Bachelor Thesis

Industrial Engineering and Management

Developing a data-analysis dashboard for Forque

Job D. G. Velthuis

S1935526

University of Twente

Faculty of Behavioural Management and Social Sciences

Supervisors University of Twente

[1] Dr. Ir. E.A. Lalla

[2] Dr. A. I. Aldea

Supervisor Forque Advisory Services B.V.

[1] Noël Pepermans

I. Executive summary

Forque Advisory Services BV is a data consultant company in Enter, Twente. They provide services to other companies such as implementing, maintaining, and optimisation of data systems. The company is specialised in AFAS and PowerBI.

Currently, Forque is making policies on how their sales division should function based on manual calculations and a dashboard that uses the company data on their projects and employees. The problem with this dashboard is that it uses broad assumptions to perform the calculations and does not include parts of the data of the company that helps it resemble reality. The problem is that the dashboard has calculations that do not represent the reality and give the management a false idea of what their goals should be. They experience the need to have this model restructured to help make it future proof and make it possible to predict the future.

In order to solve this problem, the research was designed with the goal to *design a tool to calculate the number of leads on a daily basis.* To make the problem manageable, smaller research questions were asked. They are on the current situation, deciding the problem characteristics, what data to use, and how to design a dashboard.

The selecting of the right data and primary and foreign keys was an important sub question of the designing of the tool. The company IT staff, managers, and project managers were interviewed on how the data was linked, what data was relevant, and what the data represent.

The outcome of this thesis is a tool in the form of a dashboard that analyses the data, performs calculations on the data, and visualises the outcome of the calculations and data. It gives the user insight into the data and the ability to predict future sales requirements. Besides that, the tool introduces user input variables that allow for creating different future scenarios.

II. Table of contents

III. Reader's guide

The research approach and plan are covered in the first chapter. Chapter 1 introduces the company Forque Enter and explains the problem and the preferred situation. The goal and approach of this research are explained.

The second subject, the analysis of the current situation, is reviewed in Chapter 2. This chapter reviews how the current procedures, workflows, and the data structure operate. The workflow is illustrated with a business process diagram. Based on this chapter and the previous one the theoretical framework is designed.

The fourth subject is in Chapter 4 and is on designing the tool and determining its inputs and outputs, and the variables that are used. Here, several changes are proposed to improve the tool in regards of creating scenarios and restructuring the data.

In Chapter 5 the development of the tool is discussed. Here are the decisions on coding, which data to use and how to use it, and the structuring of data are explained. In this chapter an analysis is performed on the historical data of Forque to determine how the main two inputs of the tool will develop over time.

Chapter 6 is the concluding chapter. Here the design of the dashboard is evaluated, recommendations are given to the company, and the research questions are answered.

IV. List of abbreviations

BPM	Business process modelling
CB	Customer base or regular clients
ERP	Enterprise resource planning
HRM	Human resource management
KPI	Key performance indicator
MPSM	Managerial problem-solving method
NB	New base or new clients
OLS	Ordinary least squares
TFN	Triangular fuzzy numbers
UTAUT	Unified Theory of Acceptance and Use of Technology

1. Introduction

1.1. Company context

Forque is a small data consultant company in Twente. They employ about 45 people and can provide services in several different business aspects or types of industry. This ranges from education to automotive. The service they provide can vary from collecting and analysing data or optimising business processes. For this, they use AFAS software and Microsoft Power BI. Besides this, they sell user licenses for their own in-house developed software. The business philosophy is that the order is always tailored to the customer. The company prides itself that the employees are curious, daring to explore new problems, and most of all are enthusiastic. The stance to the customer is based upon trust in their opinion so when solutions do not work for the customer the company is honest and will be transparent about it.

As mentioned above, the types of industry the company works with are many. They have different teams that are specialised in the customer branch. The company believes that by having experts the customers will be more satisfied and helped in a clear and to the point way.

1.2. Problem identification

Forque has a yearly planning of their marketing actions, in which they define a required number of leads. Leads are meetings with customers that could result in a work order, e.g., designing and implementing data collection for car workshops. This number of leads is generated at the beginning of the year by hand, based on the data of previous years and on their turnover targets. The number is what helps the divisions to distribute their attention and money. Due to ever changing circumstances and workforce changes, it is not constant throughout the year. Adapting to these unexpected events was possible for a long time as the company was small enough. Now the company wants to grow and is no longer able to do the calculations manually. They also experience events in a daily fashion that forces them to reconsider the number of leads necessary.

The company has a database and excel sheets to do the calculation. The problem is that there is no structural approach to make it an efficient analysis. The collection of data and its analysis will be examined during the phases of analysis of the problem and solution generation.

The action problem is: *Forque is currently planning the number of required leads annually by hand, which they now want to do daily in an automated way.*

1.3. Problem cluster

In the problem cluster in Figure 1 different problems. This is done to find the underlying problem that should solve the action problem. This underlying problem is the core problem. To make sure no unnecessary items were in the cluster, the pneumonia rule of Hans Heerkens and Arnold van Winden [1] was used. This rule states that it is not important to state how the company came to the current way of working. This would not help solving the problem as the cause of a problem is more important for the problem cluster. The arrow points from the cause to the effect. Core problems can be the ones with the longest chain, but sometimes the problem could be a sidestep that helps clarify or give structure to the other problems.

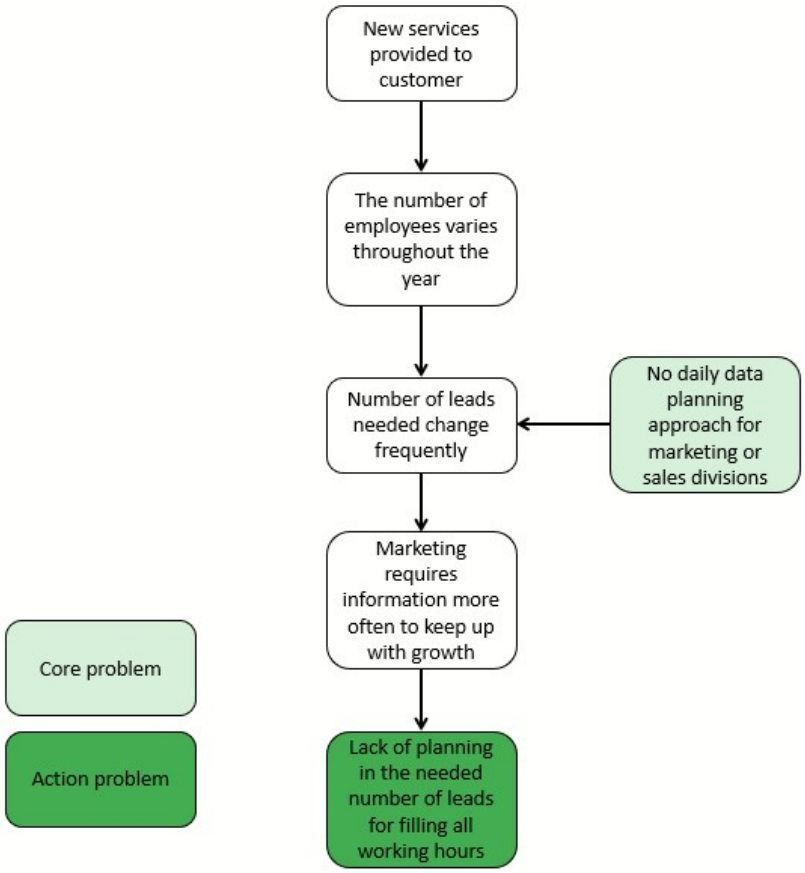

Figure 1: Problem cluster

1.4. Core problem

In Figure 1, the steps tracing back through the problems are shown. There are two problems that could be considered the core problem. They are the new services provided and no daily planning approach.

The services provided means that the company offers a range of products or services that is too big and requires more staff then expected during the introduction of the new service. Controlling the new services is difficult as it is the way the company makes revenue and is outside the scope of this research. As Hans Heerkens [1] wrote that only a problem that can be influenced must be chosen as core problem.

The problem that can be solved or influenced is that they do not have a daily planning approach to their lead procurement. This can be influenced as a tool or approach can be designed and implemented into their daily operations. Thus, the core problem is:

The company board of directors of Forque should be able to plan on a daily basis the number of leads, but now they plan the number of leads once a year.

The *norm* that they want to achieve is that they plan the number of leads they need to have every day, but the *reality* is that they know it once a year. The company had no clear structural approach to give the information to the marketing and sales divisions. This caused uncertainty as they had less

steering than is needed. A tool that would calculate the number on a daily basis would give the insight that is required. Now, the marketing and sales division will be able to plan the number of leads based upon daily input making uncertainties part of the overall daily strategy.

1.5. Motivation and approach

The reason why the tool is important is that it helps Forque to use workforce resources efficiently and to work closer to capacity. The board of directors now are able to make decisions on a daily basis regarding workforce, task distribution, and lead procurement. Besides these numbers, they also gain insight in how the marketing division is doing.

The approach for the thesis is that of the Managerial Problem Solving Method (MPSM) which is developed by Hans Heerkens and Arnold van Winden [1]. This method exists out of the following 6 stages:

1. Problem identification
2. Problem solving approach
3. Solution generation
4. Solution choice
5. Solution implementation
6. Solution evaluation

1.6. Research questions

During the research in the thesis, questions arise as a lack of knowledge about topics become clear. These questions are known as knowledge problems. The research cycle is used to provide structure to answering knowledge problems. The research cycle used will be based upon the Solving Managerial Problems Systematically by Hans Heerkens and Arnold van Winden [1].
In this thesis there is one main research question:

Which planning method is useful to help Forque plan their number of leads using their data on the company?

To answer this research question multiple sub research questions are asked for each stage of the MPSM. The second and fifth stage will be left out as they have little to do with the research questions. The second stage is about exploring the problem before the project started. The fifth stage is where the problem owner makes a decision regarding the solution. This is not a stage where a research question is relevant or useful. These two stages are more about what the approach is to the project. The latter is done by the company management and is not part of the execution of the project. This is where the requirements the company made are evaluated.

1.6.1. Problem identification

The question for this chapter will be an expansion on the company context. The goal of this to get a better understanding of the company and the process behind the marketing department. These questions are important to the core problem as before a problem in a department can be tackles, the department and it actions must first be looked into. The focus is researching what has been done before can prevent problems arising throughout the project.

The question is:

- How does Forque gather the information on leads?

This question is about exploring the core problem and the way Forque functions. To understand the core problem better, the planning methods Forque uses must be investigated. This includes what is happening now and what is limiting a future solution or change. Employees will be asked if there was an approach designed that was not implemented.
The next step in this stage is gaining an understanding about the current approach of the marketing and sales departments. For gaining the insights interviews with employees exploring their operations using a semi-structured interview will be done. The employees who will be interviewed are from the sales division and the IT staff. The analysis of the operations will exist out of reviewing the data in the excel file the company has composed and their other data on daily operations of the marketing and sales divisions. The outcome of the interviews will be compared to each other and the results will be used for the solution generation stage. The reason why it will be done in this way is that the expertise of the company will be used. This will help solving the problem for this specific company.

1.6.2. Problem analysis

After the company analysis, the literature review can start. In this part planning methods that are relevant to the problem Forque has are reviewed. Besides the methods description, an analysis of usefulness will be made.

The question is:

- What are methods available for managing and planning workforce?

The question will be approached by way of literature review. The motivation for this question is that understanding the nature of the problem can help developing the solution in a more tailored manner. The method of answering the question is an exploratory literature review. The goal of the question is to figure out what nature the problem has. The approach for the selection of the literature is in Appendix 1.

1.6.3. Solution generation

This stage is about designing the tool. To do this some, aspects have to be researched. They are inputs and outputs, how to design a tool in a user-friendly way, and variables.

The question is:

- How can a planning tool be designed for the case of Forque?

In this question, the solution and its criteria are explored. First the criteria are explored to understand what the solution should be able to do. This will be done together with the management and scientific literature. The literature will be used to find the minimum requirements of a solution to work. When the criteria are known the planning methods are examined. They will be analysed and gathered. Insights into the structure, positive and negative sides, and techniques will be summarised for a decision regarding the functions of the tool and the design of it.

1.6.4. Solution implementation

This stage is about developing and implementing the tool. This includes the coding. The tool has to be implemented into the company's daily system and linked to all relevant datasets.

The question is:

- How is the tool connected to Forque's data structure?

This question is about the development of the tool and how it should be done. The development is the coding and determining the inputs for the planning tool. There are various aspects that should be researched as the program that will be used for developing the dashboard has their own characteristics that should be taken into account. When this is done, a verification will be performed by using historical data to test the functionality and accuracy. Then, this tool still has to be implemented into the data structure of the company. The tool is made in the same program the company uses. Linking all databases to the tool makes it operational. Interviewing the employees who work daily with the program will answer the questions. The interview is a semi-structured one. Besides asking the employees on how the program functions, the online tutorials on the tool will be watched.

1.6.5. Solution evaluation

When the tool is implemented, it can be reviewed and evaluated what other uses the tool might have. This begins with the evaluation of the design and functioning.

The question is:

- How can the tool be assessed?

For this question, a review will be done on the functioning of the tool. The output will be compared to what the company is doing and see the difference. There will also be a survey on the user friendliness of the tool, the design, and recommendations of future improvement. For this method, a scientific approach will be used. Before the questionnaire will be handed out, a workshop will be held explaining what the tool does and how it functions. When the workshop is done, the attendees will have the opportunity to ask questions. After the questions, the questionnaire will be handed out. The outcome of this will be analysed. Insights of this evaluation will be summarised and put into a table to make it visible what the outcomes are.

1.7. Scope

The main focus of this thesis is on the development of a tool. This tool must calculate the number of leads required for Forque. The thesis focuses on the sales process of Forque in specific. Ideally, the company would know the exact number of leads required for the entire next year. This is not possible as the state of the company changes due to human resource management or unexpected events during the negotiation process. The study focuses only on the teams for who the sales division is negotiating. They are the teams: development, ERP, HRM 1 and 2, management, and specials. If there is time left, new features can be developed to improve the insights gained from the data analysis.

1.8. Deliverables

This thesis will have as deliverables:

- The dashboard for the calculation of number of leads needed
- A manual on how to use the dashboard
- A report on the design and development of the dashboard and the process with the recommendations on further development

2. Company analysis

In this chapter, the company will be analysed. This to gather information on how the company is currently forecasting and planning their lead procurement. In Chapter 2.1 the current gathering and processing of data is explained. In Chapter 2.2 the process of acquiring a lead is explored by way of flowchart. In Chapter 2.3 the evaluation of their planning is done. In Chapter 2.4 the wishes and requirements of a possible solution by the management are laid out. These chapters answer the sub research questions and in doing so answer the question*: How does Forque gather the information on leads?*

2.1. Current procedure

Currently, once a year a preliminary forecast is made on how to distribute the working hours amongst the employees and teams. An employee has a file that explains their productivity, distribution of hours throughout the year, and which job they perform. Each employee is allocated a number of hours that they are expected to work. That number does not include sick leave, Holiday leave, and other kinds of absence. All data is put into in an excel file for an overview. This is not a file that is linked to a dataset. The file is filled in manually. In November this excel file is changed to include the possible hiring of new employees. The quitting of employees or longer sick leaves are not used to calculate the forecast. This can be a problem as the hours are budgeted in the financial statement so other members of the team will have to work extra hours to fill the lack. Once a year, a check is made to see if the prognosis is still fitting the current status of the company and if the budget has to be adjusted.

This calculation of the prognosis above is done by hand in an excel file. This file is not linked to the company data structure. The prognosis is a snapshot of the company on a day. The fact that it is a snapshot in time does not help the reliability of the calculation. Another problem is that the distribution of work is not analysed in this prognosis. An employee can have too much work or too little as there is no way of knowing which team will get a work order.

The schedule of an employee is filled with working on projects. The amount of work a project provides depends on the amount of money it generates. A rule of thumb in the company is that 1 hour of work is equivalent to 100 euros. A different approach is that each different function in the company has its own calculation for the amount of work a project provides. Multiple calculations help because different functions have different wages. Increasing the number of calculations would complicate the planning tool as most distribution of hours worked is done per team and not per employee.

A dashboard was made to show the status and characteristics of each project per team. The workload, employees, and hours filled are analysed and displayed in graphs. The dashboard uses the real time data. The problem with this dashboard is that broad assumptions are made regarding working hours, sick and holiday leave, and how teams are connected to project groups. These assumptions do not represent the reality. Consequently, the dashboard shows data that is not usable to make decisions on. Besides that, there are missing steps in the calculations as it acts like an employee works 100% of their time for clients, which is not happening in reality.

There is a problem that is not detectable in the data. Sometimes customers want extra work outside of the initial deal done. This is not taken into consideration of the prognosis as adding extra work happens after the project is started. Adding work that was not in the project forecast means that there is more work that is not written down for scheduling. A guideline the company has made for

this is that up to 2 more days do not have to be written down in the forecast quotation. More work however has to be written down as a new project so it can be planned.

The data set for the calculations is small as the company started collecting it in a structured way a year ago. This implies that some correlations and irregularities can be harder to predict.

2.1.1. Actual organisation of the data

The company collects data on their employees' performances, project characteristics, and invoices. Forque stores it and uses it to make decisions. The decisions regarding the sales division are mostly based on the data of projects and employees. The projects have certain characteristics. The characteristics are the amount of money it generates, the average sojourn time in the company, and which team handles it. There is a distinction in their projects between which kind of customer the project belongs to. The two kinds are *customer base* and *new base*. The former are customers who have already worked before with Forque and place a new order. The new base are new clients.

In the data files the project negations are written down as follows. First the name of the customer and the company responsible are mentioned. After that, the company team is linked. Secondly the project negotiation details are determined. These are:

- **Order starting date**
- **Order expected end date**
- **Order end date for complete projects**
- **Forecasted amount paid**
- **Expected amount paid**
- **Actual amount paid**
- **Result of the negotiations**

The last one is important for filtering on what data to use. Being able to filter helps using the data that is relevant for predicting the average value and hours needed. The data of successful negotiations is used because the company wants to only use the data of negotiations that are successful.

The employees have the number of planned working hours as main feature in the currently existing dashboard. A workday has a number of hours that have to be filled. The dashboard shows how many hours still have to be filled in. In the data structure behind the dashboard, the hours are written down for every day for every year for every person. For computing the number of hours for the year that needs to be filled, a summation is done for the team and each team member.

The goal of one of the pages of that dashboard is showing what each divisions or project groups average value of a project is. This helps Forque to get an indication of what the average amount of hours is a project can take. The calculation for the average amount of hours is based on the value of the orders. The way it is calculated is taking the average order value per project group and divide that by 100. This is shown in Table 1.

Project group	Average throughput time	Total order value over period	Average order value over period	Average number of hours per order
Group 1	10	€55.000	€4.700	47
Group 2	14	€45.000	€5.000	50
Group 3	12	€50.000	€4.000	40
Total	**12**	**€50.000**	**€4.570**	**46**

Table 1: order value and calculation of hours per order per project group

A second table, depicted in Table 2, on the same page calculates the number of working hours per team. Other values in this table are the number of hours for sick leave and days off, hours for CB, and hours for NB. The number of hours for CB is an estimation by a manager and not forecasted using the data of previous orders. The number of hours for NB is calculated by subtracting the number of hours for CB from the hours to fill. There is a problem with the values in Table 2 as the hours for days off, sick leave, and the customer base are all static hard coded values that did not rely on data. This is not a problem for days off, but for sick leave it can present difficulties. Some employees get sick for a longer period of time and this must then be solved by their co-workers. Others might have no sick leave so it can balance out. This will be reviewed in a later section.

Team	Total amount of roster hours	Holiday and Sick leave hours	Hours to fill	CB hours no forecast	Number of hours NB	Number of orders year
Team 1	3.050	360	2.690	500	2.190	47
Team 2	4.500	450	4.050	0	4.050	81
Team 3	3.900	400	3.500	640	2.860	72
Total	**11.450**	**1.210**	**10.240**	**1.140**	**9.100**	**200**

Table 2: Scheduled hours and order calculation

The number of hours for customer base is not easily forecasted based on data. The company has some customers that order every year a certain amount of work. The prediction for this kind of customers is doable, but others might not be as loyal or consistent. By analysing the data some of the customer base projects can be mapped out into value ranges or must be treated like a new customer to make the dataset more complete or have the calculations make fewer assumptions. The last column of Table 2 calculates the number of projects the company needs for the entire year. This number is calculated using the average value so the number might be lower or higher depending on the projects they get.

This forecast is made by using estimates from experience of previous years for the customer base, data of previous orders for the new base, and calculations to tie them and the hours to fill together. The accuracy is only checked to see if they are still on course to get their target of the number of orders. This is done by the marketing and sales employees to see if things have to be changed. The model itself is not checked, only the output is.

2.1.2. Variables

For the projects' characteristics there are three groups the variables belong to. The first one is the project order related ones. The second is the employee related variables. Lastly there is the forecast variables that are similar to the project variables. In Table 3, they are explained.

Project order	
Project team	For each type of project there is a specialised project group that handles these projects
Monetary worth	The project is valued for a certain amount
Working hours	The number of hours that is allocated to a project is based on the worth of it. By dividing the worth by 100 the hours are calculated
Throughput time	This is calculated by taking the average of the difference between start and end date. It is necessary to determine the difference between the time allocated and the actual time that is used
Success rate	Success rate is about making deals. They call a success a scored project. They keep track of all deals that are on the table and report when they succeed or fail. By counting the number of scored projects and divide it by the total of both scored and failed the success rate is calculated. They also keep track of the likeliness of scoring a deal. This is a second success rate the company uses to determine the likeliness of future deals. This rate is not based on calculations but on experience
New customer	The projects are checked if the customer has an order for the first time or not. This helps differentiate between the two classes the company has of customer base and new base
Employee	
Team	Each employee is connected to a team. A team can have members in various project groups
Productivity	The productivity determines the amount of work an employee can do within a time unit. This also depends on the type of contract
Roster hours	An employee is expected to be available to the company for a number of hours per year. They are called roster hours. These hours include sick leave and holiday leave
Type of contract	The company offers contracts for full time and part time. This influences productivity and the number of roster hours
Start of contract	The company allocates time off per month. The start of the contract influences the amount of leave for the first year. In a similar way the end of a contract determines the same
Time off	The time off is earned per month. This helps the company determine how many hours they have to allocate to new employees who start during the year
Forecasts	
Result	A forecast receives an entry when it is closed to differentiate between forecasts that led to a deal or not. The results are written down as “closed not scored” and “closed and scored”
Potential	As mentioned in the success rate of orders the sales employee can add the change or potential a deal will be made percentage based on experience
Value	During the negotiations a project receives and estimated value that is used to determine the number of hours that should be spend and is used for the monetary examination of the company
Source	To determine the success of marketing, the way a customer reached Forque is written down

Table 3: Variables used in the dashboard

2.2. Process of acquiring leads

A method is used to analyse and visualise the process. The method is called Business Process Modelling Notation. Its goal is to notate the process in such a way that it is readable and understandable for anyone. BPM is chosen to help clarify the multiple occasions where the customer and company interact. Interacting between 2 actors is better shown in a BPM model than a flowchart. This is because BPM allows for different lanes where an agent has their actions, whereas flowcharts is one big action chain that does not represent the different actors in a clear way.

In Figure 2, four major categories are shown that are used in the notation. They are flow objects, connecting objects, swimlanes, and artefacts.

The flow objects are at the centre of any BPM Diagram. Flow objects are events, activities, and gateways. Events are the start of a program, e.g. the start of marketing campaign or a customer that calls a restaurant. Green events are the beginning of a process, and the red ones are the end. Activities are just that, activities. They can be sending an email, filing a statement, or boiling an egg. Gateways are decision points where two or more outcomes are possible, e.g. deciding which contractor to use or accepting a request.

The connecting objects connect the flow objects. There are different kinds of lines to show the different kinds of flows. Sequence flows are within an agent or one pool lane. Message flows connect pools together. Association is when two activities are connected but not in sequence with each other.

