The effect of Supply Chain Analytics on business decision-making

Heinrich Vermeulen

orcid.org/0000-0002-9021-6942

Mini-dissertation accepted in partial fulfilment of the requirements for the degree *Master of Business Administration* at the North-West University

Supervisor: Dr JA Jordaan

Graduation: May 2020

Student number: 25755412

PREFACE

Acknowledgements

My deepest appreciation goes out to:

- My wife, for her love, understanding, and support during the long hours spent on my studies and research.
- My sons for understanding when I could not play with them as much as I would have liked to due to limited time.
- Dr Johan Jordaan, my supervisor and study leader, whose guidance and wisdom have helped to shape my understanding of many aspects of my research.
- The NWU Business School for allowing me to discover new insights and hone my understanding of business during the MBA course.
- To my Lord Jesus Christ, who has guided me and has been with me every step of the way.

ABSTRACT

Data is the raw material of the business of the future and information the oil of the 21st century. Several attempts have been made to examine the use of Big Data and Data Analytics in organisations to improve increasingly complex supply chains in uncertain markets with rising levels of consumer demand. The reality is that many organisations still struggle to create a competitive advantage at the supply chain level and even those organisations that have invested in the use of Supply Chain Analytics have struggled to translate the information into effective decision-making. Previous work has failed to address the competitive advantage of Supply Chain Analytics on decision-making where Supply Chain Analytics itself is a relatively under-researched area of Data Analytics. To determine the effect that Supply Chain Analytics has on business decision-making, an empirical study was performed, making use of an online questionnaire. The various aspects that influence business decision-making when applying Data Analytics were investigated, and it was found that there is a strong contribution of Data Analytics to business decision-making. Various other aspects of Data Analytics that indirectly contribute to business decision-making also emerged. The results demonstrate that the use of Supply Chain Analytics plays a pivotal role in elevating the decision-making abilities within an organisation.

Keywords: Supply Chain Analytics, Decision-making, Data-driven Decisions, Supply Chain Management, Big Data, Big Data Analytics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF TERMS

BA	Business Analytics
BDA	Big Data Analytics
BI	Business Intelligence
DA	Data Analytics
DDS	Dimensional Data Store
DPP	Distributed and Parallel Processing
ERP	Enterprise Resource Planning
ETL	Extract, Transform, Load
GPS	Global Positioning System
HTML	Hypertext Mark-up Language
IoT	Internet of Things
KDD	Knowledge Discovery for Databases
ODS	Operational Data Store
RTI	Real-time Intelligence
SCA	Supply Chain Analytics
SCM	Supply Chain Management
SPSS	Statistical Package for Social Sciences (IBM)
RFID	Radio-Frequency Identification
XML	Extensible Mark-up Language

CHAPTER 1: NATURE AND SCOPE OF STUDY

1.1 Introduction and literature overview

The primary purpose of any for-profit organisation is to maximise profits in the long run by ensuring that all stakeholders benefit from everything it does (Arjoon *et al.*, 2018:159). To achieve this, an organisation needs to have products that customers want and be able to produce the products at a cost that will realize maximum profit (Porter & Kramer, 2019:328). To optimise profits, an organisation must, therefore, understand accurately where the costs are incurred, and for this understanding, management requires the correct information from which business decisions can be made (Bhushan *et al.*, 2017:17).

Most 21st century organisations use data from the past to make decisions that may impact the future (Amankwah-Amoah, 2016:170). However, normal hindsight models do not always consider fluctuations in demand or supply, the volatility of stock prices and availability of inventory. Due to its inherent complexity and the large cost structures normally associated with it, the supply chain increasingly lends itself to be the source of competitive advantage by using Data Analytics (Baesens *et al.*, 2016:810). Analytics, in this context, refers to the application of computer analysis of data and statistics to report, optimise or predict business processes. This fact has given rise to the study field of Business Analytics (Frizzo-Barker *et al.*, 2016:404).

Business Analytics (BA) attempts to create new perspectives and insights into the business by taking relevant data and analysing the data through quantitative analysis and then using this data to create models to predict the future (Nair, 2014:11). Laursen and Thorlund (2016:xiv) define Business Analytics as “delivering the right decision support to the right people at the right time”. Wang *et al.* (2016:99) go further to state that analytics provides the mechanism to draw insights from data through application areas such as mathematics, statistics, economics, simulations, optimisations and other proficiencies that allow a business to make better, more informed decisions.

It is not a new concept to use data to analyse a business to improve profitability. Modern-day Business Analytics can be traced back to Henry Ford at the Ford Motor Corporation, who used analytics to put the assembly line in place (Calof *et al.*, 2015). In the 2000s

Business Analytics came to the fore as the central analytical component of business intelligence (Chen *et al.*, 2012:1166). In the early 21^{st} century, Big Data became prominent, which added another dimension to Business Analytics due to the vast amounts of data that could contribute to Business Analytics (Arunachalam *et al.*, 2018:418).

Big Data can be defined as "the massive volume of both structured and unstructured data that is too large and too difficult to process using conventional database and software techniques" (Jain *et al.*, 2017:1106). Big Data is typically heterogeneous and may be structured, semi-structured or unstructured (Gandomi & Haider, 2015:138). Structured data refers to data that is found in a tabular format that can easily be processed by computers, for example, spreadsheets and databases. Unstructured data, sometimes referred to as messy data, is not structured in a way that computers can easily analyse and interpret, for example, video, text and audio.

Big Data Analytics (BDA) brings together Big Data and Business Analytics. Gandomi and Haider (2015:140) affirm that Big Data is rather meaningless in isolation, and its true value is only realised when applied to drive decision-making. BDA builds on Business Analytics but adding the ability to process and analyse both structured and unstructured data. The key outcome of BDA is decision-making that is based on complex, rapidly changing data (Nguyen *et al.*, 2018:254).

Tan *et al.* (2015:223) found that managers are increasingly viewing Big Data as a major driver of competitive advantage and the creation of value. Manyika *et al.* (2011) call Big Data "the next frontier for innovation, competition and productivity". Wang *et al.* (2016:99) concur that BDA enables value creation to provide a powerful competitive advantage to the business.

Supply Chain Analytics (SCA) is a sub-domain of Business Analytics and more specifically applies BDA to the organisational function of Supply Chain Management (SCM). The supply chain is a network of entities involved in the production and distribution of products and services. The supply chain includes all processes and activities from the supplier to the customer (Heizer *et al.*, 2016:44). Jain *et al.* (2017:1106) define Supply Chain Management (SCM) as managing the flow of commodities and services from its origin to where these get consumed.

Within the context of SCM, Arunachalam *et al.* (2018:419) define SCA as the ability of businesses to amass, organise, analyse and visualise heterogeneous data from the supply chain (both internal and external to the business) to enable proactive decision-making within the supply chain. Souza (2014:595) argues that Supply Chain Analytics "focuses on the use of information and analytical tools to make better decisions regarding material flows in the supply chain". Wang *et al.* (2016:107) argue that Supply Chain Analytics needs to be seen as an asset of strategic value by companies that should be applied holistically. They argue that successful companies compete on analytics aggressively by exploiting data. SCA is, therefore, a powerful tool to support decision-makers in the supply chain echelons of the organisation.

According to Jain *et al.* (2017:1111), the primary issues today in SCM are the ability to reduce operating as well as overall inventory costs. Reducing inventory cost entails shortening lead times, ensuring better throughput, and possibly employing lean principles. Jain *et al.* (2017:1113) further argue that by using SCA the level of inventory, as well as the associated costs, can be drastically reduced. Wang *et al.* (2016:98) affirm that Big Data can provide unique insights into ways of lowering costs and enabling business decisions that are more targeted. Yu *et al.* (2018:371) assert that Big Data has the potential to revolutionise the supply chain's performance completely.

Wang *et al.* (2016:101) further suggest that SCA can help management make better decisions by interpreting market condition changes and supply chain risks, and then capitalising on the capabilities of the supply chain to develop strategies that are competitive and geared towards being profitable yet flexible. Thus, SCA provides positive inputs at a strategic level in the organisation empowering top management to make informed strategic decisions. Souza (2014:598) further elaborates on this by stating that SCA enables not only decision-making at a strategic level but also at tactical and operational levels of the organisation.

Chae *et al.* (2014:4695) state that the use of analytics in Supply Chain Management is not a novel idea and although Souza (2014:604) agrees with this, he remarks that SCM remains a very productive area where analytics can be applied. Combining analytics with Big Data changes the scene completely, and that is a fairly new concept. Big Data brings new opportunities. Davenport and Harris (2017) suggest that data should be viewed as an asset of strategic value and be managed in such a way as to maximise this value in the organisation. He further argues that companies should compete on business analytics to differentiate themselves and to create a competitive advantage.

Information, made up of data that has been processed, translated, organised and structured, is increasingly a source of competitive advantage. Business Analytics endeavours to gain a competitive advantage through analysis of such information (Assunção *et al.*, 2015:2). Information is also getting more diverse and difficult to sieve. Big Data provides a context for such information. Big Data Analytics binds Business Analytics and Big Data. Big Data Analytics that specifically applies to the supply chain, is termed Supply Chain Analytics (SCA), which can yield a competitive advantage if executed correctly. The key outcome of applying Supply Chain Analytics in any organisation is to drive optimal and timely business decision-making (Wang *et al.*, 2018:3).

Organisations are perceived by the decisions they make (Taylor & Raden, 2007:5). The benefits of better decision-making are ample. The basic premise of better decision-making is being able to do things cheaper, faster and smarter, which has an overall benefit to an organisation's bottom line. Improving decision-making can allow an organisation to move faster and react quicker to customers who have become used to ever-evolving changing inputs from social media, thus providing an agility advantage over rivals. It allows shorter turnaround times between problem recognition and addressing the problem. Not making timely decisions might mean missed opportunities, slipping behind competitors or falling behind technology-wise. In an environment where fierce competition results in smaller profits, better decision-making may result in better resource utilisation, improved risk management and optimal use of opportunities.

Taylor and Raden (2007:8) argue that "organisations that can decide faster can move faster". Within the context of supply chains, this is especially true, and the benefits of better decision-making abound. The major benefits of optimal and timely decision-making include elimination or mitigation of risk and uncertainty and being early to detect and adapt to changes in the environment by continually evolving the organisation's operational strategy. Supply Chain Analytics facilitates elevated decision-making that will enable organisations to optimise and improve their processes at various levels of the organisation.

William Edwards Deming, the grandfather of Total Quality Management, is quoted as saying “In God we Trust, all others bring data” (Oivo, 2016:323). It is evident that access to the relevant data holds the key to effective and optimal decision-making within the domain of Supply Chain Management. SCA provides challenging new possibilities of analysis of a variety of data from various sources to act as an enabler to better decision-making to optimise operations within the organisation at a strategic, tactical and operational level.

1.2 Problem statement

The competition in the market is increasingly a primary function of the supply chain and not essentially between organisations themselves (Ai *et al.*, 2012:259). Supply Chain Analytics (SCA) can act as a vehicle to create a proactive supply chain system with sound and timely decision support which is backed by analytical data (Arunachalam *et al.*, 2018:419). SCA can further enable better decision-making by using analytical tools within the supply chain (Souza, 2014:595).

Organisations are coming under increased pressure to improve their supply chains due to increased competition and uncertainty in the market (Chae *et al.*, 2014:4695) as well as the increase in complexity of the supply chain due to rising levels of customer demand (Jain *et al.*, 2017:1109). However, the reality is that many organisations still struggle to differentiate themselves at the supply chain level (Souza, 2014:604). Even those organisations that have invested in the use of BA or SCA have struggled to translate the information into effective decision-making.

The most important reason for BDA, and specifically for SCA is to assist organisations in making better business decisions, based on sound information. It is therefore evident that analysis on the effect of SCA on business performance and decision-making is required to ascertain and quantify the competitive advantage SCA provides with a specific focus on the operational performance of an organisation. SCA is a relatively under-researched area of BA, and therefore an investigation into the effect of SCA on business decision-making is warranted.

1.3 Objectives of the study

1.3.1 Primary Objective

The main objective of this study was to perform an analysis of the impact of Supply Chain Analytics on business decision-making.

1.3.2 Secondary Objectives

The secondary objectives of the study were:

- To find theoretical evidence of a possible relationship between SCA and business decision-making.
- To quantify the effect of SCA on business decision-making.
- To recommend how SCA could be employed to improve operations.

1.4 Scope of the Study

The study tapped into the experience base of a significant sample of information officers and professionals in South African companies to understand the effect of SCA on business decision-making.

1.5 Research Methodology

1.5.1 Literature Study

The literature study covers the following areas:

- *Big Data*, which refers to the massive amounts of data internal and external to an organisation that can be analysed to assist businesses to find patterns that may assist in improving some aspects of the business.
- *Big Data Analytics*, which refers to the application of Business Analytics to Big Data.
- *Supply Chain Analytics*, where SCA is applying BDA to the field of Supply Chain Management, that includes *Data-driven Business Decision-making*, where decision-making is the process of evaluating multiple possible options to produce a set of possible outcomes and then choosing the most appropriate option.

1.5.2 Empirical Study

An empirical study was carried out that identified evidence of a relationship between SCA and business decision-making. Convenience sampling was performed using a questionnaire that was based on the literature study (Bryman & Bell, 2017:178). The sample included information officers in various companies that were involved with, and therefore knowledgeable on the topic of SCA and its effect on decision-making. Descriptive statistics and inferential statistics such as multiple regression was utilised to focus on causal relationships between the variables (Bryman & Bell, 2017:322). Only demographic information that highlighted the respondents' knowledge and experience of the topic at hand was employed.

1.6 Limitations of the Study

The study was limited to data from information officers and professionals in South African companies who were knowledgeable on the topic of SCA and was either in a decision-making position or provided others with decision-making information.

Another major limitation was inherent in the method that was used. Questionnaire data at best test people's perception of the topic, and it would have been very hard, if not impossible, to establish the real quality of decision-making from a questionnaire only. Ideally, the data should have been superimposed to business results, but that fell outside the scope of this study and may be the object of a future study. Within this limitation, the relationship between SCA and business decision-making was established.

1.7 Significance of the Study

A recent study has found that there is a limited body of empirical research that discourses how the supply chain, and SCM in particular, is impacted by the capabilities of SCA (Arunachalam *et al.*, 2018:417). A further study determined that the current literature highlights a disparity between practices of the supply chain and its inherent theory with regards to Supply Chain Analytics (Wang *et al.*, 2016:107). Yu *et al.* (2018:371) assert that empirical research done to evaluate the effect of big data-driven supply chains on supply chain capabilities is limited.

The above overview of the literature clearly shows that many empirical research studies still need to be conducted in the domain of SCA. The influence SCA has on inventory decisions, and its optimisation specifically has not been studied exhaustively and requires further empirical work.

The low-growth economy at the time of the study has put businesses in distress and made effective business decision-making critical, especially with the abundance of data available (Saleh *et al.*, 2018:450). The further importance of the study was to show how implementing SCA could have a positive optimising effect on operations processes allowing better business decisions to be made – decisions that were based on insights derived from data rather than relying on intuition (Arunachalam *et al.*, 2018:416).

1.8 Layout of the Study

Chapter 1 presented the background and setting for this research study. It further contained the problem statement and clear objectives and limitations of the study. Chapter 2 introduced the core literature study that delved into the fundamentals on which the research is based as well as the landscape of the topic in practice. Chapter 3 covered the empirical study and investigated the theory from the literature study, addressed the research objectives and discussed the analysis to the study. Chapter 4 provided conclusions to the study and recommendations based on the study.

CHAPTER 2: LITERATURE STUDY

2.1 Introduction

This chapter identifies and discusses the main underlying areas covered by this study. This literature study unfolds what is already known about these areas and determines what concepts and theories are relevant to the study.

Further to this, the strategies and research methods that have been used in the past are identified as well as highlighting controversies in such research. Inconsistencies in findings are also explored.

The main areas that are covered in this literature study are Big Data, Big Data Analytics and Supply Chain Analytics that includes the aspect of business decision-making.

2.2 Big Data

The concept of Big Data was first used in the latter part of the 1990s and was then further refined in the early 2000s through the 3V model (Blazquez & Domenech, 2018:99). Mayer-Schönberger and Cukier (2013) stated that Big Data is set to revolutionise how we live, how we work, and how we think. According to Ularu *et al.* (2012:3), Roger Magoulas was the person who invented the term 'Big Data' in 2005. He wanted to describe the vast of amounts of data that was too large, too complex and too unstructured to be processed by conventional database management systems, architectures and techniques. Scientific publications on Big Data started coming out in 2008 (Ularu *et al.*, 2012:3).

The following sections explore how Big Data is defined, the dimensions of Big Data, the sources of Big Data, the challenges and architecture of Big Data systems and the uses, objectives and benefits of Big Data.

2.2.1 Defining Big Data

Big Data is defined in the Oxford English Dictionary (2016) as "extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions". Gartner states that "Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (Gandomi & Haider, 2015:138). Ularu *et al.* (2012:4) argue that the way Big Data is defined has implications for management of companies as this pave the way for a competitive advantage strategy.

Big Data has great importance for the future. It has the potential to unlock efficiencies derived from the varied use of the enormous amounts of data from various sources, both internal and external, to organisations. It allows organisations to improve their view of their business through a better understanding of their business and all external factors that affect their business. Whether analysing data for patterns in customer service-centre log files, analysing data patterns in social media related to the industry or analysing financial data to perform a risk assessment of its customers, organisations stand to benefit from the use of Big Data in many of its operational areas (Ularu *et al.*, 2012:5). Deokar *et al.* further argues that Big Data has implications for the larger society in addition to the implications to organisations (2018:10).

2.2.2 Dimensions of Big Data

Since the inception of the term Big Data, many dimensions (or characteristics) thereof have been identified. Initially, the "three V's" of volume, variety and velocity emerged as a way to describe Big Data (Lee, 2017:294). Volume is the most common descriptor (Yang *et al.*, 2017:14).

Later Wamba *et al.* (2015:236) added more dimensions of Big Data, namely veracity and value. Gandomi and Haider (2015:139) also added variability and complexity. All these dimensions tell a story of the character of what Big Data encompasses and stands for. The following paragraphs shed some light on the details of these dimensions.

2.2.2.1 Volume

Volume refers to the sheer enormity of data, typically in extremely large data sets (Wamba *et al.*, 2015:235), with Big Data sizes typically reported in petabytes (PB) and exabytes (EB). In 2014 Facebook has already collected more than 300 petabytes since its inception (Zheng *et al.*, 2018:366) and grew by 500 TB per day (Xu *et al.*, 2016:1562). Sivarajah *et al.* (2017:263) state that 2.5 exabytes of data were generated each day worldwide in 2014.

The data sizes of Big Data are relative and vary based on elements such as type and time (Gandomi & Haider, 2015:138). What was enormous a year ago, might be just big now, which stems from organic increases in storage capacities.

2.2.2.2 Variety

Variety refers to the fact that data comes from multiple and diverse sources and exists in different formats (Wamba *et al.*, 2015:235). Data may either be structured, unstructured or somewhere in between. Structured data only makes up around 5% of all current data (Gandomi & Haider, 2015:138). Structured data refers to data that is in tabular format with explicit data types and predefined fields. All spreadsheets and relational databases contain structured data and can be easily analysed by machines.

Unstructured data is not in a format that lends itself towards being analysed by machines – at least not easily. Examples of unstructured data are text, video, audio and images.

Semi-structured data lies somewhere between structured and unstructured data. It comprises data that may contain machine-readable elements, for example, mark-up languages such as Extensible Mark-up Language (XML) or Hypertext Mark-up Language (HTML). These contain data tags and defines schemas that allow the machine to read and interpret the data (Gandomi & Haider, 2015:138).

It is notable that unstructured data has been collected by companies for many years and is not a new concept. What is new is the fact that, with the advent of Big Data technologies, this data can be analysed and interpreted computationally (Li *et al.*, 2018:301). For example, using optical recognition for interpreting video can provide intelligent analytical data regarding the movement of stock, humans or any other elements being monitored (Li *et al.*, 2018:302).

2.2.2.3 Velocity

Velocity refers not only to the speed at which data is accumulated but also to the rate at which data changes (Wamba *et al.*, 2015:235). This imposes a lower limit on the speed at which data should be processed and analysed. Otherwise, data would be collected faster than it can be acted upon negating the value that can be created through analysis. This concept is an important distinguishing factor concerning the difference between conventional database management systems and Big Data architectures and technologies.

2.2.2.4 Veracity

Veracity refers to the degree to which data is truthful, trustworthy, objective, accurate and of good quality (Yang *et al.*, 2017:14). Gandomi and Haider (2015:138) further identify the underlying unreliability of some data sources necessitating the need for uncertain and inaccurate data to be dealt with. The term 'veracity' was coined by IBM (Gandomi & Haider, 2015:139). Veracity contends that even though data might be inaccurate or uncertain, it may yet be valuable and needs to be analysed to determine its true worth. Ularu *et al.* (2012:9) state that a third of business leaders inherently mistrust the data they use in their decision-making process. This fact highlights the need to create data of good quality that is usable for decision-making bearing in mind the growth of data in both diversity and size.

2.2.2.5 Variability and Complexity

Variability refers to the fluctuation in data velocity, where data may not always change at the same rate and rates can vary over time. Complexity refers to the fact that often Big Data originates from multiple sources and needs to be cleansed, transformed and matched (Gandomi & Haider, 2015:139).

2.2.2.6 Value

Value refers to the fact that Big Data in its original form typically has a "low-value density" (Gandomi & Haider, 2015:139) as compared to the total volume of data. This concept means the value can only be elicited through rigorous analysis. Value highlights the fact that the data and its use need to bring about economic gain. Günther *et al.* (2017:191) argue that value reaped from Big Data depends on the strategic goals of the organisation and their intended use of Big Data. They further distinguish between pure economic value derived from the use of Big Data as opposed to the social value from Big Data.

2.2.2.7 Summary

Summarising the dimensions of Big Data, there exists an underlying interdependence between the dimensions. Changes in one dimension normally impact on one or more of the other dimensions. This is especially true for volume, variety and velocity where a critical point exists after which conventional analysis technologies are no longer adequate and Big Data technologies should be employed (Gandomi & Haider, 2015:139).

2.2.3 Sources of Big Data

Lee (2017:294) states that Big Data has numerous sources. Günther *et al.* corroborate this, arguing that Big Data originates from multiple distinct sources (2017:195). Big Data can come from within an organisation in the form of data from internal systems. Examples of this are Enterprise Resource Planning (ERP) systems that contain transactional data – data from its operations such as data logged on machinery or video data logged through security surveillance systems. Data can be further be generated by users such as that being logged during maintenance to identify the actual problem or data logged about who enters the premises. Organisations generate huge volumes of both structured and unstructured data at a high velocity (Günther *et al.*, 2017:195).

A McKinsey report states that the manufacturing sector generates more data than any other sector amounting to around 2 exabytes of data stored per year, as recorded in 2010 (Yin & Kaynak, 2015:143).

Big Data also originates from external sources. Big Data can come from continuous autogenerated data streams, such as the Internet of Things (IoT) sensor data (Ahmed *et al.*, 2017:459). An enormous source of Big Data also comes from social networks like Facebook, Twitter and Instagram. On these platforms, data is primarily user-generated and enhanced by algorithms that facilitate the flow of data (Ghani *et al.*, 2018:2).

IoT was born from technological advances in wireless networks, micro-electromechanical systems (MEMS) and consumer electronics (Ahmed *et al.*, 2017:458). IoT does not only consist of devices such as cell phones, laptops and tablets but also of wearable devices, smart household appliances and other internet-connected sensors. It is expected that the number of IoT devices will more than double from 23 billion devices in 2016 to a staggering 50 billion devices by 2020 (Ahmed *et al.*, 2017:459).

IoT contributes to Big Data by generating endless streams of data from mobile devices, sensors, tags and cameras located across the world (Yang *et al.*, 2017:17). It is one of the biggest sources of Big Data even though the current solutions to Big Data are in its early stages (Ahmed *et al.*, 2017:468). The range of applications of IoT in conjunction with Big Data is vast and forms part of the driving force behind the 4th industrial revolution (also termed Industry 4.0) realising self-controlled and self-optimised systems (O'Donovan *et al.*, 2015:17).

Social media is another key source of Big Data (Bello-Orgaz *et al.*, 2016:55). Social media builds on the premise that users modify content continually as opposed to the conventional approach of creating and publishing content by key individuals. Social media is, therefore, characterised by three aspects namely that users are allowed to create profiles for themselves, link to other users on the network and interact and connect with other users on the network (Ghani *et al.*, 2018:2). The well-known examples of these social media networks are Facebook, YouTube, Instagram, Twitter and LinkedIn.

2.2.4 Challenges of Big Data

As with any disruptive technology, Big Data also faces significant challenges both at technical as well as management level to organisations that wish to reap the multitude benefits offered by Big Data (Lee, 2017:294). Some of the challenges faced by these organisations are discussed in the following sections (Yang *et al.*, 2017:18).

2.2.4.1 Data Storage

Traditional storage is often inefficient to store Big Data due to its large volume and speed of change. Storage also needs to be able to rescale rapidly. Enterprise cloud storage services (like Amazon S3 or Microsoft Azure platforms) cater well for the shortcomings of traditional storage in that they scale well, possess high availability, durability and provides almost unlimited storage space (Chen *et al.*, 2014:178). However, the transmission of the data to cloud storage is costly due to the volume and speed of change required. Thus, algorithms to cater for these limitations still needs to be further developed and matured (Yang *et al.*, 2017:19).

