UNIVERSITÀ DI PISA

DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA DEI SISTEMI DEL TERRITORIO E DELLE COSTRUZIONI

RELAZIONE PER IL CONSEGUIMENTO DELLA LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE

"Fast consuming products" in Procter and Gamble: Definition of a algorithm for estimation and monitoring of their shelf availability

SINTESI

RELATORI

Prof. Ing. Riccardo Dulmin
Dipartimento di Ingegneria dell'Energia,
dei Sistemi, del Territorio e delle Costruzioni

IL CANDIDATO

Dario Tonarelli
dariok13@gmail.com

Sessione di Laurea Magistrale A.A 2020/2021

“Fast consuming products” in Procter and Gamble:
Definition of a algorithm for estimation and monitoring of their shelf Availability

Dario Tonarelli

Abstract

In the market of fast consuming products, the presence of the product in the shelf is one of the main factors to stay competitive and build or maintain strong the loyalty of the clients to the brand.

The absence of the product, caused by an out of stock, bring directly to a loss of sale and could be a cause of loss of the loyalty of the clients to the brand or to the retailer.

For this reason, in this high competitive market, the measure of On Shelf Availability (OSA) has always more importance a as index of performance of the product supply.

This pilot project has the purpose to estimate this index for the products sold in a determinate retailer, giving the opportunity to monitoring it and detect easily issues in the supply chain.

In the actual state of the art those indexes are estimated using the daily level of inventory of the retailer, this is one of the first projects that use a statistical model for compensate the miss of this data and use only the sell-out data (POS).

In this specific application in Procter and Gamble, the scope is to be able to estimate with a good level of precision the index OSA for all the products sold in the retailer's points of sales around Spain, without inventory data but using only the POS data.

The POS data, differently to the inventory data, is already shared by the big part of the retailers and doesn’t need complex informatics system.

This project could introduce a faster and cheaper way for estimate the on shelf availability of the products to be implemented easily in a large group of retailers.

1. Purpose of the project

The project has the purpose to develop a tool that, analysing only the point of sales (POS) data of a retailer, identifies the trend of the sell-out of all the products and with a statistical model estimate their on-shelf availability.

The objective was obtain an estimation of index OSA with a level of detail that gives the opportunity to analyse the performance at different level of aggregation The tool actualizes the estimation of the index with a daily frequency and gives the opportunity to monitor the index and create alarm for react rapidly in case a reduction of performance or issues.

So, the tool will be able to estimate those supply indexes:

- OSA: on shelf Availability (in percentage)
- OOS: Out of stock (days)

In the actual state of the art those indexes are estimated using the daily level of inventory of the retailer, this is one of the first projects that use a statistical model for compensate the miss of this data and use only the sell-out data (POS).

Using the only the POS allow to reduce dramatically reduces the time and cost of implementation and implement for a larger group of retailer.

In the figure 1 is shown the Top-level architecture of the tool.

The POS data received from the retailer is stored in a database from where Algorithm, realized in Knime, upload the data and run the statistical model. Once estimated the indexes, those are uploaded in a Power BI report that will few dashboards allow to explore and monitor of the results.

Figure 1 Top level architecture

The project includes also design a efficient and user-friendly dashboard that allow the sales team and the supply team that collaborate with this specific retailer to explore the results obtained with the tool.

Considering that this is a first implementation, the scope of the project was to isolate the analysis and realize the model only for the point of sales of the retailer classified as Hyper and of the whole Product and Gamble product portfolio, only the products with high volume of daily sales.

2. Work teams and project responsibility

For this project, I was first responsible and project manager, with the duty to study the retailer, the products portfolio and, with the collaboration of the IT support team, the supply team and the sales team, develop the algorithm.

The time defined for the project was of 4 months with the objective to deliver at the end of the period a first implementation and version of the tool, perfectly working.

The collaboration with the team has been weekly with the objective to validate each step of the process and take advantage of the experience of both the business teams (supply and sales).

