Expertise
and insight
for the future

Samu Timperi

Renewing Analysis Process for Catalog Data

Metropolia University of Applied Sciences

Bachelor of Engineering

Industrial Management

Bachelor's Thesis

29 May 2021

Tekijä Otsikko Sivumäärä Aika	Samu Timperi Tuotekatalogidatan analysointiprosessin uudistaminen 37 sivua 29.5.2021
Tutkinto	Insinööri (AMK)
Tutkinto-ohjelma	Tuotantotalous
Ammatillinen pääaine	Toimitusketjun johtaminen
Ohjaajat	Ere Jääskeläinen, Tuotehallinnan päällikkö Anna Sperryn, Lehtori Sonja Holappa, Lehtori

Insinöörityön tavoitteena oli rakentaa ja mallintaa interaktiivinen analytiikkanäkymä Microsoft Power BI -ohjelmalla, joka sisältää oleelliset mittarit monimuotoisten data virheiden havainnollistamiseen ja vertailuun tilaajan tuotedatassa. Työn tilaaja on elektroniikka- ja teknologiateollisuudessa toimiva yritys, joka toimii globaalisti tuotteiden valmistajana sekä kehittäjänä, lisäten datan määrää sekä sen oikeellisuuden ja laadun ehdottomuutta tehokkaan työn takaamiseksi.

Tilaajan tämänhetkinen prosessi tuotedatan tarkistamiseen toiminnanohjausjärjestelmässä on manuaalinen data taulukoiden läpikäynti ilman analyyttisten kuvaajien mahdollistamaa helpotusta mikä tarjoaisi mahdollisuuden nähdä virheiden tarkan paikan nopeasti. Näiden puuttuminen vaikuttaa datan tarkistamisen tehokkuuteen huomattavasti ja ne tullaan implementoimaan yrityksen datanhallintaprosessiin tämän insinöörityön tuloksena erillisenä Power BI raporttina. Power BI mahdollistaa tehokkaan datan analysoimisen ja visualisoimisen järjestelmästä ladattuja taulukoita käyttäen antaen käyttäjälle tiedon korjaustarpeista.

Työn tutkimusosa sisältää nykytilan analyysin sekä kohdennetun kirjallisuustutkimuksen, joissa määritellään ja havainnoidaan datan ylläpitoprosessin haasteita sekä kartoitetaan niiden minimoimista akateemisen lähestymisen avulla. Tutkimusosa yhdistää teoreettisen tiedon sekä yleiset käytännöt vastaavissa tilanteissa soveltaen niitä työn tavoitteen saavuttamiseen.

Lopputulos työstä sisältää visuaalisen raportin, joka tuottaa yksityiskohtaista informaatiota tietojohtamiselle tehostamalla datan tilan reaaliaikaista hallintaa antamalla käyttäjälle tiedon järjestelmän nykytilasta. Raportti lisää myös analyyttista tietoa kehityskohteista ja niiden priorisoinnista.

Avainsanat	Laatu, Suorituskyvyn mittari, Datan visualisointi, Analytiikka

Author Title Number of Pages Date	Samu Timperi Renewing Analysis Process for Catalog Data 37 pages 29 May 2021
Degree	Bachelor of Engineering
Degree Programme	Industrial Management
Professional Major	Supply Chain Management
Instructors	Ere Jääskeläinen, Product Management Manager Anna Sperryn, Senior Lecturer Sonja Holappa, Senior Lecturer

The objective of this thesis was to construct and model an analytics dashboard with Microsoft Power BI, that includes relevant Key Performance Indicators for comparing varying data errors in product catalog data. The tool is built to a client that operates globally in the field of electrification and automation manufacturing, which in practice means the client is using a vast amount of modern computer systems which has significantly increased the amount of data and has thus resulted in the critical need for data correctness and quality.

The current method of going through product data is to check it from a table view without any additional indicators helping to identify conspicuous data. This method consumes excessive work time and is to be replaced with an analytics program using Microsoft Power BI as a platform. Power BI is able to read data from an ERP export file through Microsoft Excel and showcase and highlight only the products and data cells that the system recognizes as the ones needing inspection.

The research section of the thesis includes a Current State Analysis and targeted literature research that tackles the weaknesses in the case company's data quality process identified through the CSA. This is where theoretical information and common practices are researched and put together to be used in a building solution.

The outcome of this thesis is an analytics dashboard with Microsoft Power BI that includes relevant Key Performance Indicators for comparing varying data errors in product catalog data.

Keywords	Data Quality, KPI, Dashboard, Analytics

Contents

List of Abbreviations

1 Introduction 1

1.1 Business Context 1
1.2 Business Challenge, Objective and Outcome 2
1.3 Thesis Outline 3
1.4 Project Framework Planning 3

2 Method and Material 5

2.1 Research Design and Project Milestones 5
2.2 Project Plan and Schedule 8
2.3 Data Collection and Analysis 9

3 Current State Analysis 10

3.1 Overview of the CSA 10
3.2 Background of Data Quality 10
3.3 Master Data Knowledge Base 12
3.4 Summary 13

4 Existing Knowledge in Literature and Common Practices with Dashboards, KPIs and Power BI 15

4.1 Business Intelligence 15
4.2 Key Performance Indicators 16
4.3 Microsoft Power BI 17
4.3.1 Power BI Desktop 18
4.3.2 Power BI Service 19
4.4 Conceptual Framework 20

5 Building Proposal for a Dashboard 23

5.1 Dashboard Building Overview 23
5.2 Preparing Data 24
5.3 Key Performance Indicators and Metrics Reviewed 25
5.4 Construction of a Prototype Dashboard for Modeling 26

5.5 Final Report Creation Process 29

6 Next Steps After Thesis 33

7 Conclusions 34

References 36

List of Abbreviations

KPI	Key Performance Indicator
CSA	Current State Analysis
BI	Business Intelligence
Dashboard	Business intelligence interface with implemented KPIs
SAP	Enterprise Resource System used by the case company
DP	Drive Products
Drive	Frequency converter

ZTECHDATA SAP-module for catalog data management

1 Introduction

We live in a vastly changing world where technology is around us constantly in everyday life. Businesses run a competitive environment where decisions are made in short cycles with a help of computers. Development is changed into data-based structures which means that companies utilizing data correctly and swiftly by applying analytics and new systems are in a privileged state when compared to those still getting into it.

Companies' management teams have shifted to make business data management an essential part of planning effective decision-making which also creates competition in skilled data-analysts that are capable of analyzing data in a manner that gives managers the right tools to create innovations that are increasingly difficult to copy. Copying technology and innovations has also become easier and faster in the last decades making strategical planning with new analytic programs a bigger priority.

This thesis was carried out as a part of Metropolia University of Applied Sciences' Bachelor's degree in the field of Industrial Management program during a period from autumn 2020 until spring of 2021.

1.1 Business Context

The thesis was done in collaboration with a global technology company's unit in Helsinki and specifically with a master data and sales Tools Development team. The client is currently operating in over one hundred countries with about 110,000 employees around the globe. The company thrives as a leading company in the field with sustainability, productivity and safety as some of the main priorities in the path that takes history back for over 130 years. (ABB 2021.)

Master Data and Sales Development team is responsible for maintaining data in several different dimensions concerning system information as well as developing and implementing new systems to the infrastructure. This work is crucial in a constant manner as the company products contain a wide selection of different variables in data both mechanically and technically which change and update as technology evolves inside the

product lines. This affects directly multiple sections such as marketing and sales departments that require correct data in the systems used to be able to provide excellent customer service. In addition, correct, properly updated data gives the possibility for managers and data analysts in their work to produce reports and analyses that are effectively moving the complex multi-layer business towards the future. All in all, master data makes the whole customer relationship driven business possible.

1.2 Business Challenge, Objective and Outcome

The Challenge in this thesis relates to analytics' building, modification and clear results in the SAP ZTECHDATA module. ZTECHDATA distributes product catalog data into several systems constantly and as new products are published or existing products updated, modifications given by product managers are to be update and added into proper sections. The study focuses on the product catalog data of the most crucial product lines in the client's frequency converter line and the scope includes product lines from their new analytics-based pricing system's first-year plan.

ZTECHDATA does not provide a proper view into the data structure to see rapidly and reliably the statuses of product-specific information from the perspective of how much is filled in comparison to empty cells. Also, priorities visible in between specific fields and product options do not fulfil wanted need. In order to achieve this goal, a comprehensive data quality process must be defined and visualized as logical errors are quite common in the system due to partly manual work and also to get the ability to build habits for a correct way of correcting mistakes that is not only in the mind of one responsible person but available for everyone involved with small to moderate guidance.

The objective of this thesis is to construct and model an analytics dashboard with Power BI that includes relevant KPIs for comparing varying data errors in product catalog data. Theories and common practices will be based on published confirmed sources such as books, studies and scientific articles to ensure correct information when building up enhancement proposals. The Power BI report resembles the main outcome of the project as it will be a re-configurable platform that can be adjusted as data changes in the enterprise resource planning system. This also enables usage of the base file in other divisions and teams that share the main characteristics in data structure. The dashboard is

planned to become a moderately used tool that will be added to the teams' internal tool selection and will also be improved to eventually become a self-updating structure that pulls data through servers daily to ensure that it is always up-to-date for quick analyzing of data quality.

1.3 Thesis Outline

This thesis contains 7 sections. Sections 1 and 2 cover the basics of the project and when going further through sections 3 to 7, a chronological order of the rest of thesis steps are explained. Section 1 is the introduction that explains the business challenge, objective, and outcome of this thesis. Section 2 describes the materials and methods used for carrying out this study and the project plan.

