TIME SERIES GROUP ASSIGNMENT

Souvenir Shop

Abstract

Time Series Analysis on Sales of Souvenir for a Fancy Store at a beach resort town in Queensland, Australia from January 1987 to December 1993.

Group #8

Vinod Kumar / Jithesh / Tarique / Shivkumar

Contents

1. (a) Visualizing the whole data, dividing the time series data into different components and plotting the different components individually

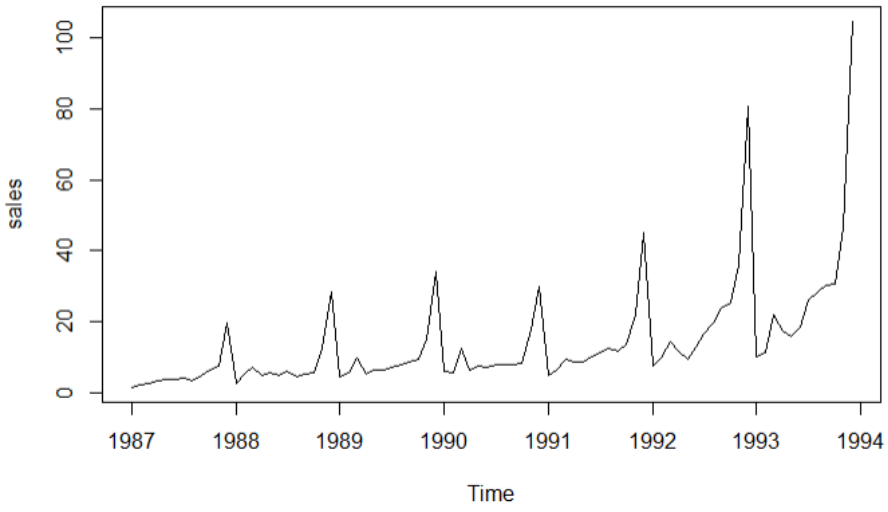
Monthly Sales Data related to Sales of Souvenir for a souvenir shop at a beach resort town in Queensland, Australia has been provided from January 1987 to December 1993.

A time series object has been created for the monthly sales data and following is the output: -

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
1987	1664.81	2397.53	2840.71	3547.29	3752.96	3714.74	4349.61	3566.34	5021.82
1988	2499.81	5198.24	7225.14	4806.03	5900.88	4951.34	6179.12	4752.15	5496.43
1989	4717.02	5702.63	9957.58	5304.78	6492.43	6630.80	7349.62	8176.62	8573.17
1990	5921.10	5814.58	12421.25	6369.77	7609.12	7224.75	8121.22	7979.25	8093.06
1991	4826.64	6470.23	9638.77	8821.17	8722.37	10209.48	11276.55	12552.22	11637.39
1992	7615.03	9849.69	14558.40	11587.33	9332.56	13082.09	16732.78	19888.61	23933.38
1993	10243.24	11266.88	21826.84	17357.33	15997.79	18601.53	26155.15	28586.52	30505.41

	Oct	Nov	Dec
1987	6423.48	7600.60	19756.21
1988	5835.10	12600.08	28541.72
1989	9690.50	15151.84	34061.01
1990	8476.70	17914.66	30114.41
1991	13606.89	21822.11	45060.69
1992	25391.35	36024.80	80721.71
1993	30821.33	46634.38	104660.67

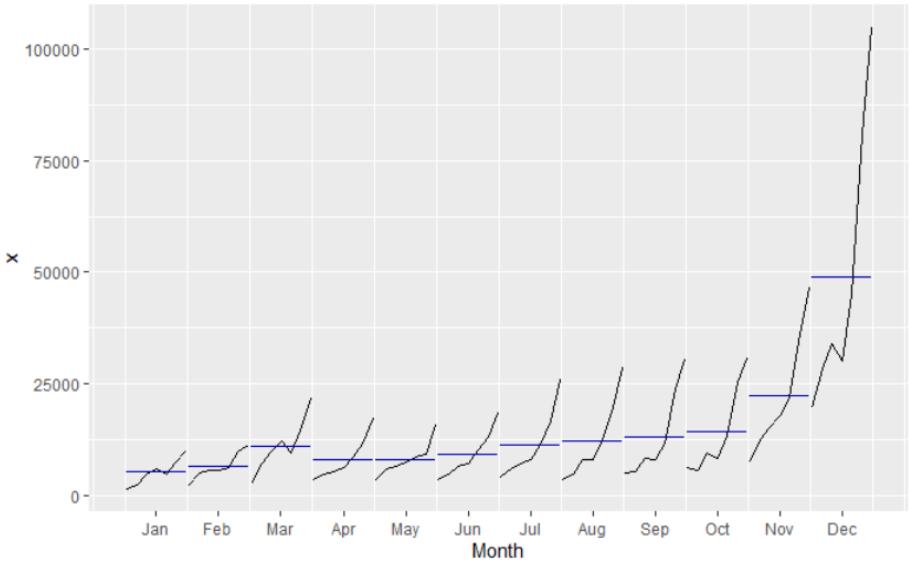
Following is the plot of the time series: -



Note – Sales Figures in '000

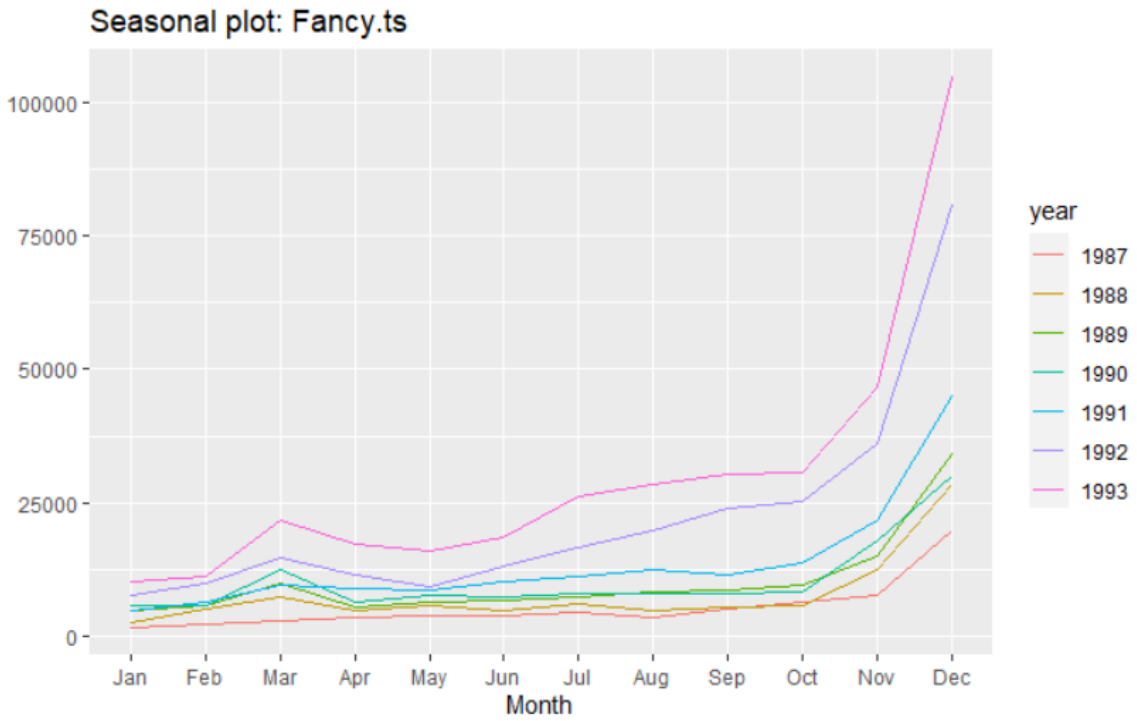
- The overall sales pattern is similar until December 1990, and gradually increases (trend) from 1991 - 1993.
- The highest sales have been recorded in December 1993.
- There is Seasonality present in the data, Nov – Dec has peaks in sales every year

Following is the monthly plot across years: -



The month of December records the highest average Sales across years.