3. Methodological scheme for the project

For realize the project has been followed this scheme:

1. Study of the retailer and the P&G product portfolio
2. Study of the supply of the retailer and the position of the main distribution platform
3. Upload, clean and understand the data
4. Study the sales trend and product rotation
5. Define the size of the dataset needed for the statistical model
6. Build the classification tree for the product rotation
7. Develop the statistical model for classify the day with no sales as OOS
8. Validate results with the supply team
9. Develop the tool in Knime
10. Develop the Power BI reports

All the steps, excluding the one of develop the statistical model, have been executed in collaboration of one of the 3 teams.

3. The algorithm

The algorithm has been realized with Knime, a platform that allow to create data analytics algorithms with flows composed by operational nodes.

As visible in the Fig. 2 the Knime flow is structured in 3 macro blocks.

The data preparation one that extracts and prepares the date for the following blocks.

The learning block, that studies the historic sales data and classify the product in rotation groups and the scoring block, that detects the OOS for high and medium rotation and estimate the OSA.

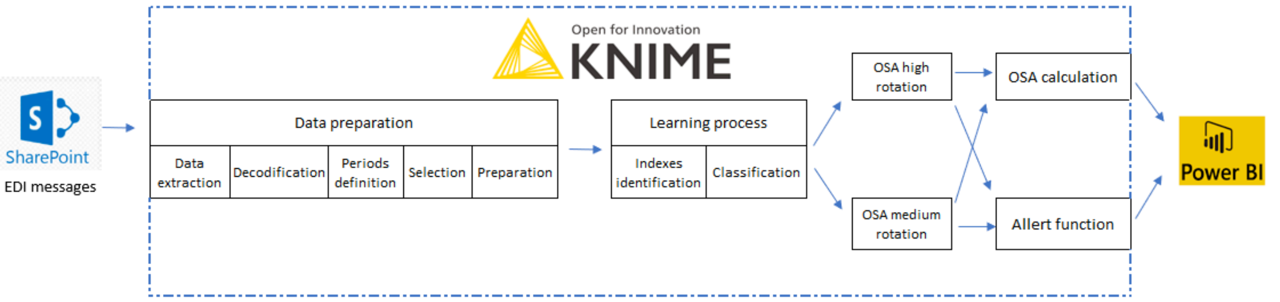

Figure 2 Architecture of the tool

Once developed the Knime flow could be uploaded in the Knime server and be launched with a daily frequency, so to run the analysis and update results every day.

3.1 Input data and data preparation block

The source of data for this project is organized in 3 different databases:

- DB POS: updated with daily frequency, contains the data of the sales performed in all the point of sales of the retailer across Spain for all the product of P&G
- DB Retailer: gathering all the information about the point of sales, as name shop, city, region and type of channel
- DB products: gathering all the detail and attributes of the products of P&G

The retailer send every day 1 or 2 EDI messages in EDI codification with the data of the sales done in the previous day.

In the Preparation Block the EDI messages are uploaded in Knime, decoded and organized in a table composed the main attributes: Date, Shop-code, Product-code and Quantity sold.

