OPEN IIT DATA ANALYTICS

Out of stock prediction And Buffer Stock Estimation

INTRODUCTION

Out Of Stock : A stock-out, or out-of-stock (OOS) event is an event that causes inventory to be exhausted. While out-of-stocks can occur along the entire supply chain, the most visible kind are retail out-of-stocks in the fast moving consumer goods industry (e.g., sweets, diapers, fruits). Stockouts are the opposite of overstocks, where too much inventory is retained.

Buffer Stock : Buffer stock corresponds to the supply of inputs held as a reserve to safeguard against unforeseen shortages or demands. Buffer Stock help to reduce the negative effects (stock-out costs) of an unusually large usage of stock. Buffer stock schemes seek to stabilize the market price of products by buying up supplies of the product when stocks are plentiful and selling stocks of the product onto the market when supplies are low.

PROBLEM STATEMENT

- **Identifying Out Of Stock situations** - In a retail chain consisting of 'n' number of stores and 'x' number of products, possible out-of-stock situations are to be identified using real world data.
- **Buffer Calculation Logic** - The best algorithm is to be found out for the calculation of buffer/safety stock in order to reduce the lost sales opportunity occurring due to the above mentioned out-of-stock situations.

CHALLENGES

A variety of challenges were faced while preparing this report. For example,while computing buffer stock, the time interval after which the inventory is renewed , better known as lead time,is an important unknown factor , so various approaches for calculating buffer stock which included the use of lead time had to be skipped.
Furthermore most of the products provided were not fast moving, So it was hard predicting the Out Of Stock Occurence.

OUT-OF-STOCK IDENTIFICATION

- **Poisson Distribution Approach**

This is what we want to get to, a rule like. “If you should have sold more than X units (and you’ve sold nothing) call an off-shelf alert.

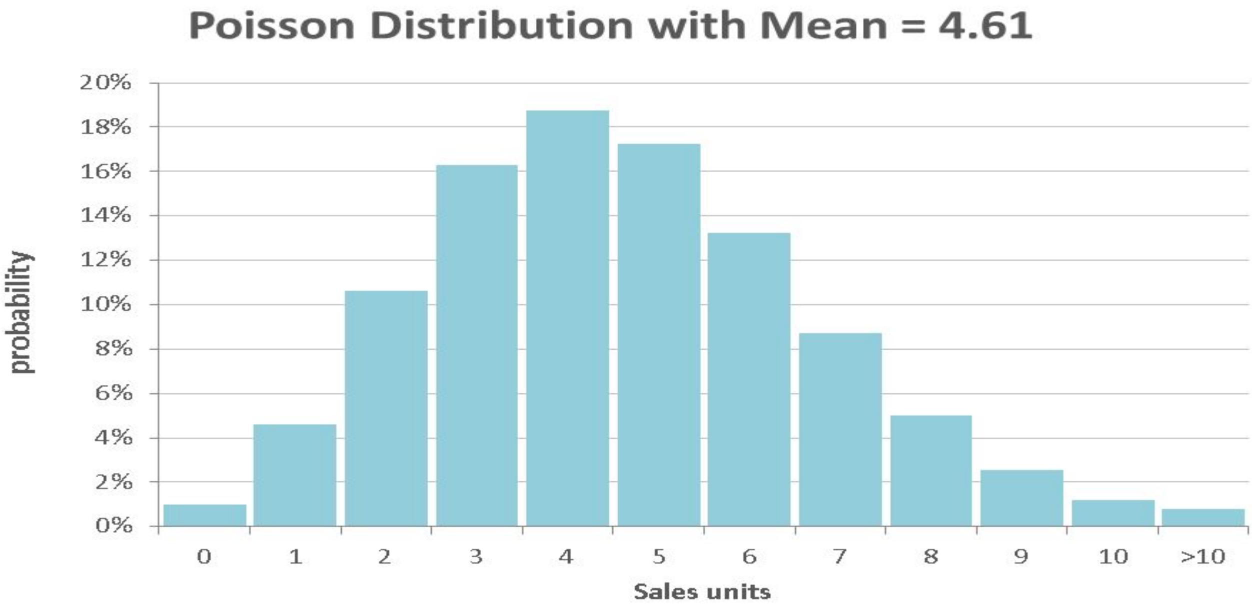
Let’s assume we have sold nothing now for product X at the target store for 3 days and that prior to this point our average sales over a 3 day period is 2.3 units. This is what a Poisson distribution looks like for average demand of 2.3 units:

It shows the probability of actually getting a sale of 1,2,3, through 10 units in the 3 day period. As we can see, there is only a little over 25% chance of selling exactly 2 units and about a 20% chance of selling 3 units. It’s also possible (but very unlikely) that you could sell 8, 9, 10 or even more than 10 units.

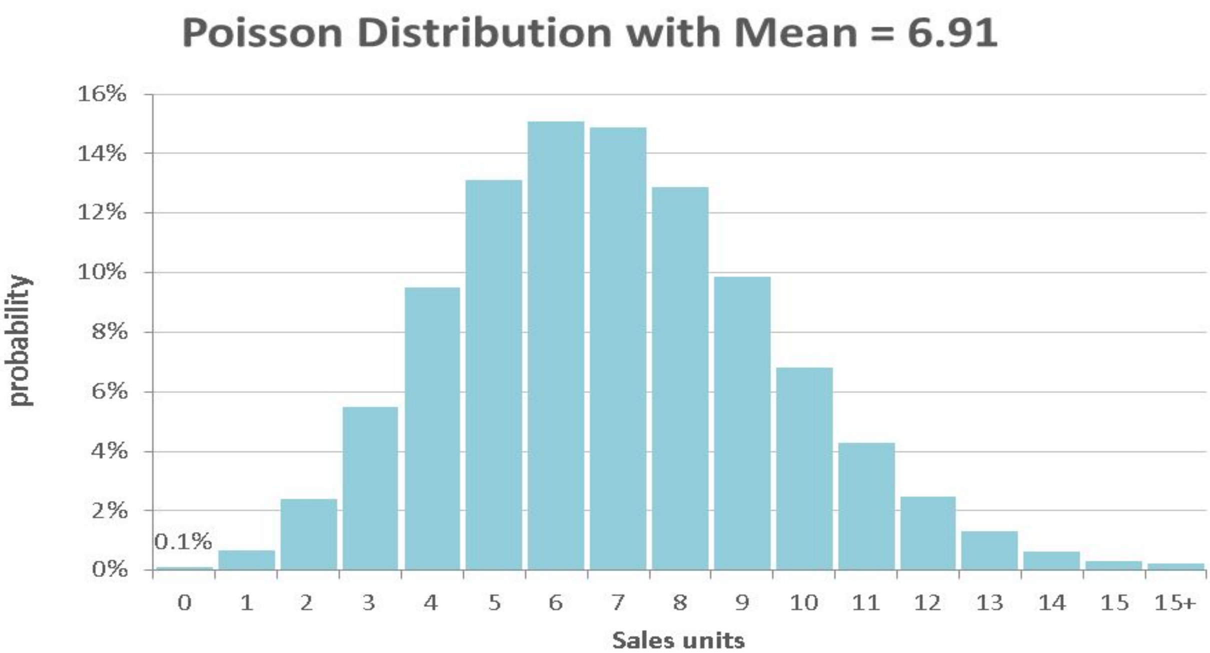
If this distribution is a good representation of reality, how odd is it that we actually sold nothing at all in the most recent 3 day period? This tell us that seeing no sales is going to happen about 10% of the time just based on random chance. We probably don’t want to call out an off-shelf alert with such a high chance that nothing is wrong.

When, a few days later, we reach the point at which you should have sold 4.6 units, and have still sold none, the probability of actually selling nothing through random chance is now just 1%.