2.2.4.2 Data Transmission

Data needs to be transferred from the source of the data to the storage facility. Then data may need to be integrated from multiple data centres from where the data needs to go to the platform on which it will be processed. Finally, the data needs to be transmitted to the platform on which it will be analysed. In each stage, transmitting large amounts of data poses challenges. There is a need for algorithms to pre-process and compress the data before sending (Yang *et al.*, 2013:2652).

2.2.4.3 Data Management

It is an arduous task for conventional computers to store, manipulate and analyse large sets of unstructured data in an efficient manner. Management of Big Data calls for a paradigm shift to embrace new technologies (for example, Apache Hadoop) for purging, storing and organising heterogeneous data (Kim *et al.*, 2014). Conventional database management systems do not provide sufficient scalability essential to Big Data processing and storage (Chen *et al.*, 2014). Although non-relational databases like Apache HBase have been developed for Big Data, the major challenge still exists on how to index the data and to provide efficient querying of the data (Li *et al.*, 2017).

2.2.4.4 Data Processing

Processing large data sets highlight the need for computing resources that are dedicated and having sufficient processing power. When specifically referring to Big Data processing, computer resources requirements are exponentially greater (Ammn & Irfanuddin, 2013). Nearly unlimited processing resources can be attained by using cloud computing, but this does not come without its challenges.

Processing Big Data in the cloud means the data needs to be transferred to the cloud through limited network bandwidths. This limitation negates the improvements offered by cloud processing power (Ammn & Irfanuddin, 2013; Mayer-Schönberger & Cukier, 2013:47).

Another aspect that is challenging is the concept of data locality. Platforms like Hadoop strives to keep computation close to the data (as a design pattern), but this is contrary to the concepts of virtualisation, which is based on the pooling of data. These two concepts make it more challenging to ensure data locality in cloud computing (Yang *et al.*, 2013:2651).

Pre-processing of data needs to be utilised to minimise the impact of veracity on Big Data. Pre-processing would further allow better quality data and ensure less data is ultimately processed (Mayer-Schönberger & Cukier, 2013:46). Zhai *et al.* (2014) argue that more efficient data reduction algorithms are necessary to eliminate data that may likely be "irrelevant, redundant, noisy and misleading". The algorithms further need to be fast enough to process streaming data from various sources (Zhai *et al.*, 2014).

2.2.4.5 Data Analysis

Analysing Big Data is a crucial step towards interpreting and extracting meaning from Big Data. This analysis is no mean feat requiring complex and scalable algorithms (Khan *et al.*, 2014). Big Data Analysis is achieved by combining processing platforms (such as Hadoop) to analysis algorithms.

There are two sides to the challenge of analysing Big Data (Yang *et al.*, 2017:18). On the one side, analysis algorithms need data that is structured and homogeneous – a characteristic that is foreign to what the bulk of Big Data is. On the other hand, there is a need for new algorithms to pre-process the unstructured Big Data so that it can be analysed using existing algorithms (Yang *et al.*, 2017:19).

2.2.4.6 Data Visualization

Visualisation of Big Data is a mechanism to expose patterns and possible cross-correlations of certain aspects of the data to enhance decision-making (Nasser & Tariq, 2015). Padgavankar and Gupta argue that visualisation is the key to find meaning in a heterogeneous set of Big Data (2014:2219). Bello-Orgaz *et al.* further identifies numerous interesting challenges related to data visualisation of social media (2016:56).

2.2.4.7 Data Integration

Value from Big Data is largely achieved through data integration from different domains (Christen, 2014). The main issues with data integration are linking different data records, combining the data and mapping data between different schemas (Dong & Divesh, 2015). The availability of metadata is crucial to facilitate such integration in an automated manner (Agrawal *et al.*, 2011). Generating such metadata from Big Data is a further challenge in itself (Gantz & Reinsel, 2011).

2.2.4.8 Data Architecture

Due to the complexity of Big Data, the architecture to capture value from Big Data is still not mature (Wright & Wang, 2011:5489). The ideal infrastructure is one where the analysis and syntheses of Big Data together with the sharing of data, networks, models and human resources would be seamless (Wright & Wang, 2011:5489).

2.2.4.9 Data Security

Conventional algorithms, encryption standards and methodologies, when applied to Big Data, become troublesome from a data security perspective (Villars *et al.*, 2011). Data security algorithms are mainly focused on structured data that is not effective in securing heterogeneous data (Villars *et al.*, 2011). New data security systems, data access control procedures and policies and storage mechanisms need to be looked into to facilitate Big Data (Chen *et al.*, 2014). Lee (2017) proposes the use of Blockchain as data security management mechanism of the future.

2.2.4.10 Data Privacy

Social media and other public platforms contain personal information that is sometimes not well protected or hidden. Where this data forms part of Big Data analysis, data privacy becomes a concern (Cheatham, 2015:334). Also, where data is gathered internally to an organisation for workforce monitoring and analysis, privacy becomes a matter that needs attention even though policies or regulations might not explicitly deal with this (Eisenstein, 2015:S4).

Bello-Orgaz *et al.* corroborates the notions above and discusses two well-known methods of privacy preservation when analysing social media but argues that more research is required to ensure the open issues to privacy are addressed (2016:56).

2.2.4.11 Data Quality

Data quality pertains to four characteristics, namely accuracy, redundancy, consistency and completeness (Chen *et al.*, 2014). Due to the inherent complexity and lack of homogeneity of Big Data, identifying and tracking data accuracy and completeness is arduous with a high risk of false positives (Mirzaie *et al.*, 2019:28). Besides, the rate of change of Big Data causes data integrity as well as data consistency to be problematic

(Khan *et al.*, 2014). Wamba *et al.* (2015:243) note that Big Data needs to be of good quality, regardless of it being messy, and irrelevant data needs to be eliminated.

Data extracted from social media varies from highly accurate to completely untrue. Times and location accuracy depend on several subjective factors. Data redundancy is also an issue as multiple versions of the same data may exist and requires correlation and correctness checking (Khayyat *et al.*, 2015:1227).

2.2.4.12 Data Skills

There is a big scarcity in data scientists who are skilled enough to understand both IT principles and can manipulate Big Data (Davenport & Dyché, 2013:14). The number of people that fall in this bracket is few. They are being enveloped by large organisations who understand the need to invest in Big Data technologies and resources (Ularu *et al.*, 2012:5).

2.2.5 Objectives, Uses and Benefits

Davenport and Dyché (2013:2) conducted a study on 20 large organisations to determine the impact and degree of Big Data in the existing data and analytical structures. It emerged that in all companies, Big Data has been integrated into the existing structures well that speaks of a new management perspective on Big Data. The key to integration is being able to analyse diverse data.

The main objectives of Big Data can be described as (Davenport & Dyché, 2013:7):

- Cost reduction
- Time reduction
- New products
- Decision support

Initially, Big Data meant that companies could reduce the cost of managing their data. Nevertheless, the objective has shifted to creating value from the use of Big Data to gain a competitive advantage in business (Ularu *et al.*, 2012:6).

Yin and Kaynak (2015:144) argue that the main objective of Big Data, specifically in industrial applications, is to realise cost-effective operations processes at the required performance levels. A McKinsey report (Yin & Kaynak, 2015:143) debates that Big Data may contribute up to 50% in cost savings for manufacturing firms as well as decreasing working capital.

Time reduction results from being able to process and analyse data so much quicker than in the past through Big Data tools and technologies. Time reduction is further evident in the real-time nature of Big Data analysis and the ability to respond to changing inputs and variables in a real-time manner (Davenport & Dyché, 2013:5).

New products that can and have originated from the use and insights brought about by Big Data abound. For example, smartphones contribute to a better understanding of customer needs and patterns that create new opportunities for those willing to dissect the available plethora of Big Data (Yin & Kaynak, 2015:144).

Big Data is not just about being big but about the insights from data previously not analysed due to emergent data types and unfamiliar content. This allows organisations to be nimble in their decision-making by providing options from a broader base of inputs and possibly support in the form of more analytical decision streams (Ularu *et al.*, 2012:12).

The benefits of Big Data will differ between sectors where the IT, finance, government, manufacturing and insurance sectors are anticipated to reap the most benefit from Big Data (Yin & Kaynak, 2015:146). Big Data can open up substantial value in product development, customer experience, decision-making and operational efficiencies. Besides, customer analysis has been rated a high benefit as an outcome from the use of Big Data analysis (Yin & Kaynak, 2015:145). This notion is further enhanced by sensor-based feedback from customers to understand needs and behaviour better (Yin & Kaynak, 2015:145).

One of the most important benefits of Big Data is not having and managing large volumes of data but the power to analyse the various data types and merging it with existing structures (Davenport & Dyché, 2013:2). Bello-Orgaz *et al.* (2016:52-54) describes several socio-economic benefits from Big Data, such as crime analysis and epidemic intelligence. Ghani *et al.* (2018:2) concur that many applications have emerged from Big Data that allows a better understanding of human behaviour, for example, when using trend analysis and sentiment analysis in marketing.

2.2.6 Conclusion

Never in history has there been such a proliferation of data. Big Data has opened doors to untapped resources more than ever before. Big Data is already playing a monumental role in the Fourth Industrial revolution (termed 'Industry 4.0') and has been identified as a major catalyst thereof (Yang *et al.*, 2017:18). Davenport and Dyché (2013:9) state that Big Data can be singled out as the most impactful trend to disrupt current IT and business infrastructure over the last decade.

The initial industrial revolution was built on water and steam power. The 2nd industrial revolution focussed on mass production and electricity while the 3rd industrial revolution laid a foundation built on semiconductors and information technology, enabling manufacturing automation. Big Data is a solid driving force for Industry 4.0 (Yin & Kaynak, 2015:144).

Many challenges still need to be faced to see the revolution through. The 4th industrial revolution utilises IoT in conjunction with Big Data to expand current products and systems to the realm of self-controlled and self-adapted systems. Cloud computing and the data integration possibilities it brings along are also key drivers of the 4th industrial revolution (Yang *et al.*, 2017:18).

In all of this, Big Data is the key dependency. Big Data Analytics is the mechanism to turn the captured data into organised and meaningful data that can be utilised for decision making (Ularu *et al.*, 2012:6).

The following section will focus on Big Data Analytics.

2.3 Big Data Analytics

Davenport and Kim (2013) define analytics as finding, interpreting and communicating patterns within data that are significant and then applying these patterns towards effective decision-making. Analytics can be seen as the glue that binds together data to achieve effective decision-making. According to Wang and He (2016:27), Big Data Analytics is the primary phase in the Big Data application chain.

Although many forms of analytics exist, for example, video analytics, web analytics, social analytics (Choi *et al.*, 2018:1869) and social big data analytics (Bello-Orgaz *et al.*, 2016:46), this section will focus on Big Data Analytics as the primary input to Supply Chain Analytics.

2.3.1 Defining BDA

BDA, which includes Business Analytics, has been identified as one of a few key technology trends of the early 21st century (Chen *et al.*, 2012:1165). BDA is often referred to as a field related to Business Intelligence and Analytics (BI&A), which is mostly concerned with data mining and statistical analysis; but BDA further builds on this (Côrte-Real *et al.*, 2017:380). Côrte-Real *et al.* (2017:380) define BDA as "a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery and/or analysis". Nguyen *et al.* (2018:254) define Big Data Analytics as a mechanism to elicit knowledge from Big Data using advanced techniques to facilitate data-driven decision-making. The emphasis on advanced analytical techniques as applied to Big Data is also raised by Rossum (Hazen *et al.*, 2018:202). BDA is seen as a mechanism of extracting value from Big Data.

Business Analytics, on which BDA is built, can be classified into three types of analytics (Souza, 2014:596) namely Descriptive Analytics, Predictive Analytics and Prescriptive Analytics.

Descriptive Analytics looks at historical data to derive insights and patterns. It attempts to answer the question of what is happening currently (Wang *et al.*, 2016:99). It feeds on real-time data from various data sources and technologies including, among other things, barcoding, RFID, GPS, IoT and other real-time tracking technologies (Souza, 2014:596). The outputs from descriptive analytics are reports that render "historical insights" (Tiwari *et al.*, 2018:320) pertaining to the organisation, for example, inventory and operational reports.

Predictive Analytics utilises techniques such as predictive modelling (forecasting) and machine learning to understand the future (Tiwari *et al.*, 2018:321). It attempts to answer the question of what will be happening (Wang *et al.*, 2016:99). It is used to assist in forecasting demand. It is also useful to predict information that is not available or where gaps exist (Souza, 2014:596).

Prescriptive Analytics employs simulation and optimisation algorithms to recommend possible options and provide details of the effect of each outcome. It attempts to answer the question of what should happen in the future (Wang *et al.*, 2016:99). It utilises descriptive and predictive analytics models to provide various alternative outcomes (Souza, 2014:598). To achieve this, it also uses “mathematical optimisation, simulation or multi-criteria decision-making techniques” (Tiwari *et al.*, 2018:320). The key advantage of prescriptive analytics is that it assists businesses by offering various alternatives with outcomes to optimise processes (Wang *et al.*, 2016:100). Of the three types, most research and organisational activity happen in the area of Prescriptive Analytics (Souza, 2014:596).

2.3.2 Techniques of BDA

Choi *et al.* (2018:1869) distinguish three different ways of processing data. These are shown together with example platforms that support such processing:

- Batch processing – Data is processed in batches. Apache Hadoop is a commonly used platform for batch processing.
- Real-time or stream processing – Data is processed as it is captured in real-time. SAP Hana supports stream processing for Big Data.
- Interactive processing – Apache Drill or Google’s Dremel are sometimes preferred for this type of processing.

The key techniques used for data analysis of Big Data are statistics, machine learning, data mining, optimisation (Choi *et al.*, 2018:1869). Other BDA techniques may include visualisation analysis, clustering analysis and social network analysis (Choi *et al.*, 2018:1869). These four techniques will now be discussed.

2.3.2.1 Statistics

The most common technique for analysing Big Data is statistics since it is a proven technique laying a scientific base for gathering, analysing and interpreting data. The statistical methods to draw correlations and regressions are well-known and often used. However, there are major challenges with using simple statistics on its own for BDA due to the heterogeneous and unstructured nature of the data (Sivarajah *et al.*, 2017:264).

Gandomi and Haider (2015:143) agree with the latter stating the need for novel statistical methods aimed at analysing Big Data.

2.3.2.2 Machine Learning

Machine Learning is the study of using algorithms and statistical models to allow computers to learn by finding patterns and logical conclusions in the data and either classify or make predictions based on these findings (L'Heureux *et al.*, 2017:7777). Machine Learning is closely related to predictive analytics (Jordan & Mitchell, 2015:255). Machine Learning is a fundamental element of Data Analytics and is also seen as a sub-domain to Artificial Intelligence (Choi *et al.*, 2018:1870).

Although various approaches to Machine Learning exist, the most common approaches are supervised and unsupervised learning. In supervised learning, a mathematical model or algorithm is built using training data that contains both the inputs and required outputs (Jordan & Mitchell, 2015:257). As the algorithm is fed new inputs, optimally, it starts to predict the outputs from what it has previously learnt. In unsupervised learning, only inputs are provided, and the algorithm uses commonalities to find the structure of the data together with patterns or trends (L'Heureux *et al.*, 2017:7778).

Machine Learning has the unique ability to learn from data to provide valuable predictions and data-driven inputs to decision-making. It is a common misconception that Machine Learning that uses more data inputs produce better results. However, the opposite is true and traditional Machine Learning algorithms, therefore, needs to be adapted to effectively work with Big Data (L'Heureux *et al.*, 2017:7777). Other paradigms of Machine Learning include Deep Learning, Online Learning and Lifelong Learning, among others (L'Heureux *et al.*, 2017:7792).

Machine Learning is extremely resourceful and adaptable to use as a technique to secure complex data problems. However, on the negative side, it has the weakness of being time-consuming to train (Choi *et al.*, 2018:1870).

2.3.2.3 Data Mining

Data mining is the process of using algorithms for discovering patterns or insights from large datasets and, according to Choi *et al.* (2018:1870), data mining is a key foundation for both BDA and Business Intelligence. It converges methods from statistics, database management and machine learning (Amato *et al.*, 2018:288).

Data mining is a key process step in the Knowledge Discovery for Databases (KDD) (Ristoski & Paulheim, 2016:2). KDD is a broad process of discovering knowledge in large data sets and forms the underlying foundation for pattern recognition, statistics, artificial intelligence and machine learning for use in data visualisation and expert systems (Ristoski & Paulheim, 2016:3). Data mining is one of five key steps in the classical KDD process.

The two primary goals of data mining are description and prediction – where description relates to discovering human-predictable patterns to describe the data and prediction involves utilising fields in the data to predict previously unknown knowledge and patterns (Fayyad *et al.*, 1996:38). The primary data mining tasks or methods include classification, regression, clustering, summarisation, dependency modelling and deviation detection (Fayyad *et al.*, 1996:42).

Data mining is very good at dove-tailing models from different techniques to allow it to deal with multiple data types but suffers from the disadvantages of the fundamental models themselves (Choi *et al.*, 2018:1870).

2.3.2.4 Optimisation

Computational optimisation is a standard technique for discovering the optimal or near-optimal result in decision-making problems of a quantitative nature (Choi *et al.*, 2018:1870). Many challenges exist when applying optimisation to BDA related to memory and processing of data, especially as it relates to real-time optimisation (Choi *et al.*, 2018:1870).

2.3.3 Strategies of BDA

Analysing Big Data poses certain challenges both from the fact that data is of high volume, variety and complexity, and from the computational perspective where current analysis methods are not able to adapt to these challenges. Different strategies, therefore, need to be adopted to cater for BDA. Choi *et al.* (2018:1871) propose some strategies for BDA that will be discussed briefly.

2.3.3.1 Divide and Conquer

Divide and Conquer is one of the primary strategies for processing Big Data (Wang & He, 2016:28). This strategy has been used for many years to process data and consist of breaking the data down to small chunks, then processing those chunks individually and then combining the individual results again (Wang & He, 2016:29).

Divide and Conquer is closely related to the emerging field of granular computing that performs a similar action to the above by breaking down large sets of data into information granules upon which processing is done to abstract data and information to derive knowledge (Choi *et al.*, 2018:1871).

2.3.3.2 Distributed and Parallel Processing

Distributed and Parallel Processing (DPP) allows Big Data to be processed using multiple parallel processing computer systems that is a similar concept to Divide and Conquer. However, DPP performs the processing on the original data without physically breaking the dataset into various chunks providing it with a very high level of flexibility (Choi *et al.*, 2018:1871).

Wang and He (2016:29) notes that DPP can function at different levels, such as bit-level, instruction-level or task level. It is further important to understand that not all data or problems can be parallelised and that parallelisation only lessons time to process and does not minimise the actual workload (Wang & He, 2016:29).

2.3.3.3 Incremental Learning

This strategy relates to algorithms where learning is employed, such as machine learning. It aims to learn one case at a time and only focuses on learning new cases under the assumption that it remembers all prior cases and outcomes in a step-by-step manner to improve the learning algorithm (Wang & He, 2016:29). Learning (or training) is performed on new data blocks that can originate from streamed or batch data (Wang & He, 2016:29).

The major advantage of this strategy is that each data block needs to be used for training only once (Wang & He, 2016:29). The primary limitation of this strategy is that good computational memory is required by the algorithm to remember all trained cases. This strategy may be very challenging, specifically as it relates to large data sets (Choi *et al.*, 2018:1871).

2.3.3.4 Statistical inference

Statistical sampling is not a novel concept and can be used to establish the relationship between the sample and the population (Choi *et al.*, 2018:1871). In BDA, the relationship can be used to understand and justify whether the sample can be used to draw an inference of the entire population, or whether a larger sample or the whole population need to be processed (Choi *et al.*, 2018:1871).

2.3.3.5 Feature Selection

Features (also referred to as attributes or random variables) in a data set increases as the data set size increases. Feature selection is a strategy to reduce the number of features to get to a principle set of features in the data set to simplify the analysis of the data (L'Heureux *et al.*, 2017:7780).

The Hughes effect holds that the higher the number of features in a data set, the lower the effectiveness of a training algorithm becomes causing machine learning algorithms, for example, to degrade in performance and accuracy (L'Heureux *et al.*, 2017:7780). This effect is termed the curse of dimensionality (Choi *et al.*, 2018:1872).

BDA is affected by the curse of dimensionality due to the size and variety of data sets, and this complicates processing during feature selection. This issue also negatively affects the time complexity of the learning algorithms exponentially (L'Heureux *et al.*, 2017:7780).

2.3.3.6 Uncertainty

Due to low veracity (accuracy or reliability) of some Big Data datasets, some datasets might have data that is missing or got lost during the process of eliciting the data. In these cases, Wang and He (2016:29) proposed an uncertainty-based learning strategy to overcome the uncertainty through what they call "fuzziness-based learning".

2.3.3.7 Heuristics

Choi *et al.* (2018:1872) note that in some cases (specifically related to operations management problems) heuristics are still being used where problems are extremely difficult and time constraints are at play. In these cases, heuristics are used to find near-optimal solutions based on numerical techniques operating within specific boundaries. Kasturi *et al.* (2016:89) illustrate using meta-heuristic algorithms as part of BDA to provide more optimal results when analysing Big Data.

2.3.4 BDA Architectures

A different kind of architecture for processing Big Data as opposed to traditional databases is essential since the elements involved are very different (Wei & Yu, 2019:18). In essence, if data is too big or unstructured, or if data needs to be captured, processed and analysed in real-time, a BDA architecture must be considered.

There is no formal architecture for BDA, and many architectures have been proposed (Kashyap *et al.*, 2015:6). Choi *et al.* (2018:1876) and Pääkkönen and Pakkala (2015:168) proposes generic architectures. Figure 2–1 illustrates a generic Big Data architecture (Pääkkönen & Pakkala, 2015:168; Choi *et al.*, 2018:1876) that consist of the major elements normally contained in a BDA architecture.

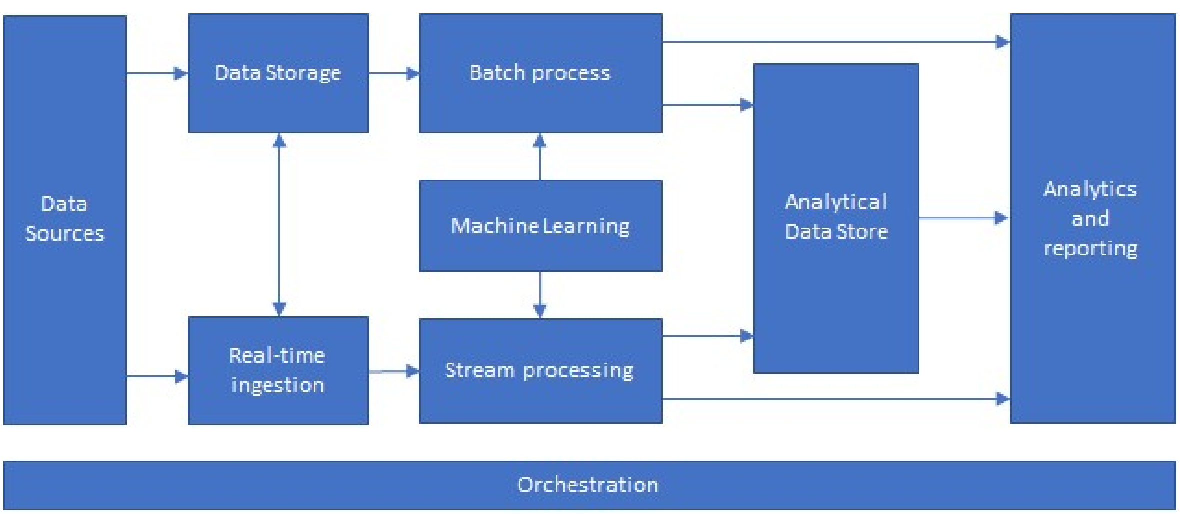

Figure 2–1: BDA Architecture (adapted from Pääkkönen & Pakkala, 2015:168; Choi *et al.*, 2018:1876)

Data sources represent the root of the data to be used and may consist of data from databases, files, logs and other sources of data. Data of various formats are stored in a data store, often termed a data lake (Wei & Yu, 2019:19). For real-time data, a method of data ingestion should be incorporated to extract the data at high velocity.

Low velocity or static data is normally processed using batch processes that may run from minutes to days to filter, sum and prepare data for analysis. Stream processing is utilised to process high-velocity data and also prepare the data for analysis (Wei & Yu, 2019:20).

The analysis datastore is where data that has been prepared is stored. The data in this store is optimised for analysis as opposed to being transaction-based – it is not normalised as in a relational database (Choi *et al.*, 2018:1876). From there the data is ready to be analysed by whatever analysis tools are available. The whole process is governed by an orchestration function that ensures that a continual repeatable process is followed and vetted, resulting in insights being derived from the data (Wei & Yu, 2019:20).

For each of the elements in Figure 2–1, there are various options to consider in terms of software frameworks to use. The data storage can be handled by platforms such as Azure Data Lake Storage. Batch processing can be performed in Apache Hive, Pig, using a HDInsight Hadoop cluster or Azure Data Lake Analytics (Pääkkönen & Pakkala, 2015:168).

Stream processing may be done in Azure Stream Analytics or one of the Apache streaming technologies. The analytics data store can utilise Azure SQL Data warehouse or an open-source option like Apache HBase or Hive (Choi *et al.*, 2018:1876). For the reporting, visualisation and analytics portion Azure Analysis services or Power BI may be used. Lastly, Azure Data Factory can be set up to cater to the orchestration layer (Choi *et al.*, 2018:1876).

2.3.5 Conclusion

Even though the BDA technologies, techniques and strategies have been relatively well-defined, it is not yet being applied to Big Data to the extent one would expect. According to Jain *et al.* (2017:1108), only 0.5% of all available data have thus far been analysed, while only a small percentage of the available structured data have been analysed

(Gandomi & Haider, 2015:137). As of 2016, roughly 33% of all data have been earmarked for analysis, and around 20% of companies have started using Big Data Analytics (Jain *et al.*, 2017:1109).

