



Mika Salonen

Data Driven Pricing and Project Planning

Metropolia University of Applied Sciences

Master's Degree

Industrial Management

Thesis

27.4.2025

Preface

Starting this thesis project was intimidating at first, since the project ahead had many new and unfamiliar aspects in it. However, it quickly became evident that Metropolia University of Applied Sciences instructors have a very strong experience in guiding students through different types of thesis projects. I want to thank Dr. James Collins, Dr. Thomas Rohweder and M.A Sonja Holappa for their insightful commenting and guidance. I want also to thank my supportive student colleagues, who provided excellent comments during the milestone presentations of the thesis.

I am grateful to my employer Mapvision, the case company for this thesis, for making the studies possible by offering flexibility for studies while working full-time and offering a great thesis project idea. I want to thank Bo Lönnqvist for giving invaluable support while being an advisor for this thesis. I express my gratitude to all my colleagues who devoted their time for interviews, feedback and chats during this project.

This thesis is the culmination of the Master's Degree in Industrial Management, and the whole study journey has been very fruitful, and I have learned so much new and have broadened my existing knowledge. I want to thank my family for always keeping the spirits high and being supportive, even during hectic times.

Vantaa, 27th of April 2025

Mika Salonen

Abstract

Author(s): Mika Salonen
Title: Data Driven Pricing and Project Planning
Number of Pages: 74 pages + 3 appendices
Date: 27th of April 2025

Degree: Master of Engineering
Degree Programme: Industrial Management
Specialisation option: Data Driven Business
Instructor(s): Dr. James Collins, Senior Lecturer
Bo Lönnqvist, Chief Business Officer

This thesis focuses on improving decision making in a case company based on its operations data, using applied research method. The research was conducted following a Metropolia University of Applied Sciences Industrial Management GATE system, which this thesis report follows closely.

The objective of this thesis is to propose a data analysis tool and its parameters to enable evidence-based decision-making in the case company. The research starts with a literature review, from which a conceptual framework for current state analysis and tool development is developed. The following current state analysis focuses on analyzing the case company operations data and its quality. The current state analysis provides a good overview of the quality of the case company data, and this overview can be used to guide further development tasks as well. After the current state review, a data analysis tool proposal is co-created with key stakeholders, where each key stakeholder shares their own input and wishes for the tool. Lastly, the co-created tool is validated by an end-user group, and feedback from the group is taken into consideration and the tool is further improved before releasing it into wide scale use.

The final improved data analysis tool is a robust and extensive Power BI report, which categorizes work hours and hardware costs from previous projects, and allows user to slice the underlying data with related parameters. The tool visually shows the meaning of data, and it allows users to comprehend the underlying data without the need for a data analyst background.

Keywords: Data Analysis, Microsoft Power BI, Data-Driven Decision Making, Data Visualization, Project Planning

The originality of this thesis has been checked using Turnitin Originality Check service.

Contents

Preface

Abstract

Contents

List of Figures

List of Tables

List of Abbreviations

1	Introduction	1
1.1	Business Context of the Case Company	1
1.2	Business Challenge, Objective, and Outcome	2
1.3	Scope and Outline of Thesis Report	2
2	Project Plan	4
2.1	Research Approach	4
2.1.1	Applied Research	4
2.1.2	Applied Action Research	4
2.2	Research Design	5
2.3	Data Plan	6
3	Design Principles for Data Analysis from Literature	8
3.1	Data Requirements and Acquisition	9
3.1.1	Data Quality Requirements	10
3.1.2	Data Mining and Visual Data Mining	12
3.2	Exploration of Data	13
3.2.1	What Data to Search and Refining the Search	13
3.2.2	Formulating Hypothesis	14
3.3	Data Preprocessing	15
3.3.1	Data Filtering, Cleaning & Generation and Integration	15
3.3.2	Missing Data	16
3.4	Model and Tool Proof-of-Concept and Validation	17
3.4.1	Data Modeling with Microsoft Power BI	17
3.4.2	Sensitivity Analysis and Model Validation	19
3.4.3	Uncertainty in Data Analysis and Visualization	19
3.5	Communication, Implementation and Quality	20
3.5.1	Data Visualization	20

3.5.2	Data Model Quality	21
3.6	Conceptual Framework	22
4	Current State Analysis of Mapvision Operations Data	25
4.1	Overview of This Data Stage	25
4.2	Operations Data Sources	27
4.3	Data Quality Analysis	32
4.3.1	CRM Data Quality Assessment	32
4.3.2	PDM Data Quality Assessment	35
4.3.3	ERP Data Quality Assessment	35
4.3.4	Dataverse data quality assessment	38
4.4	Mapvision Operations Data Current State Summary	41
4.5	Current Data Reporting and Analysis Tools	42
4.6	Data Culture	43
5	Data Analysis Tool Solution Proposal	44
5.1	Overview of This Data Stage	44
5.2	Defining the Solution Proposal	45
5.2.1	Project Hour Categorization	46
5.2.2	Hardware Cost Categorization	47
5.2.3	Relationships Between CRM Parameters and Workhours	47
5.2.4	Relationships Between CRM Parameters and Hardware Costs	48
5.2.5	Data Pre-processing and Filtering	49
5.2.6	Power BI Report Pages	50
5.2.7	Summary of The Tool Definition	50
5.3	Creating the Solution Proposal	52
5.3.1	Data Model	52
5.3.2	Analysis Page	55
5.3.3	Hour Role Categorization Function	58
5.3.4	Uncertainty	58
5.3.5	Model Sensitivity Analysis	59
5.3.6	Self-Service Analytics Pages	60
5.4	Summary of the Solution Proposal	61
6	Validation and Improvement of The Solution Proposal	63
6.1	Overview of This Data Stage	63

6.2	Feedback Received and Improvements to Solution Proposal	64
6.2.1	Visual Improvements	65
6.2.2	Functionality	67
6.2.3	New Pages	68
6.2.4	Data Model Quality	69
6.3	Summary of Final Data Analysis Tool	70
7	Conclusions	71
7.1	Executive Summary	71
7.2	Next Steps and Recommendations	71
7.3	Evaluation of Thesis Trustworthiness	73
7.4	Closing Words	74
	References	75
	Appendices	78

List of Figures

Figure 1. Research design of this thesis	5
Figure 2. Conceptual Framework for Design Principles for Data Analysis	23
Figure 3. Current State Analysis Data 1 interview form questions	26
Figure 4. Mapvision operations data overview	27
Figure 5. Filtered data sources and their relationships	28
Figure 6. Project costs categorized per routing links during data mining	29
Figure 7. Filtered project cost overview	30
Figure 8. Hours logged for projects	31
Figure 9. Hour data filtered by the 49 projects in ERP and all new projects created after that.....	31
Figure 10. Data quality assessment matrix	32
Figure 11. Filtered CRM data Completeness Simple ratio	33
Figure 12. Filtered CRM data Free-of-Error Simple ratio	34
Figure 13. CRM combined Simple ratio	35
Figure 14. ENC_INS routing link cost	36
Figure 15. INT_PRE_AC routing link cost.....	36
Figure 16. Project hour distribution per team	39
Figure 17. Project hours categorized per project phases: Project Management, Design, Assembly, Setup, Implementation and Other.....	39
Figure 18. Bundle delivery projects.....	40
Figure 19. Two copy systems compared	40
Figure 20. Data qualities in assessment matrix.....	41
Figure 21. Hour definition for the tool.....	50
Figure 22. Hardware cost definition for the tool.....	51
Figure 23. First draft of data model.....	53
Figure 24. Final data model	54
Figure 25. Data analysis tool.....	56
Figure 26. Average Base System HW Cost DAX measure.....	57
Figure 27. A sample category normal distribution	59
Figure 28. Self-service analytics page for hardware cost.....	60
Figure 29. Self-service analytics page for hours	61

Figure 30. Final data analysis tool	66
Figure 31. Improved final data model.....	69

List of Tables

Table 1. Data plan.....	6
Table 2. Literature review structure.....	8
Table 3. Subjective survey and objective assessment matrix	11
Table 4. Current state analysis interviewees	25
Table 5. ERP BOM Free-of-Error Simple ratios	37
Table 6. Zero cost items	38
Table 7. Co-creation key stakeholder meetings / chats	44
Table 8. Data stage 3 testing group members	64

List of Abbreviations

ERP	Enterprise Resource Planning
PDM	Product Data Management
CRM	Customer Relationship Management
PM	Project Management
COO	Chief Operations Officer
CBO	Chief Business Officer
PO	Production Operations
CFO	Chief Financial Officer
CPO	Chief Product Officer
PE	Project Engineering
BI	Business Intelligence
2D	Two Dimensional
3D	Three Dimensional
BOM	Bill of Material
CSWITCH	Camera Switch
INT_PRE_ACC	Internal Pre-Acceptance
FXT_ASM	Fixture Assembly
CU_ASM	Control Unit Assembly
SHUTT_ASM	Shuttle Assembly
EC_INS	Electrical Cabinet Installation
ENC_INS	Enclosure Installation
ARTEF_ASM	Artefact Assembly
MCK	Manual Control Kit
DAX	Data Analytics Expression
PROD	Production
SYS_ASM	System Assembly
IMPL	Implementation
FIFO	First In First Out
AI	Artificial Intelligence
WSI	Weld Seam Inspection

1 Introduction

Companies today produce and acquire substantial amounts of data in all business sectors. Using and analysing data gives companies a competitive edge over those who choose not to harness their data for their benefit. Data usage varies from comparing and reporting to advanced forecasting and evidence-based decision making. (Gülseçen, et al., 2020, pp. 71-92). To be able to use data efficiently, companies must ensure data quality is flawless by analysing and correcting the data at its source.

1.1 Business Context of the Case Company

The case company in this thesis is Mapvision Oy, a Finnish based company which designs and manufactures automatic optical quality inspection systems which are implemented globally in automotive customer manufacturing lines. The manufacturing lines are producing parts such as subframes, cross car beams, floors, battery components, shock towers, body in whites and other car components with low tolerances and short cycle times.

Mapvision Q-series system offering has a set of standard sizes and loading options with configurable add-on sub systems. Each system has a part-specific part holding fixture or multiple fixtures, and a camera setup. A fixture is a device used to hold the inspected part in predetermined location and shape. A camera setup is an array of cameras in the Mapvision system, used to grab images of the inspected part.

Mapvision F-series systems are fully integrated into welding lines, allowing large selection of different assemblies to be welded and inspected in the same cell without the need for part specific welding and inspection fixtures.

1.2 Business Challenge, Objective, and Outcome

Mapvision has recently conducted a change project where new Enterprise Resource Planning system (ERP), Microsoft Business Central, was implemented. Mapvision is producing large quantities of data into the new ERP, but currently the data is not used to its full potential since it is mainly used for reporting purposes only. Mapvision is not fully confident about the data quality and is not aware of all the possible use cases for the data.

The objective of this thesis is to propose a tool, or a blueprint of it, which can be used during quoting process and in project planning to make evidence-based decisions based on data. The thesis will also include assessment of Mapvision operations data and its quality. Data error correction will be done and a proposal how to produce and manage new data to avoid reproducing found errors is created.

Consequently, the outcome of this thesis is a data analysis tool, or a blueprint for a tool, for improved product pricing and project capability forecasting. The tool is fully connected to Mapvision data sources, and all newly created data is automatically refreshed into it. The tool reports artefacts in the data quality when it finds them, prompting user correction when needed.

1.3 Scope and Outline of Thesis Report

This thesis focuses on Mapvision operational data. Financial, research and development and support function data is outside the scope of this thesis. The outcome of this thesis is a supporting tool for pricing decisions and project planning; however, the pricing models are outside of the scope of this thesis. All business sensitive data, such as customer names, costs, project numbers and work hours are hidden. During the data assessment, other use cases for the data might be found. These use cases are documented, but their further analysis, development or implementation are not in the scope of this thesis.

This thesis contains seven sections. The introduction is followed by section 2, which explains in detail how the thesis is conducted. Section 3 focuses on existing knowledge and best practices in the field of business data analysis. Fourth section describes the current state of Mapvision data and its current use. Section 5 is built on the outcomes of sections 3 and 4 and describes the concept proposal for the end solution, which is validated in section 6. The closing section of this thesis provides the conclusions with executive summary and managerial implications along with an evaluation of the thesis.

2 Project Plan

This section describes the project plan on which this thesis is based on. First, the selected research approach is introduced, and reasoning why it is selected is given. Secondly, the research design presents in detail how this thesis is conducted. Lastly, the data collection methods and meaning of research data in applied action research is described.

2.1 Research Approach

Thesis research approach defines the methodologies used to conduct the research to solve a research problem. The following sections 2.1.1 and 2.1.2 will describe the selected research approach for this thesis.

2.1.1 Applied Research

Business and management research approaches can be placed on a continuum which have two extremes: basic research and applied research. Basic research, which is purely theoretical, is undertaken by people based in academia and the findings are universal principles relating to a process and its relationship to outcomes, with relatively little relevance to or impact on practice. Applied research results in solution to a problem, and the new knowledge is limited to the problem. Applied research has direct and immediate use for managers, and addresses issues they see as important, and is presented in a way that they effectively understand and take actions with. (Saunders, et al., 2019, pp. 6-11).

2.1.2 Applied Action Research

Applied action research, also known as design research, starts from where qualitative and quantitative research ends. Qualitative research aims to understand a phenomenon by providing words and sentences to open questions, and quantitative research aims to generalise a phenomenon providing numbers to structured questions, while action research aims to change the states of affairs. In addition to text, applied action research also

produces solutions which work in practice. (Kananen, 2013, pp. 27-30). This thesis aims to produce a practical solution to the case company and therefore the used research approach for this thesis is applied action research.

2.2 Research Design

This section describes in detail how the thesis is conducted with the aid of a visual presentation. This thesis is following Metropolia University of Applied Sciences Industrial Management Master's thesis Gate model, which has six gates, shown below in Figure 1. Each gate has planned tasks in them, gates 3-6 have an outcome, and gates 4-6 have a data collection section embedded in them. The data collection plan is presented in section 2.3.

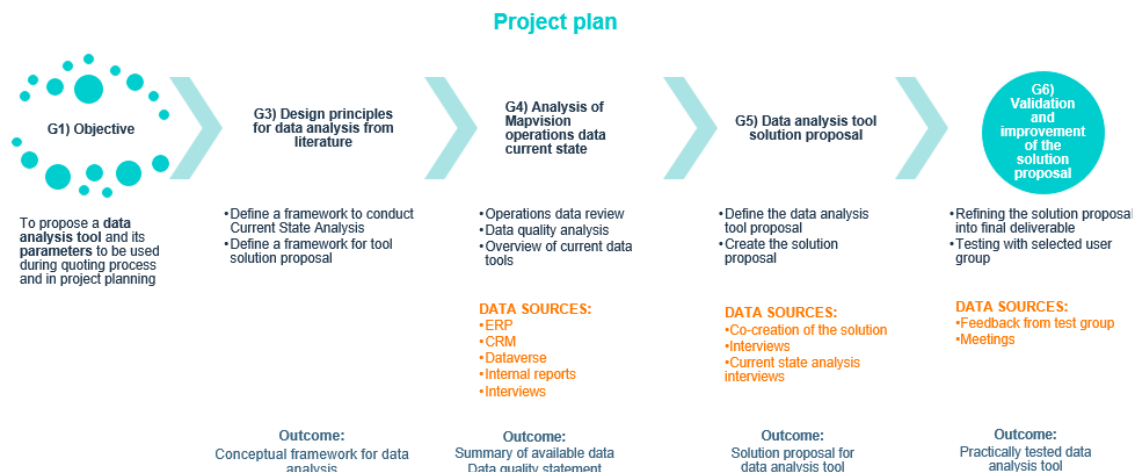


Figure 1. Research design of this thesis

As shown in Figure 1, the gates are linked to each other, and they are completed in order from gate one to gate six. The outcomes from the gates have crucial roles in building the final deliverable of this thesis. Outcomes of previous gates are used as an input for later gates in logical order. Gate two is the generation of Figure 1 and Table 1, and therefore is not shown in the illustration.

In this thesis, gate three literature review is conducted before gate four current state analysis, because the outcome, conceptual framework, is needed to be

able to analyse the current state in a meaningful way to produce necessary outcome for the later gates.

The current state of data, its quality and its use is analysed during gate four, based on conceptual framework defined in gate three. Outcomes of gates three and four are used during gate five to co-create a proposal for the final deliverable. The final deliverable is then refined and practically tested during gate six, based on all previous gate outcomes.

2.3 Data Plan

Gates four, five and six have a data collection stage in them. The data in these stages is research data used to conduct applied action research and must not be mixed with the operations data this thesis is focusing on. The data plan describes data type, source, informant timing and outcome for each three data stages below in Table 1.

Table 1. Data plan

	Data type	Source	Informant	Timing	Outcome
Data 1 Analysis of Mapvision operations data current state	<ul style="list-style-type: none"> Raw operations data 	<ul style="list-style-type: none"> ERP CRM Dataverse Interviews 	<ul style="list-style-type: none"> Head of PM COO CBO Head of PO CFO CPO 	February	List of available data and statement of its quality
Data 2 Data analysis tool solution proposal	<ul style="list-style-type: none"> Recommendations Proposals 	<ul style="list-style-type: none"> Co-creation Meetings 	<ul style="list-style-type: none"> CBO Head of PM Head of PE Head of PO CPO Business Controller 	March	Solution proposal for data analysis tool
Data 3 Validation and improvement of the solution proposal	<ul style="list-style-type: none"> Recommendations Feedback 	<ul style="list-style-type: none"> Feedback from test group 	<ul style="list-style-type: none"> Testing group (11 people) 	April	Practically tested data analysis tool

Research data is collected in three different stages of the thesis: first in current state analysis to document what data is available and what is the quality of it. Secondly, research data is collected at the beginning of the co-creation phase

of the solution to collect recommendations for the solution proposal. Lastly, after the solution proposal, a final research data collection is conducted, and these recommendations and feedback are used to refine the solution into its final form.

The informants are selected based on their possible relevant input for the data stage. In data stage 1, Head of Project Management has the best knowledge on project reporting, Chief Operating Officer has the largest knowledge on overall working hours, Chief Business Officer is the best source for current pricing and customer data, Head of Production Operations knows about hardware allocation and assembly hours, Chief Financial Officer has strong understanding on ERP, while Chief Product Officer has strong understanding on CRM.

Data stage 2 has more narrowed informant group, where the solution proposal will be co-created with relevant stakeholders. Recommendations will be collected from all informants in this stage.

In data stage 3 the solution proposal is tested with larger end user group, and their feedback and recommendations are used to finalize the tool during gate six. The data model quality will be analysed by another Mapvision Power BI developer.

This concludes the Project Plan section. The next section focuses on literature review to establish conceptual framework for the gates four, five and six.

3 Design Principles for Data Analysis from Literature

This section defines the conceptual framework which is used in Current State Analysis in the next section and in the solution proposal and validation in sections 5 and 6. The conceptual framework is presented and explained in detail in section 3.6, which is carefully crafted based on findings from literature to be best suited for the objective of this thesis. Next subsections present the findings from literature, categorized as shown in Table 2.

Table 2. Literature review structure

Current State Analysis	Tool Proposal and Validation
3.1 Data Requirements and Acquisition	3.3 Data Preprocessing
3.2 Exploration of Data	3.4 Model and Tool Proof-of-Concept and Validation
3.3 Data Preprocessing	3.5 Communication, Implementation and Quality

The subsections are based on two articles: Data science essentials in business administration: A multidisciplinary perspective (Afsharian, 2024) and Principles for data analysis workflows (Stoudt, et al., 2021).

Afsharian (2024) decodes the tasks of data scientist team in six phases: Recognize the situation & grasp objectives, Data requirements, acquisition & comprehension, Data preparation & cleansing, Data preprocessing & modeling, Model validation/ Proof-of-concept and Communication & implementation. Afsharian argues that the methods and tools used by data scientists can be categorized in three overlapping disciplines: Computer Science, Statistics and

Operations Research. He notes three different analytics emerge from combinations of these disciplines: Descriptive Analytics, which answer the question what happens and why; Predictive Analytics, which answers the question what is likely to happen; and Prescriptive Analytics, which answers the question what should be done. (Afsharian, 2024). The objective of this thesis is in the realm of Descriptive Analytics and Prescriptive Analytics.

While Afsharian (2024) focuses on the tasks of a Data Scientist, Stoudt et al. (2021) state that individuals who do not identify as Data Scientist, are working on large and complex datasets regularly. They have therefore created a basic principle for a data analysis workflow, which has three phases: Explore, Refine and Produce. They suggest that during the exploration phase, individuals “meet” their data, process it, and find potential solutions to a problem of interest. In the refine phase, the focus should be narrowed to a particularly promising approach and prototypes should be developed. The produce phase happens concurrently to the earlier phases and during this phase, the work is prepared for larger audience to be used and critiqued. (Stoudt, et al., 2021). Both articles note that domain knowledge is one of the key traits required to conduct data analysis with real world applications (Afsharian, 2024) (Stoudt, et al., 2021).

As can be seen in Table 2 above, 3.3 Data Preprocessing is in both categories. This is because data preprocessing must be conducted already in the current state analysis to visualize and analyse the findings and later to modify the data during modeling.

