Gordon Institute of Business Science
University of Pretoria

Building Sense-making Dynamic capability through Data science: A framework for strategic decision making.

Simphiwe Nxumalo
Student number: 25363931

A research project submitted to the Gordon Institute of Business Science, University of Pretoria, in partial fulfilment of the requirements for the degree of Master of Business Administration.

01 December 2020

ABSTRACT

Leaders have a duty to make decisions that advance the course of their organisations. Cognitive limitations emanating from human bounded rationality, as well as complex external environments, place a limit on a leader's ability to make quality decisions. This problem becomes pronounced when leaders have to make strategic decisions; which are often irreversible and commit scarce organisational resources.

There is, therefore creates a pressing need for organisations to develop capabilities that can aid leaders and their organisation discern their complex environment and improve decision making. This research was an inquiry into the development of dynamic capabilities that would aid sense-making, and decision making. The focal point for this inquiry was the role of Data science in enabling the creation of sense-making capabilities.

The study deployed an inductive qualitative methodology suitable for exploring the link between data science-enabled dynamic capabilities, and strategic decision making, as an emerging area of work. Part of this exploration involved reviewing existing literature from both business and academic sources. Data collection from participants ensured the infusion of domain knowledge into this study as augmented by academic literature in order to improve the reliability of the findings.

KEYWORDS

Sense-making, Data science, analytical model, decision making

DECLARATION

I declare that this research project is my own work. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration at the Gordon Institute of Business Science, University of Pretoria. It has not been submitted before for any degree or examination in any other University. I further declare that I have obtained the necessary authorisation and consent to carry out this research.

Simphiwe Nxumalo

01 December 2020

TABLE OF CONTENTS

LIST OF FIGURES AND TABLE

1. Chapter 1 - Introduction to the research problem

Introduction

This research study investigated the relationship between the creation of sense-making dynamic capability and improved strategic decision making. Organisations are increasingly employing analytics and Data science to develop and deploy their Sense-making capabilities. Thus, this study explored how this process unfolds in organisations.

Thus, the main objective of this research was to investigate Data science as an enabler of the development of Sense-making Dynamic capabilities. The study was undertaken to help leaders improve decision making. Strategic decision making is mostly the function of senior leaders, with its significant impact on the sustainability of any organisation. However, this process may be fraught with challenges, which are often exacerbated by a high-velocity business environment. As such, leaders struggle to keep pace with the changes within their environment. This negatively impacts the quality of their decisions. Therefore, there is a need to find solutions to assist leaders in improving sense-making and keeping up with this dynamic environment. Failure to do so leads to poor strategic decisions and poor performance outcomes.

The study made use of academic literature as well as collected imperial data through interviews conducted with 13 individuals. This enabled the researcher to gain a deeper understanding of the factors that affect the sense-making, and how Data science can be leveraged to create for better decision making. The study culminated into a proposed general framework for enacting Data science driven capabilities.

1.1 The problem of decision making

The fate of organisations rests on the quality of decisions that leaders make over time (Badaracco, 2016). Andrews (1987) long observed that an organisation's success is linked to consistency in the patterns of decisions that leaders make. This pattern of decisions leads to coherent strategies that facilitate organisational success (Andrews, 1987). Yet, the process of arriving at coherent and high-quality decisions appear to elude leaders. De Smet, Jost Teece, D. J. (2014). The foundations of enterprise performance: dynamic and ordinary

capabilities in an (economic) theory of firms. *Academy of Management Perspectives*

28(4), 328–352.

Teece, D. J. (2018). Business models and Dynamic capabilities. *Long Range Planning*, *51*(1), 40–49. https://doi.org/10.1016/j.lrp.2017.06.007
Weiss (2019) found that the process of decision making is fraught with convoluted processes, increasing complexity in the operating environment, leader bias, and a tendency to seek consensus among leaders. The literature revealed that more than 60% of senior managers indicated that decision-making processes are ineffective (De Smet, Jost, & Weiss, 2019). Furthermore, the inherent human shortcomings in processing high volumes of simultaneous information compound the problem (Lovallo and Sibony, 2006).

If the fate of organisations is anchored on the ability of organisations to make quality decisions; then decision making becomes an important area of inquiry in academia and business. Pursuant to this, it is worth investigating the specific constructs of this problem and where the significant difficulties lie. Studies conducted by Snowden and Boone (2007) and Uhl-Bien and Arena (2018) pointed to the fast-paced and dynamic external environment within which organisations operate, as the critical issue hampering quality decisions.

It is imperative to note that organisations operated under varying external and internal environments, as a task or contextual environments (Snowden and Boone, 2007). These contexts vary from simple, to complicated, to complex, and ultimately to chaotic. Leaders must recognise the prevailing context in their environment and apply appropriate contextual approaches (Snowden and Boone, 2007; Rosenhead, Franco, Grint, & Friedland, 2019). Failure to do so may result in leaders making an incorrect diagnosis of the environment and making poor decisions.

In simple and complicated contexts, planning and analysis are found to improve decision making and performance. However, under complex domains, foreknowing and detailed planning may not be possible, making long-term predictive planning is unrealistic (Snowden and Boone, 2007; Rosenhead, Franco, Grint, & Friedland, 2019). What is required in complex environments is for leaders to experiment, probe and sense-make as they go along (Snowden and Boone, 2007; Rosenhead, Franco, Grint, & Friedland, 2019). Therefore, this may require leaders to develop better environmental scanning or sense-making capabilities to assist them in addressing the challenges presented by dynamic environments.

Therefore, a sense-making capability should assist leaders to overcome human shortcomings (Lovallo and Sibony, 2006). Human beings have a limited cognitive capacity, which manifests into biases and satisficing (Lovallo and Sibony, 2006; Robbins and Judge, 2018; De Smet, Jost, & Weiss, 2019). Satisficing is the tendency to think in familiar patterns. This results in the production of solutions that are closely related to those that have been produced before. Robbins and Judge (2018) argued that this leads to a lack of differentiation in the market. Uhl-Bien and Arena (2018) suggested that if the leaders are struggling to make complete sense of their environment, it follows that they are also struggling to design strategies that afford their businesses a sustainable competitive advantage. These cognitive impediments add to the need for a sense-making capability to support the leaders in scanning the external environment and make well informed and quality decisions.

1.2 Dynamic capabilities as a potential solution for improving decision making

Teece, Pisano, and Shuen (1997) pioneered the construct of Dynamic capabilities as a theoretical framework for creating sustainable competitive advantage under complex environments. The Dynamic capabilities framework describes how organisations: a) make sense of the opportunities and threats prevailing in a complex environment, b) seize opportunities and deal with risks, and c) transform their organisations in a manner that creates sustainable competitive advantage in the long run. Traditionally, the Dynamic capabilities framework was premised on three capabilities, namely: i) sense-making, ii) seizing, and transforming (Teece, 2007). Helfat and Raubitschek (2018) extended the framework to include an innovation capability. Also, these authors further qualified the sense-making capability as an “environmental scanning and sense-making” capability. This extended explanation is vital as it helps clarify the role of sense-making capability within the broader Dynamic capabilities theory, which is a crucial point of departure for this study.

This study emphasised that Sense-making capability is foundational to quality decision making. That is, for leaders to improve the quality of their discussions, they need a clear and unobstructed view of the problem at hand. Horning the sense-making capability would help the organisations; then seizing and transforming also improve. This research is specifically concerned with the problem of sense-making for leaders and the development of a framework that can help address the difficulties in decisions making.

Several tools, capabilities, and methodologies have been proposed as solutions that could assist businesses to navigate complexity. These include the use of IT systems, quantitative analysis methods, lean, design thinking, and many others (Provost and Fawcett, 2013; Conboy, Mikalef, Dennehy, & Krogstie, 2020). In the past decade, innovations like AI, Robotics, and Data science had increasingly come to the fore as candidate tools for creating competitive advantage (Schmarzo and Schmarzo, 2013). This study argues that Data science presents a suitable frame for the analysis of multiple environmental variables. This aspect of Data science provides an opportunity to deal with complex environments (Schmarzo and Schmarzo, 2013).

1.3 Definition and contribution of Data science

Conboy et al. (2020) defined Data science as the science and practice of using data to improve decision making. There are many underlying technologies and factors involved in building the Data science capability, and this study elaborates on some of these in Chapter 2.

A study by Côrte-Real, Ruivo, and Oliveira (2019) established that Data science practices within the organisation play a role in the creation of dynamic capabilities. Even though there are still gaps in understanding the processes by which this happens (Mikalef, Boura, Lekakos, & Krogstie, 2019). Given that Data science provides potential solutions to leadership difficulties; there is a need to develop a framework that closes the gaps. This research project undertook to contribute in this area.

1.4 Emphasis on the need for sense-making

Sense-making is a critical aspect of strategy formulation and strategy execution. Without the ability to sense make, leaders may not be able to seize the right opportunities and may not see threats surrounding their organisations. As a result, leaders may institute transformation efforts that are misdirected with little or no value for their enterprises. This could be expected to cause the erosion of strategic advantage in the long run. Therefore, the development of a framework that can help leaders make sense of an increasingly complex environment is essential and urgent.

1.5 The purpose of the research

The purpose of this study was to investigate the potential application of Data science as a mechanism for creating Sense-making Dynamic capabilities within an organisation. The primary objective is to establish a theoretical framework that would assist in addressing the gap in understanding this phenomenon. The majority of the research on the business benefits of implementing Data science is more from practise publications than academic literature (Mikalef, Boura, Lekakos, & Krogstie, 2019). The goal of this study was to contribute to the scholarly literature by developing a framework that can be applied by leaders in exploiting Data science to create Dynamic capabilities.

The study focused on sense-making capability in relation to the practice of Data science. In employing data as a strategic resource, organisations also face ethical pitfalls to be avoided, as part of Data science practices. It is envisaged that the framework to be recommended by this study will add value towards improving leaders' ability to implement sense-making dynamic capability. The outcome is improved decision-making business performance.

2. Chapter 2 - literature review

This research project entailed the investigation of the theory of Dynamic capabilities, specifically the sense-making capability therein; the practices in Data science as applied in business; and the relationship between the two constructs. This chapter reviews the literature on how organisations create Dynamic capabilities, the practices in Data science, and how the two constructs relate; especially in aiding decision making. As such, Dynamic capabilities form the theoretical framework of this study.

This chapter is organised into three main sections chronologically covering Decision making, Dynamics Capability, and Data science. Figure 2 below depicts the map of the subject and theory that was studied to illuminate issues that were investigated in order to assist answer research questions and address the research problem.

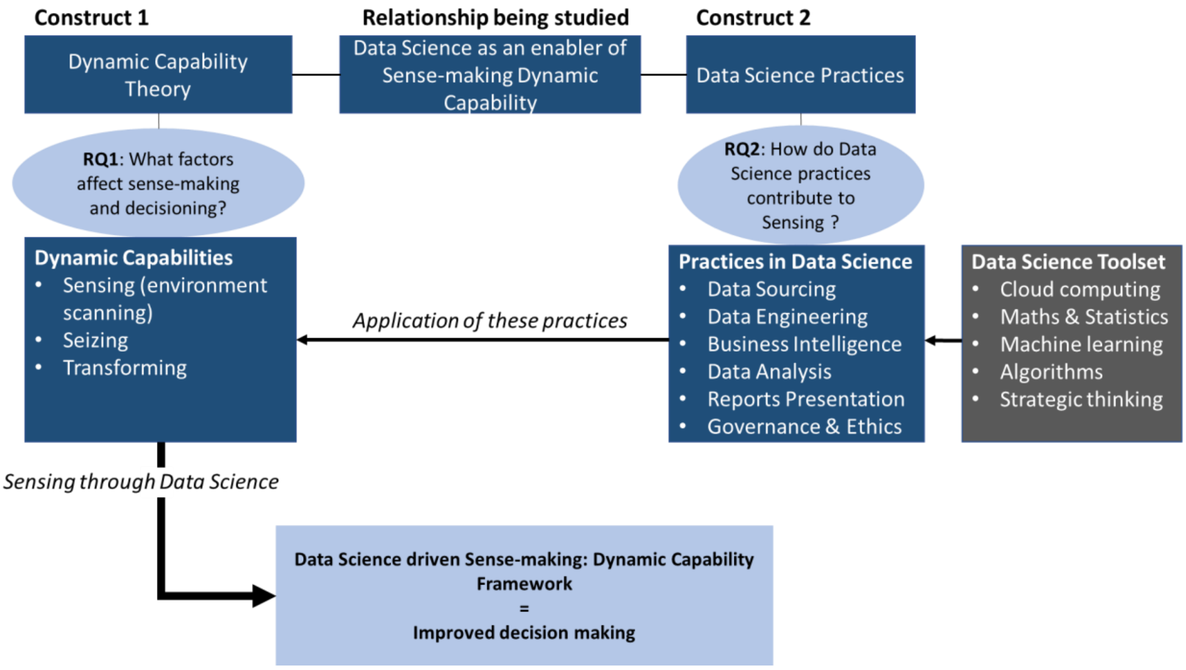


Figure 1: Map of constructs covered in the literature review (developed by the researcher)

2.1 Decision making

The process of decision making is the central tenet of every organisation’s operational and strategic mandate. Organisations live and die by the decisions they make (Colombo and Delmastro, 2008; Van den Steen, 2018). Organisations’ long-term success and competitiveness depend on the quality of the decisions they make. Constructs such as organisational structure design, governance, committees, policies, etcetera are instruments that facilitate decision making (Colombo and Delmastro, 2008).

The design of organisational structures determines the efficiency of decision making. Christensen and Knudsen (2010) found the flat structures facilitate quicker turnaround times in decision making and execution. While hierarchical structures accord better oversight, they compromise the speed of decision making (Christensen and Knudsen, 2010; Van den Steen, 2018). What may be significant is the predictable systemic issues that structural design choices create. Christensen and Knudsen (2010) found that flat design structures tend to reduce Type I errors, while hierarchical designs tend to minimise Type II error decisions. A Type I error occurs when a decision rejects a plausible alternative, and a Type II error occurs when a decision affirms a flawed alternative. Table 1 below summarises Christensen and Knudsen (2010) argument on the impact of structure on decision making. Mackay and Zundel (2017) and argue that leaders need to consider more than just the structure of an organisation; decisions are not only rational, there is often a need to address social and cultural aspects which can be oblivious to structure.

Decision making in a flat organisation	**Decision making in a hierarchical organisation**
• This structure reduces Type I errors • Reduces the likelihood of rejecting a superior strategic option • Actors (e.g. leaders) learn from peers and copy or transfer insights which form the basis of decision making • Exhibit higher risk appetite, creativity and agility • Requires less political stakeholder management.	• This structure reduces Type II errors • Reduces the likelihood of accepting an inferior strategic option • Validation in the Hierarchical structures filters out weak options • Exhibits risk averseness and emphasise self-perseveration, governance and control • Requires extensive social and political considerations to arrive at decisions.

Table 1: Influences of structure on decision making (summarised from Christensen and Knudsen, 2010; Mackay and Zundel, 2017).

Therefore, these views point to a need for leaders to be cognisant of the underlying systemic influences on decisions. Drawing from Christensen and Knudsen (2010) above, it is clear that leaders of organisations with flat structures need to build capabilities that help them reduce Type II errors. Such leaders also enact better governance and controls, which are typically poor as a consequence of the flat structure. Leaders in hierarchical organisations require capabilities that help them reduce Type I errors as well as capabilities that enable creativity and agility often get impeded by the hierarchical design (Christensen and Knudsen, 2010). Still, in both instances, leaders have to pay attention to rational, social, and political demands in the decision-making process (Mackay and Zundel, 2017).

Mackay and Zundel (2017) point of view implies that it is not enough for a decision to be sound rationally; leaders ought to balance the social and political aspects involved in decision making. However, the thrust of this study deals with the specific problem, where leaders are mainly concerned with rational decision making. This study investigates sense-making dynamic capabilities for a rational type of decision making, only. The rest of this document, therefore encompass literature that focuses on rational decisions. This does not negate the importance of balancing political and social consideration in decision making; only that the scope of the thus study does not extend that far.

The next issue in decision making is the inherent human shortcoming in coping with information flux and complexity (Van den Steen, 2018; Rosenhead, Franco, Grint, & Friedland, 2019). Human beings work under a constraint of bounded rationality, which sets an upper limit on the amount of complexity a person can process at any point in time (Robbins and Judge, 2018; Rosenhead, Franco, Grint, & Friedland, 2019). This does not discount the value of intuition in decision making, which has been acknowledged to be valuable in certain instances (Calabretta, Gemser, & Wijnberg, 2016). The purpose is to highlight the presence of the inherent human limit in applying rational thinking. Leaders ought to appreciate this constraint.

The development of a Decision Aiding (DA) framework was a direct response to the problem of bounded rationality in decision making (Tsoukiàs, 2007). DA seeks to remove or minimise ambiguity in the appraisal of decisions given the dynamic environment context. A DA is an abstract framework that can be applied across multiple domains as opposed to an ad-hoc problem-solving effort (Tsoukiàs, 2007; Meinard and Tsoukiàs, 2019). In its operation, DA mobilises human and physical resources to aid the resolution of a specific problem. Since the DA is an abstract concept, its practical manifestation must be in the form of a capability that mitigates leaders' bounded rationality and aid in decision making.

Improved decision making leads to better organisational outcomes (Van den Steen, 2017). However, it was not the purpose of this study to survey all the different categories of organisational outcomes. This study sought to uncover practices and processes that lead to decision support capability. While many types of capabilities can be built to enable better decision making, the thrust of this study was to investigate the utility of sense-making dynamic capability in addressing problems in decision making.

2.2 Dynamic capabilities

Dynamic capabilities can be defined as change processes that redesign, integrate and reconfigure the underlying organisational resources in order to create competitive advantage in dynamic markets (Eisenhardt and Martin, 2000; Kurtmollaiev, 2020). Dynamic capabilities can also be deployed to develop new business models (Eisenhardt and Martin, 2000). This study did not cover the definition of the term capability, but it was essential to differentiate between ordinary capabilities and dynamic capabilities. Winter (2003) and Teece (2014) described ordinary capabilities as those baseline resources and competencies a firm cannot compete without, such as property plant and equipment, production processes, and standard management processes. An organisation can obtain ordinary capabilities by purchasing or building them; while dynamic capabilities can only be built internally (Teece, 2014; Mikalef et al., 2018). Ordinary capabilities are sometimes equated with management routines described in the Resource-Based View of the firm (RBV) theory (Teece, Pisano, and Shuen, 1997). In contrast, dynamic capabilities operate on top of ordinary capabilities; extending, reconfiguring; and even decommissioning them according to the dictates of the prevailing market environment (Eisenhardt and Martin, 2000; Kurtmollaiev, 2020).

Teece et al.'s (1997) seminal paper on the theory of dynamic capabilities to augment and further develop the body of knowledge for the Strategic Management theory. This body of knowledge is dominated by three major theories, namely, industry analysis perspectives (as coded in Porter's Five Forces Framework); Game Theory, and RBV (Teece et al. 1997).

The industry analysis theory focused on factors mainly outside of the organisation. These factors include the industry or the industry group where the organisation operates, the threat of potential entrants, the bargaining power of buyers, the bargaining power of suppliers, the impact of close substitutes, and the intensity of rivalry among incumbent organisations (Porter, 1980). While this framework is still useful, it does not provide direct assistance on how to appraise the organisation's internal factors critical to strategy (Teece et al.,1997).

Teece at al. (1997) further pointed out that Game theory was more suitable for organisations operating in oligopolistic industries and therefore offered little assistance to other settings. However, the RBV came a long way to closing those gaps. The RBV posited that organisations could create competitive advantage by pursuing strategies that fully leveraged their resources' endowments (Wernerfelt, 1984). This particular view is vital as it locates the sources of competitive advantage internal to the organisation (Teece et al.,1997). Therefore, empowering leaders to take effective control of resource at their disposal.

Furthermore, Teece at al. (1997) pointed out that Industry Analysis and RBV are more useful when deployed together than in isolation. However, both are inadequate under complex and unstable environments or dynamic markets (Eisenhardt and Martin, 2000; Teece, 2014). In dynamic markets, existing sources of advantage are disrupted by technology, a turbulent external environment, and imitation by competitors. Therefore, integrative frameworks are more useful, especially as organisations need to focus on building management competencies and intangible assets (capabilities) that cannot be easily copied by the competition (Eisenhardt and Martin, 2000; Teece, 2014). Organisations need to reconfigure themselves and their resources in line with the changing circumstances within their environments.

As such, the dynamic capabilities theory appears to provide the guidance needed in addressing decision-making aspects, which is the underlying problem of this research study. Of particular interest, is the sense-making capability that leaders employ when navigating dynamic environments. Dynamic capabilities entail processes such as Research and Development (R&D), Strategic Decision making, Mergers & Acquisitions, Alliance formation, and Knowledge Creation (Eisenhardt and Martin, 2000). Teece (2018) further illustrated this theory by providing three pillars, namely, Sensing, Seizing, and Transforming that are foundational, as indicated in Figure 1.

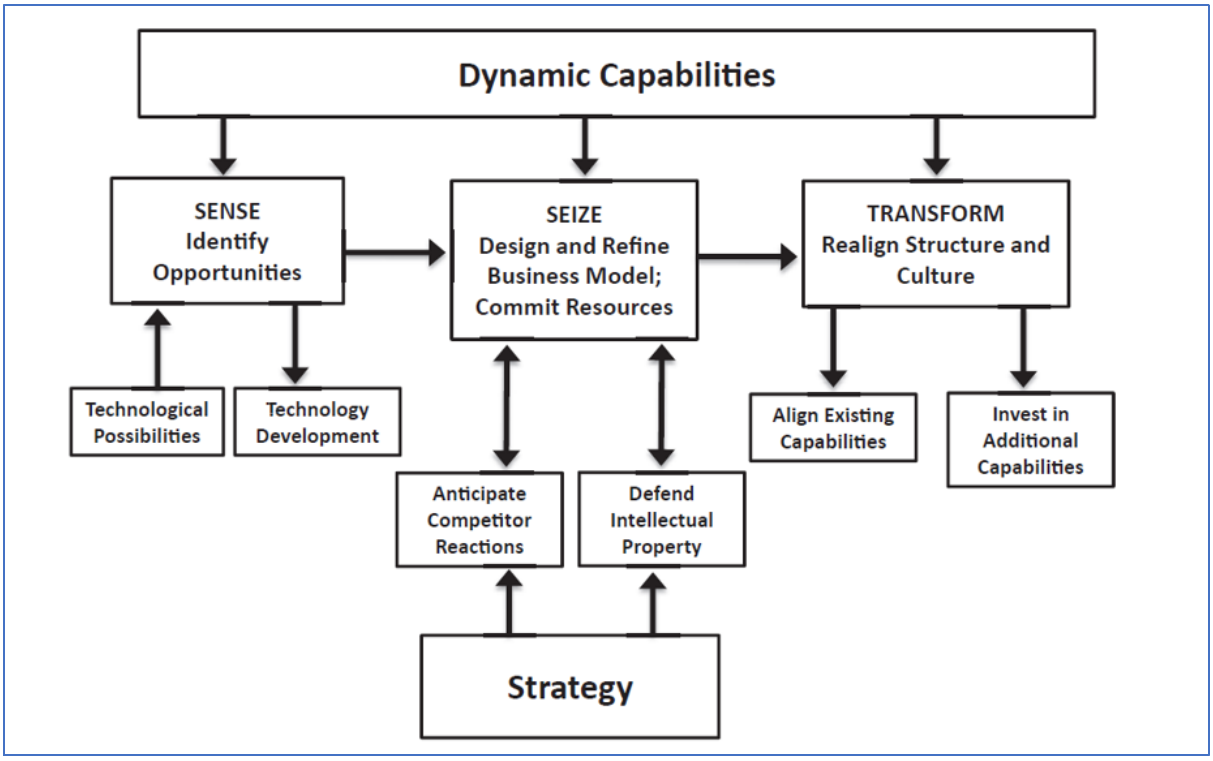


Figure 2: Business models and dynamic capabilities (Teece, 2018)

The ultimate purpose of developing Dynamic capabilities is the attainment and sustenance of competitive advantage. An organisation is said to possess competitive advantage if it controls resources that are valuable, scarce, difficult to copy, and non-substitutable (Eisenhardt and Martin, 2000; Teece, 2014; Teece, 2018; Nayak, Chia, & Canales, 2020). Notably, dynamic capabilities are not underlying resources, but they are management processes that operate on top of resources. Leaders use these processes to redesign, reconfigure, and integrate underlying resources to create a sustainable competitive advantage (Eisenhardt and Martin, 2000; Teece, 2014). Table 2 below provides a summary of critical perspectives on Dynamic capabilities.