After the introduction and defining the research methods, section 3 contains a Current State Analysis where the current state of the case company's data quality process is described. After the CSA, the report structure contains a literature research and a review of common practices that provide an academic approach with confirmed sources and opinions from experts. Section 5 encases the proposal building process with mainly Power BI content and KPI planning in conjunction with a wide variety of visual examples of dashboard construction followed by upcoming plans that can be taken after the thesis project to enhance the analytics in chapter 6. Next steps are actions that will partly be planned during the thesis and just simply do not fit the thesis schedule and level in conjunction with examples from ready-made reports of the same nature. Section 7 concludes the thesis with conclusions of the thesis process.

1.4 Project Framework Planning

The subject of the thesis is a small branch in a field of data management, which will affect the angle of approach of the study in conjunction with analytics skill level requirement for end users which will be low to moderate. The literature research consists of subjects in data management and dashboard theory, a topic already known by company workers as mandatory courses are being held to all white collars internally in the case company.

Power BI dashboard is planned with managers to provide an enhanced view of ZTECHDATA in a form that can be attached to presentations reports. This consists of right data for given purpose that can be explained to the crowd involved, as well as forming the theme of report view to the company standard meaning color usage, font usage and general appearance. The appearance is an important factor for many reasons such as faster readability with common base structure, in-line with similar reports already in use that in a long-term can affect to costs by saving valuable time with more fluid interaction. Dashboard data is limited to drive products' lines that consist over a dozen of product groups containing sub-groups in an amount that enables statistical comparison from analytical point of view. The database is built in a way that more products can be added into the report with moderate work in the future.

The framework of the report is built with a raw data table already found in the company's ERP system by exporting the data table to Excel where part of inconsistencies such as errors using a comma in the place of a dot with decimal numbers can be corrected and through that cycled in Power BI. In Power BI the first sequence of actions will be transforming the data set to a more configurable form with attribute grouping that is lost in the initial export action from SAP and by adding categorization column and relationships for multi-purpose use in analytics. The outlining is executed by narrowing product series' to a tolerable amount when mirroring the scale of the thesis with given schedule and by focusing this version of the tool to specifically find data cell values that are in a need of revisioning by using customized filters and visualizations.

In theoretical perspective, if this topic is taken into a deeper research, more studies can be added afterwards to support the angle of the report. Also, the dashboard will be left open for later improvements in a way that it touches more surfaces with a broader platform or goes deeper in a specific one. Together, the given outlines of the thesis make prioritizing an effecting element to the report build-up that must be planned carefully in order to avoid a snowball effect that would swell the thesis into a size that could not be handled in time.

2 Method and Material

There are several different points of view to choose from when making a thesis and this chapter presents an overview of the methods used in this research and the material sources used as base knowledge. Figure 1 below presents the project milestones which provide a big picture to restrictions and requirements for research design planning and the project plan chapter visualizes a walkthrough of the thesis.

2.1 Research Design and Project Milestones

The project begins with assessing business challenge, objective and outcome. After the assessment, the research design is more open to variations, while ensuring that the Current State Analysis (CSA) is built to include necessary topics towards developments. After the CSA, literature research phase is executed to ensure credibility in the proposals and to make it possible to combine different theories in practice rather than just following one straight line. The majority on this thesis' research work ends in the proposal building phase where collaboration with the company and target users takes place in a form of open development conversations and a review of the mock-up proposals.

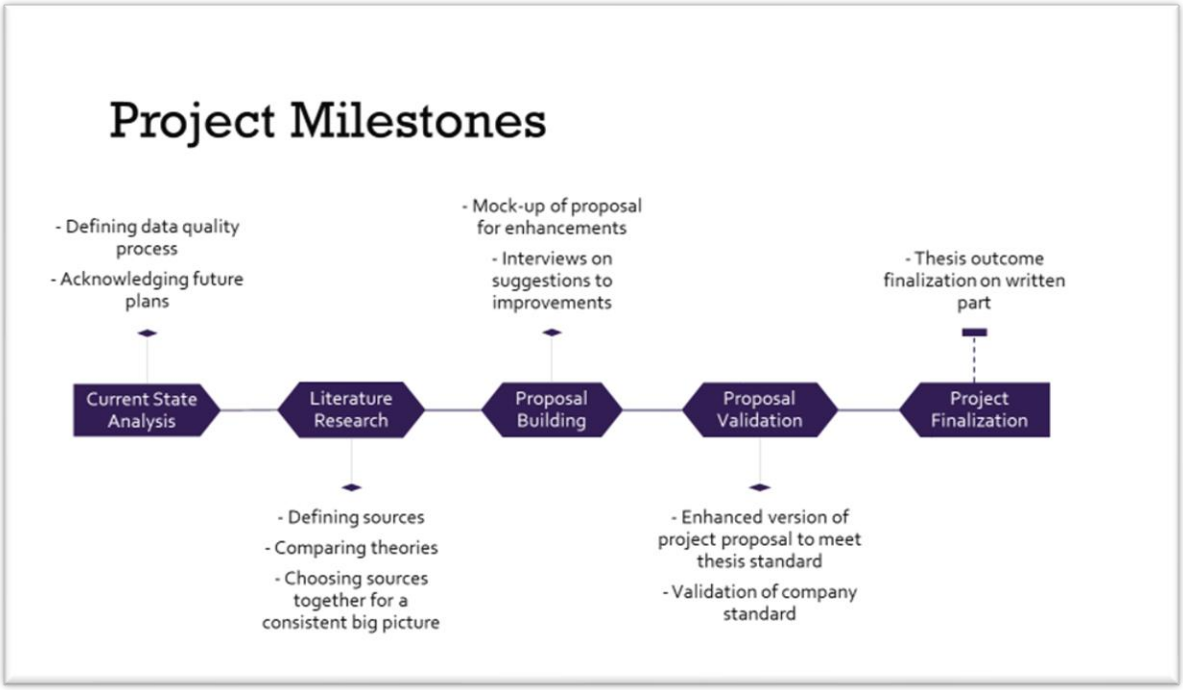

Figure 1. Project Milestones

Business challenges are a fundamental part of engineering as problem solving takes place in everyday life of modern businesses. They might seem obvious at first glance when seen on paper when they are already detected and on-going, but there is almost always an analytical mindset behind them which is either calculated with different formulas of Key Performance Indicators (KPIs), or have come from logical thinking of people working on the subject in different levels of hierarchy. The business challenge and the objective in this research are both comes from the manager of Master Data and Sales Tools team. They effect on seeing a big picture of data quality in a form of right values in right places and the very existence of the data during, and after updates in systems and product portfolios. All the necessary information concerning the starting point of the research is planned in two-person meetings with of myself and the responsible manager with guideline counseling to more specific target points.

Figure 1 above displays the CSA as the first milestone in data collection. The main source of information is uploaded directly from SAP with an ERP module made for the case company that is maintaining catalog data of Drive Products in a centralized place. Migration of this data has been done with Microsoft Excel based export function to ensure

a straight update method that the SAP module supports and that needs better monitoring. The research is made by digging into the system and going though different aspects of previous integration habits and their optimized and unoptimized sides to resolve bottlenecks that are to be enhanced. The CSA is also playing a part in taking future plans of product portfolio changes into consideration in the form of dashboard extension possibilities.

The literature research focuses on business intelligence structures and best practices used by analysts in modern infrastructures and action in conjunction with KPI selections to a specific need that can be applied to the subject in question. These are then fused into policies and practices of Power BI by researching possibilities of the program towards the thesis goal with keeping limiting factors in mind, like a possibility of adding internal only data afterwards. Sensitive data will not be used in this project mitigating the need of researching data protection procedures in Power BI Service publishing possibilities, but for future reference, the approach angle to this must be thought through. Literature research is executed using online databases with published scientific books, confirmed articles and research texts. EBSCOhost in conjunction with Emerald Insight will be the online data sources used in this thesis as they represent highly valued role in their field and are open platforms for everyone registered. In the written sources, timeline of origin is an important factor as technology in evolving constantly with data management coming along with it.

The proposal building is done with multiple layers of data collection together with employees using the tool to give a hands-on feedback of optimizations of most used shortcuts and necessary formability of the dashboard. As the dashboard with KPIs is constructed with Microsoft Power BI, the end product will be a highly interactive digital report with an ability to be updated when needed through a template in SAP, the skill level needed for efficient use must be low enough so that it can be controlled with little to moderate learning curve to avoid excess time usage.

2.2 Project Plan and Schedule

The thesis is carried during September 2020 until second quarter of 2021 for the case company in collaboration with Metropolia. All development conversations and meetings are executed remotely due to the Corona pandemic restrictions in both Metropolia and the client's side. The approach for the thesis was selected due to the need for maintaining master data in an enterprise resource planning systems' catalog data module and its suitability in the depth of knowledge needed to research enhanced ways of executing analytics inside.

The project starts from getting to know the data structure in its current state inside the system, and the methods of how the data is sustained and altered to meet any needed updates. With this structure, current methods of analyzing data is looked through with employees responsible of the work, and alternative methods that are already known to be possible, are thought and compared. This gives the best possible results with the thesis groundlines and the possibilities of enhancement afterwards. After having a big picture of the result planned, literature research will be taken into the project to educationally and scientifically prove beforehand, that correct, viable methods for this study are used and also to minimize human errors that often occur in partly manual working that has been in use for an extended period of time.

As analytics and more specifically Microsoft Power BI report is the main outcome of the thesis, the compliance of the program for this kind of data sets must be investigated to ensure smooth transitions and manageable data refreshes. Throughout the timeline, different points of view are reviewed for the report interface to ensure a pleasant interactive flow between analytic functions. Systematic identification of flaws and errors in the base data are a key factor thus making data filtering the make or break phase of the report flow.

2.3 Data Collection and Analysis

Data used in the thesis is focused on current product catalog data, found in the SAP ZTECHDATA module that resembles the base data storage for the project. ZTECHDATA stores product specifications like measurements, technical values for different parts and mechanical data which are all needed for marketing, reviewing and procurement in sales instances.

