Forecast Report

AUTHOR
Aaron Younger

PUBLISHED
December 10, 2025

Business Understanding

Business Problem

The data used in this analysis comes from the Federal Reserve Bank of St. Louis. The primary goal of this analysis is to build forecasting models to predict retail sales data. These models can be useful for retail stakeholders in anticipating future demand instead of reacting to it. These models can also help retail stakeholders with better financial planning and budgeting.

Based off this information, the **Business Problem** is to build a forecast model capable of accurately projecting six months into the future for retail stakeholders.

Two Research Questions

Along with the business problem, this report explores two research questions:

1.How accurately can different forecasting methods predict future sales?

2.Does including pre-break data harm or improve forecasting accuracy?

Now that the business problem and research questions have been clearly defined, the next step in this report is exploring the data. This **Data Understanding** phase provides the foundational insight needed for effective data preparation for modeling.

Data Understanding

Import Dataset

Rename RETAILMNSA to Sales

```
# A tibble: 6 × 2
  observation_date      Sales
  <dttm>                <dbl>
1 1992-01-01 00:00:00 126717
2 1992-02-01 00:00:00 127022
3 1992-03-01 00:00:00 137978
4 1992-04-01 00:00:00 142314
```

```
 5 1992-05-01 00:00:00 147550
 6 1992-06-01 00:00:00 146991
```

Renamed RETAILMNSA to a clearer column name, Sales.

Variables and Range of Dates

```
# A tibble: 2 × 3
  variable_name    start_date          end_date
  <chr>            <dttm>              <dttm>
1 observation_date 1992-01-01 00:00:00 2025-06-01 00:00:00
2 Sales            1992-01-01 00:00:00 2025-06-01 00:00:00
```
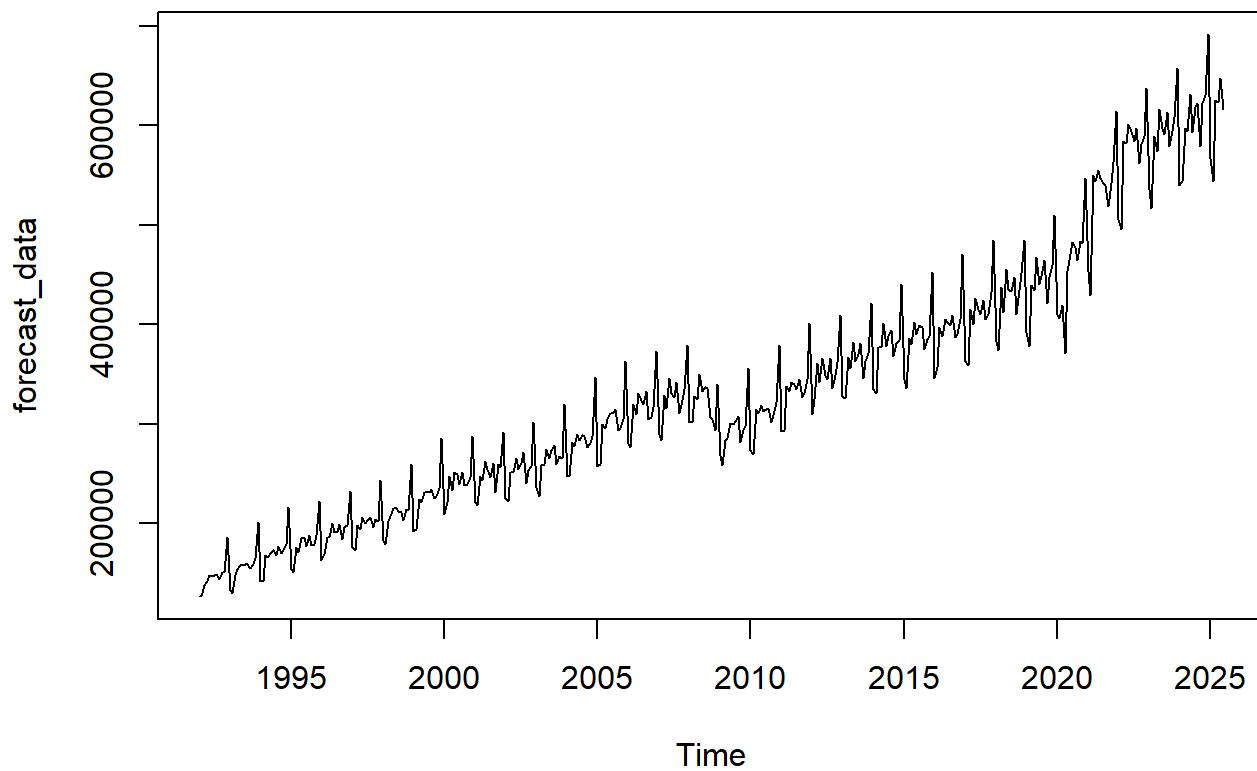
This table shows the variables in the dataset, observation_date and sales, while also showing the start and end date of the time-series dataset. The start date is 1992-01-01 and the end date is 2025-06-01. This means the time-series dataset spans 33 years and 5 months.

Create Forecast Data

The dataset is first arranged in ascending order by the observation_date variable to ensure the time series is properly ordered. Next, the code extracts the year and month of the earliest observation in the dataset, which will serve as the starting point of the time series object. Finally, a monthly (frequency = 12) time series object named forecast_data is created using the Sales variable. This converts the raw retail sales data into a structured time series format that can be used for forecasting models.

