International Journal of Mathematical, Engineering and Management Sciences
Vol. 9, No. 2, 283-304, 2024
https://doi.org/10.33889/IJMEMS.2024.9.2.015

A Data-driven Approach for Planning Stock Keeping Unit (SKU) in a Steel Supply Chain

Shivchandra Prabhat Wakle
Department of Mechanical Engineering,
Malaviya National Institute of Technology Jaipur, Rajasthan, India.
E-mail: 2020pie5313@mnit.ac.in

Ved Prabha Toshniwal
Department of Mechanical Engineering,
Malaviya National Institute of Technology Jaipur, Rajasthan, India.
E-mail: 2021rme9074@mnit.ac.in

Rakesh Jain
Department of Mechanical Engineering,
Malaviya National Institute of Technology Jaipur, Rajasthan, India.
E-mail: rjain.mech@mnit.ac.in

Gunjan Soni
Department of Mechanical Engineering,
Malaviya National Institute of Technology Jaipur, Rajasthan, India.
Corresponding author: gsoni.mech@mnit.ac.in

Bharti Ramtiyal
Department of Management Studies,
Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India.
E-mail: bharti.mnit2022@gmail.com

(Received on August 27, 2023; Revised on October 9, 2023 & November 2, 2023 & January 2, 2024; Accepted on January 4, 2024)

Abstract

In response to the growing complexities in supply chain management, there is an imperative need for a data-driven methodology aimed at optimizing inventory allocation strategies. The purpose of this research is to enhance the efficiency of allocation and operational scheduling, particularly concerning the stock keeping units (SKUs). To achieve this, one year's operational data from a specific organization's SKUs is taken and machine learning tools are employed on the data collected. These tools are instrumental in identifying clusters of SKUs that exhibit similar behaviour. Consequently, this research offers recommendations for rational inventory allocation strategies that are finely attuned to the unique characteristics of each SKU cluster. Results obtained reveals substantial disparities between the recommended strategies for the organization's SKUs and those typically found in the literature such as same strategy cannot be used for all different types for products. This underscores the critical importance of adopting a tailored approach to supply chain management. Furthermore, the research demonstrates the remarkable efficiency of unsupervised machine learning algorithms in determining the optimal number of segments within the SKUs. The current research differentiates from others in a way that in most of the research, the holistic data-driven approach is underutilized, right from the selection of the clustering algorithm to the validation of segments.

Keywords- Machine learning, Steel industry, Supply chain segmentation, Inventory optimization, Forecasting.

1. Introduction

Though supply chain management (SCM) has evolved over time many companies are operating in a costcentric approach and few of them have followed a profit-driven approach the latter have used supply chain segmentation strategies to react to the customers in a most efficient and responsive manner according to the characteristics of product portfolio (Alicke and Forsting, 2017). These companies have inclined their supply chain activities in accordance with the customer requirements so as to serve the products with tailored supply chain strategies (Protopappa-Sieke and Thonemann, 2017). In production and operations management, the organization frequently has to address a wide range of items or stock holding units (SKUs) (van Kampen et al., 2012). When there is a mismatch of fit between the product characteristics and supply chain requirements it comes up with many challenges like inventory challenges, loss of sales, etc. The features of these products or SKUs determine the manufacturing and inventory strategies of these various SKUs (van Kampen et al., 2012). During the last few years, due to the vast availability of data and the capability of tools and techniques, it has become feasible to exploit the historical supply chain data in order to gain a competitive advantage. The beauty of data-driven models is that before creating the precise model in a data-driven study first the data is collected from the company and then the meticulous study of data from an academic point of view reveals potential areas for development (Simchi-Levi, 2014). This gives us the opportunity to evaluate the fit of existing SKUs to current replenishment and order execution or inventory allocation strategy with the strategies mentioned in the literature.

Until now the research on this area is mostly based at product level but can be utilised at SKU level depending upon industry. The manufacturer and customer data can be utilised. The use of segmentation method beyond fast moving consumer goods (FMCG) sector can be explored. Until now in most of the researches only single clustering methods and that too hierarchical clustering methods have been used but the use of volume efficient and unsupervised learning methods like K-means clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) can be explored. The practitioners should try to include more features apart from current studies (Gosling and Urrutia, 2019). This research poses the following research questions:

RQ*1:* Is the SKU portfolio of XYZ homogenous in terms of supply chain characteristics?
RQ*2:* Is it possible to divide the XYZ's SKU portfolio into categories concerning their supply chain characteristics?

The following is a structure of this paper: the upcoming section examines the current literature on segmentation and unsupervised machine learning approaches. It follows by the methodology which explains the steps we have applied in this project with respect to the literature which we have discussed. Later the outcome of this project at XYZ company are explode and discuss. In this each segment obtained is properly analyzed and explained based upon literature and discussion with company team member. Section 5 discusses the conclusions followed by managerial implications and opportunities on the future scope of this project work.

2. Literature Review

2.1 Inventory Management Challenges in Metal Industry

The management of inventory in today's volatile market has become most important than ever (Tang et al., 2014). A particular firm has different types of inventories for example in the form of raw material, work-in-progress (WIP) and finished goods (Nag et al., 2014). In metal industry inventory management is one of the challenging areas (Knizek et al., 2016).

(i) Unlike the other items such as food, electronics items the metal do not get decay or obsolete rapidly. Because of this the decision makers tends to store the inventory in between any stages of manufacturing process and the consequences are often neglected. Though you can melt the unsold item but since the manufacturing time is high the holding of inventory goes into months and years and thus poor cash availability (Zaky et al., 2023).

(ii) Another challenge which is not exclusive in metal industry rather in all industries is that product spread. As the number of goods, a company produces or have in its portfolio increases it also gives invitation for complexity in the system because the amount of inventory required will be higher and also more setups need to be changed along production lines. This not only decrease the efficiency but also give the danger of items getting outdated which may left the firm with considerable volume of completed items that must be disposed at any cost maybe in the form of discounted sales or remelting (Denton et al., 2003).

Ways to address these challenges:
(i) Due to availability of data and capability to process it in recent times, the companies should explore their previous records and exploit the meaningful insights to decide how much inventory and at what stage to place in their supply chain. It has seen many a times that companies predict the orders and makes its WIP but that order actually never came resulting into excess WIP (Kalagnanam et al., 2000).

(ii) It has seen that the many times considerable inventory was being held at in the incorrect yards in the supply chain hence complete examination of firm's inventory in terms of quantity, kind and region of warehouse needs to be done (Knizek et al., 2016).

(iii) In terms of complexity of product portfolios, the metal industries can take the lesson from CPG companies on how to deal with product complexity. Firms should seek to minimise the complex and diverse product portfolio in order to eliminate extra inventory and kept manufacturing running (Denton et al., 2003).

2.2 Need of Segmentation
In today's world to stay relevant and competitive in the market companies use supply chain as a differentiating factor. In order to identify what should be the right supply chain Many researchers and scholars have given a framework. One of such is a work of Fisher (1997). The right supply chain refers to the one which is efficient, cost-effective, agile and moreover sustainable. Since all the industries are thriving towards the ideal supply chain with no inventory pile up, no shortages or back orders, minimum cost and responsive towards the market fluctuations. This section covers the points needed to be considered to move towards the right supply chain.

Lee (2004) stated in his article 'triple a supply chain' that the supply chain strategy should be adaptable, agile and there should be alignment. The researchers have stated that one size fits all strategy should not be used because each customer or market has different requirements and we should match product supply chain with the given requirement in order to act both on efficiency and responsiveness (Aitken et al., 2003; Kharlamov et al., 2013; Li and O'Brien, 2001; Payne and Peters, 2004; Shewchuk, 1998; Simchi-Levi, 2010).

In production and operations management, organizations deal with diverse stock-keeping units (SKUs) that vary in function, size, appearance, and location. These SKU differences impact manufacturing and inventory strategies, often requiring distinct approaches due to varying factors like sales volume and demand consistency. To maintain control over inventory and production, categorizing SKUs into attribute-based groups is a practical strategy, aiding decisions on manufacturing and inventory allocation.

Hence, one should try to inline the supply chain strategy and their products with respect to its characteristics (Christopher and Towill, 2002; Payne and Peters, 2004; Qi et al., 2009). By splitting the SKUs into different clusters or groups based on similar characteristics the differentiate strategies can be developed. By this the organisation or company can effectively meet the client requirements in the most feasible manner. Continuous reinforcing the need for a customized supply chain management strategy, we hope to reach practitioners and encourage them to abandon one-size-fits-all methods in favor of solutions tailored to their individual needs.

2.2.1 Segmentation Process

One should start with the organisation i.e., the context within which you have to apply segmentation and then choose the features or variables relevant for segmentation and then find the segments to match them with the required tailored operation strategies (Godsell et al., 2011).

For supply chain segmentation criteria need to be selected but there is a scarce of detail on how these criteria should be identified and selected. Each author has used different criteria based upon their intuition and context or aim of a segmentation and available data. The criteria which have been selected needs to be treated equally (Kharlamov et al., 2020).