From the above, it would seem that organisations have not harnessed the power of BDA yet. Grover *et al.* (2018:391) argue that organisations need to decide what strategic role BDA should play. Organisations should invest in capturing and analysing quality data, acquiring quality tools and employing quality data-savvy employees. This investment should be made to establish sound BDA capabilities that can create strategic value and competitive advantage to the organisation (Grover *et al.*, 2018:419).

The following section will focus on Supply Chain Analytics.

2.4 Supply Chain Analytics

Big Data Analytics (BDA) forms the basis for Supply Chain Analytics (SCA). SCA applies BDA to the supply chain to develop effective supply chain strategies and manage the supply chain at both tactical and operational levels (Wang *et al.*, 2016:101). SCA, therefore, forms the nexus between BDA and Supply Chain Management (SCM) (Kache & Seuring, 2017:10).

This section will focus on Supply Chain Analytics. SCA will first be defined and then its capabilities, challenges and opportunities will be explored. Finally, a proposed architecture for SCA will be discussed.

2.4.1 Defining SCA

Supply Chain Analytics (SCA) is a sub-domain of Business Analytics and more specifically applies BDA to the organisational function of Supply Chain Management (SCM). The supply chain is a network of entities involved in the production and distribution of products and services. The supply chain includes all processes and activities from the supplier to the customer (Heizer *et al.*, 2016:44). Jain *et al.* (2017:1106) define Supply Chain Management (SCM) as managing the flow of commodities and services from its origin to where these get consumed.

Within the context of SCM, Arunachalam *et al.* (2018:419) define SCA as the ability of businesses to amass, organise, analyse and visualise heterogeneous data from the supply chain (both internal and external to the business) to enable proactive decision-making within the supply chain. Souza (2014:595) states that Supply Chain Analytics "focuses on the use of information and analytical tools to make better decisions regarding material flows in the supply chain". Wang *et al.* (2016:107) argue that Supply Chain Analytics needs to be seen as an asset of strategic value by companies and that it should be applied holistically. The supply chain performance is to largely a function of information and therefore SCA appears to be a very useful mechanism as it offers a competitive advantage to organisations willing to make it part of their strategy (Kache & Seuring, 2017:11).

Rozados and Tjahjono (2014:6) provide several definitions of SCA but their consolidated finding is that SCA is a process where advanced analytics techniques are applied in combination with supply chain management theory to the Big Data technology stack to gain accurate and timely business insights. Chae *et al.* (2014:4696) state that SCA is collectively constituted from three different sets of resources namely data management resources, supply chain planning resources and performance management resources.

2.4.2 Capabilities of SCA

Arunachalam *et al.* (2018:424) classify capabilities of organisations according to their access to data and information as well as their analytics capabilities. They distinguish four primary capabilities of SCA that include:

- Data generation capability
- Data integration and management capability
- Advanced analytics capability
- Data visualisation capability

In addition to these key capabilities, Arunachalam *et al.* (2018:426) then define additional supporting capabilities namely data-driven culture capability, cloud computing capability and absorptive capability. These capabilities will now be discussed.

2.4.2.1 Data generation capability

This is the ability of organisations to find and access data from different types of data sources both internal and external to the organisation (Arunachalam *et al.*, 2018:425). This capability enables organisations to have data sources to generate data to drive decision-making. Data comes from a variety of data sources (refer to section 2.2.3). The main supply chain technologies are Enterprise Resource Planning (ERP) systems, Warehouse Management Systems (WMS), Customer Relationship Management (CRM) systems, and planning and scheduling systems (Haulder *et al.*, 2019:110). Collaboration between organisations is facilitated through technologies such as Electronic Data Interchange (EDI) and Electronic Supply Chain Management (e-SCM) (Haulder *et al.*, 2019:112). Organisations that can leverage their data generation capability can use it as a competitive advantage to become data-rich and information-rich (Chae *et al.*, 2013:4706).

2.4.2.2 Data integration and management capability

This is the ability to integrate various types of heterogeneous data in real-time over organisational boundaries (Arunachalam *et al.*, 2018:425). This ability increases agility, visibility and performance of the supply chain processes and the views on them as detailed by Wamba *et al.* (2015) that discussed the need to integrate internal and external data. Tan *et al.* (2015) further elaborated on the benefits of incorporating data from a multitude of sources both internal and external to the organisation to improve innovation.

Data sharing, that is implied as part of this capability, could have a major positive influence to optimise the supply chain's performance (Hu *et al.*, 2014:679). Information sharing has become very important in SCM since it added to positive customer experiences and increased supply chain visibility (Biswas & Sen, 2017:2). Information sharing further assists in monitoring certain performance indicators that magnify variances and inefficient processes, for example, the bullwhip effect (Miah, 2015:280). Some challenges do exist with respect to data sharing, namely trust among partners, sensitivity of data/information and inability to access data may become issues causing reluctance to share (Arunachalam *et al.*, 2018:426). It is critical to further ensure that trust relationships between organisations are maintained and that access is limited once relationships are severed.

2.4.2.3 Advanced analytics capability

It is imperative that organisations have advanced analytics ability as this is required to utilise BDA and is one of the most important capabilities (Arunachalam *et al.*, 2018:426). Organisations need to be equipped with the tools, techniques and skills to analyse data in both batch and real-time modes (Souza, 2014:598). As organisations progress from descriptive to predictive to prescriptive analytics (refer to section 2.3.1), the decision-making ability will increase at tactical , strategic and operational levels (Souza, 2014:599).

An analytics capability also requires analytics techniques to be employed. This was discussed in detail in section 2.3.2 and will not be repeated here. As a result of the analytics capability, numerous applications for practice have come to the fore. These will be discussed in section 2.4.3.

Essentially, an analytics capability extracts value from data that was generated in the data (Big Data) collection phase (refer to section 2.4.2.1).

2.4.2.4 Data visualisation capability

This is the ability of the organisation to use tools and techniques to provide processed and analysed information visually to decision-makers enabling them to make timely decisions (Arunachalam *et al.*, 2018:427). Visualisation aims to represent information, and more specifically knowledge, to various stakeholders in formats such as graphs, heat maps and tag clouds (Manyika *et al.*, 2011). These tools and techniques can effectively assist users to understand the information better while enabling them to see patterns more effectively than from raw data and ultimately make decisions based on the information (Manyika *et al.*, 2011).

Monitoring of key performance indicators is a prime example of the use of data visualisation. The studies of Park *et al.* (2016:94) have argued that data visualisation enhances human cognitive levels during decision-making and therefore supports the notion of this ability. Data visualisation is a widely accepted and used concept that is employed by many organisations.

2.4.2.5 Data-driven culture capability

According to Arunachalam *et al.* (2018:427) a data-driven culture is “intangible resource that represents the beliefs, attitudes, and opinion of people towards data-driven decision-making.” The key to this capability is acknowledging and treating data as an asset (Wang *et al.*, 2016:107). Aho (2015:284) further argues that the introduction of Big Data into an organisation necessitates the need to adapt the organisational culture and enforce change management to action this.

Kiron and Shockley (2011:59) proposed three primary characteristics that comprise a data-oriented culture:

- Data and analytics should be regarded as strategic assets;
- Analytics should be supported by top management; and
- Insights derived from analytics should be accessible to those who need them within the organisation.

2.4.2.6 Cloud computing capability

This is the ability to support the integration capability (see section 2.4.2.2) by providing infrastructure in the cloud to facilitate integration of internal and external data sources and storage that can be scaled (Arunachalam *et al.*, 2018:428). This ability also allows challenges (refer to section 2.4.5) of SCA to be overcome or cut down and is seen as a primary enabler to SCA capabilities (Neaga *et al.*, 2015:26). According to Neaga *et al.* (2015:27), cloud computing can enhance logistics performance, information sharing and customer service.

Cloud computing offers various types of service options namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). In addition, services deployment can be in the form of private cloud, public cloud, community cloud or hybrid cloud (Bruque Camara *et al.*, 2015:438).

2.4.3 Applications of SCA

SCA is still in its infancy and many possible applications of SCA have yet to see the light (Kache & Seuring, 2017:11). A few applications of SCA that have already been implemented will be listed and discussed briefly in this section (Arunachalam *et al.*, 2018:426; Tiwari *et al.*, 2018:324):

- Strategic sourcing aims to act as a long-term collaborative partnership with suppliers as part of the organisation's strategy and considers strategic dimensions and capabilities of the suppliers in addition to cost and quality (Jin & Ji, 2013:6824).

SCA can support this by providing specific information to match the organisation to relevant suppliers.

- Supply chain network design can be facilitated with SCA to provide an optimal supply chain based on customer demand, location services and marketing intelligence tools. This may be very useful, for example, in cases such as disaster relief and healthcare (Tiwari *et al.*, 2018:324).
- Product design and development – new innovative products can be developed based on a better understanding of customer behaviour and needs using analytics capability that can result in competitive advantage and supply chain resilience (Tan *et al.*, 2015).
- Demand forecasting can be optimised by improving the accuracy, robustness and speed (real-time). This may have a positive knock-on effect on more accurate production planning and inventory management (Tiwari *et al.*, 2018:325).
- Procurement can utilise SCA to identify and mitigate supply chain risks and manage supplier performance (Wang *et al.*, 2016:7).
- Production can use SCA to assist with shop floor logistical planning and scheduling as well as to improve production efficiencies (Zhong *et al.*, 2015:266).
- Inventory planning and optimisation can utilise SCA by linking the internal production system to the suppliers as well optimising decisions related to the order process (Wang *et al.*, 2016:8).
- Logistics and distribution can use SCA to optimise route planning based on various external inputs and is essential for transportation organisations (Ayed *et al.*, 2015). The supply chain visibility and seamless integration can further be optimised through the use of SCA.
- Data integration has been proven (through empirical studies) to increase access to information as well as data quality (Popovič *et al.*, 2012:734).

SCA provides many potential applications and subsequent benefits to organisations who would risk the few unknowns to transform their traditional supply chains into proactive data-oriented supply chains.

2.4.4 Opportunities of SCA

SCA provides numerous opportunities to organisations to provide logistics and supply chain insights. Some opportunities may be unique to certain organisations providing them with a platform to create a sustainable competitive advantage (Wamba *et al.*, 2018:481). A few studies have performed empirical and Delphi studies to determine the opportunities offered by SCA (Tiwari *et al.*, 2018:327). A study by Kache and Seuring (2015) have identified a number of opportunities at both organisational and supply chain level. Some opportunities that exist at the supply chain level will be listed and briefly discussed (Kache & Seuring, 2015:22).

At logistics level SCA provides several opportunities to optimise logistics across organisational boundaries through the abundance of information available. This enables organisations to track goods within other organisations as if it were part of their own logistics entity. The key enabler is supply chain visibility and transparency allowing end-to-end real-time access to information. Supply chain agility and efficiency are typically improved through the visibility (Kache & Seuring, 2015:24).

Operations efficiency and maintenance is also enhanced by SCA through real-time insights derived along the supply chain allowing more consistent processes to produce leaner supply chains (Zhu *et al.*, 2018:55). Maintenance is further optimised through predictive analytics and automation that results in better asset utilisation assisting in financial profitability to the organisation.

Integration and collaboration allow cross-functional approaches with major partners to tap into time saving activities that are mutually beneficial. Such collaboration requires relationships of trust that can further enhance information-sharing among partners. The key benefit is that decision-making information would be available to any partner through the integrations (Zhu *et al.*, 2018:59).

Inventory optimisation is a critical need and expectation from the use of SCA. The fact remains that due transparency, availability of information and sharing of information insights, the frequency of usable data has grown considerably (Kache & Seuring, 2015:25). This has resulted in shorter planning cycles with inventory management that is more efficient with optimised inventory levels.

2.4.5 Challenges of SCA

Even though the opportunities in SCA might be abundant, the challenges faced to implement SCA need to be carefully considered to avoid the obvious pitfalls. The challenges of BDA were discussed in section 2.2.4. This section briefly discusses the challenges of SCA with the emphasis on the supply chain environment.

Just as integration and collaboration were primary opportunities, they also hold many challenges (Kache & Seuring, 2015:27). Collaboration at a minimum requires partners who want to integrate and sees the need and benefits to do so. Not all parties may see the benefits from the outset and, therefore, might be reluctant to cooperate. The key would be for all parties to table the benefits they need from the collaboration and manage the synergy as an outcome-driven process. It has been shown that a fundamental benefit to all parties is the lowering supply chain risks through collaboration and integration (Zhu *et al.*, 2018:56).

The capabilities of IT and IT infrastructure within organisations differ vastly. Also, the skill levels of IT and data scientists within organisations are difficult to align when seeking to integrate and collaborate at a data/information level (Kache & Seuring, 2015:28). Organisations are increasingly faced with up-skilling and training their resource base to cope with the changes in technology and data management (Arunachalam *et al.*, 2018:420). A data-driven culture and organisational data strategy will assist to lessen the challenges presented.

Similar to BDA (see sections 2.2.4.9 and 2.2.4.10), SCA also suffers from privacy and security challenges. Organisational policies and procedures will need to be adapted to address these challenges and data governance initiatives will be required within the integration and collaboration activities (Arunachalam *et al.*, 2018:419).

Blackburn *et al.* (2015:422) add the challenge presented by SCA of being time-consuming. They state that due to the complexity of the supply chain, it requires the support of senior management, key stakeholders and experts from various organisational functions to build the analytics capability in the organisation.

2.4.6 SCA Architecture

An architecture for SCA that builds on the BDA architecture (shown in Figure 2–1) but incorporates various elements of the supply chain is depicted in Figure 2–2. The generic architecture was adapted from Biswas and Sen (2017:19).

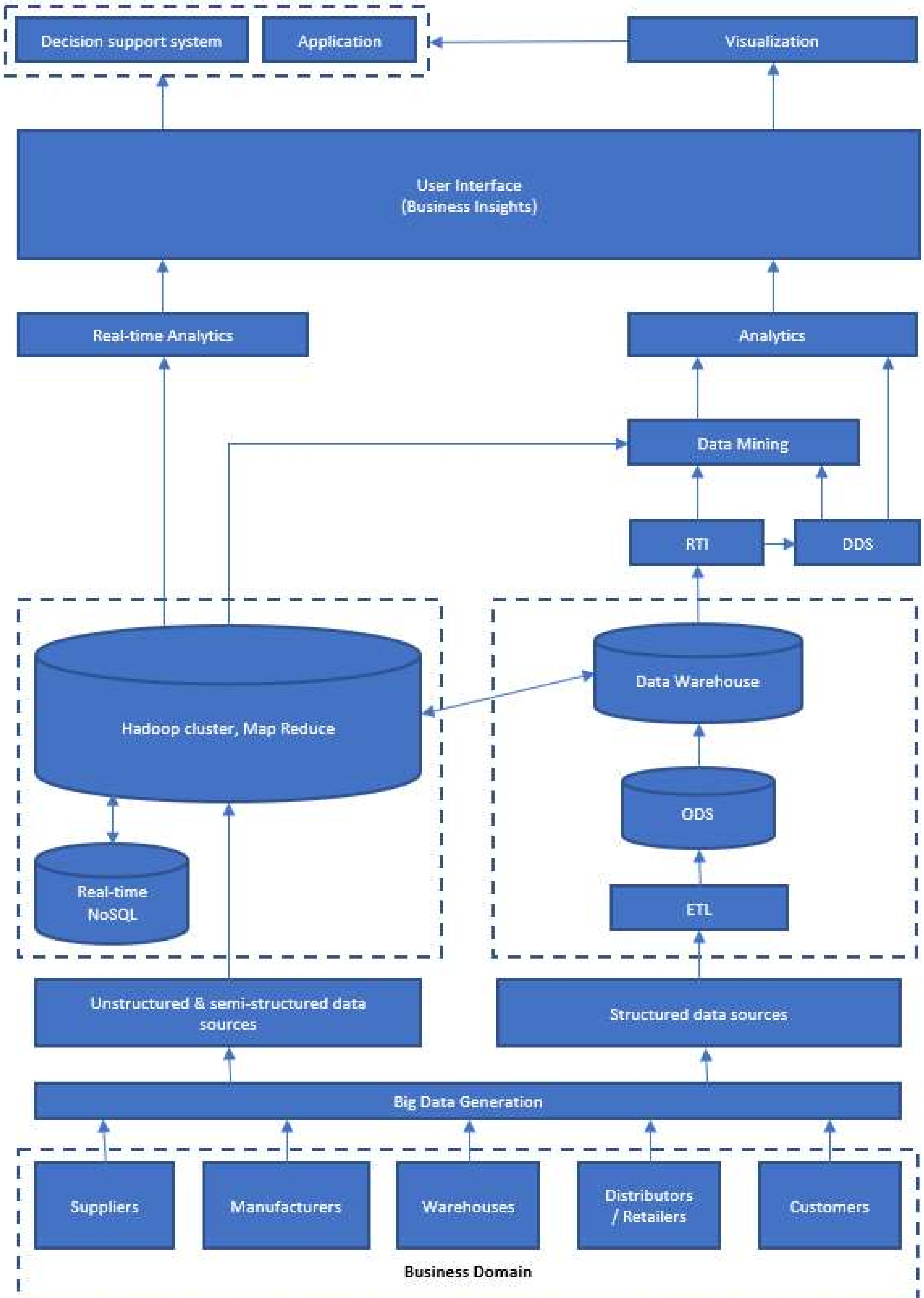

Figure 2–2: Generic SCA architecture (adapted from Biswas & Sen, 2017:19)

The architecture in Figure 2–2 shows that the key supply chain data sources provide the input data to the system (in the form of both structured and unstructured data). The supply chain elements are suppliers, warehouses, manufacturing, distributors and retailers as well as customers. The structured data originates from the typical data sources such as ERP systems. The unstructured data comes from various inputs, for example, RFID tags and other sensors. The combination of this data provides the Big Data Generation hub.

The structured data and unstructured data processing paths differ. The structured data is extracted using Extract-Transform-Load (ETL) mechanisms and then stored in the data warehouse. The Operational Data Store (ODS) allows data to be pre-processed, checked for redundancy and that the data complies with business rules (Biswas & Sen, 2017:17).

The unstructured data is extracted through the Hadoop cluster and stored in a NoSQL database from where it is either merged with structured data in the Data Mining section or analysed in real-time and provided to the user interface (Biswas & Sen, 2017:17).

The Real-time Intelligence (RTI) module accesses the data warehouse to extract the data to feed it to either a real-time warehouse or a business intelligence system for analysis. RTI utilises complex techniques to analyse the data streams in real-time to search for trends and patterns and signals when these are found. Non-real-time data from the RTI may be fed to a Dimensional Data Store (DDS) to simplify the data for analysis converting the data to a dimensional paradigm rather than a transactional paradigm. Outputs from the DDS and RTI are fed to the Data Mining section that searches for trends in the data and provides these insights to the user interface. From there the insights or information can be represented through virtualisation, for example, graphs and charts, or can be provided directly to the Decision Support System (Biswas & Sen, 2017:20).

The architecture provides all the key elements to enables SCA at organisational level although certain practical challenges will need to be addressed to implement such an architecture (Biswas & Sen, 2017:20).

2.4.7 Business decision-making

A large-scale empirical study by Schoenherr and Speier-Pero (2015:122) found that the top-ranking benefit of using Supply Chain Analytics (in the form of predictive analytics on Big Data for Supply Chain Management) was better and more informed decision-making.

A major study was conducted by Cao *et al.* (2015) to understand the effect Business Analytics (BA) has on decision-making effectiveness and what paths existed between the two variables. The findings revealed that BA had a positive effect on information processing capabilities with a data-driven environment (or data culture) being a catalyst in this respect (Cao *et al.*, 2015:28). At the same time the information processing capabilities had a positive effect on data-driven decision-making and decision-making effectiveness (Cao *et al.*, 2015:29). What is very interesting is the fact that a positive feedback was created by the use of BA in a data-driven or data-oriented environment, and that the data culture were positively influenced by BA while BA positively influenced the nurturing of the data culture (Cao *et al.*, 2015:31).

Bumblauskas *et al.* (2017) provided insights through a conceptual model of transforming Big Data Analytics into actionable knowledge. They stated that "the ultimate objective of accumulating and analysing data is to drive decision making and action while creating value across all levels of the organization" (Bumblauskas *et al.*, 2017:7). They further found that organisations created a competitive advantage where they focused on organising and analysing their data in such a way as to facilitate better decision-making (Bumblauskas *et al.*, 2017:19).

Bag (2016:3) highlights the primary purpose and goals of Business Analytics, as a superset of Supply Chain Analytics, as:

- Driving optimal business decision-making that is actionable and in real-time;
- Assisting organisations with decision-making tools at all levels within the organisation; and
- Providing insights in support of improved decision-making.

The primary purpose and core outcome of using SCA is to derive information that can be used for business decision-making (Ittmann, 2015). Kache and Seuring (2017:10) goes one step further to state that the information derived from SCA is the "driver of corporate decision-making on strategic, tactical and operational levels".

2.5 Chapter Conclusion

In this chapter the literate study was conducted. It explored Big Data as a primary input to Big Data Analytics, and Big Data Analytics as the foundation of Supply Chain Analytics.

For each of these elements the applications, challenges and opportunities were explored. The key driver and primary outcome of using Supply Chain Analytics – business decision-making – was also discussed by looking at previous studies and the findings of those.

It is evident, and has been extensively noted (Cao *et al.*, 2015:32), that limited studies exist in the area of Supply Chain Analytics and specifically its impact on business decision-making. There is general agreement that there is a positive influence on the use of SCA on business decision-making, decision-making effectiveness and decision-making quality (Bumblauskas *et al.*, 2017:20). More research, debate and development are required in this area to better understand and manage the antecedents to effective business decision-making.

CHAPTER 3: EMPIRICAL STUDY

The chapter details the research design of this study and then provides the detailed results originating from the data that was gathered and processed. The research design section covers the population, sampling techniques and data collection instruments applied. The statistical methods used to arrive at the results as well as all the results are then discussed in detail.

3.1 Research Design

The study investigated the effect Supply Chain Analytics has on business decision-making and was aimed at organisations that gather data, both from within the organisation as well as external to the organisation, for analysis for purposes of making decisions based on the data at operational and tactical levels.

A quantitative approach was followed to gather the information. A self-completion questionnaire was developed and sent to prospective respondents in various industries (Bryman & Bell, 2017:191). The questionnaire responses were categorised and analysed to find trends and patterns to determine if it underlines the argument of the problem statement and answers to the objectives.

3.1.1 The Population of the Study

The population of a study is the complete set of objects or individuals that forms the primary focus of a research study (Banerjee & Chaudhury, 2010:61). This research study was cross-sectional in nature, and the questionnaire was completed over six weeks during 2019.

The population size was limited as a result of various constraints, primarily by limited time and financial means. The study targeted information officers and professionals at senior level that are involved with and knowledgeable on Supply Chain Analytics and data-driven business decision-making. The individuals in the population were either in a decision-making position or provided others with information purposed for decision-making.

Given the above constraints, a total of 162 individuals were targeted to form the study population. The population included C-suite members, senior management, data scientists, consultants and senior IT staff.

3.1.2 Sampling Techniques

Convenience sampling was used as a means to gather information from respondents in various companies that are involved with, and therefore knowledgeable on the topic of SCA and its effect on decision-making (Bryman & Bell, 2017:178). The questionnaire was based on the literature study.

Inferential statistics in the form of multiple regression were utilised to focus on causal relationships between the variables (Bryman & Bell, 2017:322). Only demographic information that informs about the respondent's knowledge and experience on the topic at hand was employed.

3.1.3 Data Collection Instruments

A self-completion questionnaire was developed to collect data from respondents.

The questionnaire was created using Microsoft Office 365 Forms. This tool provides an online portal where questionnaires can be created. Once the questionnaire was developed, a link to the questionnaire was created and emailed to all respondents. The email included a summary and background to the intent of the research, the informed consent and anonymity statements.

The advantage of using Microsoft Office 365 Forms was that it provided a relatively simple configuration of the questionnaire without having to be concerned with the questionnaire aesthetics. Also, all responses were automatically collated in a downloadable Microsoft Excel document for rapid data analysis. This functionality ensured data integrity and reliability. Besides, the questionnaire was accessible by any respondent with internet access.

The questionnaire is provided in Annexure A: Questionnaire.

3.2 Data Analysis

This section details the data analysis of the instruments that were used. The analysis methods are discussed together with assumptions that were made. The research ethics involved in the study will also briefly be discussed.

3.2.1 Questionnaire

The questionnaire was designed to determine whether specific relationships exist between the constructs and how strong those relationships were.

The demographic questions were confined to those that may impact the conclusion of the research. The first three questions in the questionnaire related to business demographics, while questions four to six related to the demographics of the respondent.

Questions 7 to 14 were questions based on either a Likert scale. None of the Likert scale questions was reverse coded. A 5-point Likert scale was employed for these questions (with coding as follows: 1 – Strongly disagree, 2 – Disagree, 3 – Neutral, 4 – Agree, 5 – Strongly agree).

Questions 15 to 20 were questions that measured a number of different elements using multi-choice and dichotomous questions to elicit more information building on the primary constructs that were tested. Questions 21 to 25 were open-ended questions.

All respondents answered all questions because the questionnaire was set up to disallow respondents to skip questions, except for the last question. The last question was provided for general comments regarding the topic and was not completed by all respondents. Refer to section Annexure A: Questionnaire for the complete questionnaire.

The perceptions of the respondents were tested for the following primary constructs:

Data Culture: A total of four items (as part of Question 7) were used to measure the data culture in the respondents' organisations. A 5-point Likert scale was employed. Questions such as the following were included: '*The organisation values and benefits from data'* and '*The organisation treats data as an asset'*.