Artefacts represent data that is being handed over to another pool or activity. Swim lanes represent an agent who executes activities, e.g. group, customer, or company. They are used to group activities together in groups that are being executed by one agency.

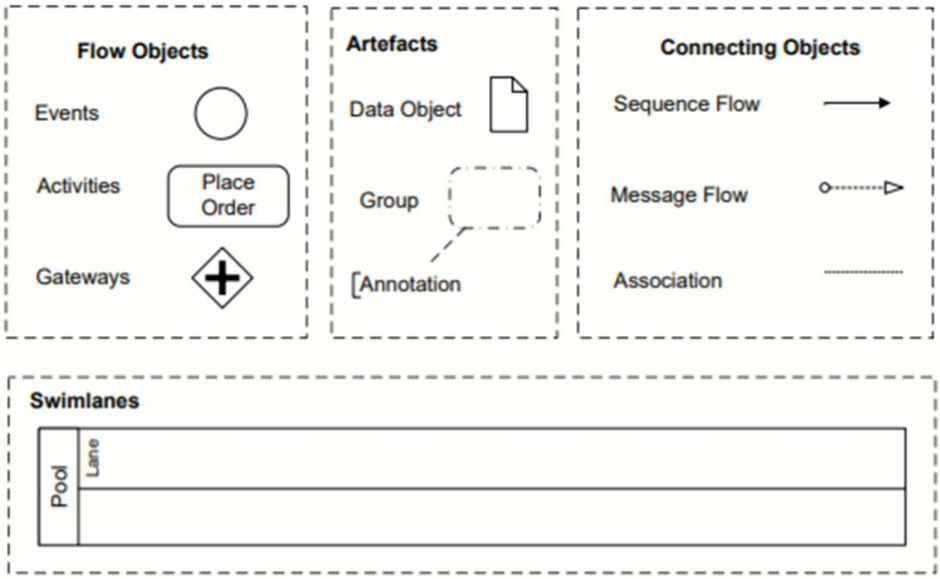

Figure 2: BPMN: categories of elements. Reprinted from "Business Process Management: Concepts, Languages, Architectures" (2nd ed.) by Weske, M. (2012), p. 209 [2]

Now the modelling method is explained, the process itself can be explored. The process in BPM notation is depicted in Figure 3. This is a schematic overview of how the process from marketing to the choice of signing a contract takes place in Forque. Before modelling the process, an interview was held with the sales manager who also checked the model afterwards as validation. The BPM

model will help create an understanding of the process and how the data is linked to each action. Because the process is now modelled, the interactions between the various actions and their corresponding data are now clear.

As mentioned, there are two divisions that are important here for the acquiring of leads. They are the sales and marketing division. The process starts by either marketing publishing a marketing campaign or a customer that calls Forque on their own accord.

When the company is contacted by a possible customer, they make a sales task. This is used for keeping track of the status of the appointment. For example, they can include that an appointment still has to be made. When the customer wants a meeting, the company schedules an appointment with someone from the sales division. During this meeting they talk about what Forque does and how they can help the customer. Here they figure out if they are a match for each other.

If they are a match, Forque makes a document for their dataset called a forecast. The forecast document is used only for administrative purposes. This document written using estimations. It includes approximations on when the project starts, how much time it will take to finish the project, what team should work on the project, and how much money the project will generate. This is not shown to the customer as they only receive a quotation. It has about the same information, but all Forque specific data is left out.

A main piece of information that is included in the forecast but not in the quotation is the potential of the negotiation. This is a success indication the sales representative makes on how likely it is a deal will be made. In the beginning of the process the indication is normally low, but closer to the end it is very likely to be high. There is no underlying method to estimating the success chance as it is purely based on the experience the employee has with negotiations.

Based on the quotation the customer receives, they can continue with the process or stop it. If they continue more details will be added to the forecast and they are also made more precise. Based on negotiated details of the project, Forque proposes a contract that is presented to the customer. When the contract is signed, the team that is linked to the project will take over. At the end of each sales task or negotiation process, an evaluation is written. This is to make sure knowledge is being stored and used for data analysis. One of the results in this evaluation is whether the deal is made. Another important data entry is how the customer met Forque. The data is used by the marketing division to analyse the effectiveness of each campaign. The analysis is then used for specialised targeted marketing adds.

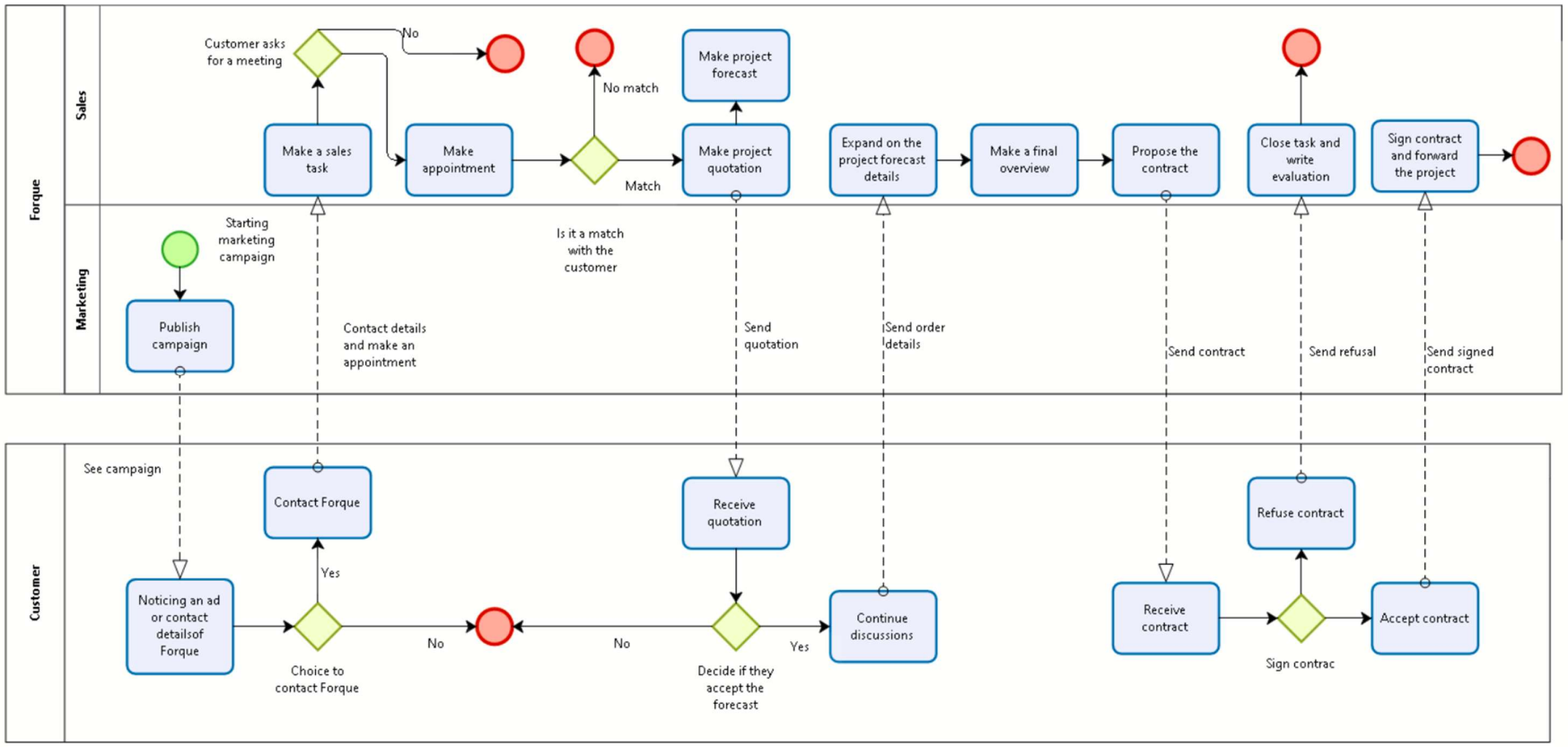

Figure 3: Process flow of lead procurement of Forque

2.3. Key Performance Indicators

The company uses three key performance indicators for the analysis of their lead acquisition. These KPIs will be used in the dashboard to monitor the performance of Forque and marketing and sales divisions. They are listed below and are explained:

- **Number of appointments achieved:** As appointments how a chance of failure, there have to be more appointments than qualitative leads. This number shows if the company is achieving its target to run a sustainable business.
- **Difference between estimated hours for a project and the real hours used:** The sales employees make assumptions on how much time a project will take when they are negotiating the deal. A smaller difference helps the company plan more efficiently and have less problems with idle time or over time.
- **Difference real time taken off and calculated:** This gives insight into how well the assumptions on how many hours of leave an employee has versus how many they use. This is of importance for the calculation as too many or too few leave hours results in not usable number of leads. By revising the calculation in such a way that the difference is minimised, the prediction will be as close to reality as possible.

2.4. Wishes, requirements, assumptions, and limitations

Forque has the following requirements:

- The tool works in real time based on the data the company has on itself.
- It focuses on the teams: development, ERP, HRM, management, and specials.
- The tool must provide insight in the marketing and sales process by reviewing the forecasts, sales actions, previous orders, and current staffing.

Forque has the following wishes:

- Forque can alter inputs or parameters to use it as a kind of simulation tool. This with the goal of doing case studies.
- Forque wants more insight into the Sales process on how the orders perform. This in terms of how much time is spent on an order and how many sick leave and holiday leave hours are used.
- A design wish is that the colour scheme of the dashboard uses darker colours.

The thesis has the following limitations:

- The time that this project should take is 10 weeks. This means that the scope of this project should be limited. The scope therefore only focuses on the calculation of the leads.
- The company only started recording the data in a structural way since 2018. This means that the data set is limited which can be troublesome for the evaluation.
- The customer base makes offers but we do not know how big they are. The same hold true for unexpected extensions of a project. This limits the ability to produce an amount that is true. To solve this ranges and variations are considered.

Assumption	Reason
Employee uses only their holiday leave in that year	Carrying over holiday hours from one year to the next increases the variability of the model exponentially. It is outside the scope of this project as it is more about HR
Employee uses all holiday and sick leave in the year	The flexibility to have a program consider all combinations of usage of leave usage is too big to program. It is better to use maximum as it produces the minimum amount of leads required to fill all working hours
Employees who have an end-date for their contract will not be rehired or given a new contract next year unless specifically put into the system	Employees are scheduled to work in the database based on their contract. Changing the rules here allows for assumptions not grounded in reality
All new employees start on the same day	Giving the fictional new employees each different start dates would overcomplicate the dashboard and would need a calculation per new employee. This flexibility is not possible in this coding language
All customer base orders do not have a pre-calculation	To distinguish between CB orders and the regular NB it has been split this way. The true division between CB and NB in the database is only in whether it has a pre-calculation
External employees are not taken into consideration	These employees have different contracts and hours than the regular employees. This was not compatible and outside the scope of the project as it was about HR
The invoiced hours are what is really spent on a project	The invoice is what is earned for the company. The hours reflect best how the revenue is made
Employees work the full hours they are expected	
Average order sizes per team do not change from current year to next year	
Productivity or invoiceable hours consultant 90%, senior consultant 80%, project manager 70%	This is the average productivity management expects from their employees

Table 4: Assumptions model

The goal of the thesis is to determine the number of leads needed for Forque. The decision was made that the following teams will be analysed: *development, ERP 1, HRM 1 and 2, management, and specials.* This is because the sales division is only looking for contracts for these teams.

Another decision regarding the main function was that the new sick and holiday leave calculation needs to be made. Forque expects that a revision of the current calculation can make the dashboard closer to reality.

The last decision regarding the main function was that the quotation must be used as the measurable variable for leads. This was done as sales action are recorded properly in the database.

The change was accepted by the management as it is a logical replacement. They would like to see the statistical analysis in the CB hours as to know how their customer retention is doing.

The decisions regarding the new features are as follows. The management came with an idea that is to analyse the price an hour was sold for in the past. The prices are analysed to see if the growth followed inflation or that reductions were taking place in the quotations. The other ideas are showing the orders in different classes of sizes, simulation of different numbers of employees, customer base hours based on previous data, and influencing success rates. These ideas were ranked as equally important as the insights are new to the management. These would help gain information on how the sales division is doing. Especially the comparison between how many hours are expected to be used on a project and actual usage was to be done as soon as possible. This comparison would help their scheduling of who does what project when. The management decided they value the possibility to influence the variables of the calculation. With the influenceable variables, Forque can sketch different scenarios based on economic situations, development, and the company's status.

Lastly, Forque wants a dashboard with variables they can influence. The management wants to see the effect it has on the number of orders, forecasts, and quotations. After that, they wanted separate reports on the sales status of both current and future forecast status, a comparison between expected and reality on hours spend on projects, and a view on usage of sick and holiday hours. Forque wants to see these reports in a new way of designing that is not their standard design strategy. The project has freedom in colour and background choices.

2.5. Conclusion

Summarising, the company keeps data on their orders and employees' rosters. By way of manual calculation, they used to make a prediction on how many leads they needed to fill the roster hours. The company transitioned a few years ago to a more structured data approach, now working with real time data. There exists a process flow to follow the lead from contact point to order. During this process the order characteristics are written down that are later used for new forecasting. The conclusion or answer to the question is that Forque gathers the information on leads by way of data analysis. This is currently a model that is not dynamic and uses the data on past orders, assumptions, and employee roster hours.

3. Literature review

In this section, the various ways how to schedule and forecast the workforce and possible solutions to the problem are reviewed. During the literature review, several methods and classifications were found. In Chapter 3.1, the problem is examined in more detail. In Chapter 3.2, concepts that are used in finding a solution are explained. In Chapter 3.3, various problems are reviewed to determine what is similar to the problem that Forque has. In Chapter 3.4, approaches are researched and evaluated to figure out what is applicable to the problem. In Chapter 3.5, the literature review is summarised, and an answer is given to the section's research question: *What method is the most suitable to the core problem for the situation at Forque?*

The modelling of the workforce helps the company management in making decisions. It maps out the difference between the current state and future needs.

3.1. Related decision problems

In this section decision problems are examined. To determine what kind of problem Forque has several problem types will be reviewed. This will help in determining the right approach and modelling method.

3.1.1. Workforce scheduling problem

Workforce scheduling concerns mobilising the workforce to perform activities that are connected to their job. To make effective use of employees, their skills have to be taken into consideration. Similar to general assignment problem, the workforce is the subject. The difference is that they are not assigned to tasks. The employees are jobs are treated as the inputs and they workers themselves are. An example of the workforce scheduling problem is ordering the engineers for a telecom company for the duration of the task, distance travelled to the problem, and the type of problem. There are characteristics that are important for scheduling [3]. They are:

- **Time window**: it can be flexible or tight depending on the task. This is the time an employee can take to complete a task.
- **Skills and qualifications**: this determines which employee can be assigned to a task. This can filter out some employees making the scheduling more restrictive. There are two approaches to describing the skill sets of employees. The first is that all employees are the same. This cannot be used in many cases and can be expensive for the company. The second is that there are levels in the skills. This is found commonly in healthcare or specialised industry, described by Cordeau et al. (2010) [4].
- **Connected activities**: this is about tasks or activities that depend on each other. This can cause a restriction that one task has to be completed before the other can start. Another way this can be seen is that they have to start at the same time.
- **Teaming**: this is sometimes required as a task cannot be done by one employee. This depends on the nature of a task explored by Li et al. (2005)[5]. However, if the team does not change, the team can be regarded as one unit or one person for modelling reasons.

A disadvantage is that the model does not allow for external input or output in a comprehensive way. For example, entering and exiting the model is difficult to depict. The option to give feedback to the system how certain elements react to moving in the model is not doable. This does not represent the problem Forque has as employees enter and leave the company.

3.1.2. Resource allocation problem

Resource allocation problems concern the optimal allocation of an amount of resources to different tasks whilst minimising the cost. This can be done by determining the optimal order of tasks, the best combinations of assigning, and scheduling resources used during the task. An example of a resource allocation problem can be budgeting the production of computers. By allocating the time and money to various stages of the production, time and money can be saved.

The objective function can be minimising time or costs, those are two common ones. The underlying goal of the problem is to make optimal use of limited resources. This all happens with a trade-off between time necessary for finishing a task and the quantity of resources needed for the it. It is common to have a resource curve that shows the quantity of resources available over a period of time. Here two variations exist about the assumptions regarding the usage of resources and the possibility of handling more than one task at the same time. It is based on the situation that is being modelled so resources can be reusable or not and overlap of tasks is allowed or not [7].

The trade-off does not have to be used if the situation or process allows it. In a nontrade-off case the resources and time are fixed. This situation makes finding a solution a simple task, as the task that can start the earliest is put at the beginning of the sequence and the others later based on their possibility to start. In the trade-off case things are not as simple. Here the time and resources are used to determine when the task could be completed at the earliest opportunity.

This problem can be used for a variety of cases like health care capacity planning [8], human resource allocation [9], and every process where resources have to be allocated to jobs.

The inputs are the costs that are associated with production like resources or the machine hours and the benefits of the product. This can be modelled in several ways, like linear, nonlinear, or time intervals. The model is about combinations of the levels of the inputs with constraints that help simulate the real world.

The benefit is that the method allows for smart use of the resources within a timeframe which helps eliminate downtime and wastes of time and material. This is especially true when the resources are limited and do not allow for failure or change.

3.1.3. General assignment problem

General assignment problems concern calculating the optimal distribution of agents to tasks [10]. This differs from the resource allocation process, as here people are assigned to tasks and not resources like money or raw materials. People can have different impacts on the process, e.g. some might be experts on assembling a chair and will use less time than a beginner. The difference between the people is of importance in the general assignment problem. The difference between the classical assignment problem and the general one is that the classical method restricts one agent to one task, whereas the general method allows one agent to have multiple tasks. The general assignment problem is similar to the knapsack or bin packing problem where the goal is to find an optimal way to pack as many useful objects into a container [11]. The agents can be assigned to a task for a cost. The task makes a profit that can depend on the combination of agent and task. Each agent is allowed a budget of time and cannot spend more than that. The decision variables here are the decision to link an agent to a task. This is often depicted by a 1, yes, or a 0, no.

The reason why it exist is that it is applicable to several important decisions like scheduling, supply chain management, and routing problems. This drives researchers to find exact or heuristic algorithms to solve these challenges [12].

The inputs of this model are the costs associated with a combination of agent(s) and task. These costs can be costs of production or transportation, but also the benefit of production like revenue or profit. The outcome is the minimal cost or maximum profit with the combination of variables that produces this outcome [10].

The downside is that all costs have to be known to allow for a functional assignment schedule. On the other side an advantage is that in the process early feasible solutions are found. Another benefit is that an imbalance between tasks and agents can be solved by creating dummy agents or allow an agent to do multiple tasks. The downsides are not relevant for Forque as the thesis does not look at each induvial employee, but the general assignment problem is of use for the thesis. The combinations between employees and tasks is influencing the total number of orders in the way it determines the forecast and data-analysis. The calculation using the various types of employees is determined by in which division each employee works, how quick they work or their productivity, and how much leave each employee has.

3.1.4. Project scheduling problem

Project scheduling problems concern the optimisation of project duration, allocation of project resources, and project cost [13]. It models, sequences, and schedules the project's activities based on constraints from resources or precedence. Instead of using the resources as determining factor like the resource allocation problem, project scheduling uses the projects as the determining factor. An example is determining the order of assembling a table. Some steps of the assembly must be done in order whilst others could be done simultaneously. Determining the right order of execution is what project scheduling does.

The reason why this problem was first researched was that projects tended to take too much time and go over budget. The reasons that were cited then were mostly about inadequate employees, poor planning, and misalignment [14]. During investigation of these problems, it was found that poor planning in the early stages is an important factor for failure. The factors for success were grouped into four area: related to the project, related to the manager and team, related to organisation, and related to external factors [15].

The inputs of the model are the resources that are required for the process to function. The output is generally the most optimal distribution of them together with how much time it will take to perform all tasks in the process. The single project scheduling problem is concerned with determining precedence of actions, resource-feasibility, and minimising duration. The multi project scheduling problem consists of several projects. The projects get their resources from a shared pool. Multiple projects can be added together to form a bigger single project. The resource constrained scheduling problem considers the limited access to resources, known project durations, and resource requests. An activity needs a certain amount of resources. There are the non-renewable resources, like raw materials, and the renewable ones, like manpower [16].

3.2. Forecasting approaches

The core problem of the thesis has forecasting aspects. It is important to review the approaches that exist and use the best approach for the forecast modelling technique. To make an educated decision, the concepts of a forecasting model are first explained. Afterwards the various approaches are reviewed to differentiate them.

3.2.1. Concepts

To understand what characteristics and concepts a forecasting model has, the following two sections are dedicated to increase the understanding of this modelling kind. A section covers how the development of a model is done.

3.2.1.1. Forecasting model

A forecasting model is used in situations when someone wants to know the development of something in the future based on what is currently known. Forecasting is used in many decision-making activities. The forecasting model has to be accurate to be useful. Therefore, the statistical approach is the most commonly used method.

To determine what the forecast of Forque's problem is several factors that must be reviewed. The first is the time horizon of the forecast. There are three types that are explained below:

- **Short term:** This is used for scheduling personnel, transportation, and production. To do this demand forecasts are used.
- **Medium term:** This is used to define resource requirements in the future to make decisions on buying raw materials, hiring practises, or acquisition of production machinery.
- **Long term:** This is used for strategic planning.

The next two factors are the subject and what it is used for. The subject is what is forecasted, e.g. production levels or number of employees needed. The subject needs to be clear and measurable. The question what it is used for has to do with limitations of the model. There are the questions if it has to be for every product sold, if it is for every team or location, and the time-interval required as in weekly, monthly, or yearly [17].

When examining the dataset to look for the necessary data it is important to look for patterns in the data. There are two common patterns, seasonality and trend. Seasonality is that there is a pattern that is related to a period in a year or month. For example, the average daily temperature depends on the period in the year. In the summer months it is higher than in winter months. This is the natural seasonality and the other is from human decisions. An example of human seasonality is the number of ice creams sold per year. This is higher during the summer months and lower when it is winter. Trend has to do with the growth or decline of a series over time. This can be the growth of a company over time as an example [18].

3.2.1.2. Steps for designing the model

To design a forecasting model, 5 steps must be taken. This is depicted in Figure 4. Below the steps are explained:

- **Step 1-Problem definition:** Defining the problem is about gaining an understanding of how the forecast will work, be used, who will use it, and how it is a forecasting model.
- **Step 2-Gathering information:** This is about gathering the historical or statistical data. It can be difficult to gather enough data for fitting a good statistical model. This is because old data can have lost its relevance due to changes of the system.
- **Step 3-Preliminary analysis of mathematical model:** This is done by making a graph or any visual representation of the data. This explains trends, seasonality, or cycles in the data. The other aspect that is examined here is the relationships between variables.
- **Step 4-Choosing and using forecast model:** The type of model is decided upon based on the information gained in Step 3. Using this a forecast is made that is used in step 5 for evaluation.

- **Step 5-Evaluation of the model:** The evaluation is done with the use of errors. This measures the difference between reality and the model. With this insight the model can be changed with modifications of the model or its parameters or by human interference.

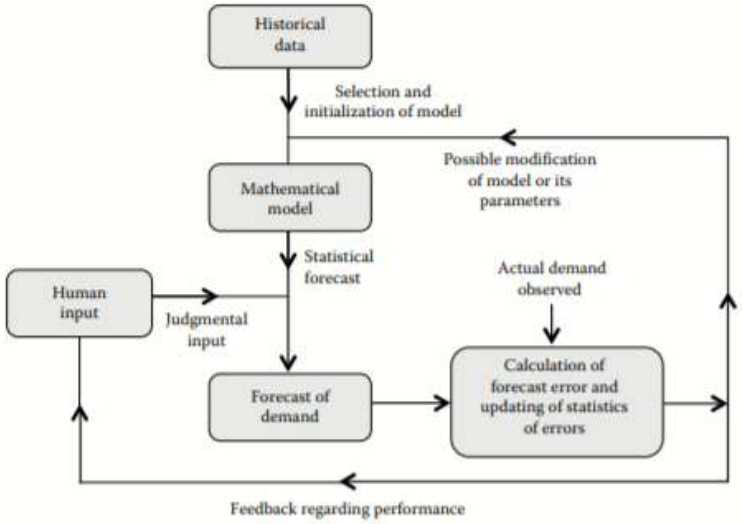

Figure 4: Steps for designing forecasting model. Collected from Inventory and Production Management in Supply Chain (4th ed.) by Silver et al. (2017). P74.[18]

3.2.2. Qualitative forecasting

Qualitative methods are an easy to use approach for forecasting the workforce. It is applicable to cases that do not have highly detailed empirical data, because it uses expert opinions, consumer surveys, and discussions [19][20][21]. A particular method that uses expert opinions is the Delphi method that is explained below. This method looks at the organisation and its workings. With that information dependant relationships are established. The data is generated by answering and combining the experts' answers on hypothetical cases.

