2. BUSINESS DATA ANALYTICS

Author: Dejan Mirčetić

In the era of digitalization, due to the enormous amount of data generated on a daily basis, traditional knowledge and approaches cannot be used to manage business processes in different areas, so also for manage logistics and supply chains (Nikoličić et al., 2019). Web 2.0, together with Industry 4.0, cloud computing, the Internet of Things (IoT), RFID and other digital technologies have led to generation, storage and transmission of large amounts of data. As the volume and complexity of data increases, so does the complexity and the time required to analyze those data and derive insight from them.

The concept of Big Data was first introduced by Cox and Ellsworth in October 1997, in an ACM digital library article (Tiwari et al., 2018). The study of Big Data and its conceptualization have evolved continuously. Initially, Big Data was characterized by the 3Vs concept, which ecnompassed **volume**, **velocity**, and **variety**, as discussed in the previous chapter. Subsequently, this characterization expanded to the 5Vs concept, incorporating two additional attributes: **veracity**, and **value** (Nguyen et al., 2018; Tiwari et al., 2018). Volume refers to the magnitude of data generated; the volume of digital data is growing exponentially (Arunachalam et al., 2018). Variety refers to the fact that data can be generated from heterogeneous internal and external sources, in structured, semi-structured, and unstructured formats. Velocity refers to the speed of data generation and delivery, which can be processed in batch, real-time, nearly real-time, or stream- lines. Veracity stresses the importance of data quality because many data sources inherently contain a certain degree of uncertainty and unreliability. Value refers to finding new value contained in the data which can be used for better business planning (Nguyen et al., 2018).

Big Data Analytics (BDA) incorporates two dimensions: **Big Data (BD)** described with the 5Vs concept and **Business Analytics (BA)** which enables to gain insight from data by applying statistics, mathematics, econometrics, simulations, optimizations, or other techniques to help business organizations make better decisions (Wang et al., 2016). Big Data Analytics (BDA) involves the use of advanced analytics techniques to extract valuable knowledge from vast amounts of data with variable types in order to draw conclusion by uncovering hidden patterns and correlations, trends, and other business valuable information and knowledge, in order to increase business benefits, increase operational efficiency, and explore new market and opportunities (Nguyen et al., 2018; Tiwari et al., 2018). BDA has attracted significant attention in different areas, both academically and business, particularly in logistics and supply chain management.

2.1. BDA in logistics and supply chain management

The supply chains (SCs) represent the network of firms and facilities involved in the transformation of raw materials to final products and distribution of final products to end customers. In SCs, there are physical, financial, and informational flows among different firms. Every day, SCs are becoming more complex, more extended and more global. Therefore, for the successful implementation and management of existing processes in the SC and their continuous alignment with market conditions, modern SC needs highly skilled experts. In order to answer these challenging tasks, SC experts need formal education, which will provide them knowledge and skills from different fields, primarily from logistics, information technology and economics.

The SC is a set of physical elements, their activities and processes through which their interaction takes place. Physical elements, which make up the chain structure, represent a fixed part of the SC. Decisions on the design of the SC structure are made at the strategic level, and at the tactical and operational level, decisions are made on the modalities and rules for the realization of particular logistics processes. Designing a fixed SC and managing an exquisite work together provide a SC management that defines the performance of the chain. Accordingly, the SC management framework consists of three basic elements: (1) SC structures; (2) business processes; (3) and the control components. Each of these elements is directly related to the objectives of the SC, that is, with the degree of fulfillment of the requirements of the end-users, while respecting the critical dimensions of the business that depends on the performance on the market (key performance indicators - KPI). In the modern world competition is no longer between organizations but among SC. Effective SC management has, therefore, become a potentially valuable way of securing competitive advantage and improving organizational performance. SC management is a fact of business, with logistics as a most powerful tool for achieving the ultimate strategic advantage.