This is public open data found from official the client's product catalogs in a different form rather than a large table when updated inside the system. The data is exported via application to an Excel-file in one transaction making it possible to import to Microsoft Power BI in one whole for a more specific examination. After all data that is to be analyzed in Power BI is transferred, they must be categorized for easier browsing, checked and fixed for errors enabling the use of every cell and also compared to the original formats, because of automatic corrections the program does in case of misfunctions of the code execution. Everything from this point forward in the mentioned dataset is done in Power BI with agreed attributes for an optimal view of the dashboard mentioned earlier.

Development conversations with the end users from client side are another part of data collection as they provide hands-on opinions of the thesis framework. Plans of ramp-ups and ramp-downs of products and their options as well as their importance in cross comparison are gotten this way in conjunction with company announcements and reports. The conversations have an advantage of offering aspects from many different angles as people differ in education, work experience and life experience. This way it is possible to avoid information blindness where reports easily drift into when one-sided opinions are focused to without a wider depth of view.

3 Current State Analysis

3.1 Overview of the CSA

Current State Analysis (CSA) is a tool that gives out necessary information for a comprehensive overview in a situation it is applied into. It combines present knowledge and actions made so far, as well as planned improvements and known issues in processes. In an ideal situation CSA would be valid inside every process of business as well as in different levels going up in complexes of company structure to provide bottlenecks for development purposes, but as making the CSA uses a lot of resources, an optimization in targeting the CSA in needed. This CSA is supposed to give the reader a picture of the data quality process in data management, meaning current practices in use with current knowledge and data handling methods.

3.2 Background of Data Quality

Data quality is a key factor in any business, and it becomes even more important when a team is responsible of maintaining master data and a sales tools. Data is often shared between systems and the related organization making a single error in data often a problem for several different layers of hierarchy. This is where the skill of avoiding and swiftly correcting data errors comes into concept as every data transaction has a possibility to develop a sale-ending error.

The company has invested widely into educating its employees by providing courses in different areas depending on the role of employees. This results into better awareness and widespread knowledge within company and increases consistency which is one of the key factors of in data management. A Lean Six Sigma course is one of these courses and because it is entirely built on several stages, it is exceptionally easy to target to specific groups when the level of knowledge digs deeper by levels of the course. Quality in general is also infused into the company values by different means which in the end in many cases is coming down to either learning upfront from data, learning from successes via data with reports or from failures and their comprehensive diagnostics afterwards.

Webinars are another way of educating staff within the company that already takes place from time to time. With varying subjects and an open invitations, internal webinars can give a comprehensive view of quality work over different functions that give tools to withstand various sudden situations and even rapid changes to data management. These lead into more consistent mindset in the work environment making it possible to share roles with better results for example in a case of vacations and momentary peaks in a specific area. With these aspects taken into consideration, the client is all the time taking steps in a form quality to its one of target, Superior Customer Service.

Currently product data completeness is checked from SAP ZTECHDATA module by filtering products and option attributes on a table and inspecting the table values' correctness manually either in the module or by exporting the table to an Excel-file and using basic data functions provided by the program. Both mentioned have advantages and disadvantages. When going though data completeness via ZTECHDATA, the advantage is that the values can be modified in real time through the editing mode if user has the rights and skills to do that. This makes it possible to make small changes when needed in a hypothetical situation where singular error occurs usually through failed sales action. The disadvantage of this is inflexibility of the systems cell-reading action that goes through every data cell visible even if they have already been read just before. This inflexibility forces the module to a slow processing procedure that practically immobilizes the user for up to minutes when loading pages making it very impractical to any more than a small group of changes.

Catalog data also needs screening when bigger changes are made. The change can for example be a new product line coming into the market or a customer specific product that is tailored to one's needs. These broader changes are made and imported from an Excel that is built with a separately for the job that reads column names, row codes and cell values and matches them with the related ones is SAP. Columns are bound by hand to attributes from a pop-up list with over 500 options in structural order that takes significant amount of time, making this sufficient only to bigger data sets as every export must be re-bound in order to avoid column disorder in binding.

3.3 Master Data Knowledge Base

The company possesses a data management library in their Intra-page called Master Data Knowledge Base. This is a hub where a large portion of master data related materials are saved for internal use and it is open globally to all employees. The main purpose of the knowledge base is, that users can become familiar in topics regarding master data and to educate themselves by learning how master data impacts in a daily work of different business assignments. The database is systematically updated and kept up to date to encourage re-visiting.

The content inside is kept well arranged in an order that guarantees the reader a soft landing when delving into a library of this size by offering video presentations of subjects first available followed by series of smaller flyers that are purposely short in length to give out targeted bulletins. These are used as supporting pillars before more in-depth knowledge cases occur as they provide more swift basic information into main business areas and the subjects. The knowledge base also presents a sharing channel where deeper information to specific areas is shared by teams globally to keep an open atmosphere in everyone's work.

Lastly, a webinar section comes to supplement the others and packs internal training sessions with longer presented but always in demand topics. These webinars are longer in length enclosing all angles investigated from a big picture and also speeches by specialists covering their designated area of work.

Figure 2 below is taken from a file in the knowledge base and resembles an example of organization collaboration that brings several areas of work closer together by illustrating the five C's that are bonded towards the same goal of quality. Master Data Organization is responsible Completeness, Compliancy and Consistency. Completeness consists of implementing all needed data into the systems by following instructions provided by an internal rule book. Compliancy means formatting data to work within different programs by checking situationally required capital letters, dots and commas and Consistency covers attribute and classification alignment with the rule book. Data owners like product managers are responsible of the fourth Correctness, consisting of checking data values

after they have been implemented to systems and before they are published to a production environment. The fifth C, being Coverage, is managed by process owners who are responsible of unifying the former four to a working package in the work environment.

This kind of reservoir of knowledge enables people from any assignment, not dependent on their role to enrich general understanding by learning from different backgrounds and cultures that makes diversity especially powerful.

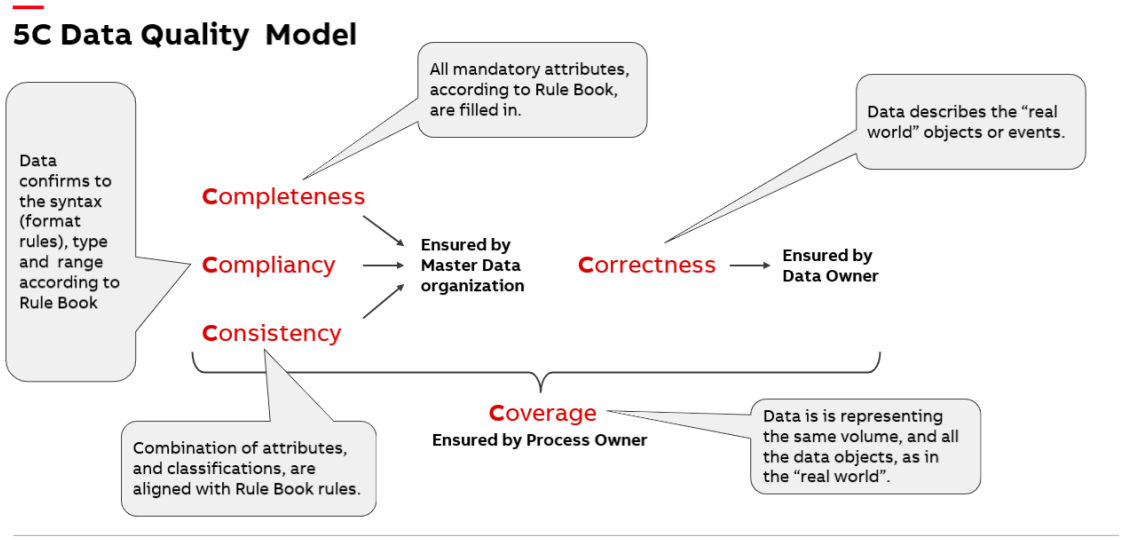

Figure 2. 5C Data Quality Model that resembles a whole kept together with master data management work

3.4 Summary

As mentioned, many occurring practices are systematically leading towards common knowledge and uniform working habits. They have been built to give better tools for teamwork with an aspect that is also emphasizing company values in all C's of Completeness, Compliancy, Consistency, Coverage and Correctness shown in figure 2. More closely, every aspect of these is tied to quality and with that, also to the goals of this thesis.

The company has invested a lot into building a steady foundation for information sharing with a presence of Master Data Knowledge Base. This internal database is kept up to date by systematically adding recent information and files from a global infrastructure to

ensure diversity in file contents. These can then be used in self-learning purposes individually or sometimes for a presentation building process as an internal file.

The weaknesses revealed by the CSA had to do with data maintenance in practice and they mainly consisted of:
1) inconsistent data tables where the system data integrations are coming from
2) the lack of integrated analytics within the ERP system and
3) unoptimized data export function in the ERP system for large exports.

Product catalog data in SAP holds a massive amount of technical data that is in a need of constant reviews as the data tables the information is coming from are being discontinued with the SAP module being its new base place. The technical data tables have been made by several different users without comprehensive specific rules which lead into data errors in data integration to the systems as the integration phase is lacking proper analytics that would notify the user uploading the data of inconsistent values. The same process weakness repeats itself when an error occurs during a sales transaction due to a wrong value which leads to manual corrections in the SAP. The lack of data analytics in the system makes the prevention of these challenging which is the main reason this thesis is executed as it is.

4 Existing Knowledge in Literature and Common Practices with Dashboards, KPIs and Power BI

This chapter helps supporting the thesis towards more scientific aspect by covering available literature and common practices used in building business intelligence solutions. First part covers business intelligence (BI) in general by opening the meaning of it in modern business, its crucial role in the rapidly growing data filled world and possibilities of keeping improvements along with competition. Second part fuses KPIs within BI as they go and grow alongside to each other by covering the main characteristics that are to be taken into consideration when planning KPI usage. Finally, the third part is focused on Power BI, which is a business analytics service developed by Microsoft as Power BI usage is in the main role of constructing the results of this thesis together.

