Beer B1

Predicting Beer Production for the next month

Jussi Juvonen, Niko Laurén, Isuru Mulle Gamage

Time series exploration with visualization and decomposition of trend, seasonality, and residuals

The time series data represents monthly beer production in Australia, with no missing months or duplicate timestamps. The dataset spans from January 1956 to August 1995. Figure 1 shows a Seasonal-Trend decomposition using LOESS (STL) analysis of the series.

The observed data shows annual seasonality and a gradual long-term increase in production until the late 1970s, followed by a slight decline. The trend component captures the long-term pattern, and both robust and non-robust decompositions produce nearly identical trends, indicating stable estimates with minimal outlier influence. The seasonal component shows a consistent, repeating yearly pattern with roughly constant amplitude. The residuals fluctuate around zero, which means that most systemic variation is explained by the trend and seasonal terms.

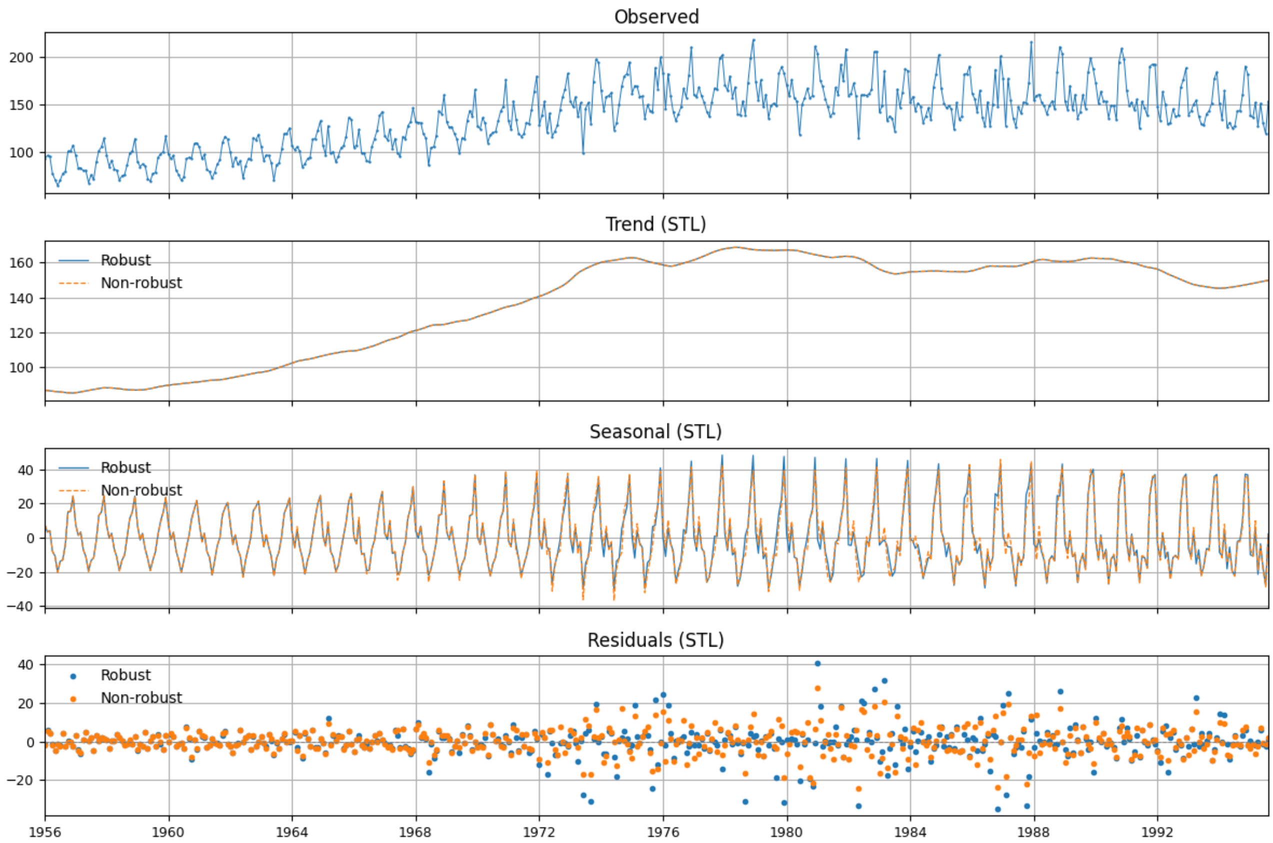

Figure 1: STL decomposition of monthly beer production time series showing the observed data, long-term trend, seasonal component, and residuals.

Figure 2 visualizes the seasonal component from the STL decomposition grouped by calendar month. The boxplots show a cyclical pattern in beer production, with the lowest seasonal values occurring during winter months (June - July) and the highest during summer (November - December). The relatively small spread of most boxes suggests that this seasonal pattern is stable across years, with only minor year-to-year variation.