	Moderate Market Pace	**High-velocity Market Pace**
Environment characteristics	• Stable to moderate pace environment	• Dynamic, high-velocity environment
Nature of capabilities	• Capabilities resemble the Routines in RBV theory • Capabilities are complicated, with highly detailed analytical rules. Little scope for creativity	• Routines morph into **Dynamic capabilities** • Dynamic capabilities are simple in design with clear boundaries.

	• They are formed in a linear path dependent manner • Highly structured and rigid • Routines are resilient • Source of competitive advantage • Competitive advantage is destroyed by outside forces	Provide management with broad scope for creativity • Formed in an iterative fashion • Flexibly structured for agility • Dynamic capability require maintenance, which is both challenging and energy intensive • They are a source of competitive advantage for a short window period. • Competitive advantage is destroyed by both external forces and internal decline
Management action	• Manager implement best practice to create routines/first-order capabilities • Develop routines over-time and encode them in organisational processes • Ordinary capabilities are either bought or built	• Managers learn by doing and use experiential methods to create Dynamic capabilities • Multiple options are created followed by an empirical selection process that evolves the best ones • Requires the involvement of cross-functional teams • Individual leader strength influences the formation of Dynamic capabilities and which ones are chosen by the firm. The impact of influential leaders by seniority is evident. • Dynamic capabilities can only be built internally
Examples	• Production processes • Sales and marketing • Financial and management accounting processes • Working capital management • Supply chain management • Human resource management	• Research and development • Strategic decision making • Mergers & Acquisitions • Alliancing (forming and managing eco-system networks)

Table 2: Dynamic capabilities: summary of key perspectives (Eisenhardt and Martin, 2000; Teece 2014; Helfat and Martin, 2015; Mikalef, 2018; Kurtmollaiev, 2020).

For Teece et al. (1997), dynamic capabilities create a competitive advantage for organisations that implement them. However, Eisenhardt and Martin (2000) and Kurtmollaiev (2020) argued against this idea. First, they contend that other organisations eventually copy Dynamic capabilities, through unique paths, which ultimately remove the advantage brought about by the scarcity factor. In this view, Dynamic capabilities are temporal in nature. Second, Dynamic capabilities are formidable and energy intensive for organisations to maintain in the long run and therefore, eventually dissipate. This view affirms short-term advantage and refutes sustainable competitive advantage. Therefore, based on both arguments, the utility of Dynamic capability is only in improving the underlying resources, while they last (Eisenhardt and Martin, 2000; Kurtmollaiev, 2020).

Yet, these arguments by Eisenhardt and Martin (2000) and Kurtmollaiev (2020) beg the question as to why any tenacious and innovative organisation should fail to sustain dynamic capabilities. Teece (2014) argues that Dynamic capabilities are evolutionary in nature and can help organisations sustain competitive advantage. Therefore, it is contingent upon an organisation's strategy, whether its Dynamic capabilities produce sustainable advantage or not (Teece, 2014). Knowledge creation and decision making are good examples of Dynamic capabilities that are being perpetually developed. There is a constant inflow of new learning and information from which knowledge is built. Therefore, these capabilities must be continually updated and renewed (Tsoukiàs, 2007; Laaksonen and Peltoniemi, 2018). Decision making is also the same; strategic decisions are continually being made. Therefore, DA processes need to evolve continuously.

Central to this study is the decision-making process, which requires environmental scanning and sense-making capabilities. Therefore, there is a need to scrutinise the sense-making capability in some detail. Sense-making is the process through which an organisation identifies opportunities and threats in a dynamic environment (Teece, 2014). In dynamic environments, opportunities and threats do not present themselves in clear and obvious ways. The pace is fast, and the window of opportunity is brief (Eisenhardt and Martin, 2000). As such, past experience does not count for much. Leaders must learn looking backwards, through experimentation and probing (Snowden and Boone, 2007).

Sense-making provides processes and tools to circumvent the influx of information faced by leaders (Teece, 2014; Helfat and Raubitschek, 2018). This serves to improve decision making. These processes and tools include the use of information technology systems, quantitative data analysis methods, lean, design thinking, and many others (Provost and Fawcett, 2013;

Conboy, Mikalef, Dennehy, & Krogstie, 2020). Also, practices that provide organisational linkages to external data providers, ecosystem partners, and industry experts are key to the development of dynamics capabilities (Eisenhardt and Martin, 2000). These tools and practices are well known and well documented for any organisation to exploit. Yet, not all organisations possess dynamic capabilities. This might be in the manner in which leaders leverage, combine and integrate these tools and practices into their organisations that creates Dynamic capability (Eisenhardt and Martin, 2000; Helfat and Raubitschek, 2018; Kurtmollaiev, 2020).

Girod Stéphane and Whittington (2017), in their longitudinal study spanning 19 years, found that proactive configuration associated with the development of Dynamic capabilities have positive outcomes on organisational performance, while configurations associated with restructuring have adverse long-term consequences. This points to the need for proactive strategies from leaders. Organisations with ineffective strategies risked being forced to make adverse changes to their detriment (Whittington, 2017). This underscores the need for leaders to continually make changes in the underlying resources and capabilities of their organisations. If they fail to do so, circumstances will eventually to force them to make the changes.

It is essential to refocus on the sense-making capability as a lens for this study, and its enablers. To this end, this study evaluates opportunities presented by Data science in the development of this capability.

2.3 Data science as a Decision Aiding process

Many organisations are turning to data for opportunities to unleash business value. However, academic literature on the subject is still at its infancy (Chiang, Grover, Liang, & Zhang, 2018). Much of the content on the topic is provided by professional publications. Therefore, there is a need for the development of academic literature in this field to guide the processes needed to harness the data resource and unlock its business value (Mikalef, Boura, Lekakos, & Krogstie, 2019).

Can organisations deploy to Data science to build decision making dynamic capabilities? In Section 2.1, the concept of a DA process was introduced as a collection of expertise, tools and practices that enable rational decision making (Meinard and Tsoukiàs, 2019). Conboy et al. (2020) defined Data science as a practice of combining data, technology and people to make decisions. Seddon and Currie (2017) described it as the use of data to make sounder, more evidence-based business decisions. Emerging studies have shown that organisations that invested in Data science improved their ability to process vast and complex information presented to them – creating an advantage for themselves (Mikalef et al., 2019). In contrast, other studies found that some organisations faced challenges in realising the value of their investments in Data science (Vidgen, Shaw, & Grant (2017). However, the main obstacle is the lack of a cohesive guiding framework (Mikalef et al., 2019). Given these views, it is clear that Data science has the potential to provide concrete processes and tools that can be leveraged as a DA (an abstract concept), to help leaders improve decision making. However, first, a cohesive framework is required to guide organisations on best practices in achieving this goal.

2.3.1 What does Data science entail?

Having established the potential contribution of Data science in creating Dynamic capabilities, it is essential to unpack its key components. Data science encapsulates the processes, principles, tools, and technologies for analysing data to improve decision making (Provost and Fawcett, 2013; Meinard and Tsoukiàs, 2019). Notably, Data science evolved from historic data practices, and this created an assortment of terminologies that are often used interchangeably. The term Business Analytics or Analytics was popularised first; however, with further improvements and augmentation of practices in the field, it has transformed into the term data science. Therefore, the two terms are often used interchangeably (Schmarzo and Schmarzo, 2013). For this study, Data science is used.

Central to Data science is Big Data. This refers to data that is varied in formats and too large in volume to be handled by traditional techniques. Big data includes forms of media that were historically not considered to be data suitable for automated analysis, including images, voice, videos, etcetera (Provost and Fawcett, 2013; Conboy et al., 2020). The scope of data accessible to Data science is extensive and includes sources inside and outside the organisation. Organisations employing Data science can consist of data from website publications, social media, proprietary, and public databases, textual reports, etcetera (Schmarzo and Schmarzo, 2013). This factor is pertinent to the goals of this research study since leaders often need to make sense of their environment using an assortment of heterogeneous information sources, which can be challenging for a human mind to process.

There are also important characteristics of Big Data that are not present in traditional operational data technologies, namely, Big versus Velocity, Volume, and Variety. Velocity refers to the high speed in which data can be generated and received. For example, data generated from electronic sensors like, weather monitoring devices, or aggregate social media posts. Volume is concerned with the size of the data, which is often far greater than the standard size of data that organisations usually kept in their databases. Variety refers to the fact that data comes in heterogeneous media formats, like video, audio, image and text. This represents a marked difference to traditional data processing technologies, which typically operate on text data only (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2016; Conboy et al., 2020).

Data science techniques implement technologies, algorithms, and mathematical techniques that make enable harnessing and analysing Big Data (Schmarzo and Schmarzo, 2013; Baesens et al., 2016). Big Data combines internal and external data sources and allows this data to be analysed simultaneously (Baesens et al., 2016). While all these forms of data are extracted and often analysed quantitatively, Braun, Kuljanin, & DeShon (2018) points out that qualitative data produced by researchers both internal and external to the organisation must not be excluded from this kind of analysis. The total of these data types forms the base resources from which leaders can build insights. These insights should then enable them to scan both internal and external environment, a key feature of sense-making dynamic capabilities.

The next crucial aspect of Data science is the multidisciplinary and cross-functional configuration of the team (Tonidandel, King, & Cortina, 2018; Lee, Inceoglu, Hauser and Greene, 2020). The point on multidisciplinary composition stems from the fact that Data science teams people together with people with different expertise, like a computer scientist, statisticians, mathematicians, business experts, etcetera (Tonidandel, King, & Cortina, 2018). The data team has to be cross-functional because for the team to function very well, it must establish close vertical and horizontal ties across the business (Lee, Inceoglu, Hauser, & Greene, 2020).

However, the most critical aspect is the business value that organisations can derive from this process. Yet, it seems as if there is no straight forward approach for ensuring successful return and investment on Data science projects (Newman et al., 2016; Chiang et al., 2018). It would be beneficial if Data science practices were crystallised into a specific methodology that can be reused by any organisation wishing to implement or improve its Data science capability. However, Hindle and Vidgen (2018) found that while it is common practice for Information Technology disciplines to follow a well-defined methodology, Data science does not have a clear and unifying methodology. While best practices were emerging and beginning to be shared across practitioners, there is still no universally accepted methodology that organisations can use to build this capability (Hindle and Vidgen, 2018; Lekakos and Krogstie, 2019). As a result, a theoretical framework is critical that will guide the efforts for the development of a standard methodology for implementing Data science capabilities in an organisation.

2.3.2 Emerging standard practices in data science

Despite the shortcomings mentioned above, some standard practices have started to emerge, which can be sequenced into distinct stages. First, there would be collaboration on the nature of a business problem(s) that the organisation would like to address. These are usually distilled into concreate research problems, research questions, or hypotheses (Hindle and Vidgen, 2018). The second stage involves collaboration among cross-functional teams or brainstorming within the Data science team. Data science efforts often require a diversity of skills involving programmers who write the software; statisticians or mathematicians who develop and validate the data models; and business stakeholders who serve as subject matter experts (Hindle and Vidgen, 2018).

Stage three involves identifying relevant data sources that can address the research questions. Unlike in traditional reporting teams, this exercise is not limited to data that the organisation possesses within its databases. The team identifies and addresses far-ranging data sets of structured and unstructured data (Schmarzo and Schmarzo, 2013). Data science can bring together data sets like trade data from the stock exchange, demographic data from institutions like Statistics South Africa, data from social media such as Twitter, data from online newspaper publications - all under one platform. Significant opportunities for insights are presented when skilled professionals mine such data with clear business purpose. Once this data has been gathered for the first time, it gets refreshed over time, with business insights are kept active (Hindle and Vidgen, 2018). Underlying mathematical models are also reused until their assumptions are no longer valid (Hindle and Vidgen, 2018).

The fourth stage involves applying Machine Learning (a branch of Statistics and Artificial Intelligence), which involves automated statistical models implemented through highly efficient algorithms, to discover patterns and insights from the Big Data (Hindle and Vidgen, 2018; Conboy et al., 2020).

The final, and perhaps the most critical step, is usually referred to as Visualisation. In this stage, the patterns uncovered from the analysis of data are presented in visual forms and presented to management. Skilful practitioners can convert insights into stories, and these are given to management to enrich their insights and improve decision making (Hindle and Vidgen, 2018).

2.3.3 Data science products

Data analytics is the essence of Data science outputs and the final product. Analytics enable competitive advantage as it aggregates complex information and recommends the best course of action (Newman et al., 2016). There are three different analytical products produced by data science, namely, descriptive analytics, predictive analytics, and prescriptive analytics (Hindle and Vidgen, 2018; Conboy et al., 2020). Descriptive analytics classify and categorise things, like placing customers into distinct segments. Predictive analytics, as the most advanced form of analytic, provide forecasts of future trends that help guide leaders concerning options that can be considered for the future. Prescriptive analytics guide leaders on actions to be taken. Prescriptive analytics can offer pointed suggestions such as what products must be provided to which customers and which markets (Newman et al., 2016; Hindle and Vidgen, 2018).

These analytics processes are informed by impersonal process and objective process. Such products assist decision-makers in arriving at precise decisions quicker. These products have the ability to process and provides data intelligence in real-time. These capabilities make Data science attractive as DA tool and the basis for sense-making dynamic capability. However, there are non-technical governance and ethical considerations that are important when developing Data science capabilities.

2.3.4 Governance and ethical considerations in Data science

Leadership has a responsibility to ensure that ethical practice permeates every aspects of their organisation (Zeni, Buckley, Mumford, & Griffith, 2016). Given this view, the ethical issues pertaining to data are therefore not an isolated issue concerning the Data team alone. The data ethical considerations are an important part of the theoretical framework for creating Data science capabilities (Baesens et al., 2016; Zeni et al., 2016). Proper data governance requires practitioners to be transparent with statistical methods used, the variables selected, and the eventual model must be clearly explained so that others may be able to review it (Baesens et al., 2016; George and Osinga, 2016). Peer review by colleagues or associates is important to reduce errors in the final products (George and Osinga, 2016). It is imperative for organisations to form data governance committees with a mandate for addressing compliance issues in data sourcing and usage (Baesens et al., 2016). Critical care must be taken around privacy issues whenever personal information is used (Tadewald, 2019).

Tadewald (2019) proposed a test that organisations should apply to assure adherence to ethical standards, this test consists of three checks:

- Verify that the data collected, and its usage is in line with what the customer expects. For example, when the customer fills in a website form, the website should not collect more information from the customer beyond what the customer provided without consent. Some websites would collect additional information such as the type and make of the device the customer is using. Some organisations have developed business models on selling customer data without the customer knowledge (Tadewald, 2019).
- Ensure the data usage conforms to legal statutes. The organisation should ensure that data protection and privacy laws are adhered to in all countries where it operates (Tadewald, 2019).
- Conduct a self-introspection test. Decision makers ought to reflect on how misuse of their personal data could harm them personally. They must then use this reflection to guide them on how they use their customers' data (Tadewald, 2019).

Any DA process must be subjected to validity checks (Meinard and Tsoukiàs, 2019). This is necessary because without this, the DA could lead to Type I and Type II errors. As such, tools in Data science must be subjected to rigorous validations and the assumptions in which they are based upon tested on an ongoing basis (Meinard and Tsoukiàs, 2019). Equally important, is the Data science ethical behaviour. The team must obtain permission to use the data sources they intend to incorporate. When scrapping internet web pages, they must ensure explicit permissions are obtained and intention of use disclosed (George and Osinga, 2016).

It is equally important to inform the management within the organisation about the sources of data used to obtain information. Ideally, compliance and legal teams within the organisation must be involved to provide advice and guide the Data science team on ethical and legal matters (George and Osinga, 2016). The majority of practices reviewed as part of this research project do not address ethical issues, which may be a critical aspect of Data science practices.

Finally, security considerations are important and must be planned and implemented appropriately. Security can be improved by proactive penetration testing. This involves the organisation simulating attacks on its own infrastructure, software, and data in order to ensure that the security is effective and adequate (Newman et al., 2016). Newman et al. (2016)

reported that 50% of the large organisations they surveyed in 2015 indicated that they spent in excess of R6.6 million a year on data cyber protection measures. This points to the crucial importance of data security and the need to include this aspect in the Data science capability framework.

2.4 Conclusion of the literature

This chapter linked together three theories, namely, decision making, dynamic capabilities, and data science. In addressing the formulated research problem, a theoretical review of these three key areas was undertaken, which assisted in gaining the theoretical foundation of the research problem.

The literature reviewed showed that there is a growing interest in the subject of Data science and its application in creating competitive advantage for organisations. However, more needs to be done to link investment on these tools to business value (Chiang et al., 2018; Mikalef et al., 2019). This study undertook to develop the framework through which Data science delivers business value.

Notably, academic journals continue to issue invitations for research papers on the subject of data science, especially focused on the creation of a unified theory for practice (Vaara, Paruchuri, Nadkarni, & Shaw, 2019). This points to a future need to develop literature on Data science and its value proposition as a management practice.

3. Chapter 3 - Research questions

This research project aimed to develop a theoretical framework that defines Data science processes and how these can support the development of dynamic capabilities, and the sense-making capability in particular. The following research questions accrued:

3.1 Research Question 1: What are the factors affecting sense-making and strategic decision making in a dynamic environment?

To improve decision making, leaders need to make sense of opportunities and risks within their operating environment. Sense-making dynamic capabilities provide a mechanism to navigate information influx and equip leaders to make better decisions (Eisenhardt and Martin, 2000; Kurtmollaiev, 2020).

This question enabled the study to focus on critical factors to sense-making and decision making in the context of dynamic markets. This study's principled position is that making decisions in dynamic markets is more challenging than less dynamic markets. Therefore, the best contribution in building the sense-making Dynamic capabilities is within the context of dynamic markets.

3.2 Research question 2: How do Data science practices contribute to the creation of sensing capability and competitive advantage within an organisation?

It is important to emphasise that Data science is not just a technology, but a set of practices that involve people, processes, and tools to achieve specific outcomes (Mikalef et al., 2019). Effective Data science practices bring cross-functional teams into collaboration (Mikalef et al., 2019). In the same way, cross-functional collaboration is required in building Dynamic capabilities (Teece; 2014; Kurtmollaiev, 2020).

This research question sought to establish the existence of a distinct set of practices in Data science that qualify the practice as an enabler of Dynamic capability in line with academic literature. Of particular focus is the Sense-making Dynamic capability in which Mikalef et al. (2019) contended could be enabled by Data science. This research question, therefore, sought to establish if these practices are generalisable, beyond the individual, organisational details, such that they can constitute a broadly applicable framework. Such a framework would then provide guidance to any organisation on how Data science can be leveraged to form Dynamic capabilities for sense-making.

The two research questions facilitated data collection and analysis. The ultimate goal was to address the stated academic and business problem dealt with in Chapter 1 as well to propose a framework which will address the stated problem – this covered in Chapter 6.

3.3 Research question 3: What are the ethical considerations emanating from collecting and using data for Data science applications?

Customer and external data provide real strategic opportunities for organisations; however, the scope of possible opportunities are limited by ethical boundaries that demarcate acceptable use (Baesens et al., 2016; George and Osinga, 2016). According to George and Osinga (2016), organisations who cross these boundaries risk breaching local of laws and brand erosion. It is, therefore, essential for organisations to be cognisant of the ethical issues attended to the use of data.

The aim of this question was to elicit information about the practices and controls that organisations put in place in order to govern the use of the data in their disposal. Based on the literature reviewed above, it is reasonable to expect that data ethics governance should be foundational in the development of Data science driven capabilities. The practices that were elicited were included in the framework for Data science driven capabilities; which was developed as part of this research project. As such, this framework will enable the user to build-in governance and ethical controls into their Data science driven dynamic capacities.

4. Chapter 4 - Research methodology

4.1 Methodological choice and design

This study employed mono qualitative methodology. This methodology was chosen for its suitability for studies that explore an area where theory is still emerging (Saunders and Lewis, 2018*)*. The research into the Data science area of work is still emerging and thus justifies the use of this methodology (Chiang et al., 2018). The study was inductive and explored the relationship between Data science and the dynamic capabilities theory, with specific on sense-making dynamic capability.

This study followed a phenomenological design strategy. Phenomenology is the study of how things work. Phenomenology involves making implicit patterns and routines explicit (Sanders, 1982). In this study, this was achieved through an analysis of the reviewed literature combined with data collected from 13 experts in the field of leadership, strategy and data science. Teece (2007) observed that dynamic capabilities are underscored by core-competencies whose formation is not obvious, often even to organisations to whom they accrue. In this study, the phenomenon of interest was the development of the sense-making capability driven by Data science practices. As such, the phenomenological design enabled the study to explicate the processes that drive the creation and development of this capability.

The study followed an inductive approach to data collection and analysis, which entailed the gathering, coding, categorisation, identifying themes, outlining relationships, and eliciting meaning. This process is intended to culminate into the development of a theoretical framework (Myres, 2020a). This choice of approach was suitable for this study as, based on qualitative data, the inductive approach has the ability to uncover nuanced insights and contribute to the existing literature on the subject (Bansal, Smith, & Vaara, 2018).

The process of data collection deployed semi-structured interviews. The semi-structured interview technique is useful for generating data for theory development. The technique also enabled flexibility in the flow of the interview process. The usefulness of this approach is in enabling authentic and perspective of the participant to be recorded (Pratima and Corley, 2011). Since interview participants were leaders with diverse roles and experiences in their respective fields of strategy and data management, some of the questions were modulated so that emphasis was placed on aspects that the participant has provided their poignant responses. However, most participants provided views on questions not related to their current professional disciplines whenever they felt they had an informed opinion.

Finally, this study was a cross-sectional research. This type of research is suitable for studies performed under limited timeframes (Saunders and Lewis, 2018). The research was constrained for time, where completion of the whole study was within a period of less than six months.

4.2 Population

The population consisted of individuals in leadership or senior management roles, with strategic decision-making responsibilities within their organisations. Janice, Michael, Maria, Karin, & Jude (2002) emphasised the importance of targeting participants with sufficient knowledge of the subject area in order to improve the rigour of the qualitative research outcome. The target population were individuals who are leaders of business units, departments, or divisions. Leaders at this level of management are ideal for this study as they assume responsibility for both strategic and operational direction of their organisations. As such, these leaders turned out to be the main consumers of Data science reports. The study focused on South Africa; however, it could have general applicability.

4.3 Unit of analysis

The unit of analysis for this study is an organisation where the development of Dynamic capabilities serves a strategic purpose. The typical profile consists of organisations which compete in high velocity dynamic markets as described in Chapter 2. As such, this profile would exclude some of the non-profit organisations where fierce competition among peers is not a defining factor of their strategy. This profile would include profit-making organisations operating in dynamic markets.

However, the unit of sampling was individual leaders who have strategic decision-making responsibilities within such organisations. Data was collected from individual leaders, in their personal capacity, to elicit their views on the subject of this research.

4.4 Sampling method and size

This study applied the purposive sampling method. This method allows the researcher a level of judgement on whom to include in the sample (Saunders and Lewis, 2018). In this study, only individuals who lead business units and have strategic decision-making responsibilities were included in the sample. These individuals hold titles such as Head, Executive, Director, Senior Manager, etcetera. A pre-screening selection criterion was included in the data collection tool to ensure that only qualifying individuals were included. The list below provides further details of the criteria that was used for selecting participants:

Senior managers:

- Senior manager level role or higher, with strategic decision-making responsibilities for a business unit or at the level of the entire organisation;
- Belong to an organisation with a Data science or business intelligence (BI) department that produces Data science reports. Some organisations expand the scope of an existing BI unit to include Data science capabilities instead of creating a separate Data science unit, that is why BI was still relevant as part of the criteria;
- Consumers of Data science reports. This is well suited to address interview questions related to Research Question 2.

Data science managers:

- Leaders of Data science or BI units;
- Producers of the Data science reports. This is well suited to address interview questions related to Research Question 1.

The sample size for the study was estimated to be 15 participants; however, saturation was reached on the 10^{th} interview. The saturation point is defined as reached when the researcher finds that additional interviews do not generate new information (Jonsen, Fendt, & Point Sébastien, 2018; Saunders and Lewis, 2018). However, the researcher decided to continue to a total of 13 participants at which point there was a significantly low number of new codes generated. At that point, the researcher concluded that a saturation point was reached.

Finally, the initial set of participants was sourced from the researcher's network of contacts, who meet the profile mentioned above. The project made use of snowball sampling (Saunders and Lewis, 2018); where most subsequent participants were referrals from the first set of participants.

4.5 Measurement instrument

A semi-structured interview questionnaire was used to record all interview questions - see Appendix A. the reviewed literature guided the development of the questionnaire. Jacob and found that good questionnaires are structured in an open-ended manner. This encourages participants to supply as much information as possible on each question (Furgerson, 2012; Jonsen, Fendt, & Point Sébastien, 2018). The measurement instrument was designed with this principle in mind, and that a few easy questions must appear at the beginning (Myres, 2020a). This approach built rapport and eased participants into the conversation.

