ORKIDEON

Supply Chain Planning Optimization with Data Science and Artificial Intelligence: A real case

by Thiago Gatti
10 min read.

Motivation

After years working as a Supply Chain consultant implementing industry leading software and creating Excel solutions for Finance and S&OP, I have decided to integrate my knowledge and experience into a single Python application.

IBP - Integrated Business Planning is the app's focus, because it synchronizes Finance and Supply Chain algorithms into a single set of numbers. This way, planning is faster and easier, which is my objective, rather than just building a "software".

With this business case, I hope to expose opportunities in Supply Chain, Sales and Finance based on real data, with digitalization and strategy.

I hope you enjoy reading this material and consider Orkideon for a project in your company.

Business Case Scenario

The company was a major chemical industry player in Brazil in 2011, with three factories, around 200 employees, over 6.000 items and close to 1,500 B2B niche clients, which were the ones considered by this study.

Low service-level was the main issue, with client complaints due to unproduced items, followed by slow moving inventory and high transfer costs. Basically, that planning team was having a hard time choosing what to produce and when.

I was hired to help solving the issue. My approach was to use data-driven decisions to properly valuate the hindrances. We have designed and implemented databases and Excel APO's - Advanced Planning and Optimizers.

The use of those APO's by the Production Planning & Control team, now under my umbrella, increased the demand forecast from around 30% to near 75%, which improved the quality of the Master-Schedule, which improved the throughput, which is a perk.

APICS MPS - Master Production Scheduling best practices and forecasting algorithms were implemented and, combined, increased service level by around 50%. Client complaints ceased.

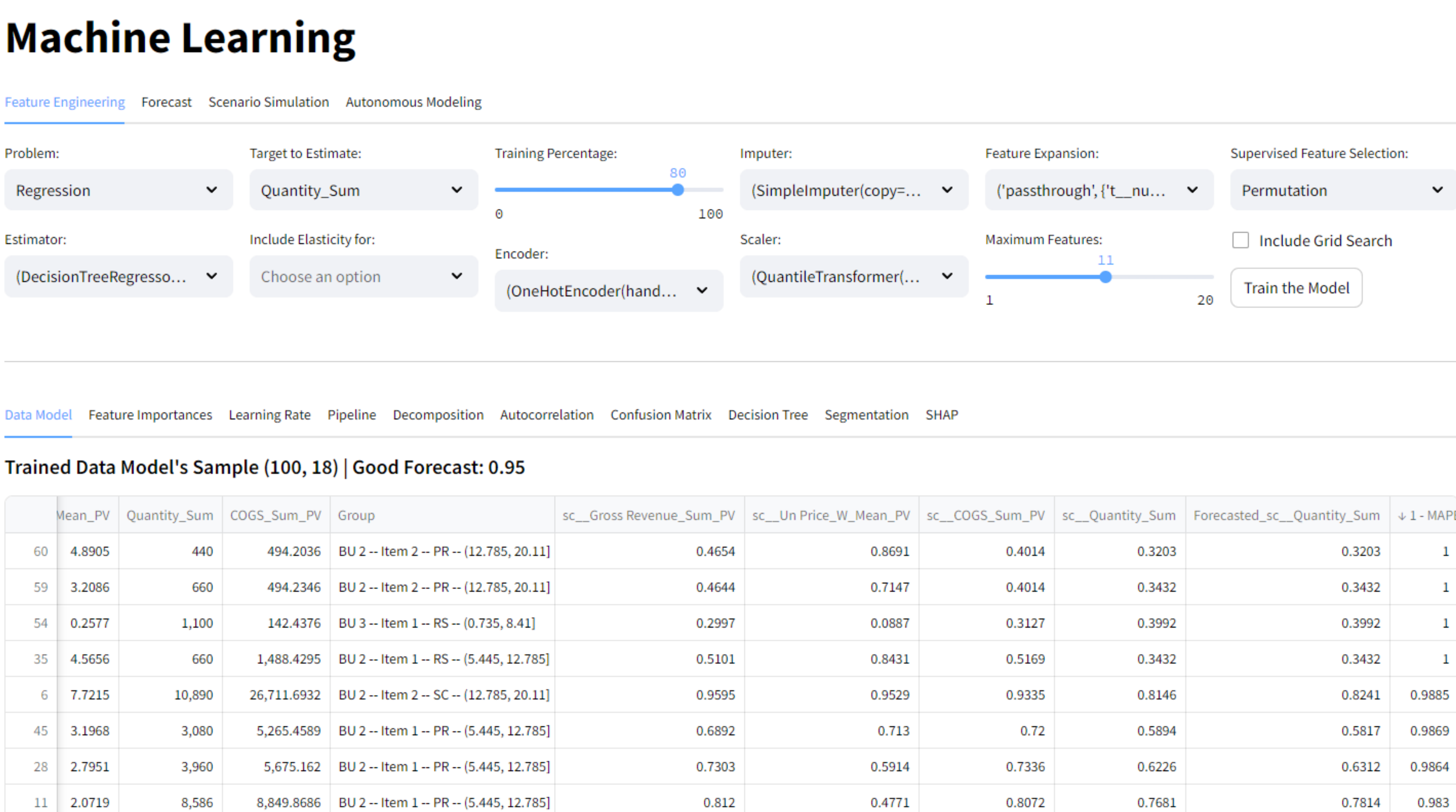
Despite the success, one question remains: Could Machine Learning do better?

Business Case Objective

1. To demonstrate the Machine Learning application being developed in Python.
2. To demonstrate some of the mentioned Excel APO's.
3. To demonstrate the use of Data Mining freeware.

These systems are used together to answer common Supply Chain questions like:

a. What database design?
b. What planning review period?
c. What demand forecast accuracy?
d. How much data is needed?
e. What features to consider?
f. What production master-schedule?
g. What inventory reduction / working capital release?
h. What warehouse capacity?
i. What service-level?
j. What inventory target?
k. What production capacity?
l. What ROI?
m. What price?
n. What surprising insights?



