Enhancing Supply Chain Operating Models Through Segmentation

by

Ibrahim AlArfaj
Bachelor of Science in Industrial Systems Engineering
(King Fahd University of Petroleum & Minerals, KSA)

and

Yalcin Arslan
Bachelor of Science in Industrial Engineering
(Bilkent University, Turkey)

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

May 2022

Signature of Author: ____

Ibrahim AlArfaj
Department of Supply Chain Management
May 6, 2022

Signature of Author: ____

Yalcin Arslan
Department of Supply Chain Management
May 6, 2022

Certified by: ____

Özden Tozanlı Yılmaz
Postdoctoral Associate
Capstone Advisor

Accepted by: ____

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

Enhancing Supply Chain Operating Models Through Segmentation

by

Ibrahim AlArfaj

and

Yalcin Arslan

Submitted to the Program in Supply Chain Management
on May 6, 2022, in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Large Food and Beverage retail chains often manage diverse sets of products and markets, where one-size-fits-all supply chain operating models are insufficient to meet their distinct requirements. In collaboration with a global retailer, the main objective of this study is to identify distinguishable supply chain segments based on product and market characteristics and design an alternative supply chain operating model (SCOM) for each segment. To achieve this, a five-step, integrated, data-driven methodology is designed. First, data is gathered and reviewed for accuracy and completeness. Second, data is analyzed to identify potential segmentation criteria and select the most relevant factors. Third, k-means clustering is applied to create the product segments. Fourth, a SCOM is designed for each segment based on the product characteristics. Finally, the SCOMs are simulated to analyze how they perform in different scenarios. Applying the methodology resulted in three segments differentiated based on the products' demand volume, demand volatility, shelf-life, item cost, and seasonality. The three segments are slow-moving, fast-moving, and complex items. Each segment was recommended to be managed using different inventory and forecasting policies. Using simulation and scenario analysis, several service level targets were tested to show how they impact inventory costs, transportation costs, and fill rate. As a result, the SCOM for each operating model brings benefits to the overall performance. Focusing on this, slow-moving products are not delivered frequently, hence eliminating their inventories in the DCs is expected to reduce the inventory holding cost without significantly increasing the transportation cost and decreasing service levels. Disaggregating the inventory in the CDCs for fast-moving items is expected to improve service levels for these items, with a low increase in inventory costs. Lastly, aggregating the demand for complex items is expected to reduce the risks of stockout and excess inventory. The methodology in this study can be generalized to other industries with high product variety to enable them to reduce inventory, improve service level, and reduce the total distance traveled.

Capstone Advisor: Dr. Özden Tozanlı Yılmaz
Title: Postdoctoral Associate at the MIT Center for Transportation & Logistics

ACKNOWLEDGMENTS

We would like to thank our incredible advisor, Dr. Özden Tozanlı Yılmaz for her constructive feedback throughout this project. We appreciate her time and support throughout this journey.

We are also thankful for the team member from our sponsoring company. They helped us understand the business, improve our results, and learn along the way.

A special thanks to Toby Gooley for her support and suggestions that included both writing and supply chain improvements. Her comments helped enhance this report significantly.

Finally, we would like to thank our professors and classmates in the Supply Chain Management program at MIT for the wonderful experience, learning, and friendship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Supply chains face many challenges, such as lead time and demand volatility, stock-keeping unit (SKU) proliferation, introduction of new channels, and diverging customer demands. These challenges occur because of the inherent complexity and uncertainty in supply chain systems, especially when operating in a global environment. The complexity intensifies even further given the continuously changing and expanding customer needs. Given these complexities and challenges, one-size-fits-all supply chain operating models are insufficient to meet the distinct requirements of products, customers, and distribution channels. Treating different supply chain segments in the same way can negatively impact operational performance. With this motivation, companies now seek ways to find different operating models to cost-effectively meet the customer demand at the right time, right place, and in the right quantity (Pereira et al., 2020).

Segmentation serves as a valuable tool to find supply chain operating models that meet the unique attributes and needs of products, customers, and distribution channels. Segmenting products by volume, delivery lead time, and product variety or segmenting customers by order size, service expectation, and strategic importance enables companies to define unique supply chain strategies for each segment. Some of these strategies involve prioritizing the orders of high-paying customers, eliminating the risk of stock-out on high-value products, and making various service-level agreements with customers. Adopting different operating models for particular segments results in increased profitability, improved service levels, and reduced delivery lead times. Lovell et al. (2005) examined segmentation to pool inventories of similar products, which led to a reduction in safety stock and warehousing costs (Lovell et al., 2005). Another study developed by Bennion (1987) showed that supply chain segmentation facilitates improvement in product quality, flexibility, and order policy. These studies emphasize the importance and the potential benefits of market and product segmentation in different industries.

Specifically, segmentation can be considered a critical strategy in the food and beverage retail industry. The products in this industry have unique characteristics that require special handling, such as seasonality, perishability, and temperature requirements. Furthermore, food retailers need to manage several products that are not for sale such as those used for cooking, packaging, or serving which are called non-point-of-sales or non-POS items. This complexity in product types and attributes requires a commensurate diversity in ways to manage these products, which emphasizes the need for a segmented supply chain strategy.

For this study, we are collaborating with a global retailer to explore how segmentation can help overcome their supply chain challenges. This company has been experiencing similar challenges due to increased supply chain complexity. It has been expanding in its locations, customer base, and product selections ever since the business started. Currently, the retail chain operates thousands of stores in North America. Within each store, a variety of products are distributed, stored, and served, including food, beverages, and paper and plastic kitchenware. Each of these products has unique characteristics that require different strategies to handle, store, and distribute. With this expansive supply chain network and product base, effective management of the supply chain operating models (SCOMs) is critical to minimizing operating costs and improving sustainability, while providing excellent service.

Given the variety of product characteristics, the company aims to design segmented SCOMs and optimize supply chain cost, sustainability, and service metrics based on differentiating factors such as point of sales (POS) vs. non-POS items, ambient vs. temperature-controlled, frozen vs. fresh, and many others. This requires identification of key distinguishing factors, segmentation based on these selected factors, and design of a unique end-to-end supply chain operating model for each segment.

1.1 PROBLEM STATEMENT AND OBJECTIVES

Despite the diverse and expanding supply chains managed by the retail chain, its SCOMs currently do not consider product and/or market segmentation. In the existing supply chain design, most products generally flow from a supplier to a Regional Distribution Center (RDC), then to a Consolidated Distribution Center (CDC), and finally to a store. Different SCOMs are currently followed based on case-by-case requirements. Stores can receive products directly from suppliers, from RDCs, or from CDCs. However, these varying distribution channels do not consider product segmentation. This lack of segmentation leads to numerous issues in cost and service level, as well as challenges in space, transportation, and capacity constraints.

This project aims to improve the supply chain performance by segmenting similar products, identifying differences and similarities between product characteristics, and designing alternative SCOM models for unique segments. The scope of product segmentation includes SKUs from the following categories: food, beverages, paper and plastic kitchenware, and packaged coffee. The differentiation in SCOMs includes various requirements and attributes such as forecasting aggregation level, temperature requirements, inventory policies, and distribution and logistics models. These SCOM designs are used to investigate performance in terms of cost, CO_2 emissions, and service level while considering trade-offs among these measures and deciding on the optimal combination based on company strategy.

2 LITERATURE REVIEW

Supply chain segmentation has been the focus of numerous studies, with the common goal of enabling efficient management of supply chains with respect to different attributes. The literature offers a vast number of studies that explored segmentation in different industries based on attributes related to products, customers, markets, or a blend of these categories. These studies delved into segmentation in several industries, such as basic (Bennion, 1987), sporting goods (Roscoe & Baker, 2014), apparel (Ptok et al., 2021), electronics (Gosling & Urrutia, 2019), automotive (Protopappa-Sieke & Thonemann, 2017), and many others. This section provides a literature review of segmentation studies conducted in the retail industry and other similar industries. Specifically, we first present different methodologies for performing product, market, and hybrid segmentation. Then, we explore how segmentation was utilized to improve sustainability, cost, and customer service metrics. Finally, we review studies that apply segmentation in the supply chains of food, beverages, kitchenware, and packaged coffee.

2.1 SUPPLY CHAIN SEGMENTATION APPROACHES

Supply chain segmentation describes an approach to designing and operating different supply chain strategies in line with value drivers such as cost, quality, and service level (Childerhouse et al., 2002). These differentiated supply chain designs are created based on segments that entail a set of products, customers, or regions grouped by defined criteria. Protopappa-Sieke and Thonemann (2017) classified supply chain segmentation frameworks in the academic literature into three categories: market-driven, product-driven, and hybrid segmentation. These three categories are discussed in detail in the following sections.

2.1.1 MARKET-DRIVEN SEGMENTATION

Market-driven segmentation is described as a methodology to develop different supply chain segments based on market factors. It primarily seeks to understand customer service expectations and behaviors, identify diverse customer segments, and design supply chain models to address the needs of each segment. Some factors for market-driven segmentation include customer service expectations for cost, product availability, delivery lead time, and customer buying behaviors such as price sensitivity and demand variability (Protopappa-Sieke & Thonemann, 2017).

Hill (1995) is one of the first articles that introduced market-driven segmentation approaches. He suggested creating customized manufacturing strategies based on the ranked importance of different criteria in varying markets. Hill breaks down those market criteria as “order winners” and “order qualifiers.” Order winners represent features that make a product or brand desirable for customers, such as quality, cost, design, or delivery reliability. Order qualifiers, on the other hand, are characterized as a set of screening criteria that a company must fulfill to even be qualified as a supplier, like the minimum required quality and safety standards. Hill proposes that companies must develop different manufacturing strategies to satisfy different order qualifiers’ and order winners’ criteria of customers, as opposed to developing a single strategy.

Other researchers further developed the market-driven segmentation approaches beyond manufacturing strategies. Walters (2006) proposed developing a differentiated customer service response for different segments based on customer service requirements. Christopher & Gattorna (2005) suggested segmenting the market based on customer purchasing behavior (including demand predictability and price sensitivity) and designing different supply chain strategies ranging from fully flexible, agile, and lean, to continuous replenishment. Lovell et al. (2005) discussed a list of market-related factors impacting decisions on market segmentation and supply chain design, such as demand location, level, variability, and service expectations. These studies analyzed different factors that are worth exploring for retailers to aid in segmenting markets or customers and developing differentiated service responses based on varying needs.

2.1.2 PRODUCT-DRIVEN SEGMENTATION

Product-driven segmentation aims to create product-based segments based on various product, demand, and supply characteristics. The characteristics studied by many researchers as part of product-driven segmentation entail product type (standard vs. special products), demand volatility and volume, replenishment lead time, and supply risk (Protopappa-Sieke & Thonemann, 2017).

