LAPPEENRANNAN-LAHDEN TEKNILLINEN YLIOPISTO LUT
School of Engineering Science
Industrial Engineering and Management
Business Analytics

Customer segmentation based on demand profiling

Master's Thesis

Arttu Pietilä

ABSTRACT

Author: Arttu Pietilä **Subject:** Customer segmentation based on demand profiling	
Year: 2021	**Place:** Espoo, Finland
Master's thesis. Lappeenranta-Lahti University of Technology, School of Engineering Science, Industrial Engineering and Management. 79 pages, 31 figures, 10 tables and 4 appendices Examiners: Christoph Lohrmann and Pasi Luukka	
Keywords: Supply chain segmentation, supply chain strategies, customer segmentation, demand profiling, clustering, supply chain performance, service models	

In the context of supply chain strategies, it is commonly stated that a one-size-fits-all supply chain is not enough in modern markets. The different customer requirements as well as the inherent characteristics of products impose certain requirements towards the supply chain and hence, customers cannot be served efficiently by a single supply chain strategy. Therefore, utilizing several separate supply chain strategies simultaneously enables meeting customer expectations while still being cost-efficient.

The aim of this thesis was to identify what kind of supply chain strategies are presented in the literature, and what characteristics influence the emergence of these strategies. Through an understanding of the literature, a clustering analysis was conducted to identify what kind of customer segments can be found in the company's data, and what kind of supply chain strategies can be used to meet these requirements.

Previous research related to supply chain strategies supports the four different customer segments identified in the clustering analysis. Three of the identified customer segments corresponded well to the known supply chain strategies, while one segment had clear features of two separate well-known supply chain strategies.

TIIVISTELMÄ

Tekijä: Arttu Pietilä

Työn nimi: Kysynnän profilointiin perustuva asiakassegmentointi

Vuosi: 2021 **Paikka:** Espoo, Suomi

Diplomityö. Lappeenrannan-Lahden teknillinen yliopisto, LUT School of Engineering Science, Tuotantotalous.

79 sivua, 31 kuvaa, 10 taulukkoa ja 4 liitettä

Tarkastaja(t): Christoph Lohrmann ja Pasi Luukka

Hakusanat: Toimitusketjun segmentointi, asiakassegmentointi, toimitusketjustrategia, kysynnän profilointi, klusterointi, toimitusketjun suorituskyky, palvelumallit, analyyttiset menetelmät

Toimitusketjustrategioiden yhteydessä on useaan otteeseen todettu, ettei yksi malli sovi kaikkeen kysyntään. Asiakkaiden erilaiset vaatimukset sekä tuotteiden luontaiset ominaisuudet asettavat tiettyjä vaatimuksia toimitusketjulle, joita ei yhdellä toimitusketjustrategialla kyetä täyttämään tehokkaasti. Tämän vuoksi usean erillisen toimitusketjustrategian hyödyntäminen samanaikaisesti mahdollistaa erilaisiin tarpeisiin vastaamisen kustannustehokkaasti.

Tämän diplomityön tavoitteena oli tunnistaa minkälaisia toimitusketjustrategioita kirjallisuudessa esitetään sekä mitkä ominaisuudet vaikuttavat näiden strategioiden syntyyn. Tutkielma koostuu kahdesta erillisestä osa-alueesta: kirjallisuuskattauksesta, joka luo kattavan teoriapohjan toimitusketjustategioiden ja klusterointimenetelmien ympärille, sekä klusterointianalyysistä, jossa teoriaviitekehyksiä sovelletaan käytännössä.

Klusterointianalyysissä tunnistetut neljä erilaista asiakassegmenttiä saavat vahvaa tukea aikaisemmin tehdyltä toimitusketjustrategioihin liittyvältä tutkimukselta. Kolme tunnistetuista asiakassegmenteistä vastasi erittäin hyvin tunnettuja toimitusketjustrategioita, kun taas yhdessä segmentissä oli selkeitä piirteitä kahdesta yleisesti tunnetusta toimitusketjustrategiasta.

ACKNOWLEDGEMENTS

Finishing this thesis means two things for me: on the one hand it means closing the door for student life behind me, but on the other hand it opens many doors in front of me. I can without any hesitation say that these years in LUT university has been the most rewarding time of my life with all the new learnings, friends, and experiences. This said, I would like to thank each and every one of you who took part in making my years in Lappeenranta as memorable as they were.

Regarding this thesis, I would like to express my gratitude for Christoph Lohrmann who really took the time and effort to share his knowledge and advise me among this project. Without the flexibility you offered, the thesis could have taken longer than it now took. Furthermore, I want to thank my supervisor Anu Metsäranta for supporting me throughout the process with excellent comments and observations. Thanks to her extensive expertise, I was able to consider perspectives that would otherwise have been overlooked.

Finally, I would like to thank my girlfriend Karoliina, who has supported me throughout these years. I can't underline enough how grateful I am to have her always being there for me. Especially in the midst of the hustle and bustle of recent months, the importance of the support of hers has risen to its own value.

Espoo, 17.06.2021

Arttu Pietilä

TABLE OF CONTENTS

1 INTRODUCTION 4

1.1 Background 4

1.2 Objectives and scope 5

1.3 Structure of the thesis 7

2 SUPPLY CHAIN SEGMENTATION 8

2.1 Lean versus agile supply chain 11

2.2 Further supply chain segmentation 17

2.2.1 Multiple variable framework 19

2.2.2 Buying behavior 23

3 CLUSTERING THEORY 27

3.1 K-means 29

3.2 Fuzzy c-means 30

3.3 Hierarchical clustering 32

3.4 Self-organizing maps 34

3.5 Customer segmentation using clustering analysis 35

3.6 Defining optimal number of clusters 36

4 CLUSTERING WITH A REAL DATASET 39

4.1 Data description 41

4.2 Data preprocessing 43

4.3 Testing different clustering algorithms 53

4.3.1 Initializing the clustering algorithms 54

4.3.2 Measuring the fit of the clusters 54

4.3.3 Obtain the results 59

5 CONCLUSIONS AND DISCUSSION 64

5.1 Answering the research questions ... 65

5.2 Future research ... 67

REFERENCES ... 69

APPENDIX ... 72

LIST OF FIGURES

Figure 1. Total value metric
Figure 2. Supply chain frontier
Figure 3. Matching products with supply chains
Figure 4. Market qualifiers and market winners with lean and agile supply
Figure 5. Supply chain segmentation based on demand predictability vs. lead-time.
Figure 6. How OW and MQ characteristics change as a function of the life cycle
Figure 7. The Pareto distribution
Figure 8. Primary customer behavior types
Figure 9. The four generic supply chain types
Figure 10. The updated five generic supply chain types
Figure 11. Demand types of different supply chains
Figure 12. Example of clustering a dataset
Figure 13. K-Means algorithm works iteratively in two phases
Figure 14. Hierarchical clustering dendrogram example
Figure 15. Illustration of agglomerative clustering using four data points.
Figure 16. The operational logic of the SOM algorithm
Figure 17. The main steps of the supply chain segmentation process used in this thesis.
Figure 18. Aggregating order-level data into customer-level
Figure 19. The pivotal attributes used in supply chain segmentation frameworks
Figure 20. Different phases of preprocessing the dataset
Figure 21. Distribution of lead-time before and after removing outliers
Figure 22. Correlation coefficients of selected variables
Figure 23. Pareto-like distributions of attributes
Figure 24. Pareto-like distributions of attributes after log-transformation
Figure 25. Attributes with other kind of distributions
Figure 26. Average Dunn index for tested clustering algorithms
Figure 27. Average Calinski-Harabasz index for tested clustering algorithms
Figure 28. Average Silhouette index for tested clustering algorithms
Figure 29. Silhouette scores for HC algorithm with different amount of clusters
Figure 30. Cluster dendogram

Figure 31. Distributions per cluster for each attribute

LIST OF TABLES

Table 1. Research questions and objectives

Table 2. Structure of the thesis

Table 3. Supply chain segmentation frameworks

Table 4. Distinguishing features of lean and agile supply

Table 5. Functional versus innovative products differences in demand

Table 6. Physically efficient versus market-responsive supply chains

Table 7. Some of the most used internal clustering validation measures

Table 8. The attributes selected for the analysis.

Table 9. Mean absolute correlations between selected attributes

Table 10. Defining the optimal clustering method

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
B2B	Business to Business
CO_2	Carbon dioxide
CVT	Customer Value Threshold
DVP	Demand Visibility Point
DWV3	Duration, Window, Volume, Variety, Variability
ERP	Enterprise Resource Planning
EU	European Union
FCM	Fuzzy C-Means
KM	K-Means
MRI	Minimum Reasonable Inventory
MTO	Made To Order
MW	Market Winner
MQ	Market Qualifier
OPP	Order Penetration Point
OQ	Order Qualifier
OQC	Order Qualifying Criteria
OW	Order Winner
OWC	Order Winning Criteria
POD-Time	Proof Of Delivery Time
R&D	Research and Development
RFM	Recency, Frequency, Monetary
RFL	Ready For Loading
RTA	Required Time of Arrival
SKU	Stock Keeping Unit
SOM	Self-Organizing Map
SSE	Sum of Squared Errors

1 INTRODUCTION

This chapter opens the background of the study and introduces the research subject in general to the reader. Furthermore, the research questions, objectives and scope of the study are presented. Finally, this chapter enlightens the research methodology used and the structure of the thesis before continuing into the actual research.

1.1 Background

For decades, companies have acknowledged that customers form different kind of groups based on their purchasing behavior, and thus the idea of market segmentation has slowly evolved (Godsell, Diefenbach, Clemmow, Towill & Christopher, 2010). The idea of recognizing customers with differing service requirements and forming separate groups based on some buying behavioral features, enables companies to answer the needs of a larger pool of customers more precisely. Even if the idea of market segmentation has matured among functions like product differentiation and marketing, the implications on supply chain strategies has not been generally recognized (Godsell, Harrison, Emberson and Storey, 2006). Since the research on supply chain strategy development based on supply chain segmentation is relatively young, no consensus has been reached on a single leading framework. For this reason, a variety of different frameworks are introduced focusing on both product characteristics as well as a more strategic approaches (Fisher 1999; Gattorna, 2009). This thesis aims to recognize customer segments by using demand profiling which pursues to capture the behavior from market segment considerations along with product characteristics.

This thesis is done for Stora Enso Oyj (later Stora Enso) Packaging Materials division. Stora Enso is a leading global provider of renewable solutions regarding packaging, biomaterials, wooden construction, and paper. Stora Enso's sales in 2019 was 10.1 billion euros and it had 26 000 employees in 30 different countries. The great majority of sales come from Europe with 73% share of total sales, 17 % comes from Asia Pacific area, 5% from Americas and the 5% from other areas. The headquarters of Stora Enso is in Helsinki Finland (Stora Enso 2019a, pp. 2-4).

The Packaging Materials division produces a wide range of different types of carton board suitable for packaging foods, liquids, pharmaceutical and luxury goods. Besides other paperboard manufacturers one of the biggest competitors for packaging materials division is the plastic packaging industry. In the past years EU legislation has worked as the trendsetter on replacing single-use plastics by renewable solutions to reach the EU's CO_2 targets. (Stora Enso 2019a, pp. 24-25) This trend has already and will also in the future support the growth of the renewable packaging materials business. Furthermore, this trend shows in Stora Enso's objectives: One of Stora Enso's focuses for the upcoming years is to grow the business created by new innovations and products (Stora Enso 2019a, pp. 27). Through innovation, it is possible to keep the product portfolio cyclical where the products in different life cycles attracts different types of customers. The variety of the products, and even further, the applications or so-called end uses of the products, has also a natural effect on the behavior of the customers. Whether the customer is a global company producing packaging for high volume markets such as food and drinks, or a local company producing packaging for a more niche market, has an effect not only to the order sizes, but also to features such as the predictability of demand, expectations of lead-time, and price sensitivity. From a customer service and supply chain perspective, the diversity of unique customer needs is a challenge, which is why Stora Enso is elaborating new service models to be able to serve the wide customer base with a more efficient manner. The service models can be considered as service bundles, where every bundle contains specific set of services and is priced accordingly. By reviewing supply chain segmentation literature and performing a data analysis respectively, we get insights on what kind of groups does the customers form, and even further, an indication of what kind of features should the service models be built on.

1.2 Objectives and scope

The objective of this thesis is to recognize the features that fits the best on modeling Stora Enso's supply chain strategies and to build a current state analysis of what kind of groups or segments do the customers form. The thesis is built on three research questions presented in **Table 1**.

Table 1. Research questions and objectives

Research question	Objective
1. What kind of frameworks are introduced in the literature regarding supply chain segmentation in business to business (B2B) markets?	To identify the underlying logic and the features being used in the introduced supply chain segmentation frameworks
2. What kind of supply chain strategies are best suited to Stora Enso's supply chain?	A proposal of the supply chain strategies that could be used in Stora Enso's cartonboard business
3. How do the identified segments based on data analysis correspond to the segments identified from the supply chain segmentation literature?	A current state analysis on what kind of customer profiles can be recognized and a discussion how these profiles match the expectations

The first question covers the whole first part of the thesis and provides the knowledge on how to approach supply chain segmentation problem. The second question aims to use the knowledge gained from the supply chain related literature and cartonboard industry in general to shape the supply chain strategies that could fit Stora Enso’s cartonboard business. The third question is formed to give understanding on how the theoretical frameworks reflects with the actual customer behavior.

The outcomes of the thesis will work as a data-based groundwork for defining the new service models for Packaging Materials division. As this thesis explores the actual customer characteristics regarding demand profiling, it will help to more fully understand different customer behavior models in the supply chain context. In addition, the findings of this thesis will give insight of the similarities and dissimilarities within customers and support growth together with customers by offering the best fitting service models for them.

This thesis will go through a variety of supply chain segmentation frameworks to give a more in-depth understanding of how the supply chain segmentation can be done and what kind of features are identified to have a meaningful effect on different supply chain strategies. This thesis will not cover all supply chain frameworks and clustering methods, but rather a set of commonly used frameworks and methods. Since many frameworks are introduced and the thesis is focused on identifying different customer profiles, the thesis will not evaluate the effects of the different strategies on the physical supply chain. Rather, the research is narrowed on discovering diverse supply chain segmenting paradigms, identifying the logic of the segmentation behind, and finally use this knowledge later in the clustering process.

1.3 Structure of the thesis

This thesis contains six main chapters that are divided further into sub-chapters. The content of the main chapters is presented in **Table 2**.

Table 2. The structure of the thesis

Chapter	Content
Chapter 1 - Introduction	Introducing the reader to the subjects, research questions, objectives, and methods as well as the structure of the thesis
Chapter 2 - Supply chain segmentation	Summarizing the research made around supply chain segmentation. Presenting the different theory frameworks, focusing on inputs and outputs.
Chapter 3 - Clustering theory	Presenting the theory behind advanced data analytic approaches on solving clustering problems. Introducing k-nearest neighbors (KNN), fuzzy c-means (FCM), hierarchical clustering (HC), and self-organizing map (SOM) clustering algorithms and their functionalities.
Chapter 4 - Clustering with a real dataset	Using the supply chain segmentation frameworks to choose Stora Enso specific supply chain segmentation features. Using the data preprocessing methods and the chosen clustering algorithms to build a real-life use case.
Chapter 5 – Conclusions and discussion	Discussing the validity and meaningfulness of the results in addition to proposing future research possibilities.

2 SUPPLY CHAIN SEGMENTATION

Competitive advantage is one of the key elements of building and maintaining a successful business. Competitive advantage means an advantage over competitors gained through acts of innovation. (Porter, M. E. 1990, p. 75) According to Porter, the innovation can be manifested across business functions such as new marketing approaches or redesigning production process.

In today's business, an effective supply chain plays an essential role in maintaining the competitiveness of firms, and as Christopher (2000) has pointed out "it is supply chains that compete not companies". Having the right products at the right place at the right time is directly reflected on customer satisfaction. An effective supply chain has its downside – high responsiveness makes it expensive to maintain. This results in a situation where it is natural to attempt to match the supply to the demand and in that way driving down the costs simultaneously with improving customer satisfaction. (Christopher, 2000; Christopher & Towill, 2001)

How customers perceive value can be difficult to explicitly quantify since customers may value different aspects. Whether the customer prioritizes fast deliveries, which are referred in supply chain context as short lead-times, or the level of service depends entirely on the customer. In a customer focused supply chain, many metrics can be used on calculating the total value of a supplied product to the final user, the end-user. However, the total value may be aggregated as a formula using four key metrics: quality, service, cost, and lead-time presented in **Figure 1** (Naylor, Naim & Berry, 1999). As a numerator, the formula has quality multiplied by service. The higher the product-related quality and service are, the more value the end-user receives. Respectively, the nominator of the formula consists of the cost multiplied by lead-time which can be considered as value reducing factors.

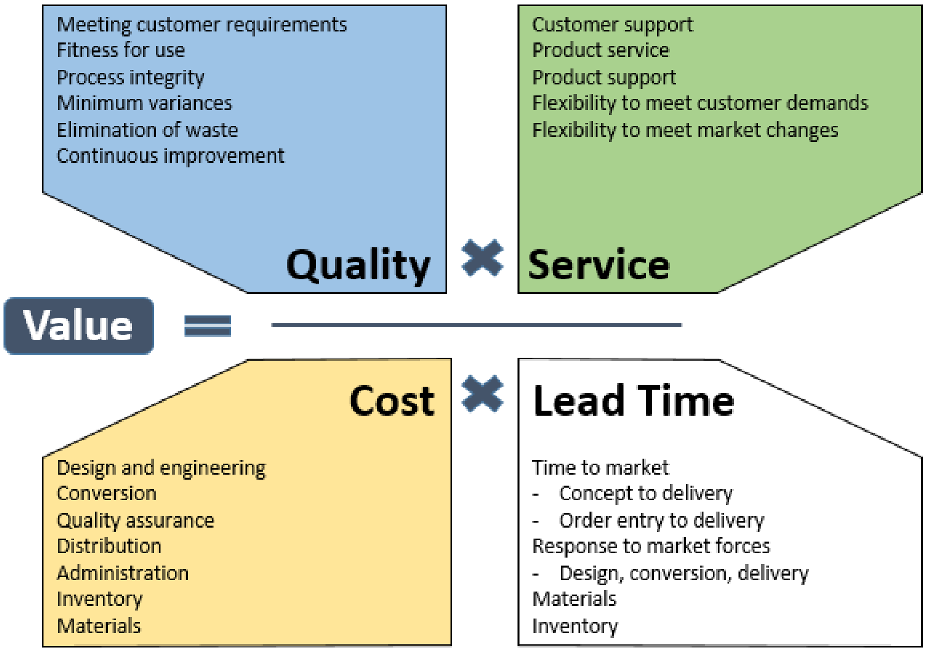


Figure 1. Total value metric (modified from Naylor et al., 1999)

Ever since the question "What is the right supply chain for your product" was asked by Fisher in 1997, companies as well as researchers have developed supply chain strategies for different needs. In our time it is already generally accepted that a one-size-fits-all supply chain solution is not enough in modern markets where customers have the possibility to switch their suppliers (Roscoe & Baker, 2014). While managing several supply chains might generate additional costs compared to having only a single supply chain, the profitability of the supply chain can increase when the different supply chains are managed properly. The reduced cost to serve can be achieved by eliminating the need for suppliers to make costly changes into existing processes due to customers varying needs (Gattorna, 2009). Pressure towards customized supply chains is engendered by the variety of customers and their unique needs. While one customer might consider the lead-time as the order winning criteria, another customer might prefer the price over all the other criteria. To satisfy the specific needs of the variety of customers, companies need to consider several parallel supply chain strategies. To build customized supply chains to match with the wide range of customer needs researchers have developed several different supply chain segmentation frameworks presented in **Table 3** (Roscoe & Baker, 2014).

The different supply chain segmentation frameworks share many similarities within one and another but has also clear differences. The objective of all these frameworks are the same: to influence on how the supply chain are segmented, but the approaches are somewhat different. Depending on the framework, the variables used in the segmentation varies from Fisher's (1997) product attributes to Christopher and Towill's (2000) order winning criteria (OWC) and to Gattorna's (2006) buying behavior.

