Bakery Sales

Matthew Edwards

2022-11-22

Summary

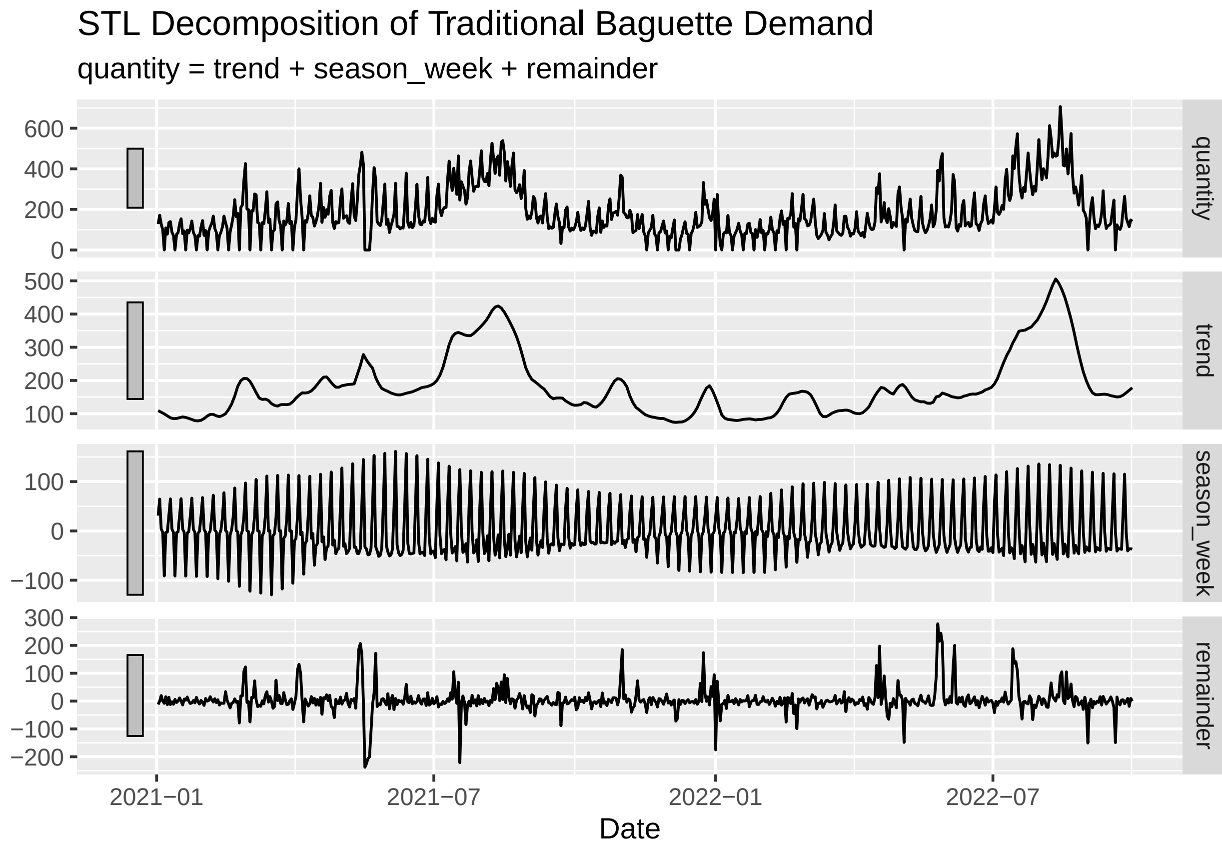
A bakery in France would like to produce the right number of products each day. Too few products would result in monetary loss and too many products would result in food waste. The baker has noticed that product demand is affected by weather. Therefore to produce the right number of products for each day the baker would like a forecasting model of product demand based on weather.

Data

There are two available data sets. The first is the `sales` data which includes the time and date each product was purchased along with the quantity and unit price. The second is the `weather` data which includes aggregate weather variables (e.g., maximum temperature) over the period the `sales` data was collected. The `sales` data suggests that there are approximately 149 different products. The five most popular products over this period are:

Article	Quantity
Traditional Baguette	117463
Croissant	29654
Pain Au Chocolat	25236
Coupe	23505
Banette	22732

For this project we will focus on prototyping a forecasting model of Traditional Baguette demand. The following figure displays an STL decomposition of the Traditional Baguette demand. It suggests non-stationary cycle-trend (upper middle panel) and seasonal (lower middle panel) components and heteroscedastic errors (lower panel). A forecasting model for Traditional Baguette demand will require exogenous variables to account for the non-stationary components. Accounting for the hetroscedastic errors is out of scope for this project.

