End-of-Life Inventory Optimization during Runout Events for a Manufacturing Automotive Company

by

Francisco Andres Calero Mantilla
BS, Industrial Engineering, Universidad San Francisco de Quito

and

Andrea Esposito
MS, Mechanical Engineering, Polytechnic University of Turin

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Signature of Author: ____
Department of Supply Chain Management
May 12, 2023

Signature of Author: ____
Department of Supply Chain Management
May 12, 2023

Certified by: ____
Dr. Matthias Winkenbach
Principal Research Scientist, Center for Transportation and Logistics
Capstone Advisor

Certified by: ____
Dr. Juan Carlos Piña Pardo
Postdoctoral Associate, Center for Transportation and Logistics
Capstone Co-Advisor

Accepted by: ____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

End-of-Life Inventory Optimization during Runout Events

for a Manufacturing Automotive Company

by

Francisco Andres Calero Mantilla

and

Andrea Esposito

Submitted to the Program in Supply Chain Management
on May 12, 2023, in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Effective and efficient inventory management is today more crucial than ever. Three factors drive its importance: (i) the ever-increasing complexity of bills of materials, causing firms to dedicate a large part of their spending to their supply base; (ii) unprecedented uncertainty, making inventory indispensable for business survival; and (iii) an emerging inflationary economy, which multiplies the cost of holding inventory. One of the most critical inventory management processes is the runout phase, which refers to stock depletion as products approach their end of life. Operational excellence during the runout phase is particularly important not only due to the negative externalities of holding excessive stock but also to the costly risk of scrapping the remaining inventory if materials become obsolete. In this capstone project, we propose an end-to-end approach to improve runout inventory management. This approach aims to guide companies in defining the problem of runout inventory management, preparing the raw data, choosing the right models, setting up the analysis, and interpreting its results. We illustrate each step of our approach by using a real-world case study from a large automotive manufacturer. Using historical data from one of their largest factories in the United States, we trained and tested several predictive models to estimate the demand during the last month of the lifecycle of materials three months in advance, when the last replenishment orders are issued. Based on our analysis, we provide several recommendations to successfully improve the company's demand forecasting and runout inventory management processes.

Capstone Advisor: Dr. Matthias Winkenbach
Title: Principal Research Scientist

Capstone Co-Advisor: Dr. Juan Carlos Piña Pardo
Title: Postdoctoral Associate

ACKNOWLEDGMENTS

Our deepest gratitude goes to those who allowed us to be here and those who worked and struggled with us throughout the year, in the effort of making this work possible. We want to thank our families for providing us with the affective, cultural, and ethical means that allowed us to be here, the MIT SCM administration, for believing in us and admitting us into the program, our capstone advisor, Dr. Matthias Winkenbach, for the quickness and depth of his understanding as well as the guidance in the most difficult phases of the project, our co-advisor, Dr. Juan Carlos Piña Pardo, for his responsiveness and precise insights, and our writing advisor, Ms. Pamela Siska, for her appreciation and encouragement.

Collaborating with the partner company was an enriching and fulfilling experience. We are grateful for the opportunity to work on the application of advanced techniques in the murky field of the real world. Along the way, we were fortunate to work with numerous professionals in the field of data science, materials planning and operations management, who learned and struggled with us in a fruitful exchange of knowledge. The company is filled with talented individuals, and we have no doubt that they will excel in building on our work and achieving the ambitious objectives that the company has set for itself.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

1.1. Motivation

Because product availability and customer demand are always uncertain, companies protect themselves from stockouts by including a safety stock in their inventory. This helps them to better serve their customers by ensuring production continuity. The main objective of the safety stock is to provide a buffer in case a sudden spike in demand or a sudden drop in supply disrupts production. If the safety stock is excessive, however, it can significantly increase inventory holding costs. This problem is particularly critical if a material is close to its end of life, i.e., about to become obsolete. In this case, excessive inventory is likely wasted if the company does not utilize its inventory before the end of life is reached. As a consequence, companies constantly try to optimize their safety stock in an effort to minimize cost while maintaining the desired service level.

Our partner company is revising its strategy to optimize safety stock, specifically focusing on the runout phase of the product cycle. In the context of inventory management, the term "runout" refers to the process of depleting the stock of a material due to the product's end of life or design change. The "runout date" denotes the last day on which the material can still be utilized. Beyond this date, any remaining inventory is considered obsolete and must be discarded. The company primarily schedules end-of-life processes during specific periods of the year, known as "runout events." This approach enables the company to systematically clear out obsolete materials, streamlining the tasks of planners and enhancing coordination with suppliers.

It is critical for our partner company to possess a reliable method for forecasting inventory consumption as materials approach their runout date, as well as a strategy that minimizes leftover inventory while avoiding stockouts. The main concern is that implementing a suboptimal policy may result in a significant waste of resources in the form of obsolete inventory after the runout date (i.e., overage cost) or excessive cost of expedited replenishment shipments in case of material shortages (i.e., underage cost). Therefore, the company is interested in exploring advanced methodologies to better manage this stage of the process.

This project focuses on one of the company's largest factories in the US. At this location, three runout events take place every year. Each of these events impacts thousands of materials, meaning that, at every runout, several materials are updated, and all the remaining inventory of the obsolete version gets scrapped. The complexity of the operations, due to the considerable number and variety of products manufactured at the company's factory, makes testing innovative methodologies to optimize inventory appealing.

Current inventory management practices at the company's factory present some opportunities for improvement. Material planners are responsible for determining the appropriate levels of inventory in the runout phase. Each planner makes the decision for all the SKUs that he manages, with the goal of ensuring production continuity. The decision is based on the planner's experience, without any system to aid him in dealing with the complexity of the task.

In the last few years, however, the company set itself in the position to step away from a purely manual process, by collecting a large amount of data throughout their global operations about plants, inventories, times, demands, backlogs, missing parts, and suppliers. For this project, the company provided such data for a time span of two years.

1.2. Problem Statement & Research Questions

Historically, the factory incurred considerable overage cost due to remaining inventory after each runout. Therefore, the company's key objective is to define a methodology to optimize their safety stock and reduce the amount of obsolete inventory after the runout dates. Specifically, they want to determine the optimal inventory required for each material three months before the runout date by forecasting demand during the material's final month of life, when safety stock consumption occurs. With a reliable forecast in place, the company anticipates improved runout planning, enabling stakeholders and material planners to make more informed business decisions based on these predictions.

In this context, the questions to be answered include:

1. What methodology should the company follow to determine three months in advance the optimal inventory needed during the runout phase?
2. What are the key components that will affect the optimal inventory and how they should be analyzed by the stakeholders or decision-makers?

1.3. Scope: Project Goals & Expected Outcomes

The project's overall goal is to provide the company with a methodology to determine an optimal safety stock for materials before a runout takes place, by forecasting the inventory consumption three months in advance and with a buffer of protection against stockouts. This method should also help the company policymakers create scenarios to support decision making about the right amount of inventory to hold.

We hypothesized that a data-driven and traditional approach to determining safety stock would be the most effective and straightforward method to help the company avoid stockouts and minimize excessive inventory after each runout. This methodological approach prioritizes interpretability, as its primary goal is to support the decision-making process rather than solely optimizing inventory. Thus, we reviewed the literature on inventory management methods for safety stock decisions, time series analysis for safety stock predictions, and the application of machine learning models for inventory optimization.

We believe that this project's success relies on using our proposed approach to connect the state-of-the-art knowledge from the literature with the daily practice of material planners, who will make the final inventory decisions. To address the needs of these decision-makers, their input was essential during the development of the methodology.

The deliverable for the company is a demonstration of our approach to framing the analysis. The immediate usability of this methodology depends on the quality of the available data. For the company's factory, the data was found to have several deficiencies; however, the method can still be applied once the data issues are resolved in the future. Upon implementing the methodology, the company is expected to reduce the overall costs associated with the runout phase, by avoiding stockouts and reducing obsolete inventory while maintaining customer service levels.

2. State of the Art

2.1. Inventory Management

2.1.1. Introduction

<u>The role of inventory and associated risks</u>

Manufacturing companies hold inventory for several reasons. First, it provides a buffer against fluctuations in demand. If the material's demand increases, firms can meet this demand by drawing on inventory rather than increasing production beyond capacity. This allows firms to respond more flexibly. Second, inventory decouples supply and demand. If production upstream is interrupted for any reason, inventory can be used to meet customer demand in the meantime. This ensures that consumers are not left without the products they need.

