CAL STATE EAST BAY

Case Study 1

→ Report

Sales Forecasting Analysis for Grocery Store Chain

Pallavi K Nair
pnair5@horrizon.csueastbay.edu
fb4097

Feb 12, 2024

Table Of → Contents

Question 1:

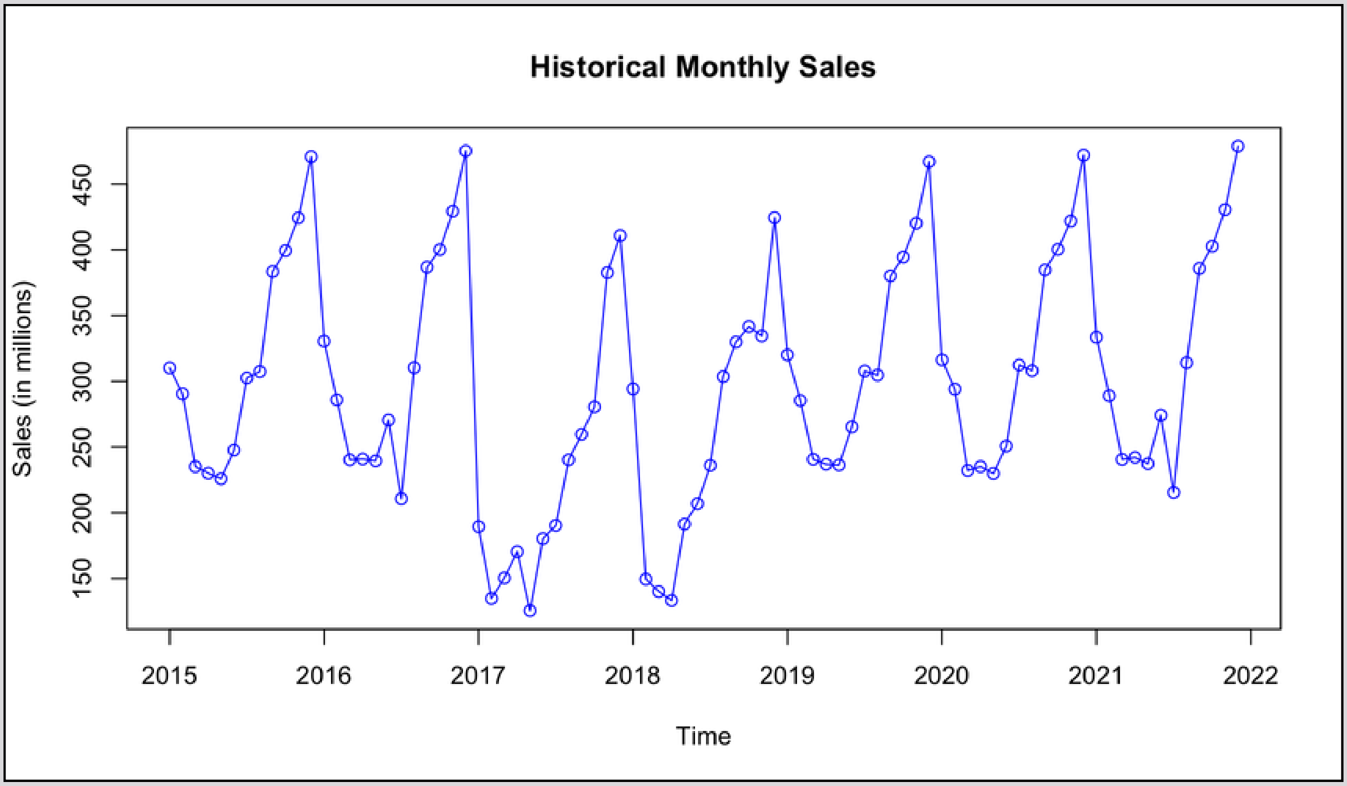
→ About the Dataset & Plots

Data Overview: The dataset consists of monthly sales data for a large chain of grocery stores from January 2015 to December 2021. Sales are reported in millions of dollars.

