Defining Future Automation Needs for Storage and Picking Operations in a Grocery Distribution Center

by

Larissa Chow

and

Meghan Wood

Submitted on May 8, 2026 in Partial Fulfillment of the Requirements for the Degree of Master of Applied Science in Supply Chain Management at the Massachusetts Institute of Technology

ABSTRACT

Picking and palletizing goods is an expensive part of warehousing. One solution is to automate this process. However, making a sizable investment into automation without knowing whether it will be able to fulfill demand or the estimated payback period of that system poses a substantial risk. In order to address this problem for a grocery distribution center, the team designed a discrete event simulation based on historical order data. This was then used to create a flexible "black box" tool that simulates potential future states of the warehouse. The tool calculates the payback period of a proposed automation solution under different demand and operational scenarios for the candidate SKU portfolio across the ambient warehouse. The solution also specifies which of these SKUs should be prioritized for automation. With this information, the company can better understand the operational benefits of their proposed solutions and projected time to profitability.

Capstone Advisor: Miguel Rodríguez García

Title: Research Scientist

TABLE OF CONTENTS

1. INTRODUCTION 3
1.1 Motivation 3
1.2 Problem Statement and Key Questions 3
1.3 Project Goals and Expected Outcomes 3
2. STATE OF THE PRACTICE 4
2.1 Current Automation Challenges 4
2.2 Quantitative Advances 4
3. METHODOLOGY 5
3.1 Methodology Selection 5
3.2 Data Collection and Preparation 5
3.3 Assumptions 6
3.4 Model Application 7
3.5 SKU Optimization and Financial Evaluation 10
4. RESULTS AND DISCUSSION 13
4.1 Results 13
4.2 Implications 17
4.3 Limitations 18
5. RECOMMENDATIONS 19
5.1 Management Recommendations 19
5.2 Future Work 19
REFERENCES 20

1. INTRODUCTION

The project sponsor is Vegalsa-Eroski, a leading grocery chain in Spain. A region-specific subsidiary of its parent company Eroski, Vegalsa-Eroski operates over 285 stores within Galicia. These range in size from small convenience stores to large, big-box stores. They advertise fresh, local goods, making them a top choice for shoppers. Core to their promise to deliver high-quality produce, Vegalsa-Eroski's main distribution center is strategically located within its region of service. This distribution center serves as a central hub for the grocer to distribute over 15,000 different types of dry goods.

1.1 Motivation

Vegalsa-Eroski has four main motivations for initiating the capstone project: ongoing labor shortages, future labor legislation that could limit productivity, continuous company growth, and rapid technological improvement in the warehousing industry. Across the globe, there is an ongoing labor shortage for manual jobs, largely because many workers have been drawn to higher-paying, sedentary options (Mangelsdorf, 2020). As a result, many companies, including Vegalsa-Eroski, are finding their current employees aging out of their physical jobs and retiring without younger workers coming in to replace them. Additionally, the sponsor is anticipating that the European Union (EU) will implement new regulations capping the amount of movement performed and weight a worker is allowed to lift each day. This potential labor legislation would put a large strain on Vegalsa-Eroski's already tight team of warehouse employees.

In tandem with Vegalsa-Eroski's labor difficulties, the grocer is expected to grow at a rapid rate of 10%–15% per year. The company expects to increase their overall Stock Keeping Unit (SKU) portfolio and experience larger order volumes. While growth is welcomed and a sign of their flourishing business, the present distribution center's capacity is not enough to keep up with current demand, let alone any increase. Vegalsa-Eroski must anticipate their increasing customer needs, which could quickly become overwhelming if the warehouse cannot push pallets out on time. In order to fulfill any future surge in demand, Vegalsa-Eroski is looking to implement new automated warehousing technology. The rapid advancement of warehouse automation provides Vegalsa-Eroski with a wide range of options to optimize their distribution centers. Incorporating cutting-edge technology into their main distribution center would bridge the gap between expanding capacity and restrictive labor requirements.

1.2 Problem Statement and Key Questions

The distribution center's process flows currently have varying degrees of automation. Vegalsa-Eroski is seeking to improve their existing automation processes to offset the ongoing and future labor shortage, capitalize on expected growth, and increase throughput. This capstone project assesses new automation scenarios to increase the ambient warehouse's capacity and sustain future expansions. The chosen automation scenario must capture demand while the company grows. The motivating question to be answered is: How can Vegalsa-Eroski define the capacity and operational requirements for a new automation system that supports the transition of fast-moving SKUs to automated storage and picking while ensuring a sound payback period?

1.3 Project Goals and Expected Outcomes

Vegalsa-Eroski hopes to obtain managerial insights and recommendations on future automation needs within their ambient distribution center. The new system needs to automate part of the labor-intensive manual picking and palletization processes while working concurrently with other areas of the warehouse.

The project deliverable is a decision-support tool that enables Vegalsa-Eroski to evaluate prospective automation systems before committing to a purchase. The tool simulates warehouse

operations against user-defined parameters and produces a financial analysis of the automation investment, giving the company a clear basis for assessing whether a given solution is viable over their planning horizon. While recommending a specific automation system was not within scope, the tool provides Vegalsa-Eroski with a clear understanding of their options and estimated payback period. Ultimately, these insights aid in picking an automation solution that aligns with the company's long-term growth strategy.

2. STATE OF THE PRACTICE

To investigate how Vegalsa-Eroski can reduce reliance on their labor force as well as reduce picking costs while increasing throughput, the team reviewed literature across several areas of warehousing. This included examining the challenges of automated warehousing to understand why technological change is pushed in the industry. The team then researched the path to implementing new automation methods and the quantitative testing that must be done to ensure that an investment in automation is profitable.

2.1 Current Automation Challenges

Since warehousing remains an expensive but necessary process, companies are looking to automation to reduce costs (Peixoto et al., 2016). Picking, which accounts for 55% of the entire cost of warehouse operations (Custodio & Machado, 2019), is a prime area for automation. Warehouse automation solutions began to be developed in the 1960s by a select few suppliers, but recently the market has exploded, and automated solutions are more accessible than ever before (de Koster, 2018).

Despite the rapid advancements in automation technology as well as the increase in automation solutions suppliers, investing in automation remains a risky decision (Peixoto et al., 2016). Return on investment often takes years, during which prospective technology may have already surpassed the machine that was purchased. Another risk is that a supplier could go out of business, leaving any companies that own their solution without a way to repair or update the automation. Vegalsa-Eroski has already dealt with the former when a company providing an automation system refused to expand the existing model from 2018 because it was out of date. The grocer needs to ensure a minimal payback period when examining different automation systems to limit these risks.

