# Tokyo's Bluefin Tuna Trends
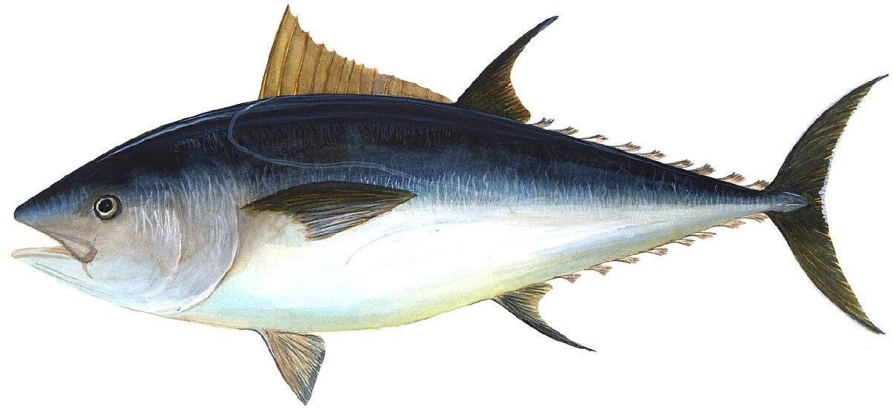
Data Source: Tsukiji Tuna Prices

*Laurynn Coleman, Justin Elwell, Thomas Cerreto, Romontae Thornhill, and Tammie Beckett*

# Overview of Data

- **Data in Focus:** A dataset on the wholesale prices of Bluefin Tuna in Tokyo's Tsukiji markets.
- **Timeframe and Granularity:** Monthly records spanning 2006–2016, capturing the ebbs and flows of market dynamics.
- **Forecasting Aim:** To apply time series forecasting methods to predict 2017–18 market trends for Bluefin Tuna.
- **Stakeholder Value:** The forecast will assist market players in making informed decisions regarding procurement, sales, and sustainability efforts.

# Introduction to the Data



Distribution of Bluefin Tuna Prices (Japanese Fleet)
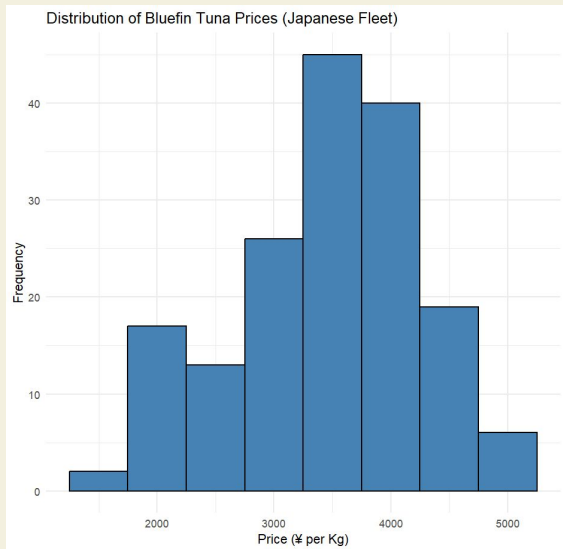
**Histogram of Bluefin Prices:**
indicates a left skewed distribution with a peak ~3,500 yen per kg. This suggests that while prices can range widely, a majority of transactions occur within this price bracket.



