

**Group Assignment**

**Time Series Forecasting- using R**

**Post-Graduate programme in Business Analytics & Business Intelligence**

**Batch-October 2019.**

**By: Group 4**

**Ashish Nath Km**

**Athul Aravind PM**

**Indu Mohan**

**Karthik R**

**Nishitha Ramesh**

**SOUVENIR SALES FORECAST CASE STUDY**

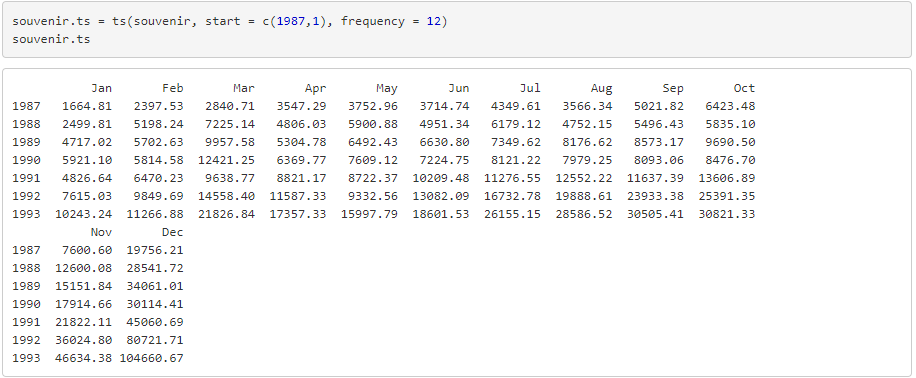
This case study attempts to forecast the Souvenir Sales for next 5 years that will help the Souvenir shops to identify any trend and seasonality in the sales.

For the analysis of the problem of Fancy dataset, we have utilised the below R-packages from the library.



The data “fancy.dat” is a txt file that consists of monthly sales of souvenirs for a souvenir shop at a beach resort town in Queensland, Australia for a time period – January 1987 to December 1993.

Using ‘scan’ function to read the data and then converting it to a timeseries object for future forecasts and model building



Few important terminologies from the point of view of analysing this case study are as follows.

**Time Series Forecast**

Time-series contains sequential data points mapped at a certain successive time duration, it incorporates the methods that attempt to surmise a time series in terms of understanding either the underlying concept of the data points in the time series or suggesting or making predictions.

Time series analysis is a method to analyse time series data so that we can derive a Statistical model and other characteristic of data. Time series forecasting make use of Time series statistical models which helps in predicting future time series values based on Historical observations.

**Exponential Smoothing**

Exponential smoothing is a time series forecasting method for univariate data.

Exponential smoothing forecasting methods make use of prediction in a weighted sum of past observations, but the model explicitly uses an exponentially decreasing weight for past observations. Specifically, past observations are weighted with a geometrically decreasing ratio

Exponential smoothing methods may be considered as peers and an alternative to the popular Box-Jenkins ARIMA class of methods for time series forecasting.

Exponential smoothing is usually used to make short term forecasts, as longer term forecasts using this technique can be quite unreliable.

* Simple (single) exponential smoothing uses a weighted moving average with exponentially decreasing weights.
* Holt’s trend-corrected double exponential smoothing is usually more reliable for handling data that shows [trends](https://www.statisticshowto.com/trend-analysis/), compared to the single procedure.
* Triple exponential smoothing (also called the Multiplicative Holt-Winters) is usually more reliable for parabolic trends or data that shows trends and [seasonality](https://www.statisticshowto.com/timeplot/" \l "seasonality).

**Holt-Winters forecasting**

Holt-Winters is a model of time series behavior. Forecasting always requires a model, and Holt-Winters is a way to model three aspects of the time series: a typical value (average), a slope (trend) over time, and a cyclical repeating pattern (seasonality). Holt-Winters uses exponential smoothing to encode lots of values from the past and use them to predict “typical” values for the present and future.

Holt-Winters is one of the most popular forecasting techniques for time series, where it’s used for purposes such as anomaly detection and capacity planning.

The three aspects of the time series behavior—value, trend, and seasonality—are expressed as three types of exponential smoothing, so Holt-Winters is called triple exponential smoothing. The model predicts a current or future value by computing the combined effects of these three influences. The model requires several parameters: one for each smoothing (ɑ, β, γ), the length of a season, and the number of periods in a season.

## Autoregressive Integrated Moving Average Model (ARIMA)

An ARIMA model is a class of statistical model for analyzing and forecasting time series data.

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a generalization of the simpler AutoRegressive Moving Average and adds the notion of integration.

This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

* **AR**: *Autoregression*. A model that uses the dependent relationship between an observation and some number of lagged observations.
* **I**: *Integrated*. The use of differencing of raw observations (i.e. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
* **MA**: *Moving Average*. A model that uses the dependency between an observation and residual errors from a moving average model applied to lagged observations.

Standard parameters (notations) of the ARIMA model are defined as follows:

* p: The number of lag observations included in the model, also called the lag order.
* d: The number of times that the raw observations are differenced, also called the degree of differencing.
* q: The size of the moving average window, also called the order of moving average.

**SARIMA**

Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component.

It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality

# **Mean absolute percentage error (MAPE)**

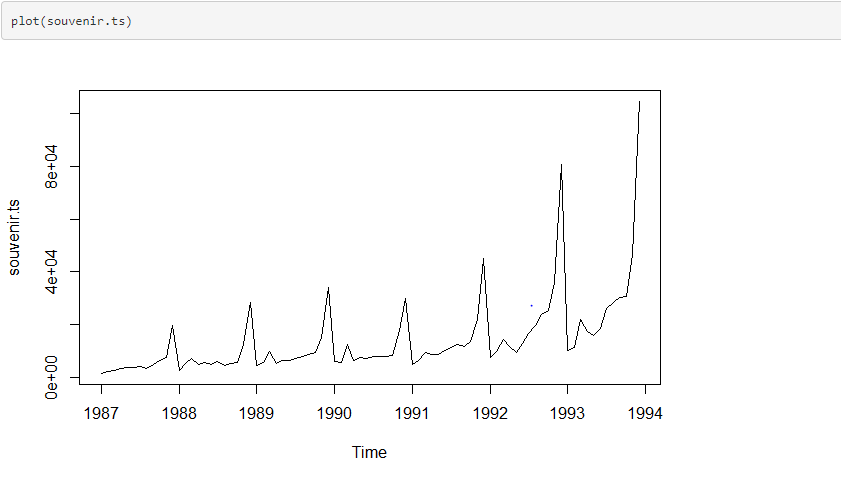
The mean absolute percentage error (MAPE) is a statistical measure of how accurate a forecast system is. It measures this accuracy as a percentage, and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values.