Product-driven segmentation was initially introduced by Fisher (1997). Fisher classified products into two categories: functional and innovative. Functional products are considered to have predictable demand, longer product lifecycle, lower contribution margin, and lower variety. In contrast, innovative products have unpredictable demand, shorter product lifecycle, higher contribution margin, and higher product variety. These two types of products must be treated differently and must have different supply chain strategies. According to Fisher, companies should prioritize supply chain efficiency for functional products and supply chain responsiveness for innovative products. An efficient supply chain is characterized by decreasing costs, and a responsive supply chain is characterized by meeting unpredictable demand.

More recent papers further developed Fisher's product segmentation approach by incorporating supply characteristics. Lee (2002) suggested segmenting products by not only demand uncertainty but also supply uncertainty based on breakdowns, quality problems, capacity constraints, reliable suppliers, and lead-time volatility. Christopher et al. (2006) also proposed segmenting products by supply (replenishment lead time), and demand characteristics (stable, volatile). These studies further expanded product-driven segmentation frameworks with a more holistic approach, resulting in segmentation that accounts for a variety of product characteristics.

2.1.3 HYBRID SEGMENTATION

Hybrid segmentation frameworks seek to create segments based on a combination of customer requirements and product characteristics. As both product characteristics and customer requirements significantly impact supply chain decisions, hybrid segmentation can be more beneficial than solely adopting product-driven and market-driven segmentation strategies (Terlunen et al., 2013)

Many researchers integrated customer and product characteristics into the segmentation approach. Godsell (2011) suggested a dual segmentation approach that incorporates product-based variables (product lifecycle duration, delivery lead times, volumes, variety, and variability) as well as market requirements (demand variability and order volume). In their comprehensive paper, Protopappa-Sieke & Thonemann (2017) suggested segmentation criteria based on product and demand characteristics, channel and customer characteristics, and supply characteristics. Those characteristics are presented in Table 1.

Table 1

Segmentation Criteria Based on Product, Customer, and Supply Characteristics

Segmentation Category	**Segmentation Criteria**
Product and demand characteristics	Demand volume
	Demand volatility
	Product lifecycle
	Forecasting ability
	Product complexity
	Product value
	Product relevance
	Contribution margin

Channel and customer characteristics	Customer type Customer specifics Customer priority Customer requirements Channel type Order type
Supply characteristics	Component supply flexibility Component lead time Component supply reliability Supply process volatility Supply capacity constraints Component value

Note. Adapted from Supply Chain Segmentation (p 16) by M. Protopappa-Sieke and U. W. Thonemann, 2017, Springer, Copyright 2017 by Springer

To achieve effective product, market, or hybrid segmentation for complex supply chains, several advanced methodologies were explored in the literature. For example, Svoboda and Minner (2021) utilized a genetic algorithm to develop decision trees that minimize total inventory costs. In addition to Svoboda and Minner's approach allowing for optimized classification, other approaches were used to improve the way each segment is managed. These methodologies include constrained optimization (Fichtinger et al., 2019), unconstrained optimization (Feng et al., 2019), and discrete rate-based simulation (Terlunen et al., 2015). While these approaches can be employed for any supply chain segmentation problem, we focused on papers that apply segmentation methodologies to food and retail supply chains. We also reviewed segmentation-related papers that aim to improve several objectives aligning with the company's goals, such as improving sustainability, cost, and service level. These applications and objectives aligning with the company's business and strategy are presented in the following sections.

2.2 SEGMENTATION APPLICATIONS ALIGNING WITH THE COMPANY'S STRATEGIC VISION

Aligning with the overall strategic direction of a company is critical to developing an effective segmentation strategy. For this study, we focused on cost, sustainability and service level metrics as the main drivers of the company's supply chain strategy. Similar to the company's strategy, Hashmi & Zhang (2016) identified Key Performance Indicators (KPIs), which they labeled as Flexibility, Efficiency, and Sustainability. Using Analytical Hierarchy Process (AHP) and Fuzzy Sets, they concluded that the most important areas that companies should focus on to improve these KPIs are sourcing and customer relationship management (Hashmi & Zhang, 2016). To dive into these strategic objectives, this section details segmentation applications that improve cost, CO_2 emissions, and service level.

2.2.1 COST

Cost is one of the most important criteria when developing a segmentation strategy. This is because segmentation enables cost efficiency through economies of scale and improvement of inventory holding and transportation processes, which was demonstrated by numerous studies. For example, Svoboda & Minner (2021) showed how segmentation in inventory classification and replenishment policies are deployed to minimize total inventory cost. Similarly, Kharlamov et al. (2020) applied a segmentation approach in the fast-moving consumer goods (FMCG) sector and achieved cost reduction by reducing demand variability. In another case study, segmentation of sporting goods reduced the costs of inventory handling and transportation (Roscoe & Baker, 2014). In a real-life implementation of segmentation, Gardena decreased logistics costs by 5%, and Philips was able to drive down the cost of its supply chain by 50% (Protopappa-Sieke & Thonemann, 2017). These studies and examples show that segmentation is a promising approach to reduce overall supply chain costs.

2.2.2 CO_2 EMISSIONS

Supply chain operations contribute to a significant portion of carbon emissions and waste creation (Bové & Swartz, 2016). With this motivation, one of the top priorities for the company is to reduce CO_2 emissions in its supply chain. Several studies have demonstrated that the network and distribution designs have a substantial impact on CO_2 emissions (Musavi & Bozorgi-Amiri, 2017; Paciarotti & Torregiani, 2021; Validi et al., 2014). Kuruvilla et al. (2012) surveyed the trend of supply chain sustainability in the United States and concluded that supply chain CO_2 emissions can be reduced through product design, waste reduction, and collaboration with vendors and service providers.

Food cooling serves as a primary factor that impacts carbon emissions at companies offering temperature-controlled food products. The challenge is that lowering cooling energy consumption may lead to food-quality degradation. Therefore, Zanoni & Zavanella (2012) developed a mathematical model to optimize the trade-off between food quality and energy consumption linked to food cooling. Their model can be replicated for different supply chain case studies to enable informed decisions given varying temperatures and batch sizes.

2.2.3 SERVICE LEVEL

Typically, in food and beverage retail, there are no backorders, since a customer can find many alternative food providers. Hence, a high service level is required to satisfy customer demand. However, high volatility in customer demands makes it very challenging to meet service level goals. To meet the service level goals, industries operating in uncertain markets need to keep a high level of safety stock (Lovell et al., 2005). As segmentation can be used to reduce demand variability (Kharlamov et al., 2020), it can also aid in improving service levels and reducing safety stocks. Additionally, service level requirements can be distinct in different products based on their relevance to the business and for unique types of customers based on their behavior (Protopappa-Sieke & Thonemann, 2017). This emphasizes the capability of segmentation strategies to improve service levels.

In Section 2.3, we present studies that focus on the food and beverage retail industry.

2.3 SEGMENTATION APPLICATIONS FOR THE FOOD & BEVERAGE RETAILER INDUSTRY

Retailers typically offer a great variety of POS and non-POS products. As the scope of this project is limited to food, beverages, packaged coffee, and paper and plastic kitchenware, we present a review of papers that studied these items.

2.3.1 SUPPLY CHAIN SEGMENTATION FOR FOOD & BEVERAGE PRODUCTS

Food products are particularly challenging to manage due to their unique characteristics, such as perishability and temperature control. The challenge amplifies in complex supply chains that require handling the products by different storage and transportation entities. Substandard handling by one entity can lead to food contamination due to delay or inadequate temperature control (Kumar, 2014). This leads to a significant loss in the form of food waste and even more severe consequences when contaminated food is consumed, resulting in serious illnesses and sometimes death (He et al., 2018). The spoilage is attributed to poor quality control and large inventories (Wang & Li, 2012).

Many studies have focused on different aspects of managing perishable foods, including inventory handling in storage facilities, transportation planning, marketing strategies, and product traceability (He et al., 2018). One proposed strategy is to develop a pricing policy based on the perishability of the product with the goal of maximizing profitability given the risk of perishability (Wang & Li, 2012). Another approach is to focus on network design. Khamsi & Stolear (2016) investigated several network designs for fresh food supply chains to evaluate cost and speed tradeoffs. Based on the different scenarios and network designs studied, they proposed a network design where the distribution centers are at the same location as the suppliers, which reduces cross-docking costs and enables consolidation with other types of goods (Khamsi & Stolear, 2016).

Food retail supply chains can widely benefit from segmentation using market characteristics. Some examples are wealth, age composition, lifestyle, and health awareness (Bahn & Granzin, 1985). Grunert (2018) reviewed how international food supply chains are commonly segmented based on markets and identified the advantages and challenges of different approaches. Grunert used the following approaches to food segmentation: product benefits, means-end market analysis, and consumer lifestyle. In general, Grunert's study revealed low utilization of market segmentation in food supply chains. In a focused study, Theodoras (2009) applied cluster analysis to segment food retailer markets based on 10 customer service elements. Using these elements, markets were segmented into three clusters: "Demanding but Satisfied," "Demanding and Unsatisfied," and "Undemanding and Unimpressed," and different customer strategies were suggested for each cluster (Theodoras, 2009).

2.3.2 SUPPLY CHAIN SEGMENTATION FOR PAPER & PLASTIC KITCHENWARE

Disposable kitchenware products are currently made with poly-coated paper or plastic (Czaika, 2010). One of the major challenges in managing the supply chain for paper and plastic kitchenware is to design a sustainable supply chain. Since this challenge is cross-functional, Czaika (2010) used "Facilitated Systems Thinking" to design cup supply chains. "Facilitated Systems Thinking" is an approach that aligns stakeholders in designing systems relevant to them. The goal of Czaika's effort was to design a circular chain for their cups, where consumers are also involved in this chain. This entails the need to design recyclable cups, develop packaging standards, work with government regulators, and improve consumer awareness to increase their contribution to recycling (Czaika, 2010). As a result, a system was designed for paper and plastic kitchenware recycling, which enabled the stakeholders to identify several supply chain challenges and opportunities and take them one step further toward eliminating waste.

Although paper and plastic kitchenware products are non-perishable, their production and delivery policies still need to be optimized due to capacity constraints (Hu et al., 2008). Accordingly, an effective supply chain strategy for these products is required.

2.3.3 SUPPLY CHAIN SEGMENTATION FOR PACKAGED COFFEE

Coffee supply chains can be complex. They can have up to 12 entities from harvesting to the coffee to be consumed (Lopez & Chaudhry, 2020). Additionally, the complexity of the supply chain depends on the desired quality of the final product. Although coffee is not highly perishable, its quality degrades over time (Bladyka, 2013). Since the company targets specialized high-quality coffee crops, these challenges must be considered. Thus, a special supply chain design is required to ensure only premium coffee beans are delivered.

