FCAS136 | A

2.25L Soda Promotion Space Sales Prediction

Rene Bravo
Jeremy Haber
Henry Izbicki
Logan Langenhuizen

Group 11

25 December 2012

Executive Summary-

The profitability of a supermarket is largely determined by efficient product display within promotional bins and shelf displays. Hence it can be argued that a sustainable competitive advantage can be achieved if demand of particular products could be accurately forecasted to allow for a customized supply of consumer demand, ultimately leading to increased profitability. In this project, we focus on one of the most basic and yet diversified products in terms of sizes within supermarkets – soda beverages. Hence, our basic business objective is to develop a more targeted sales strategy of soda bottles by leveraging the power of forecasting and predict the demand of soda bottles as to "when" soda bottles of a particular size should be supplied for an optimized equilibrium of demand and supply. Once defining the basic business objective, a first and particularly important step is an efficient data preparation. From the initial data sheet, we aggregated the most commonly consumed soda beverages to come up with four different container sizes, ranging in volume from 300ml to 2.25l. While preparing the data for forecasting models, we paid particular attention to the power of visualization, which helped us to improve and optimize our data and involved processes such as the identification of level, trend and seasonality, the identification and exclusion of outliers caused by mystery bulk buyers, the choosing of a weekly sales prediction, the focus on 2.25l sized containers and the exclusion of "noisy" weeks 49-57, all leading towards an optimized data input allowing for accurate future predictions. It has to be noted that in addition to the visualization and great effort the group put into the data preparation, external advice was also gathered from a Indian based retail consultant, giving us further insight of how we can optimize our data preparation to synergize our forecasting expertise with the initial business problem. Using this data, a partitioning was conducted under best practice, creating a training and validation period which was used for all forecasting model developments. The group saw the Holt-Winter's to be the most fit in order to capture the identified level, trend and possible seasonality but also ran a variety of other models such as linear regressions, double exponential smoothing and a naïve forecast, serving as a benchmark to compare the different measurements such as MAPE, MAD and MSE. Further, we also analyzed possible autocorrelations of residuals and incorporated any significant ARs into our models if found necessary. It turned out that the group's initially used model, namely the additive Holt-Winter's model gave the lowest of all three measures above (64.7, 37.8 and 2332.2 respectively) and beat the strong performance by merely using naïve forecasts from prior weeks. In other words, the improved performance of our model can be effectively used by interested stakeholders of this Hypermarket: The data shows that the model can predict future demand more accurately, leaving room for a wide range of applications such as the establishment of a threshold that controls when 2.25l containers should be displayed at certain promotional stands if a certain level is exceeded. However, it has to be noted that despite the model has created more than positive results, additional data over time needs to be gathered in order to further refine and improve the model to meet the dynamic demand of the 2.25l soda size containers at Hypermarket and sustain this possible competitive advantage.

Project Objective-

“To predict large volume soda sales to meet promotion supply requirements.”

Within this study, it is empirically proven that soda sale varies by time of the year. In addition, these demand patterns vary by different types of soda containers, typically ranging from 300ml to 2.25l volume containers to capture different market segments and customer demands. Hence, knowing customer buying patterns based on volume would ensure sufficient supply and targeting opportunities for supermarket chains such as Hypermarket. Therefore, accurate prediction of customer demand can indicate store profitability and customer satisfaction. An independent retail consultant based in Chennai further contributed to the power this prediction could have: He has indicated that aisle spaces and promotion bins are the key areas that determine profitability for retailers in India. This is due to government imposed price caps (MRP). Optimizing the display of both, would therefore arguably ultimately lead to increased profitability.

We assume that demand is a reflection of consumer preference across all selling points. So an accurate model of sales quantity will allow for future forecasts. Combined with the constraints of supply issues regarding products in India, our predictions will allow retailers to better time promotions to capture market share away from others. This is essential to business given the lower revenue margins at Indian retailers as an affect from government regulated price maximums. The tradeoff for a retailer is promotion allotments for margins with the brands and distributors. The result from this balance is that a retailer’s profitability is largely determined by efficient utilization of aisle and promotional space, implied by heavy reliance on sales per square meter as a key performance indicator.