Semi-structured interviews provide flexibility to the research to adopt the instrument as the situation demand (Fendt and Point Sébastien, 2018). The research also followed the same approach; some questions were slightly adjusted to allow for flexibility and extracting maximum value from the participant. As the interviews process progressed, the number of questions were reduced from twelve to ten. This enabled the researcher to improve subsequent interviews.

All interviews were conducted through an online tool - Microsoft Teams, due to COVID-19 restrictions and protocols. The Microsoft Teams recording functionality was used for recording purposes. All recordings were made with the consent of each participant.

4.6 Data gathering process

Upon approval of Ethical Clearance, a formal letter was sent to each participant, requesting their involvement in the research and outlining research objectives. After that, meetings were scheduled, and interviews conducted. Most of the 13 interviews took one hour to complete, but a few took longer. This happened whenever the participant indicated interest and availability to continue exploring the issues being discussed. Most participants honoured their commitments; only two participants could not act in line with their commitment.

A semi-structured questionnaire was used to conduct interviews. In qualitative studies, researchers iteratively adjust research questions, as they collect and analyse data from participants (Janice, 2002; Bansal et al., 2018). In this study, research questions were adjusted as the exploratory process generated more information about dynamic capabilities and practices in data science. The literature review was also refined and expanded at several points to broaden the scope based data being collected.

While the data collected from the interviews formed the key part of this study, additional data sources were included for the purpose of comparative analysis. A comparative analysis is a triangulation technique used to improve the validity of research findings (Bansal et al., 2018; Snyder, 2019). As a means of triangulation, this study used three real-life Data science projects that have been completed and analysed. The Kaggle platform was used for this purpose; it is accessible at no cost through this portal: https://www.kaggle.com/. The analysis of real-life Data science projects serves to enhance this study by illustrating typical Data science outputs and analysing the contribution of these projects to decision making.

4.7 Analysis approach

This study used an inductive approach, starting collected data to inform the theoretical framework (Bansal et al., 2018). The process was as follows: conducting interviews, record and store interviews, transcribe and capture interview data into text documents, and then load text transcripts into Atlas.ti software for coding and analysis.

The analysis of transcripts produced codes; which are succinct phrases that summarise the key points in the data (Skjott Linneberg and Korsgaard, 2019). New codes were added as each interview transcripts were analysed, but become fewer as the process progressed. On the tenth interview, fewer codes were generated, and saturation was reached. There were many codes there were consolidated into a single code as they represented the same idea, yet few codes were split to capture different facets of an idea. Subsequently, the qualitative interpretation of codes produced new categories and sub-categories. These categories were converted into broad themes for analysis.

The emergent themes were subjected to analysis and interpretation, leading to a new classification (Morse, Barrett, Mayan, Olson, & Spiers, 2002; Hsieh and Shannon, 2005). This process was followed by associating and conceptualising these themes to reveal meaningful relationships (Skjott Linneberg and Korsgaard, 2019; Myres, 2020b). This process culminated into the Data science driven framework for sense-making presented in Chapter 8, addressing the three research questions.

4.8 Quality controls

Quality controls help establish the validity and reliability of the research output (Janice et al., 2002). These controls ensure that the study's conjectures are trustworthy and truthful

(Roulston, 2010). The first quality control in the execution of the study was the pilot conducted with two participants in order to test the data collection tool. The two pilot participants are managers with knowledge of strategy processes and the field of data science. The data collected from pilot participants were not included in the results of this study. Jacob and Furgerson (2012) found that conducting pilot tests with peers improves the design of the data collection tool and thus, the outcomes of the study. This was the case in this study, as well as to psychologically prepare the researcher ahead of actual interviews. Following this, interview questions were revised, and their sequence was used to address some of the emergent issues from the pilot process.

Roulston (2010) highlighted several quality issues that negatively impact on the validity of the research output. These are listed in bullet points below together with the mitigation measures that were used in this study:

- Roulston (2010) stated that in an interview setting, participants might misrepresent what they actually do in practice. According to Janice et al. (2002), the researcher ought to address the problems of inconsistency in data by paying full attention to inconsistencies and using the subsequent data collection points to verify data already collected. In this research, careful attention was given to any inconsistencies and the researcher iterated between literature review and data collection to verify data as it came along. The sample profile was also carefully designed to include only people who have enough authority on the subject of Strategic decision making and who have sufficient exposure to Data science. In addition, outlying responses provided were highlighted and discussed as such in Chapter 6 - Discussion of results.
- The other quality issue is that a researcher may bring their subjectivity and biases in the manner questions are sequenced (Roulston, 2010). The questionnaire was carefully designed to elicit objective views of the participants based on their experience and knowledge. The researcher endeavoured to maintain objectivity throughout the data collection and analysis phase.
- The participants must be provided with an opportunity to review the written transcripts of the interview to limit unethical behaviour where the researcher may inject false information or own biases (Roulston, 2010). In line with this principle, participants in this principle, all participants in the study were promised access to their transcripts as well as the final report.

All data collected, including interview recordings and transcripts, have been safely kept on a secured online file drive and will remain there for a period of ten years. At the end of the project, the files were copied into an electronic media device and handed in with the final report.

5. Chapter 5 – Results

This study included 16 data points, consisting of 13 interviews with participants selected through purposive sampling technique, as well as three Kaggle project reports used as part of the triangulation technique. The personal details of the participants are kept confidential, both the names and the organisations they are associated with. This report refers to each of the participants as Participant 1 to Participant 13. Organisations are described only in terms of the industries they belong to, and where relevant, comments on the relative size of the organisations are provided. Table 4 provides summary information on each participant.

The detailed process which was used to produce the result is covered in the preceding methodology chapter. However, in summary, each participant was interviewed, in an effort to solicit their expertise which helped answer the research questions of this study. The interviews were transcribed into a text document and later used in an inductive analysis process. The study initially generated close to 300 codes; however, on further analysis, these were consolidated into 197 unique codes. The codes were grouped into 32 categories; from these categories, nine themes emerged.

Research questions	**Themes**
Research Question 1: What are the factors affecting sense-making and strategic decision making in a dynamic environment?	1. Strategy design considerations 2. Decision making 3. The role of leadership
Research question 2: How do Data science practices contribute to the creation of sensing capability and competitive advantage within an organisation?	4. Crafting Data science capability 5. Data science processes and practices 6. Data science skills 7. Data science technology
Research question 3: What are the ethical considerations emanating from collecting and using data for Data science applications?	8. Data Ethics and Governance 9. Data legislations

Table 3 shows the mapping between the three research questions and the nine themes that emerged as part of data analysis. The mapping of themes to categories and frequency of codes is provided in Appendix B. Finally, Appendix C provides the comprehensive list of codes, mapped to categories and to themes.

Research questions	Themes
Research Question 1: What are the factors affecting sense-making and strategic decision making in a dynamic environment?	1. Strategy design considerations 2. Decision making 3. The role of leadership
Research question 2: How do Data science practices contribute to the creation of sensing capability and competitive advantage within an organisation?	4. Crafting Data science capability 5. Data science processes and practices 6. Data science skills 7. Data science technology
Research question 3: What are the ethical considerations emanating from collecting and using data for Data science applications?	8. Data Ethics and Governance 9. Data legislations

Table 3: Research questions mapping into themes

5.1 Research participants

This section discusses the key attributes of the 13 participants who were interviewed as part of this study. The actual names of these individuals are kept confidential, and they will only be referred to as participant 1 to participant 13 throughout this report.

Participant	Gender	Industry	Details
Participant 1	Male	Energy and mining	Participant 1 is a Continuous Improvement senior specialist in a mining organisation. His role entails the use of data to guide the organisation in the implementation effort. His role is the intersection of decision making and data utilisation, which was an important area of inquiry for this study.
Participant 2	Male	Financial services	Participant 2 is the Head of Data function at a large financial services organisation. He interacts with both the technical team and strategic decision-makers as he is the key supplier of analytical reports. In his previous role he has implemented Data science capability from the ground up. He was selected due to his holistic view of data

			management and his experience in interacting with senior leaders.
Participant 3	Male	Consulting	Participant 3 is a senior executive for a South African international consultancy that specialises in Data. He is based in the United Kingdom where he is the country director for this organisation. He rose through the ranks from the role of a developer, architect and now a country leader for his organisation. He has vast experience assisting organisation implement Data science capabilities.
Participant 4	Male	Financial services	Participant 4 is a senior executive at the Financial services organisation. He is responsible for a division, include profit and losses (P&L). He has had lead engineering and IT consulting organisations prior to joining this large financial service organisation. He is a strategic decision-maker for his division. He has extensive experience managing and interacting with Data teams throughout his experience.
Participant 5	Female	Media and entertainment	Participant 5 is an executive responsible for business optimisation in a media and entertainment organisation. She has extensive experience in strategy formulation and execution. Her career spans working in multiple countries in Asia, America, and South Africa.
Participant 6	Male	Consulting	Participant 6 is a consulting data scientist and has been in this role for the past eight years when the field was nascent in South Africa. Participant 6 has implemented dozens of Data science projects for organisations across South Africa. He has deep knowledge of the subject area and

			often speaks in conferences on Data science and artificial intelligence.
Participant 7	Male	Financial services	Participant 7 is a senior manager responsible for a newly established Data science division within a large financial services organisation. Having established a Data science unit recently, his input was important in answering research questions in this study.
Participant 8	Male	Publisher	Participant 8 is a senior executive at a publishing organisation. His responsibilities include leading digital transformation and new product innovation. His career date back from other senior roles in previous organisations where he took charge of IT portfolio across the African continent. He has a strategy formulation and execution responsibility. He also has experience implementing traditional data warehouses.
Participant 9	Female	Telecommunications	Participant 9 is an executive at a large telecommunications company. She is leading a division and P&L responsibilities. She is responsible for her divisional strategy formulation and execution.
Participant 10	Male	Consulting	Participant 10 is a strategy consultant who owns and manages his own business. Prior to running his business, he had senior roles incorporates including the role of a divisional CFO in a financial services organisation.
Participant 11	Female	Tourism	Participant 11 has extensive experience as a head of Marketing and Strategy in a tourism organisation. Her role made extensive use of data and research reports to generate travel demand and improve conversion ratios on marketing spend.

Participant 12	Male	Media and entertainment	Participant 12 is a CIO and CDO for a media and entertainment organisation. He is the accountable executive for supplying strategic information. As an executive, he is a joint strategic decision-maker and interactions with executives on a matter of strategy and technology.
Participant 13	Male	Financial services	Participant 13 is the Head of Department responsible for project delivery in a financial services organisation. He is responsible for running project portfolios that span his organisation, and some of the projects his department has managed include data projects. Before this, he was held general management positions in several companies. Participant 13 was brought into the sample because many participants quoted his organisation as a model for best Data science projects. The researcher than worked on securing a leader in this organisation so that more lights could be shed on specific practices in this organisation. Fortunately, Participant 13 was identified as the relevant candidate for this study, and he agreed to the interview.

Table 4: Participants interviewed as part of this study

5.2 Triangulation data using the Kaggle projects reports

This study includes data collected from Kaggle, a public online platform for Data science projects. The platform connects organisations and data scientist. Organisations submit Data science projects, posed as challenges with specific requirements. The community of data scientists would work on these challenges and publish the results of their findings at the end. The projects are run as competitions and allow for any number of data scientists to attempt any posed challenge. Both the requirements submitted by each organisation and the corresponding solutions produced by the data scientists are openly published on the Kaggle website.

The inclusion of these projects allows for triangulation data. Triangulation is a comparative analysis method whose purpose is to improve the validity of findings (Saunders and Lewis, 2018; Snyder, 2019). There projects that were selected are listed in Table 5 below. The contents of the Kaggle projects were then analysed inductively in a similar manner as the interview transcripts. The Kaggle project reports were coded alongside the interviews of the 13 participants. Codes generated in the process are included in the same codebook as the codes generated from the interviews. The categories and themes that emerged drew from both interviews and Kaggle projects data. Finally, the researcher only selected three projects only as they were deemed to provide enough practical case studies. Adding more projects was judged by the researcher to would have resulted in little marginal benefits.

Kaggle project 1	Title: Inference on winning the Ford Stay Alert competition Highlight: this project uses sensor data to check driver alertness for Ford motor company. This information could be used to warn drivers that are in danger of falling asleep while driving.	Link: https://medium.com/kaggle-blog/inference-on-winning-the-ford-stay-alert-competition-d35f40961aee
Kaggle project 2	Title: Phil Brierley on winning tourism forecasting part two. Highlight: This project employs Data science to improve predict tourist demand for a car rental company.	Link: https://medium.com/kaggle-blog/phil-brierley-on-winning-tourism-forecasting-part-two-5aaa91b93e06
Kaggle project 3	Title: Gaining a sense of control over the COVID-19 pandemic. Highlight: The project uses Data science model to improve search of COVID 19 related articles.	Link: https://medium.com/kaggle-blog/gaining-a-sense-of-control-over-the-covid-19-pandemic-a-winners-interview-with-daniel-wolffram-1320fb2717c4

Table 5: List of Kaggle projects selected for triangulation purposes

5.3 Results for research questions one

Research Question 1: What are the factors affecting sense-making and strategic decision making in a dynamic environment?

In connection with research question 1, participants were asked questions about: their leadership backgrounds; their views on strategy design and implementation processes; inputs they consider pertinent in the strategy design process. They were also asked questions related to the decision-making process as well as the development of competitive advantage. Those participants who indicated that their organisations are operating in complex and dynamic macro environments were requested to keep that contest in mind as the address questions posed to them.

The results are presented in the following sections and organised according to the following themes: strategy design considerations; the role of leadership; and decision making – including the role of data in making decisions.

5.3.1 Environment analysis and strategy design

Creating alignment

Most participants stated that creating or enhancing dynamic capabilities must be part of the strategy design and implementation processes. Decisions on which capabilities the organisation must create are made at the top level of the organisation, beginning with the board of directors and filtering down to the executive leaders. Leaders must then align the rest of the organisation to the vision, the mission and the chosen strategy. Participant 10 stressed that "*strategy formulation is by all intents and purposes decision making, the resulting document that is called strategy is effectively a set of decisions. The strategy is decisions on what the company will do and what it will not do*."

Participant 8 emphasised several times that organisations ought to know their identity. According to Participant 8, identity is the core of an organisation's reason for existence, and everything that the organisation does must stem from there. This view is supported by most participants, as they also stated that the choice of capabilities must be driven by the vision. However, Participant 4 and Participant 9 are of the view that organisations need to change over time, and therefore their identities were not cast in stone but adaptable; this is especially so in a dynamic environment.

Most participants highlighted that the strategy design process moves from casting of a clear vision to the setting of specific goals and objectives. The strategy is then expressed in terms of goals and objectives which are expressed in a manner that is specific and measurable. This provides means for tracking the success and failure of the strategy during its implementation.

However, for Participant 8, this process always misses a critical step, which should be the first step after the vision has been clarified:

> "*Firstly, you start by defining your problem, and really articulate it clearly in terms of observations; what are the things that are working for this company? What are the things that are making this company trip and fall? And then once you have a good understanding the environmental issues at hand, then you have a point where you can work from. At this point, you will have a clear problem statement. You will now derive the goals and objectives which talk to the resources and capabilities you need in order to overcome the problems you have articulated and achieve the vision."*

Participant 8 further warned about the tendency of brainstorming goals and objectives in a manner that is detached from the real problems at hand. Most participants echo the view that the organisations' vision is an important artefact which aligns the organisation with its future aspirations. Participants said strategy then becomes the map towards the set vision.

Analytical problem-solving

Participants indicated that they apply tools and framework for problem-solving; this allows them to elicit strategic options comprehensively. However, Participant 8 criticised what he called the *mechanical* use of frameworks as a practice that was too common and ineffective. Participant 8 is over the view that the outcome of a problem-solving process and a leaders' immersion into the problem space must precede the selection of these frameworks.

However, based on the views of the majority of the participants, even the process of problem-solving requires a disciplined approach which is facilitated by the use of frameworks. A few participants mentioned the Five Whys framework as a useful framework for problem diagnosis. Perhaps the unifying point is that participant agreed that frameworks should be used for what they are suitable for, instead of being used blindly.

For example, Participant 10 said the SWOT analysis is a useful tool when used correctly; while the Porter's 5 forces and the Six forces variant are helpful for industry analysis. He added that "*you will be hard-pressed to find a framework better the PESTEL for external environment analysis*." In essence, Participant 10 felt that organisations must be pragmatic in their choice of tools and avoid box-ticking exercises. He further mentioned that while there is value in the use of consultants as facilitators of a strategic process, executive leaders must be vigilant not to outsource the thinking process to consultants.

The full list of frameworks that participants mentioned is as follows: PESTEL; SWOT; Five Whys; Jobs to be done; Porter five forces; Porters six forces; Systems thinking; and 5Cs framework. There is a consensus that frameworks provide structure in the strategy design process and decision making.

Environment dynamism

The results show that organisations were engaged in continuous scanning of their external dynamic environment. Five of the participants specifically mentioned that the biggest challenge is that the environment they find themselves was always in a state of flux. This made it difficult for them to plan for the long term. Increasingly business plans have to updated in quick successions to accommodate the dynamic environment changes.

Most participants highlighted the importance for leaders to have sufficient knowledge of their operating environment in order to inform their thinking and decisions. Participants say this involves the ability and the need for leaders to form business ties with ecosystem partners external to their organisation. It consists in scanning the environment to make sense of opportunities and threats so that the organisation can stay ahead of its competitors. Three participants, who are all in the Financial Services industry mentioned that they are on high alert for threats coming from both established and start-up organisations; often originating from outside the Financial Services industry. The three participants noted that these organisations usually apply different and novel business models. Participant 4 pointed to the blurring industry borders and emphasised the need to scan the environment for threats emanating from outside the organisation's industry.

"*Your strategy for differentiation is closely tied to that of understanding the competitive landscape in which you operate as an organisation. What has become clear in recent times is that, as we define that competitive landscape, we actually need to even look beyond what would have traditionally been your typical competitors, right? So today I work in the banking sector, but some of the biggest competitive threats would not necessarily come from a traditional bank, right? So you are most likely to be disrupted by organisations that come to a different industry."* (Participant 4)

"*it is hard enough to police the threats in the external environment, it was much harder to match the offerings presented by new entrants because their legacy systems hamstring financial Services incumbents*." Participant 3, a South African based in London made a similar observation stating that "*here we see traditional banks being disseminated by start-up Fin-tech companies.*" (Participant 7)

Participants felt that while the job of sensing competition and scanning external environment is critical, it is hard to carry out. There is an increasing realisation that the ability to do is essential to the success and even survival of their organisations. Most participants indicated that their organisations are increasingly using advanced analytics, commissioned research, and commercial publications to improve sense-making. All participants indicated that their organisations have internal Data functional unit. In a bid to enhance sense-making, most of the participants said their organisations are investing in some or all of these capabilities: advanced Analytics; Cloud computing; and Data Science. Two participants mentioned that their current organisations are building internal research practices as they see this as a capability that will give them a competitive advantage.

Contextualisation

Participant 10 further illustrated that once the problem has been defined it is important to contextualise it.

"*This involves classifying the problem as either simple, complicated, complex or a crisis. What is the antecedent of this problem? Based on this assessment, we can design suitable response measures. Therefore, leaders have got to apply the right response measure depending on the problem context.*" (Participant 10)

Most participants echoed similar views, citing that strategy and decisions must be informed by situational analysis. This involves making sure the organisation has resources and capabilities, specifically procured or built to respond to threats and opportunities.

Participants emphasised the need for listening to the voice of the customer. There were strong sentiments that leaders have a duty to speak directly with customers and understand what the customers say. Participant 11 stated that organisations must collect both qualitative and quantitative data from and about customers. Participant 1 further pointed out that qualitative data is data, and often contain rich contextual details. Data must not just be collected in vain, but it must be analysed to improve product offering and service levels to the customers. Leaders have to sense current and future needs of their customers through studying the changing trends, situational analysis and thorough appreciation of the problem, the context and the profile of their customers

Participants acknowledged the task of the leader was difficult. Participant 12 remarked that;

“*It is already hard enough to contextualise the internal environment within one’s organisation. Yet, within our organisation, we can get almost all the data we want. It is even harder to contextualise the external environment where you do not own the data; it does not reside in your own database; you may not have access to critical data when you need it. Yet, you must apply the correct strategic response.*” (Participant 12)

Most participants said there are increasingly using advanced analytics to help navigate complexity. Four participants pointed to COVID 19 pandemic as something that accelerated the need for insights. Participant 2 and Participant 3 both said several business models had been disrupted by COVID 19 pandemic and associated lockdowns. These participants cited examples like office space industries and related value chains such as bulk printing companies that sell and maintain printers and copiers in big offices – these models are disrupted. They both alluded to the fact that while events like COVID 19 are impossible to predict using data, the trajectory of recovery can already be gleaned out of existing data. Participant 1 and Participant 2 further contended that data contains enough insight to help guide decisions going forward. However, they conceded that extracting insights from this data requires advanced analytical skills and tools which leaders don’t always have in their disposal.

According to Participant 1, organisations that achieve competitive advantage have inbuilt competencies and routines that allow them to look beyond traditional places for answers.

“*a company in mining, for example, a sunset industry, must consider all its options for the next 20 to 30 years. If those options point to a need to enter other sectors, those decisions must be implemented in earnest*.” (Participant 1)

Participant 1 and Participant 5 voiced strong views that scanning the environment is not only for identifying the threats but an opportunity for the organisation to disrupt other industries as well. They contended that offence is the best defence. Participant 1 added that the strategic decisions to build or enhance resources and capabilities must be made not only as a defence mechanism but as an intentional strategy for offence and exploitation as well.

Participant 7 said it is also crucial for leaders to contextualise the internal environment with respect to the strategy they wish to pursue. It is important to take the stack of the current capabilities and processes before embarking of a process of strategic change.

“*Leaders need to know the internal complexities that could undermine the change they intend to make. Often organisations are marred by constraints such as legacy systems and* *regulatory requirements that make the process of change slow. Participant 7 agreed that customer-facing capabilities are important but insist that organisations must pay equal attention to improving internal capabilities and processes*." (Participant 7).

The importance of understanding the internal details of an organisation is confirmed by participant 8;

"*You always have to really understand the temperature of the organisation that you are dealing with. Whether you are doing an organisation-wide strategy or strategy for a Department, you need to immerse yourself in terms of understanding what the status core is. I've seen people come into an organisation; a first week or even a month, a person is already saying this is what my strategy is; this what I'm going to do here. But what is wise is to understand the people and their culture. The current culture that's there; the flow of power; what makes the organisation tick, their attitude... And one must then get to understand even the hard stuff; things like the product and services, how it all work together. What makes them tick and what does not tick - as well*." (Participant 8)

Strategy formulation

The process of designing a strategy is done at the back of a clearly defined vision and the well-articulated contextual environment, as described in the preceding sections. Participants described the strategy design process in terms of the following distinct and steps: problem identification; analysis; and creating of the solution.

Most participants who spoke specifically on problem identification step said this requires proper scanning of the external environment for opportunities and threats. Participants 7 stressed the importance of factoring the internal constraints like legacy systems into the framing of the problem. Participant highlighted the importance for leaders to immerse themselves in the internal organisation and its culture. In these views, Participant 7 and Participant 8 are drawing attention to the importance of combining the internal and external environment appraisal as part of the problem definition process step. Participant 5 stressed the importance of inviting the right people to the strategy session and ensuring that there is broad coverage of all stakeholders.

The analysis involves the process of applying the frameworks and tools to elaborate on the strategy design. The frameworks that were mentioned by participants are discussed in this Analytical problem-solving section above. Participant 8 and Participant 10 stressed the importance of applying the right framework for the right problem to arrive at better decisions and solutions. Participant 8 further warned that against the futility of force-fitting a framework into the context it is not suitable. Participants clarified that the solution is then at the back of the analysis step. Usually, several options are made available from which the final solution is selected. Participants explained that the decisions for the final solution are captured in a strategy formulation document, and this document also captures the goals and objective, as well as metrics that will be used to measure implementation progress.

Participant 10 said the organisation often miss the important step in the process, which is contextualisation. Once the problem has been identified, it needs to be framed as either a simple, complicated, complex or characteristics dimension. Participant 10 further clarified that this is important because the analysis of the problem will depend on the context it has been frame;

"*you want to avoid applying a simplistic solution to a complex problem, this created problem for implementation teams downstream as they discover that proposed solution does not work, and strategy fail, but the problem lies right in the beginning. It lies in the failure to contextualise properly*." (Participant 10)

Three participants further outlined that the strategy formulation process must produce a document that spells out the chosen solution to the identified problem. These participants said this is effectively the new organisation strategy. Participant 12 said the strategy formulation outcome could be seen as "*a cannon of distinct strategies which will be managed towards the attainment of the company's goals*". Participants explained that the strategy document must spell out the new opportunities the organisation is perusing or the defence mechanism it seeks to build. It must spell out the resources it will buy, the resources it already has, and the new capabilities to be built in pursuit of the new strategy. New capabilities are vital as they help organisations adapt to a changing environment. Participant 9 stated that; "*Failure to adopt kill organisations*".

