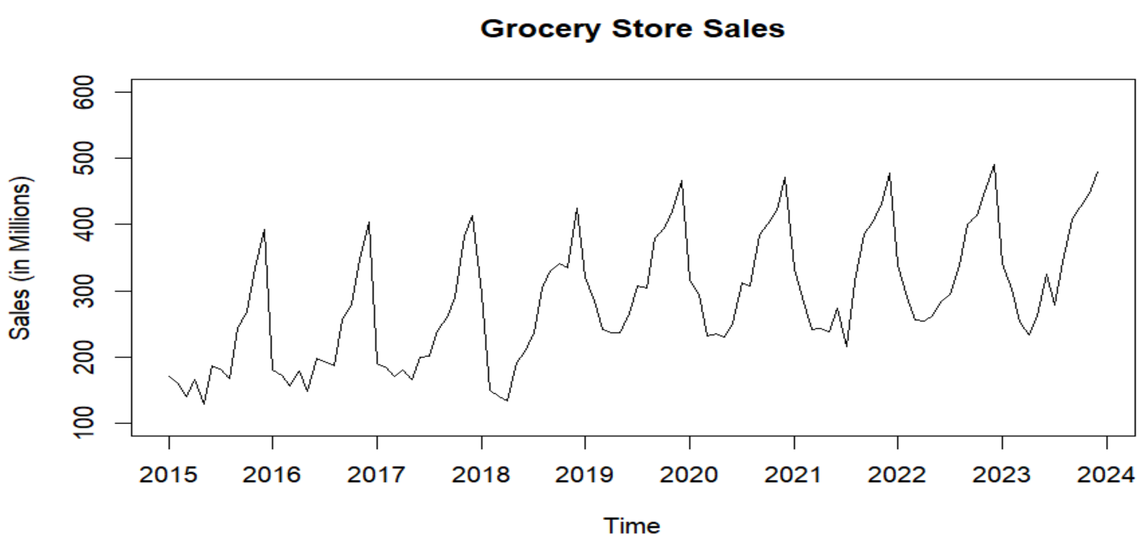
Case Study #1 - Moving Average and Exponential Smoothing

1. Identify time series components and plot the data.
   a. Create time series dataset for sales using ts()

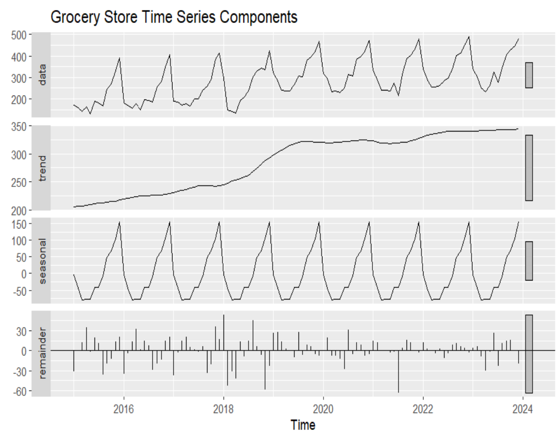
```
# a. Create time series data set sales.ts in R using the ts() function.'
sales.ts <- ts(sales.data$Sales,
               start = c(2015, 1),
               end = c(2023, 12),
               frequency = 12)
sales.ts
```

   b. In this plot for Grocery Store Sales, # we can notice that over the years there is a gradual increase in sales showing that an upward (likely linear) trend exists in this time series. We can also notice that at the start of each year, sales drop, and at the end of each year, sales dramatically increase. Since this pattern repeats consistently over the years, we can conclude that there is also seasonality in this time series for grocery store sales. Since the amplitude of the seasonal fluctuations remain relatively constant over time and the variance of the series is also relatively constant over time, we can conclude that this time series exhibits an Upward Linear Trend with Additive Seasonality.



   c. In the ACF plot for grocery store sales, *three key components are observed*: **Stationarity:** In an ACF plot, a stationary time series would typically show autocorrelation values that quickly drop to zero as the lag increases. Here, aside from a spike at lag 12, the autocorrelations are relatively low, suggesting potential stationarity. **Trend:** A trend exists when there is a long-term increase or decrease in the data. In the AFC plot, there isn't a clear indication of a trend since we do not see a gradual change in the correlation as the lags increase. The absence of a slowly diminishing autocorrelation pattern suggests there may not be a trend in the data. Despite this, if we use autoplot on the seasonal-trend decomposition (stl()), the trend plot suggests that there actually does exist an upward trend in this Grocery store time series. **Seasonality:** There is a pronounced spike at lag 12, which indicates a seasonal pattern. This suggests an annual seasonality component exists, as the sales appear to correlate with their values 12 months prior.

