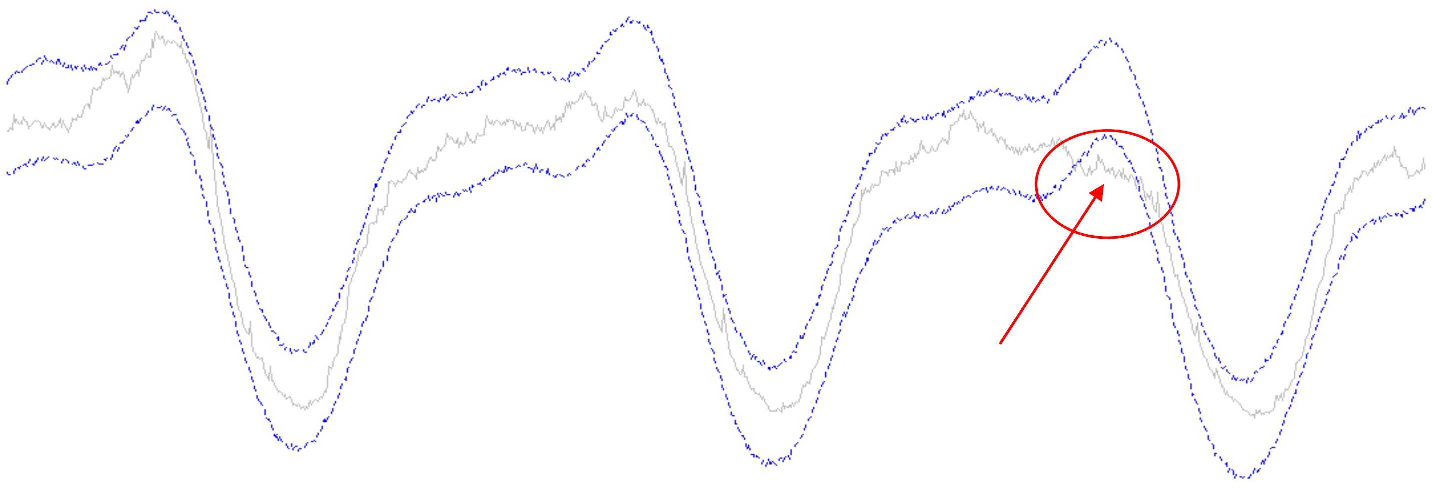
Time Series Anomaly detection

What is an anomaly/outlier?

Anomaly detection problem (for time series) can be formulated as finding outlier data points relative to some standard or usual signal



Types of outlier detection problem

- Supervised learning : This approach needs some labelled data, i.e., we already have examples in the data marked as outliers, which will help us building a model for the outlier detection
- Unsupervised learning : This approach needs only the raw data. Looking at the data, we create a generalized simplified version of the data and create bounds. Anything and everything that lies outside these bounds is an outlier.

