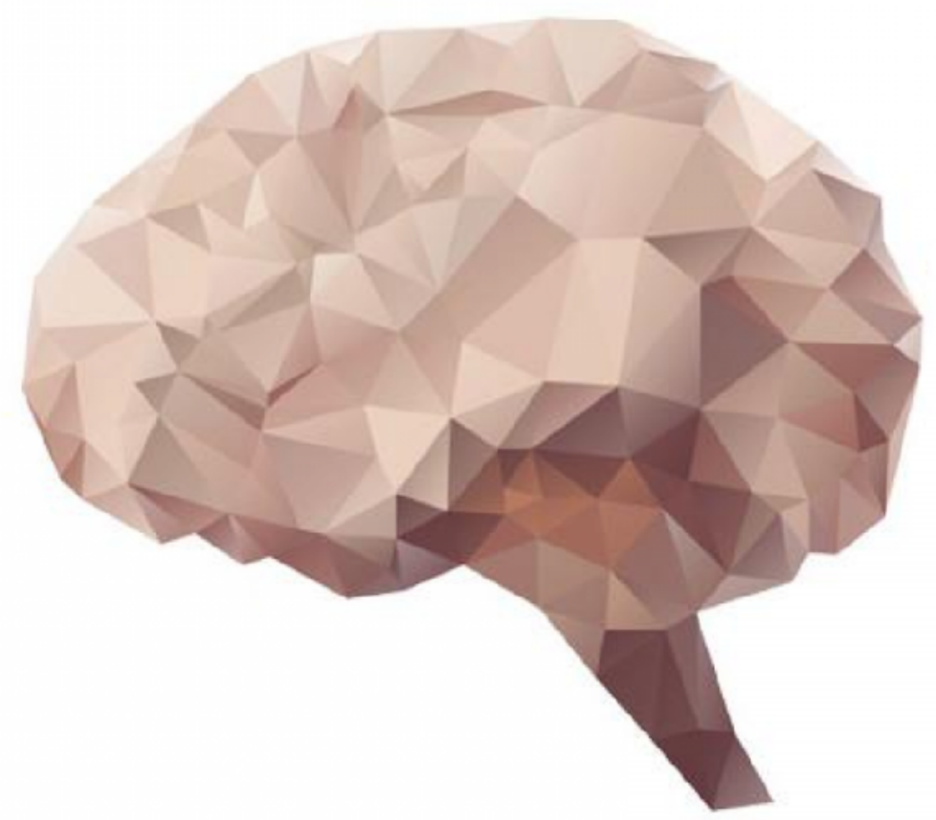


An Inquiry into the Nature and Causes of the Success of Data & Analytics in Organizations

Jonas Vandenbruaene

Academic year 2016-2017

University of Antwerp

Faculty of Applied Economics

An Inquiry into the Nature and Causes of the Success of Data & Analytics in Organizations

Jonas Vandenbruaene

Master's thesis submitted to obtain the degree of:

Master in applied economics: Business Engineer

Promoter: Prof. dr. Koen Vandenbempt
Company mentor: Joeri Arts (SAS)

Abstract

The implementation of analytics in business processes has received significant attention over the past years (Baesens, 2014; McKinsey, 2016; Provost & Fawcett, 2013). In this thesis we conduct an as-is analysis of the maturity of organizations with respect to data & analytics based on a survey. We use the DELTA model (Davenport, Harris, & Morison 2010), an often-mentioned maturity model in the data & analytics domain, as the conceptual framework. In the first part of the analysis we provide descriptive full-sample results. We find that more than 70% of respondents report data & analytics are very or extremely important to their organization. Furthermore, companies have many plans for the implementation of data & analytics. However, only 52% of those ambitions are already translated into noticeable impact. The main implementation challenges are related to organizational and human characteristics (and not to technology).

In the second part of this thesis we empirically test the stages of the DELTA model and try to obtain insight in the reasons behind maturity differences between companies. We do this by clustering the responding companies into five maturity clusters by using the k-means unsupervised segmentation algorithm. In this way we can compare the theoretical description of the maturity stages from the DELTA model with the results from the clustering. We find that an incremental increase in maturity across the maturity dimensions, as assumed in many maturity models, is not a perfect description of reality. Moreover, we observe that the relationship between ambitions and impact is non-linear. By using the Gartner Hype cycle model this behavior can be better understood. This is an important conclusion for the maturity model literature in which many models are not proficiently empirically tested.

We find that 19% of companies are “analytically impaired” meaning they lack fundamental requirements for the success of data & analytics within their organization. Almost a third of the companies in the sample are defined as “localized analytics”: they have some data & analytics projects running, but do not organize them at an enterprise level. Next, 17% of companies demonstrate “analytical hubris”. They have extreme data & analytics ambitions but faced with many challenges, they do not succeed at implementing them. Furthermore 23% of the companies in the sample are “analytical companies”. They are relatively mature and are able to translate analytics into value in a consistent way. Lastly, 9% of the companies in the sample are defined as “analytical competitors”. These companies are the data & analytics hotshots: they are very mature, innovative and data & analytics are of strategic importance to them.

Executive summary

Introduction

The implementation of analytics in business processes has received significant attention over the past years (Baesens, 2014; McKinsey, 2016; Provost & Fawcett, 2013). As the benefits of data-driven decision making are becoming increasingly clear (Bain, 2013; Brynjolfsson, Hitt, & Kim, 2011) more and more companies are developing strategies to leverage their data assets to improve business outcomes (Gartner, 2016). The confidence of executives in the power of analytics is high as 85% believes "big data" will "dramatically change the way they do business" (Accenture, 2014).

Problem statement

The technological side of data & analytics has been discussed thoroughly in the literature (Inmon, 2016). Companies can buy turnkey solutions from a myriad of software vendors to govern their data or to perform advanced analytics. However, the organizational side of analytics has received significantly less attention (Mazzei & Noble, 2017). What is the current state of data & analytics maturity in organizations? How can companies configure themselves to optimally reap the benefits of data & analytics? Do companies need a C-level executive to fully absorb the data & analytics potential? In what business activities do organizations need to implement data & analytics solutions? What are the main causes of failure of data & analytics initiatives and what are the most common solutions? These organizational issues are the key questions of this master's thesis.

Methodology

Central to the investigation of these topics is the concept of maturity modeling. Maturity models are often used in the management literature to map the capability of organizations in certain domains (Fraser, Moultrie, & Gregory, 2002; Mettler, & Rohner, 2009; Mettler, 2011). In this thesis the DELTA model (**D**ata, **E**nterprise, **L**eadership, **T**argets, **A**nalysts) developed by Davenport, Harris, & Morison (2010) is used as the theoretical backbone. The DELTA model was used to compose a data & analytics maturity survey, which was distributed to 2000 business executives via email. Because of a low participation rate and a high attrition rate only 90 surveys were useful. The resulting sample was used in two ways. First a general, full sample analysis was conducted. Next, the responding companies were clustered into five maturity clusters based on the survey data. This clustering was obtained via the k-means unsupervised segmentation algorithm. The result from this clustering was used to test the insights of the DELTA model empirically and to generate cluster-specific insights.

Results and discussion

Full sample descriptive statistics

Targets[1]

More than 70% of respondents state data and analytics are very or extremely important to their organization. Organizations have a wide range of outcomes they plan to achieve via the implementation of data & analytics in their processes. The three most mentioned ambitions are: increasing revenue, improving decision making and better understanding the customer. The most mentioned business units in which companies plan to achieve these ambitions are marketing, sales and finance. Very few companies succeed at implementing all their plans. In total only 52% of desired data & analytics outcomes are already translated into noticeable business outcomes. The most mentioned challenges are: lack of skill, organizational structure and organizational culture. Organizations plan to overcome these issues by specific data & analytics projects to prove value and effectiveness, the implementation of new technologies and external consulting. Lastly only 20% of companies report not to plan additional data & analytics investments.

Leadership

Throughout the last years, many data & analytics related C-level positions have been brought to life. The Chief Data Officer, The Chief Analytics Officer and the Chief Digital Officer are probably the best examples of this trend. However, we find that most data & analytics initiatives are still being led by the more traditional C-level officers like the CEO, COO or the CMO. In our sample, 8% of companies employ a Chief Data Officer, 4% a Chief Digital Officer and 2% has a Chief Analytics Officer on its payroll. Of the companies that do not yet have such an executive 67% indicates not to plan on hiring one. Lastly, the survey results confirm the confidence of executives in data & analytics as 63% of respondents indicate to witness strong C-level support for data & analytics initiatives.

Data

The majority of organizations indicate to have a clear view of the volume of data stored and gathered but the roles and responsibilities for the governance of this data are often not yet clearly defined. Therefore many organizations do not monitor the quality of the collected data. Less than 10% have an overview of the data quality at an enterprise level and less than 5% have data KPIs in place.

---

[1] Note that we have altered the sequence of DELTA dimensions to make the narrative as fluent as possible.

Analysts

Data-driven decision making is finding its way into the mainstream as 63% of companies indicate to use data in most or all decisions. However, only 43% of companies report they conduct post mortem analyses of the quality and accuracy of these data-driven decisions. Lastly most data & analytics teams are relatively well diversified.

Enterprise (organizational design)

Only 38% of companies organize data & analytics at an enterprise level. Moreover, 35% does not even organize analytics at the department level but relies on ad-hoc analytics by individuals for realizing their data & analytics strategies.

Empirical review of the DELTA model

By comparing the generated maturity clusters to the stages of the DELTA model we observe that the DELTA model closely describes the empiric reality for 4 out of 5 clusters. The DELTA model fails to describe the characteristics of the companies in what we call the "analytical hubris" cluster. For these organizations the relationship between ambitions and impact is non-linear. By using the insights from the Gartner Hype cycle model this behavior can be better understood.

Cluster-specific results

On the basis of the performed clustering we find that 19% of companies are "analytically impaired" meaning they lack fundamental requirements for the success of data & analytics within their organization. Almost a third of the companies in the sample are defined as "localized analytics": they have some data & analytics projects running, but do not organize them at an enterprise level. Next, 17% of companies demonstrate "analytical hubris". They have extreme data & analytics ambitions but faced with many challenges, they do not succeed at implementing them. Furthermore 23% of the companies in the sample are "analytical companies". They are relatively mature and are able to translate analytics into value in a consistent way. Lastly, 9% of the companies in the sample are defined as "analytical competitors". These companies are the data & analytics hotshots: they are very mature, innovative and data & analytics are of strategic importance to them.

Targets

The importance of data & analytics is very different across the clusters and this difference highly statistically significant. The main reasons for the implementation of data & analytics are similar across the clusters but the level of ambition, the success rate and the number of impacted business units differ significantly. When analyzing implementation difficulties we observe that the number of challenges is not a decreasing function of maturity. Moreover, the least mature organizations report the lowest number of issues but overall the difference in the number of issues is not statistically significant across the clusters. When companies mature we notice that they focus more on internal solutions (change management, new technologies and internal trainings) and less on external consulting. Lastly, more mature organizations are more likely to plan additional data & analytics related investments. This will probably increase the maturity gap between organizations even more.

Leadership

As expected, more data & analytics mature organizations witness a higher level of C-level support. When analyzing the presence of the new C-level positions we notice that the Chief Data Officer is employed in the four least mature clusters. The Chief Analytics Officer is only employed in the "analytical companies" cluster and the Chief Digital Officer is present in companies in the "localized analytics" and "analytical hubris" clusters. His role disappears once companies have reached a certain level of maturity. Surprisingly, none of these positions are present in the most mature cluster: "analytical competitors". The evidence related to the effect of these positions on the success rate of data & analytics projects is mixed. Interestingly, we find that in the cases were these officers improved the success rate it was often because of a reduction in the ambitions rather than an increase in impact. In these cases the main role of those executives is helping to focus and to prioritize.

Data

More mature companies collect a wider range of data formats from a wider range of sources. These differences are statistically highly significant. Furthermore, more mature companies have a better overview of their data and their data quality, their data governance roles are better defined and their data accessibility is higher. However these differences are not statistically significant across the five maturity clusters.

Analysts

Data-driven decision making is increasingly important for more mature organizations. All the companies in the "analytical competitors" cluster report they use data most of the time or always, compared to 0% of the "analytically impaired" companies. Furthermore, more mature companies have a higher level of team diversification.

Enterprise

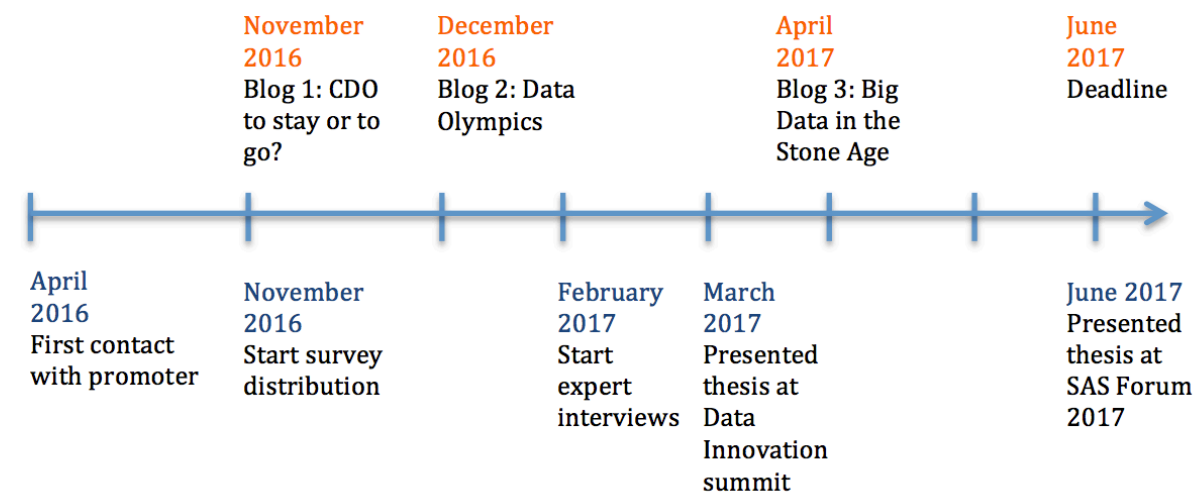
More mature organizations are more likely to organize their data & analytics efforts at an enterprise level. This relationship appears to be statistically significant.

Preface

Working on this thesis for the past year has been a great adventure for me. It was a great pleasure to do research on data & analytics, a topic that profoundly interests me, to write blogs about it and to present the findings of this thesis at several conferences. There are a number of people that I would like to thank explicitly for their support.

First of all, I would like to thank Koen Vandenbempt for giving me the opportunity to work on this topic and for his constant and insightful guidance throughout the process. Thank you Ivy Vanderheyden, Joeri Arts, Laurence Jacques, Michel Philippens, Elisabeth Versailles and Gracy Poelman of SAS. Their profound commitment to this project has been crucial. Furthermore, I deeply appreciate the possibility I was given to become a guest blogger on the SAS website and a speaker at the SAS BeLux forum 2017. Thank you Philippe Van Impe to let me present my thesis at the Data Innovation summit 2017, it was a great honor. Thank you Gregory Piatetsky-Shapiro for allowing me to publish my blogs on KDnuggets, the world's leading data science blog site. Furthermore, I would like to thank Bart De Keyser for his helpful comments and remarks. Lastly, I would like to thank Jo Coutuer, Jo Caudron, Bart Hamers, Steven Spittaels, Cédric Cauderlier and Annick Deseure for their time and the insights they shared with me during the interviews.

Timeline of milestones



List of Figures

List of tables

Table of Contents

1.0 Context

"Companies fail because they missed the future"

Larry Page

It is an understatement to say that incredible things happened on planet Earth during the last 500 years. Our society transformed from an unventilated, medieval civilization to one that is seriously considering colonizing Mars (Stockton, 2016). The scientific revolution spurred an unbelievable metamorphosis. Today, innovation goes faster than ever. Disruptive inventions like the printing press or the steam engine used to come along every century or so. Right now, new technologies of the same caliber are introduced every decade. Over the last 70 years we have witnessed the introduction of nuclear energy, the first man on the moon, the introduction of the PC, the rise of the Internet, MRI, the first cell phones, genetically modified animals, social media, bubble gum and many more (Fallows, 2013; Wolchover, 2016). Apart from transforming our society, this extreme innovation pace has profoundly impacted our way of doing business. While the average lifespan of a company in the beginning of the 1960s was above 60, it has now been brought down to 20, as depicted in figure 1. Surviving in the extremely competitive and innovative business landscape has never been more difficult. It is interesting to notice that the average lifespan has decreased almost uniformly across every sector and industry and across every size and age. "There are no safe harbors, neither scale nor experience is a safeguard" (Reeves & Pueschel, 2015).

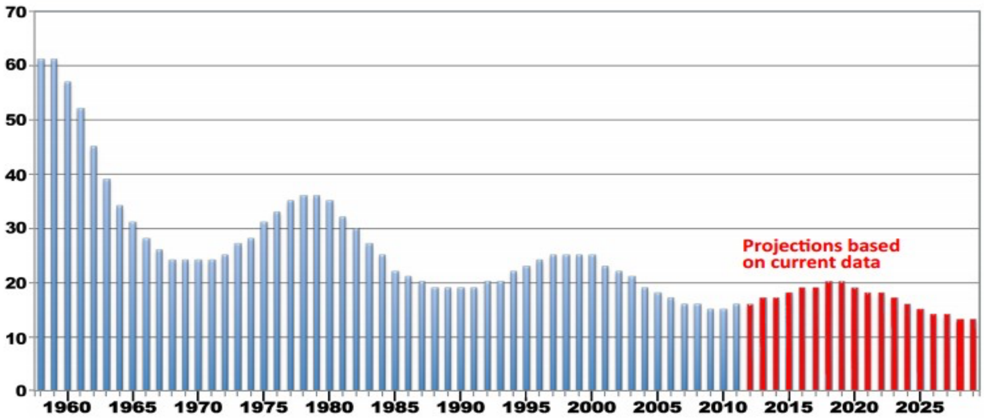


Figure 1: The average company lifespan on the S&P Index in years (rolling 7-year average).
Source: Morgan (2013)

Innovation has been quoted to be "the only way to survive and thrive in increasingly hypercompetitive markets" (Rosenbusch, Brinckmann, & Bausch, 2011). Using data & analytics in every aspect of the business cycle is one much-discussed way to stay ahead of the competition (Davenport, 2013).

1.1 The rise of data & analytics

Athens, April 10th 1896. Five athletes prepare themselves for the 100m final of the first modern Olympic games. Their hearts are racing as they walk up to the starting line. Thomas Burke is one of them. Instead of standing up straight like his competitors he crouches as he concentrates and waits for the sound of the starting pistol. Thomas is going to win the race easily. His crouch start is far more efficient; the standing start goes into the history books ("Thomas Burke (USA) - 100m", n.d.).

Mexico, October 20th 1968. Dick Fosbury gets ready to jump over 2.24m. In the next moments he is going to break the Olympic record, astonish the audience and revolutionize high jumping with his new, weird looking technique. From then on, the "Fosbury Flop" becomes the standard for professional high jumpers (Burnton, 2012).

Continuous but small improvements are inherent to all sport disciplines. Every now and then an athlete sharpens the world record by a millisecond or a centimeter because of tremendous effort and a huge portion of luck. But from time to time, a quantum leap like the crouch start or the Fosbury Flop takes place. A new technique comes along that completely outpaces the old one and shatters all previous records by a large margin. Could data & analytics be such a quantum leap in business? Could companies outrun their competitors by using data & analytics in the same way crouching Thomas Burke defeats his standing opponents any day of the week?

There are many stunning examples of how analytics is revolutionizing the business landscape and by extension the entire world. Netflix executives for example decided to make a $100 million investment in the hit show "House of Cards" after their algorithms predicted it would be a huge success (Bulygo, 2013). Back in 2002, the Oakland Athletics famously set a new winning streak record after winning 20 consecutive baseball games. While having one of the lowest budgets in the league (only about a quarter of the budget of the richest team: the New York Yankees) they managed to outrun their competitors by applying analytics to select and manage their team (Lewis, 2003). Another mind-boggling example is the anticipatory shipping system by Amazon. In 2014, the online retailer obtained a patent for a system that ships products to customers before they are even ordered. By analyzing past purchases Amazon could predict who is interested in which products and when they are going to buy them. If such a system could be successfully implemented Amazon would cut down lead times significantly (Kopalle, 2014). These applications are not even the tip of the iceberg. There are many more examples in virtually any domain ranging from cancer detection to catching terrorists and even online dating (Augur, 2016; Gandomi, & Haider, 2015).

Data & analytics are not only the talk of the town among executives. In academia, the number of papers mentioning "big data" has skyrocketed over the past few years as shown in figure 2.

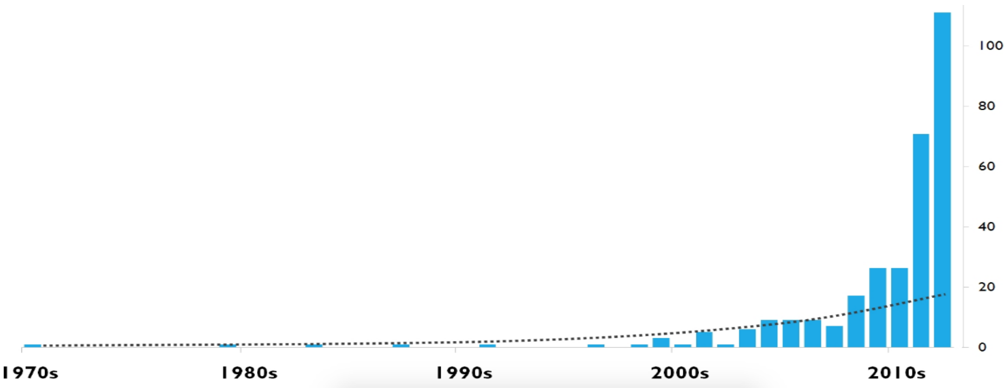


Figure 2: The number of "Big data" papers per year.
Source: Halevi & Moed (2012)

This would lead to believe that data & analytics are a recent phenomenon, but the underlying idea, the urge to understand the nature of our reality and to act accordingly, is as old as humanity itself. It is important to notice that very little of the data & analytics hype is conceptually revolutionary. For example, the first device that made efficient data analytics possible (the abacus) was introduced more than 5000 years ago (Ament, 2006). Managers as well have been trying to base their decisions on something else than gut feeling for a long time: the term business intelligence was already being used in 1865 (Devens, 1865). For a further discussion why data & analytics are nevertheless more relevant than ever we refer to extension 1.

1.2 Show me the money

On a macro level McKinsey (2013) estimates data & analytics can increase US GDP up to 1.7% ($325 billion) by 2020. On a micro level Brynjolfsson, Hitt, & Kim (2011) show strong evidence of the link between data-driven decision making and business outcomes. Data-driven decision making (or DDD) refers to "the practice of basing decisions on the analysis of data rather than purely on intuition" (Provost & Fawcett, 2013). Brynjolfsson, Hitt, & Kim (2011) show that firms with a one standard deviation higher score on a data-driven decision making index are 4.6% more productive, while controlling for other factors like IT use. The authors also find a positive relationship between data-driven decision making and market value and return on equity. Although the term data-driven decision making is already a pleonasm in many organizations (Deseure, 2017), still a surprising number of companies base most of their decisions purely on gut feeling. According to a PwC survey (2016) only 39% of companies would describe themselves as "highly data-driven" and 62% of executives rely more on intuition than on data.

This is alarming given the vast literature of evidence showing that algorithms greatly outperform human judgment on virtually every domain (Beck et al., 2011; Brynjolfsson, Hitt, & Kim, 2011; McAfee, 2013). According to Bain (2013) firms with the best data & analytics capabilities "are twice as likely to be in the top quartile of financial performance within their industries". Furthermore, a SAS study (2013) shows great savings in fuel costs for UPS after the implementation of an integrated analytics solution for route optimization (ORION). It is worth noticing that most of the benefits of data & analytics today are a result of the optimization of operational decisions. Although the impact of these decisions on the organization is several orders of magnitudes smaller than tactical or strategic decisions, their frequency is often enormous and these decisions can usually be automated by the implementation of real-time analytics (Davenport, 2013). Because of this, "the real power of data & analytics lies in the automation of operational decisions" (Philippens, 2017). The large majority of data & analytics' implementation examples are on the automation of operational decisions as shown in table 1.

Marketing	Risk	Government	Web	Logistics	Others
Response modeling	Credit risk modeling	Tax avoidance	Web analytics	Demand forecasting	Text analytics
Net lift modeling	Market risk modeling	Social security fraud	Social media analytics	Supply chain analytics	Business process analytics
Retention modeling	Operational risk modeling	Money laundering	Multivariate testing		
Market basket analysis	Fraud detection	Terrorism detection			
Recommender systems					

Table 1: Data & Analytics implementation examples.
Source: Author, based on Baesens (2014).

2.0 Problem statement

"There is a fine line between fishing and just standing on the shore like an idiot"

Steven Wright

2.1 Great Expectations

Over the last few years, companies across all industries have been implementing data & analytics in their processes. The confidence of executives in the power of analytics is high as 75% of companies are investing or planning to invest in data & analytics capacities the next two years (Gartner, 2016). Moreover, 85% believes "big data" will "dramatically change the way they do business" (Accenture, 2014) and "70% of firms view "big data" as very important or critical to their business success" (NewVantage Partners LLC, 2016). An EY & Forbes Insights study (2015) among senior executives provides another striking example of the great expectations surrounding data & analytics. The seven leading outcomes expected to be achieved via data & analytics are:

- Increased revenue
- Deeper market insights
- Increased customer satisfaction
- Improved internal operations & cutting costs
- Rapid and constant innovation
- Accelerating decision making
- Creating more partner-vendor interaction

The divergent nature of these outcomes and their pivotal importance on the survival of the company clearly demonstrate the profound faith in data & analytics. Executives are captivated by the seemingly infinite potential of data & analytics.

2.2 Don't forget the human side

Despite the extensive enthusiasm and investments, implementing data & analytics solutions is difficult. Strikingly, 92% of companies report they encounter hefty obstacles when trying to implement "big data" projects (CA Technologies, 2015). These major hurdles result in the failure of 50% of all data & analytics projects (Marr, 2016). Capgemini Consulting (2014) reposts an even lower number, "only 27% of the executives described their "big data" initiatives as successful". Of the companies that successfully implemented a data & analytics project almost 95% received external help (Accenture, 2014). Remarkably, 8 out of the 12 most mentioned obstacles are non-technological but related to people, culture or organizational design (CA Technologies, 2015). The human side of data & analytics seems to be pivotal for its success.

2.3 Key Questions

Throughout the past years, the technology and ICT infrastructure necessary to handle large amounts of data has been discussed thoroughly (Inmon, 2016; White, 2012;). Likewise, the algorithms underlying many of the most popular analytics methods have been around for decades and are well understood (Baesens, 2014; Provost & Fawcett, 2013). Today, any company can buy turnkey solutions to perform advanced analytics or to handle huge amounts of data from a myriad of software vendors. However, there is one side to this story that has not received a similar amount of attention: the organizational side (Mazzei & Noble, 2017). What is the current state of data & analytics maturity in organizations? How can companies configure themselves to optimally reap the benefits of data & analytics? Do companies need a C-level executive to fully absorb the data & analytics potential? In what business activities do companies need to implement data & analytics solutions? What are the main causes of failure of data & analytics initiatives and what are the most common solutions? These organizational issues are the key questions of this master's thesis.

2.4 Maturity modeling

Central to the investigation of these topics is the concept of maturity models. Maturity models are often used in the management literature to map the capability of organizations in certain domains (Fraser, Moultrie, & Gregory, 2002; Mettler, & Rohner, 2009; Mettler, 2011). In this thesis we will use the DELTA model (**D**ata, **E**nterprise, **L**eadership, **T**argets, **A**nalysts) depicted in figure 3 as the theoretical backbone. The DELTA model is a much-cited maturity model in the domain of data & analytics developed by Davenport, Harris, & Morison (2010). We composed a survey based on the dimensions and the stages of this maturity model. This data was used to conduct an as-is analysis. Next, we tested the theoretical insights from this model empirically by generating 5 maturity clusters from the sample. Furthermore, by comparing the characteristics of each of these 5 clusters we tried to gain insights in how companies make use of data & analytics at different maturity stages By comparing the best of class companies to the majority of firms we obtain a roadmap for further development for any organization that is looking for a quantum leap in data & analytics maturity.

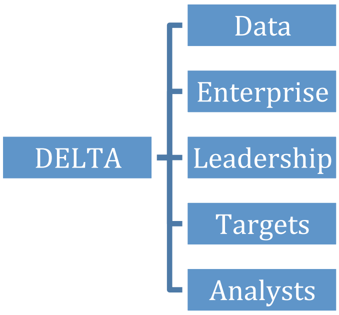


Figure 3: The DELTA model broken down into its components.
Source: Author, based on Davenport, Harris, & Morison (2010)

3.0 Research questions and outline

In the first part of this thesis some introductory but essential definitions will be discussed briefly. The following research questions are examined in the respective sections.

- **Chapter 4: Data & analytics definitions**
    - What is data?
    - What is information
    - What is analytics?

Next, maturity models are covered and critically discussed. These models are the conceptual foundation of a significant part of this thesis.

- **Chapter 5: Maturity models**
    - What are maturity models?
        - How can they be used?
        - What are their weaknesses?
    - Which maturity models exist in the domain of data & analytics?

In chapter 6 we will discuss the stages and dimensions of the DELTA model (Davenport, Harris, & Morison 2010) in depth. This model will be the theoretical foundation of our maturity research.

- **Chapter 6: The DELTA maturity model**
    - What is the DELTA model
        - Data
        - Enterprise
        - Leadership
        - Targets
        - Analysts

In chapter 7 the data & analytics maturity survey that was conducted will be introduced. We will spend a considerable amount of time examining the origins of biases related to this survey, as it is the empiric backbone of most of the conclusions of this thesis. Hereafter, we will discuss our qualitative research.

- **Chapter 7: Data & analytics maturity survey**
  - What quantitative research was conducted?
    - What are the main origins of bias?
    - In what direction does this bias influence the survey conclusions?
  - What qualitative research was conducted?

In chapter 8 we present the survey results across the five DELTA dimensions. Note that we will shake up the order of the dimensions to make the narrative as fluent as possible.

- **Chapter 8: Full sample descriptive statistics**
  - Targets
    - What are the ambitions of organizations related to data & analytics?
    - How successful are organizations in the implementation?
    - In which business activities are organizations making use of data & analytics?
    - What are the biggest challenges they face?
    - Are organizational issues a bigger challenge than technological issues?
    - How will these challenges be tackled?
  - Leadership
    - Who is leading the data & analytics initiatives?
    - How common are these new C-level positions?
    - What are their responsibilities?
    - What are the expected future trends?
  - Data
    - Which data formats are collected?
    - From which sources?
    - How well are organizations preforming data governance wise?
  - Analysts
    - How often is data used in decisions?
    - Do companies have feedback processes in place to monitor the quality of their analyses?
    - Which roles are represented in the data & analytics teams?
    - Which skills are the hardest to get by?
  - Enterprise

        - On which organizational level do companies organize their data & analytics initiatives?

In chapter 9 the theoretical insights from the DELTA framework will be combined with the survey data to generate maturity clusters of the survey respondents via the K-means algorithm. In this way we obtain an inductive, empirically grounded maturity model. The maturity clusters will be discussed briefly and will be compared to the theoretical clusters as described in the DELTA model.

- **Chapter 9: Maturity Clusters**
    - What do the generated maturity clusters look like?
        - How do they compare to the maturity stages discussed in the DELTA model?

In chapter 10 we split the data into the five previously defined maturity clusters to generate cluster-specific results. We will again cover the research questions from chapter 8 but this time compare the cluster results in order to gain insights in the behavior of the organizations at the different maturity stages. Note that we only cover the topics on which we have enough information to meaningfully compare the clusters.

- **Chapter 10: Cluster-specific results**
    - Targets
    - Leadership
    - Data
    - Analysts
    - Enterprise

- **Chapter 11: Conclusion**

4.0 Data & Analytics definitions

"Where is the Life we have lost in living?

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?"

T.S. Eliot

4.1 Data

"*Data is a set of discrete, objective **facts** about events*" (Davenport & Prusak, 2000).

*"Data is a **symbol set** that is quantified and/or qualified*" (Wersig & Neveling, 1971).

"*Data are sensory **stimuli** that we perceive through our senses*" (Davis & Olson, 1985).

*"Merely raw **facts***" (Henry, 1974).

*"Know **nothing**"* (Zeleny, 1987).

From these definitions it becomes clear that data can be:

- Facts
- Symbols
- Stimuli/Signals

Data is the raw ingredient in any epistemological process. It has no meaning *an sich*, it has to be refined in some kind of way to become valuable.

In the business informatics literature data is often divided into two categories (Davenport, 2013):

- Structured Data: refers to organized data, for example phone numbers or postal codes that always contain a fixed amount of numbers and letters. Tables and records are other frequently mentioned examples (Marr, 2015).
- Unstructured: refers to all data that is not organized in a recognizable structure (Marr, 2015). For example images, video or documents.

4.2 Information

*"Information is **data** endowed with relevance and purpose."(*Drucker, 2008)

*"Information is **data** that challenges us."* (Henry, 1974)

*"Know **what***" (Zeleny, 1987).

Information consists of data that has been better understood and contextualized, as illustrated in figure 4. Analytics provide the tools to make this transformation from data to information possible. After this processing, information can become the input for decisions.

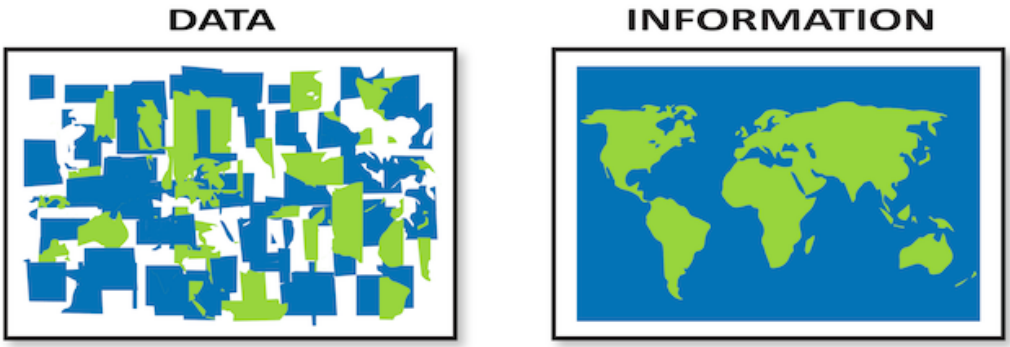


Figure 4: A graphical illustration of the difference between data and information.
Source: Systems (2014)

The Data-Information-Knowledge-Wisdom pyramid (DIKW) is an often-cited model in the information management literature that depicts an epistemological process as it materializes the interdependencies between data, information, knowledge and wisdom (Rowley, 2007; Tien, 2003; Vaes, 2013). In this chapter we have only discussed the relation between data and information, which is sufficient given the goals of this thesis. For a further discussion of these concepts we refer to extension 2.

4.3 Analytics

Every innovation wave brings about its own set of definitions and buzzwords. New terms come along that clarify and replace the existing ones but often terms are invented just for the sake of it. Especially the business world is notorious for introducing sexy new names for long-established concepts. These words do spice up the conversation, but often add no value or even create ambiguity (Green, 2011).

Along these lines, the term for applied statistics, or how managers can use data to improve their business decisions to ultimately create value has been updated many times (Davenport, 2013).

1970: Decision support systems
1980: Executive information systems
1990: Business Intelligence
2000: (Business) Analytics

The following paragraph from Davenport (2013) clearly identifies this issue. "There was much confusion about the difference between business intelligence and analytics. The CEO of a software vendor told me he thought analytics was a subset of business intelligence. Another CEO in the same industry argued business intelligence was a subset of analytics. Obviously neither term is entirely clear if each can be a subset of the other in educated executives' minds."

An inquiry into the evolution and the exact definitions of all buzzwords surrounding data & analytics could by itself be the subject of an entire master's thesis. In order not to get lost in the semantics we will only define and use the term analytics in this thesis. In extension 3 other related terms like big data, data science and data mining are discussed.

4.3.1 Definition

"Analytics refers to the discovery of meaningful patterns in data" (NIST, 2015).

Through analytics data can be refined to information. Analytics is often divided into four subcategories as depicted in figure 5 (Davenport, 2013; "Descriptive, Predictive, and Prescriptive Analytics Explained", 2016).

1. *Descriptive* analytics: refers to methods for describing the past. Descriptive analytics can provide an answer to the question: "what happened?" Possible methods are summary statistics, management reports and database queries.
2. *Diagnostic* analytics: refers to methods that reveal correlations and causality between variables. Diagnostic analytics can provide an answer to the question: "why did it happen?" Possible methods are linear regression or logistic regression.
3. *Predictive* analytics: refers to methods that use past data to predict the future as accurately as possible. Predictive analytics can provide an answer to the question: "what will happen?" Methods range from linear regression or decision trees to more complex machine learning algorithms like support vector machines or neural networks.
4. *Prescriptive* analytics: refers to methods that provide guidance towards the most optimal future scenario. Prescriptive analytics can provide an answer to the question: "what should happen?" Inventory or supply chain optimization are common examples.

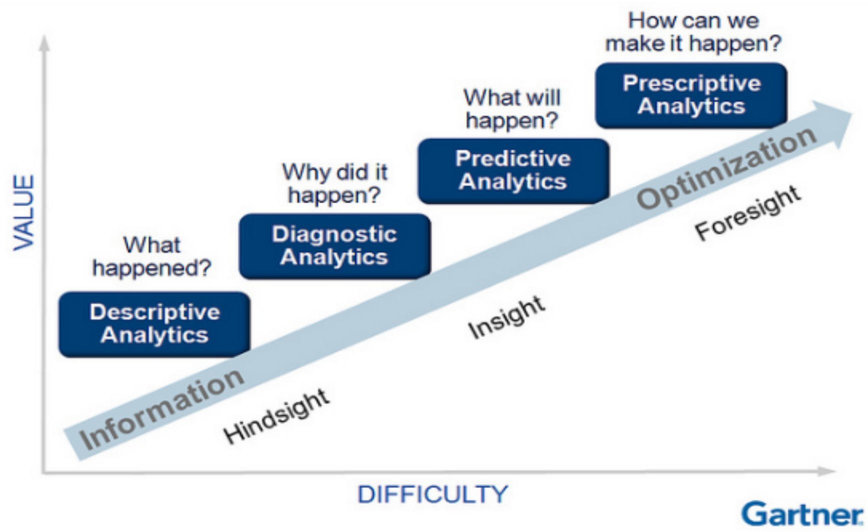


Figure 5: The Analytic value Escalator.
Source: Gartner's Analytic Value Escalator, 2016

5.0 Maturity models

In this chapter the maturity model concept will be defined and discussed. Maturity models are widely used in management domains as a tool for assessing the capability of an organization in certain business practices (Fraser, Moultrie, & Gregory, 2002; Mettler, & Rohner, 2009; Mettler, 2011). These models are very useful and pragmatic as they help substantiating the relative maturity of firms and provide a roadmap for further development. However, many maturity models are not empirically grounded and contain methodological flaws, which will be discussed explicitly throughout this chapter.

5.1 Definitions

5.1.1 Maturity

"*The state of being complete, perfect or ready*" ("Maturity", 2010).

"Maturity thus implies an evolutionary progress in the demonstration of a specific ability or in the accomplishment of a target from an initial to a desired or normally occurring end stage" (Mettler, 2011).

5.1.2 Maturity model

"Maturity models are conceptual models that outline anticipated, typical, logical, and desired evolution paths towards maturity (Becker, Knackstedt, & Pöppelbuß, 2009). *Maturity models are of normative nature* (Iversen & Nielsen & Norbjerg 1999) *and can be understood as reference models* (Herbsleb, Zubrow, Goldenson, Hayes, & Paulk, 1997)" (Becker, Niehaves, Poeppelbuss, & Simons, 2010).

Gibson and Nolan (1974) popularized maturity models in the domain of business information systems. Ever since, countless maturity models have been developed in virtually all aspects of every industry (Mettler, 2011). Many consulting firms or software companies have assembled their own model that accentuates slightly different facets of the same leitmotiv. In academia, the interest in maturity models has skyrocketed as well over the past decades as shown in figure 6. It is worth mentioning that some researchers are very skeptical about maturity models because of their lack of profound empirical grounding. "They have been characterized as step-by-step recipes that oversimplify reality and lack empirical foundation […] and neglect the potential existence of multiple equally advantageous paths" (Becker, 2010). Others argue that although these models provide a clear picture of a desired capability level they no not necessarily show how to get there. "This knowing-doing gap can be very difficult to close (Mettler, 2011)".

5.1.3 Purposes of maturity models

A maturity model can be used for three distinct purposes (Becker, 2010).

1. **Descriptive purpose**: assessing the current maturity level of an organization.
2. **Prescriptive purpose**: providing a roadmap for further development.
3. **Comparative purpose**: benchmarking the maturity of an organization relative to others or to theoretical boundaries.

Throughout the next chapters we will zoom in on each of these purposes.

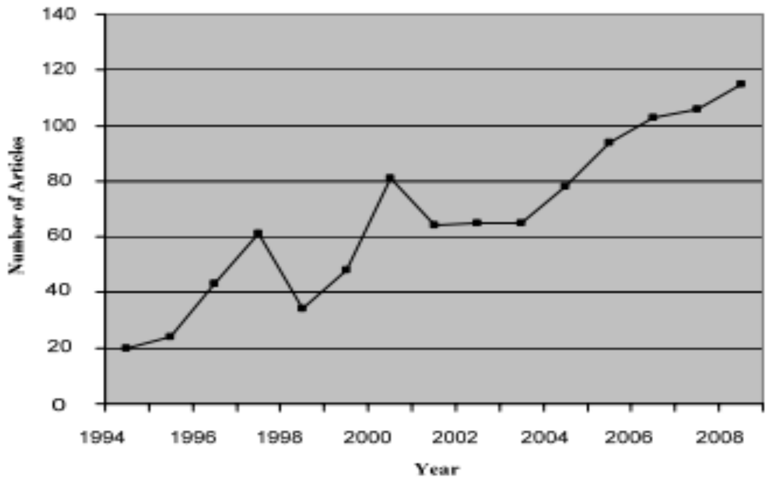


Figure 6: The number of maturity model related publications throughout the years.
Source: Becker, Niehaves, Poeppelbuss, & Simons (2010).

5.2 Maturity model properties

Although many different maturity models have been introduced in many different industries, they all share three main properties that will be discussed briefly in the next section.

5.2.1 Levels (Stages)

The levels in a maturity model are the different phases an organization goes through when it matures. Most maturity models have five maturity levels although this number tends to fluctuate between three and six (Davenport, Harris, & Morison, 2010; Fraser, Moultrie, & Gregory, 2002). Many authors argue that the number of levels is often highly arbitrary and not backed by empirical evidence (Fraser, Moultrie, & Gregory, 2002). "God has decreed that all maturity models have five stages" (Davenport, Harris, & Morison, 2010). All these stages have a name and a description of the characteristics of each of these stages is provided. Table 2 depicts the information evolution model (SAS, 2007). In this maturity model the levels are: operate, consolidate, integrate, optimize and innovate.

The Information Evolution Model

Level	Level Dimensions			
	Infrastructure	Knowledge Process	Human Capital	Culture
I. Operate	Manual systems of non-networked PCs	Personal	Individual	Me
II. Consolidate	Functional systems	Department	Functional group	Our group vs. the rest of the company
III. Integrate	Enterprise systems	Enterprise	Enterprise group	All of us
IV. Optimize	Extended enterprise systems	Extended enterprise	Extended group	Our partners and us
V. Innovate	Adaptive systems	Situations matrix	Dynamic network	Adaptive groupings

Table 2: The information evolution model.
Source: SAS (2007)

5.2.2 Dimensions

In maturity models, maturity is often a function of capability on several dimensions. Each of these dimensions captures a different aspect of the proficiency of an organization. These dimensions should be mutually exclusive and collectively exhaustive (Lahrmann, Marx, Winter, & Wortmann, 2011). Again, the chosen dimensions are often arbitrary. In the information evolution model (SAS, 2007) the dimensions are infrastructure, knowledge processes, human capital and culture as depicted in table 2.