Stocking inventory costs money, in the form of invested capital and storage space. In the simplest case, finding the optimal order quantity is equivalent to finding the policy that balances reordering costs and holding costs. Inventory also comes with some risks: most importantly, the risk of obsolescence. If a product is perishable, becomes outdated, or goes out of fashion, the firm may be left holding large amounts of inventory that it cannot sell, leading to significant financial losses (Vrat, 2014).

Inventory always represents physical objects since services cannot be stored, so the most obvious metric to measure it is units of products. Firms, however, often need to treat the inventory as an aggregate (for financial purposes) or to make comparisons between different materials (for operational purposes). Therefore, it is common to treat inventory as a financial entity, using the unit cost as a conversion mean, or as a "time of supply", using the consumption rate as a conversion mean. The first represents the total capital invested in the goods, while the second represents the time that the inventory would last if supply was interrupted today, at the current rate of consumption (Muller, 2011).

Cycle stock, pipeline inventory and safety stock

For the purpose of this capstone, it is worth making a distinction between the different types of inventories. A widely used distinction is between raw parts, work in progress, and finished goods. While being useful from a financial standpoint, this characterization does not provide any valuable insight from an operational perspective. A more convenient distinction, widely used in inventory management, is between cycle stock, pipeline inventory, and safety stock (Muller, 2011).

The cycle stock is the inventory that a firm needs to operate between two replenishments. The pipeline inventory, owned by the firm or by its supplier based on the contract terms, represents the goods in transit during the time between the order and the arrival at the firm's facilities. The safety stock, which is of primary importance for this capstone, represents the amount of inventory that the firm holds to make up for uncertainty in demand and lead time (Silver & Pyke, 2017).

The safety stock works as a decoupling point between the supplier and the operations so that if there is a spike in demand or a delay in supply the firm does not run out of inventory. The event of running out of inventory is known as stock-out and is considered a major disruption, especially in high-intensity manufacturing environments like the automotive industry. Such industries operate at a high utilization rate to minimize capital investment in equipment, and as a consequence, they do not have spare capacity. The idle time that follows a stockout event results in lost revenues and substantial financial losses (Silver & Pyke, 2017).

Common methods and schools of thought

A review of classic models for inventory management can be found in Nahmias (1982). The author reviews the models' policy and the risk of perishability based on the features of product lifetime (fixed vs random) and demand (deterministic vs stochastic) and suggests optimal and approximately optimal ordering policies according to the model. More recently, Vrat

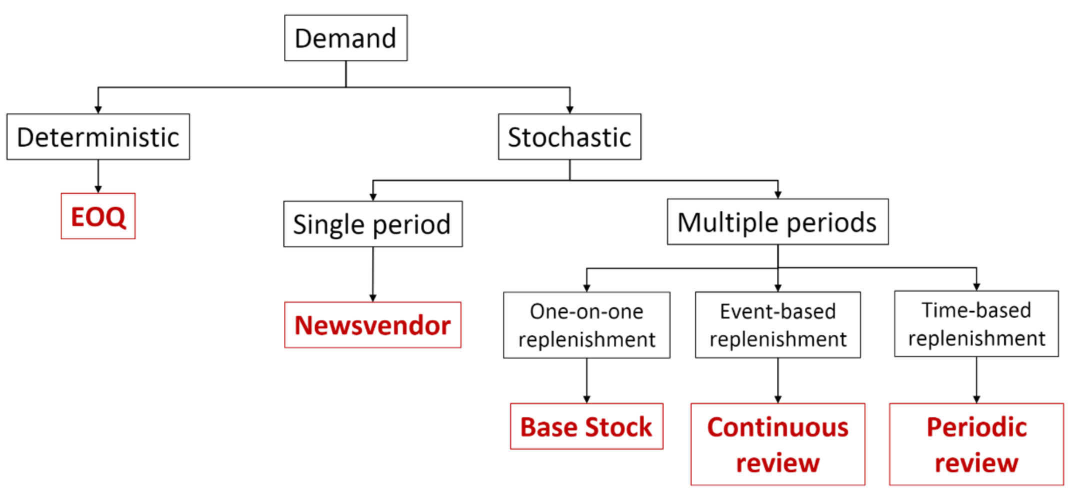
(2014) provided an extensive overview of inventory models, with a hierarchical classification based on dynamics of the decision making, stochasticity of the model, number of echelons considered and other special features of the model. Ziukov (2015) reviewed the models in the literature that deal with uncertainty, focusing his attention on whether the input parameters of each model were treated as crisp (binary or integer values) or fuzzy (real values associated with a certain probability distribution).

Examining the nuances of all the models in the literature is beyond the scope of this capstone, but it is worth reviewing the main families of inventory policies (EOQ, multiple periods models and single period models), and the factors that condition the choice towards one or another (see Figure 1). The most important distinctions are:

1. Whether demand and lead time are deterministic or stochastic.
2. Whether unused inventory can be utilized in a subsequent period or not.

Figure 1

Families of inventory policies (Silver & Pyke, 2017)



Single-period models are qualitatively different from all others. In a single-period model, all the sales are concentrated in a single event. Whatever is left unsold after the sales period loses value substantially (often completely), and there is no possibility of backordering or replenishing the inventory. The resulting optimization model does not try to minimize the total cost, but rather maximize expected profit, trading off the consequences of holding too much or too little inventory. A thorough explanation of single-period models can be found in the following section dedicated to the newsvendor model (Silver & Pyke, 2017).

2.1.2. The Newsvendor Problem

Conceptual overview

The newsvendor or newsboy problem takes its name from a simple but explanatory case of the model's application. The newsboy is the decision maker, and every morning he must decide how many copies of the newspaper to buy to maximize his profit. Newspapers are by nature a perishable product, and the copies left unsold at the end of the day are worth nothing. Therefore, the newsboy faces a dilemma: given the uncertainty of the demand for newspapers, how many copies should he buy? The solution to the problem boils down to the calculation of the so-called critical ratio. This ratio expresses how convenient it is for the decision maker to stock more than what he expects to sell for the sake of making a profit, and reflects the tradeoff between the cost of shortage and the cost of unsold inventory (Qin et al., 2011).

A more formal definition of the total profit maximization problem is:

$$maximize\ P(Q) = p\ (\min[x, Q]) - c\ Q$$

p: price per unit

x: materialized demand

c: cost per unit

P: total profit

Q: ordered units of product

The problem can be easily solved analytically. Given c_e the cost of having one unsold unit and c_s the cost of being short of one unit, the expected costs of shortage and excess are:

$$E[Excess] = c_e \, Prob[x \leq Q] \qquad E[Shortage] = c_s \, (1 - Prob[x \leq Q]),$$

where $Prob[x \leq Q]$ denotes the probability of ordering more than needed.

The optimal order quantity Q^* is such that $E[Excess] = E[Shortage]$. This condition realizes if:

$$Prob[x \leq Q^*] = \frac{c_s}{c_e + c_s}$$

This quantity is the so-called critical ratio, CR. We can break down c_e and c_s into their components:

$$c_e = c - g \qquad c_s = p - c + B$$

g: salvage value after the period has ended

B: penalty per unit short beyond lost profit

Given an order quantity Q, then, the corresponding expected profit $E[P(Q)]$ is:

$$E[P(Q)] = (p - g)\, E[x] - (c - g)\, Q - (p - g + B)\, E[US]$$

The uncertainty of the demand in the problem is accounted for by $E[US]$ expected units short. Gallego and Moon (1993) review the literature about the possible distributions for the demand in the problem. It is common in the industry, however, to consider the demand normally distributed, in which case $E[US]$ is uniquely function of the standard deviation of demand and the Critical Ratio (Silver & Pyke, 2017).

<u>Applicability to this capstone project</u>

The newsvendor model has many applications in practice. It is widely used in industries where the lifetime of the product is naturally short compared to the replenishment lead time, such as groceries, fashion, and event-specific retail. These applications have two features in common: they are subject to substantial uncertainty in demand, and once demand materializes there is no possibility to adjust the quantity that has been ordered or produced (Choi, 2012).

In high-intensity mass manufacturing industries like our partner company, inventory is typically durable, and the inventory management policy is well represented by a multi-period model. The objective of these companies, when it comes to inventory, is to run the operations efficiently while minimizing the holding cost. However, the problem submitted for this project is different. During runout events, the materials of the expiring configurations behave like perishable products with a common expiration date. After the runout, the materials have, at most, a salvage value equal to the product value subtracted from the total costs of flashing them, or they are worth nothing if they cannot be flashed. Flashing is a procedure by which an electronic component's firmware and data are cleared and overwritten, in this specific case, to update it to the next generation.