The Delphi method is built around expert opinions and forming a consensus. The forming of a unanimous opinion is done in numerous iterations. When a round or iteration is done, the results are summarised and used as a starting point for the next round. The experts are given the opportunity to change their opinions based on the new information. This goes on until there is a unanimous answer that might be correct [22][23]. The key word here is might be, as it is still an opinion and not a statistical or mathematical outcome.

The goal of the Delphi method is to shape the unknown parameters of models. Delphi methods can be used for several problems like trend forecasting, forecasting objectives, attributes, strategies, planning, etc.

The advantage of the Delphi method is that it can be used in many cases as it does not need a lot of data. A different advantage is that the Delphi method prevents hearing one opinion by using a group of an individual as the opinions are pooled anonymously [24][25]. These advantages are of no use in the case of Forque. There is enough data available and expert opinions contribute little to designing a forecasting and data-analysis tool. The downsides are that it is hinging on the experts. This has the possibility to produce an incorrect answer. A second downside is that it consumes expensive time that the managerial staff might not have [26]. The downsides are relevant to Forque's case as the management has little time available.

3.2.3. Quantitative forecasting

Quantitative models are of a numerical and statistical nature. They use high quality data to make predictions. In the modern world the value of a numerical answer is high. Therefore, these quantitative methods are widely used, e.g. for managerial problems or economic problems. There are categories in the approaches as some look for an explanation of a relation between variables whereas others are only interested in the answer. The latter is time series. It looks for the dependency on time and uses past values to forecast. The former is regression models and uses patterns and relationships to predict the future.

There are some limitations to these data driven models and their ability to accurately forecast. Because the data-analysis models rely on information of today or from the past, they do well in predicting stable enterprises, but fail when unexpected events happen that shake up the company or economy as Gordon wrote [27]. He argued that these data-analysis models should include four elements of relationships between input and output, *shape, threshold, interaction, and lags.* These elements should help recognising the existence of *chaos.* Gordon wrote that the deterministic relations formed in these models cannot reflection human society to the point of certainty. The four elements are described as followed:

- **Shape** is the mathematical form of the relation. This ranges from a simple straight line to complex algebra.
- **Threshold** are cut-off points where the relationship changes. The impact an input factor has is changes. This can manifest in different ways. For example, an input factor has the different effects in different value intervals, or it can have no effect until the factor reaches a certain value.
- **Interactions** are the causal relationships between two or more factors. The more relationships the more difficult it is to make a model.
- **Lag** happens when an output is not affected by the current input but by an earlier value.

The last challenge Gordon argues is that of *Chaos* or *the butterfly effect*. A seemingly unimportant event can have enormous consequences [27].

Concluding this all, it is difficult to give certainty to a prediction as there are many forces at work that we do not know, understand, or even experience.

3.2.3.1. Time series

Time series is an approach that uses historical data to make predictions. This data are observations that are measured through time. It can be measured continuously or discretely. When using the continuous model, it is common that the observed variable is a continuous variable recorded constantly. When the data points are analysed, it is common to sample the data series at intervals of equal size. This transforms the data into a discrete time series without losing information. The discrete time series can be made in three different ways [28]:

- **Sampled** from a continuous time series.
- **Aggregated** data over a period of time.
- **Inherent** discrete series.

The time intervals that are used for the discrete time series are normally recorded at equal time intervals. The data can be aggregated over series or across time. It is worth keeping in mind that the data is not normally independent between successive observations [28].

There are 4 main objectives for time series analysis that are listed and explained below

- **Description:** This is used to describe the data by way of statistics and graphical representation. Doing this in a time plot is particularly insightful.
- **Modelling:** There are two types of models the *univariate* that only uses past values and the *multivariate* that uses present and past data entries. This helps explaining the variation in another series when comparing present and past values.
- **Forecasting:** This is done to make an estimate of future value of the data series and is commonly called forecasting or prediction. There is the *steady-state* assumption that states future values are like the past ones, and the *what-if* assumption that is used to figure out what effect changes to policy variables have.
- **Control:** This is linked to the *what-if* assumption of forecasting as it gives control to the analyst to act regarding the process.

The methods that are used for this approach are ARIMA, Box-Jenkins, and exponential smoothing, and Markov model. Time series cannot be used when a persistent conflict does not happen or when no trends are known. Time series is used when historical arrival rates and other managerial policies change instead of using discrete-event simulation [29].

3.2.3.2. Causal or regression modelling

Regression models can measure relationships between variables where correlation is present. It considers uncertain parameters and approximate them to predict the future. These parameters inputs are based on historical data and fitted in a similar way like above with least squares or maximum likelihood. The aspects of a regression or statistical model is that the output is explored through variance, standard deviation, expected values, etc. With the regression, cause-and-effect can be found between variables or dependency. These models can be applied to a variety of cases as it is a general method.

There are many variations on the regression models that are listed below:

- **Linear**
- **Nonlinear**
- **Simple**
- **Weighted**

The main downside of regression modelling is that it can only use an isolated observation and not data over a long period of time. Regression models are generally difficult to make dynamic and thus unexpected events can discard the outcomes. Difficult obstacles in such a model can be inter dependency. This occurs when the input variables are not independent of each other. This can also be depicted in a recursive model.

To prevent interdepency, 3 conditions are specified that must be met to say that two variables have a causal relationship [30]. They are explained below:

- **Covariation:** There must be a mathematical link between the two variables to show dependency.
- **Time order:** There must be an order of time that states that one happens before the other.
- **No other causes:** The variables should only depend on each other and not any outside of the relation.

3.2.4. Discussion

Due to the nature of Forque's assignment and what the variables are that will be tested causal methods are the best choice. This is because the model is a built on a series of variables that are

dependable on each other. The data is known, and the data set is big enough to perform data-analysis without the need for creating data via expert opinions. Forque's problem has an analysis part that is founded in understanding the relations between the inputs and outputs. One of the inputs, the number of hours CB, is currently a predetermined value. This value could be changed to be dependant on previous data. The goal of the analysis of this data is to find a trend in the number of hours sold to CB. The goal here is to specifically find if it is dependant on any variable, like number of customers or time of year. Though the main goal is to find a dependency, a time series analysis will also be performed to see if a more general trend takes place.

A similar reason is used for the test that will be done on the number of hours that is spend on CB clients and the number of leave hours. It is expected that the previous number of hours spend on CB orders is related to the number of hours spend last year. This is to determine if there is growth or decline in the order amounts. This can be tested if the growth is in the form of a linear, exponential, logarithmic or any other form. This is important for the model as it is one of the 3 main inputs of the hour's calculation and the only one that is not dependant on the number of employees. This makes it relevant to ascertain what the nature is of this input.

3.3. Causal methods

Several methods will be examined on how to fit a model. The first part of the section is based on the assumptions and requirements that a model has to adhere to. The assumptions are requirements are listed and explained below:

- **Linearity** is a main rule in a linear regression model. It states that the value of a dependent variable is the outcome of a linear function of all independent variables. These variables can be modified by a coefficient and powers are also allowed. The latter is only for multicollinearity [30].
- **Causality** as mentioned in Chapter 3.2.3.2 has to do with that one variable causes another variable to change. This is never researched using a causal method but has to be stated by the model designer. This assumption can be based on theory or experience. For this to be true there are 3 requirements as stated in Chapter 3.2.3.2 [30].
- **Independence** has to do with the error terms of the model. These should be independent of each other as it would otherwise imply there is an independent variable that is left out.
- **Homoscedasticity** is another name for constant variance. The variance of the errors should be constant with regards to the independent variables or their response [30].
- **Normality** is another test for the error terms. The errors that are found should follow the Normal distribution. This is not a requirement of the tool but used for confidence intervals [30].
- Scatterplots can be used to find linearity or dependency in a quick and visual way [30].

Ordinary least squares is a model fitting method for regressions. It uses the minimisation of the sum of the squared error terms to find the model description. It works by changing the value of the coefficients as the Y is the outcome and X is the input and are thus fixed values. The formulation of the inputs and outpus is shown in the formula below. There are 3 classical assumptions about the errors:

- Linear relationship between independent and dependent variable. In other words, the expected value of an error value is 0.
- Errors have constant variance, see Homoscedasticity.
- Errors are uncorrelated.

The Gauss-Markov rule states the OLS estimators are the *Best Linear Unbiased Estimator* if the assumptions are met [31].

Recursive and non-recursive

Generalised linear models is a method that is derived from the general linear model. This method allows for variables that are not normally distributed or continuous. A linear model is described as:

$$y_i = \sum_{j=0}^{n} x_{ij}\beta_j + \varepsilon_i$$

Here y is the outcome and is the dependant variable. X symbolises the independent variables and beta the coefficient. The epsilon is the error term. This is under the assumptions mentioned above.

The generalised model relaxes these assumptions and allows for many forms of distribution, though the linearity assumptions still stands. The change in that assumption is that it is now regarding a linear predictor.

Robustness is the performance on how the model acts when the data is not normally distributed or has errors. The robustness is thus about the resistance of a statistic to errors in the outcome. When the model is robust and assumptions are met for a large enough part, the model will be efficient and of a usable level with a small bias. One instance of the troublemakers that robustness solves is extreme outliers. This will be filtered out to make it more realistic. This method of filtering is an alternative to OLS. Another common method is Robust Standard Errors (RSE). RSE is used to take patterns or error dependency into account. This is not seen as fitting the robust regression definition as it does not discount unusual data entries [31]. To assess the robustness of an estimator 4 concepts are used: efficiency, breakdown point, and influence function. They are explained below:

- **Efficiency:** This is about the minimum possible variance of an estimator in comparison to the actual variance. When the dataset is large, a goal can be to work as efficient and quickly as possible. An estimator is efficient if the ratio between the minimum and actual is one as that is the lowest possible variance.
- **Breakdown point:** This is the global degree of resistance of an independent variable. It deals with the outliers at either tail end of a distribution without producing a random result. The breakdown should not discount more than 50% of the data entries. This would mean that a majority of the data is not used.
- **Influence function:** This measures the local resistance. Another name is a sensitivity curve. This would limit the influence one data entry has on the model [31].

3.4. Assessment of forecast error

Testing the model is an important step in the validation part of the process. There are several methods that can be used to check how well the model predicts reality and if the variables are constructed in the right way. They are listed and explained below:

- **Goodness-of-fit and R-Squared:** The goodness of fit is the test that states how well the model describes the observations. The level of fit helps validating the relation between the independent and dependent variables. It measures the deviations of the observations and the predictions. A common way to measure the deviations is by using the R-Squared technique. This method scores the model using a range of 0-1 with 1 being the best. It subtracts the division of the variance in residuals from the variance of the model. The R-Squared explains the proportion of the variation in the dependant variable based on the

model's independent variable [32]. There are 2 problems with this method. The first is if more predictors are added to the model, the R-Squared will go up as it is more likely to be spot on, even with coefficients that do not represent reality. The second problem is that too many predictors create more random noise and thus throw of the R-Squared calculations [32].

- **Regression standard error:** This is related to the error term of the model. The RSE uses the variance of the error term or the variance of the dependant variable that is unexplainable. The goal is to minimise the value of variance. The range of this value is from 1 up to infinity [32].
- **Pearson's Chi-Squared test:** This test is another goodness-of-fit test. The test one uses the sum of the differences between the observations and expected outcomes divided by the number of observations. The goal is to minimise the outcome of the division as the difference between observations and predictions should be as near to 0 as possible. The resulting number can be compared with the Chi-Squared distribution to ascertain the goodness-of-fit with a number of degrees of freedom [32].

3.5. Conclusion

We will first review the findings of the answers in the previous paragraphs to answer the research question *What method is the most suitable to the core problem for the situation at Forque?* The answers help determine the most similar problem and then the most suitable solution method. The core problem of the thesis that Forque wants to know how many leads are needed to fill the roster hours of the company. This roster is compiled by the hours of every employee. Those hours are then divided over the divisions where the employee belongs to. The hours are filled by the orders Forque gets. The number of hours allocated to an order is determined by the worth of the order. The conclusion of this section is that the core problem is of a General assignment problem nature and the solution method that will be used is regression models that will be tested on how well they fit the data.

Based on the statements of the sections in Chapter 3.1, the problem that Forque has is most similar to the General assignment problem of all problems in Chapter 3.1. This is based upon the fact that it is about distributing agents to tasks. This is like what happens in the company as an order is assigned to a certain project team but that has then to be assigned to an employee who works in that project team. This problem type is also suitable for scheduling, which is a major part of the problem of this project. The assigning of orders does not happen in the calculation performed. The calculation concerns the teams as a group but does allocate hours to teams. The tool calculates the optimal allocation of orders to a team, so they have as little idle time as possible. The minimisation of idle time is best solved with the General assignment problem approach.

From the solution methods mentioned in Chapter 3.2 and Chapter 3.3 the one that is best for the project problem is of a forecasting nature. This is because the company wants to know what happens in the future and how they should plan for it. The most fitting one is regression modelling based on that cause-and-effect are present in data and the model. This is true as variables are connected and used in calculations based on previous data steps. This means that changing data somewhere has an effect on another variable. This dependency between variables is a classic characteristic of the causal or regression analysis. The data-analysis will focus on the influence of variables on the outcome, the number of leads required.

4. Solution designing

In this section, the data-analysis tool is designed. Here the inputs and outputs are chosen, new features are introduced, the visualisation of the dashboard is discussed. In Chapter 4.1, the data-analysis tool is discussed. In Chapter 4.2, more specific details of the solutions are explained. In Chapter 4.3, the specific solution parts are reviewed upon their cost of time and the various combinations are written down what is achievable with the limitations of this project. These recommendations will be further discussed in the next section where the choice has to be made what is implemented. In Chapter 4.5, a literature review on designing a dashboard is performed. This section has the research question: *How can a data-analysis tool be designed for the case of Forque?*

4.1. General solution

The main goal of this thesis is to provide insight into the lead calculation and have the outcome of the formula show how many leads are needed to fill the "hours to fill" values from Table 2. This will be done in a form that is similar to time series. The tool in the form of a dashboard will show the outputs per month and provide an overview of the insights of the data-analysis.

4.1.1. The working tool

The dashboard has a number of aspects that it should have to be considered useful and working. This comes from the requirements Forque has and that the dashboard must be able to perform the calculation and data analysis. Forque wants the model to be working in real time. The program PowerBI that the company uses works this way as it is driven by their data storage. Entering or changing a value in the data will be represented in PowerBI.

The model will use as few hard-coded values as possible. This will be achieved by using data, extracted data, or data the has been put through a calculation. This will make sure the model is future-proof and as close to reality as possible.

4.1.2. Input and output

There are a number of inputs that are chosen to be used in the model. These inputs are the data sets the company has on their timetables, tracking of orders, and on the employees. The input tables that will be used are classified in these three types. In the paragraphs below it will be explained in more detail what it contains. The reason why these inputs are chosen is that from the experience of the employees they are relevant, or that they are relevant based on the details of the calculation explained in Chapter 2.1. More inputs are chosen because they can increase insights into various stages of the calculations or processes of the sales division.

- The first type is timetables. This type is linked to the employees but the focus here is on the roster hours. The data sets contain every day an employee would work and for what amount of time. These entries are per employee per year as to make it easy for data analysis. The next data set that is of importance is the absence table. This table includes the hours an employee is allowed for days off like holidays and the number of hours they are allowed for sick leave. This again is allocated per employee per year. The last data set that is used is the data from the timetable application the company uses. In this the employees note down how many hours they worked on a task.
- The second type is orders. There are numerous tables that are connected to this type but only the ones that are of importance to this project are mentioned here. The first one is on the forecasts by the sales division as mentioned in Chapter 2.2. In this data set the leads are collected with whether the deal was made, how much the deal is worth, when the negotiations started and ended, how likely it was to close the deal, and which project group

the lead belongs to. The second data set keeps track of the financials of the order. This table is used to see if a customer is new or not. The last table is the pre-calculation. This is about the number of hours that is predicted to be spend on a lead.

- The third type is employees. As mentioned above this is closely linked to the timetables. The focus here is on the employee attributes. The first data set that is used is the employees table. This table has the information which staff member is a member of which team. The next data set is about which teams exist. The last data set is one that is mentioned before, it is the forecast data set. In this table the project groups are defined.

There are some data sets that do not fall in any of these categories but are necessary to the model. This data is needed as it provides structure to other data, e.g. the data can be grouped by date or month. The calendar data table is added. This helps the dashboard organise data per date. Another is the data set on their customers. As orders and forecasts are linked via customer identification table is another structural addition.

The main output is the number of leads required to fill the roster hours. This will be subdivided into monthly values, so the company gets insight in possible trends. Furthermore, the steps of the calculations will be shown in the output as well to make the process transparent and understandable. The values that are shown are divided into the main goal of this project and the featural outputs that are listed below in Chapter 4.2.

4.1.3. Calculations

As mentioned in Chapter 4.1.1, calculations can be done on the data within PowerBI. These formulas calculate, make a logical statement, or a mixture of both. These calculations are mentioned here as they are the backbone of calculating the numbers of orders required in a year. This leads to the number of leads via the data on sales performances. The calculations are listed and explained below:

- **Total order value:** This is the total value of all orders in a year. This is calculated by taking the sum of every deal made in an entire year.
- **Average order value:** This is the average order value in a year this is necessary to determine the number of hours.
- **Average number of hours per order:** This is the division of the average number of hours per order by the value of a working hours.
- **Total hours:** This calculation is about the total roster hours in the roster of a year for. This does not include sick leave and holiday days but does include public holidays. It sums over all hours in the data set concerning the roster hours of one year.
- **Hours unavailable:** This calculation is about the total hours of both sick leave and holiday hours. The summation is over the hours in the absence dataset.
- **Hours to fill:** This is the difference between the total hours and hours unavailable. This is the number of hours employees are expected to be productive.
- **Hours New Base:** This the number of hours that is spent on new customers. It is calculated by the subtraction of the hours for the Customer Base from the Hours to fill.
- **Number of Orders:** This is the number of orders of new customers the company needs to fill every employees' productive hours. It is calculated by the division of the hours New Base by the average numbers of hours per order.
- **Number of forecasts closed:** This is the sum of every forecast that is closed where it does not matter if it was successful or not.
- **Number of forecasts closed and scored:** This is the sum over every closed and scored forecast. This can be used to determine the success rate of a forecast.

4.2. New features

In this section, we introduce new additional features of the tool. They are for restructuring the data, creating new scenarios for forecasting, and combining data in a new way. The features are about the underlying parts of the tool that provide more information about the data. The characteristics are shown per feature in Table 5.

Feature	Type	Purpose
Classes of order size	Restructuring the dataset	Giving new insights in how certain sizes of orders perform
Simulation of different number of employees	User changeable variable	Creating new scenarios for the future
Feedback of the number of orders current	Combining data in a new way	Provide insight into how the sales division currently performs
Flexible hour of work value	User changeable variable	Creating new scenarios for the future
CB hours based on data	Combining data in a new way	Eliminating assumptions
Influence success rates	User changeable variable	Creating new scenarios for the future
Sick leave comparison	Combining data in a new way	Check for outliers amongst sick leave and check if the future calculations are accurate

Table 5: Feature characteristics

4.2.1. Classes of order sizes

Projects are classified in different order size classes based in their workload. This classification helps the sales department gaining knowledge on how the orders with different sizes impact the number of leads they have to get. For example, if the department scores 10 large orders, they might need 20 small orders to fill the same number of hours.

The company has a classification of the size of the orders, but the classifications are not reflected in the data itself. There is a way it can be extrapolated from the other data as it is based on the number of hours worked on a project. This variable is reported in the data. The classification are as follows:

- **Short** projects take 2-3 working days or 16-24 hours.
- **Medium** projects take 3-10 working days or 25-80 hours.
- **Long** projects take more than 10 working days or 80+ hours.

By grouping the projects together based on the size, we get 3 new averages on expected hours worked and real hours worked on project orders that are more specific. The Sales department gets new insights in the 3 classifications that they can use to focus their efforts.

4.2.2. Simulation of different numbers of employees

Forque would like to have a feature that simulates the impact of different numbers of employees on the number of leads needed. This feature helps provide a better forecast that the management can manipulate to show various scenarios. The management would benefit from the feature as it would explain the feasibility of the growth they want to simulate.This feature works with formulas to keep the data set clean. Entering simulated employees and removing them from the data sets is difficult and could corrupt it.

To simulate adding an employee the variables that concern employees must be investigated to figure out what their impact is and how they are structured. There are currently 3 variables that have an influence on the leads. They are listed and explained below:

- **Roster hours:** The company currently calculates the roster hour by writing down every workday and the number of hours worked on that day for then to sum over them with the year as filter. Adding an employee would be simulated by using the total hours of a year that a currently employed employee would be rostered. The public holidays are taken into account here, so they do not appear in the holiday leave calculation. There is a rule that the company has for new employees. Forque assumes that the new employee needs to be worked in and settle for 2 months until they are productive. This can be represented by not registering those hours. There is a problem with how to calculate how long this period exactly is. This is because not every 2 consecutive months have the same number of days. The assumption is made that the length is 61 days as this is the most common combination. The last part of the formula for roster hours is the starting date of the new employee. The assumption that every new employee starts on the 1st of January is false. To simulate this an input is made where the user can enter the starting date. This starting date is subtracted from the result of the subtracting of the year and 2 months introduction period. Summarising the roster hours calculation is thus the roster hours per year minus the 2 months settling in and the starting date.
- **Sick leave:** The company has the rule that every employee is allowed to have 3 sick leave days per year or 24 hours.
- **Holiday leave:** For an employee who is with the company for the entire year they get 26 working days off. To make this fair for people who start during the year, the company has a rule that states that days off are earned per month.

Leaving employees can be handled in roughly the same way. The differences are that the 2-month settling in period does not have to be taken into account and that the roster hours minus the leave data have to be subtracted from the total roster hours. To complete the simulation the sick leave and holiday days have to added back on the “hours to fill” column.

This feature can be used as an advisory tool rather than an input. This however is outside the scope of determining the number of leads as this change would focus on human resource management.

4.2.3. Feedback on the number of orders current year

A feature will be designed that counts the number of hours filled during the current year and returns how many more are needed for the rest of the year. The feature will help the sales division keep track of their work.

This feature will check the start date of the order. The check looks if the start date is in the current year. Then the number of hours that is assigned to an order is added to the total sum of filled in working hours of the current year. Subtracting this sum from the hours that should be worked in the rest of this year, returns the hours that are left to fill. The feature can be specialised in a way that it returns an overview of each team’s planned working hours, which would help the Sales division focus their efforts to market to specific types of customers.

4.2.4. Flexible hour of work value

Currently, the assumption or rule of thumb the company uses is that an hour’s work is worth 100 euros. The management and the current dashboard now see it as a hard value that is not changed.

This value of an hour's work might not always be the same due to either increased wages, inflation, or any other reason that changes the value of work or money. To compensate this changing value an input field is designed where the dashboard user enters the value of one hour of work. This feature makes the model future-proof and can help simulation of possible cases.