Sourcing and procuring high-quality coffee require special attention. Qualities and flavors differ significantly based on geography, farming practices, and post-harvesting processing (Toledo et al., 2016). While sourcing or procuring, these should be studied in comparison to the price, lead time, availability, and order quantities (Bernad & Romero, 2016). Moreover, the high-quality coffee must not be mixed with lower-quality crops at any point in the chain. Weaver (2006) studied a similar challenge involving the segregation of specialized soybean supply chains. The study proposed the separation of specialized soybeans material flow, that is linked with an accurate and updated information flow. Moreover, Weaver suggested the use of single contracts focused on the specialized crops throughout the supply chain.

In addition to the quality of coffee, on-shelf availability is critical. Khakdaman (2014) studied this problem for a coffee retailer based in Beijing. The distribution channel for this retailer was as follows: Store to Consolidated Distribution Center (CDC) to Regional Distribution Center (RDC) to store (Khakdaman, 2014). Khakdaman simulated four different alternatives that have different delivery sequences and scheduling, with the goal of finding the highest on-shelf availability. He concluded that the best alternative is to use both CDCs and RDCs. Specifically, suppliers send coffee only to CDCs, then CDCs can send it to RDCs or directly to stores, and RDCs send coffee to stores. This alternative achieved the highest on-shelf availability, but increased the total cost significantly.

To overcome these complexities, Lopez & Chaudhry (2020) used a cost-minimization model which led to the reduction of transportation and production costs. They recommended processing roasting on the farm, co-locating the milling process with roasting, distributing coffee directly to retailers, and utilizing economies of scale (Lopez & Chaudhry, 2020).

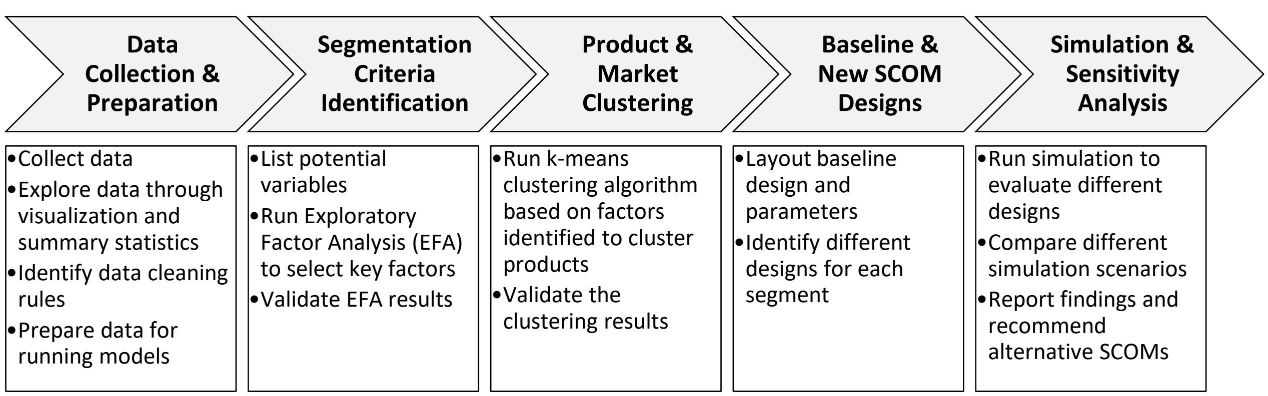
As shown in this literature review, various papers presented different approaches to designing or evaluating segmentation strategies. They separately used segmentation, simulation, and optimization to independently manage food, beverage, packaged coffee, and paper and plastic kitchenware. To further contribute to the literature, this study combines these efforts by offering a new data-driven approach that integrates exploratory data and factor analysis, unsupervised machine learning (k-means clustering), and simulation to design segmented SCOMs. Exploratory factor analysis and k-means algorithm are applied to identify product segments based on product characteristics. Then, simulation is utilized to design different SCOMs for each segment and evaluate different service level scenarios against cost, sustainability, and service level metrics. Based on simulation results, customized sourcing, forecasting, and inventory policies per segment are proposed to improve cost, sustainability, and service level metrics.

3 DATA AND METHODOLOGY

The main objective of this study is to identify distinguishable supply chain segments based on product and supply characteristics and design an alternative supply chain operating model (SCOM) for each segment to improve cost, service level, and sustainability metrics. To achieve this, a five-step, integrated, data-driven methodology is designed. First, data is gathered by working collaboratively with a team of company experts, and it is reviewed for accuracy and completeness. Second, exploratory data and factor analysis are conducted to identify potential segmentation criteria and select the most relevant factors for consideration. Then, k-means clustering is applied based on the selected factors to create the product segments. After that, a new SCOM is designed for each segment based on the characteristics of the products in each segment. Finally, the new SCOMs are simulated using discrete-event simulation where different adjustments to the service level targets for each product cluster are experimented to analyze the impact on inventory cost, transportation costs, and CO_2 emissions. A high-level description of this process is shown in Figure 1.

Figure 1

Data-driven Methodology Description



3.1 DATA COLLECTION AND PREPARATION

For a data-driven approach to work effectively, a versatile, large volume, and accurate dataset is essential. To obtain this set, recent operational data from the company's database is extracted. The extracted tables included the SKU list, items attributes, store list, DC list, and demand data, as listed in Table 2.

Table 2

Dataset Names and Descriptions

Dataset Name	Dataset Description	No. Variables
SKU List	• Includes the description and categories of SKUs in the segmentation scope • Consists of 459 SKUs across 5 product categories which are food, beverages, paper and plastic kitchenware, and packaged coffee	2
Item Attributes	• Includes the attributes of SKUs at the location level such as ABC classification, lead time, max/min order quantity, shelf life, pack/pallet size, temperature control requirements, make/buy, Make-to-Stock (MTS)/Make-to-Order (MTO), and perishability	29
Store List	• Includes the location, ownership, area (urban or rural), and RDC/CDC delivery frequency information for each store	19
DC List	• Includes the location information of RDCs and CDCs in scope	3
Demand	• Includes two years of weekly demand data per SKU per location at the RDC/CDC level	2

The data is thoroughly reviewed and prepared by finding missing or outlier entries, discussing these records with subject-matter experts to come up with data cleaning rules, and addressing data inaccuracies. To perform data cleansing, we use the following approach:

1- Identify and correct erroneous data entries.
2- Identify and correct outliers. For some outliers, we try further data extraction to correct the data points when correct data is found.
3- Fill in missing values. Expert judgment is the main resource to deal with missing values. For example, based on what makes more business sense, data can be filled using the median, most frequent, or a representative value from a similar product.

Each of the data tables in Table 2 is prepared for use in each step of this methodology as applicable. The following sections present each step in further detail.

3.2 SEGMENTATION CRITERIA IDENTIFICATION

Several initial variables are identified as potential segmentation criteria from the company's operational data. The initial variables are listed and defined in Table 3. These variables can be continuous (numeric variables that have infinite values between any two values), discrete (numeric variables that have finite values between any two values), or categorical (non-numeric variables).

Table 3

Initial List of Potential Segmentation Criteria

Variable	Definition	Type
	Product Segmentation	
Coefficient of variation of monthly demand per product	The coefficient of variation (COV) of demand in relevant periods (e.g., fall, spring, summer, winter) of product sales	Continuous
Average monthly demand per product	The demand in relevant periods (e.g., fall, spring, summer, winter) of product sales	Continuous
Shelf-life (days)	The period in days from production where a product can no longer be sold	Continuous
Lead time (days)	The average lead time days from ordering to receiving	Continuous
ABC classification	An existing classification by the company signifying the importance of a product to the operation (5 classifiers)	Categorical
Pack size (units/case)	The quantity/amount of the product in each pack	Discrete
Temperature control type	Ambient or refrigerated	Categorical
Item cost ($/unit)	The average unit cost of an item	Continuous
Seasonality	A classification that determines if a product is offered the whole year or only in special seasons	Categorical

Market Segmentation		
Average monthly demand per distribution center	The average demand for the distribution center (e.g., RDC and CDC)	Continuous
COV of monthly demand per distribution center	The coefficient of variation of demand for the distribution center (e.g., RDC and CDC)	Continuous
Urban vs. rural	A classifier based on the demographics of the location to indicate the customer behavior and preferences	Categorical
Store type	A classifier for stores based on seating and drive through availability	Categorical

Each product and market have a complex set of characteristics that explains its behavior and requirements. To have an effective and easy-to-manage segmentation strategy, segmentation should be based on a few related characteristics (Protopappa-Sieke & Thonemann, 2017). Therefore, we have employed Factor Analysis of Mixed Data (FAMD) to find how the different variables are related and hence select a few key variables as segmentation drivers. FAMD is a statistical approach that reduces the number of variables by transforming the variables into factors that contain the most information about the variables. To preserve the variability, the method finds a smaller number of linear functions of the original variables and treats these functions as new factors. In this way, the original variables in higher dimensional space can be represented by new factors in smaller dimensional space by retaining most of the variability. For this study, FAMD is particularly useful as it can be applied for a mix of categorical and numerical data whereas other dimension reduction methodologies like PCA are used for numerical data only. Considering data availability in the company's systems, only seven variables out of the ones listed in Table 3 had meaningful data. We used the following seven variables and applied FAMD to further reduce the dimensionality:

- Average monthly demand (per product)
- Coefficient of variation of monthly demand (per product)
- Shelf-life
- Seasonality
- Temperature control type
- Item cost
- Lead time

3.3 PRODUCT CLUSTERING

To segment the company's products, each product is assessed against the factors identified by FAMD. Following this, the products are clustered using the *k*-means algorithm, which aims to divide data points into *k* clusters in which each data point belongs to the cluster with the nearest mean (i.e., cluster centroid). *k*-means algorithm minimizes the total within-cluster sum of squared Euclidean distances between each data point and its corresponding cluster centroid (i.e., total within-cluster sum of squares).