Data Analytics: A total of seven items (as part of Question 8) were used to measure the data analytics culture in the respondents' organisations. A 5-point Likert scale was employed. Questions such as the following were included: '*Data Analytics is part of the organisation's strategy'* and '*Data Analytics adds to the organisation's competitive advantage'*.

Supply Chain Analytics: A total of nine items (as part of Question 10) were used to measure the current usage of Data Analytics for the supply chain in the respondents' organisations. A 5-point Likert scale was employed. Options such as the following were included: '*Demand planning'* and '*Supply chain optimisation'*.

Data Analytics Anticipated Benefits: A total of seven items (as part of Question 11) were used to measure the anticipated future benefits derived from the use of Data Analytics in the respondents' organisations. A 5-point Likert scale was employed. Options such as the following were included: '*Improving working capital'* and '*Reducing costs'*.

Data Analytics Current Benefits: A total of seven items (as part of Question 12) were used to measure the current benefits derived from the use of Data Analytics in the respondents' organisations. A 5-point Likert scale was employed. Options such as the following were included: '*Improved working capital'* and '*Reduced costs'*.

Decision-making: A total of four items (as part of Question 13) were used to measure the decision-making culture in the respondents' organisations. A 5-point Likert scale was employed. Questions such as the following were included: '*Decisions are based on intuition and experience'* and '*Data-driven decision-making is part of the organisation's culture'*.

Decision-making Quality: A total of four items (as part of Question 14) were used to measure the quality of decision-making in the respondents' organisations. A 5-point Likert scale was employed. Questions such as the following were included: '*Data-driven decisions are of good quality'* and '*Data-driven decisions are of better quality than decisions based on intuition and experience'*.

Data sources: A total of 12 items (as part of Question 15) were used to measure the data sources that are currently collected or will be collected in the next 3 years in the respondents' organisations. A 4-point Likert scale was employed (with coding as follows: 1 – Currently Collect, 2 – Collect in next 3 years, 3 – No plans to collect, 4 – Don't know or not applicable). Options such as the following were included: '*Business Activity Data'*, *'Social Media'* and '*Point of sale'*.

Value-adding data sources: For this construct (as part of Question 16) respondents could choose up to three types of data (from a total of 12 dichotomous questions) that they felt add the most value to the respondents' organisations.

3.2.2 Assumptions

The following assumptions were made for the study:

- All respondents had a good understanding of the main topics covered by the study.
- All respondents completed the questionnaire honestly and to the best of their knowledge at the time.

3.2.3 Research Ethics

The research study was preceded by ethical clearance being granted to perform the research (see ethical clearance NWU-00723-19-A4). The clearance implied that the research would be guided by sound ethical principles, among other things, obtaining informed consent and respecting respondent anonymity.

The questionnaire gathered no demographic information on respondents that could identify them or could harm or discriminate against them in any manner. Also, no organisation was named or inferred in the questionnaire and research study.

3.3 Results

A total of 74 responses from the questionnaire were received. The questionnaire responses for Question 1-20 were subjected to frequency statistics (see section 3.3.1). Factor analysis was further performed on Questions 7-15 to determine the validity and reliability of the instrument (see section 3.3.2).

Furthermore, the main constructs were subjected to regressions (see section 3.3.3) while the demographic questions were subjected to analysis of variance (see section 3.3.4).

3.3.1 Frequency Statistics

3.3.1.1 Demographics

The demographics section aimed to determine the general level of experience of both the individual respondents and the organisations as well as the industries in which they operate.

Question 1 and 3 related to how long the organisations in the sample had been in business and the size of the organisation in terms of the number of employees. The results are illustrated in Figure 3–1, and Figure 3–3. Figure 3–1 highlights the fact that most of the organisations (89%) are mature, having been in business for more than ten years. Only a handful of younger organisations formed part of the sample.

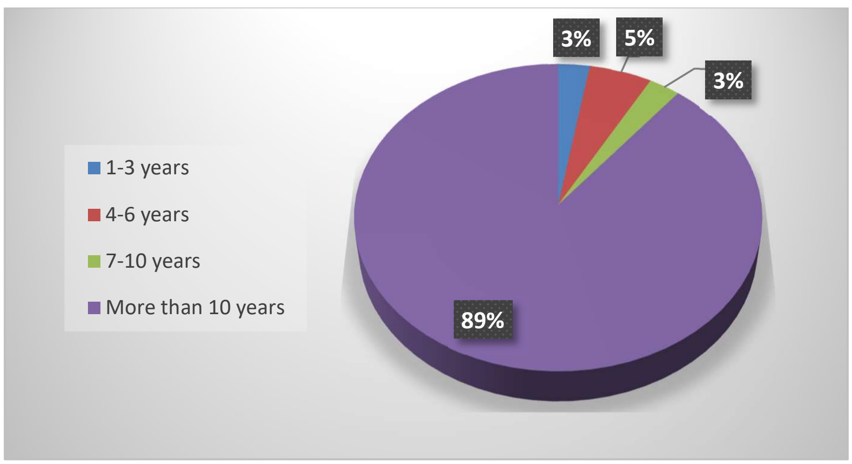

Figure 3–1: Years the organisations have been in business (Q1)

Question 2 identified the industries of respondents' organisations. The objective of the question was to gather input from a wide range of industries. If a single industry was to be used, it might have skewed the results, making it more difficult to infer the outcomes to other industries. Figure 3–2 illustrates the results. Almost 34% of the respondents came from the financial sector. Other primary sectors in the sample included the Telecommunications (14%), Professional Services (11%) and Energy (8%) sectors.

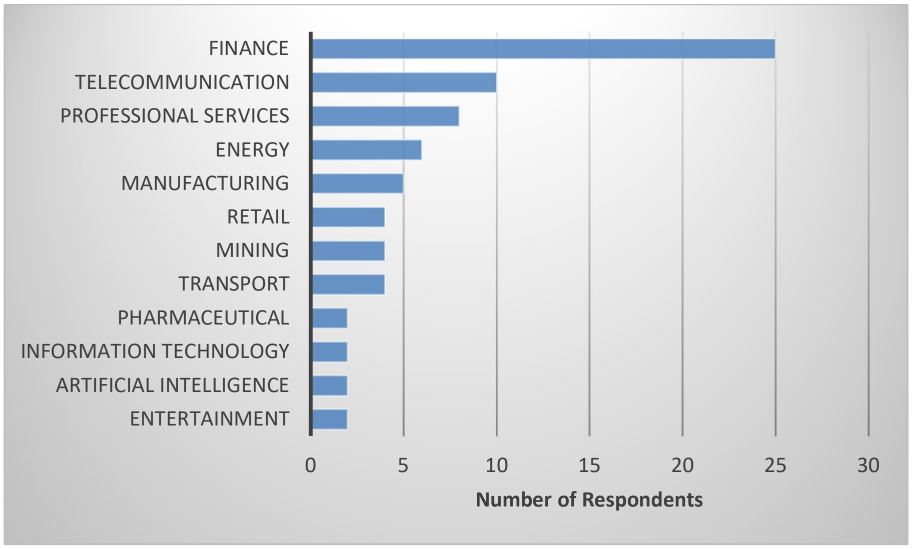

Figure 3–2: The industries of the respondents' organisations (Q2)

Figure 3–3 shows that the vast majority (81%) of the organisations in the sample had more than 500 employees, followed by very small companies with less than 50 employees (11%). The number of respondents falling in medium-sized organisations was fairly insignificant (8%).

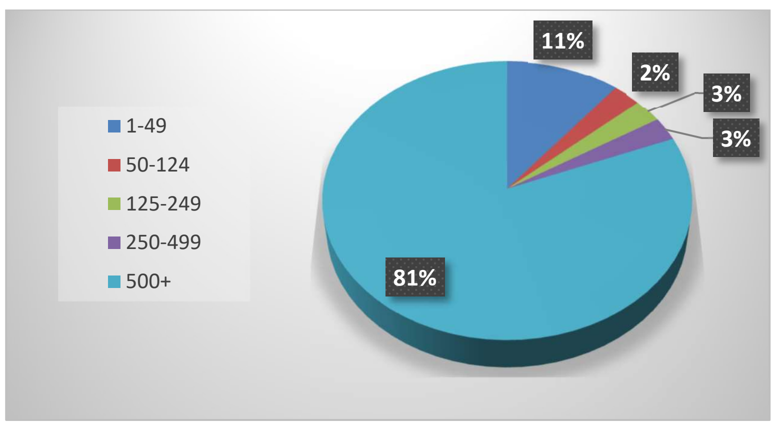

Figure 3–3: Number of employees in the organisation (Q3)

Figure 3–4 shows that a large majority of the respondents were executive heads (30%) or senior management (16%). However, the rest of the respondents had a very wide variety of job titles and were from many different areas in the organisation.

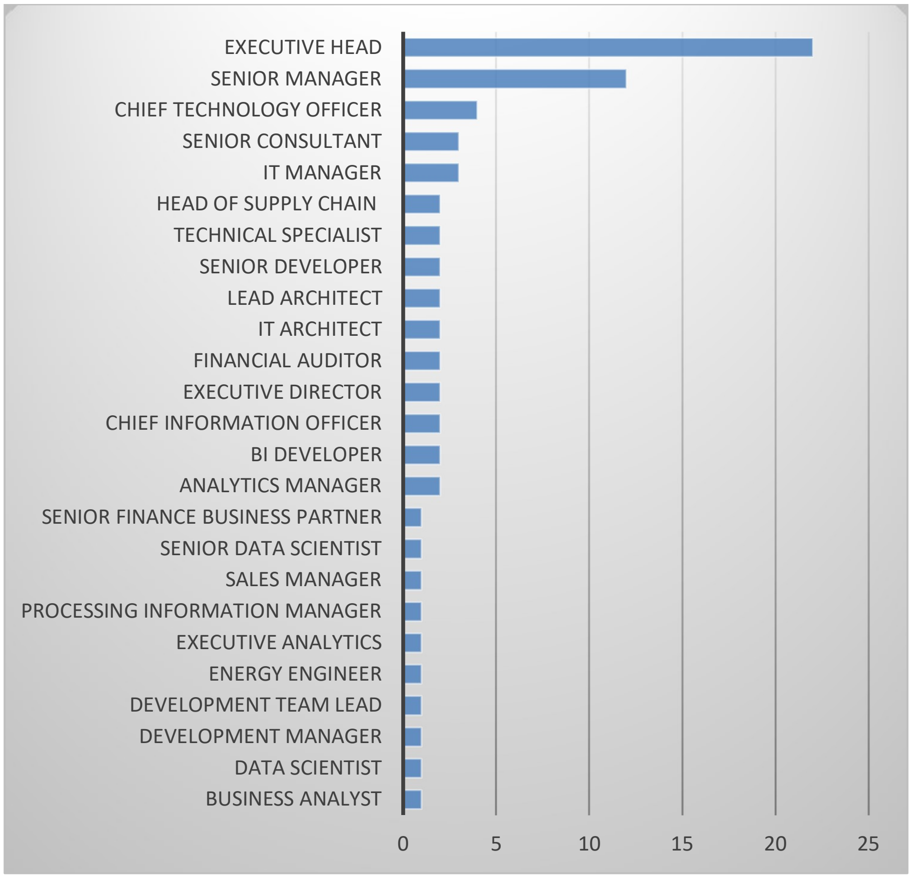

Figure 3–4: Job titles of respondents (Q4)

Question 5 aimed to determine the level of the respondent's experience in working with data in general. The results are shown in Figure 3–5. Around 59% of the respondents had more than ten years' experience working with data, while 24% had more than seven years' experience.

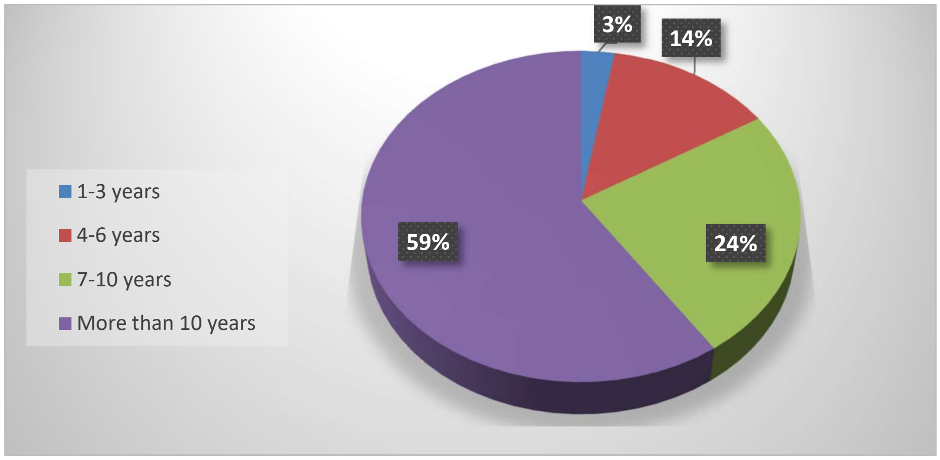

Figure 3–5: Respondent experience with data in general (Q5)

Question 6 aimed to determine the level of the respondent's experience with Big Data. The results are shown in Figure 3–6. More than half of the respondents (57%) had practical application experience with Big Data. None of the respondents had no experience with Big Data, although 20% of respondents had limited knowledge of Big Data.

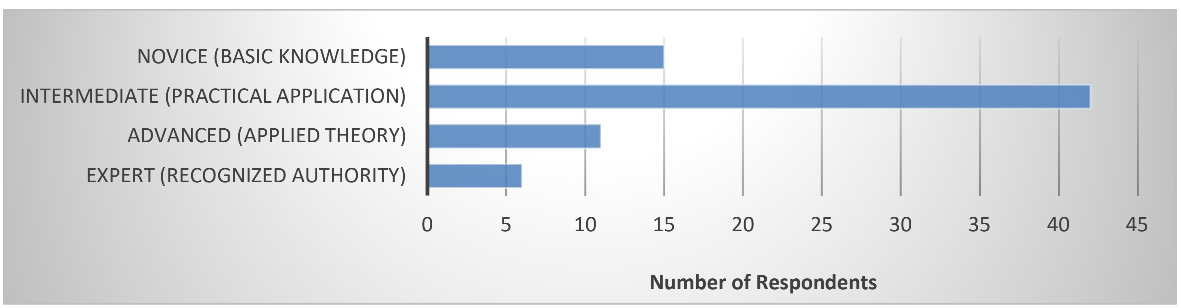

Figure 3–6: Respondent experience with Big Data (Q6)

3.3.1.2 Information culture

The information culture section aimed to determine the importance of data to the organisation as well whether an information culture existed within the organisation. Question 7 was split into four items, each looking at a different aspect of data importance and information culture. A Likert scale was used to elicit the responses. The results are illustrated in Figure 3–7. None of the respondents marked Strongly Disagree for any of the four items, and therefore, it was omitted from the chart in Figure 3–7.

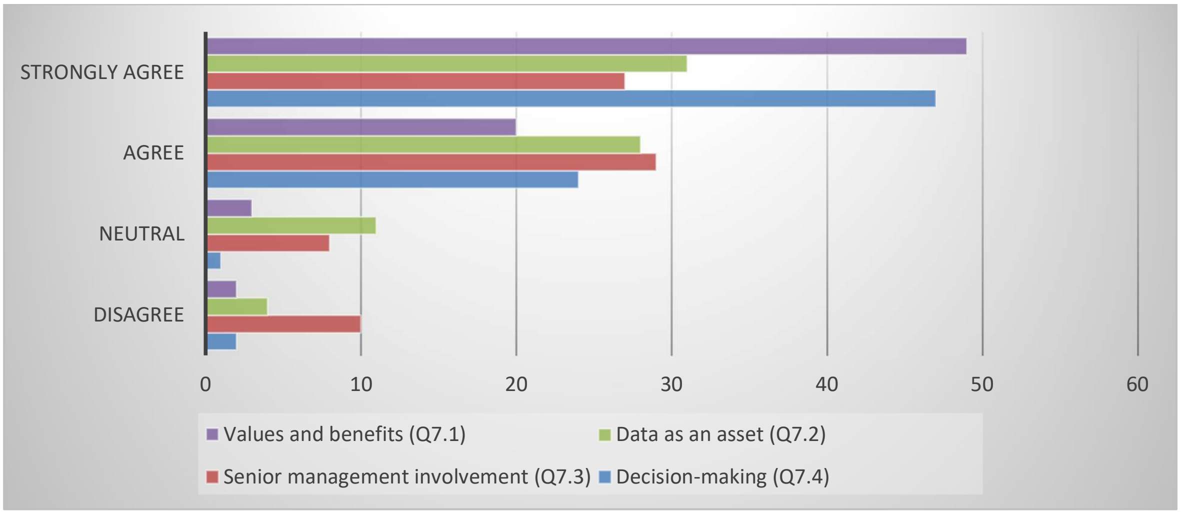

Figure 3–7: Data importance and information culture in the organisation (Q7)

Question 7.1 related to how the organisation values data and benefits from the use of data. More than 93% of respondents felt that their organisation was positive about the value of data and that they gained benefits in some manner from the use of data.

Question 7.2 aimed to determine whether the organisations viewed data as an asset. According to 80% of respondents, their organisations acknowledged the importance of data by regarding data as a valuable asset.

The level of buy-in by senior management in data projects was determined by Question 7.3. The responses revealed that 76% of respondents experienced that senior management was indeed involved in data-related projects.

As for the importance of using data in decision-making, the results of Question 7.4 show that 96% of respondents remained positive that data was a crucial driver for decisions and decision-making in their organisations.

3.3.1.3 Data Analytics

The objective of the Data Analytics section was four-fold:

- To determine whether a Data Analytics culture exists in the organisation (Question 8 and 9);
- To determine how Data Analytics are currently utilised for supply chain activities (Question 10);

- To determine what benefits have already been reaped from using Data Analytics in supply chain activities (Question 11) and
- To determine anticipated future benefits from using Data Analytics in supply chain activities (Question 12)

Question 8 was split into seven items, each looking at a different aspect of the Data Analytics culture within the organisation. A Likert scale was used to extract the responses. The results are illustrated in Figure 3–8 (part 1) and Figure 3–9 (part 2). In Figure 3–8, none of the respondents marked Strongly Disagree for any of the four items, and therefore, it was left off the chart.

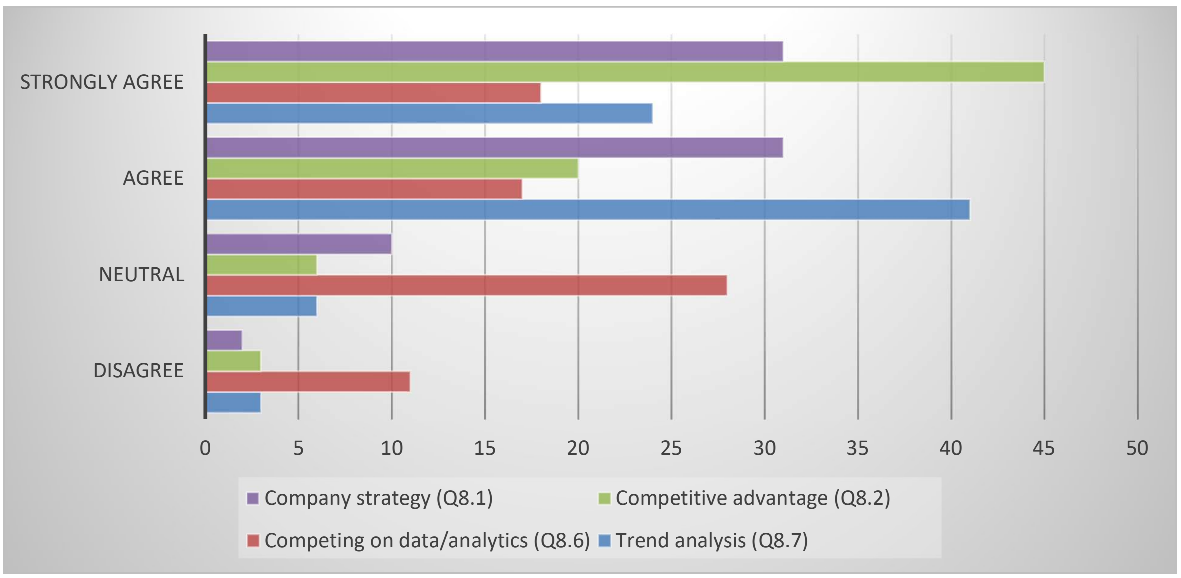

Figure 3–8: Data Analytics (Q8 Part 1)

Question 8.1 asks the question as to whether Data Analytics forms part of the organisation’s business strategy. More than 83% of respondents indicated that their organisations had a business strategy that speaks to Data Analytics. In line with this, more than 87% of respondents felt that Data Analytics extended the competitive advantage of their organisations (Question 8.2).

In contrast to this, less than half of the respondents (47%) stated that their organisations were competing based on data and analytics in general (Question 8.6). A significant portion of respondents (38%) seemed to be indecisive on their organisation's position on the matter.

Some 88% of respondents stated that their organisations utilised Data Analytics for trend analysis (Question 8.7).

Figure 3–9 portrays the elements required to make Data Analytics happen in an organisation.

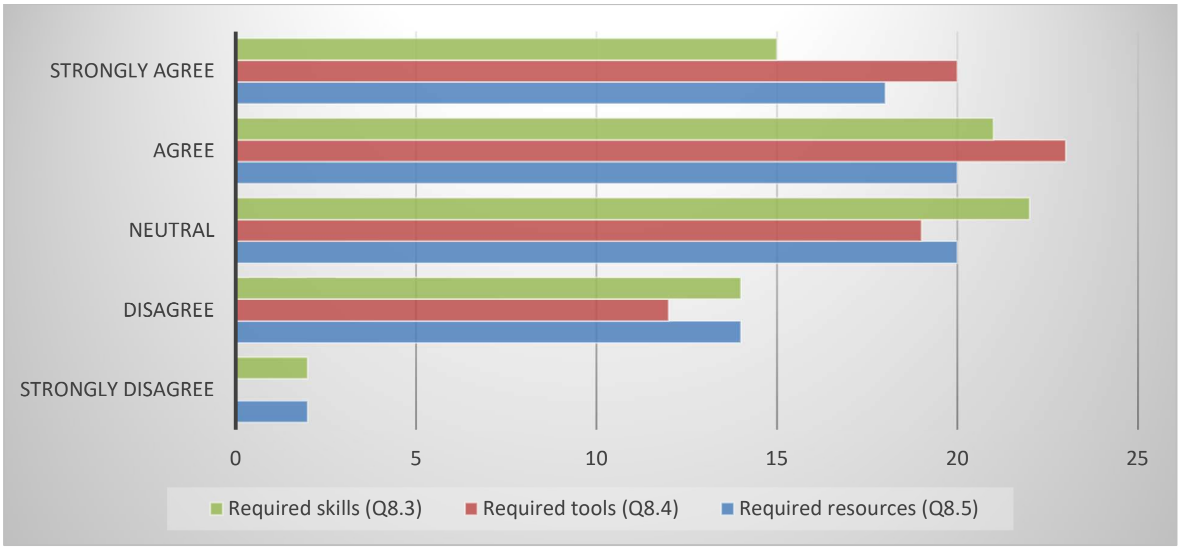

Figure 3–9: Data Analytics (Q8 Part 2)

Question 8.3 to 8.5 demanded answers to whether the organisation had the necessary skills (8.3), tools (8.4) and resources (8.5) to handle Data Analytics effectively. Only 49% of respondents felt their organisations had the required skills to handle Data Analytics, even though 58% stated they had the required tools for Data Analytics. Moreover, only 51% of respondents stated that they had the necessary resources to deal with Data Analytics. A large portion of respondents was less than positive about the prospects of addressing Data Analytics in their organisations due to a lack of resources, skills and appropriate analytical tools.

Figure 3–10 shows the responses from Question 9 that related to whether Supply Chain Analytics is used for decisions made at the supply chain level. Even though only 51% of respondents used SCA for decision-making within their supply chains, it was conclusive since many respondents (38%) seemed uncertain about this, and 11% even disagreed with this.

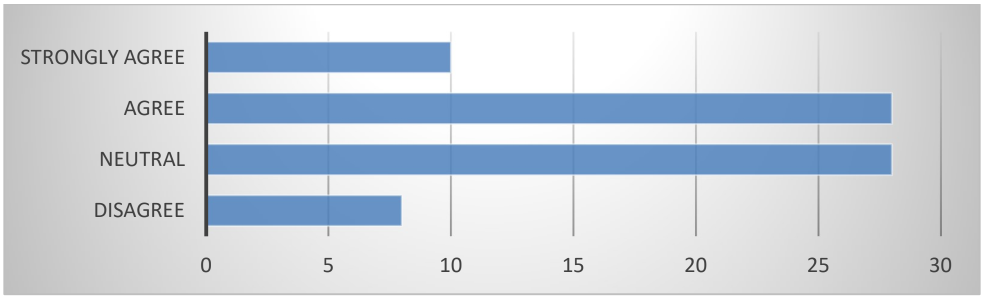

Figure 3–10: Supply Chain decisions(Q9)

Question 10 determined whether Data Analytics is used for supply chain activities given nine different activities, as shown in Figure 3–11. The numbers for 'Agree' and 'Strongly Agree' have been added together as one ('Agree') to simplify Figure 3–11. The same was done for 'Disagree'.