3.1 Data Requirements and Acquisition

Certain Mapvision operations data attributes, such as stock item quantities and their costs and overall workhours per project, are subjected to governance due to their direct impact to the company financial statements. The quality of this data can be assumed to be “good.” Other attributes, such as Routing Link Code, which defines in which exact project stage a component or work hour

belongs, is not governed by outside parties. However, this kind of detailed data could be the most valuable data for this thesis, and for the company in general.

The next sub sections will define the requirements for data quality for this thesis and the methods to collect the data and the definition of “good” data will also be described. Overall operations data quality assessment is then presented in section 4.3, based on the conceptual framework.

3.1.1 Data Quality Requirements

This section defines data quality requirements and a concept for data quality measurement to be used in this thesis. High-quality data is crucial for any decision-making model, since these models usually aggregate values together, multiplying any wrong values and skewing the results when low-quality data is used (Haiyan, 2024).

Haiyan (2024) characterizes high-quality data from business perspective as data which “is defect-free and fulfills the intended purposes in operations, decision-making, and planning”. Based on Haiyan’s research, inconsistent definition for common terms and data entries by employees are the most common sources for sources for data quality issues, followed by data migration and conversion projects. (Haiyan, 2024).

Data quality assessment should be carried out to uncover the correlation between data and data quality requirements. Data quality assessment report includes trends, processes, observations, recommendations, and suggestions for identifying and removing sources that lead to business exceptions. Haiyan (2024) notes that different organizations have different requirements for data quality assessment. (Haiyan, 2024).

Pipino et al. (2002) argue that data quality assessment should have subjective survey and objective assessment. The results should then be compared to find discrepancies and root causes for issues, and then actions should be taken to correct the root causes to eventually improve the data quality. Subjective survey

identifies what is the user's opinion of the data quality, while objective assessment measures the data "correctness". (Pipino, et al., 2002).

In this thesis current state analysis objective assessment is carried out by using Simple Ratio measurement. The Simple Ratio measures the ratio of desired outcomes to total outcomes, where outcomes represent a value of data attributes. The following quality metrics will be measured in this thesis with simple ratio: free-of-error, which means an attribute value is correct, and completeness, which means a dimension has all the required attributes. (Pipino, et al., 2002). Subjective surveys are conducted during the interviews of stakeholders in form of structured questionnaires.

According to Pipino, et al. (2002), data quality assessment analysis is done by placing results from both subjective survey and objective assessment to a 2x2 matrix, illustrated below in Table 3.

Table 3. Subjective survey and objective assessment matrix

Subjective assessment	High	II	IV
	Low	I	III
		Low	High
		Objective assessment	

Data quality can be assumed good when the outcome of the analysis falls into quadrant IV. If the outcome falls into any other of the quadrants, actions must be taken to improve the data quality. Quadrant I indicates the data is objectively and subjectively low quality, and the data must be corrected at its source. Quadrant II denotes there is an issue with the data presentation, falsely presenting the data to be high quality, or the user requires training to be able to understand the data. Quadrant III is similar to quadrant II, where the issue lies

either on low quality representation of data, or in lack of training. (Pipino, et al., 2002).

The quality measurement methods and appropriate quality metrics for the final thesis outcome are discussed in section 3.5.2.

3.1.2 Data Mining and Visual Data Mining

This section defines the methods used in this thesis to collect data for analysis. Data sources for this thesis are defined in the current state analysis section 4.2. Data mining is one area of knowledge discovery process, where useful information and knowledge is extracted automatically from large amounts of data (Espadinha-Cruz, et al., 2021) (Kopanakis & Theodoulidis, 2003).

Espadinha-Cruz, et al. (2021) argue that data mining can provide significant and valuable information from data, which was previously unidentified due to the size of the dataset. They note data mining is a process for sorting and classifying data, which then enables detection of patterns, anomalies and correlations that are hidden in the dataset. These discoveries can be used to predict outcomes. (Espadinha-Cruz, et al., 2021). Such discoveries are used in this thesis data-model together with human input to generate evidence-based analyses to help in decision making. Espadinha-Cruz, et al. (2021) note that due to the automated nature of datamining, accuracy of the data mining model is paramount, and missing data can cause issues that require human intervention (Espadinha-Cruz, et al., 2021). How missing data should be handled is discussed in section 3.3.2.

Visual data mining techniques build on traditional data mining and display the results in a visual form. Visualizing analysis results lets human interpret what the data has to say, taking advantage of human ability to perceive patterns and structure in visual form. (Kopanakis & Theodoulidis, 2003). This thesis will take advantage of data analysis visualization done in Microsoft Power BI software, which is discussed in detail in section 3.4.1.

Kopanakis & Theodoulidis (2003) argue that visual data mining could be characterised as a link between two of the most powerful information processing systems: humans and computers. They add that humans are easily overwhelmed by the amount of data available, and computerized data mining can complement the human visual capabilities. While data mining is mostly an automated process, data visualization is an interactive process between human and computer. If all data would be visualized without human input, the resulting visual could become overwhelming and fuzzy. (Kopanakis & Theodoulidis, 2003).

3.2 Exploration of Data

Stoudt et al. (2021) explains that exploration of data is usually the first phase of data intensive research project, where a domain-specific question or a particular data set is presented. In this thesis, the domain-specific question can be viewed as the objective of the thesis. The next sub sections describe first how to select the data for analysis, when the domain specific question is known, and how to change or refine the search during the analysis, and secondly how to form hypothesis from the initial discoveries.

3.2.1 What Data to Search and Refining the Search

Stoudt et al. (2021) point out that exploration phase is rarely included as an explicit step of a data intensive research, but it is an essential part of research because it allows the researcher to gain intuition of the data, informing future phases of the workflow. They also claim that it is essential to refine the research question during the exploration phase to articulate a well-defined problem, also noting the explore phase can be rather “messy”, meaning multiple avenues are considered, and many are abandoned. (Stoudt, et al., 2021). On the other hand, Stone (2002) argues that the research question cannot be too loosely defined, since then the researcher could unintentionally drift towards searching for ‘significant’ results, which could cause misleading findings.

Stone (2002) points out that if the research question is not refined and well considered, the subsequent labour-intensive steps will be wasted. Stone's statement is true for this thesis as well, since if the results from the outcome, which is a data analysis tool, give no insight or added value to the business problem at hand, the thesis work has failed to reach its objective.

The following key goals, that Stoudt et al. (2021) underline, are kept in mind during this thesis exploration phase: "determine if questions of interest could be answered with existing tools or software, without building extensive new data models" and "create extensive documentation of the explore phase for later steps of the analysis".

3.2.2 Formulating Hypothesis

Felin et al. (2021) strongly argue that there is no data exploration without hypothesis, since observation and data are always either theory- or hypothesis-loaded. They add that any data exploration is meaningless without some form of theory or hypothesis, since when exploration is done, it must be guided by some expectations. (Felin, et al., 2021). Interestingly, Felin et al. (2021) even conclude that hypothesis-free data observation is not possible, since there is no way to extract and identify anything hidden without at least a rough idea what could be found.

The arguments made by Felin et al. (2021) are not explicitly denied by Zhao et al. (2017), but their focus is on multiple hypotheses testing during data exploration, and they claim that recent tools created for interactive data exploration significantly increase the chance of false discoveries. Zhao et al. (2017) state that this problem is known in the statistics community and is referred to as "multiple comparisons problem" or "multiple hypothesis error", and the problem is stated as follows: "the more hypothesis tests an analyst performs, the higher is the chance that apparently significant phenomenon (i.e., a "discovery") is actually observed just by chance."

Interestingly Zhao, et al (2017) point out that any visualization becomes a visual hypothesis once a filter or slicer is applied to it by the user. If the applied filter or slicer does not change the distribution of the visual, it is a null hypothesis, meaning no correlation between the data and the filter exist. They instruct to make two visuals side by side, where one has a filter, or hypothesis, applied and another one does not, to visually verify if there is a difference in the data distribution. Importantly, Zhao et al. (2017) note that every heuristic can be simply wrong, and user must be able to delete already made hypotheses.

3.3 Data Preprocessing

This subsection discusses preprocessing methods employed to modify a dataset once it has been selected from the data exploration phase to enhance the data modeling efficiency and accuracy.

3.3.1 Data Filtering, Cleaning & Generation and Integration

Filtering unrelated items aims to remove data, which clearly is not related to the research question, in other words, is not interesting. Filtering can be either done manually or applying automatic filters to a dataset to remove undesired values. Downsizing the data sample lowers the computational power required later in data modeling phase. (Sirichanya & Kraisak, 2021).

Data cleansing, or data cleaning, is a preprocessing method which identifies incomplete records, outliers, incorrect values and noise and then aims to improve the data quality. Depending on the dataset size, the improvement is either done by modifying the data manually or applying automatic rules to either ignore or smooth out the identified data values by averaging them. (Alasadi & Wesam, 2017) (Sirichanya & Kraisak, 2021).

Data integration is a method used to integrate two or more data sources into one dataset, which is then used in the data model, creating more information rich dataset overall. During data generation new value expressions are generated by combining existing expressions together or applying knowledge

from the ontology that describes the dataset itself. Combination of existing values can be done before or after data integration (Alasadi & Wesam, 2017) (Sirichanya & Kraissak, 2021).

3.3.2 Missing Data

Data containing missing values is identified either by using data quality assessment described in section 3.1.1 or by recognizing it during visual data mining, described in section 3.1.2.

Larose (2014) notes that just ignoring missing data can be dangerous for the overall analysis, since it is not known if the pattern of missing data is systematic. Ignoring a systematic set of values, even without knowing it, will lead to a biased dataset. In addition, if only one or two attributes are missing from a field, it would be wasteful to discard the whole field of possibly correct and significant data. (Larose, 2014).

The following methods are commonly used to input data to missing fields: 1) input some constant, defined by the analyst, 2) input a mean value, calculated from other similar variables, 3) input a randomly generated value from the distribution of similar variables or 4) input an imputed value, based on other characteristics of the record. (Larose, 2014) (Pigott, 2001).

Larose (2014) argues that imputation is the most beneficial method of inputting missing values, but imputation requires considerable domain knowledge and judgement. In addition, imputation requires complex models such as multiple regression or classification. He also states: if any missing value is added to the dataset, the end-user must be notified about it. (Larose, 2014).

Pigott (2001) points out that before attempting to use any previously listed methods to add missing values into the dataset, the root-cause must be made clear. When the root-cause is known, the appropriate method for input can be selected. Like Larose (2014), Pigott (2001) argues that multiple imputation seems to be the most promising method to handle missing values. Multiple

imputation replaces missing values multiple times, utilizing computer processing power to do so. Pigott (2001) notes the required software packages and computational power as downsides for multiple imputation.

3.4 Model and Tool Proof-of-Concept and Validation

This subsection defines the best practices to build a data model, assuming the previously discussed steps have been completed:

1. Data quality has been assessed and improved to be good
2. Data has been selected in exploration phase
3. Data has been pre-processed

3.4.1 Data Modeling with Microsoft Power BI

Microsoft Power BI is one of many data analytics tools available in the market. Mapvision uses Microsoft platform and Power BI, and therefore this subsection focuses only on data modeling with Power BI. Data analytics tool comparison is outside the scope of this thesis.

In Power BI, data sources can be linked together using relationships, and this setup needs to be done only once per semantic model. Semantic model in Power BI is a model, which represent the structure of the data to be analysed. Semantic models can be configured to automatically update the data source queries, allowing users to always have the latest data available. End users do not need to use, or have the knowledge how to use, the Power BI Desktop semantic model, but multiple automatically updating reports can be published from a single semantic model. The reports can be interactive, if the semantic model creator chooses to do so. Multiple users can interact with the reports, allowing self-service analysis, meaning end-users do not need to rely on single or few data analyst for their data analysis needs. (Bansal, 2023) (Deckler & Powell, 2024).

Deckler & Powell (2024) note that data sources, or data tables, should be categorized as Fact tables and Dimension tables. Fact tables store the information of business processes, such as sales, while Dimension tables add information to Fact tables, such as product ID, country, colour of the product and so on. The relationships between Fact and Dimension tables should always, when possible, be many-to-one. This is because, usually in Fact tables, for example, there are records of many different sales of a product, but there is only one unique line for a product describing the characteristics of the product in the Dimension table. In Power BI, the table linking also describes the flow direction of filtering. Best practice is to have one direction flow from Dimension table to Fact table, which means when the Dimension table is filtered, the Fact table is filtered accordingly, but no other Dimension tables are filtered through the Fact table, since all the filters flow only from Dimension tables to Fact tables. Multiple Fact tables can be linked to common Dimension tables, but generally Fact tables should never be linked together (Deckler & Powell, 2024).

Deckler & Powell (2024) argue that since multiple reports can be generated from single semantic model, DAX (Data Analysis Expressions) queries should be implemented in the semantic model layer instead of separately on each report. They add that it is crucial that the logic used to build DAX queries reflect business requirements, and the logic used is clearly documented. (Deckler & Powell, 2024).

When planning and designing a Power BI report, at least the following questions should be answered:

1. Who will be using this report and why?
2. What are the priorities of this report in terms of business questions?
3. Is Power BI the right tool for the report?

To avoid convoluted user experience, one report should not try to answer all the questions. (Deckler & Powell, 2024).

Power BI has multiple in-built AI (Artificial Intelligence) features, which can be configured, for example, to find anomalies in the data, or let users ask questions about the data using natural language (Diepeveel, 2022). Adding AI features to the model is outside the scope of this thesis, but possibilities are discussed in section 7.2.

3.4.2 Sensitivity Analysis and Model Validation

Model sensitivity analysis aims to understand how sensitive the model is to fluctuations in the model input data and on the parameters the model is built on. If a model is either too sensitive or not sensitive enough, the output will not reflect the input parameters correctly. Testing and validating model sensitivity can be conducted by creating a secondary experimental model and comparing the results of them, testing the model with different dataset or changing the model assumptions and observe how the model output changes. Identifying failing model early is crucial, since failing the sensitivity analysis might require totally new model to be built. (Salciccioli, et al., 2016).

After the model has reached appropriate sensitivity for a given task, the model needs to be validated to ensure that the outputs can be trusted. Amigo (2021) argues that there is no universal quantity of sample tests to declare a model validated. He states that the model needs as many samples as necessary to verify the reliability and representativeness. (Amigo, 2021). Salciccioli, et al. (2016) state that if the dataset is large enough, it can be split into two or more datasets, and then the first one is used to build the model, and the rest of the datasets are used to validate the model. If a dataset is large enough, splitting it into two or more sub-groups should still represent the original dataset well enough for validation purposes. (Salciccioli, et al., 2016).

3.4.3 Uncertainty in Data Analysis and Visualization

When calculating with uncertain data, the uncertainty is propagated. A common method to visualize uncertainty in results, is to provide another visual next to the results, such as standard deviation, which shows the uncertainty level. If the

visual allows, without disturbing the readability, the uncertainty could be added as overlay to the result visual. (Brodie, et al., 2012).

Galvão & Mitchell (2024) discuss if communicating uncertainty is good or bad idea, and they state it depends on the communicating method and the receiver's ability to comprehend it. If the communication method is "too complex" the receiver's trust for the data decreases, but on the other hand if the uncertainty is not communicated at all, the receiver might start to debate if the data can be trusted, or they could imagine their own uncertainty approximates. Galvão & Mitchell (2024) conclude that according to their research, data uncertainty should be included along with the results, and the most effective tools are those that qualify and visualize data uncertainty. (Galvão & Mitchell, 2024).

3.5 Communication, Implementation and Quality

This section discusses how the outcome of this thesis should be communicated and implemented, and what are the best practices used to create a robust and good quality data model.

3.5.1 Data Visualization

Visualizing data aids humans to comprehend the data better when compared to seeing only numbers in tabular form. However, the visualization method must be selected according to the data at hand, and not all data can be visualized in the same format. It is also important to note that not all data must be visualized, and visualizations should be placed in a context they can be understood without added descriptions. (Healy, 2019) (Romero-Organvidez, et al., 2024).

Romero-Organvidez, et al. (2024) argue that only graphs which are represented as points, lines, bars, and boxes are recommended, and 2D and 3D shapes should be avoided. Point, line, bar and box graphs rely on visual attributes which are easily and accurately perceived. Bar graphs are excellent to visualize comparison between values and line charts are good to show time series

change. Boxes are good only in certain special conditions, while points excel in distribution, correlation and geospatial visualization. (Romero-Organvidez, et al., 2024). Graphs help to keep the data-ink ratio high, removing unnecessary clutter from the visual. Data-ink ratio means the amount of “ink” which presents the data compared to the total amount of “ink” which presents the visual. (Romero-Organvidez, et al., 2024) (Healy, 2019).

Colours affect the visual perception considerably, and the following colours meet the requirements of distinctiveness, meaning they can be used together: grey, blue, orange, green, pink, brown, purple, yellow and red. However, if red and green are used together, the visual could become hard to differentiate for colour-blind people. (Romero-Organvidez, et al., 2024). Colour themes should be selected according to the data which is visualized: sequential scales for qualitative results, which are easily distinguishable, but have the same valence for the viewer and 2) diverging colours for comparison or variance from a mean or zero point. (Healy, 2019).

3.5.2 Data Model Quality

The quality factors used for data model are different from those used for data quality. The following factors should be taken into consideration when evaluating data model quality: integration, completeness, simplicity, correctness, flexibility, implementability, integrity and understandability. (Helskyaho, et al., 2024) (Moody & Shanks, 2002).

When model has high integration level, it means the model is consistent with the rest of the data and models the organization has. Completeness indicates if the model contain all the requirements set for it during the specification phase. Simplicity implies does the model have only the essential relationships and attributes, or does it have unnecessary complexity built into it. Correctness measures how well the model follows common data modeling techniques. Flexibility denotes how well the model adapts to business environment changes. Implementability reveals how efficiently the model can be deployed to

business environment when budget and time are taken into consideration. Integrity is important in financial applications, where data integrity must be guaranteed. Understandability signifies how well the data model structure and methodology can be understood by other users or developers. (Helskyaho, et al., 2024).

Moody & Shanks, (2002) suggests that data model quality verification should be done as early possible during the development to lessen the correction workload later, when the quality issues have propagated, and the corrections must be made in shorter time. They argue that high product quality is a made possible by having high process quality, where product refers to the data model and process refers to the development of the model. (Moody & Shanks, 2002).

Data model quality assessment should be conducted by using subjective assessments from relevant and available experts, allowing the model developer to accurately know what is good and what is bad about the model. (Moody & Shanks, 2002).

3.6 Conceptual Framework

This section describes the conceptual framework on which the rest of the thesis is built on. The conceptual framework, which is shown in Figure 2 below, draws ideas from the findings described in earlier sub-sections.

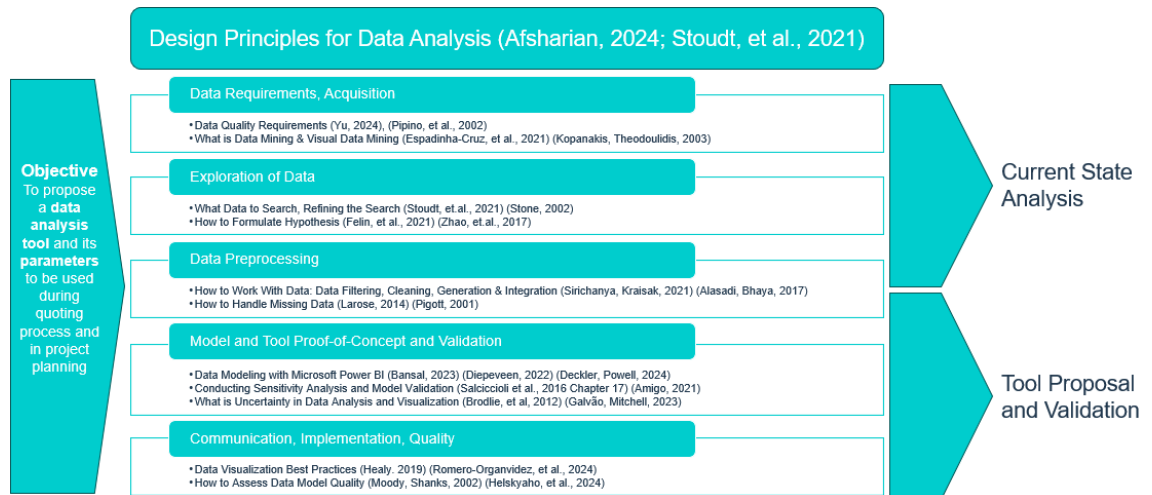


Figure 2. Conceptual Framework for Design Principles for Data Analysis

In section 4, data exploration is conducted by the researcher of this thesis using data mining techniques discussed in section 3.1.2. The researcher has extensive domain knowledge, which allows the creation and evaluation of multiple hypotheses during the search, following notes described in section 3.2.2.

Following best practices from section 3.1, available data is searched, filtered and its quality is assessed. The quality assessment on filtered data will include subjective surveys and objective assessments. Subjective surveys rank the data on a scale of 0-5 describing data quality in a scale from “bad” to “good”. Same scale is used in objective assessment. Results are considered low when the value is three or below. The data qualities are presented in a matrix described in Table 3. Data quality assessment is crucial part of the thesis since Mapvision has recently conducted data migration from old ERP to new one, and sizeable portion of data is inputted manually by employees.