	Mean_PV	Quantity_Sum	COGS_Sum_PV	Group	sc__Gross Revenue_Sum_PV	sc__Un Price_W_Mean_PV	sc__COGS_Sum_PV	sc__Quantity_Sum	Forecasted_sc__Quantity_Sum	↓ 1 - MAPE
60	4.8905	440	494.2036	BU 2 -- Item 2 -- PR -- (12.785, 20.11]	0.4654	0.8691	0.4014	0.3203	0.3203	1
59	3.2086	660	494.2346	BU 2 -- Item 2 -- PR -- (12.785, 20.11]	0.4644	0.7147	0.4014	0.3432	0.3432	1
54	0.2577	1,100	142.4376	BU 3 -- Item 1 -- RS -- (0.735, 8.41]	0.2997	0.0887	0.3127	0.3992	0.3992	1
35	4.5656	660	1,488.4295	BU 2 -- Item 1 -- RS -- (5.445, 12.785]	0.5101	0.8431	0.5169	0.3432	0.3432	1
6	7.7215	10,890	26,711.6932	BU 2 -- Item 2 -- SC -- (12.785, 20.11]	0.9595	0.9529	0.9335	0.8146	0.8241	0.9885
45	3.1968	3,080	5,265.4589	BU 2 -- Item 1 -- PR -- (5.445, 12.785]	0.6892	0.713	0.72	0.5894	0.5817	0.9869
28	2.7951	3,960	5,675.162	BU 2 -- Item 1 -- PR -- (5.445, 12.785]	0.7303	0.5914	0.7336	0.6226	0.6312	0.9864
11	2.0719	8,586	8,849.8686	BU 2 -- Item 1 -- PR -- (5.445, 12.785]	0.812	0.4771	0.8072	0.7681	0.7814	0.983

Figure 1: Python with Streamlit: The Machine Learning Module.

What database design?

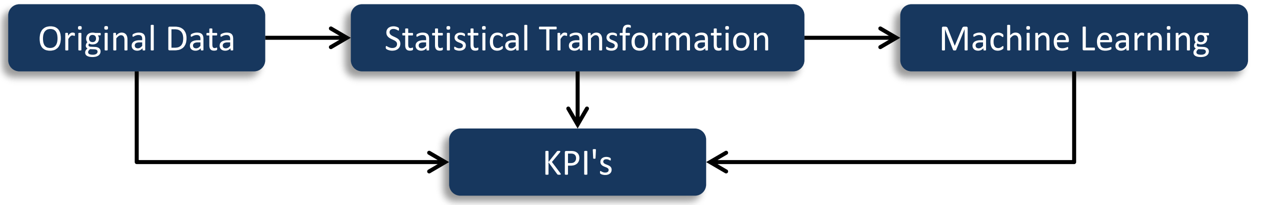


Figure 2: Simplified overview of the app's internal structure.

The original data is transformed until it is ready for Machine Learning. KPI's are present at all stages. The further the stage, the more revealing the KPI. As it will be noticed in the following pages, the available KPI's cover all Analytics layers, described by Gartner as: Descriptive, Diagnostic, Predictive and Prescriptive.

Original Features	Augmented Features	Relevant Features
Year	Period	sc__COGS_Sum_PV
Month	Group	sc__Un Price_W_Mean_PV
Day	year	sc__Gross Revenue_Sum_PV
Region	month	
Quantity	day	
Gross Revenue	BU	
Net Revenue	Item	**Final (arguably) Data Model**
COGS	Region	Occurences
Gross Profit	Un Price_Cut	Item
Reference Cost	Occurences	Salesman_Count
Replacement Cost	Gross Revenue_Sum	Un Price_W_Mean_PV_Arc_Elasticity_of_Quantity_Sum
BU	COGS_Sum	BU
Client	Quantity_Sum	Gross Revenue_Sum_PV
Item	Salesman_Count	day
Salesman	Un Price_W_Mean	month
Un Price	Interest_W_Mean	Region
	COGS_Sum_PV	Period
	Gross Revenue_Sum_PV	Un Price_W_Mean_PV
	Un Price_W_Mean_PV	Quantity_Sum
	Delta_Perc_Un Price_W_Mean_PV	COGS_Sum_PV
	Delta_Perc_Quantity_Sum	year
	Un Price_W_Mean_PV_Arc_Elasticity_of_Quantity_Sum	Group
	Un Price_W_Mean_PV_x_Quantity_Sum	
	Delta_Un Price_W_Mean_PV_x_Quantity_Sum	
	Delta_Quantity_Sum	
	Marginal_Un Price_W_Mean_PV_x_Quantity_Sum	

Figure 3: Data at different stages of the preparation process.

The transformation process (aka: pre-processing, wrangling, etc.) is dynamic, including and excluding columns (features), changing values and number of rows.

The data used in the study is real and it was completely anonymized for this business case.

What planning review period? **Monthly**

Autocorrelation (ACF)

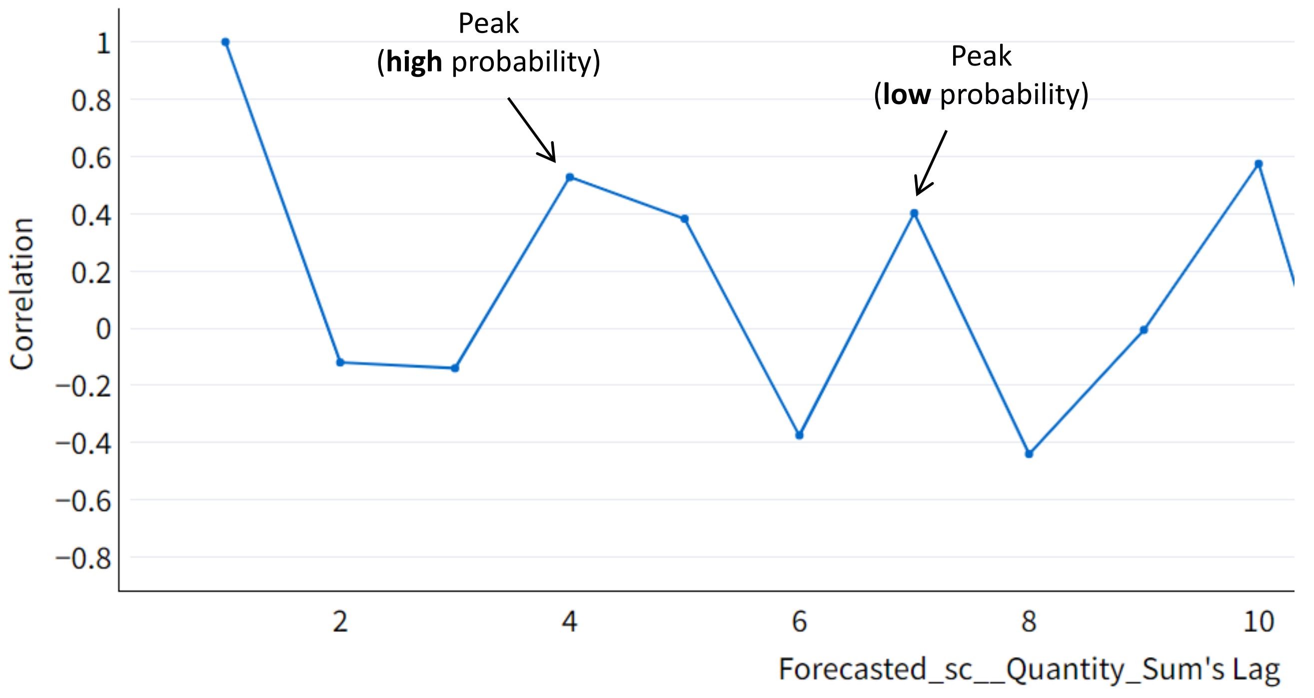


Figure 4: Python with Statsmodels and Plotly: Autocorrelation.

