Success factors for the implementation of Big Data as a tool to reduce bounded rationality in organisational decision-making

Sebastian Geisler

MSc in Technology Management

07.02.2016

Contents

Abstract

In the information age, Big Data is on the rise for organisations of all sizes to obtain a competitive advantage over market participants. It is important to executives and decision-makers to understand the current dissemination of Big Data and what organisational factors are critical to successful implementation. One of the numerous advantages is the reduction of bounded rationality in the internal decision-making processes.

The focus of this research, therefore, is in the area of technology management seeking to identify a workable practical definition of Big Data, attempting to identify the success factors that are beneficial in the introduction of Big Data to reduce bounded rationality and producing suggestions for organisations that want to implement Big Data. The research approach adopted in this thesis includes a review of relevant literature on decision-making and data-driven decision management, case studies, and the collection and analysis of empirical data. The latter is based on an open survey of 294 participants in the UK and Germany, using online questionnaires over a timeframe of five months. The research outcomes were established by revealing relevant correlations between success factors and survey participant responses.

The findings from this research show, that Big Data is more well-established for some organisations than was initially thought and there exist several cultural and factors, which can contribute to the success of a Big Data implementation. Based on the findings recommendations to organisations are made on both the success factors and the possible measurement of bounded rationality.

List of Figures and Tables

Glossary

Decision-making

The intellectual process of analysing various inputs and existing knowledge to reach a decision. This can be done in isolation or in groups.

Bounded rationality

Decisions are almost never purely rational, as knowledge might be limited, past experience might be biased or not representative, and the time and resources available to the decision-making process is limited.

Big Data

The collection and analysis process of data, defined by the four data attributes below. See also the definition section.

Data size/volume

Quantity of the data. Usually referring to sizes that go beyond the household storage capacity. Currently, this is measured in storage sizes greater than several terabytes.

Data complexity

The structure of the data might be multidimensional and it may have originated from multiple sources.

Data velocity / change rate

The rate at which the data is modified in any way. This could be by altering existing data sets, adding new data, or removing data.

Data variety

Data might not only consist of text but also of images, video, and sound.

Organisational culture

Describes how organisations and their employees act, what values are favoured, and how power is distributed and executed.

Likert scale

A survey question type that consists of multiple possible answers that place the response on a scale for further analysis. Usually this type is used to identify agreement or disagreement with a statement.

Acknowledgements

I would like to use this opportunity to express my gratitude to the institution of the Open University and especially to the tutors of which I had the privilege to learn from in recent years.
Likewise, I wish to thank my thesis supervisor Steve P. Abbott for his patience and guidance during my research.

Further, I would like to thank Michaela for her invaluable support and encouragement during my studies.

1. Introduction

1.1 Background to the problem

Prominent companies such as Google, Amazon, IBM, Facebook and eBay are known to utilise Big Data analysis for their decision-making processes (Schmidt and Rosenberg, 2014) because the outcome of Big Data is believed to help the organisation to achieve its goals more effectively and efficiently. For example Big Data is believed to help by reducing procurement costs, identifying and supplying previously uncharted markets, by analysing customer behaviour to improve existing products, by predicting future customer behaviour, or generating innovation.

However, it is not the technology that makes the decision, but the managers. They make decisions that define the organisation - both strategic and operational - in product development, market focus, project management, human resources, and production lines. Simon (1972, 1979) identified that the decision-making process is prone to “bounded rationality” because the decision-makers are limited in their processing of information due to various reasons, such as human cognitive limitation, time constraints, heuristics, personal preference, uncertainty, and bias. In addition, conformity pressure phenomena in groups (aka “groupthink”; (Janis, 1971)) has been extensively studied over the past decades. This all eventually has an effect on the outcome of the decision-making process, evidently leading to a sub-optimal decision.

Big Data analysis is said to reduce the above-named factors of bounded rationality and groupthink by including an additional scientific element that can be used as an input (Masha, 2014, Wang, 2012), thus improving the decision-making outcome and hence improve the organisations efficiency and effectiveness. Furthermore, Big Data can also act as a positive constraint to the existing process e.g. by reducing personal bias.

Although models exist for the data mining process regarding Big Data in general, e.g. CRISP-DM (Provost and Fawcett, 2013), there are no signs of quantitative studies about the actual implementation into the organisational decision-making process itself and which factors contribute (and in what amount) to the success of such projects. The lack of literature about Big Data implementation in regard to the organisational context that surrounds decision-making (for example organisational-culture influences and power distribution) leaves companies and its executives alone with the question of what factors are either prone to be problematic or may even be ultimate prerequisites for success. These should be identified and managed closely to reduce risk and increase reward in Big Data implementation. To close this gap, this research focuses on the impact of Big Data implementation to decision-making in the organisational context, and does not include the scientific models behind Big Data analysis. Given the complexity and variety of organisations and their processes, a mixed quantitative and qualitative approach was required for this research.

The target audience is the executive management of organisations that want to implement Big Data. Although the research tries to identify and analyse success factors of Big Data implementation, the findings are probably valid for a wider range of technologies surrounding Big Data that are to be implemented into business processes and the decision-making process in particular.

1.2 Justification for the research

Organisations are following a specific goal. Although most of the time the goal is to generate monetary profit, the goal itself does not matter to this research, as it is focused on how efficient and effectively the goal is reached.

Big Data is said to help organisations to achieve their goals, but it does not just appear inside an organisation and deliver the promised improvements by itself: it can be a tool, which reduces bounded rationality by utilising statistical and scientific methods. Organisations should not neglect the negative effects of bounded rationality on decision-making. Although bounded rationality is an inevitable necessity due to the constraints of available resource and human intelligence, Big Data can improve the decision-making process to mitigate the disadvantages, for example by:

1. improving information reliability – instead of following personal preference; or
2. reducing uncertainty by discovering new information that was previously unavailable – instead of relying on heuristics due to resource and time constraints,

Research shows that Big Data itself can deliver the promises made, e.g. in the area of marketing strategies (Hill et al., 2006, Martens and Provost, 2011), or as indicated by the increasing Big Data workforce requirements (Tambe, 2012) and higher market value of firms that are investing in data-driven decision-making (Brynjolfsson et al., 2011). Further research however might reveal that the actual implementation (or realisation of the benefits) in organisations is prone to failure because organisations probably face diverse challenges in the implementation of Big Data. Noticeable amounts of papers and reports exist about the technical challenges and best practices for implementing Big Data (Lazer et al., 2014, Rabl et al., 2012), but the literature lacks information about the implementation process regarding organisational context and decision-making benefits, particularly in the area of bounded rationality.

This is especially important because organisations, as sociotechnical systems, have countless variations of processes and embedded cultures. Therefore, the research aim was to find out how Big Data analysis can be successfully implemented into the organisations as a tool to reduce bounded rationality in organisational decision-making by identifying the most important factors that will benefit the successful introduction of Big Data.

The scientific measurement of the reduction is outside of the scope of this research, thus, it is limited to whether, and to what extent, bounded rationality was reduced by identifying the maturity level of the decision-making processes, e.g. process definition, integration, skills, tools, and metrics.(Elbanna, 2006, Kaner and Karni, 2004).

1.3 Definitions

Because the definition of Big Data is still the subject of discussion, and in order to make this research beneficial for a broad audience, in this thesis Big Data is defined as follows:

1. the data volume is too big, or
2. the data variety is too high, or
3. the data velocity is too high, or
4. the data is too complex

to be manually analysed periodically or in real-time because, for example, the cost is too great or the analysis would take too long.

1.4 Scope of the research

The research project was conducted to establish a practical definition of Big Data and to identify and analyse possible success factors for implementing Big Data in organisations with focus on both organisational culture and the decision-making subject, especially bounded rationality. Both primary and secondary research were used - including a survey of 294 participants of German- and English-speaking online social business networks whose employers already had Big Data in place.

Outside of the scope of this research are the technical aspects of Big Data, such as the level of control over the data and analytics processes, and the statistical groundwork that is utilised by the Big Data analysis processes.

1.5 Outline of the dissertation

Chapter 1 describes the background to the problem, establishing the definition of Big Data in this context and providing the justification for the research.

Chapter 2 lays out the research definition by providing the practical issue of success-factor identification and measurement and also provides a categorisation of these factors to form a structure for the research. The existing relevant knowledge is reviewed in the context of the practical issue, resulting in the identification of a knowledge gap that defines the research aim.

Chapter 3 explains the selected research methodology, procedures, and justification. Additionally, possible ethical issues in the research are considered and the solution to mitigate these is explained.

Chapter 4 presents a summary of the data collected, explains the analysis, including statistical methods, and gives an interpretation in regard to the research objectives and research aim.

Chapter 5 summarises the research with a conclusion about the research objectives and research aim, and recommends possible further research in the area. The implications of the research are outlined and a reflection on the researchers experience during the project is given.

2. Research definition

2.1 The practical problem/issue

The implementation of technology into organisations and their processes can be a difficult task. Specifically, with Big Data, there might also be high entry and maintenance costs associated – either with the sourcing of the tasks and technology or the allocating of internal resources. Furthermore, the risk of not being able to exploit the results of the implementation might also be significant. Therefore, for an organisation, it is essential to identify the success factors that allow the organisation to take an appropriate course of action to reduce the risk of financial loss and to increase the likelihood of potential positive outcomes.

Some of the potential challenges on the organisational and technological dimensions to the implementation of Big Data have been identified by Russom (2011), although these are to be taken cautiously, as the independence of the results might be questionable. A more reliable reference is Biehl (2007) who identified several success factors for the implementation of Global Information Systems; which are technologically and organisationally similar to some extent.

The general definition of factors that help to achieve success is that they are elements or capabilities that, if delivered, will enable the organisation to achieve the successful accomplishment of projects, strategies, goals, and missions. Freund (1988) coined the term "critical success factors" for these, and defined that they are measurable and controllable. Although this definition might be applicable to some degree to the factors that support the implementation of Big Data, this research requires us to go beyond Freund's definition. The reason for this is that we are looking at organisations' decision-making processes themselves, which are strongly linked to the organisational culture, power distribution and, of course, bounded rationality. These factors are not easily measured from the outside and are prone to the interpretative perception of the researcher and the research subjects. For this reason, I used the term "success factor" but not strictly limited to Freund's definition, rather including all identified factors that could contribute to the successful implementation of Big Data. For example, according to Biehl (2007) top-management support was a success factor in all conducted case studies. Furthermore, several factors that are closely tied to organisational culture have been named.

The possible success factors that are distilled from interviews are organised into four dimensions (Yeoh and Koronios, 2010).This is a model that is borrowed from the programme management field and is sometimes referred to as "POTI":

- Processes, e.g. process knowledge, transparency and maturity;

- Organisation, e.g. strategic management, culture, goals, and power distribution;
- Technology, e.g. IT factors, skills, and tools; and
- Information, e.g. data-related factors.

Along at least two of these four dimensions, the existing literature shows that little is known about the actual impact of the distinct factors for successful implementation of Big Data that can rely on statistical relevancy, especially in the processes and organisations dimension (Ang and Teo, 2000, Biehl, 2007). A better understanding of the various success factors for the implementation is expected to provide guidance to organisations so that they are able to focus on the success factors that have an impact on the implementation and thus achieve the expected returns on the investment.

For example, it might be of interest to organisations if companies with strong hierarchical structures are more successful in the implementation than others, or not. Armed with this knowledge, it might be possible either to find the relevance of more concrete factors by digging deeper into the existing knowledge of organisational theory, or, at least, to direct further studies into the significant areas. Furthermore, organisations might be able to incorporate the findings into their risk assessment and increase the chance of success.

2.2 Existing relevant knowledge

Along the aforementioned four POTI dimensions, the following existing knowledge was identified:

Organisation

The organisational culture does influence and is being influenced by decision-making and expected action and behaviour of employees and management (Mintzberg et al., 1998, Weick, 1985). Miller et al. (1999), as well as Pettigrew (1973) further explicated that decision-making is a political process, which makes it possible that the scientific output of Big Data analysis and its unpolitical reasoning might be conflicting with the particular interests of existing power distributions within an organisation. Therefore, it is reasonable to conduct the research with awareness of the context of the organisational culture and power dimensions to identify possible cultural success factors of the organisation. For example, powerful agents in the organisation might pursue their own agenda, which might contradict the direction in which Big Data might drive the decision-making process. Therefore, the survey also included questions about the power distribution and internal conflicts during Big Data implementation and especially about power conflicts in the decision-making processes after the implementation.

The theoretical decision-making process prescribes that decision-making follows problem analysis. However, Laroche (1995) argues that often, decisions are mere rationalisations of already-taken actions. In such cases, the undoubtable result of a Big Data analysis will confront the decision maker with a potential incoherency between decision and analysis, introducing a cognitive dissonance. In addition to the new technology - that might result in a shift of power from gut thinking to scientific analysis – decision-makers could be confronted during the implementation with quasi-resolution of conflict (satisficing), uncertainty avoidance, and problemistic search (Cyert and March, 1963) – each of which might negatively affect the success of the implementation. Although it may be interesting to research how this has affects the organisational decision-making process regarding process improvement and organisational learning, cognitive-dissonance analysis might be more suitable for a psychological and behavioural study and is not therefore given priority in this research.

Also, on the subject of power distribution, deliberate manipulation of the Big Data processes or outputs could be considered to be an issue in the decision-making process. As the analyses are unlikely to be checked manually again by the decision-makers due to the specific attributes of Big Data - and thus acting like a black box - a blind reliance on Big Data might lead to unintentional decisions. This probability is not addressed in existing publications; therefore, I cannot refer to existing knowledge. In this regard, the general scepticism of Big Data results is also addressed in this research.

The measurement of the organisational-culture dimensions is difficult, as the observation and interpretation of these aspects will differ between the observers (Robbins, 1989). However, it might be feasible to measure the perceived culture with common established and shared characteristics that can be isolated in the research. Thus, the following broad dimensions seem to be worthwhile to follow from which the maturity level might be interpreted:

1. direction, “the degree to which the organisation creates clear objectives and performance expectations”;
2. integration, “the degree to which units within the organisation are encouraged to operate in a coordinated manner”; and
3. management support.

I have dismissed the other characteristics that Robbins identified because the perceptions are likely to be highly varying across the individual employees of a single organisation and are therefore better researched in case studies or surveys that focus on specific organisations to obtain more reliable results. The classification itself, therefore, is best done with Boisots “I-Space” (1998), in which the culture characteristics can be located (Figure 1).

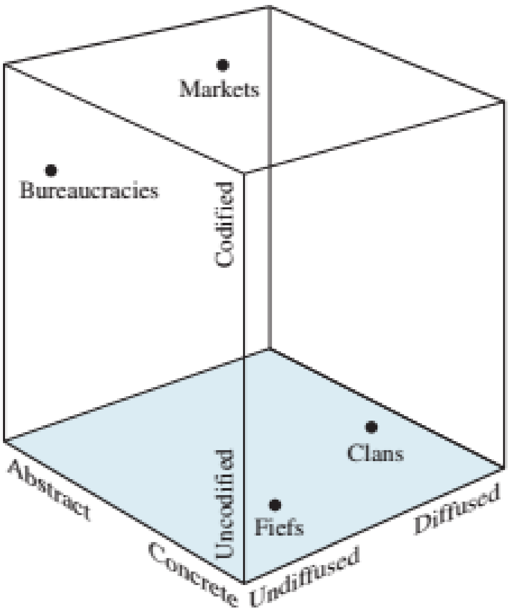

Figure 1: Boisots I-Space, 1998

Processes

Mintzberg (1976) suggested four steps in the decision-making process. For each of these steps, the outcome of the overall decision-making process could be improved by the application of Big Data analysis:

1. Problem awareness
   Continuous analysis of data can provide an early warning system for organisations. For example, exchange platform providers would be able to identify upcoming technical issues in real time based on the trending latency of trading messages, global market activity, and other factors that go beyond the commonly known IT monitoring processes (Geisler 2012). In this specific case, the decision-making process to take action against the issue would be triggered before the incident actually happened, together with information about the affected systems, exchanges, and customers. Because of the amount of variables, dependencies, and correlations, this could not be handled manually in real-time. Without such an early warning system, the issue would occur, first level support would then be informed by a customer that an issue exists, and only after the subsequent analysis, could a decision-making process be started – all while the service impact is ongoing.

2. Problem diagnosis

Correlation between various data types and sources can provide insight into the root cause of a problem when a diagnosis is made, such as in production line and product quality issues. For example, Chien and Chuang (2014) provide an empirically tested framework, where they are able to identify the factors that provide a significant impact to the yield of that specific production line, from an initial input of about 4000 available factors. Arriving at such conclusions in a resourceful way, as Chien and Chuang did, seems almost impossible without Big Data analysis and especially the corresponding data-dimension reduction, considering the vast quantity of factors.

3. Finding solutions and 4. selecting a solution
   In some occurrences, it might be feasible to simulate multiple possible solutions, for example by A/B testing (statistical validation) or historical or seasonal comparison to verify assumptions to some degree of certainty (Kohavi et al., 2013). Although most of this may also be achieved by manual analytics, Big Data stands out because of the efficient reuse of analysis models to analyse several solutions in sequence over a period of time by using updated input data.

Therefore, competitors who have not yet materialised any benefits from Big Data analysis will face increasing pressure to do so to close the competitive gap with organisations that already exploit Big Data (Masha, 2014). The gap might be reduced or even closed by copying the techniques that the successful first-movers are using. Even in other contexts outside of the multinational business-sphere, examples exist to prove this, as was seen in the US presidential election campaigns of 2012 and in the near future for local owner managed companies that could analyse customer behaviour to provide better market-orientated services and products.
This can be accomplished by either analysing their own collected data or by buying analytics and data from specialised service providers. Well-known similar external Big Data services have, for years provided references to financial credit ratings, for example, Schufa in Germany or Equifax in the UK and US. However, these examples lack the technology and possibly even the information dimension for research, as the analysis process as well as the data sourcing itself is a black box to the organisation.

Mandinach, et al. (2006a, 2006b) created a conceptual framework for data-driven decision-making , which is based upon Ackoffs data to wisdom transformation process (1989) and Checkland’s systems thinking and soft systems methodologies (1998, 1999a, 1999b, 1989). The framework itself is not included in this research, but their findings that data-driven decision-making skills are relevant as a success factor is included.

The research is applicable for organisations that want to exploit the benefits of analysing internally or externally available data to provide a competitive advantage, regardless of

1. whether internal processes are to be optimised,
2. the establishment of early warning systems, or
3. for gaining a better understanding of customer behaviour and prediction through social media, etc.

The research aims to show whether or not Big Data can be better integrated into formal decision-making processes. The initial assumption was that the problem diagnosis and solution finding in formal processes is usually done in a planned and conscious manner, which might work well together with a Big Data analysis setup, as it can require thoughtful modelling of the issue at hand. In addition, the contract between the sourcing party (the organisation that wants to utilise Big Data) and the service provider (an external entity that provides analytics services or data collections) might be a significant success factor after the implementation. The level of control over the data and analysis models, for example, could influence the decision-making processes that utilise the Big Data analysis. Although relevant, the latter question is outside of the scope of this project due to the tight timeframe of the project.

Technology & information

Miller and Mork (2013) suggest that Big Data is a collection of functions that can be visualised as a value chain, which helped to break the Big Data analysis process down, which further allowed to group and categorise the findings of success factors in the research (Figure **2**).

Figure 2: Data value chain, adapted from Miller and Mork (2013)

At each of these functions, pitfalls exist for organisations that want to implement Big Data. For example:

- **Data discovery**

Typical questions in this step are “what data can be collected?”, “where can it be collected from?” and “what is the quality of the collected data? (e.g. complete or random samples, non-ambiguous)”. The discovery requirements depend on whether a specific decision-making question has to be answered, e.g. on how to reduce waste in a specific production process, or if correlations are being searched within the available data to further identify possible causations which might help in improving the organisations performance by identifying and eventually modifying as-yet-unknown factors.

- **Data integration**

An example here is the format in which the data is to be saved and whether the data is to be stored inside the organisation or with a subcontractor. The data-storage strategy will eventually determine the type and pace of analysis that can be designed and processed. In addition, because of the cost of building, maintaining, and operating an appropriate infrastructure, it is likely to be a matter of discussion whether it might be financially beneficial to outsource to specialised contractors. This does, however, introduce other concerns to be considered such as vendor lock-in, service provider lock-in, personal data privacy, data security, trade secrets, and technology-path dependency.

- **Data exploitation**

One important issue is that the combination of data sources and the uncovering of correlations might introduce spurious correlation, false positives, and false negatives etc. In addition, integration into existing technology-supported information systems could likely be desired, e.g. decision support systems, which might come with technical or procedural issues. In particular, the implementation of the exploitation process into the various stages of the decision-making process is of interest in this research. The choice of which data-mining technique the organisation will use might depend upon the specific data that is available and the outcome that is pursued – both which bay be unique to the organisation. Therefore, the techniques themselves are initially given low priority in this research.

Maturity

The Decision Making Capability Maturity Model (DM-CMM) describes a model that allows evaluation of the organisation's decision-making processes on several levels (Kaner and Karni, 2004). Also, the evaluation of the decision support system type (e.g. from printed-paper towards automated decision systems) can be used as a frame for the survey. Although possibly relevant to the research findings, the interpretation of the maturity level of the processes might be highly subjective. Therefore, it is essential to find and use generic objective measurements for each level, which can then be asked in the survey.

Although Kaner builds a framework in which Big Data can either act as a part of the described support system, or as the support system itself, there is no data available that suggests benefits to a specific maturity level for successful exploitation of Big Data. However, it may be assumed that the maturity level in organisations could have a relevance to the implementation success.

When viewed together with organisational learning, the maturity level is already being considered to be significant due to the instalment of regular feedback loops. As decision-making is already proven to benefit from such learning cycles (Simon, 1991), this is also likely to be true for Big Data supported decision-making. Referring back to Miller and Mork's data value chain, the learning cycle would not only improve the Big Data processes and thus increase the value and validity of the input to the decision-making process, but also would improve the integration into the decision-making process itself.

There is a gap in current literature: there are no attempts assemble these previously well-researched but isolated pieces together to form the knowledge that would allow us to answer the question of which success factors are relevant, to what extent, and to which organisations. This is why the aforementioned theoretical models are only of limited use to organisations that want to introduce and utilise Big Data for their goals.

2.3 Aim, objectives, methods, tasks, and deliverables

The aim for this research is to identify possible success factors for implementing Big Data to reduce bounded rationality in organisational decision-making.

The objectives are as follows:

1. give an overview of the current practical definition, usage and distribution of Big Data;
2. identify and analyse the success factors (or success-factor groups, if no distinct factors can be identified) for the implementation of Big Data across the surveyed organisations that provide a positive or negative effect on the reduction of bounded rationality achieved by the implementation;
3. identify statistically relevant correlations and test these against causality through existing case studies; and,
4. produce guidance for decision-makers about the success factors for implementing Big Data to reduce bounded rationality.

3. Methodology

3.1 Methods and techniques selected

Qualitative research stage: interviews & secondary research

The interview phase was prepared with consideration to the existing knowledge, and was required to obtain an insight into the actual challenges of Big Data implementation in organisations, going beyond the theoretical frameworks and findings of the existing literature. The interviews and research were combined to form the survey. For this, 16 participants from 13 organisations of between 10 and 1500 employees were interviewed. The hierarchy of these participants is 5 C-level executives (meaning the highest organisational hierarchy), 8 middle line managers and 3 Big Data specialists/data analysts. These results were only part of the foundation of the survey and were used in addition to the literature review to achieve the overall research goal of identifying, interpreting, and grouping the possible implantation challenges and the factors and features of organisational culture that affect implementation. Because of this, the interviews are not required to be statistically significant on their own.

Quantitative research stage: survey – cross-sectional - statistical

A survey was used as the major primary research source to produce a statistically sound result. The survey was published in an electronic form (online survey with yes/no and Likert-scale questions), to increase the response and flow rate. The audience surveyed consisted in its majority of mid- to senior level managers and C-Level executives that were contacted through business social networking channels on XING and LinkedIn directly (systematic sampling) and especially in target-specific message groups/forums. Examples of these forums/groups are Energy & Management, CIO Forum, IT Management, IT Finance, CouldComputing-Insider, B2B Cluster, IT-Connection, Big Data Expert, CTO Forum, CIO Exchange, Fintech, and Predictive Analysts.

This approach is likely to be biased towards participants that actually use business social media channels. However, it is assumed that the bias via such message forums is not much different from a random or simple random sampling approach (Biggam, 2015) if compared to selection on multiple streets, for a specific time and day. It might be argued that this approach might be even less biased in this regard with surveys accessible 24 hours a day for continuous weeks – in contrast to several hours per day that are highly geographically bound.

The questions in the survey ask about the aims, goals, and objectives that are laid out in the previous section. In the design phase, several isolated prototype surveys with the previously-interviewed target group were conducted to find an unbiased, appropriate, and meaningful way to create survey questions that are calibrated to the actual situation of the participants. For example, several tests with the 5- and 7-type Likert scale (7-type was dismissed as many of the participants stated later that the granularity was too fine) and the choice of words, especially for non-native speakers (Ervin and Bower, 1952) were carried out. For each iterative cycle, the answers to the questions were compared against the interview results and adjustments to the survey were made.

As surveys depend upon third parties, I produced and sent out the survey as fast as possible to reduce the risk of late submission and to allow sufficient time to analyse the results.

As compensation, a 100€ Amazon voucher was offered which was randomly given to the winning participant. As the survey took about 15 to 20 minutes to complete, it is believed that larger monetary compensation, compared to what other surveys are offering, was appropriate. To accord with the "Ethics Principles for Research involving Human Participants" (The Open University, 2006) which states that "no inducement to participate should be offered prior to seeking consent, either in the form of payments or of gifts", the compensation was first mentioned at the end of the survey, assuming that the participation implied consent.

3.2 Justification

It was assumed that in the first stage (which consisted of unstructured interviews) the areas which the research had to examine would be processes, organisation, technology, and information.

These areas also included fragments about legislation, organisational strategy, organisational culture, financial aspects, technological aspects, organisational power and politics, organisational learning and change, improvement and innovation, as well as project management. Further, the decision-making process itself, especially in the matter of formality/maturity was assumed to have a significant impact on the implementation success of Big Data.

This required a qualitative approach to identify the possible success factors, and a quantitative follow-up to measure these enabling factors for the desired positive impact of the implementation of Big Data, by finding statistically significant results for the various factors among the questioned organisations. Existing literature about success factors is mostly limited to general views of organisational goals and is not specific to the details regarding to the special technologies or their requirements in terms of processes, technology management, organisational culture, and so on. The weighting of the success factors among the various organisation types and Big Data utilisation modes suggests a variety of special requirements for which a one-fits-all approach seems questionable.

The discovery and data-collection processes by means of interview and survey are likely to be influenced by the participants, as well as by the researcher. Therefore, I am aware that the underlying paradigm for that part of the research tends to be constructivistic, as organisations, processes and of course, the interviews/surveys themselves are influenced by social interaction. In particular, where people are questioned about experiences and factors that contributed to the implementation of Big Data, these are reconstructed from memory and told from the personal view of the participant. Because of this, efforts were made to include multiple participants from each organisation to mitigate these negative effects. However, this was only possible for a small subset of organisations. Furthermore, the interpretation of the interview answers by the researcher also falls into the constructivistic paradigm.