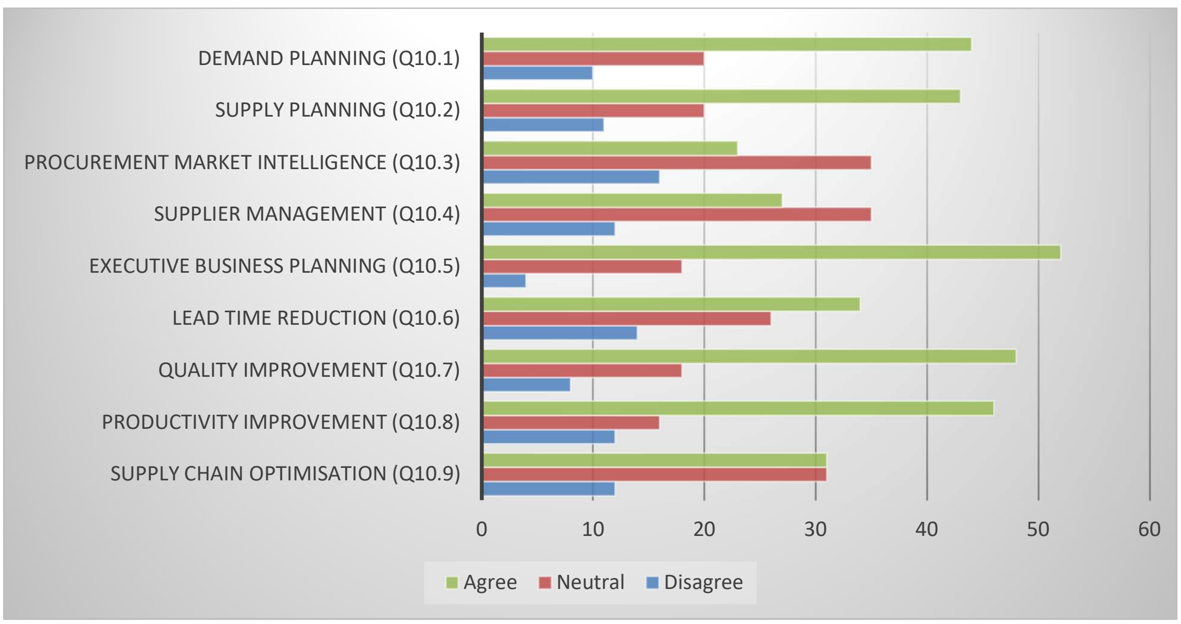

Figure 3–11: Supply Chain Analytics (Q10)

The most prominent use of Data Analytics appeared to be in Executive business planning (70%), followed by Quality Improvement (65%), Productivity Improvement (62%), Demand Planning (59%) and Supply Planning (58%). The Neutral responses under Procurement Market Intelligence (47%) and Supplier Management (47%) possibly indicated that these respondents' organisations did not use Data Analytics for supply chain activities in these areas.

Question 11 and 12 purported to understand the current benefits that have been gained (Question 12) from the use of Data Analytics in supply chain activities versus the anticipated future benefits (Question 11) that it hoped to reap.

Figure 3–12 portrays the current benefits of Data Analytics in supply chain activities. The numbers for ‘Agree’ and ‘Strongly Agree’ have been added together as one (‘Agree’) to simplify Figure 3–12. The same was done for ‘Disagree’.

The major current benefits were improved service (72%), followed by reduced costs (64%). According to the majority of respondents, clear benefits were being reaped from the use of Data Analytics in supply chain activities.

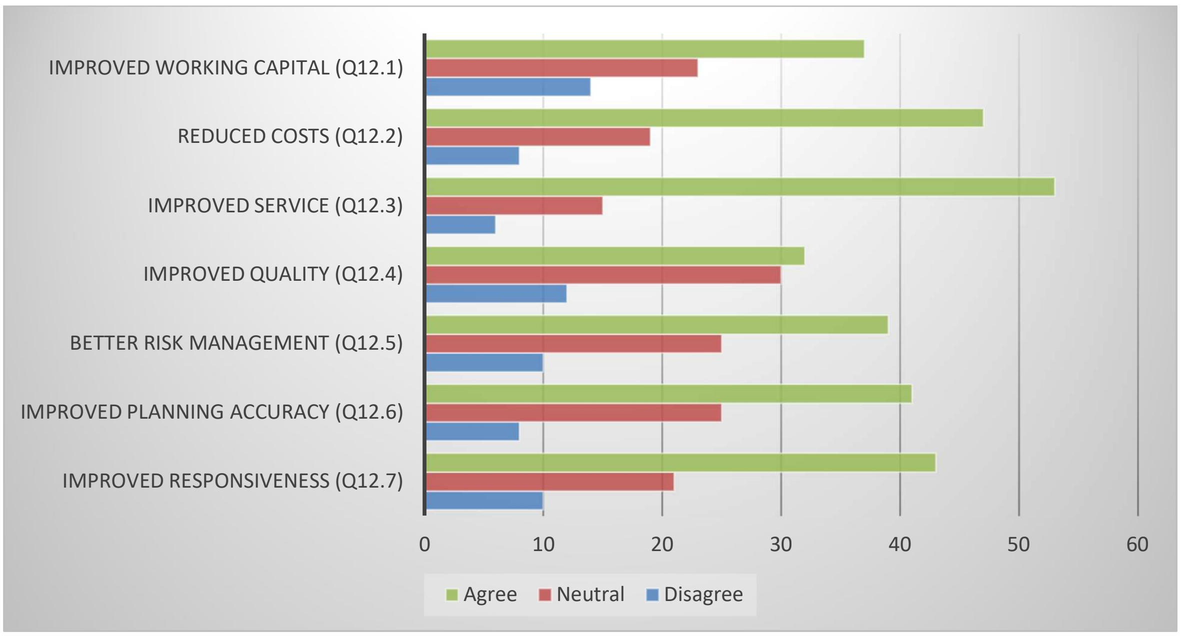

Figure 3–12: Data Analytics current benefits in supply chain activities (Q12)

Figure 3–13 portrays the anticipated benefits of Data Analytics in supply chain activities. The major anticipated benefits of using Data Analytics were seen to be the reduction of costs (82%) followed closely by service improvement (80%) and planning improvements (78%). The vast majority of respondents considered the value of Data Analytics in all areas of the supply chain.

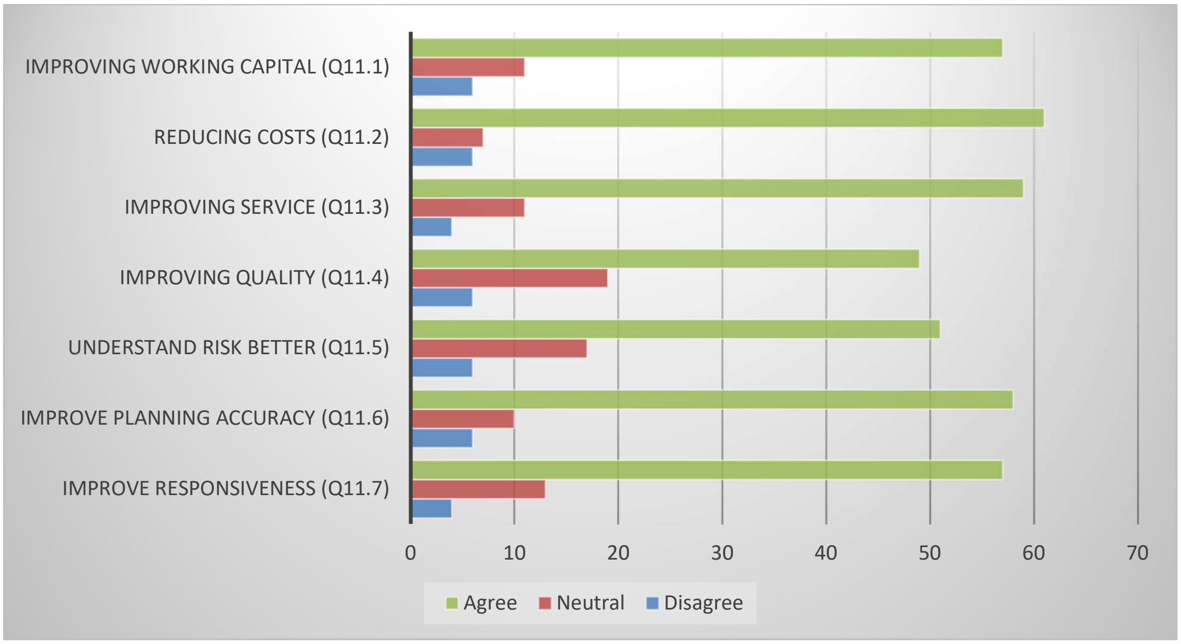

Figure 3–13: Data Analytics anticipated benefits in supply chain activities (Q11)

When comparing Figure 3–13 to Figure 3–12, it was noteworthy that there was a shift from the current benefits to the anticipated future benefits that resulted in an overall move towards seeing more value in using Data Analytics for supply chain activities.

3.3.1.4 Decision-making

The objective of the decisions section was two-fold:

- To determine the basis of decisions (Question 13); and
- To determine the quality of decisions (Question 14);

Question 13 was split into four items, each looking at a different aspect of decision-making within the organisation. The results are illustrated in Figure 3–14.

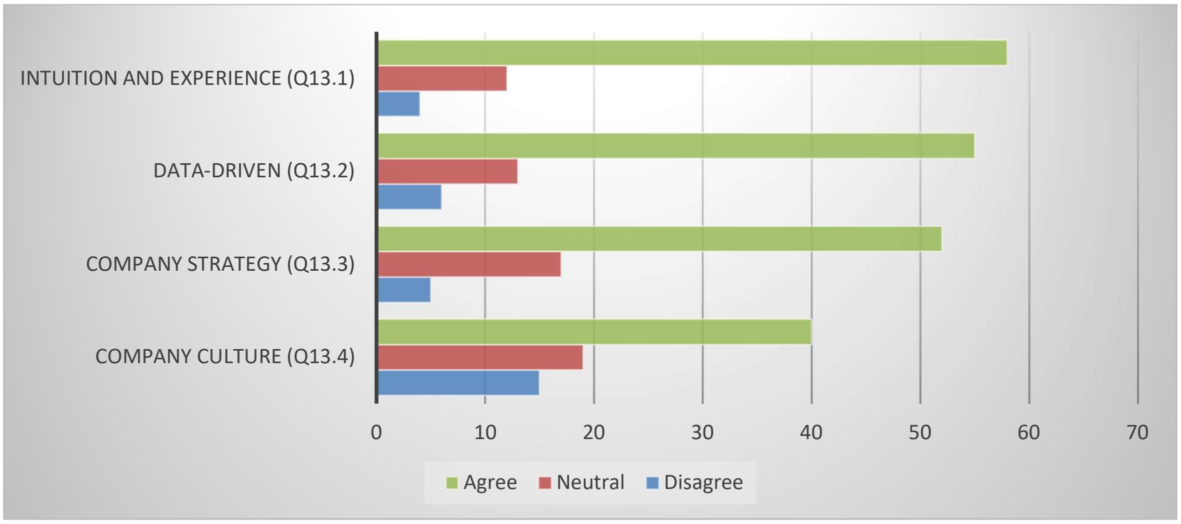

Figure 3–14: Basis for decision-making (Q13)

Question 13.1 and 13.2 asked the question as to whether decisions are based on intuition and experience (Question 13.1) or are data-driven (Question 13.2). More than 78% of respondents agreed that their organisations make decisions based on intuition and experience as opposed to 74% that agreed that decisions are data-driven.

Question 13.3 and 13.4 aimed to determine whether the organisations view data-driven decisions as being part of the organisational strategy (Question 13.3) and organisational culture (Question 13.3). According to 70% of respondents, decision-making is part of their organisational strategy, whereas only 54% felt that decision-making is part of their organisational culture.

Question 14 has four items, each looking at a different aspect of the quality of decision-making within the organisation. The results are illustrated in Figure 3–15.

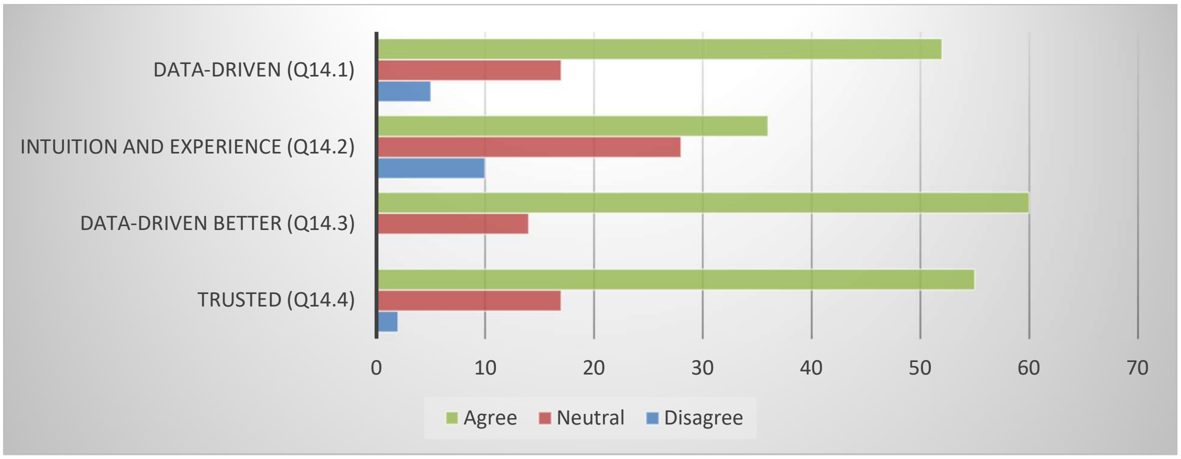

Figure 3–15: Quality of decision-making (Q14)

Question 14.1 and 14.2 aimed to determine the quality of decisions when it is data-driven (Question 14.1) and when it is based on intuition and experience (Question 14.2). More than 70% of respondents agreed that decisions based on data are of good quality in their organisations, while only 49% of decisions based on intuition and experience are of good quality.

Question 14.3 attempted to determine whether data-driven decisions are of better quality than decisions driven by intuition and experience. According to 81% of respondents, data-driven decision-making is of better quality than intuition and experience-based decisions in their organisations.

Question 14.4 further aimed to elicit whether data-driven decisions can be trusted. More than 74% of respondents indicated that they felt data-driven decisions are trustworthy.

3.3.1.5 Data Collection and Data Types

The data sources section aimed to reveal:

- The data source collection methodology being used (Question 15);
- Which data sources add the most value to the organisation (Question 16); and
- How much data is actually being processed (Question 17).

Question 15 determined which data sources were currently being collected and which would be collected in the next three years. It also indicated whether organisations had no plans to collect a particular type of source data. The results are illustrated in Figure 3–16.

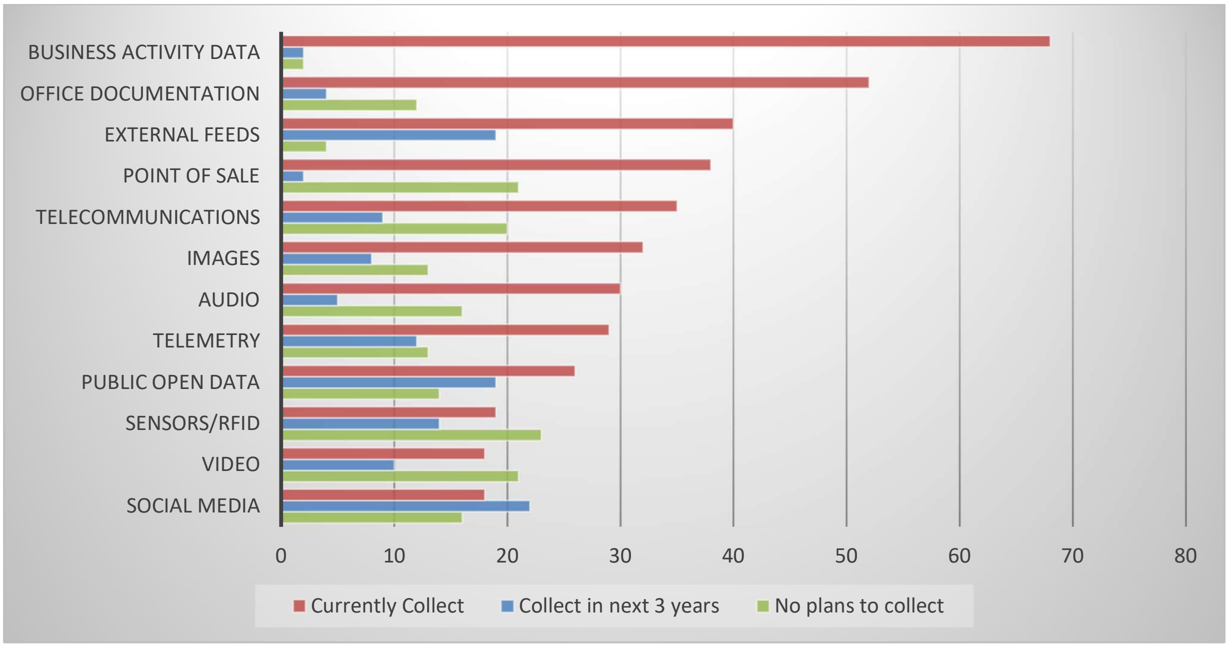

Figure 3–16: Data Source collection (Q15)

Figure 3–16 clearly illustrates that Business Activity Data is the most common data source currently collected, as indicated by 92% of respondents. This was followed by Office Documentation (70%), External Feeds (54%) and Point of Sale data (51%). Respondents further indicated that in the next few years, Social Media would be the most sought after (30%) followed by moves to collect Public Open Data and External Feeds (both at 26%). It would also seem that some companies do not see the need to gather certain data sources. These data sources are primarily Sensor/RFID data (31%), Video (28%) and Point of Sale data (28%).

Question 16 focussed on the importance of the data sources and sought to find the primary data sources of organisations as it relates to Supply Chain Analytics. Respondents could choose up to three data sources. The results are illustrated in Figure 3–17.

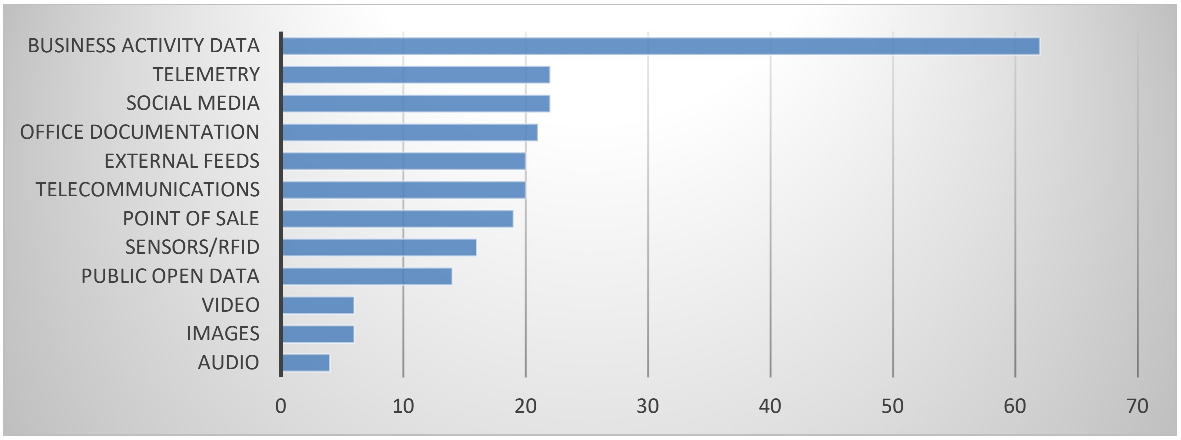

Figure 3–17: Data Source importance (Q16)

Respondents indicated that Business Activity Data was the most valued data source according to 84% of respondents. However, many other data sources were deemed similarly valuable according to about 25% of respondents. This may be as a result of different needs by different industries.

Question 17 aimed to determine how much of the total analysable data in an organisation was currently being analysed. Also, how much of the total analysable data in the organisation came from external sources. Besides, it then asked the question of how much it would like to analyse in three years. The results are illustrated in Figure 3–18.

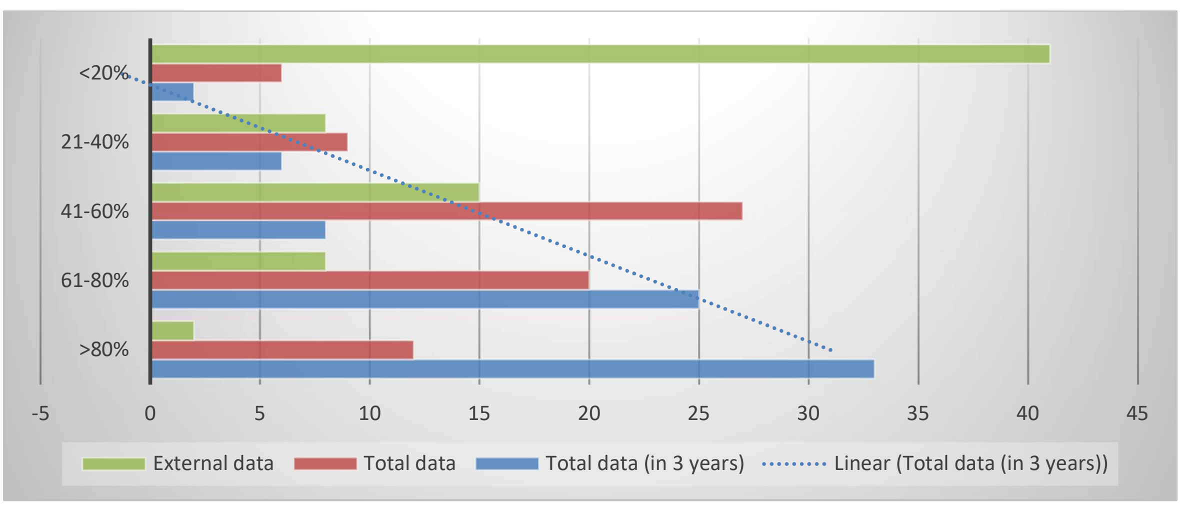

Figure 3–18: Data Collection (Q17)

More than 55% of respondents indicated that less than 20% of all data collected are sourced externally, while around 20% of respondents acknowledged that between 41%-60% of their data comes from external sources.

It is promising that 45% of respondents felt that their organisation will be increasing the total data analysed to more than 80% over the next three years. The trendline for the anticipated total data analysed in three years clearly shows a steep growth trend. This could be indicative that organisations are ramping up to accumulate more information by analysing more of their total data.

3.3.1.6 Decision-making Support

The decision-making support section aimed to understand:

- What would best support data-driven decision-making abilities (Question 18);
- The amount of data available for decision-making (Question 19); and
- The primary challenges presented when making data-driven decisions (Question 20).

Question 18 determined the key aspects that would support data-driven decision-making in the best possible way. The results are illustrated in Figure 3–19.

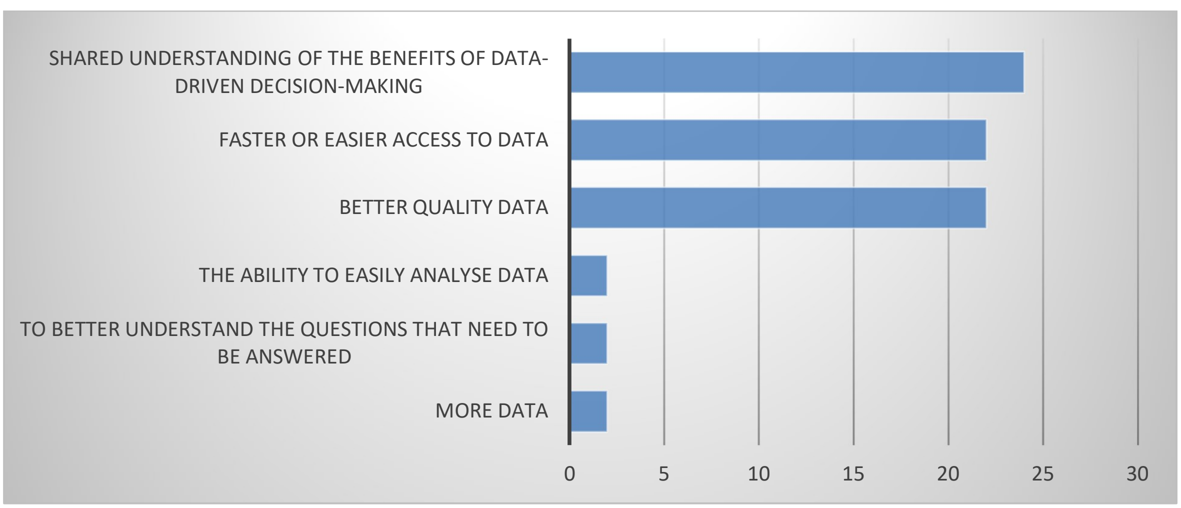

Figure 3–19: Decision-making support aspects (Q18)

In Figure 3–19 respondents have highlighted three key aspects that would support data-driven decision-making. Having a shared understanding of the benefits of data-driven decision-making was critical to 32% of respondents. The rest of the votes were more or less evenly split between faster or better access to data, and to have better quality data available (30% each).

Question 19 sought to determine the perception of the amount of data available in support of decision-making. The results are illustrated in Figure 3–20.