5.2.3 Maturity principle

Lastly, maturity models can be classified into two categories: models with a continuous maturity principle and models with a staged maturity principle. For models with a continuous maturity principle the maturity of an organization is the weighted sum of the maturity on each of the dimensions. A company can thus increase its overall maturity by increasing its maturity on one of the dimensions. In models with a staged maturity principle the maturity of an organization is equal to that of the dimension with the lowest maturity. The maturity of an organization is equal to that of its weakest link. To increase its overall maturity an organization must raise its maturity on each of the dimensions (Lahrmann, Marx, Winter, & Wortmann, 2011). Although this division between maturity models with a continuous and a staged maturity principle is discussed in the maturity model literature, many authors of maturity models do not make this split. Therefore, we will not explicitly mention this characteristic when we compare maturity models in the next section.

5.3 Data & analytics maturity models

As an illustration of the proliferation of maturity models in the domain of data & analytics we discuss and compare some much-cited models in table 3. Note that this list is by no means exhaustive, for a more in depth discussion of the different developed models we refer to Lahrmann, Marx, Winter, & Wortmann (2011) or Lismont, Vanthienen, Baesens, & Lemahieu (2017).

Name	Levels	Dimensions
SAS information evolution model (2007)	5	4 (Infrastructure, knowledge process, human capital, culture)
Data warehousing maturity model (Watson, Ariyachandra, & Matyska, 2001)	3	9 (data, architecture, stability of the production environment, warehouse staff, users, impact on users' skills and jobs, applications, costs and benefits, and organizational impacts)
Data governance maturity model (DataFlux, 2007)	4	4 (People, policies, technology, risk and reward)
DELTA model (Davenport, Harris, &Morison, 2010)	5	5 (Data, enterprise, leadership, targets, analysis)
The Digital Maturity Model 4.0 (Forrester, 2016)	4	4 (Culture, organization, technology, insights)
Big data maturity model (Hortonworks, 2016)	4	5 (Sponsorship, data and analytics, technology and infrastructure, organization and skills, process management)
Master data management (KPMG, n.d.)	5	4 (Governance, systems, processes, content)
The IBM Data Governance Council Maturity Model (IBM, 2007)	5	11 (Organizational Structures & Awareness, Stewardship, Policy, Value Creation, Data Risk Management & Compliance, Information Security & Privacy, Data Architecture, Data Quality Management, Classification & Metadata, Information Lifecycle Management, Audit Information, Logging & Reporting)
TDWI Analytics Maturity Model Guide (TDWI, 2014)	5	6 (Infrastructure, data management, analytics, governance, organization)
BA capability maturity model (Cosic, Shanks, & Maynard, 2012)	5	4 (Governance, culture, technology, people)

Table 3: A list of common maturity models in data & analytics.
Source: Author, based on the sources cited in the table.

6.0 DELTA maturity model

In the rest of this thesis we will use the DELTA model as our conceptual maturity framework. This model has been selected over the others because of the following reasons. Firstly, this model is purely non-technological which coincides perfectly with the goal of this thesis. Secondly, the model is very thoroughly documented by the authors (Davenport, Harris, & Morison, 2010). Thirdly, this maturity model has already been the subject of an empirical review (Lismont, Vanthienen, Baesens, & Lemahieu, 2017).

6.1 Maturity stages

The DELTA maturity model (**D**ata, **E**nterprise, **L**eadership, **T**argets and **A**nalysts) is defined over five stages and in five dimensions as depicted in table 4.

	Stage 1: Analytically impaired	Stage 2: Localized Analytics	Stage 3: Analytical Aspirations	Stage 4: Analytical Companies	Stage 5: Analytical Competitors
Data	Poor	Usable	Consolidated	Integrated	Innovative
Enterprise	-	Islands	Early stages of an enterprise wide approach	Some key analytical resources are centrally managed	All key analytical resources centrally managed
Leadership	None	Local	Aware	Supportive	Passionate
Targets	-	Disconnected targets	A small set of clear targets	Analytical activity centered on a few key domains	Analytics support the firm's distinctive capability and strategy
Analysts	Few skills	Isolated pockets of analysts	Influx of analysts in key target areas	Highly capable analysts in central or networked organization	World class professional analysts and attention to analytical amateurs

Table 4: The DELTA model over 5 stages and 5 dimensions.
Source: Davenport, Harris, & Morison (2010).

Stage 1: Analytically Impaired. The organization lacks one or several of the prerequisites for serious analytical work, such as data, analytical skills, or senior management interest.

Stage 2: Localized Analytics. There are pockets of analytical activity within the organization, but they are not coordinated or focused on strategic targets.

Stage 3: Analytical Aspirations. The organization envisions a more analytical future, has established analytical capabilities and has a few significant initiatives under way, but progress is slow –often because some critical DELTA factor has been too difficult to implement.

Stage 4: Analytical Companies. The organization has the needed human and technological resources, applies analytics regularly and realized benefits across the business. But its strategic focus is not grounded in analytics and it hasn't turned analytics to competitive advantage.

Stage 5: Analytical Competitors. The organization routinely uses analytics as a distinctive business capability. It takes an enterprise-wide approach, has committed and involved leadership, and has achieved large-scale results. It portrays itself both internally and externally as an analytical competitor.

6.2 Dimensions

In this section we will introduce the five dimensions of the DELTA model. Note that we have altered the sequence of the dimensions to make the narrative as fluent as possible.

6.2.1 Targets

The targets dimension of the DELTA model deals with the organization's strategic plan related to data & analytics. Davenport, Harris, & Morison (2010) do not provide a clear-cut definition of what they mean by targets or a data & analytics strategy but they point out "finding your opportunities" and "setting your ambition" are the key activities underlying smart targets and a solid data & analytics strategy. In later work Davenport and Dallemule (2017) elaborate on data & analytics strategies and provide a very workable framework by dividing these strategies into two groups: defense and offence. "Data defense is about minimizing downside risk. Activities include ensuring compliance with regulations, using analytics to detect and limit fraud, and building systems to prevent theft (Davenport, & Dallemule, 2017)." Data offense on the other hand "focuses on supporting business objectives such as increasing revenue, profitability, and customer satisfaction. It typically includes activities that generate customer insights or integrate disparate customer and market data to support managerial decision making (Davenport, & Dallemule, 2017)."

Many examples of both offensive and defensive applications were shown in chapter 1, table 1. To work out a solid data & analytics strategy and to set smart targets companies have to find a good balance between the offensive and defensive aspects. The authors stress the importance of sector specific characteristics on this tradeoff. In heavily regulated industries like the financial sector, companies rightly tend to put the emphasis on defense while companies in the retail industry tend to invest more heavily in offensive capabilities. Throughout the rest of this thesis we will often refer to this offense/defense tradeoff.

6.2.2 Leadership

During the past few years many data & analytics related C-level positions have been brought to life. The Chief Data Officer, the Chief Analytics Officer (CAO) or the Chief Digital Officer[2] are the most common examples of this trend. In this section we will give a very brief overview of these new positions in order to make comparisons between the literature and our findings possible. A thorough discussion of the existing literature on these positions can be found in extension 4.

Chief Data Officer

"The CDO is the executive who holds the keys to help an organization both protect and unlock the full value of its data assets" (Deloitte, 2016).

Data is increasingly being regarded as a strategic corporate asset. Just as other crucial assets like money or people have their specific business executives some companies feel it makes sense to create a data related C-level position: the Chief Data Officer (Coutuer, 2017). Although the first Chief Data Officers were appointed in the early 2000s, the position only became mainstream after the financial crisis of 2007-2008. Many financial institutions hired a Chief Data Officer to comply with tightened regulatory requirements concerning data quality or privacy that were brought to life after the crisis (defensive purposes in our terminology) (Bean, 2016).

Deloitte (2016) and PwC (2015) identify four main responsibilities for the Chief Data Officer: governance, technology & architecture, analytics and innovation as depicted in figure 7.

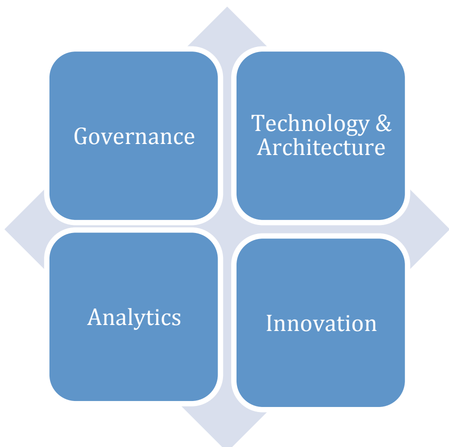


Figure 7: Four dimensions of responsibilities for the Chief Data Officer
Source: Deloitte (2016), PwC (2015)

2 To avoid confusion the acronym CDO will never be used as it refers to both the Chief Data Officer and the Chief Digital Officer.

Note that this responsibility spectrum is peculiarly broad and encompasses both offensive and defensive goals. The number of companies with a Chief Data Officer on their payroll increased significantly during the last few years. NewVantage Partners (2016) finds that more than 50% of companies now have a Chief Data Officer in their ranks. Forrester (2015) nuances these results slightly; they find that 45% of firms have a Chief Data Officer. Interestingly they also find top firms are 65% more likely to appoint a Chief Data Officer. This result is consistent with Xu, Zhan, Huang, Luo, & Xu (2016), who find that companies that appoint a Chief Data Officer have superior financial results in comparison with companies that do not. An enormous increase in Chief Data Officers is also the conclusion of research by Gartner (2016). Remarkably, they make the bold prediction that 90% of large firms will have appointed a Chief Data Officer by 2019 but only half of those will be successful. Lastly, it is worth pointing out that 85% of Chief Data Officers are active in the US or the UK as depicted in figure 8.

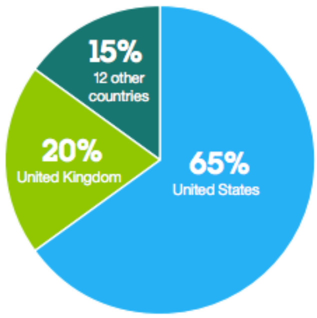


Figure 8: Chief Data Officers by geography
Source: IBM (2014)

Chief Analytics Officer

The Chief Analytics Officer is the latest data & analytics related C-level position that has been brought to life. Therefore, significantly less research on this position has been conducted (Agarwal, 2015). The Chief Analytics Officer mainly emerged as an answer to the width of the responsibility spectrum of the Chief Data Officer as shown in figure 7. Many authors agree that it is very difficult for one and the same person to encompass all these responsibilities effectively (O'Regan, 2014; Suer, 2015). Therefore, a Chief Analytics Officer is appointed to deal with the offensive side of analytics while the Chief Data Officer can fully focus on defense as depicted in figure 9. “Peanut butter and chocolate may work in a Reese’s cup but it will not work here—the orientations are too different (Suer, 2015).”

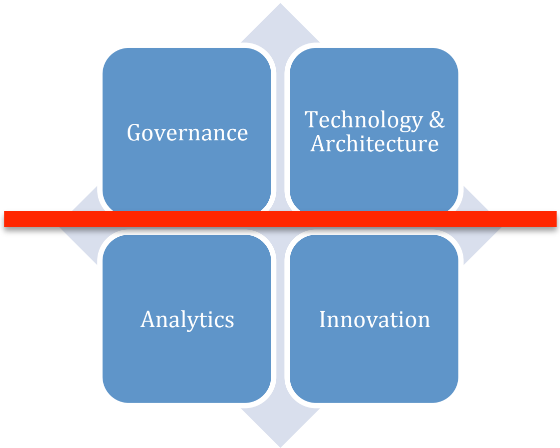


Figure 9: Four dimensions of responsibilities for the Chief Data Officer revisited. The Chief Data Officer focuses on defense (governance and technology & architecture), the Chief Analytics Officer deals with offense (analytics & innovation).
Source: Author

Chief Digital Officer

The last new C-level position that we are going to discuss is the Chief Digital Officer. Although this role is not as closely linked with data & analytics as the previous ones it is an often-mentioned term in the literature and in the industry. It is valuable to clearly identify its responsibilities as its acronym is identical to that of the Chief Data Officer and the two terms could, mistakenly, be used interchangeably.

"[The Chief Digital Officer] takes care of digital innovation both externally, in the companies' interactions with customers, partners, and suppliers, and internally, collecting and analyzing data, improving efficiency through the use of digital technologies, and transforming organization and culture to enable their companies to compete successfully in the digital age" (Strategy&, 2015).

As the abovementioned definition implies, the Chief Digital Officer is a transformational manager that leads the digital changeover of a company, as digital has become an inextricable aspect of any business (Johnston, 2016). A major difference between the Chief Data Officer and the Chief Analytics Officer on the one hand, and the Chief Digital Officer on the other hand is the temporary nature of the latter. Once a company has encoded digital into its DNA the need for a Chief Digital Officer disappears as depicted in figure 10. "Digital will become so infused with the business that it will make no more sense to have a separate leader and separate team than it does now to have a Chief Email Officer" (Deloitte, 2015).

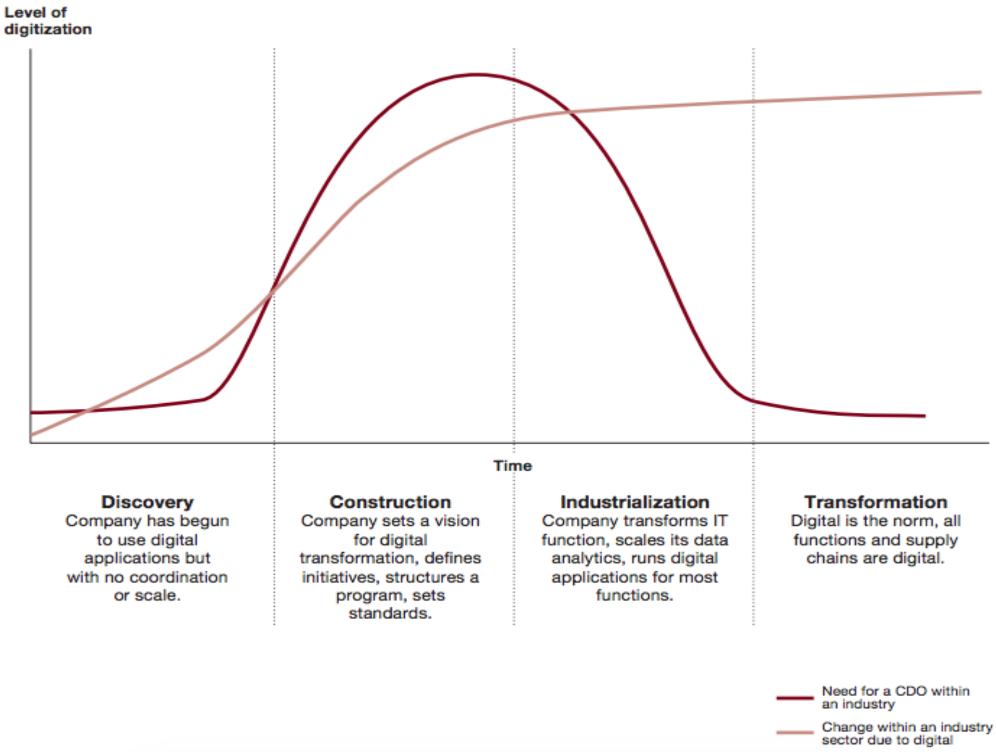


Figure 10: The need for a Chief Digital Officer during each phase of the digital revolution.
Source: Strategy&, 2015

In 2015, Strategy & concluded that 6% of the world's 1500 largest businesses had a Chief Digital Officer among their ranks (Strategy &, 2015).

6.2.3 Data

This definition was already discussed in chapter 4.

6.2.4 Analysts

Davenport, Harris, & Morison (2010) define analysts as *"workers who use statistics, rigorous quantitative or qualitative analysis, and information modeling techniques to shape and make business decisions".* Furthermore, they subdivide analysts into four categories: analytical champions that who analytics initiatives, analytical professionals who build analytical models and algorithms, analytical semi-professionals who apply analytical models to business problems and analytical amateurs who put the output of analytical models to work as depicted in figure 11. In our analysis culture is interpreted as a sub-dimension of analysts meaning we will also report findings on for example data-driven decision making in this dimension.

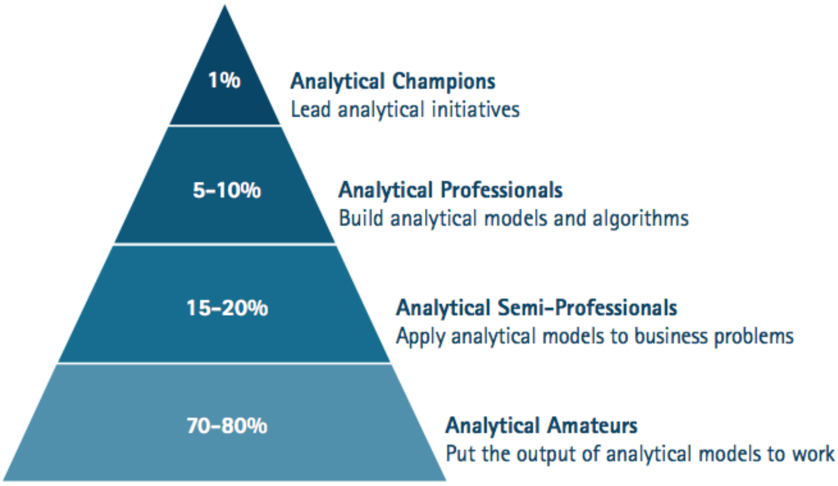


Figure 11: Types of analysts
Source: Accenture (2010)

6.2.5 Enterprise (Organizational design)

The design spectrum

Davenport, Harris, & Morison (2010) describe five common organizational models for organizing data & analytics. These models constitute a spectrum ranging from centralized to decentralized as depicted in figure 12.

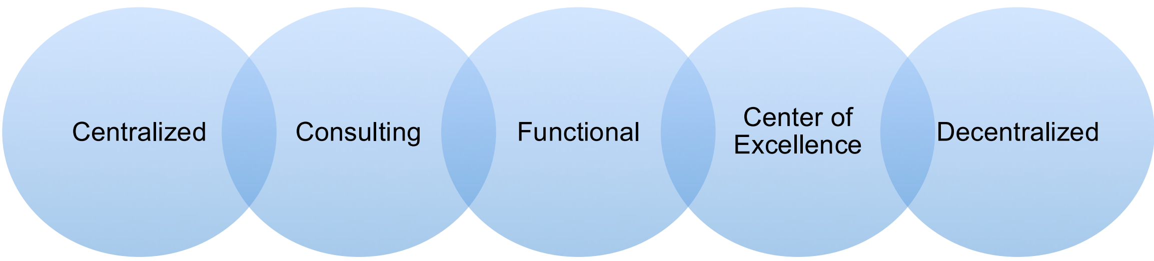


Figure 12: The design spectrum
Source: Author, inspired by Davenport (2013)

Broadly speaking, centralization fosters coordination while decentralization fosters flexibility as discussed oftentimes in the literature (Jones, 2010). In the next paragraphs we will discuss the differences between these five organizational models.

Centralized model

In a centralized design all data scientists are part of the same corporate unit, hence the name. This unit in turn carries out activities for the other departments. This way the leader of the analytics group can allocate his resources strategically and optimally across the different divisions. The centralized model fully leverages synergies like for example cross-pollination of ideas, the standardization of analytics methods and software, the development of skills or the formation of an analytics community. The major drawback of this design is the distance between analytics and the business. This distance hampers communication and can lead to analytics projects that do not fully satisfy the needs of the business units because the data scientists lack relevant business knowledge (Davenport, 2013).

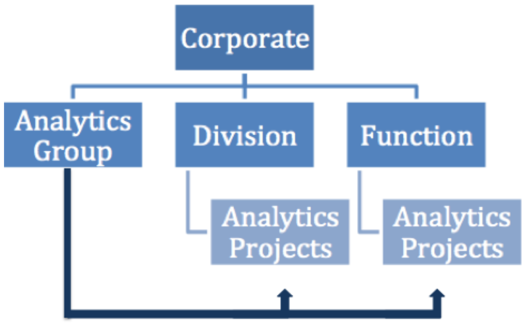


Figure 13: The centralized model
Source: Davenport (2013)

Consulting model

The consulting model as shown in figure 14 strongly resembles the centralized model. The key difference is that in this design the departments "hire" data scientists for analytics projects making this design more market driven. A potential drawback of this design is the allocation of projects across the departments. The projects could be allocated to the departments with the largest budgets instead of to those where the most value could be generated (Davenport, 2013).

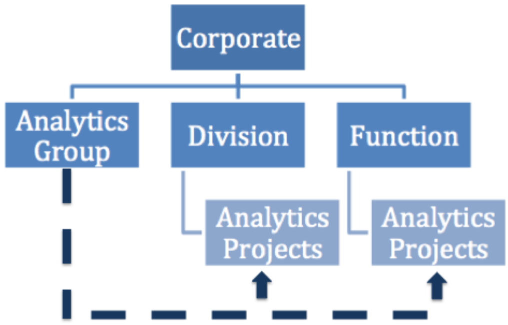


Figure 14: The consulting model
Source: Davenport (2013)

Functional model

In a functional model the analytics group is stationed in an analytics intensive department. Next to performing analytics for this main business unit the analytics team can execute consulting assignments for the rest of the company. A major advantage of this

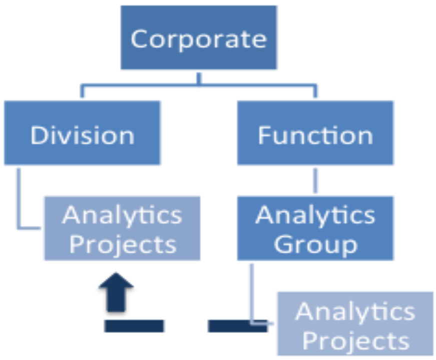


Figure 15: The functional model
Source: Davenport (2013)

model is the proximity between the business and the analytics group. Many organizations have this functional model in one or two departments, often marketing and manufacturing (Davenport, 2013).

Center of Excellence

The center of excellence model is an extension of the functional model in which the different analytics groups are coordinated by a center of excellence. Because of this center of excellence the decentralized analytics groups can take advantage of the synergies already mentioned in the description of the centralized model (Davenport, 2013).

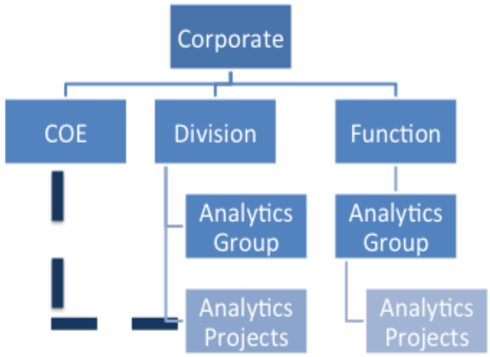


Figure 16: The center of excellence model
Source: Davenport (2013)

Decentralized

The last model we will discuss is the decentralized design. In this model all analytics groups are scattered throughout the organization with little or no communication between these silos as depicted in figure 17. This design is only worth considering when the analytics going on in the different business units have little to nothing in common. "This is the model we are least likely to endorse" (Davenport, 2013).

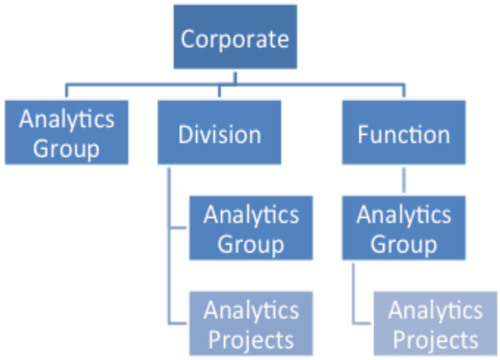


Figure 17: The decentralized model
Source: Davenport (2013)

The optimal configuration

All the discussed organizational designs have their pros and cons. The optimal company layout should be constructed in such a way that it "keeps analysts both close to the business and close to each other (Accenture, 2010)". There is no one size fits all architecture to achieve this outcome in every organization (Stubbs 2014). The optimal configuration is extremely dependent of company characteristics like, among others, size, industry and data & analytics ambitions. However, there seems to be a consensus that centralization to some degree is necessary to fully leverage the data & analytics potential (Accenture, 2010; Miller, Gerlach, & Bräutigam, 2006).

6.3 Critical remarks

Overall, the DELTA model is a solid and useful maturity model although some major setbacks have to be stressed. Firstly, it is convenient to have a maturity model of which the dimensions form a pronounceable acronym. However, the names of the dimensions seem to be molded with exactly this goal in mind, which compromises the prerequisites of collectively exhaustiveness and mutually exclusiveness. The culture dimension for example (included in many other data & analytics maturity models) is nowhere to be found. Whether this dimension is interpreted as a sub-dimension of Analysts, Enterprise, or not taken into account at all is never mentioned. Furthermore the Enterprise dimension is very poorly defined. Many of the elements mentioned in this dimension are also part of other dimensions, which compromises the mutually exclusiveness prerequisite. For example organizational design is mentioned in both Enterprise and Analysts. Strategies are discussed in both Enterprise and Targets. Applications are both brought up in Enterprise and Targets, Data governance is both mentioned in Enterprise and Data. To avoid the abovementioned confusion we will interpret the Enterprise dimension in this thesis solely as organizational design. Lastly, we would like to stress again that the DELTA model is, as almost all maturity models based on the experience on the authors and many qualitative interviews. However, as discussed in section 5.1.2, it is not based on solid quantitative research.

7.0 Data & analytics maturity survey

In this chapter we discuss the data & analytics maturity survey. We will explicitly stress the main sources of bias that influence the conclusions of this thesis.

7.1 Data Collection

The survey was distributed via email to a sample of 2000 business executives. The contacts are a combination of SAS Belgium & Luxembourg clients and personal contacts of the promoters of this thesis (LinkedIn contacts). In total, 169 managers participated in the survey resulting in a participation rate of 8%. However, only the surveys that had a completion rate of more than 50% were used in the analysis, resulting in a dataset of 90 useful surveys. To optimize the validity of the answers all respondents were given the option to participate anonymously.

7.2 Survey questions

In total, the survey consisted of 37 questions (36 closed questions and one open question). Four of these questions assess general company characteristics (number of employees, annual revenue, age and sector), the other 33 questions gauge data & analytics maturity across the five DELTA dimensions. The full list of all questions and answer possibilities is included in appendix 1. It is important to note that the knowledgeability of the respondents with respect to data & analytics could be substantial leading to biased results.

7.3 Data Description

7.3.1 Sector

The sample contains 90 data points. In total, 25 sectors are represented although we observe a clustering around a few key sectors as shown in table 5. Financial services firms are strongly represented as they account for almost 30% of total respondents. IT companies come in second and account for 12% of total responders. The top 6 industries (chemical, financial services, food & beverage, human resources, IT services and research) represent approximately two thirds of all respondents. Note that some respondents did not specify the sector of their company hence the 5 missing values.

7.3.2 Age

Older firms are strongly represented in the sample as the lion's share of companies (around 50%) was founded before 1950 as shown in table 6. Companies that were founded between 1950-1990 and between 1990-2010 account for roughly the same number of respondents (20%). Lastly, 10% of the observations are start-ups and other young companies.

7.3.3 Number of employees

The sample largely consists of companies with more than 1000 employees (55%). The segments of firms with a number of employees between 0-10, 10-50 and 250-1000 are almost uniformly represented in the data set as shown in table 6. The segment of firms with 50-250 employees are strikingly underrepresented. In hindsight and with further research in mind, it could be advisable to further subdivide the category >1000 employees as it represents more than half of the respondents.

	Frequency	Percent
Missing values	5	5.6
Agriculture & Agribusiness	1	1.1
Art	2	2.2
Automotive	2	2.2
Chemical	5	5.6
Communication	1	1.1
Construction	2	2.2
Creative Industries	1	1.1
Education	1	1.1
Electronics	1	1.1
Energy	2	2.2
Entertainment	1	1.1
Financial services	26	28.9
Food & beverage	6	6.7
Health	1	1.1
Human resources	6	6.7
IT services	11	12.2
Legal services	2	2.2
Maintenance	1	1.1
Manufacturing	2	2.2
Professional Services	1	1.1
Public sector	2	2.2
Real estate	1	1.1
Recycling	1	1.1
Research	5	5.6
Web services	1	1.1
Total	**90**	**100.0**

Table 5: Sector of respondents
Source: Author

7.3.4 Annual revenue

The distribution of the annual revenue of the companies in the sample logically strongly resembles that of the number of employees. Almost half of the companies have an annual revenue that exceeds €1 billion. Again, the critical remark can be made that the last category should be subdivided into more categories.

Founding year	Annual Revenue	Number of employees
<1950: 49%	<€1m: 15%	1-10: 13%
1950-1990: 19%	€1m-€10m: 13%	10-50: 15%
1990-2010: 23%	€10m-€50m: 5%	50-250: 3%
>2010: 9%	€50m-€250m:13%	250-1000: 15%
	€250m-€1b: 8%	>1000: 54%
	>€1bn: 46%	

Table 6: Descriptive statistics of the dataset
Source: Author

7.3.5 Job title of respondents

Figure 18 plots the job titles that appeared at least twice in the data set. More than a third of the respondents hold a C-level position. The CEO is the most represented C-suite member, before the CIO and the CDO (Chief Data Officer). Approximately 20% of respondents hold a Data/Analytics Manager position.

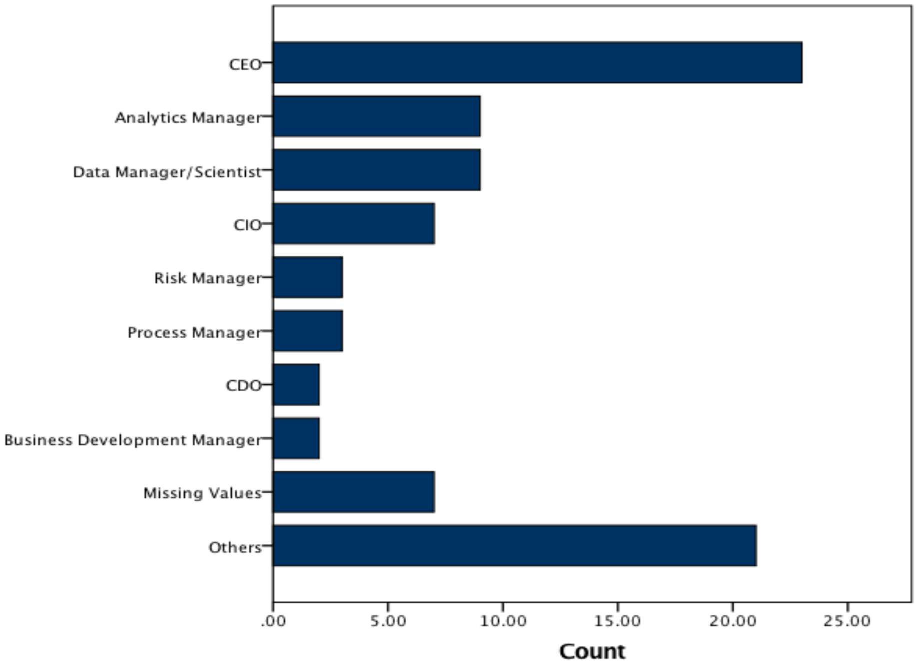


Figure 18: Bar chart of the job titles of the respondents
Source: Author

7.4 Survey Bias

In this section a few major methodological issues will be brought to light. The most relevant possible biases are discussed and their existence is tested.

7.4.1 Sampling bias

The sample of 2000 business executives that was used is clearly non-random and could heavily underestimate the number of companies with a very low data & analytics maturity. Organizations that are already in contact with a software firm like SAS are very likely to be implementing more advanced data & analytics solutions than organizations that are not.

Therefore the number of companies with a very low data & analytics maturity is probably underestimated.

7.4.2 Self-selection bias/ Non-response bias

Only 8% of all selected managers started the survey. The rest never opened the survey because it either ended up in their spam folder, a lack of interest or any other reason. It is dangerous to assume that the sample of responders is stratified. If the sample is stratified, the non-responders are Missing Completely at Random (MCAR) and there are no effects on the conclusions (Ghosh, Little, &Rubin, 1988). It is reasonable to assume that companies with a higher data & analytics maturity are more interested in the subject and therefore more likely to respond. Similar to the sampling bias, the conclusions will overestimate the organizations with a high data & analytics maturity.

7.4.3 Attrition

Only 53% of all participants finished the entire survey. The sign of the bias will again be in the direction of the organizations with a high data & analytics maturity. The argumentation is similar to that of the non-response bias. Figure 19 gives a graphical representation of these three discussed sources of bias.

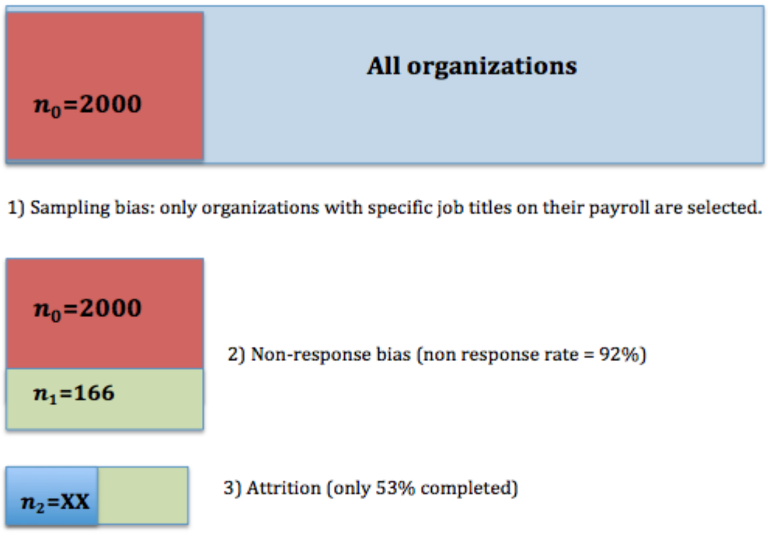


Figure 19: A graphical representation of the biases
Source: Author

7.4.4 Test for bias

To get a grip of the size and the sign of the biases we have discussed in the previous sections we compare the distribution of annual revenue of the companies in the sample with the distribution of all European companies in the Amadeus database for which annual revenue data is available (10,958,090 companies).

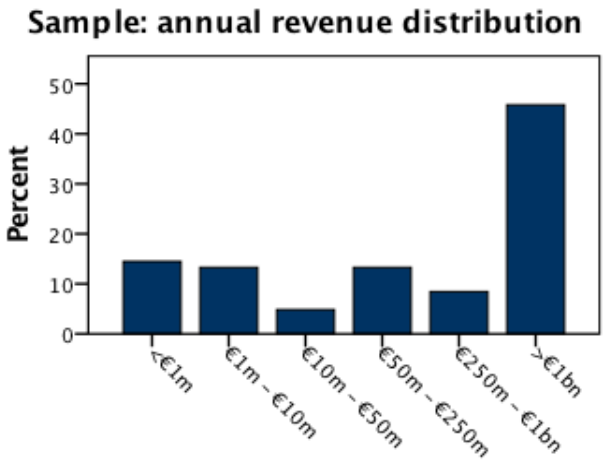


Figure 20: Annual revenue distribution of the sample
Source: Author

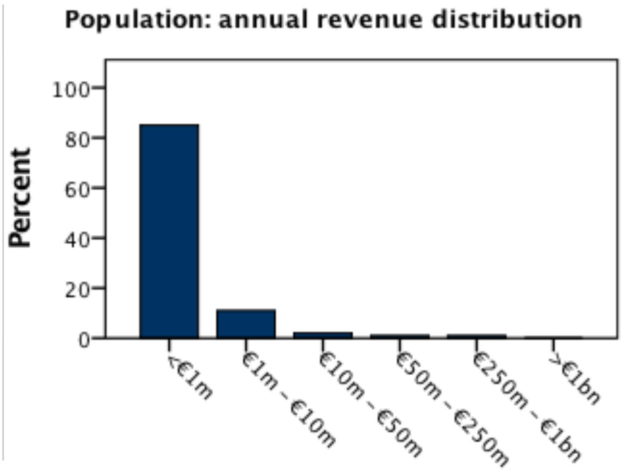


Figure 21: Population annual revenue distribution
Source: Amadeus

By comparing figure 20 and 21 it is clear that our sample is extremely biased in the direction of large companies. Only 0.04% of European companies have an annual revenue of more than €1billion while this segment represents 46% of our sample. This analysis provides clear evidence that the size of the biases cannot be underestimated.

7.4.5 Modest sample size

Only 169 respondents participated. Although this number is comparable to other data & analytics survey studies (Lismont, Vanthienen, Baesens, & Lemahieu, 2017) it is low in absolute terms. Generalizing findings should thus be done with great care.

7.4.6 Likert scales

In the survey, 10 out of 37 questions were Likert scale questions. The methodological issues of Likert scales have been discussed many times in the literature (Ogden, & Lo, 2012). As an exhaustive overview of these problems, and the sign and the magnitude of the bias they cause, could by itself be the subject of an entire master's thesis we will just refer to Carifio, & Perla (2007) and Lee, Jones, Mineyama, & Zhang (2002) who cover many of these issues.

7.5 Qualitative research methodology

Next to the quantitative survey a much smaller qualitative research has been conducted. Experienced business executives in the field of data & analytics have been interviewed. The purpose of these interviews is two-sided. First: to dig deeper into the role of the Chief Data Officer, Chief Digital Officer and the Chief Analytics Officer as a preparation for the CDO panel discussion on the SAS BeLux Forum 2017 moderated by the author. Second: to explore other analytics related topics that were difficult to bring to light by using a survey with closed questions. The results from these interviews have been used both implicitly throughout the entire thesis and quoted in the relevant sections. Note that the interviews were conducted before the survey results were analyzed. In retrospect and with future research in mind it would have been better to plan these interviews after the quantitative research was entirely finished. This could have increased the added value of the interviews with respect to the second purpose dramatically.

Name	Role	Company	Date	Duration	**Topic**
Michel Philippens	Analytics and decision management consultant	SAS	01/03/17	1 hour	Data & analytics in general
Jo Coutuer	Chief Data Officer	BNP-Paribas Fortis	06/03/17	1 hour	CDO panel preparation + organizational design
Bart Hamers	Senior lead in Analytics & Information Management	Deloitte	09/03/17	1 hour	CDO panel preparation, biggest implementation challenges and solutions
Steven Spittaels	Senior Partner	McKinsey	13/03/17	30 min	Future trends in data & analytics
Jo Caudron	Founding Partner	Union Duval	14/03/17	30 min	CDO panel preparation Digitalization + data & analytics strategy
Cédric Cauderlie	Co-founder	Mountainview	23/03/17	30 min	Digitalization + data & analytics strategy
Annick Deseure	Customer intelligence manager	Mediahuis	11/04/17	1 hour	CDO panel preparation, organizational design, applications and biggest challenges + solutions

Table 7: Planning of interviews
Source: Author

8.0 Full sample descriptive statistics

8.1 Targets

8.1.1 The importance of data & analytics

Data & analytics are crucial for the success of companies. More than 70% of respondents state that data & analytics are extremely important or very important to their organization, as depicted in figure 22. Interestingly, there appears to exist a strong relationship between the annual revenue of a company and the importance of data & analytics as shown in figure 23. More than 90% of companies with an annual revenue above €1 billion rate data & analytics as very or extremely important to their organization compared to just 40% of firms with an annual revenue of less than €10 million. Many other scholars have also noted the existence of this relationship (Neef, 2015; Ogbuokiri, Udanor, & Agu, 2015; Paskach, & Johnston, 2017; Simon, 2013). Despite an apparently lower interest in data & analytics and the different nature of this segment of companies many authors agree that SMEs can greatly benefit from the implementation of data & analytics in their processes (Donnelly, & Simmons, 2013; Paskach, & Johnston, 2017). Often mentioned reasons why SMEs lag behind are, among others, a lack of awareness of the potential of data & analytics, having too little data, lack of budget and a lack of skill (Sen, Ozturk, & Vayvay, 2016). However, because of the inherently higher flexibility of SMEs many of these issues can be overcome relatively easily (Simon, 2013; Sen, Ozturk, & Vayvay, 2016). For a further discussion of data & analytics strategies for SMEs we refer to Kowalke, (2016) and Marr (2015).

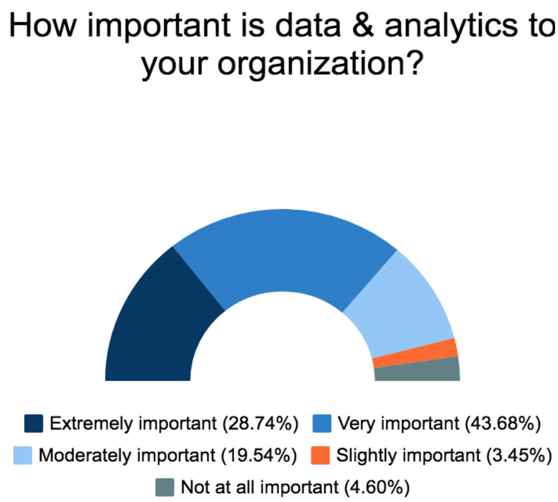


Figure 22: The importance of data & analytics
Source: Author

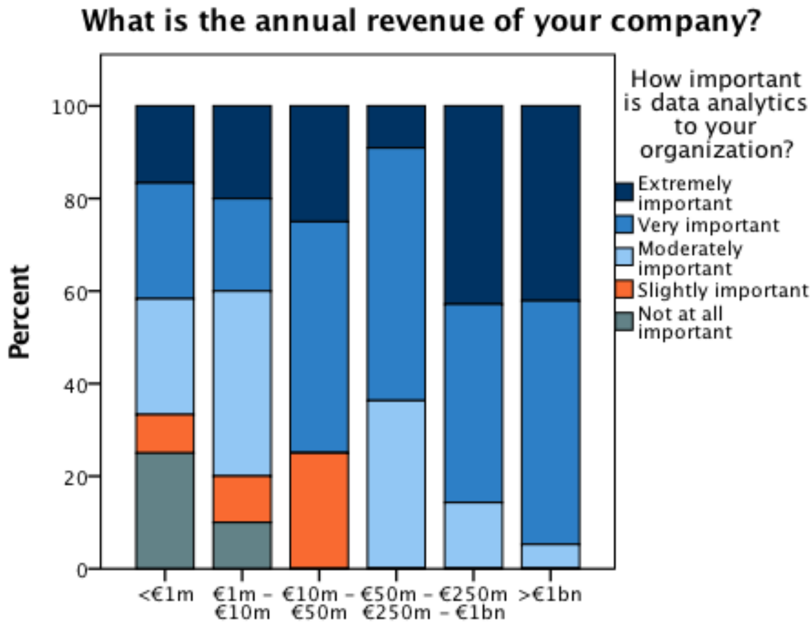


Figure 23: The relationship between the importance of data & analytics and annual revenue
Source: Author

Because of the modest sample size we cannot make any statements about the relative importance of data & analytics across industries as the number of indices per industry is often very low. Research by Accenture (2016) across six industries shows that the financial services industry considers data & analytics to be the most important, followed by communications and healthcare as shown in figure 24. (Note that the absolute percentages of figure 24 cannot be compared to the results of our data & analytics maturity survey as Accenture (2016) used a Likert scale with just four categories.)