The quantitative research stage, which was built upon the previously-described discovery stage, has been conducted without the previously-described issues and cane hence be attributed as positivist paradigm.

The following research methods and techniques were dismissed:

1. Experiments
   Because of the high variability of organisations attributes like the variation in the decision-making process, the on-the-day performance of the decision-makers themselves, etc., the repeatability of an experiment would be low and would probably provide limited insights at best and be of limited use for the wider audience of organisations.
2. Observation
   Although I would think that observation on multiple case studies would generate a rich set of data to analyse, resource constraints did not allow such an approach. Furthermore, organisations tend to be secretive to outsiders about details of their decision-making, and thus are unlikely to allow outsiders to observe management meetings, which would reduce the number of possible study participants.
3. Case studies
   It seems that there does not exist sufficient case studies for analysis in order to answer the research questions with statistical relevance. Conducting primary case studies is not possible due to time and resource constraints. However, the existing case studies in the area of data-driven decision-making and Big Data implementation may be used as a basis for verification of the analysis results.

3.3 Research procedures

At the beginning of the survey, the participants were grouped into a hierarchy: management or non-management; the level of participation in decision-making processes (which also acts as a control question); and the area of work (IT focus or not). In addition, the organisation itself was categorised based on sector, size, and internationalisation.

The survey then proceeded by asking multiple-choice questions, where the possible checkpoints have been synthesised from the qualitative stage by evaluative Likert-style questions.
The Likert-style questions required careful thought about the question itself because they are required to be one-dimensional. The order in which these are presented, as well as their evaluation, and confirmation (e.g. by using the Spearman, Kendall, Wilcoxon-Mann, T-test, confidence interval, or split-half reliability test) also required careful work. Furthermore, Schumann (2012) suggests mixing standard and inverted questions to reduce approval bias, as well as introducing control questions (Table 1).

Main question types	**Control questions**	**Question formulation**	**Answer formulation**
recall	predefined outcome answers	one-dimensional	exhaustive
reason	use of combinations	unambiguous	similar width categories
filter	duplicates	simple	"don't know" never as middle answer
		no double negations	symmetric/asymmetric scale
		precise	multiple choice
		not suggestive	reduce primacy- and recency effects
			halo effects

Table 1: Considerations for survey creation

These requirements were incorporated into the survey. Furthermore, additional free-text fields were provided below the questions because, in the previously conducted interviews for some items none of the respondents gave similar answers to the others, which might indicate that the variation in response for some of the questions could become higher than anticipated. To analyse these, it was planned to produce a category schema based on the answers to synthesise the entries gathered in the free-text fields (Schumann, 2012) to provide further detail to the analysis.

By analysing the results of the survey, I produced the rest of the research using statistical analysis and verification.

3.4 Ethical considerations

Possible issues identified were:

a) Confidentiality

   To counter this issue, the names of survey participants or companies have not been collected.

b) Openness and integrity

   Participants have been offered to receive the research results.

c) Informed consent

   The nature and purpose of the research have been explained. The time required to participate in the survey and interviews was stated beforehand. The identified ethical considerations above were explained to the participants.

The Code of Practice for Research at The Open University (2013) has been followed. Data Protection Laws of Germany (BDSG) and the UK (DPA) have been adhered to. No difficulties were encountered.

4. Analysis and interpretation

4.1 Summary of data collected

The research was conducted in two stages. The first stage took place in Germany during May 2015 and consisted of unstructured interviews of 16 participants from 13 organisations with between 10 and 1500 employees. The hierarchy of these participants consisted of 5 C-level executives, 8 middle line managers and three Big Data specialists/data analysts.

The second stage was held online through an online-survey tool and was open for participation from July to December 2015. In total, 294 participants completed the survey. As the tool did not register incomplete attempts, the bounce rate (participants who did not complete the survey) remains unknown. There was no possible differentiation of the 40 personal invites to the survey and the social business-network postings. Therefore, the response rate could not be calculated.

The margin of error is 6.1% with a confidence level of 95% for the population group of Germany and the UK with their combined population group of 7.2 million organisations (Companies House, 2015, Statistisches Bundesamt, 2013). The data collection process is believed to be unbiased.

4.2 Data analysis

Conditioning / filtering

All questions were mandatory, thus, there are no empty responses. Further, the survey had to be completed with all questions to be recorded. Therefore, unanswered or technically invalid answers do not exist.

For the complete set of 294 participants (or "cases"), the following filter was used:

a) Removal of all cases that stated they used data analysis for more than 25 years in the organisation. The reason for this is, that, although predecessors of Big Data may have existed and may, for many years, have matched the definition in this research for years, the probability that the participants actually have the insights required to answer the questions is likely to be negatively correlated to the number of years. For example, the longer the actual introduction is in the past, the less accurate the answers about influences, bounded rationality, and culture are. The 25 years' limit was chosen based upon my professional experience in IT.
   After applying this filter, 259 cases remain.

b) Removal of all cases where question #9, “For what is data analysis being used” was not answered with at least one option. As every participant answered with at least one option, none were removed.
c) Answers that indicated a “don’t know” or “unsure” response variable were treated as missing for the statistics.

Limitations

The sample was not weighted to match the population representation. Therefore, this research will not establish correlations of business sectors to the findings.

Correlations leave cause and effect unconsidered. Therefore, the results that were uncovered in this research are likely to require a controlled experiment approach to identify possible causations.

Diagrams

The diagrams use the German number format, in which decimal point is described with a comma.

Participant categorisation

About half of the participants were employed in an IT-focused function. Three-quarters (76.1%) reported, that their positions included management tasks (Figure 3).

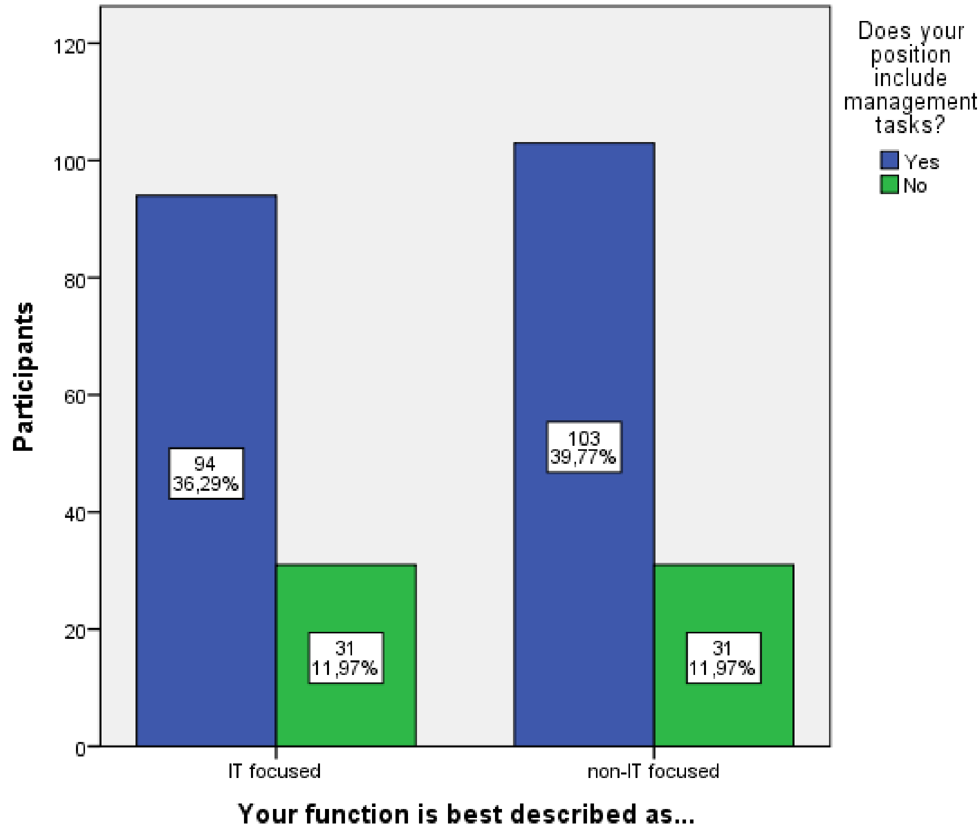

Figure 3: Participants per function and position

The organisations of the participants were diverse in sector, size, and internationality. One-third had international subsidiaries (Figure 4 & Figure 5).

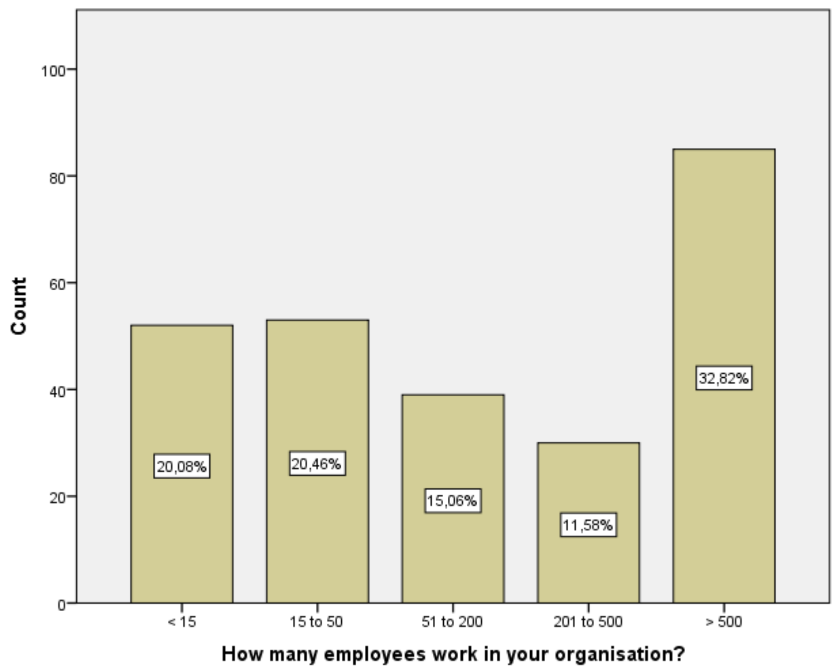

Figure 4: Organisation's employee count

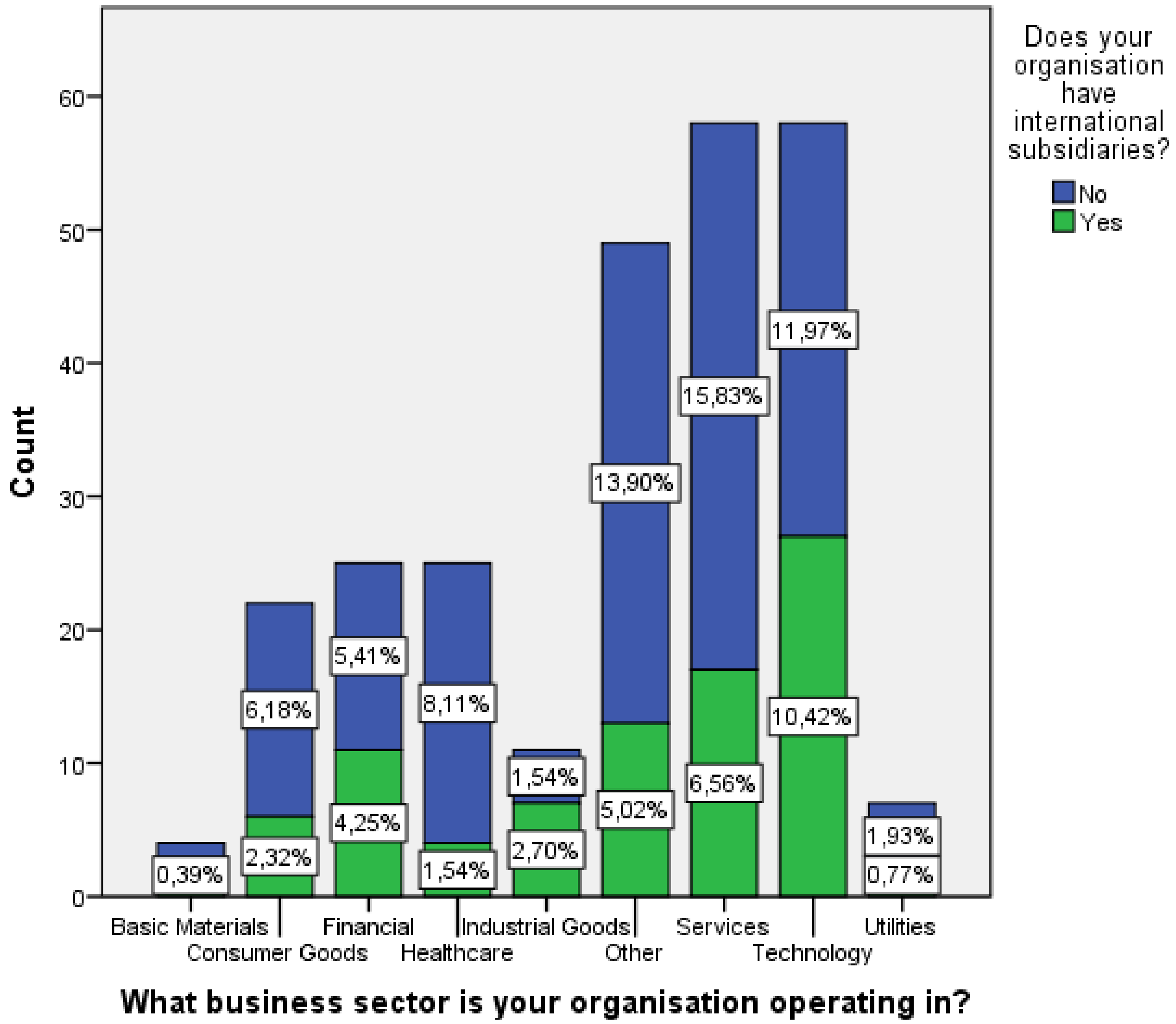

Figure 5: Categorisation of business sectors and internationality

One-third (34%) of the surveyed organisations rely on Big Data for their business models. Unsurprisingly, the services, technology, and financial sectors are leading in this regard, as seen in Figure 6.

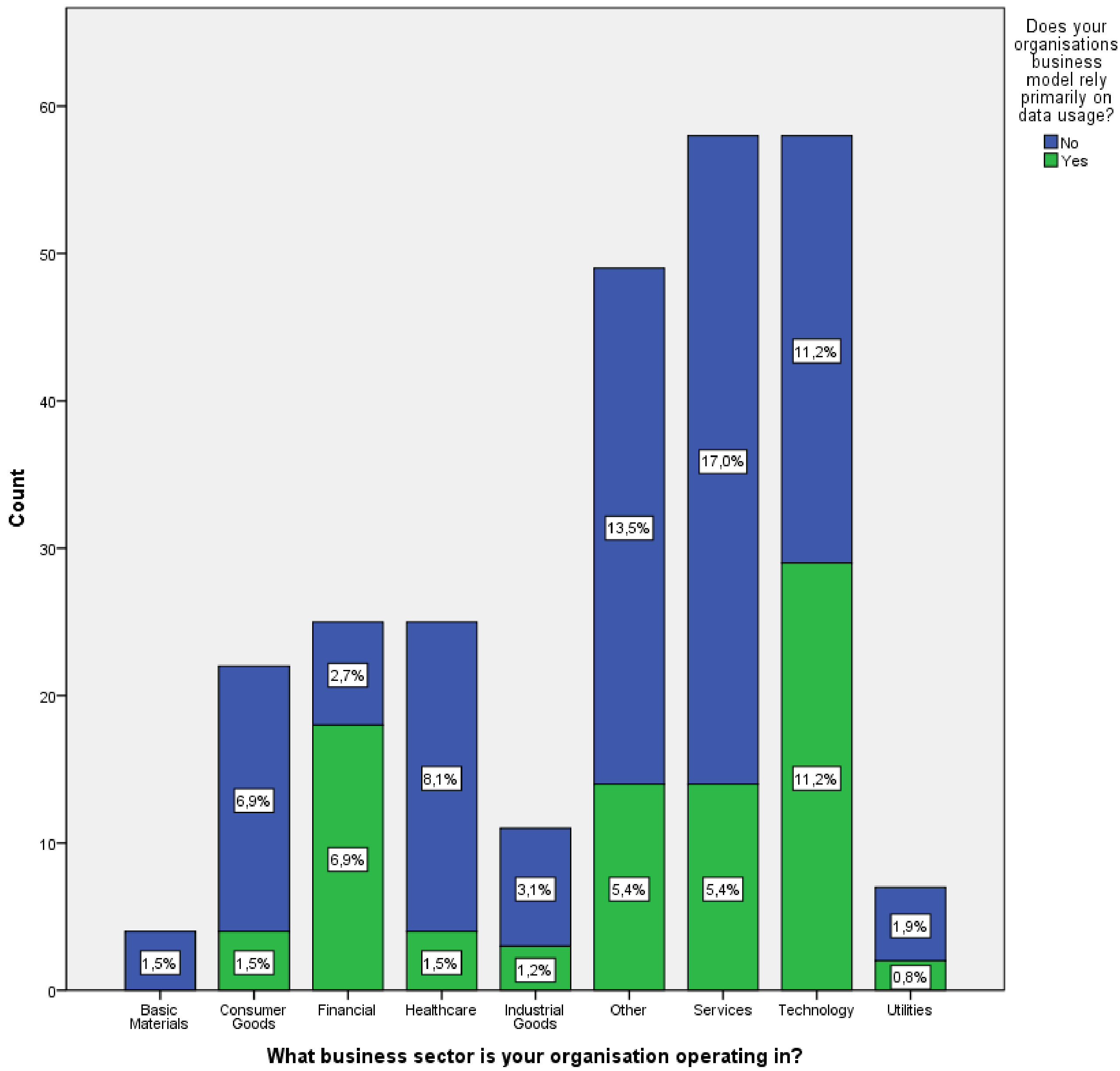

Figure 6: Business sector vs. Big Data business models

Usage and definition of Big Data

The analysis of the classification for what Big Data is used for suggests that the most prevalent usage is in the solution-identification and solution-selection phase of decision-making (Figure 7). Multiple choices were possible in this survey question.

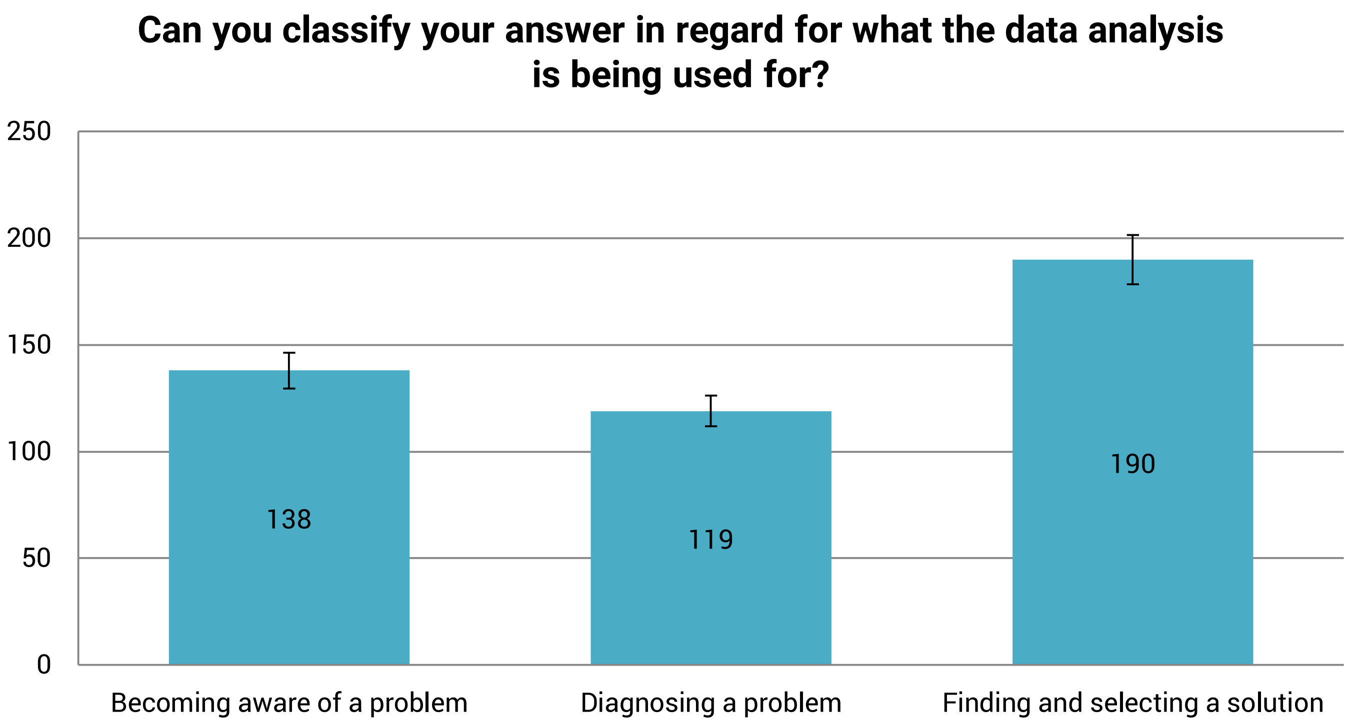

Figure 7: Big Data usage classification

The participants were asked how difficult a set of issues would be, which generally describe the four attributes of Big Data if their data analysis had to be done manually in an Excel sheet. The question was asked in the Likert-scale, coded with 1 = "strongly agree", 3 = "neither", and 5 = "strongly disagree". The mean values suggest an agreement that the Big Data research definition is matching the practical perception (Figure 8).

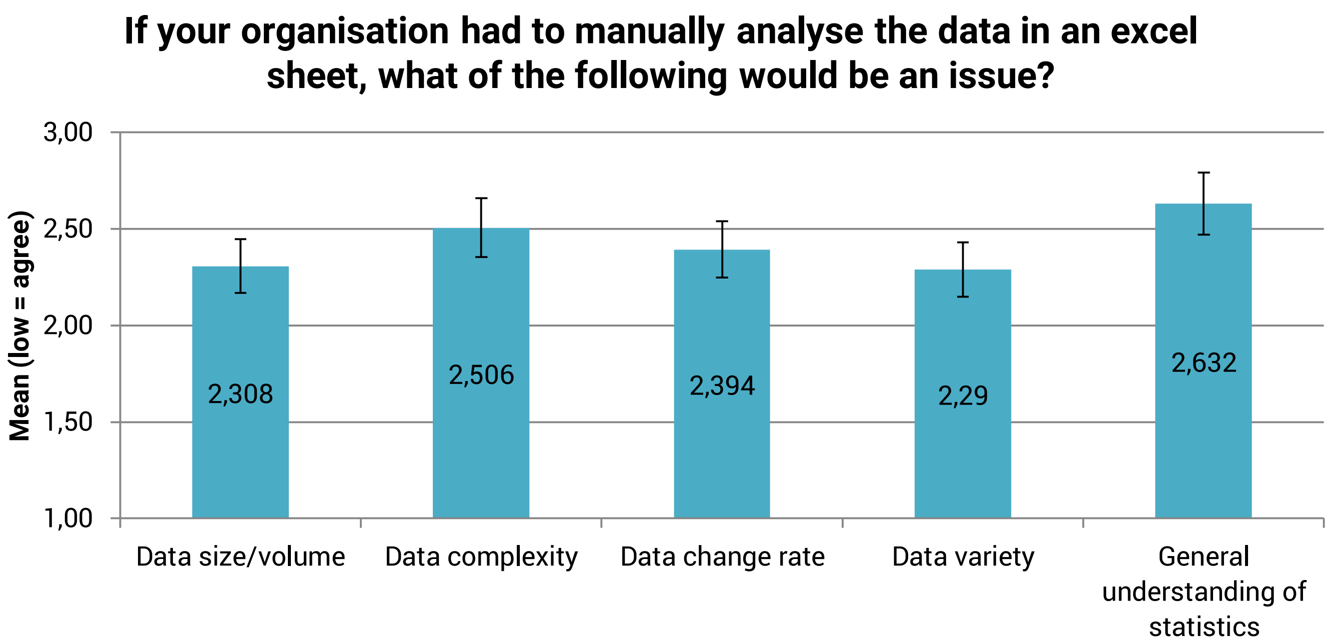

Figure 8: Big Data definition

The Big Data definition-match between the definition that was established in the research and within the surveyed organisations was arrived at by taking the above Likert values for the four Big Data definition measurements and then using this to this calculate the approval percentage to the definition by using the formula

$$x = 100 - 5(a + b + c + d - 4)$$

to account for the inverse Likert scaling. The calculation results for the participants can be seen in Figure 9. The average approval rate over all four measurements combined was 73 %.

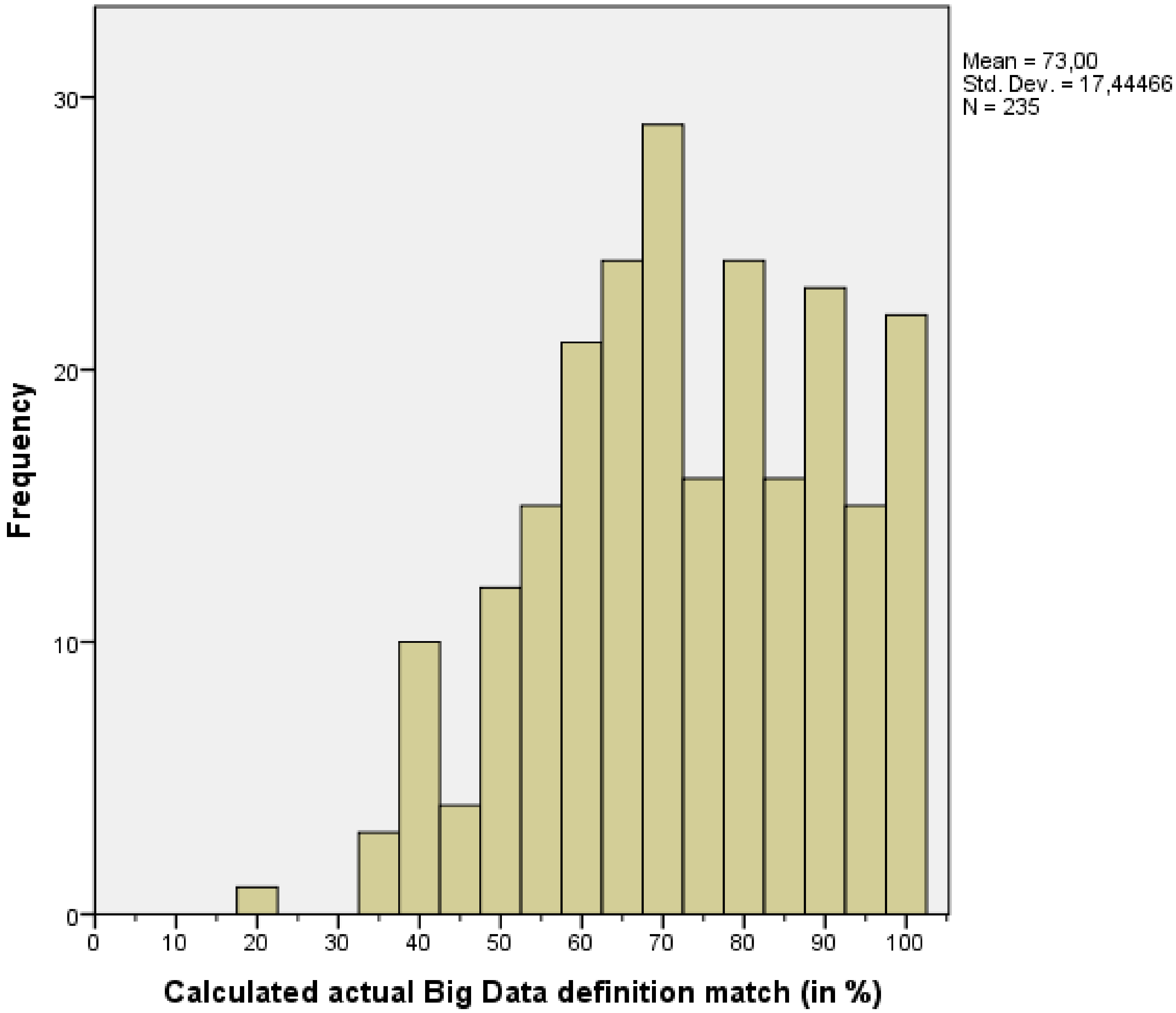

Figure 9: Calculated actual Big Data definition match

The strongest driving force for Big Data was that it would be uneconomical to conduct data analysis manually, as seen in Figure 10.