Monthly Price Variation of Bluefin Tuna (Japanese Fleet)
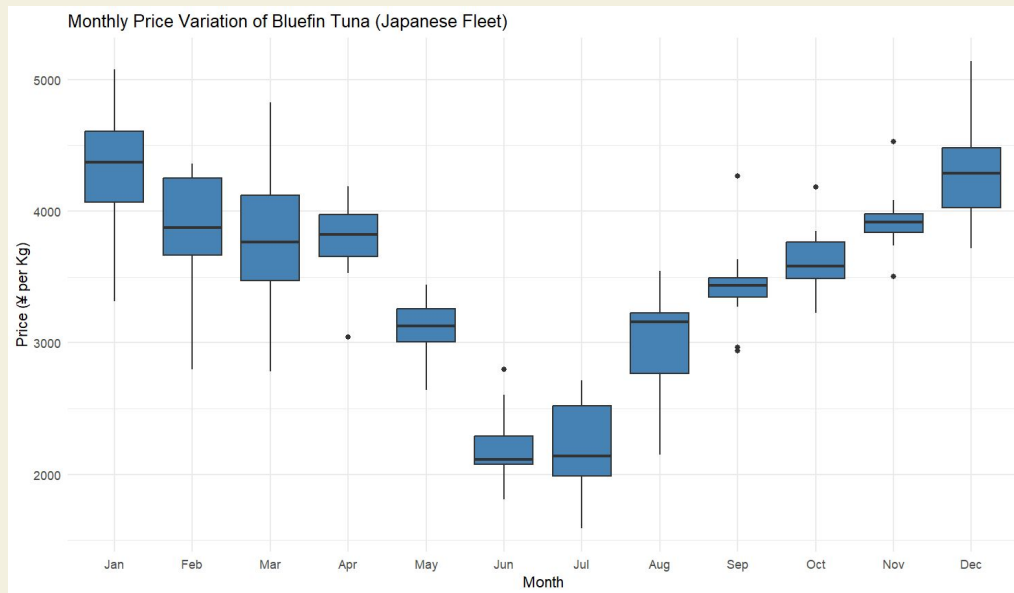
**Boxplot of Monthly Price Variation:** Prices tend to be higher and more volatile during the winter months (December & January to March), which could indicate increased demand or limited supply during this period.

Lower and more stable prices in the summer months (May to August) could suggest a seasonal increase in supply or possibly a lower demand.

# Bluefin Tuna Trends



Fresh Bluefin Tuna Caught by the Japanese Fleet
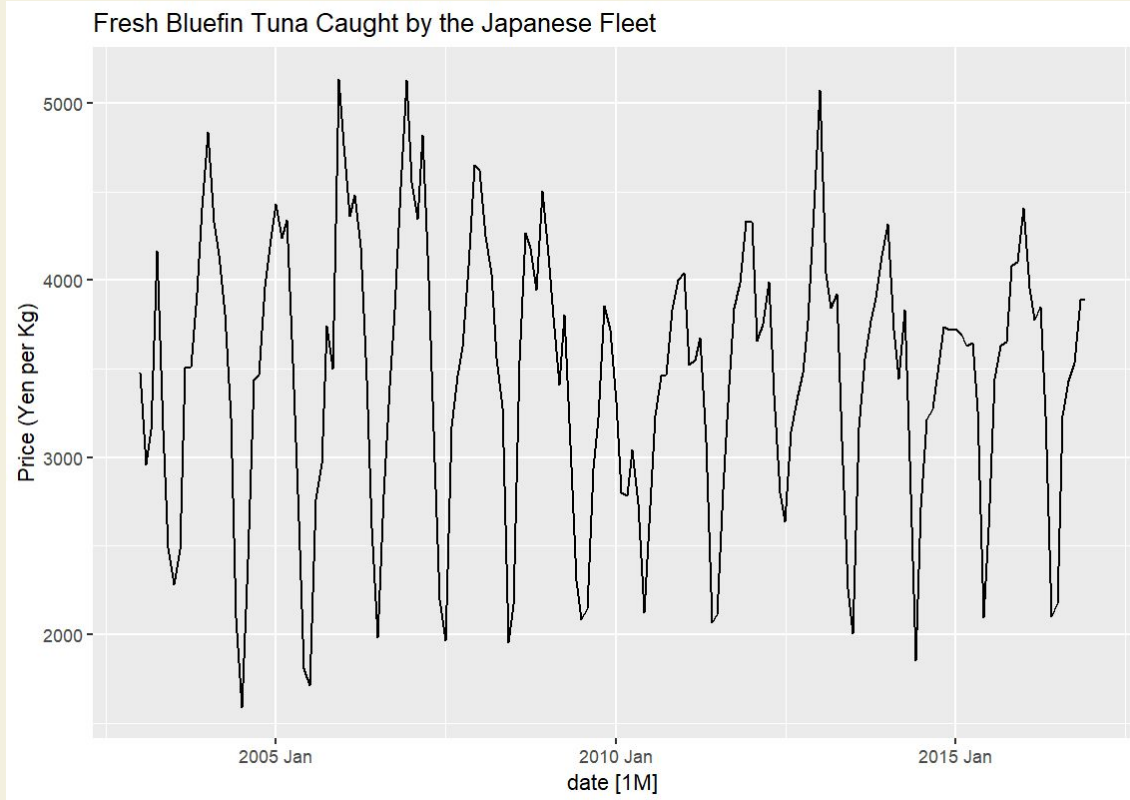
There seems to be a pattern of seasonality with the peaks and valleys, but let's look into it further.

