Executive Summary

The project utilizes monthly truck sales data from a truck selling company spanning 2003 to 2014. By applying time series analytics, we forecasted truck sales for 24 future periods covering 2015 and 2016. The dataset reveals an upward linear trend combined with multiplicative seasonality—sales typically increase until mid-year and then decline sharply by the end of the year. Additionally, statistically significant autocorrelation across all 12 lags indicates strong dependence within the data.

To forecast truck sales, we implemented four distinct models:

1. A two-level forecasting approach that combines a regression model with a linear trend and seasonality along with a trailing moving average.
2. Holt-Winter's exponential smoothing model.
3. A model incorporating a quadratic trend with seasonal adjustments.
4. An Auto ARIMA model.

For each forecasting model, we assessed the performance using both training/validation splits and the complete dataset. By comparing accuracy metrics—specifically, RMSE and MAPE—we determined that Holt-Winter's model is the optimal choice for forecasting truck sales over the 24 future periods (January 2015 to December 2016).

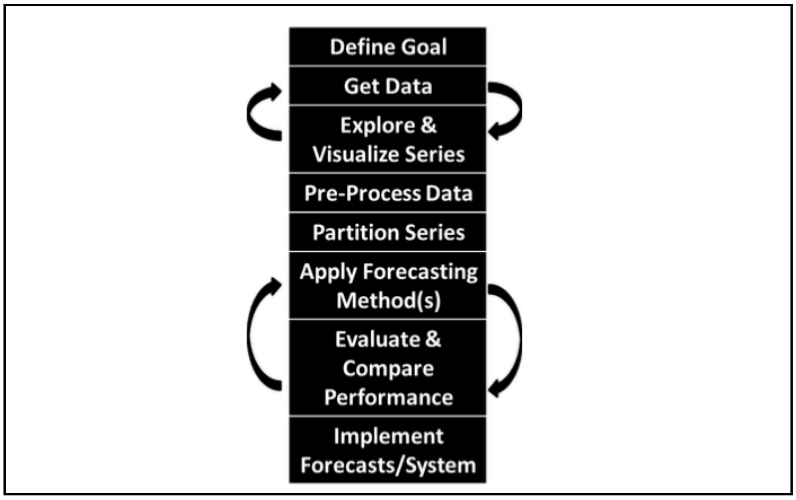
Introduction

This Kaggle-sourced dataset provides a detailed monthly record of truck sales for a specific company from January 2003 to December 2014, comprising 144 observations. Designed for time series analysis, it offers a solid foundation for exploring temporal trends and sales dynamics. The data enables a deep dive into seasonal fluctuations, cyclical trends, and long-term patterns, helping to pinpoint peak sales periods and understand industry-specific events.

By analyzing these temporal patterns, the dataset becomes a strategic tool for decision-making. It supports accurate forecasting of future sales, which in turn allows the company to adjust production schedules, optimize inventory, and streamline production planning. This proactive approach not only enhances market responsiveness—enabling swift adjustments to factors such as fuel prices and regulations—but also provides a competitive edge. Moreover, it lays the groundwork for risk mitigation strategies, equipping the company with the insights needed to prepare contingency plans for economic shocks or supply chain disruptions.

Ultimately, this dataset transforms reactive decision-making into a forward-thinking strategy, steering the company toward sustainable growth and operational excellence in the dynamic truck sales market.

For Analysis we follow this 8 Step forecasting process



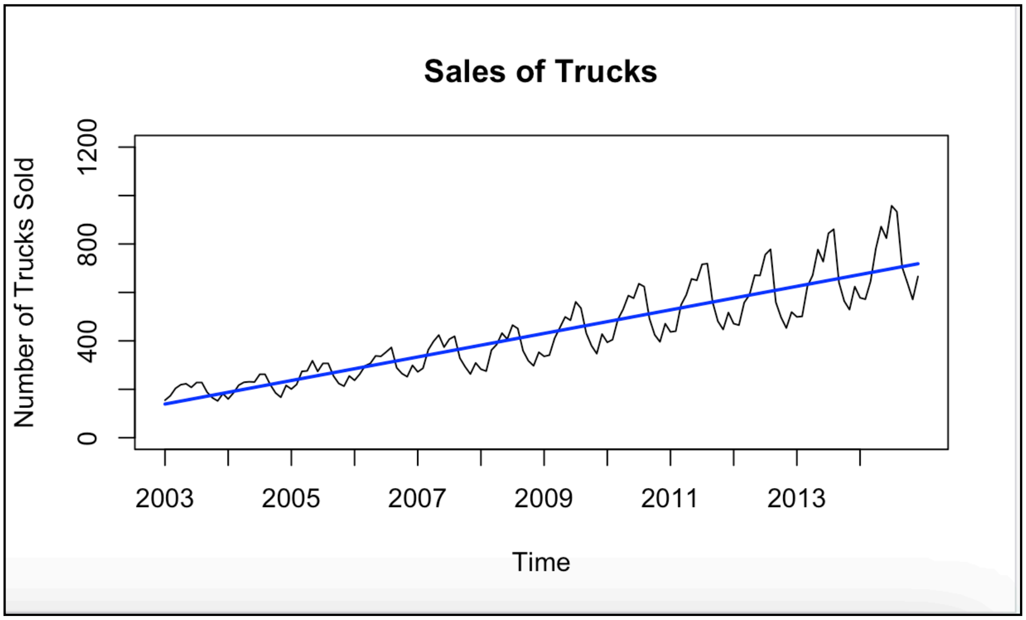
1.Define Goal

The objective of this project is to utilize the historical truck sales from January 2003 to December 2014 to forecast sales for the 24-month period of 2015 and 2016. We will test and compare a range of forecasting models, make predictions, and evaluate the performance of the models with accuracy measures such as RMSE and MAPE. The objective is to determine the most appropriate model to utilize to make truck sales forecasts. By utilizing these findings, the company can make informed choices, optimize its sales strategies, and improve overall operational efficiency

2. Get Data

The Kaggle-sourced dataset is specifically structured for time series analysis and is formatted as a CSV file that logs the end-of-month dates alongside corresponding truck sales figures. It consists of 144 monthly observations, covering January 2003 through December 2014. For effective model training and validation, the dataset is segmented into distinct periods: the training set spans from January 2003 to December 2011, while the validation set includes data from January 2012 to December 2014. Additionally, a future period from January 2015 to December 2016 is provided to facilitate forecasting and evaluate model performance beyond the historical data. This thoughtful segmentation supports robust analysis and ensures reliable insights into the dynamics of truck sales.

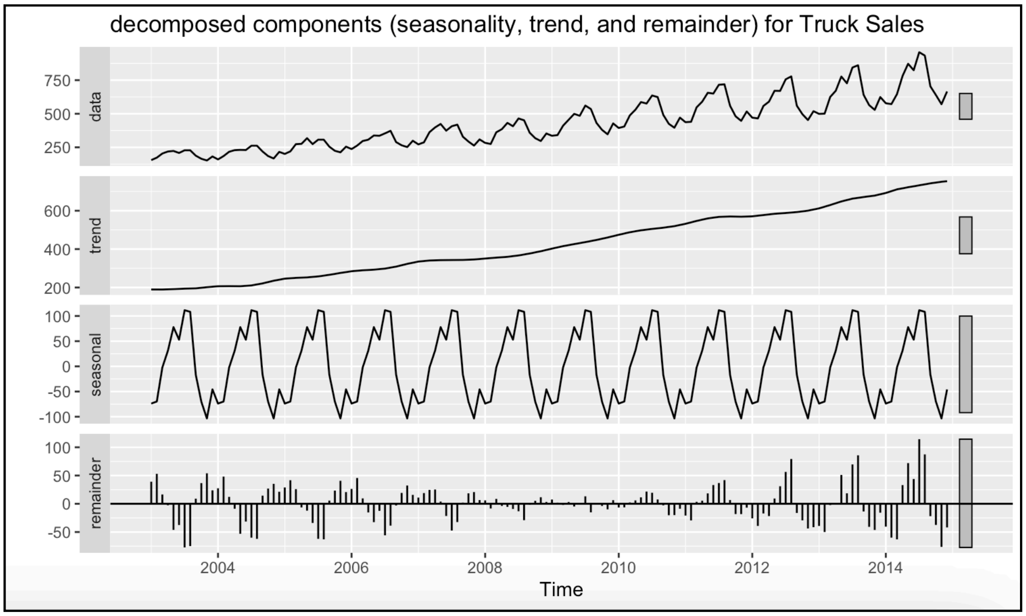
Step 3: Explore and visualize series



The analysis shows that truck sales have been steadily increasing from 2003 to 2014, with noticeable peaks and valleys each year, likely due to seasonal or cyclical factors. The linear trend model, shown by the blue line, captures the general upward movement but does not account for the seasonal fluctuations, which suggests that more sophisticated forecasting models (e.g., seasonal models like Holt-Winter's) may be needed for more accurate future sales predictions

