Table of Contents

Contents

1. Project objective

- Using the demand data for two products to build a forecasting models and choose the best model based the accuracy on the validation data and test scores.
- Based on the accuracy of the models choose the best model to make forecast, along with the confidence intervals, for the test dataset.

2. Exploratory Analysis

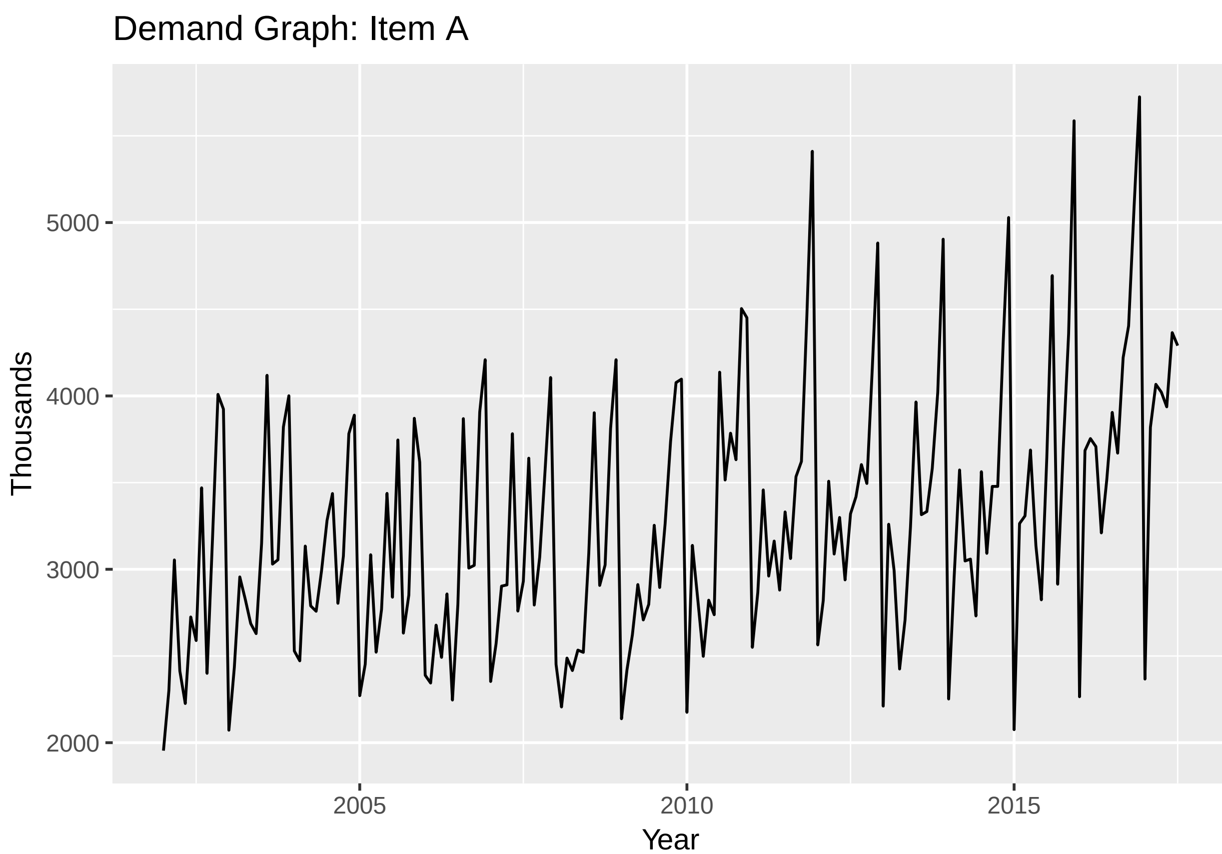
- An explanatory model is useful because it incorporates information about other variables, rather than only historical values of the variable to be forecast. However, there are several reasons a forecaster might select a time series model rather than an explanatory or mixed model.
- First, the system may not be understood, and even if it was understood it may be extremely difficult to measure the relationships that are assumed to govern its behaviour.
- Second, it is necessary to know or forecast the future values of the various predictors in order to be able to forecast the variable of interest, and this may be too difficult.
- Third, the main concern may be only to predict what will happen, not to know why it happens. Finally, the time series model may give more accurate forecasts than an explanatory or mixed model.

```
Warning: package 'forecast' was built under R version 3.4.4
```

```
# A tibble: 6 x 4
   Year Month `Item A` `Item B`
  <dbl> <dbl>    <dbl>    <dbl>
1  2002  1.00     1954     2585
2  2002  2.00     2302     3368
3  2002  3.00     3054     3210
4  2002  4.00     2414     3111
5  2002  5.00     2226     3756
6  2002  6.00     2725     4216
```

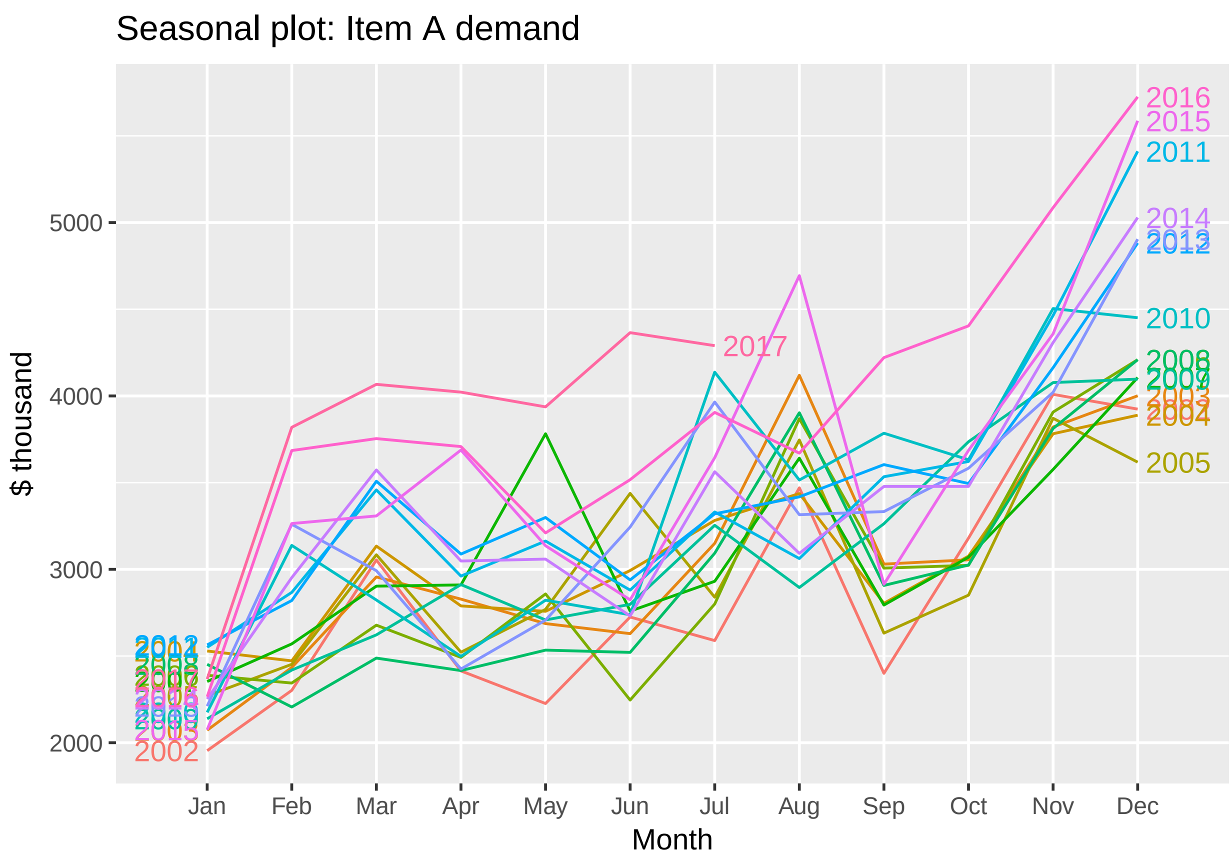
Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis?

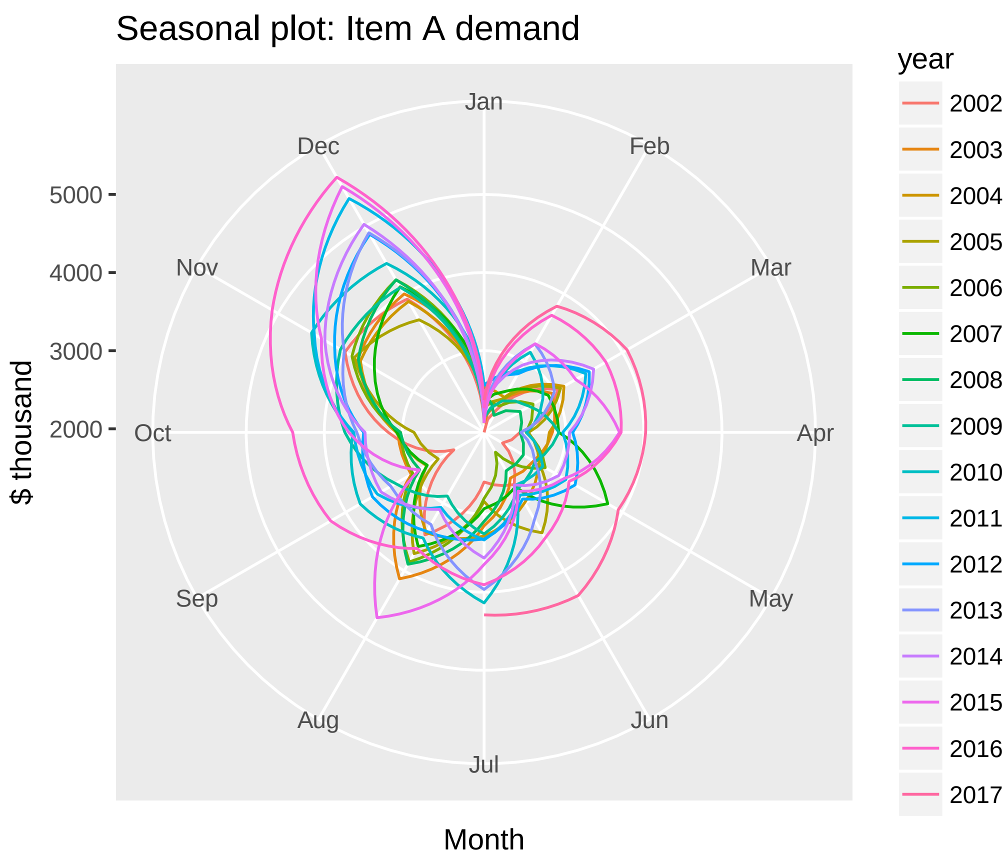
Item A visualization



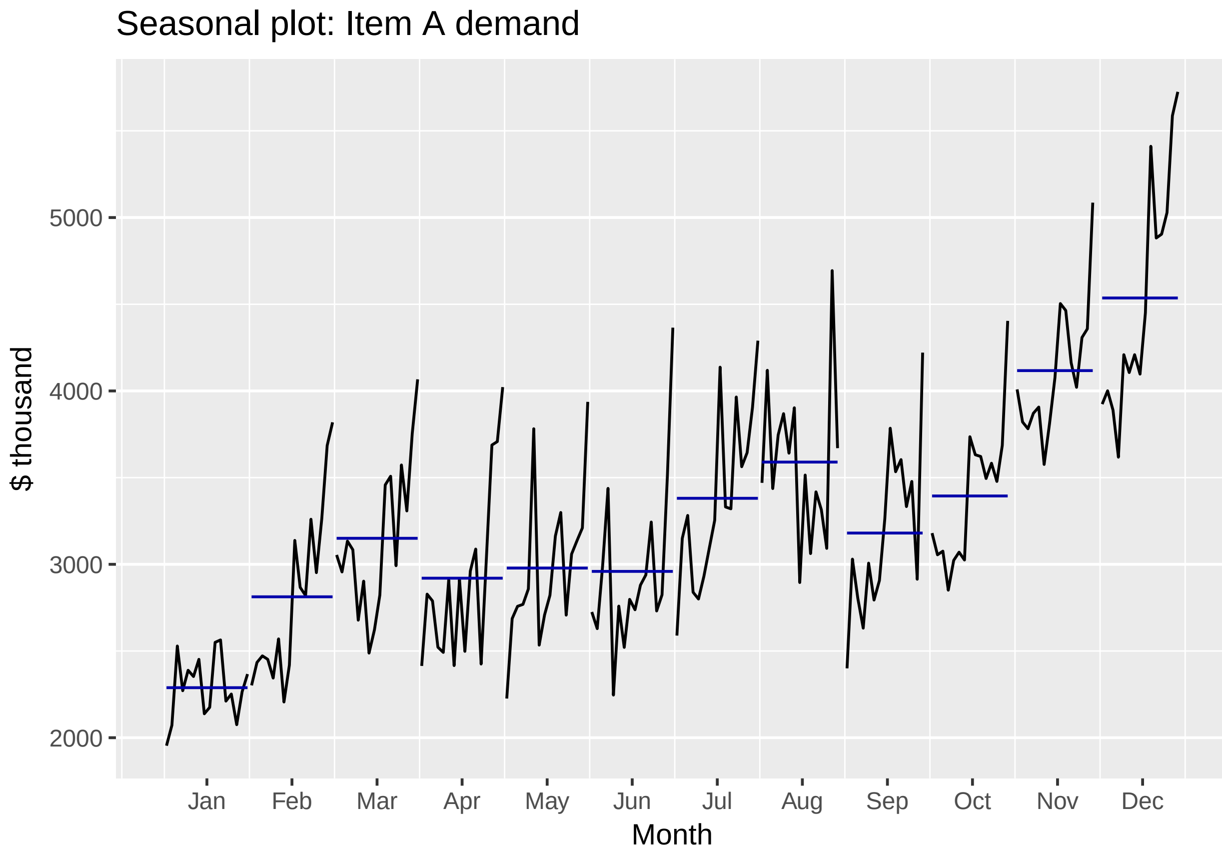
Observation:

- There is every slow trend in the data but definitely there is seasonality, with increase in season demand towards the latter part
- The Seasonality is additive for the demand of Item A


Seasonal plot: Item A demand
$ thousand
Month
5000
4000
3000
2000
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2016
2015
2011
2014
2010
2005
2017
2002



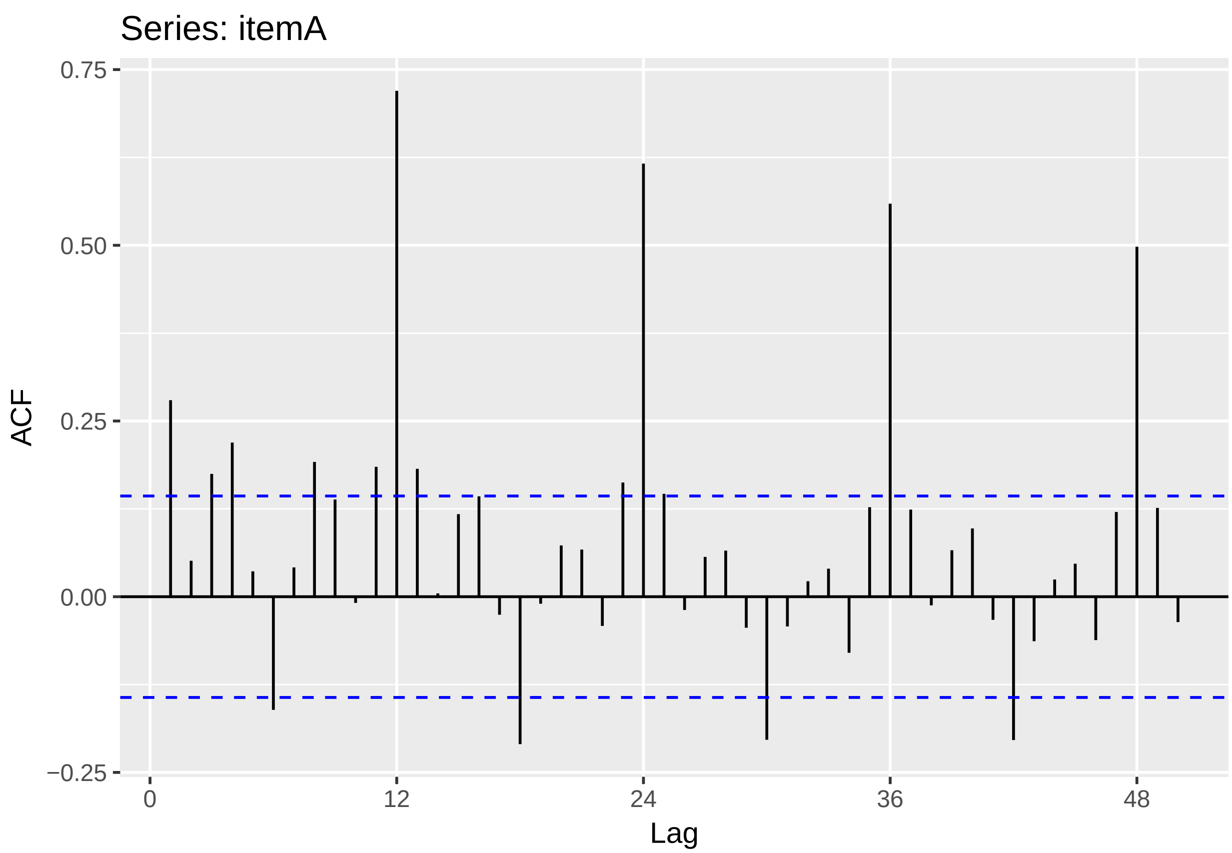
Seasonal sub-series plots



Observation:

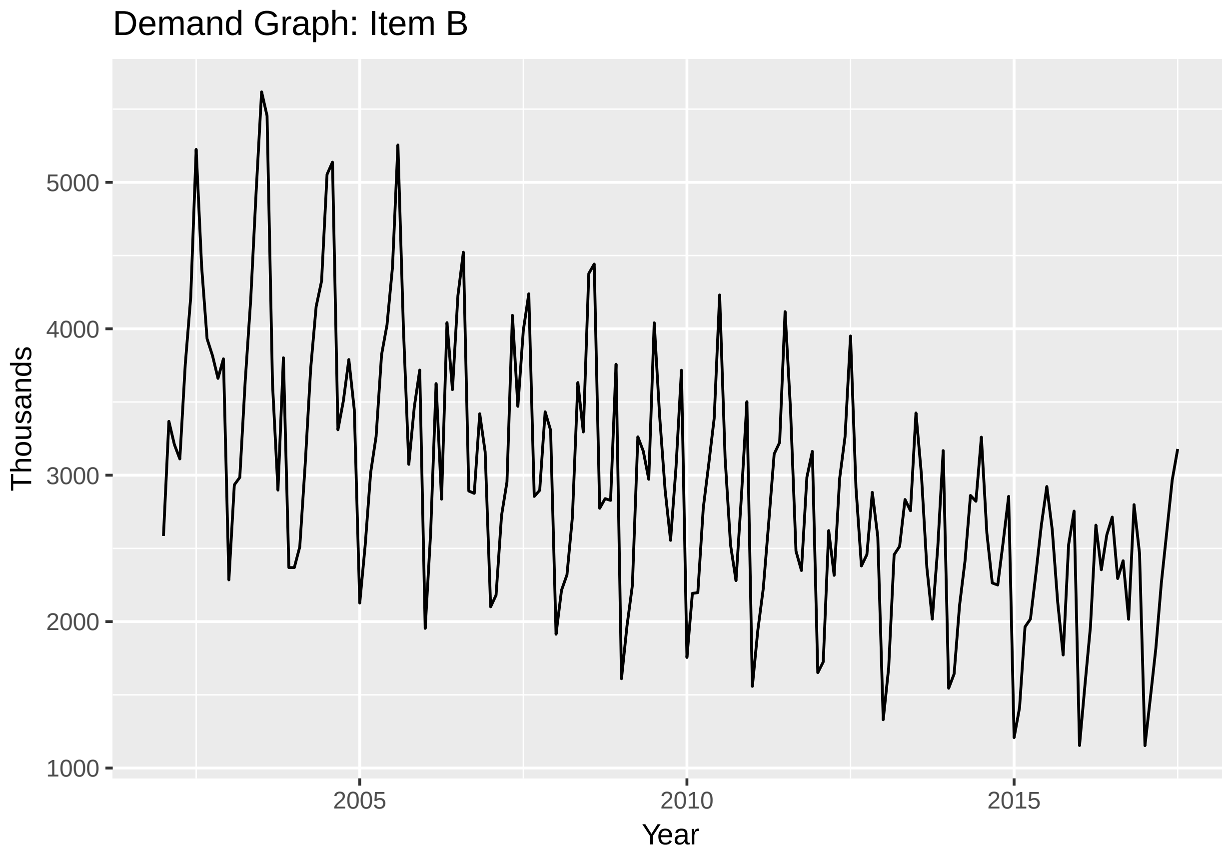
- The month of December has higher sales compared to all other months
- The month of January has lower sales compared to all other months

Correlogram of Item A



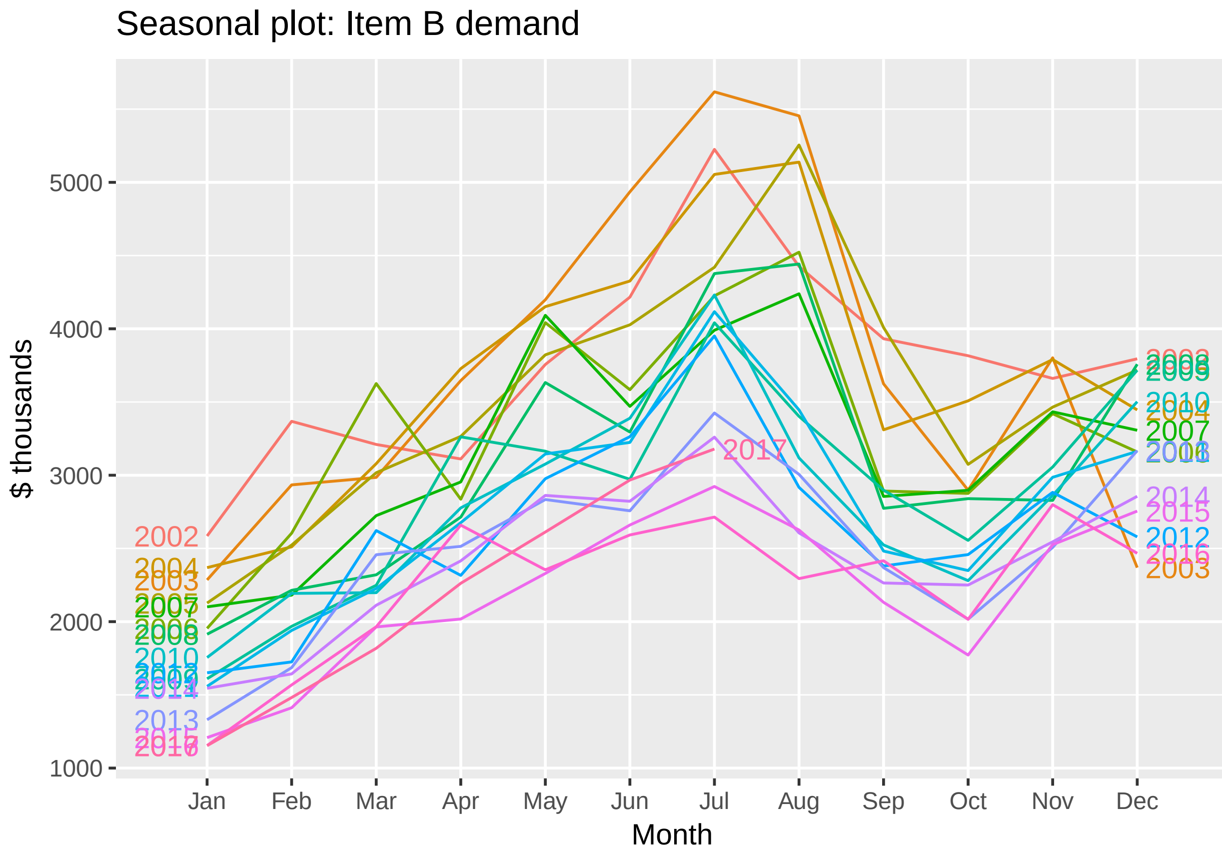
Observation:

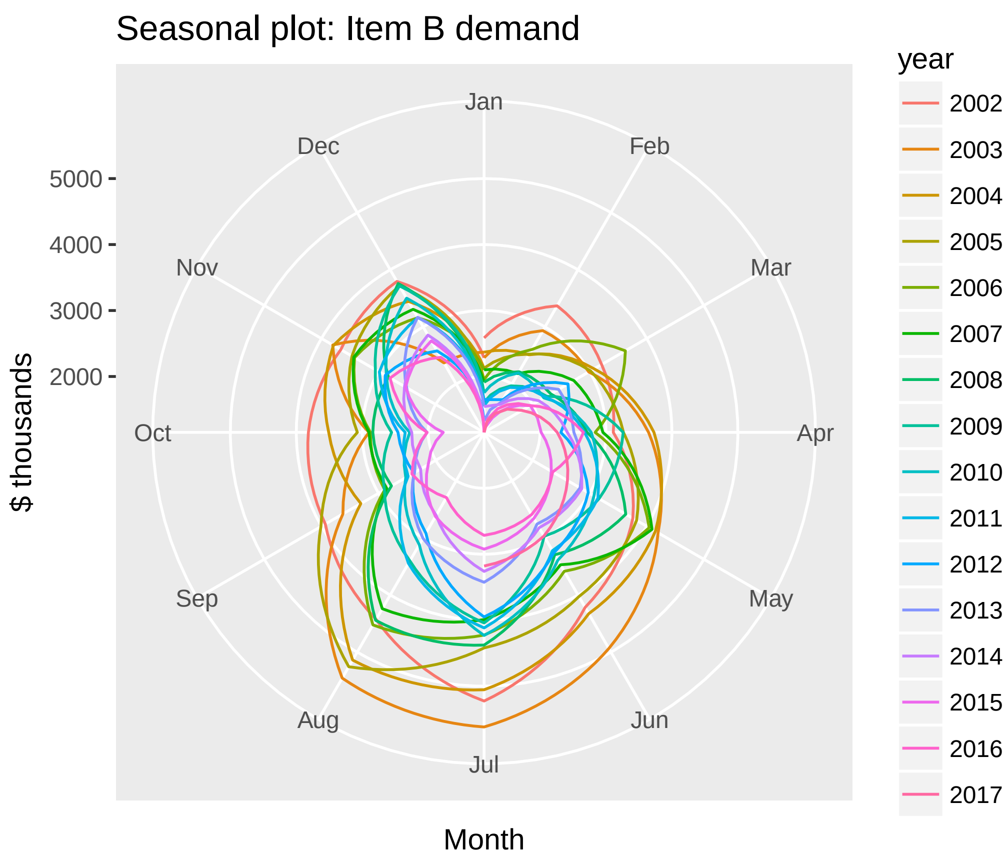
- Lag 6 and one apart multiple of 6 is lower than for the other lags. This is due to the seasonal pattern in the data.
- Lag 12 and its multiple is higher than for the other lags this due larger effect of the seasonality as there is no trend in data the correlations seems to similar and continuing the same pattern