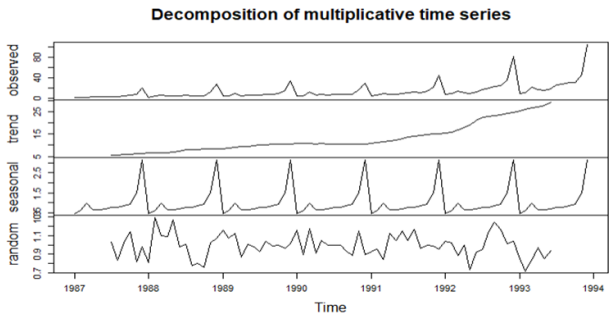
Following is the seasonal plot: -

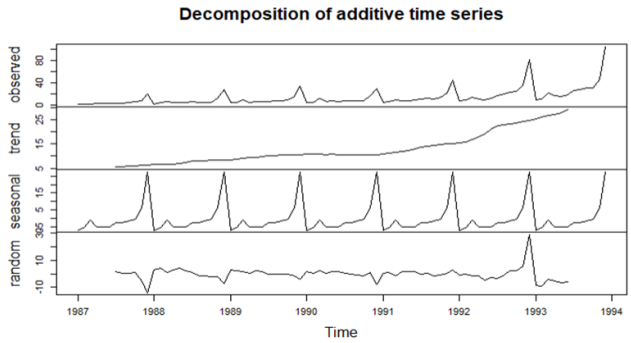


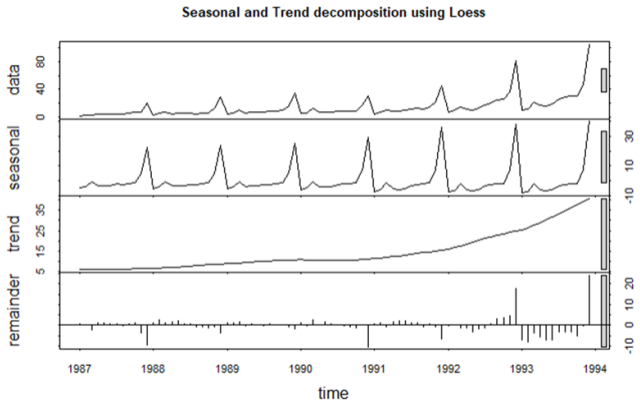
The highest Sales has been recorded in the Year 1993, across months.

Decomposition

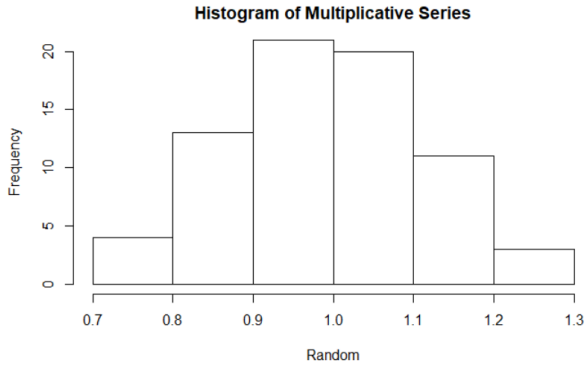
Following is the decomposition plots of the time series: -





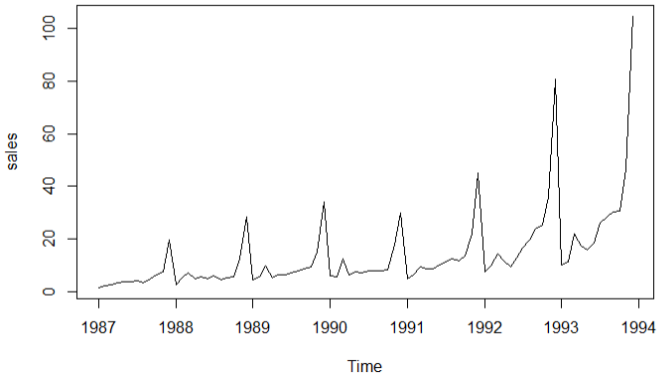


- There are two spikes towards the end of the series (STL plot) which is influenced by randomness. More information about data is required for conclusion on the spike.
- The random errors are normally distributed in the multiplicative time series as compared to the additive time series as shown by the below histograms: -



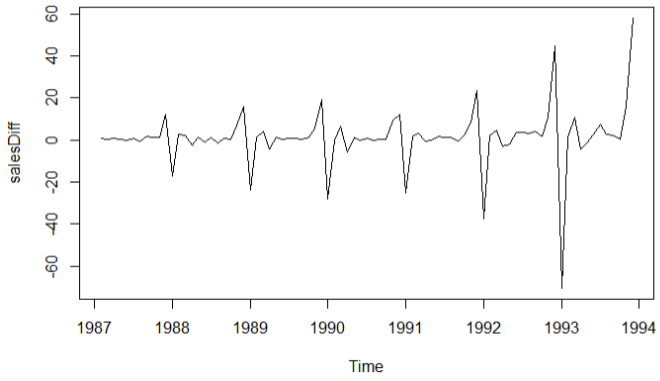
1.b) Time series stationarity check

The original time series is not stationary. Augmented Dickey Fuller Test on the original time series gives a p-value of greater than 0.5 to justify non-stationarity.




```
        Augmented Dickey-Fuller Test

data:  Fancy.ts
Dickey-Fuller = -2.0809, Lag order = 4, p-value = 0.5427
alternative hypothesis: stationary
```

The first order difference of the time series is represented by the below plot: -



Augmented Dickey Fuller Test on the first order difference gives a p-value of less than 0.5 to justify stationarity.

```
        Augmented Dickey-Fuller Test

data:  diff(Fancy.ts)
Dickey-Fuller = -3.7374, Lag order = 4, p-value = 0.02655
alternative hypothesis: stationary
```

2.Splitting data into Test/Train or Dev/hold-out

The train data is considered from January 1987 to December 1991.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
1987	1664.81	2397.53	2840.71	3547.29	3752.96	3714.74	4349.61	3566.34	5021.82	6423.48
1988	2499.81	5198.24	7225.14	4806.03	5900.88	4951.34	6179.12	4752.15	5496.43	5835.10
1989	4717.02	5702.63	9957.58	5304.78	6492.43	6630.80	7349.62	8176.62	8573.17	9690.50
1990	5921.10	5814.58	12421.25	6369.77	7609.12	7224.75	8121.22	7979.25	8093.06	8476.70
1991	4826.64	6470.23	9638.77	8821.17	8722.37	10209.48	11276.55	12552.22	11637.39	13606.89

	Nov	Dec
1987	7600.60	19756.21
1988	12600.08	28541.72
1989	15151.84	34061.01
1990	17914.66	30114.41
1991	21822.11	45060.69

The test data is considered from January 1992 to December 1993

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
1992	7615.03	9849.69	14558.40	11587.33	9332.56	13082.09	16732.78	19888.61	23933.38
1993	10243.24	11266.88	21826.84	17357.33	15997.79	18601.53	26155.15	28586.52	30505.41

	Oct	Nov	Dec
1992	25391.35	36024.80	80721.71
1993	30821.33	46634.38	104660.67

3.a) Holt Winter Model creation

Applying Holt-Winters model on the Train data using <u>additive method</u>. Following output gives the default smoothing parameters and the error measures: -