Algorithms to solve an outlier detection problem

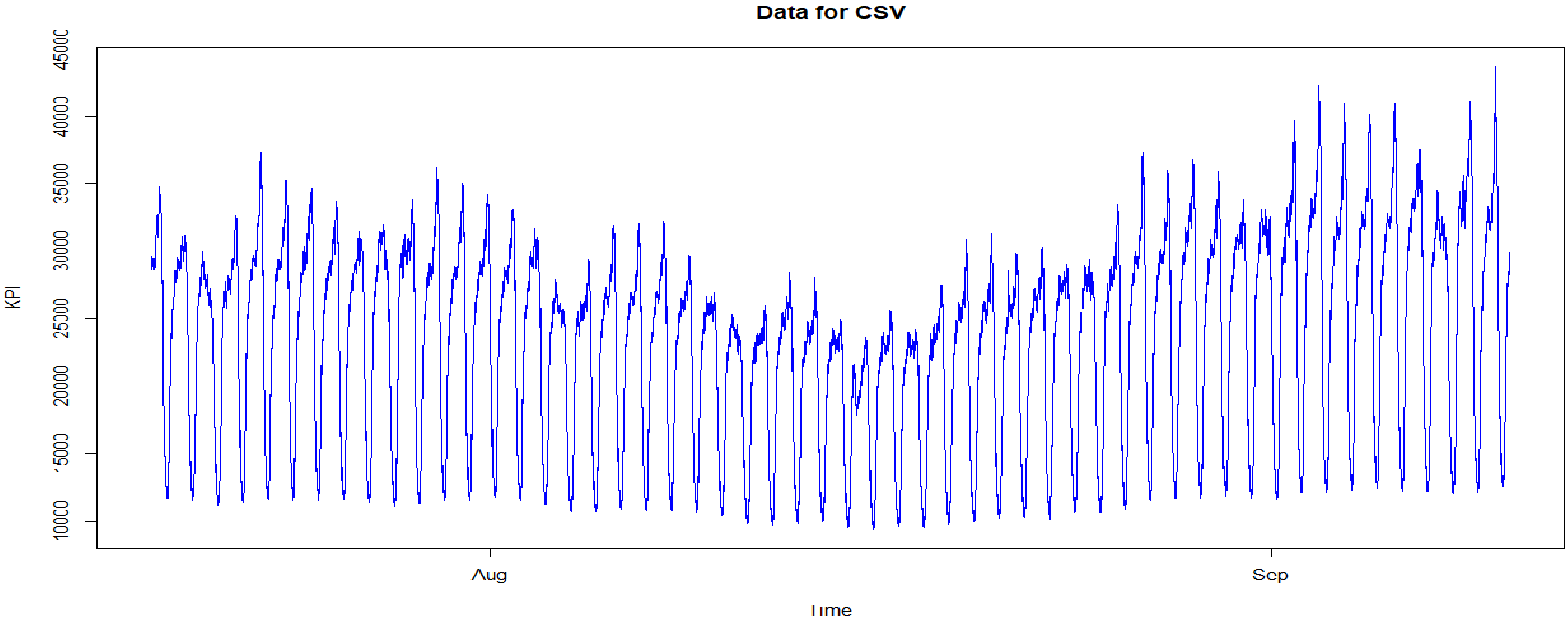
- ARIMA : *ARIMA is a very simple method by design, but still powerful enough to forecast signals and to find anomalies in it.*
- *Exponential Smoothing : Exponential smoothing techniques are very similar to the ARIMA approach. Uses exponential window function.*
- *CART : Classification and regression trees.*
- *LSTM : Long short term memory is a type of neural network (an RNN)*
- *Prophet : An open source library by Facebook*

This area is still ongoing research.