Time-Series Plot
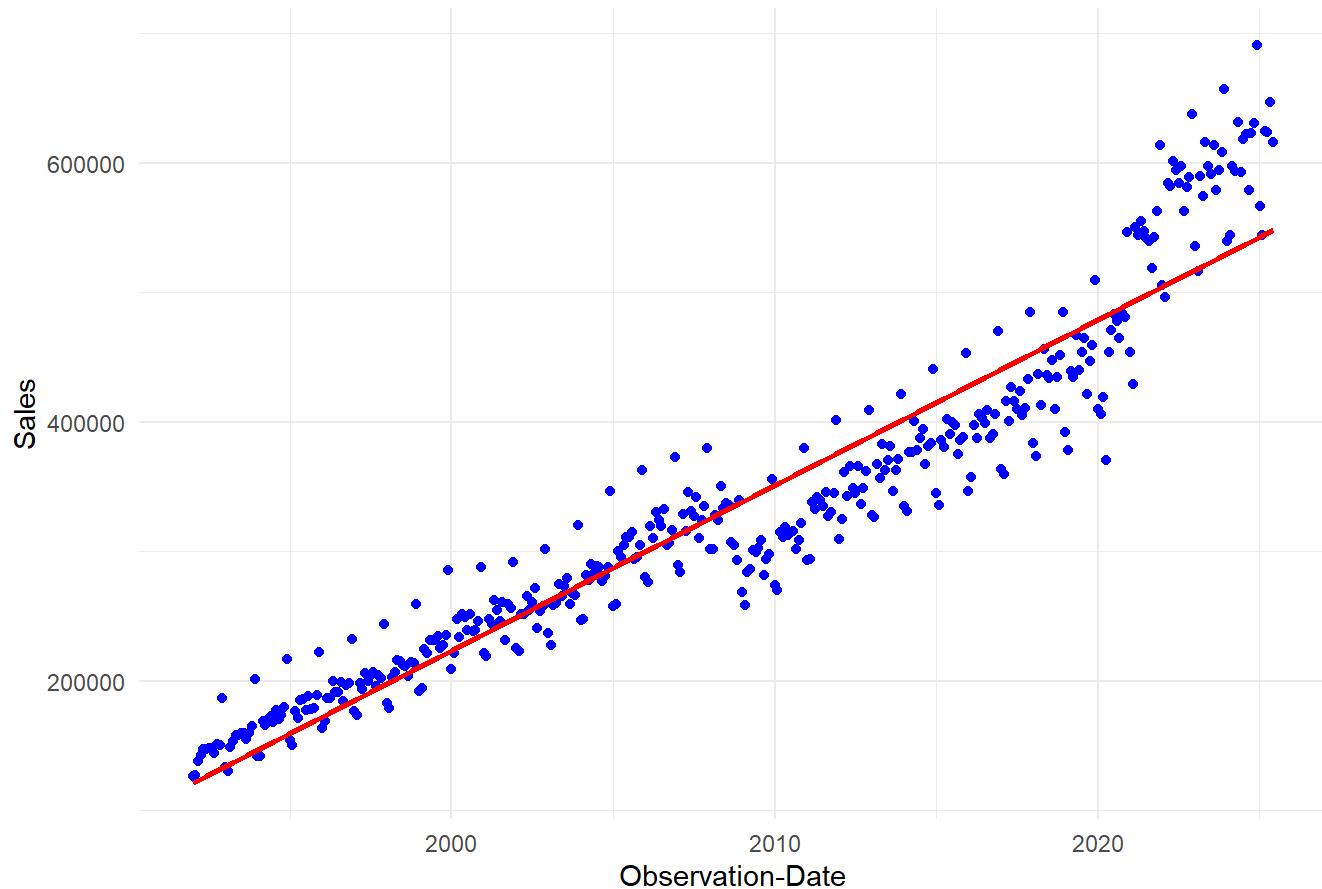
Comments on Time-Series plot:

From the initial view of the plot it looks like there is both trend and seasonality. However, this will be explored further in the modeling phase.