As pointed out in section 3.3, data filtering is done during current state analysis, but also during solution development phase do avoid creating results that are overwhelming and require filtering on a larger scale. Data cleaning could be seen as a part of data quality analysis, aiming to improve either the overall quality of the data, or the quality of selected dataset. Best practice is always to correct the data at its source, to make it available to be used in next analysis.

Ignoring incomplete data can be dangerous, since there might be a pattern in the data — a reason why certain values are incomplete, as discussed in section 3.1.2. Decision to impute values must be considered carefully by examining the root cause of the missing values. A value in Mapvision’s case might be missing for a logical reason: a cost of a component might be missing since the invoice transaction has not been recorded in the system yet.

Data integration, as discussed in section 3.3.1, is used to combine all required data sources into one semantic model in Microsoft Power BI according to findings in section 3.4.1, since all the required data is not available in the same Business Central table, or even in the same data location. The report is distributed into multiple pages, and the underlying data is always accessible in the report, on a separate page, to let the user scrutinize the results given by the tool.

Sensitivity and validation of the tool, described in section 3.4.2, can in this case be done with historical data, i.e. modeling already delivered projects in the tool, and correlating the output to these already delivered and realized projects. Uncertainty level will be included in the results. The best way to visualize the uncertainty, will be further discussed in section 5.3.

The results from the tool are visualized according to best practices from section 3.5.1, and colours are selected to be best suited for the visualization from Mapvision theme. Data-Ink ratio is maximised, and descriptions are added only when Power BI tooltips are found not be enough.

Data model development will follow best practices from section 3.5.2, and its quality is verified by other Mapvision domain experts in section 6.2. Documentation of the methods employed will be done directly into the data model, using commenting and read-me section which is hidden for the end-user.

This concludes the conceptual framework section. The next section focuses on current state analysis, using this section as guidance.

4 Current State Analysis of Mapvision Operations Data

This section reports the current state of Mapvision data, and it is divided in six sub-sections. First, the research data and its collection methods are described. Then Mapvision operations data sources are identified with stakeholders in interviews, and the filtered results from further data mining are presented, according to the conceptual framework. After that, the quality of filtered data is analysed. Then current data reporting and analysis tools are described, to verify if they can achieve the same results as the planned outcome of this thesis. Current State Analysis findings are then summarized, after which Mapvision Data Culture is discussed in the last sub-section.

4.1 Overview of This Data Stage

Research data was collected in structured interviews held at Mapvision headquarters in Helsinki during February 2025 and via datamining by the researcher. The interviewees have broad understanding of Mapvision data, and each hold a position in Mapvision management team. The interviewees and their area of expertise related to this thesis are listed below in Table 4.

Table 4. Current state analysis interviewees

Title	Code	Expertise in area (related to this thesis)
Head of Production Operations	A	ERP Routings, Bill of Materials, Sourcing
Chief Finance Officer	B	Power BI reports, ERP
Chief Business Officer	C	Pricing strategy, customer requirements, CRM data
Chief Operating Officer	D	Resource utilization, Customers, Work hours, Power BI reports
Head of Project Management	E	Project structures, Costs, ERP Routings, Work hours
Chief Product Officer	F	System configurations, Quotes

All the interviews were recorded, and they were conducted using an interview template. The questions of the template are shown in Figure 3 below.

<p>Question 1a What <u>operations</u> related data MV currently has / is collecting (if multiple, list each)</p> <p>Question 1b What is the quality of the data in scale of 0-5 (if multiple datasets, grade each separately) 0: Provides no value and cannot be used 5: Is high quality, available, understandable and is / can be used</p>
<p>Question 2a Where is the data stored?</p>
<p>Question 2b Describe the accessibility of the data storage locations</p>
<p>Question 3a How is the data collected (if multiple methods, describe each)</p> <p>Question 3b What is the quality of the collection method in scale of 0-5 (if multiple, grade each separately)</p>
<p>Question 4a How MV currently uses its operations data</p> <p>Question 4b Grade the usage of the data in grade 0-5 (if multiple uses, grade each separately) 0: Use does not provide any value, or provides negative value 5: Use provides significant value for decision making and/or forecasting</p>
<p>Question 5a Describe the company data culture (hint: strong data driven culture: all decision makers are <u>required</u> to focus on the information provided by data; no data culture: decisions are made based on experience and or gut-feeling)</p>
<p>Question 5b Who are the primary users of data, if any</p>

Figure 3. Current State Analysis Data 1 interview form questions

The interview questions were selected to get a broad understanding of the current state, and they were open in nature, allowing discussion between the interviewee and interviewer. The interviewees were asked to grade data quality, data collection method and the usage of data in grade 0-5. Data quality grade is used in section 4.3 together with objective quality assessment to conduct a data quality assessment analysis and grades from data collection method and usage

of data are used together with questions 5a and 5b to discuss the company data culture in section 4.6. Full interview notes can be found as Appendix 1.

4.2 Operations Data Sources

In question 1a, interviewees listed total of 71 operations related data sources, out of which 48 are unique. The data was identified to be stored in HubSpot CRM, Teamcenter PDM, Business Central ERP, PowerApps Dataverse, SharePoint, Cloud and emails, Excel files or Power BI reports. Unique data sources are categorized per location in Figure 4 below.

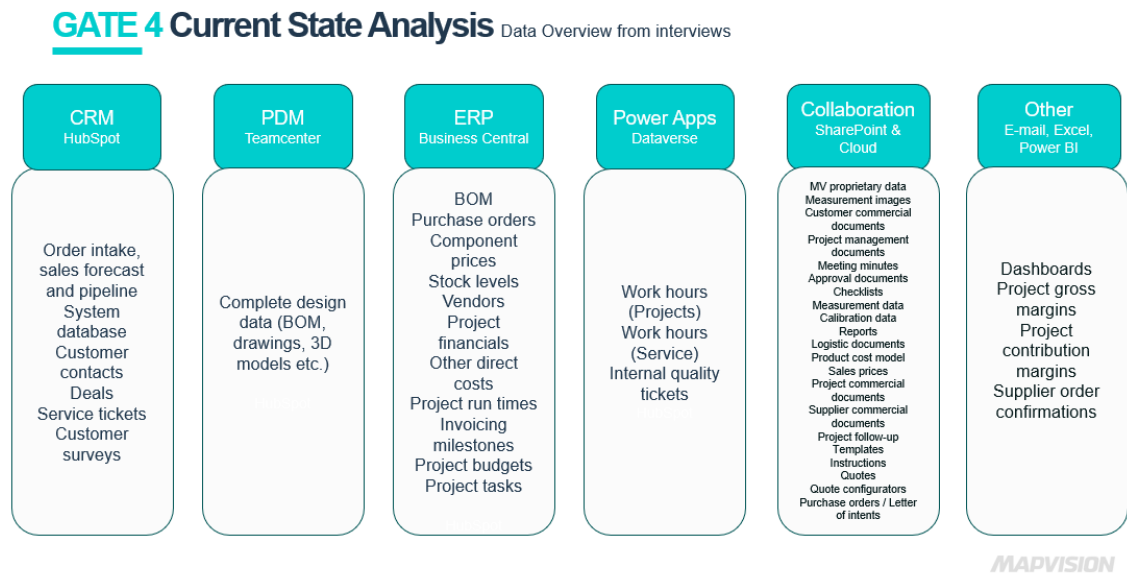


Figure 4. Mapvision operations data overview

The interviewees were not asked to filter or limit their data listings to the scope of this thesis during the interviews, since it could have left significant data sources out of the overview due to personal thoughts or opinions. As can be seen from the categorization in Figure 4, a significant portion of the data is stored in ERP and in SharePoint or Cloud.

The overall data scope was mined and filtered by the researcher according to best practices discussed in conceptual framework in section 3.6. The filtered data sources, which are relevant for this thesis, are shown in Figure 5 below.

GATE 4 Current State Analysis Explored and filtered for this thesis

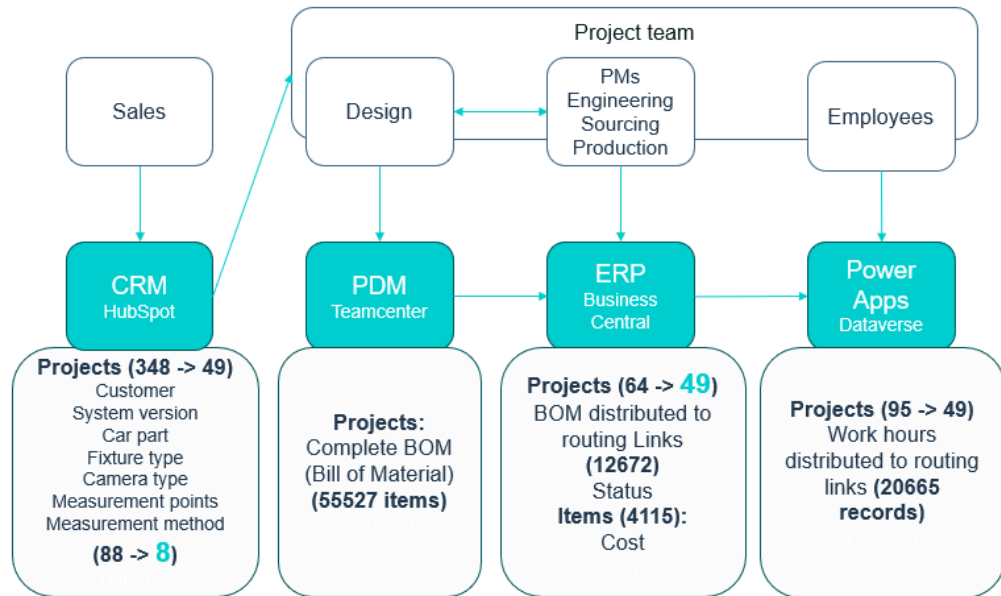


Figure 5. Filtered data sources and their relationships

As demonstrated in the Figure 5, the operations data generation starts from Sales organization. In the interview, informant C described the CRM data collection methods as follows:

“CRM data is logged manually into the system, and the system guides the user by offering attributes from drop down menus and the system has preconfigured data fields in it. The original logging is done by Sales, and it is then updated during the project by a Project Manager and during the operational lifecycle by a Service Manager or Service Engineer.”

Significant information stored in CRM for this thesis are customer, system version, car part, fixture type, camera type and part type, measurement points and measurement method (relative or absolute). CRM holds a total of 348 project records, each having 88 different columns of data. It was identified during data mining that the CRM system cannot be queried by Power BI, but a mirrored table of selected system information exists in Dataverse, which is updated periodically by manually inputting the data. Dataverse can be queried with Power BI.

Project teams use CRM, SharePoint and Cloud data during projects, and project related data is generated or collected into PDM, ERP and Dataverse. PDM data is significant for this thesis, but PDM cannot be queried by Power BI. However, actual project BOMs (Bills of Materials) are mirrored from PDM to ERP during project sourcing phases via excel import, and the full BOMs and component costs can then be queried with Power BI directly from ERP. The Project BOMs are structured per Routing Links in ERP, which means a project total and substructure cost can be data mined and visualized as illustrated in Figure 6.

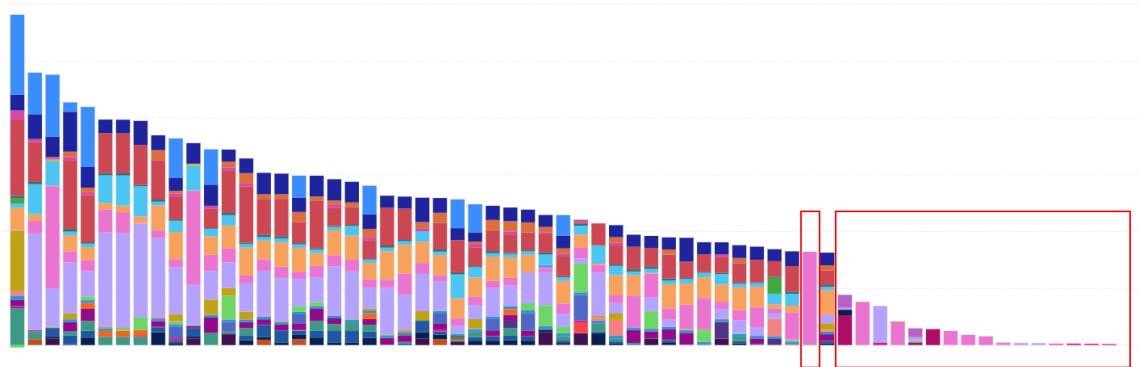


Figure 6. Project costs categorized per routing links during data mining

The highlighted projects were inspected in more detail, and it was identified that the ERP has records of 64 projects, out of which 49 have complete BOM, or the projects are being currently built. The complete or currently being built project amount is only 49 due to the recent ERP upgrade, where older records were not carried over from the old system. These 49 projects in current ERP and all new projects will be used as main filters for other data sources in this thesis. The filtered project cost breakdown is shown in Figure 7.

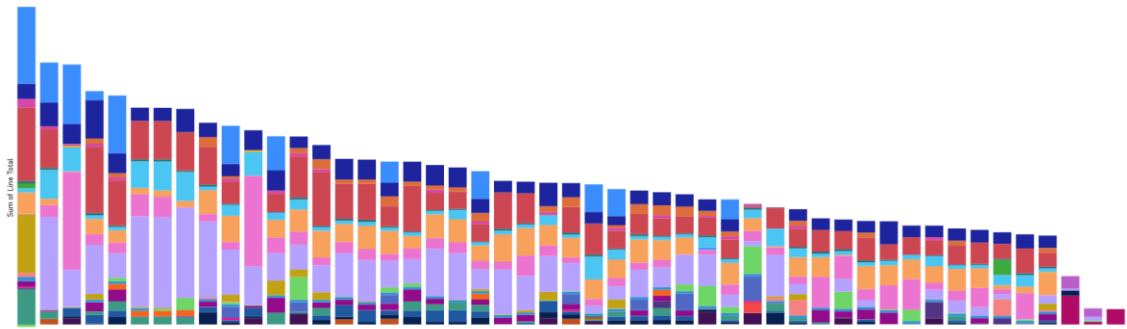


Figure 7. Filtered project cost overview

As seen from Figure 7, the rightmost projects have received some hardware for them, but are not complete yet, and therefore their cost in the system is not yet comparable to others.

Work hours are logged by each operations employee into Dataverse via a Mapvision designed logging tool using Microsoft PowerApp application. The app fetches projects and their routing links from ERP, and users input their hours into correct routing links after each day. Informant D, who oversees the hour logging, described the collection method as follows:

“The hour logging has taken significant leaps forward in its coverage and correctness. There are weekly and monthly checks in place to verify that all the logs are correct. It has been identified that a partition of employees tend to log their hours only end of month, even though the guideline instructs to log hours daily. Project hour logging is on a good overall level, but service-related hour logging needs more granularity in it, and it is being actively developed”

Dataverse has 95 projects and 20655 log records available, but all the records are not linked to projects since other activities are also logged into the system. The hours logged to projects are shown in data mining visual in Figure 8. Figure 9 illustrates the hour data when only the 49 projects from ERP and all new projects are shown.

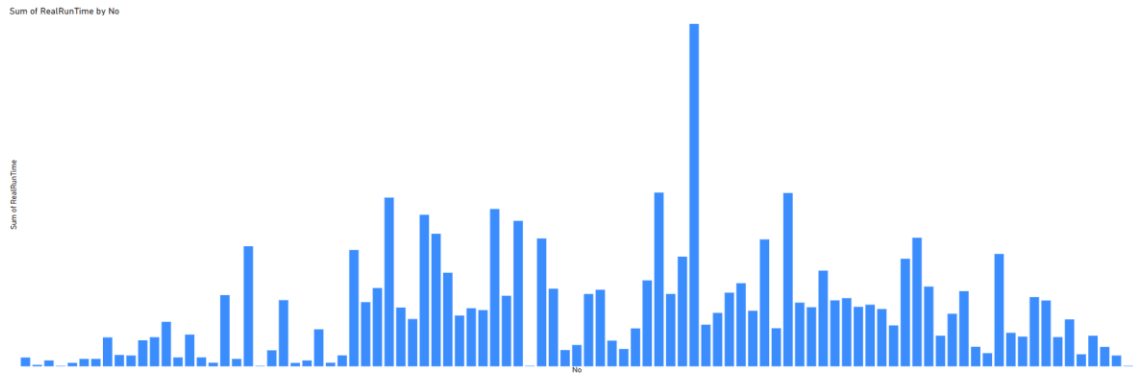


Figure 8. Hours logged for projects

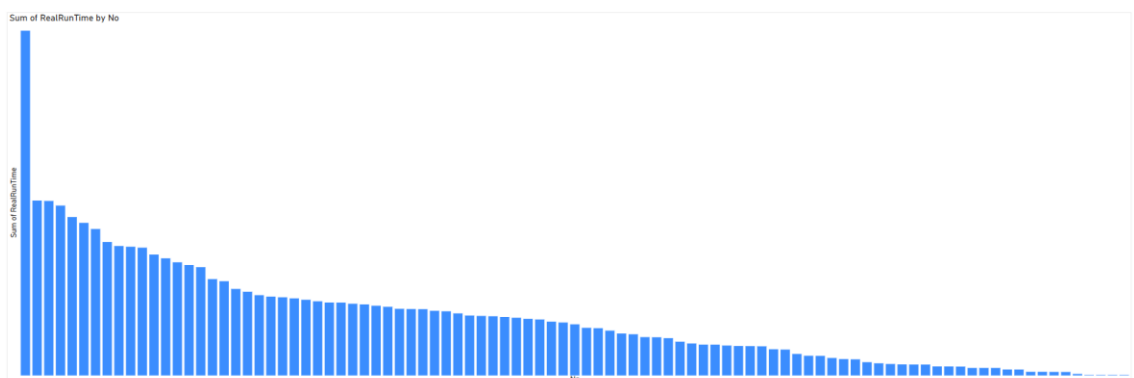


Figure 9. Hour data filtered by the 49 projects in ERP and all new projects created after that

The leftmost project was inspected in detail to verify that it is showing the records correctly, which it does. The rightmost projects have just been started, and it is normal they have work hours but no hardware records in them. A Power BI weakness was discovered during work hour data mining: the visual cannot be categorized in stacked columns, because the hours are distributed into too many routing links.

As a conclusion, it can be said that Mapvision currently has sufficient data for the outcome of this thesis to be partially tested and validated. The data amount will increase over time, and the tool will automatically include the new data in its queries. Section 6 will describe how the tool is developed to be future proof so when new, currently unused data categories emerge, they will work automatically. The filtered data quality will be analysed in the next subsection.

4.3 Data Quality Analysis

This section analyses the data quality of the filtered data shown in Figure 5. Data quality assessment is done by using a quality assessment matrix described in section 3.1.1. The matrix grading has been selected according to the needs of the outcome of this thesis, and it is shown below in Figure 10.

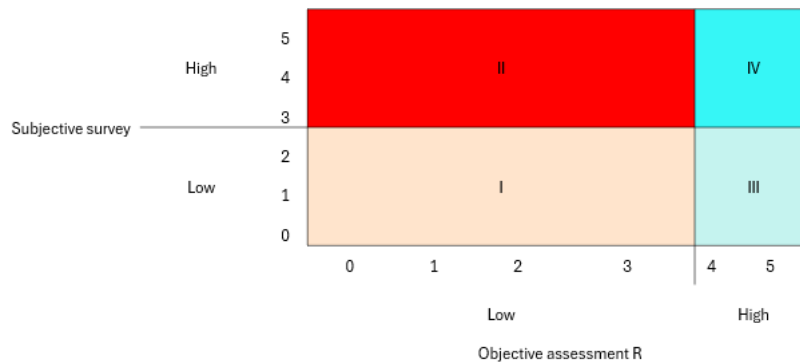


Figure 10. Data quality assessment matrix

Simple ratio R for objective assessment is calculated with the following formula:

$$R = 5 \times \frac{Od}{Ot},$$

Where R is simple ratio, Od is desired outcomes, meaning the quantity of outcomes that satisfy the criteria used and Ot is total outcomes, meaning the total quantity of evaluated records. Where possible, data is analysed using Free-of-Error simple ratio, completeness simple ratio and combination of both.

4.3.1 CRM Data Quality Assessment

Four out of six interviewees identified useful CRM data for this thesis, and three of them graded it with grade 3 each. Informant A did not grade the quality of the data, since he does not know enough about it. Therefore, the average Subjective survey grade for CRM data is 3.

Dataverse CRM mirror was found out to be out of date and therefore CRM data was analysed directly from CRM system via Excel export. Objectively, CRM data was analysed first by calculating Completeness simple ratio for each data column and then calculating Free-of-Error simple ratio for data which was available, meaning empty values were not considered. The calculations are shown in Figure 11 and Figure 12, rightmost bar demonstrates the full record ratio, considering all the columns at once.



