DS745: Project One

John Woodward

2023-09-27

Introduction

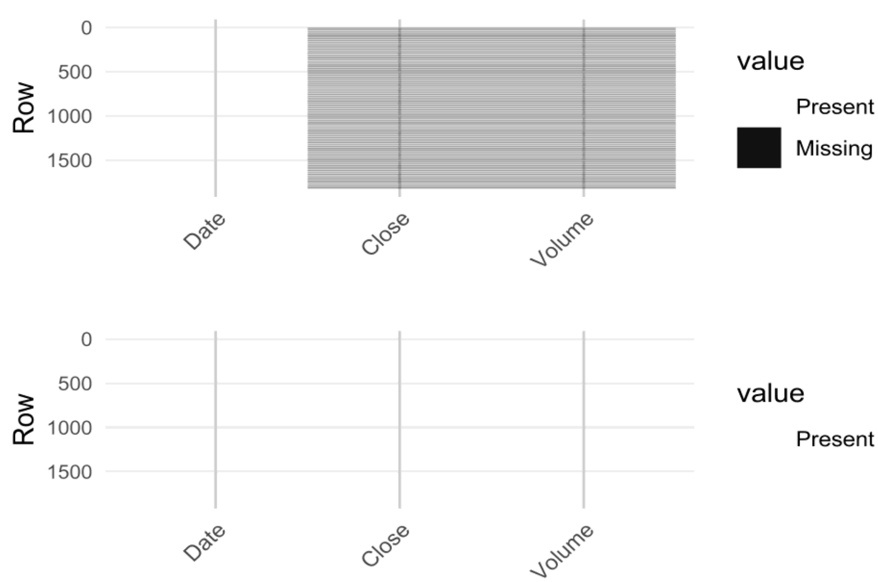
In forecasting, for my project, I will apply the Autoregressive Integrated Moving Average (ARIMA) to five years of GameStop (GME) stock to inform whether I should sell my holdings now or wait until a better time. So, let's embark on this stocky endeavor.

Data Preparation

Although the GME stock data (Yahoo! Finance 2023) is relatively tidy and simple in column structure and size, there are nuances to it ("GameStop (GME) - Stock Split History" n.d.):

1. `Date` - Stock exchange dates do not include all calendar dates by default.
2. `Close` - The stock price per day at the closing time, with stock splits accounted for.
3. `Volume` - The volume of the stock bought or sold daily, with stock splits accounted for.