Trend Investigation

## Scatterplot Showing Potential Trend
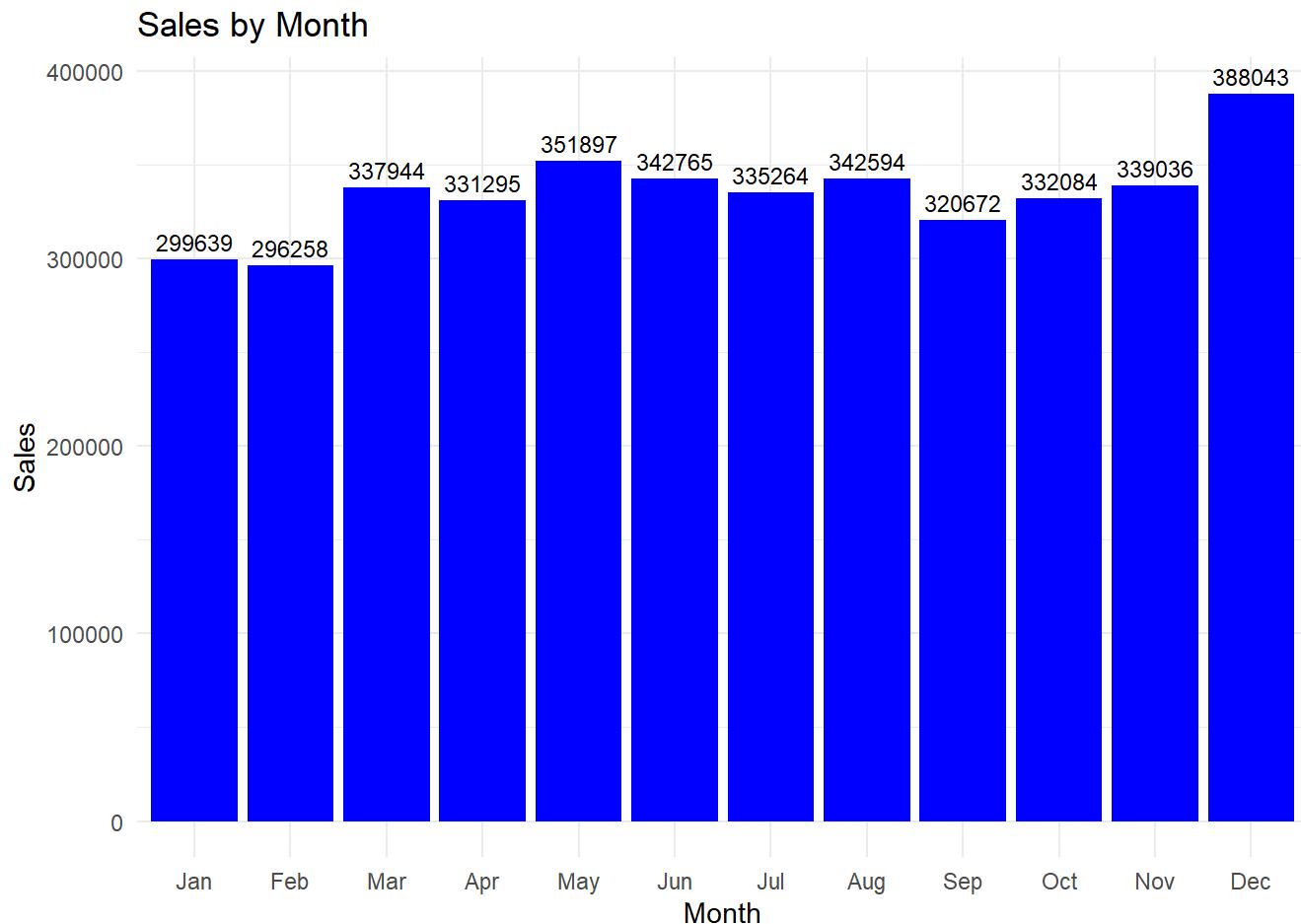
Comments on Scatterplot exploring trend:
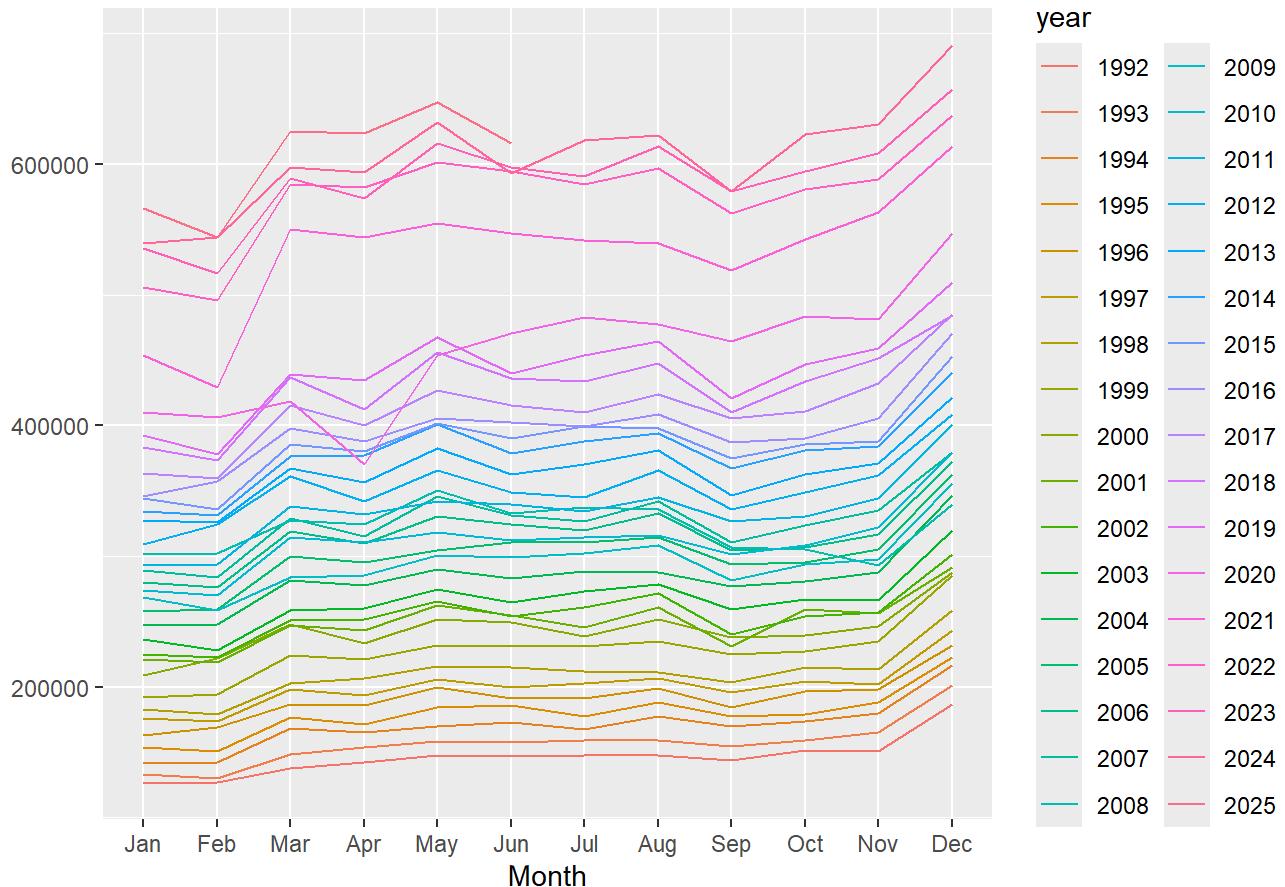
Looking at a scatterplot of the sales by observation date there is a clear trend. As observation date goes up so do sales. So this time-series plot has trend.

Seasonality Investigation

Sales by Month

Seasonal plot: forecast_data

Comments on Seasonality investigation plots:

To determine whether the time series exhibited seasonality, I evaluated two key aspects. First, I examined whether certain months consistently showed higher or lower sales than others. This pattern is clearly visible in the monthly column chart, where December repeatedly shows the highest sales, while January–February tend to be the lowest, and March–May consistently rise above winter months. These predictable differences across months indicate a recurring seasonal pattern.
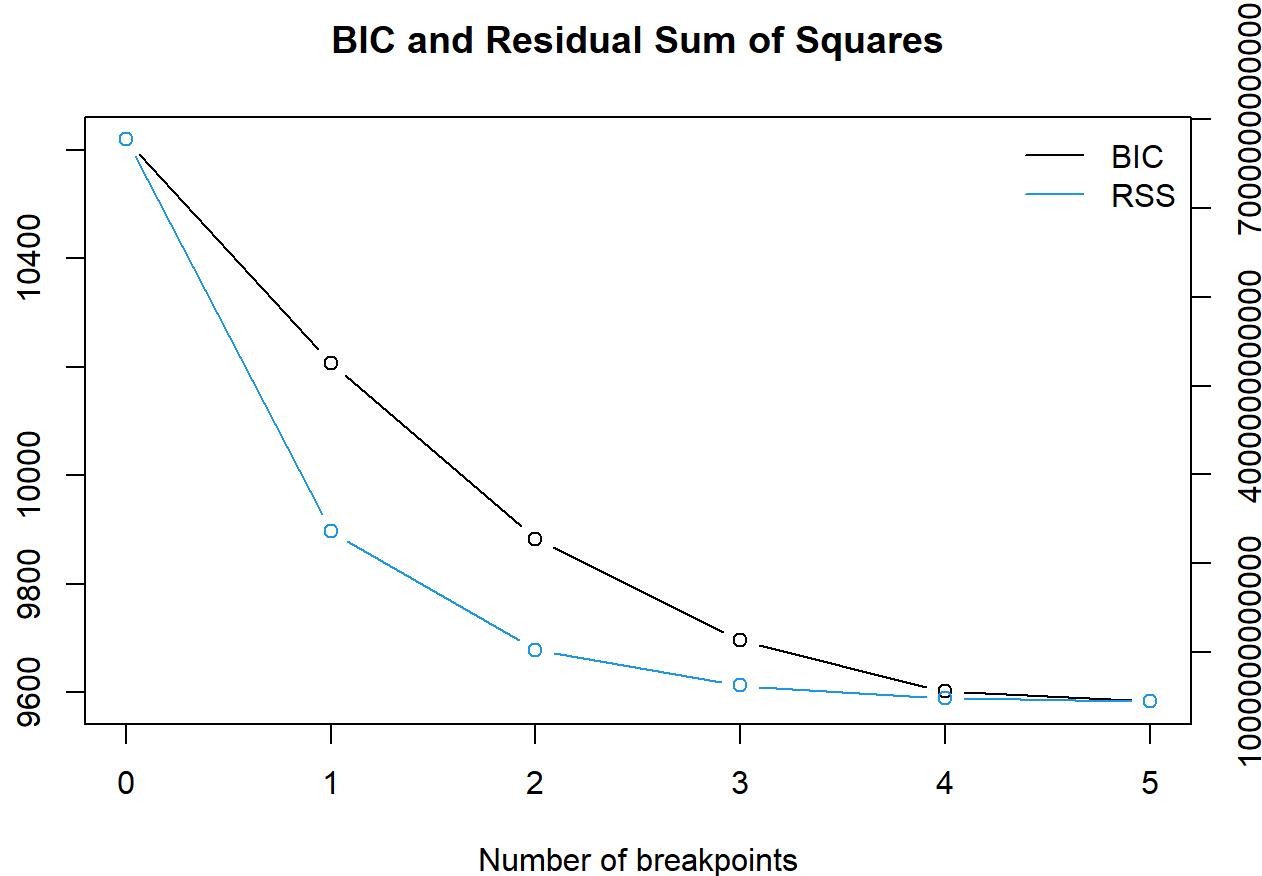
Second, I assessed whether these monthly patterns were consistent across multiple years, which is a defining feature of seasonality. The seasonal line plot confirms this: each year follows a similar shape, with sales increasing in spring, softening in late summer, and rising sharply again in December. The fact that the lines maintain nearly the same month-to-month pattern year after year demonstrates strong and stable seasonality in the data.

