UNIVERSITY OF TURKU
Turku School of Economics

A DESIGN STUDY TO ENHANCE PERFORMANCE DASHBOARDS TO IMPROVE THE DECISION-MAKING PROCESS

Master thesis:
IT Enterprise Management
Supervisors:
E. Caron & E. Koskivaara

Author (s):
Eric Beem (U429259)
10.03.2020 Tilburg

FOREWORD

I have written this thesis as a final piece of my double Master program Information Technology for Enterprise Management (ITEM). I was one of the lucky ones who had the opportunity to participate in this program. I had the opportunity to visit an amazing country and meet lovely people who are still in my heart. Due to the short time period and the lack of knowledge in programming languages it took me way more time than I expected. Furthermore, I have experienced that there is not a good overview of the available performance dashboards.

I have also experienced that writing a master thesis is not an easy task. It is a learning process with multiple lessons for my future life. I have made some wrong decisions, but I have learned from every decision. This process changed my way of thinking and communicating for the future.

I want to utilize this opportunity to thank the ones who have supported me while writing this thesis. First, I want to thank the University of Tilburg for giving me the opportunity to participate in this program. I want to thank my supervisor E. Caron for his guidance and patience during my Master thesis. I also want to thank E. Koskivaara, my second supervisor for her support during my time in Finland and her flexibility during my Master thesis. Finally, I want to thank my friends and family for their support during this period. I know I was not the friendliest person in this period, and I am grateful I have such great friends and family. Even in bad times they tried to help me with everything.

Kind regards,

Eric Beem

MANAGEMENT SUMMARY

Performance dashboards are tools that can be used to improve the decision making in an organisation (Henke et al., 2016). Nevertheless, organisations have trouble finding the right person to integrate and analyse the data in an organisation (Henke et al., 2016). This is not solely because the data analyst does not have the capabilities, but also because there is an information imbalance between the management board and the data engineer. Nowadays we live in a digital era and data plays an important role for organisations (McGee, Prusak and Pyburn, 1993). This thesis aims to solve this problem by creating an artefact to improve performance dashboards with explanatory business diagnoses. This will solve the imbalance between the management board and the data engineers and will improve the decisions in the organisation.

The first chapter is starts with the practical and scientific relevance and gives reasons why an artefact is needed. The research and sub questions are formulated, and the scope of this thesis is described. The second chapter focuses on the history of business intelligence (BI) and the role of BI in performance dashboard. Business intelligence and performance dashboards are related. Furthermore, the characteristics of performance dashboards and different performance dashboards are discussed. Multiple articles are combined to form four important characteristics for performance dashboards.

1. Flexibility: a performance dashboard needs to be easy to modify, used by multiple users and the ability to personalize the overview page.
2. Interactive: a performance dashboard needs to have the ability to drill down, monitor KPI's and show not solely graphs.
3. Visual: a performance dashboard needs to give a visual overview of accurate data from the past and the present day in time.
4. External benchmarking: a performance dashboard needs to have the ability to compare the results with competitors and make prescriptive and predictive analysis based on the data.

This chapter ends with a comparison of different performance dashboards to find the most suitable tool for this research. Power BI is the most suitable tool for this research because it is easy to use and free. The focus of the third chapter is on the decision-making process. The articles of Mintzberg (1970) Endsley and Garland (2000) and Eppler and Mengis (2004) form the basis of this chapter. Information influences the decision-making process, but information can also lead to information overload (Eppler and Mengis, 2004). This

chapter gives an overview of important factors in the decision-making process. These factors are used to improve the performance dashboard. Chapter four is about business diagnoses and explains the model of the artefact. The artefact is based on an article of Daniels and Feelders (2001). This article states that a good business diagnosis is based on six different steps.

1. determine the actual data (normalised/absolute and scaled or not scaled);
2. determine the reference data (normalised/absolute and scaled or not scaled);
3. get model relations from star scheme;
4. compute influence of reference data to determine causes;
5. filter causes to avoid information overload;
6. visual explanation tree of causes.

These steps are used to create the artefact. Chapter five analyses a new business diagnosis tool from Power BI. This tool is called the decomposition tree. This chapter finds out if it is useful for automated business diagnosis. The artefact is described in chapter six and different graphs and outcomes are displayed. The research ends with a conclusion about the advantages of the artefact, limitations and future research.

Table of contents

List of figures

Table of Tables

1 INTRODUCTION

The world is continuously changing. The last twenty-five years the world has changed from an industrial economy to an information economy (McGee, Prusak and Pyburn, 1993). The article is from 1993 and it already predicts the digital era we are living in right now. This modern society is focused on technology. With the rise in technology and the need for understandable data we can speak of a modern society. Every year, organizations are improving on a scientific and technological level. This leads to an exponential rise in knowledge, data and new technologies (Goldie, 2016). An increase in the amount of data can create chaos, because data itself has no value. Information with no value creates chaos and leads to information overload. Data is something that can be useful for organisations, but only if data can be transferred into knowledge (Cheston, Flickinger and Chisolm, 2013). The ability to access relevant information and harness the resources offered by the organisation and opinions of others have become an important skill, particularly as the need for lifelong learning, both formal and informal is increasingly recognized by individuals, organizations and institutions (Cheston et al., 2013). In this digital era, it is important to focus on information. Organizations compete on the ability to collect, manipulate, interpret and use this information in such a way that it increases the competitiveness (McGee et al., 1993). If an organisation is not trying to move forward, it will be overtaken by their competitors (McGee et al., 1993). There are a lot of advantages of big data. Knowing what your business is doing in a fast and cheap way. Furthermore, it is possible to understand the market and control the online reputation (Shim, French, Guo and Jablonski, 2015).

The problem right now is that there are multiple organisations who collect a lot of data, but do not fully benefit from this data collection. This is related to the fact that it is not solely about collecting the data, but also about having a good overview of what to do with this data and how to present is. The report of McKinsey shows that companies have trouble finding the right person to integrate and analyse the data in the organisation (Henke et al., 2016). This leads to high demand of data scientists and this increases the wages of the data specialists. In the United States the average wage of data scientists increased with approximately 16 percent per year between 2014 and 2016.The average wage of other employees increased with two percent in the United States (Henke et al., 2016). The increase in data science programme graduates is not in line with the increase of needed data scientists. McKinsey predicts that there needs to be an increase of twelve percent to keep up with the demand of employees with technology knowledge. The predictions only expect seven percent growth of data scientists. This will lead to a shortfall of 250.000 data scientists in the world. A countervailing force could solve this problem. It is possible to automate the data preparation part. This data preparation part is around 50% of the process (Henke et al., 2016).

Furthermore, organisations focus on hiring data scientists. The organisations assume that a data scientist can make an analytical transformation and business reports that will improve the business processes (Henke et al., 2016). This is not always the most important person for an organisation. An important role is for business translators. A middleman who serves as the link between the business side and the analytical side. This person needs to ask the right questions to stir the analytical side into the right directions to make analytical reports of the key factors for the organisation (Henke et al., 2016). This middleman needs to have knowledge of the organisation and industry expertise. This was one of the reasons for this design thesis. Because the analytical side and the business side are two different islands in an organisation, it is extremely important to have a good explanation of the given graphs and performance dashboards. The two teams need to be on the same page and there needs to be more information from both sides. For example, if the business side asks for an overview of the profit of last year. Then the analytical part of the organisation can do this in multiple ways. The easiest way to understand the data is by showing a graph with the total profit of the year. This graph can show a profit of 6 million dollar a year and the profit of every quarter. This sounds as a good year, but it does not give the real picture of the data. The data scientist knows that every month has a different weighted factor and that in some months there was a huge loss. This is not included in the picture and there is no written information in the graph. This is an example of information imbalance. The business side can think that everything is going well, but it is important to drill down on the data and give an explanatory analysis of the graphs and data in performance dashboards. These dashboards are being used more and more in an organisation. Growing is important for organisations and this can only be done if the organisation understand the data in the organisation. If they know that December is one of the most important months because of Christmas, they need to make sure that they are fully prepared in that month. If the data shows that March is a month where there are not many productions, they need to find new ways of advertising and making deals to increase the production in this month.

The article of Shim et al. (2015) shows that there are multiple organisations investing in big data and aiming for a return of 3,50$ per one dollar spent. Nevertheless, there are organisations who lose money by investing in big data. There are several reasons why investing in big data is not always profitable. Organisations currently start by installing sensors or other devices which can collect data but do not have the right storage and network requirements, people or budget (Katal, Wazid and Goudar, 2013). In addition to these issues, even if all these challenges are fixed it is possible that an organization still cannot fully benefit from big data and business intelligence. One of the reasons is that the front-end is not fully clear for all the stake- and shareholders. Because a performance dashboard is mostly small and does not give details of the graphs and pie charts, it is unclear for the reader. The performance dashboards show not enough information and what the performance dashboard show is poorly (Few, 2006). This is a real problem for data analysts and managers. You want to have a good overview of all the important information, but it needs to be clear enough for everybody. Performance dashboards are made to solve this problem. Nevertheless, performance dashboards can be improved to prevent miscommunication and give data analysts the tools to analyse the correct data.

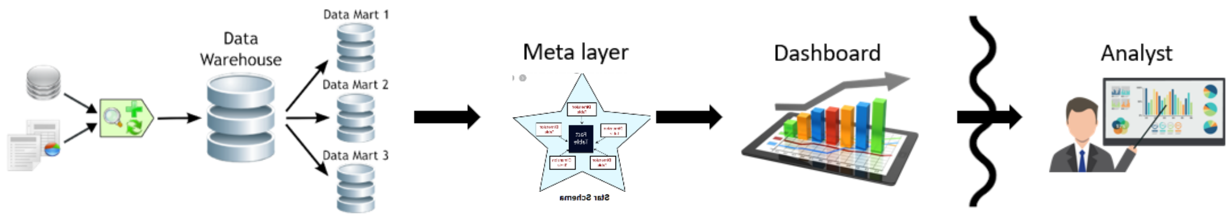

Figure 1: visual example of the addressed problem

Figure one gives a good overview of the process of using information to improve the decision making of the management board. It starts with collecting data in the organisation. This can be done with installing sensors and other devices to collect the data. This data will be stored in a data warehouse with different data marts to improve the data distribution for different departments. When the data is placed in a data mart it is easier for departments to find the right data. To analyse the data, there needs to be a meta layer. A star model can support the explanatory analytics. When there is a sort of meta layer it is possible to build a performance dashboard based on data. The line in the last arrow shows the critical point that will be addressed in this thesis. An analyst could understand the data wrong, because he has no knowledge about the data side and the dashboard gives not enough information.

1.1 Goal

The goal of this research is to improve the performance dashboards. Nowadays, the dashboards are too complex for different stakeholders. This leads to an information imbalance between the creators and the readers. The information is clear for the IT department but not for the rest of the organisation. Furthermore, the information delivered in those dashboards is not always the right information (Few, 2006). This research is design based and will provide an improved version of a performance dashboard. This is called an artefact. This dashboard will have an option to ask for textual and visual clarification on certain points to reduce the information disbalance between the creator and the reader.

1.2 Research question

The main research question for this research is: How can a performance dashboard be extended with features for explanatory analytics? This question is divided into two sub questions that relate to the literature review and the empirical part of the research. These sub questions will help the researcher to answer the research question. The following questions cover the topics in the literature review:

1. What are performance dashboards, and which performance dashboard is most suitable?
2. What is the relation between diagnostic analysis and performance dashboards?

1.3 Scope

Before addressing the empirical questions, it is important to explain the scope of this research. This research is done to complete my Master's degree, ITEM, on the Tilburg University and Turku University. This research is not conducted in partnership with another firm. Power BI is used in multiple firms to have to monitor their data. With the help of Python, it is possible to adjust certain graphs and tables. This design research will be tested in a real Power BI dashboards with a dataset to test the advantages and disadvantages of the performance dashboard 2.0. In this research the central performance dashboard will be Power BI. The reason for this is that it is a cheap or free performance dashboard tool from Microsoft. Multiple big organizations are using this tool and will keep using this performance dashboard in the future. Furthermore, it is possible to use the programming language Python to adjust the performance dashboard. Because a programming language is needed to create an automated business diagnosis a combination of Python and Power BI is used.

1.4 Contribution

1.4.1 Practical contribution

As stated in the introduction, data usage is increasing in almost every organisation. Even the small companies can gain advantage from analysing their data to improve their performance. Business intelligence is one of the terms used for improving the decisions in organisations and a lot of times performance dashboards are used to visually summarize the most important information in an organisation. Nevertheless, organisations are using performance dashboards to inform data analysts, stakeholders and shareholders. To prevent miscommunications, it is important that every employee can get additional information in the performance dashboard. This artefact reduces the costs of meetings for explaining the performance dashboards or training employees to understand the dashboards. This saves money and time for the organisation.

1.4.2 Scientific contribution

Every year there is an increase in information. Nevertheless, the increase in information is higher than the ability to make use of this information (Koomey and Holdren, 2008). The literature states that there is a problem with this increased usage of performance dashboards. Such a dashboard is clear for the creator, but the dashboard can be unclear for the readers (stakeholders or other people with less BI knowledge). In addition, a dashboard does not give a good overview of the most important attributes in an organisation. To give an example, it is possible that there is an increase in performance, but that the sick days also increased in an organisation. This relation is not visible in the performance dashboard but can be a sign of people working too hard. There is not much information about extending performance dashboards and if there is an article about performance dashboard it doesn't focus on business intelligence. For the practical contribution it can help organisations to improve their performance and reduce the time to explain performance dashboards to stakeholders and readers. Nowadays, it is normal that a performance dashboard is explained in a one-hour meeting. On the one hand it is easy to implement features in dashboards but on the other, getting text in a dashboard is difficult. Performance dashboards are visual based and not text based. You can add text in a dashboard, but you need another program, such as Python or R, to include automated text in the dashboard. This research tries to add value to the literature by implementing an automated business diagnosis in a performance dashboard.

1.5 Research methodology and data

The research is design based. A design study delivers an artefact. The artefact will be an updated performance dashboard with extra information to decrease the information disbalance and to help employees understand their organisation in a better way. To create a good performance dashboard a literature study is necessary. What are the important features of a performance dashboard and what are the different performance dashboards. It is important that this research is in line with other research to make sure it will have scientific and practical relevance.

The literature provide multiple articles to improve performance dashboard. These include explanatory analytics, storytelling and improved analytical analyses. This research will focus on a combination of those three because they have the same basis. The focus on those articles is to improve the performance dashboard of organisations. In this design study the performance dashboard Microsoft Power BI will be used. Power BI is easy to use, cheap and multiple organisations use this performance dashboard. In the next chapter the reason for using Power BI will be described in detail. A performance dashboard uses data to generate reports. Therefore, data is needed to test the prototype and these databases are collected from the Tilburg University and the high school R.S.G. Pantarijn in Wageningen. The dataset from Tilburg University is based on the sales data of a foodmart operating in the USA, Canada and Mexico. This dataset is already cleaned and consists of a star model with two fact tables. The researcher used this dataset, because it was already cleaned, consists of more than 100.000 sales data cells and it consists a star model.

The other dataset is from a high school in Wageningen and is focused on a bike shop. This data is much smaller and there is not a data model behind it. Because this research is focused on creating an artefact for explanatory analytics it is not necessary to explain the extraction, transform and load steps of data science. As stated above, performance dashboards are not likely to have a text option combined with R of Python. Because a performance dashboard is visual based, it is possible to create a table or other visuals to display data. This table can be modified so that it is possible to show solely text. This research will use Python instead of R because the researcher has more knowledge about Python compared to R. To compare the dataset, a reference value is needed. This reference value can be created with data from inside the organisation or from external data. This depends on the availability of the data of competitors.

1.6 Structure of the thesis

This study consists of seven chapters. The first chapter explains the purpose and the context of this study. Chapter two is focused on the characteristics and different performance dashboards. Chapter three highlights the managerial decision making because a performance dashboard is used for decisions. Chapter four is focused on automated business diagnosis. Chapter five analyses the new business diagnosis tool of Power BI and chapter six shows the improved performance dashboard based on the literature review, Python and Power BI. Finally, the thesis ends with the conclusion and limitations and future research in chapter six, followed by the references and the appendices.

2 THE EVOLUTION OF PERFORMANCE DASHBOARDS

This chapter analyses topics that relate to the research question. To improve a performance dashboard, it is necessary to understand the key factors of different performance dashboards. Furthermore, it is important to focus on the user opinions, because the users are the employees who are working with the dashboards. Performance dashboards are supportive tools for organisations to improve their decision making. A dashboard is a tool but making decisions based on data is called business intelligence. Therefore, it is not possible to write about performance dashboards without talking about business intelligence. Business intelligence has changed rapidly in the past twenty years. This change in business intelligence has had impact on the use of performance dashboards. Because performance dashboards are used to make decisions based on data (business intelligence), they are closely related. This chapter shows the evolution of business intelligence and performance dashboards.

2.1 Business intelligence

Performance dashboards and business intelligence are related topics. Business intelligence is a hot topic nowadays. Because of the increase in data usage, there is a higher demand for people who can understand the data, transform the data and use the data. Companies use a lot of data and this can be used to make better business decisions. Business intelligence is focused on making good choices based on this data, which can then positively influence the organisation. Business intelligence is focused on shifting from big data to big impact. There are multiple research topics, trends and other factors that shape the business intelligence research directions. In the book, the World is Flat (2005) by Thomas Friedman, he predicted that several factors, such as high-speed internet, globalisation, outsourcing and international travel will create opportunities for IT advancement. These factors can be summarized to ultra-fast global IT connections. These connections, in combination with the development of business standards, databases and electronic data interchange formats have facilitated the possibility for creating and using data. The internet as of 1970 and the World Wide Web has increased the possibility of collecting and generating data. After the internet the web-based, mobile and sensor generated data arrived and created even more data. Nevertheless, collecting data does not lead to a lasting impact on the industry or organisation (The Economist, 2010). This information needs to be structured and used to make decisions based on facts and trends. Organisations shift to being data driven and business intelligence plays an important role in this change. Researchers and practitioners need to analyse the applications and characteristics of data and then use the right analytical tools to derive useful information. There needs to be a good connection between the technical side and the business side. Effective communication is necessary to complete business intelligence projects successfully. The IT departments of organisations play an important role in developing education programmes for the employees and new ways to generate and understand data (Chen, Chiang and Storey, 2012). Business intelligence is mainly focused on understanding the structured data. Business intelligence has a rich history and it can be separated in different versions. These versions will be discussed in the following paragraph. Transforming unstructured data to structured data is based on data science instead of business intelligence. That is the reason that there will be no information about transforming unstructured data to structured data.

2.1.1 Business intelligence 1.0

In the past two decades there has been an increase in the relevance of business intelligence (Chen et al., 2012). Business Intelligence started in the year 1990 and in the nineties it was combined with business analytics. Business analytics was the key factor of business intelligence (Davenport, 2006). Business intelligence relies on data collection, technologies to analyse and other sources of information (Chaudhuri, Dayal and Narasayya, 2011; Turban, Sharda, Aronson and King, 2008; Watson and Wixom 2007). Business intelligence 1.0 was mainly focused on structured data collected through legacy systems and stored in relational databases (Chen et al., 2012). Business intelligence 1.0 was focused on statistical methods and data mining techniques from the 70's and 80's (Chen et al., 2012). Business intelligence 1.0 is mainly focused on data management and warehousing. The design of data and tools for ETL (extraction, transformation and load) are important to start using queries OLAP analysis (Chen et al., 2012). This analysis can be used to create dashboards to show the most important information. The Gartner report of Sallam, Richardson, Hagerty and Hostmann, (2011) considered 13 essential capabilities for business intelligence platforms. Business intelligence 1.0 has the following eight out of thirteen capabilities: reporting, dashboards, ad hoc query, search-based BI, OLAP, interactive visualization, scorecards, predictive modelling, and data mining. A few BI&A 1.0 areas are still under active development based on the Gartner BI Hype Cycle analysis for emerging BI technologies, which include data mining workbenches, column-based DBMS, in-memory DBMS, and real-time decision tools (Bitterer, 2011). Furthermore, this wave is also focused on statistical analysis and techniques for data mining to improve association analyses, clustering, segmentation of data, regression analysis, classification and predicting models (Chen et al., 2012). Almost all these processing and analytical technologies are already in the business intelligence IT vendors (Sallam et al., 2011).