2.2 Quantitative Advances

Due to the high level of investment that automation solutions require, determining a solution that balances key performance standards and costs poses a challenge. The weight of such decisions means it is imperative to test whether proposed solutions meet any requirements the sponsor specifies. The most popular technique for assessing an automation solution's performance is simulation. Simulation allows the user to analyze warehouse design by testing different metrics and subsequently making operational decisions (Saderova et al., 2022; Peixoto et al., 2016). The simulation technique most suited for modeling warehouse processes, including picking and palletizing, is discrete event simulation (DES). According to Antoniades and Stamelos (2022), DES is a good fit for this type of project since it can model complex systems with concurrent events such as the multiple overlapping processes in Vegalsa-Eroski's warehouse.

DES also allows the user to adjust metrics, such as throughput rate, in the simulation when performing iterative runs within the "black box." The black box acts as a simplified stand-in for an automation solution where a user can easily adjust inputs, run the simulation, and see an estimated evaluation of that automation solution.

With profitability top of mind for the company, it is crucial to study the costs associated with both manual and automated picking and palletizing. This project will examine the payback period, or the time required for the money saved from using an investment to equal the initial expenditure (Ceroni, 2009). Specifically, when the projected baseline annual labor cost equals the cost of the machine and maintenance, payback is achieved. Vegalsa-Eroski wants to minimize the payback periods of the automation solutions they will consider to ensure a sound investment.

3. METHODOLOGY

To address the complex warehouse operations and the financial specifications provided by Vegalsa-Eroski, the team employed a hybrid quantitative framework. This approach utilized DES as the primary engine and a black box analytical optimizer with a built-in financial evaluation to assess the payback period. The tool was implemented as a web-based interface using Gradio.

3.1 Methodology Selection

Since the final automation solution is unknown, the team opted for a black box modeling approach. This approach does not need to know the specifics of how the automation system's hardware works. The model takes input parameters and outputs projected labor savings, a multi-year cash flow analysis, machine utilization rates, and the estimated breakeven period for the automation investment. The input parameters are user-defined controls that prompt Vegalsa-Eroski to adjust strategic variables, such as the demand-growth rate, the estimated total cost of automation (CapEx), and the estimated automation production (cases per hour). If the user opts not to input values for some of the prompted metrics, such as labor cost per hour, the simulation is designed to default to current operational and financial baseline metrics. By structuring the optimizer tool with the ability to run variable inputs, the model functions as a dynamic tool for decision-making. This sets up Eroski with a tool that can be adaptable as their automation venture develops.

DES is a good choice for simulating the picking and palletizing process because it allows for the analysis of system behavior over time. Other methods may assume a continuous state of operation, which lends itself to continuous event simulation. However, warehouse environments are often in flux due to changing demand, providing an appropriate reason for using DES. This approach is more realistic for automating picking and palletizing since it also allows for easy identification of throughput capabilities and operational limits.

3.2 Data Collection and Preparation

The ambient section of this large distribution center houses the three processes this capstone focuses on. There is an Automated Storage and Retrieval System (AS/RS) stacker crane that feeds picking aisles with high-turnover goods and a supplementary high-turnover liquid area (HTLA) for products such as beer and soda. There is also a conventional picking area where pallets are packed based on the aisle of the store they are heading to. This is a manually intensive process with a lower, variable turnover rate. Tertiary among the main processes is the small cross-docking area where pallets are separated by their individual SKU. The pallets are then transferred to the dispatch dock.

The data used in this simulation was broken down into six sources: historical orders, SKU master list, automation candidate list, flow logic table, customer schedule, and pallet stacking table. For historical orders, the team received one year's worth of raw sales order data for all unique SKUs, from November 2024 to October 2025. The twelve-month dataset serves as the simulation's base period, from which all annual projections are derived. This data was then filtered to only include order data from the 992 SKUs on the automation candidate list. These designated SKUs were a

selection from across the AS/RS, conventional, and cross-docking flows and were chosen by Vegalsa-Eroski based on factors such as size and shape that would allow for automation.

The flow logic table defines the standard pick rate in cases per picker per hour for each warehouse flow. The customer schedule maps each customer to their expected order transmission and completion information by day of week and hour. This information was used to derive picking deadlines. Finally, the pallet stacking table assigns a stacking layer (1–8) to each SKU, which was used to determine the picking order for SKUs within a rainbow pallet.

Across all data, the SKU references and flow names were normalized to ensure consistent joining across order history, master data, and the automation candidate list. All data was cleaned to impute any missing values based on averages, industry standards, and sponsor-specified information.

3.3 Assumptions

In order to manage the scope of the simulation, the team made the following assumptions:

Labor. Headcount is always rounded up to the nearest integer, with a minimum of one worker per flow. Labor cost is uniform across all flows, shifts, and worker types, meaning there is no difference for night shifts, weekends, or diverse skills. Moreover, the simulation applies the wage growth percentage as a flat compound annual rate equal for all workers.

Workforce. To quantify the labor needed for the baseline simulation, the model assumes there is no cap on the number of workers available. If an order arrives and requires the workhours of 10 simultaneous pickers, then 10 pickers are assumed available. This means the model is computing total labor hours needed to pick these SKUs rather than a staffing constraint.

Shift operation. The simulation has three defined shifts: morning (06:00–14:00), afternoon (14:00–22:00), and night (22:00–06:00). The shift schedule varies by day of the week: Monday through Friday operates three daily shifts, Saturday operates two (morning and afternoon), and Sunday operates one (night). Each shift runs for 7.5 hours and no partial shifts are modeled.

Pallets. All pallets in the model are assumed to be European-sized pallets (1.2m x 0.8m), which have a maximum weight capacity of 1,500 kg and a maximum height capacity of 2 meters.

Order release. For base task release and execution within the simulation, each sales order is assigned a completion date and time based on the customer schedule. The order release time is set to one shift prior to the completion shift, which provides pickers one shift to complete the order. Look-ahead smoothing is also employed in the simulation to allow orders to be pulled forward into an earlier shift if the capacity is available, reducing idle time. Any order that is picked across multiple shifts is split, and labor minutes are allocated to each shift proportionally.

Operational assumptions. Pick rates remain constant per flow per shift with no variability. The AS/RS picking aisles and corresponding HTLA have a manual pick rate of 190 cases per hour, the conventional process has a manual pick rate of 130 cases per hour, and the cross-docking area also has a manual pick rate of 130 cases per hour. Automation throughput is constant at the user-input specified cases per hour, with one half hour of downtime per shift to represent setup time and maintenance windows. The automation machine has a 90% utilization cap, factoring in any error that may occur during the daily processes. The numeric differences in pick rates across flows directly determine the labor minutes per SKU in each area. As a result, the automation optimizer prioritizes SKUs by total labor intensity instead of designated flow or SKU count. This relationship is examined further in Section 4.1.

Inter-order transition. An inter-order transition time of 10 minutes (4 minutes to pallet wrap + 6 minutes to transfer to outbound dock) is added to each pallet in the shift once the tasks are completed. This movement time is assumed to have no variance.