Original data comes consolidated by week. Autocorrelation is at peak at 4 and 10 weeks. Since 4 has less error, the data model will be re-consolidated monthly and the planning review period shall follow.

What demand forecast accuracy? **95%**

How much data is needed? **2 years of movement**

Learning Rate (Estimator Accuracy)

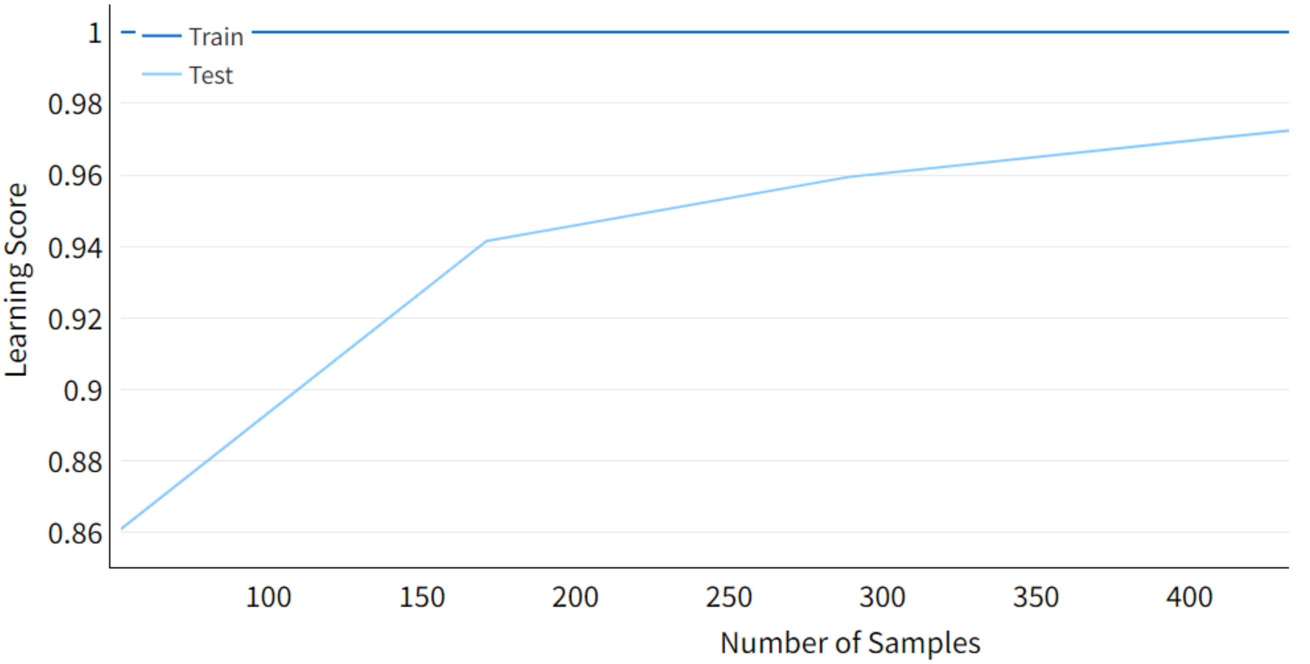


Figure 5: Python with Scikit-Learn and Plotly: Learning Rate.

95% is the new global demand forecast accuracy in comparison to the previous 75%. That is almost 30% improvement. High accuracy is achieved past around 300 samples or the equivalent of 2 years of data.

The metric used in this case was 1 - WMAPE (1 - Weighted Mean Average Percentage Error).

The winning model is a Decision Tree Regressor with One-Hot encoding and Quantile transformation.

Both Python and Excel apps were measured against test data, which is a subset of the original database.

What features to consider? COGS

Feature Importances / Predictive Power

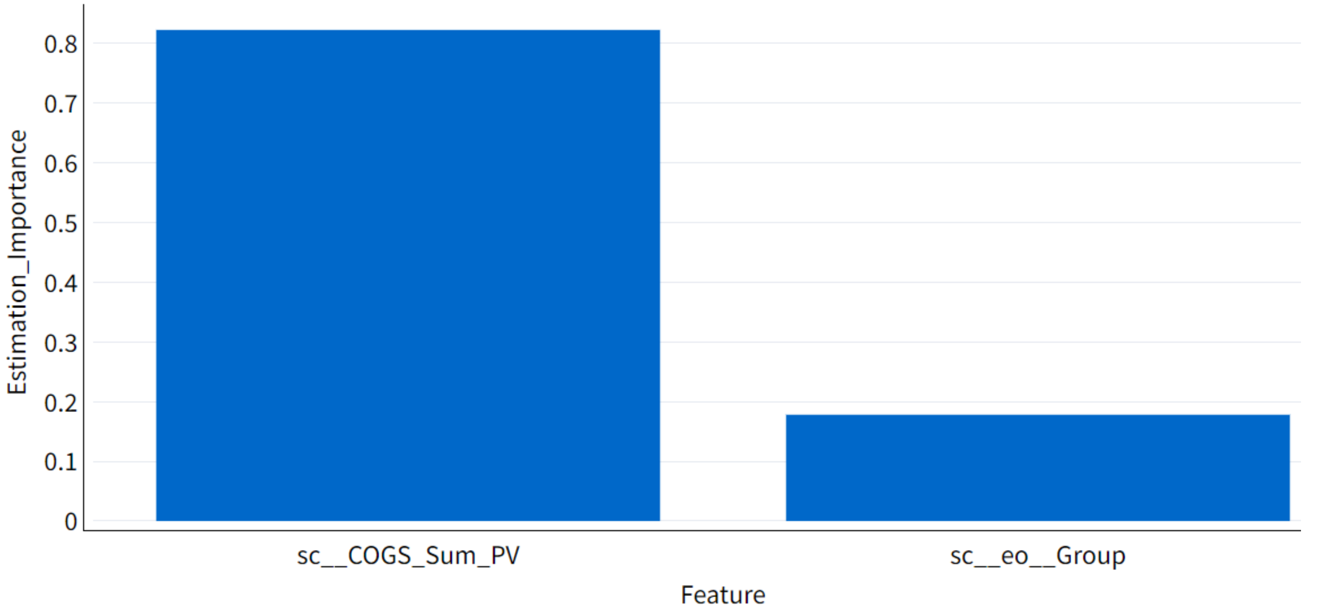


Figure 6: Python with Scikit-Learn and Plotly: Estimator Feature Importances.