The Identification of characteristics or features is important and can differ from industry to industry (Fisher, 1997; Frohlich and Westbrook, 2001; Schnetzler et al., 2007). It is important to understand product and market characteristics in order to develop supply chain strategy (Fisher, 1997; Lamming et al., 2000; Kharlamov et al., 2013). Different researchers have given different models by considering different product features for example functional and innovative in which they suggested to focus on efficiency for functional products and on responsiveness for innovative ones (Kharlamov et al., 2013). Lamming et al. (2000) considered uniqueness of product, complexity of product and further enhance the fisher's model and (Lee, 2002) worked on uncertainty or volatility of supply and demand. Based on this (Christopher and Towill, 2002; Martin and Towill, 2000) considered variables such as volume, variety, lead time, coefficient of variation, life cycle to create a model (Childerhouse et al., 2002; Vitasek et al., 2003).

According to most researchers the identification of segments is usually by predefined level which came from intuition (Godsell et al., 2011; Holweg, 2005; Payne and Peters, 2004; Vitasek et al., 2003). Selection of many levels makes the segments and analysis difficult to interpret (Christopher et al., 2009).

After identification of segments, it comes to a operationalize those segments for which tailored strategies and practices reviewed from the literature of operations and supply chain management and also the existing managers input is taken (Kharlamov et al., 2020). There is a shift from generalized to hybrid methods of combining data with intuition to gain insights by the deductive abductive methods in order to get differentiated segments (Godsell et al., 2013).

All these researches have acted as inspiration for researcher and as a basic foundation and motivated to go further deeper (Godsell et al., 2011; Qi et al., 2011; Selldin and Olhager, 2007).

2.2.2 Data Driven Segmentation

Marketing does have a strong record of using advanced data mining techniques for consumer segmentation to acquire insights. Such findings could then be utilised to help design promotional activities (Gosling and Urrutia, 2019; Payne and Frow, 2006; Sanders, 2016). The challenge of comprehending variability within huge groups of current and prospective customers depending on individual attributes is solved using data mining technologies. With so many suppliers, goods, and consumers, SCM has a similar challenge.

In the context of SCM, data mining driven approaches have been used to minimise production costs (Zhao et al., 2017) optimise product designs (Song and Kusiak, 2009) analyse forecast patterns (Altintas and Trick, 2014) and choose vendors in mass customisation, among other things (Ni et al., 2007). Data mining approaches have been used to certain level in quality control in sustainable and environmentally friendly supply chains, using the readily available logistical data (Ting et al., 2014). Data mining technologies address the challenge of recognising variation within huge groups of existing and prospective consumers based on their unique qualities. SCM confronts a similar issue because of its different vendors, goods, and consumers.

In priori segmentation the segmentation is carried out based on the criteria or features which are came from experience or intuition of managers (Green, 1977). Here objective is to gain quick insights and association between features. It is effective in case of less features and simple groups (Kharlamov et al., 2020). Post-hoc methods are data driven approaches in which data is clustered based upon similar characteristics to obtain more unique groups (Kharlamov et al., 2020).

Clearly the post-hoc methods give more meaningful insights and trends in the data and their accuracy is also high due to which tailored marketing strategies can be developed (Clarke, 2009; Kazbare et al., 2010; Kharlamov et al., 2020; Martínez-López and Casillas, 2009). It is less typical, nevertheless, to use data mining techniques to facilitate SC segmentation. This provides a chance to use marketing techniques to comprehend differences between groups in the context of SCM (Kharlamov et al., 2020).

2.2.3 Supply Chain Segmentation Approaches
There are three types of approaches companies usually adopt (Protopappa-Sieke and Thonemann, 2017). market driven segmentation which considers external factors like customers, region, market etc. (Hill, 2000) give this kind of framework where he mainly introduced differentiated or segmented manufacturing strategies to catering requirement of specific market segment rather than a one size fit all strategy to market. After that other researcher added for the capabilities other than manufacturing. Another one is product driven segmentation, in this the major emphasis is on internal information of product, its sales, performance and demand patterns (Protopappa-Sieke and Thonemann, 2017). The product-driven method focuses on internal product details, sales and performance history, and demand characteristics, as opposed to the market-driven approach. Fisher (1997) pioneered the product-driven approach with his research of functional versus inventive goods. The last one is hybrid approach segmentation, here both the above two stated factors used in order to segment products. Irrespective of approach the end result is tailored strategy for each segment. As explained in various sections of the article, this personalized perspective not only improves operational efficiency but also develops adaptation in dynamic business contexts.

2.3 Systematic Clustering Procedure
Authors (Kaya and Schoop, 2022; Xu and Wunsch II, 2005) have discussed a very systematic way of Clustering exercise as shown in figure which consists of following major steps: Feature selection selects distinctive characteristics from a group of candidates, whereas feature extraction employs modifications to produce usable and unique features out from originals (Bishop, 1995; Jain et al., 1999, 2000; Xu and Wunsch II, 2005). Then Selection of clustering technique among many available can be be done with the help of cluster validity metrices. Next is Validation of cluster because for a given input dataset the clustering algorithm will always create a cluster irrespective of whether there exists a pattern or not. Furthermore, various techniques often result in different clusters. In fact, for the same method the input parameters might have an impact on the final findings. Due to which there is a need of effective assessment criteria in order to provide stakeholders a level of trust on results obtained. The evaluations should be unbiased i.e., should not favour specific algorithm as well as objective (Xu and Wunsch II, 2005). And at last Interpretation of Results in which the data division is interpreted by experts in the appropriate domains. Additional analyses

and experimentation, may be necessary to ensure the dependability of retrieved knowledge (Xu and Wunsch II, 2005). Figure 1 explains the process in detail.

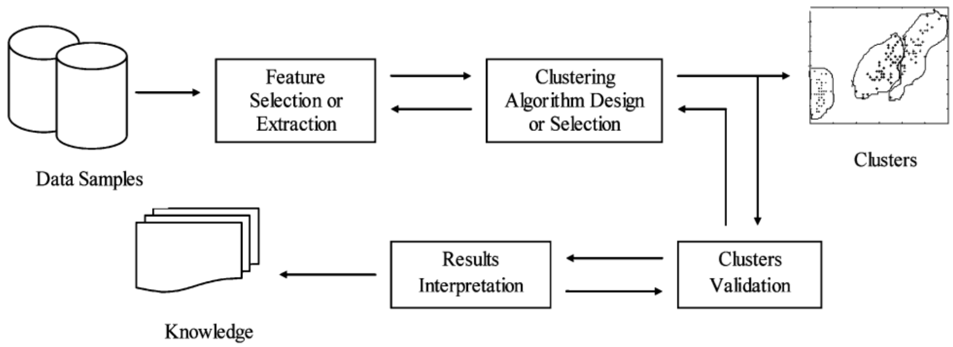

Figure 1. Systematic clustering process (Xu and Wunsch II, 2005).

The objectives of this research project are:
(i) To identify XYZ's SKU clusters by data-driven methodology and,
(ii) To propose a tailored inventory allocation strategy for individual clusters with referring to the standard practices in literature and discussion with the team members.

3. Methodology

Planning is like adhesive which helps in the proper working of the supply chain (Kharlamov et al., 2020). XYZ was following the same order execution and replenishment strategy for all these SKUs irrespective of carefully studying and identifying their characteristics based upon previous data. It is understood that these SKUs need a unique or differentiated execution process but it is unclear among many SKUs which SKU needs what process i.e., we want to find out ''known unknown''. In order to do that we took the help of data mining techniques to form different clusters.

The Figure 2 shows the steps we have taken to answer the research questions:

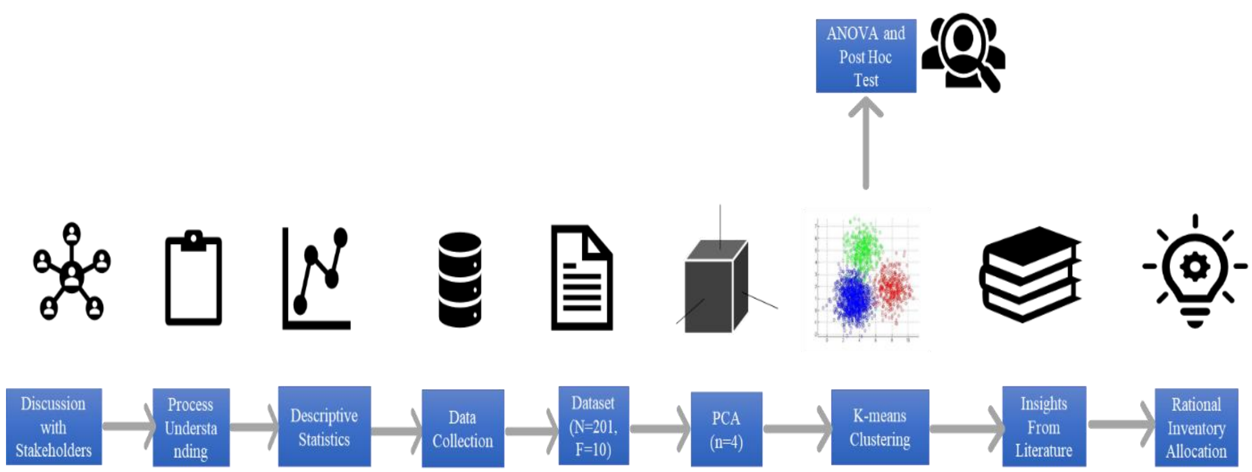

Figure 2. Steps taken to address research questions developed.