Time Series decomposed components (seasonality, trend, and remainder)

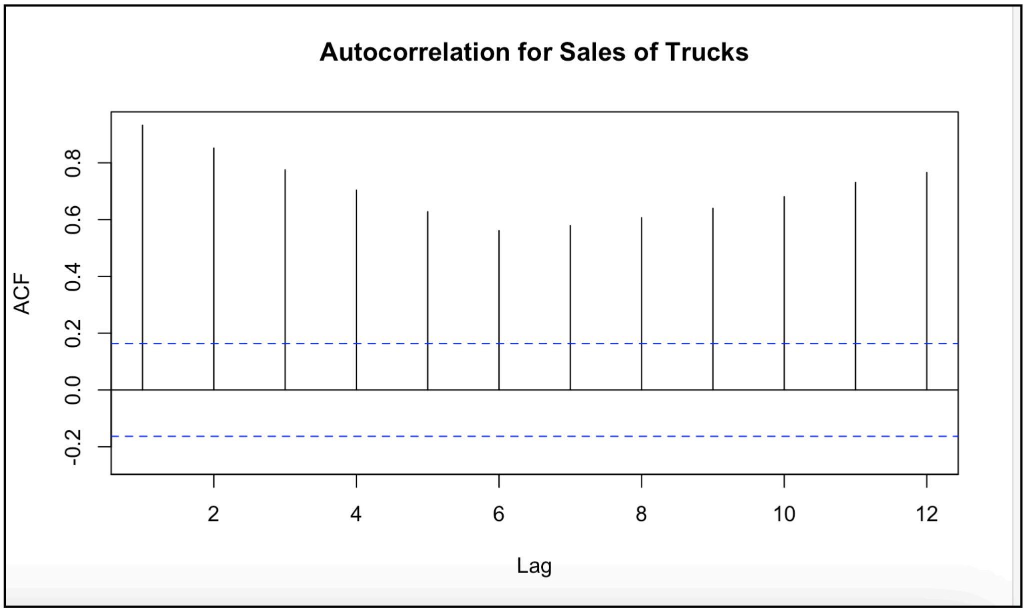
The plot provides a decomposition of truck sales data into three components: trend, seasonality, and remainder. The **data** shows the original sales numbers, highlighting an overall upward trend with



noticeable seasonal fluctuations. The **trend** component reveals a consistent increase in truck sales over time, indicating long-term growth. The **seasonal** component illustrates periodic variations, with peaks and dips occurring at regular intervals each year, showing strong seasonal patterns in the sales. The **remainder** component captures the residual noise or irregularities after removing the trend and seasonality, revealing random fluctuations in the data. This decomposition offers valuable insights into the underlying patterns of truck sales, which can assist in more accurate forecasting and strategic decision-making.

Autocorrelation for truck sales

From the plot, we observe significant peaks at specific lags, indicating strong correlations at those time intervals, suggesting that the truck sales



are highly seasonal, with regular fluctuations occurring each year. The ACF plot helps in identifying such seasonal patterns, which are important for forecasting future truck sales

Time series predictability

AR(1) model (Autoregressive model) to the truck sales time series data

```
> summary(slaes.ar1)
Series: sales.ts
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1      mean
      0.9494  426.1107
s.e.  0.0258   90.5668

sigma^2 = 3870:  log likelihood = -799.27
AIC=1604.55   AICc=1604.72   BIC=1613.46

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 3.005351 61.77734 46.36556 -1.244375 11.0898 0.9457973 0.1732178
```

Based on the summary output and the plot from your AR(1) model, the AR(1) coefficient of **0.9494** indicates a strong persistence in the data, meaning the current value is highly dependent on the previous value. The

intercept value of 426.1107 represents the expected level of the time series when the AR component is not considered (i.e., for $Yt-1 = 0$).

The model's error metrics indicate a reasonable predictive accuracy. The **MAPE (Mean Absolute Percentage Error)** of **11.0898** suggests that the model has moderate reliability in forecasting truck sales.

The AR(1) Model Equation:

$$\mathbf{Yt = 426.1107 + 0.9494\ Yt-1 + \epsilon t}$$

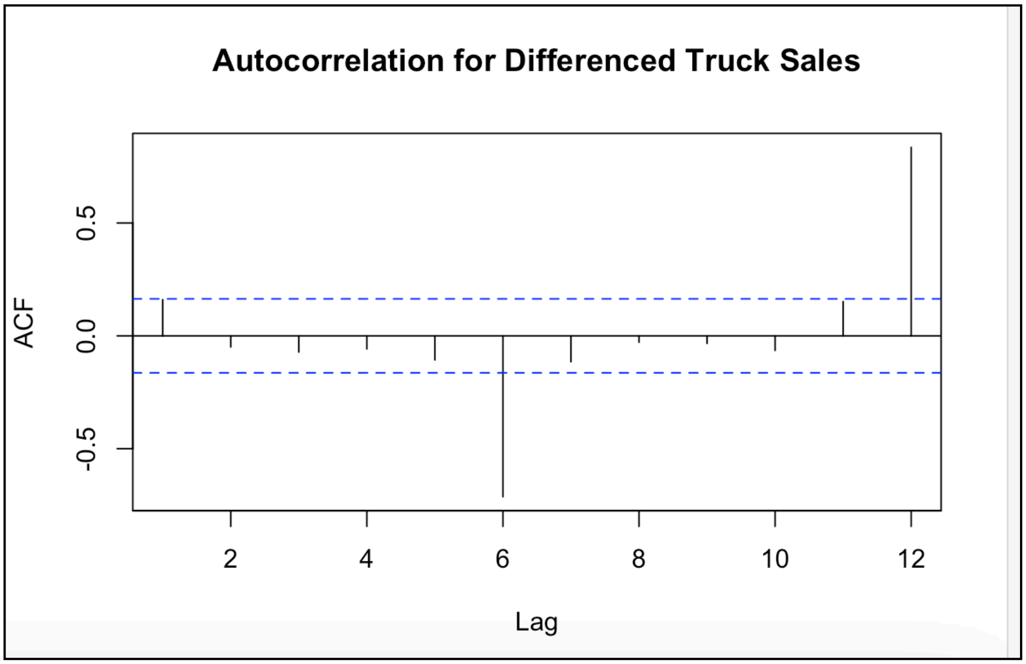
This equation suggests that each month's truck sales is approximately **94.94%** influenced by the sales from the previous month, with the intercept providing the baseline adjustment.

