Data science in Sweden

- Exploring the state of data science use in Swedish businesses

Master's thesis in the Master's Programme Management and Economics of Innovation
by Tom Pettersson and Adam Prytz

Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018
Report No. E 2018:041

REPORT NO. E2018:041

Data science in Sweden

Exploring the state of data science use in Swedish businesses

TOM PETTERSSON
ADAM PRYTZ

Examiner: Henrik Berglund

Department of Technology Management and Economics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018

Data science in Sweden
Exploring the state of data science use in Swedish businesses
TOM PETTERSSON
ADAM PRYTZ

Master's Thesis E2018:041
Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone: + 46 (0)31-772 1000

Chalmers Reproservice
Gothenburg, Sweden 2018

Abstract

Data science brings with it vast opportunities. New technology enables advanced analytics to an increased extent and organizations seek to extract business value by becoming increasingly data driven. On behalf of the Technology Management and Economics department at Chalmers University of Technology this thesis seeks to explore the current adoption of data science and its implications among Swedish businesses. The focus of the study lies on mapping the current state, aiming to understand what the future holds and which challenges firms will have to deal with in the process of adopting data science into their organizations. Drawing upon data from interviews with representatives from a diverse set of Swedish firms, along with current research in the field of data science, opportunities and challenges within the three areas competence, organization, and business impact are identified. With regard to business impact, there is a substantial gap between potential and actual value extracted from data science among Swedish businesses. And while it is clear that Swedish firms seek to develop their capabilities within the field, the gap between potential and real value extracted is deemed to grow unless considerable action is taken within two areas: (1) develop the right managerial competence to bridge between technology and management, and (2) address the organizational challenges related to integrating data science to the operations of the firm. Drawing upon literature on data science and change management, the study provides clarity in what competencies should be focused on in order to foster 'data translators', and it also conceptualizes a seven step organization process which should be leveraged as a source of orientation for firms under the process of transition from 'no data science operations' to 'full-scale data science operations'.

Contents

Executive summary

In its most general sense, data science concerns the transformation of data to information which is used as input to decision making. During the last decade the volume, speed, and variety of data has consistently increased, and consequently new generations of technologies have emerged from it. This has eluded no one; while everyone does not necesssarily know artificial intelligence, for example, almost everyone knows about artificial intelligence. Gartner Inc., an American research and advisory firm, provides an overview of data science's development, explaining the different levels of potential business value extraction in figure 1 below.

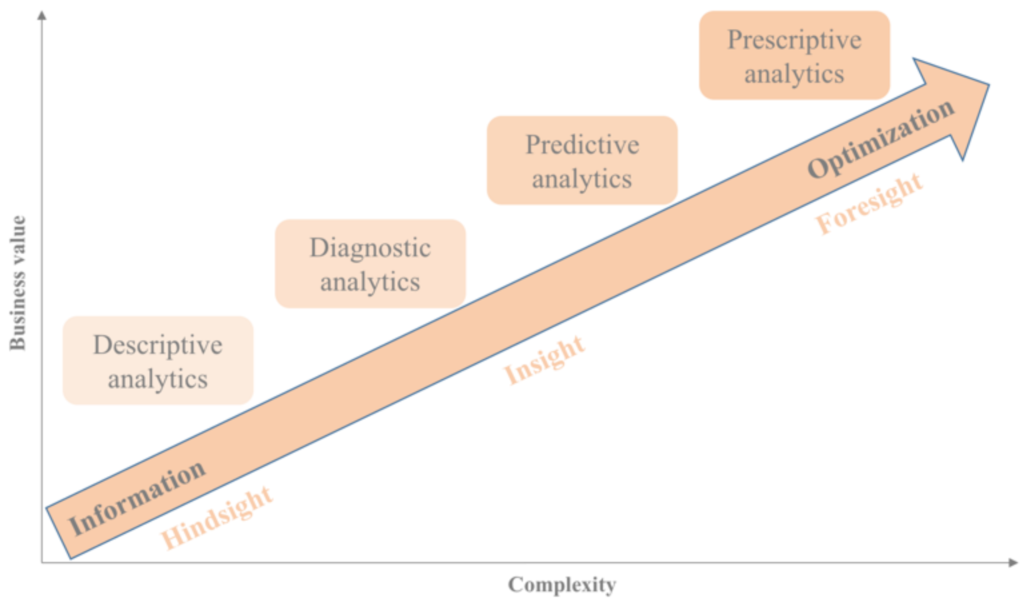

Figure 1: The Gartner levels of business value extraction.

Businesses have a lot to gain from 'climbing the ladder' as conceptualized above, but most businesses fail to do so even if the potential is clear. The McKinsey Global Institute (MGI) first investigated the discrepancy in potential and utility of analytics back in 2011 and five years later, a similar investigation concluded that there was still a large gap between potential and reality. Consequently, this thesis picks up the trail of MGI and embarks upon a deeper investigation of the drivers of success in data science practice with a local focus on the Swedish business landscape.

In essence, the objective has been to describe the 'state of the data science art' in Sweden, and what challenges that firms generally face when setting out to become more data driven as an organization. Furthermore, the goal has been to provide a holistic perspective of the Swedish business landscape as a whole, which is why data has been gathered from a diverse sample of business representatives, with regard to company size, industry, role of interviewee, and also based on a binary segmentation of the degree to which the business is inherently technology oriented. The latter segmentation differentiates by label, where firms inherently rooted in digital technology are labelled 'Digital', and those who are not are labelled 'Industry'. The sample has the following properties:

- 23 interviews with 28 representatives: 4 data scientists, 2 C-level executives and 22 Head of Analytics or similar

- 9 digital firms ranging from 25 - 3,000 employees
- 14 industry firms ranging from 1,500 - 200,000 employees
- 14 represented industries/sectors: retail, telecommunication, airline, hotel, healthcare, manufacturing, construction, streaming, transportation, advertising, venture capital, media, government authority, and software engineering

During the interaction with these individuals, focus was put on gathering data related to three areas of data science: (1) the newly emerged competence requirements, from primarily a managerial but also a technical point of view. (2) The organizational challenges that a corporation typically faces when setting out to become more data driven, and (3) the level of business impact that businesses are able to extract as an effect of their data science operations.

The verdict: Swedish businesses master descriptive analytics and are moving into the predictive field, but need to increase the pace

Historically, Swedish businesses have leveraged data for descriptive analytics practices, i.e. collecting information into reports that summarizes business outcomes. Today, most firms are progressing into predictive analytics, and a few have advanced even further than that into prescriptive analytics as defined by Gartner. This leads to the conclusion that there is a significant gap between the potential value to be extracted from data science, and the actual value currently being derived. Interviewed businesses agree with the assertion, and unanimously resonate the expectations that the field's potential will continue to increase, allowing processes to be improved even further and products development to better align with the needs of customers. Essentially, this means that the currently existing gap between potential and reality will become even larger if considerable action is not taken. The different levels of analytics in Swedish businesses is outlined in figure 2 below.

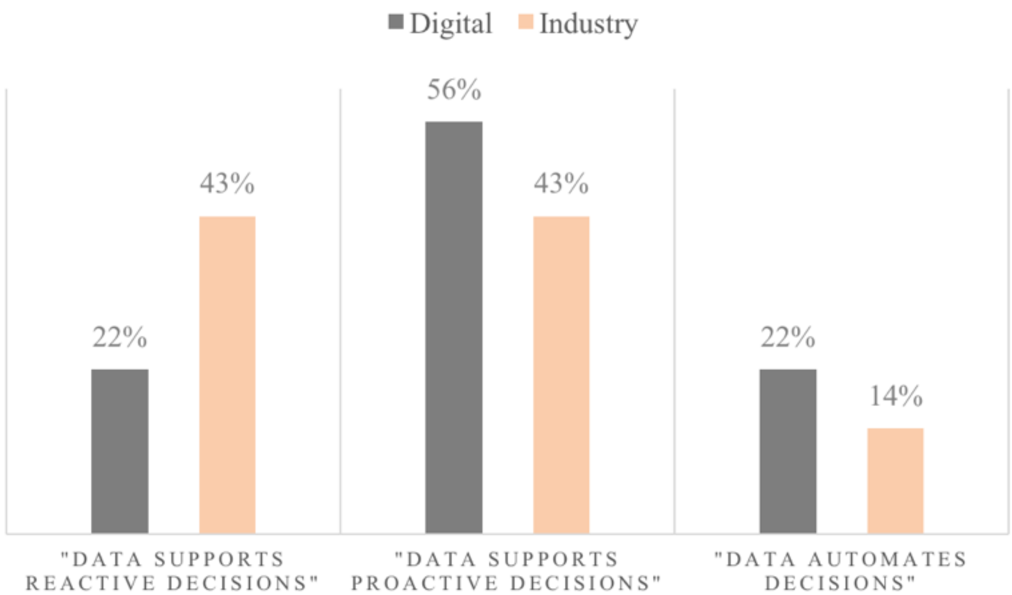

Figure 2: Percentages of Swedish businesses that have progressed into the respective levels of business value extraction from data science.

An important note to be made regarding business impact is that it is highly contingent on success in organizational development and competence pursuit - the other factors investigated in this study. This is to say that if an arbitrary firm does not actively pursue the right competence and an enabling organizational structure, business impact will undoubtedly suffer. While the notion might seem obvious, it is warranted to highlight as it produces points of entry for a future organization seeking to become better at leveraging data science in their decision making. More specifically, it illuminates that an organization must successfully enact two strategies in order to extract business value out of data science: one of competence pursuit, and another of organizational transformation.

Strategy #1: Seek and develop the data translators by emphasizing five crucial competencies

The biggest challenge of competence demand is its incredibly fast paced development. Historically speaking, Swedish businesses have not generally pursued technical competence related to data science; i.e. data scientists or similar. Even fewer considered the need for management roles integrated within the field. However, this has completely changed as of late; the foundation of a data driven organization are data driven individuals. This has undoubtedly sparked the immense demand for data scientists, but businesses also agree that there is a need for managerial communities to embrace data science into their competence portfolios as well. More specifically, five skills (illustrated in figure 3) have emerged as crucial for a manger to possess in connection to data science: (1) business knowledge, referring to being familiar with how the business works. (2) Programming, emphasizing the practical skills in different programming languages. (3) Statistics & math, not surprisingly referring to knowledge in statistics and mathematics. (4) Utility of technology tools, which concerns the knowledge of what techniques data scientists utilize to address various questions and problems, and doing what might be referred to as 'speaking the data science language'. Finally, (5) change management which emphasizes the ability to inspire change.

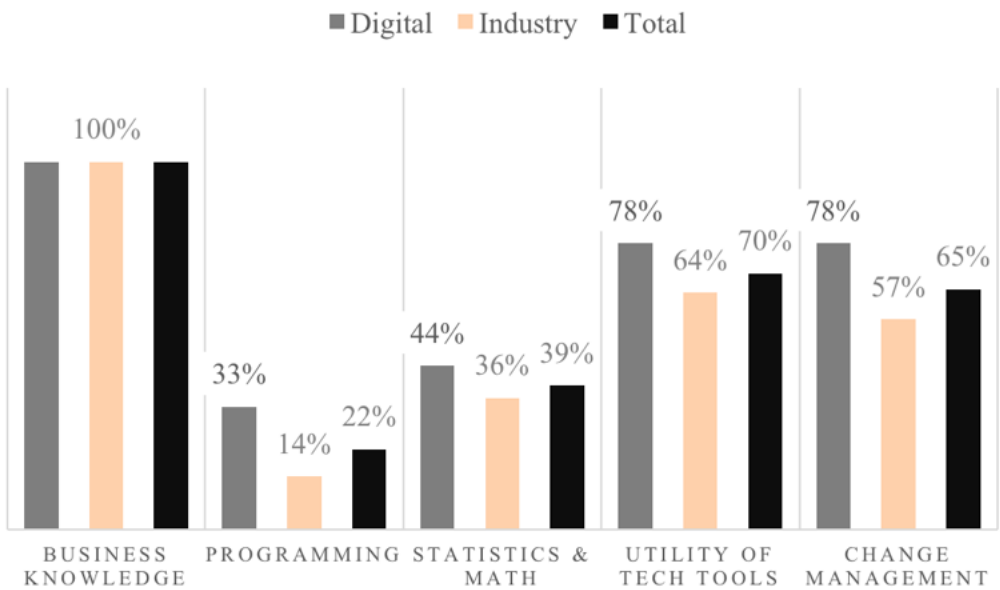

Figure 3: Competencies and their criticality for management in data science.

The analysis suggests that competence demands appear 'universal' since the discrepancy in emphasis between industry and digital firms does not seem significantly high. Theory often refers to the need of 'data translators'; individuals who understand both the technology as well as the business. Given that these skills span between inherent business and technology skills, they might provide a strong basis for what such a data translator actually should know. Figure 4 illustrates these notions in the form of a competence spectrum.

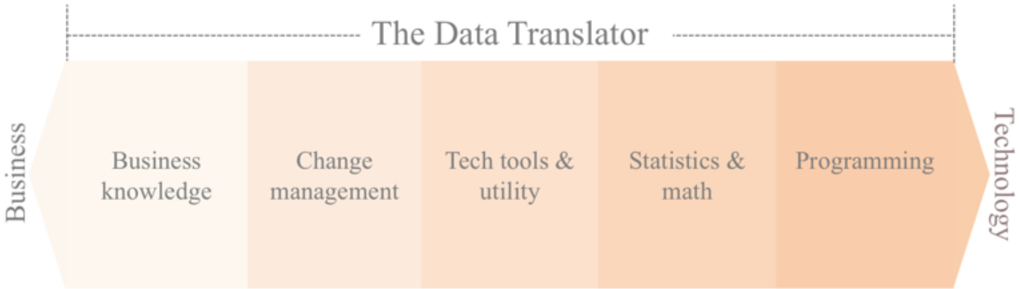

Figure 4: The spectrum properties of the five crucial competencies.

Strategy #2: Address the organizational challenges and transform the firm in seven steps

The major finding of the thesis is the conceptualized process of seven steps that serves to depict the progress of an arbitrary firm transitioning from 'no practice' to 'full-scale practice' of data science operations. It is encouraged to be used by the individual or enterprise in their respective processes of development, providing clarity in what path to take as well as in relative progression compared to peers.

The seven steps (illustrated in figure 5) are the following: (1) become the active change agent, referring to the first individual realizing the need to change, as well as being willing enough to take action on it. (2) Show people what can be done, which is when the active change agent gains support among their peers to scale up the number of active change agents and create strength in numbers. (3) Draw executive attention and support, which is when the active change agent has drawn sufficient attention to gain an audience with senior management, enabling the critical leadership anchoring for continued progress in the process.

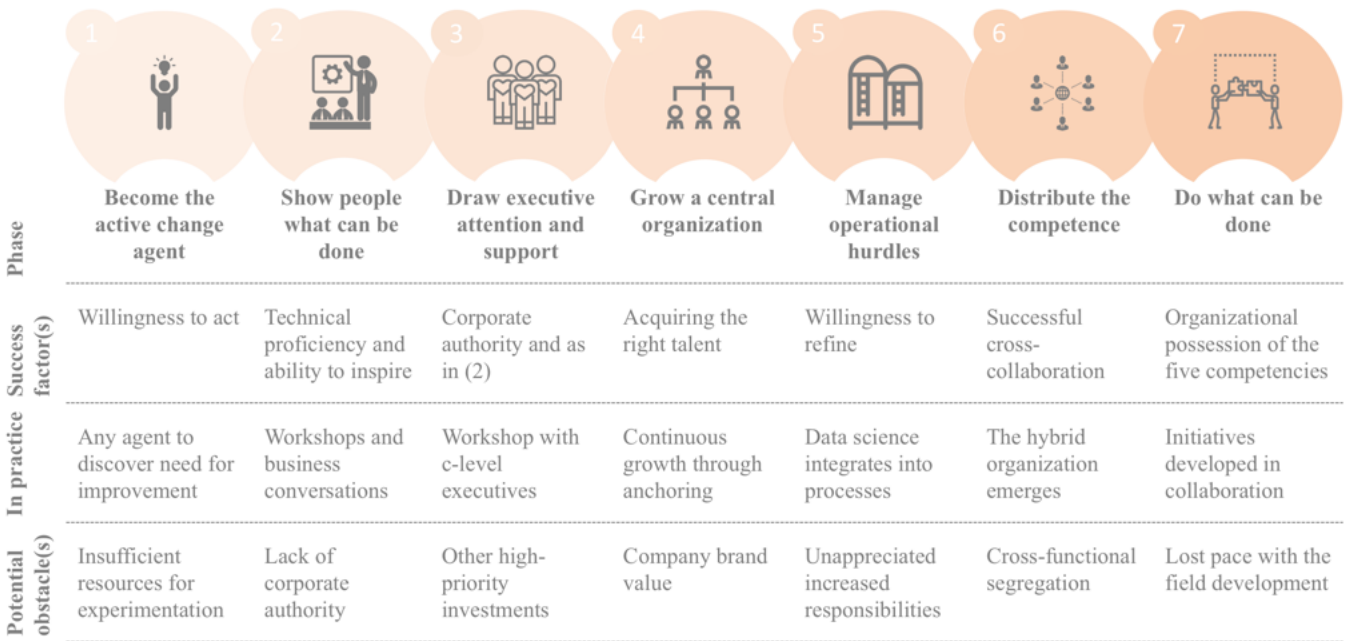

Figure 5: Proposed process to become data driven as an organization.

(4) Grow a central organization is the step in which the unit is created and given the chance to grow by proving its worth, changing the mindset of people in doubt and enabling a 'Center of Excellence' to emerge within the organization. (5) Manage operational hurdles, a minor or major pit stop depending on the organization, when issues of data and process management are dealt with to enable future development and cross-functional collaboration. (6) Distribute the competence, which is when the center of excellence has gained critical mass in volume, the operational hurdles have been overcome, and the central unit can start leveraging their competence into cross-functional teams across the organization. Lastly, (7) do what can be done, which is the state of optimal data science practice; when there is an organizational mindset to be data driven, the firm collectively possesses the five crucial competencies outlined above, and data science initiatives are developed in collaboration between roles.

The rationale behind the conceptualization is based on an analysis which has answered two questions: (1) what steps should be included, and (2) how those steps should be ordered. As for question one, the seven steps have been created on the basis that they should lay out a reasonable road map towards what is considered the best organizational structure for data science, while concurrently addressing the major organizational challenges in the course of following said road map. Below follows a brief contextualization of what that desired structure is, as well as the challenges that may stand in the way to accomplish it.

As for the desired design of the organization, a vast majority of Swedish businesses agree that they consider a hybrid data science organization the best structure, i.e. data scientists and technologically savvy people that organize in a 'Center of Excellence' while leveraging their competence into cross-functional teams in the organization. This has been deemed better than organizing centralized or decentralized as it offers knowledge sharing within the unit and resource efficiency, while maintaining flexibility to distribute the competence into teams throughout the organization where and when it is needed. However, setting up such an organization is not arbitrary; five major challenges (illustrated in figure 6) have emerged to generally affront most companies when doing so.

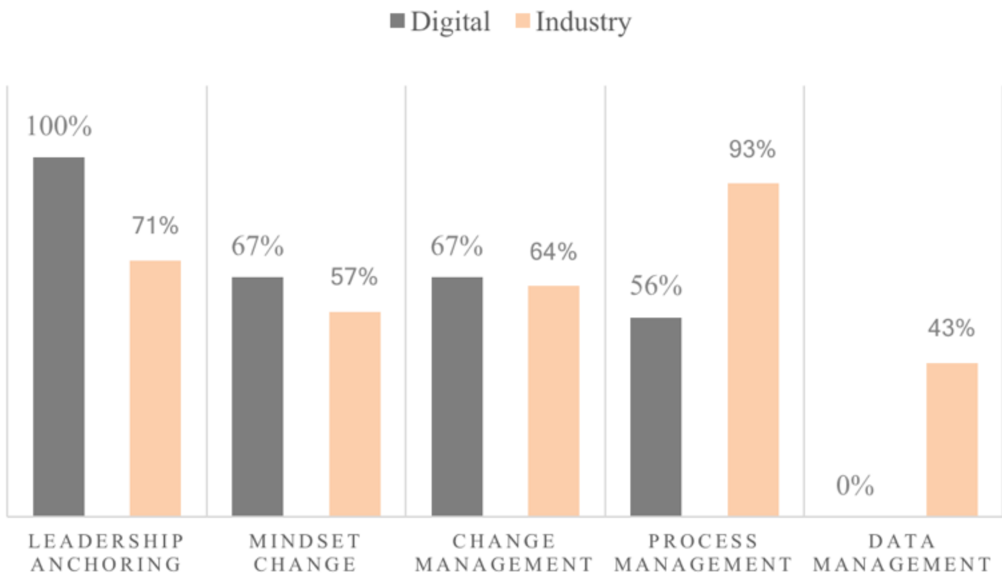

Figure 6: The five major organizational challenges, comparing rate of emphasis between digital and industry firms.

