**Sales Forecasting**

**1. Identify time series components and plot the data.**

**a. Create time series data set sales.ts in R using the ts() function.**

To create the time series data set, we loaded the sales data from the CSV file and created the data set *‘sales.ts’* using the ‘*ts()*’ function. We specified the start and end date of the time series, along with the frequency of the data.

Text

Description automatically generated

The resulting time series data set has the monthly sales data from January 2015 to December 2022.

**b. Employ the plot() function to create a data plot of the historical data, provide it in your report, and explain what data patterns can be visualized in this plot.**

To create a data plot of the historical sales data using *‘plot()’* function:

**A picture containing text

Description automatically generated**

This will produce a plot that shows the monthly sales from 2015 to 2022.

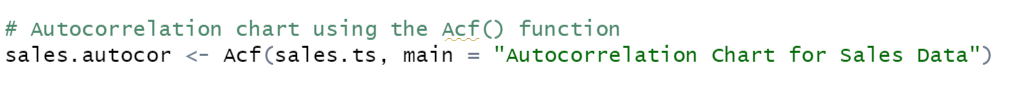
Chart, histogram

Description automatically generated

In the above plot, we can see that there is an upward linear trend in sales over time, with additive seasonality. Overall, this plot suggests that the data exhibits both trend and seasonal components, as well as some degree of randomness or volatility.

**c. Apply the Acf() function to identify possible time series components. Provide in the report the autocorrelation chart and explain time series components existing in the historical data.**

To create an autocorrelation chart using the Acf() function:



From the chart shown below, apart from level component, we can see that there is a significant positive autocorrelation at lag 1, indicating a strong linear relationship between each month's sales and the sales from the previous month. This suggests the presence of a trend component in the data.

There is also a significant positive autocorrelation at lag 12, which confirms the presence of a seasonal component in the data with a period of 12 months. This makes sense given that the data has a frequency of 12 (monthly data). The seasonal component is likely related to the regular peaks and troughs observed in the data plot.

There is some evidence of a more complex autocorrelation pattern, with significant negative correlations at some lags (such as lag 2 and 13) which suggests that there may be some monthly sales values that are negatively correlated with each other, although this is not as strong as the positive autocorrelations.

Overall, we can see that the sales data has both a trend and a seasonal component, as well as some random noise and possible correlation between certain time periods.

Chart, box and whisker chart

Description automatically generated

**2. Use trailing MA for forecasting time series.**

**a. Develop data partition with the validation partition of 24 monthly periods (2 years) and training partition of 72 monthly periods (6 years). Provide the data partition’s R code in your report.**

The R code to develop data partition with the validation partition of 24 monthly periods (2 years) and training partition of 72 monthly periods (6 years):

Text

Description automatically generated

It then uses the *‘window()’* function to split the time series into the two partitions.

**b. Use the rollmean() function to develop 3 trailing MAs with the window width of 3, 8, and 12 for the training partition. Present the R code for these MAs in your report.**

**A picture containing application

Description automatically generated**

The *align = "right"* argument specifies that the moving averages should be right-aligned, and the *fill = NA* argument specifies that any missing values resulting from the rolling window should be filled with NA. These moving averages are stored in the *ma3.trail*, *ma8.trail*, and *ma12.trail* variables.

Output:

Text

Description automatically generated

Text

Description automatically generated with low confidence

Text

Description automatically generated

**c. Use the forecast() function to create a trailing MA forecast for each window width from question 2b in the validation period, and present one of them, e.g., with window width of 3, in your report.**

To create a trailing MA forecast for window width of 3, 8 and 12:

Text

Description automatically generated

The forecasted values for the validation period for width of 3 are presented below:

Table

Description automatically generated

**d. Apply the accuracy() function to compare accuracy of the three trailing MA forecasts in the validation period. Present the accuracy measures in your report, compare MAPE and RMSE of these forecasts, and identify the best trailing MA forecast.**

To calculatethe accuracy measures for each of the three trailing MA forecasts using *‘accuracy()’* function:

A picture containing text

Description automatically generated

Output:

Calendar

Description automatically generated

From the output shown above, we can fetch the following values of MAPE and RMSE:

|  |  |  |
| --- | --- | --- |
|  | **MAPE** | **RMSE** |
| **ma3.trail.pred** | 4.181 | 12.462 |
| **ma8.trail.pred** | 1.184 | 3.987 |
| **ma12.trail.pred** | 0.674 | 2.818 |

Based on the accuracy measures, we can see that the MA forecast with a window width of 12 has low MAPE and RMSE as compared to window width of 3 and 8. Therefore, it is safe to conclude that the MA forecast with a window width of 12 is the best trailing MA forecast.

