Forecasting Alcoholic Beverages Wholesale Trade: A Time Series Analysis

Sarath Kumar Vatyam (ss2405)

Abitha Pasupuleti (gz8641)

Soundarya Chandra Mohan (rp7088)

Urmila Desetty (qg2774)

EXECUTIVE SUMMARY

"I have seen the future, and it is very much like the present, only longer."

-Kehlog Albran, *The Profit*

This project involves a comprehensive examination of historical wholesale trade: sales and inventory data with respect to the Beer, Wine and Distilled Alcoholic Beverages domain to uncover recurring patterns and trends. Using time series analysis techniques, we seek to uncover valuable insights into the dynamics of sales behavior over time, including seasonal fluctuations and long-term trends, to better understand the evolution of sales in this sector.

The data encompasses an 11-year span from 2013 to 2023. A variety of models, such as ARIMA, Auto ARIMA, and regression-based models featuring linear trend and seasonality, were applied. Visual examination revealed predominantly upward trend. Statistical significance was verified. Additionally, model evaluation was conducted, with accuracy measures including Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) taken into consideration. The best and optimized results were obtained using the Auto ARIMA model. The evaluation is done based on RMSE and MAPE metrics. By understanding historical sales patterns and forecasting future trends, businesses can make data-driven decisions to optimize operations and capitalize on emerging opportunities.

INTRODUCTION

The wholesale trade of Beer, Wine, Alcoholic Beverages are integral components of the United States economy, playing pivotal roles in driving consumption and economic growth. Accurate forecasting of monthly sales within these sectors is paramount for operational success and strategic decision-making. Leveraging time series analysis methodologies offers a robust framework for understanding sales data dynamics over time, unveiling trends, patterns, and seasonal fluctuations crucial for informed predictions and optimized sales strategies.

This project embarks on a comprehensive time series analysis of monthly sales data spanning approximately 11 years within the wholesale trade of alcoholic beverages. The primary objective is to dissect the dataset, identifying underlying patterns in monthly sales trends, understanding their evolution, uncovering recurring seasonal fluctuations, long-term trends, and anomalies impacting sales performance.

The methodology adheres to the eight steps of forecasting, employing various models including Regression-based models, Holt-Winter's Exponential smoothing model, and autoregressive integrated moving average models (ARIMA). Comparative analysis of forecasts utilizing parameters such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) will be conducted to determine the most suitable model for sales forecasting and trend identification. Ultimately, this holistic approach to time series analysis not only enhances short-term sales forecasting accuracy but also empowers businesses to foster long-term sustainability and resilience amidst evolving market dynamics.

THE FORECASTING JOURNEY BEGINS

STEP 1 - GOAL DEFINITION

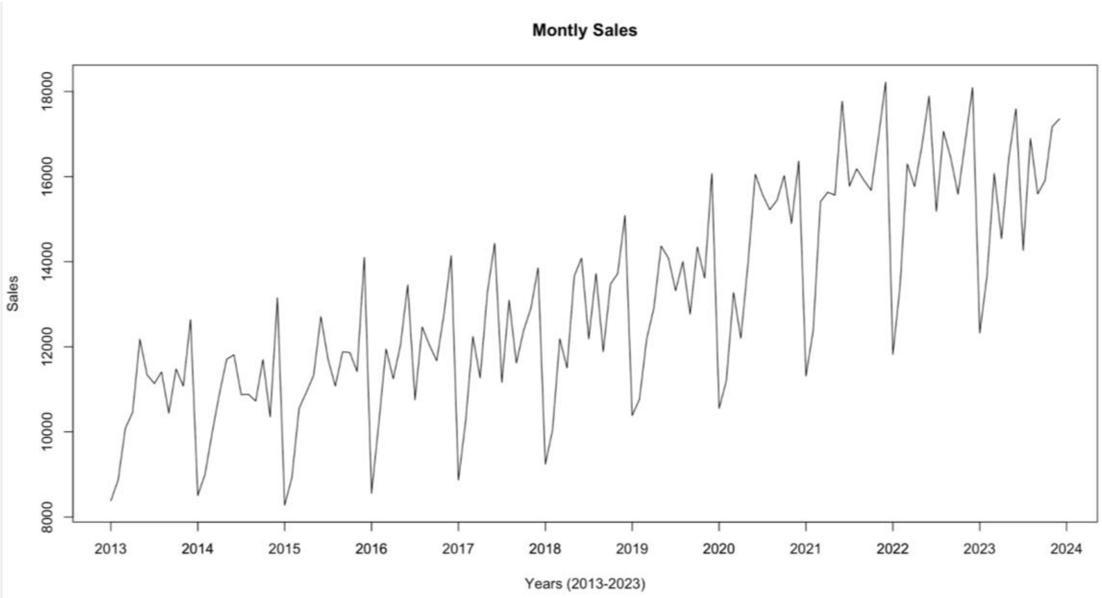
The primary aim of this time series analysis project is to delve into the trends and patterns of sales within the Wholesale Trade of Alcoholic Beverages Industry over the specified timeframe. This analysis is geared towards furnishing valuable insights to aid decision-making, forecasting, and strategic planning endeavors. The key objective is to identify and scrutinize any seasonal fluctuations or patterns present in sales data, enabling businesses to anticipate and effectively prepare for peak and off-peak periods. Additionally, anomalies will be pinpointed, allowing for the detection of any unusual spikes or drops in sales that may signify exceptional events or factors impacting the Wholesale Trade of Alcoholic Beverages. Understanding these anomalies is paramount for proactive decision-making. The dataset is segmented monthly and will be utilized for forecasting future trends. The project will culminate in a succinct summary of key findings and insights, accompanied by actionable recommendations tailored to stakeholders, aimed at enhancing decision-making processes and strategic planning initiatives.

STEP 2 - INFORMATION OF DATA & DATASET

The provided data is organized monthly. Each row in the dataset corresponds to a specific month, and the column contains the month and year in a formatted text, such as "Jan-13" for January 2013. This monthly granularity allows for a detailed examination of trends, patterns, and fluctuations in the Wholesale Trade of Alcoholic Beverages over the entire period covered by the dataset, from January 2013 to December 2023.

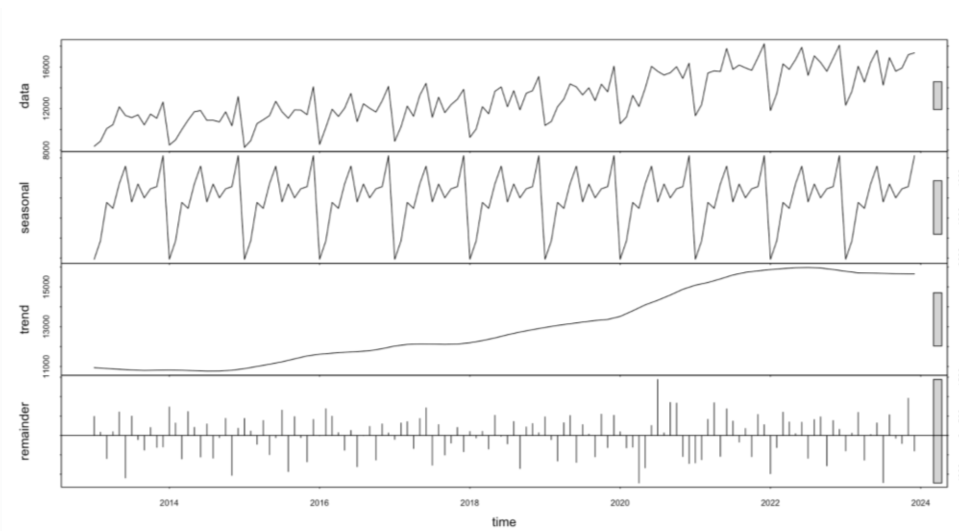
STEP 3 - Explore and Visualize Data (Descriptive Analytics)

Time series plot



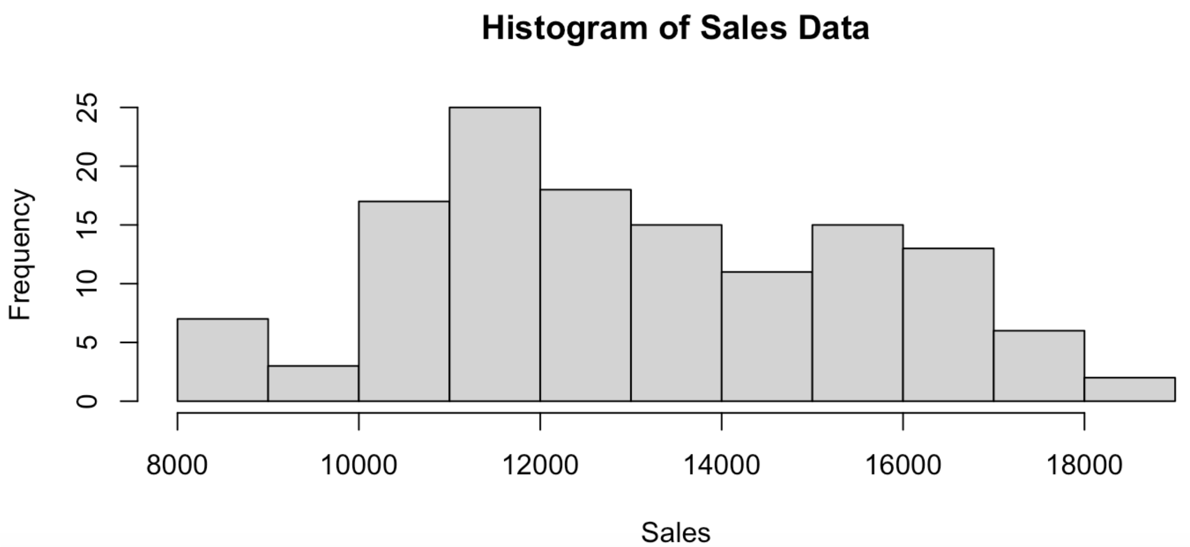
The time series plot reveals an upward trend in monthly sales. The trend is almost constant from 2013 to 2016, but then the trend is going up from 2017 all the way to 2023. There's a seasonal pattern with lower sales at the beginning of the year (January) and higher sales at the end of the year (December). This indicates that the data has both trend and seasonality.

