Introduction:

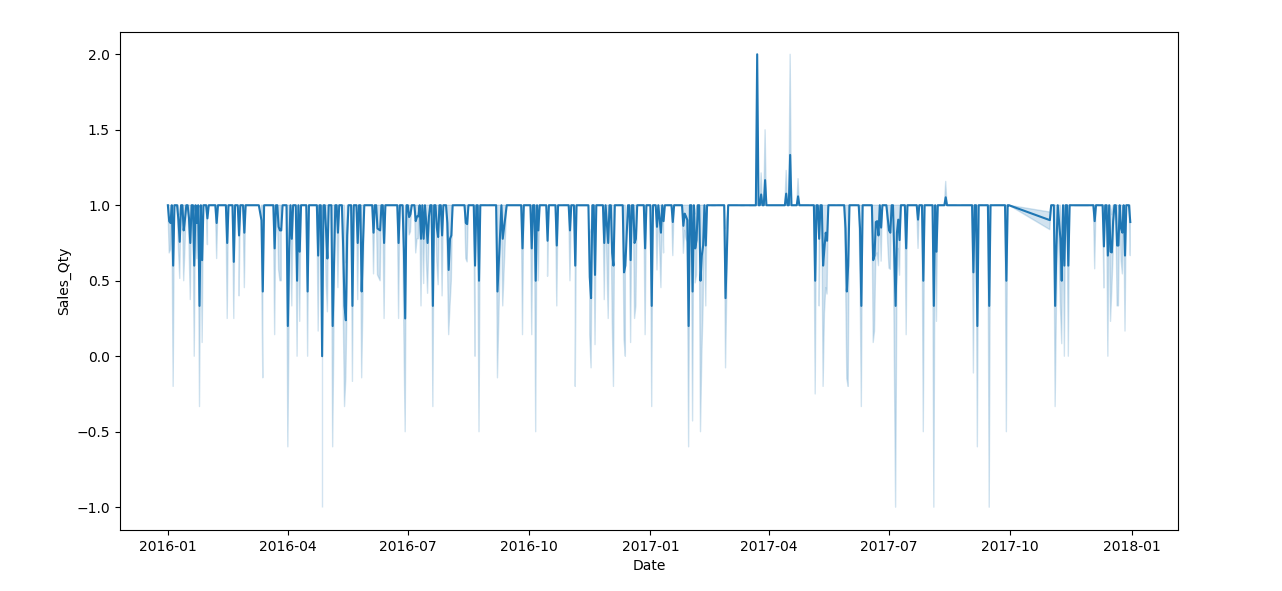
**1. Aggregate the Sales\_Qty for each Store-SKU at a month level; detect any Outliers in the**

**Sales\_Qty for each Store-SKU combination and apply an outlier treatment on the same.**

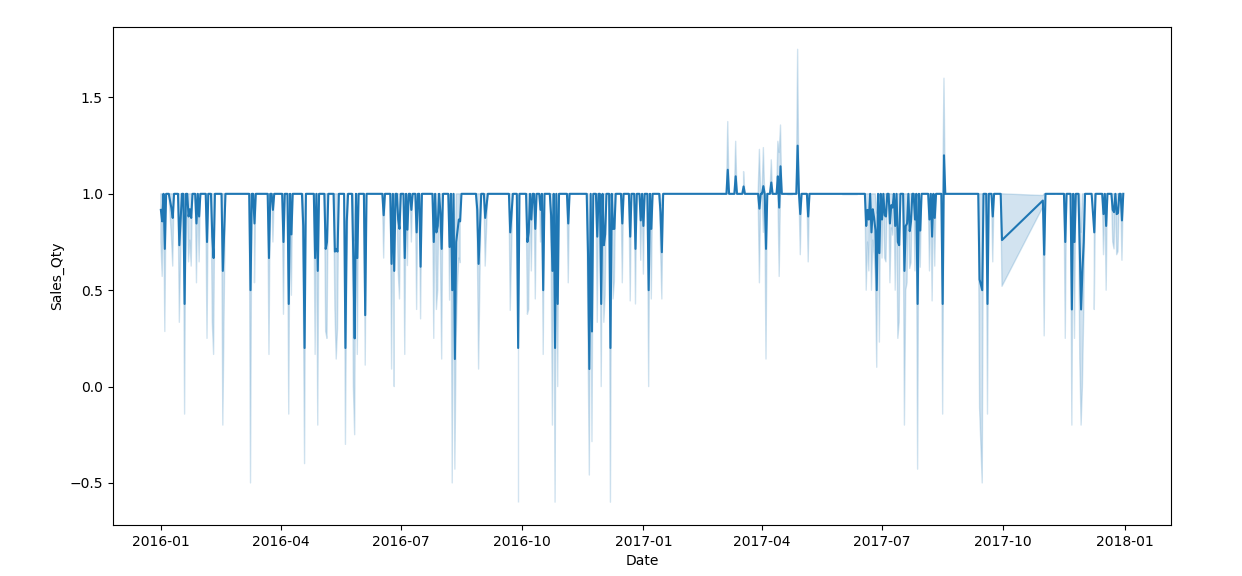
**Specify the outlier treatment technique.**