# Bluefin Tuna Trends



STL decomposition
value = trend + season_year + remainder



Seasonal plot: Value of Bluefin Tuna
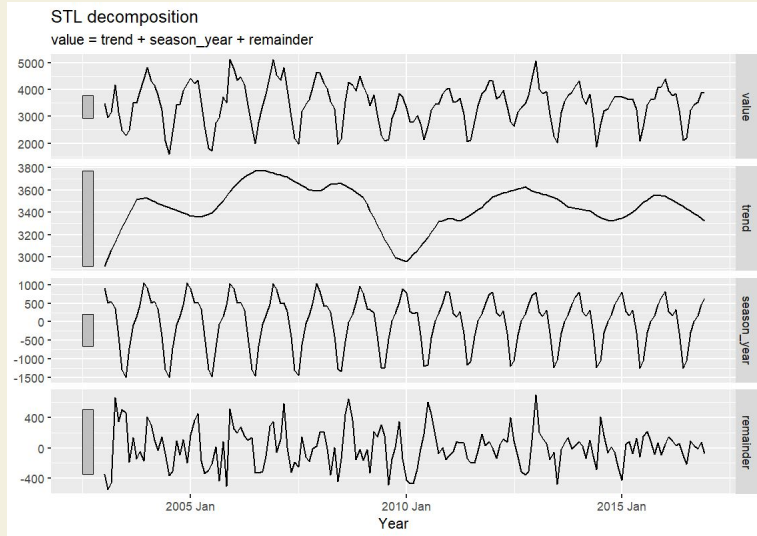
The trend line in the decomposition does not show a trend. There is one very large price decrease in 2009. The seasonal pattern (season_year) suggests multiplicative seasonality.
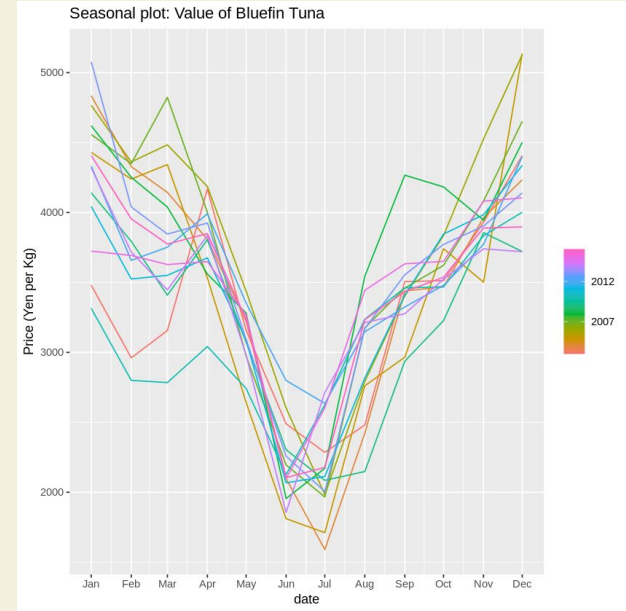
This shows a consistent seasonal pattern that may change in depth over the years. Overall, there is seasonality, which will affect what model we use.