Observation:

- Here, there is a clear and decreasing trend. There is also a strong seasonal pattern that decreases in size as the level of the series decreases.
- The trend is changing slowly and seasonality is multiplicative

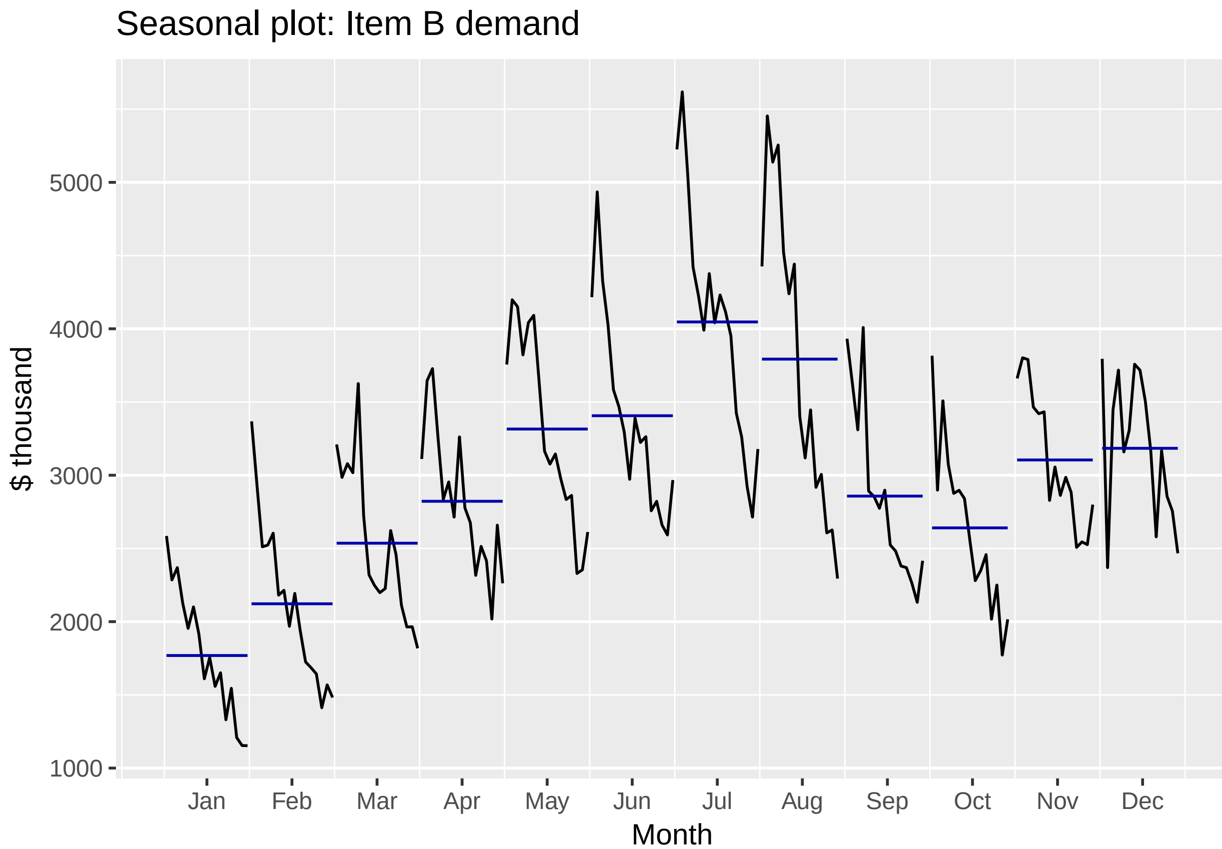

Seasonal plot: Item B demand
$ thousands
Month
5000
4000
3000
2000
1000
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2002
2004
2003
2007
2010
2012
2013
2017
2014
2015
2016



Observation:

- There is large jump in demand in July and decrease in demand in October.
- There is large drop in demand in January

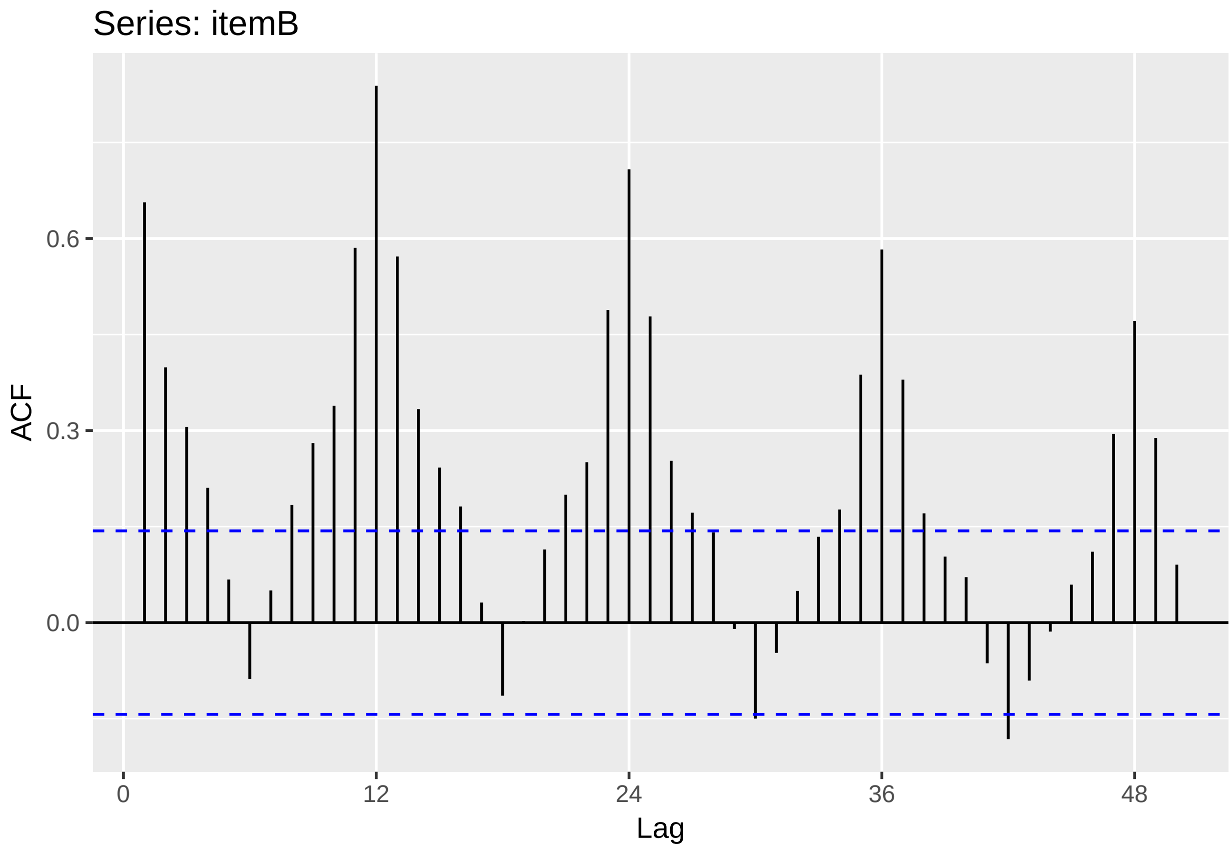
Seasonal sub-series plots



Observation

- The month of July has higher sales compared to other months
- The month of January has lower sales compared to other months

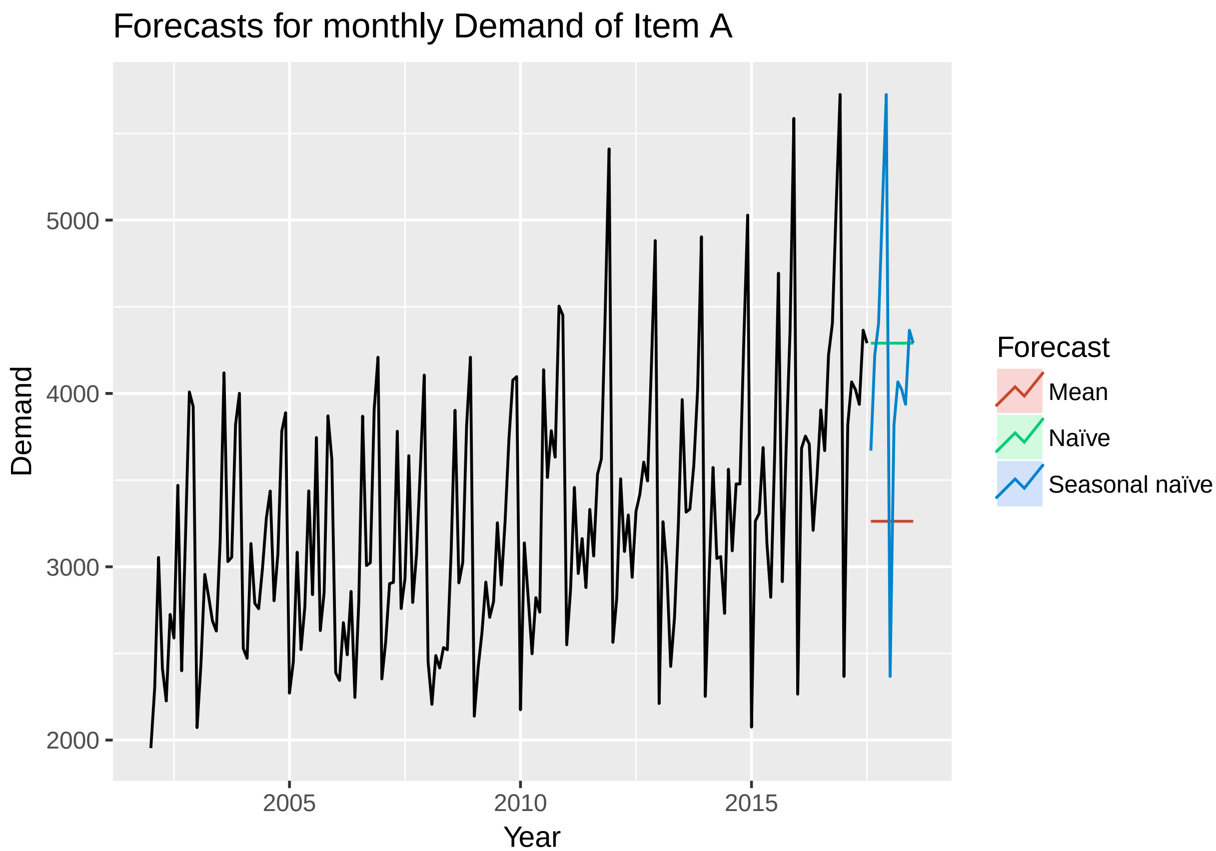
Correlogram of Item B

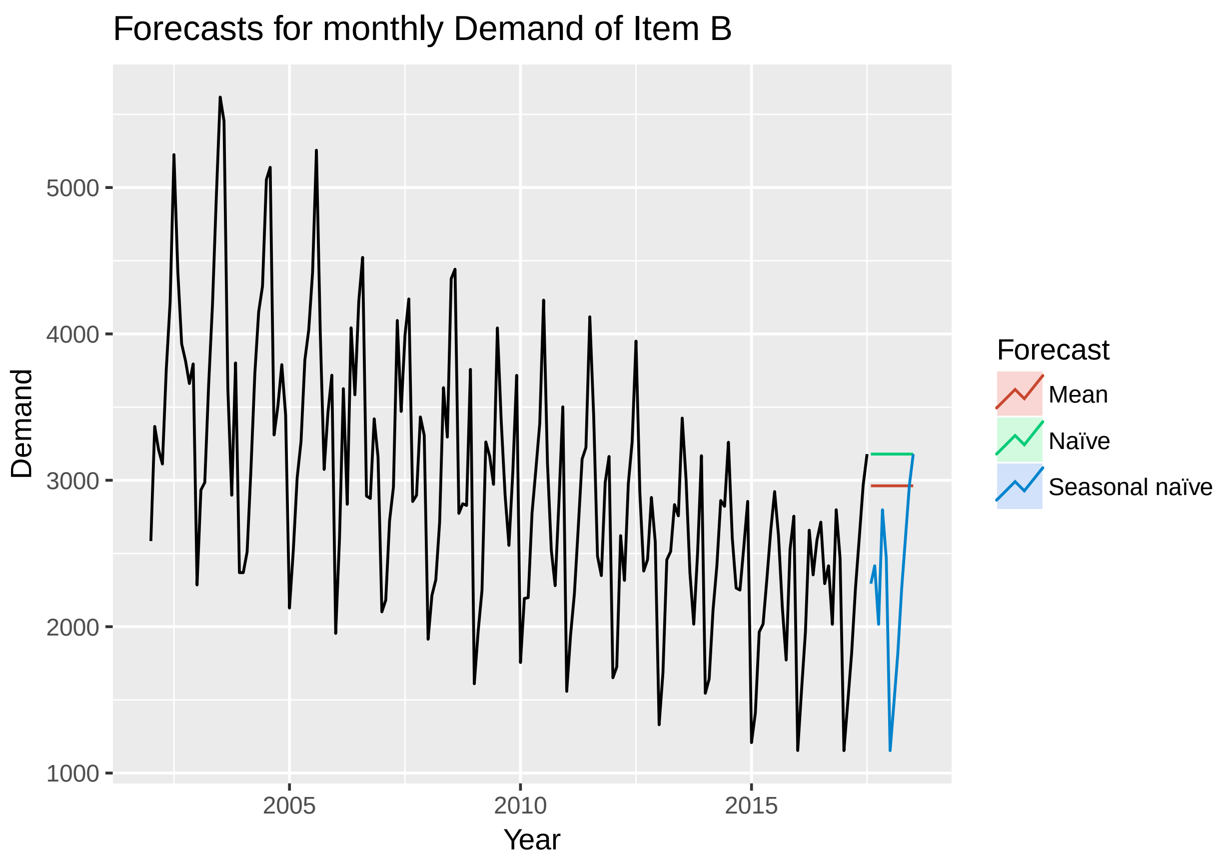


Observation:

- The lag 6 through is negative and same for multiples of 6 as the series has decreasing trend the throughs seems to increasing
- The decreasing highs are because of the seasonality which keeps decreasing the trend
- The slow decrease in the ACF as the lags increase is due to the trend, while the "scalloped" shape is due the seasonality.

