

Module 7. Individual Assignment, Yana James

Ultimately, I have decided to use the time series regression model where I have log-transformed the weekly-sales (y) variable and have not used any lag variables. I have evaluated all of the regression models based on the following criteria and assigned the highest weight to MAE:

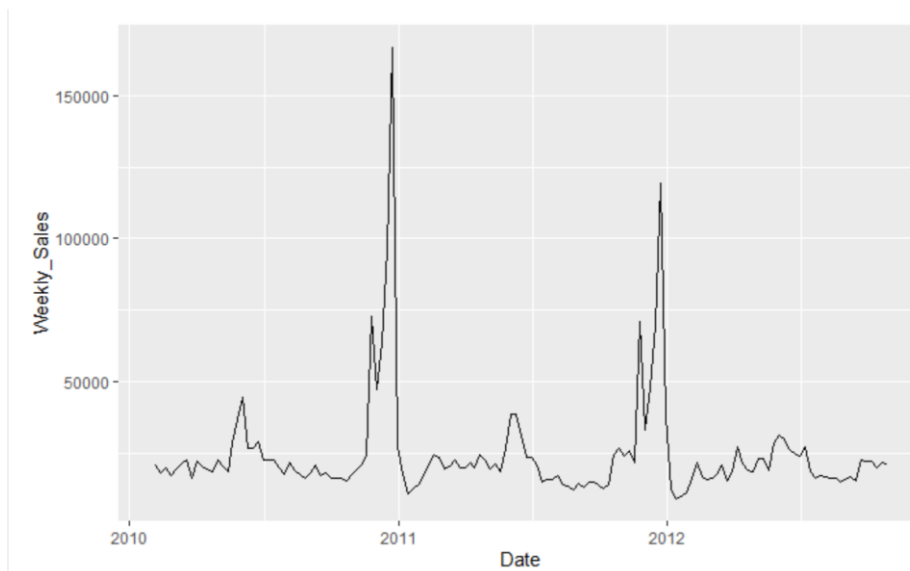
Model Name	Autocorrelation	p-value	R squared	MAE
lm1	-0.228	0.184	0.4722	3735
lm2	-0.151	0.872	0.4848	5037
lm3	-0.182	0.322	0.6749	3768
lm4	-0.150	0.592	0.66	2970

It was clear that in order to capture the seasonality both the month and the fact if it was a holiday week needed to be included in the model. To capture the trend the dummy variable of the week number was created. Since my first model was not the best, I have enhanced the second model with the lag variables for the past 7 weeks but that did not improve the performance. Log-transforming the y variable and removing the log variables made the mode more accurate and reduced the MAE to under 3,000.

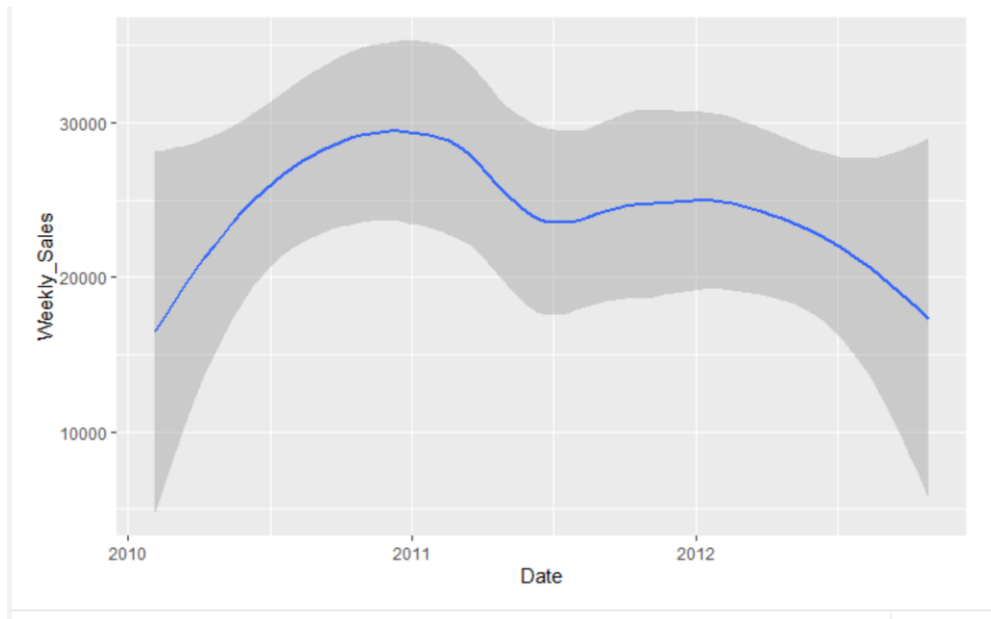
I was not 100 confident in my R skills to create the new DF for the upcoming 52 weeks, so I have used the alternative way of plotting my predictions for the next 52 weeks. I have used the TBATS method which is recommended when dealing with weekly seasonality as opposed to the Holt-Winters methods (which I still tried and received pretty high MAE of 4707 for the additive method and 4313 for the multiplicative).

