Vaasan yliopisto
UNIVERSITY OF VAASA

Miika Hahto

Supplier Data Analysis and Utilization in Supply Chain Management

Case ABB Smart Power

School of Technology and Innovations
Master's thesis in Automation and
Information Technology

Vaasa 2023

ACKNOWLEDGMENTS

I wish to thank all of the personnel at ABB Smart Power who helped me during this thesis. Without their help and knowledge this thesis would not have been possible. I want to especially thank my instructor at ABB Veli-Pekka Sydänoja for all of the guidance and help in the building and solving the problems of this thesis. In addition, I wish to thank the instructor of this thesis from the University of Vaasa Teemu Mäenpää for the guidance and support.

Additionally, I would like to thank my friends and family for the support and understanding during this strenuous process. Special thanks I place to Tikru and Nipsu, the best cats to force me to take a break, even when I did not want a break.

Seven years is a long time to study, but all good things come to an end and in the end the bad times will be forgotten like tears in the rain while the good times stay with you forever.

Vaasa, 10.4.2023

Miika Hahto

UNIVERSITY OF VAASA
School of Technology and Innovations
Author: Miika Hahto
Title of the Thesis: Supplier Data Analysis and Utilization in Supply Chain Management: Case ABB Smart Power
Degree: Master of Science in Technology
Programme: Automation and Information Technology
Instructor: Teemu Mäenpää, KTT
Supervisor: Teemu Mäenpää, KTT
Graduation Year: 2023 **Pages:** 97

ABSTRACT:
This thesis' subject is utilizing supplier data analysis for supply chain management at ABB Smart Power. The research problem is the lack of information on the supply chain. Suppliers and materials are divided into three categories based on their relationship to the supply chain. The main information to be learned from this data is material consumption, material spend, material volume, and movement of materials in the supplier network. The Purpose of this thesis is to build a supply data analysis system to calculate and visualize this information for the supply chain management team. This thesis excludes electronic components due to their volatility in the market. In addition, machine learning algorithm k-nearest neighbours is tested for material price forecasting.

The thesis focuses on the research question "how can the program Power BI be used for gathering, analysing, and utilizing the supplier data". The solution proposed is to automate supplier data extraction from SAP ERP utilizing Microsoft Excel with VBA-programming and utilize Microsoft Power BI for data analysis, visualization, and machine learning to provide the information required to solve the research problem.

The Development process for the solution to the research problem can be divided into four parts which are data extraction, data analysis, visualization and utilization, and machine learning. This solution is built through multiple prototypes and developed based on theory, testing, and feedback. The end product is released to Power BI Service ABB workspace for use.

The thesis is divided into seven chapters according to the constructive research process steps. The first three chapters focus on the background, research question, theory, and technologies utilized in the thesis. The fourth chapter focuses on the research approach and process. The fifth chapter focuses on the solution construct's development process and the sixth chapter on the results of this solution. The final chapter focuses on the conclusions, discussion, and future development of this research.

The development and results of this thesis conclude that the combination of Microsoft Excel with VBA-programming and Microsoft Power BI for data analysis presents an efficient method for gathering, analysing, and utilizing supplier data. Through the data analysis capabilities of Power BI, the data can be analysed, calculated, and visualized efficiently. The Machine learning implementation is possible for Power BI, however utilizing DAX-programming caused technical problems which could not be solved during the thesis.

KEYWORDS: Data analysis, Machine Learning, Supply Chain Management, Automation, Visual Basic For Applications, SAP ERP, Power BI, Data Gathering

VAASAN YLIOPISTO
Tekniikan ja innovaatiojohtamisen yksikkö

Tekijä:	Miika Hahto
Tutkielman nimi:	Supplier Data Analysis and Utilization in Supply Chain Management: Case ABB Smart Power
Tutkinto:	Diplomi-insinööri
Oppiaine:	Automaatio- ja tietotekniikka
Työn valvoja:	Teemu Mäenpää, KTT
Työn Ohjaaja:	Teemu Mäenpää, KTT
Valmistumisvuosi:	2023 **Sivumäärä:** 97

TIIVISTELMÄ:

Tämän tutkielman aiheena on toimittajadatan käyttö ja analysointi toimitusketjun hallinnassa ABB Smart Powerilla. Tutkimusongelma on toimittajatiedon puute toimitusketjussa. Toimittajat ja materiaalit on jaettu kolmeen kategoriaan perustuen toimitusketju suhteisiin. Tärkeimmät hyödynnettävät tiedot tästä datasta ovat materiaalien kulutus, materiaalikulut, materiaalimäärät ja materiaalien liikkuminen toimittajaverkostossa. Tämän tutkielman päämääränä on rakentaa järjestelmä toimittajadatan analysoimiseen, laskemiseen sekä visualisointiin toimitusketjun hallintatiimille. Tutkielma poissulkee elektroniset komponentit, johtuen niiden markkinoiden epävakaudesta. Lisäksi koneoppimisalgoritmi k-nearest neighboursia testataan materiaalien hintojen ennustamiseen.

Tutkielma keskittyy tutkimuskysymykseen “Kuinka Power BI ohjelmaa voidaan hyödyntää toimittajadatan keräämiseen, analysointiin ja käyttöön”. Ehdotettu ratkaisu on automatisoida toimittajadatan keräys käyttämällä Microsoftin Excel-ohjelmaa VBA-ohjelmoinnin kanssa ja käyttää Microsoft Power BI:tä datan analysointiin, visualisointiin sekä koneoppimiseen tarvittavan informaation saamiseen tutkimuskysymyksen ratkaisemiseksi.

Ratkaisun kehitysprosessi voidaan jakaa neljään osaan, jotka ovat datan keräys, data-analyysi, visualisointi ja koneoppiminen. Ratkaisu rakennetaan prototyyppien kautta, jotka kehittyvät teorian, testauksen ja palautteen perusteella. Lopputuote julkaistaan Power BI Servicessa ABB:n workspacessa.

Tutkielma on jaettu seitsemään osaan konstruktiivisen tutkimuksen vaiheiden mukaisesti. Ensimmäiset kolme kappaletta keskittyvät taustatietoihin, tutkimuskysymykseen, teoriaan sekä käytettyihin teknologioihin. Neljäs kappale keskittyy tutkimuksen lähestymistapaan sekä prosessiin. Viides kappale keskittyy ratkaisun konstruktion kehitysprosessiin ja kuudes kappale keskittyy tämän ratkaisun tuloksiin. Viimeinen kappale keskittyy lopputuloksiin, pohdintaan sekä kehitykseen.

Tutkielman kehityksen ja tuloksien perusteella VBA-ohjelmoinnin käyttö Microsoft Excel ohjelmoinnilla yhdistettynä Power BI:n data analyysin kanssa tuottaa tehokkaan metodin toimittajadatan keräämiseen, analysointiin sekä käyttöön. Power BI:n data-analyysiominaisuuksien avulla data voidaan tehokkaasti analysoida, laskea ja visualisoida. DAX-ohjelmoinnin hyödyntäminen koneoppimisessa aiheutti toisaalta teknisiä ongelmia, joita ei pystytty ratkaisemaan tässä tutkielmassa.

Avainsanat: Data analyysi, Koneoppiminen, Toimitusketjun hallinta, Automation, Visual Basic For Applications, SAP ERP, Power BI, Datan keräys

Contents

Pictures

Tables

Abbreviations

ABB	Asea Brown Boveri
DAX	Data Analysis Expressions
ERP	Enterprise Resource Planning
k-NN	k-Nearest Neighbours Algorithm
MDF	Material Description Code
SCM	Supply Chain Management
VBA	Visual Basic for Applications

1 Introduction

Information is a key in managing the modern world. Information is gathered from data which is crucial in today's business and production. ABB has realized the need for data analyzation and the knowledge on how to use it effectively. The utilization of data analysis in supply chain management has grown over the past decade rabidly developing the field (Ahmadi & et. al, 2020). With the current supply shortages of components and raw materials the need for accurate real time information on the movement of raw materials and components has become more and more needed.

1.1 Background

ABB Smart Power has for a long time known the value of data and its use. It has used many methods for data analysis to enhance its business before. Most recently the Microsoft program Power Bi has been utilized in a greater capacity. Power Bi has been mostly utilized for different kind of reports inside the company. This thesis came from the idea to use Power Bi more broadly for data utilization and analyzation especially in supply chain management.

ABB Smart Power is part of Asea Brown Boveri (ABB). ABB is a Swedish-Swiss technology company which operates in over one hundred different countries. ABB has operated for over 130 years, has four business lines and employs over 147 000 people from around the world. These four business lines are Electrification, Industrial Automation, Motion, and Robotics & Discrete Automation. ABB Smart Power is a part of Electrification and produces switches. ABB Smart Power is a technology leader using data driven insights to optimize energy efficiency, reliability, and management of electrical assets. (ABB, 2022)

In supply chain management the different levels of suppliers and materials are referred to as Tiers. These Tiers vary from company to company with Tier 1 being the most significant and closest to the business within a supply chain. In this thesis the Tier 1 materials are purchased directly to ABB Smart Power and the suppliers of these parts are Tier 1 suppliers. Tier 2 suppliers supply Tier 2 materials are components which are purchased by Tier 1 suppliers and are a part of a Tier 1 material. The division of the suppliers is not clear due to suppliers supplying materials in different Tiers. Tier 3 are raw materials used in Tier 1 and Tier 2 materials.

1.2 Purpose of the thesis

The purpose of this thesis is to study and build an easy-to-use system for collecting, analyzing, and utilizing supplier data. The system will enable the end users to utilize the supplier data to get the full picture on the share of the material movement and spend per supplier and a picture of the whole supplier network. Machine learning is tested for the system for predicting future prices for materials by suppliers.

The thesis problems are the lack of information in the supply chain on the supplier material movement, information on Tier 2 material spend, quantities, and volume, the share of suppliers per different Tiers, spread of the supplier network, and the raw material consumption. This thesis aims to find a solution to these problems through this research.

Data analysis on suppliers has been researched from different points of view. However, this research provides focused research on one company's database analyzation from the point of view of the supply chain network and supplier data to offer information on the supplier network. Most of the research done on data analysis in supply chain management utilizes the program MATLAB or nonspecific programs. In addition, data analysis in supply chain management has been researched more on the utilization of big data while this research focuses on small data utilization. The focus of previous research has been on risk management and optimization of supply chains while this research focuses on the utilization of supplier data in the day-to-day operations in supply chain management.

Utilizing machine learning in supply chain management has been researched from different points of view similar to this research. However, it has not been thoroughly researched with the use of Power BI. Previous research has focused on forecasting utilizing different programming languages and optimization while this research focuses more on the utilization of these techniques within Power BI. Overall, there is a limited amount of research on the use of Power BI in supply chain management presenting a research gap. This thesis aims to add to and fill the gap in research by providing new information on the utilization of Power BI for both data analysis and machine learning in supply chain management.

1.3 Research Questions and Objectives

The purpose of this thesis is to study the use and utilization of supplier data in supply chain management. The need for information on material consumption, material spend, material volume, and movement of materials in the supplier network are important information needed at ABB. The information regarding Tier 2 and 3 suppliers is especially emphasised.

The thesis focuses on the research question "*how can the program Power BI be used for gathering, analysing, and utilizing the supplier data"*. Related to these questions this thesis looks at using machine learning to forecast material price changes, maximizing business value with this system, and how to further develop this system.

1.4 Research methods and limitations

Research methods and limitations are important parts of any thesis. Utilizing the correct research methods helps in completing a thesis efficiently. Without limitations a thesis would never be done. With correct limitations a thesis stays focused on the subject and is efficient in its purpose.

The thesis research method is the constructive research method. Constructive research method according to Kari Lukka is an innovative construction producing method for solving real-world problems. Which produce contributions to the discipline it is applied to. Constructions are anything people have created, for example models, diagrams, organizational structures, and information system models. (Lukka, 2001)

Limitations of this thesis are related to data sources, analysis tool capabilities, and utilization of machine learning algorithms. The limitations on data sources are related to how the data is extracted and which data can be used. The analysis tool capabilities are related to the data analysis tool Power BI's limitations. Machine learning limitations are related to how it can be utilized within the limitations presented by the data sources and data analysis tool. This thesis excludes electronic components from the materials and products used due to difficulties in their procurement and the use of alternative materials. Raw materials which are considered in this thesis are copper, silver, zinc, steel, and plastic.

1.5 Thesis Structure

This thesis consists of seven chapters each focusing on a different part of constructive research method. The first chapter introduces the subject, the purpose, the problem, the research questions, limitations, and the methods to solve the questions. The second chapter goes over the theory in behind this thesis and the third chapter explains the technologies utilized in this thesis and prior research. Fourth chapter goes through the research process while reflecting to theory. The Fifth chapter goes through the development process for the solution to the research problem while the sixth chapter presents the results of this development. The final chapter explores the conclusions of this thesis and the future development for it. The Development process for the solution to the research problem can be divided into four parts which are data extraction, data analysis, visualization and utilization, and machine learning.

2 Theoretical background

To understand the thesis' subject, the theoretical background of the subject needs to be explained. This chapter focuses on the theoretical background of the thesis, focusing on the different aspects and background information. This thesis utilizes data analysis and machine learning to research the use of data in Supply Chain Management. Data Analysis is the main method and process used for this thesis' research while machine learning is utilized as an addition. Supply Chain Management is the area this research is applied to. All of these subjects are gone through in their own chapters.

2.1 Data Analysis

Data is a collection numbers and letters which can be utilized in different ways. The difference between data and information is that when value and meaning are found in data, it becomes information, and information becomes knowledge when information and data turn into a set of rules to assist in decisions. (Cuesta, 2013).

Data analysis is an important method of understanding past and future using collected data. It is the process in which raw data is organized to get information. However, data analysis is not about numbers it is about questions, developing explanations, and testing hypotheses (Cuesta, 2013). According to Tiffany Bergin in the book An Introduction to Data Analysis (2018) the ability to analyse raw data is an increasingly important skill in numerous different fields.

Data analysis is a part of statistical analysis (Govaert, 2009). Data analysis uses multiple methods and tools for analysis. Data is spread into categorical and numerical data. Categorical data is values that can be sorted into groups or categories while numerical data is values which can be measured (Cuesta, 2013). Both data types are needed for data analysis due to the different information they uncover from raw data (Bergin, 2018). This thesis uses both categorical and numerical data for different analysis. Using both of these methods is called mixed-methods according to Bergin (2018).

Data analysis is driven by questions and knowing which questions to ask are the main driving force behind successful data analysis (Bergin, 2018). The data analysis process is similar through-out all types of data analysis and according to Hector Cuesta (2013) data analysis process follows these steps:

1. The statement of problem
2. Obtain your data
3. Clean the data
4. Normalize the data
5. Exploratory visualization
6. Predictive modeling
7. Validate your model
8. Visualize and interpret your results
9. Deploy your solution

Data analysis combines and uses multiple different fields including Computer Science, Artificial intelligence, statistics, and mathematics. Data analysis is done in either quantitative or qualitative data analysis. Quantitative data analysis uses numerical measurements while qualitative data analysis uses categorical measurements. This thesis focuses mostly on quantitative data analysis. (Cuesta, 2013).

The concept of big data is a definition often occurring when discussing data analysis. Big data is not synonymous with a large amount of data. Some of the general differences between big data and traditional data are different goals, location, and data structure. The goal of small data is to answer specific questions while big data grants multitude answers. Big data is spread throughout multiple locations while small data is contained within one place. Big data is a collection of both structured and unstructured data while small data is typically structured. Big data is defined by three different factors which are volume, variety of data including databases, images, documents, and complex records, and velocity of the changes within the data (Berman, 2018). This thesis does not utilize big data.

Data analysis relies on computer science due to the vast amounts of data used require computational power for efficient analysis. Machine learning is used regularly in data analysis to help with answering questions, testing hypothesis, and for predicting future based on past data.

2.2 Machine Learning

Machine learning is a form of Artificial Intelligence which uses past data or experiences to make predictions and decisions using historical data. Machine learning uses past data to unveil possible hidden structures and regularity patterns in the data (Theodoridis, 2015). It was best described by Arthur Samuel in 1959 (Theodoridis, 2015), "Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed".

2.2.1 Basics of Machine Learning

Machine learning uses algorithms which in turn use past data to model, predict, and help with analysis of large amounts of data (Theodoridis, 2015). Machine learning relies on data to optimize and learn how to perform tasks using algorithms such as linear regression, k-NN, and decision trees. Excluding linear regression, the other algorithms are classifiers.

Machine learning categorizes data into four types which are supervised learning, semi-supervised learning, unsupervised learning, and reinforcement learning. According to Andrew Ng 99% of all machine learning is supervised. (Campesato, 2020).

Supervised learning uses labeled training data for learning. Labeled data means that for every example in the training data there is an input and an output object. This means that the algorithm learns the correct outputs for certain types of inputs. (Bell, 2014).

Unsupervised learning the algorithm itself tries to find the hidden patterns in data. Unlike supervised learning there is no training data required. The outputs of unsupervised learning are unknown. (Bell, 2014).

Reinforcement learning is a method focused on goal directed learning from interaction. In reinforcement learning the algorithm learns by trying different interactions and what these interactions cause. The three most important features of reinforcement learning are a closed loop system, no direct instructions to begin choices with, and consequences of actions play out over extended time periods. (Barto & Sutton, 2015).

2.2.2 Data Forecasting Using Machine Learning

Forecasting is a method to embrace future changes. Predictions on demands and trends are a crucial part of business and there are three general types of forecasting which are qualitative techniques, time series analysis and projection, and causal models. Qualitative techniques are used primarily when data is scarce. Time series analysis is used when several years' data is available, and relationships and trends are clear and stable. Causal models are used when there is historical data, and the data has already been analysed to know the relationships between the different factors. (Chambers & Mullick, 1971).

This thesis focuses on time-series analysis. This is due to the abundance of data available and clear relationships. In addition, the algorithm k-NN is used as basis for forecasting in this thesis.

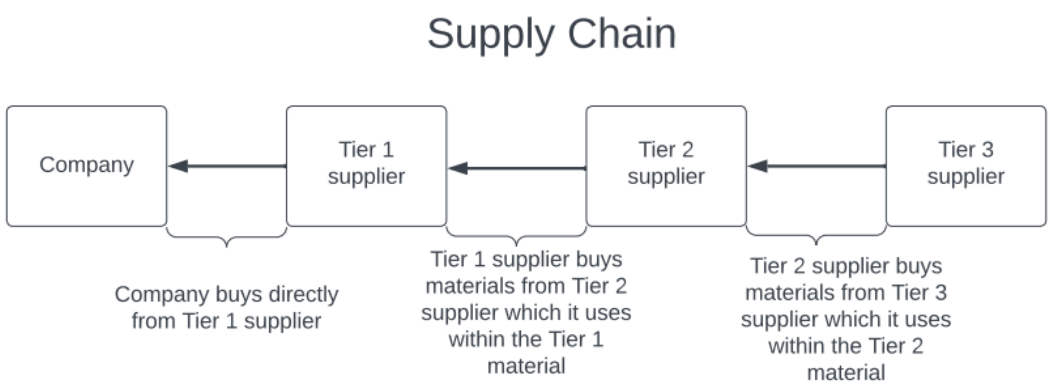
Time series is a type of data which measures how things change over time. However, the time column does not represent a variable and is the primary structure to order a data set. Time series forecasting uses machine learning models which are trained on historical time series data for future predictions. There are four main categories components which may affect the values of data. These components are long-term movement or trend, seasonal short-term movements, cyclic short-term movements, and random or irregular fluctuations. These types of components need to be considered in order to build an accurate model. (Lazzeri, 2020).