Participant 7 said companies must not be limited by the capabilities they currently have but must also look at sourcing the capabilities they do not have in order to enable their strategies. This view is essential because and differs from other participants who viewed the internal capabilities as a constraining factor in the strategy design and therefore, the opportunities an organisation can pursue. Participant 7 and Participant 8 said a more aggressive way of acquiring capabilities is through Mergers and Acquisitions. However, both participants noted this was a high-risk strategy, and strong leadership is required to realise value from mergers and acquisition strategy.

Strategy implementation

Most participants said strategy design is simpler than implementation; however, this process is often relegated to a lower level of management. Participants differ in their views of how much support implementation efforts get from senior leaders. Some participant reported consistent high-level support and sponsorship. Other participants felt that less attention was accorded to implementation efforts as compared to strategy formulation efforts.

Participant 5 and Participant 9 stressed that the strategy must be agile. Both participants mentioned that the feedback from implementation efforts must be used to make decisions about the suitability of the chosen strategy and even to change it. Participant 12 said data plays a crucial role in being able to gauge the success of organisation strategy. Participant 12 explained that the organisation needs both internal and external data to measure the impact of its strategy and must not hesitate to make changes where there are failures.

5.3.2 The role of leadership

Participant 10 specifically pointed out that strategy is the responsibility of the board of directors for that organisation. However, in practice, this responsibility is shared with executive leadership and senior management of the organisation. Participant 5 said leaders often neglect to include a fair representation of all organisational stakeholder, especially at the lower rungs of the organisation. Most participants stressed the importance of prior preparation ahead of strategy meetings. Participants 5 emphasised that the providers of data and input for such preparation are found across the level of the organisation and leaders must ensure they are involved. When queried on the practicability of this view, Participant 5 clarified that this is not about the democracy of involving everyone. Still, leaders ought to know the critical input providers and ensure their inclusion without blotting the meeting.

Participant 3, Participant 4, and Participant 12 pointed out that the critical input that must be provided to leadership is data to aid their decision making. Participant 3 added that.

"*data requires a competent interpreter so that one does not add complexity to the picture but helps the organisation to truly declutter and filter out the noise in the environment.*"

(Participant) 3

"*Therefore, if you believe that technology is playing a key role in shaping the direction of the organisation, and then you believe that Data science is playing a critical role in shaping the strategy of the organisation. Then it would make sense that you would want to bring those* *minds on the table where the strategy is being defined... It does not matter where this person reports to, right? You just bring that capability to the table."* (Participant 3 and Participant 4).

Still, Participant 13 finds there are limits to what data could be used for;

" Decision making, in my experience, is seldom based on data, right? Decision-making is experience and guts, right? But for me to make the decisions, I need to understand the problem, and the data helps me to understand the problem so that I can make a better decision. Data is not used to make a decision but to support the process of decision-making." (Participant 13).

Participant 4 mentioned that organisations in matured industries tended to rely less on advanced analytics and more on their operational data. Participant 4 also added that older leaders tend to trust their knowledge above data. Participant 8 took the point further stating that, usually, highly paid leaders dominate decision making and even ignore the facts presented through data.

" In some organisations, I have seen strategies coming shot; they usually develop a negative culture within an organisation where opinions of a highly paid and highly experienced person are priced over those of everyone else in the room. And in many cases they are not giving facts, they're just giving their opinions. Their opinions are so honoured within an organisation, and sometimes it will force an organisation to be blind to the new things that are coming in the market because of the historical experience of this person… the best predictor of the future is through the use of data." (Participant 8)

Participants who have or currently undergoing Data Science implementation reported that they had seen the successful performance of analytical models that can make actual decisions around: effective products portfolios and product mix; market viability; and customer segmentation. They reported that, in some cases, these analytical models provided better predictive and prescriptive decisions than managers can. There is, however, a consensus that leaders can and should override some of these decisions when there is a good course to do so. An example pointed out by Participant 3, the decision that scores high on the model has significant risk compared to the business risk appetite. In such cases, leaders should override the decision of the model. Participant 12 provides a moderating view which reconciles the constructing perspectives reported above;

“*Data Science and Artificial Intelligence is not about replacing human beings with machines. It is about humans working alongside with machines. Machines assisting humans in areas where humans are no longer efficient and cost effective in carrying out a task.*”
(Participant 12).

Yet, while Participant 12’s view captures the dominant thread reported by most participants, there are ethical questions attached to these views, which are discussed further in Section 5.5.1 below.

In essence, respondents described the role of the leader as a complex one and vast in scope. They painted a list of knowledge areas that successful leaders must be vested in: leaders must have a deep understanding of the operating environment; leaders must understand the organisational drivers and constraints; leaders must be fully aware of the resources and capabilities in their disposal; leaders must understand and be close to customers, and leaders must understand the culture within their organisations.

Participant 11 said in order to gain a deep understanding of the environment leaders;
“*the first thing that you do is that you start off with an environment scan. Leaders can use published industry report, external data, and commissioned reports. These sources help with providing market trends and the broad direction the industry is taking… leaders must apply frameworks and tools like PESTEL to capture their understanding of the environment and compare notes with fellow leaders.”* (Participant 11).

“*So if you're looking at, say, growing the customer base from 1 million to 2 million, you have got to understand the customer enough to identify what is going to move the dial; otherwise you are not going to make it. Close proximity to customers will provide you with the data points you need to inform your growth strategy.”* (Participant 11).

Attributes and roles of leaders

When participants were asked where the Data capability should be positioned with an organisation, given its strategic importance, most answered: “*directly to the CEO*”. Participant 2 stated that establishing the office of the Chief Data Officer (CDO) is the first step towards making Data Science a strategic capability and this office must report directly to the Chief Executive Officer (CEO). Participants mentioned that in traditional organisation structure, the Data function is located in the office of the Chief Information Officer (CIO) or the Chief Financial Officer (CFO). However, most participants are of the view that, given the impetus accorded to data and the strategic importance of the modern data capability, there is a need to position this capability at the centre of decision making. This translates to a need for CDO.

Participant 2 further clarified that the role played by the CIO and CDO are different;

*"CDO and CIO are complementary but different competencies, ok. I will give an analogy that I read some years back. Where someone said if you want to know the difference between the CIO and the CDO think of water and plumbing. You have got all the pipes, that is the CIO. The water that flows through the pipes is data. They are totally different, but they need one another." (*Participant 2)

Participant 3 made the remarks below concerning why younger leaders embrace data readily compared to older generation of leaders. Participant 3 also summarises the evolution of data from being a resource to becoming a strategic capability;

"The young generation of leaders started in business in the 1980s or the 1990s where they were exposed to data; and exposed to large volumes of data. They have an advantage in that, for them, data has always existed. Older leaders started back in the 1970s when computerised data was not readily available. I mean, yes, Finance has always been a data game for 2000 years. There has always been some business data set, in manual form, for 2000 years used by Finance. But ever since the 90s, data expanded beyond just the finance numbers to broader numbers. Operational numbers and Market numbers. You know, sharing of information with computer systems became standard. So it is the leaders that have come through the business from that time who are now sitting around the strategic decision-making table and actually understands the value of data. And, I think that's what's driving the shift where a lot more people appreciate and will invest in this asset which is data."

No other participant, except Participant 3, acknowledged the leader's age as the driver for their propensity to embrace data. Most participants said organisations that do well in implementing the Data Science capability were those that invest the money into its development. Participants emphasised the point that Data Science unit requires investment, and that is a strong indication by leadership that they take data driven insights seriously. Participants further noted that any capability that helps them as leaders to connect the dots across industries and confer a competitive advantage on their organisations; such capability that pays for itself.

5.3.3 Decision making

As captured in the sections above, participants clarified that the whole process of strategy formulation is about making decisions. However, Participant 3, Participant 5, and Participant 13 all stated that there was a need to differentiate strategic decisions from operational decisions. Participant 3 made the point that operational decisions are made daily and help guide the day to day operations of the business; in contrast, strategic decisions are made through planned occasions and help guide the long term direction of the organisation. As a consequence of this, strategic decisions do not need real-time data to support them; strategic decisions require aggregated data that feed analytics.

Making decisions, whether operational or strategic, is hard enough. However, Participant 12 said strategic decisions were more demanding. This is because strategic decisions involve significant investments in the chosen decision. They require taking one course of action instead of another, and often those decisions are not reversible. Participant 8 said this is the reason why the organisation must rely on more than just gut feel to make strategic decisions. Participant added that:

"The key thing is that if you don't invest in getting that data, that would be required for you to make decisions, then you most likely will make decisions that are informed by gut feel and all of that. And we all know that gut feel does not always give you the right answers." (Participant 8).

Participant 12 said there is often a large number of relevant and pertinent variable that must be considered when making strategic decisions;

"*These involve changes in the market, and how they affect the demand for the organisation's products, the substitute products, the competitor's products. How changes in the market affect inputs costs is a relevant consideration. Also, important considerations include the macro-economic impact on consumers, and customers and how this is likely to govern their purchasing behaviours in the future. The internal environment bears its complexities. It is an extremely complex task. A modern leader must, therefore, leverage tools, resources and capability that help them to declutter and improve decisions." (*Participant 12).

"*One of our clients in the UK is a large services business that runs many services contracts like a Bidvest type of business, but it's a global business. They have a cleaning business, and the other one is people transport business...so they do many different things for different organisations and government. So they actually go and model the operational drivers that feed the economics of the business so that they can start making strategic choices around* *which products or which service contracts are toxic contracts; which contracts were poorly set up initially and under contracted... What are the drivers that created that scenario? And how they can better bid for work in future so that they don't get themselves into a situation where they have loss-making contracts." (*Participant 3).

It is important to note that while operational data is not suitable for strategic decisions, data analytics, on the other hand, is required at all levels of the organisation. Participant 2 and Participant 7 made several examples to support this argument. Participant 2 said, in his current role, they make product pricing decisions weekly and sometimes daily. They use analytics to predict the behaviour of the markets, customers and also use analytics to get ahead of the competition.

The findings show that where analytics and data capability is missing or where the culture of the organisation does not place reliance on data-driven decision making, then democracy and gut feel decisions prevail. Participant 11 explained that in her experience, there had been too many decisions that were made without a thorough analysis of data. This would typically be due to the organisations' culture rather than the scarcity of data. Participant 12 warned that not leveraging data can lead to decisions that are contrary to prevailing trends and market signals. It also leaves room for the politicisation of decisions. Participant 12 described the politicisation of decisions as:

"*the appropriation of the strategic decision-making process in a manner that produces an individual leader's predetermined outcomes. Leaders who do this support their point of view through the use of a combination of anecdotal evidence, personal experience and hunch.*" (Participant 12).

In this section, the findings covered a sweeping landscape of strategy formulation and the thought process. Participants went into depths in describing factors that impact decision making are weaved into leadership thinking and action — this helps to address the research question 1 through empirical data. The next section covers the results for research question 2.

5.4 Results for research questions two

Research question 2: How do Data science practices contribute to the creation of sensing capability and competitive advantage within an organisation?

In connection with research question 2, participants were asked questions about: the role of Data unit in organisations; the evolution of the data capability in recent years; the embeddedness of data and reliance leaders put on data when making decisions. Participants who are practising data scientists and those are leading Data science teams were asked more profound technical questions related to the process and technologies involved in setting up and running a Data unit.

The results to research question 2 are presented in four subsections below. These subsections cover the following: crafting Data science capability; Data science processes and practices; Data science skills; and Data science technology.

5.4.1 Data science processes and practices

This section focusses on Data science process and practices as described by the research participants. The section covers the configuration of Data Science functional area, sometimes called Business Intelligence and sometimes called Data unit. This report refers to it as the Data unit.

Data unit and team configuration

The findings reseal that few of the participants have had an experience or have been part of establishing a Data unit from the ground up. However, most of the participants have had ample opportunity to be directly involved or consulted in the transformation effort of Data units, in the past five years, in their respective organisations. Participant 3 said the process of establishing a data science capability is essentially a process of reconfiguring and transforming the traditional data warehouse function into an advanced Data unit. Participant 3 added that the data warehouse units have been in existence for more than 20 years, and they are still a good idea to have as a foundation. However, Participant 6 vehemently oppose that the data warehouse is still a good idea. Participant 6 said data warehouses were a hindrance that slows down the development of advanced analytics because they consume financial and human capital which could be invested in the Data science efforts.

Yet, the findings show that the transformation efforts to Data science and data analytics include the old technology stack, which is dominated by data warehouses. Participants reported that advanced Data science technologies coexist with old ones and their organisations may retire some of them but only in a gradual manner. Therefore, the findings reveal that the establishment of Data science capability is often a transformation project instead of a greenfield project because most organisations are building on top of existing capabilities. In essence, these findings show that the building of the Data science capability entails the configuration of Data unit to add advanced technologies and processes which culminate into a Data science capability.

Most participants stated that the selection of people is the key and foremost consideration when transforming the Data unit. Section 5.4.2 below provides comprehensive findings on skills set required to set up a Data science capability. The finding shows that cross-functional skills are more valuable than deep technical skills. In practice, the transformed Data unit will typically include people with technical skills and those with business skills which is a departure from a traditional configuration where business stakeholders sat outside the Data unit. Participant 6 even mentioned, in jest, that:

"*There is no such thing as a Data Scientist, a single person cannot single-handedly manage the outcomes required of a Data science project, this should always be thought of as a team*" (Participant 6)

This section presented findings which demonstrate the importance of constituting the team with the right skills. The Data team makes up for the skills it does not have by co-opting people from other business units to complement its skills. Participant 4 indicate that it was normal for Data science to have extensions that involve working committees that pull skills complement from the broader organisation. The next section presents the Business analysis practices that are part of the Data team process.

Business analysis

The findings show that Data science practices are more robust and detailed than the traditional practices data warehouse practices. Participant 3 narrated that the traditional business intelligence reports were developed through gathering requirements from business users and documenting these in specification documents after that reports will be developed and provided to users. However, Data science reports development follow a different process. First, there is a problem statement that gets written, according to Participant 3 and Participant 6, problem elaboration is a robust process which involves both business and the Data team. In this case, the Data team is not merely documenting what the business is saying but actively analysing the problem and elaborating it. Once the problem has been clearly articulated, to the satisfaction of the teams involved, a set of hypotheses are then developed. The Data team then investigate candidate data sources to help test the hypotheses.

Once the data has been sourced, it is ingested into the data environment, prepared, and cleansed to improve quality where necessary. The Data team would then perform analysis, develop models to test the hypotheses and produce results. Participant 6 said the team members rely on each other to peer review each team member's work in order to minimise mistakes, "good teams are highly collaborative and open". The final stage would be the reporting of the results first to the rest of the team and then to the target business audience. The findings show that some Data science projects produce reports that are useful only once-off while others would produce reusable results - as long as the concerned problem persists. Participant 3 and Participant 6 see the sequence of events as iterative and consists of four distinct phases which are: 1. Problem definition; 2. Data sourcing; 3. Analysis; and 4. Reporting. It is important to note that Data science projects do not only produce reports; as such, some participants referred to Data science output as a product.

"*Sometimes the output is a mathematical model; sometimes the result is a data stream which can be fed to a line of the business system so that it can make, intelligent, automated decisions*" (Participant 3)

Each of the distinct four steps has its own complexity, and according to Participant 7, the organisation's competitive advantage in this area is determined by the competence of the Data team and how thorough it is in executing each of the sub-steps involved in producing data products.

"The problem-solving step requires interdisciplinary collaboration. Collaboration with subject matter experts is essential. Participant 4 emphasised that technical skills alone are not sufficient;

"What becomes powerful is if you take Data science, and you combine it with industry insights. If you take someone, for example, that understands the Telco or Financial services business. You must take this person, and you bring them together with somebody who's a data scientist. You must remember that there are certain stories that the data might be telling you, but if you don't have that insight of the business or the industry, you will not know what this data actually tell you. So, you need somebody that has got the industry expertise or industry insight to be able to show you deeper meaning in data."

Findings are clear that the problem definition step requires the collaboration of interdisciplinary teams. Participant 6 emphasised the need for a cross-functional and multidisciplinary team when zooming into the problem. According to Participant 6, this is critical because all problems have a boundary space. Data science models provide answers that are relevant only within the problem boundary space. Therefore, people with an intimate understanding of the problem space must be involved in the problem definition. An illustration was given about solving a customer churn problem. The Data team would identify a set of drivers that cause customers to churn; customers who churn for reasons other than those defined in the set should fall outside of the boundaries of the problem - as defined by the team. The final model may suggest an action that customers should be incentivised or sent communications to reduce churn. It is critical that the suggested actions are applied only to those customers in the problem set, which represents the boundary of the problem. It is clear from these findings cement the Data science is a collaborative multi-disciplinary endeavour.

It also came through that as par the problem-solving process; the team may discover that the problem at hand has a solution that can be managed using operational reporting or an Excel-based model. Such problems are typically referred to as teams dealing with operational reporting as they do not require in-depth analysis or advanced analytics.

This section narrated the findings that pointed to the importance of problem definition and inclusions of the multi-disciplinary team. It was also stated that this is the first of four steps in a typical Data science effort. The next subsection deals with the second step, which is Data Sourcing.

Data Sourcing

The findings revealed that data could be found in internal or external data sources and that it could be structured or unstructured in form. Internal data is sourced from the company's line of the business system (LOBs) through the integration with the data platform. Participant 6 pointed out that internal data is often subjected to business rules and is stored in the data platform in a structured form. Sourcing internal data is straight forward, but not necessarily straightforward as it often required integration hen dome for the first time.

According to Participant 2, Participant 3, and Participant 6, the key distinguishing feature Data science environments is that unlike the data warehouses, which predominately contain structured internal data, Data science environments accommodate both external and internal, structured and unstructured data sets.

To source external data, the Data team must establish external ties, collaborate with ecosystem partners, which include ties with outside organisations. Furthermore, participants pointed out that establishing ties that reach outside the organisation may need facilitation by senior leaders. It is common practice in matured Data teams to include leaders in all hierarchical levels. Senior leaders are co-opted into the semi-permanent Data committees, and their expertise solicited as the need arise.

Some of the data is sourced market transactions from companies specialising in providing data. According to Participant 7, it is often a good idea to procure data from data providers as it is often of acceptable quality, which saves the team time as minimal quality checks are required. The findings also show that Social media organisations like Facebook, Twitter, Instagram, etcetera, provide means for organisations to extract social data through integration interfaces called application programming interfaces (APIs). Participant 3 reported that in advanced use cases, companies collect streaming data from sensors; for example, some insurance companies collect weather patterns data overtime to develop models for predicting ling term climate. They then use these models to build and price insurance products for agricultural business and commercial farmers.

Another insurance organisation uses devices in customers' vehicles to collect information about customer driving behaviour. This information is then fed into a Data science model which calculates monthly incentives and rewards based on the driving risk profile of each customer.

Another distinguishing feature of Data science technology is its ability to process varied, massive data sets, that are generated at high speed, often referred to as velocity:

"Real-time data comes in high frequencies, it is just sprayed at you, and you need to capture it, and that is a much bigger concept you need to think about… the format in which the information arrives at your environment will also give you the biggest challenge. A lot of your external data would come in that way because you are tapping into the world around you, and it continues to happen around you. The world around you is not going to send you bits of information packaged at the end of the day. You need to read that and log it and capture it while it is happening." (Participant 3)

The findings point out that some data sources are freely available. These are often accessed through APIs or data file dumps. Yet, other techniques involve data scrapping, which utilises software to extract information from websites automatically. Participant 1 said it is important to note that qualitative data is also data, and it has a useful role to play in advanced analytics.

Participant 3 and Participant 6 mentioned that qualitative data generated through research or taken from published reports are used to gain an understanding of a problem, to set up hypotheses, and to inform assumptions underlying the mathematical models.

It is observed that Data science has the ability to uses both qualitative and quantitative data in all its four stages. Data is collected from various sources and is made of a variety of formats like texts, images, videos, etcetera. This enables the Data science team to process multiple variables at the same time. Several participants mentioned that Data science models could address dozens and even hundreds of variables simultaneously. An amount of information that an individual cannot handle at once.

Ultimately the organisation must store this data. Participants said the proliferation of Cloud services in the past ten years has made it cheaper and possible to handle the large amount of data required for data science projects. Participant 12 mentioned that the recent arrival of Microsoft cloud, Azure, in South Africa is a game changer. The cloud reduces the upfront costs associated with buying servers. The cloud environment allows organisations to pay for storage services and computing services only when they use those services. Participant 6 mentioned that cloud based solutions like Azure Data lake is geared to support Data science initiatives, which makes the development of Data science capabilities easier than before. Other participants mentioned that companies like Google and Amazon offer similar cloud services as Microsoft Azure.

The findings in this section clarified the differences between the sourcing of internal and external data. Several options for sourcing external data were presented, including the challenges involved. The insights given by participants help illuminate the processes involved in sourcing data, which helps partially address Research Question 2. The next section presents findings on the analytical model design.

Analytical model design and execution

The findings show that once the data has been gathered, a team of data scientists begin constructing mathematical models that test hypotheses which were set up during the problem definition. There is usual data that is ring-fenced, not processed and transformed by models, which Participant 6 referred to as a control experiment or AB testing set. Participant 6 provided an illustration using customer churn models. Suppose 10 000 customers are identified by the model as a churn risk, and therefore targeted with a mitigation scheme like a price discount. Then, a small percentage of the qualifying customers would be set aside, say 10%, and would be excluded from the churn mitigation scheme. When the scheme has been rolled out, the

Data team will then study the behaviour of customers in the 10% set versus the behaviour of the customers in the bigger set. If there is a significant change in behavioural patterns in the bigger set, and this pattern is not observed in the 10% set, then the model is taken to be valid – assuming other validation criteria pass.

Participant 3 and Participant 6 mentioned that Data science models are predominantly mathematical in nature; they are built using technics in Applied Mathematics, Statistics, and Computer Science to solve problems. Data science model also use techniques in Machine Learning extensively. Machine Learning derives from both Statistics and Computer Science - Artificial Intelligence. Participant 6 added that models are then coded into algorithms in a programming language; these algorithms are optimised to process large volumes of data at high speed.

Participants described the concept of analytical model training. Training involves making a model recognise a pattern, and then improving on its ability to recognise other similar patterns by passing through a large amount of data. Eventually, its ability to recognise these patterns improve as a result of processing a large volume of data. The availability of large volumes of data is key to model training.

The results illustrate that Data science can produce three categories of analytical models – these categories are based on the function that these models fulfil, they include descriptive models; predictive models; and prescriptive models. The first and most basic is the descriptive category. Descriptive models simply present results, which are often visualised in a report or a dashboard. The users of the prescriptive model dashboard draw their conclusion about the results presented to them. Participant 13 described descriptive models as backwards-facing;

"*There is limited value to be derived in operational reporting; all that these reports give you is a description of what has happened. That is all there is to descriptive analytics. We have always had this. It is not artificial intelligence.*"(Participant 3)

However, most participants acknowledged the value of descriptive reports for operational needs. Examples given include reports that show sales trend over time; products revenue contribution by segment and by time period. The findings show that management at all levels of the organisation have used these reports over the year and will continue to do so. Still, descriptive models can achieve advanced outputs compared to traditional reporting, and this appears to due to advance mathematical models and algorithms underpinning these models. Participant 5 mentioned there had been successful use cases where media organisations deploy descriptive analytics to filter out fake news from online content aggregated across hundreds of websites.

The second category of analytical models is predictive analytics. The findings show that this represents an advanced level compared to descriptive analytics. Predictive models produce analytical reports that use data to predict future events or future trends. Participant 7 explained that this is the same as forecasting, which has always been there historically. However, Participant 7 added that Data science-based predictive models could combine internal data with external data. It is now common to combine macro-economic forecasts and social media sentiments to produce a single forecast model; like a revenue forecast model. Participants 7 further commented that this combination is contemporary and represents advancement brought about by data science.

The findings show that predictive models are useful for aiding decision making as they process large volumes of data and paint a picture of the future. Participant 9 stated that predictive models can assist organisations in predicting looming disruption by predicting the future performance of new entrants in an industry.

The third category of analytical models is prescriptive analytics. This represents the most novel of Data science products. Prescriptive models also operate on an extensive data set and produce an analytical report that tells users what the best cause of action is. The findings show that organisations that have implemented prescriptive models can automate advanced decisions. Participant 4 narrated a project they implemented where the prescriptive model produced decision best target markets for their products. The model also provided decisions on how and when raw materials had to be sourced based on predicted global commodity demand and price levels. It is this category of advanced models that help leaders declutter complexity and inform better decisions. Participants 6 warned that both predictive and prescriptive analytics do not work under and environment of chaos; for example, COVID 19 pandemic has invalidated many analytical models. The team must rewrite those models or deprecate them entirely.