The algorithm starts with identifying *k* random centroids. Then, it assigns each data point to a single centroid that has the smallest squared Euclidean distance to itself. After that, centroid locations are updated for each cluster by taking the mean of data points in the cluster. Assignment of data points to clusters and update of cluster centroids continues until there is no change in the total within-cluster sum of squares. The algorithm steps can be mathematically written as follows:

1. Initialize centroids $z^{(1)}, \dots, z^{(k)}$
2. Repeat until there is no further change in the total within-cluster sum of squares: $\sum_{j=1}^{k} \sum_{i \in C_j} ||x^{(i)} - z^{(j)}||^2$
    a. For each $j = 1, \dots, k$: $C_j = \{i \in \{1 \dots n\}\ s.t.\ x^{(i)}\ is\ closest\ to\ z^{(j)}\}$ where distance is defined by $||x^{(i)} - z^{(j)}||^2$
    b. For each $j = 1, \dots, k$: $z^{(j)} = \frac{1}{|C_j|} \sum_{i \in C_j} x^{(i)}$

where,

There are n data points and k cluster centroids

$z^{(j)}\ where\ j \in \{1, \dots, k\}$ represents cluster centroids

$x^{(i)}\ where\ i \in \{1, \dots, n\}$ represents data points

$C_j\ where\ j \in \{1, \dots, k\}$ represents the set of data points assigned to cluster j

For this study, each data point represents products where data point coordinates are shaped by products' characteristics. The *k*-means algorithm is used to identify the product segments. Different experiments are performed to find the ideal number of clusters by varying *k* values, calculating the total within-cluster sum of squares, and assessing the trade-off between decreased variability and increased number of clusters.

3.4 BASELINE AND ALTERNATIVE SCOM DESIGNS

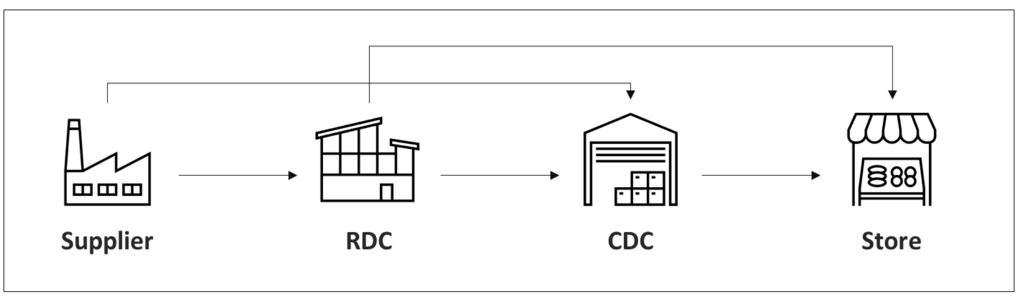
As explained in Section 3.3, identified product segments have unique attributes that require different SCOM strategies. With this motivation, alternative SCOM designs must be developed to minimize cost and CO_2 emissions and maximize service level, in comparison with the current situation.

To create new SCOMs, it is essential to develop a simplified baseline model representative of the operating model of the company. In this baseline model, the company's supply chain includes suppliers, RDCs, CDCs, and stores. Suppliers are responsible for providing all products. RDCs are used for storing and shipping ambient products. CDCs are closer to stores than RDCs and they are used for storing and shipping both ambient and frozen products. Finally, stores are responsible for meeting consumer demand.

The company has sourcing rules to identify the product flow between different locations in their network. RDCs are replenished from suppliers and CDCs are replenished from suppliers and/or RDCs depending on the sourcing rules. Stores can be served by RDCs and/or CDCs based on sourcing rules. This baseline SCOM design including the distribution channel and product flow is illustrated in Figure 2.

Figure 2

Representative Baseline Supply Chain Operating Model for the Company



Based on the characteristics of product segments identified by FAMD and k-means clustering, new SCOM designs for each segment are developed as an alternative to the baseline model. For the new SCOM of each product segment, three strategies are considered. The strategies are sourcing policies (where should the RDCs, CDCs, and stores get their demand from), forecasting policy (the level of forecast aggregation), and the inventory policy (reorder point and order quantity). These strategies are used as inputs in the simulation model as explained in Section 3.5.

3.5 SIMULATION AND SENSITIVITY ANALYSIS

To evaluate the possible impact of new SCOM designs on the performance of each segment, three scenarios for different service level targets are simulated. The simulation model includes all RDC, and CDC locations in North America. As there is no supplier data available, the model represents suppliers as a single entity. Similarly, since there is no information about the store demand nor about precise store locations, thousands of stores are aggregated at the CDC level (i.e., there is only one store entity per CDC).

The simulation model represents the company's product flow from suppliers to stores. Whenever an order from a store is issued, items are pulled from the corresponding DC based on the company's sourcing rules, which leads to inventory reduction in the DC. Then, the inventory in that DC is replenished from the other DCs or the supplier based on sourcing rules and inventory policy, which depends on the review cycle and the target service level. The model then measures the unfulfilled store orders, the number of expired perishable items in DC, average fulfillment lead time, distance traveled, and average inventory holding days, which are all factored into the calculations for cost, emissions, and service level. Table 4 defines the simulation setup including policies and assumptions in detail and Table 5 defines the output metrics.

Table 4

Simulation Setup

Simulation Policies	**Description**	**Inputs**	**Assumptions***
Store Orders	It includes the weekly store demand per product per store.	Weekly demand per product per store	• The company provided the historical demand for RDCs and CDCs from 2020 to 2021. This demand is then translated into the pooled stores demand.

Inventory Policy	It includes the inventory policy per product per location. Inventory at each location is replenished based on inventory policies.	Reorder point, safety stock, order quantity, inventory holding cost	• For each product and location, inventory is replenished in the amount of a fixed order quantity when it reaches reorder point. • The order quantity is equal to the average weekly demand per product per location. • The safety stock is equal to the service level target multiplied by the standard deviation of weekly demand per product per location. • The reorder point is equal to the safety stock plus the demand over lead time and review period per product per location. • The review period is 1 week for all products across all locations. • There is no storage at supplier locations. • Inventory cost is equal to $15/pallet/month for all locations.
Sourcing Policy	It includes the rules for product flow between each location per product. Products can only flow between defined locations.	Sourcing rules, lead time	• Sourcing rules between CDCs & stores and RDCs & CDCs for each product are provided by the company. • The lead time between supplier & CDCs and supplier & RDCs for each product is given by the company. • RDCs serve all the customers that their CDCs serve. • For RDCs, all product replenishments come from suppliers. • For CDCs, all product replenishments come from either suppliers or RDCs. When replenishment is necessary, this is selected each time randomly with equal probability. • The lead time between RDCs and CDCs is considered five days. • The lead time between RDCs/CDCs and stores is considered one day.
Transportation Policy	It includes the transportation flow information between each location.	Transportation lead time, truck capacity, CO_2 emissions, transportation cost per mile, distance between each location	• The location of RDCs and CDCs are provided by the company. • Total distance between CDC and pooled stores depends on the dispersion of the actual number of stores in that location. Based on the actual number of stores, a weight is calculated and multiplied by the estimated travel distance for the CDC serving the least number of actual stores (please see Appendix A for more details). • The distance from the RDC to the store is equal to the distance from the RDC to the CDC plus

	the distance traveled from the CDC to its stores. • Any distance and transportation cost that occurs from the supplier to RDCs/CDCs/stores is neglected. • Each truck can carry 28 pallets. • Transportation cost per mile is $3 (regardless of the products). • There is no constraint on the number of trucks. • CO_2 emissions for trucks is 87g/tm.

* All assumptions are representative and not actuals.

Table 5

Output Metrics for the Simulation and Sensitivity Analysis

Output Metrics	**Definition**
Cost	$ value that includes costs of delivery and inventory
Emissions	The weight of CO_2 emitted by the trucks, which depends on the distance traveled
Service-level	% of store orders fulfilled

We first simulate the baseline model explained in Section 3.4, where no segmentation is performed. After that, different scenarios are simulated and analyzed for the different clusters. Specifically, we experiment with different service levels for each cluster, which provides an understanding of the trade-off between fulfillment and logistics costs. This scenario analysis aims to explore opportunities for improvements to the segmentation strategy. The results for each step of the methodology are presented in Section 4.

4. RESULTS

The analysis led to a comprehensive product segmentation strategy. The strategy consists of identifying key product characteristics for segmentation, clustering products based on selected characteristics, providing different sourcing, inventory, and forecasting policies for each product cluster, and analyzing the potential improvements in cost, service level, and sustainability metrics.

4.1 DATA PREPARATION

To ensure we obtain meaningful results, data must be clean, complete, and accurate. Therefore, a rigorous effort was exerted to fix erroneous entries, correct outliers, and fill missing data.

4.1.1 FIXING ERRORS

We modified several data entries to correct errors and remove irregularities. The following fixes were implemented:

1- In different observations, each item had many different ABC classifications. We used the most frequent ABC classification across all observations for each product.

2- There were multiple similar product categories. To reduce the complexity of analysis, we combined several product categories that are all managed as kitchenware.

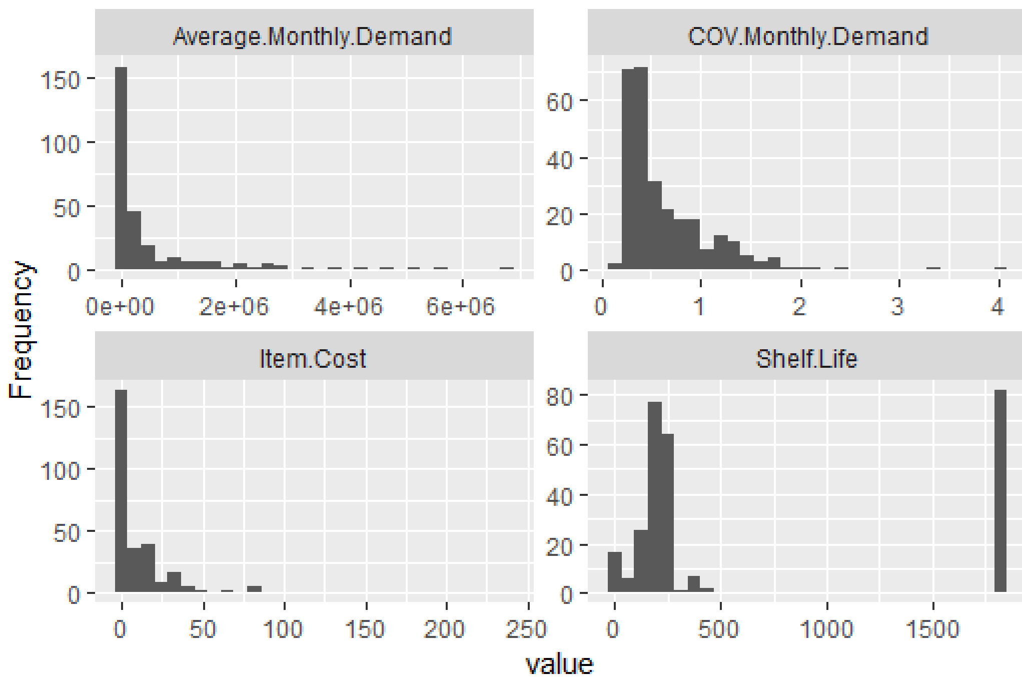
3- Several paper and plastic SKUs have the same features and can be used interchangeably. Thus, we grouped similar items as a single SKU. This reduced the number of SKUs from 459 to 238 for the purpose of product segmentation.