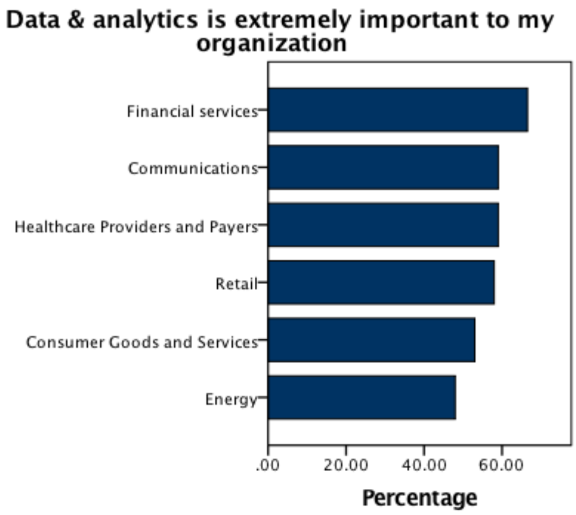


Figure 24: The importance of data & analytics across industries
Source: Accenture

8.1.2 The data & analytics plans

There are many reasons why businesses are implementing a data & analytics strategy, as depicted in figure 25. The three most mentioned reasons are: increasing revenue, improving decision-making, and better understanding the customer. These results reflect the findings of EY & Forbes Insights (2015). The divergent nature of these reasons underlines the great expectations companies have in data & analytics as discussed in chapter 2. Note that offensive data & analytics applications are mentioned far more often. The defensive applications (to improve risk & compliance, to improve the management of existing data and to increase cyber security) all rank very low on the list.

Few companies succeed at implementing all proposed data & analytics projects. In total, only 52% of desired data & analytics outcomes are already translated into noticeable business outcomes. This number corresponds to Marr (2016) who estimates that 50% of all data & analytics projects fail to deliver. The companies in the sample appear to be very good at leveraging data & analytics to improve internal efficiency and cut costs. More than 86% of companies that implement those projects are already witnessing a noticeable impact. Data governance comes in second place with almost 70% of those projects already generating noticeable differences. Thirdly, 68% of firms that implement data & analytics for improved decision making are already noticing benefits. On the contrary companies seem to have the most difficulties with leveraging data & analytics to monitor competitor behavior. Less than 30% of companies that are implementing a data & analytics strategy for this reason are already witnessing noticeable impact. Other major difficulties are observed at monetizing data and leveraging data & analytics as a competitive differentiator. Interestingly, companies seem to be better at realizing their defensive data & analytics strategies. Almost 64% of those initiatives are already being translated into noticeable impact versus only 50% of the offensive plans. Notably, less than 10% of firms indicate not to witness any noticeable impact on any domain.

8.1.3 Impacted business units

Figure 26 shows in which business-units the companies in the sample plan to realize the abovementioned ambitions. Marketing, sales and finance are the departments were data & analytics are used most often.

8.1.4 Implementation challenges

Figure 27 depicts the main reasons why many companies fail to turn their data & analytics strategies into noticeable business outcomes. The three most mentioned issues are: lack of skill, organizational structures and organizational culture. It is striking that these top reasons are all related to human and organizational characteristics. This observation confirms CA Technologies

(2015), EY & Forbes Insights (2015) and our hypothesis from chapter 3 stating that human and organizational challenges are harder to solve than technological issues. Technological issues like low data quality or data security issues appear to be less of a problem. Interestingly, the lack of C-level support, a common mentioned issue in the literature (KPMG, 2016) appears to be overcome as it is only mentioned by 2% of respondents.[3]

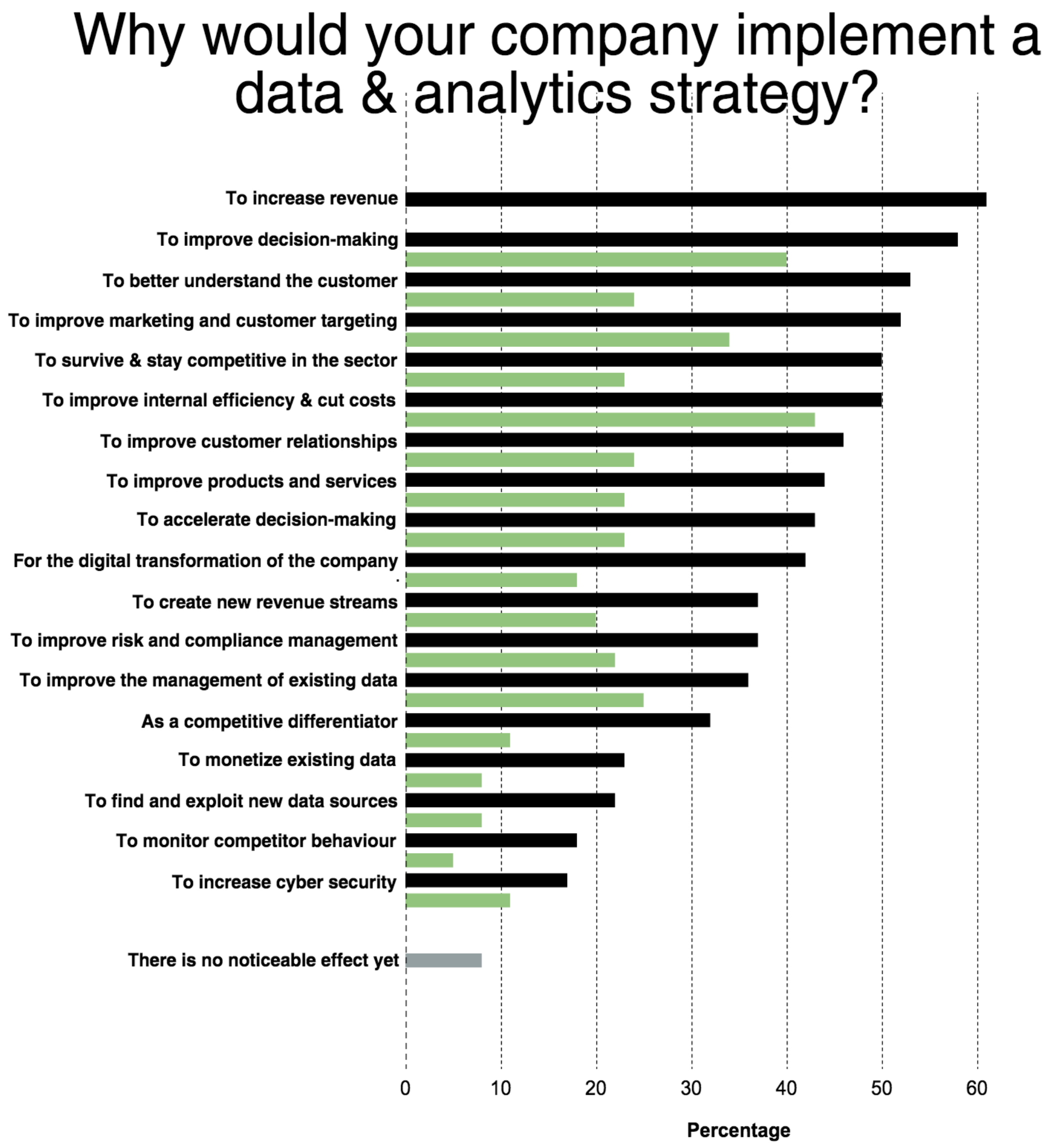


Figure 25: Data & analytics plans versus impact. Black bars indicate the plans, the green bars indicate on which of those plans a noticeable impact is already realized. Note that data about the number of realized plans for increased revenue was not collected due to an error in the online survey.
Source: Author

[3] Note that more than a third of the respondents hold a C-level position. Their answer to the C-level support question is highly subjective.

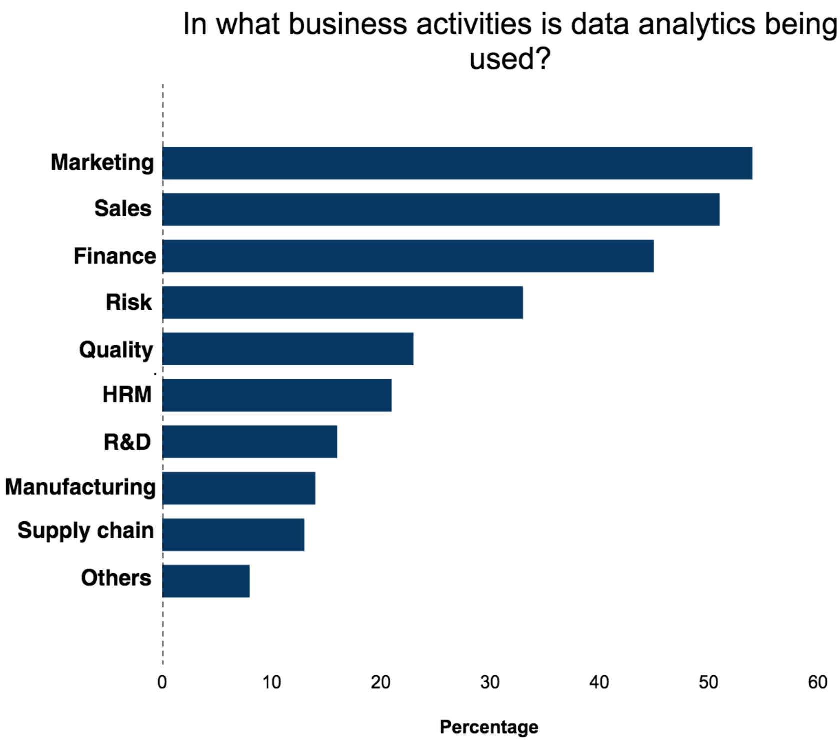


Figure 26: The business activities in which data & analytics are being used
Source: Author

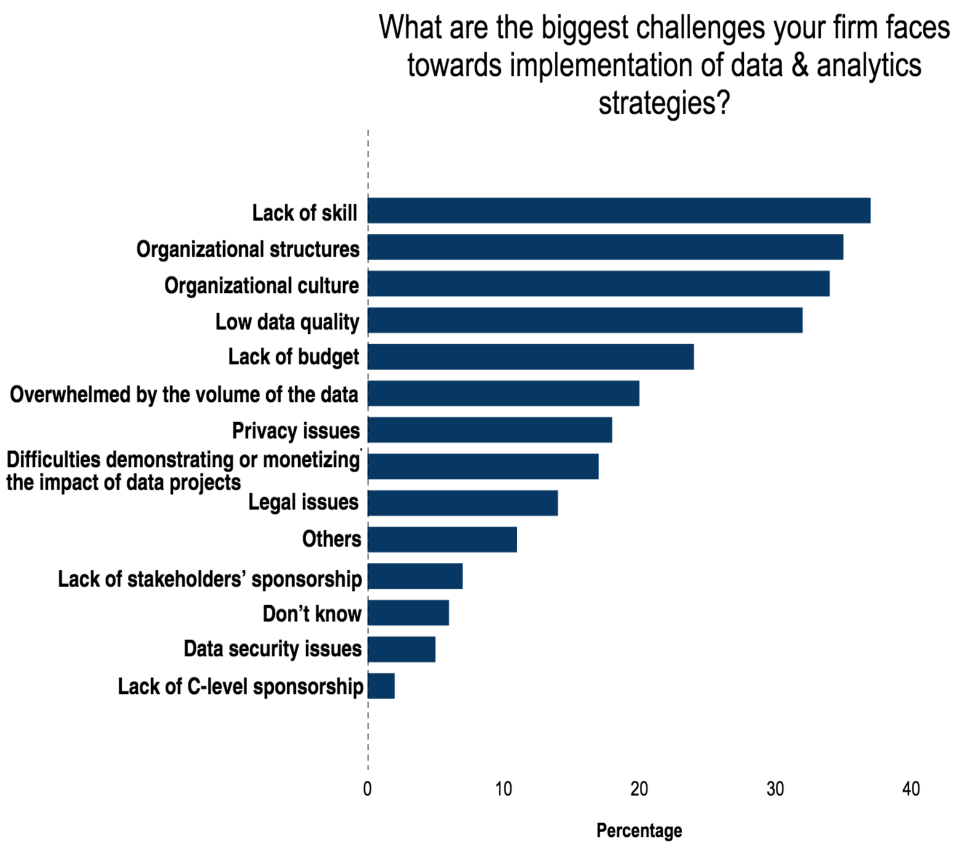


Figure 27: The biggest implementation challenges ranked by prevalence
Source: Author

Figure 28: The most mentioned solutions to the discussed implementation challenges
Source: Author

8.1.5 Solutions

Figure 28 shows how companies plan to tackle the abovementioned challenges. The top three methods are: specific data & analytics projects to prove value and effectiveness, the implementation of new technologies and external consulting. This first solution is an often-mentioned way to initiate a transformation of the company culture and to get more sponsorship for data & analytics from both C-level managers and other employees (Stubbs, 2014). As many companies clearly are struggling with company culture issues it is a logical solution. Surprisingly, the implementation of new technologies is the second most mentioned solution to the implementation challenges. This seems to be counterintuitive, as technological issues seem to be much less critical. The third most mentioned solution is external consulting, which reflects the results of Accenture (2014) who found that 95% of companies that are successful in their data & analytics initiatives received external help. By leveraging external consulting companies they are able to scale their data & analytics platforms more quickly and cheaply than by in-house development (McKinsey 2013).

8.1.6 Data & Analytics investments

In this section we zoom in on additional data & analytics investments. At the moment of crystallization of the survey this was expected to be one of the major solutions to the implementation challenges (McKinsey, 2016). This seemed not to be the case as a larger budget is only the eight-mentioned solution but the in-depth analysis of this topic nevertheless yields interesting results. Figure 29 confirms the significant interest of companies in data & analytics. More than half will make additional data & analytics related investment during the course of 2017. Furthermore, only about 1 in 5 say they do not plan on making additional investments. In addition, research by NewVantage Partners (2016) shows that the percentage of firms planning to invest more than $50 million in data & analytics quintupled between 2014 and 2017 as shown in figure 30, which further underlines the significant interest in data & analytics. We again notice a difference in data & analytics interest between small and large companies as shown in figure 31. Of the firms with an annual revenue over €1 billion about 70% plans additional data & analytics investments within the next year. For firms with an annual revenue below €1m only 20% has this ambition.

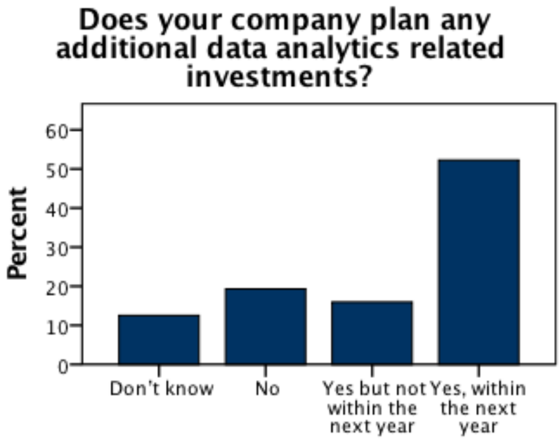


Figure 29: Planned data & analytics investments
Source: Author

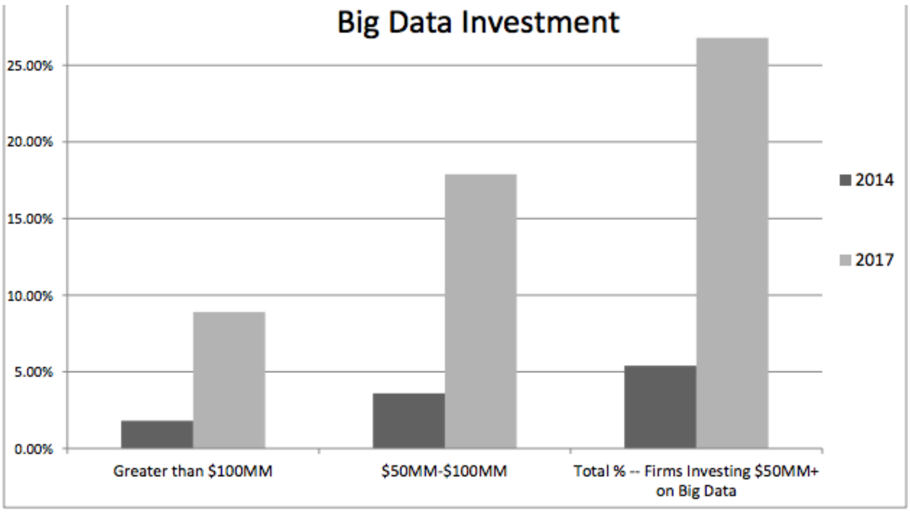


Figure 30: The percentage of firms investing in data & analytics in 2014 and 2017
Source: NewVantage Partners (2016)

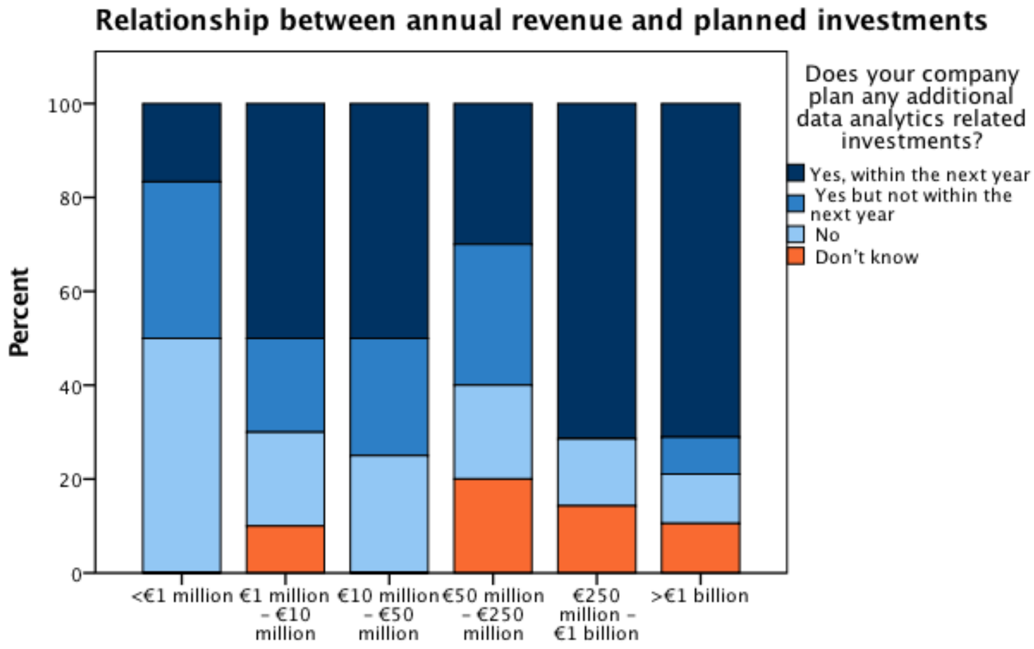


Figure 31: The relationship between annual revenue and planned investments
Source: Author

8.2 Leadership

8.2.1 C-level support for data & analytics

As discussed in the introduction, there is great interest in data & analytics among executives. The survey data confirms this statement as 27% of respondents strongly agree and 36% agree with the statement: "there is strong C-level support for data analytics projects" as depicted in figure 32.[4]

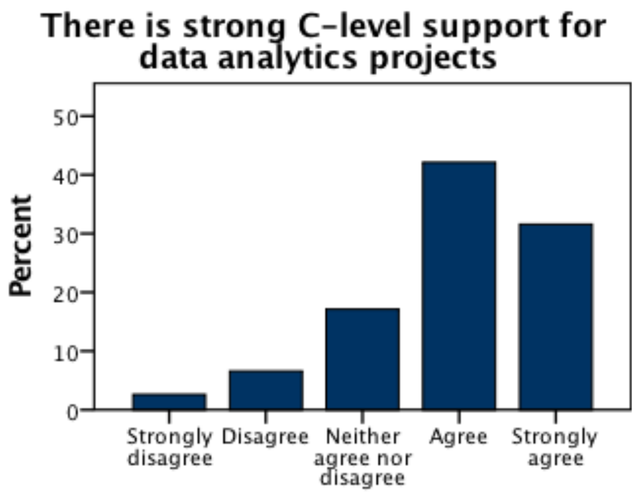


Figure 32: C-level support for data & analytics in the sample
Source: Author

[4] Note again that a large portion of the respondents are C-levels themselves.

8.2.2 Leader of data & analytics initiatives

Despite the increased interest in data & analytics related C-level positions most initiatives are currently still being led by the more traditional executives, which is somewhat surprising. The Chief Operating Officer, the Chief Executive Officer and the Chief Marketing Officer are all frequently mentioned to be in charge of the data & analytics projects. By analyzing the prevalence of the specific data & analytics related C-level positions we find the same ranking as we obtained by the research discussed in section 6.2.2: the Chief Data Officer is the most common position followed by the Chief Digital Officer and the Chief Analytics Officer. Note that the prevalence of these positions in our sample is much lower than the results from the literature research. Only 8% of firms in the sample employ a Chief Data Officer, only 4% a Chief Digital Officer and 2% a Chief Analytics Officer. Possibly geographical factors play a role as 85% of Chief Data Officers are employed in the US or the UK (IBM, 2014) and many firms in our sample are Belgian companies. Furthermore, many projects seem to be led by non-C-level positions. Both the data scientist and data analyst roles are frequently mentioned as being in charge of data & analytics initiatives as depicted in figure 33.

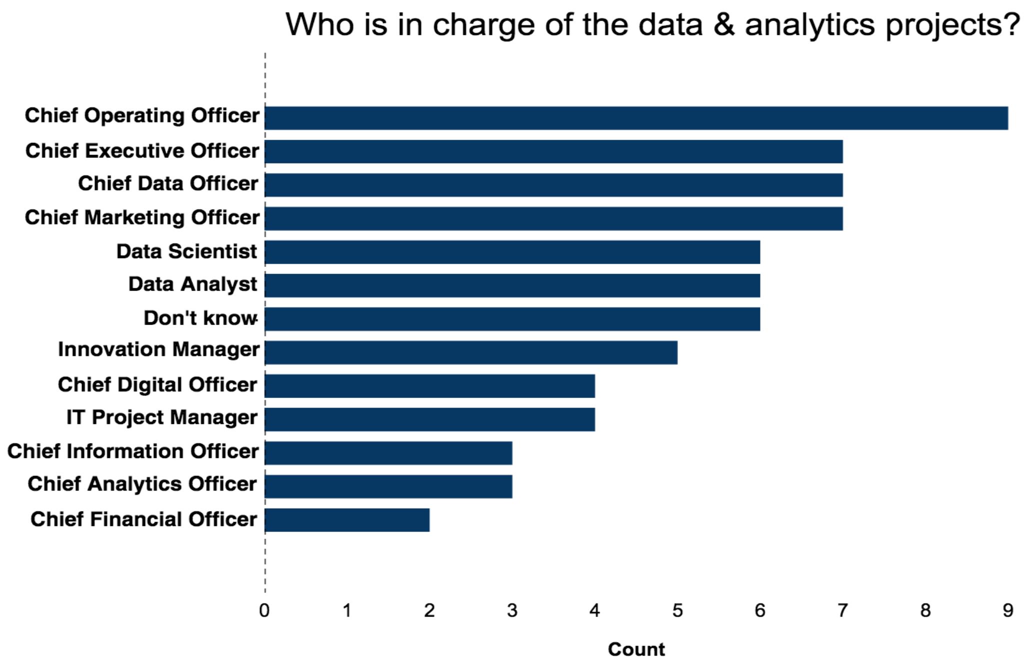


Figure 33: The leader of the data & analytics initiatives
Source: Author

8.2.3 Responsibilities of the data & analytics related positions

In this section we analyze the responsibilities of the data & analytics related positions as indicated by the survey respondents. Note that we will not cover the responsibilities of the traditional C-level roles, as this is not the focus of this project. As the number of these data &

analytics related positions in the sample is very low, the results from the next section should be generalized with great care.

The results from table 8 are somewhat surprising. To start, the responsibilities of the Chief Data Officer are focused on offensive applications: analytics and generating value. The management of data in general (defense) seems to be less important as data governance, data integration, data quality and data security are not often mentioned as a responsibility. It is also interesting to observe that the general influence of this position seems to be rather low. Only 14% of Chief Data Officers participate on the executive board and help determining the strategic direction of their firm.

	Chief Data Officer	Chief Digital Officer	Chief Analytics Officer	Data Scientist	Data Analyst
Data collection	43%	50%		67%	50%
Data governance	29%	25%	50%	33%	67%
Data exploitation	43%	25%		50%	67%
Data security	14%	25%		33%	17%
Data analytics	86%	50%	50%	83%	83%
Data integration	14%			50%	67%
Monitoring data quality	14%	50%		67%	67%
Translating analysis into value	71%		100%	83%	50%
Finding new data sources	43%	50%		17%	33%
Managing the analytics department	43%		100%	17%	50%
Monitoring data analytics initiatives	43%	25%	100%	17%	17%
Finding new business opportunities	14%	25%	50%	17%	17%
Participating on the executive board	14%	50%			
Establishing a data culture in the organization	29%	75%		17%	17%
Democratizing data tools across the organization	14%			17%	33%
Leading workshops/seminars on data analytics	14%	25%	50%	17%	33%
Helping to determine the strategic direction of the company	14%	50%		17%	17%
Improving the online presence of the firm	14%	25%			17%
Communicating the firms data analytics strategy internally and externally	29%	25%	50%	17%	17%
Change management	29%	50%			17%
Sponsoring digitalization or automation	29%	100%		17%	
Develop ways to attract and retain highly skilled data talents	14%	25%			
Coordinating data related investments	71%	50%			33%
Others					
Don't know	14%				

Table 8: The responsibilities of data & analytics related positions. The three most mentioned responsibilities are highlighted.
Source: Author

In the survey data the Chief Digital Officer has much more influence than the Chief Data Officer as 50% of the Chief Digital Officers participate on the executive board. The Chief Digital Officer is a transformational leader that sponsors digitization, establishes a data culture and helps determining the strategic direction of the company. Surprisingly, Chief Digital Officers are also responsible for many rather technical areas like data collection, analytics and data quality. This description of the Chief Digital Officer is somewhat broader and more technical than the definition from section 6.2.2.

The Chief Analytics Officers from our sample are mainly focused on managing the analytics department, monitoring analytics initiatives and turning analytics into value. These positions are very technical: they are not involved in change management, establishing a data culture or participating on the executive board.

8.2.4 Educational background of the data & analytics related C-level positions

Table 9 depicts the educational background of the data & analytics C-levels in the sample. The difference in education of the people in these positions is notably large. Firstly, the educational variance is the highest for the Chief Data Officer role as all four major domains depicted in table 9 are represented. Secondly, all CAOs have a background in statistics. However, it is worth repeating that there were only 2 CAOs in the data so we cannot expect this result to be generalizable. Thirdly, making conclusions about the background of the Chief Digital Officer is not straightforward as 50% of respondents indicate they do not know the background of their Chief Digital Officer.

	Chief Data Officer	Chief Digital Officer	Chief Analytics Officer
Math/physics	14%	25%	
IT/computer science	14%		
Data Science/statistics	29%		100%
Economics/business	29%	25%	
Others			
Don't know	14%	50%	

Table 9: The educational background of data & analytics related positions. The most mentioned educational background is highlighted.
Source: Author

8.2.5 Presence of the data & analytics related C-level positions

In this section we investigate how long the data & analytics related C-level positions have been in place in the companies in the sample. Overall we can conclude that this timespan distribution as depicted in figure 34 is relatively uniform: no distinct hiring trends can be discovered. Table 10 shows a breakdown of these timespans over the three different C-level positions. All CAOs were hired more than 2 years ago, half of the Chief Digital Officers were hired less than a year ago and most of the Chief Data Officers were hired between 4 and 1 years ago. Note again that these results are not generalizable because of the very small number of C-levels in the data.

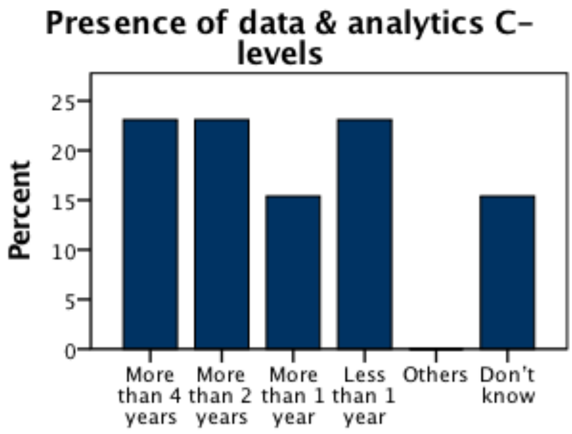


Figure 34: The presence of the data & analytics related C-level positions in years. 2016=t0.
Source: Author

	Chief Data Officer	Chief Digital Officer	Chief Analytics Officer
Less than 1 year	14%	50%	
More than 1 year	29%		
More than 2 years	29%		50%
More than 4 years	14%	25%	50%
Don't know	14%	25%	

Table 10: The presence of data & analytics related positions. Most mentioned timespan is highlighted. 2016=t0
Source: Author

8.2.6 Recruitment intentions

Figure 35 shows the recruitment intentions of the companies in the sample that do not yet have a data & analytics related C-level officer. Surprisingly 67% of these firms do not plan on hiring a data & analytics related C-level officer. Only 10% plan to recruit such an officer within the next year.

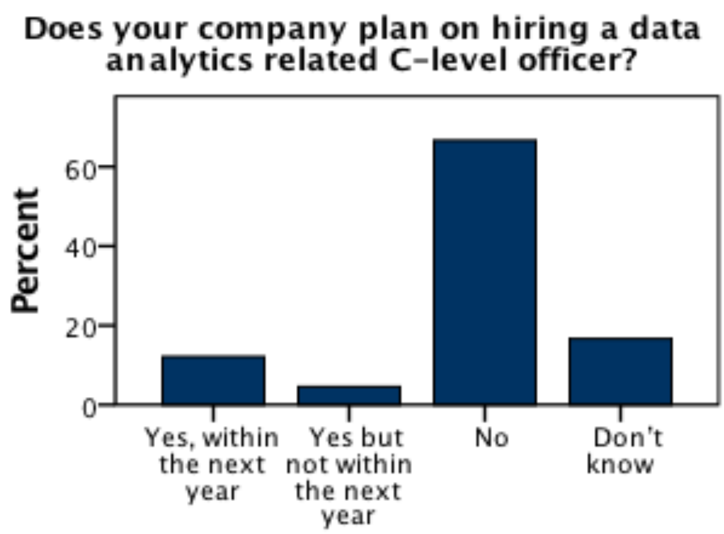


Figure 35: Recruitment intentions of the companies without a data & analytics related C-level
Source: Author

8.3 Data

8.3.1 Data collection

Structured data formats like tables and records are still the most widely collected format as almost 80% of the companies in the sample indicate to gather this kind of data. Roughly 50% of companies collect text data and a third collect clickstream data as shown in figure 36. Note that these figures say nothing about which data sources are analyzed or contribute to decisions. In further research this consideration should be taken into account. The vast majority of data collected is originating from back office systems as depicted in figure 37. Other often mentioned data sources are: website, point of sale, 3rd parties and social media.

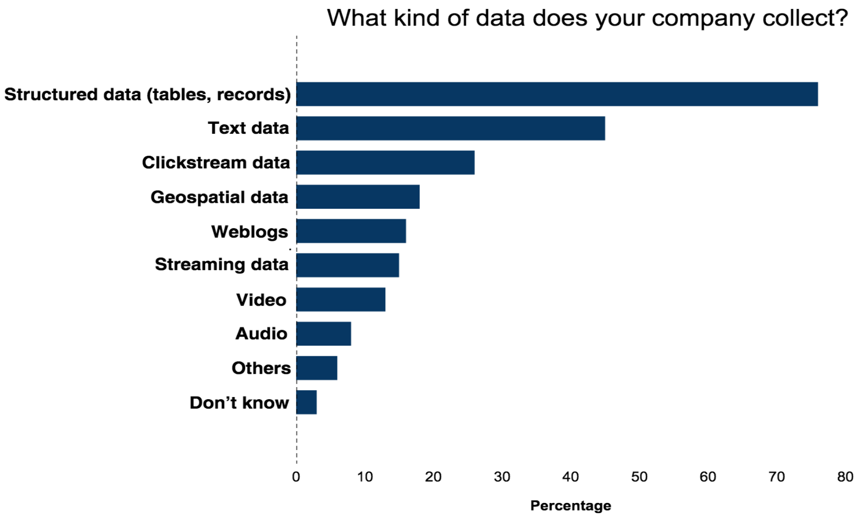


Figure 36: Collected data formats
Source: Author

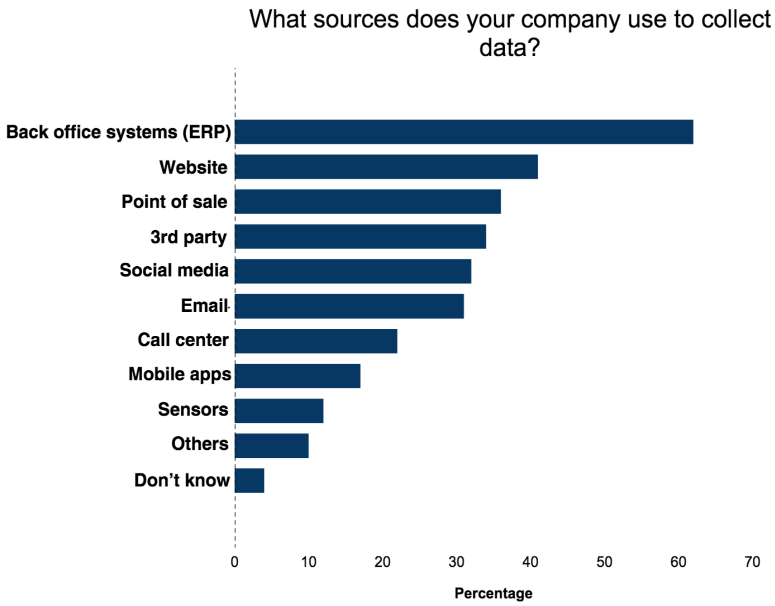


Figure 37: Utilized data sources
Source: Author

8.3.2 Data governance

The majority of firms have a clear view of the volume of data stored and gathered as depicted in figure 38. However, the roles and responsibilities for data governance are not yet clearly defined in most organizations as depicted in figure 39. This observation follows the statement from chapter 2 that most organizations have the necessary data & analytics technology implemented but have not yet thought about all its organizational implementations. Furthermore many organizations seem to have a progressive data culture as the majority indicates that all employees have easy access to data. The largest data governance related issue seems to be the quality of the collected data. Less than 10% have an overview of data quality at the enterprise level and less than 5% have data quality KPIs in place (see figure 42).

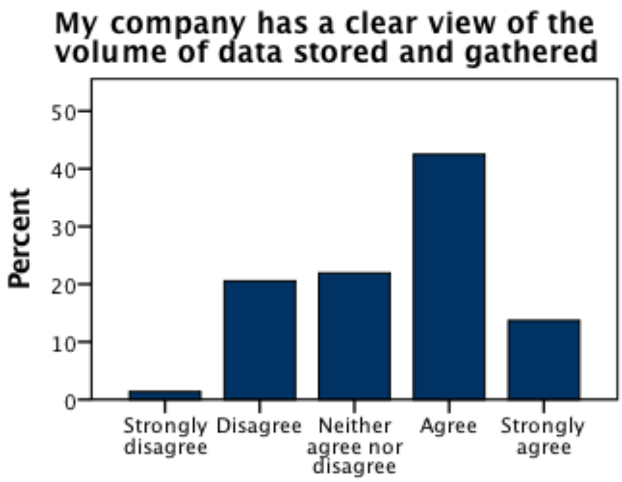


Figure 38: My company has a clear data overview
Source: Author

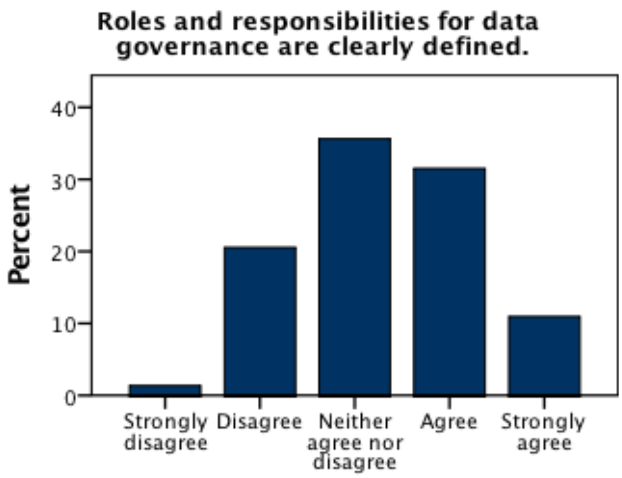


Figure 39: Roles and responsibilities for data governance are clearly defined.
Source: Author

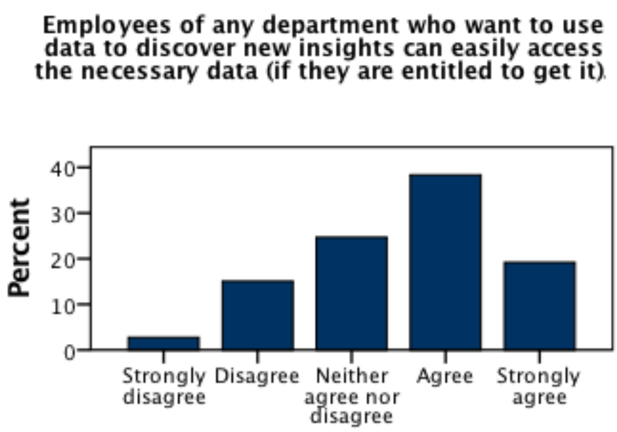


Figure 40: In my company, employees have easy access to data
Source: Author

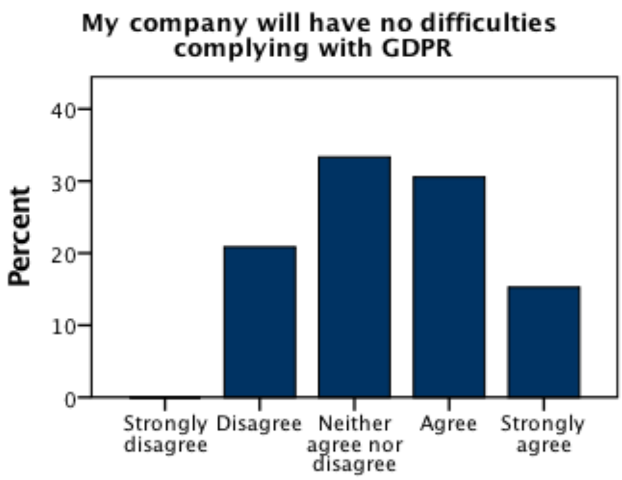


Figure 41: My company will have no difficulties complying with GDPR
Source: Author

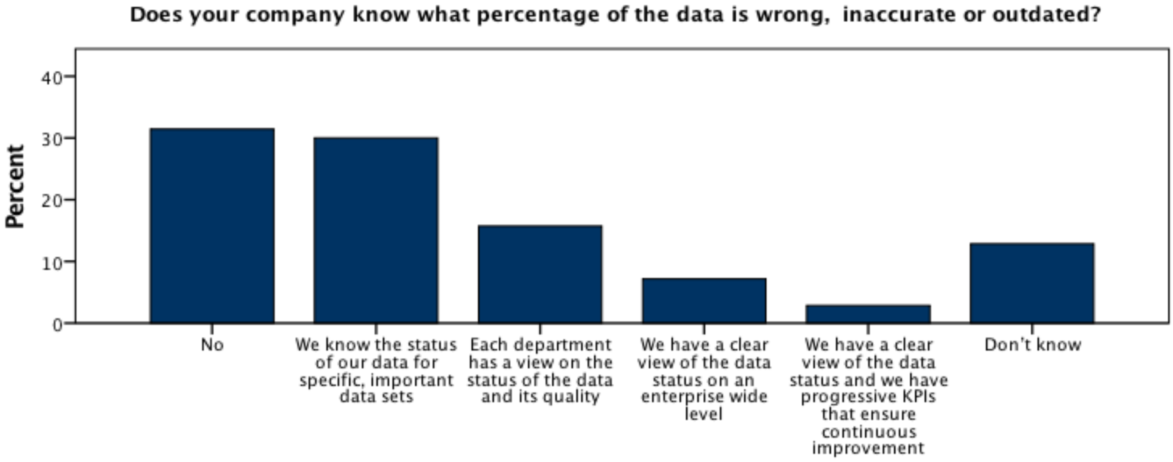


Figure 42: Does your company know what percentage of data is wrong, inaccurate or outdated?
Source: Author

8.4 Analysts

8.4.1 Data-driven decision making

The survey results suggest that most firms are aware of the importance of data-driven decision making (as discussed in chapter 1). 17% of all firms indicate data & analytics are involved in all decisions and 46% indicate data & analytics are involved in most of the decisions as depicted in figure 43. Interestingly, only 43% of companies suggest they have feedback processes in place to assess the quality and accuracy of the analyses they use as an input for their decisions as shown in figure 44.

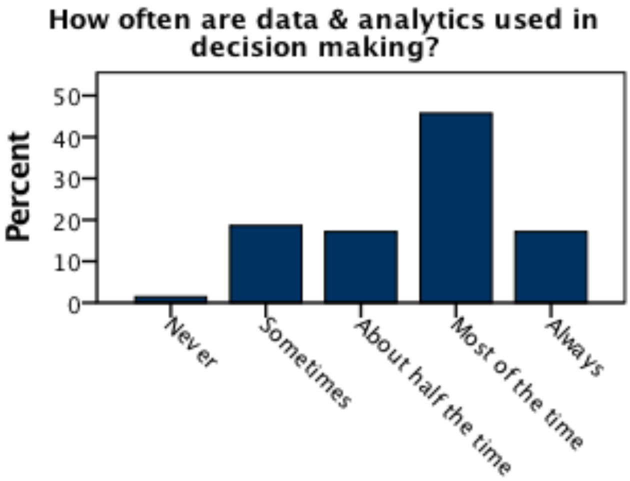


Figure 43: How often are data & analytics used in decision making?
Source: Author

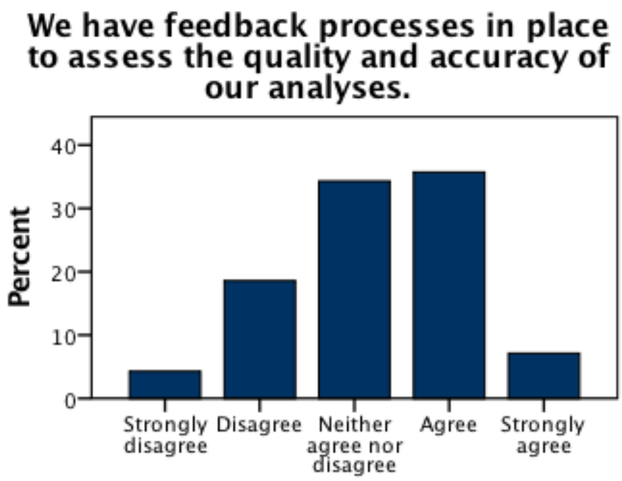


Figure 44: We have feedback processes in place to assess the quality and accuracy of our analyses
Source: Author

8.4.2 Analytics teams

As depicted in figure 45, analytics teams tend to be diversified. Business analysts are the most represented roles followed by data scientists, IT experts and system architects. In the category "others" mainly consultants and marketers are mentioned. In section 8.1.3 we discussed that lack of skill is the most mentioned implementation challenge. Figure 46 shows what skills are in general hardest to get by. Skills that bridge IT and business are most mentioned followed by analytical skills and creativity.