3.1 Case Organization

XYZ is one of the top stainless-steel producers in the world. It manufactures various products like slabs, blooms, hot rolled, cold rolled, sheets, plates, etc. Many of which are also exported worldwide. The company is committed to keeping high-quality standards, effective delivery schedules, competitive pricing, and the best after-sales services to remain competitive in the market. The company offers several grades, thickness, width, and quality requirements to its customers and has a wide range of SKU portfolios. Considering the broad product portfolio, managing the entire product portfolio is challenging while keeping customer satisfaction at the highest level.

The company does not have any central warehouse rather it has regional warehouses only through which the SKUs are served to the customers. The company also has service centres which converts the products into the slit form according to the requirements and the WIP inventory is mainly kept at Hot Rolled Annealed and Picked (HRAP) stages, single echelon from where it is then served as and when order comes. It is because in steel manufacturing the planning of orders is usually by matching the customer order to existing finish goods inventory and others by allocating unfinished one from the inventory stock. Apart from that those orders do not match with existing inventory need to be scheduled.

3.2 Data Collection

For analysis data related to sales and operations from last one years which was mainly available in XYZs enterprise resource planning (ERP) system was taken. Among those data few features were not suitable for target purpose and that was not used. The data taken from the ERP systems cover the sales and operations related to one plant and its warehouses of 201 data points that makeup XYZ's 2021 SKU profile. The data chosen is mainly after Covid-19 pandemic restrictions were lifted since there was irregularity in data during the lockdown times.

The XYZ has robust data record system in which the generated data is saved on a regular basis. The individual records of each month where extracted and then all are merged to have a single one. In this data set we identified 26 features which was used for defining transactions. We identified 20 of those as key features which were characterising the SKU. One thing to note that is many of the variables was not in the datasheet we extracted but were instead computed with the help of already present data as metrics. Some variables are categorical while others were numeric. The analysis was done on a monthly basis as the production planning is on monthly basis.

The variables or features which are irrelevant for the purpose for example coil number, delivery number are dropped off and finally the remaining features which was in the dataset were considered. The resulted dataset focused on SKUs.

According to the literature there are many variables which we can consider for e.g., Volatility, Volume, Price, Margin, Average inter-demand interval (ADI), Manufacturing lead time (LT) etc. (Gosling and Urrutia, 2019; Kharlamov et al., 2020). Table 1 shows features used to characterize SKUs in this research, their description and data type and their references from literature.

3.3 Data Analysis

3.3.1 Modelling and Evaluation

Before applying the principal component analysis (PCA) and clustering algorithms we need to normalize the data. As some of the algorithms are sensitive to distance and the unit of the feature for e.g., the price is in rupees per kg while average sales are in tonnes thus having different units and this influence their weightage in cluster forming. To solve this problem normalization is done which converts the feature values in the range of 0 to 1. Once we normalized the data, we calculate Kaiser-Meyer-Olkin (KMO) value in order to test for sample adequacy. We got KMO value of 0.643 which suggested that data is factorable (Kharlamov et al., 2013, 2020). PCA is ued to in order to reduce the number of features based on which segmentation to be done and also for better cluster formation. For finding out how many numbers of principal components to extract we used cumulative explained variance and found that first four principal components are explaining almost 90% of variation in the data as shown in Table 2. There is no specific rule for how much explained various is good rather it depends upon the intended application and judgement (Berthold and Hand, 2003; Kharlamov et al., 2020). Up to 95% can be used instead of overall variance (Jolliffe, 2011; Rahim et al., 2021). The PCA gave us four principal components as output as shown in Table 2.

Table 1. Features used to characterize SKUs in this research, their description and data type.

Features used to characterize SKUs in this research	Description	Data type	References
Volatility	Coefficient of variation of monthly dispatches	Numeric	Kharlamov et al. (2013, 2020), Lamming et al. (2000), Lee (2002), Simchi-Levi and Timmermans (2021)
Volume	Average dispatches per month	Numeric	Kharlamov et al. (2020), Lee (2002), Simchi-Levi and Timmermans (2021)
Profit margin	Profit per kg on SKU	Numeric	Fisher (1997), Lee (2002), Simchi-Levi and Timmermans (2021)
Average inter-demand interval (ADI)	Average interval between two demands	Numeric	Boylan et al. (2008), Ptok and Camargo Henao (2021)
Price	Price per kg of SKU	Numeric	Bandyopadhyay et al. (2021), Fuller et al. (1993), Jackson et al. (2019), Ptok and Camargo Henao (2021), Simchi-Levi and Timmermans (2021)
Number of destinations	Count of unique ship to city per SKU	Numeric	Gosling and Urrutia (2019)
Number of orders	Number of orders per year	Numeric	Gosling and Urrutia (2019), Payne and Peters (2004)
Number of customers	Count of unique ship to customer per SKU	Numeric	Gosling and Urrutia (2019), Payne and Peters (2004)
Average order size	Average quantity per order line	Numeric	Gosling and Urrutia (2019), Payne and Peters (2004)
Number of yards	Count of unique ship to yard per SKU	Numeric	-

Table 2. Total variance explained.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.017	40.175	40.175	4.017	40.175	40.175	3.466	34.656	34.656
2	2.31	23.1	63.275	2.31	23.1	63.275	2.012	20.119	54.775
3	1.441	14.407	77.682	1.441	14.407	77.682	1.989	19.892	74.666
4	1.231	12.312	89.994	1.231	12.312	89.994	1.533	15.327	89.994
5	0.634	6.342	96.336						
6	0.143	1.434	97.77						
7	0.112	1.116	98.886						
8	0.084	0.842	99.728						
9	0.026	0.262	99.99						
10	0.001	0.01	100						
Extraction Method: Principal Component Analysis.									

3.3.2 Choosing Clustering Method

Now once we identify the number of principal components, we will be going to do segmentation based upon these four principal components. Researchers select the type of clustering method to be used by their experience or prior knowledge (Han et al., 2011). This method of selecting algorithm is subjective since the performance will depend upon the algorithm utilised. We need objective selection criteria for choosing clustering algorithm (Jun, 2006; Park et al., 2003). Jun and Lee (2010) made use of objective selection criteria in which the algorithms are compared with each other by their silhouette score (Jun and Lee, 2010). Kaya and Schoop (2022) have discussed the research approach in which they mentioned how the clustering algorithm can be chosen objectively with the help of Silhouette index, Davies-Bouldin index etc.

As we have seen the researchers have started working towards choosing clustering methods based on subjective approaches but since our research is based upon data driven approach we have chosen the best suited algorithm between the well tested algorithms which researchers have used successfully (Jun and Lee, 2010; Theodoridis and Koutroumbas, 2009; Xu and Wunsch II, 2005). This is due to the fact that we cannot focus on each and every possible algorithm out there. While there are number of clustering algorithms but K-means, Agglomerative, Density-based spatial clustering of applications with noise (DBSCAN) etc., are well established and tested clustering algorithms (Shi et al., 2017). Shi et al. (2017) compared these well tested algorithms on different data sets where they used silhouette score for choosing the best algorithms.

Rousseeuw (1987) introduced silhouette index. Silhouette Score is complicated with respect to running time but at the same time it provides better information than elbow and Davies-Bouldin score (Kaya and Schoop, 2022). This index lies between -1 to 1. Ideally the silhouette index should be closed to 1 i.e., high value is preferable (Wiersma et al., 2021). The minimum value Davies–Bouldin index is ideal for clusters to be compact and well separated. (Milligan and Cooper, 1985) in their paper got one of the best results for Calinski Harabasz score metrics and for this metrics the highest score is preferred. Also, in order to implement the algorithm in python the number of clusters needed to be entered. For which we use cluster validation Metrices such as Davies Bouldin, elbow score, Calinski Harabasz and Silhouette score (Tibshirani et al., 2001). Cluster validity metrics for different algorithms can be seen in Table 3.

Table 3. Cluster validity metrics summary table for different clustering algorithms.

Clustering Type	Algorithm	Number of clusters "k"	Silhouette score	Davies Bouldin score	Calinski Harabasz score
Partitional	K-Means	5	0.4285	0.7641	111.7908
Hierarchical	Agglomerative	5	0.4243	0.7628	107.9704
Density Based	DBSCAN	20	0.2834	3.1033	18.3227
Non-Hierarchical	GMM	12	0.2393	1.2028	65.8400

After comparing these algorithms with their cluster validity metrices we confirmed that K means clustering is outperforming all the rest for our intended application. Furthermore the K means clustering algorithm is better in terms of scalability (Rahim et al., 2021). So, we choose K means clustering for the clustering purpose.