# Model Selection

**Training Data Isolation:** Segregated all but the last two years of data (24 months) to form the training set for model building.

Data has monthly multiplicative seasonality, therefore is non-stationary.

- auto_ARIMA
- auto_ETS
- Log transformed TSLM

Dip in 2009 (shown in STL decomposition), therefore a recession indicator variable added.

# Forecast Output



Forecast on Test Set for Value of Fresh Bluefin Tuna
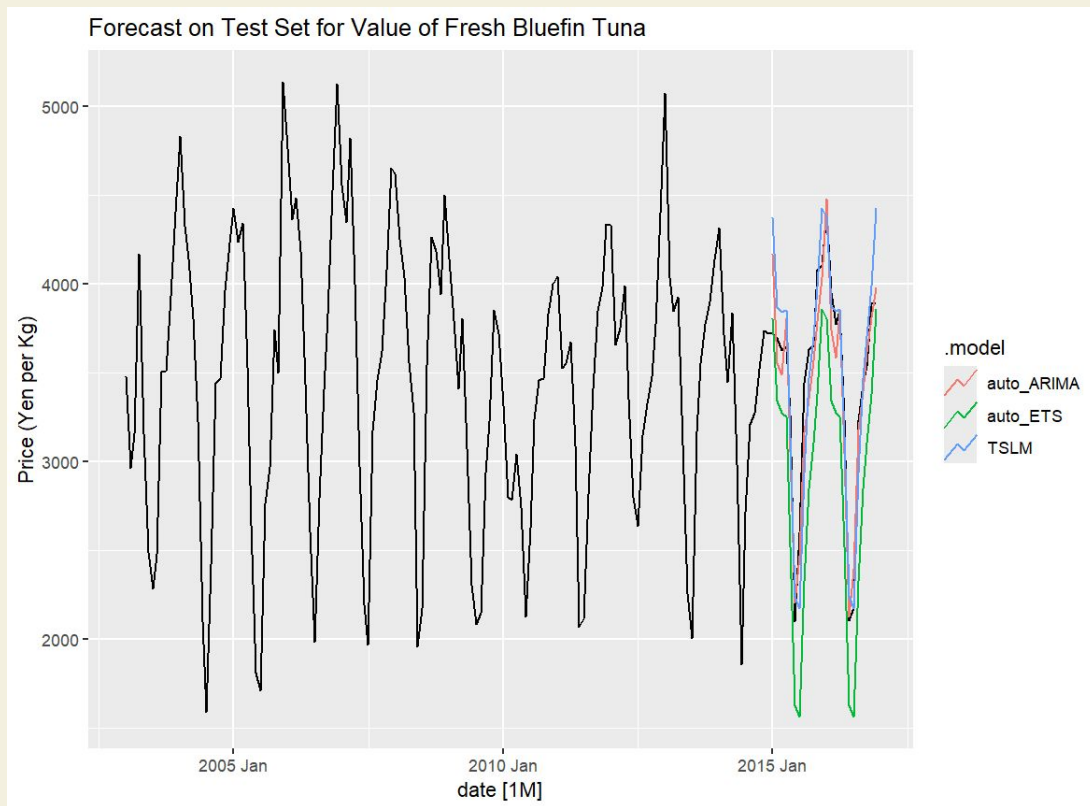
# Accuracy Measures

| | Models | RMSE | MAE | MAPE |
|---|---|---|---|---|
| #1 | auto_ARIMA (1,0,0)(2,1,0)$_{12}$ | 181. | 149. | 4.41 |
| #2 | TSLM | 260. | 194 | 5.73 |
| | auto_ETS (ANA) | 615 | 560 | 17.2 |

Based on the lowest RMSE, MAE, and MAPE, models Auto–ARIMA and TSLM performed the best.