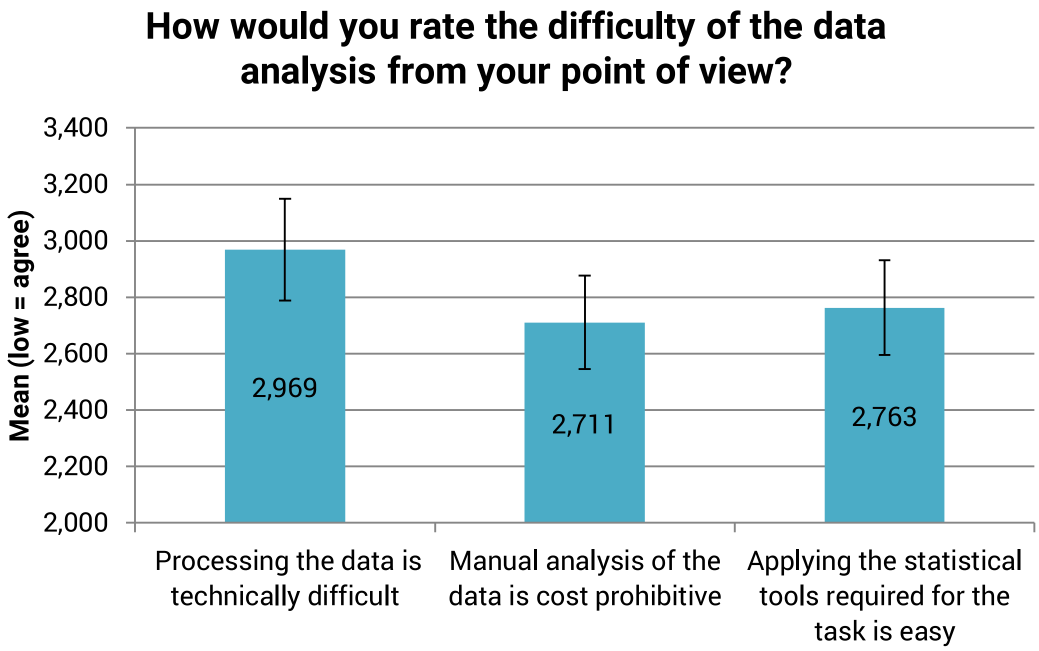

Figure 10: Big Data driving forces

Cross-correlation between the difficulties and the Big Data definition match that was calculated, showed, that with a higher agreement to the definition, the participants also have been more sensitive to both the actual technical and financial challenges of the data analysis, as seen in Table 2.

Correlations c

How would you rate the difficulty of the data analysis from your point of view?			Processing the data is technically difficult	Manual analysis of the data is cost prohibitive	Applying the statistical tools required for the task is easy
Kendall's tau_b	Calculated actual Big Data definition match	Correlation Coefficient	-,258**	-,230**	0,039
		Sig. (2-tailed)	0,000	0,000	0,450
Spearman's rho	Calculated actual Big Data definition match	Correlation Coefficient	-,329**	-,288**	0,056
		Sig. (2-tailed)	0,000	0,000	0,410

**. Correlation is significant at the 0.01 level (2-tailed).

c. Listwise N = 220

Table 2: Correlation between data analysis difficulties and Big Data definition match

According to the participants, the process to identify the required data is more difficult than the data-obtaining process (Figure 11). Of the valid responses, 44.8% stated that the identification process was "somewhat hard" or "very hard", in contrast to only 29.2% in the other category.

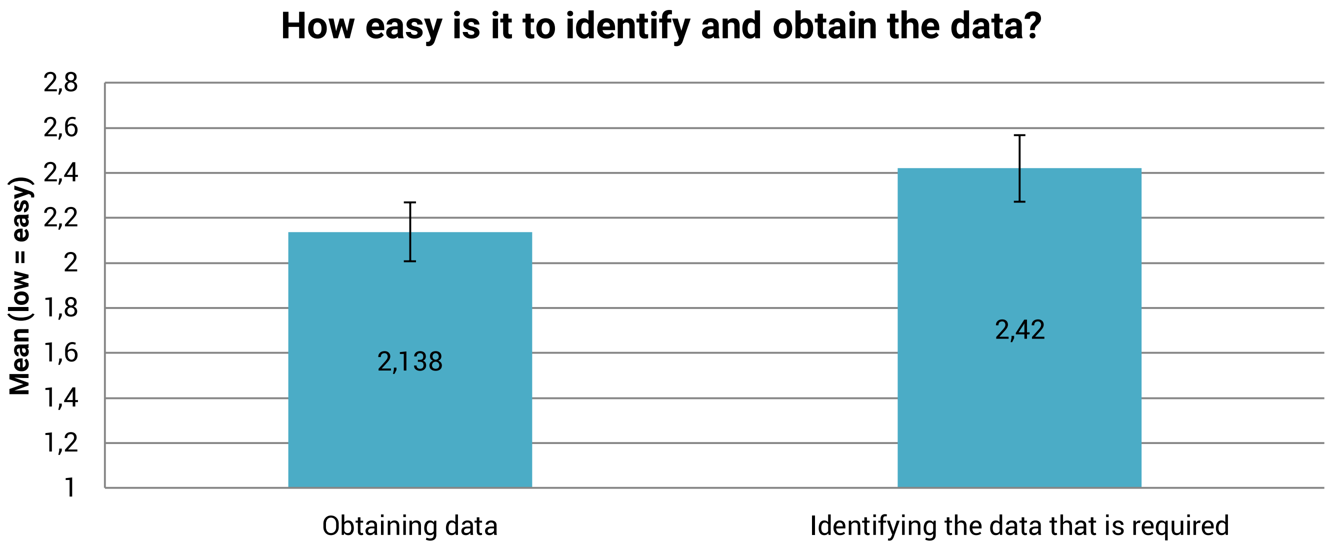

Figure 11: Data-collection challenges

Regarding the data collection frequency, 66.2% answered that the organisation "often" or "always" collects the required data in real-time (Figure 12).

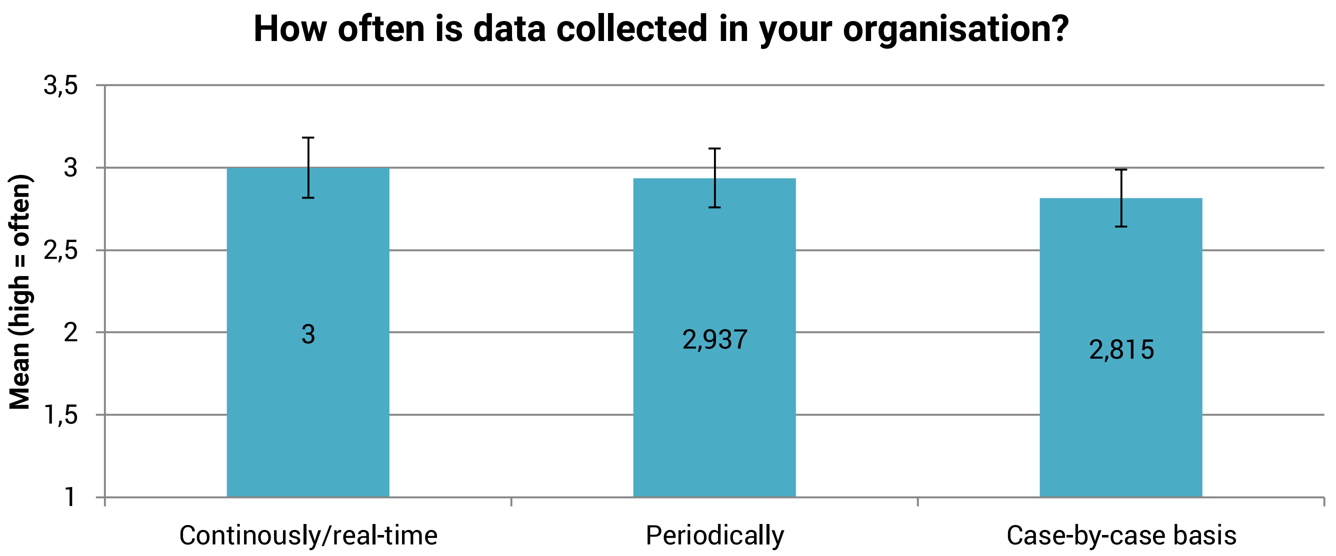

Figure 12: Data-collection frequency

The hypothesis was that real-time data collection correlates with a higher difficulty of manual analysis. Due to the coding of the questions scale, when calculating the correlation, it was expected that the coefficient would be negative. Both Kendall's Tau-b and Spearman's rho tests (Table 3) verified this assumption and the hypothesis (r_{tau} = -0.118; r_{rho} -0.141, $p < 0.05$). A scatterplot summarises the results (Figure 13).

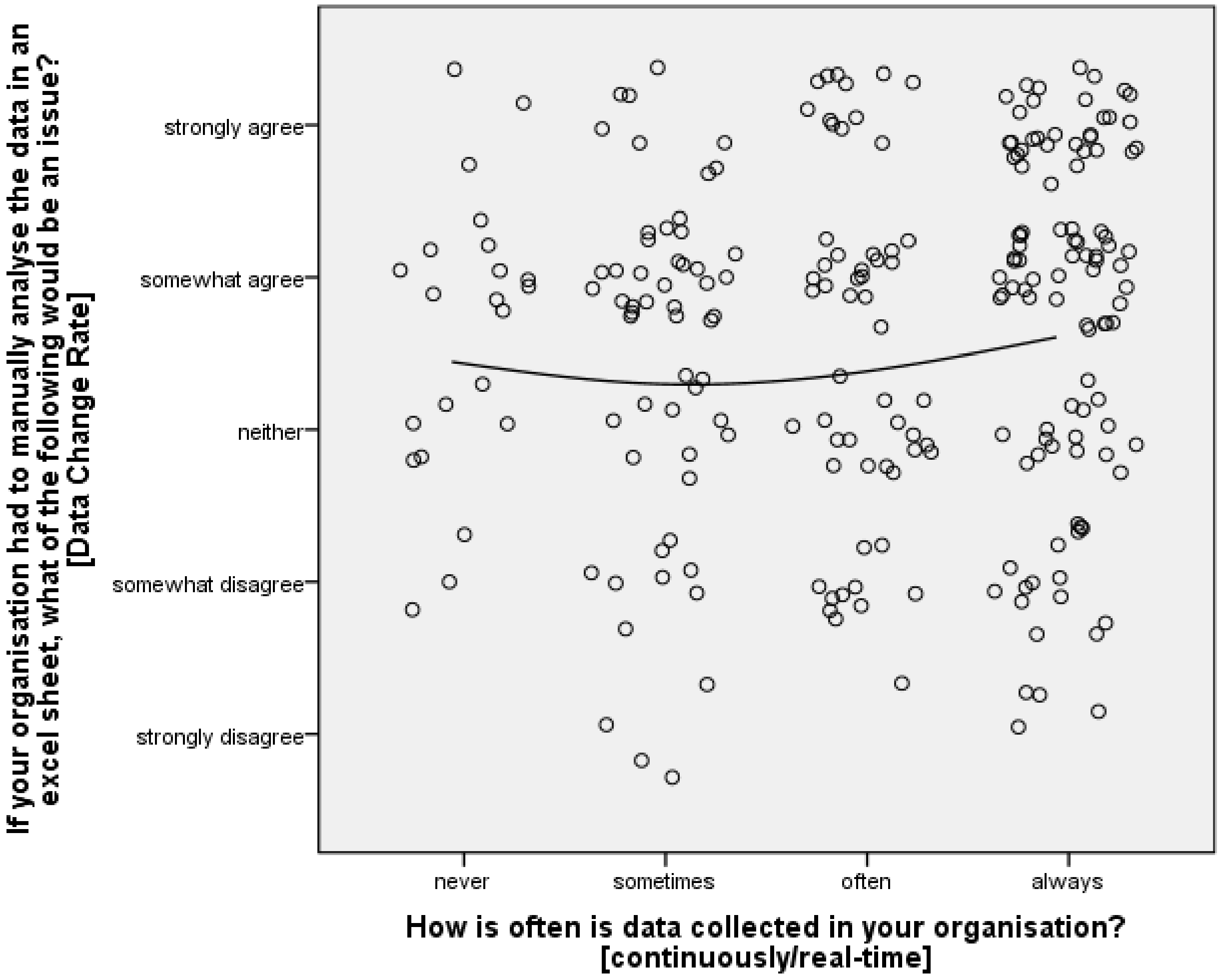

Figure 13: Scatterplot - real-time collection vs. data-change rate

Correlations c

How is often is data collected in your organisation?			continuously/real-time	periodically	on a case basis
Kendall's tau_b	Data Change Rate	Correlation Coefficient	-,118*	-0,097	-0,058
		Sig. (2-tailed)	0,035	0,084	0,307
Spearman's rho	Data Change Rate	Correlation Coefficient	-,141*	-0,114	-0,067
		Sig. (2-tailed)	0,032	0,085	0,310

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

c. Listwise N = 230

Table 3: Correlation – real-time data collection vs. data change rate

Further, it was found that only a few of the participants use external service providers to conduct data analysis and only 28% of the organisations have had specialised teams available, as can be seen in Figure 14.

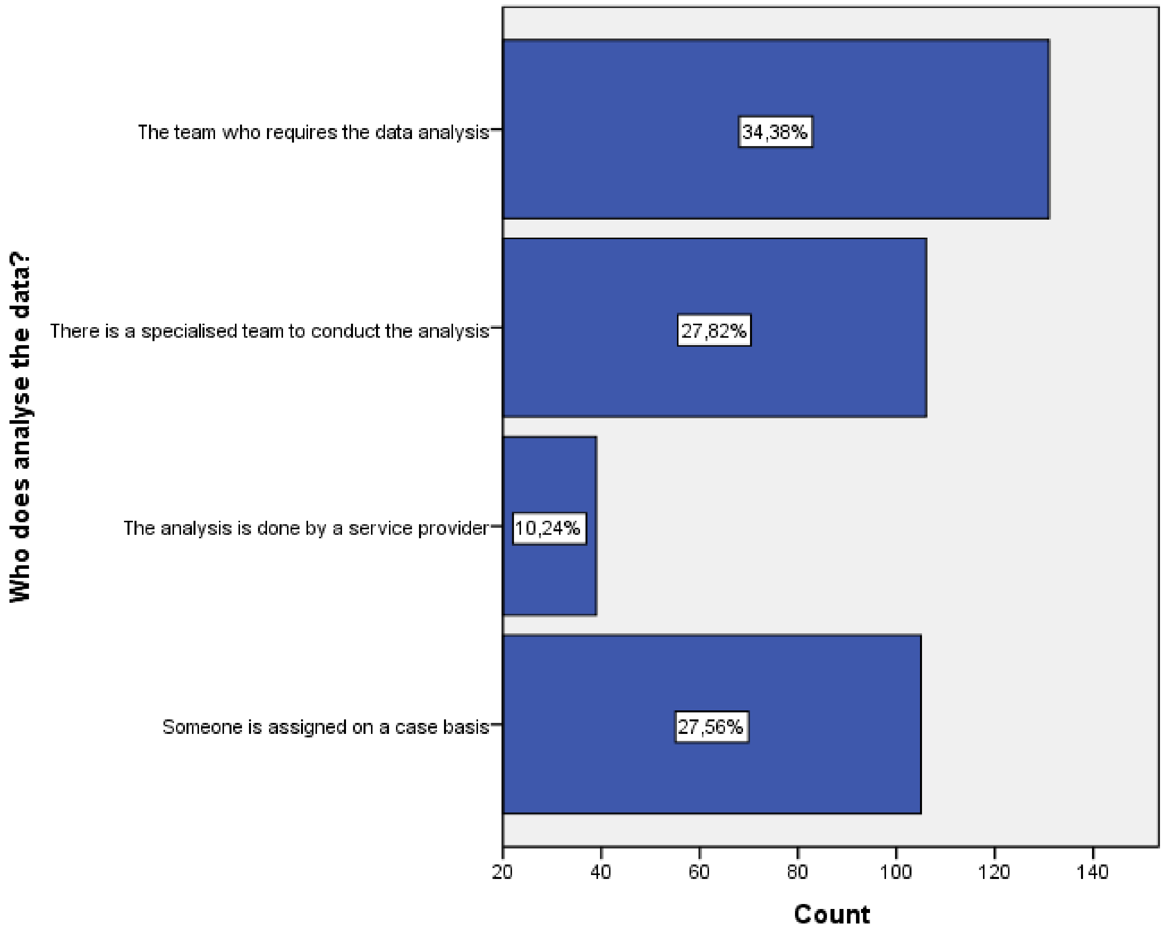

Figure 14: Who conducts data analysis

Analysis quality

As for the question of how often the data is collected, the research also investigated the frequency of the data analysis process in the organisations. The participants were asked Likert-scale questions to determine how prevalent a specific frequency is in the organisation.

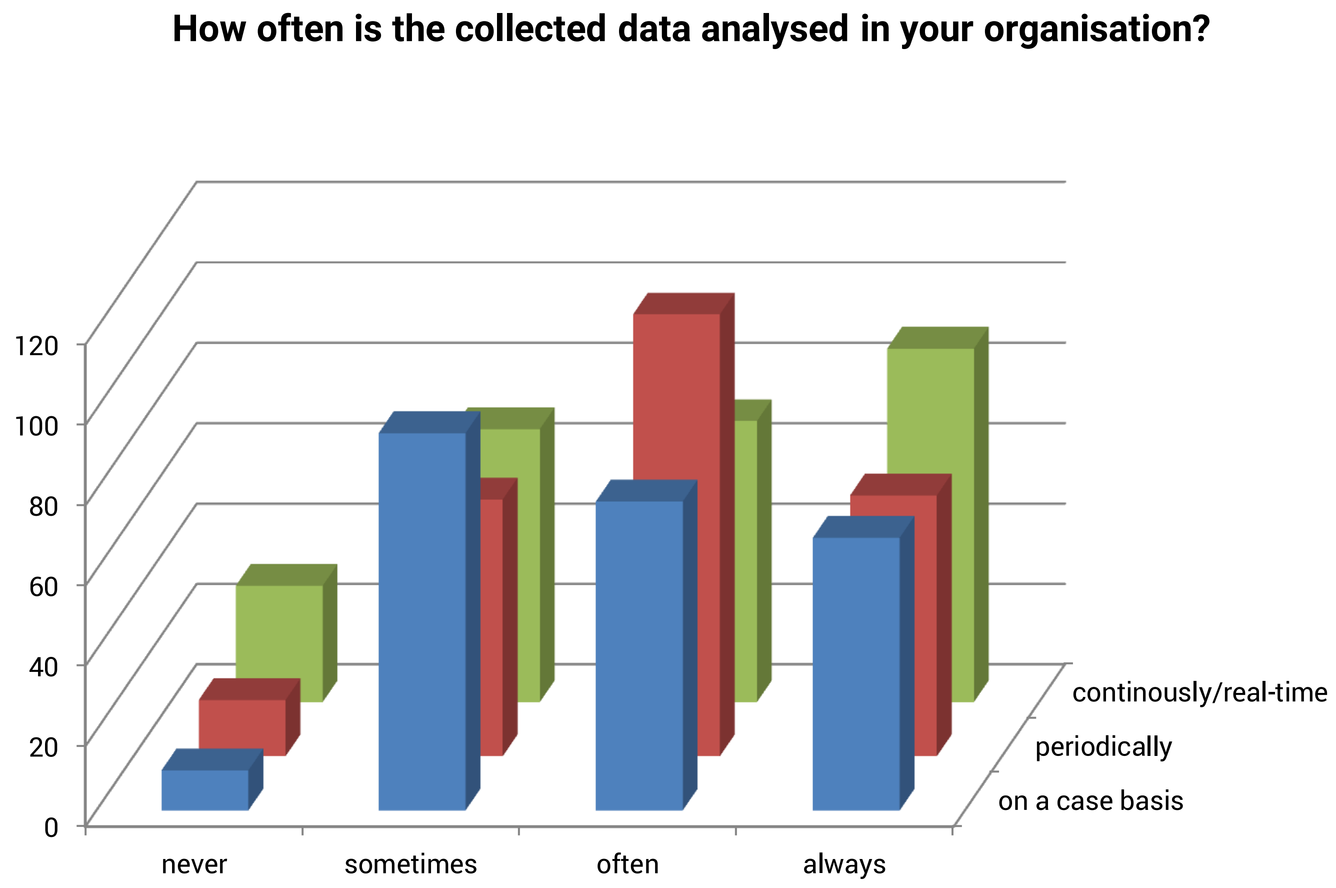

Figure 15: Data analysis frequency (detailed)

This result was tested for correlation against the results of the data collection frequency. It was discovered that the strongest correlation lies in the equidistant frequency of collection and analysis (r_{tau} > 0.655; r_{rho} > 0.691, p < 0.001). The other correlations that emerged can be explained by, for example, warehousing of the data so that the organisation can rely on short notice analyses or because of procedural requirements (Table 4). The scatterplots in Figure 16, Figure 17, and Figure 18 summarise the results.

Correlations c

			How is often is the collected data **analysed** in your organisation? [continuously/real-time]	How is often is the collected data **analysed** in your organisation? [periodically]	How is often is the collected data **analysed** in your organisation? [on a case basis]
Kendall's tau_b	How is often is data **collected** in your organisation? [continuously/real-time]	Correlation Coefficient	,672**	,155**	0,086
		Sig. (2-tailed)	0,000	0,005	0,121
	How is often is data **collected** in your organisation? [periodically]	Correlation Coefficient	,137*	,655**	,316**
		Sig. (2-tailed)	0,013	0,000	0,000
	How is often is data **collected** in your organisation? [on a case basis]	Correlation Coefficient	0,035	,286**	,709**
		Sig. (2-tailed)	0,525	0,000	0,000
Spearman's rho	How is often is data **collected** in your organisation? [continuously/real-time]	Correlation Coefficient	,711**	,177**	0,096
		Sig. (2-tailed)	0,000	0,006	0,138
	How is often is data **collected** in your organisation? [periodically]	Correlation Coefficient	,157*	,691**	,351**
		Sig. (2-tailed)	0,015	0,000	0,000
	How is often is data **collected** in your organisation? [on a case basis]	Correlation Coefficient	0,038	,322**	,749**
		Sig. (2-tailed)	0,554	0,000	0,000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

c. Listwise N = 243

Table 4: Correlation - data analysis frequency

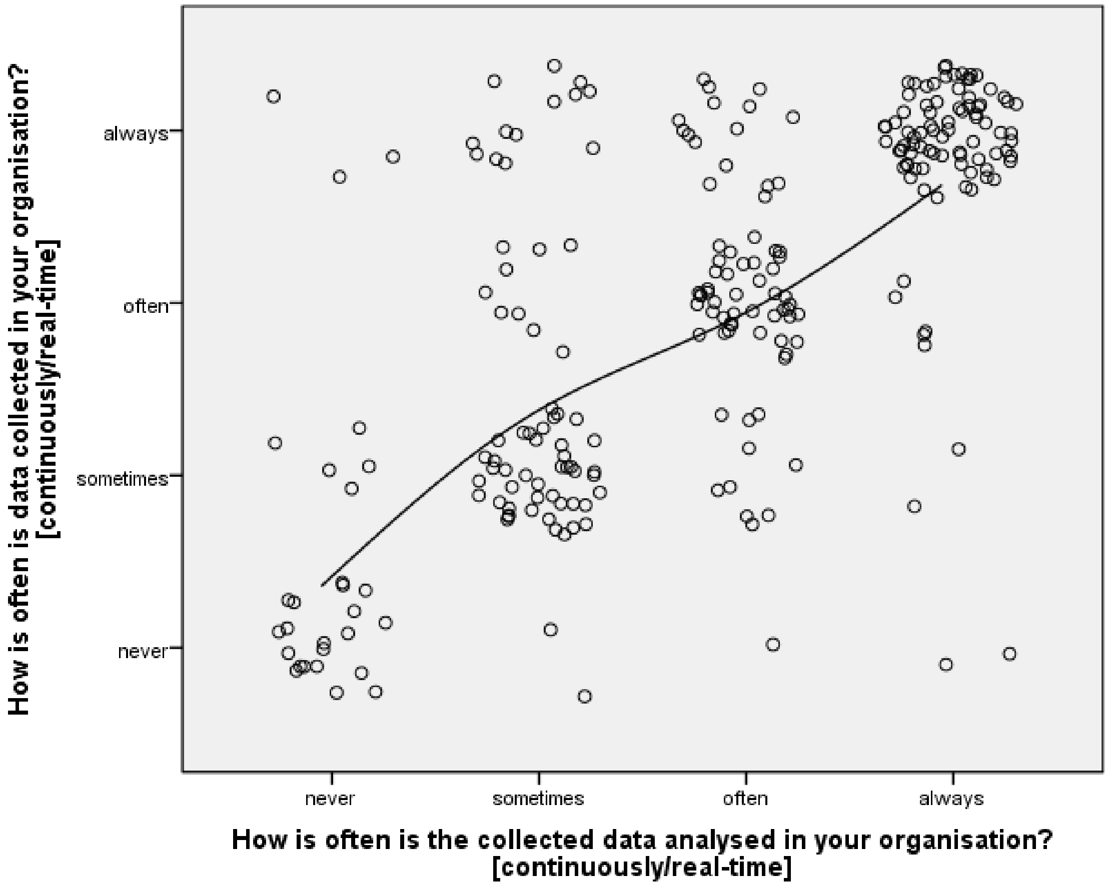

Figure 16: Scatterplot - data collection vs. data analysis (real-time)