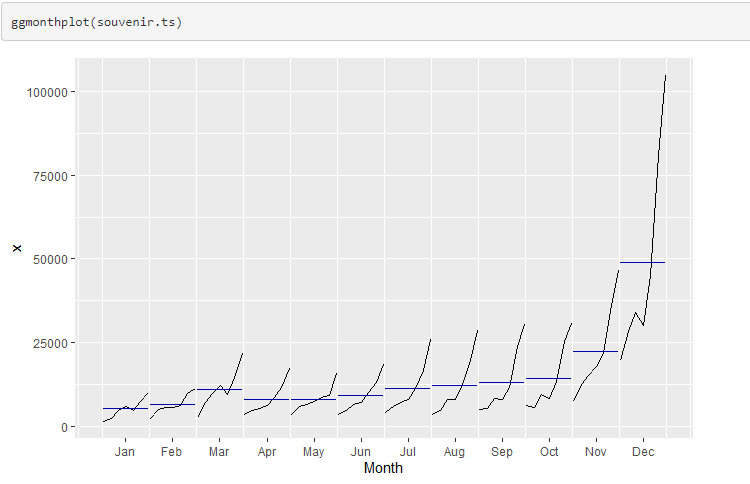
Where At is the actual value and Ft is the forecast value, this is given by:

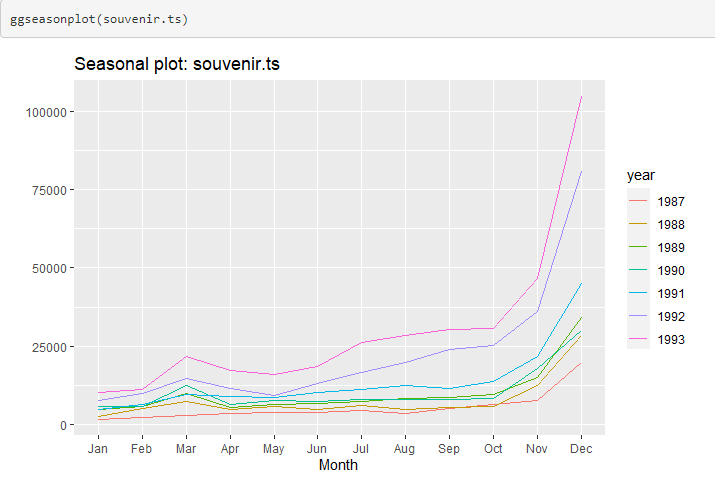
[](https://www.statisticshowto.com/wp-content/uploads/2017/09/mape.jpeg)  
  
  
The mean absolute percentage error (MAPE) is the most common measure used to forecast error, and works best if there are no extremes to the data (and no zeros).

1. **Visualizing the data and dividing the timeseries data into different components**

**Plot the data:**

****





**Observations:**

From the above plots, we can see that in this time series there seems to be seasonal variation in the sales in one year. The sales is high in the end of the year and almost steady for the first 10 months each year.

The seasonal sales is changing over time, and the random fluctuations also seem to be change in size over time.

There seems to be a clear upward trend and also a clear seasonality in the sales with higher sales in November and December, preceeded by a small spike in sales in late summer / early autumn (end of season sale perhaps).

**Decompose the timeseries data:**

A seasonal time series like above consist of trend component, a seasonal component and an irregular component.

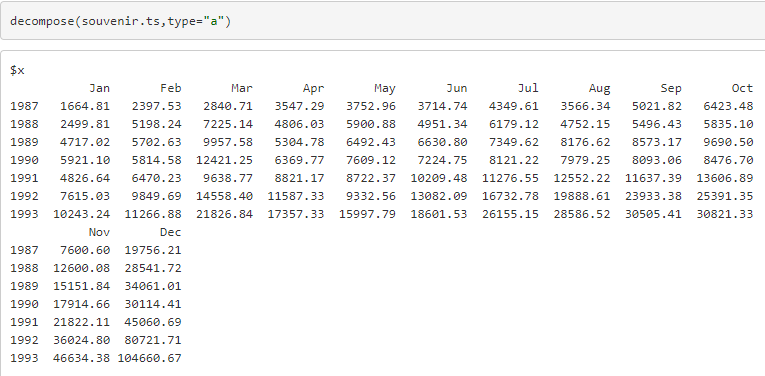
Decomposing time series means separating the time series into these three component, that is estimating the three component. To decompose seasonal data, use “decompose()” command in R.

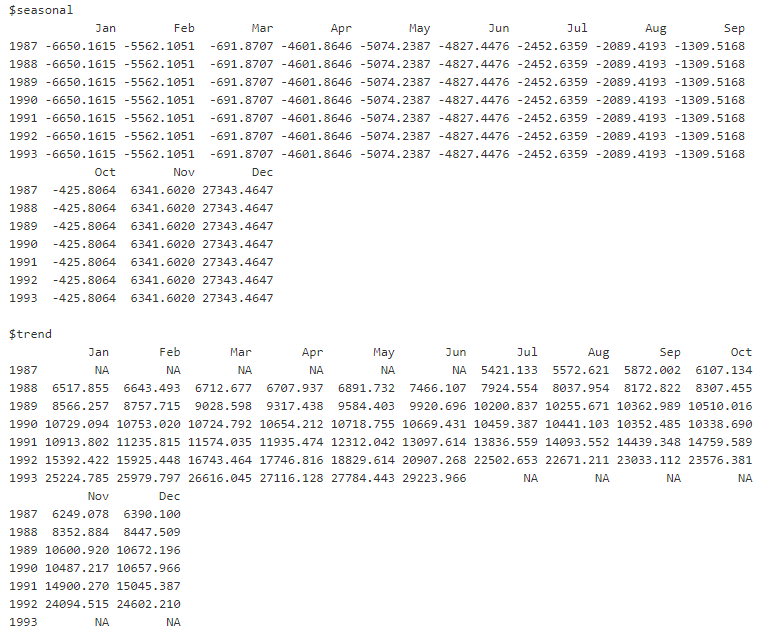
Decomposing the timeseries data can be done in 2 ways