However, some participants warned that models should be periodically reviewed to check the validity of assumptions and review the boundaries of their problem space. If the underlying assumptions shift, the model needs to be updated and in certain cases, deprecated.

"*Ultimately, every model has a confidence level. A data scientist must disclose the confidence level to end users so that management can decide for themselves how much* *trust to put in a certain model… I have seen people try to rescue intractable models… we should be courageous enough to throw away inaccurate models instead of trying to justify them*" (Participant 6)

Participants clarified that a model with a confidence level of 90% is more reliable than one with an 80% confidence level. As such, the visualisation and presentation of analytical models are essential and require a competent presenter to help convey the meaning correctly to the end-users. There are visualisation tools that are part of the Data science toolkit; these are discussed in Section 5.4.3 below. Participants made it clear that Data science analytics provide great assistance in sorting out complex information and decision making. However, leaders make the final choice on the cause of action.

This section presented the findings on different categories of models. This revealed that there are three categories of analytical models; in order of novelty: descriptive, predictive, and prescriptive. Several illustrations of how these models are used in practice were provided. The next section presents quality management practices in Data science operations.

Quality management

The findings reveal that data quality is a critical consideration that impacts the reliability of the analytical models if not appropriately managed. Both internal and external data can introduce quality issues, although external data was reported to be proportionally more prone to quality problems. These results make sense as external data combine data collected from various providers who may not apply the same standards in storing their respective data sets. Typical quality issues include incompleteness, lack of accuracy, invalid records, inconsistencies, and redundancies. Participants emphasise different quality issues. The reason for this could be that different organisations experience each type of quality issue in varying degrees. But what does each of these quality issues mean?

Participants explained typical data quality issues. Incompleteness refers to instances where vital information is missing in a data record. Examples given include a contact number in customer or lead data set; missing transaction value in product sales data

. Invalid data occurs when values appear where they should not, examples include a date appearing under a surname field, or an id number appearing under an email address field. Inconsistencies refer to a situation where there is no uniformity in data format under a given field. One example is a currency field containing venues that are expressed in various currencies. The Kaggle project 2 shows that how data is presented can create difficulty for analysis. The data scientist working on Kaggle project 2 highlighted that the data site provided was aggregated at the level of the month; on close investigation, it was discovered that the weekdays' customer demand differed to weekend demand. However, using the data aggregated at a month level could not reveal these details, which the data scientist said they are critical to the model. Therefore, this reveals the need for the data team to pay close attention to the nature of the data set and the limitations they pose on downstream analysis.

Data quality issues make it difficult and even impossible to manipulate data for analysis. One cannot simply add the total balance of field that contains data reported in various currencies. To resolve this would require additional information which points out the specific currency of each value – this information may not be available for specific data sets. Participant 3 stated that all Data science projects require data quality management as the issues are unavoidable. According to Percipient 3 "*good quality management practices assume errors in data until proven otherwise.*"

Participants said the process for identifying a correcting data quality is often referred to as Data Engineering; some participants call it data wrangling. It is part of the data sourcing ad storing phase, and it occurs before the data is fed into analytical models. Participant 6 illustrated data quality scoring as a useful good practice. Data quality scoring involves assigning a score for each data source and creating an aggregate score for the quality of the entire data platform. The score is often expressed as a percentage. As the team continue to close known data quality problem, the overall score improves. According to Participant 6, this increases the credibility accorded to reports and models produced by the Data team as team's transparency and commitment to quality improve stakeholders trust. The ultimate goal of the data quality management practices is to improve the quality of the decisions downstream.

The findings show that operating a successful Data unit requires a good set of skills. However, a Data unit that implements Data science requires significantly more advanced skills sets. The typical skillsets are discussed in the next section.

5.4.2 Data science skills

Roles and cross-functional skills

As presented earlier, Data science implementation requires the involvement of people with diverse skills set. People are co-opted from various departments within the organisation to provide business or industry expertise which facilitates problem-solving. These could include people from Finance, Marketing, Operations, or services business units like Research and Development (R&D).

Based on the findings, the core Data team consists of people with mathematical and statistics skills, computer science skills, and domain expertise. A typical data scientist has computer science, engineering or mathematics educational backgrounds; sometimes even advanced degrees. According to Participant 6, the majority of the practising data scientist started with a qualification in either mathematics, engineering or computer science and augmented their skills over the years. However, there is a growing Data science tailed qualifications offered at both undergraduate and postgraduate levels.

Participant 3 said the two most critical roles inside the core Data team are those of data architect and data scientist. Participant 3 added that these roles are strategic and help build competitive Data science capability. Over time, architects and data scientists internalise vast intellectual property of their organisation, and therefore organisations must retain them. A data architect is concerned with making sure that the design of the end to end data platform support efficient Data science processes. Architects also designed the governance process and set data standards based on benchmarked best practices.

It is clear from the findings that traditional data warehouse and operational reporting skills are often retained. These consist of data engineers and developers who are skilled in what is called Transactional SQL - structured query language. These skills add a lot of value in processing and organising internal data as well as the processing and addressing data quality issues. However, according to Participant 3, these skills are not strategic. Organisations can outsource them when there is a need.

Participant 3 said organisations should never outsource the Data science roles as much of the organisational intellectual property accumulate in them. However, Participant 7 and Participant 11 said organisations that cannot afford to build their Data science capability should look to outsource the entire data function to organisations specialising in offering the capability. In this way, they gain the benefit without the substantial upfront investment required to build the Data science capability internally.

5.4.3 Data science technologies

As a collective, the Data science team must possess knowledge of data processing algorithms, mathematical modelling, statistics, machine learning, and more. They need to be skilled in programming languages like SQL, R, Python, Scala, Java, etcetera. They need to be adept at working with technologies Amazon web services (AWS) or Microsoft Azure, Spark, Hadoop, etcetera. The lists below sort out the artefacts of traditional and contemporary Data practices in two categories:

Traditional set-up and technologies

- Business intelligence produced descriptive and predictive reports
- Based on structured internal data
- Textual data format
- Advanced analytics + Data Architecture + Data
- Cube
- SQL and MDx query language
- Data Warehouse

Contemporary set-up and technologies

- Data science produces predictive and prescriptive analytics
- Use of internal and external data with structured or unstructured forms.
- Variety of data formats, text, images, voice, and video.
- SQL is still relevant in the contemporary toolset
- AWS, MS Azure - Cloud is scalable and generally more secure than on-premise infrastructure
- Programming language: SCALA, SQL, R, Python
- Data Science Tools - Spark, Java, Hadoop
- Gradient Boosting Machine (GBM)
- Spark to handle multiple data sets
- Skills – mathematics, statistics, machine learning, algorithms design, business domain knowledge

This section provided a reference list the participated pointed out as important in the Data science implementation. With the appreciation of the input in this section and all the preceding sections, the next section presents how Data science capability can be built. The next section covers the elements and processes required to implement a Data science capability based on these research findings.

5.4.4 Crafting Data science capability

This section discusses vital facets involved in implementing a Data science capability as reported by participants. Implementing a Data science capability requires a project effort; whether the team is implementing the capability from a zero baseline or upgrading a traditional data warehouse, some form of a project is required. The findings demonstrate that the successful implementation of Data science projects need to be aligned to company strategic goals. It is crucial for the implementation team to ensure alignment to these goals and the overall vision of the organisation. In certain organisations, visionary leaders champion the idea of implementing a Data science capability. However, participants say in most cases; it is the Data team that proposes the transition to senior leaders. In the latter case, the Data team is expected to produce a solid business case to justify the project. Therefore, the first huddle the team has to get through is securing investment for the implementation project.

Investment in Data science capability

The Data science capability is underpinned by a technology platform which needs to be built to the specification of the respective organisation. Implementing data platform requires investments. The findings show that these investments can be more than R10 million, and significantly higher for a large organisation like in big banks in South Africa.

Participants reported that despite the advantages demonstrated by advanced analytics over the years, organisations exhibit inertia in adopting data science. Data teams struggle to produce business cases that justify the required investment in terms of monetary return on investment (ROI).

> "*It is easier to use third-party providers these days and commission analytics reports tailored for your specific needs. However, for organisations that want to build their own internal Data science platform, they find it is much more complex to justify the business case for this. It becomes this chicken or egg scenario. You need the platform to justify the value of Data science, but in the beginning, you do not have it. Therefore, you cannot justify the business case. Organisations that have the insight in terms of what they're going to get out of it invest into data platforms, and then they see the benefit when that platform is there*" (Participant 6)

In order to justify the investment, the Data team needs the platform and a sandbox environment to conduct experimental analysis and produce prototype analytical models. The prototypes are then used to quantify potential ROI. Until the sandbox environment or platform is available, and the Data science skills are brought on board, it is a difficult task for Data teams to prove a financially viable business case.

Participants said some organisations overcome this inertia by convincing executives, through the use of industry success documented in research reports or case studies. In cases where there are no success case studies in similar industries or when industry partners are reluctant to share their successes, this approach may also fail.

There these findings reveal that many organisations are accomplice in their own failure to adopt advanced analytics. By insisting on effective methods for justifying projects, these organisations fail to seize opportunities for advancement and setting themselves apart - until it is too late. However, the findings point out that, often, organisations that have a strategy that embraces technology, find it easier to adopt data science. These organisations apply less red-tape for technology projects, making it easier for their respective Data teams to justify Data science investment.

However, Participant 12 advises organisations should not invest in technology for the sake of it. There must still be a compelling road map of how the Data team plans to use the platform for the development of a real strategic capability. Participant 13 once the platform has been put in place, the Data team must begin tracking the ROI on an ongoing basis. In essence, while is difficult to justify the business case before the platform is built, the Data team have a responsibility to put in place metrics that show how the capability is tracking against overall organisations' strategic goals and vision.

Implementation project considerations

It turns out that Data science projects are not merely walk in the park. Participant 6 mentioned that about 70% to 90% Data Science projects fail. Participant 12 concurred with this view, mentioning that the information technology project failure rate is above 85%. However, Participant 12 contextualised this statistic: the failure rate is calculated based on the number of projects that miss the original deadlines; exceed the initial budget or do not stick to the original scope and quality. However, most of these projects are eventual pulled through and completed only not according to initial metrics. The clear message presented by participants is that Data science projects, like information technology projects, are prone to failures and must be executed with utmost commitment.

The findings show that projects that are implemented progressively and in an agile manner can take long; however, they have a better chance of success. Most participants said it was simple to transition from the traditional data warehouse to Data science platform than to start from nothing. The reason appears to be related to the fact that some of the essential practices, like data engineering and data governance, can be transferred to the new Data science capability. Therefore, the learning curve for the Data team is not too steep. Also, in the case of a transition from the data warehouse to data science, the team is simple adding technology and new roles on top of an existing team structure. As such, the team keeps its knowledge about how the business works and the relationship with business stakeholders, which are critical in Data science operations as described in the previous sections.

The findings show that the Data science platform represents a significant improvement from the data warehouse platforms. Data science platforms are designed to consolidate fragmented data capabilities scattered across the organisations. It follows that technology consolidation should also pull together silo Data teams under one capability. However, it could be argued that in large groups with divisions that operate as an independent unit, there could be a justification for each division to operate its data capability. Still, although autonomous by design, such data units would do best to collaborate and find synergies to accelerate the creation of competitive advantage.

Participants cited the advancement of cloud technology as a shift forward and enabler for Data science platforms. Cloud is reported to lower the costs of adoption as there is no upfront cost for acquiring servers as it is often the case in traditional data warehouse platforms. Technologies like Data Lake in the Microsoft Cloud (Azure) is specifically geared for Data science platforms according to Participant 6; Google and Amazon are the next most popular providers of data platform technologies.

The critical message emerging from the findings is that the implementation of Data science platform is best approached in progressive stages. This allows both the Data team and business stakeholders to go through a manageable maturity curve. The maturity stages allow the team to add features, technology and skills set in a pace that they can both manage and afford.

In terms of governance and compliance, the findings demonstrate a need for increased governance to ameliorate increased ethical compliance risks introduced by reliance on external data sources. Pertinent ethical compliance risks are discussed in Section 5.5.1. However, it makes sense that ethical considerations and governance processes are inculcated into the design of the Data science platform technology and team practices, from the beginning.

Team selection considerations and activation

Findings show that successful adoption of Data science capability requires the involvement of strategic, tactical & operational levels. The findings presented above show that Data teams often co-opt key stakeholder across the organisation into Data science committee. These stakeholders assist with providing business and industry knowledge that the team use to define problems and model solutions. Participants 6 said Data science teams need advanced skills which include Data science and mathematical skills.

The findings show that successful and excelling Data science teams are empowered to experiment and have the latitude to propose analytical project portfolio they want to work in any given year, as long as that portfolio aligns with strategic business objective.

"*The Data team is, the strategic capability for decisioning; to the extent that they are successful, the business increases its competitive advantage*" (Participant 13)

Once Data science has been implemented, a change management effort is required to socialise the new capability and improve user adoption. Participant 10 argued that Data science analytics could not be a capability until leaders begin to leverage it for its intended purposes. This point to the critical role change management should play in ensuring that leaders are fully aware of the capability and are enabled to use it for strategic decision making. Leaders must be shown how data Data science products help decrease complexity and improve the ability to sense make. The change management effort is successful if it helps the organisation adopt a culture of data-driven decision making.

However, it is clear that change management effort cannot be a once-off initiative like projects. The change management drive must be part of the Data science operations. The next section discusses key elements that must be operationalised once the Data science platform has been installed.

Data science operations

The findings show that the change management efforts must be driven not only by the core Data science team but must be sponsored by executive leaders at the highest level of the organisation. Business stakeholders co-opted into the Data team through working committees should also play the role of change management and advance the culture of data driven decision making.

Participants mentioned that Data science projects are still scarce in South Africa. Therefore, Data science teams do well when they establish communities of practice and share learnings beyond the organisation, even establishing international ties. However, teams need to be discerning and balance sharing enough and sharing too much, which could risk disclosing the organisation intellectual property and compromise competitive advantage. The caveat is that teams that do participate in communities of practice may struggle to keep abreast of technical developments locally and internationally.

The development of analytical models is an iterative process that participants have described as experimental. Participants mentioned that the Data science platform must be built with a sandbox environment which its sole purpose is to experiment with ideas. This provides a safe environment where data scientist can create and test model and delete them when required without affecting business.

"*You need people who are playful. You need people to experiment because it is all about experimenting. You have to emphasize that making a mistake is not the end of the world*." (Participant 2)

The findings point to the need for quality management processes that are introduced during the project phase to be operationalised. The team must take charge of ongoing improvements. The key aspect of this is the ongoing improvement of the data quality score. The findings revealed that the score would taper down when new data sets with low quality are introduced. However, the overall trend would improve if the team puts due effort in improving data quality. Over time, with the assistance of the Data architect, the team must get to automate validations that prevent new data quality issues coming into the platform, which also help improve the overall quality of the platform.

The team, guided by the data architect, work on strengthening data engineering practices, which include sourcing, integration with data providers, automation of the process, and implementation of data quality validations. As seen above, Data engineering typically transitions from a data warehouse platform, it is therefore important that data engineers learn new skills made available by Data science toolkit. The findings show that there is a tendency for some teams to stick to what is familiar and fail to exploit advanced tools for solving problems. Data science teams often compliment the skills they do not have internally with the use of external consultancies. The use of consultancies must be managed well to ensure effective transfer of skills to internal Data team.

5.5 Results for research questions three

Research question 3: What are the ethical considerations emanating from collecting and using data for Data science applications?

In connection with research question 3, participants were asked questions about: their views on critical ethical considerations connected to data; legal and regulatory requirements that organisations are expected to comply with; and risks posed by ethical risks to the customer as well company brands.
Leveraging data exposes organisations to a range of moral and ethical issues. In this section, a range of ethical and legal issues are presented as narrated by the participants. It can be argued that identifying and addressing ethical concerns is important to Data unit success because failure to do so may lead an organisation to moral conflict with customers and society and damage the organisation’s brand. The data capability must be set up to enable a competitive advantage for an organisation and not to harm it. It, therefore, follows that if leaders and practitioners of Data science do not pay careful attention to ethical issues, there will be collateral damage

Participants were asked questions about their knowledge and experience of moral and ethical issues involved in data practices. The research questions focussed on practice within organisations, but some questions extended to participants personal experience and attitude towards personal data abuse to amplify the issues involved.

5.5.1 Data Ethics and Governance

Ethical and Moral considerations

As described earlier, Data Science is a field of artificial intelligence. Many participants alluded to a stigma attached to artificial intelligence (AI) as some people view it as a threat to jobs. Participant 10 confirmed this view “*in fact that is the reality of AI implementation - it takes away jobs*”. Participants acknowledged the need for organisations to improve efficiencies to ensure competitiveness; however, it was pointed out that this must be counterbalanced by a due regard for the goals and livelihoods of employees. The stigma around AI can translate into Data science as some participants pointed out that the two fields are related. The change management effort discussed above must manage this risk. Furthermore, this is a challenge that calls for leaders to lead with integrity, and to recognise their triple bottom line responsibilities.

The findings revealed several ethical issues and risks pertaining to how organisations use the data they collect. Organisations have the responsibility to conduct themselves with integrity in how they use the data they collect. Participants said ethical conduct and integrity means that organisations must obtain permission for collecting data; disclose the reasons for collecting the data; disclose the data they are collecting; they must use the data only for the reasons they specified. The organisation must develop governance processes for enforcing these practices without being forced to do so by regulators. Clearly, adhering to these recommendations strengthen the Data science capability. It can be further argued that a Data science capability that is effective while adhering to ethical practices will be sustainable in the long term. Adhering to ethical practice will also protect the organisation from reputational risks.

Participants pointed out that there are organisations whose whole business models are based on selling people's data in a manner that does not observe acceptable ethical conduct and integrity. With the introduction of data privacy laws, these organisations business models are under threat.

"*A lot of countries are introducing regulations that talk to that data. And remember, this is going to upset the applecart, right? There are businesses whose business models revolve around selling data about me and you, and that cannot continue"* (Participant 4)

It is therefore critical that the ethical practice is built into the foundation a Data science capability. Furthermore, the findings pointed to the need for organisations to be trained in ethical conduct. One way of advancing this would be to include data ethical conduct awareness as part of the change management efforts.

There are many specific issues that the findings uncovered the key ones include nuanced issues that can be hard to for organisations to pick unless they are alert to them. Some data set could contain data that would result in prejudice if left unchecked. As an illustration, voice recognition models trained on data obtained from overseas English speakers may fail to recognise South African voices, due to differences in accents. If such a model is used to grant some form of a service, like a call centre menu, certain people could be unintendedly denied service. Some models can fail to recognise the faces of a particular race group because the underlying training data is heavily skewed towards faces of another racial group. The Data team must study the underlying data in order to remove potential ethical land mines.

5.5.2 Data legislations

The findings highlighted the need for organisations to adhere to laws governing data conduct in all the countries where they operate. Laws like GDPR in Europe and POPIA in South Africa stipulate strict measures that organisations under the jurisdiction of these laws need to comply with. Both the GDPR and POPIA mainly focus on the protection of data of persons. These laws deal with issues of privacy and security. Privacy laws are geared to safeguard individuals from having their data exploited without their express concern. Security laws provide for a strict measure that the organisations that collect data must adhere to in order to ensure safety and protection of people’s data from risks like theft and misuse.

“The GDPR is a very comprehensive set of laws. We have observed it forced the companies in Europe to improve the maturity of their data practices much faster than other countries around the world... in fact, the South Africa POPIA act borrows heavily from GDPR, so you can expect South African companies to begin making significant improvements in their data practices” (Participant 4)

Overall, a resonating message echoed by Most participants is that compliance to legislations should be the minimum standard. Organisations must hold themselves to the highest moral standards as a matter of cause, without being coerced by law. Companies that are found short in their ethical practices do not only risk the consequences of breaking the law, but they risk damaging their brand equity as well.

6. Chapter 6 – Discussion of Results

6.1 Introduction

This chapter discusses the findings presented in Chapter 5 in details. The results presented in Chapter 5 were based on 13 interviews conducted with subject matter experts. Additional data was collected from three Kaggle community projects which are relevant to the research problem. In this chapter, these results are discussed in conjunction with the literature reviewed in Chapter 2.

The discussion will be framed in the context of the problem statement discussed in Chapter 1; pertinent elements of the research methodology; as well the research questions posed in this study. The chapter is structured with the three research questions forming the major sections and the detailed finding discussion falling within this structure. The goal of this chapter is to establish the extent to which both the theory reviewed and data collected, address the research questions.

6.2 Research Question 1: What are the factors affecting sense-making and strategic decision making in a dynamic environment?

Research question 1 sought to uncover the factors that drive leadership sense-making, specifically in a dynamic environment. Most participants took a long-winded view when providing answers related to this question. Participants stressed that first and foremost, the vision must be in place. The rest of the organisation must be aligned with this vision. The strategic goals must be set in alignment and pursuit of this vision. The findings revealed that the complexity in the dynamic environment is exerted from outside-in. Therefore the organisation must be stable in terms of its strategy and alignment on how all internal structures should respond to the external forces exerted by the environment.

6.2.1 Strategy formulation processes and the environment factors

Most participants emphasised the need for leaders to create a common identity. The organisation was found to be at the core of an organisation's reason for existence, and participant said everything that the organisation does must stem from there. The findings showed that strategy formulation is the ultimate decision-making process; it is vital to keep these decisions in sync with the vision of the organisation.

The participants said the entire process of setting strategic goals and objectives involves making decisions on what the organisation will do and what it will not do. All though decision making permeate all levels of the organisation; the participants emphasised that decision made at a strategic level affects everything else, including the future of the organisation. These views are consistent with Colombo and Delmastro (2008); and Van den Steen (2017) who mentioned it is the strategic decisions that have a lasting effect and determine the success and competitive advantage of the organisation. The long term nature of these decisions often meant that they are irreversible (Van den Steen, 2017). Clearly, the factor of the irreversibility of strategic decisions should provide the impetus for improving their accuracy. While it could be argued that aiming for absolute accuracy is unreasonable, organisations should still for high level of accuracy.

Participants identified two kinds of interventions that improve the quality of decisions leaders make. The first is the practice of following a formal process and applying frameworks and tools in making decisions. The second is leveraging data to improve the leaders' ability to make sense of the environment; declutter the noise, and focus better at the real problems at hand.

Human beings have an inherent limitation is the amount of complexity they can deal with at one point in time; as such, leaders as any human beings, struggle to process information influx (Van den Steen, 2018; Rosenhead, Franco, Grint, & Friedland, 2019). The literature showed that their human beings have a limit in rational reasoning (Robbins and Judge, 2018; Rosenhead, Franco, Grint, & Friedland, 2019). This does not discount the value of intuition in decision making, which has been acknowledged to be valuable in certain instances (Calabretta, Gemser, & Wijnberg, 2016). The purpose is to highlight the presence of the inherent human limit in applying rational thinking. Leaders ought to appreciate this constraint.

The results showed that companies were engaging in continuous assessment and analysis of their external dynamic environment. The external environment was reported to be in a constant state of flux. This made it difficult for organisations to put a long term plan. Increasingly plans have to updated in quick successions to accommodate changes that are continually emerging. The findings show that this is the second part of the source of complexity, in addition to inherent human limitations discussed in the preceding section. Both human limitations and environmental turbulence impedes decision making.

Participants emphasised that the process of strategy formulation is essentially about making decisions. Related to this point, participants said strategic decisions are different from operational decisions. It was clarified that operational decisions are made daily and help guide the day to day operations of the business; in contrast, strategic decisions are made through a planned process that helps guide the long term direction of the organisation.

The findings showed that making decisions, whether operational or strategic, is hard enough. However, strategic decisions were significantly more challenging and require mode support. This is because strategic decisions involve significant investments in the chosen decision, and the risk of choosing poor options have long term negative consequence on the organisation. Participants added that strategic decisions require choosing one course instead of another, and often those decisions are not reversible. One participant said this was the reason why the organisation must rely on more than just gut feel to make strategic decisions.

These findings corroborate the literature reviewed as part of this study. The process of decision making is described as the central tenet of every organisation's operational and strategic mandate (Colombo and Delmastro, 2008; Van den Steen, 2018). Organisations are said to live and die by the decisions they make (Colombo and Delmastro, 2008; Van den Steen, 2018). Furthermore, as pointed out participants, the theory agrees that organisations' long-term success and competitiveness depend on the quality of decisions they make. To support decision making, organisations have devised constructs such as organisational structure design, governance, committees, policies, etcetera are instruments that facilitate decision making (Colombo and Delmastro, 2008).

The theory advanced by Christensen and Knudsen (2010) suggest that organisational structure has an impact on decision making. According to this theory, hierarchical structures accord better oversight; however, compromise the speed of decision making (Christensen and Knudsen, 2010; Van den Steen, 2018). Central to this theory is the predictable systemic issues that structural design choices create - flat design structures tend to reduce Type I errors. In contrast, hierarchical designs tend to reduce Type II error decisions. Christensen and Knudsen (2010) described a Type I error as a decision that rejects a plausible option, and a Type II error as a decision that affirms a flawed option (Christensen and Knudsen, 2010).