Table 3. Supply chain segmentation frameworks

Author	Segmentation is based on	Identified supply chain segments	
Fisher (1997)	Product characteristics	Efficient process Responsive process	Lean/Agile supply chain strategies
Naylor, Naim, & Berry (1999)	Variability of products vs. variability on production	Agile supply chain Lean supply chain	
Christopher & Towill (2000); Mason-Jones, Naylor, & Towill (2000); Aitken, Childerhouse, Christopher & Towill (2005)	Market qualifying (MQ) and order winning (OW) criteria (cost, service level)	Agile supply chain Lean supply chain	
Christopher, Peck, & Towill (2006)	Demand predictability vs. lead-time	Agile supply chain Leagile supply chain (postponement) Lean supply chain Lean (continuous replenishment)	Mixture of agile and lean strategies
Juttner, Godsell, & Christopher (2006)	Demand chain alignment and product lifecycle management	Lean supply chain Agile supply chain Leagile supply chain	
Gattorna (2009)	Buying behavior	Continuous replenishment Lean supply chain Agile supply chain Fully flexible supply chain	
Childerhouse, Aitken & Towill (2002); Christopher, Towill, Aitken & Childerhouse (2009)	DWV3 variables (Duration of lifecycle, delivery window (lead-time), volume, product variety and variability of demand)	Rapid time-to-market Agile supply chain Lean supply chain	
Godsell, Harrison, Emberson & Storey (2006)	Buying behavior	Options by plan, source, make and deliver	More complex supply chain strategies
Payne & Peters (2004)	Product clustering (DVW3 variables + order-line value, -frequency, -weight, number of customers and substitutability of product	Centralized stock model, dispersed stock model, and finish to order	
Collin, Eloranta & Holmstrom (2009)	Customer requirements and demand visibility	Make to order, pack to order, ship to plan, ship to inventory replenishment, and ship to plan	

As we see from **Table 3** the influence on the supply chain is similar between the different frameworks. Many of the frameworks results on having supply chain strategies such as lean and agile (Naylor et al., 1999; Mason-Jones et al., 2000; Christopher & Towill, 2000; Atiken et al., 2005), but some frameworks go even further on recognizing more specific segments like leagile (Christopher et at. 2006; Juttner et al., 2006), a fully flexible supply chain (Gattorna, 2009) or differentialized stock models (Collin, Eloranta & Holmstrom, 2009). Despite the number of different supply chain strategies, the supply chain strategies are identified to be a trade-off between efficiency and service level (Selldin & Olhager, 2007). Basnet and Seuring (2016) exemplifies the trade-off between the efficiency of the supply chain and the service level between some of the popular supply chain strategies with a supply chain frontier, which illustrates how supply chain strategies are positioned on the frontier (**Figure 2**).

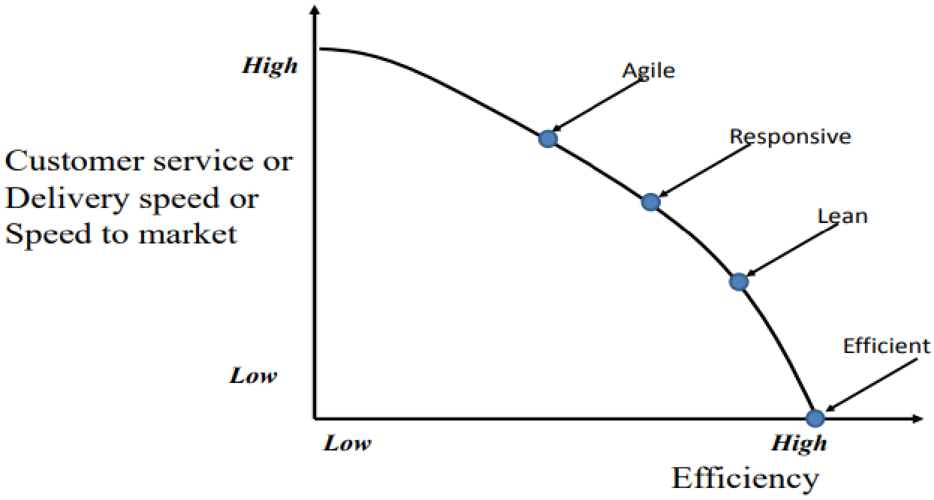


Figure 2. Supply chain frontier (reproduced from Basnet & Seuring, 2016 with the permission of the publisher)

When the supply chain frontier is interpreted trough the total value metric presented in **Figure 1**, we observe how different supply chain strategies generate value for customers. An agile supply chain seems to value customer service and lead-time over the efficiency (cost) whereas the lean supply chain strategy generates value by maximizing the efficiency by the expense of the other variables.

2.1 Lean versus agile supply chain

The two most widely used paradigms in the supply chain segmentation literature, lean and agile supply chain strategy, appears in many frameworks. Lean and agile methodologies were not originally invented for supply chain segmentation but rather for manufacturing from where the philosophies are adopted and implemented into the supply chain context. (Fisher 1997, pp. 99, Christopher & Towill, 2001)

Lean is a widely known paradigm where the focus is to eliminate waste. Waste means in lean context all non-value adding operations in a process such as time or material. Lean philosophy was originally invented for manufacturing, but it is used in a wide range of different applications (Christopher & Towill, 2001). In an ideal lean supply chain, all waste including inventories and part loads would be pruned off, leading to extremely cost-effective supply chains. Unpredictable demand among with lead-time requirements makes such "pure" lean supply chain with absolutely no waste and inventories difficult to achieve. Rather than pursuing an ideal lean supply chain, companies focus more on defining a minimum reasonable inventory (MRI), which acts as thresholds, after which reduction of stocks levels has a negative impact on profitability (Naylor et al., 1999). In manufacturing, lean concept works the best when the demand is stable and hence predictable, and the variety of product form is low (Christopher, Towill, 2001). The continuous and predictable demand enables many cost reducing operations such as manufacturing in bigger batches by optimizing the production cycles and filling up the supply vehicles to their maximum capacity. Even if lean manufacturing and supplying generally leads to lower costs, the cost-reducing factors of lean such as minimizing the inventories and extending production cycles often prolongs the lead-times and impairs the overall service level.

Agile manufacturing is unlike lean a methodology suitable for more volatile demand and hence it can be used to respond to customer demand where variability is high (Christopher & Towill, 2001). To be able to match the supply to the volatile demand, the organization's processes must be built on flexibility (Christopher & Towill, 2001). Flexibility often requires a departure from the most efficient way of operating, for example by splitting production cycles into smaller batches or by supplying only partially full trucks. In an agile supply chain, the decoupling point, meaning the point in the supply chain network where the flow of materials changes from factory push to customer pull, is often pushed closer to the customer and hence it is more responsive than lean manufacturing (Naylor et al., 1999). An alternative way of achieving a proper agile supply chain strategy involves building redundant capacity along the supply chain. The unused capacity enables small batches to be supplied alongside the regular production plan and thereby keeps the lead-times short (Gattorna, 2009). The main features of lean and agile supply chain shown in **Table 4**.

Table 4. Distinguishing features of lean and agile supply. Adapted from Mason et al. (2000), Naylor et al. (1999), Christopher & Towill (2000), Atiken et al. (2005)

Distinguishing attributes	Lean supply chain	Agile supply chain
Typical products	Functional	Innovative
Marketplace demand	Predictable	Volatile
Product variety	Low	High
Product life-cycle	Long	Short
Customer drivers	Cost	Availability
Profit margin	Low	High
Lead-time	Long	Short
Price sensitivity	Low	High

Fisher was the first author to note that problems plaguing supply chains is fundamentally due to the mismatch of the supply chain and product type (Roscoe & Baker, 2014). Fisher's (1997) supply chain segmentation framework defines the exact chain that should be used to supply the products based on product attributes. The idea is premised on defining whether the product is a functional- or an innovative product and respectively choosing the appropriate supply chain for it. The functional products have steady demand, long lifecycles, and usually low margins whereas innovative products are essentially the opposite. To simplify these categories even more: functional products satisfy the basic need of a customer while innovative products satisfy a specific need (Fisher, 1997). The difference between basic and specific need can be clarified by an example: having paper as the product category, a basic need could be something like plain white printing paper whereas a specific need could be tailored greeting cards. Additional features of the two product categories are gathered in **Table 5** to help managers classify the products correctly into functional and innovative products.

Table 5. Functional versus innovative products differences in demand (reproduced from Fisher, 1997 with the permission of the publisher)

	Functional (Predictable demand)	Innovative (Unpredictable demand)
Aspects of demand		
Product life cycle	more than 2 years	3 months to 1 year
Contribution margin*	5% to 20%	20% to 60%
Product variety	low (10 to 20 variants per category)	high (often millions of variants per category)
Average margin of error in the forecast at the time production is committed	10%	40% to 100%
Average stockout rate	1% to 2%	10% to 40%
Average forced end-of-season markdown as percentage of full price	0%	10% to 25%
Lead time required for made-to-order products	6 months to 1 year	1 day to 2 weeks

* The contribution margin equals price minus variable cost divided by price and is expressed as a percentage.

From the features presented in **Table 5** we get a general understanding of the properties of the two product categories. Because of the different characteristics of the two product categories, we need more than just one supply chain to deliver products to the market. The two supply chain strategies Fisher (1997) propounds for these two product categories are efficient and a responsive supply chain. The key difference between these two supply chains is to supply products either as efficiently and at low costs as possible or with as short lead-times as possible (**Table 6**). Even if Fisher didn’t yet in 1997 name these two supply chains as lean and agile supply chain, the analogy between the naming conventions of efficient and responsive versus lean and agile supply chain is apparent.

Table 6. Physically efficient versus market-responsive supply chains (reproduced from Fisher, 1997 with the permission of the publisher)

	Physically efficient process	Market-responsive process
Primary purpose	supply predictable demand efficiently at the lowest possible cost	respond quickly to unpredictable demand in order to minimize stockouts, forced markdowns, and obsolete inventory
Manufacturing focus	maintain high average utilization rate	deploy excess buffer capacity
Inventory strategy	generate high turns and minimize inventory throughout the chain	deploy significant buffer stocks of parts or finished goods
Lead-time focus	shorten lead time as long as it doesn't increase cost	invest aggressively in ways to reduce lead time
Approach to choosing suppliers	select primarily for cost and quality	select primarily for speed, flexibility, and quality
Product-design strategy	maximize performance and minimize cost	use modular design in order to postpone product differentiation for as long as possible

After classifying the products based on their characteristics in functional or innovative products, the next step in Fisher's (1997) framework is to build 2x2 matrices to define whether the correct supply chain to be used is a responsive or an efficient one. The matrices itself (**Figure 3**) is straightforward to use and gives an understandable overview of the current situation.

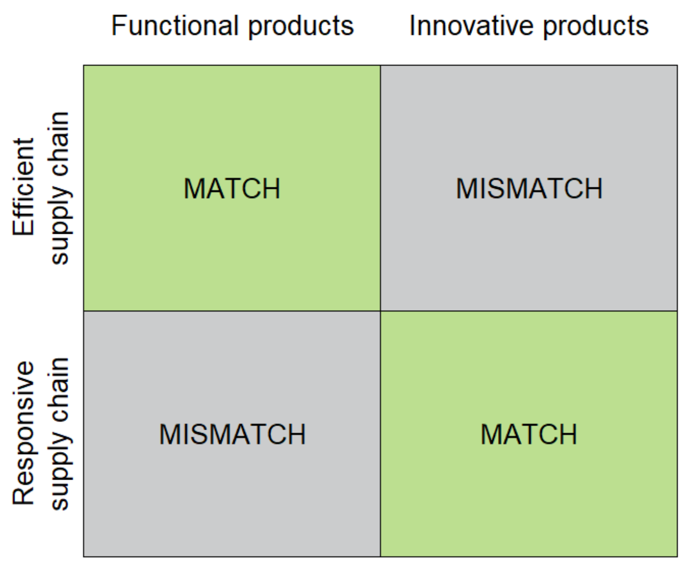


Figure 3. Matching products with supply chains (modified from Fisher, 1997)

It was not until Naylor et al. (1999) set the naming convention of the two main supply chain strategies as lean and agile. The framework of Naylor et al. (1999) was very similar to the one Fisher proposed in 1997. Rather than classifying products into functional or innovative products, Naylor et al. (1999) suggest comparing the demand for a variety of products and the demand for variability in production in 2x2 matrices to find the applications of leanness and agility. The low variability in products and production indicates a lean supply chain strategy should be used whereas high variability indicates the need for an agile supply chain (Naylor et al., 1999).

Another way authors have segmented the supply chain into agile and lean is to use order winning criteria rather than only product-related characteristics. In the framework introduced by Mason-Jones et al. (2000), Christopher and Towill (2000), and Aitken et al. (2005) the right supply chain is selected according to marketplace requirements. This framework provides an alternative way to define the right supply chain strategy without having to use only product- and production-centric attributes. In this concept, the customer is set in the center and the question is asked: “How to win this order”. The two types of supply strategies have different “order qualifiers” (OQ) and “order winners” (OW) based on the needs of the customers. Depending on the framework, OQ and OW might also be referred as market qualifiers (MQ) and market winners (MW). For agile supply, the market winner is according to Mason-Jones et al. (2000) the service level and for lean supply respectively the cost. The market qualifiers define other preferences in the supply chain model (**Figure 4**). (Mason-Jones et al., 2000)

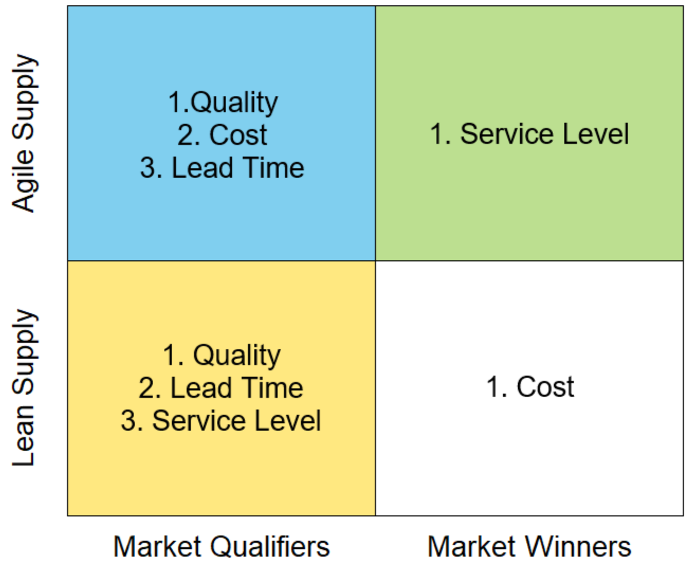


Figure 4. Market qualifiers and market winners with lean and agile supply (modified from Mason-Jones et al., 2000)

Since in lean manufacturing excess time is considered as waste and leanness thrives to eliminate all waste, lead-times must also be shortened by definition. The key difference between lean and agile supply chain in terms of total value provided to the customer is therefore directly related to the market winners: lean supply focuses on reducing costs at the expense of other attributes whereas the critical factor in agile supply is the service level. (Mason-Jones et al., 2000)

2.2 Further supply chain segmentation

While many of the first supply chain segmentation frameworks introduced in the early 2000s by numerous authors identify only two major supply chains, lean and agile, many authors have since identified more specific supply chain segments (Roscoe & Baker, 2014). While having two separate supply pipelines to serve customers is superior to having only one fixed supply chain, the variety of customer needs requires more diverse supply chain strategies.

The framework by Christopher et al., (2006) takes the supply chain segmentation further and proposes two additional segments in addition to lean and agile supply. Within this framework, the used supply chain is decided similarly with a 2x2 matrices as in Fisher's (1997) framework with the difference that the attributes compared are lead-time and predictability of the demand and each area of the matrices presents a specific supply chain (**Figure 5**). The dominant variable in this framework is the predictability of the demand. Products with steady or easily predictable demand can be supplied by waste eliminating and efficient lean process while products with unpredictable demand are supplied focusing on other factors over cost. The framework by Christopher et al. (2006) considers besides the predictability of demand, also the lead-time, which specifies the actual supply chain to be used.

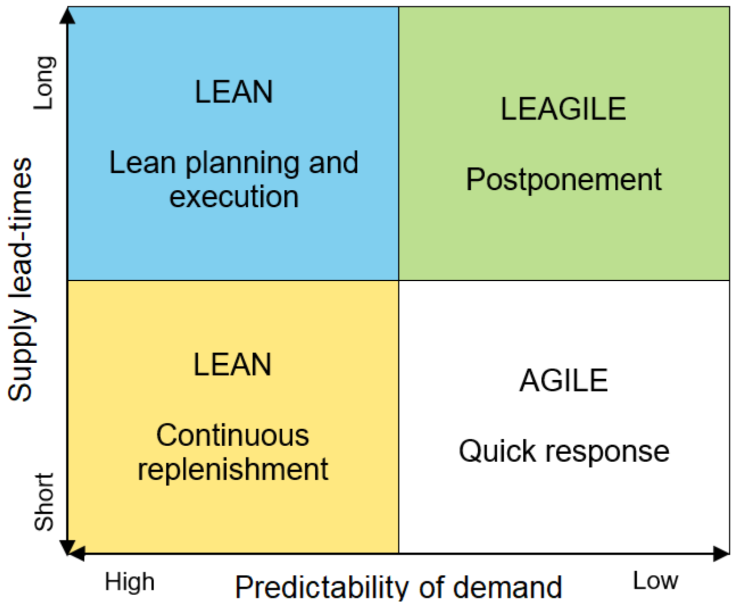


Figure 5. Supply chain segmentation based on demand predictability vs. lead-time. (modified from Christopher et al., 2006)

The "pure" lean supply strategy in Christopher et al. (2006) framework is used when the predictability of demand is high, and the supply lead-times are long. This combination provides the circumstances to manufacture and supply the products with an efficient cost-minimizing process. The steady demand and long lead-times make it possible to plan the production and supply chain further in the future and therefore manufacture or ship the products in larger batches and thus minimize the waste.

When shorter lead-time is required for steady demand products, a lean supply chain with continuous replenishment should be selected. Shorter lead-time may be desirable when products have a short life cycle or when possible stock out would be expensive for the customer. Christopher et al. (2006) A way to reach lean supply with continuous replenishment can be by storage or so-called ready for load (RFL) location closer to the markets. This buffer storage allows short lead-times to the customer location while still maintaining the advantages of lean production and supply. According to the total value equation (Naylor et al., 1999) (**Figure 1**), while quality and service are remaining the same in a lean supply chain with continuous replenishment but lead-time in the nominator decreases, the cost of this supply chain will respectively increase to provide the same value for the customer.

The quick and responsive agile supply chain is selected in Christopher et al. (2006) framework when the lead-time is short, and the demand is volatile or hard to predict. Respectively, when the demand is unpredictable and the lead-time required is rather long, Christopher et al. (2006) suggests a mixture of lean and agile supply, leagile. In this supply chain strategy, the lead-time is not an order winning criterion which allows the postponement of the production until a reasonable size batch can be manufactured and shipped.

2.2.1 Multiple variable framework

While many of the paradigms listed in **Table 3** approaches the supply chain segmentation problem with 2x2 matrices or with only a few different variables, Childerhouse et al. (2002) considered that more factors must be taken into consideration when planning supply chain strategies. The importance of variability of demand and volume has been generally identified for some time but using only product-related characteristics to shape supply chain strategies does not necessarily reflect the actual needs of different market segments. The idea of combining both market-related- and product-related factors to define the supply chain strategies led to the DWV3 key variables (Christopher, et al., 2009):

- Duration (of product lifecycle)
- Window (time window for delivery/lead-time)
- Volume (Pareto classification)
- Variety (of product form)
- Variability (of demand).

As the supply chain strategies are context-specific, the DWV3 variables shouldn't be used by default together. Instead, the idea of the framework is to identify which ones of the DWV3 variables are dominant and use these key variables to define the supply chain segments (Godsell et al., 2011). To be able to critically analyze the fit of the DWV3 variables on the design of supply chain strategies, the variables need to be explained in more detail.