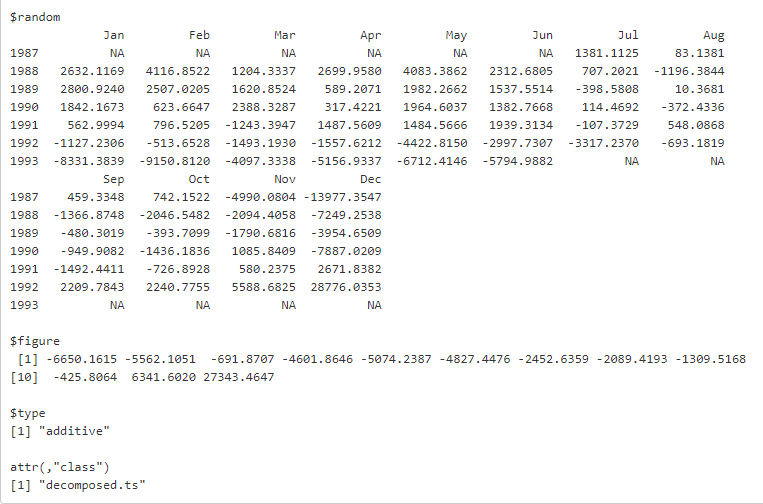
1. Additive method

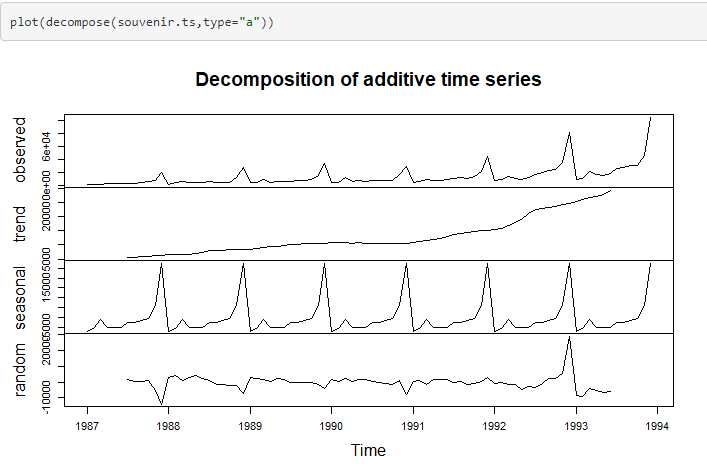
2. Multiplicative method

Additive method of decomposition

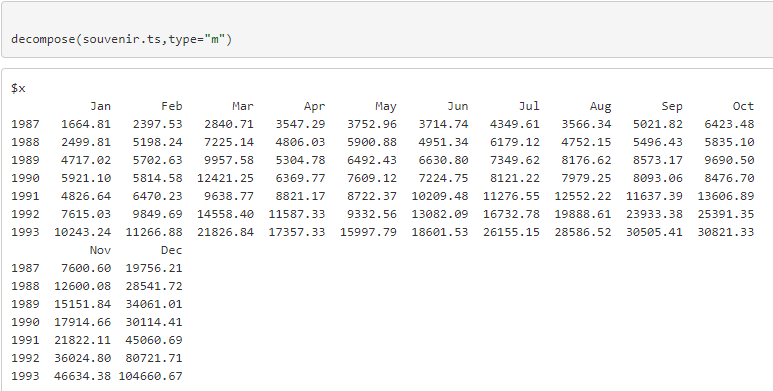


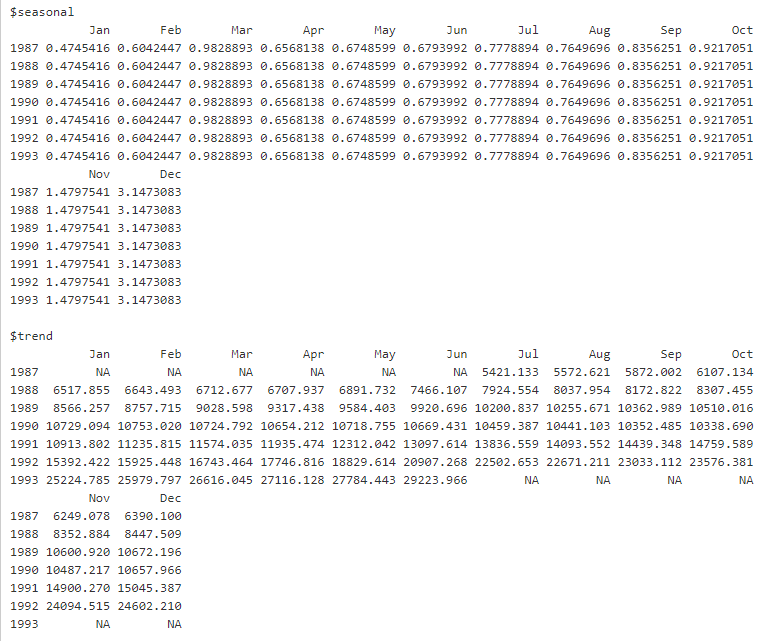


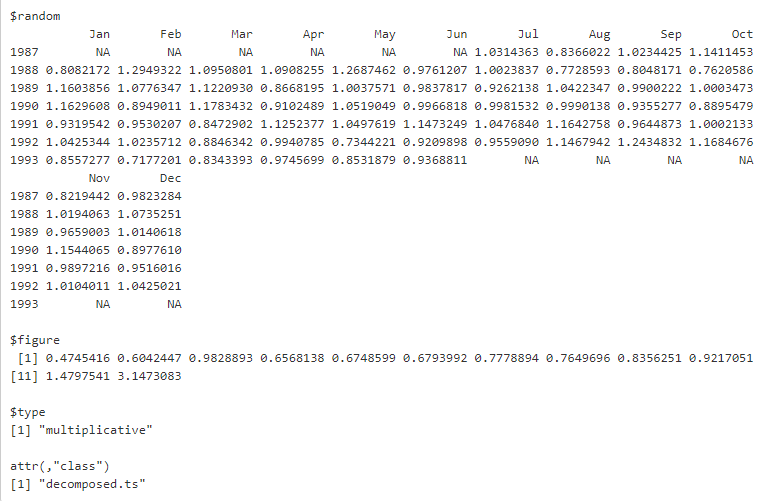


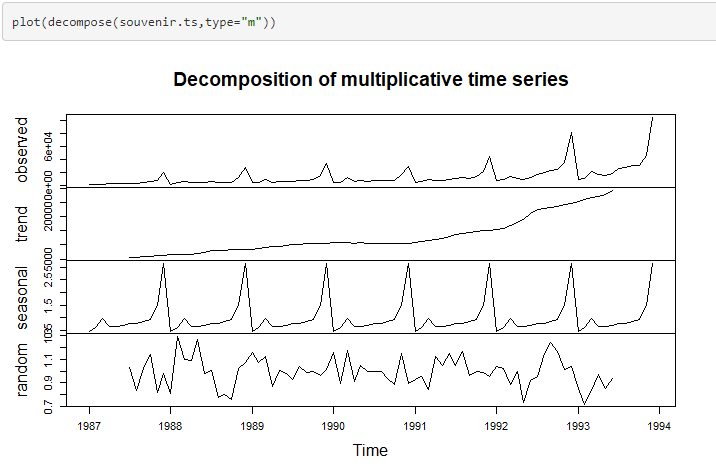


Multiplicative method of decomposition





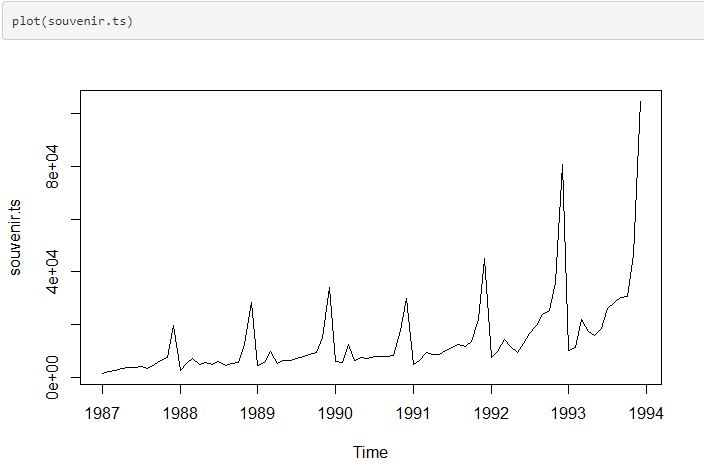


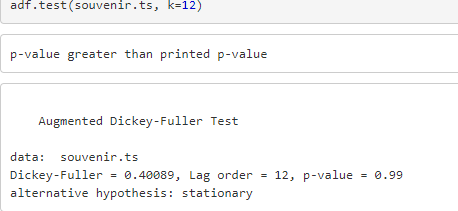
Looking at the decomposition plots for both additive and multiplicative type, we can see that there is a good upward trend over the last couple of years, and we can see the regular, seasonal component of our data.