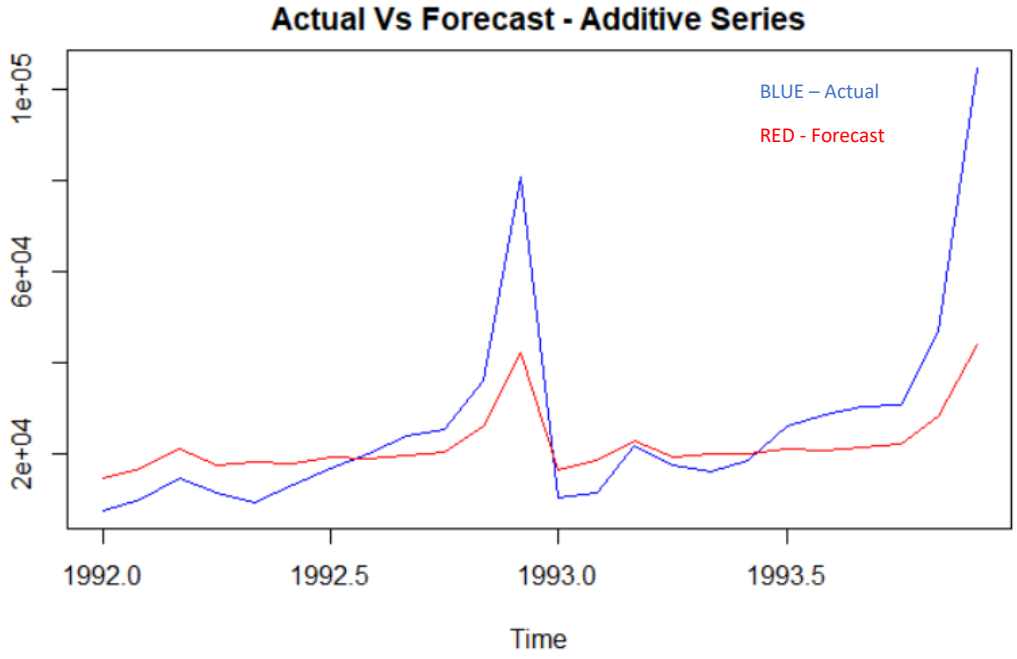
```
Forecast method: Holt-Winters' additive method

Model Information:
Holt-Winters' additive method

Call:
 hw(y = Fancy_Train, h = 24, seasonal = "a")

  Smoothing parameters:
    alpha = 0.4644
    beta  = 1e-04
    gamma = 2e-04

  Initial states:
    l = 4913.2629
    b = 151.2208
    s = 20258.77 4484.072 -1118.708 -1803.284 -2367.86 -1908.666
           -2944.362 -2605.234 -3235.492 392.5435 -3462.8 -5688.984

  sigma:  2487.174

     AIC     AICc      BIC
1199.320 1213.891 1234.924

Error measures:
                    ME     RMSE      MAE      MPE     MAPE      MASE      ACF1
Training set 217.7606 2129.889 1391.689 3.617414 21.86565 0.5538379 0.072869
```

3.b) Predicting the values for test dataset using HW model

Following output gives the forecast for 24 months of Test Data using the Additive method for different confidence levels: -

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Jan 1992	14515.45	11328.01	17702.90	9640.682	19390.23
Feb 1992	16892.79	13378.30	20407.28	11517.837	22267.74
Mar 1992	20900.40	17086.66	24714.14	15067.789	26733.01
Apr 1992	17426.08	13334.79	21517.36	11168.991	23683.16
May 1992	18208.88	13857.58	22560.17	11554.148	24863.61
Jun 1992	18022.62	13425.88	22619.35	10992.516	25052.72
Jul 1992	19210.70	14380.87	24040.53	11824.109	26597.29
Aug 1992	18904.48	13852.19	23956.78	11177.663	26631.30
Sep 1992	19621.42	14355.95	24886.89	11568.584	27674.27
Oct 1992	20458.99	14988.54	25929.43	12092.666	28825.31
Nov 1992	26214.84	20546.73	31882.94	17546.217	34883.46
Dec 1992	42142.53	36283.34	48001.73	33181.665	51103.40
Jan 1993	16348.19	10303.72	22392.65	7103.978	25592.40
Feb 1993	18725.52	12501.34	24949.71	9206.451	28244.59
Mar 1993	22733.13	16334.19	29132.08	12946.787	32519.48
Apr 1993	19258.81	12689.67	25827.96	9212.171	29305.45
May 1993	20041.61	13306.49	26776.74	9741.132	30342.10
Jun 1993	19855.35	12958.16	26752.54	9307.014	30403.69
Jul 1993	21043.44	13987.83	28099.04	10252.815	31834.06
Aug 1993	20737.22	13526.59	27947.84	9709.519	31764.92
Sep 1993	21454.16	14091.71	28816.61	10194.261	32714.06
Oct 1993	22291.72	14780.44	29803.01	10804.205	33779.24
Nov 1993	28047.57	20390.28	35704.87	16336.749	39758.39
Dec 1993	43975.27	36174.62	51775.91	32045.212	55905.32

Plotting a comparison plot between the Test Data and the Forecasted Data using Holt-Winters



Applying Holt-Winters model on the Train data using multiplicative method. Following output gives the default smoothing parameters and the error measures: -

```
Forecast method: Holt-Winters' multiplicative method

Model Information:
Holt-Winters' multiplicative method

Call:
 hw(y = Fancy_Train, h = 24, seasonal = "m")

  Smoothing parameters:
    alpha = 0.2178
    beta  = 0.0198
    gamma = 1e-04

  Initial states:
    l = 4208.3295
    b = 195.0312
    s = 3.0591 1.5073 0.8884 0.7983 0.7929 0.7956
           0.7177 0.7055 0.6299 1.0177 0.6002 0.4875

  sigma:  0.1746

     AIC     AICc      BIC
1124.164 1138.735 1159.767

Error measures:
                     ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 59.63395 1206.105 934.3848 -1.783939 12.38556 0.3718488 0.05860524
```

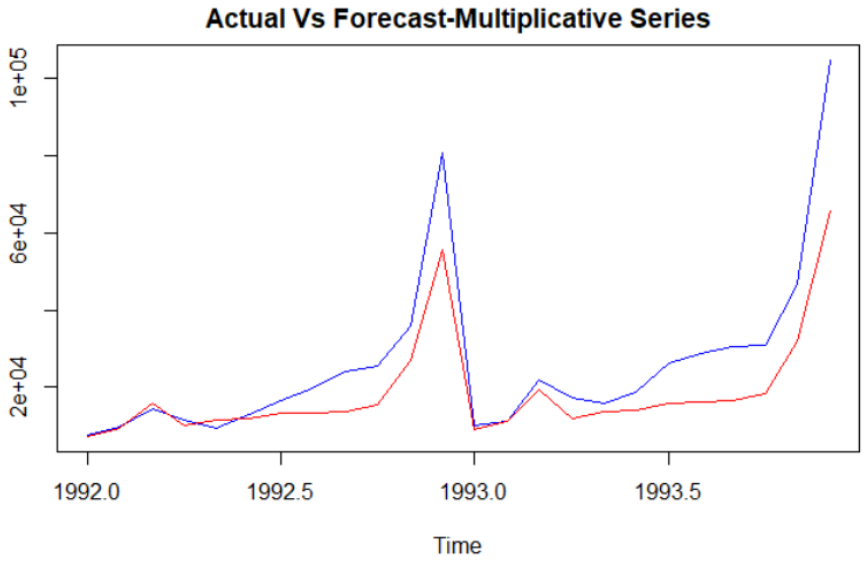
Forecasting the Sales for January 1992 – December 1993 leveraging HW Multiplicative Model and validating against Test Data available for the same period

Following output gives the forecast for 24 months of Test Data using the multiplicative method for different confidence levels: -