4.1 Business Intelligence

Business Intelligence in present business is rapidly evolving as data environment expands constantly, creating a need for analytics improvements. Public and private organizations are forced to be under pressure forcing them to respond quickly when conditions are changing and to innovate operation changes with limited planning. This form on action requires organizations to be agile in making quick and strategic operational, tactical and strategic decisions. In such decision making, considerable amount of relevant information is required in order to be successful. Processing these in real time to meet a framework in plan usually need computerized assistance in a form of analytic software specifically developed to this and proper personnel capable in efficient use of the programs. (Sharda 2015.)

Design thinking is a major part of business intelligence as it provides theoretical planning tools in visualization work. It can be successfully implemented in several fields, however, in information systems research many studies are focused to enhance strategic, operational and managerial challenges to provide updated possibilities in analytics design. Embedding learning processes of design thinking to employees is a subject that has only minor research done, making design-oriented BI and business analytics a niche skill because the demand towards analysts often lacks knowledge of what can and what cannot be achieved. Design thinking is in many cases copied from reports used inside of a company before and changes to graphs are only implied after seeing examples as a by-

product from system development updating projects' consults leaving internal innovative thinking to minimum. This can be mitigated with more open atmosphere towards changes in old habits and should be thought through is schedules to make it possible. (Chongwatpol 2020.)

With systems and strategies evolving in BI, companies with success in implementations are achieving up a thousand percent cost savings in designated areas while also giving them an edge to competitors. On the other side of the equation, companies with great challenges in BI are losing more in revenue than the investments toward analytics. (Jourdan; Rainer; Marshall 2008.). These statistics support in giving planning and drafting of business intelligence projects a massive importance as delivering proper outcome is often a big investment. The strategy of prototyping with smaller test samples can be executed in data driven companies with tolerable costs, assuming that the will towards these are seen with a long-term value even if the initial results would be underwhelming. With this formula, the investment is also done towards developing common knowledge of employees thus it can be seen as general quality increase. (Sharda 2015.)

4.2 Key Performance Indicators

Key Performance Indicators (KPIs) are still a relevant subject and well used even when in modern strategies big data, artificial neural networks and blockchains are given more and more attention over them, as measuring each of the latter is in the end based on KPIs (Kuhfahl et al 2018, 37-40.). KPI is a measurable value that can be used to illustrate and demonstrate effectiveness of objectives while evaluating success at multiple levels. They are often used to measure progress towards strategic goals in businesses and can be modified to each and every purpose separately. Managing KPIs can be taken as a way of improving indicators that will later drive benefits. They are most often illustrated as measures and charts to give out results in more effective and easily readable way.

KPIs possess an ability to provide day-to-day operation insights that will make changes in approach methods after abnormalities are noticed. They can be used to alert employees in emerging issues before critical failures happen thus mitigating the need of any major emergency fixing. These operational measures are designed for a use of front-line employees and can be highly detailed because they often need fast-phased fixing and

therefore the owner of a tool they are designated to is highly known to them. Operational measures in master data management give an opportunity to set custom thresholds for data value changes that when surpassed let the people responsible know. (Kuhfahl et al 2018, 37-40.)

Setting and planning KPIs is always dependable of data relationships. As relationships are multi-layered, it is an optimization game that is always incomplete in some extension. This is given away in the word *indicator* that means indicating a direction of something rather than telling the absolute truth. Overcoming the challenge of optimizing KPIs grows as a skill with experience caused by trial and error with error part more effectively teaching personnel in how to act in certain sudden need of change which also makes KPI planning as teamwork to be the most reliable way. (Graham 2020.)

4.3 Microsoft Power BI

Power BI is a business analytics solution developed by Microsoft that is data analytics and a visualization tool for business performance monitoring while also being data structuring enabler. Built in functions are letting the user to modify the imported data in individual cells or by groups by using DAX code language. Power BI has been developed by Microsoft and published as an open concept in 2015 to expand Excel capabilities towards more business intelligence themed whole while keeping Excel a working solution to lighter and simpler need.

Main difference between Excel and Power BI is the latter's possibility to read data structures with greatly more advanced way by combining knowledge and visualizing them in a way that constructs more diverse and complicated graphs to a specific need with an addition of making it possible to download open source add-ons via published, Microsoft led channel. In addition, as Power BI is a visualization themed platform compared to Excel's more mathematical platform, for a beginner it is much easier to stitch together working dashboards as long as the base data is prepared correctly.

4.3.1 Power BI Desktop

Desktop version of Power BI is free, open application made to Windows systems that can be downloaded from Microsoft channels that is often pre-loaded to Microsoft operated systems. The program lets user import and combine data from dozens of different common sources like Microsoft programs and different databases that can be re-structured for analyzing needed and after preparation used together to fuse everything into interactive analytics. This version is used to construct reports with all of its functions and capabilities that are constantly updated and upgraded with an addition of a publishing and sharing function that connects the report to Power BI Service. (Microsoft 2020.)

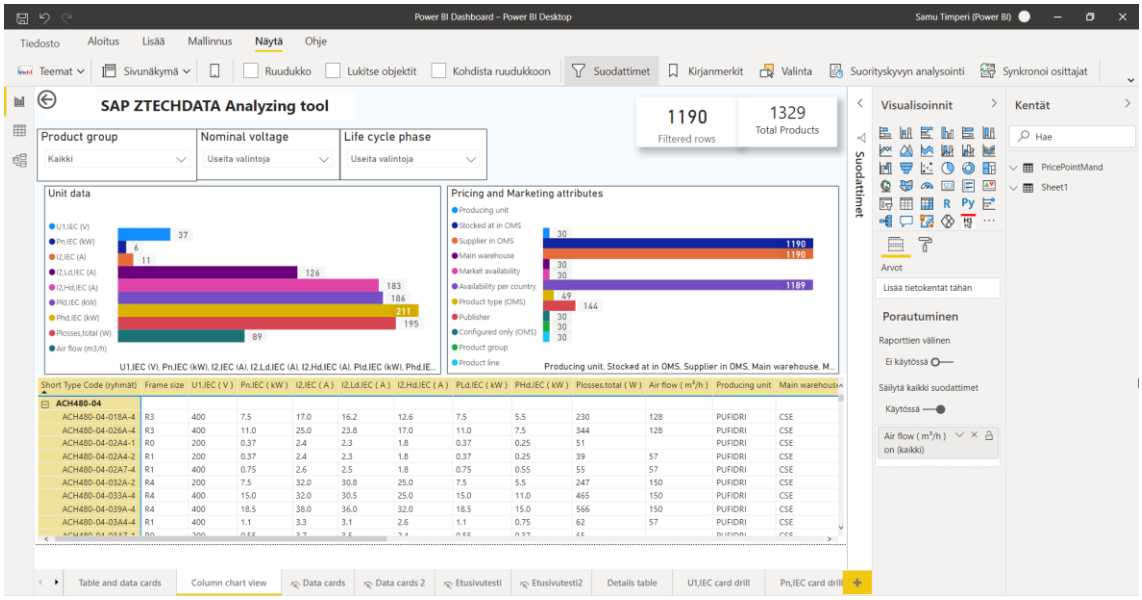

Figure 3. Power BI Desktop example view

Figure 3 shows an example view of the desktop version with data tables and bar charts that interactively change as user is setting parameters through filtering options. To open up the main functionalities, on the left side bar there are three views to choose from. First one from the top is the report view which is showcased in the picture main view. Here you can modify and maneuver though visualizations of the report and change them to suit the report that is to come. Second one is a data set view where all imported data is in a form of traditional data tables where data grouping is made with an option to add, modify, or delete data without Power Query capabilities followed by relationships tab that fuses different data together and gives a possibility to unite repeating data (Microsoft 2011.)

4.3.2 Power BI Service

Power BI Service at its core, is a web-based, cloud operated publishing and sharing channel consisting a collection of software services, connectors and applications to help user in business insight creating processes. Users of the service can share, conditionally modify or simply go through reports providing them an open way to create analytics. Power BI Service is often referred as the Software of a Service (SaaS) of Power BI to give it more contrast towards end user utility. (Microsoft 2020.)