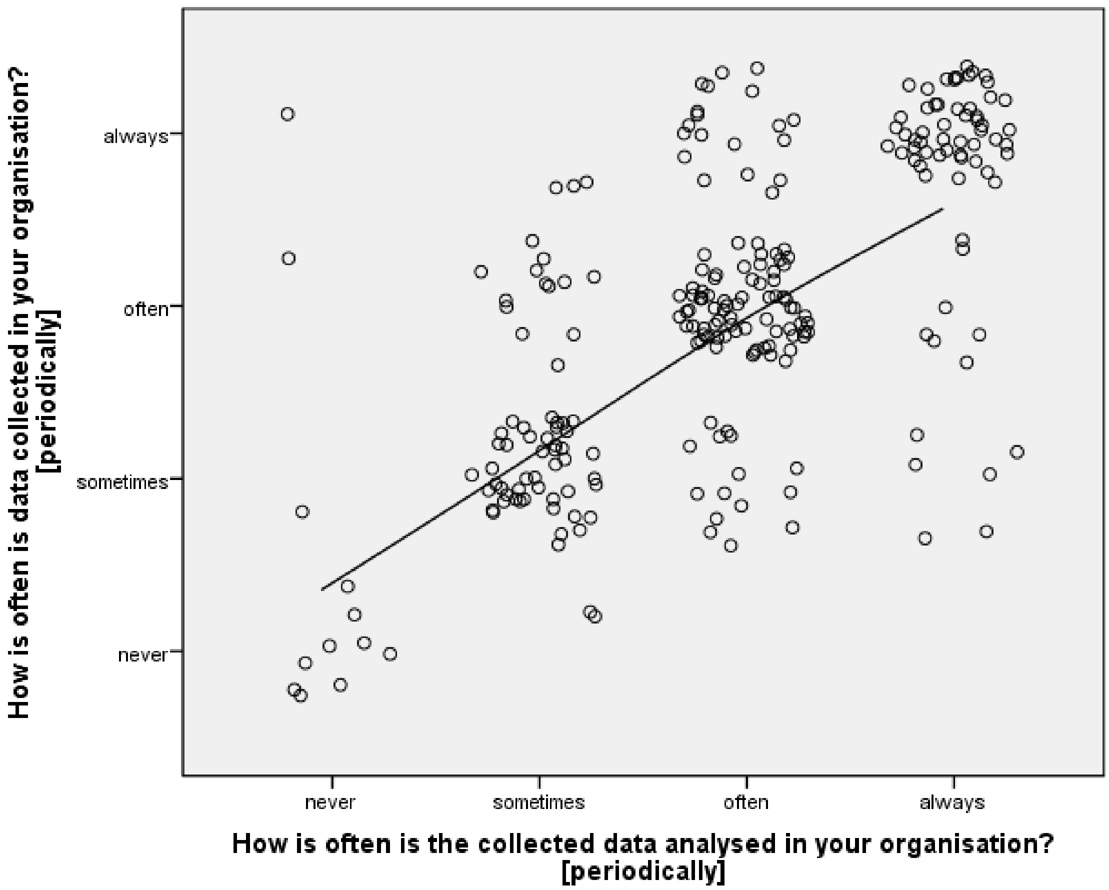

Figure 17: Scatterplot - data collection vs. data analysis (periodically)

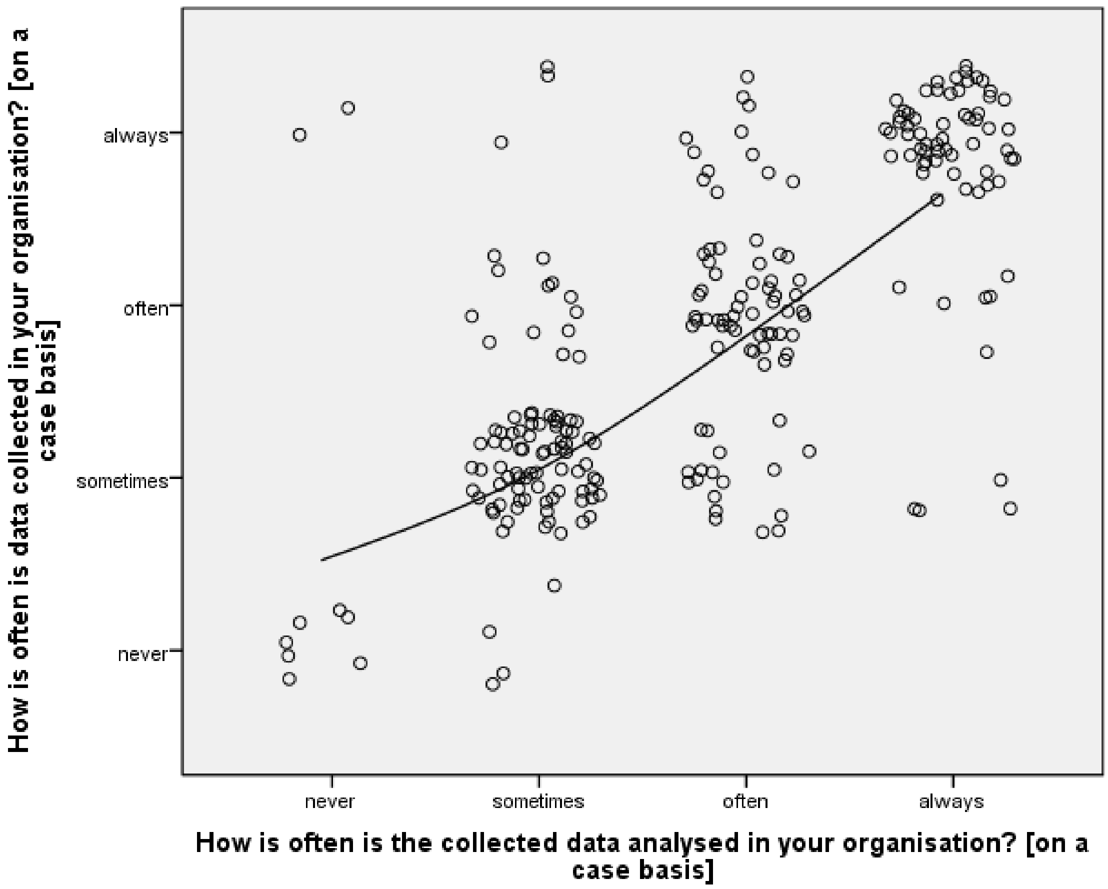

Figure 18: Scatterplot - data collection vs. data analysis (on a case-by-case basis)

The analysis task quality is seen as mostly positive, but with room for improvement in the input data verification and quality (Figure 19).

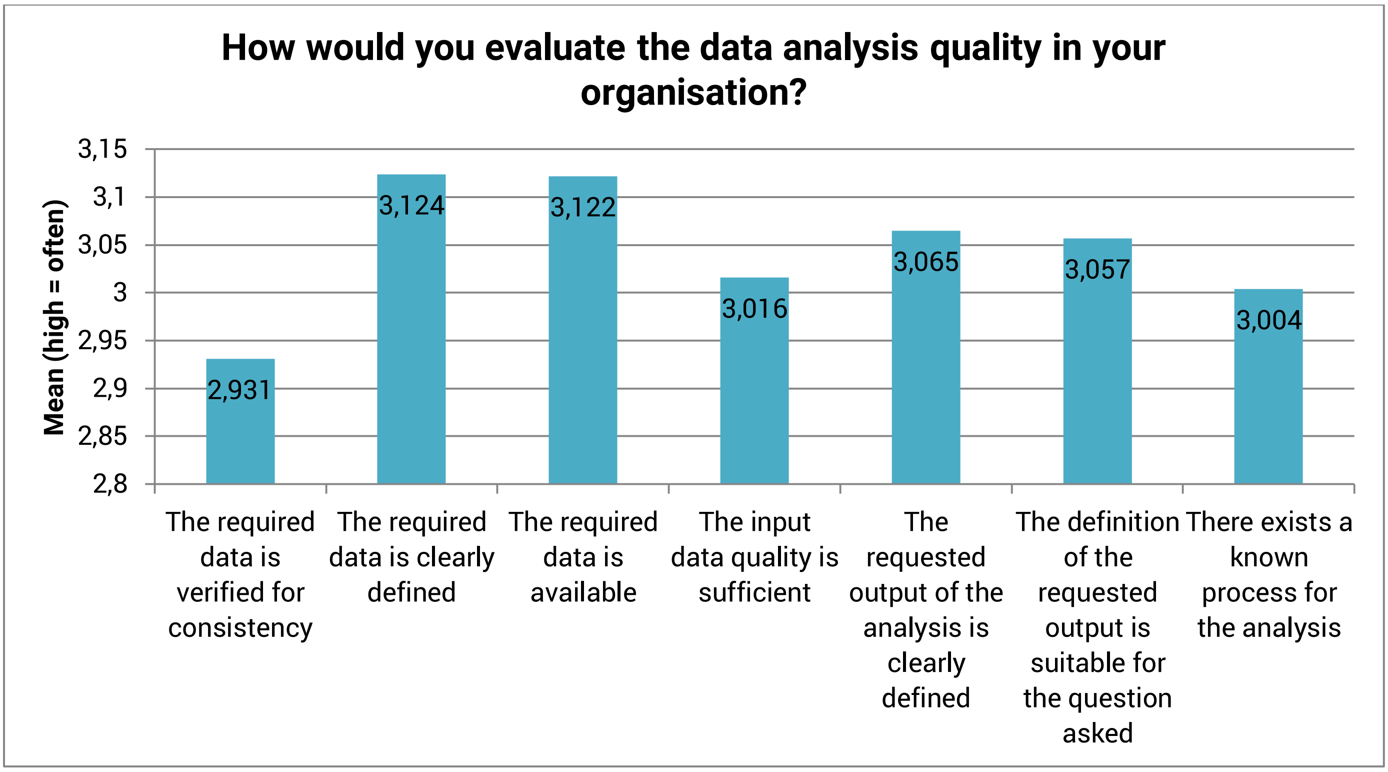

Figure 19: Data analysis quality

If the variables of data verification and data quality are plotted against each other, a pattern emerges that indicates a correlation (Figure 20). This assumption stands true after calculating the correlations and significance (Table 5), with r_{tau} = 0.348; r_{rho} = 0.395, p < 0.01. Therefore, the statement can be made that organisations that verify input data more often, achieve a better data quality. This unsurprisingly stands true when viewed in the light of process improvement and organisational learning.

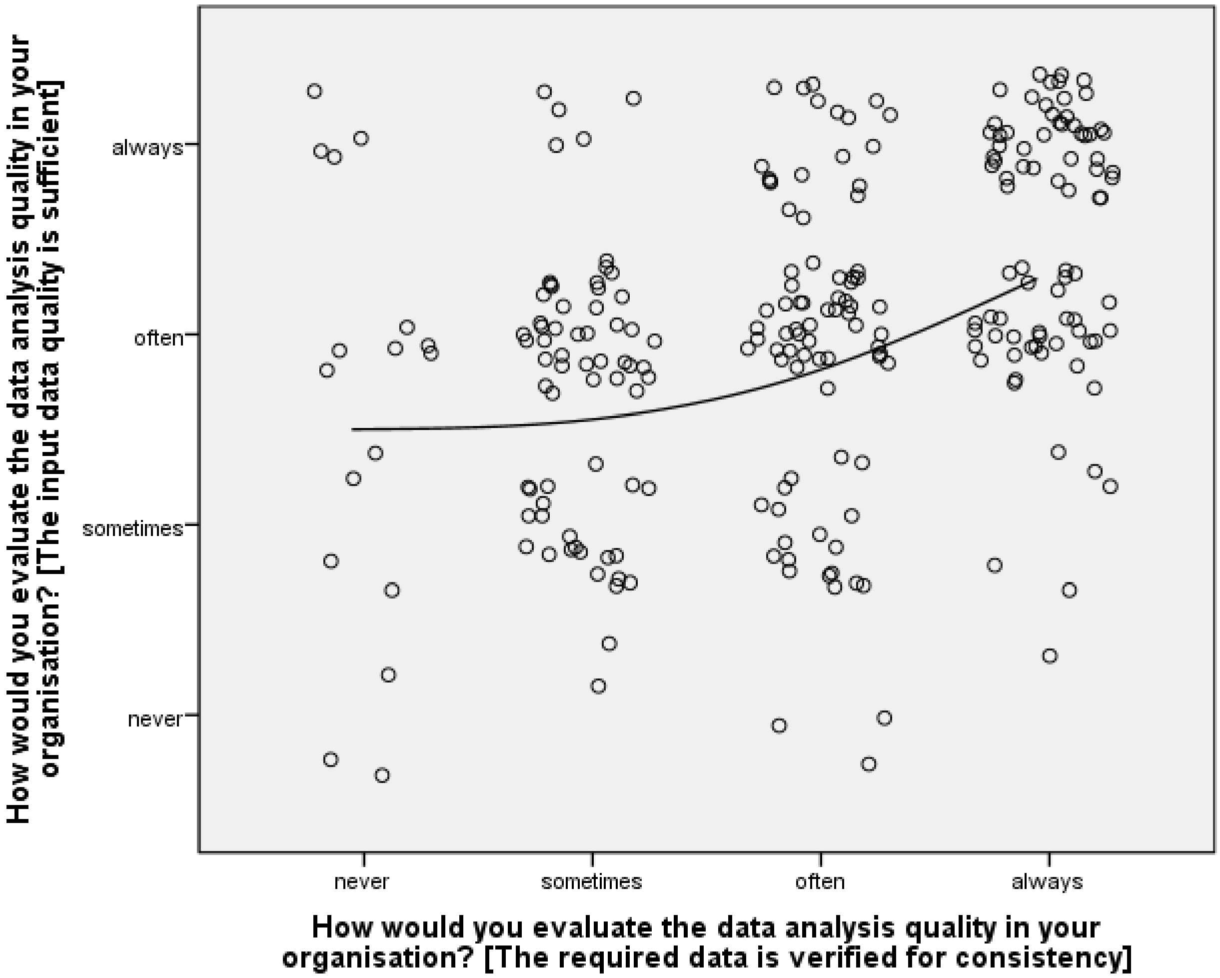

Figure 20: Scatterplot - data quality vs. data verification

Correlations b

			How would you evaluate the data analysis quality in your organisation? [The required data is verified for consistency]
Kendall's tau_b	How would you evaluate the data analysis quality in your organisation? [The input data quality is sufficient]	Correlation Coefficient	,348**
		Sig. (2-tailed)	0,000
Spearman's rho	How would you evaluate the data analysis quality in your organisation? [The input data quality is sufficient]	Correlation Coefficient	,395**
		Sig. (2-tailed)	0,000

**. Correlation is significant at the 0.01 level (2-tailed).

b. Listwise N = 245

Table 5: Correlations - data quality vs. data verification

The analysis quality results, on the other hand, were more interesting. The output of Big Data was, on average, “often produced timely” (mean around a Likert-value of “3”) and “often answered the question [that was expected to be solved]” (Figure 21). The other variables are in the middle of the possible answer-spectrum.

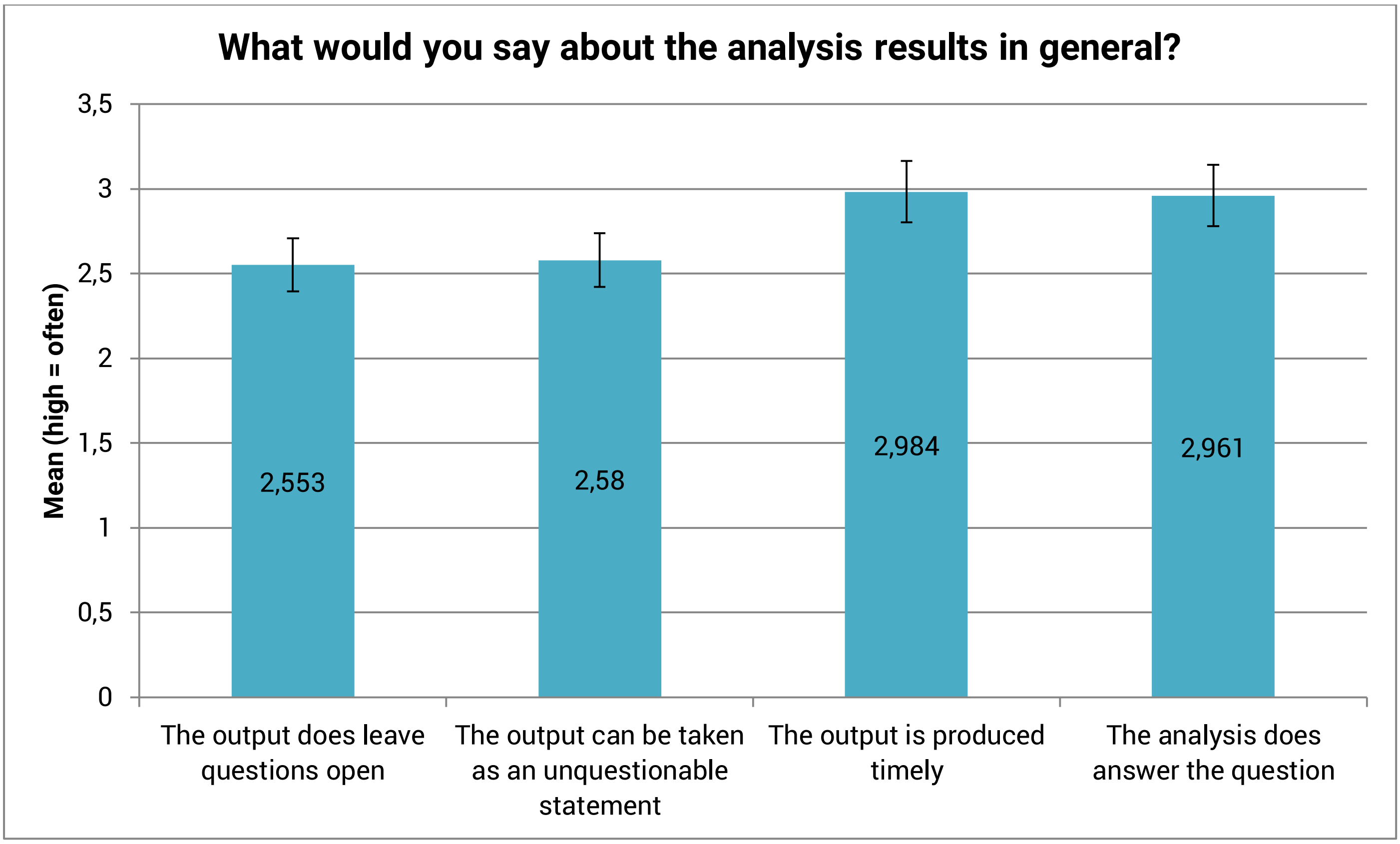

Figure 21: Analysis results quality

Decision-making in the organisation

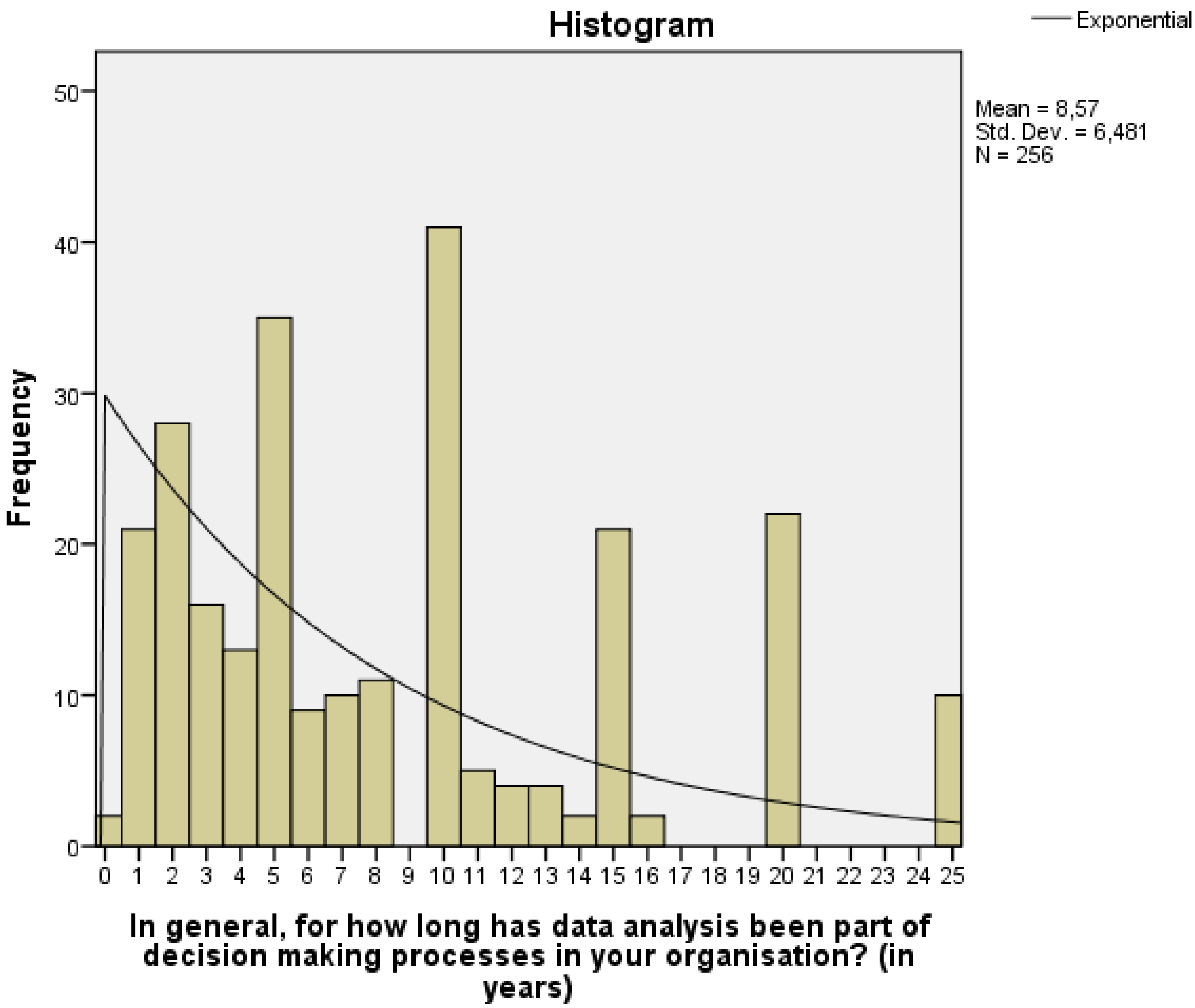

Figure 22: For how long is Big Data utilised

Figure 22 shows that Big Data might be around for longer than some might have expected. The peaks at the five-year intervals were expected, as these are commonly used by participants as approximations (Engel et al., 2012). After ignoring these peaks, it seems that the chart is following an exponential growth curve, which might be reasonable as Big Data is still a relatively new phenomenon to the majority of organisations.

To differentiate from one of the earlier questions of whether a management position was occupied, it was asked if the participant was included as an active part in the decision-making processes. This for one instance acted as a control question. However, as the question was not identically asked, the variance was expected, and can be seen in Figure 23. Additionally, it might be argued that management tasks do not necessarily include decision-making involvement and further, the interpretation of the question will differ between participants, e.g. supervisory tasks could be seen as management without decision-making power. Overall, I suggest that the differences are within reasonable parameters.

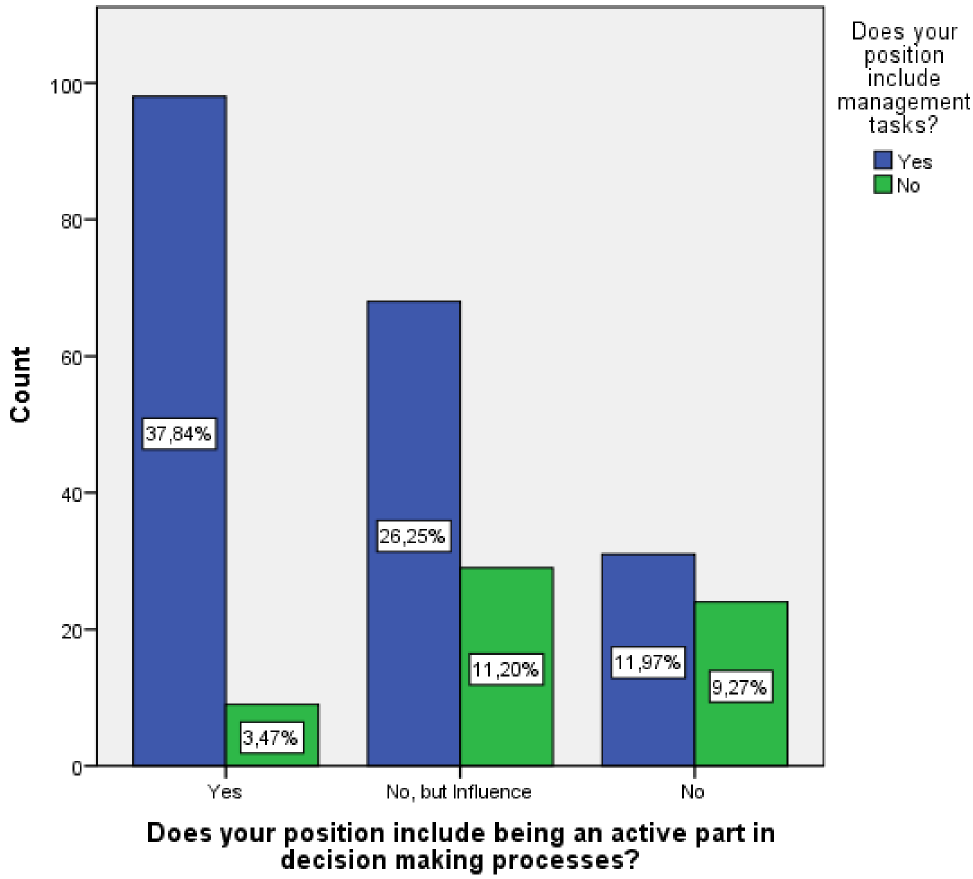

Figure 23: Crosschecking management vs. decision-making involvement

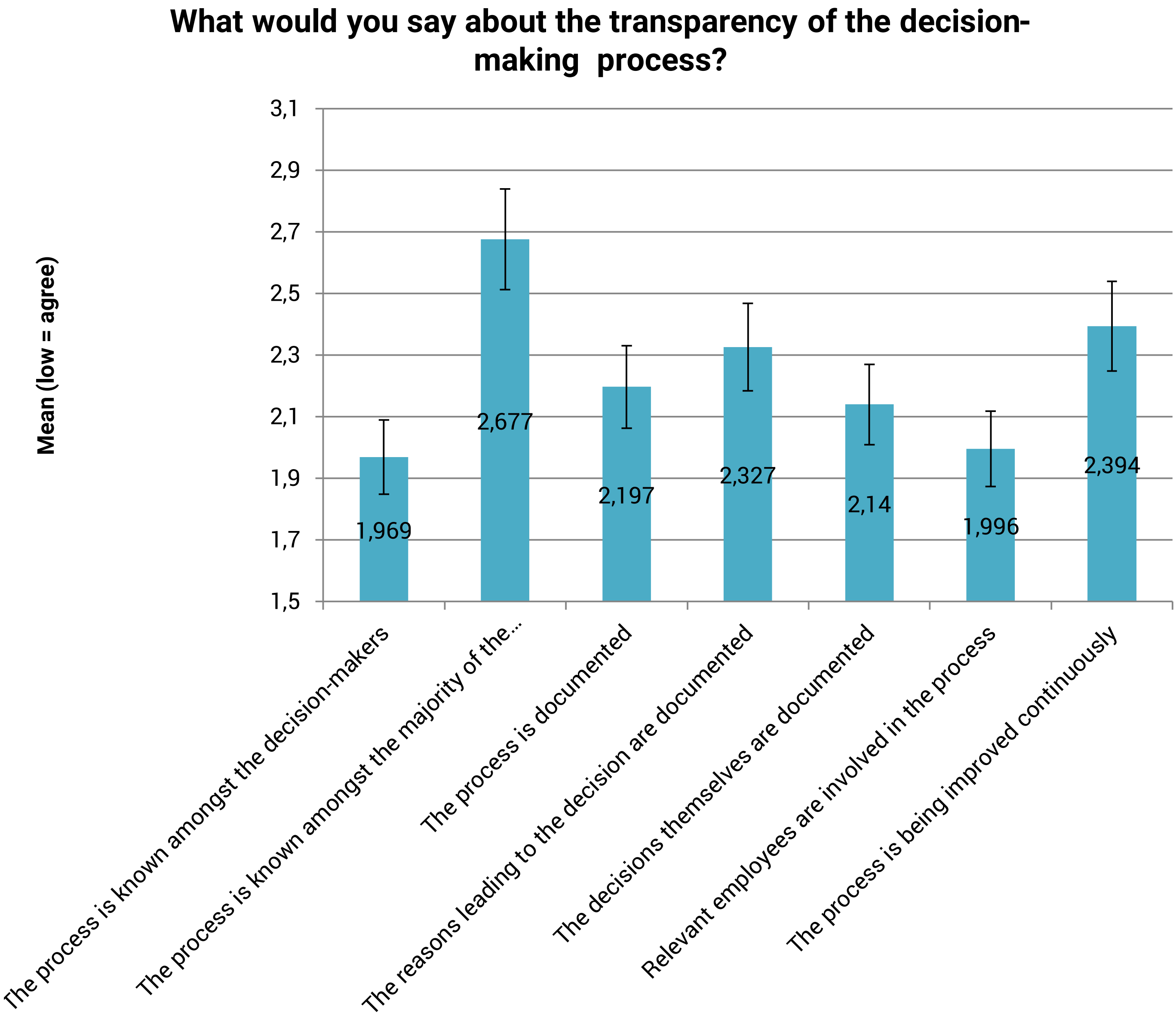

Figure 24: Decision-making process transparency

Figure 24 shows the responses to the decision-making process transparency questions. This initially was to be used as a variable for determining if there is a correlation between the transparency and bounded rationality reduction due to Big Data. Unfortunately, no significant correlation could be established. Tests of bounded rationality against hierarchy change also were inconclusive.