Additionally, if we use autoplot on the seasonal-trend decomposition (stl()), the seasonal plot shows repetitive patterns that occur at regular intervals which is a key sign that there is strong seasonality in the sales data. This is typical for many businesses that have annual cycles influenced by factors such as holidays or events. More people tend to buy groceries as the holiday season progresses.



2. Use trailing MA for forecasting time series.
   a. Data partition with the validation partition of 24 monthly periods (2 years) and training partition of 84 monthly periods (7 years)

```
nValid <- 24
nTrain <- length(sales.ts) - nValid
train.ts <- window(sales.ts, start=c(2015, 1), end=c(2015, nTrain))
valid.ts <- window(sales.ts, start=c(2015, nTrain + 1), end=c(2015, nTrain + nValid))
```

   b. Use the rollmean() function to develop 3 trailing MAs with the window width of 4, 6, and 12 for the training partition.

```
ma.trailing_4 <- rollmean(train.ts, k = 4, align = "right")
ma.trailing_6 <- rollmean(train.ts, k = 6, align = "right")
ma.trailing_12 <- rollmean(train.ts, k = 12, align = "right")
```

   c. Trailing MA forecast with window width of 4 in the validation partition.

```
> ma.trail_4.pred
         Point Forecast     Lo 0     Hi 0
Jan 2022       415.7363 415.7363 415.7363
Feb 2022       390.1537 390.1537 390.1537
Mar 2022       345.4365 345.4365 345.4365
Apr 2022       287.1568 287.1568 287.1568
May 2022       269.7001 269.7001 269.7001
Jun 2022       269.0454 269.0454 269.0454
Jul 2022       284.8164 284.8164 284.8164
Aug 2022       302.3676 302.3676 302.3676
Sep 2022       334.1653 334.1653 334.1653
Oct 2022       360.0781 360.0781 360.0781
Nov 2022       393.9507 393.9507 393.9507
Dec 2022       437.6319 437.6319 437.6319
Jan 2023       428.8682 428.8682 428.8682
Feb 2023       403.2856 403.2856 403.2856
Mar 2023       358.5684 358.5684 358.5684
Apr 2023       300.2887 300.2887 300.2887
May 2023       282.8319 282.8319 282.8319
Jun 2023       282.1773 282.1773 282.1773
Jul 2023       297.9482 297.9482 297.9482
Aug 2023       315.4995 315.4995 315.4995
Sep 2023       347.2972 347.2972 347.2972
Oct 2023       373.2100 373.2100 373.2100
Nov 2023       407.0826 407.0826 407.0826
Dec 2023       450.7638 450.7638 450.7638
```

d. Based on MAPE and RMSE, the trailing MA that performs the best (exhibits the least error) is the trailing MA forecast with window width of 4.

```
The MAPE and RMSE accuracy measures for a trailing MA forecast in the validation peri
od with window width 4 are respectively as follows:  16.198 59.89
 The MAPE and RMSE accuracy measures for a trailing MA forecast in the validation per
iod with window width 6 are respectively as follows:  21.529 77.742
 The MAPE and RMSE accuracy measures for a trailing MA forecast in the validation per
iod with window width 12 are respectively as follows:  18.847 80.55
```

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

a. The following equation represents a regression model with linear trend and seasonality for this Grocery Store Sales time-series: *y = 191.4617 + 1.7910*trend - 40.1767*season2 - 72.7106*season3 - 67.4016*season4 - 73.9783*season5 - 40.8121*season6 - 33.2746*season7 - 9.6084*season8 + 47.7720*season9 + 65.7953*season10 + 105.9900*season11 + 158.7133*Season12158.7133*season12*