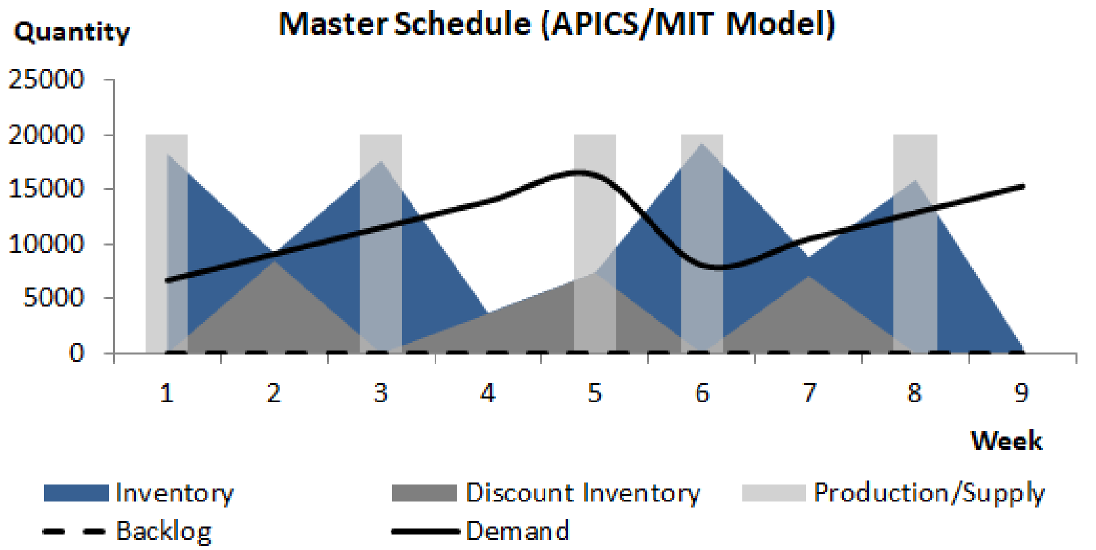
When COGS estimation is present, it is possible to achieve higher forecast accuracy.

COGS forecasting is commonly required by Finance departments. It could also contribute to the quality of Supply Chain forecasts.

Future Opportunity

To forecast COGS before forecasting demand.

What production master-schedule? ... for item 1



1 - Master Schedule (APICS/MIT Model)

Week	Inventory Before	Demand Forecast	Active Orders	Production /Supply	Inventory After	Backlog	Available To Promise	To Discount
1	5000	6700	0	20000	18300	0	18300	0
2	18300	9110	0	0	9190	0	9190	8475
3	9190	11525	0	20000	17665	0	17665	0
4	17665	13935	0	0	3730	0	3730	3655
5	3730	16345	0	20000	7385	0	7385	7385
6	7385	8075	0	20000	19310	0	19310	0
7	19310	10490	0	0	8820	0	8820	7100
8	8820	12900	0	20000	15920	0	15920	0
9	15920	15315	0	0	605	0		

2 - P&L [R$]

Week	Revenue	COGS	Backlog	Inventory	Warehousing	EBITDA	Service	ROI
1	13,552	7,088	0	19,361	5,121	1,343	100%	0.1
2	18,427	9,638	0	9,723	2,572	6,217	100%	0.6
3	23,312	12,193	0	18,689	4,943	6,176	100%	0.3
4	28,187	14,743	0	3,946	1,044	12,400	100%	3.1
5	33,062	17,293	0	7,813	2,067	13,702	100%	1.8
6	16,334	8,543	0	20,430	5,404	2,387	100%	0.1
7	21,219	11,098	0	9,331	2,468	7,652	100%	0.8
8	26,093	13,648	0	16,843	4,455	7,990	100%	0.5
9	30,978	16,203	0	640	169	14,606	100%	22.8
Sum:	211,164	110,448	0		28,242	72,474		
Average:				11,864			100%	
Std Dev:	27%							
							ROI:	0.7

Figure 7: Excel: APO's Master-Schedule Simulation.

What inventory reduction?
What working capital release? ... for item 1

50% Working Capital released

An APICS/MIT Master-Schedule model was adopted to receive the forecast for item 1 and calculate necessity. Then, a P&L is calculated for every line and the final ROI is presented.

Simulation Parameters

Reactor capacity is 20 ton, the produced item is stored in drums weighting 200 kg each when full, 4 drums per pallet. Warehousing cost is 25% of the COGS. Backlog cost is double the COGS. Data on active orders is not available so, it is presumed zero. Starting inventory for item 1 is unknown so 5,000 kg is assumed.

Reconsolidating the data model from month to week increases granularity, which decreases accuracy. However, a tailor made model, specific for weekly forecasting, managed to achieve 91% accuracy, when a COGS estimation is present and 85% without COGS. Alternatively, the monthly forecast could be divided by 4.3 for a weekly demand approximation.

What warehouse capacity? ... for item 1
Bonus: What EOQ - Economic Order Quantity? ... for item 1

The Excel APO calculates some common Supply Chain metrics like EOQ - Economic Order Q

t-Q Algorithms for 20 ton capacity

Cost per Order/Setup [R$] = 200.00
EOQ/EPQ [kg] = 12,565
Order Frequency [periods] = 1.08
Confidence Interval = 99%
Statistical Safety Stock [kg] = 26,984
Expected Average Inventory [kg] = 27,950

Req. Warehouse Capacity [kg] = 37,618
Req. Warehouse Capacity [pallets] = 47
POQ [kg] = 9,668

47 pallets

12,000 kg

Figure 8: Excel: APO's Supply Chain Metrics.

What service-level?
What inventory target?
What capacity?
What ROI?