a. Creating Time Series Dataset

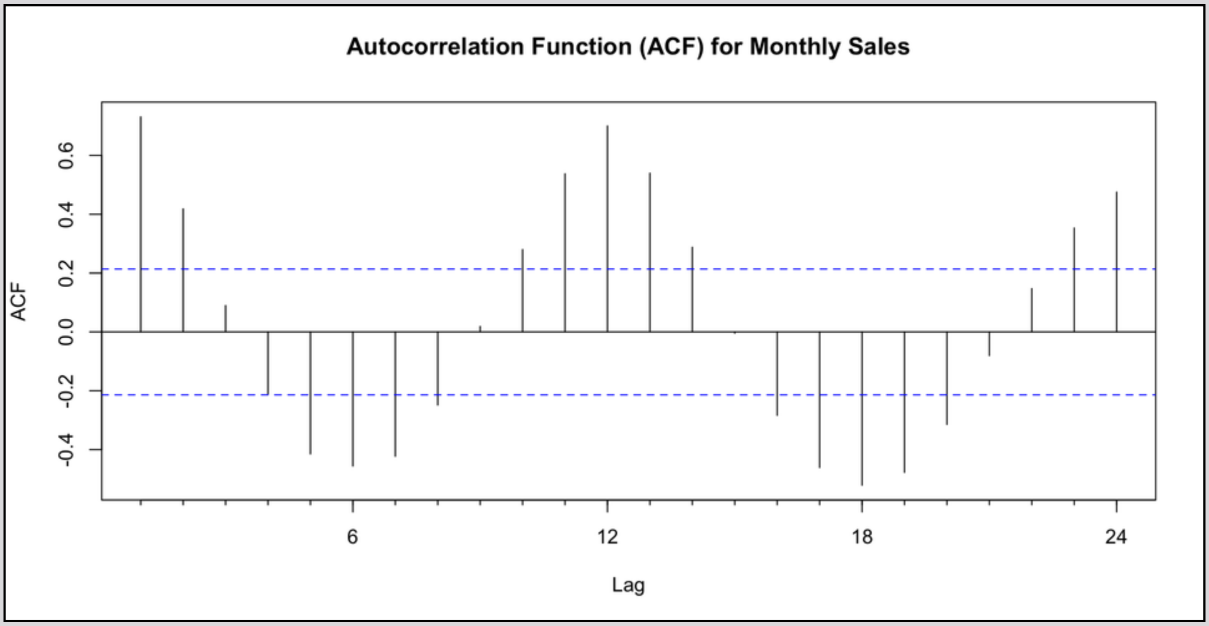
- In R, this would be done using the ts() function.

b. Data Plotting:



- **Trend**: There appears to be an overall increasing trend in sales over the years, indicating growth in the grocery chain's monthly sales.
- **Seasonality**: There are patterns that suggest seasonality within each year, with certain months consistently showing higher or lower sales than others.
- **Volatility**: The sales data show some degree of volatility, with occasional spikes and drops, which may be attributed to specific events or periods of high demand.

c. Autocorrelation Analysis:
An autocorrelation plot (ACF) will be generated to identify time series components.



- **Significant Lagged Correlation:** There are several spikes outside the confidence bands at specific lags, indicating significant autocorrelations at those points. This suggests seasonality in the data, as the autocorrelation pattern repeats annually.
- **Decaying Correlation:** The gradual decay in autocorrelation as the lag increases suggests a trend component, where sales in closer months are more correlated than sales further apart.

Question 2:

a. Partitioning the Data

- The dataset was partitioned into a training set covering five years (2015-2019) and a validation set for the last two years (2020-2021), enabling the development and validation of forecasting models.

```
##2 (a)# Partitioning the data
# Define the numbers of months in the training and validation sets,
# nTrain and nValid, respectively.
training  <- window(sales.ts, start=c(2015,1), end=c(2019,12))
validation <- window(sales.ts, start=c(2020,1))
nValid <- length(validation) # Number of periods to forecast
nTrain <- length(sales.ts) - nValid
```

→ *b. Trailing Moving Averages*

```
install.packages("zoo")  # Install the zoo package if you haven't already
library(zoo)             # Load the zoo package to use the rollmean() function

# 2 (b) Calculate moving averages
ma2 <- rollmean(training, 2, align='right')
ma6 <- rollmean(training, 6, align='right')
ma12 <- rollmean(training, 12, align='right')
ma2
ma6
ma12
```

```
> ma2
        Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct    Nov    Dec
2015        300.35 262.80 232.60 228.00 236.80 275.15 305.00 345.45 391.45 411.85 447.55
2016 400.65 308.20 263.10 240.60 240.20 255.00 240.60 260.55 348.65 393.50 414.75 452.30
2017 332.35 162.15 142.60 160.40 148.00 153.00 185.40 215.35 250.00 270.10 331.60 396.75
2018 352.55 221.90 144.80 136.70 162.35 199.20 221.50 269.80 316.75 335.75 338.05 379.50
2019 372.30 302.75 263.05 238.80 236.60 250.85 286.60 306.25 342.45 387.35 407.35 443.60
> ma6
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov
2015                                           256.5833 255.3167 258.1333 282.8667 311.0833 344.1500
2016 385.9833 382.4000 358.5333 332.1167 301.3167 267.9333 247.9667 252.0500 276.4833 303.0167 334.6667
2017 365.2500 335.9833 296.5667 258.2833 207.6500 158.5167 158.6667 176.2500 194.4667 212.8167 255.6667
2018 311.3833 296.2500 276.3167 251.7833 219.9000 185.9333 176.2167 201.9000 233.5333 268.2500 292.1000
2019 342.3667 339.3333 324.4667 307.0167 290.6500 264.1333 262.0833 265.3000 288.5500 314.8167 345.4667
          Dec
2015 381.3333
2016 368.7833
2017 294.0667
2018 328.3500
2019 379.0667
> ma12
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov
2015
2016 320.6500 320.2667 320.7000 321.6000 322.7333 324.6333 316.9750 317.2250 317.5083 317.5667 317.9917
2017 306.6083 294.0167 286.5250 280.6500 271.1583 263.6500 261.9583 256.1167 245.5167 235.5500 231.6583
2018 235.0250 236.2500 235.3917 232.3000 237.7833 240.0000 243.8000 249.0750 254.9250 260.0167 256.0000
2019 259.2917 270.6167 279.0000 287.6333 291.3750 296.2417 302.2250 302.3167 306.5083 310.9167 318.0583
          Dec
2015 318.9583
2016 318.3583
2017 226.2917
2018 257.1417
2019 321.6000
```