Ignoring errors in data, your estimation of average sales or your assumptions (perhaps it’s not a Poisson) you will be wrong about 1 in 100 times.



To get more security, we wait a little longer. At the point that we should have sold 6.91 units, there is only 0.1% chance that the zero sales you are seeing is due to random chance: far more likely in fact that there really is some issue inhibiting sales at the shelf.Waiting helped us gain accuracy but it also cost us in lost sales.



Combining this approach with time segment average we built a model and termed three threshold values as per experimental data:

Harsh: Will generate more off-self alerts, on even a little of sales dwindling.

Easy: Will generate less off-self alerts, only in extreme cases

- **Our Model**

We considered the average sales of previous three weeks into account then as per its poisson distribution we calculated the probability of getting an out of stock.

 If it came less than the threshold then it is surely an out of stock situation.

 If it is not less than the threshold but is near to it, we term as not out of stock but add the lost sales number in check for the next point of sales data and so on.

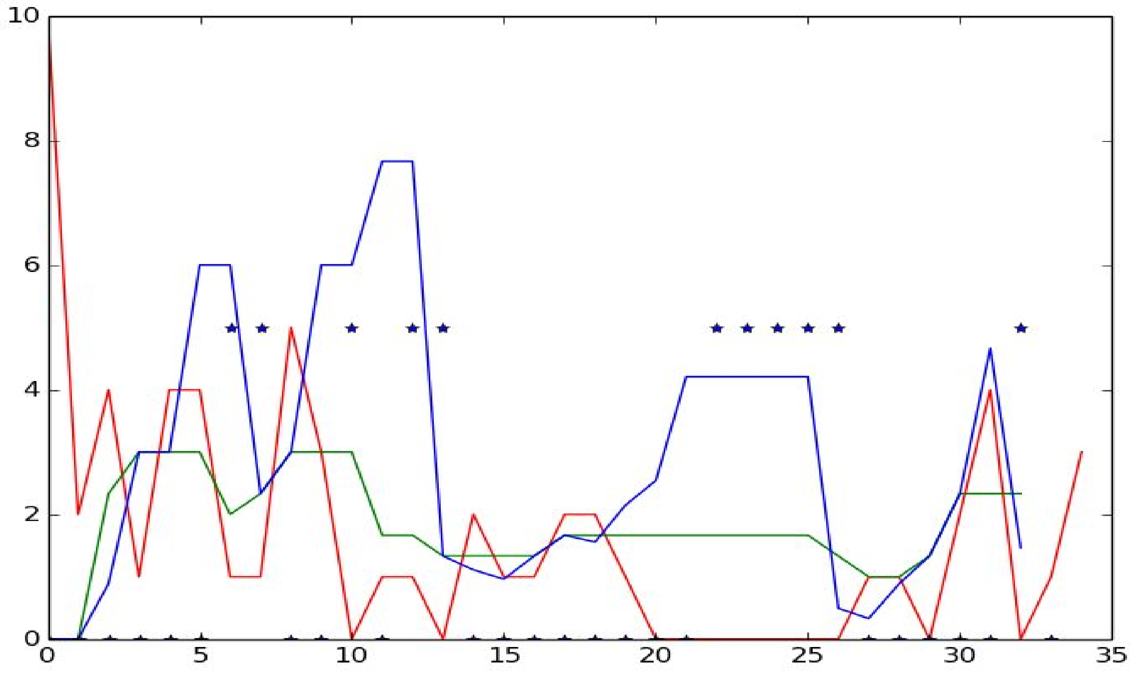
 If it is far greater than the threshold, we term it as not out of stock.

Then to make the model more realistic we added a smooth normalizing function depending on number of consecutive out of stock days.