Availability and demand. While a case is in the system, it is assumed that it will always be available when a worker goes to pick the case. The annual demand-growth percentage is uniform across all SKUs, with no seasonality or customer variation. The entirety of modeled demand is associated with the 992 company-specified SKUs; other sales orders are excluded entirely. All cases from each SKU will be automated if they are in the optimal subset.

Financial assumptions. There are no installation, training, or transition costs added to the simulation. CapEx is assumed to be a single up-front payment at Year 0 with no financing costs, interest, or lease structure. Annual maintenance costs are a fixed percentage of the original CapEx each year, meaning these costs do not adjust with age or utilization.

3.4 Model Application

Depending on the stage of vendor engagement, the model operates in two different modes: pre-vendor and post-vendor. One of the two modes activates depending on the level of information that is input by the user. The pre-vendor mode activates when the user knows the total automation budget but is not aware of the vendor's exact machine pricing. This mode derives an implied cost per machine based on the amount of throughput capacity needed for Year 1. It then calculates the breakeven point. Because pre-vendor mode is designed to determine an allowable machine cost rather than constrain the project to a pre-priced machine, it will always recommend the full candidate portfolio for automation. This gives decision-makers a clear view of the maximum savings potential, total capacity needs, and an implied budget threshold before entering vendor negotiations.

The post-vendor mode activates once the user inputs both an automated machine count and a unit cost greater than zero. For example, a user could input the specifications for two picking arms priced at 200,000 € each. Instead of recommending the full candidate portfolio for automation, post-vendor mode iterates through the ranked SKU list and adds machines only when labor savings from the additional automated volume justify the cost of that machine within the simulation time period. This logic ensures the optimal subset of SKUs is selected, reflecting real-world capital constraints. When using both the pre- and post-vendor mode sequentially, the user can reach a holistic procurement decision. The pre-vendor analysis will establish the economic potential and capacity requirements, whereas the post-vendor analysis will evaluate the price per machine to identify the optimal automation scope.

In either mode, users can choose to override default pick rates with custom cases-per-hour (cph) values for each warehouse flow. The tool also allows users to selectively include warehouse flows in the analysis, enabling focused evaluation of automation opportunities in specific areas.

The model is executed as a sequential pipeline in five steps. The first step takes in user inputs. The second step reads the order history data from the simulation period to construct picking activity at the SKU and shift level. The third step is the manual process (or baseline) construction via discrete event simulation. The fourth step is the Pareto-ranked optimization calculation, which lists all 992 automatable SKUs in order, from highest labor hours to lowest. This list then determines the cutoff point for the optimal subset of SKUs that minimizes the payback period. The last step is a multi-year financial projection that applies demand and wage growth to produce a cash-flow table and breakeven analysis along with accompanying visual charts. The financial projection output will assist Vegalsa-Eroski in making three key decisions: how much automation capacity to procure, whether the investment will see a payback within an acceptable planning horizon, and which SKUs to automate (in post-vendor mode).

The mathematical formulations in the discrete event simulation are as follows:

Time and capacity variables:

- η = Automation utilization cap = 0.90 (accounts for setup, errors, maintenance)
- R_a = Manual pick rate for flow a (cases / picker / hour)
- R_{auto} = Automation throughput (cases / hour, user input)
- C_M = Annual machine capacity per machine (cases)

Pallet building variables:

- w_i, l_i, b_i, h_i = Weight, length, breadth, and height of case i
- W = Cumulative weight of cases placed on a given pallet (kg)
- W_{max} = Maximum allowable pallet weight = 1,500 kg
- A_{max} = Maximum allowable pallet footprint area = 0.96 m²
- H_{max} = Maximum allowable pallet height = 2.0 m
- P_{min} = Theoretical minimum number of pallets pre-allocated
- A_L = Cumulative footprint area of the layer currently being built
- H_L = Cumulative stacked height of the current pallet

Operational and flow variables:

- $a \in F$ = flow index, where $F = \{AS/RS, HTLA, Conventional, Cross-Docking\}$
- B_p = Set of (flow, count) pairs on rainbow pallet p
- $n_{p,a}$ = Boxes from flow a on pallet p
- $T_{rainbow}^{(p)}$ = Manual pick time for rainbow pallet p (minutes)
- $T_{auto}^{(p)}$ = Automated pick time for rainbow pallet p (minutes)
- t_{auto} = Automated processing time per case = $60/R_{auto}$ (minutes)
- L_a = Total simulated manual labor minutes accruing to flow a (over the simulated period)
- L_{tot} = Total simulated manual labor minutes $= \sum_a L_a$

Daily Capacity Calculations. The model treats workers' availability as unlimited within each shift (Section 3.3), so per-area human capacity is calculated post-simulation. Annual machine capacity is calculated upfront at the start of the pipeline and carried forward into the financial evaluation described in Section 3.5. The average number of shifts per labor day across the operating week is $S^- = (5 \cdot 3 + 1 \cdot 2 + 1 \cdot 1)/7 \approx 2.57$, with $H_s = 7.5$ hours per shift and $D_y = 242$ labor days per year. Annual machine capacity per machine is therefore: $C_M = D_y \cdot S^- \cdot H_s \cdot \eta \cdot R_{auto}$

Pallet Building and Volumetric Constraints. Cases are sorted by stacking layer (8 = bottom, 1 = top) and packed onto pallets subject to three constraints: weight, footprint area, and stack height. The starting pallet count is estimated from weight alone. Cases are then packed sequentially by stacking priority onto the lightest eligible pallet, building layer by layer until the constraints of the pallet are reached. The layer is then closed, and the pallet moves on to the next layer. A new pallet is added as needed when a box cannot fit on any existing pallet without breaching any other pallet's weight and height limitations.

Minimum theoretical pallets (pre-allocation)	$P_{min} = \lceil \frac{\sum w_i}{W_{max}} \rceil$
Layer footprint (running)	$A_L = \sum(l_i \cdot b_i)$
Vertical height tracking	When $A_L > A_{max}$, the layer closes: $H_L \leftarrow H_L + max(h_i)$, $A_L \leftarrow 0$. A new pallet opens when no existing pallet satisfies $W \leq W_{max}$ and $H_L + h_i \leq H_{max}$

Task-Duration Math. Order lines are split into two flows. Full supplier pallets go straight to dispatch with a 10-minute transfer. Partial-pallet cases are picked and combined into rainbow pallets. Rainbow pallets are mixed-SKU and may contain boxes from multiple flows; their pick time is the total time needed to pick every box on the pallet across all flows.