Figure 1: Missingness before and after imputing

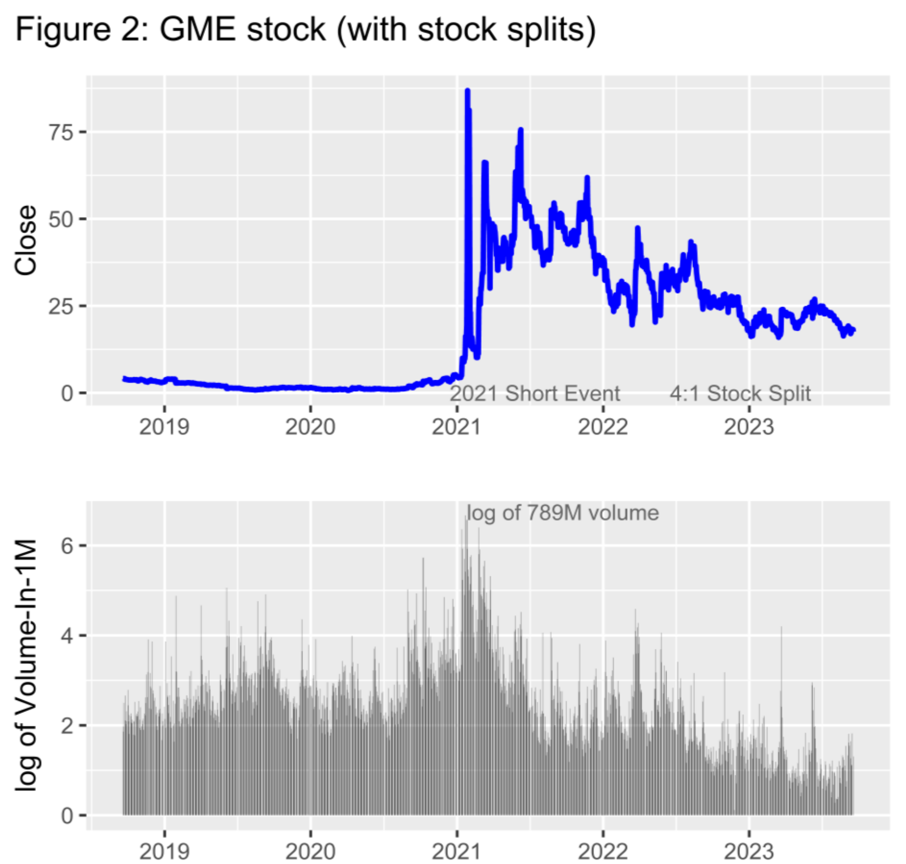


The `Close` and `Volume` have missing dates on Saturdays and Sundays because stock markets close on weekends. Therefore, I imputed `Close` by filling it down as the price remains fixed over the weekend. Whereas `Volume` is typically at 0 during the offline periods when no one can trade at the stock markets, I imputed NAs to 0. Figure 1 is the before and after heatmap of missingness (Woodward 2023).

Next, let's visualize the `Close` and `Volume` time series in Figure 2:

Figure 2 illustrates the extreme price surge in early 2021, which is challenging for prediction models.

Also, in 2022, GameStop underwent a stock split, quadrupling its supply ("GameStop (GME) - Stock Split History" n.d.). Despite adjustments, this event may complicate the model, causing disruptions in 'Close' and 'Volume' data.



Forecasting with ARIMA

In ARIMA forecasting and variable accounting, I included the time series of closing prices (e.g., `Close`), `Volume`, and post-transformed volume. The hypothesis behind this was that volume correlates directly to closing prices.

Hyperparameter tuning/seasonality/customer lifetime value

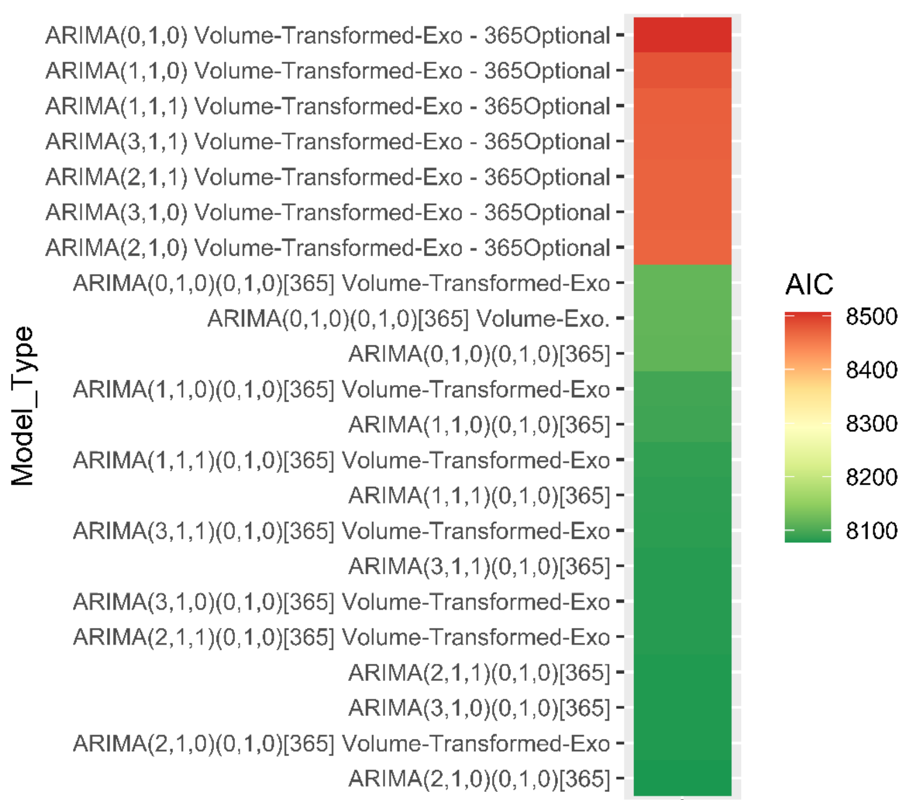
With **customer lifetime value**, I needed to learn more before applying it to data in the project; likewise, applying it in this project did not seem relevant and cohesive, so I decided not to.

However, **tuning hyperparameters** and considering **seasonality** seemed appropriate, given the scope. Therefore, to accomplish both, I used `auto.arima` from R's *`forecast`* library and varied my forecasting ARIMA models between four scenarios: (1) ARIMA; (2) ARIMA with coerced seasonality (Pierre and Kolassa 2016); (3) #2 with exogenous variable of volume; (4) #2 with exogenous variable of volume transformed.