2.3 Supply Chain Management

Supply Chain Management is a management process for production systems. The purpose for Supply Chain Management is to design, plan, execute, control, and monitor Supply Chain activities with the goal of creating value, competitive infrastructure, and synchronizing supply with demand (Renko, 2011). The use of Supply Chain Management is to assist businesses in the dynamic international market (Habib, 2011) meaning through Supply Chain Management the flow of goods is optimized in the ever-changing market.

Supply Chain Management can be defined in operational terms as the flow of materials and products, or it can be seen as a management philosophy or as a management process. (Habib, 2011), However according to Drake (2012) traditionally Supply Chain Management's business functions consist of procurement, forecasting, production, transportation, warehousing, customer service, and order management. Supply Chain Management combines three different preceding business functions materials management, physical distribution, and business logistics into one function which further evolves the functions (Drake, 2012).

In the twenty-first century the need for Supply Chain Management has substantiated due to global supply networks and supplies being affected by global changes. Supply Chain Management seeks to coordinate this field and the operations within the chain. Due to this chain the terms Tier 1, Tier 2, and onwards have been created to illustrate the relationships within this supply chain which is shown in picture one. The main part which sets Supply Chain Management apart from traditional business functions is the emphasis on interorganizational element meaning the focus on interactions and collaborations with suppliers and customers with the goal of ensuring the end customers' requirements are fully satisfied. (Drake, 2012).



Picture 1. Supply chain visualization.

At ABB Supply Chain Management is used to handle strategic purchasing and logistic problems. The team handles the acquiring of suppliers for different purposes and handles the contracts which the company's supply chain is dependent upon.

2.4 Prior Research

The three subjects related to this thesis whose theory was explained in this chapter have each multiple research papers written about the subjects. The prior research is looked at from the point of view of how they relate and assist with the research of this thesis.

The use of data analysis for inventory management in a supply chain was studied in the paper Big Data Analysis on Supply Chain Management (Garg & et. al, 2018) which concluded that the use of big data and data analysis can decrease many generated problems in the supply chain. Quality analytics in a supply chain with big data sources has been studied in such papers as Quality Analytics in a Big Data supply chain: Commodity data analytics for quality engineering by Ang and et. al (2016) which concluded that utilizing big data and data analysis effectively means winning the marketplace. The main point of this study was the realization of the help data visualizations and dashboards offer in the use of data on the choices made and the reporting required at businesses. These main points highly effect how this study looks at visualization of the data in use.

In the research paper Big Data Analytics Role in Managing Complex Supplier Networks and Inventory Management by Anusha and et. al (2022) different data analysis tools were tested to manage big data to handle supply chain processes such as inventory management. According to the study the use data analytics tools are reliable for reducing overall production costs and possibility of errors within inventory management and supply chains. The use of big data has been explored in other research papers on different aspects of supply chains such as how to improve the security of a supply chain. Big data can be utilized to develop different metrics and machine learning can use the data to detect deceptive behaviour and potential risks within a supply chain (Glass & et. al, 2013). Other uses in supply chain management for big data have been the use of big data for decision trees for investment decision-making (Liu, 2019) and the information flow interpreting using big data platforms such as MATLAB compared to traditional information transmission models (Hongxiong & Mingrui, 2022).

The process of data acquisition and extraction for data analysis has been studied using different methods. One method for the acquisition of big data for data analysis has been studied utilizing internet of things (IoT) as the tool to gather data from different sources such as file logs, network big data collection, and databases (Li, 2021). Data extraction utilizing browsers for interface and XML data extraction of SAP data was explored by Chen and et al. in the paper A Web-based data extraction system for supply chain management using SAP R/3 (2001). This research subject closely resembles both the subject matter, data sources used, and the interface use of this research.

Machine learning, deep learning, and fuzzy logic have all been utilized in the research on supply chain management and supply chains. Risk management optimization has been the most researched subject relating to supply chain management and supply chains utilizing these methods. In the research paper A fuzzy-random optimization approach using fuzzy measure and fuzzy integral for emergency risk coordination of supply chain, fuzzy logic is used to create an emergency coordination system proposed to combat out-sourcing risks in supply chains (He & et. al, 2014). Deep learning approach has also been utilized in studies for risk assessment and supply chain optimization. Risk assessment was researched in the study Application of Deep Learning Neural Network in Online Supply Chain Financial Credit Risk Assessment by Xu and He (2020). The study on how to utilize deep learning for optimizing supply chain under demand uncertainty was studied by Feng and et. al (2019). Deep learning approach was deemed to be effective approach for both problems based on the results of the studies.

Overall machine learning tools have been utilized in multiple different methods as explained in the literature review paper A Systematic Literature Review of Machine Learning Tools for Supporting Supply Chain Management in the Manufacturing Environment (Breitenbach & et. al, 2021) which details multiple different methods machine learning can be utilized in supply chain management. The paper aims to review the machine learning methods for improved demand forecasting methods and for supply chain disruption detection. According to this paper machine learning methods improve the supply chain management work, and the major findings include the most dominant applications of machine learning for forecasting and prediction to reduce uncertainty from previously used methods. This research paper’s findings direct the solution for how to utilize machine learning in this research even though the research paper focuses more on big data solutions.

This study focuses on the forecasting capability of machine learning which has been researched on different fields and subjects. The forecasting research utilizing machine learning in supply chain management has been looked at from such views as demand forecasting and order lead time. In the study A Novel Dynamic Demand Forecasting Model for Resilient Supply Chains using Machine Learning the researchers looked at how to react to the bullwhip effect in supply chains by utilizing demand forecasting with machine learning. The paper proposes the use of a novel dynamic data driven demand forecasting method using a machine learning approach to bypass complexities associated in these types of forecasts. This method allows for accurate and simpler calculations of demand forecasts with the risks affecting supply chains weighted in (Appadoo & et. al, 2021). Demand forecasting was looked at as a comparison between machine learning algorithms nearest neighbours and Bayesian network in a study by Gaur and et. al in 2015. This study's use of the k-NN algorithm was similar to the proposed method the algorithm was to be used in this thesis. The study concludes from these algorithms that the Bayesian method outperforms the k-NN in both accuracy and efficiency (Gaur & et. al, 2015).

Different look at forecasting was presented in the case study Predicting Order Lead Time for Just in Time production system using various Machine Learning Algorithms: A Case Study by Sing and Soni (2019). This study looked at production lead times instead of demand. This study's subject more closely resembled this thesis' forecasting of price changes as it looked at how to reduce costs by predicting order lead time. The study utilized multiple machine learning algorithms to find the most accurate one. Included in these algorithms was k-NN which is utilized in this thesis. The findings were that machine learning can be utilized effectively and accurately for order lead time prediction excluding various conditions such as traffic, and other factors that change in real time.

Power BI has not yet been utilized for many research papers on its use within supply chain management. Power BI has been utilized on data analysis studies and quantitative and causal analysis which closely apply to the subjects of this thesis. In a study on the use of social media during the SARSCOVID19 pandemic Power BI and R-programming were utilized on data analysis of social media information (Luna & Mora-Arciniegas, 2022). The study concluded that these tools were an efficient way to collect and transform data into presentable enterprise level. The study on quantitative and causal analysis on Power BI focused on optimising Power BI data models when dealing with large amounts of data. Similar to this thesis the optimization of a Power BI model is an important step in fully utilizing the analysis system and making sure the system runs efficiently. The study proposes the use of different techniques to optimize Power BI files. The techniques proposed are reconfiguration of time intelligence functionality, vertical filtering, horizontal filtering, switch storage Mode to DirectQuery or composite mode, hybrid tables, and application of multiple or all techniques proposed. The study concludes that even though the dynamics of every Power BI file differ the use of these different techniques will have a cascading effect to reduce the size of the file and allow the program to run more efficiently (Gogate & Khalwadekar, 2022).

When looking at the research on this thesis a research gap rises. This solution offers new research into the gap regarding the use of Power BI for small data, supply chain management supply network analysis, the use of Power BI for supply chain supplier value review, and the utilization of machine learning within Power BI. The use of Power BI in this research adds to the growing research on the use of the application.

3 Technologies and Data

This thesis uses multiple different technologies to advance toward its goal. The use of SAP data and Power BI for data analysis are the main components to solve this thesis' research questions. This chapter presents the theoretical information on SAP ERP and Microsoft's Power BI program.

3.1 Enterprise Resource Planning & SAP

Enterprise resource planning (ERP) is an integrated system for major business functions like sales, production, and financial accounting. SAP is the most popular ERP program in the world (Sharma & Mutsaddi, 2010). SAP has had many different versions over the years released and currently there are two SAP systems in use. The current newest versions of these systems are SAP ERP 6.0 which is a legacy system (Rau, 2016) and the new cloud-based SAP S/4HANA (Evans, 2021).

Enterprise resource planning is primarily an enterprise-wide system which includes corporate mission, objectives, attitudes, beliefs, values, operating style, and organization (Parthasarathy, 2007). ERP systems form the infrastructure of many organizations. They support the core business functions such as procurement, production, sales, accounting, cost management, and human resources (Loos & et. al, 2005).

ERP systems have been found to directly improve efficiency, information integration, and faster response times to customers. Indirectly ERP systems have improved corporate image, customer goodwill, and customer satisfaction. These improvements are due to the benefits an ERP system generates with business integration, flexibility, improved analysis and planning capabilities, and the use of the latest technology. ERP systems benefit from integrating different business functions and automatically updating the business functions in real-time. (Parthasarathy, 2007).

SAP ERP can be implemented in different ways based on the company using it. SAP ERP offers three different major application areas which are accounting, human resources, and logistics. The accounting application area includes applications for financial accounting, financial supply chain management, and project systems. Human resources application area focuses on human resources e.g., personnel management, payroll, and time management. The logistics application area includes such applications as materials management, production, and plant maintenance. (Sharma & Mutsaddi, 2010).

The limitations of SAP ERP are primarily based on how the system has been built over the years. The biggest disadvantages of SAP ERP are the large memory requirements, the complex structure of the system, and that familiarization requires a lot of time due to many complex functionalities and confusing user interface (Gambit, 2023). According to Peter Catt in the book *Sales Forecasting with SAP Enterprise Resource Planning* (2010) other limitations related to using SAP ERP are the lack of tools for time series feature identification, no ability to perform forecast comparisons, lack of qualitative or quantitative event management processing, and no standard reposting of historical error measures. These limitations are mainly related to real time data processing, data analysis, and end user use. Due to the methods used in this thesis these limitations do not cause problems for the system as the data analysis is handled by Power BI and the data used is updated only once a month meaning the need for real time processing is eliminated.

ABB uses the SAP ERP which is a legacy system. ABB, however, utilizes some HANA databases but does not yet fully utilize the full S4/HANA system. This thesis utilizes the SAP logistics application material management for the SAP data.

3.2 Microsoft Power BI

Power BI is a collection of software services and apps developed by Microsoft released in 2015. It was developed with data analysts and business reports in mind. BI in the name stands for Business Intelligence and the first concept for Self-Service Business Intelligence was introduced by Microsoft in 2009 and adapted into use inside the Excel application as Power Pivot in 2010. Microsoft developed two more business intelligence add-ins for Excel: Power Query and Power View. All these add-ins were combined into Power BI with the addition of new features, visualizations, and use of different data sources. (Ferrari & Russo, 2016).

3.2.1 Basics of Power BI

Power BI is used for data analysis and for transforming data into information. According to Ken Puls and Miguel Escobar in the book *Master Your Data with Excel and Power BI* (2021) Power BI users use Power BI to extract data from data sources, transformation of data to needs, appending data sets, merging multiple data sets together, and enriching data for better analysis.

Power BI can be used in multiple different ways by utilizing DAX-programming, Power Query, and visualizations to transform data into information. The basic process for building a Power BI model according to the book *Expert Data Modeling with Power BI* by Soheil Bakhshi and Christian Wade (2021) is:

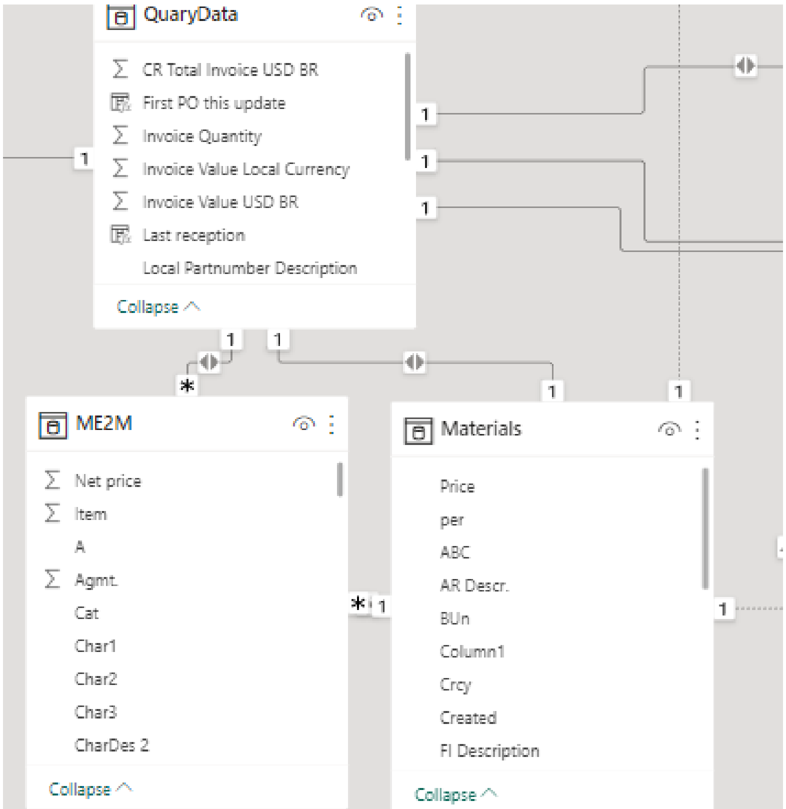
1. Get Data
2. Transform Data
3. Load Data
4. Model Data
5. Visualize Data

Power BI contains four different views for the user to use which are Power Query, Report View, Data View, and Model View. Power Query is called the data preparation layer where data is made available for other layers. Data view and Model view are called the data model layer as they are the views which show the data in use. After Power Query has transformed the data in use to tables, the data view layer allows the user to see the underlying data. The data model view is where created tables and the relationships between the tables are shown. The report view is the data visualization layer and is used to create visualizations of the data. (Bakshi & Wade, 2021).

3.2.2 Data Sources, Analysis, and Modelling in Power BI

Power BI can utilize several different data sources. All the sources of data must be transformed and loaded using Power Query. The purpose of Power Query is to collect and reshape data into the desired format before loading the data into tables for use of Power BI (Escobar & Puls, 2021). These data sources include for example Excel files, Text files, Web Files, Team Desk, and SAP HANA. In addition, Power BI supports direct queries for many sources, for example, SAPBW, SQL Servers, and Azure Data Explorer (Microsoft, 2023).

Data flows in one direction in Power BI. This direction is Power Query -> Data Model -> Visualization. This data flow allows the system to be easily corrected in case of problems and highlights the importance of the data loading phase of building a data model (Bakshi & Wade, 2021). When data is loaded into tables in Power BI the data model view shows the tables and how they are connected. Below is an example picture of connected tables using data relationship connections.



Picture 2. Example of Power BI data view**.**

In the picture above the lines are the connections between the tables and the marking in front of the connection to the table tells if the table's values occur only once in the connected column or multiple times. One (1) in front of the connection means each of the values occur only once in the column while the star (*) means the values occur multiple times in the table. The arrows in the connection lines inform which direction the connection is applied. The direction of the connection means that the table the connection is directed to can utilize data from the other table. When the arrows are both ways the connection is both-ways and both tables can utilize data from each other. Only one relationship connection at a time can be active between two tables in Power BI. (Microsoft, 2023).

Power BI was built with data analysis in mind and offers many different tools for this. Data analysis in Power BI can take many forms such as grouping, creating aggregations, averages, complex calculations, correlations, and forecasting. One of the main methods is the use of key performance indications (Deckler, 2019). The main purpose with any Power BI model is to answer the questions it is supposed to answer efficiently and quickly. One of the main points when building a model is for the model to be easily understood (Bakshi & Wade, 2021). Power BI allows users to analyse data before the data is loaded into Power BI using Power Query or by utilizing the DAX-programming. DAX-programming allows users to create calculations in the form of calculated columns, measures, and calculated tables. DAX-programming is very efficient in measures such as gross margins and percentage of totals but can be utilized furthermore for more complex measures (Deckler, 2019).

Data visualizations in the Report View are created by dragging and dropping fields on to the report “Canvas” where the visualizations are shown as graphical representations of the data within the model. There are 32 default visualizations for example pie charts, waterfall charts, and data tables in Power BI however, additional visualizations can be imported according to needs from the AppSource. In the Report View multiple different sheets can be created to house different visualizations. The visualizations and sheets can interact with each other in Power BI. The Report View can be shared for use through a browser by utilizing Power BI Service. (Deckler, 2019).

3.2.3 Power BI Service

Power BI Service is a hosted web application of Power BI which runs on Microsoft’s cloud platform Azure. This Service is mostly used for viewing and editing reports, creating dashboards, sharing, and refreshing data. The Power BI reports are published in the Service in a workspace only available to users with access. (Deckler, 2019).

Power BI Service is an easy method to share the data model. The main point of using Power BI Service is the security of the model. Power BI Service allows the report to be accessed and utilized but does not allow the core data to be modified through the Service. Access to the workspace is granted to user accounts which can be an outlook account or a work account. (Bakshi & Wade, 2021).

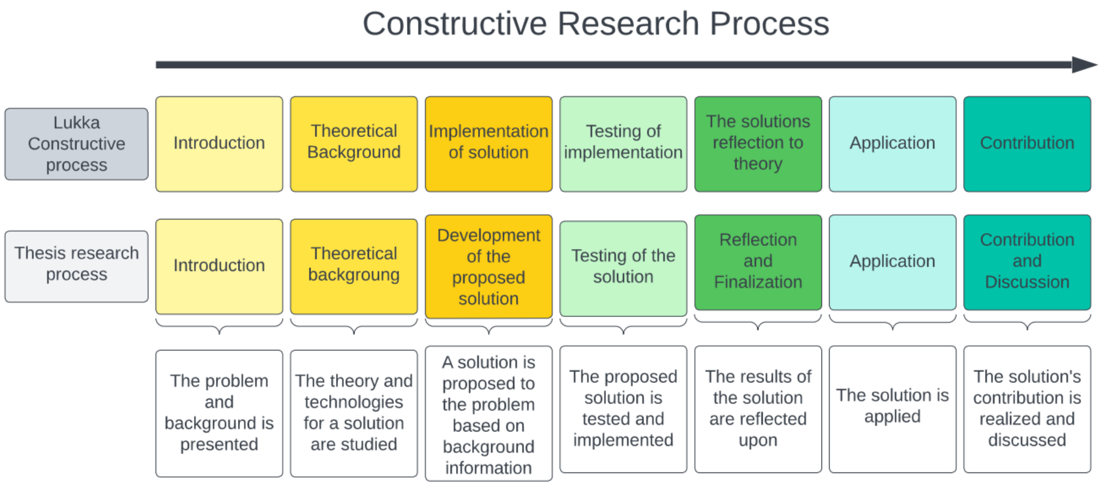
4 Research Approach

The research process for this thesis was chosen to be the constructive research process. This chapter explains the process and what is constructive research process. According to Kari Lukka in the book Case study research in logistics (2003) constructive research is a research procedure for producing innovative constructions, intended to solve problems faced in the real world.