Overall, both plots show that the time-series data has seasonality.

Test for Structure Breaks

Bai-Perron Test

BIC and RSS Plot

BIC and Residual Sum of Squares

Number of breakpoints

The BIC and RSS plot evaluates how model fit improves as additional structural breakpoints are added. Both the Bayesian Information Criterion (BIC) and Residual Sum of Squares (RSS) decrease as the number of breakpoints increases, indicating that models with more breaks better capture shifts in the underlying data-generating process. The BIC curve reaches its lowest point at five breakpoints, suggesting this is the optimal model complexity before overfitting becomes a concern.

Breakpoint Confidence Intervals

```
[1] 1998.250 2004.083 2010.417 2015.417 2020.417


        Confidence intervals for breakpoints
        of optimal 6-segment partition:

Call:
confint.breakpointsfull(object = sb)

Breakpoints at observation number:
  2.5 % breakpoints 97.5 %
1    74          76      78
2   144         146     148
3   217         222     226
4   278         282     285
5   340         342     343
```
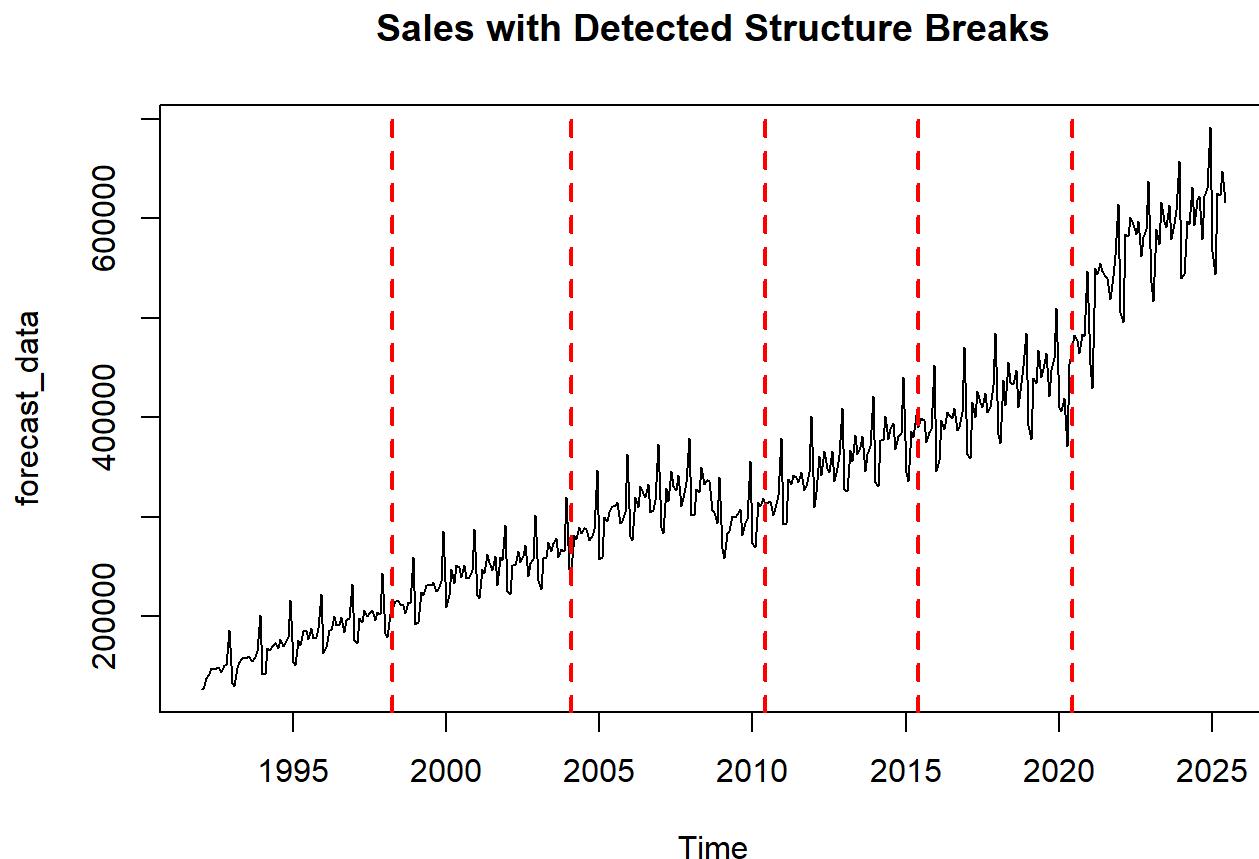
```
Corresponding to breakdates:
  2.5 %     breakpoints 97.5 %
1 1998(2)  1998(4)      1998(6)
2 2003(12) 2004(2)      2004(4)
3 2010(1)  2010(6)      2010(10)
4 2015(2)  2015(6)      2015(9)
5 2020(4)  2020(6)      2020(7)
```

The breakpoint summary reports the estimated locations of structural changes in the sales series along with 95% confidence intervals for each break. The detected breakpoints occur around late 1998, late 2004, early 2010, mid-2015, and mid-2020. These results indicate five major periods where the statistical properties of the retail sales time series shifted significantly. The narrow confidence intervals reflect strong evidence supporting these structural changes.