2.1.2 Business intelligence 2.0

When the internet started to play a more dominant role and started to offer unique data collection methods, opportunities to develop and analytical research, business intelligence started to shift from the 1.0 version to the 2.0 version (Chen et al., 2012). With the rise of web search engines such as Yahoo and Google and the rise of companies such as eBay and Amazon it became possible to start online businesses and interact with their customers directly (Chen et al., 2012). With the knowledge gained from cookies and server logs it is possible to customize their products to customer needs and to identify new business opportunities (Chen et al., 2012). With these new technologies there was a shift of importance in the business intelligence. The focus in business intelligence 2.0 is more towards text and web analytics to understand the unstructured web contents (Doan, Ramakrishnan and Halevy, 2011; O'Reilly 2005). Around the year 2000 it became possible to gather industry, customer, product and company information and to visualize this information with the help of web and text mining techniques. With the new online web shops, it became possible to analyse the clickstream data logs (Chen et al., 2012). Google analytics can be used to analyse the user's browsing and purchasing patterns. In addition, it is possible to use all the advantages of web analytics. With web analytics it is possible to design a website, optimize product placement, analyse customer transactions and market structure (Chen et al., 2012). After 2004 more 2.0 business intelligence web applications have been developed. This creates an abundance of new sorts of websites, such as forums, blogs, social media sites, virtual worlds etc (O'Reilly, 2005). It is almost impossible to live without these "new" websites in this era. With the new 2.0 web applications it is also possible to gather amounts of data from different customers and for different businesses at the same time (Chen et al., 2012). The article of Lusch, Liu and Chen (2010) states that analysing social media give businesses a unique opportunity to change the business-to-customer (one way) marketing to a two-way conversation between the business and customer. Another difference between business intelligence tools of the 1.0 and 2.0 versions is that the 1.0 version technologies are already integrated into the IT systems (Chen et al., 2012). Business intelligence 2.0 systems require integration of scalable mature text mining techniques, social network analysis and spatial-temporal analysis with existing DBMS-based business intelligence 1.0 systems (Chen et al., 2012).

2.1.3 Business intelligence 3.0

The next generation of business intelligence is focused on the mobile and sensor-based-content. Where the 2.0 version of business intelligence was mainly focused on the abundance of new information due to the world wide web, the 3.0 version focuses on the increase in mobile phones and tablets (Chen et al., 2012). In October of 2011, the Economist stated that the number of tablets and mobile devices (480 million) that are sold exceeded the number of desktops and laptops (380 million). Furthermore, the same article predicted that the number of mobile connected devices will reach 10 billion in the year 2020. The increase is exponential, and it showed that business intelligence needed to change. All these mobile devices connect to the internet and share and create data with their ecosystem of applications such as games, social media, shopping, education, healthcare, etc. These applications and mobile devices change and improve every month. It is now also possible to connect devices such as fridges and cars to the internet (internet of things), have barcodes and devices that are equipped with radio frequency identification (Chen et all., 2012). This creates new challenges for researchers and business intelligence professionals. Human computer interactions, mobile interfaces and new analytics tools for collecting, processing and analysing the new mobile and sensor data are topics which are incorporated in business intelligence 3.0 (Chen et all., 2012). In 2011 Bitterer (2011) said that these new technologies had the possibility to disrupt the business intelligence market significantly, but as of 2019 it has become more of an addition then a disruption. The uncertainty associated with business intelligence 3.0 presents another unique research direction for the IS community. Key characteristics for this wave are the following: Location-aware, person-centred, context-relevant analyses and the visualization of mobile and human computer interaction (Chen et all., 2012).

2.1.4 Business intelligence 4.0

The fourth generation is mainly focused on the adaptation of artificial intelligence. Artificial intelligence is an autonomous machine which is capable of human-like thinking (McCorduck, 2004). In 1956 McCarthy started with artificial intelligence and the goal was to make applications and machines that could determine languages, concepts and solve different problems (McCarthy, 2006). Nowadays, there are multiple articles describing AI, but the basic is still the same. The goal of artificial intelligence is to mimic the cognitive function of the human brain and to be able to learn of the past and to solve problems (Lee, Suh, Roy and Baucus, 2019). Artificial intelligence can help business intelligence professionals by producing results that can be used to make successful decisions (Lee et all., 2019). The article of Lee et all. (2019) states some key characteristics of artificial intelligence that can have positive effect on business intelligence. Neuro linguistic programming, deep learning, recommendations systems and artificial neural network are all key characteristics that can be used in the relation of business intelligence and artificial intelligence.

The table below summarizes the key characteristics of the evolution of business intelligence based on the above information. The future will have new waves and business intelligence will change because technology is always evolving. The business community is starting to adopt business intelligence and see the advantages of business intelligence. It all starts with collecting all your data and use this data to make good decisions. Nowadays organisations collect a lot of data but after the collection phase the business intelligence phase stops. The reason for this is the lack of knowledge of what to do with this data. There is a high demand for business intelligence professionals to work with all the available data in a senseful way. These challenges and opportunities in making impact on a scientific and societal level will be addressed soon (Chen et all, 2011). It is important that research and educational programs carefully evaluate the directions of the future, action plans and the curricula from business intelligence 1.0 to 4.0.

BI wave	Key characteristics
BI 1.0	*DBMS-based, structured content* • RDBMS and data warehousing • ETL and OLAP • Dashboards and scorecards • Data mining and statistical analysis
BI 2.0	*Web-based, unstructured content* • Information retrieval and extraction • Opinion mining • Question answering • Web analytics and web intelligence • Social media analytics • Social network analysis • Spatial-temporal analysis
BI 3.0	*Mobile and sensor-based content* • Location-aware analysis • Person-centred analysis • Context-relevant analysis • Mobile visualization and HCI
BI 4.0	*Artificial intelligence content and machine learning* • Neuro linguistic programming • Deep learning • recommendations systems • Artificial neural network

Table 1: Different BI waves

2.2 Essential components of a performance dashboard

After getting a clear overview of business intelligence it is possible to show the components of performance dashboards. Every organization has data and want to visualize this data in such a way that the employees, stakeholders and other actors benefit from it. Not everybody can read and understand data analytics tools. That is the reason that there are multiple tools to visualize your data to be understandable for people with no knowledge in that segment. A Performance dashboard is a diagnostic tool to gives a quick overview of the organisation (Velcu-Laitinen and

Yigitbasioglu, 2012). Performance dashboards are widely used and are replacing ad-hoc analyses (Sallam et al., 2011). These dashboards can help organisation with making decisions. Performance dashboards are a type of decision support systems (Arnott and Pervan, 2005). Performance dashboards provide the most important data in a simple one-page overview (Velcu-Laitinen and Yigitbasioglu, 2012). A dashboard can contain individual or organisational goals and gives the user the opportunity to identify and communicate critical areas that need to be changed (Velcu-Laitinen and Yigitbasioglu, 2012). A performance dashboard contains visual and functional features which can help the interpretation and the cognition (Velcu-Laitinen and Yigitbasioglu, 2012). These dashboards can be broadly used in different levels of the organisation from line workers, by analysing the production process, to executive managers to evaluate strategical choices. Different factors have changed the performance dashboards that are used today. The technology development that is stated earlier has provided the opportunity for a technical infrastructure which can be linked effectively with the business intelligence part of an organisation (Velcu-Laitinen and Yigitbasioglu, 2012). Furthermore, the article of Kaplan and Norton (1992) is focused on balanced scorecards. This article states that there is a need for multidimensional measurements based on the performance of an organisation. Lastly, the book of LaPointe (2005) states that there are two factors that drive the adoption of performance dashboards. The first one is that organisations need a performance reporting tool that is focused on multiple departments instead of different silos. Communication and integration are important to make the organisation one streamlined machine. Secondly, managerial bias leads to a high level of adoption of dashboards. Managerial bias can have negative effect for an organisation because the manager is focused on his own ideas (Wang and Wong, 2012). Performance dashboards can prevent this bias by showing a one-page overview of facts. In the past ten years there is shift in the dashboards. There is a change in purpose, from intrinsic monitoring to strategic analytics (Velcu-Laitinen and Yigitbasioglu, 2012). This change also leads to new features. The new features are a new flexible format presentation, the capabilities to drill down and up, scenario planning and analysis and the integration with other systems such as the workflow systems (Velcu-Laitinen and Yigitbasioglu, 2012). The basic of performance dashboards lays in the theory of information systems, psychology and accounting. A dashboard needs to communicate hard to understand data in an easy way to managers or other persons in charge of decisions with the help of visualization (Velcu-Laitinen and Yigitbasioglu, 2012). Visualization is the presentation of abstract data to amplify cognition

(Card, 1999). The reason for using visualization in performance dashboards is based on the cognitive fit theory. This cognitive fit theory is based on the relation between the skills of making decisions, the format of the visualization and the task itself (Vessey and Galletta, 1991). Different formats are useful for different tasks. Accountants use more tabular information and therefore it is more suitable for symbolic tasks. Forecasters use more graphs because it is better for spatial tasks (Vessey and Galletta, 1991). This theory states that decisions with a high level of fit are better than decisions with a low level of fit. Because it is possible to show different key performance indicators in performance dashboards it is justified to present these in graphical format. Nevertheless, the users of the dashboards differ in knowledge and cognitive profiles. Therefore, it is useful to have some sort of drill down and up capabilities and flexibility to make the dashboard understandable for all users (Velcu-Laitinen and Yigitbasioglu, 2012). Another important feature is that dashboards needs to be focused on the tasks and users instead of the other way around, this is called the task-technology fit (Goodhue and Thompson, 1995). A performance dashboard needs to support the employees instead of producing extra workload.

2.2.1 Key characteristics for performance dashboards

Because there is a lot of information about performance dashboards and technology is changing rapidly it is hard to conclude the main characteristics of performance dashboards. Pauwels et all. (2009) states that a performance dashboard has four purposes. The four purposes, monitoring, communicating, planning and consistency, are confirmed in surveys of Clark, Abela and Amber (2006). The purpose or a performance dashboard is based on the literature and evolves over time. The literature gives the following functions for a good performance dashboard: drill-down and up capabilities, external benchmarking, scenario analysis, real-time notifications and presentation flexibility (Pauwels et al., 2009; Ying et al., 2009, Velcu-Laitinen and Yigitbasioglu, 2012). The drill down function is a key factor for performance dashboards. It allows the user to use the slice and dice option on data. This will give a more detailed analysis without using any other tools (Velcu-Laitinen and Yigitbasioglu, 2012). Furthermore, the drill function will also help users with aggregate data and tasks with a high uncertainty (Bariff and Lusk, 1977; Benbasat and Dexter, 1979). External benchmarking is focused on the relation with their competitors. How well is the organisation really performing? It is possible that your organisation doubled their profit, but this is useless if the profit of you competitors is tripled. Scenario planning can be used in performance dashboards if the dashboard got the possibility to change variables and see the effect on other variables. This is useful when performance dashboards are used as a planning tool (Velcu-Laitinen and Yigitbasioglu, 2012). Performance dashboards can track data daily. It is possible to have real time notifications to alert the users when a measure deviates from its targets. This can be done with an alarm, colour or audio signal (Velcu-Laitinen and Yigitbasioglu, 2012). Lastly, it is important that the visualization part of a dashboard got the ability to show data in different ways (Huber, 1983). A performance dashboard needs to have the ability to quickly change when the user adds, deletes or changes data. This is in relation with the earlier mentioned cognitive fit theory. A performance dashboard should give the user the best option for different tasks instead of letting the user select it (Huber, 1983). Other articles are in line with the above-mentioned literature. A performance dashboard needs to be SMART and IMPACT (No, 2015). Smart means synergetic, monitor, accurate, responsive and timely (No, 2015). Synergetic means that a performance dashboard needs to be visual in one glance, monitor is about monitoring the KPI's of an organisation (No, 2015). This data needs to be accurate, on time and not solely graph based (No, 2015). Impact means that a dashboard needs to be interactive, based on data history, personalised, analytical, collaborative and be able to track data. This means that a performance dashboard needs to be interactive with the possibility to drill down and up, consist of data from now and the past which can be used with one click (No, 2015). The performance dashboard needs to have the ability to personalize the page and it needs to have tools to analyse the data and trace the most important KPI's per user (No, 2015). SMART, IMPACT, the five characteristics of Pauwels et al. (2009) Ying, Lijun and Wei, (2009) Velcu-Laitinen and Yigitbasioglu (2012) and the four purposes of Pauwels et all. (2009) can be used to compare the multiple available performance dashboards. These theories can be combined to the following FIVE characteristics:

1. Flexibility: a performance dashboard needs to be easy to modify, be able to be used by multiple users and the it needs to have the ability to personalize the overview page.
2. Interactive: A performance dashboard needs to have the ability to drill down, monitor KPI's and show not solely graphs.
3. Visual: A performance dashboard needs to give a visual overview of accurate data from the past and the present day in time.

4. External benchmarking: a performance dashboard needs to have the ability to compare the results with competitors and make prescriptive and predictive analysis based on the data.

2.3 Comparing available performance dashboards

Chapter 2.1 and 2.2 describes the characteristics and components of BI and performance dashboards. With the information from these chapters it is possible to compare the available dashboards. There are multiple performance dashboards on the market and the common ones are discussed in this paragraph. The following top four were found on the internet: Microsoft Power BI, Tableau, ClicData and Sisense (Banguy, 2017 and Gartner, 2018).

Figure 2: Magic quadrant for analytics and business intelligence platform from Gartner (February 2018)

Another dashboard that is described is Excel. Excel is one of the first actors in the performance dashboard segment and is still used by multiple organisations and business intelligence employees. Every performance dashboard has positive and negative points. There are not much literature articles about the positive and negative features of different performance dashboards. Therefore, websites will be used to compare the different performance dashboards. Every business intelligence tool is described with a short history, components and the positive and negative features. This research starts with the top four based on the internet. The first performance dashboard that will be described is Microsoft Power Bi (2.3.1), the second one is Qlik (2.3.2), the third one is Tableau (2.3.3), the fourth one is Sisense (2.3.4) and the last one is Excel (2.3.5).

2.3.1 Microsoft Power BI

Microsoft is one of the biggest organisations in the world and it is no surprise that this organisation got their own performance dashboard. It all started in 2009 when Microsoft introduced Power Pivot for Excel (Cohen-Dumani, 2016). It was a tool for self-service business intelligence in organisations. Nevertheless, it was not a big success. The reason for this was that Microsoft did not use any sort of marketing to promote their new business intelligence tool. The new business intelligence tool had a slow start and only some experienced users of Excel used this tool. The business intelligence professionals were surprised by this approach of Microsoft. In their eyes it was a useful tool for people to get insights in their data. For years Microsoft did not respond to the request of the business intelligence professionals to improve this new tool. In addition to this the professionals also asked for an addition to this tool to share reports within a team. Microsoft declined this option and responded with the possibility of using SharePoint in this situation. The reason for this response was that behind the scene Microsoft worked on project “Crescent” (DataTeamFlair, 2018). This project was the beginning of the new self-service business intelligence tool of Microsoft. In 2011 Microsoft made project “Crescent” available in combination with another project, “Denali”, which was focused on SQL coding. This was later renamed to Power BI and since September 2013 part of the Office 365 of Microsoft. After 2013 Microsoft added new features to upgrade their self-service business intelligence tool. In 2018 Gartner named Power BI as the leader in "2018 Gartner Magic Quadrant for Analytics and Business Intelligence Platform" (Howsen et al., 2018). This was not the first year for Microsoft, Microsoft won this price for the twelfth time in a row.

Power BI has different components that form the tool. The following tools are in the literature: Power Query, Power Pivot, Power View, Power Map, Power Q&A and Power BI Desktop (Rad, 2019). These components form the basis of the business intelligence tool. Power Query got the possibility to transform and mash up data (Rad, 2019). Power query reads, transforms and load data in the business intelligence tool. Power Query also uses a formula language M that can perform measures. Power Pivot is a data modelling component working with xVelocity in a memory tabular engine (Rad, 2019). This engine produces fast outcomes and the opportunity to build star schemes, build relationships, create columns and create measures. Data Analysis eXpression language (DAX) is the functional language and this can be used to run functions (Rad, 2019). These functions are available in the library. Power View is the visualization component. This component gives the opportunity to show your data in a visual way. There are multiple options and you can slice and dice through your data (Rad, 2019). The next component is Power Map, this component is for visualizing Geo-spatial information in 3D mode (Rad, 2019). This component works together with Bing maps to give Power BI the possibility to use the world map and 3D measures in the reporting tool. The Power BI desktop and website are both components that can be used by analysts and gives employees the possibility to use the tool everywhere. The last component is the Power Q&A. This component gives the users the ability to ask questions about the report (Rad, 2019). This is still complex to use and not every report got the ability to use this option. For analytical programming Power BI connects to OLAP cubes with the SQL servers (Educba, 2019). This gives the opportunity for multidimensional analytics.

After the history and components of Microsoft Power BI it is now time to address the advantages of this tool. As stated above it is one of the best self-service business intelligence tools. Microsoft Power BI is updated on regular basis and has an active community (Howsen et al., 2018, Heller, 2018). Power BI has predefined charts which can be useful for inexperienced users and provide good looking visualizations (Howsen et al., 2018). Customers place Power BI in the top three for ease to use (Howsen et al., 2018). Microsoft aims to attract customers with products that can be used by everybody and gives enough support when a user has questions (Howsen et al., 2018). The strength of this tool is the low price. It is possible to have a free trial and when an organisation wants to have the premium version it costs around ten dollar a month (Howsen et al., 2018). Microsoft made the price of the analytics and business intelligence market go down. Because Power BI is a tool that is embedded in the Microsoft package, which is already adopted by most of the organisations, it is an easy tool to use. (Howsen et al., 2018). Lastly, Microsoft is improving in the comprehensive product vision. Investment is made in a set of visionary capabilities that are integrated in Power BI (Howsen et al., 2018). These investments contain enhancements to augmented analytics which leads to new features in machine learning. Artificial intelligence, reporting services, open data model and preparation with flows of data are also available for inexperienced users (Howsen et al., 2018). It is also possible to use machine learning with the help of Microsoft Azure to analyse trends and patterns within the data. Power BI is starting to improve the artificial intelligence function (Phocassoftware.com/business-intelligence-blog, 2019). This function is only for premium members and consists of three main topics: detect languages, extract key phrases and calculate sentiment scores. Detect languages is based on detecting the language in the text. Extract key phrases got the ability to highlight important data and calculate sentiment score is focused on the sentiment in the data, is it positive or negative (Phocassoftware.com/business-intelligence-blog, 2019). In November 2019 Power BI updated Power BI to add the decomposition tree. This is a new function with the ability to AI split. This gives the user the ability to see the influence of certain variables on one variable.

Nevertheless, there are also downsides of Power BI. Although Microsoft offers Power BI as one tool, you need additional products for different tools. For example, if you want to use conversational analysis you need Cortana Assistant (Howsen et al., 2018). Another example of this downside is that you need Microsoft Power BI and SQL server reporting service for making reports (Howsen et al., 2018). There is also a difference in on-premise and cloud. It is possible to share Power BI reports, but not the dashboards. It is missing some of the features that can be found in Power BI SaaS and Microsoft does not give the freedom to their users to choose a cloud infrastructure as a service (Howsen et al., 2018). It is only running in Azure because it is connected to Microsoft. The article of Educba states that Microsoft Azure is not the best tool to analyse trends and patterns (Educba, 2019). Other business intelligence tools use better options such as Python or R (Educba, 2019). Furthermore, if the user wants to load data into Power BI there are limited options. Not every data source is connected (Educba, 2019). This is not as bad as it sounds because there are still a lot of data sources available. Nevertheless, it has a maximum storage space of 10GB, or the user needs to put data in the cloud of Microsoft (Educba. 2019).

Power BI is one of the best choices for organisations that use Windows, Office or Azure. Furthermore, it is a good choice for cost-sensitive organisations that want to have business intelligence for every employee in the organisation. It is not as good as other tools, but the relative ease to use, cheap price and active community make it one of the best tools nowadays (Howsen et al., 2018).

2.3.2 Qlik

Qlik saw that the world needed a new type of software. A software that can embody and reflect in the same way as the human mind and is be able to create detailed data analysis (Oresundstartups, 2012). In 1993 Berg and Gestrelius started Qlik in Lund, in the south of Sweden. The first software Berg and Gestrelius made was QUIK (in 1996 changed to Qlikview). QUIK stands for quality, understanding, interaction and knowledge. These four words are still the cornerstones of Qlik nowadays. Qlik's business strategy was to venture where no other organisation was active. Qlik provided business intelligence to small and medium sized enterprises. Qlik actively contacted small organisations to offer Qlikview. At the beginning of the 21st century Qlik changed its CEO and DFO to safe the organisation from bankruptcy. Because Qlik was offering multiple software's it was losing money. The CEO decided to only focus on business intelligence and this was successful. Qlik started to grow again, but it still did not have enough capital and expertise. Qlik earned 12.5 million dollars in venture capital. Qlik used this money to expand in other countries. This made the company grow with 35% and it made a profit of 13 million dollar. In 2010 is was one of the hottest technology organisations in the United States (Oresundstartups, 2012). In 2010 Qlik also made their appearances on the stock market. In the past 8 year the company started to grow and build new headquarters in different continents.