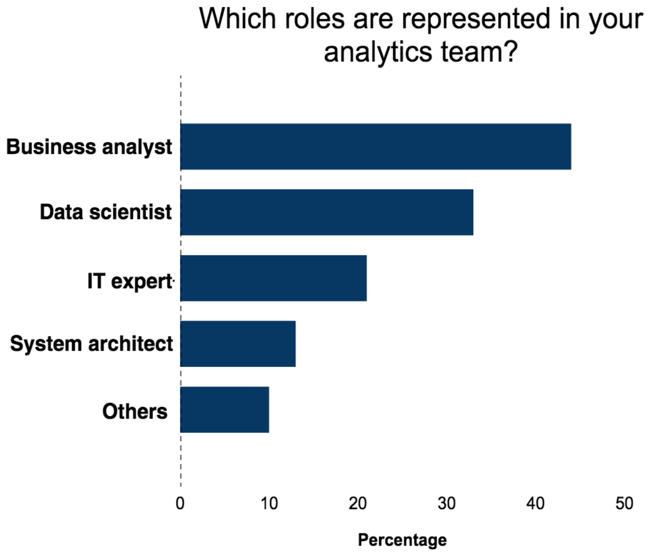


Figure 45: Which roles are represented in your analytics team?
Source: Author

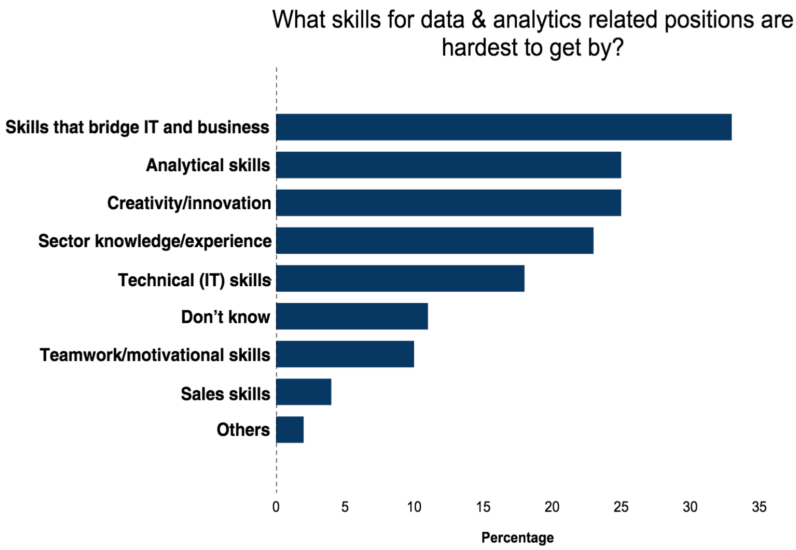


Figure 46: What skills for data & analytics related positions are the hardest to get by?
Source: Author

McKinsey (2013) mentions that companies often fail in the implementation of data & analytics because the front-line employees who use models or their results (analytical amateurs in the DELTA jargon) are insufficiently trained. McKinsey (2013) further suggests that 50% of the data & analytics budget should go to training. Our survey data confirm this concern as less than 20% indicate to spend a significant part of their investments on training their analytical amateurs.

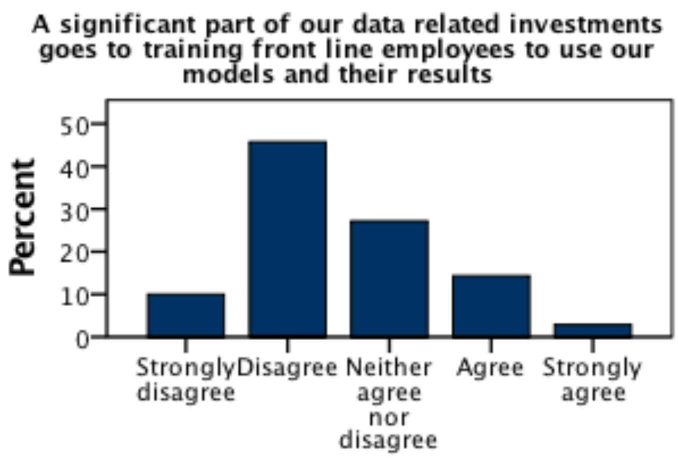


Figure 47: A significant part of our data related investments goes to training front-line employees to use our models and their results
Source: Author

8.5 Enterprise (organizational design)

The discussion of the enterprise dimension is much shorter than the other dimensions because the nature of this topic is much more difficult to investigate on the basis of a survey. Almost 35% of the companies in the survey rely on ad-hoc analytics by individual employees for realizing their data & analytics strategy as shown in figure 48. Around 20% organize data & analytics at the department level and 13% at the enterprise level. Lastly, 25% of the companies report to have both automated analytics processes at an enterprise level and experimental teams to innovate in analytics. The qualitative research revealed most companies are using a center of excellence model to organize data & analytics. Oftentimes, business units are involved in very specific data & analytics initiatives that require in-depth knowledge of the respective business domain. A center of excellence design allows for this kind of far-reaching specialization while still leveraging the advantages of a centralized design as discussed in section 6.2.5.

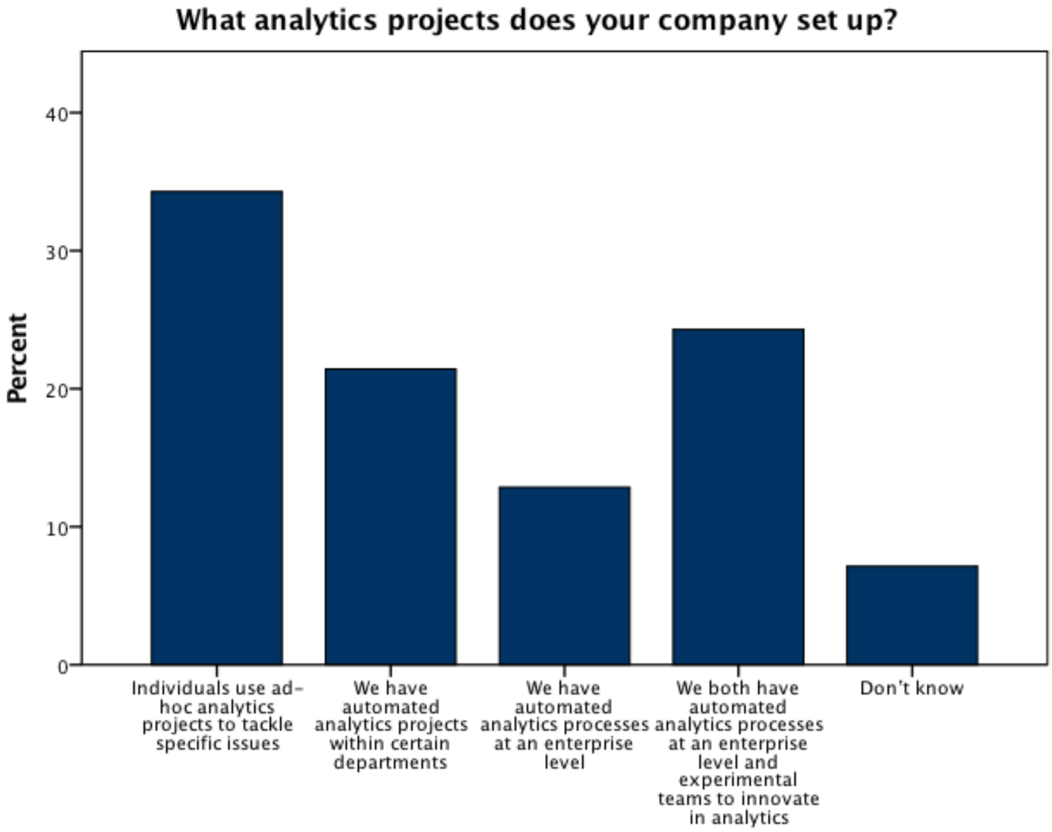


Figure 48: What analytics projects does your company set up?
Source: Author

9.0 Maturity clusters

In this chapter we will introduce the k-means clustering that was performed on the survey data to generate an inductive maturity model. First we will briefly introduce the methodology and its pitfalls. Next, we will give an overview of the characteristics of each cluster and compare this to the theoretical description of these clusters in the DELTA model.

9.1 Methodology: k-means

The organizations in the survey data were clustered into 5 maturity clusters via the k-means algorithm. K-means is an often-used unsupervised segmentation algorithm that clusters data around k cluster centers (centroids). A further discussion of this algorithm is included in extension 5. As the DELTA model has 5 maturity stages we opted for k=5. Convergence was reached after 12 iterations.

9.2 Data preparation

The answers on the 37 questions from the survey were encoded into dummy variables for nominal data and into thermometer variables for ordinal data (Martens, 2017). This resulted in a dataset of 217 variables. The k-means output of this dataset was not meaningful due to methodological issues discussed that will be discussed in the next section. Therefore we introduced nine additional continuous variables (plans, impact, challenges, solutions, data sources, data formats, analytics techniques, impacted business activities and teams). These variables are simply the total number of elements in each of these categories per survey respondent. For example, a company that indicated they collect data from 5 sources has value 5 for the variable "data sources". If we did not include these additional variables two companies that were both very mature in data & analytics but used totally different data sources were not clustered together. After the inclusion of these variables the clustering became highly intelligible. Lastly, the general company statistics (industry, annual revenue, number of employees and age) were left out because we only want the clustering to be influenced by maturity indicators.

9.3 Critical remarks

It is important to stress some major methodological issues of the analysis we conducted. Firstly, the questions that were used to assess data & analytics maturity are highly arbitrary. Although these questions were based on the dimensions of the DELTA model (Davenport, Harris, & Morison, 2010) and inspired by other, similar maturity questionnaires (Accenture, 2014; CA Technologies, 2015; EY & Forbes Insights, 2015; Lismont, Vanthienen, Baesens, & Lemahieu, 2017; NewVantage Partners LLC, 2016) another set of questions could have led to a different clustering.

Secondly, the number of defined clusters in the k-means algorithm (here k=5) is highly arbitrary as well. Although this number is again based on the number of stages in the DELTA model there are other maturity models that define 3, 4 or even 6 stages, as discussed in section 5.3. A different choice for k would have led to a different clustering. When performing a TwoStep cluster analysis to determine the optimal number of clusters only two clusters were deemed optimal. This conclusion was not satisfactory, as a maturity model with only two stages would not provide firms with a sufficiently specific and detailed roadmap.

Thirdly, there are technical issues with applying a k-means clustering to our survey data. As described in section 9.2, many of the variables in our data set are binary. The k-means algorithm by default uses the Euclidian distance as a similarity measure. Calculating the Euclidian distance between two binary vectors comes down to simply counting the variables on which the two vectors do not coincide (IBM, 2016). Furthermore, the k-means algorithm uses means to calculate the centroids. Having a mean value between 0 and 1 for a binary variable is not defined, as its scale is not continuous (Bradshaw, n.d.; Khan, 2016). These problems could lead to unmeaningful clusters and the algorithm failing to reach convergence after 10 to 20 iterations. These problems could possibly be overcome by using a different similarity metric like the Gower distance (Gung, 2015) or using the k-modes algorithm that uses modes instead of means to define the centroids (Huang, 1998). Another strategy could be to first run a principal component analysis on the data and then run k-means on the (continuous) factor scores (IBM, 2016). The first two options were disregarded because the functionality was not included in the statistical package that was used for writing this thesis. The PCA option was tested, but because the number of significant principal components was high (31) the new variables were extremely hard to interpret.

Fourthly, the k-means algorithm implicitly assumes all variables to be equally important as they all contribute in the same extent to the distance calculations. Because of this, the answer to a question assessing a minor data governance issue contributes the same weight to the clustering as a major question about data & analytics in general. Therefore it could be a good idea to give different weights to the more general and important questions and less weight to the detailed questions. However, this again is highly arbitrary. Which questions are the most decisive in determining data & analytics maturity? What weights should be used?

All these methodological issues deserve a great deal of attention and should always be kept in mind throughout the rest of the analyses. However, we believe the obtained clustering is nevertheless valuable and useful because of the following reasons.

- The clustering is highly intelligible as will be discussed later
- The clustering resembles the theoretical maturity stages of the DELTA model
- Convergence was reached after 12 iterations
- "All models are wrong, some are useful" (Box, 1976). We believe our clustering can be of help for companies willing to increase their data & analytics maturity.

9.4 High-level cluster variables

Table 11 summarizes the cluster centers for each of the obtained clusters. Note that only some high-level variables are shown, the full output can be found in appendix 2 and will be discussed throughout the next chapters. More information on the variables depicted in table 11 can be found in extension 6. Furthermore, it is worth stressing that the results shown in table 11 are only average values for the depicted variables for each of the clusters. This does not mean all the instances in the cluster necessarily have this value for these variables.

Dimension		A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
	Percentage of instances	19%	32%	17%	23%	9%
	Mode annual revenue **	<€1m	>€1bn	>€1bn	>€1bn	>€1bn
	Mode founding year *	1990-2010	<1950	<1950	<1950	<1950-1990
Targets	Importance of data & analytics ***	Moderately important	Very important	Extremely important	Very important	Extremely important
Targets	Self indicated strategy maturity ***	Low	Low	Medium	Medium	High
Targets	# Plans ***	2	4.90	10.40	10.48	13.88
Targets	# Impact ***	1	2.72	2.87	6.24	8.88
Targets	# Impacted business activities ***	0.29	2.72	1.87	4.05	7.13
Targets	# Challenges ***	1.06	2.76	3.47	2.95	3
Targets	# Solutions ***	0.82	2.55	2.67	3.52	2.50
Data	# Data formats ***	0.24	2.34	1.40	2.48	6.63
Data	# Data sources ***	0.18	3.21	1.80	4.62	6.00
Leadership	Strong C-level sponsorship ***	Neither	Agree	Agree	Agree	Strongly agree
Enterprise	Data analytics projects ***	Don't know	Medium	High	Medium	High
Analysts	# Analysis techniques ***	0.12	2.79	0.87	3.19	5.50
Analysts	Team diversification ***	0.12	1.76	0.33	2.43	4.25

Table 11: Cluster centers of the 5 generated clusters
Source: Author

The statistical significance of the categorical variables across the groups is assessed via a Chi-squared test. In some cases the expected frequency of one of the cells is below five. In these cases we use the Fisher's exact test to assess the statistical significance. For the interval variables one-way ANOVA was used.

9.5 Cluster Description

Figure 49 depicts the 5 generated clusters on 4 dimensions: number of plans, number of impacted business activities, data & analytics importance and the self-indicated data & analytics strategy maturity. After this first graphical analysis the maturity stages as described in the DELTA model do not seem to fit the data entirely. In the DELTA model maturity increases incrementally over all dimensions as companies move to a higher maturity stage. If we exclude cluster C, this predicted evolution is exactly what we witness in the data. As depicted in table 11, all high-level variables on the 5 DELTA dimensions are without any exceptions an increasing function of maturity. Cluster C is the only cluster for which this behavior is not observed. In what follows we will dig deeper into the description of these 5 generated clusters. During the discussion of the clusters in the next sections it can be a good idea to turn back to chapter 6 from time to time to closely compare the theoretical description of the maturity stages with the descriptions of the obtained cluster. We will focus our attention on the non-linearity of the maturity stages in the data, as the DELTA model does not predict this.

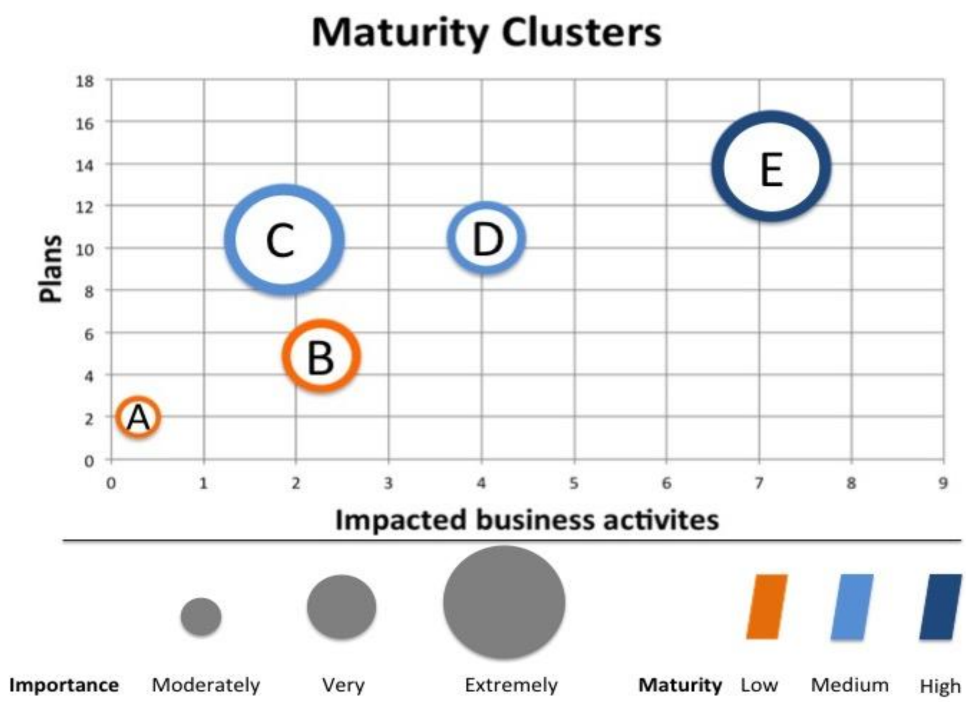


Figure 49: A graphical representation of the generated maturity clusters
Source: Author

9.5.1 Cluster A: Analytically impaired

This cluster contains the least data & analytics mature organizations and it represents roughly one fifth of the total sample as shown in table 11. Data & analytics are only of moderate importance in this cluster (the lowest value of the five clusters) and they indicate to have an ad hoc data & analytics strategy in certain business units (lowest strategy maturity). Companies in this cluster only have minor plans for the implementation of data & analytics in a very limited number of business activities. However, they are already witnessing a noticeable impact of data & analytics on 50% of those plans. Organizations in this cluster seem to be experiencing very few obstacles, which could indicate these companies are focusing on "low hanging fruit". Furthermore, the diversification in data formats, analytics techniques and analytics roles is very low and they only collect data from a limited number of sources. These companies lack strong C-level support for data & analytics. As shown in figure 50, most companies in cluster A are SMEs as 80% of these firms have an annual revenue smaller than €50m. To conclude, organizations in this cluster are small and have very limited data & analytics ambitions. However, this should not immediately be interpreted as an issue. Some of these companies could indeed be totally unaware of the potential of data & analytics and could be facing a great competitive disadvantage in the near future. On the other hand, there are companies in this cluster for which data & analytics will never be of great importance. As shown in table 12, cluster A contains, amongst others, 2 independent artists and a start-up communication firm for which the importance of data & analytics can indeed be negligible. The description of cluster A strongly resembles that of the "analytically impaired" companies, the first maturity stage in the DELTA model.

9.5.2 Cluster B: Localized analytics

Cluster B contains about a third of the sample. For the organizations in this cluster, data & analytics are very important although they only have an ad hoc data & analytics strategy. These firms have quite a few data & analytics plans and are already witnessing noticeable effects on about 55% of those plans. Overall, these organizations score relatively low on the data, analytics and team variables but still significantly better than the analytically impaired companies in cluster A. Companies in this cluster have strong C-level support for data & analytics. It is worth noticing that the average company profile for this cluster is very different from that of cluster A. As the companies in the analytically impaired cluster are mainly relatively young SMEs, the focus in cluster B lies on bigger and older firms as shown in figures 52 and 53. As 16 industries are represented, cluster B has the highest sector variance of all the clusters as shown in table 13. To conclude, companies in cluster B have some data & analytics ambitions in specific business activities but they are still far away from an enterprise wide and mature data &

analytics strategy. This description coincides closely with the "localized analytics" enterprises, the second stage in the DELTA model.

9.5.3 Cluster C: Analytical hubris

Cluster C is a very interesting cluster to analyze, as it is a cluster that is not described in the DELTA model. These companies stress the extreme importance of data & analytics and indicate to have "an enterprise wide data & analytics strategy, but not yet fully aligned with all business units". Firms in this cluster have definitely picked up on the potential of data & analytics as they indicate to have many plans for its implementation. However, the impact they are already witnessing is remarkably low. At this time, they are only able to translate 28% of their plans into noticeable impact. This is by far the lowest impact to plan ratio. This is surprising because they indicate to have "both automated analytics processes at an enterprise level and experimental teams that explore innovative analytic methods": the highest enterprise maturity level. Furthermore, they seem not to have figured out in which business units they wish to achieve this impact as the number of planned impacted business activities is even lower than that of the localized analytics companies from cluster B. Not surprisingly, this is also the cluster with the highest number of implementation challenges. Companies in cluster C score very bad on the amount of data formats, data sources, analysis techniques and data teams compared to their closest competitors in clusters B and D. As shown in figures 54 and 55 cluster C contains mainly large companies that were founded before 1950. Furthermore, mainly financial services companies and human resource firms are occupying this cluster as shown in table 14. To conclude, these companies feel they must catch the data & analytics train to stay competitive. They make ambitious plans, but faced with many challenges they are not able to turn these ideas into value yet. This cluster description does not seem to match any of the maturity stages as defined in the DELTA model. Their self-indicated strategy maturity level is relatively high but they score very badly on the variables necessary to translate this vision into value. Furthermore, it is not clear whether this cluster is more or less mature than the companies in cluster B when all high-level variables are taken into account. We will come back to these observations later in this chapter. We will name cluster C the "analytical hubris" cluster.

9.5.4 Cluster D: Analytical companies

This cluster contains the companies with the second highest overall data & analytics maturity and it represents roughly a quarter of the sample as depicted in table 11. These organizations have the same strategy maturity as the analytical hubris companies and roughly the same data & analytics ambitions. However, companies in this cluster are much more successful in their efforts as 60% of their plans are already delivering a noticeable impact and data & analytics are already implemented in many more business activities. As these companies also have roughly the same annual revenue & founding year distribution as the companies from cluster C, we could interpret cluster D as the matured version of the analytical hubris firms. Interestingly, for companies in this cluster data & analytics are less important than for companies in cluster C (very important vs. extremely important). We will come back to this remark later in this chapter. It is also worth noting that very little SMEs are reaching this stage of data & analytics maturity as only 5% of this cluster has an annual revenue below €50m as shown in figure 56. Moreover, most companies in this cluster are in the financial services sector as shown in table 15. To conclude, the companies in cluster D have a solid data & analytics maturity. They have big ambitions and seem to have the capabilities to translate this vision into business value. The description of the 4th maturity stage in the DELTA model "analytical companies" seems to apply to the firms in this cluster.

9.5.5 Cluster E: Analytical competitors

The companies in this last cluster are the data & analytics hotshots and represent roughly 10% of the sample. This is the only cluster that indicates to already have "a mature and enterprise wide data & analytics strategy with a strong focus on continuous experimentation and improvement". These companies are the most ambitious and at the same time have the highest impact to plan ratio. Furthermore, they have been implementing data & analytics throughout the entire organization as they have by far the highest number of impacted business activities. The other variables already give an indication of how these companies are so successful in their efforts. They collect more data formats from more sources, apply more analytics techniques and have very heterogeneous teams. Moreover, this is the cluster with the strongest C-level sponsorship. Lastly, these companies have a very explicit profile in terms of annual revenue and age: the vast majority has a turnover of more than €1bn and they are all founded before 1990 as shown in figures 58 and 59. These companies are what Davenport, Harris, & Morison (2010) define as "analytical competitors" and they constitute the fifth and final maturity stage.

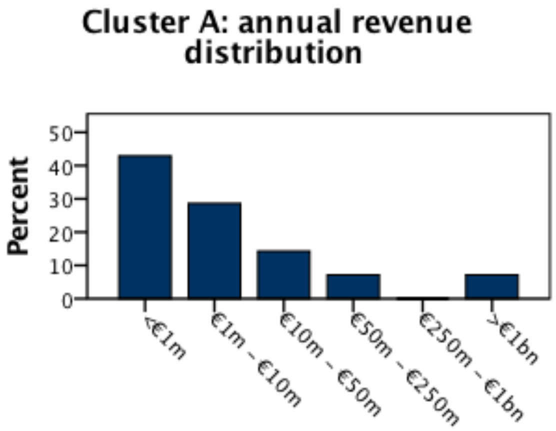


Figure 50: The distribution of annual revenue of cluster A
Source: Author

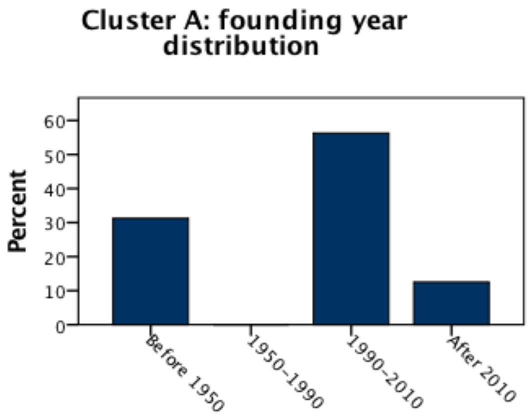


Figure 51: The distribution of founding year of cluster A
Source: Author

Cluster A: sectors	Count
Art	2
Automotive	1
Chemical	1
Communication	1
Financial services	3
Food & Beverage	1
Human resources	1
IT services	2
Research	3
Missing values	2

Table 12: Sector distribution of cluster A
Source: Author

Custer B: sectors	Count
Automotive	1
Chemical	3
Construction	2
Creative Industries	1
Education	1
Entertainment	1
Financial services	4
Food & beverage	2
Health	1
IT services	3
Maintenance	1
Public sector	1
Real estate	1
Recycling	1
Research	2
Web services	1
Missing values	3

Table 13: Sector distribution of cluster B
Source: Author

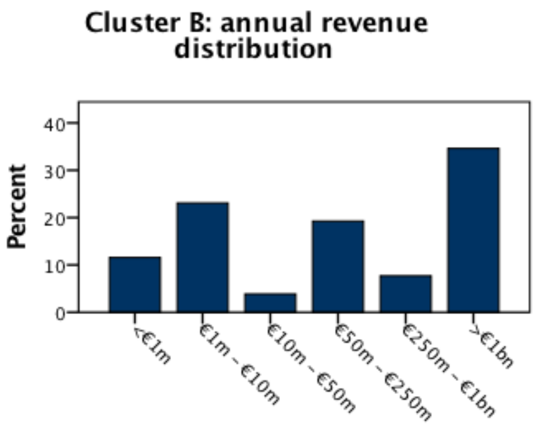


Figure 52: The distribution of annual revenue of cluster B
Source: Author

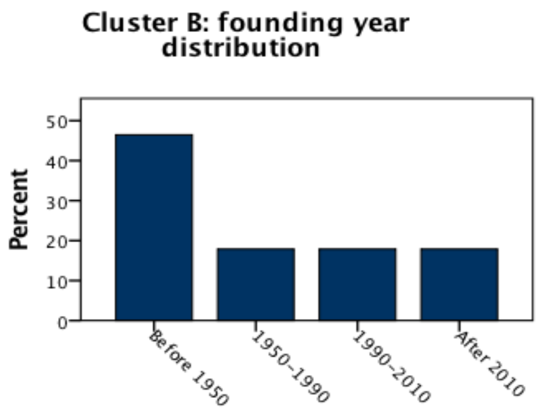


Figure 53: The distribution of founding year of cluster B
Source: Author

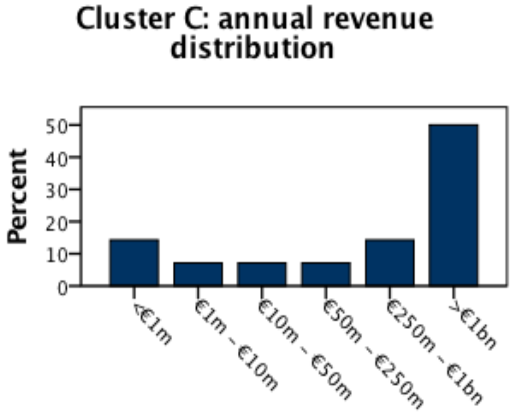


Figure 54: The distribution of annual revenue of cluster C
Source: Author

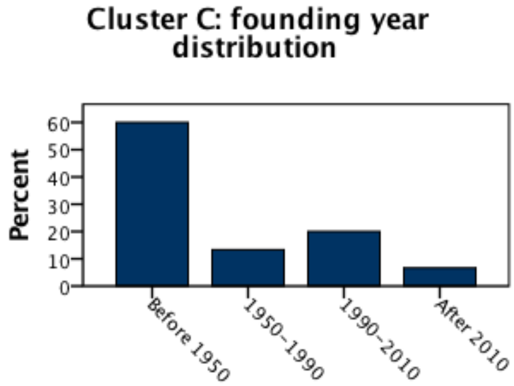


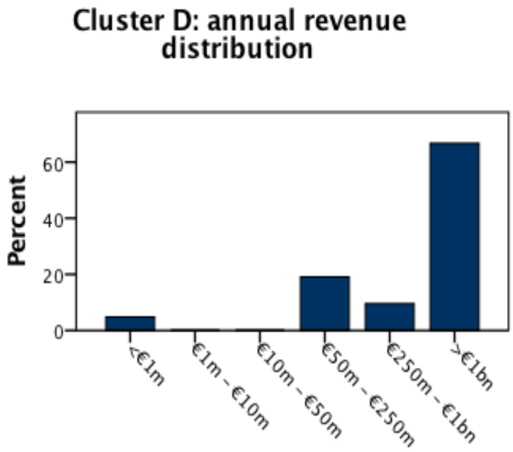
Figure 55: The distribution of founding year of cluster C
Source: Author

Cluster C: sectors	Count
Energy	1
Financial services	5
Food & beverage	2
Human resources	4
IT services	1
Legal services	1
Professional Services	1

Table 14: Sector distribution of cluster C
Source: Author

Cluster D: sectors	Count
Energy	1
Financial services	13
Food & beverage	1
Human resources	1
IT services	1
Electronics	1
Legal services	1
Manufacturing	1
Public sector	1

Table 15: Sector distribution of cluster D
Source: Author



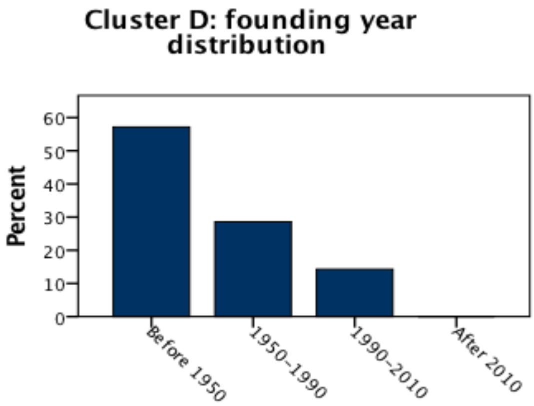
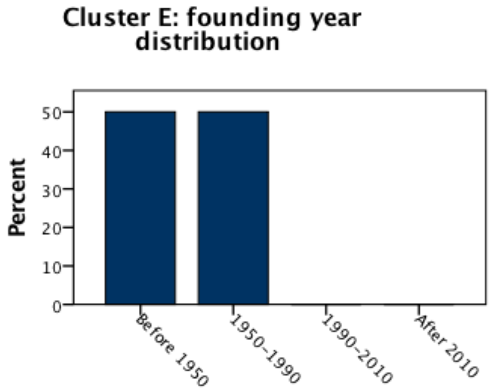


Figure 56: The distribution of annual revenue of cluster D
Source: Author

Figure 57: The distribution of founding year of cluster D
Source: Author



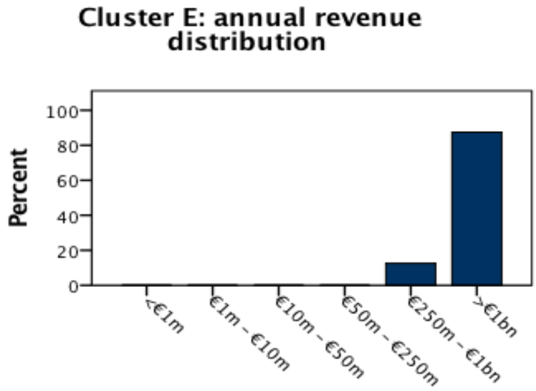


Figure 58: The distribution of annual revenue of cluster E
Source: Author

Figure 59: The distribution of founding year of cluster E
Source: Author

Cluster E: sectors	Count
Agriculture & Agribusiness	1
Chemical	1
Financial services	1
IT services	4
Manufacturing	1

Table 16: Sector distribution of cluster E
Source: Author

9.6 Discussion

During the analysis of the clusters it became clear that the characteristics of the analytical hubris companies (cluster C) display anomalies that are not predicted by the DELTA model.

1. The importance of data & analytics in cluster C is higher than that of the more mature D cluster as shown in figure 49. The DELTA model would predict a linear increase in data & analytics importance.
2. Cluster C scores better than cluster B on the targets and enterprise dimensions, but lower on all the other dimensions. In the DELTA model companies mature incrementally over all dimensions making such an outcome impossible.
3. Cluster C has the most implementation issues as shown in table 11. The DELTA model would predict cluster A has the most implementation issues as “analytically impaired organizations lack one or several prerequisites for serious analytical work” (Davenport, Harris, & Morison, 2010).

At this point it is helpful to introduce the Gartner hype cycle. The hype cycle is "a graphic representation of the maturity and adoption of technologies and applications, and how they are potentially relevant to solving real business problems and exploiting new opportunities" (Gartner, 2017). It depicts the expectations surrounding new technologies as a function of their maturity (time) as shown in figure 60 and describes five maturity stages of technology: innovation trigger, peak of inflated expectations, trough of disillusionment, slope of enlightenment and the plateau of productivity. Although the hype cycle is a well-known model, "it is practically an institution in high tech" (Mullany, 2016), it is worth mentioning that it has received some serious criticism. Steinert, & Leifer (2010) argue the model does not correspond to the empirical evidence and that it lacks an underlying quantitative model. Veryard (2005) disagrees with the shape of the graph and mentions the model is rather a hype curve instead of a hype cycle. As this criticism certainly deserves some thought, the model seems to fit our data remarkably well as it explains the anomalies witnessed in the analytical hubris companies of cluster C.

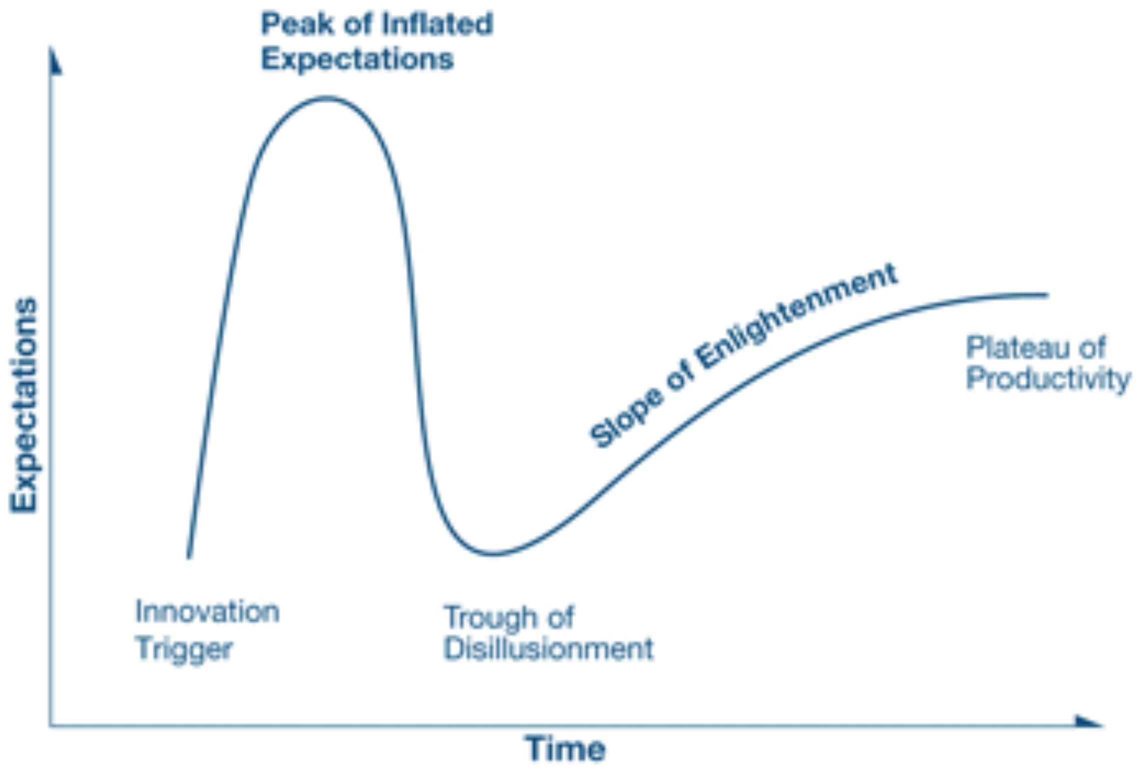


Figure 60: The Gartner hype cycle
Source: Gartner (2017)

The clustering depicted in figure 49 closely resembles the hype cycle and its narrative seems to be applicable as well. The non-linear hype cycle thus better describes the behavior of the companies in the data than the linear DELTA model. The Analytical hubris companies are at the "peak of inflated expectations" of the hype cycle. They have made extremely ambitious plans, which they cannot turn into value yet. According to the hype cycle model these companies will go through the "trough of disillusionment" during which they adjust their expectations. Based on our data we cannot make similar statements, a longitudinal study would be necessary to chart those dynamics. Companies in cluster C could indeed scale down their ambitions and focus on realizing impact on specific domains and move toward the B cluster. However, as the company profile of the C cluster is different than that of cluster B (cluster C has more large firms) it could also be the case that those companies will invest heavily in data & analytics to achieve their goals and move immediately toward cluster D. In this case impact lags behind ambition.

9.7 Conclusion

In general, our empirical research confirms the legitimacy of the DELTA model. When applying k-means clustering to our dataset the resulting clusters resemble the maturity stages of the DELTA model in 4 out of 5 cases. However the DELTA model does not describe the characteristics of cluster C: analytical hubris companies. Companies in this cluster are extremely ambitious and stress the extreme importance of data & analytics but they are unable to translate this to concrete impact. By including non-linear relationships between plans and impact as described in the Gartner Hype Cycle, this anomaly can be explained.

10.0 Cluster-specific results

In this section we will split the data into the five previously defined maturity clusters: analytically impaired (A), localized analytics (B), analytical hubris (C), analytical companies (D) and (E) analytical competitors. This way we can closely compare the characteristics of companies at each of the maturity stages.

10.1 Targets

10.1.1 The importance of data & analytics

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Percentage of instances	19%	32%	17%	23%	9%
Importance of data & analytics ***	Moderately important	Very important	Extremely important	Very important	Extremely important

Table 17: The importance of data & analytics across the 5 maturity clusters.
Source: Author

Table 17 recapitulates the results obtained in the previous chapter. Companies in cluster A rate data & analytics as moderately important, companies in cluster B and D as very important and companies in clusters C and E as extremely important. Note that companies in cluster C are potentially overestimating the impact of data & analytics on their organization as discussed in the previous chapter. The differences between the clusters are statistically significant.

10.1.2 Data & analytics plans

Table 18 depicts the data & analytics ambitions across the 5 generated clusters. The first number is the percentage of companies of each cluster that indicates to pursue the respective plan. The number in between brackets is the percentage of companies of each cluster that indicates to already witness noticeable impact on the respective domain.[5] The top 5 plans for each cluster are marked in yellow. The three domains with the highest impact/plan ratio for each cluster are marked in green. Overall, we can conclude that the main reasons for the implementation of data & analytics in companies are similar throughout all the clusters. To increase revenue, to improve decision making, to better understand the customer and to improve marketing and customer targeting are all within the top 5 across four or more clusters. Note that these strategies are offensive and marketing and sales driven. For most of the domains, the difference in number of plans is statistically significant across the clusters.

[5] Note that in two cases the percentage of firms witnessing impact is higher than the percentage of firms planning data & analytics initiatives in the respective domain. This is actually not possible. A lack of attention while filling out the survey could be an explanation.

In cluster A, analytically impaired companies, we note a considerable interest in improving internal efficiency and improving products and services. Furthermore 18% indicates not to know why its company is implementing data & analytics and 35% declares not to know what impact data & analytics already has. Lastly, 14% of cluster A firms are not yet witnessing a noticeable impact, the highest number of all clusters. This is also the cluster with the lowest absolute number of plans and impact as discussed in section 9.5.1

	A	B	C	D	E
Percentage of instances	19%	32%	17%	23%	9%
To increase revenue ***	24%	48%	73%	86%	100%
To improve decision making ** (ns)	24% (18%)	48% (41%)	80% (33%)	76% (52%)	88% (75%)
To better understand the customer * (***)**	6% (0%)	34% (17%)	100% (7%)	71% (47%)	88% (88%)
To improve marketing and customer targeting * (***)**	12% (0%)	34% (21%)	80% (27%)	76% (67%)	88% (88%)
To survive & stay competitive in the sector * (*)**	12% (6%)	34% (24%)	73% (13%)	66% (29%)	100% (63%)
To improve internal efficiency and cut costs * (***)**	29% (18%)	28% (38%)	60% (33%)	76% (62%)	100% (100%)
To improve customer relationships * (***)**	6% (12%)	31% (14%)	80% (7%)	66% (43%)	75% (75%)
To improve products and services ** (*)	18% (6%)	31% (14%)	53% (33%)	62% (33%)	88% (50%)
To accelerate decision-making * (***)**	12% (12%)	24% (7%)	66% (13%)	62% (38%)	88% (100%)
For the digital transformation of the company * (**)**	6% (6%)	21% (10%)	60% (7%)	67% (38%)	100% (50%)
To create new revenue streams * (***)**	6% (0%)	21% (17%)	53% (7%)	57% (29%)	88% (75%)
To improve risk and compliance management ** (*)	6% (0%)	28% (21%)	40% (26%)	67% (38%)	63% (25%)
To improve the management of existing data * (**)	12% (0%)	31% (17%)	60% (46%)	47% (38%)	38% (38%)
As a competitive differentiator * (***)**	6% (0%)	14% (3%)	60% (0%)	38% (29%)	88% (50%)
To monetize existing data * (*)**	6% (6%)	7% (0%)	40% (7%)	33% (24%)	75% (13%)
To find and exploit new data sources ** (ns)	0% (0%)	14% (7%)	20% (0%)	43% (24%)	50% (13%)
To monitor competitor behavior ns (ns)	0% (0%)	17% (0%)	26% (13%)	24% (10%)	38% (13%)
To increase cyber security * (***)	6% (0%)	10% (3%)	13% (7%)	24% (24%)	63% (50%)
No noticeable effect yet (ns)	18%	14%	7%	0%	0%
I don't know * (***)	18% (35%)	0% (0%)	0% (13%)	0% (0%)	0% (0%)
Impact/Plan ratio	**50%**	**55%**	**28%**	**60%**	**64%**

Table 18: The data & analytics plans and (impact) across the 5 maturity clusters.
Source: Author

In cluster B, localized analytics, we observe, apart from the four main reasons, plans to implement data & analytics to survive and stay competitive. These companies seem to be particularly good at leveraging data & analytics for improving decision making, efficiency and risk and compliance management. Moreover, 14% indicates to have no noticeable impact yet.

Companies in cluster C, analytical hubris, have roughly the same ambitions as the companies in cluster B although the absolute level of ambitions is significantly higher. Overall these companies are not very successful in their efforts. Relatively, they score best on improving products and services, risk and compliance and data governance. In this cluster, 7% of respondents indicate not to already witness a noticeable impact and 13% do not know.

Cluster D, analytical companies, focuses on improving internal efficiency next to the four main reasons. These companies appear to be good at improving marketing, data governance and cyber security. In this cluster all respondents indicate to already witness noticeable effects.

Lastly, cluster E (analytical competitors) scores well on virtually all domains as expected.

10.1.3 Impacted business activities

Table 19 depicts the business activities in which data & analytics are used across the five clusters. Overall, there appears to be little difference between the clusters in terms of relative importance of each of the departments. Marketing, sales and finance appear to be the most impacted business activities across clusters B, C, D and E. In absolute numbers, the differences between the clusters are remarkable and statistically significant. The most mature E cluster leverages data & analytics over almost all departments resulting in a cluster average of 7 impacted business units. Interestingly, 88% of those organizations use data & analytics in their HRM department. For the other clusters this application seems not to be a priority.