NOTE: All the analysis is done for Store1 .

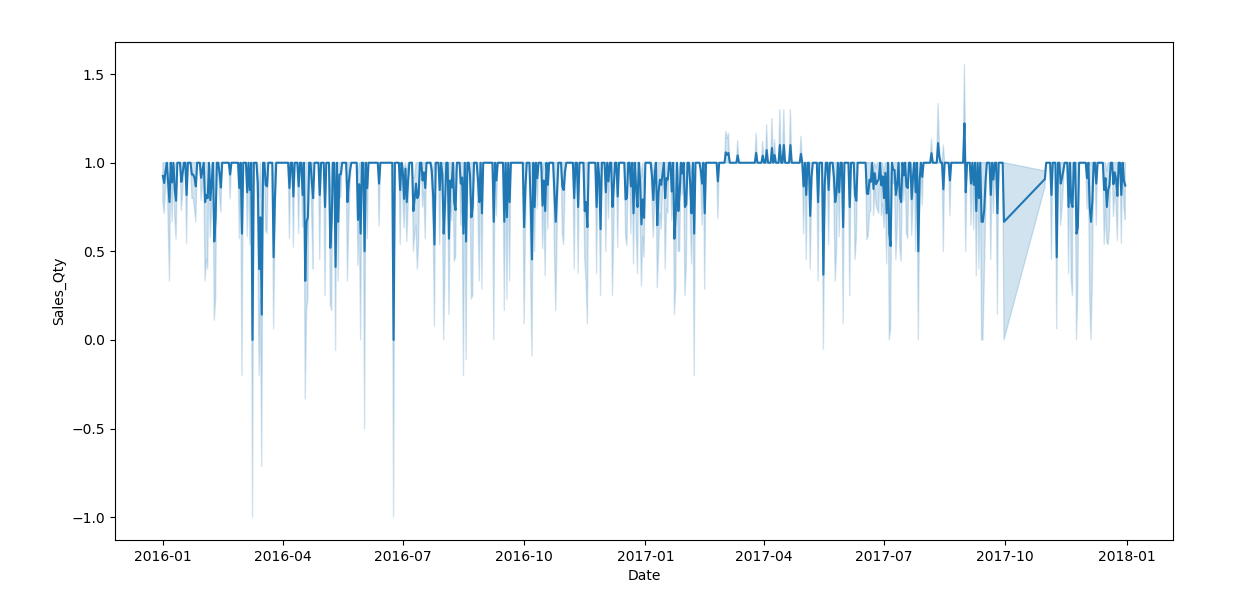
Plot for sales-Qty for Store1



Plot for sales\_qty for Store2

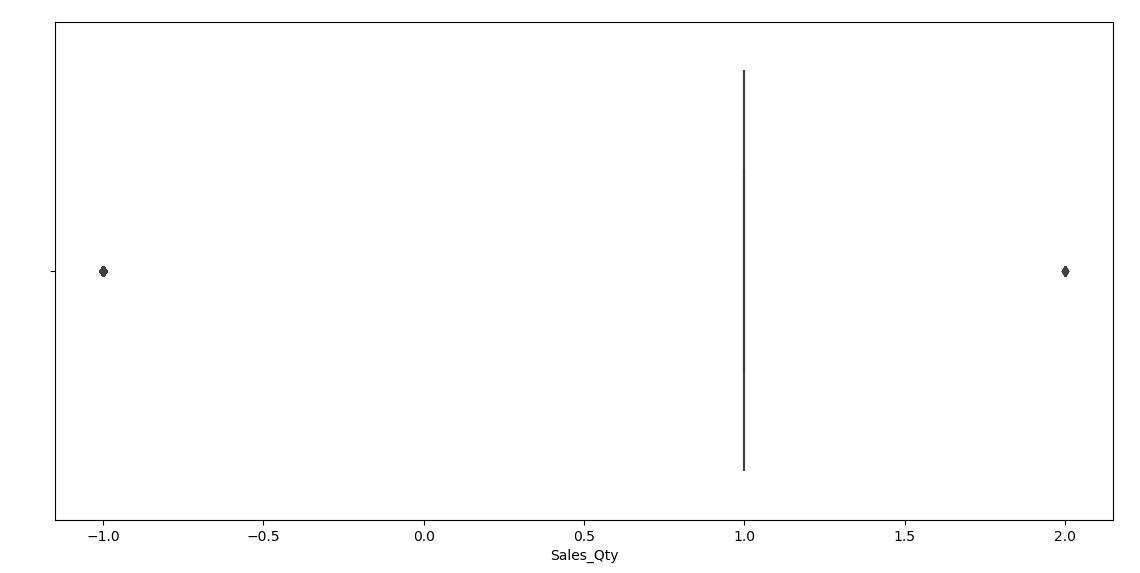