3. Forecast using simpler methods

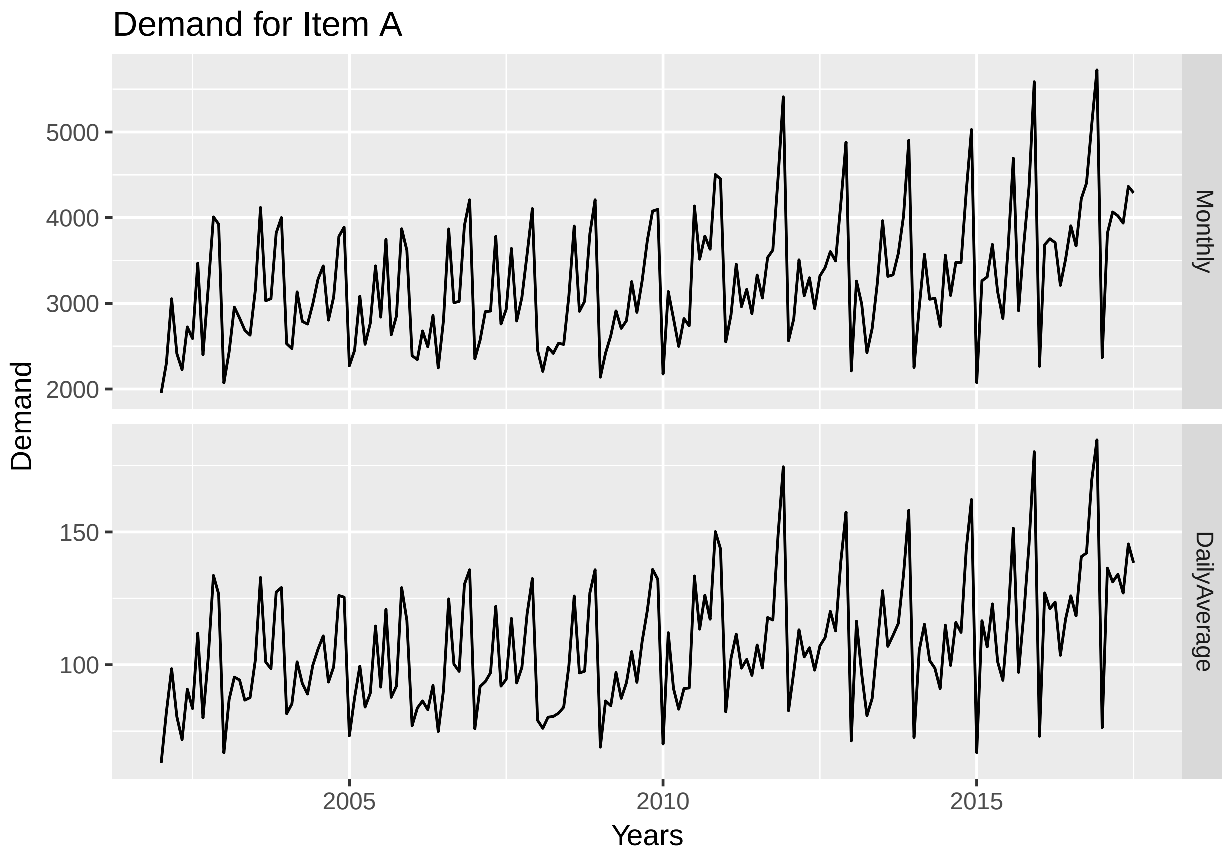


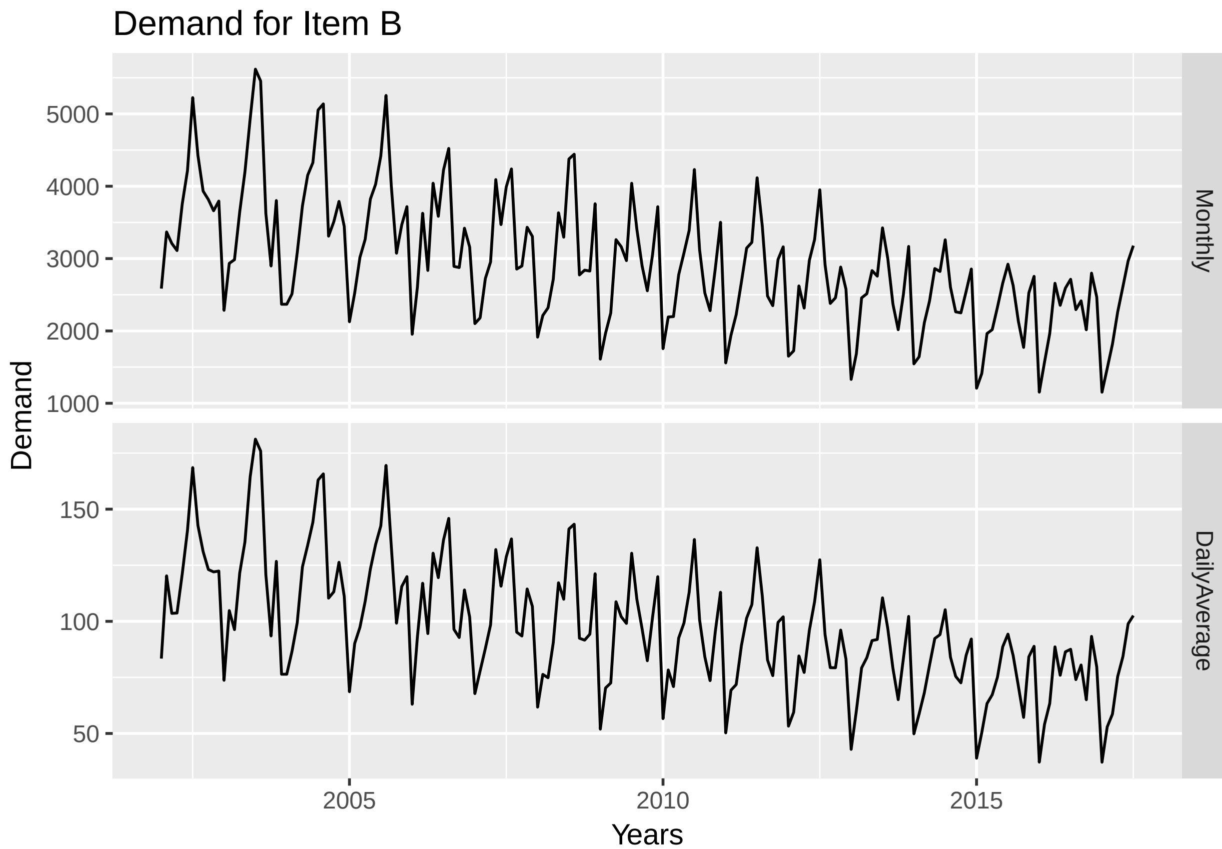


Observation:

- will serve as benchmarks for the forecast given by other methods for the series

Calender adjusted demand

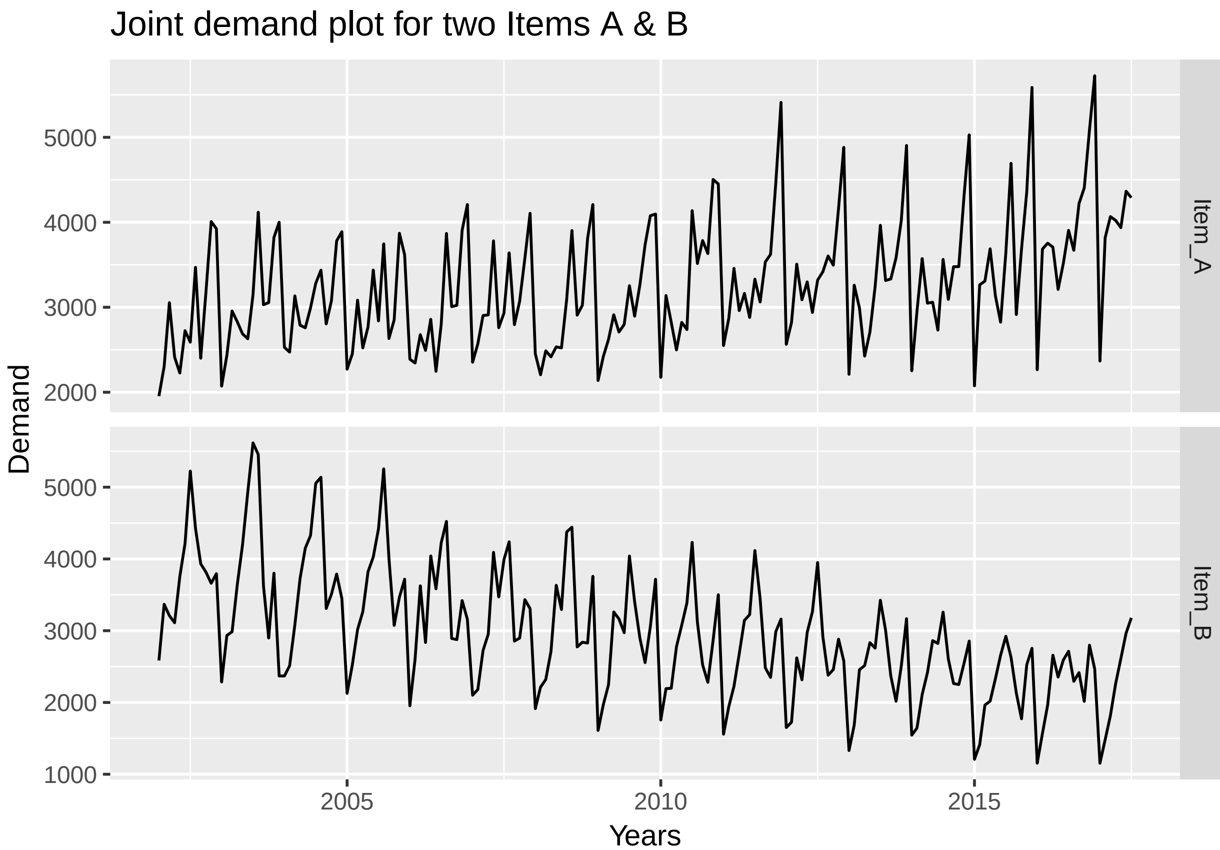




Observation:

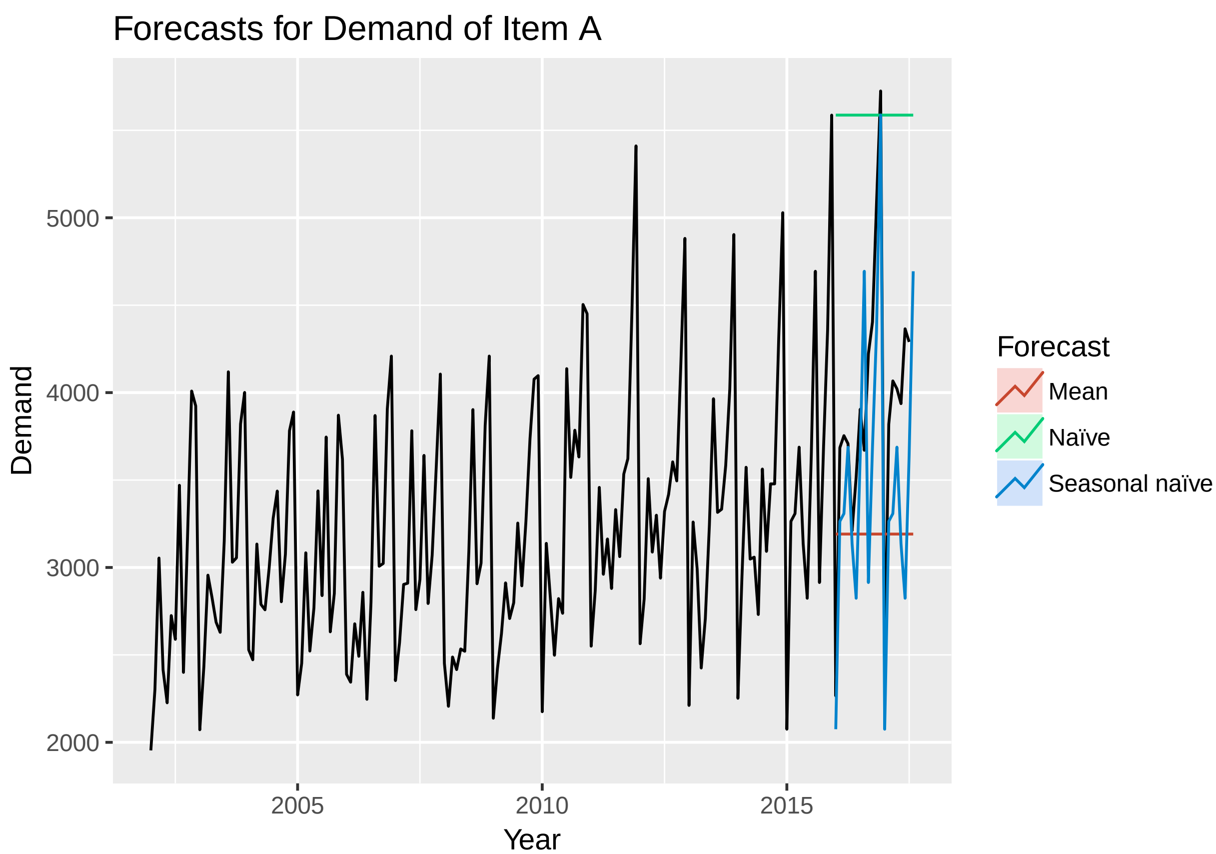
- The difference in number of days in some months has no major effect on the series, the pattern of the monthly and calender adjusted series for difference in number of days looks closely the same.