Moving Average Outputs:

The calculated values for each month across the years in the training set are as follows (selected examples to illustrate the trend):

- MA2 Example: For January 2019, the 2-month MA was 372.30, showing a smoothing over the immediate past two months.
- MA6 Example: For December 2019, the 6-month MA was 379.07, averaging sales over the latter half of the year.
- MA12 Example: For December 2019, the 12-month MA was 321.60, indicating the average sales over the entire year.

c. Forcasted trairing MA

```
> forecast_ma2
         Point Forecast        Lo 80    Hi 80     Lo 95    Hi 95
Jan 2020       388.9827 358.30328 419.6621 342.06258 435.9028
Feb 2020       271.9981 228.61302 315.3831 205.64637 338.3497
Mar 2020       226.5765 173.44177 279.7112 145.31395 307.8390
Apr 2020       219.7166 158.36246 281.0708 125.88354 313.5497
May 2020       221.8603 153.26458 290.4559 116.95223 326.7683
Jun 2020       239.2137 164.07115 314.3562 124.29309 354.1343
Jul 2020       255.7416 174.57857 336.9046 131.61346 379.8697
Aug 2020       288.2169 201.45013 374.9837 155.51858 420.9152
Sep 2020       339.1769 247.14696 431.2068 198.42926 479.9245
Oct 2020       373.2919 276.28400 470.2999 224.93110 521.6528
Nov 2020       397.9211 296.17838 499.6638 242.31905 553.5231
Dec 2020       443.6000 337.33244 549.8676 281.07778 606.1223
Jan 2021       388.9827 278.37603 499.5894 219.82441 558.1410
Feb 2021       271.9981 157.21623 386.7799  96.45440 447.5417
Mar 2021       226.5765 107.76610 345.3868  44.87169 408.2812
Apr 2021       219.7166  97.00989 342.4233  32.05288 407.3803
May 2021       221.8603  95.37716 348.3433  28.42106 415.2995
Jun 2021       239.2137 109.06374 369.3636  40.16653 438.2608
Jul 2021       255.7416 122.02531 389.4579  51.24019 460.2430
Aug 2021       288.2169 151.02696 425.4068  78.40300 498.0308
Sep 2021       339.1769 198.59909 479.7547 124.18171 554.1720
Oct 2021       373.2919 229.40605 517.1778 153.23747 593.3464
Nov 2021       397.9211 250.80145 545.0407 172.92103 622.9211
Dec 2021       443.6000 293.31560 593.8844 213.75984 673.4402
```

Forecasting Results for 2-Month Window Width

Focusing on the 2-month window width as an example, the forecasted values along with their lower and upper confidence intervals at 80% and 95% levels are as follows:

- January 2020 Forecast: 388.98 with a high 95% confidence interval of 435.90 and a low of 342.06.
- December 2021 Forecast: 443.60 with a high 95% confidence interval of 673.44 and a low of 213.76.

The point forecasts and confidence intervals provide a range of expected sales, reflecting the uncertainty inherent in forecasting. The confidence intervals widen as the forecast horizon extends, indicating increasing uncertainty in the forecasted values over time.

d. Accuracy Comparison and Best Forecast

- The 2-month MA forecast demonstrates the lowest RMSE and MAPE values among the three models, indicating it has the highest accuracy in forecasting sales for the validation period. It also has the lowest Theil's U statistic, suggesting better forecast accuracy relative to a naïve benchmark.
- In contrast, the 12-month MA forecast shows the highest RMSE and MAPE values, indicating lower accuracy in capturing the validation period's sales variability.
- The 6-month MA forecast occupies an intermediate position in terms of RMSE and MAPE but is significantly less accurate than the 2-month MA forecast.