Qlik got different components. Nevertheless, it is important to explain QlikView and QlikSense first. There are similarities in the analysis engine and calculation engine of these two. Therefore, the same codes work in both engines (Qlik community, 2019; Heller, 2018). QlikView is better for guided analytics. End-users have the freedom to analyse, select and explore the dataset. Nevertheless, QlikView has limitations in creating new visualizations (Qlik community, 2019). It is more pre-canned and a good tool if you do not want to make new visualizations. Qlik Sense is better for self-service data discovery and touchscreen. This tool gives more freedom for visualization and is adaptive to different screens. Qlik Sense is better for users who want to have freedom. QlikSense and QlikView are two components, the other components are Analytics Platform, Data Catalyst and Core. There are also some add-on products for Qlik such as GeoAnalytics, DataMarket and Connectors. Qlik Analytics Platform offers access to Qlik's data engine through open and standard application programming interfaces (Analytics8, 2019). Qlik Data Catalyst is focused on enterprise data management. It is focused on transforming the data from raw to clean data in one place for all employees (Qlik, 2019). Qlik Core is a platform for analytical development based around the engine of Qlik. This component gives the ability to build, extend and quickly deploy custom interactive data-driven solutions (Core.Qlik.com, 2018). Qlik GeoAnalytics gives the possibility for mapping capabilities and use GeoAnalytics use cases (Help.Qlik.com, 2019). Qlik DataMarket is focused on the relationship with the competitors. It gives the opportunity to gain access to external data (Qlik.com, 2019). The last component is Qlik Connectors, this tool gives the user the opportunity to interact with external data sources that are not accessible through the normal ODBC or OLE-DB data connections (Qlik.com, 2019.)

There are multiple advantages of Qlik. Qlik is a capable interactive data analyse tool that can connect to almost every SQL database (Heller, 2018). Furthermore, it has the possibility to deal with a huge amount of data and provide more storage space compared to its competitors. It has an in-memory storage. Because of this storage it is possible to scale up your data. Qlik also has a front- and back-end supportive requirement and it is useful for combining machine learning and artificial intelligence (Howsen et all, 2018. The active community and conference series help this organisation to follow the demands of the organisation (Howsen et all, 2018). Another useful option is the possibility to email the reports in PDF files to other employees, (Activewizards.com, 2018). Lastly, Qlik is continuously improving its augmented analytic roadmap, its data preparation and embedded analytics capabilities (Howsen et all., 2018). Qlik is trying to be the best business intelligence tool and tries to work on the negative parts and improving the positive parts in the tool.

A negative part of Qlik is that QlikSense can be hard to understand compared to Tableau and Power BI (Heller, 2018). Furthermore, there is an active community. But the interest in Qlik started to slow down in the year 2018. The reason for this is the cheaper options and because Qlik tries to move the QlikView customers to use the, in their eyes, better QlikSense tool (Howsen et all., 2018). Qlik is also a tool that is harder to understand and a background in data science is needed for this tool. Because the back end is hard to understand. Some developers reported a lot of syntax errors and the formulas showed strange errors (Activewizards.com, 2018). It is hard to work with Qlik without a data science background (Activewizards.com, 2018). Another negative part of Qlik is that you need all the different components to make one good workflow (Howsen et all., 2018). Lastly Qlik does not have ergonomics and visualization quality and the latest reviews of Qlik are showing bad scores (data-flair.training, 2019).

Qlik does not give information about explanatory business diagnosis. Therefore, the researcher concluded that this is not one of the main points of Qlik. The only possibility to give explanatory analytics is using external sources such as Python or R.

2.3.3 Tableau

In the past the numbers and data were for IT professionals. It was their job and the organisation expected reports from them. In 2003 a group started a small software company to make this process more transparent and revolutionary (Intellipaat.com/blog/, 2019). Hanrahan, Chabot and Stolte started the organisation Tableau in 2003 to incorporate data visualisation to make databases interactive and comprehensive (Intellipaat.com/blog/, 2019). Tableau had a rough start because IT professionals did not find it attractive as it was not accepted to use self-service business intelligence tools, but after business analysts and marketers saw the potential in business intelligence of Tableau the organisation started to grow. In 2013, Tableau launched its initial public offering and raised around 250 million dollars (Tableau.com, 2013). In the past four years Tableau analytics has been utilized in more than 150 countries and the organisation is still growing (Intellipaat.com/blog/, 2019). In 2018, there was a report of the performance of Tableau since the initial public offering and the market value exceeded 8000 million dollars in 2015 (Intellipaat.com/blog/, 2019). In 2010, the organisation was almost sold but nowadays around 60.000 organisations use Tableau (Intellipaat.com/blog/, 2019). This number will grow in the future, because it offers strong business intelligence visualization and support for their users (Tableu.com, 2019).

Tableau consists of the following components: Tableau Desktop, Tableau Server, Tableau Online, Tableau Prep Builder, Tableau Visible, Tableau Public and Tableau Reader. Tableau Desktop is a component that gives the user the possibility to make visual analytical reports (Tableau.com, 2019). Tableau Server gives the organisation the possibility to share reports across the whole organisation (Tableau.com, 2019). Tableau Online is a SaaS analytics tool that gives the user the possibility to share your reports and data with other employees (Tableau.com, 2019). Tableau Prep Builder is based on the extract, transform and load process of data (Tableau.com, 2019). This process is focused on getting clean data in Tableau. Tableau Visible is focused on transforming the data to visual reports, with the possibility to drill down and show the right data

(Tableau.com, 2019). The last two components are Tableau Public and Tableau Reader. Tableau Reader is the free lite version of Tableau. With this free version it is possible to make reports and make business decisions, but the lite versions lacks some important features (Tableau.com, 2019). Tableau Public is focused on the use of public data. It does not use the consumer's personal data, but data available through public sources. This component is broadly used by journalists and bloggers (Geekwire.com, 2013).

Tableau presents itself as a product offering analytical tool to give users the possibility to spot patterns quickly and benefit from these findings (Heller, 2018). This is something every self-service business intelligence tool can say. Tableau offers an interactive, visual, intuitive experience that gives managers the ability to present, analyse and prepare data without hard coding skills (Howsen et al., 2018). Tableau is mostly on-premises, but it is also possible to have a stand-alone desktop version and it has a SaaS cloud-based server (Howsen et al., 2018). Tableau introduced a lower budget version, Viewer Role in 2018, for organisation with a low revenue. This version is lacking some features for in-depth analyses but is still good enough for business intelligence analysis (Howsen et al., 2018). Furthermore, Tableau released empirical systems to improve their augmented analytics and Tableau Prep to improve the data preparation of the desktop version (Howsen et al., 2018). The visual graphs and possibilities to show data is top level compared to other tools (Clusterdata, 2018; Howsen et al, 2018). It is also possible to do cross-database joins in Tableau, blend these joins and make visualizations fast (Howsen et al., 2018). Tableau sets the bar for their smooth and easy to use software (Clusterdata, 2018). Tableau is constantly updating and improving their software and is able to adapt to the changes of growing businesses (Howsen et al., 2018). It is also possible to manipulate your data quickly during the visualization process. Tableau got a fanbase of customers. Tableau offers meetings, roadshows and tutorials (Howsen et al., 2018). Nevertheless, there are mixed signals about the customer satisfaction. The same article of Howsen et al., (2018) states that there is a decline of customer support and fixing the problems. One eight of the reference customers states the lack of support is a problem and that Tableau cannot handle big data.

There are also negative parts of Tableau. Tableau is one of the best business intelligence tools on the market right now, but the main problem is the price. Power BI offers around 80% of the features of Tableau for 25% of the price (Clusterdata, 2018). This is a deal breaker for multiple organisations. Working with a potentially inferior tool can result in extra labour cost. Tableau tried to fix this problem by offering a cheaper version, but this cheaper version is only available with a subscription license (Howsen et al., 2018). Perpetual customers need to change to a new subscription model to buy this cheaper license (Howsen et al., 2018). Another issue with Tableau is the negative customer support reviews. There reviews state that there is a lack of support in some parts for querying fact tables and schemas (Howsen et al., 2018). Tableau is trying to fix these issues by placing it on their agenda (Howsen et al., 2018). Poor customer service can break an organisation and give a bad reputation, no matter how advanced their software is. The business intelligence part in most of the organisations is still in the development phase and support is needed to fully benefit from these tools. Lastly, it is useful to have some IT consultancy to fully understand Tableau (activewizards.com, 2018).

One of the good features of tableau is the exploratory data analysis (EDA). This sounds like something that is in line with this research. Nevertheless, this is not fully developed. This EDA gives a summary of the dataset based on the different variables (towardsdatascience.com, 2019). An EDA groups the variables based on classes to give an overview of the data. It will not give an explanatory business diagnosis, although the name suggests differently. This EDA is useful in situation where the data is not clean and needs to be grouped first. In this research the data is already grouped, and the relations are clear.

2.3.4 Sisense

Sisense is an organisation founded in 2004 by Farkash, Harell, Boyangy, Azaria and Israeli (Orpaz, 2016). From 2004 till 2010 Sisense worked on research and development behind the radar. In 2010 this changed because the organisation gained four million dollars in stock funding (Crunchbase.com, 2019). In July 2012 a new CEO was appointed and in 2013 Sisense announced another funding round of 10 million dollar (Bizjournals.com, 2012). In 2014, Sisense announced another funding round of 30 million dollar. This last 30-million-dollar funding was needed to analyse over 10 years of data from CrunchBase: a database consisting of approximately 100.000 organisations. This new dashboard provided the organisations the possibility to query the data in different categories to see trends in the past decade (Forbes.com, 2013). In the next years, Sisense has become a big data group and has announced two more funding rounds: one in 2016 for 50 million dollar and one in 2018 for 80 million dollars. Sisense started to develop itself to provide software for organisation in different countries. In 2018, Sisense operated in 49 different countries (Thenextweb.com, 2013). Nowadays, Sisense is one of the best organisations for business intelligence (Gartner.com, 2012). In 2019, Sisense acquired Perscope Data. The merge of these two organisations resulted in a revenue of 100 million dollar and over 700 employees in 2019 (Techcrunch.com, 2019).

The positive features of Sisense are the embedded use cases. Sisense focused a lot on this segment. This results in the fact that Sisense is one of the best tools for embedded use cases (Howsen et all., 2018). It is possible to ingest data from different data sources and clean and load it into Sisense (Howsen et al., 2018). It is also possible to bring the data in the Sisense ElastiCube. This Cube uses a combination of in-memory and in-chip processing to increase the speed and performance (Howsen et al., 2018). This is useful because the user does not need a data warehouse in this situation.

The negative part of Sisense is that it isn't focused on the workflow compared to the other business intelligence tools (Technologyevaluation.com, 2019). Sisense only covers ten percent of the available workflow functions (Technologyevaluation.com, 2019). Furthermore, Sisense is mainly used in northern America and just started with offices in the United Kingdom and Japan. Therefore, it is a tool that is not widely used all over the world (ActiveWizards). The article of Howsen et all. (2018) states that the functions of Sisense are not good enough. Sisense lacks some grouping functions and advanced analytics for citizen data analyst (Howsen et all., 2018). The reason for this could be that Sisense focused on embedded use cases, a small segment of the business intelligence market (Howsen et all., 2018). This creates less room for developing other segments. This can also be one of the reasons that the support of Sisense is below average compared to other business intelligence tools (Technologyevaluation.com, 2019).

2.3.5 Excel

Excel is the first business intelligence tool that was widely used by organisations. It all started in 2009 when Microsoft introduced Power Pivot for Excel (Cohen-Dumani, 2016). It was originally designed as a tool for self-service business intelligence in organisations. In the next couple of years, new business intelligence tools started to grow, and Microsoft decided to make a new tool. This tool is called Power BI and Microsoft thinks this is the best tool to battle in the business intelligence segment. Although Excel isn't as big as other business intelligence tools it is still a tool widely used by other organisations. It is easy to use, cheap, with an active community and it gives organisations the possibility of using visual data (Phocassoftware.com/business-intelligence-blog, 2019). This is because Excel is based on spreadsheets. This gives the user the possibility to use text and figures. It is also possible to use machine learning in Excel because Excel can run Python scripts.

Nevertheless, it is not the best tool to use. It is slow compared to its bigger brother, Power BI, and it is hard to see errors in the report. You have a limited amount of choices for visual reports and it is not suitable for complex analytics (Educba.com, 2019). Lastly, because it is possible to put spreadsheets on USB sticks or sent them by email this isn't the safest option. If you are not dealing with confident information this is not an issue at all. However, bigger organisations often process a lot of data that is not suitable for everybody.

Excel is a useful tool for smaller organisation with limited amount of data and where complex reports are not needed. It is free and easy to use; the active community and tutorial make it easy for every employee and it is possible to share all your spreadsheets with other employees. There is a possibility to do small explanatory business diagnoses using Excel. These diagnoses are mainly focused on grouping the dataset (support.office.com, 2019). Furthermore, other tools are outperforming Excel and in the future this difference will increase. This is also one of the reasons why Microsoft is focusing more and more on Power BI instead of Excel.

2.3.6 Overview of performance dashboards.

Page 40 shows a table with the top four performance dashboards and Excel. This table gives a simplified overview of the pros and cons of every tool. Based on the table, this research will focus on the Power BI performance dashboard because it cheap to use and it has all the attributes for this research. Excel is not used by bigger organisations and therefore not possible for this research. Sisense is a good tool but it is not globally used, not easy to use without data science knowledge and it doesn't have an active community. Therefore, this isn't suitable for this research. Tableau is also hard to understand and is the most expensive version of the options. The in-build Python Interpreter is something that is useful for this research, but it is impossible for the researcher to get a Tableau license. This leaves two suitable performance dashboards, Power BI and Qlik. There is not a big difference between the two tools, but Power BI is cheaper and easier to use. Furthermore, Power BI got ergonomics and visualization quality and Qlik doesn't provide this. The last reason for the choice for Power BI is that SmartDataPeople (organisation in the Netherlands) recommended Power BI. SmartDataPeople works with a lot of organisations and most of these organisations use Power BI. Microsoft provide Power BI in the Office 365 package, and most organisations got this package.

	Free Trial	Huge volumetric	Active community	Price	Ease to use	Programming connections	Machine learning (ML)
Power Bi	Yes	Yes	Yes	Free/$9.99	High	Python, R and SQL	AI in early stage
Qlik	No	Yes	Yes	Free/$20	Medium	C++, C#, Python and R	AI and ML related
Tableau	Yes	Yes	Medium	$70/$42/$15	Medium	Python, R	Python inbuild ML
Sisense	Yes	Yes	Medium	Quote based	Medium	Python, R and SQL	Yes, early ML innovators
Excel	Yes	No	Yes	Free	High	Python	Python support

Table 2: Overview of performance dashboards (1)

	Supportive requirements	Online analytical programming	speed	Explanatory business analytics
Power Bi	Power BI desktop, gateway	OLAP cubes with SQL for multidimensional analysis	Good, smart recovery	Detect language, extract key phrases and calculate sentiment score (Only for premium users)
Qlik	Qlik consists of both front-end (Qlik Developer) and back-end (Qlik Publisher)	Access to OLAP for encapsulated data views	Fast, works with in-memory storage and ram	No available information. External sources can be used to solve this issue
Tableau	Front end tool such as R	Tableau plus OLAP cube measures at the deepest level	Speed depends on RAM and data sets.	EDU
Sisense	Front end desktop	OLAP cubes	Speed depends on RAM and data sets.	No info available
Excel	Front end desktop, gateway	Simple analytics	Slow	Small number of tools. Mainly based on grouping

Table 3: Overview of performance dashboards (2)

3 MANAGERIAL DECISION MAKING

Business intelligence and performance dashboards are used to improve the managerial decision making in organisations. Every day multiple decisions are made. This can be something small, such as changing the beans of the coffee machine or something big about outsourcing. This decision-making process is based on different steps. Most of those steps are the same in every process. This process will be described in this chapter and will lead to the need of automated business diagnosis's for organisations. Because automated business diagnoses can improve the decision making in organisations this will have impact on the decision-making process. The literature describes strategic decision making as gathering substantial resources, setting a precedent based on these resources and reduce the amount of decision possibilities (Mintzberg, Raisinghani and Théorêt, 1976). Understanding the strategic decision emerging process helps to influence and control these decisions. The figure below shows an overview based on Mintzberg et al. (1976) decision model with small adjustments of Kask (2010).

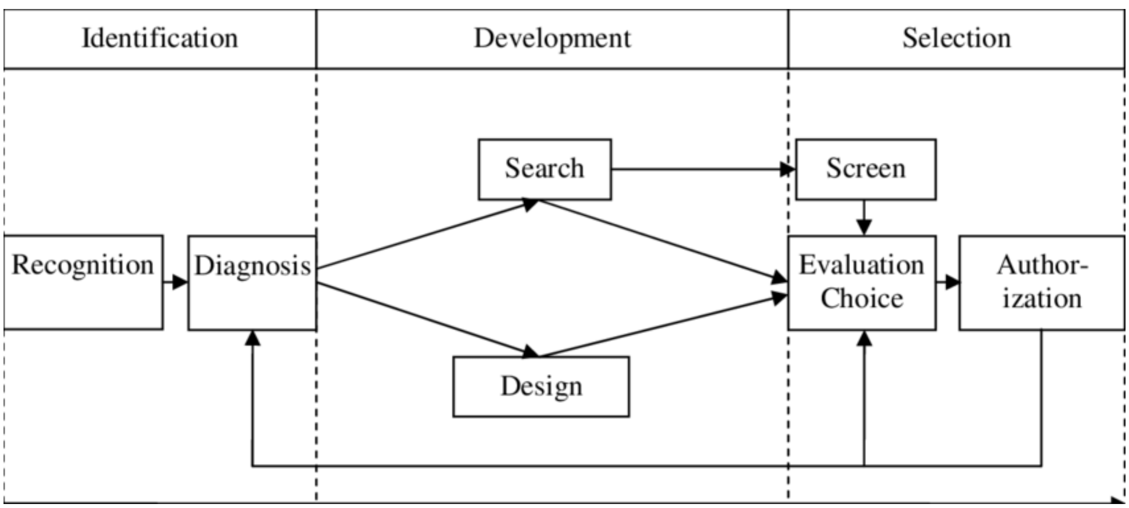

Figure 3: Strategic decision model (Mintzberg et al., 1976; Kask, 2010)

Mintzberg et al. (1976) states that business decisions are based on three steps. These steps are identification, development and selection. Identification is focused on recognizing the need for a decision. A diagnosis follows the recognition phase. It is important to have a defined and clarified overview of the need for a decision (Mintzberg et al., 1976). The second phase is to search for already existing solutions for the decision. If there is not an available solution pre-existing solution needs to be modified or the design department needs to provide a new decision (Mintzberg et al., 1976). The last step is selection, and this phase is based on evaluate and screen the different solutions. If there is no good solution after the authorization phase the organisation needs to go back to step number one (Mintzberg et al., 1976). The model of Mintzberg et al. (1976) shows that solutions needs to be analysed with care and that the problems needs to be clear for the organisation. Performance dashboards can help organisations to get a good visual overview of the data and can help to create scenarios to form different solutions.

3.1 Situational awareness

Managerial decisions are based on information. Nevertheless, it is not easy to see all information. Situational awareness is about understanding the surrounding. What is happening around you and what could happen soon. This can be physical, for examples with cars and bikes, or non-physical in means of data and information (Gonzalez and Wimisberg (2007). Wickens (2000) describe situational awareness as the ability to extract and integrate information in a continuously changing environment and the ability to use this information to guide future actions. This is related to the amount of information available right now. The literature divides situational awareness in three different levels (Endsley and Garland, 2000). The figure below shows the model of Endsley (2000). The model can be used to make decisions. The model consists of situation awareness, individual factors and task and environmental factors.

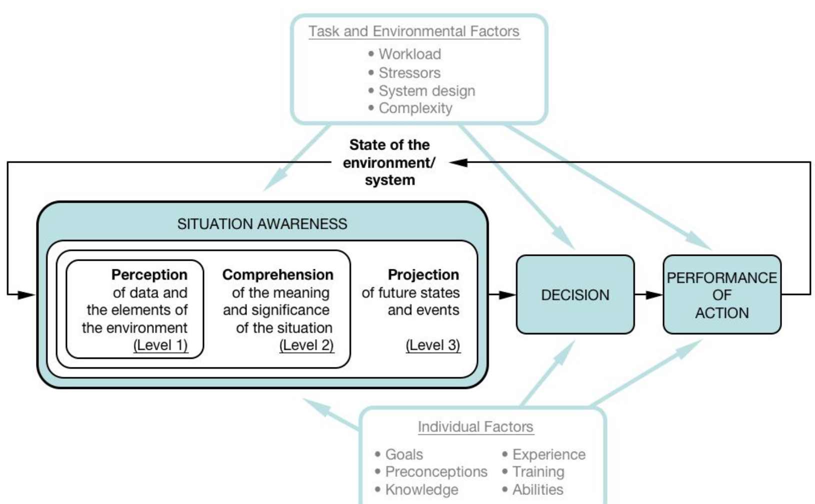

Figure 4: Endsley and Garland decision model, 2000

The first component is perception, the possibility of seeing information, cues and other situations in the environment (Endsley and Garland, 2000). This can best be described with seeing different words (getting data). The second one is comprehension; this is not solely about the perceptions of activities in your environment but also understanding and linking this information (Endsley and Garland, 2000). This can best be described with sentences, combing the words and make a meaningful sentence (transforming data into knowledge). The last level is projection, this means that you can use the information to anticipate future events (Endsley and garland, 2000). This can best be described as using the sentence to know what will happen in the future (using the information to get knowledge). A business diagnosis is focused on giving the most important information in an organisation. Business diagnosis can use a performance dashboard. A performance dashboard tries to reduce the amount of available data in a one-page overview. A performance dashboard can be used to improve the business diagnosis by focusing on the most important data and information and therefore improve the decision of the management board. As stated before, it is important to not only see all the information but also link this information and predict the future. Performance dashboards can connect to all the available data in an organisation transform this data into useful information. This information can be used to predict the future and therefore improve the decision making within the organisation. A performance dashboard is situational aware and can extract and integrate information in a continuously changing environment and uses this information to guide future actions.

