Pattern Recognition in Consumer Packaged Goods Data

Topic Area(s): Inventory, Database Analytics, Supply Chain Planning

By: Hanin Aboutaleb
Capstone Advisor: James B. Rice

Summary: One of the significant problems in retail is the lack of on-shelf availability (OSA) of high demand products. Retailers are continuously working on providing the product to the shoppers whenever and wherever they need it. In the efforts of improving their product availability, a leading CPG manufacturer decided to focus on recognizing patterns within the inventory data of its biggest retail partner. Upon analysis of the data, two patterns were identified: (1) replenishments preceded by out of stocks (OOS) and (2) short-lasting replenishments. Observations on which specific stock keeping units (SKUs) and distribution centers (DCs) experienced the biggest issues, were provided. These observations could potentially be leveraged to reduce out of stocks, increase on-shelf availability and improve performance in retail.

Prior to MIT, Hanin Aboutaleb worked as a Supply Chain Associate Manager with a consumer packaged goods manufacturer, Egypt. She received her Bachelor of Science in Mechanical Engineering from the American University in Cairo, Egypt.

KEY INSIGHTS

1. Pattern recognition could highlight opportunities to solve persistent problems in the operations of both the manufacturing and the retail industries.

2. The lack of on-shelf availability could potentially be alleviated by frequent analysis of the SKUs inventory per DC.

3. During the first three months after a product's launch, there is a major challenge in maintaining a safety stock level.

Introduction

The speed at which a manufacturing company analyzes big data and reacts to the market trends can be key to its success. In consumer packaged goods (CPG) companies, a delay in analyzing the huge amounts of data produced could result in missing opportunities to identify patterns that indicate problems, and therefore, delay response initiatives to prevent problems. This was the case with a leading CPG company that was experiencing lack of on-shelf availability of some of its profitable products. It was necessary to analyze internal data to identify where and why this problem existed.

Data collection, selection and analysis were conducted over the inventory data and sales data of a leading brand at a large retail partner. This analysis enabled us to identify patterns in inventory records, including short replenishments cycles and frequency of out of stocks. It was significant that, during a few months following product introduction, there was a high frequency of out of stocks. This reduces the product's availability on the retailer's shelf; hence, the service level drops. A few months later, replenishment policies of some products were changed to maintain their inventory above safety level. An example of this improvement is shown in Figure 1.

We provided observations on which SKUs and DCs had the highest frequencies patterns including replenishments preceded by OOS and short-lasting replenishments. Such patterns indicate that with a few changes in replenishment policies, out of stocks could potentially be reduced, increasing on-shelf availability.