# Cross Validation: Auto_ARIMA vs TSLM

| | Models | RMSE | MAE | MAPE |
|---|---|---|---|---|
| #1 | auto_ARIMA | 278. | 216. | 6.98 |
| | TSLM | 430. | 302. | 9.53 |

**Cross validation:** the data set was split sequentially to train and test the models multiple times for stability.

Parameters: There are 168 available observations. We used 80 observations in the training period, stepping forward 3 observations at a time resulting in 30 different training periods.

The formula to calculate the number of training periods is:

$$Training\ Periods = \frac{Total\ Observations - Initial\ Training\ Set\ Size}{Step\ Size} + 1$$

# Model Overview: Auto-ARIMA

AR    I    MA

**Seasonal »**

(1,  0,  0)
(2,  1,  0) [12]
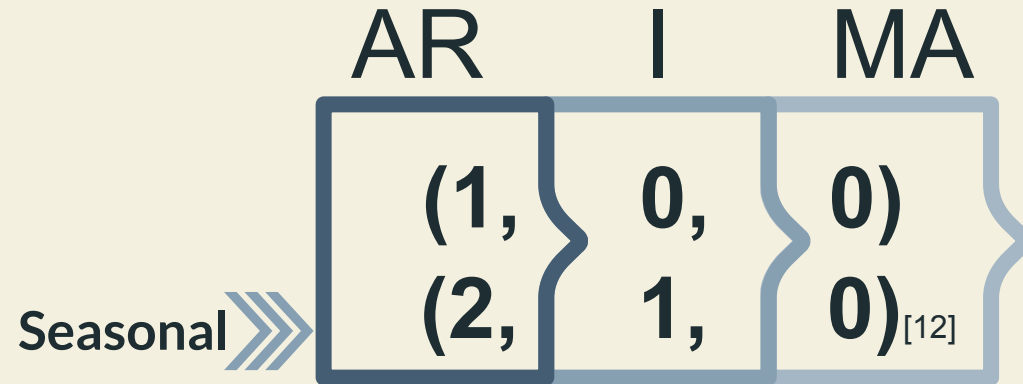
**AR (Autoregressive)**

- Models current values based on previous values.
- Uses a linear combination of past values ($p$ terms).

**I (Integrated)**

- Differencing raw observations to remove trends and seasonality.
- Denoted by $d$, represents the change from one observation to the next.

**MA (Moving Average)**

- Models current values as a function of past forecast errors.
- Number of terms ($q$) shows how past errors influence predictions.

# Residual Analysis

| Term | Est. | P-Value |
|------|------|---------|
| AR1 | 0.651 | 6.47e-19 |
| SAR1 | -0.558 | 2.96e-10 |
| SAR2 | -0.323 | 2.66e- 4 |