As a result of tuning, I found the best ARIMA model for GME forecasting (Figure 3):

Figure 3 shows that AIC values fall within the 8000s range. While there are subtle distinctions between coerced and non-coerced seasonality, there is generally minor variability among various autoregressions and exogenous variables.

Figure 3: GME ARIMA models

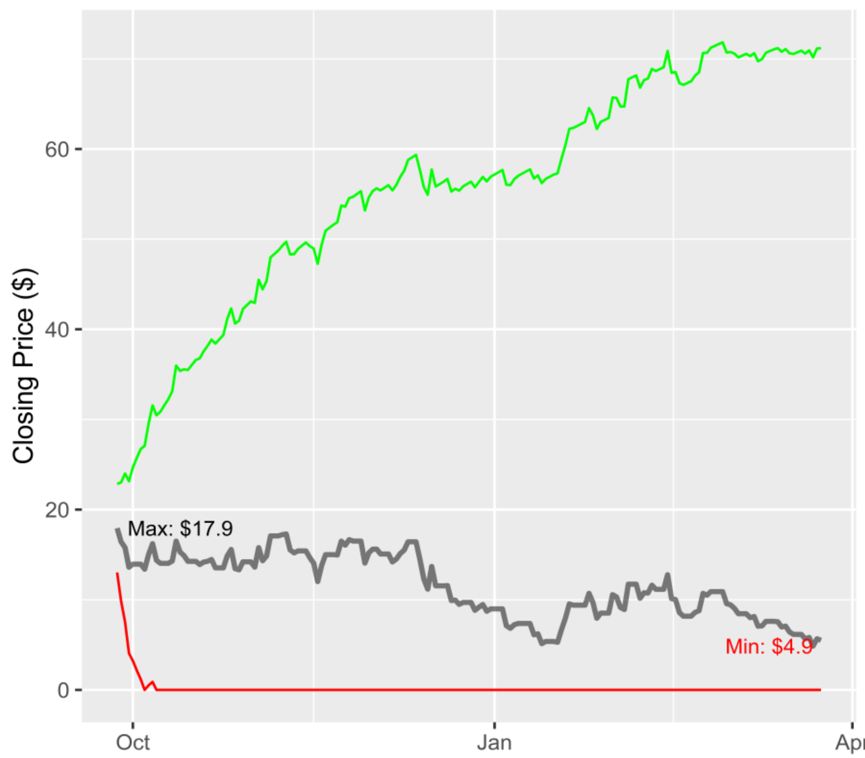


Model-fitting and Conclusions

Due to the high AIC value (e.g., 8000s), our model may not accurately capture the intricate behavior of GME stock. As a result, we should approach its forecasts cautiously.

Figure 4 shows some interesting information about the forecasting model and expected prices in the future:

Figure 4: GME ARIMA(2,1,0)(0,1,0)[365] forecasting



- The maximum `Close` price predicted is $17.90 right now, and the minimum is $4.90 at the end of the forecast horizon of 6 months. This indicates a pessimistic forecast that the price is just rolling downhill.
- Concurrently, the 80% prediction interval (Hyndman and Khandakar 2008) on the lower end broached the negative realm, so I had to cut it off at 0; this adds to a pessimistic tale—conversely, the upper bound paints a maximum of around $70 per share, which seems unlikely.

The prediction model suggests that selling now is a **reasonable choice**. However, if the model isn't reliable and I sell just before the stock gains significantly, I may miss out on potential profits. Forecasting volatile stocks is tricky!

References

"GameStop (GME) - Stock Split History." n.d. CompaniesMarketCap.com. n.d. https://companiesmarketcap.com/gamestop/stock-splits/.

Hyndman, Rob J., and Yeasmin Khandakar. 2008. "Automatic Time Series Forecasting: The forecast Package for r." *Journal of Statistical Software* 27 (3). https://doi.org/10.18637/jss.v027.i03.

Pierre, L., and S. Kolassa. 2016. "Is There a Way to Force Seasonality from Auto.arima." Stack Overflow. May 2016. https://stackoverflow.com/questions/37046275/is-there-a-way-to-force-seasonality-from-auto-arima/37046276#37046276.

Woodward, John. 2023. "Final Project for Ds740: Machine Learning." GitHub repository. https://github.com/jrwoodwa/Masters-Data-Science-DS740-Machine-Learning.

Yahoo! Finance. 2023. "GameStop Corp. (GME) Stock Historical Prices & Data." 2023. https://finance.yahoo.com/quote/GME/history?period1=1537315200&period2=1695081600&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true.