(1) Leadership anchoring refers to the support from senior management. This challenge is affronted in step three of figure 5: draw executive attention and support. (2) The organizational mindset change essentially concerns the need to change the organizational perception of the value of data science practices, and consequential necessary changes in their work to enable it. This is continually addressed throughout the process, but mostly in step two and six, as those phases seek to show people what can be done at different scales. The third challenge is (3) the active practice of change management, underlining the importance of available change managers that are enabled to show people what can be done with data science, which is important throughout the entire process and undoubtedly in the first two steps. Continuing, the fourth challenge is (4) process management refers to the need to facilitate cross-collaboration between competencies, implement agile work methods, and simplify any arbitrary operational process related to data science integration. And finally, (5) data management, concerns the availability and cooperation across functions with regard to data. The purpose of the process' fifth step is specifically to address both these challenges.

As for the second question to be answered when conceptualizing the process, the order of the steps is motivated by the perceived sequential nature of the challenges. This is to say that the authors of the study have interpreted some challenges more relevant to address earlier than others. For instance, there are several arguments to be made for why leadership anchoring should be focused on early. For example, an intuitive train of thought would suggest that dealing with any obstacle related to growing a central organization of data science will never precede gaining the support from management to set up said organization in the first place. However, it should be mentioned that the process outcome for an arbitrary firm can vary and the current design should be viewed as the most general of processes. Regardless, each part of the process matters and needs to be addressed one way or another.

In conclusion, the goal of mapping the Swedish business landscape in data science has been addressed by investigating three factors: competence requirements, organizational demands, and business impact. This thesis draws upon the insights within each aspect to conceptualize frameworks and processes representing what businesses agree are the desirable end goals within each aspect.

1 Introduction

Over the past two decades, ‘Data Science’ has grown tremendously in importance and relevance to both practitioners in business and academia (Chen, Chiang, & Storey, 2012). Self-driving cars, ’Industry X.0’, and the Saudi-Arabian robot citizen Sophia are all examples of mankind’s increased proficiency within the field (Schaeffer, 2017;Goertzel, Mossbridge, Monroe, Hanson, & Yu, 2017). Data science is conventionally defined as *“the collection, management, processing, analysis, visualization and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and interdisciplinary applications”* (Donoho, 2017). While analyzing data to make informed decisions has been common among businesses for a long period of time, what has altered the dynamics of the field lately is the accessibility and capacity to process enormous amounts of data (Chaffin et al., 2015). The described development subsequently affects the business and academic landscapes in two ways: (1) through increased accuracy when analyzing common problems, and (2) an increase in the number of problems that can be analyzed (George, Osinga, Lavie, & Scott, 2016). In essence, more problems can be solved, and the answers are better than ever.

Recognizing the development that has taken place for the past decades solicits considerations of implications for businesses. For any organization wishing to remain competitive in their respective line of business, unravelling the potential of emerging scientific fields is crucial for the future of the business. Yet understanding the potential of an opportunity requires an understanding of its value; in essence being capable to map the relative difference between what the said potential offers, and present reality.

1.1 Background

Historically, thorough mappings of the extent to which analytics and data science have been adopted throughout businesses has been conducted extensively abroad, looking at what business impact it has the potential to generate (Henke et al., 2016). Furthermore, there have been parallel studies which addressed the extent to which academia had progressed within the field at the time (Chen et al., 2012. These studies systematically addressed the situation by investigating demands on competencies, on the organization, and ultimately what business impact said organizations had managed to extract from working more intimately with data (Chen et al., 2012; Henke et al., 2016). Half a decade later, a similar bibliometric study was conducted in Sweden in 2017 which found significant differences in how countries such as the US and Sweden had progressed within the field of artificial intelligence (AI) (Hanson, Jeppsson, & Nordlund, 2017), which is inherently connected to the field of data science.

While errors of conflating causation with correlation are desirable to avoid, an arguably intuitive train of thought would attribute some of the foreign success to its thorough historic self-examination; constructive criticism not uncommonly provokes an urge to improve. Meanwhile, the same train of thought suggests that without the process of such a self-reflection, the improvement may never happen yet alone appear desirable or necessary. Consequently, this thesis will constitute one perspective of self-reflection.

1.2 Context description

This thesis is not conventional in the sense that it does not serve a specific client and is purely academic. Therefore, a brief overview of the context in which the thesis was formed is deemed warranted.

There are three stakeholder parties involved in this thesis. First there are the authors: two students majoring Industrial Engineering and Management at Chalmers University of Technology, and in the creation of the thesis both students were pursuing a Master's degree within Management and Economics of Innovation. The second stakeholder is Chalmers University of Technology; the aspirations and ambitions of said institution is to align educational supply with the ever developing business demand is the raison d'être of this thesis. This is to say that the exclusive source of the initiative stems from Chalmers and the institution is ultimately the primary stakeholder. Third, and last, are the interviewed people and the firms they represent. These are representatives of 23 Swedish businesses, of different size and industry in which they operate.

1.3 Purpose & research questions

The purpose of this thesis is to provide a source of orientation to the Swedish landscape of data science, with regard to the extent that it has been incorporated in businesses to drive growth and business impact. The aim is to address the purpose systematically, why the following research questions have been formulated:

(i) How can the landscape of Swedish data science practice be described with regard to the following three aspects: (1) competence requirements, (2) organizational demands, and (3) business impact?

(ii) How has the landscape described in (i) looked historically, and what can it be expected to become in the future?

1.4 Delimitations

In an effort to obtain a scope for the study which is reasonable to cover given the resources allocated to the project, three major limitations have been made: (1) the sole focus is on Swedish businesses, (2) business representatives are the exclusive sources for data, and (3) the study abstains from recommendations on how to develop the landscape. Below, these delimitations are explained in brief.

Firstly, the study is limited to covering exclusively Swedish businesses. While other countries would be interesting to cover, resources of time were too scarce to expand further than the current scope. Secondly, the empirical data collection is limited to 23 interviews exclusively from business representatives, meaning that academic representatives indirectly have been excluded. This is because the initial purpose was to solely describe the Swedish business landscape, but during many interviews the discussions drew upon close connections to the role of academia. These data points were judged to contribute strongly to the value of the thesis, and consequently they were included. As will be discussed in the methodology chapter, expanding the data set in such a manner is reasonable as long as it enrichens and contributes to the study (Maxwell & Chmiel, 2005). However, given more time and resources, it would undoubtedly contribute value to the study to interview representatives of the Swedish academic community. Lastly, the study is delimited from providing recommendations on how to improve the situation described in the thesis. Ideas or suggestions for improvement that appear within the thesis exclusively originates from the interviewed people, and are included to the extent that they are considered to enrichen the mapping of the data science landscape.

2 Theoretical framework

In this chapter the theoretical framework is presented. To help steer the method of choice for this thesis, the following three sections of theory were studied: (1) data science, which provides a general description of the field and its development, (2) data science and organizations, in which the challenges and opportunities faced by organizations looking to adopt data science are looked into, and lastly (3) change management, which describes a general process for organizations facing major change. Since chapter one testifies to data science being an emergent phenomenon and increasingly important for organizations to embrace, theory on change management has been deemed warranted to equip the framework with ideas of how to manage the necessary transitions.

2.1 Data science

This chapter will provide insight to the data science area and what it includes on a general level. Theory describing the development of data science together with an attempt at categorization of the analytics landscape will also be presented.

2.1.1 An interdisciplinary field

Data science is an interdisciplinary field which aims to provide insight through the analysis of data. On a high level data science can be defined as follows: *"a set of fundamental principles that support and guide the principled extraction of information and knowledge from data"* (Provost & Fawcett, 2013, p. 52). In similar fashion Mahmood (2016) argues that data science mainly aims to assist strategic decision making through the provision of tools, techniques, and scientific methods that facilitate the process of capturing and analyzing data.

As data science encompasses a wide range of technologies and methods, the boundaries of the term become somewhat blurry, causing uncertainties and divided opinions regarding the necessity of the term. An activity is not necessarily characterized under the data science label simply because it involves data, which is a common misconception (Provost & Fawcett, 2013). With its broad definition and wide range of applications one must not make the mistake of considering data science in isolation, as it is highly a matter of considering a mix of other existing disciplines (Dhar, 2013).

2.1.2 Field development

The amount of data available for analysis has dramatically increased in terms of volume; everything is recorded and stored, be it something that is said or bought and the advent of big data has brought with it new challenges and opportunities when it comes to analyzing data. The three V's used to define big data (illustrated in figure 7): volume, variety, velocity, imply that the data is highly heterogeneous, and as a result the analysis of such data has to deal with complex relationships (Dhar, 2013). Data science requires an extended focus, rather than purely analyzing data the focus has shifted towards understanding the surrounding environment where data is gathered; Cao (2017, p. 8) phrases this in the following way: *"data science is the new generation of statistics, is a consolidation of several interdisciplinary fields, or is a new body of knowledge"*.

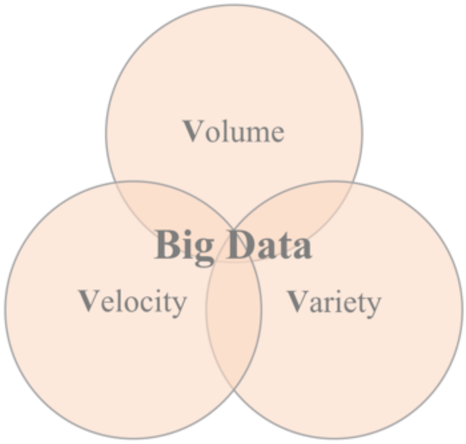

Figure 7: Properties of big data.

The advancement of data science has been significantly attributed to the development of frameworks and softwares capable of processing immense data sets in parallel (Delgado, 2016). Such examples are Spark and Hadoop, which have enabled widespread adoption of more advanced data analytics by providing means of processing such data sets (Delgado, 2016). In essence, the increase of data availability and the possibility to properly analyze it have given rise to increased business applications for organizations looking to leverage data to guide strategic decision making.

2.1.3 The analytics landscape

Due to the nature of data science it is not easy to define boundaries or set clear categories for what it encompasses, there exist however some commonalities for how to describe the analytics landscape of data science. Classifying the analytics landscape after degree of complexity is common in research and provides a picture relatable to the world of business (Akerkar, 2013). Adopting this approach, a model popularised by Gartner, Inc. consisting of four levels of complexity is often brought forward, the four levels being: (1) descriptive analytics, (2) diagnostic analytics, (3) predictive analytics, and (4) prescriptive analytics (Gartner, 2014).

Descriptive analytics in its most basic form encompasses analysis of historical data with the intention of extracting insight from it (Gupta, 2016). While descriptive analytics aim to describe what has happened, diagnostic analytics tries to answer the question of why it happened. These two lower levels of complexity are seen as the more basic forms of advanced analytics and represents the bulk of data analytics taking place in the business world today, often present in the form of reports providing historical insight and the analysis of such reports (Seddon, Constantinidis, Tamm, & Dod, 2017).

Moving onto predictive analytics the focus shifts from understanding the past to forecasting the future. As with the lower levels of complexity, predictive analytics also makes use of historical data, although with a focus on developing and training models able to predict future values and new data (Chen et al., 2012). The highest level of complexity is represented by prescriptive analytics. Moving on from predicting what is going to happen, prescriptive analytics aims to provide actions for how to handle the predicted outcomes in order to ensure the optimal outcome (Akerkar, 2013). An interpretation of Gartner's model for analytics complexity can be seen in figure 8. It is important not to consider the different levels of complexity in isolation, just as there is no easy way to define data science the same is true when describing the different levels of complexity present in advanced data analytics.

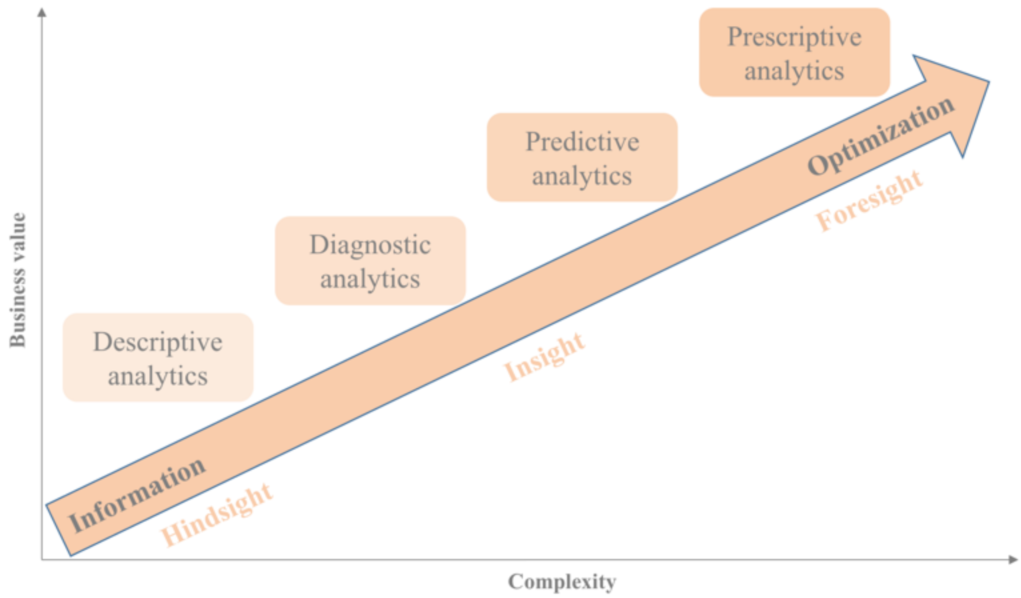

Figure 8: The Gartner levels of business value extraction.

2.2 Data science and organizations

In this chapter the implications and opportunities for businesses looking to embrace data science methodologies and technologies are presented. A brief description of data science's emergence in the world of business is followed by a deeper look into the three focus areas: Competence, Organization, and Business impact.

2.2.1 From business intelligence to data science

For a long time, business intelligence has been common practice for businesses; the term was popularized back in the 1990s and is still highly relevant (Chen et al, 2012). Business intelligence can arguably be regarded as an umbrella term encompassing tools and methodologies related to the utilization of information to improve decisions for businesses, Gartner (2013) describes business intelligence as: *"an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance"*. Business analytics is a term commonly used in an attempt to try describe the increase in complexity and use of advanced analytics in businesses. Larson and Chang (2016) identify the trend of data science as a part of more advanced business intelligence, while Schmarzo (2015) highlights clear differences between data science and business intelligence.

Regardless of which view one takes, it is clear that the advent of new technologies and big data has transformed the analytics landscape, creating vast opportunities and challenges for businesses in the process. Looking at traditional business intelligence it has mainly revolved around descriptive analytics, with reporting being the main outcome of data analysis in organizations. High technology firms started to emerge in the 00's following the emergence of big data and technologies that enabled more complex forms of analytics (Davenport, 2014). However, Davenport (2014) notes that the most basic descriptive analytics has remained the most prominently used form of analytics. This is not necessarily an indication of development lag, as older technologies will remain an integral part for businesses; it is not a matter of adopting one or the other, but rather absorbing new capabilities and focus has increasingly shifted towards predictive and prescriptive analytics (Larson & Chang, 2016). To highlight the shift towards more advanced analytics, Schmarzo (2015) categorizes data science as the more advanced form of analytics in relation to the descriptive nature of business intelligence. An interpretation of the difference between traditional business intelligence and data science can be seen in figure 9.

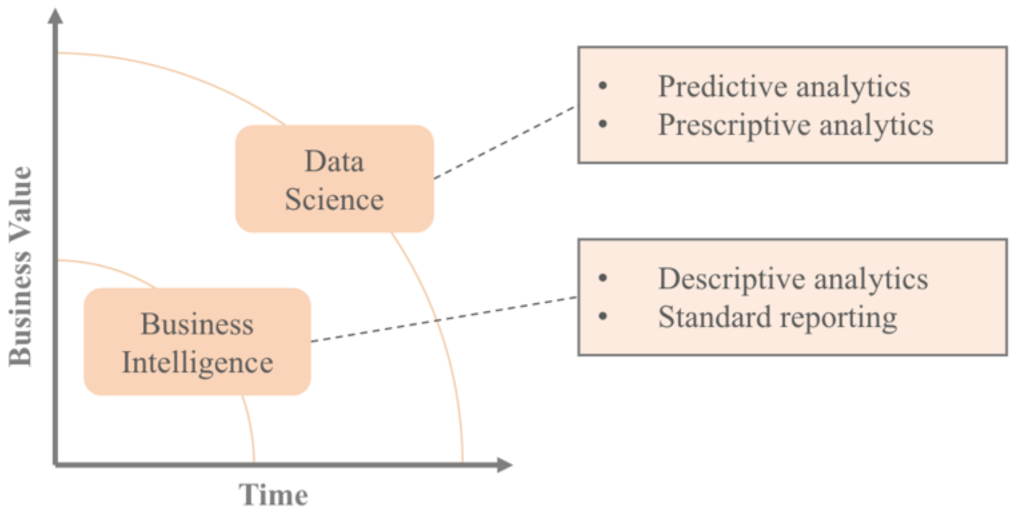

Figure 9: Schmarzo's (2015) distinction between business intelligence and data science.

Firms look to structure their organization in a way that enables them to leverage the opportunities presented by data science. Although, it is important to recognize that moving from reporting to more advanced analytics is not a matter of merely increasing volume of established analytics operations. Being data driven is not about putting more data into reports, it is a matter of embracing a new way of tackling problems and making decisions (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). Companies born digital have an advantage over more traditional companies and are normally further ahead in the process of utilizing advanced analytics, not due to superior potential but rather because of better organizational alignment.

2.2.2 Competence

Finding the right competence is crucial if organizations are to successfully embrace data science practices, and it is also one of the main challenges. The analytics capability cannot function in isolation and thus the skill set requirements become broad (Vidgen, Shaw, & Grant, 2017). Data scientists not only need to be proficient with numbers and statistics, they also need to have a curious mindset and be able to cooperate efficiently.

This part contains theory related to two areas on the topic of competencies: (1) demand, and (2) supply of competence. They will be handled in turn below.

Demand

The competence demands of data science is evident all across an organization, just as the data scientists will need to understand the business aspect the business stakeholders needs to think more like a data scientist (Schmarzo, 2015). If data science is to make businesses increasingly data driven by supporting decision making, employees throughout all business units must be able to efficiently interact with data science teams (Provost & Fawcett, 2013). The cross-functional nature of data science brings with it demand for translators, tasked with the roles to facilitate the integration between analytics and business. According to Gartner (2016) these roles often exist unknowingly in the form of individuals doing extra work. For an organization to become truly data-driven the value of translators needs to be recognized and actively managed (Gartner, 2016).

Retaining competence is another matter of great importance. If organizations are to develop their analytics capabilities properly a solid core is necessary, requiring the retention of core competencies. A survey performed by MIT Sloan Management Review showed that about 40% of organizations do not only struggle in the matter of finding the appropriate competence, but also in retaining it (Ransbotham, Kiron, & Prentice, 2015). As analytics is becoming an acknowledged path to value the competition for competence is increasing (Ransbotham et al., 2015). Competence is also highly a managerial matter. McAfee et al. (2012) emphasize the potential to revolutionize management through data driven decisions. Those in positions of influence need to rely on data rather than intuition and experience if improved decision making is to be achieved through the use of data science (McAfee et al., 2012). Organizations invest in analytics to gain competitive advantage and/or to improve efficiency, and it is necessary that managers understand the influence analytic capabilities can have on organizational performance in order to realize the potential benefits (Seddon et al., 2017).