Plot for Sales\_qty for Store3



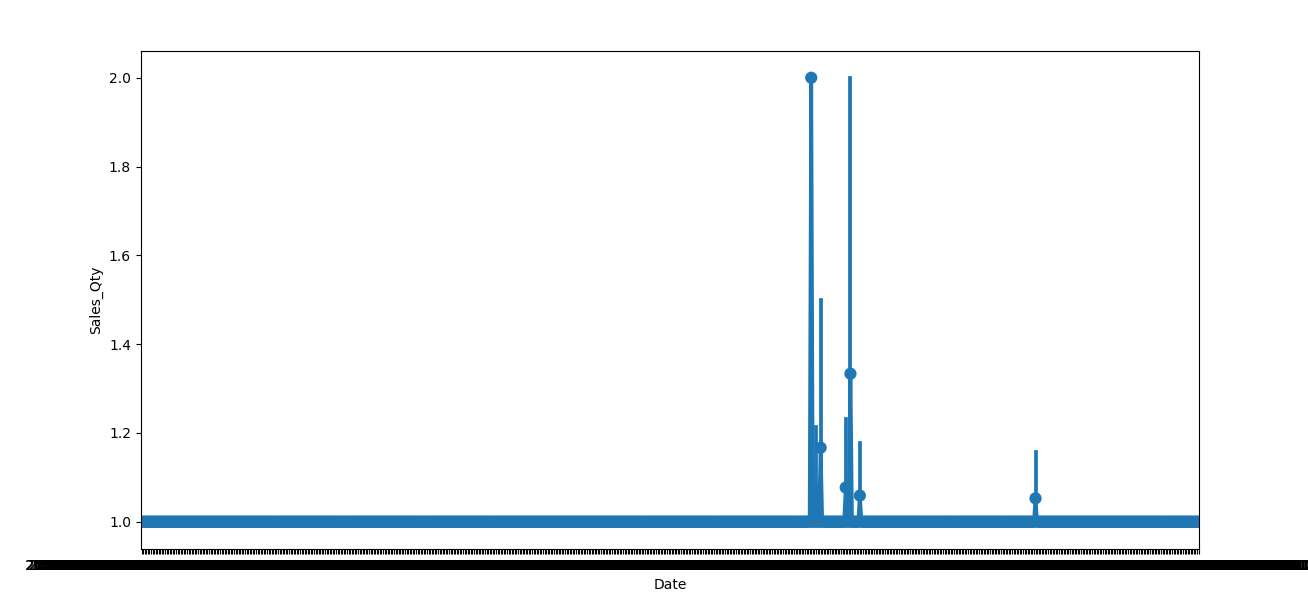
Outlier detection in Sales\_Qty

A good statistic for summarizing a non-Gaussian distribution sample of data is the Interquartile Range.