Firms are under heavy pressure to improve SC planning and performance because of factors such as increasing uncertainty and competition. Improving SC performance has become a continuous process that requires an analytical performance measurement system. Considering the number and diversity of logistics processes and SC processes, the resources used for their realization, the parameters that characterize them, as a basis for determining SC performance, a large number of data is used on: geographic, time and quantity determinants of goods, transport means, transport - manipulative assets, warehouse capacities, employees, etc. Data generated through internal operations, as well as transactions with suppliers and customers, can be used to uncover small changes that can make a big impact on an organization with regard to efficiency gains and even cost savings. The other words, the volume of data in every SC is exploding from different data sources, business processes, and IT systems. As the volume and complexity of data increases, so does the complexity and time taken to analyze that data and derive insight from it. Determining, monitoring, and improving logistics and performance SC becomes more complex and involves many processes such as identifying measures, defining targets, planning, communication, monitoring, reporting and feedback. Consequently, conventional approaches cannot be used to make SC decisions and SC management.

In SC management, there is a growing interest in Business Analytics, which is also called **Supply Chain Analytics (SCA)**. SCA is used synonymously with the terms such as 'Big Data Analytics' and 'Business Analytics' within the business and academic communities (Srinivasan and Swink, 2018). SCA refers to the use of data and quantitative tools and techniques to improve operational performance, often indicated by such metrics as order fulfillment and flexibility. Analytics in SCs is not necessarily a new idea since various quantitative techniques and modeling methods have long been used in manufacturing firms. The recent surge of interest in SCA is accompanied by new challenges and opportunities in both business and information technology environments. These challenges include issues arising from managing large amounts of data (e.g. data availability and data quality) and dealing with environmental uncertainties. The properly applied SCA can impact several areas in SC and it can generate significant benefits in logistics performances: improved planning and scheduling; improved responsiveness; improved demand planning; order optimization; optimized inventory management; improved replenishment planning. In recent decades, under the influence of technological development, globalization and increasingly demanding customers, business paradigms have also changed. Figure 2.1 shows typical periods (with a brief description) in the evolution of logistics, SCM and BDA.

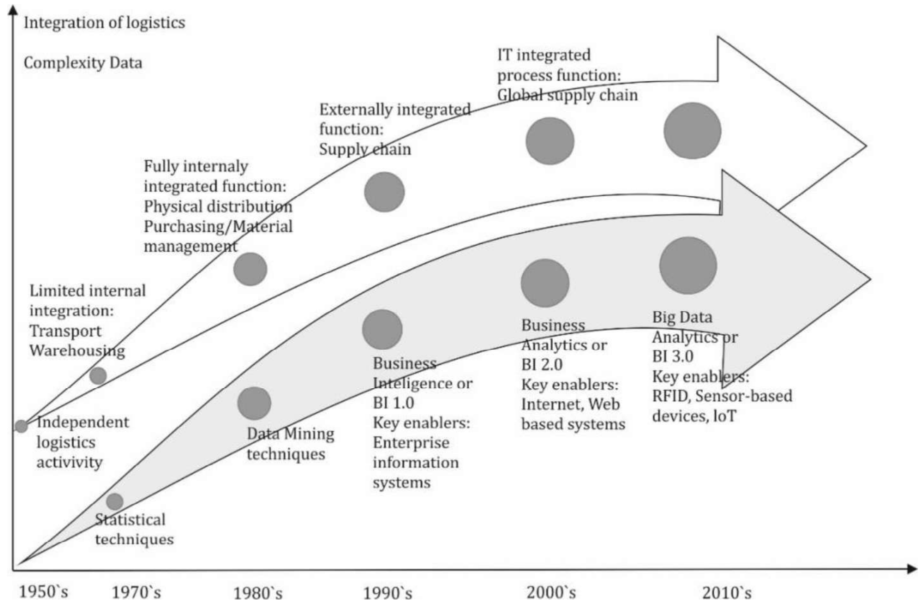

Figure 2.1 Evolution of logistics, SCM, and BDA

Source: Adapted from Arunachalam et al. (2018).

2.2. Tools in business data analytics

Figure 2.2 presents different trends, tools & benefits used in BDA or SCA. All presented analytics techniques can be categorized into three types: descriptive, predictive, and prescriptive. **Descriptive analytics** looks at data and analyzes past events for insight as to how to approach the future. There are looking for the response behind past failures and successes. **Predictive analytics** uses historical data to determine the probable future outcome of an event or the likelihood of a situation occurring. It exploits patterns found in the data to identify future risks and opportunities. **Prescriptive analytics** automatically synthesizes Big Data, business rules, and machine learning to make future predictions. It goes beyond predicting the future by suggesting actions, which needs to be taken in order to achieve desired goals. Also, they are able to demonstrate the implications of each possible decision and act as a decision support tool for SC experts. In the following sub-chapters, we will introduce and describe the various analytical tools used for the BDA in SCM. Additionally, we will focus on strategies on how to enhance knowledge in BDA for SC experts of the 21st century, via several case studies.