Supply

Data analytics is no longer a matter isolated to the specialists, and the trend points towards data science becoming a fundamental skill for the twenty-first century workforce (Schuff, 2018). As data availability along with tools enabling proper handling and analysis is increasing drastically all parts of our society is being affected, changing jobs across several industries in the process (Van der Aalst, 2014). The data science programs and courses that do exist are mainly focused on graduates and specialists, the need is much broader than that and academia must adapt (Dichev & Dicheva, 2017). As a consequence there exists a lack of general knowledge regarding how data can generate value for organizations (Wixom et al., 2014). Gil (2014) adds on that note, stating that a lack of accessible courses for non-programmers exists, limiting students ability to assist in the development of data science once they enter the workforce. Current education systems are not equipped to produce individuals with the wide array of skills and knowledge necessary to make best use of the analytical capabilities which new technologies are enabling (Education Development Center [EDC], 2016).

The interdisciplinary nature of data science is important to consider. To best meet future needs and enable students to navigate data-rich environments courses in data science should be made available across several disciplines (Dichev & Dicheva, 2017). Throughout literature data literacy is highlighted as the main competence academia should seek to teach students. Dichev and Dicheva (2017) describe a data literate individual as someone capable of collecting, evaluating, analyzing, and interpreting data, presenting results, and making decisions based on them. Building on the same description of data literacy, EDC (2016) emphasize that no single individual is expected to perform them all, rather it is a matter of cross-functional collaboration across organizations. The interdisciplinary nature of data science helps underscore data literacy as a proficiency more important than statistics and technology (Schuff, 2018). Rather than sophisticated analytics techniques, data literacy needs to be recognized as a core skill for students looking to make an impact when entering the workforce (Schuff, 2018).

If academia is to truly embrace data science the field needs to be acknowledged as a new discipline. Data science must be given a distinct place in academia through formal definition of research areas in the field (Van der Aalst, 2014; Schutt & O'Neil, 2013). Organizations emphasize that students often lack perspective with regards to how data and analytics is handled in the real-world (Wixom et al., 2014). Kollwitz, Dinter, and Krawatzeck (2018) suggest that this issue can be dealt with through the introduction of more hands-on exercises, enabling students to attain knowledge relevant for usage in their future worklife. The argument of practical hands-on projects helping bridge the existing gap between industry and universities is shared by Schiller, Goul, Lyer, Sharda, and Schrader (2014) who underline the benefits of using real data sets as it helps students understand the challenges associated with large quantities of data sources from different sources. Regardless of method used, the importance of addressing it in any arbitrary way is apparent since it is clear that academia is trailing their industry counterparts in the current state (Jin, Wah, Cheng, Wang, 2015).

2.2.3 Organization

The organizational infrastructure decides the limits and potentials for what data science can accomplish. Embracing advanced analytics requires infrastructure able to keep up with the speed and rapid change demanded (Gartner, 2015). Traditional view of analytics as an isolated manner is not sufficient and the underlying infrastructure needs to adapt (Vidgen et al., 2017). Whereas traditional business intelligence deals with analyzing the past, newer data science technologies are about shaping future decisions and require organizations to have infrastructure that supports the three V's of big data. Real-time analytics of data stemming from various sources demands reactive and agile processes (Delgado, 2016). This requires organizations to rethink the way data is managed, as siloed practices, both in regards to storage and analysis, is a common phenomena (Schmarzo, 2015; Gartner, 2016). As data is truly everywhere in organizations, a coordinated approach with regards to data infrastructure is necessary in order to avoid embedded inertia.

Mastering data management and improving data quality is a matter of effective communication and sharing across organizational functions (Asadi Someh, Wixom, Davern, & Shanks, 2017). Schmarzo (2015) highlights the necessity to move on from traditional data warehousing and embrace the concept of data lakes. A 'data lake' is a new type of data repository enabling advanced analytics of big data by providing sufficient storage and processing power (Larson & Chang, 2016). While proper data management is necessary it is also a prerequisite for coordination; technology enables data science, it does not guarantee successful application of it (Schmarzo, 2015). As analytics aims to improve business decisions, cooperation and understanding across functions becomes paramount (Gartner, 2016). Shared language as well as a common understanding of data and the goals of analytics is important all across the organization (Asadi Someh et al., 2017). To become truly data driven analytics and business need to work closely together as they depend on each other to create value in an effective manner (Schmarzo, 2015).

When facing the issue of organizing analytics organizations need to plan for how to structure. According to Grossman and Siegel (2014) organizations need to consider three main challenges when organizing analytics operations; (1) identifying and resourcing analytics, (2) obtaining the necessary data, (3) deploying the models. Three distinct ways of structuring the analytics function exist, organizations can choose to gather their analytics in a central unit or go the other direction and have analytic competence spread across business units in a decentralized manner. The third alternative implies a hybrid option between the two, where a central unit, often called a center of excellence, houses a critical level of analytics competence with the remaining competence spread across the organizational business units. An illustration of the hybrid model can be seen in figure 10. While no single structure is perfect they each bring specific traits and it is much a matter of balancing trade-offs between the various options (Davenport, 2013).

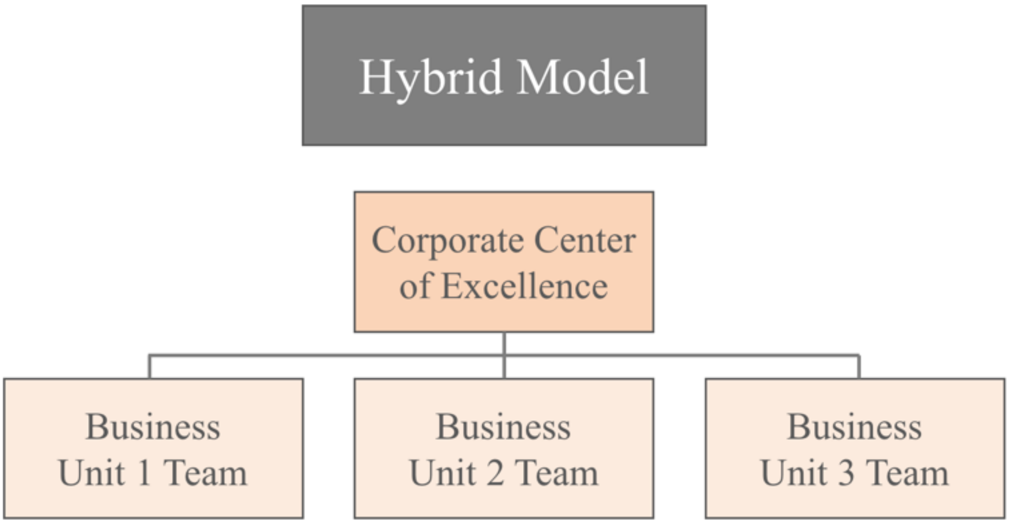

Figure 10: Conceptualized dynamics of hybrid structures in organizations.

A decentralised structure has the advantage of putting data analytics right where the business takes place, making it easier to achieve collaboration and identifying key impact areas (Grossman & Siegel, 2014). Due to the nature of how the analytics process oftentimes is initiated, through the initiatives of individual employees or units, a decentralized structure is commonly found within organizations where no clear organization wide analytics strategy has been established (Franks, 2014). In order to achieve efficiency and enable organization wide efforts, moving away from the fully decentralized structure is necessary (Grossman & Siegel, 2014). A centralized structure suggests that analytics becomes a function on its own, separate from other existing business units. This allows for the required competence and data to be gathered in one place, enabling efficient development and testing of analytics models (Grossman & Siegel, 2014). The hybrid model enables central coordination through a critical mass of centralized competence, along with domain-specific insight through close collaboration with decision makers of individual business units (Davenport, 2013).

All of the options for structuring the analytics process exist in various forms across different organizations, it does not mean that they are all preferable. It is much a matter of matching the organizations maturity in the field with the proper structural solution (Dallemule & Davenport, 2017). If there exist no clear analytics strategy, efforts concerned with consolidating the competence will have limited effect. In the end a hybrid structure is generally deemed the most viable solution for organizations aiming to become truly data driven (Davenport, 2013). However, a hybrid structure puts high demands on the ability of collaboration between analytics and business, it is therefore crucial to have the right competence and processes in place for it to be beneficial (Franks, 2014).

2.2.4 Business impact

Data science and more advanced analytics have the potential to deliver business benefits all across organizations. By presenting cases from e-commerce and market intelligence, e-government and politics, science and technology, smart health and well-being, as well as security and public safety, Chen et al. (2012) emphasizes how domain-specific analytics in various areas can result in significant positive impact. According to Seddon et al. (2017) the analytical resources of an organization contributes to business value either by providing insights affecting decisions to take action with the use of existing organizational resources, or by influencing actions to change the organizational resources. On a high level business analytics enable new insights and effective use of resources (Seddon et al., 2017).

In an attempt to categorize the vast potential of data science Meulen (2018) presents five potential areas of impact: (1) Innovation; data analytics can help unveil new or better solutions. (2) Exploration; exploring data can reveal before unknown patterns resulting in added value. (3) Prototyping; data science can allow for agile testing of potential new solutions. (4) Refinement; analyzing data of current processes and products allows for continuous refinements to be made, and (5) Firefighting; dealing with undesirable situations data analytics can help facilitate the process of identifying root causes. Potential applications within each category outlined above is seemingly infinite; when dealing with analytical solutions there is no 'one-size-fits-all' as it is much a matter of adapting to the situation or problem at hand (Arora & Malik, 2015).

When assessing the potential impact of analytics Franks (2014) emphasizes the importance to differentiate between traditional artisanal analytics processes and operational applications. Whereas both add value, operational analytics refers to a state where analytics is embedded into processes assisting decision making in a predictive and prescriptive manner, in relation to traditional batch analytics whose nature is 'one-off'; one solution for one problem, without much notions in mind to reuse the created solutions. Similarly to how Gartner classifies the analytics landscape with regards to complexity, Arora and Malik (2015) argue that the potential to deliver business value through analytics depends on the depth of analysis. As has been mentioned before, on the most basic level descriptive analytics allows for extraction of useful information through analysis of historical data, and further predictive analytics enables predictions of future outcomes, while finally prescriptive analytics aims to assist decision making directly.

Realizing the potential benefits of data science is not simply a matter of adopting the appropriate resources and then moving on to execution. If businesses are to unfold the vast opportunities which data science enables it is a matter of moving through a process of development (Hürtgen & Mohr, 2018). Morabito (2015) emphasizes organizations maturity in data analytics as key to extract business impact, stating that there exists phases necessary to go through in order to reach maturity. In essence, it is not possible for organizations to simply move straight to more complex analytics such as predictive and prescriptive methods without first mastering traditional analytics of descriptive nature (Franks, 2014). Before being able to derive insights and optimize business decisions through the analysis of data, sufficient resources and processes for simply monitoring must be in place (Morabito, 2015). Fundamentally, as the potential benefits derived from data science depends on the analytical resources (Seddon et al., 2017), business must first make sure that proper generation and collection of data is in place before insights can be derived and more advanced forms of analytics adopted (Hürtgen & Mohr, 2018).

2.3 Change management

In this chapter the area of change management is explored. The chapter starts of by defining the area, followed by arguments for the relevance of change managements in businesses. Lastly, challenges and recommended approaches when looking to adopt change management is presented.

2.3.1 Managing change

Change management is the systematic approach of managing the transition or transformation of processes, structures, and people within an organization Tamilarasu (2012). The argument is that change management is basically about shifting from a current state to a future desired one. Any organization looking to influence their future state needs to consider change management, *"Organizational change cannot be separated from organizational strategy, or vice versa"* - Todnem By (2015, p. 369). After carrying out extensive literature research, Rosenbaum, More, and Steane (2018) conclude that change management can vary in several dimensions, it can be implemented through continuous or step-wise change, it may be planned for or emergent, the initiatives can originate top-down or bottom-up and lastly, depending on size and impact it can be either incremental or transformational. Change management is not a novel concept, but the need for it is of utmost relevance in the modern world of business.

There should be no hesitation regarding the need for change management, having an idea and target for where one wants and needs to be in the future is paramount for any organization (Todnem By, 2015). Franklin (2014) emphasizes how the need for continuous change does not tolerate hierarchical organizational structures and lengthy decision making processes. The often unpredictable need for change explains why reactive approaches to change management are common (Todnem By, 2015), but if organizations are to embrace all aspects of change management a proactive approach is necessary (Tamilarasu, 2012). Tamilarasu (2012) mentions adapting to change, controlling change, and effecting change as the three core aspects of change management.

2.3.2 The change process

For organizations approaching change management several challenges become relevant to consider. A vast array of frameworks for how to tackle these challenges exist and in the end they mainly highlight similar approaches and recommendations (Rosenbaum et al., 2018). In his well-known eight step framework, Kotter (1996) identifies several challenges associated with change efforts in organizations: (1) establishing a sense of urgency, (2) forming a guiding coalition, (3) creating a vision, (4) communicating the vision, (5) empowering people, (6) planning for short-term wins, (7) consolidating success, and (8) anchoring changes. Below follows a deeper explanation of Kotter's (1996) steps, complemented with contributions from other scholars on the topics mentioned.

(1) First of all it is important to establish a sense of urgency if change is to be enabled, as people are more inclined to deal with pressing problems and opportunities (Kotter,

1996). Mento, Jones, and Dirndorfer (2010) extend on this and state that if the sense of urgency is lost, focus should be put on creating something new rather than fixing existing problems, to increase motivation.

(2) If change is to take place it is important to reach a critical mass, sufficient resources need to be in place and properly coordinated for change programs to succeed (Gill, 2002). Without the commitment of top management and the people affected, initiatives of transformative nature are likely to produce substandard outcomes (Sirkin, Keenan, & Jackson, 2005).

(3) Willingness to change comes from the underlying vision (Mento et al., 2010). In essence, a vision serves as the foundation on which surrounding idea for change can be initiated. Consequently, a firm should set up a vision to fulfill these purposes (Kotter, 1996).

(4) Having a clear vision is important, but it is also useless without proper communication, if an organization is to truly embrace change the employees must be convinced of its importance (Tamilarasu, 2012). Translating the vision into actions showing tangible results can act as a powerful motivator (Luecke, 2003).

(5) Empowerment is brought up repeatedly across literature as a crucial factor for overcoming resistance existing within organizations (Rosenbaum et al., 2018). When individuals are able to make an impact through their own actions they are much more likely to embrace change. Empowerment is as much about motivating as it is making sure knowledge, skills, and resources necessary for change are spread across an organization (Gill, 2002).

(6) Another way to make sure that motivation persists is through "small" wins (i.e. specific and visible results stemming from the change effort) (Mento et al., 2010). Without visible results there is a risk of losing any sort of established momentum, especially if the change program spans over a long period of time (Sirkin et al., 2005).

(7) Any kind of momentum established in a change effort should be used as fuel for further change as small wins can help convince organizations to aim for larger change projects (Kotter, 1996).

(8) Lastly, if change management is to have any lasting results institutionalization is key. It is necessary to develop sustainable shared values, which support the overall vision, continuously throughout the change process if a desired organizational culture is to be reached (Gill, 2002).

3 Method

This chapter provides an overview of the approach to the collection, processing, and analysis of data that has been implemented to address the research questions. It will consist of four parts: (1) research strategy, (2) research process, (3) trustworthiness, and (4) ethics.

3.1 Research strategy

Selection of research strategy concerns the extent to which qualitative and quantitative research is used (Bryman & Bell, 2015). The nature of a quantitative method revolves around the idea of numerically testing the relationships between a set of circumstances, or 'independent variables', and the effect which said circumstances have on an investigated outcome, often referred to as 'dependent variable' (Lakshman, Sinha, Biswas, Charles, & Arora, 2000). In turn, qualitative analysis seeks to interpret, understand, and make sense of social contexts and phenomena, assessing what meaning they bring to people (Greenhalgh & Taylor, 1997). As the desired outcome of the thesis aligns strongly with what qualitative analysis generally accomplishes it was evidently selected as the proper approach to the study.

3.2 Research process

The following section covers the research process, which consisted of four parts: (1) sampling, (2) data collection, (3) data processing, and (4) analysis. Below follows an outline of the work done in each part.

3.2.1 Sampling

When defining the sample universe (i.e. all the companies possible to interview with regards to criteria for inclusion and exclusion), the vision was to achieve a representative view of the Swedish industry, which is desirable as Robinson (2013) argues that sample heterogeneity allows for findings to be more widely applicable. Here, the trade-off between accuracy and breadth in the data should be recognized: the holistic perspective on the Swedish business landscape that this thesis takes most likely comes at a cost of specificity and detail in the possible analysis. An alternative approach would consequently be to focus on a niche segment of businesses, and enable a higher level of findings accuracy relative to reality.

Having acknowledged the trade-off above, this thesis prioritizes breadth over accuracy in an aspiration to provide a holistic perspective of the data science use in Swedish firms. Hence, two tactics were employed: (1) seek a diverse sample with regard to the represented business sectors and corporate sizes, and (2) seek representatives of samples that span the entire spectrum of data science proficiency, generalizable to the Swedish industry as a whole. In order to systematize the second tactic, two conceptual segments of data science maturity were formed for the purposes of gaining representation in both: (1) industry firms, enterprises with business models not inherently rooted in digital technologies, and (2) digital firms, companies that are digital by nature. The rationale behind the creation of these two segments is inarguably subjective; data science practices were simply expected to be more sophisticated and developed by firms whose business models revolve around digitalization. The validity of these notions has been verified by scholars mentioned in the theoretical framework (McAfee et al., 2012). Needless to say, while the majority of interviewed firms could be assigned to one of the segments without much confusion or debate, some allocations called for a higher degree of subjectivity and the possibility of sub-optimal decisions are certainly recognized here.

Furthermore, the specific firm representatives to interview were selected on the basis of three factors: (1) corporate rank, (2) ability to represent the data science operations, and (3) availability. In essence, this resulted in the desire to find representatives of the given data science practice, as high up in the corporate ladder as possible, that had time to meet for an interview. Given the efforts outlined above, the turnout was a sample of 23 companies with the following properties:

- 28 representatives: 4 data scientists, 2 C-level executives and 22 Head of Analytics or similar
- 9 digital firms ranging from 25 - 3,000 employees
- 14 industry firms ranging from 1,500 - 200,000 employees
- 14 represented industries/sectors: retail, telecommunication, airline, hotel, healthcare, manufacturing, construction, streaming, transportation, advertising, venture capital, media, government authority and software engineering

3.2.2 Data collection: Interviews

Data was collected through the 23 interviews. Below, two aspects of the interviews are outlined: (1) the structure, and (2) the generic process through which they were conducted.

The interview structure

Each interview was initiated with the premise of gathering data related to three core aspects: competence, organization and business impact. The rationale behind the selection of these particular aspects merits clarification. Data science is a broad field to survey, and consequently the authors sought to break the field down into factors that span it well for purposes of systematization and structure. Three activities were performed which led to eventually focusing on the three mentioned aspects: (1) a literature study overviewing what scholars perceive as important factors in the field of data science, (2) a subsequent creation of a draft questionnaire based on the literature study, touching on a vast list of issues that potentially could be of interest when interviewing, and (3) a workshop with the author's thesis supervisor. During said workshop, the outcome of activity (1) and (2) were summarized and it was eventually concluded that all questions formulated in the questionnaire could either be allocated to competence, organization or business impact, which led to the focus on these specific areas.

Consequently, a semi-structured approach to the interviews was selected as it was desirable to leave room within the three core aspects for flexibility, given the high diversity of interview companies that likely had diverse experiences to share. A semi-structured approach is the intermediary between structured and unstructured interviews (Bryman & Bell, 2015). In practice, interviews were conducted with the support of two tools: a questionnaire (see Appendix A), and a case framework, which essentially was the conceptualization of the research questions in the form of a matrix. See figure 11 below:

	Competence requirements	Organizational challenges	Business impact
Historic			
Current			
Future			

Figure 11: Case framework used for interviews.