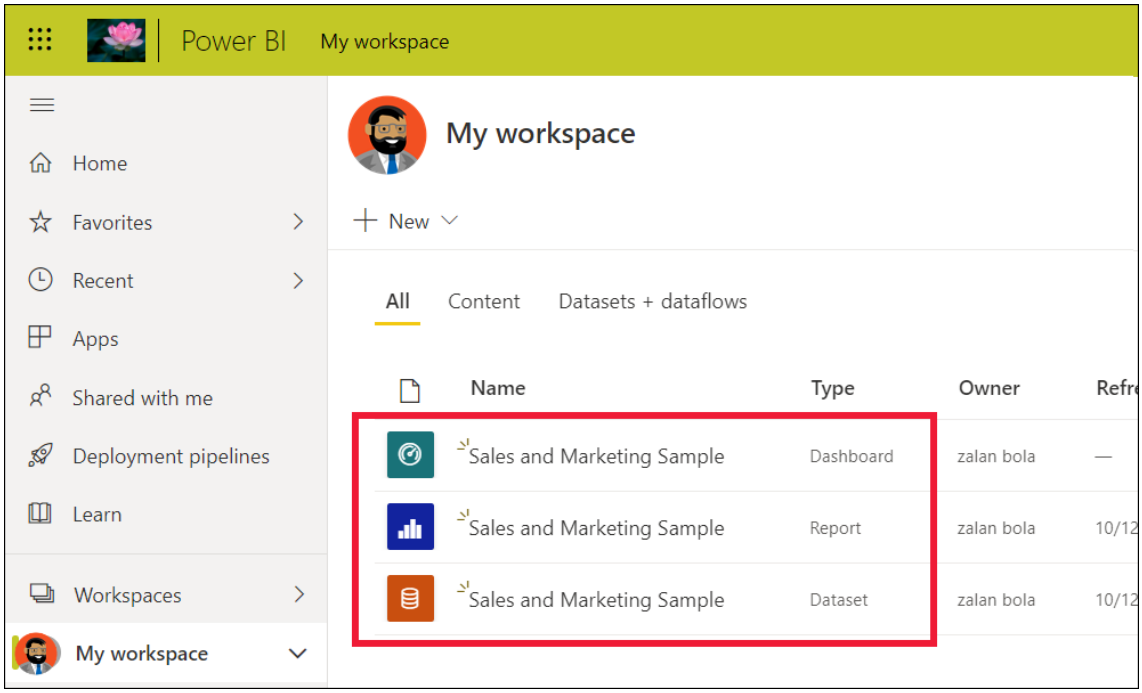

Figure 4. Power BI Service view from *My workspace* which shows user all available reports

Power BI Service view is designed with less interactions than the desktop version minimizing unintended interface moving and frustration when beginners are starting to get to know to the interface and logic (Microsoft 2020.). The main design softeners are the use of a similar overview that is used in Microsoft Windows systems menus and folders in conjunction with interactive parts like *My workspace* -section that is personalized with a help of Office profile data giving out rights to specific contents, sharing and report browsing. Figure 4 demonstrates *My workspace* view with quick menus on the left side giving out similar tools and view than modern web browsers and on the main section, all files that are available and the datasets linked to them.

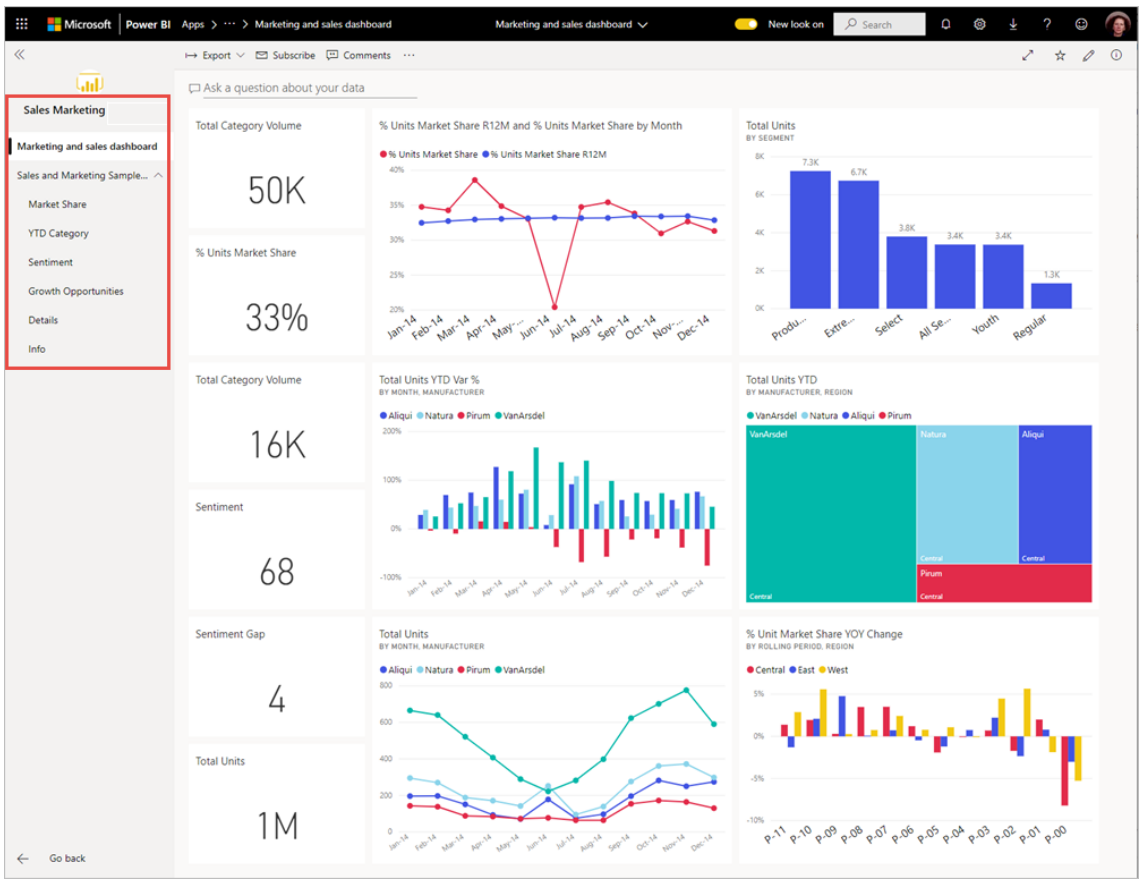

Figure 5. Example report view in Service illustrating the simple design used

Service report view, shown in Figure 5 is mainly a simplified version of the desktop version where all editing rights have been taken off from the user leaving only clean report to investigate. Left side menu is switched from tool view to a tab list that can be used to proceed within the report pages opening different dashboards that can often also be opened through assigned filters and graphs and coded buttons that are constructed with a drill-through function enabling swift custom page changes with a click of bar or a specific point. (Microsoft 2020.).

4.4 Conceptual Framework

Conceptual framework is an analytical tool that can be applied into situations where overall picture needs structuring. The purpose of the tool in this thesis is to merge literature research into the thesis infrastructure by uniting relationships of different subjects investigated in chapter 4, with a use of graph that presents key points and subject merging.

This framework helps understanding how the literature research will be implemented to final proposal.

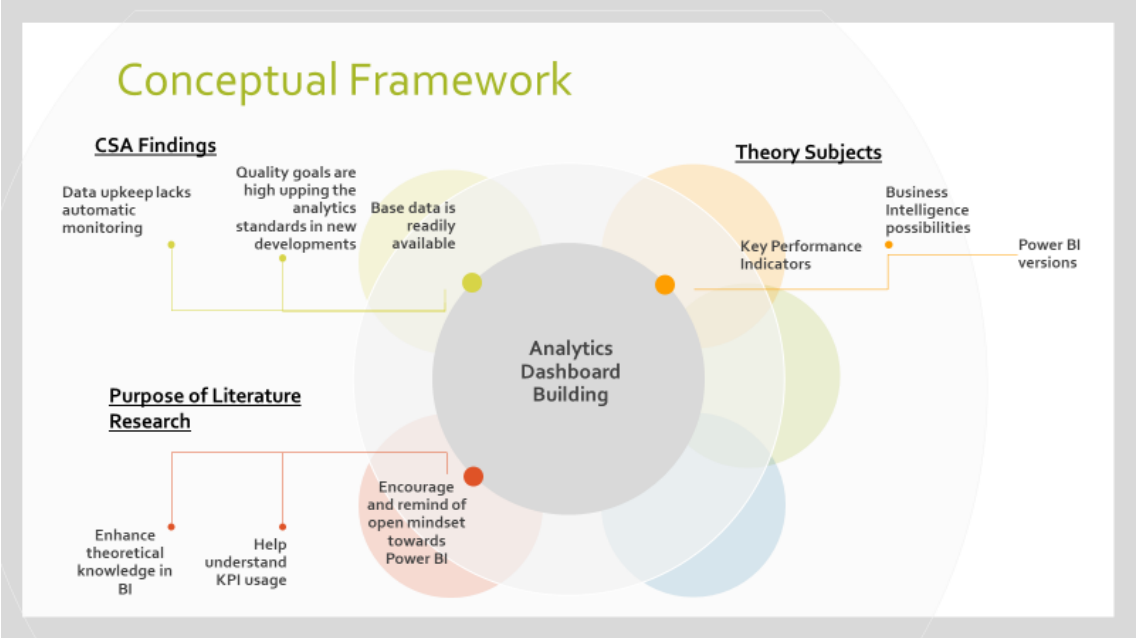

Figure 6. Three main topics approaching dashboard building from different perspectives

Figure 6 depicts the conceptual framework with the approach towards functional theory with illustrative key points. CSA findings are used as a subject finding tool with three main points that make the greatest affect when literature topics are evaluated. These consist of automatic monitoring upkeep absence with present system, as most values are checked by hand from normal table view followed by quality goals that are aimed high with continuous quality programs like 5C Data Quality Model mentioned earlier. Last point on the CSA findings path comes the fact that all needed raw data is already available, therefore completely avoiding the need for finding information about them and accelerating analytics assembly. Theory subject is on the second path leading towards center keeping inside the main bulletins of literature research. Power BI version research is coming first, meaning Power BI Desktop program that is used in dashboard making and Power BI Service where reports can be published and shared with designated people while also making it possible to give different kind of rights to the report dependent on the employee. Business Intelligence possibilities are next on the line as they have been written from general point of view that editorializes in common knowledge of BI

while also looking at it from different angles through researches. Finally, KPI opportunities were examined to give them a comprehensive big picture before selecting the ones used in the report.

5 Building Proposal for a Dashboard

The objective of this thesis was to construct and model an analytics dashboard with Power BI that includes relevant KPIs for comparing varying data errors in product catalog data. To systematically progress towards the goal, development meetings with team leader were kept periodically after improvements in report prototypes and functionalities. Theory was implemented into the tool as complexity raised, and more research was to be underway which made natural growth of the thesis possible. An improvement for data management is often somewhat subjective as it relates to opinions and ways of doing things that are once before learned and kept the same way. Because of this, it is important for all sides to be open towards new mindsets and breaking borders in habits they are used to. As an enhancement proposal, an analytical report was built to modify visual and logical experience of data management. The report is built with an interactive user interface that gives out different results depending on the preferences and choices made by the user towards the end goal.