We can attempt to build a forecast model using Holt-Winters exponential smoothing using both additive and multiplicative seasonal effects.

**Time Series Stationary Check**

A stationary time series is one whose properties do not depend on the time at which the series is observed. Time series are stationary if they do not have trend or seasonal effects. The trend and seasonality will affect the value of the time series at different times.

****

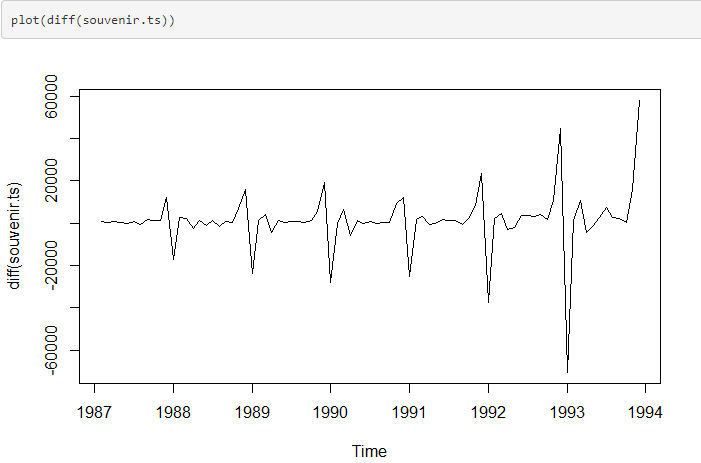
****

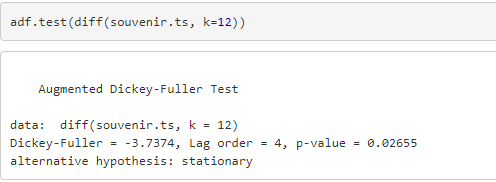
Augmented-Dickey Fuller Test is a statstical test for timeseries stationary check.

The Null Hypothesis states that the timeseries is not stationary.

The Alternate Hypothesis states that the timeseries is stationary.

The original timeseries (souvenir.ts) is not stationary because thje p-value is greater than 0.05.So we take the difference of the timeseries to eliminate the trend and seasonality

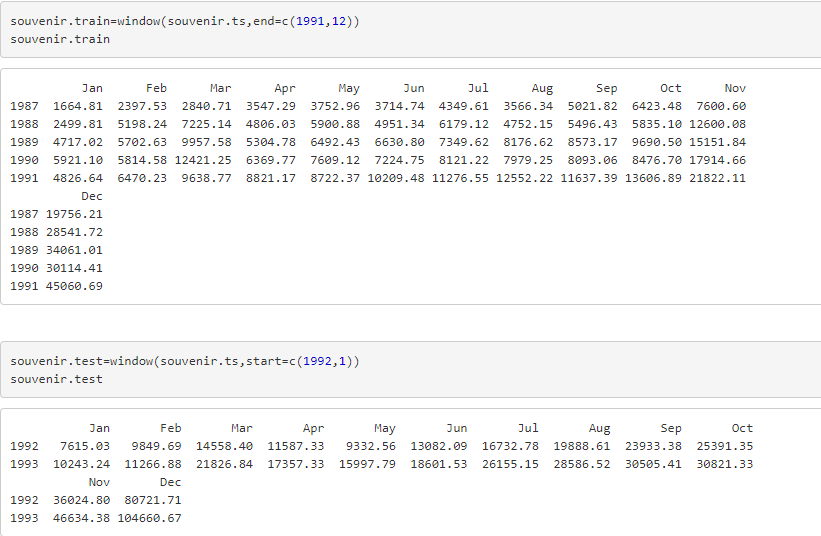
****

****

The Augmented Dickey-Fuller Test for the difference of the timeseries (diff(souvenir.ts)) is stationary as the p-value is lesser than 0.05. Thus the timeseries is not stationary for the difference of souvenir.ts

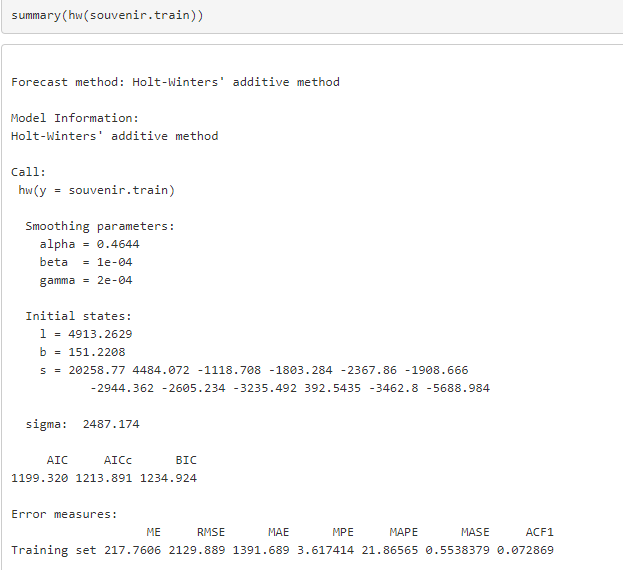
1. **Splitting Data into Train/Test**

Splitting the data into Train/Test is a most common practice for evaluating the performance of the Machine learning Algorithm.

****

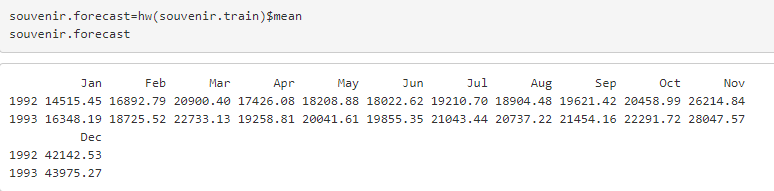
1. **Holt Winter Model (Triple Exponential Smoothing)**