Hypothesis Testing

```
> ar1 <- 0.9494
> s.e. <- 0.0258
> null_mean <- 1
> alpha <- 0.05
> z.stat <- (ar1-null_mean)/s.e.
> z.stat
[1] -1.96124
> p.value <- pnorm(z.stat)
> p.value
[1] 0.0249255
> if (p.value<alpha) {
+   "Reject null hypothesis"
+ } else {
+   "Accept null hypothesis"
+ }
[1] "Reject null hypothesis"
```

From the hypothesis testing we clearly reject the null hypothesis (i.e Ho=1) which mean the trucks data not a random walk .So we can conclude that past appliance trucks sales influence future truck sales. This means the dataset exhibits a predictable pattern, making AR(1) a useful model for forecasting appliance shipments.

Autocorrelation for first differenced Truck Sales Some autocorrelation values at certain lags exceed the confidence bounds, indicating significant dependencies in the data. The presence of a significant spike at lag 6



and Lag12 suggests that past values still have an influence on future values .Since there are some significant autocorrelations at specific lags, the appliance shipments dataset retains predictability

Step 4: Data preprocessing/Step 5: Data partitioning

To ensure precision and consistency in time series forecasting, a subset of the original dataset, spanning from January 2003 to December 2014, was selected for analysis. As part of the data preparation process, the date format was adjusted to reflect actual end-of-month dates, aligning with real-time data while retaining the corresponding sales figures for each month. This adjustment was necessary because the original dataset dynamically altered the year for each month. Before starting the forecasting procedure, the data was divided into training and validation sets.

```
> train.ts
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2003 155 173 204 219 223 208 228 228 188 165 152 182
2004 160 185 217 229 231 230 262 262 219 185 167 216
2005 201 220 274 276 318 274 307 307 255 224 213 255
2006 237 263 297 307 338 336 354 373 289 265 252 299
2007 272 287 363 398 424 374 407 419 329 293 263 309
2008 283 275 362 385 432 407 465 451 359 318 297 353
2009 336 341 411 455 499 485 561 535 432 380 347 428
2010 394 405 488 530 587 576 636 624 492 425 396 471
2011 437 440 548 590 656 650 716 719 560 481 447 517
> valid.ts
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2012 471 465 558 590 671 670 756 778 560 497 453 519
2013 499 501 625 671 777 727 844 861 641 564 529 624
2014 578 572 646 781 872 824 958 933 704 639 571 666
```

Out of 144 total data points, 108 (approximately 75%) were assigned to the training set, and the remaining 36 (about 25%) were designated for validation. This partitioning ensured that the model was properly trained and validated, leading to reliable forecasting results

Step 6: Applying forecasting models

Model 1: Two-level forecasting using regression model with linear trend and seasonality along with trailing moving average (k = 12)

The two-level forecasting model combines a linear trend and seasonality model with a trailing moving average (MA) forecast, using a window width of 12 for the residuals. This time series forecasting technique first employs the linear trend and seasonality model to capture the underlying patterns in the data. The trend component identifies long-term changes, while the seasonality component accounts for recurring patterns with a fixed period (monthly frequency). In the second level, the residuals—representing the differences between the actual values and the Level 1 forecast—are forecasted using a trailing MA approach. This dual-level approach enhances forecasting accuracy by accounting for both the primary patterns in the data and the residual variations

```
> summary(trend.seas)

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

Residuals:
    Min      1Q  Median      3Q     Max 
-58.606 -25.532  -4.181  18.112  99.911 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  87.4162    12.1239   7.210 1.34e-10 ***
trend         3.8282     0.1028  37.256  < 2e-16 ***
season2       8.8384    15.5976   0.567  0.57229    
season3      68.8991    15.5986   4.417 2.65e-05 ***
season4      90.0708    15.6003   5.774 9.73e-08 ***
season5     121.6870    15.6027   7.799 7.96e-12 ***
season6      99.1921    15.6057   6.356 7.15e-09 ***
season7     139.3639    15.6094   8.928 3.23e-14 ***
season8     133.5356    15.6138   8.552 2.04e-13 ***
season9      41.3741    15.6189   2.649  0.00946 ** 
season10     -5.4542    15.6247  -0.349  0.72781    
season11    -31.7269    15.6311  -2.030  0.04518 *  
season12     19.5560    15.6382   1.251  0.21418    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.09 on 95 degrees of freedom
Multiple R-squared:  0.947,	Adjusted R-squared:  0.9403 
F-statistic: 141.5 on 12 and 95 DF,  p-value: < 2.2e-16
```

Above is the summary for the regression model with linear trend and seasonality generated on training data

Equation for above model is

$$\hat{Y_t} = 87.4162+3.8282 \cdot trend+8.8384 \cdot season2 +68.8991 \cdot season3 +90.0708 \cdot season4+121.6870 \cdot season5 +99.1921 \cdot season6+139.3639 \cdot season7 +133.5356 \cdot season8+41.3741 \cdot season9-5.4542 \cdot season10-31.7269 \cdot season11 +19.5560 \cdot season12$$