By accurately forecasting the allotment mixture of soda on shelves the retailer will be able to manage inventory, shelve space, and marketing campaigns more effectively. Our forecasts can be used to determine whether a promotion should be applied that week, should expected demand surpass an established threshold. We chose to forecast the demand of 2.25L soda bottles since they are most likely to be used for promotion according to our external consultant. Our inclination was that soda sales would be relatively stable. Since visualization indicated possible seasonality and trend, we would chose models such as Multiple Linear Regression and Holt-Winter’s to capture the entire signal. This led us to begin compiling our assumptions, which include stable prices, promotions and existing supply and demand. Our forecasting horizon was set to one week after learning that the replenishment frequency for soda at hypermarket retailers was of that frequency.

Data Preparation

Starting from the original hypermarket data set and after defining the business problem, we filtered the hypermarket data to include only subclass “Aerated Beverages.” Next, we selected the top 8 selling brands of soda, which included Coke, Fanta, Sprite, Thumsup, 7 up, Pepsi, Diet Coke, and Miranda. We needed to clean the data and extract bottle size from the Item_Description field, thus creating a new field: “Size.” The next step was to standardize size records, and group them into 4 size categories: 300ML, 600ML, 1.25L, and 2.25L. It is important to note that soda brands may not have 4 different sizes (i.e. Diet Coke is only sold in 300ML and not 2.25L). This however didn't affect our prediction because the next step was to aggregate data based on size on transaction date.

After aggregating drink size on daily transaction date, we visualized the data based on quantity sold and noticed interesting results from 2.25L. First, we noticed spikes on certain days (see exhibit 1). We considered 5 days outliers, did a deep dive into the data, and found single shoppers purchasing many 2.25L bottles of the same brand on the same day. Before excluding these shoppers we visualized the data by week and also noticed our data to be affected (exhibit 2). To see data by week we added a week# column, in addition we added semi-monthly, monthly, and DayofWeek columns for further analysis and visualization. We noticed larger variability in sales and too few data points, thus concluded not to predict on a semi-monthly or monthly basis. (see exhibit 5-6)

We further noticed weekly seasonality on weekends, and ran a model to predict next weekend sales, but then asked the retail consultant if promotions only last a weekend, which turned out to be not often. Most of the time promotions run for a week. Thus we chose to predict weekly sales but kept weekend dummy. In running our model later, the weekend dummy proved to be insignificant.

Dummy variables were created on the brand field, showed no influence on predictions during the analysis.

Last to note, the team decided to ignore the 300ML, 600ML, and 1.25L size data series. Upon asking our industry expert, 1.25L and 2.25L drink sizes are most common in middle aisle promotions, however we noticed 1.25L data to be uninteresting in recent months (see exhibit 7), possibly drink suppliers pushing 2.25L sales more and customer preferences changing to 2.25L.

Before proceeding to the models, we deleted 49-57 weeks (see data issues).

Data Issues

The team decided to exclude outliers from large purchasers of 2.25L bottles of the data (exhibit 3). We assumed these purchasers were small shop owners, buying in bulk and not the "typical customer." Although these customers may return in the future, we will reassess the model based on occurrence of stockouts resulting from these shoppers. The timeframe of the data provided, however, was too short to assess a possible reoccurrence of these shoppers.

In addition, we noticed summer months (week 49-57) to be quite volatile and difficult to predict. The team decided many reasons could cause the volatility including bad tracking of data, data recording mistakes and change in consumer behavior that is unpredictable with only one year of data. Thus, weeks 49-57 were excluded and instead focused on predicting until week 48 of 2.25L quantity sold (exhibit 4).

As a small note, we also noticed week 57 was not a full week because it excluded the weekend, thus week 57 was excluded from the data in earlier analysis in order to ensure consistency.

3. Methods

Partitioning

The data was partitioned according to our forecasting objective which was to predict the following weekly sales of 2.25l soda bottles. After aggregating the top selling soda brands (coke, diet coke, Pepsi, diet Pepsi) we then categorized by weekly total sales up until the summer months (June/July). Our validation period was set to week 48 and our training period is week 1 through 47.

Models

Before running any models for the forecast we first looked at the visualization of the weekly 2.25l sales (exhibit 2). From the visualization we noticed that there is trend, level and possible seasonality in the data. We chose the holt-winters multiplicative and additive formulas to test for both seasonal and non-seasonal trend. Multiple linear regression and double exponential smoothing were chosen to compare accuracy against the holt-winters models testing for trend and level without the apparent seasonality seen in the visualization.

In the linear regression model we ran a regression of the total sales on the weeks 1 through 47 to see if our model would be best fit adjusting for the peaks and troughs by capturing the level and trend of the data (exhibit 8). After running the regression we then ran an auto-correlation on the residuals of the forecast and found that the period before (lag-1) was significantly correlated to the current period. The model was then ran again capturing the auto-correlation in the linear regression and analyzed against the results of our other models.