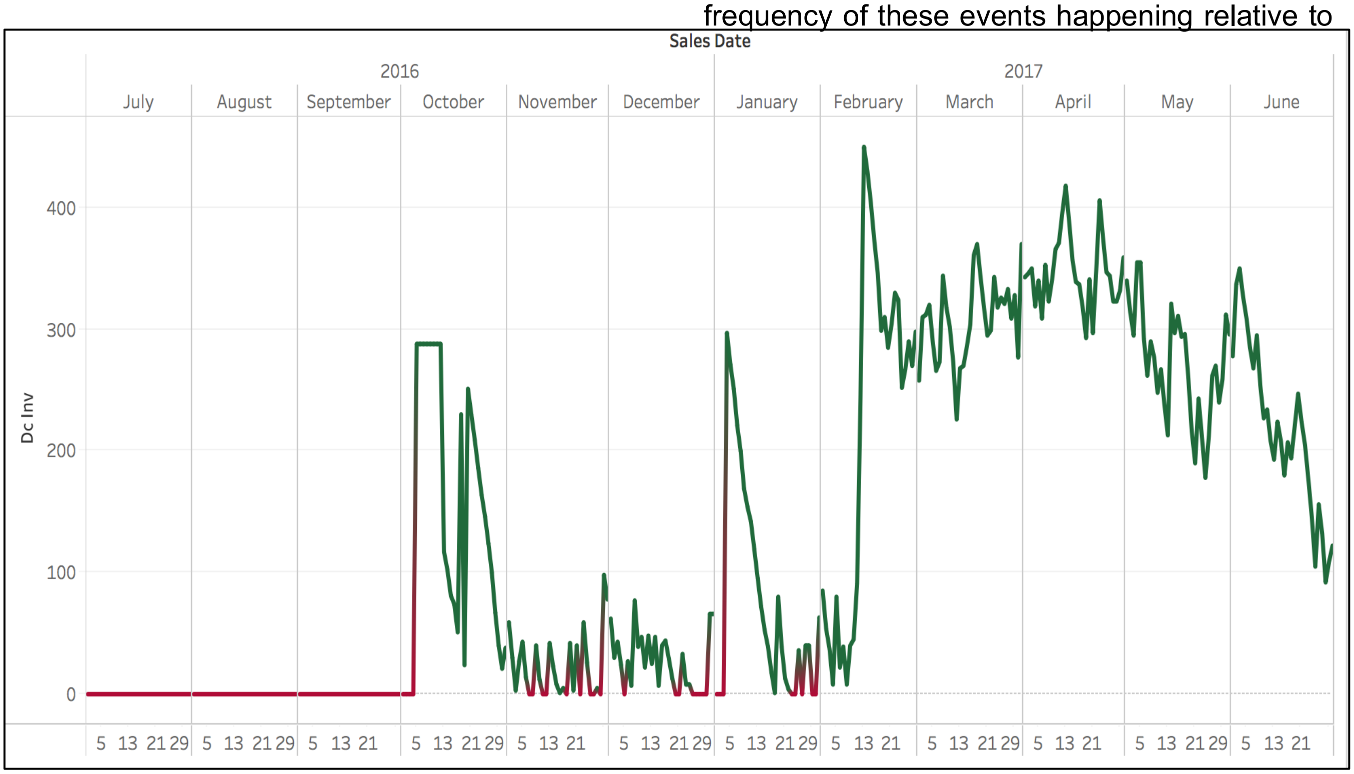

Figure 1: Inventory level of the SKU 101 in one of the DCs

Analysis

Inventory of products, in DCs, should be maintained above a safety level to mitigate the risk of out of stocks (OOS) in case of fluctuations in demand or supply. The two patterns we focused on, 'replenishments preceded by OOS' and 'short-lasting replenishments,' were of concern because in an ideal situation they both should not occur. To quantify the existence of these patterns, the frequency of these events happening relative to the total number of replenishments was measured. each SKU at each DC, the frequencies were variable according to how efficient the replenishment cycle was. The heat map shown in Figure 2, enabled us to spot specific SKU/DC combinations with high pattern frequencies, highlighting an opportunity of improvement. We aggregated these issues by SKUs, DCs, replenishment streams and time of the year.

DC/SKU	101	102	103	104	105	106	107	108
1	14%	6%	10%	3%	7%	1%	1%	1%
2	17%	3%	6%	4%	2%	2%	3%	4%
3	17%	5%	8%	5%	8%	6%	10%	5%
4	10%	5%	6%	3%	1%	2%	9%	5%
5	21%	7%	1%	2%	1%	1%	1%	1%
6	13%	7%	4%	2%	1%	1%	1%	1%
7	12%	2%	8%	5%	6%	6%	6%	5%
8	18%	10%	7%	3%	1%	2%	1%	2%
9	15%	2%	5%	2%	7%	7%	5%	4%
10	18%	9%	5%	2%	5%	4%	5%	2%
11	29%	5%	3%	1%	1%	1%	1%	6%
12	17%	7%	8%	2%	1%	1%	1%	1%
13	19%	2%	5%	5%	4%	2%	5%	4%
14	18%	2%	3%	1%	1%	2%	1%	3%
15	15%	2%	2%	1%	3%	1%	3%	5%
16	12%	2%	2%	8%	4%	4%	4%	2%
17	10%	2%	6%	4%	2%	3%	4%	6%
18	2%	4%	4%	3%	3%	4%	13%	11%
19	12%	2%	6%	3%	8%	1%	4%	4%
20	19%	6%	6%	4%	1%	2%	2%	1%
21	31%	3%	8%	4%	17%	7%	6%	4%
22	13%	5%	6%	3%	1%	1%	3%	9%
23	11%	4%	2%	3%	3%	1%	6%	4%
24	14%	9%	5%	4%	2%	2%	2%	3%
25	20%	2%	2%	4%	3%	1%	4%	12%
26	3%	15%	3%	3%	3%	3%	1%	4%
27	15%	6%	8%	9%	9%	9%	8%	6%
28	16%	2%	4%	7%	1%	1%	1%	2%
29	11%	2%	3%	1%	1%	1%	1%	1%
30	16%	2%	3%	2%	2%	2%	3%	3%
31	16%	2%	6%	1%	1%	3%	1%	2%
32	11%	2%	2%	3%	2%	2%	3%	3%
33	9%	2%	4%	1%	1%	1%	1%	1%
34	16%	2%	3%	4%	3%	1%	1%	1%
35	4%	5%	5%	6%	5%	1%	2%	4%
36	21%	2%	6%	1%	2%	1%	6%	4%
37	1%	3%	4%	3%	1%	1%	1%	6%
38	19%	6%	4%	2%	7%	1%	15%	2%
39	4%	4%	2%	2%	1%	1%	1%	1%
40	12%	2%	2%	1%	1%	5%	1%	2%
41	2%	2%	9%	4%	2%	2%	5%	2%
42	1%	2%	4%	4%	6%	2%	1%	1%