**AR1** suggests that a higher price in the previous period is associated with a higher price in the current period.
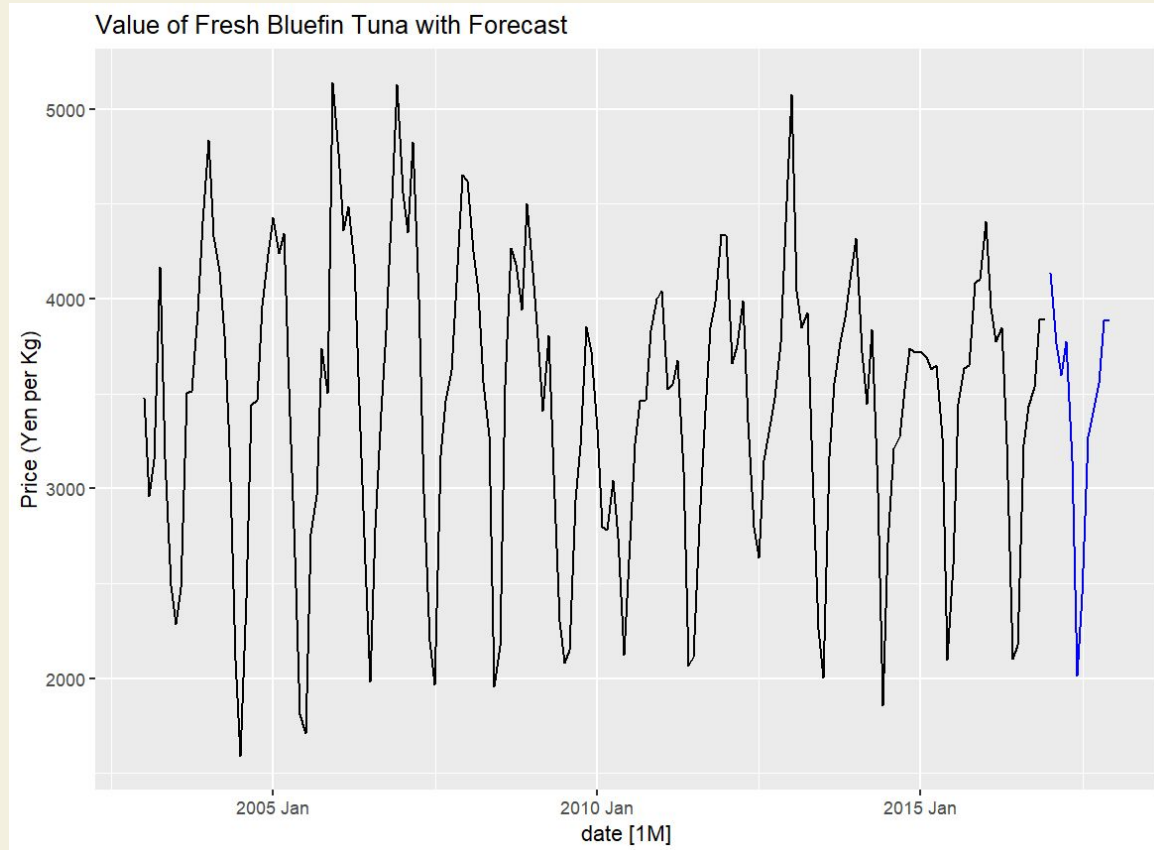
**SAR1** suggests that there is a negative relationship between the current value and the value from one season ago