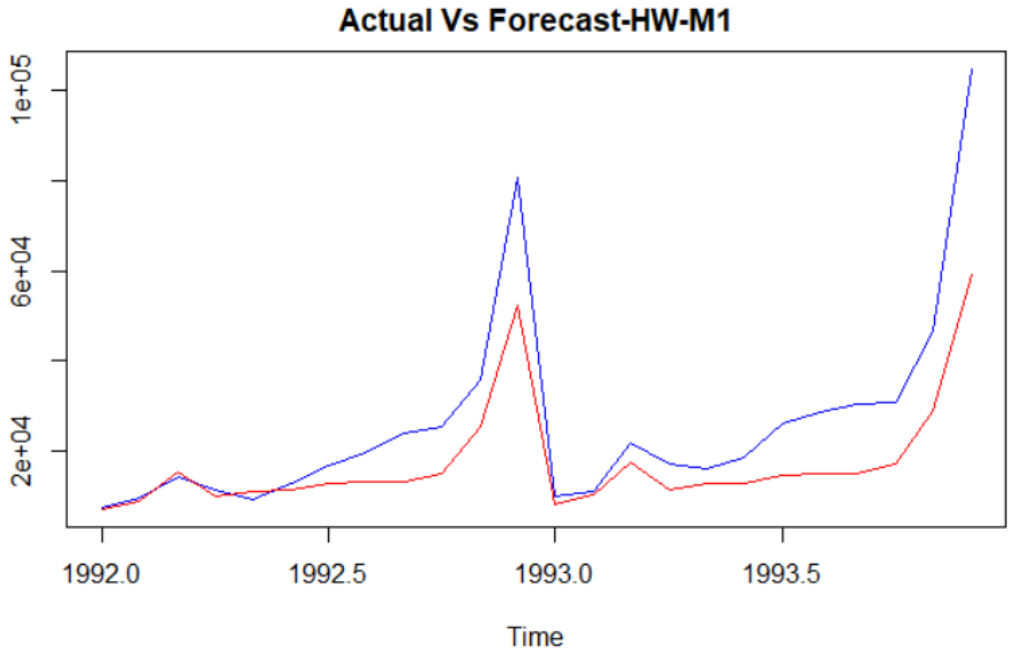
	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Jan 1992	7398.234	5743.181	9053.286	4867.049	9929.418
Feb 1992	9270.708	7139.314	11402.103	6011.022	12530.395
Mar 1992	15993.626	12207.158	19780.095	10202.723	21784.530
Apr 1992	10070.057	7610.864	12529.249	6309.046	13831.067
May 1992	11468.651	8575.632	14361.670	7044.161	15893.141
Jun 1992	11860.745	8766.828	14954.662	7129.008	16592.482
Jul 1992	13363.384	9755.654	16971.115	7845.837	18880.932
Aug 1992	13532.931	9749.475	17316.386	7746.634	19319.227
Sep 1992	13840.576	9831.948	17849.203	7709.909	19971.243
Oct 1992	15642.251	10947.952	20336.550	8462.941	22821.561
Nov 1992	26945.902	18566.544	35325.260	14130.779	39761.025
Dec 1992	55516.695	37629.287	73404.102	28160.266	82873.123
Jan 1993	8978.435	5981.692	11975.178	4395.312	13561.558
Feb 1993	11216.226	7339.318	15093.134	5287.007	17145.446
Mar 1993	19292.305	12389.036	26195.574	8734.666	29849.944
Apr 1993	12111.908	7627.214	16596.603	5253.160	18970.657
May 1993	13755.448	8487.473	19023.423	5698.776	21812.120
Jun 1993	14187.070	8570.187	19803.952	5596.789	22777.350
Jul 1993	15942.281	9420.563	22463.998	5968.175	25916.386
Aug 1993	16103.212	9300.198	22906.226	5698.900	26507.524
Sep 1993	16428.331	9264.850	23591.812	5472.732	27383.930
Oct 1993	18521.995	10190.497	26853.494	5780.068	31263.922
Nov 1993	31831.697	17069.103	46594.291	9254.258	54409.137
Dec 1993	65433.075	34162.769	96703.382	17609.268	113256.882

24 rows

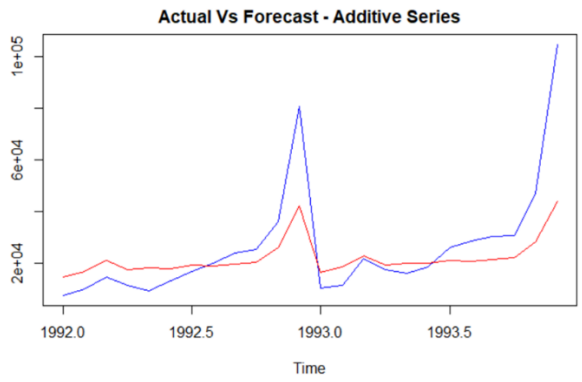
Plotting a comparison plot between the Test Data and the Forecasted Data using Holt-Winters (multiplicative) for the period January 1992 to December 1993: -

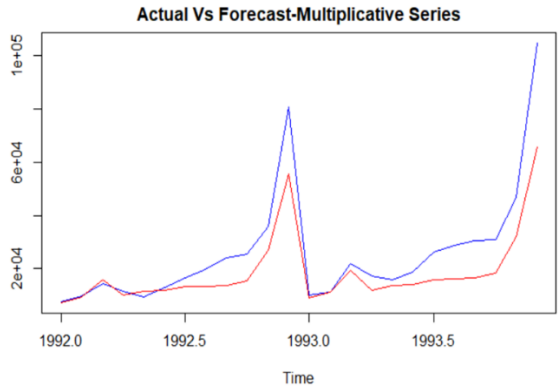


Following is the prediction on Test Data with smoothing parameters taken as alpha = 0.8, beta = 0.02, gamma = 0.0001: -



Comparison of the Additive and Multiplicative output based on the forecasting model arrived on the basis of the Train data: -





3.c) validation against actual values using MAPE

Following table summarizes the error measures for the additive and multiplicative model and the related smoothing parameters: -

Smoothing Parameters	**Additive**	**Multiplicative**	**Multiplicative (M1)**
Alpha	0.4644	0.2178	0.8
Beta	0.0001	0.0198	0.02
Gamma	0.0002	0.0001	0.0001
Error Measures – Train Data	**Additive**	**Multiplicative**	**Multiplicative (M1)**
RMSE	2129	1206	1421.397
MAPE	21.87	12.38	11.95364
Error Measures – Test Data	**Additive**	**Multiplicative**	**Multiplicative (M1)**
MAPE	37.34	24.31	27.63

The multiplicative models give the best results, with error measures being lesser than the additive model.

4.a) ARIMA Model creation

Stationarity of Train Data

```
        Augmented Dickey-Fuller Test

data:  Fancy_Train
Dickey-Fuller = -3.1882, Lag order = 3, p-value = 0.09786
alternative hypothesis: stationary
```

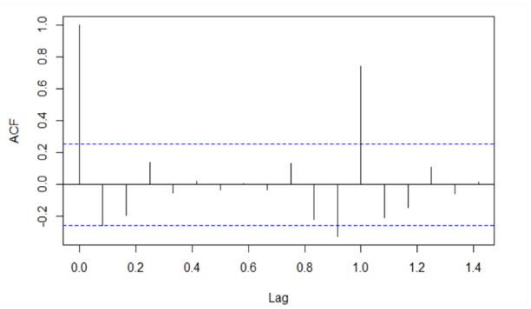
The p-value is greater than 5%, hence the Train Data is not stationary

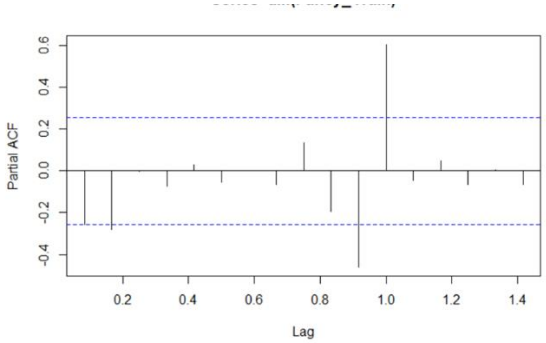
Taking the first difference in the train data and carrying out an ADF test gave the following output: -

```
p-value smaller than printed p-value
        Augmented Dickey-Fuller Test

data:  diff(Fancy_Train)
Dickey-Fuller = -4.5431, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Thus, the first difference of Train data, is stationary as the p-values are low.

Following are the ACF & PACF plots related to Autocorrelation based on the First-Difference Train Data (d=1): -





- Based on above ACF plot it looks like MA(q) order of 1 is significant (almost touching the significance dotted line)
- From PACF plot AR (p)order of 2 is significant

The ARIMA[p,d,q] model for Fancy Train Data is ARIMA[2,1,1]. Following are the coefficients and the error measures: -

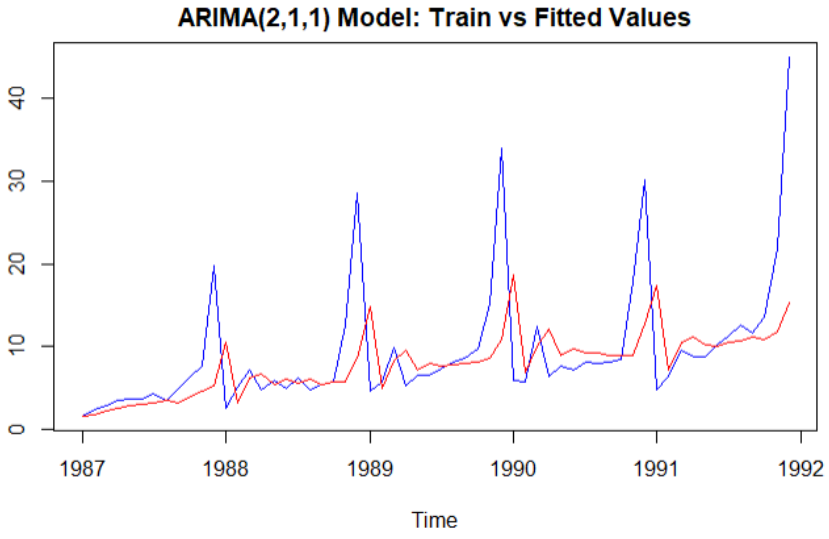
```
arima(x = salesTrain, order = c(2, 1, 1))