3.2 Information overload

The first level of situational awareness is about perceiving the information. Getting the information isn't a problem anymore. The amount of information increases every year, but this does not lead to an increase in better decisions (Adeoti-Adekeye, 1997). Almost every decision is made with the available information (Adeoti-Adekeye 1997). There are employees with a limited capacity to process information (Owen, 1992). These employees are unable to cope with the available information (O'Reilly, 1980). This is called information overload. This is critical in a world where there is more and more available information. In the literature there are different names for this phenomenon, such as cognitive overload (Jacoby 1984; Kirsh 2000, Potter and Choi 2006) or information fatigue syndrome (Wurman, 2001). The article of Lee and Lee (2004) states that information overload leads to less satisfied and confused consumers of information and therefore bad decision making. This thesis will have the following definition of information overload: *"the notion of receiving too much information too quickly"* (Kutty et al., 2007). In the article of Eppler and Mengis (2003 and 2004), the authors studied articles focused on information overload from the past thirty years and gave an overview of definitions of information overload, causes, the effects, measures and counter measures. The basis of the article of Eppler and Mengis (2004) is about five factors which causes information overload problems. The factors are the following: personal factor information (PFI), information characteristics (IC), tasks and processes (TP), organizational design (OD) and information technology (IT). Eppler and Mengis (2004) note countermeasures for the five factors. These countermeasures are placed in the table below.

Construct	**Countermeasures**
Personal factor information (PFI)	Time management skills and techniques, training programs, processing skills (file handling and using e-mail), documents classification, priority setting and screening skills.
Information characteristics (IC)	Quality of information, value-added information; communication design (e.g., e-mail etiquette), compress, aggregate, categorize, structure information, brand names for information, customisation, intelligent interfaces and interlink various information.
Task and processes (TP)	Standardize procedures, define decision models, exception-reporting system, allow more time for task performance, selection of media, handle incoming information, collaboration with information specialists, qualitative information, regulate the rate of information flow and measurement system.
Organisation design (OD)	Coordination, processing capacity, creation of lateral relationships (integrate roles, create liaisons between roles, teamwork etc.), goal setting, creation of self-contained tasks (reduced division of labour, authority structures), reduce divergence, install appropriate measures of performance, hire additional employees and create slack resources.
Information technology (IT)	Intelligent information management, prefer push to pull technologies, facilitator support through (e-)tools, DSS, search with artificial intelligence, intelligent data selectors and filling systems

Table 4: Information overload factors and countermeasures (Eppler and Mengis, 2004)

Because this thesis is focused on the intersect between decision making and information technology not every factor of Eppler and Mengis (2004) is related. The three factors that play a role in this design study are information characteristics (IC), task and processes (TP) and information technology (IT). Information characteristics is about the information. This factor is focused on the quantity, quality and frequency of the information (Kutty et al., 2007). One of the countermeasures to prevent information overload is to make the information visualized and have aggregated steps (Ackoff, 1967; Meyer, 1998). This relates to the need for visualized performance dashboard that give a clear overview of the available information. The factor, task and processes, is focused on the standardisation procedures of information (Bawden, 2001). It is about defining rules and having a decision model (Chewning and Harell, 1990), handling the incoming information in collaboration with specialists in information technology (Edmunds and Morris, 2000). Qualitative information is the basis for a good information flow (Grise and Gallupe, 2000). This factor is about the collaboration between the business and the IT side to make sure the right information is processed and used. This is something important for performance dashboards because the right information is needed to make good decision in an organisation. The last important factor is information technology. This is about the technologies that can be used to prevent information overload (Kutty et al., 2007), such as artificial intelligence. Making intelligent machines that can influence the decision making or can process information without the use of humans (Bawden, 2001). Furthermore, intelligent information management decision support systems and push and pull technologies can be used to prevent information overload. This is something that can be added to performance dashboards. With the help of artificial intelligence, it is possible to make automated business diagnoses based on the available information in organisations.

Manager use information before making decisions (Cook, 1993). This needs to be done with good care to prevent information overload. The article of Kutty et al. (2007) shows 80 percent of the decisions are made alone. The factor task and processes show that decisions need to be made in collaboration and that there needs to be a decision model. The three factors above can help manager to reduce the 'noise' in the information stream. Performance dashboards can be used to reduce this noise by visualize the most important information and give a clear overview of the available data. It is important that a business diagnosis is based on the right information that is understood by the business and it side of the organisation. A performance dashboard plays an important role in this process by providing the right information in a clear overview that filters the ‘noise’ out of the information stream.

3.3 Individual, task and environmental factors

The last part in the model is about different external factors of information. Making decision is not only based on the amount of understandable information. The workload, stress and complexity in an organisation also have influence on the situational awareness and decision making (Endsley and Garland, 2000). When the level of complexity is high and the worker has a high level of stress it is harder to cope with information (Endsley and Garland, 2000). Therefore, it is useful to aim for a low level of complexity and workload in an organisation and hire employees that can deal a big amount of information. It is possible to reduce this complexity and workload with supportive tools that can cope with huge amount of information. Performance dashboards can be a solution for this factor.

Implementing a new system, obligated use of performance dashboard or another new tool to reduce the complexity and workload can lead to a higher workload and complexity if there is no support (Endsley and Garland, 2000). An employee without experience, support and training in a new system will have more workload and complexity because the system requires new steps instead of automaticity. This is an example of individual factors in the decision-making process (Endlsey and Garland, 2000). Individual factor contains experience, training, goals, abilities, knowledge and preconditions. Training and knowledge will improve the ease of use of the new system and lead to the ability to gain experience in the new system or tool (Endlsey and Garland, 2000). This is not related to a performance dashboard but needs to be addressed because using a performance dashboard to create a business diagnosis is something that is not widely used. Training, support, goals and preconditions are needed to improve the knowledge about business diagnosis based on performance dashboards to prevent information overload and improve the situational awareness within the organisation. This will lead to improved decisions. The dashboard needs to be easy to understand.

4 AUTOMATED BUSINESS DIAGNOSIS

Chapter two (the evolution of performance dashboards) and chapter three (managerial decision making) are related to use data to improve the decision making. Combined they can create business diagnosis of the current situation and predict future events. Business diagnosis are focused on the business model of an organisation (Feelders and Daniels, 2001). This chapter explains the concept of explanatory business diagnosis (4.1), how to automate this process (4.2) and the relation between the automated business diagnosis and a performance dashboard (4.3).

4.1 Explanatory business diagnosis

An organisation shows particular behaviour and provides data. Data itself cannot provide a business diagnosis. A diagnosis is an explanation of the observed behaviour of an organisation (Feelders and Daniels, 2001). This means that a business diagnosis is focused on linking the causes and effects of a particular behaviour. Organisations focus on irregular or negative behaviour and tries to solve this behaviour. A business diagnosis tries to explain this behaviour. Explanations are based on rules that express the relation between events. Every business model exists of qualitative and quantitative relations (Feelders and Daniels, 2001). Quantitative relations are based on equations. For example, profit can be calculated with extracting the costs from the revenues (Feelders and Daniels, 2001). The relation in qualitative relations is unknown. Therefore, the focus in this research is on quantitative relations. Saariluoma (2005) states that explanatory is based on the idea that the problems needs to be related to the right knowledge to provide solutions. With the right knowledge it is possible to have solutions for the problems. The picture below is an explanatory framework (Saariluoma, 2002 and 2003).

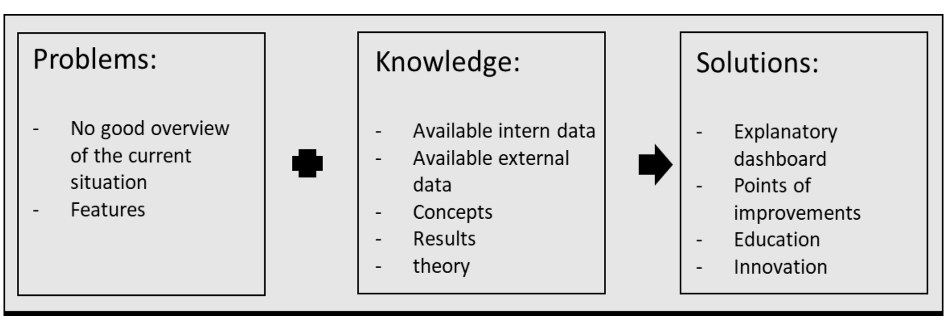

Figure 5: Explanatory framework (Saariluoma, 2002 and 2003)

This framework can be used to explain and solve problems with the help of a unified system of theoretical knowledge (Saariluoma, 2005). This framework is based on the behaviour of an organisation. The first step is to identity irregular behaviour and try to link the causes and the effects. This is called the problem(s). In multiple situations there is enough knowledge within the organisation to provide a solution that is good enough. The data, theoretical papers, concepts and results can be used to solve the problems. Education and innovation increase the change of having the right knowledge and therefore the solution for the problem(s) (Saariluoma, 2002). Performance dashboards can play an important role in this process by providing the available data in a systematic way and show irregular behaviour of the organisation. Furthermore, it is possible to automate this process to reduce the time for analysing the solutions. This is explained in section 4.2.

4.2 Automated business diagnoses

This section describes business diagnosis and the need to automate this process. It is possible to make (partly) automated business diagnoses. This form of diagnosis's can be beneficial for organisations because it saves time and costs. The reason for this is that less time is spend to activities that have zero value to the organisation (finding irregular behaviour), and more time is spent to solving the problem(s). There are multiple articles about automated business diagnosis. This research focuses on the article of Feelders and Daniels (2001) because it concludes a model that can be used in a performance dashboard. This model forms the basis of this chapter. The formalisation process of automated business diagnoses are a combination of artificial intelligence and operations research. As stated in 4.1, a business diagnosis is an explanation of the observed behaviour of an organisation. The article of Feelders and Daniels (2001) is focused on a causal model of explanation of an event. An event is the value of a variable that changes from time to time (t to $t^{\wedge t}$), in this case variable y. Because a business diagnosis is focused on the difference between irregular behaviour over time, time needs to be included. The behaviour is explained by the causes of this behaviour. An event occurs with the following explanation:

$$< a, F, r > \text{ because } C^{+}, \text{despite } C^{-} \quad (1)$$

The first part of the explanation contains the letter *a*, *F* and *r*. Hesslow (1984) states that all explanations are based on an object (*a*) with a property (*F*) and a reference class (*r*). For example, object *a* is the local supermarket with the property *F* of having a low profit margin and a reference class that includes the competitors of the local supermarket. There are multiple ways to create a reference class. Feelders and Daniels describes three different reference classes. This research uses the temporally normal case reference class. The focus is on the temporal difference between the actual state and a certain state in the past (t to $t^{\wedge t}$). For a business diagnosis of the performance of an organisation it is important to focus on the difference between the actual and norm behaviour.

a = the actual behaviour of an organisation
F = a variable that deviates from the norm value
r = the norm behaviour of the organisation

The second part of the explanation contains the influences on an event. The articles highlight the notion of aleatory explanations of Humphreys (1989). This notion states that there are two causal influences on an event, contributing (C^{+}) and counteracting (C^{-}) (Humphreys, 1989). Counteracting causal influences are a possible empty set and got a negative impact on the event (Feelders and Daniels, 2001). Contributing causal influences is a non-empty set and got a positive impact on the event of the organisation. A business diagnosis aims to find the causes that have the highest influence on the behaviour of the organisation.

This event (*y*), the difference between the actual and norm behaviour, occurred despite the presence of, the earlier mentioned, counteracting (C^{-}) and contributing (C^{+}) causal influences. The set of counteracting and contributing causes consists of the components x_i of $\mathbf{x}$. This can be formulated in the following formula:

$$inf\ (\ x_i, y\ )\ *\ \Delta y\ >\ 0 \quad (2)$$

This difference can have a high, normal or low value based on the difference between y^a and y^r.

	σy
$y^a > y^r$	High
$y^a = y^r$	Normal
$y^a < y^r$	Low

Table 5: *Mapping to qualitative value*

Furthermore, it is important to create an influence matrix to see the effect of the causes on the event. To calculate the measure of influence this research used the following formula:

$$\inf\ (x_i, y) = f(\boldsymbol{x}^r_{-i}, x^a_i) - y^r \quad (3)$$

This measure of influence can be simplified if the function *f* is additive to:

$$inf\ (x_i, y) = x^a_i - x^r_i \quad (4)$$

This formula uses all the same values except for one (x_i). This formula shows the difference between the actual and norm value if only x_i would deviate from its norm value. Filtering is an important part for a business diagnosis because there is a lot of data in organisation. This amount of data can be too much for managerial decision making and can create information overload (3.2). Feelders and Daniels (2001) use a parsimonious set of contributing and counteracting causes for the filtering process.

$$\frac{\inf\ (C^+_p, y)}{\inf\ (C^+, y)} \geq T^+ \quad (5)$$

This parsimonious set is one of the options to filter the causes but there are different options for the filtering process. It is possible to have a minimal percentage of impact on the event. For example, that the highest causes are listed until a percentage of 50% is reached. It is also possible to select the *n* highest number of causes. Another option is to filter the causes per dimension. For example, filter the causes based on the time, location or product dimension. In this dimension it is possible to filter on quarter to month to week. It is possible to use other filter options, but that depends on the needs of the organisation.

The article of Feelders and Daniels (2001) is focused on a business model and can be summarized into six steps. These six steps are necessary to make automated business diagnosis of organisations based on quantitative relations. The six steps are:

1. determine the actual data (EQ 1 and 2);
2. determine the reference data (EQ 1 and 2);
3. get model relations from star scheme;
4. compute influence of reference data to determine causes (EQ 3 and 4);
5. filter causes to avoid information overload (EQ 5);
6. visual explanation tree of causes.

The first two steps can be changed with logistic steps. Data can be based on normalised, absolute, scaled or non-scaled data. Furthermore, a sixth step is added to visualise the causes in an explanation tree. This explanation tree gives a good overview of the most important causes

4.3 Explanatory business diagnosis and performance dashboards

The previous section gives an overview of how to make automated business diagnosis based on the literature (Feelders and Daniels, 2001). This research is based on creating automated business diagnosis within a performance dashboard to improve the decision making of the management board. The basis of a performance dashboard is a data warehouse and a data scheme for this data. The literature states that there are different data warehouse models and schemes. For example, an entity relationship model, dimensional model, star scheme, snowflake scheme and star flake scheme (Sulaiman and Yahaya, 2013). The two common models for data warehouse are entity relationship and dimensional model. Dimensional modelling is focused on measures, facts and dimensions. Dimensional modelling is simpler, more powerful in representing the requirements of the business user in the context of database tables and easier to interpret compared to entity relationship modelling (Güratan, 2005 and Ballard et al., 1998). This research is focused on the dimensional modelling because the data is dimensional based. There are different ways to represent multidimensional schemes such as star, snowflake and a combination of both (Sulaiman and Yahaya, 2013). This research focuses on the star scheme because this scheme is widely accepted and the available data is star scheme based (Sulaiman and Yahaya, 2013). A star scheme contains at least one fact table and multiple dimensions tables. This star scheme is based on relations that influence the behaviour of an organisation. The counteracting and contributing casual influences of Feelders and Daniels (2001) are based on the structure of the star scheme. Model relations got influence on the behaviour of the organisation. These relations cause a change in y.

4.3.1 Feelders and Daniels (2001) in performance dashboards

Step (1) is based on the actual data (a). The actual data is one of the components of EQ (1). These components are necessary to create an automated business diagnosis. The actual data is based on the front-end of a dashboard. It is possible to point out a cell in the star scheme. For example, a = sales (2019, y). The data can be seen and retrieved. This can be done with Python and Power BI. If the data is in a star scheme it is possible to apply measures in a performance dashboard. These measures can be used to calculate the behaviour of an organisation (a). For example, profit consists of different measures. Profit is the revenues of a given year in a given region minus the costs of that year in the given region. All these variables got different dimensions to get more in-depth information of the organisation. The actual data can be highlighted in Power BI to calculate the behaviour of the organisation. There is also a difference between absolute and relative data. This can play a big role in generating automated business diagnosis. Absolute data is based on the numbers you can see directly. For example, the profit is 638.000 or the organisation is selling twelve products. Relative data is based on this absolute data. For example, the profit is a quarter of the revenues or the number of products depends on the sales of last year. In some cases, it would be better to focus on relative data because absolute data can give a wrong image of an organisation. It is possible that an organisation got 50 shops in the Netherlands and 4 shops in Finland. The revenue based per country would be an unfair indicator to compare the two countries. An overview of the average profit per shop per country would give a good insight of the two countries.

Step (2) is based on the reference class (r). This step is still based on the front-end of a performance dashboard. There are two methods to determine the reference data. The first method is to take the average of all competitors. Take the annual reports and make a norm value based on all competitors. This gives a good overview of the position of the organisation compared to the competitors. For example, r = sales (2019, org1/org2/org/3) divided by the number of organisations. Nevertheless, it is not always possible to find suitable data from organisation in the same branch due to legal restrictions or other factors. The second option is to use the data from the organisation as reference data. It is possible to point out the same cell but use the data from other years. For example, r = sales (2018, 2017, 2016). This data can be combined and divided by the amount of years to get a reference class. In this research the second option is used. The data of this research is based on one organisation and not every annual report is available of its competitors.

Step (3) is focused on model relations. Model relations are necessary for the model of Feelders and Daniels (2001). The model will not work without these relations. The model relations are in the back end of the performance dashboard. The star model consists of different relations. There is at least one central fact table that is related to dimensional tables. For example, the fact table can be related to time, region, product, promotion, inventory etc. These tables are all related in different ways. It is possible to have one-to-one, one-to-many and many-to-many relationships. There are two types or relations in a star scheme. Dimensional hierarchy and organisational measures. Dimensional hierarchy is based on the hierarchy within a dimension. For example, sales (2019) = sales (2019.Q1) + sales (2019.Q2) + sales (2019.Q3) + sales (2019.Q4). Every quarter exist of months and the months consists of weeks. There are multiple levels of relations in one dimension. The other type of relation, organisational measure, is focused on measures within the fact table. For example, profit (2019) = revenues (2019) – costs (2019). This is just an example of one of the many relations in a performance dashboard.

Step (4) is based on computing the influence of the reference data to determine causes. This is an important factor because a business diagnosis is based on finding the causes for irregular behaviour. A difference between y^a and y^r is based on contributing (C^+) and counteracting (C^-) causes. Because the model relations are available in a performance dashboard it is possible to use additive measures (count and sum) and aggregations to make interactive analyses. This can be used to calculate an influence matrix based on equation (3) and (4). This depends on the complexity of the data. Additive data can use EQ (3) and when the data is more complex EQ (4) is used. This influence matrix will list all the causes that influence y^a. An influence matrix can be created in a performance dashboard with a Python script. The star scheme-based data needs to be transferred to a data frame and the EQ (3) and (4) create the influence matrix.

Step (5) is based on filtering the data. There is too much data available to take care of every causes that influence the behaviour within an organisation. This relates to the information overload concept. Data needs to be reduced to a certain level to prevent information overload so that data analysts and the management board can focus on the important information of data. Although this can be done with EQ (5), this research will use another approach for the filtering process. As stated in section 4.2 the filtering can also be done by setting up rules in a performance dashboard. For example, the seven highest influence percentages based on the influence matrix or select all causes from high to low until a minimum of 80 percent is reached. This filtering can be done within the Python data frame. This can be done with the influence matrix of Feelders and Daniels (2001) (EQ 2). Rank these influences from high to low and add rules to filter the causes.

Step (6) is focused on presenting the causes in a visual and textual way. This can be done with a Python script and an explanation tree. This tree gives a visual overview of the most important causes. Colour can be added to create differences between the cases. This can be done with different tools within the performance dashboard. It is possible to generate an explanation tree with the decomposition tree, but more information will be following about this tool. The text message can be created with a Python script. Section 6.2 gives more information about Python scripts.

4.3.2 Numeric example

Step (1) for using a performance dashboard for automated business diagnosis is based on EQ (1). This numeric example uses a practice dataset from the R.S.G. Pantarijn in Wageningen. This dataset is used in Economic courses and can be requested from the school or the researcher of this study. Organisation Z is an organisation that repairs broken bikes. The profit of organisation Z in 2019 is 7.495 and in 2018 the profit was 18.915. Z wants to have a business diagnosis to see which weeks had the most influence on this difference. First the actual data (a) needs to be determined. It is possible to point out a cell in the star scheme. For example, a = profit (2019, org Z). In this example the data is absolute and non-scaled because it is a one-man shop with absolute data. In the table below the eight lowest value of profit in 2019 are displayed ranked by week. The appendix gives an overview of all the other weeks of organisation Z in 2019.