Stl() PLOT



The STL plot reveals there is an upward trend and seasonality in the data, there is yearly seasonality in the data which indicates minimum sales at the beginning of the year and maximum sales at the end of the year. The Remainder/Residuals explain the non-systematic noise and systematic level component of the data.

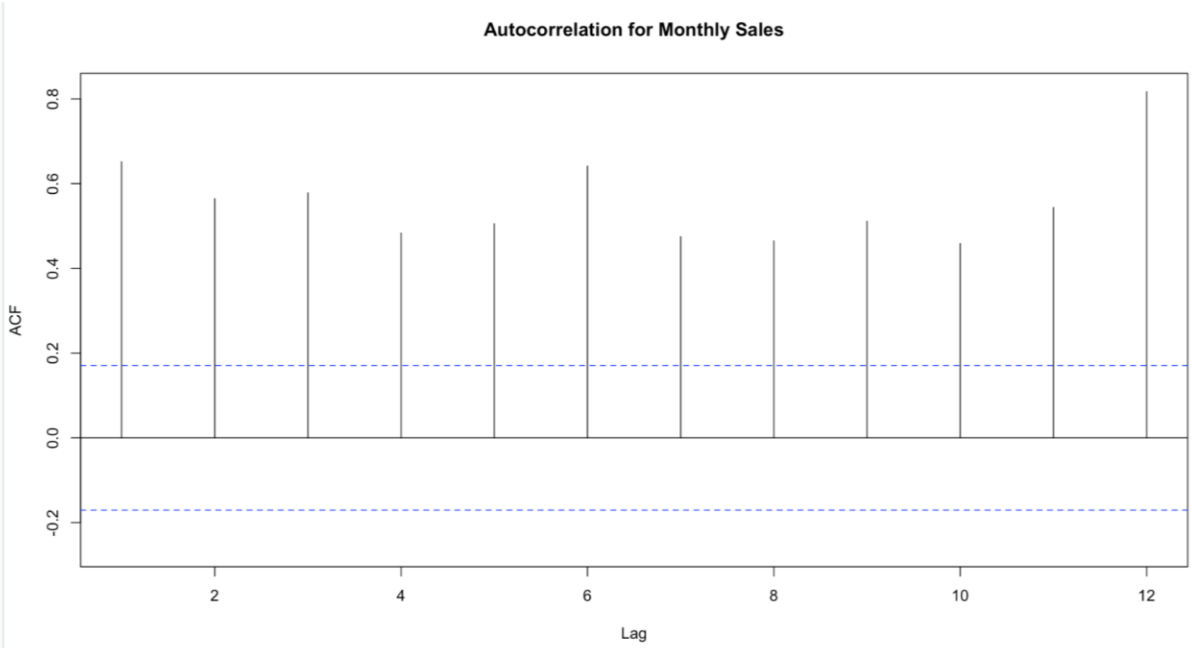
HISTOGRAM OF SALES DATA



The sales data histogram shows a central tendency around 8,000-12,000 units, with a slight right skew suggesting occasional high sales months (possibly exceeding 18,000 units) and low sales months (below 4,000 units).

CORRELOGRAM - ACF ()

The correlogram reveals strong positive autocorrelation at all lags and higher than the horizontal threshold (significantly greater than zero) and so the autocorrelation coefficients are significant. There is very strong positive autocorrelation at lag 12 indicating seasonality of the data. And the strong positive autocorrelation at lag 1 indicates there is an upward trend. Overall, the positive autocorrelation suggests that the current values of the series are strongly correlated with its recent historical data.



STEP 4 - DATA PRE PROCESSING

The dataset, named sales.ts, underwent rigorous preprocessing to ensure data integrity. Duplicates were systematically eliminated to streamline the dataset, while missing values were addressed meticulously to ensure data completeness. To standardize temporal information, the "Period" column was converted to a date format using the `as.Date()` function in R. Concurrently, numerical precision was optimized by converting the "Value" column to numeric format, with extraneous commas removed beforehand. These steps facilitated seamless integration into the analysis pipeline, enhancing the accuracy of subsequent analyses. The meticulous attention to data preprocessing underscored a commitment to data quality and integrity, laying a robust foundation for further analysis and model development. The data used in the project is from Jan 2013 to Dec 2023.

STEP 5 – PARTITION SERIES

The dataset is partitioned into two parts – Training and Validation. The training partition is used to train the forecasting models and the set consists of 132 records from the period of January 2013 to December 2023. The validation partition is used to validate the performance of the forecasting models and has 36 records from the period of January 2021 to December 2023.

```
> # Print Date Range and Count of Records for Training Set
> cat("Training Data:\n")
Training Data:
> cat("Start Date:", start(train.ts), "\n")
Start Date: 2013 1
> cat("End Date:", end(train.ts), "\n")
End Date: 2020 12
> cat("Number of Records:", length(train.ts), "\n\n")
Number of Records: 96
```

```
> # Print Date Range and Count of Records for Validation Set
> cat("Validation Data:\n")
Validation Data:
> cat("Start Date:", start(valid.ts), "\n")
Start Date: 2021 1
> cat("End Date:", end(valid.ts), "\n")
End Date: 2023 12
> cat("Number of Records:", length(valid.ts), "\n")
Number of Records: 36
```

STEP 6 - APPLYING FORECASTING METHODS

ADVANCED EXPONENTIAL SMOOTHING: HOLT'S-WINTER METHOD

The next method is focusing the sales data using advanced smoothing techniques. The study employs the Holt-Winters method, serves as the cornerstone of our time series analysis and forecasting approach. This method extends basic exponential smoothing by incorporating trend and seasonality components into our models.

By utilizing three smoothing parameters - alpha (α), beta (β), and gamma (γ) - the Holt-Winters method captures both the underlying trend and seasonal patterns in our sales data. This allows us to generate forecasts that not only reflect recent observations but also adapt to changes in trend and account for recurring seasonal variations. In R, the Holt-Winter's model is considered using model = 'ZZZ' in the ets () function, that represents the additive or multiplicative error, trend, and seasonality.

Holt-Winter's for Training and Validation

The Holt-Winter's model with the automated selection of the model an their smoothing parameters for the training period is as shown as below:

```
ETS(M,A,M)

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

  Smoothing parameters:
    alpha = 0.1283
    beta  = 0.0109
    gamma = 2e-04

  Initial states:
    l = 10813.4495
    b = 22.3133
    s = 1.1701 1.028 1.0456 0.9983 1.052 0.9988
           1.1177 1.0654 0.9595 0.9674 0.8347 0.7625

  sigma:  0.0478

     AIC     AICc      BIC
1672.814 1680.661 1716.408
```

The model has ETS (M, A, M) that is multiplicative error, additive trend, and multiplicative seasonality. As the smoothing parameters should be between 0 and 1 here, the alpha parameter which represents high level of smoothing is very close to zero, this means that the historical data is having a great influence on the forecast. A smaller beta value assigns less weight to recent observations when updating the trend, resulting in a smoother trend line. Similar to beta, a smaller gamma value dampens the effect of seasonal fluctuations, resulting in a smoother seasonal component.

A sigma 0.0478 estimate was made, which aids in determining the degree of variability in the data. To assess how well the model fits the data, various metrics including AIC, AICc, and BIC values are considered. Lower the values better is the model.

Holt-Winter's model for the entire data set

The Holt-Winter's model with the automated selection of the model and their smoothing parameters for the entire data set is as shown as below:

```
ETS(M,A,M)

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

  Smoothing parameters:
    alpha = 0.1431
    beta  = 0.0291
    gamma = 1e-04

  Initial states:
    l = 10788.45
    b = 18.3647
    s = 1.1597 1.0393 1.0413 0.9999 1.05 0.9948
           1.1204 1.0538 0.9644 0.9827 0.8339 0.7598

  sigma:  0.0463

     AIC     AICc      BIC
2348.197 2353.565 2397.205
```

The model has ETS (M, A, M) that is multiplicative error, additive trend, and multiplicative seasonality for the entire data set. Even here it is observed that the smoothing parameters are within 0 and 1.

Forecast for the future data

	Point Forecast	Lo 0	Hi 0
Jan 2024	11663.66	11663.66	11663.66
Feb 2024	12759.49	12759.49	12759.49
Mar 2024	14986.13	14986.13	14986.13
Apr 2024	14659.17	14659.17	14659.17
May 2024	15965.15	15965.15	15965.15
Jun 2024	16918.45	16918.45	16918.45
Jul 2024	14971.69	14971.69	14971.69
Aug 2024	15750.12	15750.12	15750.12
Sep 2024	14947.87	14947.87	14947.87
Oct 2024	15513.62	15513.62	15513.62
Nov 2024	15431.82	15431.82	15431.82
Dec 2024	17162.01	17162.01	17162.01

Accuracy measures:

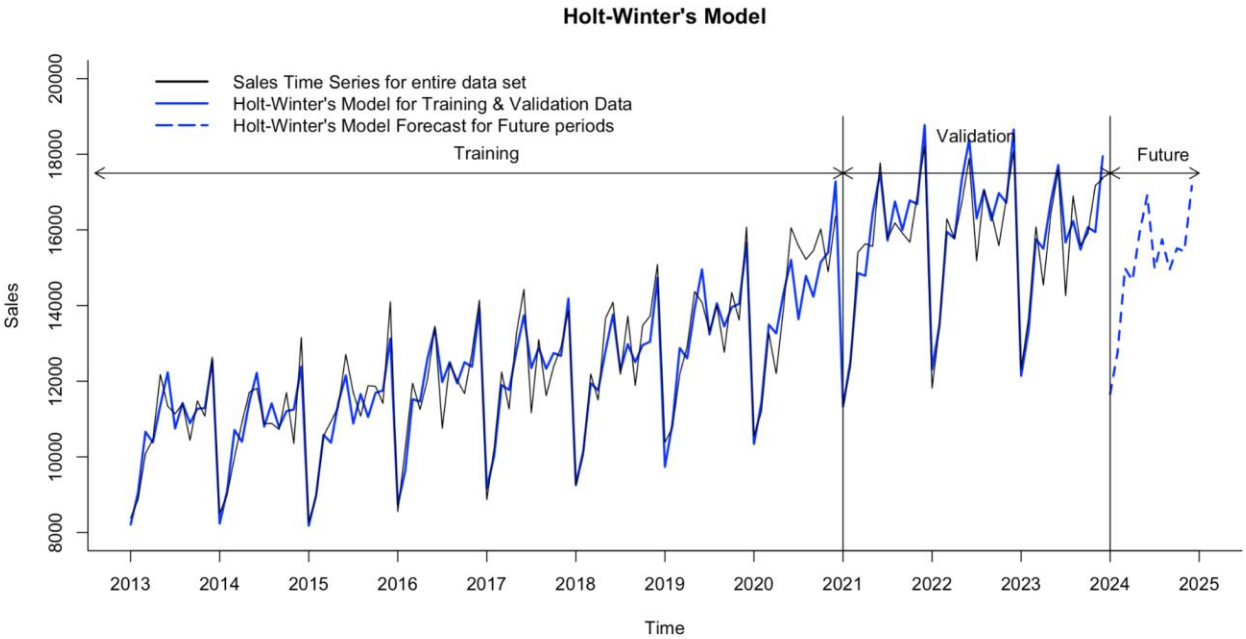
```
> round(accuracy(holtw.predict$mean, valid.ts), 3)
                ME     RMSE     MAE    MPE  MAPE  ACF1 Theil's U
Test set -512.116 1227.906 905.718 -3.224 5.697 0.498     0.576
> round(accuracy(holtw2.predict$fitted, sales.ts), 3)
               ME    RMSE     MAE    MPE MAPE   ACF1 Theil's U
Test set -19.471 580.914 457.899 -0.235 3.49 -0.078     0.302
```

Holt-Winter's	RMSE	MAPE
for validation set	1227.906	5.697
for entire data set	580.914	3.49

Comparing both the validation set with the entire data set, the entire data set has a substantially better MAPE (3.49%) and RMSE (508.914) for the entire set than the validation set respectively. The RMSE and MAPE values provide valuable insights into the accuracy and performance of the Holt-Winter's forecasting.

Overall, the evaluation results suggest that the Holt-Winters models exhibit reasonable predictive accuracy, with both RMSE and MAPE values indicating relatively low forecast errors. However, further analysis and fine-tuning of the models may be warranted to improve forecast accuracy and refine the modeling approach.

Holt-Winter's Plot



In the plotted graph depicting the Holt-Winter's model, we observe various elements illustrating the model's performance in forecasting sales data. The blue dashed line represents the forecasted sales values generated by the Holt-Winter's model, extending into future periods beyond the data's historical range. Overall, this plot offers a clear visualization of the Holt-Winter's model's ability to capture trends and seasonality in the sales data and its effectiveness in forecasting future sales figures.

REGRESSION – BASED MODELS

The next model used in this analysis is Regression – Based Model.

Regression models in time series use one or more independent variables like time, to predict the dependent variable values like sales over a period. Basically, it models the relationship between the dependent and the independent variables and uses this relationship to predict the future values of the dependent variable. This model is used because it is simple and since it considers both trend and seasonality it displays strong and relevant results. This model can be further improved with autoregressive components and trailing moving average for residuals. The model was first evaluated based on training and validation partitions, before running on the entire dataset. We have trained several regression models like the Regression model with linear trend, quadratic trend, model with seasonality, linear trend with seasonality and quadratic trend with seasonality forecasted future periods using them. But the Regression model with linear trend and seasonality proved to be the best for this dataset.

Time Series 2013-2023 (Training / Validation)

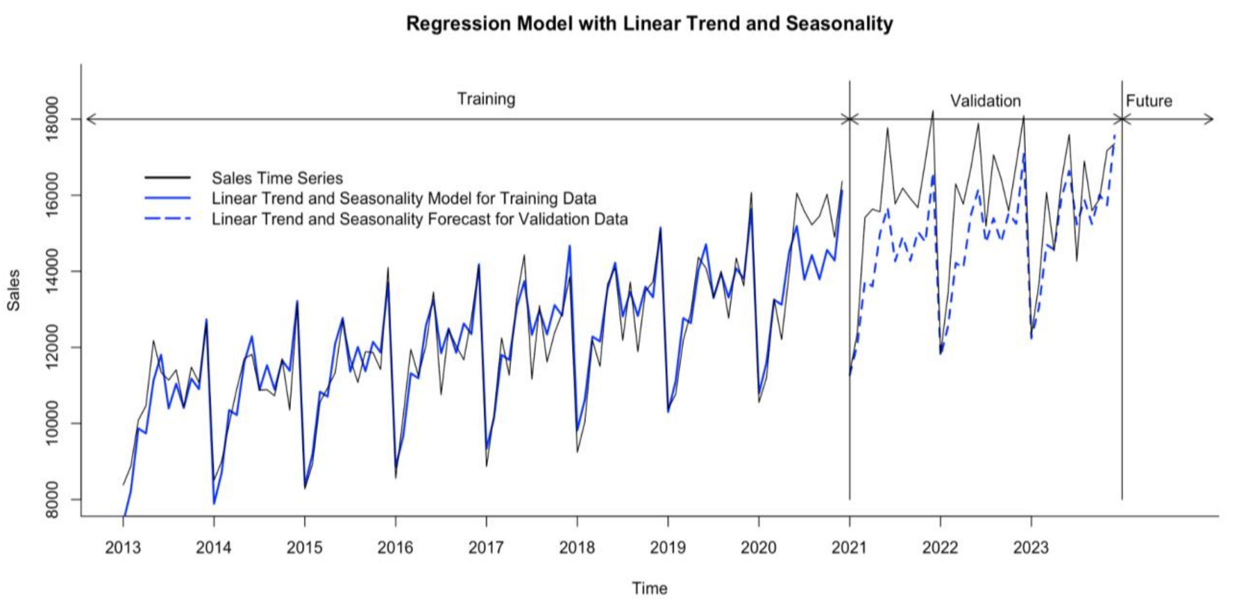
```
Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
     Min       1Q   Median       3Q      Max 
-1165.16  -452.85    22.02   337.26  1796.85 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7364.827    243.026  30.305  < 2e-16 ***
trend         40.277      2.326  17.320  < 2e-16 ***
season2      770.848    313.256   2.461   0.0159 *  
season3     2381.195    313.282   7.601 4.03e-11 ***
season4     2213.168    313.325   7.063 4.59e-10 ***
season5     3559.266    313.385  11.357  < 2e-16 ***
season6     4198.739    313.463  13.395  < 2e-16 ***
season7     2751.086    313.558   8.774 1.85e-13 ***
season8     3356.059    313.670  10.699  < 2e-16 ***
season9     2681.782    313.799   8.546 5.29e-13 ***
season10    3408.879    313.946  10.858  < 2e-16 ***
season11    3093.852    314.109   9.850 1.31e-15 ***
season12    4885.325    314.290  15.544  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 626.5 on 83 degrees of freedom
Multiple R-squared:  0.9041,	Adjusted R-squared:  0.8902 
F-statistic: 65.21 on 12 and 83 DF,  p-value: < 2.2e-16
```

After looking at the data series plot, it is clear that the data has a trend with seasonality. The summary above represents a regression model with linear trend with seasonality. The model has an R-squared value of 0.9041. Indicating 90.41% of variance is explained by the predictors. The adjusted R-square of the model is 89.02%. It is statistically significant since the F - statistic p-value is very low 2.2e-16 which is a lot lower than alpha of 5%. The intercept for the model is

7364.827. There are 1 trend predictor and 11 seasonal predictors (indicating monthly periods) which are dummy variables. Therefore, this model is a good fit to the data.

<u>Model Equation</u>: 7364.827 + 40.277t + 770 D2 + 2381.195 D3 + 2213.168 D4 + 3559.266 D5 + 4198.739 D6 + 2751.086 D7 + 3356.059 D8 + 2681.782 D9 + 3408.879 D10 + 3093.852 D11 + 4885.325 D12.

Plotting



The above plot is a line graph displaying a regression model for monthly sales of distilled alcoholic beverages with linear trend and seasonality, the model is fitting well by taking the data's trend and seasonality into consideration. There is a slight under prediction in the validation partition by the model. However, this model forecasts the sales for the unseen data (year 2024).

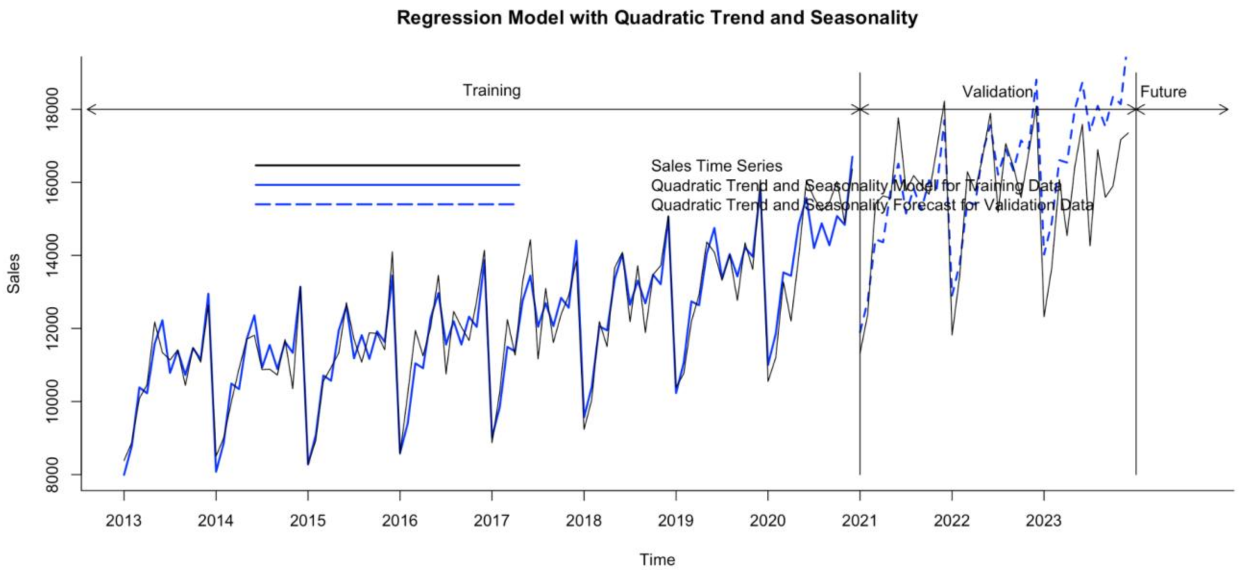
Regression model with quadratic trend and seasonality