Holt-winters multiplicative and additive models (see exhibit 9 and 10) were used to see if the seasonality could be captured accurately in the model. We ran the formula using .2 alpha for faster learning level and .2 beta to give enhanced adjustment to the most recent given data. When choosing how many seasons were in our data set we examined the daily sales data and noticed weekly seasonality with increasing sales on the weekends. To account for the weekly seasonality we used a period of 1 totaling 47 seasons in our training data and 1 in our validation set.

Double exponential smoothing was used in our comparison to capture the level and trend of the data using a learning formula that can adapt with the volatile data thereby giving more significance to the most recent information (exhibit 11). Also, double exponential smoothing would include the noticeable increasing trend in the last few months of our data.

4. Performance

We determined model performance based on a comparison of errors against a benchmark and against each other in both the training and validation periods. A side-by-side comparison chart (see below) was generated for the smoothing methods while multiple linear methods are scored independently. We found a smaller MAPE for the additive method of Holt-Winter’s compared to multiplicative. This indicated to us that perhaps there was no multiplicative pattern and instead we are seeing a more aggressive trend. Our RMS errors were better for the Linear project after conducting an autocorrelation and identifying a lag-1 trend.

Training		Linear Lag-1	Holt-Winter's(A)	Holt-Winter's (M)	Double exp	Naïve
	MAPE		64.6863	80.6512	84.24	85.0124
	MAD		37.8271	41.1321	38.98	46.1739
	MSE		2332.2210	2913.0736	2663.74	3272.0870
	RMS	50.5422				50.545226
	AV ERROR	0.00000179				0.00000178
	TSS	117521.7125				150516
Validation						
	MAPE		9.69104	10.93994	6.700	8.2353
	MAD		26.74727	30.19423	18.51	21.0000
	MSE		715.41626	911.69130	342.800	441.0000
	RMS	50.9749				21.0000
	AV ERROR	50.9749				21.0000
	TSS	2598.9017				441.0000

We used a naïve forecast for our benchmark across all methods. The naïve performed very well giving almost equivalent predictions to the linear lag-1 since we found a strong AR signal from the prior week. However, a smoothing method was able to beat the naïve in the training set on every deployment. Within the validation set, the naïve was able to beat all methods. We expected this for the any simple smoothing method because it provides equal weight to data regardless of age. Compared to linear regression, we found a lot of noise in the data such that additional forecasts could correct for any direction bias from a single measurement.

For each method, we plotted the residuals to visually inspect for missing signal. Once we incorporated the lag-1 element into our models, we found the residuals to behave more as noise.

5. Conclusion

We successfully created a model using the Holt-Winter's method that beats the Naïve forecast and allows the retailer to predict when 2.25L sales will be high enough to put on a promotion. Hence, it can be argued that by using our model, the Hypermarket could create a sustainable competitive advantage if utilized correctly. However, in order to sustain the competitive advantage, a critical evaluation is needed as outlined in the following:

Our model works for weeks not including June/July while it is constrained by a lack of additional data. Additional data collection on a week-by-week basis will allow us to determine if the model should be rolled forward or only used to predict a single large period. We suggest using a different model for the excluded weeks. Moreover, the given timeframe of merely one year did not allow us to see any inter-annum seasonality but merely for intra-annum seasonality.

Given that seasonality might not be present within the weekly pattern, we would be interested in using both a simple smoothing method such as moving average and a learning method such as Holt Winter's for future testing.