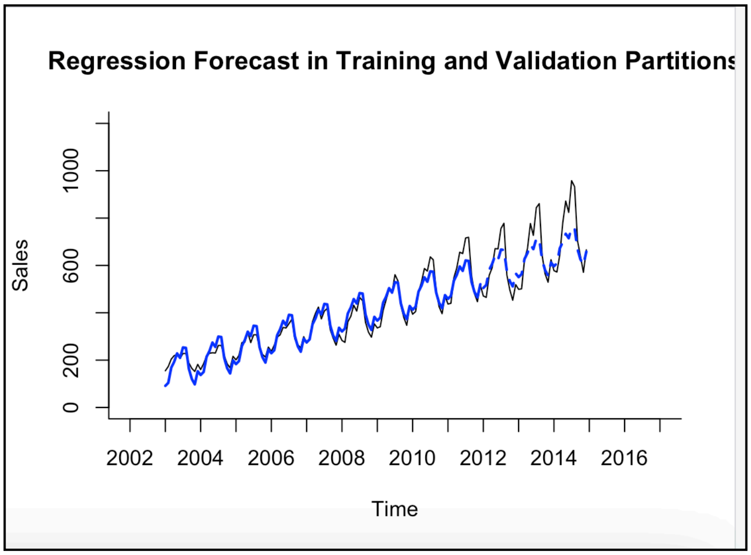
The model demonstrates strong predictive capability, with an R-squared value of 0.947, explaining 94.7% of the variability in sales. The overall statistical significance (p-value < 2.2e-16) confirms the reliability of the model, making it a useful tool for forecasting sales trends while accounting for both long-term growth and seasonal fluctuations.

```
> trend.seas.pred
         Point Forecast     Lo 0     Hi 0
Jan 2012       504.6944 504.6944 504.6944
Feb 2012       517.3611 517.3611 517.3611
Mar 2012       581.2500 581.2500 581.2500
Apr 2012       606.2500 606.2500 606.2500
May 2012       641.6944 641.6944 641.6944
Jun 2012       623.0278 623.0278 623.0278
Jul 2012       667.0278 667.0278 667.0278
Aug 2012       665.0278 665.0278 665.0278
Sep 2012       576.6944 576.6944 576.6944
Oct 2012       533.6944 533.6944 533.6944
Nov 2012       511.2500 511.2500 511.2500
Dec 2012       566.3611 566.3611 566.3611
Jan 2013       550.6333 550.6333 550.6333
Feb 2013       563.3000 563.3000 563.3000
Mar 2013       627.1889 627.1889 627.1889
Apr 2013       652.1889 652.1889 652.1889
May 2013       687.6333 687.6333 687.6333
Jun 2013       668.9667 668.9667 668.9667
Jul 2013       712.9667 712.9667 712.9667
Aug 2013       710.9667 710.9667 710.9667
Sep 2013       622.6333 622.6333 622.6333
Oct 2013       579.6333 579.6333 579.6333
Nov 2013       557.1889 557.1889 557.1889
Dec 2013       612.3000 612.3000 612.3000
Jan 2014       596.5722 596.5722 596.5722
Feb 2014       609.2389 609.2389 609.2389
Mar 2014       673.1278 673.1278 673.1278
Apr 2014       698.1278 698.1278 698.1278
May 2014       733.5722 733.5722 733.5722
Jun 2014       714.9056 714.9056 714.9056
Jul 2014       758.9056 758.9056 758.9056
Aug 2014       756.9056 756.9056 756.9056
Sep 2014       668.5722 668.5722 668.5722
Oct 2014       625.5722 625.5722 625.5722
Nov 2014       603.1278 603.1278 603.1278
Dec 2014       658.2389 658.2389 658.2389
```

```
> trend.seas.res
            Jan         Feb         Mar         Apr         May         Jun         Jul         Aug
2003  63.7555556  69.0888889  36.2000000  26.2000000  -5.2444444  -1.5777778 -25.5777778 -23.5777778
2004  22.8166667  35.1500000   3.2611111  -9.7388889 -43.1833333 -25.5166667 -37.5166667 -35.5166667
2005  17.8777778  24.2111111  14.3222222  -8.6777778  -2.1222222 -27.4555556 -38.4555556 -36.4555556
2006   7.9388889  21.2722222  -8.6166667 -23.6166667 -28.0611111 -11.3944444 -37.3944444 -16.3944444
2007  -3.0000000  -0.6666667  11.4444444  21.4444444  12.0000000 -19.3333333 -30.3333333 -16.3333333
2008 -37.9388889 -58.6055556 -35.4944444 -37.4944444 -25.9388889 -32.2722222 -18.2722222 -30.2722222
2009 -30.8777778 -38.5444444 -32.4333333 -13.4333333  -4.8777778  -0.2111111  31.7888889   7.7888889
2010 -18.8166667 -20.4833333  -1.3722222  15.6277778  37.1833333  44.8500000  60.8500000  50.8500000
2011 -21.7555556 -31.4222222  12.6888889  29.6888889  60.2444444  72.9111111  94.9111111  99.9111111
            Sep         Oct         Nov         Dec
2003  24.7555556  44.7555556  54.2000000  29.0888889
2004   9.8166667  18.8166667  23.2611111  17.1500000
2005  -0.1222222  11.8777778  23.3222222  10.2111111
2006 -12.0611111   6.9388889  16.3833333   8.2722222
2007 -18.0000000 -11.0000000 -18.5555556 -27.6666667
2008 -33.9388889 -31.9388889 -30.4944444 -29.6055556
2009  -6.8777778 -15.8777778 -26.4333333  -0.5444444
2010   7.1833333 -16.8166667 -23.3722222  -3.4833333
2011  29.2444444  -6.7555556 -18.3111111  -3.4222222
```

```
> ma.trail.res
            Jan         Feb         Mar         Apr         May         Jun         Jul         Aug
2003                                     48.8111111  31.5611111  13.8944444  -1.5500000 -13.9944444
2004  37.7152778  35.3138889  22.5791667  12.8722222  -3.6277778 -18.7944444 -28.9888889 -35.4333333
2005  19.2763889  20.6250000  18.3902778  11.9333333   6.9333333  -5.9833333 -19.1777778 -26.1222222
2006  13.3375000  15.6861111   7.7013889  -0.7555556  -9.7555556 -17.9222222 -25.1166667 -23.3111111
2007   7.1486111   5.2472222   4.0125000   7.3055556  11.0555556   6.3888889  -4.0555556 -13.5000000
2008 -23.7902778 -35.6916667 -39.9263889 -42.3833333 -39.3833333 -32.8000000 -28.4944444 -26.6888889
2009 -30.7291667 -32.3805556 -32.8652778 -28.8222222 -22.3222222 -12.7388889   3.3166667   8.6222222
2010 -15.4180556 -16.5694444 -10.3041667  -6.2611111   7.7388889  24.0722222  39.6277778  48.4333333
2011 -16.3569444 -20.0083333 -10.9930556  -2.7000000  17.8000000  43.8833333  64.4388889  81.9944444
            Sep         Oct         Nov         Dec
2003  -6.4944444   5.0888889  25.0333333  38.2000000
2004 -22.1833333 -11.1000000   4.0944444  17.2611111
2005 -25.6222222 -15.7888889  -0.3444444  11.3222222
2006 -19.3111111 -14.7277778  -1.2833333   4.8833333
2007 -21.0000000 -18.9166667 -15.9722222 -18.8055556
2008 -28.6888889 -28.6055556 -31.6611111 -31.4944444
2009   8.1222222   4.2055556 -10.3500000 -12.4333333
2010  40.9333333  25.5166667   4.4611111  -9.1222222
2011  74.2444444  54.3277778  26.0222222   0.1888889
```

```
> fst.2level <- trend.seas.pred$mean + ma.trail.res.pred$mean
> fst.2level
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov
2012 484.2447 480.1236 530.5822 544.8381 571.6872 546.1442 584.6432 578.2424 486.3884 440.5719 415.8742
2013 452.0129 463.5259 526.4919 550.7536 585.6073 566.4681 610.0901 607.7877 519.2124 476.0189 453.4196
2014 492.5799 505.1673 568.9928 593.9421 629.3459 610.6468 654.6208 652.6000 564.2501 521.2368 498.7817
          Dec
2012 469.1827
2013 508.4068
2014 553.8843
```



Above is the plot for the model in training and validation data:

We then find the regression residuals and run a trailing MA model with window width of 12 to forecast the residuals. This incorporates autocorrelation of residuals, if any.

```
> round(accuracy(trend.seas.pred$mean, valid.ts), 3)
             ME   RMSE    MAE   MPE  MAPE  ACF1 Theil's U
Test set 26.617 76.142 57.704 2.135 8.161 0.689     0.758
> round(accuracy(fst.two.level, valid.ts), 3)
             ME   RMSE    MAE    MPE  MAPE  ACF1 Theil's U
Test set -0.121 71.323 59.745 -2.115 9.169 0.689     0.787
```

Based on **RMSE** and **MAPE**, the two-level forecasting model (fst.two.level) is the better choice.

- **RMSE**: The two-level model has a lower RMSE (71.323) compared to the trend.seas model (76.142), indicating better fit and less error in predictions.
- **MAPE**: The trend.seas model has a slightly better MAPE (8.161) compared to the two-level model (9.169), but the difference is minimal.

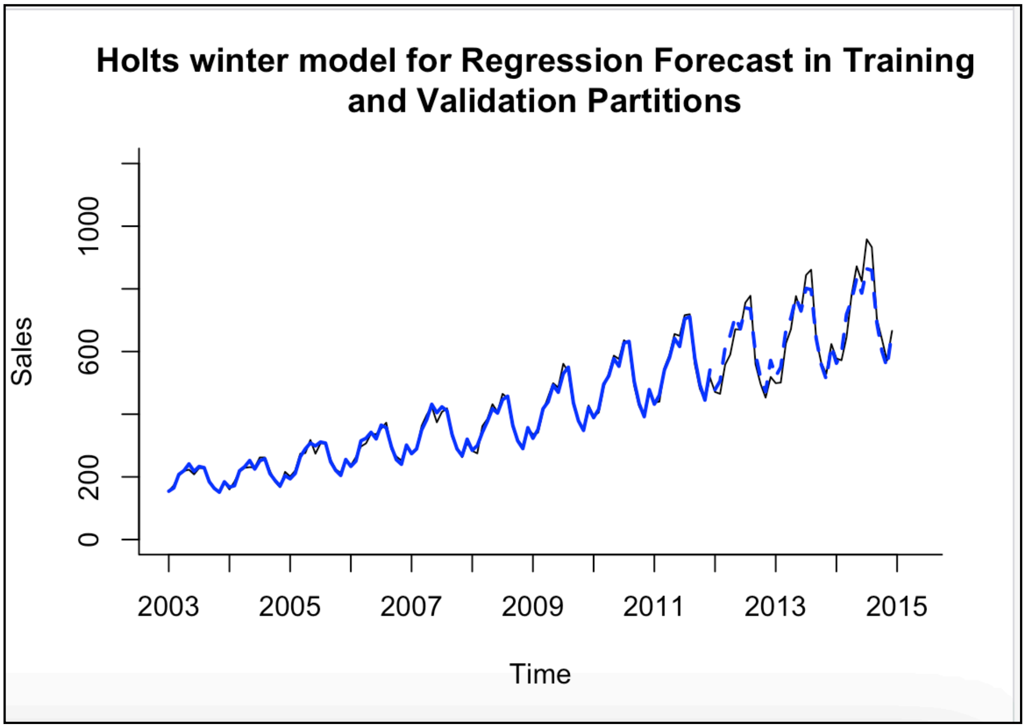
Given that RMSE is a more reliable measure of overall error and the two-level model has a lower RMSE, **the two-level model** should be one of the models for forecasting truck sales