Constructive research characteristic is the researcher's strong and explicit empirical intervention. Opposite to typical objective research, constructive research relies on intervention as a part of the method. Constructive research method is an experimental research method as the new construction should be thought of as a new test instrument to illustrate, test, and define a theory or to develop a new theory. (Lukka, 2003).

Constructive research method's core principles are the research focuses on real-world problems needing to be solved, the research produces an innovative construction meant to solve the real-world problem, close relations between researcher and problem representatives, the research is connected to theoretical information, and the focus of the research is especially on the reflection of empirical findings to theory. Due to these core principles this research is most suited to use the constructive research method. (Lukka, 2001).

Constructive research process is a seven-step process in which the solution is found by going through the steps. These steps are shown in picture three by comparing the research process according to Lukka (2003) to how they are applied to in this thesis. This chapter goes through each of the steps according to this thesis' subject.



Picture 3. Research process compared to the constructive research process according to Lukka (2003).

These steps aim to find the construction which serves as the solution to the research problem. These steps are in the constructive research process theory as follows introduction, examination of potential long-term research, acquiring of deep understanding of the topic practically and theoretically, innovating a solution for the problem, implementing, and testing the solution, pondering the scope of the applicability of the solution, and identifying and analysing the theoretical contributions (Lukka, 2003). This thesis is written and mapped based on these constructive steps.

4.1 Problem & Research

The problem is how to know how materials move through the supplier network, what are the quantities and volume through the network, the total spend of both Tier 1 and Tier 2 suppliers, how much raw materials are consumed by different parts, suppliers, and consumption during a time period. In addition to this the problem is how to present the supplier network in both a visual and importance of supplier way. Most of this data is not available anywhere in the company directly and some of it is not available at all. This problem is focused on from a data analysis angle.

Supplier Data is to be utilized in opening supplier relations and their power levels compared to other suppliers meaning how much value, quantity, and volume there are actually moving between companies. This data is also utilized in reporting on raw material consumption and provide a whole picture on how much activity there is during which parts of the year.

The use of data can help the supply chain management team in negotiations and show the importance of suppliers through the whole supply chain. Currently this information is not known by the team in a detailed manner based on data and to study this demands massive amounts of work which is not feasible for the team members. The aim of this research is to automize the extraction of data, data analysis, and visualization in order to create an efficient system for the use of this information.

This research can be divided into four different parts based on the part of the problem they impact. The problems are data extraction, data transformations and calculations, visualization and data use, and machine learning. The solution to these problems must be one system which is automated to generate the information required from the data used. The solution must be easy to use and accurate for maximum benefit. To find the solution for these problems multiple research papers are looked at and the solution is created based on theory, tools in use, and the restrictions within the company's systems.

4.2 Theoretical background & Solution

There are many research papers written about the use of data analysis and machine learning in supply chain management and in material management as shown in the prior research explored in chapter 2.4. Many research papers have delved into this topic in the past years. The research papers relating to this research can be divided by their focus on data analysis, big data, machine learning, and forecasting.

Supply chain management itself has been researched from multiple angles over the years from different points of view. Such as risk management, quality control & analytics, inventory management, and supply chain supplier selection. The use of data analysis has been studied in supply management on how to efficiently utilize data sources and data analysis techniques however, most of the studies have focused on problems generated using big data.

The solution proposed is to use both Microsoft Excel and Microsoft Power BI programs to automate data extraction in a fast and efficient way and to analyse the data extracted and provide a solution to the problem presented in chapter 4.1. Most of the data comes from SAP which is used as the main data source. This data is transformed and analysed by using Power BI's tools and visualized to be used by end users in an easy and efficient way. Machine learning is to be utilized using k-NN algorithm for price forecasting inside Power BI.

4.3 Implementation & Testing

The implementation of this solution is to build a system in both Microsoft Excel and Microsoft Power BI. The implementation will start by building prototypes to test data connections and solutions to data transformations and connections which is important when dealing with large amounts of data. When building these models, the optimization and efficiency of the model must be taken into account and can be achieved utilizing the research referenced in chapter 4.2. The process needs to be well documented to allow for the future use and service of the system. The service and documentation are documented for the end users benefit. The implementation of the data analysis and data visualization are presented and changed based on feedback and testing.

The testing of this system will be a main focus point to validate the calculations and results the system presents. Through testing the system can be directed towards the wanted direction based on the requirements given by the end users. The usability of the system depends on the end users understanding of the system and the systems ease of use. Testing is required in addition to eliminate possible errors and bugs within the system. Testing will be conducted by utilizing spot testing, comparisons to other information sources, end user testing, and presentations during each phase of the model.

4.4 Application & Contribution

This research can be applied in many ways for supplier data analysis. The main application of this research and construct is to the supply chain management team by supplying them with this tool and the information learned from this research process. This application can be stated based on the testing and feedback from the end user group. At the company the model can be used to in the future help in developing other data analysis reports and data automation methods by allowing others to explore the built model. This research produces a system which allows the supply chain management team to utilize the massive amounts of data within the company in their work and allows for a clearer picture on the supply chain. This product is moulded to the requirements of the end users based on constant testing and feedback. This research looks at the possibilities in the future development of these methods and offers different ideas on how to advance the field in the future.

The main contribution is a new construction in the data analysis and Supply Chain Management world. The utilization of supplier data in this way offers a new way to look at suppliers within a supply chain. The other contribution is the utilization of Power BI in this field. Power BI is a tool with many different uses which can be greatly expanded upon. The utilization of machine learning within the field of supply chain management can be made into a more used resource. Further developing optimizing methods for Power BI researched by Gogate & Khalwadekar (2022) the development of Power BI models can be taken further while the visualization and combination of different data sources utilizing Power BI is explored from a supply chain point of View adding to the research by Luna & Mora-Arciniega (2022).

5 Construct Development Process

This chapter presents the development process for the construct of the thesis. It is spread into six parts. Each part explains the chosen method which is used in this thesis. Through this chapter the solution is presented and described which has been chosen to solve the research question of this thesis "*how can the program Power BI be used for gathering, analysing, and utilizing the supplier data".*

The main sources for data are from ABB Smart Power's SAP. In addition to data collected from SAP a master database Excel file "Plastic component Masterfile" is used to provide additional data to components and parts made from plastic.

SAP provides data for bill of materials, purchasing orders, receptions, and current active contracts and prices for parts. Out of this set of information bill of materials is separately updated due to the data source being different from the other data. Using Excel macros data collection is automatised and will be done once a month when prices have been updated.

Data analysis is executed utilizing Power BI software which is designed for massive datasets. Power BI holds an assortment of different tools for data manipulation and connections while also providing tools for visualizations of the datasets.

Machine learning is to be applied and tested for forecasting use. The main algorithm to be used is k-NN which is a simple machine learning algorithm. This thesis test whether machine learning can be a useful asset for Supply Chain Management in this context and will provide information on whether to further develop the idea.

Testing is mainly done through using a prototype model of the system first. This system is tested by the supply chain management team and the instructor for this thesis at ABB. Through this testing technical difficulties can be found, and the layouts and data accuracy can be optimized for the users. Among the testing done for data and information accuracy is the user interface layout and functionality.

The system will be released into ABB’s workspace in Power BI Service which is a browser-based version on Power BI. Users using the system in the browser version cannot modify the system’s data build or design. The main program will be locked in for only certain users to modify and will be stored on ABB’s network drives.

The prototype is released first for feedback and testing. Through this feedback and testing the final version can be built and tested for additional functionality by the users. This testing allows for feedback on the usefulness of the version and what is desired by the end users.

5.1 Automated Data Gathering and Other Data Sources

Automating the data gathering is done utilizing Excel Macros which utilize VBA-programming language. Using macros exporting the correct data in the correct form from SAP can be automated. The data from SAP is exported from each transaction by creating text files which can be then imported into either Excel or Power BI using the Power Query tool. Text files are small compared to Excel or CSV files and can easily be transformed to the correct form for the application which use the files through Power Query. This transformation creates columns, rows, removes unneeded data, and transforms data types to correct datatypes i.e., number values to decimal number data types, and text values to text value data types.

5.1.1 Original Design and Problems

Originally data was to be gathered in real-time utilizing SAP Hana database which was to be connected to the Power BI model directly utilizing Power Query (Microsoft, 2023). However, this method was not feasible due to the company not using SAP Hana. The next option was to use SAP Business Warehouse connector which would update the data with a lead time of one day (Microsoft, 2023). However, this method did not work due to missing data cubes inside SAP. These data cubes need to be created for each datatype and they did not exist at the company during the development of this thesis. Due to these problems the choice was made to use Excel Macros and VBA programming language to automate the data extraction.

5.1.2 Masterfile

The macro is built into a “Master” Excel file which contains the data used by the Power BI model. In the Excel there will be sheets for the “Master Control”, the data used in queries and sheets for each of the data tables used by the Power BI.

The Macro uses three different SAP transactions and refreshes connections inside the Excel. The Macro uses data gathered from the eSmart system which houses the data on which materials have been bought at the company during which time as the query data. A connection to the eSmart system is built through the “Masterfile”. From this connection the query data is the materials which have been bought the previous year. This data is used to gather data from SAP on the quantities and prices of these materials.

The SAP transactions which the Masterfile uses are ME2M – Purchase Orders by Material, ZEMBK – conditions by contract, and MB51 – Material Doc. Transaction ME2M provides all of the purchase orders for materials which are used as parameters over a chosen period of time, the transaction ZEMBK provides prices per materials per vendor which are in the system, and the transaction MB51 provides the goods receipt information on materials which are used as parameters over a period of time.

The Excel macros run from a Master Excel file and gather the data from SAP to text files which are transformed to correct datatypes and used stored in the Masterfile. The other data sources were connected to the Master file through Power Query tool.

5.1.3 Other Excel Sources

New sources are created to help with data organization and connections for different sources. These data sources are either permanent or are data which cannot be obtained directly from SAP and due to this cannot be automated in an efficient way. These files need to be updated manually by users.

The main new source is Excel file “Material List” which is created using Power BI’s compiling to help create a list of all materials which were assembled and by who the assembly was done and who they bought the parts from. This Excel will be the company’s only source for accurate supplier network. This information is checked and updated by the people responsible for each category of product.

Second most important source file is the Excel file “Z030_12” which houses the data on all of the bill of materials for products. This file is used to get the quantities of Tier 2 materials per Tier 1 material. It is also used to get the information on metal raw material use per Tier 1 materials. This file is created by extracting the bill of material information from SAP transaction Z030_C12 using an older Excel macro used at the company. This information needs to be manually updated in case there are any changes. Due to the sheer amount of data in this file automating updates would prove to be very inefficient. This is also due to how the SAP transaction works as it cannot be exported directly to a text file from SAP.

Third file is “Plastic component masterfile” which is an Excel file created during 2021 to contain all of the information on plastic components used by the company. This file is used to get the information on which plastic raw material each plastic component uses and how much is the weight of the raw material per component.

The fourth source is an Excel file “Material Codes For BOMS”. The Excel file is a list of material codes called “MDF” which describe what class the material belongs to. This data is gathered by hand from SAP and is not required to be updated frequently due to MDF-codes being permanent.

5.2 Analyzation Model

This chapter explains the process on how the Power BI model is built. It explains how the model is built from the first idea to prototype to the final version, the visualizations, and the user interface. In Power BI a model is the Power BI report which is built using different data sources which creates a data model for analyzation. The model uses the data sources to create new data tables, transforms data, and combines the different sources to create a Power BI report. This reports intention is to use this Power BI report to calculate and visualize material consumption, material spend, material volume, and movement of materials in the supplier network for the different Tiers of products. In addition, the supply chain is to be visualized in clear way.

5.2.1 Basic Data Model

Basic Power BI model is built by importing data from sources. After this each data table is transformed through Power Query to the correct data type. Then different tables are connected and combined to get information from the data. This information can then be used to build visualizations which will be what the end users will use in Power BI Service.

When building the data model, the efficiency and size of the model must be taken into consideration. To allow for optimization of the file size and efficiency techniques from the study by Gogate and Khalwadekar (2022) are utilized between the new model builds. The most important part of the optimization will be the reduced recomputing of the same calculations, connections, and transformations.

5.2.2 Data model Prototype 1

At first a prototype version of the model is built in order to test and map out the different connections in Power BI. Due to the high amounts of data the relationships between different data source tables need to be tested. The protype allows testing on how the data can be utilized and how the whole system is built from the ground up. Through the prototype the needed and unneeded data can be acknowledged.

The first prototype utilizes more DAX-programming and relationship connections between the different tables. With this prototype the connections needed, and which data is useful and which data is not can be studied. In addition, data still required by the system can be detected.

The visualization style and control are tested through this prototype to find the direction the development should go. The more different styles and choices are tested the more fine-tuned and useful the visualizations will grow during the development process. Through the visualization testing the calculations of the model can in addition be easily verified.

5.2.3 Data Model Protype 2

Using the lessons learned from prototype 1 the prototype 2 is built from the ground up. The main lessons learned from prototype 1 are how to handle calculations and data connections between tables when dealing with big amount of data. The prototype 1 relied on DAX-programming and relationship connections between tables while the second prototype decreases the use of these techniques mostly due to limitations with the use of data, inaccuracy of the calculations, and the heaviness of running the calculations and connections using relationships and DAX. The second prototype utilizes the Power Query tool more to combine and connect different tables and to create new tables. This method will be the main difference between prototype 1 and prototype 2.

Utilizing Power Query creating new tables, calculations, and connections between tables is more accurate and precise. Building the new model allows for easy use of older data from SAP for comparisons between values from past years. Older values from SAP transactions are kept as their own table which can be appended with the newer values from their own table. This model's main improvement on the previous model is the removal of the circular logic issue which is a problem when utilizing DAX-programming.

The overall data use between tables is tested and established during this prototype. When there is a lot of transformations done to the data the model becomes heavier and more complex to update. Updating the data cannot take too long for the model to be useful and easy to use. Eliminating the connections and transformations which are not needed helps in decreasing the model's heaviness and streamlining the transformations done to the tables helps in the efficiency of the model.

The visualizations of the data in the prototype are shown to the end users for input on their usefulness and which visualizations are needed and which are unneeded. The use of slicers, search options, and pie charts are emphasized in this version of the model. The visual sheets are chosen based on the feedback received from end users.

5.2.4 Data Model Final Version

The final version of the model is built from what is learned from the previous prototypes. This version will be the version used by the end users. This version is similar to the prototype 2, but it is the most streamlined version. All the connections and data transformations are done in an efficient way to limit the amount of computing needed meaning same data is not transformed same way multiple times. In the final version the visualizations will be finalized and fine-tuned for end users while considering the useability and visual style.

5.2.5 Visualization of data and User Interface

The visualization of the model was highly affected by the version of the model in use. The main point in the visualizations is to present information gathered from data for the end user in an unambiguous way. The visualizations controls must be clear to allow the end users to efficiently use this tool for information. The chosen visualizations for the final version must be useful and fulfil this thesis purpose.

The visualizations are what is used as the user interface for the report tool. In the Power BI Service, the visualization view is the only view from Power BI which the user can see. This means that all the controls for using the tool must be created through the visualizations. These controls are different slicers and filters to modify what information is shown by the visualizations.

5.3 Machine Learning Utilization

Machine learning utilization is to be tested in this thesis by implanting material price forecasting by utilizing k-NN algorithm. Power BI can use Machine learning by using R-programming, Python programming, or DAX-programming (Wade, 2021). The use of R or Python requires external programs to run the programming language and have a limit of 150 000 rows which the program will take into account (Microsoft, 2023). Due to this the chosen approach was to use DAX-programming which is used inside Power BI. The use of external programs restricts the update process of the Power BI report. In addition, the limit on the number of rows which can be used in R and Python programming restricts their utilization further.

5.4 Testing

Testing for both the Power BI and the Excel “Supplier Data Analysis Masterfile” are done both as spot testing, data comparisons, and through end user testing. The spot testing is done with the Excel by changing date values and by running the Excel macros multiple times to eliminate errors. Testing for the Power BI through spot testing is done by using a Tier 1 material which has multiple suppliers who use different Tier 2 suppliers and contain multiple Tier 2 and Tier 3 materials to see if the information is correct on regarding the product. Data comparisons are done by utilizing an older Power BI tool developed for raw material consumption reporting and by manually calculating values exported directly from SAP. End user testing is done by releasing each version of the Power BI for end users to test whether there are discrepancies or errors, and to test the useability of the user interface in the visualizations.

5.5 Release for Internal Use

The Power BI is to be released through Power BI Service for use inside the company. Power BI Service provides a browser-based use for the Power BI report only showing the visualization view of the report. The report needs to be manually uploaded each upgrade cycle into the Service meaning it is not updated in real time. This reupload requires the user updating the data to have both Power BI desktop version and the Power BI pro licence. (Microsoft, 2023)

The use of Power BI through the Service is very limited. The users can modify the visuals only by using filters or controls provided in the visualization. This means only the data shown is changed in the current visual. No other data can be modified or seen through the service which secures the program from being altered. Power BI Service can be used by any user without the need for Power BI desktop edition. User will only need approved access to the Power BI Service server from an authorized user.

5.6 Documentation

The project is heavily documented to allow the system to be supported, maintained by other users, and to allow the system to be used for years to come. Six different documents are written for how the system is created and how to use and update it. The six documents are FAQ, Basic Instructions & Examples, Material list & BOM Instructions, Plastic Database Instructions, Supplier_Data_Analysis_Masterfile full documentation, and Supplier Data Analysis Power BI Full documentation. In addition, the VBA-programming is documented thoroughly. The tables in Power BI data model view are explained in the table's properties descriptions to provide a fast way to identify the use of the tables.

6 Construct Results

This chapter explores and explains the results of this study. The results are presented in similar chapters as chapter five. The changes made to choices chosen in chapter five are explained in this chapter. The development of each segment went through multiple changes during the application and implementation. More information was added through each version of the system based on feedback received during the development.