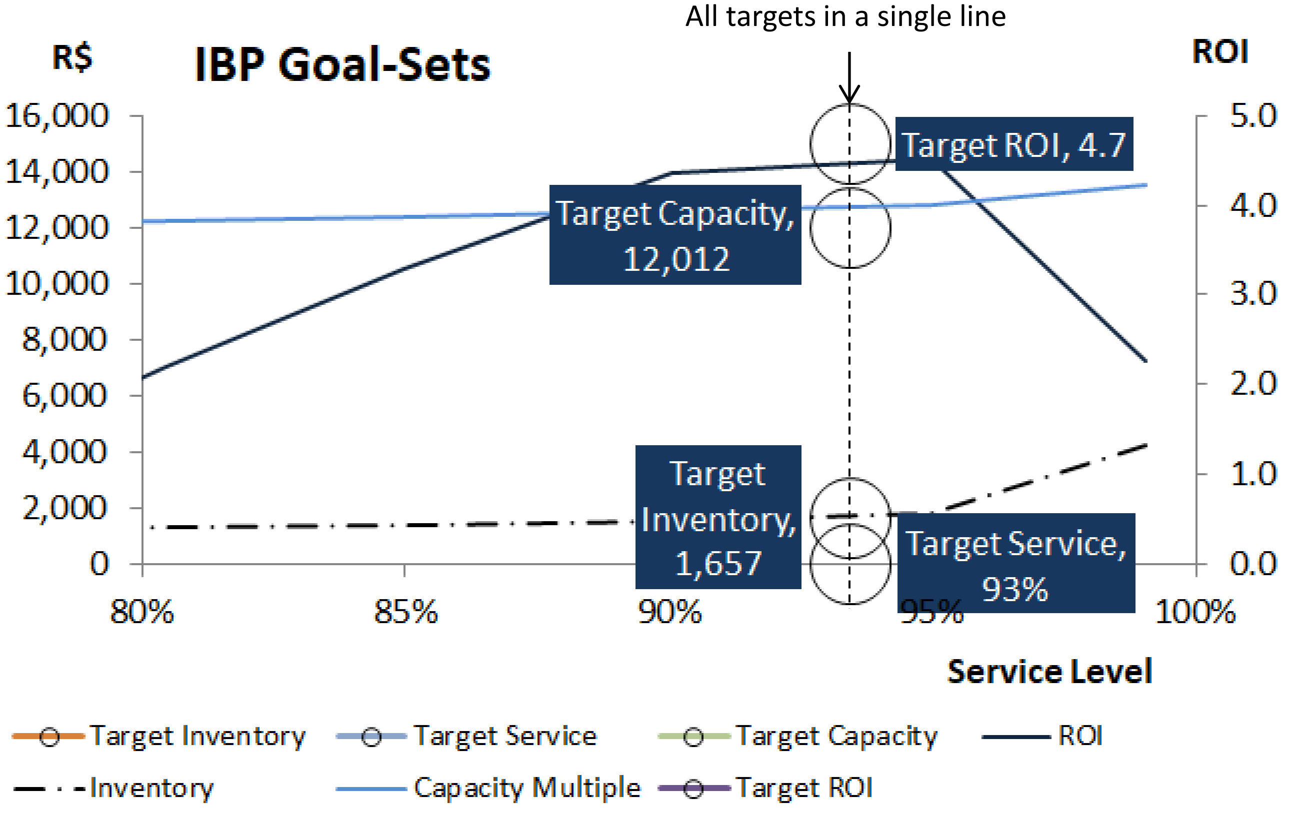


Figure 9: Excel: APO's Goal-Setting Module.

All IBP targets at once!

Continuing the exercise for item 1, a reactor capacity of 12 ton is able to deliver 4.7 ROI in comparison to 0.7 ROI from the current 20 ton reactor. It happens because big batches lead to slow movers.

The simulation optimum is better than the EOQ in terms of ROI because it takes into account the "hidden" constraints of that particular operation, which are not included in the EOQ algorithm. Although both quantities are close, the same is not true for ROI, which is partially explained by the fact that item 1 is a commodity.

Judging by item 1 alone, the suggestion is to replace the current reactor by a smaller one.

Corporate Goal-Setting (OKR/MBO)

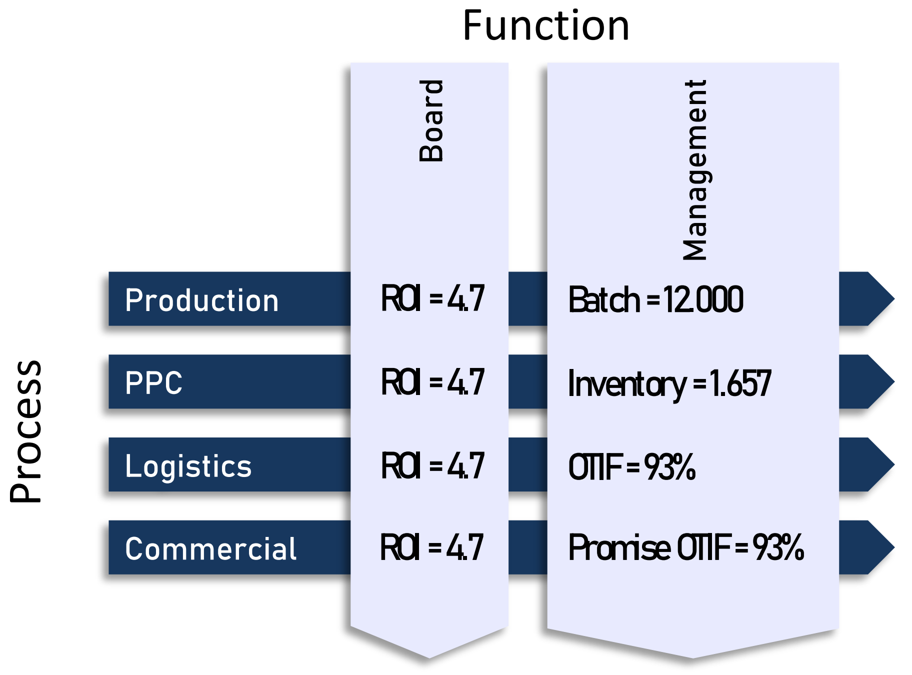


Fig. 10: Simplification of the Corporate Goals System Matrix corresponding to the IBP.

IBP targets were derived into corporate goal-sets and distributed according to the available structure.

What price?