Holt-Winter Additive Model

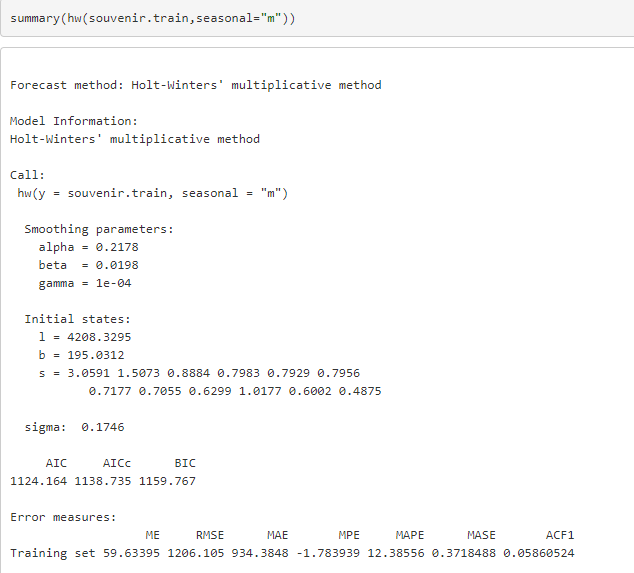
****

Holt-Winters smoothing uses 3 parameters i.e., alpha beta and gamma as the smoothing parameters. Alpha=0.46 indicates the estimate of level, beta=1e-04 indicates the trend and gamma= 2e-04 indicates the seasonality

**Forecast on Test data**

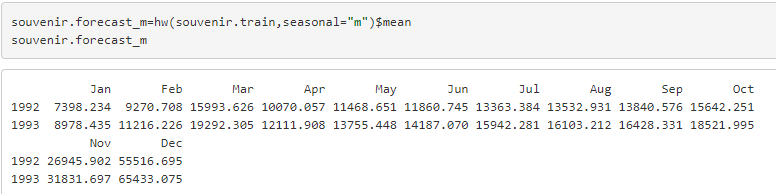
****

**Holt-Winters Multiplicative Model**

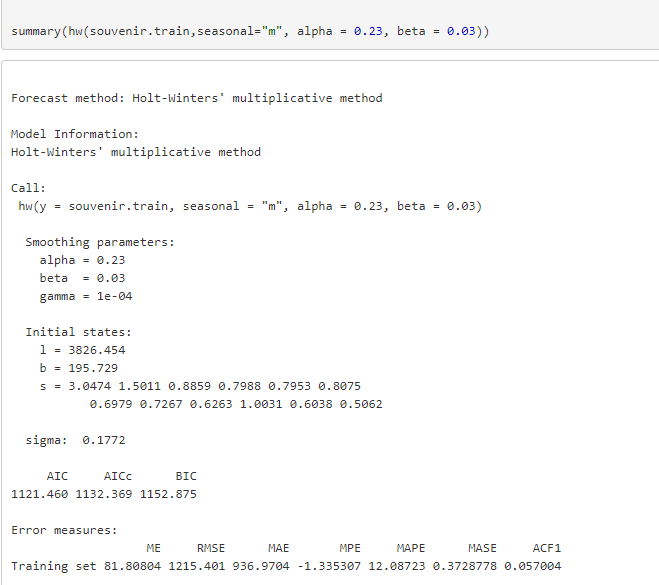
****

Holt-Winters smoothing uses 3 parameters i.e., alpha beta and gamma as the smoothing parameters. Alpha=0.217 indicates the estimate of level, beta=0.0198 indicates the trend and gamma= 1e-04 indicates the seasonality

**Forecast on Test data**

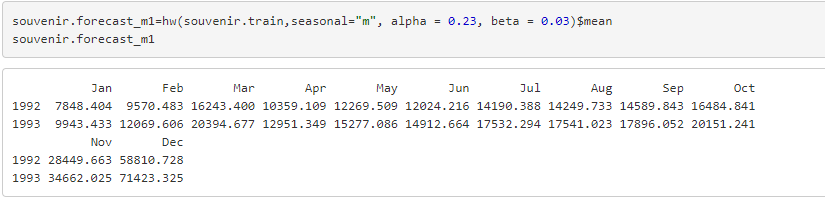
****

**Holt-Winters Multiplicative Model with tweaked alpha and beta values**

****

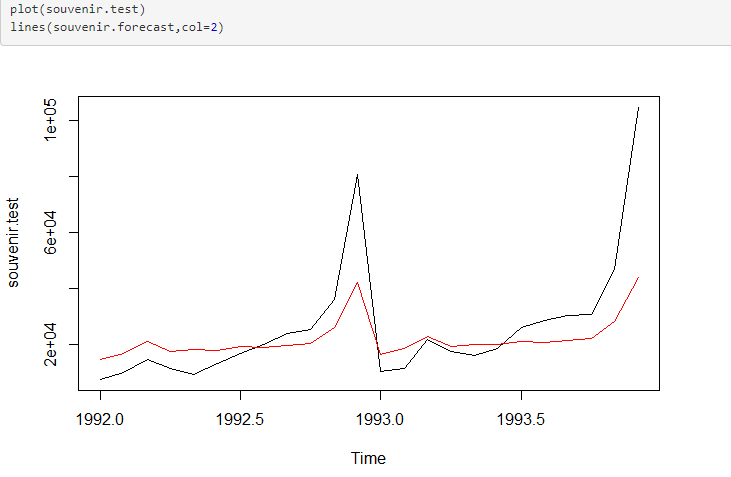
Alpha=0.23 indicates the estimate of level, beta=0.03 indicates the trend and gamma= 1e-04 indicates the seasonality

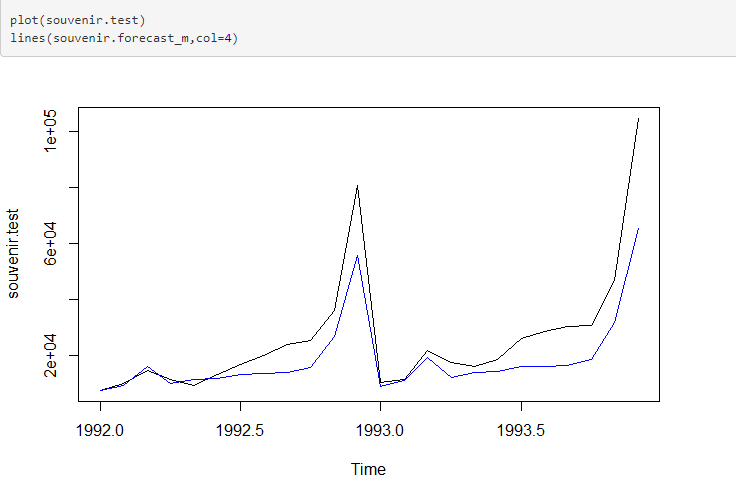
**Forecast on Test data**

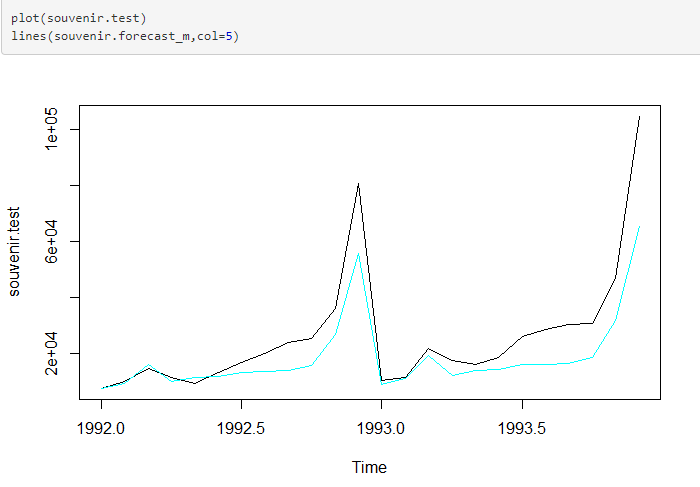
****

Comparing the RMSE and MAPE values, the Multiplicative Model performs better than the additive model as RMSE and MAPE values are low for multiplicative model.