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Marketing ***		55%	53%	80%	100%
Sales ***	6%	45%	60%	76%	88%
Finance ***		51%	20%	71%	100%
Risk ***		28%	20%	61%	75%
Quality **	6%	20%	7%	38%	63%
HRM ***		14%	13%	33%	88%
R&D ***	6%	17%		14%	75%
Manufacturing ***		17%		14%	63%
Supply chain ***		14%	7%	5%	75%
Others (ns)	11%	10%	7%	10%	
Cluster Average	**0.29**	**2.72**	**1.87**	**4.05**	**7.13**

Table 19: The importance of data & analytics across the 5 maturity clusters.
Source: Author

10.1.4 Implementation challenges

Table 20 depicts the implementation challenges across the 5 generated clusters. These implementation challenges are in general uniform across the 5 clusters. Lack of skill, organizational structures, organizational culture and low data quality are often indicated issues in all clusters. Surprisingly, clusters B, C, D and E are reporting roughly the same number of issues although their maturity levels are very different. Furthermore, the number of issues is often not statistically significant across the clusters. Implementing data & analytics seems to remain an exhaustive undertaking, even for the most mature organizations. Moreover, when companies mature the spectrum of issues they are facing seems to be widening instead of shrinking. Many organizations in cluster E are also reporting data security issues and legal issues, challenges that are less pronounced in less mature organizations. Cluster A has the lowest cluster average in terms of faced challenges. It could be that companies in this cluster are focusing on "low hanging fruit" and therefore not witnessing many challenges. Another possibility is that the respondents from this cluster had too little knowledge about the data & analytics initiatives in their organizations as 30% reports not to know what challenges their organization is having.

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Lack of skill *	6%	48%	40%	47%	38%
Organizational structures (ns)	24%	34%	46%	38%	38%
Organizational culture **	12%	48%	53%	14%	50%
Low data quality (ns)	12%	34%	46%	38%	25%
Lack of budget (ns)	12%	28%	40%	24%	13%
Overwhelmed by the volume of the data *		21%	33%	33%	
Privacy issues *	12%	10%	6%	38%	38%
Difficulties demonstrating or monetizing the impact of data projects (ns)	6%	28%	27%	10%	13%
Legal issues (ns)	6%	7%	27%	14%	38%
Others (ns)	6%	10%	20%	14%	
Lack of stakeholder' sponsorship (ns)	6%	3%	6%	14%	13%
Don't know **	30%				13%
Data security issues **		3%		5%	50%
Lack of C-level sponsorship (ns)	6%			5%	
Cluster Average	**1.06**	**2.76**	**3.47**	**2.95**	**3**

Table 20: The implementation challenges across the 5 maturity clusters.
Source: Author

10.1.5 Solutions

Table 21 depicts the reported solutions for the implementation challenges. For firms in clusters B, C and D, specific data analytics projects to prove value and effectiveness is the most mentioned solution. Remarkably, the number of companies in cluster E that use this solution is much lower as the value of data & analytics is probably already agreed upon. As companies mature they are leveraging the implementation of new technologies, change management, internal workshops and a larger budget to an increasing extent as a solution for their implementation challenges. The reverse seems to be true for external consulting as only 13% of companies in cluster E mention this option. As data & analytics are crucial (and often of strategic importance) for these firms it is often a good idea to develop and implement their tools themselves, which explains these 2 evolutions. McKinsey (2013) further discusses this make or buy tradeoff. The differences across the clusters are often not statistically significant.

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Specific data analytics projects to prove value and effectiveness * **		62%	73%	66%	38%
**Implementation of new technologies ** **	12%	28%	33%	62%	63%
External consulting (ns)	24%	38%	33%	48%	13%
**Change management ** **	6%	31%	33%	42%	75%
Hiring new employees (ns)	18%	28%	20%	52%	38%
First steps to making this a strategic priority are taken (ns)	12%	31%	40%	24%	
Internal trainings/workshops (ns)	6%	24%	13%	38%	50%
Larger budget (ns)	6%	14%	13%	10%	25%
Don't know *	30%	3%	6%		13%
Others (ns)			6%	10%	

Table 21: The reported solutions across the 5 maturity clusters.
Source: Author

10.1.6 Additional data & analytics investments

Table 22 depicts the planned additional data & analytics investments across the 5 clusters. Surprisingly, more mature companies are more likely to plan additional investments. This will probably further increase the maturity differences between the clusters. Companies in cluster C have the second highest number of additional investments during the course of this year. They possibly feel they have to make an additional effort to realize their high ambitions. Lastly, 34% of companies in cluster A reports not to plan any additional investments. Via Kendall's tau coefficient we find that the overall relationship between maturity and investment plans is not significant as depicted in the left top corner of table 22.

NS[6]	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Yes, within the next year **	18%	48%	67%	57%	100%
Yes but not within the next year ns	24%	10%	13%	24%	
No ns	34%	21%		19%	
Don't know ns	24%	21%	20%		

Table 22: Additional data & analytics investments across the 5 maturity clusters.
Source: Author

10.2 Leadership

10.2.1 C-level support

Table 23 displays the C-level support across the clusters. In particular it shows the distribution of the answers to the statement "there is strong C-level support for data analytics projects". The results are not surprising. Companies in the mature E cluster all indicate to witness significant C-level support. Furthermore, the vast majority of cluster D agrees or strongly agrees with the abovementioned statement. The answer distribution for cluster C resembles that of cluster D with the exception that 13% of the companies in cluster C indicate to strongly lack C-level support. As mentioned before, these companies fail to turn their ambitions into value. A lack of C-level support could be one of the reasons for a part of the companies in this cluster. The B cluster has companies in all of the 5 C-level support categories but a majority of 62% agrees or strongly agrees with the abovementioned statement. Lastly, companies in cluster A seriously lack C-level support, which can explain their low data & analytics ambitions.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Strongly agree ns	6%	31%	20%	29%	75%
Agree ***		31%	60%	52%	25%
Neither agree not disagree ns	12%	21%	7%	19%	
Disagree ns	6%	14%			
Strongly disagree ***	76%	3%	13%		

Table 23: C-level support for data & analytics across the 5 clusters.
Source: Author

10.2.2 Leader of data & analytics initiatives

Table 24 shows who is in charge of data & analytics across the five maturity clusters. The companies in cluster A do not have any specific data & analytics related executives. The COO

[6] Note that for the ordinal variables we investigate significance via the Kendall's tau coefficient. Next to this we also investigate significance at a categorical level via a Chi-squared test.

and the CEO are leading data & analytics initiatives. Furthermore 45% of respondents in this category declare not to know who is responsible for data & analytics in their company.

In cluster B the spectrum of data & analytics executives is considerably larger. In these companies mainly data scientists and innovation managers are in charge of data & analytics but in total 11 different roles have been identified. Note that 7% of these companies have a Chief Data Officer and 7% have a Chief Digital Officer. In companies in cluster C the CEO is most often indicated as the executive in charge of data & analytics. Note that also in this cluster there are many different executives that lead the data & analytics projects. Furthermore, 15% of companies in this cluster have a Chief Data Officer and 8% of companies have a Chief Digital Officer. Cluster D is mostly similar to cluster C although some interesting trends can be identified. Firstly, the Chief Data Officer is most strongly represented in this cluster as 19% of companies indicate to employ one. Secondly, the Chief Digital Officer has disappeared at this stage of maturity. This observation confirms the theoretical predictions as described in 6.2.2 namely that the Chief Digital Officer has a transformational role and his/hers importance diminishes once the company has reached a certain stage of maturity. Also note that this is the only cluster that employs Chief Analytics officers. In the most mature E cluster, 62% of the companies report their CEO is in charge of data & analytics. Surprisingly, this cluster employs none of the previously discussed data & analytics C-level officers. For these companies the data & analytics agenda seems to coincide with the corporate strategy. Therefore it is not necessary to have an additional C-level officer.

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
COO	22%	10%		10%	12%
CEO		3%	23%	5%	62%
Chief Data Officer	11%	7%	15%	19%	
CMO		7%	15%	14%	
Data Scientist		21%			
Data Analyst		10%		14%	
Innovation manager		14%			13%
Chief Digital Officer		7%	8%		
IT project manager		3%	8%	5%	13%
CIO		3%	8%	5%	
CAO				10%	
CFO			15%		
Others	22%	3%	15%	10%	
Don't know	45%	10%		5%	

Table 24: Leader of the data & analytics initiatives across the clusters.
Source: Author

10.2.3 Recruitment intentions: data & analytics related C-levels

When analyzing the recruitment intentions across the five maturity clusters some interesting trends are observed. First of all, none of the companies in cluster A plans on hiring a data & analytics related C-level officer. Clusters B, C and D all show a limited interest in hiring such an executive. Surprisingly, the strongest recruitment intentions can be observed for the companies in cluster E. Of these companies 43% report to plan on hiring a data & analytics related C-level within the next year. Because of this, our hypothesis from the previous section that the most mature companies do not need a separate data & analytics leader is rejected. However, overall the relationship between maturity and recruitment intentions is not significant as indicated in the top left corner.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Yes, within the next year ns		12%	9%	6%	43%
Yes, but not within the next year ns		4%	9%	6%	
No ns	71%	72%	45%	75%	57%
Don't know ns	29%	12%	36%	13%	

Table 25: recruitment intentions across the clusters.
Source: Author

10.2.4 The effect of a dedicated C-level officer on the success rate

In this section we investigate the effect of a dedicated C-level officer on the success rate (plan/impact ratio) of data & analytics projects. Overall, companies with such a C-level executive have a slightly higher success rate as they are already witnessing a noticeable impact on 56% of their data & analytics effort compared to just 52% of the entire sample. Table 26 depicts the average number of plans, the average number of impacted domains and the success rates of the companies with a dedicated C-level officer. To make a fair comparison we have put the success rates of the respective maturity clusters in between brackets to only compare companies with the same maturity level. The fields for which no information was available in the sample have been shaded.

The evidence for this analysis is mixed. In clusters A and C having a Chief Data Officer or Chief Digital Officer clearly pays off. In cluster B these positions, counter intuitively, seem to dampen the success rate. In cluster D both an increase and a decrease in success rate can be observed. Interestingly, in the cases were a C-level officer increased the success rate this was mainly due to a decrease in the number of plans, and not a miraculous increase in impact. This can be important insight: hiring a dedicated C-level officer can help your company focus on the initiatives that really matter.

However, because the number of C-levels in every cluster is very low (often even only 1) the results are not expected to be generalizable for the entire population.

		A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Chief Data Officer	# Plans	1 (2)	6.5 (4.9)	9 (10.40)	12 (10.48)	
	# Impact	1 (1)	2.5 (2.72)	6 (2.87)	5 (6.24)	
	Success rate	100% (50%)	38% (55%)	67% (28%)	43% (60%)	
Chief Digital Officer	# Plans		6.5 (4.9)	4.5 (10.40)		
	# Impact		3 (2.72)	3.5 (2.87)		
	Success rate		46% (55%)	77% (28%)		
Chief Analytics Officer	# Plans				8 (10.48)	
	# Impact				6.5 (6.24)	
	Success rate				67% (60%)	

Table 26: Success rate of data & analytics projects in companies with a dedicated C-level officer. The numbers in between brackets are the cluster averages.
Source: Author

10.3 Data

10.3.1 Data collection

Table 27 provides interesting insights on the data formats collected by the companies in the different clusters. The differences in collected data formats are all statistically significant. Companies in cluster A barely collect any data at all. Only 18% indicate to collect structured data. Organizations in cluster B have a much more diversified portfolio of data sources although most of the organizations still heavily rely on structured data formats. Companies in cluster C collect considerably less data formats as companies in cluster B, which could in part explain the disability of this cluster to fulfill their ambitions. The main difference between clusters D and E lies in the unstructured data collection. In both clusters all companies indicate to collect structured data, but the number of unstructured data sources is considerably larger in cluster E.

As shown in table 28, companies in cluster A only collect data from back office systems and 3rd parties. Companies in cluster B on the other hand have a much larger number of data sources. Companies in cluster C again collect from roughly the same sources as companies in cluster B, but the absolute number of sources is remarkably lower. Companies in the mature D cluster collect data from a wide range of sources although the focus on the use back office systems remains. This focus disappears in cluster E. These organizations make intensive use of a wide array of data sources. When comparing these two most mature clusters there are in particular big differences in the use of social media, mobile apps and sensors as data sources.

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Structured ***	18%	90%	73%	100%	100%
Text data ***		55%	13%	61%	100%
Clickstream data ***		28%	6%	38%	88%
Geospatial **		20%	13%	19%	63%
Weblogs ***		10%	6%	19%	88%
Streaming ***		10%	6%	19%	75%
Video ***		10%		19%	63%
Audio ***			6%	14%	50%
Others *	6%	3%		5%	13%
Don't know ns	6%		6%	5%	
Average # of formats	0.24	2.34	1.40	2.48	6.63

Table 27: Collected data formats across the clusters.
Source: Author

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Back Office Systems (ERP) ***	12%	66%	60%	90%	88%
Website ***		48%	27%	62%	75%
Point of sale ***		34%	27%	67%	63%
3rd party ***	6%	38%	7%	57%	75%
Social media ***		38%	13%	43%	88%
Email ***		34%	20%	57%	38%
Call center ***		14%		52%	63%
Mobile apps **		10%	7%	33%	63%
Sensors *		14%	13%	5%	50%
Others ns	6%	17%	7%	10%	
Don't know ns	6%		13%		13%
Average # of sources	0.18	3.21	1.80	4.62	6.00

Table 28: Data sources across the clusters.
Source: Author

10.3.2 Data overview

Table 29 depicts the distribution of the answers to the statement "my company has a clear view of the volume of data stored and gathered". In general, the level of data overview seems to be an increasing function of maturity. However, a much higher number of companies in the analytical hubris cluster strongly agree with the statement in comparison to the more mature D cluster. It is not clear whether these companies genuinely have a clear view or whether they are overestimating their abilities because they are not able to translate their ambition into value. Furthermore, note that none of these differences are statistically significant.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Strongly agree ns		10%	17%	5%	25%
Agree ns		24%	42%	33%	50%
Neither agree nor disagree ns	67%	34%	25%	43%	25%
Disagree ns	33%	31%	17%	14%	
Strongly disagree ns				5%	

Table 29: The roles and responsibilities for data governance are clearly defined.
Source: Author

10.3.3 Data governance

Table 30 shows the distribution of the answers to the statement "the roles and responsibilities for data governance are clearly defined". The analysis of table 30 is very similar to that of table 29: in more mature firms the roles and responsibilities for data governance tend to be better defined. Again cluster C is a remarkable outlier and none of the differences are statistically significant.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Strongly agree ns		7%	25%	10%	38%
Agree ns		41%	42%	48%	50%
Neither agree nor disagree ns	67%	28%	25%	10%	12%
Disagree ns	33%	24%		33%	
Strongly disagree ns			8%		

Table 30: My company has a clear view of the volume of data stored and gathered
Source: Author

10.3.4 Data accessibility

Table 31 depicts the distribution of answers to the statement: "employees of any department who want to use data to discover new insights can easily access the necessary data (if they are entitled to get it)". All respondents from cluster A indicated "neither agree nor disagree".[7] For clusters B and C, 50% or more agrees or strongly agrees with the statement. Surprisingly, companies in cluster D seem to be doing a worse job in allowing easy access to data. Lastly, all companies in cluster E agree or strongly agree with the statement.

[7] It is worth mentioning that this question did not have "don't know" as an answer possibility. It could be possible that respondents from this cluster picked this answer in lack of another feasible option. With further research in mind, the "don't know" option should be included.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Strongly agree ns		24%	17%	5%	50%
Agree ns		38%	33%	43%	50%
Neither agree nor disagree **	100%	14%	25%	38%	
Disagree ns		24%	17%	10%	
Strongly disagree ns			8%	5%	

Table 31: Employees of any department who want to use data to discover new insights can easily access the necessary data (if they are entitled to get it**).**
Source: Author

10.3.5 GDPR compliance

A third of the companies from cluster A agree to the statement: "my company will have no difficulties complying with GDPR by May 2018 (General data protection regulation by the European Commission)." In cluster B, 38% of companies agree of strongly agree with the statement. Surprisingly 45% of companies from cluster C strongly agree with the statement. This is suspicious given the fact that only 25% of companies in the most mature E cluster strongly agree. This could be a strong indication that companies in cluster C are not yet fully aware of the complexity surrounding the legislation and are thus overestimating their abilities. The answers from cluster D and E are in line with the expectations. However, note that 25% of companies from cluster E report to have difficulties complying with GDPR. The legislation seems to be a challenge even for very mature organizations and it is worth mentioning that most of the differences are not statistically significant.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Strongly agree ns		7%	45%	10%	25%
Agree ns	33%	31%	9%	38%	38%
Neither agree nor disagree *	67%	48%	27%	19%	12%
Disagree ns		14%	18%	33%	25%
Strongly disagree ns					

Table 32: My company will have no difficulties complying with GDPR by May 2018.
Source: Author

10.3.6 Data related issues

Table 33 depicts the issues that arise because of low data quality. Half of companies in cluster A say to experience data quality issues. These issues are solved on a department level. In cluster B on the other hand we notice a tendency to organize data quality governance more centrally and to device solutions for data quality issues at an enterprise level. The same observation can be made for companies in cluster C. Interestingly, 18% of these companies report not to experience any challenges due to poor data quality. This seems counterintuitive as these companies experience the highest number of challenges. When comparing cluster D and E, we note that less companies in the most mature cluster indicate not to experience challenges. This is also counterintuitive at first but could be explained by the much higher number of data formats and sources used by companies in cluster E, which could strongly impact data quality issues. Note that none of the differences are statistically significant.

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
We experience a wide range of issues that impact our organization. ns		24%		10%	38%
We experience some challenges but it is up to each department to solve them. ns	50%	28%	36%	35%	12%
We experience some issues, but we are always able to solve them at an enterprise wide level. ns		34%	36%	50%	38%
We have not experienced or no longer experience any challenges due to poor data quality at our organization. ns			18%	33%	12%
Don't know. ns	50%	14%	9%	5%	

Table 33: Data quality issues across the five clusters.
Source: Author

10.3.7 Data quality overview

Table 34 gives an overview to what extent companies have a clear view on the accuracy of their data. Surprisingly, this overview is often non-existent or very limited. Even in the most mature E cluster 50% of companies report not to be aware the status of their data at all or to only know the status of specific important data sets. The rest of the analysis of table 34 is in line with the initial description of the clusters from section 9.5. Companies in cluster A either have no overview of their data quality or report not to know. Companies in cluster B mainly report to have no overview although some of these companies are already having an overview on an enterprise wide level. Companies in cluster C mainly indicate "don't know" as an answer and none of these companies already have a department or enterprise wide overview of the data quality confirming the low data & analytics maturity of this cluster. Cluster D is the first cluster in which companies report to have KPIs in place to continuously improve their data quality. Cluster E has the best data quality overview of all the clusters, yet only 12% has progressive KPIs in place.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
No ns	50%	38%	18%	35%	12%
We know the status of our data for specific, important data sets. ns		31%	36%	25%	38%
Each department has a view on the status of the data and its quality ns		17%		20%	25%
We have a clear view of the data status on an enterprise wide level. ns		10%		5%	12%
We have a clear view of the data status and we have progressive KPIs that ensure continuous improvement. ns				5%	12%
Don't know *	50%	3%	45%	10%	

Table 34: Data quality overview across the five clusters.
Source: Author

10.4 Analysts

10.4.1 Data-driven decision making

More data & analytics mature companies use the output of their analyses to an increasing extent in their decisions as depicted in table 35. 60% of companies in cluster E report to always use data & analytics in their decisions compared to 0% and 3% for the least mature firms. These differences are statistically highly significant, but the overall relationship between maturity and data-driven decision making is not.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Never *			9%		
Sometimes ns	50%	24%	27%	10%	
About half the time ns	50%	24%	18%	10%	
Most of the time **		48%	9%	70%	40%
Always ***		3%	36%	10%	60%

Table 35: The frequency of which data & analytics are used in decisions.
Source: Author

10.4.2 Feedback processes for assessing the accuracy of analyses

Table 36 depicts the distribution of answers to the statement: "we have feedback processes in place to assess the quality and accuracy of our analyses". Even the "analytical competitors" seem to have some difficulties regarding these feedback processes as only 13% strongly agrees. The rest of the answers are in line with the expectations. Note that 36% of the firms in the analytical hubris cluster disagree with the statement. A lack of solid post mortem analyses of their analyses could be a reason why these firms fail to fulfill their ambitions. Note that the differences between the groups are not statistically significant for any of the answer possibilities.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Strongly agree ns		7%	9%	5%	13%
Agree ns		28%	18%	55%	50%
Neither agree nor disagree ns	100%	41%	27%	20%	38%
Disagree ns		17%	36%	20%	
Strongly disagree ns		7%	9%		

Table 36: We have feedback processes in place to assess the quality and accuracy of our analyses.
Source: Author

10.4.3 Data & analytics teams

Table 37 depicts the roles represented in the data & analytics teams across the five maturity clusters. Companies in cluster A virtually have no teams. Both clusters B and D have well diversified teams with a focus on business analysts. For the most mature firms this focus seems to shift to data scientists although business analysts are still very well represented. Companies in cluster C report a low number of team diversification. Note that the differences in roles represented in data & analytics teams between the groups are all highly significant.

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Business analyst ***		66%	40%	70%	63%
Data Scientist ***		45%	20%	50%	75%
IT expert ***		24%		45%	38%
System architect ***		14%		25%	38%
Others ns	100%	7%	40%	15%	

Table 37: The roles represented in data & analytics teams across the five clusters.
Source: Author

10.4.4 Rare skills

In table 38 the skills related to data & analytics that are the hardest to get by are reported. As already discussed in section 8.4.2 skills that bridge IT and business are reported to be the hardest to get by. This observation seems to be true for all clusters with the exception of cluster A. Apart from IT skills, none of the differences between the clusters are statistically significant.

	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Skills that bridge IT and business ns		45%	36%	43%	50%
Analytical skills ns		31%	36%	33%	38%
Creativity/innovation ns		38%	27%	29%	38%
Sector knowledge/experience ns		31%	18%	38%	25%
IT skills *		35%	18%	10%	38%
Don't know ns	50%	14%	27%	10%	0%
Teamwork/motivational skills ns		7%	9%	19%	25%
Sales skills ns	50%	7%	0%	5%	0%
Others ns		3%	0%	5%	0%

Table 38: The data & analytics skills that are the hardest to get by.
Source: Author

10.4.5 Front-line employee training

Table 39 depicts the responses to the statement: "a significant part of our data & analytics budget goes to training front-line employees to work with the models and their results". As discussed in section 8.4.2, McKinsey (2013) reports this as a reason for failure for many data & analytics initiatives. We observe that even for the most mature firms only 50% agrees or strongly agrees with the statement. The other clusters are doing even worse as "disagree" is the mode for each of the clusters. The differences are not statistically significant.

NS	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Strongly agree ns		3%			13%
Agree ns		14%		15%	38%
Neither agree nor disagree ns	50%	24%	36%	30%	13%
Disagree ns	50%	48%	45%	45%	38%
Strongly disagree ns		10%	18%	10%	

Table 39: We invest in training front-line employees to work with the models and their results.
Source: Author

10.5 Enterprise

Table 40 tries to give an insight at which level data & analytics are organized across the five maturity clusters. Companies in clusters A and B mainly indicate to rely on ad-hoc analytics projects by individuals. This remains true for companies in cluster C although they also organize their projects at a department level and at an enterprise level to an increasing extent. For companies in cluster D more than 50% organizes data & analytics at an enterprise level and for the most mature E cluster this number increases to 76%. The relationship between maturity and the level data & analytics is organized at is statistically significant.

***	A Analytically impaired	B Localized analytics	C Analytical hubris	D Analytical companies	E Analytical competitors
Individuals use ad-hoc analytics projects to tackle specific issues *	50%	55%	27%	15%	13%
We have automated analytics projects within certain departments ns		24%	27%	20%	13%
We have automated analytics processes at an enterprise level ns		3%	9%	30%	13%
We both have automated analytics processes at an enterprise level and experimental teams to innovate in analytics ns		17%	18%	25%	63%
Don't know ***	50%		18%	10%	

Table 40: At which level does your company organize its data & analytics efforts?
Source: Author

11.0 Conclusion

In this master's thesis we empirically reviewed the DELTA maturity model by clustering survey data into five maturity clusters and conducted an as-is analysis of the maturity of companies with respect to data & analytics. First we find that an incremental increase in maturity across the maturity dimensions, as assumed in many maturity models, is not a perfect description of reality. Moreover, we observe that the relationship between ambitions and impact is non-linear. By using the Gartner Hype cycle model this behavior can be better understood. This is an important conclusion for the maturity model literature in which many models are not proficiently empirically tested.

In the rest of the conclusion we will highlight the main findings of the full sample as-is analysis and the cluster-specific results across the five DELTA dimensions.

11.1 Targets

The vast majority of firms in the dataset have picked up upon the data & analytics potential. For more than 70% data & analytics are extremely or very important. The three most mentioned targets are increasing revenue, improving decision making and better understanding the customer. The most impacted business units are sales, marketing and finance. Furthermore, it is worth noting that companies put heavy emphasis on offensive applications. The defensive applications appear to be put lower on the agenda although implementing these defensive applications seems to be easier as they have a higher success rate. In total, only 52% of ambitions are already translated into noticeable impact. People and organizational characteristics substantiate the main causes of failure as 4 out of 5 most mentioned challenges are non-technological. The interest of the vast majority of companies in data & analytics is expected to endure and even increase in the future as only 20% of organizations indicate not to plan additional data & analytics related investments.

When analyzing the cluster-specific results we observe that data & analytics are increasingly more important for more mature organizations, with the exception of cluster C. Furthermore, more mature firms have a wider range of ambitions and a higher impact/plan ratio, again with the exception of cluster C. Leveraging data & analytics to improve business outcomes remains difficult even for the most mature firms as the number of challenges faced is an increasing function of maturity. It is the least mature A cluster that has the lowest number of implementation issues on average. Evaluating the proposed solutions for these problems also generates insightful results. The most mature companies in cluster E mainly solve their challenges in-house as they focus on internal trainings, change management and the implementation of new technologies. Companies in the other clusters rely more on 3rd party support as they often mention external consulting or hiring new employees as their top solutions. As 88% of firms in cluster E indicate to leverage data & analytics as a competitive differentiation it makes sense to internally develop knowledge and skills.

11.2 Leadership

The C-level sponsorship for data & analytics projects is substantial. In total, 63% of survey respondents report to have strong support for these initiatives in their company. Furthermore, several new, data & analytics related C-level positions have been brought to life during the past years. The Chief Data Officer, the Chief Analytics Officer and the Chief Digital Officer are the most common examples of this trend. However, we observe that most of the data & analytics initiatives in companies are still being led by more traditional C-levels, namely the CEO, the COO and the CMO. Only 8% of firms employ a Chief Data Officer, only 4% a Chief Digital Officer and only 2% a Chief Analytics Officer. Furthermore, we report that the recruitment intentions of the companies in the sample with respect to these new C-level executives are limited. Only 10% of respondents plan to hire one of these positions within the next year.

When analyzing the clusters we observe a relationship between the intensity of C-level support and the maturity of a company. Only 6% of the least mature companies indicate to have strong C-level support compared to 100% of the most mature firms. When analyzing the prevalence of the new data & analytics related officers we see that Chief Data Officers are present in clusters A, B, C and D: a wide range of companies seems to be benefitted by his presence. Chief Analytics Officers on the other hand are only employed by companies in the mature D cluster. Lastly, Chief Digital Officers are employed in cluster B and C: their importance seems to diminish once companies reach a certain maturity level, as predicted in the literature. The evidence whether these executives have a positive effect on the impact/plan ratio is mixed.

11.3 Data

We find that traditional, structured data formats are still the most widely collected as almost 80% indicate to gather this data. Back office systems are the most mentioned data source followed by website and point of sale. Furthermore, we observe that the majority of firms have a clear view of the volume of data they store and gather. However, roles and responsibilities for governing this data are often not yet defined. Monitoring the quality of the gathered data seems to be difficult: less than 5% have an overview of data quality at an enterprise level and have data quality KPIs in place.

When analyzing the cluster-specific result we observe that the number of collected data sources is an increasing function of maturity with the exception of cluster C. Companies in cluster A barely collect any data at all while the most mature companies have a very well diversified portfolio of data formats. The difference between the mature D cluster and the most mature E cluster lies in unstructured data collection. The conclusion of the analysis of the used data sources is similar to that of the data formats: more mature companies have a more diversified portfolio of data sources. Furthermore, the level of data overview, data governance, data accessibility and GDPR compliance are, as expected, all positively correlated with an increase in maturity. This cannot be said of the number of data related challenges. We report that even the most mature companies seem to encounter hefty data quality related issues.

11.4 Analysts

Most firms in the sample are profoundly data-driven as 63% indicate data & analytics are involved in most or all the decisions they make. Interestingly only 43% of companies suggest they have feedback processes in place to assess the quality and accuracy of the analyses they use as an input for their decisions. The business analyst is the most common role within a data & analytics team followed by the data scientist and the IT expert. Furthermore, companies seem to have the most difficulties with finding employees that have skills that bridge IT and business. Lastly, less than 20% of firms indicate to spend a significant part of their data & analytics related investments on the training of their front-line employees to work with the models or their results.

More mature companies are more data-driven. In the most mature E cluster, 60% of companies indicate to always use data & analytics in their decisions compared to 0% and 3% for the least mature firms. The same is true for the feedback processes to assess the quality and accuracy of analyses: more mature companies conduct more post-mortem analyses of their analyses. When looking at the data & analytics teams we again observe that more mature firms have more roles represented in their teams. In cluster B, C and D, the business analyst is the most mentioned role. For cluster E the data scientist is most mentioned. Furthermore, companies across all maturity levels seem to encounter the same challenges when looking for valuable skills. Skills that bridge IT and business are most mentioned in 4 out of 5 clusters as the hardest skill to get by. Lastly, more mature companies invest more in training their front-line employees to work with models and their results.

11.5 Enterprise

We observe that almost 35% of companies rely on ad-hoc analytics by individual employees for realizing their data & analytics ambitions. Around 20% organize data & analytics at the department level and 13% at the enterprise level. Lastly, 25% of the companies report to have both automated analytics processes at an enterprise level and experimental teams to innovate in analytics.

When comparing the clusters we observe that there is a relationship between maturity and the level at which data & analytics is organized. None of the companies in cluster A organize their analytics at an enterprise level. A fifth of companies in cluster B organize data & analytics at an enterprise level, 27% of companies in cluster C, 55% of companies in cluster D and 76% of companies in cluster E.

Extension 1: The drivers of the hype

1. Moore's law: the gigantic increase in computing power

Moore's law famously observes an exponential relationship between the number of transistors in a circuit and time. Specifically, it states that the amount of transistors in a chip doubles every two years. For the past 50 years, Moore's law has been a good approximation of reality, although growth seems to be slowing down recently (Simonite, 2016). This incredible increase in computing power has made complex computing activities possible, as our devices are millions of times more powerful than 50 years ago. Advanced data analytics operations that would have taken years can now be done in milliseconds. The servers at Google for example are now able to process 5.5 billion Google searches each day (Sullivan, 2016).

2. Moore's law: the gigantic decrease in cost of computing power

Moore's law is often interpreted from a performance perspective, but the economic perspective is the flipside of the same coin. A doubling of the number of transistors also means halving the cost of the chip. When the transistor density increases, more transistors can be printed on the same silicon wafer and the price of each unit decreases almost proportionally ("Beyond Moore's law", 2015). Figure 61 illustrates this phenomenon. It would take $1,000,000,000,000 in 1950 to buy the computing power of an iPad 2 ($500 at launch). This trend caused the democratization of computing. For a long time, only research facilities were able to afford the super computers needed for advanced analytics. Nowadays every company can afford the necessary technological infrastructure.

3. The data explosion

Today, more than 3.5 billion people are connected to the Internet ("Number of Internet Users - Internet Live Stats", 2017). By 2020 this number will have increased to 6 billion (Marr, 2015). Furthermore, because of the Internet of things evolution, it is expected that by 2020 more than 50 billion smart devices will be around to gather, analyze or share data (Marr, 2015). All these users and devices create an unimaginable amount of data. Let's consider the following statistics:

- In 2012, IBM estimated that 90% of the total data (from beginning of civilization until 2012) has been generated between 2010 and 2012 ("IBM - What is big data?", 2012).
- The amount of data generated worldwide increases by 40% every year (AT Kearny, 2013).

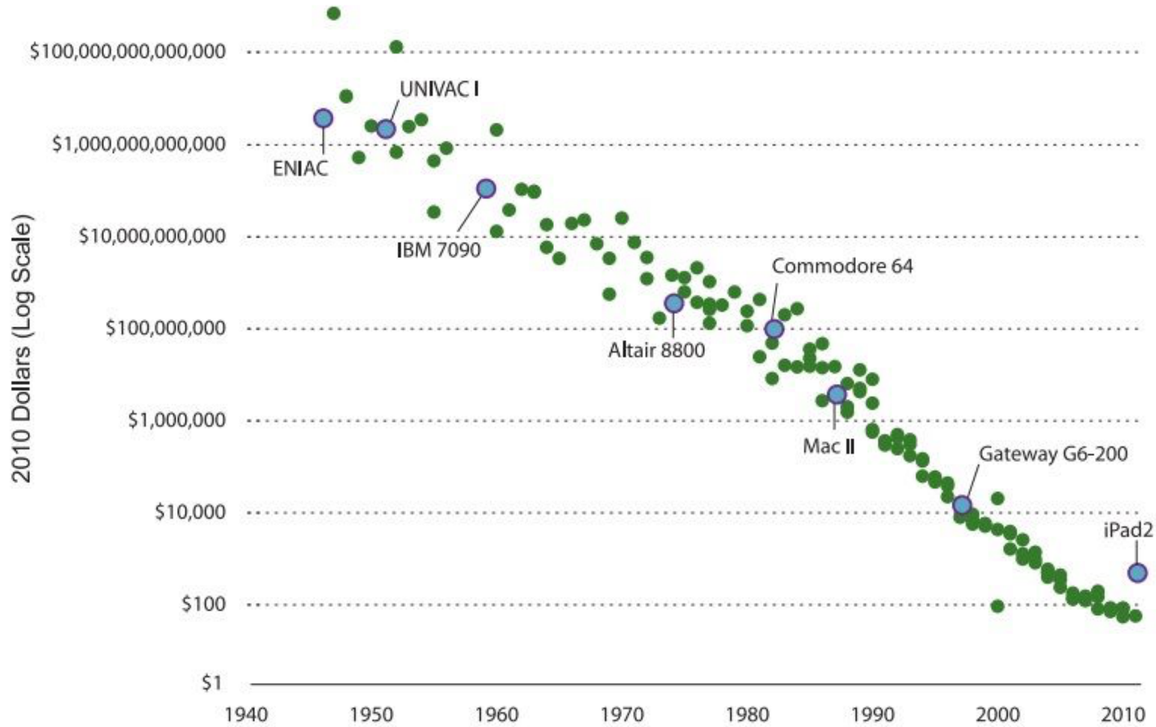


Figure 61: Cost of computing power equal to an iPad2
Source: "Cost of computing power equal to an iPad2 | The Hamilton Project", 2011

- 205 billion emails per day in 2015 (Tschabitscher, 2016)
- 432,000 hours of video uploaded on YouTube per day in 2016 ("STATS | YouTube Company Statistics - Statistic Brain", 2016)
- 1 trillion photos were taken in 2015 (Marr, 2015).

4. Kryder's law

Kryder's law (often referred to as Moore's law of storage) predicts disk storage density doubles every year (Rouse, 2014). This allows storing ever more data as shown in figure 62. This incredible amount of data generated and stored can be of tremendous value for a company if it knows how to extract information from it.

To end this section, there is one more essential statistic that needs to be considered. According to Regalado (2013) only 0.5% of all data is analyzed. In other words 99.5% of all data is never used. The percentage of data analyzed is even expected to keep on shrinking because of the enormous increase in data (Guess, 2015). This statistic singlehandedly justifies the high expectations surrounding data & analytics. There is a vast ocean of undiscovered data waiting to be analyzed and turned into value.

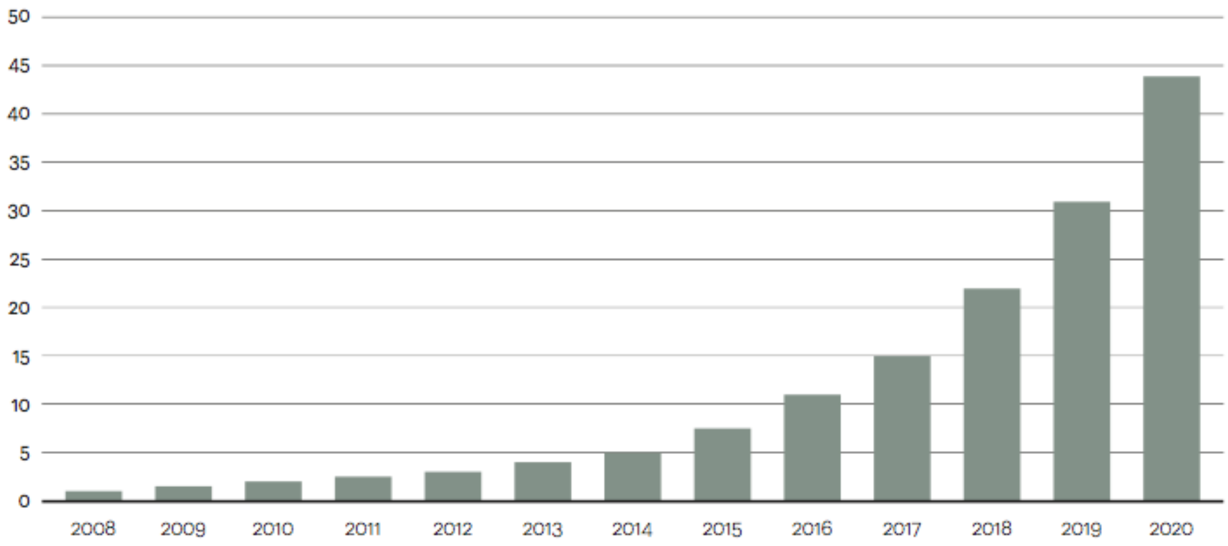


Figure 62: Total amount of data in zettabytes (ZB)
Source: AT Kearny (2013)

Extension 2: the DIKW-Pyramid

The Data-Information-Knowledge-Wisdom pyramid depicts an epistemological process as it materializes the interdependencies between data, information, knowledge and wisdom. It shows the conceptual blueprint of any data & analytics initiative. This hierarchical model is often cited, but mostly implicitly used, in the knowledge management literature (Rowley, 2007; Tien, 2003; Vaes, 2013). As shown in figure 63, a large amount of data is consumed in the creation of a small quantity of wisdom and this process requires energy. The higher up in the pyramid, the better the initial data has been understood and contextualized (Rowley, 2007). In the next section the epistemological tiers mentioned in the DIKW pyramid will be briefly defined. Note that these core concepts have many, sometimes surprisingly different, definitions and are often, mistakenly, used interchangeably. However, for a solid understanding of data & analytics it is vital to be aware of the distinctions between these concepts.

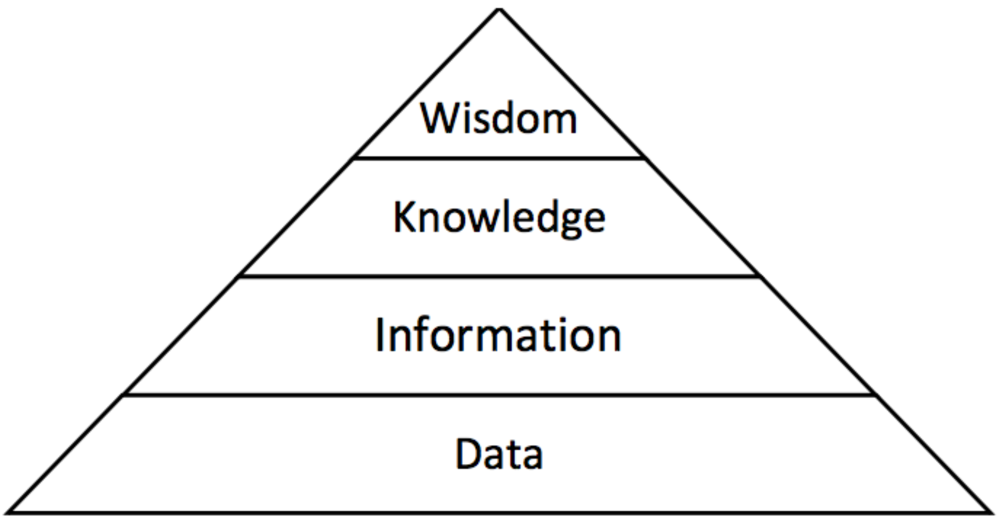


Figure 63: The DIKW pyramid
Source: Rowley (2007)

1. Data

"*Data is a set of discrete, objective **facts** about events*" (Davenport & Prusak, 2000).

*"Data is a **symbol set** that is quantified and/or qualified*" (Wersig & Neveling, 1971).

"*Data are sensory **stimuli** that we perceive through our senses*" (Davis & Olson, 1985).

*"Merely raw **facts***" (Henry, 1974).

*"Know **nothing**"* (Zeleny, 1987).

Form these definitions it becomes clear that data can be:

- Facts
- Symbols
- Stimuli/Signals

Data is the raw ingredient. It has no meaning *an sich*, it has to be refined in some kind of way to become valuable. In the business informatics literature data is often divided into two categories (Davenport, 2013):

- Structured Data: refers to organized data, for example phone numbers or postal codes that always contain a fixed amount of numbers and letters. Tables and records are other frequently mentioned examples (Marr, 2015).
- Unstructured: refers to all data that is not organized in a recognizable structure (Marr, 2015). For example images, video or documents.

2. Information

*"Information is **data** endowed with relevance and purpose."(*Drucker, 2008)

*"Information is **data** that challenges us."* (Henry, 1974)

*"Know **what**"* (Zeleny, 1987).

Information is the second tier of the DIKW pyramid. It consists of data that has been processed in some kind of way, as illustrated in figure 64. Because of this, information can become the input for decisions.

3. Knowledge

*"Knowledge is **structured** and **organized information** that has developed inside of a cognitive system or is part of the cognitive heritage of an individual*" (Hartshorne & Weiss, 1934).

"Knowledge is a justified, true belief" (Plato & Gallop, 1988).

*"Know **how**"* (Zeleny, 1987).

Knowledge is the third level in the DIKW pyramid. Many authors agree that the difference between knowledge and information resides in actionability and value (Rowley, 2007). It consists of information that has been more contextualized and better understood. In this way, it is a more valuable input to decisions.

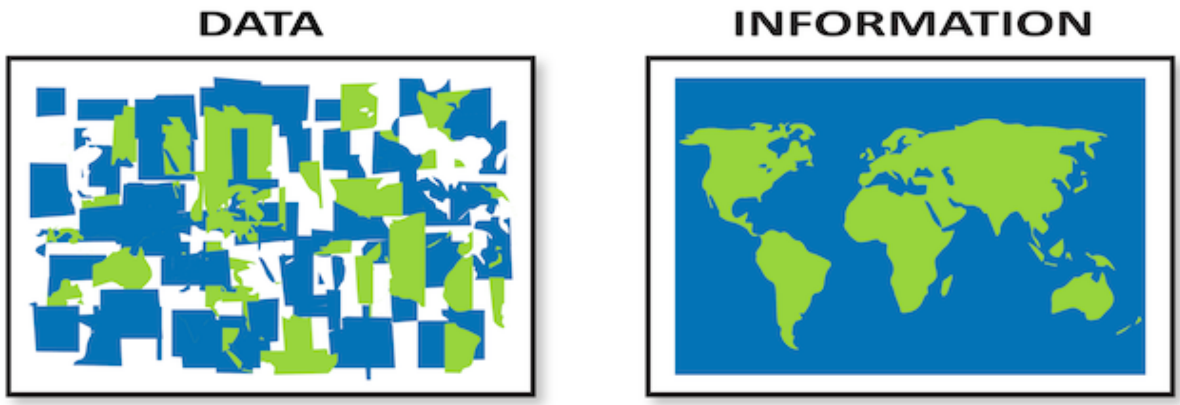


Figure 64: A graphical illustration of the difference between data and information
Source: Systems (2014)

4. Wisdom

"Wisdom is the ability to increase effectiveness. Wisdom adds value, which requires the mental function that we call judgment. The ethical and aesthetic values that this implies are inherent to the actor and are unique and personal." (Ackoff, 1989)

*"Know **why**"* (Zeleny, 1987)

*"Wisdom is **integrated knowledge**, information made super useful by theory, which **relates bits and fields of knowledge to each other**, which in turn enables me to use the knowledge to **do something**" (Cleveland, 1985)*

Wisdom occupies the top of the DIKW pyramid but the concept has been discussed significantly less thoroughly in the literature compared to the other levels (Rowley, 2007). Therefore the term remains somewhat esoteric but this is not a problem for our purposes. We interpret wisdom as the most valuable output of a data refining process.