Pricing Fine Tuning for Maximum Profit at Different What-If COGS Scenarios

year	month	Quantity	BU	Item	Elasticity	Commissions and Taxes	Un COGS	Interest	Profit Margin	Un Price
2012	1	189,507	BU 1	Item 1	1	0.33	3.00	0.50	0.30	3.48
2012	1	196,621	BU 1	Item 1	1	0.33	3.00	0.50	0.25	3.42
2012	1	190,912	BU 1	Item 1	1	0.33	3.00	0.50	0.20	3.41
2012	1	189,507	BU 1	Item 1	1	0.33	3.00	0.60	0.30	3.48
2012	1	196,621	BU 1	Item 1	1	0.33	3.00	0.60	0.25	3.42
2012	1	190,912	BU 1	Item 1	1	0.33	3.00	0.60	0.20	3.41
2012	1	189,507	BU 1	Item 1	1	0.33	3.30	0.50	0.30	3.48
2012	1	196,621	BU 1	Item 1	1	0.33	3.30	0.50	0.25	3.42
2012	1	190,912	BU 1	Item 1	1	0.33	3.30	0.50	0.20	3.41
2012	1	189,507	BU 1	Item 1	1	0.33	3.30	0.60	0.30	3.48
2012	1	196,621	BU 1	Item 1	1	0.33	3.30	0.60	0.25	3.42
2012	1	190,912	BU 1	Item 1	1	0.33	3.30	0.60	0.20	3.41

Fig 11: Orange: Price Suggestions.

There is a different maximum profit price suggestion for each scenario. And it can be adjusted in real time.

Price variation affects the demand, which affects the master-schedule, which affects inventory, which affects ROI.

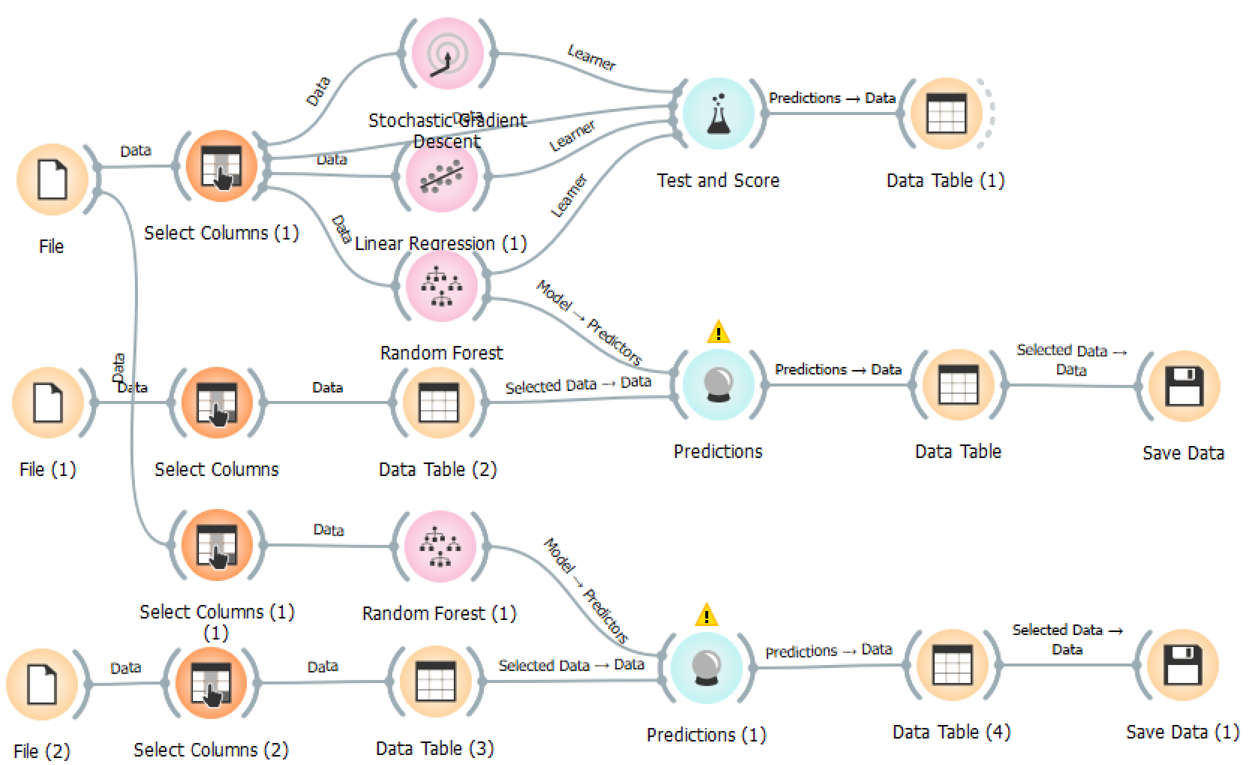


Fig 12: Orange data mining freeware canvas.

What surprising insights?