Full-pallet handling (per pallet)	$T_{full} = 10$ min
Manual pick time for rainbow pallet p	$T_{rainbow}^{(p)} = \sum_{a \in B_p} \frac{n_{p,a}}{R_a} \cdot 60$
Automated pick time for rainbow pallet p	$T_{auto}^{(p)} = \left(\sum_{a \in B_p} n_{p,a}\right) \cdot t_{auto}$, where $t_{auto} = 60/R_{auto}$

Inter-Order Transition. Each rainbow pallet adds $T_{wd} = 10$ min. (4 min. wrap + 6 min. dock transfer) to its primary flow when it completes. Full pallets use T_{full} , which equals T_{wd} numerically but represents a different physical activity. For a rainbow pallet p with primary flow a: $L_a \leftarrow L_a + T_{rainbow}^{(p)} + T_{wd}$. For a full pallet routed to flow a: $L_a \leftarrow L_a + T_{full}$.

Resource Utilization. Manual utilization is computed per shift after the DES assigns all labor minutes to (shift, flow) buckets. Values above 100% mean workload exceeds capacity; look-ahead smoothing reduces this by pulling work forward from shifts with surplus. Automation utilization uses annual cases against capacity instead of minutes, since automated processes don't follow shift schedules. M is the number of machines installed, and α is the annualization factor defined in Section 3.5.

$$Q_{auto}^{annual} = \alpha \cdot \sum_{p} \frac{T_{auto}^{(p)}}{t_{auto}} \qquad U_{auto} = \frac{Q_{auto}^{annual}}{M \cdot C_M}$$

3.5 SKU Optimization and Financial Evaluation

The model evaluates user inputs and determines the most viable automation split by optimizing around the Pareto principle. The work hours required, or labor-intensity metric, for each SKU is calculated by summing up the number of picking minutes needed to move all of the cases for that SKU over the base year. Once all 992 SKUs have been sorted by labor intensity, the simulation calculates their potential payback by moving down the list of SKUs and calculating the labor savings achieved by automating that item. As the model moves farther down the list, lower-impact SKUs add smaller value savings to the calculation. The optimal number of SKUs automated ensures the lowest payback period.

In pre-vendor mode, the optimal SKU selection will be 992 since the implied machine cost is derived from the full-scope capacity requirement. In post-vendor mode, a more exact number of optimal SKUs will be identified to justify the vendor's machine pricing within the simulation horizon. For most scenarios in the post-vendor mode, the optimal number of SKUs automated will not be

100%. The model calculates the number of machine hours required for the case volume being automated in the simulation. For the SKUs remaining in the manual process, the model re-runs the simulation similar to the baseline, now with any relevant simulation inputs. A related metric from the automated process is the utilization rate, which is calculated as the fraction of the machine's monthly capacity consumed by the determined automated SKU case volume.

Since the simulation uses one year's worth of historical sales data, the model must calculate financial projections for the user-dictated number of years. Both the baseline and automated scenarios are multiplied by inputted percentage growth for demand and labor.

The cash flow for each year is calculated as annual savings minus annual maintenance. Cumulative cash flow is tracked starting from the negative initial CapEx at Year 0, and continuously incremented each succeeding year. The breakeven point occurs where the cumulative cash flow total flips from negative to positive, indicating that the investment has been paid off.

The simulation does not require the automation solution to fulfill all years of demand in the financial evaluation. This allows the user to make a more informed decision about when to invest in additional machine capacity. Here, the user can examine a range of sensitivity analyses that display information such as the minimum number of machines needed to fill all demand across the desired time period. This empowers decision-makers to decide between an investment with a faster breakeven period or an investment with greater longevity.

The mathematical formulations underlying these calculations are as follows:

Variables (additional to Section 3.4)

- α = Annualization factor $= 12/n_{months}$ (scales simulated period to one year)
- l_s = Simulated labor minutes attributable to SKU s (rainbow pick + share of wrap)
- q_s = Simulated cases of SKU s
- ρ_s = Labor-per-case score for SKU $s = l_s/max(q_s, 1)$
- π_s = Rank index of SKU s (0 = highest priority for automation)
- i = Step index in the Pareto walk = number of SKUs automated
- Q_i^{auto} = Cumulative annualized cases automated through step i
- L_i = Cumulative annualized labor minutes saved through step i
- $T_{BE,i}$ = Single-year analytical breakeven at step i
- i^* = Optimal step $= argmin_i\, T_{BE,i}$ among finite values
- K^0 = Annualized baseline labor cost $= (L_{tot} \cdot \alpha/60) \cdot c_h$
- $K_{i^*}^{auto}$ = Annualized labor cost in the automated scenario at i^* (DES re-run + oversight)
- Q_{scope}^{auto} = Annualized cases for the full 992-SKU candidate scope
- t = Year index in the multi-year projection ($t = 1, 2, \ldots, Y$)

SKU ranking (labor efficiency). All 992 candidate SKUs are ranked by labor-per-case efficiency. Each SKU s gets a score $\rho_s = l_s/max(q_s, 1)$; it is then sorted in descending order by $(\rho_s, l_s, order\ count)$, yielding rank $\pi_s \in \{0,1,\ldots,991\}$. The primary key ρ_s ensures that high-volume, labor-intensive SKUs are prioritized for automation. SKUs with no demand in the simulated period sort to the bottom and are not automated unless capacity remains.

Pareto walk and optimal cut. Walking down the rank list, at each step i, the model adds the next SKU to the automated set and updates cumulative quantities. The single-year analytical breakeven $T_{BE,i}$ is used as the search criteria; the rigorous multi-year breakeven is computed only at the chosen i^* using a fresh DES run.

Cumulative annualized auto cases at step i	$Q_i^{auto} = \alpha \cdot \sum_{s:\pi_s<i} q_s$
Cumulative annualized labor minutes saved at step i	$L_i = \alpha \cdot \sum_{s:\pi_s<i} l_s$
Machines needed at step i	$M_i = \lceil Q_i^{auto}/C_M \rceil$
Cost per machine	Pre-vendor: $c_{mach} = I_{budget}/M_{full}$, where $M_{full} = \lceil Q_{scope}^{auto}/C_M \rceil$. Post-vendor: c_{mach} = user-supplied price; walk caps $Q_i^{auto} \leq M^{cap} \cdot C_M$
CapEx at step i	$I_i = M_i \cdot c_{mach}$
Annual labor savings at step i (analytical)	$\Delta K_i = (L_i/60) \cdot c_h$
Single-year breakeven at step i	$T_{BE,i} = \frac{I_i}{\Delta K_i - m \cdot I_i}$
Optimal step	$i^* = argmin_i\, T_{BE,i}$ (over steps where $\Delta K_i > m \cdot I_i$)

Annual labor cost (workload-based). Once i^* is fixed, the DES is re-run with only the SKUs in the automated subset routed to automation. Annual cost is then computed directly from the simulated labor minutes (not from headcount), multiplied by the hourly wage rate c_h. Oversight full-time equivalents (FTEs) (technicians or operators added to run the automation) are then added.