Alternative versions

The DIKW pyramid has been described many times in the literature. This leads to many, slightly different iterations of the same concept as depicted in figure 65. Some authors add "Action" (Davenport & Prusak, 2000) or "Enlightenment" (Zeleny, 2013) on top of the pyramid, while others dedicate the basis of the pyramid to "Signals" (Choo, 1996). Others argue about the shape (a circle instead or an inverted pyramid would be more appropriate). Anyhow, for our purposes we have defined these concepts extensively. We gladly leave the more in depth discussion to epistemologists.

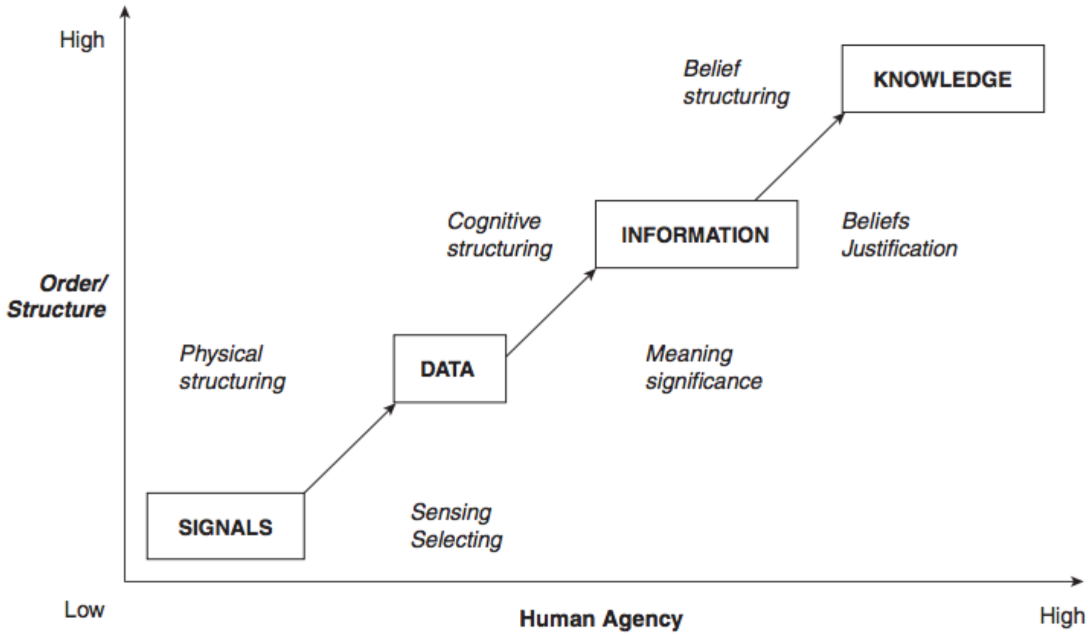


Figure 65: An alternative representation of the DIKW pyramid
Source: Choo, 1996

Extension 3: Beyond the buzz: definitions

1. Data Science

"Data science is the extraction of actionable knowledge directly from data through a process of discovery, or hypothesis formulation and hypothesis testing." (NIST, 2015)

As figure 66 shows, data science is at the intersection of three distinct disciplines:

- **Mathematics/Statistics**: above all a data scientist needs a deep, fundamental understanding of the statistical models used to extract information from data. His/her knowledge should range from simple summary statistics to complex machine learning algorithms.

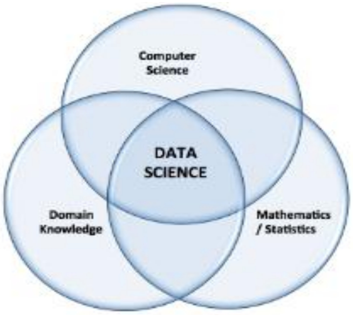


Figure 66: Data Science Venn-Diagram.
Source: Dartmouth (2016)

- **Domain knowledge**: a data scientist should know his way around the business he is working for. He/she can interpret the data in the relevant context.

- **Computer Science**: a data scientist has to be a fabulous programmer. For example: working with very big data sets requires strong coding skills to massage the raw data into a workable format.

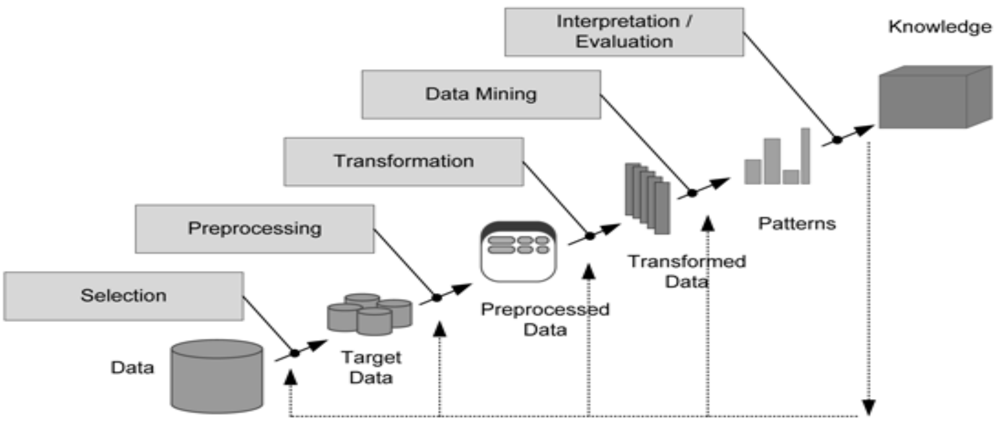


Figure 67: The data science process
Source: Costagliola, Fuccella, Giordano, & Polese (2009)

Figure 67 shows a process often mentioned in the data science literature that transforms raw data to knowledge. We will not go into the details as stages of the process have been discussed thoroughly in the literature (Baesens, 2014) (Provost & Fawcett, 2013).

However, there is a debate going on the added value of the term data science. Is data science really that different from statistics? Some observers belief it is not.

- “Data-scientist is just a sexed up term for a statistician. Statistics is a branch of science. Data scientist is slightly redundant in some way and people shouldn’t berate the term statistician. "(Nate Silver: What I need from statisticians - Statistics Views", 2013)

- "Data science is applied statistics, but in San Francisco." (Taylor, 2016)

- “There is very little truly new under the sun but the paradigms under which we collect and analyze data are beginning to change and it is incumbent upon us to recognize that.” (Wolfe, 2015)

However, there seems to be a growing consensus in the industry that the two terms are distinct enough to justify the existence of both of them. By analyzing the job openings for statisticians and data scientists we observe an incredible surge in the demand for data scientists. The traditional statisticians on the other hand seem to become old-fashioned. Why is it then that businesses are looking to hire this new role, what superpowers does a data scientist possess?

Figure 68: Job trends related to data scientists and statisticians
Source: "What’s the Difference Between Data Science and Statistics?" (2015)

It are mainly programming abilities that make a data scientist a much more valuable asset than a traditional statistician. "The largest part of the problem of getting value from data is the collection, manipulation and clean up of the data. Then there is a small amount of statistics and afterwards there is a large amount of interpretation and presenting results to business people. Let's not overestimate the role of statistics in this process. There is a lot of other messy stuff that we better think about as well … we have to start thinking about the entire pipeline, going from raw messy data to insights" (Ghahramani, 2015).

As data scientists spend up to 80% of their time on collecting and preprocessing data, it is important that he has all the necessary skills to do this as efficiently as possible.

Statistics has been developed in a pre-computing era and therefore its methods and techniques are most appropriate for dealing with small and structured data sets. Data science has been developed more recently and is therefore more suited to accommodate problems and needs that arise from current datasets. ("What's the Difference Between Data Science and Statistics?", 2015)

Data science is thus the field that takes the entire data cycle into account. From raw data to business value. Data scientists develop tools to clean data, to process data, to apply models to the data and to visualize the results. Examples are distributed computing with Hadoop to handle huge data sets or incorporating real-time social media data into models.

- Data scientists: better statisticians than most programmers & better programmers than most statisticians (Driscoll, 2012)

Conclusion

Conceptually, there is nothing new to data science. Both statistics and data science try to extract valuable knowledge from data. However, data science has a very strong focus on new techniques that take the entire data life cycle into account and make it possible to work with enormous, unstructured datasets.

2. Big Data

"Nowadays, all data is big"

Joeri Arts (2016), Business Development manager SAS

*"Big data is high **volume**, high **velocity**, and/or high **variety** information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization"* (Laney, 2001).

Laney (2001) was the first scholar to properly define the term big data. His definition, categorizing big data into three dimensions (3V's) has become the quintessential textbook definition.

- ***Volume***: refers to the extremely large amount of data stored, unworkable for traditional techniques. Facebook for example has a data warehouse of 300 Petabytes to store all incoming data (Wilfong & Vagata, 2016).

    Note: Since the volume of data created as well as its storage capacity is expanding massively year after year, it makes little sense to define big in a fixed amount of (peta)bytes. A more pragmatic approach to define the size of big data is given by Kohlschütter (2011) *"Data is "Big Data" if you need to hire a team of smart engineers to just handle its storage."*

- ***Velocity:*** refers to the high speed of new data entering the system. Walmart for example processes up to 40 Petabytes per day from their 250 million weekly customers (van Rijmenam, 2015).
- ***Variety:*** while traditional databases contain structured data in a very limited number of formats, big data refers to data in any format imaginable, including social media data, clickstream data, video, audio…

Common extensions to Laney (2001) include:

- ***Veracity:*** included in IBM's definition of big data (Zikopoulos, deRoos, Bienko, Buglio, & Andrews, 2015). This refers to the quality and correctness of the data. It poses the question to what extent the data can be trusted to base important decisions on.
- ***Variability***: included in SAS's definition of big data. "Refers to changes in rate and nature of data gathered by use case" (NIST, 2015).
- ***Value:*** refers to what extent the data is useful for the company (Biehn, n.d.).

As Laney (2013) and Grimes (2013) correctly point out these last three V's are not just characteristics of big data, they are characteristics of all data. Therefore, it does not make much sense to include them in the definition; however, these V's do pinpoint some of the crucial issues every company involved with big data should keep in mind. A similar argument could be made with respect to the original 3V's (Volume, Velocity, Variety). Without a clear benchmark (which is impossible for the abovementioned reason) there is no way of objectively differentiating between data and big data.

"Big data is the emerging field where innovative technology offers new ways of extracting value from the tsunami of new information." (Cavanillas, Curry, & Wahlster, 2016)

"Big data used to mean data that a single machine was unable to handle. Now big data has become a buzzword to mean anything related to data analytics or visualization." (Swanstrom, 2014)

Ultimately, the academic contribution of the term big data could be questioned. It has become an umbrella term for the data itself, but also for all techniques and applications related to it. Current definitions fail to clarify the ambiguity surrounding the term. In our view, the term is a convenient way to communicate differences between the obsolete way of dealing with data & analytics and its modern equivalent. The term will become redundant once companies get used to the new size of their databases, the new methods to govern them and the new techniques extract value from them. In 2015, Gartner removed the term Big Data from their Hype Cycle for emerging Technologies. "However, this is not a sign that this field is obsolete it just means big data is now the new normal" (Douglas, 2016).

Conclusion: in this thesis we will not distinguish between data and big data. However, the V's offer insight to common problems in these datasets that cannot be disregarded.

Extension 4: Data & Analytics related C-levels

1. The Chief Data Officer

As discussed in chapter 1, data is increasingly being regarded as a strategic corporate asset. Just as other crucial assets like money or people have their specific business executives some companies feel it makes sense to create a data related C-level position: the Chief Data Officer (Coutuer, 2017).

1.1 The emergence of the position

According to Leaper (2014) the Chief Data Officer position is created out of a long-lasting dispute between IT and other departments. These other departments generated most of the data and used them for their operations but it was unclear who owned the data and thus was responsible for its governance. This inconvenience was tackled by introducing a new position that explicitly governs the data and makes it available for all business units by bridging the gap between IT and the functional leaders (Gambill, 2016). Although the first Chief Data Officers were appointed in the early 2000s, the position only became mainstream after the financial crisis of 2007-2008. Many financial institutions hired a Chief Data Officer to comply with tightened regulatory requirements concerning data quality, accuracy or privacy that were brought to life after the crisis (defensive purposes in our terminology) (Bean, 2016).

1.2 Definition

"The Chief Data Officer is a senior executive who bears responsibility for the firm's enterprise wide data and information strategy, governance, control, policy development, and effective exploitation. The Chief Data Officer's role will combine accountability and responsibility for information protection and privacy, information governance, data quality and data life cycle management, along with the exploitation of data assets to create business value (McCall, 2015)."

"This individual [the Chief Data Officer] is charged with establishing and maintaining data governance, quality, architecture, and analytics—enabling firms to harness information to manage risk and create revenue generating opportunities (PwC, 2015)."

"The CDO is the executive who holds the keys to help an organization both protect and unlock the full value of its data assets" (Deloitte, 2016).

PwC (2015) divides the responsibilities from the abovementioned definition into three categories. Deloitte (2016) defines four categories (shown between brackets) of which the first three coincide with PwC (2015).

1. **Data governance (=operator):** developing and monitoring a company wide data governance strategy and policy. This includes defining metadata, data quality standards and implementing procedures that ensure correct and timely data is easily accessible. This category also includes compliance with data related legislation like the GDPR.

2. **Data architecture and technology (=technologist):** this refers to processes that standardize the collection, aggregation, storage and consumption of data across all business units and the technology necessary to support these mechanisms. Deloitte (2016) presses the importance of an enterprise wide approach that connects departments and dismantles silos.

3. **Data analytics (=strategist):** molding the processes that allow a company to extract value from data effectively and efficiently. For the success of a Chief Data Officer it is of upmost importance to work closely together with the different departments in which these solutions will be implemented.

4. **(=catalyst):** this category is not defined by PwC (2015). Deloitte (2016) explicitly identifies this aspect of the Chief Data Officer's job description to indicate that his/her responsibilities go further than only data governance or the implementation of analytics software. The Chief Data Officer has a highly innovative and strategic role that is reflected in the development of new products or services or the reinvention of operational processes. The Chief Data Officer has to think outside the box and be a great change manager.

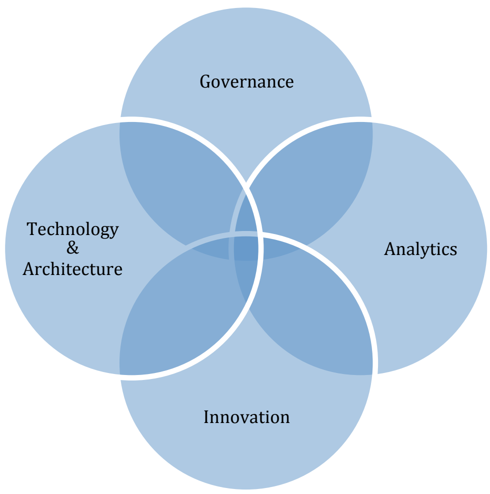


Figure 69: Four dimensions of responsibilities for the Chief Data Officer
Source: Author

This vanilla Chief Data Officer description will be fine-tuned according to the defensive/offensive strategy tradeoff. For example, some of our interviewees pointed out Chief Data Officers in the banking sector are strongly focused on data governance (defensive strategy) as this sector faces very demanding regulatory requirements (such as BCBS 239). One of the advantages of this legislation is that banks are forced to quickly reach a high data maturity. This characteristic can in turn be used to leverage other data & analytics capabilities in non-compliance related activities.

The previous breakdown reveals the width of the responsibility spectrum of a Chief Data Officer. On the one hand, he or she is responsible for compliance and governance (defense), which require strong organizational and detail-oriented skills. On the other hand he or she is responsible for innovation (offense), which requires strong creative and visionary skills. Bean (2016) describes this tension field as a challenging balance between "defensive practicality and offensive transformation". Finding someone who can encompass this responsibility spectrum is not evident. Hamers (2017) interprets this straddle as an opportunity: a Chief Data Officer can leverage the proven value of data & analytics substantiated by innovative pilot projects to convince other departments and stakeholders of the great importance of conscientiously data governance. Although this strategy is potentially very effective, it is not the path Chief Data Officers follow in practice. In 2015, almost 80% of the Chief Data Officers indicated their initial focus was on data governance (PwC, 2015). This leads to the conclusion that Chief Data Officers are mainly hired for compliance related reasons rather than for innovation related reasons, which is in line with the evolution of the Chief Data Officer as described by Deloitte (2016) and shown in figure 70. This approach is initially somewhat limiting the potential of a Chief Data Officer, but it is understandable given that non-compliance sanctions for the GDPR (that will enter into application in May 2018) can reach up to 4% of the annual worldwide turnover. Gartner (2016) finds that the first priority of Chief Data Officers is revenue generation rather than governance; this is again in line with the Chief Data Officer's evolution depicted in figure 70. In the future the strategic responsibilities of a Chief Data Officer are expected to increase as data becomes a key competitive differentiator (Deloitte, 2016; Spittaels 2017).

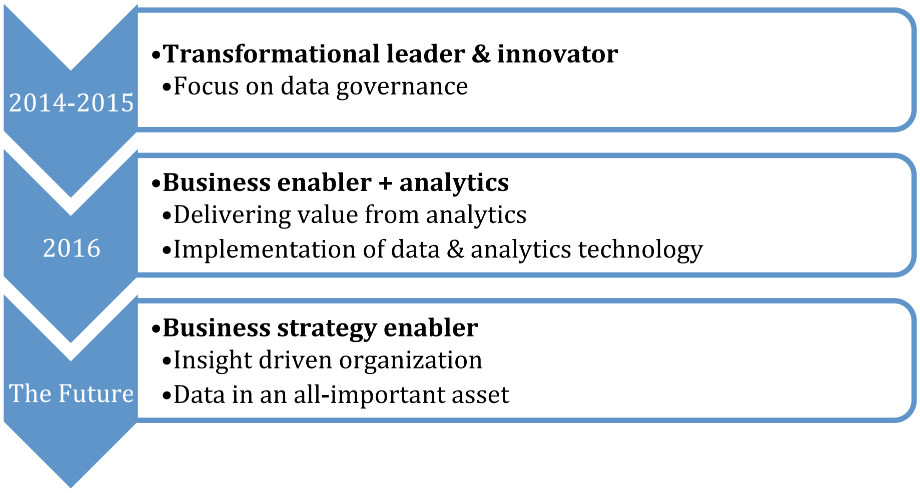


Figure 70: The role of the Chief Data Officer continues to evolve
Source: Deloitte, 2016

1.3 The rise of the Chief Data Officer

The number of companies with a Chief Data Officer on their payroll increased significantly during the last few years as shown in figure 71. NewVantage Partners (2016) finds that more than 50% of companies have a Chief Data Officer in their ranks. It is worth mentioning that their sample largely consists of firms in the financial sector. Forrester (2015) nuances these results slightly; they find that 45% of firms have a Chief Data Officer. Interestingly they also find top firms are 65% more likely to appoint a Chief Data Officer. This result is consistent with Xu, Zhan, Huang, Luo, & Xu (2016), who find that companies with a Chief Data Officer have superior financial results in comparison with those who do not. An enormous increase in Chief Data Officers is also the conclusion of research by Gartner (2016). Remarkably, they make the bold prediction that 90% of large firms will have appointed a Chief Data Officer by 2019 but only half of those will be successful. Lastly, it is worth pointing out that 85% of Chief Data Officers are active in the US or the UK as depicted in figure 72.

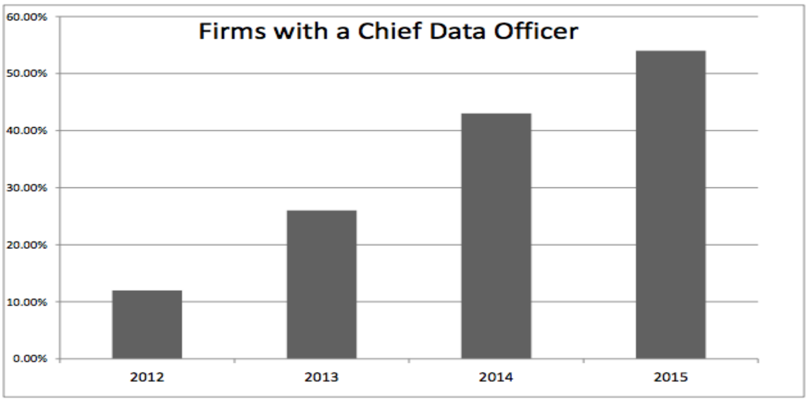


Figure 71: Firms with a Chief Data Officer
Source: NewVantage Partners (2016)

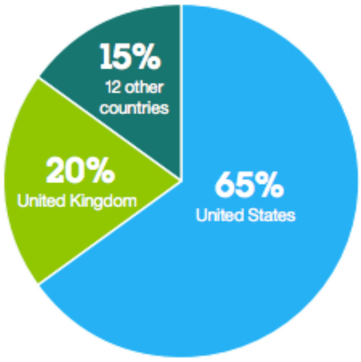


Figure 72: Chief Data Officers by geography
Source: IBM (2014)

2. The Chief Analytics Officer (CAO)

The Chief Analytics Officer is the latest data & analytics related C-level position that has been brought to life. Therefore, significantly less research on this position has been conducted (Agarwal, 2015). Figure 73 backs up this claim. In March 2017, the Google searches on the Chief Analytics Officer represented less than 20% of those of the Chief Data Officer or Chief Digital Officer. In this section we will first discuss the emergence of the position and its definition then shed a light on the difference between the CAO and the Chief Data Officer and lastly discuss its further evolution.

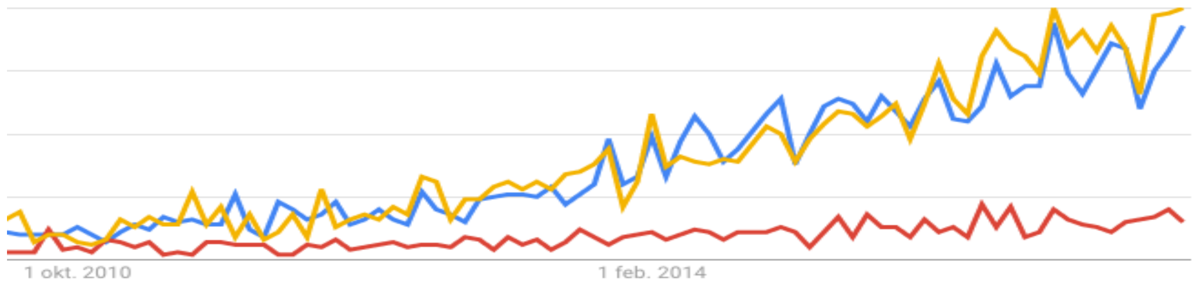


Figure 73: the relative interest in CAOs (red line), Chief Data Officers (blue line) and Chief Digital Officers (yellow line) between 2010 and 2014.
Source: Google trends (2017)

2.1 The emergence of the position

As discussed previously, the Chief Data Officer became mainstream after the financial crisis. Many companies invested in technology and infrastructure to better govern their data assets and appointed a Chief Data Officer to facilitate this process. After the initial compliance needs were covered the Chief Data Officer expanded its scope by leveraging the data for other business goals. This is where the Chief Analytics Officer comes into play. Suer (2015) and O'Regan (2014) do not believe one and the same person can encompass the broad responsibility spectrum described in the previous section. Therefore, a Chief Analytics Officer is appointed to deal with the innovative side of analytics while the Chief Data Officer can fully focus on governance as depicted in figure 74. "Peanut butter and chocolate may work in a Reese's cup but it will not work here—the orientations are too different (Suer, 2015)." Others interpret the CAO as a natural evolution of the Chief Data Officer (Bien, 2014). As data governance will become less of a challenge the former Chief Data Officer will focus more on solving business problems with data & analytics and become a Chief Analytics Officer. Note that except from a difference in nomenclature this transformation is identical to that described in the previous section.

2.2 Definition

"The CAO is a business strategist who knows the flow of information, understands its context, and is aware of how it links across the enterprise. He or she uses analytics to capitalize on the data to make sound decisions and achieve better outcomes for the business (Foo, 2013)."

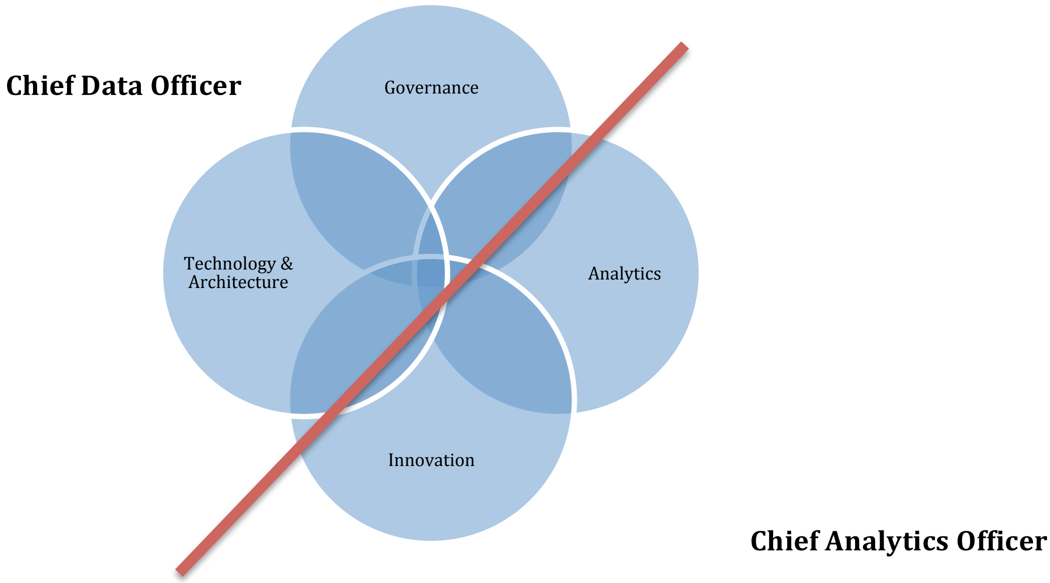


Figure 74: Four dimensions of responsibilities for the Chief Data Officer revisited
Source: Author

3. The Chief Digital Officer

"By 2019, 41% of revenue will come from digital marketing and ecommerce."

(Burtchell, 2016)

The last new C-level position that we are going to discuss in debt is the Chief Digital Officer. Although this role is not as closely linked with data & analytics as the previous ones it is an often-mentioned term in the literature and in the industry. It is valuable to clearly identify its responsibilities as its acronym is identical to that of the Chief Data Officer and the two terms could, mistakenly, be used interchangeably.

3.1 The emergence of the Chief Digital Officer

Since the introduction of the Internet, digital has become an increasingly more important aspect of doing business. In the early 2000s digital platforms were just a gimmick that added little value. This changed dramatically during the last decade. Because of the democratization of the Internet, the explosion of e-commerce and later the introduction of smartphones and apps digital has become such a crucial aspect of the corporate strategy that it cannot be treated as a side project (Caudron, & Peteghem, 2015). Furthermore, a C-level executive is sometimes considered necessary to fully leverage the potential of this portfolio of new, digital technologies. Figure 75 clearly illustrates the abovementioned evolution. Deloitte (2015) predicts that the Chief Digital Officer will have been disappeared by 2020 as the digital and corporate strategy have become indistinguishable. We will return to this point later in this chapter.

Figure 75: The convergence of corporate and digital strategy and the role of the Chief Digital Officer
Source: Deloitte, 2015

3.2 Definition

"The Chief Digital Officer is an individual who helps a company drive growth by converting traditional "analog" businesses to digital ones, and by overseeing operations in the rapidly-changing digital sectors like mobile applications, social media and related applications, virtual goods, as well as "wild" web-based information management and marketing." (Conneally, 2013)

"The Chief Digital officer is in charge of the digital business strategy" (Gartner, 2012)

"[The Chief Digital Officer] takes care of digital innovation both externally, in the companies' interactions with customers, partners, and suppliers, and internally, collecting and analyzing data, improving efficiency through the use of digital technologies, and transforming organization and culture to enable their companies to compete successfully in the digital age." (Strategy&, 2015)

Although these definitions cover the main responsibilities of the Chief Digital Officer many observers agree there is no one size fits all definition. The position has been interpreted many different times in different industries (Deloitte, 2015).

Deloitte (2015) tries to overcome this ambiguity by introducing three types of Chief Digital Officers.

> **Ex Agency: "***Traditional interactive marketing leaders that view digital as digital marketing and engagement with the customer.***"**
>
> This Chief Digital Officer interpretation is most common in fast moving consumer goods.

Digital Transformation Strategist: "*Change agents chartered with the reinvention of their organizations.*"

This Chief Digital Officer interpretation is most common in media, high tech and to a less extent retail.

Technologists: "*Those who view digital primarily from an enterprise perspective – most often reporting to the CIO.*"

This Chief Digital Officer interpretation is most common in manufacturing, oil, healthcare and other heavy industries.

Strategy& (2015) does not make this distinction between three types of Chief Digital Officers, but the three categories of responsibilities they attribute to the Chief Digital Officer strongly resemble the breakdown by Deloitte (2015), as depicted in figure 76. The Chief Digital officer is responsible for introducing new products and services (≈ Digital transformation strategist), digitizing the customer relationship (≈ Ex Agency) and revising operations (≈Technologist). The main levers to the disposal of a Chief Digital Officer to accomplish these goals are:

- Culture and change management
- New organizational structures and governance
- Technical infrastructure and data

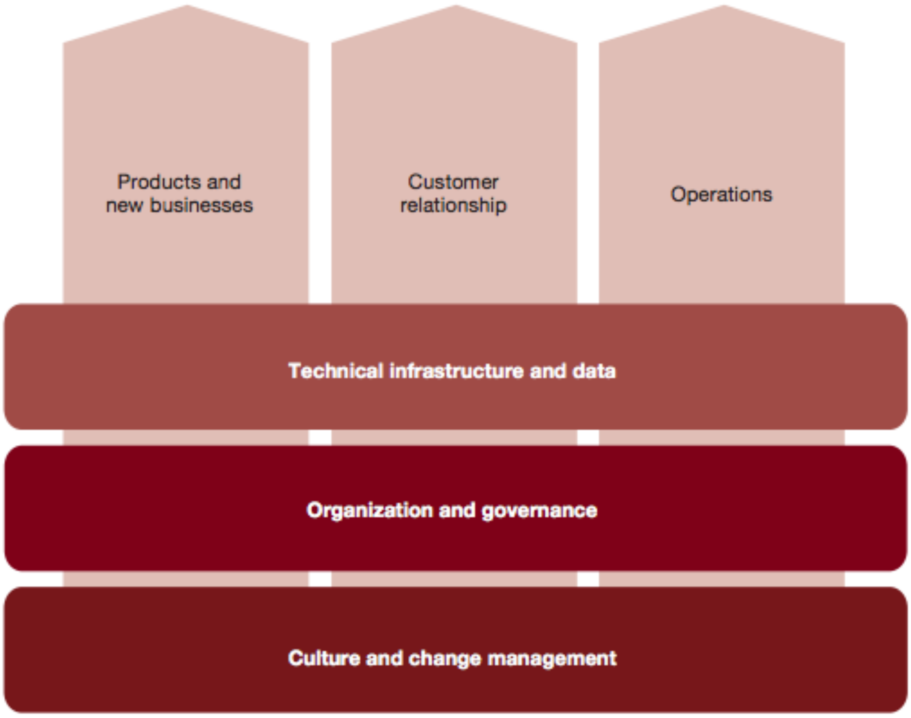


Figure 76: The responsibilities of the Chief Digital Officer
Source: Strategy&, 2015

3.3 The rise of the Chief Digital Officer

The number of Chief Digital Officers increased dramatically during the last decade as shown in figure 77. Gartner (2012) predicted that 25% of all organizations would have a Chief Digital Officer by 2015. This prediction turned out to be widely optimistic as only 6% of the world's 1500 largest businesses had a Chief Digital Officer among their ranks in 2015 (Strategy &, 2015).

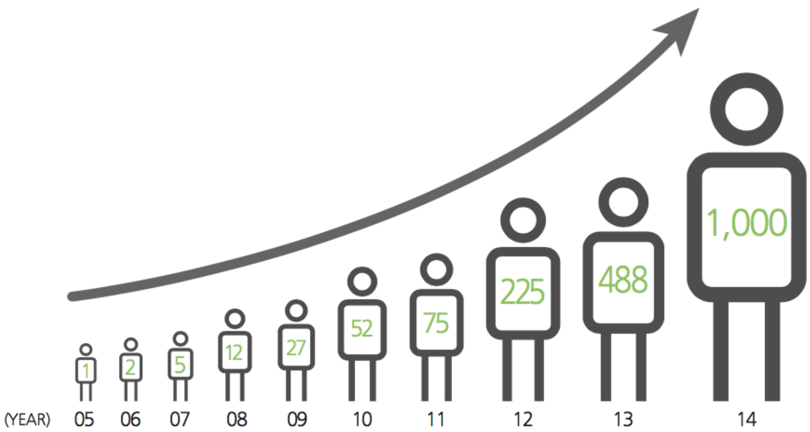


Figure 77: The rise of the Chief Digital Officer
Source: (Deloitte, 2015)

The difference between the Chief Digital Officer adoption rates across industries is significant as illustrated in figure 78. Chief Digital Officer positions are most frequent in the communications, media and entertainment industry (13% of those companies have a Chief Digital Officer) followed by the food/beverage/agriculture industry (11% of companies have a Chief Digital Officer). The metals/mining industry is in last place: only 1% of those companies have a Chief Digital Officer. Remarkably only 3% of tech companies have a Chief Digital Officer in their ranks. This seems counterintuitive but as these companies typically have a very high digital maturity, it is very possible that all the responsibilities of a Chief Digital Officer are already distributed among other C-suite members. Next, a clear correlation exists between company size and Chief Digital Officer adaptation: the bigger the firm, the higher the probability of having a Chief Digital Officer. 8% of large companies (>100.000 employees) employ a Chief Digital Officer while only 2% of small companies (< 1000) employees have one. The digital transformation process is deemed to be far more complex in huge companies and they therefore require a dedicated c-level executive to facilitate change (Strategy&, 2015).

The Chief Digital Officer is often interpreted as a transformational manager that leads the digital changeover of a company, as digital has become an inextricable aspect of doing business (Johnston, 2016). A question that can be asked is whether this manager is still an added value for a company once digital is flowing through all its arteries.

Industry cluster	No	Yes	Total	Percentage
Communications/media/entertainment	75	11	86	13%
Food/beverages/agriculture	73	9	82	11%
Consumer products/retail/wholesale	133	13	146	9%
Insurance	105	9	114	8%
Transport/travel/tourism	96	8	104	8%
Banking	178	13	191	7%
Pharma/health/chemicals	107	7	114	6%
Technology/electronics	88	3	91	3%
Automotive/engineering/machinery	106	3	109	3%
Utilities/oil/gas	196	5	201	2%
Other	161	4	165	2%
Metals and mining	96	1	97	1%
Total	**1,414**	**86**	**1,500**	**6%**

Figure 78: Percentage of companies with a Chief Digital Officer by industry
Source: Strategy&, 2015

3.4 The disappearance of the Chief Digital Officer

"Digital will become so infused with the business that it will make no more sense to have a separate leader and separate team than it does now to have a Chief Email Officer"

(Deloitte, 2015).

Samuels (2016), Deloitte (2015) and Strategy& (2015) agree that the surge of Chief Digital Officers was mainly due to the lack of digital maturity in many organizations. Now that a better understanding of digital has been developed and integrated into many business processes the need for a transformational manager disappears as illustrated in figure 79. For example a Chief Digital Officer that was responsible for the development of digital marketing strategy hands these activities over to the Chief Marketing Officer now the road has been paved. A similar logic can be followed for all the other responsibilities of the Chief Digital Officer. They are peeled off and given to other C-level members leading to a dismantling of the position. A crucial remark that has to be made is that digital innovation is far from over. New technologies and applications, like for example virtual reality/Internet of things and many others are already finding their way into the mainstream. In order to keep up with the competition it is crucial that companies have someone that is actively monitoring these new developments and implementing them in the business processes where possible.

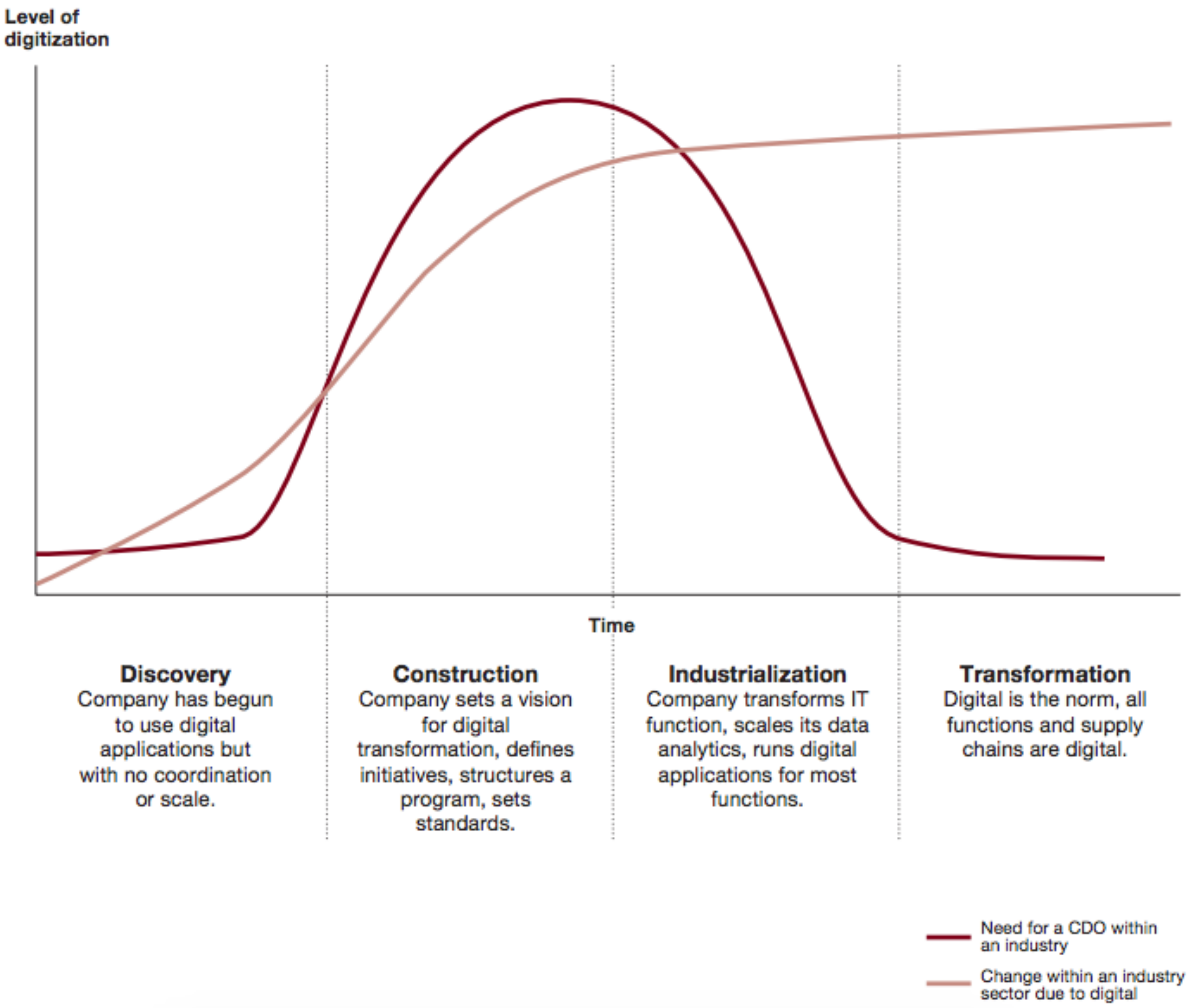


Figure 79: The need for a Chief Digital Officer during each phase of the digital revolution
Source: Strategy&, 2015

5. Conclusion

To wrap up this chapter we will give a brief overview of the characteristics of each of the new data & analytics related C-level positions.

Chief Data Officer

The chief Data Officer is the oldest and most common data & analytics related C-level officer. Research estimates that around 50% of companies have this officer in their ranks. Possibly geographical factors play a role as 85% of these executives are employed in the US or the UK.

His/her responsibilities are both offensive and defensive although the initial emphasis lies on defense and shifts to offense when the data governance related challenges have been solved. This narrative could be confirmed by our data as the Chief Data Officers in the sample are focused on offence. Companies in all clusters except cluster E hire Chief Data Officers.

Chief Analytics Officer

The Chief Analytics Officer is the most recent and least common data & analytics related position. In the literature, this role is interpreted as the sidekick of the Chief Data Officer. By having a Chief Analytics Officer that fully focuses on offensive data & analytics initiatives the Chief Data Officer can concentrate on data governance. As mentioned before this is not what we see in practice as offensive initiatives are the top priority of both the Chief Data Officer and the Chief Analytics Officer although the responsibility of the latter are more technical. Lastly, only the mature D cluster employs Chief Analytics Officers.

Chief Digital Officer

Around 28% of companies employ a Chief Digital Officer. This executive is less directly related to data & analytics than that of the previous two positions. The Chief Digital Officer is a transformational manager that has to guide a company through the digital revolution. Strategy& (2015) predicts the disappearance of this position once a company has become digitally mature. This evolution is in line with the empiric analysis, as the most mature companies do no hire Chief Digital Officers.

General remarks

The C-level support for data & analytics initiatives is significant in almost 65% of the companies in the sample. Throughout the last few years, the Chief Digital Officer, The Chief Data Officer and the Chief Analytics Officer have received significant attention. However, most data & analytics initiatives are still being led by the more traditional C-level officers, in particular the Chief Marketing Officer, the Chief Operations Officer and the Chief Executive Officer. The evidence whether these new positions increase the effectiveness of companies in their data & analytics efforts is mixed and requires further research.

Ultimately the job titles do not matter. What is truly vital however is that a company has the necessary skills somewhere to deal with an ever changing business environment (Morgan, 2016).

Extension 5: K-means algorithm

Figure 80 illustrates the mechanics of this algorithm. Firstly, k random data points are appointed as centroids (in this example k=2) as shown in (b) (red x and blue x). Then the (Euclidian) distance between each data point and these k centroids is calculated. Next, clusters around the centroids are formed as each data point is assigned to its closest centroid as shown in (c). After this first clustering, the cluster centers (centroids) are recalculated by averaging the values of all the data points in the clusters as shown in (d). Now the second iteration starts with these new centroids and the data points are again assigned to the closest centroid (e). The algorithm stops when there is no longer any change in cluster centers (f).