We used **func(x) = e^(-x/100)**

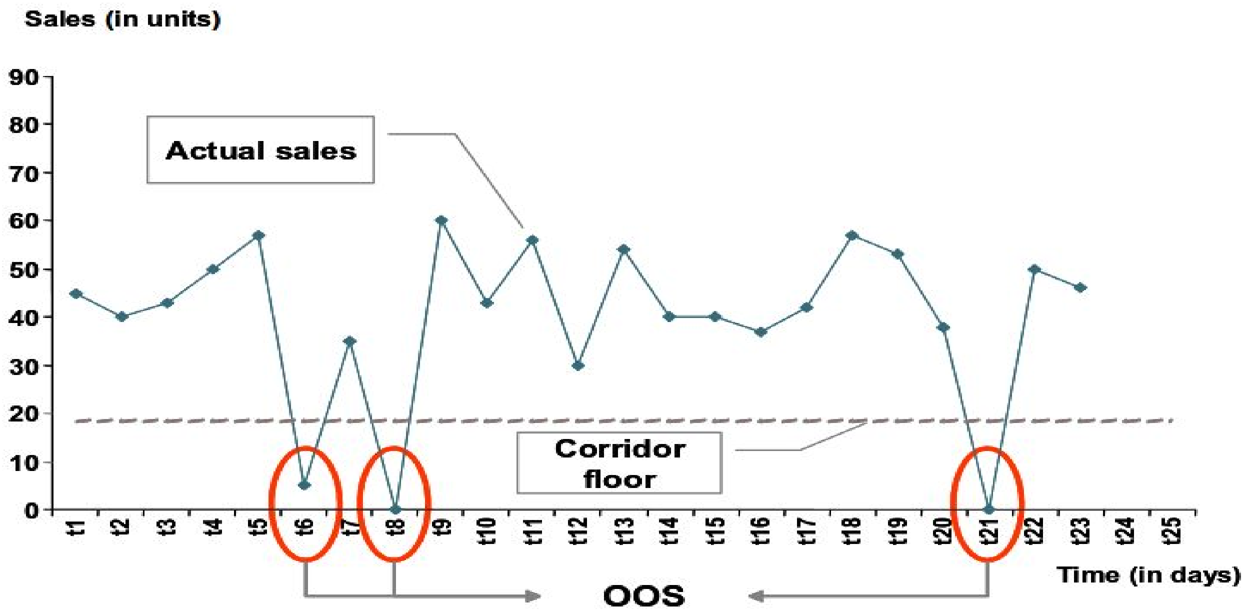
Where x is the consecutive out of stock occurence.

Because more the number of days with out of stock increase more the mood of customer will turn away from the product. So this model took into account the seasonal factor and mood of customers both.

