FEBRUARY 2024

CASE STUDY #2

Shipment of Appliances in U.S.

Report by:
Pallavi Nair
fb4097

BRIEF GOALS & OBJECTIVES OF THE REPORT

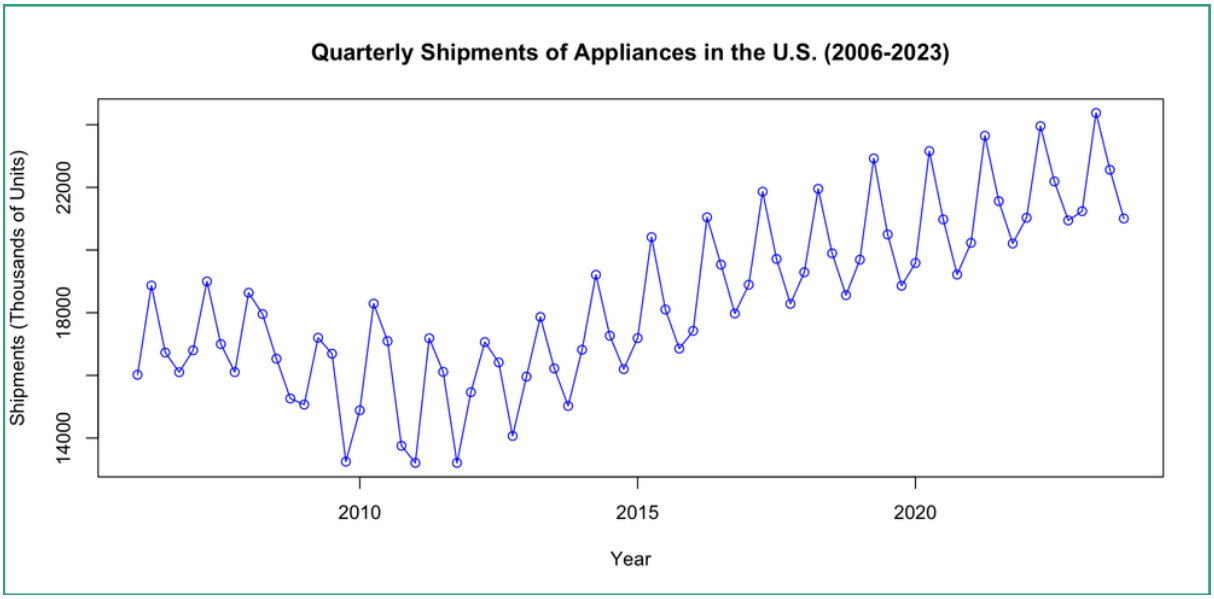
This report will cover the visualization of historical data, the development and evaluation of various regression models, and the forecasting of appliance shipments for 2024 and 2025.

1A. TIME SERIES DATASET CREATION

The dataset was transformed into a time series object, shipments.ts, using the ts() function in R. This object encapsulates quarterly shipments data from 2006 to 2023, with each observation representing the number of appliances shipped in thousands of units.

1B. DATA PLOT

A plot of the quarterly shipments data was generated to visualize the historical trends and patterns.



Analysis of Time Series Components:

- Trend: The plot exhibits a generally upward trend over the years, indicating an increase in appliance shipments over time.
- Seasonality: There are patterns that repeat annually, suggesting seasonal fluctuations in appliance shipments.
- Irregularity: Variations not explained by the trend or seasonality are observed, which could be due to external factors or random fluctuations.

2A. DATA PARTITIONING

The data was partitioned into a training set (up to the end of 2018) and a validation set (starting from 2019), with the latter consisting of 20 quarters.

```
nValid <- 20
nTrain <- length(shipments.ts) - nValid
# Define the training and validation sets using window()
train.ts <- window(shipments.ts, end = c(2018, 4)) # Adjusted to ensure correct partitioning
valid.ts <- window(shipments.ts, start = c(2019, 1)) # Starting from 2019 Q1 for validation
```

2B. MODEL DEVELOPMENT AND EVALUATION

Five regression models were developed using the tslm() function:

Linear Trend Model
Model Summary:

- Equation: Shipments = 15482.24 + 65.63 * Trend
- Coefficients: Intercept (15482.24), Trend (65.63)
- Significance: Both coefficients are statistically significant ($p < 0.05$).
- Goodness of Fit: R-squared = 0.2339, indicating that approximately 23.39% of the variance in shipments is explained by the model.

Forecast Evaluation:

- **MAPE:** 8.283%
- **RMSE:** 2330.402

Quadratic Trend Model
Model Summary:

- Equation: Shipments = 17832.614 - 195.523 * Trend + 4.927 * Trend^2
- Coefficients: Intercept (17832.614), Trend (-195.523), Trend^2 (4.927)
- Significance: All coefficients are statistically significant ($p < 0.05$).
- Goodness of Fit: R-squared = 0.4712, a significant improvement over the linear model, indicating a better fit to the data.

Forecast Evaluation:

- **MAPE**: 17.856%
- **RMSE**: 4309.635

Seasonality Model
Model Summary:

- Equation: Shipments = 16588.3 + Seasonal Adjustments
- Coefficients: Intercept (16588.3), Seasonal Adjustments (varied)
- Significance: Season2 is significant; Season3 and Season4 are not statistically significant at the 0.05 level.
- Goodness of Fit: R-squared = 0.3664, indicating seasonality accounts for some of the variance in shipments.