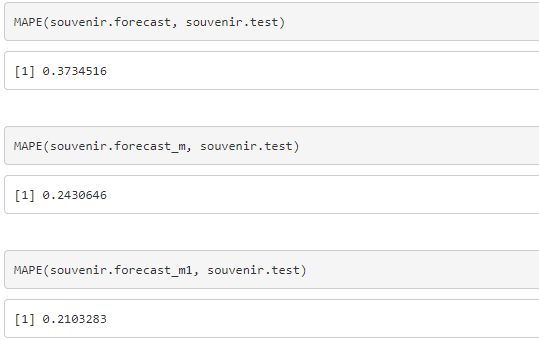
**Plotting the Forecasted values on the Test Data**

****

****

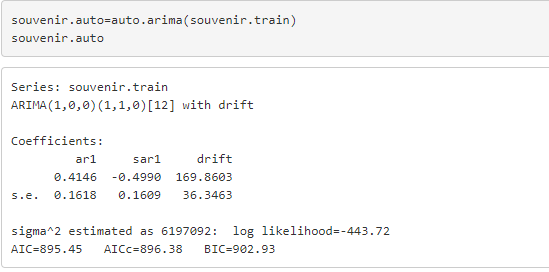
On plotting the Forecasted values using 3 forecasts (Additive, Multiplicative and Multiplicative-tweaked), we can observe that the Multiplicative model forecasts perform a better job at forecasting.

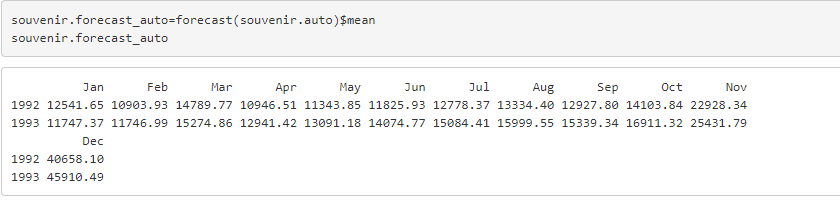
**Validation against Actual values using MAPE**



The Mean Absolute Percent Error (MAPE) is least for the Multiplicative model with tweaked alpha and beta values

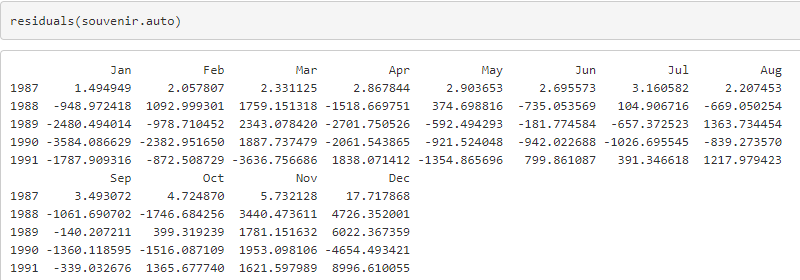
1. **ARIMA (SARIMA) Model**

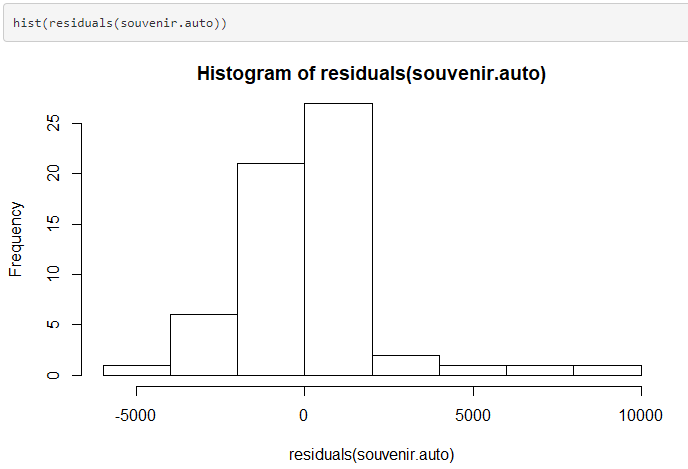
****

****

A good model, ideally the error will follow normal distribution with mean zero. To see this, it is necessary to check for the histogram of the error(residuals)

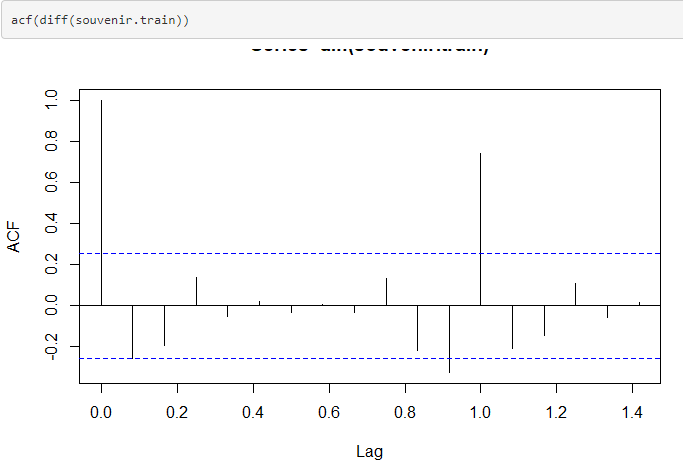
Residuals:

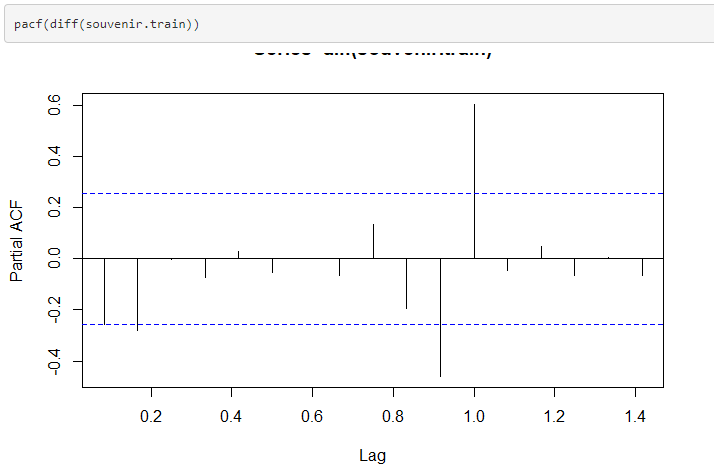


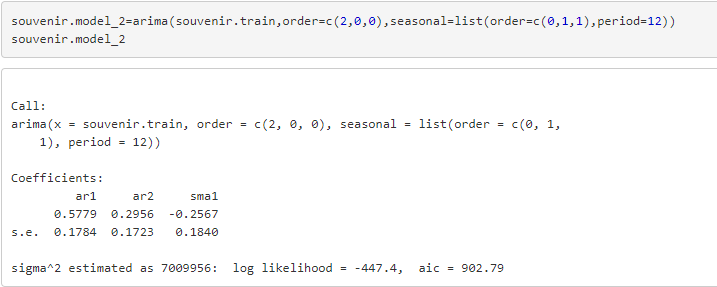
****

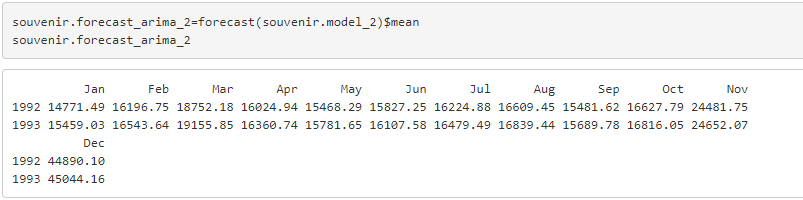
From the residul histogram plot above, it can be seen that the error roughly follow normal distribution with mean zero.

Manual ARIMA:

****

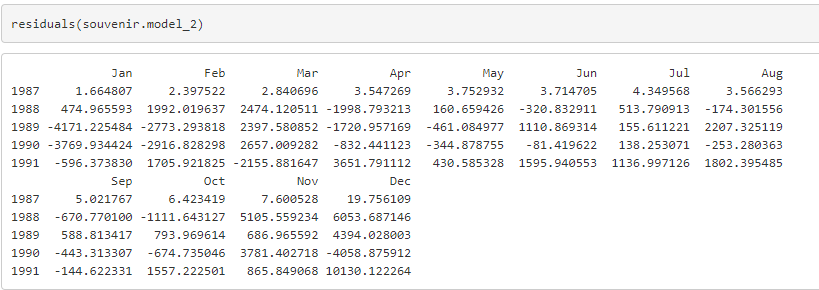
****

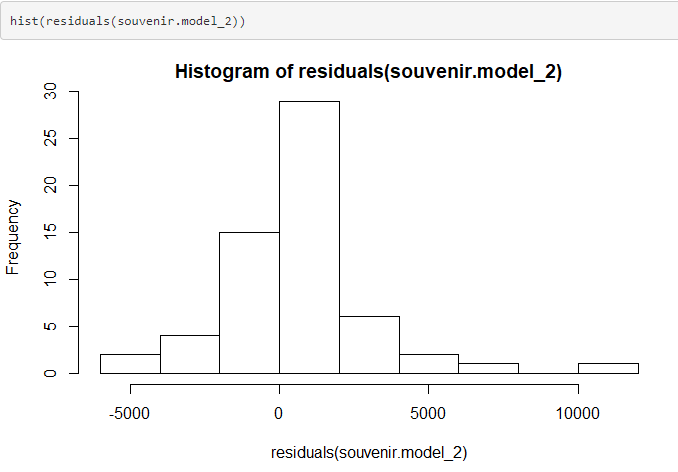
****

****

A good model, ideally the error will follow normal distribution with mean zero. To see this, it is necessary to check for the histogram of the error(residuals)

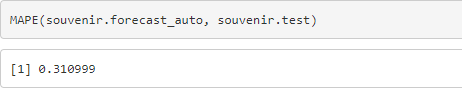
Residuals:

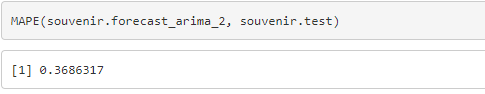




From the residul histogram plot above, it can be seen that the error roughly follow normal distribution with mean zero.

**Validation against Actual values using MAPE**

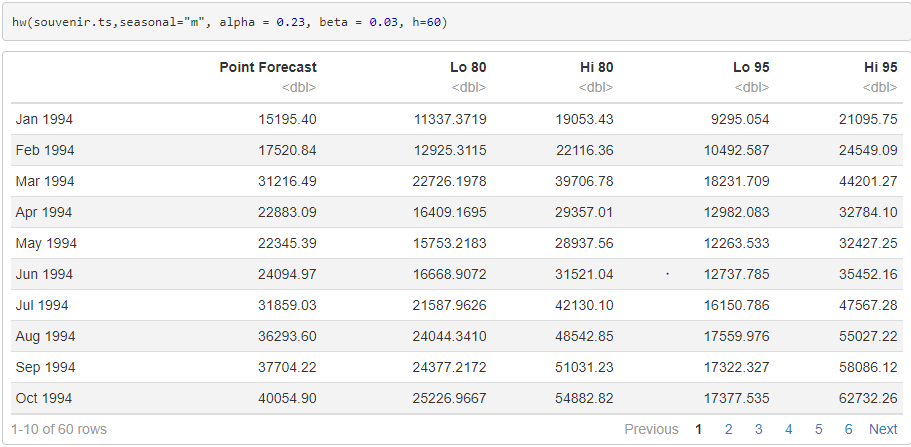
****

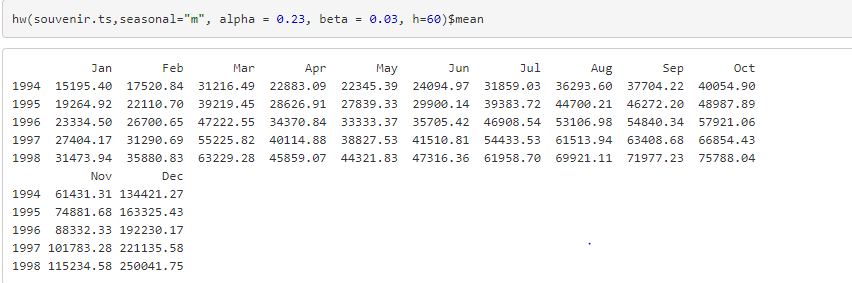
****

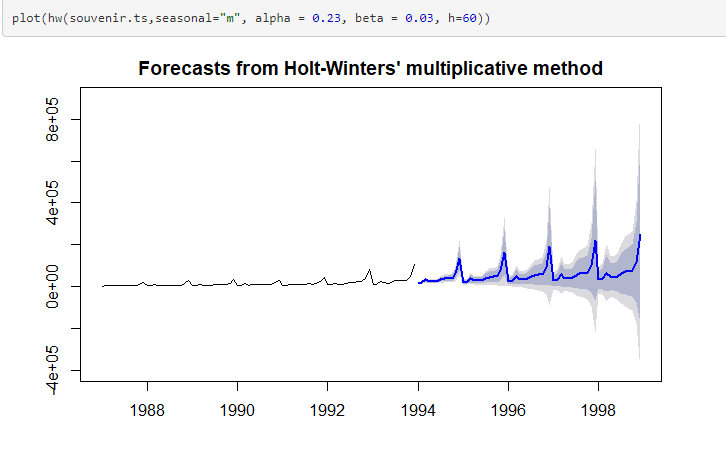
The Mean Absolute Percent Error (MAPE) is lesser for the auto arima Model than the Manual Arima, hence we can consider the auto.arima to be a better model.