4.2.5. Customer Base hours based on data

The current dashboard only calculates the averages for their new base, not customer base. Forque does keep track of whether a customer is new or not in one data set, but that is not used in analysis. The data that is currently used in the dashboard for calculating the number of hours for customer base orders are values that are based on assumptions. Combining the available data sets, it is possible to calculate the hours that are spend on customer base orders. The new values are thus based on data instead of assumptions. There is a problem, some of the customer base orders do not follow the process-flow in Chapter 2.2 but are accepted right away. This data skipping the process can have an influence on the average value, as it might be lower than it should be due to missing data. The data that does get stored is the information on invoices so the data might not be as incomplete. The missing data or not consistent storing data could be a reason why it might not be possible to work this feature out fully but give it as a recommendation.

4.2.6. Influence success rate

Now with the social lockdown, clients could be less likely to accept an offer. This would affect the number of leads needed as the company would now need more leads to achieve the same number of orders. To simulate this, a slider will be implemented that influences the success rate of a lead, quotation, or forecast. The main purpose of the feature is to help simulate possible future scenarios. To calculate the current success rate, certain rules must be set. The success rate could be per team, project team, or order. The choice on rules will be made together with team Specials to understand which values are most useful. The slider will then add to or subtract from the success rate to represent the expectations

4.2.7. Sick leave comparison

The employees are allowed 3 days of sick leave or 24 hours. There are employees who do not use the full amount whereas others use more. There will be a comparison between expected sick leave and real sick leave with groupings based on employee, team, and the company. This will help the company determine if they have to change the sick leave hours as they are a loss of revenue potential.

4.3. Achievable features and recommendations

To ascertain whether the proposed features were doable and achievable in the time available for the thesis, the specialised PowerBI team of Forque was interviewed.

The team specials members were asked about their opinion on these features. They were approving of them. Their main concern was to support why which data was used and that for some features not all data might be present. This is especially the case for the feature on the Customer Base hours. They believed everything should be possible within the timeframe of the thesis and that the features should be ranked on importance or priority. This recommendation will be put forward to the sales manager for the decisions on which features are important.

4.4. Tool framework

As seen in the conclusion of Chapter 3.5, the tool is not of a true forecasting nature. It has some forecasting aspects as looking into the future for calculating the number of leads needed next year, but this is in essence only a simple calculation that exists out of adding, subtracting, and division

without any unknows. It also occurs in a linear way that has a one-to-one relationship. Here coefficients are not logical to use due to the one-to-one relationship. This is another reason why it is not a classical example of a forecasting model. To help structure the dashboard, the following sections are written that explain the approach and function.

4.4.1. Function

The tool has as function calculating the number of leads needed for a certain number of employees. Each employee has a number of hours that they have to scheduled and a number of hours they do not. The latter is scheduled by the employees themselves. Apart from that, the company has a number of clients who do not follow the normal sales procedure but make deals directly without the negotiation. These customer orders do not have a pre-calculation. The company is not interested in calculating the number of this customer base clients and only wants to know the number of new clients needed.

This analysis is important for the scheme that is used. Due to this dashboard not being one of the traditional methodologies or frameworks, that part is left out. The framework that is used is the following. In Chapter 2.1.1, the table is shown that states the calculationon how the number of orders is calculated, not predicted. To calculate the number of orders, the average number of hours that is spent on an order is calculated for each team and is then used as the denominator and the number of hours NB as numerator. This computes the number of orders and the next step is use the success rate of forecasts and sales quotations to compute these.

Linear regression is the only aspect of methodology used from Chapter 3. The regression is used on the number of hours spent on CB orders. This is the only variable of the calculation on how many orders are needed next year that can be considered unknown. The other parts that are integral to the calculation is the number of employees and the number of employees added. These are known and are deterministic in their effect. The number of hours spent on CB is analysed to see if a trend is present in how it develops year to year. Time-series is not used as the matter researched is how the customers react year to year, for example if they order more than the previous year. To analyse the data, the hours that are sold to a client in a year are aggregated and compared between the years 2018 and 2019. These years are the only two complete datasets of an entire year. By comparing the numbers of hours sold to each client a possible trend can be discovered. This trend will then be used in the calculations.

4.4.2. Approach

The data that will be used dates back to 2018. This was the first year that Forque recorded its data in the structure it still uses now. This small data set does pose the problem that was mentioned in Chapter 2.4, as there is not much to compare or use to validate the model concerning CB hours.

The calculation stays mostly the same as the one in Chapter 2.1. A newt step is determining how many hours should be spend on NB clients. This is done by following the findings of the analysis of CB hours and use the sum of all hours spent last year and possible coefficient and subtract that of the hours to fill. This is done by looking at all orders where a client has more than one order placed at the company. Then the assumption is made that such a client does not have a pre-calculation as the goal is to know the orders that happen without the whole sales process. Another new step is the introduction of productivity. The hours to fill is multiplied by a productivity level that is between 0 and 1. The outcome will be called “hours to sell”.

The success rates are calculated by using the data on failed and achieved negotiations. The classes of size require the pre-calculation, which in turn is made by the sales employees. The feedback on orders uses the start date of an order and then how many are started in a specific year.

4.4.3. Transforming the process into the dashboard

To determine how the process can be mapped into the tool, Chapters 2.1 and 2.2 and Figure 3 are used to describe the process. The process starts with a sales action where a customer contacts Forque. This part of the process can be divided into what is the current status and what the future needs are. The parts are relevant to all teams and detailing this will be needed. These two parts can be done on two pages where the steps are visualised in graphs or in tables.

The next step in the process is the calculating. This is the step between making contact and closing the deal. This is depicted in the data structure as mentioned in Table 3 under forecast. The result of the forecast, the value of an order, and the success rate are of importance to this step. The process begins when the negotiation starts and with it the throughput time. During this process, the sales employee fills in the data on how the talks are coming along. They decide on which team the customer's order belongs to, how much hours it will take, and when it starts. When the negotiations are over, a deal is made or not. This outcome is depicted as the result. All forecast data can be mapped out by graphs and numerical depictions.

The process of calculating how many orders are needed also requires calculations. This was described in Chapter 2.1. This can be mapped in the tool by way of tables and calculations on the data sets. This provides for the opportunity to implement the features of Chapter 4.2.

4.5. Dashboard design

The design of the dashboard on how it operates and looks like is important for the user. It is important that it is easy to use and logical in how it is built. On this topic the following aspects are of importance for the design stage:

- **Task conformance:** According to Kristoffersen (2008) [33] task conformance is about how well a system helps the user doing their tasks. He mentions two aspects of the conformance, namely completeness and adequacy. The former is about how much the systems covers the different facets of the programmed task. The latter is about how the task is modelled in a manner the user thinks is right
- **Consistency**: This concerns that the user does not have to think about different words, actions or other situations where the same thing is meant. This prevents confusion. Gallitz (1992) states that consistency can be reached by abiding by the following:
    - *Similar functions should be performed in the same way*
    - *The system should look, react, and operate in the same way in its entirety. Similar aspects should be designed, used, and operated in a similar way*
    - *The same action should end in the same result*
    - *The function of an aspect should not change*
- **Simplicity**: According to [34] the dashboard should be designed in a way that does not bombard the user with information. The tasks that are done on a regular basis should be designed in a simple way and, if possible, provide adequate shortcuts. Galitz (1992) states the following three principles on simplicity:
    - *Design common tasks in a simple way even if the uncommon tasks are made more difficult*
    - *Offer default settings*

    - *Hide aspects until the user needs them. Provide the important information and common features in a highly visible way and hide the less regular used functions*
- **Familiarity:** This is about the natural and logical order of the dashboard. The design should be in accordance with the terms and concepts used by the user or organisation. Galitz (1992) [35] suggests three standards to assure the familiarity to the user:
    - *Use familiar terms and concepts that are well-known by the user*
    - *Keep the model as close to the order of the process as possible. It provides for a natural and familiar environment that the user can use instinctively*
    - *Use metaphors from the real-world*

4.6. Data analysis

The data on the spending of CB clients is analysed to see if a trend is present to say if they increase their order size over the years. In Appendix 2 are the results of this analysis. The first thing that is apparent is that there are a number of orders whose size have increased with a considerable amount. On the other side there are also several clients who decreased their order sizes. To find a mathematical trend several functions were used to predict future values. The test measure R-squared however is in each case lower than 0.2 which concludes that none of the function types fit the data. Therefore, an average value of the last year is good enough for the predictions.

The data on the order size growth over the last two years shows that there is no specific growth. The number of outliers has increased but the bulk of the data stays about the same. The test value R-squared is never higher than 0.1 which lets us conclude that no specific function is useful. The increase of outliers and other values does show an increase and together with the company statements of the last two years, it can be said that the coefficient of 1.2 will be used to simulate the growth of the company.

4.7. Conclusion

Summarising, the design of the dashboard the case of Forque can be done in 2 parts. The first one is about the main objective of this thesis, namely calculating the number of leads needed per year. This calculation is done by designing a dashboard that works in real time with the input from Forque's databases. These data inputs are classified into three categories timetables, orders, and employees. Each category focuses on a different aspect of the calculation and works together by linking the datasets through foreign and primary keys. Having determined the inputs, the calculations were explored to figure out which are needed for the main objective of calculating and analysing the leads needed. When all this was done, grouping based on sales, forecasting, and information was done to design the pages of the dashboard.

The second part of designing the tool was adding new features to gain deeper insight into the data sets. To this end 7 new features were designed. The features have different uses and natures that are explained in Table 5. They are listed below:

- **Classes of order size**
- **More and fewer employees**
- **Feedback on the number of orders current year**
- **Flexible worth of an hours**
- **Customer Base hours based on data**
- **Influence success rate**
- **Sick leave comparison**

The team that is specialised in PowerBI and the data of Forque agreed that all new features are doable in the timeframe of the thesis. The only limitation can be that the required data does not exist. This is mainly true for the CB hours analysis as this often does not follow standard documenting procedures.

5. Implementation

In this section, the forecasting tool is structured and implemented. In Chapter 5.2, the data sets are reviewed and structured. In Chapter 5.3, the complete tool is described. This section has the research question: *How is the tool connected to Forque's data structure?*

5.1. Requirements of the model and usage

The requirements of the tool can be split in two ways, the functioning of the model and the appearance. The functioning requirements were that the model calculates the leads, calculate based on data of forecasts, quotations, orders, and employees, and that the tool can predict the future. The appearance requirement was that the tool has a dark theme.

The functioning on the calculation of leads is executed in the following way. The calculation is made on the basis of calculating the number of hours rostered, unavailable, to sell, and NB.

The tool will be used by the management team and the sales division. For them it is important to have clear values and be able to spot trends in the data that indicate growth or decline. It is also important that they have insight in what their current status is in comparison to the target of the year.

The appearance requirements were executed by designing the dashboard based on the research of Chapter 4.5 and the colour scheme the sales manager wanted.

The KPIs are executed in the way of making counters and analyses of the data. The KPI on the achieved appointments is executed by counting the finished appointments. The KPIs on the differences are executed by comparing between the expected calculated value versus the actual hours used.

5.2. Data sets

For Power BI to function, the data sets must be uploaded into the tool to be able to use them. Picking the right data with the right structure is of importance for any data analysis tool. First, the data is reviewed on their relevance and then the data can be transformed into more user-friendly data that can be analysed.

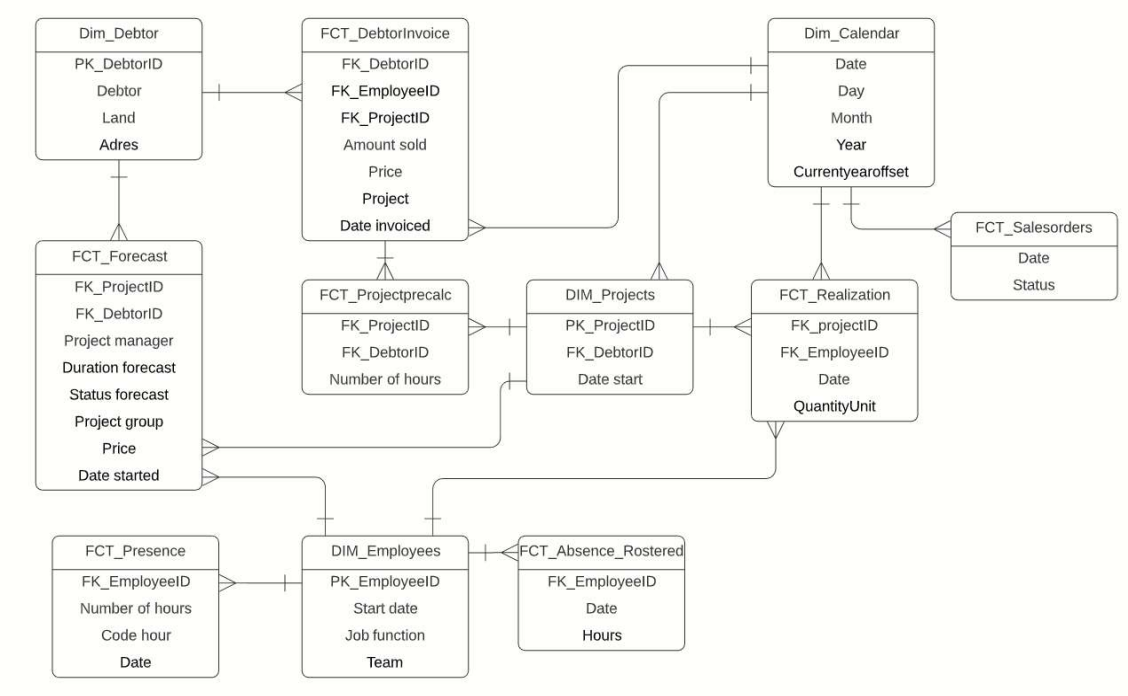

Figure 5: Data model of the tool

In Figure 5 the data model is shown and how each table is related to another. As the figure shows, there 3 primary keys with their corresponding foreign keys. Besides these keys is the employee used as key for the FCT_Forecast table as that makes it possible to link projects to teams. Besides these is the date field used to make sorting based on day, month, or year possible. The three primary keys are listed below:

- **Employee ID**
- **Project ID**
- **Debtor ID**

5.2.1. Relevant data

For the data to be relevant it has to relate to any of the features or the main goal of calculating the leads. We will first discuss the main goal and the related data and secondly review the data for the features.

To determine what is relevant for the tool, the calculations must first be explored. They are reviewed in Chapters 2.1 and 4.1.3. From this we can conclude that the following inputs and their data set are required for the calculations in the order of Chapter 4.4:

- **Sales total needed**
- **Hours total needed**
- **Forecasts total needed**
- **Order total needed**
- **Forecasts closed**
- **Forecasts closed and scored**
- **Forecasts closed and not scored**
- **Order value**

- **Throughput time**
- **Order value over period**
- **Roster hours**
- **Hours unavailable**
- **Customer Base hours**
- **Hours invoiced**

The data that is relevant to the features are now explored. The features are explored in Chapter 4.2. From this we can determine the inputs and their data sets in the order of Chapter 4.2:

- **Sales achieved**
- **Hours achieved**
- **Hours pre-calculation**
- **Hours realisation**
- **Sick leave realisation**

The data thus selected based on its characteristics and use in the calculations of the tools.

5.2.2. Filters used

The filters are on data that is in the datasets but are not needed or are unwanted in the model. By filtering the data out, the datasets are both clean and useful. There are 4 filters that are used. They are listed and explained below:

- Employees working: In the dataset the employees who were working for the company in the past are still present. They have to be filtered out as they give a false idea of the company size. To this end “InDienst”, or employed is set to 1. This means that an employee is still employed at the company.
- Teams: Not all teams have jobs that are filled by customer tasks. There are teams that are concerned with the internal functioning of the company as HR or IT. To this end “Teams” is set to include specials, HRM 1 and 2, development, and ERP 1 and 2.
- Years: The company switched in their data storage in 2018 and grew considerably from 2018 onward. As the data before 2018 is thus incomplete and not representative of the current situation the filter is set to include data from 2018 to today.
- Administration code: The company uses various administration codes to signal where the orders come from. The use “DataDone” as their company code that is relevant to this assignment. So all other administration codes are filtered out.

5.2.3. Restructuring data: Holiday hours

The first implementation of the changes was in the data set of the holiday hours. This was previously based on a single value that was used for every employee for every year. This is not in accordance with the company policy. The way holidays are earned at the company is that an employee receives holiday leave worth about 2.2 working days per month he or she is with the company. This means that the day an employee starts working for the company and when one ends his contract are of importance to the calculation. Based on this the code in Appendix 4 was written. The code checks whether an employee is hired or leaves the company in the year that is being reviewed. If this is false, an employee will receive the full year worth of holiday leave. If it is true, the code checks which month the contract is started or ending as to then accurately calculate the true holiday leave allowed. This calculation is based on the assumption that an employee will use their full amount of holiday leave in that year. There is an arrangement to use leftover hours for the next year, but that

was deemed outside the scope of this project as that has more to do with Human Resource management.

5.3. Forecasting tool

In the following sections the design and the interpretation of the data is discussed. These sections are on why the dashboard looks like as it does and what we can learn from the data. The interpretation of the data is in Appendix 9. The designing of each page is explained in Appendix 10.

5.3.1. Design

For the design of the dashboard the 4 rules of Chapter 4.5 were used. To this end the graphs were designed to show values of similar nature in similar ways. The way this is executed is structuring the panels to go from left to right in the order forecasts, quotations, and orders. This helps the user with structure as the values are in the same section on each panel. With the rule of simplicity information is hidden or aggregated. In the case of sick leave usage, it is possible to view the hours on team basis, but if needed also on employee basis. This simplifies the first view of the dashboard.

The colours were chosen to have a dark theme on the background and the colours in the graphs are highlighted to make them stand out. The dark theme was a requirement of the company. The colours were chosen in cooperation with the sales manager.

6. Evaluation of the tool

This chapter is dedicated to evaluating the dashboard through a workshop and questionnaire. In Chapter 6.1, the method is explained that will be used. In Chapter 6.2, the workshop and approach of taking the questionnaire are detailed. In Chapter 6.3, the results of the questionnaire are reviewed.

6.1. Unified Theory of Acceptance and Use of Technology

To validate the tool in practice, the Unified Theory of Acceptance and Use of Technology (UTAUT) method was used [36]. This method determines the likelihood of success of new technological artifacts, like models, dashboards, or other forms of user-technology. The method, depicted in Figure 6 uses a questionnaire with 6 constructs:

- **Performance expectancy:** The expectation that using the artifact will help the user in improving in their job performance.
- **Effort expectancy:** The expected ease to use the artifact.
- **Social influence:** The perceived notation that the user thinks others believe he/she should use the system. This is not relevant in the Netherlands.
- **Facilitating conditions:** The expectation that organisational and technical infrastructure is implemented to support using the artifact.
- **Behavioural intention:** This is about the intention to work with the artifact and accepting the tool in the daily operations.
- **Use behaviour:** This is about the way the users will work with the artifact.

Besides that, the model uses 4 other variables for understanding the inputs from the interviewee point of view, gender, age, experience, and voluntariness of use [36].

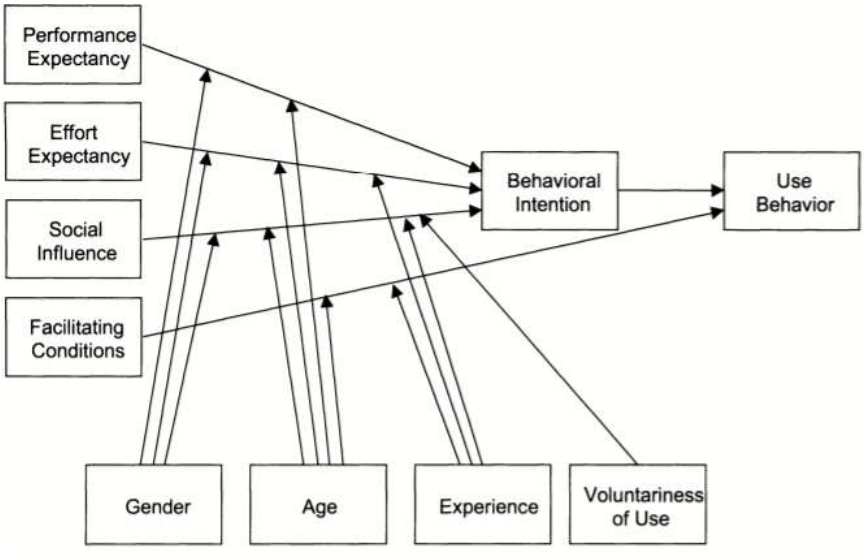

Figure 6: UTAUT model. Collected from Venkatesh, Morris, Davis, and Davis (2003). User Acceptance of Information Technology: Toward a Unified View

6.2. Workshop

To have an evaluation, a workshop was first held to explain what the dashboard does and how it works. After 20 minutes of explaining, a question round was held. When all questions were answered, the participants had time to fill in the questionnaire, which is shown in Appendix 6.

In the questionnaire were two question about the participants' background and knowledge of PowerBI. The questions were constructed with the help of UTAUT and then translated into Dutch. The complete findings are in Appendix 7.

6.2.1. Background

The participants of the workshop are three members of the management team and one data consultant who is also IT staff for the company. The knowledge the participants have on how PowerBI works is varying. Two members mentioned they are very experienced in how PowerBi functions, whilst the other two are not experienced and answered they only use the dashboard and not construct them.

6.3. Evaluation results

For the questionnaire, a five-level Likert scale was used. The answers the participants could give were strongly disagree, disagree, neutral, agree, strongly agree. The strongly disagree had a score of 1 and the strongly agree a score of 5. Thus, values between 1 and 2 are negative feedback, the value 3 means neutral feedback, and 4 till 5 means positive feedback. The results are in Appendix 7.

Question	Minimum value	Maximum value	Mean	Standard deviation
1 PE-1	3	4	3,5	1,73
2 PE-2	2	3	2,75	0,50
3 PE-3	2	3	2,75	0,50
4 PE-4	1	3	2	0,82
5 EE-1	3	4	3,5	0,58
6 EE-2	2	4	3	0,82
7 EE-3	2	4	2,5	1,00
8 EE-4	3	4	3,5	0,58
9 ATT-1	3	5	4,25	0,96
10 ATT-2	3	5	4	0,82
11 ATT-3	3	5	4	0,82
12 ATT-4	3	5	4,25	0,96
13 FC-1	4	5	4,5	0,58
14 FC-2	1	5	3,75	1,89
15FC-3	3	5	4	0,82
16 FC-4	1	5	3,25	1,71
17 SE-1	1	5	2,5	1,91
18 SE-2	3	5	4,25	0,96
19 SE-3	3	4	3,25	0,50
20 SE-4	1	5	3,25	1,71
21 BIU-1	2	5	3,75	1,26
22 BIU-2	2	5	3,75	1,26
23 BIU-3	2	5	3,75	1,26
Average Performance Expectancy	**2**	**3,25**	**2,75**	**0,89**
Average Effort Expectancy	**2,5**	**3,13**	**3,13**	**0,74**
Average Attitude Towards Technology	**3**	**5**	**4,13**	**0,89**

Average Facilitating Conditions	**2,3**	**5**	**3,88**	**1,25**
Average Self-Efficacy	**2**	**4,75**	**3,31**	**1,27**
Average Behavioural Intention of Use	**2**	**5**	**3,75**	**1,26**

Table 6: Evaluation questionnaire results

In Table 6, the summary of the results is shown. The mean value shows the average opinion and the standard deviation shows the difference in opinions between participants. A higher value for the standard deviation means that the participants do not agree in their opinions amongst themselves. The outcomes of the means and standard deviation are shown in Figure 7.

The mean value for the questions has a range of 2 to 4.5. For 5 questions was the mean below 2. For the other 18 questions was the value 3 or higher, with 17 being above 3.25. This shows that there is a positive opinion of the proposed dashboard and that it is generally accepted. The question with the most positive feedback was FC-1 with 4.5 and the most negative feedback was PE-4 with 2. The most positive feedback was on whether the participant had the resources necessary for using the dashboard. An explanation for this could be that they have integrated PowerBI in their daily operations and have adopted them into their tasks. This means they are provided the tools to operate the dashboard on their computers, smart devices, and screens. Besides that, they are provided the data that they input themselves so they can check what they see in the dashboard. The most negative feedback are on expecting a raise in their salary. This might be explained by the fact that the participants are of the management team and that an increase in salary is not based on their personal performance but rather that of the company as a whole. They also do not have a manager above them who could compliment them on their productivity and could increase their salary.