Combined plot



Observation:

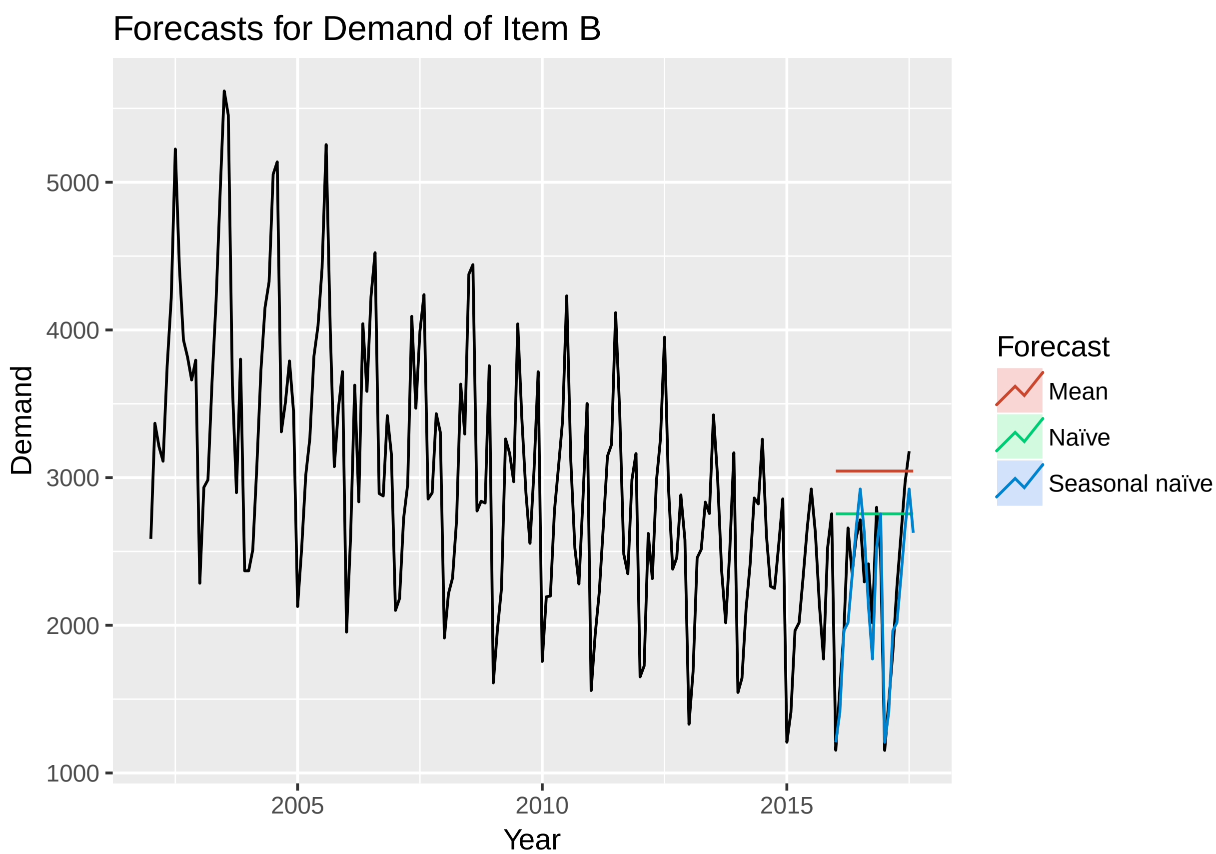
- The two series have evident seasonality, and Item A in particular has its series majorly influenced by seasonality
- Item A series has no evident trend, whereas Item B has downward trend
- There is no cyclicality present in the series of both the Items
- Both the series exhibit multiplicative seasonality which changes with time and trend




```
                       ME      RMSE      MAE         MPE     MAPE     MASE
Training set -1.949975e-13  685.2301 548.8301  -4.407797 17.61875 1.835592
Test set     -9.557021e+02 1112.8129 955.7021 -55.276710 55.27671 3.196397
                  ACF1 Theil's U
Training set 0.2507729        NA
Test set     0.4697821  2.627487
                     ME      RMSE       MAE          MPE     MAPE      MASE
Training set    21.75449  815.0209  599.4311    -2.89952  20.2945  2.004829
Test set     -3351.68421 3399.8196 3351.6842 -171.86654 171.8665 11.209887
                   ACF1 Theil's U
Training set -0.2993938        NA
Test set      0.4697821   7.54672
                   ME      RMSE       MAE          MPE      MAPE   MASE
Training set    57.250  400.6822  298.9936   0.9460036  9.479604 1.0000
Test set     -1191.895 1411.7852 1206.9474 -59.8300081 60.337343 4.0367
                  ACF1 Theil's U
Training set 0.1741347        NA
Test set     0.1190562  3.150099
```

Observation

- The accuracy scores are not good on the test set from any of the simple methods for Item A




```
                        ME     RMSE      MAE        MPE     MAPE     MASE
Training set -1.085922e-13 867.2231 687.2523  -8.971564 25.37329 2.353087
Test set     -8.092080e+02 989.8504 823.3634 -48.148228 48.59351 2.819119
                  ACF1 Theil's U
Training set 0.6368737        NA
Test set     0.4697821  2.349504
```

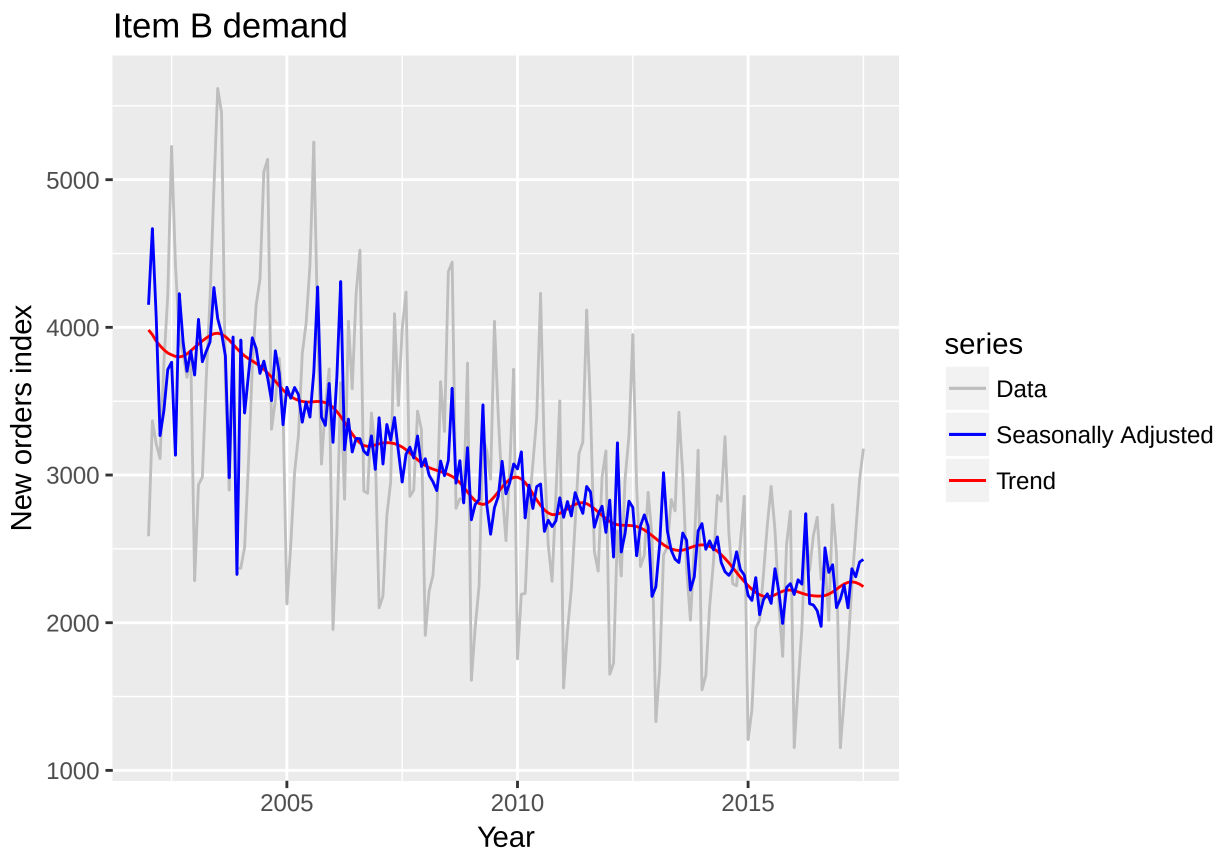
```
                      ME     RMSE      MAE        MPE     MAPE     MASE
Training set    1.017964 740.0684 577.3653  -3.829999 21.69407 1.976844
Test set     -519.684211 771.4000 591.2632 -34.059838 36.38139 2.024429
                   ACF1 Theil's U
Training set -0.1204740        NA
Test set      0.4697821  1.831417
```

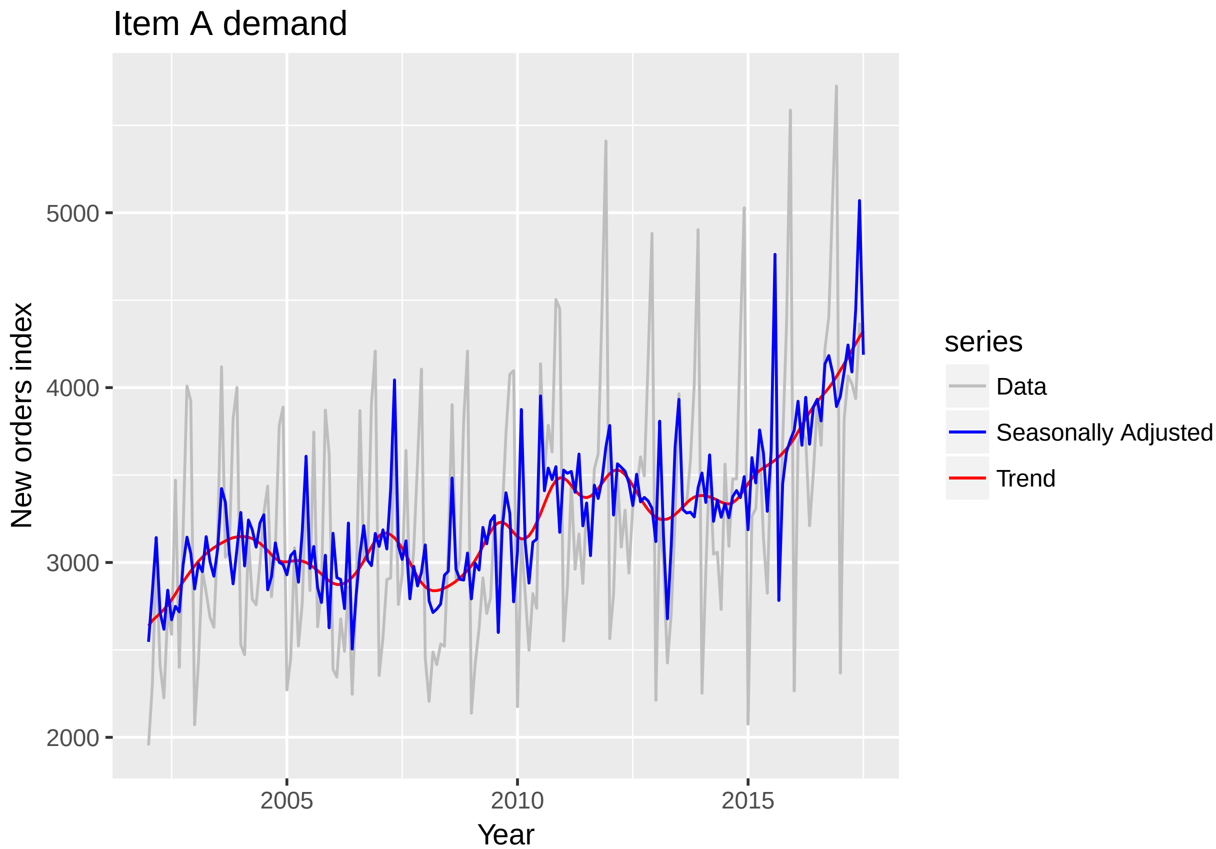
```
                    ME     RMSE      MAE       MPE      MAPE      MASE
Training set -120.35897 382.7072 292.0641 -5.072653 10.121136 1.0000000
Test set       85.89474 253.7720 207.1579  3.135471  8.814549 0.7092891
                   ACF1 Theil's U
Training set 0.08266974        NA
Test set     0.11422827 0.4989471
```

Observation

- The accuracy scores are somewhat good on the test set given by seasonal naive method from the simple methods for Item B