The process of the interviews

Each interview was initiated by one of the interviewers hand-drawing the above matrix, explaining the expectations and desired areas of information to the interviewee. This was motivated by two reasons: (1) a desire to build trust and cohesiveness with the interviewees, which is recommended by Alvesson (2003) to ensure 'correct data' to be gathered, and (2) to create a systematic manner of taking notes during the interview with readiness to expand the data collection. Said readiness did prove useful during almost all interviews as the interviewers consistently asked the interviewees for complementary data if they had anything more to add outside the scope of the case framework. By following the rationale of Maxwell and Chmiel (2005) it should be acceptable to expand the data collection in such a manner as long as it contributes to the overall purpose of the study. Eisenhardt (1989) agrees, stating that new data collection opportunities should be exploited if they provide new theoretical insight. However, this flexibility should not be regarded as an excuse to be unsystematic, a risk which was mitigated by the practice of handwriting the framework above and taking notes within it. Generally speaking, the interviews lasted as long as the framework remained unfinished. To clarify: the implicit argument is that the use of the matrix and active note taking within it allowed the interview to proceed systematically forward.

The interviews were recorded which was beneficial as it generally allows for capturing of the interviews in their entirety without interrupting the free flow of conversation (Krishnaswamy & Satyaprasad, 2010). The average interview lasted 60 minutes and whenever possible (19 times out of 23) the interviews were carried out in person as it allows for more accurate interpretation of the information shared, while also facilitating the ability to steer the direction of the interviews for the interviewer (Sreejesh, Mohapatra, & Anusree, 2014).

3.2.3 Processing & Analysis

Frameworks are useful to employ in order to structure and make sense of qualitative data (Nigatu Haregu, 2012). For the specific case of this thesis, it proved particularly true as the selected framework served as the driving force throughout the entire research process. Data was collected according to the sections of competencies, organization and business impact, and subsequently analyzed in the same manner through the content analysis technique. Content analysis is a method of analyzing written, verbal or visual communication messages (Cole, 1988).

Content analysis can be performed in two ways: inductively or deductively (Elo & Kyngäs, 2008). Inductive content analysis is used in cases where there are no previous studies dealing with the phenomenon or when it is fragmented. A deductive approach is useful if the general aim was to test a previous theory in a different situation or to compare categories at different time periods. The selected approach was a combination of the two; the nature of the case framework constituted the deductive part of the analysis, as data was consistently gathered with the purpose to segment it into one of the three buckets. The inductive analysis was subsequent to this as categorization of the data within each bucket had not been done, and thus required novel open coding and interpretation of the content. Consequently, the data was allocated to either competence, organization, or business impact, and then it was openly coded, the codes were grouped and categorized, which eventually produced a set of areas upon which the empirical findings could be based.

The goal has been to illuminate the coded categories as clearly as possible in the empirical findings. For example, the section on organizational challenges distinguishes between them through subsequently bolded subheadings. Such transparency is consistently sought throughout the presentation of the empirics, but it was not deemed possible to maintain that structure at all times; if the volume of data regarding a certain category was not large enough, it seemed excessive to dedicate an entire subheading to it. An example of such is in the empirical findings related to the five identified competencies; apart from explaining what competencies the categories referred to, there was not much more to add on and consequently they were covered in bulk.

3.3 Trustworthiness

In order to address the trustworthiness of the study, four aspects should be addressed: (1) credibility, (2) dependability, (3) confirmability and (4) transferability, with credibility being the most important criterion (Bryman & Bell, 2015; Polit & Beck, 2014). Credibility is the extent to which the findings confidently can be expected to reflect the truth according to Connelly (2016), who further suggests that prolonged engagement with participants is one of several means to increase credibility. To that account, meeting with the interviewees physically and engaging them in the case framework (described above) attributed to the credibility of the data collected, as interviewees were physically engaged in the goals of the interview. However, although a strong majority of the interviews were conducted physically, not all of them were which damages credibility in this sense. Another way to increase credibility is through triangulation (Woodside, 2010), as it involves cross-examining data from different sources. For that purpose, the collected data from the interviews was checked and compared with the learnings gathered from literature.

While efforts have been made to increase credibility, it undoubtedly suffers from the limited contact with each firm. At most, two people from a given firm were interviewed and the majority of interviews had just one representative. This decreases credibility because of two reasons: (1) the subjectivity of the interviewee goes untriangulated, and (2) different firms give their testimonies to the research questions based on different perspectives; some companies have provided their data scientist perspective, others have spoken from the point of view of c-level executives. Needless to say, this will affect the credibility when making comparisons between statements of individuals with completely different organizational roles since they respond according to their own perspectives, ambitions and ideas. In order to address these challenges of credibility, the thesis will contextualize and connect the findings to the role and background of the individuals that contributed to the given statements, in an attempt to present the findings as transparently as possible.

Furthermore, dependability refers to the stability of the data over time and the conditions of the study (Polit & Beck, 2014). Dependability is to an extent achieved through triangulation, while it should be stated that what the thesis aims to cover is the current state of data science practice in Sweden, the very nature of such a task implies possible and likely change in the future. Therefore, striving for dependability to a certain extent undermines the purpose of the study.

Continuing, conformability is the neutrality or the degree to which findings are consistent and could be repeated (Connelly, 2016). Essentially, conformability challenges the objectivity of the study and during different phases of the research process, even mild degrees of objectivity have been unfeasible expectations. An event in which this proved true was the sampling; the companies selected to interview with were allocated to one of two segments with varying maturity within the field of data science, based on the expectation of the authors of the study. Needless to say, this was a subjective differentiation on behalf of the authors. However, while impossible to perform with objectivity, the activity was deemed useful and necessary for the purposes of striving towards a representative and credible sample collection. In any case, minimization of subjectivity has been strived for in two ways: (1) transcriptions have been shared with the interviewees and they have been given the chance to comment and (2) drafts of the report have been iterated with supervisors, student colleagues and the participating firms for comments.

Finally, transferability concerns the extent to which findings are useful to persons in other settings according to Connelly (2016), who also suggests that a means of accomplishing transferability is to focus on the informants and their stories without suggesting that it is everyone's story. In the empirical findings, this will be strived for by sharing quotes from the interviewees, omitting intermediary interpretations and rewritings on behalf of the authors.

3.4 Ethics

The authors of the thesis have given their best efforts to provide clarity and transparency through not only this chapter of methodology, but in every interaction with the interview subjects. While Chalmers is undoubtedly the stakeholder with the most power and influence on the outcome of this thesis, in the perspective of the authors the interviewed companies stand at the center of importance. They are ultimately the ones that have been the most generous to contribute to the fruitfulness of the study and without them the value of this study would be nonexistent.

For the reasons mentioned above, the following four actions have been taken to engage and protect the interests of the interviewed subjects: (1) continuous efforts to communicate the expected outcome of the thesis, prior to as well as during the time of the interview. This was done through phone conversations, e-mail exchanges and sharing the questionnaire prior to each respective interview. Regrettably, all interview subjects were not given the opportunity to investigate the questionnaire beforehand, at the fault of the author's mistakes. This is not believed to have mentionably affected the outcome of the result as all the 'unprepared' interviewees were able to address the questions, seemingly without difficulty. (2) Quotes included in the report have been anonymized, and if the quote involved mentions of the specific firm said name was consistently censored to either 'the firm' or 'the company'. (3) Transcripts and drafts of the report has been shared with the interviewees for feedback prior to its publication. (4) For specific situations in which it has been deemed that data or findings easily could be deducted to originate from specific companies, these findings have been rephrased in an intentionally vague manner to allow the value of the findings to remain clear while censoring the source of the finding.

While the above mentioned actions that have been taken could be improved and accompanied by more initiatives, it has been the intent of the authors to give their best efforts, at all times, to avoid harming, invading the privacy of, or unintentionally deceive any stakeholder involved in the creation of this thesis.

4 Empirical findings

The following chapter includes empirical data collected from the 23 interviews. Some figures distinguish between firms labelled 'industry', and 'digital', as defined previously in the report. The chapter will be structured into four parts. The first three parts contain the findings related to each of the investigated aspect: competencies, organization, and business impact. The last part summarizes these findings.

The findings will more often than not discuss 'historic' and 'future' states, which merit clarification. They refer to a period of five years back or forward in time, or an arbitrary time close to that if the company did not exist five years ago. Furthermore, at times quotes will appear with a description of the person who said it and an indication of what company they represent. The said company may at times be referred to as a small, medium-sized or large. For purposes of context, the following benchmarks apply for these segments:

- Small firms: 1-250 employees
- Medium-sized firms: 251-10,000 employees
- Large firms: More than 10,000 employees

4.1 Competence

This section contains findings related to two areas on the topic of competencies: (1) demand, and (2) supply of competence. In essence, the former topic concerns the 'what' of competencies that firms need, and the latter supply part concerns the 'how' those competencies can be acquired. Each part will present the findings of historic, current and future state of demand and supply to the extent the findings are relevant.

4.1.1 Demand

When addressing historic competence requirements for data science most, if not all, interviews leaned towards whether or not the given firm have had practices in place to actively recruit said competence. Therefore, the empirical findings in this sub-topic most fittingly presents the distribution of firms that did one of three things: (1) actively pursued technical competence (i.e: data scientists or similar), (2) actively pursued both technical competence and managerial competence (i.e: management roles with required data science competencies), or (3) if the said firm did not actively recruit either. Figure 12 below illustrates how firms distributed according to these three tactics.

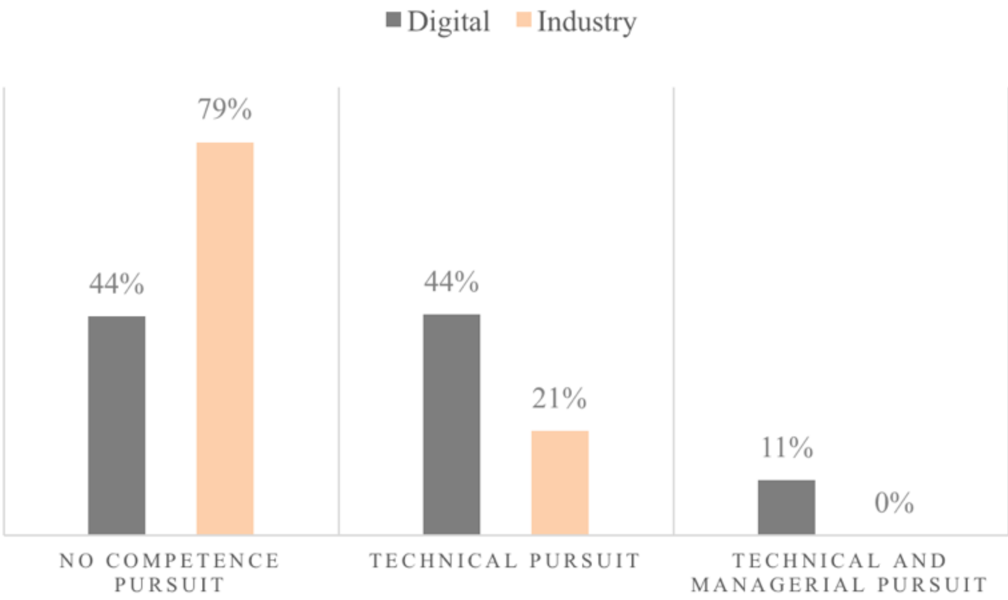

Figure 12: Historic competence pursuit for digital and industry firms.

Since five years have passed, competence requirements have changed dramatically to the extent that firms can discuss desired competencies for management roles connected to data science. The findings primarily underlines the essentiality of five: (1) business knowledge, (2) programming, (3) statistics and math, (4) utility of tech tools, and (5) change management. The interviewees most often underlined some of these, if not all, as critical for people working in management to effectively cooperate with data scientists. Figure 13 below presents the percentages of interviewees that explicitly underlined importance of each competence, distinguishing between digital and industry firms. Admittedly, the level of competence that each interviewee subject referred to was seldom addressed.

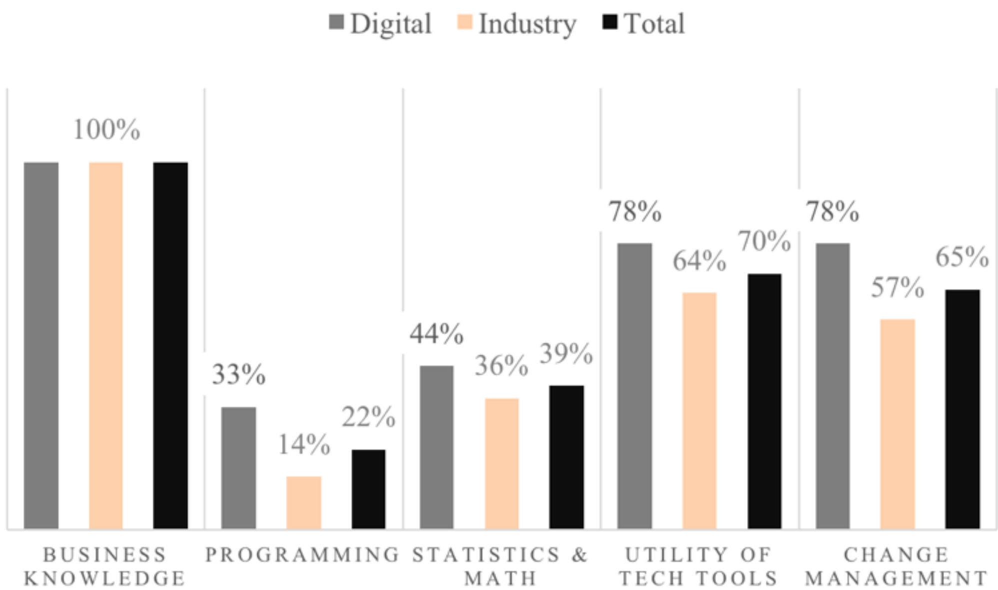

Figure 13: Competencies and their criticality for management.

'Business knowledge' refers to being familiar with how the business works; the customer needs, the properties of the value proposition, and how the individual work connects to the business objective. 'Programming' refers to practical skills in different programming languages, i.e: being able to work with implementations related to the technical data science practice.

'Statistics & Math', not surprisingly, refers to statistics and mathematics. Both did not necessarily have to be mentioned in order to count. 'Utility of tech tools' refers to management people having knowledge of what techniques data scientists utilize to address various questions and problems, and doing what many interviewees refer to as 'speaking the data science language'. Speaking, or trying to speak, the data science language is contextualized by a head of analytics at a large retail industry company: *"I joke that you sometimes hear a senior business leader say: 'we need to put everything in the cloud and get some data scientists to do some algorithms and then plug it into an API and everything will be solved.' They have just strung a sentence together with a bunch of words that they kind of know are important but they don't really understand them."*

'Change management' refers to the capacity of people to enthuse and inspire others to change. The relevance of this competence is explained by a director of analytics at a software engineering company, within the digital segment: *"If we want to become best in the world at using it, we must first become best in the world to implement it. Otherwise, people will just not use it. It does not matter how high quality the data science provides, if people cannot use it, people may perceive it to negatively affect the business and consequently avoid using it."*

When addressing future competence demands, no interviewee could outline a specific new competence or trait that organizations working with data science in the future must possess, apart from the ones already mentioned. Therefore, the findings suggest that new demanded competencies are not expected to arise in the future, and the currently demanded ones will likely remain in demand. However, while the competence profile may remain intact, the number of people that will need to fulfil it is expected to increase. The equivalent role to head of analytics at a medium-sized digital streaming services company reflects: *"More and more parts of people's life will be connected to data, so the number of areas of the world which will benefit from a good understanding of data science will increase."* And similarly, by a business developer at a large transportation company, within the industry segment: *"Competence requirements will likely be the same, but I think that they will grow in volume. We need to become more and better at changing things, more agile."*

4.1.2 Supply

The following section contains empirical findings related to the supply side of competencies. While many firms focus on training their employees in data science technologies, academia was consistently discussed as the foundational source of talent, and thus this section focuses on said institution. There are three segments of findings: (1) descriptions of the current state, (2) what needs to be done, and (3) how it could be done. These follow below.

The current state

The findings suggest that Swedish universities trail their business counterparts when it comes to advancing within the field of data science. A director of analytics of a medium-sized digital company, with long prior experience within the academic community, explicitly voiced the discrepancy in development: *"The sphere of universities has not kept pace to the extent that the industry of Sweden should expect from the universities, given that we are a country that aspires to lead industrially."*

The empirical data of this thesis substantiates the above statement on two observations: (1) the lag in academic scientific progress relative to businesses, and (2) the lack of clear instances to foster data science talent domestically. The first observation relates to representatives of the interviewed companies that have experienced instances of interactions with academia, when scholars systematically have dedicated resources and investment to non-novel research questions. One example of such an instance is given by an innovation manager at a medium-sized, industry-labelled, transportation company: *"When I visited a university and presented what we were working on, one of the lecturers said: 'Oh, our students are currently researching this.' I replied: 'Why? This is already available on GitHub! Just have them download the library and they will be ready to go. You have to do things that have not been done yet.' What do I know about academia, but that's how it works in research, right?"*

The second observation follows from the tendency of Swedish businesses to recruit beyond the borders; firms need skilled business developers that proficiently connects business and technology, and they see potential in academia to contribute in developing this competence domestically, but currently the majority of pursuits are conducted abroad. The perception is that there is no clear school, education or instance of any sort to which employers within the field can go looking for talent.

What to do

The critique from the business representatives regularly came acquainted with an idea of what academia should be aspiring towards, in order to address the present challenges. Three major areas of improvement were identified: (1) facilitate the career path for prospective talent, (2) create an attractive university environment to prolong the retention of talent, and (3) find people who can enthuse change throughout academia. With regard to the student career facilitation process, one idea involved Swedish universities to cooperate and construct guidelines for students who wish to combine analytics and management in their careers. This is explained by a director of analytics in a medium-sized digital firm: *"If all you have is a portfolio of courses, you have no idea what courses to take in order to become what you want. So what you need is a portal where you add in: 'This is all the courses available, and if you want to become a business developer within services, for example, these courses are relevant.'"*

Furthermore, another point of entry for universities aspiring to address the current state may come from increasing the attraction of the university environments, in the perspective of data science prospect talent. This is for the purpose of talent retention; if Swedish leading post-docs and doctorates stay in Sweden, they will be easier to recruit compared to if they do their post-docs abroad. Finally, earlier in the perspective of businesses, acquiring strong managers of change was earlier considered a factor of success, and the findings suggest that the same holds for academia; identifying and encouraging people who see potential for change to advocate the change will be crucial for success.

How to do it

This part outlines a set of potential approaches to the described challenges within academia. It is not by any means an exclusive collection of all possible recommendations, but merely a consolidation of the ideas collected during the data collection.

In order to facilitate processes related to course and career selection, businesses want to contribute as they have as much an interest in academia to develop their capabilities to foster data science talent as academia does itself. Therefore, collaboration between the parties could find mutual appreciation and one interview subject mentioned an example of such an initiative currently ongoing with a Swedish university: the idea is to create a digital portal to serve as an instance of educational consolidation to which other universities can connect, together creating a digital environment orientateable and illuminative enough for students that wish to study attractive fields within data science. However, consolidating the current and existing educations for the purposes of eased logistics will not be enough; new majors also need to be developed. Fellowships and further development careers within business model development connected to data science are explicitly demanded by businesses.

Previously, a willingness to change was mentioned as important. In order to accomplish this, the findings stress the value of illustrating the potential through demonstration, which can be accomplished either by engaging in pilot initiatives, or by drawing upon learning from historic activities. An example of change management through active engagement is the nature of this thesis, as explained by a business angel and industry expert within the field: *"This initiative by you, and by Chalmers, is great. This can lead to ideas formulating and educations being created within the space of data science, so that we can get people directly from school to fill these roles."*

Furthermore, a potential source of change management through drawing upon historic learnings is the one from historic academic success internationally. By systematically studying the accomplishments of academia abroad, and carefully exploring underlying key success factors within the field one is likely to provoke self-reflection, and a reorientation where academia currently is in relation to where it wants to be. A head of analytics of a medium-sized digital firm provides a straight forward-approach on how to lead this change in academia: *"Just show the facts. This is what we are surrounded by and this is how we look. Ask businesses: "What universities do you cooperate with, and why do you do it? What are we missing at home?" I mean, we send people abroad to take business development courses, because there are no universities here that have what we need."*

4.2 Organization

This subsection covers the findings related to organizational demands for firms. It consists of two parts: (1) findings related to organizational structures, and (2) findings related to organizational challenges.