Time Series Plot with Structural Breaks

Sales with Detected Structure Breaks

The time-series plot overlays the fitted structural breakpoints on the retail sales series. The vertical red dashed lines mark the locations where significant changes in trend or growth rate occur. These breaks visually align with periods of noticeable shifts in the sales pattern, such as changes in overall growth slope or volatility. The plot confirms that retail sales do not follow a single continuous trend but evolve through distinct regimes over time.

Structural breaks will not be a concern for the regression forecasting model in data preparation but will need to be addressed for the Holt-Winters model.

Data Preperation

Regression

Create Financial Global Crisis dummy variable

A dummy variable was created to capture the impact of the economic crisis period on retail sales. The crisis window was defined from January 2009 to March 2020. For every month within this range, the dummy variable takes the value 1, and 0 otherwise. The dummy was then converted into a time-series object aligned with the same start date and monthly frequency as the retail sales data, ensuring compatibility with the regression model.

Create Covid Dummy Variable

A second dummy variable was constructed to account for the structural effects of the COVID-19 pandemic. The COVID period was defined as beginning in April 2020 and continuing through the end of the sample. All observations from April 2020 onward are assigned a value of 1, while earlier periods are assigned 0. This dummy variable was also converted into a monthly time-series format consistent with the main dataset to allow inclusion in the regression forecasting model.

Partition Forecast Data and Dummy Variables

The dataset was divided into a training set and a test set to evaluate forecast accuracy. The training set consists of observations from January 1992 through May 2024, while the test set includes the final 12 months of data, from June 2024 through June 2025. This split allows the models to be trained on the historical pattern and then assessed on their ability to predict unseen future values. The dummy variables were also divided into corresponding training and testing periods to ensure they aligned with the sales data, allowing the model to evaluate whether these structural events have a measurable effect on forecasting accuracy.

Check Lenghts of Partitioning

```
trainset has 389 observation.
```

```
testset has 13 observation.
```

```
crisis trainset has 389 observation.


crisis testset has 13 observation.


covid trainset has 389 observation.


covid testset has 13 observation.
```
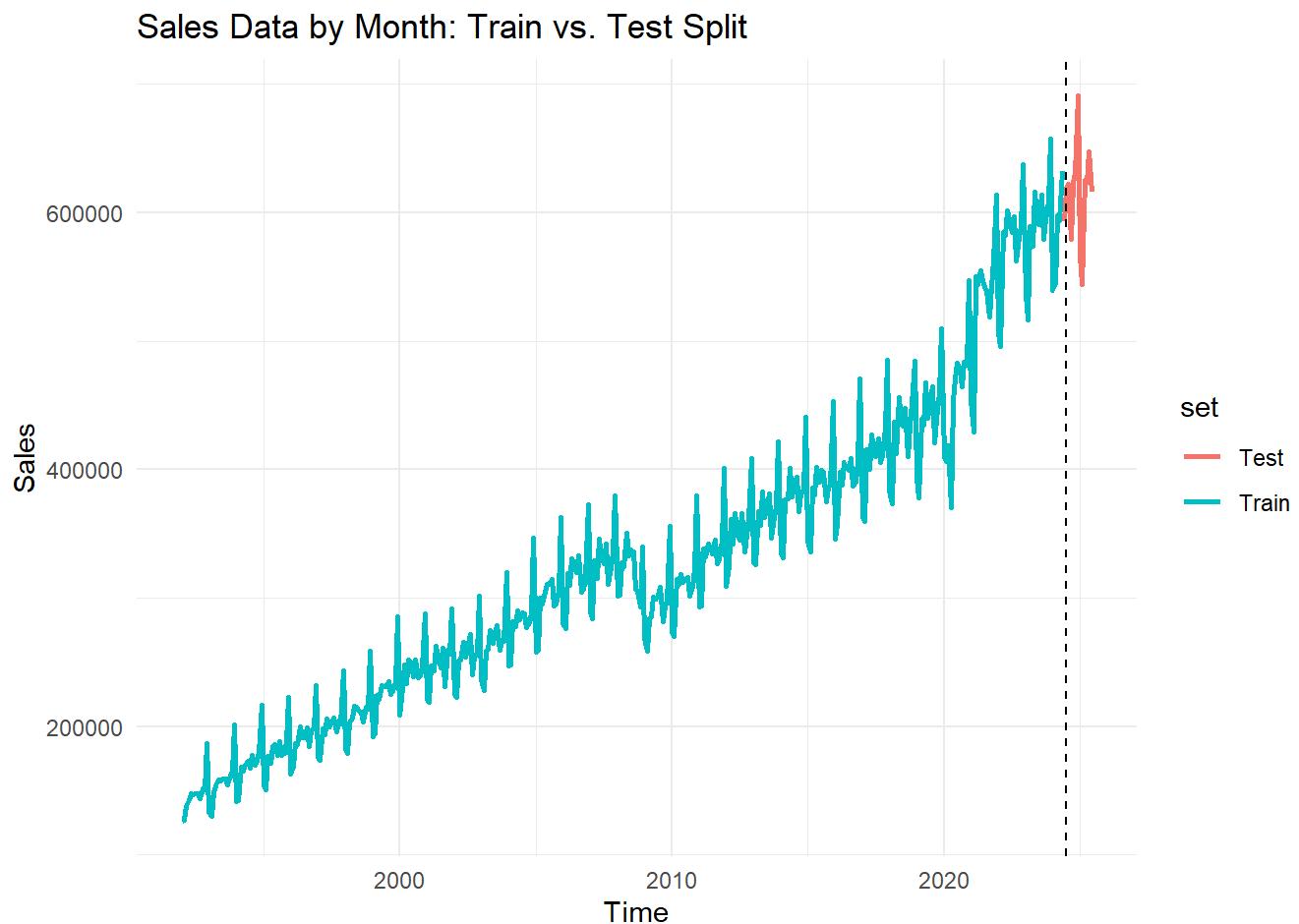
This shows that the partitioning was done correctly as each train and test set have the same amount of observations.