The coded Likert-scale questions that ask about the current and previous influences in decision-making can be subtracted from each other to calculate the difference, e.g. the change, between each. This calculated delta value was used as a variable for further correlation tests. To ensure that the delta itself is significant, a paired sample T-test was conducted, as seen in Table 6.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	c_Intuiton - p_Intuition	0,118	0,813	0,055	0,010	0,226	2,155	219	**0,032**
Pair 2	c_Data - p_Data	-0,339	0,808	0,054	-0,446	-0,232	-6,246	220	**0,000**
Pair 3	c_ExternalConsultants - p_ExternalConsultants	-0,014	0,764	0,052	-0,117	0,089	-0,268	214	**0,789**
Pair 4	c_PersonalExperience - p_PersonalExperience	0,093	0,750	0,050	-0,006	0,191	1,859	226	**0,064**
Pair 5	c_Creativity - p_Creativity	0,084	0,748	0,050	-0,014	0,183	1,693	224	**0,092**
Pair 6	c_Logic - p_Logic	-0,075	0,734	0,049	-0,171	0,021	-1,537	226	**0,126**
Pair 7	c_Reasoning - p_Reasoning	0,035	0,777	0,052	-0,066	0,137	0,685	225	**0,494**
Pair 8	c_Skills - p_Skills	0,026	0,837	0,055	-0,083	0,135	0,474	228	**0,636**

Table 6: Paired sample T-test for decision-making influence factors

The paired sample T-test for the pre- and post-values of the influencing factors show that there is a significant change ($p < 0.05$) in intuition towards weaker influence. Further, there is also a significant change ($p < 0.001$) in the factor data towards higher influence. The other factors are not significant. Therefore, it might be suggested that the initial assumption, in which Big Data reduces bounded rationality as an effect of reducing the "intuition" factor in decision-making (Kahneman, 2002), is true.

Analysis between the intuition factor and decision-making transparency itself revealed no correlations.

Another dimension that was analysed is the change in organisations for several generic attributes including the organisational processes and culture. On average, the change has been perceived to be positive, as shown in Figure 25. However, for every category there were about 5 - 15% of participants that reported that the change had led to worse conditions.

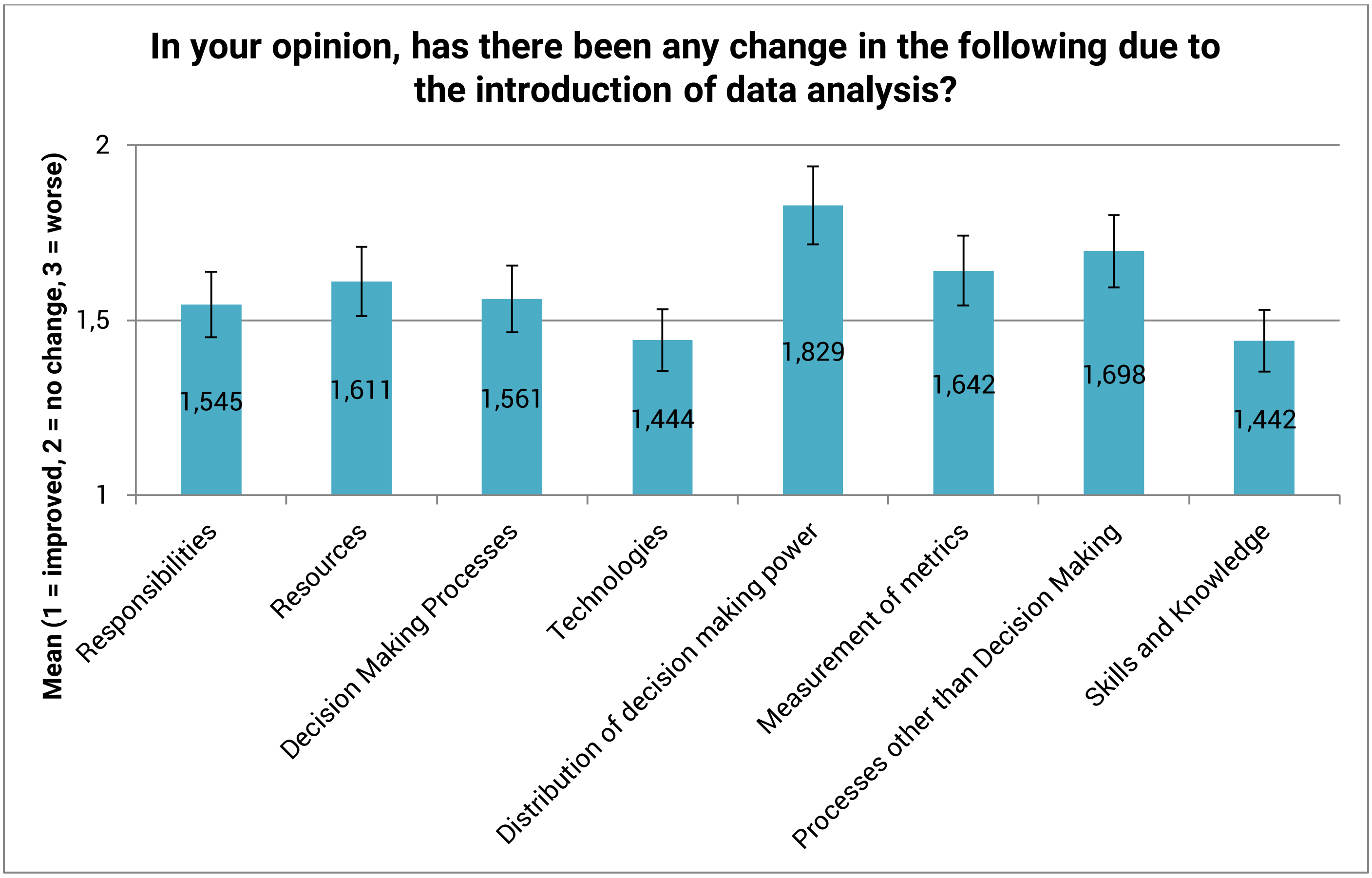

Figure 25: Change of generic attributes

Additionally, the changes in organisational hierarchy were evaluated against the previous findings to find out if there is any correlation between the surveyed variables. Only two variables showed a significant correlation, as shown in Table 7.

Correlations c

			Responsibilities	Resources	Decision-Making power	Power over the processes	Setting success metrics	Checking success against metrics	Skills and Knowledge
Kendall's tau_b	Distribution of decision-making power in the organisation	Correlation Coefficient	0,096	-0,016	0,098	,173**	,219**	0,088	0,008
		Sig. (2-tailed)	0,148	0,806	0,139	0,009	0,001	0,185	0,908
Spearman's rho	Distribution of decision-making power in the organisation	Correlation Coefficient	0,103	-0,018	0,103	,183*	,234**	0,094	0,008
		Sig. (2-tailed)	0,163	0,812	0,165	0,013	0,001	0,202	0,909

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

c. Listwise N = 185

Table 7: Correlations - distribution of decision-making power vs. generic attributes

After producing a scatter plot for these correlations, an interesting picture emerges (Figure 26 & Figure 27). The figures show that the introduction of Big Data was perceived more positively when the decision-making power shifted towards the lower hierarchies or at least when there was no change in power.

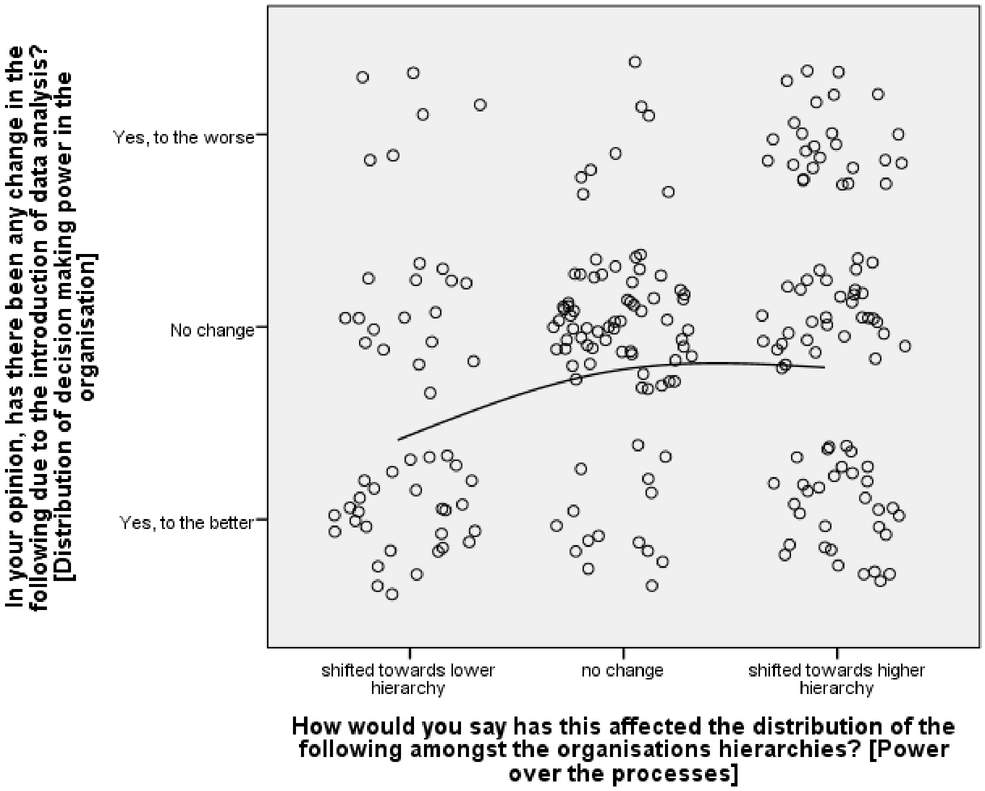

Figure 26: Scatterplot - distribution of decision-making power vs. power over the processes

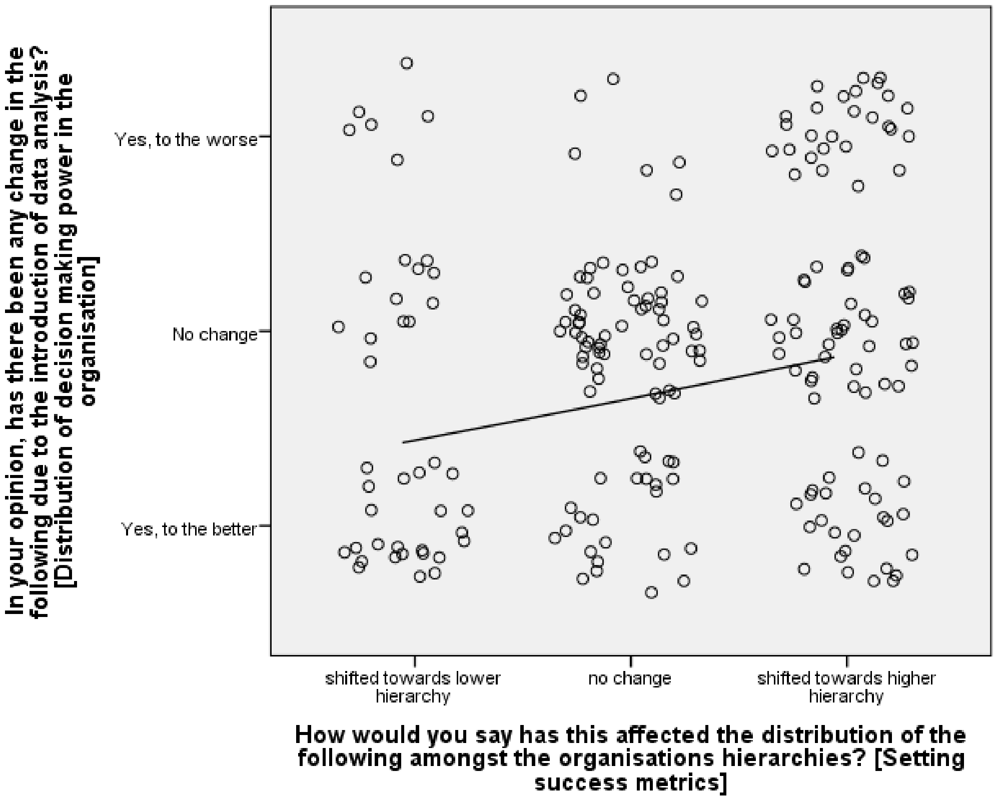

Figure 27: Scatterplot - distribution of decision-making power vs. setting success metrics

The measurement of influence-change to the decision-making process was mostly successful for the organisations, as seen in Figure 28. There was no significant correlation between the measurement success and measurement metrics.

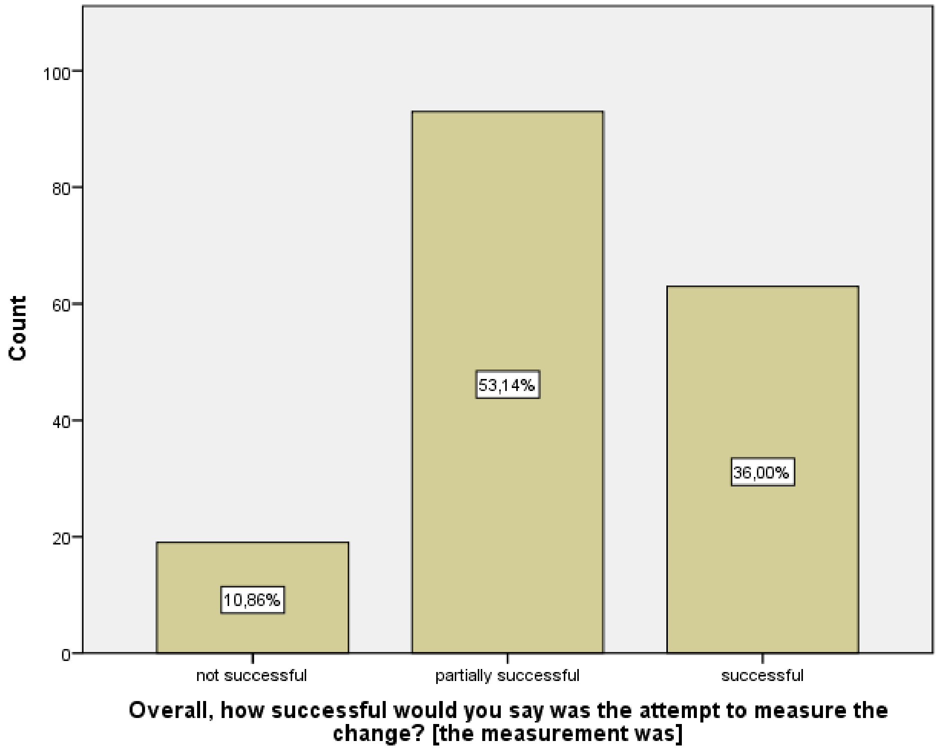

Figure 28: Measurement success of decision-making influences

The majority of executives were supporting the implementation of Big Data, as seen in Figure 29.

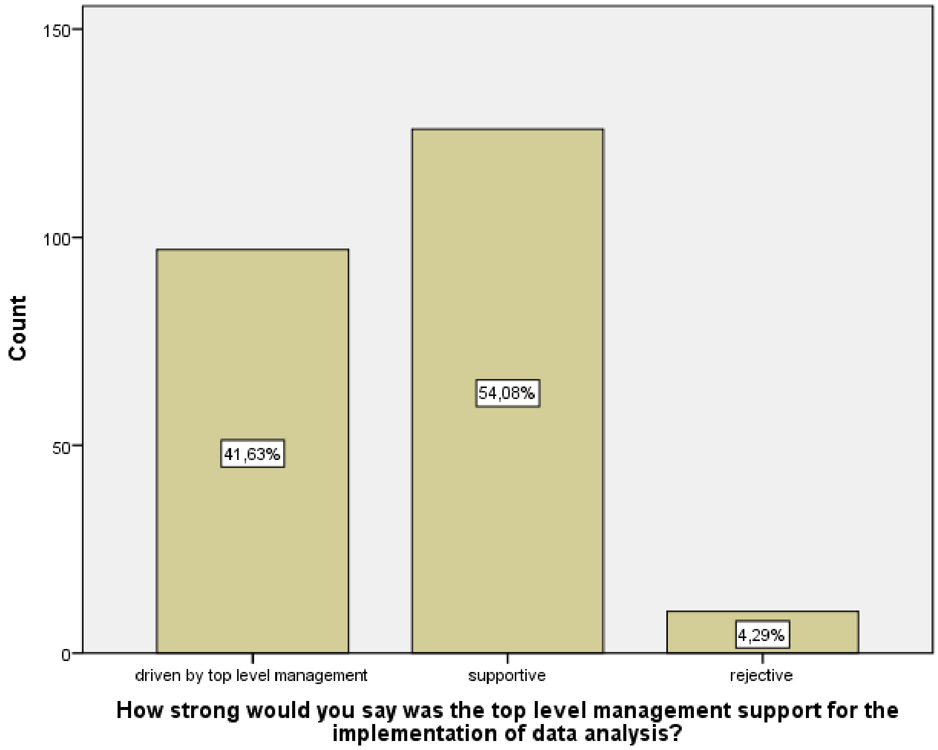

Figure 29: Executive support

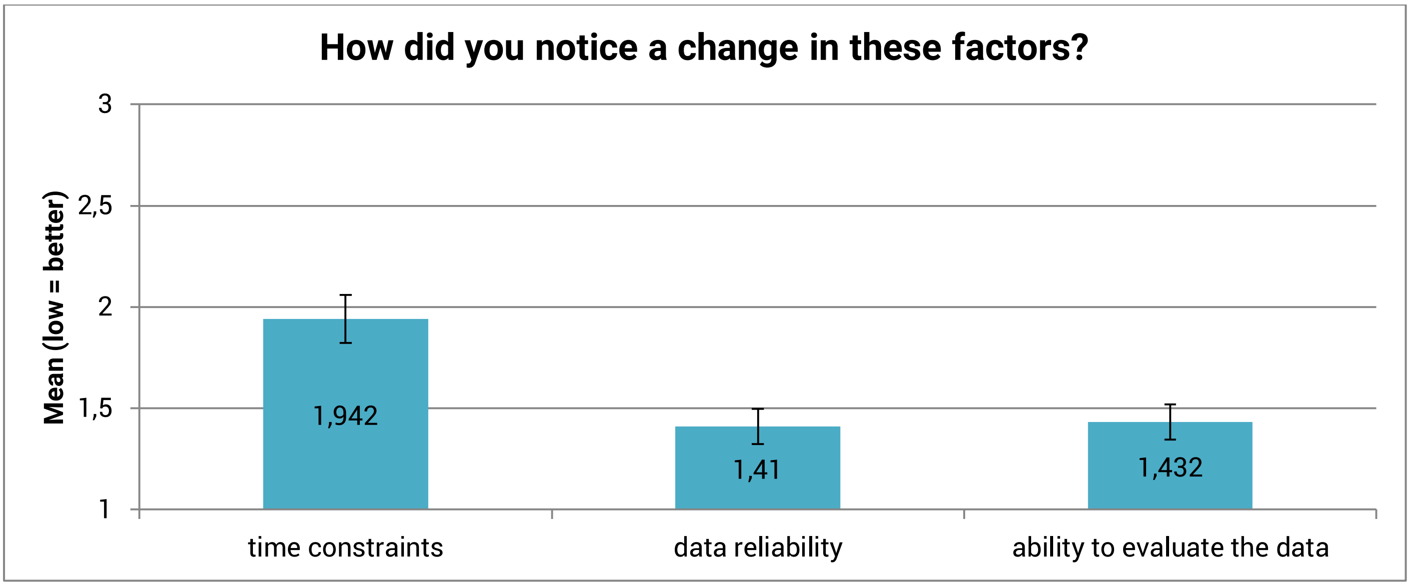

Figure 30: Changes in constraints due to Big Data

In Figure 30 the average change (from 1 = “better” to 3 = “worse”) is shown. About 23% of the participants stated that the time constraints got worse due to the implementation of Big Data. On average, time constraints did not change. The negative responses in the other categories are negligible (< 6%).

Organisational culture

Additionally, three dimensions of the organisations culture were measured by asking questions about information flow, relationship and goals that are derived from the I-Space definition (Boisot, 1998). From this, a categorisation of the culture was planned, that would allow for further correlation tests against the previously-uncovered findings. Unfortunately, for some dimensions the corresponding control pair variables showed no significant correlation to each other, as is shown in Table 8 and Table 9 for the more obvious control questions.

Correlations

			What do you think about the information flow in your organisation? [information is abstract]	What do you think about the information flow in your organisation? [information is concrete]
Kendall's tau_b	What do you think about the information flow in your organisation? [information is abstract]	Correlation Coefficient	1,000	-0,094
		Sig. (2-tailed)		0,076
		N	248	248
	What do you think about the information flow in your organisation? [information is concrete]	Correlation Coefficient	-0,094	1,000
		Sig. (2-tailed)	0,076	
		N	248	256
Spearman's rho	What do you think about the information flow in your organisation? [information is abstract]	Correlation Coefficient	1,000	-0,100
		Sig. (2-tailed)		0,117
		N	248	248
	What do you think about the information flow in your organisation? [information is concrete]	Correlation Coefficient	-0,100	1,000
		Sig. (2-tailed)	0,117	
		N	248	256

Table 8: Correlation of information - abstract vs. concrete

Correlations

			What do you think about the information flow in your organisation? [information flows freely]	What do you think about the information flow in your organisation? [information flow is controlled]
Kendall's tau_b	What do you think about the information flow in your organisation? [information flows freely]	Correlation Coefficient	1,000	,040
		Sig. (2-tailed)	.	,458
		N	254	250
	What do you think about the information flow in your organisation? [information flow is controlled]	Correlation Coefficient	,040	1,000
		Sig. (2-tailed)	,458	.
		N	250	252
Spearman's rho	What do you think about the information flow in your organisation? [information flows freely]	Correlation Coefficient	1,000	,041
		Sig. (2-tailed)	.	,515
		N	254	250
	What do you think about the information flow in your organisation? [information flow is controlled]	Correlation Coefficient	,041	1,000
		Sig. (2-tailed)	,515	.
		N	250	252

Table 9: Correlation of information flow - controlled vs. free

It is believed that the reason for this is the length of the survey - most of the participants likely rushed the last few difficult questions and therefore all culture-specific answers are assumed to be flawed. In retrospect, these questions would have been much better placed towards the beginning of the survey, as answering them requires some effort. Additionally, the survey could have been conducted with fewer questions to reduce the time required. Because of this reasoning, the research cannot rely upon the results of these questions regarding the organisational culture.

Cross-correlations

The research tried to identify cross-correlations between the various findings. The results are as follows:

a) A positive correlation has been found between the assignment of specialised data analyst teams and an improvement in technology (Table 10).
b) Organisations that assign the data analysis task on a case-by-case basis have a more difficult time collecting and identifying the required data for the analysis (Table 11).
c) Organisations that established transparent decision-making processes find it easier to obtain and identify the required data (Table 12).
d) Data verification, quality, and availability have a link to analysis confidence, being able to generate results in time and more often meeting the analysis requirements. Established well-known analysis processes better correlated with meeting the requirements of the analyses (Table 13).
e) Decision-making transparency is positively tied with analysis quality (Table 14).
f) Executive driven and supported Big Data implementations correlated with better analysis processes and results (Table 15).
g) Certain aspects of transparent decision-making are tied to data reliability, ability to evaluate data and reduction of time constraints (Table 16).