Baseline (100% manual)	$K^0 = (L_{tot} \cdot \alpha/60) \cdot c_h$
Automated scenario at i^*	$K_{i^*}^{auto} = \left(L_{i^*}^{manual} \cdot \alpha/60\right) \cdot c_h + C_{ov}$, where $L_{i^*}^{manual}$ = total simulated labor minutes for SKUs not in the automated set, and $C_{ov} = F_{ov} \cdot D_y \cdot S^- \cdot H_s \cdot c_h$
Year-1 annual savings	$\Delta K = K^0 - K_{i^*}^{auto}$

Multi-year cash flow and breakeven. Each year of the projection is a scaled version of the base case: demand grows at g, and labor wage grows at g_l, both compounding annually. CapEx is incurred once at Year 0; maintenance is a fixed percentage m of CapEx each year, independent of age. The breakeven year is when cumulative cash flow first turns positive, with linear interpolation used to estimate the exact point within that year.

Year-t baseline labor cost	$K_t^0 = K^0 \cdot (1+g)^{t-1} \cdot (1+g_l)^{t-1}$
Year-t automated labor cost	$K_t^{auto} = K_{i^*}^{auto} \cdot (1+g)^{t-1} \cdot (1+g_l)^{t-1}$
Year-t savings	$\Delta K_t = K_t^0 - K_t^{auto}$
Year-t maintenance	$M_a = m \cdot I_{i^*}$ (constant across years)
Year-t net cash flow	$CF_t = \Delta K_t - M_a$
Cumulative cash flow	$\sum CF_0 = -I_{i^*}$; $\sum CF_t = \sum CF_{t-1} + CF_t$
Breakeven year (linear interpolation)	Let t^* = first year with $\sum CF_{t^*} \geq 0$. $T_{BE} = (t^* - 1) + \frac{-\sum CF_{t^*-1}}{CF_{t^*}}$. If $\sum CF_Y < 0$, breakeven exceeds the horizon.

Capacity sensitivity (full-horizon coverage). Because optimal-breakeven sizing M_{i^*} does not need to cover demand in later years, the model also reports the minimum machine count M^{min} that keeps automation utilization at or below 100% across all Y projection years. Re-running the cash flow with M^{min} in place of M_{i^*} gives a second breakeven figure that the user can weigh against the first:

$$M^{min} = \lceil \frac{Q_{i^*}^{auto} \cdot (1+g)^{Y-1}}{C_M} \rceil$$

When evaluating the formulations as a whole, the user gains a functional understanding of how the simulation works. This is the mathematical basis of understanding when the investment sees a pay back, and when additional machine capacity will be required long-term.

4. RESULTS AND DISCUSSION

The model is a flexible tool that can be used to gain managerial insights about prospective automation solutions as many times as necessary. The user enters inputs and new results are generated each time. To evaluate the robustness of the automation investment and understand how sensitive its financial viability is to key operating assumptions, the team conducted a structured analysis. Five operational and financial parameters were tested: demand-growth rate, labor cost per hour, labor cost growth rate, machine productivity, and total capital expenditure. For each parameter, three trials were run: a conservative Trial A, a reference Trial B, and an upper-bound Trial C. All other inputs were held constant at determined reference values.

4.1 Results

The analysis was conducted in pre-vendor mode, meaning the total automation budget serves as the primary cost input and the model derives an implied cost per machine from the full-scope capacity requirement. Since no real machine price constrains the scope, the model recommends all 992 candidate SKUs in every trial, establishing the full savings potential and capacity requirements to inform subsequent vendor engagement. The reference inputs, from which all five parameter sweeps were derived, are detailed in Table 1.

Input	Reference Value (Trial B)
Years	10
Flows	Conventional, AS/RS and HTLA, Cross-Docking
Labor cost per hour	16 €
Labor cost growth	4%
Demand growth	12%
Total investment (budget)	2,000,000 €
Machine productivity	1,500 cph
Annual maintenance	2% of CapEx
Automation support FTE	1

Table 1: Reference Input Parameters

Consistent with the pre-vendor mode design, all 992 candidate SKUs were selected. In this case, three machines operating at 1,500 cph each, were required to serve the SKU set. Each machine had an implied capital expenditure of 666,667 €. This was derived when the 2 million € budget was allocated across the three machines' productivities needed to cover the full 992-SKU scope. It is important to note that this logic changes in the post-vendor model. With the amount and cost per machine set as an input, the model will perform the aforementioned SKU optimization algorithm and recommend an optimal subset of the 992 SKUs to automate.

The automation solution reduces the warehouse workforce requirement from 25.5 FTEs to 2.0 FTEs, generating Year 1 labor savings of 1,829,939 €. A maintenance cost of 40,000 € was factored into annual costs. The resulting breakeven period is 1.10 years, as illustrated in Figure 1. Figure 2 presents the Year 1 labor cost waterfall, detailing the savings generated by routing the optimal SKU subset to automation. Figure 3 shows the 15 SKUs with the highest potential for labor cost savings. Since the AS/RS flow has the most demanding case volume per SKU, transferring these to an automated solution and decreasing the cost of the pick unlocks the most cost savings.