Visualize traind and test split

Sales Data by Month: Train vs. Test Split

This is a visual way to see that our partitioning worked correctly. The blue part of the line represents the amount of time that will be used for the train set, the blue line spans the correct times. The red line represents the amount of time that will be used for the test set, the red line spans the correct times.

Holt-Winters Data Preparation

Since the time-series data has both trend and seasonality, the Holt-Winters exponential smoothing method is the appropriate choice, as it is specifically designed to model series with both trend and seasonality.

To test the second research question, the Holt–Winters exponential smoothing model will be trained on two datasets that differ in their time span. The first dataset will include data after the most recent structure break, this ensures the model will be trained and tested on a stable level of both trend and seasonal pattern. The second dataset will have all observation dates. This is to test and see if structural breaks affect forecast accuracy for this data.

Create Post-Structure Break Data

Because the most recent structural break occurred around July 2020, only data from 2020-07-01 onward is used to train the Holt-Winters model.

Convert to Time Series object

Modeling: Regression

Rolling-Origin Cross-validation

```
   Fold                         Model       RMSE       MAPE
1     1                         trend   98342.62 16.433593
2     1                 trend + season   93741.15 16.469002
3     1         trend + season + crisis  61096.63 10.354668
4     1          trend + season + covid  65902.85 11.253176
5     1 trend + season + crisis + covid  63988.35 10.907190
6     2                         trend  102047.38 16.483178
7     2                 trend + season   98237.66 16.652156
8     2         trend + season + crisis  55934.70  9.236111
9     2          trend + season + covid  52114.79  8.579710
10    2 trend + season + crisis + covid  49352.26  8.070224
11    3                         trend   91660.56 14.121568
12    3                 trend + season   87393.08 14.456157
13    3         trend + season + crisis  41698.11  6.542055
14    3          trend + season + covid  38897.35  6.076912
15    3 trend + season + crisis + covid  35353.58  5.415304
```

Average Error per Model

```
# A tibble: 5 × 3
  Model                        mean_RMSE mean_MAPE
  <chr>                            <dbl>     <dbl>
1 trend                           97350.      15.7
2 trend + season                  93124.      15.9
3 trend + season + covid          52305.       8.64
```

```
4 trend + season + crisis              52910.        8.71
5 trend + season + crisis + covid      49565.        8.13
```
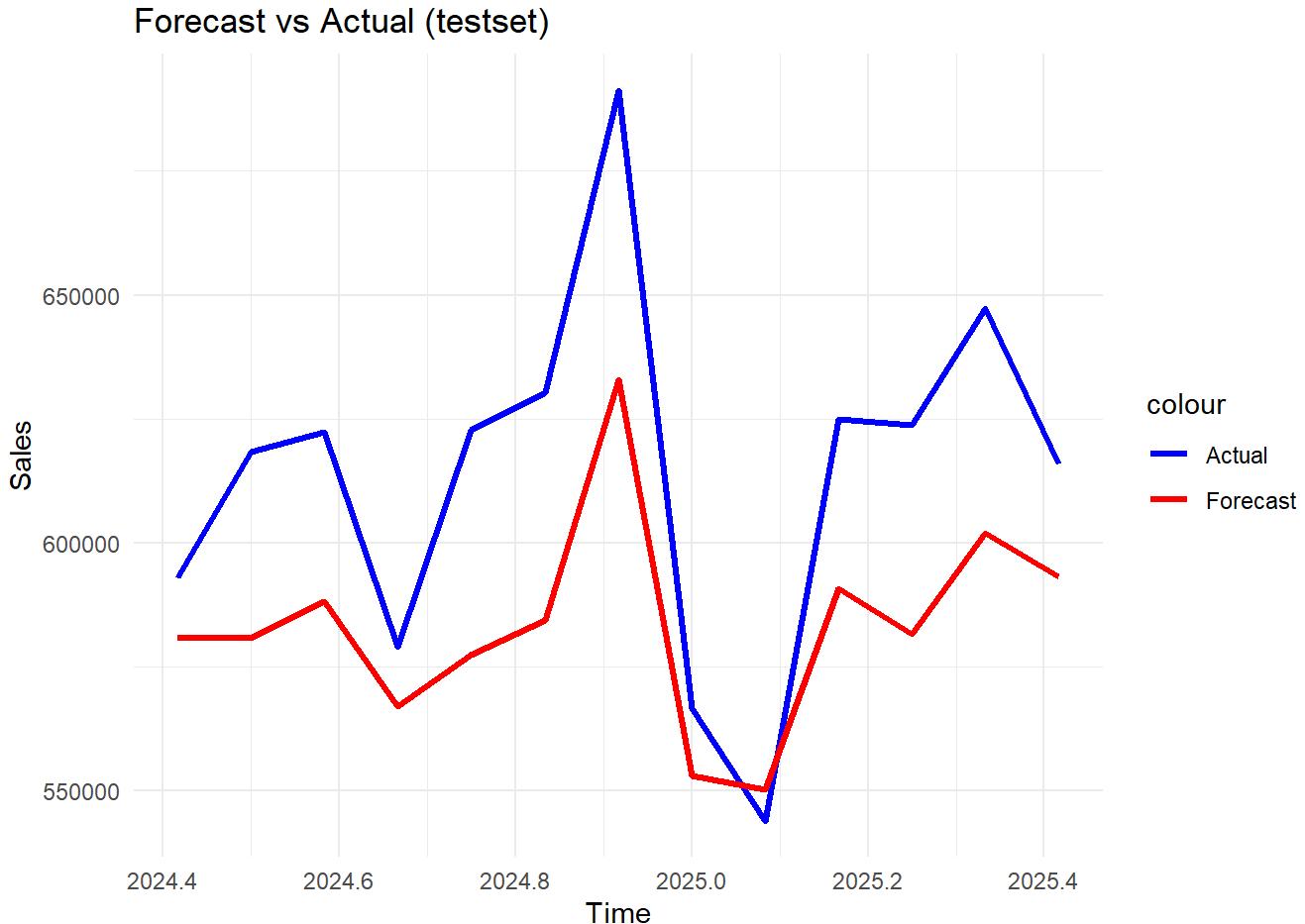
From the Rolling-Origin Cross Validation the model that performs the best with the lowest average prediction error is the model with trend + season + crisis + covid.