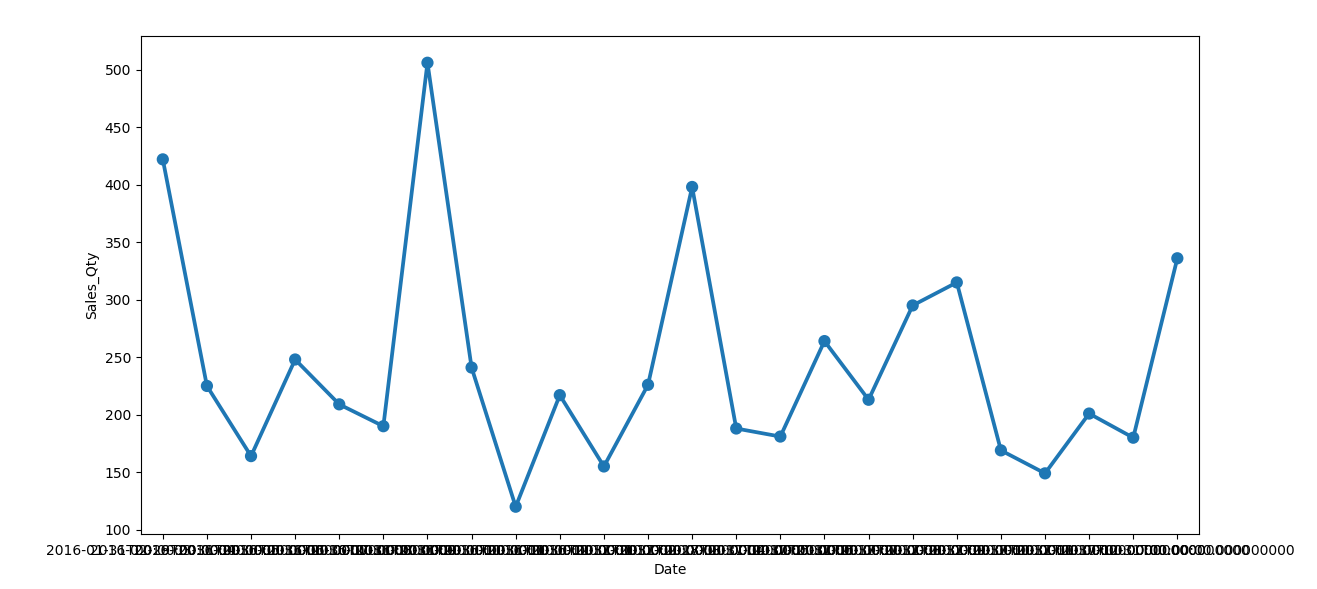


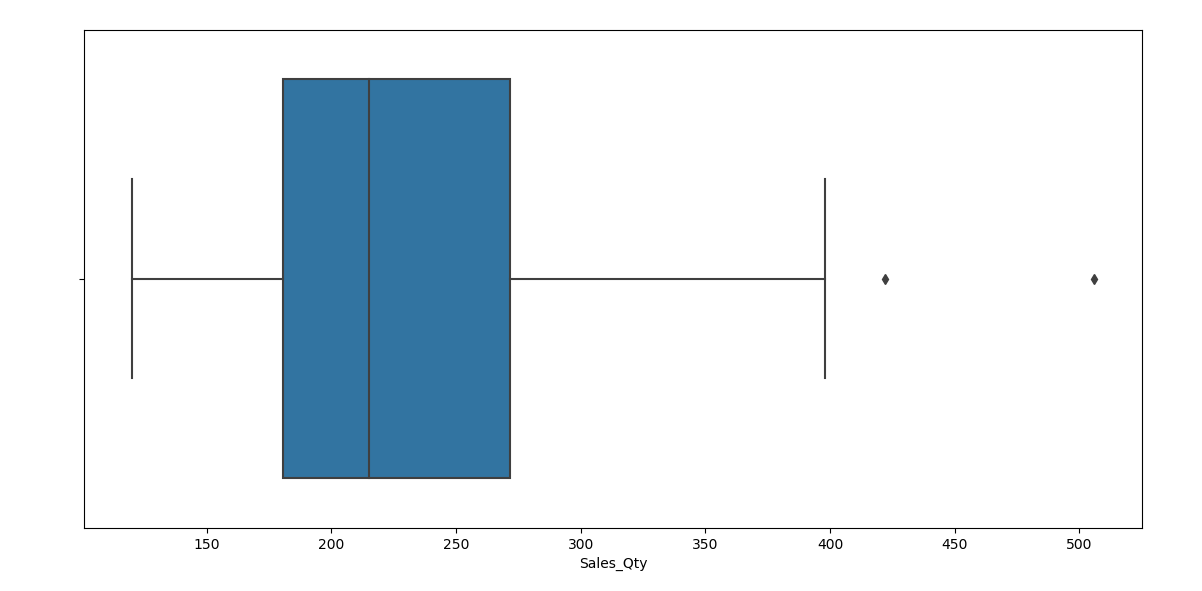
Box plot shows there are some outliers. And secondly most of the observations are lies on 1. And there are some items that must be returned and there are some items which sales are more than 1.