We used the elbow curve method shown in Figure 3 to determine the optimal number of clusters in our K-means analysis. The "elbow point" on this curve, where the distortion score levels off, is crucial, but it doesn't precisely pinpoint the ideal k value. Fit time analysis suggests k = 5 among options like 4, 5, and 6, but the elbow method is subjective and adding too many clusters can harm cluster quality. To address this, alternative methods like the Silhouette score, Davies-Bouldin index, and Calinski-Harabasz index can be considered.

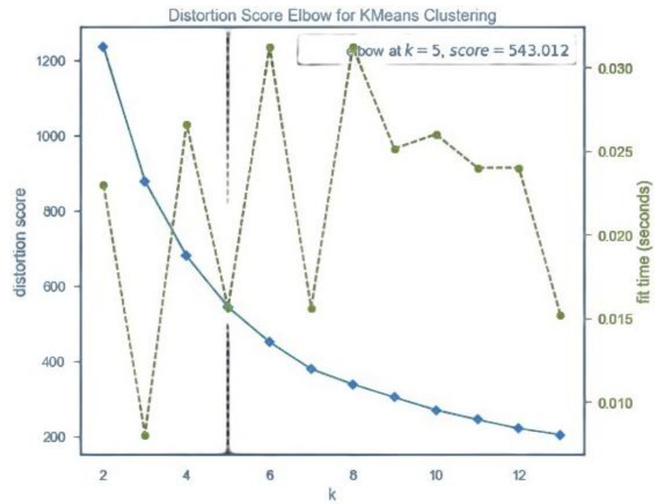

Figure 3. Elbow curve for K-Means clustering.

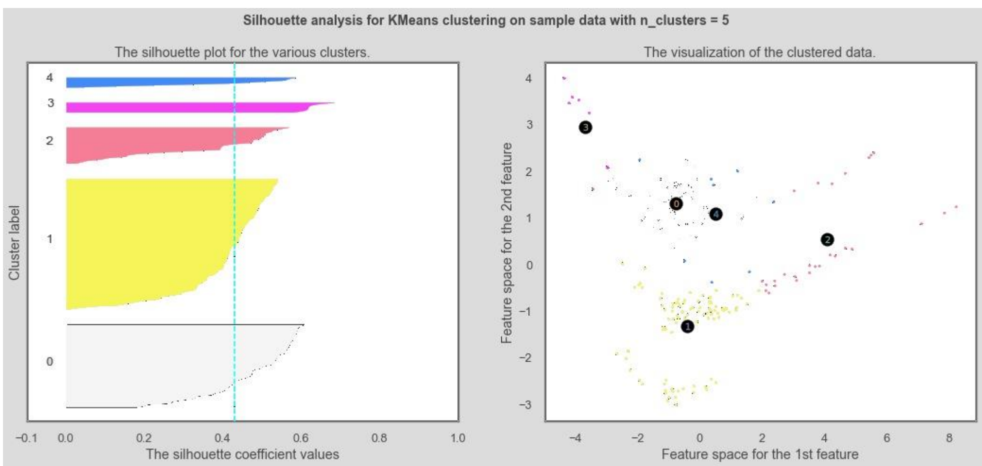

Figure 4. Silhouette analysis for K-Means clustering with five clusters.

Table 4. Clustering results.

Classification variables	Segments				
	Cluster: 0	Cluster:1	Cluster:2	Cluster:3	Cluster:4
Variety (number of SKU's)	28	95	8	8	62
Volume (mean tons)	138.723	27.8107	25.4766	412.881	26.0816
Volatility (mean)	0.6919	0.82584	3.4641	0.58599	1.38462
Margin (mean)	18.0357	21.5263	14.625	15.625	11.9355
Price (mean)	257.143	308.537	208.75	222.75	169.355
Number of yards (mean)	3.67857	1.53684	1	1	1.59677
ADI (mean)	1.02403	1.4465	12	1.03636	2.37237
Number of customers (mean)	78.0714	21.6	9.375	26	22.8226
Number of destinations (mean)	27.8214	9.03158	4.25	9.75	9.85484
Number of orders (mean)	364.786	55.7474	34.25	245.25	69.6452
Average order size (mean)	4.31837	6.14915	11.9854	20.9894	6.30917

The silhouette score graph in Figure 4 serves as an essential tool for evaluating the separation and quality of clusters generated by K-means clustering. The number of clusters is represented on the graph's x-axis by various k values, and the scores for each k are shown on the y-axis. Silhouette scores, which range from -1 to 1, assess the coherence and distinctiveness of a cluster. While a score close to 0 denotes probable overlap or ambiguity, one close to +1 denotes well-defined clusters with distinct boundaries. A score that is nearly -1 indicates potential misassignments. The k value with the highest silhouette score, which achieves the optimum balance between separation and cohesion, is used to determine the number of clusters that should be used. Analysis of the silhouette score is essential for evaluating the quality of the clusters and guiding cluster selection.

The clustering results obtained are summarized in following Table 4. To characterise the clusters we use centroids i.e. Mean (Kharlamov et al., 2013).

3.3.3 Analysis of Variance (ANOVA) and Post-hoc Analysis

ANOVA, as Milligan (1996) suggests, can be used to confirm clustering solutions. The robustness of the Clustering Analysis performed can be verified using ANOVA (Cagliano et al., 2022). A null hypothesis of ANOVA is there are no significant differences between the groups under examination. If a p-value produced from a test is less than a cut-off point, which is normally equivalent to 5%, then null hypothesis must be rejected, so it is feasible to confirm a difference between the groups studied (Rezaei et al., 2018). ANOVA is used to test the difference between means or centroids of clusters (Ciric Lalic et al., 2022; Dikmen et al., 2009; Lee et al., 2022; Saayman et al., 2012; Sudol et al., 2022; Zancanaro et al., 2007) and Tukey post-hoc test is used for multiple comparison (Cao et al., 2015; Flynn et al., 2010; Gosling and Urrutia, 2019; Revilla et al., 2013).

Table 5. Games-Howell multiple comparisons post-hoc test for feature volatility.

Characteristics	(I) cluster	(J) cluster	Mean Difference (I-J)	Sig.
Volatility	Cluster 0	Cluster 1	-0.13394	0.183
		Cluster 2	-2.77221*	<.001
		Cluster 3	0.1059	0.734
		Cluster 4	-.69273*	<.001
	Cluster 1	Cluster 0	0.13394	0.183
		Cluster 2	-2.63827*	<.001
		Cluster 3	0.23984	0.141
		Cluster 4	-.55879*	<.001
	Cluster 2	Cluster 0	2.77221*	<.001
		Cluster 1	2.63827*	<.001
		Cluster 3	2.87811*	<.001
		Cluster 4	2.07948*	<.001
	Cluster 3	Cluster 0	-0.1059	0.734
		Cluster 1	-0.23984	0.141
		Cluster 2	-2.87811*	<.001
		Cluster 4	-.79863*	<.001
	Cluster 4	Cluster 0	.69273*	<.001
		Cluster 1	.55879*	<.001
		Cluster 2	-2.07948*	<.001
		Cluster 3	.79863*	<.001

ANOVA has one important assumption that the variances are assume to be equal and then only we can conduct the test for post-hoc analysis but in case of unequal variances the Welch's ANOVA can be utilised. If we use classical ANOVA test in case of unequal variances the results might be wrong. Welch test is less sensitive in case of unequal variances. Since we have unequal variances so we can't use tukey post-hoc test rather we can use games-howell post-hoc test which is a more advanced form of Tukey-Kramer method and used in case of unequal variances (Amemiya et al., 2018). If we have number of items greater than equal to 6, games-howell post-hoc test can be applied (Lee and Lee, 2018). Welch test is used in case of unequal variances irrespective of size of the groups (Dederichs et al., 2021; Lix et al., 1996). In our case the cluster sizes were not same and also the equality of variance assumptions is not satisfied hence we used Welch Robust Tests of Equality of Means and for post-hoc we used Games-Howell. The ultimate goal of the ANOVA in this research is to see if the distinct clusters formed are indeed different for each variable considered. As a result, the test is performed for each research variable. The results for feature volatility are shown in Table 5. All of the tests performed for ANOVA are significant, indicating that the clustering was successful in accurately grouping the SKUs in the datasets.

4. Results

***RQ*1**

Once we have got the results after clustering and one-way ANOVA. It is evident that XYZ's SKU portfolio differentiates in variety of features. From the ANOVA results we can further conclude that the clusters are heterogeneous and different from each other. This answers our *RQ*1 and *RQ*2. SKU portfolio of XYZ is not homogenous in terms of supply chain characteristics and it is possible to divide the XYZ's SKU portfolio into categories as we have got five different clusters.