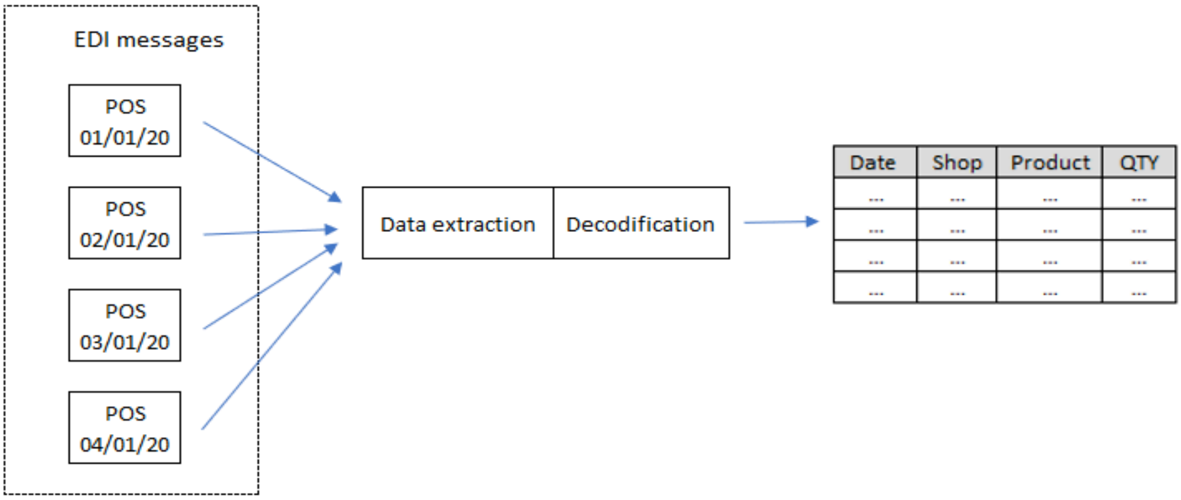

Figure 3 Data extraction and Decodification

3.2 The learning block

The Analysis of the daily sales is done with a learning process that study the history of the elements "shop-product", so each product in each shop, for classify them in the 3 different types of rotation and estimate the indexes necessary for describe the trend.

For the project, we had available only 6 months of data history with a daily update and increase of the dataset with the recent data.

The OSA index is a temporal index, it is referred to a determinate period.

In this project, we decided to take in consideration for estimate the OSA just the last 30 days, starting from the day of the last data received.

Figura 4 Dataset periods split

3.2.1 Daily sales trend analysis

This portfolio is composed by different types of product, each one with a different trend of sales. In the fast consuming market the products are generally classified by their daily sales volume in 3 groups:

- High rotation, high volume of daily sales and really rare day with zero sales.
- Medium rotation, medium volume of sales and some day with no sales.
- Low rotation, low volume of sales and frequent periods with no sales.

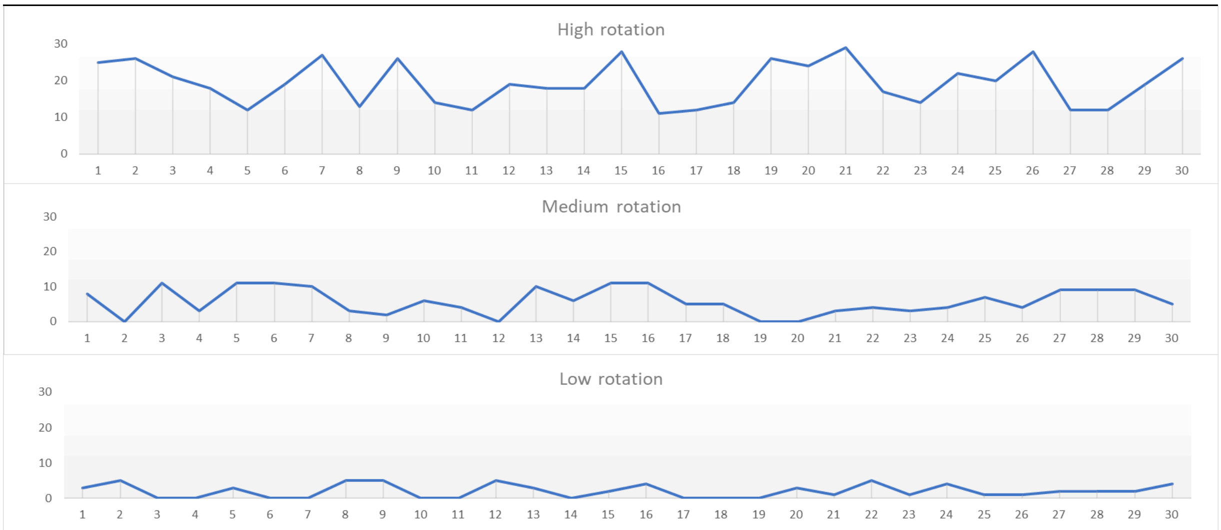

Figure 5 Types of product rotation

Observing the distribution of the daily sales of the singular products in the different point of sales, we have noticed that, taking a period of 3 months we find a distribution of frequency with the shape of a Gauss curve cut on the zero (see Fig 6)

The products with higher volumes recreate a curve more centred in the average and a distribution that we could associate to the T-student one cut in the zero.