4.2.1 Structures

When investigating how firms organize for data science operations, three potential structures emerged from the findings: centralized, decentralized and hybrid organizations, in which the latter most often relates to having a central function which branches out to company specific functions. Figure 14 below outlines the development of structure among firms, comparing historically to the present status.

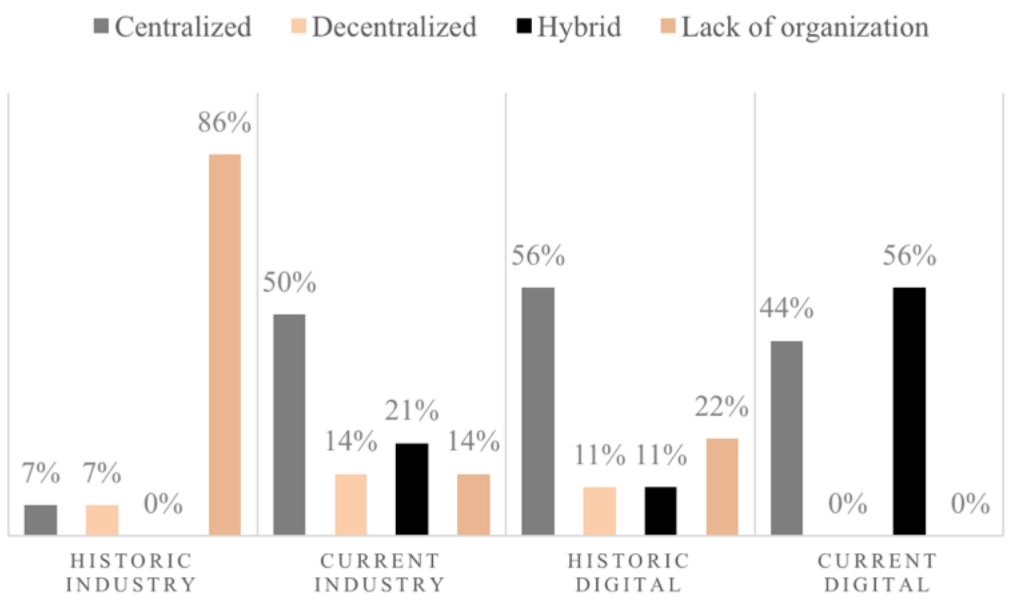

Figure 14: Development of the data science organ's organizational structures , compared between digital and industry firms.

While industry and digital firms organize somewhat differently, the arguments for each structure of organization are consistently similar between representatives of the two segments. Therefore, the below explanation of the empirical data focuses on addressing the arguments of each organizational structure without distinguishing notably between the two segments of firms.

The majority of industry firms have not historically organized around data science operations. Here, admittedly, the five year period do many companies a disservice, as some firms can testify to initiatives for shorter periods back in time. However, the five year period was kept for the sake of consistency. An analytics manager of a medium-sized industry company reflects on the development: *"Before 2 years ago, data science was not even on the map. Then I entered and started 'pounding the drum', trying to convince people about the importance of data."* Or similarly, an IT manager of a large manufacturing company testifies to their development: *"We reorganized for this back in 2016. Before that, data science was not on our minds. We worked with information in a traditional manner; you looked at what you needed for your own purposes, your own processes."*

While some firms interviewed still lack a clear organization for data science, today it is no longer for lack of understanding its value and necessity. Rather, explanations reside in the nature of the business model of the company which presents obstacles and raises difficult questions of corporate politics. Reasons of corporate politics and business models also hold as an explanation for why a minority have opted to organize according to a decentralized structure. This is to say that for some firms it has not been deemed suitable to organize centrally or in a hybrid way, because such a constellation would synchronize badly with the rest of the organizational structure. To contextualize with an example, one interviewed firm has opted for the decentralized structure because said firm is an 'umbrella' organization monitoring and controlling several sub-companies, each in different areas of business. Because of the nature of that organization, decentralized data science was deemed necessary and the other firms with decentralized data science functions generally gave similar accounts when explaining the rationale behind it.

50% of the industry companies describe their data science organization as centralized.

Motivations revolve around two arguments: (1) resource efficiency and (2) the sharing of knowledge and experiences. As an example, resource efficiency through centralization is achieved by not having several decentralized units work on similar tasks. Then, it is argued, the task can be solved centrally and replicated to different domains where it is needed, exemplified by a head of analytics at a large industry company: *"I can see a situation where the center of excellence gets slightly replicated in markets, and I get schizophrenic because I worry that eventually Sweden and China will work on the exact same problem and that is very inefficient!"*

Furthermore, organizing centrally was not uncommonly rationalized as a means to an end; achieving a strong central unit is paramount to subsequently developing a strong hybrid organization. A vast majority considers hybrid organizations to be difficult to construct from scratch, for reasons of challenges related to change management; the organization does not necessarily embrace new people entering their project teams from outside, enforcing new methods, processes and analyses. Consequently, interviewees resonate the idea of initially growing a centralized organization and leverage that to achieve anchoring and understanding continuously. By also allowing such a 'Center of Excellence' to increasingly branch out to functions, it will transform into a hybrid organization over time. The head of analytics of a small but strongly growing digital start-up reflects on their organizations internal structure: *"Since we are so few, we work closely together and attack analytical issues across the entire organization. We collect the needs and prioritize centrally. But this will change as we scale up."* Another interviewee subject, in the role of analytics director at a large industry corporation share the ambitions to reach a state in which competence is distributed: *"The data science unit has not reached enough volume to be distributed. But we continuously recruit more of these competencies so that eventually we can distribute them into functions and have them join cross-functional teams."*

When addressing the future of organizational structure, the unanimous belief is that a hybrid structure that embraces the benefits of both decentralized and centralized work will dominate the field. This partially follows from the highly subscribed rationale behind currently centralized units in Swedish businesses; many are currently centralized because they aspire to become hybrid. It also follows from the findings to expect future dominance of hybrid organizations when analyzing the strengths and weaknesses of the model more holistically; there are positive aspects to both centralized and decentralized units and hybrid organizations can incorporate them both. Centralized units have been argued for previously, while the benefits of decentralized units are mostly attributed to two aspects: decentralization enables the capacity to specialize in closer connection to the function-specific business problems, and it provides higher exposure to the rest of the organization, thus enabling cross-functionality. All these benefits are deemed desirable by a partner at a digital venture capital firm, who discussed the issue holistically: *"It is a highly relevant question whether to centralize or decentralize per function. My strong recommendation is a hybrid method with distributed teams in which a centralized team distributes technically savvy people. Essentially, they will function as a team there while being distributed out to different functions and join different teams. What you will get is the synergies of working together."*. A head of analytics at a large industry firm explains the idea and value of the hybrid structure: *"You feed them [cross-functional teams] competence from the center of excellence, they work in agile teams, and I leave them alone. They work on the problem, deliver results with the team and come back to the center. And the beauty of that is that another entity might then come up with a similar need - okay, send the same person out! She has already learned something here, she has done the model, it is probably mostly transferable. So there's that value and then there's the value of her coming to two parties and saying: 'I solved this problem twice - you guys need to talk more to each other.'"*

4.2.2 Challenges related to data science

When addressing the key success factors for firms when developing it's data science practice, five major challenges emerge from the findings: (1) leadership anchoring, (2) mindset change, (3) change management, (4) process management, and (5) data management. Figure 15 below outlines the percentages of interviews which explicitly underlined these factors as critical factors for success in data science, after which each of these factors are explained in depth.

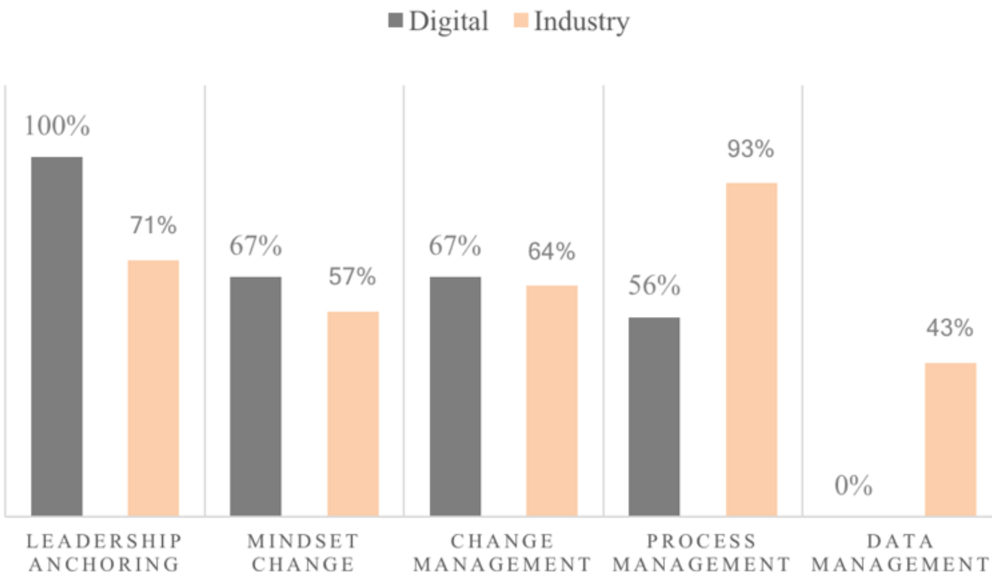

Figure 15: The five major organizational challenges, comparing rate of emphasis between digital and industry firms.

Leadership anchoring

The first factor refers to sponsorship from the board of directors. A large majority underlined this as important, for example one IT manager at a large industry firm attributed a lot of their success to leadership anchoring: *"Having made this shift and this turnaround is because of our excellent CIO and other directors. They are incredibly engaged and truly make sure that this happens."*. The director of analytics of a medium-sized digital firm resonates the value of executive support: *"Before anything, there has to be a demand from upper management to do something. Otherwise, nothing will happen."* At the other end of the spectrum, the head of analytics at a large industry construction company contextualizes the value of leadership anchoring by explaining their setbacks attributed to the lack of support from senior management: *"The biggest obstacle standing in our way right now is senior management. We need their support to invest more, and we don't have that right now."*

The benefits of engaging the leadership are various, but mostly revolve around two areas: (1) contribution from the top to lead change and work with change management, and (2) authority to invest. These follow implicitly from the above quotes, and in general from the empirics. There is also a distinction to be made between convincing the senior management and truly engaging them; interviewees agree that the leadership anchoring exists to the extent that they understand the value and potential of data science, but not to the extent that they act upon it. The explanation for this is that senior management may hold a vast list of necessary investments and initiatives, which creates a situation where investments have to compete against one another. This is a challenge for people advocating data science investments, with the return on investments often relatively hard to determine the process of persuading senior management to prioritize them is consequently complicated.

Mindset change

'Mindset change' refers to the cultural change across an entire organization, to open mindedly refine and adjust historically rooted standards to more fitting practices incorporating data science. The CEO of a small digital retailer summarizes the organizational challenges related to data science: *"I think a changed mindset summarizes well what is the lodestar in this whole process."* In similar fashion the head of analytics of a medium-sized industry firm highlights the importance of an organizational mindset: *"90% is about the right mindset. It is simply the most important thing to have across the organization."* Another head of analytics at a large industry firm communicates the value of a mindset as it contributes to create an attractive working environment: *"As long as we can maintain people within this so that our data pipeline is maintained by the community and so that we have a data science community which is strong and contributes to reinforcing it, then we will find ourselves within an environment which fosters further improvement and that people like working in."*

Change management

'Change management' refers to the competence mentioned in the 'Competence'-section. The difference here is that interview subjects refer to the practice of change management, e.g. when employees of the organization engage people through workshops, presentations, or similar, to visualize the potential of data science and enthuse initiatives related to it. Setting a strong change management movement in play is crucial to overcome the challenges related to corporate culture as an obstacle for change, according to the findings. The innovation manager of a medium-sized industry firm underlined this and continued to exemplify when they gained an audience with senior management: *"Before, our firm was confused when trying to take action. See, they knew they had to take action, but they did not know what actions to take. So I told them: 'No wonder, you lack the insights! And in order to obtain insights, you need to obtain data.' See, the insights they had back then came from individuals who stated what they thought with reference to experience: "I have done this for 20 years. It is this way." I knew that we could not have it that way."*

Furthermore, another typical stage of the organizational development in which change management proves to be critical is when the data science function is developing from being solely centralized to becoming hybrid. As mentioned previously, when the rationale behind temporary centralization was explained, hybrid data science organizations are recommended to grow by continuously anchoring its usefulness across the organizations via a centralized unit that increasingly branches out.

Process management

Continuing, 'Process management' refers to a wide array of operational issues that should be continuously addressed and developed. Such examples of operational issues relate to becoming increasingly effective at working agile, dealing with old infrastructure that hinders flexibility, and enabling technology-oriented people to integrate better with their business-oriented counterparts. A partner of a venture capital company reflects on the need for firms to be more agile in their processes: *"What you learn from the high performers in this field [data science] is that 80 percent of all ideas generate no effect on the end result. That is an intimidating reality for* *most people, but when you face that reality you also understand why you need to be experimental and iterative.".* A data scientist of a large industry firm introduces another element of process management challenges: *"One big challenge for the firm, and other traditional companies, are with old legacy processes and systems.".* Finally, an IT manager of a medium-sized service company within the industry segment underlines the need for process integration between technology and management: *"Our end goal is to have several technical roles to work incredibly tight with the business specialists. In fact, we want to erase existing boundaries between the fields completely."*

Data management

Finally, 'Data management' refers to specific issues related to accessibility of data. An example of such issues is what many interview subjects refer to as 'data silos', meaning that data access is inherently dependent on the function which the employee belongs to. Data from other functions resides elsewhere, locations that the said employee likely does not have access to. This is an issue that traditional firms typically have to deal with, as illustrated by the head of analytics of a large industry firm: *"Because we have not worked too hard to drive data driven decisions, we have not put so much effort into making sure that our data is a hundred percent correct."*

Other challenges

While the five challenges mentioned above were the major ones that systematically were addressed in most, if not all, interviews there were three organizational challenges that were not addressed systematically enough to gain credibility in quantification, but still merits mentioning as a side note. These are (1) firm attraction, (2) unappreciated increased responsibilities, and (3) organizational data science proficiency skewness. Firm attraction represents the concerns that some representatives underlined related to talent acquisition: its success can be limited by the inherent brand value of the company, and how talent perceives the attraction of said work place. The head of analytics at a small digital firm in the growth stage explains the value of peoples perception about the company when recruiting: *"We're a fun company, which is good employer branding. It's fun to work here and make stuff happen."* In contrast, a data scientist of a large industry corporation testifies to the opposite in their case: *"It is hard to get good data scientists into the company, or data scientists at all. Industry companies might have problems to compete with more modern, hip companies."*

In turn, unappreciated increased responsibilities relate to the situation when the organization is heavily dependent on specific individuals, not uncommonly people who are early obtainers of the five crucial competencies outlined above in section 4.2.1. These people create crucial bridges between technology and business, and consequently encounter themselves with much more responsibilities than they expected when taking on their role. They also make the organization extraordinarily vulnerable if these individuals are lost to other employers, as explained by an IT manager of a medium-sized industry firm: *"It has become very individual dependent, for example at one function there is this business manager who is skilled in analytics. She is one of few power-users and she will be incredibly important as we drive data science forward."*

Lastly, organizational data science proficiency skewness relates to the suggestions that different functions of an arbitrary firm may perceive the value of data science differently, and consequently embrace the technologies to various extents. This creates a scenario where functions develop different levels of proficiency to leverage the techniques in their practices, and possibly also gain access to a disproportionate amount of the available resources as they request them more.

4.3 Business impact

Historically, the business impact from data science was either related to describing business outcomes, such as in sales reports, or it was used to diagnose and explain why certain events had occurred. From the empirical findings there is close to no deviation from this general description of the purposes of data as something firms used to 'describe and diagnose'. A business intelligence architect of a large industry firm reflects: *"The area has developed much since the 90s, which was when I started to do these things: collect data to create reports, observe sales statistics and follow-ups which historically has been the most common implementations until as of late."*

Today, the findings testify to varied advancement and levels of value extraction from data science. The testimonies of each firm to where they currently stand have been subjectively coded and subsequently assigned to one of three stages: (1) "data supports reactive decisions", (2) "data supports proactive decisions, or (3) "data automates decisions". These stages have been formulated based on frequency. The result is indicated in figure 16 below. Worth clarifying is that an organization absorbs new capabilities, rather than shift from one capability to another. In essence, this means that a company that for instance has progressed to have data support proactive decisions is likely able to also support reactive decisions from it. Furthermore, the findings suggest two major sources of business value generation when using

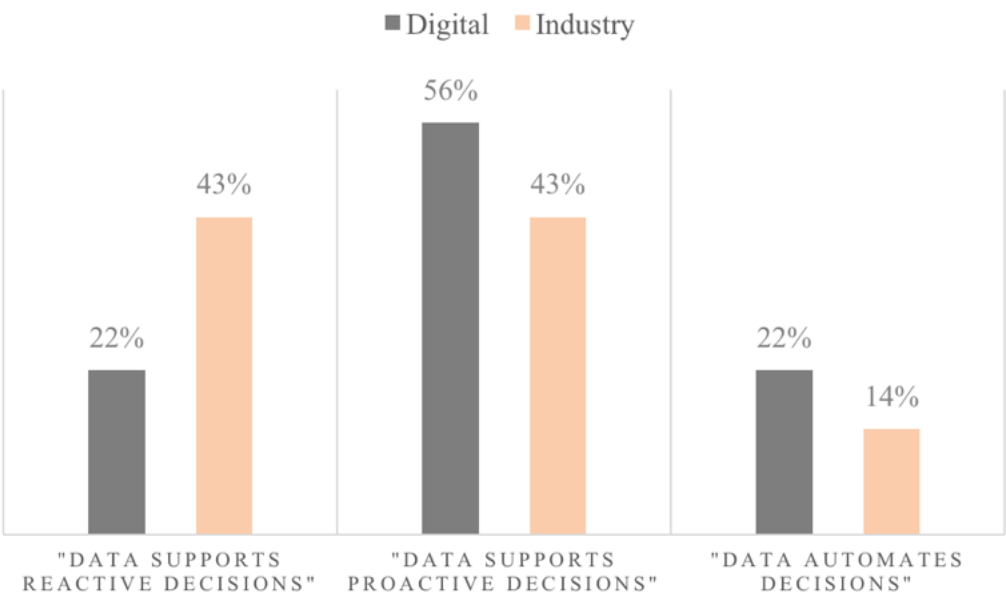

Figure 16: Current levels of business impact of data science in Swedish business.

data science: enhanced customer understanding and operations improvement. The CEO of a small digital firm contextualizes the former source of value: *"The whole point with having data scientists is to understand our customers better. We wish to create connections between why they behave a certain way, and what we can do to support behavior which is positive for us and them."* In turn, operations improvement is explained by a data scientist at a large industry firm: *"Don't waste time on, let's say, simple stuff. This kind of efficiency is what we are after and I think that it applies to a lot of domains. Data science is important in this regard."*

For the future, the findings suggest that two things will happen with regard to the business impact of data science: (1) firms will continue to climb the ladder of capabilities towards higher degrees of automation, and (2) new business opportunities will arise from the data. The former expected outcome specifically means that problems are believed to be solved more efficiently in the future by leveraging the possibility of available techniques. This means that the large advancements may not necessarily be data scientists and management to work on novel and unique problems, but rather in the extent to which machine learning and artificial intelligence automates processes of reporting and analyzing. At the same time, there are also the expectations that the spectrum of problems that data science can solve will become larger, as explained by an analytic insights manager of a medium-sized digital firm: *"Things that we don't use data for now at all we will start to use data for, and within each problem the depth of that data will increase. We will be much more specific, even more than we are today in terms of understanding behaviors."*

Apart from providing increasingly better operational support, data science is also expected to present new business opportunities and value propositions disruptively. The conventional notion of products is that they are either hardware, software or a service. Now, data monetization is on the verge of creating a new space to the extent of disruption and fundamental questions of how to sell, distribute and deliver data-based products will demand answers. The head of analytics at a medium-sized digital firm addresses this with positivity: *"Data monetization, as we call it, is something I think will grow. It is becoming increasingly accepted that if you, as a customer, experience greater value from it you are more willing to hand out your data. This is different from a few years back when customers were generally more sensitive to these things."*

4.4 Summary

This section provides an overview of the findings within the aspects competence, organization, and business impact. Furthermore, it also incorporates the differentiation of time: the findings segment on historic, current, and future state. Table 1 below presents the summary.

