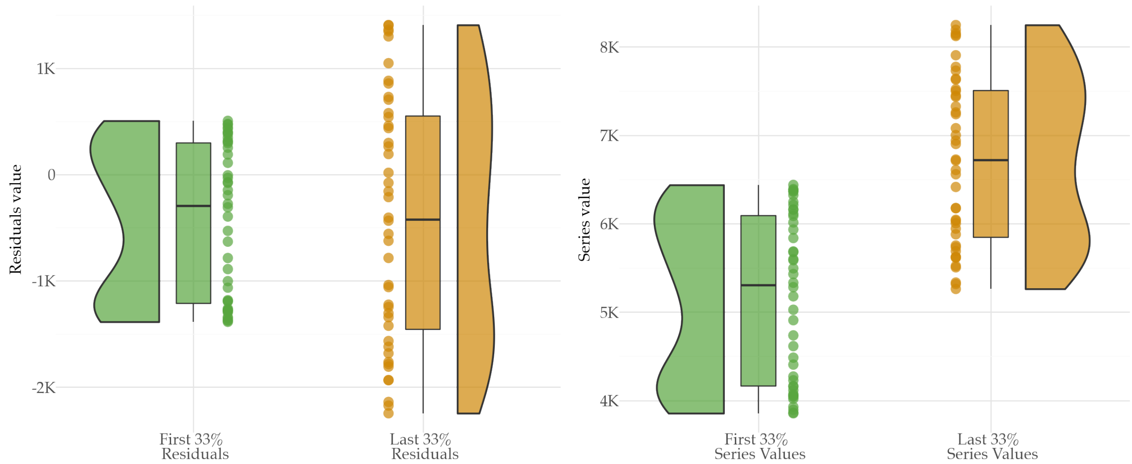


Figure 9: Time series residuals analysis. Difference in the distribution of the residuals (left) and the series (right) in the first and last thirds of the series, following a Goldfeld-Quand partition.

- In the analysis of seasonality we did not find significant differences in the dispersion among periodic 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 Gaussian transformation, a Power-law 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

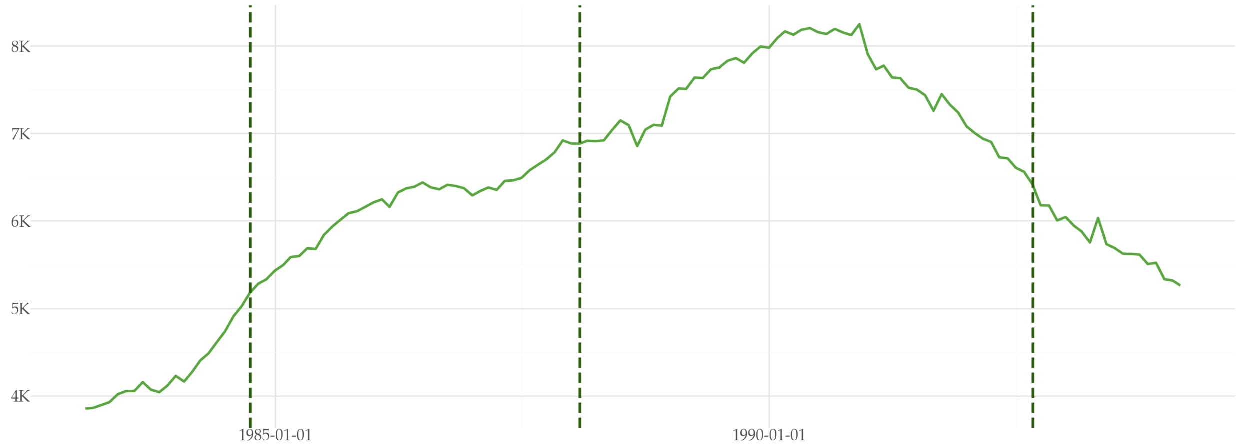


Figure 10: Time series plot with marked change points according to the PELT method.

- There are a total of 3 change points over the time series
- The first change point was found at September 1984 where the time series shows an increasing tendency.

## Changes in Distribution

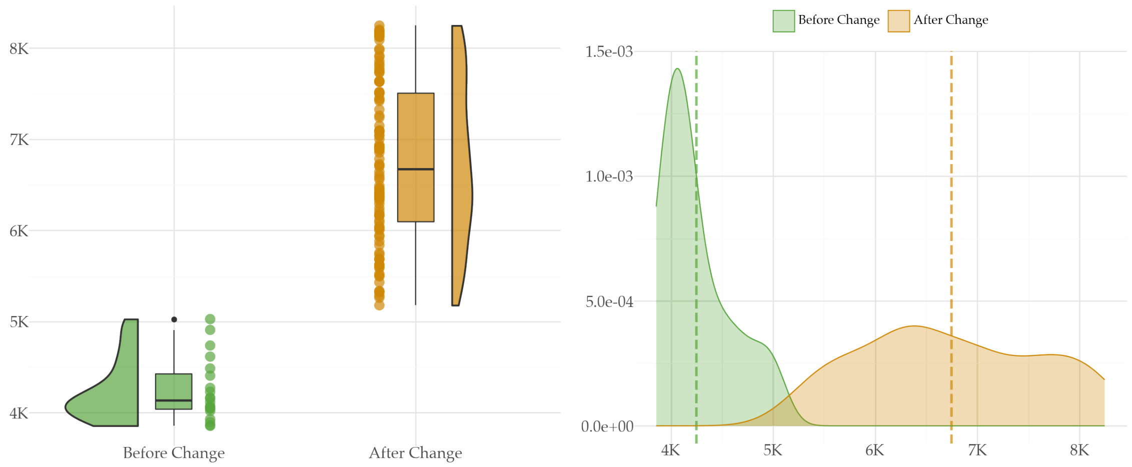


Figure 11: Time series analysis before and after the first detected change point occurs. Paired distributions before and after change (left), and overlapped density plot (right).

- Before the change point, the data follows a Log-Normal distribution. But, after the first change point, a Gamma distribution is a better fit.