5.1 Dashboard Building Overview

A dashboard is built in several stages with a logic between the steps. This section explains the planning process of the analytics logic for the initial report prototypes that were analyzed and discussed in meetings within different growth stages of thesis construction. Improvements are thought through in meetings with the thesis client to give opportunities for several structural corrections as well as smaller tunings to be executed towards client's preferences.

Raw data set that is used in the report contains a table of technical values ranging from empty cells and figures of different scale classes into written values and combinations of these three. Table is taken from an enterprise resource planning system with an export function and tied to specific product data and their technical data columns which vary by product. Data set in the thesis is starting from an empty table and the included product series are limited for easier consistency control and simpler infrastructure.

It is necessary for Microsoft Power BI to have large enough sample size of values that are analyzed to be able to work properly with its functions, and in general, data analyzing

will always be more efficient with larger amount data to work with. This is the main benefit that gives precise results when compared to manual scavenging and comparison of traditional data tables. The structure of the table in this report consists of 12 products series with subcategories differing from each other by power ratings, sizes, compliances to a specific need and mounting configurations. When measured by cells, the table size is nearly a million units which enables comparable results as an outcome.

5.2 Preparing Data

Data is taken from an ERP system as raw table to Microsoft Excel via SAP export function where from it is further imported to Power BI. In this phase, it is crucial that the naming of the export file remains as the same and the destination folder is thought with future use in mind. When updating data sets, new export is taken and saved to the original folder with a same name which enables automatic updating of the dashboard whenever new freshly updated spreadsheet is exported. This way of keeping the dashboard up to date makes it possible for the data team to refresh metrics when wanted, in comparison to cloud service-based timed refreshing. Both of the mentioned are viable solutions with main difference being that manual refreshing possesses the advantage to investigate data from a working older version in case of a failure in the system by using previous logs.

Modeling work begins with data formatting. In order for Power BI to be able to read and actuate with the data, it must be formatted consistently and to match the settings offered inside the program. One of the most frequent misconception in this situations is a use of either a dot (.) and comma (,) without realizing the data separation rules and regional settings of Power BI. Power BI uses an engine called Power Query to transform, check and prepare data before report visualization begins. Use of Power Query is not mandatory but, in this case column transformation is needed to identify the top row as attribute list before the first implementation and afterwards, to form data groups for customized analyzing as well as specific extra columns to further classify values.

	Short Type Code	Frame size	Options	No. of input phases (-)	No. of output phases (-)
1	ACH480-04-02A4-1	R0		1	3
2	ACH480-04-03A7-1	R0		1	3
3	ACH480-04-04A8-1	R1		1	3
4	ACH480-04-06A9-1	R1		1	3
5	ACH480-04-07A8-1	R1		1	3
6	ACH480-04-09A8-1	R2		1	3
7	ACH480-04-12A2-1	R2		1	3
8	ACH480-04-02A4-2	R1		3	3
9	ACH480-04-03A7-2	R1		3	3
10	ACH480-04-04A8-2	R1		3	3

Figure 7. Power Query view from Power BI preparations

5.3 Key Performance Indicators and Metrics Reviewed

Defining optimized KPIs is rather subjective topic, as people's personal experiences are often giving opinions only about metrics they have most experience of which multiplies in large companies with similar work tasks. Approach in this thesis gave more freedom to new eyes and insights as only base lines were defined with a couple examples from previous management reports. Way of indicating the first KPI visualizations was purposely leaning to my preferences as optimizations would be made several times as a result of project meetings and planning within the timeline of thesis.

The base line was to first construct indicators into selected small portion of attributes that have the biggest effect current need in data analyzing. These around 30 attributes consisted of general technical product data like powers and electrical currents with a group of basic pricing and marketing data that is also giving information to several sales support tools and where to categorize these three.

Figure 8 below shows the initial filters that can be applied separately to data card visualizations of the prototype. Filtering is projected to three fundamental slicing possibilities; product group, nominal voltage and life cycle phase that enable effective result finding for product managers as in many cases data alternations are made to product subcategory or to a specific voltage range. Product group filter consists of all products in the company division in question with possibility to go down a product tree all the way to a single specific product. This makes it faster to find errors by narrowing selection while following interface data cards that show the amount or hits within the search. Nominal voltage filter can also be applied through product group filter with a caveat of human

error in selecting incomplete results which makes it insufficient way. By selecting main product group and a voltage range separately, user can be convinced that every wanted product is selected with less time of scrolling through options and therefore increasing work efficiency as there are also a lot less options in the voltage group. Life cycle phase filter contains only four options being a blank, active, classic and in design. This is used to lessen false error quantities with data updates in active products as in design -products are commonly having large number of empty cells in them.

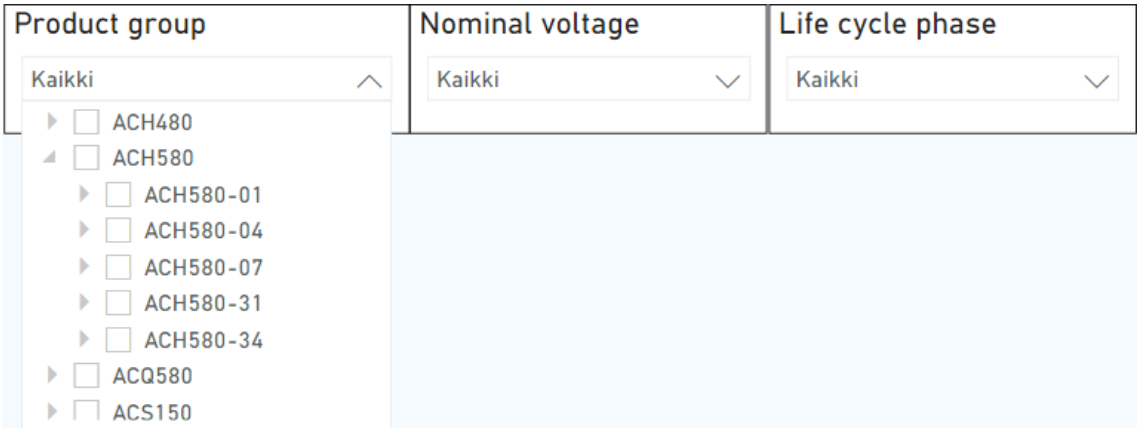

Figure 8. Showing an example of filtering system with product group selection open

5.4 Construction of a Prototype Dashboard for Modeling

When it comes to a business intelligence dashboards, there are certain guidelines to follow depending on the usage. User interface scaling is dependent on screen size, affecting to different presenting scenarios. Group presentations and meeting room presentations enable a use of projectors with massive canvas screens that makes it possible to fit more graphs into one page with smaller font and figure size while with on another perspective, if the end users are mainly using the dashboard as personal analyzing tool, everything has to thought with a mindset of only laptop screen coverage or in other words less graphs in one page and font sizes scaled up to ensure comfortable examination.

With comfortable examination comes also clear and intuitive view of everything in sight without overwhelming the user, which can be made and enhanced in several ways. Data visualizations should only contain mandatory information, but always the information needed to get onwards with a possibility to drill through a specific graph, to get to a more

specific view. This way the user can be led towards the result that is currently being searched.

Power BI has a built-in function that enables linking of tables, filters and data sets to get to a bound pre-fixed page for more information like in many programs and web pages where the user is led through menus. To make these into an interactive and intuitive whole, it is preferred that user is guided to an assumed goal that helps towards to a more fluid experience. The prototype is be built with highlighting procedures that take place according to users' actions. Little contrast changes that depend on what and when something is selected to guide to next intended selection like filtering pane that is be highlighted at the start to activate users' eyes to find that section of the report first. Graph coloring is another comfort factor that is taken into consideration. Data bars are more easily separated by user if they differ in color, as it sends a message that a specific color means one thing and as the same theme continues throughout the report, intuitive perspectives are created. Data labels which sometimes consist of nearly the same text with multiple rows are now completely different to the reader.

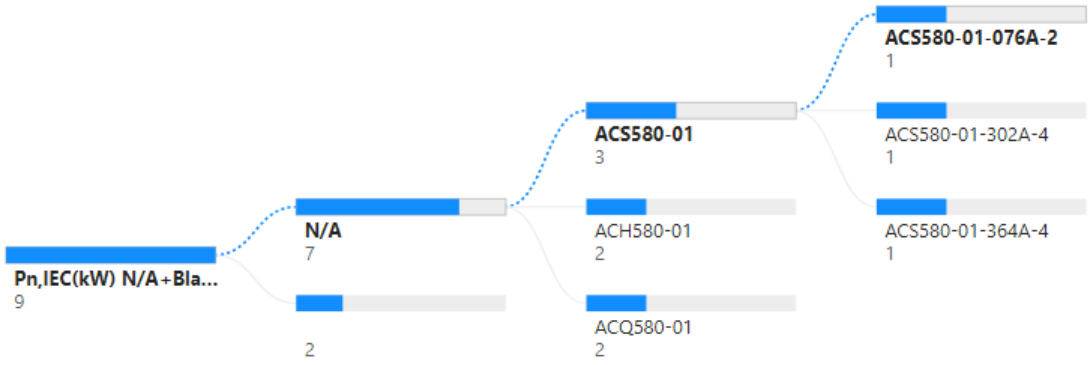

Figure 9. Hierarchy tree showing which products have the searched value in ERP system

In the prototype phase, several models were built. First versions consisted of data card view seen in figures 10 and 11 with different interfaces. In the main page with filtering options and a drill through button to get from singular attribute data card to a more in-depth hierarchy view seen in figure 9. The advantage on this version was simplicity as a whole with a consistent view of rolling values when adding more filters, and the disadvantage was that if a small amount of filtering was used, the view would remain just full of boxes with varying big numbers making it unpleasant to follow. This way every data

card should be investigated as a variable with no intuitive line to follow. Adding an interactive table to this view would have made it easier to follow, as it would have shown the concerning products directly but scaling issues occurred with room simply running out. These were the reasons the first prototype was discarded.