Case framework	Competence	Organization	Business impact
Historic	• Minority of industrial firms pursued technical knowledge • Minority of digital pursued technical and managerial competence	• 9/10 companies lacked a clear organization structure • Centralization most common when initiating organization	• Data summarized and diagnosed outcomes
Today	• Emerged *demand* for five competencies: business, programming, statistics & math, utility of tech tools and change management • Lack of a strong competence *supply* from domestic academia	• 2/10 lacks clear structure • Centralization most common for industry firms • Hybrid structure most common for digital firms	• Data summarizes and diagnoses outcomes • Firms increasingly move to data predicting outcomes
Future	• Technology and business roles to be integrated into one • Academia to face necessary change	• Hybrid organizations expected to dominate	• Data to summarize and diagnose outcomes • Data to predict outcomes • Data to cause outcome

Table 1: Overview of the empirical findings.

5 Analysis

The following chapter interprets the empirical findings and analyzes them in relation to the theoretical framework. It follows the same logic as in previous chapters, where the three aspects of data science in organizations are handled in turn.

5.1 Competence

Below follows an analysis on competencies, discussing the demand and supply side as presented in the empirical findings.

5.1.1 Demand

Given the empirical findings, three major point of entries to the demand analysis emerge: (1) the lag in competence pursuit of industrial firms relative to digital firms, (2) the ongoing shift towards requiring the five competencies, and (3) the implications for future competence demands. These will be handled subsequently below.

The lag in pursuit

There is a clear indication that focus has shifted from mainly the pursuit of technical competencies to also including recruitment of managerial data science competence. In this shift, digital firms have been ahead and still are, while industrial firms seem to have started to catch up. This progress discrepancy merits a brief analysis.

There are likely several contributing factors to why digital firms are doing better, and it does not necessarily have to be an isolated competence issue. This becomes clear through a simple line of reasoning: if a firm reorganizes for data science and undergoes the process of addressing the organizational challenges that may arise, it is certainly more likely to have processes in place that enable consistent competence pursuit. Therefore, the competence issue becomes an organizational issue and the observed discrepancy in progress between the two segments of firms can partially be explained to different organizational maturity in this field, which will be touched upon later in the analysis.

Furthermore, one can analyze the difference from a business impact point of view. Traditionally, data has primarily impacted businesses through the use of descriptive analytics, with reports of different kinds as the most common product of it (Seddon et al., 2017). Given that industry firms have longer history than their digital counterparts, it is consequently reasonable to assume that historic sources of value extraction to an extent have 'locked in' industry firms in traditional business intelligence, and all the inertia it brings. Given the existence of this 'lock-in' effect, it provides another source of explaining the lag in competence pursuit between the two segments of firms: if industry firms to a higher extent are stuck in older data science technologies, a consequence might be that they focus more of their recruiting on the competencies of older technologies.

Besides organizational and business impact aspects, it is certainly possible that there are more ways to explain the discrepancy highlighted above. That would be interesting to explore in future research.

The emerged demand for five competencies

One of the major discoveries made in the empirical findings is that of the five emerged skills required of management to possess: (1) business knowledge, (2) programming, (3) statistics & math, (4) utility of tech tools, and (5) change management. Below, the implications of these findings will be discussed, preceded by a brief comment on the small, but consistent, differences between digital and industrial firms in the results.

As was mentioned above, there is a small but clearly consistent difference between digital firms and industrial firms, specifically with regard to the extent to which representatives of each segment underline the criticality of the five competencies. Essentially, a few percentage points more of the digital sample believe that each skill is important for management to possess, relative to the industry sample. The biggest difference seems to be in 'change management' at 21 percentage units, but it is indeed smaller for the rest and in some cases negligible. This tendency warrants two comments: (1) the historic lag in competence pursuit likely drives a significant chunk of this difference, and (2) the differences are small enough to indicate that these competence requirements are relatively generalizable and not dependent on industry or technology segment.

The first comment connects to the discussion in the previous section in a 'cause and effect' line of reasoning. The cause is the historic lag of industry firms in data competence pursuit, the effect it produces is less recruitment experience of said competence and consequently less awareness of what to look for. This does not necessarily have to completely explain the entire discrepancy, but it should certainly contribute. Another factor that likely also plays a role is the human error of the authors of the study; answers might be interpreted wrong, questions not consistently asked the same way or similar. The second comment is an important insight to make for three reasons: (1) any reader, regardless of background and industry in which they operate, will likely benefit from increasing their skill in the five areas of competence, and (2) the academic community can leverage the findings as a framework to produce the needed talent, and (3) there emerges a potential to conceptualize a general approach for businesses of what skills to acquire, and how to do it. Below follows an attempt at doing the latter.

While the five skills are all considered important, the extent to which they are underlined vary. These differences can be considered as proxies of criticality, and thus provide basis for a line of priority among the competencies; essentially saying which competence is more important to have than the other, which is useful if one cannot acquire them all. In figure 17 this line of order is illustrated through the means of a five step process. After the figure follows an explanation of its rationale and the idea behind each step.

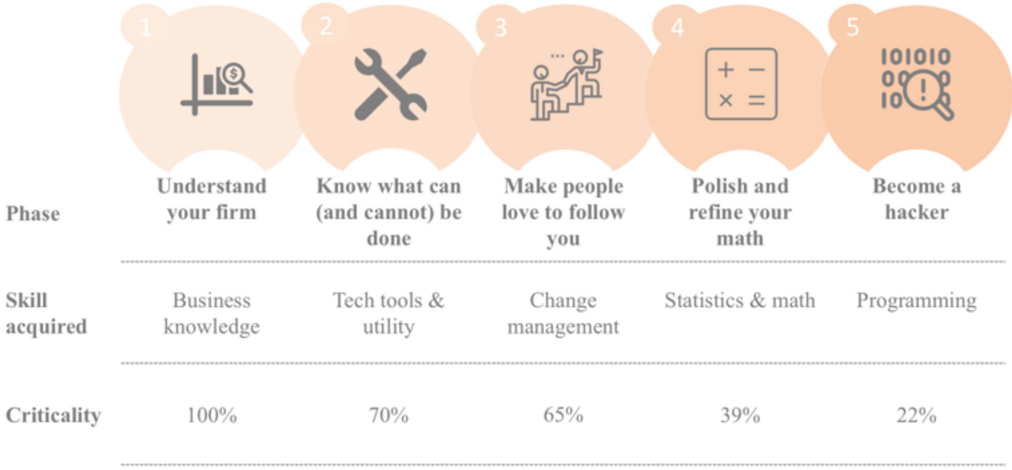

Figure 17: Line of priority on competence acquisition.

Drawing on the idea of data literacy as described by Dichev and Dicheva (2017), any manager or employee must know their business; identifying and translating actual business problems into feasible analysis is key, which is also clearly resonated in the empirical findings. The whole point of data science is to better understand customers, or to improve processes, and so for managers seeking to bridge into the data science environment it is crucial to not lose the business understanding (Schmarzo, 2015). The statement may seem obvious, and to some extent it is, but it is important to underline as a constant reminder that focus can never be lost on what data science actually contributes to the business.

With strong knowledge of the business, the manager can enter the field of science through orientating themselves within the landscape of tools, technologies and their respective utilities. It is clear from the empirical findings that this is important so that the manager can know what can be done, and what cannot be done. The theoretical framework on competence discusses the tendency that students lack insight on the utility of analytics, and that the appropriate remedies are to incorporate more hands-on exercises that allow for practical orientation within the utility of different techniques (Wixom et al., 2014; Kollwitz et al., 2018). These assertions are interesting to draw upon for managers, as the situation is likely similar and the remedies as suitable. Managers need to enter the field and do things to learn things. Not necessarily so that they can do everything that the specialized data scientist does, but so that they 'speak the data science language', as phrased in the empirical findings.

In times of turmoil, the need for change is evident which has implications for managers on the road towards data literacy. Change needs to be managed and it is a competence in itself, as pointed out both in the empirical findings as well as in the studied literature. There will be a return to change management in the organization section, but on behalf of competence it is important to recognize it as such and crucial to practice. In terms of how one does it depends on the given scenario, but an open mindset and willingness to take on responsibilities in leading change should provide a strong source of exposure that allows for continuous practice.

Continuing on to statistics, math, and programming: these competencies can be considered complementary to the previous three on the basis that only a minority underlined them as crucial. However, they are undoubtedly valuable for two reasons: (1) they are mentioned consistently enough to be considered among the five most important competencies, and (2) becoming a data literate manager is contingent on the ability to handle numbers in an analytic context, thus these skills matter (Dichev & Dicheva, 2017).

The last two areas of competencies discussed above are connected to the second one, utility of technology tools, in the sense that they all serve to immerse the manager within the technical nature of data science. They can likely also be acquired through similar practical methods as outlined by Kollwitz et al. (2018). The difference lies in the level of detail and marginal value added from acquiring the competencies; becoming knowledgeable about tools and technologies produces a high level perspective on the field, with high marginal value for the manager seeking to become data literate and start speaking the data science language. It is a Pareto principle of sorts; 80 % of the output is derived from 20% of the input. An eventual refinement and polishing of mathematics and programming allows the manager to increase their data literacy as it goes in more practical depth and rigor, but for the purposes of speaking the data science language the marginal value is not necessarily as high as the first point of entry into the field. Therefore, criticality is lower while the skills still remain highly relevant to obtain and maintain.

These five steps should be viewed as an attempt to generalize the process of competence acquisition on the highest of levels, and consequently prone to be flexible for more practical scenarios. It should not be viewed as an inherently linear process; competencies can likely be acquired simultaneously or in different ways. The purpose of the visualization is to facilitate for the manager who wishes to set up an action plan for the future, and it should be recognized that such an action plan can take many shapes depending on each individuals condition and situation.

A final interesting remark on the competencies is that they neatly span a spectrum from pure business to technology skills. If 'business knowledge' constitutes one extreme point of that spectrum, programming will arguably reside in the other with the other three in between, creating a 'bridge' of sorts between business and technology. This is interesting to connect to the academic coverage about the need of 'Data Translators'; individuals who are able to integrate the business side with the technology side of a firm (Gartner, 2016). The possible connection to be made is to suggest that a data translator, more specifically, arguably is an individual in possession of the five highlighted competencies. It is certainly possible that the spectrum does not depict the profile of a data translator with complete accuracy; some skills may be missing or excessive, but it should certainly contribute to provide clarity in what a data translator actually is. Figure 18 below illustrates these notions.

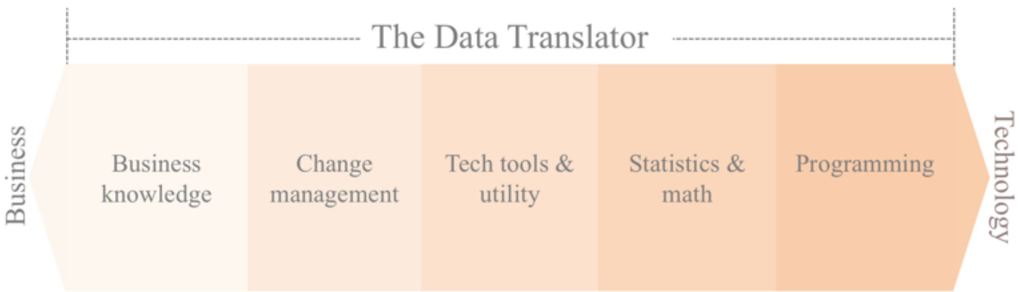

Figure 18: The spectrum properties of the five crucial competencies.

Implications for the future

The empirical findings does not predict any future addition of skills that will be critical for management and data science. This is reasonable; if a competence could be predicted to be demanded in the future, it would indirectly be demanded in present time as well in order to prepare for said future. Therefore, little can be said with regard to an increasingly wider requirement for management people in data science.

However, what merits a brief comment is the projected scale-up in the amount of people that need to be data savvy in the future. The empirical findings suggest that the boundaries between technology and business will break, which indicates an incredible increase in competence demand. The implication is that everyone will need to possess the five skills above, and possibly more depending on what the future holds. Rather than having specific individuals acting as translators between technology and business everyone will need to possess competencies along the spectrum and become data translators, to different extents. Scholars that discuss this agree with the assertion as the belief is that all future work roles will prerequisite data literacy of the individual (Dichev & Dicheva, 2017). In essence, the numbers need to scale up and it will put tremendous pressure on communities to teach the competencies to the required amount of people. This conclusion provides a suitable bridge to analyzing the supply of competence, which is what follows below.

5.1.2 Supply

The rest of the competence section analyzes the findings related to the supply side, exclusively focusing on academia because of its natural role in competence supply. It will focus on the issues holistically, rather than on the specific recommendations outlined by the interviewees in the empirical findings. Those serve for two other purposes: (1) the recommendations underline the eagerness of Swedish businesses to see an improvement in supply, and (2) they are potentially a source of inspiration for future change managers in academia, seeking to take action.

Moving on to what the analysis will discuss, three aspects are analyzed below: (1) the current state of academia, (2) the effect academic supply has on businesses, and (3) generic ideas of what to focus on moving forward.

Current data science talent supply

There are many points of entry to making the case that Swedish academia is at a point in which it is lagging in discipline progress with respect to Swedish businesses. One is simply to recognize that it is not exclusively an issue in Sweden; on a global level academia is generally not keeping pace with industry in research and applications (Jin et al., 2015). The consistency of these relative differences are certainly interesting, and while the pursuit of their explanation is outside the scope of this thesis, they undoubtedly warrant investigation in the future.

Continuing on to a second basis of justification to the assertions made above, the empirical findings strongly support the idea that Swedish academia trails Swedish businesses. On these notions, the empirics communicate confusion, frustration and desperation; confusion because Swedish businesses struggle to identify value in the current data science academia, frustration because they need significantly more than they are currently receiving, and desperation because they find sub-optimal solutions to the issues at hand, for instance by recruiting abroad or conducting in-house training of employees, undermining the value of the academic institution when the teaching there is not enough. In essence, it is fundamentally clear that the Swedish businesses expect and need better sources of supply from academia.

Finally, a third point of entry is that which compares Swedish academia relative to its international counterparts. There is an arguably clear value in making that comparison since academia as a whole trails the global industry, but such a conclusion provides no indication as to how Swedish academia is positioned within the global academic community itself. The empirical findings does not contain sufficient data points to contribute on this comparison, but prior research supports the motion that Swedish academia lags relative to other countries (Hanson et al., 2017). It may or may not be reasonable to expect Sweden to compete with the leading countries China, the US and India, but it should certainly be possible to face off against the leading European countries, which Sweden currently is failing at.

The conclusion is that Swedish academia is not by any means or measure in the forefront of the data science discipline, and could certainly do better.

The effect on businesses

The state of academia affects the businesses because of two reasons: (1) they are an important source for talent acquisition, and (2) their contributions to research development can potentially improve the ways that businesses apply technology in practice. The first reason is rather straightforward and also follows from the empirical observation that universities are considered as the primary competence source. The second reason is indirectly communicated in the empirical findings, as interviewees questioned the purpose of universities if not to contribute to new scientific discoveries.

Prior research and theory also resonate the industrial need of a strong domestic academic organ. A study conducted by the McKinsey Global Institute found that, in the US, the domestic labor market will lack 190 thousand data scientists and 1.5 million 'data savvy' managers as a result from challenges pertaining in the academic supply of talent (McKinsey Global Institute, 2011). Relative to Sweden in terms of population, this translates to 6.3 thousand data scientists and 50 thousand data savvy managers. There should not be an overly high emphasis put on such a comparison, as the difference between Sweden's and the US's respective labor markets are not likely simply a matter of scale, but it should contribute to strengthen the idea that academia needs to foster more competence within this field, and whether they do or not will heavily impact the Swedish business landscape.

Given that the levers of talent and research development determines the effect that academia has on the domestic businesses, and that Swedish academia currently performs sub-optimally in both aspects, the conclusion follows that Swedish businesses are hurting as a consequence.

Moving forward

Data science competence, overall, needs to increase. The role of academia should be to bring data science literacy to the universities, and offer to teach it to any student regardless of background and intended major (Dichev & Dicheva, 2017). While the empirical findings mostly contain specific ideas on what to do rather than overviewing focus points, the spirit of those suggestions is similar to the ideas outlined by scholars and researchers within the data science field. To an extent, the ultimate purpose of creating the suggested platforms for career navigation and setting up new majors within the principle of data science, as desired by the Swedish business community, is to achieve a larger volume of people in Sweden that are able to identify, collect, evaluate, analyze, interpret, present, and protect data - essentially bringing data literacy to more people. Consequently, both empirics and theory suggest improved data literacy as the focal point moving forward.

Given that initiatives are put in place to accomplish a higher level of data science literacy in Swedish universities, three consequences can be speculated as reasonably expected to happen: (1) a channel for influx of data savvy managers will emerge to Swedish businesses, (2) a higher volume of students will consider further majoring in data science, leading to a higher number of data scientists, and (3) a higher volume of graduates will consider an academic career in data science, enabling academia with more resources to invest in further research and development of the principle. These are direct responses and remedies to the present problems, that have been outlined above, and while they all may not occur at all or to the extent expected, it still remains reasonable to conclude that academia possesses the potential to substantially improve its own state as well as that of Swedish businesses, by focusing on increased data literacy in the curriculums.

Needless to say, major change is being advocated in this subsection, and change must be managed. In fact, the necessity and criticality of change management is the overarching success factor throughout this entire chapter and it holds true for academia as well if it is to be able to initiate anything at all (Dichev & Dicheva, 2017). Therefore, the question whose answer most determinedly will predict success here is one of the questions from the empirical findings: Does Swedish academia have the world class change managers required to change its own mindset?

5.2 Organization

While organizations are different and face a vast number of challenges, two common aspects seem to emerge as generally generalizable for organizations under transition towards becoming more data driven:(1) what the desired design of their organization is, and (2) what the road map which leads there looks like. These will be handled in turn below.

5.2.1 The desired design

Firms that have been labelled as industry in this thesis are mostly organized in a centralized fashion, but it does seem that this current state is merely a means to an end; the end goal for a clear majority is to continuously implement a hybrid organization. Certainly, not all organizations share these ambitions, as the empirical findings as well as the theoretical base indicates that it does not lend itself to all types of companies (Davenport, 2013). However, the generalizability of Swedish business to strive for a hybrid organization has arguably gained credibility from the empirical findings, and it consequently becomes an important milestone to reach for organizations, which will become evident later.

One consequence of recognizing the hybrid organization as the end goal for many companies is the assertion of digital firms as the leaders in the process of conversion; a larger portion of them has progressed to that state. Such a conclusion should not be entangled in too much controversy as it is consistent with the conducted analysis on competencies above, identifying the digital firms to be in pole position. In fact, the pattern should be embraced and discussed: if digital firms are ahead, relative to what specifically are they leading? For instance, does the relation between the two segments of firms imply that digital companies have progressed through a process of data science development, and if so: what does that process look like? The empirical findings does seem to suggest the existence of a process, and in the section below follows an attempt to conceptualize it.

5.2.2 The road to get there

The empirical findings presented five major challenges, or key success factors, that often are entailed in a given firms journey toward becoming data driven. The process presented below is an attempt to conceptualize a road map to overcome these five challenges through seven steps, and while all organizations will likely not need to address all of these steps, or even do so linearly, the analysis suggests that they need either to be addressed directly or indirectly. The seven steps are the following: (1) become the active change agent, (2) show people what can be done, (3) draw executive attention and support, (4) grow a central organization, (5) manage operational hurdles, (6) distribute the competence, and (7) do what can be done. Figure 19 below presents an overview to this process, outlining the key success factors of each step, the practical elements of the step, and potential obstacles related to it. In depth explanations follow after the figure.