The other feature that makes this problem similar to a newsvendor problem is the uncertainty about the demand. The replenishment lead time for materials is, in many cases, eight to ten weeks for the plant, and the company chooses the quantities for the last replenishment three months in advance. Customers, on the other hand, may change the configuration of the optional features of the finished product down to 2 weeks before it gets delivered to the client.

It is clear, then, that the problem presents all the features necessary to qualify as a newsvendor problem, and the tradeoff is between the cost of stocking out and the cost of scrapping the inventory left after the runout.

2.2. Advanced/Novel Approaches to the Newsvendor Problem

In the traditional Newsvendor Problem (NVP), demand is considered stochastic, and its distribution and parameters are known. However, this assumption does not always hold. When applying these methods, the real challenge is overcoming changing demand, disruption, and incomplete information, among others (Rolf et al., 2022). Therefore, several methods are available for dealing with unknown demand distribution for inventory decision-making: parametric and non-parametric (de Castro Moraes & Yuan, 2021).

The parametric methods rely on the assumption that the demand distribution corresponds to a parametric family of distribution, and its parameters are unknown. This approach is presented in Section 2.1.2. An extension of the common approach is found in Liyanage & Shanthikumar (2005), where the authors integrate the parameter estimation step with the optimization task arguing that not doing both of them together leads to sub-optimal results. However, the assumption about the demand distribution is problematic, as it is complex to confirm beforehand. Hence, additional methods, non-parametric methods, have been developed to tackle this problem.

The non-parametric methods do not assume any distribution for the demand, but they rely on empirical information to solve the task at hand (de Castro Moraes & Yuan, 2021). The following approaches are part of non-parametric methods: robust optimization, sample average approximation, and machine learning techniques.

For robust optimization, Scarf (1959) presented an NVP for a single-product, single-period problem setting, where only the product demand’s mean and variance are known. Then, Bertsimas & Thiele (2006) extended the research on robust optimization by using demand observations instead of demands' mean and variance. The authors argued that their approach outperformed other models that used parametric methods. Additionally, Bertsimas and Thiele (2006) reformulated the NVP and several extensions of it as linear programming models.

Moreover, for sample average approximation, Levi et al. (2007) proposed using a sample of the demand to solve the NVP. Specifically, the authors provide an upper bound on the number of samples required to achieve almost optimal solutions without depending on the demand distribution.

Finally, several approaches have leveraged the power of machine learning by joining the demand estimation and optimization steps into a single procedure. Ban and Rudin (2019) proposed exactly this method by developing algorithms to solve multi-feature NVPs based on empirical risk minimization and kernel-weights optimization. The authors also argued that including additional features (besides demand and inventory costs) helped achieve more robust and better decisions. In addition, Oroojlooyadid et al. (2020) proposed an algorithm based on deep learning that optimizes the order quantities for several products using features of the demand data. Their algorithm also integrates forecasting and optimization into a single step and is applied for solving a multi-feature NVP. The authors argue that their model outperforms other approaches, especially for demands with high volatility. Punia et al. (2020) extended multi-term NVP by including a capacity constraint and using machine and deep learning methods to find order quantities and a heuristic to perform multi-item inventory optimization when a capacity constraint is active.

To sum up, several of these novel approaches aim to optimize the solution of the NVP. These methods expand the NVP usage into different scenarios and integrate available data which is not supported in the original problem definition. For the purpose of our capstone, NVP will be the main mode of analysis for the materials with the most stable demand and the baseline model for the others.

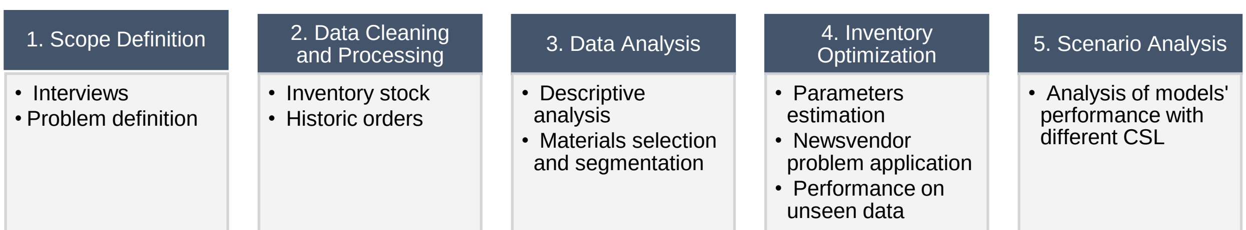
3. Methodology

After defining our research problem and looking for state-of-the-art methodologies, we concluded that the model that best applies to our objective is the Newsvendor Problem (NVP). Specifically, as we are dealing with stochastic demand and the order period will be unique (single), the NVP fits our objectives and needs.

As stated in Section 2, solving an NVP involves two objectives forecasting demand and optimizing the quantity to order. The company has a team specialized in demand forecasting of the finished product. However, the forecasts were not available, nor had they been disaggregated to a material level. Therefore, we focused on both steps. Our methodology is divided into five steps (see Figure 2).

Figure 2

Methodology



We started by defining the company's problem. Then we cleaned and processed the information provided, which included data on inventory and historical orders. After that, we investigated lead times, order quantities, and demand fluctuation. Finally, by using different variations of the Newsvendor Problem, we provided an optimal order quantity for the most problematic materials and assessed the historical performance such a method would have achieved if applied. Finally, we performed a scenario analysis by quantifying the cost of the order quantity suggested for different customer service levels (CSL).

3.1 Data Selection

3.1.1 Importance of Data Selection

Data exploration and selection play a crucial role in any data analysis project, particularly when working with large multinational corporations. Various plants, functions, and processes store information in separate tables, resulting in numerous databases containing vast amounts of data. The integration of these tables poses some challenges. During our exploration of the data, we encountered multiple issues, rendering a significant portion of the original dataset unusable. A comprehensive report outlining the interventions is available in section 4.1.

3.1.2 Data Sources

We analyzed two datasets to gain insights into the company's inventory management practices. The first dataset consists of all the records of inventory transactions in the factory between January 2021 and July 2022. The second dataset contains the materials master data, i.e., a table with the anagraphic data of materials, such as the item's description, associated planner, and lot size.

Dataset 1: Inventory transactions records from the factory

- Number of observations: 2.754B
- Relevant features of the dataset:
    - material: unique identifier for each material
    - plant: location of the plant where the material is stored
    - demand_date: date when the demand is expected according to the MRP
    - mrp_element_cd: code indicating the type of MRP transaction
    - change_quantity: inventory change in quantity associated with the transaction
    - total_quantity: the total quantity of inventory available after the transaction
    - supplier_nr: unique identifier for the supplier of the material

- planner: person responsible for planning the inventory of the material
- snapdate: date on which the inventory record was taken
- snaptime: time at which the inventory record was taken

We used this dataset to extract insights about inventory consumption, both under normal conditions and as materials approach their end-of-life.

Dataset 2: Materials Master Table

- Number of observations: 11.480M
- Relevant features of the dataset:
    - initial_creation_date: the date when the material was initially created in the system
    - last_change_date: the date when the material data was last modified
    - material: unique identifier for each material
    - psm: the type of replenishment used for the material
    - lot_size: the minimum order or production quantity for the material
    - safety_stock: the minimum amount of inventory to be held for the material
    - bulk_material: indicates if the material is a bulk material
    - active_material: indicates if the material is active
    - runout_date: the date after which the material becomes obsolescent

We used this dataset to extract the end-of-life date of the materials. It was also useful for retrieving other information, such as the type of replenishment or the lot size.

3.1.4 Data Aggregation

During this phase of our data analysis, we aimed at gaining a thorough understanding of the consumption patterns of our inventory, both during regular operations and during the runout phase. To achieve this aim, we computed the historical daily inventory consumption by calculating the differences between the total inventory available on a specific day and the inventory available on the previous day. We then utilized the materials master table to identify the end-of-life of each material in our inventory. This approach allowed us to evaluate whether consumption patterns were impacted as the material approached its end-of-life.