4.1.2 HANDLING OUTLIERS

We analyzed data distributions to detect outliers in each item attribute. The distributions for each attribute are shown in Figure 3.

Figure 3

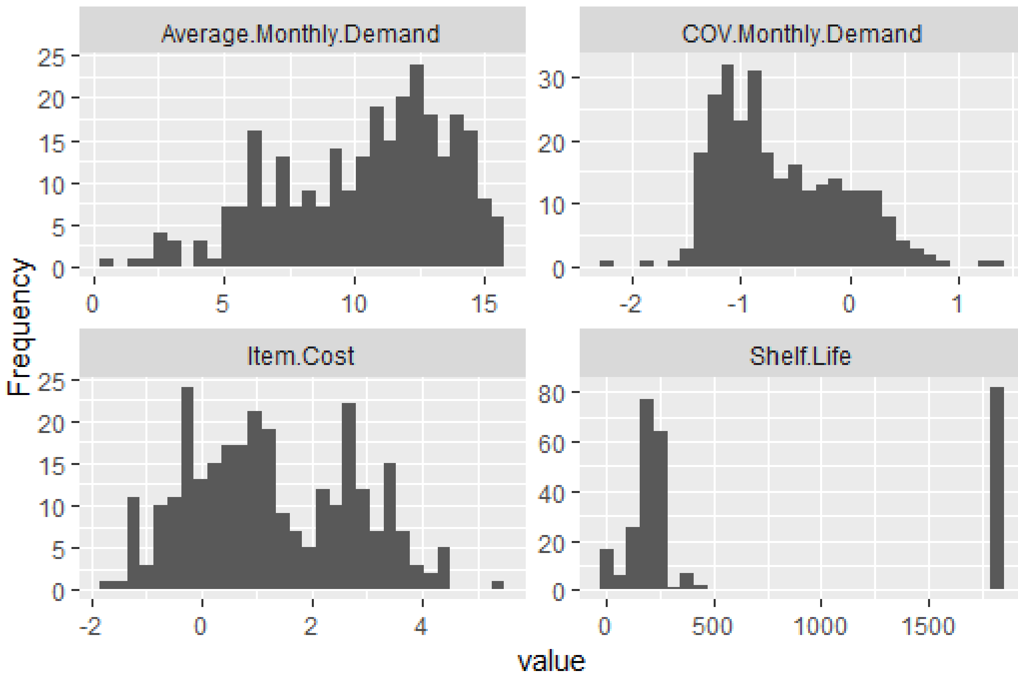
Distribution of Numerical Item Attributes



To handle the large number of outliers, we applied logarithmic transformation to average monthly demand, coefficient of variation of monthly demand, and item cost variables. This resulted in reduced variability and skewness in data as shown in Figure 4.

Figure 4

Distribution of Numerical Attributes after Log Transformation



4.1.3 HANDLING MISSING DATA

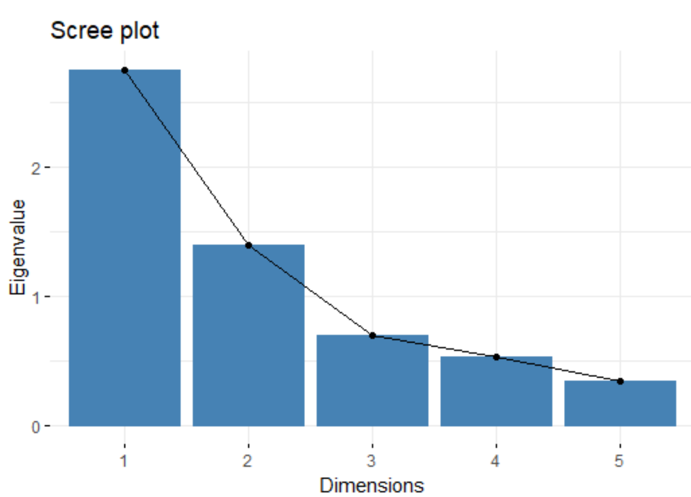
After exploring the data, we found out that some SKUs did not have demand data. Focusing on this, we filled the missing values for the average and coefficient of variation of monthly demand. For both, we filled the missing data with the median of items within the same category (e.g., a coffee bag demand would be the median average monthly demand and median coefficient of variation of monthly demand for the packaged coffee category). This is because items from the same category generally have similar demand patterns, as discussed with company experts.

4.2 SEGMENTATION FACTORS

After cleaning and preparing data for seven variables listed in Section 3.2, we used FAMD to reduce the variable dimensionality. A key input for FAMD is to provide a number of dimensions to be used in the model. Scree plots are commonly used to select the number of dimensions to be used for FAMD. Scree plots show the relative impact of adding a new factor by comparing the eigenvalues (a measure of the amount of variability explained by a factor) against the number of factors. As our goal is to have a smaller number of factors that explain most of the variability in data, the Kaiser rule suggests that we should pick the factors that have eigenvalues more than 1.0. This is because an eigenvalue of 1.0 means that the factor has information equal to a single vector (Kassambara, 2017). Figure 5 clearly shows that only two factors have an eigenvalue of more than 1.0. Therefore, we can capture most of the variability by using two factors. We proceeded with using two factors to represent all original variables, resulting in total explained variability of 69.11%.

Figure 5

Scree Plot of Eigenvalues with Respect to the Number of Components from FAMD Analysis



In addition, we analyzed the squared loadings of each variable for a given factor. The squared loadings provide a measure of the proportion of the variance in the original variables captured by a given factor. If a variable has a high squared loading value for a factor, it indicates that the factor better explains the variability of that variable. Based on our analysis, we found out that lead time variable has low squared loadings for both factors (a commonly used cut-off score for squared loadings is 0.4). As none of these factors can explain the variability in those variables, we removed them from the model.

Table 6 shows the squared loadings of each variable for a given factor. In Table 6, squared loadings are bolded based on their magnitude to indicate factor representation of each variable. This shows that Factor 1 represents average monthly demand, shelf-life, and item cost variables. Factor 2 represents the coefficient of monthly demand and seasonality variables.

Table 6

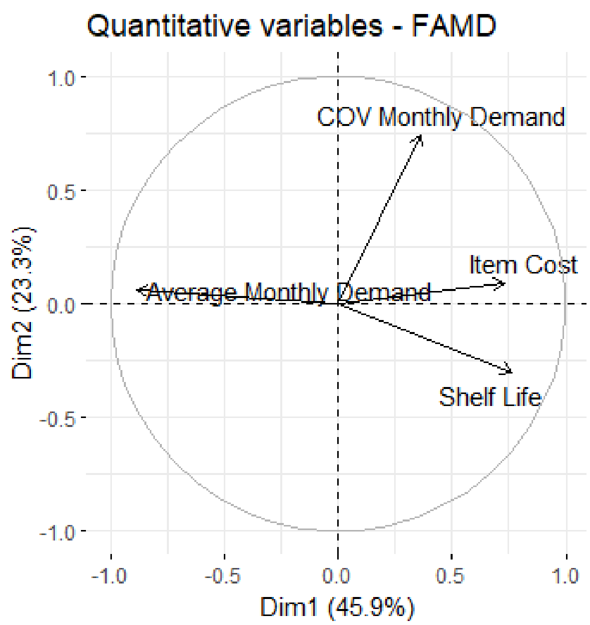
Squared Loadings Between the Original Variables and the Factors

Variables	Factor 1	Factor 2
Average Monthly Demand	**0.782369**	0.000577
COV of Monthly Demand	0.085488	**0.596994**
Shelf Life	**0.625559**	0.053583
Item Cost	**0.514290**	0.025528
Seasonality	0.033456	**0.715124**
Temperature Control Type	**0.698728**	0.015199

The correlations between variables and factors are shown in Figure 6. While squared loadings show the importance of each variable to each factor, correlations plot helps to see the positive or negative relationship between variables and factors. As we can see, average monthly demand is negatively correlated with Factor 1 whereas item cost and shelf-life are positively correlated with Factor 1. This is an expected result, because the company has a higher demand for products that have low shelf-life and item cost whereas lower demand for products that have high shelf-life and item cost. On the other hand, we see that the coefficient of monthly demand and seasonality are positively correlated with Factor 2. Based on squared loadings and correlation analysis, we concluded that Factor 1 represents how fast the product moves and Factor 2 represents how complex it is to manage the products. Therefore, we named Factor 1 as Product Speed and Factor 2 as Demand Complexity.

Figure 6

Correlations Between Factors and Variables

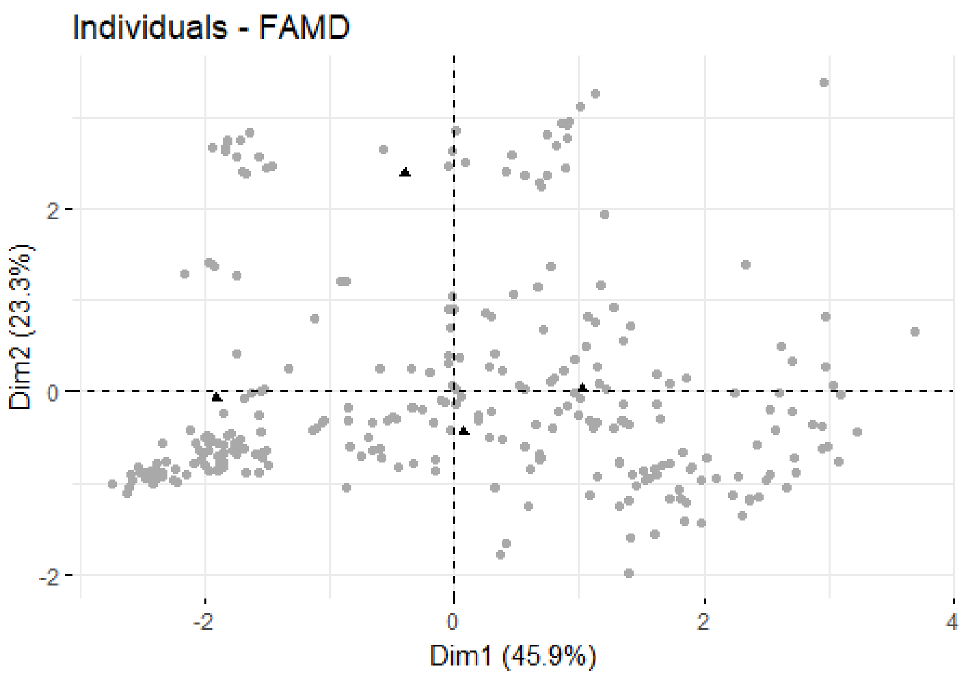


As FAMD reduced the dimensionality to two factors representing the original five variables, we can plot the distribution of products with respect to two factors. Figure 7 shows the distribution of products based on two factors, namely, Product Speed and Demand Complexity. In Figure 7, the x-axis represents Product Speed, the y-axis represents Demand Complexity, and each point represents a product. The products that have similar characteristics are located close to each other. The characteristics of each product in the scatter plot can be interpreted by their position in the graph. For example, products in the bottom left quadrant have high Product Speed and low Demand Complexity. As average monthly demand is strongly negatively correlated with Product Speed whereas item cost and shelf-life are strongly positively correlated per Figure 6, we can infer that those products have high demand, low cost, and low shelf-life. On the other hand, as the coefficient of monthly demand and seasonality are positively correlated with

Demand Complexity per Figure 6, we can infer that those products also have low demand volatility and low seasonality. These interpretations inferred from Figures 6 and 7 are critical for understanding k-means clustering results and supply chain design recommendations.