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

Residuals:
     Min       1Q   Median       3Q      Max 
-1236.96  -350.04     8.11   356.83  1377.69 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.990e+03  2.516e+02  31.757  < 2e-16 ***
trend       1.554e+00  8.280e+00   0.188  0.85155    
I(trend^2)  3.992e-01  8.267e-02   4.829 6.28e-06 ***
season2     7.748e+02  2.781e+02   2.786  0.00662 ** 
season3     2.388e+03  2.781e+02   8.588 4.76e-13 ***
season4     2.223e+03  2.782e+02   7.991 7.28e-12 ***
season5     3.570e+03  2.782e+02  12.834  < 2e-16 ***
season6     4.211e+03  2.783e+02  15.131  < 2e-16 ***
season7     2.763e+03  2.784e+02   9.926 1.05e-15 ***
season8     3.367e+03  2.785e+02  12.092  < 2e-16 ***
season9     2.691e+03  2.786e+02   9.661 3.50e-15 ***
season10    3.416e+03  2.787e+02  12.257  < 2e-16 ***
season11    3.098e+03  2.788e+02  11.109  < 2e-16 ***
season12    4.885e+03  2.790e+02  17.510  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 556.2 on 82 degrees of freedom
Multiple R-squared:  0.9253,	Adjusted R-squared:  0.9135 
F-statistic: 78.17 on 13 and 82 DF,  p-value: < 2.2e-16
```

The summary above represents a regression model with quadratic trend with seasonality. The model has an R-squared value of 0.9253. Indicating 92.53% of variance is explained by the predictors. The adjusted R-square of the model is 91.35%. It is statistically significant since the F - statistic p-value is very low 2.2e-16 which is a lot lower than alpha of 5%. The intercept for the model is 7.990e+03. There are 2 trend predictors and 11 seasonal predictors (indicating monthly periods) which are dummy variables. Therefore, this model is a good fit to the data.

Model Equation: **7.990e+03 + 1.554e+00 t + 3.992e-01 t^2 + 7.748e+02 D2 + 2.388e+03 D3 + 2.223e+03 D4 + 3.570e+03 D5 + 4.211e+03 D6 + 2.736e+03 D7 + 3.367e+03 D8 + 2.691e+03 D9 + 3.416e+03 D10 + 3.098e+03 D11 + 4.885e+03 D12.**

Plotting



The above plot is a line graph displaying a regression model for monthly sales of distilled alcoholic beverages with quadratic trend and seasonality, the model is fitting well by taking the data's trend and seasonality into consideration. There is a slight over prediction in the validation partition by the model. However, this model forecasts the sales for the unseen data (year 2024).

Regression model	RMSE	MAPE	ACF1
Linear Trend with Seasonality	1238.637	6.35	0.002
Quadratic Trend with Seasonality	1236.43	6.429	0.488

After considering RMSE and MAPE values, Linear Trend with Seasonality has a slightly low MAPE of 6.35% compared to Quadratic Trend with Seasonality of 6.429%. While RMSE of Linear Trend with Seasonality is slightly high with 1238.637 compared to Quadratic Trend with

Seasonality value of 1236.43. Since MAPE is the superior metrics, Linear trend with seasonality would be the best fitted model to forecast the future period sales.

<u>Regression model with Linear Trend and Seasonality on Entire Dataset</u>

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

Residuals:
     Min       1Q   Median       3Q      Max
-1539.32  -527.28   -41.75   433.54  1514.15

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 6890.699    242.364  28.431  < 2e-16 ***
trend         48.346      1.672  28.908  < 2e-16 ***
season2      907.017    310.914   2.917  0.00422 **
season3     2813.217    310.928   9.048 3.35e-15 ***
season4     2504.870    310.950   8.056 6.87e-13 ***
season5     3714.797    310.982  11.945  < 2e-16 ***
season6     4575.542    311.022  14.711  < 2e-16 ***
season7     2773.650    311.072   8.916 6.83e-15 ***
season8     3641.849    311.130  11.705  < 2e-16 ***
season9     2932.230    311.198   9.422 4.37e-16 ***
season10    3371.248    311.274  10.830  < 2e-16 ***
season11    3462.811    311.359  11.122  < 2e-16 ***
season12    4998.283    311.454  16.048  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 729.1 on 119 degrees of freedom
Multiple R-squared:   0.92,     Adjusted R-squared:  0.9119
F-statistic:   114 on 12 and 119 DF,  p-value: < 2.2e-16
```

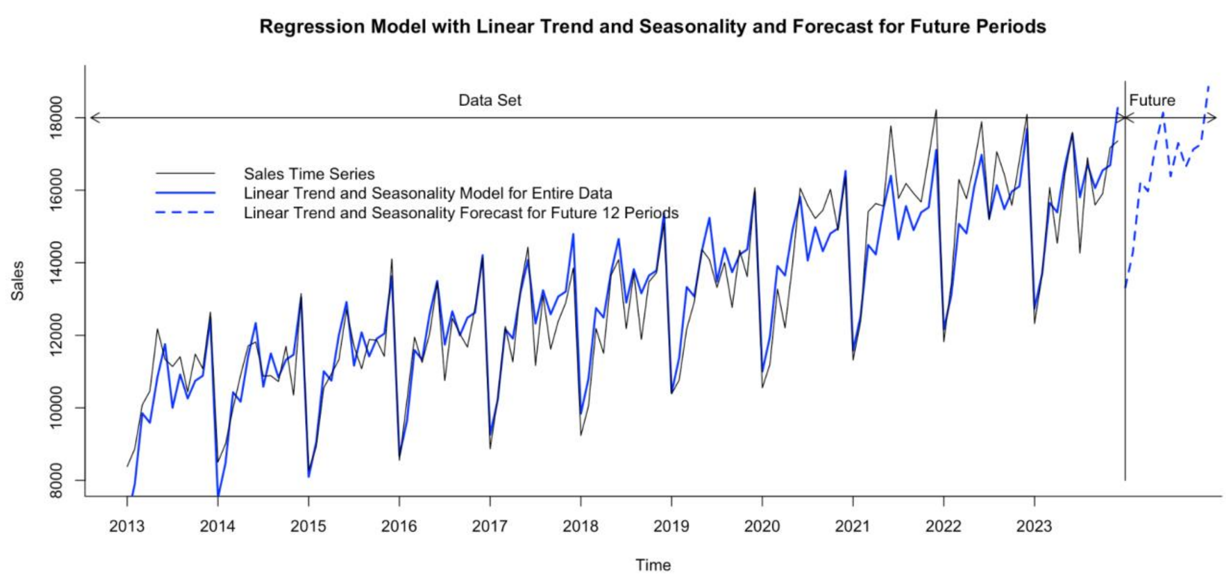
For forecasting the future period sales, the training and validation partitions must be recombined into an entire (time series) dataset. The above summary shows a regression model for the entire dataset. The model has an R-squared value of 0.92. Indicating 92% of variance is explained by the predictors. The adjusted R-square of the model is 91.19%. It is statistically significant since the F - statistic p-value is very low 2.2e-16 which is a lot lower than alpha of 5%. The intercept for the model is 6890.699. There are 1 trend predictor and 11 seasonal predictors (indicating monthly periods) which are dummy variables. Therefore, this model is a good fit to the data

<u>Model Equation</u>: 6890.699 + 48.346 t + 907.017 D2 + 2813.217 D3 + 2504.870 D4 + 3714.797 D5 + 4575.542 D6 + 2773.650 D7 + 3641.849 D8 + 2932.230 D9 + 3371.248 D10 + 3462.811 D11 + 4998.283 D12.

Forecasting for future 12 periods

	Point Forecast	Lo 0	Hi 0
Jan 2024	13320.75	13320.75	13320.75
Feb 2024	14276.11	14276.11	14276.11
Mar 2024	16230.65	16230.65	16230.65
Apr 2024	15970.65	15970.65	15970.65
May 2024	17228.93	17228.93	17228.93
Jun 2024	18138.02	18138.02	18138.02
Jul 2024	16384.47	16384.47	16384.47
Aug 2024	17301.02	17301.02	17301.02
Sep 2024	16639.75	16639.75	16639.75
Oct 2024	17127.11	17127.11	17127.11
Nov 2024	17267.02	17267.02	17267.02
Dec 2024	18850.84	18850.84	18850.84

Plotting using entire dataset



The above graph is plotting the linear trend and seasonality model for the entire dataset and future 12 periods. It is clear that there is little over prediction in the forecasting period. The data has an upward trend which indicates that the sales are increasing on an average over the period of time. There is also seasonality in the data. This forecasted trend for 2024 shows the increase in sales.

Regression model with Linear Trend and Seasonality	**RMSE**	**MAPE**	**ACF1**
Training & Validation data	3685.344	23.001	-0.213
Entire Data Set	692.313	4.358	0.308

From the performance measures, the RMSE and MAPE values are 3685.344 and 23.001% for training and validation periods respectively. The autocorrelation is -0.213 displaying negative autocorrelation. The RMSE, MAPE values for the entire dataset are 692.313 and 4.358% respectively. The autocorrelation is 0.308 displaying positive autocorrelation.

The performance is better when the entire dataset was considered, compared to individual training and validation periods. The positive autocorrelation for the entire dataset indicates there are underlying trends and patterns in the data.

Regression models	**RMSE**	**MAPE**	**ACF1**
Linear Trend with Seasonality	555.545	4.358	0.308
Naive	1974.397	12.172	-0.4

SNaive	915.749	5.086	0.351

Comparing MAPE, RMSE values of Linear Trend and Seasonality with base models like naive and snaive indicates that Linear Trend with Seasonality is the best model to forecast the sales in 2024 using this data.

TIME SERIES PREDICTABILITY

Predictability Test for US Beverage Sales data

Approach 1:

Here we are fitting the AR (1) model to the time series and test the hypothesis that the slope coefficient Beta.