Try to remove the outliers on the daily Sales\_Qty . The result look bad. As it removes all the sales\_qty which are -1 and 2.

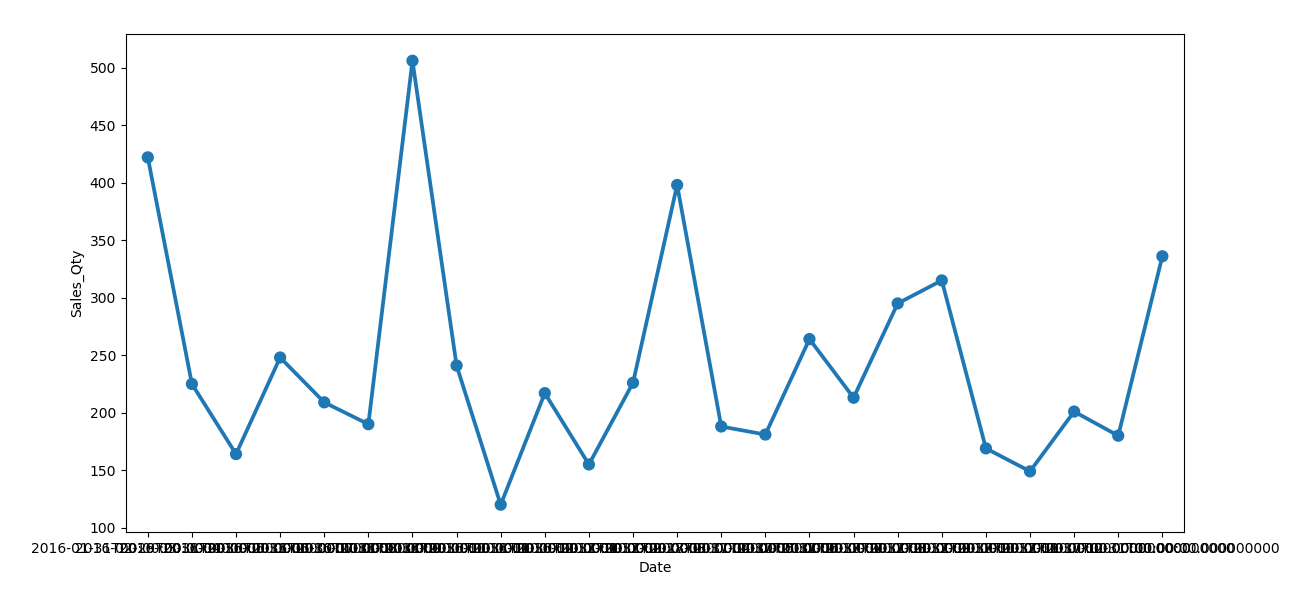


Now we are resampling the sales data at month level to see if any outliers is present in the data.