Correlations[b]

			In your opinion, has there been any change in the following due to the introduction of data analysis? [Technologies]
Kendall's tau_b	[There is a specialised team to conduct the analysis] Who does analyse the data?	Correlation Coefficient	,142*
		Sig. (2-tailed)	0,023
Spearman's rho	[There is a specialised team to conduct the analysis] Who does analyse the data?	Correlation Coefficient	,146*
		Sig. (2-tailed)	0,023

*. Correlation is significant at the 0.05 level (2-tailed).

b. Listwise N = 243

Table 10: Correlation between specialised data analysts and technology improvement

Correlations c

			How easy is it to identify and obtain the data? [obtaining data is...]	How easy is it to identify and obtain the data? [identifying the data that is required is...]
Kendall's tau_b	Who does analyse the data? [Someone is assigned on a case basis]	Correlation Coefficient	-,172**	-,149*
		Sig. (2-tailed)	0,004	0,012
Spearman's rho	Who does analyse the data? [Someone is assigned on a case basis]	Correlation Coefficient	-,183**	-,159*
		Sig. (2-tailed)	0,004	0,012

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

c. Listwise N = 250

Table 11: Correlation between analyst and data collection/identification

Correlations c

			How easy is it to identify and obtain the data? [obtaining data is...]	How easy is it to identify and obtain the data? [identifying the data that is required is...]
Kendall's tau_b	What would you say about the transparency of decision-making process? [The process is known amongst the decision makers]	Correlation Coefficient	,180**	,113*
		Sig. (2-tailed)	0,002	0,047
	What would you say about the transparency of decision-making process? [The process is known amongst the majority of the employees]	Correlation Coefficient	,116*	,121*
		Sig. (2-tailed)	0,036	0,029
Spearman's rho	What would you say about the transparency of decision-making process? [The process is known amongst the decision makers]	Correlation Coefficient	,204**	,129*
		Sig. (2-tailed)	0,002	0,047
	What would you say about the transparency of decision-making process? [The process is known amongst the majority of the employees]	Correlation Coefficient	,134*	,140*
		Sig. (2-tailed)	0,039	0,031

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

c. Listwise N = 238

Table 12: Correlation between decision-making transparency and process knowledge

Correlations c

What would you say about the analysis results in general? / How would you evaluate the data analysis quality in your organisation?			The required data is verified for consistency	The required data is clearly defined	The required data is available	The input data quality is sufficient	The requested output of the analysis is clearly defined	The definition of the requested output is suitable for the question asked	There exists a known process for the analysis task
Kendall's tau_b	The output does leave questions open	Correlation Coefficient	-0,014	-0,015	-0,053	0,027	0,059	0,008	-,139*
		Sig. (2-tailed)	0,819	0,807	0,379	0,656	0,324	0,892	0,019
	The output can be taken as an unquestionable statement	Correlation Coefficient	,311**	,392**	,331**	,319**	,346**	,347**	,309**
		Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	The output is produced timely	Correlation Coefficient	,238**	,287**	,334**	,294**	,305**	,351**	,328**
		Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	The analysis does answer the question	Correlation Coefficient	,217**	,295**	,235**	,208**	,262**	,303**	,230**
		Sig. (2-tailed)	0,000	0,000	0,000	0,001	0,000	0,000	0,000
Spearman's rho	The output does leave questions open	Correlation Coefficient	-0,017	-0,017	-0,058	0,029	0,066	0,009	-,157*
		Sig. (2-tailed)	0,806	0,806	0,394	0,674	0,334	0,901	0,021
	The output can be taken as an unquestionable statement	Correlation Coefficient	,357**	,443**	,376**	,361**	,384**	,386**	,347**
		Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	The output is produced timely	Correlation Coefficient	,269**	,318**	,369**	,334**	,338**	,387**	,364**
		Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	The analysis does answer the question	Correlation Coefficient	,243**	,325**	,258**	,230**	,288**	,331**	,259**
		Sig. (2-tailed)	0,000	0,000	0,000	0,001	0,000	0,000	0,000

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

c. Listwise N = 216

Table 13: Correlation between data quality and analysis quality

Correlations c

What would you say about the analysis results in general? \ What would you say about the transparency of decision making process?			The process is known amongst the decision makers	The process is known amongst the majority of the employees	The process is documented	The reasons leading to the decision are documented	The decisions itself are documented	Relevant employees are involved in the process	The process is being improved continuously
Kendall's tau_b	The output does leave questions open	Correlation Coefficient	0,119	,138*	,149*	,145*	0,073	0,091	0,074
		Sig. (2-tailed)	0,050	0,019	0,012	0,014	0,220	0,133	0,209
	The output can be taken as an unquestionable statement	Correlation Coefficient	-,139*	-,196**	-0,115	-,150*	-,145*	-,196**	-,268**
		Sig. (2-tailed)	0,022	0,001	0,053	0,011	0,015	0,001	0,000
	The output is produced timely	Correlation Coefficient	-,226**	-,228**	-,182**	-,122*	-0,081	-,192**	-,232**
		Sig. (2-tailed)	0,000	0,000	0,002	0,040	0,175	0,002	0,000
	The analysis does answer the question	Correlation Coefficient	-,134*	-,140*	-0,115	-,229**	-,144*	-,213**	-,180**
		Sig. (2-tailed)	0,029	0,019	0,057	0,000	0,017	0,001	0,003
	The required data is verified for consistency	Correlation Coefficient	-,183**	-,157**	-,256**	-,263**	-,175**	-,191**	-,224**
		Sig. (2-tailed)	0,002	0,007	0,000	0,000	0,003	0,002	0,000
	The required data is clearly defined	Correlation Coefficient	-,162**	-,170**	-,165**	-,212**	-,141*	-,164**	-,244**
		Sig. (2-tailed)	0,008	0,004	0,006	0,000	0,019	0,007	0,000
	The required data is available	Correlation Coefficient	-,281**	-,182**	-,196**	-,178**	-,169**	-,190**	-,292**
		Sig. (2-tailed)	0,000	0,002	0,001	0,003	0,005	0,002	0,000
	The input data quality is sufficient	Correlation Coefficient	-,205**	-,182**	-,183**	-,220**	-,161**	-,178**	-,272**
		Sig. (2-tailed)	0,001	0,002	0,002	0,000	0,008	0,004	0,000
	The requested output of the analysis is clearly defined	Correlation Coefficient	-,178**	-,199**	-0,108	-,186**	-,138*	-,214**	-,182**
		Sig. (2-tailed)	0,004	0,001	0,070	0,002	0,022	0,000	0,002
	The definition of the requested output is suitable for the question asked	Correlation Coefficient	-,161**	-,190**	-,152*	-,152*	-,123*	-,193**	-,159**
		Sig. (2-tailed)	0,009	0,001	0,012	0,011	0,042	0,002	0,008
	There exists a known process for the analysis task	Correlation Coefficient	-,257**	-,244**	-,246**	-,223**	-,234**	-,214**	-,265**
		Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

c. Listwise N = 210

Table 14: Correlations between decision-making transparency and analysis quality

Correlations b

		How would you evaluate the data analysis quality in your organisation?	The required data is verified for consistency	The required data is clearly defined	The required data is available	The input data quality is sufficient	The requested output of the analysis is clearly defined	The definition of the requested output is suitable for the question asked	There exists a known process for the analysis task
Kendall's tau_b	How strong would you say was the top level management support for the implementation of data analysis?	Correlation Coefficient	-0,097	-,180**	-,219**	-0,078	-,178**	-,252**	-,212**
		Sig. (2-tailed)	0,125	0,005	0,001	0,225	0,006	0,000	0,001
Spearman's rho	How strong would you say was the top level management support for the implementation of data analysis?	Correlation Coefficient	-0,106	-,195**	-,238**	-0,084	-,193**	-,272**	-,230**
		Sig. (2-tailed)	0,128	0,005	0,001	0,228	0,005	0,000	0,001

**. Correlation is significant at the 0.01 level (2-tailed).

b. Listwise N = 207

Table 15: Correlation between executive support and analysis quality

Correlations c

Decision-making is done under certain constraints, […] How did you notice a change in these factors? / What would you say about the transparency of decision-making process?			The process is known amongst the decision makers	The process is known amongst the majority of the employees	The process is documented	The reasons leading to the decision are documented	The decisions itself are documented	Relevant employees are involved in the process	The process is being improved continuously
Kendall's tau_b	time constraints	Correlation Coefficient	0,089	,164**	,147*	0,054	0,036	,144*	,187**
		Sig. (2-tailed)	0,148	0,006	0,015	0,368	0,552	0,019	0,002
	data reliability	Correlation Coefficient	,143*	,206**	,168**	0,109	,169**	,203**	,247**
		Sig. (2-tailed)	0,024	0,001	0,007	0,079	0,007	0,001	0,000
	ability to evaluate the data	Correlation Coefficient	,215**	,201**	,277**	,166**	,204**	,232**	,255**
		Sig. (2-tailed)	0,001	0,001	0,000	0,008	0,001	0,000	0,000
Spearman's rho	time constraints	Correlation Coefficient	0,101	,193**	,169*	0,061	0,041	,159*	,215**
		Sig. (2-tailed)	0,145	0,005	0,014	0,376	0,552	0,021	0,002
	data reliability	Correlation Coefficient	,156*	,231**	,188**	0,122	,186**	,222**	,274**
		Sig. (2-tailed)	0,024	0,001	0,006	0,077	0,007	0,001	0,000
	ability to evaluate the data	Correlation Coefficient	,232**	,224**	,306**	,183**	,223**	,248**	,280**
		Sig. (2-tailed)	0,001	0,001	0,000	0,008	0,001	0,000	0,000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

c. Listwise N = 211

Table 16: Correlation between decision-making transparency and constraints

4.3 Interpretation in relation to the objectives

I. **<u>Give an overview of the current practical definition, usage and distribution of Big Data</u>**

Big Data is not a novel phenomenon, although it has an exponentially growing user base. On average organisations have been exploiting Big Data for their decision-making processes since seven years ago.

Organisation size does not seem to be an indicator of whether Big Data comes into play, nor is the internationalisation of an organisation. However, it was revealed that the majority of organisations in the technology and financial sectors have already generated business models that rely on Big Data.

The most prevalent usage was seen in the last step of the decision-making process: finding and selecting a solution. The usage of Big Data in the previous process steps were each mentioned by about half of the surveyed organisations.

The practical definition of Big Data (e.g. that data volume, variety, velocity, and complexity combined makes it unfeasible to manually analyse data) scored 73% acceptance among the surveyed organisations with a significant correlation for both the technical and financial challenges that come with Big Data.

II. **<u>Identify and analyse the success factors (or success factor groups, if no distinct factors can be identified) for the implementation of Big Data across the surveyed organisations that provide a positive or negative effect on the reduction of bounded rationality by the implementation</u>**

The major findings of this research concerning the success factors were:

1) It was established that Big Data reduces bounded rationality by reducing the intuition factor in decision-making.

2) Decision-making transparency
Regarding organisational culture, decision-making transparency does not seem to correlate with the reduction of bounded rationality.
Organisations that established transparent decision-making processes found it easier to obtain and identify the required data. The analysis further showed that organisations that enacted transparent decision-making processes, decision and process documentation, employee involvement, and continuous improvement scored higher on the data-reliability, data-evaluation and time-constraint scales. Decision-making transparency is also positively tied with analysis quality, e.g. that the analysis is meeting the requirements and is produced in a timely manner. As

these attributes all provide a benefit to Big Data, it could be argued that there is an additional indirect link between decision-making transparency and the reduction of bounded rationality. No correlation was established regarding the reduction of bounded rationality when observing decision-making changes along the organisation's hierarchies.

3) Hierarchies and power
   The implementation of Big Data was perceived more positively by the employees when the power over processes and success metrics either did not change or shifted towards the lower hierarchies.

4) Continuous improvement (CI)
   Continuous verification of input data might increase data quality.

5) Data input
   Data verification, quality, and availability of the input data are linked to analysis confidence, ability of the analysis to generate results in time, and ability of the analysis to more often meet the requirements. Also, established well-known analysis processes (process transparency) further correlated with the analysis meeting the requirements. The task of identifying which data is required for analysis in order to meet the requirements is mostly perceived to be difficult.

6) Responsibilities
   Organisations that assign the data analysis task on a case-by-case basis have a more difficult time collecting and identifying the required data for the analysis and time constraints are more pressing.
   A positive correlation has been found between the assignment of specialised data analyst teams and an improvement in organisational technology.

7) Top-level support
   Executive driven and supported Big Data implementations correlated with better analysis processes and results. In addition, more executives were supportive to, or driving the, implementation than were not.

8) Bounded rationality measurement

No correlation could be established between bounded rationality measurement metrics and measurement success.

9) Organisational culture

   No analysis could be done due to the unreliability of the corresponding survey answers.

<u>Identify statistically relevant correlations and test these against causality through existing case studies</u>

The following case studies were found to support or reject the correlations identified:

1) Big Data vs. bounded rationality

   No case study was found that addressed the intuition factor in bounded rationality in relation to Big Data.

2) Decision-making transparency

   The observed correlations between the decision-making transparency and the positive factors, which seem to go hand-in-hand with the former, might be related to the feedback loop, which revolves around trust, transparency, and decision-making quality that is mentioned in Akkermans case study (2004). Although based on supply chain planning research, the validity of his findings are likely to be valid for decision-making processes. In regard of the "trust" in decision-making processes and organisational culture, this has not been a part of the research; further research with quantitative statistics is recommended in this area.

3) Hierarchies and power

   A similar example of improved perception of change was given by Moilanen (2008) in which he explains a case of negotiation in which the power over the processes was given to a lower hierarchy, whereas the power over the resources remained unchanged. However, it may be argued that the actual perception remains debatable as it might be situation-specific and bound to the organisation culture and social culture group norms.

4) Continuous improvement

   Although this research is aware that the correlation between continuous improvement and data quality does not equal causation, these two factors are highly interrelated, which allows for a cautious suggestion that causation exists. The reason for this is that the quality control will benefit the continuous improvement of the relevant processes (Shewhart and Deming, 1939).

5) Data input

The reasoning here is generally identical with that of No. 4 (continuous improvement) and No. 2 (feedback loop), as data verification is seen as part of the improvement and feedback loop. In particular, the link to analysis confidence provides a certain strength to this possible correlation. However, critical voices researched the matter of overconfidence that arises due to more data, in which the decision-making outcome was not improved (Geraldi and Arlt, 2015).

6) Responsibilities

With changing responsibilities and process assignments, the organisation is hindered in building up expert knowledge in the analysis tasks. Although evidence exists that knowledge transfer in groups by rotation is an effective and efficient means by which to distribute knowledge among the employees (Kane et al., 2005), it is, on the one hand, required that at least one of the rotation members has superior skills, and it obviously also requires that a group or team exists in the first place. As the research did not include an analysis of analyst team sizes, no statement can be made in this regard.

7) Top-level support

Jarvenpaa (1991) conducted a study in which he analysed executive support and involvement in IT management and identified, that this factor correlates with the "progressive use of IT within the firm". Even if the scope of Big Data goes way beyond the traditional information technology subject, his case study took a broad approach by defining all executive time investments towards all IT related matters as support and involvement, therefore again including Big Data. Therefore, it does not seem accidental that top-level support is highly visible throughout the surveyed organisations, as these are more likely to implement Big Data in the first place.

8) Bounded rationality measurement

To measure bounded rationality without relying on subjective interpretation, a test setup is required that can simulate decision trees in a controlled environment that excludes the probability of chance (McKinney Jr and Van Huyck, 2006). But even with this, the separation of the factors that contribute to bounded rationality might not be isolatable due to organisational structures.

9) Organisational culture

As no analysis was conducted, a comparison against case studies was omitted.

III. **<u>Produce guidance for decision makers about the success factors for implementing Big Data to reduce bounded rationality</u>**

After identification of the success factors, the following recommendations to decision-makers can be given:

a) Decision-making transparency has been proven much too influential in the success of Big Data implementation and exploitation to be ignored. Although it might sometimes be difficult to increase transparency due to the prevalent organisational culture and existing processes, a critical view on the existing situation might be beneficial. The transparency could be improved in the beginning by, for example, documenting the relevant processes and decisions and making these accessible for the wider employee base.

b) The implementation of Big Data should not be accompanied by changes that shift power over the processes to higher hierarchies, because this was universally seen as a dissatisfactory change.

c) A dedicated team should be used to conduct the analyses to create process knowledge and a rotation plan be produced in order to benefit from organisational learning, if the analyses are planned to be conducted inside the organisation.

d) Continuous improvement is the key in achieving high quality results, but it requires a working feedback loop. Therefore, it is suggested to supplement the use of Big Data with a feedback process, e.g. by asking the analysis requester and the analysis conductor to produce and share a quality assessment on the output and input data respectively. The further improvement process itself requires knowledge and consideration of the actual organisation; therefore, no suggestion can be made here.

The resemblance to some Kaizen features, as described by Imai (1993), is interesting. continuous improvement is probably the most well-known of Kaizen's modules, however Imai also claimed that organisational learning, employee involvement, and decision-making transparency are highly beneficial in his book. The reason for this might be that Kaizen and other methodologies, such as six-sigma, rely heavily on the Deming-cycle for improvement. However, the behavioural, organisational culture, and decision-making transparency focus seems to be a major feature in the Japanese approach only.

4.4 Interpretation in relation to the research aim

The aim for this research was to identify possible success factors for implementing Big Data to reduce bounded rationality in organisational decision-making. Several factors have been correlated against positive variables and found to be linked to the implementation success, as laid out before. However, no relevant correlation between the reduction of bounded rationality and these factors was found.

The results are assumed to be valid for organisations in Germany and the UK; it might be difficult to transfer the results to organisations in other culture groups and countries, as the environmental influences, work ethics, and management styles are likely to differ; as such, the research results might not match the distinct success factors.

Unfortunately, not all possible factors could be identified and evaluated. For example, the attempt to map the answers to Boisot's I-Space dimensions of organisational culture was not successful due to the survey participants failing the control questions on that particular survey section.

5. Conclusions

5.1 Conclusions about the objectives

The research has given an overview of the current practical definition of Big Data, as well as the usage and distribution in different business sectors. The findings that emerged during the analysis of the survey showed, that multiple success factors exist for the implementation of Big Data and that it is worthwhile for organisations to monitor these to ensure/improve the quality of the data analysis processes and general acceptance of the technology.

Even though bounded rationality cannot be measured objectively without experimental research methods, the measurement of bounded rationality by comparing the perception of change by employees and managers has allowed for analysis of many factors in broad scope including decision-making transparency, responsibilities, organisational culture, and processes. It is noteworthy that not all initially-assumed cross-correlations could be established with the required significance. It might be that there is no actual correlation, or just that no correlation could be established due to the unavoidably superficial survey questions that tried to cover the wide ground.

Guidance was produced for organisations, which, without the researcher's intention, resemble many attributes of the Kaizen methodology. The suggestions included increasing decision-making transparency, building process knowledge by creating a dedicated team, following the continuous improvement route by establishing feedback loops, and considering specific power changes.

5.2 Conclusions about the research aim

The research revealed evidence that Big Data in fact reduces bounded rationality, but was unable to identify relevant direct correlations between the identified success factors and the reduction of bounded rationality. However, several implementation and exploitation success factors have been identified:

- Decision-making transparency had a positive correlation with the quality of the Big Data output and organisations generally had less issues with time constraints for the analyses.
- Employees perceived the implementation of Big Data more positively when power over the processes was not shifted to higher hierarchies.
- Continuous improvement was identified to be beneficial to the data input and output quality and timeliness of the analyses.
- Organisations that assign the data analysis task on a case-by-case basis have a more difficult time collecting and identifying the required data for the analysis and time constraints are more pressing.

- Executive driven and supported Big Data implementations correlated with better analysis processes and results.

Unfortunately, the survey results for the majority of the organisational-culture aspects have been of low quality, therefore no analyses were conducted on these.

5.3 Further work

This research could be validated and extended in a number of areas:

- Survey and testing enhancements
  The chosen survey methodology still seems appropriate for the research goal. However, by reducing the scope through reduction in the survey length, e.g. through the means of isolating distinct dimensions and rethinking the survey variables, features that have shown no significant correlation in the current research may actually be found to be success factors for reducing bounded rationality. Additionally, the culture-dimension survey results were of low quality, which indicates that the survey was too long for most of the participants.
  Further, different target areas, such as other culture groups and countries could be chosen as the survey sample population to find out if there are environmental differences between these groups that affect the findings.
- Analysis enhancements
  As the researcher was previously inexperienced with enhanced statistical models, it is very likely that there are technical issues in the analysis work, regardless of the effort that was put into the statistical analysis. Verification of the results before using this research's results is therefore recommended.
- Established frameworks
  The resemblance to Kaizen and possibly other frameworks might indicate that there are similarities between the implementation and exploitation success factors of Big Data to other technologies. A follow up research might verify, if this assumption is true.
- Correlation -/- causation
  This research only identified possible correlations between the variables and factors. Further research might want to identify how any causation is established.
- Overconfidence
  Researching to what extend Big Data is tied with overconfidence in the decision-making process is suggested.

5.4 Implications of the research

Executives and managers are continuously going forward to establish a competitive advantage for their organisations. Implementing Big Data might be a new technology for these organisations, with distinct features and requirements that go beyond the usual IT systems. With this research, some success factors for implementing and handling Big Data have been identified and a guide based upon these findings has been produced that might allow organisations to improve their chances of success in adopting the new technology. In addition, further research might be conducted to verify and extend the findings. In particular, the identification of any existing frameworks may help to put the identified success factors into a generalised pattern that might be valid for other existing or upcoming technologies.

5.5 Reflection on the experience of the research process

This research was the researcher's first experience with the use of advanced statistical analysis to identify and verify correlations of this amount of variables as part of academic research. In the duration of the research project, the challenges and difficulties, that all researchers face, to identify, interpret, and consolidate existing knowledge have been especially noticeable. The most difficult part might have been defining the scope of the research aim, as in the beginning, even after a thorough literature-research phase, much uncertainty remained about the existing relevant literature and how it can aid the survey outcomes as a reference or verification method.

Therefore, the path chosen - to include as many assumed success factors as possible - might have hindered the research in multiple ways. Firstly, the literature review took most of the effort in creating a loose framework for all the dimensions that were examined in the survey. Secondly, the survey size was too big to rely on the attention span of the participants towards the end, which has resulted in bogus answers for the final category of the research. Thirdly, because of the tight time constraints, the wide scope of the research objectives took its toll in regard to the depth of the analysis, the critical discussion of the existing literature, and the analysis results.

In hindsight, a preferable approach would have been to define a very narrow scope, conduct critical analyses on the outcomes and extend the research from that point on over time. Also, more academic preparation to create and evaluate surveys, including statistics, may have helped to achieve better and more concrete results.

References

Ackoff, R.L. (1989) 'From data to wisdom: Presidential address to ISGSR, June 1988', *Journal of applied systems analysis,* vol. 16, no. 1, pp. 3-9 [Online].

Akkermans, H., Bogerd, P. and Van Doremalen, J. (2004) 'Travail, transparency and trust: A case study of computer-supported collaborative supply chain planning in high-tech electronics', *European Journal of Operational Research,* vol. 153, no. 2, pp. 445-456 [Online].

Ang, J. and Teo, T.S. (2000) 'Management issues in data warehousing: insights from the Housing and Development Board', *Decision Support Systems,* vol. 29, no. 1, pp. 11-20 [Online].

Biehl, M. (2007) 'Success factors for implementing global information systems', *Communications of the ACM,* vol. 50, no. 1, pp. 52-58 [Online].

Biggam, J. (2015) *Succeeding with your master's dissertation: a step-by-step handbook* [Online], McGraw-Hill Education (UK).

Boisot, M.H. (1998) *Knowledge assets: Securing competitive advantage in the information economy: Securing competitive advantage in the information economy* [Online], OUP Oxford.

Brynjolfsson, E., Hitt, L.M. and Kim, H.H. (2011) 'Strength in numbers: How does data-driven decisionmaking affect firm performance?', *Available at SSRN 1819486,* [Online].

Checkland, P. and Holwell, S. (1998) *Information, Systems and Information Systems*, West Sussex, Wiley.

Checkland, P. (1999a) 'Systems thinking', *Rethinking management information systems,* pp. 45-56 [Online].

Checkland, P. (1999b) 'Systems thinking, systems practice: includes a 30-year retrospective', [Online].

Checkland, P.B. (1989) 'Soft systems methodology', *Human Systems Management,* vol. 8, no. 4, pp. 273-289 [Online].

Chien, C. and Chuang, S. (2014) 'A Framework for Root Cause Detection of Sub-batch Processing System for Semiconductor Manufacturing', *IEEE Transactions for Semiconductor Manufacturing,* vol. 27, no. 4, [Online].

Companies House (2015) [Online]. Available at https://www.gov.uk/government/organisations/companies-house/about/statistics (Accessed 26.01.2016).

Cyert, R.M. and March, J.G. (1963) 'A behavioral theory of the firm', *Englewood Cliffs, NJ,* vol. 2, [Online].

Elbanna, S. (2006) 'Strategic decision-making: Process perspectives', *International Journal of Management Reviews,* vol. 8, no. 1, pp. 1-20 [Online].

Engel, U., Bartsch, S., Schnabel, C. and Vehre, H. (2012) *Wissenschaftliche Umfragen - Methoden und Fehlerquellen*, Frankfurt am Main, Campus Verlag GmbH.

Ervin, S. and Bower, R.T. (1952) 'Translation problems in international surveys', *Public opinion quarterly,* vol. 16, no. 4, pp. 595-604 [Online].

Freund, Y.P. (1988) 'Critical success factors', *Planning Review,* vol. 16, no. 4, pp. 20-23 [Online].

Geisler, S. (2012) *Einfuehrung eines Marktmonitorings zur Erkennung von aufkommenden Leistungsengpaessen [en: Introduction of a market monitoring to predict upcoming performance bottlenecks]*, Frankfurt am Main, Unpublished.

Geraldi, J. and Arlt, M. (2015) 'Confident and" wrong"? Towards a mindful use of visuals in project portfolio decisions', *IRNOP 2015 Conference* [Online]. . Available at http://orbit.dtu.dk/ws/files/118435405/Confident_and_wrong.pdf

Hill, S., Provost, F. and Volinsky, C. (2006) 'Network-based marketing: Identifying likely adopters via consumer networks', *Statistical Science,* pp. 256-276 [Online].

Imai, M. (1993) 'Kaizen: der Schlüssel zum Erfolg der Japaner im Wettbewerb', [Online].

Janis, I.L. (1971) 'Groupthink', *Psychology today,* vol. 5, no. 6, pp. 43-46 [Online].

Jarvenpaa, S.L. and Ives, B. (1991) 'Executive involvement and participation in the management of information technology', *MIS quarterly,* pp. 205-227 [Online].

Kahneman, D. (2002) 'Maps of bounded rationality: A perspective on intuitive judgment and choice', *Nobel prize lecture,* vol. 8, pp. 351-401 [Online].

Kane, A.A., Argote, L. and Levine, J.M. (2005) 'Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality', *Organizational behavior and human decision processes,* vol. 96, no. 1, pp. 56-71 [Online].

Kaner, M. and Karni, R. (2004) 'A capability maturity model for knowledge-based decisionmaking', *Information, Knowledge, Systems Management,* vol. 4, no. 4, pp. 225-252 [Online].

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y. and Pohlmann, N. (2013) 'Online controlled experiments at large scale', *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* [Online]. ACM.

Laroche, H. (1995) 'From decision to action in organizations: decision-making as a social representation', *Organization Science,* vol. 6, no. 1, pp. 62-75 [Online].

Lazer, D., Kennedy, R., King, G. and Vespignani, A. (2014) 'The parable of Google Flu: traps in big data analysis', *Science,* vol. 343, no. 14 March, [Online].

Mandinach, E.B., Honey, M. and Light, D. (2006a) 'A theoretical framework for data-driven decision making', *annual meeting of the American Educational Research Association, San Francisco, CA* [Online]. .

Mandinach, E.B., Rivas, L., Light, D., Heinze, C. and Honey, M. (2006b) 'The impact of data-driven decision making tools on educational practice: A systems analysis of six school districts', *annual meeting of the American Educational Research Association, San Francisco, CA* [Online]. .

Martens, D. and Provost, F. (2011) 'Pseudo-social network targeting from consumer transaction data', *Faculty of Applied Economics, University of Antwerp, Belgium,* [Online].

Masha, E.M. (2014) 'The Case for Data Driven Strategic Decision Making', *European Journal of Business and Management,* vol. 6, no. 29, pp. 137-146 [Online].

McKinney Jr, C.N. and Van Huyck, J.B. (2006) 'Does seeing more deeply into a game increase one's chances of winning?', *Experimental economics,* vol. 9, no. 3, pp. 297-303 [Online].

Miller, H.G. and Mork, P. (2013) 'From data to decisions: a value chain for big data', *IT Professional,* vol. 15, no. 1, pp. 57-59 [Online].

Miller, S.J., Hickson, D.J. and Wilson, D.C. (1999) 'Decision-making in organizations', *Managing organizations: Current issues,* pp. 43-62 [Online].

Mintzberg, H., Ahlstrand, B. and Lampel, J. (1998) *Strategy Safari – The complete guide through the wilds of strategic management*, Harlow, Pearson Education Limited.