Appendix

Exhibit 1 – Daily Quantity Sold by Size w/ Outliers

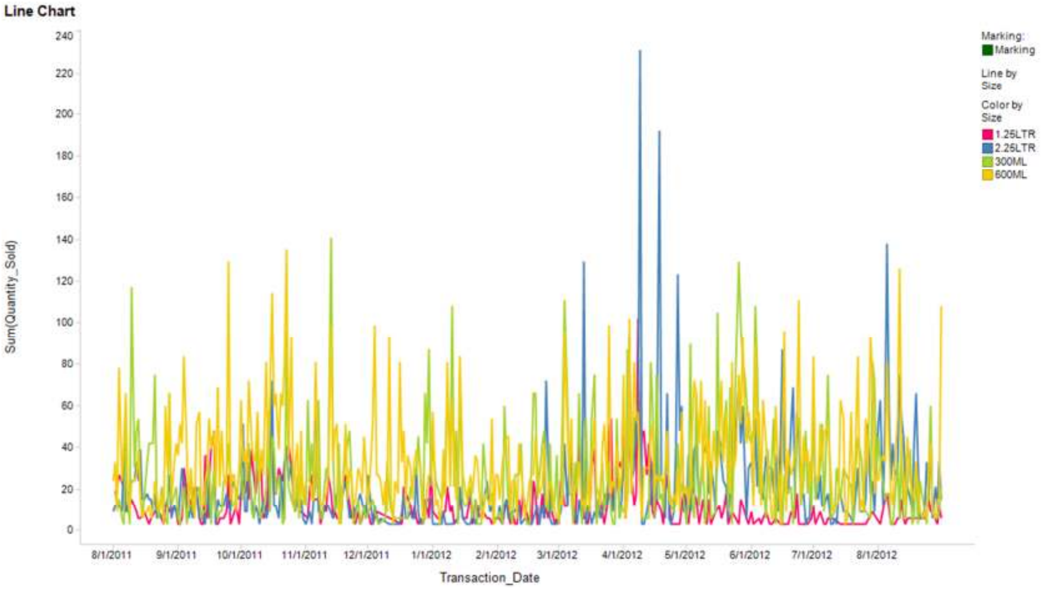


Exhibit 2 – Weekly Quantity Sold of 2.2L Bottles w/ Outliers

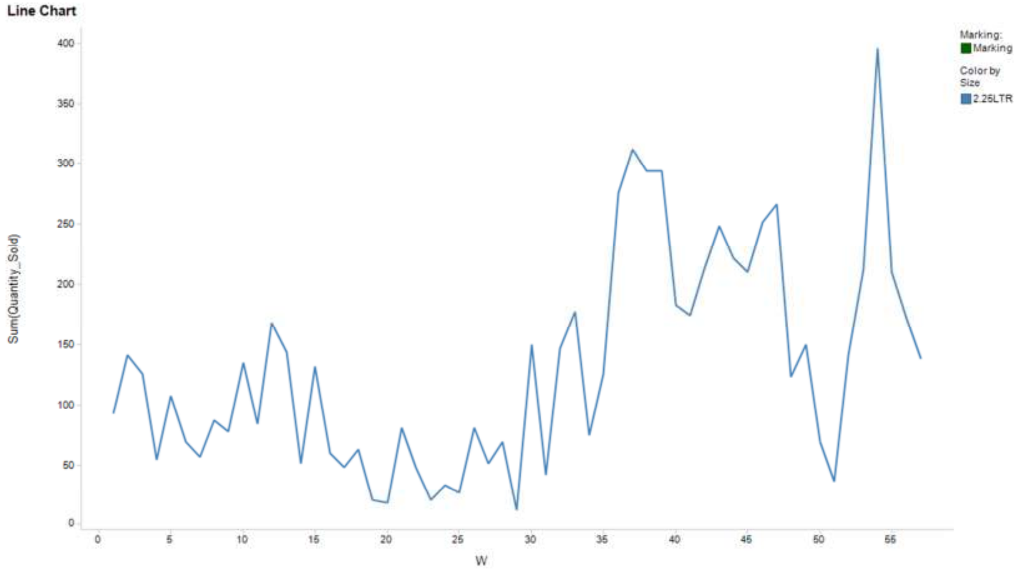


Exhibit 3 – Weekly Quantity Sold of 2.2L Bottles w/o Outliers

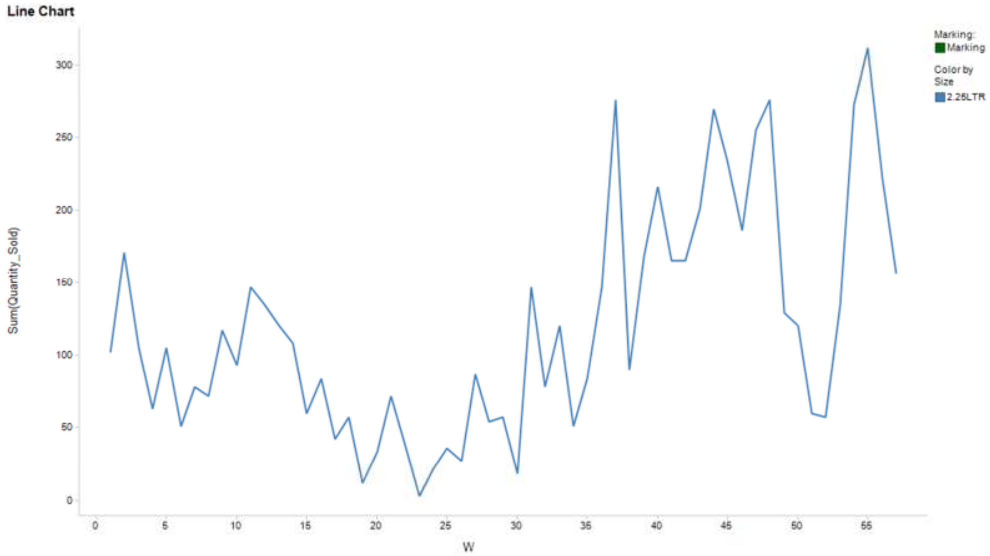


Exhibit 4 – Weekly Quantity Sold of 2.2L Bottles no summer (predict week 48)