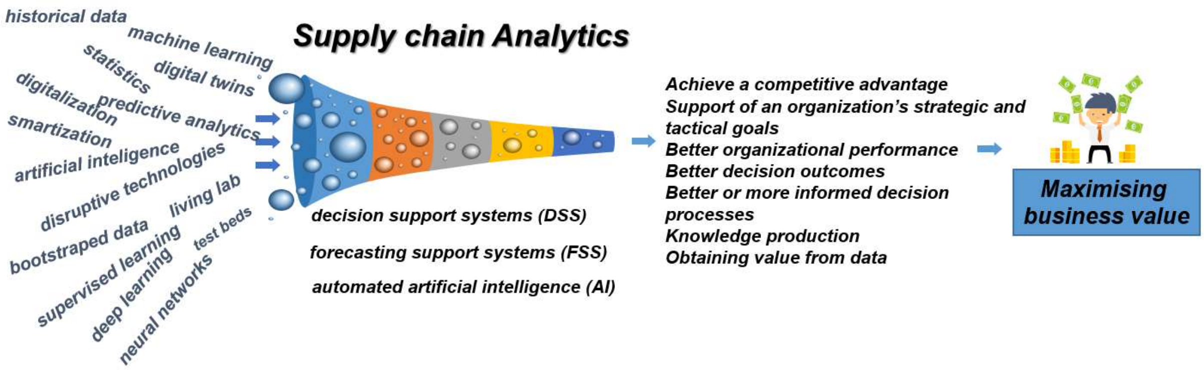

Figure 2.2 Trends, tools & benefits of SCA

Source: Author.

2.2.1. Descriptive analytics

Descriptive analytics provides a summary of descriptive statistics for a given data sample, for example: mean, mode, median, range, histogram and standard deviation. Descriptive analytics describes what happened in the past and derives information from significant amounts of data to answer the question of what is happening. On the basis of real-time information about locations and quantities of goods in the supply chain, managers make decisions at the operational level (e.g. they adjust the schedule of shipments, deploy vehicles, issue orders for restocking products, etc.) (Souza, 2014). It attempts to identify opportunities and problems using online analytical processing system and visualization tools supported by real-time information and reporting technology (e.g. GPS, RFID, transaction bar-code). Common examples of descriptive analytics are reports that provide historical insights regarding the company's production, financials, operations, sales, finance, inventory, and customers (Tiwari et al., 2018).

2.2.2. Predictive analytics

Predictive analytics uses historical data to determine the probable future outcome. Predictive analytics in supply chains derives demand forecasts from past data and answers the questions related to what will be happening or what is likely to happen (Tiwari et al., 2018). It use artificial intelligence, optimisation algorithms and expert systems to predict future behaviors based on patterns uncovered in the past and the assumption that history will repeat. It exploits patterns found in the data to identify future risks and opportunities and predict the future. This is used to fill in the information that is missing and to explore data patterns using statistics, simulation, and programming.

2.2.3. Prescriptive analytics

Prescriptive analytics derives decision recommendations based on descriptive and predictive analytics models as well as on mathematical optimization, simulation or multi-criteria decision-making techniques. It goes beyond predicting future outcomes by also suggesting action to benefit from the predictions and showing the decision maker the implications of each decision option. Prescriptive analytics answers the question of what should be happening.

2.3. BDA ecosystem

The main purpose of the BDA ecosystem is to deliver value for the decision-maker. Accordingly, the BDA has a primary goal of providing insight into business processes and leading to the possible answer on how to reduce costs and increase the service level for the final customers. To fulfil its goal, the BDA solutions are usually delivered in the form of Decision Support System (DSS) or Expert System (ES). Therefore, in this and upcoming chapters we will dive into the key pillars of BDA: **business data**, **data mining** and **knowledge discovery** (data analytics, DSS, ES platforms, etc).