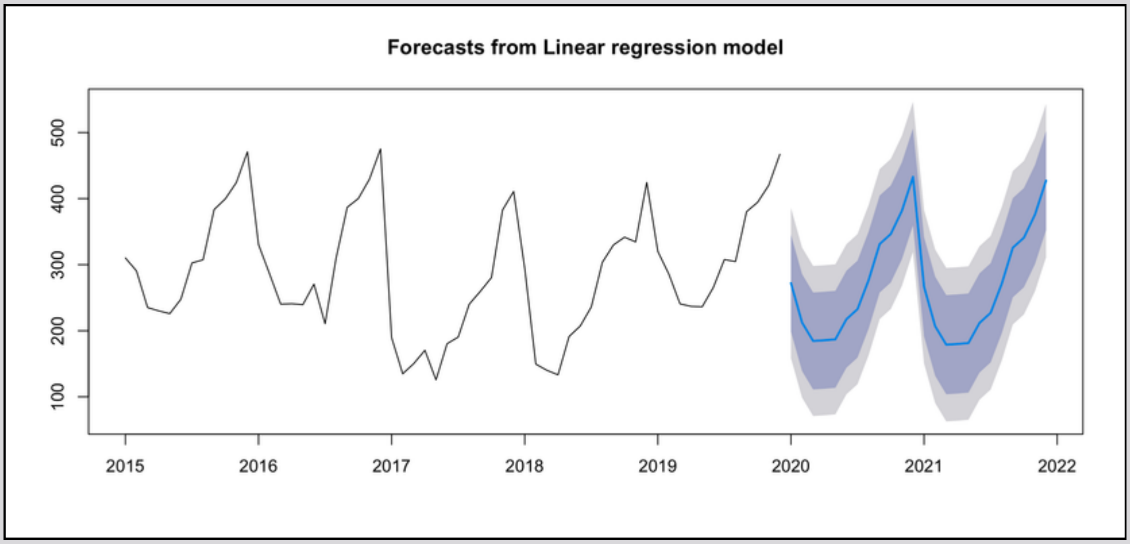
```
> acc_ma2
                    ME     RMSE      MAE        MPE     MAPE      MASE        ACF1 Theil's U
Training set -0.195556 20.90698 16.03396 -0.3999217 6.394312 0.3165041  0.28354974        NA
Test set     15.396155 33.66436 29.43007  4.3330554 9.190482 0.5809380 -0.05910627 0.5852469
> acc_ma6
                    ME      RMSE       MAE         MPE     MAPE      MASE      ACF1 Theil's U
Training set 0.1032791  7.880149  6.136341 -0.02769707  2.39370 0.1313992 0.4750483        NA
Test set     3.9276497 78.459362 73.426035 -3.69894323 23.51445 1.5722920 0.7557434  1.479102
> acc_ma12
                     ME      RMSE       MAE        MPE      MAPE       MASE      ACF1 Theil's U
Training set   0.1901474  4.389761  2.895964   0.112793  1.072254 0.06357928 0.1273538        NA
Test set     -57.8735238 94.511323 78.830192 -24.544847 29.276089 1.73067268 0.7138867  2.143902
```

Question 3:

a. Regression model with linear trend and seasonality

Model Summary

This model assumes that sales can be explained by a linear trend over time (trend) and by seasonal effects (season), capturing the systematic changes in sales across different months.



Key features of the plot:

- **Black Line:** Represents the actual historical sales data up to the point where the forecast begins. This line demonstrates the variability and the underlying patterns in the historical sales data.
- **Blue Line:** Indicates the point forecasts from the linear regression model for the validation period. This line provides the expected value of sales for each month, as predicted by the model.
- **Shaded Areas:** The lighter and darker shaded regions around the forecasts represent the 80% and 95% prediction intervals, respectively. These intervals reflect the range within which the actual sales figures are expected to fall, with a given level of confidence.
- **Forecast Horizon:** The forecast horizon covers the validation period and extends into 2022, allowing us to evaluate the model's performance against unseen data and to anticipate future sales trends.

```
> accuracy_measures
                 ME   RMSE    MAE    MPE   MAPE  MASE   ACF1 Theil's U
Training set  0.000 44.253 37.541 -3.825 16.274 0.686  0.702        NA
Test set     52.028 55.376 52.996 16.884 17.334 0.969 -0.055       1.1
```

Accuracy Measures:

The RMSE and MAPE values indicate the average magnitude of the forecast errors, with lower values representing more accurate forecasts. The Theil's U statistic provides a comparison to a naïve benchmark model, with values closer to 0 indicating better forecast accuracy.