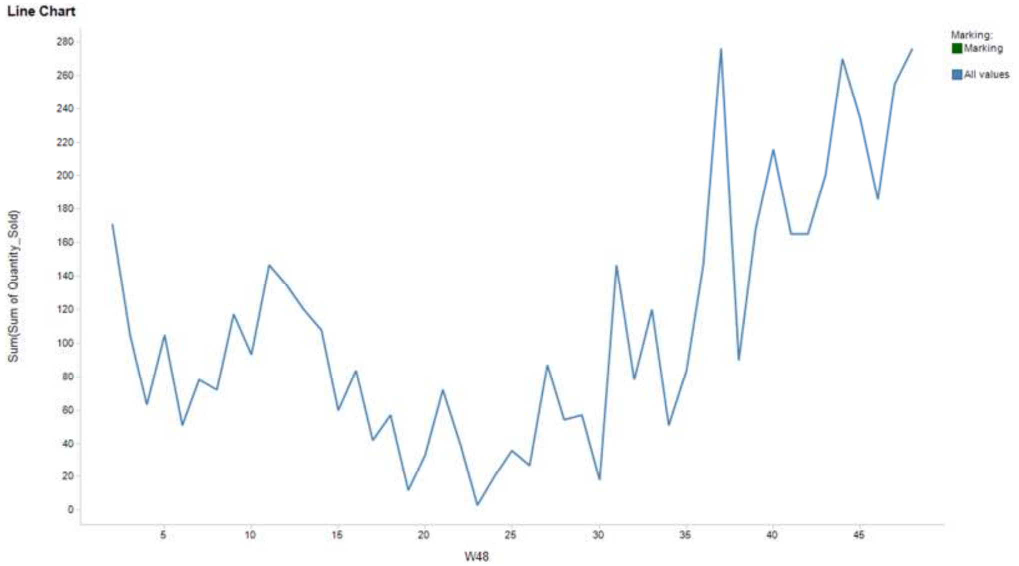


Exhibit 5 – Semi-monthly Quantity Sold of 2.2L Bottles

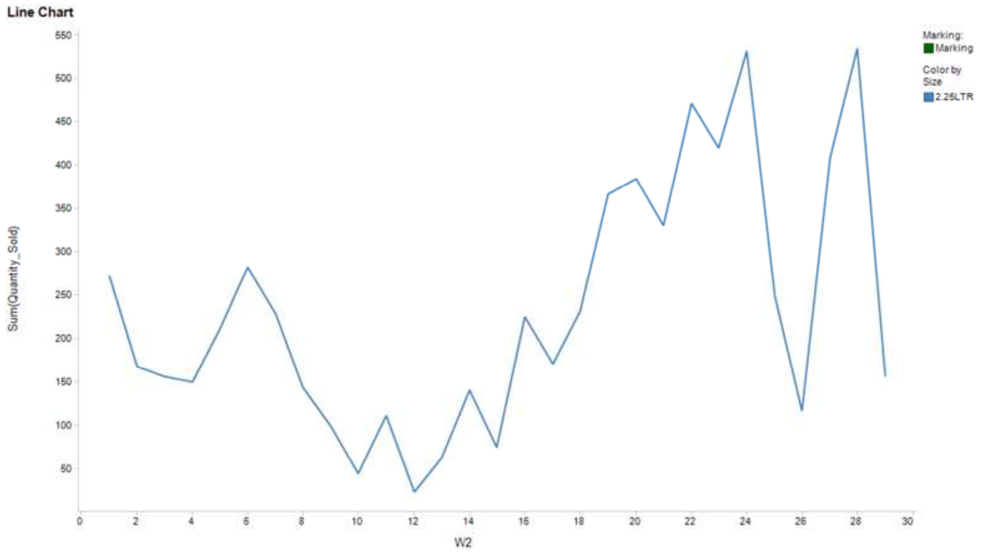


Exhibit 6 – Monthly Quantity Sold of 2.2L Bottles

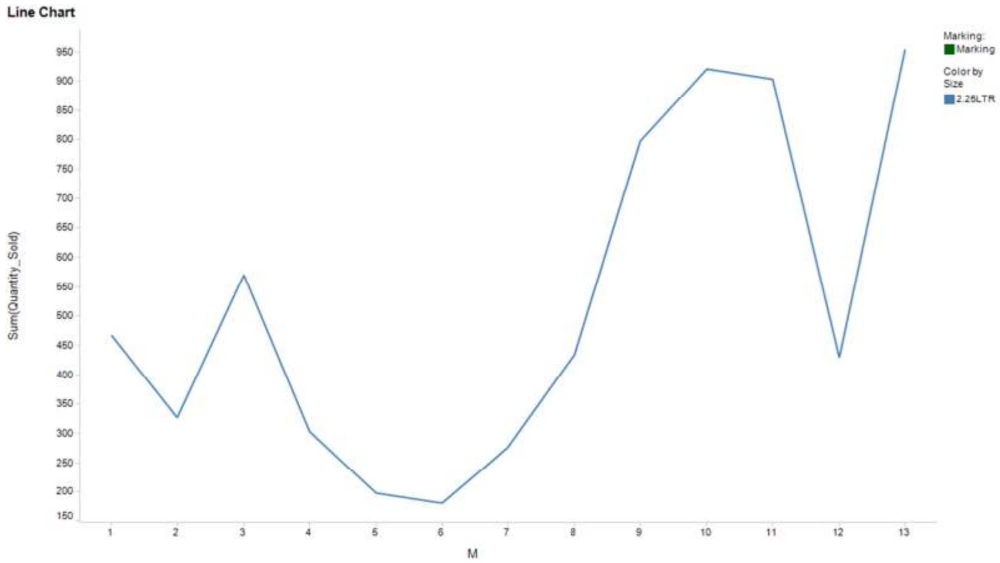