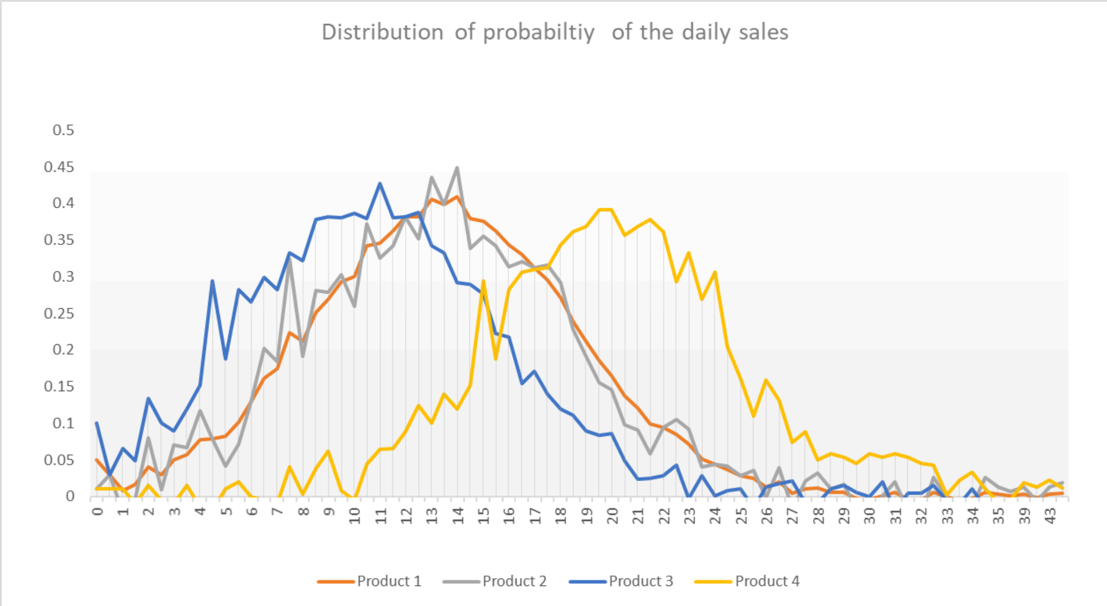

Figure 6 Distribution of probability of the daily sales for 4 different products

Higher is the daily volume of sales and lower is the frequency that in a day has been performed zero sales. This curve of frequencies represents, with a high number of days, a good approximation to the distribution of probability of the decision of the costumers about buy the product in a whole day.

In the index identification process of the learning block are calculated, from the training set, the indexes needed for describe the distribution of probability of the daily sales (the first 2) and two indexes that describe the frequency and size of the periods with no sales (the last 2)

- Average daily sales: μ sales
- Sales deviation standard: δ sales
- Average size of no sales periods
- Number of no sales periods

In this way, we could have defined the fluctuation of daily sales in the period studied and build the distribution.

3.2.2 Product classification

The classification process is structured mainly in two phases:

- Detecting the lower rotation products
- Dividing the remnant products in high and medium rotation