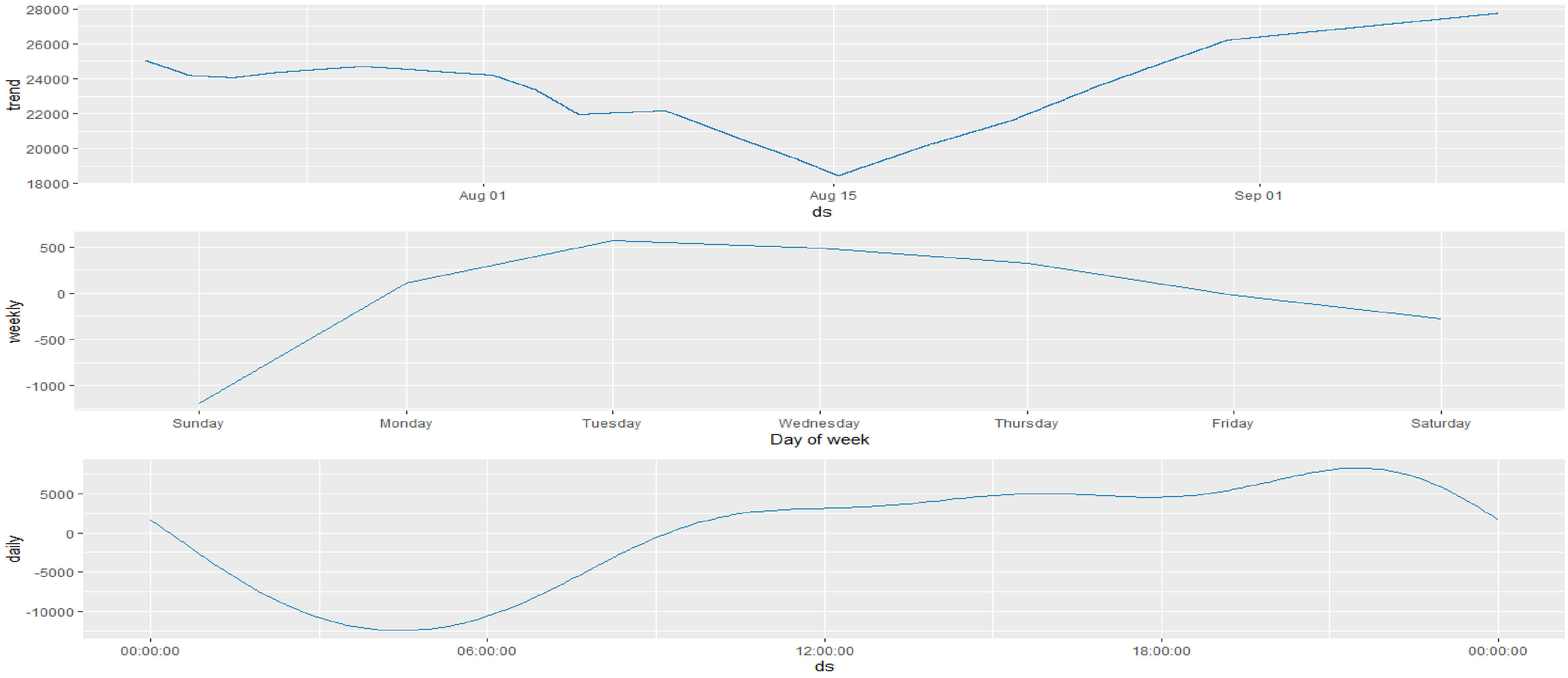
The process

1. Get and clean file
   a. Check for missing data and impute if needed.
   b. Add various other columns (Hour, minutes, weekday etc...) and make data ready for modelling
2. Fit the data to an appropriate model (that, hopefully, will describe the data at hand) and create the threshold bands.
3. Identify outliers based on whether they lie outside the threshold bands.