Mintzberg, H., Raisinghani, D. and Theoret, A. (1976) 'The structure of" unstructured" decision processes', *Administrative Science Quarterly,* pp. 246-275 [Online].

Moilanen, S. (2008) 'The role of accounting in the management control system: a case study of a family-led firm', *Qualitative Research in Accounting & Management,* vol. 5, no. 3, pp. 165-183 [Online].

Pettigew, A. (1973) 'Decision -making as a Political Process', in *The Politics of Organisational Decision-Making*, Tavistock.

Provost, F. and Fawcett, T. (2013) 'Data science and its relationship to big data and data-driven decision making', *Big Data,* vol. 1, no. 1, pp. 51-59 [Online].

Rabl, T., Gómez-Villamor, S., Sadoghi, M., Muntés-Mulero, V., Jacobsen, H. and Mankovskii, S. (2012) 'Solving big data challenges for enterprise application performance management', *Proceedings of the VLDB Endowment,* vol. 5, no. 12, pp. 1724-1735 [Online].

Robbins, S.P. (1989) *Organizational behavior: Concepts, controversies, and applications* [Online], Prentice Hall Englewood Cliffs, NJ.

Russom, P. (2011) 'Big data analytics', *TDWI Best Practices Report, Fourth Quarter,* [Online].

Schmidt, E. and Rosenberg, J. (2014) *How Google Works*, Hachette UK.

Schumann, S. (2012) *Repräsentative Umfrage - Praxisorientierte Einführung in empirische Methoden und statistische Analyseverfahren*, 6th Edition edn. München, Oldenbourg Wissenschaftsverlag GmbH.

Shewhart, W.A. and Deming, W.E. (1939) *Statistical method from the viewpoint of quality control* [Online], Courier Corporation.

Simon, H.A. (1972) 'Theories of bounded rationality', *Decision and organization,* vol. 1, no. 1, pp. 161-176 [Online].

Simon, H.A. (1979) 'Rational decision making in business organizations', *The American Economic Review,* pp. 493-513 [Online].

Simon, H.A. (1991) 'Bounded rationality and organizational learning', *Organization science,* vol. 2, no. 1, pp. 125-134 [Online].

Statistisches Bundesamt (2013) [Online]. Available at https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/UnternehmenHandwerk/Unternehmensregister/Tabellen/UnternehmenBeschaeftigteUmsatzWZ08.html; (Accessed 26.01.2016).

Tambe, P. (2012) 'How the IT workforce affects returns to IT innovation: Evidence from Big Data analytics', [Online].

The Open University (2006) 'Ethics Principles for Research Involving Human Participants', [Online]. Available at http://www.open.ac.uk/research/main/sites/www.open.ac.uk.research.main/files/files/OU%20Ethics%20Principles%20for%20Research%20Involving%20Human%20Participant.pdf

The Open University (2013) 'Code of Practice for Research at The Open University', [Online]. Available at http://www.open.ac.uk/research/main/sites/www.open.ac.uk.research.main/files/files/ecms/web-content/CoP-amended-after-Senate-Feb-2014-Final-version-updated-Dec-2014-for-external-use-FINAL.pdf; (Accessed 20.04.2015).

Wang, F. (2012) 'A big-data perspective on ai: Newton, merton, and analytics intelligence', *Intelligent Systems, IEEE,* vol. 27, no. 5, pp. 2-4 [Online].

Weick, K.E. (1985) 'The significance of corporate culture', *Organizational culture,* pp. 381-389 [Online].

Yeoh, W. and Koronios, A. (2010) 'Critical success factors for business intelligence systems', *Journal of computer information systems,* vol. 50, no. 3, pp. 23-32 [Online].

Extended Abstract

Background

Executives and managers are continuously going forward to establish a competitive advantage for their organisation. One possible way to achieve such advantage is by utilising Big Data. As prominent companies such as Google, Amazon, IBM, Facebook and eBay are known to utilise Big Data analysis for their decision-making processes (Schmidt and Rosenberg, 2014) because the outcome of Big Data is believed to help increase the efficiency and effectiveness of the organisation in achieving its goals. It does this by reducing procurement costs, identifying and supplying previously uncharted markets, analysing customer behaviour to improve existing products, predict future customer behaviour, and aiding innovation.

However, it is not the technology that makes the decision, but the managers. They make decisions - both strategic and operational - that define the organisation in areas such as product development, market focus, project management, human resources, and production lines. Simon (1972, 1979) identified that the decision-making process is prone to “bounded rationality” because the decision-makers are limited in their processing of information due to various reasons such as human cognitive limitation, time constraints, heuristics, personal preference, uncertainty, and bias. In addition, conformity pressure phenomena in groups, also known as “groupthink” (Janis, 1971), has been extensively studied over the past decades. This all eventually has an effect on the outcome of the decision-making process, evidently leading to a sub-optimal decision.

Big Data analysis is said to reduce the above-named factors of bounded rationality and groupthink in this regard by including an additional scientific element that can be used as an input (Masha, 2014, Wang, 2012), thus improving the decision-making outcome and hence improving the organisation’s efficiency and effectiveness.

Aims and objectives

The aim for this research was to identify possible success factors for implementing Big Data to reduce bounded rationality in organisational decision-making.

The objectives were as follows:

1. give an overview of the current practical definition, usage and distribution of Big Data;

2. identify and analyse those success factors for the implementation of Big Data throughout the surveyed organisations that provide a positive or negative effect on the reduction of bounded rationality achieved by the implementation;
3. identify statistically relevant correlations and test these against causality through existing case studies; and
4. produce guidance for decision makers about the success factors for implementing Big Data to reduce bounded rationality.

Research methodology

The research was conducted in two stages. The first stage took place in Germany during May 2015 and consisted of unstructured interviews of 16 participants from 13 organisations in the size of 10 and 1500 employees. The hierarchy of these participants consisted of 5 C-level executives, 8 middle line managers and three Big Data specialists/data analysts.

The second stage was held place through an online-survey tool that picked up the previous interview findings and was open for participation from July to December 2015. In total, 294 participants completed the survey. As the tool did not register incomplete attempts, the bounce rate (participants who did not complete the survey) remains unknown. No differentiation of the 40 personal invites to the survey and the social business-network postings was possible.

The margin of error is 6.1% with a confidence level of 95% for the population group of Germany and the UK with their combined population group of 7.2 million organisations (Companies House, 2015, Statistisches Bundesamt, 2013). The data collection process is believed to be unbiased.

About half of the participants were employed in an IT-focused function. Three-quarters (76.1%) reported that their positions included management tasks. The organisations of the participants were diverse in sector, size, and internationality. Organisation size does not seem to be an indicator of whether Big Data comes into play, nor does the internationalisation of an organisation. However, it was revealed that the majority of organisations in the technology and financial sectors have already generated business models that rely on Big Data.

Big Data is not a novel phenomenon, although it has an exponential-growing user base. On average, organisations have been exploiting Big Data for their decision-making processes since seven years ago. The practical definition of Big Data (that data volume, variety, velocity, and complexity combined makes it unfeasible to manually analyse data) scored 73% acceptance among the surveyed organisations with a significant correlation for both the technical and financial challenges that come with Big Data.

Conclusions

Several factors have been correlated against positive variables and found to be linked to implementation success. However, no relevant correlation between the reduction of bounded rationality and these factors was found.

The findings that emerged during the analysis of the survey showed, that multiple success factors exist for the implementation of Big Data and that it is worthwhile for organisations to monitor these to ensure/improve the quality of the data analysis processes and general acceptance of the technology.

Even though bounded rationality cannot be measured objectively without experimental research methods, the measurement of bounded rationality by comparing the perception of change by employees and managers has allowed for analysis of many factors in broad scope including decision-making transparency, responsibilities, organisational culture, and processes. It is noteworthy that not all initially-assumed cross-correlations could be established with the required significance. It might be that there is no actual correlation or just that no correlation could be established due to the unavoidably superficial survey questions that attempted cover a wide ground.

Guidance was produced for organisations, which, without the researcher's intention, resemble many attributes of the Kaizen methodology. The suggestions included increasing decision-making transparency, building process knowledge by creating a dedicated team, following the continuous improvement route by establishing feedback loops, and considering specific power changes.

Unfortunately, not all possible factors could be identified and evaluated. For example, the attempt to map the answers to organisational culture was not successful due to the survey participants failing the control questions on that particular survey section.

The results are assumed to be valid for organisations in Germany and the UK; it might be difficult to transfer the results to organisations in other culture groups and countries, as the environmental influences, work ethics, and management styles are likely to differ; as such, the research results might not match the distinct success factors.

Word Count: 1.021

Appendix A – Survey Data

a) Participant Categorisation

Q1 – Coding #1

Your function is best described as...

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	IT focused	125	48,3	48,3	48,3
	non-IT focused	134	51,7	51,7	100,0
	Total	259	100,0	100,0	

Q2 - Coding #2

Does your position include management tasks?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	62	23,9	23,9	23,9
	Yes	197	76,1	76,1	100,0
	Total	259	100,0	100,0	

Q3 – Coding #3

How many employees work in your organisation?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	< 15	52	20,1	20,1	20,1
	15 to 50	53	20,5	20,5	40,5
	51 to 200	39	15,1	15,1	55,6
	201 to 500	30	11,6	11,6	67,2
	> 500	85	32,8	32,8	100,0
	Total	259	100,0	100,0	

Q4 – Coding #2

Does your organisation have international subsidiaries?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	171	66,0	66,0	66,0
	Yes	88	34,0	34,0	100,0
	Total	259	100,0	100,0	

Q5 – Coding #4

What business sector is your organisation operating in?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Basic Materials	4	1,5	1,5	1,5
	Consumer Goods	22	8,5	8,5	10,0
	Financial	25	9,7	9,7	19,7
	Healthcare	25	9,7	9,7	29,3
	Industrial Goods	11	4,2	4,2	33,6
	Other	49	18,9	18,9	52,5
	Services	58	22,4	22,4	74,9
	Technology	58	22,4	22,4	97,3
	Utilities	7	2,7	2,7	100,0
	Total	259	100,0	100,0	

Q6 – Coding #2

Does your organisations business model rely primarily on data usage?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	171	66,0	66,0	66,0
	Yes	88	34,0	34,0	100,0
	Total	259	100,0	100,0	

Q7

Free text field

b) Usage and definition of Big Data

Q8

Free text field

Q9 – Coding #2

[Becoming aware of a problem] Can you classify your answer in regard for what the data analysis is being used?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	138	53,3	53,3	53,3
	No	121	46,7	46,7	100,0
	Total	259	100,0	100,0	

[Diagnose a problem] Can you classify your answer in regard for what the data analysis is being used?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	119	45,9	45,9	45,9
	No	140	54,1	54,1	100,0
	Total	259	100,0	100,0	

[Finding and selecting a solution] Can you classify your answer in regard for what the data analysis is being used?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	190	73,4	73,4	73,4
	No	69	26,6	26,6	100,0
	Total	259	100,0	100,0	

Statistics

		[Becoming aware of a problem] Can you classify your answer in regard for what the data analysis is being used?	[Diagnose a problem] Can you classify your answer in regard for what the data analysis is being used?	[Finding and selecting a solution] Can you classify your answer in regard for what the data analysis is being used?
N	Valid	259	259	259
	Missing	0	0	0
Mean		1,467	1,541	1,266
Skewness		,132	-,164	1,063
Std. Error of Skewness		,151	,151	,151
Kurtosis		-1,998	-1,989	-,877
Std. Error of Kurtosis		,302	,302	,302

Q10 – Coding #6

If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [Data Size/Volume]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	89	34,4	35,2	35,2
	somewhat agree	71	27,4	28,1	63,2
	neither	34	13,1	13,4	76,7
	somewhat disagree	44	17,0	17,4	94,1
	strongly disagree	15	5,8	5,9	100,0
	Total	253	97,7	100,0	
Missing	Unknown	6	2,3		
Total		259	100,0		

If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [Data Complexity]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	61	23,6	24,1	24,1
	somewhat agree	89	34,4	35,2	59,3
	neither	37	14,3	14,6	73,9
	somewhat disagree	46	17,8	18,2	92,1
	strongly disagree	20	7,7	7,9	100,0
	Total	253	97,7	100,0	
Missing	Unknown	6	2,3		
Total		259	100,0		

If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [Data Change Rate]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	55	21,2	22,8	22,8
	somewhat agree	90	34,7	37,3	60,2
	neither	51	19,7	21,2	81,3
	somewhat disagree	36	13,9	14,9	96,3
	strongly disagree	9	3,5	3,7	100,0
	Total	241	93,1	100,0	
Missing	Unknown	18	6,9		
Total		259	100,0		

If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [Data Variety (e.g. data from different sources are not uniform)]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	73	28,2	29,4	29,4
	somewhat agree	89	34,4	35,9	65,3
	neither	40	15,4	16,1	81,5
	somewhat disagree	33	12,7	13,3	94,8
	strongly disagree	13	5,0	5,2	100,0
	Total	248	95,8	100,0	
Missing	Unknown	11	4,2		
Total		259	100,0		

If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [General understanding of statistics]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	55	21,2	21,7	21,7
	somewhat agree	76	29,3	30,0	51,8
	neither	48	18,5	19,0	70,8
	somewhat disagree	55	21,2	21,7	92,5
	strongly disagree	19	7,3	7,5	100,0
	Total	253	97,7	100,0	
Missing	Unknown	6	2,3		
Total		259	100,0		

Statistics

		If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [Data Size/Volume]	If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [Data Complexity]	If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [Data Change Rate]	If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [Data Variety (e.g. data from different sources are not uniform)]	If your organisation had to manually analyse the data in an excel sheet, what of the following would be an issue? [General understanding of statistics]
N	Valid	253	253	241	248	253
	Missing	6	6	18	11	6
Mean		2,308	2,506	2,394	2,290	2,632
Skewness		,608	,502	,513	,701	,281
Std. Error of Skewness		,153	,153	,157	,155	,153
Kurtosis		-,860	-,869	-,550	-,435	-1,041
Std. Error of Kurtosis		,305	,305	,312	,308	,305

Q11 – Coding #7

How easy is it to identify and obtain the data? [obtaining data is...]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very easy	52	20,1	20,6	20,6
	somewhat easy	127	49,0	50,2	70,8
	somewhat hard	61	23,6	24,1	94,9
	very hard	13	5,0	5,1	100,0
	Total	253	97,7	100,0	
Missing	Unknown	6	2,3		
Total		259	100,0		

How easy is it to identify and obtain the data? [identifying the data that is required is...]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very easy	33	12,7	13,2	13,2
	somewhat easy	105	40,5	42,0	55,2
	somewhat hard	86	33,2	34,4	89,6
	very hard	26	10,0	10,4	100,0
	Total	250	96,5	100,0	
Missing	Unknown	9	3,5		
Total		259	100,0		

Statistics

		How easy is it to identify and obtain the data? [obtaining data is...]	How easy is it to identify and obtain the data? [identifying the data that is required is...]
N	Valid	253	250
	Missing	6	9
Mean		2,138	2,420
Skewness		,362	,113
Std. Error of Skewness		,153	,154
Kurtosis		-,248	-,572
Std. Error of Kurtosis		,305	,307

Q12 – Coding #8

How is often is data collected in your organisation? [continuously/real-time]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	never	25	9,7	9,8	9,8
	sometimes	61	23,6	23,9	33,7
	often	58	22,4	22,7	56,5
	always	111	42,9	43,5	100,0
	Total	255	98,5	100,0	
Missing	Unknown	4	1,5		
Total		259	100,0		

How is often is data collected in your organisation? [periodically]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	never	11	4,2	4,3	4,3
	sometimes	69	26,6	27,2	31,5
	often	99	38,2	39,0	70,5
	always	75	29,0	29,5	100,0
	Total	254	98,1	100,0	
Missing	Unknown	5	1,9		
Total		259	100,0		

How is often is data collected in your organisation? [on a case basis]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	never	12	4,6	4,8	4,8
	sometimes	97	37,5	39,0	43,8
	often	65	25,1	26,1	69,9
	always	75	29,0	30,1	100,0
	Total	249	96,1	100,0	
Missing	Unknown	10	3,9		
Total		259	100,0		

Statistics

		How is often is data collected in your organisation? [continuously/real-time]	How is often is data collected in your organisation? [periodically]	How is often is data collected in your organisation? [on a case basis]
N	Valid	255	254	249
	Missing	4	5	10
Mean		3,000	2,937	2,815
Skewness		-,537	-,292	,004
Std. Error of Skewness		,153	,153	,154
Kurtosis		-1,025	-,785	-1,214
Std. Error of Kurtosis		,304	,304	,307

Q13 – Coding #9

Statistics

		[There is a specialised team to conduct the analysis] Who does analyse the data?	[Someone is assigned on a case basis] Who does analyse the data?	[The analysis is done by a service provider] Who does analyse the data?	[The team who requires the data analysis] Who does analyse the data?
N	Valid	259	259	259	259
	Missing	0	0	0	0
Mean		1,591	1,595	1,849	1,494
Skewness		-,371	-,388	-1,965	,023
Std. Error of Skewness		,151	,151	,151	,151
Kurtosis		-1,877	-1,864	1,877	-2,015
Std. Error of Kurtosis		,302	,302	,302	,302

Q14

Free text field.

c) Analysis Quality

Q15 – Coding #8

Statistics

		How is often is the collected data analysed in your organisation? [continuously/real-time]	How is often is the collected data analysed in your organisation? [periodically]	How is often is the collected data analysed in your organisation? [on a case basis]
N	Valid	255	253	249
	Missing	4	6	10
Mean		2,851	2,893	2,815
Skewness		-,340	-,341	,017
Std. Error of Skewness		,153	,153	,154
Kurtosis		-1,094	-,560	-1,108
Std. Error of Kurtosis		,304	,305	,307

Q16 – Coding #6

Statistics

		How would you evaluate the data analysis quality in your organisation? [The required data is verified for consistency]	How would you evaluate the data analysis quality in your organisation? [The required data is clearly defined]	How would you evaluate the data analysis quality in your organisation? [The required data is available]	How would you evaluate the data analysis quality in your organisation? [The input data quality is sufficient]	How would you evaluate the data analysis quality in your organisation? [The requested output of the analysis is clearly defined]	How would you evaluate the data analysis quality in your organisation? [The definition of the requested output is suitable for the question asked]	How would you evaluate the data analysis quality in your organisation? [There exists a known process for the analysis task]
N	Valid	248	250	255	252	245	245	249
	Missing	11	9	4	7	14	14	10
Mean		2,931	3,124	3,122	3,016	3,065	3,057	3,004
Skewness		-,387	-,498	-,408	-,463	-,311	-,322	-,399
Std. Error of Skewness		,155	,154	,153	,153	,156	,156	,154
Kurtosis		-,895	-,653	-,828	-,260	-,884	-,591	-,812
Std. Error of Kurtosis		,308	,307	,304	,306	,310	,310	,307

Q17 – Coding #6

Statistics

		What would you say about the analysis results in general? [The output does leave questions open]	What would you say about the analysis results in general? [The output can be taken as an unquestionable statement]	What would you say about the analysis results in general? [The output is produced timely]	What would you say about the analysis results in general? [The analysis does answer the question]
N	Valid	253	245	250	254
	Missing	6	14	9	5
Mean		2,553	2,580	2,984	2,961
Skewness		,213	,066	-,552	-,304
Std. Error of Skewness		,153	,156	,154	,153
Kurtosis		-,739	-,598	-,198	-,190
Std. Error of Kurtosis		,305	,310	,307	,304

Q18 – Coding #6

Statistics

		How would you rate the difficulty of the data analysis from your point of view? [Processing the data is technically difficult]	How would you rate the difficulty of the data analysis from your point of view? [Manual analysis of the data is cost prohibitive]	How would you rate the difficulty of the data analysis from your point of view? [Applying the statistical tools required for the task is easy]
N	Valid	256	242	249
	Missing	3	17	10
Mean		2,969	2,711	2,763
Skewness		,193	,303	,219
Std. Error of Skewness		,152	,156	,154
Kurtosis		-,977	-,866	-,846
Std. Error of Kurtosis		,303	,312	,307

Q19

Free text field.

Q20

Statistics

In general, for how long (years) has data analysis been part of decision-making processes in your organisation?

N	Valid	256
	Missing	3
Mean		8,570
Skewness		,879
Std. Error of Skewness		,152
Kurtosis		-,056
Std. Error of Kurtosis		,303

Q21 – Coding #10

Statistics

Does your position include being an active part in decision-making processes?

N	Valid	259
	Missing	0
Mean		1,799
Skewness		,359
Std. Error of Skewness		,151
Kurtosis		-1,217
Std. Error of Kurtosis		,302

Does your position include being an active part in decision-making processes?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	107	41,3	41,3	41,3
	No, but Influence	97	37,5	37,5	78,8
	No	55	21,2	21,2	100,0
	Total	259	100,0	100,0	

Does your position include being an active part in decision-making processes? * Does your position include management tasks? Crosstabulation

Count

		Does your position include management tasks?		
		No	Yes	Total
Does your position include being an active part in decision-making processes?	Yes	9	98	107
	No, but Influence	29	68	97
	No	24	31	55
Total		62	197	259

Q22 – Coding 6

Statistics

		What would you say about the transparency of decision-making process? [The process is known amongst the decision makers]	What would you say about the transparency of decision-making process? [The process is known amongst the majority of the employees]	What would you say about the transparency of decision-making process? [The process is documented]	What would you say about the transparency of decision-making process? [The reasons leading to the decision are documented]	What would you say about the transparency of decision-making process? [The decisions itself are documented]	What would you say about the transparency of decision-making process? [Relevant employees are involved in the process]	What would you say about the transparency of decision-making process? [The process is being improved continuously]
N	Valid	254	251	249	251	250	254	254
	Missing	5	8	10	8	9	5	5
Mean		1,969	2,677	2,197	2,327	2,140	1,996	2,394
Skewness		1,162	,474	,843	,803	,932	1,160	,611
Std. Error of Skewness		,153	,154	,154	,154	,154	,153	,153
Kurtosis		1,047	-,811	-,288	-,262	,108	,731	-,428
Std. Error of Kurtosis		,304	,306	,307	,306	,307	,304	,304

Low mean = agree; This will be another DV for further correlation tests.

What would you say about the transparency of decision-making process? [The process is known amongst the decision makers]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	89	34,4	35,0	35,0
	somewhat agree	116	44,8	45,7	80,7
	neither	23	8,9	9,1	89,8
	somewhat disagree	20	7,7	7,9	97,6
	strongly disagree	6	2,3	2,4	100,0
	Total	254	98,1	100,0	
Missing	Unknown	5	1,9		
Total		259	100,0		

What would you say about the transparency of decision-making process? [The process is known amongst the majority of the employees]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	34	13,1	13,5	13,5
	somewhat agree	106	40,9	42,2	55,8
	neither	40	15,4	15,9	71,7
	somewhat disagree	49	18,9	19,5	91,2
	strongly disagree	22	8,5	8,8	100,0
	Total	251	96,9	100,0	
Missing	Unknown	8	3,1		
Total		259	100,0		

What would you say about the transparency of decision-making process? [The decisions itself are documented]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	80	30,9	32,0	32,0
	somewhat agree	103	39,8	41,2	73,2
	neither	29	11,2	11,6	84,8
	somewhat disagree	28	10,8	11,2	96,0
	strongly disagree	10	3,9	4,0	100,0
	Total	250	96,5	100,0	
Missing	Unknown	9	3,5		
Total		259	100,0		

What would you say about the transparency of decision-making process? [Relevant employees are involved in the process]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	93	35,9	36,6	36,6
	somewhat agree	110	42,5	43,3	79,9
	neither	18	6,9	7,1	87,0
	somewhat disagree	25	9,7	9,8	96,9
	strongly disagree	8	3,1	3,1	100,0
	Total	254	98,1	100,0	
Missing	Unknown	5	1,9		
Total		259	100,0		

What would you say about the transparency of decision-making process? [The process is being improved continuously]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	58	22,4	22,8	22,8
	somewhat agree	99	38,2	39,0	61,8
	neither	49	18,9	19,3	81,1
	somewhat disagree	35	13,5	13,8	94,9
	strongly disagree	13	5,0	5,1	100,0
	Total	254	98,1	100,0	
Missing	Unknown	5	1,9		
Total		259	100,0		

Q23 – Coding #15

Statistics

		Please rate the influence of the following to the decision-making process: [intuition]	Please rate the influence of the following to the decision-making process: [data]	Please rate the influence of the following to the decision-making process: [personal experience]	Please rate the influence of the following to the decision-making process: [external consultants]	Please rate the influence of the following to the decision-making process: [creativity]	Please rate the influence of the following to the decision-making process: [logic]	Please rate the influence of the following to the decision-making process: [reasoning]	Please rate the influence of the following to the decision-making process: [skills]	Please rate the influence of the following to the decision-making process: [knowledge from previous decisions]
N	Valid	248	253	254	242	247	250	252	252	250
	Missing	11	6	5	17	12	9	7	7	9
Mean		1,911	1,332	1,516	2,083	1,968	1,392	1,496	1,421	1,408
Skewness		,158	1,420	,937	-,142	,046	1,374	,983	1,221	1,165
Std. Error of Skewness		,155	,153	,153	,156	,155	,154	,153	,153	,154
Kurtosis		-1,357	1,077	-,293	-1,291	-1,000	,711	-,183	,373	,343
Std. Error of Kurtosis		,308	,305	,304	,312	,309	,307	,306	,306	,307

Please rate the influence of the following to the decision-making process: [intuition]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	88	34,0	35,5	35,5
	weak influence	94	36,3	37,9	73,4
	no influence	66	25,5	26,6	100,0
	Total	248	95,8	100,0	
Missing	Unknown	11	4,2		
Total		259	100,0		

Please rate the influence of the following to the decision-making process: [data]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	179	69,1	70,8	70,8
	weak influence	64	24,7	25,3	96,0
	no influence	10	3,9	4,0	100,0
	Total	253	97,7	100,0	
Missing	Unknown	6	2,3		
Total		259	100,0		

Please rate the influence of the following to the decision-making process: [personal experience]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	148	57,1	58,3	58,3
	weak influence	81	31,3	31,9	90,2
	no influence	25	9,7	9,8	100,0
	Total	254	98,1	100,0	
Missing	Unknown	5	1,9		
Total		259	100,0		

Please rate the influence of the following to the decision-making process: [external consultants]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	62	23,9	25,6	25,6
	weak influence	98	37,8	40,5	66,1
	no influence	82	31,7	33,9	100,0
	Total	242	93,4	100,0	
Missing	Unknown	17	6,6		
Total		259	100,0		

Please rate the influence of the following to the decision-making process: [creativity]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	66	25,5	26,7	26,7
	weak influence	123	47,5	49,8	76,5
	no influence	58	22,4	23,5	100,0
	Total	247	95,4	100,0	
Missing	Unknown	12	4,6		
Total		259	100,0		

Please rate the influence of the following to the decision-making process: [logic]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	172	66,4	68,8	68,8
	weak influence	58	22,4	23,2	92,0
	no influence	20	7,7	8,0	100,0
	Total	250	96,5	100,0	
Missing	Unknown	9	3,5		
Total		259	100,0		

Please rate the influence of the following to the decision-making process: [reasoning]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	150	57,9	59,5	59,5
	weak influence	79	30,5	31,3	90,9
	no influence	23	8,9	9,1	100,0
	Total	252	97,3	100,0	
Missing	Unknown	7	2,7		
Total		259	100,0		

Please rate the influence of the following to the decision-making process: [skills]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	165	63,7	65,5	65,5
	weak influence	68	26,3	27,0	92,5
	no influence	19	7,3	7,5	100,0
	Total	252	97,3	100,0	
Missing	Unknown	7	2,7		
Total		259	100,0		

Please rate the influence of the following to the decision-making process: [knowledge from previous decisions]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	162	62,5	64,8	64,8
	weak influence	74	28,6	29,6	94,4
	no influence	14	5,4	5,6	100,0
	Total	250	96,5	100,0	
Missing	Unknown	9	3,5		
Total		259	100,0		

Q24

Free text field

Q25 – Coding #12

Statistics

How strong would you say was the top-level management support for the implementation of data analysis? [it was...]