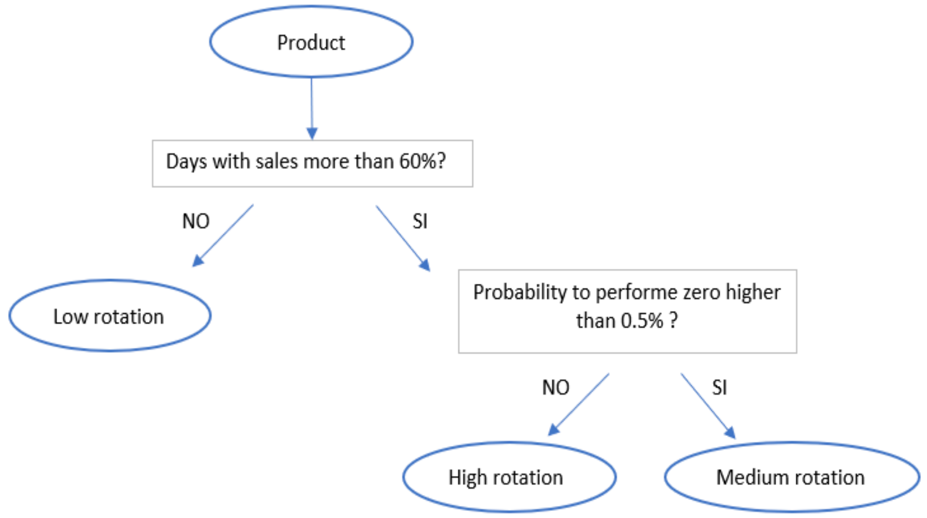

Figure 7 Three of classification for the products rotation

For classify the product in high and medium rotation is taking in consideration their distribution of probability for the daily sales. The products that have a probability to perform zero sales in a day lower than 0.5% are classified as high rotation and the ones that have it higher and are not low rotation are classified as medium rotation.

3.2.3 Scoring block

In the scoring block, the Knime flow will classify the days with no sales in OOS or not and will calculate the index of OSA with the 2 levels of aggregation.

3.2.3.1 OOS classification for High Rotation products

In the High rotation products, considering that have the probability of perform no sales in a day with the distribution of probability identified less than 0.5%, if there is a day of zero sales we have a high confidence that it is due a supply issue and not a decision of the costumer. So, all the days with zero sales are classified as Out of Stock days.

3.2.3.2 OOS classification in Medium Rotation products

In this case the days with no sales are more common and cannot have classified them directly in OOS. In medium rotation products, a day is classified as OOS when is part of a period with 0 sales (days in row) that has a length equal or higher than the average size of the no sales periods calculated for that product in that shop. During the learning block is calculated the average size of the periods with no sales performed in the historic data, so to have a dimension of the regular period with no sales driven by decision of the costumer and not a product shelf availability issue.

If for example a product has an average of 3 days of no sales, looking at figure 8, is classified as OOS only the period from the 28th to 29th

Period																													
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
10	7	7	0	14	12	12	13	14	12	11	10	8	8	10	7	0	0	12	9	10	12	11	13	9	8	0	0	0	5

Figure 8 Medium rotation OOS classification example

The first day of the period is not classified as OOS because the probability of false positive is higher and get lower and lower for each day with no sales that are in row. If day after day there are no sales of that product, increase the probability that is due to a availability issue than a costumer decision. So classifying all the days of the period, excluding the first, reduce the error in the estimation of the OSA index.

3.2.3.4 OSA calculation

Once identified for each element shop-product the number of day of OOS in the period and knowing the size of this period, that is the number of days in which the shop in analysis stayed open for sales, we can calculate the index OSA.

In this application, the index OSA will be calculated with two different levels of aggregation:

- OSA Shop-Product: index of OSA of a product in a determinate shop/point of sales.
- OSA Product: index of OSA of a product in whole Spain for the product.

The OSA Shop-Product estimate the level of availability of the product in the shop and help for identify issues or inefficiencies isolated a one determinate shop or, aggregating the data, in a region. Aggregating the data we will be able to observe, with different geographical detail (shop, city, region) the overall performance of the products or the categories.

The index is calculated as the percentage of days with no OOS divided by the days in which the shop has been open.

$$\text{OSA Shop-Product} = 1- (\text{ OOS})/\text{Day of sales}$$

4. Results visualization

The Power BI has the porpuse to visualize the results and the perfomance in the way to optimize and be easly understandable from the managers.

The report has been realized to be accessible and easly understandable for:

- The Supply team: that will be able to track the OSA performance and detect supply issues fastly so to be able to actuate corrections or solve the problems
- Sales team: check how is the avaiability of the products with the higher priority in the portfolio and directing their analysis across all the spanish cities and all the retailers.

So, the most important caratteristic of the Power BI is the opportunity of aggregate the data of the performance and be able to filter them easly and fastly using all the main actributes necessary for the analysis.

In this Fig9. we could see the first dashboard of the power BI with the overall performances for each point of sales and reagon of each product category.