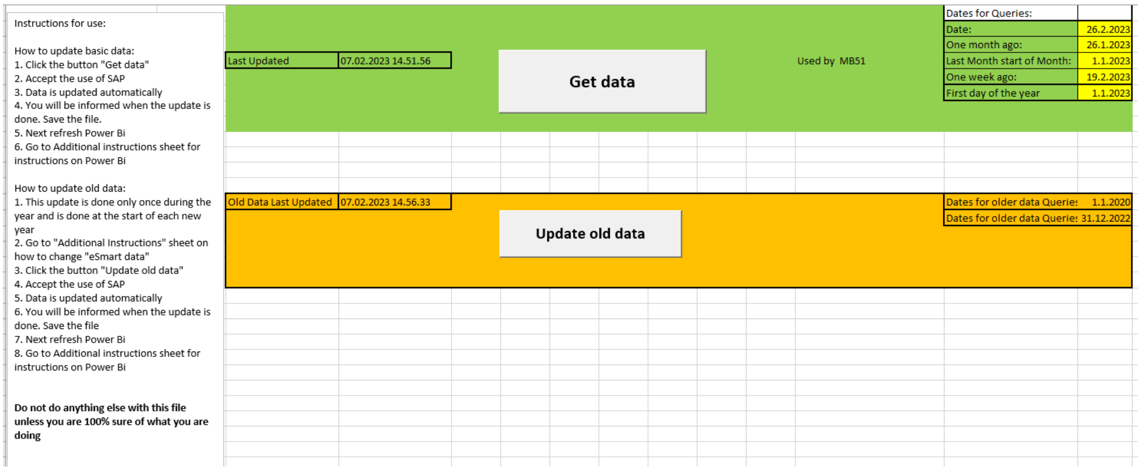
6.1 Automated Data Gathering and Other Data Sources

Automated data gathering was accomplished using Microsoft Excel Macros utilizing VBA-programming, manually creating source files for data, and older databases made at the company. Data gathering from the sources was automated through Excel with VBA programming and macros for data on movement of materials and the quantities. Other data sources were collected by hand or were older data master files which have been created at the company. The different parts of extracting data and sources are explained in this chapter. This chapter is spread between the Master file and the other data sources.

6.1.1 Masterfile

The Masterfile was originally built as the main source for Power BI data and only utilized one VBA macro for data extraction which was the macro “Get data”. There were many changes to the original model due to technical errors during the update process which this new version corrected making the Masterfile more stable. The other macro “Update old data” "was added to allow for deeper analysis.

The design and visual style of the Masterfile was streamlined to simplify the update process. The original versions contained multiple different sheets for different data. In the picture below is shown the final version of the Masterfile’s user interface.



Picture 4. Supplier_Data_Analysis_Masterfile user interface.

The original version of the Masterfile consisted of seven different sheets. The version used only the “Get data” macro to extract data from SAP. The Macro saved three different files from SAP to text files which were then imported to the Masterfile Excel using Power Query tool creating three sheets ME2M, MB51, and ZEMBK. The other data used by the Power BI was connected using Power Query as well.

This version was overhauled later when new data needs arose and due to constant problems occurring during the update process. Problems arose due to data first being transformed in the Masterfile Excel and then transformed again through in Power BI’s Power query. When there were any changes in SAP layout the Power Query transformations did not work properly. When there were two different transformations of the same data the problems were doubled.

Similar problems arose from connecting the other Excel files to the Masterfile. The Power Query transformations were done first in the master file and then through Power BI. This method did not cause similar problems as the data files from SAP but made the Masterfile too cumbersome. The choice was made to directly connect the other source files to the Power BI without a connection to the Masterfile.

The Excel Masterfile uses Power Query to connect to ABB’s eSmart database through SQL Server Analysis Services Database. The eSmart database holds data on all of the directly purchased components to all of ABB Smart Power’s different organizations. This database is used to easily get the data on the components which were purchased to ABB Smart Power during the previous year. This data is used as the query data for the SAP queries. This addition to the Masterfile was done to get a list of the components which have been purchased and to allow the data to be updated in almost real-time. The data used for queries had to be focused due to limitations within SAP transactions on how many values could be used as parameters in the queries. In addition, the limitation on how many parameters were used expedited the macros’ running times.

The “Get data” macro is the main macro in use. The macro is used every month to update the data. Originally the macro did three queries in SAP using three different transactions ME2M, MB51, and ZEMBK and produced three different text files. These text files are used in the Power BI as sources. Below is shown an algorithmic presentation of the “Get data” macro.

```
Begin
Save file
Connect to SAP GUI
IF error
     then return to start
else
Copy column A Values from Sheet QueryData

Pop-up "Getting data from ME2M"
Start transaction MB51 // Purchase Orders
input parameters
Change layout
Save as text file

Pop-up "Getting data from MB51"
Start transaction MB51 // Goods receipts
input parameters
Change layout
Save as text file

Pop-up "Getting data from ZEMBK"
Start transaction ZEMBK //Component prices
input parameters
```

```
Change layout
Save as text file
Update connections
Pop-up text box "Data retrieved"
end
```

The macro works by copying the query values extracted from the eSmart connection and connecting to SAP. The Macro then goes through each transaction by inputting parameters which are the material values from the clipboard and date values “to” and “from” which are retrieved from the Masterfile’s Master sheet. These values are the first day of current year and the value 31.12.9999. The transaction ZEMBK retrieves all the data from the database on the materials used as parameters and due to this the transaction does not require any other parameters. The transaction MB51 requires additional parameters which remain static and are written directly into the VBA program. After the data is searched in each transaction the layout of the data is changed to the correct form. The data is saved as a text file in the same folder under the same name each time in other words overwriting the original text file.

During prototype 2 of the Power BI model the need for data from transaction ME2M-Purchase Orders by Material was noticed to be unnecessary and the use of this transaction was eliminated from the macro. During the same prototype the need for more data from transaction ZEMBK arose. The need for price data on materials which are a part of the materials which were first retrieved. These materials for the query had to be manually retrieved from the “Material List” Excel file’s list. Due to this these materials are not updated automatically in any operation. This new query was added to the macro and the data was saved as its own text file. This version of the macro is presented as an algorithm below.

```
Begin
Save file
Connect to SAP GUI
IF error
     then return to start
else
Copy column A Values from Sheet QueryData
```

```
Pop-up "Getting data from ZEMBK"
Start transaction ZEMBK // Direct material prices
input parameters
Change layout
Save as text file

Pop-up "Getting data from MB51"
Start transaction MB51 // Goods receipts
input parameters
Change layout
Save as text file

Pop-up "Getting data from ZEMBK"
Start transaction ZEMBK // Component prices
input parameters
Change layout
Save as text file

Pop-up text box "Data retrieved"
end
```

The macro “Update Old data” is used once a year to get data from the past three years. Originally this macro exported the data from two different transactions ME2M and MB51. The macro runs similarly to the macro “Get data” and is shown below as an algorithmic presentation.

```
Begin
Save file
Connect to SAP GUI
IF error
then return to start
else
Delete Columns A, B, C from Sheet QueryData
Copy Columns A, B, C from Sheet eSmart
Paste clipboard to Sheet QueryData
Copy column A Values from Sheet QueryData

Pop-up "Getting data from ME2M"
Start transaction MB51 // Purchase Orders
input parameters
Change layout
Save as text file

Pop-up "Getting data from MB51"
Start transaction MB51 // Goods receipts
input parameters
Change layout
Save as text file
```

```
Pop-up text box "Data retrieved"
end
```

The macro starts by copying the data from the QueryData sheet then the macro goes through the two transactions similarly as in the other macro with the main difference being the use of different values for the to and from dates to the queries from the Master sheet in the Excel. The Master sheet uses a formula to get the first day of the year three years from the current moment for the “from date” to the queries. For the “to date” it uses a formula to get the last day of the last year. Using these values only the data from the past three full years is searched. After the data is searched using the correct parameters similar to the other macro the data’s layout is changed to the correct form and the data is saved as text files “ME2M_OldData” and “MB51_OldData”.

Similarly, to the “Get data” macro the need for data from transaction ME2M was noticed to be unnecessary. After prototype 2 phase this part of the macro was deleted. Deleting the additional data extraction allowed the macro to become lighter to run as the data amount extracted reduced by 50 percent. Before this the macro required a long time to run due to how much time was required to first search the data in SAP and to save the data. This version of the macro is shown below.

```
Begin
Save file
Connect to SAP GUI
IF error
     then return to start
else
Delete Columns A, B, C from Sheet QueryData
Copy Columns A, B, C from Sheet eSmart
Paste clipboard to Sheet QueryData
Copy column A Values from Sheet QueryData

Pop-up "Getting data from MB51"
Start transaction MB51 // Goods receipts
input parameters
Change layout
Save as text file

Pop-up text box "Data retrieved"
end
```

During each phase of each macro a pop-up window will appear to inform the user what is currently happening. When the macro is finished a pop up informs the user and a timestamp is created in an Excel cell. An error handler was added to the macro to stop the macro and return everything to start in case of an error occurring during running.

The final version of the Masterfile consists of four different sheets which are Additional instructions, Master, QueryData, and eSmart. These sheets were chosen for the final version due to their necessity and usefulness. The Master sheet is needed for control and overall instructions, Additional instructions sheet contains more instructions for use and the additional instructions for the "Old data" macro and error situations. The QueryData sheet is required for the macros to run and perform the queries in SAP. The eSmart sheet is required to update the query data for QueryData sheet.

6.1.2 Other Excel sources

New sources were created to help with data organization and connections to different sources. There were four main new sources "Material List", "Z030_12", "Plastic component masterfile", and "Material Codes for BOMS". In addition, other sources were created during different prototype phases of the Power BI system. However, these sources were later discarded.

"Material List" Excel file contained over 75 000 rows which contained all the products which had a bill of materials, who makes the assembled product and who supplies the components which belong into to the product. Excluded from this list are basic plastic components whose bill of materials contained only a raw material data which was not up to date. This file was created based on the file "Z030_12" utilizing Power BI. The file was loaded into Power BI where data on who produces the assembled product from text file ME2M was used to create the first version of this list. Then the list was gone through and filled utilizing SAP data on who makes the product and who makes which components. Later this list was verified with SCM team members checking the validity of the data in the file by going through all the materials, who makes them, and who supplies the components to who. This Excel file is the only place at the company where this data is stored at this date.

The file "Z030_12" was continually grown through different prototypes when missing data was noticed. This files final size contained over 100 000 rows containing the current revisions of the bill of materials for ABB materials excluding plastic components. Plastic components were excluded due to the data contained on their bill of materials has not been kept up to date. This file was used to get the data on assembled products and metal components on the number of components per assemblies and the weight of metal raw materials per components. In addition, this file was used to create the "Material List" file. This file was continually updated by utilizing a macro made at ABB for extracting the bill of materials. However, due to limitations in SAP this was slow and time-consuming operation. Materials were added when missing materials were noticed which had been received at ABB during the past three years. To find all of the materials which were needed for accurate analysis a list of all materials which had a bill of materials was gone through and materials which had an active contract during the last three years were added to the file" Z030_12". This file needs to be manually updated in case of changes in the bill of materials and when new materials enter production.

"Plastic component masterfile" was used to calculate the plastic raw material consumption and add data on plastic components used at ABB to other calculations. The original file which was created in 2021 was updated during the creation of the Power BI system due to missing data on which raw material which plastic component used. Through the Power BI system the "Plastic component masterfile" can be kept up to date.

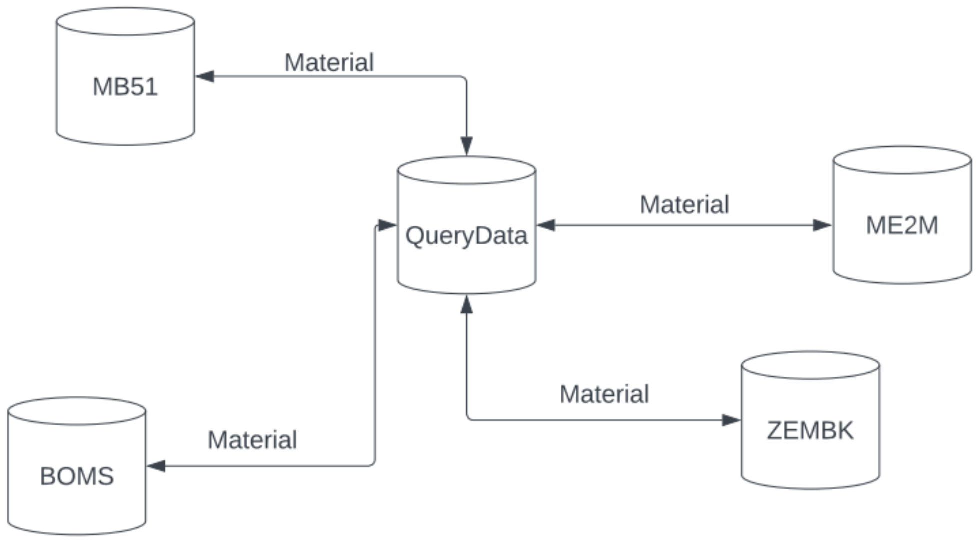
"Material Codes for BOMS" file was used to get the data on which material group each material was a part of. This data was used differently in different versions of the Power BI system. Originally the data was utilized to allocate which materials were plastic components and which were which metal components. Later the data was used to differentiate different Tiers of parts. However, these uses were inaccurate and non-efficient, and the data was used for information on the spend, quantity, and volume based on the share between different material groups.

6.2 Analyzation Data Model

The main operational ability of this system is the Power BI model used for data analysis to answer the research question. The Power BI model's build and functionality was changed multiple times during the project when the use and understanding of the system was improved upon. The analyzation, data transformation, connections, formulas, and visualizations were tested and improved through each prototype phase. The data used for calculations and connections in each of the prototypes is elaborated upon during each version's chapter.

6.2.1 Basic Data Model

Prototype 1 was created based on the basic model which was started by laying out the different data tables on pen and paper. This was done to map out what data was needed from which table and how the connections and relationships between different tables were to be built. The original plan was to rely on data relationships between the different tables. This meant planning the connections based on columns in different tables. This original plan is shown in the picture below. This plan was implemented as the starting point for the for the whole system's development.



Picture 5. Data model relationships between tables draft layout.

The main point when creating connections in Power BI is to create universally useful connections between the tables. Due to Power BI limiting the number of active connections in a data model each connection created had to be useful. In the picture above all of the tables connect using a similar data column which was chosen to be the Material column found in each of the tables. When the connections had been made data, could be extracted from any of the tables to any of the other tables for example, material price data from table ZEMBK could be extracted to table BOMS.

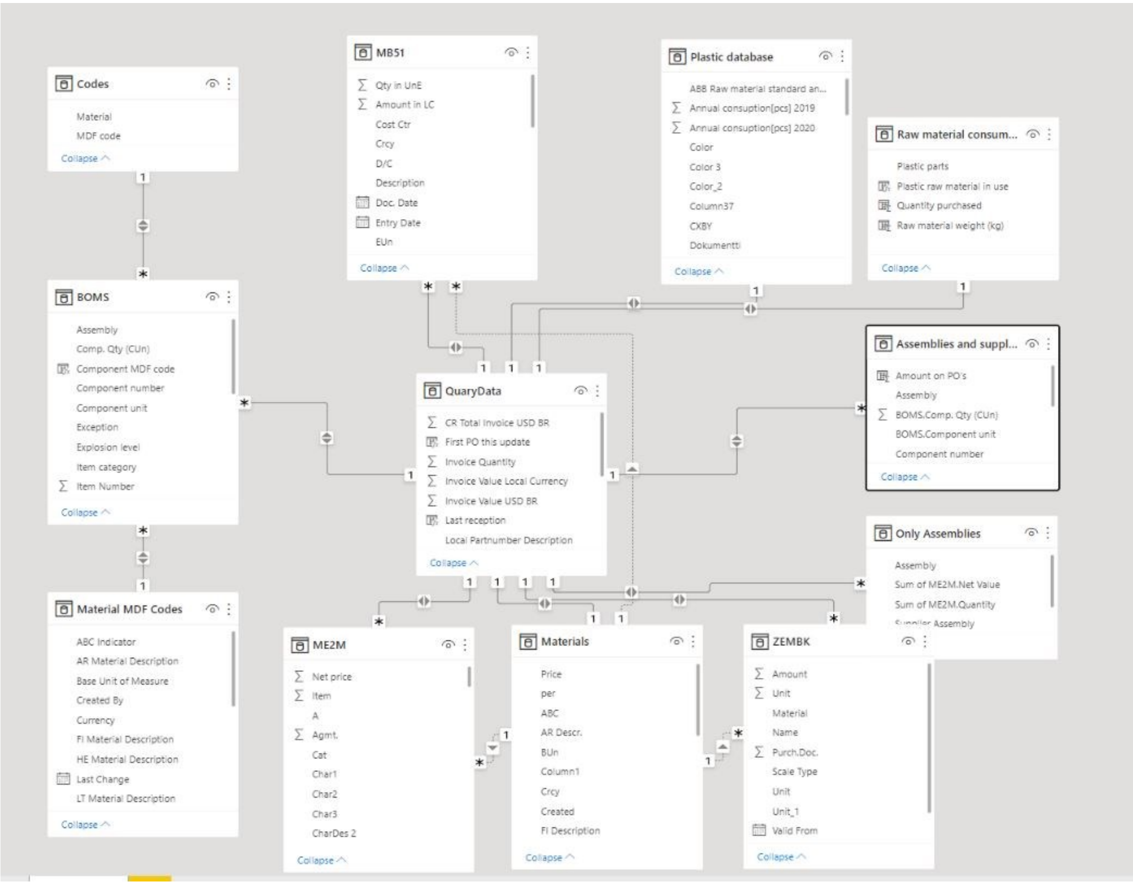
6.2.2 Data Model Protype 1

The first prototype was built based on the basic model's plan. This prototype used nine different source files which were transformed and used to create twelve different tables. These tables and the data they contained are shown in the table below.

Table 1. Prototype 1 data tables.

Data Table	Data Source	Description
ME2M	ME2M text file	Purchase orders based on materials from transaction ME2M
MB51	MB51 text file	Goods receipts based on materials from transaction MB51
QueryData	QueryData Sheet from Supplier_Data_Analysis_Masterfile Excel	List of materials used for SAP queries
Codes	Sheet Data from Excel file MDF codes 2022	Description of MDF codes
BOMS	Sheet BOMS from Excel Z030_12	Bill of materials for materials used at ABB
Material MDF Codes	Sheet 1 from Excel Material Codes for BOMS	Table containing Material codes for materials
Materials	Materials text file	Table containing information on materials
ZEMBK	ZEMBK text file	Conditions by contract data for materials. All of the contracts and prices for materials from transaction ZEMBK
Plastic database	Sheet Plastic database from Excel file Plastic component masterfile	Table on all plastic parts in use at ABB and their information
Assemblies and suppliers	Sheet Assemblies and suppliers from Excel file Material List	Table containing suppliers of assembled parts and the components which belong to them and their suppliers
Only Assemblies	Sheet Assemblies and suppliers from Excel file Material List	Table containing only assembled materials and their suppliers
Raw Material consumption	Table created using Dax-programming	Table created in Power BI to calculate plastic raw material consumption

This protype relied heavily on the QueryData table which contained the list of materials purchased to ABB Smart Power during the year 2021. This table was used as the basis because it contained the values used for the queries and the values occurred in the table only once. Connecting the different tables to the QueryData table allowed the table to be used as a gateway to connect all of the different tables together in a chain. This table was the centre where the other tables were connected based on similar columns. Table Raw material consumption was created using DAX-programming to calculate the raw material consumption. Table Only Assemblies was created using the same source file as Assemblies and suppliers for additional calculations and to help separate different calculations and results for the visualizations. In the picture below are shown the data connections using the relationship method in Power BI's data model view.