1. **Predicting the values for next 5 years**

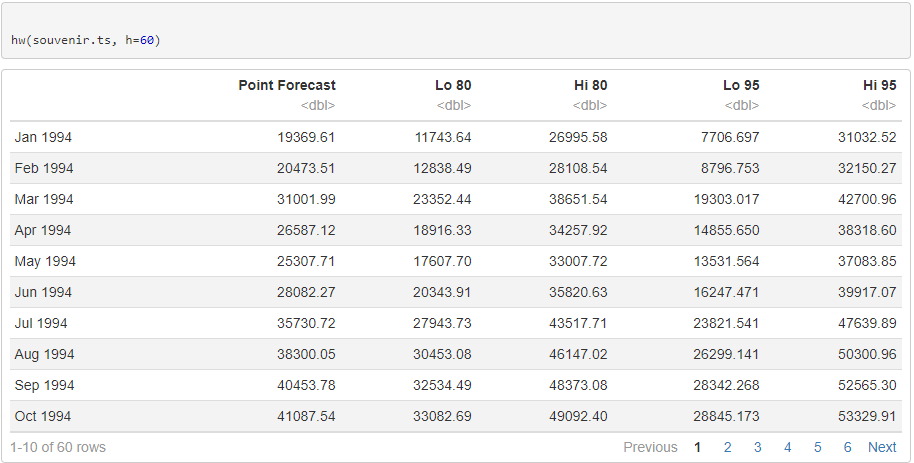
Predicting the sales for next 5 years using the Holt-Winters Multiplicative model

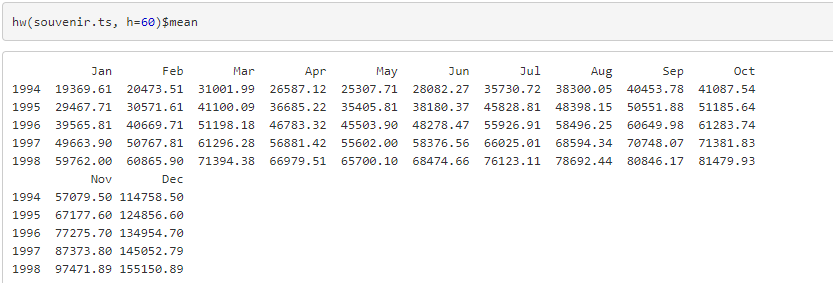
****

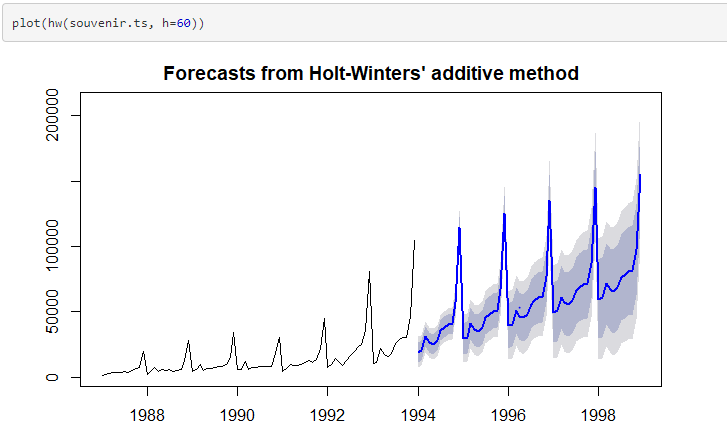
****

****

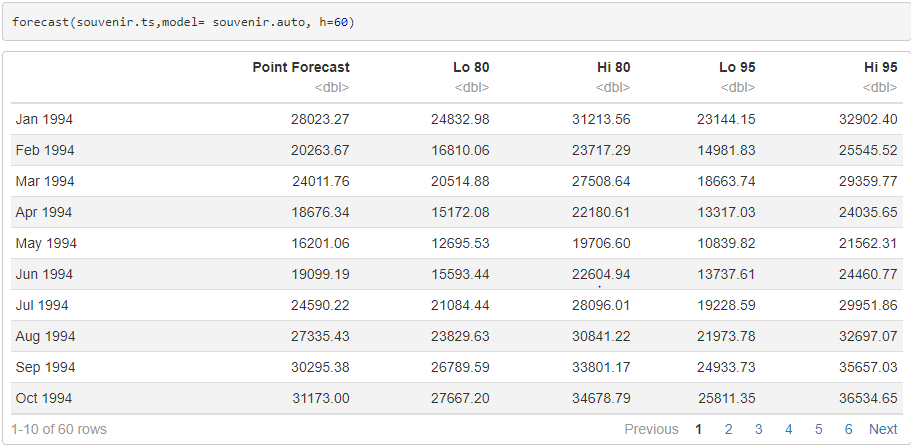
Predicting the sales for next 5 years using the Holt-Winters Additive model

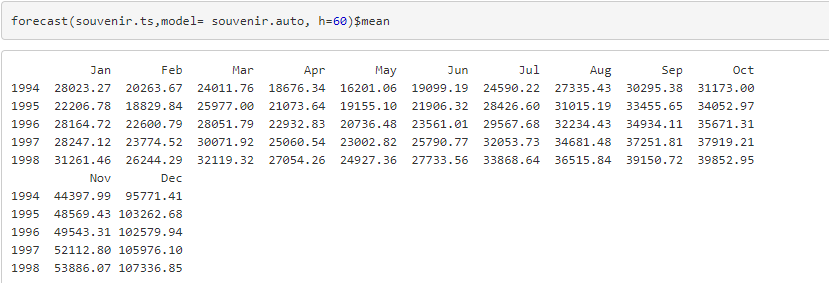


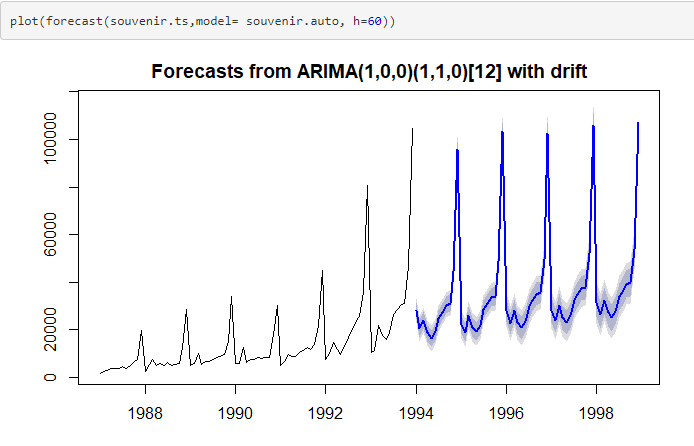




Predicting the sales for next 5 years using the Auto ARIMA model







Predicting the sales for next 5 years using the Manual ARIMA model