However, this perspective did not emerge from the findings. No participant acknowledged the impact of the organisational design aa a critical contributor to the quality of decisions leaders make. Even the few comments made some of the participants in this regard were superficial and did not address the impact of structure on decision making.

In addition to cognitive limitations raised above, participants cited environmental factors as another key issue in decision making. Participants said it was imperative that leaders have sufficient knowledge of their operating environment in order to inform their decisions. This involves the need for leaders to form business ties with ecosystem partners external to their organisations. It also involves scanning the environment to make sense of opportunities and threats therein, so that the organisations exploit opportunities and stay ahead of the competition.

In the Financial Services industry, organisations are also constantly on the lookout for threats emanating from Fin-tech and other start-up organisations. What compounds the issues is that most threats originate from organisations that are outside of the Financial Services industry. Often these organisations apply novel business models that address customers’ problems. The findings show that it is often difficult for incumbent organisations to tackle those challenges due to legacy systems constraints. However, some participants argued that despite the constraint, the leaders in financial services are not alert to these threats until it is too late. The findings are clearer, however, that this is a general issue that extends beyond financial services.

In order to mitigate this challenge, the literature states that organisations require environmental scanning and sense-making capabilities. Sense-making is defined as the process through which an organisation identifies opportunities and threats in a dynamic environment (Teece, 2014). In a dynamic environment, opportunities and threats do not present themselves in clear and obvious ways (Eisenhardt and Martin, 2000). This is due to the rapid pace of change and the short window of opportunity characterising the dynamic environments (Eisenhardt and Martin, 2000). Therefore, this theory may explain why leaders in financial services are reported to struggle with identifying threats in their environments until it is too late.

Furthermore, the literature showed that past experience does not count for much in a dynamic and complex environment (Snowden and Boone, 2007). Therefore, leaders must also learn as they do, through experimentation and probing (Snowden and Boone, 2007). However, leaders do have recourse on these issues; they can either secure tools from the market or build them internally. One category of tools falls within what has been referred to in the literature as a Decision Aiding process (DA). A DA is a framework for problem-solving that seeks to mitigate the impact of bounded rationality in decision making. The DA remove or minimise ambiguity in the appraisal of decisions given the dynamic environment context that participants have alluded to (Tsoukiàs, 2007). A DA is an abstract framework which is useful for providing theoretical guidance but would not directly address the practical concerns until it is made concrete (Tsoukiàs, 2007; Meinard and Tsoukiàs, 2019). The findings support the implementation of such tools with most participants having stated that organisations need to increase their reliance on data, tools, and technology and begin to adapt advanced analytics to help mitigate these problems.

At this point, it should be pointed out that both findings from the results and literature review have addressed the research question 1. Research question one sought to uncover the critical factors that impact on sense-making and thereby decisions making. The findings presented in the section and preceding chapter agree with literature that the key factors the dual issues of human cognitive limitation and environmental dynamism (Van den Steen, 2018; Rosenhead, Franco, Grint, & Friedland, 2019).

6.2.2 The role of leadership

The findings demonstrated that the role of a leader is a complex role and vast in scope. Participants painted a list of knowledge areas that successful leaders must be vested in to be successful. These include: a deep understanding of the operating environment; understand organisational drivers and constraints; advanced awareness of the resources and capabilities in their disposal; understanding and close proximity to customers; and leaders must understand the culture within their organisations. Once they understand these areas, it is the leader's job to influence all of these areas for the attainment of organisational goals.

The findings pointed out that since leaders are accountable for their organisations, the strategy is, therefore, their responsibility. While the strategy is the mandate of the leaders who serve in the board for directors, in practice, this responsibility was shared with executives and senior management. However, the findings highlighted the need for leaders to seek input from all relevant all key individuals when formulating strategy - irrespective of the level of these individuals. Related to this, most participants stressed the importance of prior preparation ahead of strategy meetings. Reports and data are the critical inputs that participants said leaders needed to make sure are brought into decision making. As such, leaders ought to know the critical input providers and ensure their inclusion.

The findings showed that there was a strong case for leaders to rely on data to help them deal with the complexity of making decisions. The findings further showed that if the leaders at senior level do not demonstrate the importance of data in decision making, the culture of relying on data would not be entrenched. The literature stipulates that the culture that eventually emerges is directly related to what leaders do; if the leaders tend to base their strategic decision on data, the rest of the organisation learn to follow suit (Christensen and Knudsen, 2010; Mackay and Zundel, 2017)

Participants pointed out that there are leaders who make decision predominantly based on their experience and gut and feel. Yet, some participants pointed out that the older generation of leaders tended to not data for making the decision, instead rely mainly on experience. However, it was not clear from the literature reviewed whether the leader's age is a driver for propensity to use data in decision making or not. However, the literature does cover the concept of gut-feel based decisions. Calabretta, Gemser, & Wijnberg (2016) argued that there was a space for intuition in decision making, in as much as there is a space for rational decision making. However, the authors leave it to the leaders to discern when to use which approach.

It is clear that while data and tools can be made available to assist in decision making, leaders need to take the lead in incorporating that practice into their organisations. The action of leaders plays an important role in developing a particular culture; include the culture of making decision driven decision. As such, the leaders' own behaviour is the third factor that affects the development of sense-making capability in organisations.

6.3 Research question 2: How do Data science practices contribute to the creation of sensing capability and competitive advantage within an organisation?

This question was intended to solicit, from both theory and data collected from participants, the set of practices within the field of data science, that results into creation of dynamic capability – this is the explicit part of the question. The question also contains an implicit proposition that the existence of the sense-making capability should lead to competitive advantage. The study endeavoured to investigate answers to both the explicit and implicit parts of the equation, and the findings are presented in this section.

6.3.1 Data science processes and practices

This section focusses on Data science process and practices as articulated participants, and as recorded in literature. The next section presents the findings on dynamic capabilities and the role of Data science in enabling them.

Configuration the Data science capability

The findings show that the foremost consideration in configuring Data science capability is the setting of the Data unit. However, only a few participants have had an of establishing a new Data unit although most of them reported having been involved with some form of transformation of existing Data units in their careers. Most of the participants stated that their organisations were in transition, from the old data technology involving the data warehouses to the new Data science and its suite of technologies. Furthermore, the findings show that the transformation efforts to Data science and data analytics include the old technology stack, which is dominated by data warehouses. Participants reported that advanced Data science technologies coexist with old ones. Participants also mentioned that it was essential to implement the change in a gradual manner. This ensured that the environment was operational while new technologies and features were being added. In essence, the findings demonstrated that the building of the Data science capability entailed the configuration of existing Data unit to add advanced technologies and processes which culminated into a Data science capability.

These findings confirm the theory which stated that Data science encapsulates the processes, principles, tools, and technologies for analysing data (Provost and Fawcett, 2013; Meinard and Tsoukiàs, 2019). The journey to Data science is an evolution of existing technologies data practices (Schmarzo and Schmarzo, 2013). However, there are cases where some organisations do not have an existing data unit and wish to establish data science, like in the case growing small enterprises and start-up organisation.

The findings highlighted the selection of the data team as key consideration in Data science implementations. The finding demonstrated that cross-functional skills set were more valuable than deep technical skills. The transformed Data unit should include people with technical skills as well as those with business skills. This represents a departure from traditional data warehouse data teams' configuration, where business stakeholders were outside the Data unit. Additionally, it was customary for Data teams to have extensions that involve working committees which consolidates skills, expertise, power and influence across the organisation. The importance of this multidisciplinary configuration was significant for problem solving and collaboration as well for change management and marketing of the Data science capability across the organisation. This last point is dealt with in detail in the next sections.

These findings support the literature which established that effective teams form horizontal and vertical ties across the organisation (Tonidandel, King, & Cortina, 2018; Lee, Inceoglu, Hauser, & Greene, 2020). Additionally, the theory state that Data science team cannot function without multidisciplinary skills as the work required to involve fields like mathematics, statistics, computer science and business skills, which is hard to find in one individual (Tonidandel, King, & Cortina, 2018).

Problem identification and analysis

The findings show that Data science practices are more robust and detailed than the traditional practices. Data warehouse driven business intelligence reports were developed through gathering requirements from business users and documenting these in specification documents after that reports would be developed and provided to users. However, Data science emphases collaborative problem-solving. With a clear definition of the problem as the first step.

The findings showed that the problem solving process involved a clear definition and analysis of the problem. This often pulls various collaborator with the team from business stakeholders, who have a deep understanding of the domain area where the problem exists. Once the problem has been clearly articulated, hypotheses are developed. The Data team then work at sourcing the relevant data to help test the hypotheses.

Participants added that the Data team would then perform analysis, develop models to test the hypotheses and produce results. The final stage would be the reporting of the results first to the target business audience. In summary, the findings distilled the Data science product development cycle into the following sequence of steps: 1. Problem definition; 2. Data sourcing; 3. Analysis; and 4. Reporting or visualisation. The findings clarified that Data science projects do not only produce reports; as such, some participants referred to Data science output as a product. The findings support what Hindle and Vidgen (2018) in that the problem definition is typically distilled into concreate research problems, research questions, or hypotheses. The findings also identified similar stages in development Data science products: following the problem definition, the second stage involves collaboration among cross-functional teams or brainstorming within the Data science team (Hindle and Vidgen, 2018). Data science efforts often require a diversity of skills involving programmers who write the software; statisticians or mathematicians who develop and validate the data models; and business stakeholders who serve as subject matter experts (Hindle and Vidgen, 2018).

The findings showed that the end result or the output of Data science effort could be a mathematical model; a report, a dashboard or a data stream which can be fed to the line of business systems to automated decisions. These outputs are sometimes called data products as a generic term. This assertion is consistent with literature which stated that data analytics is the essence of Data science outputs and its final product (Newman et al., 2016). There are three distinct analytical products produced by data science, namely, descriptive analytics, predictive analytics, and prescriptive analytics (Hindle and Vidgen, 2018; Conboy et al., 2020).

Findings emphasised that the problem definition step requires the collaboration of interdisciplinary teams. One of the key contributions of the cross-functional and multi-disciplinary team is the identification of the problem boundary space. Data science models provide solutions that are relevant only within the problem boundary space. Therefore it is important to involve people with an intimate understanding of the problem space in the problem definition. As an illustration, a boundary space would specify which customers are affected by the stated problems, like high propensity to churn. Therefore, the rest of the customers, not identified as having a *high propensity to churn*, fall outside of the boundaries of the problem. The final model may suggest an action for the *high propensity to churn* customers. This could include incentives or SMS communications to reduce churn. It is critical that the suggested actions are applied only to those customers in the problem set, which represents the boundary of the problem. The participants stated that addressing a target, like customers, who fall outside the problem boundaries is wasteful to the organisation and may lead to an unintended reaction from the target. The finding further emphases the importance of collaboration in the Data science practice and business stakeholders who serve as subject matter experts (Hindle and Vidgen, 2018).

6.3.2 Analytical model design and execution

The findings show that once the data has been gathered, a team of data scientists begin constructing mathematical models that test hypotheses which were set up during the problem definition. Data science models are predominantly mathematical; they are built using technics in Applied Mathematics, Statistics, and Computer Science to solve problems. Data science model also use techniques in Machine Learning extensively. Machine Learning derives from both Statistics and Computer Science - Artificial Intelligence. These findings are aligned to literature; Hindle and Vidgen (2018) stated that Data science is built on mathematical models. These models are reused until their assumptions are no longer valid (Hindle and Vidgen, 2018)

Participants described the concept of analytical model training. Training involves machining learning techniques which progressively improve the action of a model by feeding it a large amount of data. Eventually, its ability to recognise these patterns improve as a result of processing a large volume of data. The availability of large volumes of data is key to model training. Again this view conforms what literature says; applying Machine Learning which involves automated statistical models implemented through highly efficient algorithms, to discover patterns and insights from the Big Data (Hindle and Vidgen, 2018; Conboy et al., 2020).

The literature state that the final step in the Data science execution process reporting, which is sometimes referred to as Visualisation. In this stage, the patterns uncovered from the analysis of data are presented in visual forms and presented to management. Skilful practitioners are able to convert insights into stories, and these are presented to management to enrich their insights and improve decision making (Hindle and Vidgen, 2018). The findings also described the final output as including visualised report or a dashboard. However, the finding added that the final output sometimes includes a data set or a data stream which is passed to the line of business systems to automate a process.

The major categories of analytical reports were shown to be descriptive analytics; predictive analytics and prescriptive analytics. The findings support the literature which also states that there are three different analytical products produced by Data science efforts: descriptive analytics, predictive analytics, and prescriptive analytics (Hindle and Vidgen, 2018; Conboy et al., 2020). Descriptive analytics classify and categorise things, like placing customers into distinct segments. Predictive analytics, as the most advanced form of analytic, provide forecasts of future trends that help guide leaders concerning options that can be considered for the future. Prescriptive analytics guide leaders on actions to be taken. Prescriptive analytics offer a direct course of action that should be taken, such as spelling out which products must be offered to which customers and in which markets (Newman et al., 2016; Hindle and Vidgen, 2018).

Participants reported the same analytical reports as being the significant categories of Data science products. Descriptive analytics was defined as the simplest product but one which adds value in terms of operational reporting. Predictive analytics was mentioned as representing an advanced level compared to descriptive analytics. Predictive models produce analytical reports to predict future events. The findings show that predictive models are useful for aiding decision making as they process large volumes of data and paint a picture of the future. The third category of analytical models is prescriptive analytics. This represents the most novel of Data science products. Prescriptive analytics provides the ability to make concrete decisions about the reasonable cause of action.

At this point, the ability of Data science to enable sense-making has been demonstrated. The predictive and prescriptive models are argued to be the most suitable for the development of sense-making capability. Through the effective deployment of these tools and underlying practices, organisations can create a Data science driven capability. This assessment begins to address research question 2.

However, there are additional factored that must be considered in the creation of sense-making capability through data science. These include applicable technologies and investment.

6.3.3 Data science skills

Roles and cross-functional skills

As demonstrated earlier in this chapter, the findings confirmed that effective Data science team are multi-disciplinary and cross-functional. People are co-opted from various departments within the organisation to provide business or industry expertise which facilitates problem-solving. These could include people from Finance, Marketing, Operations, or services business units like Research and Development (R&D). The need for collaboration across the organisation is articulated in literature and teams that do well tend to put emphasis strong ties with stakeholders (Hindle and Vidgen, 2018).

It was observed that a typical data scientist has qualifications that include: computer science; engineering; or mathematics educational backgrounds. Many data scientist possess advanced degrees, like masters or doctorate. Participants said there are two critical roles inside the core Data team are they are that of a data architect and a data scientist. These roles were said to be strategic and crucial in building competitive Data science capability. Architects and data scientists internalise vast intellectual property of their organisation, and therefore organisations needed to retain them.

6.3.4 Data science technologies

The technology was observed to be part of the foundational aspect of data science. The participants listed what they said were essential technologies in data science. As a collective, the Data science team must possess knowledge of data processing algorithms, mathematical modelling, statistics, machine learning, and more. They need to be skilled in programming languages like SQL, R, Python, Scala, Java, etcetera. They need to be adept at working with technologies Amazon web services (AWS) or Microsoft Azure, Spark, Hadoop, etcetera. Table 6 below sorts out the artefacts of traditional and contemporary data practices:

Traditional set-up and technologies (Data warehouse based)	**Contemporary set-up and technologies (Data science)**
• Business intelligence produced descriptive and predictive reports • Based on structured internal data • Textual data format • Advanced analytics + Data Architecture + Data • Cube • SQL and MDx query language • Data Warehouse	• Data science produces predictive and prescriptive analytics • Used internal and external data with structured or unstructured forms. • Variety of data formats, text, images, voice, and video. • SQL is still relevant in the contemporary toolset • AWS, MS Azure - Cloud is scalable and generally more secure than on-premise infrastructure • Programming language: SCALA, SQL, R, Python • Data Science Tools - Spark, Java, Hadoop • Gradient Boosting Machine (GBM) • Spark to handle multiple data sets • Skills – mathematics, statistics, machine learning, algorithms design, business domain knowledge

Table 6: Comparison between data warehouse and Data science technologies and artefacts

6.3.5 Investment in Data science capability

Emerging studies have demonstrated that investing in Data science improve organisation ability to deal with a complex problem and therefore make informed decisions (Mikalef et al., 2019). The findings support this literature; most participants said there was business value in investing in advanced analytics and data science. The Kaggle projects discussed above reveal practical anecdotes of how organisations have derived benefits by employing data science. In the Kaggle project 1, commissioned by Ford motor company, the Data science model analysis sensor data installed in a car to decide driver alertness level. Based on the outcome of the model, an alert system could be developed to warn the driver of the issue. Kaggle 2 uses Data science to assist a company in the rental business to project the level of tourists demand correctly. Kaggle project 3 was built for the medical research community to help them find the most relevant COVID 19 related articles to help them fight the surge of COVID-19. The team that worked on the Kaggle project 3 built a search engine that shows the best article for intended search alongside closely related articles to the top article. The project team claim their search works better than some of the academic search engines available.

The researcher did not verify the validity of the claims made in these projects, however, relied on the fact that each of these projects won the competition in their categories and were paid by the organisations that commissioned the projects. This provides a level of confidence on the value they added. The evidence provided literature, the findings obtained from interviews and the anecdotal evidence from the Kaggle projects, point to the value of Data science that cannot be easily achieved by other tools readily available today. Therefore, it can be concluded that the value of Data science is demonstrable.

The findings revealed that Data science capability is underpinned by a technology platform which can cost in excess of R10 million and significantly higher for large organisations like in big banks in South Africa. However, participants reported that despite the advantages demonstrated by advanced analytics, organisations are slow in accelerating adoption. Added to this, Data teams struggle to produce business cases that justify the required investment in terms of monetary return on investment (ROI).

The findings showed that Data team faced a dilemma when trying to justify project investment. In order to justify the investment, the Data team needs the platform and a sandbox environment to conduct experimental analysis and produce prototype analytical models. The prototypes are used to quantify potential ROI. Until the sandbox environment or platform is available, and the Data science skills are brought on board, it is almost impossible to prove a financially viable business case. Participants called this dilemma a chicken and an egg paradox.

According to literature, organisations are not totally unjustified in their hesitancy to invest in Data science projects. Studies have found that organisations faced challenges getting a return on investment on Data science projects (Vidgen, Shaw, & Grant (2017). However, the main obstacle is the lack of a cohesive guiding framework (Mikalef et al., 2019).

Some participants corroborated literature on this view, stating that about 70% to 90% Data Science projects fail. This is, however, found to be in line with the information technology project failure rate, which was reported by participants to be above 85%. However, the findings contextualised this statistic: the failure rate is calculated based on the number of projects that miss the original deadlines; exceed the initial budget, or do not stick to the original scope and quality. However, most of these projects are eventual pulled through and completed only not according to initial metrics. The clear message presented by participants is that Data science projects, like information technology projects, are prone to failures and must be executed with utmost commitment.

Both the findings and literature, Mikalef et al. (2019), raise issues that are pertinent to the aims of this research project which sought to develop a theoretical framework that will aid the development of effective Data science enabled capability. Section 6.5 discusses the proposed framework for the creation of Data science enabled sense-making capability.

6.4 Research question 3: What are the ethical considerations emanating from collecting and using data for Data science applications?

Research question 3 sought to uncover ethical risks facing organisations as a consequence of leveraging data. This question was though necessary as failure to mitigate the related risks, could undermine the sustainability of the Data science enabled capability and potentially expose the organisation to intractable risks. Therefore this section draws from theory and the findings to frame the ethical issues and help answer research question 3.

The findings showed that ethical consideration should take high priority in the leadership agenda and how organisations leverage data. Furthermore, the findings showed that mistakes made when leveraging data could expose organisations to brand damage and legal risks. This finding supports the literature, which provides that leadership has a responsibility to ensure that ethical practice permeates every aspects of their organisation (Zeni, Buckley, Mumford & Griffith, 2016).

6.4.1 Data Ethics and Governance

Ethical and Moral considerations

As described earlier, Data Science is related to the field of artificial intelligence. Many participants alluded to a stigma attached to artificial intelligence (AI) as some people view it as a threat to jobs. Participants acknowledged that organisations do need to improve efficiencies in order to ensure competitiveness. However, the findings pointed out that this ought to be counterbalanced by a due regard for the goals and livelihoods of employees. This is a challenge that calls for leaders to lead with integrity, and to recognise their triple bottom line responsibilities.

The findings stressed the importance for organisations to conduct themselves with integrity in how they leverage the data they collect. Integrity was described as the practice that includes disclosure, transparency and honest, in terms of using data for only the purposes the organisation promised it would use it for. It is evident from the findings that adhering to these recommendations improve the strengthen the Data science capability. It can, therefore, be inferred that a Data science capability that is effective, while adhering to ethical practices, will be sustainable in the long term. Adhering to ethical practice will also protect the organisation from reputational risks.

These findings confirm literature, which implores organisations to put measures in place to prevent the abuse of data in their disposal whether those abuses are intentional or not; the duty lies within the organisation to safe keep and applies fair use principles concerning data (George and Osinga, 2016; Meinard and Tsoukiàs, 2019). The findings highlighted that those organisations whose business models are based on inappropriate use of people's data are unsustainable. With the introduction of data privacy laws, these organisations business models risk being put out of operations by regulators.

The findings revealed nuanced issues that can be hard for organisations to pick. This, therefore, calls for advanced controls to be put in place. Participants pointed out that some data set could contain data that would result in prejudice if left unchecked. The analytical model may fail to recognice local faces and voices for a particular ethnical group, if that model was trained on a data set that is biased towards another ethical group. The findings cautioned that both the data team and data committees must continuously interrogate and weave out these unintended problems to protect their brand image and mitigate the risks of legal challenges associated with these issues.

6.4.2 Data legislations

Increasingly data use and data operations are being legislated to safeguard the interest of the owners of data from harm. However, organisations must promote best practices that address data privacy, security and integrity issues above and beyond legislation (Baesens et al., 2016; Tadewald, 2019). The findings supported these views and indicated that many standards and practices have been published over the years, which set high standards than the prevailing laws on data.

Participants stipulated that Laws like GDPR in Europe and POPIA in South Africa stipulate strict measures that organisations under these respective jurisdictions need to comply with. Both the GDPR and POPIA focus on the protection of data of persons. These laws deal with issues of privacy and security. Privacy laws are geared to safeguard individuals from having their data exploited without their express concern. Security laws provide for a strict measure that the organisations that collect data must adhere to in order to ensure safety and protection of people's data from risks like theft and misuse.

The proposed framework for the creation of Data science enabled Data science capability, includes ethical consideration in its foundation. It proposes, in line with findings, that compliance to law should be the minimum standard. It proposes that organisations must adhere to higher standards from the inception of implanting a Data science capability

6.5 Data science enabled capability framework

The section describes the proposed framework for creating a Data science enabled capability for sense-making developed by the author of this research. The design and the content of this framework draws heavily from the literature reviewed as part of this study and similarly from the results of the data collected as from research participants. To a lesser, extent aspects of the triangulation data from the Kaggle projects have been included.

The literature revealed that organisations were going through rapid adoption of Data science as a strategic capability. However, these projects were marred with issues of delays, budget overrun, and not meeting the needs of the project sponsors; these issues are viewed as project failures (Zeni, Buckley, Mumford, & Griffith, 2016; Mikalef et al., 2019). The findings from imperial results placed the failure rates of Data science projects in a range of 70% to 90%. Perhaps, owing to these issues, certain companies were reported by the findings to be reluctant to adopt data science, despite the highlighted advantages.

Furthermore, literature has demonstrated that the field of Data science is still emerging (Hindle and Vidgen, 2018; Lekakos and Krogstie, 2019). This can be taken to imply that organisations and their Data science teams are still undergoing a learning curve, thus the reasons for the high rate of project failure. It is therefore clear from both the theory and empirical data that there is an urgent need for an intervention. There is a need for a framework that is grounded theory and empirical data, that would help guide Data science teams on the best practices for transforming their data practices into fully-fledged Data units. The framework spells out the methodology for transforming legacy data units into an effective Data science unit. This is the goal of the Data science capability framework explained below.

This framework applies to transformation efforts where the organisation is starting with a baseline data capability. The model may not be suitable for cases where an organisation is assembling a data capability from the ground up. This limitation is partly a consequence of the limit in empirical data obtained during the interviews. Only two participants, who are both consultants, have had the experience of building the advanced data capability from the ground up. This was deemed insufficient imperial evidence to develop a holistic framework. The research is satisfied that most established organisation face a scenario where they are transitioning or wish to transition from one baseline state to advanced data capability.