• Duration of the product life cycle

The expected duration of a product life cycle is an important factor to consider when developing supply chain strategies. The phase of the lifecycle as well as the length affects the requirements of the supply chain. The different phases of the life cycle: introduction, growth, maturity, saturation, and decline, all have unique characteristics regarding the order winners and market qualifiers shown in **Figure 6**. The OW for the maturity and saturation phase is cost while the service level is assigned as the OW for the growth and decline phase of the life cycle. (Atiken et al., 2005)

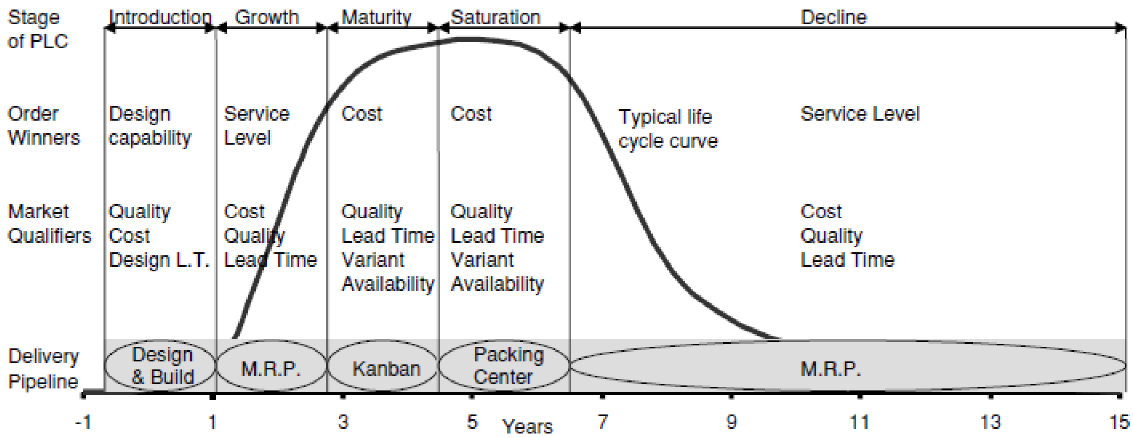


Figure 6. How OW and MQ characteristics change as a function of the life cycle (reproduced from Childerhouse et al., 2002 with the permission of the publisher)

• Time window for delivery

The time window for delivery or "window" as it is defined within the DWV3 variables, is also an essential factor in segmenting the supply chain. Generally, the window for delivery refers to the lead-time, which is a more widely used term in the supply chain segmentation context. Lead-time has usually a strong influence on the separation between agile and lean supply chain strategies (Atiken et al., 2005). Already in Fisher's (1997) framework, the lead-time was in a key role in separating the functional and innovative products from each other where the lead-time for innovative make-to-order (MTO) products was marked out as one day to two weeks and for functional MTO-products respectively from six months to one year.

• Volume

Products with a high level of demand are often aimed at mass markets and therefore the condition often allows a lean type of production and supply. The high volumes enable the use of strategies like make-to-forecast, where production is based on forecast rather than individual orders. Furthermore, the constant and forecastable demand allows the economies of scale to lower the total costs. Conversely, products with low volume enable more flexibility in the production and the wider supply chain. (Atiken et al., 2005; Christopher, et al., 2009)

One way to bind the lean and agile paradigms together regarding the volume is by Pareto classification (**Figure 7**). In many manufacturing companies, Pareto Law applies meaning that 20% of products generate 80% of sales and vice versa (Christopher and Towill, 2001). Reasonably the wide gap in the demand characteristics with the high-volume products and the long tail of low volume affects the best fitting supply chain strategy for these products. Occasionally it is appropriate to use lean strategies for 20% of high-volume products and other strategies like agile for the remaining 80% of low volume products (Atiken et al., 2005).

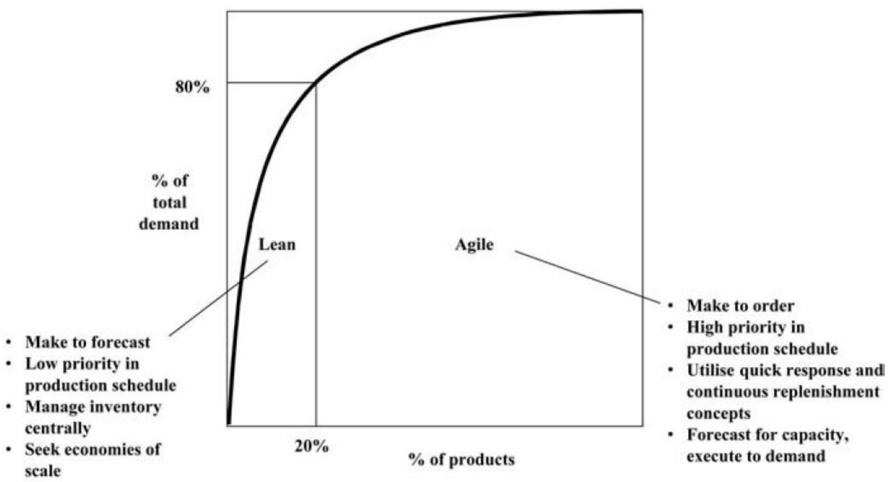


Figure 7. The Pareto distribution (reproduced from Christopher & Towill, 2001 with the permission of the publisher)

• Variety

The level of variety demanded in the marketplace correlates with the number of stock-keeping units (SKU's). The higher the level of variety demanded by the marketplace is, the lower the average volume per SKU is. Eventually, a higher variety of products reflect the number of changeovers and set-ups of the machines making the production and supply more fragmented and therefore less efficient. With fragmented production, the complexity and waste of time grow which is why a pure lean strategy is challenging to implement to production with high variety. (Atiken et al., 2005; Christopher, et al., 2009)

• Variability

Variability refers to the unpredictability or "spikiness" of the demand. Predictable demand fits the best for lean supply chain strategies where unpredictable demand requires more flexibility from the production and hence more agile strategies should be used. As a measure of variability Atiken et al., (2005) uses the coefficient of variation (CV) defined in (1), where the standard deviation is divided by the mean. The higher the CV is, the harder the demand is to forecast and therefore agile-like strategies might work the best. (Atiken et al., 2005; Christopher, et al., 2009)

$$\text{Coefficient of Variation} = \frac{\text{standard deviation}}{mean} \tag{1}$$

By moving from two-dimensional supply chain segmentation into five-dimension segmentation with the DWV3 variables, the possible number of supply chain pipelines grow drastically. Where supply chain is defined as a network of interdependent organizations working together to provide a flow of products to the markets, pipeline is defined as the operational mechanisms and procedures in service specific product or market contexts. Even when modifying the DWV3 variables into binary variables, the number of possible pipelines grows from 4 to 25. From a company perspective, managing 25 different supply chain pipelines is not expedient which is why the pipelines can be reduced by clustering the set of variables into supply chain strategy specific clusters. (Atiken et al., 2005)

2.2.2 Buying behavior

Another way researchers have approached the supply chain segmentation problem, besides using product related characteristics to segment the supply chain into a mixture of lean and agile strategies, is by a more strategic approach. Gattorna (2009) made an argument that also other parameters than just product characteristics, demand and supply attributes, and order winners and -qualifiers must be used in supply chain segmentation. Gattorna's (2009) framework approaches supply chain segmentation by segmenting the customers by their buying behavior and reverse-engineering the supply chain strategies from there. The idea itself is kind of reverse to the product characteristic centered approach. Rather than defining supply chain strategies based on the known product and demand related features, the focus is on identifying customer segments and aligning the supply chain strategies to fulfill each segment's needs. The four key customer behavior types are: "Understand me", "surprise", "be consistent", and "respond". These behavior types are described in more detail in **Figure 8**. (Gattorna, 2009)

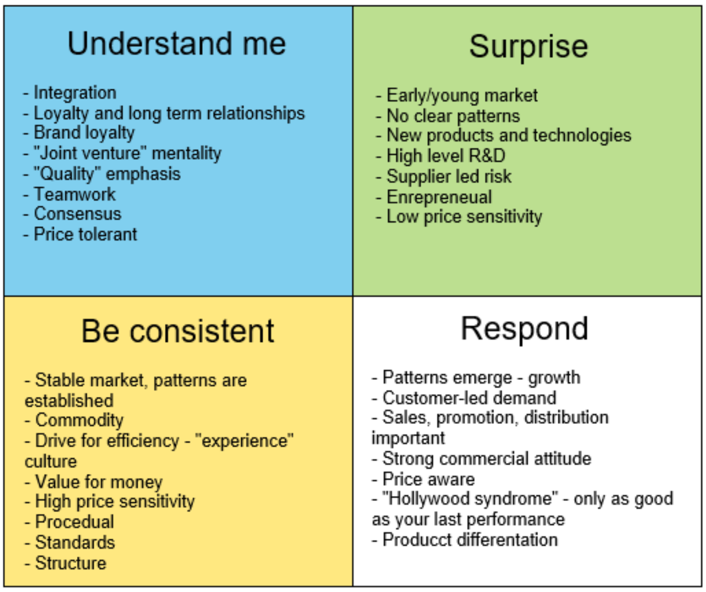


Figure 8. Primary customer behavior types. Modified from Gattorna (2009) and Gattorna (2015)

Even if the features of the behavior types have similarities within the supply chain strategies presented previously (e.g. Christopher et al., 2006 framework), Gattorna's approach is much more customer orientated. According to Gattorna, these behavior models have each a unique way of experiencing customer service. Customers classified in the "understand me" –category are looking for empathy and understanding from the customer service and this way they are willing to grow together and build a long-lasting relationship. Customers segmented in the "surprise"-category on the other hand finds good customer service as understanding their unique needs and surprising them constantly with innovative and creative solutions, at speed. For customers belonging in the "respond"-group good service means mostly being able to fulfill their demanding requirements regarding short lead-times with great responsiveness. Finally, customers in the "be consistent" group are seeking a reliable and predictable service with the focus on consistency. (Gattorna, 2009) **Appendix 1** shows a more detailed listing of service logics related to the specific customer behavior types.

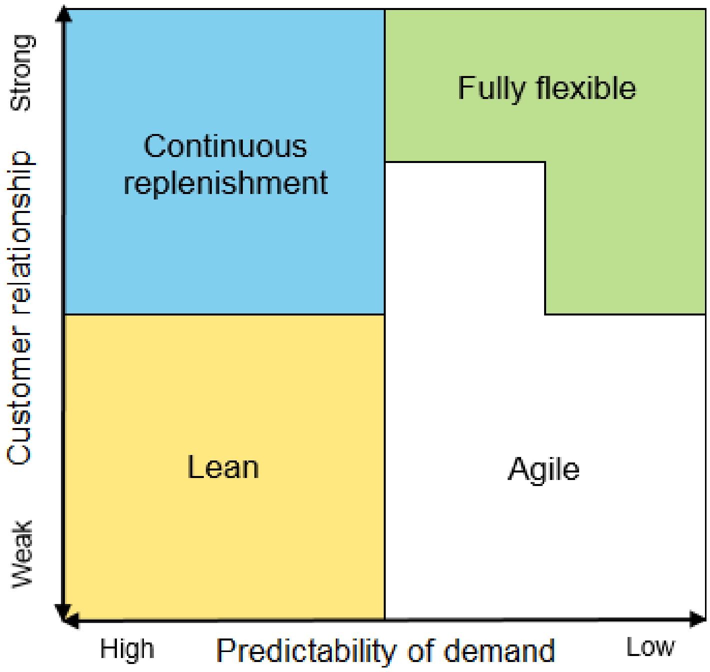


Figure 9. The four generic supply chain types. (Modified from Gattorna, 2009)

To match the buying behaviors into actual supply chain strategies, Gattorna uses again a 2x2 matrices (**Figure 9**) and four unique supply chain strategies, most of which are familiar from other frameworks as well. In this framework, the x-axis is the predictability of demand just like in the Christopher et al., (2006) framework, but instead of lead-time on the y-axis, Gattorna proposes the relationship to the customer. Gattorna suggests that the focus of the supply chain should be with continuous replenishment on the retention of customer relationship, with fully flexible supply chain on providing creative solutions for premium price, with lean the focus should be on efficiency and finally, with agile supply chain the focus should be on speed and capacity (Gattorna 2009). In 2015, Gattorna updated its framework by adjusting the boundary between agile and fully flexible supply chain, re-naming the "continuous replenishment" supply chain type as "collaborative" as well as introducing yet another supply chain type – campaign (**Figure 10**). Campaign is an overlapping supply chain type as it has elements of lean, agile, and fully flexible supply chain.

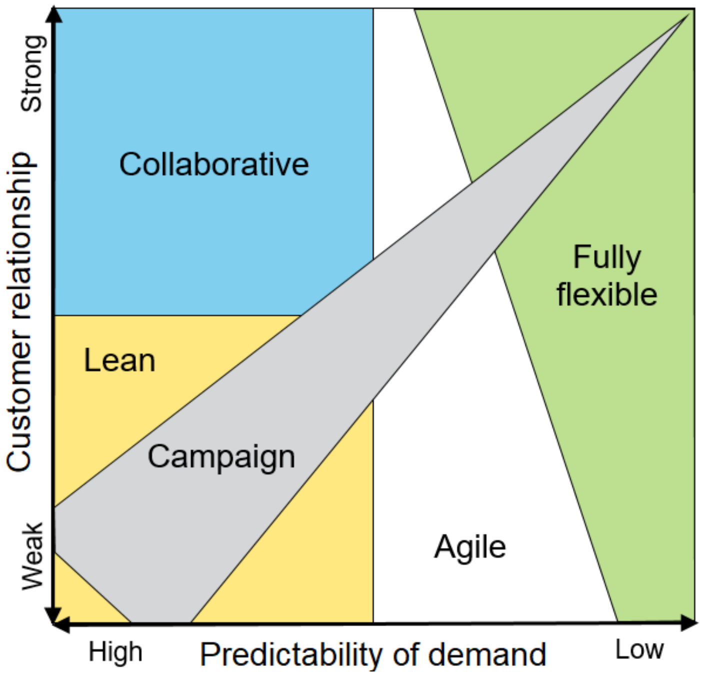


Figure 10. The updated five generic supply chain types. (Modified from Gattorna, 2015)

The five supply chain types, the collaborative, campaign, lean, agile, and fully flexible, are described in **Figure 11**. Even though different variations of these supply chains exist, the four generic supply chains presented cover most of the cases. (Gattorna, 2015)

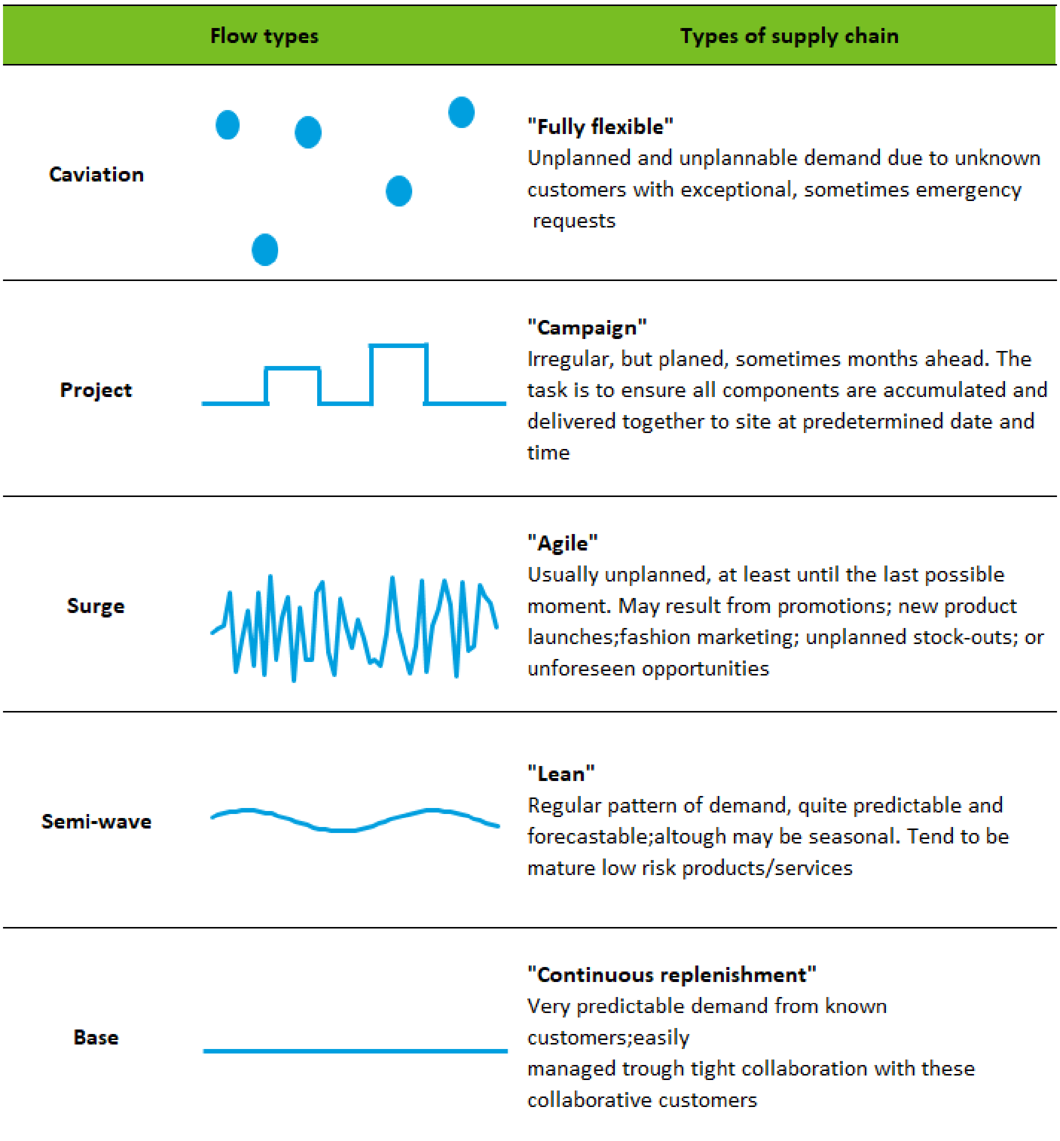


Figure 11. Demand types of different supply chains. (Modified from Gattorna, 2015)

A pivotal feature of Gattorna's (2009) framework is a so-called dynamic alignment. This means that the selected supply chain strategies must be always aligned with the customers buying behavior. Consequently, the provided supply chain strategies can change together with customers buying behavior changes. Due to dynamic alignment, it is possible and even relatively usual that a given customer might exhibit all four different buying behaviors at different times of their lifecycle. Properly managed, the dynamic alignment brings many additional perks to the supply chain management. The ability to match the supply chain strategy with changing customer buying behavior, enables companies to answer better market expectations, without having to make continuously costly exceptions from a fixed process. Furthermore, the tailored service bundles for dynamically changing customer behaviors enable fulfilling the customer requirements better in long term, and hence it shows as better service level. Having the possibility to switch the supply chain strategy within a single customer based on their changing needs, makes also the developing and charging of additional value-adding supply chain services easier. It is precisely the dynamism and the integration of strategy that makes Gattorna's (2009) framework so versatile. The constant alignment of the supply chain design and customer needs eventually results in functional excellence within the different supply chain strategies. (Gattorna, 2009)

3 CLUSTERING THEORY

Digitalization has triggered a movement towards data-centric world. Alongside with the exponential growth of the amount of data, the growing interest towards data processing and analyzing techniques is obvious. Raw data becomes meaningful insights only when it is processed it into an understandable format with suitable methods. Grouping and clustering data is one of the most important data analysis activity when it comes to obtaining information of the data (Xu & Wunsch, 2009).

Essentially all classifying systems are either supervised or unsupervised, depending on whether they classify the new data objects into discrete predefined supervised classes or respectively into unsupervised clusters. Due to the increasing amount of data, manual labeling has come extremely difficult and time consuming, and therefore automatic labeling has become crucial step in data mining (Aggarwal & Reddy, 2013). The objective of clustering algorithms is to separate an unlabeled dataset into a finite and discrete number of clusters by identifying the unseen data structures, and therefore clustering methods falls into the family of unsupervised learning (Xu & Wunsch, 2009). A simple clustering result with two-dimension dataset is represented in **Figure 12**.