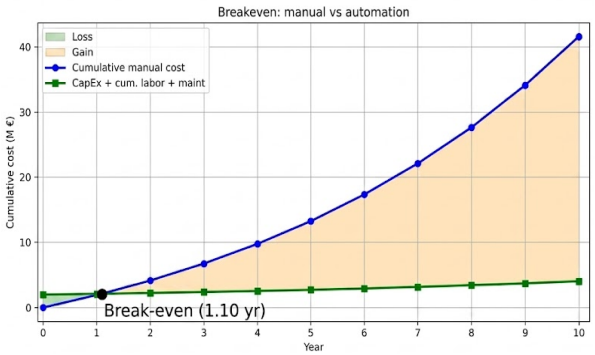


Figure 1: Reference Breakeven Chart

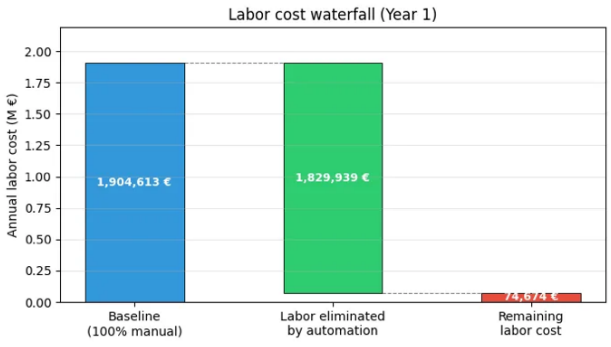


Figure 2: Reference Cost Savings Waterfall

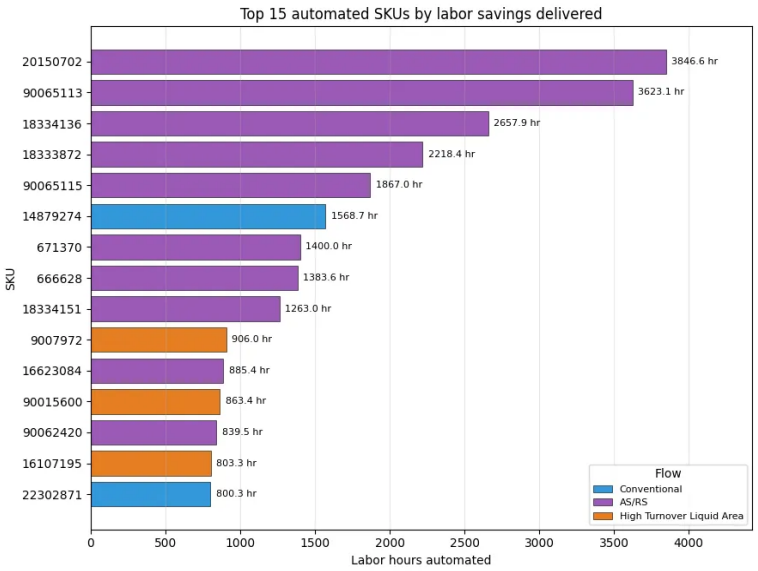


Figure 3: Most Labor-Intensive SKUs

As seen in Table 2, although Conventional accounts for 49.1% of the 992 candidate SKUs, it only generates 23.7% of total labor minutes across the simulated period. On the other hand, AS/RS represents 9.5% of the candidate scope but accounts for 35.9% of total labor, averaging 23,851 labor minutes per SKU compared to Conventional's 3,049 labor minutes per SKU. This large discrepancy, despite the AS/RS having a faster pick rate, is due to pure volume. Notably, volume dominates the labor intensity calculation, which is why 10 of the 15 most labor-intensive SKUs are AS/RS items. The optimizer correctly identified these SKUs as the highest priority to automate since such SKUs would produce the greatest savings proportional to the scope.

Flow	SKUs	% of scope	% of total labor minutes	Avg labor minutes per SKU
Conventional	487	49.1%	23.7%	3,049
HTLA	280	28.2%	39.8%	8,894
Cross-docking	131	13.2%	0.6%	269
AS/RS	94	9.5%	35.9%	23,851

Table 2: Flow Distribution of the 992 Candidate SKUs

One important capacity planning consideration is that the optimal three-machine solution covers projected automated volume for nearly two projection years under 12% annual demand growth, as shown in Figure 4. From Year 3 onwards, the projected volume of automated SKUs exceeds the machine's annual capacity of 18,901,929 cases. This does not compromise the financial case, as labor savings are still captured. It signals that an investment in additional machine capacity should be evaluated before the start of the third operating year. The machine productivity analysis, discussed below, provides the relevant framework for that decision.

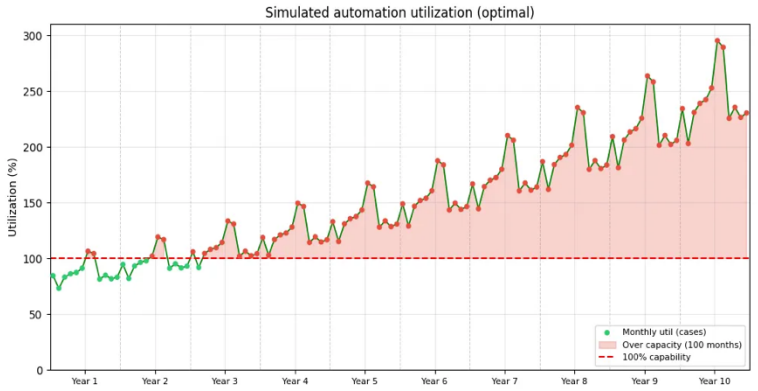


Figure 4: Demand Coverage Chart

The analysis results across all five parameters and three trials are summarized in Table 3. For each trial, the reported outputs are the Year 1 labor savings, the breakeven period, and the 10-year cumulative cash flow, which detail both short-term viability and long-run financial return of the automation investment.

Metric		Trial A	Trial B (Reference)	Trial C	Breakeven Range	10-Yr CF Range
Demand Growth	Adjusted Input	8%	12%	15%	0.00 yr	13,959,787 €
	Cum. Cash Flow	30,214,148 €	37,545,675 €	44,173,935 €		
	Payback Period	1.10 yr	1.10 yr	1.10 yr		
Labor Cost / hr	Adjusted Input	14 €	16 €	18 €	0.25 yr	9,986,418 €
	Cum. Cash Flow	32,552,466 €	37,545,675 €	42,538,884 €		
	Payback Period	1.24 yr	1.10 yr	0.99 yr		
Labor Cost Growth	Adjusted Input	2%	4%	6%	0.00 yr	8,793,788 €
	Cum. Cash Flow	33,402,181 €	37,545,675 €	42,195,969 €		
	Payback Period	1.10 yr	1.10 yr	1.10 yr		
Machine Productivity	Adjusted Input	1,000 cph	1,500 cph	2,500 cph	0.00 yr	0 €
	Cum. Cash Flow	37,545,675 €	37,545,675 €	37,545,675 €		
	Payback Period	1.10 yr	1.10 yr	1.10 yr		
CapEx	Adjusted Input	1.5M €	2.0M €	3.0M €	0.76 yr	1,800,000 €
	Cum. Cash Flow	38,145,675 €	37,545,675 €	36,345,675 €		
	Payback Period	0.83 yr	1.10 yr	1.59 yr		

Table 3: All Trials Analysis Results

The results of these trials are as follows:

Demand Growth. Demand growth generates the largest 10-year cash flow swing of any variable tested in the pre-vendor mode. Across the three demand growth trials (8%, 12%, and 15% per year) the breakeven period is invariant at 1.10 years. Demand growth significantly affects the 10-year cumulative cash flow, which spans from 30,214,148 € at 8% growth to 44,173,935 € at 15% growth. At the company's stated expectation of 10–15% annual growth, Vegalsa-Eroski is operating near the upper bound of the tested range. This corresponds to the most financially favorable demand scenario tested. It is also worth noting that higher demand growth accelerates the year in which projected automated volume exceeds fleet capacity, reinforcing the need for a phased procurement plan.

Labor Cost per Hour. Labor cost per hour is the only variable in the analysis that directly shifts the breakeven period in both directions. At 14 € per hour (Trial A), the breakeven extends to 1.24 years and 10-year cumulative cash flow falls to 32,552,466 €. At 18 € per hour (Trial C), the breakeven compresses to 0.99 years and cumulative cash flow rises to 42,538,884 €. The relationship is essentially linear: each 2 € per-hour increase in the labor rate shortens breakeven by approximately 0.125 years and adds approximately 5 million € in 10-year savings. Given an ongoing labor shortage in the warehousing sector and potential increase in warehousing costs in the EU, the trajectory of this variable is more likely to move toward Trial C than toward Trial A.