Figure 11. Filtered CRM data Completeness Simple ratio

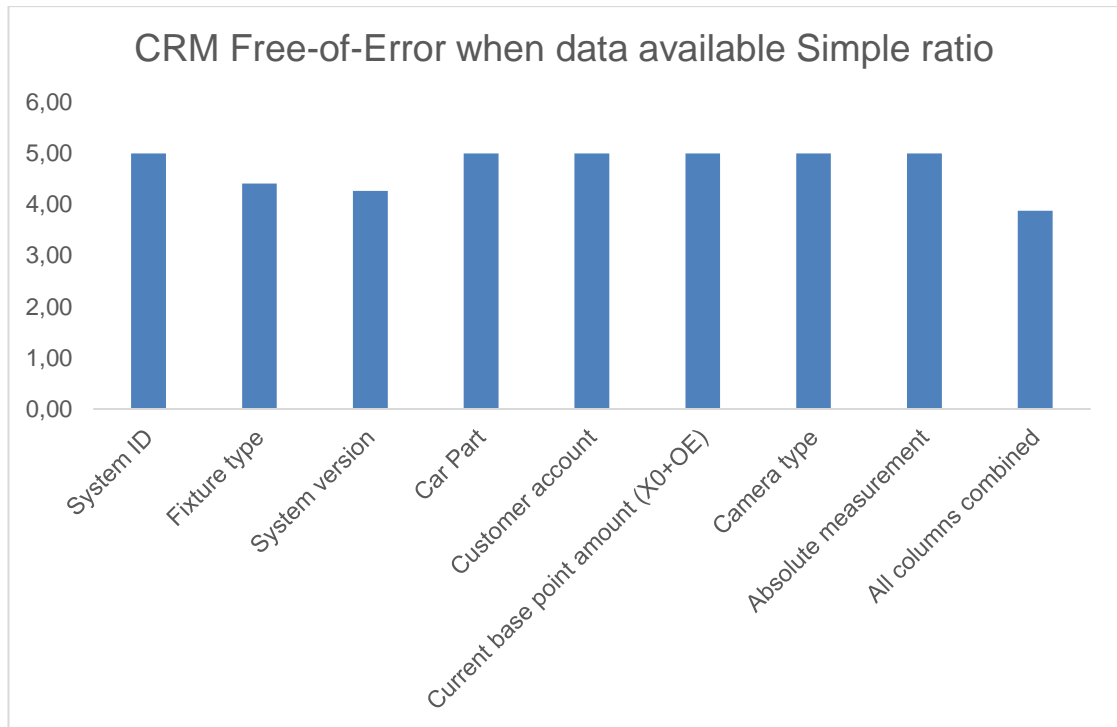


Figure 12. Filtered CRM data Free-of-Error Simple ratio

As can be seen in Figure 11, Current base point amount (X0+OE) column is missing the most values, followed by Camera type. Current base point amount (X0+OE) was identified to be a value that might be modified during the system lifetime after the project delivery. This modification will skew project phase analysis, and therefore it must be separated into two values; one for project phase amount, and amount which is used during the lifecycle of the system.

Figure 12 shows that when data is available, its Free-of-Error ratio is over 4, the lowest being System Version with ratio 4,27. When combining Completeness and Free-of-Error simple ratios, the overall ratio becomes 0,71, as shown in Figure 13.

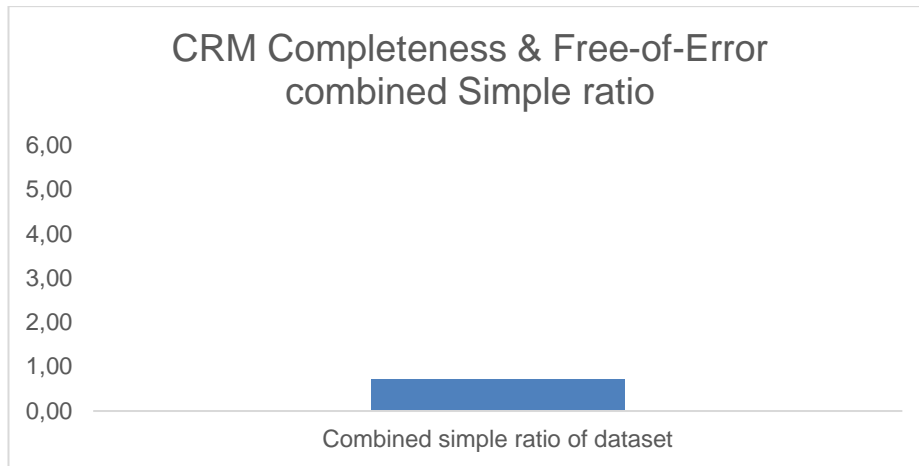


Figure 13. CRM combined Simple ratio

Ratio 0,71 means that only 7 projects of the total 49 observed have all 8 data columns in them and they are correct. When placing the average Subjective survey grade and objective assessment grade in the data quality assessment matrix shown in Figure 10, the conclusion is that CRM data quality falls in quadrant II.

4.3.2 PDM Data Quality Assessment

Half of the interviewees identified and graded PDM related data with grades 5, 5 and 4, resulting in an average grade of 4,67. PDM is the source for ERP BOM, and PDM will not be directly used in this thesis. Therefore, no objective assessment was conducted.

4.3.3 ERP Data Quality Assessment

Four interviewees identified and graded ERP BOM data with grades 4, 3, 5 and 4 and component cost data with grades 3,5, 4, 4 and 4. Therefore, the average ERP BOM grade is 4 and average cost data grade is 3,88.

Visual data mining methods, shown in Figure 14 and Figure 15, were used to find outliers in ERP BOM data. It was identified that project specific BOM is 100% correct, but the distribution inside project routing links are not. The lowest

identified ratios are for a machine enclosure routing link ENC_INS, which was found to have Free-of-Error simple ratio 3,54, and for one of the last routing links in project assembly, INT_PRE_AC, which has Free-of-Error simple ratio 3,45.

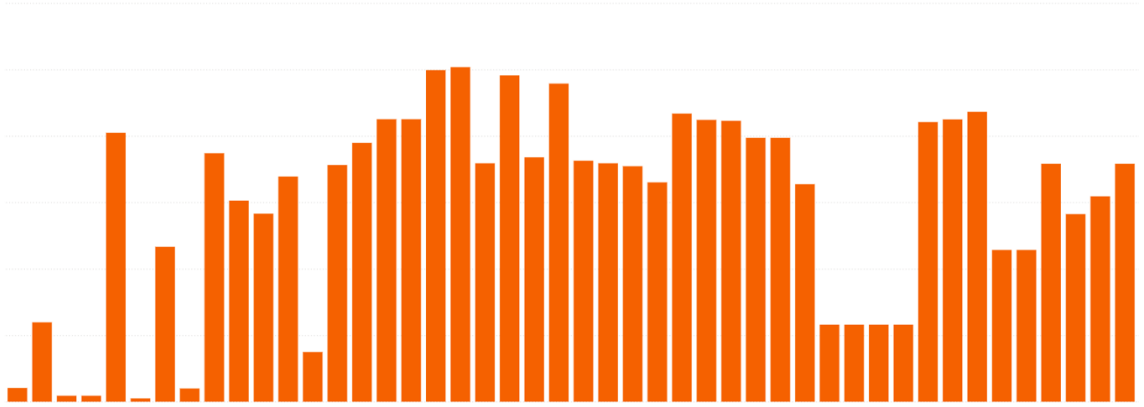


Figure 14. ENC_INS routing link cost

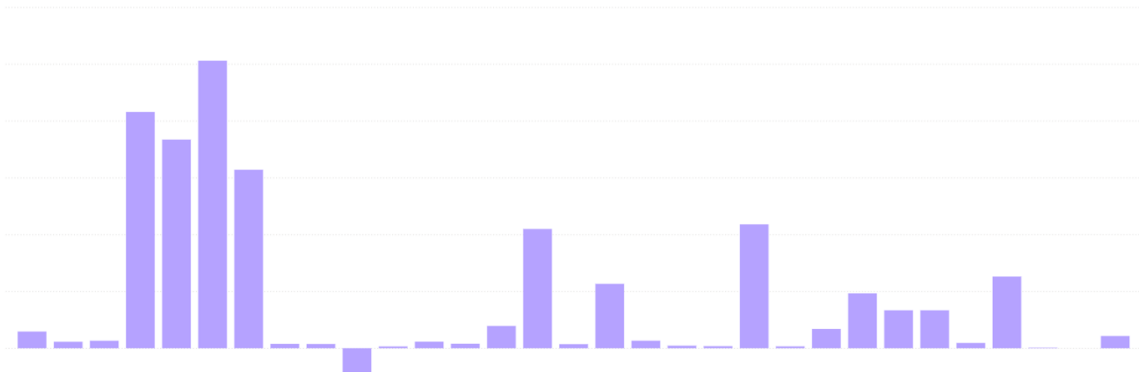


Figure 15. INT_PRE_AC routing link cost

All the routing links were inspected using a similar method as shown in Figure 14 and Figure 15, and a drill down method was used for the suspected outliers. Drilling down in data in this case means showing the underlying data of selected bar in tabular form. All the inspected routing links are listed below in Table 5. Routing links that are not listed in the table, are evaluated to be correct.

Table 5. ERP BOM Free-of-Error Simple ratios

CSWITCH_CONF: 4,9	ENC_INS: 3,54
INT_PRE_AC: 3,45	ARTEF_ASM: 4,58
FXT_ASM: 4,89	LONG_LIGHT: 4,89
CU_ASM: 4,13	MCK: 4,29
SHUTT_ASM: 4,66	SPAREBOX: 4,17
EC_INS: 4,02	Average of all inspected: 4,32

The simple ratios listed in Table 5 are for the data available in each routing link, and it does not consider the completeness of the routing links. This method is selected to avoid duplicating errors in the simple ratio; if a component which should be in FXT_ASM is currently in ENC_INS, it is reported only in ENC_INS simple ratio, not in both.

When placing the combined subjective survey grade and objective assessment grade in the data quality assessment matrix shown in Figure 10, the conclusion is that ERP BOM average data quality falls in quadrant IV.

Item costs completeness was analysed per project by sorting full BOM by item cost column. It was identified that some items have 0 cost, because a certain vendor occasionally combines its deliveries into one invoice line as shown in Table 6.

Table 6. Zero cost items

Description	Sum of Quantity_per	Sum of UnitCost From Items table	Sum of Line Total
Locating pin tower machined	1,00	0,00	0,00
locating pin tower 2 machined	1,00	0,00	0,00

Due to zero cost items shown in Table 6, the overall routing link cost remains correct, but single item cost can be substantially wrong when for example fixture frame component has 2 or more other assemblies combined into its item cost. When placing the average Subjective survey grade and objective assessment grade in the data quality assessment matrix shown in Figure 10, the conclusion is that ERP cost data quality falls in quadrant IV

4.3.4 Dataverse data quality assessment

This chapter concerns only Dataverse hour logs, since CRM system database - Dataverse mirror was found out to be out of date. The currently manual CRM mirroring should be automated to erase the possibility for input errors, and to keep the data up to date automatically.

Five out of six interviewees identified hour data in Dataverse and graded the quality with grades 4.5, 4, 4, 4 and 4. The average subjective survey grade is therefore 4,1.

Objective assessment Completeness simple ratio is 5, and this is due to the fact that the PowerApp -tool used to collect the data does not allow a log to be made if any value is left empty.

As discussed earlier, Power BI has a limitation on how many categories it can show in stacked bar charts, and currently hour logs are distributed into too many routing links. The overall project hours, shown in Figure 9, were therefore categorized per team and per project phase. Project work hour distribution per team is shown in Figure 16.

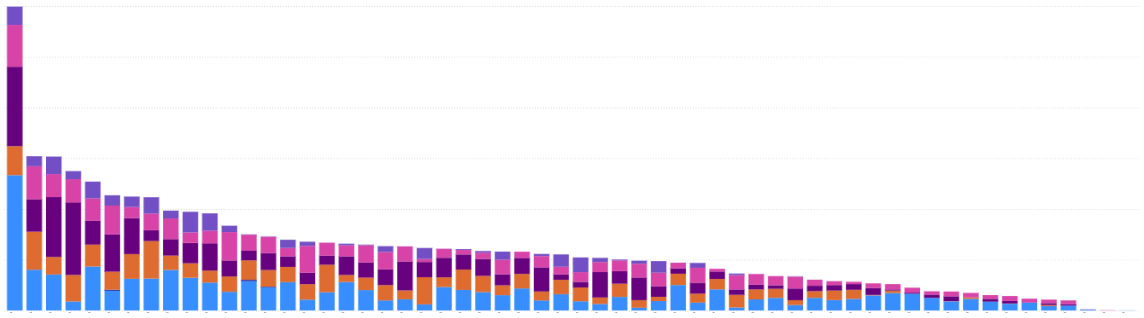


Figure 16. Project hour distribution per team

As seen in the Figure 16, the rightmost projects do not yet have contribution from as many teams as the leftmost projects. There were no outliers to be identified in this categorization. Project hours categorized per project phase is shown in Figure 17.

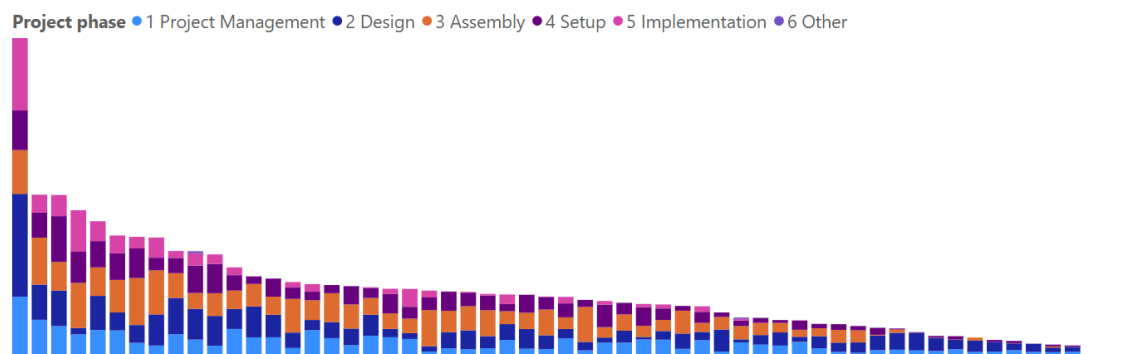


Figure 17. Project hours categorized per project phases: Project Management, Design, Assembly, Setup, Implementation and Other

As seen in Figure 17, each project, other than the rightmost ones which have just been started, have hours in project management and in design. There are no projects which are missing hours from any previous phase. The proportions of phase-hours are changing, but this is normal in project deliveries.

Some projects have suspiciously low Design phase hours compared to Assembly phase hours. After further analysis, it was noted that these systems are part of bundle delivery, and they are listed next to each other in Figure 18.

Project phase ● 1 Project Management ● 2 Design ● 3 Assembly ● 4 Setup ● 5 Implementation



Figure 18. Bundle delivery projects

The leftmost project in Figure 18 was designed first, and then the others were copied and modified from it. According to Mapvision hour logging guidelines, the hours of such bundle deliveries should be evenly distributed to every project. However, in these projects, it is not the case, as can be seen from the non-even Project Management and Design hours. The rightmost project have more Assembly hours than the rest, which partly might be explained due to it being assembled in a different Mapvision assembly plant.

Another similar issue can be identified with two other bundle projects, which are compared in Figure 19.

Project phase ● 1 Project Management ● 2 Design ● 3 Assembly ● 4 Setup ● 5 Implementation

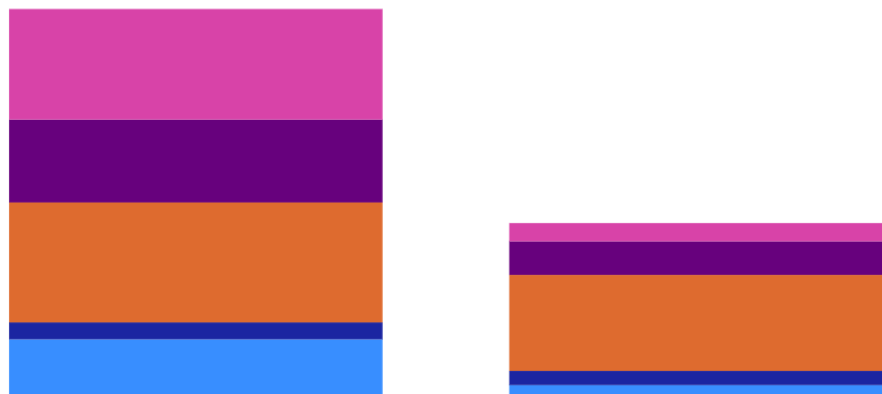


Figure 19. Two copy systems compared

As can be seen in the Figure 19, the left project has significantly more hours logged into it, even though the Project Management, Design, Assembly and Setup should be even.

Taken these six discoveries into account, the correctness simple ratio for Dataverse hours is 4,4. When placing average Subjective survey average grade and objective assessment grade in the data quality assessment matrix shown in Figure 10, the conclusion is that Dataverse hour quality falls in quadrant IV.

4.4 Mapvision Operations Data Current State Summary

A large portion of available data from ERP and Dataverse is sorted out based on available BOMs in ERP, but this is a must due to the time schedule and scope of this thesis. As discussed, the filtered data is rich enough for testing and partial validation of the thesis outcome data analysis tool. In Figure 20 below, data quality assessments from previous sections are placed in one matrix.

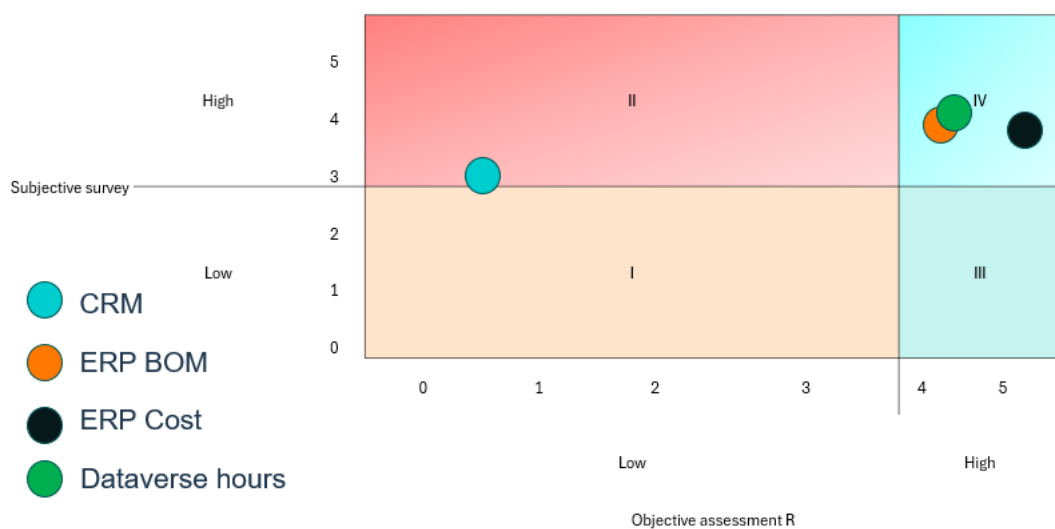


Figure 20. Data qualities in assessment matrix

As seen in Figure 20, CRM data cannot be used as is for this thesis and it must be corrected to improve the objective assessment grade. ERP BOM

assessment is acceptable for this thesis, but all the routing links listed in Table 5 should be corrected to increase the accuracy of the results produced by the outcome of this thesis, especially when working with major project specific routing links, such as FXT_ASM. ERP cost quality is good for this thesis, but in general it is advised to have a cost for each order line separately. Uneven hour reporting in Dataverse for duplicate projects will be shown in the outcome of this thesis as uncertainty when displaying the results, but the actual values will be averaged between all similar projects and does not require immediate correction.

CRM current base point amount (X0+OE) was identified to be a value that might be modified during the system lifetime after the project delivery. This modification will skew the project phase analysis which the outcome of this thesis is focusing on. Therefore, the current base point amount (X0+OE) record must be separated into two discreet records; one record which states the point amount after system delivery, and one record which is used during the lifecycle of the system.

4.5 Current Data Reporting and Analysis Tools

As stated by Stoudt, et al (2021) in section 3.2.1, one key goal during data exploration is to determine if existing tools or software can be utilized to answer the questions of interest. This section therefore describes the currently available data reporting and analysis tools and why they cannot solve the objective of this thesis.

The interviewees were asked to discuss how Mapvision uses its operations data in question 4a, and it was found out that a model exists which estimates the future workload divided per team. However, this model is static with preset parameters, and it does not consider the differences between different projects, and therefore it does not satisfy the business problem of this thesis.

An existing Power BI report is available for project reporting, but this report does not forecast future projects. The data model of this report will be reviewed

during the creation of the outcome of this thesis, and already developed solutions will be taken from it, if possible.

4.6 Data Culture

Interviewees were asked to describe data culture at Mapvision in question 5a to identify the readiness of the company to implement the outcome of this thesis into use. The data analysis tool structure and complexity are shaped according to the maturity of the company data culture, using best practices from section 3.4.1.