Coefficients:
         ar1      ar2      ma1
      0.2771  -0.1206  -0.8617
s.e.  0.1723   0.1607   0.1033

sigma^2 estimated as 54.66:  log likelihood = -202.31,  aic = 412.62

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE         ACF1
Training set 1.531089 7.331218 4.016114 -8.602968 41.15481 0.8776473 -0.03115535
```

Coefficients	**Ar1**	**Ar2**	**Ma1**
t-Score	1.61	0.75	8.33

Moving Average part of the equation carries more weightage as compared to the rest, because of the high t-score

Equation based on the ARIMA (2,1,1):

$X_t = 0.2771X_{(t-1)} - 0.1206X_{(t-2)} - E_t - 0.8617E_{(t-1)}$



The fitted values are one period ahead of the actual data (blue line), as the ARIMA model does NOT account for SEASONALITY aspect, as exhibited by the decomposition plot

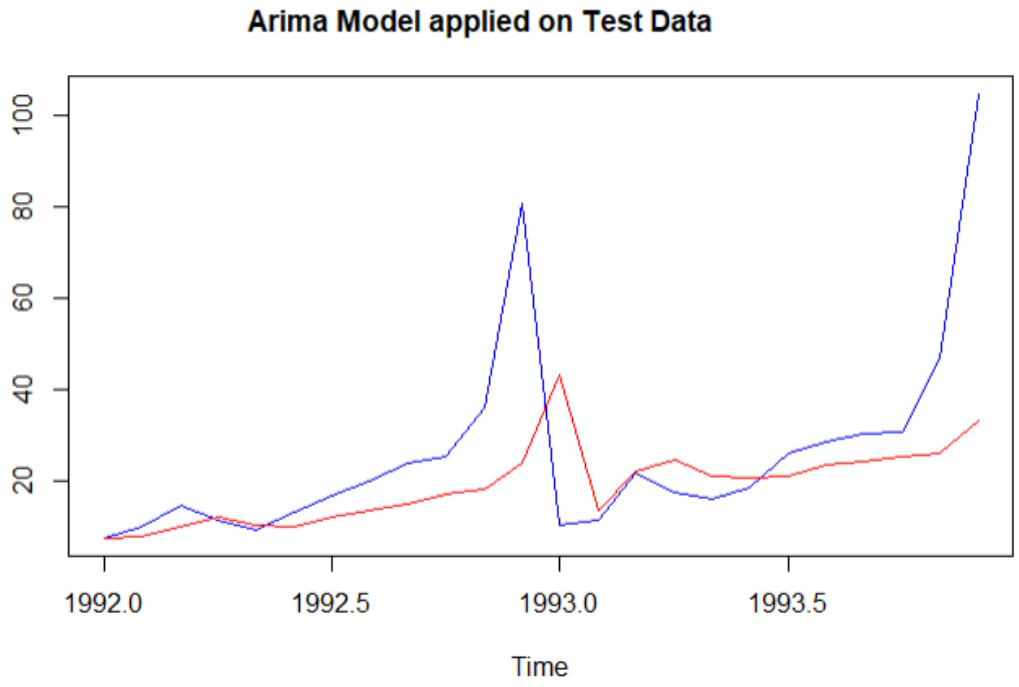
Box-Ljung test to check if autocorrelation exist for residuals on the ARIMA(2,1,1) model:

```
        Box-Ljung test

data:  arima.fit$residuals
X-squared = 1.377, df = 5, p-value = 0.9268
```

The p-value is high, which endorses Null Hypothesis that no auto-correlation exist in residuals and model can be used.

4.b) Predicting the values for test dataset using ARIMA model

Following is the forecast on the test data using ARIMA [2,1,1]



- The fitted values are one period ahead of the actual data (blue line), as the model does not account for Seasonality aspect, as exhibited by the decomposition plot

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
1992	7.607385	7.976262	9.854930	12.004836	10.468568	10.102637	12.030959	13.452804
1993	43.043819	13.480987	21.953093	24.726864	21.180416	20.619598	21.227355	23.692279

	Sep	Oct	Nov	Dec
1992	14.986946	17.136860	18.378324	23.970821
1993	24.136252	25.258560	25.889222	33.114869

*Values are in '000

4a.1) Seasonal – ARIMA Model Creation

Following ARIMA models of different orders were created. ARIMA (2,1,2) (1,1,2) [12] gives the best results in terms of having least values of error measures compared to the rest.

Model	MAPE	RMSE
ARIMA(1,1,1)(1,1,1)[12]	15.12	2231
ARIMA(2,1,1)(0,1,0)[12]	14.80	2349
ARIMA(2,1,1)(1,1,2)[12]	11.10	1661
ARIMA(2,1,2)(1,1,2)[12]	11.21	1653

ARIMA Model (2,1,2) (1,1,2) [12] Output: -

```
Call:
arima(x = salesTrain, order = c(2, 1, 2), seasonal = list(order = c(1, 1, 2),
    period = 12))

Coefficients:
          ar1      ar2     ma1      ma2    sar1     sma1    sma2
      -0.6347  -0.2046  0.0572  -0.2911  0.4473  -1.1381  0.9996
s.e.   0.5406   0.3242  0.5427   0.4968  0.6812   0.8358  0.7334

sigma^2 estimated as 3.491:  log likelihood = -108.85,  aic = 233.71

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 0.1502816 1.653725 1.066512 -2.223948 11.21494 0.2330664 -0.01925558
```

Coefficients	Ar1	Ar2	Ma1	Ma2	Sar1	Sma1	Sma2
t-score	1.174	0.6310	0.105	0.586	0.657	1.362	1.363

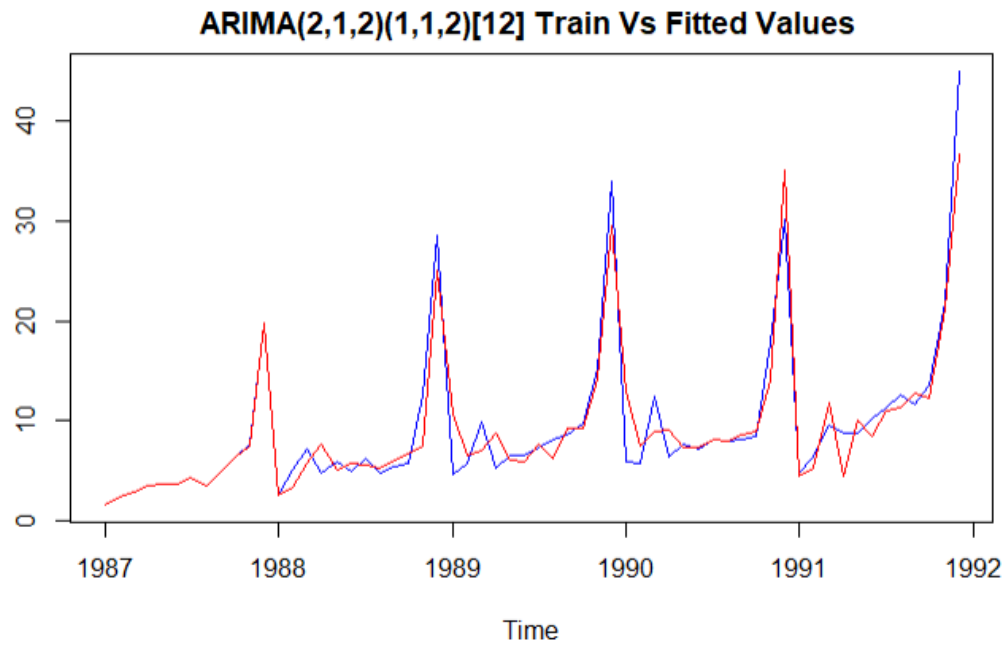
Seasonality related moving average parameters have higher weightage compared to the rest of the coefficients based on the t-score

Based on the ARIMA model, monthly order of difference is 1 (d) and seasonal order of difference is 1 (D). Hence: -

$$X_t = (Y_t - Y_{(t-1)}) - (Y_{(t-12)} - Y_{(t-13)})$$

Equation for ARIMA(2,1,2)(1,1,2)[12] is: -

$$X_t = -0.6347X_{(t-1)} - 0.2046X_{(t-2)} + E_t + 0.0572E_{(t-1)} - 0.2911E_{(t-2)} + 0.4473X_{(t-12)} + F_t + 0.4473F_{(t-1)} - 1.1381F_{(t-2)}$$



Box-Ljung test to check if autocorrelation exist for residuals on the