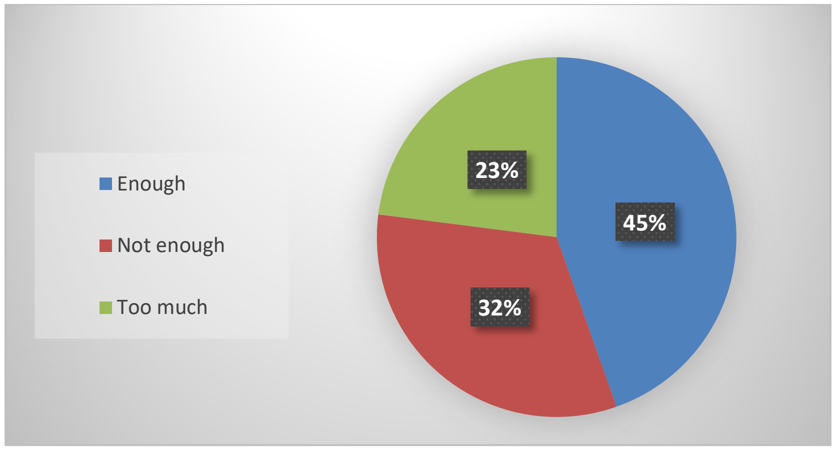

Figure 3–20: Data available for Decision-making (Q19)

From Figure 3–20 it can be seen that most respondents (45%) felt that enough data is available. However, 32% of respondents believed that there is not enough data supporting decision-making.

Question 20 aimed to reveal the most important challenges of data-driven decision-making. Respondents could choose up to three data sources. The results are illustrated in Figure 3–21.

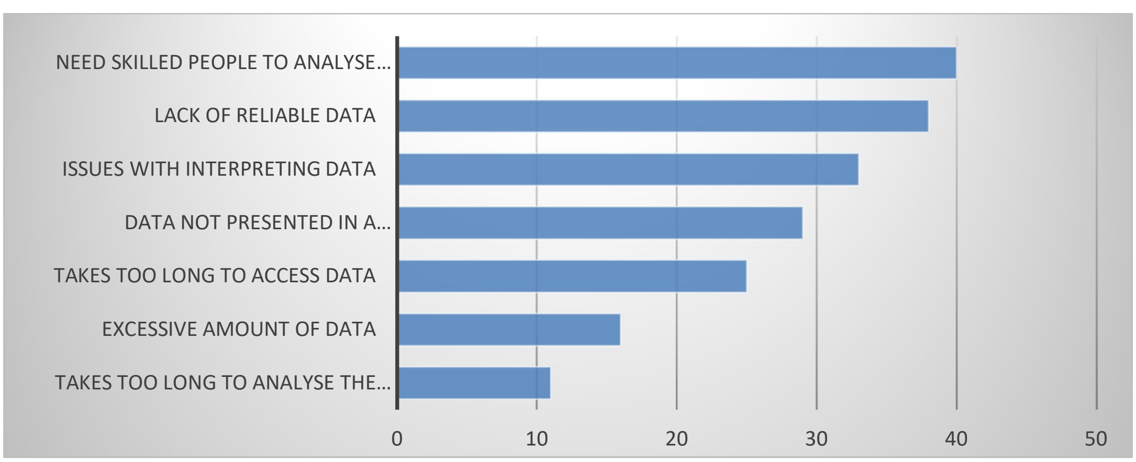

Figure 3–21: Significant challenges of data-driven decision-making (Q20)

By a small margin, the need for skilled people to assist in analysing data that is purposed for supporting decisions was the most important challenge, according to 54% of respondents. This was followed by the challenge of potentially unreliable data (51%), difficulty in interpreting data (45%) and being able to present the data (information) purposefully (39%).

3.3.1.7 Summary

This section covered the frequency statistics of the questionnaire in detail and highlighted the main constructs that were tested. The following section will indicate how factor analysis was used to determine the validity and reliability of the study.

3.3.2 Validity and Reliability

Factor analysis was used to establish validity. Initially, all questionnaire items were subjected to a principal component analysis to see whether the factors that were extracted corresponding to those included in the questionnaire. The results were inconclusive, as only one major factor was extracted, explaining 77.5% of the variance, as can be seen from Table 3-1 and confirmed by the scree plot in Figure 3-22. This was expected because the sample was possibly not large enough to enable a sufficient factor analysis and the respondents were a reasonably homogeneous group.

Table 3-1: Principal component analysis for Questions 7-15

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	42.637	77.523	77.523	42.637	77.523	77.523
2	2.331	4.238	81.761	2.331	4.238	81.761
3	1.482	2.695	84.456	1.482	2.695	84.456
4	1.125	2.046	86.502	1.125	2.046	86.502
5	0.996	1.812				

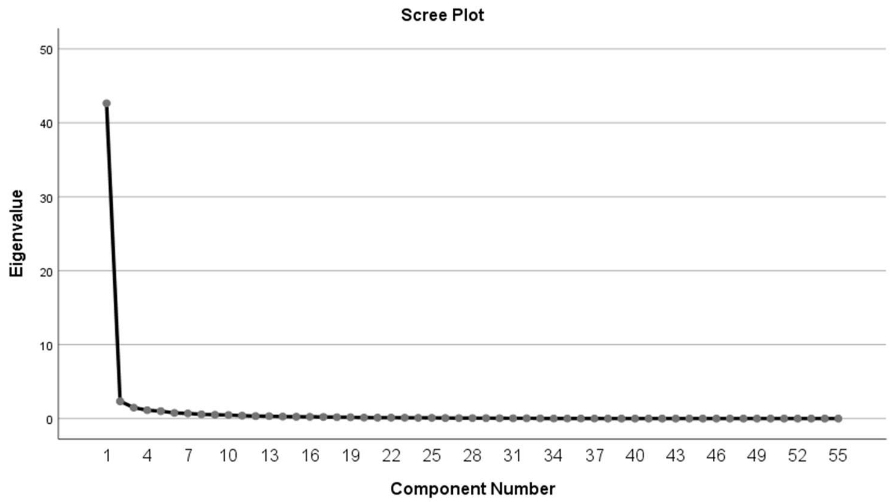

Figure 3-22: Scree plot for Questions 7-15

Consequently, it was decided to do a principal component analysis on each of the constructs included in the questionnaire to see whether each of them would yield only one factor. This was indeed the case, as reported below.

Although a principal component analysis is not strictly the same as factor analysis, the term “factor analysis” is sometimes used for principal component analysis. All analyses were carried out using a principal component analysis with direct oblimin rotation. For ease of reading, the term “factor analysis” will be used for the rest of this study.

The factor analysis was conducted using SPSS software. For each factor that was extracted, Cronbach’s alpha value was used to determine reliability. The different constructs tested in the questionnaire were included in Question 7 to Question 15. Each of the questions was therefore separately subjected to factor analysis.

The following sections detail the results of the factor analysis performed.

3.3.2.1 Construct: Data Culture

Table 3-2, Table 3-3 and Table 3-4 detail the results of the factor analysis performed for the Data Culture construct.

Table 3-2: Factor Analysis for Question 7 (Construct: Data Culture)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.852
Bartlett's Test of Sphericity	Approx. Chi-Square	378.967
	df	6
	Sig.	0.000

From Table 3-2, the KMO value of 0.852, which is above the cut-off value of 0.6 as proposed by Pallant (2013:90). Also, the fact that Bartlett's test of sphericity yielded a significance of less than 0.001 indicates that the data was suitable for factor analysis.

Table 3-3: Total Variance Explained for Question 7 (Construct: Data Culture)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	3.499	87.481	87.481	3.499	87.481	87.481
2	0.284	7.089	94.570			
3	0.137	3.432	98.003			
4	0.080	1.997	100.000			

From Table 3-2 and Table 3-3, it is evident that only one factor explains 87.4% of the variance and it could, therefore, be inferred that this factor is valid. The specific items of the factor are shown in Table 3-4.

Table 3-4: Items forming part of Question 7 (Construct: Data Culture)

Number of factors extracted	Variance on number of factors explained	Cronbach's alpha
1	87.481	0.952
Component Matrix		**Factor = 1**
Data is important for decision-making		0.956
Senior management is involved in data-related projects		0.956
The organisation treats data as an asset		0.944

The organisation values and benefits from data	0.884

Cronbach's alpha value on standardised items was calculated to establish reliability. A value of 0.952 was obtained, and it could thus be concluded that the factor is reliable.

3.3.2.2 Construct: Data Analytics

Table 3-5, Table 3-6 and Table 3-7 detail the results of the factor analysis performed for the Data Analytics construct.

Table 3-5: Factor Analysis for Question 8 (Construct: Data Analytics)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.956
Bartlett's Test of Sphericity	Approx. Chi-Square	2367.778
	df	171
	Sig.	0.000

From Table 3-5, the KMO value of 0.956 and the fact that Bartlett's test of sphericity yielded a significance of less than 0.001 indicate that the data was suitable for factor analysis.

Table 3-6: Total Variance Explained for Question 8 (Construct: Data Analytics)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	14.82	78.01	78.01	14.82	78.01	78.01
2	0.858	4.515	82.528			
3	0.566	2.978	85.506			
4	0.475	2.499	88.005			

From Table 3-5 and Table 3-6, it is evident that only one factor explains 78% of the variance and it could, therefore, be inferred that this factor is valid. The specific items of the factor are shown in Table 3-7.

Table 3-7: Items forming part of Question 8 (Construct: Data Analytics)

Number of factors extracted	Variance on number of factors explained	Cronbach's alpha
1	78.013	0.954
Component Matrix		**Factor = 1**
The organisation uses data for trend analysis		0.932
The organisation has the required resources to handle Data Analytics		0.906
Data analytics is part of the organisation's strategy		0.905
The organisation is competing on data and analytics		0.887
Data analytics adds to the organisation's competitive advantage		0.860
The organisation has the required analytical tools to handle Data Analytics		0.776
The organisation has the required skills to handle Data Analytics		0.758

Cronbach's alpha value on standardised items was calculated to establish reliability, and a value of 0.954 was obtained. It could thus be concluded that the factor is reliable.

3.3.2.3 Construct: Supply Chain Analytics

Table 3-8, Table 3-9 and Table 3-10 detail the results of the factor analysis performed for the Supply Chain Analytics construct.

Table 3-8: Factor Analysis for construct Supply Chain Analytics (Q10)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.929
Bartlett's Test of Sphericity	Approx. Chi-Square	964.908
	df	36
	Sig.	0.000

From Table 3-8, the KMO value of 0.929 and the fact that Bartlett's test of sphericity yielded a significance of less than 0.001 indicate that the data was suitable for factor analysis.

Table 3-9: Total Variance Explained for construct Supply Chain Analytics (Q10)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	7.316	81.286	81.286	7.316	81.286	81.286
2	0.463	5.145	86.430			
3	0.342	3.805	90.235			
4	0.206	2.290	92.525			

From Table 3-8 and Table 3-9, it is evident that only one factor explains 81.3% of the variance and it could, therefore, be inferred that this factor is valid. The specific items of the factor are shown in Table 3-10.

Table 3-10: Items forming part of Construct: Supply Chain Analytics (Q10)

Number of factors extracted	Variance on number of factors explained	Cronbach's alpha
1	81.286	0.971
Component Matrix		**Factor = 1**
Procurement market intelligence		0.944
Lead time reduction		0.928
Supply planning		0.923
Executive business planning		0.920
Productivity improvement		0.916
Quality improvement		0.914
Demand planning		0.879
Supply chain optimisation		0.862
Supplier management		0.821

Cronbach's alpha value on standardised items was calculated to establish reliability, and a value of 0.971 was obtained. It could thus be concluded that the factor is reliable.

3.3.2.4 Construct: Data Analytics Anticipated Benefits

Table 3-11, Table 3-12 and Table 3-13 detail the results of the factor analysis performed for construct Data Analytics Anticipated Benefits.

Table 3-11: Factor Analysis for construct Data Analytics Anticipated Benefits (Q11)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.890
Bartlett's Test of Sphericity	Approx. Chi-Square	592.602
	df	21
	Sig.	0.000

From Table 3-11, the KMO value of 0.890 and the fact that Bartlett's test of sphericity yielded a significance of less than 0.001 indicate that the data was suitable for factor analysis.

Table 3-12: Total Variance Explained for construct Data Analytics Anticipated Benefits (Q11)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	5.361	76.585	76.585	5.361	76.585	76.585
2	0.553	7.893	84.478			
3	0.400	5.709	90.187			
4	0.278	3.976	94.162			

From Table 3-11 and Table 3-12, it is evident that only one factor explains 76.6% of the variance and it could, therefore, be inferred that this factor is valid. The specific items of the factor are shown in Table 3-13.

Table 3-13: Items forming part of construct Data Analytics Anticipated Benefits (Q11)

Number of factors extracted	Variance on number of factors explained	Cronbach's alpha
1	76.585	0.948
Component Matrix		**Factor = 1**
Understand risk better		0.927
Reducing costs		0.926
Improving working capital		0.909
Improving service		0.889
Improving quality		0.869
Improve responsiveness		0.816
Improve planning accuracy		0.779

Cronbach's alpha value on standardised items was calculated to establish reliability, and a value of 0.948 was obtained. It could thus be concluded that the factor is reliable.

3.3.2.5 Construct: Data Analytics Current Benefits

Table 3-14, Table 3-15 and Table 3-16 detail the results of the factor analysis performed for the Data Analytics Current Benefits construct.

Table 3-14: Factor Analysis for construct Data Analytics Current Benefits (Q12)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.933
Bartlett's Test of Sphericity	Approx. Chi-Square	831.488
	df	21
	Sig.	0.000

From Table 3-14, the KMO value of 0.933 and the fact that Bartlett's test of sphericity yielded a significance of less than 0.001 indicate that the data was suitable for factor analysis.

Table 3-15: Total Variance Explained for construct Data Analytics Current Benefits (Q12)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	6.003	85.760	85.760	6.003	85.760	85.760
2	0.385	5.494	91.254			
3	0.188	2.687	93.941			
4	0.162	2.318	96.259			

From Table 3-14 and Table 3-15, it is evident that only one factor explains 85.8% of the variance and it could, therefore, be inferred that this factor is valid. The specific items of the factor are shown in Table 3-16.

Table 3-16: Items forming part of construct Data Analytics Current Benefits (Q12)

Number of factors extracted	Variance on number of factors explained	Cronbach's alpha
1	85.760	0.972
Component Matrix		**Factor = 1**
Improved working capital		0.886
Reduced costs		0.922
Improved service		0.959
Improved quality		0.872
Better risk management		0.943
Improved planning accuracy		0.942
Improved responsiveness		0.956

Cronbach's alpha value on standardised items was calculated to establish reliability, and a value of 0.972 was obtained. It could thus be concluded that the factor is reliable.

3.3.2.6 Construct: Decision-making

Table 3-17, Table 3-18 and Table 3-19 detail the results of the factor analysis performed for the Decisions construct.

Table 3-17: Factor Analysis for construct Decision-making (Q13)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.811
Bartlett's Test of Sphericity	Approx. Chi-Square	211.287
	df	6
	Sig.	0.000

From Table 3-17, the KMO value of 0.811 and the fact that Bartlett's test of sphericity yielded a significance of less than 0.001 indicate that the data was suitable for factor analysis.

Table 3-18: Total Variance Explained for construct Decision-making (Q13)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	3.035	75.874	75.874	3.035	75.874	75.874
2	0.431	10.777	86.651			
3	0.366	9.149	95.800			
4	0.168	4.200	100.000			

From Table 3-17 and Table 3-18, it is evident that only one factor explains 75.9% of the variance and it could, therefore, be inferred that this factor is valid. The specific items of the factor are shown in Table 3-19.

Table 3-19: Items forming part of construct Decision-making (Q13)

Number of factors extracted	Variance on number of factors explained	Cronbach's alpha
1	75.874	0.893

Component Matrix	Factor = 1
Decisions are based on intuition and experience	0.918
Decisions are data-driven	0.900
Data-driven decision-making is part of the organisation's culture	0.842
Data-driven decision-making is part of the organisation's strategy	0.820

Cronbach's alpha value on standardised items was calculated to establish reliability, and a value of 0.893 was obtained. It could thus be concluded that the factor is reliable.

3.3.2.7 Construct: Decision-making Quality

Table 3-20, Table 3-21 and Table 3-22 detail the results of the factor analysis performed for the Decision Quality construct.

Table 3-20: Factor Analysis for construct Decision-making Quality (Q14)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.863
Bartlett's Test of Sphericity	Approx. Chi-Square	295.675
	df	6
	Sig.	0.000

From Table 3-20, the KMO value of 0.863 and the fact that Bartlett's test of sphericity yielded a significance of less than 0.001 indicate that the data was suitable for factor analysis.

Table 3-21: Total Variance Explained for construct Decision-making Quality (Q14)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	3.372	84.289	84.289	3.372	84.289	84.289
2	0.259	6.471	90.759			
3	0.207	5.164	95.923			

4	0.163	4.077	100.000			

From Table 3-20 and Table 3-21, it is evident that only one factor explains 84.3% of the variance and it could, therefore, be inferred that this factor is valid. The specific items of the factor are shown in Table 3-22.

Table 3-22: Items forming part of construct Decision-making Quality (Q14)

Number of factors extracted	Variance on number of factors explained	Cronbach's alpha
1	84.289	0.938
Component Matrix		**Factor = 1**
Data-driven decisions are of good quality		0.932
Decisions based on intuition and experience are of good quality		0.930
Data-driven decisions are trusted		0.906
Data-driven decisions are of better quality than decisions based on intuition and experience		0.905

Cronbach's alpha value on standardised items was calculated to establish reliability, and a value of 0.938 was obtained. It could thus be concluded that the factor is reliable.

3.3.2.8 Construct: Data Sources

Table 3-23, Table 3-24 and Table 3-25 detail the results of the factor analysis performed for construct Decision Quality.

Table 3-23: Factor Analysis for construct Data Sources (Q15)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.941
Bartlett's Test of Sphericity	Approx. Chi-Square	1441.303
	df	66
	Sig.	0.000

From Table 3-23, the KMO value of 0.941 and the fact that Bartlett's test of sphericity yielded a significance of less than 0.001 indicate that the data was suitable for factor analysis.

Table 3-24: Total Variance Explained for construct Data Sources (Q15)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	% of Variance	Cumulative %	% of Variance	Cumulative %
1	9.776	81.466	81.466	9.776	81.466	81.466
2	0.491	4.090	85.556			
3	0.451	3.756	89.312			
4	0.270	2.247	91.559			

From Table 3-23 and Table 3-24, it is evident that only one factor explains 81.5% of the variance and it could, therefore, be inferred that this factor is valid. The specific items of the factor are shown in Table 3-25.

Table 3-25: Items forming part of construct Data Sources (Q15)

Number of factors extracted	Variance on number of factors explained	Cronbach's alpha
1	81.466	0.979
Component Matrix		**Factor = 1**
Office documentation (events, emails, documents)		0.946
Business Activity Data		0.938
Audio		0.935
Sensors/RFID		0.913
Public Open Data		0.912
Telemetry		0.905
Video		0.903
External feeds		0.899
Telecommunications (phone or data traffic)		0.894
Images		0.890
Point of sale		0.876

Social Media	0.815

Cronbach's alpha value on standardised items was calculated to establish reliability, and a value of 0.979 was obtained. It could thus be concluded that the factor is reliable.

The Cronbach Alpha value in all cases is above 0.70, indicating that the reliability of the data for Questions 7 to 15 (Pallant, 2013:90).

3.3.3 Correlations and Regressions

The first analysis was to look for relationships between the different factors. For this purpose, Pearson's product-moment correlations were calculated. Between all the factors there were very large positive correlations, all significant at the 0.01 level. These are shown in Table 3-26.

Table 3-26: Pearson Correlations

Pearson Correlations								
	Culture	**DA**	**SCA**	**Anti. Benif.**	**Curr. Benif.**	**Decis.**	**Source**	**Value**
Data Culture	1							
Data Analytics	.943**	1						
Supply Chain Analytics	.919**	.904**	1					
DA Anticipated Benefits	.935**	.955**	.950**	1				
DA Current Benefits	.988**	.948**	.882**	.925**	1			
Decisions	.918**	.977**	.877**	.928**	.922**	1		
Data Sources	.939**	.933**	.991**	.975**	.910**	.902**	1	
Value-adding data sources	.857**	.892**	.941**	.927**	.823**	.859**	.949**	1
*** Correlation is significant at the 0.01 level (2-tailed).*								

Although the strong positive correlations were encouraging in answering the main research question, no causal relationship was proven by this, and, therefore, multiple linear regressions were performed to establish whether there are any causal relationships between the different factors.

Regressions were used to determine the causal relationships between the various constructs. For each regression, the effect of the other constructs on the construct under scrutiny was determined. Although the sample size was small, the other assumptions of multiple regression (outliers, multicollinearity, normality, linearity homoscedasticity and independence of residuals) were all satisfied for all the regression analyses performed for this study.

3.3.3.1 Construct: Decision-making as the dependent variable

The purpose of the study was to establish whether the use of Data Analytics (and more specifically Supply Chain Analytics) leads to improved decision-making. The first multiple regression analysis, therefore, used Decision-making as the dependent variable. Table 3-27 shows regressions for the construct Decision-making as the dependent variable. All the assumptions necessary for regressions have been satisfied.

Table 3-27: Regression for the Decision-making construct

Component	Unstandardized Coefficients		Standardised Coefficients		
	B	Std. Error	Beta	t	Sig.
Data Culture	0.215	0.168	0.264	1.281	0.204
Data Analytics	1.026	0.093	1.111	10.984	0.000
Supply Chain Analytics	0.383	0.179	0.430	2.142	0.035
DA Anticipated Benefits	0.202	0.135	0.210	1.492	0.140
DA Current Benefits	-0.206	0.155	-0.264	-1.333	0.186
Value-adding data sources	-0.060	0.073	-0.068	-0.829	0.409
Data Sources	-0.624	0.262	-0.710	-2.382	0.020

In line with the objective of this study, the aim was to determine the constructs that contributed to the Decision-making construct. From Table 3-27, it is evident that Data Analytics, in general, had a large contribution towards decision-making, since it had a statistical significance lower than 0.05% and a large positive beta coefficient of greater than 1. More specifically, Supply Chain Analytics also had a statistically significant effect on Decision-making with a positive beta coefficient of 0.430.

Even though Data Sources had a higher statistical significance of 0.020%, it had a negative beta coefficient of -0.710. This was possibly the result of the fact that most respondents had different data sources listed.

This is an important result, as it proves that the use of Data Analytics, as a general category, and the use of Supply Chain Analytics, specifically, do lead to improved decision-making, which is the main research question being answered.

3.3.3.2 Construct: Data Analytics as the dependent variable

The next step was to ascertain which of the factors contributed to the use of Data Analytics. Table 3-28 shows regressions for the construct Data Analytics as the dependent variable. All the assumptions necessary for regressions have been satisfied.

Table 3-28: Regression for the Data Analytics construct

Component	Unstandardized Coefficients		Standardised Coefficients		
	B	Std. Error	Beta	t	Sig.
Data Culture	-0.237	0.125	-0.269	-1.892	0.062
Supply Chain Analytics	-0.295	0.135	-0.307	-2.184	0.032
DA Anticipated Benefits	0.038	0.104	0.036	0.362	0.718
DA Current Benefits	0.382	0.110	0.450	3.453	0.001
Value-adding data sources	0.137	0.053	0.141	2.574	0.012
Decision-making	0.589	0.054	0.544	10.984	0.000

The regression in Table 3-28 aimed to determine the constructs that contributed to Data Analytics construct. From Table 3-28, it evident that the Data Analytics Current Benefits had a fairly large contribution towards Data Analytics since it had a statistical significance lower than 0.05% and a fairly large positive beta coefficient of 0.450. Value-adding data sources also had a statistical significance lower than 0.05% and a positive beta coefficient of 0.141.

Similarly, Supply Chain Analytics had a statistical significance lower than 0.05% but had a negative beta coefficient of -0.307. This last result was totally unexpected but seems to indicate that the use of Data Analytics applications, other than Supply Chain Analytics, are still more prevalent than the mere use of Supply Chain Analytics.

The implication is that the use of Data Analytics and Supply Chain Analytics depend largely on participants' past experience of the benefits of using Data Analytics for decision-making, as well as on use of the correct sources for data.

3.3.3.3 Construct: Supply Chain Analytics as dependent variable

The next question asked was whether Supply Chain Analytics, specifically, depends on other factors. Table 3-29 shows regressions for the construct Supply Chain Analytics as the dependent variable. All the assumptions necessary for regressions have been satisfied.

Table 3-29: Regression for the Supply Chain Analytics construct

Component	Unstandardized Coefficients		Standardised Coefficients		
	B	Std. Error	Beta	t	Sig.
Data Culture	0.022	0.104	0.025	0.216	0.829
Data Analytics	-0.193	0.088	-0.186	-2.184	0.032
DA Anticipated Benefits	-0.288	0.078	-0.266	-3.704	0.000
DA Current Benefits	-0.031	0.096	-0.035	-0.322	0.748
Decision-making	0.143	0.067	0.128	2.142	0.035
Value-adding Data Sources	0.023	0.045	0.023	0.518	0.606

The regression in Table 3-29 aimed to determine the constructs that contributed to the Supply Chain Analytics construct. From Table 3-29, Data Analytics had a contribution to the use of Supply Chain Analytics by respondents since it had a statistical significance of 0.032% and a beta coefficient of -0.186. Although the Data Analytics Anticipated Benefits construct had a very low statistical significance of 0.000%, it had a negative beta coefficient of -0.266.

Table 3-29 shows that improved decision-making also increases the use of Supply Chain Analytics. This led to the next critical causal relationship: What really contributes to anticipated benefits of using Data Analytics.

3.3.3.4 Construct: Data Analytics Anticipated Benefits as dependent variable

Analysis of the factors that act as predictors to the anticipated benefits (as the dependent variable) followed to ascertain mainly what would motivate organisations to start using Data Analytics for improved decision-making. Table 3-30 shows regressions for the construct Data Analytics Anticipated Benefits as the dependent variable. All the assumptions necessary for regressions have been satisfied.