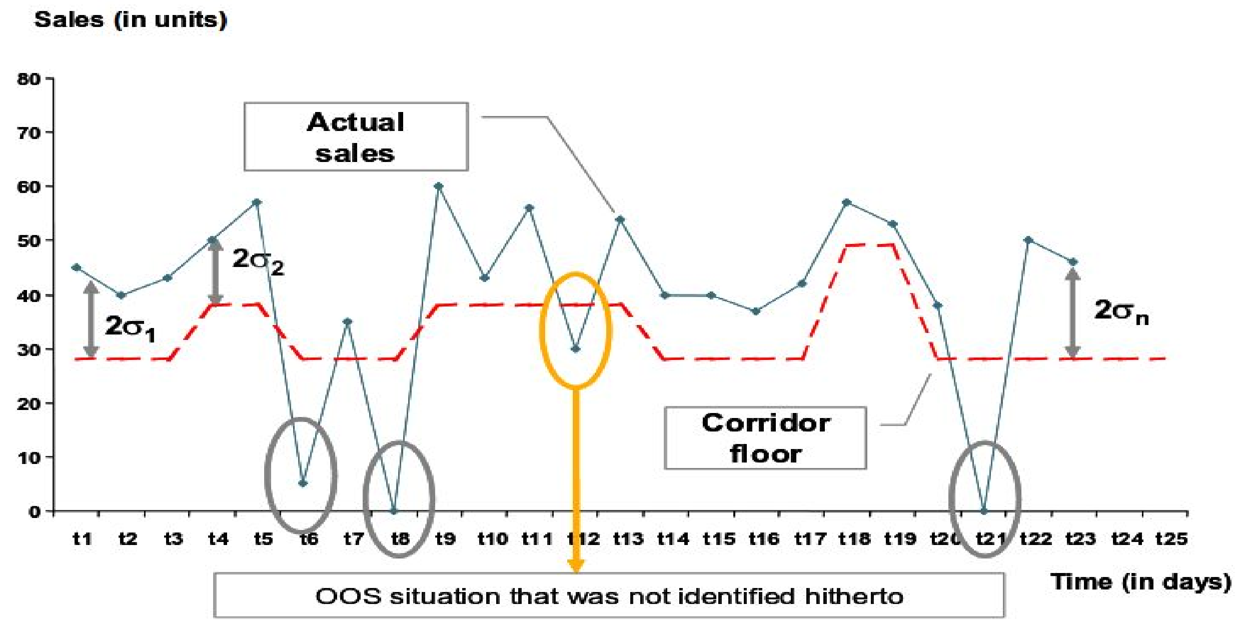


- **Mean and Standard Deviation of Sales Approach**

On the Sales vs Time graph, we plot a Line curve of (mean +– Standard Dev of sales). We get a lower bound and upper bound as per that. Firstly we were doing it over all the data provided(Total Time). There this model was useless as it did not take into account the seasonal factors and gave a very bad result, like the one used below:

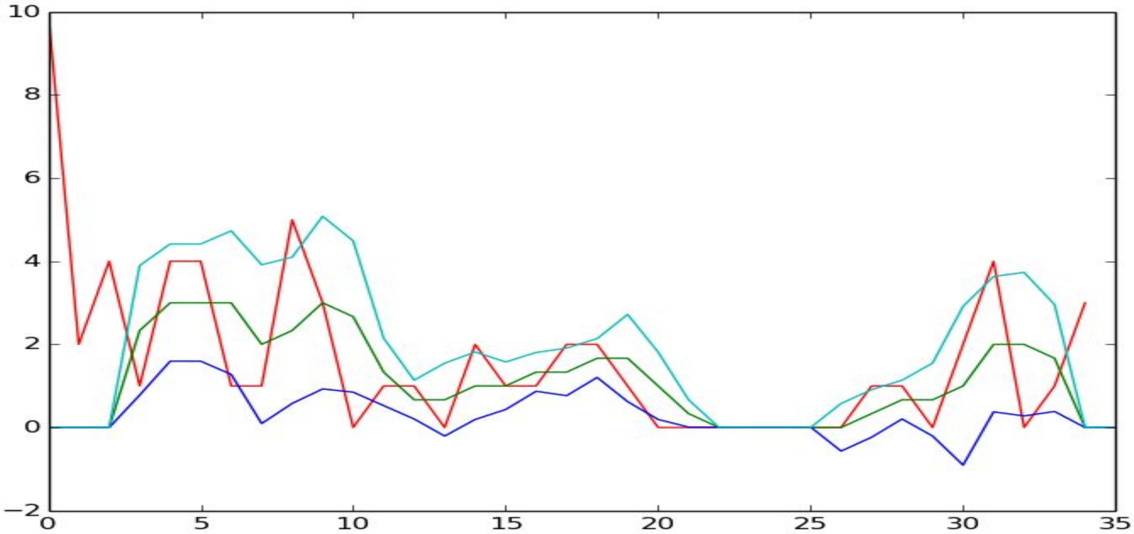


Then we calculated the same on a time segmented basis, like the mean of previous three sales and standard deviation calculated accordingly. This covered our data strongly, like the one used below:



Now if the point-of-sales data is less than the lower bound at that point we can term it as out of stock(OOS) and if the sales is more than the upper-bound at that point we can term it as an unusual sale. It could be due to some seasonal demand which can further be helpful for buffer stock calculation.As the demand rises unusually it is more likely that our stocks and buffer both get depleted before our previous calculated safe-zone.