Picture 6. Prototype 1 data connection model view.

The data connection used in this prototype are one-to-many and both-ways. Through these connections data can be used in different tables utilizing DAX-programming. The choice of which columns were used for connections were based on if the two tables have a column which has a column with similar values and if one of the tables has the values occurring only once. If the values occur in both tables more than once the connections will be many-to-many which will scramble the accuracy of the data.

Using these connections, the main requirements for the system could be calculated. Data from SAP was connected to the QueryData table to get data on how many purchase orders there were in the previous year, what where the quantities of the materials, and who they were purchased from. The table Raw Material consumption was based on the data on plastic components from the Plastic database which connected to the SAP data to calculate the amount of raw material consumption during the year based on plastic components, suppliers, and the raw material. The tables Only Assemblies and Assemblies and suppliers were used to get data on metal raw material consumption and the different Tiers of materials. Several other calculations were done before testing to look at how the data could be utilized in the current version.

Calculating the material consumption, material spend, material volume, and visualising the movement of materials in the supplier network was done using different methods for each information required. DAX-programming was used to calculate data between the tables. In this prototype the material consumption, spend, and volume were calculated for Tier 1 materials directly by getting the data from purchase orders from the source file ME2M. This source contained the information on the quantity of the material, the net value of the order, and from which supplier the material was from.

This same data could be used for calculating the raw material consumption for metals and plastics. Using the quantities from the purchase orders, the amount of raw material per component from the plastic database, and the bill of materials file the raw material consumption could be calculated for each of the components. In order to get the data for the raw material weight per components and which raw material was in use the data had to be extracted from one table to another using DAX-programming function *LOOKUPVALUE*. This extracting of values from one table to another caused problems with DAX-programming, due to the circular logic error. To circumvent the error different values were calculated using different tables.

Data on Tier 2 materials was found to be difficult to calculate. The quantities of the Tier 2 materials could be calculated by extracting the data on the number of components per Tier 1 material. However, the values on these materials were found to be nigh impossible using these methods. The visualization of the supplier network was done using a table to show which suppliers' assembled product used which suppliers' component.

After the data was connected and calculated the utilization proved to be difficult using the method chosen for the prototype. The data connections were not accurate and showed multiple errors with values when compared to values calculated by hand. The supplier network could not be shown accurately with the data on quantities and spend of each supplier, due to the connections being inaccurate. This was caused by how the data connects using the relationship model. This method caused the data to be calculated using summaries and data for different time periods was not possible to be shown. The other main problem with this version of the system was circular dependency which arose when multiple columns or values were used in a table from another table. This limited the use of data from one table to another and would have had to be corrected by building multiple auxiliary tables. However, this method would not have solved the inaccuracy of the different connections. The main reason for the inaccuracy of the connections between different columns and rows was the limit of how many columns could be used as the search values, when searching data from one table to another. Multiple times when values were looked up using the DAX-function *LOOKUPVALUE* the values were scrambled or not found.

The testing of the accuracy of the values and results of the model were noticed when the data was visualized. When the data was utilized during the visualization process the values were noticed to be inaccurate while, other required values and results for the thesis were nigh impossible to create using this method for the system. Below is shown one of the sheets of the prototype 1 visualization view which was used for testing in the Power BI Service tool.

Picture 7. Prototype 1 sheet Plastic part on BOMS.

The picture above shows the sheet Plastic Parts on BOMS with the values covered. The sheet is used by using the Power BI tools. This sheet is shown as an example of the sheets in this prototype as the other sheets work in a similar manner. If a user would want to look at for example how many certain plastic components are used during the year, they would have to filter the data using the table BOMS column Weight of parts on BOMS from the Power BI tools. This same method needs to be used in the other sheets for detailed information.

Most of the visualizations were done as different kind of tables which according to end users was useful, but more visualization was asked for the next versions. This protype relied on the Power BI controls and not controls built into the visualization itself. During testing this version of control proved to be cumbersome and hard to use. Especially filtering or searching for information was found to be difficult.

During this prototype the purchase order data from the SAP transaction ME2M was used profoundly with most of the values calculated utilizing these values. This was however noticed to be inaccurate data for this system due to the fact that purchased materials were just orders and were not yet received. The material goods receipts could be received a long time later which falsifies the data calculated. Due to this the text file MB51 was chosen to be used in place of the ME2M file in the next version.

The problems, the futility, inaccuracy, and complexity of the DAX-programming method caused the whole system to be recreated using a different method. The creation of this prototype however, showed how the visualizations were to be handled and the testing and presentation of the prototype to end users allowed for input on how to proceed with the next prototype. This prototype in addition validated the testing methods for the following prototypes.

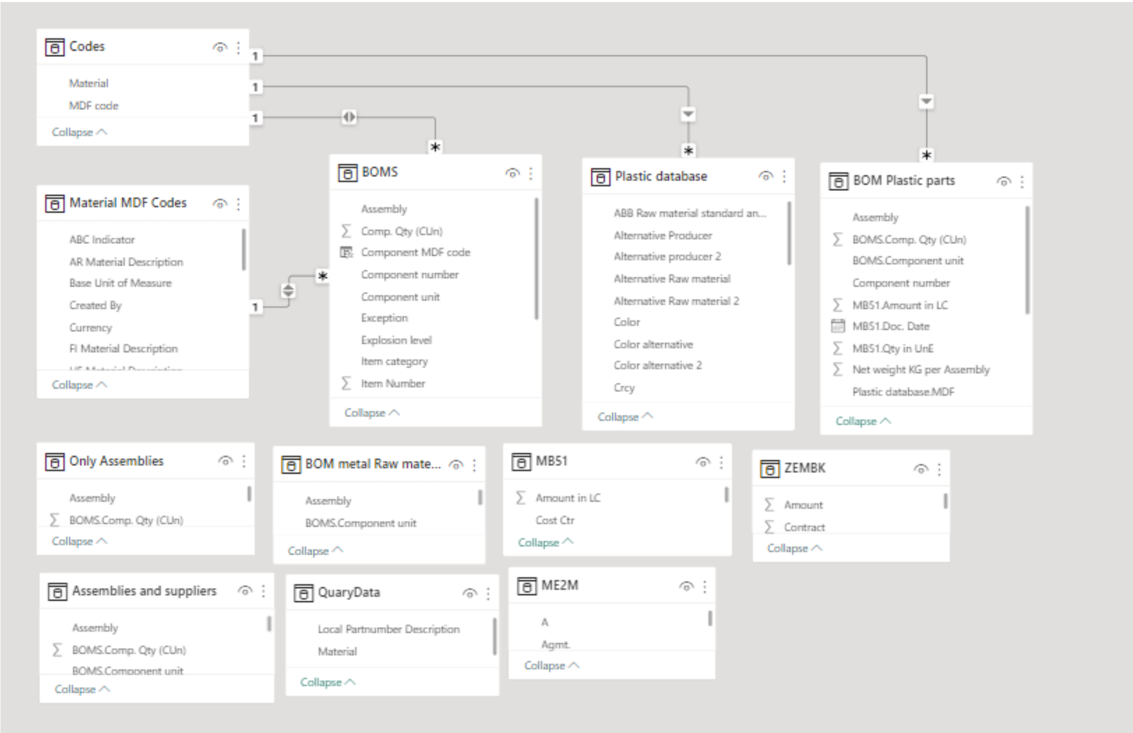
6.2.3 Data Model Prototype 2

Prototype 2 was built from scratch and was based on utilizing the Power Query tool and M-programming instead of data relationships in Power BI or DAX-programming. This was due to the problems with data accuracy, circular logic, and incorrect calculations noticed when these methods were utilized in prototype 1. This version started by adding the source data to Power BI and creating tables using Power Query according to needs. This version used seven different source files and has twelve different tables. The explanation and source of each table is presented in the table below.

Table 2. Prototype 2 data tables.

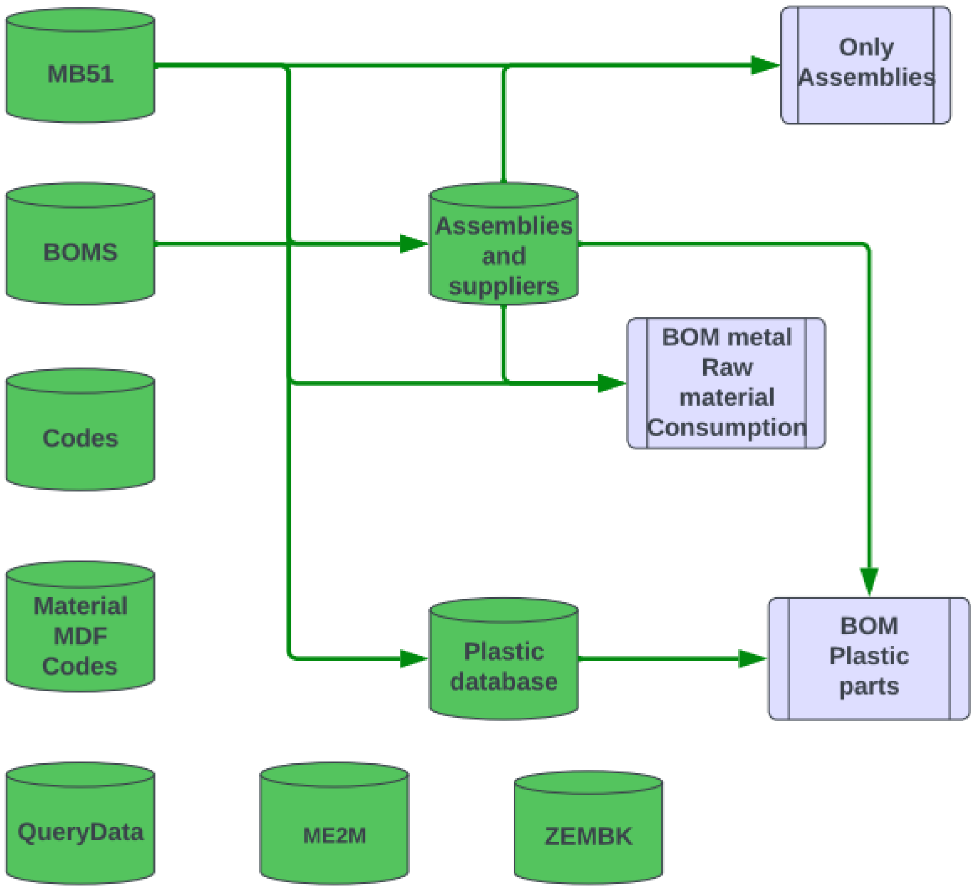
Data Table	Data Source	Description
ME2M	ME2M text file	Purchase orders based on materials from transaction ME2M
MB51	MB51 text file	Goods receipts based on materials from transaction MB51
QueryData	Sheet QueryData from Supplier_Data_Analysis_Masterfile Excel	List of materials used for SAP queries
Codes	Sheet Data from Excel file MDF codes 2022	Description of MDF codes
BOMS	Sheet BOMS from Excel Z030_12	Bill of materials for materials used at ABB
Material MDF Codes	Sheet 1 from Excel Material Codes for BOMS	Table containing Material codes for materials
ZEMBK	ZEMBK text file	Conditions by contract data for materials. All contracts and prices for materials from transaction ZEMBK
Plastic database	Sheet Plastic database from Excel file Plastic component masterfile	Table on all plastic parts in use at ABB and their information
Assemblies and suppliers	Sheet Assemblies and suppliers from Excel file Material List	Table containing suppliers of assembled parts and the components which belong to them and their suppliers
BOMS metal raw material	Tables Assemblies and suppliers and MB51 merged	Table containing materials and suppliers with metal raw material consumption
Only Assemblies	Sheet Assemblies and suppliers from Excel file Material List	Table containing only assembled materials and their suppliers
BOM Plastic parts	Tables Assemblies and suppliers and Plastic database merged	Table containing assembled parts which contain plastic parts and the plastic raw material consumption

Due to this prototype utilizing Power Query the connections between tables are not as apparent as they were in prototype 1. The connections, transformations, and calculations are only shown in Power Query while the data model view only shows the tables and relationship connections. In the picture below the data table for prototype 2 is shown with relationship connections between five tables. These connections were connected through the MDF-code columns for all other tables except for tables BOMS and Material MDF Codes which were connected through the column Material.



Picture 8. Prototype 2 data connection model view.

In Power Query many of the tables use data from other tables by merging or appending the different tables. These transformations and combinations are explained in the figure below. Tables created directly from source data are green while tables created using Power Query are grey and the connections between the tables are merges.



Picture 9. Table data connections in Power Query.

Using this method provides accurate data by connecting different tables based on multiple values and creating new rows for instance based on each goods receipt per material. This method increased the data amount exponentially but provided more accurate and useful data.

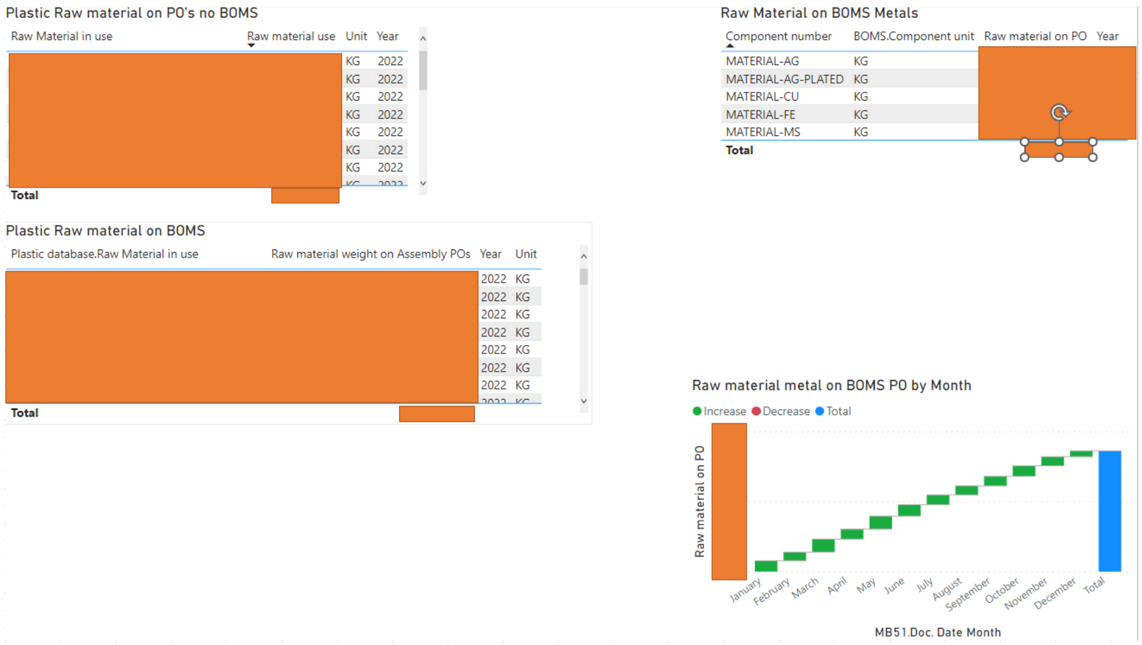
To provide answers and solutions to the main questions for this thesis this prototype utilized the Power Query tool and M-programming. Through Power Query the different tables could be merged using multiple columns as the basis for the merge. In addition, other tables could be appended and through this, data from different tables could be combined into a single table.

This version of the prototype focused on first getting the raw material consumption calculations and visualizations correct. This was done by creating new tables for both raw material groups plastic, and metal components. The plastic raw material consumption had to be divided into two different tables. The first table was based on the plastic database table and calculated the Tier 1 plastic components' raw material consumption. The second table was based on the Assemblies and suppliers table and was used to calculate the plastic raw material consumption for assembled products which contain plastic. This table was named BOM plastic parts and was then merged with the plastic database table using the columns for components to extract the weight of raw material per plastic component and the raw material in use. The consumption was then calculated by merging the MB51 table with both of these tables to get the data on the quantity and dates for each component sold containing plastic. In addition, the table Plastic database extracted the column supplier from the table MB51. The table BOM plastic raw material was merged using the supplier code and material while, the table plastic raw material used the component number column as the connecting column. For the BOM plastic parts' quantity, the quantity of the assembled product was multiplied with the quantity of the plastic material in the assembled product. The quantity of the plastic components was then multiplied with the weight of raw material per plastic component. This provided the information on plastic component consumption per supplier, plastic component, assembled product, date, and raw material in use.

Metal raw material consumption was calculated by using the bill of materials data from the table BOMS which, contains the metal in use and the weight of the raw material per component. The table BOM metal raw material used the table Assemblies and suppliers as the basis for the new table. This table filtered out every other row excluding, the rows containing metal raw materials. The table was then merged with the table BOMS through the columns assembled product and component to get the weight of the raw material per component. Then the table was merged with the table MB51 to get the data on the quantities, and dates of the components sold. The quantity was then used to calculate the total raw material. Due to the metal raw material amount per component being provided in different ways additional IF-statements had to be added to calculate the weight of the raw material in kilograms.

The spend and quantity of Tier 1 materials was gathered directly from the MB51 table while the information on Tier 2 materials needed to be calculated. The quantities of Tier 2 materials were calculated by using the table Assemblies and suppliers. The table was merged with the table BOMs through the assembled product column and the component column to get the quantities of the components per assembled product. The table was then merged with the table MB51 through the column assembled product supplier and assembled product to get the data on the quantities, the spend, and the date of the goods receipts per assembled product. The quantity data was then multiplied with the quantity of the component per assembled product to get the information on the quantity of the component per a goods reception. This data however could not provide information on the spend of Tier 2 materials. This table however provided clearer picture on the true amount of quantities different suppliers provide.

The visualizations were done more sparingly in this version due to time constraints when the system was unveiled for the end users for feedback and testing. This was done out of time restraints and to see whether the chosen method would be the one to continue with. This version contained only five sheets which were seen as the most crucial for the feedback and usefulness of the system. Calculating and creating this information from the raw data also directed the way to create the next parts. The new version relied more heavily on charts and figures while containing tables for more accurate data. The amount of control and accuracy to the data was improved as well. Below is shown the sheet Raw Material from the visualization view with business information removed.



Picture 10. Prototype 2 sheet Raw Material.

This data however was not accurate in this model and needed work to correct the data for the next version. This prototype was not completed as completely as the first prototype due to testing and feedback from end users. This prototype's main purpose was to show how the next version was to be built and it allowed to see which data useful and which data was not needed. The user interface in the visualization was made clearer with more visual aids and controls within the visualization itself which allowed for easier, clearer control, and use for users.

Before the final version was started a version prototype 2.5 was created. This version continued the evolution of the prototype 2 by further implementing additional data and testing on how to connect different tables in order to see the most efficient way for the system to run. During this stage the need for additional data to be gathered arose which was discussed in chapter 6.1. The most major part of this further developed version was the implementation of past years' data and the need for price data for Tier 2 materials. During this testing in addition, the need for certain data was found to be unnecessary which resulted in the data being deleted from the system.