3.1.5 Unexpected Results

During this phase of data analysis, we discovered that:

1. Inventory transactions stop long before the planned end-of-life or continue long after, or runout dates happen in unexpected periods. This refutes the idea of runout dates as events that recur three times per year in predetermined periods, when a large number of materials would suddenly become obsolescent and never be utilized again.
2. The runout date of a specific material, when there is one, does not necessarily happen in a specific month. Having dates when many materials run out at once is a good practice, but not necessarily a rule.
3. There are two types of runouts: planned and unplanned. During an unplanned runout, certain materials are suddenly deemed obsolescent for safety or quality reasons. This kind of runout does not allow the time to plan a phase-out for obvious reasons.
4. A material may be utilized after its runout date has passed if the new material that it has been substituted with is not available (for production or procurement reasons), provided that the obsolescent material is still compatible with the rest of the vehicle.

Accordingly, we decided to continue searching for a general method to characterize materials' historical demand approaching their end-of-life, systematically discerning between those that became obsolescent when they were supposed to and those that did not. Furthermore, the optimization model to be developed will focus on a smaller selection of representative materials (see Section 3.1.6) that have a known runout date.

3.1.6 Selection Criteria

Due to their importance for the company, we concentrated on the materials with long procurement lead times, identified through their replenishment process. Specifically, we focused on "Bulk" and "Kanban" materials, which are the materials that are shipped from overseas. To characterize their value, we utilized an ABC classification provided by the company. To characterize their performance, we selected the scrap rate caused by obsolescence, expressed as the ratio between the number of units left after the end of life of a material and its average demand.

3.2 Data Analysis

In our data analysis we outlined the methodologies that we think the company should pursue to manage its inventory in the runout phase, based on the demand variability of the materials analyzed. We recommend traditional statistical models, such as the Newsvendor model, for those with stable demand, and more advanced approaches, like Time Series and Machine Learning models, for the volatile ones (with highly variable demand). This chapter concludes with an exploration of the training and testing processes for these models, emphasizing performance optimization despite the limited available history.

3.2.1 Stable and Volatile Materials

The most basic applications of the conventional Newsvendor model employ two characteristics to depict the material's demand pattern: its mean and standard deviation. When demand follows a normal distribution, this pattern can be defined solely by the average demand and a measure of its variability. In numerous real-world scenarios, the variability is represented through the mean square error of the forecast, which accounts for the remaining uncertainty after all predictable aspects of demand have been characterized in the forecast. However, since the company does not monitor its forecast error and deletes forecasts older than six weeks due to storage memory constraints, we had to use an alternative measure. The standard deviation of demand is a suitable metric when demand is not influenced by a significant trend or seasonality, which appears true for many of the materials studied. Therefore, we chose this method.

While standard deviation is a good proxy to model the uncertainty of the demands we analyzed, it is not a good metric to make comparisons between materials. Standard deviation measures absolute error, which depends on the magnitude of the demand. To obtain a measure of demand uncertainty that is normalized for magnitude, we assessed the coefficient of variation, which is the ratio of the standard deviation to the average demand. Our analysis revealed that the coefficient of variation for the materials in scope ranges between 1 and 6, with two primary clusters centered around 1.5 and approximately 3.

The considerable coefficient of variation observed prompted us to reassess some of our assumptions, particularly for materials with highly variable demand. In many instances, such elevated coefficients of variation signify sporadic demand that cannot be approximated by a normal distribution, requiring a more sophisticated model to characterize the demand pattern. Thus, we decided to create two separate modes of analysis. For materials with a stable demand, we utilized a traditional newsvendor model, assuming these materials follow a normal distribution. For materials with a more erratic demand, we opted for advanced methodologies recommended in the literature, originating from the fields of robust optimization and machine learning. The outlines of these methodologies can be found in our review of the literature in Section 2.2.

3.2.2 Assessment of the company's Performance on Stable Materials

The application of the newsvendor model to the most stable materials in our analysis was carried out following the conventional methodology outlined in Section 2.1.2. We tested the model using three distinct critical ratios: 95%, 99%, and 99.9%. These values represent the Cycle Service Level (CSL) of the model, which indicates the probability of avoiding stockouts if the model is employed. The selected values aimed to offer a wide range of models, from a high-risk, aggressive approach (95%) to an extremely conservative one (99.9%).

Under the assumption that the company had ordered the recommended quantity during the material's final replenishment before its end of life, we examined how historical consumption would have impacted inventory levels. This enabled us to determine whether the company would have faced a stockout and, if not, the amount of inventory remaining after the runout date. As expected, the 95% model resulted in lower leftover inventory but more stockout occurrences, while the 99.9% model experienced minimal stockouts but significantly larger remaining inventory.

Unfortunately, comparing our model results with the historical performance described by the company's databases was not possible. This is because operational staff members can manually adjust the runout date in their databases, pushing it forward in time to reduce scrapped inventory (usually until nearly all inventory is depleted). As a result, 60% of all materials have no remaining inventory after the runout date, and for the remaining materials, 85% possess inventory levels amounting to less than one-tenth of the average weekly demand. Stockouts are avoided because the flexibility of the runout date encourages materials planners to order more than necessary, since any excessive inventory can be consumed anyway.

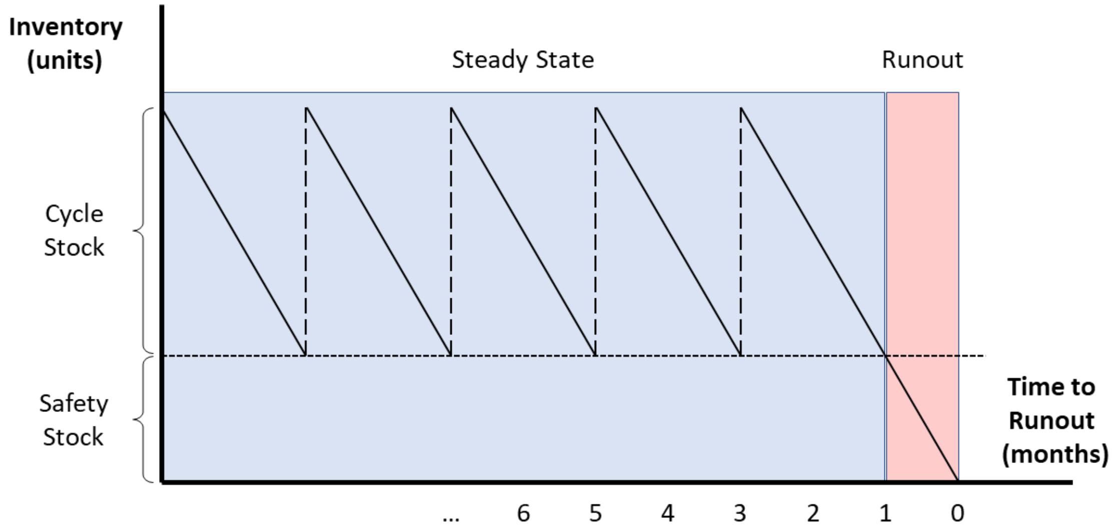
3.2.3 Objective Reformulation

Since the provided databases are unreliable for measuring excess inventory after a runout, in consultation with the company, we redefined the problem and established a new goal for the capstone project: to assess various end-of-life inventory management methodologies and determine which can deliver the best performance. The company aims to eliminate the practice of extending the runout date as part of its efforts to streamline operations; selecting a more effective policy will facilitate this change. Additionally, we decided to use the newsvendor model developed for stable materials as a reference point for evaluating the performance of more complex methodologies.

During the redefinition of the objective, we further clarified the feature to focus on in our prediction, defining the target metric for our optimization. Early in the project, the company had requested that we disregard their day-to-day inventory management policies and concentrate on the runout phase. To clarify the optimization objective, we agreed to distinguish between regular operations and runout at the point when cycle stock is depleted, and safety stock consumption begins (Figure 3). Since the company maintains approximately a month's worth of safety stock, our model aims to predict and optimize inventory consumption during the material's final month of life, using information available three months before the runout, which is when the last replenishment is typically scheduled.

Figure 3

Phases of Inventory Management



3.2.4 Prediction and Optimization Models

The chosen model should achieve two objectives: a good prediction of the inventory consumption and a subsequent optimization of the replenishment quantity. A good prediction is crucial to minimize the remaining inventory but is not sufficient for our purpose. Our objective, in fact, is to optimize for cost, and the large difference between the cost of holding too much inventory and the cost of holding too little means that the right replenishment quantity should keep the level of inventory as low as possible, but also avoid stock-outs.

The candidate models drawn from the literature fall into the categories of Newsvendor, Time Series, and Machine Learning models. By comparing the performance of these models, we sought to identify the most effective approach for preventing stockouts and minimizing the inventory to be discarded after the runout date. The following is a brief overview of the models that we considered.