Figure 7

Representation of Products with Respect to Two Factors from FAMD Analysis



4.3 PRODUCT CLUSTERS

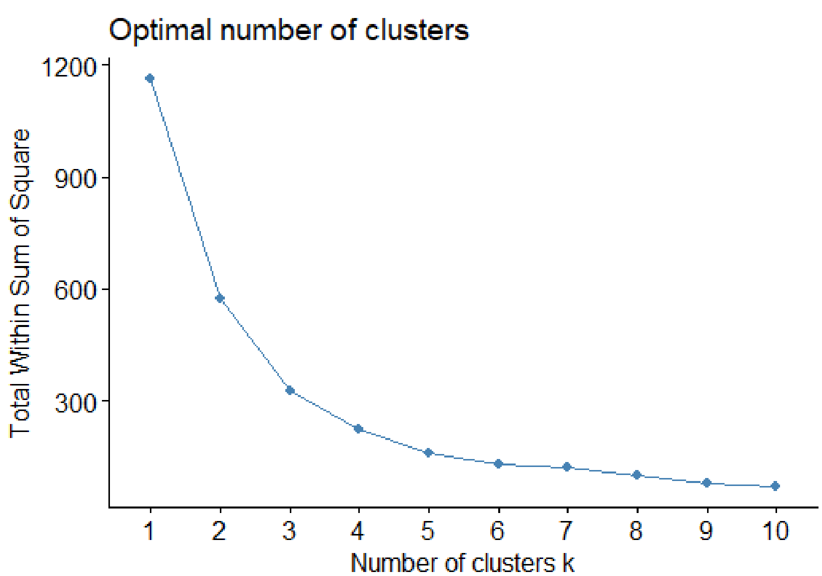
To cluster the products based on Product Speed and Demand Complexity factors identified by FAMD, *k*-means clustering was performed. A key input for the k-means algorithm is to provide the number of clusters (i.e., *k* value) to be used in the model. One way to identify the number of clusters is to look at the change in the total within-cluster sum of squares as the k-value changes. It is calculated by measuring the Euclidean distance between each data point and its centroid and summing these squares across each cluster. We aim to achieve a small total within-cluster sum of squares. We also want to obtain a small number of clusters to reduce the complexity associated with creating and managing multiple SCOMs.

However, the total within-cluster sum of squares is inversely proportional to the number of clusters. Therefore, we need to balance the trade-off between decreasing the total within-cluster sum of squares and increasing the number of clusters.

Figure 8 shows the total within-cluster sum of squares for each *k* value (number of clusters). Based on this graph, we can use the elbow method to determine the number of clusters. The elbow method suggests that we should pick the number of clusters according to the inflection point. This is the point where marginal reduction in the total within-cluster sum of squares decreases as we increase the number of clusters. From Figure 8, we can conclude that the inflection point occurs when the number of clusters is chosen as *k*=3.

Figure 8

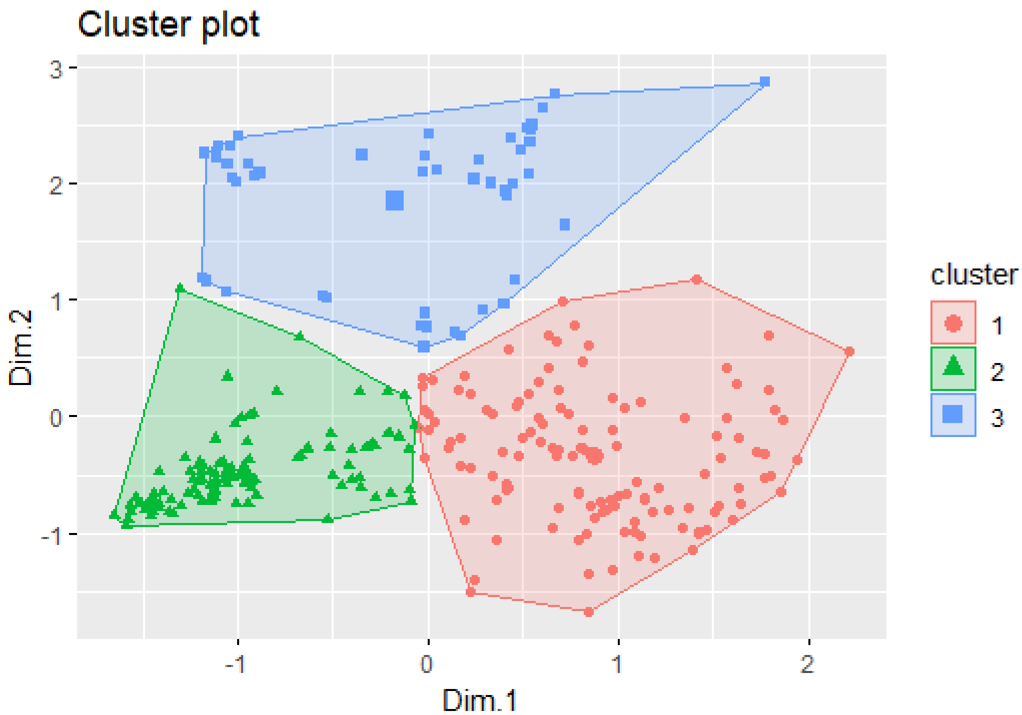
Total Within-cluster Sum of Squares Error for Different Numbers of Clusters (k)



Using *k*=3 clusters, the *k*-means algorithm generates the clustering shown in Figure 9. As shown in the graph, each point is assigned to one of three clusters (i.e., product segments) based on their distance to cluster centroids.

Figure 9

k-means Clustering Results



All three product segments have unique characteristics. These different characteristics require a unique approach to managing each segment. These approaches are detailed in Section 4.4.

4.4 NEW SCOM DESIGNS FOR EACH SEGMENT

Section 4.3 shows three unique product segments identified by k-means clustering. The characteristics and recommended supply chain strategies for each product segment are described below:

- **Slow-moving items segment (red cluster):** Consists of low-volume, medium-volatility, high-cost, and high-shelf-life items. Slow-moving items stay in the DCs for a long time due to low demand. For these items, just-in-time inventory is recommended, where the reorder point is managed from the store inventory and the materials are sent directly from suppliers to the stores. For some high-value items, Vendor Managed Inventory (VMI) can be considered to ensure their availability.

Similarly, forecasting should be done at the store level, and orders are made based on the store's forecast.

- **Fast-moving items segment (green cluster):** Consists of high-volume, low-volatility, low-cost, low shelf-life items. For these items, the inventory to fulfill store demand should be stored in the CDCs to be close to the customers. Also, since the variability of demand is low for this segment, detailed forecasts at the CDC level are recommended. The company can further consider using micro-fulfillment centers that are even closer to the stores to store and fulfill these items, which will help reduce lead time and improve the service level of these items.
- **Complex items segment (blue cluster):** Consists of high-volatility and seasonal items. To reduce inventory risks, the inventory of these items should be pooled in the RDCs or stored at supplier locations for high-cost items. For seasonal items, the inventory can be moved to the CDC in their respective seasons and be managed as a fast-moving item in that season. Furthermore, due to high variability, forecasting can be improved by aggregating the demand forecasts at the RDC level and forecasting for monthly demand instead of weekly demand, while keeping the same forecasting horizon.

The segments, their characteristics and recommended management policy are summarized in Table 7.

Table 7

Segmentation Policies

Cluster	Inventory Policy	Forecasting Policy
Slow-moving items segment	• Do not hold inventory at DCs • Make/ procure to order • Consider VMI	Forecast at the store level
Fast-moving items segment	• Hold inventory in CDCs close to the stores • Consider Micro-Fulfillment centers	Detailed disaggregate forecasts at the CDC level
Complex items segment	• Pool inventory in RDCs • Consider VMI for high-cost products	Aggregate demand forecasts at the RDC level

For each segment, additional strategies were tested using simulation based on their characteristics and their recommended management policy. The simulation results will be presented in Section 4.4.

4.5 SCENARIO ANALYSIS AND SIMULATION RESULTS

To explore additional strategies that can improve the segmented SCOMs, we use Coupa's Supply Chain Guru software to analyze how modifying the service-level targets (which directly affect the level of safety stock requirements) for each segment impacts the inventory cost, transportation cost, and fill rate. For this experiment, we ran the simulation of the baseline model. In this scenario, the service level targets are assigned based on the item's ABC classification (provided by the company), where A items have a target service level of 99.5% and other items have a target service level of 98.5%.

Next, we investigate assigning different service levels for slow-moving, fast-moving, and complex items. Since complex products have high variability, they require a high safety stock to achieve a high service level. This leads to a high inventory cost and a high risk of surplus. Thus, we are assigning a high service level for the fast-moving products, a medium service level for slow-moving products, and a low service level for the complex products. The resulted performance is summarized in Table 8. The results show that reducing the target service levels has a significant impact on the carrying cost, but a low impact on the transportation cost and fill rate.

Table 8

Service Level Scenarios

	Service Level Targets			**Cost (In Millions)***			**Fill Rate**
	Slow-Moving	Fast-Moving	Complex	Total Transportation Cost	Total Carrying Cost	Total Cost	Avg. Fill Rate
No Segmentation	-	-	-	$52.2	$61.4	$113.6	96.39%
Scenario 1	98.5%	99.5%	95.0%	$52.4	$58.1	$110.5	96.36%
Scenario 2	97.5%	99.5%	92.5%	$52.5	$56.8	$109.3	96.36%
Scenario 3	97.5%	97.5%	92.5%	$51.9	$53.1	$105.0	96.26%

* Costs are representative and not actuals.