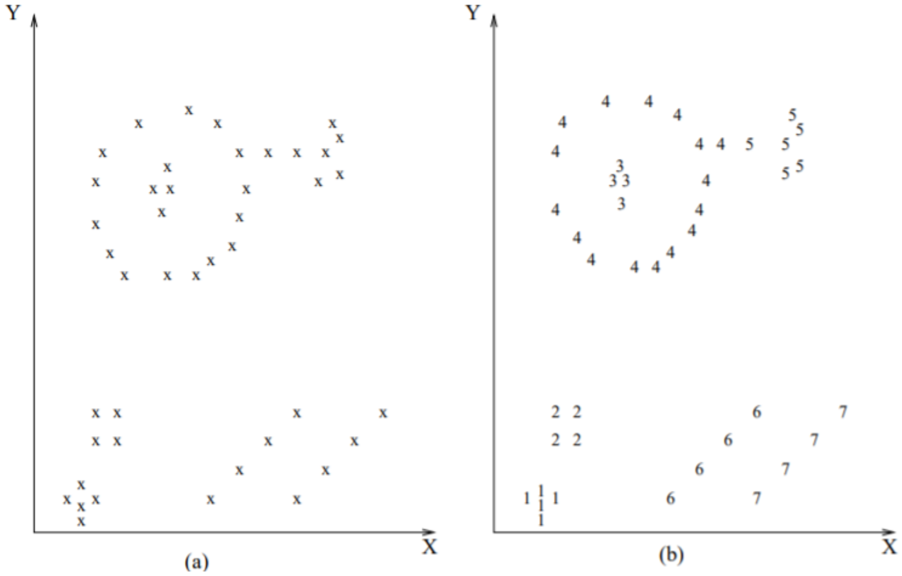


Figure 12. Example of clustering a dataset (reproduced from Jain, Murty & Flynn, 1999, pp. 2 with the permission of the publisher)

Clustering in useful technique in several data mining and machine learning tasks such as data summarization, network analysis, and customer segmentation. There are a variety of different clustering algorithms each with their unique logic and characteristics. Generally clustering methods can be categorized into five groups: partitioning methods, hierarchical methods, density-based methods, grid-based methods, and probabilistic methods. Algorithms belonging to partition methods uses the distance between data points to capture the similarities and dissimilarities within the data. Hierarchical methods on the other hand divides the data into levels and forms a hierarchy-like structure of the data. Algorithms using density-based clustering can detect arbitrarily shaped data structures from the data and therefore can separate objects based on their density variation. Grid-based methods faces clustering problems by dividing the data space into a finite number of grids, and dividing data into clusters based on the density of the surrounding regions. Finally, the probabilistic methods assume that the data is represented by a combination of probability distributions and hence the clustering problem is converted into a parameter estimation problem. (Aggarwal & Reddy, 2013) Later in this chapter, a set of commonly used clustering methods in customer segmentation applications are presented.

3.1 K-means

The k-means (KM) algorithm is the most known and widely used clustering algorithm in scientific and industrial applications (Aggarwal & Reddy, 2013; Omran, Engelbrecht & Salman, 2007). K-means algorithm was introduced already in 1967 by McQueen and hence it has been widely researched and generally approved (Berkhin, 2002). K-means is a distance-based clustering method where the objective of the algorithm is to minimize the sum of squared errors (SSE). The objective function (2) assigns all data points to the nearest cluster center based on distance (Hamerly & Elkan, 2014). As a distance measure, Euclidian distance is usually used among with the algorithm, but any other suitable distance measure may also be used (Aggarwal & Reddy, 2013; Omran, Engelbrecht & Salman, 2007).

$$KM(X,C) = \sum_{i=1}^{n} \min_{j \in \{1 \dots k\}} \left\| x_i - c_j \right\|^2 \tag{2}$$

X = n elements$\{x_1,\dots,x_n\}$
C = k clusters $\{c_1,\dots,c_k\}$

K-means algorithm works in three steps (**Figure 13**) and hence it is straightforward and easy to understand (Hamerly & Elkan, 2014). The first step is to initialize the k-cluster centers defined as an input parameter. The second step is to calculate the distances of each data point to other k-cluster centers and furthermore assign the points to the closest cluster center. In the third step the new cluster centers are recomputed using all data points assigned to the clusters in step 2. K-means is an iterative algorithm as it iteratively repeats steps two and three until the assignment no longer changes and the convergence of the centroids is reached (Aggarwal & Reddy, 2013; Omran et al., 2007).

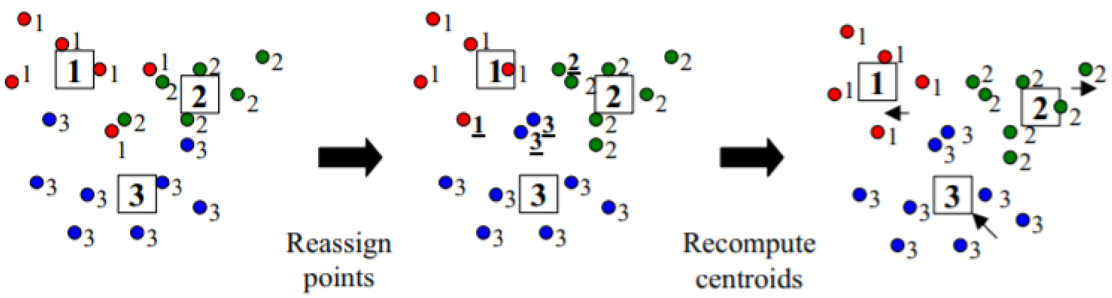


Figure 13. K-Means algorithm works iteratively in two phases (reproduced from Berkhin, 2002 with the permission of the publisher)

Despite the popularity and numerous implementations, *k*-means algorithm has also its downsides. Since the objective function cannot be solved directly but rather by an iterative approach, the found solution is only a local optimum. Therefore, k-means algorithm may give different results depending on what initialization method is being used (Hamerly & Elkan, 2014). Whether the selected initialization method is Hartigan and Wong (1979), MacQueen (1967), Lloyd (1957) or Forgy (1965) effects on how the initial cluster centers are selected and therefore affects the outcomes. Another weakness of k-means algorithm is the sensitivity to outliers (Berkhin, 2002). In every iteration when the new cluster centers are computed, the outliers pull the cluster centers further away from the actual cluster centers and hence weakens the results. Since k-means assigns the datapoints to the nearest cluster center, the results might also be unbalanced due to different distances among datapoints. The need for user input for defining the number of clusters prior applying the algorithm, as well as the support for only numerical variables are also considered as a weakness of the k-means algorithm. (Berkhin, 2002; Aggarwal & Reddy, 2013)

3.2 Fuzzy c-means

Fuzzy c-means, sometimes called as fuzzy k-means, is an adaption of k-means algorithm developed to its current format by Bezdek in 1981 (Hamerly & Elkan, 2014). Fuzzy c-means (FCM) is similar to the k-means algorithm with the difference that the membership function is soft rather than hard as it is in KM (Omran et al., 2007) (3). This means that unlike KM, FCM allows the data points to belong into several clusters with a certain degree. The possibility for data points to belong simultaneously into several clusters within a certain degree is especially useful when a crisp clustering result is not desired.

When working with complex datasets with overlapping clusters, it is not feasible to assign points into crisp clusters, but rather to allow the clusters to overlap to some extent (Aggarwal & Reddy, 2013). The possibility to belong into several clusters simultaneously with a certain membership degree, resolves some of the problems that comes up with hard clustering methods like KM. On the other hand, FCM requires an additional fuzzifier parameter to be defined by the user which might be challenging (Omran et al., 2007). Eventually due to the similarities with both algorithms, the KM and FCM, they fall into same pitfalls regarding the sensitivity to outliers and finding only the local optimum (Aggarwal & Reddy, 2013).

$$FKM(X,C) \quad = \quad \sum_{i=1}^{n}\sum_{j=1}^{k} u_{ij}^{r}||x_i - c_j||^2 \tag{3}$$

u_{ij} = membership value
r = fuzzifier parameter
k = number of clusters
n = number of samples

The objective function of FCM presented in equation 3 differs from equation 2 by having the membership values and fuzzifier values involved. The membership value stands for the membership value of data point x_i belonging in j cluster and the fuzzifier value ($r \geq 1$) defines how much fuzziness is allowed in the clusters. The fuzzifier is defined by the user and the higher the value is, the lower the membership degrees are, and hence the fuzzier the clusters are. Small fuzzifier values instead result to less fuzziness and eventually in the lower limit fuzzifier being 1, the membership converges to either 0 or 1 resulting a crisp partitioning. (Hamerly & Elkan, 2014)

3.3 Hierarchical clustering

Hierarchical clustering algorithms are among partitional clustering algorithms the most widely studied unsupervised clustering methods (Omran et al., 2007). Just like with partitional clustering algorithms, hierarchical clustering algorithms have become widely used among different applications primarily because their simplicity and ease of implementation (Omran et al., 2007). The specialty of hierarchical clustering algorithms is that as they execute the task, they create a binary tree-like data structure called the dendrogram (**Figure 14**) (Jain et al., 1999, pp. 14). The dendrogram presents the nested groupings of similar data points which can be used to decide the optimal number of clusters. As the dendrogram has assigned the datapoints into different number of clusters from one cluster to the amount of datapoints, there is no need to rerun the algorithm to obtain an output with different number of clusters like in partitional- or density-based algorithms. Rather, the dendrogram can be split from a different level to obtain a solution with different number of clusters. The drawback from computing the whole dendrogram is that hierarchical clustering methods are generally computationally expensive algorithms to use. Besides the computational expense, another disadvantage of hierarchical clustering methods is that they struggle to separate overlapping clusters (Omran et al., 2007).

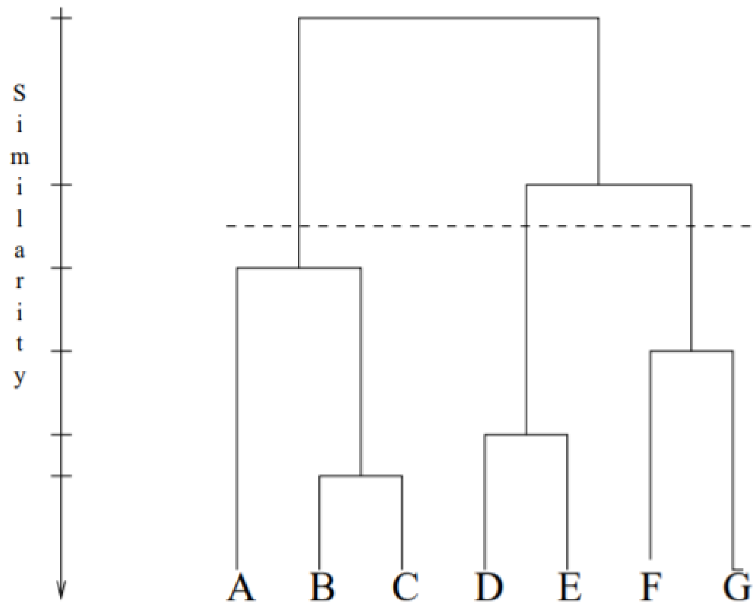


Figure 14. Hierarchical clustering dendrogram example (reproduced from Jain et al., 1999, pp. 15 with the permission of the publisher)

Hierarchical clustering can be executed using either divisive (top-down) or agglomerative (bottom-up) approach. Several different similarity measures can be used among with agglomerative clustering. The most studied algorithms are single link, complete link, average linkage, and Ward’s criterion, from which single link and complete link are the most used ones (Jain et al., 1999, pp. 15; Aggarwal & Reddy, 2013). As agglomerative clustering starts by having all singleton data points separated and combining them one by one into bigger clusters, the different similarity measures influences the outcome (**Figure 15**). The similarity of two clusters in single link clustering method is the similarity between their nearest members. This approach gives by its nature weight for datapoints being close to each other and hence it is capable to detect nonelliptical shapes. The drawback on the other hand is that the single link approach finds only local optimum. The complete link clustering on the other hand tackles the problem of local optimum by combining clusters together by the similarity of their most dissimilar members. Combining the clusters by their most dissimilar datapoint leads to more compact shaped clusters. Both agglomerative clustering methods, the single link method and the complete link method, are sensitive for outliers (Aggarwal & Reddy, 2013).

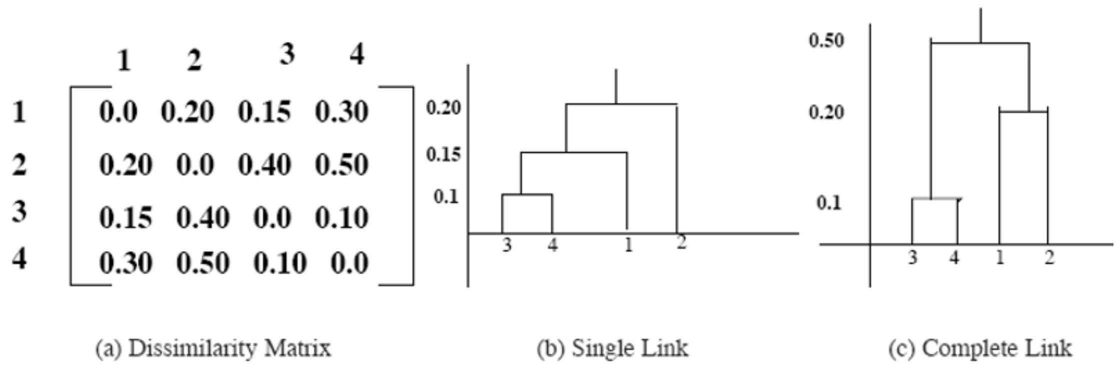


Figure 15. Illustration of agglomerative clustering using four data points (reproduced from Aggarwal & Reddy, 2013 with the permission of the publisher)

Divisive hierarchical clustering algorithm starts inversely from the top having all data points in one cluster and then beginning to split the dataset into smaller clusters until singleton level is reached. Neither agglomerative- or divisive hierarchical clustering methods require initialization since the algorithms starts by having either all datapoints in one cluster or all datapoints in individual clusters. When it comes to computational cost, divisive clustering is more efficient than agglomerative clustering, especially when there is no need to run the algorithm all the way to the individual data point level. As divisive approach has the complete information about the datapoints before splitting the clusters into smaller clusters, it can also be considered as a global approach (Aggarwal & Reddy, 2013).

3.4 Self-organizing maps

Self-organizing map (SOM) is an unsupervised clustering-like algorithm originally invented by Kohonen (1982) and have since been applied widely in clustering problems and data exploration in variety of different applications (Kohonen, 2013). Self-Organizing maps are an effective way to visualize high-dimensional data as the SOM maps the data into low-dimensional, usually two-dimensional grid of nodes. The simple and understandable semantic-map-like visualization where similar models are closer to each other in the grid than the more dissimilar ones, makes SOM a great way to find patterns and relations from the underlying high-dimensional data. (Kohonen, 1998; Omran et al., 2007)

SOM uses an artificial neural network (ANN) to associate the inputs with the output nodes, but unlike other types of ANN, SOM has no output function which enables SOM to map the vectorial input data with similar patterns into a lower dimensional grid of neurons in a topology-preserving manner (Herbst & Casper, 2008). The logic behind SOM is described below.

The SOM algorithm's iterative process (**Figure 16**):

Step 1: The reference vectors of the units are initialized to the grid

Step 2: Input vector is randomly picked and the best matching unit (BMU) is selected by calculating Euclidean distance to each reference vector

Step 3: The input vector is assigned to its BMU and the reference vectors are updated within neighborhood

Step 4: Steps 2 and 3 are repeated with the new weight vector until a fixed number of iterations is reached

(Herbst & Casper, 2008)

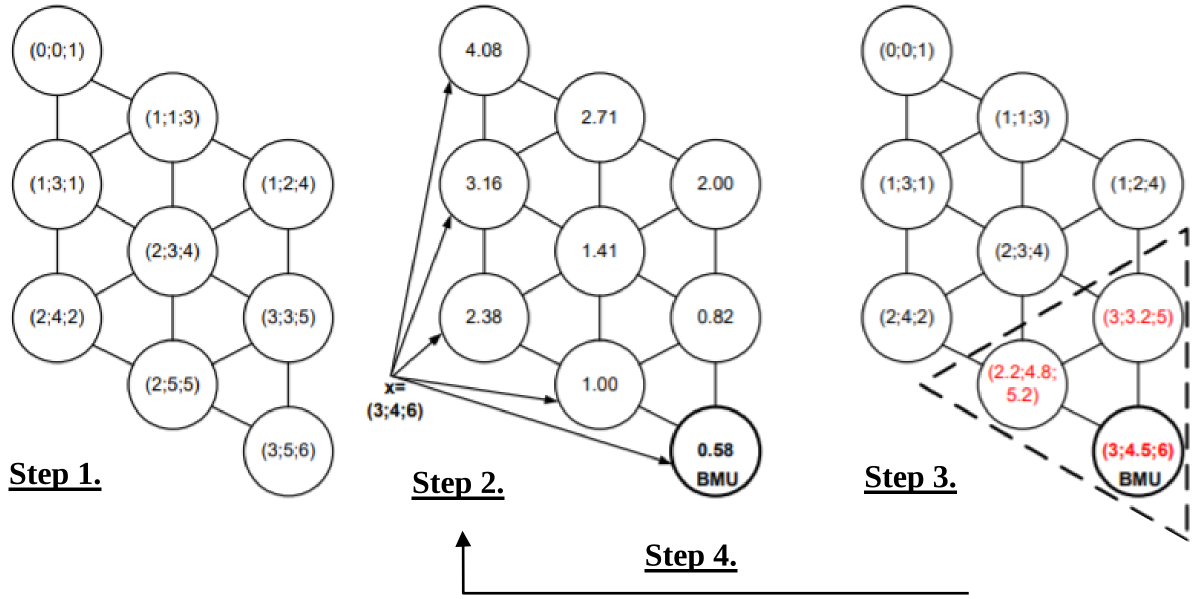


Figure 16. The operational logic of the SOM algorithm (reproduced from Herbst & Casper, 2008 with the permission of the publisher)

As the fit of SOM has been researched and accepted in applications in variety of different fields such as biomedical analyses, industrial analyses and financial applications it has also been used to segment customers (Kohonen, 2013; Bloom, 2014; Li & Lin, 2008). It is not a surprise since SOM has identified to have several advantages compared to other clustering methods such as robustness, ability of handling missing data and the lack of the need to define the number of clusters in advance or to study the underlying distributions (Bloom, 2014).

3.5 Customer segmentation using clustering analysis

While the supply chain segmentation frameworks introduced in chapter 2 mainly uses rule-based approaches and only few different parameters to define the suitable supply chain strategies, customer segmentation as such is often done by different clustering algorithms. One of the most used ways of segmenting the customers by marketers and widely also in academic studies is by using the RFM-variables (recency, frequency, monetary) (Olson & Chae, 2012). The RFM-variables describes how recently the customers has made purchases (recency), how often customers make purchases (frequency) and how much money does the customers spend on purchases (monetary). In this thesis the objective of the segmentation is to define what kind of supply chain strategies should be used with the customers and thus, the customers are segmented using the variables and the principals used in the supply chain segmentation frameworks introduced in chapter 2 rather than using the RFM variables.

Since the efficiency of the various clustering techniques depends on the application and the characteristics of the dataset, it is impossible to know beforehand which algorithm fits the clustering task the best (Zakrzewska & Murlewski, 2005). The introduced algorithms have been used widely in academic customer segmentation studies. K-means is the most often used clustering algorithm for market segmentation due to its simplicity and efficiency on large dataset (Zakrzewska & Murlewski, 2005). FCM on the other hand is similar to KM with the difference that FCM enables through soft clustering some overlapping with the clusters and hence describes the customers in a more realistic manner (Yuliari, Putra & Rusjayanti, 2015; Munusamy & Murugeasan, 2020; Omran et al., 2007). The hierarchical clustering method is believed to provide better quality clusters than k-means especially with smaller datasets (Hung, Lien & Ngoc, 2019). Finally, the SOM algorithm is used generally because of the clear representation of the similarities of the customers without a need of further analysis (Bloom, 2014; Li & Lin, 2005).

3.6 Defining optimal number of clusters

The number of clusters play an essential role in clustering analysis. As an unsupervised learning method, it is crucial to be able to validate the goodness of partitioning after the clustering (Aggarwal & Reddy, 2013). Whether the specific dataset is clustered into three or five clusters might result to two completely different outcomes. Unfortunately, there is no consistent and conclusive solution to find the optimal number of clusters. Rather defining the optimal number of clusters as case-specific attribute meaning that the optimal number of clusters must be defined separately for every clustering application. However, defining the optimum number of clusters might be challenging since it requires priori knowledge, or the ground truth about the data. Most of the clustering algorithms cannot determine the optimal number of clusters by themselves and thus the number of clusters must be specified in advance. (Mahamed, Engelbrecht & Salmam, 2007)

Two criteria measuring the quality of data partitioning are generally accepted. A high-quality cluster is believed to be both *compact* within individual clusters and *separated* from other clusters. Compactness means that the patterns in one cluster are similar to each other and different from the patterns of the other clusters. In other words, having small variance within a cluster is equivalent to having a compact cluster. Simultaneously with compactness, a good cluster must be well-separated from other clusters. The Euclidean distance between the cluster centroids along with other distance measures can be used as an indicator of the separation of the clusters. (Mahamed et al., 2007; Aggarwal & Reddy, 2013; Liu, Li, Xiong, Gao & Wu, 2010; Omran et al., 2007)

Numerous of different clustering validation measures have been proposed for different clustering applications since no consensus has been reached. The validation measures can be roughly categorized into two families: internal- and external clustering validation measures. The internal cluster validation measures evaluate the goodness of the partitioning using only the internal information present in the dataset, whereas the external cluster validation measures use external information such as labels or categorization made by experts. (Aggarwal & Reddy, 2013)

In this thesis the only prior external information about the amount of the clusters and the labeling comes from the supply chain segmentation literature. Even if the supply chain segmentation literature reviewed in chapter 2 gives a good indication of the typical supply chain strategies, their characteristics and even a rough amount of the plausible different strategies, segmenting the supply chain is not unambiguous. Therefore, the external clustering validation measures cannot be used. Instead, the internal clustering validation measures are used in this context. A wide set of different internal clustering validation measures are used to define the optimal number of clusters. Liu et al., 2010 introduces the 11 most used internal clustering validation measures from which three measures are introduced in more depth in **Table 7**.