Overall, the interviewees state that the company is evolving towards data driven business, and data driven decision making is now possible due to recently upgraded ERP and the possibilities Power BI provides. Informant B stated:

“Data "drivenness" is increasing and employees are willing to lead with data, but company data quality and availability have only recently improved. Data-entry culture requires improvement - everyone does not have the visibility how entries effect downstream reports and operations. “

What informant B said about data-entry can be seen to be true especially in the CRM data quality.

Five out of six interviewees said everyone at Mapvision uses data in their work, and one stated management on all levels uses data. Everyone agreed that when available, data is used to support all decision making. It was noted that strategic business decisions do not necessarily rely only on data, but they require expert judgement.

As a conclusion, the data culture is in a level where the outcome of this thesis does not need to be limited to be a simple tool, but it can be extensive and provide all the available information for data driven decision making. This concludes the current state analysis, and the next section will focus on the solution proposal.

5 Data Analysis Tool Solution Proposal

This section is focusing on the analysis tool co-creation with key stakeholders, and it is divided into sub-sections. First the research data is presented, then the definition of the tool is described. After that, the decisions and methods to create the tool in Power BI are discussed. Lastly, a summary of the tool proposal is described.

5.1 Overview of This Data Stage

This section describes the research data collected during the tool proposal co-creation phase. The data was collected in meetings and via Teams chat in various phases of the development of the tool. The data collection points are listed below in Table 7.

Table 7. Co-creation key stakeholder meetings / chats

Informant	Code	Meeting / Chat Focus Area	Date
Head of Production Operations	A	Assembly hours, Design hours, Hardware cost, Tool parameter relationships	10.3.2025
Chief Business Officer	C	1st: Tool data preprocessing 2nd: Tool usage during quoting process	26.2.2025 20.3.2025
Head of Project Management	E	Project Management hours, Tool parameter relationships	10.3.2025 12.3.2025
Chief Product Officer	F	Tool usage during quoting process	20.3.2025
Head of Project Engineering	G	Project Engineering hours, Design hours, Tool parameter relationships	5.3.2025
Business Controller	H	Synchronizing hours categorization to existing Project Report	5.3.2025

The key stakeholders listed in Table 7 were selected to represent the user group of the tool from a wide perspective. Informants A, E and G are function heads, providing expert insight to relationships between CRM parameters and workhours and hardware costs. Informants C and F will use the tool during quote review process, and their viewpoints are considered during the tool

development. Informant H is a Power BI developer, who authors currently used Power BI reports and can provide expert technical support. All meeting minutes from this data stage can be found in Appendix 2.

5.2 Defining the Solution Proposal

This section describes the definition and expected functionality of the tool. The tool building in Power BI is discussed in section 5.3. The tool core functionality is to divide projects into smaller categorized subsets of hours and hardware costs, which are then used as building blocks for new projects. The proposed tool functionality follows currently used Excel based forecasting configurator. The current configurator is manually updated annually, and while it has accurate hardware module costs in it, the hours are only best estimates made by Project Manager and the budget approval group. The hardware modules do not consider project specific customizations, but they are instead representing standard solutions. The tool aims to provide more realistic view of future hours and costs, based on actual completed projects.

The data filtered from current state analysis was corrected manually, except the Base Point Amount (X0+OE) for 15 projects, because Project Managers were not able to get this information for all unfinished projects. The data quality used in this stage can be considered excellent for all other data than the Base Point Amount (X0+OE). The corrected and filtered data was presented to informant C, and he noted:

“The CRM data should be used from exported Excel file for the time being. The Excel can then later be used as a guideline to improve CRM data quality. The automatic link between CRM and Dataverse will be established later.”

The System Type column was decided to be separated into two, one listing the version of the system, and second listing the size of the system. This separation allows filtering between fully customized systems, and systems which have only customized size, but are based on standard system version. Informant C also noted that Base Point Amount (X0+OE) is a value that might be changed during

the lifetime of the system, and this would skew the results when it is used to filter project costs. It was decided that new column is created in CRM, one describing the project stage Base Points, and one which is updated during the lifecycle of the system. The rest of the CRM parameters are used as is, and their relationships to hours and hardware costs are discussed in sections 5.2.3 and 5.2.4.

The modified CRM parameters are System Type, System Size, Customer, Car Part, Fixture Type, Camera Type, Absolute Measurement, and Base Point Amount.

5.2.1 Project Hour Categorization

Same project hours categorization, which was used in the current state analysis, was proposed. This categorizes the hours into larger project phases, since working with all the available routing links would make the data analysis tool hard to comprehend. The proposed categorization is shown in Figure 17 in section 4.3.4.

Informant G noted that the hour categorization does not follow the current Project Reporting Power BI report, and we should aim to harmonize the categorization to allow clear comparison between project reporting and project planning. Informant H, who has developed the Project Reporting Power BI report, was inquired about the hour categorization logic, and eventually it was decided that the queries and DAX columns used to categorize the hours will be copied from the existing report to the analysis tool. The hours are categorized per role: Project Management, Project Engineering Setup, Design, Assembly and Logistics, and Implementation.

Informant A suggested dividing Assembly hours into two different categories; System Base Assembly hours and Fixture Assembly hours, to increase the granularity of Assembly hour data. This was agreed to be a good idea, and the change was implemented. Informant A also proposed separating Artefact Design hours from overall Design hours, but it was agreed that this change will

be carried out later, since currently Mapvision has very limited hour data for the relatively new ARTEF_DES -routing link, since it was introduced only recently. The logic how the hours are divided into agreed roles in Power BI, is described in section 5.3.

5.2.2 Hardware Cost Categorization

Hardware cost categorization follows the same categories used in current Excel based project planning configurator: Base System, Fixture, Cameras, Projectors, Artefact, Spare Part Box and Calibration Kit. All the stake holders agreed keeping this categorization as is. Informant A noted that smaller modules, such as extra PC, scanner, printer, temperature compensation kit, extra user interface, safety lock and freight type should not be used as filters to retain larger number of projects per selected filtering condition. It was agreed that the averaging done by the tool is a satisfactory result, and these smaller modules will be included in the Base System category.

5.2.3 Relationships Between CRM Parameters and Workhours

The expected relationships between project parameters and project roles were mapped with stakeholders. Informant E noted that Customer Account has largest effect on Project Management hours, and she has a feeling Absolute Measurement also influences Project Management hours, and that it increases Project Engineering and Implementation hours. Project Engineering and Implementation hours were discussed with informant G, and he noted that Fixture type and Camera type does not affect Project Engineering or Implementation hours, and Car Part, Absolute Measurement and Base Point Amount influence both Project Engineering and Implementation hours, while System Type influences Implementation hours only. He noted that system type influences Implementation hours because more complex systems require more on-site installation hours.

Informant A suggests that Fixture type and System type has the biggest effect in Assembly hours. He wonders if Base Point Amount could have some effect in

the assembly hours, due to it being indirectly linked to camera quantity, which would require more assembly. This matter was discussed with informant G, and he invalidated this possible relationship, arguing that Camera Amount is purely linked to the locations of the Base Points, instead of the quantity of the Base Points. This relationship could not be fully tested, since several projects are still missing Base Point values in CRM.

Informant A listed that System Type, Car Part, Fixture type and Absolute Measurement influence Design hours. Informant G also noted that Absolute Measurement influences Design hours, because Absolute Measurement requires a design for custom Artefact.

5.2.4 Relationships Between CRM Parameters and Hardware Costs

It is self-evident that Base System category is affected only by System Type and System Size parameters and Fixture category is affected only by Fixture Type parameter.

Artefact, however, is affected by Car Part and Absolute Measurement. Absolute Measurement effectively enables or disables this category, since Artefact is included only in systems which measure independently, without a reference to any other measurement system. Fixture Type could be added as a slicer, since same car parts have multiple different Fixture Types for them, and Fixture Type affects the complexity of the Artefact in certain cases. However, considering the amount of data at the tool disposal, Fixture Type is not added as a slicer for Artefact currently.

Cameras and Projectors are both affected by same slicers, since it was noted by informant G that 'Surface Feature Count' is missing in CRM as a parameter, and therefore, no proper way of filtering Projectors exists now. The parameters affecting Cameras and Projectors are Car Part, Fixture Type, Camera Type and Absolute Measurement. Car Part and Fixture Type categorizes the baseline camera amount, while Absolute Measurement usually adds a certain % of

cameras on top of this, and Camera Type switches the camera to different type, which has a different cost per camera.

5.2.5 Data Pre-processing and Filtering

To allow more data points for the tool, endpoints for project hour logging roles were identified with the stakeholders. This means that certain roles from projects could be considered in the tool, even if the project has not been closed yet. The project milestones, after no more hours are logged for each role are Design – Ready to Ship, Assembly and Logistics – Ready to Ship, Project Engineering Setup – Installed and calibrated, Implementation – Handover to Care and Project Management – Handover to Care. Handover to Care is the final milestone after project closing, so this means that Design, Assembly and Logistics and Project Engineering Setup hours can be considered in the tool long before the project has been completed, which gives more timely data for the tool's disposal.

The Hardware modules can be considered in the tool after Ready to Ship milestone, except for the Artefact, which is considered only after ARTEF_ASM routing link completion. It was noted by informant A, that currently, it is possible that all Artefact Assembly hours are not accrued yet when Ready to Ship milestone is completed, but informant A stated:

“Due to current supplier challenges we might receive Artefact components late, but we will continue as is. The supplier issues will be fixed sooner or later, and it does not make sense to build the tool around this temporary issue.”

Double Shuttle systems AS2 and AS2 MCK have two fixtures in them, but both fixture hours and costs are recorded for same routing link. Therefore, the data is pre-processed in case of AS2 or AS2 MCK projects by dividing FXT_ASM hardware costs by two and FXT_ASM hours by two. Design is not divided by two, since in most of the cases, the fixtures are copies of each other, resulting only in minimal overall hour increase.

5.2.6 Power BI Report Pages

Since the tool is created in Power BI, multiple report pages can be added. The following additions were requested by the stakeholders to supplement to the main analysis page:

1. Projects that are currently considered in the tool and their milestone statuses.
2. Self-analytics pages for Hardware Cost and Project Hours, where user could see the projects as a whole, without the tool categories being applied.
3. Self-analytics page where certain parameter and its correlation to projects is shown. This page contains business sensitive information, and it is not discussed further in this thesis.

5.2.7 Summary of The Tool Definition

Hour categorized into roles, milestones after they are considered and CRM parameters affecting them are shown below in Figure 21.

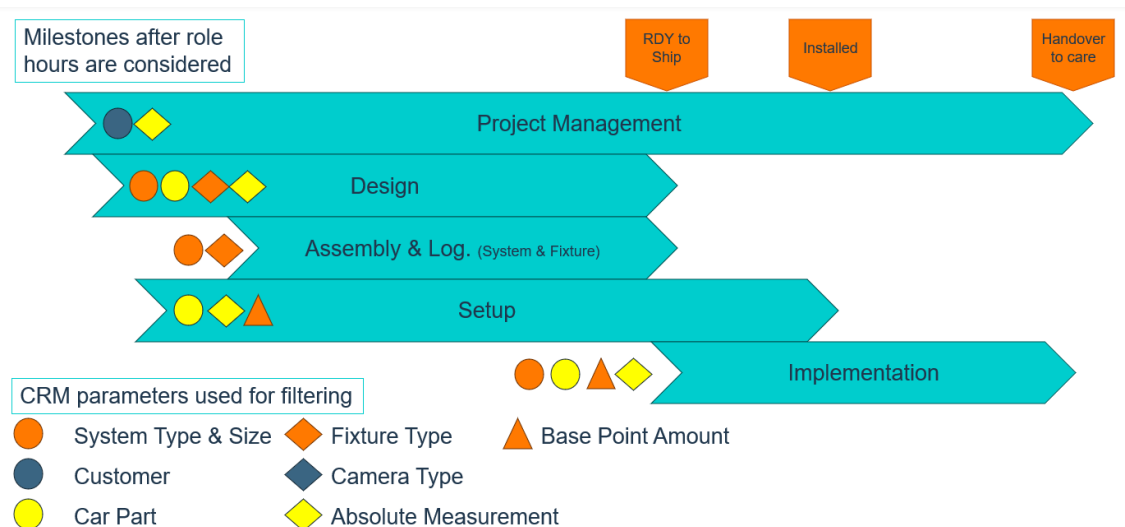


Figure 21. Hour definition for the tool

As can be seen in Figure 21, camera type does influence any hour role, and this is due to the reason that the functionality which different camera type enables, is not currently done during project phase, but only later after project closing.

Hardware cost categorized into modules, milestones after they are considered and CRM parameters affecting them are shown below in Figure 22.

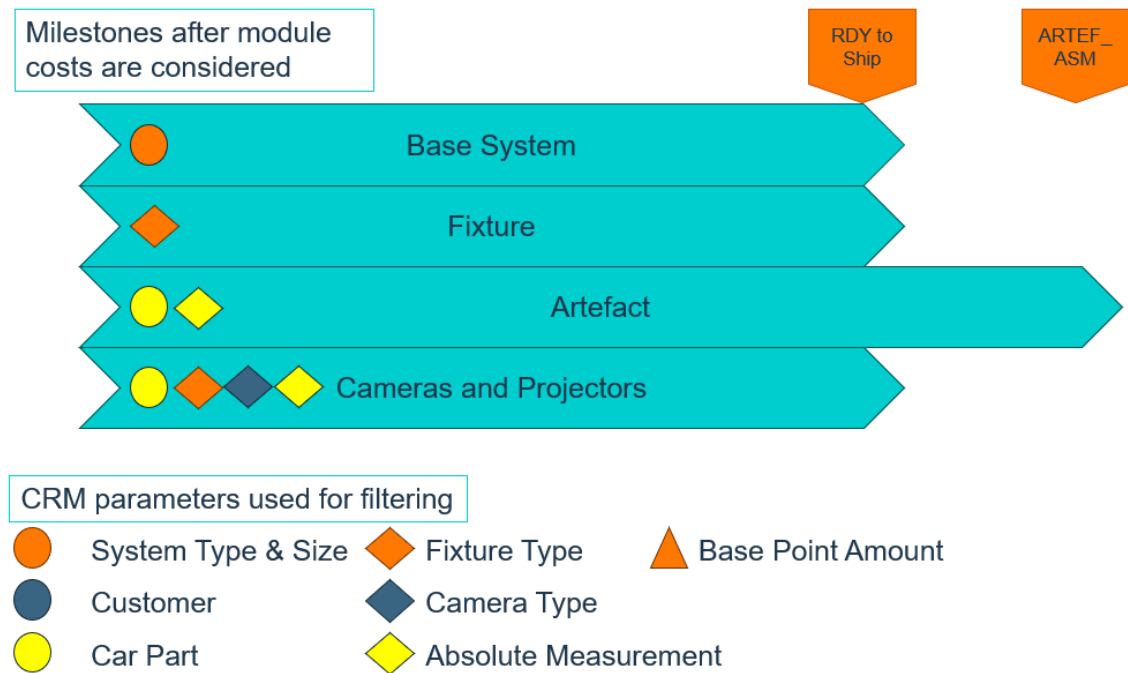


Figure 22. Hardware cost definition for the tool

As can be seen in Figure 22, Customer currently does not influence the hardware cost. While this might be not be completely correct in some special cases, it was decided that averaging the costs over multiple projects is satisfactory for the possible customer specific increases in hardware cost. As discussed earlier, Base Point Amount was not added as a relationship for Cameras, and therefore it is not affecting any Hardware Costs at this time.

All the hypotheses provided by the key stakeholders, in other words the proposed relationships, were tested by evaluating a change in the dataset when using proposed relationship as a filter and comparing it to unfiltered result. All the hypotheses were found out to be correct, meaning they have significant change

in the data when used as filters. Few outliers remain, especially in certain hour roles. This can be explained by the nature of project business, where in theory everything effects everything, and project scopes are different, or they change during the project. The uncertainty of the results will be discussed in section 5.3.4.

5.3 Creating the Solution Proposal

This section describes the tool building in Power BI software, and it focuses on the significant choices and methods used. The basic software functionalities and settings are not in the scope of this thesis and will not be discussed.

5.3.1 Data Model

Data model in Power BI refers to the relationships between different data tables. In this case, there is a need for two fact tables: work hours from 'sula_workcenterdataprod' table and component BOMs from 'Prod_Order_Components_Excel' table. These tables are then sliced and filtered with dimension tables, such as CRM parameters and milestone statuses from 'prod_order_routing_excel' table. Initial data model, which was tested in the beginning of tool creation is shown in Figure 23.

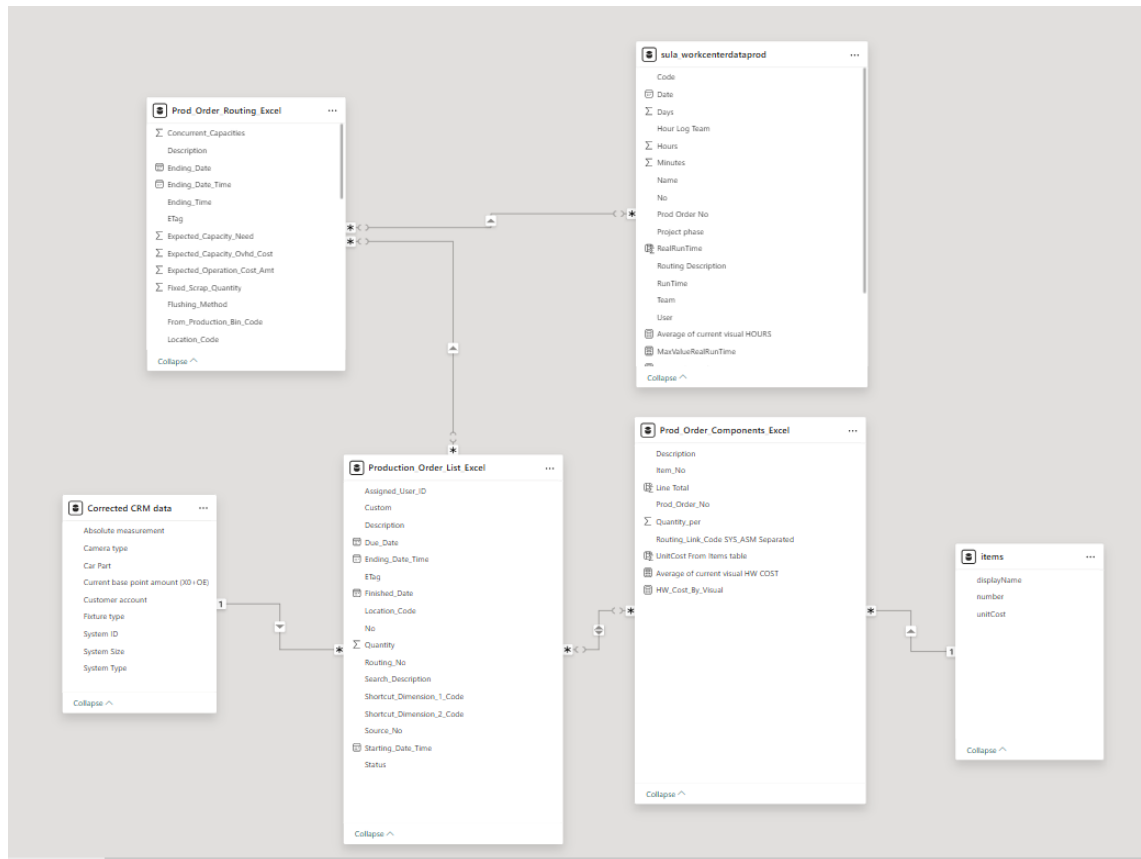


Figure 23. First draft of data model

The model shown in Figure 23 would work for simple visual slicing, but one major problem was identified during testing, and another issue was found, which would emerge when advanced DAX measures would be used. The major problem is the use of many-to-many relationships between the tables. These relationships cause issues with filter propagation from table to table, and in this case, there was an issue when slicing work hours from table 'sula_workcenterdataprod' with dimensions from table 'Corrected CRM data'. The sliced result would show correctly with some combination of slicers, and then with another combination it would not work. Another issue is that 'prod_order_components_excel' fact table has a relationship to 'sula_workcenterdataprod' fact table via 'production_order_list_excel' and 'prod_order_routing_excel'. Linking two fact tables together will cause problems in combined measure calculations, where both fact tables are referenced, since they are sharing same dimensions.

The many-to-many relationship issues were fixed by completely re-evaluating the data model and by creating bridge tables. At the same time, Corrected CRM data was split into smaller dimension tables, each representing one dimension of a project. The final and working data model is shown in Figure 24.



Figure 24. Final data model

As can be seen in Figure 24, the 'production_order_list_excel' table was removed since it was not needed after all, and the two fact tables in the middle have only many-to-one relationships with the surrounding dimension tables, with cross filtering direction pointing to fact tables only, effectively disconnecting fact tables from each other. 'Prod_Order_routing_Excel' and 'User_Location (manually update)' tables do not have unique values, and therefore a bridge tables in between them and fact tables were created. The bridge tables hold only unique values, and both sides are connected with one-to-many relationships, effectively removing all filter propagation issues created by many-to-many relationships. 'Prod_Order_Components_Excel' table uses costs for components in Bill of Materials from 'Items' table. Cameras and Projectors were recently moved to a combination SYS_ASM routing link, but they are separated

back to their own, old routing links in the tool query to allow more detail data representation.

