Name: Andrea Astudillo

Business Forecasting

25 November 2024

Tesla Forecasting Analysis: Model Y

1. **Forecasting question and its importance**

Which forecast method performs better for next year's sales demand using Tesla’s historical data?

1. **Describe the data**

Model Y is an electric crossover vehicle that production began in early 2020 in worldwide factories, in China and Germany, and began its deliveries to the market in mid-2020. Model Y shares 75% of its components with Model 3 but with more cargo space. The main objective of Model Y was to outsell model 3. In 2023 Model Y was the top 5 most selling cars in the US surpassing the sales of Model 3. The dataset for Model Y is based on monthly sales units from July 2022 until December 2023.

1. **Insights from Exploratory Data Analysis**

In the first three months, Tesla sold an average of 10,000 cars per month. Suddenly, in the next three months, sales dropped dramatically to 1,000 units per month. After this, from 2021, a constant growth period was seen in Model Y. However, in the months of April to June of 2023, sales peaked more than usual, reaching 35,000 pieces per month.

In the central tendency results, the box blot shows a right-skewed distribution. Indicating production planning can be underestimated in certain cases. With lower predicted sales than the actual outcomes, understocking can result in unmet demand. In the Acf plot, some lags crossed the confidence intervals, implying unexpected behaviors were not captured.

1. **Accuracy measure and its importance**

Mean Absolute Percentage Error (MAPE) explains how the forecast deviates from the actual sales as a percentage error. It penalizes large errors with a greater penalty and not only fits the historical data but also predicts based on significant fluctuations. Since the objective of this analysis is to understand the market, interpreting the results of MAPE will guide decisions related to inventory management.

1. **Insights from different forecasting methods and their residual analysis**
2. The Naïve Method uses the last observed value as the next forecast value assuming no trend, seasonality, or any fluctuations.

* The residual results could be analyzed through different plots.
  + The histogram is normally distributed as most of its values have a mean of zero. The right side of the histogram has a perfect bell shape, however, on the left side a gap is formed. Implying that the model has missed some important values.
  + In the Acf plot, all the values are close to zero without crossing the confidence interval.

1. The Simple Exponential Smoothing Method uses a weighted average to forecast the next value. The alpha value of the simple exponential smoothing is 0.8758, implying the model is giving more weight to recent values.

* The residual results could be analyzed through different plots.
  + The histogram for the simple exponential smoothing is very similar to the naïve. One side of the plot has a perfect bell shape and the other a gap, indicating some missing values were left out of the analysis.
  + In the Acf plot, all the values are close to zero without crossing the confidence interval.

1. The Holt-Winters Additive Method is effective when seasonality fluctuations are constant over time. The alpha value of the Holt-Winters method is 0.7158, implying more weight to recent values. For beta, the result was 1e-04, suggesting the trend component remains constant over time. Lastly, the gamma value is 0.2842, indicating more weight is given to past seasonal values.

* The residual results could be analyzed through different plots.
  + The histogram for the Holt-winter is very similar to the naïve and the simple exponential smoothing. One side of the plot has a perfect bell shape and the other a gap, indicating some missing values were left out of the analysis.
  + In the Acf plot, one of the values slightly crosses the confidence interval, implying the model has not fully captured all patterns.

1. The Arima Method combines the autoregression (AR), differencing (I), and moving averages (MA). It captures patterns and correlations between sales and future sales and considers periods of increases or decreases. The result of the model is ARIMA(1,0,0)(0,1,0)[12]. AR(1), the model uses the previous month’s sales to forecast the next period, and D(1) which removes any patterns.

* The residual results could be analyzed through different plots.
  + The histogram for ARIMA as compared to the other model is normally distributed with a perfect bell shape.
  + In the Acf plot, all the values at the lags are close to zero.

1. **Prediction and Accuracy summary from different forecasting methods**
2. In the Naïve Method prediction, the best single estimate is 34,689 sales units without any uncertainty. The confidence interval of 80% and 95% provides a range of values when there is uncertainty. The wider the interval, the more uncertainty. The accuracy result using MAPE is 7.193973.
3. In the Simple Exponential Smoothing predictions, the best single estimate is 34,519 sales units without any uncertainty. The confidence interval of 80% and 95% provides a range of values when there is uncertainty. The wider the interval, the more uncertainty. For the accuracy results using MAPE are 6.588294. This model performs better than the Naïve method.
4. In the Holt-Winters predictions, the best estimate range is from 39,261 to 43,618 sales units without any uncertainty. The confidence interval of 80% and 95% is a range of values when there is uncertainty. The wider the interval, the more uncertainty. For the accuracy results using MAPE is 4.794672. Performing better than Naïve and Simple Exponential Smoothing methods.
5. In the Arima predictions, the best estimate range is from 39,377 to 48,138 sales units without any uncertainty. The confidence interval of 80% and 95% is a range of values when there is uncertainty. The wider the interval, the more uncertainty. For the accuracy results using MAPE are 3.225937. Performing better than all the other forecast methods.
6. **Decision-based on the analysis**

Arima has the smallest mean absolute error, implying the predicted values are closer to the actual values. Regarding the residual plot, no values have crossed the confidence interval and were normally distributed.

1. **Provide some ideas to improve your forecasts.**

Incorporate events such as holidays or promotions that can impact sales. One example will be a peak during the months of April to June in the year 2023. This data can be excluded or adjusted from the analysis.