Table 3-30: Regression for the Data Analytics Anticipated Benefits construct

Component	Unstandardized Coefficients		Standardised Coefficients		
	B	Std. Error	Beta	t	Sig.
Data Culture	-0.348	0.133	-0.412	-2.621	0.011
Data Analytics	0.044	0.121	0.046	0.362	0.718
Supply Chain Analytics	-0.515	0.139	-0.556	-3.704	0.000
DA Current Benefits	0.312	0.123	0.384	2.534	0.013
Decision-making	0.136	0.091	0.130	1.492	0.140
Value-adding Data Sources	0.034	0.060	0.037	0.571	0.570

The regression in Table 3-30 aimed to determine the constructs that contributed to the Data Analytics Anticipated Benefits construct. From Table 3-30, the Data Analytics Current Benefits construct had a statistical significance of 0.013% and a positive beta coefficient of 0.384.

Interestingly, the Supply Chain Analytics construct provided a fairly large negative beta coefficient of -0.556 and had a statistical significance lower than 0.01% (Pallant, 2013:91). This may indicate that organisations still do not believe in the value of Supply Chain Analytics.

The effect of an organisation's culture on the anticipated benefits of using Data Analytics was a surprising result, apart from the fact that present benefits are predictors for expected benefits and that those who currently use Data Analytics extensively would expect more future benefits. This led to the following question: Are certain sources of data (or information) more valuable than others?

3.3.3.5 Construct: Value-adding Data Sources

It was expected that the list of data sources used might not be consistent with the other findings, as data sources are often industry-specific. For example, a bank would use telemetry more than a restaurant would, the latter possibly preferring point-of-sale data (Kim *et al.*, 2005). Table 3-31 shows regressions for the construct Value-adding Data Sources as the dependent variable. All the assumptions necessary for regressions have been satisfied.

Table 3-31: Regression for the Value-adding Data Sources construct

Component	Unstandardized Coefficients		Standardised Coefficients		
	B	Std. Error	Beta	t	Sig.
Data Culture	0.173	0.260	0.190	0.666	0.507
Data Analytics	0.565	0.219	0.547	2.574	0.012
Supply Chain Analytics	0.146	0.282	0.147	0.518	0.606
DA Anticipated Benefits	0.120	0.211	0.112	0.571	0.570
DA Current Benefits	-0.548	0.233	-0.626	-2.356	0.021
Decision-making	-0.143	0.172	-0.128	-0.829	0.409

The regression in Table 3-31 aimed to determine the constructs that contributed to the Value-adding Data Sources construct. From Table 3-31, the Data Analytics had a fairly large contribution towards data sources regarded as value-adding to respondents since it had a statistical significance of 0.012% and a positive beta coefficient of 0.547. Although the Data Analytics Current Benefits construct had a slightly higher statistical significance of 0.021%, it still contributed to a negative beta coefficient of -0.626.

The first relationship was expected and commensurate with what was found in chapter 2, that people who use Data Analytics extensively would likely use more data sources. The negative regression coefficient was surprising, indicating that people experiencing bigger benefits from Data Analytics use less sources. This could be attributed to a more focused approach in selecting data to use based on previous benefits experienced.

3.3.3.6 Construct: DA Current Benefits as dependent variable

Table 3-32 shows regressions for the construct DA Current Benefits as the dependent variable. All the assumptions necessary for regressions have been satisfied.

Table 3-32: Regression for the DA Current Benefits construct

Component	Unstandardized Coefficients		Standardised Coefficients		
	B	Std. Error	Beta	t	Sig.
Data Culture	0.985	0.050	0.947	19.564	0.000
Data Analytics	0.344	0.100	0.291	3.453	0.001
Supply Chain Analytics	-0.043	0.132	-0.037	-0.322	0.748
DA Anticipated Benefits	0.241	0.095	0.196	2.534	0.013
Decision-making	-0.107	0.080	-0.083	-1.333	0.186
Value-adding Data Sources	-0.120	0.051	-0.105	-2.356	0.021

The regression in Table 3-32 aimed to determine the constructs that contributed to the Data Analytics Current Benefits construct. From Table 3-32, the Data Culture had a very large contribution towards current benefits gained from using Data Analytics to respondents since it had a statistical significance of 0.000% and a beta coefficient of 0.947. Also, Data Analytics and Data Analytics Anticipated Benefits constructs had a slightly statistical significance of 0.001% and 0.013%, with low positive beta coefficients of 0.291 and 0.196 respectively.

3.3.4 Comparison between different demographic groups

To ascertain whether there are significant differences between the different demographic groups, the following analyses were carried out.

3.3.4.1 ANOVA: Age of the organisation

The question to be answered is whether there is a relationship between the time that the organisation is in existence and (1) whether they use Data Analytics, (2) whether they use Supply Chain Analytics and (3) how good their decision-making is.

The low levels of significance ($p<0.01$ for all constructs) indicate that there is a statistically significant difference between organisations of different ages. Where the difference lies, was investigated using post-hoc tests, where these could be performed. For the sake of brevity, only the three main constructs, namely decision-making, Data Analytics and Supply Chain Analytics, were included.

Levene's test for homogeneity of variances revealed that the variances are not homogeneous, with the significance values all being smaller than 0.01. The ANOVA is given in Table 3-33.

Table 3-33: ANOVA: Age of the organisation

		Sum of Squares	df	Mean Square	F	Sig.
Data Analytics	Between Groups	204.475	6	34.079	136.449	0.000
	Within Groups	19.981	80	0.250		
	Total	224.456	86			
Supply Chain Analytics	Between Groups	221.089	6	36.848	141.394	0.000
	Within Groups	20.848	80	0.261		
	Total	241.937	86			
Decision-making	Between Groups	170.337	6	28.390	108.580	0.000
	Within Groups	20.917	80	0.261		
	Total	191.254	86			

The level of significance for all three factors is below 0.01, therefore there are some statistical significance between businesses of different ages regarding how they perceive these values. Post-hoc results could not be calculated due to missing values. It can, however, be deduced that the age of a business has an influence on whether it uses Data Analytics and Supply Chain Analytics, and also on the decision-making process in the organisation.

3.3.4.2 ANOVA: Industry representation

The question to be answered is whether there is a relationship between industry in which the organisation is and (1) whether they use Data Analytics, (2) whether they use Supply Chain Analytics and (3) how good their decision-making is.

The low levels of significance ($p<0.01$ for all constructs) indicate that there is a statistically significant difference between organisations in different industries. Where the difference lies, was investigated using post-hoc tests, where these could be performed. For the sake of brevity, only the three main constructs, namely decision-making, Data Analytics and Supply Chain Analytics, are included and post-hoc test results are only reported where the results would warrant specific conclusions and recommendations.

Levene's test for homogeneity of variances revealed that the variances are not homogeneous, with the significance values all being smaller than 0.01. The ANOVA is given in Table 3-34.

Table 3-34: ANOVA: Industry representation

		Sum of Squares	df	Mean Square	F	Sig.
Data Analytics	Between Groups	197.080	5	39.416	116.622	.000
	Within Groups	27.376	81	.338		
	Total	224.456	86			
Supply Chain Analytics	Between Groups	191.279	5	38.256	61.168	.000
	Within Groups	50.659	81	.625		
	Total	241.937	86			
Decision-making	Between Groups	163.547	5	32.709	95.622	.000
	Within Groups	27.708	81	.342		
	Total	191.254	86			

The ANOVA revealed statistically significant differences between different industries regarding all three these constructs. Although post-hoc tests were performed for this analysis, the specific differences between the different industries is beyond the scope of this study.

3.3.4.3 ANOVA: Size of the organisation (Number of employees)

The question to be answered is whether there is a relationship between the number of employees working in the organisation and (1) whether they use Data Analytics, (2) whether they use Supply Chain Analytics and (3) how good their decision-making is.

The low levels of significance ($p<0.01$ for all constructs) indicate that there is a statistically significant difference between organisations of different sizes. Where the difference lies, was investigated using post-hoc tests, where these could be performed. For the sake of brevity, only the three main constructs, namely decision-making, Data Analytics and Supply Chain Analytics, are included and post-hoc test results are only reported where the results would warrant specific conclusions and recommendations.

Levene's test for homogeneity of variances revealed that the variances are not homogeneous, with the significance values all being smaller than 0.01. The ANOVA is given in Table 3-35.

Table 3-35: ANOVA: Size of the organisation

		Sum of Squares	df	Mean Square	F	Sig.
Data Analytics	Between Groups	178.923	5	35.785	63.658	.000
	Within Groups	45.533	81	.562		
	Total	224.456	86			
Supply Chain Analytics	Between Groups	174.356	5	34.871	41.795	.000
	Within Groups	67.581	81	.834		
	Total	241.937	86			
Decision-making	Between Groups	148.767	5	29.753	56.723	.000
	Within Groups	42.488	81	.525		
	Total	191.254	86			

The low level of significance indicates that there is a statistically significant difference between different size organisations with regards to all three these constructs. Bonferroni's post-hoc test confirmed that companies with more employees tend to use Data Analytics and Supply Chain Analytics more and base their decisions on this, relative to smaller companies.

3.3.4.4 ANOVA: Experience of respondents with big data

The question to be answered is whether there is a relationship between the experience level of respondents with big data and (1) whether they use Data Analytics, (2) whether they use Supply Chain Analytics and (3) how good their decision-making is.

The low levels of significance ($p<0.01$ for all constructs) indicate that there is a statistically significant difference between respondents with different levels of experience with big data. Where the difference lies, was investigated using post-hoc tests, where these could be performed. For the sake of brevity, only the three main constructs, namely decision-making, Data Analytics and Supply Chain Analytics, are included and post-hoc test results are only reported where the results would warrant specific conclusions and recommendations.

Levene's test for homogeneity of variances revealed that the variances are not homogeneous, with the significance values all being smaller than 0.01. The ANOVA is given in Table 3-36.

Table 3-36: ANOVA: Experience of respondents with big data

		Sum of Squares	df	Mean Square	F	Sig.
Data Analytics	Between Groups	196.709	6	32.785	94.527	.000
	Within Groups	27.747	80	.347		
	Total	224.456	86			
Supply Chain Analytics	Between Groups	214.962	6	35.827	106.249	.000
	Within Groups	26.976	80	.337		
	Total	241.937	86			
Decision-making	Between Groups	166.671	6	27.778	90.397	.000
	Within Groups	24.584	80	.307		

		Sum of Squares	df	Mean Square	F	Sig.
	Total	191.254	86			

The level of significance for all three factors is below 0.01. Therefore, there are some statistical significance between respondents with different levels of experience with big data regarding how they perceive these values. Post-hoc results could not be calculated due to missing values. It can, however, be deduced that level of experience with big data has an influence on whether it uses Data Analytics and Supply Chain Analytics and also on the decision-making process in the organisation.

It can be summarised that the use of Data Analytics and Supply Chain Analytics and the level of decision-making ensuing from this differs considerably (statistically significant) between different demographic groupings and different organisations. However, it does not alter the relationship between these constructs as established using multiple regression analysis.

3.3.5 Narrative Questions

Five narrative questions were included in the questionnaire (four of which were compulsory) as a contingency in the event that the quantitative analysis was not conclusive and required further support. However, since the statistical analysis of the data supports the objectives conclusively, it seemed unnecessary to analyse the narrative questions further, and it may be a subject for future research.

The most prominent comments from respondents that support the main theme of the study came from Question 24, "*How does your organisation expect to create value from Big Data?*" The following comments have been extracted from the responses based on relevance:

"Data literacy is key to moving forward in a data-driven organization"

"My company is quite large with some business units starting to see the value of data-driven-decisions, but some are still stuck in the dark ages. Getting access to most of our data is a problem and things like GDPR and POPI makes it even more difficult."

“Get dedicated data scientists to analyse the data”

“Translate Big data into actionable insight”

“Honestly, we have acquired all the tools, have access to many data sources and are well funded but we have yet to extract true value.”

“Big data appears to be the new "gold" for business. If you have access to data, you can control the customer’s needs. By being able to control the customer, you will be better able to control your value-chain activities right up to the raw material suppliers.”

“Our company needs to focus on the education of the value of data first. Create a demand for data-driven decision-making before building the solutions.”

“A data-driven culture within an organization is driven from the top-down. If EXCO does not place a high priority on data and data-driven decision making, then neither will 90% of the organization's staff.”

3.4 Chapter Conclusion

This chapter discussed the different elements of the empirical study including discussions on the research design and data analysis techniques employed. It then went into great length to provide the details of the results that were obtained through statistical analysis. In addition, comments from the narrative questions were discussed.

The following chapter will summarise the findings, provide implications of the study to managers, as well as recommendations and concluding remarks.

CHAPTER 4: CONCLUSIONS AND RECOMMENDATIONS

4.1 Introduction

The primary objective of this study was to research the effect that the use of Supply Chain Analytics (SCA) has on business decision-making. A literature study was performed in Chapter 2 to understand the various elements that contribute to SCA and how SCA is used in practice. Chapter 3 details the results from the empirical study performed quantitatively using a new questionnaire designed to measure the constructs of the study.

This chapter aims to draw conclusions from the results presented in the previous chapter. Recommendations are offered regarding insights from this study for organisations as well as implications for managers. The success of the study is evaluated against the primary and secondary objectives that were established in Chapter 1. Recommendations are also made for further research endeavours related to the topic at hand.

4.2 Conclusions

The purpose of this study was to determine whether there is a relationship between using Data Analytics with specific reference to the supply chain (Supply Chain Analytics) in an organisation and its business decision-making outcomes. From Table 3-27, it is evident that there is a statistically significant positive causal relationship between the use of Data Analytics and business decision-making in an organisation. More importantly, there is a statistically significant positive causal relationship between the use of Supply Chain Analytics and business decision-making in an organisation.

When looking at all the causal relationships that exist between the major constructs, as can be seen from Figure 4–1, there is a definite pattern that can be distinguished. Refining the constructs slightly to combine the related constructs of Data Analytics Current Benefits and Data Analytics Anticipated Benefits, a more streamlined causal relationship diagram transpires, as can be seen from Figure 4–2.

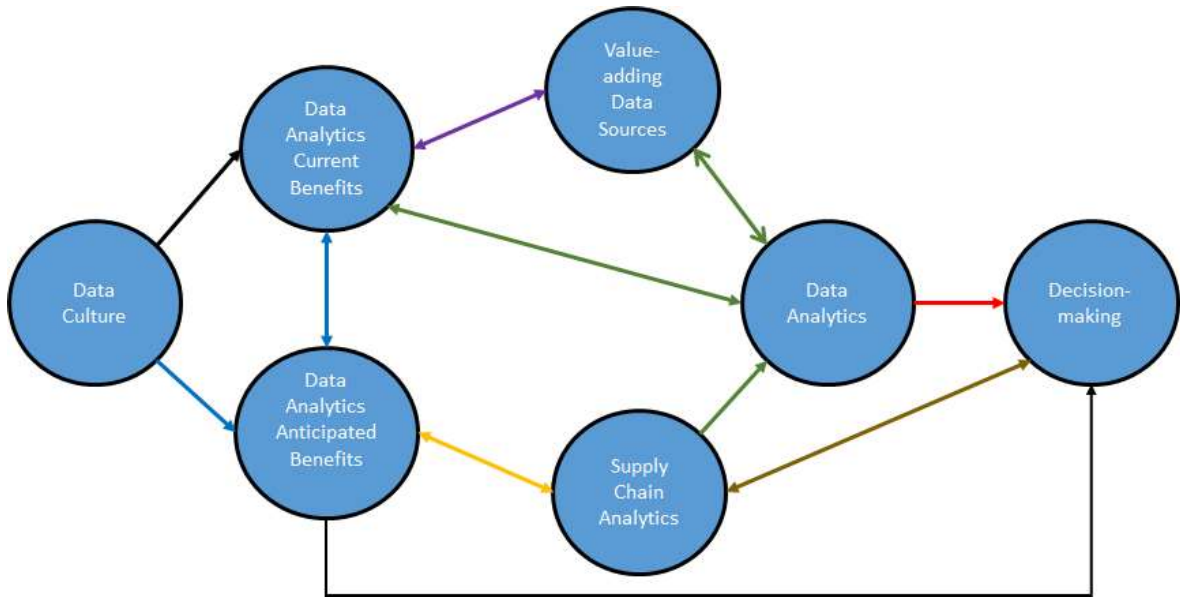

Figure 4–1: Causal relationships of constructs (as determined through the regression analyses in section 3.3.3)

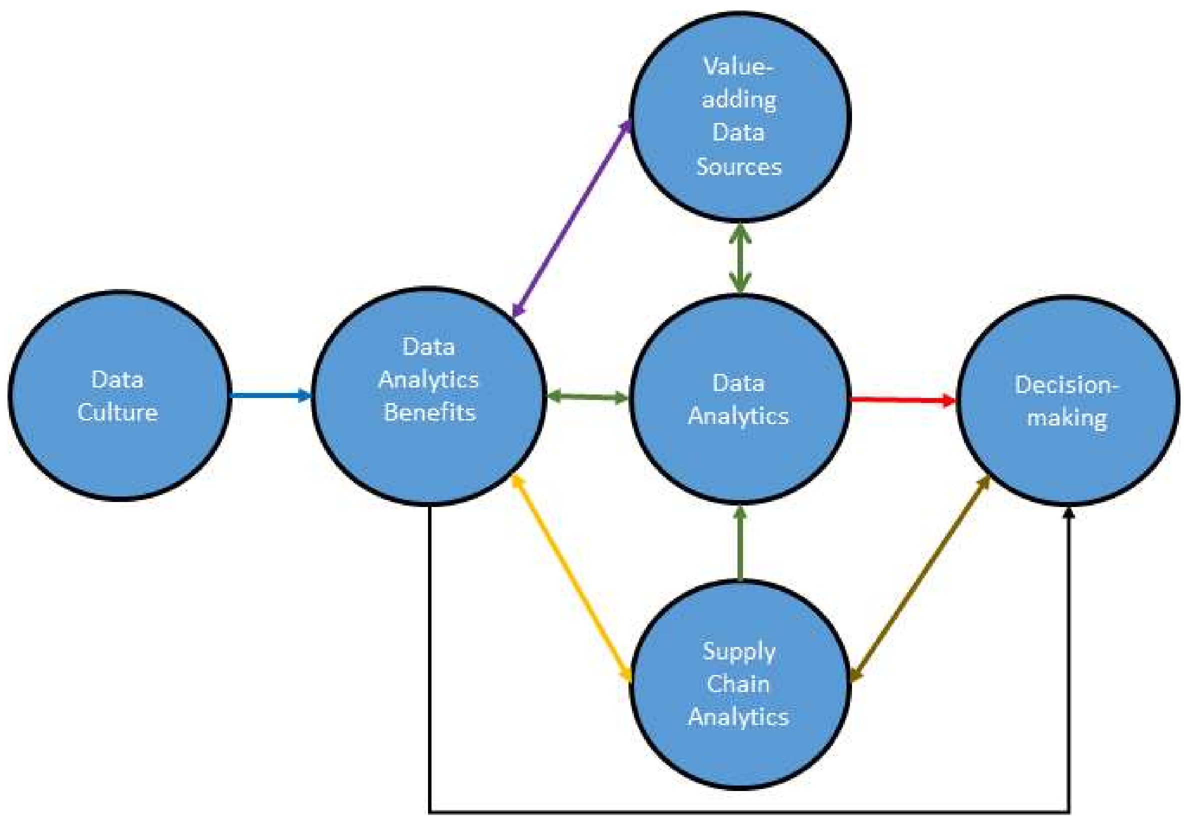

Figure 4–2: Causal relationships of constructs (streamlined)

From Figure 4–2, Data Analytics and Supply Chain Analytics contribute to Decision-making (red and brown arrows). This has already been established through statistical analysis (see section 3.3.3.1). The three primary constructs contributing to Data Analytics (green arrows) are Data Analytics Benefits, Supply Chain Analytics and Value-adding Data Sources (see section 3.3.3.2). This means that the fact that organisations benefit from Data Analytics and uses it, contributes to it being used more and eventually impacting on business decision-making. Therefore, organisations will use Data Analytics if it provides them with value and benefits them now with possible future opportunities. Note in Figure 4–2, where the arrow points both ways, the contribution is mutual and goes both ways.

The yellow arrow highlights the relationship between Data Analytics Benefits and Supply Chain Analytics. The is a mutual relation between these two since the more Supply Chain Analytics is used, the more the organisation will benefit from it and the more it benefits, the more the organisation will actually utilise Supply Chain Analytics capabilities (see section 3.3.3.3).

The purple arrow highlights the relationship between Data Analytics Benefits and Value-adding Data Sources. Again, there exists a mutual relation between these two since the more Value-adding Data Sources are used, the more the organisation will benefit from it and the more it benefits, the more the organisation will explore, use and grow Value-adding Data Sources (see section 3.3.3.5).

Similar to the red arrow, both the brown and the black arrows also contribute, to a lesser extent, to decision-making. This results from using Data Analytics and benefitting from it that has a positive impact on and relationship to better and more effective decision-making.

The primary driving force of Data Analytics and Supply Chain Analytics, and ultimately decision-making, is Data Culture. Organisations that have established a culture that is data-driven and where decisions based on intuition are tested against data-driven decisions, are those that benefit most from Data Analytics. These organisations have established an inherent trust in the data to guide them because the facts cannot lie. These organisations further go out and use all the available data to strengthen their ability to make decisions. In essence, there appears to be unseen positive feedback where organisations start using Data Analytics and Supply Chain Analytics to drive decision-making that further enhances the notion of an organisational data culture and builds a foundation for effective data-driven decision-making.

4.3 Implications for Management

Managers and leaders of companies can glean a lot from looking at industry trends and how digital transformations are changing the business landscape. This study has highlighted the need for using data and analytics in the area of the supply chain to drive decision-making and the factors that contribute to decision-making. In a recent study by McKinsey, this notion is taken further by pointing to key elements that organisations can use to elevate themselves above the competition through the use of data and analytics (2019). These elements echo the conclusions from this study and are given below (McKinsey, 2019):

1. *Strategy* – Organisations should make digital transformation part of the organisation's strategy by nurturing a data culture that is used to drive the transformation. Allocating more resources to digital transformation will speed up to change to a data-driven culture.
2. *Talent* – Hiring employees that are well-versed in data and technology fosters a digital workforce. Adding digital experts deepens and broadens the technological skills of the organisation as a whole and raises the bar of the organisation's competitive advantage.
3. *Agile* – Changing the way people work to provide organisations with agility and flexibility to meet market demands more rapidly based on decisions that are rapid and data-driven.
4. Analytics – Organisations are encouraged to embed analytics throughout the organisation by defining strategic opportunities, strengthening data-management practices, and enabling more employees to make analytics-driven decisions.
5. Evolve – Organisations that appreciate the evolving nature of technology and embraces the agility of innovation, while still enabling basic operations, are seeing more digital transformation that spurs growth.
6. Operations – Organisations realise that digital transformation requires bringing in technology solutions for everyday business operations that require and stimulate the transformation of operational processes to align with technology.

From the above it is, therefore, imperative that managers and leaders take heed to embrace Data Analytics in their organisations, to upskill their employees and permeate data-driven decision-making as a norm throughout the organisation. Only through consistent application of these principles will organisations stimulate a data culture that will give them a sustainable competitive advantage and help them accelerate into the digital era.

4.4 Recommendations

Based on the conclusions discussed in the preceding sections the following recommendations are offered:

- Digital transformation needs to be at the heart of the organisation's strategy and not an add-on or afterthought. To establish and foster a data culture takes hard work and commitment firstly from those in C-suite positions who make data and data-driven decision-making a priority within the organisation. Establishing a clear vision and goals for such a strategy will enable the organisation as a whole to understand what is required and attract buy-in from all stakeholders.
- In line with the above, organisations need to align the data strategy with business objectives in the form of data-centric goals and actionable Key Performance Indicators (KPIs) that deliver a sustainable competitive advantage to the organisation.
- Organisations further need to invest resources to foster a data culture. They will need to purchase the right tools, hire the right employees and consultants (the best in the industry?), and establish the right skills within the organisation. This may be a hard sell but will reveal whether top management is committed to the data strategy.
- Finally, organisations need to force themselves to move away from gathering huge amounts of data while still depending heavily on opinion or intuition to make decisions. Creating a data culture would imply making decisions that are data-driven and feeding the results back into the system to continually improve business decision-making and performance within the organisation.

4.5 Evaluation of the Study

To determine the success of the study a critical evaluation of the achievement of the primary and secondary objectives is required.

4.5.1 Primary Objective

The main objective of this study was to perform an analysis of the impact of Supply Chain Analytics on business decision-making. To gain an appreciation for the different elements involved, a literature study was conducted (see Chapter 2). To understand the impact the independent variable had on the dependent variable an empirical study was conducted that clearly established a causal relationship and highlighted the effect on business decision-making (see section 3.3.3).

4.5.2 Secondary Objectives

The secondary objectives of the study were:

- To find theoretical evidence of a possible relationship between SCA and business decision-making.
- To quantify the effect of SCA on business decision-making.
- To recommend how SCA could be employed to improve operations.

The objective to find theoretical evidence of a possible relationship between SCA and business decision-making was achieved through the literature study and the practical outflow of that was achieved through the empirical study. Section 2.4 elaborated in detail on Supply Chain Analytics and concludes with the effect this has on decision-making from the literature (see section 2.4.7).

The objective to quantify the effect of SCA on business decision-making was achieved through the empirical study that established that a causal relationship exists (see section 3.3.3) and also quantifies the size of the relationship (see section 4.2).

The objective to recommend how SCA could be employed to improve operations was provided as part of both the Implications for Management (section 4.3) and the Recommendations (section 4.4).

4.6 Limitations of the Study

A major limitation of the study was that the sample size of the population may have been insufficient to do more advanced statistical analyses that could have provided more insight into causal relationships between constructs. This was clearly seen from the statistical analysis performed in Chapter 3 as indicated by the initial validity and reliability analysis. Even though the sample elicited feedback from multiple industry sectors and respondents from various levels within the organisation, the responses proved to be very homogeneous with abnormally high reliability.