6.2.4 Data Model Final Version

The final version started from the basic structure of the prototype 2.5 which was a further developed version of the prototype 2, containing almost all of the information required from the system. The previous version of the system provided a connection map on which tables needed to be appended and which were to be merged to cut down on unnecessary operations in Power Query. The total amount of data and tables was exponentially grown during the testing of the final version, when more ways to provide new information was discovered, and by considering additional request on the system from the end users.

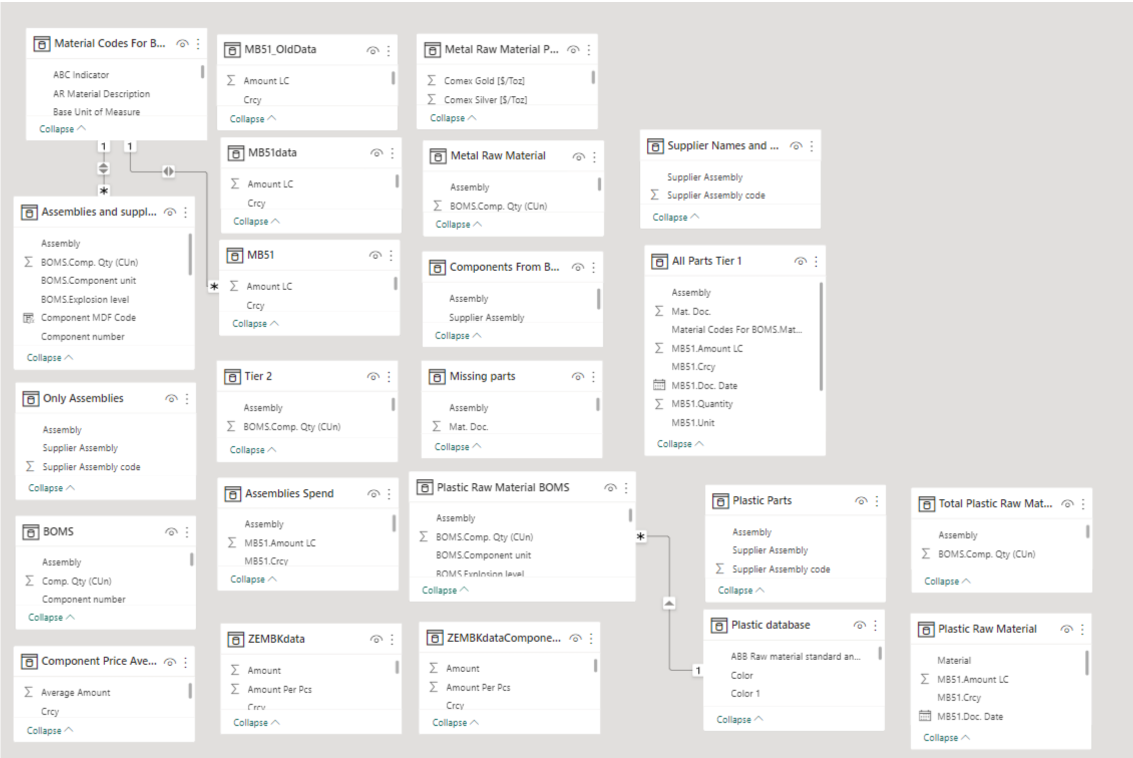
The final version's data amount is far greater than the amount of data used in previous versions. This is due to the added data for previous years and the added price data on Tier 2 materials. The final version used eight source files and had 23 tables. In the table below the source and the description of the tables in the final version are listed.

Table 3. Final version data tables.

Data Table	**Data Source**	**Description**
MB51data	MB51data text file	Goods receipts based on materials from transaction MB51 from the start of the current year
MB51_OldData	MB51_OldData text file	Goods receipts based on materials from transaction MB51 from the past three years
MB51	Tables MB51data and MB51_OldData appended	Table containing all goods receipt data from tables MB51data and MB51_OldData
Material Codes For BOMS	Sheet 1 from Excel Material Codes for BOMS	Table containing Material codes for materials
Assemblies and suppliers	Sheet Assemblies and suppliers from Excel file Material List	Table containing suppliers of assembled parts and the components which belong to them and their suppliers
Only Assemblies	Sheet Assemblies and suppliers from Excel file Material List	Table containing only assembled materials and their suppliers
BOMS	Sheet BOMS from Excel Z030_12	Bill of materials for materials used at ABB

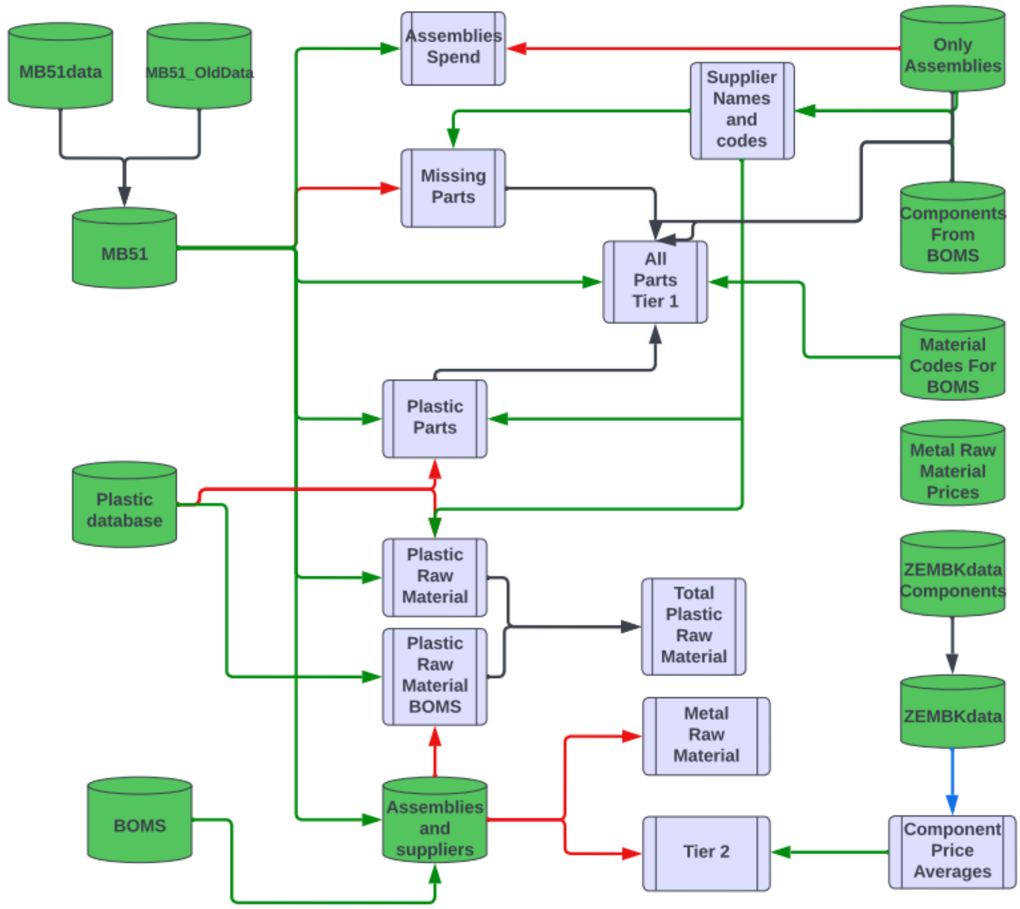
ZEMBKdata	ZEMBKdata text file	Conditions by contract data for materials. All the contracts and prices for Tier 1 materials from transaction ZEMBK
ZEMBKdataComponents	ZEMBKdataComponents text file	Conditions by contract data for materials. All the contracts and prices for Tier 2 materials from transaction ZEMBK
Component Price Averages	Table ZEMBKdata summarised	Average prices for materials per supplier
Plastic database	Plastic component masterfile sheet Plastic database	Table on all plastic parts in use at ABB and their information
Plastic Raw Material	Table plastic database	Table containing all Tier 1 plastic parts and suppliers, and the plastic raw material consumption
Plastic Raw Material BOMS	Table Assemblies and suppliers	Table containing assembled parts which contain plastic parts and the plastic raw material consumption
Total Plastic Raw Material	Appended table of tables Plastic Raw Material and Plastic Raw Material BOMS	Table containing all plastic parts both Tier 1 and Tier 2 and their plastic raw material consumption
Metal Raw Material Prices	Sheet 1 from Excel file Raaka-aine hinnat	Metal Raw Material Prices from the London Metal Exchange
Metal Raw Material	Table Assemblies and suppliers	Table containing materials and suppliers with metal raw material consumption
Tier 2	Table Assemblies and suppliers	Table containing suppliers of assembled parts their components and component suppliers. Spend, quantities, and volumes calculated based on components
Supplier Names and Codes	Tables Only Assemblies and Components From BOMS Appended	List of supplier names and codes
Components from BOMS	Sheet Assemblies and suppliers from Excel file Material List	Table containing Tier 2 materials and their suppliers
Plastic Parts	Table Plastic database	Table containing a list of plastic parts
Missing Parts	Table MB51	Table to get the materials missing from other tables
All Parts Tier 1	Only assemblies table appended with tables Components from BOMS, Plastic Parts, and Missing Parts	Table containing all Tier 1 materials and the suppliers
Assemblies Spend	Table Only Assemblies	Table containing only the assembled parts and suppliers with the spend, quantity, and volume

The number of connections and tables created in Power Query increased further from the previous version of the system. Due to this the need for direct relationships in Power BI was decreased and the final version only contains three different relationship connections. These connections were utilized only for simple data needs in tables such as names of suppliers and MDF codes. In the below picture is shown the data view model of the system. Due to the connections being in Power Query the need for descriptions for the tables shown in this view were required.



Picture 11. Final version data connection model view.

The merges and appends between different tables are a complex system which can be explained in a figure. The picture below explains how the different tables connect with each other. Tables created directly from source data are green while tables created using Power Query are grey. The connections are colour coded based on the type of connection. Green connections are merges between tables, red connections mean the table was created by referencing the other table, blue connections mean a table was created by grouping, and black connections mean appends of tables.



Picture 12. Final version Power Query connections.

This system was planned during the prototype 2.5 phase to decrease the number of transformations, merges, appends, and calculations in tables. Starting from the table MB51 the system was built by adding new tables and combinations in the mapped order.

The final versions calculations were taken further than in the prototypes to correctly provide essential information for the company. The calculations and data combinations done in by the system were developed during the whole prototyping phase to provide accurate information on the material consumption, the material spend, the material volume, and the movement of materials in the supplier network for the two material groups. The calculations and data combinations utilized multiple methods for the system to provide accurate information.

Material consumption is divided by the three categories Tier 1, Tier 2, and raw material consumption. Tier 1 material consumption was calculated by utilizing table MB51. Tier 2 material consumption was calculated by utilizing tables Assemblies and suppliers, MB51, and BOMS. The consumption of Tier 2 materials was calculated inside the table Tier 2 which was created to separate the calculations for Tier 2 materials from the other calculations. The table Tier 2 was created by referencing the table Assemblies and suppliers to provide a copy of the table. The table was already merged with the table MB51 through the columns on assembled product and supplier, through which the table was provided with data on the spend, quantities, and dates of receptions for the assembled products per supplier. The table was then merged with the table BOMS trough columns on assembled product and assembled product component. This merge provided to the table with the data on the quantities of the components per assembled product and the unit in use. The consumption of Tier 2 materials was then calculated by multiplying the quantity on per reception and the quantity of the components to provide the information on the total consumption of Tier 2 materials.

During the previous prototype the raw material consumption of plastic and metal components was compared to previously used sources to test whether the calculations were correct. This testing provided missing data and solutions to accurately calculate the consumption. The plastic components had to be calculated using two tables which were then appended to create the table Total Plastic Raw Material Consumption which contained the total amount of plastic raw material consumption. The consumption of plastic raw materials for Tier 1 plastic components and assembled products with plastic components was calculated the same way as in prototype 2. When the tables were combined the consumption was noticed to be less than in the comparison calculations. This problem was found to be caused by the source file “Plastic component masterfile” sheet

Plastic database which was missing information on the raw material in use for many plastic components. The raw material in use information had to be manually searched and added to the Excel file, which corrected the calculations for the consumption. Metal Raw Material was conversely calculated in the same way as in the previous prototype. All the metal components were contained in the Assemblies and suppliers' source which allowed the system to contain all the metal raw material calculation in one table. During the comparison to other calculations on metal raw material consumption errors were noticed. These errors were caused by incorrect information and missing components on the Assemblies and suppliers' source and the BOMS' source. This incorrect and missing information was manually corrected which exponentially grew both files' sizes.

Material spend was calculated for Tier 1, Tier 2, metal components, plastic components, and assembled products. Material spend for Tier 1 materials was calculated similarly to the previous prototype using table MB51 and the spend per receptions. Calculating the Tier 2 spend was crucial for this thesis and proved to be the most challenging. Tier 2 spend is information not found anywhere directly at the company and through this information the true strength of a supplier can be seen in the supplier network. Tier 2 materials are not purchased directly, which meant the price per component had to be searched. However, due to most parts having multiple price updates during a year the correct price for a material had to be connected by not just supplier and part, but in addition the correct timeframe. The main reason table Tier 2 was created was to calculate the spend on Tier 2 materials. The quantity of Tier 2 materials was already in the table through earlier merges. The need then arose to get the price per materials from table ZEMBKdata.

The price data for components in table ZEMBK is presented in a date timeframe "from" and "to". The reception data had to fall between this timeframe for the price data to be accurate. The merge to get the correct price for the correct component per correct supplier during the correct timeframe was tested using Power Query's M-programming function *Table.SelectRows()* and *Each Keyword Debugging*. This method was further developed using M-programming function *Table.Buffer()* due to the data amount processed being massive. The table buffer function buffered the tables which were merged to help with runtime speed. Both tables contained millions of rows of data which Power Query had to go through and compare if the date was correct. Even with the buffering, this method proved to be too heavy to run crashing and freezing the computer in use. The data was filtered to cut out unneeded price data to decrease the number of rows. In addition, unneeded columns were removed, and Power Query settings were modified to help with the optimization. However, these efforts were futile. This method was tested in a separate Power BI report to see if the other calculations, connections, and data in the system were too heavy for the system, but this test version had similar runtime problems. Due to this method being too burdensome for the system to run another method was used to get the prices.

The chosen method was to use the average prices for components from suppliers. Using the Power Query tool *Grouping* a new table was created, Component price averages, which housed the supplier code, component, and average price per component. Before this grouping was done the price data on ZEMBKdata was filtered for the data to be as accurate as possible. ZEMBKdata contained data from 2006 onward. The prices were filtered to exclude older prices than 2018. Further filtering was done to filter out incorrect prices stemming from users updating prices to SAP and incorrectly announcing the price per piece i.e., a price was input per 1000 pieces and the per piece number was input as per one.

Tier 2 table was then merged with this Component price averages table through columns component and component supplier. This merge provided the average price per piece for each Tier 2 component per supplier. This data was then used to calculate the spend per goods receipt by multiplying the total quantity of a component on a goods reception, with the average price which provided the calculation for the Tier 2 spend. This method was tested using averages per years, but this method proved too inaccurate due to many component prices not having any price updates for many years.

Material volumes were a collection of calculations through the other calculations. The volumes could be seen through the quantity, consumption, and supplier network presentation. Through the previous calculations the total volumes per timeframes, material groups, components, and suppliers could be shown and calculated.

The movement of materials through the supplier network was based on the connections and source data collected. The most important part of providing this information was on how to visualize the supplier network. Calculations related to the movement of materials were the quantities, spend, and volume per suppliers and materials from one Tier to another Tier. This information was provided by the previous calculations.

Through the use slicers all of the information gathered could be changed to show information per materials, time, suppliers of different Tiers, and explosion levels per assemblies. This method allowed through the visualizations greater control over the information shown and for more detailed analysis. In addition to the calculations required by the thesis, an additional calculation on the stock prices of metal raw materials was created during this final phase of the system.

The visualization and control of the system was crucial to provide an efficient method for data analysis. These methods are explained in more detail in the next chapter.

6.2.5 Visualization and use of the final version

Visualization of the data calculated, combined, and gathered in this Power BI system is the main point where the data becomes information. The visualizations were developed during each phase based on feedback and the needs for the system. This chapter focuses on the final version of the system while the visualizations of the prototype versions were gone over in the previous chapters.

The final version consisted of eleven different sheets Summary, Tier 1, Tier 2, Plastic Raw Material, Metal Raw Material, Network Assembler Side, Network Component Side, Assemblies, Supplier All Parts, Price Fluctuation, and All Goods Receipts. These sheets were chosen for the need, usefulness, and based on the clarity of the data. One of the problems with Power BI visualization was how to fit the data into a screen. This caused the visualization to be planned not just based on needed information, but also on how to clearly present the data. The chosen sheets are explained based on the information, visualization, importance, and the control they provide to the end user. Due to these sheets containing trade secrets they are not shown in full in this report.

The sheet Summary contains summarised information of the spend and quantities. The sheet contains control for what year’s data it shows which controls the shown graphs. This sheet is used for simple summarised information and the most important graph is an indicator which shows a chosen year’s spend compared to the current year.

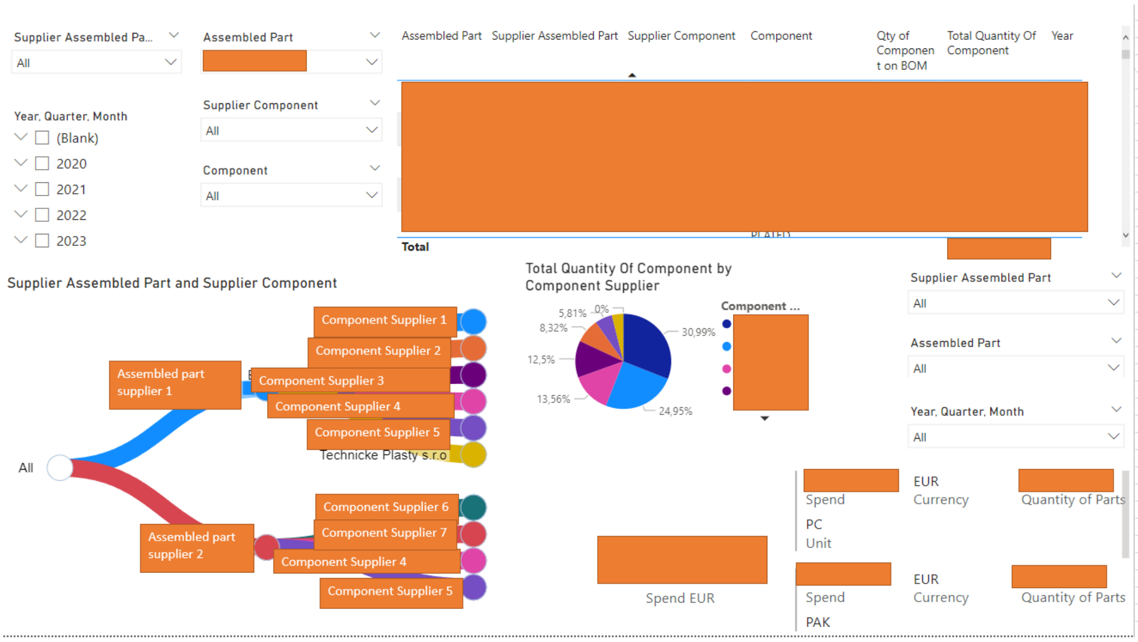
Tier 1 sheet is used to show information on Tier 1 materials. This sheet is used to get the required information on Tier 1 materials using multiple visualizations. sheet shows a table of Tier 1 materials who makes them, their amount in euros and quantity per month and year. The sheet contains three pie charts for the spend in euros per supplier, quantity of materials per supplier, and spend in euros based on share MDF code. The sheet contains a bar chart which shows the spend in euros per suppliers for each month of a year. In addition to these visualizations the sheet contains a table which shows all the Tier 1 materials, their suppliers, the quantities received, the spend per the part, and the month and year the information is based on. The sheet contains four slicers for controls which are the date, supplier name, material, and unit of material. The date can be chosen by the end user by year, quarter, or month. This sheet uses table All Parts Tier 1 as the data source.