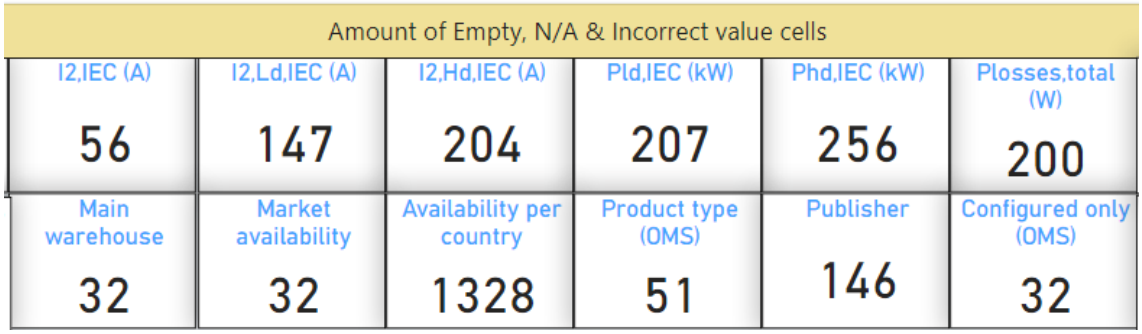

Figure 10. Screen capture of data card function on the downside of the report with amount of attribute values to be revise

Prototyping at a starting point was focusing a lot of potential towards implementing data cards to page view. Figures 10 and 11 showcase two different versions of the data cards which from the prior was selected to stay in the final version but as such wide range of numbers can be hard to follow, it was moved to a secondary page of final report structure.

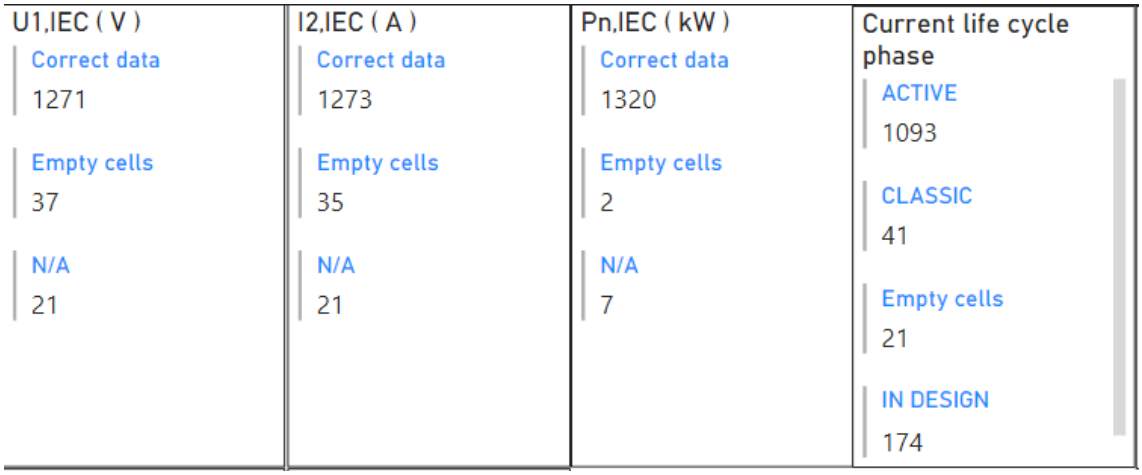

Figure 11. First version of data cards that were eventually discarded

In the following prototypes filtering was changed towards more visual view with interactive bar charts working as selectable filters to data as figure 12 below illustrates. User would pick preferred selection of products and voltages, and as with the first mock-up

the data card set from figure 10 would show the number of cells that need deeper investigating. This prototype style was eventually discarded due to the lack of intuitive walkthrough to the users. This kind of model where multiple bar charts must be selected to effectively get onwards is not common to most employees and also in general the model did not seem as appealing to design thinking wise as main page interface.

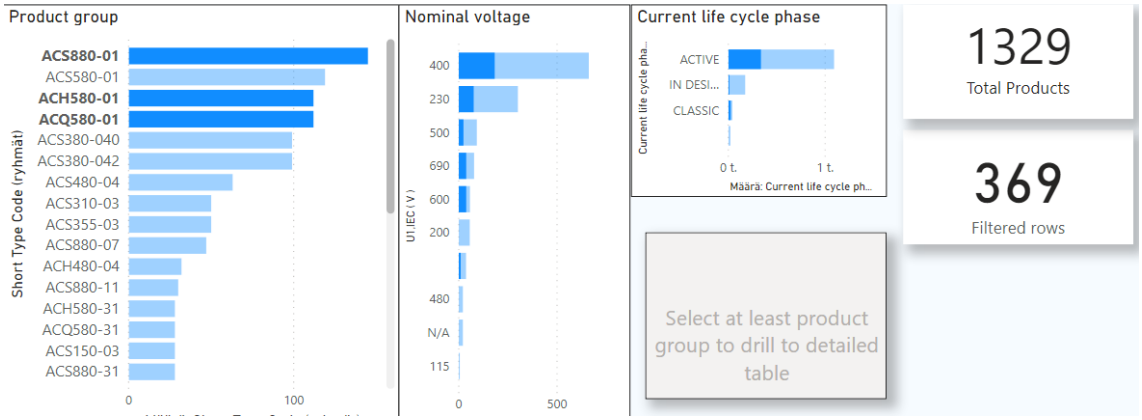

Figure 12. Prototype of the report view with interactive bars working as filters

5.5 Final Report Creation Process

The final dashboard building consisted of combining prototypes into a comparison and optimization work with functionalities, as well as visualization enhancement work. More data bar coloring was taken into use with conditional formatting, quantity labeling being changed from mouse hover-over pop-up to separate individual plates. Power BI bar chart modifications were made, and the general appearance was transformed into clearer and simpler page to inspect, with an addition of more pleasurable user experience when going through report pages by drill through options from bars as seen in figure 13.

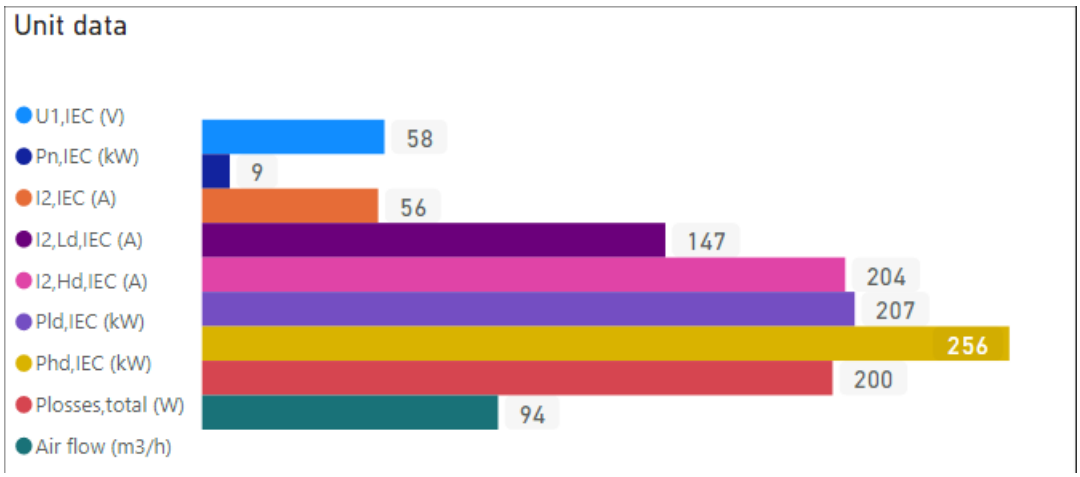

Figure 13. Final model from the dashboard unit data section showing the amount of attribute data cells that need to be revised

To give the dashboard more diversity, two main pages were made to serve varying preferences with front page in figure 14 weighted more towards visualization side and a secondary opening page with same functionalities but visually leaning more towards traditional table view shown in figure 15. Given the purpose of this tool, it was decided after prototype testing to leave out graphs with more technical look available in Power BI selection, like pie charts and waterfall charts as they are developed to be used in illustrations of lists or data with smaller amount of columns.