2.3.1. Business data

The concept of data is explanied in detail in the first chapter of this book. Data is the key factor for conducting any kind of analysis. **Business data** is generated as a result of the execution of processes in a given business environment. In the case of the SC, there are many processes & subprocesses involved in delivering the service to the final customer (Figure 2.3).

Figure 2.3 represents the structure of the SC where in the case of SCA each of the given processes can be observed as a generator of business data. Generated data is different in its importance and influence on the final goals of the company. Accordingly, the business data can be divided into **internally driven** and **externally driven** data. The internally driven data is the data which emerges as a result of company structure, hierarchy and the way the company has decided to operate (for example, manufacturing data, human resource data, delivery data, accounting data, etc). This data is different for each company and it serves for reporting, analyzing and legal reporting to the authorities (accounting data). What is interesting about this data is that the companies are directly controlling and influencing this data and it only has a value for a given company.

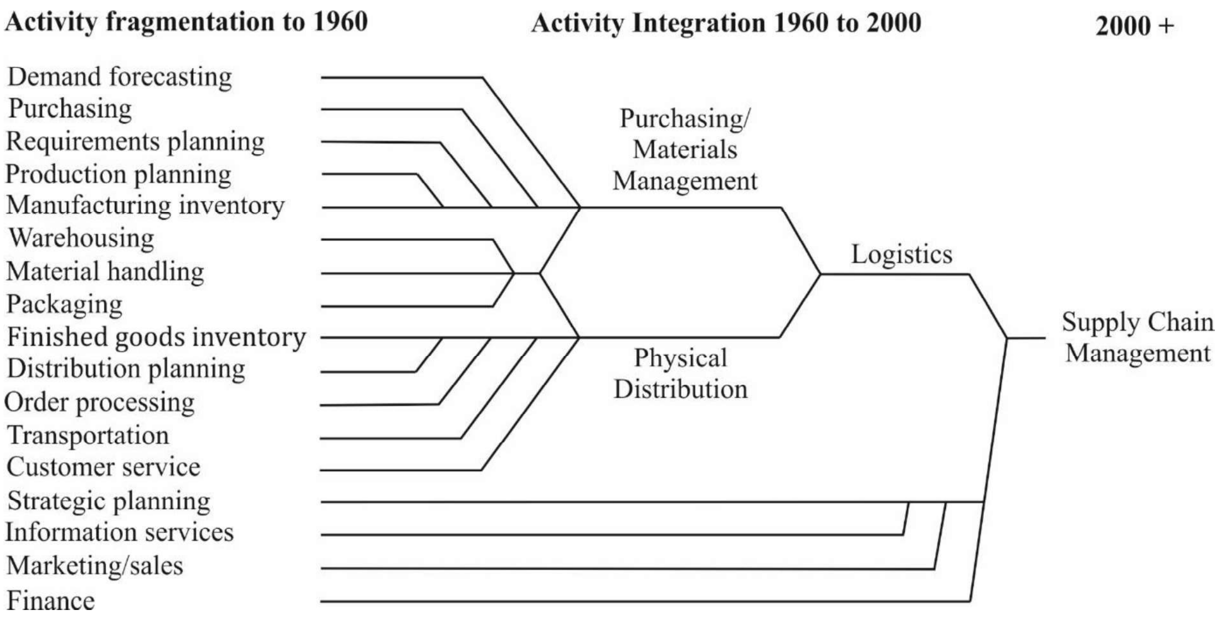

Figure 2.3 Evolution of the logistics and SCs

Source: Hesse & Rodrigue (2004).

External data refers to data generated outside an organisations, which can be public, unstructured, or collected by third-party organisations (private data). For SC analysts, particulary significant is external data shared between companies within supply chains, including market demand data. This kind of externally driven data is important for the companies since it is a result of the market response to the company products and services. The company doesn't have any direct influence on a given data, although companies try via the demand process and its subprocess demand planning to indirectly improve the market response of the customers. Moreover, companies try to model the market engagement to their products via demand planning activities like packaging, promoting the product, making sale promotions, using several distribution channels, etc.