Phase	Become the active change agent	Show people what can be done	Draw executive attention and support	Grow a central organization	Manage operational hurdles	Distribute the competence	Do what can be done
Success factor(s)	Willingness to act	Technical proficiency and ability to inspire	Corporate authority and as in (2)	Acquiring the right talent	Willingness to refine	Successful cross-collaboration	Organizational possession of the five competencies
In practice	Any agent to discover need for improvement	Workshops and business conversations	Workshop with c-level executives	Continuous growth through anchoring	Data science integrates into processes	The hybrid organization emerges	Initiatives developed in collaboration
Potential obstacle(s)	Insufficient resources for experimentation	Lack of corporate authority	Other high-priority investments	Company brand value	Unappreciated increased responsibilities	Cross-functional segregation	Lost pace with the field development

Figure 19: Proposed process to address the five organizational challenges.

Become the active change agent

The sense of urgency to improve precedes the actions taken (Kotter, 1996). Essentially, change starts with a realization that it is desirable and within the mind of a single frustrated soul at the firm, asking: *"We should be doing this. Why are we not doing it?"*.

Undoubtedly, the change agent is not particularly esteemed or invaluable by its nature if one can simply become it by wanting change; most people at most firms likely always want something new changed at all times. Therefore, step one emphasizes activity - becoming an active change agent implies the willingness and capability to start experiment, to start unravelling the potential value that the active change agent believes to be associated with something, being data science in this particular case. The empirical support for the value of this step became apparent from the stories of many interviewees. Not uncommonly, the subjects had been the active change agents themselves in their respective organizations, drawing back on one example of an innovation manager that 'pounded the drum' for a year before the organization started to embrace the importance of data.

Who is fit to be the active change agent? The findings do not seem to suggest the need of the agent to possess a particular role or level of management, yet since the successful change agent is an active one, access to experimentation resources is important. Therefore, insufficient resources for experimentation presents a possible obstacle in becoming the active change agent.

Show people what can be done

The importance of achieving critical mass when setting out to make a change is supported in theory (Gill, 2002). In the field of data science, the empirical findings have clearly underlined that enthusing people through visualization, or in other words, showing people what can be done is a great technique to gain support and sponsorship from the organization. This particular tactic also finds support in theory, for example by Luecke (2003) who emphasizes the motivating power of actions showing tangible results.

Essentially, the change agent needs to engage their colleagues in creative ways. Countless approaches are likely feasible, yet most seem to revolve around one of two tactics: (1) conducting pilot projects that tweaks some process, product characteristic, or any arbitrary valuable improvement that can engage another individual or individuals, (2) drawing inspiration from external lodestars within the field, leveraging their success onto the own organization. Needless to say, both can be integrated simultaneously, as one tactic does not exclude the feasibility of the other. The empirical findings does exemplify this as interviewees not uncommonly hold organized workshops for the purposes of showing people what can be done, which drew upon internal potential as well as external success.

Success in phase two relies heavily on two factors: (1) data science proficiency and (2) ability to inspire. The first requirement follows partially from arguments of credibility, but is also closely connected to the tactic of pilot projects; if the change agent wishes to conduct pilot projects, moderate technical competence to do so is presumed. The second requirement follows from theory as well as empirics; Tamilarasu (2012) represents the academic community in this regard, stating that if an organization is to truly embrace change the employees must be convinced of its importance. In turn, the 65% of the interviewed subjects who emphasized change management as critical to the organizations development provide empirical validation to the outlined theory.

Potential obstacles related to showing people what can be done is the access to platforms to utilize for these purposes. Gaining the audience is likely contingent on the change agent possessing a moderate corporate rank, so that a chance to present ideas is granted. This analysis does not follow from direct observations or testimonies of the interviewees, but rather from from an intuitive train of thought; it is reasonable to expect that an arbitrary manager will have greater possibilities to speak in front of people compared to the manager's subordinate employees, for example.

Draw executive attention and support

Up until phase three, the idea is that a critical mass of support has been gained from people so that attention of senior management has been drawn, and consequently the active change agent is granted the chance to show the board of directors what can be done. The importance of gaining this audience and making the most of it cannot be understated; both literature and the empirical findings has put large emphasis on the criticality of leadership anchoring from top management. As such, it undoubtedly seems warranted to suggest that achieving it is one of the most fundamental pillars of success in the entire process. Sirkin et al. (2005) suggests similar notions, and the empirical findings shares various stories that attributes much of their success or failure to the respective level of support that they have received from top management. In essence, leadership anchoring is a major gatekeeper on the road towards a data driven organization, as the unanimous conviction is that close to no progress can be achieved without it.

The goal of phase three is not only to enthuse the leaders for support in the same way as the whole organization in phase two, but also to convert that support into investments in setting up an organization for data science. In order to accomplish this, competence within the said field and an ability to inspire remain highly relevant skills for the active change agent to have, while corporate authority becomes increasingly important when drawing executive attention, as it supposedly increases the likelihood of gaining an audience with top management.

It is worth noting that phase two and three are not by any means static in the line of order; persuading the board of directors can certainly proceed persuading the organization. Furthermore, becoming the active change agent could be absorbed into phase three if the agent is a member of top management. The process is flexible and here the lack of required linearity, as well as the rationale behind the current conceptualization, is worth motivating and it follows in the paragraph below.

While the outcome and line of order may vary, each part of the process matter and needs to be addressed one way or another. As it is currently presented, the process seeks to represent the most reasonable 'passage without privileges', which means the following: imagine the average employee at an arbitrary company as the emergent active change agent. This individual is not a member of top management, and they most likely do not possess the corporate authority to immediately present ideas to the board of directors at the moment of idea conception. What the average employee most likely will have to face is a process of continuous sponsorship acquisition, building critical mass to the point when it draws sufficient attention from top management to gain the audience. Undoubtedly, this is not a representation of all scenarios, or even necessarily the majority of scenarios, but it represents the 'passage without privileges', which is argued to most suitably depict a basic process.

Returning back to the topic of gaining executive attention and support, a potential hurdle in progressing from step three is an overwhelmed back-log of c-level initiatives that compete with the investments sought to set up the data science organization. The problem of difficult prioritization is represented in the empirical findings and should be recognized by the active change agent in their attempt to persuade top management.

Grow a central organization

With authorization from management to construct the central organization, focus shifts to gaining critical mass within the newly formed 'center of excellence'. Another option is to invest on full scale immediately, and create the desired size of organization. There are positive and negative aspects to both approaches: for example, the full-scale investment offers speed at the cost of likely higher investment costs, and organically growing the organization offers continuous anchoring at the cost of speed. The conceptualized process advocates organic growth, for purposes of change management and agile methodologies: the field is continuously tested with limited resources, and if success is achieved it can grow. This allows the data science organ to continuously impact the mindset of the whole organization; the growth is connected to increased contribution and if it is clear that it does contribute, the organization is more likely to have a positive mindset (Luecke, 2003).

The goal of phase four is to grow and acquire talent. Therefore, success relies on the organization to continuously prove its worth to the rest of the firm, and to find the right data science and management talent. The desired competencies of management people has been discussed extensively above, and the data scientists places higher emphasis on the technical aspects. A potential obstacle in doing this, inarguably, is the lack of competence supply; the empirical findings as well as available theory testifies clearly to the discrepancy in competence supply and demand. This negatively impacts some firms more than other, as some interviewees testified to company brand value affecting their ability to recruit data science talent. Essentially, the idea is that if the firm does not appear to be an attractive environment for a data scientist to work in, the overall supply of this competence becomes even smaller. It seems to create a negative spiral: the firm that lacks brand value struggles to improve its brand value. However, the cyclicality of that situation is not by any means impossible to disrupt, and firms should embrace the challenge to find creative ways to make it attractive to work with data science within their organizations.

Manage operational hurdles

As the function gains size and exposure to others, it becomes increasingly important to focus on operational efficiency so that it integrates well into the organization as a whole. The empirical findings segments operational efficiency into two parts: (1) process management and (2) data management. The former segment commonly refers to the need to shift towards agile methodologies and iterative development which is supported by the empirics; 93% of industry firms and 56% of digital firms underline it as paramount to success. The relevance to make the shift has also been extensively noted and discussed by the academic community (Schmarzo, 2015; Vidgen et al, 2017).

While it does not contribute to the relevance of step five, a brief analysis on the empirical discrepancy in support between digital and industry firms for process management is warranted: a possible explanation is the that process management is a bigger challenge for industry firms as they need to untangle the organization from legacy systems and deep-rooted cultures to work conservatively. As it was put in the empirical findings, for firms with long history it is generally hard to reform because of organizational culture and deep entanglement in older systems. This likely drives the difference between the two segments of firms, with the general implication being that industry firms will have to spend more time overcoming these issues.

Continuing to unwrap operational efficiency, the empirical findings also touched upon data management and the common issue of siloed function practices leading to siloed data. This is primarily an issue for industry firms as illustrated by the findings, and it is likely so for the same reasons as process management is predominantly a challenge for industry firms. Regardless, these issues clearly exist and requires some organizations to rethink their practices (Schmarzo, 2015).

Phase five does not necessarily require extensive focus from all firms in the process, as it becomes evident that the operational hurdles outlined above are not necessarily issues that every organization has to face. However, the step can present a substantial gateway to progression, and in that case success will rely on capabilities of change management as the organizational mindset will likely need to change. While it is outlined as a separate challenge in the empirical findings, changing mindsets is consistently described as the key success factor in transforming process and data management for the better. The theoretical framework of this thesis also presents a close connection to mindsets and organizational practices, with Schmarzo (2015) as an example who emphasizes the importance of companies rethinking their organizational practices. For these purposes, the previously grown data science function should focus on convincing people to rethink their practices which undoubtedly can present itself as a difficult challenge for some firms with certain operational history. Another potential challenge to be considered is unexpected new responsibilities; with processes reconsidered and restructured, some individuals may prove to become crucial bridging points between functions and competencies, and subsequently create a high dependency on these individuals for success. In the empirical findings, this was exemplified by a head of analytics discussing the value of their 'power-users', as phrased by the subject. Such a development may generally be hard to completely avoid; sometimes there just has to be one lodestar to lead the way. However, specific individual dependence exposes the firm to unwanted vulnerability, and operationally there should be processes and activities in place to mitigate these scenarios however possible.

Distribute the competence

The purpose of the penultimate step is to allow time and space for the prior preparatory work to bear fruit. The whole process of seven phases is revolved around continuously anchoring the value of data science, and as the grown center of excellence gains the preconditions to start delivering value on full scale, distributing the competence is the process in which the hybrid organization emerges, data science is leveraged in cross-functional business problems, and the proficiency of the organization is continuously improved until the point it is fully immersed across the organization. An analogue to the phase is a bird flapping its newly developed wings to the point it gains full flying proficiency; the potential of flight was 'known' prior to the full proficiency state, but the bird required practice and anchoring within its mind and muscle before it could 'do what can be done'.

An interesting example of 'distributing the competence' in play is represented in the empirical findings, and it is warranted to repeat it. It originates from an IT manager in a large industry firm: *"I feed them[cross-functional teams] competence from the center of excellence, they work in agile teams, and I leave them alone. They work on the problem, deliver results with the team and come back to the center. And the beauty of that is that another entity might then come up with a similar need - okay, send the same person out! She has already learned something here, she has done the model, it is probably mostly transferable. So there's that value and then there's the value of her coming to two parties and saying: "I solved this problem twice - you guys need to talk more to each other.""* Here, three things are occurring: (1) competence is being distributed, which seems obvious but merits identification when arguing that the example represents phase six of the process. (2) Continuous improvement is achieved through the data scientist replicating the application of her developed solutions, and (3) the data science function is getting anchored as an intermediary between units that in turn can cooperate better. These are three indications of an organization in change as a consequence of distributing competence, and the third occurrence also resonates with Kotter (1996) notions on how to overcome the challenges related to change management.

In order to become fully immersed as an organization and proceed to the final step, a potential obstacle is skewness in data science practice proficiency across different functions. As it was mentioned as a possible challenge in the empirical findings, it should be recognized as a risk when distributing the competence and mitigated fittingly by each respective firm facing it.

On a final note, depending on the philosophy of the company distributing the competence can be the desired end state; the hybrid organization is finding its place, the operational hurdles have been overcome, and the organization is becoming increasingly more sophisticated at making data informed decisions. Arriving to a state of optimal practice is possibly thus merely a vision rather than possible state of reality. Regardless of the perspective of the reader, defining such a state does seem justified and it follows below.

Do what can be done

The primary distinction between distributing the competence and doing what can be done is that in the ultimate step of the process the organization has developed its data science maturity to the extent that one can consider the competence to be exchanged, rather than distributed one way. Exchanging competence implies that ideas are interchanged in both ways; business and technology becomes one.

Based exclusively on the empirical findings, the analysis suggests a criterion of measuring the extent to which an organization is doing what can be done: address the extent to which organization wide adoption of data translation skills has occurred. By distributing the competence continuously, it is believed that each employee eventually is to gain understanding about (1) the business, (2) conceptual knowledge of programming, (3) the landscape of data science regarding modern technologies and tools, (4) reasonable knowledge in statistics and mathematics relative to the role of the employee, and (5) ability to drive change. When the organization has reached that state, it is effectively doing what can be done. A potential challenge when reaching this state, hypothetically, is to keep up with the pace of field development; the field of data science has seen tremendous development during the past decades and should be expected to develop even further in the future. This puts pressure on the organization to keep up; if one aspires to do what can be done and what can be done develops, then so too must the aspirant align its capabilities.

5.3 Business impact

Relative to the other aspects which map the Swedish business landscape, business impact is undoubtedly the area which reveals the least interesting insights. Overall, it is clear that businesses want to create business value out of data science, and realize how it can be done. Different firms testify to different levels of value extraction, yet this difference cannot be analyzed in isolation as it is clear that the variable is not independent of the others. As an example, if a given organization has progressed far into the development process conceptualized in 5.2.2, they are expected to derive higher business value from data science compared to another entity that has not yet initiated the process.

Having acknowledged the lack of variable independence and thus analysis value, two observations can still be made: (1) Gartner's models can be used to describe the historic, present, and projected future state of business value extraction in Swedish businesses, and (2) with the business impact from data science and technology expected to continuously increase in the future, every firm must consider itself a technology firm and start making up for lost ground.

In chapter 2.1.3 the Gartner model was introduced as a tool to map the analytics landscape. Employing this, the analysis suggests that historically Swedish businesses have developed solid descriptive and diagnostic tools. Today, firms are generally moving into performing predictive analysis, although all firms are certainly not there yet. For the future, Swedish businesses unanimously targets a move into the field of prescriptive analytics. While some firms have developed enough to explore the field, the Swedish business landscape as a whole cannot reasonably be considered to have achieved such a high level of analysis just yet. Firms, industry as well as digital, acknowledge that realizing business value from analytics is a process in itself, with forerunners testifying to having started with the 'simpler' analytics before moving onto solutions of predictive and prescriptive nature. In figure 20 below, the analytics landscape over time is summarized.

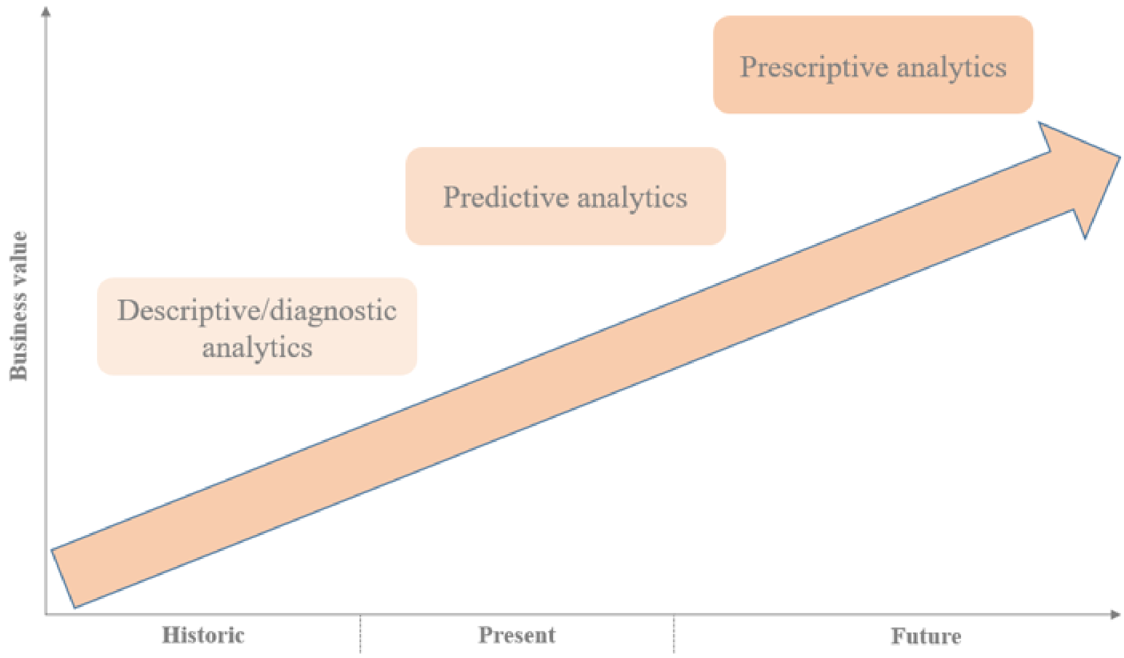

Figure 20: Mapping of the state of data science business impact in Swedish businesses.

Finally, the second observation that businesses must become increasingly oriented around technology is not likely subject to much controversy, but it does serve as fuel to the criticality of overcoming the other challenges outlined in this chapter. The empirical findings suggest a move away from isolated analytics efforts, towards a desired state where analytics is operationalized across the organization. Such a shift towards 'operational analytics' is highlighted by Franks (2014) as necessary for firms looking to truly become data driven. Firms are already acknowledging the tremendous value in data science proficiency, and several challenges have been outlined throughout this chapter. Given the advancement that the analysis of business impact suggests, the criticality and relevance to address these challenges will increase even more. In essence, the analysis of competencies and organization highlights respective challenges, and these will become even greater challenges in the future when the potential business impact of data science increases even more, given the relationships between the aspects.

5.4 Summary

This section summarizes the analysis on the three aspects in table 2 below.

	Competence	Organization	Business impact
Insight #1	• The lag in competence pursuit from industrial firms is likely attributed to historic organizational and business value extraction inertia	• Hybrid data science organizations are the desired goal for most firms, which drives the conclusion that digital firms are in pole position in the process of transition	• Value extraction from data science is highly contingent on success in adhering to competence requirements and addressing organizational challenges
Insight #2	• The varying criticality of the five crucial competencies suggests the following line of skill priority: (1) business knowledge, (2) utility of tech tools, (3) change management, (4) statistics & math, and (5) programming	• Successful approach to the five challenges deemed ultimately decisive in data science progress: (1) leadership anchoring, (2) mindset change, and (3)(4)(5) change, process and data management	• Swedish businesses are currently entering the analytics of prediction after historic residence in descriptive analytics
Insight #3	• The academic community needs to rise to the challenge and teach data literacy holistically	• In order to address the five most common organizational challenges a 7-step process is conceptualized	• Future offers higher business impact and consequently more pressure to act fast

Table 2: Overview of the conducted analysis.

6 Conclusion

The purpose of this thesis was to survey the use of data science in Swedish businesses, and provide an understanding about the general extent to which it has been incorporated in businesses to drive growth and and business value. What has been found are three insights that together span the space of data science use well: (1) data translation is a must-have competence for all, with increased supply from the academic community revolved around five critical competencies as paramount to success. (2) Organizations must have world-class change management to succeed, and the thesis suggests a seven step process to address the challenges related to it. And finally, (3) with incredible and growing potential to drive business impact, data science is demanding increasingly more attention from businesses which elevates the importance of insights (1) and (2) even more as technology and business integrates into one, and fast-paced process progress likely to become a matter of life and death in the future of competition.