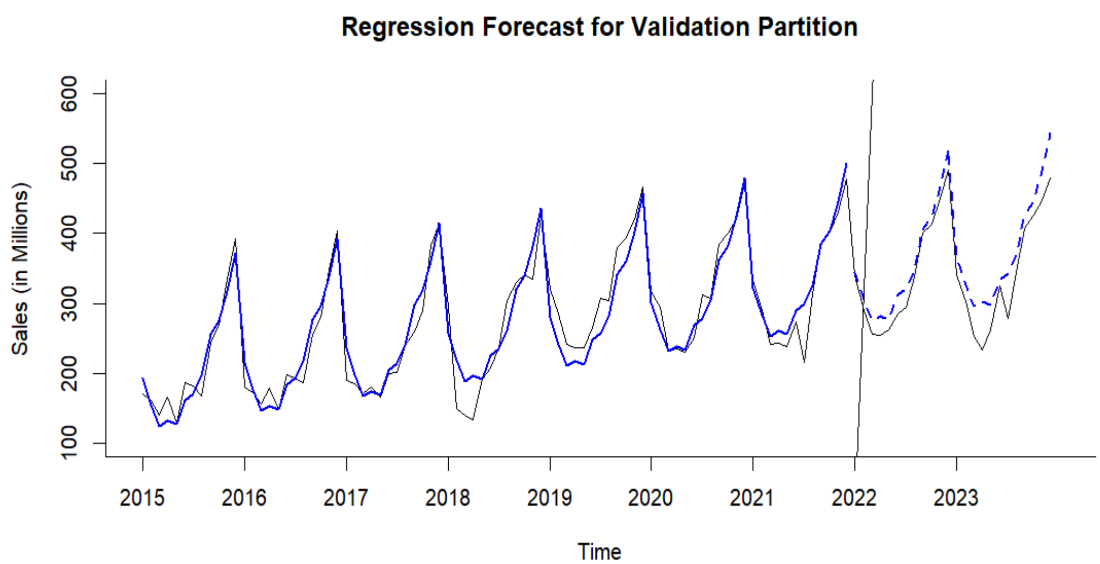
```
Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-84.176 -13.857   0.967  17.228  51.108 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 191.4617    11.6113  16.489  < 2e-16 ***
trend         1.7910     0.1276  14.036  < 2e-16 ***
season2     -40.1767    15.0027  -2.678  0.00920 ** 
season3     -72.7106    15.0044  -4.846 7.16e-06 ***
season4     -67.4016    15.0071  -4.491 2.68e-05 ***
season5     -73.9783    15.0109  -4.928 5.24e-06 ***
season6     -40.8121    15.0157  -2.718  0.00825 ** 
season7     -33.2746    15.0217  -2.215  0.02996 *  
season8      -9.6084    15.0288  -0.639  0.52466    
season9      47.7720    15.0369   3.177  0.00220 ** 
season10     65.7953    15.0461   4.373 4.12e-05 ***
season11    105.9900    15.0564   7.040 9.94e-10 ***
season12    158.7133    15.0677  10.533 3.71e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28.07 on 71 degrees of freedom
Multiple R-squared:  0.9216,	Adjusted R-squared:  0.9083 
F-statistic: 69.55 on 12 and 71 DF,  p-value: < 2.2e-16
```

```
> trend.seas.pred
         Point Forecast     Lo 0     Hi 0
Jan 2022       343.6964 343.6964 343.6964
Feb 2022       305.3107 305.3107 305.3107
Mar 2022       274.5679 274.5679 274.5679
Apr 2022       281.6679 281.6679 281.6679
May 2022       276.8821 276.8821 276.8821
Jun 2022       311.8393 311.8393 311.8393
Jul 2022       321.1679 321.1679 321.1679
Aug 2022       346.6250 346.6250 346.6250
Sep 2022       405.7964 405.7964 405.7964
Oct 2022       425.6107 425.6107 425.6107
Nov 2022       467.5964 467.5964 467.5964
Dec 2022       522.1107 522.1107 522.1107
Jan 2023       365.1884 365.1884 365.1884
Feb 2023       326.8027 326.8027 326.8027
Mar 2023       296.0598 296.0598 296.0598
Apr 2023       303.1598 303.1598 303.1598
May 2023       298.3741 298.3741 298.3741
Jun 2023       333.3313 333.3313 333.3313
Jul 2023       342.6598 342.6598 342.6598
Aug 2023       368.1170 368.1170 368.1170
Sep 2023       427.2884 427.2884 427.2884
Oct 2023       447.1027 447.1027 447.1027
Nov 2023       489.0884 489.0884 489.0884
Dec 2023       543.6027 543.6027 543.6027
```