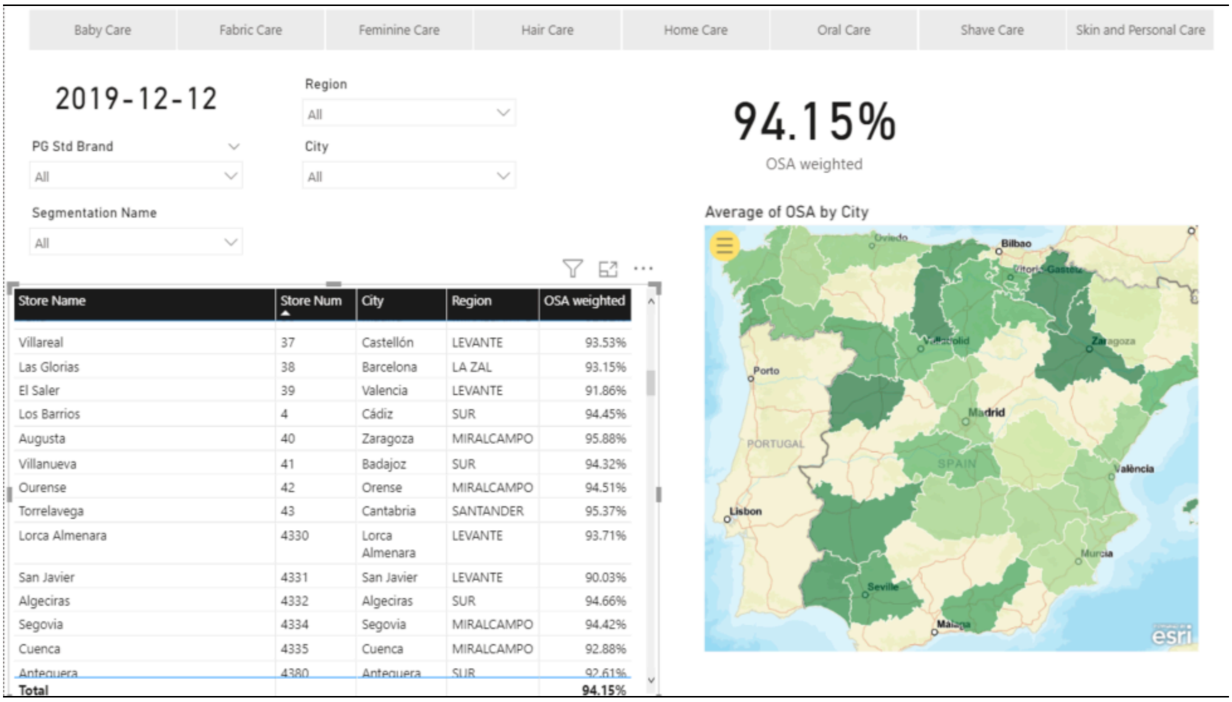

Figure 9 Power BI Dashboard geography view

The Report, structured 4 dashboard that give the opportunity to deep dive to the indexes results at a geography level and product level. This report has also integrated a specific dashboard that evidence and flag the products that are having right now supply issue, giving the opportunity to take direct action for identify the cause and resolve the problem.

5. Conclusions

The overall project has been a success, with a small dataset, is possible to realize estimations about the requested indexes of performance with a really good level of precision, a great detail and a daily update. The deadline of 4 months had been respected and the results of the OSA indexes after few checks with the teams were in line with the value expected and the information that we got from the operative issues.

A real unique value of this algorithm is the opportunity to implement it on all the retailers that agree to share EDI data, nowadays practically all the most important clients of P&G, in really short times and near zero cost.

This was just a pilot project in which could be implemented more functionalities, but in the future if applied also on other retailers could be possible add aggregation of data in a higher level or observe the supply performance of a product with a all clients overview in the Spain territory.

The need of monitoring the supply performance in this last mile of the service is getting always and always more important and statistical or machine learning model will be fundamental in the next years for guide the company in optimize their supply chain and increase the competitive advantage

This project gave me a unique opportunity to learn a lot for the market, the retailer, the supply chain and how to work collaborating with 3 different teams.

Have been a real pleasure work with P&G and develop a such interesting solution in a so complex and competitive market.