→ b. Residuals & Trailing Moving Average for Regression Residuals

```
> sales_train_res
           Jan        Feb        Mar        Apr        May        Jun        Jul        Aug        Sep        Oct        Nov        Dec
2015  10.093333  50.093333  22.593333  16.593333  10.973333   2.313333  41.913333   2.933333  24.273333  24.993333  14.893333   9.953333
2016  35.986667  51.086667  33.386667  32.986667  30.166667  30.706667 -44.393333  11.526667  33.266667  31.286667  25.586667  19.946667
2017 -99.420000 -94.420000 -50.920000 -31.920000 -78.140000 -53.800000 -59.100000 -52.980000 -88.340000 -82.720000 -15.520000 -38.860000
2018  10.973333 -74.126667 -55.626667 -63.426667  -6.746667 -21.606667  -7.906667  15.913333 -12.546667 -16.026667 -58.126667 -19.566667
2019  42.366667  67.366667  50.566667  45.766667  43.746667  42.386667  69.486667  22.606667  43.346667  42.466667  33.166667  28.526667
> # Trailing MA on residuals
> training_ma_res <- rollmean(sales_train_res, 2, align='right')
> #Print
> training_ma_res
           Jan        Feb        Mar        Apr        May        Jun        Jul        Aug        Sep        Oct        Nov        Dec
2015             30.093333  36.343333  19.593333  13.783333   6.643333  22.113333  22.423333  13.603333  24.633333  19.943333  12.423333
2016  22.970000  43.536667  42.236667  33.186667  31.576667  30.436667  -6.843333 -16.433333  22.396667  32.276667  28.436667  22.766667
2017 -39.736667 -96.920000 -72.670000 -41.420000 -55.030000 -65.970000 -56.450000 -56.040000 -70.660000 -85.530000 -49.120000 -27.190000
2018 -13.943333 -31.576667 -64.876667 -59.526667 -35.086667 -14.176667 -14.756667   4.003333   1.683333 -14.286667 -37.076667 -38.846667
2019  11.400000  54.866667  58.966667  48.166667  44.756667  43.066667  55.936667  46.046667  32.976667  42.906667  37.816667  30.846667
> # Regression residuals in validation period.
> sales_valid_res <- validation - training_forecast_tslm$mean
> #Print
> sales_valid_res
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2020 44.16000 81.66000 47.66000 49.56000 42.94000 33.38000 79.68000 31.60000 53.44000 53.86000 40.46000 39.12000
2021 66.95333 82.15333 61.65333 62.15333 55.93333 62.37333 -11.62667 43.19333 60.23333 61.85333 54.75333 51.51333
```

- Point Forecast: The forecasted residuals for each month are consistently around 30.85, suggesting a systematic underestimation or overestimation by the regression model that the trailing MA aims to correct.
- Confidence Intervals: The intervals widen over time, starting with a range from approximately 3.98 to 57.72 in January 2020 and reaching a range from -100.77 to 162.47 in December 2021 at the 80% confidence level. This widening reflects increasing uncertainty in the residuals' forecasts as we move further away from the training data.

```
> sales_valid_res_forecast
         Point Forecast       Lo 80     Hi 80      Lo 95     Hi 95
Jan 2020       30.84736    3.977511  57.71722  -10.24653  71.94126
Feb 2020       30.84736   -7.150447  68.84517  -27.26527  88.96000
Mar 2020       30.84736  -15.689484  77.38421  -40.32460 102.01933
Apr 2020       30.84736  -22.888311  84.58304  -51.33426 113.02899
May 2020       30.84736  -29.230647  90.92537  -61.03402 122.72875
Jun 2020       30.84736  -34.964580  96.65931  -69.80332 131.49805
Jul 2020       30.84736  -40.237491 101.93222  -77.86754 139.56227
Aug 2020       30.84736  -45.145407 106.84013  -85.37355 147.06828
Sep 2020       30.84736  -49.755029 111.44976  -92.42336 154.11809
Oct 2020       30.84736  -54.114924 115.80965  -99.09124 160.78597
Nov 2020       30.84736  -58.261754 119.95648 -105.43327 167.12800
Dec 2020       30.84736  -62.224004 123.91873 -111.49301 173.18774
Jan 2021       30.84736  -66.024325 127.71905 -117.30510 178.99983
Feb 2021       30.84736  -69.681083 131.37581 -122.89763 184.59236
Mar 2021       30.84736  -73.209415 134.90414 -128.29375 189.98847
Apr 2021       30.84736  -76.621971 138.31670 -133.51280 195.20753
May 2021       30.84736  -79.929450 141.62418 -138.57115 200.26588
Jun 2021       30.84736  -83.141000 144.83573 -143.48279 205.17752
Jul 2021       30.84736  -86.264513 147.95924 -148.25980 209.95452
Aug 2021       30.84736  -89.306855 151.00158 -152.91266 214.60738
Sep 2021       30.84736  -92.274043 153.96877 -157.45058 219.14531
Oct 2021       30.84736  -95.171386 156.86611 -161.88168 223.57641
Nov 2021       30.84736  -98.003597 159.69832 -166.21317 227.90790
Dec 2021       30.84736 -100.774878 162.46961 -170.45148 232.14621
>
```