b. The trailing MA forecast for residuals in the validation period

```
> ma.trail.res.pred
         Point Forecast      Lo 0      Hi 0
Jan 2022      -12.91321 -12.91321 -12.91321
Feb 2022      -12.91321 -12.91321 -12.91321
Mar 2022      -12.91321 -12.91321 -12.91321
Apr 2022      -12.91321 -12.91321 -12.91321
May 2022      -12.91321 -12.91321 -12.91321
Jun 2022      -12.91321 -12.91321 -12.91321
Jul 2022      -12.91321 -12.91321 -12.91321
Aug 2022      -12.91321 -12.91321 -12.91321
Sep 2022      -12.91321 -12.91321 -12.91321
Oct 2022      -12.91321 -12.91321 -12.91321
Nov 2022      -12.91321 -12.91321 -12.91321
Dec 2022      -12.91321 -12.91321 -12.91321
Jan 2023      -12.91321 -12.91321 -12.91321
Feb 2023      -12.91321 -12.91321 -12.91321
Mar 2023      -12.91321 -12.91321 -12.91321
Apr 2023      -12.91321 -12.91321 -12.91321
May 2023      -12.91321 -12.91321 -12.91321
Jun 2023      -12.91321 -12.91321 -12.91321
Jul 2023      -12.91321 -12.91321 -12.91321
Aug 2023      -12.91321 -12.91321 -12.91321
Sep 2023      -12.91321 -12.91321 -12.91321
Oct 2023      -12.91321 -12.91321 -12.91321
Nov 2023      -12.91321 -12.91321 -12.91321
Dec 2023      -12.91321 -12.91321 -12.91321
```

c. Table that contains validation data, regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the validation period.

```
   Sales Regression.Fst MA.Residuals.Fst Combined.Fst
1  338.8        343.696          -12.913      330.783
2  290.3        305.311          -12.913      292.398
3  255.7        274.568          -12.913      261.655
4  253.9        281.668          -12.913      268.755
5  262.1        276.882          -12.913      263.969
6  284.0        311.839          -12.913      298.926
7  294.2        321.168          -12.913      308.255
8  338.6        346.625          -12.913      333.712
9  401.4        405.796          -12.913      392.883
10 414.4        425.611          -12.913      412.698
11 450.1        467.596          -12.913      454.683
12 491.5        522.111          -12.913      509.198
13 340.8        365.188          -12.913      352.275
14 302.2        326.803          -12.913      313.889
15 253.1        296.060          -12.913      283.147
16 233.2        303.160          -12.913      290.247
17 262.3        298.374          -12.913      285.461
18 325.6        333.331          -12.913      320.418
19 278.2        342.660          -12.913      329.747
20 343.8        368.117          -12.913      355.204
21 406.5        427.288          -12.913      414.375
22 426.2        447.103          -12.913      434.189
23 447.4        489.088          -12.913      476.175
24 478.9        543.603          -12.913      530.689
```

```
The MAPE and RMSE accuracy measures for the regression model with linear trend and seas
onality are respectively as follows:  8.638 32.526
 The MAPE and RMSE accuracy measures for the two-level (combined) model with the regres
sion and trailing MA for residuals are respectively as follows:  5.313 22.907
```