As mentioned before, a standard deviation close to 0 shows agreement between the participant whilst a value higher than 1 shows a high variation. The values for the standard deviation show that there is on average a higher degree of variation in the answers. The highest disagreement can be found in the questions SE-1, FC-2, and PE-1. There is strong disagreement on knowing how to operate the dashboard if the participant is alone and cannot ask for help. This can be explained as two of the participants were involved in the development of the tool and the others were not. To combat this, a detailed manual was written. The last strong disagreement in PE-1 could be explained by the fact that the participants all fulfil a different role within the company so the usefulness of a dashboard designed for the sales division is likely not to be useful for the operational manager.

On the other hand, the participants strongly agree that they could learn how to operate the dashboard if given time. The participants provide positive feedback that the dashboard is a good idea. They strongly agree that the interactions with the dashboard is slightly positive and that it would be easy to learn how to operate it. A reason for this might be that some of the participants were involved in the development or that there has been a knowledge transfer with an employee who is specialised in PowerBI.

Figure 7: Mean and standard deviation summary

Performance expectancy

Looking at the answers of the performance expectancy questions the participants agreed that it is in general not helping their job performance. They agreed that the dashboard helpful for their job, but that it will not increase their productivity, salary, or speed.

The question regarding increasing the speed of performing jobs was answered with 3 neutrals and 1 disagree. The same is seen when asked if the dashboard would help their productivity. This can be caused by the fact that the tool is designed to provide insight into the situation and not performing any tasks. It can also depend on the fact that not every manager has to know as much about the sales situation as they are concerned with their own division. The general concept has a mean of 2.75 and a standard deviation of 0.89 showing a high value of disagreement.

Summarising, the dashboard is considered to be useful for the company management, however the participants are disagreeing that the tool will increase their performance due to the nature of the dashboard.

Effort expectancy

These questions are about how easy it is to operate the tool.

As the data shows the participants are agreeing that the model is easy to use in general as the mean is above 3. The standard deviation of this concept is the lowest as this is the subject most participants agree on. When asked if the dashboard is clear and understandable 2 participants were neutral and 2 agreed. A reason for this might be the previous lack of structural insight into the status of the sales division and their required targets. The second question with most positive feedback, the same as the interaction was clear, was on if the dashboard was easy in learning how to use it. This received 2 agrees and 2 neutrals. This is a good sign that the difficulties can be learned and that the users will be experienced without too much trouble or time. This could also help the other participants who did say that they would need more time to learn the dashboard and its functions.

The most negative feedback was given to the question if the dashboard is easy to use. There were 3 answers disagree and 1 agree. The requirements of the dashboard were difficult to achieve in a

structural easy way and needed a new and difficult way of programming. The reason why 1 of the response was positive could be that the person was involved in the development whilst the other 3 were not.

This was not as problematic as the participants agreed that it is easy to learn how to use the dashboard. As it was mentioned before that there were 2 answers that were positive about learning the dashboard was easy and 2 neutral.

Concluding, the data implies that the dashboard is not easy to use, but the participants answered that the dashboard will be easy to learn how to use it and be familiar with the tool. This is based upon their answers.

Attitude towards technology

This concept is about what the participants attitude is towards the dashboard. With a mean value of 4.125 it is the most positive feedback given of any concept. All participants agreed that the dashboard was a good project. This concept is also the only one were all participants agreed that the dashboard is positive with no score below 3.

The participants agreed on the fact the dashboard is a good idea and like working with it. These two questions scored the highest with both having an average value of 4.25 but a reasonably high standard deviation of 0.96. These two have the highest standard deviation. The most positive feedback came from the question if the dashboard is a good idea. There were 2 responses with strongly agree, 1 with agree, and 1 with neutral. This shows that there was a need for gaining insight into the data that the dashboard visualises and that the participants are enthusiastic about the way the dashboard is working. The other question received the same results. This might be explained by the fact that the managers can gain insight in an interactive way that allows for an entertaining form of data analysis with many options they can chose or change. The interactive part of the future calculations might be a reason for this.

It is interesting that since the questions on how the participants liked working with the dashboard, the score on if the dashboard makes work more interesting scored lower. This might be because the dashboard is perceived to be difficult to use or that the participants who have to work on maintaining and developing the dashboard further do not use the dashboard in a similar way. The lower scored questions did share a lower standard deviation. This could be explained in a way that they work with multiple different dashboards and thus are less interested in the dashboard as concept but do like the visualisation of the data.

Concluding, the attitude toward the dashboard is perceived as positive based on the answers to the questions. The participants are of the opinion the dashboard is a good tool and enjoy using it.

Facilitating conditions

This concept is on the facilitating conditions that help the user working with the dashboard. The questions were on how well the dashboard would fit in with other systems used by the company and whether the participants feel confident they have what they need to use the dashboard. The mean value of this section of questions is 3.88 and a standard deviation of 1.25. Overall, the participants were agreeing they can use the dashboard in a way they feel comfortable with.

The results of the data show that there is a large disagreement amongst the participants if they think they can use the dashboard and if there are people who can help them. This is mainly caused by 1 strongly disagree entry in 2 questions where most answers were very positive. This can be caused in two ways as the first is that the management team has different functions for each member and that

they thus possess little knowledge on the sales division or that they do not know what the data means as they lack knowledge if the process of the dashboard.

On the question if the participants had the required resources the participants were agreeing. 2 answered with strongly agree and the others agree. This can be explained by the fact that the company uses PowerBI for various different dashboards and is thus integrated into their daily operations. A second reason why this could be is that they are using the input data themselves or put the data into the system and are thus provided the tools for understanding the data.

The reasons can be validated by the answers to the questions if the dashboard is compatible with other systems the participants use. The answers was 1 strongly agree, 2 agree, and 1 neutral. The neutral could be caused by the fact that this dashboard is specifically designed for the sales division.

On the question if the participants had the required knowledge for using the dashboard, they responded on average very positive. There was one participant who strongly disagreed. This could be caused that this person does not use dashboards frequently, is not familiar to the sales process, or does not work with the sales data and knows what it means. As a solution the manual was made in a detailed manner to explain every function of the tool together with possible interpretations of the data and the graphs. Besides that, an employee was transferred the knowledge of this dashboard who will be maintaining it. This person can help the user of the dashboard if more detailed questions are asked.

In the end, the participants have the required resources and knowledge for using the dashboard. Although there might be need for additional information on how the tool functions. They also consider it to be compatible with their other systems.

Self-efficacy

This concept has the highest standard deviation and thus the strongest disagreement between participants. It has a mean value of 3.31 which is not the lowest value.

The disagreement can be caused by the preference on how and when people need help when performing a task. This is shown in SE-1 where people say they can perform a task when someone helps them every step of the way which has a mean value of 2.5, but with a standard deviation of 1.91 which is the highest of all questions. This disagreement shows that some people like help in one way but not another.

In general, the participants think they can fulfil their tasks if they receive any form of help or amount of time. There is a preference for a built-in function that will help the user complete a task. Here they answered with 1 strongly disagree, 1 neutral, 1 agree, and 1 strongly agree. The strongly disagree might be caused by the function the participant has as this person has a different relation to the tool.

Behavioural intention of use

The overall opinion of the participants was that they expect to use the tool within a period of 6 months. They answered every question in the same way. This could be caused by the fact that the questions were translated into Dutch and that the words have a more similar meaning than in English. There is a level of disagreement on the usage, but this can be explained by the various functions the participants have within the company. The conclusion of these questions is that the management and staff are positive in their opinion of the dashboard as a whole and the function of the dashboard.

Concluding, the participants gave positive feedback on the dashboard and its functions in the form of the UTAUT model questionnaire. They were most positive in their attitude towards the dashboard in regard to if it is a good idea to develop and use it. The lowest score is for the performance expectancy as they think it is a good addition as a tool, but that it will not increase their productivity or salary. The roles of the participants can be seen as an influencing factor as the people who are tasked with the sales division or IT score the dashboard higher than other participants. The feedback on the facilitating conditions, self-efficacy, and behavioural intent all scored positive. Overall, the model is found helpful and the company management expects to use it.

7. Conclusion

In Chapter 7.1, we discuss the answers to the research questions briefly. In Chapter 7.2, the recommendations that resulted from the research are written down and explained. In Chapter 7.3, the limitations of the thesis will be reviewed. In Chapter 7.4, the contributions to theory and practice are mentioned. In Chapter 7.5, the future development is reviewed.

7.1. Research questions

In Chapter 1.7, 5 research questions were asked to answer the main research question: *Which planning method is useful to help Forque plan their number of leads using their data on the company?* The questions have been answered throughout the thesis. This section is a summary of the answers.

How does Forque gather the information on leads?

Currently, Forque has two methods to decide and gather information on leads. The first method is a manual one that is done by filling in an excel file. Forque's management enters all hours that employees should work and divide that by an average hours per order. The second method is more structured and uses data that is being recorded by the company. The method uses the data Forque has on all their previous projects and forecasts together with the data on employee rosters. This calculation uses the rostered hours, holiday and sick leave, CB hours, and average hours per order to calculate the number of leads needed. The calculation is roughly designed as it has numerous hard values, which are programmed into the calculation. These values do no depend on data which would help representing reality, and broad assumptions.

What are methods available for managing and planning workforce?

The discussion in Chapter 3.1 on similar problems indicated that the problem of Forque is on the general assignment problem. This is similar because the company wants to plan the future. The planning of the leads should be based on the information Forque has on their employees and the projects. In Chapter 3.2, the approaches qualitative and quantitative were reviewed. The conclusion was that the most similar problem was the general assignment problem and that the quantitative approach regression modelling was the best for the situation at Forque. This was used as the cause-and-effect was important to how the variables interact with each other.

How can a planning tool be designed for the case of Forque?

The design for the tool was split in three parts. The first was on the main objective of calculating the number of leads. This calculation was done by using the existing way of calculating the number of leads and make it more detailed. These details include aspects of the employees and processes, decreasing the number of assumptions, and cleaning up the data sets.

The second focus was on increasing the insights into the dataset and make it possible to create different scenarios for in the future. This was done by adding 7 new features that focused on increasing insights creating new scenarios.

Lastly, it was needed to design a dashboard. From reviewing the literature on dashboards, the following 3 criteria came forward:

- The dashboard should present similar information in the same way
- The dashboard should be structured in a logical way or represent the workflow
- The information that is shown should not bombard the user

How is the tool connected to Forque's data structure?

The tool is connected to the data structure by real time updates from the database. After that, the model is made in a way to reflect the workflow of the sales division. The data that is needed is on the following subjects:

- Rostered working hours
- Holiday and sick leave
- Forecasts
- Quotations
- Clients
- Realisation of projects
- Pre-calculation of projects

The datasets are linked together by primary and foreign keys and they are: projectID, employeeID, debtorID, and date. Which are used in the coding language DAX as that is the method used by PowerBI.

How can the tool be assessed?

For assessing the tool, it has been decided that the interviewees should be the management team of Forque and the employees who work with dashboards in PowerBI. The evaluation will be performed by a questionnaire. The method for designing the questionnaire is UTAUT. The questionnaire is about how the dashboard functions, looks, and experienced by the interviewees. To assess the functioning of the tool itself, data is compared between the expected values in the tool versus the actual values. An example of this is the holiday and sick leave hours as the calculation is closer to reality in the current tool compared to the previous tool of the company.

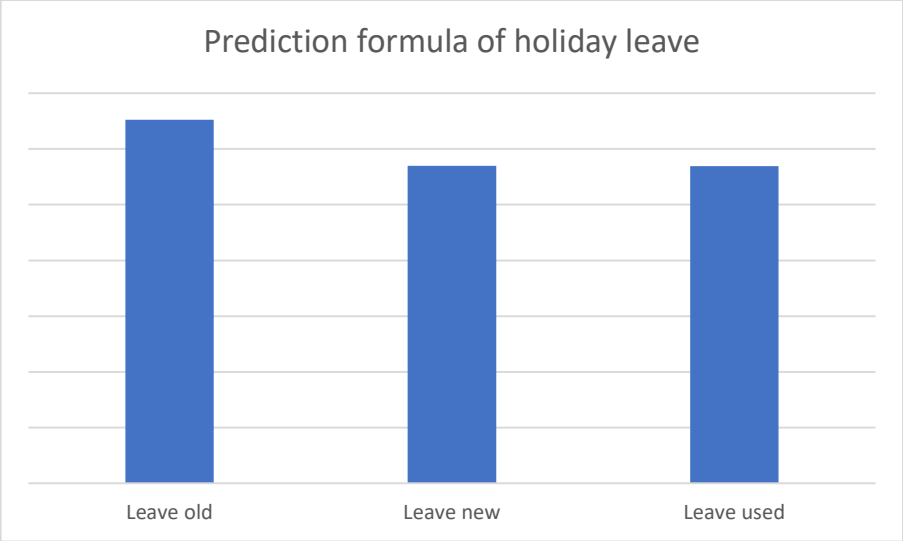

Figure 8: Performance of the holiday leave calculation

In Figure 8, the differences between the old calculation used previously, the new or current calculation made during this thesis, and the actual usage of leave hours is shown. For the accuracy test the absolute percentage error is used. If it is completely accurate the percentage is 0 which goes up as the accuracy goes down. The old calculation has an absolute percentage error of 14,58% and the current calculation has an absolute percentage error of 0,05%. This is an increase in accuracy of 291%.

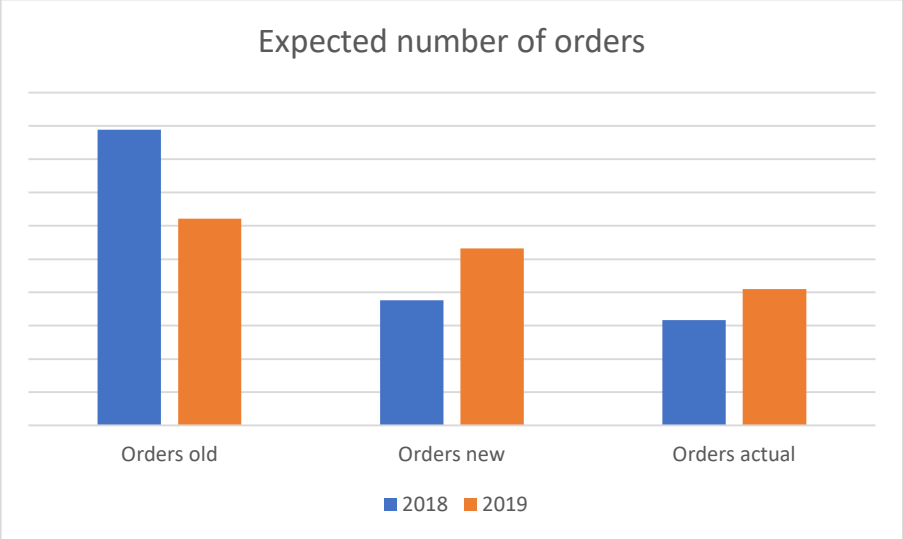

Figure 9: Performance of the forecast of orders

In Figure 9, the goal of this forecasting is tested for its accuracy. The absolute error percentage of the old order calculation previously used for the year 2018 is 181,15% and the year 2019 51,58%. The current calculation has an error in 2018 of 19,00% and in 2019 an error of 29,76%. This is an increase in accuracy of 954% in 2018 and of 173% in 2019.

7.2. Recommendations

The tool is producing values that help the sales division with their jobs. The tool is not yet working optimally. In this section, recommendations for improvement of the tool and the data are written discussed. These recommendations will improve the tool when it is further developed.

The first recommendation is on the data that is used in the calculations. The sales division currently records their sales actions, quotations, and forecasts in different ways then their workflow states. To improve the quality of the data, it is recommended to perform future sales actions and enter data according to the workflow the sales division made every time. Entering data in the same way each time will limit the number of incomplete entries. Furthermore, the sales actions data will improve if the sales actions will be connected to the projects themselves. Currently, no link is present, and this recommendation would help bring more structure to the model.

The second recommendation is to split the projects and support orders from each other. Upselling takes place now with numerous orders. This means that the hours that will be recorded for a project might be different than the expected prognosis. To close this gap, it is recommended to split the project that can end from the support service. The implication would be that there have to be more data entries, but it would pay out in having clearer data and give more structure to the dataset.

The third recommendation is on customer base orders. Forque has deals with some customers that they will have projects every year. It would be beneficial if the customer base projects were recorded in a separate way from projects that were accepted from new customers who do not have a pre-calculation. As CB projects do not have a pre-calculation or forecasts it is difficult to separate them from the “hot” leads, leads that are decided upon within a day. Yet these customer base orders are used in the prognosis for next year. To solve the filtering problem, it is recommended to give a separate identifier to every project to determine if it is customer base or not. The current way to model the CB orders is that the orders of last year are taken and multiplied with a growth factor.

This is a robust way to calculate the hours, yet it is currently not the most PowerBI friendly way as it does cooperate with the year filters. It is also possible to assign a certain percentage of the total number of hours to CB orders, but this might not be representative of the real situation.

The fourth recommendation is about improving the dashboard. As it currently stands, the dashboard can be changed to make it resemble reality better. The first improvement is on the holiday leave prediction. It does consider the rules on how holidays are earned, but the rule that hours can be transferred to next year is not yet implemented. To do the change, it is advised to incorporate the "Absence saldo" into the prediction calculation. The next improvement is on the way newly hired employees are put into the calculation. The calculation that is used now only allows for an average starting date for the entire new group. By modelling starting employees in excel and usage a file as input, the calculation can work with more details. The ones that are recommended are: individual starting dates, include the team, greater number of new hires. What is important here that also date fields are added to help give structure to the rostering. An excel file already exists, so it is recommended to implement it until a table is built into the PowerBI dashboard.

The fifth recommendation is to clean up the database regarding Datadone. The database is already properly implemented, but the administration code is not yet fully implemented in other datasets like forecasts. As Datadone is a feature the company values, it is advised to not leave it an empty field when entering data. When this happens, the company can link the Datadone database with all relevant tables and start filtering based on the desired administration code.

The final recommendation is about changing the productivity calculation. The recommendation is on the productivity of employees. This is not the productivity of the user of the dashboard but a variable that is assigned to an employee in the calculation to determine the number of hours the employee works for clients of their total amount of hours worked. The value for the productivity is currently a predetermined number based on the job of an employee. This value does not represent the reality as close as it could. It is therefore recommended that the calculation should be the same as the one in the management dashboard.

7.3. Limitations

The main limitation of the thesis is that the dataset of the company is small. A small sample can skew data analyses which influence the coefficients used in the calculations. This means that the outcome might not be close to reality and should be different.

The second limitation is that the data links between the project groups and teams is not a one on one relation. This relation is that way as a project group can serve multiple teams and vice versa. No clear relation made analysis more difficult and convoluted.

The third limitation is that the datasets have values for customer base that are difficult to separate. Difficulty filtering data means that the customer base hours can be more than the actual value. The value would currently include leads that did not have a pre-calculation. This separation of customer base hours is not what the company currently sees as customer base but there is no distinction possible.

7.4. Contributions to theory and practice

The contribution to theory is the overview of determining the nature of the problem that needs to be modelled. The research on the types of problems combined with the approaches and the designing of the dashboard gives an overview of the development from problem to solution. A comprehensive overview of what each type of problem does and when it is considered useful is not researched and discussed in depth yet. Especially the differences between the problems and when

to use them is not clearly discussed in a paper. With this research, the distinctions are made clearer and the foundation of the distinctions are made.

The contribution to practice is the blueprint of building a data-analysis dashboard of the sales operations. The model of the dashboard can be used to give other companies an idea how to structure their own dashboards. This blueprint includes the aspects of the process that need to be considered when building a dashboard. The path that was taken in the thesis can be used as a starting point or roadmap for other companies as it has a generic nature that can be adapted to other situations or problems.

7.5. Future development

For the further development, it is advised to make the dashboard easier to work with. From the results of the questionnaire the dashboard was deemed to be cluttered and that some functions seem to be confusing. For further development of the tool, it is advised to review how the year filters function and how they can be incorporated into the graphs. Revising the dashboard could be done in a way that the three calculations of orders needed can be depicted in one graph that is filtered by the year filter. This is important to the usability of the dashboard as the year filter would work more intuitively.

Sources

1 Heerkens, H., & Van Winden, A. (2017). *Solving Managerial Problem Systematically (1st edition).* Groningen: Noordhoff

2 Weske, M. (2012). *Business Process Management: Concepts, Languages, Architectures" (2nd ed.). Springer verlag*

3 Castillo-Salazar, J.A., Landa-Silva, D., & Qu, R., (2016). Workforce scheduling and routing problems: literature survey and computational study. *Annals of Operations Research,* 238(1), 39-67.

4 Cordeau, J.F., Laporte, G., Pasin, F., & Ropke, S., (2010). Scheduling technicians and tasks in a telecommunications company. *Journal of Scheduling,* 13(4), 393-409.

5 Li, Y., Lim, A., & Rodrigues, B., (2005). Manpower allocation with time windows. *Journal of Operational research Society,* 55(11), 302-311.

6 Safarishahrbijari, A., (2018). Workforce forecasting models: A systematic review. *Journal of Forecasting*. 2018(37), 739–753.

7 Litsios, S. (1965.) A Resource Allocation Problem. *Operations Research,* 13(6), 960-988.

8 Bretthauer, K.M., Shetty, B., & Syam, S. (2003). A Specially Structured Nonlinear Integer Resource Allocation Problem *Naval Research Logistics,* 50(7) 770 –792.

9 Bouajaja, S., & Najoua Dridi, N. (2017). A survey on human resource allocation problem and its applications. *Operations Research International Journal,* 2017(17), 339–369. DOI: 10.1007/s12351-016-0247-8

10 Gottlieb, E.S., & Rao, M.R. (1990). The Generalized Assignment Problem: Valid Inequalities and Facets. *Mathematical Programming,* 46 (1990), 31-52.

11 Krumke, S.O., & Thielen, C. (2013). The generalized assignment problem with minimum quantities. *European Journal of Operational Research,* 228(1), 46-55.

12 Yagiura, M., & Ibaraki, T. (2004). Recent Metaheuristic Algorithms for the Generalized Assignment Problem. *Proceedings of the 12th International Conference on Informatics Research for Development of Knowledge Society Infrastructure (ICKS'04),* 229-237. DOI: 10.1109/ICKS.2004.1313429

13 Ortiz-Pimiento, N.R., & Francisco Javier Diaz-Serna, F.J. (2018). The Project Scheduling Problem with Non-Deterministic Activities Duration: A Literature Review. *Journal of Industrial Engineering and Management JIEM*, 2018 - 11(1), 116-134. https://doi.org/10.3926/jiem.2492

14 Winter, F. (2003) Gantthead: Why projects succeed and fail. Accessed https://www.gantthead.com

15 Belassi, W., &Tukel, O.I. (1996). A new framework for determining critical success/failure factors in projects. *International Journal of Project Management,* 14(3), 141–151.

16 Herroelen, W., De Reyck, B., & De meulemeester, E. (1998) Resource constrained project scheduling—A survey of recent developments. *Computers and Operations Research,* 25(4) 279–302.

17 Hyndman, R.J., & Athanasopoulos, G. (2014). *Forecasting Principles and Practice*, OTEXTS.com

18 Silver, E. A., Pyke, D. F., & Peterson, R. (2017). *Inventory and Production Management in Supply Chain (4th ed.), Wiley*

19 Shemin, R. J., Dziuban, S. W., Kaiser, L. R., Lowe, J. E., Nugent, W. C., Oz, M. C., & Wallace, J. K. (2002). Thoracic surgery workforce: Snapshot at the end of the twentieth century and implications for the new millennium. *Annals of Thoracic Surgery*, 73(6), 2014–2032.