Forecast Evaluation:

- **MAPE**: 19.434%
- **RMSE**: 4243.646

Linear Trend and Seasonality Model
Model Summary:

- Equation: Shipments = 14880.87 + 68.30 * Trend + Seasonal Adjustments
- Coefficients: Intercept (14880.87), Trend (68.30), Seasonal Adjustments (varied)
- Significance: All coefficients except for season3 are statistically significant.
- Goodness of Fit: R-squared = 0.6183, indicating a strong model that captures both trend and seasonality.

Forecast Evaluation:

- **MAPE**: 7.94%
- **RMSE**: 1775.152

Quadratic Trend and Seasonality Model
Model Summary:

- Equation: Shipments = 17238.7825 - 194.2423 * Trend + 4.9536 * Trend^2 + Seasonal Adjustments
- Coefficients: Intercept (17238.7825), Trend (-194.2423), Trend^2 (4.9536), Seasonal Adjustments (varied)
- Significance: All coefficients are statistically significant.
- Goodness of Fit: R-squared = 0.8582, indicating a very strong fit to the data, capturing both trend and seasonality effectively.

Forecast Evaluation:

- **MAPE**: 17.472%
- **RMSE**: 4149.346

2C.IDENTIFICATION OF THE MOST ACCURATE MODELS

- **Most Accurate Model:** The **Linear Trend and Seasonality Model** shows the best forecasting accuracy with the lowest MAPE (7.94) and RMSE (1775.152). This model effectively captures both the linear trend and seasonal patterns in the data, making it highly suitable for forecasting future shipments.
- **Second Most Accurate Model:** The **Linear Trend Model** follows, with a slightly higher MAPE (8.283) and RMSE (2330.402) than the Linear Trend and Seasonality Model. While it captures the overall trend, it does not account for seasonal variations as effectively as the combined model.
- **Third Most Accurate Model:** Surprisingly, the more complex **Quadratic Trend and Seasonality Model,** despite its higher R-squared value indicating a good fit

to the training data, does not translate into higher forecast accuracy. It has a higher MAPE (17.472) and RMSE (4149.346), possibly due to overfitting or the model capturing noise along with the actual signal.

3A. ENTIRE DATASET TIME SERIES FORECASTING FOR 2024 AND 2025

Model Summaries and Forecasts
Linear Trend Model:

- Equation: Shipments = 14848.42 + 96.77 * Trend
- R-squared: 0.5636, indicating that around 56.36% of the variance in shipments is explained by the model.

Linear Trend and Seasonality Model:

- Equation: Shipments = 14199.51 + 98.15 * Trend + Seasonal Adjustments
- R-squared: 0.8046, showing a significant improvement in explaining the variance in shipments.

Quadratic Trend and Seasonality Model:

- Equation: Shipments = 16035.5869 - 50.8852 * Trend + 2.0416 * Trend^2 + Seasonal Adjustments
- R-squared: 0.8913, indicating that this model has the highest explanatory power among the three.

```
> lin.trend.pred
        Point Forecast     Lo 0     Hi 0
2024 Q1       21912.30 21912.30 21912.30
2024 Q2       22009.07 22009.07 22009.07
2024 Q3       22105.83 22105.83 22105.83
2024 Q4       22202.60 22202.60 22202.60
2025 Q1       22299.36 22299.36 22299.36
2025 Q2       22396.13 22396.13 22396.13
2025 Q3       22492.90 22492.90 22492.90
2025 Q4       22589.66 22589.66 22589.66
```

```
> lin.season.pred
        Point Forecast     Lo 0     Hi 0
2024 Q1       21364.55 21364.55 21364.55
2024 Q2       24061.91 24061.91 24061.91
2024 Q3       22346.39 22346.39 22346.39
2024 Q4       20667.57 20667.57 20667.57
2025 Q1       21757.16 21757.16 21757.16
2025 Q2       24454.51 24454.51 24454.51
2025 Q3       22739.00 22739.00 22739.00
2025 Q4       21060.18 21060.18 21060.18
```

```
> quad.season.pred
        Point Forecast     Lo 0     Hi 0
2024 Q1       23200.63 23200.63 23200.63
2024 Q2       26053.14 26053.14 26053.14
2024 Q3       24492.79 24492.79 24492.79
2024 Q4       22969.13 22969.13 22969.13
2025 Q1       24222.04 24222.04 24222.04
2025 Q2       27090.89 27090.89 27090.89
2025 Q3       25546.87 25546.87 25546.87
2025 Q4       24039.55 24039.55 24039.55
```

FORECAST ACCURACY COMPARISON

Forecast Accuracy Comparison

- **Linear Trend Model**
  - RMSE: 1769.768
  - MAPE: 8.18
- **Linear Trend and Seasonality Model**
  - RMSE: 1184.034
  - MAPE: 5.375
- **Quadratic Trend and Seasonality Model**
  - RMSE: 883.314
  - MAPE: 3.942
- **Naive Forecast**
  - RMSE: 2050.86
  - MAPE: 10.003
- **Seasonal Naive Forecast**
  - RMSE: 930.796
  - MAPE: 4.125

Analysis:

- The Quadratic Trend and Seasonality Model is the most accurate, showing the lowest RMSE and MAPE values. This indicates it's the best model among those tested for forecasting future shipments, capturing both the trend and seasonal patterns effectively.
- The Linear Trend and Seasonality Model also performs well, significantly better than the simple Linear Trend Model and the naive forecasts, making it a viable alternative for forecasting, especially if simplicity and interpretability are preferred.
- Compared to the naive models, all regression models offer better accuracy

Recommendations for Forecasting 2024 and 2025 Shipments:

- Quadratic Trend and Seasonality Model should be used for the most accurate forecasts, as it best captures the complex patterns in the shipments data.
- The improved accuracy of these models over naive approaches underscores the importance of using sophisticated time series analysis techniques for business forecasting.