Now we can see we have some outliers in the sales data. Either we can remove these points, or we can keep it. Let’s check with z-score. At the month level with z-score we do not get any outliers in the data. So, our final graph of sales resembles the same before only.



Technique Used:

Box-plot

Z-score

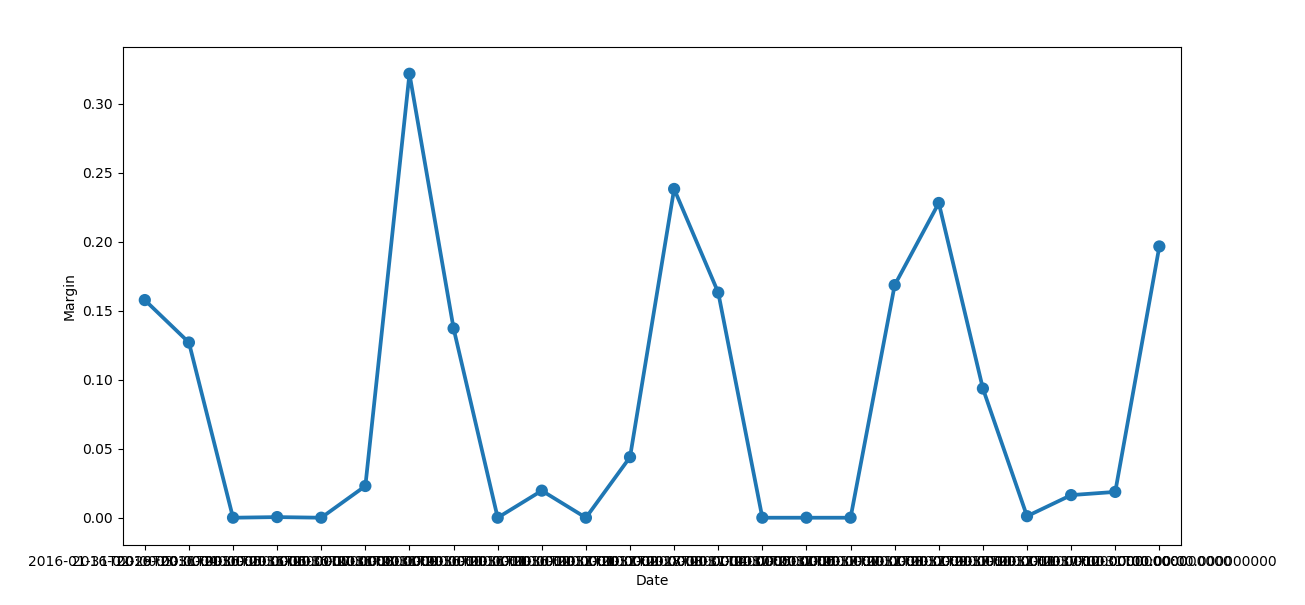
**2. Estimate the level of promotions (Discount%) for each Category-Store level at a month**

**level - remove any outliers / inconsistencies from this, and specify the technique used;**

**the level of promotions is defined as Discount% = (1 - sum of SP / sum of MRP)**

Again, when we test with z-score for the promotion also there were no outliers in this also.

Plot for margin in Selling price and Mark price over the period of time for store1.



**3. Estimate the inventory levels at a weekly level for each Store-SKU by interpolating**

**missing values from data on secondary and primary sales; the following equation holds**

**true in general: ( you can do this for a shorter period of Jan 2017 to Mar 2017 )**

**Closing inventory on day [t] = Closing inventory on day [t-1]**

**- Secondary (sales - returns) on day [t]**

**+ Primary (sales - returns) on day [t]**

**NOTE:**

**a. Secondary sales is the file named “WC\_DS\_Ex1\_Sec\_Sales.csv” - and it refers**

**to sales from stores to customers (and returns by customers)**

**b. Primary sales is the file name “WC\_DS\_Ex1\_Pri\_Sales.csv” - and it refers to**

**stock movements from retailer WH to stores (and returns back to WH)**

**c. Returns in both datasets are indicated by negative values in ‘Sales\_Qty’ and ‘Qty’**

**fields respectively**



**The calculation is done using Excel, for 3 months only.**

**4. The inventory estimations in Question 3 will have data inconsistencies - take any**

**assumption to resolve them and explain that assumption.**

First of all, the calculation is done for only 3 months. Primary sales data is generally available two times in a month for each SKU\_Code . And it is also no properly sampled, not taken at properly interval also. Lot of inconsistency was there to make the data in the pattern itself. To interpolate the primary data , I used Store1\_sales\_month.ffill(inplace=True). This basically replace the 0 with previous sales\_value when sample the data on weekly basis.