Tier 2 sheet presented the spend and the quantities for Tier 2 materials. This sheet focused just on the information required from Tier 2 materials. The sheet presents the information on which parts are Tier 2 materials and who supplies them. The most important information is the quantity of the materials, the spend, and the volume of the Tier 2 materials. The sheet shows in a table the information on the Tier 2 parts including the price and the explosion level within the bill of materials. Pie charts in the sheet show the shares of spend per materials, suppliers, and MDF codes per date. The sheet contains controls for the assembled part, the supplier of assembled part, component, supplier of component, and the date. Through this information the true share of the network per suppliers can be seen.

Plastic Raw Material sheet provides all information on plastic raw material consumption. The sheet contains information on the consumption of the different raw materials in use, the raw materials per plastic component, and the consumption of plastic components per date and supplier. The sheet contains stacked columns to represent the consumption of different raw materials per different timeframes and supplier consumption based on dates. The sheet contains pie charts showing how big the different raw materials in use, components, and suppliers share is in consumption. The sheet contains four different controls for the user which are three slicers and one list. The slicers are for raw material in use, supplier name, and material. The list is for the date which can be chosen by year, quarter, or day. The data used for this sheet comes from the table Total Plastic Raw Material Consumption.

Metal Raw Material sheet shows data on metal raw material consumption. The sheet contains two tables, with one showing the total raw material consumption by kg per year and per raw material, and the other table shows the total raw material consumption by component per supplier per year. There are three pie charts which show the total raw material consumption by component, the total raw material consumption by supplier, and the percentage of all metal raw material consumption by raw material. A clustered bar chart shows the total raw material consumption by raw material. These visualizations are controlled using four different slicers for the date, raw material, supplier name, and the component containing metal. The date slicer allows to the user to choose the information based on year, quarter, and month. This sheet mostly uses table Metal Raw Material as the data source. However, it also uses the table Metal Raw Material Prices as a data source. The sheet contains a combined chart to show the London Metal Exchange prices over time per raw materials primary aluminium, zinc, copper, and silver which has only one control for the year the information is based on.

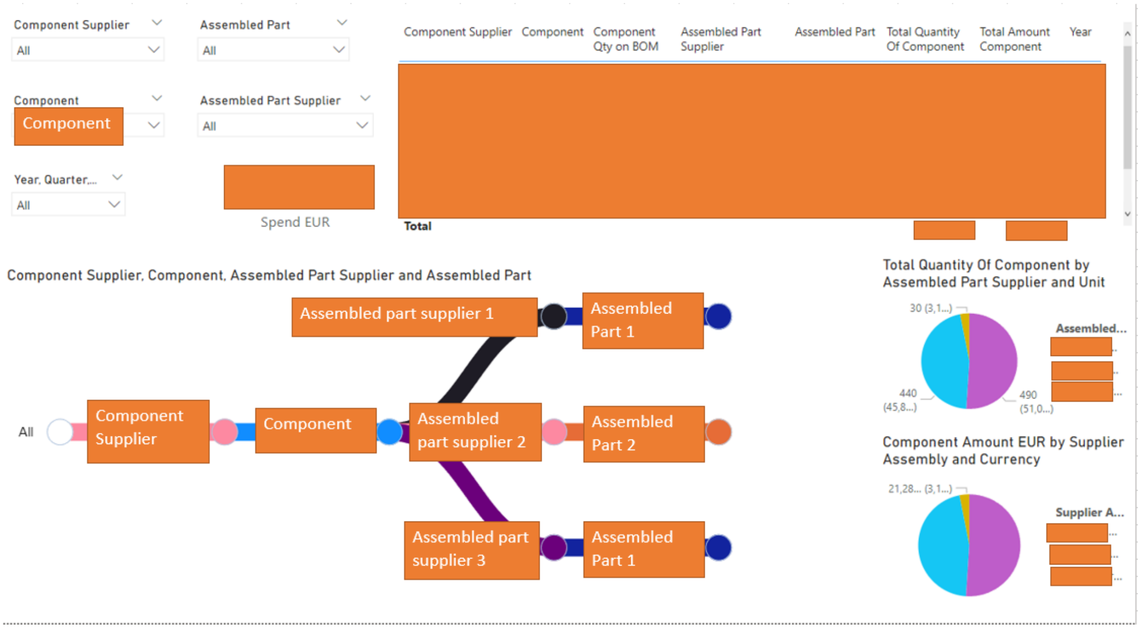
Network Assembler Side shows the supplier network from the assembled product side. The main graph for the sheet is a tree graph to show how the supplier network is built from the assembled product supplier's side. This graph is shown in picture 12. Through this graph the whole network per assembled product supplier, assembled product, component, and component supplier is shown in a clear way. The sheet in addition shows in a table listing the assembled products, the suppliers, component, supplier components, the quantity of the components per assemblies, quantity of the assembled products, total quantity of the components, and the dates they were received. The sheet contains pie charts showing the shares of the suppliers per assembled products, the component suppliers per component, and components per MDF codes. The sheet contains 4 slicers and a list as the controls for the information. The slicers used are supplier of assembled product, assembled product, component supplier, and component while the list is used for the date which can be chosen by year, quarter, or year. This sheet uses for this information the Assemblies and supplier table as a source. Below the sheet is shown with business values removed and with one assembled product's information chosen.



Picture 13. Network Assembler Side sheet visualization for one assembled product.

The sheet Network Assembler Side in addition, contains additional information and controls for the information on the spend per assemblies. This information is shown on the lower right corner of the picture above without filters meaning all data is shown however with the business information removed. This information comes from the table Assemblies Spend, and contains a card which shows the spend per currency and the quantity of the assembled product. This information is not connected to the controls mentioned before as this information is controlled using three slicers shown in the picture on the right side of the screen. These slicers are for assembled product supplier, assembled product, and the date which can be chosen based on year, quarter, or month.

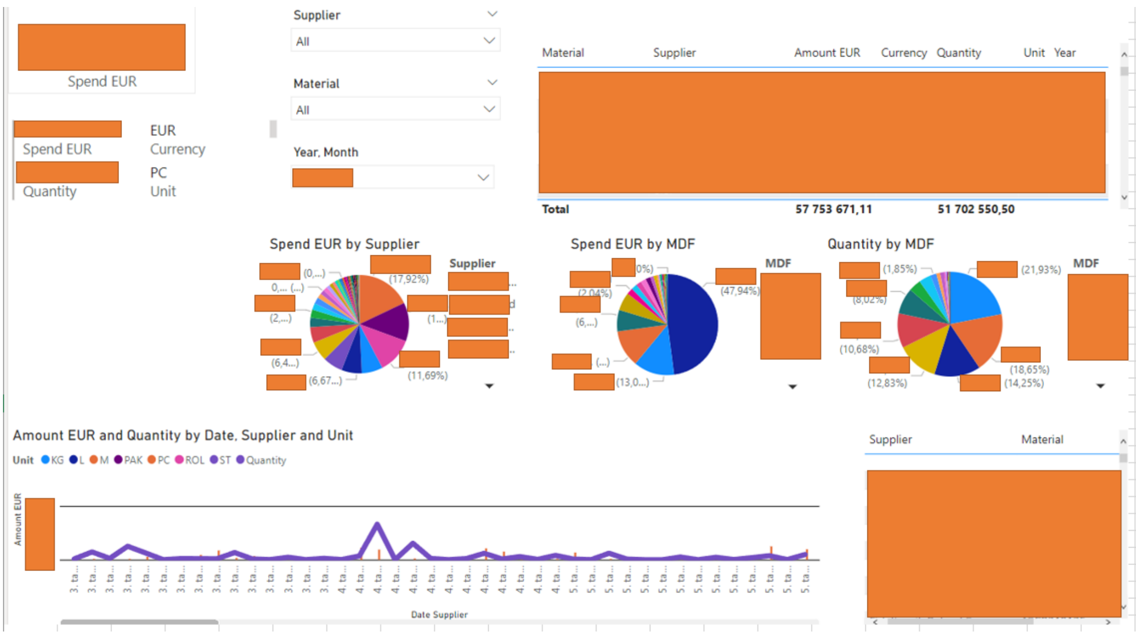
Network Component Side presents the supplier network from the component supplier side. This sheet is used look at the supplier network from the component supplier's side and the main graph used is the tree graph which presents a component supplier and who uses their components in which assembled parts. In addition to this the sheet contains a card showing the spend and two pie charts informing the share of the quantity of the components per supplier and the amount of spend per the supplier. The sheet contains a table which contains a list of the components, component suppliers, assembled products, assembly suppliers, total quantity of the component, and the total spend per component by year. The sheet contains 5 slicers for filtering which are component supplier, component, assembled product, assembled product supplier, and time. These controls were deemed to be the ones required for this information. Below picture shows the sheet with the business values removed. The picture below is filtered to only show how one component moves in the supply network.



Picture 14. Network Component Side sheet visualization for one component.

The sheet Assemblies was an additional sheet created based on the data available and the user feedback. This sheet provides additional information on the assembled products showing how a network of the components with their explosion levels, the listing of assembled products, their suppliers, components, component suppliers, the quantity of component per assembly, the quantity and spend per year of assembled product, and the dates of these receptions. The sheet shows with pie charts how different products quantities and spend directly on the assembled product is spread between different suppliers and MDF Codes. The sheet includes controls as slicers for the assembled product, supplier of assembled part, component, supplier of component, and the date. The sheet uses the table assemblies and suppliers as the source for the information.

Supplier All Parts sheet was also created based on end user needs and the data available. The sheet's main function is to show all the materials a supplier supplies to the network. The sheet shows per supplier if there are direct purchases the amount of spend and quantity per material per year. The sheet also shows the spend and quantity for materials which have direct purchases. The sheet uses the table All Parts Tier 1 data as the source for the visualizations. The sheet contains three pie charts which show the spend in euros per supplier, spend per euros for MDF codes, and quantity per MDF codes for the materials with direct purchases. A clustered bar chart shows over time the amount per euros, and quantity per supplier. The sheet has two tables which the first one shows the material, supplier, the spend, currency, quantity, and unit, based on the year while the second table of the sheet shows all materials a supplier produces even if the materials do not have direct purchases. The information shown is controlled using three different slicers which are supplier name, material, and year and month. The sheet is shown in the picture below with the business values removed and with only one year's information shown.



Picture 15. Supplier All Parts sheet information.

Price Fluctuation sheet was created based on feedback, as an easy method to review how prices have changed over a time. In addition, the sheet allows the user to see whether a material has been deleted from supplier's contract and what currency is in use. This sheet uses the table ZEMBK as the data source for these visualizations. The table contains a table which shows the price per time frame and per supplier, currency, and the status of the contract. The sheet contains a card informing the average price per component and supplier. The sheet uses a stacked column graph to visually inform the changes in price. The controls for the sheet are 4 slicers and 2 selection boxes. The four slicers are valid from date, valid to date, material, and supplier while the selection boxes are for the validity of the contract and currency.

All Goods Receipts sheet shows all data on goods receipts from current year and past three years. The sheet has a table which tells the information on the goods receipt including the purchase order number. Pie charts are for the amount in euros per supplier and the quantity per supplier. There are two clustered columns which tell by time, the quantity per supplier and the amount in euros per supplier. The user can use four slicers to filter by material, time, supplier, and MDF code the information shown. The time can be shown by year, quarter of month. This table uses the table MB51 as the source of data for these visualizations,

These visualizations and information shown were deemed to be required information while the system was built. Additional information and sheets could be produced if necessary. The controls and the detail offered by the controls was changed during the testing phase, when more detail or less detail was required from the information shown. The system originally included a sheet Forecasting which was designed to forecast future prices for the different materials per suppliers using machine learning. However, this sheet was deleted when it was clear that with the deadlines and required changes the forecasting could not be created in an accurate fashion. The feedback from the end users during the final phase reflected the thought that this functionality was not a must need requirement.

6.3 Machine Learning Utilization

The machine learning utilization was taken out of the finished Power BI report due to technical difficulties with DAX-programming. The amount of data used proved too much for the system to handle through DAX, and while limiting the number of materials whose prices were to be forecasted to just 20 out of 7000 allowed for the program to run, the results were not usable. In addition, due to time constraints the algorithms parameters could not be finetuned and the values could not be tested. The company updates the prices every month. Due to this testing the algorithm would have required much more time than was available to use for this thesis.

The algorithm was tested with data from table ME2M and table ZEMBK. Data from both tables were referenced to create tables for training and testing of the algorithm. The algorithm itself is simple compared to many other machine learning algorithms and could easily be implemented into the programming language. Due to this the conclusion that the system was not being able to handle the data load was unexpected.

In the future the use of Python or R-programming would be a method to implement k-NN for forecasting. Similar method to the one shown in the instruction Prediction via KNN (K Nearest Neighbours) KNN Power BI: Part 3 (Etaati, 2017) could be used for this Power BI report if the need for forecasting would arise, and the solution for this forecasting was the same as used here.

Other algorithms such as the Bayesian network could benefit the company in efficiency and accuracy better than the use of k-NN (Gaur & et. al, 2015). This method would require more work compared to k-NN, but after building the system and testing it the algorithm would not require more work.

6.4 Testing

The testing was done for both the Excel Masterfile and the Power BI system. With the Power BI system testing was done differently for different versions created during this thesis. The testing was done as spot testing, user feedback, and user testing. In addition, the calculated values for spend and consumption of raw materials were tested during the testing of the prototypes and the final version.

“Suplier_Data_Analysis_Masterfile” Excel’s testing was done mostly by simply running the system’s macros as frequently as possible. During the first build of the macro and the first prototype, this testing showed problems with the method the system was built upon due to SAP changing layouts, and how text files were saved from the system. Due to this testing, the later prototypes were able to correct these problems.

The spot testing was done after the first prototype was ready weekly for each version of the system. The spot testing was done by using a certain assembled product as the material to look at how information for it was shown and whether the information was correct. This particular material was chosen due to it containing multiple different components and it was produced by two different suppliers. This spot testing provided the information on the accuracy of the Tier 1 and Tier 2 spend and quantity, and the supplier network visualization. Every version of the system was further developed due to the occurrence of the incorrect information and errors.

Through user feedback which was organized both during the weekly spot testing and meetings at the end of development for each prototype the information shown, needed, the control of the system, and additional request were received. The meetings were organized to show the current state of the system and how it works. During these meetings the participants were able to ask questions and ask how to find certain kind of information. When the information was unavailable in the version, notes were taken on it and the information was made available in later versions. Through these queries from end users the information could be verified, and errors could be found. In addition, the feedback on how visualizations were to be handled from the end users allowed the visualization process to be streamline, due to there being clear needs and demands. This feedback allowed the system to truly evolve towards what was required for the system the end users had asked for.

The user testing was executed by a smaller group of testers than was originally planned, due to many of the proposed users not having time to test the system. The limited testing however was enough to find the errors, and incorrect information in the system. These found errors were marked and solutions to these errors could be found by the testing, while further testing allowed the system to become more robust and reliable.

To test whether the raw material consumption of plastic and metal components was correct the values were compared to previously used sources to test whether the calculations were correct. The source for the plastic raw material consumption comparison were plastic raw materials files from the suppliers of plastic components which had data on plastic raw material consumption. The source for metal raw material consumption was an older Power BI report developed at the company to calculate the raw material consumption of metal raw materials. Through this testing problems with calculations and source files could be identified and fixed. Working through the fixes and missing data the final versions values were found to be correct and even more accurate than in the older Power BI report.

The final test for this system was the final presentation held to the end users when the system was complete. During this presentation however an update to SAP resulted in an error in the Power BI's connection to the updated text files. This error was quickly reconciled, and a stable solution was found. Due to this error presenting itself despite the testing an expert at the company was shown the system and how it works to allow for more robust maintenance in the future in case of errors.

6.5 Release for Internal Use

The release for internal use for the system was handled during all the protype phases through the Power BI Service. This service allowed the system to easily be implemented for internal use through the browser, only requiring an access to the system. The Power BI Service version of the system allowed the users full ability to modify the information shown, while not allowing any changes to the build, data, or sources which protected the system.

Each time the system was updated, meaning monthly the system had to be uploaded to the Service. The downside of the Service is that it requires the user implementing the update to have a Power BI Pro licence (Microsoft, 2023). This downside is, however, small when compared to the easy-of-use of the update process and the use of the system. The Use in browser in addition, allows for any user to use any computer to access the system.

Before the final version of the system was implemented for use the older prototypes were available to use in the Power BI Service for testing. These versions were deleted from the Service when the final version was fully implemented for use, to ensure the users utilize the final version of the system.

6.6 Documentation

The documentation for this project was done during and after the project. The end product was six different documents with the total amount of 180 pages across these documents. These documents were created as instructions for use, instructions for problem situations, documentation on how each part of the project was done, and explanations for each of the files and parts. These six documents were Basic Instructions & Examples, FAQ, Material list & BOM Instructions, Plastic Database Instructions, Supplier Data Analysis Power BI full documentation, and Supplier_Data_Analysis_Masterfile full documentation.

Basic Instructions & Examples this document is the main document created to show how the basic use of the Power BI Service version of the Supplier Data Analysis report is used and what different figures and values mean. The document is divided into the basic use, data updating, and examples and explanations and is 35 pages long. In the data updating part each of the 11 different sheets are explained briefly on what data they show and what they are used for. The updating data part is the same instruction on how to update the data in the report, which is featured in most of the documents. The Examples and explanation's part explains in more detailed fashion what all the different tables, values, charts, filters and slicers do in each of the sheets. This part in addition provides examples on how to get different kinds of information from the system using the different sheets for example how much is the spend of a certain supplier on Tier 2 materials during the year 2022.

The document FAQ (Frequently Asked Questions) was created to be a fast and efficient way for users to find out in which of the documents certain instructions and documentation is stored. The document is three pages long and feature 13 different questions and their answers. These questions were thought of during the making of the whole project and were guesses on what kind of questions would come up during the use of this system.

Material list & BOM Instructions document provided the instructions on how to update these source files in case of any kind of update. This document was 15 pages long detailing everything about these source files. The updates to just different suppliers for assemblies or components in assembled products, the changes to a bill of material, and the addition of new suppliers for older products already on the files or new products. In addition, the document provided explanation on how these sources are used, where they are used, and their importance.

Plastic Database Instructions document provides explanation for the source of the data in the file "Plastic component masterfile", how to update and maintain the data inside the file, importance of the layouts, the original version of the file, and the use of this data over the course on 4 pages. Through this document the maintenance updating of this file can be handled efficiently, without errors to the Power BI system.