AR (1) model for historical data

```
> sales.ar1<- Arima(sales.ts, order = c(1,0,0))
> summary(sales.ar1)
Series: sales.ts
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1        mean
      0.6810  13072.3260
s.e.  0.0656    488.6903

sigma^2 = 3361613:  log likelihood = -1178.45
AIC=2362.89   AICc=2363.08   BIC=2371.54

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE     MASE       ACF1
Training set 34.1703 1819.527 1384.127 -1.837669 11.22567 1.999871 -0.1954321
>
```

ARIMA (1, 0, 0) is an autoregressive (AR) model with an order 1, along with no differencing, and no moving average.

Model equation: yt = 13072.3260 + 0.6810 Yt-1

The intercept of the model is 13072.3260. The coefficient of the ar1 (Yt-1) variable, Beta1 = 0.6810. The latter parameter could be used for hypothesis testing on the value of AR (1) regression coefficient.

Z-test to test the null hypothesis

Applying z-test to test the null hypothesis that the beta coefficient of AR (1) is equal to 1.

```
> # Apply z-test to test the null hypothesis that beta
> # coefficient of AR(1) is equal to 1.
> ar1 <- 0.6810
> s.e. <- 0.0656
> null_mean <- 1
> alpha <- 0.05
> z.stat <- (ar1-null_mean)/s.e.
> z.stat
[1] -4.862805
> p.value <- pnorm(z.stat)
> p.value
[1] 5.786696e-07
> if (p.value<alpha) {
+   "Reject null hypothesis"
+ } else {
+   "Accept null hypothesis"
+ }
[1] "Reject null hypothesis"
```

Hypothesis Testing: Z- Test

Null hypothesis Ho: Beta1 = 1

Alternative hypothesis H1: Beta1! = 1

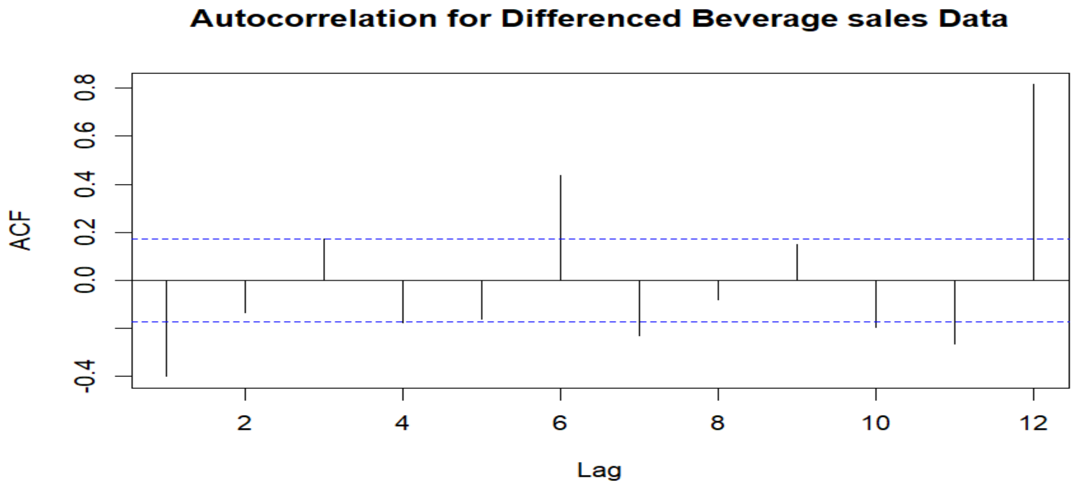
z-stat = (Beta1 - 1)/s.e. = (0.6810 - 1)/0.0656 = -4.8628, p-value for z-stat = 5.786696e-07

As the p-value is less than 0.05, we cannot accept the null hypothesis. Thus, the time series data for Beverage sales (sales.ts) is predictable.

Approach 2:

Applying ACF function for differenced US Beverage sales data.

Autocorrelation for First Differenced Beverage Sales Data



On Examining the ACF values, the differenced series seems to be exceeding the significant horizontal threshold levels at lag 1,6,7,10,11 and 12. Also there seems to be greater seasonality expressed at lag 12. So, we could infer by using the first differencing that the time series is Predictable and not a random walk.

Two-level forecast with regression model and AR model for residuals

Develop a regression model with linear trend and seasonality – Autocorrelation for Residuals

Regression model with linear trend and seasonality for the training data set and forecasting

for the validation period.

```
> # See summary of linear trend equation and associated parameters.
> summary(train.lin.season)

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

Residuals:
     Min       1Q   Median       3Q      Max 
-1165.16  -452.85    22.02   337.26  1796.85 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7364.827    243.026  30.305  < 2e-16 ***
trend         40.277      2.326  17.320  < 2e-16 ***
season2      770.848    313.256   2.461   0.0159 *  
season3     2381.195    313.282   7.601 4.03e-11 ***
season4     2213.168    313.325   7.063 4.59e-10 ***
season5     3559.266    313.385  11.357  < 2e-16 ***
season6     4198.739    313.463  13.395  < 2e-16 ***
season7     2751.086    313.558   8.774 1.85e-13 ***
season8     3356.059    313.670  10.699  < 2e-16 ***
season9     2681.782    313.799   8.546 5.29e-13 ***
season10    3408.879    313.946  10.858  < 2e-16 ***
season11    3093.852    314.109   9.850 1.31e-15 ***
season12    4885.325    314.290  15.544  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 626.5 on 83 degrees of freedom
Multiple R-squared:  0.9041,	Adjusted R-squared:  0.8902 
F-statistic: 65.21 on 12 and 83 DF,  p-value: < 2.2e-16
```

The residuals exhibit a range of values, which indicates the variability in the data that is beyond model explanation. Here is the training part. The model explains around 90.41 % of the variance in the training data, which is indicated by the multiple R-squared value. The adjusted R-squared refers to the number of predictors and suggests a slightly lower explanatory power of 89.02%. The F-statistic is highly significant, supporting the overall significance of the model. The p-values of all the components are statistically significant.

Auto Regressive (AR (1)) models for Regression Residuals

ARIMA (1, 0, 0) is an autoregressive (AR) model with order 1, with no differencing, and no moving average

```
> res.ar1 <- Arima(train.lin.season$residuals, order = c(1,0,0))
> summary(res.ar1)
Series: train.lin.season$residuals
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1     mean
      0.2012   3.2008
s.e.  0.1010  72.7604

sigma^2 = 332691:  log likelihood = -745.55
AIC=1497.09   AICc=1497.35   BIC=1504.79

Training set error measures:
                    ME     RMSE      MAE      MPE     MAPE      MASE        ACF1
Training set -2.253249 570.7538 458.8712 82.07834 140.3027 0.8391925 -0.04726052
```

AR (1) model equation: yt = 3.2008 + 0.2012 yt-1

The coefficient of the ar1 (Yt-1) variable, β1 = 0.2012, and standard error of estimate, s.e. Is 0.1010. These two parameters would be used for hypothesis testing about the value of the AR (1) regression coefficient.

Autocorrelation for Sales Training Residuals:

A correlogram must be created which shows the autocorrelation exited between the regression models and their corresponding residuals.

The above chart infers that the significant autocorrelation of residuals in lags 2,3,9,10, which means that these autocorrelations between residuals are not infused into the regression model.

Therefore, modeling of these residual autocorrelations with an AR model and thereby developing a two-level model might improve the overall forecast. Hence, it is a good idea to add AR models for residuals in the forecast.

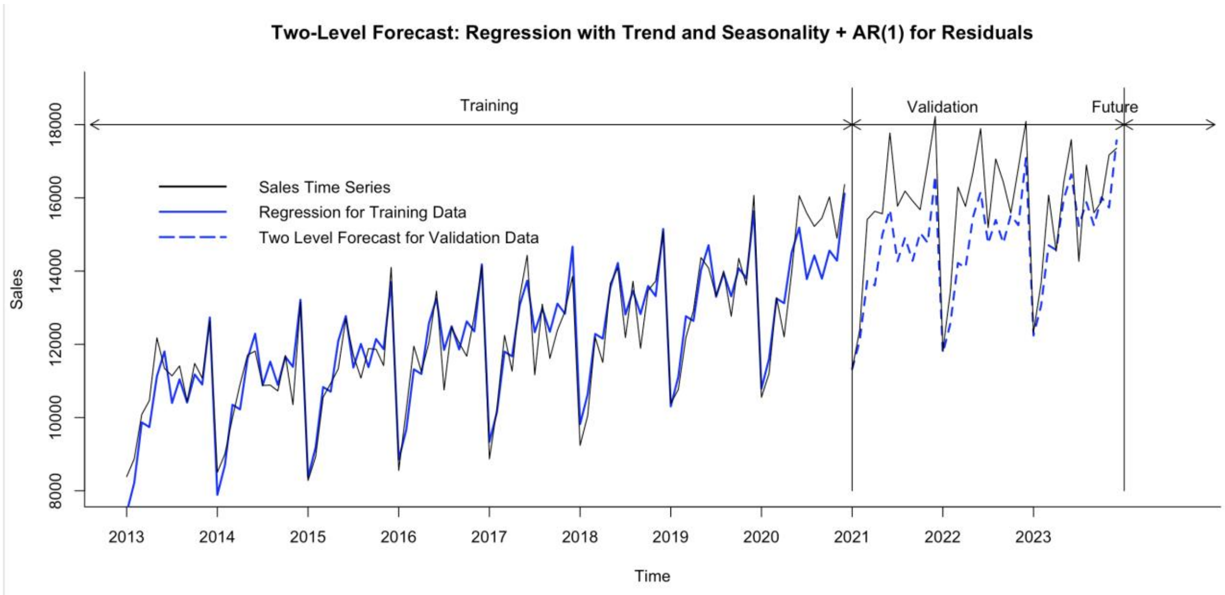
<u>Two-level model for linear trend and seasonality and AR (1) model for residuals</u>