Week	***Profit***	***Week***	***Profit***
4	-4.320	18	-3.545
5	-3.230	48	-725
6	-1.890	51	-5.210
17	-1.250	52	-5.250

Table 6: Profit of organisation Z

Step (2) is based on determining the reference data (r). EQ (2) states that the reference data can be determined based on pointing out a cell and year. In this case the focus is on profit. In this numeric example this would be r = profit (2018, org Z). The table has a new format by adding the reference data. In the table below the eight lowest value of profit in 2019 are displayed ranked by week. This filter can be applied in Power BI. The data can be seen and retrieved in the following table.

Week	***Profit 2019***	***Profit norm***	***Week***	***Profit 2019***	***Profit norm***
4	-4.320	-4.625	18	-3.545	-765
5	-3.230	-2.890	48	-725	-755
6	-1.890	-2.125	51	-5.210	-5.475
17	-1.250	400	52	-5.250	-4.980

Table 7: Profit of organisation Z of 2019 and norm

Step (3) is about the model relations. There are two types of relations in a star scheme. Dimensional hierarchy and organisational measures. Profit consists of organisational measures. Profit = revenues – costs. An example of hierarchical relation is revenue = revenue.Q1 + revenue.Q2 + revenue.Q3 + revenue.Q4 and costs = cost.Q1 + cost.Q2 + cost.Q3 + cost.Q4. There are more organisational measures and levels in dimensional relations. These relations form the basis for the next steps in formulating an automated business diagnosis.

Step (4) is focused on computing the influence of reference data to determine causes. For this step EQ (3), (4) or a combination is used. The focus in this case is on absolute data because it is focused on the whole organisation and it is based in one country with the same amount of shops and time. Because additive measures are used it is possible to select EQ (4). In this example, the equation uses all the reference data expect for one week. It is also possible to see the influence percentage with this formula. This influence percentage shows how much the y value derive from its actual value when one-week changes. Step (4) is focused on contributing (C^{+}) and counteracting causes

(C^-). Contributing causes got positive effect on the behaviour and counteracting causes got negative effect on the behaviour. In this case the profit decreased, this means that causes that explain the decrease are contributing and causes that have a positive effect on the profit are counteracting. If the profit showed an increase this would be the other way around. The table below gives an overview of the influence per week.

Week	***Profit 2019***	***Profit norm***	***Inf***	***Week***	***Profit 2019***	***Profit norm***	***Inf***
4	-4.320	-4.625	1.61%	18	-3.545	-765	-14.70%
5	-3.230	-2.890	-1.80%	48	-725	-745	0.16%
6	-1.890	-2.125	1.24%	51	-5.210	-5.475	1.40%
17	-1250	400	-8.72%	52	-5.250	-4.980	-1.43%

Table 8: Profit of organisation Z of 2019, norm and influence:

The table has a new format by adding the influence column. In the table the eight lowest value of profit in 2019 are displayed ranked by week. The table doesn't give a good overview of the important weeks because the focus is on the profit instead of the influence percentage. Week 4, 5, 6, 51 and 52 got a low influence percentage and are not the main reason for a decrease in profit for organisation Z. The next step will apply a filter to solve this problem.

Step (5) is focused on filtering the causes to prevent information overload and increase situational awareness. Table 8 shows a part of the full table of organisation Z. The owner of the shop doesn't need to work on weeks with a low influence percentage because those weeks are not the reason for a decrease in profit. There are multiple ways for filtering. For example, using EQ (5), based on the *n* highest value or until a certain percentage is reached. In this example the *n* highest numbers filter is used. This can be done with a Python script. The number of hits is *n*, and in this case four. Df.round is about showing the decimals for the influence percentage, in this case two decimals are chosen. A new data frame is created (df1) to show a new table, df.nbiggest(number_of_hits, ['variable one', 'variable two']. This gives the user the ability to filter based on the influence percentage and use a second variable to rank if the influence variable is the same.

Number_of_hits = 4
df = df.round(2)

df1 = (df.nbiggest(Number_of_hits, ['Influence','Profit Norm']))

print(df1)

This code is used two times, one time for the highest and one time for the lowest percentage. The left side of the table displays the four lowest values and the right side displays the highest influence value. The table below gives an overview of the results with filtering:

Week	***Profit 2019***	***Profit Norm***	***Influence***	***Week***	***Profit 2019***	***Profit Norm***	***Influence***
9	-700	3.459	-21.99%	4	-4.320	-4.625	1.61%
18	-3.545	-765	-14.70%	51	-5.210	-5.475	1.40%
17	-1.250	400	-8.72%	6	-1.890	-2.125	1.24%
39	245	650	-2.14%	44	1055	925	0.69%

Table 9: Filtered causes organisation Z

This table gives a better overview of the most important weeks for organisation Z. This table does not have the same weeks as the other tables. Week 9, 39 and 44 are new and replaced week 5, 48 and 52. This is the strength of focusing on the influence matrix instead of absolute numbers. There are counteracting causes and contributing causes that influence the profit of organisation Z that are hidden in the data. Using the filter in both ways gives the user the possibility to see both.

Step (6) is based on making the causes visual. This can be done with an explanation tree. The picture below shows a possible way for this step. This can be done in Power BI. The counteracting and contributing causes are separated to see the worst and the best weeks. Red means the most important week and yellow the least important week in the top four.

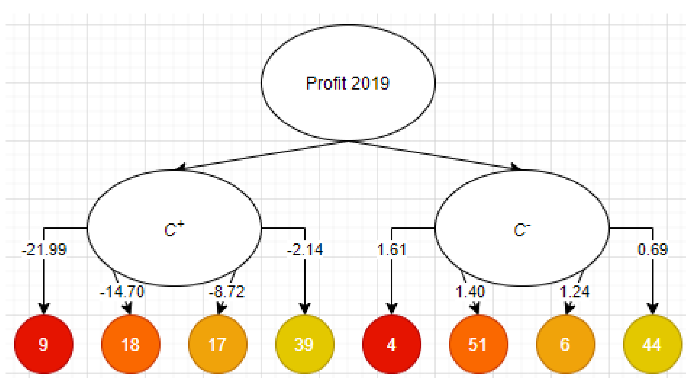

Figure 6: Explanation tree of organization Z

The next step in creating an automated business diagnosis is to use a Python script in a performance dashboard to not only give a visual overview but also a textual. As stated earlier, text is needed to make data understandable throughout the whole organisation. The figure below shows an automated business diagnosis. Section 6.2 describes this process in detail.

All zero values are removed	All zero values are removed
After filtering the most important *C+ w*eek is 9 with a profit difference of 4.159 and an influence percentage of -21.99%	After filtering the most important *C- w*eek is 4 with a profit difference of 305 and an influence percentage of 1.61%
After filtering the most important *C+ w*eek is 18 with a profit difference of 2.780 and an influence percentage of -14.70%	After filtering the most important *C- w*eek is 51 with a profit difference of 265 and an influence percentage of 1.40%
After filtering the most important *C+ w*eek is 17 with a profit difference of 1.650 and an influence percentage of -8.72%	After filtering the most important *C- w*eek is 6 with a profit difference of 235 and an influence percentage of 1.24%
After filtering the most important *C+ w*eek is 39 with a profit difference of 405 and an influence percentage of -2.14%	After filtering the most important *C- w*eek is 44 with a profit difference of 130 and an influence percentage of 0.69%

Figure 7: Automated text message for organisation Z

5 POWER BI AND DECOMPOSITION TREE

This chapter focuses on the decomposition tree. Because it is a new feature there is not much information about it yet. Two blogs of Power BI are used to gather more information about this new tool. The blog of Cofsky (2019) and an anonymous blog (2019). In 2019 Power BI released features for the November 2019 update (Cofsky, 2019). The decomposition tree can be used to create a tree as below.

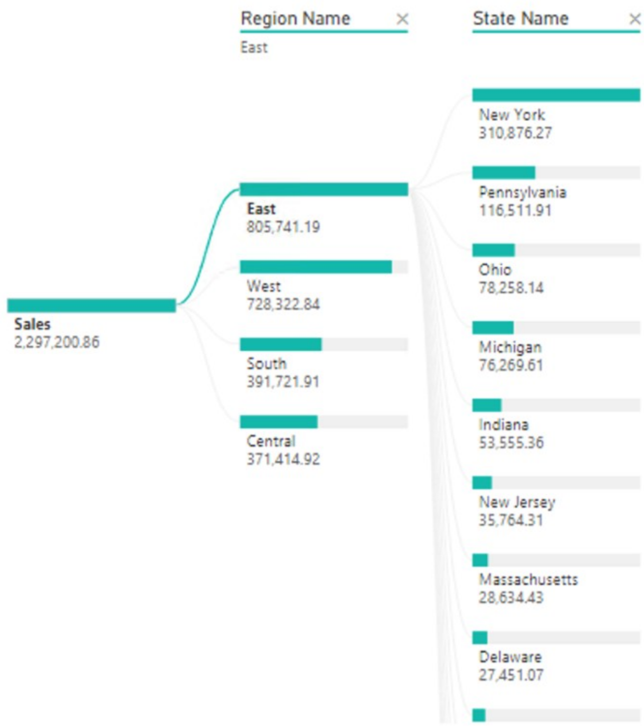

Figure 8: Decomposition tree

The decomposition tree is an analysis tool that uses the root-cause analysis by viewing how individual categories contribute to the whole. The decomposition tree gives the user the ability to decompose or divide a group in order to show the individual categories. It is also possible to see how these categories can be sorted according to a selected measure. In this case sales. This tool can make predictions and simulations to improve the decision-making process. The tool got the ability to determine the data in absolute or relative way. It is also possible to visualize data across several dimensions. The decomposition tree automatically aggregates data and gives the user the ability to get in-depth information into dimensions. Furthermore, the drill down function shows the taken path, from the first drill down to the last. This is new compared to other visualisation tools, where a drill down created a new graph or table with no information from the previous drill downs. Another option is the artificial intelligence split. This AI split can be used to find the next dimension based on specific criteria. This is one of the reasons why this tool could be useful for ad-hoc exploration, predicting, simulating and root cause analytics.

There are also some negative sides of the decomposition tree. There is maximum of 50 levels in the tree. This isn't a big problem at first, but in some cases more than 50 level are needed. Furthermore, the tree cannot be used if the data of an organisation contains more than 5000 data points. This tool does not support on-premises analysis services, Power BI Report Server, Direct Query and Azure Analysis Services. The decomposition tree is still a preview feature and therefor it is not possible to use this tool on the mobile application, show data functionality and pinning to the dashboard. Because it is a preview feature while writing this article it is possible that Power BI will change some features in the upcoming releases.

At first sight this looks like a tool for business diagnosis, finding root causes that have influence on a certain behaviour. This chapter uses the different steps from Feelders and Daniels (2001) to see the advantages and flaws for using this visual for automated business diagnosis. Can this decomposition tree replace this research or is it something different?

5.1 Decomposition tree versus explanation tree

The basis of the article of Feelders and Daniels (2001) is equation (1). This equation states that a behaviour occurs despite the effect of contributing and counteracting causes. Furthermore, the article formulates six different steps in creating an automated business diagnosis (see page 53). These six steps are used to find out if the decomposition tree is suitable for creating an automated business diagnosis.

Step (1) is based on EQ (1), determine the actual data (a). The actual data is based on the front-end of a dashboard. The decomposition tree uses the same data structure as other visual tools. This means that the star scheme is still the basis. It is possible to point out a cell in the star scheme. For example, a = sales (2019, orgY). The data can be seen and retrieved. It is possible to decompose this variable into smaller variables that explain the first variable. Figure seven is an example of this feature. A drill down on sales shows the different regions. Furthermore, an AI split drill down can be used to find the next dimension based on specific criteria. Step (2) is based on another component of EQ (1), the reference class (*r*). This step is also based on the front-end of a performance dashboard. The process is the same as step (1) but instead of the current year, another cell is selected. This will give the same data about another year compared to the actual value. If the same drill down functions are used it is possible to see the difference between the current year and the chosen year. It is not possible to see the difference in one tree. For this function two different trees are needed, because it is not possible to choose two routes in the tree next to each other. It is also possible to create a reference class based on different years or data from the competitors by adding measures in the Power BI software. This gives the opportunity to get a reference class based on multiple years or the competition.

Step (3) is focused on the model relationship in the decomposition tree. The decomposition tree uses the same relations as all the other tools in Power BI. Because the data is still in the star scheme there are two different relationships. Hierarchical relations and organisational measures. Hierarchical relations consist of getting more in-depth information within a dimension. For example, the quarter in a year. Organisational measures are measures based in the fact table. By connecting different measures, it is possible to calculate the profit per month. The reference class with multiple years uses hierarchical, multiple years of the total profit, and organisational relations. Because multiple dimensions are needed to calculate the average profit of multiple years. The new AI split consists of new relationships. The AI split can be used to find the next dimension based on specific criteria. This is based on machine learning. The AI split work in two ways, high and low value. With the high value Power BI considers all available fields and determines which variable got the highest value of the measure being analysed. Power BI drill down into this variable. The low value works the same except Power BI determines the variable with the lowest value. It is possible to use the AI split multiple times. The relations behind this AI split are not clear yet. It is a black box right now. The reason for this is that Power BI did not gave information about it and it is still in the preview phase.

Step (4) is focused on computing the influence of reference data to determine causes. For this step EQ (3), (4) or a combination of both is used. This is something that is not possible in the decomposition tree. It is not possible to use EQ (3) or (4) to change one week in the actual data to see the effect on the variable. It is possible to break up your data but not changing data or compare different sources in one tree. A solution for this problem would be to adjust the data frame to make

extra columns with the influence per week. This needs to be done with measures in Power BI. This is a risk because it can influence other data in the data frame. New relations need to be made and dimensions are adjusted with measures in Power BI.
Step (5) is based on filtering the causes. Nevertheless, as stated in step (4), it is not possible to use the reference and actual data in the same tree. This is a problem for step (5) because this step filters the causes. The filtering can be done with the decomposition tree. It is possible to select a variable and list these filters from high to low. Although, it is not possible to filter the causes the decomposition tree got the ability to filter variables.
Step (6) is based on visualize the data. The decomposition tree can do this by filtering the data. Nevertheless, without step 4 and 5 it is not possible to filter the most important causes. Other tools in Power BI are better to visualize the data in an explanation tree compared to the decomposition tree.

The decomposition tree is a useful tool to get a cross functional overview of your data. It gives an organisation the ability to quickly see the causes and the influence on a variable. The AI split can be used to find the next dimension based on specific criteria. The decomposition tree can be used to determine the actual data, reference data and the model relations are partly clear. Nevertheless, it is not possible to use the equations from Feelders and Daniels (2001) in the decomposition tree. Because these equations are essential for the last two steps it is not possible to determine the causes based on the measure of influence. The decomposition tree is a drill-down tree with no relation between the actual and reference data. Furthermore, it is not possible to use this extension to generate textual business diagnosis.

This new tool can play a supportive role. The decomposition tree makes it easy to decompose data. It can be used to get more in-depth information after the business diagnosis. The results of section 4.3.2 can be analysed with the decomposition tree. The AI split can be used to analyse a specific week. Nevertheless, the AI splits are not supported by every server yet. For example, Azure Analysis Services, Direct Query, Power BI Report Server, web application and complex measures and measures from extensions schemas in analyse. Other tools are needed for creating an automated text message based on the data and the influence matrix needs to be made with a tool that supports Python. Therefore, this tool can be useful for organisation. Nevertheless, in this case, it can only be used to get in depth analyses after the causes are determined and filtered with another tool. The decomposition tree can play a role after the automated business diagnosis. The table below gives an overview of the decomposition tree and the explanation tree.

Step	***Decomposition tree***	***Explanation tree***
1	Selecting a cell to determine the actual data. Front-end based, easy to select a cell	Selecting a cell to determine the actual data. Front-end based, easy to select a cell
2	Pointing out the same cell but with a different year. The possibility to generate a reference class based on different years is possible with the help of Power BI (hard).	Pointing out the same cell but with a different year. Calculating the reference class based on different years is possible with help of Power BI (hard) or Python (easy).
3	New tool with a black box AI split. Hierarchical and organisational relations are clear because the data is star scheme based. Relations can be added within Power BI, this can have negative impact on other relations.	Hierarchical and organisational relations are clear because the data is star scheme based. Relations can be added within Power BI, this can have negative impact on other relations.
4	Not possible until all relations are clear, and the decomposition tree has the possibility to combine the reference and actual data. The decomposition tree is a drill-down tree with no relation between actual and reference data.	Determining the causes is possible by using a Python script. EQ (3) and (4) can be used to generate an influence matrix within a performance dashboard.
5	Possibility to filter is good. Ability to show the highest value is possible. Not every filter process is possible because it is not possible to generate Python scripts. It is also not possible to generate text in the decomposition tree.	Filtering can be done based on EQ (5) or other filter options. It is possible to use Python to apply scripts and generate an automated business diagnosis based on text and data.
6	Decomposition tree can be used for a visual overview of the causes	Other tools are used for this part, one option is the decomposition tree

Table 10: comparison between the decomposition tree and a performance dashboard based on Feelders and Daniels (2001)

6 NEW PERFORMANCE DASHBOARD

This chapter describes the prototype of a new performance dashboard and explains how this artefact will help managers to make better decisions. This chapter is based on the steps of the article of Feelders and Daniels (2001) on page 53 and the FIVE model from chapter two. This design research focuses on the sales data of a foodmart in 2015 and 2016 and the bike shop of organisation Z. The foodmart data consists of two fact tables and is star scheme based. The foodmart got different stores in the USA, Canada and Mexico. Organisation Z is a small bike shop for students. The sales data of organisation Z is based on the year 2018 and 2019. It is hard for employees to get insights in all information and process this information. This artefact tries to prevent information overload and enhance the main points of situational awareness. The past chapters showed the requirements for a performance dashboard. This chapter will modulate these requirements to form a basis for the artefact. Section 6.1 describes the artefact based on an UML class diagram. Section 6.2 focuses on the different Python scripts that are necessary based on the requirements and section 6.3 used all the information to build the new artefact based on data from the foodmart. This new artefact is tested in section 6.4.

6.1 UML class diagram for automated business diagnosis

This section is focused on modelling a general model for an automated business diagnosis. This modelling is based on an UML class diagram. An UML class diagram is a type of static structure diagram that gives a description of the structure of a system. The UML class diagram shows the classes, attributes and operations of these classes and the relationships among objects. The UML diagram on the next page contains the different components of a performance dashboard with the possibility to give an explanation based on an influence matrix. A performance dashboard is used to give a visual overview of a dataset. This database consists of different data cells, every data cell contains organisational data. The focus is on a part of the dataset, the part that shows irregular behaviour and got the highest influence on the variable. This small part of the dataset is called an exceptional cell. This cell is analysed to find causes. An exceptional cell can be explained by an explanation tree to show the causes. These causes can be contributing or counteracting. This explanation model uses hierarchical, organisational measures and an aggregation table to give an overview of the causes. The reason for integrating an aggregation table is to reduce the input, output, RAM, CPU, swapping requirements and the need for dynamic calculations. This table minimize the amount of data that must be aggregated and sorted when the dashboard needs to perform an explanation. A dashboard got the possibility to visualize data and generate an automated text message based on this general model. This model is used to generate automated text messages in Power BI with the help of Python visuals.