The two-level (combined) model with regression and trailing MA for residuals performs better in forecasting for the validation period compared to the regression model with linear trend and seasonality, as it has lower MAPE (5.313 vs 8.638) and RMSE (22.907 vs 32.526) values. Thus, **the two-level (combined) model with regression and trailing MA for residuals is the best forecasting model for the validation period.**

d. A table that contains the regression forecast, trailing MA forecast for residuals, and two-level (combined) forecast in the 12 months of 2024.

```
> future12.df
   Regression.Fst MA.Residuals.Fst Combined.Fst
1         366.802          -22.219      344.583
2         327.269          -22.219      305.050
3         294.058          -22.219      271.839
4         297.169          -22.219      274.950
5         297.591          -22.219      275.372
6         334.247          -22.219      312.028
7         337.369          -22.219      315.150
8         369.391          -22.219      347.172
9         429.358          -22.219      407.139
10        448.402          -22.219      426.183
11        487.380          -22.219      465.161
12        537.880          -22.219      515.661
```

e. The Two-level (combined) model with regression and trailing MA for residuals performs the best in forecasting monthly sales in 2024 based on the lowest MAPE and RMSE values. Specifically, it has the lowest MAPE of 5.227 and the lowest RMSE of 17.232 compared to the other forecasting models.

```
The MAPE and RMSE accuracy measures for the Seasonal Naïve forecast for the entire dataset are respectively as fo
llows:  8.9 36.324
 The MAPE and RMSE accuracy measures for the Regression model with linear trend and seasonality for the entire da
taset are respectively as follows:  7.752 25.193
 The MAPE and RMSE accuracy measures for the Two-level (combined) model with the regression and trailing MA for r
esiduals for the entire dataset are respectively as follows:  5.227 17.232
```

4. Use advanced exponential smoothing methods.

a. The Holt-Winters (HW) model with automated selection was developed using the ets() function in R to forecast the sales time series data. The model uses the following parameters: Alpha (Level Smoothing) of 0.1951, Beta (Trend Smoothing) of 1e-04, and Gamma (Seasonal Smoothing) of 1e-04. The initial states of the model are Level (l) at 201.4349, Trend (b) at 1.569, and Seasonal Components (s) ranging from 157.8048 to -7.2129. The model has an error standard deviation (Sigma) of 27.9821 and information criteria of AIC 948.1414, AICc 957.4142, and BIC 989.4653. The model's performance on the training set shows a Mean Error (ME) of -0.2771, Root Mean Squared Error (RMSE) of 25.17647, and Mean Absolute Percentage Error (MAPE) of 8.377149.

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

  Smoothing parameters:
    alpha = 0.1951
    beta  = 1e-04
    gamma = 1e-04

  Initial states:
    l = 201.4349
    b = 1.569
    s = 157.8048 103.5611 60.5357 42.4259 -13.9408 -25.3492
           -46.1451 -78.2233 -72.2383 -78.0027 -43.2154 -7.2129

  sigma:  27.9821

     AIC     AICc      BIC
948.1414 957.4142 989.4653

Training set error measures:
                     ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set -0.2771099 25.17647 19.35964 -1.131419 8.377149 0.7336283 0.1654209
```