```
        Box-Ljung test

data:  arima.fit.s$residuals
X-squared = 0.49122, df = 5, p-value = 0.9924
```

The p-value is high, which endorses Null Hypothesis that no autocorrelation exist in residuals and model can be used.

4.c) Predicting the values for test dataset using SARIMA model

We are using a One-step forecast on test data method which leads to decrease the RMSE and MAPE.

Code:

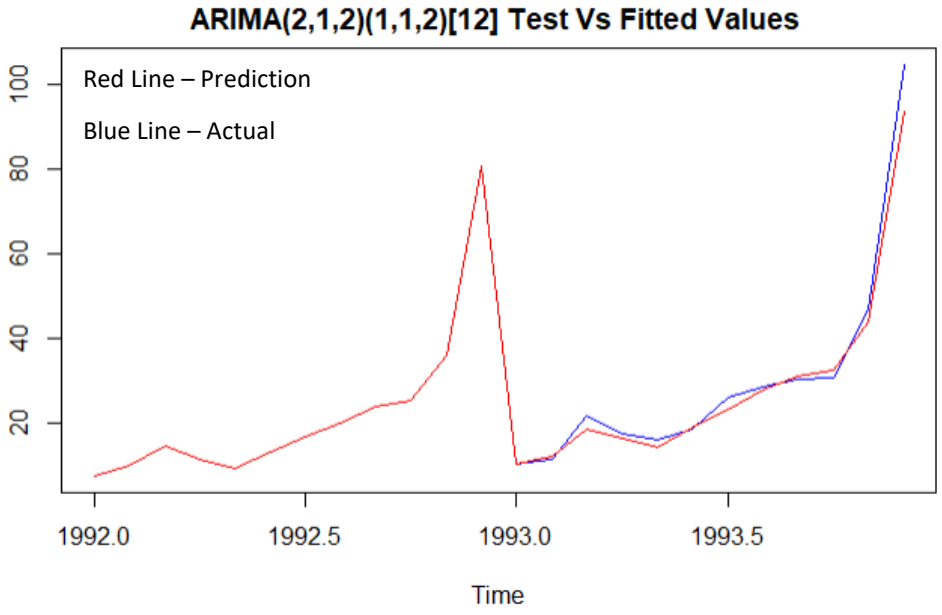
```
library(forecast)
ntest = Arima(Fancy_Test, model=Fancy_Train_ArimaSeasonal)
Fancy_ntest = cbind(Fancy_Test, ntest$fitted)
ts.plot(Fancy_ntest, col = c("blue", "red"), main = "ARIMA(2,1,2)(1,1,2)[12] Model Test Data Validation
using Arima func.")
```

Based on various parameter tuning of pdq, PDQ, the model with the least MAPE value is ARIMA (2,1,2)(1,1,2)[12]

Model	MAPE	RMSE
ARIMA(1,1,1)(1,1,1)[12]	4.02	3234
ARIMA(2,1,1)(0,1,0)[12]	4.6	3639
ARIMA(2,1,1)(1,1,2)[12]	3.26	2580

ARIMA(2,1,2)(1,1,2)[12]	3.22	2553

Forecast on the test data using ARIMA Model (2,1,2) (1,1,2) period 12



5.a) Use both the models to predict the values for next 5 years using HW model

Error Measures and Smoothing Parameters using Multiplicative Method

```
Model Information:
Holt-Winters' multiplicative method

Call:
 hw(y = Fancy.ts, h = 60, seasonal = "m")

  Smoothing parameters:
    alpha = 0.4165
    beta  = 0.0243
    gamma = 6e-04

  Initial states:
    l = 3957.9459
    b = 186.0517
    s = 3.176 1.4775 0.9811 0.9504 0.88 0.8064
           0.6413 0.5762 0.6646 0.914 0.5146 0.4179

  sigma:  0.1946

     AIC     AICc      BIC
1661.938 1671.211 1703.262

Error measures:
                    ME    RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 111.4845 1573.73 1264.728 -1.223233 13.64245 0.2857896 0.1985004