Traditional Newsvendor Models:

- Parametric Approach: A statistical model that utilizes known (Normal) probability distribution for demand and optimizes order quantity based on cost parameters. This will also be the benchmark against which we compare the other models.
- Sample Average Approximation: Utilizes historical demand data to estimate the optimal order quantity without assuming a specific distribution.

Time Series Models:

- ARIMA: An acronym for AutoRegressive Integrated Moving Average, a statistical model that captures the autocorrelation, differencing, and moving average aspects of time series data to make predictions.

Machine Learning Models:

- Decision Tree: A hierarchical model that splits data into subsets based on feature values, leading to a decision or prediction.
- Random Forest: An ensemble method that combines multiple decision trees, improving generalization and reducing overfitting.
- XGBoost: An abbreviation for eXtreme Gradient Boosting, a scalable and efficient gradient boosting algorithm that constructs decision trees sequentially, minimizing a loss function and enhancing prediction accuracy.

3.2.4 Training and Testing over a Limited Time Interval

For the same reason of wanting to make the most out of the limited data available, we trained our models multiple times on the same time interval using a technique called sliding windows (Zhang et al., 2009). This approach leverages the ordered structure of events in a time series by training and testing on a certain time interval and then sliding it by a fraction of the interval's length.

A common concern associated with the use of sliding windows is data leakage. When evaluating the accuracy of a data model, it is crucial to ensure that the test data is entirely new to the model. If this condition is not met, it implies that the model has been partially trained on the same outcomes it is attempting to predict. This scenario can be likened to examining a student by asking a question that was previously discussed during a class session. The test results, in such a case, cannot be considered reliable. Data leakage is especially concerning when it occurs from the testing to the training set. Thus, we ensured that the time periods used for testing and training our models were kept distinct. Ideally, if the available data covered a sufficient time span, it would be a good practice to maintain separation between the intervals constituting individual training data points. However, in our case, the available observations are so scarce that implementing such a separation would result in an inadequate amount of training data. Consequently, we decided to be very strict about avoiding data leakage between testing and training set, but we allowed some data leakage among the cross-validation series within the training set.

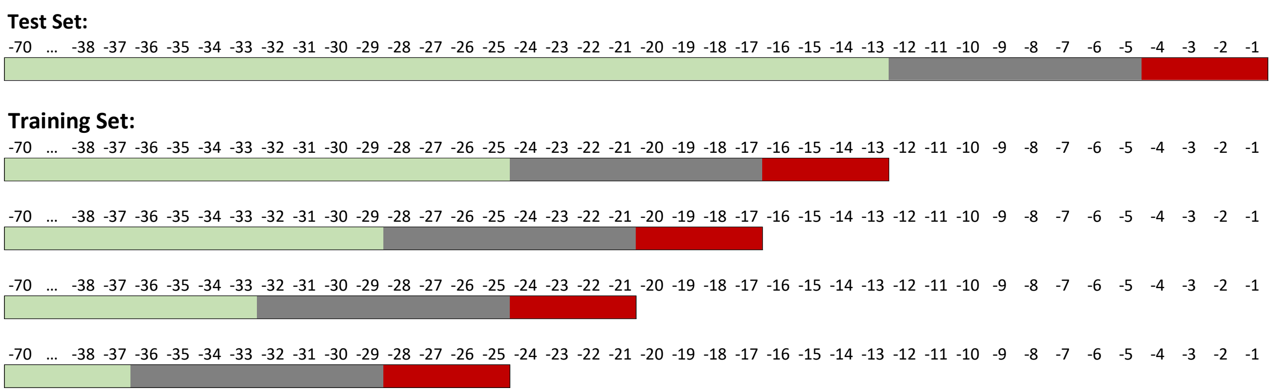
In Figures 4 and 5, we offer a graphical explanation of how the model manages the historical data. The colored bar represents historical inventory consumption for a specific material, and the numbers indicate the number of weeks before runout when the observation took place. The model utilizes the data differently based on the number of weeks remaining until runout. The green bar signifies the data is used by the model to make predictions by identifying patterns and adjusting hyperparameters. The last week of the green interval marks the point at which the reorder quantity decision must be made. The information about the subsequent eight weeks (in grey) is not used in the model. In fact, they cannot be used to improve the prediction, because at this point in time the reorder quantity has already been picked. The final four weeks (in red) represent the period during which the forecast is tested; if the model has accurately captured the patterns, it should yield a reliable prediction. While the underlying concepts are very similar, the way in which the Time Series model and the Machine Learning models divide the data into training and testing is slightly different, as explained in the following.

Time Series models

For the test set, the models utilize all available data. All observations up to week -13 are used to predict inventory consumption during the final four weeks before runout. As previously discussed in the context of data leakage, it was crucial to ensure that none of the data used for testing is available in the training set. Consequently, for the first cross-validation, the model was trained by shifting the time window 13 weeks ahead. Data leakage between cross-validation sets within the training set is not as critical an issue, so for the remaining cross-validation sets, we shifted the time window by only four weeks at a time.

Figure 4

Test and Training Set for Time-Series Models

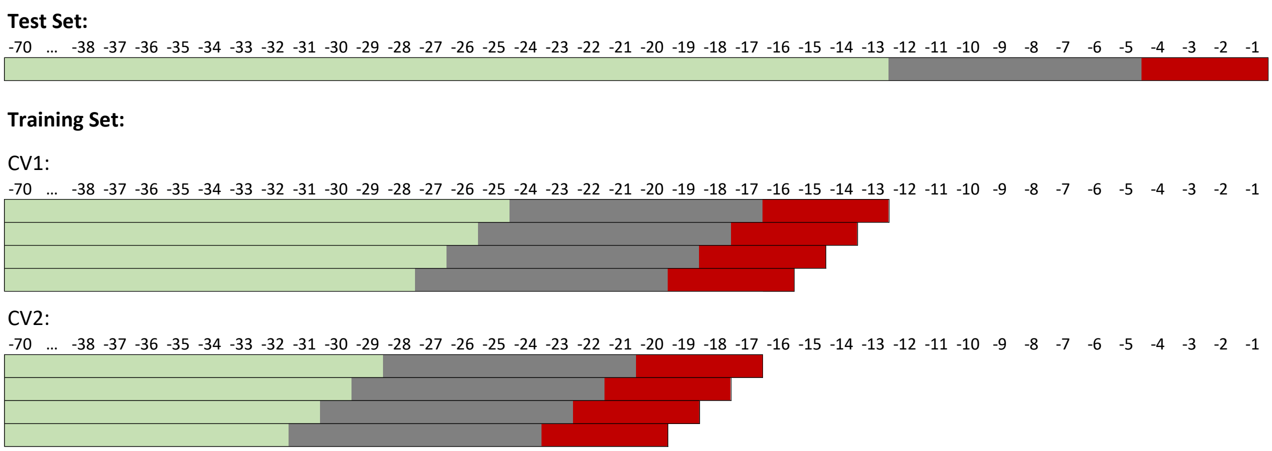


Machine Learning models

The reasoning behind the division between training and testing for Machine Learning models is similar to that of time series models. The only distinction between the two categories is that we used 4-fold cross-validation sets for Machine Learning, as these models typically demand higher data density, which could not be supplied with the partition used for Time Series. Consequently, for every data point in the time series models, the Machine Learning models can utilize four.

Figure 5

Test and Training Set for Machine Learning Models



The pseudo code in Figure 6 describes the algorithm to be implemented for every material. Therefore, three months before the intended runout date, for each material, the following process should be followed. First, based on a pre-defined set of models and hyperparameters' ranges, conduct a grid search by creating models that use a different set of hyperparameters. For every fold, train each model with the training set available and predict the demand for the validation set. With this prediction calculate inventory policy cost and obtain the average units short when a stockout occurred at the validation set. Repeat this for every fold and every parameter combination. After this, select the best model by minimizing the average inventory policy cost in the validation set. Then, obtain the average units short that the best model incurred when stocking out. Continue with training the best model with all the data available up to that point and predict the demand for the last month of the material's lifetime. Then, for every CSL of interest, include a buffer at each prediction by adding the average units short of the best model on unseen data. Finally, when the data is available, calculate the inventory policy cost.