**SAR1** suggests that the value from two seasons ago has a negative relationship with the current value.
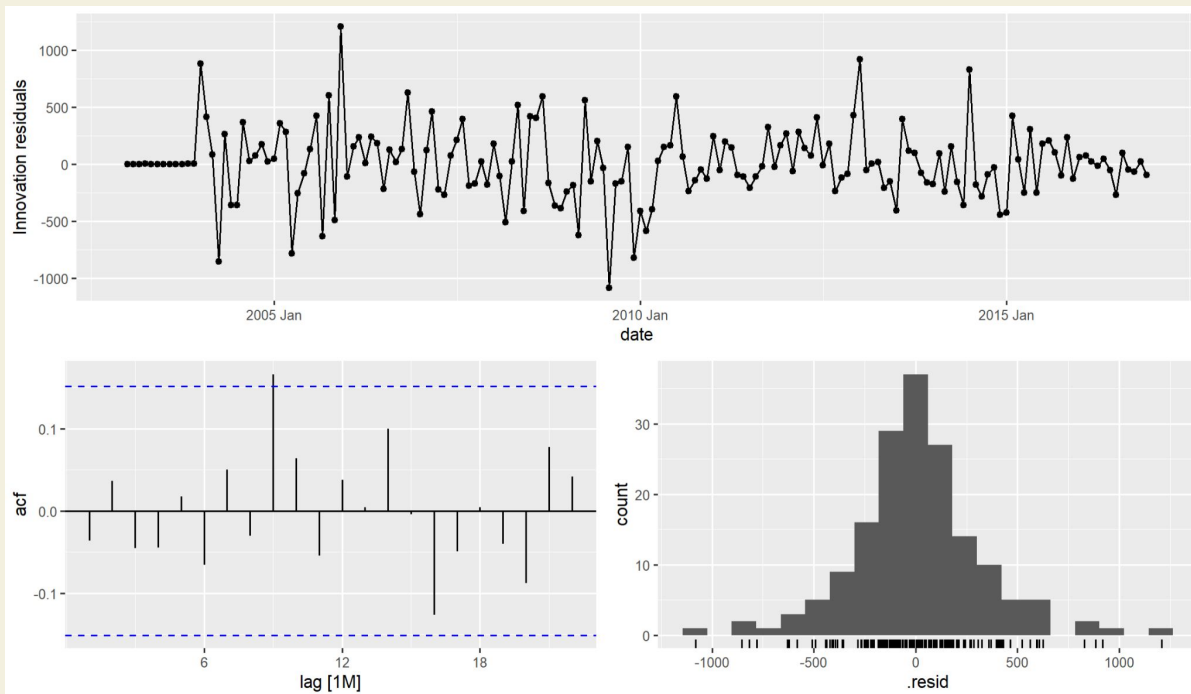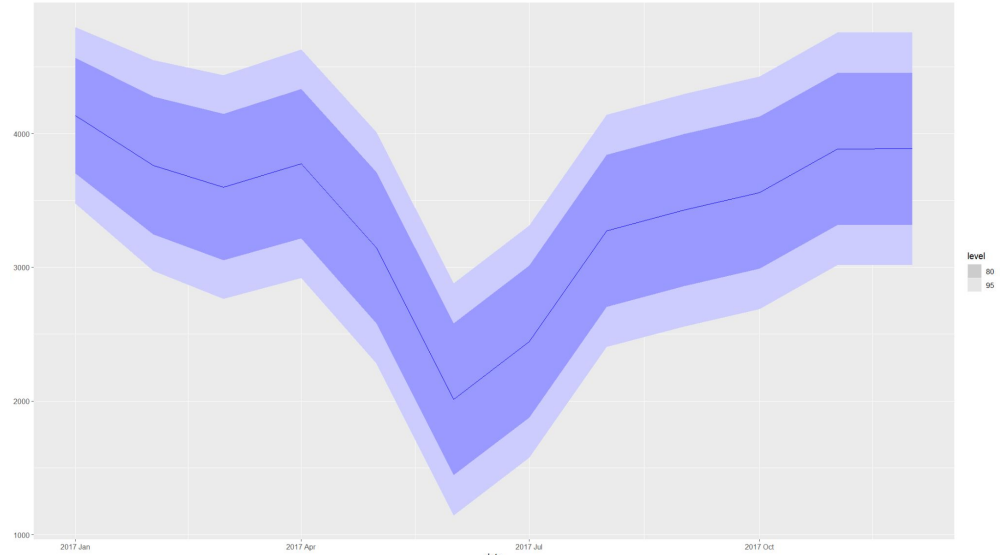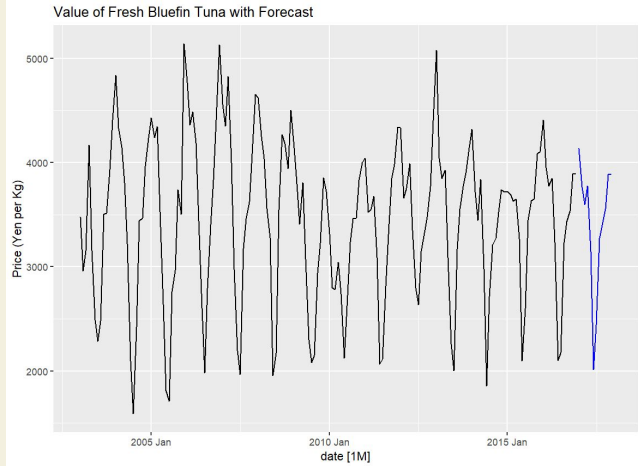


Residuals appear to be homoscedastic and normally distributed. There is one significant autocorrelation at lag 9, but it is not large enough to be a major concern and not in lags 1 or 2.

# The Final Forecast



Value of Fresh Bluefin Tuna with Forecast

# The Final Forecast



Value of Fresh Bluefin Tuna with Forecast

Thank You!