The below table explains how the Beverage sales data and forecasts in the validation partition (Regression forecast in the validation period (Reg.Forecast), AR (1) model's forecast of the regression residuals in the validation period (AR (1) Forecast), and combined forecast (Combined. Forecast) as a sum of the regression and AR (1) models forecasts are being demonstrated.

```
> names(valid.df) <- c("Sales", "Reg.Forecast",
+                      "AR(1)Forecast", "Combined.Forecast")
> valid.df
   Sales Reg.Forecast AR(1)Forecast Combined.Forecast
1  11318     11271.72        52.296          11324.02
2  12384     12082.85        13.078          12095.93
3  15410     13733.47         5.188          13738.66
4  15633     13605.72         3.601          13609.32
5  15564     14992.10         3.281          14995.38
6  17770     15671.85         3.217          15675.07
7  15772     14264.47         3.204          14267.68
8  16185     14909.72         3.201          14912.92
9  15914     14275.72         3.201          14278.92
10 15676     15043.10         3.201          15046.30
11 16909     14768.35         3.201          14771.55
12 18222     16600.10         3.201          16603.30
13 11822     11755.05         3.201          11758.25
14 13439     12566.18         3.201          12569.38
15 16295     14216.80         3.201          14220.00
16 15766     14089.05         3.201          14092.25
17 16708     15475.43         3.201          15478.63
18 17891     16155.18         3.201          16158.38
19 15186     14747.80         3.201          14751.00
20 17062     15393.05         3.201          15396.25
21 16434     14759.05         3.201          14762.25
22 15584     15526.43         3.201          15529.63
23 16818     15251.68         3.201          15254.88
24 18090     17083.43         3.201          17086.63
25 12324     12238.38         3.201          12241.58
26 13661     13049.50         3.201          13052.70
27 16074     14700.13         3.201          14703.33
28 14542     14572.38         3.201          14575.58
29 16419     15958.75         3.201          15961.95
30 17592     16638.50         3.201          16641.70
31 14265     15231.13         3.201          15234.33
32 16896     15876.38         3.201          15879.58
33 15593     15242.38         3.201          15245.58
34 15903     16009.75         3.201          16012.95
35 17173     15735.00         3.201          15738.20
36 17356     17566.75         3.201          17569.95
```

Two level forecasting plots

Below is the plot of the Two-level model for linear trend and seasonality and AR (1) model and residuals for residuals for training and validation data. We could infer that the model is fitting well into data, where the trend and seasonality components are captured.



Accuracies

```
> round(accuracy(valid.two.level.pred, valid.ts), 3)
          ME     RMSE      MAE   MPE  MAPE  ACF1 Theil's U
Test set 956 1235.975 1030.046 5.822 6.322 0.003     0.614

> round(accuracy(train.lin.season.pred$mean, valid.ts), 3)
              ME     RMSE      MAE   MPE MAPE  ACF1 Theil's U
Test set 960.908 1238.637 1033.908 5.857 6.35 0.002     0.615
```

```
> cat("Two-level model")
Two-level model
> round(accuracy(lin.season$fitted + residual.ar1$fitted, sales.ts), 3)
              ME    RMSE     MAE    MPE MAPE   ACF1 Theil's U
Test set -4.076 657.066 538.336 -0.231 4.22 -0.097     0.345
> cat("Linear trend and seasonality")
Linear trend and seasonality
> round(accuracy(lin.season$fitted, sales.ts), 3)
         ME    RMSE     MAE    MPE  MAPE  ACF1 Theil's U
Test set  0 692.313 555.545 -0.205 4.358 0.308     0.364
> cat("Seasonal naive forecast")
Seasonal naive forecast
> round(accuracy((snaive(sales.ts))$fitted, sales.ts), 3)
              ME    RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 485.808 915.749 692.108 3.503 5.086 0.351     0.451
```

Two-Level Forecast	**MAPE**	**RMSE**	**ACF1**
Training and Validation	6.322	1235.975	0.003
Entire Data Set	4.22	657.066	-0.097

On considering the accuracy measures, the RMSE and MAPE values for Two-Level Forecast model for the entire dataset seems to be 657.066 and 4.22. From this we could infer the magnitude of the model. The ACF1 value is -0.097 that gives the degree of autocorrelation in the residuals.

AUTO REGRESSIVE INTEGRATED MOVING AVERAGE MODEL (ARIMA MODEL)

ARIMA (Auto Regressive Integrated Moving Average) model is a famous time series forecasting method which generally is a combination of auto regression, differencing, and moving averages. This method could be used for forecasting data containing various components such as level, trend, and seasonality. Seasonal ARIMA (Seasonal Auto Regressive Integrated Moving Average), is an augmented extension of the basic ARIMA model which is designed to deal with time series data exhibiting seasonal patterns. In general, Seasonality refers to the variations in data with regard to periods occurring at regular intervals. The season considered here is monthly.

Seasonal Auto Regressive Integrated Moving Average Model (Seasonal ARIMA Model)

ARIMA for Training and Validation data

```
> summary(train.arima.seas)
Series: train.ts
ARIMA(2,1,2)(1,1,2)[12]

Coefficients:
          ar1      ar2     ma1      ma2    sar1     sma1     sma2
      -1.0394  -0.2875  0.3884  -0.6116  0.2793  -0.7030  -0.2921
s.e.   0.1834   0.1929  0.2469   0.2121  0.3437   1.2025   0.4640

sigma^2 = 301653:  log likelihood = -648.19
AIC=1312.37   AICc=1314.32   BIC=1331.72

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 47.32937 488.6805 340.7445 0.1990541 2.731159 0.5285982 -0.0148474
```

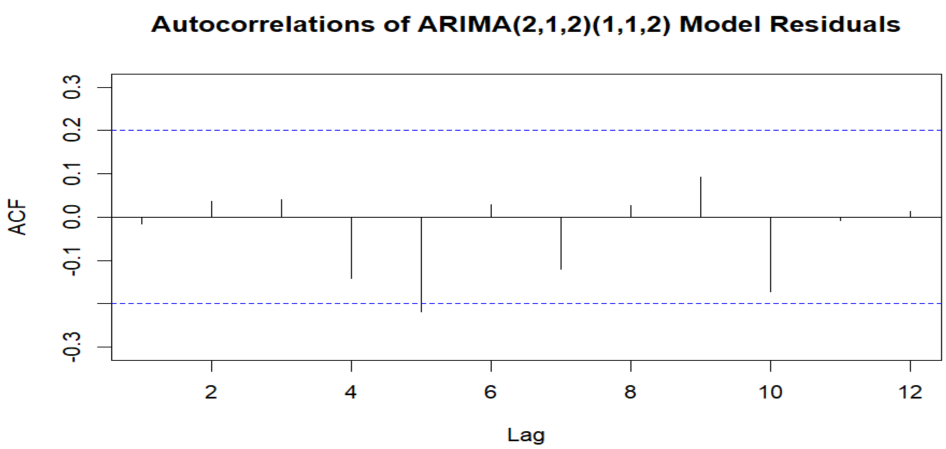
This is a seasonal ARIMA model, ARIMA (p, d, q)(P, D, Q)[m], where:

- p = 2, order 2 autoregressive model AR (2)
- d = 1, first differencing
- q = 2, order 2 moving average MA (2) for error lags
- P = 1, order 1 autoregressive model AR (1) for the seasonal part
- D = 1, first differencing for the seasonal part
- Q = 2, order 2 moving average MA (2) for the seasonal error lags
- m = 12, for yearly seasonality.

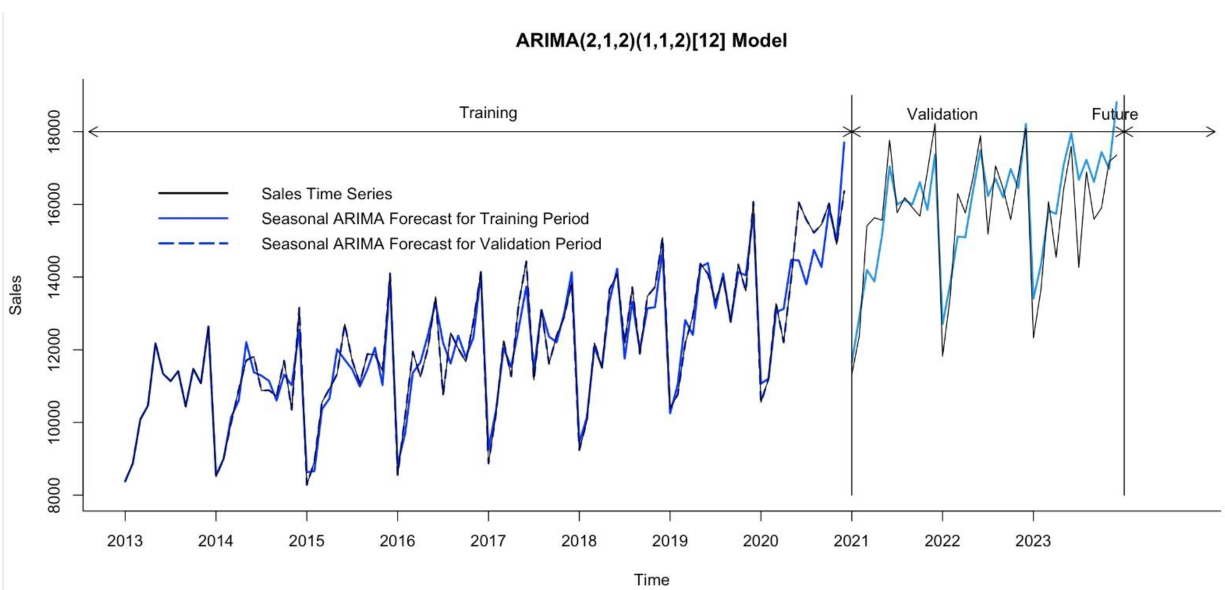
Model Equation: $y_t - y_{t-1} = -1.0394 (y_{t-1} - y_{t-2}) - 0.2875 (y_{t-2} - y_{t-3}) + 0.3884\ \varepsilon_{t-1} - 0.6116\ \varepsilon_{t-2} + 0.2793 (y_{t-1} - y_{t-13}) - 0.7030\ \rho_{t-1} - 0.2921\ \rho_{t-2}$

The variance of the error term (sigma^2) is 301653, and the log likelihood is -648.19. The model is statistically significant. The model selection criteria AIC, AICc, and BIC are 1312.37, 1314.32, 1331.72 respectively.