**3. Apply the two-level forecast with regression and trailing MA for residuals.**

**a. Develop using the tslm() function a regression model with linear trend and seasonality. Present the model summary in your report. Present and briefly explain the model equation in your report. Using this model, forecast monthly sales in the validation period with the forecast() function. Present the forecast in your report.**

The R code to develop a regression model with linear trend and seasonality using the tslm() function:

Text

Description automatically generated with low confidence

Below is the printed model summary:

Text

Description automatically generated

The model equation:

**y =** 181.8003 + 2.0419\*trend - 39.4086\*season2 - 69.2505\*season3 - 63.2591\*season4 - 70.0844\*season5 - 37.4930\*season6 - 18.8682\*season7 - 7.6435\*season8 + 47.3813\*season9 + 65.6560\*season10 + 107.9474\*season11 + 161.4722\*season12

To forecast monthly sales in the validation period with the forecast() function:

Text

Description automatically generated with medium confidence

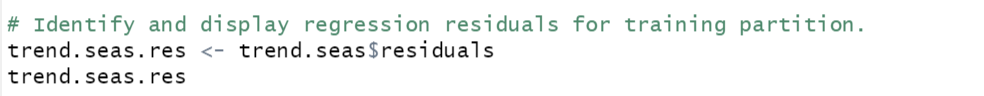
The forecasted values of the monthly sales for the validation period are presented below:

Text, table

Description automatically generated with medium confidence

**b. Identify regression residuals in the training period, apply a trailing MA (window width of 3) for these residuals using the rollmean() function, and identify trailing MA forecast of these residuals in the validation period (use the forecast() function). Provide the trailing MA forecast for residuals in the validation period in your report.**

To identify and display the regression residuals in the training period:

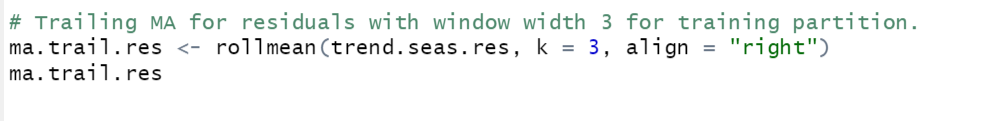


Outcome:

Text

Description automatically generated with low confidence

To apply the trailing MA for residuals with window width k = 3 for these residuals using the rollmean() function:



Output:

Text

Description automatically generated with low confidence

To identify trailing MA forecast of these residuals in the validation period using the ‘*forecast()’* function:

Text

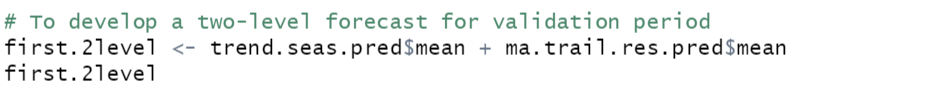
Description automatically generated with medium confidence

This will produce the forecast values for the trailing MA of residuals for the validation period.:

Table

Description automatically generated

**c. Develop two-level forecast for the validation period by combining the regression forecast and trailing MA forecast for residuals. Present in your report a table that contains validation data, regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the validation period. Apply the accuracy() function to compare accuracy of the regression model with linear trend and seasonality and the two-level (combined) model with the regression and trailing MA for residuals. Present the accuracy measures in your report, compare MAPE and RMSE of these forecasts, and identify the best forecasting model for the validation period.**

To develop a two-level forecast for the validation period, we need to combine the forecasts obtained from the regression model with the trailing MA forecast of residuals. We can do this by adding the two forecasts together. 

Output:

A picture containing text

Description automatically generated

To create a table that contains validation data, regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the validation period:

Text

Description automatically generated

Output:

Table

Description automatically generated

To compare accuracy of the regression model with linear trend and seasonality and the two-level (combined) model with the regression and trailing MA for residuals:Text

Description automatically generated

Output:

Text

Description automatically generated

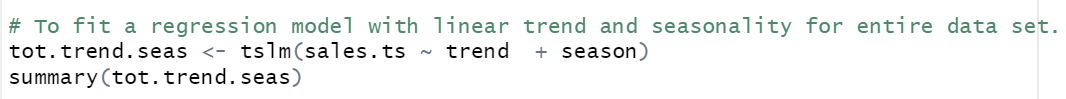
From the output shown above, we can fetch the following values of MAPE and RMSE:

|  |  |  |
| --- | --- | --- |
|  | **MAPE** | **RMSE** |
| **Regression** | 10.55 | 37.212 |
| **Two-Level** | 8.533 | 31.625 |

Based on the accuracy measures, we can see that the two-level model has low MAPE and RMSE as compared to regression model. Therefore, it is safe to conclude that the two-level model is the best forecasting model for the validation period.