Supplier_Data_Analysis_Masterfile full documentation file contains the full explanation on how the Masterfile was created, how it works, how to fix errors, and how it connects to the Power BI system. The document provides the documentation through 50 pages on the whole VBA-programming for both macros, going through each part of the program and explaining how to recreate them in the correct way. The layouts are explained in detail to confirm that they are not changed. The possible errors which might arise mostly due to updates to SAP are explained in detail and instructions on what the errors mean, and how to correct them can be found in the document. The connection to the eSmart database is explained and which data must be extracted from there to get the correct information for this system. Through this document the system can be maintained, updated, and even rebuilt in the correct way if the need arises. The document was written and illustrated in an easy-to-understand way to provide the tools for any user to maintain and fix this Masterfile, in case there is an error in the future.

Supplier Data Analysis Power BI full documentation file contains the full documentation on the Power BI system. Through this document's 73 pages is explained how the system is built, how the connections work, how the calculations are achieved, how to connect and transform tables, what different tables and columns mean, how the visualizations are done, and how to fix errors. The document was written to explain how the Power BI system works and to document the build of the system, connections, and calculations in order for the system to be further developed and maintained in the future. The whole system's logic behind the choices in the system are explained for the reader to wholly comprehend how and why this system was created in this way. Through this document a user can learn how to use Power BI, how to modify the system, and how to add new calculations and visualizations to the system. The document in addition explains what the most common errors in Power BI mean and how to fix these.

The main point behind these documents was to explain how everything works and to allow the system to be maintained and further developed without the creator. These documents were written densely to allow the recreation of the system and teach how to use these systems for other actions to automate, analyse, and develop data use, analysis, and use of information systems.

7 Conclusions and Discussion

The end of this thesis shows that the use of Power BI for data utilization for supplier data analysis in Supply Chain Management is a useful tool for ABB. Through the long research and development process of this thesis new information on the systems in use and the possibilities and limitations within them could be noticed and addressed. In the end this research could answer the research question discarding a working implementation of machine learning. This chapter focuses on the conclusions and the discussion on furthering the study on this subject. These discussions and conclusions are based on the feedback, discussions, past research, and the writer's own reflection on this project.

Chapter 7.1 looks at the conclusions to the research, the related question to the research question on, and the answer to the research question *"how can the program Power BI be used for gathering, analysing, and utilizing the supplier data"* and the related questions on the use of machine learning to forecast material price changes and maximizing business value with this system. Chapter 7.2 looks at the future development of the system.

7.1 Conclusions & Discussion

The original plan for the thesis was fairly simple when reflecting on the end product. The most change caused the lack of knowledge and lack of support in data gathering. This study started with the use of SAP Hana database as its source for data, but after much research it was found out that ABB Smart Power does not use SAP Hana. The study then evolved into using SAP Business Warehouse application server for data, but this too was unusable due to there being no data needed available through this application.

Using Excel Macro's with VBA programming language was the only way the data collection could be automated. This however lessened the planned streamlining of the final program. Using Excel Macro's required many different adjustments for a less demanding update of information for Power BI. The Excel Masterfile was thoroughly tested all the time due to errors appearing constantly which were caused by layout changes in SAP and the final working version was implemented during the last week of development. However, due to these problems the Masterfile required extensive documentation on how it functions and how to deal with possible errors. With the future use of SAP data cubes and Business Warehouse connections the system could be updated to update data in real-time providing even more value to the company.

The Power BI system itself proved to be an excellent choice for the data analysis, visualization, and use for the information required from the system. The model required multiple different versions to fine tune and optimise the solution. However, through this research duplicating this kind of system is simplified and the research offers new methods for the end user team to utilize Power BI.

From the original requirement of this thesis were to have a system which provides information on material consumption, material spend, material volume, and movement of materials in the supplier network in Tier 2 and 3 suppliers was achieved through this research. In addition, the detailed information on the suppliers could be seen from the system including all of the materials in different Tiers a supplier provides to the supplier network. The suppliers could be divided into Tiers based on which Tier they most supply materials and have a biggest effect. For Tier 1 these information requirements were achieved with great accuracy and through feedback from end users more information could be offered by the system than was originally required. For Tier 2 the material consumption, material volume, and material spend could be achieved as required. The material spend on Tier 2 materials had to be an estimate due to the limitations on the systems efficiency. The results were deemed to be close to the actual numbers but could not be a quantitative result. This is due to how the price data which was used to calculate the spend on the assembled products components was in the SAP system. With added time a more accurate result could be achieved, however this was seen as unnecessary by the end users for the use of this system. For Tier 3 which were raw materials the system achieved the results and could replace older systems used previously due to the added accuracy of this system. Tier 3 spend could not be directly calculated with the data used however due to multitude of factors affecting the price. Additional information was added on the use of raw materials which according to end users could be more utilized in the future. The requirement of the movement of materials through the supplier network was achieved and expanded upon to show the supply network from multiple points of view. The result was clear and allowed for the end users to clearly understand how the supply chain was built from the different Tiers. The use of the tree visual was the clearest method to communicate the relations between suppliers.

Testing, documentation, and user feedback were the most important part of this research. Through these methods the end system could be developed even further than was originally designed and it could provide even more benefit to the company. The feedback allowed the system to be more accurate, errors could be noticed, and new information from the data could be thought of. The end system allows a user to update the data with few clicks and requires no knowledge of the inner working allowing for easier updates.

The main research question "*how can the program Power BI be used for gathering, analysing, and utilizing the supplier data"* was answered through this research. The answer was that Power BI can be utilized for this purpose. The answer to the related question on how to maximize business value with this system Power BI could be used for added business benefit through the information provided by the system on supplier relations, importance, and size through the whole network. The end users can utilize this system in an easy way for added information regarding a supplier when negotiating and considering changes within the network. Main suppliers can be chosen and ranked by importance when looking at the supply chain to further the control on the supply chain for the supply chain management team. The added information provides additional views and support for choices made within supply chain management at ABB Smart Power.

The related question on using machine learning was researched however, the solution researched in this thesis could not provide a working version of forecasting of price changes. The theoretical side of the solution could work however, the algorithm could not be created to be used in this system in an accurate or an efficient way. This forecasting could be achieved in the future using this system however, the utilization of DAX-programming should be excluded as a method due to the in efficiency of it.

This research has allowed for greater understanding of data, information, and data analysis for the writer. This research expanded the knowledge on how to use information from these systems and how to further use these technologies. The understanding of the research within the field and how to further it by exploring the gaps within them.

The gaps in the use of Power BI in supply chain management were filled partly by this research. When looking at previous research at data analysis in supply chain management it has previously been executed by utilizing nonspecific technologies for analysis by studies by Glass & et. al (2013) and Liu (2019). The program MATLAB has in addition been utilized for supply chain management data analysis research (Hongxiong, 2022). This research's focus on the utilization of Power BI proposes a new method for the research not fully utilized yet. Machine learning utilization in supply chain management research was expanded by this research by studying the utilization with Power BI. Previous research utilized nonspecific programs as in the studies by Peng & et. al (2019) and Gaur & et. al (2015). Machine learning applications for Power BI are a new field which can be expanded and utilized through research in the future.

7.2 Future Development

In the future this system and data queries could be further developed. This is especially true due to the advancements in the use of SAP data at ABB. When ABB builds data cubes inside SAP the warehouse application server query can be utilized which will eliminate the use of Excel as an intermediary to export data from SAP to Power By streamlining and downsizing the program files sizes and complexity. The most useful part of this development would be the real-time updates allowed by this method of data extraction which would allow for even greater reaction to events within the supply chain.

The system could be further developed with the currently chosen data and could be made into an even more important tool within supply chain at ABB Smart Power. The current version created can however be utilized to build more similar systems for the company's other teams and functions to utilize the abundance of data at the company. Example of what the system could in the future provide is the ability to clearly calculate the where the cost of an assembled product comes by looking at component prices compared to the assembled product price. This would allow for even greater benefit for future negotiations.

Machine learning can be utilized in the system itself or other functions for forecasting from this data. Utilizing k-NN algorithm can be used for both price forecasting, but when looking at research done on the field even to material shortages and the planning of order within the supply chain. The use of another algorithm for example the Bayesian network could allow for even more utilization of machine learning within this system and similar systems at the company in the future. The use of Bayesian would be recommended based on the research from Gaur and et. Al (2015). Machine learning should be implemented into this system or a similar system utilizing Python or R-programming instead of DAX-programming. This would allow for a more efficient and developed system but would add an extra requirement for the updates. However, if the updating would be done through data cubes from SAP this problem would be solved.

This researches findings and conclusions offer fertile ground for future development. The utilization data analysis at the company should be focused on in the future, based on the prior research and the lessons learned from this thesis.

References

ABB Website. Retrieved on 12.6.2022 from https://new.abb.com/en.

Ahmadi, E., Ghalehkhondabi, I., & Maihami, R. (2020). *An overview of big data analytics application in supply chain management published in 2010-2019*. Prod. 2020. Vol. 30. https://doi.org/10.1590/0103-6513.20190140

Ang, A., Corral, M., Gan, S., Lu, L., & Tan, J. (2016). *Quality Analytics in a Big Data supply chain: Commodity data analytics for quality engineering*. 2016 IEEE Region 10 Conference (TENCON), Singapore, 2016, pp. 3455-3463. https://doi.org/10.1109/TENCON.2016.7848697

Anusha, D., Panga, M., Hadi Fauzi, A., Sreeram, A., Issabayev, A. & N. Arailym. (2022). *Big Data Analytics Role in Managing Complex Supplier Networks and Inventory Management*. 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 533-538. https://doi.org/10.1109/ICSCDS53736.2022.9761008

Appadoo, S., Banitalebi, B., Hoque, M., Thavaneswaran, A., & Thulasiram, R. (2021). *A Novel Dynamic Demand Forecasting Model for Resilient Supply Chains using Machine Learning*. 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 2021, pp. 218-227. https://doi.org/10.1109/COMPSAC51774.2021.00040

Bakshi, S & Wade, C. (2021). *Expert Data Modeling with Power BI* (Chapter 1 & Chapter 12). Packt Publishing.

Bell, J. (2014*). Machine learning: Hands-on for developers and technical professionals* (Chapter 1). John Wiley & Sons, Incorporated.

Bergin, T. (2018). *An Introduction to Data Analysis Quantitative, Qualitative and Mixed Methods* (Preface & Chapter 1). SAGE Publications Ltd.

Berman, J. (2018). *Principles and Practice of Big Data* (Chapter 1). Academic Press.

Breitenbach, J., Buettner, R., Haileselassie, S., Schuerger, C., & Werner, J. (2021). *A Systematic Literature Review of Machine Learning Tools for Supporting Supply Chain Management in the Manufacturing Environment*. 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 2875-2883. https://doi.org/10.1109/BigData52589.2021.9672013

Campesato, O. (2020*). Artificial Intelligence, Machine Learning, and Deep Learning* (Chapter 1). Mercury Learning & Information.

Catt, P. (2010). *Sales Forecasting with SAP Enterprise Resource Planning An Empirical Study* (Chapter 2.8). LAP LAMBERT Academic Publishing.

Chen, R., Wang, H. & Wang, C. (2001). *A Web-based data extraction system for supply chain management using SAP R/3*. 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37 236), Tucson, AZ, USA, 2001, pp. 2557-2562 vol.4. https://doi.org/10.1109/ICSMC.2001.972948

Cuesta, H. (2013). *Practical data analysis* (Chapter 1). Packt Publishing, Limited.

Decker, G. (2019). *Learn Power BI* (Section 1). Packt Publishing.

Loos, P., Rosemann, M., & Themistocleous, M. (2005). *Enterprise Resource Planning and Enterprise Application Integration* (p. 387-439). Emerald Group Publishing Limited.

Drake, M. J. (2012). *Global Supply Chain Management* (Chapter 1). Business Expert Press.

Escobar, M & Puls, K (2021). *Master Your Data with Excel and Power BI* (Chapter 0). Holy Macro! Books.

Etaati, L. (2017). *Prediction via KNN (K Nearest Neighbours) KNN Power BI: Part 3*. Radacad. Retrieved on 28.3.2023 from https://radacad.com/prediction-via-knn-k-nearest-neighbours-knn-power-bi-part-3.

Evans, B. (7.9.2021) *SAP Buries the Past: Calls Out Legacy ERP, Welcomes Modular Cloud ERP*. Acceleration Economy. Retrieved on 19.3.2023 from https://acceleration-economy.com/cloud/sap-buries-the-past-calls-out-legacy-erp-welcomes-modular-cloud-erp/.

Feng, Y., Su, H., Peng, Z., Zhang, T., & Zhang, Y. (2019). *Deep Reinforcement Learning Approach for Capacitated Supply Chain optimization under Demand Uncertainty*. 2019 Chinese Automation Congress (CAC), Hangzhou, China, 2019, pp. 3512-3517. https://doi.org/10.1109/CAC48633.2019.8997498

Ferrari, A & Marco Russo. (2016) *Introducing Microsoft Power BI* (Chapter 1). Microsoft Press.

Gambit. *Benefits and disadvantages of SAP S/4HANA and SAP ERP*. GAMBIT SAP Consulting. Retrieved on 19.3.2023 from https://www.gambit.de/en/wiki/benefits-disadvantages-sap-s4hana-and-sap-erp/.

Gaur, M., Goel, S., & Jain, E. (2015). *Comparison between Nearest Neighbours and Bayesian Network for demand forecasting in supply chain management*. 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New

Delhi, India, 2015, pp. 1433-1436. Retrieved on 28.3.2023 from https://ieeexplore-ieee-org.proxy.uwasa.fi/document/7100485.

Garg, V., Pathak, P., Rajyashree, R. & Upadhayay, S. (2018). *Big Data Analysis on Supply Chain Management*. 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2018, pp. 1145-1150. https://doi.org/10.1109/CESYS.2018.8723889

Govaert, Govaert, G. & Govaert, G. (2009). *Data Analysis* (Preface). Wiley-ISTE.

Gogate, U & Khalwadekar, R. (2022). *Quantitative and Causal Analysis of Techniques of Microsoft Power BI File Optimisation*. 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2022, pp. 35-39. https://doi.org/10.1109/ICAST55766.2022.10039578

Habib, M. (2011) *Supply Chain Management* (s.2-10). InTech.

He, M, & Xu, R. (2020). *Application of Deep Learning Neural Network in Online Supply Chain Financial Credit Risk Assessment*. 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, China, 2020, pp. 224-232. https://doi.org/10.1109/CIBDA50819.2020.00058

Hongxiong, Y & Mingrui, L. (2022). *Collaborative delivery of information flow in big data supply chain.* 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, 2022, pp. 312-314. https://doi.org/10.1109/EEBDA53927.2022.9744900

Lazzeri, F. (2020). *Machine Learning for Time Series Forecasting with Python* (Chapter 1). John Wiley & Sons, Inc.

Liu, P. (2019). *Decision-making of a three-stage supply chain based on Big Data information costs*. 019 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE), Hangzhou, China, 2019, pp. 190-194. https://doi.org/10.1109/SMILE45626.2019.8965318

Lukka, K. (2001). *Konstruktiivinen tutkimusote*. Metodix. Retrieved 12.06.2022.
From https://metodix.fi/2014/05/19/lukka-konstruktiivinen-tutkimusote/

Lukka, K. (2003). *Case study research in logistics (pp.83-101).* Publications of the Turku School of Economics and Business Administration.

Luna, G & Mora-Arciniegas, M. (2022). *Paper Smart Cities data analysis with Power BI and R*. 2022 IEEE Global Engineering Education Conference (EDUCON), Tunis, Tunisia, 2022, pp. 1824-1828. https://doi.org/10.1109/EDUCON52537.2022.9766385

Microsoft. (2023). *Active vs inactive relationship guidance*. Learn Microsoft. Retrieved on 23.3.2023 from https://learn.microsoft.com/en-us/power-bi/guidance/relationships-active-inactive.

Microsoft. (2023). *Create and manage relationships in Power BI Desktop.* Learn Microsoft. Retrieved on 20.3.2023 from https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-create-and-manage-relationships.

Microsoft. (2023). *Create Power BI visuals with Python*. Learn Microsoft. Retrieved on 20.3.2023 from https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-visuals.

Microsoft. (2023). *Licenses and subscriptions for business users*. Learn Microsoft. Retrieved on 26.3.2023 from https://learn.microsoft.com/en-us/power-bi/consumer/end-user-license.

Microsoft. (2023). *Power BI data sources.* Learn Microsoft. Retrieved on 20.3.2023 from https://learn.microsoft.com/en-us/power-bi/connect-data/power-bi-data-sources.

Microsoft. (2023). *SAP HANA database*. Learn Microsoft. Retrieved on 20.3.2023 from https://learn.microsoft.com/en-us/power-query/connectors/sap-hana/overview.

Microsoft. (2023). *Use the SAP Business Warehouse connector in Power BI Desktop*. Learn Microsoft. Retrieved on 22.3.2023 from https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-sap-bw-connector.

Parthasarathy, S. (2007). *Enterprise Resource Planning: A Managerial & Technical Perspective* (Preface & Chapter 1). New Age International.

Rau, S. (13.4.2016). *Delta Scope – SAP Enhancement Package 8 for SAP ERP 6.0.* Blogs SAP. Retrieved on 19.3.2023 from https://blogs.sap.com/2016/04/13/relased-in-the-first-quarter-2016-enhancement-package-8-what-is-new/.

Renko, S. (2011). *Supply Chain Management* (Chapter 1). InTech.

Sharma, K. & Mutsaddi, A. (2010). *Configuring SAP® ERP Sales and Distribution* (Chapter 1 & Chapter 2). Sybex.

Singh, S & Soni, U. (2019). *Predicting Order Lead Time for Just in Time production system using various Machine Learning Algorithms: A Case Study*. 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2019, pp. 422-425. https://doi.org/10.1109/CONFLUENCE.2019.8776892

Barto, A & Sutton, R. (2014-2015). *Reinforcement Learning: An Introduction. Second Edition, in progress* (Preface & Chapter 1). The MIT Press. Cambridge, Massachusetts, London, England. Retrieved 29.7.2022 from https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf.

Theodoridis, S. (2015). *Machine learning: A bayesian and optimization perspective* (Preface & Chapter 1). Academic Press.

Wade, R. (11.1.2021). *ADVANCED ANALYTICS IN POWER BI WITH R AND PYTHON*. Blue Granite. Retrieved on 23.3.2023 from https://www.bluegranite.com/blog/advanced-analytics-in-power-bi-with-r-and-python.

Zage, D., Glass, K. & Colbaugh, R. (2013). *Improving supply chain security using big data*. 2013 IEEE International Conference on Intelligence and Security Informatics, Seattle, WA, USA, 2013, pp. 254-259. https://doi.org/10.1109/ISI.2013.6578830

Zhang, G., He, L. & Li, S. (2014). *A fuzzy-random optimization approach using fuzzy measure and fuzzy integral for emergency risk coordination of supply chain*. 2014 International Conference on Machine Learning and Cybernetics Lanzhou, China, 2014, pp. 789-795. https://doi.org/10.1109/ICMLC.2014.7009710