Figure 2: Frequency of replenishments preceded by OOS per SKU/DC (full-year)

The initiatives stream that was expected to have a higher percentage of patterns turned out to have the least. While SKU 101 is the one that had a higher percentage of base demand versus other streams, it still had a high percentage of replenishments preceded by OOS. This occurrence could be avoided by having more frequent replenishments to decrease the percentage of OOS.

The SKUs 101 and 103 had the highest average frequency of replenishments preceded by OOS. They both had high frequencies throughout the year and specifically in the first three months of the products' introduction. Referring to Figure 3, the huge difference between the frequencies in both SKUs shows that the first three months comprise most of the events. Therefore, if this was detected early on, the pattern could have been avoided in these few months, and hence, the overall frequencies of these SKUs drop.

The DCs 27 and 21 had the highest average frequencies among the 42 DCs studied. The persistence of the problem among all SKUs, in these two DCs specifically, indicates a problem with the replenishment policies of these DCs.

To be able to figure out why some DCs are performing better than others and why the issues are persistent in the first few months, more information is needed. This information should include details about the replenishment policies of the DCs and what challenges are they facing within their operations and logistics.

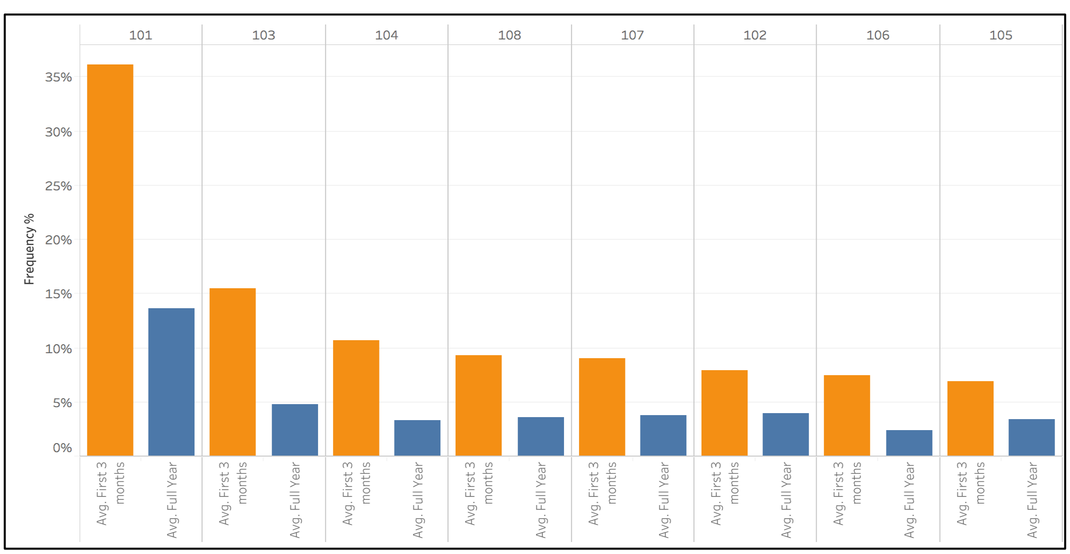

Figure 3: Comparison between frequencies during first 3 months and full year (per SKU)

Limitations

All the observations and insights developed after the data analysis were prone to a few limitations. Once the problem was identified, it was necessary to develop a 'What? Where and when? Why?' thought process. In other words, what is the problem being tackled? where and when does this issue appear most frequently? and finally why does it exist and what triggers it?