Overall, it is clear that we have a slightly downward trend with really high sales around Thanksgiving/Black Friday and Christmas and rather small spike around Easter. The TBATS model predicts a steady increase in sales around end of November 2013 to end of December with average sales of 90,000 and max sales of 140,000 and min of 80,000. The model also predicts higher than normal sales around Easter of 2013 (close to 40,000, as high as 50,000 and as low as 25,000).

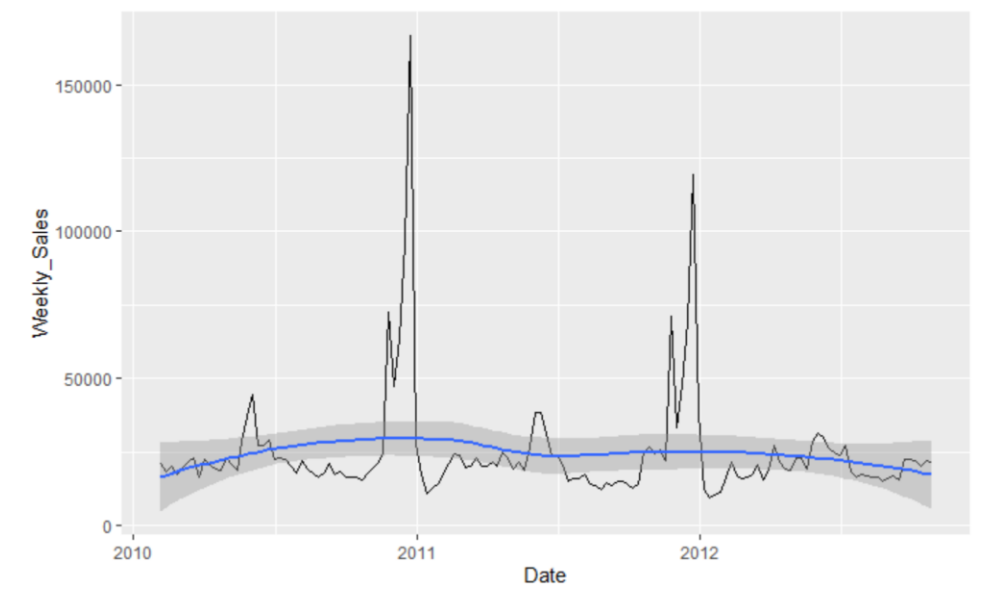
Weekly Sales per Year



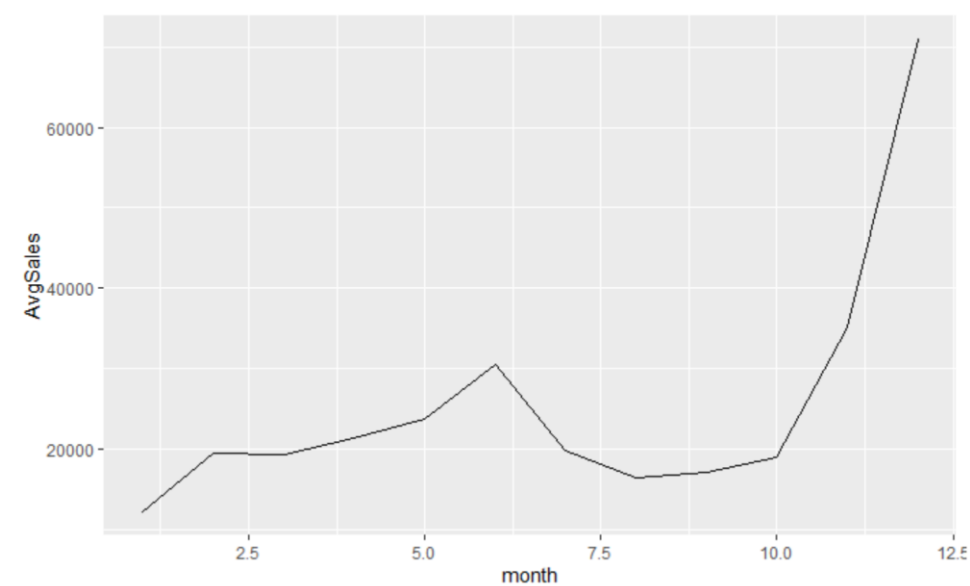
Weekly Sales Trendline (smooth)



Weekly Sales per Year and Trendline (smooth)