In order to determine with respect to a feature which cluster are different from one another we conducted games howels post-hoc analysis. From these test results, we can interpret below findings: post-hoc analysis showed that cluster 0 and 1, cluster 0 and 3, cluster 1 and 3 fail to reject the null hypothesis on the feature volatility while rest cluster combinations do not. Cluster 1 and 2, cluster 1 and 4, cluster 2 and 4 fail to reject the null hypothesis on the feature volatility while rest cluster combinations do not. Cluster 0 and 3, cluster 0 and 4, cluster 0 and 1, cluster 1 and 2, cluster 1 and 3, cluster 1 and 4 fail to reject the null hypothesis on the feature volatility while rest cluster combinations do not. On feature number of yards cluster 1 and 4 fail to reject the null hypothesis while rest all do not. On feature ADI cluster 1 and 4 fail to reject the null hypothesis while rest do not.

We can take the clusters having coefficient of variation greater than 1 as high volatile cluster and for smaller than one we can take it as low volatile (Slivinskiy, 2005). Which is also supported by the fact that from post-hoc analysis cluster 0,1 and 3 fail to reject the null hypothesis thus we can take them as low volatile. Similarly for cluster 2 and 4 fail to reject the null hypothesis and we can take both of them as high volatile clusters. While low and high volatile clusters do not fail to reject the null hypothesis. Based on similar approach for other features we can get final Table and we can summarise the results as shown in Table 6. Table 7 shows the clusters, their features and recommended strategies according to literature:

Table 6. Characteristics of each segment.

Segments						
Classification variables	**Cluster:0**	**Cluster:1**	**Cluster:2**	**Cluster:3**	**Cluster:4**	**Significance**
Variety (number of SKU's)	28	95	8	8	62	<0.001
Volume (mean tons)	Low	Low	Low	High	Low	<0.001
Volatility (mean)	Low	Low	High	Low	High	<0.001
Margin (mean)	Low	High				<0.001
Number of yards (mean)	3.67857	1.53684	1	1	1.59677	<0.001
ADI (mean)	1.02403	1.4465	12	1.03636	2.37237	<0.001

Table 7. Illustrative matching of SCM planning practices with SKU clusters.

Cluster	key feature	Recommended strategy from literature	References
Cluster 0	Low volume, Low volatility and Low margin	Lean towards efficiency/economy of scale	Simchi-Levi and Timmermans (2021)
Cluster 1	Low volume, Low volatility and High margin	Lean towards responsiveness	Ciechańska and Szwed (2020), James and John (2010), Simchi-Levi and Timmermans (2021)
Cluster 2	Low volume, High volatility and High ADI	MTO	Aitken et al. (2003), Kharlamov et al. (2013, 2020), Ramdas (2003), Simchi-Levi and Timmermans (2021), Tenhiälä and Ketokivi (2012)
Cluster 3	High volume, Low volatility and Single yard	MTS	Aitken et al. (2003), Ciechańska and Szwed (2020), Fisher and Raman (1996), Kharlamov et al. (2013, 2020), Simchi-Levi and Timmermans (2021), Syntetos et al. (2016)
Cluster 4	Low volume, High volatility and High ADI	Complexity Reduction	Schürmann et al. (2012), Simchi-Levi (2010), Simchi-Levi et al. (2013) Simchi-Levi and Timmermans (2021)

The volatility of sales can be said as inversely related to forecast accuracy and it is directly related to SKU inventory holding risk and as sales volatility becomes higher there will be a chance of frequent stock outs and the service level will be low which affect the customer satisfaction. On the other hand, the profit margin is directly proportional to risk of holding inventory. The reason behind this is that a single order will create a bigger impact on the bottom line or baseline. In case of volume, it is inversely proportional to the risk associated because as volume gets increased it will have the lower impact on any missed order (Simchi-Levi and Timmermans, 2021).

5. Discussion

Cluster 0

This cluster is characterized by its low volume as well as low volatility. Since the drivers are conflicting with each other the decision would be based upon third variable which is margin. Since the margin is low the risk associated with holding of the inventory would be low. The replenishment of these SKUs should be done in full truck loads from manufacturing locations to yards. Hence in this case the strategy needs to be leaned towards efficiency. This cluster has ADI of 1.02 which shows a stable demand (Ghobbar and Friend, 2002, 2003; Regattieri et al., 2005) and also it has Number of yards value equal 3.67 which is also high. The SKU's are going to many yards and hence we could position them at later stage in the supply chain (Simchi-Levi and Timmermans, 2021). In this sense we could position them at just before finishing to serve orders with minimum lead time i.e., at CRAP stage in the supply chain considering their characteristics. By positioning them at CRAP stage the manufacturing lead time could be reduced which will reflect into decrease in inventory at yards (Slivinskiy, 2005).

Cluster 1

This cluster is characterized by its low volume and as well as low volatility. Since both the drivers are conflicting with each other the decision would be based upon third variable which is margin since the margin is high it is better to replenish these SKU's according to actual sales because risk associated with holding of the inventory would be high and hence the availability of these SKUs at a proper yard will be the key factor (Simchi-Levi and Timmermans, 2021). This cluster has ADI of 1.44 which shows a stable demand and also it has Number of yards value equal 1.53 which is lower than cluster 0. The availability of these SKUs at a proper yard will be the key factor. Here we can use continuous replenishment strategy that we are following in a current state (James and John, 2010). In majority of circumstances this methodology allow consumer needs to be met but this will come at an expense of keeping high stock levels and more often release of production orders with smaller individual sizes (Ciechańska and Szwed, 2020). In terms of demand volatility, there may be advantages to manage inventory at a higher aggregated level (i.e., at parent level). Managing inventory at this level rather than final product parameters (finished SKU level) may result in the conversion of certain more volatile SKUs to less volatile SKUs (Slivinskiy, 2005). It is supported by Iocco (2009) for steel industry in which they concluded that for low volatile items when the holding risk is less the inventory can be allocated in downstream or nearer to the customer on the other hand when the holding risk is high it is allocated in upstream echelons. Single echelon models do not perform as well as stand-alone multi-echelon methods (Iocco, 2009). In order to determine exactly at what echelon as shown in Figure 5 and Figure 6 the inventory should be positioned the simulation model can be run for the clusters 0 and 1.

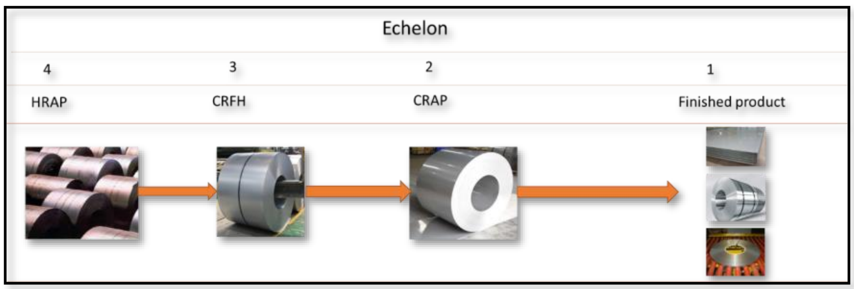

Figure 5. Four-echelon system.

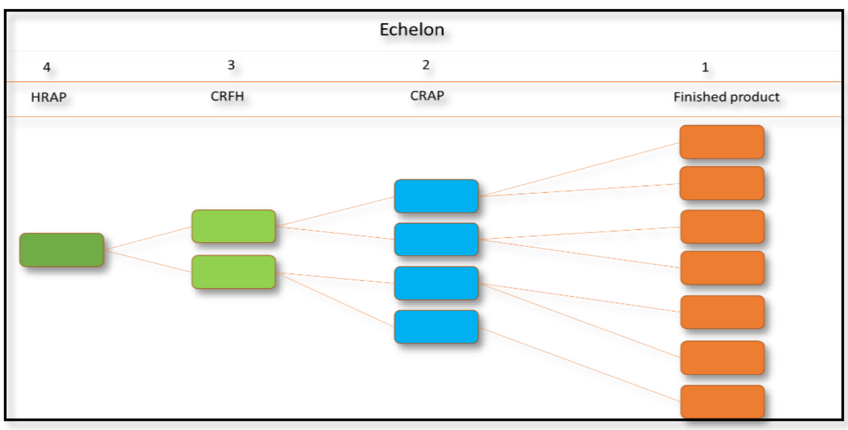

Figure 6. Four-echelon system detail. each echelon shows its SKUs.

Cluster 2

This cluster is characterized by high volatility, low demand and highest ADI. Along with that these SKUs are serving less numbers of customers as well as destinations. These are the SKU's ideal under MTO strategy (Aitken et al., 2003; Kharlamov et al., 2013, 2020; Ramdas, 2003; Simchi-Levi and Timmermans, 2021; Tenhiälä and Ketokivi, 2012). If we apply the make to stock system to cluster 2 and 4 then we will need to carry excess amount of inventory in order to counter the high volatility and it has seen that in most of the cases the carrying cost of inventory is high on these types of SKUs. Hence these SKUs should be manufactured on make to order strategy. The make to stock can be a good strategy for these items only in case if the market requirements are of a high service level and the key point here will be to set the inventory level which should be mostly driven by experience and data (Slivinskiy, 2005).