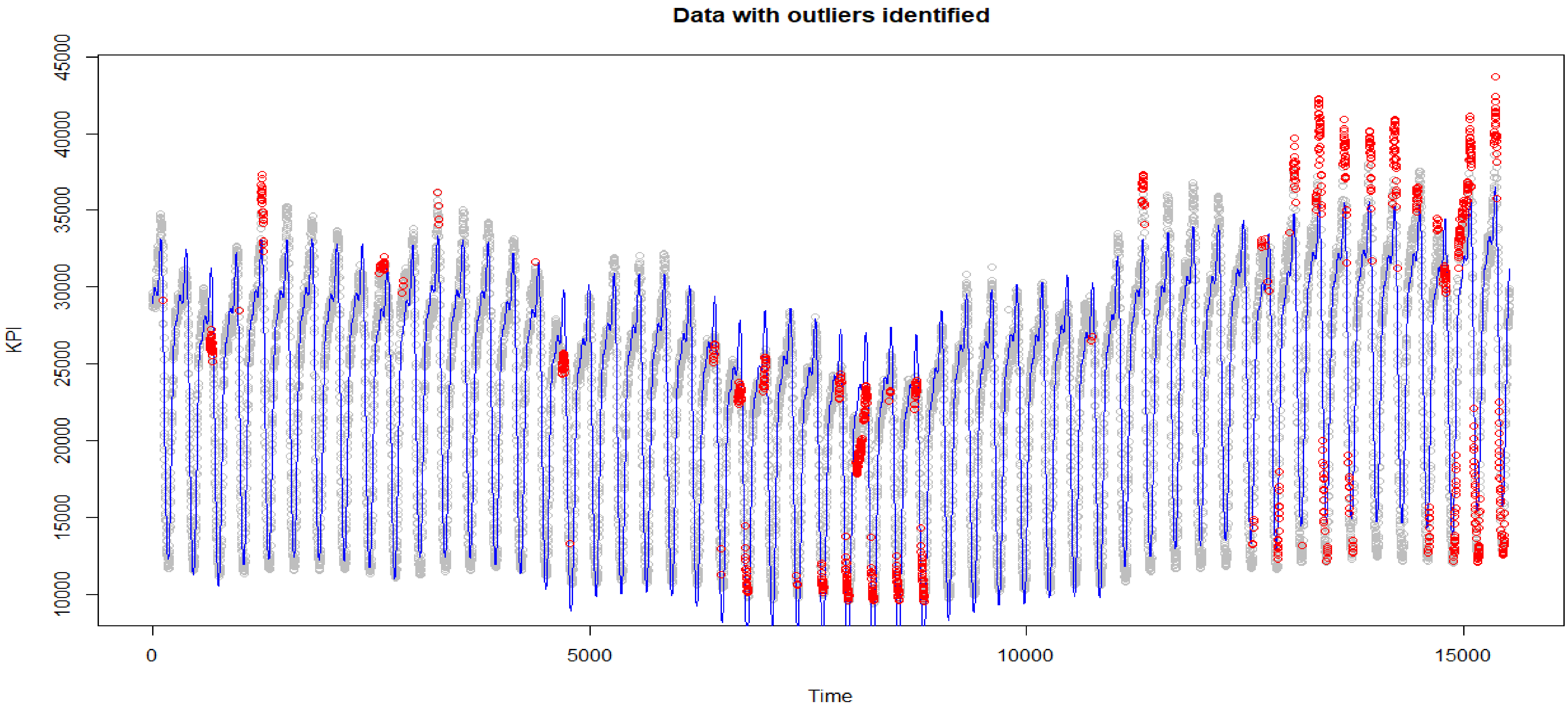
First set of data: Full data



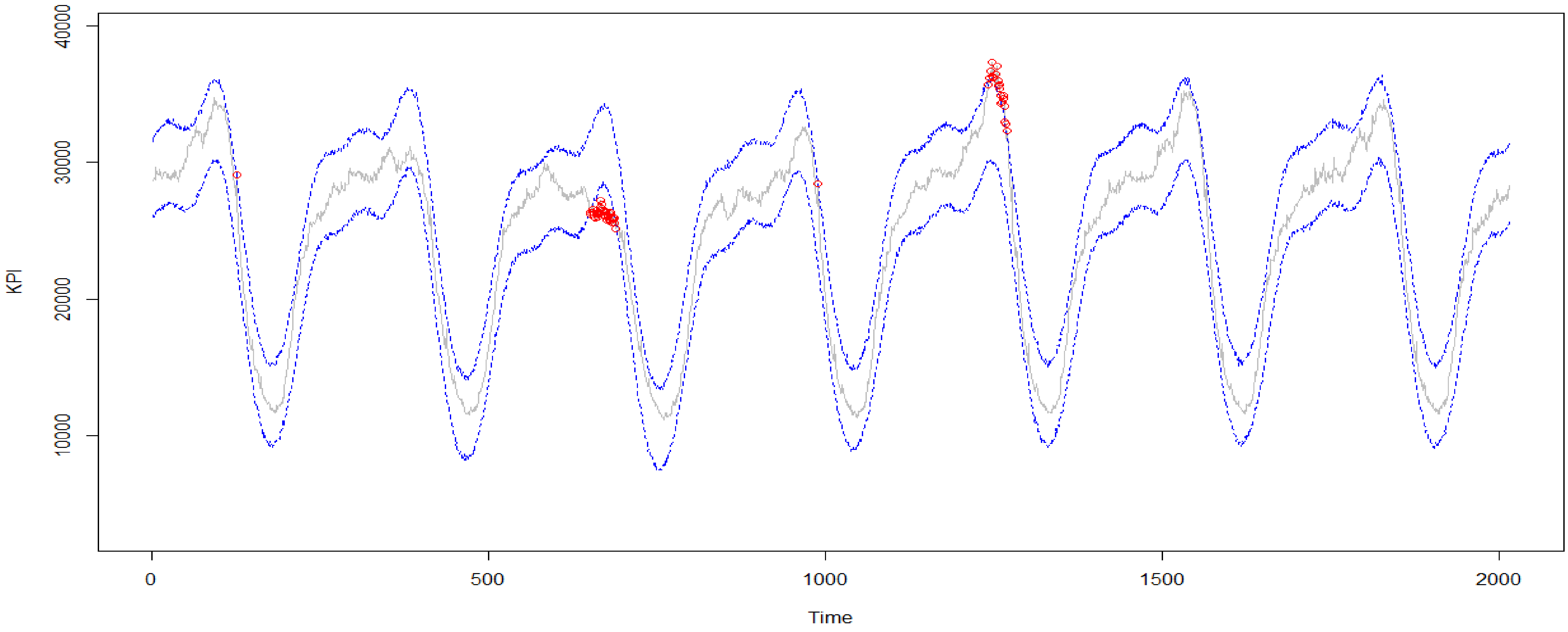
Seasonalities and Trend in the data



Full data with outliers identified



Zooming in to the first week



Model used

- The model used in order to achieve outlier detection was **Prophet** with weekly and daily seasonality with 95 % prediction bands.
- Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily (and sub daily) seasonality, plus holiday effects.
- It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.
- Other models such as ARIMA and exponential smoothing were also tried but they did not yield as much accuracy as Prophet did.

Types of outliers

- <u>Additive Outlier:</u> An additive outlier appears as a surprisingly large or small value occurring for a single observation. Subsequent observations are unaffected by an additive outlier. Consecutive additive outliers are typically referred to additive outlier patches.
- <u>Innovational Outlier:</u> An innovational outlier is characterized by an initial impact with effects lingering over subsequent observations. The influence of the outliers may increase as time proceeds.
- <u>Level Shift Outlier:</u> For a level shift, all observations appearing after the outlier move to a new level. In contrast to additive outliers, a level shift outlier affects many observations and has a permanent effect.
- <u>Transient Change Outlier:</u> Transient change outliers are similar to level shift outliers, but the effect of the outlier diminishes exponentially over the subsequent observations. Eventually, the series returns to its normal level.
- <u>Seasonal Additive Outlier:</u> A seasonal additive outlier appears as a surprisingly large or small value occurring repeatedly at regular intervals.
- <u>Local Trend Outlier:</u> A local trend outlier yields a general drift in the series caused by a pattern in the outliers after the onset of the initial outlier.