Exhibit 7 – Weekly Quantity Sold of 1.25L Bottles

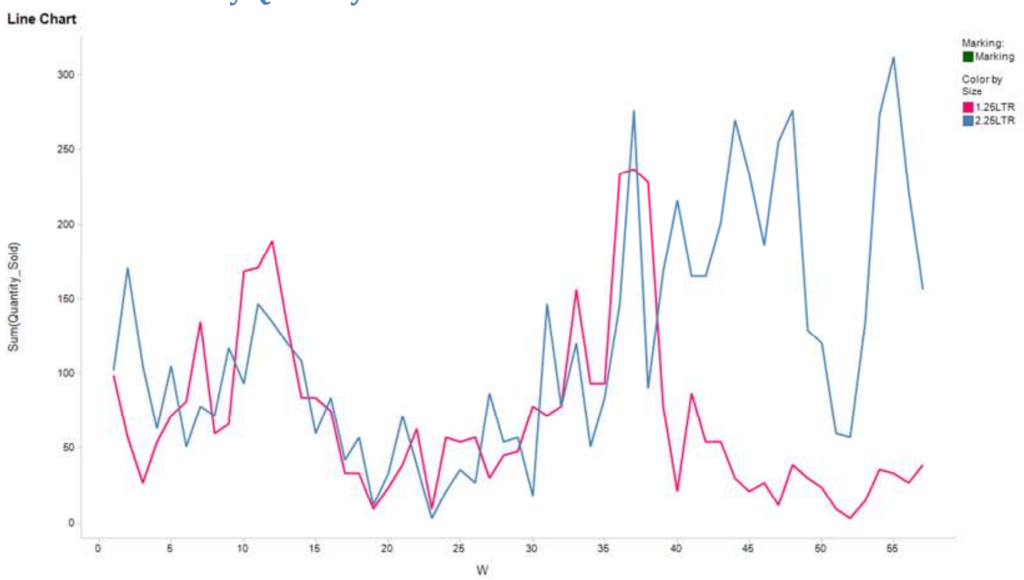


Exhibit 8 – Linear regression Lag t-1

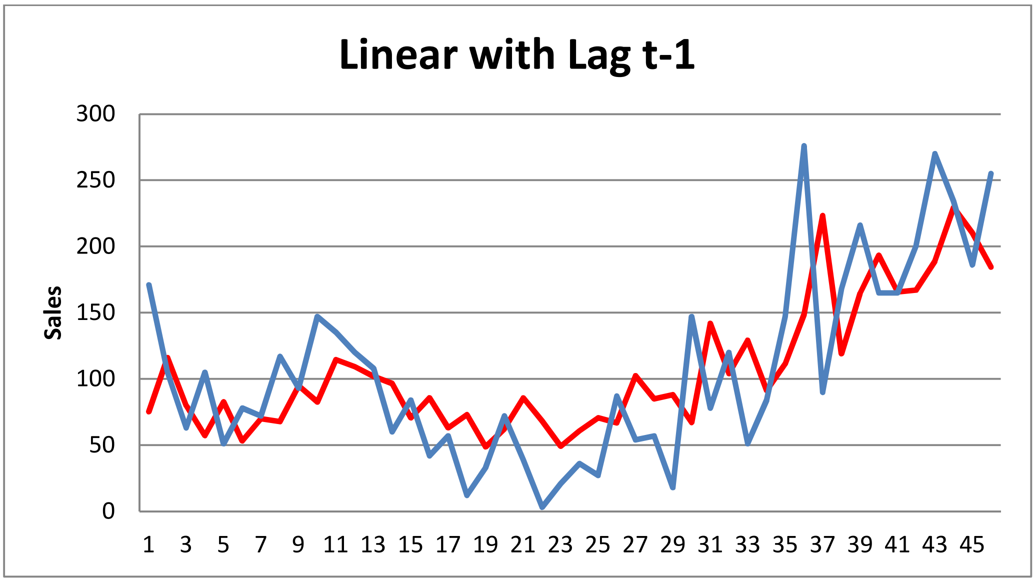

Linear with Lag t-1
Sales
300
250
200
150
100
50
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45