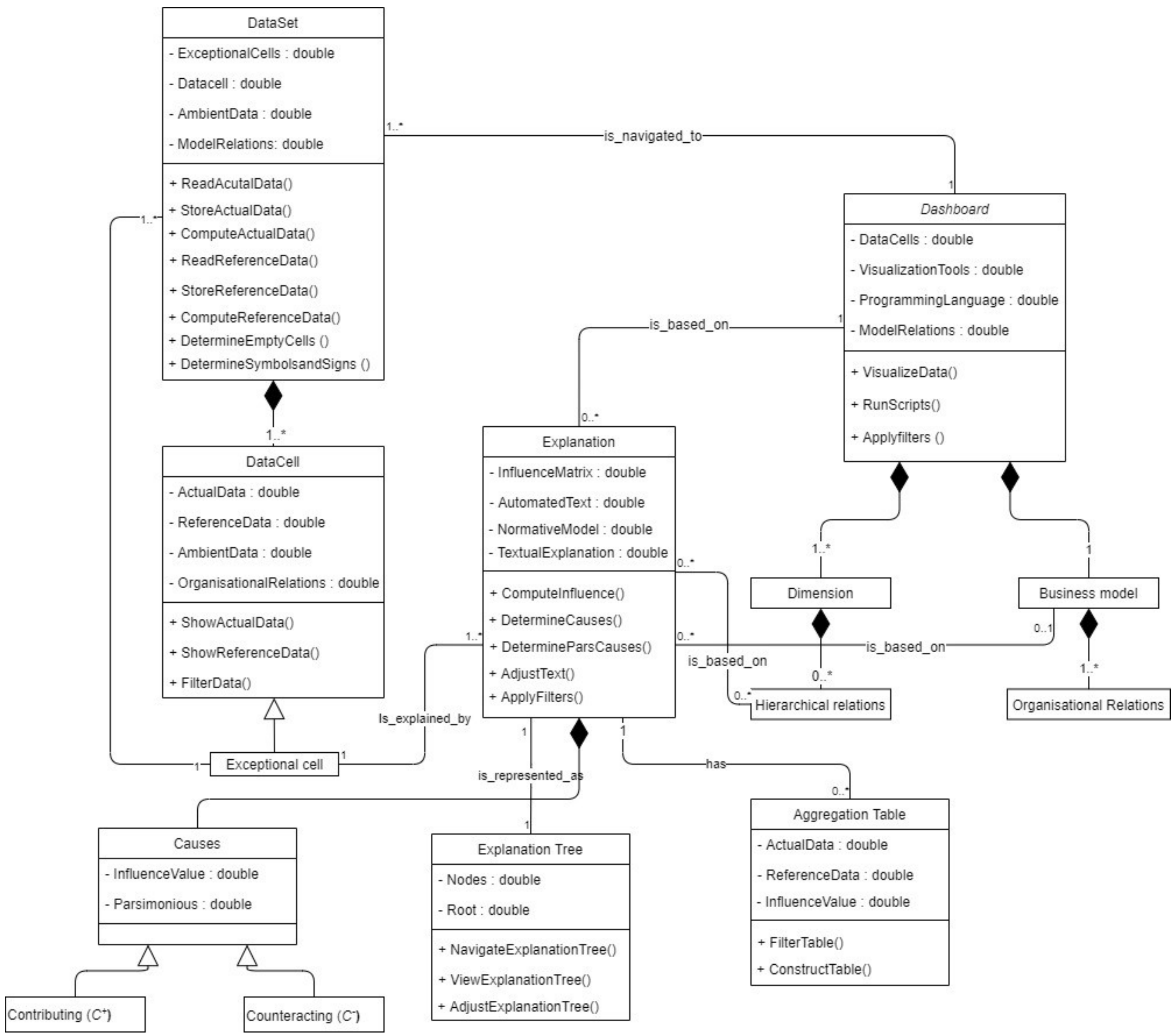

Figure 9: UML class diagram for the artefact for diagnosis

6.2 Python scripts for automated business diagnosis

The UML class diagram is a general model for the new performance dashboard. This UML model is used to generate Python scripts that work in performance dashboards to execute automated business diagnosis. Power BI is a visual based performance dashboard. Before creating Python scripts for analysing data there needs to be a possibility to create text messages. It is possible to create a text message in Power BI with the help of Python visual. A programming language is needed because Power Bi is primarily focused on visuals instead of making text. If a user wants to have text in a performance dashboard it is necessary to use the Python visual or R visual tool. These tools can create graphs, tables, plots and other visuals to show data in a visual way. It is possible to create automated text in a Python visual if the programmer uses the table visual. This table visual gives the programmer the ability to use data variables and text. First the programmer needs to find the right cell and removes the cell lines and adjust the font. This is done with the following code:

```
import pip
import matplotlib.pyplot as plt
Table = plt.table(cellText=[[text]],loc='center',cellLoc='left',edges='open',
)
Table.set_fontsize(24)
plt.axis("off")
plt.show()
```

This was the first adjustment that needs to be made. Now this problem is solved it is possible to improve the performance dashboards even more. Step (1) and (2), the actual data and reference data, can be displayed in the performance dashboard by pointing out a cell. This can be done in table format or in visual tools. Step (3) is based on the model relations and within the database there are already hierarchical and dimensional relations. It is possible to adjust model relations in Power BI, but it is better to leave the data relations in the performance dashboard. For step (4) and (5) an influence matrix is needed. It is not possible to create an influence matrix in Power BI without a programming language such as R or Python. A new data frame needs to be created for using EQ (3) or (4) of Feelders and Daniels (2001). The Python script below shows how to create a new data frame within Python with an influence matrix. In the script below there are different parts red. The red parts are different for every user. For example, select dataset is based on the name of the dataset of the organisation. This influence matrix can be created with the following Python script.

```
import pandas as pd
import numpy as np
import matplotlib as mat
import pip
import matplotlib.pyplot as plt

df = pd.read_csv('(Select dataset).csv')
df = df.rename(columns={ '(t)' : 't', '(a)' : 'a', '(r)' : 'r'})
df = df.fillna(0)
df = df.replace('\€', '', regex=True, inplace=False)
df = df.replace('\%', '', regex=True, inplace=False)
df = df.replace('\,', '.', regex=True, inplace=False)
df.a = df.a.astype(float)
df.r = df.r.astype(float)
(r) = df['r'].sum()
df['prof_diff'] = df.apply(lambda row: row.a - row.r, axis = 1)
df['inf'] = df.apply(lambda row: row.r - row.a + norm_value, axis = 1)
df['inf_perc'] = df.apply(lambda row: 100 - ((row.inf / (r) *100), axis = 1)

Numbre_of_hits = (n number of hits)
df = df.round(n number of decimals)

df1 = (df.n(smallest/highest values)(Numbre_of_hits, ['filter on variable’]))
print(df1)
```

This Python script is based on step (4) and (5) of Feelders and Daniels. The influence matrix is generated and used to filter the causes. In this case the *n* highest number is used for the filtering process. It is possible to use other filters instead of smallest/highest.

Step (6) is based on making the causes visual and textual. The filtering process above creates a new data frame that can be used as the basis for an explanation tree. This tree highlights the most important causes and the effect of the causes. To create an automated text another Python script is needed. The following script is added to the Python script of the influence matrix to create a text message:

```
values = [[df1.iloc[i][0],df1.iloc[i][4],df1.iloc[i][6]] for i in range(7) if
df1.iloc[i][1]>0]
```

```
text=('After filtering the most important cause is cause {} \nwith a difference
of {} \nand an influence percentage of {}'
'\n\nThe second most important cause is cause {} \nwith a difference of {} \nand
an influence percentage of {}'
'\n\nThe third most important cause is cause {} \nwith a profit difference of {}
\nand an influence percentage of {}')
(4th, 5th, etc cause can be added with the same sentences)
.format(values[0][0], values[0][1], values[0][2],values[1][0], values[1][1],
values[1][2],values[2][0], values[2][1], values[2][2], values[3][0],
values[3][1], values[3][2], values[4][0], values[4][1], values[4][2],
values[5][0], values[5][1], values[5][2], values[6][0], values[6][1],
values[6][2], values[7][0], values[7][1], values[7][2],)

Table = plt.table(cellText=[[text]],loc='center',cellLoc='left',edges='open',)
Table.set_fontsize(24)
plt.axis("off")
plt.show()
```

This Python script is based on the three most important causes. If the filter is adjusted more weeks can be added by adjusting the same sentence and change the format with another new values based on the column and row.

6.3 New performance dashboard based on data

As stated before, employees have a limited capacity to process information. Organisations use performance dashboards to prevent information. This section uses the information from section 6.1 and section 6.2 to create an automated business diagnosis based on a star scheme dataset. This is a different dataset then the bike shop dataset of organisation Z. The foodmart dataset is retrieved from the Tilburg University and can be requested from the researcher or the University. The reason for a new dataset is the difference in volume. The dataset based on the foodmart is way bigger and the explanation tree needs to work on huge volumetric to be beneficial for the management board. Step (1) and (2) are focused on determine the actual data (a) and reference data (r). the numeric example showed that it is possible to select a cell in the star scheme and point out actual or reference data. In this case the focus is on profit, the year 2016 will be the actual value and the year 2015 the reference value because there is only data about 2015 and 2016. The next picture gives the sales data of the foodmart in 2015 and 2016. The graph shows the profit of 2015 and 2016 with the formula of sales costs minus sales revenues.

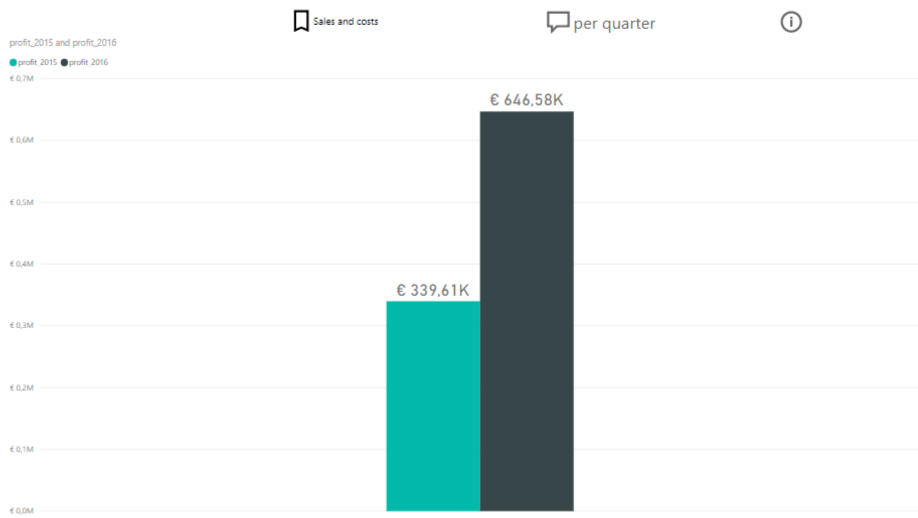

Figure 10: Sales data difference between 2015 and 2016

This graph shows that there is a huge difference between 2015 and 2016. There is a growth, but it is not really clear how much this growth of profit is and it doesn't give any reasons for this. As stated earlier, in a competitive world you need to outperform your competitors. What is your percentage growth? What are the causes for this growth and are there also causes that had negative influence? These are important questions that cannot be answered in one graph. An automated business diagnosis could solve this problem. To help users an information button is added to give a textual overview of the image. This information button is added to reduce the information barrier between the data analyst and the management board. The management board can use this button to generate an automated textual business diagnosis. This is possible with the help of Python coding. The following text is created with a Python script from the appendix.

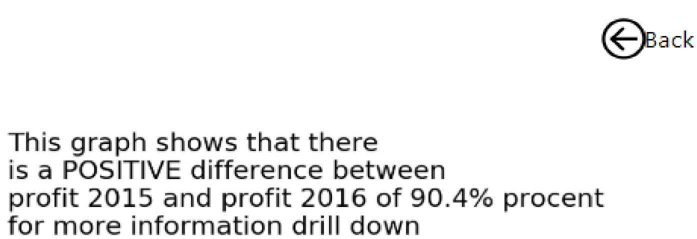

Figure 11: Automated text for profit difference between 2015 and 2016

This gives more information about the graph but is still easy to understand. The information is limited and only focused on the most important numbers of the organisation to reduce information overload. The combination of visual and textual data improves the situational awareness. Figure nine compares 2015 with 2016. One key factor of the FIVE models shows that a performance dashboard also needs to have the function to drill down. The artefact still got the possibility to drill down. From year it is possible to drill down to quarter and from quarter to month and from month to week. A year consists of four quarters and the profit of every quarter can change drastically compared to another quarter. Some months are more beneficial than other months. For example, ice creams are mainly sold in the spring, summer and partly in the autumn. High temperatures will have a positively impact on the sales of ice creams. Therefore, it is useful to drill down from year to quarter. The picture below shows the profit per quarter of 2015 and 2016.

Figure 12: Sales difference between year 2015 and year 2016 based on quarters

With the script from the appendix it is possible to generate a business diagnosis that is text based. The Python script gives the following output:

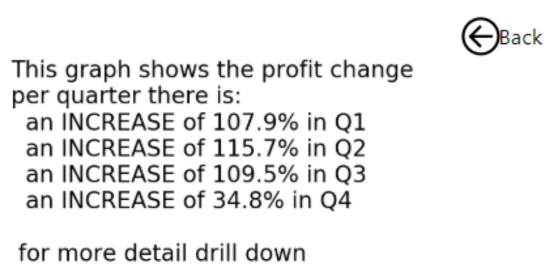

Figure 13: Automated text for profit difference between 2015 and 2016 based on quarters

The text shows that in every quarter the profit increased. Nevertheless, quarter four shows strange behaviour. Every quarter shows an increase of more than 100% except quarter four. Seasonal products can be an explanation for this phenomenon but to have more information about this it is possible to use step (4), (5) and (6) of the explanation model.

Step (4) and (5) are focused on finding the causes and filter these causes. The profit of 2016 was great but as stated before it was not clear why there was such a great increase and what happened in the last quarter. For step (4) and (5) the actual and reference data is needed. The reference and actual data are used to compute the influence matrix with EQ (3) and/or (4). The following data frame is created with the Python script from section 6.2. The reference model below shows the first five rows of the influence matrix, the full matrix is in the appendix.

t	a	r	influence_value	prof_diff	inf	inf_perc
1	0.00	3531.19	0.72	-3531.19	496627.51	-0.72
2	0.00	5925.36	1.20	-5925.36	499021.68	-1.20
3	4844.72	8411.02	1.71	-3566.30	496662.62	-0.72
4	15950.27	11180.62	2.27	4769.65	488326.67	0.97
5	18424.46	12919.89	2.62	5504.57	487591.75	1.12

Figure 14: influence matrix

T stands for the week of the year, *a* is the actual value, *r* is the normal value, influence_value is the influence of the week for 2016, profit difference is the difference between the actual value and the normal value, influence is the profit of year 2016 when every week is the same as the year value except the week and influence_percentage is the percentage of influence on the variable. This influence matrix is used to determine the most important weeks of the year for the organisation. There are different filtering options to highlight the worst and best weeks. In this case the researcher used the *n* highest/smallest number filter from section 6.2. The following pictures show the seven best weeks and the seven worst weeks for the organisation based on the profit.

t	a	r	influence_value	prof_diff	inf	inf_perc
41	16445.03	9594.73	1.95	6850.30	486246.02	1.39
32	20824.63	14850.40	3.01	5974.23	487122.09	1.21
5	18424.46	12919.89	2.62	5504.57	487591.75	1.12
50	17626.15	12167.65	2.47	5458.50	487637.82	1.11
47	18291.21	12867.41	2.61	5423.80	487672.52	1.10
21	16548.37	11173.07	2.27	5375.30	487721.02	1.09
23	13668.97	8380.00	1.70	5288.97	487807.35	1.07

Figure 15: Seven largest influence percentage of profit difference

The seven weeks of figure fourteen got a big profit difference and have a high level of influence on the total profit of the organisation. These weeks needs to be analysed to see why these weeks were so profitable.

t	a	r	influence_value	prof_diff	inf	inf_perc
2	0.00	5925.36	1.20	-5925.36	499021.68	-1.20
3	4844.72	8411.02	1.71	-3566.30	496662.62	-0.72
1	0.00	3531.19	0.72	-3531.19	496627.51	-0.72
51	3909.30	5515.95	1.12	-1606.64	494702.96	-0.33
52	0.00	559.90	0.11	-559.90	493656.22	-0.11
7	7154.31	6989.17	1.42	165.14	492931.18	0.03
26	10319.17	10085.54	2.05	233.64	492862.68	0.05

Figure 16: Seven smallest influence percentage of profit difference

Picture fifteen shows the seven smallest values of the influence percentage based on the actual and norm value per week. Not every week is important. Figure fifteen shows the value of week 7, 26 and 52. These weeks are in the top seven based on profit difference, but the influence is too low to be worried about. It is possible to adjust the filter to show weeks with at least an influence percentage of -0.10. It is more important to focus on week 1, 2 and 51. These weeks needs to be analysed to why the profit is below the norm value.

Step (6) is based on giving a textual and visual overview of the causes. The data frame from the influence matrix can be used to generate automated business diagnosis and create an explanation tree. The following text is based on the Python script from section 6.2 and can be found in the appendix:

All zero values are removed

After filtering the worst week is week 3.0
with a profit difference of -3566.3
influence percentage of -0.72

The second worst week is week 51.0
with a profit difference of -1606.64
influence percentage of -0.33

The third worst week is week 7.0
with a profit difference of 165.14
influence percentage of 0.03

All zero values are removed

After filtering the best week is week 41.0
with a profit difference of 6850.3
influence percentage of 1.39

The second best week is week 32.0
with a profit difference of 5974.23
influence percentage of 1.21

The third best week is week 5.0
with a profit difference of 5504.57
influence percentage of 1.12

Figure 17: Automated text for worst and best weeks

The following explanation tree is generated to visualize the data in Power BI:

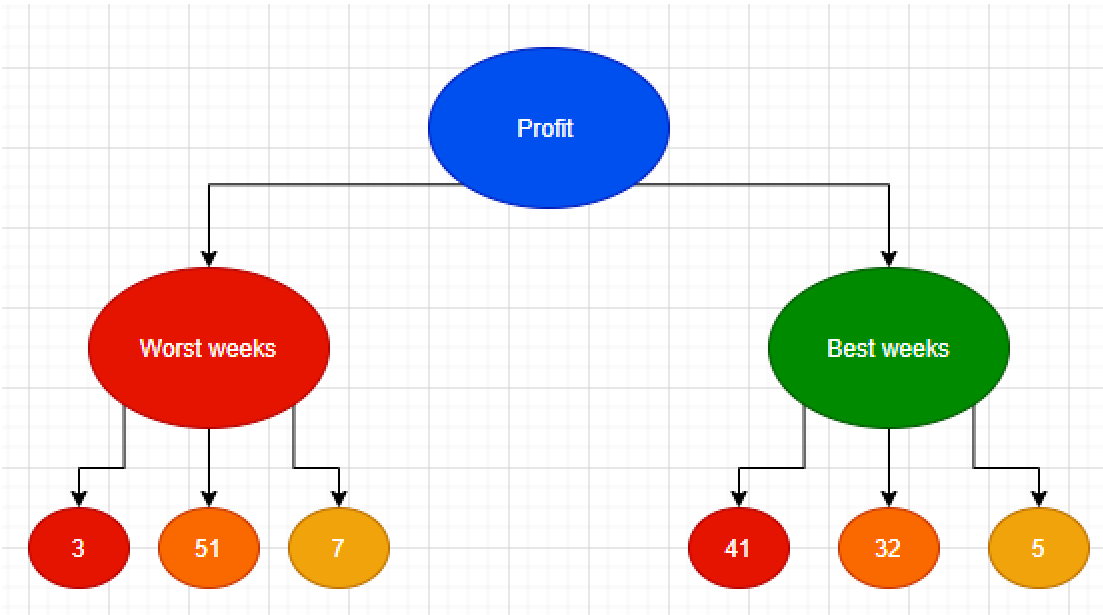

Figure 18: Explanation tree based on the dataset of the foodmart

This new artefact can be an addition to the existing performance dashboards because it is focused on the literature and it adds value. The three steps mentioned in the literature review are see, understand and anticipate in the future. Understanding data becomes easier for every employee because it is a visual graph with text that explains the situation. This helps organisations anticipate in the future by making changes in the current business plan to prevent losses or increase the profit margin. Furthermore, the FIVE model from the literature states that a performance dashboard needs to be flexible, interactive, visual and external focused. A normal performance dashboard is already visual, flexible and got the ability to change quickly with new or changed data. This new artefact got the same abilities and improved some features. The text will change automatically if the data is changed. Interactive is focused on the ability to drill down, monitor KPI's and not solely show graphs. The artefact adds text that explains the graphs. This improves the interactive part of the FIVE model because before this extension it was purely focused on graphs and other visuals. The artefact also improved the external benchmarking part of the FIVE model. The organisation gets a good overview of the current state in the market if the reference data is based on competitors. Performance dashboards can be used to compare the data of an organisation with the competitors of this organisation. Nevertheless, multiple graphs were needed for this process in the past. This leads to information overload. With the new extension it is possible to benchmark the results of the organisation. The artefact is an extension for the existing Power BI performance dashboard and the good parts of the performance dashboards are still available. Some parts are changed to improve the dashboards based on the FIVE model, situational awareness and information overload.

6.4 Testing of the explanation tree

In this section the results of the testing phase are displayed. Testing is an important feature of a design study. The artefact needs to add value for the users. The testing is based on two persons. One with IT and performance dashboard experience (person I) and the other without any experience in data (person N). The reason for having one person with IT experience and the other without is to see if there is any difference between the two groups. The explanation tree needs to be understandable for every employee in an organisation and not only for people with knowledge of IT. The test persons need to find eight causes. Four different causes that have a contributing effect and four causes with a counteracting effect. The test persons need to filter these causes to select the one with the highest influence on the variable. The test persons will use the dataset of organisation Z. The reason for using the dataset of organisation Z is that this is a small dataset with the possibility to find causes in three different ways. The first phase is based on finding causes in the dataset without any tools. In the second phase the test persons need to use the new decomposition tree from Power BI to find the causes. The last the last step is based on the artefact, the explanation tree. The reason for the dataset of organisation Z is that the dataset of the foodmart is too big to find causes without tools. The results of the test persons can be found in the next table. The table gives an overview of the weeks of the test persons. The weeks are ordered from left to right. Left is the most important week and right the least important week.

	C^+ *I*	C^- *I*	C^+ *N*	C^- *N*
No tool	5, 9, 17 and 18	4, 26, 45 and 51	4, 18, 51 and 52	10, 11, 12 and 13
Decomposition tree	9, 17, 18 and 39	4, 6, 44 and 51	4, 18, 51 and 52	11, 13, 35 and 37
Explanation tree	9, 17, 18 and 39	4, 6, 44 and 51	9, 17, 18 and 39	4, 6, 44 and 51

Table 11: Contributing and counteracting weeks test persons

The first phase was based on finding the causes without any tools. Because person I had some experience with data the focus was not only on the biggest loss or profit. Nevertheless, not all the correct weeks were highlighted. The reason for this could be that it is easy to make a human mistake in a dataset. Person N had no experience with data and was only focused on the weeks with the highest profit and loss. This example shows the reason for this artefact. It is not always about the absolute numbers; it is about the difference based on an influence matrix. Because the two test persons had a different approach there was a small overlap between the most important causes between the two test persons.