Table 7. Some of the most used internal clustering validation measures (Modified from Liu et al., 2010; Hassani & Seidl, 2016)

Measure	Notation	Definition	Optimal value
Dunn-index	D	$min_i\{min_j\left(\frac{min_{x\in c_i, y\in c_j} d(x,y)}{max_k\{max_{x,y\in c_k} d(x,y)\}}\right)\}$	Max
Calinski-Harabasz index	CH	$\frac{\sum_i n_i d^2(c_i, c)/(NC-1)}{\sum_i \sum_{x\in c_i} d^2(x, c_i)/(n-NC)}$	Max
Silhouette index	S	$\frac{1}{NC}\sum_i\{\frac{1}{n_i}\sum_{x\in c_i}\frac{b(x)-a(x)}{\max[b(x), a(x)]}\}$	Max

Where:

D: dataset; c: center of D; n: number of objects in D; $d(x, y)$: distance between x and y; k: cluster number; NC: number of clusters; C_i: the i:th cluster; c_i: center of C_i; n_i: number of objects in C_i

The typical procedure of defining the optimal number of clusters using the internal clustering validation measures goes as follows. First the applied clustering algorithms are initialized and if needed, tested with different parameters such as fuzzifier value in FCM, to get several results. Afterwards the corresponding internal clustering validation indexes are computed for each partition and finally the best partition is selected according to the criteria. All measures presented in **Table 7** are based either on measuring the compactness, separation or both simultaneously. Even if all clustering validity measures are built on the same quality criteria, they measure the compactness and separation using slightly different underlying logic. Dunn's index uses the minimum pairwise distance between objects in different clusters as a separation measure (inter-cluster) and the maximum diameter as the compactness measure within the clusters (intra-cluster). The Calinski-Harabasz index on the other hand uses the average of inter-cluster- and intra-cluster- sum of squares to evaluate the cluster validity. Both clustering validity indices takes therefore a form of *Index=(a*Separation)/(b*Compactness),* where *a* and *b* stand for weights. Slightly differing from the validation indices mentioned above, the Silhouette index is based on the pairwise difference of inter-cluster and intra-cluster distances. (Liu et al., 2010) The Silhouette index formula can be also written as follows:

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1, & \text{if } a(i) > b(i) \end{cases} \tag{3}$$

As $a(i)$ measures to dissimilarity of i to its own cluster, a small value indicates that i is well matched to its cluster. Respectively, a large $b(i)$ implies that i fits poorly into its neighboring cluster. Thus, the closer the silhouette score $s(i)$ is to 1, the more properly the dataset is clustered. A negative silhouette score, on the other hand, implies that i should rather belong to its adjacent cluster than the cluster it is assigned and therefore indicates of a poor clustering result. Moreover, silhouette score equal to 0 indicates the border of two separate clusters. Despite the differences of the indices listed above, all of these determines the optimal number of clusters by maximizing the value of the indices (Liu et al., 2010).

4 CLUSTERING WITH A REAL DATASET

The purpose of this chapter is to cluster the customers into different segments using the theory of the previous sections among with Stora Enso specific idiosyncrasies. The variables used in the clustering analysis are selected to match the vital characteristics recognized in the numerous supply chain segmentation frameworks presented in chapter 2. As supply chain strategies are always content-specific (Godsell et al., 2011), it is not intended to implement directly any single supply chain segmentation framework with Stora Enso's customers, but rather to engage best practices from the frameworks and enrich the segmentation with suitable clustering algorithms to outcome fully tailored service models for the customers. The main steps of creating the tailored supply chain segments are described in **Figure 17**.

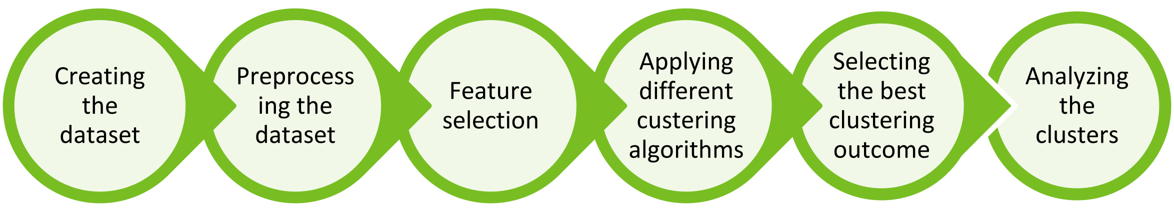


Figure 17. The main steps of the supply chain segmentation process used in this thesis.

The initial dataset is a one-year order data where each row stands for a single order. The order-level data is later aggregated in such a way that every customers delivery location (consignee) is one entity that is defined by a set of attributes **Figure 18**. A single customer might have several different consignees, but the entity is aggregated on the consignee -level rather than customer level since the behavior of the consignees might drastically differ from each other even within a single customer. This is because the consignees might be producing packaging for entirely different end-users and end-uses, and therefore the varying pull of the different markets affect the customers ordering behavior and requirements significantly. Therefore, aggregating the data on the customer level would give biased results, which would not serve the segmentation application. For the sake of consistency and simplicity, later in this thesis consignees are referred as customers.

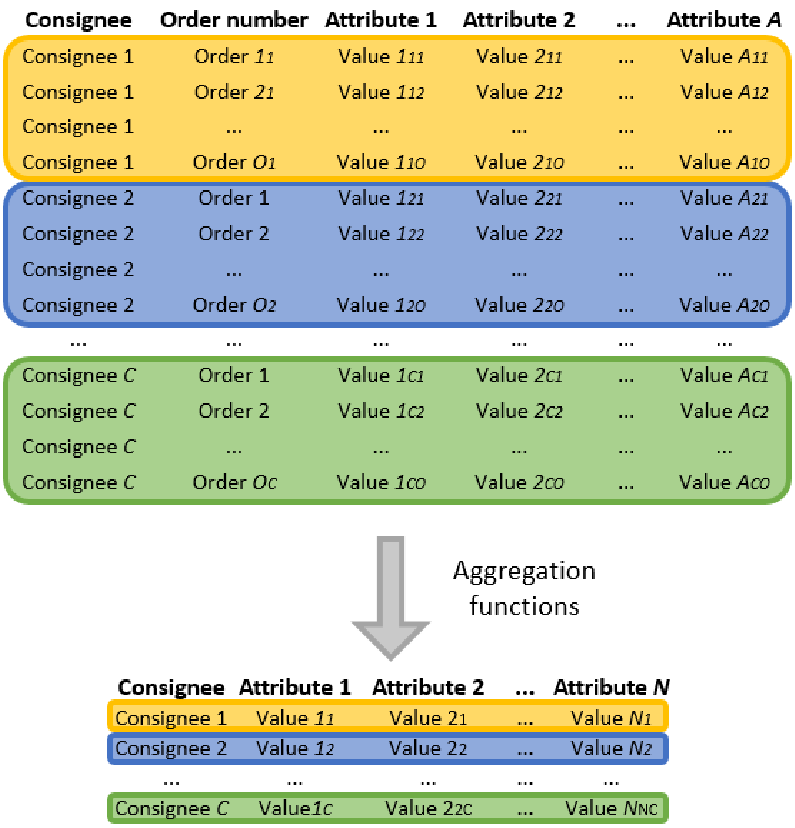


Figure 18. Aggregating order-level data into customer-level

Where:

C: number of unique customers; O_n: number of orders for n:th customer; A: number of different attributes before aggregation; N: number of attributes after aggregation

4.1 Data description

To summarize the findings of chapter 2, the central attributes used in the supply chain segmentation frameworks are listed in **Figure 19.** The attributes are mapped in a 3x3 matrices by evaluating their importance to the corresponding supply chain segments and by observing the incidence of the attributes used in the different frameworks.

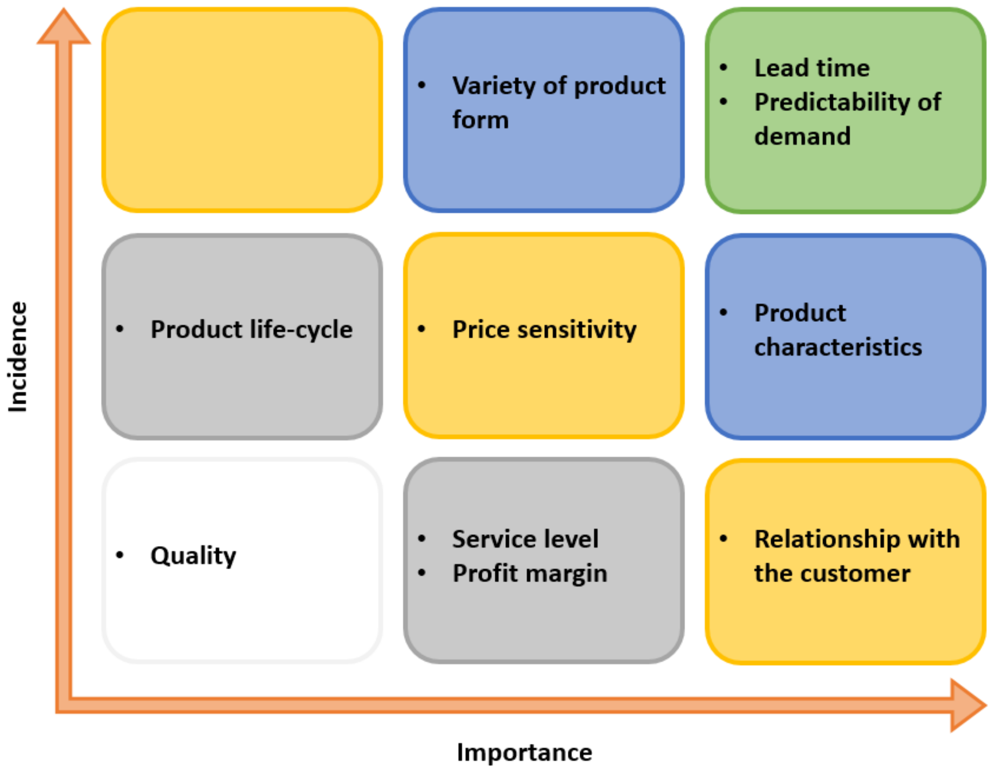


Figure 19. The pivotal attributes used in supply chain segmentation frameworks

As perceived from **Figure 18**, the lead-time and predictability of demand are the most critical attributes used in the supply chain segmentation as they are used in several frameworks and usually in a pivotal role (high incidence and importance). Lead-time is a commonly used measure of responsiveness usually measured from the time of order placement to the discharge on the customer site. The predictability of demand in business to business markets correspondingly gives an indication of the continuity of orders and therefore enables the production planning further to the future and reduces costs. These two attributes alone describe a considerable amount of customer behavior and their expectations regarding supply chain performance. Customers with predictable demand can more likely be served with cost optimizing lean procedures whereas unpredictable demand leads to higher production costs through exceptions in the production. The differences in lead-time reflect not only the production costs, but also the supply related costs through the exceptions in transportation modes. Lead-time and predictability of demand therefore reveals a lot of customers business in general, their planning capability, and price sensitivity.

Other fundamental attributes used in the supply chain segmentation frameworks among lead-time and predictability of demand, are the product characteristics and the relationship with the customer, where the product characteristics appear in several frameworks and the relationship with the customer appears only in one. The product characteristics is widely used in the product-driven supply chain strategies whereas the relationship with customer is used in the frameworks focusing on dynamic alignment. Regardless of the different schools of supply chain segmentation, these attributes play an essential role in the frameworks they are used in. The product characteristics used in the segmentation frameworks pursue to describe the nature of the product and this way it steers towards the suitable supply chain strategies. Characteristics such as the partitional demand of the SKU gives an indication of the level of customization of the product and how can economies of scale be utilized in the production and supply chain. The relationship with the customer becomes a key attribute when the supply chain segmentation strives for dynamic alignment. To be able to consider the supply chain as a living organism that evolves over time rather than a fixed mechanical structure, the relationship with the customer, as well as other aspects of customer behavior, needs to be considered as the reference point (Gattorna, 2009). Whenever customer behavior changes, the supply chain should dynamically adjust to the new needs. The ability to adjust the supply chain to the changes in the relationship with the customer allows a more flexible and precise service portfolio.

Besides the attributes already mentioned, other important attributes that are used in the supply chain segmentation literature are the variety of product form, price sensitivity, service level, and profit margin. The variety of product forms occurs in several product-centric segmentation frameworks since it gives a relatively good indication of how effectively the production lines can be run. The price sensitivity of a customer in turn, is a significant feature that defines the willingness to pay for additional services. However, it is not necessary to consider the price sensitivity in isolation, since it is by its nature incorporated in several other attributes. In a large-scale segmentation application, it might be even hard or impossible to define the price sensitivity for all customers. Just like with price sensitivity, the service level might include aspects that are hard to measure like customer service quality or flexibility. As the expected service level is hard to directly model, we can use simplifications and measure the customers expected service level indirectly by observing the usage of additional services.

Just like Godsell et al., (2011) discovered, the best practice creating supply chain strategies is to use only the case-specific dominant variables and leave the less important variables out of the analysis. Delimiting the less important variables out of the analysis improves the clustering result and creates more meaningful segments. For Stora Enso specific supply chain segmentation purpose the profit margin, product life cycle, and quality are considered irrelevant and therefore they are left out of the analysis. Even though the profit margin is used as one of many features to create product clusters in supply chain segmentation frameworks, it is left out of this analysis because of the sensitivity of profit information in general in business. The product lifecycle, on the other hand, is disregarded since different SKU's in carton business have very similar lifecycles so there is no reason to take this metric into consideration when segmenting the supply chain. The product quality is likewise omitted of the analysis since all products ordered are primarily first grade.

The attributes selected in the analysis are picked to describe compendiously the most pivotal features recognized from the supply chain segmentation literature. The limited availability of data along with the nature of Stora Enso's cartonboard business sets the definitive framework for the feature modeling. As the supply chain segmentation literature is mostly general with some deep-dives into specific companies or businesses, a one-size-fits-all segmentation solution isn't neither valid nor desirable. The attributes used in the analysis are modeled to describe the desired features case-specifically from a Stora Enso point of view.

4.2 Data preprocessing

Data preprocessing is a crucial step in every data analysis application. The quality of the underlying data is directly linked to the quality of the analysis. Data preprocessing is a necessary and time-consuming task that pursues to process the dataset for the clustering algorithms by cleaning the data from erroneous values such as noise, missing values, or inconsistency (Zhang & Yang, 2003). Data inaccuracies may be due to, for an example, loose data quality management, false measurements, or some other irregularity in the data creation process. In this thesis the preprocessing of the data is done in two separate phases: before aggregating the order -level data into customer-level, and after the aggregation. The first part of the data preprocessing focuses entirely on cleaning the data to ensure a high-quality result whereas the latter part addresses processing the features in a suitable way to get more informative distributions. **Figure 20** shows the different areas covered in the preprocessing phase.

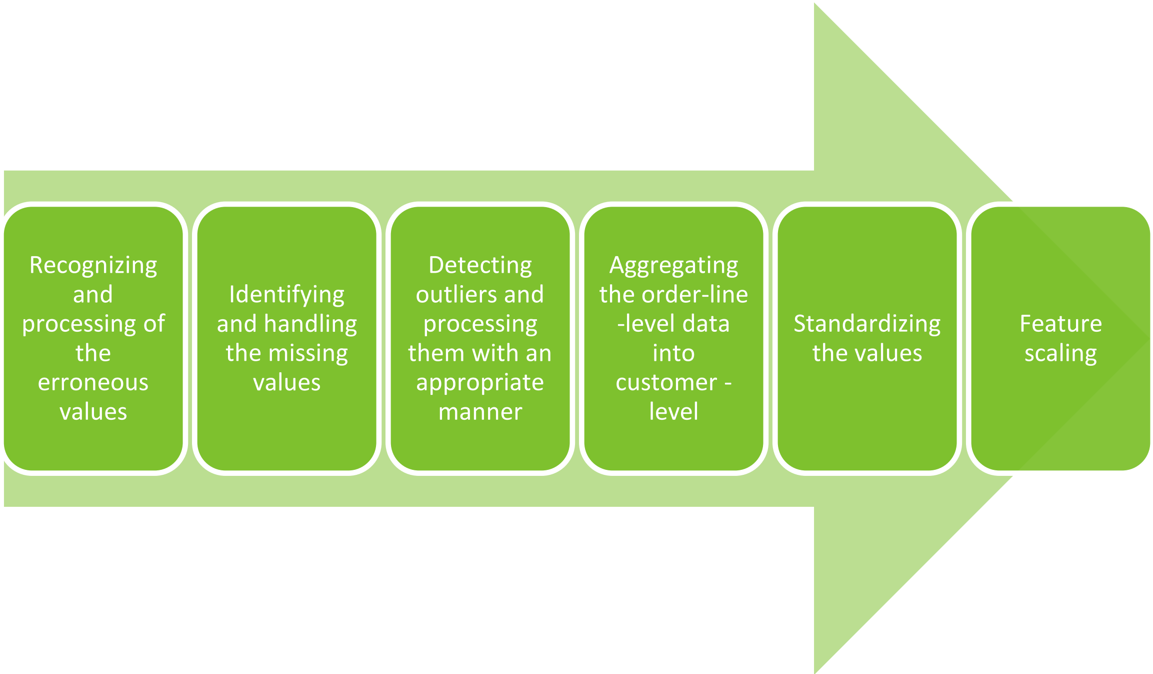


Figure 20. Different phases of preprocessing the dataset

The nature of the data, given the objectives of the analysis, influences the way the data can be pre-processed. As the dataset is eventually aggregated into a customer-level, deleting any records of the dataset after the aggregation will exclude the whole customer out of the analysis, which is not a desirable outcome since one of the objectives of the thesis is to find out what kind of customer profiles can be found from the current state analysis. Another case-specific special feature of the aggregated customer-level data is the high variability between customer characteristics. The ordering patterns between the customers are drastically different due to the different needs and sizes of the companies.

Even though the initial dataset is generally high-quality data, it doesn't change the need for preprocessing. A real-life business is always prone to incorrect data especially if the data is created by humans or it is coming from multiple business processes such as production, logistics, or sales as it is in this case. Even though processes are well planned and executed, businesses may face exceptional situations that shows in the data as inconsistency. In the first part of the preprocessing the non-plausible values are defined for each of the attributes. The records containing erroneous values, such as negative quantities or negative lead-times are an outcome of an exception or an error, which indicates that the whole order line might contain corrupted information and therefore these rows were deleted from the dataset.

The next step after removing the biased orders from the dataset was to handle the missing values in an appropriate manner. By studying the dataset, we find out that the lead-time is the only attribute containing missing values and therefore the only attribute that requires processing at this point. As described in **Table 8**, lead-time is defined in the framework of this thesis as the time from order entry to the delivery on the requested location. From a data integrity perspective, the lead-time measure is sensitive when it comes to data preciseness and completeness. Unlike other attributes used in this analysis, lead-time is generated not only from one, but from two totally different sources: the sales company and the logistics carrier. This makes the lead-time measure the most challenging attribute, and hence it needs the most preprocessing. The last leg of the supply chain is not in all cases fully transparent for Stora Enso since it uses external carriers to deliver the orders. This leads to a noticeable amount of missing values of the actual proof of delivery (POD) date which is used in the lead-time calculation. A commonly used method for handling missing values is to replace the missing values with a value obtained from using some aggregation function over the other values within the same group (Auguinis, Gottfredson & Joo, 2013). Using an aggregation function for this specific use is slightly problematic since some specific legs are especially prone for the missing POD-times, and the aggregation for these customers would populate numerous values based on only a few actual records. Rather than replacing the missing values using some aggregation function, the lead-time is calculated for the records based on the difference between order entry and the planned POD-time.