Autocorrelation for model residuals



Plot



The plot of the Seasonal ARIMA model for training and validation data sets infers that the model seems to be fitting well into data, by capturing the trend components and thereby it is said to be statistically significant. Thereby, we can proceed for further forecasting and for the entire dataset.

<u>Seasonal ARIMA for Entire DataSet</u>

```
> arima.seas <- Arima(sales.ts, order = c(2,1,2),
+                     seasonal = c(1,1,2))
>      summary(arima.seas)
Series: sales.ts
ARIMA(2,1,2)(1,1,2)[12]

Coefficients:
          ar1      ar2     ma1      ma2    sar1     sma1     sma2
      -0.9720  -0.2884  0.3011  -0.6560  0.0857  -0.5638  -0.2808
s.e.   0.1279   0.1383  0.1182   0.1293  0.3252   0.3642   0.2161

sigma^2 = 332533:  log likelihood = -929.48
AIC=1874.95   AICc=1876.26   BIC=1897.19

Training set error measures:
                   ME     RMSE      MAE        MPE     MAPE      MASE       ACF1
Training set 23.96745 531.1776 391.8125 0.03607532 2.949553 0.5661144 0.01138115
```

Model Equation: yt - yt-1 = -0.9720 (yt-1)- 0.2884 (yt-2 -yt-3) + 0.3011 εt-1 - 0.6560 εt-2 + 0.0857(yt-1 -yt-13) - 0.5638ρt-1 -0.2808 ρt-1

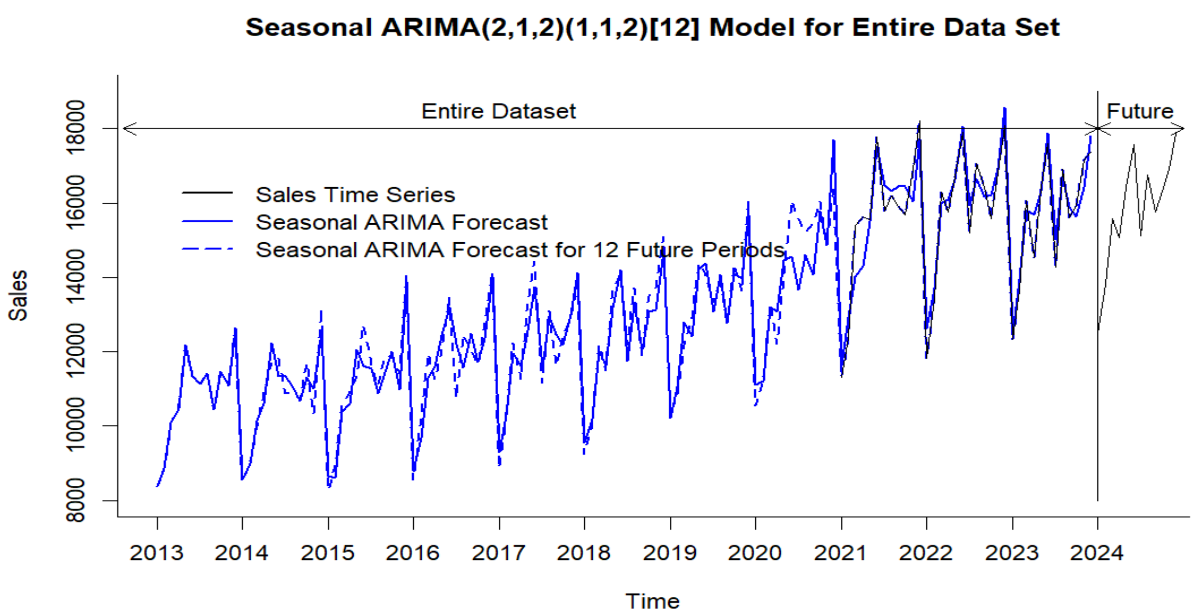
The variance of the error term (sigma^2) is 332533, and the log likelihood is -929.48. The model selection criteria AIC, AICc, and BIC are 1874.95, 1876.26, 1897.19 respectively.

Forecast

```
> arima.seas.pred <- forecast(arima.seas, h = 12, level = 0)
>       arima.seas.pred
         Point Forecast     Lo 0     Hi 0
Jan 2024       12501.35 12501.35 12501.35
Feb 2024       13821.37 13821.37 13821.37
Mar 2024       15589.47 15589.47 15589.47
Apr 2024       15069.88 15069.88 15069.88
May 2024       16422.05 16422.05 16422.05
Jun 2024       17571.11 17571.11 17571.11
Jul 2024       15129.54 15129.54 15129.54
Aug 2024       16755.41 16755.41 16755.41
Sep 2024       15766.69 15766.69 15766.69
Oct 2024       16314.73 16314.73 16314.73
Nov 2024       16918.29 16918.29 16918.29
Dec 2024       17868.44 17868.44 17868.44
```

Here, we can see that many lags of autocorrelation are insignificant for all but for lag 4,9 the autocorrelation is significant, so, we can say there is no room for improvement.

<u>Future forecast of the year 2024-2025 by ARIMA model</u>



From the plot of the Seasonal ARIMA model for the entire dataset, we could infer that the model seems to fit well into data, by capturing the components and thereby it is said to be statistically significant.

AUTO ARIMA for Training and Validation data:

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

Coefficients:
          ma1
      -0.8343
s.e.   0.0780

sigma^2 = 430585:  log likelihood = -656.24
AIC=1316.48   AICc=1316.63   BIC=1321.31

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 64.57657 606.4582 447.4124 0.3412448 3.564934 0.6940726 -0.0619345
```

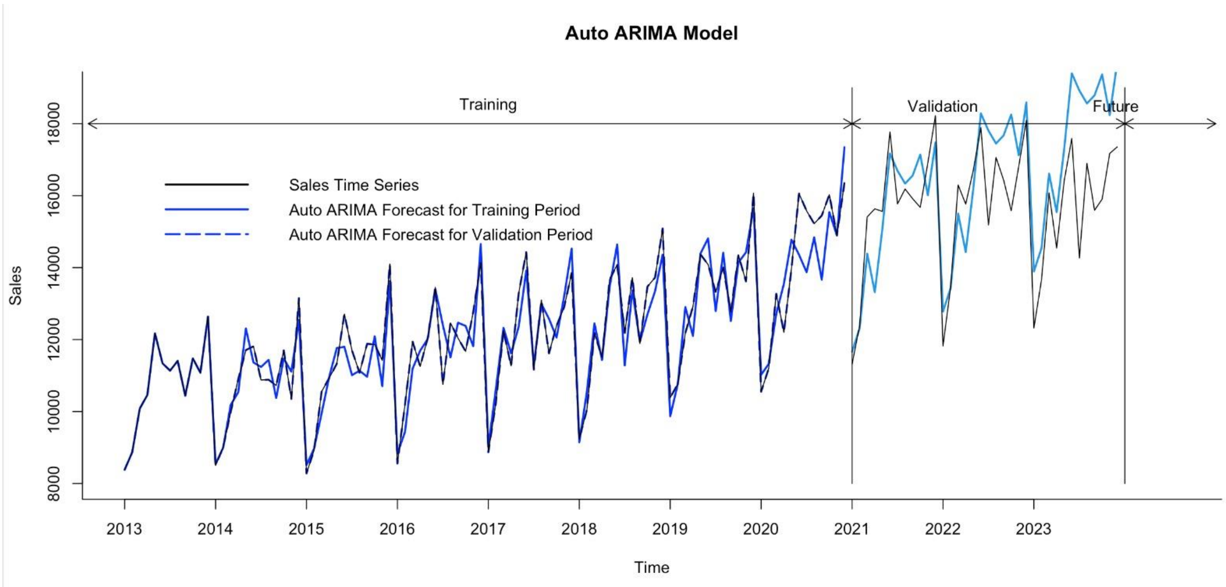
This is a seasonal Auto ARIMA model, ARIMA(p, d, q)(P, D, Q)[m], where:

- p = 0, zero order autoregressive model
- d = 1, first differencing
- q = 1, order 1 moving average MA (1) for error lags
- P = 0, order 1 autoregressive model AR (1) for the seasonal part
- D = 1, first differencing for the seasonal part
- Q = 0, order zero moving average MA
- m = 12, for yearly seasonality.

Model equation: yt-yt-1 = -0.8343 εt-1

The model is statistically significant, as The variance of the error term (sigma^2) is 430585, and the log likelihood is -656.24. The model selection criteria AIC, AICc, and BIC are 1316.48, 1316.63, 1321.31 respectively.

Plot



The plot of the Auto ARIMA model for training and validation data sets infers that the model seems to be fitting well into data, by capturing the trend components and thereby it is said to be statistically significant. Thereby, we can proceed for further forecasting and for the entire dataset.

Auto ARIMA Model for Entire Data Set

```
> auto.arima <- auto.arima(sales.ts)
>      summary(auto.arima)
Series: sales.ts
ARIMA(3,0,1)(2,1,1)[12] with drift

Coefficients:
          ar1     ar2     ar3     ma1    sar1     sar2     sma1    drift
      -0.1702  0.2314  0.4624  0.4447  0.3863  -0.2264  -0.8636  45.2689
s.e.   0.1909  0.1045  0.0922  0.2201  0.1420   0.1080   0.1912   4.6387

sigma^2 = 329473:  log likelihood = -935.91
AIC=1889.83   AICc=1891.46   BIC=1914.91