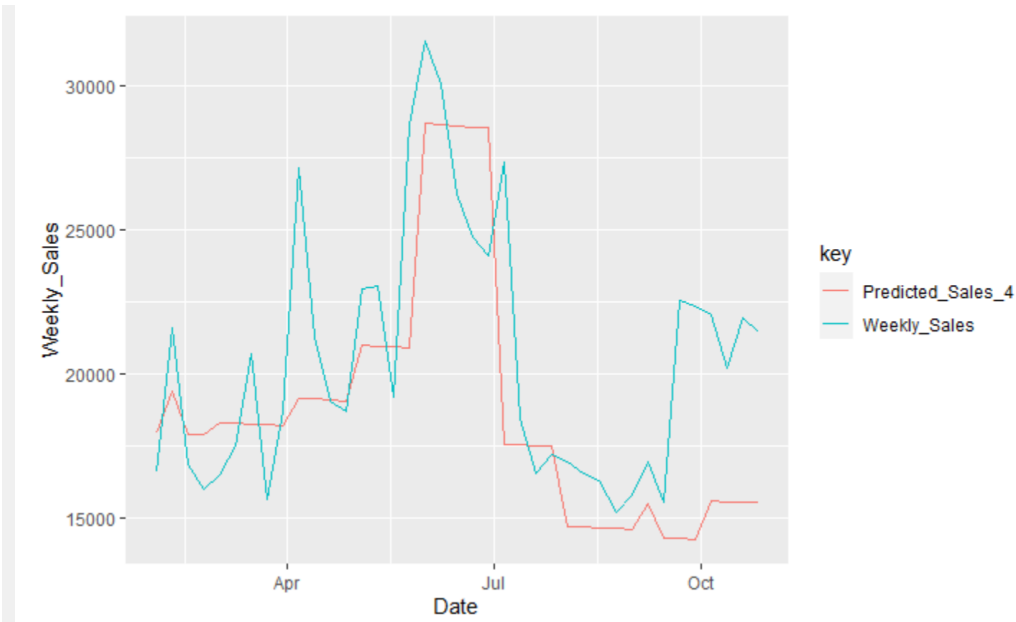
Average Sales per Month



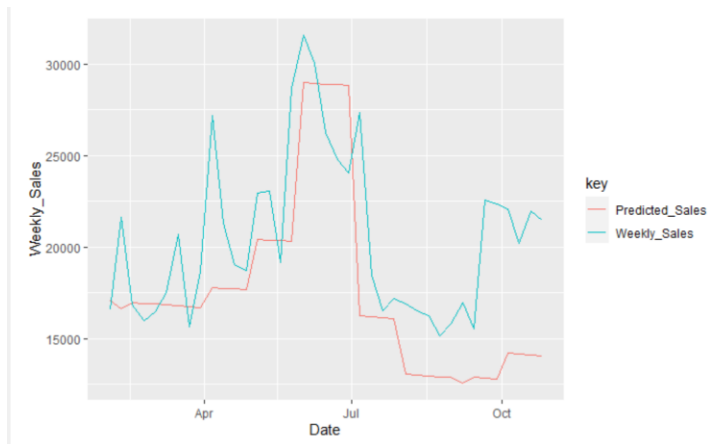
Average Sales by Holiday

IsHoliday	AvgSales
FALSE	23995.02
TRUE	32167.06

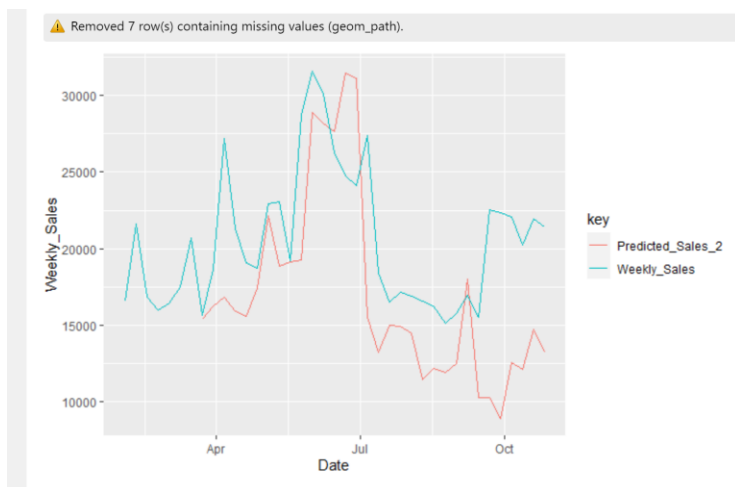
Best Time Series Regression Model – MAE 2970



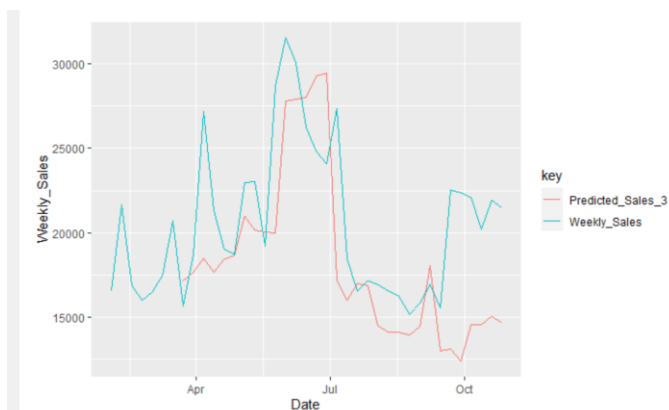