'Machine_Center_Card_Excel' is used in hour role categorization, which is described in detail in section 5.3.3. 'System ID' table at the bottom left is used to filter all available projects to a list which displays only the projects that are considered in this tool. This list is shown in one of the auxiliary pages of the report, as requested by stakeholders in section 5.2.6.

5.3.2 Analysis Page

This is the first and main page of the Power BI report, earlier referred to as the "tool". The tool has the previously mentioned CRM parameters at the top, where user configures the desired project configuration. These parameters then filter the hour and cost categories, which are illustrated in bar charts where X axis is a project ID and Y axis is either own cost in euros or hours. An average of the current visual is shown in the bar chart as a line, and each average value from the bar charts are listed in summary tables. The tool, as it is when opened by a user, is shown below in Figure 25.



Figure 25. Data analysis tool

As seen in Figure 25, relationships of each CRM parameter is described below the slicers, and these relationships are the same as shown in Figure 21 and Figure 22. A 'Project Filter slider' slicer has been added for future use, which can be used to filter older projects out from the tool.

The summary tables list the category average hours and costs, sums them and shows the percentage of total for each. The percentage of total is a good indicator for quote reviews, providing insightful information about the characteristics of the cost structure. The average values for the tables are calculated with DAX measures as shown in Figure 26 below.

```

1 Average Base System HW Cost =
2 CALCULATE(
3     AVERAGEX(
4         VALUES('Customer Account'[System ID]),
5         CALCULATE(SUM(Prod_Order_Components_Excel[Line Total]))
6     ),
7
8     // Hardcoded filters
9     'Customer Account'[System ID] <> BLANK(),
10    Prod_Order_Components_Excel[Routing_Link_Code SYS_ASM Separated] <> "ARTEF_ASM",
11    Prod_Order_Components_Excel[Routing_Link_Code SYS_ASM Separated] <> "CALI_P_ASM",
12    Prod_Order_Components_Excel[Routing_Link_Code SYS_ASM Separated] <> "CAM_INS",
13    Prod_Order_Components_Excel[Routing_Link_Code SYS_ASM Separated] <> "FXT_ASM",
14    Prod_Order_Components_Excel[Routing_Link_Code SYS_ASM Separated] <> "PROJ_ASM",
15    Prod_Order_Components_Excel[Routing_Link_Code SYS_ASM Separated] <> "SPAREBOX",
16
17    // Remove filters on specific Tables from visual. When commented, the line slicer affects this measure:
18    // REMOVEFILTERS('System Type'[System Type]),
19    // REMOVEFILTERS('System Size'[System Size]),
20    REMOVEFILTERS('Customer Account'[Customer account]),
21    REMOVEFILTERS('Car Part'[Car Part]),
22    REMOVEFILTERS('Fixture Type'[Fixture type]),
23    REMOVEFILTERS('Camera Type'[Camera type]),
24    REMOVEFILTERS('Absolute Measurement'[Absolute measurement]),
25    REMOVEFILTERS('Base Points'[Current base point amount (X0+OE)]),
26
27    // Set Project Stage
28    FILTER(
29        ALL(Prod_Order_Routing_Excel),
30        Prod_Order_Routing_Excel[Routing_Link_Code] = "READYTOSH" &&
31        Prod_Order_Routing_Excel[Routing_Status] = "FINISHED"
32    )
33 )

```

Figure 26. Average Base System HW Cost DAX measure

As seen from the Figure 26, the DAX code is well commented and easy to follow, and similar coding style is used for all the average hardware cost values, average hour average values and elsewhere where required. In total, the report has 19 measures, 4 calculated tables and 10 calculated columns.

A “notes for use” text box provides basic information of the report, such as if a filtering selection provides no projects for certain bar chart, this is because currently there are no projects that satisfy the milestone filtering criteria. The projects which are considered, and their milestone statuses, can be seen on 2nd page of the Power BI report. Top right shows the current version of the tool, and in this point of time the version is 20-3-2025 Validation Release.

5.3.3 Hour Role Categorization Function

As noted in section 5.2.1, the hours are divided into roles in Power BI with a DAX function in a column. The column is the same as in Project Reporting Power BI, and the function is 73 lines of DAX code. The function first checks if a routing line type in Business Central is 'Work Center', representing a team, or 'Machine Center', representing project member. If a type is 'Work Center', the function categorizes the hours into Assembly & Logistics when the Work Center is equal to "Prod", into Site Inst. & Impl when the Work Center is equal to "SITE IMP" and into Design when the Work Center is equal to "EPLAN". When the type is 'Machine Center', the function checks the allocated project member default role from 'User_Location (manually update)' table and categorizes the hours into this role. 'User_Location (manually update)' table holds the current roles of operations members, and it has a 'valid from' column, which is used when an employee changes roles.

5.3.4 Uncertainty

The bar charts are shown in the report, instead of just displaying average values in the tables to visualize the current uncertainty of the tool output. The 'Notes For Use' text box has a note that clear outliers can be manually removed from consideration by excluding them. It also states that if a selection results in less than 3 projects, the average can be considered highly uncertain for future projects. Removing suspected outliers, however, can cause problems, since all projects are different, and it is possible the suspected outlier would represent the actual future project that is being simulated in the tool. Adding a normal distribution curve to increase the uncertainty understanding, as shown in Figure 27, was tested.

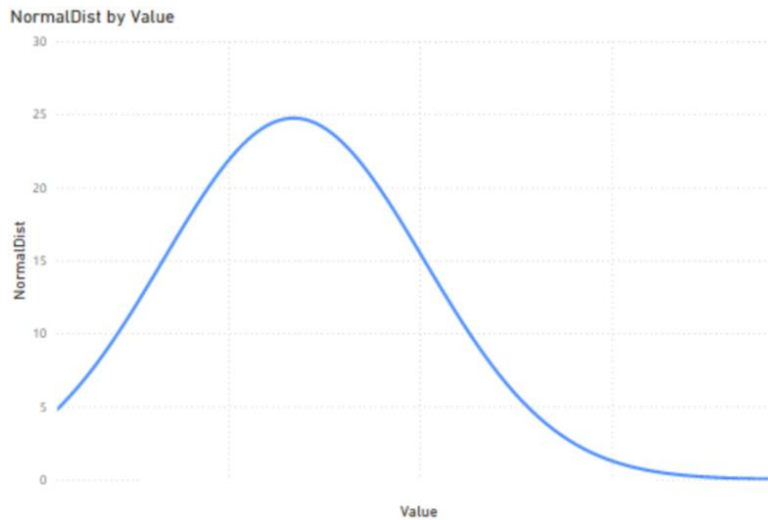


Figure 27. A sample category normal distribution

It was found out that adding such curve for all the categories would harm the readability of the report and provide minimal value when compared to bar charts only, and therefore the idea was discarded.

In the end, the uncertainty evaluation is left to be the end user's responsibility, since the tool aims to aid in decision-making by visualizing data, not to provide 100% accurate future projections.

5.3.5 Model Sensitivity Analysis

The sensitivity of the tool was brought up by informants A and E, when they mentioned new working methods which would reduce hours, or new standard components or changed suppliers, which would affect the own costs of hardware costs would not be reflected in the tool immediately. It is true that there is zero linking from new efficient working methods for hours, but hardware costs are linked to the tool, when same component numbers are used. The tool uses the current cost from 'Items' table, and when a cost is updated into the table in Business Central, it is reflected in the tool. The costing method used in Business Central is FIFO (First In First Out), which means the current Unit Cost is based on the current stock level and its cost. When standard hardware structures change because of new Hardware Releases, these new items and

their costs are not reflected in the history data, which this tool uses, immediately. The 'Project Filter' slicer was added to filter older projects out from the tool in the future for this reason.

5.3.6 Self-Service Analytics Pages

Overview of all project costs and hours can be analysed on separate report pages by the user, using his or her own filtering. These pages are made available for self-service analytics purposes, and the data can then be exported to Excel for further processing or used in the Power BI report directly. These report pages use stacked bar chart visuals without predetermined filtering, as shown in Figure 28 and Figure 29.

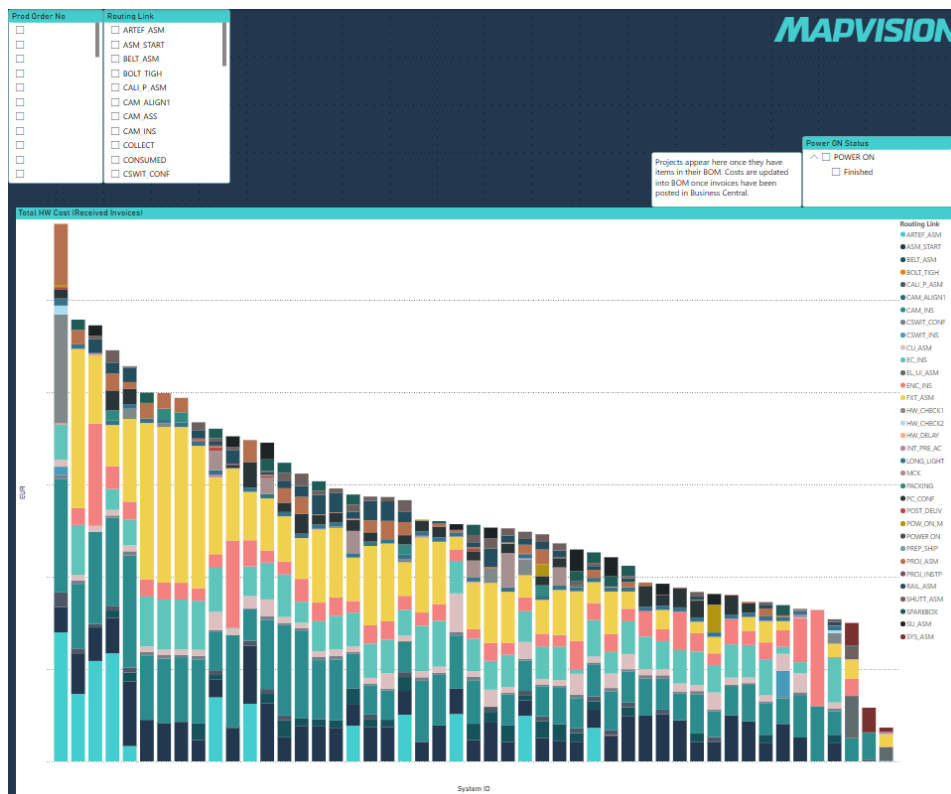


Figure 28. Self-service analytics page for hardware cost

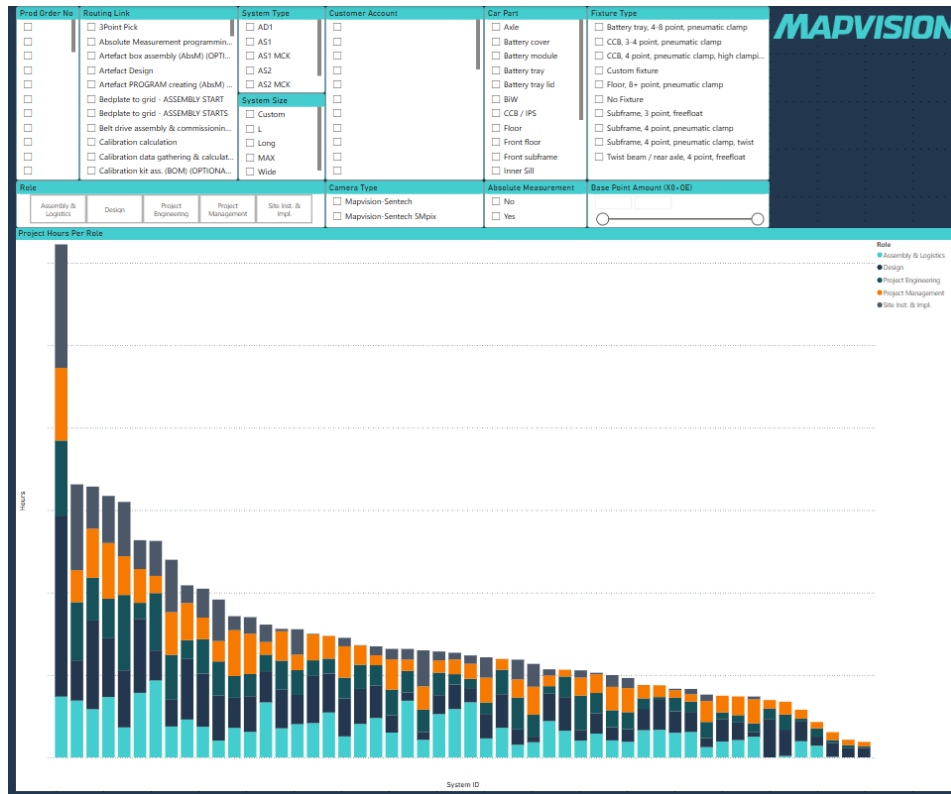


Figure 29. Self-service analytics page for hours

As seen in Figure 28 and Figure 29, no average values are shown. This is because there is no milestone filtering present in these pages, so the averages would not represent any actual values, since projects in different stages would be considered. The data bars can be drilled down by the user, presenting the underlying data in a tabular form.

5.4 Summary of the Solution Proposal

The tool is comprehensive for its intended purposes, and all end users were represented in the creation of the tool. Therefore, it should provide value to all its users by clearly presenting underlying business data in a categorized manner by using relationships, which could remain unseen without the tool. The Power BI report with all its pages were published in Power BI Service and its refresh rate was set to be once a day. The report was shared with a testing group, which consists of 11 employees, and they will validate the tool for two

weeks, and the validation results and recommendations are discussed in the next section.

6 Validation and Improvement of The Solution Proposal

This section focuses on validating and improving the co-created solution described in previous section. This section is divided into three sub-sections; first the research data collection methods are described, then the received feedback and implemented improvements to the tool proposal are presented, and lastly the final, improved data analysis tool is described as the outcome of this thesis.

6.1 Overview of This Data Stage

Feedback to the solution proposal presented in the previous section was collected via e-mail after two weeks of testing by a testing group, which represents the end-users of the tool. An e-mail-based feedback method was selected as validation method because it allows gathering of multiple viewpoints in a relatively short time. In addition to email feedback, one meeting was held with informant H, a Mapvision Power BI developer, to evaluate the tool model quality. The testing group is shown below in Table 8.

Table 8. Data stage 3 testing group members

Informant	Code	User Role during Feedback	Date received
Head of Production Operations	A	Budget approval	3.4.2025
Chief Finance Officer	B	Financial forecasting	6.4.2025
Chief Business Officer	C	Quote approval	6.4.2025
Chief Operating Officer	D	Budget approval	4.4.2025
Head of Project Management	E	Budget approval	31.3.2025
Chief Product Officer	F	Quote approval	4.4.2025
Head of Project Engineering	G	Budget approval	4.4.2025
Business Controller	H	Power BI model expert	4.4.2025
Director, Core Applications	I	Cost and hour analyst	1.4.2025
Senior Project Manager	J	Budget planner	27.3.2025
Project Manager	K	Budget planner	3.4.2025

As can be seen in Table 8, the testing group has the same persons who were involved in the tool co-creation. In addition, relevant stakeholders who were not part of the co-creation are involved, to avoid the possibility of receiving only biased feedback from co-creators for their own suggestions. All the feedback received can be found as a summarized table in Appendix 3.

6.2 Feedback Received and Improvements to Solution Proposal

73% of respondents describe the tool as good or value-adding and none of the respondents describe the tool as bad or waste of time. Informants B, D and I all valued the data drill-down feature, where user can select a data bar and list the underlying data in a tabular form, showing detailed information. Informants C, F, J and K requested more data to be analysed or visualized, while three of them also stated the tool is great, which tells that the tool proposal has raised their interest in this particular data domain. In total, 38 improvement ideas were

proposed, out of which 24 were implemented, 7 were left to be implemented outside of this thesis and 7 were rejected.

6.2.1 Visual Improvements

Informants D and J had the first impression that the tool is overwhelming at the first glance, because it is showing so many different slicers, information boxes, data graphs and tables as shown in Figure 25. However, they noted that after reading the information boxes and using the slicers for a while, the tool functionality became clear. Informant A and H noted that to improve the clarity, the information boxes should be shown as tool tips instead, which pop up only when needed, and that the two tables should be separated from each other and placed next to the corresponding graph groups. Informant A found out that the table rows are sorted according to the values, and therefore they change places when new slicers are applied, making comparison more difficult. Informant H suggested changing the colour schema of slicers to distinguish them from the data charts, and to separate hour charts from hardware cost charts with a line. He also suggested trying different colours for hardware and hour visuals, and to see if it is possible to add total lines for each table. Informant I had an issue where the Project Slicer was out of the view when zoomed in, and the selection and the consequences were not recognized. Informant B recognized that when selecting a value from Customer Account or Part type slicers, and then scrolling down or up, the selection is not visible anymore, creating a possibility where the user does not recognize that a selection has been made. Informant E noted that the relationships between slicers and the data is not fully clear, and that a “Clear all slicers” button should be added. Informant J proposed reorganizing the CRM parameters to be in order from System Type, System Size, Car Part, Fixture Type and Customer Account, to make the configuration more logical. Lastly, informant C suggested adding a visual showing the distribution of average hours and hardware costs from the two tables. All the proposed visual improvements were implemented, and the improved analysis page is shown in Figure 30.



Figure 30. Final data analysis tool

As can be seen in Figure 30, the “Notes For Use” text box has been removed and a dedicated Read Me page has been added, which is described in section 6.2.3. The slicer section at the top is now more compact due to the free space gained from changing the relationship boxes to tooltips. Free space at the top right from “Notes For Use” box removal has been used for Clear all slicers button and two card visuals, which show the current selection of Car Part and Customer Account slicers. The Slicers at the top have slightly different colour gradient than the data visuals, and the hour section has been divided with a line from the hardware cost section. The summary tables have been separated and placed next to the corresponding data visuals, and summary visuals have been added below the summary tables to visualize the distributions shown in the tables.

6.2.2 Functionality

Informant A recognized that toggling Spare part box and Calibration portal did not influence assembly hours, while they actually require assembly hours to collect the physical items to the spare part box and assemble the portal. This relation was added to the assembly hour calculation in the DAX code, by using SWITCH() function, which loops through four different scenarios, depending on the slicer condition. First, the DAX measure checks if neither spare part box or calibration portal is selected, and excludes their hours from the calculation, secondly and thirdly it checks if only one of them is selected and includes only the selected in the calculation and lastly it checks if both of them are selected and includes them both. Informant A also asked if a visual level data bar exclusion could be carried over to the summary table calculations, but this request was rejected because the current tool functionality does not support visual level relationships to calculated measures, which are used in the summary table. A note about exclusions not affecting the summary tables was added to the Read me page. He also noted that System Sizes “long” and “wide” could be combined into one, but this idea was put into backlog and might be implemented later if no detailed analysis are needed between the two sizes.

Informant C stated that the Analysis page could have a default slicing option selected when the report is opened, while also stating that when a domain expert user uses the report, he or she might know exactly what kind of configuration is needed and the default slicing option would not be necessary in these cases. This idea was rejected because the analysis page should be used to analyse exact configurations, and it should not guide the user to certain preset slicing direction.

Informant J stated that there are three similar car part selections available in the Car Part slicer: Subframe, Front Subframe and Rear Subframe. In reality, there are only the last two that Mapvision inspects, but all three values are in the CRM. Since front and rear subframes are very similar in terms of project hours

and costs, all three were combined into one “Subframe” in Power BI Power Query, resulting in only one showing in the slicing selection.

Informant F noted that Camera Type relation must be added in the future, and that a new WSI (Weld Seam Inspection) parameter must be added from the CRM. Both of these requests were placed in a backlog to be implemented later, because according to informant G, the Camera Type does not currently affect project hours, but it will affect them later once the WSI functionality will be implemented during projects and because adding a new WSI parameter requires thorough data quality analysis and relationship determination, which are not possible during the timeframe of this thesis.

6.2.3 New Pages

As discussed earlier, a read me page was added. This page contains a description of the functionality of the tool and its different pages, and several other key notes, such as when no selections are done in the slicer, Power BI determines that all of the slicers are selected.

Informant D requested two new pages, which would show hardware cost based on supplier and another one which would show individual component costs based on their different suppliers. The first request was implemented, and now a user can analyse the different routing link costs based on different suppliers, allowing detailed supplier price analysis for project specific components. The individual component analysis page was rejected, because key stock components are ordered from one supplier at a time and their costs are based on agreements, which are valid for a period of time.

Informants J and K requested a milestone duration page, which would show the previous project milestone durations based on the CRM parameters. This page would provide significant value in project planning, but it will be implemented later outside of this thesis, because the development and implementation of such page requires considerable amount of time and would not fit into the timeframe of this thesis.

6.2.4 Data Model Quality

As discussed in section 3.5.2, The following factors should be taken into consideration when evaluating data model quality: integration, completeness, simplicity, correctness, flexibility, implementability, integrity and understandability. (Helskyaho, et al., 2024) (Moody & Shanks, 2002).