Training set error measures:
                    ME     RMSE      MAE        MPE     MAPE      MASE
Training set -21.54363 528.7278 392.1741 -0.4458712 2.984804 0.5666368
                    ACF1
Training set 0.002001979
```

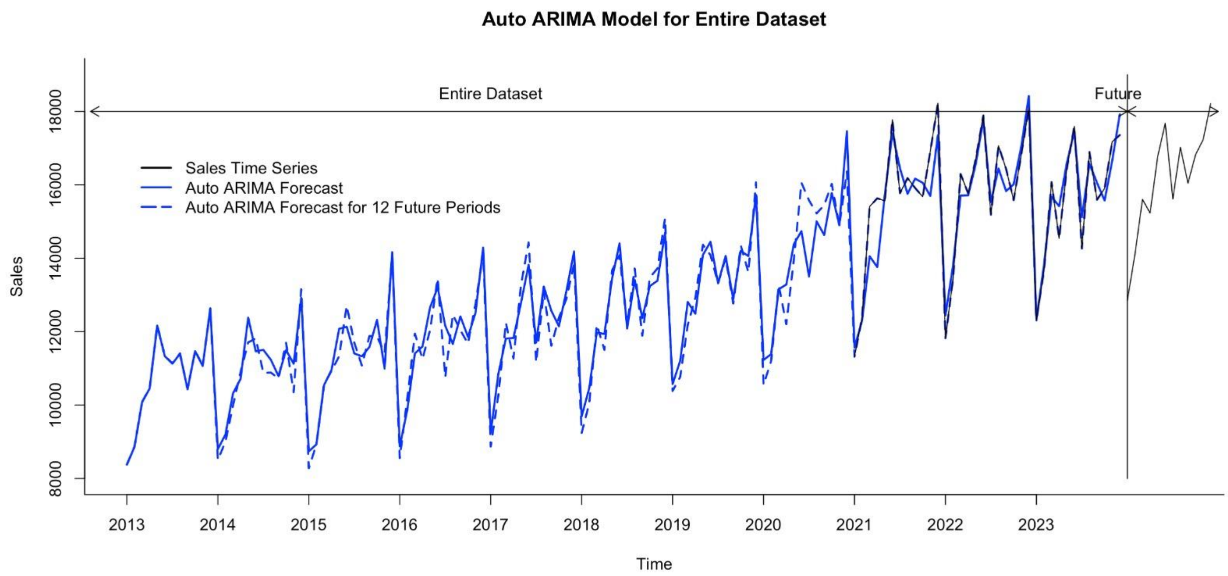
Model equation: yt - yt-1 = - 0.1702yt-1+ 0.2314 yt-2 +0.4624 yt-3 + 0.4447 εt-1 + 0.3863 (yt-1 - yt-13) - 0.2264 (yt-2 - yt-14) -0.8636 ρt-1

The variance of the error term (sigma^2) is 329473, and the log likelihood is -935.91. The model selection criteria AIC, AICc, and BIC are 1889.83, 1891.46, 1914.91 respectively. Here the Arima Model comes with a drift parameter, which means that this model has an intercept.

Future Forecast for Auto Arima Model

```
>      auto.arima.pred
         Point Forecast     Lo 0     Hi 0
Jan 2024       12849.29 12849.29 12849.29
Feb 2024       14104.91 14104.91 14104.91
Mar 2024       15604.26 15604.26 15604.26
Apr 2024       15231.02 15231.02 15231.02
May 2024       16767.45 16767.45 16767.45
Jun 2024       17677.91 17677.91 17677.91
Jul 2024       15617.14 15617.14 15617.14
Aug 2024       17018.55 17018.55 17018.55
Sep 2024       16045.58 16045.58 16045.58
Oct 2024       16822.01 16822.01 16822.01
Nov 2024       17229.89 17229.89 17229.89
Dec 2024       18213.91 18213.91 18213.91
```

Future forecast by Auto ARIMA model



The plot of the Auto ARIMA for entire data sets seems to fit well into data, thus capturing the trend components.

Accuracy and Performance Measures

For Validation Period Forecast

```
>      round(accuracy(train.arima.seas.pred$mean, valid.ts), 3)
              ME     RMSE      MAE    MPE  MAPE  ACF1 Theil's U
Test set -182.087 913.597 743.535 -1.441 4.858 0.167     0.437
>      round(accuracy(train.auto.arima.pred$mean, valid.ts), 3)
              ME     RMSE      MAE    MPE  MAPE  ACF1 Theil's U
Test set -747.349 1612.422 1235.558 -4.933 7.955 0.507     0.744
```

On considering the MAPE and RMSE accuracy values, the best model is the Seasonal ARIMA model, ARIMA (2,1,2)(1,1,2) with a measure of m=12, has the lower values of MAPE 4.858 and RMSE of 913.597 in comparison with the corresponding measures of other models.

For Entire Dataset:

```
>  # Use accuracy() function to identify common accuracy measures for:
>      # (1) Seasonal ARIMA (2,1,2)(1,1,2) Model,
>      # (2) Auto ARIMA Model,
>      # (3) Seasonal naive forecast, and
>      # (4) Naive forecast.
>      round(accuracy(arima.seas.pred$fitted, sales.ts), 3)
             ME     RMSE     MAE    MPE MAPE  ACF1 Theil's U
Test set 23.967 531.178 391.812 0.036 2.95 0.011     0.275
>      round(accuracy(auto.arima.pred$fitted, sales.ts), 3)
              ME     RMSE     MAE     MPE  MAPE  ACF1 Theil's U
Test set -21.544 528.728 392.174 -0.446 2.985 0.002     0.272
>      round(accuracy((snaive(sales.ts))$fitted, sales.ts), 3)
              ME     RMSE     MAE    MPE  MAPE  ACF1 Theil's U
Test set 485.808 915.749 692.108 3.503 5.086 0.351     0.451
>      round(accuracy((naive(sales.ts))$fitted, sales.ts), 3)
             ME      RMSE     MAE     MPE   MAPE ACF1 Theil's U
Test set 68.489 1974.397 1482.55 -0.761 12.172 -0.4         1
```

On considering the MAPE and RMSE accuracy values for the entire dataset, the best model is the Seasonal ARIMA model, ARIMA (2,1,2) (1,1,2) with a measure of m=12, has the lower values of MAPE 2.95 and RMSE of 531.178 in comparison with the corresponding measures of other models. However, to make comparison for the entire data set, we would consider the auto ARIMA model as well.

ARIMA Models	**MAPE**	**RMSE**	**ACF1**
Seasonal ARIMA Model	2.95	531.178	0.011
Auto Arima Model	**2.985**	**528.728**	**0.002**

COMPARING MODEL PERFORMANCE

```
> #Comparing Model Perfomance usinf accuracy()
> round(accuracy(holtw2.predict$fitted, sales.ts), 3)
              ME    RMSE     MAE    MPE MAPE   ACF1 Theil's U
Test set -19.471 580.914 457.899 -0.235 3.49 -0.078     0.302
> round(accuracy(lin.season.pred$fitted, sales.ts),3)
         ME    RMSE     MAE    MPE  MAPE  ACF1 Theil's U
Test set  0 692.313 555.545 -0.205 4.358 0.308     0.364
> round(accuracy(arima.seas.pred$fitted, sales.ts), 3)
             ME    RMSE     MAE   MPE MAPE  ACF1 Theil's U
Test set 23.967 531.178 391.812 0.036 2.95 0.011     0.275
> round(accuracy(auto.arima.pred$fitted, sales.ts), 3)
              ME    RMSE     MAE    MPE  MAPE  ACF1 Theil's U
Test set -21.544 528.728 392.174 -0.446 2.985 0.002     0.272
> round(accuracy((snaive(sales.ts))$fitted, sales.ts), 3)
              ME    RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 485.808 915.749 692.108 3.503 5.086 0.351     0.451
> round(accuracy((naive(sales.ts))$fitted, sales.ts), 3)
             ME     RMSE     MAE    MPE   MAPE ACF1 Theil's U
Test set 68.489 1974.397 1482.55 -0.761 12.172 -0.4         1
```

The Auto ARIMA model has the lowest RMSE, and MAPE values, which suggests it has the best overall fit compared to the other models. The Seasonal ARIMA model also performs well having the next lowest RMSE, and MAPE. Based on these metrics, the **Auto ARIMA model** would be the best choice for implementation in this project, as it strikes a good balance between minimizing forecast errors (RMSE and MAPE) and achieving the best overall fit among the models evaluated.

CONCLUSION

After evaluating several forecasting models, including Holt-Winters' Exponential Smoothing, Regression Model with Linear Trend and Seasonality, ARIMA with Seasonal Component, Automatic ARIMA, Seasonal Naive Method, and Naive Method, **the Auto ARIMA model** emerged as the **best performing model** for forecasting based on multiple accuracy metrics. This model achieved the lowest values for **Root Mean Squared Error (RMSE) at 528.728** and **Mean Absolute Percentage Error (MAPE) at 2.985%**, indicating its ability to minimize forecast errors effectively. Moreover, the Auto ARIMA model exhibited the lowest Theil's U statistic of 0.272, suggesting the best overall fit compared to the other models evaluated. While the ARIMA with Seasonal Component model also performed well, with the lowest RMSE, and MAPE values, its slightly higher Theil's U statistic made the Auto ARIMA model a more suitable choice for implementation. Therefore, based on the comprehensive analysis of these accuracy metrics, **the Auto ARIMA model** is recommended for forecasting in this project due to its superior performance and overall fit.

Possible benefits of using Forecasting methods in this case: Accurate sales forecasting can offer several advantages for wholesalers of Beer, Wine and Distilled Alcoholic Beverages. It enables improved inventory management by optimizing stock levels, reducing overstocking or stockouts, and enhancing customer satisfaction. Forecasts also facilitate efficient resource allocation, such as personnel, transportation, and production capacities, to match anticipated demand. Additionally, they provide insights into seasonal patterns and trends, aiding in promotional planning and pricing strategies. Furthermore, forecasts can foster better supply chain coordination among suppliers, distributors, and wholesalers, ensuring a smooth product flow and minimizing disruptions.

Limitations: However, forecasting methods face limitations in this context. Alcohol consumption can be influenced by unpredictable factors like economic conditions, social trends, regulations, and consumer preferences, which may not be fully captured by models. External events like natural disasters, pandemics, or policy changes can significantly impact sales patterns. Data quality and availability are crucial, as incomplete or inaccurate data can compromise forecast accuracy. Different forecasting models have inherent assumptions and limitations, and their performance may vary. Over Reliance on forecasts without considering other factors, such as market intelligence and qualitative inputs, may lead to suboptimal decision-making.

APPENDICES

- The dataset from following website: https://www.census.gov/econ/currentdata/datasets/?programCode=MWTS&startYear=1992&endYear=2024&categories[]=4248&dataType=SM&geoLevel=US&adjusted=1&notAdjusted=1&errorData=0