Model 2: Holt-Winter's Model

```
> summary(hw.ZZZ)
ETS(M,A,M)

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

  Smoothing parameters:
    alpha = 0.6673
    beta  = 0.0329
    gamma = 1e-04

  Initial states:
    l = 184.8992
    b = 2.7523
    s = 0.9087 0.7563 0.8323 0.9621 1.2027 1.217
           1.1146 1.1882 1.0951 1.0379 0.8631 0.8221

  sigma:  0.0369

     AIC     AICc      BIC
1061.294 1068.094 1106.890

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 0.5343425 11.86894 9.563203 0.0337578 2.801755 0.2045605 0.1106735
```

The ETS model (Exponential Smoothing State Space Model) appears to be a good fit for the time series data, as indicated by the relatively low error metrics. The **RMSE** (11.86894) and **MAE** (9.563203) suggest that the model's predictions are fairly accurate, while the **MAPE** (2.801755) indicates that the forecast error is about 2.8% on average, which is considered reasonable. The small **Mean Error (ME)** (0.5343425) shows minimal bias in the predictions, and the **ACF1** value of 0.1106735 suggests that there is no significant autocorrelation in the residuals, indicating that the model has captured most of the underlying patterns in the data. Additionally, the model's information criteria (AIC, AICc, and BIC) are within reasonable ranges, further supporting its adequacy.

Overall, the model seems to fit the data well and is likely to provide reliable forecasts.



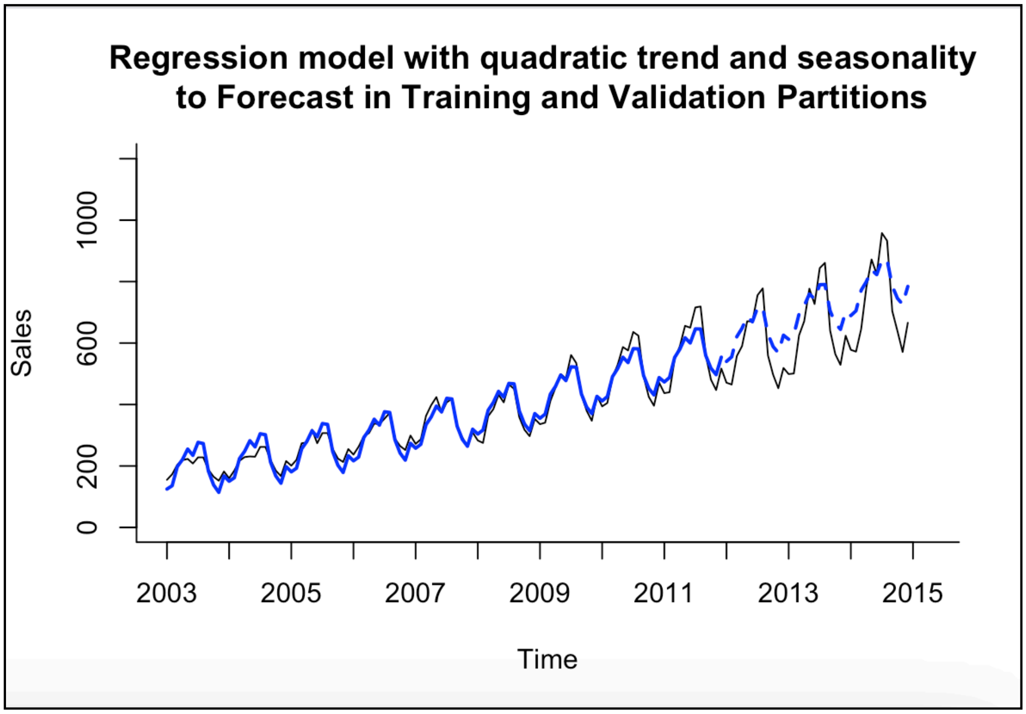
The forecast captures the seasonal and trend patterns of the data. The model fits the training data well and continues the forecast into the validation period, indicating how the model projects future sales trends. The overall trend is upward, with seasonal fluctuations present, and the model shows reasonable accuracy in predicting future sales.

Model 3: Quadratic trend and seasonality

```
Call:
tslm(formula = train.ts ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-51.147 -19.609  -1.225  20.074  73.222 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 122.929950  12.280594  10.010  < 2e-16 ***
trend         1.873191   0.358387   5.227 1.04e-06 ***
I(trend^2)    0.017936   0.003184   5.633 1.84e-07 ***
season2       9.017788  13.558214   0.665  0.50760    
season3      69.221926  13.559180   5.105 1.72e-06 ***
season4      90.501303  13.560746   6.674 1.72e-09 ***
season5     122.189252  13.562882   9.009 2.35e-14 ***
season6      99.730217  13.565572   7.352 7.14e-11 ***
season7     139.901976  13.568807  10.311  < 2e-16 ***
season8     134.037863  13.572585   9.876 3.38e-16 ***
season9      41.804544  13.576915   3.079  0.00272 ** 
season10     -5.131314  13.581815  -0.378  0.70643    
season11    -31.547489  13.587310  -2.322  0.02240 *  
season12     19.556019  13.593437   1.439  0.15357    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28.76 on 94 degrees of freedom
Multiple R-squared:  0.9604,	Adjusted R-squared:  0.9549 
F-statistic: 175.3 on 13 and 94 DF,  p-value: < 2.2e-16
```

The time series linear regression model (tslm) appears to be a good fit for the truck sales data. The model explains approximately 96% of the variance in the data, as indicated by the high R-squared value (0.9604) and the adjusted R-squared value (0.9549), suggesting it captures the underlying patterns effectively. Most of the model coefficients, including the trend, quadratic trend, and seasonal components, are statistically significant, indicating that these variables contribute meaningfully to the model. The F-statistic (175.3) and its associated p-value (less than 2e-16) further confirm the model's overall significance. However, the non-significant p-value for **season2** (0.50760) suggests that it may not contribute significantly to the model and could potentially be excluded.