Figure 6

Pseudo Code for the Implementation of Machine Learning Models

```
for each material:
    for each parameter combination in the grid search:
        for each fold k:
            Train model with training set.
            Evaluate model using validation set.
            Calculate cost of inventory policy.
            Store the average units short during a stockout.
    Select the model that minimizes the cost.
    Obtain the average units short during a stockout.
    Train best model with all the data available.
    Predict demand for the last month (F_m).
    for each CSL:
        Include buffer based on the error: F = F_m + ϵ Φ(CSL)
        Calculate the inventory policy cost.
```

4. Results and Recommendations

4.1 Analysis Results

4.1.1 Data Cleaning Report

In this section, we outline the data cleaning process of the database provided by the partner company. Our goal was to identify and address any issues in the data that could undermine our subsequent analysis, such as duplicate records and missing values. The dataset provided by the company contains 7 million records for recent movements (after 2022-06) and 23 million records for historical movements (before 2022-06). The table of runout dates included 25 thousand unique materials. 428 materials have more than one runout date.

Of these 25 thousand materials, 12 thousand have a runout date before December 2022 and are hence suitable for data analysis. During the analysis, we encountered a series of issues:

- 9.0 thousand materials (75%) do not have an assigned procurement strategy.
- 8.5 thousand materials (71%) did not have cost information.
- 7.5 thousand materials (60.6%) have zero inventory after runout date. The runout date has been adjusted accordingly.

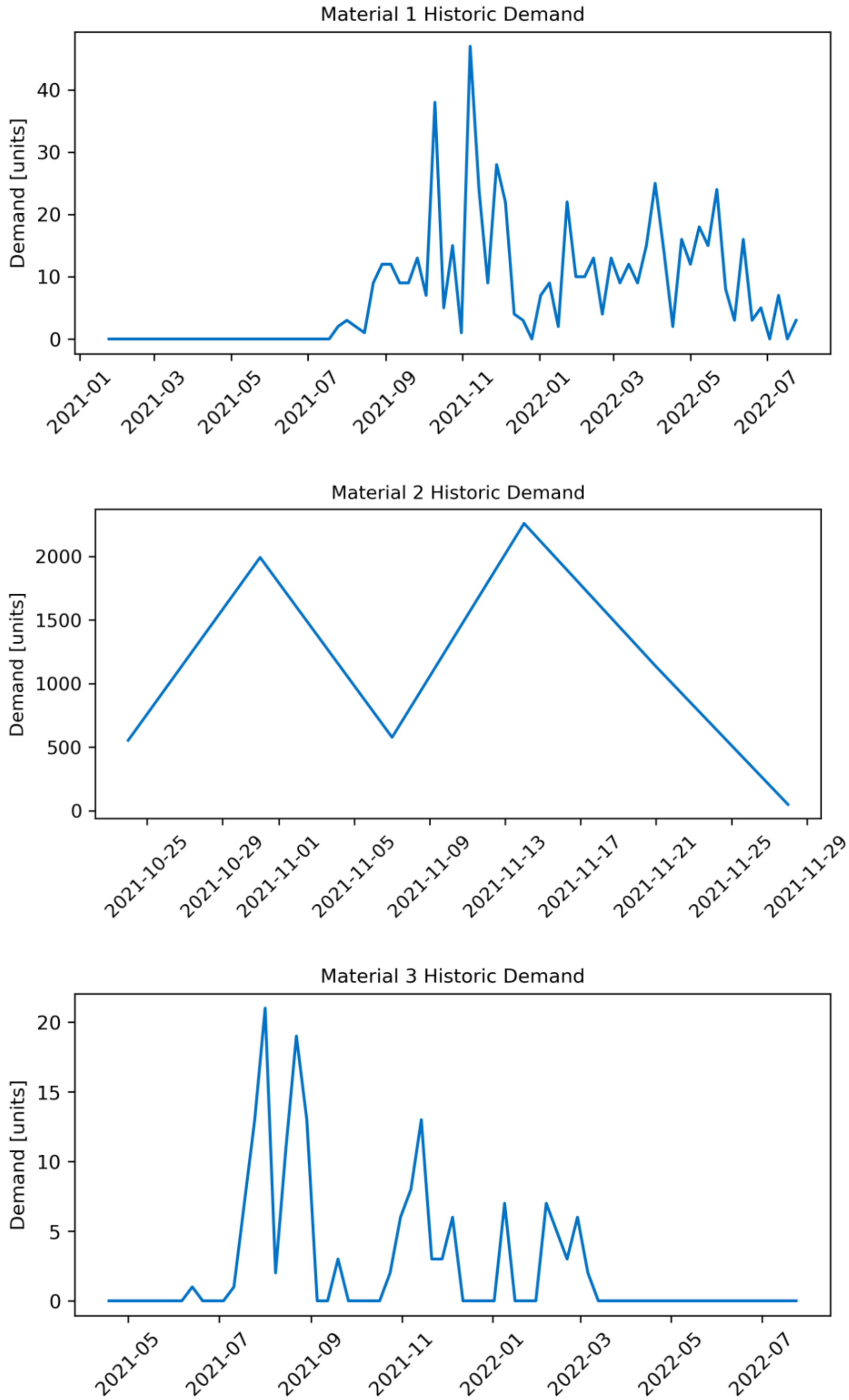
Many of these 12 thousand materials, moreover, present demand history issues.

- 75% of materials presents an initial period of almost null demand. See Figure 7(a)
- 37% of materials present less than 24 weeks of historical data. See Figure 7(b)
- 38% of materials present null demand during their last month of life. See Figure 7(c)

For the purpose of the analysis, we selected data with at least 70 weeks of history, and modified the data to avoid the data quality problems just mentioned.

Figure 7

Demand history issues. Null initial demand (a), Short history (b), Null final demand (c).



(a)

(b)

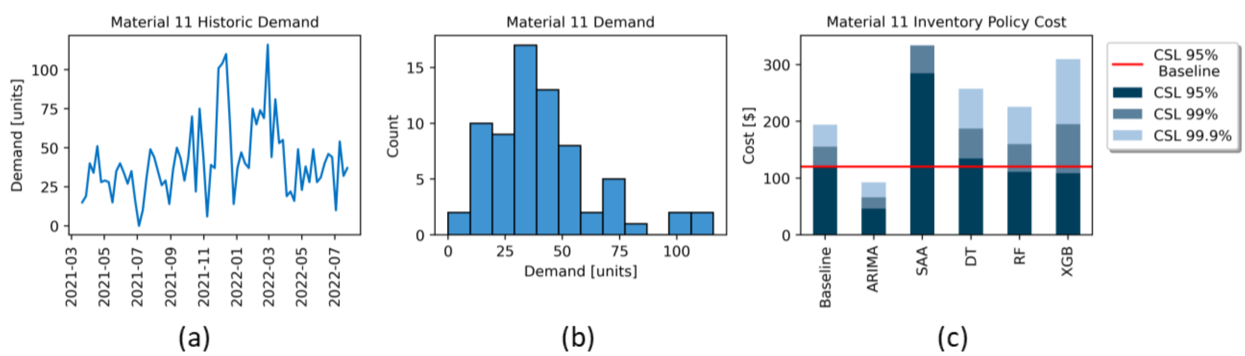
(c)

4.1.2 Analysis Report

By implementing the models outlined in Section 3.2.4 to the small set of materials qualified for the analysis (see Section 4.1.1), we were able to estimate the relative costs at various Cycle Service Levels (95%, 99%, 99.9%) associated with each model. An example of these results can be seen in Figure 8. Figure 8(a) displays the historical recorded demand for the analyzed material. Figure 8(b) shows the distribution of the daily demand of the material. Figure 8(c) illustrates the cost for each applied model. The Parametric Newsvendor Model cost with a 95% CSL was chosen as the baseline (red line). For this particular material, we observe that the ARIMA model outperforms the baseline (lower cost) at any service level, while the Random Forest model performs slightly better at a 95% CSL. This example aims to explain two important ideas. First, that different models will result in significantly different policies, with different costs. Choosing the optimal strategy for every material is especially important. Second, the higher the service level, the higher the cost.

Figure 8:

Results of the study for a sample material



When generalizing our results across the entire set of available materials, we found that no single model consistently outperforms the others. Instead, the best model appears to depend on the specific demand history of the material in question. Figure 9 shows the remaining inventory after the material runouts, expressed as a ratio of the remaining inventory divided by the total demand during the last month of the material's lifetime. Some models consistently stock much more than needed. This is especially true for materials like the one previously showed in Figure 7(c), which experience a sudden drop in demand towards the end of their life. Since our models are solely based on historical data, there was no way to foresee such a sudden decrease in inventory consumption in the last month. Unless we incorporate external information into the models (such as vehicle-level forecasting), our approach remains unprotected against such events.