```
> hw.ZZZ.pred
         Point Forecast     Lo 0     Hi 0
Jan 2022       323.1783 323.1783 323.1783
Feb 2022       288.7394 288.7394 288.7394
Mar 2022       255.5209 255.5209 255.5209
Apr 2022       262.8517 262.8517 262.8517
May 2022       258.4331 258.4331 258.4331
Jun 2022       292.0789 292.0789 292.0789
Jul 2022       314.4322 314.4322 314.4322
Aug 2022       327.4177 327.4177 327.4177
Sep 2022       385.3516 385.3516 385.3516
Oct 2022       405.0276 405.0276 405.0276
Nov 2022       449.6175 449.6175 449.6175
Dec 2022       505.4270 505.4270 505.4270
Jan 2023       341.9785 341.9785 341.9785
Feb 2023       307.5396 307.5396 307.5396
Mar 2023       274.3210 274.3210 274.3210
Apr 2023       281.6519 281.6519 281.6519
May 2023       277.2332 277.2332 277.2332
Jun 2023       310.8791 310.8791 310.8791
Jul 2023       333.2323 333.2323 333.2323
Aug 2023       346.2178 346.2178 346.2178
Sep 2023       404.1517 404.1517 404.1517
Oct 2023       423.8278 423.8278 423.8278
Nov 2023       468.4177 468.4177 468.4177
Dec 2023       524.2271 524.2271 524.2271
```

b. The Holt-Winters (HW) model with automated selection was developed using the ets() function in R to forecast the sales time series data. The model's smoothing parameters are Alpha (Level Smoothing) at 0.2017, Beta (Trend Smoothing) at 1e-04, and Gamma (Seasonal Smoothing) at

1e-04. The initial states of the model are Level (l) at 202.496, Trend (b) at 1.3772, and Seasonal Components (s) ranging from 156.7496 to -3.3328. The model has an error standard deviation (Sigma) of 25.1588 and information criteria of AIC 1218.998, AICc 1225.798, and BIC 1264.594. The model's performance on the training set shows a Mean Error (ME) of -0.2093662, Root Mean Squared Error (RMSE) of 23.2205, and Mean Absolute Percentage Error (MAPE) of 7.241739.

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

  Smoothing parameters:
    alpha = 0.2017
    beta  = 1e-04
    gamma = 1e-04

  Initial states:
    l = 202.496
    b = 1.3772
    s = 156.7496 105.029 65.6201 47.3008 -11.9245 -38.2964
           -43.5536 -77.6554 -77.1076 -80.0095 -42.8196 -3.3328

  sigma:  25.1588

     AIC     AICc      BIC
1218.998 1225.798 1264.594

Training set error measures:
                     ME    RMSE      MAE        MPE     MAPE      MASE      ACF1
Training set -0.2093662 23.2205 17.42257 -0.7390224 7.241739 0.7390601 0.1537264
```

```
> tot.hw.ZZZ.pred
         Point Forecast    Lo 0     Hi 0
Jan 2024       345.1928 345.1928 345.1928
Feb 2024       307.0794 307.0794 307.0794
Mar 2024       271.2673 271.2673 271.2673
Apr 2024       275.5425 275.5425 275.5425
May 2024       276.3692 276.3692 276.3692
Jun 2024       311.8473 311.8473 311.8473
Jul 2024       318.4755 318.4755 318.4755
Aug 2024       346.2273 346.2273 346.2273
Sep 2024       406.8265 406.8265 406.8265
Oct 2024       426.5198 426.5198 426.5198
Nov 2024       467.3017 467.3017 467.3017
Dec 2024       520.3950 520.3950 520.3950
```

c. The Holt-Winters (HW) model with automated selection outperforms the Seasonal Naïve forecast in terms of forecasting accuracy for the entire dataset, with a lower MAPE of 7.242 compared to 8.9 and a lower RMSE of 23.221 compared to 36.324. Thus, the Holt-Winters (HW) model with automated selection is the better forecasting model for the entire dataset.

```
The MAPE and RMSE accuracy measures for the Seasonal Naïve forecast for the entire dataset are respectively as
follows:  8.9 36.324
 The MAPE and RMSE accuracy measures for the Holt-Winters (HW) model with automated selection forecast for the
entire dataset are respectively as follows:  7.242 23.221
```

d. The Two-level (combined) model with regression and trailing MA for residuals remains the best forecasting model for the entire dataset, with the lowest MAPE of 5.227 and the lowest RMSE of 17.232. While the Holt-Winters (HW) model with automated selection showed improved performance over the Seasonal Naïve forecast with a MAPE of 7.242 and an RMSE of 23.221 compared to a MAPE of 8.9 and an RMSE of 36.324, it was outperformed by the Two-level model. Given that the dataset had a significant trend over a short period, the Two-level model's ability to capture both trend and seasonality likely contributed to its superior performance. Therefore, the final choice for the forecasting model in this case would still be the Two-level (combined) model with regression and trailing MA for residuals due to its superior accuracy in predicting the monthly sales for the entire dataset, especially when dealing with a significant trend.