Companies invest a lot of time, money and effort to better understand and model its processes according to the market demand data. This is a very challenging task for several points. First of all, companies need to establish infrastructure, procedures and contracts with the retailers to track & record the demand data. Usually, companies use sales data from the downstream partner in the SC as a proxy for demand data. In reality, this is not demand data, rather it is the procurement data which can significantly distort the demand data. This is a very common practice since companies don't want to share its data and a large share of companies do not know that this is a bad practice. One of the downsides of this approach because it causes a bullwhip effect among the partners in SC. Another approach is that companies use retailers' point of sale (POS) data as a proxy for demand data (Syntetos et al., 2016). This also has its merits and downsides, since it doesn't take into account out of stock situations on the retail shelves. This approach also needs strong IT infrastructure and contracts with retailers.

The second „problem” with market-generated data is that it doesn't follow the usual statistical generating processes. This is a problem because the majority of the mathematical and statistical methods assume that data is following some statistical generating process. This is very notable in SC, where inventory models assume that demand during the lead time follows the Normal distribution and develop equations for calculating safety stock based on that assumption. According to Mirčetić et al. (2017), Mirčetić et al. (2022) & Mirčetić et al. (2018), 90% of the data from the pool of 97 series in the empirical study of the beer industry, doesn’t follow the Normal distribution. Also, inventory models assume that the demand is deterministic and uniformly distributed through all periods, in reality, this is hardly the case. For example in Figure 2.4, the demand for pre-sliced salami in the period of 2015-2022 is presented for Italian manufacturer. The demand shows clear non-deterministic, i.e. stochastic behaviour (with random fluctuations and trend).

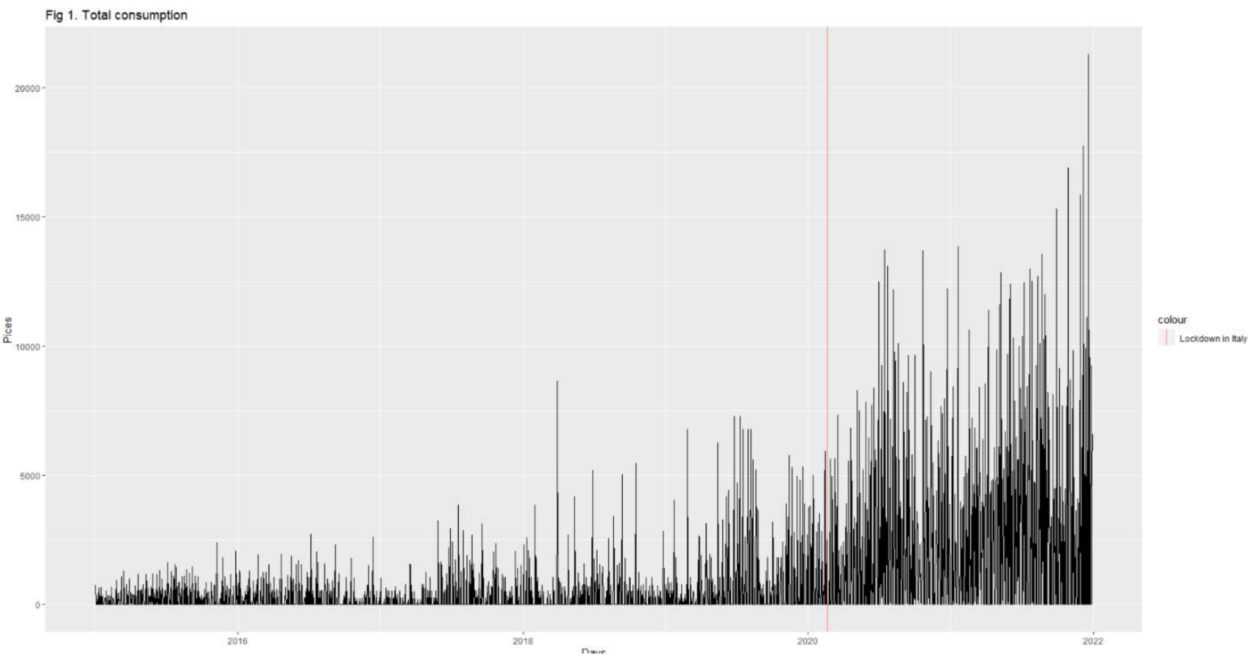

Figure 2.4 The demand for the pre-sliced salami in the period of 2015-2022

Source: Author.