→ c. Two-Level Forecast: Combining Regression and Trailing MA for Residuals

```
> validation_df
   Actual_Sales Regression_Forecast Trailing_MA_Residuals Combined_Forecast
1         316.3             272.140                30.093           302.987
2         294.1             212.440                36.343           243.287
3         232.2             184.540                19.593           215.387
4         235.1             185.540                13.783           216.387
5         229.9             186.960                 6.643           217.807
6         250.8             217.420                22.113           248.267
7         312.4             232.720                22.423           263.567
8         308.1             276.500                13.603           307.347
9         384.7             331.260                24.633           362.107
10        400.3             346.440                19.943           377.287
11        421.9             381.440                12.423           412.287
12        472.0             432.880                22.970           463.727
13        333.5             266.547                43.537           297.394
14        289.0             206.847                42.237           237.694
15        240.6             178.947                33.187           209.794
16        242.1             179.947                31.577           210.794
17        237.3             181.367                30.437           212.214
18        274.2             211.827                -6.843           242.674
19        215.5             227.127               -16.433           257.974
20        314.1             270.907                22.397           301.754
21        385.9             325.667                32.277           356.514
22        402.7             340.847                28.437           371.694
23        430.6             375.847                22.767           406.694
24        478.8             427.287               -39.737           458.134
```

Validation Data Comparison:

A table was constructed to compare the actual sales data with the forecasts. This table includes:

- Actual sales data from the validation period.
- The regression forecast, which is the sales prediction based on trend and seasonality.
- The trailing MA for residuals, which represents the smoothed prediction errors from the training period.
- The combined forecast, which is the final forecast after adjusting the regression forecast with the trailing MA for residuals.

```
> round(accuracy(training_forecast_tslm$mean, validation), 3)
             ME   RMSE    MAE    MPE   MAPE   ACF1 Theil's U
Test set 52.028 55.376 52.996 16.884 17.334 -0.055       1.1
> round(accuracy(fst_2level, validation), 3)
            ME  RMSE   MAE   MPE MAPE   ACF1 Theil's U
Test set 21.18 28.43 24.72 6.678 8.32 -0.055     0.569
>
```

Conclusion of the accuracy Measures between the the regression model vs combined model:

- The two-level combined model outperforms the regression model in terms of RMSE (28.43) and MAPE (8.32%) indicating improved accuracy in predicting future sales for the validation period.
- Theil's U value (0.569) suggests that the two-level combined model provides better forecasts compared to a naive model.
- Overall, based on the accuracy measures provided, the two-level combined model with regression and trailing MA for residuals appears to be the better forecasting model for the validation period.

→ *d. Two-level forecast for the 12 future months*

```
> future12_df
   Regression_Fst MA_Residuals_Fst Combined_Fst
1         314.308           11.748      326.056
2         262.137           11.748      273.885
3         226.451           11.748      238.199
4         227.794           11.748      239.542
5         227.380           11.748      239.127
6         257.394           11.748      269.142
7         268.737           11.748      280.485
8         313.480           11.748      325.227
9         373.794           11.748      385.542
10        389.265           11.748      401.013
11        421.337           11.748      433.085
12        472.123           11.748      483.870
```

- The table is the forecast summary for the next 12 months, generated from a two-level forecasting approach. This approach combines a regression forecast with a trailing moving average (MA) of residuals.

→ *e. Comparison of the accuracy measures:*

- The combined model exhibits the lowest RMSE (28.43) and MAPE (8.32) among the models, indicating better overall fit to the data in terms of prediction accuracy.
- The autocorrelation of errors (ACF1) is low for all models, indicating that the residuals are not significantly correlated, which is desirable.
- Considering the trade-off between precision (MAPE) and accuracy (RMSE), the combined model may be the most suitable for forecasting monthly sales in 2022, especially if minimizing forecasting errors is critical.

```
> # Use accuracy() function to identify common accuracy measures.
> # Use round() function to round accuracy measures to three decimal digits.
> round(accuracy(tot_forecast_tslm$fitted, sales.ts), 3)
         ME   RMSE    MAE    MPE   MAPE  ACF1 Theil's U
Test set  0 41.132 33.079 -3.261 14.192 0.697     0.933
> round(accuracy((naive(sales.ts))$fitted, sales.ts), 3)
            ME   RMSE    MAE    MPE   MAPE  ACF1 Theil's U
Test set 2.031 63.357 44.834 -2.101 16.593 0.116         1
> round(accuracy((snaive(sales.ts))$fitted, sales.ts), 3)
            ME   RMSE    MAE    MPE   MAPE  ACF1 Theil's U
Test set 0.233 57.729 40.067 -3.278 17.372 0.601     1.188
> round(accuracy(fst_2level, sales.ts), 3)
            ME  RMSE   MAE   MPE MAPE   ACF1 Theil's U
Test set 21.18 28.43 24.72 6.678 8.32 -0.055     0.569
```