20 Masnick, K. (2009). Narrowing the gap between eye care needs and service provision: A model to dynamically regulate the flow of personnel through a multiple entry and exit training programme. *Human Resources for Health*, 7(1), 42–47.
21 Hagopian, A., Mohanty, M. K., Das, A., & House, P. J. (2012). Applying WHO's "workforce indicators of staffing need" (WISN) method to calculate the health worker requirements for India's maternal and child health service guarantees in Orissa State. *Health Policy and Planning*, 27(1), 11–18.
22 Bryant, D. R., Maggard, M. J., & Taylor, R. P. (1973). Manpower planning models and techniques: A descriptive survey. *Business Horizons*, 16(2), 69–78.
23 Sackman, H. (1974). *Delphi assessment, expert opinion, forecasting, and group process*, Santa Monica: Rand Corporation
24 Chang, P.T., Huang, L.C., & Lin, H.J. (2000). The fuzzy Delphi method via fuzzy statistics and membership function fitting and an application to the human resources. *Fuzzy Sets and Systems*, 112 (2000), 511–520.
25 Ishikawa, A., Amagasa, M.,Shiga, T., Tomizawa, G., Tatsuta, R., & Mieno, H. (1993). The max–min Delphi method and fuzzy Delphi method via fuzzy integration, *Fuzzy Sets and Systems,* (1993) 55, 241–253.
26 Caine, D.J., & Parker, B. J. (1996). Linear programming comes of age: a decision-support tool for every manager, *Management Decision, (34)4, 46-53.* DOI: https://doi.org/10.1108/00251749610115161
27 Gordon, A.V.(2017) The boundaries of quantitative forecasting methods: respecting the limits of determinism. *Foresight, (2011)19, 9-15.*
28 Chatfield, C. (2000). *Time-series forecasting.* London: Chapman & Hall/CRC
29 Zais, M. A., & Laguna, M. (2017). Simulation Optimization Approach to Estimate Workforce Requirements. Manuel Laguna Military Operations Research, (22) 1, 19-20. DOI: 10.5711/1082598322119
30 Asher, H. B. (1983) *Causal Modeling.* Thousand oaks: SAGE Publications. DOI: https://dx-doi-org.ezproxy2.utwente.nl/10.4135/9781412983600
31 Andersen, R. (2008). *Modern Methods for Robust Regression.* Thousand oaks: SAGE Publications. DOI: https://dx.doi.org/10.4135/9781412985109
32 Maindonald, J. H., & Braun, W. J. (2010*). Data Analysis and Graphics Using R. An Example-Based Approach (Third ed.)*. New York: Cambridge University Press.
33 Kristoffersen, S. (2008). Flexibility, Completeness and Soundness of User Interfaces: Towards a Framework for Logical Examination of Usability Design Principles. *Proceedings of the 10th International Conference on Enterprise Information Systems HCI*, 346–351.
34 Picking, R., Grout, V., McGinn, J., Crisp, J., & Grout, H. (2010). Simplicity, Consistency, Universality, Flexibility and Familiarity: The SCUFF Principles for Developing User Interfaces for Ambient Computer Systems. *International Journal of Ambient Computing and Intelligence*, 2(3), 40–49.
35 Galitz, W. O. (1992). User-Interface Screen Design. Wellesley, MA: QED Publishing Group.
36 Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly, (27)3, 425-478.*

Appendix 1 Literature review

Research question

The question is: *What are methods available for managing and planning the workforce?* The research will be on how scheduling workforce can provide a solution to the lack of planning in the lead procurement.

Concepts

Quantitative and qualitative forecasting

Workforce forecasting

Optimization planning methods

Inclusion criteria

The first criterium is that the paper discusses the method in a *managerial or organisational* way. This is to make sure the paper is related to the nature of the core problem.

The second criterium is that the paper is *explanatory*. In this review the goal is to get to know each method.

Exclusion criteria

The first criterium is *non-open access.* If I cannot read the paper via either free access or the university library, the paper will be of no use.

The second criterium is *neural networks.* This method requires time as the program is learning by analysing the data sets. The data set of the company is too small to do this.

The third is *not applicable to scheduling or planning problems.*

The fourth is *not using data.*

Databases

The database that will be used is Web of Science. It has easy to use filters for research field, date range, subject area, and others. This will help me find all papers that are of use to this review.

Search terms and strategy

Variables, methods, planning workforce, scheduling workforce, forecasting, appointment allocation, Human resources, manpower

Method AND ((Planning OR scheduling) AND (manpower OR workforce OR “human resources” OR HR)

Variables AND ((Planning OR scheduling) AND (manpower OR workforce OR “human resources” OR HR))

Methods AND forecasting AND (manpower OR workforce OR “human resources” OR HR)

Variables AND forecasting AND (manpower OR workforce OR “human resources” OR HR)

Method AND “appointment allocation” AND manpower OR (workforce OR “human resources” OR HR)

Variables AND “appointment allocation” AND (manpower OR workforce OR “human resources” OR HR)

Search string	Scope	Date of search	Date range	Number of entries
Method AND ((Planning OR scheduling) AND (manpower OR workforce OR “human resources” OR HR)	Topic	12/5/2020	1945-2020	6354
Variables AND ((Planning OR scheduling) AND (manpower OR workforce OR “human resources” OR HR))	Topic	12/5/2020	1945-2020	1111
Methods AND forecasting AND (manpower OR workforce OR “human resources” OR HR)	Topic	12/5/2020	1945-2020	411
Variables AND forecasting AND (manpower OR workforce OR “human resources” OR HR)	Topic	12/5/2020	1945-2020	99
Method AND “appointment allocation” AND manpower OR (workforce OR “human resources” OR HR)	Topic	12/5/2020	1945-2020	285104
Variables AND “appointment allocation” AND (manpower OR workforce OR “human resources” OR HR)	Topic	12/5/2020	1945-2020	1838
Total in Endnote				294917
Removing duplicates				-2398
Selecting based on inclusion				-289706
Selecting based on exclusion				-2359
Removed after abstract reading				-398
Removed after complete reading				-45

Total selected				11

Article	Authors	description	Method	Variables
Workforce forecasting models: A systematic review	Safarishahrbijari, A.	There is a division in forecasting models between high data and low data methods	Qualitative, time series, optimisation, statistical, simulation	Dependent and independent variables
A forecasting method for manpower requirement and related optimal decisions	Guo, Y.J., Pan, D.H., & Zheng, L.H.	Differences amongst tasks influences the manpower scheduling	Dynamic time-varying parameter forecasting	Manpower, technology, employee skills
Application of Markov Model in Human Resource Supply Forecasting in Enterprises	Du, W., & Li, S.	Staff can grow in the company	Markov model, gray model	Movement in the company
Workforce scheduling and routing problems: literature survey and computational study	Castillo-Salazar, J. Arturo; Landa-Silva, Dario; Qu, Rong	For a diverse workforce the activities should be distributed in time windows if there is time constraints	Heuristics, constraint programming, multi-commodity network flow problem, integer programming model	Time windows, skills and qualifications, service time, connected activities, and teaming
A Simulation Optimization Approach to Estimate Workforce Requirements	Mark Zais and Manuel Laguna	Aggregate models can save cost and limits flexibility	Simulation Optimization, Forecasting, Time Series	Workers with skill sets, demand
Identification of Variables and their Influence on the Human Resources Planning in the Territorial Level	Rodobaldo Martínez-Vivar , Alexander Sánchez-Rodríguez , Reyner Pérez-Campdesuñer , Gelmar García-Vidal	Worker welfare influences productivity and willingness to continue working	empirical	Grow possibilities for employees,
Skilled workforce scheduling in Service Centres	Vicente Valls a , A´ ngeles Pe´rez b , Sacramento Quintanilla b,*	For daily scheduling, urgency and critical tasks have an impact on the schedule	Multi-stage scheduling, generic algorithms	Staff skills, urgency
A successive convex approximation method for multistage workforce capacity planning problem with turnover	Haiqing Song a,b, Huei-Chuen Huang a,*	Advantage of multi-stage is that it can handle large problems whereas linear programming would not be possible	Multi-stage scheduling, stochastic, successive convex approximation	Number of employees, turnover, cost of hiring and transferring
Variable employee productivity in workforce scheduling	Gary M. Thompson a,1, John C. Goodale b,*	Linear representation of the workforce is inadequate	Nonlinear representation	Productivity, types of employees, cost working, work schedules

The fuzzy Delphi method via fuzzy statistics and membership function fitting and an application to the human resources	Ping-Teng Chang a ;∗, Liang-Chih Huang b, Horng-Jiun Lin c	For incomplete understanding of the reality a fuzzy model can be used. It uses statistics to predict.	Fuzzy delphi method	-
Input±output analysis for organizational human resources management	Hector Correa a, *, James Craft b a Graduate School of Public and International A airs, University of Pittsburgh, Pittsburgh, PA 15	Input-output models based on organisational structure help forecast personnel requirements	Input-output models	Employee characteristics, tasks, skills, function

Multi-objective multi-mode resource constrained project	Erfan Babaee Tirkolaee1,2 · Alireza Goli3 · Milad Hematian1 · Arun Kumar Sangaiah4 · Tao Han5	
Toward an integrated workforce planning framework using structured equations	Marie Doumic b , Benoît Perthamec , Edouard Ribes d,∗ , Delphine Salort a , Nathan Toubiana	
Integer programming models for hierarchical workforce scheduling problems including excess off-days and idle labour times	Cemal Özgüven ⇑ , Banu Sungur	
Workforce planning incorporating skills: State of the art	Philippe De Bruecker a,∗ , Jorne Van den Berghb, Jeroen Beliënb, Erik Demeulemeester a	

Appendix 2 Data analysis

Redacted due to sensitive data

Figure 1: Power formula

Redacted due to sensitive data

Figure 2: Polynomial formula

Redacted due to sensitive data

Figure 3: Logarithmic formula

Redacted due to sensitive data

Figure 4: Exponential formula

Redacted due to sensitive data

Figure 5: Linear formula

Redacted due to sensitive data

Figure 6: Power formula

Redacted due to sensitive data

Figure 7: Exponential formula

Redacted due to sensitive data

Figure 8: Linear formula

Redacted due to sensitive data

Figure 9: Logarithmic formula

Redacted due to sensitive data

Figure 10: polynomial formula

Appendix 3 Dashboard manual

Manual dashboard sales-marketing Forque

Overview panel

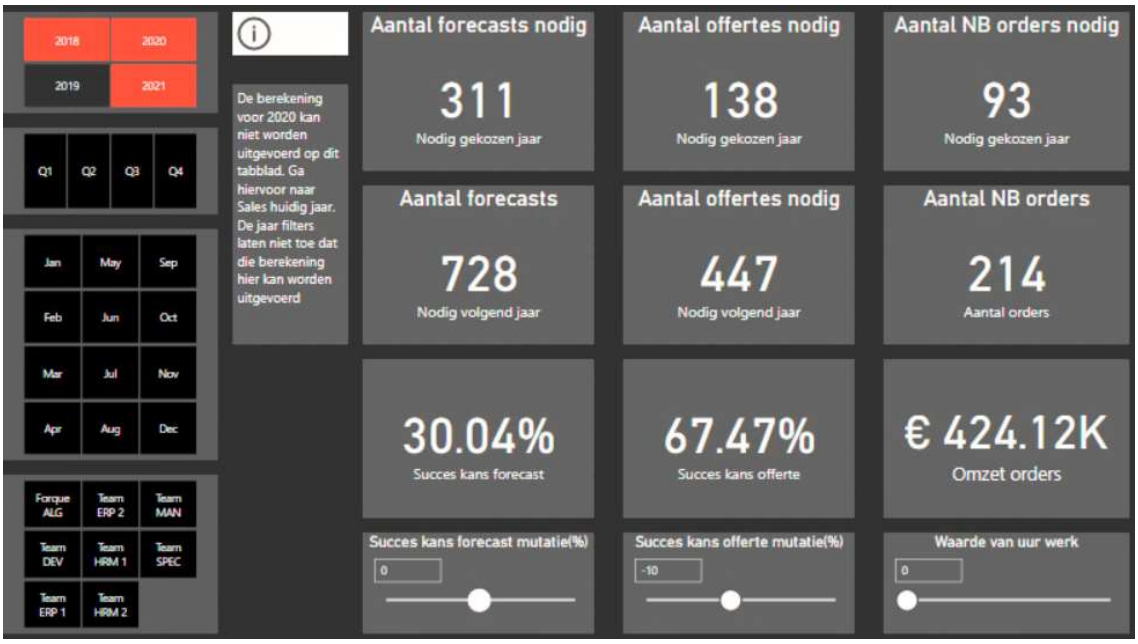

Figure 1: Overview panel

This panel in Figure 1 is a general overview and place to influence 3 factors of the process. On the left side of all panels is a set of filters that allow the user to pick a year, month, quarter, or team that they want to know more of in detail. The one aspect that should be remembered is that the calculations for the current year is only shown in the "Sales current year" panel. This is written down in a card that can be viewed by clicking on the i button in the top row of the panel. The card is shown in Figure 2.

Figure 2: Information on overview panel

The numbers along the top show from left to right are the number of forecasts, quotations, and orders needed for the chosen year. In the second them are the same values but for the next year or the future.

The numbers in the third row are the success rates of left forecasts, here 30.04% and in the middle quotations, here 67.47%. The right value is the revenue based on the number of orders and the average order value. The success rate is calculated by dividing the number of forecasts or quotations that ended in success by the total number of forecasts or quotations in that year.

Below the values for future years are the parameters the user can change. The left two influence the success rates of forecasts and quotations. If it is above 0 then the change will increase from the current value and if it is below 0 it is decreased from the current value. So, for example if the mutation of forecasts is set to -11, the success rate would be 30.04 – 11 = 19.04%. The right input box is for the average value of an order. If this is above 100 then fewer orders are needed, then now as there is more revenue made per order.

The reason why the current year does not work here is because the way the year filter interacts with the calculation. The filter acts when a year is chosen like other years do not exist. As the calculation makes use of the roster hours of the current years and the prediction of customer base hours is calculated based on the previous years, it is not possible to filter on one year. This means that when one of the years is selected the other parts of the calculation see no data and therefore does not work.

Sales current year

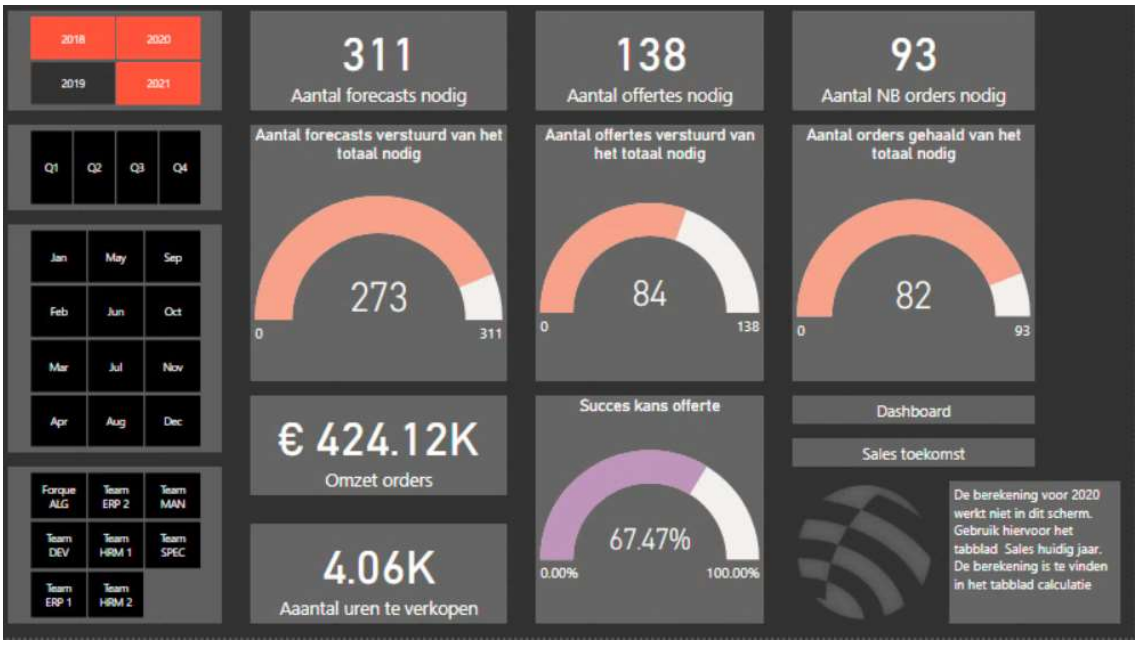

Figure 3: Sales current year panel

This panel shows insight into the state of the sales division in the past years. The past can be viewed by selecting a past year on the left-hand side. The top row of values is the same as on the "Overview" panel.

The three half circle diagrams are on the status of forecasts, quotations, and orders. The left shows the number of forecasts that where sent versus all forecasts that were needed to be sent. So, in 2019 were 273 forecasts sent of the 311 that were needed. This is the same for the other two diagrams in this row. This is made to give the sales division an easy overview on how close they are to their target. Below these three is one half circle diagram more. This is the success rate of quotations. In 2019 a quotation had a success rate of 67.47%.

In the lower left corner are two boxes with values in them. They are the revenue of that year and the number of hours that need to be sold. The expected revenue was 424k euros and 4.060 hours had to be sold in 2019.

Sales current year

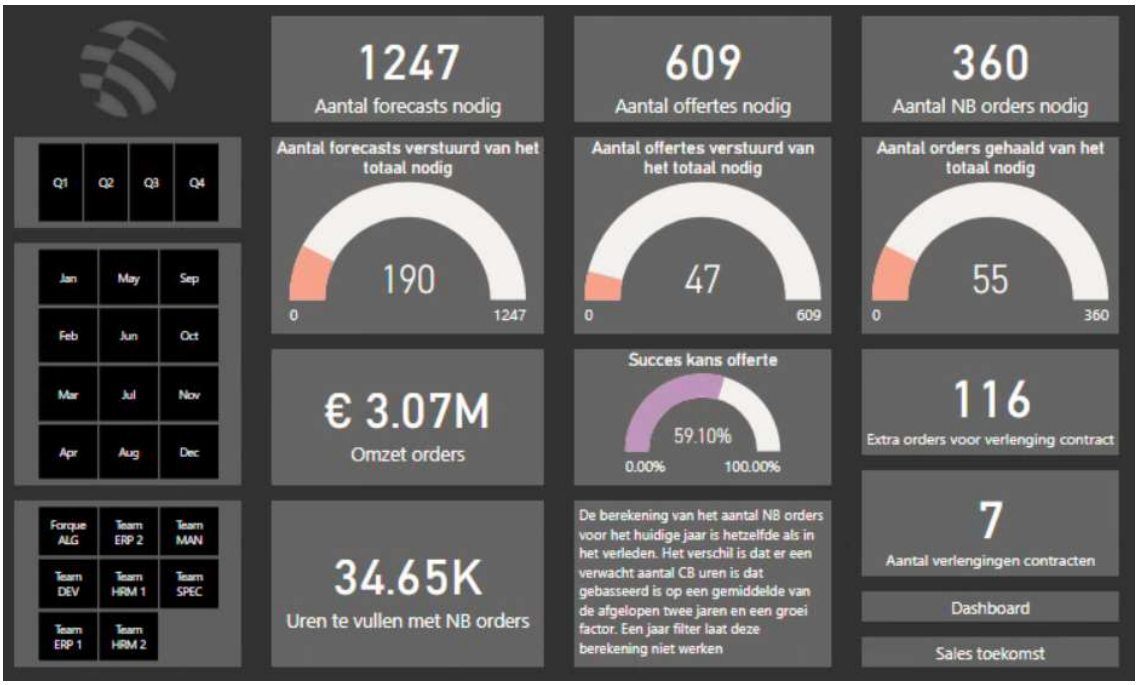

Figure 4: Sales current year panel

This panel shows the values for the sales division in the current year.

In the top row are the forecasts, quotations, and orders needed in this year and how many have been sent. So of the 609 quotations needed to be sent only 47 have been.

Beneath these three values are 2 more numbers on the right had side of the panel. They represent the impact of extending the contracts that end this year. They are the number of orders that are needed extra if contracts are extended and the number of contracts that could be extended. During the year contracts will end. There is the possibility that a contract will be extended, but this extension is not yet taken into consideration in the main calculation. To that end is this value. By using the hours that could be made if a contract is extended an extra number of orders is calculated. So, in this case 7 contracts can be extended which results in a need for an extra 116 orders on top of the 360.

In the lower left corner are the revenue and number of hours that need to be sold.

Sales next year

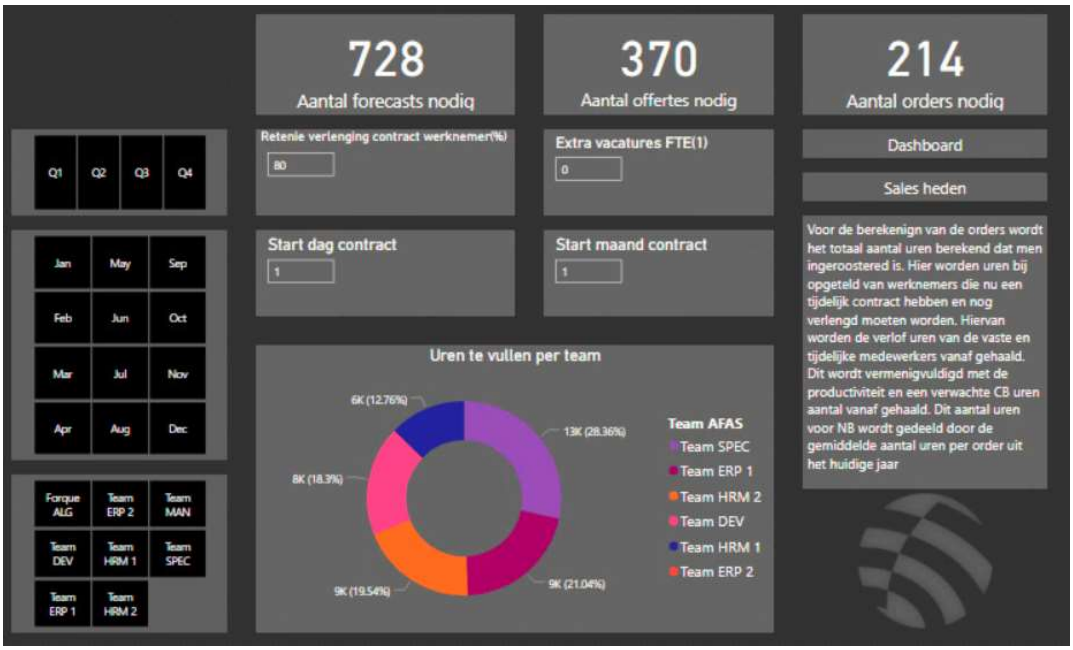

Figure 5: Sales next year panel

This panel is designed to give the sales division and the management team. It presents what the future requirements of the sales division will be. In the top row are the same values as in the second row on the “Overview” page.

Below them are 4 input fields that the user can use to create various scenarios. They are the retention of employees during extension of the contract, number of new employees, starting day of the contract, and starting month of the contract. The top left one is on how many people will want to extend their contract if their contract ends the previous year as a percentage. The one on the right is how many new people will be hired that year with FTE(1) meaning it is a full contract. The date when the newly hired start can be written down in the two input boxes below.

The donut diagram shows how many hours each team has that need to be filled. This way the sales division can see where they have to focus. As can be seen that team specials needs 13k hours to be filled whereas team HRM 1 only needs 6k. So, more work might be needed for team specials to fill these hours and shows that the attention of the sales division cannot be equally divided amongst the teams but that some need more than others.

Forecasts pre/after calculation

Figure 6: Forecast pre-/after-calculation panel

This panel paints a picture on how the deals sales make play out in reality. When a deal is made an expected number of hours are allocated and with this diagram it shows how many hours are actually spend on projects. This helps making future estimates be more precise.

In the upper row are two values, the average order size and the comparison between orders expected and actually used. The left shows how many hours are actually spent on average on projects, in this case 20.75 hours. The right shows a comparison where if it would be 100% the estimation is equal to reality, below 100% the reality is lower than expected, and above 100% it means reality is higher than expected. In this case it is lower than 100%, so that means fewer hours are spent on projects than expected.