Figure 9:

Remaining inventory after runout date

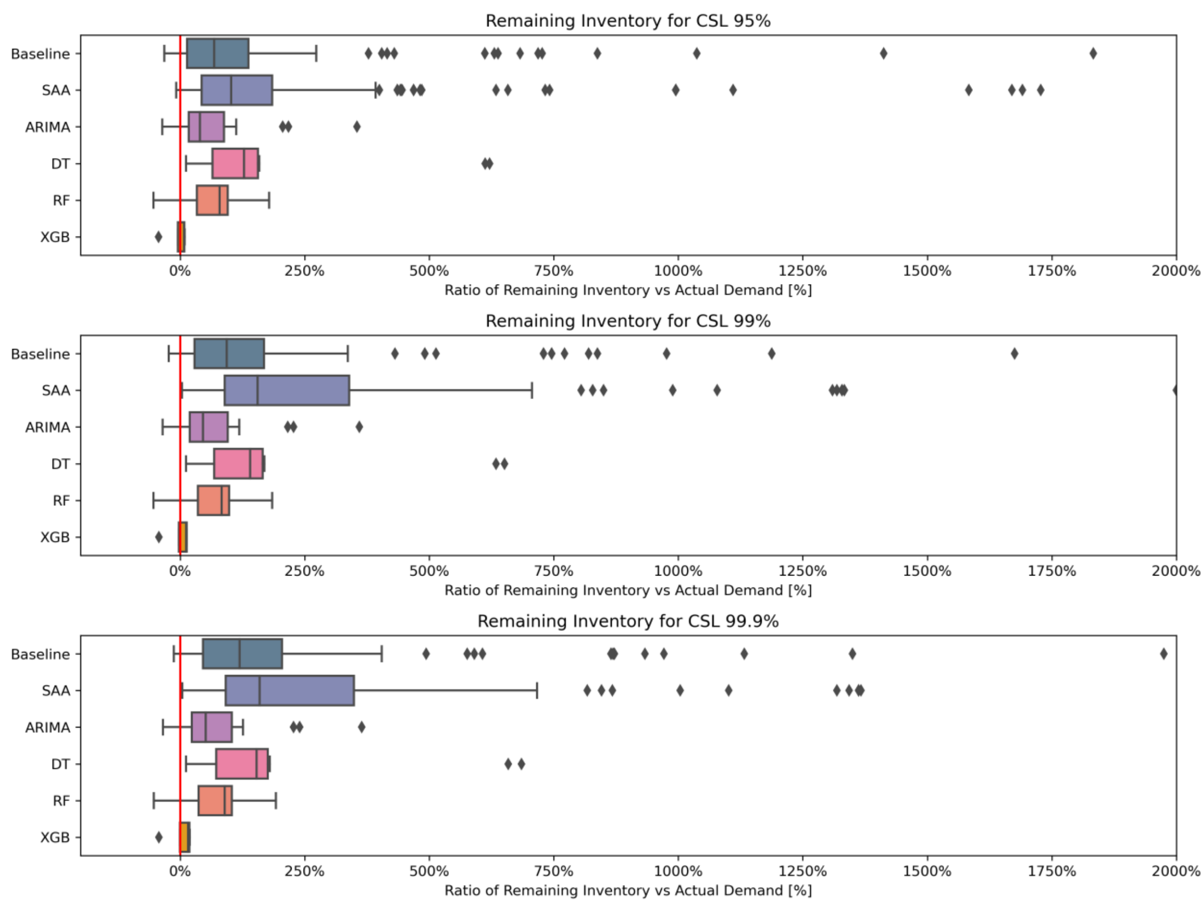


Figure 10:

% Cost with respect to baseline.

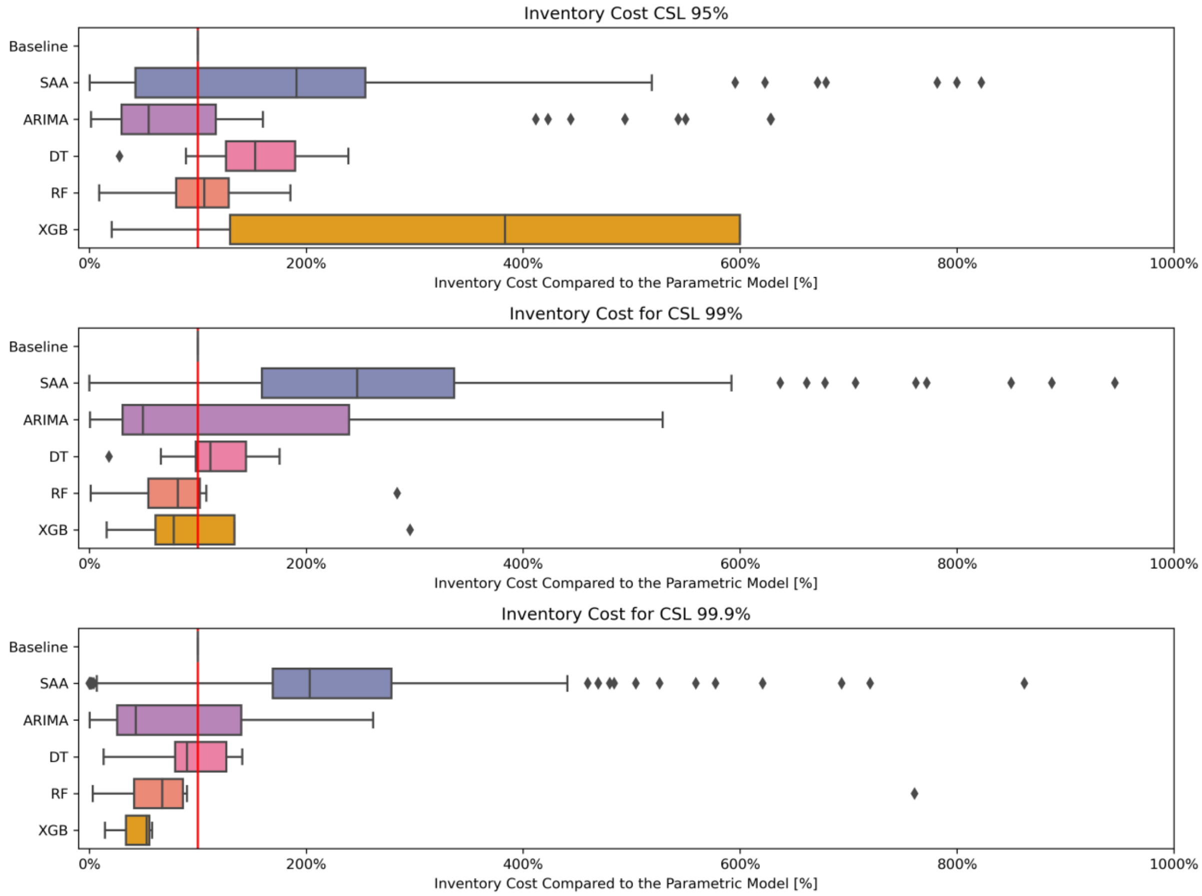


Figure 10 displays the performance of various models compared to the baseline (the Newsvendor Parametric model). Some models demonstrate unusual behavior. For instance, XGBoost exhibits subpar performance at a 95% CSL but progressively improves as the service level increases. However, due to the limited size of the dataset, we cannot conclusively determine the repeatability of such behavior.

In addition to cost savings, another aspect of the models' performance is their ability to prevent stockouts. Table 1 presents the number of times each model was selected as the most cost-effective option. It indicates the percentage of times, out of the instances when the model was chosen, that the selection resulted in a stockout.

Table 1:

Relative model performance and stockout avoidance

	Parametric (Baseline)	ARIMA	SAA	Decision Tree	Random Forest	XGBoost
Number of materials where the model achieved the best performance	101	45	186	12	10	4

Percentage of materials that stocked out when the model was chosen	**Parametric**	**ARIMA**	**SAA**	**Decision Tree**	**Random Forest**	**XGBoost**
95% Service Level	17%	11%	2%	0%	20%	25%
99% Service Level	6%	11%	0%	0%	10%	25%
99.9% Service Level	1%	9%	0%	0%	10%	25%

We suggest caution in the interpretation of the results of this analysis: because of the size of the dataset analyzed, our results may not possess the necessary level of reliability to support definitive conclusions. The objective of presenting these findings is to show a working example of the application of the methodology and to explore the use of different models and their differential impacts on number of stockouts and overall cost.