Question 4:

→ a. HW Model

HW Model Summary:

- The model selected by the automated process is ETS(A,N,A).
- Smoothing parameter alpha = 0.4948 for the level
- Seasonal smoothing parameter gamma = 0.0002
- The sigma value, representing the standard deviation of the forecast errors, is 36.6192.

```
> summary(hw_model_ZZZ)
ETS(A,N,A)

Call:
 ets(y = training, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.4948
    gamma = 2e-04

  Initial states:
    l = 321.6044
    s = 162.6948 111.6296 72.6742 59.9108 9.2893 -43.0291
           -52.2506 -83.0115 -84.7847 -89.119 -62.7936 -1.2102

  sigma:  36.6192
```

Model fit is evaluated using:

- AIC (Akaike Information Criterion) = 691.7871,
- AICc (Corrected Akaike Information Criterion) = 702.6962,
- BIC (Bayesian Information Criterion) = 723.2022,
- Lower values suggest a better fit to the data, balancing model complexity and goodness of fit.

```
     AIC     AICc      BIC
691.7871 702.6962 723.2022
```

Training Set Error Measures : Error measures on the training set include:

- ME (Mean Error) = -0.4136574
- RMSE (Root Mean Squared Error) = 32.06355
- MAE (Mean Absolute Error) = 22.5914
- MPE (Mean Percentage Error) = -1.679396
- MAPE (Mean Absolute Percentage Error) = 9.513858
- MASE (Mean Absolute Scaled Error) = 0.4130212
- ACF1 (First Autocorrelation of Errors) = 0.1205568

Lower values generally indicating better performance.

```
Training set error measures:
                     ME     RMSE     MAE       MPE     MAPE
Training set -0.4136574 32.06355 22.5914 -1.679396 9.513858
                  MASE      ACF1
Training set 0.4130212 0.1205568
```

Forecast for the Validation Period:

- The forecast for the validation period shows the point forecasts for each month, along with 80% and 95% prediction intervals. For example, the point forecast for January 2020 is 308.1148 with a 95% prediction interval between 236.34256 and 379.8870. This forecast is based on the model's understanding of the seasonal pattern and level observed in the training data.

```
> hw_forecast_validation
         Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
Jan 2020       308.1148 261.1854 355.0441 236.34256 379.8870
Feb 2020       246.5327 194.1736 298.8918 166.45638 326.6090
Mar 2020       220.2053 162.9289 277.4817 132.60857 307.8020
Apr 2020       224.5372 162.7335 286.3409 130.01661 319.0579
May 2020       226.3116 160.2903 292.3328 125.34070 327.2824
Jun 2020       257.0730 187.0878 327.0582 150.03993 364.1061
Jul 2020       266.2998 192.5635 340.0360 153.52993 379.0697
Aug 2020       318.6095 241.3039 395.9150 200.38085 436.8381
Sep 2020       369.2364 288.5193 449.9536 245.79019 492.6826
Oct 2020       382.0012 298.0109 465.9915 253.54916 510.4533
Nov 2020       420.9524 333.8118 508.0930 287.68239 554.2224
Dec 2020       472.0189 381.8361 562.2017 334.09624 609.9416
Jan 2021       308.1148 214.9910 401.2386 165.69420 450.5354
Feb 2021       246.5327 150.5579 342.5075  99.75195 393.3135
Mar 2021       220.2053 121.4618 318.9487  69.19019 371.2204
Apr 2021       224.5372 123.1006 325.9738  69.40335 379.6711
May 2021       226.3116 122.2515 330.3716  67.16544 385.4577
Jun 2021       257.0730 150.4540 363.6920  94.01334 420.1326
Jul 2021       266.2998 157.1819 375.4177  99.41835 433.1812
Aug 2021       318.6095 207.0486 430.1703 147.99183 489.2271
Sep 2021       369.2364 255.2849 483.1879 194.96266 543.5102
Oct 2021       382.0012 265.7083 498.2942 204.14650 559.8560
Nov 2021       420.9524 302.3642 539.5406 239.58739 602.3174
Dec 2021       472.0189 351.1776 592.8602 287.20810 656.8297
```

Model Evaluation on the Validation Set

- For the validation period, the forecasted values show relatively small errors compared to the actual sales, as indicated by the training set error measures.
- The model's accuracy measures, such as RMSE, MAE, and MAPE, suggest that the forecasts are close to the actual sales values.

```
> round(accuracy(hw_forecast_validation$mean, validation), 3)
             ME   RMSE    MAE   MPE  MAPE   ACF1 Theil's U
Test set 11.596 22.804 17.606 3.527 6.104 -0.016     0.461
```