Exhibit 9 – Holt-Winter's Multiplicative

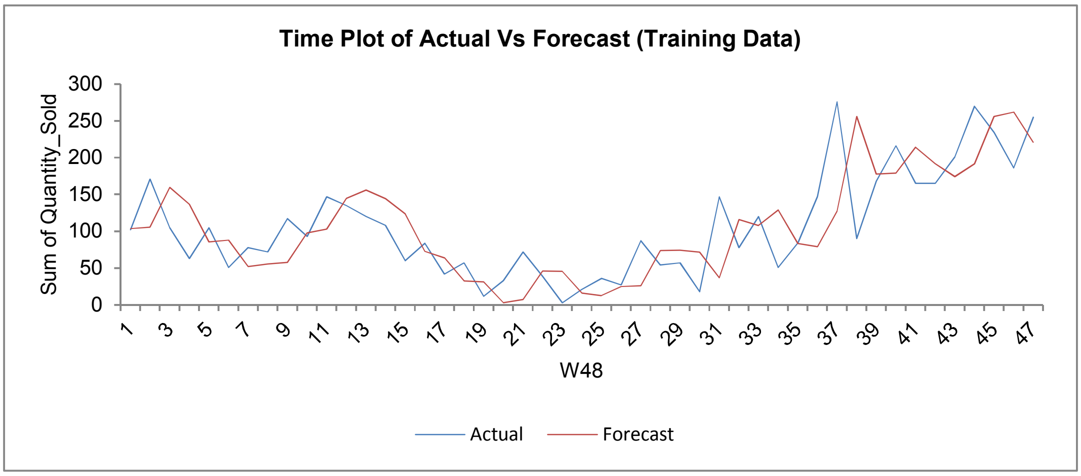


Exhibit – 10 Holt-Winter's Additive

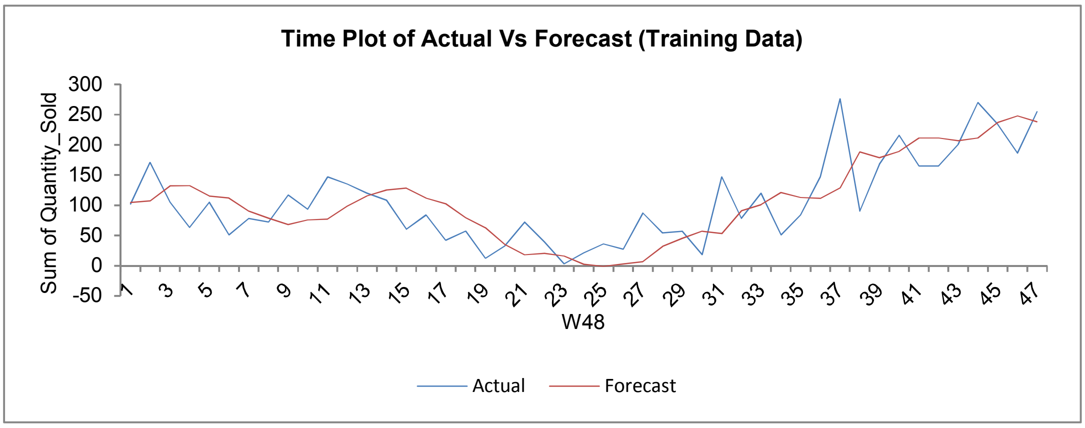


Exhibit – 11 Double exponential smoothing