The Power BI semantic model was sent to informant H for analysis, and he noted that the data model is robust, without any many-to-many relationships, and contains only the data that is needed. He noted that to increase clarity, the data model page could be re-organized and that the data tables could be renamed to be more informative than their original system names. He also noted that in addition to renaming, a description could be added to each data table to describe why the table is in this model. The improved data model is shown in Figure 31.

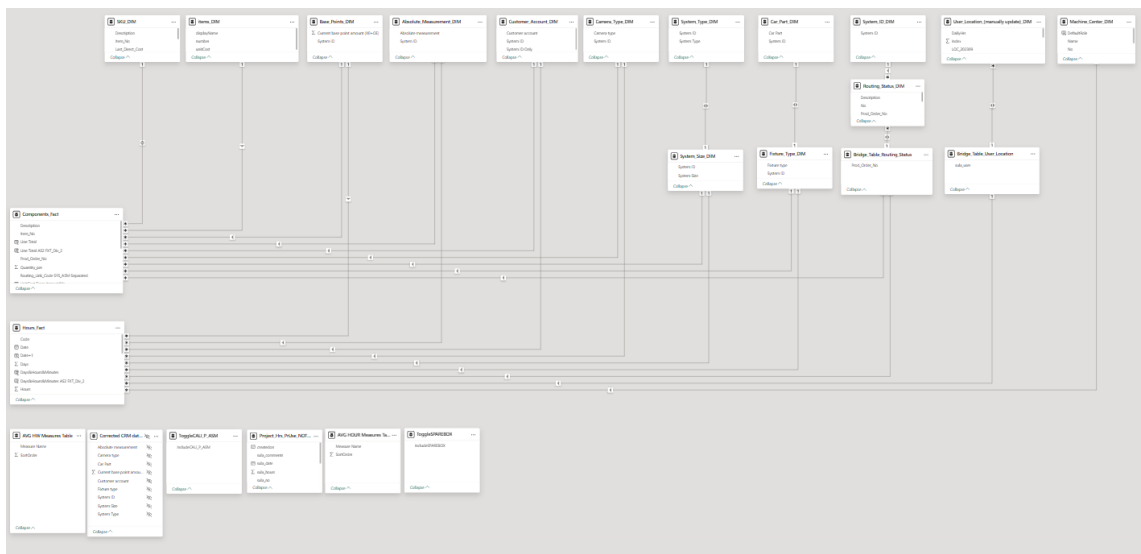


Figure 31. Improved final data model

As can be seen in Figure 31, the tables have been reorganized and renamed. The fact tables, which hold the measurable data, have a suffix `_FACT` in their names and the dimension tables, which are used for slicing and filtering, have a suffix `_DIM` in them. The unlinked tables, such as CRM master table and auxiliary tables are placed at the bottom, clearly showing there are no

relationships from them. Each table has a description in them which described why the table is in this model.

Informant H noted that the distributed CRM tables should be references of the master table instead of duplicates to increase the performance of the model. It was agreed that once the Dataverse link has been established between CRM and Dataverse, the new Dataverse table which will replace the current manually exported excel, will be constructed using reference tables. He also noted that all query steps that modify the data should be renamed, and a description should be added into them to increase the understandability of the model. Each “Custom Added” query step was renamed to have a unique name, and a description was added to explain what it is doing.

6.3 Summary of Final Data Analysis Tool

The improved Data Analysis Tool is now visually easy to understand by different stakeholders, and in addition to comprehensive project hour and cost analysis, it allows detailed self-service analytics. After implementing the previously discussed improvements, the tool should provide even more value in a clearer manner. The tool version 1.0 has now been published to Power BI Service, and all new additions, changes and improvements will be done as revisions, and the revision descriptions will be added at the bottom of the read me page.

This concludes the tool validation section, and the next section will end this thesis report in conclusions of the whole research project.

7 Conclusions

This section ends the thesis, and it first describes the executive summary of the whole research, provides next steps and recommendations for the case company and evaluates the trustworthiness of the thesis.

7.1 Executive Summary

The objective of proposing a data analysis tool and its parameters to increase the data usage in evidence-based decision making was reached by the release of a comprehensive stakeholder validated data analysis tool. The research project started with the literature-based creation of a conceptual framework for data analysis, which was closely followed throughout the current state analysis and tool development and validation phases. The current state analysis was limited to data which the tool uses, and it exposed major data quality issues in CRM and minor data quality issues in Business Central. Data amount was found out to be sufficient for the development of the tool, which was co-created together with key stakeholders. The tool proposal was then validated and commented by a larger group of experts, which lead to several improvements.

According to solution proposal validation group, the outcome of this thesis matches the project objective. The outcome provides value to many different stakeholders in the case company and supports them in data-driven decision making by visualizing the underlying data and its relations in understandable and clear format.

7.2 Next Steps and Recommendations

There are several development topics which can help Mapvision to further increase its data driven decision making, some of which are related to this tool, and some which are larger topics outside of the scope of this thesis.

Surface feature count should be added as a CRM parameter, to allow slicing of projector hardware cost in a meaningful way in the data analysis tool. Currently,

the projectors follow the same parameters as cameras, except the Camera Type -parameter. Hardware release level should be also added as a CRM parameter, not only to allow filtering in the tool, but also to aid operations and sales in case of retrofits to efficiently know the overall system configuration without the need for consulting design department. The WSI CRM parameter must be integrated to the tool, as discussed in section 6.2

As proposed by informant A, Artefact Design hours should be separated from overall design hours when more data is available for relatively new routing link ARTEF_DES, to allow more accurate hour estimation in the tool.

The milestone duration page, which was discussed in section 6.2.3 should be added, and in addition, it should be discussed if the tool could visualize planned and actual budgets from previous projects based on the same filtering criteria. This could give insights of problematic estimates during project planning.

The development of integrated Power BI AI features must be closely monitored, and if a useful feature is added by Microsoft, it should be tested in the tool with low threshold. Such features could be language-based analysis, where user can prompt the AI to analyse the data, based on normal language.

The whole CRM data should be analysed and corrected, using the methods from this thesis as a guideline. As seen in the current state analysis, the quality of the filtered data used for this thesis was not in usable condition, and it can be expected that the quality of the rest of the CRM data is has the same issues in it. Data quality metrics for all Mapvision data could be established, allowing future data analysis tools to be created more efficiently knowing the underlying data can be used as is.

Mapvision could connect the company sales forecast to the thesis outcome, to allow richer forecasting of future workload and cash flow. Similar parameters are already used in the sales forecast, and therefore the link creation between the analysis tool and sales forecast would not require great amounts work.

7.3 Evaluation of Thesis Trustworthiness

Well-planned research yields better results, and before the research can start, following three questions should be answered and documented in the research report: a problem statement, which defines the “what?” to research, a purpose, which defines the “why?” to research and research design, which defines the “how?” to research (Aityan, 2022). The business problem statement, the objective and the research design were defined and documented in detail in the beginning of this research to allow having a clear focus on the task at hand.

All the findings, including major data quality issues, are reported in the current state analysis section, enabling the creation of a comprehensive and accurate view of the current state. Minimal censoring is kept throughout the thesis to leave the reader with as few questions as possible after going through the whole report. Only the most business sensitive data is hidden, which includes customer accounts, project numbers, actual hour values and actual cost values. However, these are illustrated as comparable bar charts, where applicable.

The researcher of this thesis is not a data scientist but has considerable domain knowledge of each subject covered in this thesis. Research data was collected from key-stake holders in three different data stages, and all the collected research data is included in the appendices of this report. Even though the researcher works as a function head, he was not an informant at any point in the research to avoid possible biased answers. The research data was collected in open settings, allowing the informants to formulate their own and honest view of the discussed topics. The tool proposal evaluation was done by a large group of experts, each scrutinizing and examining the tool from different points of view, to achieve as extensive evaluation as possible in the given timeframe.

The progress of the thesis was presented six times after each gate to Metropolia University of Applied Sciences teachers and student colleagues, who scrutinized the outcomes of each gate. These sessions made sure the

research was on track and did not miss any crucial aspects during different phases of the project.

7.4 Closing Words

Releasing the outcome of this thesis was very rewarding since it provides practical value to multiple stakeholders, and it supports the organisation in evidence-based decision making. The current state analysis provided insight to data quality challenges that are currently present in CRM and in Business Central. Mapvision data culture is evolving, and this thesis has been a great possibility to expose multiple colleagues to data in a more detail level.

References

- Afsharian, M., 2024. Data science essentials in business administration:. *Decision Analytics Journal*, Issue 11.
- Aityan, S. K., 2022. *Business Research Methodology, Research Process and Methods*. 1st ed. s.l.:Springer Cham.
- Alasadi, S. A. & Wesam, B. S., 2017. Review of Data Preprocessing Techniques in Data Mining. *Journal of Engineering and Applied Sciences*, Issue 16, pp. 4102-4107.
- Amigo, M. J., 2021. Data Mining, Machine Learning, Deep Learning, Chemometrics, Definitions, Common Points and Trends. *Brazilian Journal of Analytical Chemistry*, VIII(32), pp. 45-61.
- Bansal, A., 2023. POWER BI SEMANTIC MODELS TO ENHANCE DATA ANALYTICS AND DECISION-MAKING. *International Journal of Management (IJM)*, XIV(5), pp. 136-142.
- Brodlie, K., Allendes, O. R. & Lopes, A., 2012. A review of uncertainty in data visualization. *Expanding the Frontiers of Visual Analytics and Visualization*, pp. 81-110.
- Deckler, G. & Powell, D., 2024. *Microsoft Power BI Cookbook: Convert Raw Data into Business Insights with Updated Techniques, Use Cases, and Best Practices*. Birmingham: Packt Publishing.
- Diepeveel, M.-J., 2022. *Artificial Intelligence with Power BI*. Birmingham: Packt Publishing.
- Espadinha-Cruz, P., Godina, R. & Rodrigues, E. M. G., 2021. A Review of Data Mining Applications in Semiconductor. *Processes*, Issue 9.

Felin, T. et al., 2021. The data-hypothesis relationship. *Genome Biology*, Issue 22.

Galvão, A. B. & Mitchell, J., 2024. Communicating Data Uncertainty: Multiwave Experimental Evidence for UK GDP. *Journal of Money, Credit and Banking*, 56(1).

Gülseçen, S., Sushil, K. S. & Akadal, E., 2020. *Who Runs the World: Data*. Istanbul: University Press.

Haiyan, Y., 2024. *Data Quality Management in the Data Age: Excellence in Data Quality for Enhanced Digital Economic Growth*. 1 ed. s.l.:Springer Cham.

Healy, K., 2019. *Data Visualization: A Practical Introduction*. 1st ed. Princeton, Woodstock: Princeton University Press.

Helskyaho, H., Ruotsalainen, L. & Männistö, T., 2024. Defining Data Model Quality Metrics for Data Vault 2.0 Model Evaluation. *Inventions*, 1(9).

Kananen, J., 2013. *Design Research (Applied Action Research) as Thesis Research: A Practical Guide for Thesis Research*. 1st ed. s.l.:JAMK University of Applied Sciences.

Kopanakis, I. & Theodoulidis, B., 2003. Visual data mining modeling techniques for the visualization of mining outcomes. *Journal of Visual Languages and Computing*, Issue 14, pp. 543-589.

Larose, D. T., 2014. *Discovering Knowledge in Data: An Introduction to Data Mining*. 2nd ed. s.l.:John Wiley & Sons, Inc.

Moody, D. L. & Shanks, G. G., 2002. Improving the quality of data models: empirical validation of a quality management framework. *Information Systems*, Issue 28, pp. 619-650.

Pigott, T. D., 2001. Educational Research and Evaluation. *Educational Research and Evaluation*, VII(4), pp. 353-383.

Pipino, L. L., Lee, Y. W. & Wang, R. Y., 2002. Data Quality Assessment. *COMMUNICATIONS OF THE ACM*, 45(4).

Romero-Organvidez, D., Horcas, J.-M., Galindo, J. A. & Benavides, D., 2024. Data visualization guidance using a software product line approach. *The Journal of Systems & Software*.

Saliccioli, D. J., Crutain, Y., Komorowski, M. & Marshall, D. C., 2016. *Chapter 17. in: MIT Critical Data (ed.) Critical Data: Secondary Analysis of Electronic Health Records*. Cambridge: Springer Cham.

Saunders, M., Lewis, P. & Thornhill, A., 2019. *Research methods for business students*. 8th ed. s.l.:Pearson Education Limited.

Sirichanya, C. & Kraissak, K., 2021. Semantic data mining in the information. *International Journal of Intelligent Systems*, Issue 36, pp. 3880-3916.

Stone, P., 2002. Issues in research: Deciding upon and refining a research question. *Palliative Medicine*, Issue 16, pp. 265-267.

Stoudt, S., Vásquez, V. N. & Martinez, C. C., 2021. Principles for data analysis workflows. *PLOS Computational Biology*.

Zhao, Z. et al., 2017. Controlling False Discoveries During. *SIGMOD '17*, pp. 527-540.

Appendices

Appendix 1 Data Stage 1 Current State Analysis interview notes

Interview form	
Master's Thesis, Mapvision data current state analysis (Operations data)	
Interviewer Mika Salonen	
Name (Code) of the informant A	
Position in the company Head of Production Operations	
Date of the interview 05/02/2025	
Duration of the interview 1 hour	
<p>Question 1a What operations related data MV currently has / is collecting (if multiple, list each)</p> <p>Question 1b What is the quality of the data in scale of 0-5 (if multiple datasets, grade each separately) 0: Provides no value and cannot be used 5: Is high quality, available, understandable and is / can be used</p>	<p>Commercial documents, customer input data, 3D, spec, requirements: 3,5 Project management documents, project plan, budget, customer contact information etc.: 3,5 Meeting minutes, emails, Teams chat: 4 Design Data: 3D, 2D, BOM (Teamcenter), eplan: 5 Approval documents / protocol / agreements: 4 BOM (BC): 4 Purchase orders: 4 Order confirmations: 3 Invoices -> components prices: 3,5 Stock levels: 4 Mapvision proprietary data (Measurement specification): 4 QTY of cameras, projectors, optics: 4 Checklists: 2,5 Measurement data: 4 Calibration data, calibration results, certificate: 5 Measurement images, measurement results: 4 Reports: 4 Logistics documents: 4 Work hours: 4,5 Hubspot data: customer / system details: - Vendor data: 3,5</p>
<p>Question 2a Where is the data stored?</p>	Sharepoint, OneDrive, Teamcenter, Business Central, Hubspot, Email, Teams, MapCloud, Dataverse
<p>Question 2b Describe the accessibility of the data storage locations</p>	Taking into account the amount of data, the accessibility is on acceptable level. Not perfect level due to: data is distributed into multiple locations. Some data might be only in personal emails. Acceptable because: currently available reports. Data locations policies are well communicated and followed
<p>Question 3a How is the data collected (if multiple methods, describe each)</p> <p>Question 3b What is the quality of the collection method in scale of 0-5 (if multiple, grade each separately)</p>	<p>Commercial documents, customer input data, 3D, spec, requirements: trough sales via different transfer emails or through customer data portal: 3 Project management documents, project plan, budget, customer contact information etc: Manual : 4 Meeting minutes, emails, Teams chat: Manual: 3 Design Data: 3D, 2D, BOM (Teamcenter), eplan: Teamcenter & Eplan: 4,5 Approval documents / protocol / agreements: Manual: 4 BOM (BC): Via excel import (manual): 4 Purchase orders: Automatic via BC: 5 Order confirmations: email: 2 Invoices -> components prices: email + scanning service Apix. Component prices: (manual or automatic?): 4 Stock levels: Automatic: 4,5 Mapvision proprietary data (Measurement specification): Created in Mapvision Software: 4,5 QTY of cameras, projectors, optics: same as above: 4,5 Checklists: manual: 2 Measurement data: Generated by MV SW: 5 Calibration data, calibration results, certificate: Generated by MV SW: 5, cert: manual: 4 Measurement images, measurement results: Generated by MV SW: 5 Reports: Generate by MV SW: 5 Logistics documents: Manual / partly automatic via BC: 3 Work hours: Via Power App (manual input): 4 Hubspot data: customer / system details: manual: - Vendor data: Initial creation manually, statistics automatically in BC: 4</p>
<p>Question 4a How MV currently uses its operations data</p> <p>Question 4b Grade the usage of the data in grade 0-5 (if multiple uses, grade each separately) 0: Use does not provide any value, or provides negative value 5: Use provides significant value for decision making and/or forecasting</p>	<p>Data is used: As input for next process step: 5 For statistics: 5 Reporting: 4 Used in decision making: 4 Resourcing: 4</p>
<p>Question 5a Describe the company data culture (hint: strong data driven culture: all decision makers are <u>required</u> to focus on the information provided by data; no data culture: decisions are made based on experience and or gut-feeling)</p>	Company is highly data driven. Company is transparently communicated and well available throughout the organization. Operations are strongly data driven, Finance operates only based on data, R&D is less data driven than operations since it is based on opinions in addition to data. Resourcing cannot be 100% data driven, since the company business model is based on project deliveries, and they fluctuate by nature (customer dependency, customer input data varies).
<p>Question 5b Who are the primary users of data, if any</p>	Everyone. Decision making, reporting & statistics are primarily used by function leaders and management

Interview form	
Master's Thesis, Mapvision data current state analysis (Operations data)	
Interviewer Mika Salonen	
Name (Code) of the informant B	
Position in the company Chief Financial Officer	
Date of the interview 06/02/2025	
Duration of the interview 1 hour	
<p>Question 1a What <u>operations</u> related data MV currently has / is collecting (if multiple, list each)</p> <p>Question 1b What is the quality of the data in scale of 0-5 (if multiple datasets, grade each separately) 0: Provides no value and cannot be used 5: Is high quality, available, understandable and is / can be used</p>	<p>Project data: total income & cost (actual & forecast): 4 Project timeline, resource consumption on timescale: 2 (=Based on general model, not on project specific Routings) Procurement & inventory management (DIO, stock levels, project needs): 3 Product cost model: 3 Workhours: 4 Order intake, Sales forecast and pipeline: 4</p>
<p>Question 2a Where is the data stored?</p>	ERP + BI reporting, Onedrive Excel files, Dataverse
<p>Question 2b Describe the accessibility of the data storage locations</p>	Relatively good, especially everything in BI and ERP. Certain excel files are not easily available for all persons requiring data (certain model based reports are done manually in excel)
<p>Question 3a How is the data collected (if multiple methods, describe each)</p> <p>Question 3b What is the quality of the collection method in scale of 0-5 (if multiple, grade each separately)</p>	Manual field inputting into systems: 4, manual excel inputs based on templates: 4, fully automatic (reports): 5
<p>Question 4a How MV currently uses its operations data</p> <p>Question 4b Grade the usage of the data in grade 0-5 (if multiple uses, grade each separately) 0: Use does not provide any value, or provides negative value 5: Use provides significant value for decision making and/or forecasting</p>	Used for: Supporting management at all levels decisions: 3,5 Stake holder reporting: 4
<p>Question 5a Describe the company data culture (hint: strong data driven culture: all decision makers are <u>required</u> to focus on the information provided by data; no data culture: decisions are made based on experience and or gut-feeling)</p>	Data "driveness" is increasing, employees are willing to lead with data, but company data quality and availability have only recently improved. Data-entry culture requires improvement - everyone does not have visibility how entries effect downstream reports / operations
<p>Question 5b Who are the primary users of data, if any</p>	Management on all levels