Some relevant information was needed about the operations of the DCs and how they handle their logistics in order to have a profound analysis. The unavailability of the data about the replenishment policies, lead times, transportation etc., hindered the identification of reasons for many of the observations. The what and where questions were answered; the problem was the lack of OSA and frequent OOS and mainly exists among specific SKUs and DCs over the time series. However, a root-cause analysis was not yet developed because not enough information was available on hand. The existing opportunities for improvements were highlighted in this project, and the CPG company can use their internal information to further analyze and determine causes and actions to be done.

Moreover, the analysis was mainly conducted using Microsoft Excel; data selection and visualization was done using MySQL and Tableau. In this case, excel was suitable for the few SKUs being analyzed. For the model to be expanded among hundreds of SKUs, it is best to use a different tool like Python. Developing a model that embraces the same concepts used here, using Python, would save a lot of the computation time needed to generate the data tables on Microsoft Excel.

Lastly, this project was focused on only a baby products' brand among all the brands that the CPG company produces. Exploring if other brands and products have the same patterns or behave similarly would have been beyond the scope of this project. Therefore, the focus was on one brand which might not be representative to all the products and how they behave.

Future exploration

As implemented in this project's approach, using heat maps and histograms to identify DCs and SKUs that need investigation of replenishment policies is recommended. Once the SKUs and DCs of focus are determined, analysis of the internal operations data is necessary. The companies' operations team could look into the drivers of replenishment that would lead to OOS. For instance, the lead times (order and transit times), lead time variability, order quantities and forecast errors. Analyzing these factors would imply some changes like enforcing a higher safety stock at the DC or store. It could also recommend having shorter transit time or cutting down lead time uncertainty by shipping only by truck for example. The company could incorporate several transportation and operations management changes to handle the plethora of incoming products at the DCs.

Furthermore, it is necessary to quantify how much an OOS at the DC is worth, even after the frequency of replenishment preceded by OOS is identified. It's also necessary to identify how an OOS affects the store sales. The model used in this project outputs the average number of days that an OOS event lasts for. This could be used to evaluate the responsiveness of the DC to the stock out and whether an OOS lasts for long enough to question the DCs sensitivity to issues. Before implementing any drastic changes, the quantification of the impact of the problem is important to determine whether it is worth solving.

Conclusions

Pattern recognition could highlight opportunities to solve persistent problems in the operations of both the manufacturing and the retail industries. In the scope of this project, a leading CPG manufacturer provided the inventory data of its products in a retailer's DC. The data was scaled down to the leading baby product brand, produced by the manufacturer, across all the retailers' DCs. Based upon the analysis of the inventory data of the highest selling SKUs of the brand, some patterns were recognized. The first pattern was the high frequency of replenishments that are preceded by out of stocks (OOS) out of the total replenishments. The second pattern was short-lasting replenishments, where the inventory drops back to zero within three days after the replenishment.

It was found that many SKUs have high frequency of these patterns, especially in the first few months after the product is introduced. Specific DCs seem to have a problem with setting up a replenishment policy that maintains a safety stock while fulfilling the customers' demand. For the first three months after a product's introduction, some DCs where able to fix the issues they face and decrease the frequency of OOSs and the frequency of which the patterns appear. However, other DCs had the same problem persistent along the time series.

Upon shedding the light on specific DCs and SKUs that need focus, the limitations of the approach used to arrive at certain insights were described. A few recommendations were provided about the internal data measures that need further analysis before deriving the causes of the problem and developing an action plan. The lack of on-shelf availability could potentially be alleviated by frequent analysis of the SKUs inventory per DC. Identifying patterns in the inventory data could highlight improvement opportunities and impact performance in retail.