First Time Series Regression Model (lm1) – MAE 3735



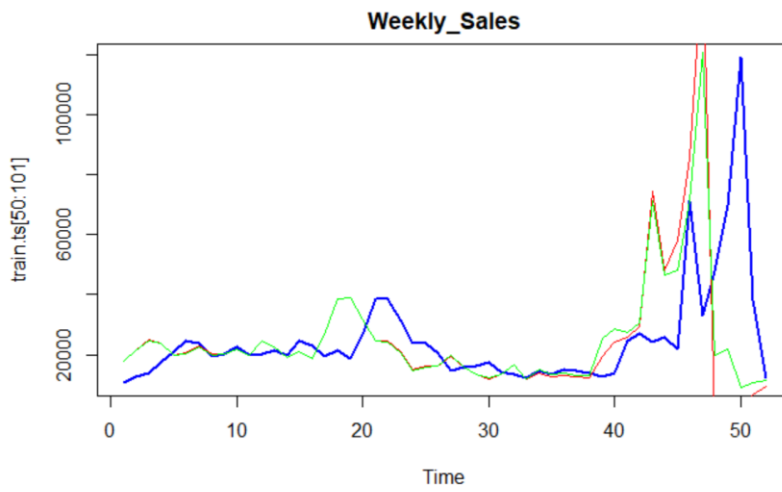
Second Time Series Regression Model (lm2) – MAE 5037



Third Time Series Regression Model (lm3) – MAE 3768



Smoothing Models (MAEs: Additive 4707, Multiplicative 4313)



Weekly Sales Forecast via TBATS method