4.7 Suggestions for Further Research

Using this research study as a basis, the following suggestions for further work are suggested:

- The study focused on the effect of SCA on decision-making without looking at the performance benefits gained from its use. It would be very interesting to see whether the use of SCA has a measurable impact on the performance of the supply chain (for example, performance and throughput) and the organisation as a whole (for example, earnings and performance).
- Another aspect that would prove interesting would be to explore the effect of SCA on decision-making by critically comparing various industries and examining which industries benefit more from SCA and why this is the case.
- In the literature and in some responses, respondents indicated that Big Data is more a buzz word without real application due to various constraints of Big Data (such as security and privacy). A longitudinal case study where Big Data has been used to effectively solve real-world applications would also prove a benefit to ensure applicability of technology to practical applications.
- It is proposed that a full qualitative analysis be performed on the narrative (qualitative) questions of the study in support of the quantitative analysis that could possibly extract more insights into the value of the topic to the industry from the respondents' viewpoint. However, such a qualitative study fell outside the scope of the current study.
- Finally, a study that utilises Artificial Intelligence in conjunction with SCA to optimise Data Analytics for improved decision-making is a fairly under-researched

topic that may provide further insights that could be used by organisations to address supply chain complexities.

4.8 Overall Conclusion

As the primary objective, this study set out to explore the effect that Supply Chain Analytics has on business decision-making. The results of the quantitative study highlighted the impact of Data Analytics and Supply Chain Analytics on Decision-making but also revealed the impact Data Culture and the benefits of using Supply Chain Analytics have on Data Analytics and ultimately decision-making. The causal relationships of these constructs were discussed in this chapter and conclusions were provided. An important factor that came out was that organisations can stimulate effective decision-making through Data Analytics by fostering a data culture within the organisation.

Recommendations based on the findings and implications to managers were offered to develop and nurture data-driven decision-making within the organisation. Suggestions were also made for future research related to the topic of this study.

The study was critically evaluated against the primary and secondary objectives and was found to be successful based on both primary and secondary objectives being achieved.

REFERENCES

Ahmed, E., Yaqoob, I., Hashem, I.A.T., Khan, I., Ahmed, A.I.A., Imran, M. and Vasilakos, A.V. 2017. The role of big data analytics in Internet of Things. Computer Networks, 129, pp.459-471.

Aho, A.M. 2015. Product data analytics service model for manufacturing company. In International Conference on Knowledge Management in Organizations. pp.282-296.

Ai, X., Chen, J., Zhao, H. & Tang, X. 2012. Competition among supply chains: implications of full returns policy. International Journal of Production Economics, 139(1), pp.257-265.

Amankwah-Amoah, J. 2016. Emerging economies, emerging challenges: Mobilising and capturing value from big data. Technological Forecasting and Social Change, 110, pp.167-174.

Amato, G., Candela, L., Castelli, D., Esuli, A., Falchi, F., Gennaro, C., Giannotti, F., Monreale, A., Nanni, M., Pagano, P. and Pappalardo, L. 2018. How data mining and machine learning evolved from relational data base to data science. In A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years. pp.287-306.

Arjoon, S., Turriago-Hoyos, A. and Thoene, U. 2018. Virtuousness and the Common Good as a Conceptual Framework for Harmonizing the Goals of the Individual, Organizations, and the Economy. Journal of Business Ethics, 147(1), pp.143-163.

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: Logistics and Transportation Review, 114, pp.416-436.

Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A. and Buyya, R. 2015. Big Data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing, 79, pp.1-15.

Ayed, A.B., Halima, M.B. and Alimi, A.M. 2015. Big data analytics for logistics and transportation. In 2015 4th International Conference on Advanced Logistics and Transport (ICALT). pp.311-316.

Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J. and Zhao, J.L. 2016. Transformational Issues of Big Data and Analytics in Networked Business. MIS Quarterly, 40(4).

Bag, D. 2016. Business analytics. Routledge.

Banerjee, A. and Chaudhury, S. 2010. Statistics without tears: Populations and samples. Industrial psychiatry journal, 19(1), pp.60-65.

Bello-Orgaz, G., Jung, J.J. and Camacho, D. 2016. Social big data: Recent achievements and new challenges. Information Fusion, 28, pp.45-59.

Bhushan, U., Gujarathi, R., Banerjee, A., Sharma, H. and Seetharaman, A. 2017. The Impact of Hidden Costs on Production and Operations. Journal of Accounting-Business & Management, 24(1), pp.1-20.

Biswas, S. and Sen, J. 2017. A proposed architecture for big data-driven supply chain analytics. International Journal of Supply Chain Management, 6, pp.1–24.

Blackburn, R., Lurz, K., Priese, B., Göb, R. and Darkow, I.L. 2015. A predictive analytics approach for demand forecasting in the process industry. International Transactions in Operational Research, 22(3), pp.407-428.

Blazquez, D. and Domenech, J., 2018. Big Data sources and methods for social and economic analyses. Technological Forecasting and Social Change, 130, pp.99-113.

Bruque Camara, S., Moyano Fuentes, J. and Maqueira Marin, J.M. 2015. Cloud computing, Web 2.0, and operational performance: the mediating role of supply chain integration. The International Journal of Logistics Management, 26(3), pp.426-458.

Bryman, A. & Bell, E. 2017. Research Methodology: Business and Management Contexts, Oxford University Press: Southern Africa, ABC Press: Cape Town.

Bumblauskas, D., Nold, H., Bumblauskas, P. and Igou, A., 2017. Big data analytics: transforming data to action. Business Process Management Journal, 23(3), pp.703-720.

Calof, J., Richards, G. and Smith, J. 2015. Foresight, competitive intelligence and business analytics – tools for making industrial programmes more efficient. Foresight-Russia, 9(1), pp.68-81.

Cao, G., Duan, Y. and Li, G. 2015. Linking business analytics to decision making effectiveness: A path model analysis. IEEE Transactions on Engineering Management, 62(3), pp.384-395.

Chae, B., Olson, D. and Sheu, C. 2014. The impact of supply chain analytics on operational performance: a resource-based view. International Journal of Production Research, 52(16), pp.4695-4710.

Cheatham, M. 2015. Privacy in the Age of Big Data. In The 2015 International Conference on Collaboration Technologies and Systems (CTS), pp.334–335.

Chen, H., Chiang, R.H. and Storey, V.C. 2012. Business intelligence and analytics: from big data to big impact. MIS quarterly, pp.1165-1188.

Chen, M., Mao, S., Zhang, Y. & Leung, V.C. 2014. Big Data: Related Technologies, Challenges and Future Prospects. Heidelberg: Springer.

Choi, T.M., Wallace, S.W. and Wang, Y. 2018. Big data analytics in operations management. Production and Operations Management, 27(10), pp.1868-1883.

Côrte-Real, N., Oliveira, T. and Ruivo, P. 2017. Assessing business value of Big Data Analytics in European firms. Journal of Business Research, 70, pp.379-390.

Davenport, T.H. and Dyché, J. 2013. Big data in big companies. International Institute for Analytics, 3, pp.1-31.

Davenport, T.H. & Kim, J. 2013. Keeping up with the quants: Your guide to understanding and using analytics. Harvard Business Review Press: Massachusetts.

Davenport, T. and Harris, J. 2017. Competing on Analytics: Updated, with a New Introduction: The New Science of Winning. Harvard Business Press.

Deokar, A.V., Gupta, A., Iyer, L.S. and Jones, M.C. 2018. Analytics and Data Science.

Eisenstein, M. 2015. Big Data: The Power of Petabytes. Nature 527 (7576): S2–S4.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., 1996. From data mining to knowledge discovery in databases. AI magazine, 17(3), pp.37-37.

Frizzo-Barker, J., Chow-White, P.A., Mozafari, M. and Ha, D. 2016. An empirical study of the rise of big data in business scholarship. International Journal of Information Management, 36(3), pp.403-413.

Gandomi, A. & Haider, M. 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), pp.137-144.

Ghani, N.A., Hamid, S., Hashem, I.A.T. and Ahmed, E. 2018. Social media big data analytics: A survey. Computers in Human Behavior.

Grover, V., Chiang, R.H., Liang, T.P. and Zhang, D. 2018. Creating strategic business value from big data analytics: A research framework. Journal of Management Information Systems, 35(2), pp.388-423.

Günther, W.A., Mehrizi, M.H.R., Huysman, M. & Feldberg, F. 2017. Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems, 26(3), pp.191-209.

Hazen, B.T., Skipper, J.B., Boone, C.A. and Hill, R.R. 2018. Back in business: Operations research in support of big data analytics for operations and supply chain management. Annals of Operations Research, 270(1-2), pp.201-211.

Haulder, N., Kumar, A. and Shiwakoti, N., 2019. An analysis of core functions offered by software packages aimed at the supply chain management software market. Computers & Industrial Engineering, 138, p.106-116.

Heizer, J., Render, B. & Munson C. 2016. Operations Management: Sustainability and Supply Chain Management. 12th edition. Boston: Pearson.

Hu, H., Wen, Y., Chua, T.S. and Li, X. 2014. Toward scalable systems for big data analytics: A technology tutorial. IEEE access, 2, pp.652-687.

Ittmann, H.W. 2015. The impact of big data and business analytics on supply chain management. Journal of Transport and Supply Chain Management, 9(1), pp.1-9.

Jain, A.D.S., Mehta, I., Mitra, J. & Agrawal, S. 2017. Application of big data in supply chain management. Materials Today: Proceedings, 4(2), pp.1106-1115.

Jin, Y. and Ji, S. 2013. Partner choice of supply chain based on 3d printing and big data. Information Technology Journal, 12(22), pp.6822-6826.

Jordan, M.I. and Mitchell, T.M. 2015. Machine learning: Trends, perspectives, and prospects. Science, 349(6245), pp.255-260.

Kache, F. and Seuring, S. 2017. Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. International Journal of Operations & Production Management, 37(1), pp.10-36.

Kashyap, H., Ahmed, H.A., Hoque, N., Roy, S. and Bhattacharyya, D.K. 2015. Big data analytics in bioinformatics: A machine learning perspective. pp.1-20.

Kasturi, E., Devi, S.P., Kiran, S.V. and Manivannan, S. 2016. Airline Route profitability analysis and Optimization using BIG DATA analytics on aviation data sets under heuristic techniques. Procedia Computer Science, 87, pp.86-92.

Khan, N., Yaqoob, I., Hashem, I.A.T., Inayat, Z., Mahmoud Ali, W.K., Alam, M., Shiraz, M. and Gani, A. 2014. Big Data: Survey, Technologies, Opportunities, and Challenges. The Scientific World Journal 2014:pp.1–18.

Khayyat, Z., Ilyas, I.F., Jindal, A., Madden, S., Ouzzani, M., Papotti, P., Quiané-Ruiz, J.A., Tang, N. and Yin, S. 2015, May. Bigdansing: A system for big data cleansing. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. pp.1215-1230.

Kim, D.S., Hosain, S.Z., Ho, H.D., Pedersen, S.D. & Pederson, D.V. 2005. Methods and apparatus for implementing telemetry applications on a subscriber identity module. (Patent: US 10079942B2).

Kim, G. H., Trimi, S. & Chung, J. H. 2014. Big-data Applications in the Government Sector. Communications of the ACM 57(3): pp.78–85.

Kiron, D. & Shockley, R. 2011. Creating business value with analytics. MIT Sloan Management Review, 53(1), p.57-63.

Laursen, G.H. & Thorlund, J. 2016. Business analytics for managers: Taking business intelligence beyond reporting. John Wiley & Sons.

Lee, I. 2017. Big data: Dimensions, evolution, impacts, and challenges. Business Horizons, 60(3), pp.293-303.

L'Heureux, A., Grolinger, K., Elyamany, H.F. and Capretz, M.A. 2017. Machine learning with big data: Challenges and approaches. IEEE Access, 5, pp.7776-7797.

Li, Z., Hu, F., Schnase, J.L., Duffy, D.Q., Lee, T., Bowen, M.K. and Yang, C. 2017. A spatiotemporal indexing approach for efficient processing of big array-based climate data with MapReduce. International Journal of Geographical Information Science, 31(1), pp.17-35.

Li, J., Xu, L., Tang, L., Wang, S. and Li, L. 2018. Big data in tourism research: A literature review. Tourism Management, 68, pp.301-323.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A.H. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Mayer-Schönberger, V. and Cukier, K. 2013. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt.

McKinsey & Company. 2019. How top companies excel with digital and analytics. McKinsey Global Institute.

Miah, S.J. 2015. A Cloud-based Business Analytics for Supply Chain Decision Support. Journal of Information Sciences and Computing Technologies, 4(1), pp.274-280.

Mirzaie, M., Behkamal, B. and Paydar, S. 2019. Big Data Quality: A systematic literature review and future research directions. arXiv preprint arXiv:1904.05353.

Nair, P.R. 2014. Supply Chain Analytics. CSI Communications, 33(9), pp.11-12.

Nasser, T., and Tariq, R. S. 2015. Big Data Challenges. Journal of Computer Engineering & Information Technology 4(3): 1–10.

Neaga, I., Liu, S., Xu, L., Chen, H. and Hao, Y. 2015. Cloud enabled big data business platform for logistics services: a research and development agenda. In International Conference on Decision Support System Technology. pp.22-33.

Nguyen, T., Li, Z.H.O.U., Spiegler, V., Ieromonachou, P. and Lin, Y. 2018. Big data analytics in supply chain management: A state-of-the-art literature review. Computers & Operations Research, pp.254-264.

O'Donovan, P., Leahy, K., Bruton, K. and O'Sullivan, D.T. 2015. Big data in manufacturing: a systematic mapping study. Journal of Big Data, 2(1), pp.2-20.

Oivo, M. 2016. Data science revolution in process improvement and assessment. Perspectives on Data Science for Software Engineering, pp.323-325.

Oxford Dictionary, 2016. Oxford Dictionary.

Pääkkönen, P. and Pakkala, D. 2015. Reference architecture and classification of technologies, products and services for big data systems. Big data research, 2(4), pp.166-186.

Padgavankar, M. H., and Gupta, S. R. 2014. Big Data Storage and Challenges. International Journal of Computer Science and Information Technologies. 5(2), pp.2218-2223.

Pallant, J. 2013. SPSS Survival Manual. A step by step guide to data analysis using SPSS, 4th edition. Allen & Unwin.

Park, H., Bellamy, M.A. and Basole, R.C., 2016. Visual analytics for supply network management: System design and evaluation. Decision Support Systems, 91, pp.89-102.

Popovič, A., Hackney, R., Coelho, P.S. and Jaklič, J. 2012. Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. Decision Support Systems, 54(1), pp.729-739.

Porter, M.E. and Kramer, M.R. 2019. Creating shared value. In Managing sustainable business. Springer, Dordrecht. pp.323-346

Ristoski, P. and Paulheim, H. 2016. Semantic Web in data mining and knowledge discovery: A comprehensive survey. Web semantics: science, services and agents on the World Wide Web, 36, pp.1-22.

Rozados, I.V. and Tjahjono, B. 2014. Big data analytics in supply chain management: Trends and related research. In 6th International Conference on Operations and Supply Chain Management, Bali. pp.1-12.

Saleh, A.S., Halili, E., Zeitun, R. and Salim, R. 2017. Global financial crisis, ownership structure and firm financial performance: An examination of listed firms in Australia. Studies in Economics and Finance, 34(4), pp.447-465.

Schoenherr, T. and Speier-Pero, C. 2015. Data science, predictive analytics, and big data in supply chain management: Current state and future potential. Journal of Business Logistics, 36(1), pp.120-132.

Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V. 2017. Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, pp.263-286.

Souza, G.C. 2014. Supply chain analytics. Business Horizons, 57(5):595-605.

Tan, K.H., Zhan, Y., Ji, G., Ye, F. & Chang, C. 2015. Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. International Journal of Production Economics, 165:223-233.

Taylor, J. & Raden, N. 2007. Smart Enough Systems: How to Deliver Competitive Advantage by Automating Hidden Decisions. 1st edition. Boston: Pearson Education, Inc.

Tiwari, S., Wee, H.M. and Daryanto, Y. 2018. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering, 115, pp.319-330.

Ularu, E.G., Puican, F.C., Apostu, A. and Velicanu, M. 2012. Perspectives on big data and big data analytics. Database Systems Journal, 3(4), pp.3-14.

Wamba, S.F., Akter, S., Edwards, A., Chopin, G. & Gnanzou, D. 2015. How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics, 165:234-246.

Wamba, S.F., Gunasekaran, A., Papadopoulos, T. and Ngai, E. 2018. Big data analytics in logistics and supply chain management. The International Journal of Logistics Management, 29(2), pp.478-484.

Wang, G., Gunasekaran, A., Ngai, E.W. & Papadopoulos, T. 2016. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics, 176:98-110.

Wang, X. and He, Y. 2016. Learning from uncertainty for big data: future analytical challenges and strategies. IEEE Systems, Man, and Cybernetics Magazine, 2(2), pp.26-31.

Wang, Y., Kung, L. and Byrd, T.A. 2018. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126, pp.3-13.

Wei, H. and Yu, C. 2019. Big Data Architecture and Reference Models. On the Move to Meaningful Internet Systems: OTM 2018 Workshops. Vol. 11231, p.15-24.

Wright, D.J. and Wang, S. 2011. The emergence of spatial cyberinfrastructure. Proceedings of the National Academy of Sciences, 108(14), pp.5488-5491.

Xu, Z., Frankwick, G.L. and Ramirez, E. 2016. Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. Journal of Business Research, 69(5), pp.1562-1566.

Yang, Y., Long, X., & Jiang, B. 2013. K-Means Method for Grouping in Hybrid MapReduce Cluster. Journal of Computers 8(10), pp.2648–2655.

Yang, C., Huang, Q., Li, Z., Liu, K. and Hu, F. 2017. Big Data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth, 10(1), pp.13-53.

Yin, S. and Kaynak, O. 2015. Big data for modern industry: challenges and trends [point of view]. Proceedings of the IEEE, 103(2), pp.143-146.

Yu, W., Chavez, R., Jacobs, M.A. & Feng, M. 2018. Data-driven supply chain capabilities and performance: A resource-based view. Transportation Research Part E: Logistics and Transportation Review, 114, pp.371-385.

Zhai, Y., Y. S. Ong, and I. W. Tsang. 2014. "The Emerging "Big Dimensionality"." IEEE Computational Intelligence Magazine, 9(3), pp.14–26.

Zheng, Z., Xie, S., Dai, H.N., Chen, X. and Wang, H. 2018. Blockchain challenges and opportunities: A survey. International Journal of Web and Grid Services, 14(4), pp.352-375.

Zhong, R.Y., Huang, G.Q., Lan, S., Dai, Q.Y., Chen, X. and Zhang, T. 2015. A big data approach for logistics trajectory discovery from RFID-enabled production data. International Journal of Production Economics, 165, pp.260-272.

Zhu, S., Song, J., Hazen, B.T., Lee, K. and 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.

Annexure A: Questionnaire

The following is the questionnaire that was sent out to respondents.

The effect of Supply Chain Analytics on business decision-making

The purpose of this questionnaire is to gain insights into how data and its analytics are used in influencing decisions related specifically to supply chain processes. The information, including any demographic information, obtained from the questionnaire will be used solely for research for the topic above as an input to an MBA study at the Northwest University. Your anonymity will be maintained, and all information will remain confidential.

All fields are required except for the last question.

1. How long has your organisation been in business?
    a. Less than 1 year
    b. 1-3 years
    c. 4-6 years
    d. 7-10 years
    e. More than 10 years
2. In what industry is your organisation?
    a. AI
    b. Energy
    c. Entertainment
    d. Finance
    e. IT
    f. Manufacturing
    g. Mining
    h. Pharmaceutical
    i. Professional services
    j. Retail
    k. Telecommunication
    l. Telecoms
    m. Transport
3. How many employees do you have?

a. 1-49
b. 50-124
c. 125-249
d. 250-499
e. 500+

4. What is your job title?
5. How long have you personally been involved in data?
   a. Less than 1 year
   b. 1-3 years
   c. 4-6 years
   d. 7-10 years
   e. More than 10 years
6. What is your personal experience with Big Data?
   a. No experience
   b. Novice
   c. Intermediate
   d. Advanced
   e. Expert
7. The following questions relate to the importance of data and information culture in your organisation (Likert scale with coding 1 [Strongly disagree], 2 [Disagree], 3 [Neutral], 4 [Agree], 5 [Strongly agree])
   a. The organisation values and benefits from data
   b. The organisation treats data as an asset
   c. Senior management is involved in data-related projects
   d. Data is important for decision-making
8. The following questions relate to Data Analytics in your organisation (Likert scale with coding 1 [Strongly disagree], 2 [Disagree], 3 [Neutral], 4 [Agree], 5 [Strongly agree])
   a. Data analytics is part of the organisation's strategy
   b. Data analytics adds to the organisation's competitive advantage
   c. The organisation has the required skills to handle data analytics
   d. The organisation has the required analytical tools to handle data analytics
   e. The organisation has the required resources to handle data analytics
   f. The organisation is competing on data and analytics

    g. The organisation uses data for trend analysis

9. The following questions relate to Supply Chain Analytics in your organisation (Likert scale with coding 1 [Strongly disagree], 2 [Disagree], 3 [Neutral], 4 [Agree], 5 [Strongly agree])
    a. Data analytics is used for supply chain decisions
10. In your organisation, indicate whether Data Analytics is used for the following supply chain activities (Likert scale with coding 1 [Strongly disagree], 2 [Disagree], 3 [Neutral], 4 [Agree], 5 [Strongly agree])
    a. Demand planning
    b. Supply planning
    c. Procurement market intelligence
    d. Supplier management
    e. Executive business planning
    f. Lead time reduction
    g. Quality improvement
    h. Productivity improvement
    i. Supply chain optimisation
11. Indicate what the anticipated benefits of using Data Analytics in Supply Chain activities are in your organisation (Likert scale with coding 1 [Strongly disagree], 2 [Disagree], 3 [Neutral], 4 [Agree], 5 [Strongly agree])
    a. Improving working capital
    b. Reducing costs
    c. Improving service
    d. Improving quality
    e. Understand risk better
    f. Improve planning accuracy
    g. Improve responsiveness
12. Indicate what benefits have already been reaped from using Data Analytics in Supply Chain activities in your organisation (Likert scale with coding 1 [Strongly disagree], 2 [Disagree], 3 [Neutral], 4 [Agree], 5 [Strongly agree])
    a. Improved working capital
    b. Reduced costs
    c. Improved service

    d. Improved quality
    e. Better risk management
    f. Improved planning accuracy
    g. Improved responsiveness

13. The following questions relate to decision-making in your organisation (Likert scale with coding 1 [Strongly disagree], 2 [Disagree], 3 [Neutral], 4 [Agree], 5 [Strongly agree])
    a. Decisions are based on intuition and experience
    b. Decisions are data-driven
    c. Data-driven decision-making is part of the organisation's strategy
    d. Data-driven decision-making is part of the organisation's culture
14. The following questions relate to the quality of decision-making in your organisation (Likert scale with coding 1 [Strongly disagree], 2 [Disagree], 3 [Neutral], 4 [Agree], 5 [Strongly agree])
    a. Data-driven decisions are of good quality
    b. Decisions based on intuition and experience are of good quality
    c. Data-driven decisions are of better quality than decisions based on intuition and experience
    d. Data-driven decisions are trusted
15. In your organisation, what are the sources of the data that you collect? (Likert scale with coding 1 [Currently Collect], 2 [Collect in next 3 years], 3 [No plans to collect], 4 [Don't know or not applicable])
    a. Business Activity Data
    b. Office documentation (events, emails, documents)
    c. Social Media
    d. Sensors/RFID
    e. Public Open Data
    f. Telecommunications (phone or data traffic)
    g. External feeds
    h. Point of sale
    i. Audio
    j. Images
    k. Video

l. Telemetry

16. In your organisation, which types of data do you see as adding the most value to the organisation? [Select up to 3 options]
    a. Business Activity Data
    b. Office documentation (events, emails, documents)
    c. Social Media
    d. Sensors/RFID
    e. Public Open Data
    f. Telecommunications (phone or data traffic)
    g. External feeds
    h. Point of sale
    i. Audio
    j. Images
    k. Video
    l. Telemetry
17. The following questions relate to how collected data is used (Likert scale with coding 1 [<20%], 2 [21-40%], 3 [41-60%], 4 [61-80%], 5 [>80%])
    a. External data collected compared to total data?
    b. Total data currently analysed?
    c. Total data expected to be analysed in 3 years?
18. What would most support your ability to make data-driven decisions? Choose 1 that applies from
    a. More data
    b. Better quality data
    c. A shared understanding of the benefits of data-driven decision-making
    d. Faster or easier access to data
    e. Other
19. In your organisation, how would you characterize the amount of data available to support decision-making? Choose 1 that applies from
    a. Too much
    b. Enough
    c. Not enough
    d. Don't know

20. What is the most significant challenge faced when making data-driven decisions? [Select up to 3 options]
    a. Lack of reliable data
    b. Takes too long to access data
    c. Issues with interpreting data
    d. An excessive amount of data
    e. Data not presented in a meaningful way
    f. Need skilled people to analyse the data
    g. Takes too long to analyse the data
    h. Other
21. What type of data does your organisation find relevant but has not yet been able to analyse?
22. What analytical tools do you consider particularly important?
23. Which operational areas in your organisation are involved in using data technologies and data analytics?
24. How does your organisation expect to create value from Big Data?
25. Please add any comments you feel might add value to the questionnaire [optional]