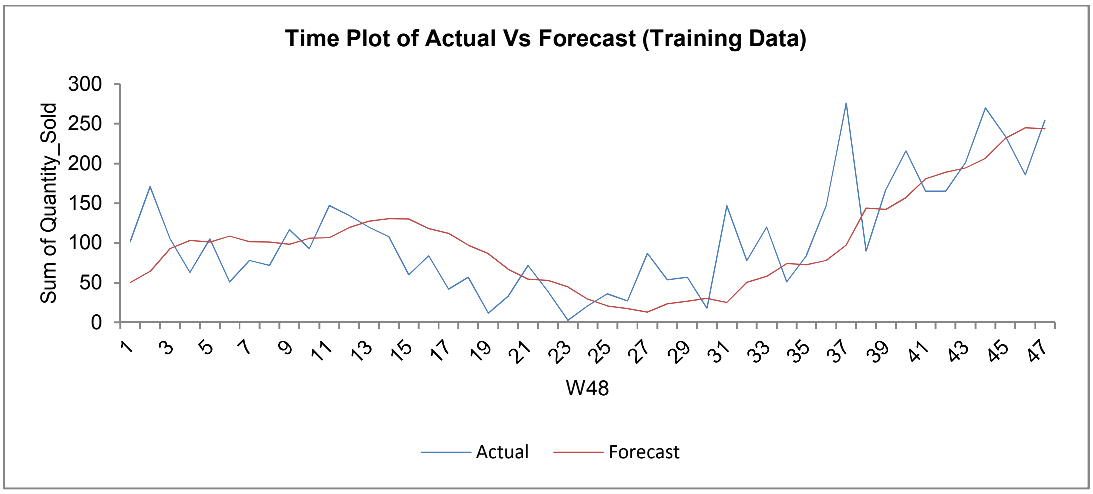
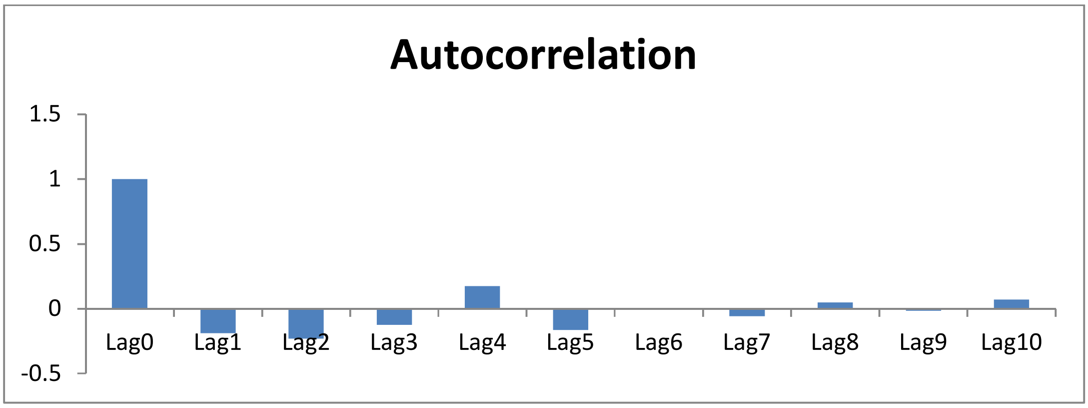


Exhibit – 12 Autocorrelation on Holt-Winters additive


Autocorrelation
1.5
1
0.5
0
-0.5
Lag0
Lag1
Lag2
Lag3
Lag4
Lag5
Lag6
Lag7
Lag8
Lag9
Lag10