Insight number one follows from the work done on competence requirements from both a demand as well as a supply point of view. The analysis eminently indicates that increased competence requirements in data science bypasses no one; everyone must acquaint themselves in what this field of science is about and data literacy is transforming into a prerequisite for entering and staying within the working population. The thesis suggests orientation around five competencies as general areas of emphasis in order to facilitate this and become a better data translator: (1) business knowledge, (2) utility of technology tools, (3) change management, (4) statistics & math, and (5) programming. Acquisitions of these skills rests on supply from two sides: the given company and academia. Business knowledge is likely best gained at said business, and change management proficiency likely presupposes continuous exposure to practical scenarios within industry, but for the rest academia resides on teaching responsibilities that cannot be understated moving forward.

Insight number two is related to the gathered empirics on organizational demands and challenges. Many challenges have been illuminated, but the major ones are the following five: (1) leadership anchoring, (2) mindset change, (3) change management, (4) process management, and (5) data management. While they are not necessarily mutually exclusive in their nature, they have been consistently regarded as critical challenges to overcome and the proposed seven step process seeks to address them in a sequential, but flexible way. The steps are the following: (1) become the active change agent, which is when an arbitrary individual identifies the need for change with willingness to act upon it. (2) Show people what can be done, which is when the active change agent seeks support and subscribers to their ideas, building critical mass in support to proceed to phase three. (3) Draw executive attention and support, which involves same actions as in the previous step but specifically directed to the board of directors, seeking both their engagement and financial sponsorship to invest in an organization. (4) Grow a central organization, which is essentially the seed of what eventually is to organically grow into a hybrid organization. This process of organic growth is important to change mindsets about the practices and build anchoring throughout the organization. (5) Manage the operational hurdles, effectively serving as the 'housekeeping' phase of the process when issues of data and process management are dealt with. (6) Distribute the competence, which is when the hybrid organization continuously emerges as competence from the central center of excellence is distributed into cross-functional teams. This model is suggested to best leverage and develop the data science competence. Finally, (7) Do what can be done, which is the state in which the entire organization embraces the practice, possesses the mindset and five crucial competencies, and continuously leverages data science to make data informed decisions.

Lastly, insight number three concerns the investigations on business impact. The verdict is that Swedish businesses master descriptive analytics and are moving into the predictive field, but need to increase the pace. This is because of the identified gap between potential and actual value extracted from the practice, a gap which is doomed to increase unless considerable action is taken as the potential of the field increases. Data science is simply becoming an increasingly stronger force to be reckoned with, and impossible to ignore for any organization within any industry that wishes to survive. Technology and business roles are to merge into one, and no one is exempt from this reality. Consequently, the development is expected to put an increasingly high pressure on the competence demands and organizational challenges; the number of people and organization that must adhere to them will continue to increase, making it even more crucial to act than it ever seemed before.

In summary, the goal of mapping the Swedish business landscape in data science has been addressed by investigating three factors: competence requirements, organizational demands, and business impact. This thesis draws upon the insights within each aspect to conceptualize frameworks and processes representing what businesses agree is the desirable end goals within each aspect. This is subsequently encouraged to be used by the individual or enterprise in their respective processes of development, providing clarity in what path to take as well as in relative progression compared to peers.

Note on sustainability

As has been mentioned, data science can fulfill two purposes: (1) increased customer understanding, and (2) increased operational efficiency. This is to say that optimally customers get better products and businesses consume less resources in the process of producing said products. Consequently, the field clearly has a role in the pursuit of a more sustainable society and if this thesis can result in more firms advancing their practices within data science, it should undoubtedly be beneficial from an environmental point of view.

Call for future research

While more feasible opportunities likely exist, the authors would suggest four areas that would benefit from future research.

Firstly, this thesis approaches businesses on the most holistic of levels. While it has proved useful to the extent that several generalizable patterns and factors have emerged, a more niche investigation on specific sectors of businesses would provide more nuanced and concrete insights for representatives of said industries. Therefore, the first call for future research is made upon the various possible ventures into industry specific data science investigations.

Secondly, there has been an identified need of competence supply from academia. Theory and empirics clearly resonate this importance. However, there lacks a clear action plan of what the change leaders of the academic community should take in order to do as effectively as possible create this necessary change. Therefore, a second call for research concerns a more detailed investigations on academia, what type of individuals that are fit to be change leaders in academia, how they can identify what areas that require the most attention, and what they can do to turn the tides.

Thirdly, the thesis has stumbled upon the consistent relative difference between academia and business in data science development, in which business generally leads academia. This consistency has been noted as interesting, and undoubtedly warrants investigation in the future.

Finally, there has been an identified difference between industry and digital firms in competence pursuit which illuminated the tendency of industry firms to generally recruit competence connected to less modern technologies. A few sources of explanations have been provided in the analysis in the thesis, but it is certainly possible that more exist. Therefore, the authors of this study would like to dedicate a final call for future research dedicated to unravelling the factors that drive competence pursuit.

References

[1] Akerkar, R. (2013). Advanced data analytics for business. Big data computing, 373-400.

[2] Alvesson, M. (2003). Beyond Neopositivists, Romantics, and Localists: A Reflexive Approach to Interviews in Organizational Research. The Academy of Management Review, 28(1), 13.

[3] Arora, D., Malik, P. (2015). Analytics: Key to go from generating big data to deriving business value. Paper presented at the 446-452. doi:10.1109/BigDataService.2015.62

[4] Asadi Someh, I., Wixom, B., Davern, M., & Shanks, G. (2017). Enablers and Mechanisms: Practices for Achieving Synergy with Business Analytics.

[5] Bryman, A. and Bell, E. (2015). Business research methods. 4th ed. New York: Oxford University Press.

[6] Cao, L. (2017). Data science: a comprehensive overview. ACM Computing Surveys (CSUR), 50(3), 43

[7] Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2015). The promise and perils of wearable sensors in organizational research. Organizational Research Methods. Published online ahead of print. doi: 10.1177/1094428115617004.

[8] Chen, H., Chiang, R. H., & Storey, V. C. (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. MIS Quarterly, 36(4), 1165-1188.

[9] Cole, F. L. (1988). Content analysis: Process and application. Clinical Nurse Specialist, 2(1), 53-57.

[10] Connelly, L. M. (2016). Trustworthiness in qualitative research. Medsurg Nursing, 25(6), 435-436.

[11] Dallemule, L., Davenport, T. H. (2017). What's Your Data Strategy?. Harvard Business Review, 95(3), 112-121.

[12] Davenport, T. (2014). Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business review Press.

[13] Davenport, T. (2013). "Organizing Analytics and Big Data," International Institute for Analytics, September 23, 2013, at https://iianalytics.com/research/organizing-analytics-and-big-data

[14] Delgado, J. C. M. (2016). An Interoperability Framework and Distributed Platform for Fast Data Applications. In Data Science and Big Data Computing (pp. 3-39). Springer, Cham.

[15] Dhar, V. (2013). Data science and prediction. Communications of the ACM, 56(12), 64-73.

[16] Dichev, C., & Dicheva, D. (2017). Towards Data Science Literacy. Procedia Computer Science, 108, 2151-2160.

[17] Donoho, D. (2017). 50 Years of Data Science, Journal of Computational and Graphical Statistics, 26:4, 745-766, DOI: 10.1080/10618600.2017.1384734

[18] EDC. (2016). Building Global Interest in Data Literacy. Education Development Center, Inc. Retrieved from http://oceansofdata.org/sites/oceansofdata.org/files/ODI

[19] Eisenhardt, K. M. (1989). Building Theories from Case Study Research. Academy of Management Review, 14(4), 532-550.

[20] Elo, S., & Kyngäs, H. (2008) The qualitative content analysis process. Journal of Advanced Nursing 62(1), 107–115 doi: 10.1111/j.1365-2648.2007.04569.x

[21] Franklin, M.(2014). Agile change management: A practical framework for successful change planning and implementation (1st ed.). London;Philadelphia;: Kogan Page.

[22] Franks, B. (2014). The analytics revolution: How to improve your business by making analytics operational in the big data era (1st ed.). Hoboken, New Jersey: Wiley.

[23] Gartner. (2016). Why You Need to Rethink Your Data and Analytics Roles Now. Retrieved from https://www.gartner.com/doc/3136418/itscore-overview-bi-analytics

[24] Gartner. (2015). Advancing Business With Advanced Analytics. Retrieved from https://www.gartner.com/doc/3090420?srcId=1-6470978268

[25] Gartner. (2014). Gartner Says Advanced Analytics Is a Top Business Priority. Retrieved from https://www.gartner.com/newsroom/id/2881218

[26] Gartner IT Glossary. (2013). Definition of business intelligence.. Retrieved from:http://www.gartner.com/it-glossary/business-intelligence-bi/ (accessed on 10.04.18.)

[27] George, G., Osinga, E., Lavie, D., & Scott, B. (2016). Big data and data science methods for management research. Academy of Management Journal, 59(5), 1493-1507. doi:10.5465/amj.2016.4005

[28] Gil, Y. (2014). Teaching parallelism without programming: a data science curriculum for non-CS students. In Proceedings of the Workshop on Education for High-Performance Computing (pp. 42-48). IEEE Press.

[29] Gill, R (2002) Change management–or change leadership?, Journal of Change Management, 3:4, 307-318, DOI: 10.1080/714023845

[30] Goertzel, B., Mossbridge, J., Monroe, E., Hanson, D., & Yu, G. (2017). Humanoid Robots as Agents of Human Consciousness Expansion. Cornell University Library.

[31] Greenhalgh, T., & Taylor, R. (1997). How to Read a Paper: Papers That Go beyond Numbers (Qualitative Research). BMJ: British Medical Journal, 315(7110), 740-743. Retrieved from http://www.jstor.org.proxy.lib.chalmers.se/stable/25175748

[32] Grossman, R. L., Siegel, K. P. (2014). Organizational models for big data and analytics. Journal of Organization Design, 3(1), 20-25. doi:10.7146/jod.9799

[33] Gupta, B. (2016). Interview questions in business analytics. Bangalorei, Karnataka, India: Bhasker Gupta/ Apress. doi:10.1007/978-1-4842-0599-0

[34] Hanson, M., Jeppsson, T., & Nordlund, M. (2017). Strategisk analys av forsknings- och innovationslandskapet inom Artificiell Intelligens (AI): Hur ligger Sverige till och finns vi med i utvecklingen?

[35] Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Sethupathy, G. (2016). THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD. McKinsey Global Institute.

[36] Hürtgen, H., & Mohr, N. (2018). Achieving business impact with data. McKinsey Company. Retrieved 14 April 2018, from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/achieving-business-impact-with-data

[37] Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. Big Data Research, 2(2), 59-64.

[38] Kollwitz, C., Dinter, B., & Krawatzeck, R. (2018). Tools for Academic Business Intelligence and Analytics Teaching: Results of an Evaluation. In Analytics and Data Science (pp. 227-250). Springer, Cham.

[39] Kotter, J. P. (1996). Why transformation efforts fail. The Journal of Product Innovation Management, 2(13), 170.

[40] Krishnaswamy, O. R., Satyaprasad, B. G., Ebook Central (e-book collection), & ebrary, I. (2010). Business research methods (1st ed.). Mumbai [India]: Himalaya Pub. House.

[41] Lakshman, M., Sinha, L., Biswas, M., Charles, M., & Arora, N.K. (2000). Quantitative vs Qualitative Research Methods. The Indian Journal of Pediatrics, 5(67), 369-377. doi:10.1007/BF02820690

[42] Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management, 36(5), 700-710. doi:10.1016/j.ijinfomgt.2016.04.013

[43] Luecke, R. (2003). Managing change and transition (Vol. 3). Harvard Business Press.

[44] Mahmood, Z., SpringerLink (Online service), & SpringerLink (e-book collection). (2016). Data science and big data computing: Frameworks and methodologies. Cham: Springer International Publishing. doi:10.1007/978-3-319-31861-5

[45] Maxwell, J., & Chmiel, M. (2005). Qualitative Research Design. Oxford Bibliographies Online Datasets.

[46] McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard business review, 90(10), 60-68.

[47] McKinsey Global Institute. (2011). Big data: the next frontier for innovation, competition and productivity. McKinsey Global Institute.

[48] Mento, Jones & Dirndorfer (2010) A change management process: Grounded in both theory and practice, Journal of Change Management, 3:1, 45-59, DOI: 10.1080/714042520

[49] Meulen, R. (2018). 5 Ways Data Science and Machine Learning Impact Business. Smarter With Gartner. Retrieved 15 April 2018, from https://www.gartner.com/smarterwithgartner/5-ways-data-science-and-machine-learning-impact-business/

[50] Morabito, V., SpringerLink (Online service), SpringerLink (e-book collection). (2015). Big data and analytics: Strategic and organizational impacts (2015th ed.). Cham: Springer International Publishing. doi:10.1007/978-3-319-10665-6

[51] Nigatu Haregu, T. (2012). Qualitative data analysis. [video] Available at: https://www.slideshare.net/tilahunigatu/qualitative-data-analysis-11895136 [Accessed 17 Jan. 2018].

[52] Robinson, O.C. (2013) Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide, Qualitative Research in Psychology, 11:1, 25-41, DOI: 10.1080/14780887.2013.801543

[53] Polit, D.F., & Beck, C.T. (2014). Essentials of nursing research: Appraising evidence for nursing practice (8th ed.). Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins.

[54] Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big data, 1(1), 51-59

[55] Ransbotham, S., Kiron, D., & Prentice, P. K. (2015). Minding the analytics gap. MIT Sloan Management Review, 56(3), 63.

[56] Rosenbaum, D., More, E., & Steane, P. (2018). Planned organisational change management: Forward to the past? an exploratory literature review. Journal of Organizational Change Management, 31(2), 286.

[57] Todnem By, R. (2005) Organisational change management: A critical review, Journal of Change Management, 5:4, 369-380, DOI: 10.1080/14697010500359250

[58] Schaeffer, E. (2017). Industry X.0: Realizing digital value in industrial sectors. Munchen: Redline Verlag.

[59] Schiller, S., Goul, M., Iyer, L., Sharda, R., & Schrader, D. (2014). Build your dream (not just big) analytics program.

[60] Schmarzo, B. (2015). Big data mba : driving business strategies with data science. Retrieved from https://ebookcentral.proquest.com

[61] Schuff, D. (2018). Data science for all: a university-wide course in data literacy. In Analytics and Data Science (pp. 281-297). Springer, Cham.

[62] Schutt, R., & O'Neil, C. (2013). Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc.".

[63] Seddon, P. B., Constantinidis, D., Tamm, T., & Dod, H. (2017). How does business analytics contribute to business value?. Information Systems Journal, 27(3), 237-269.

[64] Sirkin, H. L., Keenan, P., & Jackson, A. (2005). The hard side of change management. Harvard business review, 83(10), 108.

[65] Sreejesh, S., Mohapatra, S., Anusree, M. R. (2014). Business research methods: An applied orientation. Springer.

[66] Tamilarasu, V. (2012). Change management. International Journal of Management Prudence, 4(2), 26.

[67] Van der Aalst, W. M. (2014). Data scientist: The engineer of the future. In Enterprise interoperability VI (pp. 13-26). Springer, Cham.

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

[69] Wixom, B., Ariyachandra, T., Douglas, D. E., Goul, M., Gupta, B., Iyer, L. S., & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. CAIS, 34, 1.

[70] Woodside, A. G. (2010). Case study research: Theory, methods and practice. Emerald Group Publishing.

Appendices

A - Questionnaire

Introduction

2. Tell us a little about yourself and your role at the company.
   (a) How does your role relate to data science?
   (b) What competencies are demanded for your role?
3. How is data science defined in the industry in which your company operates?
   (a) What areas of the business do you think potentially could be improved with data science?
   (b) What effect could implementation of data science in these practices have?
4. How developed do you consider the industry currently to be, with respect to adoption of data science practices?
   (a) Given maximal utilization, where do you think the industry as a whole could be?
   (b) What do you think causes the discrepancy between current situation and the optimal extent to which data science could be adopted in your industry?
   (c) What differences do you see between now and five years ago, with regards to adoption of data science?
   (d) What explains this difference?

Competencies

6. Competence requirements
   (a) From a data scientists point of view, what competencies are important to have?
   (b) From a management/executive point of view, what competencies are important to have?
   (c) How does this compare to five years ago? What will change in the next five?
7. Acquiring competence
   (a) How are the required competencies acquired?
   (b) What challenges do you see related to competence acquisition?
   (c) What do you do to overcome these challenges?
8. Does the progress of data science affect the demanded competence of newly recruited graduates of the Industrial Engineering and Management programme?
   (a) If so, how should his changed demand be handled by universities?

Organization

10. What demands are put on organizational infrastructure when looking to adopt data science?

(a) How do you structure in order to make data accessible and possible to act on?

(b) How does this compare to five years ago? What will change in the next five?

11. Where do data science initiatives originate within your organization?

    (a) How do you enable company-wide initiatives, spanning multiple parts of the organization?

    (b) How does this compare to five years ago? What will change in the next five?

12. How are employees involved and encouraged to take initiatives related to data science?

    (a) How do you ensure that employee initiatives drive value for the business as a whole?

Business impact

14. How is data science connected to your business strategy?

    (a) How does this compare to five years ago? What will change in the next five?

15. Can you conceptually describe an example of when data science was used to successfully drive substantial business impact?

    (a) What was the key success factors for this example to succeed? Competencies? Infrastructure? Collaboration?

16. How has the opportunity to derive business value from data science evolved over time?

    (a) How do you think it will evolve in the next five years?

B - Guidelines for IEM

The authors of this study deem it highly relevant for developers of IEM to evaluate and integrate data science into the program. The CEO of a small digital firm agrees: *"If IEM does not integrate and teach this type of knowledge, it will not be an education for the future"*. Consequently, an overview of suggested guidelines are outlined below, which the targeted reader is encouraged to incorporate for future considerations. It should be read as a discussion of speculative nature as limited empirical data has been gathered on the topic; these are merely the thoughts of two people who recently have both graduated from the IEM program, as well as acquired relatively strong knowledge about the state of data science use in Sweden. The outline will touch on two topics: (1) what should be taught, and (2) how it could be implemented.

What should be taught: In essence, IEM should strive to supply courses that teach students the five important competencies outlined in the thesis: (1) business knowledge, (2) change management, (3) utility of technology tools, (4) statistics & mathematics, and (5) programming. The first and second skills are already well trained, but the other three more technologically oriented ones are simply not taught enough or at all. These notions are resonated by a partner of a venture capital firm, with a background in the IEM program: *The program should definitely include courses within every part. There's tons of business courses already, but the analytical is also important where that to some extent entails pure mathematics, but it should be complemented with more modern methods of conducting analysis as well."*

IEM simply needs to equip aspiring students with knowledge of modern analysis. Such knowledge could be broken into two areas: (1) Familiarity about what tools, frameworks and techniques that exist for different analytical purposes and knowledge of how those analyses are conducted. Examples of relevant topics to cover are given by a head of analytics at a small digital firm: *"You most definitely should have at least 5 credits of A/B testing, perhaps another 5 credits on growth analysis. And practice these things on real cases with real data."* (2) Becoming knowledgeable about the work methodologies of analytics, in order to effectively lead teams within it. The head of analytics at a medium-sized digital firm contextualizes: *The fact that you don't clearly emphasize agile work methodologies in courses related to leading innovation [referring to the MEI program], well that tells me that someone does not have their head in the game."*. In essence, the vision should be to develop students that are able to conceptually understand and assure the quality of a data science solution, but not necessarily create said solution themselves.

How it could be implemented: While it would be a radical change, the authors view removing the four technical bases in favor of four courses in data science as the change with highest marginal value added to the program in general. This is motivated by two reasons: (1) data science related topics would bring immense and crucial value to the program, and (2) the four technical bases bring limited value. The second reason is undoubtedly subjective, yet the authors perceive that obtaining empirical support by asking I-alumni about their perceptions of the technical bases could strengthen that claim.

The advocated change would undoubtedly come at considerable costs. The authors are likely not aware of them all, but primarily identify the one of lost flexibility for graduating bachelor's students selecting a master's; each technical base enables entry into other programs than those of the Industrial Engineering faculty. This would evidently be lost should the technical bases be removed. While the marginal value of adding data science courses is deemed higher than the marginal value lost from such flexibility, it should still be considered and possible means of mitigating such a loss evaluated.