Interview form	
Master's Thesis, Mapvision data current state analysis (Operations data)	
Interviewer	Mika Salonen
Name (Code) of the informant	C
Position in the company	Chief Business Officer
Date of the interview	
Duration of the interview	1 hour
<p>Question 1a What <u>operations</u> related data MV currently has / is collecting (if multiple, list each)</p> <p>Question 1b What is the quality of the data in scale of 0-5 (if multiple datasets, grade each separately) 0: Provides no value and cannot be used 5: Is high quality, available, understandable and is / can be used</p>	<p>System/project BOM: 5 Component purchase prices: 4 Other direct project costs (packing, shipping, sub contracting, travel costs): 4 Work hours: 4 Project run times: 4 Sales price: 5 Invoicing milestones: 4 Contribution margin: 4 Gross margin: 4 System Data base (CRM): 3 Customer contacts (CRM): 4 Deals (CRM): 4 Stock quantities: 4</p>
<p>Question 2a Where is the data stored?</p>	<p>Teamcenter (PDM), Business Central (ERP), HubSpot (CRM), subsidiary bookkeeping systems, Dataverse, One Drive Excel files</p>
<p>Question 2b Describe the accessibility of the data storage locations</p>	<p>PDM, ERP, CRM, Dataverse (=Power BI reports) availability is good for employees with required access rights. Subsidiary bookkeeping & OneDrive Excel files requires support from Finance team</p>
<p>Question 3a How is the data collected (if multiple methods, describe each)</p> <p>Question 3b What is the quality of the collection method in scale of 0-5 (if multiple, grade each separately)</p>	<p>System/project BOM: Generated by Design Engineer from system guided alternatives: 5 Component purchase prices: No Knowledge of current method. Other direct project costs (packing, shipping, sub contracting, travel costs): From invoices automatically: 5. Subsidiary invoices are collected manually: 4 Work hours: PowerApp (manually) -> Dataverse: 3 Project run times: Semi-automatic system guided: 4 Sales price: Manual: 3 Invoicing milestones: Manual: 3 Contribution margin: Calculated: 5 Gross margin: Calculated: 5 System Data base (CRM): Manual (drop-down menus): 3 Customer contacts (CRM): Manual: 3 Deals (CRM): Manual: 3 Stock quantities: Automatic + manual periodic checks: 4</p>
<p>Question 4a How MV currently uses its operations data</p> <p>Question 4b Grade the usage of the data in grade 0-5 (if multiple uses, grade each separately) 0: Use does not provide any value, or provides negative value 5: Use provides significant value for decision making and/or forecasting</p>	<p>System/project BOM: Documentation, assembly, shipping documentation, system cost (together with purchase prices), system cost analysis per configuration: 5 Component purchase prices: Used in above mentioned, component purchase price trend analysis and bench marking, Inventory value: 5 Other direct project costs (packing, shipping, sub contracting, travel costs): Same as above: 4 Work hours: Flex balance, work hour reporting (legal), revenue recognition, project effort analysis, capacity forecasting and planning, gross margin: 5 Project run times: Project planning: 2 (data could be used more effectively) Sales price: Contribution and gross margin calculation, cash flow forecasting: 5 + (company financial figures) Invoicing milestones: Invoicing, cash flow forecasting: 5 Contribution margin: Profitability analysis and assessment, pricing: 5 Gross margin: Profitability analysis and assessment: 5 System Data base (CRM): Market analysis (what sold where and to whom (feature counts, system types etc.)), Internal documentation (for example: verifying system configuration): 4 Customer contacts (CRM): Marketing (contacting, newsletters): 5 Deals (CRM): Progress follow-up, pipeline value (weighted and un-weighted): 5 Stock quantities: Inventory value, stock refill decisions: 4</p>
<p>Question 5a Describe the company data culture (hint: strong data driven culture: all decision makers are <u>required</u> to focus on the information provided by data; no data culture: decisions are made based on experience and or gut-feeling)</p>	<p>Data culture is evolving, currently we are in the "middle" of no and strong culture. Business cannot rely only on data, but it requires expert judgement (but data is used to ground the decisions)</p>
<p>Question 5b Who are the primary users of data, if any</p>	<p>Everyone in Operations, Product Business Management, Finance, Sales uses atleast some amount of above mentioned data in their daily work. R&D uses some parts of the data, but not daily</p>

Interview form	
Master's Thesis, Mapvision data current state analysis (Operations data)	
Interviewer	Mika Salonen
Name (Code) of the informant	D
Position in the company	Chief Operating Officer
Date of the interview	06/02/2025
Duration of the interview	1 hour
<p>Question 1a What <u>operations</u> related data MV currently has / is collecting (if multiple, list each)</p> <p>Question 1b What is the quality of the data in scale of 0-5 (if multiple datasets, grade each separately) 0: Provides no value and cannot be used 5: Is high quality, available, understandable and is / can be used</p>	<p>Work hours per project: 4 Work hours used for service: 2 Project specific Bill of Material: 4 Component cost: 4 Commercial documents (project related): 4 Commercial documents (supplier related): 3 Design documents: PDM (Product Data Management): 4 System database (CRM): 3 Service tickets (CRM): 4 Sales pipeline (CRM), actual & forecast: 4 Inventory (BC): 4 Project specific data (cloud): 4 Internal quality tickets: 3</p>
<p>Question 2a Where is the data stored?</p>	PDM, CRM, ERP, Float, SharePoint, OneDrive, MapCloud, Dataverse. Some work files might be local, but end product is in some of the listed locations
<p>Question 2b Describe the accessibility of the data storage locations</p>	Role based access to relevant data is available to all of the above
<p>Question 3a How is the data collected (if multiple methods, describe each)</p> <p>Question 3b What is the quality of the collection method in scale of 0-5 (if multiple, grade each separately)</p>	<p>In house developed Power App for hour logging: 4 MS office tools / files based on templates: 4 Data transfer (BOM): 4 Automatic email transfer (ticket system): 4 CRM database (field inputs): 3 CRM pipeline (field inputs): 4 Float userface input: 4 Upload (SharePoint, OneDrive, Mapcloud): 4 Power App (quality ticket): 4</p>
<p>Question 4a How MV currently uses its operations data</p> <p>Question 4b Grade the usage of the data in grade 0-5 (if multiple uses, grade each separately) 0: Use does not provide any value, or provides negative value 5: Use provides significant value for decision making and/or forecasting</p>	<p>Forecasting (resourcing, financial): 3 Business reporting, incl. project reporting: 4 Operational excellence / development: 3</p>
<p>Question 5a Describe the company data culture (hint: strong data driven culture: all decision makers are <u>required</u> to focus on the information provided by data; no data culture: decisions are made based on experience and or gut-feeling)</p>	Decisions are based on data - but previous experience and customer relationships are strong additional driver
<p>Question 5b Who are the primary users of data, if any</p>	Everyone uses data - different roles for different purposes. Example: higher management uses for decision making and engineers use to get the job done as efficiently as possible

Interview form	
Master's Thesis, Mapvision data current state analysis (Operations data)	
Interviewer	Mika Salonen
Name (Code) of the informant	E
Position in the company	Head of Project Management
Date of the interview	06/02/2025
Duration of the interview	1 hour
<p>Question 1a What <u>operations</u> related data MV currently has / is collecting (if multiple, list each)</p> <p>Question 1b What is the quality of the data in scale of 0-5 (if multiple datasets, grade each separately) 0: Provides no value and cannot be used 5: Is high quality, available, understandable and is / can be used</p>	<p>Project Follow-up excel: 4 Other project management related documents: 4 MV specific project data: 3 Templates folder: 4 G/L Budgets: 5 Project timelines / tasks: 4 Dashboards: 4 Workhours: 4 Instructions: 4 Project overview excel: 4 System database (CRM): 3 Customer surveys: 5</p>
<p>Question 2a Where is the data stored?</p>	Sharepoint, ERP, Mapcloud, Dataverse, CRM
<p>Question 2b Describe the accessibility of the data storage locations</p>	Accessibility is on good level
<p>Question 3a How is the data collected (if multiple methods, describe each)</p> <p>Question 3b What is the quality of the collection method in scale of 0-5 (if multiple, grade each separately)</p>	Templates - manual input: 2, automatic (dashboards + overview excel): 5
<p>Question 4a How MV currently uses its operations data</p> <p>Question 4b Grade the usage of the data in grade 0-5 (if multiple uses, grade each separately) 0: Use does not provide any value, or provides negative value 5: Use provides significant value for decision making and/or forecasting</p>	<p>Data is used for: Project management: 4 Financial steering: 5 Forecasting, hours: 3 Forecasting, HW: 4 System assembly: 4 System programming: 4</p>
<p>Question 5a Describe the company data culture (hint: strong data driven culture: all decision makers are <u>required</u> to focus on the information provided by data; no data culture: decisions are made based on experience and or gut-feeling)</p>	Data is the basis for decision making in the company
<p>Question 5b Who are the primary users of data, if any</p>	Everyone

Interview form	
Master's Thesis, Mapvision data current state analysis (Operations data)	
Interviewer	Mika Salonen
Name (Code) of the informant	F
Position in the company	Chief Product Officer
Date of the interview	11/02/2025
Duration of the interview	1 hour
<p>Question 1a What <u>operations</u> related data MV currently has / is collecting (if multiple, list each)</p> <p>Question 1b What is the quality of the data in scale of 0-5 (if multiple datasets, grade each separately) 0: Provides no value and cannot be used 5: Is high quality, available, understandable and is / can be used</p>	<p>Customer measurement specification, 3D models, general terms, standards, requirements: 3,5 (=in certain projects, we have too much unrelevant data) Mapvision Quotes: 4 Configured price configurators (source for quotes): 3,5 Price configurator template (Xlsx): 3,5 (working OK, some items missing / needs to be updated) Quote template (Word): 4 Purchase Orders / Letter of Intents: 3,5 (not always available in correct location, or POs are referring to different Quote version than expected)</p>
<p>Question 2a Where is the data stored?</p>	Sharepoint
<p>Question 2b Describe the accessibility of the data storage locations</p>	Is on very good level
<p>Question 3a How is the data collected (if multiple methods, describe each)</p> <p>Question 3b What is the quality of the collection method in scale of 0-5 (if multiple, grade each separately)</p>	Manually created or moved from other sources: 3 (could be better = more structured - guide is "too" structured, but users sometimes missplace data)
<p>Question 4a How MV currently uses its operations data</p> <p>Question 4b Grade the usage of the data in grade 0-5 (if multiple uses, grade each separately) 0: Use does not provide any value, or provides negative value 5: Use provides significant value for decision making and/or forecasting</p>	<p>Feasibility analysis during presales: 5 (without data this cannot be done. If data is not available, generated quote labeled as budgetary or with added disclaimers) GTC, Standards and Requirements used during projects: 5 Quotes are approved through approval process: 5</p>
<p>Question 5a Describe the company data culture (hint: strong data driven culture: all decision makers are <u>required</u> to focus on the information provided by data; no data culture: decisions are made based on experience and or gut-feeling)</p>	In presales, quotes are not sent without data. As discussed, if data is missing, quotes are sent as budgetary. In certain cases, experience drives the decision, but when we have data, data is used. Data is also used in product development
<p>Question 5b Who are the primaly users of data, if any</p>	Everyone

Appendix 2 Data stage 2 Co-creation meeting minutes

Data Analysis Tool Co-Creation Meeting Minutes

Meeting Room Taiichi, 26.2.2025, 15:30-16:30

Attendance:

Mika Salonen, Head of System Design

Bo Lönnqvist, Chief Business Officer, Thesis Advisor, Informant C

1. The meeting was opened by Mika 15:30.
2. Mika presented the findings from Current State Analysis
 - a. Bo suggests using Excel for CRM data for the time being. Corrections should be done into this excel, and it can then be later used as guideline to correct the CRM data. It was agreed that later we make decision on acquiring CRM – Dataverse link license.
3. Bo identified that Current Base Point (X0 + OE) is a value that might change during the lifetime of a project. A decision was made to make a new column in CRM, called "Handover to Care Current Base Point Qty", which is not modified after system delivery, and will be used in this tool.
4. It was agreed that System Type and Version columns will be modified as follows:
 - a. System Type: AD1, AD2, AS1, AS2, AS1 MCK, AS2 MCK, Custom
 - b. System Size: Long, Wide, S, L, 2L, W, 2W, 3W, 4W, 5W, MAX, Custom
 - c. This column separation allows better control of the filtering conditions over the system "general" type and its Q-series modular size. Custom type and size allow filtering out "custom-custom" projects.
5. An additional column in CRM describing the system Hardware Release level was discussed. It could separate the system HW release into major releases: old, 16, 18, 23, 25. This was decided to be added later.
6. General: Bo noted that the tool will be useful in quote verifications and the tool will be very useful in project planning when estimating actual work hours.
7. The meeting was closed at 16:30

Data Analysis Tool Co-Creation Meeting Minutes

Meeting Room Taiichi, 5.3.2025, 14:30-15:00

Attendance:

Mika Salonen, Head of System Design

Informant G, Head of Project Engineering

1. The meeting was opened by Mika 14:30.
2. Project categorization and hour filtering idea was presented to informant G, and the matter was discussed:
 - a. Implementation hours accrue until Handover to Care.
 - b. Fixture type does not affect implementation hours.
 - c. System type affects implementation hours.
 - d. Camera type does not affect setup hours.
 - e. Camera type does not affect implementation hours.
 - f. Measurement type affects design hours, since artefact design requires considerable effort. Could artefact design hours be separated from other design hours?
 - g. Discussed if base points have an effect in Camera Amount. Agreed they do not, and Car Part alone is a good filter for Camera Amount.
3. General notes about the tool:
 - a. The tool must show clearly which parameter affects what.
 - b. Surface features were discussed. It was agreed that they are currently not being recorded in CRM, and the possibility of adding them will be noted for further development.
4. The meeting was closed at 15:00.

Data Analysis Tool Co-Creation Meeting Minutes

Meeting Room Taiichi, 10.3.2025, 13:30-14:00

Attendance:

Mika Salonen, Head of System Design

Informant E, Head of Project Management

1. The meeting was opened by Mika at 13:30.
2. The current draft of the tool and its relationships and filters were presented by Mika.
 - a. Informant E noted that the focus is different than in current project planning excel. It was noted that it needs to be thought through how we can use this effectively.
 - b. Informant E notes that Customer Account has the largest effect on Project Management hours, followed by Measurement type.
3. Discussed about the tool sensitivity. It was noted that new modifications which affect hours and hardware cost will not be reflected in the tool immediately.
4. Hour categories do not match the current Project Reporting Power BI report, and it was agreed they should match. Participants agreed to call a second meeting to discuss it again once the categorization matches.
5. The meeting was closed at 14:00.

Data Analysis Tool Co-Creation Meeting Minutes

Meeting Room Taiichi, 10.3.2025, 13:30-14:00

Attendance:

Mika Salonen, Head of System Design

Informant A, Head of Production Operations

1. The meeting was opened by Mika at 13:30.
2. The draft of the tool was presented by Mika.
 - a. A categorization proposal by Informant A via Teams Chat earlier was implemented already for this meeting:
 - i. Fixture Type affects Assembly hours.
 - ii. System Type affects Assembly hours.
3. Informant A noted that Artefact Assembly hours might not be currently present when RDYTOSH project milestone is finished, due to current supplier challenges, but we will continue as is. The supplier issues will be fixed later.
4. Informant A noted that Measurement type affects Design Hours due to Artefact design.
5. It was agreed that additional items, such as extra PC, Scanner, Printer, Temperature compensation kit, Extra UI, Safety lock and Freight type will not be separately selectable, but they will be averaged in the tool instead.
6. It was noted that changes in HW costs are not immediately reflected in the tool.
7. A discovery was made; Project milestones for recently shipped systems, which have different Car Part in them, were not updated to RDYTOSH = Finished. This update was done manually during the meeting, and it was observed that part type indeed affects camera cost.
8. Informant A suggested separating total Assembly hours to System and Fixture hours
 - a. This was agreed to be a good idea, and implementation was planned
9. Separating Artefact design to its own category from general Design hours was discussed. It was agreed that this separation will be done later, when we have more hour data for ARTEF_DES routing link.
10. The meeting was closed at 14:00.

Data Analysis Tool Co-Creation Chat notes

Teams Chat, 5.3.2025

Mika Salonen, Head of System Design

Informant H, Business Controller

1. Mika inquired about Project Reporting Power BI hour categorization and explained the Analysis Tool should match the Project Report.
 - a. Informant H described the functionality of the categorization, which is done with a DAX measure in Power Bi.
 - b. It was agreed that Mika will copy and paste the required queries from Project Reporting Power Bi report to the Analysis Tool Power Bi and use the same DAX measure for the categorization of hours.

Data Analysis Tool Co-Creation Meeting Minutes

Meeting Room Taiichi, 12.3.2025, 09:00-10:00

Attendance:

Mika Salonen, Head of System Design

Informant E, Head of Project Management

1. The meeting was opened by Mika at 09:00.
2. A new version of the tool draft was presented, and it was explained that the hour categorization now follows the Project Reporting Power Bi report.
 - a. The categorization was agreed to be OK
3. Informant E asked could we add general Project hardware and hour summaries to the same report for self-analytics purposes?
 - a. This was agreed to be a good addition, and implementation was agreed.
4. Informant E noted that it is very important to add notes in the report showing which filters what.
5. Customer Account grouping was discussed, but no clear grouping was found. It was agreed to leave Customer Accounts ungrouped for now.
6. The meeting was closed at 10:00

Data Analysis Tool Co-Creation Meeting Minutes

Teams Meeting, 20.3.2025, 12:00-12:40

Attendance:

Mika Salonen, Head of System Design

Bo Lönnqvist, Chief Business Officer, Thesis Advisor, Informant C

Informant F, Chief Product Officer

1. The meeting was opened by Mika at 12:00.
2. The current version of the tool was presented. It was explained that all the previously requested changes and improvements have already been implemented.
3. Project slider should be added, using the project numbers. Today it is not as critical, but in the future, it would be beneficial to be able to filter old projects out. It was agreed that this will be implemented.
4. Self-service analytics could include comparison between Camera Type & Measurement Type. It was agreed this will be implemented.
5. It was agreed with Bo that the tool can be released to Validation Phase after #3 and #4 are completed. The validation group was earlier discussed with Head of Project Management, and it was now agreed to be:
 - a. Two Project Managers
 - b. Project Budget Approval Group
 - c. Bo Lönnqvist
 - d. Chief Product Officer
 - e. Business Controller
 - f. Chief Financial Officer
 - g. Director, Core Applications
6. The meeting was closed at 12:40

Appendix 3 Data Stage 3 Validation feedback

Informant	Improvement ideas	General
A	1) Could System Size Long & Wide be combined? 2) Could visual exclusion affect also the summary table? 3) Could Spare part box and Calibration portal selection also affect workhours? 4) Summary table line order changes based on values, could they be static? 5) Could the relationship boxes be elsewhere to increase clarity? 6) Move the Hardware summary table down next to hardware bar charts.	<ul style="list-style-type: none"> • Relationships between CRM parameters and hours and hardware costs make sense and are good • Relationships are not too detailed, keeping the amount of datapoints high • The report (=the tool) makes perfect sense and provides value for decision making, especially during project planning • The tool is logical, and the layout is mainly clear
B	1) Slicer selection to be clarified; nothing selected = everything shown 2) Currently selected slicers to be shown. In long lists of slicers, some selections might be "hidden" when the list is scrolled	<ul style="list-style-type: none"> • Tool is logical to follow • Drill down feature provides value • Data accuracy in the long term is questionable. Data quality check should be added to the delivery process, especially for Hardware. Perhaps all routing links could be double checked after shipping?
C	1) Could there be a default slicing option. Now many slicer settings are needed 2) Add total Hardware and Hour graphs next to the tables 3) Could we add for example Fixture types as columns instead of projects?	<ul style="list-style-type: none"> • Not many PM hours shown • Freight, packing and other direct project costs missing, which are required for financial analysis
D	1) Clarify routing links in Hardware cost Self-service analytics page 2) Add filtering by supplier in Hardware cost Self-service analytics page 3) Add supplier-based analytics for components (same component, different supplier)	<ul style="list-style-type: none"> • Tool is comprehensive, and filters are correct, allowing detailed analysis • Drill through is a good feature • Tool provides value to team leaders in operations when forecasting resources
E	1) Relationships not 100% clear to understand, maybe add a "how-to" page to the report describing how the report is working 2) Change selection of multiple Customer Accounts to be similar as others (no ctrl click needed for	<ul style="list-style-type: none"> • Is the tool data shown in real time? Could this info be added in the "how-to" page? • Tool provides value in second stage of project planning, and it can be used to validate the output

	multiple selection) 3) Add a "Clear slicers" button	of the Excel based planning configurator
F	1) Camera type relation to hours must be added, not now, but later 2) WSI parameter should be added, separate from Camera type	<ul style="list-style-type: none"> • Tool is good • Artefact costs are interesting topic to monitor
G	1) Difference between two self-service pages not clear	<ul style="list-style-type: none"> • Base points missing from many projects • Tool works well. Gives more information about PE and Implementation hours when more projects are closed
H	<ol style="list-style-type: none"> 1) Change relationships to tooltips instead 2) Separate summary tables to increase readability 3) Define different colour theme for slicers 4) Separate hours and hardware with a line 5) See if hardware and hours could use a different colour theme 6) Maybe add totals to the table bottom rows via measure 7) Restructure data model page 8) Rename data tables 9) Add description to data tables 10) Rename and add information in query step properties 11) Tables should be reference tables instead of duplicates 	<ul style="list-style-type: none"> • Overall, the report makes sense • Data model is robust without m2m relationships • Bidirectional relationships are purposely made, and they make sense • Hidden pages are not actually the best method to document model work. However, before a new method is defined, we continue like this • CRM data tables can be left as duplicates, but they should be made as references once Dataverse link is established • User Location table is currently created in Query, which is not good. This should be either in online Excel, or better alternative would be a Dataverse table created in PowerApps
I	1) Change layout in a way that project filter is clearly visible when zoomed in	<ul style="list-style-type: none"> • Overall, a good tool, drill down is a good function for analysis

J	<p>1) Combine Subframe, Front Subframe and Rear Subframe into one 'Subframe'</p> <p>2) Could milestone durations in weeks be added?</p> <p>3) The CRM parameter selection should be reorganized to: System Type - System Size - Car Part - Fixture type - Customer account</p>	<ul style="list-style-type: none"> • The tool provides value to me. However, it should not replace the Excel based configurator since the tool shows only data which we have, so some system configurations are not visible • The tool should be used to verify the Excel based configuration for accurate plan • The tool makes sense • At the moment, the Customer account does not provide much value, since we do not have that many projects in closed state
K	<p>1) Could the self analytics routing link list be simplified, it has quite many items in it due to our routing</p> <p>2) Could we drill down time plan accuracy per account or project</p> <p>3) Could we add information for service use, such as milestones and dates?</p>	<ul style="list-style-type: none"> • Can we filter in the analysis page by project? • Overall, the tool is great