Labor Cost Growth. The 10-year cumulative cash flow ranges from 33,402,181 € at 2% annual wage growth to 42,195,969 € at 6%, a spread of 8,793,788 €. This variable functions as a long-run amplifier: slower wage escalation moderates the savings advantage of automation over time, while faster wage growth increases that advantage year over year. By Year 10, the gap in annual cash flow between the 2% and 6% wage growth scenarios exceeds 2,500,000 €.

Machine Productivity. Unlike the other variables tested, machine productivity does not affect labor savings, breakeven period, or 10-year cumulative cash flow in pre-vendor mode. Because the total investment is fixed at 2,000,000 €, the model adjusts machine count and implied unit cost in response to throughput while holding the automated SKU scope at 992 and annual labor savings at 1,829,939 € across all three trials. At 1,000 cases per hour (Trial A), five machines are required at 400,000 € each, producing a fleet utilization rate of 98.3% and a capacity breach in Year 4. At 2,500 cases per hour (Trial C), two machines suffice at 1,000,000 € each, with the same total fleet capacity and breach year, but a lower utilization rate of 78.6%, providing a greater operational capacity buffer. The reference case breaches one year earlier at Year 3 since its three-machine fleet yields lower total capacity than either alternative.

The pre-vendor mode holds total investment constant at 2,000,000 €, meaning the implied cost per machine adjusts inversely with machine count across the productivity trials. This causes the financial outputs to appear unaffected by productivity. To fully understand the financial impact of throughput, the analysis was furthered by holding the price per machine constant at the reference implied cost of 666,667 €. Under this assumption, a 1,000 cph system would require five machines at a total investment of 3,333,335 €, producing a breakeven of 1.89 years and a 10-year cumulative cash flow of approximately 35,959,918 €. A 2,500 cph system would require two machines at 1,333,334 €, shortening the breakeven to 0.74 years and improving the cumulative cash flow to approximately 38,359,919 €. These results showcase a breakeven range of 1.15 years and a cash flow range of approximately 2,400,000 €. Such results can be evaluated more directly using the tool's post-vendor mode with a fixed price per machine.

Capital Expenditure. The greatest payback period range in this analysis was in the CapEx trials. Doubling the total budget from 1.5 million € to 3.0 million € stretches the breakeven period from 0.83 years to 1.59 years, a 91.6% increase in payback time, while changing the 10-year cumulative cash flow by only 1,800,000 €, or roughly 4.7%. This shows that CapEx affects only

the payback side of the investment, not the savings side. The optimizer selects all SKUs regardless of budget in the pre-vendor mode, while only the implied machine price changes. At a 1.5 million € budget, the model derives an implied machine cost of 500,000 €, producing a fast payback against the same labor savings. At 3.0 million €, the implied cost rises to 1,000,000 €, stretching payback by approximately nine months with virtually no corresponding benefit to long-run returns. Figure 5 presents the breakeven curves for all three CapEx trials, illustrating how sharply the payback timeline diverges while cumulative savings remain nearly constant.

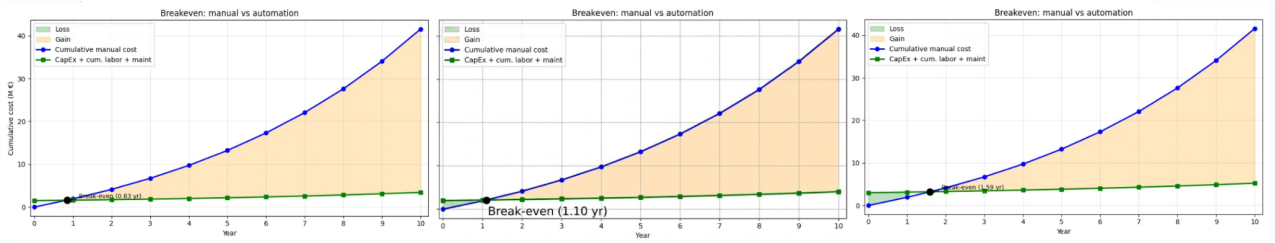


Figure 5: CapEx Breakeven Chart Comparison (Trials A, B, C)

4.2 Implications

The analysis results offer two overarching conclusions for Vegalsa-Eroski. First, the automation investment is financially viable across every scenario tested, including the most conservative combination of inputs. The business case does not depend on favorable conditions; it holds under stress. Second, the variables that most influence the long-run value of the investment are largely outside the company's direct control, while the variable with the sharpest near-term effect on payback and capital expenditure is directly negotiable. This distinction has important consequences for how the company should prioritize its efforts ahead of a vendor engagement.

Demand growth and labor cost per hour were the dominant drivers of 10-year cumulative return, with combined value swings of 23,946,205 € across their tested ranges. Neither is completely controllable by Vegalsa-Eroski: demand is a function of market conditions and the company's commercial performance, while labor cost per hour is set by the market. Labor cost growth is also beyond the company's direct control but moves predictably in a direction that strengthens the investment case, due to the structural nature of the warehouse labor shortage. Capital expenditure, by contrast, is the only major lever that management can actively influence. This analysis shows that pulling it in the wrong direction, by budgeting upward rather than negotiating the machine price down, creates a meaningful payback penalty with almost no corresponding benefit to long-run returns.

From a capacity planning perspective, the reference results confirmed that the optimal three-machine solution will be over capacity by Year 3 under the reference assumption of 12% demand growth. At 15% growth, the demand-capacity gap widened even more significantly in subsequent years. This did not impair the financial returns, but it means that Vegalsa-Eroski should treat the initial investment as the first stage of a phased procurement rather than a standalone solution. The machine sensitivity analysis shows how automation scope and breakeven evolve across tiers: one machine covers 31 SKUs (3.1% of scope) at a breakeven of 1.23 years, two machines expand coverage to 169 SKUs (17.0%) at 1.19 years, and three machines, the recommended configuration, automates the full 992-SKU scope at 1.23 years. Beyond three machines, additional units do not expand the automated SKU scope but do extend demand coverage. A fourth machine would raise total annual fleet capacity to approximately 25,202,572 cases, pushing the capacity breach from Year 3 to Year 5 under 12% demand growth, and from Year 2 to Year 5 under the 15% growth scenario.

These pre-vendor results showcase the economic ceiling and capacity requirements Vegalsa-Eroski should understand prior to communicating with vendors. Once machine count and unit price are known, the model can be run in post-vendor mode to assess whether each machine in the proposal is justified by the labor savings it enables. This mode will then determine the optimal SKU automation scope for the input capital constraints. Used sequentially, the two modes support a complete path from planning to procurement.

4.3 Limitations

This tool provides a comprehensive overview that aids in planning the purchase and implementation of an automation solution. It includes some simplifications to make the tool more user-friendly. One of these is that there is no discount rate. If a supplier prices their automation machines with incremental discounts, the discounts would not be reflected in the model when recommending how many machines to purchase. These discounts could be substantial and should be considered in the decision-making process. The overall budget may still be inputted accurately, even if the cost per machine is uniform.