Missing values are not the only challenge that we face when processing the lead-time. The other attributes used in this analysis, such as ordered quantity or number of different SKU's ordered, are unambiguous since they are derived from customer orders and hence represents the customer behavior. But when it comes to lead-time, the measure is calculated by the difference from order entry to the delivery date, which is not always explicit due to example amendments to the orders. The long tail of orders with unreasonably long lead-times (**Figure 21**) might be on account of different reasons such as a simple human or system mistake to some exceptions in the normal order delivery process. By processing this small portion of distorted records already before aggregating the dataset into the customer-level, we have the change to replace the biased records by some aggregation over the group and therefore get a more reliable result. The exceptionally long lead-times are detected by using the interquartile rule defined in (4) iteratively. These outliers are replaced by calculation the mean value of all the other lead-times within the same customer. For a small portion of orders, a mean value couldn't be calculated due to having only distorted records and therefore these rows were deleted from the dataset.

$$
\begin{aligned}
Lower\ Outlier &= Q1 - (1.5\ x\ IQR) \\
Upper\ Outlier &= Q3 + (1.5\ x\ IQR)
\end{aligned}
\tag{4}
$$

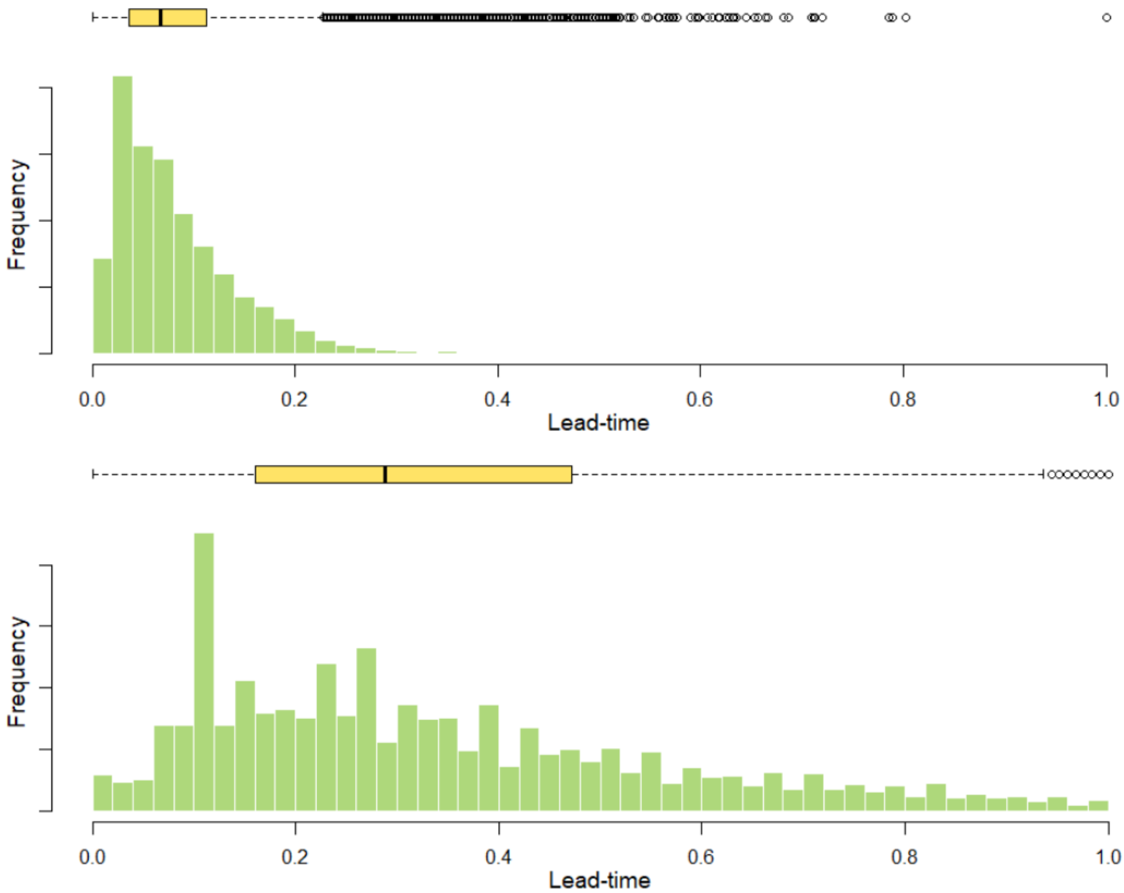


Figure 21. Distribution of lead-time before and after removing outliers

After the initial preprocessing of the order-level data, we can aggregate the dataset into customer-level. Even though the supply chain segmentation frameworks points out the features that are generally being used to define supply chain strategies such as lead-time and releation with the customer, they do not necessarily take any kind of stand on how should these features be used to fit the purpose. Therefore, the supply chain segmentation frameworks leaves it for the analyst to decide how the lead-time or relationship with the customer metrics are calculated. **Table 8** brings together all the attributes used in the further analysis along with the descriptions of how the attributes are built.

Table 8. The attributes selected for the analysis.

Attribute	Describes the feature	How is it built
Ordered quantity	Relationship with customer	The relationship with customer is captured in a simplified way by taking the sum of the ordered quantity.
VarOrderQuanQTR (Variability of Ordered Quantity between quarters)	Predictability of demand	The variability of demand per quarter is captured by using coefficient of variation. Rather than using all orders to calculate the mean and standard deviation, the ordered amounts are first summarized for quarters and the values are calculated from quarterly demand. This way the predictability is captured over longer period.
SKUCount (Distinct Count of SKU's)	Variety of product form	Distinct count of the unique SKU's per customer.
OrderFRQ (Order Freaquency)	Predictability of demand	Ordering frequency is calculated by dividing the total number of orders by 365
AVGLeadtime (Average Lead-time)	Lead-time	The average lead-time is calculated as the mean of all customer orders lead-times. Lead-time is defined as the time from order entry to the discharge in ordered location.
Q1Leadtime (First Quartile Lead-time)	Lead-time	The Q1 lead-time is the first quartile of all customer order lead-times.
AVGSkuType (Average SKU Type)	Product characteristics	The average SKU type is calculated by labeling first the SKU's according to pareto distribution into interval]0,1[and calculating the mean value from all the SKU's ordered within one year.
RFLRTA-Ratio (Ratio of RFL Orders)	Service level/Price sensitivity	Dividing the count of RFL-orders by the count of all orders to get a ratio of how much interim warehousing is used.

When using unsupervized learning methods like clustering, it is not always easy to determine what features are relevant for the result since an explicit accuracy measure cannot be calculated. One way to remove redundancy is to investigate the correlations between the features. Plotting a correlation matrix (**Figure 22)** gives immidiately an indication of whether the different features describe the same behavioural aspect. The high correlation between two separate variables indicates that the variables moves in the same direction.

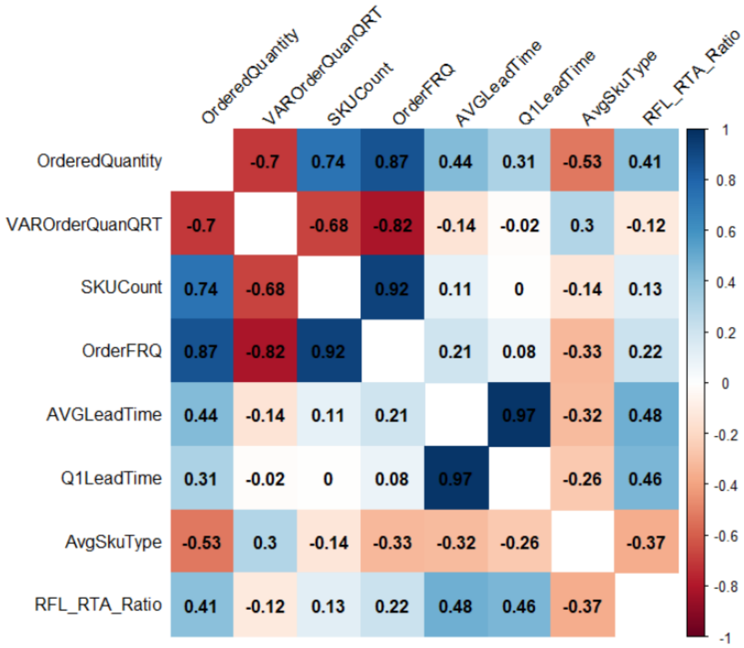


Figure 22. Correlation coefficients of selected variables

By analyzing the correlogram (**Figure 25)**, we discover that "*SKUCount*" and "*OrderFRQ*" strongly correlates between each other (correlation coefficient 0,92) as well as "*AVGLeadTime*" and "*Q1LeadTime*" (correlation coefficient 0,97). The strong correlation between these variables indicates that these variables does not bring any additional information to the model and rather might disort the significance of a single behavioral aspect. To avoid giving too much weight for strongly correlating variables, we can drop one of the correlating variables out of the dataset. To choose which variable to drop out of the dataset, we calculate the mean absolute correlation between the selected attributes and the remaining attributes (**Table 9)**.

Table 9. Mean absolute correlations between selected attributes

Variable	Mean of absolute correlations
SKUCount	0.46
OrderFRQ	0.55
AVGLeadTime	0.46
Q1LeadTime	0.39

From **Table 9** we see that the mean of absolute correlation between *SKUCount* and *OrderFRQ* as well as between *AVGLeadTime* and *Q1LeadTime* doesn't differ considerably which supports the hyphothesis that these variables describes the same beavioral aspect. However, the existing difference between the mean of absolute correlations can still be used to define what variable to drop out of the dataset. In this case, *OrderFRQ* and *AVGLeadTime* are dropped out of the dataset since these variables have higher mean of absolute correlations.

Where the primary preprocessing focuses on cleaning the initial dataset for the aggregation, the second part of the preprocessing aims to process the attributes for the clustering algorithms. Exploratory data analysis provides more information on the distributions of the data, and thus helps us understand the data better. By observing the distributions both visually and statistically, we can draw conclusions about how the data should be processed for the algorithms.

Falkenhausen, Fleischmann, and Bode (2019) noticed that the distributions of the DWV3 (Duration, Window of delivery ,Volume, Variety & Variability), variables are usually highly skewed, and the knowledge of cartonboard business also supports this hypothesis. Not surprisingly, the highly unbalanced distributions can be immidiately confirmed from the exploratory data analysis. Like in many other companies, the distribution of customers in terms of sales follow a pareto-like distribution (**Figure 7**) meaning that roughly 20% of the customers makes 80% of sales and vice versa. As the vast majority of customers are small regarding sales, the attributes such as the count of ordered unique SKU's and ordered quantity are weighted near the minima of the distribution. On the other hand, the values of the attributes for the biggest customers are extreamly high making the tail of the distributions really long (**Figure 23**). Since the use of actual production data is considered to be rather sensitive, the data is always normalized before plotting by min-max-normalization into scale [0,1], to avoid expozing sensitive information.

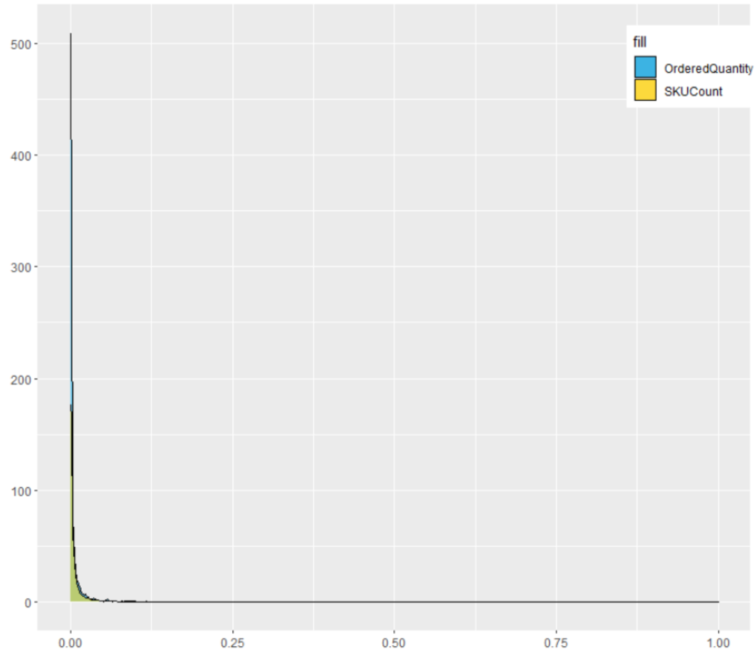


Figure 23. Pareto-like distributions of attributes

Without any standardization the distance based clustering algorithms would create own clusters around the few extreme outliers and the vast majority of customers would be clustered as one huge cluster. From a supply chain strategy perspective, it is not a desired outcome since the majority of customers would be served on the same service model even if their behaviour and expectations would vary greatly from one an other. The information itself, whether the ordering freaquency for an example is high or low, is significant and reveals much about the customer ordering patterns, but the skewness of the distribution prevents the usage of these attributes in clustering algorithms directly after normalization. Falkenhausen et al., (2019) overcome the skew in the distributions by using log-transformation over the attibutes. By taking natural logarithm over the attributes the skewness of the distribution can be reduced and hence the clustering algorithms gives more meaningfull results. From **Figure 24** we can see the distributions of the attributes having pareto-like distributions after the logarithmic transformation.

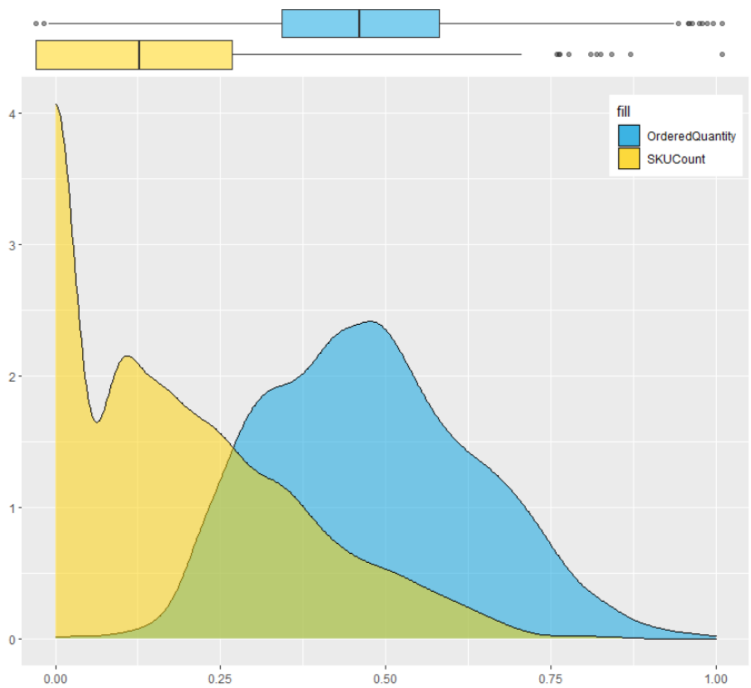


Figure 24. Pareto-like distributions of attributes after log-transformation

Not all attributes have pareto-like distributions and hence can be processed by log-transformation. Some of the attributes are aggregated from binary data which results to having binary-like distributions also in the aggregated dataset like shown in **Figure 25**. As the attributes are by their nature binary-like, modifying the distributions by removing or processing the outliers would skew the information content and hence these attributes are not processed anyway in this phase.

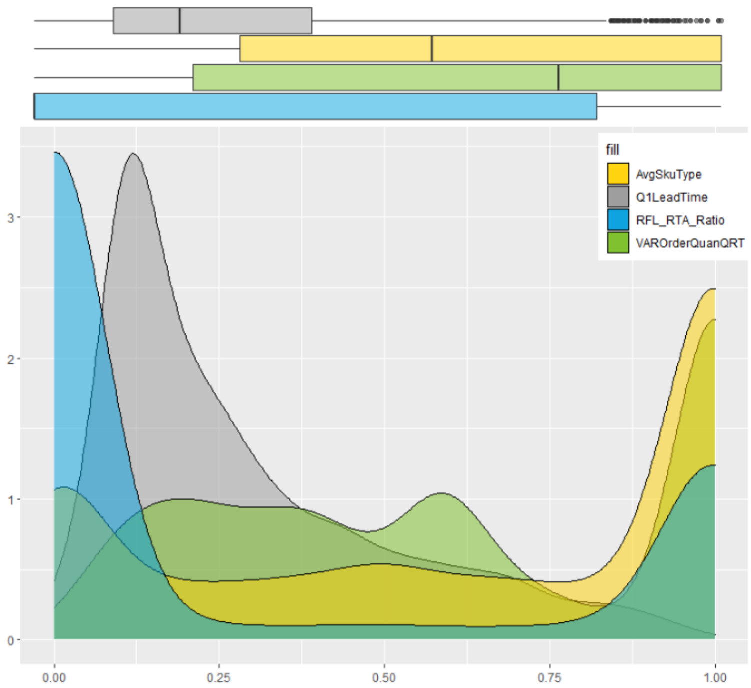


Figure 25. Attributes with other kind of distributions

Since the erroneus values were processed already in the initial dataset, we can with confidence state that the outlier values in the modelled attributes truly describe the customer behaviour and hence there is no reason to cap them outside of the analysis. The outlier processing in this analysis is balancing between removing information from the dataset and keeping statistical significance high. Excluding or capping the outliers might disort the result by removing information about the distributions and hence forcing the distribution to seem less variable than they in reality are. On the other hand, keeping the outliers in the analysis might reduce statistical significance of the analysis. For the attributes that resembles pareto-distribution the log-transformation modified most of the outliers as meaningful values, leaving only a few outliers.

4.3 Testing different clustering algorithms

The dataset is run through KM, FCM, HC, and SOM algorithms to find out the best fitting algorithm for this specific dataset. The goodness of the clustering results is evaluated by using the internal clustering validity indexes presented earlier in chapter 3. Testing the different clustering algorithms is important for estimating the algorithms ability to generalize to future observations.

4.3.1 Initializing the clustering algorithms

Each of the clustering algorithms require their own set of parameters that define how the algorithms work. As mentioned in chapter 3, hierarchical clustering (HC) can be run by agglomerative or divisive method. The agglomerative hierarchical clustering algorithm can be run by several different linkages, from which single link and complete link were introduced previously. To choose which hierarchical clustering method to use, divisive clustering method, as well as agglomerative method with both single and complete linkage, are applied to the dataset. The quality of the clusters is evaluated by using the Silhouette index (**Appendix 2**). From the comparison we can see that the divisive method creates higher quality clusters than either one of the agglomerative methods, and hence the divisive method is selected as the HC method that will be compared to the other clustering algorithms.

In a similar way than with hierarchical clustering, the k-means algorithm is applied to the dataset using *Hartigan-Wong*, *Lloyd*, *Forgy*, and *MacQueen*- algorithms and the best performing one is selected for further comparison with other algorithms. From **Appendix 3** we see that the differences between the algorithms are marginal, as the Silhouette index for *Lloyd* method is slightly better that the other ones, it is selected as the k-means algorithm to use. Fuzzy c-means on the other hand, is initialized by using squared Euclidean distance as a distance metric. Even though 2 is a generally used default fuzzifier value, fuzzifier values equal to [2,3,5,10] are also tested against silhouette score for consistency (**Appendix 4)** (Schwämmle & Jensen, 2010). The silhouette score is the highest when fuzzifier value m = 2, and therefore it is selected as the fuzzifier value. With SOM, a hexagonal 20x20 grid is used, and the dataset is presented to the neural network 100 times. Learning rate for SOM is set to decline from 0.05 to 0.01, which is the default learning rate range for the algorithm. To cluster the grid of nodes generated by SOM further into clusters, divisive hierarchical clustering is used.