Evaluation Metrics: Best Regression Model

```
                               ME       RMSE       MAE        MPE      MAPE
Training set     -0.0000000000001505291  17119.07  10431.24  0.08585422  3.257943
Test set      30589.4571014014436514117  35311.26  31558.41  4.84080477  5.018931
                    MASE       ACF1  Theil's U
Training set  0.601290  0.68659688         NA
Test set      1.819128  0.08302784  0.7257939
```

The regression model fits the training data very well. The Mean Error (ME) is essentially zero, indicating no systematic over or under prediction during model estimation. The RMSE (17,119) and MAE (10,431) are reasonable and the low MAPE (2.26%) shows that the model predicts historical values with strong accuracy. This Model will be used to compare against the holt-winters model to test and see which model should be used in addressing the business problem.

Graph Forecasted vs Actual Test Set

The graph shows that the forecast values capture the general direction of sales movement but consistently underestimate the magnitude of actual sales.

Modeling: Holt-Winters (Post-Structure Break Data)

3-fold CV with AAA and AAM

```
  Fold Model      RMSE       MAPE
1    1    AAA 43264.84 6.255587
2    1    AAM 43264.84 6.255587
3    2    AAA 39120.83 5.936093
4    2    AAM 43212.41 6.270994
5    3    AAA 36336.80 5.735455
6    3    AAM 36336.80 5.735455
```

Comparison of Average Error

```
# A tibble: 2 × 3
  Model mean_RMSE mean_MAPE
  <chr>     <dbl>     <dbl>
1 AAA      39574.      5.98
2 AAM      40938.      6.09
```

Comments on Post-structure Break Data 3-Fold CV with AAA and AAM: The Holt-Winters models evaluated on the post-structural break dataset included both the AAA and AAM specifications. The AAA model assumes constant seasonal effect, whereas AAM assumes seasonal effects that scale proportionally. Three-fold time series cross validation was used to compare both models, using RMSE and MAPE as evaluation metrics. Based on the average errors across the folds, the AAA model outperformed the AAM model. The AAA model will used to compare against the best holt-winters model run on the full dataset.

Modeling: Holt-Winters (Full Dataset)

3-fold CV with AAA and AAM

```
  Fold Model      RMSE       MAPE
1    1    AAA 21300.79 2.915173
2    1    AAM 14947.39 2.081649
3    2    AAA 16659.65 2.558220
4    2    AAM 11283.81 1.251424
5    3    AAA 10607.14 1.608037
6    3    AAM 10607.14 1.608037
```

Comparison of Average Errors

```
# A tibble: 2 × 3
  Model mean_RMSE mean_MAPE
```

```
       <chr>        <dbl>        <dbl>
 1 AAA          16189.         2.36
 2 AAM          12279.         1.65
```

Comments on Full Data 3-Fold CV with AAA and AAM: Similar to the post-structure break analysis, both the AAA and AAM Holt-Winters models were evaluated on the full dataset using three-fold time series cross validation. Based on the average RMSE and MAPE values across the folds, the AAM Model showed superior accuracy. The full-data AAM model will be used in comparison against the best-performing model from the post-structural break dataset.

Compare Best Holt-Winters Models

```
# A tibble: 2 × 4
  Model mean_RMSE mean_MAPE Data
  <chr>      <dbl>     <dbl> <chr>
1 AAA       39574.       5.98 Post-Break
2 AAM       12279.       1.65 Full Dataset
```

The AAA model from the post-structural break data and the AAM model from the full observation date data were compared using the evaluation metrics RMSE and MAPE. The AAM model from the full dataset had the lowest RMSE and MAPE indicating it has superior accuracy. This suggests that using the full dataset will lead to improved forecasting accuracy results.

Evaluation Metrics: Best Holt Winters Model

```
                   ME       RMSE       MAE         MPE      MAPE      MASE
Training set   92.36211   9416.832   6053.883  -0.05412857  1.844098 0.3496515
Test set     16933.81783 19198.314 16933.818   2.73017922  2.730179 0.9780392
                  ACF1 Theil's U
Training set 0.04783256        NA
Test set     0.05675312 0.3701533
```

The AAM Holt-Winters Model trained over the full dataset has a relatively low RMSE and MAE value compared to the scale of the data. The MASE value is slightly below one (0.978), meaning the model performs better than an naive seasonal benchmark. Also the Theil's U statistic of 0.37 is well below one meaning the forecast outperforms a simple random-walk.

The Evaluation phase will compare the best regression model against the AAM Holt-Winters model trained over the full dataset.