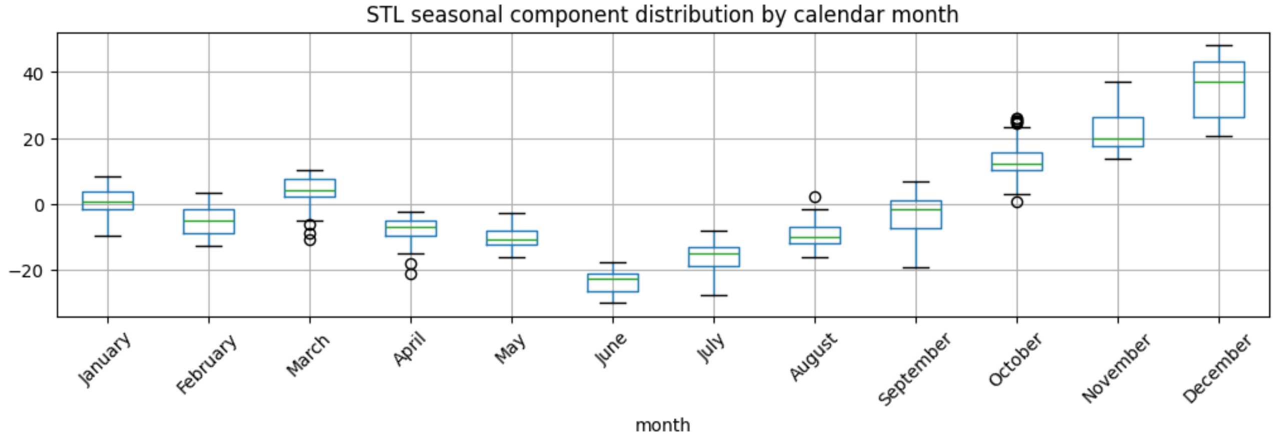

Figure 2: Distribution of the STL seasonal component by calendar month.

Autocorrelation analysis

Figure 3 shows the autocorrelation structure of the time series. The Autocorrelation (ACF) plot displays a strong positive correlation that gradually decays over many lags, with a noticeable seasonal pattern repeating approximately every 12 months, confirming yearly seasonality in the data. The Partial Autocorrelation (PACF) plot shows significant spikes at lag 1 and around lag 12. This suggests short-term dependence and annual cyclical effects. Together the plots confirm that beer production is highly autocorrelated and strongly seasonal over time.

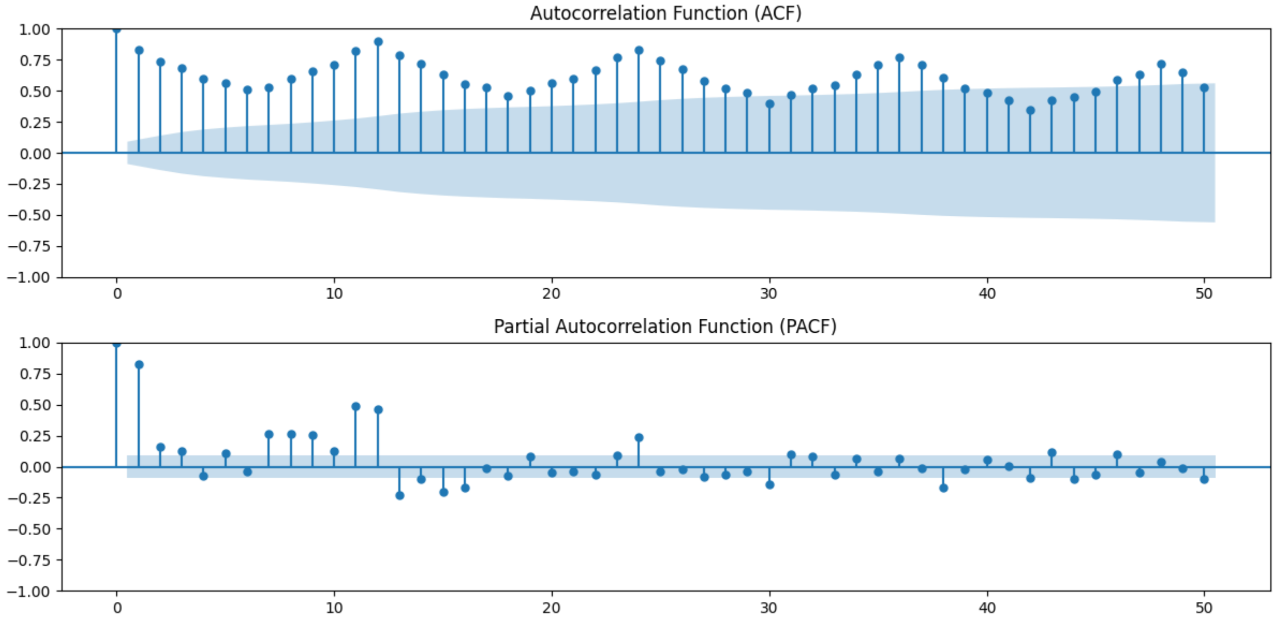

Figure 3: ACF and PACF plots of monthly beer production

Partitioning time series data

The dataset is partitioned chronologically into three subsets. The training subset is ~70% of the data covering period from January 1956 to September 1983 consisting of 333 samples. This subset is used to train the model and identify seasonal patterns in beer production. The validation subset is the following 71 samples or ~15% of the data from October 1983 to August 1989. The validation subset is used to adjust and fine-tune the parameters of the model. The remaining 72 samples (~15%) from September 1989 to August 1995 is for the test subset. The test subset is used to evaluate the model's predictive performance on unseen future data.