4.3.2 Measuring the fit of the clusters

While k-folds cross validation is usually used in classification models to evaluate the model accuracy more precisely by avoiding overfitting, k-folds is used in this context to divide the dataset in five samples. The idea is to run the clustering algorithms five times, each time with a different sample of the dataset. The internal clustering validation measures are then calculated for each fold, and the average values of the validation measures are presented in **Figures 26-**

28, where the higher (lighter) values indicate the best performing number of clusters for each of the four clustering algorithms. **Figure 26** shows the Dunn index, **Figure 27** shows the Calinski-Harabasz index, and the **Figure 28** shows the Silhouette index for the different clustering algorithms and number of clusters**.** As mentioned in chapter 3, for each of the following clustering validity indexes used in **Figures 26-28**, the optimum value is the maximum value, meaning that higher values indicate on average higher quality clusters. As the clustering validity measures are not comparable to each other as such, we can't compare the values between Dunn-, Calinski-Harabasz-, and Silhouette indexes with each other. Rather, we can compare the validity scores of different clustering algorithms and number of clusters within a single clustering validity index. This way, we find out which of the clustering algorithms are performing the best and with what number of clusters.

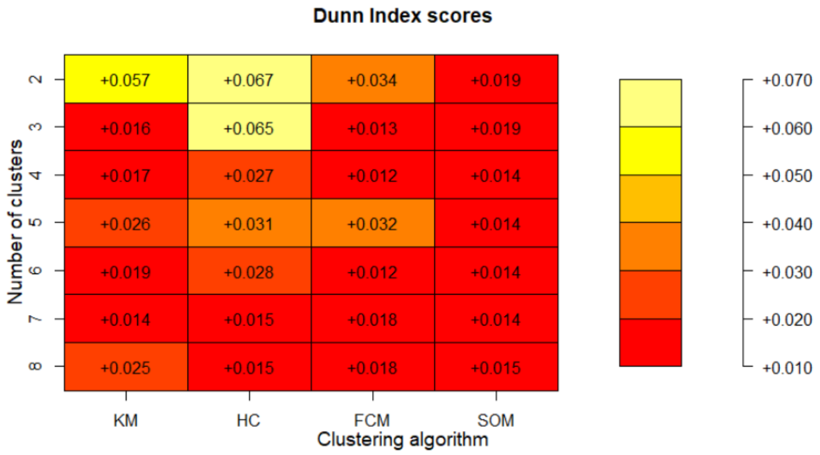


Figure 26. Average Dunn index for tested clustering algorithms

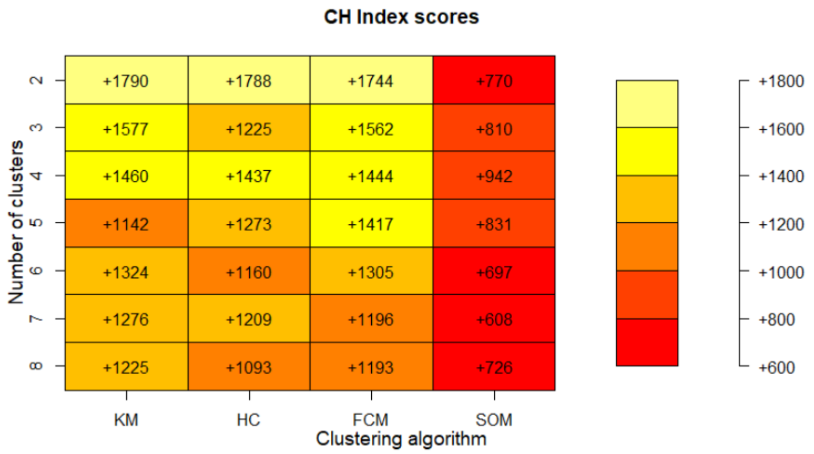


Figure 27. Average Calinski-Harabasz index for tested clustering algorithms

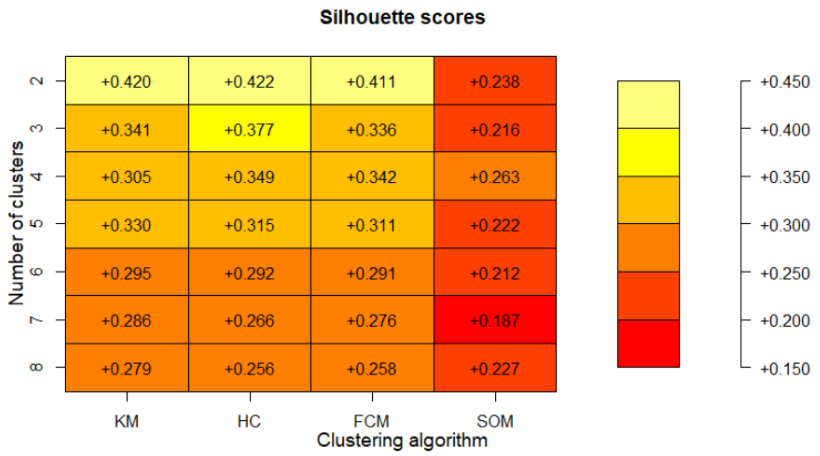


Figure 28. Average Silhouette index for tested clustering algorithms

As we can see from **Figures 26-28,** the four clustering algorithms provides a wide range of possible outcomes. Depending on what clustering algorithm and clustering validity measure is used, the quality of the clusters varies noticeably. Noticeable is that three out of four algorithms, the k-means, fuzzy c-means, and hierarchical clustering methods returns relatively similar results in addition to which they clearly outperform the self-organizing maps regarding the internal clustering validation measures. As a consensus of the best performing algorithm based on **Figures 26-28** is hard to define, the decision is done based on an aggregation score. Firstly, the average index values are computed for each clustering algorithm and index, followed by normalization of each index over the same interval [0,1]. Lastly, an average over the normalized averages is taken, which leaves us with one value per clustering algorithm, which are presented in **Table 10.** Based on the aggregated score, the HC seems to be clustering the dataset slightly better than the KM or FCM and is hence chosen as the algorithm to proceed with. Another observation from **Table 10** is that the aggregation score for SOM is zero, meaning that it was the worst-performing clustering algorithm based on each clustering validity index.

Table 10. Defining the optimal clustering method

Clustering method	Aggregated clustering validity index
KM	0.80
HC	0.95
FCM	0.71
SOM	0

Another observation from **Figures 26-28** is that the clustering algorithms tend to prefer lower number of clusters. Two is the most preferred number of clusters, but also from three to five clusters seem to be a plausible solution especially based on Calinski-Harabasz index and Silhouette index. This is in line with the supply chain segmentation literature, where the proposed number of supply chain strategies varied from two to five depending on the framework. Purely based on the three cluster validity indices, two would be an obvious number of clusters to proceed with. However, as the trend in the supply chain segmentation literature shows a movement from simple lean-agile-split towards identifying more specific and even overlapping segments, the approach focusing entirely on clustering validity indexes may not provide the desired result. For this reason, the number of clusters is selected by a combination of clustering validation indices and the supply chain literature. As the most recent supply chain segmentation frameworks indicates that a sufficient number of clusters would be between four to five, the decision of the number of clusters to use is done after analyzing the silhouette scores for HC algorithm visually (**Figure 29)**.

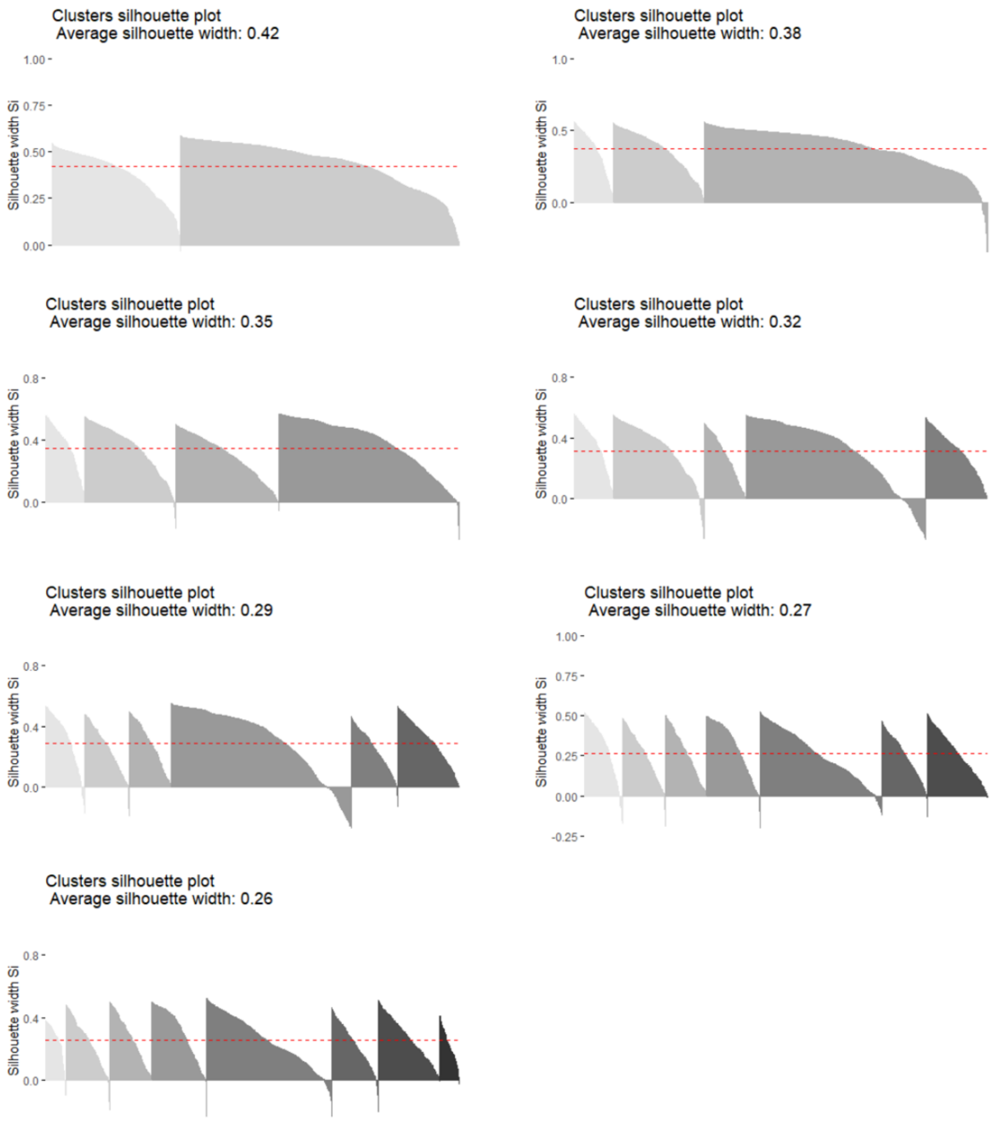


Figure 29. Silhouette scores for HC algorithm with different amount of clusters

As mentioned previously, the Silhouette index can return values between [-1, 1], where high values indicate that the datapoints are similar to the cluster they belong in and dissimilar to the other clusters, and negative values indicates that the datapoints might not be assigned to the optimal cluster. In the Silhouette plot (**Figure 29**), all the datapoints (customers) are listed in the x-axis and the y-axis represents the corresponding silhouette score for the specific customers. To make the plot easier to read, the datapoints are grouped by the cluster and within each cluster sorted by the silhouette width in descending order. The different shades of gray stand for the different clusters and the length of the colored area represents the number of customers assigned to the specific cluster. While the colored shapes represent the silhouette scores of all individual customers, the red dotted line shows the average silhouette width, which is also presented numerically above each visualization. Due to the sensitivity of the customer related data, the colors and cluster names are removed from **Figure 29** and **Figure 30** to prevent matching the cluster size and the cluster descriptions together.

As we can see from the **Figure 29** the average silhouette widths decrease gradually from 0.42 to 0.26 when increasing the number of clusters from two to eight. The differences in the silhouette scores when increasing the number of clusters by one is minor, which is why maximizing the average Silhouette width should not be considered as the ground truth, but rather as an indication of what solutions provide a plausible result. From **Figure 29** we see that in the two-cluster solution, cluster 1 is noticeably smaller than cluster 2. When moving to the three-cluster solution, the differences in the sizes of the clusters grows even more as it looks like the previous cluster 1 is further split into two clusters. Another observation is that part of the datapoints assigned to the cluster 3 forms now a short tail, where the silhouette score is below 0, indicating that the customers might not be assigned to the optimal cluster. When clustering with four clusters, the previous cluster 3 is in turn split into two separate clusters, and at the same time the short tail of negative silhouette scores is now shrunk into smaller tails. Because of this, even if the silhouette score of the three-cluster solution is 0.03 higher than the four-cluster solution, the four-cluster solution seems to be a more appropriate solution. If we continue to increase the number of clusters into five clusters, the tails of negative silhouette scores grow again, and together with the slightly lower average Silhouette width, this solution doesn't seem suitable. The average Silhouette width of the solutions with five or more clusters are already noticeably lower than the four-cluster solution, and while they don't get any support either from the supply chain segmentation literature, they can be ruled out. Based on the recent trends in supply chain segmentation identifying more than two supply chain strategies, and the comparison made with different number of clusters and different clustering algorithms, the further clustering analysis uses divisive hierarchical clustering method with four clusters.

4.3.3 Obtain the results

While the sizes of the clusters can already be seen from the Silhouette plots, dendrogram gives us more information about the similarity of the clusters. When analyzing the dendrogram presented in **Figure 30**, we see how similar the customers are in different clusters. As the dendrogram is built using divisive method, the algorithm has divided the most heterogenous clusters from the top and ended up in the singular customer level. Hence, we can analyze which clusters are the most dissimilar to one another. The first split divides the dataset into clusters where the first cluster has the clusters 1 and 2, and the second cluster has the clusters 3 and 4, meaning that the dissimilarity between these pairs of clusters is higher than within each of the pairs. The second split in turn, splits the first cluster into clusters 1 and 2, and the third split is divides clusters 3 and 4 into separate clusters.

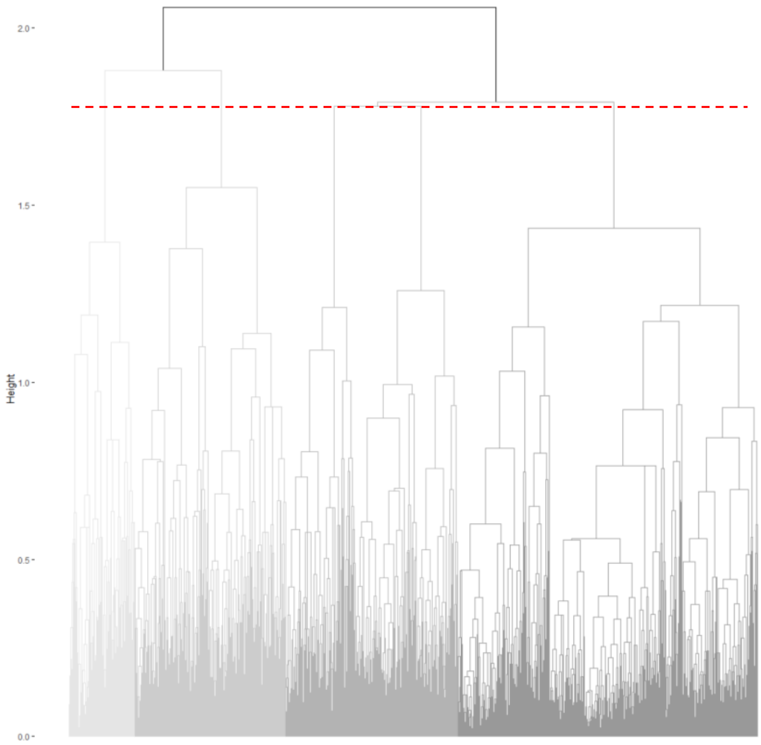


Figure 30. Cluster dendogram

To be able to understand the clusters in a more detail, we need to understand the distributions of each cluster and attribute. In **Figure 31,** the cluster-specific distributions are plotted for each attribute. As already mentioned, the order of the clusters as well as the colors are at this point randomized to prevent matching the cluster description to the cluster size. On top of the renaming of the clusters, the plot showing the distribution of *OrderedQuantity* attribute is removed due to its sensitivity. Despite not showing the *OrderedQuantity* distributions, the distributions are still discussed when describing the clusters. In **Figure 31,** each color stands for a specific cluster, and while the attributes are normalized between 0 and 1, we can easily understand and compare the relative differences of the clusters. The first observation from the fourth sub-plot, the distributions of *SKU Count*, is that the distributions between clusters are relatively similar and hence, it doesn't seem to affect the clustering result much. Another general observation is that the attributes *RFL_RTA_Ratio* (second sub-plot) and *AvgSkuType* (fifth sub-plot) seems to have the highest impact on the clustering results as the distributions between clusters differs significantly. This is understandable since the original distributions of these two attributes are U-shaped, meaning that most of the customers belongs in either end of the distribution. The *Q1LeadTime* (first sub-plot), *VAROrderQuanQRT* (third sub-plot) and *OrderedQuantity* (sub-plot not shown) on the other hand seems to have wide distributions with long, overlapping whiskers, which indicates that none of these attributes define the clusters alone, but rather together with other attributes. In addition, the long whiskers reveal that a single customer might not have just one, but several buying behaviors. Because the values are aggregated over all orders of specific customers, the values are not necessarily describing only one buying behavior, and hence the distributions aren't compact shaped. To understand the features of the clusters better, the clusters are discussed in more detail one at a time.

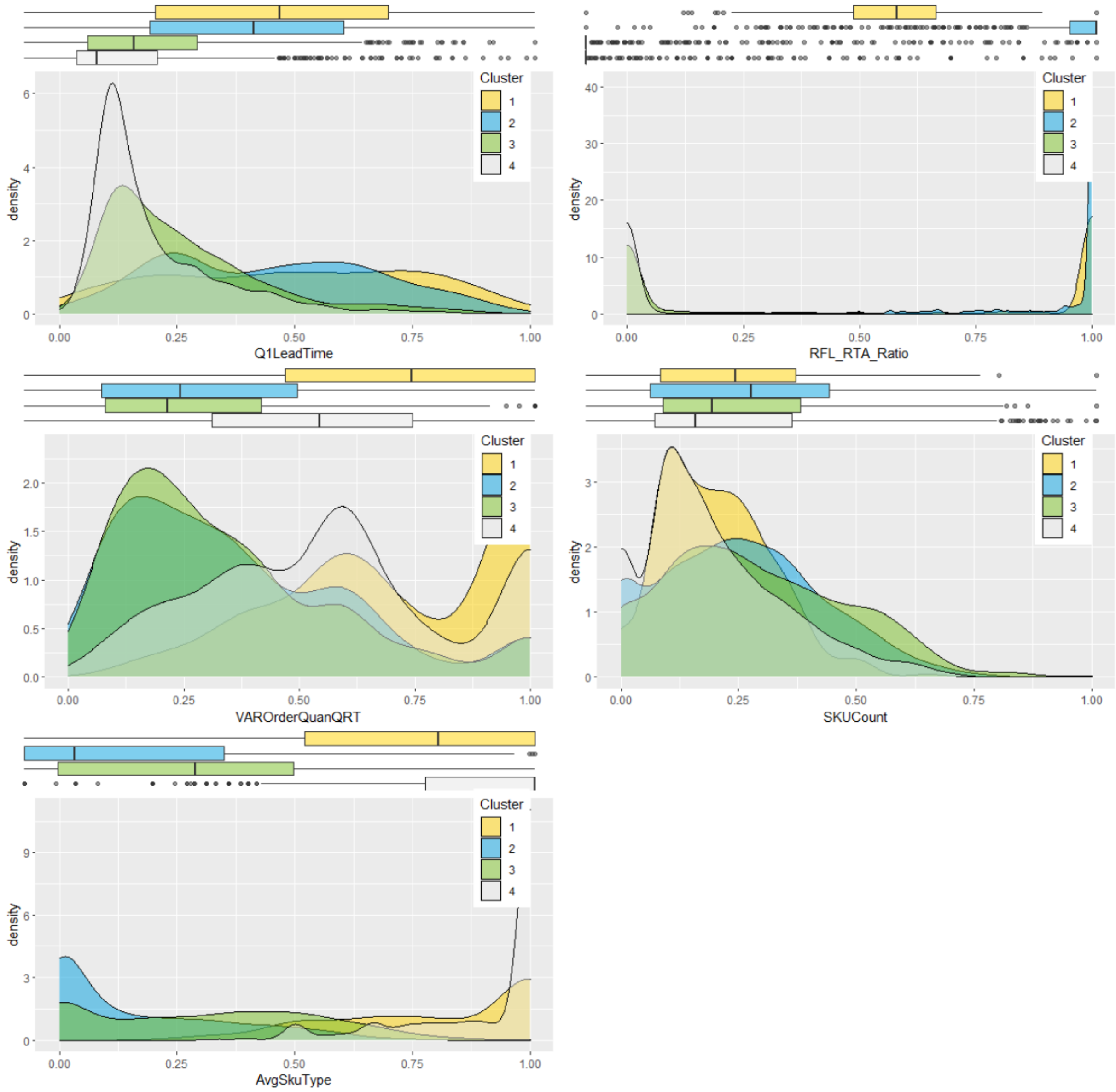


Figure 31. Distributions per cluster for each attribute

Cluster 1 (yellow)– "Collaborative/demanding"

From **Figure 31** we can see that customers belonging to cluster 1 has a high *RFL_RTA_Ratio* (sub-plot 2) indicating that they have a clear need for interim warehousing. The use of interim warehousing indicates that the order winning criteria for these customers is something else than cost, since the interim warehousing can be considered as an additional service. The usage of additional services seems to walk hand in hand with the customers excpectations towards the company. The demand of these customers are very unpredictable on quarterly basis (sub-plot 3) which might be due to, for an example, the seasonal nature of the customer business.