Similarly , I sampled the secondary sales data on weely basis . Take the sum of sales\_qty over the week.

**5. Using the Secondary sales data and inventory series from Question 3, determine**

**average out-of-stock percentage (OOS%) for each Category-Store combination at a**

**monthly level; the OOS % is defined as:**

**OOS % = 1 - {Average of no. of unique SKUs in stock each day**

**/ No. of unique SKUs in stock over the entire month}**

**(for each Category-Store combination each month)**

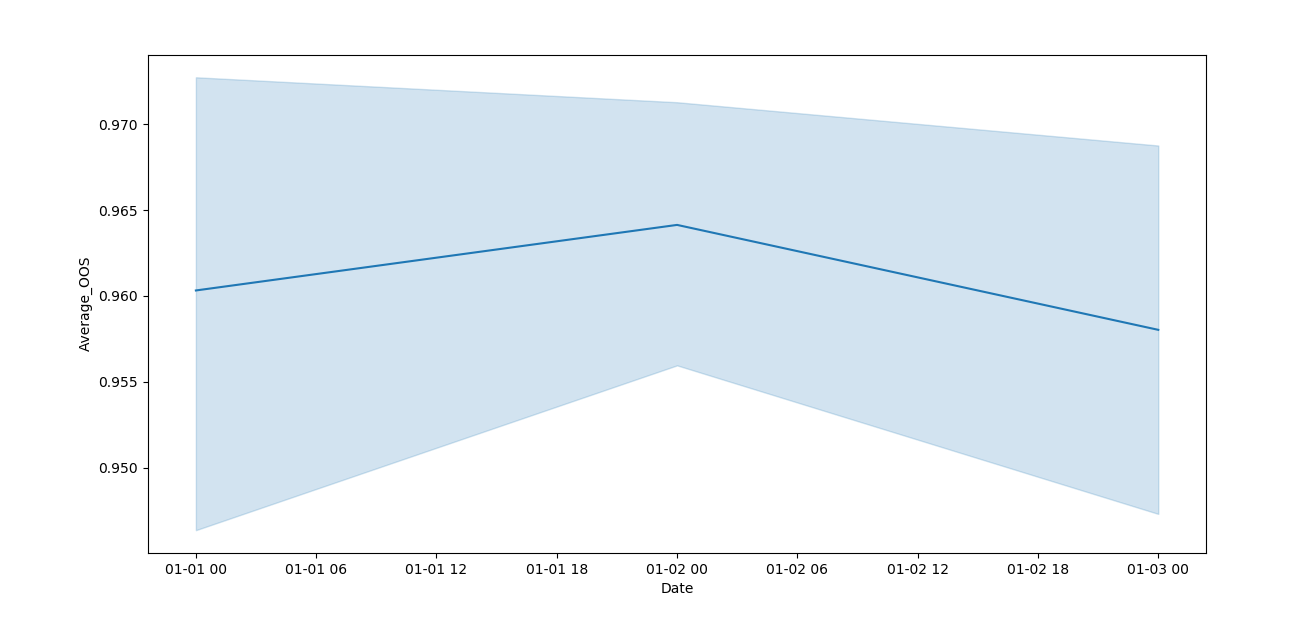
This calculation is also done only for 3 month.

Code sinpet for this is .

0.9666666666666665

0.9599999999999997

0.9565217391304348



**6.Using the historical secondary sales, inventory series, OOS% levels and promotion**

**levels, determine the demand for each Store-SKU combination at a monthly level for the**

**forecast period; use any forecasting technique that you’re comfortable with (you may use**

**multiple techniques.**