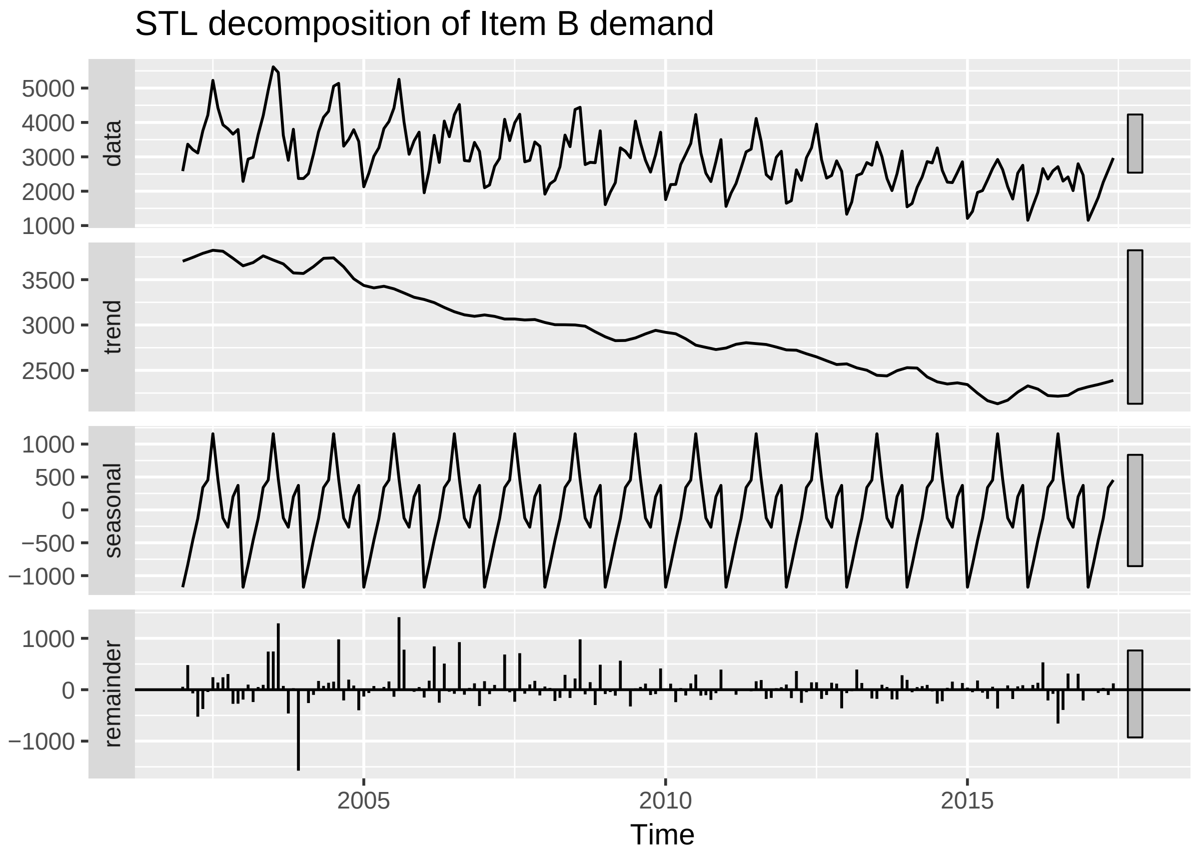
4. Decomposition of times series

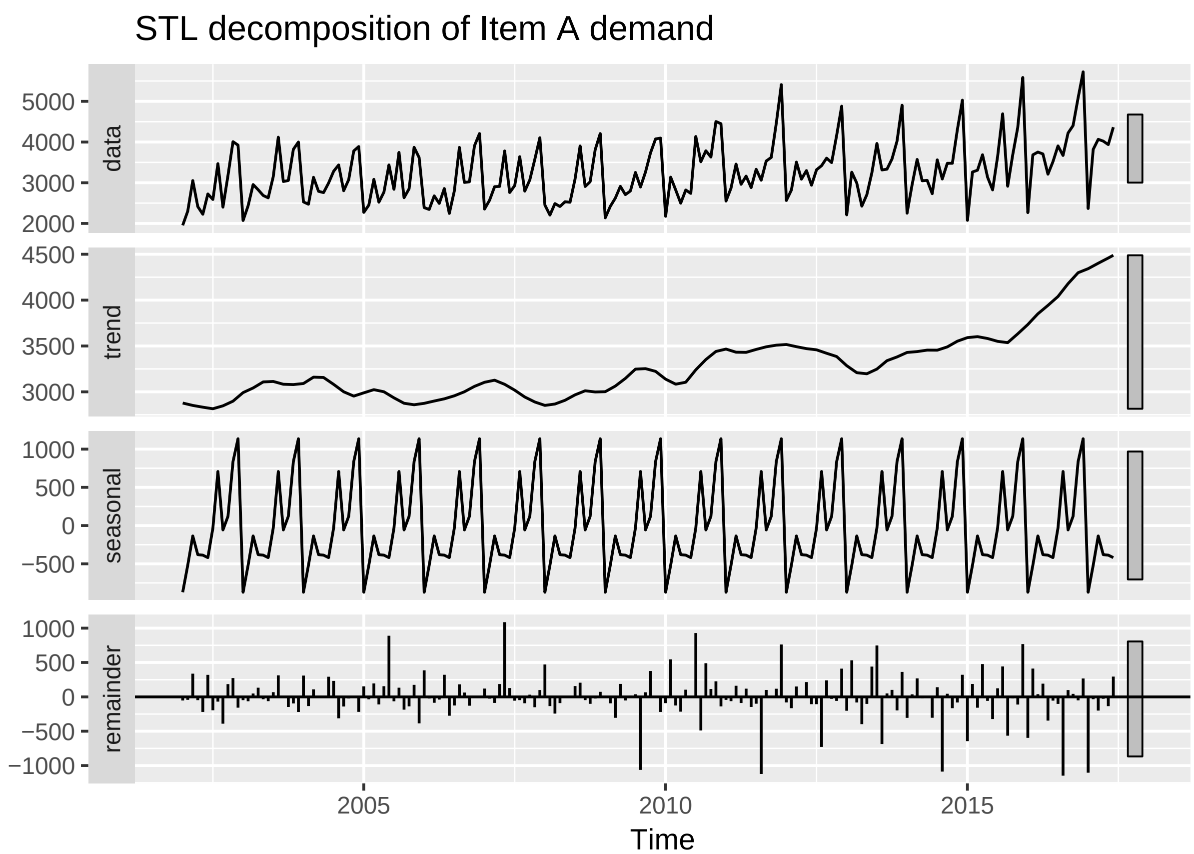




Observation:

- The Seasonally adjusted plots provides clear visuals of the trend and remainder in the each of the series
- The Trend in Item B is moving downwards from the start of the series
- The Trend in Item A is not that evident at the begining but then a slow upward trend picks in the mid-section of the series


STL decomposition of Item B demand
data
trend
seasonal
remainder
5000
4000
3000
2000
1000
3500
3000
2500
1000
500
0
−500
−1000
1000
0
−1000
2005
2010
2015
Time



Observation:

- The Item A series is a stable series with only seasonality initial but turns to have slow increasing trend in mid-section
- The Item B series has both trend and seasonality evident from the initial data points, having a medium decreaing trend with varying seasonality with time
- The seasonality in the both the series is multiplicative as it changes with time clearly evident for series B but visible for series A as from the mid-section of the series

5. Stationarity of the residuals

```
Box.test(remainder(itemA_decom), lag=10, type="Ljung")
```

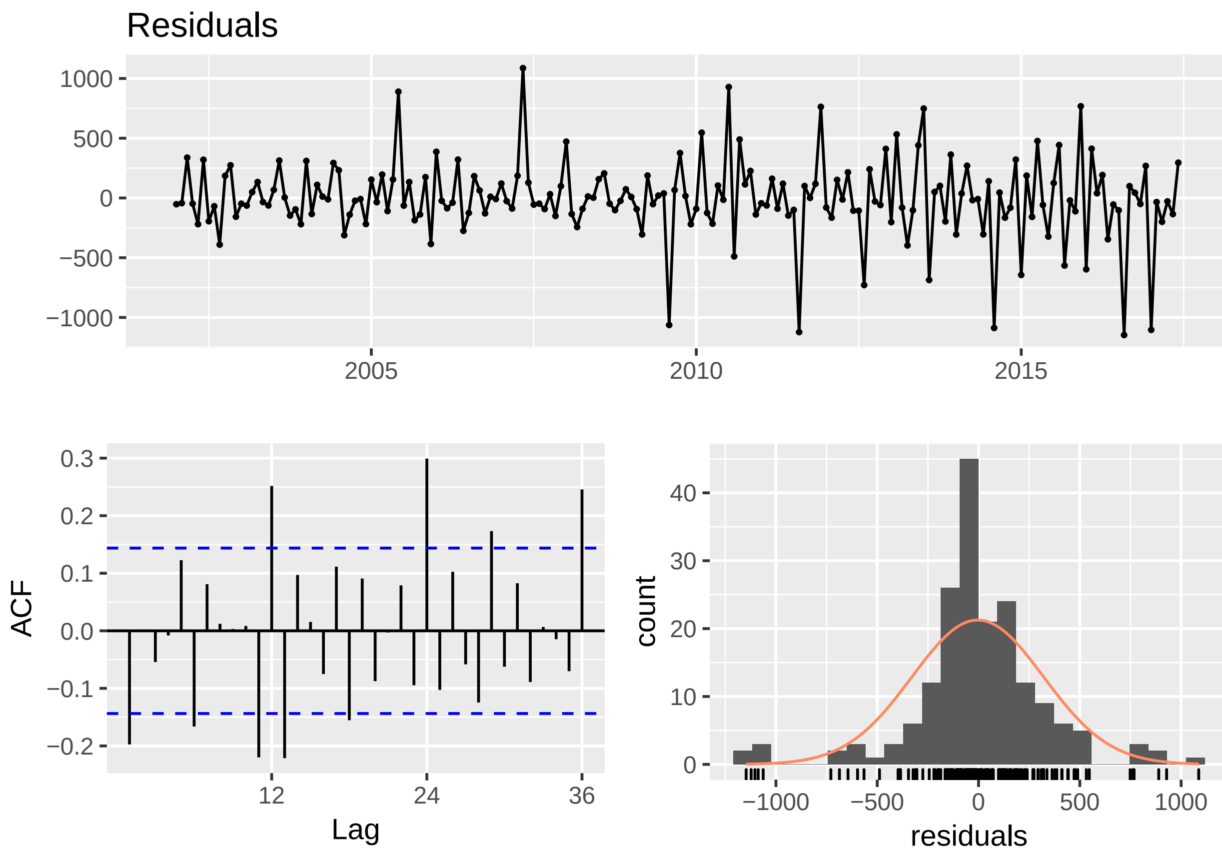
```
	Box-Ljung test

data:  remainder(itemA_decom)
X-squared = 17.546, df = 10, p-value = 0.06311
```

```
Box.test(remainder(itemB_decom), lag=10, type="Ljung")
```

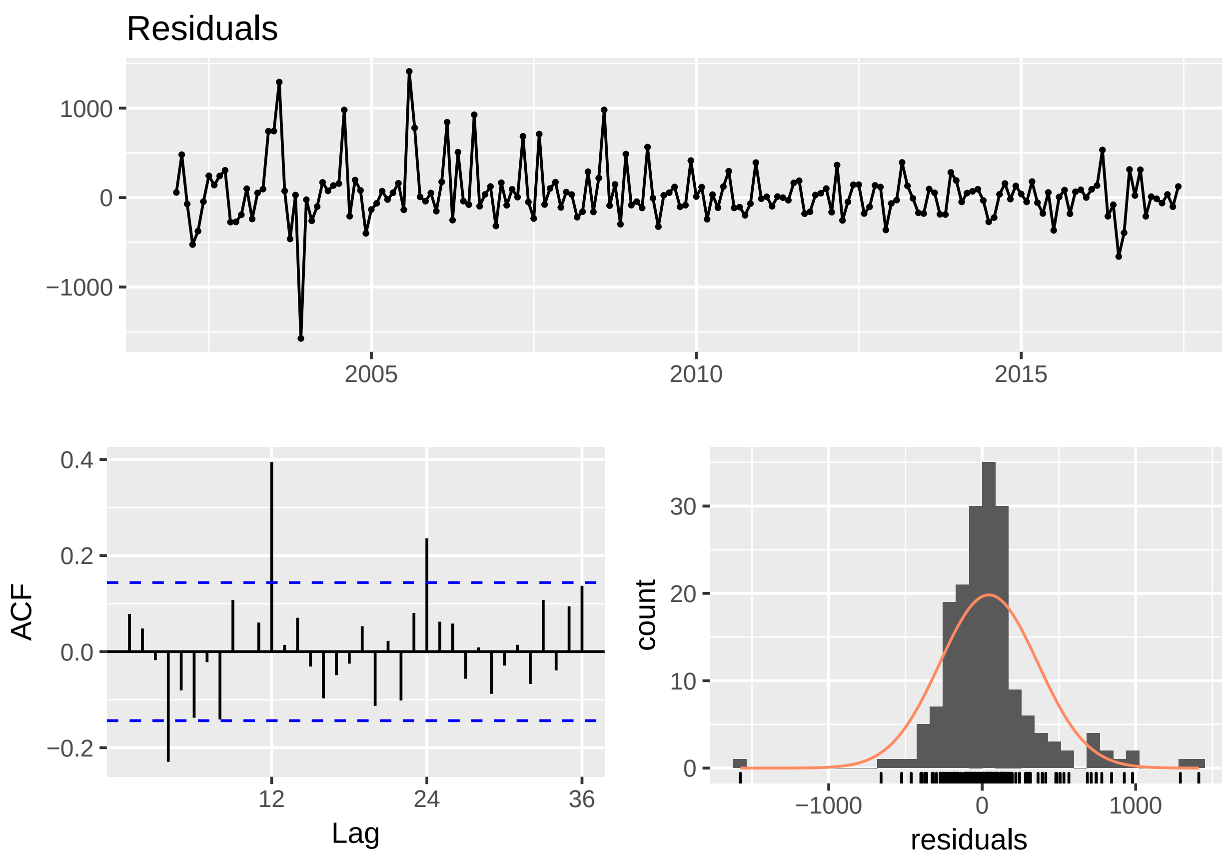
```
	Box-Ljung test
```

```
data:  remainder(itemB_decom)
X-squared = 22.998, df = 10, p-value = 0.01075
```

```
Item A residulas plots
```




```
Item B residulas plots
```