Additionally, the daily demand shows highly volatile behavior, therefore aggregation of the demand through different time horizons (weekly, monthly, etc) demonstrate a clear galloping demand trend (Figure 2.5).

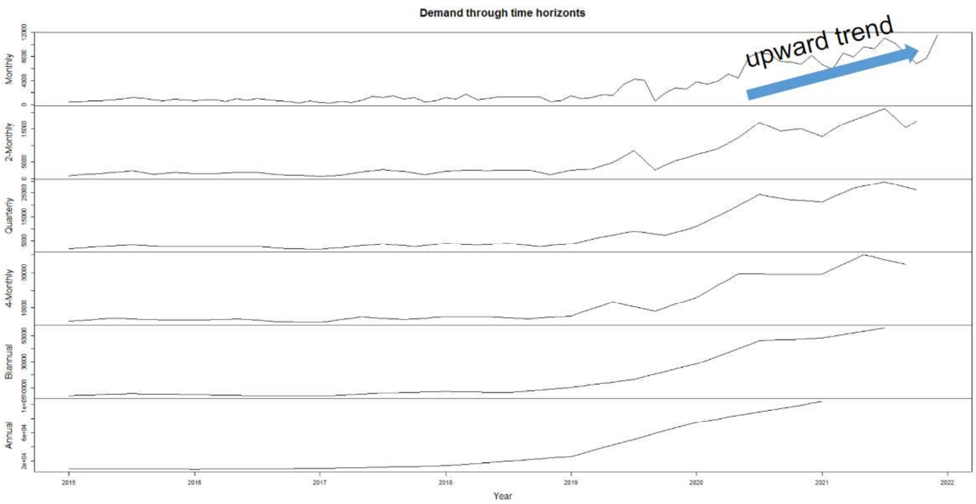

Figure 2.5 Demand aggregation through different time horizons

Source: Author.

Demand aggregation demonstrates a clear new reality from the start of the COVID-19 (an upward trend in the demand consumption process)! These are bulletproof evidence that demand is not deterministic. Regarding the assumption of normality, Figure 2.6 demonstrates a significant discrepancy from the Normal distribution, bearing in mind that it is an extremely right-skewed distribution.

Figure 2.6 demonstrates, that there has been a notable change in the salami demand since the start of the COVID-19 pandemic. The first lockdown in Italy was 2020-02-21 (vertical red line in Figure 2.4), after which demand for presliced salami erupted and reached a historical maximum. The demand was constantly trending, reaching the highest peak on 2021-12-20, with 21280 sliced pieces sold in one day.