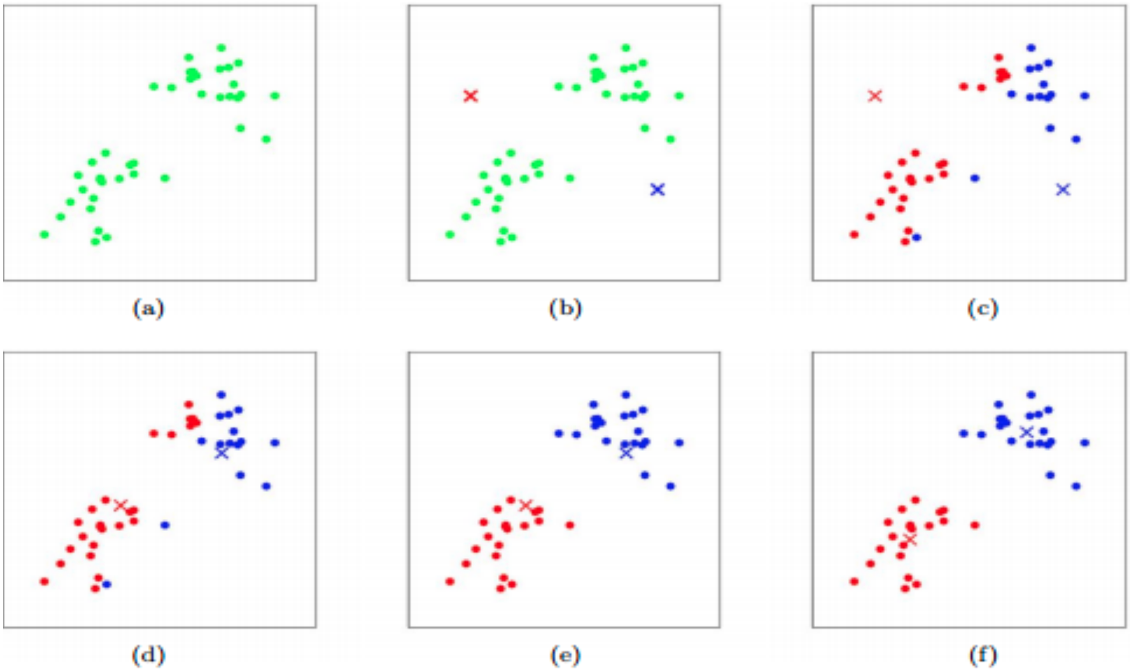


Figure 80: K-Means clustering algorithm explained
Source: Piech, 2013

Extension 6: variable description

Self indicated strategy maturity: this variable corresponds to the question "how would you describe your current data analytics strategy?" There were five possible answers:

- A mature and enterprise wide data analytics strategy with a strong focus on continuous experimentation and improvement **(High)**
- A mature and enterprise wide data analytics strategy
- An enterprise wide data analytics strategy exists, but not yet fully aligned with all business units **(Medium)**
- Ad hoc data analytics strategies in certain business units **(Low)**
- No data analytics strategy

To save space in the table we recoded these answers to high, medium of low. This variable corresponds to the "targets" dimension.

 # Plans: this variable corresponds to the question "why would your company implement a data analytics strategy?" There we several possible answers including:

- To create new revenue streams
- To improve internal efficiency & cut costs
- To better understand the customer
- To improve risk and compliance management
- To increase cyber security
- …

The value for this variable is simply the sum of reasons for data & analytics implementation and reflects the ambition and vision of a company. This variable corresponds to the "targets" dimension.

 # Impact: this variable corresponds to the question "on which of the following domains does the current data analytics strategy already have a noticeable impact?" The possible answers were similar to those of the previous question. The value for this variable is the sum of domains that were already impacted (for example increased cyber security, improved risk and compliance management). This variable corresponds to the "targets" dimension.

 # Impacted business activities: this variable corresponds to the question "in what business activities are data & analytics already being used?" Possible answers include marketing, finance, HRM, R&D… The value for this variable is again the sum of business activities in which data & analytics are already implemented. This variable corresponds to the “targets” dimension.

 # Challenges: this variable corresponds to the question “what are the biggest challenges your firm faces towards the implementation of data analytics strategies?” Possible answers include a lack of budget, a lack of skill, low data quality… The value for this variable is the sum of challenges listed by the respondent. This variable corresponds to the “targets” dimension.

 # Solutions: this variable corresponds to the question “how will these challenges be tackled?” Possible answers include hiring new employees, internal trainings/workshops, external consulting… The value for this variable is the sum of solutions listed by the respondent. This variable corresponds to the “targets” dimension.

 # Data formats: this variable corresponds to the question “what kind of data does your company collect?” Possible answers include text, audio, geospatial data, clickstream data… The value for this variable is the sum of data formats listed by the respondent. This variable corresponds to the “data” dimension.

 # Data sources: this variable corresponds to the question “what sources does your company use to collect data?” Possible answers include email, call center, website, 3rd party… The value for this variable is the sum of data sources listed by the respondent. This variable corresponds to the “data” dimension.

Strong C-level sponsorship: this variable corresponds to the question “there is strong C-level support for data analytics projects.” There were five possible answers: strongly agree, agree, neither agree nor disagree, disagree, strongly disagree. This variable corresponds to the “leadership” dimension.

Data analytics projects: this variable corresponds to the question “what analytics projects does your company set up?” There were 5 possible answers:

- Individuals use ad-hoc analytics projects to tackle specific issues
- We have automated analytics projects within certain departments
- We have automated analytics processes at an enterprise level **(Medium)**
- We both automated analytics processes at an enterprise level and experimental teams that explore innovative analytic methods **(High)**
- Don’t know

To save space in the table we have recoded these values to medium and high. This variable corresponds to the "enterprise" dimension.

 # Analysis techniques: this variable corresponds to the question: "which data analysis techniques are being used in your company?" Possible answers include descriptive statistics, predictive statistics, prescriptive statistics… The value for this variable is the sum of data techniques listed by the respondent. This variable corresponds to the "analysts" dimension.

Team diversification: this variable corresponds to the question "which roles are represented in your analytics team?" Possible answers include business analysts, data scientists,…The value for this variable is the sum of roles present in the data & analytics team. This variable corresponds to the "analysts" dimension.

Bibliography

Accenture,. (2014). *Big Success With Big Data*. Retrieved from https://www.accenture.com/ie-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_14/Accenture-Big-Data-POV.pdf

Accenture,. (2016). *Analytics Everywhere Smarter Actions, Happier Customers, Greater Value*. Retrieved from https://www.accenture.com/us-en/~/media/PDF-26/Accenture-CMT-Analytics-Everywhere-July18.pdf

Accenture. (2010). *Counting on Analytical Talent*. Retrieved from http://www.criticaleye.com/insights-servfile.cfm?id=2125&view=1

Achenwall, Gottfried. (2012). *Infoplease.com*. Retrieved 30 December 2016, from http://www.infoplease.com/encyclopedia/people/achenwall-gottfried.html

Ackoff, R. (1989). From data to wisdom. *Journal Of Applied Systems Analysis*, *16*, 3-9.

Agarwal, G. (2015). *5 Megatrends Driving the Chief Analytics Officer Role*. *GoodData*. Retrieved 21 March 2017, from https://www.gooddata.com/blog/5-megatrends-driving-chief-analytics-officer-role

Ament, P. (2006). *Abacus History - Invention of the Abacus*. *Ideafinder.com*. Retrieved 28 February 2017, from http://www.ideafinder.com/history/inventions/abacus.htm

Analysis. (2016). *Oxford Dictionary*. Retrieved from https://en.oxforddictionaries.com/definition/analysis

Analytics. (2016). *Oxford Dictionary*. Retrieved from https://en.oxforddictionaries.com/definition/analytics

Andrews, K. (1980). *The concept of corporate strategy* (1st ed.). R. D. Irwin, Homewood, Ill.

Armstrong, J. & Overton, T. (1977). Estimating Nonresponse Bias in Mail Surveys. *Journal Of Marketing Research*, *14*(3), 396. http://dx.doi.org/10.2307/3150783

Arts, J. (2016). Antwerpen.

AT Kearny,. (2013). *Big Data and the Creative Destruction of Today's Business Models*. Retrieved from https://www.atkearney.com/documents/10192/698536/Big+Data+and+the+Creative+Destruction+of+Todays+Business+Models.pdf/f05aed38-6c26-431d-8500-d75a2c384919

Augur, H. (2016). *Can Big Data Save Your Love Life? Online Dating Apps Say "Yes" - Dataconomy*. *Dataconomy*. Retrieved 11 April 2017, from http://dataconomy.com/2016/02/can-big-data-save-your-love-life/

Baesens, B. (2014). *Analytics in a big data world* (1st ed.). Hoboken (N.J.): J. Wiley & Sons.

Bain. (2013). *Big Data: The organizational challenge*. Retrieved from http://www.bain.com/Images/BAIN_BRIEF_Big_Data_The_organizational_challenge.pdf

Bean, R. (2016). *Big Data And The Emergence Of The Chief Data Officer*. *Forbes.com*. Retrieved 13 March 2017, from https://www.forbes.com/sites/ciocentral/2016/08/08/big-data-and-the-emergence-of-the-chief-data-officer/#571819bbba94

Beck, A., Sangoi, A., Leung, S., Marinelli, R., Nielsen, T., & van de Vijver, M. et al. (2011). Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival. *Science Translational Medicine*, *3*(108), 108ra113-108ra113. http://dx.doi.org/10.1126/scitranslmed.3002564

Becker, J. (2010). Maturity Models in IS Research. *ECIS 2010 Proceedings*, *42*. Retrieved from http://aisel.aisnet.org/ecis2010/42/

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management. *Business & Information Systems Engineering*, *1*(3), 213-222. http://dx.doi.org/10.1007/s12599-009-0044-5

Becker, J., Niehaves, B., Poeppelbuss, J., & Simons, A. (2010). Maturity Models in IS Research. *ECIS 2010 Proceedings*, *42*. Retrieved from http://aisel.aisnet.org/ecis2010/42

Beyond Moore's law. (2015). *The Economist*. Retrieved 6 February 2017, from http://www.economist.com/news/science-and-technology/21652051-even-after-moores-law-ends-chip-costs-could-still-halve-every-few-years-beyond

Biehn, N. *The Missing V's in Big Data: Viability and Value*. *Wired*. Retrieved 14 December 2016, from https://www.wired.com/insights/2013/05/the-missing-vs-in-big-data-viability-and-value

Bien, F. (2014). *It's Time To Welcome The Chief Analytics Officer To The C-suite | Fast Company*. *Fast Company*. Retrieved 21 March 2017, from https://www.fastcompany.com/3033590/the-future-of-work/its-time-to-welcome-the-chief-analytics-officer-to-the-c-suite

Box, G. (1976). Science and Statistics. *Journal Of The American Statistical Association*, *71*(356), 791-799. http://dx.doi.org/10.1080/01621459.1976.10480949

Bradshaw, J. *Why k-modes*. *Daylight.com*. Retrieved 8 April 2017, from http://www.daylight.com/meetings/mug04/Bradshaw/why_k-modes.html

Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.1819486

Bulygo, Z. (2013). *How Netflix Uses Analytics To Select Movies, Create Content, & Make Multimillion Dollar Decisions*. Retrieved 5 March 2017, from https://blog.kissmetrics.com/how-netflix-uses-analytics/

Buringh, E. & Van Zanden, J. (2009). Charting the "Rise of the West": Manuscripts and Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth Centuries. *The Journal Of Economic History*, *69*(02), 417. http://dx.doi.org/10.1017/s0022050709000837

Burnton, S. (2012). *50 stunning Olympic moments No28: Dick Fosbury introduces 'the flop'*. *the Guardian*. Retrieved 28 February 2017, from https://www.theguardian.com/sport/blog/2012/may/08/50-stunning-olympic-moments-dick-fosbury

Burtchell, M. (2016). *The Rise of The Chief Digital Officer: Why You Need One and What to Look For*. *Salsify.com*. Retrieved 8 March 2017, from http://www.salsify.com/blog/the-rise-of-the-chief-digital-officer-why-you-need-one-and-what-to-look-for

CA Technologies,. (2015). *The State of Big Data Infrastructure: Benchmarking global Big Data users to drive future performance*. Retrieved from https://www.ca.com/content/dam/ca/us/files/industry-analyst-report/the-state-of-big-data-infrastructure.pdf

Capgemini Consulting,. (2014). *Cracking the Data Conundrum: How Successful Companies Make Big Data Operational*. Retrieved from https://www.capgemini-consulting.com/resource-file-access/resource/pdf/cracking_the_data_conundrum-big_data_pov_13-1-15_v2.pdf

Carifio, J., & Perla, R. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal Of Social Sciences*, *3*(3), 106-116. http://dx.doi.org/10.3844/jssp.2007.106.116

Caudron, J., & Peteghem, D. (2015). *Digital transformation* (1st ed.). Duval Union Consulting.

Cavanillas, J., Curry, E., & Wahlster, W. (2016). New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe. *Springer Open*. http://dx.doi.org/10.1007/s40558-016-0061-4

Choo, C. (1996). The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International Journal Of Information Management*, *16*(5), 329-340. http://dx.doi.org/10.1016/0268-4012(96)00020-5

Cleveland, H. (1985). *The knowledge executive* (1st ed.). New York: Truman Talley Books.

Conneally, T. (2013). *'Chief Digital Officer' is the next hot executive title, says Gartner. BetaNews*. Retrieved 8 March 2017, from https://betanews.com/2012/10/22/chief-digital-officer-is-the-next-hot-executive-title-says-gartner/

Cosic, R., Shanks, G., & Maynard, S. (2012). Towards a Business Analytics Capability Maturity Model. Retrieved from https://pdfs.semanticscholar.org/7ff9/78d5dd970101c681383fcb168b49efee654a.pdf

Cost of Computing Power Equal to an iPad2 | The Hamilton Project. (2011). *Hamiltonproject.org*. Retrieved 6 February 2017, from http://www.hamiltonproject.org/charts/cost_of_computing_power_equal_to_an_ipad2

Costagliola, G., Fuccella, V., Giordano, M., & Polese, G. (2009). Monitoring Online Tests through Data Visualization. *IEEE Transactions On Knowledge And Data Engineering*, *21*(6), 773-784. http://dx.doi.org/10.1109/tkde.2008.133

Coutuer, J. (2017). Data & Analytics maturity interview. Brussels.

Dartmouth, U. (2016). *Master of Science - UMass Dartmouth. Umassd.edu*. Retrieved 31 December 2016, from http://www.umassd.edu/datascience/masterofscience/

Data. (2016). *Oxford Dictionary*. Retrieved from https://en.oxforddictionaries.com/definition/data

DataFlux,. (2007). *The Data Governance Maturity Model Establishing the People, Policies and Technology That Manage Enterprise Data*. Retrieved from http://www.fstech.co.uk/fst/whitepapers/The_Data_Governance_Maturity_Model.pdf

Davenport, T. (2013). *Enterprise analytics* (1st ed.). Upper Saddle River, N.J.: FT Press.

Davenport, T. & Patil, D. (2012). *Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review*. Retrieved 29 December 2016, from https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

Davenport, T. & Prusak, L. (2000). *Working knowledge* (1st ed.). Boston, Mass.: Harvard Business School Press.

Davenport, T., & Dallemule, L. (2017). What's Your Data Strategy?. *Harvard Business Review*, (May-June), 112-121. Retrieved from https://hbr.org/2017/05/whats-your-data-strategy

Davenport, T., & Harris, J. (2007). *Competing on analytics* (1st ed.). Boston (Mass.): Harvard Business School Press.

Davenport, T., Harris, J., & Morison, R. (2010). *Analytics at work* (1st ed.). Boston, Mass: Harvard Business Press.

Davidian, M. & Louis, T. (2012). Why statistics?. *Science*, *336*. Retrieved from https://www.stat.wisc.edu/~wahba/science.whystat.pdf

Davis, G. & Olson, M. (1985). *Management information systems* (1st ed.). New York: McGraw-Hill.

de Brí, F. (2009). An e-Government Stages of Growth Model Based on Research Within the Irish Revenue Offices. *Electronic Journal Of E-Government*, *7*(4), 339 - 348.

Deloitte,. (2015). *The rise of the Chief Digital Officer*. Retrieved from http://www.deloittedigital.ca/chief-digital-officer

Deloitte,. (2016). *The evolving role of the chief data officer in financial services*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/uy/Documents/strategy/gx-fsi-evolving-role-of-chief-data-officer.pdf

DeLucia, N. (2016). *The History of Printing. Inksterinc*. Retrieved 30 December 2016, from http://www.inksterprints.com/inksterprints-blog/the-history-of-printing

Descriptive, Predictive, and Prescriptive Analytics Explained. (2016). *Halo*. Retrieved 31 December 2016, from https://halobi.com/2016/07/descriptive-predictive-and-prescriptive-analytics-explained/

Deseure, A. (2017). Data & Analytics maturity. Antwerpen.

Devens, R. (1865). *Cyclopaedia of commercial and business anecdotes* (1st ed.). New York: D. Appleton and Company.

Donnelly, C., & Simmons, G. (2013). Small Businesses Need Big Data, Too. Harvard Business Review. Retrieved 17 March 2017, from https://hbr.org/2013/12/small-businesses-need-big-data-too

Douglas, T. (2016). *The rise and fall of big data hype—and what it means for customer intelligence*. *Vision Critical*. Retrieved 29 December 2016, from https://www.visioncritical.com/fall-of-big-data-hype/

Drexler, K. (2013). *Radical abundance: How a Revolution in Nanotechnology Will Change Civilization* (1st ed.). PublicAffairs,U.S.

Driscoll, M. (2012). *Michael E. Driscoll on Twitter*. *Twitter*. Retrieved 15 January 2017, from https://twitter.com/medriscoll/status/225333351526051842?ref_src=twsrc%5Etfw

Drucker, P. (2008). *Classic drucker* (1st ed.). Boston: Harvard Business Press.

EY & Forbes Insights,. (2015). *Analytics: don't forget the human element*. Retrieved from http://www.ey.com/Publication/vwLUAssets/EY-Forbes-Insights-Data-and-Analytics-Impact-Index-2015/$FILE/EY-Forbes-Insights-Data-and-Analytics-Impact-Index-2015.pdf

Fallows, J. (2013). *The 50 Greatest Breakthroughs Since the Wheel*. *The Atlantic*. Retrieved 28 February 2017, from https://www.theatlantic.com/magazine/archive/2013/11/innovations-list/309536/

Finnemann, N. (2001). *The Internet – A New Communicational Infrastructure* (1st ed.). Skrifter fra Center for Internetforskning. Retrieved from http://web.archive.org/web/20040328165322/http://cfi.imv.au.dk/pub/skriftserie/002_finnemann.pdf

Foo, A. (2013). *The Emerging Role of the Chief Analytics Officer*. *IBM Big Data & Analytics Hub*. Retrieved 21 March 2017, from http://www.ibmbigdatahub.com/blog/emerging-role-chief-analytics-officer

Forrester,. (2015). *Top Performers Appoint Chief Data Officers*. Retrieved from https://www.forrester.com/report/Top+Performers+Appoint+Chief+Data+Officers/-/E-RES123064

Forrester,. (2016). *The Digital Maturity Model 4.0*. Retrieved from https://www.forrester.com/report/The+Digital+Maturity+Model+40/-/E-RES130881

Fraser, P., Moultrie, J., & Gregory, M. (2002). The use of maturity models / grids as a tool in assessing product development capability: a review. *IEEE International Engineering Management Conference*. Retrieved from http://www.academia.edu/3391259/The_use_of_maturity_models_grids_as_a_tool_in_assessing_product_development_capability

Gambill, M. (2016). *What Is A Chief Data Officer And Why Do Firms Need Them?*. *Forbes.com*. Retrieved 13 March 2017, from https://www.forbes.com/sites/kimberlywhitler/2016/10/22/what-is-a-chief-data-officer-and-why-do-firms-need-them/#5de1e61bc94a

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal Of Information Management*, *35*(2), 137-144. http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007
Gartner,. (2012). *Gartner Says Every Budget is Becoming an IT Budget*. *Gartner.com*. Retrieved 8 March 2017, from http://www.gartner.com/newsroom/id/2208015

Gartner,. (2016). *Gartner CDO Survey Reveals That Chief Data Officers Drive Both Data Management and Analytics for Maximum Impact*. Retrieved from https://www.gartner.com/doc/3514317

Gartner,. (2016). *Gartner Estimates That 90 Percent of Large Organizations Will Have a Chief Data Officer by 2019*. *Gartner.com*. Retrieved 14 March 2017, from http://www.gartner.com/newsroom/id/3190117

Gartner,. (2017). *Hype Cycle Research Methodology | Gartner Inc.*. *Gartner.com*. Retrieved 13 April 2017, from http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp
Gartner's analytic value escalator. (2016). *Flickr*. Retrieved 31 December 2016, from https://www.flickr.com/photos/27772229@N07/8267855748/in/photostream

Ghahramani, Z. (2015). *Data Science and Statistics: different worlds?*. *YouTube*. Retrieved 15 January 2017, from https://www.youtube.com/watch?v=C1zMUjHOLr4

Ghosh, S., Little, R., & Rubin, D. (1988). Statistical Analysis with Missing Data. *Technometrics*, *30*(4), 455. http://dx.doi.org/10.2307/1269814

Gibson, C. & Nolan, R. (1974). Managing the Four Stages of EDP Growth. *Harvard Business Review*. Retrieved from https://hbr.org/1974/01/managing-the-four-stages-of-edp-growth

Granville, V. (2015). *Data Scientist vs Statistician*. *Datasciencecentral.com*. Retrieved 2 January 2017, from http://www.datasciencecentral.com/profiles/blogs/data-scientist-vs-statistician

Green, S. (2011). *Business Jargon Is Not a "Value-Add"*. *Harvard Business Review*. Retrieved 28 February 2017, from https://hbr.org/2011/12/business-jargon-is-not-a-value

Grimes, S. (2013). *Big Data: Avoid 'Wanna V' Confusion - InformationWeek*. *InformationWeek*. Retrieved 14 December 2016, from http://www.informationweek.com/big-data/big-data-analytics/big-data-avoid-wanna-v-confusion/d/d-id/1111077?page_number=1

Guess, A. (2015). *Only 0.5% of All Data is Currently Analyzed - DATAVERSITY*. *DATAVERSITY*. Retrieved 7 February 2017, from http://www.dataversity.net/only-0-5-of-all-data-is-currently-analyzed/

Gung,. (2015). *How to use both binary and continuous variables together in clustering?*. *Stats.stackexchange.com*. Retrieved 8 April 2017, from http://stats.stackexchange.com/questions/130974/how-to-use-both-binary-and-continuous-variables-together-in-clustering

Halevi, G. & Moed, H. (2012). *The Evolution of Big Data as a Research and Scientific Topic* (1st ed.). Research Trends. Retrieved from https://www.researchtrends.com/wpcontent/uploads/2012/09/Research_Trends_Issue30.pdf

Hamers, B. (2017). Data & Analytics maturity. Brussels.

Harari, Y. (2011). *Sapiens* (1st ed.). Vintage Books.

Harris, J. (2016). *History of Storage from Cave Paintings to Electrons (Infographic) - Remo Software - Info*. *Remo Software - Info*. Retrieved 17 January 2017, from http://www.remosoftware.com/info/history-of-storage-from-cave-paintings-to-electrons

Hartshorne, C. & Weiss, P. (1934). Collected Papers of Charles Sanders Peirce. *The Journal Of Philosophy*, *31*(21), 582. http://dx.doi.org/10.2307/2015395

Henry, N. (1974). Knowledge Management: A New Concern for Public Administration. *Public Administration Review*, *34*(3). http://dx.doi.org/10.2307/974902

Herbsleb, J., Zubrow, D., Goldenson, D., Hayes, W., & Paulk, M. (1997). Software quality and the Capability Maturity Model. *Communications Of The ACM*, *40*(6), 30-40. http://dx.doi.org/10.1145/255656.255692

Hortonworks,. (2016). *Big data maturity model*. Retrieved from http://hortonworks.com/wp-content/uploads/2016/04/Hortonworks-Big-Data-Maturity-Assessment.pdf

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining And Knowledge Discovery*, *2*, 283-304. Retrieved from http://www.cs.ust.hk/~qyang/Teaching/537/Papers/huang98extensions.pdf

IBM - What is big data?. (2012). Retrieved 30 January 2017, from https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

IBM,. (2007). *Data Governance Council Maturity Model*. Retrieved from https://www-935.ibm.com/services/uk/cio/pdf/leverage_wp_data_gov_council_maturity_model.pdf

IBM,. (2016). *IBM Clustering binary data with K-Means (should be avoided) - United States*. *Www-01.ibm.com*. Retrieved 8 April 2017, from http://www-01.ibm.com/support/docview.wss?uid=swg21477401

IBM. (2014). *The new hero of big data and analytics*. Retrieved from http://www-07.ibm.com/au/pdf/GBE03607USEN9.pdf
Ifrah, G. (2001). *The universal history of computing* (1st ed.). New York: John Wiley.

Information. (2003). *Merriam-Webster*. Retrieved from https://www.merriam-webster.com/dictionary/information

Information. (2016). *Oxford Dictionary*. Retrieved from https://en.oxforddictionaries.com/definition/information

Inmon, W. (2016). *Data lake architecture* (1st ed.). Basking Ridge, NJ: Technics Publications.

Iversen, J., Nielsen, P., & Norbjerg, J. (1999). Situated assessment of problems in software development. *ACM SIGMIS Database*, *30*(2), 66-81. http://dx.doi.org/10.1145/383371.383376

Jayaraman, S. (2012). *Enabling Strategic Decision Making: The role of Decision Support Systems in banks | Fintellix Blog*. *Fintellix.com*. Retrieved 4 April 2017, from http://www.fintellix.com/blog/?p=117

Johnston, C. (2016). *The Rise of the Chief Digital Officer - Disruption*. *Disruption*. Retrieved 8 March 2017,

Jones, G. (2010). *Organizational theory, design, and change* (1st ed.). Upper Saddle River, N.J.: Pearson.

Khan, S. (2016). *Can we use K-means for clustering binary vectors?*. *Quora*. Retrieved 8 April 2017, from https://www.quora.com/Can-we-use-K-means-for-clustering-binary-vectors

Knowledge. (2016). *Oxford Dictionary*. Retrieved from https://en.oxforddictionaries.com/definition/knowledge

Knudson, K. (2016). *Use Your Fingers: The Abacus Just Might Improve Your Arithmetic Abilities*. *Forbes.com*. Retrieved 29 December 2016, from http://www.forbes.com/sites/kevinknudson/2016/04/28/use-your-fingers-the-abacus-just-might-improve-your-arithmetic-abilities/#4d54a989590d

Kohlschütter, C. (2011). *How much data is "Big Data"?*. *Quora*. Retrieved 14 December 2016, from https://www.quora.com/How-much-data-is-Big-Data

Kopalle, P. (2014). *Why Amazon's Anticipatory Shipping Is Pure Genius*. *Forbes.com*. Retrieved 5 March 2017, from https://www.forbes.com/sites/onmarketing/2014/01/28/why-amazons-anticipatory-shipping-is-pure-genius/#744fee184605

Kowalke, P. (2016). *Six Ways to Use Big Data If Your Company is Small*. *It.toolbox.com*. Retrieved 18 March 2017, from http://it.toolbox.com/blogs/inside-erp/six-ways-to-use-big-data-if-your-company-is-small-73563

KPMG,. (2016). *Building trust in analytics*. Retrieved from https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2016/10/building-trust-in-analytics.pdf

KPMG,. *Effective master data management*. Retrieved from https://www.compact.nl/articles/effective-master-data-management/#Maturity_model_for_master_data_management

Lahrmann, G., Marx, F., Winter, R., & Wortmann, F. (2011). Business Intelligence Maturity: Development and Evaluation of a Theoretical Model. *Proceedings Of The 44Th Hawaii International Conference On System Sciences*. http://dx.doi.org/DOI: 10.1109/HICSS.2011.90

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. *Technical Report, META Group*. Retrieved from https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Laney, D. (2013). *Batman on Big Data - Doug Laney. Gartner*. Retrieved 14 December 2016, from http://blogs.gartner.com/doug-laney/batman-on-big-data/

Leaper, B. (2014). *The Rise of the Chief Data Officer. WIRED*. Retrieved 13 March 2017, from https://www.wired.com/insights/2014/07/rise-chief-data-officer/

Lee, J., Jones, P., Mineyama, Y., & Zhang, X. (2002). Cultural differences in responses to a likert scale. *Research In Nursing & Health*, *25*(4), 295-306. http://dx.doi.org/10.1002/nur.10041
Lewis, M. (2003). *Moneyball* (1st ed.). New York: W.W. Norton.

Lismont, J., Vanthienen, J., Baesens, B., & Lemahieu, W. (2017). Defining analytics maturity indicators: A survey approach. *International Journal Of Information Management*, *37*(3), 114-124. http://dx.doi.org/10.1016/j.ijinfomgt.2016.12.003

Marr, B. (2015). *Big Data: 20 Mind-Boggling Facts Everyone Must Read. Forbes.com*. Retrieved 6 February 2017, from http://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#4724296d6c1d

Marr, B. (2015). *Big Data: Using SMART Big Data; Analytics and Metrics To Make Better Decisions and Improve Performance* (1st ed.). Wiley.

Marr, B. (2015). *How small businesses can get started with big data. Hiscox Business Blog*. Retrieved 18 March 2017, from https://www.hiscox.co.uk/business-blog/bernard-marr-column-small-businesses-can-get-started-big-data/

Martens, D. (2017). Methodological question thesis.

Maturity. (2010). *Webster's New World College Dictionary*. Retrieved from http://www.yourdictionary.com/maturity#websters

Mazzei, M. & Noble, D. (2017). Big data dreams: A framework for corporate strategy. *Business Horizons*. http://dx.doi.org/10.1016/j.bushor.2017.01.010

McAfee, A. (2013). *Big Data's Biggest Challenge? Convincing People NOT to Trust Their Judgment. Harvard Business Review*. Retrieved 4 April 2017, from https://hbr.org/2013/12/big-datas-biggest-challenge-convincing-people-not-to-trust-their-judgment

McCall, T. (2015). *Understanding the Chief Data Officer (CDO) Role - Smarter With Gartner. Smarter With Gartner*. Retrieved 11 March 2017, from http://www.gartner.com/smarterwithgartner/understanding-the-chief-data-officer-role/

McElheran, K., & Brynjolfsson, E. (2016). *The Rise of Data-Driven Decision Making Is Real but Uneven. Harvard Business Review*. Retrieved 4 April 2017, from https://hbr.org/2016/02/the-rise-of-data-driven-decision-making-is-real-but-uneven

McKinsey,. (2013). *Game changers: Five opportunities for US growth and renewal*. Retrieved 21 March 2017, from http://www.mckinsey.com/global-themes/americas/us-game-changers

McKinsey,. (2013). *Mobilizing your C-suite for big-data analytics. McKinsey & Company*. Retrieved 21 March 2017, from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/mobilizing-your-c-suite-for-big-data-analytics

McKinsey. (2013). *Big data: What's your plan?. McKinsey & Company*. Retrieved 4 May 2017, from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-whats-your-plan

McKinsey. (2016). *Big data: Getting a better read on performance. McKinsey & Company*. Retrieved 29 April 2017, from http://www.mckinsey.com/industries/high-tech/our-insights/big-data-getting-a-better-read-on-performance

McLaughlin, P., & Sherouse, O. (2016). *The McLaughlin-Sherouse List: The 10 Most-Regulated Industries of 2014. Mercatus Center*. Retrieved 21 March 2017, from https://www.mercatus.org/publication/mclaughlin-sherouse-list-10-most-regulated-industries-2014

McMillan, R. (2012). *How Social Security Rescued IBM From Death by Depression. WIRED*. Retrieved 30 December 2016, from https://www.wired.com/2012/06/how-social-security-saved-ibm/

Mettler, T. (2011). Maturity assessment models: a design science research approach. *International Journal Of Society Systems Science*, *3*(1/2), 81. http://dx.doi.org/10.1504/ijsss.2011.038934

Mettler, T., & Rohner, P. (2009). Situational maturity models as instrumental artifacts for organizational design. *Proceedings Of The 4Th International Conference On Design Science Research In Information Systems And Technology - DESRIST '09*. http://dx.doi.org/10.1145/1555619.1555649

Miller, G., Gerlach, S., & Bräutigam, D. (2006). *Business intelligence competency centers* (1st ed.). Hoboken, N.J.: John Wiley & Sons, Inc.

Morgan, J. (2013). *Why Big Company Doesn't Mean Job Security. Forbes.com*. Retrieved 21 February 2017, from http://www.forbes.com/sites/jacobmorgan/2013/11/14/why-big-company-doesnt-mean-job-security/#35b01c0779fb

Morgan, L. (2016). *Rise And Fall Of The Chief Data Officer - InformationWeek. InformationWeek*. Retrieved 13 March 2017, from http://www.informationweek.com/strategic-cio/it-strategy/rise-and-fall-of-the-chief-data-officer/a/d-id/1324280

Mullany, M. (2016). *8 Lessons from 20 Years of Hype Cycles. LinkedIn*. Retrieved 13 April 2017, from https://www.linkedin.com/pulse/8-lessons-from-20-years-hype-cycles-michael-mullany

Nate Silver: What I need from statisticians - Statistics Views. (2013). *Statisticsviews.com*. Retrieved 30 December 2016, from http://www.statisticsviews.com/details/feature/5133141/Nate-Silver-What-I-need-from-statisticians.html

Neef, D. (2015). Why haven't SMEs cashed in on big data benefits yet?. SearchBusinessAnalytics. Retrieved 17 March 2017, from http://searchbusinessanalytics.techtarget.com/feature/Why-havent-SMEs-cashed-in-on-big-data-benefits-yet

NewVantage Partners,. (2016). *Big Data Executive Survey 2016*. Retrieved from http://newvantage.com/wp-content/uploads/2016/01/Big-Data-Executive-Survey-2016-Findings-FINAL.pdf

NIST,. (2015). *NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements*. Retrieved from http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf

Northouse, P. (2013). *Leadership* (1st ed., p. 5). Los Angeles [u.a.]: SAGE.

Number of Internet Users (2016) - Internet Live Stats. (2017). *Internetlivestats.com*. Retrieved 6 February 2017, from http://www.internetlivestats.com/internet-users/

O'Regan, R. (2014). *Chief analytics officer: The ultimate big data job?. Computerworld*. Retrieved 21 March 2017, from http://www.computerworld.com/article/2688352/chief-analytics-officer-the-ultimate-big-data-job.html

Ogbuokiri, B., Udanor, C., & Agu, M. (2015). Implementing bigdata analytics for small and medium enterprise (SME) regional growth. IOSR Journal Of Computer Engineering, Volume 17(Issue 6), 35-43. http://dx.doi.org/10.9790/0661-17643543

Ogden, J., & Lo, J. (2012). How meaningful are data from Likert scales? An evaluation of how ratings are made and the role of the response shift in the socially disadvantaged. *Journal Of Health Psychology*, *17*(3), 350-361. http://dx.doi.org/10.1177/1359105311417192

Paskach, C., & Johnston, D. (2017). Data Analytics Now Cost-Effective For Smaller Firms - Law360. Law360.com. Retrieved 17 March 2017, from https://www.law360.com/articles/893647/data-analytics-now-cost-effective-for-smaller-firms

Philippens, M. (2017). Data & Analytics maturity. Tervuren.
Piech, C. (2013). *CS221. Stanford.edu*. Retrieved 7 April 2017, from http://stanford.edu/~cpiech/cs221/handouts/kmeans.html

Plato, & Gallop, D. (1988). *Phaedo* (1st ed.). Oxford [Oxfordshire]: Clarendon Press.

Provost, F. & Fawcett, T. (2013). *Data science for business* (1st ed.). Sebastopol, Calif: O'Reilly.

PwC,. (2015). *Great expectations: The evolution of the chief data officer*. Retrieved from https://www.pwc.com/us/en/financial-services/publications/viewpoints/assets/pwc-chief-data-officer-cdo.pdf

PwC,. (2016). *PwC's Global Data and Analytics Survey 2016*. Retrieved from http://www.pwc.com/us/en/advisory-services/data-possibilities/big-decision-survey.html

Reeves, M. & Pueschel, L. (2015). *Die Another Day: What Leaders Can Do About the Shrinking Life Expectancy of Corporations. www.bcgperspectives.com*. Retrieved 15 January 2017, from https://www.bcgperspectives.com/content/articles/strategic-planning-growth-die-another-day/

Regalado, A. (2013). *Big Data Gets Personal. MIT Technology Review*. Retrieved 7 February 2017, from https://www.technologyreview.com/s/514346/the-data-made-me-do-it/

Rouse, M. (2013). *What is maturity grid (maturity model)? - Definition from WhatIs.com. WhatIs.com*. Retrieved 20 January 2017, from http://whatis.techtarget.com/definition/maturity-model

Rouse, M. (2014). *What is Kryder's Law? - Definition from WhatIs.com. SearchStorage*. Retrieved 5 March 2017, from http://searchstorage.techtarget.com/definition/Kryders-Law

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal Of Information Science*, *33*(2), 163-180. http://dx.doi.org/10.1177/0165551506070706

Samuels, M. (2016). *Is this really the end for the chief digital officer? | ZDNet. ZDNet*. Retrieved 9 March 2017, from http://www.zdnet.com/article/is-this-really-the-end-for-the-chief-digital-officer/

SAS,. (2007). *Five steps to evolving into an intelligent, high- performance enterprise*. Retrieved from http://www.eurim.org.uk/activities/ig/voi/IEM.pdf

SAS,. (2013). *Big Data in Big Companies*. Retrieved from http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf

Sen, D., Ozturk, M., & Vayvay, O. (2016). An Overview of Big Data for Growth in SMEs. *Procedia - Social And Behavioral Sciences*, *235*, 159-167. http://dx.doi.org/10.1016/j.sbspro.2016.11.011

Signal. (2016). *Oxford Dictionaty*. Retrieved from https://en.oxforddictionaries.com/definition/signal

Simon, P. (2013). Even Small Companies Can Tap Big Data If They Know Where to Look. Harvard Business Review. Retrieved 17 March 2017, from https://hbr.org/2013/12/even-small-companies-can-tap-big-data-if-they-know-where-to-look

Simonite, T. (2016). *The foundation of the computing industry's innovation is faltering. What can replace it?. MIT Technology Review*. Retrieved 11 April 2017, from https://www.technologyreview.com/s/601441/moores-law-is-dead-now-what/

STATS | YouTube Company Statistics - Statistic Brain. (2016). *Statistic Brain*. Retrieved 30 January 2017, from http://www.statisticbrain.com/youtube-statistics/

Steinert, M., & Leifer, L. (2010). Scrutinizing Gartner's hype cycle approach. *Technology Management For Global Economic Growth (PICMET)*. Retrieved from https://www.researchgate.net/publication/224182916_Scrutinizing_Gartner%27s_hype_cycle_approach

Stockton, N. (2016). *Elon Musk Announces His Plan to Colonize Mars and Save Humanity. Wired.com*. Retrieved 29 April 2017, from https://www.wired.com/2016/09/elon-musk-colonize-mars/

Strategy&,. (2015). *Adapt, disrupt, transform, disappear: The 2015 Chief Digital Officer Study*. Retrieved from http://www.strategyand.pwc.com/media/file/The-2015-chief-digital-officer-study.pdf

Strategy&,. (2015). *Companies without a dedicated leader of digital transformation may lose competitive advantage. Strategyand.pwc.com*. Retrieved 8 March 2017, from http://www.strategyand.pwc.com/uk/home/press_contacts/displays/cdo-studiy-uk

Stubbs, E. (2014). *Big Data, Big Innovation: Enabling Competitive Differentiation through Busi* (1st ed.). John Wiley & Sons.

Suer, M. (2015). *Is it the CDO or CAO or Someone Else?. The Informatica Blog - Perspectives for the Data Ready Enterprise*. Retrieved 21 March 2017, from https://blogs.informatica.com/2015/01/26/chief-data-officer-cao-someone-else-data-analytics/#fbid=tevSDKjvCQM

Sullivan, D. (2016). *Google now handles at least 2 trillion searches per year. Search Engine Land*. Retrieved 6 February 2017, from http://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247

Swanstrom, R. (2014). *What Is Big Data? - Blog. Datascience.berkeley.edu*. Retrieved 14 December 2016, from https://datascience.berkeley.edu/what-is-big-data

Systems, E. (2014). *You Have The Data, But You Need INFORMATION - Effectual Systems. Effectual Systems*. Retrieved 2 January 2017, from http://effectualsystems.com/data-need-information/

Taylor, D. (2016). *Battle of the Data Science Venn Diagrams. Kdnuggets.com*. Retrieved 30 December 2016, from http://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html

TDWI,. (2014). *TDWI Analytics Maturity Model Guide*. Retrieved from https://tdwi.org/whitepapers/2014/10/tdwi-analytics-maturity-model-guide.aspx

Teigen, R. (1997). *Supply Chain Management. Eil.utoronto.ca*. Retrieved 4 April 2017, from http://eil.utoronto.ca/wp-content/static/profiles/rune/node5.html

The Abacus: A Brief History. Ee.ryerson.ca. Retrieved 29 December 2016, from http://www.ee.ryerson.ca/~elf/abacus/history.html

The destruction of the Great Library of Alexandria. (2016). *Ancient Origins*. Retrieved 30 December 2016, from http://www.ancient-origins.net/ancient-places-africa-history-important-events/destruction-great-library-alexandria-001644

The Royal Swedish Academy of Sciences,. (1997). *Press Release*. Retrieved from http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1997/press.html

Thomas Burke (USA) - 100m. International Olympic Committee. Retrieved 28 February 2017, from https://www.olympic.org/news/speed-merchant-burke-shines-in-the-sprint

Tien, J. (2003). Toward a decision informatics paradigm: a real-time, information-based approach to decision making. *IEEE Transactions On Systems, Man And Cybernetics, Part C (Applications And Reviews)*, *33*(1), 102-113. http://dx.doi.org/10.1109/tsmcc.2003.809345

Tschabitscher, H. (2016). *Ever Wonder How Many Emails Get Sent Worldwide Every Day. Lifewire*. Retrieved 30 January 2017, from https://www.lifewire.com/how-many-emails-are-sent-every-day-1171210

Vaes, K. (2013). *Understanding ; Data, Knowledge, Information & Wisdom. Karim Vaes*. Retrieved 11 December 2016, from https://kvaes.wordpress.com/2013/05/31/data-knowledge-information-wisdom/

van Rijmenam, M. (2015). *Big Data at Walmart is All About Big Numbers; 40 Petabytes a Day!. Linkedin*. Retrieved 14 December 2016, from https://www.linkedin.com/pulse/big-data-walmart-all-numbers-40-petabytes-day-mark-van-rijmenam

Veryard, R. (2005). *Technology Hype Curve. Demandingchange.blogspot.be*. Retrieved 13 April 2017, from https://demandingchange.blogspot.be/2005/09/technology-hype-curve.html

Walker, M. (2012). *Data Veracity. Datasciencecentral.com*. Retrieved 14 December 2016, from http://www.datasciencecentral.com/profiles/blogs/data-veracity

Watson, H., Ariyachandra, T., & Matyska, R. (2001). Data Warehousing Stages of Growth. *Information Systems Management*, *18*(3), 42-50. http://dx.doi.org/10.1201/1078/43196.18.3.20010601/31289.6

Wersig, G. & Neveling, U. (1971). *Terminology of documentation* (1st ed.). Paris: UNESCO.

What Is Big Data? - Gartner IT Glossary - Big Data. (2012). *Gartner IT Glossary*. Retrieved 14 December 2016, from http://www.gartner.com/it-glossary/big-data/

What's the Difference Between Data Science and Statistics?. (2015). *Priceonomics*. Retrieved 2 January 2017, from https://priceonomics.com/whats-the-difference-between-data-science-and/

White, T. (2012). *Hadoop: The Definitive Guide* (1st ed.). O'Reilly Media.

Wiggins, C., Hand, D., Bennett, F., Wolfe, P., & Ghahramani, Z. (2015). *Data Science and Statistics: different worlds?. YouTube*. Retrieved 15 January 2017, from https://www.youtube.com/watch?v=C1zMUjHOLr4

Wilfong, K. & Vagata, P. (2016). *Scaling the Facebook data warehouse to 300 PB. Facebook Code*. Retrieved 14 December 2016, from https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/

Wisdom. (2016). *Oxford Dictionary*. Retrieved from https://en.oxforddictionaries.com/definition/wisdom

Wolchover, N. (2016). *Top 10 Inventions that Changed the World. Live Science*. Retrieved 28 February 2017, from http://www.livescience.com/33749-top-10-inventions-changed-world.html

Wolfe, P. (2015). *Data Science and Statistics: different worlds?. YouTube*. Retrieved 15 January 2017, from https://www.youtube.com/watch?v=C1zMUjHOLr4

Xu, F., Zhan, H., Huang, W., Luo, X., & Xu, D. (2016). The value of Chief Data Officer presence on firm performance. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1364&context=pacis2016

Zeleny, M. (1987). Management support systems: towards integrated knowledge management. *Human Systems Management*, *7*(1), 59-70.