N	Valid	233
	Missing	26
Mean		1,627
Skewness		,205
Std. Error of Skewness		,159
Kurtosis		-,767
Std. Error of Kurtosis		,318

How strong would you say was the top-level management support for the implementation of data analysis? [it was...]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	driven by top level management	97	37,5	41,6	41,6
	supportive	126	48,6	54,1	95,7
	rejective	10	3,9	4,3	100,0
	Total	233	90,0	100,0	
Missing	Unknown	26	10,0		
Total		259	100,0		

Q26 – Coding #13

Statistics

		In your opinion, has there been any change in the following due to the introduction of data analysis? [Responsibilities]	In your opinion, has there been any change in the following due to the introduction of data analysis? [Resources]	In your opinion, has there been any change in the following due to the introduction of data analysis? [Decision Making processes]	In your opinion, has there been any change in the following due to the introduction of data analysis? [Technologies]	In your opinion, has there been any change in the following due to the introduction of data analysis? [Distribution of decision-making power in the organisation]	In your opinion, has there been any change in the following due to the introduction of data analysis? [Measurement of metrics]	In your opinion, has there been any change in the following due to the introduction of data analysis? [Processes other than Decision Making]	In your opinion, has there been any change in the following due to the introduction of data analysis? [Skills and Knowledge]
N	Valid	244	244	244	243	240	226	225	240
	Missing	15	15	15	16	19	33	34	19
Mean		1,545	1,611	1,561	1,444	1,829	1,642	1,698	1,442
Skewness		,926	,740	,864	1,090	,245	,588	,494	1,199
Std. Error of Skewness		,156	,156	,156	,156	,157	,162	,162	,157
Kurtosis		-,483	-,740	-,543	,121	-,927	-,727	-,862	,224
Std. Error of Kurtosis		,310	,310	,310	,311	,313	,322	,323	,313

In your opinion, has there been any change in the following due to the introduction of data analysis? [Responsibilities]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes, to the better	143	55,2	58,6	58,6
	No change	69	26,6	28,3	86,9
	Yes, to the worse	32	12,4	13,1	100,0
	Total	244	94,2	100,0	
Missing	Unknown	15	5,8		
Total		259	100,0		

In your opinion, has there been any change in the following due to the introduction of data analysis? [Resources]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes, to the better	129	49,8	52,9	52,9
	No change	81	31,3	33,2	86,1
	Yes, to the worse	34	13,1	13,9	100,0
	Total	244	94,2	100,0	
Missing	Unknown	15	5,8		
Total		259	100,0		

In your opinion, has there been any change in the following due to the introduction of data analysis? [Decision Making processes]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes, to the better	138	53,3	56,6	56,6
	No change	75	29,0	30,7	87,3
	Yes, to the worse	31	12,0	12,7	100,0
	Total	244	94,2	100,0	
Missing	Unknown	15	5,8		
Total		259	100,0		

In your opinion, has there been any change in the following due to the introduction of data analysis? [Technologies]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes, to the better	152	58,7	62,6	62,6
	No change	74	28,6	30,5	93,0
	Yes, to the worse	17	6,6	7,0	100,0
	Total	243	93,8	100,0	
Missing	Unknown	16	6,2		
Total		259	100,0		

In your opinion, has there been any change in the following due to the introduction of data analysis? [Distribution of decision-making power in the organisation]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes, to the better	82	31,7	34,2	34,2
	No change	117	45,2	48,8	82,9
	Yes, to the worse	41	15,8	17,1	100,0
	Total	240	92,7	100,0	
Missing	Unknown	19	7,3		
Total		259	100,0		

In your opinion, has there been any change in the following due to the introduction of data analysis? [Measurement of metrics]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes, to the better	107	41,3	47,3	47,3
	No change	93	35,9	41,2	88,5
	Yes, to the worse	26	10,0	11,5	100,0
	Total	226	87,3	100,0	
Missing	Unknown	33	12,7		
Total		259	100,0		

In your opinion, has there been any change in the following due to the introduction of data analysis? [Processes other than Decision Making]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes, to the better	99	38,2	44,0	44,0
	No change	95	36,7	42,2	86,2
	Yes, to the worse	31	12,0	13,8	100,0
	Total	225	86,9	100,0	
Missing	Unknown	34	13,1		
Total		259	100,0		

In your opinion, has there been any change in the following due to the introduction of data analysis? [Skills and Knowledge]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes, to the better	156	60,2	65,0	65,0
	No change	62	23,9	25,8	90,8
	Yes, to the worse	22	8,5	9,2	100,0
	Total	240	92,7	100,0	
Missing	Unknown	19	7,3		
Total		259	100,0		

Q27 – Coding #14

Statistics

		How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Responsibilities]	How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Resources]	How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Decision Making power]	How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Power over the processes]	How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Setting success metrics]	How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Checking success against metrics]	How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Skills and Knowledge]
N	Valid	244	237	243	237	225	231	234
	Missing	15	22	16	22	34	28	25
Mean		1,877	1,979	2,259	2,173	2,231	2,130	1,880
Skewness		,232	,036	-,488	-,316	-,412	-,227	,213
Std. Error of Skewness		,156	,158	,156	,158	,162	,160	,159
Kurtosis		-1,485	-1,295	-1,191	-1,316	-1,146	-1,276	-1,335
Std. Error of Kurtosis		,310	,315	,311	,315	,323	,319	,317

How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Responsibilities]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	shifted towards lower hierarchy	99	38,2	40,6	40,6
	no change	76	29,3	31,1	71,7
	shifted towards higher hierarchy	69	26,6	28,3	100,0
	Total	244	94,2	100,0	
Missing	Unknown	15	5,8		
Total		259	100,0		

How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Resources]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	shifted towards lower hierarchy	72	27,8	30,4	30,4
	no change	98	37,8	41,4	71,7
	shifted towards higher hierarchy	67	25,9	28,3	100,0
	Total	237	91,5	100,0	
Missing	Unknown	22	8,5		
Total		259	100,0		

How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Decision Making power]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	shifted towards lower hierarchy	50	19,3	20,6	20,6
	no change	80	30,9	32,9	53,5
	shifted towards higher hierarchy	113	43,6	46,5	100,0
	Total	243	93,8	100,0	
Missing	Unknown	16	6,2		
Total		259	100,0		

How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Power over the processes]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	shifted towards lower hierarchy	56	21,6	23,6	23,6
	no change	84	32,4	35,4	59,1
	shifted towards higher hierarchy	97	37,5	40,9	100,0
	Total	237	91,5	100,0	
Missing	Unknown	22	8,5		
Total		259	100,0		

How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Setting success metrics]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	shifted towards lower hierarchy	44	17,0	19,6	19,6
	no change	85	32,8	37,8	57,3
	shifted towards higher hierarchy	96	37,1	42,7	100,0
	Total	225	86,9	100,0	
Missing	Unknown	34	13,1		
Total		259	100,0		

How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Checking success against metrics]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	shifted towards lower hierarchy	55	21,2	23,8	23,8
	no change	91	35,1	39,4	63,2
	shifted towards higher hierarchy	85	32,8	36,8	100,0
	Total	231	89,2	100,0	
Missing	Unknown	28	10,8		
Total		259	100,0		

How would you say has this affected the distribution of the following amongst the organisations hierarchies? [Skills and Knowledge]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	shifted towards lower hierarchy	87	33,6	37,2	37,2
	no change	88	34,0	37,6	74,8
	shifted towards higher hierarchy	59	22,8	25,2	100,0
	Total	234	90,3	100,0	
Missing	Unknown	25	9,7		
Total		259	100,0		

Q28

Free text field.

d) Changes in Decision making

Q29 – Coding #15

Statistics

		If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [intuition]	If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [data]	If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [personal experience]	If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [external consultants]	If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [creativity]	If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [logic]	If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [reasoning]	If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [knowledge from previous decisions]	If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [skills]
N	Valid	228	224	230	226	231	234	229	234	235
	Missing	31	35	29	33	28	25	30	25	24

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [intuition]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	98	37,8	43,0	43,0
	weak influence	86	33,2	37,7	80,7
	no influence	44	17,0	19,3	100,0
	Total	228	88,0	100,0	
Missing	Unknown	31	12,0		
Total		259	100,0		

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [data]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	107	41,3	47,8	47,8
	weak influence	84	32,4	37,5	85,3
	no influence	33	12,7	14,7	100,0
	Total	224	86,5	100,0	
Missing	Unknown	35	13,5		
Total		259	100,0		

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [personal experience]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	161	62,2	70,0	70,0
	weak influence	52	20,1	22,6	92,6
	no influence	17	6,6	7,4	100,0
	Total	230	88,8	100,0	
Missing	Unknown	29	11,2		
Total		259	100,0		

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [external consultants]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	59	22,8	26,1	26,1
	weak influence	83	32,0	36,7	62,8
	no influence	84	32,4	37,2	100,0
	Total	226	87,3	100,0	
Missing	Unknown	33	12,7		
Total		259	100,0		

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [creativity]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	87	33,6	37,7	37,7
	weak influence	90	34,7	39,0	76,6
	no influence	54	20,8	23,4	100,0
	Total	231	89,2	100,0	
Missing	Unknown	28	10,8		
Total		259	100,0		

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [logic]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	148	57,1	63,2	63,2
	weak influence	64	24,7	27,4	90,6
	no influence	22	8,5	9,4	100,0
	Total	234	90,3	100,0	
Missing	Unknown	25	9,7		
Total		259	100,0		

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [reasoning]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	147	56,8	64,2	64,2
	weak influence	60	23,2	26,2	90,4
	no influence	22	8,5	9,6	100,0
	Total	229	88,4	100,0	
Missing	Unknown	30	11,6		
Total		259	100,0		

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [knowledge from previous decisions]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	164	63,3	70,1	70,1
	weak influence	49	18,9	20,9	91,0
	no influence	21	8,1	9,0	100,0
	Total	234	90,3	100,0	
Missing	Unknown	25	9,7		
Total		259	100,0		

If you think about the past, before data analysis was a part of the decision-making process, how would you say was the influence of the following: [skills]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strong influence	143	55,2	60,9	60,9
	weak influence	68	26,3	28,9	89,8
	no influence	24	9,3	10,2	100,0
	Total	235	90,7	100,0	
Missing	Unknown	24	9,3		
Total		259	100,0		

Q30 – Coding 16

Statistics

		Has there been an attempt to measure the change in influence when data analysis was introduced? [intuition]	Has there been an attempt to measure the change in influence when data analysis was introduced? [data]	Has there been an attempt to measure the change in influence when data analysis was introduced? [experience]	Has there been an attempt to measure the change in influence when data analysis was introduced? [external consultants]	Has there been an attempt to measure the change in influence when data analysis was introduced? [creativity]	Has there been an attempt to measure the change in influence when data analysis was introduced? [logic]	Has there been an attempt to measure the change in influence when data analysis was introduced? [reasoning]	Has there been an attempt to measure the change in influence when data analysis was introduced? [knowledge from previous decisions]	Has there been an attempt to measure the change in influence when data analysis was introduced? [skills]
N	Valid	211	211	207	204	214	214	206	209	212
	Missing	48	48	52	55	45	45	53	50	47
Mean		2,351	1,934	2,097	2,343	2,341	2,154	2,180	2,115	2,142
Skewness		-,669	,132	-,184	-,703	-,640	-,295	-,335	-,215	-,275
Std. Error of Skewness		,167	,167	,169	,170	,166	,166	,169	,168	,167
Kurtosis		-,903	-1,782	-1,548	-1,105	-,911	-1,480	-1,351	-1,475	-1,558
Std. Error of Kurtosis		,333	,333	,337	,339	,331	,331	,337	,335	,333

High mean = no measurements;

According to the participants, the least measured influencing factors are Intuition, External Consultants and Creativity.

Has there been an attempt to measure the change in influence when data analysis was introduced? [intuition]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	34	13,1	16,1	16,1
	yes, but without metrics	69	26,6	32,7	48,8
	no	108	41,7	51,2	100,0
	Total	211	81,5	100,0	
Missing	Unknown	48	18,5		
Total		259	100,0		

Has there been an attempt to measure the change in influence when data analysis was introduced? [data]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	94	36,3	44,5	44,5
	yes, but without metrics	37	14,3	17,5	62,1
	no	80	30,9	37,9	100,0
	Total	211	81,5	100,0	
Missing	Unknown	48	18,5		
Total		259	100,0		

Has there been an attempt to measure the change in influence when data analysis was introduced? [experience]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	63	24,3	30,4	30,4
	yes, but without metrics	61	23,6	29,5	59,9
	no	83	32,0	40,1	100,0
	Total	207	79,9	100,0	
Missing	Unknown	52	20,1		
Total		259	100,0		

Has there been an attempt to measure the change in influence when data analysis was introduced? [external consultants]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	43	16,6	21,1	21,1
	yes, but without metrics	48	18,5	23,5	44,6
	no	113	43,6	55,4	100,0
	Total	204	78,8	100,0	
Missing	Unknown	55	21,2		
Total		259	100,0		

Has there been an attempt to measure the change in influence when data analysis was introduced? [creativity]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	34	13,1	15,9	15,9
	yes, but without metrics	73	28,2	34,1	50,0
	no	107	41,3	50,0	100,0
	Total	214	82,6	100,0	
Missing	Unknown	45	17,4		
Total		259	100,0		

Has there been an attempt to measure the change in influence when data analysis was introduced? [logic]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	59	22,8	27,6	27,6
	yes, but without metrics	63	24,3	29,4	57,0
	no	92	35,5	43,0	100,0
	Total	214	82,6	100,0	
Missing	Unknown	45	17,4		
Total		259	100,0		

Has there been an attempt to measure the change in influence when data analysis was introduced? [reasoning]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	50	19,3	24,3	24,3
	yes, but without metrics	69	26,6	33,5	57,8
	no	87	33,6	42,2	100,0
	Total	206	79,5	100,0	
Missing	Unknown	53	20,5		
Total		259	100,0		

Has there been an attempt to measure the change in influence when data analysis was introduced? [knowledge from previous decisions]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	59	22,8	28,2	28,2
	yes, but without metrics	67	25,9	32,1	60,3
	no	83	32,0	39,7	100,0
	Total	209	80,7	100,0	
Missing	Unknown	50	19,3		
Total		259	100,0		

Has there been an attempt to measure the change in influence when data analysis was introduced? [skills]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, with formal metrics	63	24,3	29,7	29,7
	yes, but without metrics	56	21,6	26,4	56,1
	no	93	35,9	43,9	100,0
	Total	212	81,9	100,0	
Missing	Unknown	47	18,1		
Total		259	100,0		

Q31 – Coding #17

Statistics

Overall, how successful would you say was the attempt to measure the change? [the measurement was]

N	Valid	175
	Missing	84
Mean		2,251
Skewness		-,274
Std. Error of Skewness		,184
Kurtosis		-,663
Std. Error of Kurtosis		,365

Overall, how successful would you say was the attempt to measure the change? [the measurement was]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	not successful	19	7,3	10,9	10,9
	partially successful	93	35,9	53,1	64,0
	successful	63	24,3	36,0	100,0
	Total	175	67,6	100,0	
Missing	Unknown	84	32,4		
Total		259	100,0		

Q32

Free text field

Q33 – Coding #18

Statistics

		Decision-making is done under certain constraints, e.g. time might be pressing, or the available data could be unreliable. Also, it might be that there is too much data to consider. How did you notice a change in these factors? [time constraints]	Decision-making is done under certain constraints, e.g. time might be pressing, or the available data could be unreliable. Also, it might be that there is too much data to consider. How did you notice a change in these factors? [data reliability]	Decision-making is done under certain constraints, e.g. time might be pressing, or the available data could be unreliable. Also, it might be that there is too much data to consider. How did you notice a change in these factors? [ability to evaluate the data]
N	Valid	226	222	222
	Missing	33	37	37
Mean		1,942	1,410	1,432
Skewness		,088	1,143	,966
Std. Error of Skewness		,162	,163	,163
Kurtosis		-1,084	,302	-,052
Std. Error of Kurtosis		,322	,325	,325

Decision-making is done under certain constraints, e.g. time might be pressing, or the available data could be unreliable. Also, it might be that there is too much data to consider. How did you notice a change in these factors? [time constraints]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	got better	66	25,5	29,2	29,2
	did not change	107	41,3	47,3	76,5
	got worse	53	20,5	23,5	100,0
	Total	226	87,3	100,0	
Missing	Unknown	32	12,4		
	System	1	,4		
	Total	33	12,7		
Total		259	100,0		

Decision-making is done under certain constraints, e.g. time might be pressing, or the available data could be unreliable. Also, it might be that there is too much data to consider. How did you notice a change in these factors? [data reliability]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	got better	143	55,2	64,4	64,4
	did not change	67	25,9	30,2	94,6
	got worse	12	4,6	5,4	100,0
	Total	222	85,7	100,0	
Missing	Unknown	36	13,9		
	System	1	,4		
	Total	37	14,3		
Total		259	100,0		

Decision-making is done under certain constraints, e.g. time might be pressing, or the available data could be unreliable. Also, it might be that there is too much data to consider. How did you notice a change in these factors? [ability to evaluate the data]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	got better	136	52,5	61,3	61,3
	did not change	76	29,3	34,2	95,5
	got worse	10	3,9	4,5	100,0
	Total	222	85,7	100,0	
Missing	Unknown	36	13,9		
	System	1	,4		
	Total	37	14,3		
Total		259	100,0		

Q34

Free text field.

e) Organisational Culture

Q35 – Coding #6

Statistics

		What do you think about the information flow in your organisation? [information flows freely]	What do you think about the information flow in your organisation? [information is abstract]	What do you think about the information flow in your organisation? [information is concrete]	What do you think about the information flow in your organisation? [information flow is controlled]	What do you think about the information flow in your organisation? [information is written down]
N	Valid	254	248	256	252	255
	Missing	5	11	3	7	4
Mean		2,583	2,992	2,254	2,115	2,078
Skewness		,553	,112	,748	,907	,925
Std. Error of Skewness		,153	,155	,152	,153	,153
Kurtosis		-,751	-1,060	-,192	,185	,086
Std. Error of Kurtosis		,304	,308	,303	,306	,304

What do you think about the information flow in your organisation? [information flows freely]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	45	17,4	17,7	17,7
	somewhat agree	106	40,9	41,7	59,4
	neither	36	13,9	14,2	73,6
	somewhat disagree	44	17,0	17,3	90,9
	strongly disagree	23	8,9	9,1	100,0
	Total	254	98,1	100,0	
Missing	Unknown	5	1,9		
Total		259	100,0		

What do you think about the information flow in your organisation? [information is abstract]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	24	9,3	9,7	9,7
	somewhat agree	79	30,5	31,9	41,5
	neither	51	19,7	20,6	62,1
	somewhat disagree	63	24,3	25,4	87,5
	strongly disagree	31	12,0	12,5	100,0
	Total	248	95,8	100,0	
Missing	Unknown	11	4,2		
Total		259	100,0		

What do you think about the information flow in your organisation? [information is concrete]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	67	25,9	26,2	26,2
	somewhat agree	108	41,7	42,2	68,4
	neither	39	15,1	15,2	83,6
	somewhat disagree	33	12,7	12,9	96,5
	strongly disagree	9	3,5	3,5	100,0
	Total	256	98,8	100,0	
Missing	Unknown	3	1,2		
Total		259	100,0		

What do you think about the information flow in your organisation? [information flow is controlled]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	78	30,1	31,0	31,0
	somewhat agree	108	41,7	42,9	73,8
	neither	32	12,4	12,7	86,5
	somewhat disagree	27	10,4	10,7	97,2
	strongly disagree	7	2,7	2,8	100,0
	Total	252	97,3	100,0	
Missing	Unknown	7	2,7		
Total		259	100,0		

What do you think about the information flow in your organisation? [information is written down]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	91	35,1	35,7	35,7
	somewhat agree	96	37,1	37,6	73,3
	neither	33	12,7	12,9	86,3
	somewhat disagree	27	10,4	10,6	96,9
	strongly disagree	8	3,1	3,1	100,0
	Total	255	98,5	100,0	
Missing	Unknown	4	1,5		
Total		259	100,0		

Q36 – Coding #6

Statistics

		What do you think about the relationships in your organisation? [relationships are personal]	What do you think about the relationships in your organisation? [relationships are hierarchical]	What do you think about the relationships in your organisation? [relationships are competitive]
N	Valid	256	255	255
	Missing	3	4	4
Mean		2,359	2,082	2,616
Skewness		,726	,988	,411
Std. Error of Skewness		,152	,153	,153
Kurtosis		-,269	,184	-,889
Std. Error of Kurtosis		,303	,304	,304

What do you think about the relationships in your organisation? [relationships are personal]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	49	18,9	19,1	19,1
	somewhat agree	125	48,3	48,8	68,0
	neither	32	12,4	12,5	80,5
	somewhat disagree	41	15,8	16,0	96,5
	strongly disagree	9	3,5	3,5	100,0
	Total	256	98,8	100,0	
Missing	Unknown	3	1,2		
Total		259	100,0		

What do you think about the relationships in your organisation? [relationships are hierarchical]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	89	34,4	34,9	34,9
	somewhat agree	103	39,8	40,4	75,3
	neither	25	9,7	9,8	85,1
	somewhat disagree	29	11,2	11,4	96,5
	strongly disagree	9	3,5	3,5	100,0
	Total	255	98,5	100,0	
Missing	Unknown	4	1,5		
Total		259	100,0		

What do you think about the relationships in your organisation? [relationships are competitive]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	43	16,6	16,9	16,9
	somewhat agree	101	39,0	39,6	56,5
	neither	39	15,1	15,3	71,8
	somewhat disagree	55	21,2	21,6	93,3
	strongly disagree	17	6,6	6,7	100,0
	Total	255	98,5	100,0	
Missing	Unknown	4	1,5		
Total		259	100,0		

Q37 – Coding #6

Statistics

		What do you think about the goals and coordination between teams in your organisation? [goals are existent and known]	What do you think about the goals and coordination between teams in your organisation? [goals are prescribed]	What do you think about the goals and coordination between teams in your organisation? [coordination is hierarchical]	What do you think about the goals and coordination between teams in your organisation? [coordination comes from negotiation]	What do you think about the goals and coordination between teams in your organisation? [coordination comes from self-regulation]
N	Valid	254	254	253	249	252
	Missing	5	5	6	10	7
Mean		2,000	2,079	2,146	2,639	2,306
Skewness		1,096	,921	,751	,382	,778
Std. Error of Skewness		,153	,153	,153	,154	,153
Kurtosis		,399	,279	,108	-,578	,339
Std. Error of Kurtosis		,304	,304	,305	,307	,306

What do you think about the goals and coordination between teams in your organisation? [goals are existent and known]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	104	40,2	40,9	40,9
	somewhat agree	90	34,7	35,4	76,4
	neither	26	10,0	10,2	86,6
	somewhat disagree	24	9,3	9,4	96,1
	strongly disagree	10	3,9	3,9	100,0
	Total	254	98,1	100,0	
Missing	Unknown	5	1,9		
Total		259	100,0		

What do you think about the goals and coordination between teams in your organisation? [goals are prescribed]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	84	32,4	33,1	33,1
	somewhat agree	103	39,8	40,6	73,6
	neither	37	14,3	14,6	88,2
	somewhat disagree	23	8,9	9,1	97,2
	strongly disagree	7	2,7	2,8	100,0
	Total	254	98,1	100,0	
Missing	Unknown	5	1,9		
Total		259	100,0		

What do you think about the goals and coordination between teams in your organisation? [coordination is hierarchical]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	67	25,9	26,5	26,5
	somewhat agree	114	44,0	45,1	71,5
	neither	44	17,0	17,4	88,9
	somewhat disagree	24	9,3	9,5	98,4
	strongly disagree	4	1,5	1,6	100,0
	Total	253	97,7	100,0	
Missing	Unknown	6	2,3		
Total		259	100,0		

What do you think about the goals and coordination between teams in your organisation? [coordination comes from negotiation]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	34	13,1	13,7	13,7
	somewhat agree	93	35,9	37,3	51,0
	neither	65	25,1	26,1	77,1
	somewhat disagree	43	16,6	17,3	94,4
	strongly disagree	14	5,4	5,6	100,0
	Total	249	96,1	100,0	
Missing	Unknown	10	3,9		
Total		259	100,0		

What do you think about the goals and coordination between teams in your organisation? [coordination comes from self-regulation]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	44	17,0	17,5	17,5
	somewhat agree	126	48,6	50,0	67,5
	neither	50	19,3	19,8	87,3
	somewhat disagree	25	9,7	9,9	97,2
	strongly disagree	7	2,7	2,8	100,0
	Total	252	97,3	100,0	
Missing	Unknown	7	2,7		
Total		259	100,0		

Q38

Free text field

Appendix B – Coding Table

Coding #1	
non-IT focused	1
IT focused	2

Coding #2	
Yes	1
No	2

Coding #3	
< 15	1
15 to 50	2
51 to 200	3
201 to 500	4
> 500	5

Coding #4	
Utilities	1
Financial	2
Industrial Goods	3
Technology	4
Consumer Goods	5
Healthcare	6
Services	7
Basic Materials	8
Other	9

Coding #5	
Becoming aware of a problem	1 = Yes, 2 = No
Diagnose a problem	1 = Yes, 2 = No
Finding and selecting a solution	1 = Yes, 2 = No

Coding #6	
strongly agree	1
somewhat agree	2
neither agree or disagree	3
somewhat disagree	4
strongly disagree	5
Unsure	9

Coding #7	
very easy	1
somewhat easy	2
somewhat hard	3
very hard	4
Unsure	9

Coding #8

never	1
sometimes	2
often	3
always	4
don't know	9

Coding #9	
There is a specialised team to conduct the analysis	1 = Yes, 2 = No
Someone is assigned on a case basis	1 = Yes, 2 = No
The analysis is done by a service provider	1 = Yes, 2 = No
The team who requires the data analysis	1 = Yes, 2 = No

Coding #10	
Yes	1
No, but I influence the decisions	2
No	3

Coding #11	
strong influence	1
weak influence	2
no influence	3
don't know	9

Coding #12	
driven by top level management	1
supportive	2
rejective	3
unsure	9

Coding #13	
Yes, to the better	1
No change	2
Yes, to the worse	3
unsure	9

Coding #14	
shifted towards lower hierarchy	1
no change	2
shifted towards higher hierarchy	3
unsure	9

Coding #15	
strong influence	1
weak influence	2
no influence	3
don't know	9

Coding #16

Yes, formal approach with metrics	1
Yes, but without metrics	2
No	3
don't know	9

Coding #17	
not succesful	1
partially successful	2
successful	3
don't know	9

Coding #18	
got better	1
did not change	2
got worse	3
don't know	9