Time Series Project

Table of Contents

## 1. Data Exploration

* Describe your time series and the source of the series.
* Import the series. Note: Make sure you specify the correct starting time and frequency for the series.
* Plot the time series.
* Comment on the trend of the series (does the series have trend? upward or downward? from when to when, etc).
* Comment on the seasonality of the series (does the series have seasonal component? if so, what is the seasonal period, etc.)
* Comment on the seasonality of the series. Notice that trend or seasonal series is not stationary, being stationary implying having constant variance and constant mean.

### 1.1 Smoothing

* *Moving Average*: Explain the main idea of Moving average. Apply the MA method on the series. Plot the original series and its MA smoothing. Try a few different moving average and comment on the trend of the series through the smoothing curve. What values of best reveal the series pattern?
* *Exponential Smoothing* (ES): Explain the main idea of exponential smoothing. Apply the ES method on the series. Plot the original series and its ES smoothing. Try a few different weights comment on the trend of the series through the smoothing curve. What values of the weight best reveal the series’ pattern? Compare the ES curve and the MA curve above.

### 1.2 Decomposition

* *Classical Decomposition*. Explain the idea of Classical Decomposition in your own words. Apply the classical decomposition on the series for additive model and multiplicative model.
* *STL Decomposition*. Explain the idea of STL Decomposition in your own words. What is the main difference between the classical decomposition and STL Decomposition. Apply STL Decomposition on the series.

### 1.3 Auto-correltion

* Explain what the Auto-correlation Function (ACF) is and plot the ACF of the series. What is the first value of the ACF (ACF when the lag is 0) Is the series stationary according to the look of the ACF?
* Explain what the Partial Auto-correlation Function (PACF) is and plot the PACF of the series.

## 2. Modelling

In addition some baseline models, We will consider two models the HoltWinters and the ARIMA models. The baseline models are

* Mean/Average model: forecast by the average of the training series
* Naive Model: forecast by the last observation of the series
* Seasonal Naive Model: forecast by the values of the last season
* Random Walk with drift Model: Drawing the line from the first to the last observation

To evaluate the models, We will use a machine learning approach, which is use a portion of the data for training and the remainder data for testing. We also perform residual analysis for the models.

### 2.1 Model Training

Follow the sample codes to.

* Split the original series into the training series and the testing series
* Train all six models

### 2.2 Residual Analysis

A good model would have residuals looks like white-noise of a mean zero normal distribution, which means

* The residuals should have no autocorrelation.
* The residuals should have a mean zero
* The residuals should have constant variance
* The residuals should be normally distributed

Report the residuals analysis of all the models. Which model would you consider good models? Some indications of a good model could be

* The ACF is within the blue strip
* The distribution of the residual follows a bell curve
* The p-value of the Ljung-Box test is large

### 2.3 Testing Accuracy

* Calculate the forecast of all the models on the validation period. Calculate the Mean absolute percentage error (MAPE) of all the models. Plot the forecast of all models.
* Which model gives the lowest MAPE? Which model gives the greatest MAPE?

## 3. Forecasting

Based on the MAPE above, decide the best model. Retrain the best model on the entire series. Use the retrained model to forecast the next values. Plot the series and the forecast values.