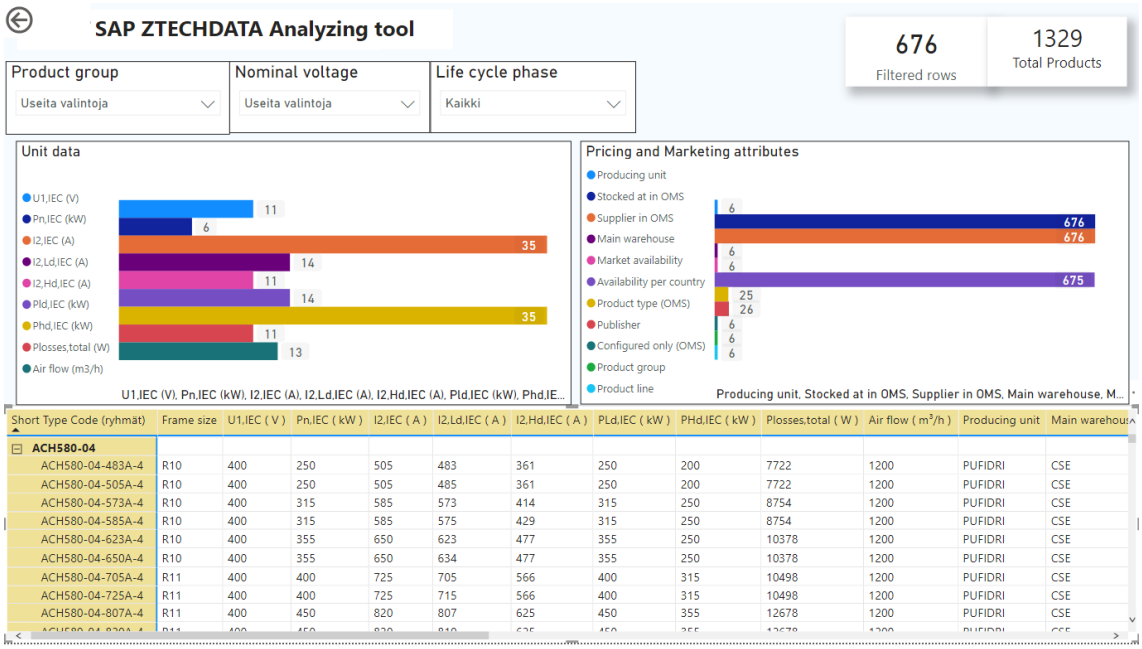

Figure 14. Whole page picture of the final model's front page with two classes of attributes in bar charts and an interactive table on the lower part of the page

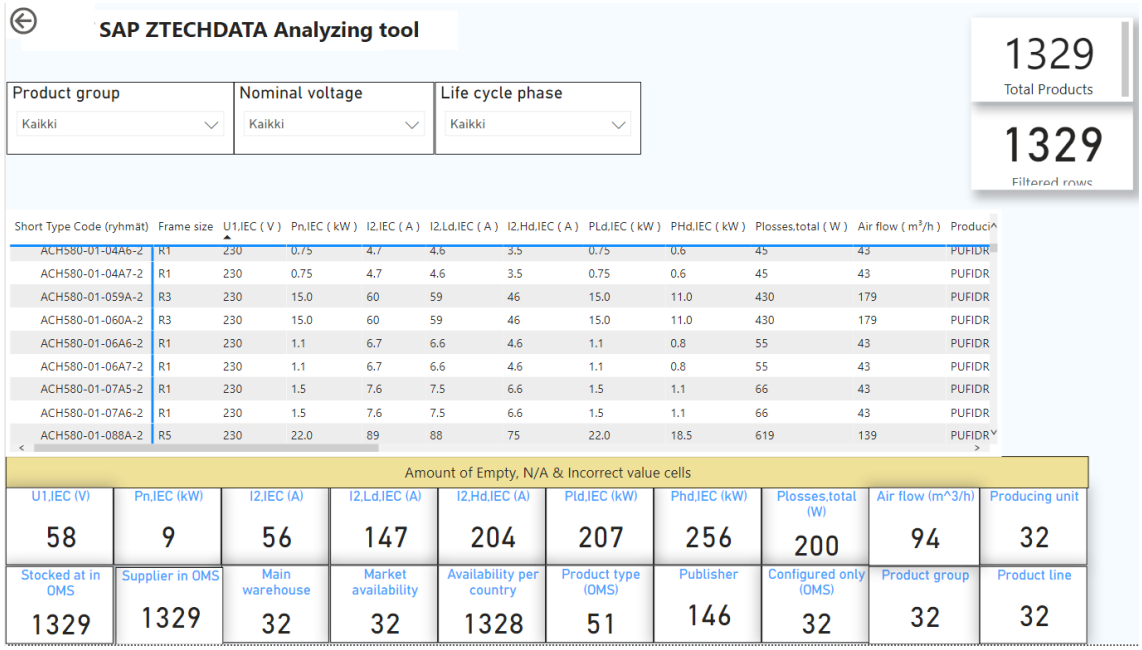

Figure 15. Table page of the final model if user is in a need of more numeric view

In data updating side of the thesis, SAP module where from base data is exported from was taken under development work as its original purpose was not to export tables of this size but rather a lot smaller with lower frequency. Cell reading procedure of the module in original form was to read data from first cell onwards until new information of an individual cell is found and then start over again in a manner that in 500,000 cell size files the first value would be checked again 499,999 times. This slowed down the data set refreshing time significantly leading to a re-programming of the module export function to reduce down-time and excess load in server engine.

6 Next Steps After Thesis

After the thesis project, the development of the constructed analytics will continue in a form of researching possibilities to share the functionalities with other teams' master data management, to develop more unified way of identifying need for revisions in catalog data. As the tool is developed in Masted Data and Sales Tool Management team, it lowers the bar to alter and add functionalities into the Power BI code.

Following the export function re-programming in SAP, data refreshing will be updated and moved into an SQL-server based import, that will automatically check and update the report according to database alternations. The transition to this will happen during the following quarters after the original release. First steps towards this are already taken in planning meeting within the team and the plan is to update the master file once a day during off-hours via cloud service.

More attributes and value thresholds can be added to enrich diversity in conjunction to a distinct page that notifies user if threshold is surpassed in form of misspelling or another equivalent. Product manager's section is also a possibility that would work in slightly customizable way with modules and personal, named paths to get into those with a use configurations. Especially product manager's page would speed up data revisions when product information is checked as a result of an urgent need.

The report in its customizable form can be added to Master Data teams' internal Microsoft Teams page branch and modifying rights can be given to selected people that benefit from it. Report usage instructions are to be implemented inside the tool with tooltips and a separate menu with more in-depth insight versions of the program logic.

7 Conclusions

The objective of this thesis is to construct and model an analytics dashboard with Power BI that includes relevant KPIs for comparing varying data errors in product catalog data. The old method of going through product catalog data in an Enterprise Resource Planning system was changed from checking an individual table view to a customized analytics program using Microsoft Power BI as a platform. Power BI would read data from an ERP export file through Microsoft Excel and showcase and highlight in visualization view only the products and data cells that the system recognizes as ones needing to further inspection.

The research section of the thesis included a Current State Analysis and a targeted literature research that would match the knowledge demand found through the CSA. This is where theoretical information and common practices were researched and put together in order to achieve the theoretical credibility needed for a thesis. After these, a report building process with Power BI was started and this would keep on going for a majority of the thesis time period with improvements being made in small batches consecutively that would be based on feedback given in thesis meetings with the client company.

Data management development solutions of this thesis are based on researched theories and experienced opinions and understandings of people working with data improvement, giving these a foundation to rely on. Finding the right sources in BI report enhancement was a key factor as Power BI functions are focused more towards financial side of businesses making workarounds a necessity and a good challenge to industrial management student.

This thesis work merged a challenge of theoretical, reliable argument finding and implementation of those in practice into an analytical tool with certain limitations together in order to steer manual work towards more automated decision making.

The goals from the client were achieved in the form of analytics that are open for further developments. In addition to a report, ways to go forward from this point got mapped for the client to be able to focus on specific points from report final stage to achieve greatly

more automated inspection process in catalog data quality. The report results will be utilized in future as part of solidifying quality processes in data inspection.

References

ABB. 2021. About ABB. [Online] Available at: https://global.abb/group/en/about. Read: 17.1 2021.

Becker, Louis T. & Gould, Elyssa M. 2019. Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data. Jul-Sep2019, Vol. 45 Issue 3, p184-188. 5p. [Online]. Available at: http://search.ebscohost.com.ezproxy.metropolia.fi/login.aspx?direct=true&db=bsh&AN=138575416&site=ehost-live. Read 22.4.2021.

Chongwatpol, Jongsawas. 2020. Operationalizing Design Thinking in Business Intelligence and Analytics Projects. Decision Sciences Journal of Innovative Education, v18 n3 p409-434 Jul 2020. [Online]. Available at: http://dx.doi.org.ezproxy.metropolia.fi/10.1111/dsji.12217. Read: 10.5.2021.

Graham, Kenny. 2020. What Are Your KPIs Really Measuring?. Harvard Business Review Digital Articles. 9/14/2020, p2-4. [Online]. Available at: http://search.ebscohost.com.ezproxy.metropolia.fi/login.aspx?direct=true&db=bsh&AN=146335740&site=ehost-live. Read: 7.5.2021.

Jourdan, Zach; Rainer, R. Kelly; Marshall, Thomas. 2008. Business Intelligence: An Analysis of the Literature. Information Systems Management. Spring2008, Vol. 25 Issue 2, p121-131. [Online]. Available at: http://search.ebscohost.com.ezproxy.metropolia.fi/login.aspx?direct=true&db=lxh&AN=31560811&site=ehost-live. Read: 7.2.2021.

Kuhfahl, Raquel; Sehlke, Chad; Sones, Jake; Howard, Norbert. 2018. Armed Forces Comptroller. Spring2018, Vol. 63 Issue 2, p37-40. Key Performance Indicators: What Can They Do for You?. [Online]. Available at: http://search.ebscohost.com.ezproxy.metropolia.fi/login.aspx?direct=true&db=bsh&AN=129621367&site=ehost-live. Read: 2.5.2021.

Microsoft, 2020. What is the Power BI service?. [Online]. Available at: https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-service-overview. Read: 23.4.2021.

Microsoft. 2020. What is power query? [Online]. Available at: https://docs.microsoft.com/en-us/power-query/power-query-what-is-power-query. Read: 10.5.2021.

Sharda, Ramesh. 2015. Business Intelligence: A Managerial Perspective on Analytics, International Edition: A Managerial Perspective on Analytics. [Online]. Available at: http://search.ebscohost.com.ezproxy.metropolia.fi/login.aspx?direct=true&db=nlebk&AN=1418680&site=ehost-live. Read 22.4.2021.

Van Winkle, Kevin. 2021. Above All Made by Themselves: The Visual Rhetoric of W. E. B. Du Bois's Data Visualizations. Mar2021, p1-16. 16p. [Online]. Available at: http://search.ebscohost.com.ezproxy.metropolia.fi/login.aspx?direct=true&db=bsh&AN=149390663&site=ehost-live. Read 13.3.2021