Cluster 3

This cluster is characterized by high volume, low volatility and these SKUs are serving only one yard. The forecast accuracy is high for these SKUs. Here the focus should be on operations as well as transportation efficiency. The inventory should be replenished based on a fixed schedule and in full truck loads to save on ordering cost and logistics cost (Aitken et al., 2003; Fisher and Raman, 1996; Kharlamov et al., 2013, 2020; Syntetos et al., 2016). These SKUs should be stored at regional warehouses (in our case yards). There is no need to keep the inventory of these SKUs at plant level rather it should be replenished based on fixed schedule which will also save transportation cost (Simchi-Levi and Timmermans, 2021). MTS model will not necessitate the regular production orders to replenish the stock (Ciechańska and Szwed, 2020) and hence the reduction in ordering cost can be achieved.

Cluster 4

This cluster is characterized by low volume and high volatility. Here we need to take insights from average inter-demand interval and inputs by discussions with the company team members. After discussion with company team members and carefully analysing these SKUs in this cluster we found that most of the SKU's have their inventory norms too less. Thus, facing challenges in planning. They are replenished in quantity greater than required “to order quantity” and since their consumption rate is low it takes almost 2 to 3 months to empty the inventory at yard and hence the ADI is high. In order to tackle these issues company can act on pricing strategy (Knizek et al., 2016; Schürmann et al., 2012) , complexity reduction (Simchi-Levi, 2010) or can try to increase the norms by serving them from a central warehouse which is in our case single manufacturing site. This will help XYZ to aggregate the demand and hence increase in norms (Simchi-Levi and Timmermans, 2021). For this analysis need to be done in order to get the overall cost of supply chain of serving these SKUs.

Cluster 2 and 3, cluster 3 and 4 are fail to reject the null hypothesis on the feature Price which indicates that the XYZ is following the almost same pricing strategies for these clusters irrespective of their opposite nature of characteristics. Here it will be beneficial for XYZ to do a market analysis to gain the advantage for these SKUs.

With all these insights this research address *RQ*3 that there is a difference between the current execution and inventory allocation strategy and the strategies recommended by the literature.

6. Conclusion

The aim of this study was to identify whether there exist different segments of SKUs in order to check for current execution strategy suitability. However, it was unclear how to do it eventually it was data which revealed that there exist different clusters and features are driving each cluster in order to drive that segment. This insight prompted to identify the strategies for each segment. At the same time equally, significant it was data only who indicated the importance or contribution of features in a given factor. This study adds to the knowledge and implementation of segmentation of SKUs at XYZ by taking insights from the previous studies. It revealed that the current “one size fit all” strategy is not appropriate and each cluster should be dealt with its own exclusive strategy to serve the customer in a better way (by having proper inventory availability at right location, at right time and in right quantity). In such a way the groups of SKUs with specific requirements example low cost, low lead time, high responsiveness can be dealt with capabilities for example aggregation, transportation in full truck loads which can properly satisfy their needs.

Until now the criteria selection and also the boundary conditions were set arbitrarily and intuitively but now it can be done via data driven and statistical insights. We used many features in this study but the number of yards and ADI had very little mentioned in the previous studies but through this study we identify the significance of these features. One of the main features of this research is its wide applicability, flexibility and scalability. The suggested approach may be utilised as a frequent diagnosing assessment model to regularly re-evaluate identified groups with the latest consumption history, which will help to address dynamic supply and demand issues.

First it is important to understand the present SKU portfolio their characteristics and the correct execution strategy according to their characteristics. Now once we identified that these SKU's need to keep close to the customer or not, we can proceed to identify exactly in which echelon they should be placed for which the simulation should be carried out which can be taken as a future scope of this study. Thus, combining this data driven approach with the simulation approach which is nothing but a type of prescriptive analytics will lead to a complete supply chain solution. This study focuses on a first part and the latter part can be built over the insights from first one in the future.

7. Managerial Implications

The findings of this study hold crucial managerial implications for SKU segmentation and execution strategy. It is evident that adopting a "one size fits all" approach is not suitable for the company's diverse range of SKUs. Instead, leveraging data-driven insights for segmentation can lead to more effective strategies tailored to the specific needs of each SKU cluster. By identifying distinct clusters, managers can implement exclusive strategies that ensure better customer service through optimized inventory availability, location, timing, and quantities. For example, addressing the unique requirements of SKU groups with specific characteristics, such as low cost, low lead time, or high responsiveness, can be achieved by deploying appropriate capabilities like aggregation and full truck load transportation. Moreover, the reliance on data and statistical insights for criteria selection and setting boundary conditions offers a more objective and informed decision-making process. This approach can significantly enhance operational efficiency and customer satisfaction by meeting the diverse demands of different SKU segments effectively.

8. Future Scope

There are several limitations on this research. To quantify the effects on costs and efficiency and to compare them to the current situation, more refinement is needed. The model's capabilities are limited by the unavailability of some data, which is caused by inadequate tracking. While our study primarily focuses on inventories, there is potential to expand the focus to other supply chain components like procurement and customer policies, ideally with the use of global supply chain data for deeper insights. Another way to improve is to investigate different machine learning techniques for evaluating cluster outcomes. Additionally, extending the research beyond the stainless steel industry to the service sector can broaden the study's focus and improve its scientific validity.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgments

This study did not receive any particular support/funding from public, commercial, or non-profit funding bodies. The authors respectfully thank the editor and reviewers for their time and insightful suggestions.

References

Aitken, J., Childerhouse, P., & Towill, D. (2003). The impact of product life cycle on supply chain strategy. *International Journal of Production Economics*, *85*(2), 127-140. https://doi.org/10.1016/S0925-5273(03)00105-1.

Alicke, K., & Forsting, M. (2017). Supply chain segmentation scientific frameworks. In: Protopappa-Sieke, M., Thonemann, U.W. (eds) *Supply Chain Segmentation.* Springer, Cham, pp. 5-13. https://doi.org/10.1007/978-3-319-54133-4_2.

Altintas, N., & Trick, M. (2014). A data mining approach to forecast behavior. *Annals of Operations Research*, *216*(1), 3-22. https://doi.org/10.1007/s10479-012-1236-9.

Amemiya, N., Terada, R., & Asahi, Y. (2018). Characteristic analysis of each store in Japanese hair salon. In: Yamamoto, S., Mori, H. (eds) *Human Interface and the Management of Information. Information in Applications and Services*. Springer, Cham, pp. 12-30. https://doi.org/10.1007/978-3-319-92046-7_2.

Bandyopadhyay, S., Thakur, S.S., & Mandal, J.K. (2021). Product recommendation for e-commerce business by applying principal component analysis (PCA) and K-means clustering: Benefit for the society. *Innovations in Systems and Software Engineering*, *17*(1), 45-52. https://doi.org/10.1007/s11334-020-00372-5.

Berthold, M., & Hand, D.J. (2003). *Intelligent data analysis: An introduction*. Springer.

Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford University Press.

Boylan, J.E., Syntetos, A.A., & Karakostas, G.C. (2008). Classification for forecasting and stock control: A case study. *Journal of the Operational Research Society*, *59*(4), 473-481. https://doi.org/10.1057/palgrave.jors.2602312.

Cagliano, A.C., Mangano, G., Rafele, C., & Grimaldi, S. (2022). Classifying healthcare warehouses according to their performance. A cluster analysis-based approach. *International Journal of Logistics Management*, *33*(1), 311-338. https://doi.org/10.1108/ijlm-02-2020-0110.

Cao, Z., Huo, B., Li, Y., & Zhao, X. (2015). The impact of organizational culture on supply chain integration: A contingency and configuration approach. *Supply Chain Management*, *20*(1), 24-41. https://doi.org/10.1108/scm-11-2013-0426.

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., & Towill, D.R. (2002). Developing market specific supply chain strategies. *The International Journal of Logistics Management*, *13*(1), 1-14. https://doi.org/10.1108/09574090210806324.

Christopher, M., Towill, D.R., Aitken, J., & Childerhouse, P. (2009). Value stream classification. *Journal of Manufacturing Technology Management*, *20*(4), 460-474. https://doi.org/10.1108/17410380910953720.

Ciechańska, O., & Szwed, C. (2020). Characteristics and study of make-to-stock and make-to-availability production strategy using simulation modelling. *Management and Production Engineering Review, 11*(4), 68-80. https://journals.pan.pl/dlibra/publication/136121/edition/119032.

Ciric Lalic, D., Lalic, B., Delić, M., Gracanin, D., & Stefanovic, D. (2022). How project management approach impact project success? From traditional to agile. *International Journal of Managing Projects in Business*, *15*(3), 494-521. https://doi.org/10.1108/ijmpb-04-2021-0108.

Clarke, A.H. (2009). Bridging industrial segmentation theory and practice. *Journal of Business-to-Business Marketing*, *16*(4), 343-373. https://doi.org/10.1080/10517120902762427.

Dederichs, M., Fahmy, M.D., An, H., Guentsch, A., Viebranz, S., & Kuepper, H. (2021). Comparison of wear resistance of prefabricated composite veneers versus ceramic and enamel. *Journal of Prosthodontics*, *30*(8), 711-719. https://doi.org/10.1111/jopr.13303.