Specialty Items

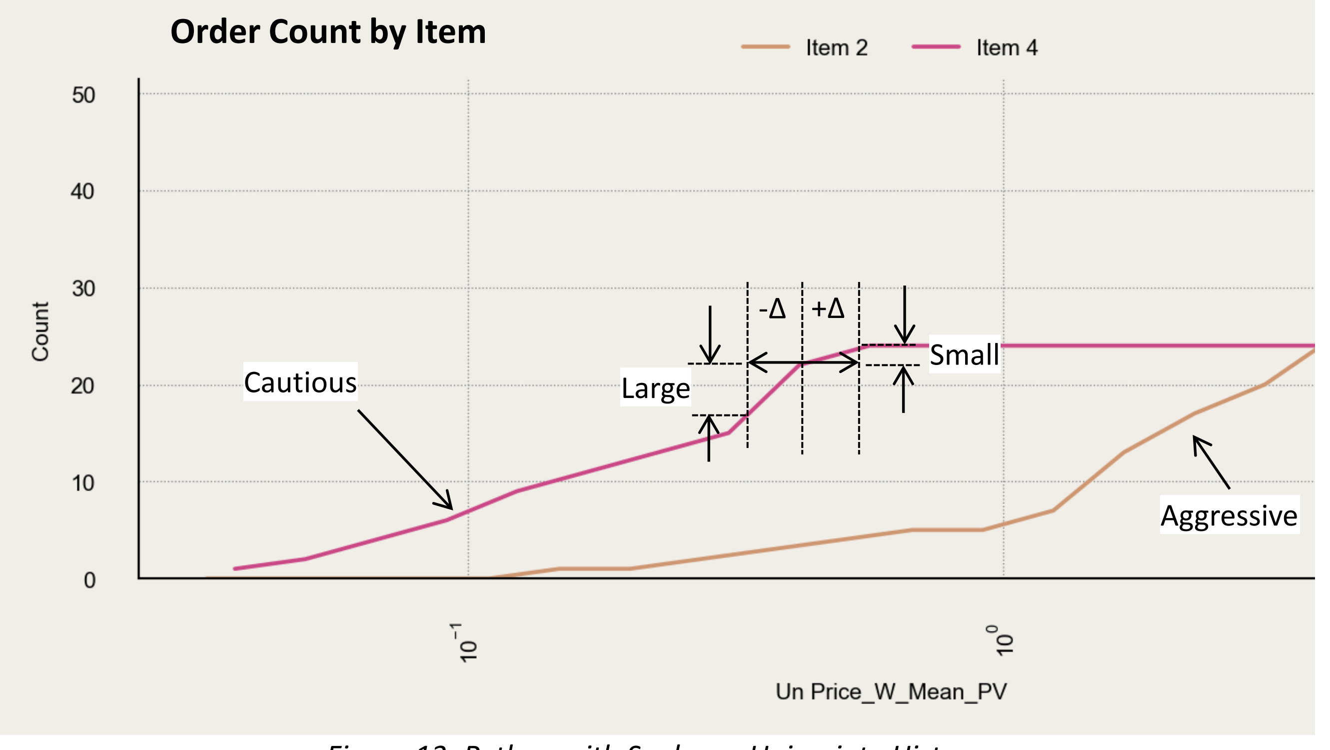


Figure 13: Python with Seaborn: Univariate Histogram.

For item 4, a delta increase in price represents a small number of sales events, while the same delta decrease represents a large number of sales events. So, it is preferred to decrease prices to generate more sales events. This is considered a cautious profile.

The aggressive profile allows prices to increase, indicating differentiation potential, or room for the designation of specialty items.

What surprising insights? Sales Insights

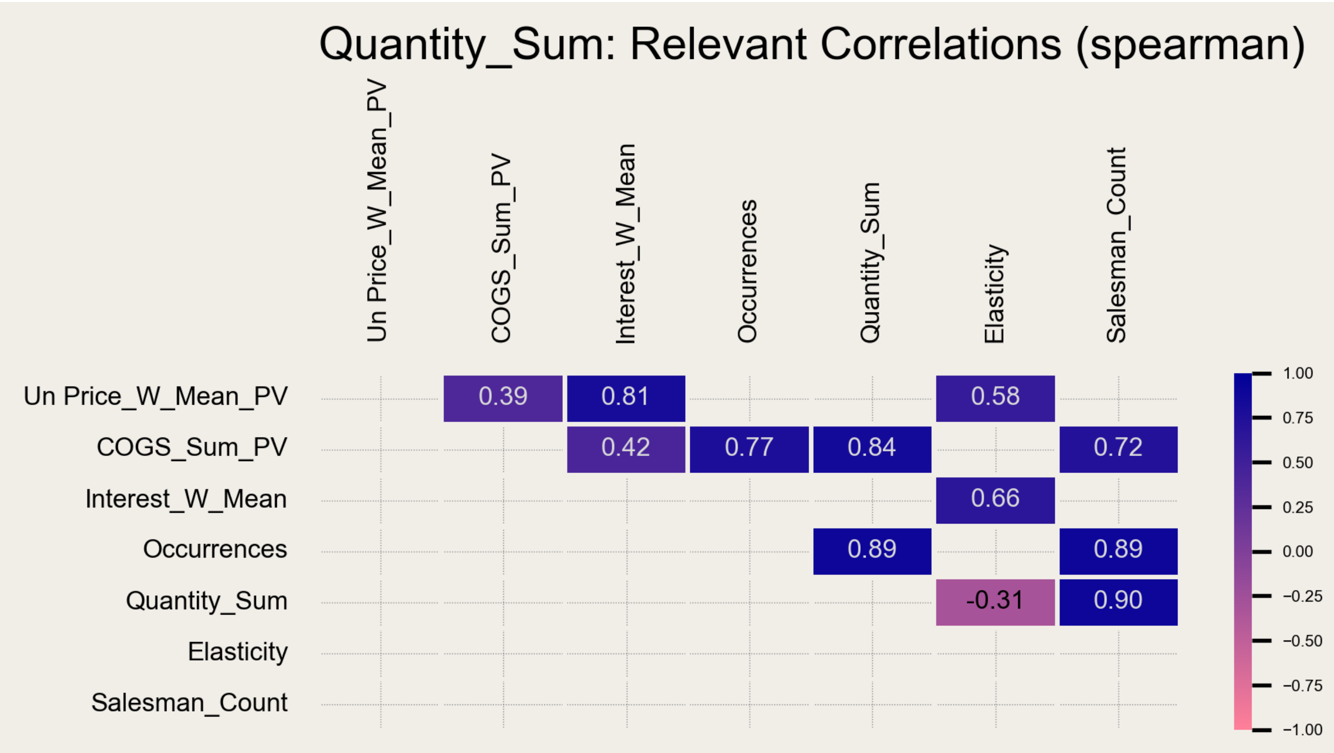


Figure 14: Python with Scipy: Spearman Correlations.

Conditional Probabilities

		Salesman_Count	
Un Price_W_Mean_PV	Quantity_Sum	(1.0, 8.0]	(8.0, 15.0]
{"left": 0.07, "right": 2.1}	{"left": -77.2, "right": 148830}	0.535	0.095
{"left": 0.07, "right": 2.1}	{"left": 148830, "right": 297440}	0	0.477
{"left": 2.1, "right": 4}	{"left": -77.2, "right": 148830}	0.465	0.095
{"left": 2.1, "right": 4}	{"left": 148830, "right": 297440}	0	0.333

+43%

Figure 15: Python with Pandas: Conditional Probabilities.

Sales headcount associated with volume could indicate sales concentrated at big clients with high volumes and/or better commissions, given that, it is easier to sell at lower prices. It is also more convenient to do so in less visits.

On the high quantity scenario, there are 43% more representatives than when prices are lower.

Is this practice minimizing the company's profitability?