Forecasts:
```

*

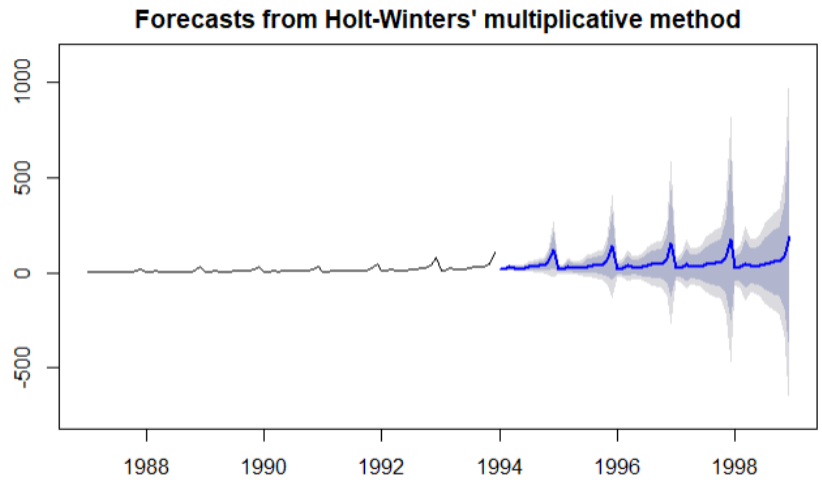
Plot of the forecasted data for the next 5 Years

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Jan 1994	14032.09	10533.0425	17531.14	8680.7577	19383.43
Feb 1994	17595.08	12799.7838	22390.38	10261.3060	24928.86
Mar 1994	31786.50	22391.2096	41181.78	17417.6456	46155.35
Apr 1994	23524.41	16030.9603	31017.86	12064.1678	34984.65
May 1994	20766.35	13675.1365	27857.55	9921.2771	31611.41
Jun 1994	23489.06	14929.1900	32048.93	10397.8691	36580.25
Jul 1994	30022.75	18392.0569	41653.45	12235.1382	47810.37
Aug 1994	33294.49	19629.5334	46959.45	12395.7408	54193.25
Sep 1994	36549.81	20704.3917	52395.22	12316.3359	60783.28
Oct 1994	38344.84	20832.0876	55857.60	11561.3954	65128.29
Nov 1994	58666.22	30506.2568	86826.19	15599.2703	101733.18
Dec 1994	128038.45	63584.2598	192492.64	29464.2738	226612.62
Jan 1995	17112.62	8095.0184	26130.22	3321.3881	30903.85
Feb 1995	21388.42	9612.1094	33164.74	3378.1050	39398.74
Mar 1995	38518.45	16394.8349	60642.07	4683.2996	72353.60
Apr 1995	28420.18	11417.0270	45423.32	2416.1041	54424.25
May 1995	25014.46	9446.9627	40581.95	1206.0295	48822.88
Jun 1995	28213.62	9971.8636	46455.37	315.2615	56111.98
Jul 1995	35961.96	11833.6395	60090.28	-939.1213	72863.04
Aug 1995	39774.13	12111.5701	67436.68	-2532.1020	82080.35
Sep 1995	43549.47	12184.3343	74914.61	-4419.3672	91518.31
Oct 1995	45572.94	11615.2375	79530.65	-6360.8855	97506.77
Nov 1995	69553.96	15981.5482	123126.37	-12377.9720	151485.89
Dec 1995	151438.98	30965.7317	271912.22	-32808.9517	335686.90
Jan 1996	20193.30	3613.3648	36773.23	-5163.5213	45550.12
Feb 1996	25181.95	3858.6956	46505.20	-7429.1535	57793.05
Mar 1996	45250.74	5760.8867	84740.59	-15143.7781	105645.26
Apr 1996	33316.18	3370.4391	63261.93	-12481.8806	79114.25
May 1996	29262.78	2188.5734	56336.99	-12143.6458	70669.21
Jun 1996	32938.42	1586.9306	64289.91	-15009.5453	80886.39
Jul 1996	41901.47	893.4745	82909.47	-20814.8472	104617.79
Aug 1996	46254.09	-267.5017	92775.69	-24894.5472	117402.74
Sep 1996	50549.51	-1675.6249	102774.64	-29321.9402	130420.96
Oct 1996	52801.43	-3209.2097	108812.06	-32859.4497	138462.30
Nov 1996	80442.27	-7133.8453	168018.40	-53493.8430	214378.39
Dec 1996	174840.78	-20433.2352	370114.79	-123805.0525	473486.60
Jan 1997	23274.14	-3383.5404	49931.83	-17495.2657	64043.55
Feb 1997	28975.69	-5046.0361	62997.41	-23056.0480	81007.42

1-38 of 60 rows

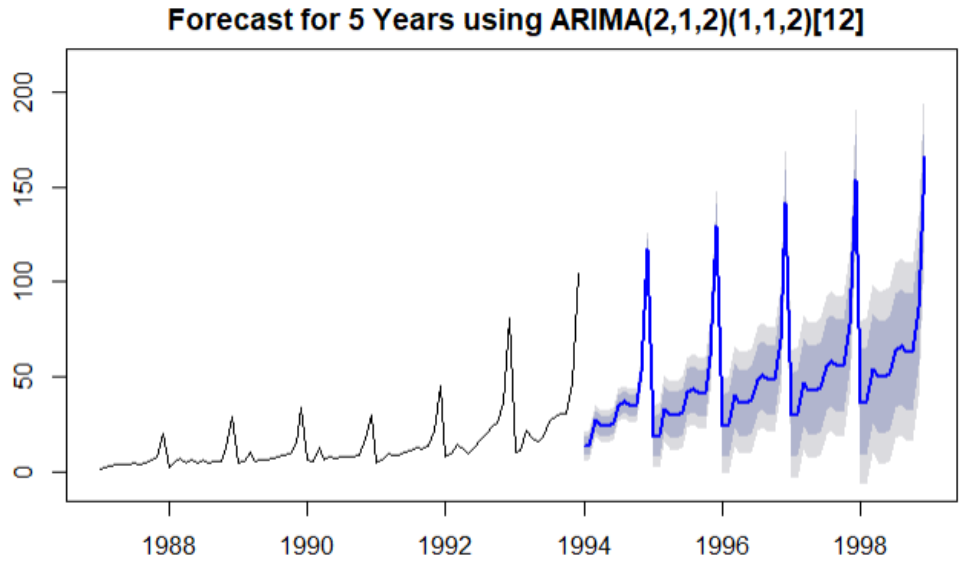
Previous 1 2 Next

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Mar 1997	51983.40	-10564.1452	114530.95	-43674.8170	147641.62
Apr 1997	38212.47	-8888.5477	85313.49	-33822.3202	110247.26
May 1997	33511.35	-8790.5867	75813.28	-31183.8778	98206.57
Jun 1997	37663.49	-11011.0879	86338.07	-36777.8548	112104.84
Jul 1997	47841.33	-15439.8387	111122.49	-48938.8649	144621.52
Aug 1997	52734.45	-18639.1306	124108.02	-56422.0194	161890.91
Sep 1997	57549.96	-22129.7329	137229.65	-64309.6140	179409.53
Oct 1997	60030.35	-24971.0383	145031.73	-69968.0527	190028.74
Nov 1997	91331.25	-40897.2335	223559.74	-110894.7658	293557.27
Dec 1997	198244.01	-95155.0899	491643.10	-250471.1901	646959.20
Jan 1998	26355.18	-13510.1086	66220.47	-34613.5173	87323.88
Feb 1998	32769.66	-17879.4730	83418.79	-44691.5040	110230.82
Mar 1998	58716.49	-33997.8052	151430.79	-83077.7886	200510.77
Apr 1998	43109.07	-26419.1298	112637.26	-63225.1336	149443.27
May 1998	37760.19	-24434.7526	99955.12	-57358.7644	132879.14
Jun 1998	42388.87	-28900.9714	113678.72	-66639.5350	151417.28
Jul 1998	53781.57	-38559.2607	146122.41	-87441.5453	195004.69
Aug 1998	59215.22	-44564.1182	162994.56	-99501.5828	217932.03
Sep 1998	64550.87	-50909.4267	180011.18	-112030.4181	241132.17
Oct 1998	67259.75	-55506.0625	190025.56	-120494.3567	255013.85
Nov 1998	102220.97	-88147.5809	292589.51	-188922.6018	393364.53
Dec 1998	221648.83	-199463.2147	642760.88	-422386.4662	865684.13



5.b) Use both the models to predict the values for next 5 years using ARIMA model

```
Call:
arima(x = sales, order = c(2, 1, 2), seasonal = list(order = c(1, 1, 2), period = 12))

Coefficients:
NaNs produced        ar1     ar2      ma1     ma2    sar1     sma1     sma2
      0.749  0.0060  -1.4269  0.4440  0.9832  -0.2523  -0.6214
s.e.    NaN  0.0013   0.0727  0.0695     NaN   0.1036   0.1080