The graph shows the same information but visually. If the orange bar is higher than the blue one it means more hours are spent on projects than expected, or when it is lower fewer hours are spent.

Forecasts

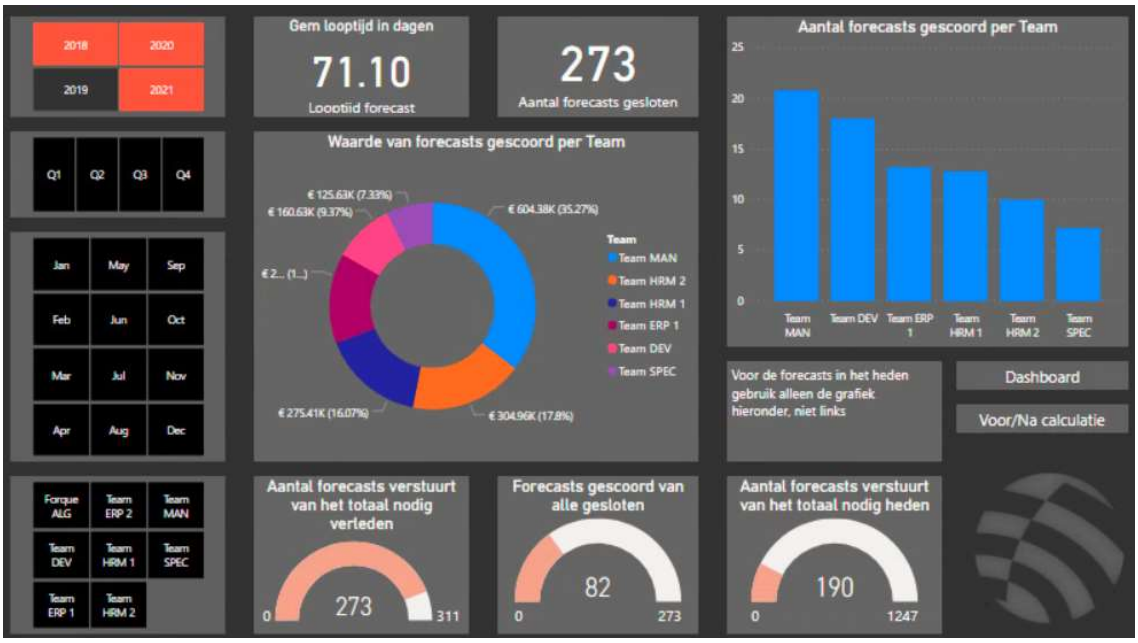

Figure 7: Forecast panel

This panel is on how the status is of the forecasts. In the top row the average length of negotiations, here 71.10 days, and the number of negotiations ended is shown, here 273 for 2019.

Below that is the breakdown of the average value per order per team in the three classes of order size. This shows which team has a higher average project value. This combined with the graph in the upper right corner which shows the number of forecasts scored per team shows which team has the most revenue. In this case team MAN has the highest average order value and the highest number of forecasts scored.

In the bottom row is a breakdown of orders sent of the should be sent and how many ended in success. The half circle diagram on the right is only for the current year, the one on the left is only for the past years and will not work for the current year. So, the chart on the right shows that for the current year 1247 forecasts are needed to be sent and 190 have been. The chart on the left shows that for 2019 311 forecasts were needed to be sent and 273 have been sent. The closer the number of sent is to the number should be sent the fewer idle hours employees have which means more revenue. The chart in the middle shows the success of forecasts. Of the 273 sent in 2019 82 ended in success. If the number of forecasts that ended in success is close to the number sent, then less work is needed as the number of forecasts that should happen is reached earlier.

Leave usage

Figure 8: Leave usage panel

This panel is about how the holiday and sick leave are for the company and how the value of an hour's work develops over time. The breakdown on leave usage is used to check if the calculations are close to reality and if some teams need fewer hours due to prolonged sick leaves.

The graph in the top right corner shows the expected leave, the used leave, and the used maternity leave, this last one is as maternity leave is seen as sick leave. If the orange bar is higher than the purple one, then more sick leave is used then expected. If it is lower, then fewer hours are used. This is important because when fewer sick leave hours are used then there has to be more work to fill these unexpected hours. To help the management with the estimation a table is made on the right side that shows the total value of sick leave versus the expected to show if the company as a whole uses more or less sick leave than expected.

In the upper right corner is the average price of a working hour per year. This helps the management with determining if they can still increase the price, have to change the financial budget, or expect anything regarding the price. Currently the price is still increasing, but at a lower rate than in the past. It can be expected that the price does increase but not as significantly as in the past.

In the lower left corner is the breakdown of holiday leave expected versus used. This is similar as the sick leave chart as when the red bar is higher than the blue bar more holiday leave is used than expected. If it is higher than fewer hours are needed to be filled so fewer orders are needed. The opposite is true if the red bar is lower than the blue one.

Calculations

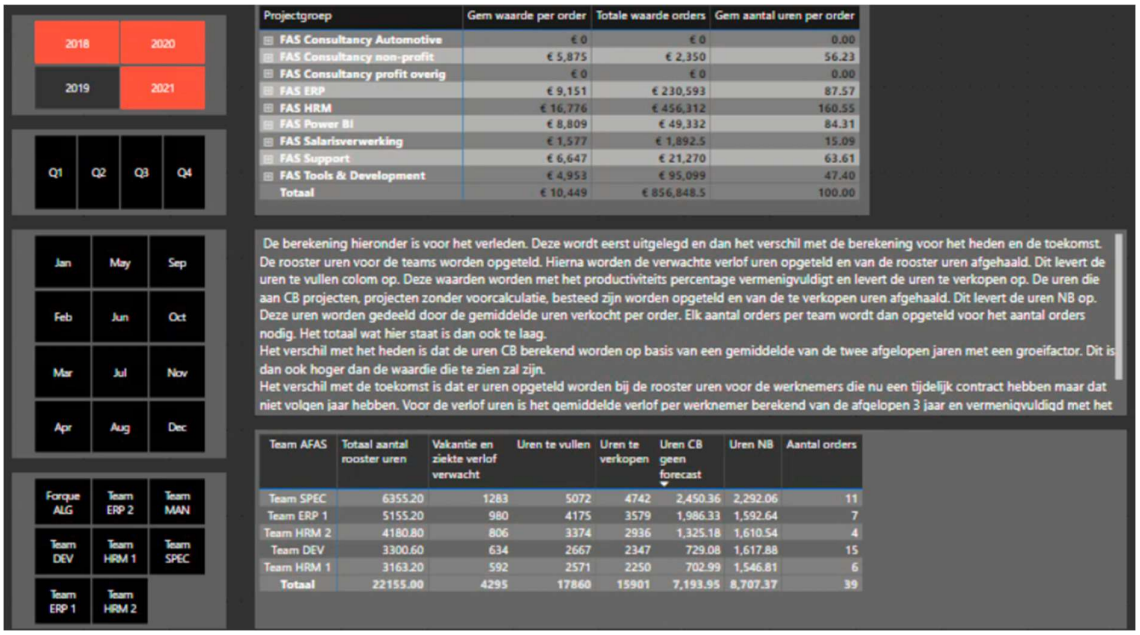

Figure 9: Calculations panel

This panel shows insight into the calculation. The depicted calculation is for the past years. Here the total number of roster hours is summed and allocated to the teams. The same happens for the leave hours. The leave is subtracted from the roster hours which results in the hours to fill. Then a productivity factor is used as not every hour is spent on the clients so by multiplying the hours to fill with the productivity, the result is the hours to sell. Then the hours that are spent on customer base or projects that did not have a pre-calculation are summed and subtracted from the hours to sell, resulting in the hours new base. This hours new base is divided by the average number of hours per order which calculates the number of orders needed.

The calculation for the current year is different in the customer base part of the calculation. This formula uses the past customer base hours and multiplies it with a growth factor.

The calculation for the future uses the same customer base approach as the current year formula, but uses the input values for extra employees in the roster hours and leave hours. An average number of leave hours per employee is determined based on past years and is then multiplied by the number of employees in the future.

Appendix 4 code

Measures

```
#Average Hours Per Order =

DIVIDE([$Average Order Scored], [$Hour's work], 0)


#Average Order Size AfterCalculation =

Calculate(

        AVERAGE(DIM_Projects[OrderSizeAfterCal]), DIM_Projects[OrdersizePreCal] <> 0,
        DIM_Projects[OrdersizePreCal] <> BLANK ()

        )
```

--Average ordersize for orders that did have a precalculation

```
#Average Order Size PreCalculation =

Calculate(

        AVERAGE(DIM_Projects[OrdersizePreCal]), DIM_Projects[OrdersizePreCal] <> 0,
        DIM_Projects[OrdersizePreCal] <> BLANK()

        )
```

--Average expected order size for projects with precalculation

```
#Employee y =

Calculate(

        DISTINCTCOUNT(FCT_Presence[NaamMW]), DIM_Calendar[CurYearOffset] = 0)


#Employee y+1 =

Calculate(

        DISTINCTCOUNT(FCT_Presence[NaamMW]), DIM_Calendar[CurYearOffset] = 1)


#Extensions = [#Employee y] - [#Employee y+1]


#Forecast General = COUNTROWS(FCT_Forecasts)
```

```
#Forecasts Closed =

IF(

        CALCULATE( COUNT( FCT_Forecasts[Resultaat] ) , FCT_Forecasts[Resultaat] <> "Open") =
        BLANK()
        , 0
        , CALCULATE( COUNT( FCT_Forecasts[Resultaat] ) , FCT_Forecasts[Resultaat] <> "Open")
)


#Forecasts Closed Scored =

IF(

CALCULATE(

    COUNT(FCT_Forecasts[Resultaat])

    , FILTER( FCT_Forecasts, FCT_Forecasts[Resultaat] = "Gesloten en gescoord")

    ) = BLANK()

  ,0

  ,CALCULATE(

    COUNT(FCT_Forecasts[Resultaat])

    , FILTER( FCT_Forecasts, FCT_Forecasts[Resultaat] = "Gesloten en gescoord")

    ))

#Forecasts Needed =

DIVIDE([#Orders Needed],[%Forecasts Scored], 0)
```

--Use success rate to determine forecasts from orders

```
#Forecasts Needed Future =

IF(

        DIVIDE([#Orders Future], [%Forecasts Scored], 0) = BLANK(),0 ,DIVIDE([#Orders Future],
        [%Forecasts Scored], 0)

        )
```

--Use success rate to determine forecasts from orders

```
#Holiday Hours Extension =

8*(26/12)*

(
```

```
   12*([#Employee y]-[#Employee y+1])
   -
   CALCULATE(SUM(DIM_Employees[End_Month]),DIM_Employees[End_Year] = Year(UTCNOW()))
)

#Hours CB Calculated Invoice =
CALCULATE(
   SUM(FCT_DebtorInvoicesLineItems[DebtorInvoicesLineItems_Aantal_eenheden]),
   DIM_Debtors[CB (Y/N)] = "Yes", DIM_Projects[OrdersizePreCal] = Blank(),
FCT_DebtorInvoicesLineItems[Eenheid] = "Uren"
      ) --Determine hours CB based on the hours invoiced to projects from CB clients without pre-calculation

#Hours CB Estimate =
SUM(FCT_CB_Hours_Estimate[CB uren geen forecast])
--manual estimation based on data from 11/11/2019

#Hours Holiday Current =
CALCULATE(SUM(FCT_Presence[Presence_AantalUren]), FCT_Presence[CodeRegel] = "V")
--Sum all hours holiday that have code V

#Hours Holiday Future =
DIVIDE(
      Calculate(Divide([#Total Hours Holiday
Future],DISTINCTCOUNT(FCT_Presence[NaamMW]),0),DIM_Calendar[CurYearOffset] = -2)
      +
      Calculate(Divide([#Total Hours Holiday
Future],DISTINCTCOUNT(FCT_Presence[NaamMW]),0),DIM_Calendar[CurYearOffset] = -1)
      +
      Calculate(Divide([#Total Hours Holiday
Future],DISTINCTCOUNT(FCT_Presence[NaamMW]),0),DIM_Calendar[CurYearOffset] = 0)
      ,3
      ,0
```

```
)
```

--Take Sum of hours holiday last 3 years and divide by number of employees who used them to determine average holiday hours per employee

```
#Hours NB =
[#Hours to Fill] - [#Hours CB Calculated Invoice]
```

--Hours NB is the number of hours left from hours to fill without CB

```
#Hours NB calculated =
CALCULATE(
        SUM(FCT_Realization[Realization_QuantityUnit]),FCT_Realization[InitId] = "UUR")

#Hours NB Future =
[#Hours to Fill Future] - Calculate([#Hours CB Calculated Invoice], DIM_Calendar[CurYearOffset] = 0)
```

--Hours NB Future is the hours to fill future with the assumption that CB does not get higher since last year

```
#Hours Sick Current =
CALCULATE(SUM(FCT_Presence[Presence_AantalUren]),FCT_Presence[CodeRegel] = "Z")
```

--Sum all hours sick that have code Z

```
#Hours Sick Future =
DIVIDE(
  Calculate( Divide( [#Total Hours Sick Future] , DISTINCTCOUNT(FCT_Presence[NaamMW] ) , 0 ) , DIM_Calendar[CurYearOffset] = -2)
  +
  Calculate( Divide([#Total Hours Sick Future], DISTINCTCOUNT(FCT_Presence[NaamMW]), 0), DIM_Calendar[CurYearOffset] = -1)
  +
  Calculate( Divide( [#Total Hours Sick Future] , DISTINCTCOUNT( FCT_Presence[NaamMW] ) , 0 ) , DIM_Calendar[CurYearOffset] = 0 )
 , 3
, 0
```

```
)

#Hours to Fill =
[#Total Hours] - SUM(FCT_Absence_Rostered[Hours Unavailable])

#Hours to Fill Extension =
[#Total hours extension] - [#Holiday Hours Extension]

#Hours to Fill Future =
(
   [#Total Hours Future]              --# of hours work for employees with contract start of year

   - [#Hours Unavailable Future]
)

#Hours Unavailable Current =
[#Hours Holiday Current] + [#Hours Sick Current]
*
(
[Change Number of Employees Added]+Divide([Waarde van retentie],100,0)
*
(Calculate(DISTINCTCOUNT(FCT_Presence[NaamMW]), DIM_Calendar[CurYearOffset] = 0)-
[#Employee y+1])
)

#Orders =
RoundUp(DIVIDE([#Hours NB], [#Average Hours Per Order], 0),0)
--Number of orders is the number of hours to fill divided by the number of hours per order. As
orders must be a whole number it must be rounded up to prevent shortage

#Orders Extension =
Divide([#Hours to Fill Extension],[#Average Hours Per Order],0)
```

```
#Orders Future =

IF(Divide([#Hours NB Future], Calculate([#Average Hours Per Order], DIM_Calendar[CurYearOffset] = 0)) = BLANK()

, 0

, Divide([#Hours NB Future], 1.2*Calculate([#Average Hours Per Order], DIM_Calendar[CurYearOffset] = 0))

)
```

--Expected number of hours per order is not going to change

```
#Orders Future New Employees =

DIVIDE(

    (

      (

        (Calculate(DISTINCTCOUNT(FCT_Presence[NaamMW]), DIM_Calendar[CurYearOffset] = 0)-[#Employee y+1])*[Waarde van retentie]

      +[Change Number of Employees Added]

      ) --Number of employees added

      *

      8

      *                                                    --8 hour working day for all new employees

      (Calculate(DISTINCTCOUNT(FCT_Presence[Presence_StartDate]), FCT_Presence[CodeRegel] = "R", DIM_Calendar[CurYearOffset]= 1)
```

--Count the number of days from start contract to end of year

```
      - 61                                                    --Simulate training period

      - Datediff(DATE([Year +1],1,1),DATE([Year +1],[#Month Start Contract],[#Day Start Contract]),DAY)

      )

    )

  -     --Subtract the number of hours unavailable new employees from Hours total

    (24

      +
```

```
        8                       ----Number of hours leave per day
        *
        (26/12                  --Amount of holiday days per month
        *
        (12-[#Month Start Contract])    --Number of months employee earns holiday leave
        )
    )
    *
    (
        [Change Number of Employees Added]+
        [Waarde van retentie]*(Calculate(DISTINCTCOUNT(FCT_Presence[NaamMW]),
        DIM_Calendar[CurYearOffset] = 0)
        - [#Employee y+1])
    )
, [#Average Hours Per Order]
,0)

#Orders Needed =
CALCULATE([#Orders], DIM_Teams[Team AFAS] = "Team DEV")
+
CALCULATE([#Orders], DIM_Teams[Team AFAS] = "Team ERP 1")
+
CALCULATE([#Orders], DIM_Teams[Team AFAS] = "Team HRM 1")
+
CALCULATE([#Orders], DIM_Teams[Team AFAS] = "Team HRM 2")
+
CALCULATE([#Orders], DIM_Teams[Team AFAS] = "Team MAN")
+
CALCULATE([#Orders], DIM_Teams[Team AFAS] = "Team SPEC")
```

--Add the #orders per team to get a number closer to reality, otherwise number of orders uses the overall average number of hours per order which is aggregating

```
#Orders Needed Future =
CALCULATE([#Orders Future], DIM_Teams[Team AFAS] = “Team DEV”)
+
CALCULATE([#Orders Future], DIM_Teams[Team AFAS] = “Team ERP 1”)
+
CALCULATE([#Orders Future], DIM_Teams[Team AFAS] = “Team HRM 1”)
+
CALCULATE([#Orders Future], DIM_Teams[Team AFAS] = “Team HRM 2”)
+
CALCULATE([#Orders Future], DIM_Teams[Team AFAS] = “Team MAN”)
+
CALCULATE([#Orders Future], DIM_Teams[Team AFAS] = “Team SPEC”)
+
[#Orders Future New Employees]
```

--Add the #orders per team to get a number closer to reality, otherwise number of orders uses the overall average number of hours per order which is aggregating

```
#Sales Quotations Needed =
Divide([#Orders Needed], [%Sales Quotations Scored], 0) --Use the success rate of quoatitons to determine #quotations from orders

#Sales Quotations Needed Future =
DIVIDE([#Orders Future], [%Sales Quotations Scored Future], 0) --Use the success rate of quoatitons to determine #quotations from orders

#Sales_Orders Current = DISTINCTCOUNT(FCT_Sales_Orders[Dossieritem])

#Total Hours =
CALCULATE(
    SUM(FCT_Presence[Presence_AantalUren])
    , FCT_Presence[CodeRegel] = “R”
    )
```

```
*
CALCULATE(AVERAGE(DIM_Employees[Productivity]))
--Total number of hours that is being worked without leave on all weekdays of 8 hours per day


#Total hours extension =
 8*
(([#Employee y]-[#Employee y+1]) --#rehires
   *
   CALCULATE(DISTINCTCOUNT(FCT_Presence[Presence_StartDate]),DIM_Calendar[CurYearOffset] =
0) --calculate number of days in a year
    -
   CALCULATE(COUNT(FCT_Presence[Presence_StartDate]),DIM_Employees[End_Year] =
YEAR(UTCNOW())) --calculate number of days rehires are already working
)


#Total Hours Future =
CALCULATE(
      SUM(FCT_Presence[Presence_AantalUren])
      , FCT_Presence[CodeRegel] ="R"
      , DIM_Calendar[CurYearOffset] = 1
      )
--Total number of hours that is being worked without leave on all weekdays of 8 hours per day
next year for people who have a contract
+
   (
      ([#Employee y]-[#Employee y+1])*(Divide([Waarde van retentie],100,0)
      +[Change Number of Employees Added]
   ) --Number of employees added
      *
      8
      *                                                    --8 hour working day for all new employees
      (
```

```
        [#Working Days y+1] --Count the number of days from start contract to end of year
        - 61        --Simulate training period
    - Datediff(DATE([Year +1],1,1),DATE([Year +1],[#Month Start Contract],[#Day Start Contract]),DAY)
    )
  )                 --Number of days between start of the year and start of contract

#Total Hours Holiday Future =
CALCULATE(
    SUM(FCT_Presence[Presence_AantalUren])          --Total number of holiday leave hours
    , FCT_Presence[CodeRegel] = "V"                 --Holiday leave
    , MONTH(DIM_Calendar[CurMonthOffset])           --group per month
    , FCT_Presence[Presence_StartDate]>= UTCNOW() -730  --2 years in the past
    )

#Total Hours Sick Future =
CALCULATE(
    SUM(FCT_Presence[Presence_AantalUren])          --Total hours sick leave
    , FCT_Presence[CodeRegel] = "Z"                 --sick leave
    , MONTH(DIM_Calendar[CurMonthOffset])           --group per month
    , FCT_Presence[Presence_StartDate]>= UTCNOW() -760  --2 years past
    )

#TravelTime =
CALCULATE(SUM(FCT_Realization[Realization_QuantityUnit]), FCT_Realization[ItemCodeId] = "7")

#Working Days y+1 =
Calculate(DISTINCTCOUNT(FCT_Presence[Presence_StartDate]), FCT_Presence[CodeRegel] = "R", DIM_Calendar[CurYearOffset]= 1)

%Forecasts Scored =
IF(
```

```
   DIVIDE([#Forecasts Closed Scored], [#Forecasts Closed], 0) -- number success and number scored
for success rate
   +
   Divide([Change Success Rate Forecasts], 100, 0 )          --Input from dashboard
   = BLANK()
   , 0
   ,DIVIDE([#Forecasts Closed Scored], [#Forecasts Closed], 0) -- number success and number scored
for success rate
+
Divide([Change Success Rate Forecasts], 100, 0 ))


%Fullfillment Pre/After Calculation =
DIVIDE([#Average Order Size AfterCalculation], [#Average Order Size PreCalculation],0)


%Sales Quotations Scored =
CALCULATE(Divide(
      CALCULATE(COUNTROWS(FCT_Sales_Orders)          --Filter all quotations for success
            , FCT_Sales_Orders[Status_code] = 4)             --Code 4 means success
      ,CALCULATE(COUNTROWS(FCT_Sales_Orders)         --All quotations finished
            , FCT_Sales_Orders[Status_code] = 4)             --Filter all quotations for success
      +
      CALCULATE(COUNTROWS(FCT_Sales_Orders)          --Filter all quotations for failure
            , FCT_Sales_Orders[Status_code] = 6)
      , 0
   ))


%Sales Quotations Scored Future =
DIvide(
   DIVIDE(
   CALCULATE(COUNTROWS(FCT_Sales_Orders_FutureCalc)       --Filter all quotations for success
            , FCT_Sales_Orders[Status_code] = 4)
   ,CALCULATE(COUNTROWS(FCT_Sales_Orders_FutureCalc)       --Filter all quotations for success
```

```
            , FCT_Sales_Orders[Status_code] = 4)
      +
      CALCULATE(COUNTROWS(FCT_Sales_Orders_FutureCalc)        --Filter all quotations for failure
            , FCT_Sales_Orders[Status_code] = 6)
   ,0)
+
DIVIDE(CALCULATE(COUNTROWS(FCT_DossierItems_Overig)        --Filter all quotations for success
            , FCT_DossierItems_Overig[Type_dossieritem] = "37",
FCT_DossierItems_Overig[Codetabelwaarde] ="Offerte is terug van klant")
   ,CALCULATE(COUNTROWS(FCT_DossierItems_Overig)        --Filter all quotations for success
            , FCT_DossierItems_Overig[Type_dossieritem] = "37",
FCT_DossierItems_Overig[Codetabelwaarde] = "Offerte is terug van klant")
      +
   CALCULATE(COUNTROWS(FCT_DossierItems_Overig)        --Filter all quotations for failure
            , FCT_DossierItems_Overig[Type_dossieritem] = "37",
FCT_DossierItems_Overig[Codetabelwaarde] = "Offerte is afgekeurd")
   ,0)
   ,2
   ,0)
+
Divide([Change Success Rate Sales], 100, 0)           --Input from dashboard


$Average Order Scored =
DIVIDE([$Sum Order],[#Forecasts Closed Scored],0)
--Divide the total revenue by number of orders to get average order value


$Average working hour = --Average value of an hour's work
CALCULATE(
         AVERAGE(FCT_DebtorInvoicesLineItems[DebtorInvoicesLineItems_Prijs_per_eenheid]) --take
the average invoiced price of an hour
         , FCT_DebtorInvoicesLineItems[Eenheid] = "Uren" --only average over hours that are invoiced
      )
```