Here we see how this approach not only solves the problem of OOS situation prediction but also warns us about possible increase of Sales opportunity.



In the graph above :

- The Red Line is actual point of sales data.
- The Green Line is each time segment's average of point of sales.
- The Blue Line is average - standard deviation.
- The Green Line is average + standard deviation.

We can see how beautifully the model covers the data, predicting Out of Stock intelligently.

It works better than previous model and at **real time** it predicts Out Of Stock more responsibly.

BUFFER STOCK CALCULATION

The buffer stock is calculated using the following formula:
BS = σ*Z
where BS = Buffer Stock,
Z = Safety Factor,
σ = Standard Deviation of Demand

We see in the formula that we need to calculate the standard deviation of demand, but we cannot conclude the actual demand (or lost sales) from the data given to us. We know predicted sales is the quantity that is forecasted by the company as a demand using previous data. So we have assumed the predicted sales given in the case to be the demand rate.

Safety factor is a measure of how much service level we want(in how many cases stockouts are avoided). For each percentage there exists a value for safety factor calculated from normal distribution. Higher the service level higher is the safety factor. However, applying this model caused a lot of overstocking situations and the correct safety stock was not predicted.

Initial Model: When our sales exceed the predicted sales(when the sales is found to be greater than the estimated demand) , we need to take buffer stock in account to satisfy the customers demand.Let us define the Buffer error function as :

$$\sum (PS - S - BS)^2$$

where PS = Predicted Sales,
S = Sales,
BS = Buffer Stock.

Target: Minimize this cost function to get the optimal buffer stock We observed that, for more accurate buffer stock calculation we should define this buffer error function for a short period of time(seasonally) to consider the seasonal and trends. For example in October sales are expected to be higher because of Diwali Season rather than in normal days. So we need to decide the safety factor accordingly.

Method: We use the gradient descent model. Gradient descent is a first-order optimization algorithm. Convergence means that the result is not going to change significantly if you continue. The above instructions generalize to an arbitrary number of dimensions. Let's suppose we want to fit a straight line $y = w_1 + w_2 x$ to a training set of two-dimensional points $(x_1, y_1), \ldots, (x_n, y_n)$ using least squares. The objective function to be minimized is:

$$Q(w) = \sum_{i=1}^{n} Q_i(w) = \sum_{i=1}^{n} (w_1 + w_2 x_i - y_i)^2 .$$

The update of parameters for this specific problem will become:

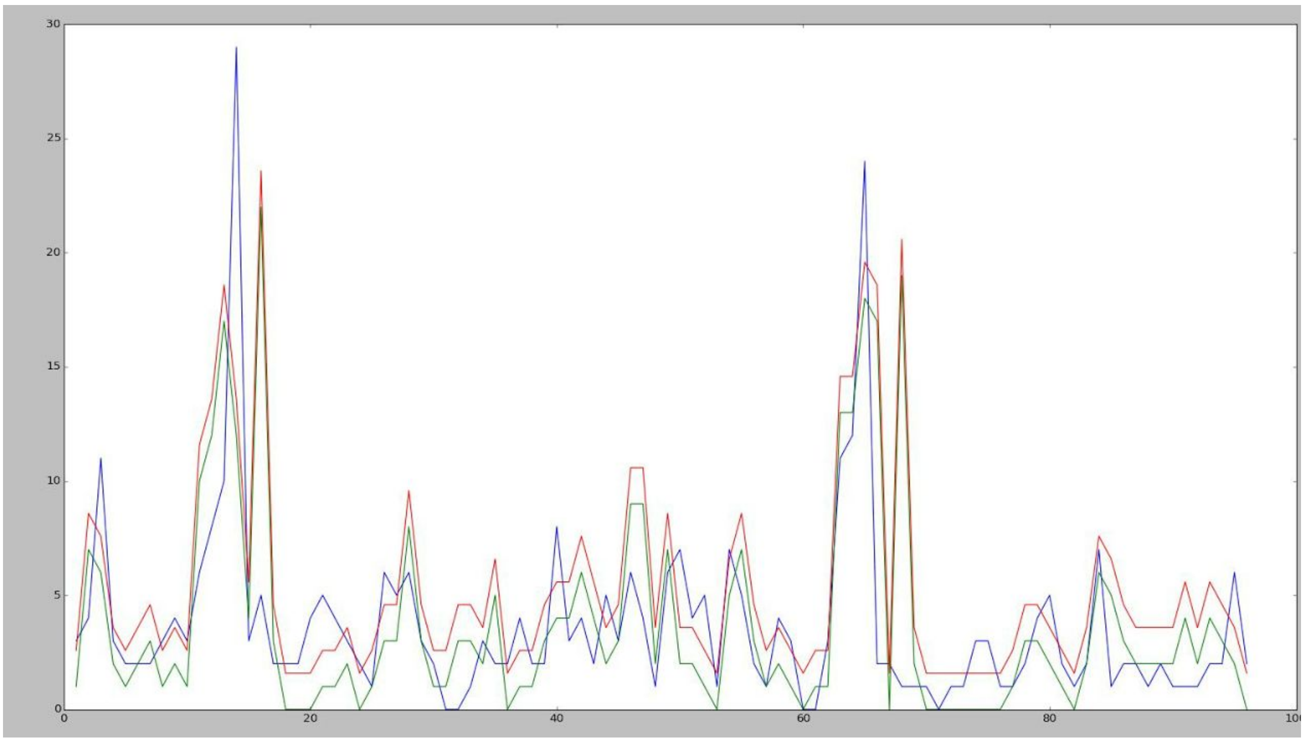
$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} := \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \eta \begin{bmatrix} 2(w_1 + w_2 x_i - y_i) \\ 2x_i(w_1 + w_2 x_i - y_i) \end{bmatrix}.$$

Our function is defined by a set of parameters, gradient descent starts with an initial set of parameter values and iteratively moves toward a set of parameter values that minimize the function. Buffer error is function of safety factor Z. The Cost Function here thus becomes:

Q(w) : $\sum (S - PS - BS)^2$

To minimize this cost function we update the values of variables in every iteration and again calculate the value of buffer error for the new value of Z. This is done till a constant value of error is achieved or for a certain number of iterations specified for a certain service level.

Here Z = Z - α*∂BE/∂Z.



In the above graph, the blue line represents sales data, the green line represents predicted sales data and the red line represents the sum of predicted sales and buffer that we calculated.

This graph pertains to a particular product ID - 15664 and City - C.

RESULTS

After applying the first model which used Poisson Distribution approach we got 38.06% Out of Stock situation while when we used the Second model which used Mean - Standard Deviation Approach we got 12.35% Out of Stock situations.

The approach used for buffer calculation when applied to the given data, the number of units which we were short by is equal to 13576(11.70% of total demand). If the buffer is taken to be zero then out of stock units is 21234(18.29% of total demand). So we see there is improvement of 36 percent.

DISCUSSIONS

Even though Poisson Distribution Approach strictly predicts Out of Stock and takes into account Probability of different Sales, But it can not be termed realistic as it keeps on predicting Out of Stock even if there is no sales for more than a month. While the Mean - Standard Deviation Approach intelligently predicts Out of Stock situation as per real trends.

Even though both the models take into account the seasonal trends as we are calculating average over previous three sales but we get an edge while we count on Standard Deviation Approach as it makes **Two Corridors :**

- One for Out Of Stock Prediction (Blue Line in Our Case)
- Second one predicts an upsurge of unusual sales, And possible gives a warning for Buffer Inventory Stocking.(Cyan Line in Our Case)

Furthermore, If a train data had been provided It would have been easier to create a more precise model.