⟶ *b. 12 future months of 2022 forecast*

Model Summary:

- Smoothing parameter for the level (alpha): 0.4261
- Seasonal smoothing parameter (gamma): 0.0001

```
> summary(tot_hw_model_ZZZ)
ETS(A,N,A)

Call:
 ets(y = sales.ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.4261
    gamma = 1e-04

  Initial states:
    l = 318.8679
    s = 158.2748 106.7147 74.2088 58.6614 0.3927 -35.9219
           -54.337 -85.8696 -84.9243 -86.7219 -51.1724 0.6946

  sigma:  32.2874
```

The sigma value, which is the standard deviation of the forecast errors, is 32.2874.
Model fit metrics include:

- AIC (Akaike Information Criterion): 970.6194
- AICc (Corrected AIC): 977.6782
- BIC (Bayesian Information Criterion): 1007.0816

These metrics are useful for model comparison, where lower values suggest a better balance between model fit and complexity.

```
     AIC      AICc       BIC
970.6194  977.6782 1007.0816
```

Training Set Error Measures:

- Error measures for the model based on the entire dataset are provided, indicating the accuracy of the fitted model:
- The model's accuracy measures on the training set indicate relatively small errors, suggesting a good fit to the historical data.
- The forecast for the next 12 months is provided, allowing for future sales predictions based on the model's estimation of trends and seasonality in the data.

```
Training set error measures:
                     ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 0.06220546 29.47424 19.21026 -1.404785 8.019432 0.4794575 0.1248909
```

```
> hw_forecast_future
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jan 2022       321.7889 280.4110 363.1669 258.5068 385.0711
Feb 2022       269.9217 224.9440 314.8993 201.1343 338.7091
Mar 2022       234.3722 186.0623 282.6821 160.4886 308.2558
Apr 2022       236.1695 184.7429 287.5961 157.5193 314.8197
May 2022       235.2249 180.8600 289.5899 152.0809 318.3690
Jun 2022       266.7562 209.6037 323.9086 179.3491 354.1633
Jul 2022       285.1656 225.3555 344.9758 193.6939 376.6374
Aug 2022       321.4905 259.1358 383.8451 226.1272 416.8537
Sep 2022       379.7595 314.9601 444.5588 280.6574 478.8616
Oct 2022       395.3053 328.1501 462.4604 292.6004 498.0102
Nov 2022       427.8099 358.3789 497.2409 321.6244 533.9955
Dec 2022       479.3690 407.7331 551.0049 369.8114 588.9266
```

→ *c & d. Accuracy Comparison*

```
> #4 c # Identify performance measures for HW forecast.
> round(accuracy(hw_forecast_future$fitted, sales.ts), 3)
            ME   RMSE    MAE    MPE  MAPE  ACF1 Theil's U
Test set 0.062 29.474 19.21 -1.405 8.019 0.125     0.548
> round(accuracy((naive(sales.ts))$fitted, sales.ts), 3)
            ME   RMSE    MAE    MPE   MAPE  ACF1 Theil's U
Test set 2.031 63.357 44.834 -2.101 16.593 0.116         1
> round(accuracy((snaive(sales.ts))$fitted, sales.ts), 3)
            ME   RMSE    MAE    MPE   MAPE  ACF1 Theil's U
Test set 0.233 57.729 40.067 -3.278 17.372 0.601     1.188
```

Comparison and Model Selection

MAPE and RMSE Comparison:

- The Holt-Winters (HW) model shows significantly better performance with the lowest MAPE (8.019) and RMSE (29.474). These metrics indicate higher accuracy and reliability in forecasting, with lower average percentage error and lower root mean squared error, respectively.
- The seasonal naïve forecast has a higher MAPE (17.372) and RMSE (57.729) compared to the HW model, indicating it is less accurate.
- The naive forecast performs the worst among the three, with the highest MAPE (16.593) and RMSE (63.357), suggesting it is the least accurate model for forecasting sales in this context.

Final Choice: The final choice for the forecasting model, in this case, would be the Holt-Winters (HW) model. This decision is based on its superior performance in terms of both MAPE and RMSE, suggesting that it more accurately captures the patterns in the data, including seasonality, and translates them into more precise forecasts. The lower Theil's U statistic for the HW model (0.548) compared to the seasonal naïve forecast further supports this choice, indicating better predictive performance relative to a simple benchmark.

In conclusion, the Holt-Winters (HW) model's ability to account for seasonality with additive errors and its lack of a trend component (as determined by the automatic selection process) makes it particularly well-suited for forecasting the sales data in question. Its superior accuracy metrics make it the preferred model for predicting future sales in this scenario.