Another limitation is that demand and labor growth are applied as a uniform percentage increase each year to all SKUs and all flows. In other words, any difference in growth rates due to seasonal demand or product life cycle is not factored in under this flat rate. While these nuances are not captured, the percentage increase serves as a sufficient generalization to estimate these variables. The cost of workers needed to handle any overflow volume is also not factored into the cost analysis. Any additional manual labor would increase the effective cost of automation in later years of the projection. Labor costs are also assumed to be a flat rate across all workers and shifts, with no shift premiums or part-time workers, which generally would result in a higher modeled labor cost if accounted for.

Similarly, maintenance costs are applied as a fixed percentage of the original capital expenditure each year. In practice, maintenance costs are typically lower in the first year and increase with age, as wear and tear accumulates. This model overestimates upfront maintenance costs but underestimates those in later years, which makes the long-term breakeven appear marginally more favorable than it may be in practice.

In pre-vendor mode, throughput does not affect labor savings or automated SKU scope. This is because the budget is fixed and the model scales machine count accordingly. In this mode, the operationally significant effects of throughput are machine utilization rate and per-machine implied cost. Additionally, since the pre-vendor machine productivity analysis holds total budget constant, the implied cost per machine varies inversely with throughput. This structure means the financial outputs, breakeven and cumulative cash flow, are invariant to productivity in pre-vendor mode. Users seeking to isolate the impact of throughput on financial returns should conduct this analysis in post-vendor mode with a fixed price per machine.

One final limitation is that the automatable SKU list is fixed at the original 992 candidates identified by Vegalsa-Eroski prior to the analysis. While the model determines which subset of these SKUs is most cost-effective to automate in the post-vendor mode, it cannot evaluate the thousands of remaining SKUs not included in this list. Similarly, once the model finds the optimal SKU subset, demand growth will only include that subset. Expanding the candidate scope in a future iteration of the tool could identify additional labor savings opportunities not captured in the current analysis.

5. RECOMMENDATIONS

The purpose of this project was to give Vegalsa-Eroski the capability to evaluate a variety of automation solutions. To enable this, the team built a flexible tool using a DES with a black box

area where the user could input specifications. The strength of this tool is that it has the capability to run as many scenarios as needed to gain proper insight into investing in automation. Insights gained from the results are generalized to give management the clarity on how to use this tool. These recommendations are made within the bounds of the stated limitations.

5.1 Management Recommendations

The team recommends leadership use the simulation not just as an internal validation tool, but as a direct asset during vendor negotiations. To extract the maximum value from the capstone solution, the following sequence is suggested during procurement. Prior to issuing a request, management must run the tool in the pre-vendor mode using their internally approved capital budget. This establishes an implied maximum machine cost ceiling. Once vendor quotes are received, these specific pricing structures should be inputted into the tool using the post-vendor mode. This comparative workflow clearly identifies whether a vendor's proposal is priced to deliver an acceptable return. By entering negotiations with a definitive, mathematically backed cost ceiling, Vegalsa-Eroski can ensure a vendor’s pricing is justified against their proven threshold for financial viability.

While the simulation outputs model a clear financial return, they also highlight the gravity of projected capacity breaches. The model identifies that under a 12% annual growth scenario, the current facility will breach its maximum throughput capacity by Year 3. However, Year 3 should not be viewed as the decision deadline. Scaling up the chosen automated system may require significant lead time. By working backward from the projected capacity breach, leadership should target finalizing the procurement decision with plenty of time to avoid stockouts, facility gridlock, and subsequent losses in market share. This phased approach also aligns with the company’s goal of scaling with demand instead of over-investing ahead of it. Rather than committing to a full system up front, the company can validate real-world performance against simulation projections in Year 1. They can subsequently update and re-run the model before placing the next phase’s order.

The reference simulation demonstrates a dramatic reduction in required labor for the automated SKUs, transitioning from an estimated 25.5 FTEs to just 2.0 FTEs. While this is reported as a highly successful financial and operational outcome, it represents a material workforce management challenge. A workforce transition strategy, drafted ahead of deployment, is recommended. Instead of relying on layoffs, the sponsor should evaluate natural attrition rates. Furthermore, human resources should develop frameworks for redeploying displaced workers to other value-added areas of the distribution center. Additionally, a select cohort of the existing workforce should be upskilled for the technical oversight, system management, and maintenance roles required by the automation.

5.2 Future Work

There are several areas for future expansion of the tool. For example, the simulation model could be updated to more realistically model the processes in the warehouse by stochastically determining pick times and pallet availability. The scope could also be widened to include an expanded list of SKUs to be examined for automation. This may also consist of other areas within the distribution center, such as Vegalsa-Eroski’s Microshuttle process and non-ambient areas. Another route to investigate would be to incorporate a discount rate into the financial evaluation, which would provide a more conservative estimate of the net present value of the investment and allow for more nuanced comparisons between phased procurement options.

An additional potential enhancement is a forward-looking optimizer that solves machine count and SKU scope in conjunction across the entire planning horizon. This extension of the model

could assist in more complex automation projects down the line since the current model focuses specifically on minimizing the payback period.

While this simulation model was developed specifically for the ambient area in Vegalsa-Eroski's main distribution center, the lessons learned here apply to warehouse automation planning more broadly. The results of this project suggest that forecasting how well an automation solution will perform in a specific operational context yields more informed investment decisions. Rather than assuming that automation will increase productivity and generate cost savings, companies should first assess their specific demand profiles, labor cost structures, and growth trajectories to ensure they are making a sound and well-calibrated investment.

REFERENCES

Antoniades, I. P., & Stamelos, I. (2022). Discrete event simulation of an automated warehouse inspection system with drones. *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 1–8. https://doi.org/10.1109/IISA56318.2022.9904417

Ceroni, J.A. (2009). Economic Rationalization of Automation Projects. In: Nof, S. (eds) Springer Handbook of Automation. Springer Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78831-7_40

Custodio, L., & Machado, R. (2019). Flexible automated warehouse: a literature review and an innovative framework. *International Journal of Advanced Manufacturing Technology*, *106*, 533–558. https://doi.org/10.1007/s00170-019-04588-z

de Koster, R. (2018). Automated and robotic warehouses: Developments and research opportunities. *Logistics and Transport*, *38*(33), 4–18.

Peixoto, R., Dias, L., Carvalho, M. S., Pereira, G., & Geraldes, C. A. S. (2016). An automated warehouse design validation using discrete simulation. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 199–204. https://doi.org/10.1109/ITSC.2016.7795554

Saderova, J., Rosova, A., Behunova, A., Kubekova, M., Hlatka, M., & Melisova, A. (2022). Case study: the simulation modelling of selected activity in a warehouse operation. *Wireless Networks*, *28*, 431–440. https://doi.org/10.1007/s11276-021-02574-6