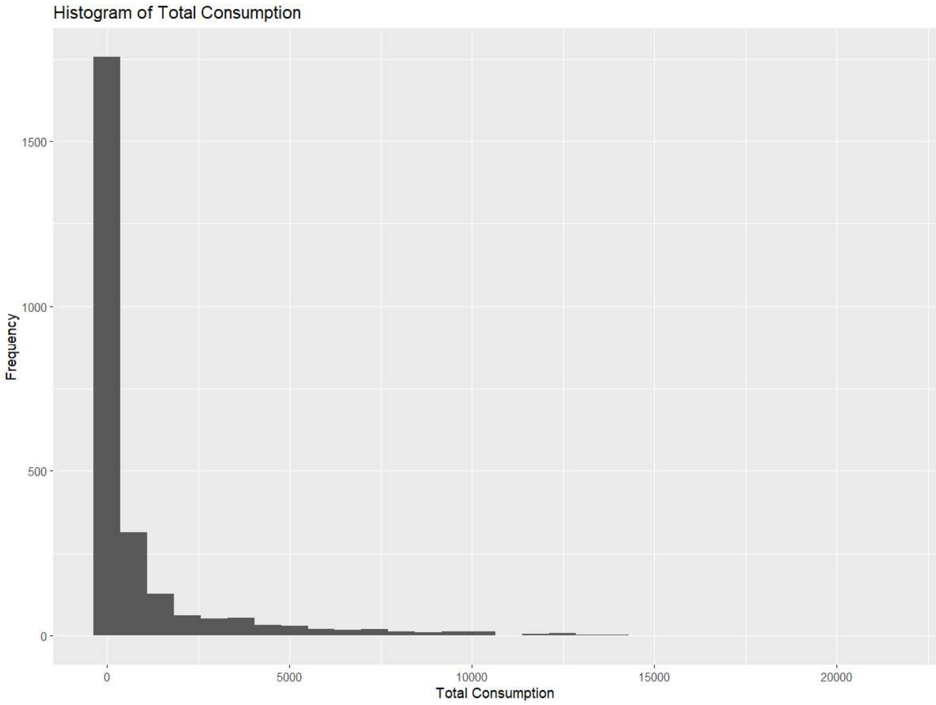

Figure 2.6 Empirical distribution of the demand

Source: Author.

Besides problems with the aforementioned theoretical assumptions about the data, the current situation in a worldwide economy and consequently SCs is posing another question in front of business analysts. These questions emerge as a result of pandemics, war crises, resource shortages, growing inflation, broken worldwide SC, etc. The question that contemporary business analysts need to solve when dealing with the data after the start of the COVID-19 is how long period and data horizon are now valid for observation and modeling? This is clearly seen in Figure 2.4. If we take a closer inspection of the figure we will notice that before the COVID-19 pandemic and lockdown, consumers never consumed the pre-sliced salami in levels like from the first lockdown. It is noticeable that consumption of a given product erupted. Several questions now emerge:

- Is this just a hype in consumption due to specific conditions during the pandemic;
- Will this trend continue in the future and the company needs to increase its production;
- Does pre-COVID-19 data (from 2015-2020) have any value today and need to be discarded when modelling the demand data for pre-sliced salami?

These are all very hard questions to answer without proper data analytics procedure, which will be presented in the following Chapters.

REFERENCES

1. Arunachalam, D., Kumar, N. & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. Transportation Research Part E, 114, pp. 416-436.
2. Hesse, M. & Rodrigue, J.-P. (2004). The transport geography of logistics and freight distribution. Journal of Transport Geography, 12(3), pp. 171–184.
3. Mircetic, D., Rostami-Tabar, B., Nikolicic, S. & Maslaric, M. (2022). Forecasting hierarchical time series in supply chains: an empirical investigation. International Journal of Production Research, 60(8), pp. 2514-2533.
4. Mirčetić, D. (2018). Unapređenje top-down metodologije za hijerarhijsko prognoziranje logističkih zahteva u lancima snabdevanja (Doctoral dissertation), University of Novi Sad, Serbia.
5. Mirčetić, D., Nikoličić, S., Stojanović, Đ. & Maslarić, M. (2017). Modified top-down approach for hierarchical forecasting in a beverage supply chain. Transportation Research Procedia, 22, pp. 193-202.
6. Mirčetić, D., Ralević, N., Nikoličić, S., Maslarić, M. & Stojanović, Đ. (2016). Expert system models for forecasting forklifts engagement in a warehouse loading operation: A case study. Promet-Traffic&Transportation, 28(4), pp. 393-401.
7. Nikoličić, S., Maslarić, M., Mirčetić, D. & Artene, A. (2019). Towards more efficient logistic solutions: Supply chain analytics. In Proceedings of the 4th Logistics International Conference, Belgrade, Serbia (pp. 23-25).
8. Souza, G. C. (2014). Supply chain analytics. Business Horizons, 57, pp. 595-605.
9. Srinivasan, R. & Swink, M. (2018). An investigation of visibility and flexibility as complements to supply chain analytics: an organizational information processing theory perspective. Prod. Oper. Manag. 27(10), pp. 1849–1867.
10. Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S. & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. European Journal of Operational Research, 252(1), pp. 1-26.
11. Tiwari, S., Wee, H. M. & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering, 115, pp. 319-330.

12. Wang, G., Gunasekaran, A., Ngai, E. W. T. Papadopoulos Th. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. Int. J. Production Economics, 176, pp. 98-110.

13. Zhu, S., Song, J., Hazen, B. T., Lee, K. & Cegielski, C. (2018). How supply chain analytics enables operational supply chain transparency: An organizational information processing theory perspective. International Journal of Physical Distribution & Logistics Management, 48(1), pp. 47-68.