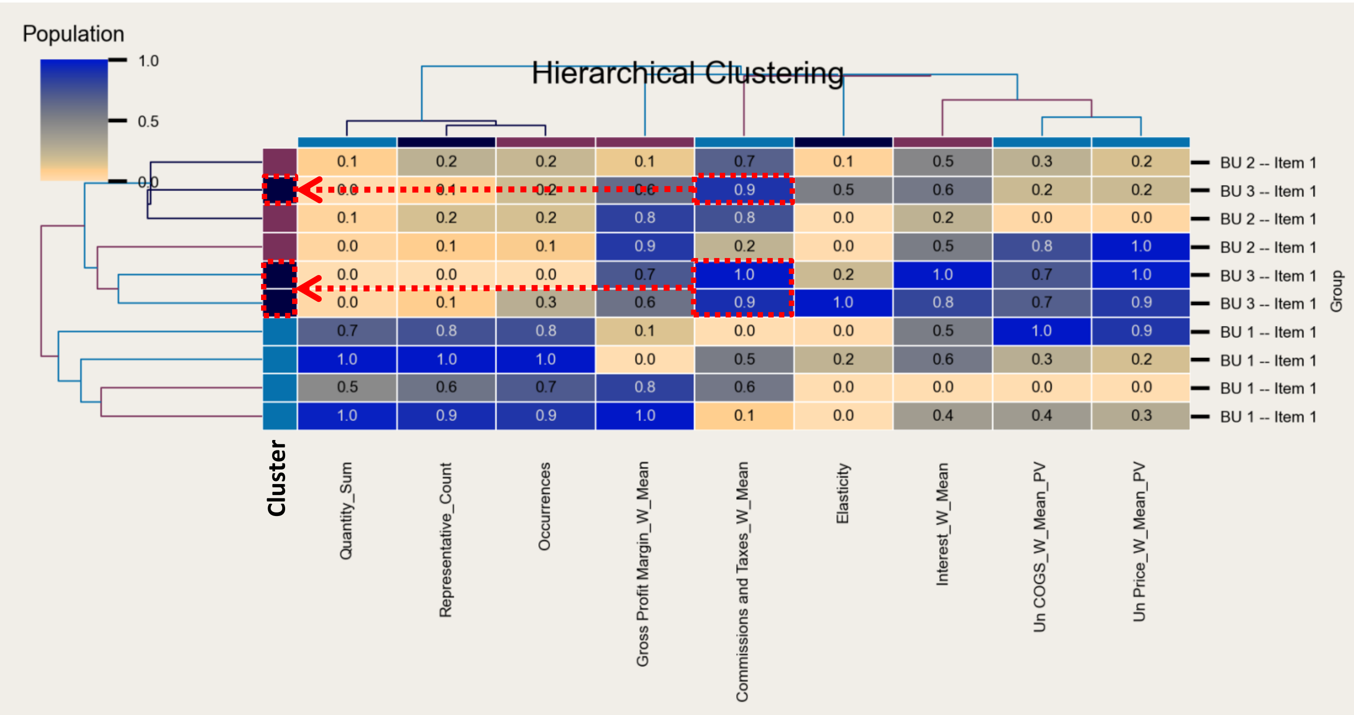


Figure 15: Python with Seaborn: Hierarchical Clustering.

Commissions and Taxes is the "hidden pattern" suggesting a cluster at BU's 2 and 3.

In the other hand, at BU 1, representative count and total volume play a more significant role.

What surprising insights?

Complementary / Cannibalization

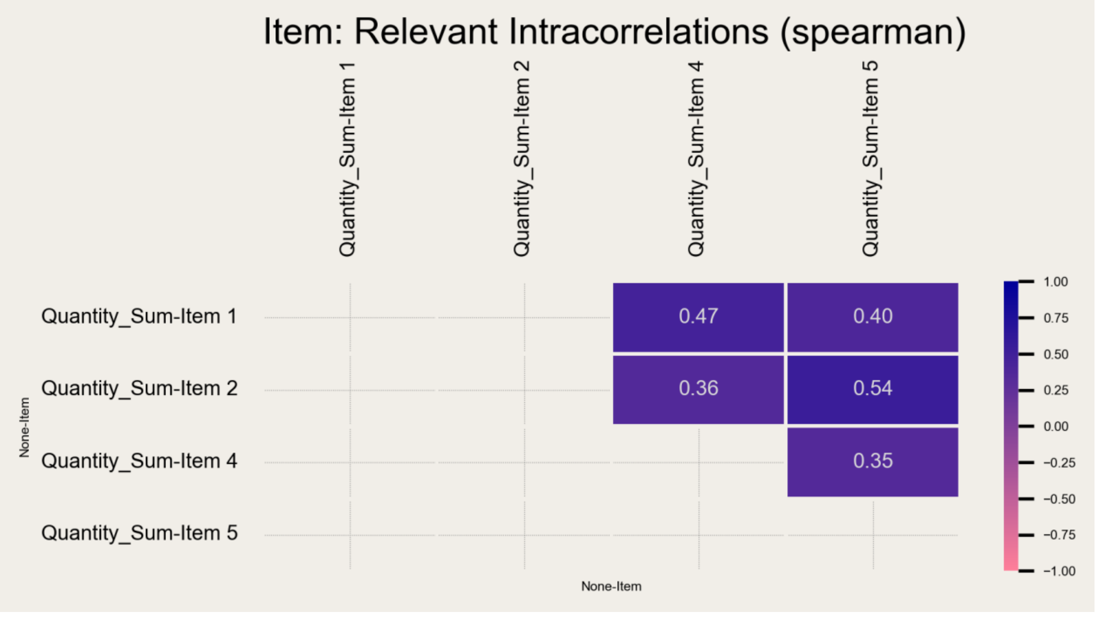


Figure 16: Python with Seaborn: Cannibalization versus complementary items.

Items 2 and 5 are complementary. The sales of one helps to sell the other. Should this relation be negative and one would be cannibalizing the other.

Final Considerations

As mentioned before, the objective of this business case was to demonstrate Orkideon´s capacity to answer Supply Chain questions and to establish goal-sets. All proposed questions were answered and surprising insights emerged. Finally, new doors were open for further exploration.

Machine Learning applied to Supply Chain Planning is a green field because knowledge is still being borrowed from other areas.

With the right knowledge, methodology and tools, decision making is fast and easy. Companies can become more responsive and one can go back and forth from operations to strategy like never before.

This whole case took only a few minutes but, the project allowing it to happen is already 2 years old. It is well known that most executives currently don´t have the "patience" to invest in long-term development projects like this, but the results are evident and the payback is unquestionable. The good news is that, now companies don´t have to go through all the process because Orkideon fills the gap.

I hope you enjoyed reading this business case as much as I did preparing it.

Contact us and know more.

Sincerely,

Thiago

Python app
https://www.orkideon.com/app-1