The second phase was based on finding the causes with the help of the decomposition tree. The researcher gave a short introduction about the decomposition tree and how you can add measures to improve the usage. Person I used his knowledge to add some measures before using the decomposition tree. He applied measures and changed some variables to find the correct weeks. This dataset is small and simple measures can be used to find the right weeks. Person N tried to use the tool. During the introduction person N already said that he was going to try something to find the causes. Nevertheless, this went not so well and almost the same weeks were noted. During the interview person N complained about the usage of the decomposition tree. There was again a small overlap between the two groups.

The last phase was based on the explanation tree. After a short introduction both test persons got the same task and got the same answers. These weeks are the same as in 4.3.2. The time to find the causes decreased significantly compared to the other phases. A small interview was scheduled to see if there are any tips for the explanation tree and to get clarification of the results. The interview questions can be found in the appendix and the answers are used in the conclusion of this section.

The result of the testing phase shows that the explanation tree can be beneficial for organisation. Employees with no IT knowledge can use the tool to find the most important causes within an organisation. Person I showed that the decomposition tree can be beneficial. The decomposition tree is a tool that can be used in simple cases when the users have IT knowledge. Nevertheless, this testing phase used a small dataset where it is possible to apply measures and use this in the decomposition tree. When the dataset consists of more data this can lead to problems because the decomposition tree is a black box. It is not possible to use the equations of Feelders and Daniels

(2001). In this case the decomposition tree gave the same results. Nevertheless, this is an exception. Both test persons where happy with the ease of use of the explanation tree. It was possible to change the filters and after a small instruction both persons could work with the artefact. Person N was happy that even without IT knowledge it was easy to use because it is based on basic knowledge. The decomposition tree was too hard to understand because this tool had multiple options. The script of the explanation tree is hard to understand but the user only needs to fill in different words to make the script run. Nevertheless, there are also some negative point in the explanation tree. The design could be increased of the artefact to improve the front-end of the tool. The test persons missed some extra colours or highlights because they would not use the automated business diagnosis in their presentation. Also, the automated text could be improved.

7 CONCLUSION

The last chapter of this research answers the research question, the research question is:

How to extend performance dashboards with features for explanatory analytics?

In this research the question is answered with an artefact. This artefact is an extension for a performance dashboard and is created based on a literature review. Firstly, it is important to have a good overview of which performance dashboard is suitable for explanatory analytics. There are multiple performance dashboard tools available on the market. These dashboards are discussed in this research. A performance dashboard is a diagnostic tool to give a quick overview of the organisation and improve the decision making in the organisation. Performance dashboards displays the most important data in a simple one-page overview and contains visual and functional features which can help the interpretation and the cognition of data.

7.1 Performance dashboards

The research question is divided into two sub questions. The first sub question is focused on performance dashboards and clarifies the reason for the chosen dashboard. The sub question is:

What are performance dashboards, and which performance dashboard is most suitable?

Not every performance dashboard is as good as its creator's state. The literature states that a performance dashboard needs to be FIVE. FIVE means the following: flexible, a performance dashboard needs to be easy to modify, and must be used by multiple users and needs to have the ability to personalize the overview page. Interactive, a performance dashboard needs to have the ability to drill down, monitor KPI's and show not solely graphs. Visual, a performance dashboard needs to give a visual overview of accurate data from the past and the present day in time. External benchmarking, a performance dashboard needs to have the ability to compare the results with competitors and make prescriptive and predictive analysis based on the data. These four characteristics, in combination with other key characterises, are used as a basis for developing the artefact. The decision for Power BI is based on the ease of use, low costs and high level of usage around the world. Furthermore, Power BI gave the researcher the opportunity to create an artefact with a small amount of programming knowledge. Lastly, in the active community, Power BI is widely used and with all the available tools for making diagnostic business analyses, this tool is suitable for this research.

7.2 Explanatory analytics

The next step was to focus on the diagnostic analytics in relation to performance dashboards. The following sub question relates to this step:

What is the relation between diagnostic analysis and performance dashboards?

Every year, performance dashboards plays a bigger role in organisations. Every year there is more information available and organisation needs to make multiple decisions based on this information. This decision process is not easy and performance dashboards can provide support for the management board to justify their decisions. Nevertheless, there is a knowledge gap between the data scientists and the management board that can result in wrong decisions. The management board doesn't have any knowledge about data model, graphs and underlying relations. The data scientist has no knowledge about the business side of the organisation. Performance dashboards have the potential to reduce this gap by providing a one-page overview of the KPIs of an organisation. This one-page overview will reduce the information overload and increases the situational awareness. This overview will improve the identification, development and selection steps in the decision making. Nevertheless, the performance dashboards of today are not good enough in the eyes of the researcher. It gives a one-page overview, but it is still hard to understand. Not every cause can be seen in a one-page overview and there is a difference between absolute and relative data. The article of Feelders and Daniels is based on creating automated business diagnosis. This article can be summarized into six steps that are integrated in the artefact to make automated business diagnosis. There needs to a reference dataset to compare the actual data. This reference dataset can be created with the actual data from the past and the current year or based on the data from competitors. Power BI can add measures to the current dataset and create new data frames based on the already existing columns. With the help of this reference dataset is it possible to create an influence matrix to find the most important causes for an organisation. This will improve the decision-making process within the organisation. The one page-overview is related to this data table, and with the help of Python is possible to generate text-based messages to inform the management board with facts that are based on the dataset and influence matrix. The combination of diagnostic analytics and performance dashboards enhance the power of the two subjects only by themselves. They are complementary to each other and fill in each other's blank spots. Diagnostic analytics is focused on why something happened. Performance dashboards have the data but don't give a reason for why something happens. Together it is possible to show the results and give reasons for why this happened.

7.3 The extension

The focus of this research is on creating an artefact that can improve performance dashboards. These dashboards have the function to show data in an easy way and therefore improve the decision-making process in an organisation. The result is an artefact that improves the performance dashboards of today. The focus of performance dashboards is on presenting data in a simple way. Nevertheless, this "simple" way is not always that simple. This artefact makes a performance dashboard simpler by adding automated text based on the most important causes for a variable and highlighting the most important causes. Because performance dashboards are visual based, and the extension is text based, an external tool is needed. It was possible to use the programming language R or Python. Python was the most suitable to solve this task, because the researcher has previous experience with this language and there was more information available. With the help of Python, it is possible to generate automated text messages. Research and the trial and error method are used to make the artefact. This automatically generated text message changes when the data is changed and therefore can be reused in the future. It is also possible to use the artefact for other KPIs instead of profit. The only thing that needs to be changed, is the column name in the beginning of the script. This artefact will help the management board and the data analyst to improve the decision making in the organisation. The new artefact gives the management board the opportunity to drill down to get more information to see the causes that have a high level of influence on the result. The data engineers will have more time to analyse data because they do not have to focus on explaining the graphs and table anymore. The artefact is compatible with Power BI and more information about the scripts can be found in the appendices and in chapter six. Lastly, the four characteristics are all in the new artefact. This artefact improved the interaction and external benchmarking characteristic with increasing the ability to drill down based on causes and to compare your data on a reference model.

7.4 Recommendations

The goal of this research was to improve the performance dashboards to reduce complexity and the imbalance between the creators and the readers. The information delivered in the old performance dashboards is not always the right information. This research formulated some recommendations based on the goal and research question.

1. To provide the right information it is important to have clean data. The data needs to be clean before entering the dataset. If the data is not clean it is possible that the artefact will not be beneficial for organisations because not all data is observed. To prevent this problem, it is possible to train employees to use the artefact and data loading in a correct way. A small training and a support group can increase the success rate of the artefact.
2. The complexity is reduced by providing a one-page overview that is understandable for every employee in the organisation. The possibility to create automated text reduces the imbalance between the creators and the readers. It is important that organisations invest in good performance dashboards because decisions based on data are the future.
3. The text message and the filter can be adjusted to be in line with the organisation. It is recommended to let the data analyst in combination with the management board be in charge for this adjustment to make sure the right information is displayed. Research shows that performance dashboards needs to be adjustable and usable for the whole organisation. The artefact can easily be adjusted.
4. The testing phase showed that the artefact is understandable for every employee in an organisation. Nevertheless, the front-end could be improved to increase the chance of usage. This depends on the organisations needs.
5. Running scripts in Power BI takes a long time and finding the wrong sentence is hard. It is recommended to improve this in Power BI or use an external programme for running scripts.

A software organisation can use this research to improve a performance dashboard. This research provides an overview of the components and formulas for an automated business diagnosis based on the literature. There is room for improvements and a professional organisation can fine tune the artefact. It is recommended that the artefact is implemented in collaboration with an organisation to ensure that the dashboard can be used by every employee.

7.5 Practical and scientific contribution

The practical contribution of this research is that organisation can be more transparent to stake and shareholders. Less time is needed to make a business diagnosis and the data in an organisation can add value. This artefact reduces the costs of meetings for explaining the performance dashboards or training employees to understand the dashboards. This saves money and time for the organisation.

The scientific contribution of this research is to solve the knowledge gap about extending performance dashboards with automated business diagnosis. Furthermore, this research provides an overview of the different available performance dashboards. This wasn't available in the literature. Lastly, this research provides further research for other researchers to improve the artefact.

7.6 Limitations and implications for further research

This study is focused on Power BI. It is one of the best performance dashboard tools available but even during this study multiple things changed within Power BI. The artefact works with the Power BI version of January 2020. One of the limitations is that updates in the future can influence the artefact. It is possible that these changes have a positive or negative effect on the artefact. For example, during this research Power BI added the decomposition tree that had the possibility to replace the artefact. Nevertheless, it was not suitable for automated business diagnosis. Another limitation of this research is that it is not possible to use Python Visual in Power BI when there are two fact tables in the dataset. This problem can be solved with the merge option in Power BI. Nevertheless, it is not possible to merge two fact tables if there is no overlap between the two tables. In this dataset it was possible to merge the two fact tables into another fact table based on the time_id. Most of the datasets will have overlap but if this is not possible all the data needs to be put in another table manually. Furthermore, if the data is not clean or there is not a data model behind it then there is a possibility that the artefact doesn't work. A lot of organisations got dirty data instead of clean data. Lastly, this design study is focused on a literature review instead of qualitative or quantitative research. The risk of a literature review is that the artefact is too much focused on the literature instead of the employees. The researcher tried to prevent this by testing the artefact. Nevertheless, this sample group was based on two persons. Two persons is not a big sample, and this leads to an inflated false discovery rate and a low reproducibility. And finally, this study was done in a relative short time frame. To improve the speed of coding a good laptop or computer would be useful. The computer of the researcher had no SSD and this resulted in crashes and some frustrating moments. Halfway during the research, the scripts moved from Power BI to PyCharm because it was possible to test codes faster. In the future it would be a good idea to start with running your scripts in PyCharm and with a good computer. This will improve the speed and chance of success.

Future research is needed to improve the artefact. A part of the artefact is automated, but it is possible to automate a bigger part of the business diagnosis process. This research can be used as a basis for to improve the artefact. The Python scripts can be improved to speed up the business diagnosis. Furthermore, the artefact could be tested in a real organisation. The testing phase was based on two test persons and a dataset from the Tilburg University. Real life testing would be beneficial to find any flaws of the artefact. Lastly, artificial intelligence will play a big role in the future. This artefact can be combined with artificial intelligence to make decisions based on predictions of the future.

REFERENCES

Books and Articles:

Adeoti-Adekeye, W. B. (1997). The importance of management information systems. Library Review, 46(5), 318-327.

Ackoff, R. L. (1967). Management misinformation systems. Management science, 14(4), B-147.

Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. Journal of information technology, 20(2), 67-87.

Ballard, C., Herreman, D., Schau, D., Bell, R., Kim, E., & Valencic, A. (1998). Data modeling techniques for data warehousing (p. 25). IBM Corporation International Technical Support Organization.

Bariff, M. L., & Lusk, E. J. (1977). Cognitive and personality tests for the design of management information systems. Management Science, 23(8), 820-829.

Bawden, D. (2001). Information overload. Library & information briefings, (92), 1-15.

Benbasat, I., & Dexter, A. S. (1979). Value and events approaches to accounting: an experimental evaluation. Accounting Review, 735-749.

Bitterer, A. (2011). Hype cycle for business intelligence. Gartner, Inc., Stamford, CT.

Buchanan, J., & Kock, N. (2001). Information overload: A decision making perspective. In Multiple criteria decision making in the new millennium (pp. 49-58). Springer, Berlin, Heidelberg.

Card, M. (1999). Readings in information visualization: using vision to think.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. Communications of the ACM, 54(8), 88-98.

Chen, H. (2011). Design science, grand challenges, and societal impacts. ACM Transactions on Management Information Systems (TMIS), 2(1), 1.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 36(4).

Cheston, C. C., Flickinger, T. E., & Chisolm, M. S. (2013). Social media use in medical education: a systematic review. *Academic Medicine*, *88*(6), 893-901.

Chewning Jr, E. G., & Harrell, A. M. (1990). The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task. Accounting, Organizations and Society, 15(6), 527-542.

Clark, B.H, Abela A.V, Ambler T. (2006). "Behind the wheel", Marketing Management, vol. 15, n. 3: 18-23.

Cohen-Dumani, D. (2016). Microsoft Power BI: A Look Under the Hood. Retrieved from https://www.cmswire.com/information-management/microsoft-power-bi-a-look-under-the-hood/

Cuddley, M. O. (2016). Introduction To Microsoft Power BI: Bring Your Data To Life!. CreateSpace Independent Publishing Platform.

DataFlairTeam. (2018). Microsoft POWER BI Tutorial – Learn Power BI (Beginner to Expert). Retrieved from https://data-flair.training/blogs/power-bi-tutorial/

Davenport, T. H. (2006). Competing on analytics. Harvard business review, 84(1), 98.

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. Communications of the ACM, 54(4), 86-96.

Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. International journal of information management, 20(1), 17-28.

Endsley, M. R., & Garland, D. J. (Eds.). (2000). Situation awareness analysis and measurement. CRC Press.

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. Journal of Cognitive Engineering and Decision Making, 9(1), 4-32.

Eppler, M. J., & Mengis, J. (2003). A framework for information overload research in organizations. Università della Svizzera italiana.

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. The information society, 20(5), 325-344.

Feelders, A. J., & Daniels, H. A. (2001). A general model for automated business diagnosis. European Journal of Operational Research, 130(3), 623-637.

Few, S. (2006). Information dashboard design.

Friedman, T. L. (2005). The world is flat: A brief history of the twenty-first century. Macmillan.

Goldie, J. G. S. (2016). Connectivism: A knowledge learning theory for the digital age?. *Medical teacher*, *38*(10), 1064-1069.

Gonzalez, C., & Wimisberg, J. (2007). Situation awareness in dynamic decision making: Effects of practice and working memory. Journal of Cognitive Engineering and Decision Making, 1(1), 56-74.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. MIS quarterly, 213-236.

Grisé, M. L., & Gallupe, R. B. (1999). Information overload: Addressing the productivity paradox in face-to-face electronic meetings. Journal of Management Information Systems, 16(3), 157-185.

Güratan, I. (2005). The design and development of a data warehouse using sales database and requirements of a retail group (Master's thesis, İzmir Institute of Technology).

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Sethupathy, G. (2016). The age of analytics: Competing in a data-driven world. McKinsey Global Institute, 4.

Howson, C., Sallam, R. L., Richardson, J. L., Tapadinhas, J., Idoine, C. J., & Woodward, A. (2018). Magic quadrant for analytics and business intelligence platforms. Retrieved Aug, 16, 2018.

Huber, G. P. (1983). Cognitive style as a basis for MIS and DSS designs: Much ado about nothing?. Management Science, 29(5), 567-579.

Jacoby, J.; Speller, D. E.; Berning, C. K. 1974. 'Brand choice behavior as a function of information load: Replication and extension. Journal of Consumer Research 1:33– 43

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: measures that drive performance.

Kask, T. (2010). Innovation as a result of strategic decisions in the context of organizational environment: The case of Estonian information and communication technology companies (Doctoral dissertation, University of Tartu).

Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In *2013 Sixth international conference on contemporary computing (IC3)* (pp. 404-409). IEEE.

Kirsh, D. (2000). A few thoughts on cognitive overload.

Koomey, J., & Holdren, J. P. (2008). Turning numbers into knowledge: Mastering the art of problem solving. Analytics Press.

Kutty, A., Shee, H., & Pathak, R. (2007). Managing information overload for effective decision making. An Empirical Study of Organizational Practices, Monash Business Review, 3(3), 8-9.

LaPointe, P. (2005). Marketing by the dashboard light: How to get more insight, foresight, and accountability from your marketing investments. MarketingNPV.

Lee, B. K., & Lee, W. N. (2004). The effect of information overload on consumer choice quality in an on-line environment. Psychology & Marketing, 21(3), 159-183.

Lee, J., Suh, T., Roy, D., & Baucus, M. (2019). Emerging Technology and Business Model Innovation: The Case of Artificial Intelligence. Journal of Open Innovation: Technology, Market, and Complexity, 5(3), 44.

Lusch, R. F., Liu, Y., and Chen, Y. 2010. "The Phase Transition of Markets and Organizations: The New Intelligence and Entrepreneurial Frontier," IEEE Intelligent Systems (25:1), pp. 71-75

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine, 27(4), 12-12.

McGee, J. V., Prusak, L., & Pyburn, P. J. (1993). *Managing information strategically: Increase your company's competitiveness and efficiency by using information as a strategic tool* (Vol. 1). John Wiley & Sons.

McCorduck, P. (2004). Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence. Natick. MA: AK Peters, 576.

Meyer, J. A. (1998). Information overload in marketing management. Marketing Intelligence & Planning, 16(3), 200-209.

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of'unstructured'decision processes. Administrative science quarterly, 21(2).

Nilsson, N. J. (2014). Principles of artificial intelligence. Morgan Kaufmann.

No, J. K. S. (2015). Applying Performance Dashboard in Hospitals. International Journal Of Software Engineering and Its Applications, 9(1), 213-220.

O'Reilly III, C. A. (1980). Individuals and information overload in organizations: is more necessarily better?. Academy of management journal, 23(4), 684-696.

O'reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. Communications & strategies, (1), 17.

Owen, R. S. (1992). Clarifying the simple assumption of the information load paradigm. ACR North American Advances.

Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., ... & Wiesel, T. (2009). Dashboards as a service: why, what, how, and what research is needed?. Journal of service research, 12(2), 175-189.

Potter, R. F., & Choi, J. (2006). The effects of auditory structural complexity on attitudes, attention, arousal, and memory. Media psychology, 8(4), 395-419.

Saariluoma, P. (2005). Explanatory frameworks for interaction design. In Future interaction design (pp. 67-83). Springer, London.

Sallam, R. L., Richardson, J., Hagerty, J., & Hostmann, B. (2011). Magic quadrant for business intelligence platforms. Gartner Group, Stamford, CT.

Shim, J. P., French, A. M., Guo, C., & Jablonski, J. (2015). Big Data and Analytics: Issues, Solutions, and ROI. *CAIS*, *37*, 39.

Sulaiman, N. S., & Yahaya, J. H. (2013). Development of Dashboard Visualization for Cardiovascular Disease Based on Star Scheme. Procedia Technology, 11, 455-462.

The Economist. 2010. "The Data Deluge," Special Report on Managing Information, Technology Section, February 25 (http://www.economist.com/node/15579717).

Troyansky, O., Gibson, T., & Leichtweis, C. (2015). QlikView Your Business: An Expert Guide to Business Discovery with QlikView and Qlik Sense. John Wiley & Sons.

Turban, E., Sharda, R., Aronson, J. E., & King, D. (2008). Business intelligence: A managerial approach (pp. 58-59). Upper Saddle River, NJ: Pearson Prentice Hall.

Velcu-Laitinen, O., & Yigitbasioglu, O. M. (2012). The Use of Dashboards in Performance Management: Evidence from Sales Managers. International Journal of Digital Accounting Research, 12.

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. Information systems research, 2(1), 63-84.

Wang, H., & Wong, K. F. E. (2012). The effect of managerial bias on employees' specific human capital investments. Journal of Management Studies, 49(8), 1435-1458.

Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. Computer, 40(9), 96-99.

Wurman, R.S. (2001). Information anxiety 2, Indianapolis, IN: Macmillan

Ying, L., Lijun, X., & Wei, S. (2009, July). Designing supply chain KPIs for upper-level management. In 2009 IITA International Conference on Services Science, Management and Engineering (pp. 19-21). IEEE.