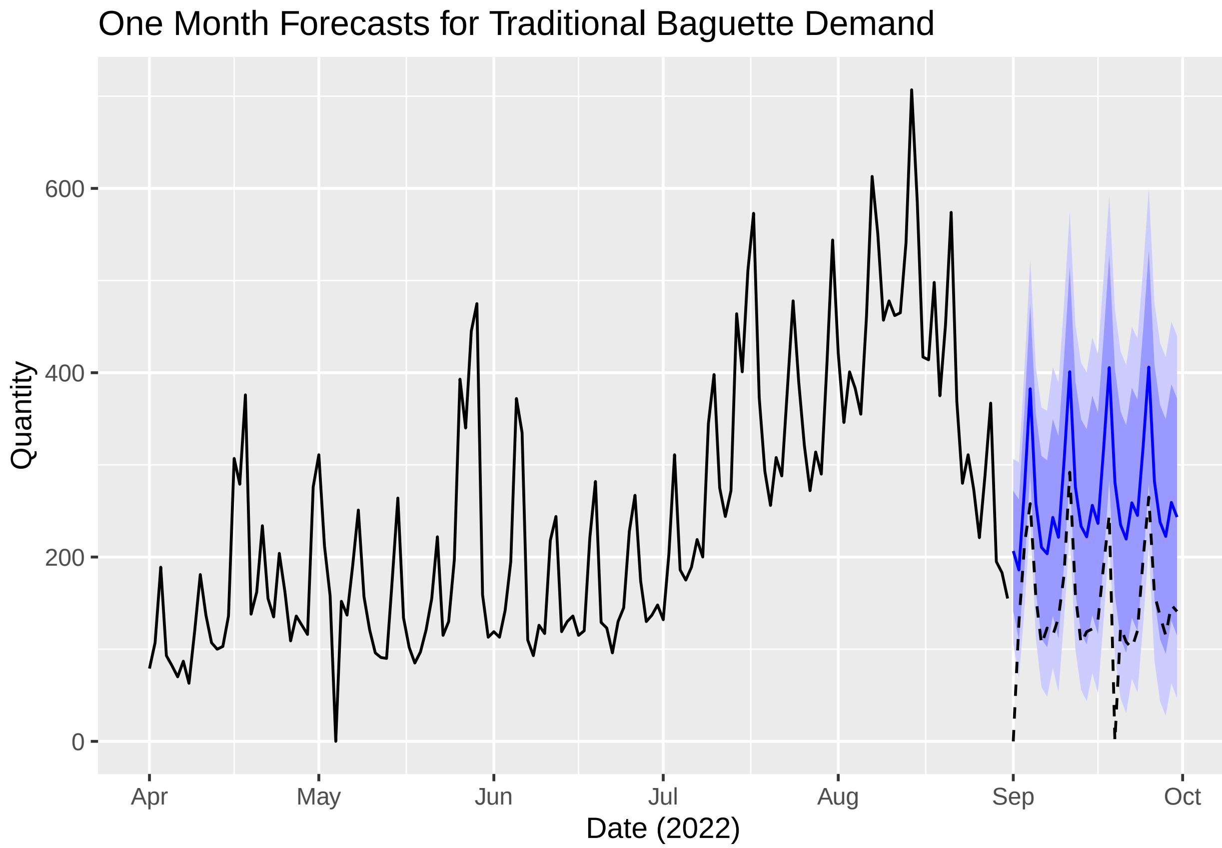


Model

To account for the non-stationary trend-cycle component the average temperature, wind speed and pressure variables where added as exogenous variables. To account for the non-stationary seasonal component weekly dummy variables where added as exogenous variables. Furthermore, the ARIMA model was used with both auto regressive and moving average components to account for temporal correlation. Residual diagnostics (non-included) suggest that most of the temporal correlation and non-stationary components are accounted for. However, there is still unaccounted for heteroscedasticity in the errors, which is out of scope.

Results

The following plot displays the last six months of data for the Traditional Baguette demand. The training data has a solid line and the test data has a dashed line. The predictions are provided in blue with 80% and 95% confidence intervals. The test data is mostly contained within the 95% confidence intervals. The predictions in the last month of the data set had a mean absolute error of approximately 122.



Conclusion

The prototype forecasting model for the Tradition Baguette demand is probably not accurate enough for production. We suggest that any future work should include extensive exploratory data analysis to gain a deeper understand the the mechanisms behind the non-stationary components and the hetroscedastic errors. Only then will a forecasting model be accurate enough for production.