Overall, the model is robust and provides reliable insights for forecasting truck sales



The model captures the trend and general seasonal patterns in the training data, but in the validation period, the **seasonality is not captured perfectly**, as seen in the forecast deviations. This suggests that while the model handles the trend well, it may struggle with accurately predicting seasonal fluctuations in the future

Model 3: Quadratic trend and seasonality

```
> summary(train.auto.arima)
Series: train.ts
ARIMA(1,1,0)(1,1,0)[12]

Coefficients:
          ar1     sar1
      -0.2033  -0.1708
s.e.   0.1018   0.1027

sigma^2 = 247.4:  log likelihood = -395.76
AIC=797.51   AICc=797.78   BIC=805.18

Training set error measures:
                    ME     RMSE      MAE        MPE     MAPE      MASE         ACF1
Training set 0.5270378 14.59548 10.61751 0.01951244 2.981792 0.2271124 0.007045764
```

The output shows the results of fitting an **ARIMA(1,1,0)(1,1,0)[12]** model to the training time series data. Here's a brief explanation of the key components:

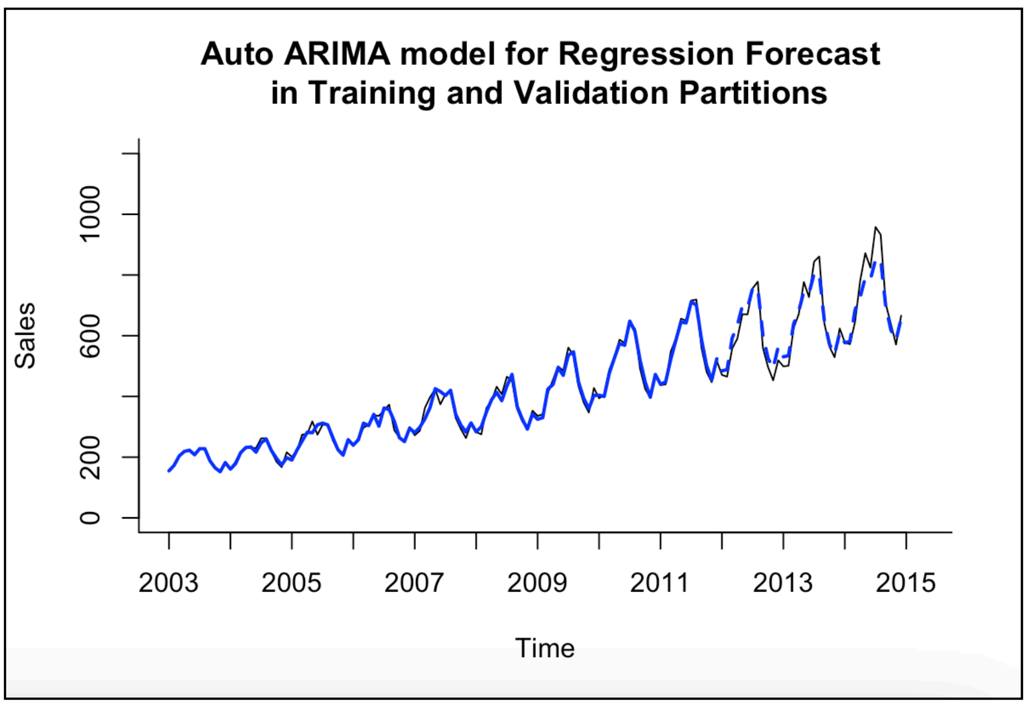
Model Overview:

- **ARIMA(1,1,0)(1,1,0)[12]**:
  - **AR(1)**: One autoregressive term (lag 1).
  - **I(1)**: One order of differencing (to make the series stationary).
  - **MA(0)**: No moving average term.
  - **Seasonal components (1, 1, 0, 12)**: One seasonal autoregressive term, one seasonal difference, and no seasonal moving average term with a period of 12 (monthly data).

Coefficients:

- **ar1 = -0.2033** and **sar1 = -0.1708**: The coefficients for the autoregressive (AR) and seasonal autoregressive (SAR) terms, both of which are statistically significant (since their standard errors are relatively small).

The ARIMA(1,1,0)(1,1,0)[12] model fits the training data well, capturing both trend and seasonal patterns. The coefficients for the autoregressive and seasonal terms are statistically significant. The model shows low error metrics, with a **MAPE** of 2.98%, indicating an average forecast error of about 3%. The **ACF1** value is near zero, suggesting no significant autocorrelation in the residuals. Overall, the model provides a good fit, effectively forecasting the time series data with minimal bias and error.



The graph shows the forecast from an **Auto ARIMA model** applied to truck sales data. The model effectively captures the overall trend and seasonal patterns in the training data, and the dashed lines show the forecast for the validation period (2015 onwards). The forecast follows the seasonal fluctuations well but may slightly deviate in the validation period, indicating that while the model captures the key patterns, there may be some room for improvement in forecasting accuracy.

Step7: Evaluate and compare performance for validation data

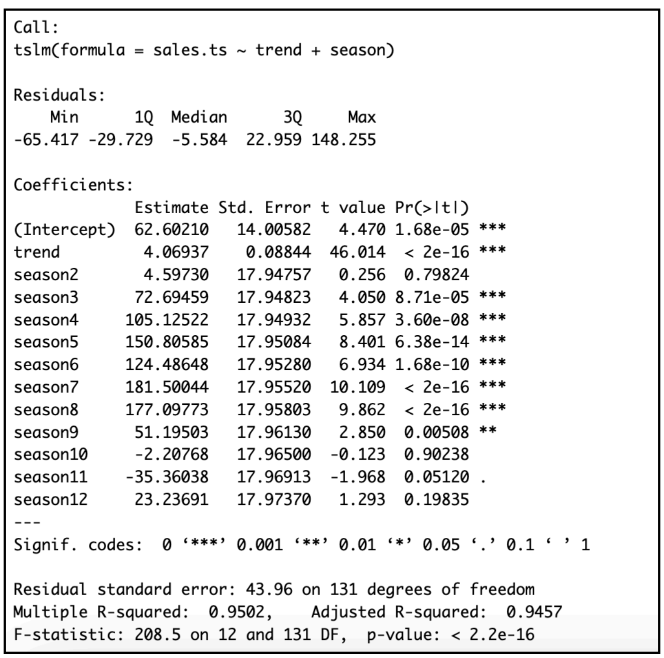
```
> round(accuracy(fst.two.level, valid.ts), 3)
             ME   RMSE    MAE    MPE  MAPE  ACF1 Theil's U
Test set -0.121 71.323 59.745 -2.115 9.169 0.689     0.787
> round(accuracy(hw.ZZZ.pred$mean, valid.ts), 3)
             ME   RMSE    MAE    MPE MAPE  ACF1 Theil's U
Test set -0.841 38.927 31.302 -0.941 4.78 0.605     0.459
> round(accuracy(train.quad.season.pred$mean, valid.ts),3)
              ME   RMSE    MAE    MPE  MAPE ACF1 Theil's U
Test set -50.868 84.511 74.803 -9.791 12.58 0.65     1.068
> round(accuracy(train.auto.arima.pred$mean, valid.ts), 3)
            ME   RMSE    MAE    MPE MAPE  ACF1 Theil's U
Test set 1.417 37.964 29.846 -0.726 4.44 0.584      0.41
>
```

The **Auto ARIMA model** indeed appears to be the best choice, as it has the **lowest RMSE (37.964)** and **MAPE (4.44)** among all models, indicating better forecasting accuracy. The **MAE (29.846)** is also lower than most of the other models, which further supports its strong performance. While other models like the **two-level forecasting model** and **Holt-Winter's model** perform reasonably well, the **Auto ARIMA** model outperforms them in terms of both RMSE and MAPE, making it the most accurate choice for forecasting truck sales.

Running models for the entire data set

To forecast future periods, we first combine the training and validation data and re-run the models on the complete dataset. We then compare the performance of each model, and based on their accuracy, we select the best one for forecasting truck sales for the next 24 months (2016 and 2017). In the following sections, the models will be re-executed on the entire dataset.