sigma^2 estimated as 14.45:  log likelihood = -201.95,  aic = 419.89

Training set error measures:
                     ME     RMSE      MAE        MPE     MAPE      MASE         ACF1
Training set 0.3533508 3.494717 2.176887 -1.845148 15.91193 0.3183582 -0.01129329
```



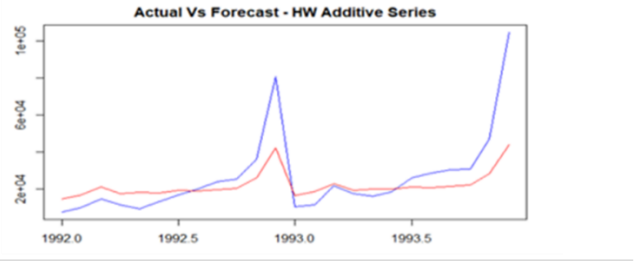
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Jan 1994	13.33797	8.392986	18.28296	5.7752673	20.90068
Feb 1994	13.80360	8.607111	19.00009	5.8562562	21.75094
Mar 1994	26.98418	21.624423	32.34394	18.7871378	35.18123
Apr 1994	23.93854	18.471123	29.40596	15.5768455	32.30024
May 1994	23.72341	18.181871	29.26495	15.2483565	32.19847
Jun 1994	25.39677	19.801994	30.99155	16.8402966	33.95325
Jul 1994	34.80157	29.166899	40.43625	26.1840837	43.41906
Aug 1994	36.84803	31.182187	42.51387	28.1828716	45.51318
Sep 1994	34.73545	29.044262	40.42663	26.0315318	43.43936
Oct 1994	35.15651	29.443959	40.86906	26.4199190	43.89310
Nov 1994	52.81881	47.087703	58.54992	44.0538368	61.58379
Dec 1994	117.22639	111.478808	122.97397	108.4362229	126.01655
Jan 1995	18.53216	7.787606	29.27671	2.0997831	34.96454
Feb 1995	19.09951	7.888383	30.31065	1.9535689	36.24546
Mar 1995	33.34079	21.811249	44.87034	15.7078762	50.97371
Apr 1995	29.90234	18.150822	41.65385	11.9299450	47.87473
May 1995	29.70881	17.795392	41.62222	11.4888106	47.92880
Jun 1995	31.42442	19.387672	43.46116	13.0158045	49.83303
Jul 1995	41.56623	29.431527	53.70093	23.0078053	60.12465
Aug 1995	43.70616	31.490528	55.92179	25.0239629	62.38836
Sep 1995	41.34028	29.055319	53.62525	22.5520519	60.12851
Oct 1995	41.59676	29.250475	53.94305	22.7147462	60.47878
Nov 1995	60.84527	48.443286	73.24725	41.8780735	79.81246
Dec 1995	129.14929	116.695726	141.60285	110.1032089	148.19537
Jan 1996	24.15364	7.915183	40.39210	-0.6809372	48.98822
Feb 1996	24.79200	8.049696	41.53430	-0.8131440	50.39714
Mar 1996	40.06012	22.952273	57.16797	13.8959256	66.22431
Apr 1996	36.22328	18.843901	53.60266	9.6438117	62.80275
May 1996	36.04177	18.451708	53.63184	9.1400891	62.94346
Jun 1996	37.79194	20.031598	55.55229	10.6298384	64.95405
Jul 1996	48.65309	30.749813	66.55637	21.2723903	76.03379
Aug 1996	50.88093	32.853490	68.90838	23.3103383	78.45153
Sep 1996	48.26299	30.124364	66.40163	20.5223524	76.00364
Oct 1996	48.35536	30.114541	66.59619	20.4584327	76.25230
Nov 1996	69.16166	50.824900	87.49842	41.1180064	97.20531
Dec 1996	141.29527	122.867092	159.72344	113.1118065	169.47873
Jan 1997	30.10321	8.235313	51.97110	-3.3408498	63.54726
Feb 1997	30.81062	8.385742	53.23550	-3.4852722	65.10652
Mar 1997	47.08774	24.241820	69.93366	12.1479212	82.02756
Apr 1997	42.85879	19.686782	66.03080	7.4202610	78.29732
May 1997	42.68878	19.253619	66.12393	6.8477965	78.52975
Jun 1997	44.47268	20.817017	68.12834	8.2944657	80.65089
Jul 1997	56.04087	32.194029	79.88772	19.5702717	92.51148
Aug 1997	58.35500	34.337308	82.37270	21.6231078	95.08690
Sep 1997	55.48914	31.314692	79.66358	18.5175127	92.46076
Oct 1997	55.42008	31.098579	79.74158	18.2235547	92.61661
Nov 1997	77.75789	53.295916	102.21987	40.3465276	115.16926
Dec 1997	153.65662	129.058760	178.25449	116.0374371	191.27581
Jan 1998	36.37235	8.469419	64.27528	-6.3014988	79.04619
Feb 1998	37.14763	8.628335	65.66693	-6.4688698	80.76414
Mar 1998	54.41676	25.417502	83.41601	10.0662247	98.76729

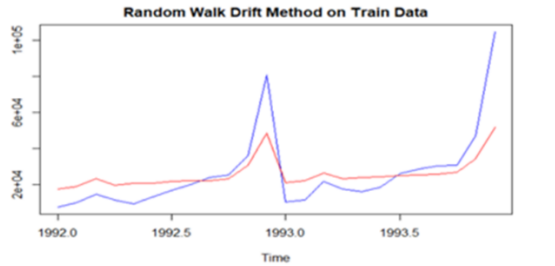
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 1998	49.80228	20.420167	79.18439	4.8662177	94.73834
May 1998	49.64355	19.944016	79.34308	4.2220329	95.06507
Jun 1998	51.46061	21.488572	81.43265	5.6223345	97.29888
Jul 1998	63.72395	33.510596	93.93731	17.5166114	109.93129
Aug 1998	66.12291	35.689963	96.55587	19.5797312	112.66610
Sep 1998	63.01329	32.375768	93.65081	16.1572451	109.86933
Oct 1998	62.78551	31.953613	93.61741	15.6321907	109.93884
Nov 1998	86.62908	55.609466	117.64870	39.1886734	134.06949
Dec 1998	166.22960	135.026673	197.43252	118.5088414	213.95036

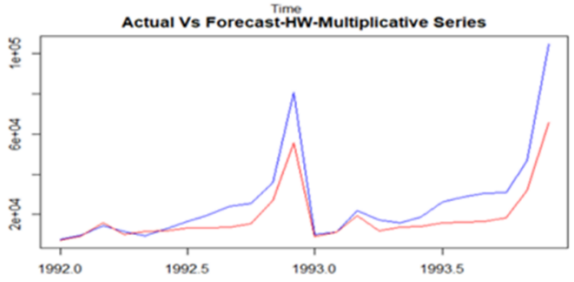
- Both the models Holt Winters and S ARIMA model are forecasting the series with very similar plot,
- However, with each forecast year (1996 & beyond) the variation in the confidence interval is increasing significantly for Holt Winter's model compared to ARIMA model
- It is always a good practice of not forecasting too ahead into the future.

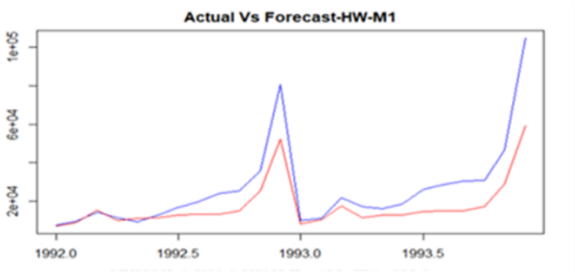
6.Summary

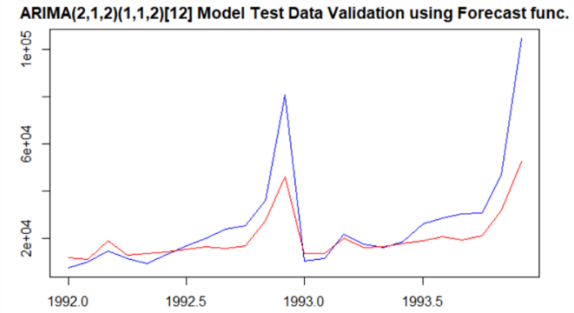
Actual Test Data Vs Forecasted Data for all models:

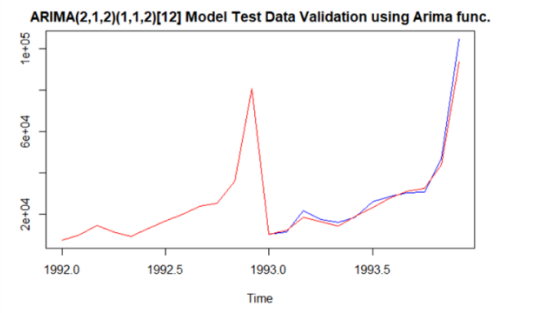












Sr.No:	TestData	HW-M	HW-M1	STL	S-Arima	S-Arima (Arima Method having a Drift term – one step forecast method)
1	7615.03	7398.233712	7338.964	17638.8005	11678.92	7610.633
2	9849.69	9270.708393	9101.909	18813.77781	10948.86	9845.993
3	14558.4	15993.62639	15325.25	23053.28466	19073.84	14552
4	11587.33	10070.05687	9978.898	19705.06507	12712.85	11585.18
5	9332.56	11468.65099	11185.34	20534.75688	13734.92	9332.884
6	13082.09	11860.745	11434.66	20812.76664	14445.93	13078.91
7	16732.78	13363.38434	12905.38	21807.78003	15327.47	16726.66
8	19888.61	13532.93064	13194.97	22157.94248	16604.09	19880.29
9	23933.38	13840.57584	13278.32	22217.23362	15809.2	23922.16
10	25391.35	15642.25103	15076.48	23318.56716	16875.37	25379.87
11	36024.8	26945.90199	25436.67	30791.39045	27428.26	36004.21

12	80721.71	55516.69457	52163.3	48412.0224	46075.23	80654.33
13	10243.24	8978.434944	8472.109	20990.1329	13544.36	10340.27
14	11266.88	11216.22624	10489.4	22165.11021	13411.89	11977.86
15	21826.84	19292.30522	17632.13	26404.61706	20092.49	18396.37
16	17357.33	12111.90817	11462.39	23056.39747	16230.24	16523.26
17	15997.79	13755.44811	12827.84	23886.08929	16487.12	14352.79
18	18601.53	14187.06955	13093.48	24164.09904	17992.28	19068.03
19	26155.15	15942.28055	14755.18	25159.11243	18952.33	23315.21
20	28586.52	16103.21199	15063.96	25509.27488	20664.26	27840.25
21	30505.41	16428.33096	15137.17	25568.56602	19359.3	30943.73
22	30821.33	18521.99527	17162.74	26669.89956	21054.92	32410.79
23	46634.38	31831.69732	28916.42	34142.72285	31802.15	43951.17
24	104660.7	65433.07514	59218.83	51763.3548	52499.43	93610.97

Seasonal Arima Model **ARIMA (2,1,2) (1,1,2) [12]** is the best performing model for the following reasons: -

- The model validates the test data in the best possible manner, with the forecasted values closely fitting the actual test data
- The Mean absolute percentage error is the lowest amongst all models
- The Root Mean Square error is the lowest for this model

The Arima function having a drift term is predicting the test period more accurately than the forecast function when applied on the train data for a forecasting period of 24 months (test period).