Observation:

- The ACF of the residuals of Item A looks just like that of a white noise series. There are no autocorrelations lying outside the 95% limits, and the Ljung-Box Q statistic has a p-value of 0.06311 for item A, but whereas it is not stationery for item B with p-value 0.01075. However at lag one both the series seems to be stationery
- The residuals of Item A is normally distributed, though with extended tails and is stationary
- The residuals of Item B is also normally distributed, but skewed to the right ans can be considered stationary as the lags in ACF are not significant
- A more clear confirmation can be obtained usin the Dickey-Fuller test

```
Warning: package 'tseries' was built under R version 3.4.4

Warning in adf.test(remainder(itemA_decom), k = 10): p-value smaller than
printed p-value


    Augmented Dickey-Fuller Test

data:  remainder(itemA_decom)
Dickey-Fuller = -5.4341, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary

Warning in adf.test(remainder(itemB_decom), k = 10): p-value smaller than
printed p-value


    Augmented Dickey-Fuller Test
```

```
data:  remainder(itemB_decom)
Dickey-Fuller = -5.6753, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

Observation:

- The H0: presence of a unit root; Ha: stationary series
- The Dickey-Fuller test statistic is very low, providing us with a low p-value. We can likely reject the null hypothesis of the presence of a unit root and conclude that we have a stationary series for both the residuals of ItemA and ItemB

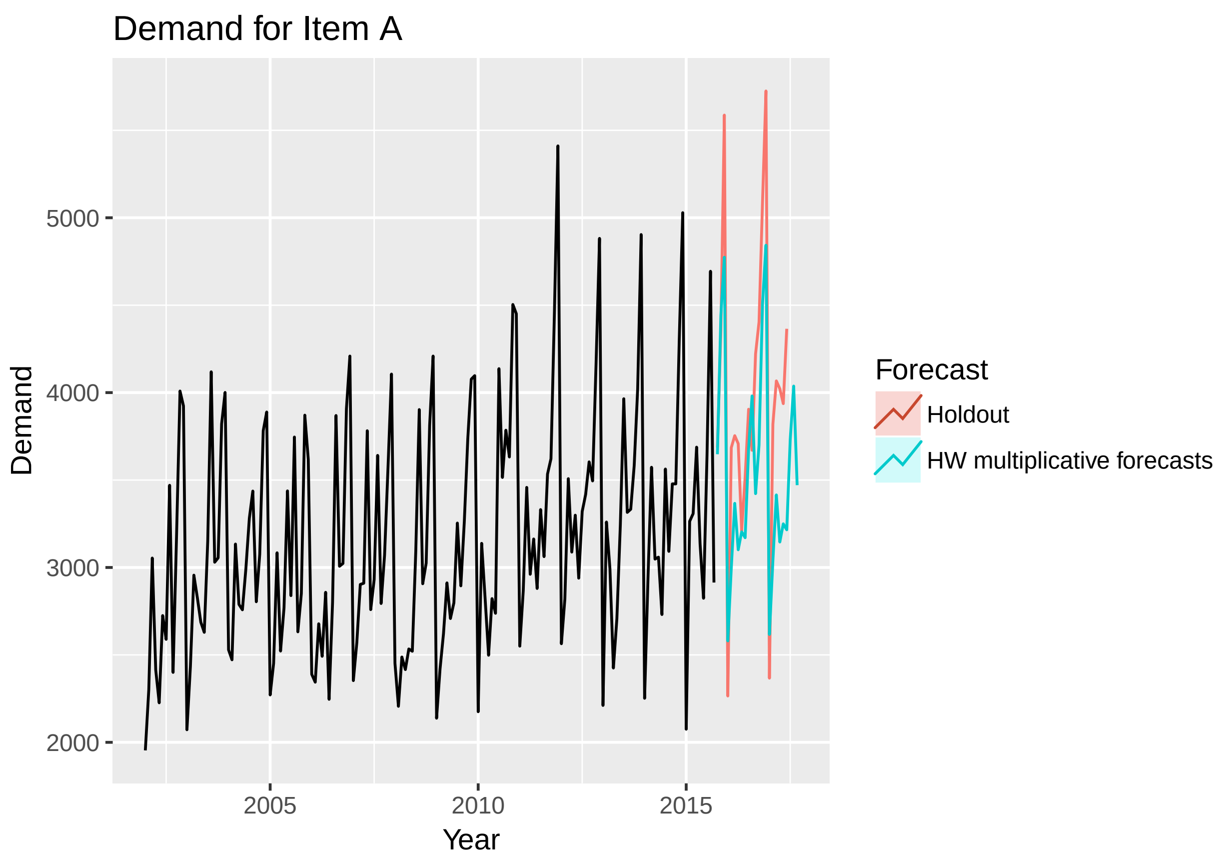
6. Fitting the models on Train set and testing

Train & Test split

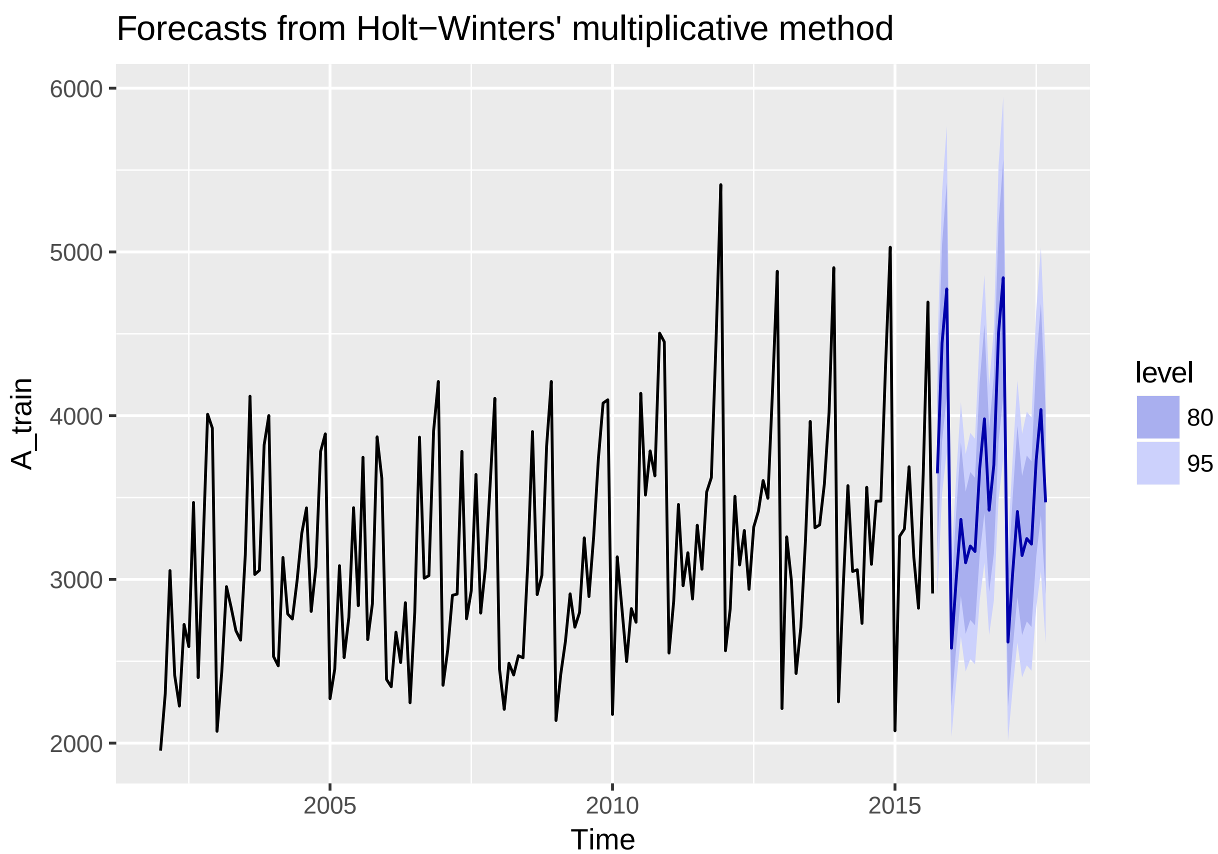
```
A_train <- window(itemA, start = c(2002,1), end = c(2015,9))
A_test <- window(itemA, start = c(2015,10), end = c(2017,6))

B_train <- window(itemB, start = c(2002,1), end = c(2015,9))
B_test <- window(itemB, start = c(2015,10), end = c(2017,6))
```

```
fit1 <- hw(A_train,seasonal="multiplicative")
autoplot(A_train) +
  autolayer(A_test, series="Holdout", PI=FALSE) +
  autolayer(fit1, series="HW multiplicative forecasts",
    PI=FALSE) +
  xlab("Year") +
  ylab("Demand") +
  ggtitle("Demand for Item A") +
  guides(colour=guide_legend(title="Forecast"))
```

```
Warning: Ignoring unknown parameters: PI
```

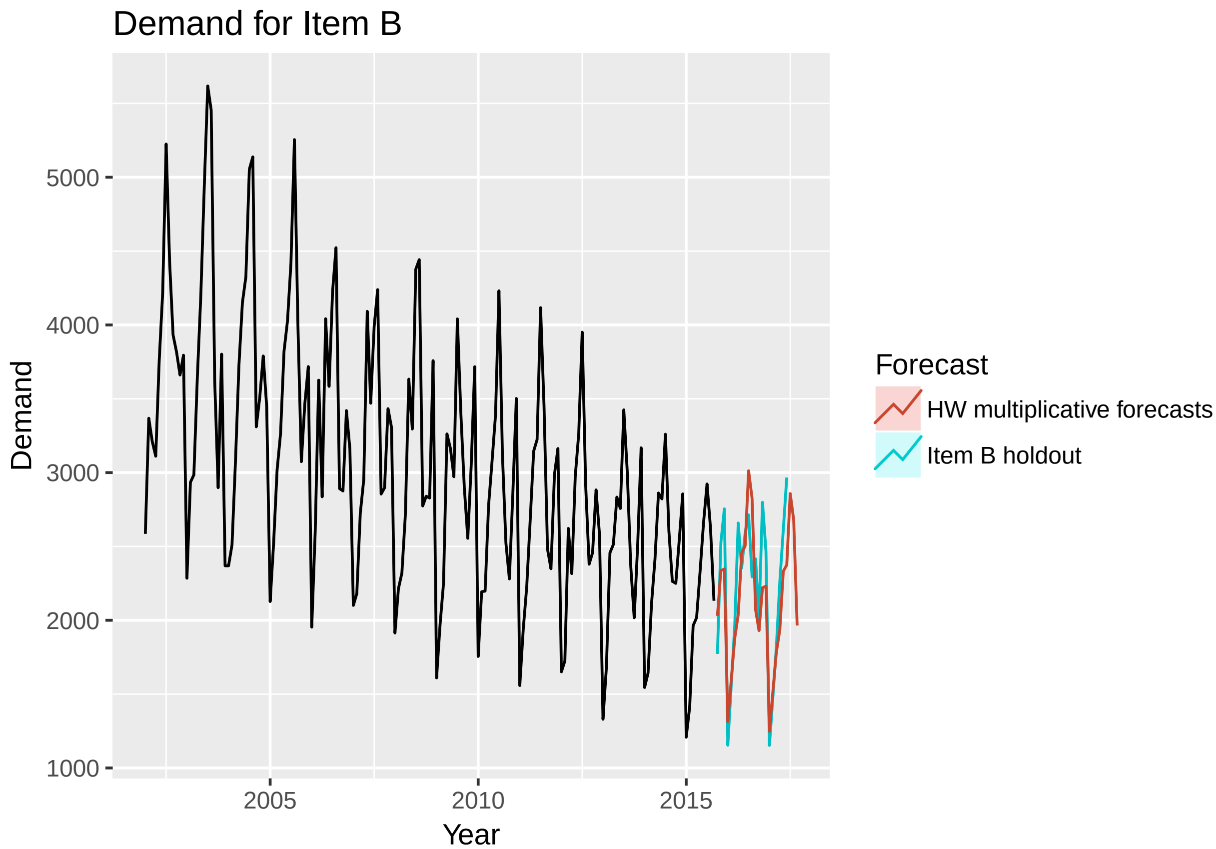


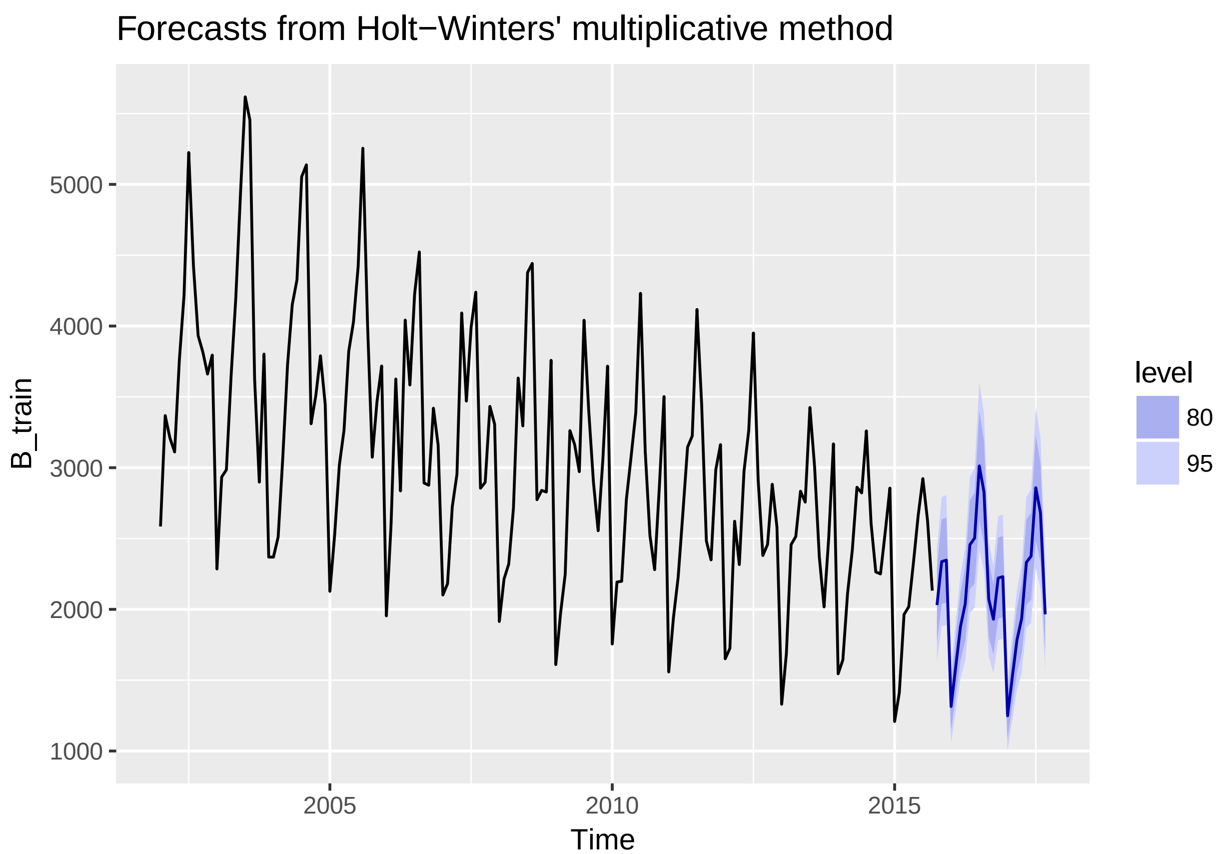
Demand for Item A
5000
4000
3000
2000
Demand
2005
2010
2015
Year
Forecast
Holdout
HW multiplicative forecasts