Model 1: Two-level forecasting using regression model with linear trend and seasonality along with trailing moving average (k = 12)



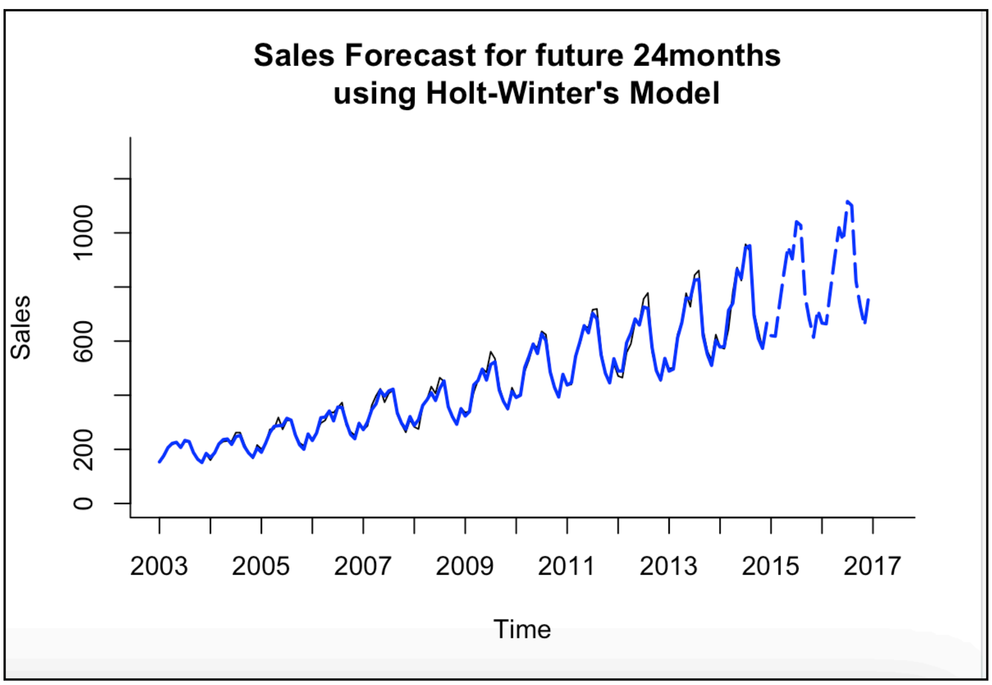
```
Call:
tslm(formula = sales.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-65.417 -29.729  -5.584  22.959 148.255 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  62.60210   14.00582   4.470 1.68e-05 ***
trend         4.06937    0.08844  46.014  < 2e-16 ***
season2       4.59730   17.94757   0.256  0.79824    
season3      72.69459   17.94823   4.050 8.71e-05 ***
season4     105.12522   17.94932   5.857 3.60e-08 ***
season5     150.80585   17.95084   8.401 6.38e-14 ***
season6     124.48648   17.95280   6.934 1.68e-10 ***
season7     181.50044   17.95520  10.109  < 2e-16 ***
season8     177.09773   17.95803   9.862  < 2e-16 ***
season9      51.19503   17.96130   2.850  0.00508 ** 
season10     -2.20768   17.96500  -0.123  0.90238    
season11    -35.36038   17.96913  -1.968  0.05120 .  
season12     23.23691   17.97370   1.293  0.19835    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.96 on 131 degrees of freedom
Multiple R-squared:  0.9502,	Adjusted R-squared:  0.9457 
F-statistic: 208.5 on 12 and 131 DF,  p-value: < 2.2e-16
```

The model appears to be a good fit for the truck sales data, as indicated by the high **Multiple R-squared** value of **0.9502**, meaning it explains approximately 95% of the variance in the data. The **Adjusted R-squared** value of **0.9457** further confirms that the model is robust and not overfitting. The **F-statistic** of **208.5** and its p-value (< 2e-16) indicate that the model is statistically significant, and the coefficients for the trend and seasonal components are largely significant, with p-values well below 0.05 (many even less than 0.001). While some seasonal components, such as **season2**, are not statistically significant, the overall model performance is strong. The **residual standard error** of **43.96** and low residuals (with minimum, median, and maximum values ranging from -65.417 to 148.255) suggest that the model fits the data well, capturing both the trend and seasonal effects accurately

Model 2: Holt-Winter's Model

```
> summary(HW.ZZZ)
ETS(M,A,M)

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

  Smoothing parameters:
    alpha = 0.3738
    beta  = 0.0084
    gamma = 0.4603

  Initial states:
    l = 184.794
    b = 2.0932
    s = 0.9015 0.7461 0.8173 0.9495 1.1635 1.1883
           1.0684 1.1734 1.1584 1.0834 0.9287 0.8217

  sigma:  0.0396

     AIC     AICc      BIC
1518.383 1523.240 1568.870

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 2.465506 16.87606 11.89977 0.3780356 2.776921 0.2427399 0.2153029
```

The **ETS model (Exponential Smoothing State Space Model)** with the automatic "ZZZ" configuration is a good fit for the truck sales data. The **AIC (1518.383)**, **AICc (1523.240)**, and **BIC (1568.870)** values suggest the model is well-calibrated. The smoothing parameters, particularly the seasonal component (gamma = 0.4603), show that the model effectively captures seasonality. The training set error measures, including **RMSE (16.87606)**, **MAE (11.89977)**, and **MAPE (2.77691)**, indicate that the model's predictions are close to the actual sales, with a forecast error of about 2.78%. The **ACF1** value of **0.2153029**, close to zero, confirms that the residuals do not exhibit significant autocorrelation. Overall, the model provides a reliable and accurate forecast, making it a strong fit for the data.

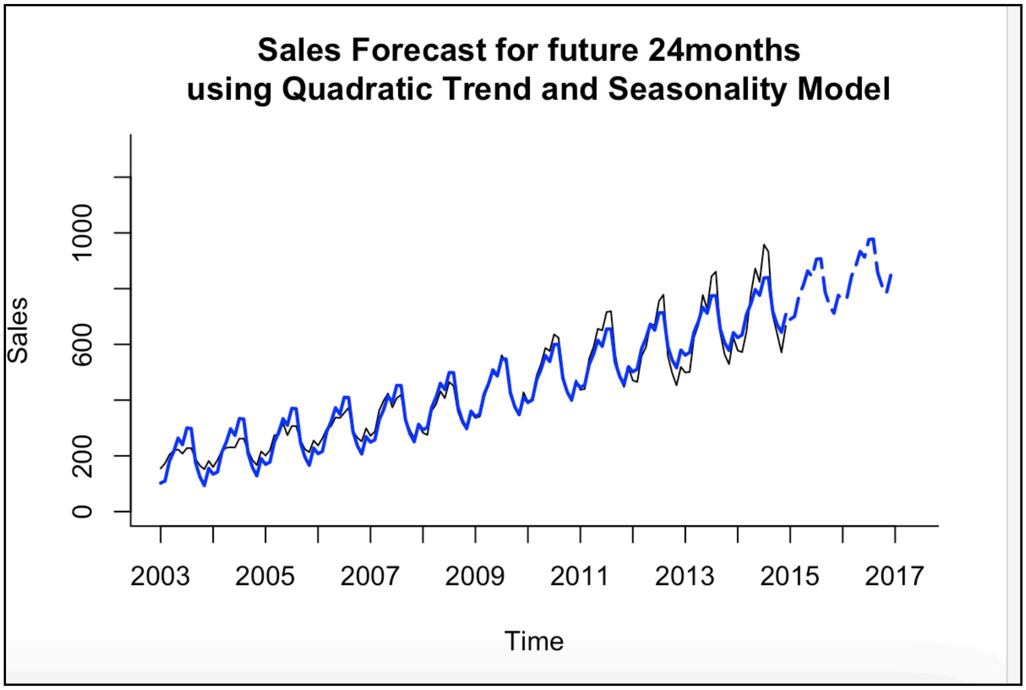


The plot Above shows the historical data and future forecasts

Model 3: Quadratic trend and seasonality

```
Call:
tslm(formula = sales.ts ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-72.519 -28.886  -1.213  26.834 119.507 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  99.89190   15.04803   6.638 7.82e-10 ***
trend         2.52891    0.32786   7.714 2.82e-12 ***
I(trend^2)    0.01062    0.00219   4.852 3.44e-06 ***
season2       4.70353   16.57805   0.284  0.77708    
season3      72.88582   16.57869   4.396 2.27e-05 ***
season4     105.38019   16.57973   6.356 3.22e-09 ***
season5     151.10332   16.58117   9.113 1.26e-15 ***
season6     124.80519   16.58300   7.526 7.70e-12 ***
season7     181.81915   16.58521  10.963  < 2e-16 ***
season8     177.39520   16.58781  10.694  < 2e-16 ***
season9      51.45000   16.59080   3.101  0.00236 ** 
season10     -2.01645   16.59418  -0.122  0.90347    
season11    -35.25414   16.59797  -2.124  0.03556 *  
season12     23.23691   16.60217   1.400  0.16401    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.61 on 130 degrees of freedom
Multiple R-squared:  0.9579,	Adjusted R-squared:  0.9537 
F-statistic: 227.4 on 13 and 130 DF,  p-value: < 2.2e-16
```