**d. For the entire data set, identify the regression model with linear trend and seasonality and trailing MA with the window width of 3 for the regression residuals. Use these models to forecast the 12 months of 2023 and develop a two-level forecast for the 12 future months as a combination of the specified forecasts. Present in your report a table that contains the regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the 12 months of 2023.**

To identify the regression model with linear trend and seasonality



Output:

Table

Description automatically generated

To identify the regression residuals and the trailing MA with the window width of 3 for theses regression residuals:

Text

Description automatically generated with medium confidence

Output:

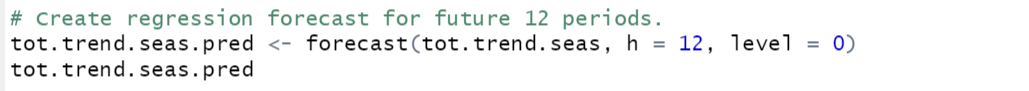
Text

Description automatically generated

Text

Description automatically generated

Next, we can use these models to forecast the 12 months of 2023:



Output:

Text

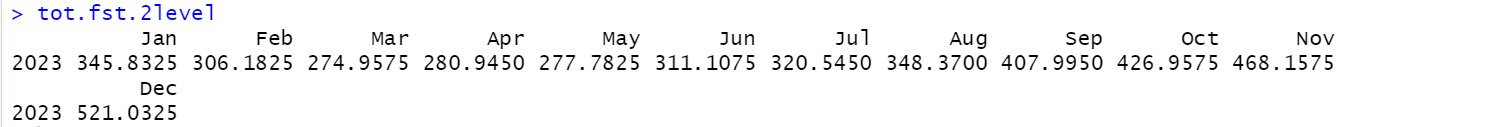
Description automatically generated

To develop 2-level forecast for future 12 periods by combining regression forecast and trailing MA for residuals for future 12 periods:

Graphical user interface, text

Description automatically generated

Output:

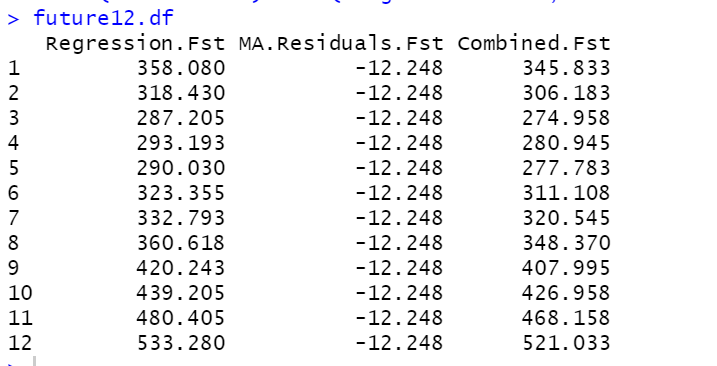


Finally, to prepare a table that contains the regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the 12 months of 2023:

Text

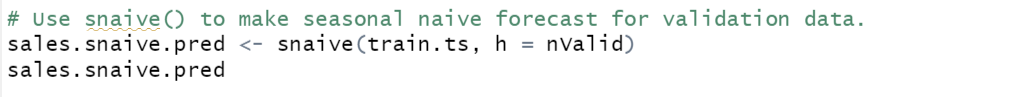
Description automatically generated

Output:

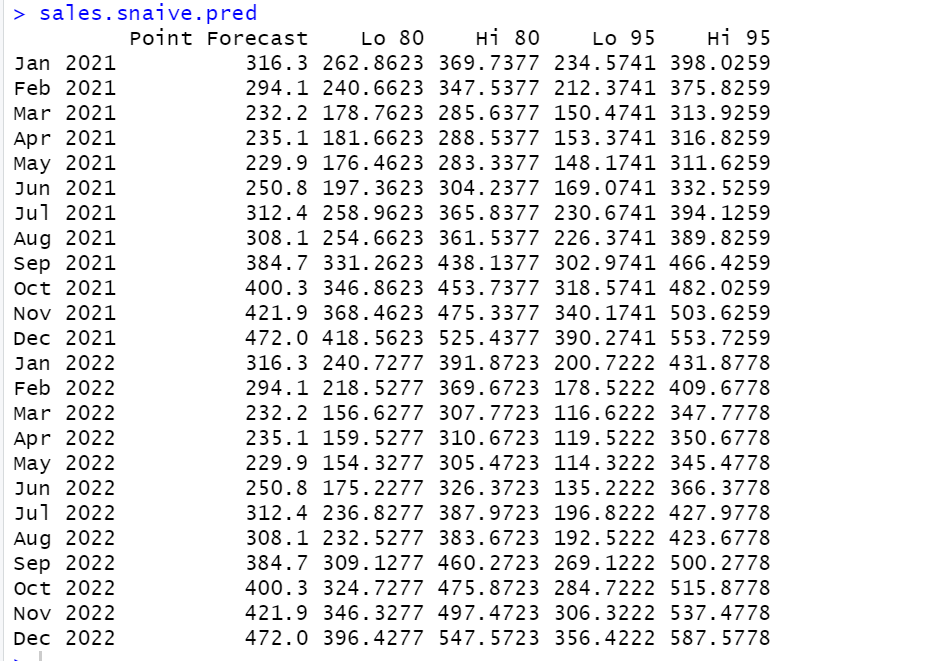


**e. Develop a seasonal naïve forecast for the entire historical data set and apply the accuracy() function to compare accuracy of the three forecasting models: seasonal naïve forecast, regression model with linear trend and seasonality, and two-level (combined) model with the regression and trailing MA for residuals. Present the accuracy measures in your report, compare MAPE and RMSE of these forecasts, and identify the best forecasting model for forecasting monthly sales in 2023.**

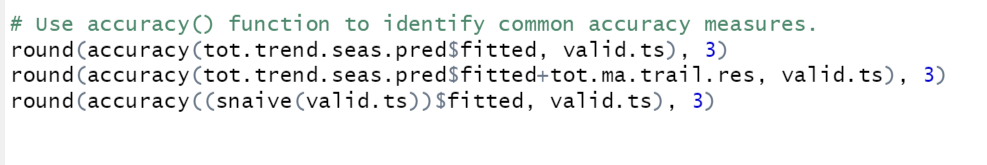
To develop a seasonal naïve forecast for the entire historical data set, we can use snaive() function in R:



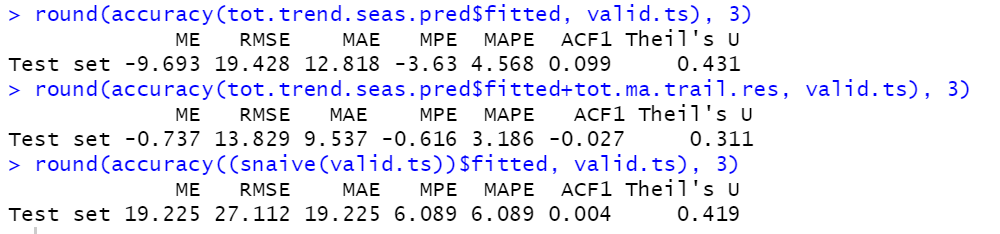
Output:



To compare accuracy of the three forecasting models: seasonal naïve forecast, regression model with linear trend and seasonality, and two-level (combined) model with the regression ad trailing MA for residuals:



Output:



From the output shown above, we can fetch the following values of MAPE and RMSE:

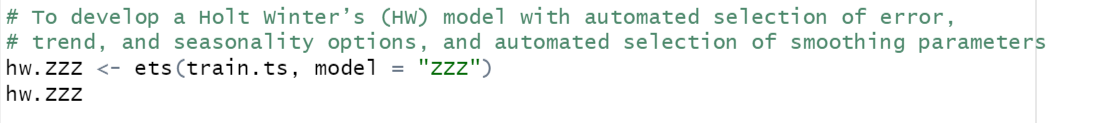
|  |  |  |
| --- | --- | --- |
|  | **MAPE** | **RMSE** |
| **Snaive** | 6.089 | 27.112 |
| **Regression** | 4.568 | 19.428 |
| **Two-Level** | 3.186 | 13.829 |

Based on the accuracy measures, we can see that the two-level model has low MAPE and RMSE as compared to regression model and Snaive model. Therefore, it is safe to conclude that the two-level model is the best forecasting model for forecasting monthly sales in 2023.