Zeleny, M. (2013). Integrated Knowledge Management. *International Journal Of Information Systems And Social Change*, *4*(4), 62-78. http://dx.doi.org/10.4018/jissc.2013100104

Zikopoulos, P., deRoos, D., Bienko, C., Buglio, R., & Andrews, M. (2015). *Big data beyond the hype* (1st ed.). Mc Graw Hill Education.

Appendix 1: Full Survey

General questions

Q1 What is your job title?

Q2 What is the name of your company?

- (1) ________________
- I would rather do this survey anonymously. Note: the data will only be used for academic purposes and will never be shared with any third party. Company names will never be mentioned in the report, only general sector-wide conclusions will be made. (2)

Q3 What is the number of employees in your company?

- 1-10 (1)
- 10-50 (2)
- 50-250 (3)
- 250-1000 (4)
- >1000 (5)

Q4 What is the annual revenue of your company?

- (1)
- €1 million - €10 million (2)
- €10 million - €50 million (3)
- €50 million - €250 million (4)
- €250 million - €1 billion (5)
- >€1 billion (6)

Q5 What sector is your company in?

Q6 When was your company founded?

- Before 1950 (1)
- 1950-1990 (2)
- 1990-2010 (3)
- After 2010 (4)

Definitions:

Data analytics: refers to techniques and processes to discover, interpret and communicate patterns in information.

Targets

Q7 How important is data analytics to your organization?

- ❍ Extremely important (1)
- ❍ Very important (2)
- ❍ Moderately important (3)
- ❍ Slightly important (4)
- ❍ Not at all important (5)

Q8 Why would your company implement a data analytics strategy? (Multiple answers are possible)

- ❑ To survive & stay competitive in the sector (1)
- ❑ To increase revenue (21)
- ❑ To create new revenue streams (2)
- ❑ To improve internal efficiency & cut costs (3)
- ❑ To better understand the customer (4)
- ❑ To improve customer relationships (5)
- ❑ To improve marketing and customer targeting (6)
- ❑ To improve products and services (7)
- ❑ To improve the management of existing data (8)
- ❑ To monetize existing data (9)
- ❑ To find and exploit new data sources (10)
- ❑ To improve risk and compliance management (11)
- ❑ To increase cyber security (12)
- ❑ To monitor competitor behavior (13)
- ❑ To improve decision-making (14)
- ❑ To accelerate decision-making (15)
- ❑ For the digital transformation of the company (16)
- ❑ As a competitive differentiator (17)
- ❑ Others: (18) ____________________
- ❑ I don't know (19)

Q9 How would you describe your current data analytics strategy?

- ❍ A mature and enterprise wide data analytics strategy with a strong focus on continuous experimentation and improvement (1)
- ❍ A mature and enterprise wide data analytics strategy (2)
- ❍ An enterprise wide data analytics strategy exists, but not yet fully aligned with all business units (3)
- ❍ Ad hoc data analytics strategies in certain business units (4)
- ❍ No data analytics strategy (5)

Q10 On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible)

- ❑ There is no noticeable impact yet (20)
- ❑ On the ability to survive & stay competitive in the sector (1)
- ❑ On the ability to create new revenue streams (2)
- ❑ On internal efficiency & costs (3)
- ❑ On the ability to understand the customer (4)
- ❑ On customer relationships (5)
- ❑ On marketing and customer targeting (6)
- ❑ On the improvement of products and services (7)
- ❑ On data governance (8)
- ❑ On monetizing existing data (9)
- ❑ On the ability to find and exploit new data sources (10)
- ❑ On risk and compliance management (11)
- ❑ On cyber security (12)
- ❑ On the ability to monitor competitor behavior (13)
- ❑ On decision-making (14)
- ❑ On accelerating decision-making (15)
- ❑ On the digital transformation of the company (16)
- ❑ On the ability to differentiate from competitors (17)
- ❑ Others: (18) ____________________
- ❑ I don't know (19)

Q11 In what business activities are data & analytics already being used? (Multiple answers are possible)

- ❑ Sales (1)
- ❑ Marketing (2)
- ❑ HRM (3)
- ❑ Finance (4)
- ❑ Manufacturing (5)
- ❑ R&D (6)
- ❑ Supply chain management (7)
- ❑ Quality management (8)
- ❑ Risk management (9)
- ❑ Others: (10) ____________________
- ❑ Don't know (11)

Q12 What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible)

- ❑ Lack of budget (1)
- ❑ Lack of skill (2)
- ❑ Organizational structures (3)
- ❑ Organizational culture (4)
- ❑ Overwhelmed by the volume of the data (5)
- ❑ Low data quality (6)
- ❑ Difficulties demonstrating or monetizing the impact of data projects (7)
- ❑ Lack of C-level sponsorship (8)
- ❑ Lack of stakeholders' sponsorship (9)
- ❑ Data security issues (10)
- ❑ Privacy issues (11)
- ❑ Legal issues (14)
- ❑ Others: (12) ____________________
- ❑ Don't know (13)

Q13 How will these challenges be tackled? (Multiple answers are possible)

- ❑ Hiring new employees (1)
- ❑ Internal trainings/workshops (2)
- ❑ External consulting (3)
- ❑ Specific data analytics projects to prove value and effectiveness (4)
- ❑ First steps to make this a strategic priority are taken (5)
- ❑ Bigger budget (6)
- ❑ Implementation of new technologies (7)
- ❑ Change management (8)
- ❑ Others: (9) ____________________
- ❑ Don't know (10)

Q14 Does your company plan any additional data analytics related investments?

- ❍ Yes, within the next year (1)
- ❍ Yes but not within the next year (2)
- ❍ No (3)
- ❍ Don't know (4)

Q15 Which investments in which department?

Leadership

Q16 Who is in charge of the data analytics projects?

- ❍ Data scientist (1)
- ❍ Data Analyst (2)
- ❍ IT project manager (3)
- ❍ Innovation manager (4)
- ❍ Operations manager (5)
- ❍ Chief risk officer (6)
- ❍ Chief digital officer (7)
- ❍ Chief information officer (8)
- ❍ Chief data officer (9)
- ❍ Chief analytics officer (10)
- ❍ Chief marketing officer (11)
- ❍ Others: (12) ________________
- ❍ Don't know (13)

Q17 What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible)

- ❑ Data collection (1)
- ❑ Data governance (2)
- ❑ Data exploitation (3)
- ❑ Data security (4)
- ❑ Data analytics (5)
- ❑ Data integration (6)
- ❑ Monitoring data quality (7)
- ❑ Translating analysis into value (8)
- ❑ Finding new data sources (9)
- ❑ Managing the analytics department (10)
- ❑ Monitoring data analytics initiatives (11)
- ❑ Finding new business opportunities (12)
- ❑ Participating on the executive board (13)
- ❑ Establishing a data culture in the organization (14)
- ❑ Democratizing data tools across the organization (15)
- ❑ Leading workshops/seminars on data analytics (16)
- ❑ Helping to determine the strategic direction of the company (17)
- ❑ Improving the online presence of the firm (18)
- ❑ Communicating the firms data analytics strategy internally and externally (19)
- ❑ Change management (20)
- ❑ Sponsoring digitalization or automation (21)
- ❑ Develop ways to attract and retain highly skilled data talents (22)
- ❑ Coordinating data related investments (25)
- ❑ Others: (23)
- ❑ Don't know (24)

Q18 What is the educational background of your Chief Digital Officer/Chief Data Officer or Chief Digital officer?

- Math/physics (1)
- IT/computer science (2)
- Data science/Statistics (3)
- Economics/Business (4)
- Others: (5) ____________________
- Don't know (6)

Q19 How long has his role been in place in your company?

- More than 4 years (1)
- More than 2 years (2)
- More than 1 year (3)
- Less than 1 year (4)
- Others (5) ____________________
- Don't know (6)

Q20 Does your company plan on hiring a Chief data officer, a Chief analytics officer, a Chief digital officer or another equivalent position?

- Yes, within the next year (1)
- Yes but not within the next year (2)
- No (3)
- Don't know (4)

Q21 There is strong C-level support for data analytics projects

- Strongly agree (1)
- Agree (2)
- Neither agree nor disagree (3)
- Disagree (4)
- Strongly disagree (5)

Q22 Who is the sponsor of data analytics projects? (Highest in the corporate hierarchy)

- ❍ Chief executive officer (1)
- ❍ Chief marketing officer (2)
- ❍ Chief financial officer (3)
- ❍ Chief human resource officer (4)
- ❍ Chief operations officer (5)
- ❍ Chief risk officer (6)
- ❍ Chief technology officer (7)
- ❍ Vice president (8)
- ❍ Others: (9) ____________________
- ❍ Don't know (10)

Data

Q23 What kind of data does your company collect? (Multiple answers are possible)

- ❑ Structured data (tables, records) (1)
- ❑ Text data (2)
- ❑ Geospatial data (3)
- ❑ Audio (4)
- ❑ Video (5)
- ❑ Weblogs (6)
- ❑ Clickstream data (7)
- ❑ Streaming data (8)
- ❑ Others: (9) ____________________
- ❑ I don't know (10)

Q24 What sources does your company use to collect data? (Multiple answers are possible)

- ❑ Point of sale (1)
- ❑ Social media (2)
- ❑ Back office systems (ERP) (3)
- ❑ Email (4)
- ❑ Call center (5)
- ❑ Website (6)
- ❑ 3rd party (7)
- ❑ Mobile apps (8)
- ❑ Sensors (9)
- ❑ Others: (10) ____________________
- ❑ Don't know (11)

Q25 Employees of any department who want to use data to discover new insights can easily access the necessary data (if they are entitled to get it).

- Strongly agree (1)
- Agree (2)
- Neither agree nor disagree (4)
- Disagree (5)
- Strongly disagree (8)

Q26 Roles and responsibilities for data governance are clearly defined.

- Strongly agree (1)
- Agree (2)
- Neither agree nor disagree (4)
- Disagree (5)
- Strongly disagree (8)

Q27 My company has a clear view of the volume of data stored and gathered

- Strongly agree (1)
- Agree (2)
- Neither agree nor disagree (4)
- Disagree (5)
- Strongly disagree (8)

Q28 My company will have no difficulties complying with GDPR by May 2018 (General data protection regulation by the European Commission).

- Strongly agree (1)
- Agree (2)
- Neither agree nor disagree (4)
- Disagree (5)
- Strongly disagree (8)

Q29 Are there issues in your organization that arise as a result of poor data quality?

- We experience a wide range of issues that impact our organization. (1)
- We experience some challenges but it is up to each department to solve them. (2)
- We experience some issues, but we are always able to solve them at an enterprise wide level. (3)
- We have not experienced or no longer experience any challenges due to poor data quality at our organization. (4)
- Don’t know (5)

Q30 Does your company know what percentage of the data is wrong, inaccurate or outdated?

- ❍ No (1)
- ❍ We know the status of our data for specific, important data sets (2)
- ❍ Each department has a view on the status of the data and its quality (3)
- ❍ We have a clear view of the data status on an enterprise wide level (4)
- ❍ We have a clear view of the data status and we have progressive KPIs that ensure continuous improvement (5)
- ❍ Don't know (6)

Enterprise

Q31 What analytics projects does your company set up?

- ❍ Individuals use ad-hoc analytics projects to tackle specific issues (1)
- ❍ We have automated analytics projects within certain departments (2)
- ❍ We have automated analytics processes at an enterprise level (3)
- ❍ We both automated analytics processes at an enterprise level and experimental teams that explore innovative analytic methods (4)
- ❍ Don't know (5)

Analysts

Q32 We have feedback processes in place to assess the quality and accuracy of our analyses.

- ❍ Strongly agree (1)
- ❍ Agree (2)
- ❍ Neither agree nor disagree (4)
- ❍ Disagree (5)
- ❍ Strongly disagree (8)

Q33 Which data analysis techniques are being used in your company? (Multiple answers are possible)

- ❑ Descriptive statistics (1)
- ❑ Predictive statistics (2)
- ❑ Prescriptive statistics (3)
- ❑ Data mining (4)
- ❑ Machine learning (5)
- ❑ Data visualization (6)
- ❑ Others: (7)
- ❑ Don't know (8)

Q34 We have the right skills in place to address new big data infrastructure technologies for our big data efforts.

- ❍ Strongly agree (1)
- ❍ Agree (2)
- ❍ Neither agree nor disagree (4)
- ❍ Disagree (5)
- ❍ Strongly disagree (8)

Q35 What skills for data related positions are hardest to get by?

- ❑ Sector knowledge/experience (1)
- ❑ Creativity/innovation (2)
- ❑ Teamwork/motivational skills (3)
- ❑ Skills that bridge IT and business requirements (4)
- ❑ Analytical skills (5)
- ❑ Technical skills (6)
- ❑ Sales skills (7)
- ❑ Others: (8) ____________________
- ❑ Don't know (9)

Q36 How often is data used in decision-making?

- ❍ Always (1)
- ❍ Most of the time (2)
- ❍ About half the time (3)
- ❍ Sometimes (4)
- ❍ Never (5)

Q37 What decisions are influenced by analytics?

- ❍ None (1)
- ❍ Day-to-day decisions (2)
- ❍ Short to mid term tactical decisions (3)
- ❍ Long term strategic decisions (4)

Q38 A significant part of our data related investments goes to training front-line employees to use our models and their results

- ❍ Strongly agree (1)
- ❍ Agree (2)
- ❍ Neither agree nor disagree (3)
- ❍ Disagree (4)
- ❍ Strongly disagree (5)

Q39 Which roles are represented in your analytics team?

- ❑ Business analyst (1)
- ❑ Data scientist (2)
- ❑ IT expert (3)
- ❑ System architect (4)
- ❑ Others (5) ____________________
- ❑ Don't know (6)

Q40 If you want to receive the results of this research, please leave your e-mail below.

- ❍ Email: (1) ____________________
- ❍ I do not want to receive the results of this survey (2)

Appendix 2: Full output k-means clustering

Note that the cluster numbers do not correspond to the cluster letters. For the letters we have altered the sequence from low maturity to high maturity.

Final Cluster Centers

	Cluster				
	1	2	3	4	5
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To improve risk and compliance management	0	1	0	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To increase cyber security	0	0	0	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To monitor competitor behaviour	0	0	0	0	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To improve decision-making	0	1	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To accelerate decision-making	0	1	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice For the digital transformation of the company	0	1	1	1	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice As a competitive differentiator	0	0	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice Others:	0	0	0	0	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice I don't know	0	0	0	0	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To survive & stay competitive in the sector	0	1	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To create new revenue streams	0	1	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To improve internal efficiency & cut costs	0	1	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To better understand the customer	0	1	1	1	0

Final Cluster Centers

	Cluster 1	2	3	4	5
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To improve customer relationships	0	1	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To improve marketing and customer targeting	0	1	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To improve products and services	0	1	1	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To improve the management of existing data	0	0	1	0	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To monetize existing data	0	0	0	1	0
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To find and exploit new data sources	0	0	0	1	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
Why would your company implement a data analytics strategy? (Multiple answers are possible) - Selected Choice To increase revenue	0	1	1	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On risk and compliance management	0	0	0	0	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On cyber security	0	0	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On the ability to monitor competitor behaviour	0	0	0	0	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On decision-making	0	1	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On accelerating decision-making	0	0	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On the digital transformation of the company	0	0	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On the ability to differentiate from competitors	0	0	0	1	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice Others:	0	0	0	0	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice I don't know	0	0	0	0	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On the ability to survive & stay competitive in the sector	0	0	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On the ability to create new revenue streams	0	0	0	1	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On internal efficiency & costs	0	1	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On the ability to understand the customer	0	0	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On customer relationships	0	0	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On marketing and customer targeting	0	1	0	1	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On the improvement of products and services	0	0	0	1	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On data governance	0	0	0	0	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On monetizing existing data	0	0	0	0	0
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice There is no noticeable impact yet	0	0	0	0	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
On which of the following domains does the current data analytics strategy already have a noticeable impact? (Multiple answers are possible) - Selected Choice On the ability to find and exploit new data sources	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Privacy issues	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Others:	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Don't know	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Legal issues	0	0	0	0	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Lack of budget	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Lack of skill	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Organizational structures	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Organizational culture	0	0	1	1	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Overwhelmed by the volume of the data	0	0	0	0	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Low data quality	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Difficulties demonstrating or monetizing the impact of data projects	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Lack of C-level sponsorship	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Lack of stakeholders' sponsorship	0	0	0	0	0
What are the biggest challenges your firm faces towards implementation of data analytics strategies? (Multiple answers are possible) - Selected Choice Data security issues	0	0	0	1	0

	Cluster				
	1	2	3	4	5
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice Hiring new employees	0	1	0	0	0
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice Internal trainings/workshops	0	0	0	1	0
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice External consulting	0	0	0	0	0
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice Specific data analytics projects to prove value and effectiveness	0	1	1	0	1
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice First steps to make this a strategic priority are taken	0	0	0	0	0
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice Bigger budget	0	0	0	0	0
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice Implementation of new technologies	0	1	0	1	0
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice Change management	0	0	0	1	0

	Cluster 1	2	3	4	5
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice Others:	0	0	0	0	0
How will these challenges be tackled? (Multiple answers are possible) - Selected Choice Don't know	0	0	0	0	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Develop ways to attract and retain highly skilled data talents	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Others:	0	0	0	0	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Don't know	0	0	0	0	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Coordinating data related investments	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Managing the analytics department	0	0	0	0	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Monitoring data analytics initiatives	0	1	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Finding new business opportunities	0	0	0	0	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Participating on the executive board	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Establishing a data culture in the organization	0	0	0	0	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Democratizing data tools across the organization	0	0	0	0	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Leading workshops/seminars on data analytics	0	0	0	0	0

	Cluster				
	1	2	3	4	5
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Helping to determine the strategic direction of the company	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Improving the online presence of the firm	0	0	0	0	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Communicating the firms data analytics strategy internally and externally	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Data collection	0	1	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Data governance	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Data exploitation	0	0	0	1	0

	Cluster				
	1	2	3	4	5
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Data security	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Data analytics	0	1	0	1	1
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Data integration	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Monitoring data quality	0	0	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Translating analysis into value	0	1	0	1	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Finding new data sources	0	0	0	0	0
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Change management	0	0	0	1	0

	Cluster				
	1	2	3	4	5
What are the responsibilities of the person in charge of the data analytics projects? (Multiple answers are possible) Sponsoring digitalization or automation	0	0	0	1	0
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Structured data (tables, records)	0	1	1	1	1
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Text data	0	1	0	1	1
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Geospatial data	0	0	0	1	0
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Audio	0	0	0	1	0
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Video	0	0	0	1	0
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Weblogs	0	0	0	1	0
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Clickstream data	0	0	0	1	0
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Streaming data	0	0	0	1	0

	Cluster				
	1	2	3	4	5
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice Others:	0	0	0	0	0
What kind of data does your company collect? (Multiple answers are possible) - Selected Choice I don't know	0	0	0	0	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Don't know	0	0	0	0	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Point of sale	0	1	0	1	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Social media	0	0	0	1	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Back office systems (ERP)	0	1	1	1	1
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Email	0	1	0	0	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Call center	0	1	0	1	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Website	0	1	0	1	1

Final Cluster Centers

	Cluster 1	2	3	4	5
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice 3rd party	0	1	0	1	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Mobile apps	0	0	0	1	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Sensors	0	0	0	1	0
What sources does your company use to collect data? (Multiple answers are possible) - Selected Choice Others:	0	0	0	0	0
Which data analysis techniques are being used in your company? (Multiple answers are possible) Descriptive statistics	0	1	0	1	1
Which data analysis techniques are being used in your company? (Multiple answers are possible) Predictive statistics	0	1	0	1	1
Which data analysis techniques are being used in your company? (Multiple answers are possible) Prescriptive statistics	0	0	0	1	0
Which data analysis techniques are being used in your company? (Multiple answers are possible) Data mining	0	1	0	1	0
Which data analysis techniques are being used in your company? (Multiple answers are possible) Machine learning	0	0	0	1	0

Final Cluster Centers

	Cluster				
	1	2	3	4	5
Which data analysis techniques are being used in your company? (Multiple answers are possible) Data visualization	0	1	0	1	1
Which data analysis techniques are being used in your company? (Multiple answers are possible) Others:	0	0	0	0	0
Which data analysis techniques are being used in your company? (Multiple answers are possible) Don't know	0	0	0	0	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Don't know	0	0	0	0	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Sales	0	1	1	1	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Marketing	0	1	1	1	1
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice HRM	0	0	0	1	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Finance	0	1	0	1	1

Final Cluster Centers

	Cluster				
	1	2	3	4	5
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Manufacturing	0	0	0	1	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice R&D	0	0	0	1	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Supply chain management	0	0	0	1	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Quality management	0	0	0	1	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Risk management	0	1	0	1	0
In what business activities are data analytics procedures being used? (Multiple answers are possible) - Selected Choice Others:	0	0	0	0	0
What skills for data related positions are hardest to get by? - Selected Choice Sector knowledge/experience	0	0	0	0	0

Final Cluster Centers

	Cluster 1	2	3	4	5
What skills for data related positions are hardest to get by? - Selected Choice Creativity/innovation	0	0	0	0	0
What skills for data related positions are hardest to get by? - Selected Choice Teamwork/motivational skills	0	0	0	0	0
What skills for data related positions are hardest to get by? - Selected Choice Skills tha bridge IT and business requirements	0	0	0	1	0
What skills for data related positions are hardest to get by? - Selected Choice Analytica skills	0	0	0	0	0
What skills for data related positions are hardest to get by? - Selected Choice Technica skills	0	0	0	0	0
What skills for data related positions are hardest to get by? - Selected Choice Sales skills	0	0	0	0	0
What skills for data related positions are hardest to get by? - Selected Choice Others:	0	0	0	0	0
What skills for data related positions are hardest to get by? - Selected Choice Don't know	0	0	0	0	0
Which roles are represented in your analytics team? - Selected Choice Business analyst	0	1	0	1	1

	Cluster				
	1	2	3	4	5
Which roles are represented in your analytics team? - Selected Choice Data scientist	0	1	0	1	0
Which roles are represented in your analytics team? - Selected Choice IT expert	0	1	0	1	0
Which roles are represented in your analytics team? - Selected Choice System architect	0	0	0	1	0
Which roles are represented in your analytics team? - Selected Choice Others	0	0	0	1	0
Extremely_important	0	0	1	1	0
Very_important	0	1	1	1	1
Moderately_important	1	1	1	1	1
Slightly_important	1	1	1	1	1
A mature and enterprise wide data analytics strategy with a strong focus on continuous experimentation and improvement	0	0	0	1	0
A mature and enterprise wide data analytics strategy with a strong focus on continuous experimentation and improvement	0	0	0	1	0
An enterprise wide data analytics strategy exists, but not yet fully aligned with all business units	0	1	1	1	0
Ad hoc data analytics strategies in certain business units	1	1	1	1	1
Yes, within the next year	0	1	1	1	0
Yes but not within the next year	0	1	1	1	1
No	1	1	1	1	1
Strongly agree	0	0	0	1	0
agree	0	1	1	1	1

Final Cluster Centers

	Cluster				
	1	2	3	4	5
neither	0	1	1	1	1
disagree	0	1	1	1	1
Yes	0	0	0	0	0
Yes but not in the next year	0	0	0	0	0
No	0	1	0	1	1
Don't know	0	0	0	0	0
CEO	0	0	1	1	1
CMO	0	0	0	0	0
CFO	0	0	0	0	0
COO	0	0	0	0	0
Risk	0	0	0	0	0
technology	0	0	0	0	0
Vice	0	0	0	0	0
others	0	0	0	0	0
Strongly_Agree	0	0	0	1	0
Agree	0	0	0	1	1
Neither	0	1	1	1	1
Disagree	0	1	1	1	1
Strongly Agree	0	0	0	0	0
Agree	0	0	0	1	0
Neither	0	1	1	1	1
Disagree	0	1	1	1	1
Strongly Agree	0	0	0	0	0
Agree	0	1	1	1	1
Neither	0	1	1	1	1
Disagree	0	1	1	1	1
Individuals	0	0	0	0	1
Departments	0	0	0	0	1
Enterprise	0	1	0	0	1
Innovation	0	1	1	1	1
strongly agree	0	0	0	0	0
Agree agree	0	0	0	1	0
Neither agree	0	1	1	1	1
Wide range of issues	0	0	0	0	0
Some issues departments solve	0	0	0	1	1
Some issues enterprise solve	0	1	1	1	1

Final Cluster Centers

	Cluster				
	1	2	3	4	5
No	0	1	1	1	1
Clear view, progressive KPIs	0	0	0	0	0
Clear view	0	0	0	0	0
department view	0	0	0	1	0
specific datasets	0	1	0	1	1
No	0	1	0	1	1
Strongly Agree	0	0	0	0	0
Agree	0	1	0	1	0
Neither	0	1	0	1	1
disagree	0	1	1	1	1
Strongly Agree	0	0	0	0	0
Agree	0	0	0	1	0
Neither	0	1	0	1	0
Disagree	0	1	1	1	1
Always	0	0	0	1	0
Most of the time	0	1	0	1	1
Half of the time	0	1	0	1	1
Sometimes	0	1	1	1	1
Long term strategic	0	0	0	0	0
Tactical strategic	0	1	0	1	1
Operational	0	1	1	1	1
Strongly Agree	0	0	0	0	0
Agree	0	0	0	1	0
Neither	0	0	0	1	0
Disagree	0	1	1	1	1
Q4_Reasons	2.00	10.48	10.40	13.88	4.90
Q6_Impact	1.00	6.24	2.87	8.88	2.72
Q8_Challenges	1.06	2.95	3.47	3.00	2.76
Q9_Solutions	.82	3.52	2.67	2.50	2.55
Q24_Data	.24	2.48	1.40	6.63	2.34
Q25_Sources	.18	4.62	1.80	6.00	3.21
Q32_Techniques	.12	3.19	.87	5.50	2.79
Q35_Activities	.29	4.05	1.87	7.13	2.72
Q41_Team	.12	2.43	.33	4.25	1.76

Appendix 3: Blogs

Blog1: CDO: to stay or to go? Help me out!

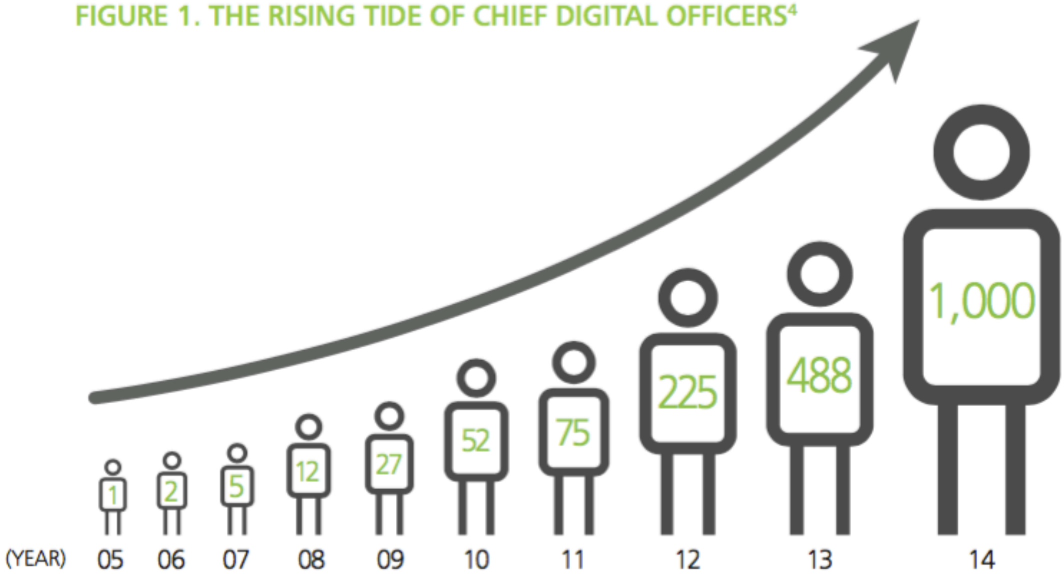
At the end of the 19th century, one of the greatest revolutions in history took place: electrification. Companies could now make use of a new, convenient form of energy: electricity. This innovation induced exceptional productivity gains, but the transition was never easy. Therefore many companies hired a CEO, a chief *electricity* officer to facilitate this transformation. However, just a few decades later the position became obsolete. Electricity quickly became well established and the introduction of a centralized electricity grid with synchronized voltages and frequencies solved all problems that kept a CEO awake at night. After just a few years, this great business innovator got sacked.

A new kid in town

Today, we witness a revolution of the same order of magnitude: digitalization. While digitalization offers awesome innovation opportunities for virtually every company, well-established businesses can easily lose their competitive advantage if they are reluctant to change. To take full advantage of digitalization without getting lost, business leaders are hiring experienced Sherpas to get them up the digital slopes: chief digital officers (CDO).

The CDO is a new kid in town that is quickly gaining popularity. The number of Chief digital officer positions worldwide has increased from 1 to 1000 in the last 9 years

FIGURE 1. THE RISING TIDE OF CHIEF DIGITAL OFFICERS[4]



Observers draw a parallel between the electric and the digital revolution. While the CEO was introduced to make sense of electrification, the Chief digital officer is introduced to make sense of digitalization. Question is, will the CDO suffer the same fate? Will the CDO become an invaluable and permanent C-suite member, or will the function become obsolete once digital has been incorporated into the company DNA?

Two questions come to mind:

1. Will a digital transition be as definitive as an electrical one? Meaning virtually no attention or innovation is needed after a digital integration.

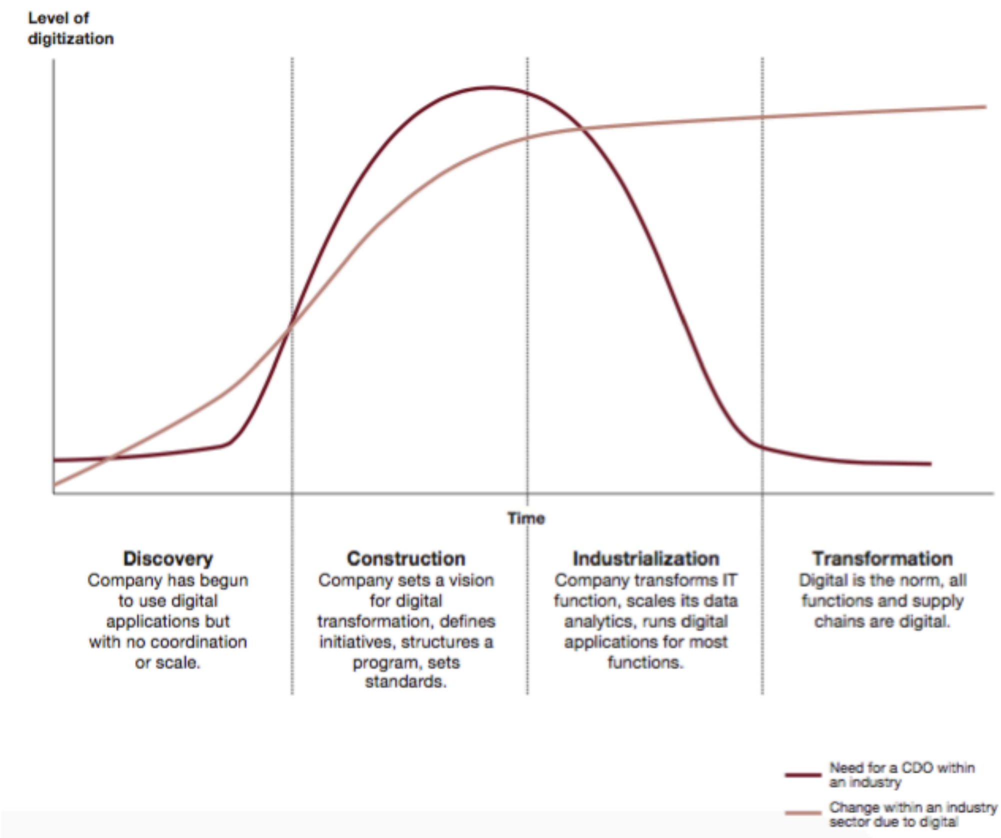
2. When constant attention and innovation is needed, who is going to make sure it happens?

The answer to the first question is straightforward. The electrification story is a nice analogy of today's revolution, but electricity is merely an energy input. Digitalization is a vast hodgepodge of new developments in multiple directions that will impact companies on a strategic level. Continuous improvement will be necessary as the technology will continue to change and the stakes are too high.

Will the real CXO please stand up?

The last question I want to address in this blog post is who is going to get his/her hands dirty. There are mainly three possible scenarios.

The need for a CDO during each phase of the digital revolution



1. The entire C-suite will integrate all aspects of digitalization in their responsibilities; there is no need for a CDO. The CEO will strategically navigate the company through the digital seas, the CMO will construct the appropriate big data systems to create a 360° view of the customer and so on. Digital will become a fundamental part of every corporate strategy and change the way we do business. Creating a separate role for it would be as silly as creating a separate Chief Smartphone Officer. It makes no sense to establish a new team for innovation, as it is a key responsibility for every member of the C-suite.

   By creating a CDO position the responsibility for continuous innovation could even shifted from the entire C-suite to the CDO. In this scenario, innovation is being taken care of in one department so everyone else can keep on doing business as usual. This is not an improvement; it is a denial of the true nature of digitalization

2. In this scenario the role of the CDO is temporary and transitional. The CDO is a strong motivation and a change manager with a clear vision. He transforms a traditional, cumbersome company into a dynamic and agile organization. The need for a CDO diminishes with digital maturity: once he has induced a culture change in the organization his position becomes unnecessary. After the disappearance of the CDO, the entire C-suite will become responsible for the challenges of digitalization like in scenario 1.

(PWC, 2015)

3. A new, permanent C-level position will be created. Question then quickly becomes: which one? Over the last decade several new digital related positions have been brought into existence i.e. Chief Digital Officer, Chief Data officer, Chief Analytics officer, Chief innovation officer... There still exist a lot of ambiguity about the exact definitions. Schmitz predicts the Chief Digital Officer will replace the Chief Information Officer over time, others predict otherwise. What is certain however is that companies will have to continuously reinvent themselves in order to survive. Time will tell which new C-level positions will be most suited to take on this challenge. For the others, there will be no other choice than to join the collection of outdated jobs, just like the man in the picture at the top of this article. He was an aurally gifted sky surveillant before Radar technology ruthlessly took his job. But hey, we can have a laugh at it now!

What scenario do you prefer and why? What are your thoughts? Leave your comments in the section below!

Blog 2: Data Olympics and the profitability of data investments

Athens. April 10th 1896. Five athletes prepare themselves for the 100m final of the first modern Olympic games. Their hearts are racing as they walk up to the starting line. Thomas Burke is one of them. Instead of standing up straight like his competitors he crouches as he concentrates and waits for the sound of the starting pistol. Thomas is going to win the race easily. His crouch start is far more efficient; the standing start goes into the history books.

Mexico. October 20th 1968. Dick Fosbury gets ready to jump over 2.24m. In the next moments he is going to break the Olympic record, astonish the audience and revolutionize high jumping with his new, weird looking technique. From then on, the "Fosbury Flop" becomes the standard for professional high jumpers.

Gold medals and quantum leaps

Continuous but small improvements are inherent to all sport disciplines. Every now and then an athlete sharpens the world record by a millisecond or a centimeter because of tremendous effort and a huge portion of luck. But from time to time, there is a quantum leap like the crouch start or the Fosbury Flop. A new technique comes along that completely outpaces the old one and shatters all previous records by a large margin.

Data analytics is such a quantum leap in business. By using advanced algorithms and smart strategies companies can easily outrun their competitors in the same way crouching Thomas Burke defeats his standing opponents any day of the week.

As high jumpers started practicing the "Flop" the day after they saw Dick Fosbury win the Olympic gold medal, businesses too are keen on rapidly adopting superior techniques from their competitors.

However, completely changing the way of doing business is very difficult. It requires a lot of money and persistence plus it is super risky. Therefore, it makes a lot of sense to analyze the effect of data investments on profitability in depth before taking serious action.

How much faster will I run if I use the crouch start? How much higher will I jump when I "Flop" over the bar? Does the benefit of the new technique really outweigh its costs? Is there evidence proving data analytics makes companies more profitable?

Show me the money!

Ironically, data analytics promises managers more data-driven decisions but the decision to invest in these data capacities are often quite intuitive.

Of course there are a lot of breath taking examples of the great opportunities of data analytics: self-driving cars, catching terrorists or even online dating. But how does this translate to the profitability of a company?

"Asking digital natives about the effect of analytics on the profitability of their companies is as meaningless as asking headmasters about the effect of teachers on the success of their schools. "

According to McKinsey, big data analytics can increase US GDP up to 1.7% ($325 billion) by 2020. This figure is a good start in the empirical assessment of the profitability of data analytics, but it will not make the investment decision of a CEO easier.

On a micro level, we, data fanboys, often evangelize the so-called *digital natives.* Companies like Google, Facebook or Uber that established enormous business empires within just a few years by using advanced algorithms and smart strategies. The stories of these companies are inspiring lighthouse examples of how any company should deal with digitalization and data analytics. They are great to give a sense of direction to any transformation plan. However, they will not pave the road for the complete digital metamorphosis of cumbersome, obsolete businesses. These examples do not show how to bridge the gap between an outdated, analogue way of doing business and its updated, digital equivalent. They do not answer the question of how to achieve data maturity, nor provide insight in the expected profitability of data related investments. Asking digital natives about the effect of analytics on the profitability of their companies is as meaningless as asking headmasters about the effect of teachers on the success of their schools.

For CEOs struggling with data strategies, it makes more sense to figure out how other big and old companies are tackling the problem. IBM, Caterpillar but especially GE are good examples of thoughtful but profound innovation in enormous companies.

Solid evidence

The profitability of data investments in those cases is usually very clear. A SAS study for example shows great savings in fuel costs for UPS after the implementation of an integrated analytics solution for route optimization (ORION). More generally, McKinsey estimates the increase in profits from big data related investments at 6% on average, but there are major profitability differences between data projects across different companies and industries. Bernard Marr estimates that half of these investments fail to reach the expectations.

Conclusion

1. Dick Fosbury showed us what to do, but it is up to us, our peers and our coaches to figure out how to get there. Google is a great example, but CEOs better study the transformation of old companies when they devise their own digital strategy.
2. Data investments have enormous potential, but the difference in profitability is enormous as well. Some companies manage to reap the full benefits of data analytics while others fail.

What are your thoughts on the profitability of data investments? Leave your thoughts in the comment section below!

Did you know cavemen were already dealing with big data issues?

Blog 3: Struggling with Big Data in the Stone Age

Nowadays everyone is talking about data & analytics. Executives are figuring out how to turn it into value, researchers are writing huge amounts of papers about it and students are enrolling in massive numbers in data science programs.

This would lead to believe that data & analytics are a recent phenomenon. And indeed, there are convincing reasons that show analytics is more important and relevant than ever. However, the underlying idea of analytics- the urge to understand the nature of our reality and to act accordingly- is as old as humanity itself.

What's even more surprising is that people seem to have solved an enormous big data problem several millennia ago. To see how, let us summarize a chapter from Yuval Noah Harari"s great book "A Brief History of Humankind'".

Memory Overload

For millennia our ancestors have been living as foragers hunting for animals and looking for sweet berries. At this point in history data was already crucial. If you always forgot that snakes are dangerous and red mushrooms are poisonous your chances of survival were slim. Luckily, the amount of important data was small so people could store everything they needed to know in their brains. However, when humans started living in bigger settlements and communities the limits of our brainpower became painfully obvious.

Living in these larger societies required an entirely different mental skillset. Many problems that arose were data related and could not be solved with the conventional technique of the day: memorization. For example, local rulers had to know how much grain they had in their inventories and whether it was enough to survive the winter. There were taxes to be collected from thousands of inhabitants and thus payment data that had to be stored. There were legal systems that had to handle property data and ever growing families that had to keep track of their family tree. There was just so much more data around than before that just learning everything by heart, like we had been doing for millennia, was never going to work out.

What we needed was a quantum leap. Thus our ancestors conceived a genius system roughly 5000 years ago that encoded their thoughts into abstract symbols, which in turn could be interpreted by others who were familiar with the rules of this system. This made storing and processing information outside early humans' brains possible for the first time ever. Writing was born.

Figure 1: early writing on clay tablets, Mesopotamia 3000 BC.

Big data 2.0

It is astonishing how closely these data issues in early human civilizations are related to the current struggle of companies and governments to store and process evermore data.

Today, the total amount of data is increasing at the astonishing rate of 40% per year as depicted in figure 2.

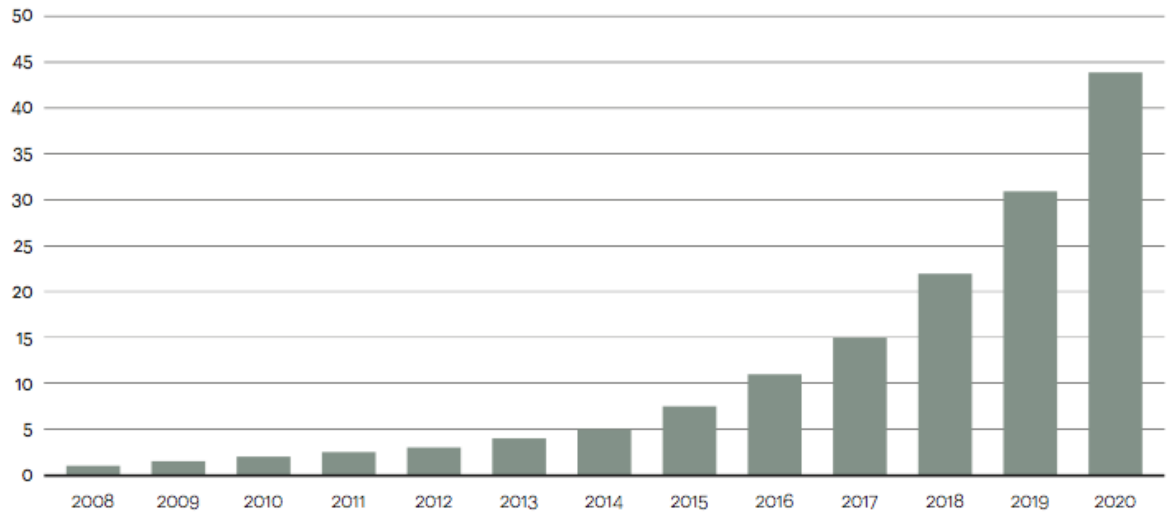


Figure 2: total amount of data in ZB.

To get a sense of how huge this is, let's consider the following statistics:

In 2015, 205 billion emails were sent every day!

In 2016, 432,000 hours of video were uploaded on YouTube, every day!

By 2020, more than 6GB of data will be created every hour for every human being on the planet!

This incredible amount of data can be of tremendous value for a company if it knows how to extract information from it. However, researchers estimate only 0.5% of all data is ever analyzed. Clearly, there is a vast ocean of undiscovered data waiting to be analyzed and turned into value.

Just like 5000 years ago, when simple memorization was not longer able to store and analyze data efficiently, companies today are witnessing their traditional analytics systems are not longer effective to tame the big data tsunami.

Just like memorization was replaced by writing, traditional databases are replaced by distributed systems. The quantum leap necessary to make data analytics on a vast scale possible.

Conclusion

Big data turns out to be a 5000-year-old problem! The big data architect could turn out to be the second oldest profession after all…

Universiteit
Antwerpen

Declaration on word of honor

I hereby declare that I know what plagiarism entails, namely to use another's work and to present it as my own without attributing the sources in the correct way.

I acknowledge that copying someone else's assignment or essay, or part of it, is wrong and declare that this is my own work.

I have used the American Psychological Association (APA) as the convention for citation and referencing. Each significant contribution to, and quotation in, this thesis from the work, or works of other people has been attributed and has been cited and referenced.

This Master's thesis is my own original work and has not yet been handed in at any other university, nor had it been published.

I am aware of the consequences of fraud as stated in the exam regulation of the University of Antwerp.

Date

Name

Signature