Although the products are grouped based on their statistical correlation, looking into the products in each segment in detail can help ensure the product segments are logical for the business. The breakdown of the item categories in each segment is shown in Table 9. The summary statistics for the item categories per cluster are shown in Appendix B.

Table 9

Product Categories per Segment

Product	SKU Count
Cluster 1 (Slow-Moving)	
Beverage Component	16
Packaged Coffee	25
Paper/Plastic Kitchenware	84
Cluster 2 (Fast-Moving)	
Beverages	22
Packaged Coffee	3
Savory Food	48
Sweet Food	32
Cluster 3 (Complex)	
Beverages	15
Packaged Coffee	16
Savory Food	3
Paper/Plastic Kitchenware	1
Sweet Food	15
Total	280

As shown in Table 9, three savory food items are assigned to the complex segment. When investigating these products, we found that they are fresh food items with a shelf-life of 2 days. Hence, they cannot be pooled and stored in the RDCs. Furthermore, they are seasonal items that have high consistent demand in their respective seasons. Therefore, we propose that they should be managed as fast-moving items in their seasons.

Additionally, we found that around 50% of complex items are classified as important items (A or AA items in the ABC classification), so it may be reasonable to consider storing these items closer to all stores despite their volatility. In that case, the items can be manually modified to be managed using another segment's operating model. This requires an analysis of the trade-off between margin loss and inventory cost, which was not a part of this project’s scope.

5 DISCUSSIONS

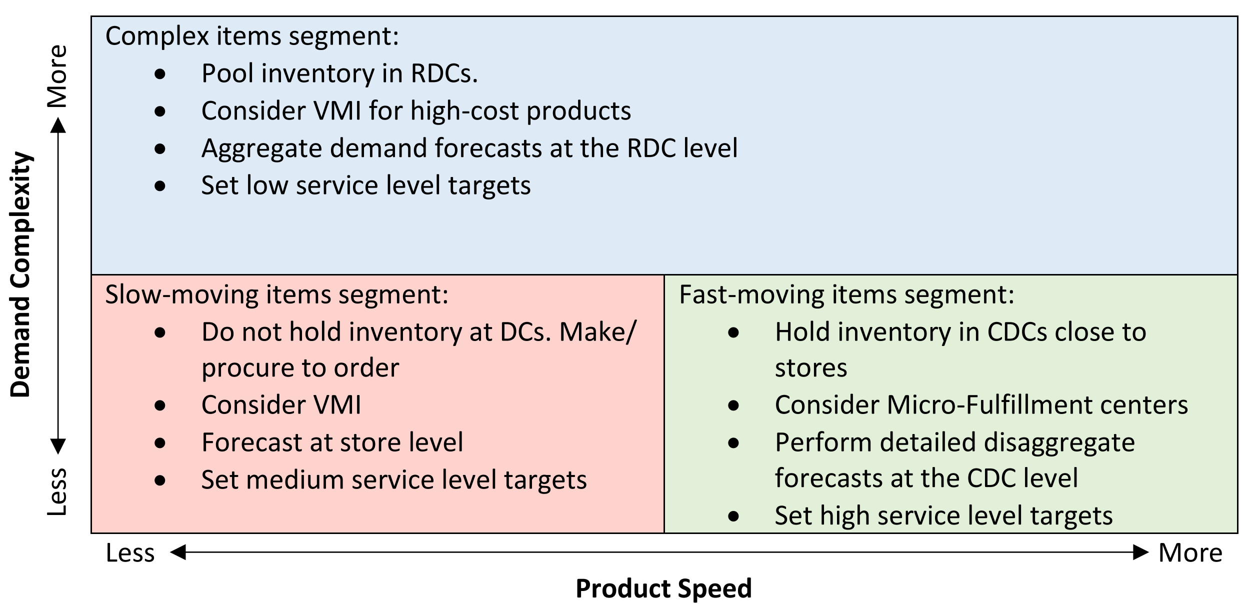
As the results in Section 4 reveal, several unique inventory, forecasting, and distribution policies can be applied to each segment to enhance overall performance in the supply chain. There are trade-offs between cost and service level, but these tradeoffs are different for each segment. Applying the policies that are fit for each segment can improve total supply chain cost, service level, and sustainability metrics.

5.1 MANAGERIAL INSIGHTS

In Sections 4.4 and 4.5, we provided independent recommendations that can be obtained from the clustering analysis and scenario analysis, respectively. To put these results together, Figure 10 depicts a matrix that lists the segments and the recommended model for each segment.

Figure 10

Final Segment Policies



The SCOM for each operating model brings benefits to the overall performance. First, slow-moving products are not delivered frequently, so eliminating their inventories in the DCs is expected to reduce the inventory holding cost without significantly increasing the transportation cost, and without significantly decreasing service levels. Second, disaggregating the inventory in the CDCs for fast-moving items is expected to improve service levels for these items. While this can increase the inventory holding cost, this increase is not expected to be high, as the fast-moving nature of these items means that they are not stored for long, and their low volatility means a reduced risk of surplus stock. Third, aggregating the demand for complex items is expected to reduce the risks of stockout and excess inventory.

To successfully implement this strategy, we recommend reviewing the items in each cluster to ensure they can be managed with the designated segment's design given the current capabilities. In case the current capabilities do not allow some items to be managed as recommended, the company should consider either investing in adding the capability or moving the item to another segment. For example, we found some sweet food items that are assigned to the complex segment. These items require frozen storage. Although the current RDCs may not be capable of frozen storage, the sponsoring company can consider investing in RDC freezers to cover complex frozen food items. Alternatively, they can consider off-site frozen storage locations. At this time, these items can be managed as slow-moving or fast-moving products, depending on the demand volume of each one.

Going forward, applying the segmentation strategy can have a positive impact on the whole supply chain. First, product types in each DC type will be reduced and more focused. This enables DC managers to organize and optimize the storage in those facilities. It will also facilitate orders from stores by reducing the variability of sources. Second, additional supply chain strategies can be applied uniquely for each segment, such as building micro-fulfillment centers for fast-moving products and signing VMI contracts for slow-moving items. These strategies will further improve key metrics including cost and service level. Finally, applying the segmentation methodology in different markets will allow the company to decide on which products to introduce to which new markets and have a targeted growth strategy.

5.2 LIMITATIONS

While this study yields numerous insights on how segmentation improves supply chain operations, it has some limitations that can be further examined. Specifically, this study performs segmentation based only on products and their limited set of characteristics.

5.2.1 SEGMENTATION CRITERIA

This study draws insights from product characteristics alone. The supply chain can be segmented using other criteria as well. For example, market-driven segmentation can focus on the customers to include factors such as customer types, customer priorities, customer demand patterns, and customer income. Another aspect of market-driven segmentation is the city characteristics. For example, the city can be distinguished by its size, density, average income, modernization, and urbanization.

Another criterion to look at is store-driven segmentation. Supply chain operations can be differentiated for different stores using store demand, size, type, and ownership model. To improve customer experience in the stores, each store segment can have different products served based on its characteristics.

5.2.2 PRODUCT ATTRIBUTES

Based on data availability, this study segmented products based on their demand volatility, demand volume, shelf-life, seasonality, temperature control type, item cost, and lead time. Other important product attributes include product margin, product use (e.g., point of sales, packaging, ingredient, or long-term equipment), and product size, which can lead to additional policies for each product segment. For example, high-margin products may be pushed close to stores despite its volatility, due to the high cost of missing a sale of those items. Also, product use and size can provide insight into how frequently products should be delivered.

5.2.3 SIMULATION ASSUMPTIONS

The simulation model in this study had several assumptions that simplify the real process of the company, as explained in Section 3.4. For example, it assumes that the size, transportation costs, and inventory holding costs are the same for all products. It also does not consider capacity constraints. Inventory policies may be impacted based on the capacities of the suppliers, RDCs, CDCs, or stores. Moreover, transportation policies and frequencies may change given the limited number of trucks and allowed truck schedules. Finally, supply constraints can also be a factor in determining ordering policies and frequencies.

6 CONCLUSION

Building upon studies utilizing segmentation strategies to improve supply chain performance, this paper provides a data-driven methodology that integrates machine learning and simulation techniques to apply supply chain segmentation. The methodology was tested on a portion of the global retailer's products, containing food, beverages, packaged coffee, and paper and plastic kitchenware. The segments were created depending on their demand volume, demand volatility, item cost, lead-time, shelf-life, temperature control type, and seasonality.

Three segments were identified based on the characteristics of the products: slow-moving, fast-moving, and complex products. For each segment, a unique SCOM was designed. Slow-moving products' inventory should be ordered by the store directly from suppliers. Fast-moving products should be forecasted and managed by CDCs or micro-fulfillment centers. Finally, complex products should be pooled in RDCs to reduce inventory risk. The facilities to handle each segment should have the required capabilities and equipment to be able to implement these strategies.

The SCOMs for each segment were simulated using discrete-event simulation in Supply Chain Guru. Based on historical demand and lead-time, the model measures each SCOM performance using inventory cost, transportation cost, and service level. For each segment, three different service level targets were experimented. As a result, it was recommended to set the target service level to 99.5% for fast-moving items, 97.5% for slow-moving items, and 92.5% for complex items. This reduces the total cost without a significant impact on the fill rate.

6.1 FUTURE RESEARCH

This study focused on segmentation based on seven factors of product and supply characteristics. In future research, additional factors and designs can be explored. For example, retailers can design segmented models based on the type of store service, store ownership, and store size. Also, market segmentation can be performed using market characteristics of different locations such as average income, location demand pattern, and urban vs. rural locations. Analyzing these factors can help make new design decisions that enable enhancements in managing different stores and markets.

Future studies can also test additional scenarios in the simulation. For example, they can test the impact of improving the forecast by reducing the standard deviation in the demand distributions. This analysis can help compare the cost of investing in an advanced forecasting tool with the benefits of reducing forecasting error. Additionally, the simulation can test different distribution channels for some products, such as the impact of storing frozen products in the RDC versus in the CDC. Finally, some additional recommendations, like micro-fulfillment centers, VMI, or inventory in the stores, can be tested in the simulation to assess feasibility.

Finally, this study focused on segmentation in the food & beverage retail industry. Similar studies can use the integrated methodology proposed in this study to apply supply chain segmentation to numerous other industries. Because of their differences, it is expected that each industry will use different product attributes, reach different segments, and create different policies for each segment. This enables them to reduce inventory, improve service level, and reduce the total distance traveled, hence improving cost and sustainability performance.