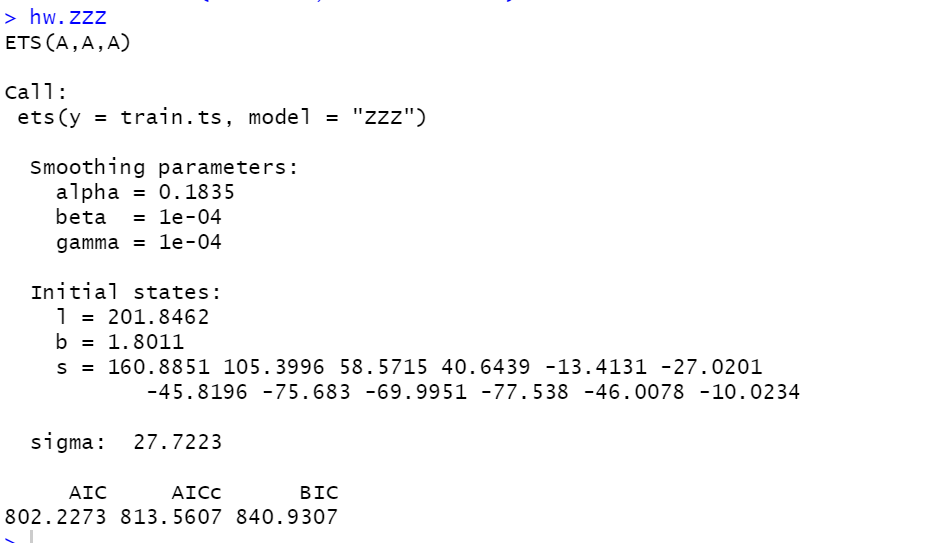
**4. Use advanced exponential smoothing methods.**

**a. For the training partition (from question 2a), use the ets() function to develop a Holt Winter’s (HW) model with automated selection of error, trend, and seasonality options, and automated selection of smoothing parameters for the training partition. Present the model summary (output) and explain the model in your report. Use the model to forecast monthly sales for the validation period using the forecast() function, and present this forecast in your report.**

To develop a Holt Winter’s (HW) model, using ets(), with automated selection of error, trend, and seasonality options, and automated selection of smoothing parameters for the training partition:

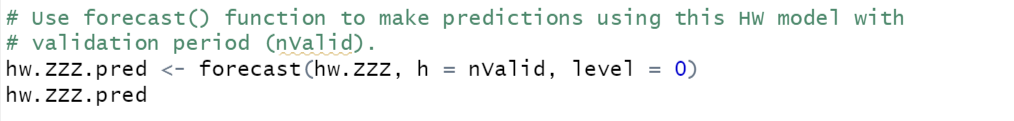


Output:

****

The model selected is ETS(A,A,A), which means that the model includes additive error, additive trend, and additive seasonality. The smoothing parameters selected are alpha = 0.1835, beta = 1e-04, and gamma = 1e-04. The initial states for level, trend, and seasonality components are also provided.

To forecast monthly sales for the validation period using the forecast() function:



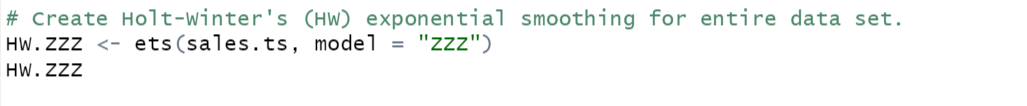
Output:

Table

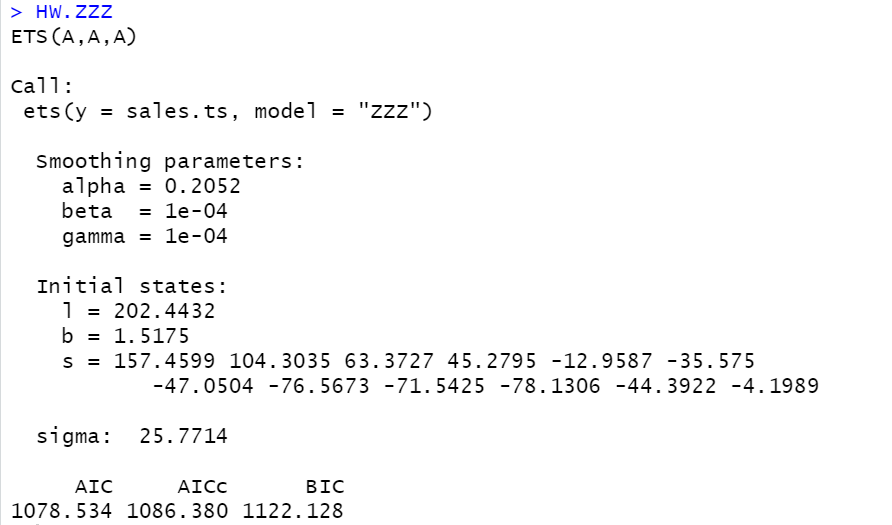
Description automatically generated

**b. To make a forecast in the 12 months of 2023, use the entire data set (no partitioning) to develop the HW model using the ets() function for the model with the automated selection of error, trend, and seasonality options, and automated selection of smoothing parameters. Present the model summary (output) and explain this model in your report. Use the model to forecast monthly sales in the 12 months of 2023 using the forecast() function, and present the forecast in your report.**

To make a forecast in the 12 months of 2023, using the entire data set (no partitioning) to develop the HW model, using the ets() function, for the model with the automated selection of error, trend, and seasonality options, and automated selection of smoothing parameters:

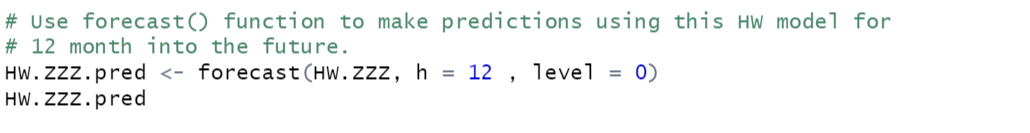


Output:



The model selected is ETS(A,A,A), which means that the model includes additive error, additive trend, and additive seasonality. The smoothing parameters selected are alpha = 0.2052, beta = 1e-04, and gamma = 1e-04. The initial states for level, trend, and seasonality components are also provided.

To forecast monthly sales in the 12 months of 2023 using the forecast() function:



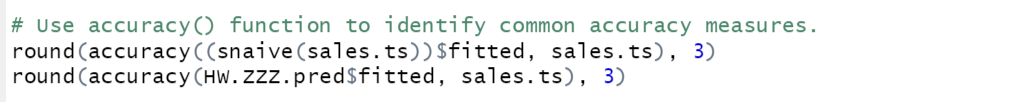
Output:

Text

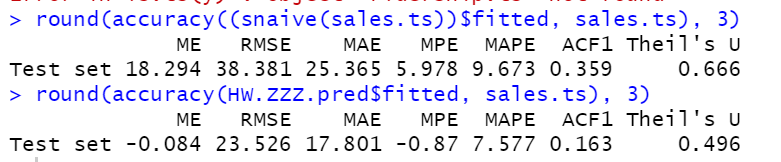
Description automatically generated

**c. Apply the accuracy() function to compare the two models: seasonal naïve forecast (applied in question 3e) and the HW model developed in question 4b. Present the accuracy measures in your report, compare MAPE and RMSE of these forecasts, and identify the best forecasting model.**

To compare the two models: seasonal naïve forecast and the HW model, using accuracy() function:



Output:



From the output shown above, we can fetch the following values of MAPE and RMSE:

|  |  |  |
| --- | --- | --- |
|  | **MAPE** | **RMSE** |
| **Snaive** | 9.673 | 38.381 |
| **HW model** | 7.577 | 23.526 |

Based on the accuracy measures, we can see that the HW model has low MAPE and RMSE as compared to seasonal naïve forecast model. Therefore, it is safe to conclude that the HW model is the best forecasting model.

**d. Compare the best forecasts identified in questions 3e and 4c. Explain what your final choice of the forecasting model in this case will be.**

The best forecasts identified in 3e:

|  |  |  |
| --- | --- | --- |
|  | **MAPE** | **RMSE** |
| **Snaive** | 6.089 | 27.112 |
| **Regression** | 4.568 | 19.428 |
| **Two-Level** | 3.186 | 13.829 |

Best Model: Two-Level Model

The best forecasts identified in 4c:

|  |  |  |
| --- | --- | --- |
|  | **MAPE** | **RMSE** |
| **Snaive** | 9.673 | 38.381 |
| **HW model** | 7.577 | 23.526 |

Best Model: HW Model

After comparing the accuracy measures of the best forecasts identified in questions 3e and 4c, we can see that the Two-Level Model developed in question 3e has a lower MAPE and RMSE compared to the HW Model developed in question 4c. Therefore, we can conclude that the Two-Level Model is a better forecasting model for the monthly sales data of the given company.