Internet sources

A Comparative Analysis Of Top 6 Bi and Data Visualization Tools in 2018: ActiveWizards: Data Science and Engineering Lab
https://activewizards.com/blog/a-comparative-analysis-of-top-6-bi-and-data-visualization-tools-in-2018/

Diagnostic Data in Office
https://support.office.com/en-us/article/diagnostic-data-in-office-f409137d-15d3-4803-a8ae-d26fcbfc91dd

Difference Between Qlikview and Qliksense
https://community.qlik.com/t5/New-to-QlikView/Difference-between-QlikView-and-Qliksense/td-p/1291986

Qlik Sense Server Vs. Qlik Analytics Platform
https://www.analytics8.com/insights/qlik-sense-server-vs-qlik-analytics-platform/

Qlikview and Qlik Sense
https://community.qlik.com/t5/Qlik-Design-Blog/QlikView-and-Qlik-Sense/ba-p/1469130

https://core.qlik.com/, Henschen, Doug. "Qlik hits reset button, rolls out new cloud, AI, and developer capabilities". ZDNet. Retrieved 2019-10-09

Power Bi Vs Tableau - 7 Most Valuable Differences You Should Know
https://www.educba.com/power-bi-vs-tableau/

Tableau: Business Intelligence and Analytics Software
https://www.tableau.com/

Data Analytics For Modern Business Intelligence
https://www.qlik.com/us

About Qlik GeoAnalytics
https://help.qlik.com/en-US/geoanalytics/Content/Qlik_GeoAnalytics/Home.htm

The Success Story Of Tableau Software
18th Sep-Name * - https://intellipaat.com/blog/success-story-tableau-software/

Qlikview Advantages and Disadvantages: Qlikview Benefits
DataFlair Team - https://data-flair.training/blogs/qlikview-advantages-and-disadvantages/

Journalists, Look Out: Tableau Public Projects Will Soon Be Even More Public
Todd Bishop - https://www.geekwire.com/2013/tableau-cautions-journalists-tableau-public-projects-public/

Israeli Start-up Makes Big Data Bite-sized
Inbal Orpaz - https://www.haaretz.com/israel-news/business/.premium-start-up-sisense-makes-big-data-bite-sized-1.5273498

Sisense
https://www.crunchbase.com/organization/sisense

https://www.bizjournals.com/sanjose/news/2012/07/17/sisense-hires-amit-bendov-as-ceo.html

Crunchbase Crunching: Sisense and the Big Data Group on the State Of the Vc Industry
Gil Press - https://www.forbes.com/sites/gilpress/2013/03/15/crunchbase-crunching-sisense-and-the-big-data-group-on-the-state-of-the-vc-industry/

Sisense Raises $30m More To Bring Big Data Analytics To Businesspeople
Ingrid Lunden - https://techcrunch.com/2014/06/12/sisense-raises-30m-more-to-take-big-data-analytics-to-businesspeople/

https://www.gartner.com/doc/reprints?id=1-4PUK79W&ct=180130&st=sb%2F

Sisense Closes $10m Series C Round Of Funding Led By Battery Ventures
Ken Yeung - https://thenextweb.com/insider/2013/04/03/big-data-company-sisense-raises-10m-round-from-battery-ventures-opus-capital-and-genesis-partners/

Sisense Acquires Periscope Data To Build Integrated Data Science and Analytics Solution
Ron Miller - https://techcrunch.com/2019/05/14/sisense-acquires-periscope-data-to-build-integrated-data-science-and-analytics-solution//

Sisense: Modules and Features
https://www3.technologyevaluation.com/store/benchmark/sisense.html

Excel Vs Business Intelligence Software
Phocas Software - https://www.phocassoftware.com/business-intelligence-blog/excel-vs-business-intelligence-software

Exploratory Data Analysis Of Classification Problems with Tableau
Sergey Medvedev - https://towardsdatascience.com/exploratory-data-analysis-with-tableau-for-classification-problems-9dea20ab8c01

https://documentation.sisense.com/latest/using-dashboards/exploration-paths.htm

Cofsky, A. (2019, 12 November). *Power BI Desktop November 2019 Feature Summary.* https://powerbi.microsoft.com/en-us/blog/power-bi-desktop-november-2019-feature-summary/

Decomposition Tree - Power Bi
https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-decomposition-tree

Online book

Rad, R. (2019). Power BI From Rookie to Rockstar Edition 4: 7th of Jan 2019

APPENDICES

Appendix I Interview questions

Questions before the testing phase:

Do you have any experience with IT?

Have you ever worked with a performance dashboard?

Questions after the testing phase:

What is your option about finding causes without any tools?

What is your opinion about finding the causes using the decomposition tree?

What is your opinion about finding the causes using the explanation tree?

How can the explanation tree be improved?

Appendix II: Data frame organisation Z

Week	Profit 2019	Profit Norm	Difference	Inf percentage
1	845	800	18960	0.24%
2	1450	1450	18915	0.00%
3	525	600	18840	-0.40%
4	-4320	-4625	19220	1.61%
5	-3230	-2890	18575	-1.80%
6	-1890	-2125	19150	1.24%
7	425	700	18640	-1.45%
8	1245	1456	18704	-1.12%
9	-700	3459	14756	-21.99%
10	2865	3000	18780	-0.71%
11	2355	2255	19015	0.53%
12	1745	1895	18765	-0.79%
13	2315	2265	18965	0.26%
14	1550	1650	18815	-0.53%
15	1450	1645	18720	-1.03%
16	875	865	18925	0.05%
17	-1250	400	17265	-8.72%
18	-3545	-765	16135	-14.70%
19	-680	-740	18975	0.32%
20	-625	-650	18940	0.13%
21	-590	-620	18945	0.16%
22	-675	-550	18790	-0.66%
23	185	245	18855	-0.32%
24	295	355	18855	-0.32%
25	860	865	18910	-0.03%
26	1015	925	19005	0.48%
27	755	915	18755	-0.85%
28	1065	1185	18795	-0.63%
29	845	1085	18675	-1.27%
30	1415	1485	18845	-0.37%
31	1015	985	18945	0.16%
32	1140	1285	18770	-0.77%
33	1025	1135	18805	-0.58%
34	800	945	18770	-0.77%
35	1495	1485	18925	0.05%
36	1165	1285	18795	-0.63%
37	1195	1185	18925	0.05%
38	1275	1285	18905	-0.05%
39	245	650	18510	-2.14%

40	285	615	18585	-1.74%
41	235	545	18605	-1.64%
42	540	545	18910	-0.03%
43	645	745	18815	-0.53%
44	1055	925	19045	0.69%
45	915	825	19005	0.48%
46	845	850	18910	-0.03%
47	-500	-250	18665	-1.32%
48	-725	-755	18945	0.16%
49	-655	-800	19060	0.77%
50	-615	-650	18950	0.19%
51	-5210	-5475	19180	1.40%
52	-5250	-4980	18645	-1.43%

Appendix III: Script for automated text of sales data difference 2015 and 2016

```
    # dataset = pandas.DataFrame(sales 2015, sales 2016)

    # Paste or type your script code here:
    import pandas as pd
    import numpy as np
    import matplotlib as mat
    sales_2016_sum = np.sum(dataset.loc[0:0])[1]
    sales_2015_sum = np.sum(dataset.loc[0:0])[0]
    sales_incr = 100*sales_2016_sum/sales_2015_sum-100.0
    sales_sign = "NEGATIVE"
    if(sales_incr>0):
        sales_sign = "POSITIVE"
data = [sales_sign,"{0:.1f}%".format(sales_incr)]

names = dataset.columns

text = "This graph shows that there\nis a {} difference between \n{} and {} of
{} procent \nfor more information drill down\n".format(data[0], names[0],
names[1], data[1])

import pip
import matplotlib.pyplot as plt
Table = plt.table(cellText=[[text]],loc='center',cellLoc='left',edges='open',
)
Table.set_fontsize(24)
plt.axis("off")
plt.show()
```

Appendix IV: Automated text script of sales data based on quarters of year(x)

```
# dataset = pandas.DataFrame(profit 2015, profit 2016, quarter)

# Paste or type your script code here:
import pandas as pd
import numpy as np
import matplotlib as mat
q4_2015 = dataset.loc[0][0]
q4_2016 = dataset.loc[0][1]

q3_2015 = dataset.loc[1][0]
q3_2016 = dataset.loc[1][1]

q2_2015 = dataset.loc[3][0]
q2_2016 = dataset.loc[3][1]

q1_2015 = dataset.loc[2][0]
q1_2016 = dataset.loc[2][1]

incr_q = [0,0,0,0]
incr_q[3] = (q4_2016/q4_2015)*100-100
incr_q[2] = (q3_2016/q3_2015)*100-100
incr_q[1] = (q2_2016/q2_2015)*100-100
incr_q[0] = (q1_2016/q1_2015)*100-100

profit_sign_Q = [0,0,0,0]

for i in range(0,4):
    profit_sign_Q[i] = "INCREASE"
    if(incr_q[i]<0):
        profit_sign_Q[i] = "DECREASE"


sentences = []

for i in range(0,4):
    sentences.append(" an {0} of {1:.1f}% in Q{2}
\n".format(profit_sign_Q[i],incr_q[i],i+1))

names = dataset.columns

text = "This graph shows the profit change \nper quarter there is:\n {} {} {}
{}\n for more detail drill
down".format(sentences[0],sentences[1],sentences[2],sentences[3])

import pip
```

```
import matplotlib.pyplot as plt
Table = plt.table(cellText=[[text]],loc='center',cellLoc='left',edges='open', )
Table.set_fontsize(24)
plt.axis("off")
plt.show()
```

Appendix V: Automated text script of sales data based on months in year(x)

```
# dataset = pandas.DataFrame(the_month, profit 2016, profit 2015)

# Paste or type your script code here:

import pandas as pd
import numpy as np
import matplotlib as mat
import string

month = []
for row in dataset:
    month.append(row)

month_names = [dataset.loc[i][0] for i in range(0,12) if i!=2]
profits_2016 = [dataset.loc[i][1] for i in range(0,12) if i!=2]
profits_2015 = [dataset.loc[i][2] for i in range(0,12) if i!=2]

sum_profits_2015 = np.nansum(profits_2015)
sum_profits_2016 = np.nansum(profits_2016)

profits_2015_perc = [(x/sum_profits_2015)*100 for x in profits_2015]
profits_2016_perc = [(x/sum_profits_2016)*100 for x in profits_2016]
profits_2016_delta = [profits_2016_perc[i]-profits_2015_perc[i] for i in
range(11)]

perc_sign = [0 for i in range(11)]

for i in range(0,11):
    perc_sign[i] = "INCREASE that is"
    if(profits_2016_delta[i]<0):
        perc_sign[i] = "DECREASE that is"


sentences = []
for i in range (11):
    sentences.append("THE MONTH: {0}, THE {1}
{2:.2f}%\n".format(month_names[i],perc_sign[i],profits_2016_delta[i]))

text = "{}\n{}\n{}\n{}\n{}\n{}\n{}\n{}\n{}\n{}\n{}\nTHE MONTH: December, No Data
From 2016".format(sentences[3],sentences[2],sentences[6],
sentences[0],sentences[7],sentences[5],
sentences[4],sentences[1],sentences[10],
sentences[9],sentences[8],)
```

```
import pip
import matplotlib.pyplot as plt
Table = plt.table(cellText=[[text]],loc='center',cellLoc='left',edges='open', )
Table.set_fontsize(16)
plt.axis("off")
plt.show()
```

Appendix VI: Reference model

t	a	r	influence_value	prof_diff	inf	inf_perc
1	0.00	3531.19	0.72	-3531.19	496627.51	-0.72
2	0.00	5925.36	1.20	-5925.36	499021.68	-1.20
3	4844.72	8411.02	1.71	-3566.30	496662.62	-0.72
4	15950.27	11180.62	2.27	4769.65	488326.67	0.97
5	18424.46	12919.89	2.62	5504.57	487591.75	1.12
6	12542.35	7912.25	1.60	4630.10	488466.22	0.94
7	7154.31	6989.17	1.42	165.14	492931.18	0.03
8	16471.70	12069.10	2.45	4402.61	488693.71	0.89
9	16639.96	11641.15	2.36	4998.80	488097.52	1.01
10	11971.43	8768.78	1.78	3202.65	489893.67	0.65
11	11528.23	8097.73	1.64	3430.50	489665.82	0.70
12	13613.24	10734.08	2.18	2879.16	490217.16	0.58
13	13229.25	12073.47	2.45	1155.79	491940.54	0.23
14	15244.42	10351.59	2.10	4892.83	488203.49	0.99
15	12476.91	9100.48	1.85	3376.43	489719.89	0.68
16	9551.53	7220.61	1.46	2330.92	490765.40	0.47
17	16122.62	12379.48	2.51	3743.14	489353.18	0.76
18	13120.22	8969.92	1.82	4150.30	488946.02	0.84
19	8997.37	7509.22	1.52	1488.14	491608.18	0.30
20	15638.79	12060.06	2.45	3578.73	489517.59	0.73
21	16548.37	11173.07	2.27	5375.30	487721.02	1.09
22	11487.58	8438.63	1.71	3048.95	490047.37	0.62
23	13668.97	8380.00	1.70	5288.97	487807.35	1.07
24	13009.22	9921.85	2.01	3087.38	490008.94	0.63
25	10852.74	7411.89	1.50	3440.85	489655.47	0.70
26	10319.17	10085.54	2.05	233.64	492862.68	0.05
27	16220.91	11063.37	2.24	5157.54	487938.78	1.05
28	15922.02	11676.12	2.37	4245.90	488850.43	0.86
29	12577.32	9614.50	1.95	2962.82	490133.50	0.60
30	10307.57	6780.23	1.38	3527.34	489568.98	0.72
31	13675.43	9969.53	2.02	3705.90	489390.42	0.75
32	20824.63	14850.40	3.01	5974.23	487122.09	1.21
33	7535.95	6783.83	1.38	752.12	492344.20	0.15

34	14514.56	11236.40	2.28	3278.16	489818.16	0.66
35	13052.82	9469.79	1.92	3583.04	489513.28	0.73
36	15037.59	10694.48	2.17	4343.11	488753.21	0.88
37	12937.84	9173.77	1.86	3764.07	489332.25	0.76
38	10527.43	9304.60	1.89	1222.83	491873.49	0.25
39	15446.17	11286.26	2.29	4159.92	488936.40	0.84
40	12521.97	8846.31	1.79	3675.66	489420.66	0.75
41	16445.03	9594.73	1.95	6850.30	486246.02	1.39
42	11602.21	8457.55	1.72	3144.66	489951.66	0.64
43	15804.48	12144.53	2.46	3659.95	489436.37	0.74
44	10357.37	8800.65	1.78	1556.72	491539.60	0.32
45	5963.49	4386.63	0.89	1576.87	491519.45	0.32
46	17781.12	12719.00	2.58	5062.12	488034.20	1.03
47	18291.21	12867.41	2.61	5423.80	487672.52	1.10
48	17505.33	13043.31	2.65	4462.01	488634.31	0.90
49	10785.98	8833.26	1.79	1952.72	491143.60	0.40
50	17626.15	12167.65	2.47	5458.50	487637.82	1.11
51	3909.30	5515.95	1.12	-1606.64	494702.96	-0.33
52	0.00	559.90	0.11	-559.90	493656.22	-0.11

Appendix VII: Script for seven lowest weeks based on profit

```
import pandas as pd
import numpy as np
import matplotlib as mat
import pip
import matplotlib.pyplot as plt

df = pd.read_csv('thesisdata.csv')
df = df.rename(columns={'%GT Measure': 'influence_value', 'week_of_year' : 't',
'actual value' : 'a', 'norm value' : 'r'})
df = df.fillna(0)
df = df.replace('\€', '', regex=True, inplace=False)
df = df.replace('\%', '', regex=True, inplace=False)
df = df.replace('\,', '.', regex=True, inplace=False)
df.a = df.a.astype(float)
df.r = df.r.astype(float)
df.influence_value = df.influence_value.astype(float)
norm_value = df['r'].sum()
df['prof_diff'] = df.apply(lambda row: row.a - row.r, axis = 1)
df['inf'] = df.apply(lambda row: row.r - row.a + norm_value, axis = 1)
df['inf_perc'] = df.apply(lambda row: 100 - ((row.inf / norm_value) *100), axis = 1)


Numbre_of_hits = 7
df = df.round(2)

df1 = (df.nsmallest(Numbre_of_hits, ['inf_perc','prof_diff']))
print(df1)
```

Appendix VIII: Script for seven highest weeks based on profit

```
import pandas as pd
import numpy as np
import matplotlib as mat
import pip
import matplotlib.pyplot as plt

df = pd.read_csv('thesisdata.csv')
df = df.rename(columns={'%GT Measure': 'influence_value', 'week_of_year' : 't',
'actual value' : 'a', 'norm value' : 'r'})
df = df.fillna(0)
df = df.replace('\€', '', regex=True, inplace=False)
df = df.replace('\%', '', regex=True, inplace=False)
df = df.replace('\,', '.', regex=True, inplace=False)
df.a = df.a.astype(float)
df.r = df.r.astype(float)
df.influence_value = df.influence_value.astype(float)
norm_value = df['r'].sum()
df['prof_diff'] = df.apply(lambda row: row.a - row.r, axis = 1)
df['inf'] = df.apply(lambda row: row.r - row.a + norm_value, axis = 1)
df['inf_perc'] = df.apply(lambda row: 100 - ((row.inf / norm_value) *100), axis = 1)


Numbre_of_hits = 7
df = df.round(2)

df1 = (df.nbiggest(Numbre_of_hits, ['inf_perc','prof_diff']))
print(df1)
```

Appendix IX: Automated text script of seven lowest weeks based on profit

```
import pandas as pd
import numpy as np
import matplotlib as mat
import pip
import matplotlib.pyplot as plt

df = pd.read_csv('thesisdata.csv')
df = df.rename(columns={'%GT Measure': 'influence_value', 'week_of_year' : 't',
'actual value' : 'a', 'norm value' : 'r'})
df = df.fillna(0)
df = df.replace('\€', '', regex=True, inplace=False)
df = df.replace('\%', '', regex=True, inplace=False)
df = df.replace('\,', '.', regex=True, inplace=False)
df.a = df.a.astype(float)
df.r = df.r.astype(float)
df.influence_value = df.influence_value.astype(float)
norm_value = df['r'].sum()
df['prof_diff'] = df.apply(lambda row: row.a - row.r, axis = 1)
df['inf'] = df.apply(lambda row: row.r - row.a + norm_value, axis = 1)
df['inf_perc'] = df.apply(lambda row: 100 - ((row.inf / norm_value) *100), axis = 1)


Numbre_of_hits = 7
df = df.round(2)

df1 = (df.nsmallest(Numbre_of_hits, ['inf_perc','prof_diff']))
values = [[df1.iloc[i][0],df1.iloc[i][4],df1.iloc[i][6]] for i in range(7) if
df1.iloc[i][1]>0]
text=('After filtering the worst week is week {} \nwith a profit difference of {}
\nand a influence percentage of {}'
      '\n\nThe second worst week is week {} \nwith a profit difference of {} \nand a
influence percentage of {}'
      '\n\nThe third worst week is week {} \nwith a profit difference of {} \nand a
influence percentage of {}'
      .format(values[0][0], values[0][1], values[0][2],values[1][0], values[1][1],
values[1][2],values[2][0], values[2][1], values[2][2]))

Table = plt.table(cellText=[[text]],loc='center',cellLoc='left',edges='open',)
Table.set_fontsize(24)
plt.axis("off")
plt.show()
```

Appendix X: Automated text script of seven highest weeks based on profit

```
import pandas as pd
import numpy as np
import matplotlib as mat
import pip
import matplotlib.pyplot as plt

df = pd.read_csv('thesisdata.csv')
df = df.rename(columns={'%GT Measure': 'influence_value', 'week_of_year' : 't',
'actual value' : 'a', 'norm value' : 'r'})
df = df.fillna(0)
df = df.replace('\€', '', regex=True, inplace=False)
df = df.replace('\%', '', regex=True, inplace=False)
df = df.replace('\,', '.', regex=True, inplace=False)
df.a = df.a.astype(float)
df.r = df.r.astype(float)
df.influence_value = df.influence_value.astype(float)
norm_value = df['r'].sum()
df['prof_diff'] = df.apply(lambda row: row.a - row.r, axis = 1)
df['inf'] = df.apply(lambda row: row.r - row.a + norm_value, axis = 1)
df['inf_perc'] = df.apply(lambda row: 100 - ((row.inf / norm_value) *100), axis = 1)


Numbre_of_hits = 7
df = df.round(2)

df1 = (df.nlargest(Numbre_of_hits, ['inf_perc','prof_diff']))

values = [[df1.iloc[i][0],df1.iloc[i][4],df1.iloc[i][6]] for i in range(7) if
df1.iloc[i][1]>0]
text=('All zero values are removed\n\nAfter filtering the best week is week {} \nwith
a profit difference of {} \nand a influence percentage of {}'
      '\n\nThe second best week is week {} \nwith a profit difference of {} \nand a
influence percentage of {}'
      '\n\nThe third best week is week {} \nwith a profit difference of {} \nand a
influence percentage of {}'
      .format(values[0][0], values[0][1], values[0][2],values[1][0], values[1][1],
values[1][2],values[2][0], values[2][1], values[2][2]))

Table = plt.table(cellText=[[text]],loc='center',cellLoc='left',edges='open',)
Table.set_fontsize(24)
plt.axis("off")
plt.show()
```