Denton, B., Gupta, D., & Jawahir, K. (2003). Managing increasing product variety at integrated steel mills. *Interfaces*, *33*(2), 41-53. https://doi.org/10.1287/inte.33.2.41.16373.

Dikmen, I., Birgonul, M.T., & Budayan, C. (2009). Strategic group analysis in the construction industry. *Journal of Construction Engineering and Management*, *135*(4), 288-297. https://doi.org/10.1061/(asce)0733-9364(2009)135:4(288).

Fisher, M. (1997). What is the right supply chain for your product? *Harvard Business Review*. https://hbr.org/1997/03/what-is-the-right-supply-chain-for-your-product.

Fisher, M., & Raman, A. (1996). Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research*, *44*(1), 87-99. https://doi.org/10.1287/opre.44.1.87.

Flynn, B.B., Huo, B., & Zhao, X. (2010). The impact of supply chain integration on performance: A contingency and configuration approach. *Journal of Operations Management*, *28*(1), 58-71. https://doi.org/10.1016/j.jom.2009.06.001.

Frohlich, M.T., & Westbrook, R. (2001). Arcs of integration: An international study of supply chain strategies. *Journal of Operations Management*, *19*(2), 185-200. https://doi.org/10.1016/S0272-6963(00)00055-3.

Fuller, J., O'Conor, J., & Rawlinson, R. (1993). Tailored logistics: The next advantage. *Harvard Business Review*. https://hbr.org/1993/05/tailored-logistics-the-next-advantage.

Ghobbar, A.A., & Friend, C.H. (2002). Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, *8*(4), 221-231. https://doi.org/10.1016/S0969-6997(01)00054-0.

Ghobbar, A.A., & Friend, C.H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: A predictive model. *Computers & Operations Research*, *30*(14), 2097-2114. https://doi.org/10.1016/S0305-0548(02)00125-9.

Godsell, J., Diefenbach, T., Clemmow, C., Towill, D., & Christopher, M. (2011). Enabling supply chain segmentation through demand profiling. *International Journal of Physical Distribution & Logistics Management*, *41*(3), 296-314. https://doi.org/10.1108/09600031111123804.

Godsell, J., Kharlamov, A.A., Vasishta, S., & Burdett, J. (2013). Challenging the volume-variability paradigm: An empirical investigation in a global beverage company. In *Proceedings - 2013 20th International Annual EurOMA Conference*. Dublin, Ireland. https://wrap.warwick.ac.uk/75404/.

Gosling, L., & Urrutia, J. (2019). *Analytics driving supply chain segmentation for Lenovo*. https://dspace.mit.edu/handle/1721.1/121297.

Green, P.E. (1977). A new approach to market segmentation. *Business Horizons*, *20*(1), 61-73. https://doi.org/10.1016/0007-6813(77)90088-X.

Han, J., Kamber, M., Pei, J. (2011). *Data mining: Concepts and techniques*. Elsevier, USA.

Hill, T. (2000). *Manufacturing strategy: Text and cases*. Irwin/McGraw-Hill. ISBN 9780256230727(p), 9780256230722(e).

Holweg, M. (2005). The three dimensions of responsiveness. *International Journal of Operations & Production Management*, *25*(7), 603-622. https://doi.org/10.1108/01443570510605063.

Iocco, J.D. (2009). *Multi-echelon multi-product inventory strategy in a steel company*. Thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/53049.

Jackson, I., Avdeikins, A., & Tolujevs, J. (2019). Unsupervised learning-based stock keeping units segmentation. In: Kabashkin, I., Yatskiv, (Jackiva), I., Prentkovskis, O. (eds) *Reliability and Statistics in Transportation and Communication* (vol. 68, pp. 603-612). Springer, Cham. https://doi.org/10.1007/978-3-030-12450-2_58.

Jain, A.K., Duin, R.P.W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4-37. https://doi.org/10.1109/34.824819.

Jain, A.K., Murty, M.N., & Flynn, P.J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264-323. https://doi.org/10.1145/331499.331504.

James, F.C.III, & John, G.S. (2010). *Theory of constraints handbook, 1st Edition*. McGraw-Hill Education, New York. ISBN: 9780071665544. https://www.accessengineeringlibrary.com/content/book/9780071665544.

Jolliffe, I. (2011). Principal component analysis. In: Lovric, M. (ed) *International Encyclopedia of Statistical Science*. Springer, Berlin, Heidelberg, pp. 1094-1096. https://doi.org/10.1007/978-3-642-04898-2_455.

Jun, S.H. (2006). An optimal clustering using hybrid self organizing map. *International Journal of Fuzzy Logic and Intelligent Systems*, *6*(1), 10-14. https://doi.org/10.5391/ijfis.2006.6.1.010.

Jun, S.H., & Lee, S.J. (2010). Empirical comparisons of clustering algorithms using Silhouette information. *International Journal of Fuzzy Logic and Intelligent Systems*, *10*(1), 31-36. https://doi.org/10.5391/ijfis.2010.10.1.031.

Kalagnanam, J.R., Dawande, M.W., Trumbo, M., & Lee, H.S. (2000). The surplus inventory matching problem in the process industry. *Operations Research*, *48*(4), 505-516. https://doi.org/10.1287/opre.48.4.505.12425.

Knizek, Ganeriwalla, Schuler, Wolfgang, Akins, Frost, & lee. (2016). Metals manufacturing: Four rules for managing inventory better. BCG.com. Retrieved January 16, 2024, from https://www.bcg.com/publications/2016/supply-chain-management-metals-manufacturing-four-rules-managing-inventory-better.

Kaya, M.F., & Schoop, M. (2022). Analytical comparison of clustering techniques for the recognition of communication patterns. *Group Decision and Negotiation*, *31*(3), 555-589. https://doi.org/10.1007/s10726-021-09758-7.

Kazbare, L., van Trijp, H.C.M., & Eskildsen, J.K. (2010). A-priori and post-hoc segmentation in the design of healthy eating campaigns. *Journal of Marketing Communications*, *16*(1-2), 21-45. https://doi.org/10.1080/13527260903342712.

Kharlamov, A.A., Ferreira, L.M.D.F., & Godsell, J. (2013). The power of analytical approaches towards the development of differentiated supply chain strategies: Case study. In: Azevedo, A. (ed) A*dvances in Sustainable and Competitive Manufacturing Systems*, *Lecture Notes in Mechanical Engineering.* Springer, Heidelberg, pp. 1223-1235. https://doi.org/10.1007/978-3-319-00557-7_100.

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.

Lamming, R., Johnsen, T., Zheng, J., & Harland, C. (2000). An initial classification of supply networks. *International Journal of Operations & Production Management*, *20*(6), 675-691. https://doi.org/10.1108/01443570010321667.

Lee, H.L. (2002). Aligning supply chain strategies with product uncertainties. *California Management Review*, *44*(3), 105-119. https://doi.org/10.2307/41166135.

Lee, H.L. (2004). The triple-a supply chain. *Harvard Business Review*. https://hbr.org/2004/10/the-triple-a-supply-chain.

Lee, S., & Lee, D.K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology*, *71*(5), 353-360. https://doi.org/10.4097/kja.d.18.00242.

Lee, S., Lim, J., Lee, S., Heo, Y., & Jung, D. (2022). Group-tailored feedback on online mental health screening for university students: Using cluster analysis. *BMC Primary Care*, *23*(1), 19. https://doi.org/10.1186/s12875-021-01622-6.

Li, D., & O'Brien, C. (2001). A quantitative analysis of relationships between product types and supply chain strategies. *International Journal of Production Economics*, *73*(1), 29-39.

Lix, L.M., Keselman, J.C., & Keselman, H.J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "F" test. *Review of Educational Research*, *66*(4), 579-619. https://doi.org/10.3102/00346543066004579.

Martin hristopher, C., & Towill, D.R. (2000). Supply chain migration from lean and functional to agile and customised. *Supply Chain Management*, *5*(4), 206-213. https://doi.org/10.1108/13598540010347334.

Martínez-López, F.J., & Casillas, J. (2009). Marketing intelligent systems for consumer behaviour modelling by a descriptive induction approach based on genetic Fuzzy systems. *Industrial Marketing Management*, *38*(7), 714-731. https://doi.org/10.1016/j.indmarman.2008.02.003.

Milligan, G.W. (1996). Clustering validation: Results and implications for applied analyses. In: Arabie, P., Hubert, L.J., & De Soete, G. (eds) *Clustering and Classification*. World Scientific, pp. 341-375. https://doi.org/10.1142/9789812832153_0010.

Milligan, G.W., & Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*(2), 159-179. https://doi.org/10.1007/bf02294245.

Nag, B., Han, C., & Yao, D.Q. (2014). Mapping supply chain strategy: An industry analysis. *Journal of Manufacturing Technology Management*, *25*(3), 351-370. https://doi.org/10.1108/jmtm-06-2012-0062.

Ni, M., Xu, X., & Deng, S. (2007). Extended QFD and data-mining-based methods for supplier selection in mass customization. *International Journal of Computer Integrated Manufacturing*, *20*(2-3), 280-291. https://doi.org/10.1080/09511920601150651.