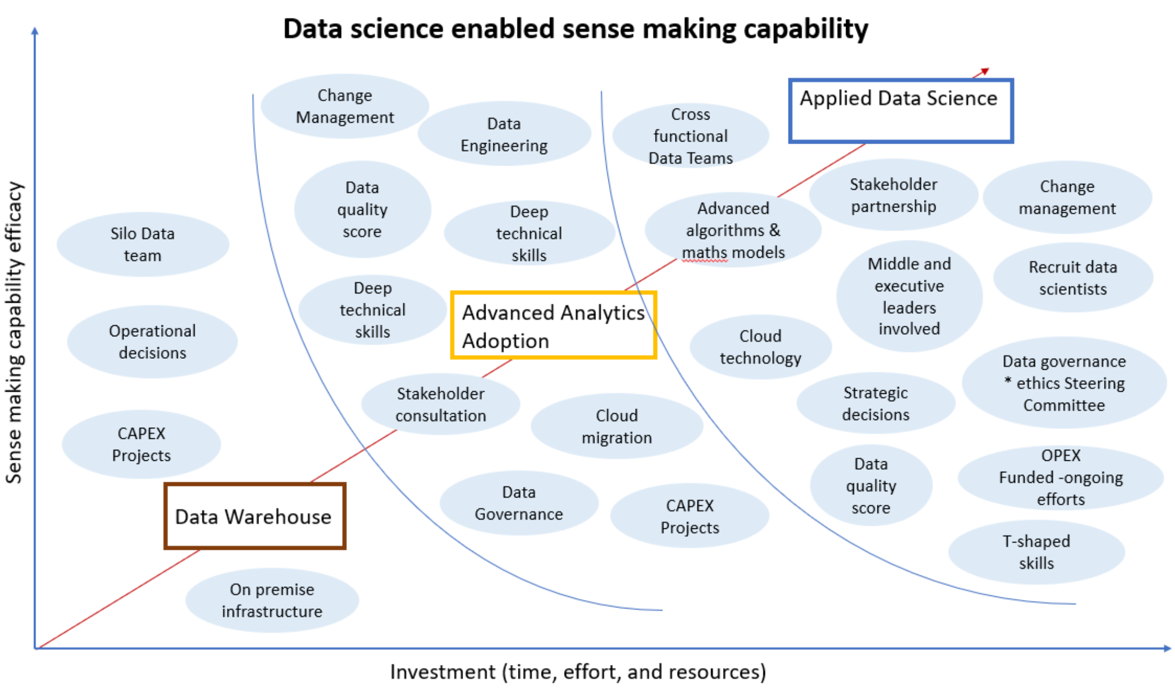
6.5.1 Developing Data science enabled sense-making capability

The Data science enabled capability (DSEC) framework for strategic sense-making is exhibited in Figure 3 below. The DSEC frameworks spell out a methodology for transforming traditional data capabilities, that are based on data warehouse and transactional reporting solutions, into fully-fledged Data science environments. The DESC encompasses requisite transformation on people from, process front and technology front – this aspect is capture in Figure 3: Data science capability framework below.

It should be noted that both Figure 3 and **Error! Reference source not found.** illustrate aspects of the same framework. Primary constructs within the DESC are represented by blue elapses. They are shown in Figure 3, while Figure 4 serves the purpose of elaborating the details that could not be presented in Figure 3 for fear of cluttering the model. To clarify, the DESC framework is contained in Figure 3 with the next figure serving as details.

The DESC is a qualitative framework constructed using theoretical and empirical qualitative information. The progression along both the vertical, and horizontal axis represent progress from traditional to advanced data capabilities. However, this is the best effort to estimate based on theory and results findings. On the vertical axis (Y axis), the DESC depicts the increasing efficacy of the sense-making capability. On the horizontal axis (X axis), the DESC depicts the increasing investment in terms of time, effort, and resources, that the organisations may commit on the transformation journey.

Figure 3: *Data science capability framework*



Looking at Figure 3, the typical transition, as reported in the finings consist of advancing the organisation along these lines: *data warehouse* is the baseline platform; *advanced analytics adoption* is the second, advanced stage; and *applied Data science* stage is the most advance stage. The red straight-line represents the theoretical transition path. The two blue curves that move from top to bottom in rightward direction are demarcations that represent boundaries between the maturity stages. The blue ellipses represent the processes, technologies, skills set, and all artefacts that are applicable within the stage. In brief, there are three distinct stages of maturity: *data warehouse, advanced analytics adoption,* and *applied Data science* – in that order.

The investment dimension represents the time, effort, and resources the organisation can commit to a transition project. The more investment is committed, the more the data capability moves from left to right along the red line. The implication, which was established form theory and empirical data, is that investing in data capability increases the sophistication of reports that the Data team can produce, which help improve the efficacy of decision leaders make. Therefore the input variable is an investment.

The Data Warehouse - maturity stage 1

As stated above, this is the baseline data capability the DESC considers. The data warehouse is reported to have been around for more than 20 years. They are useful data technologies that consolidate data from several line-of-business systems into a single reporting platform. The data warehouses predominantly contain organisations' internal data. The processes and artefacts involved in this stage are discussed next.

The data warehouse stage is characterised by a data team that is self-reliant and work as a silo. While the team consults and receive inputs from stakeholders in the business, the team is essentially independent of all those stakeholders. The team most of its decision and bring others into its design decisions on discretion. The independency provides a sense of autonomy and may allow the team to deliver relatively rapidly; however, the deep business expertise that resides in business stakeholders do not filter in as a result. That put limitations in terms of holistic development on the team and individual within.

The reports the team deliver are used for operational decision making. They are often not geared for strategic decision making. However, the data warehouse platform as a technology has the ability to support analytical reports, but that represents progression towards maturity stage 2 and will be discussed in the next section.

The data warehouse platform is often deployed on the servers that are owned, hosted, and maintained by the organisation. This represents substantial infrastructure capital expenditure (CAPEX) investment. However, the cost of ownership does come down in well-maintained environments where breakdowns are few.

To transition from maturity stage 1 to maturity stage 2 requires the organisation to begin implementing the process, technology and artefacts depicted in stage 2; see Figure 3 and

Figure 4**Error! Reference source not found.**. An organisation at any stage encapsulate the artefacts of the previous stage. That is, if an organisation is in stage 2, it still has the processes, technology and artefacts if stage 1. Therefore the DESC represent accumulative progression. An organisation must deprecate redundant artefact and technology, as an illustration, an organisation in maturity level 3 that has implemented a cloud platform should deprecate it is an on-premise data warehouse, as this is redundant and costly.

Governance is an application to all maturity stages. Implementation of governance standards and ethics controls cannot wait until the organisation has progressed up the stages. However, it is was deducted from empirical data that in stage 3, ethical risks are more pronounced, and there is a need to strengthen controls at this level.

Advanced analytics adoption - maturity stage 2

Stage 2 represents a commitment by the organisation to rely on data for decision making. While this is not the most advanced stage, capabilities at this stage allow for strategic decision making. At this stage, the team exhibit more significant collaborative behaviour than in the previous stage. Based on the data collected and literature review, a high number of large organisations can be thought to be at this level; with many putting plans in place to transition onward to stage 3.

Data quality a data engineering feature strongly in stage 2. Data quality encapsulates processes for ensuring the reliability of data used to build analytics and make decisions. The level of data quality is critical for the credibility of the data team. Data quality should be disclosed to all stakeholders using the data quality score metric. Data quality score should not be viewed as an indictment on the data team. The findings pointed out that the quality score fluctuates over time. In certain instances, as the team strengthen the data capability by bringing additional data sets, the quality score may decline – and this should not be a cause for alarm.

What is imperative is that the team should implement Data engineering practices for continuous quality improvement. Data engineering practices must be applied to validate initial data ingestions where possible. Additionally, the data team must implement governance processes that set out architectural standards and best practices that the data team should strive for.

At this stage, it is ideal for organisations to start migrating to cloud technologies which serve the organisation well when it transitions to stage 3.

Applied Data science - maturity stage 3

Stage 3 represents the most advanced data capability and demonstrates the utmost commitment by the organisation to leveraging data for competitive advantage. At this stage, the data team moves from collaboration with business to integration with business stakeholders through the formulation of working committees that encourage corporative problem-solving. Also, change management effort ensure that the data team projects are socialised across the organisation and at all levels of management. This encourages leaders to adopt advanced analytics for decision making.

At stage 3, all advanced analytics are possible – predictive and prescriptive analytics. This implies that at this stage, the organisation can leverage data to: solve complex problems; declutter complexity in the external environment; automate complex decisions that are untraceable for a human mind. Used correctly, this capability should afford the organisation a competitive advantage.

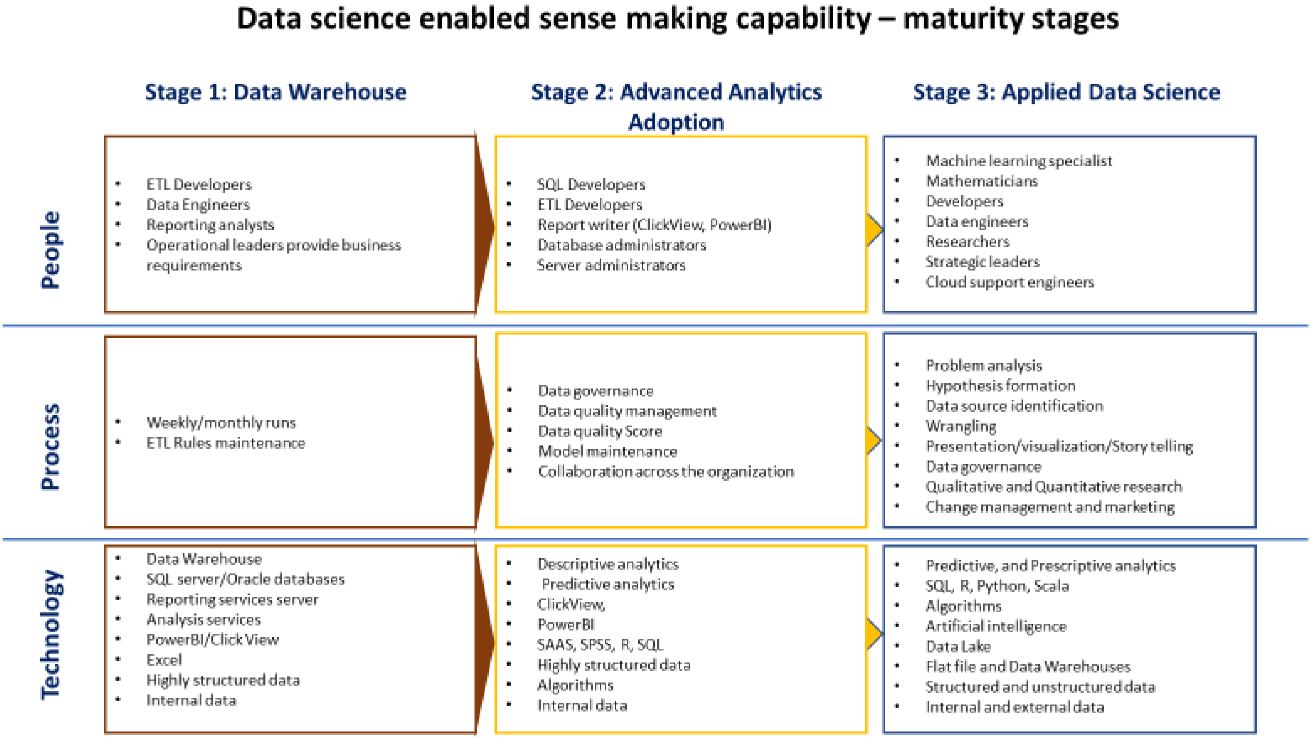
This advancement is enabled by the inclusion of advanced skills in the core data team. These include mathematicians, statisticians, machine learning and computer scientist; as well as domain experts with deep business skills. This team become the centre of a differentiated competency over time, as they solve more and more problems related to their organisations; they become the target of external poaching. The organisations must do what it can reasonably do to keep these employees.

At stage 3, the infrastructure is almost always hosted in the cloud. Much of the costs are ongoing operational expenditure cost (OPEX) which also indicate the ongoing commitment to the capability.

Stage progression

The next diagram, Figure 4, provides the details of each maturity stages grouped by the three stages, as well as by people, process and technology dimensions. See Figure 4 below. This dimension supports the DESC framework presented in *Figure 3* by providing additional clarifying details that practitioners and researchers can reference to gain more understanding of the framework.

Figure 4: *Data science capability framework – content per stage*



6.5.2 Intended outcome

Ultimately, the reason for the implementation of this framework is to help the organisation and its leaders improve decision making. The change element is included in the framework to maintain that thread thorough the framework, such that while the implementation proportionally technological, the meaning should not be lost as to the strategic purpose of this framework. Change management process within the framework is that glue that keeps all stakeholders informed and involved throughout the transformation process.

6.5.3 Conclusion

The DESC framework was constructed using the literature on capability building and empirical data collected through interviews. The DESC provides a straightforward methodology for both the practical implementation of Data science capability in organisations as well as point of leverage for future academic research in the subject.

The DESC can also be used by the organisation to assess where they currently stand in terms of their maturity stage and to inform their plan in moving to the next stages. While organisations may tailor the framework to meet their specific needs by adding more artefacts, deletion of artefacts must be done with due regard as the researcher only included those that literature and the findings pointed were crucial.

7. Chapter 7 - Conclusions and Recommendations

7.1 Principal conclusions

The success and failure of organisations hinge on the quality of the strategic decision making that leaders make (Colombo and Delmastro, 2008; Van den Steen, 2018). Decision making was compounded by human limitation in processing vast information as well as complexity prevalent in the dynamic environment (Snowden and Boone, 2007; Uhl-Bien and Arena (2018)

In order to cope with this challenge, organisations need to build dynamic capabilities for better decision making (Eisenhardt and Martin, 2000; Kurtmollaiev, 2020. Unlike resources, dynamic capabilities cannot be bought but must be built internally in the organisations. Dynamic capabilities are valuable as they provide the organisation competitive advantage. The empirical findings showed that organisations are increasingly relying on data and advanced analytics that are enabled by Data science to improve sense-making and decision making.

With this background, this research project sought to investigate how sense-making could be improved by applying Data science in order to improve leaders' decisions making and increase competitive advantage. Three research questions were commissioned to help guide the investigation and arrive at the answers to the research problem.

Research Question 1: What are the factors affecting sense-making and strategic decision making in a dynamic environment?

The process of decision making was described as the foremost duty of leaders. The long term success and competitiveness of organisations rely on the quality of decisions they take (Colombo and Delmastro, 2008; Van den Steen, 2018). However, require assistance to keep up with the rate of change and to improve the quality of their decision (Van den Steen, 2018). The complexity in the environment compounds the problem (Snowden and Boone, 2007; Uhl-Bien and Arena, 2018).

Both the literature and the findings confirmed two critical factors as impacting sense-making and decision making the most: cognitive limitations arising from human bounded rationality which puts a limit on the amount of information an individual can process; and complexity arising from the external environmental dynamism which makes it difficult to scan the environment. These conclusions address the essence of research question 1. However, what should leaders do about these issues?

The findings in the literature illustrated the potential of using data and advanced analytics to help leaders cope with both their inherent limitation and to help declutter the environment. What the leaders need is a decision aiding process, that mediates complexity (Tsoukiàs, 2007; Meinard and Tsoukiàs, 2019). The findings highlighted that organisations need to implement decision aiding processes and tools, and those who get better at this task will attain a competitive advantage.

Research question 2: How do Data science practices contribute to the creation of sensing capability and competitive advantage within an organisation?

This question was addressed by exploring, through literature and empirical data, the practices that other organisations have implemented as part of their data evolution projects. The study was able to obtain considerable details even though Data science field is still emerging (Hindle and Vidgen, 2018; Lekakos and Krogstie, 2019).

The Data science capability could be found at different maturity stages, with each organisation either positioned at a particular stage or transitioning between stages. Three stages were conceptualised based on literature and findings: the basic stage which is described as a data warehouse stage; the analytics adoption stage; finally, the applied Data science stage. Transitioning between stages require investment; however, this investment translates into increased sense-making capability, which is taken to confer competitive advantage for the organisation.

The findings also showed that effective Data science capability required advanced skills that include professionals with advanced qualifications in computer science; mathematics and business subject matter experts. These findings support the same view in the literature that mathematics skills set have become mainstream in Data science (Schmarzo and Schmarzo, 2013; Baesens et al., 2016). Traditional skills that were prevalent in the data warehouse environments are no longer sufficient. Data teams need to upgrade their skills and onboard new people with the required skill set.

In Chapter 5 and Chapter 6, the process of Data science products is explained in detail. This process entailed the following sequence of steps: 1. Problem definition; 2. Data sourcing; 3. Analysis; and 4. Reporting or visualisation. At the advanced stages, like stage 3, the Data team form strong collaborative ties with multiple business stakeholders throughout the four process steps. The collaboration helps to improve the adoption of the team's products.

The data team also implements data engineering processes to ensure the quality of the end product. Good data teams maintain a quality score for their data and continuously work on improving it. Data products consist of analytics reports which include descriptive analytics; predictive analytics; and prescriptive analytics – the most advance of the three.

Research question 3: What are the ethical considerations emanating from collecting and using data for Data science applications?

This question sought to uncover the ethical issues that emerge as companies increase their reliance on data. The findings showed that several ethical and legal risks arise which organisation must manage in order to prevent breaking the law or damaging their brands. The findings pointed out that ethical risks must be effectively governed by the leaders of the organisation. This responsibility cannot be simple delegated to lower levels without proper oversight. In South Africa, laws like POPIA has emerged for the purpose of protecting the rights of individuals from harm caused by unethical practices. Leaders to build effective and sustainable competitive advantage, they need to address ethical risks.

The three research questions guided the inquiry throughout this study. The final analysis enabled the researcher to produce DESC, which is a framework that is intended to guide organisations in implanting Data science driven sense-making capabilities.

7.2 Implications for management and other relevant stakeholders

As discussed above, leaders are entrusted with the responsibility of leading their organisations to success. While they face significant challenges in these tasks, the tool at their disposal. Leaders must find ways and means to overcome the challenges they are faced ahead of the competition. Decision making has been cited several times as one of the key issues' leaders were struggling with. Literature and empirical data have shown that using data can help leaders improve their organisations by first improving their decision making.

However, leveraging Data science requires investment in terms of resources and time. Leaders who take a plunge and explore stand a chance of repeating the reward of superior competitive advantage compared to their peers. Admittedly, the investment in data is a tough decision to make for leaders because, first, Data science projects are prone to failure. Second, it is hard for data time to build a business case with clear ROI ahead of the Data science platform being built; creating what has been termed an egg and chicken scenario.

However, this need not paralyse visionary leaders. Enough empirical evidence and literature have shown tangible benefits of implementing data science. The DESC framework hopes to make a contribution to this problem. If correctly applied, it would guide organisations of the broad considerations that must be in place when creating the Data science capability. Following DESC is hoped to mitigate the rate of project failure as the model spells out a methodology created from best practice.

The DESC framework systematically guides the organisations on how to implement data capability at of the three stages: data warehouse; advanced analytics adoption; and applied data science. The model contains specific processes, technologies, and artefacts applicable at each stage.

7.3 Limitations of the research

It has been shown that some leaders may have only limited experience employing Data science capabilities. Participants indicated that their organisations were migrating from traditional Business Intelligence to Advanced Analytics and Data science. Many proved to be varying phases of implementation. The challenge with this is that few have experienced the full journey of implementing Advanced analytics end to end. It is therefore not clear how these views will change over time as leaders get more exposure to Data science and its potential. . Furthermore, Shaw (2017) emphasised the importance of building a research project on top of solid theory. However, South African specific literature in the field of Data science is limited (Walker and Brown, 2019). The limiting impact of this was that the study relies on global theories while the data collection through interviews were mostly confined to South African participants, with the exception of one individual.

The second limitation was not being able to conduct face-to-face interviews due to COVID-19 social distancing protocols. This removed the personal connection and the ability to read visual cues from the participants. The researcher did not see it appropriate to request the interviews to be on video as some participants could have been inconvenienced through high bandwidth or their personal environment set up.

The participants in this study were chosen based on their expertise and knowledge of strategy and data management capabilities. However, the of most participants have worked in organisations where the data units were already established. As a result, only two have had the experience of building data capability from scratch. This placed a limit on the DESC framework as the research did not have enough empirical data to ascertain the inclusion of this aspect into the framework. As such, the DESC is applicable to assist organisations that are transitioning from one maturity stage to another.

Although the DESC was developed to help organisations enact decision-making capabilities in dynamic markets, it may not be suitable for all categories of dynamism. Dynamic environments can be classified into four categories: simple; complicated; complex; and chaotic (Snowden and Boone, 2007). The literature consulted in this study dealt with a dynamic environment which appears to fit within the complicate and complex realms of dynamism. The implication of this is that the DESC may not be suitable for a chaotic environment, and might be excessive for organisations who operate in simple environmental context.

7.4 Suggestions for future research

The DESC is intended for use by organisations that already have some baseline data capability. Further research is required to describe and include those process necessary to build the capability from the ground up. Future research should also consider whether there should be variations of this model based on the level of dynamism in the environment, that is, nuances that should be considered for each of dynamisms: simple; complicated; complex; and chaotic.

8. References

Badaracco, J. L. (2016). How to tackle your toughest decisions. *Harvard Business Review*, 2016 (September)

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational issues of big data and analytics in networked business. *Mis Quarterly, 40*(4)

Bansal, P. T., Smith, W. K., & Vaara, E. (2018). New ways of seeing through qualitative research. *Academy of Management Journal, 61*(4), 1189–1195. https://doi.org/10.5465/amj.2018.4004

Braun, M. T., Kuljanin, G., & DeShon, R. P. (2018). Special considerations for the acquisition and wrangling of big data. *Organizational Research Methods, 21*(3), 633–659. https://doi.org/10.1177/1094428117690235

Calabretta, G., Gemser, G., & Wijnberg, N. M. (2016). The interplay between intuition and rationality in strategic decision making: a paradox perspective*. Organisation Studies*, (2016).

Castro, A., Machado, J., Roggendorf, M., & Soller, H. (2020). How to build a data architecture to drive innovation—today and tomorrow. *McKinsey Digital*. Retrieved from: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow

Chiang, R. H. L., Grover, V., Liang, T.P., & Zhang, D. (2018). Special issue: strategic value of big data and business analytics. *Journal of Management Information Systems, 35*(2), 383–387.

Christensen, M., & Knudsen, T. (2010). Design of decision-making organisations. *Management Science, 56(*1), 71–89.

Colombo, M. G., M. Delmastro. 2008. *The Economics of Organisational Design*. Palgrave MacMillan, New York.

Conboy, K., Mikalef, P., Dennehy, D., & Krogstie, J. (2020). Using business analytics to enhance Dynamic capabilities in operations research: a case analysis and research agenda. *European Journal of Operational Research*, *281*(3), 656–672

Côrte-Real, N., Ruivo, P., & Oliveira, T. (2019). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value? *Information & Management*.

De Smet, S. A., Jost, G., & Weiss, L. (2019). Three keys to faster, better decisions. *Mckinsey Quarterly, 2019*(2)

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic Management Journal*, *21*(10-11), 1105–1121.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and Data science methods for management research. *Academy of Management Journal, 59*(5), 1493–1507.

Girod Stéphane J. G, & Whittington, R. (2017). Reconfiguration, restructuring and firm performance: Dynamic capabilities and environmental dynamism. *Strategic Management Journal*, *38*(5), 1121–1133. https://doi.org/10.1002/smj.2543

Helfat, C. E., & Raubitschek, R. S. (2018). Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems. *Research Policy, 47*(8), 1391–1399. https://doi.org/10.1016/j.respol.2018.01.019

Hindle, G.A., & Vidgen, R., (2018). Developing a business analytics methodology: a case study in the foodbank sector. *European Journal of Operational Research 268*(3), 836–851.

Hsieh, H., & Shannon, S. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 16*(9), 1277-1288.

Jacob, S., & Furgerson, S. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research*. Qualitative Report, 17*.

Janice, M. M., Michael, B., Maria, M., Karin, O., & Jude, S. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, *1*(2), 13–22.