Moreover, the need for interim warehousing as an balancing element in the supply chain supports the hypothesis of seasonality in demand. Besides the unpredictable demand, the type of products customers in cluster 1 tend to order focuses mainly on the more rarely used SKU's, which can be seen in the distribution plots as high *AvgSkuType*-values (sub-plot 5). Due to the unpredictable demand and special SKU's, customers belonging in cluster 1 can by no means be served by a lean supply chain strategy. Rather, as this cluster seems to have elements from agile and continuos replenishment supply chain strategies, this cluster requires special attention from the supply chain. In addition, as the ordered quantity is rather high for cluster 1, the possible sales losses in case of not meeting the customer expectations are noticable. On the other hand, the customers high expectations together with the willingness to pay, can be seen as a possibility to build close relationships with the customers, as well as to build service bundles around the supply chain to serve the customer needs even better.

Cluster 2 (blue)– "Continuos replenishment"

By analyzing the cluster 2 from **Figure 31**, we notice that cluster 2 is relatively similar to cluster 1. Even though clusters 1 and 2 have some similarities with the usage of intreim warehouses, lead-time and the number of SKU's, these two clusters differ considerably from one another. Unlike cluster 1, the predictability of the demand for customers belonging in cluster 2 is relatively high (sub-plot 3). On top of the predictable demand, from sub-plot 5 we notice that customers belonging in cluster 2 tend to order mainly high demand products, which can be produced more efficiently. As the customers use interim warehouses to further stable the demand, these customers are "easy" in terms of managing the supply chain. Customers belonging in cluster 2 meet the criteria generaly identified in supply chain literature for lean supply chain, with the difference that interim warehousing causes additional costs. Noticable observation is that the upper whisker in the ordered quantity boxplot has the highest value indicating that this cluster holds inside very big customers. Customers belonging in cluster 2 can be considered to be generating the basic demand for the supply chain, and hence these customers should be able to be served as effortlessly as possible but still efficiently. Nevertheless, the capability of increasing the service offering for these cusomers has to be in place to be sure that the needs of key customers aren't in any situation overlooked.

Cluster 3 (green)– “Lean”

The cluster 3 meets almost all lean supply chain definitions. The demand is predictable, the ordered products are high volume products, the leadtime is not the shortest, and customers belonging in this cluster don’t use additional warehouse services. Noticable is that cluster 3 is the only cluster out of the four clusters where the lower whisker in the ordered quantity boxplot reaches 0, indicating that the smallest customers should be served by default with a lean strategy. On the other hand, as the distribution for the ordered quantity is wide and also overlapping with other clusters, we cannot state that all smaller customers would be served by the lean supply chain strategy. However, the lean supply chain can be considered as the default supply chain strategy to use if the customer doesn’t have any special requests. In the case of special criteria or requests from customers side, the customer naturally moves away from this cluster and is hence served with a more suitable supply chain strategy. “Growing out of the segments” is natural according to Gattorna (2009), and if the supply chain is dynamic, the used supply chain strategies can be changed whenever needed.

Cluster 4 (gray)– “Agile”

The fourth identified cluster differs from other clusters especially by short lead-times (sub-plot 1). On top of the short lead-times, customers belonging to cluster 4 orders almost exclusively specialized products. From the *AvgSkuType* boxplot (sub-plot 5) we see that with the exception of a few individual customers, all of these customers order mostly the pareto distributions long tail products. As the predictability of demand is also relatively low (sub-plot 3), it is clear that the cluster 4 stands for agile customers. For these customers the responsiveness is the most crucial part of the supply chain. While the combination of non-standard products together with rather unpredictable demand and no warehousing services cannot be supplied fast, while still remaining cost-effective, it is clear that these customers are willing to pay additional fees for fast and responsive delivery. On the other hand, to be able to utilize the full potential of agile-like buying behavior, the supply chain needs to be built on such level of flexibility that the more demanding needs can also be fulfilled.

5 CONCLUSIONS AND DISCUSSION

The objective of this study was to create supply chain segments based on the knowledge gained from the comprehensive supply chain segmentation literature review. On top of recognizing the distinguishing features that are used globally to shape supply chain strategies, the objective was to use advanced analytics methods to transform the customer-level set of attributes into meaningful “as-is” clustering analysis. The motivation was to find out whether the known supply chain segments can be identified using clustering analysis, and if not, to identify what kind of clusters do the customers form. Answers to the three research questions are presented below.

5.1 Answering the research questions

What kind of frameworks are introduced in the literature regarding supply chain segmentation in business to business (B2B) markets?

In summary, the area of supply chain segmentation has evolved through the past two decades substantially. The more recent frameworks are identifying more precise segments whereas the early 20th century supply chain frameworks were mainly focused on identifying two major kind of segments: lean and agile. The general idea of lean supply chain is to supply the products as cost-efficiently as possible by removing all non-value adding elements from the supply chain. Having the supply chain trimmed from all waste, it is possible to supply with minimal costs on the compromise of lead-time. Agile supply chain in comparison, focuses on a rapid time to market and high service level making this strategy more expensive to manage. Agile and Lean strategies are generally believed to be suitable for supplying different kind of products, which causes many frameworks to have high weight on the product characteristics when determining the right supply chain strategies. High volume products by default tend to favor lean supply chain whereas agile supply chain should be used to the long tail of low volume products.

Even if the more recent supply chain segmentation frameworks recognize more precise segments such as continuous replenishment or leagile, the supply chain segments are generally known to be always a tradeoff between efficiency and service level. As a result, the known supply chain strategies tend to position themselves in a curve describing the dependency between efficiency and lead-time, the so-called supply chain frontier. Instead of identifying more specific supply chain strategies, a more noteworthy transition in supply chain segmentation literature is the shift towards a more customer-oriented and dynamic segmentation. While the first supply chain segmentation frameworks defined the supply chain strategies exclusively based on product characteristics, the more recent frameworks emphasize the importance of the customer needs. Having frameworks where the supply chain strategies are defined by customers order winning criteria or even by customer behavior, indicates that the supply chain strategies are being tailored to better fit customer requirements. Furthermore, the shift towards more customer centric supply chain strategies supports the conclusion of moving towards more dynamic supply chain management. As the customer requirements change through time, the supply chain strategies must dynamically follow the change to keep on filling the new customer requirements. When the product as a determinant of supply chain strategies is replaced by a customer, the supply chain strategies are more likely to meet market requirements and thus the supply chain strategies support the creation of sustainable competitive advantage.

What kind of supply chain strategies are best suited to Stora Enso's supply chain?

Since cartonboard business is process industry, where the end products are either cartonboard reels or sheets, the features of the products do not drastically differ from one another. Hence, defining supply chain strategies entirely based on the product characteristics might not result to value adding supply chain strategies. Even though the different cartonboard types are used in their own end uses and therefore might have naturally a differing market pull, supply chain strategies based on only product characteristics might be an unnecessary compromise among the market requirements. Instead, creating the supply chain strategies on top of customer profiling, supports customer business by supplying the products as per customers preferences. Tailoring the supply chain strategies for certain customer segments enables the creation of additional services and provides the circumstances for additional business.

Based on the current state analysis and the four clusters identified, there is clearly a need for several supply chain strategies. Four different customer segments were identified, each of which could be served by a customized supply chain. The foundation of the supply chain should be built on top of efficiency, which can answer the steady and easily predictable demand of the “lean” and “continuous replenishment” customers. On top of the efficient base of the supply chain, more customized solutions can be built to fulfill the needs of “collaborative/demanding” and “agile” customers. Whether the need is for short lead-times, or a large order of specialized products, the supply chain needs to be built on that level of flexibility, that these requirements can be fulfilled without disrupting the steady primary production.

How do the identified segments based on data analysis correspond to the segments identified from the supply chain segmentation literature?

The identified clusters corresponded well to the ones identified generally in the supply chain segmentation literature. Three out of four clusters were so similar to those identified in the literature that they were named using the same commonly used naming conventions as “lean”, “agile”, and “continuous replenishment”. Meanwhile, the fourth cluster, the “collaborative/demanding”, has strong elements from both “agile” and “continuous replenishment” supply chain strategies, and hence it cannot be matched to any single strategy. This customer behavior type may be a specialty related to the cartonboard business and therefore does not correspond to any commonly used strategy. Even if this cluster isn’t a commonly recognized supply chain segment, it has its own idiosyncrasies and hence it should be taken into account in a similar way as the commonly used strategies.

5.2 Future research

This study identified a movement towards a more dynamic and customer orientated supply chain strategies, but as the scope of this study was limited to providing only a current-state analysis, the dynamicity of customer requirements and supply chain offering were not captured. Future research should study this dynamicity more by analyzing data from longer period of time. This way, the business-specific trends could be captured which would help understanding how supply chain strategies should be created and managed now as well as in the future. Moreover, a longer period of time would reveal valuable information about the underlying business in general. “The best way to predict the future is to study the past” is a commonly used phrase which, when placed in a supply chain strategy context, helps to create more long lasting and value-creating supply chain strategies.

Another suggestion for future research is related to the tradeoff between efficiency and service-level when designing supply chain strategies. As the different supply chain strategies were recognized to position themselves in a so-called supply chain frontier (**Figure 2**), future research could investigate how the supply chain frontier could be pushed further by improving the efficiency and service level of the supply chain by digital solutions. Hypothesis is that digital solutions related to system integrations, steering the production, or having better visibility in the supply chain could improve efficiency or service-level without necessarily compromising the other. If the shape of the frontier could be modified, it would be interesting to find out how will the supply chain strategies position in the new frontier and will this have an effect to the customer behavior?

REFERENCES

Aggarwal, C. and Reddy, C., (2013). Data Clustering. Boka Raton: Chapman and Hall/CRC, pp.2-22

Aguinis, H., Gottfredson, R. and Joo, H., (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. Organizational Research Methods, 16(2), pp.270-301.

Basnet, C. and Seuring, S., (2016). Demand-Oriented Supply Chain Strategies A Review of the Literature. SSRN Electronic Journal,.

Berkhin, P. (2002). Survey of Clustering Data Mining Techniques.

Bezdek, J., (1987). Pattern recognition with fuzzy objective function algorithms. New York: Plenum Press.

Childerhouse, P., Aitken, J. and Towill, D., (2002). Analysis and design of focused demand chains. Journal of Operations Management, 20(6), pp.675-689.

Christopher, M. & Towill, D. (2001). An integrated model for the design of agile supply chains. International Journal of Physical Distribution & Logistics Management, 31(4), pp. 235-246.

Christopher, M., and D. R. Towill. (2000). “Supply Chain Migration from Lean and Functional to Agile and Customized.” Supply Chain Management: An International Journal 5(4): 206–213.

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

Christopher, M., Towill, D., Aitken, J. and Childerhouse, P., 2009. Value stream classification. Journal of Manufacturing Technology Management, 20(4), pp.460-474.

Collin, J., E. Eloranta, and J. Holmstrom. (2009). “How to Design the Right Supply Chains for Your Customers.” Supply Chain Management: An International Journal 14(6): 41–417.

Fisher, M. L. (1997). “What Is the Right Supply Chain for Your Product?” Harvard Business Review, March–April, 105–116.

Gattorna, J. (2006). Living Supply Chains. Harlow: Pearson.

Gattorna, J. and Friends (2009). Dynamic Supply Chain Alignment. Harlow: Pearson.

Gattorna, J. (2015). Dynamic Supply Chains, 3rd Edition. FT Publishing International.

Godsell, J., Diefenbach, T., Clemmow, C., Towill, D. and Christopher, M., (2011). Enabling supply chain segmentation through demand profiling. International Journal of Physical Distribution & Logistics Management, 41(3), pp.296-314.

Godsell, J., Harrison, A., Emberson, C. and Storey, J., (2006). Customer responsive supply chain strategy: An unnatural act?. International Journal of Logistics Research and Applications, 9(1), pp.47-56.

Hamerly, B., Elkan, C., (2002). Alternatives to the k-means algorithm that find better clusterings. In Proceedings of the Eleventh International Conference on Information and Knowledge Management, McLean, VA, USA, 4–9 November 2002; pp. 600–607.

JAIN, A.K., MURTY, M.N. and FLYNN, P.J., (1999). Data clustering: A review. ACM

Computing Surveys, 31(3)

Jüttner, U., Godsell, J. and Christopher, M., (2006). Demand chain alignment competence — delivering value through product life cycle management. Industrial Marketing Management, 35(8), pp.989-1001.

Lovell, A., R. Saw, and J. Stimson. (2005). “Product Value-Density: Managing Diversity Through Supply Chain Segmentation.” International Journal of Logistics Management 16(1): 142–158.

Mason-Jones, R., Naylor, B. and Towill, D., (2000). Engineering the leagile supply chain. International Journal of Agile Management Systems, 2(1), pp.54-61.

MCQUEEN, J. (1967). Some Methods for Classification and Analysis of Multivariate

Observations. Computer and Chemistry, 4, 286-296.

Naylor, J. B., M. M. Naim, and D. Berry. (1999). “Leagility: Integrating the Lean and Agile Manufacturing Paradigms in the Total Supply Chain.” International Journal of Production Economics 62(1–2): 107–118.

Omran, Engelbrecht, and Salman. (2007). An overview of clustering methods. Intell. Data Anal.. 11. 583-605. 10.3233/IDA-2007-11602.

Payne, T. & Peters, M., (2004). What Is the Right Supply Chain For Your Products?. The International Journal of Logistics Management, 15(2), pp.77-92.

Porter, M. E. (1990). The competitive advantage of nations. New York: Free Press.

Samuel Roscoe & Peter Baker (2014). Supply chain segmentation in the sporting goods industry, International Journal of Logistics Research and Applications, 17(2), 136-155.

Schämmle, V. and Jensen, O., (2010). A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. Bioinformatics, 26(22), pp.2841-2848.

Selldin, E. and Olhager, J., (2007). Linking products with supply chains: testing Fisher's model. Supply Chain Management: An International Journal, 12(1), pp.42-51.

Stora Enso. (2019). Annual Report 2019 [www-document]. Stora Enso -website [accessed 3.7.2020]. Available: https://www.storaenso.com/en/investors/reports-and-presentations

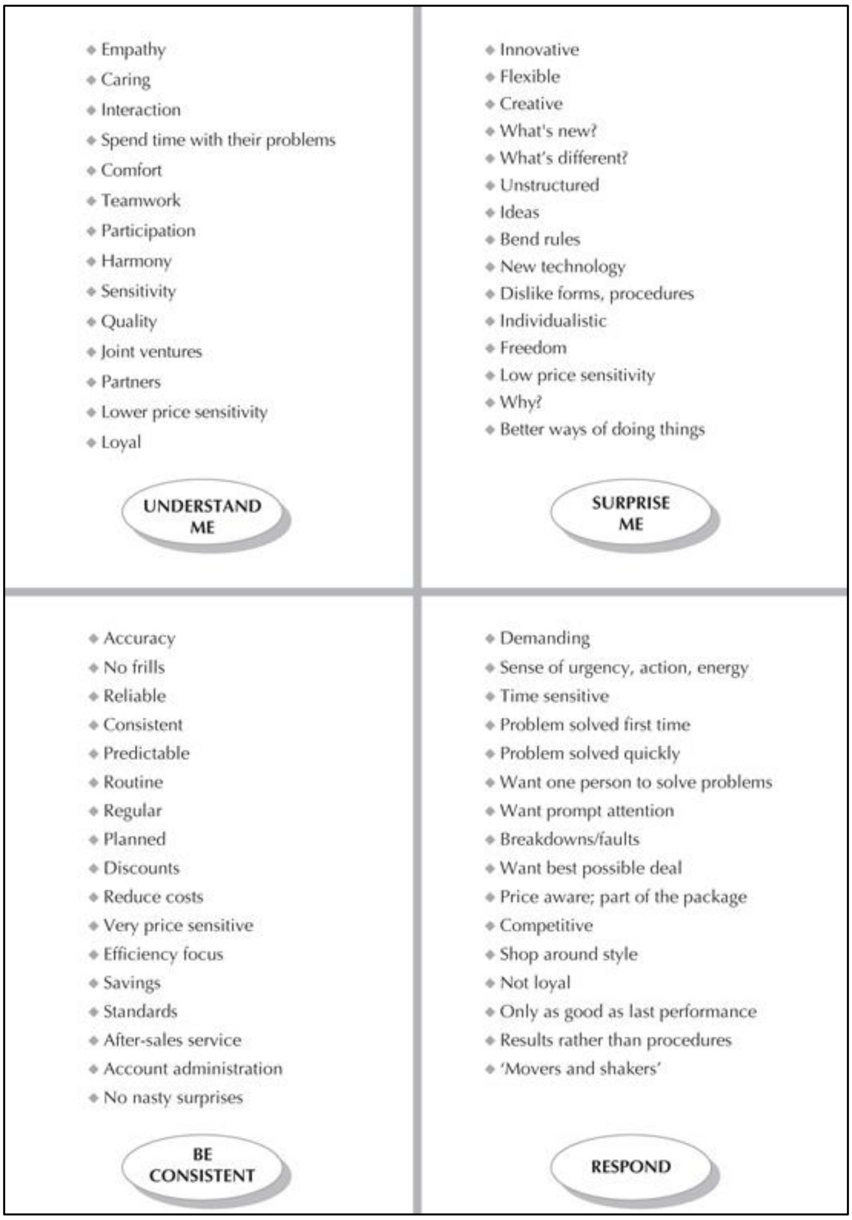
Xu, R. & Wunsch, D., (2009). Clustering. John Wiley & Sons, Inc., pp.1-16.

Zakrzewska, D. & Murlewski, J., (2005) Clustering algorithms for bank customer segmentation. Intelligent Systems Design and Applications, 2005. ISDA '05. Proceedings. 5th International

Zheng, D., (2013). Application of Silence Customer Segmentation in Securities Industry Based on Fuzzy Cluster Algorithm. Journal of Information and Computational Science, 10(13), pp.4337-4347.

APPENDIX

Appendix 1. Service logics related to customer behavior types (reproduced from Gattorna, 2015 with the permission of the publisher)



Appendix 2. Average Silhouette index for tested HC methods

Silhouette index			
Number of clusters	Agglomerative complete link	Agglomerative single link	Divisive
2	0.27	0.33	**0.41**
3	0.26	0.06	**0.37**
4	0.25	0.05	**0.36**
5	0.24	0.01	**0.35**
6	0.27	-0.09	**0.33**
7	0.27	-0.08	**0.30**
8	0.26	-0.20	**0.33**

Appendix 3. Average Silhouette index for different k-means initializations

Silhouette index				
Number of clusters	Hartigan-Wong	Lloyd	Forgy	MacQueen
2	0.42	0.42	0.42	0.42
3	0.34	0.34	0.34	0.34
4	0.31	0.31	0.33	0.31
5	0.33	0.33	0.27	0.33
6	0.31	0.31	0.31	0.31
7	0.29	0.30	0.30	0.30
8	0.28	0.28	0.29	0.29
Average	0.322	0.334	0.322	0.333

Appendix 4. Average Silhouette index for different fuzzifier value m

Silhouette index				
Number of clusters	m = 2	m = 3	m = 5	m = 10
2	0.41	0.35	0.34	0.34
3	0.36	0.34	0.35	0.27
4	0.29	0.33	0.32	0.26
5	0.31	0.32	0.32	0.27
6	0.29	0.31	0.33	0.14
7	0.28	0.24	0.31	0.28
8	0.26	0.24	0.18	0.20
Average	**0.314**	0.304	0.307	0.251