Park, M.J., Jun, S.H., & Oh, K.W. (2003). Determination of optimal cluster size using bootstrap and genetic algorithm. *Journal of the Korean Institute of Intelligent Systems*, *13*(1), 12-17. https://doi.org/10.5391/jkiis.2003.13.1.012.

Payne, A., & Frow, P. (2006). Customer relationship management: From strategy to implementation. *Journal of Marketing Management*, *22*(1-2), 135-168. https://doi.org/10.1362/026725706776022272.

Payne, T., & Peters, M.J. (2004). What is the right supply chain for your products? *The International Journal of Logistics Management*, *15*(2), 77-92. https://doi.org/10.1108/09574090410700310.

Protopappa-Sieke, M., & Thonemann, U.W. (2017). *Supply chain segmentation: Best-in-class cases, practical insights and foundations*. Springer. ISBN: 978-3-319-54132-7(p), 978-3-319-54133-4(e). https://doi.org/10.1007/978-3-319-54133-4.

Ptok, F.L., & Camargo Henao, J.E. (2021). *Supply chain segmentation in the apparel industry*. https://dspace.mit.edu/handle/1721.1/130980.

Qi, Y., Boyer, K.K., & Zhao, X. (2009). Supply chain strategy, product characteristics, and performance impact: Evidence from Chinese manufacturers. *Decision Sciences*, *40*(4), 667-695. https://doi.org/10.1111/j.1540-5915.2009.00246.x.

Qi, Y., Zhao, X., & Sheu, C. (2011). The impact of competitive strategy and supply chain strategy on business performance: The role of environmental uncertainty. *Decision Sciences*, *42*(2), 371-389. https://doi.org/10.1111/j.1540-5915.2011.00315.x.

Rahim, M.S., Nguyen, K.A., Stewart, R.A., Ahmed, T., Giurco, D., & Blumenstein, M. (2021). A clustering solution for analyzing residential water consumption patterns. *Knowledge-Based Systems*, *233*, 107522. https://doi.org/10.1016/j.knosys.2021.107522.

Ramdas, K. (2003). Managing product variety: An integrative review and research directions. *Production and Operations Management*, *12*(1), 79-101. https://doi.org/10.1111/j.1937-5956.2003.tb00199.x.

Regattieri, A., Gamberi, M., Gamberini, R., & Manzini, R. (2005). Managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management*, *11*(6), 426-431. https://doi.org/10.1016/j.jairtraman.2005.06.003.

Revilla, E., Sáenz, M.J., & Knoppen, D. (2013). Towards an empirical typology of buyer–supplier relationships based on absorptive capacity. *International Journal of Production Research*, *51*(10), 2935-2951.

Rezaei, J., van Roekel, W.S., & Tavasszy, L. (2018). Measuring the relative importance of the logistics performance index indicators using best worst method. *Transport Policy*, *68*, 158-169.

Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65. https://doi.org/10.1016/0377-0427(87)90125-7.

Saayman, M., Saayman, A., & Joubert, E.M. (2012). Expenditure-based segmentation of visitors to the wacky wine festival. *Tourism Recreation Research*, *37*(3), 215-225. https://doi.org/10.1080/02508281.2012.11081710.

Sanders, N.R. (2016). How to use big data to drive your supply chain. *California Management Review*, *58*(3), 26-48. https://doi.org/10.1525/cmr.2016.58.3.26.

Schnetzler, M.J., Sennheiser, A., & Schönsleben, P. (2007). A decomposition-based approach for the development of a supply chain strategy. *International Journal of Production Economics*, *105*(1), 21-42. https://doi.org/10.1016/j.ijpe.2006.02.004.

Schürmann, Izaret, Schuler, & Koppitz. (2012). Long-tail pricing in business-to-business markets. In BCG . BCG.com. Retrieved January 16, 2024, from https://www.bcg.com/publications/2012/pricing-go-to-market-strategy-long-tail-pricing-business-to-business-markets.

Selldin, E., & Olhager, J. (2007). Linking products with supply chains: Testing Fisher's model. *Supply Chain Management*, *12*(1), 42-51. https://doi.org/10.1108/13598540710724392.

Shewchuk, J.P. (1998). Agile manufacturing: One size does not fit all. In: Bititci, U.S., Carrie, A.S. (eds) *Strategic Management of the Manufacturing Value Chain*. Springer, Boston, MA. (vol. 2, pp. 143-150). https://doi.org/10.1007/978-0-387-35321-0_16.

Shi, X., Wang, W., & Zhang, C. (2017). An empirical comparison of latest data clustering algorithms with state-of-the-art. *Indonesian Journal of Electrical Engineering and Computer Science*, *5*(2), 410-415. https://doi.org/10.11591/ijeecs.v5.i2.

Simchi-Levi, D. (2010). *Operations rules: Delivering customer value through flexible operations*. The MIT Press, London, England.

Simchi-Levi, D. (2014). OM forum-OM research: From problem-driven to data-driven research. *Manufacturing & Service Operations Management*, *16*(1), 2-10. https://doi.org/10.1287/msom.2013.0471.

Simchi-Levi, D., Clayton, A., & Raven, B. (2013). When one size does not fit all. *MIT Sloan Management Review*, *54*(2), 15-17.

Simchi-Levi, D., & Timmermans, K. (2021). A simpler way to modernize your supply chain. *Harvard Business Review*. https://hbr.org/2021/09/a-simpler-way-to-modernize-your-supply-chain.

Slivinskiy, A.L. (2005). *Delivery optimization for a make to order custom wire mill*. Thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/34865.

Song, Z., & Kusiak, A. (2009). Optimising product configurations with a data-mining approach. *International Journal of Production Research*, *47*(7), 1733-1751. https://doi.org/10.1080/00207540701644235.

Sudol, P.E., Ochoa, G.S., Cain, C.N., & Synovec, R.E. (2022). Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. *Analytica Chimica Acta*, *1209*, 339847. https://doi.org/10.1016/j.aca.2022.339847.

Syntetos, A.A., Babai, Z., Boylan, J.E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, *252*(1), 1-26. https://doi.org/10.1016/j.ejor.2015.11.010.

Tang, L., Meng, Y., Wang, G., Chen, Z.L., Liu, J., Hu, G., Chen, L., & Zhang, B. (2014). Operations research transforms Baosteel's operations. *Interfaces*, *44*(1), 22-38. https://doi.org/10.1287/inte.2013.0719.

Tenhiälä, A., & Ketokivi, M. (2012). Order management in the customization-responsiveness squeeze. *Decision Sciences*, *43*(1), 173-206. https://doi.org/10.1111/j.1540-5915.2011.00342.x.

Theodoridis, S., & Koutroumbas, K. (2009). Chapter 16-cluster validity. In: Theodoridis, S., Koutroumbas, K. (eds) *Pattern Recognition (Fourth Edition)*. Academic Press, Boston. pp. 863-913. https://doi.org/10.1016/b978-1-59749-272-0.50018-9.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *63*(2), 411-423. https://doi.org/10.1111/1467-9868.00293.

Ting, S.L., Tse, Y.K., Ho, G.T.S., Chung, S.H., & Pang, G. (2014). Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry. *International Journal of Production Economics*, *152*, 200-209. https://doi.org/10.1016/j.ijpe.2013.12.010.

van Kampen, T.J., Akkerman, R., & Pieter van Donk, D. (2012). SKU classification: A literature review and conceptual framework. *International Journal of Operations & Production Management*, *32*(7), 850-876. https://doi.org/10.1108/01443571211250112.

Vitasek, K.L., Manrodt, K.B., & Kelly, M. (2003). Solving the supply. *Supply Chain Management Review*, *7*, 58-64.

Wiersma, S., Just, T., & Heinrich, M. (2021). Segmenting German housing markets using principal component and cluster analyses. *International Journal of Housing Markets and Analysis*, *15*(3), 548-578. https://doi.org/10.1108/ijhma-01-2021-0006.

Xu, R., & Wunsch II, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, *16*(3), 645-678. https://doi.org/10.1109/tnn.2005.845141.

Zaky, N., Ahmed, M.Z., Alarjani, A., & Attia, E.A. (2023). Lean manufacturing implementation in iron and steel industries: effect of wastes management on the production costs. *Journal of Engineering, Design and Technology*. https://doi.org/10.1108/jedt-01-2023-001. (In press).

Zancanaro, M., Kuflik, T., Boger, Z., Goren-Bar, D., & Goldwasser, D. (2007). Analyzing museum visitors' behavior patterns. In: Conati, C., McCoy, K., & Paliouras, G. (eds.), *User Modeling 2007* (Vol. 4511, pp. 238-246). Springer. https://doi.org/10.1007/978-3-540-73078-1_27.

Zhao, C., Johnsson, M., & He, M. (2017). Data mining with clustering algorithms to reduce packaging costs: A case study. *Packaging Technology and Science*, *30*(5), 173-193. https://doi.org/10.1002/pts.2286.

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.