Jonsen, K., Fendt, J., & Point Sébastien. (2018). Convincing qualitative research : what constitutes persuasive writing? *Organizational Research Methods*, *21*(1), 30–67. https://doi.org/10.1177/1094428117706533

Kurtmollaiev, S. (2020). Dynamic capabilities and where to find them*. Journal of Management Inquiry, 29*(1), 3–16. https://doi.org/10.1177/1056492617730126

Laaksonen, O., & Peltoniemi, M. (2018). The essence of Dynamic capabilities and their measurement. *International Journal of Management Reviews, 20*(2), 184–205. https://doi.org/10.1111/ijmr.12122

Lee, A., Inceoglu, I., Hauser, O., & Greene, M. (2020). Determining causal relationships in leadership research using machine learning: the powerful synergy of experiments and data science. *Leadership Quarterly*, *(2020)*. https://doi.org/10.1016/j.leaqua.2020.101426

Lovallo, D. P., & Sibony, O. (2006). Distortions and deceptions in strategic decisions. *McKinsey Quarterly*, (1), 18–29

Mackay, D., & Zundel, M. (2017). Recovering the divide: a review of strategy and tactics in business and management: recovering the divide. *International Journal of Management Reviews, 19*(2), 175–194. https://doi.org/10.1111/ijmr.12091

Meinard, Y., & Tsoukiàs, A. (2019). On the rationality of decision aiding processes. *European Journal of Operational Research, 273*(3), 1074–1084. https://doi.org/10.1016/j.ejor.2018.09.009

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019a). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research, 98*, 261–276.

Morse, J., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1*(2).

Myres, K. (2020a). Seminar 3: *Qualitative Research Skills* [PowerPoint slides]. Retrieved from

GIBS Blackboard: http://gibs.blackboards.com. GIBS.

Myres, K. (2020b). Seminar 4: *Second General Qualitative Research Workshop* [PowerPoint slides]. Retrieved from GIBS Blackboard: http://gibs.blackboards.com. GIBS.

Nayak, A., Chia, R., & Canales, J. I. (2020). Noncognitive micro foundations: understanding Dynamic capabilities as idiosyncractically refined sensitivities and predispositions. *Academy of Management Review, 45*(2), 280–303. https://doi.org/10.5465/amr.2016.0253

Newman, R., Chang, V., Walters, R. J., & Wills, G. B. (2016). Model and experimental development for business Data science. *International Journal of Information Management, 36*(4), 607–617. https://doi.org/10.1016/j.ijinfomgt.2016.04.004

Porter, M. E. (1980). Industry structure and competitive strategy: keys to profitability. *Financial Analysts Journal*, *36*(4), 30–41.

Pratima, T. B., & Corley, K. (2011). From the editors: the coming of age for qualitative research: embracing the diversity of qualitative methods. *The Academy of Management Journal, 54*(2), 233–237.

Provost, F., & Fawcett, T. (2013). *Data science for business: what you need to know about data mining and data-analytic thinking*. Retrieved from https://ebookcentral-proquest-com.uplib.idm.oclc.org

Robbins, S. P. & Judge, T. A. (2018). *Perception and individual decision making.* In Organisational behaviour (pp. 170 – 190). Sandton: Pearson.

Rosenhead, J., Franco, L. A., Grint, K., & Friedland, B. (2019). Complexity theory and leadership practice: a review, a critique, and some recommendations. *The Leadership Quarterly, 30*(5), 101304–101304.

Roulston, K. (2010). Considering quality in qualitative interviewing. *Qualitative Research, 10*(2), 199-228.

Sanders, P. (1982). Phenomenology: a new way of viewing organisational research. *The Academy of Management Review*, *7*(3), 353–360.

Saunders, M. N., & Lewis, P. (2018). *Doing research in business & management*: An essential guide to planning your project. (2nd ed.). Harlow: Pearson.

Schmarzo, B., & Schmarzo, B. (2013). *Big data: Understanding how data powers big business*. Retrieved from https://ebookcentral-proquest-com.uplib.idm.oclc.org

Shaw, J. D. (2017). Advantages of Starting with Theory. Academy of Management Journal, 60(3), 819–822.

Skjott Linneberg, M., & Korsgaard, S. (2019). Coding qualitative data: a synthesis guiding the novice. *Qualitative Research Journal, 19*(3), 259–270. https://doi.org/10.1108/QRJ-12-2018-0012

Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. a leader's framework for decision making. *Harvard Business Review, 85*(11), 68–76.

Snyder, H. (2019). Literature review as a research methodology: an overview and guidelines. *Journal of Business Research*, *104*, 333–339.

Tadewald, J. I. M. (2019). Data ethics: doing the right thing with data. *Strategic Finance*, Jun2019, P19.

Teece, D.J. (2007). Explicating Dynamic capabilities: the nature and micro foundations of (sustainable) enterprise performance. *Strategic Management Journal (28)*

Teece, D. J. (2014). The foundations of enterprise performance: dynamic and ordinary capabilities in an (economic) theory of firms. *Academy of Management Perspectives 28*(4), 328–352.

Teece, D. J. (2018). Business models and Dynamic capabilities. *Long Range Planning*, *51*(1), 40–49. https://doi.org/10.1016/j.lrp.2017.06.007

Teece, D.J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal 18*(7)

Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational Research Methods, 21*(3), 525–547. https://doi.org/10.1177/2F1094428116677299.

Tsoukiàs Alexis. (2007). On the concept of decision aiding process: an operational perspective. *Annals of Operations Research, 154*(1), 3–27. https://doi.org/10.1007/s10479-007-0187-z

Uhl-Bien, M., & Arena, M. (2018). Leadership for organisational adaptability: A theoretical synthesis and integrative framework. *Leadership Quarterly, 29*(1), 89–104.

Vaara, E., Paruchuri, S., Nadkarni, S., & Shaw, J. D. (2019). New ways of seeing big data. *Academy of Management Journal, 62*(4), 971–978. https://doi.org/10.5465/amj.2019.4004

Van den Steen, E. (2017). A formal theory of strategy. *Management Science, 63*(8), 2616–2636. https://doi.org/10.1287/mnsc.2016.2468

Van den Steen, E. (2018). Strategy and the strategist: how it matters who develops the strategy. *Management Science*, *64*(10), 4533–4551. https://doi.org/10.1287/mnsc.2017.2857

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626–639. https://doi.org/10.1016/j.ejor.2017.02.023

Walker, R. S., & Brown, I. (2019). Big data analytics adoption: a case study in a large South African telecommunications organisation. *SA Journal of Information Management, 21*(1).

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171–180.

Winter, S. G. (2003). Understanding Dynamic capabilities. *Strategic Management Journal*, *24*(10), 991–995.

Zeni, T. A., Buckley, M. R., Mumford, M. D., & Griffith, J. A. (2016). Making “sense” of ethical decision making. *The Leadership Quarterly, 27*(6), 838–855. https://doi.org/10.1016/j.leaqua.2016.09.002

1. Appendices

Appendix A: Measuring instrument

Question	Comments	Relevant research question
1. Please provide some background of your leadership roles and responsibilities.	The unit of analysis is the individual leader. The question serves as a screening question, will help identify participants who do not meet the sample profile.	Research Question 1
2. Can we please important elements and aspects of the strategy making process? Who is responsible for those elements?	The description of the strategy process will clarify how and how the leader approaches Sense-making, Seizing, and Transforming.	
3. The idea of sense-making, seizing and transforming. How do you Sense Make? How do you identify the major opportunities and challenges presenting themselves in your industry and perhaps for your organisation specifically?	This question is intended to unpack how sense-making takes place	Research Question 1.
4. What inputs are critical to be prepared and made available ahead of the strategy session?	Gaging the importance of Data science as input to the process of sense-making to the leader and their organisation.	Research question 2

5. How do you satisfy yourself that you have an accurate assessment of opportunities and threats when making strategic decisions?	Important for leaders who indicate that their environment is complex/often in flux.	Research question 2.
6. Do you have a built-in mechanism that provides feedback on strategy execution; does Data science play any role in this process?	Is the leader using data and metrics to track the execution of their strategy?	Research question 2
7. What do you find to be the special contribution of the Data science capability?	The definition of Data science will be provided, and a distinction made with the traditional Data function. This is not to test for knowledge of Data science but to simply establish the extent of the capability and its use.	Research question 2
8. What is the composition of the Data team/extended team? Roles, skills set. What interface with business do you have?	Is this just a supplier/customer handshake or is there integration? E.g. Data Committee.	Research question 2
9. What does the Data team do well that assists your organisation with decision making?	To assess important Data science team routines and practices.	Research question 2
10. What the key ethical legal considerations observed by	To assess presence of quality controls and ethical practices.	Research question 3

organisations in your experience		
11. What are the critical moral issues that arise from leveraging data		Research question 3
END		

Appendix B: Themes to category mapping

Themes	**Categories**	**Code count**
1. Environment analysis and strategy design	Analytical problem solving	9
	Contextualisation	12
	Creating alignment	5
	Environment scanning	7
	Strategy formulation	9
	Strategy implementation	4
2. Decision making	Data informed decisions	4
	Gut feel decisions	6
3. Leadership	Attributes and roles of leaders	5
	Leadership action	10
4. Crafting Data science capability	Attributes of Data science capability	2
	Deployment of Data science capability	3
	Goal alignment	3
	Implementation project consideration	9
	Investment in data capability	7
	Operating model	16
	Team selection considerations and activation	5
5. Data science processes and practices	Analytical model design and execution	15
	Business analysis	9
	Data Sourcing	9
	Quality management	5
	Reporting and visualisation	1
	Team configuration	1
6. Data science skills	Cross functional skills	5
	Technical skills	8
7. Data science technology	Contemporary technologies	6
	Traditional technologies	2
8. Data Ethics and Governance	Good practice	6
	Impact brand equity	2
	Moral considerations	7
9. Data legislations	Legal requirements	7

Appendix C: Code book

Theme	Category	Code
Crafting Data science capability	Attributes of Data science capability	Analytics is based on Models
Crafting Data science capability	Attributes of Data science capability	Data is resource
Crafting Data science capability	Deployment of Data science capability	Data capability internally built
Crafting Data science capability	Deployment of Data science capability	Data capability undergoes maturity stages
Crafting Data science capability	Deployment of Data science capability	Desperate data sources internally
Crafting Data science capability	Goal alignment	Analytics enables differentiation
Crafting Data science capability	Goal alignment	Data can be transformed into a capability
Crafting Data science capability	Goal alignment	Data insights built to support strategy
Crafting Data science capability	Implementation project consideration	About 70% to 90% Data Science projects fail
Crafting Data science capability	Implementation project consideration	Cloud technology is a shift forward
Crafting Data science capability	Implementation project consideration	Data lake has many types of data
Crafting Data science capability	Implementation project consideration	Data Warehouses are still in use
Crafting Data science capability	Implementation project consideration	Fragmented data capability across business units
Crafting Data science capability	Implementation project consideration	Hybrid data store required - Data Lake
Crafting Data science capability	Implementation project consideration	Iterate in building the capability - maturity curve
Crafting Data science capability	Implementation project consideration	One company X - cited by many as model for data driven decisions
Crafting Data science capability	Implementation project consideration	Technology must be sustainable

Crafting Data science capability	Investment in data capability	10% of budget is for research in one company
Crafting Data science capability	Investment in data capability	Cloud can reduce upfront investment
Crafting Data science capability	Investment in data capability	Data and Analytics capability requires investment
Crafting Data science capability	Investment in data capability	Do not invest in technology for the sake of it
Crafting Data science capability	Investment in data capability	Justifying investment on data platform difficult
Crafting Data science capability	Investment in data capability	Small businesses lack resources to build data capability
Crafting Data science capability	Investment in data capability	Value come after the effect - ROI not immediate
Crafting Data science capability	Operating model	Create sandbox for experimenting
Crafting Data science capability	Operating model	Data project must be supported by organisation culture to be sustainable
Crafting Data science capability	Operating model	Data science projects still scarce - emergent
Crafting Data science capability	Operating model	Finance departments historical data driven
Crafting Data science capability	Operating model	Implements Data Management processes
Crafting Data science capability	Operating model	Implements Data Quality processes
Crafting Data science capability	Operating model	Legacy systems impedes tech and business improvements
Crafting Data science capability	Operating model	Operational data is ordinary capability
Crafting Data science capability	Operating model	Outsourcing of Data Science capability
Crafting Data science capability	Operating model	Predictive analytics do not work in chaotic environment
Crafting Data science capability	Operating model	SA Cloud adoption leg behind EU

Crafting Data science capability	Operating model	Technology now shapes business strategy
Crafting Data science capability	Operating model	Telecoms are data driven
Crafting Data science capability	Operating model	Use of consultancies in analytics
Crafting Data science capability	Operating model	Use of open competitions online for some of Data Science projects
Crafting Data science capability	Operating model	You need playfulness and experimentation
Crafting Data science capability	Team selection considerations and activation	Adoption requires involvement of strategic, tactical & operational levels
Crafting Data science capability	Team selection considerations and activation	Analytics driven by research team
Crafting Data science capability	Team selection considerations and activation	Data team is made up of cross functional team
Crafting Data science capability	Team selection considerations and activation	Data team needs empowerment and voice
Crafting Data science capability	Team selection considerations and activation	Data team requires vertical integration
Data Ethics and Governance	Country specific law	Apply law and Codes of Good practice
Data Ethics and Governance	Good practice	Ask for customer consent before collecting data
Data Ethics and Governance	Good practice	Data governance and stewardship
Data Ethics and Governance	Good practice	Law is minimum standard - we must do more
Data Ethics and Governance	Good practice	Protect the right of customers

Data Ethics and Governance	Impact brand equity	Customers punish organisations with unethical data practices
Data Ethics and Governance	Impact brand equity	Young customers weary about personal data
Data Ethics and Governance	Moral issues	AI takes way jobs
Data Ethics and Governance	Moral issues	Business models based on selling people's data
Data Ethics and Governance	Moral issues	Data ethics must be observed
Data Ethics and Governance	Moral issues	Ethical data model
Data Ethics and Governance	Moral issues	Ethics - automated bio data collection by sensor & cameras
Data Ethics and Governance	Moral issues	Prejudice in models due to data set they trained on
Data Ethics and Governance	Moral issues	Unethical to use data to benefit company but disadvantage customer
Data legislation	Country specific law	Ensure data privacy and security
Data legislation	Country specific law	EU Data legislation comprehensive
Data legislation	Country specific law	EU GDPR
Data legislation	Country specific law	Law drives ethics and governance improvements
Data legislation	Country specific law	Law forces improvement in data maturity
Data legislation	Country specific law	POPI Act
Data legislation	Country specific law	Provide ability to opt in and out
Data legislation	Good practice	Data team must be trained on ethics and law
Data science processes and practices	Analytical model design and execution	AB Testing - applying this
Data science processes and practices	Analytical model design and execution	Analytical models use algorithms and maths to process volumes of data

Data science processes and practices	Analytical model design and execution	Analytics come from data
Data science processes and practices	Analytical model design and execution	Data analytics enables prediction
Data science processes and practices	Analytical model design and execution	Data reveals looming disruption
Data science processes and practices	Analytical model design and execution	Data Science can be used to filter facts from fake news
Data science processes and practices	analytical model design and execution	insights from data is competitive advantage
Data science processes and practices	Analytical model design and execution	Join dots across industries using data
Data science processes and practices	Analytical model design and execution	Training of analytical model
Data science processes and practices	Analytical model design and execution	use analytics for customer retention
Data science processes and practices	Analytical model design and execution	use analytics to prevent churn
Data science processes and practices	Analytical model design and execution	Use customer data to personalise the service
Data science processes and practices	analytical model design and execution	Use Data Science to declutter information
Data science processes and practices	Analytical model design and execution	Use data to cross sell current products

Data science processes and practices	Analytical model design and execution	Use of data to show trends
Data science processes and practices	Business analysis	Data use has expanded across the organisation
Data science processes and practices	Business analysis	Excel heavily used for modelling
Data science processes and practices	Business analysis	Find data for variables that drive the business
Data science processes and practices	Business analysis	Operational data for day to day decisions
Data science processes and practices	Business analysis	Predictability only applies within the problem boundaries
Data science processes and practices	Business analysis	Real time data
Data science processes and practices	Business analysis	Streaming data requires good data management techniques
Data science processes and practices	Business analysis	Throw away inaccurate model – do not justify it
Data science processes and practices	Business analysis	Your own data contains critical signals
Data science processes and practices	Data Sourcing	Data ecosystem include outside parties
Data science processes and practices	Data Sourcing	Data is not only numeric

Data science processes and practices	Data Sourcing	Data Science uses both structured and unstructured data
Data science processes and practices	Data Sourcing	Data use is industry dependent
Data science processes and practices	Data Sourcing	Data warehouse requires structured data
Data science processes and practices	Data Sourcing	Market data is sourced from the industry
Data science processes and practices	Data Sourcing	Timely access to data is important
Data science processes and practices	Data Sourcing	Use macro-economic data
Data science processes and practices	Data Sourcing	Use of external data sources important
Data science processes and practices	Quality management	Assume errors in data until proven otherwise
Data science processes and practices	Quality management	Assumptions in models must be re-evaluated
Data science processes and practices	Quality management	Data can be bias - e.g. pre 94 black had no credit
Data science processes and practices	Quality management	Data quality requires Data Management
Data science processes and practices	Quality management	Poor data quality, wrong decisions

Data science processes and practices	Reporting and visualisation	Data requires competent interpreter
Data science processes and practices	Team configuration	Collaboration across organisation
Data science skills	Cross functional skills	Data Science requires formal education
Data science skills	Cross functional skills	Data scientist a strategic skill
Data science skills	Cross functional skills	Data team understands behavioural patterns
Data science skills	Cross functional skills	Strategic data skills are a competitive advantage
Data science skills	Cross functional skills	Strategic data skills hold the company IP
Data science skills	Technical skills	Data Architect a strategic skill
Data science skills	Technical skills	Data Science Tools - Spark, Java, Hadoop, SCALA, SQL, R, Python
Data science skills	Technical skills	Data skills include mathematicians and statisticians
Data science skills	Technical skills	Data team understands algorithms
Data science skills	Technical skills	Data team understands systems
Data science skills	Technical skills	Key skills are rare
Data science skills	Technical skills	SQL developer not strategic skill
Data science skills	Technical skills	Without skill you can not exploit data
Data science technology	Contemporary technologies	Cloud more secure than on premise infrastructure
Data science technology	Contemporary technologies	Computer vision is advanced AI
Data science technology	Contemporary technologies	Data Science Tools - Spark, Java, Hadoop, SCALA, SQL, R, Python
Data science technology	Contemporary technologies	Gradient Boosting Machine (GBM)
Data science technology	Contemporary technologies	Science +Data product

Data science technology	Contemporary technologies	Spark can handle multiple data sets
Data science technology	Traditional technologies	Business intelligence + Advanced analytics + Data Architecture + Data
Data science technology	Traditional technologies	Data Warehouse
Decision making	data informed decisions	Analytics based decisions are used at all levels
Decision making	data informed decisions	Strategy design is decision making
Decision making	data informed decisions	Use data to find drivers of the business
Decision making	data informed decisions	Use of data in strategic decisions
Decision making	gut feel decisions	Decisions based on qualitative and democratic process
Decision making	gut feel decisions	Flying blind if you do not have data analytics
Decision making	gut feel decisions	Gut feel based decisions
Decision making	gut feel decisions	Highly paid opinion overshadowing others
Decision making	gut feel decisions	Overwhelmed with the amount of data
Decision making	gut feel decisions	Politicisation of decisions
Leadership	Attributes and roles of leaders	Chief Data Officer is a strategic position
Leadership	Attributes and roles of leaders	Chief Data Officer must report to CEO
Leadership	Attributes and roles of leaders	CIO and CDO complementary but different competencies
Leadership	Attributes and roles of leaders	Leaders has technical background
Leadership	Attributes and roles of leaders	Young leaders embrace data
Leadership	Leadership action	Involve people from all levels in strategy design
Leadership	Leadership action	Leader must prep for strategy

Leadership	Leadership action	Leaders coming to understand the value of data
Leadership	Leadership action	Leaders must immerse themselves in organisation processes
Leadership	Leadership action	Older leaders and those in mature industries trust own knowledge above data
Leadership	leadership action	Understand operating environment
Leadership	Leadership action	Understand organisation drivers and constraints
Leadership	leadership action	Understand your strengths, resources and capabilities
Leadership	Leadership action	Understanding customer needs leads to differentiation
Leadership	Leadership action	Use a framework to lead strategy discussion
Strategy design considerations	Analytical problem solving	5 Whys framework
Strategy design considerations	Analytical problem solving	Define problem , establish context, analysis and solution
Strategy design considerations	Analytical problem solving	Jobs to be done
Strategy design considerations	Analytical problem solving	PESTEL Analysis to scan external environment
Strategy design considerations	Analytical problem solving	Porter five forces framework
Strategy design considerations	Analytical problem solving	Problem can be complicated, complex, crisis
Strategy design considerations	Analytical problem solving	Problem solving should factor in CONTEXT
Strategy design considerations	Analytical problem solving	SWOT Framework for internal analysis
Strategy design considerations	Analytical problem solving	Use 5Cs framework
Strategy design considerations	Contextualisation	Assessment of own current product offerings

Strategy design considerations	Contextualisation	Collect voice of the customer qualitative data
Strategy design considerations	Contextualisation	Contextual or situational analysis
Strategy design considerations	Contextualisation	COVID 19 accelerated the need for insights
Strategy design considerations	Contextualisation	COVID disrupted business models
Strategy design considerations	Contextualisation	Look beyond traditional places for answers
Strategy design considerations	Contextualisation	New business model can disrupt your company
Strategy design considerations	Contextualisation	Organisational dynamics impact strategy adoption
Strategy design considerations	Contextualisation	Profile and know your customer
Strategy design considerations	Contextualisation	Sensing current and future customer needs important
Strategy design considerations	Contextualisation	Start by defining the problem not objectives - strategy design
Strategy design considerations	Creating alignment	Organisations need to know their identity
Strategy design considerations	Creating alignment	Setting of business objectives
Strategy design considerations	Creating alignment	Strategy comes from identity and vision
Strategy design considerations	Creating alignment	Strategy is map towards the vision
Strategy design considerations	Creating alignment	Vision is an aspiration of an organisation
Strategy design considerations	Environment scanning	competition from companies with totally different business models
Strategy design considerations	Environment scanning	Environment is very dynamic - changes all the time
Strategy design considerations	Environment scanning	Operating environment knowledge key in strategy

Strategy design considerations	Environment scanning	Sensing of competition
Strategy design considerations	Environment scanning	Sensing of competition coming from outside your industry
Strategy design considerations	Environment scanning	Sensing opportunities and threats through data
Strategy design considerations	Environment scanning	use data to scan environment
Strategy design considerations	Strategy formulation	Culture is important in strategy design and success
Strategy design considerations	Strategy formulation	Define target state and desired capabilities
Strategy design considerations	Strategy formulation	Do not force fit framework into an organisation
Strategy design considerations	Strategy formulation	Know and involve your stakeholders
Strategy design considerations	Strategy formulation	Outsource to compensate for internal impediment
Strategy design considerations	Strategy formulation	Strategic change must consider internal complexity
Strategy design considerations	Strategy formulation	Strategic decisions require scenario modelling tool
Strategy design considerations	Strategy formulation	Strategy a function of board and executives
Strategy design considerations	Strategy formulation	Strategy creates market relevancy
Strategy design considerations	Strategy formulation	User Mergers and Acquisition to acquire capabilities
Strategy design considerations	Strategy implementation	Failure to adopt kill organisations
Strategy design considerations	Strategy implementation	Monitor KPIs
Strategy design considerations	Strategy implementation	Strategy design is simpler than implementation
Strategy design considerations	Strategy implementation	Strategy must be agile

Appendix E: Plagiarism declaration

Declaration

I declare that this research project is my own work. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration at the Gordon Institute of Business Science, University of Pretoria. It has not been submitted before for any degree or examination in any other University. I further declare that I have obtained the necessary authorisation and consent to carry out this research.

Simphiwe Nxumalo

01 December 2020

Appendix F: Certification of Data Analysis Support Form

(Additional support retained or not - to be completed by all students)

Please note that failure to comply and report on this honestly will result in disciplinary action

I hereby certify that (please indicate which statement applies):

- ***I DID NOT RECEIVE*** any additional/outside assistance (i.e. statistical, transcriptional, and/or editorial services) on my research report:

  ...None……………………………………………

- ***I RECEIVED*** additional/outside assistance (i.e. statistical, transcriptional, and/or editorial services) on my research report:
  I received assistance for editing but no personally identifiable information was included in the report
  ………………………………………………………………………………………..

If any additional services were retained– please indicate below which:

□ Statistician

□ Transcriber

□ Editor

□ Other (please specify:N/A……………………….)

Please provide the name(s) and contact details of all retained:

NAME:…Simphiwe Nxumalo………………………………………………………...

EMAIL ADDRESS: …zwides@gmail.com…..……………………………………….

CONTACT NUMBER: ..0725887697…………………………………………………

TYPE OF SERVICE: …Editing …………………………………………….….

I hereby declare that all *statistical write-ups and thematic interpretations of the results for* **my study were completed by myself without outside assistance**

NAME OF STUDENT:

Simphiwe Nxumalo…………………………………………………………………….

SIGNATURE:

…………………………………………………………………………………………….

STUDENT NUMBER:

25363931…………………………..……………………………………………………

STUDENT EMAIL ADDRESS:

zwides@gmail.com ……………………..……………………………………………….