Evaluation

Compare Best Forecasting Models

```
=== Holt-Winters Test Accuracy ===


              ME              RMSE              MAE              MPE              MAPE
16933.81782723 19198.31353622 16933.81782723      2.73017922       2.73017922
            MASE              ACF1         Theil's U
    0.97803923      0.05675312      0.37015330



=== Regression Test Accuracy ===


              ME              RMSE              MAE              MPE              MAPE
30589.45710140 35311.26172502 31558.40881437      4.84080477       5.01893071
            MASE              ACF1         Theil's U
    1.81912787      0.08302784      0.72579395
```
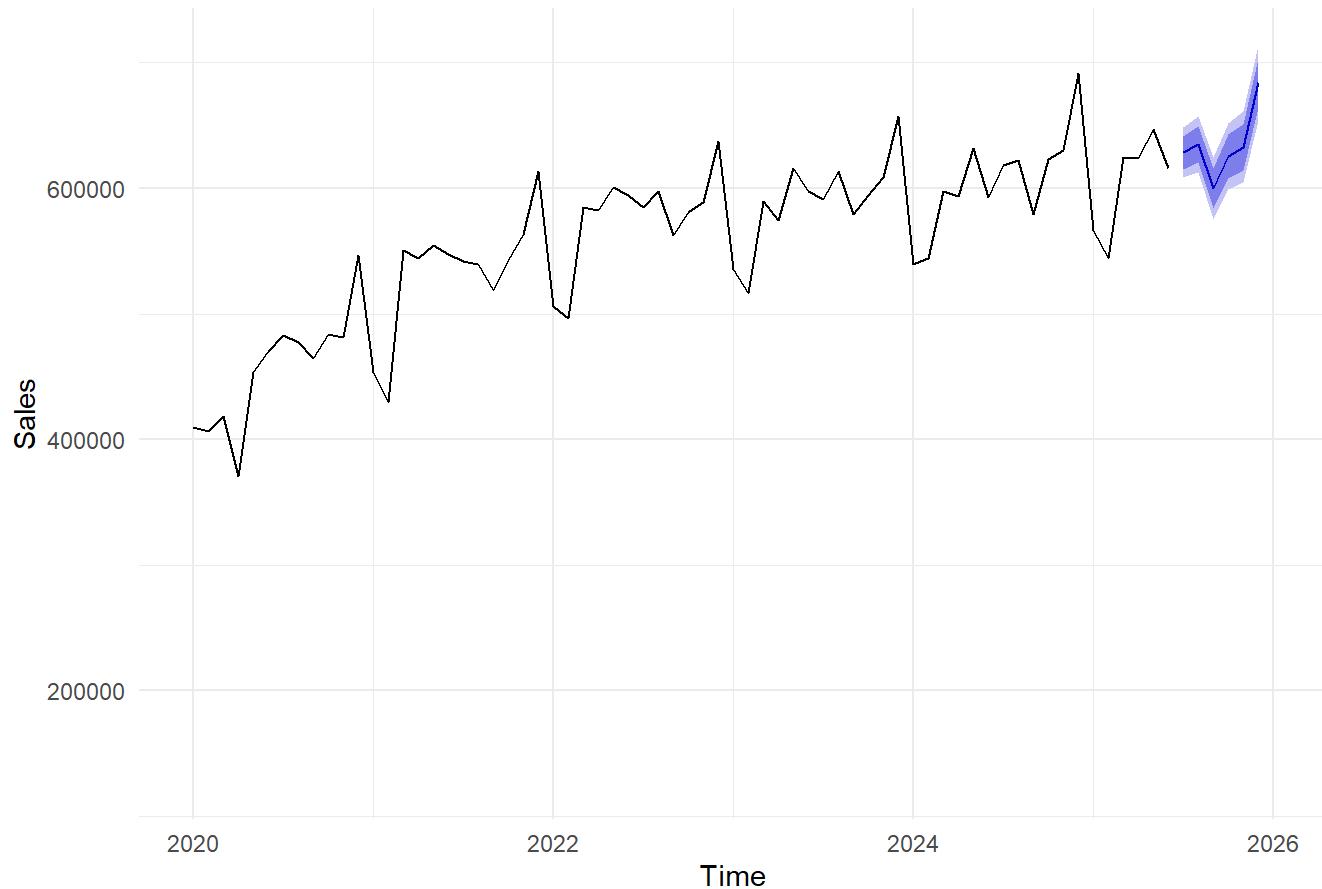
The Holt-Winters AAM model should be deployed as the final forecasting solution because it delivers substantially lower forecast errors than the regression model and demonstrates superior predictive performance across all evaluation metrics, including RMSE, MAPE, MASE, and Theil's U. This indicates that Holt-Winters captures the underlying structure of retail sales more effectively and provides more reliable forward-looking predictions.

Deployment: Holt-Winters

Forecasting 2025-07 - 2025-12

Retail Sales Forecast – Full Year 2025

The graph presents a six-month forecast generated by the Holt-Winters model. The projected values follow a seasonal pattern and overall trend that closely resemble those observed in previous years. This consistency indicates that the Holt-Winters model is effectively capturing the underlying structure of the retail sales data and provides confidence in its forecasting accuracy.

Answer to Business Question

This business problem guiding this analysis was to is to build a forecast model capable of accurately projecting six months into the future for retail stakeholders. Through modeling and evaluation the AAM Holt-Winters smoothing model trained on the entire dataset offered the most accurate predictions. This is supported by the graph above as the 6 months of forecasted values follow similar seasonal patterns and overall trend from previous years. This tool can be used by retail stakeholders in predicting future dates of retail sales data.

Answer to Research Questions

1.How accurately can different forecasting methods predict future sales?

The analysis demonstrates that forecasting methods vary in their ability to accurately predict future sales. While both models provide reasonable estimates certain forecasting models will perform better than others

so it is worth exploring multiple different methods. However, based on this analysis, forecasting methods perform far better than a random or naïve approach.

2.Does including pre-break data harm or improve forecasting accuracy?

For this analysis including pre-break structure helped improve forecast accuracy for the Holt-Winters model. In theory, the opposite is often expected, models typically perform better when trained only on post-break data, since older structural patterns may no longer reflect current dynamics. However, in this case the earlier data still contained valuable seasonal and trend information that strengthened the model's ability to forecast future values.

Business Recommendations

If retail stakeholders were to implement the Holt-Winters AAM model I have two business recommendations. The first recommendation is to use it for short-term operational planning. Given that the Holt-Winters AAM model consistently produced the most accurate six-month forecasts and successfully captured seasonal retail patterns, retail stakeholders could adopt this model as a primary tool for short-term sales planning. The second business recommendation is to retain long run historical data. Although structural break theory suggests older data may sometimes distort model accuracy, this analysis showed that retaining pre-break historical data improved the Holt-Winters model's performance. This indicates that long-run seasonal cycles and trend components remain valuable for retail stakeholders to produce more accurate forecasts.

References

Citation of orignial data source with authors

U.S. Census Bureau. Retailers Sales [RETAILSMNSA]. FRED, Federal Reserve Bank of St. Louis. Retrieved from https://fred.stlouisfed.org/series/RETAILSMNSA

Citation of ChatGP

OpenAI. (2025). ChatGPT (GPT-5.1 Thinking) [Large language model].https://chat.openai.com/

Version of R

```
Version R version 4.5.2 (2025-10-31 ucrt)
```

R packages used during analysis

Wickham H, Bryan J (2025). readxl: Read Excel Files. R package version 1.4.5, https://readxl.tidyverse.org.

Cui B (2025). DataExplorer: Automate Data Exploration and Treatment. R package version 0.8.4, http://boxuancui.github.io/DataExplorer/.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686.

Wickham H, François R, Henry L, Müller K, Vaughan D (2025). dplyr: A Grammar of Data Manipulation. R package version 1.1.4, https://dplyr.tidyverse.org.

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, https://ggplot2.tidyverse.org.

Grolemund G, Wickham H (2011). "Dates and Times Made Easy with lubridate." Journal of Statistical Software, 40(3), 1–25. https://www.jstatsoft.org/v40/i03/.

Hyndman RJ, Khandakar Y (2008). "Automatic Time Series Forecasting: The forecast Package for R." Journal of Statistical Software, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03.

Zeileis A, Leisch F, Hornik K, Kleiber C (2002). "strucchange: An R Package for Testing for Structural Change in Linear Regression Models." Journal of Statistical Software, 7(2), 1–38. https://doi.org/10.18637/jss.v007.i02.