```
fit2 <- hw(B_train,seasonal="multiplicative")
autoplot(B_train) +
  autolayer(B_test, series="Item B holdout", PI=FALSE) +
  autolayer(fit2, series="HW multiplicative forecasts",
    PI=FALSE) +
  xlab("Year") +
  ylab("Demand") +
  ggtitle("Demand for Item B") +
  guides(colour=guide_legend(title="Forecast"))
```

```
Warning: Ignoring unknown parameters: PI
```


Demand for Item B
Demand
5000
4000
3000
2000
1000
2005
2010
2015
Year
Forecast
HW multiplicative forecasts
Item B holdout



Accuracy forecast of HW

Acutal Values for Item A series:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2015										3686	4358	5587
2016	2265	3685	3754	3708	3210	3517	3905	3670	4221	4404	5086	5725
2017	2367	3819	4067	4022	3937	4365						

Forecasted Values :

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Oct 2015	3647.681	3157.504	4137.858	2898.020	4397.342
Nov 2015	4443.077	3842.217	5043.936	3524.142	5362.012
Dec 2015	4774.100	4124.248	5423.952	3780.237	5767.963
Jan 2016	2580.139	2226.566	2933.712	2039.396	3120.883
Feb 2016	2996.561	2583.085	3410.036	2364.205	3628.917
Mar 2016	3366.788	2898.935	3834.641	2651.269	4082.308
Apr 2016	3101.274	2667.189	3535.359	2437.398	3765.149
May 2016	3203.868	2752.092	3655.643	2512.937	3894.799
Jun 2016	3170.221	2719.797	3620.645	2481.357	3859.085
Jul 2016	3677.791	3151.202	4204.380	2872.443	4483.139
Aug 2016	3980.811	3406.330	4555.291	3102.219	4859.402
Sep 2016	3422.215	2924.370	3920.060	2660.827	4183.603
Oct 2016	3700.154	3157.457	4242.851	2870.171	4530.137
Nov 2016	4506.915	3840.381	5173.449	3487.539	5526.290
Dec 2016	4842.612	4120.365	5564.859	3738.031	5947.194

```
Jan 2017       2617.122 2223.437 3010.807 2015.033 3219.211
Feb 2017       3039.461 2578.257 3500.666 2334.110 3744.813
Mar 2017       3414.932 2892.173 3937.691 2615.440 4214.423
Apr 2017       3145.568 2659.731 3631.405 2402.544 3888.591
May 2017       3249.572 2743.123 3756.022 2475.025 4024.120
Jun 2017       3215.392 2709.677 3721.108 2441.967 3988.817
```

```
Acutal Values for Item B series:
```

```
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2015                                                1772 2526 2755
2016 1154 1568 1965 2659 2354 2592 2714 2294 2416 2016 2799 2467
2017 1153 1482 1818 2262 2612 2967
```

```
Forecasted Values :
```

```
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Oct 2015       2030.005 1770.674 2289.335 1633.392 2426.617
Nov 2015       2336.665 2038.077 2635.253 1880.014 2793.316
Dec 2015       2347.347 2047.302 2647.393 1888.468 2806.227
Jan 2016       1314.015 1145.996 1482.034 1057.052 1570.978
Feb 2016       1598.651 1394.158 1803.143 1285.907 1911.395
Mar 2016       1878.410 1638.031 2118.789 1510.782 2246.038
Apr 2016       2037.955 1777.039 2298.872 1638.917 2436.993
May 2016       2456.173 2141.552 2770.794 1975.002 2937.344
Jun 2016       2502.898 2182.114 2823.683 2012.301 2993.496
Jul 2016       3012.727 2626.366 3399.088 2421.839 3603.615
Aug 2016       2826.050 2463.390 3188.709 2271.410 3380.690
Sep 2016       2071.993 1805.910 2338.077 1665.054 2478.933
Oct 2016       1929.898 1681.870 2177.926 1550.572 2309.224
Nov 2016       2220.961 1935.288 2506.633 1784.063 2657.858
Dec 2016       2230.632 1943.460 2517.805 1791.439 2669.826
Jan 2017       1248.408 1087.533 1409.283 1002.371 1494.445
Feb 2017       1518.498 1322.617 1714.380 1218.923 1818.073
Mar 2017       1783.836 1553.474 2014.198 1431.527 2136.144
Apr 2017       1934.916 1684.751 2185.082 1552.321 2317.512
May 2017       2331.463 2029.651 2633.275 1869.881 2793.045
Jun 2017       2375.276 2067.384 2683.169 1904.395 2846.158
```

Smoothing parameters for Holt-Winters models

- The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations - one for the level Lt, one for the trend Bt, and one for the seasonal component St, with corresponding smoothing parameters alpha, beta and gamma.
- The small value of gamma for the multiplicative model means that the seasonal component hardly changes over time.

```
Exponential model for Item A: model information:
```

```
Holt-Winters' multiplicative method

Call:
 hw(y = A_train, seasonal = "multiplicative")

  Smoothing parameters:
    alpha = 0.11
    beta  = 0.0025
```

```
    gamma = 1e-04

  Initial states:
    l = 2981.0809
    b = 15.5355
    s = 1.3577 1.2651 1.0399 0.9629 1.1214 1.0373
           0.8952 0.9057 0.8778 0.9541 0.8502 0.7329

  sigma:  0.1049

     AIC     AICc      BIC
2772.863 2777.026 2825.664
```

Observaton:

- The smoothing parameters for Item A series are
- alpha = 0.11, beta = 0.0025 and gamma = 1e-04

```
Exponential model for Item B: model information:
```

```
Holt-Winters' multiplicative method

Call:
 hw(y = B_train, seasonal = "multiplicative")

  Smoothing parameters:
    alpha = 0.0219
    beta  = 0.0013
    gamma = 1e-04

  Initial states:
    l = 4061.2037
    b = -12.7268
    s = 1.0492 1.0402 0.8999 0.962 1.3066 1.3869
           1.1473 1.1211 0.9263 0.8502 0.7206 0.5898

  sigma:  0.0997

     AIC     AICc      BIC
2732.999 2737.162 2785.800
```

Observaton:

- The smoothing parameters for Item B series are
- alpha = 0.0219, beta = 0.0013 and gamma = 1e-04

Accuracy for the Holt-Winters models

```
MAPE for Item A for the fitted model:
```

```
[1] 7.914132
```

```
MAPE for Item B for the fitted model:
```

```
[1] 7.32645
```

Accuracy for holdout data

```r
cat('Accuracy scores for the Item A: \n\n')
```

```
Accuracy scores for the Item A:
```

```r
accuracy(A_test, A_fore$mean)
```

```
                ME     RMSE      MAE       MPE     MAPE       ACF1
Test set -441.0355 614.6804 532.5735 -12.42268 15.42209 -0.1288378
         Theil's U
Test set  1.014193
```

```r
cat('Accuracy scores for the Item B: \n\n')
```

```
Accuracy scores for the Item B:
```

```r
accuracy(B_test, B_fore$mean)
```

```
                ME     RMSE      MAE      MPE     MAPE       ACF1
Test set -112.2963 319.3895 256.4416 -5.13389 11.77697 0.06421008
         Theil's U
Test set 0.7423873
```

7. Forecasting for July 2017 to December 2018

```r
fit1 <- hw(itemA,seasonal="multiplicative")
fit2 <- hw(itemB,seasonal="multiplicative")
```

```r
A_forecast <- forecast(fit1, h=17)
B_forecast <- forecast(fit2, h=17)
```

```
Forecasted value for Item A: July 2017 to December 2018:
```

```
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Aug 2017       4478.749 3856.675 5100.823 3527.370 5430.128
Sep 2017       4280.627 3682.090 4879.164 3365.244 5196.010
Oct 2017       4574.393 3930.282 5218.503 3589.310 5559.475
Nov 2017       5430.485 4660.198 6200.772 4252.433 6608.537
Dec 2017       6314.394 5411.842 7216.945 4934.060 7694.727
Jan 2018       2834.952 2426.492 3243.411 2210.267 3459.637
Feb 2018       4091.394 3497.006 4685.782 3182.356 5000.432
Mar 2018       4347.871 3710.783 4984.958 3373.529 5322.212
Apr 2018       4164.703 3549.023 4780.382 3223.102 5106.303
May 2018       4008.567 3410.531 4606.603 3093.950 4923.184
Jun 2018       4086.023 3470.675 4701.372 3144.929 5027.118
Jul 2018       4647.025 3940.403 5353.648 3566.339 5727.711
Aug 2018       4753.241 3985.056 5521.427 3578.403 5928.080
Sep 2018       4541.664 3800.913 5282.415 3408.783 5674.544
Oct 2018       4851.954 4053.136 5650.771 3630.267 6073.640
Nov 2018       5758.359 4801.182 6715.535 4294.483 7222.234
Dec 2018       6693.754 5570.160 7817.349 4975.365 8412.144
```

```
Forecasted value for Item B: July 2017 to December 2018:
```

```
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Aug 2017       2732.642 2373.071 3092.213 2182.725 3282.559
```

```
Sep 2017       2032.116 1764.692 2299.541 1623.126 2441.107
Oct 2017       1875.124 1628.325 2121.923 1497.677 2252.571
Nov 2017       2186.956 1899.066 2474.846 1746.666 2627.246
Dec 2017       2226.922 1933.712 2520.131 1778.496 2675.347
Jan 2018       1218.828 1058.313 1379.343  973.341 1464.315
Feb 2018       1480.410 1285.393 1675.426 1182.157 1778.662
Mar 2018       1747.191 1516.960 1977.422 1395.083 2099.299
Apr 2018       1960.994 1702.499 2219.488 1565.661 2356.327
May 2018       2296.872 1993.983 2599.760 1833.644 2760.099
Jun 2018       2366.972 2054.703 2679.242 1889.397 2844.548
Jul 2018       2818.731 2446.679 3190.783 2249.727 3387.735
Aug 2018       2627.383 2280.398 2974.369 2096.715 3158.051
Sep 2018       1953.589 1695.434 2211.745 1558.774 2348.404
Oct 2018       1802.429 1564.091 2040.767 1437.923 2166.936
Nov 2018       2101.897 1823.759 2380.035 1676.522 2527.273
Dec 2018       2140.027 1856.620 2423.434 1706.594 2573.461
```

8. Conclusion

- The forecast will be used to plan and set the goal which will align with the forecast and if possbile will try to identify additional insights which will substantiate the forecast. Based on the accuracy received on the holdout set, the forecast will moderately for Item A and to some extent closely for Item B shadow the actual observation in the future for the forecasted period.
- In addition to taking into account the past demand, lead time and planned advertising and other marketing activity will also be incorporated into forecast horizon to make decisions in real suituation which would have an impact on the stading of the business