4.2 Recommendations

Based on our findings, we present a set of recommendations for the company to improve its demand forecasting and inventory management. Our suggestions address the limitations identified during our capstone project and leverage the insights gained from our analysis. Particularly, we focus on three areas: data richness and completeness, the relationship between forecasting and inventory management policies, and the utilization of our methodology.

The data cleaning process conducted in this capstone project highlighted the importance of data richness and completeness for accurate and reliable analysis. We recommend that the company invest in data management practices, including data validation, standardization, and consolidation. This will ensure that future projects have access to higher-quality data, leading to more accurate and reliable forecasts and decision-making. A notable limitation encountered in the analysis was the brief historical data available for many of the examined materials. One approach the company could employ to better utilize their current data is to combine the historical data of different materials as if they were a single entity. For instance, the company could use expert knowledge to determine which material replaced another during an update, clarifying the relationships between SKUs across multiple generations. This would facilitate the creation of "chains of materials' versions", enabling models to be trained on a more extensive history. Another complementary approach could involve aggregating materials that are alternative to one another or that belong to the same product family. Aggregating would make our forecasts less error prone by capturing relationships and interdependencies between materials.

To fully capitalize on the potential benefits of improved demand forecasting, the company should focus on the integration of its inventory management and forecasting systems. As demonstrated in our capstone project, these two functional domains are closely interconnected. One immediate improvement that the company could implement is tracking the forecasting error. This would not only facilitate the use of models such as the newsvendor model, but also have cross-functional benefits on inventory management. Our observations indicate that because of the excessive scrap inventory, the company's operations often result in the postponement of materials' runout dates and delays in utilizing new SKU versions in production. Poor predictions of inventory consumption led to suboptimal performance in terms of inventory management. Monitoring forecasting errors and enhancing the accuracy of demand forecasts appear to be crucial steps for improving this aspect of the operations. We also think that the company should take advantage of finished product forecast information. The vast number of configurations that the company produces makes it difficult to link individual material forecasts to finished product forecasts. However, assuming that demand is independent across materials is inaccurate, and excluding finished product forecasts altogether weakened the power of the predictions.

Regarding the deliverables of our capstone project, we suggest that the company should utilize expert knowledge and a model selection framework to choose the most suitable forecasting model. Since no single forecasting model consistently outshines the others, the company should assess multiple forecasting models for each material, considering their demand history and specific characteristics. Further research should be conducted to identify the criteria determining their success and develop guidelines for selecting the most appropriate model for each material. To enhance the selection and application of forecasting models, the company should involve individuals who have firsthand experience in handling the materials in question. These experts could evaluate the results of the study and offer insights into the distinct demand patterns of specific materials. By incorporating their knowledge, the company can gain a better understanding of the nuances of each material's demand and refine the forecasting models accordingly.

The absence of cost information restricted our ability to prioritize materials based on their economic impact, treating all materials in the analysis as equally important for the company, which is likely not the case. If cost information were available, we could better align inventory management policies with the value of the materials, exercising greater caution with high-value materials and being less conservative for cheaper materials.

5. Conclusions

Nowadays, managing inventory effectively and efficiently is mandatory for any company that wants to stay competitive. A careful balance of forecasting and optimization, as well as complex decision-making processes, are required to find the right compromise between minimizing costs and providing an appropriate service level to customers, all within a highly uncertain environment. The complexity of the bills of materials, intense competition, and narrow profit margins, all of which are defining features of the industry in which our partner company operates, magnify the importance of robust inventory management. In this industry, improving inventory management policies can make a difference in companies' competitive advantage.

One of the most crucial aspects of inventory management is the runout stage. Maintaining excessive inventory not only leads to increased holding costs but also incurs significantly higher costs associated with scrapping the full value of excessive inventory. Mastering runout inventory management means finding the optimal balance between minimizing stockouts and overstocks while maximizing profitability and customer satisfaction. Although it can be challenging, improving runout inventory management is particularly valuable in industries like the one in which our partner company operates, where the product life cycle for raw materials is relatively short and new versions are constantly being introduced. To address these challenges, developing innovative tools and strategies for managing runout inventory effectively is essential. By doing so, our partner company can reduce the costs associated with rapidly changing iterations, streamline its processes, and ultimately secure a significant competitive edge in the industry.

Throughout this report, we have illustrated an end-to-end approach that companies should follow to improve their runout inventory management. Specifically, we presented a study about how to define the problem of runout inventory management, prepare the raw data, choose the models, set up the analysis, and interpret its results. Furthermore, we outlined several recommendations for our partner company to improve its demand forecasting and inventory management processes successfully: greater richness and completeness of the data, and better integration between the forecasting and inventory management functions.

The proposed methodology can be improved and strengthened. The inclusion of additional information could significantly improve the prediction accuracy. For instance, adding data on external factors or consumption correlations between materials, which might play a role in how the demand per material occurs, and increase the model's predictive power. Moreover, reducing the prediction time horizon would reduce complexity and increase accuracy. Even though planning for an aggregate of materials is a lower-resolution decision (does not provide the same number of details), these process constraints play a significant role and should not be overlooked.

In summary, our contribution to the existing literature is a comprehensive and pragmatic guide for incorporating advanced analytics into inventory management practices. By presenting an authentic example, we not only demonstrated the application of these techniques in real-world scenarios but also provided valuable insights into how to navigate the challenges that may arise.

References

Ban, G.-Y., & Rudin, C. (2019). The Big Data Newsvendor: Practical Insights from Machine Learning. *Operations Research*, *67*(1), 90–108. https://doi.org/10.1287/opre.2018.1757

Bertsimas, D., & Thiele, A. (2006). A Robust Optimization Approach to Inventory Theory. *Operations Research*, *54*(1), 150–168. https://doi.org/10.1287/opre.1050.0238

Choi, T.-M. (Ed.). (2012). *Handbook of newsvendor problems: Models, extensions and applications*. Springer.

de Castro Moraes, T., & Yuan, X.-M. (2021). Data-Driven Solutions for the Newsvendor Problem: A Systematic Literature Review. In A. Dolgui, A. Bernard, D. Lemoine, G. von Cieminski, & D. Romero (Eds.), *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems* (Vol. 633, pp. 149–158). Springer International Publishing. https://doi.org/10.1007/978-3-030-85910-7_16

Levi, R., Roundy, R. O., & Shmoys, D. B. (2007). Provably Near-Optimal Sampling-Based Policies for Stochastic Inventory Control Models. *Mathematics of Operations Research*, *32*(4), 821–839. https://doi.org/10.1287/moor.1070.0272

Liyanage, L. H., & Shanthikumar, J. G. (2005). A practical inventory control policy using operational statistics. *Operations Research Letters*, *33*(4), 341–348. https://doi.org/10.1016/j.orl.2004.08.003

Muller, M. (2011). *Essentials of inventory management* (2nd ed). AMACOM.

Nahmias, S. (1982). Perishable Inventory Theory: A Review. *Operations Research*, *30*(4), 680–708. https://doi.org/10.1287/opre.30.4.680

Oroojlooyjadid, A., Snyder, L. V., & Takáč, M. (2020). Applying deep learning to the newsvendor problem. *IISE Transactions*, *52*(4), 444–463. https://doi.org/10.1080/24725854.2019.1632502

Punia, S., Singh, S. P., & Madaan, J. K. (2020). From predictive to prescriptive analytics: A data-driven multi-item newsvendor model. *Decision Support Systems*, *136*, 113340. https://doi.org/10.1016/j.dss.2020.113340

Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., & Seref, M. M. H. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research*, *213*(2), 361–374. https://doi.org/10.1016/j.ejor.2010.11.024

Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2022). A review on reinforcement learning algorithms and applications in supply chain management. *International Journal of Production Research*, 1–29. https://doi.org/10.1080/00207543.2022.2140221

Scarf, H. (1959). Bayes Solutions of the Statistical Inventory Problem. *The Annals of Mathematical Statistics*, *30*(2), 490–508.

Silver, E. A., & Pyke, D. F. (2017). *Inventory and production management in supply chains* (Fourth Edition). CRC Press, Taylor & Francis Group.

Vrat, P. (2014). *Materials management: An integrated systems approach*. Springer.

Zhang, T., Yue, D., Gu, Y., Wang, Y., & Yu, G. (2009). Adaptive correlation analysis in stream time series with sliding windows. *Computers & Mathematics with Applications*, *57*(6), 937–948. https://doi.org/10.1016/j.camwa.2008.10.083

Ziukov, S. (2015). A literature review on models of inventory management under uncertainty. *Business Systems & Economics*, *5*(1), 26. https://doi.org/10.13165/VSE-15-5-1-03