The output shows the results of a **time series linear regression model** with a quadratic trend and seasonal components. The model explains a significant portion of the variance in the data, with a **Multiple R-squared** of **0.9579** and an **Adjusted R-squared** of **0.9537**, indicating a strong fit. Most of the coefficients are statistically significant, as reflected by their **p-values** (many less than 0.001), except for **season2**, **season9**, and **season12**, which have higher p-values, indicating less significance. The **F-statistic (227.4)** and its very low p-value (**< 2e-16**) show that the model is statistically significant overall. The **residual standard error** is **40.61**, suggesting that while the model fits well, there is still some error in the predictions. Overall, the model captures both the trend and seasonal variations in truck sales effectively, providing reliable forecasts

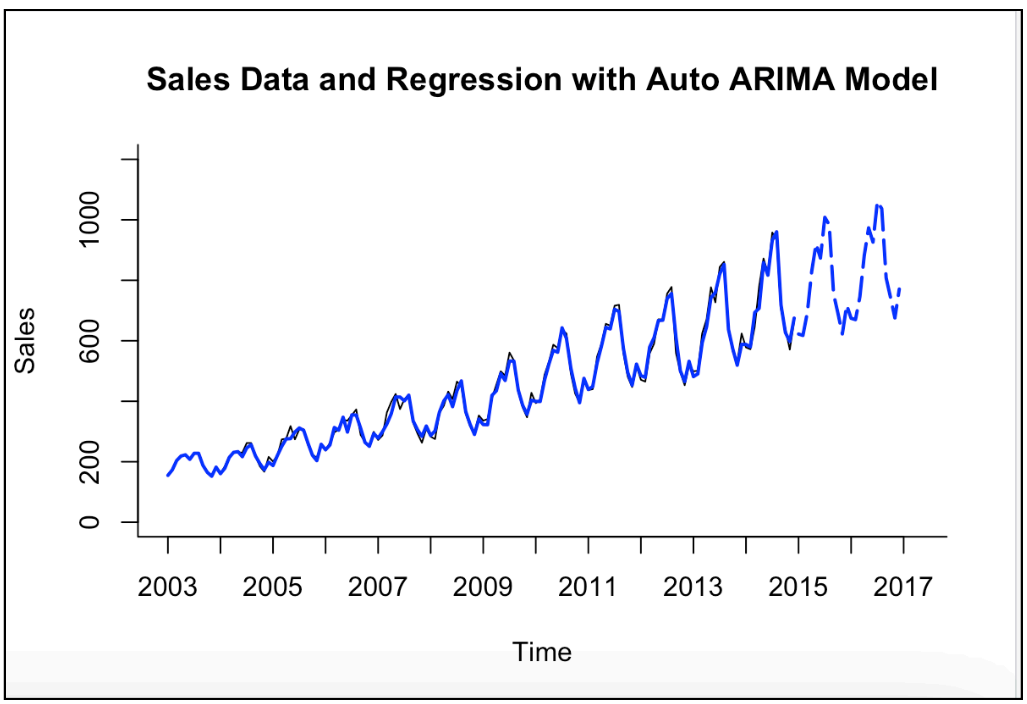


The plot Above shows the historical data and future forecasts

Model 4: Auto ARIMA

```
> summary(auto.arima)
Series: sales.ts
ARIMA(2,1,1)(0,1,0)[12]

Coefficients:
         ar1     ar2      ma1
      0.5988  0.1955  -0.9816
s.e.  0.0889  0.0882   0.0287

sigma^2 = 349.1:  log likelihood = -568.49
AIC=1144.98   AICc=1145.3   BIC=1156.48

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE     MASE         ACF1
Training set 2.176789 17.61612 12.83449 0.3775811 2.966608 0.261807 -0.002502738
```

The **ARIMA(2,1,1)(0,1,0)[12]** model provides a good fit for the truck sales data. The **AIC** (1144.98), **AICc** (1145.3), and **BIC** (1156.48) values suggest that the model effectively balances fit and complexity. The coefficients for the autoregressive and moving average terms (**ar1 = 0.5988, ar2 = 0.1955, ma1 = -0.9816**) are statistically significant, indicating that the model captures the key dependencies in the data. The training set error measures, including **RMSE (17.61612)**, **MAE (12.83449)**, and **MAPE (2.966608)**, show moderate prediction accuracy, with the forecast error being about 2.97%. Additionally, the **ACF1** value of **-0.0025** suggests minimal autocorrelation in the residuals, indicating that the model captures the patterns in the data well. Overall, the model is a reliable choice for forecasting truck sales.



The plot Above shows the historical data and future forecasts

Comparing performance of Models for entire dataset:

Based on the accuracy metrics, the **Auto ARIMA model** appears to be the best choice for forecasting. It has the lowest **RMSE (17.616)** and **MAE (12.834)** compared to the other models, indicating superior prediction accuracy. The **MAPE (2.967)** is also relatively low, suggesting that the forecast error is small. In contrast, while the **Holt-Winter's model** (HW.ZZZ) has good performance with a **RMSE of 16.876**, its **ACF1 (0.296)** indicates some residual autocorrelation, which is not ideal. The **two-level forecasting model** (tot.trend.seas) shows reasonable performance but has a higher **RMSE (36.889)**, making it less accurate than the Auto ARIMA model. Similarly, the **quadratic seasonality model** has a high **MAPE (8.686)**, which indicates less accuracy compared to the Auto ARIMA model. Overall, the **Auto ARIMA model** provides the best fit for the data with the lowest error metrics

Step 8: Implementation

Overall, the **Auto ARIMA model** provides the best fit for the data with the lowest error metrics.When forecasting the sales trends for the next two years (2015 and 2016) using the **Auto ARIMA model**, we obtain the following forecast:

```
> auto.arima.pred
         Point Forecast      Lo 0      Hi 0
Jan 2015       622.2831  622.2831  622.2831
Feb 2015       617.6502  617.6502  617.6502
Mar 2015       692.9151  692.9151  692.9151
Apr 2015       828.9398  828.9398  828.9398
May 2015       920.8006  920.8006  920.8006
Jun 2015       873.5164  873.5164  873.5164
Jul 2015      1008.1133 1008.1133 1008.1133
Aug 2015       983.6107  983.6107  983.6107
Sep 2015       755.0252  755.0252  755.0252
Oct 2015       690.3706  690.3706  690.3706
Nov 2015       622.6585  622.6585  622.6585
Dec 2015       717.8984  717.8984  717.8984
Jan 2016       674.3814  674.3814  674.3814
Feb 2016       669.9151  669.9151  669.9151
Mar 2016       745.3189  745.3189  745.3189
Apr 2016       881.4593  881.4593  881.4593
May 2016       973.4166  973.4166  973.4166
Jun 2016       926.2128  926.2128  926.2128
Jul 2016      1060.8766 1060.8766 1060.8766
Aug 2016      1036.4298 1036.4298 1036.4298
Sep 2016       807.8908  807.8908  807.8908
Oct 2016       743.2750  743.2750  743.2750
Nov 2016       675.5952  675.5952  675.5952
Dec 2016       770.8620  770.8620  770.8620
```

DataSet

https://www.kaggle.com/datasets/ddosad/dummy-truck-sales-for-time-series/data