REFERENCES

Bahn, K. D., & Granzin, K. L. (1985). Benefit segmentation in the restaurant industry. *Journal of the Academy of Marketing Science*, *13*(3), 226–247. https://doi.org/10.1007/BF02729948

Bennion, M. L. (1987). Segmentation and positioning in a basic industry. *Industrial Marketing Management*, *16*(1), 9–18. https://doi.org/10.1016/0019-8501(87)90003-4

Bernad, L., & Romero, L. (2016). *Analyzing Sourcing Networks for a Coffee Retailer*. [Capstone project, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/102952

Bladyka, E. (2013). Coffee Staling Report. Specialty Coffee Association of America.

Botero Lopez, S., & Salman Chaudhry, M. (2020). Designing an Efficient Supply Chain for Specialty Coffee from Caldas-Colombia. [Capstone project, Massachusetts Institute of Technology] https://dspace.mit.edu/handle/1721.1/126376

Bové, A.-T., & Swartz, S. (2016). *Starting at the source: Sustainability in supply chains | McKinsey*. https://www.mckinsey.com/business-functions/sustainability/our-insights/starting-at-the-source-sustainability-in-supply-chains

Childerhouse, P., Aitken, J., & Towill, D. R. (2002). Analysis and design of focused demand chains. *Journal of Operations Management*, *20*(6), 675–689. https://doi.org/10.1016/S0272-6963(02)00034-7

Christopher, M., Peck, H., & Towill, D. (2006). A Taxonomy for Selecting Global Supply Chain Strategies. *International Journal of Logistics Management*, *17*(2), 277–287.

Czaika, E. G. (2010). *Starbucks Cups: Trash or Treasure? An Example of Facilitated Systems Thinking Assisting Stakeholders in Designing Their Own System to Recycle Take-Away Cups* [Master's thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/59229

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory Factor Analysis*. Oxford University Press

Feng, L., Li, Y., Xu, F., & Deng, Q. (2019). Optimal pricing and trade-in policies in a dual-channel supply chain when considering market segmentation. *International Journal of Production Research*, *57*(9), 2828–2846. https://doi.org/10.1080/00207543.2018.1551636

Fichtinger, J., Chan, C. (Wan C., & Yates, N. (2019). A joint network design and multi-echelon inventory optimisation approach for supply chain segmentation. *International Journal of Production Economics*, *209*(September 2017), 103–111. https://doi.org/10.1016/j.ijpe.2017.09.003

Fisher, M. L. (1997). What Is the Right Supply Chain for Your Product? *Harvard Business Review*, *75*(2), 105–116. http://mba.teipir.gr/files/Fisher_What_is_the_right_SC_for_your_product.pdf

Godsell, J., Christopher, M., Diefenbach, T., Clemmow, C., & Towill, D. (2010). Enabling supply chain segmentation through demand profiling. *International Journal of Physical Distribution & Logistics Management*, *41*(3), 296–314. https://doi.org/10.1108/09600031111123804

Gosling, L., & Urrutia, J. (2019). *Analytics Driving Supply Chain Segmentation for Lenovo*. [Capstone project, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/121297

Hashmi, S. A., & Zhang, W. (2016). *Performance Metrics for Food & Beverage Supply Chains*. [Capstone project, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/103089

He, Y., Li, D., Wu, S. J., & Shi, C. (2018). Quality and operations management in food supply chain. *Journal of Food Quality*, *2018,* 9871379. https://doi.org/10.1155/2018/9871379

Hill, T. (1995). *Manufacturing Strategy Text and Cases*. Macmillan Education UK. https://doi.org/10.1007/978-1-349-13724-4

Hu, X., Duenyas, I., & Kapuscinski, R. (2008). Optimal Joint Inventory and Transshipment Control Under Uncertain Capacity. *Operations Research*, *56*(4), 881–897. https://doi.org/10.1287/OPRE.1080.0515

Khakdaman, M. (2014). *On-shelf Availability Strategies in Coffee Industry*. [Master's thesis, Malaysia Institute for Supply Chain Innovation]. https://dspace.mit.edu/handle/1721.1/102137

Khamsi, C., & Stolear, V. (2016). *Efficient Supply Chain Design for Highly-Perishable Foods*. [Masters thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/107516

Kharlamov, A. A., Ferreira, L. M. D. F., & Godsell, J. (2020). Developing a framework to support strategic supply chain segmentation decisions: a case study. *Production Planning & Control*, *31*(16), 1349–1362. https://doi.org/10.1080/09537287.2019.1707896

Kumar, P. (2014). *Cold Chain Evaluation Framework in a Fragmented Market*. [Master's Thesis, Malaysia Institute for Supply Chain Innovation]

Kuruvilla, A., Norton, S., & Gee, M. (2012). *Greening The Supply Chain – Trends in Corporate America*. International Journal of Business and Management Studies, *4*(2), 1–10.

Lee, H. L. (2002). Aligning Supply Chain Strategies with Product Uncertainties. *California Management Review*, *44*(3), 105–119.

Lovell, A., Saw, R., & Stimson, J. (2005). Product value-density: Managing diversity through supply chain segmentation. *The International Journal of Logistics Management*, *16*(1), 142–158. https://doi.org/10.1108/09574090510617394

Musavi, M. M., & Bozorgi-Amiri, A. (2017). A multi-objective sustainable hub location-scheduling problem for perishable food supply chain. *Computers & Industrial Engineering*, *113*, 766–778. https://doi.org/10.1016/J.CIE.2017.07.039

Paciarotti, C., & Torregiani, F. (2021). The logistics of the short food supply chain: A literature review. *Sustainable Production and Consumption*, *26*, 428–442. https://doi.org/10.1016/J.SPC.2020.10.002

Protopappa-Sieke, M., & Thonemann, U. W. (2017). *Supply Chain Segmentation*. Springer. https://doi.org/10.4018/978-1-4666-8228-3.ch004

Ptok, F. L., Eduardo, J., & Henao, C. (2021). *Supply chain segmentation in the apparel industry By Supply chain segmentation in the apparel industry*. [Capstone project, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/130980

Roscoe, S., & Baker, P. (2014). Supply chain segmentation in the sporting goods industry. *International Journal of Logistics Research and Applications*, *17*(2), 136–155. https://doi.org/10.1080/13675567.2013.837869

Svoboda, J., & Minner, S. (2021). Tailoring inventory classification to industry applications: the benefits of understandable machine learning. *International Journal of Production Research, 60*(1), 388–401. https://doi.org/10.1080/00207543.2021.1959078

Terlunen, S., Fischer, J.-H., & Hellingrath, B. (2013). *Introducing Supply Chain Segmentation Procedures into Flexibility Management*. *Proceedings of the 24th Annual Conference of the Production and Operations Management Society (POMS 2013)*.

Terlunen, S., Horstkemper, D., Wölck, M., & Hellingrath, B. (2015). Simulation-based Decision Support for Vertical Supply Chain Segmentation Scenarios. *Proceedings of the Simulation in Production and Logistics, Dortmund, 157–166*.

Theodoras, D. (2009). Customer Service Strategy and Segmentation in Food Retailing using the Importance-Performance Paradigm. *Supply Chain Forum: An International Journal*, *10*(2), 64–66. https://doi.org/10.1080/16258312.2009.11517218

Toledo, P. R. A. B., Pezza, L., Pezza, H. R., & Toci, A. T. (2016). Relationship Between the Different Aspects Related to Coffee Quality and Their Volatile Compounds. *Comprehensive Reviews in Food Science and Food Safety*, *15*(4), 705–719. https://doi.org/10.1111/1541-4337.12205

Validi, S., Bhattacharya, A., & Byrne, P. J. (2014). A case analysis of a sustainable food supply chain distribution system—A multi-objective approach. *International Journal of Production Economics*, *152*, 71–87. https://doi.org/10.1016/J.IJPE.2014.02.003

Wang, X., & Li, D. (2012). A dynamic product quality evaluation based pricing model for perishable food supply chains. *Omega*, *40*(6), 906–917. https://doi.org/10.1016/J.OMEGA.2012.02.001

Zanoni, S., & Zavanella, L. (2012). Chilled or frozen? Decision strategies for sustainable food supply chains. *International Journal of Production Economics*, *140*(2), 731–736. https://doi.org/10.1016/j.ijpe.2011.04.028

APPENDIX A – CALCULATIONS FOR DISTANCE TRAVELED TO CUSTOMERS USED FOR SIMULATION

Weight formula (rounded to one decimal place):

Let $CDC_{i,W}$ indicate the weight assigned to CDC_i, $CDC_{i,S}$ indicate the number of stores assigned to CDC_i and $CDC_{i,D}$ indicate the distance assigned to CDC_i where $i \in \{All\ CDCs\}$

$$CDC_{i,W} = 1 + \left(\frac{CDC_{i,S}}{\min(CDC_{1,W}, \dots, CDC_{47,W})} - 1\right) * \frac{1}{10} for\ \forall\ i \in \{All\ CDCs\}$$

Below table includes a hypothetical sample of distance calculation from the CDCs to stores:

CDC	# of Stores	Weight	Distance
CDC 1	350	1.3	78
CDC 2	694	1.8	108
CDC 3	100	1	60
CDC 4	80	1	60

APPENDIX B – STATISTICS FOR ITEM CLUSTERING*

* Representative values, not actuals, are used for each field.

Row Labels	Count of Items	Average of Average Weekly Demand	Average of COV Weekly Demand	Average of Item Cost	Average of Shelf Life
Cluster 1 (Slow Moving Products)	**304**	**42596.06**	**0.52**	**24.31**	**1522.56**
Beverages	16	49968.50	0.70	3.91	216.75
Equipment	1	465.04	0.41	24.46	1825.00
Packaged Coffee	25	34744.82	0.68	13.85	252.56
Paper/Plastic Product	212	51586.01	0.44	27.45	1698.12
Service Part	3	1011.57	0.25	24.38	1825.00
Smallware	46	7414.46	0.78	23.02	1825.00
Supplies	1	288.84	0.34	5.71	1825.00
Cluster 2 (Fast Moving Products)	**105**	**1084063.89**	**0.39**	**1.81**	**162.32**
Beverages	22	691781.15	0.45	2.16	239.75
Packaged Coffee	3	513393.37	0.26	10.02	238.00
Savory Food	48	933406.42	0.39	1.94	104.07
Sweet Food	32	1633244.83	0.35	0.60	189.38
Cluster 3 (Complex Products)	**50**	**316752.48**	**1.15**	**7.59**	**234.97**
Beverages	15	207157.23	1.25	2.97	195.30
Packaged Coffee	16	17028.81	1.09	14.56	238.00

Print	1	1317.08	3.42	56.28	1825.00
Savory Food	3	587968.85	0.32	3.11	2.00
Sweet Food	15	712838.73	1.11	2.43	212.00
Grand Total	**459**	**310704.87**	**0.56**	**17.34**	**1071.14**