$Hour's work =

100 **--Standardised value of an hour's work**

+

[Change Value Hour's work] **--Input from dashboard**

$Sum Order = S

UM(FCT_Forecasts[Forecasts_Orderbedrag_basisvaluta]) **--Total value of all orders based on the forecasts**

$travel =

CALCULATE(SUM(FCT_Realization[Realization_SalesAmount]), FCT_Realization[ItemCodeId] = "7")

Data sets

Dim projects

OrdersizePreCal =

CALCULATE(SUM(FCT_ProjectPrecalculation[Aantal_eenheden]))

OrderSizeAfterCal =

CALCULATE(SUM(FCT_Realization[Realization_QuantityUnit]))

CB (Y/N) =

IF(SUM(FCT_LedgerEntries[# Entry])>2, "Yes", "No")

aftercalInvoice =
CALCULATE(SUM(FCT_DebtorInvoicesLineItems[DebtorInvoicesLineItems_Aantal_eenheden]))

FCT rostered absence

Holiday days =

(26/12) **--Number of holidays earned per month**

*

```
(12
  -
  IF(                                    --Checks the number of holidays from the beginning of the year since hiring
    CALCULATE(SUM(DIM_Employees[Start_Year])) = FCT_Absence_Rostered[Start]    --Checks if hired this year
    , CALCULATE(SUM(DIM_Employees[Start_Month]))--If yes then subtract the number of months
    , 0                                    --if no nothing
  )
  -
  IF(          --Checks the number of holidays from the end of the year from retiring or leaving
      Calculate(SUM(DIM_Employees[END_Year])) = 1900     --Does not leave and has permanent contract
      , 0                                  --If yes nothing
      , IF(                                --If no next check
          CALCULATE(SUM(DIM_Employees[End_Year])) <> FCT_Absence_Rostered[Start]
--Checks if not leaving in the year
          , 0                                        --If yes nothing
          , 12-CALCULATE(SUM(DIM_Employees[End_Month]))
--If not calculate the number of months from the end of the year that should be subtracted
        )
    )
)


Hours Unavailable =
FCT_Absence_Rostered[Budget Sick Hours] --Number of sick leave hours
+
  (
  8                        --Number of hours leave per holiday day
  *
  FCT_Absence_Rostered[Holiday days]  --Number of holidays allowed
  *
```

FCT_Absence_Rostered[Average FTE]) **--Determines contract type**

Holiday hours =

FCT_Absence_Rostered[Holiday days]*FCT_Absence_Rostered[Average FTE]*8

Appendix 5 dashboard design

Page 1

Forecasts/orders pre/after-calculation

Goal of the page:

Identify the difference between estimation and reality

Divide in 3 classes: small, medium, large

Parameters:

- Number of hours estimated
- Number of hours spent

Page 2

Forecasts/orders open/closed/scored

Goal of the page:

Show numerical values regarding forecasts

Parameters

- Average order value
- Throughput time
- Number of forecasts closed, open, and scored
- Success rate of closing forecasts

Page 3-4

Sales Orders this year and next year

Goal of the page:

Provide insight into current status of orders needed to fill

Parameters

- Number of hours to fill per team
- Number orders needed
- Total number orders needed next year

Page 5

Calculations

Goal of the page:

Provide insight into underlying values

Parameters

- Hours rostered, unavailable, to fill, CB, NB
- Order value

Page 6

Dashboard

Goal of the page:

Provide simple overview of other pages and enable creating scenarios of the future

Inputs

- Success rate forecasts and quotations
- Order value
- Value of an hour's work

Parameters

- Forecasts needed
- Orders needed
- Quotations needed

Appendix 6 Questionnaire

My job is:

I have experience with PowerBI and how it works

Performance expectancy

PE-1: I would find the system useful in my job.

PE-2: Using the system enables me to accomplish tasks more quickly.

PE-3: Using the system increases my productivity.

PE-4: If I use the system, I will increase my chances of getting a raise.

Effort expectancy

EE-1: My interaction with the system would be clear and understandable.

EE-2: It would be easy for me to become skillful at using the system.

EE-3: I would find the system easy to use.

EE-4 Learning to operate the system is easy for me.

Attitude toward using technology

ATT-1: Using the system is a bad/good idea.

ATT-2: The system makes work more interesting.

ATT-3: Working with the system is fun.

ATT-4: I like working with the system.

Facilitating conditions

FC-1: I have the resources necessary to use the system.

FC-2: I have the knowledge necessary to use the system.

FC-3: The system is not compatible with other systems I use.

FC-4: A specific person (or group) is available for assistance with system difficulties.

Self-efficacy I could complete a job or task using the system...

SE-1: If there was no one around to tell me what to do as I go.

SE-2: If I could call someone for help if I got stuck.

SE-3: If I had a lot of time to complete the job for which the software was provided.

SE-4: If I had just the built-in help facility for assistance.

Behavioral intention to use the system

BIU-1: I intend to use the system in the next months.

BIU-2: I predict I would use the system in the next months.

BIU-3: I plan to use the system in the next months.

Mijn functie is:

Ik heb ervaring met PowerBI en hoe het werkt

PE-1 Ik vind het dashboard handig voor mijn baan

PE-2 Het gebruiken van het dashboard helpt mij mijn taken sneller te doen

PE-3 Het gebruiken van het dashboard verhoogt mijn productiviteit

PE-1 Als ik het dashboard gebruik, verhoog ik de kansen van een salaris verhoging

EE-1 Mijn interactie met het dashboard is duidelijk en begrijpbaar

EE-2 Het is makkelijk voor mij om ervaren te worden met het dashboard

EE-3 Ik vind het dashboard makkelijk te gebruiken

EE-4 Het leren gebruiken van het dashboard is gemakkelijk voor mij

ATT-1Het dashboard gebruiken is een goed idee

ATT-2Het dashboard maakt werk interessanter

ATT-3 Het gebruiken van het dashboard is leuk

ATT-4 Ik vind werken met het dashboard leuk

FC-1 Ik heb de middelen die nodig zijn voor het gebruik van het dashboard

FC-2 Ik heb de kennis die nodig is voor het gebruik van het dashboard

FC-3 Het dashboard is te verenigen met andere systemin dat ik gebruik

FC-4 Een specifiek person of groep is bereikbaar voor hulp met problemen van het dashboard

SE-1 Ik kan een taak volbrengen als: Een Als niemand aanwezig was om te vertellen wat ik stap voor stap moet doen

SE-2 Ik kan een taak volbrengen als: Een Als ik iemand kan bellen als ik vast zit

SE-3 Ik kan een taak volbrengen als: Een Als ik veel tijd had voor het voltooien van mijn taak waarvoor het dashboard voor gemaakt is

SE-4 Ik kan een taak volbrengen als: Een Als ik een ingebouwde hulp functie ingebouwd heb

BIU-1 Ik heb de intentie het dashboard te gebruiken in de komende 6 maanden

BIU-2 Ik voorspel het dashboard te gebruiken in de komende 6 maanden

BIU-3Ik ben van plan het dashboard te gebruiken in de komende 6 maanden

Appendix 7 Results of questionnaire

ID	Mijn taakbeschrijving is	Ik heb ervaring met PowerBI en zijn opbouw	PE-1		PE-2	PE-3	PE-4	EE-1	EE-2	EE-3	EE-4
1	Manager		2	3	3	2	1	3	3	2	3
2	technisch support	5	5		3	3	3	5	5	5	5
3	Manager	5	3		3	3	2	3	3	2	3
4	Manager	2	5		2	3	2	5	2	2	5

Table 1: Results of the questionnaire

ATT-1	ATT-2	ATT-3	ATT-4	FC-1	FC-2	FC-3	FC-4	SE-1	SE-2	SE-3	SE-4	BIU-1	BIU-2	BIU-3
5	5	3	3	5	5	3	1	1	5	3	5	5	5	5
5	3	5	5	5	5	5	5	5	5	5	1	5	5	5
3	5	5	5	5	1	5	3	3	3	3	3	2	2	2
5	5	5	5	5	5	5	5	1	5	3	5	5	5	5

Appendix 8 variables

Variables calculation orders

- **Total hours:** The total amount of working hours are the hours that are rostered in the database. This number reflects the hours that would be worked if there were no holiday, national ones are taken into account, or sick leave. This is done by using the SUM function over all days in that year for every employee.
- **Hours holiday current:** This is the number of holiday hours the employees use in a year. It works by using the SUM function over all hours that are marked as holiday leave for every employee in a year.
- **Hours holiday future:** This is the expected number of holiday hours an average employee will use. This is based on how many hours were used in the past 3 years and how many employees were employed in that year.
- **Hours sick current:** This is the number of sick hours the employees use in a year. It works by using the SUM function over all hours that are marked as sick leave for every employee in a year.
- **Hours sick future:** This is the expected number of sick hours an average employee will use. This is based on how many hours were used in the past 3 years and how many employees were employed in that year.
- **Hours to fill current:** This is the number of hours that employees are supposed to be effectively working. It is calculated by subtracting the number of hours unavailable from the total number of hours.
- **Hours to fill future:** This is the expected number of hours that employees are supposed to be effectively working. It is calculated by subtracting the number of hours unavailable future from the total number of hours future.
- **Hours to sell:** This is the effective number of hours that need to be sold to customers based on the available hours times the productivity of an employee.
- **Number of orders needed:** This is the number of orders needed for each team based on an average order size and number of hours that needs to be filled. It is calculated by dividing the number of hours of NB by the average order size per team.
- **Forecasts “closed”:** This is the number of forecasts that are finished irrespective if it was a success or not. It is counted by filtering on forecasts that are “closed and not scored” and “closed and scored”.
- **Forecasts “closed and scored”:** This is the number of forecasts that are closed with an order assignment. It is counted by filtering on forecasts that are “closed and scored”.
- **Average hours per order:** This is the average expected number of hours allocated to an order. It is calculated by dividing the average value of an order by the worth of an hour’s work.
- **Average value of an order:** This is the average value of an order. It is calculated by dividing the total value by the number of orders in a year.
- **Total sum of all order values:** This is the total value of all orders in a year. It is calculated with the SUM function.

- **Number of employees next year:** This is the number of employees that are employed next year. Due to the different contracts, only the ones with a permanent contract are counted as those are sure to be employed next year. It is calculated by counting the distinct employee names in the next years presence table.
- **Number of holiday days:** This is calculated as is mentioned in Chapter 5.1.2.
- **Percentage forecasts scored:** This illustrates the success rate of a forecast. It is calculated by dividing the number of forecasts "closed and scored" by the forecasts "closed".
- **Percentage sales quotations scored:** This is the success rate of a quotation. It is calculated by dividing the signed quotations by all quotations written.

Variable calculations

- **Hours unavailable reality:** This reflects the number of hours that employees were not available as experienced in the past. It is computed by adding the hours sick current to hours holiday current.
- **Hours unavailable estimated:** This is explained in Chapter 5.1.2.
- **Hours CB reality:** This determines the number of hours that is really used for CB projects. It is calculated by adding all hours invoiced to projects that do not have a pre-calculation and are commissioned to CB clients.
- **Hours CB invoiced:** As mentioned previously as relevant data. This is the number of hours that are invoiced for projects for CB.
- **Total orders needed:** This is the number of orders that are needed to fill the workflow for the teams: development, ERP, HRM, managing, and specials. It is calculated by adding all separate number of orders needed per team.
- **Forecasts needed:** This is the number of forecasts that is needed to achieve the total orders needed. It is determined by dividing the total orders needed by the success rate of a forecasts.
- **Sales quotations needed:** This is the number of quotations that is needed to achieve the total orders needed. It is determined by dividing the total orders needed by the success rate of a quotation.
- **Total orders needed future:** This is the number of orders that are needed to fill the workflow for the teams: development, ERP, HRM, managing, and specials. It is calculated by adding all separate number of orders needed per team in the next year.
- **Forecasts needed future:** This is the number of forecasts that is needed to achieve the total orders needed future. It is determined by dividing the total orders needed future by the success rate of a forecasts with the possibility to lower or increase it using the input on the dashboard.
- **Sales quotations needed future:** This is the number of quotations that is needed to achieve the total orders needed future. It is determined by dividing the total orders needed future by the success rate of a quotation with the possibility to lower or increase it using the input on the dashboard.
- **Value of an average working hour:** This is the average amount charged for an hour's work. It is calculated by taking the average over all invoiced hourly tariffs.
- **Total hours future:** The total amount of working hours are the hours that are rostered in the database. This number reflects the hours that would be worked if there were no holiday, national ones are taken into account, or sick leave. This is done by using the SUM function over all days in the next year for every employee who has a contract for that period.

- **Hours unavailable future:** This is the number of hours when employees are not available for work due to sick leave or holiday leave. It is calculated by adding the number of hours sick and holiday future together and add the number of hours leave for potential new employees based on the code from Chapter 6.1.2 where it only uses the start of the contract. This means that the number of holiday days an employee has is the number of holidays per month times the number of months the employee is hired for during the next year.
- **Average order size pre-calculation:** This is the average number of hours that is spent on orders based on the estimation of the pre-calculations. It is calculated by adding all number of hours estimated of orders together and divide them by the number of orders.
- **Average order size after-calculation:** This is the average number of hours that is spent on orders based on the estimation of the after-calculations. It is calculated by adding all number of hours used for orders together and divide them by the number of orders.
- **Success rate forecasts:** This is the percentage of forecasts that were "closed and scored" of all forecasts "closed". It is calculated by dividing the "closed and scored" by "closed" and a percentage can be added via an input from the dashboard to allow for testing of different scenarios.
- **Success rate sales quotations:** This is the percentage of quotations that were "offerte is terug van klant" of all quotations. It is calculated by dividing the former by all quotations and a percentage can be added via an input from the dashboard to allow for testing of different scenarios.
- **Order size type:** This gives a name to the size of each project. This is done as written in Chapter 4.2.1.
- **Hours NB:** This is the number of hours that should be spend on new orders from new customers or for orders that get a pre-calculation. It is calculated by subtracting the hours for CB from the hours to sell.
- **Hours NB future:** This is the number of hours that should be spend on new orders from new customers or for orders that get a pre-calculation. It is calculated by subtracting the hours for CB of the current year from the hours to sell future.
- **Travel time:** This is the time that is spent on travelling to clients. It is determined by using the SUM function over all hours recorded that were filtered on travel time.
- **Travel value:** This is the value of all the time that is spend travelling to clients. It is determined by adding all invoiced hours together filtered on that it is travel time.

Argumentation of variables

Most of the previously mentioned variables are self-explanatory. There are some variables that need more argumentation on why and how they are used. They are listed below with their argumentation:

- **CB hours invoiced:** This is the number of hours sold to CB for their projects. As the after-calculation includes both administration and work hours, it was deemed necessary to use the invoiced hours to have a accurate representation on the hours that are spent on CB orders that make revenue. This variable is calculated based on a summation of the hours in the debtor_invoice dataset and the pre-calculation dataset to distinguish between NB and CB.
- **Extended contracts**: This is used to find the gap between the current expectations of the company versus when they extend the ending contracts. It is calculated by performing the

same calculation for orders again but only for the number of employees whose contract is ending.

Number of forecasts/quotations/orders needed: This is a better representation of the actual required amount as the normal calculation would only calculate the number needed on the company level. The change here was that the calculation is performed on the teams level and then added together. This gives a better insight in how the company actually stands with the needs.

Appendix 9 Interpretation of the data

Due to sensitive data the results of the data calculations have been changed by multiplying with random coefficients. These figures do not represent the company in any way. The idea of this section is to explain what the data means in a more general sense. The interpretation will be per panel.

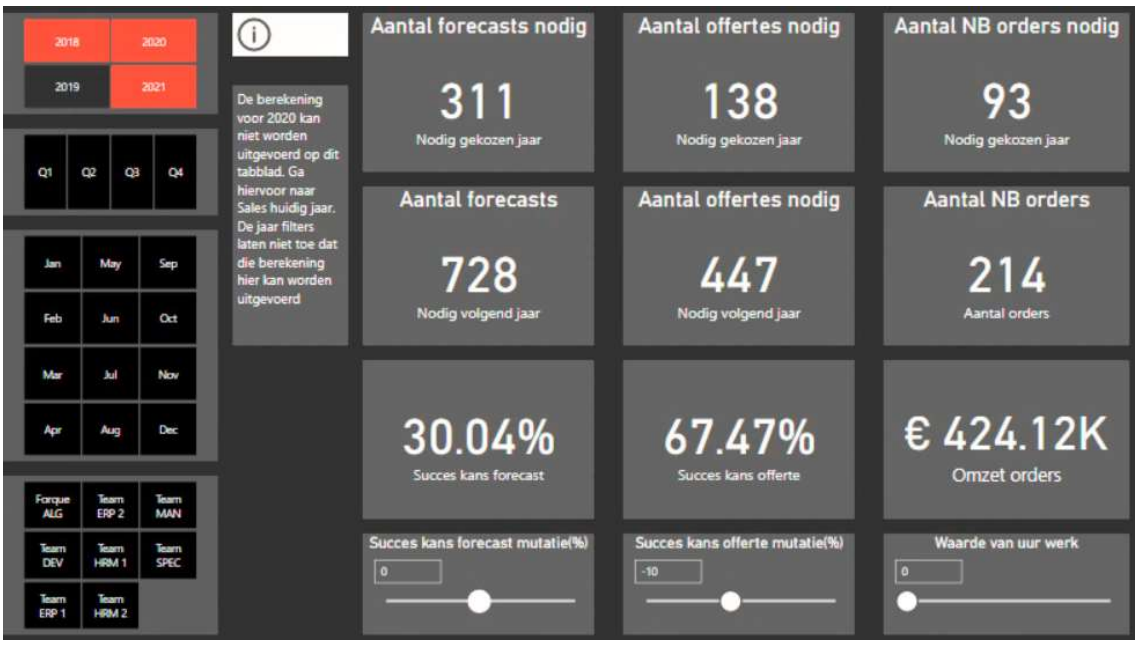

Figure 1: Overview panel

The data in Figure 6 shows how the needs of the company play out in terms of forecasts, quotations, and orders needed. In the selected year, here 2019, the company needs 93 orders which leads to 138 quotations and 311 forecasts. It shows here and in the success rate of 30.04% that forecasts are less likely to succeed than quotations with 67.47%. This shows that the company needs to send many forecasts to fill the orders. When the success rate is high, the workload is low as fewer actions have to be done to have the same number of orders.

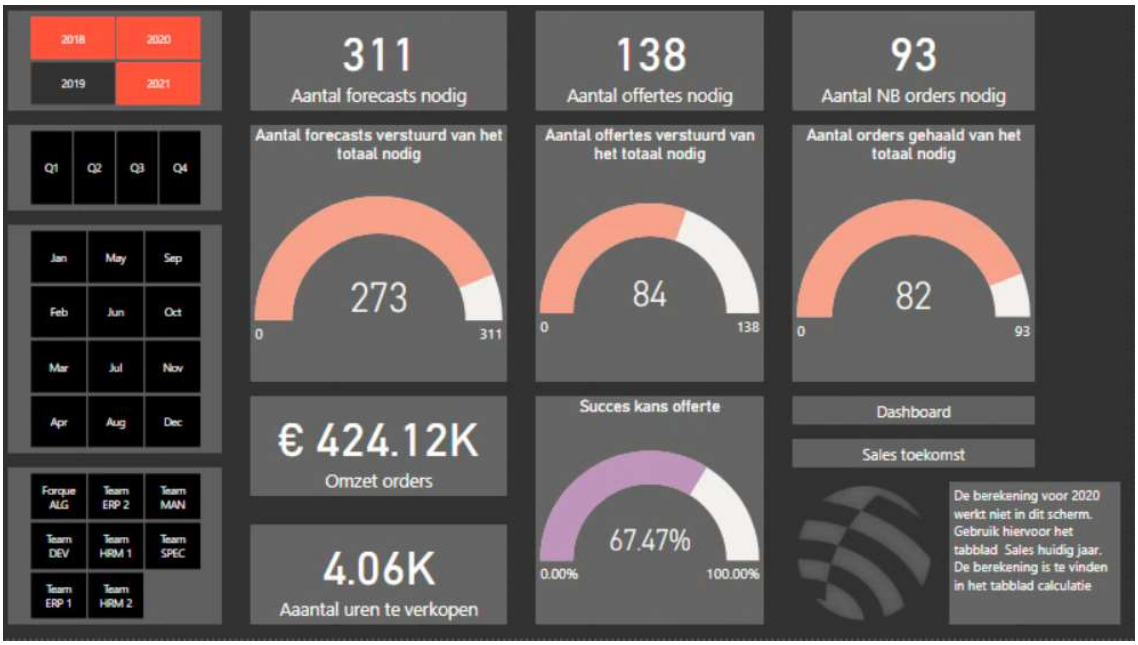

Figure 2: Sales past years panel

The panel in Figure 7 shows how the sales division did in the past. The two half circle diagrams show that the quotations have a success rate that is 67.47%. If this would be higher less work has to be done. The 84 quotations sent of the 138 needed shows that there are quotations that need to be sent. If this number would be close to the target, the company has enough work to fill all hours that people work. The less this number is the more idle time employees have and less revenue is made. The interpretations of the panel in Figure 8 are the same as in the previous panel in Figure 7. The main point of interest here is that there are 116 more orders needed if 7 contracts are extended.

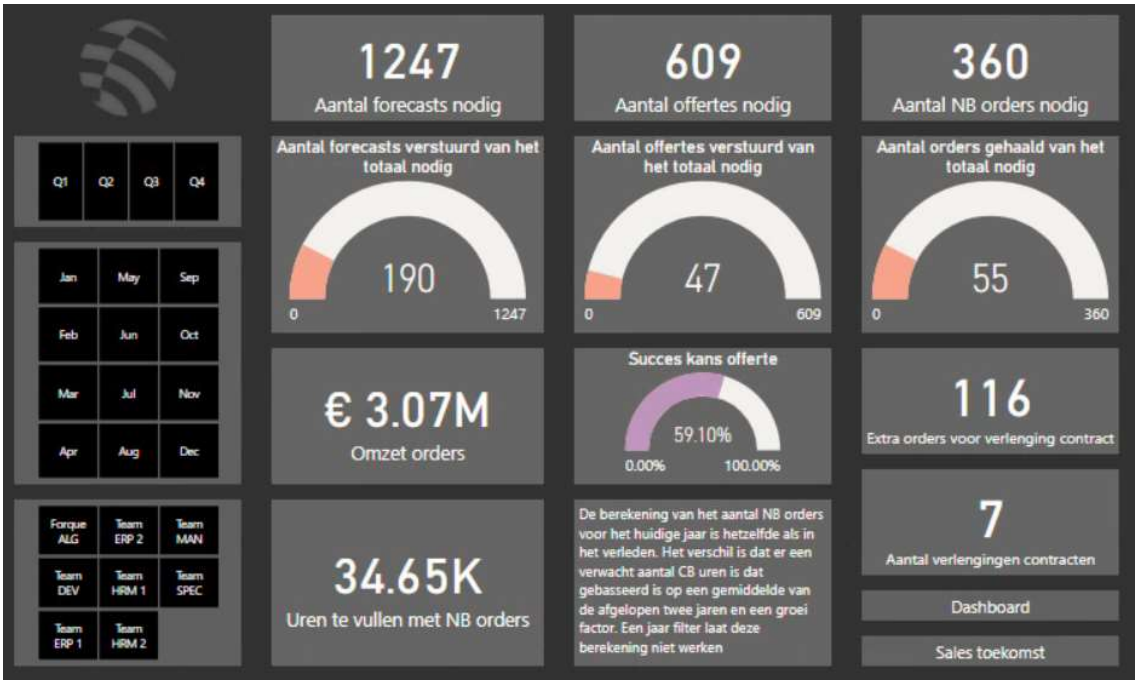

Figure 3: Sales current year panel

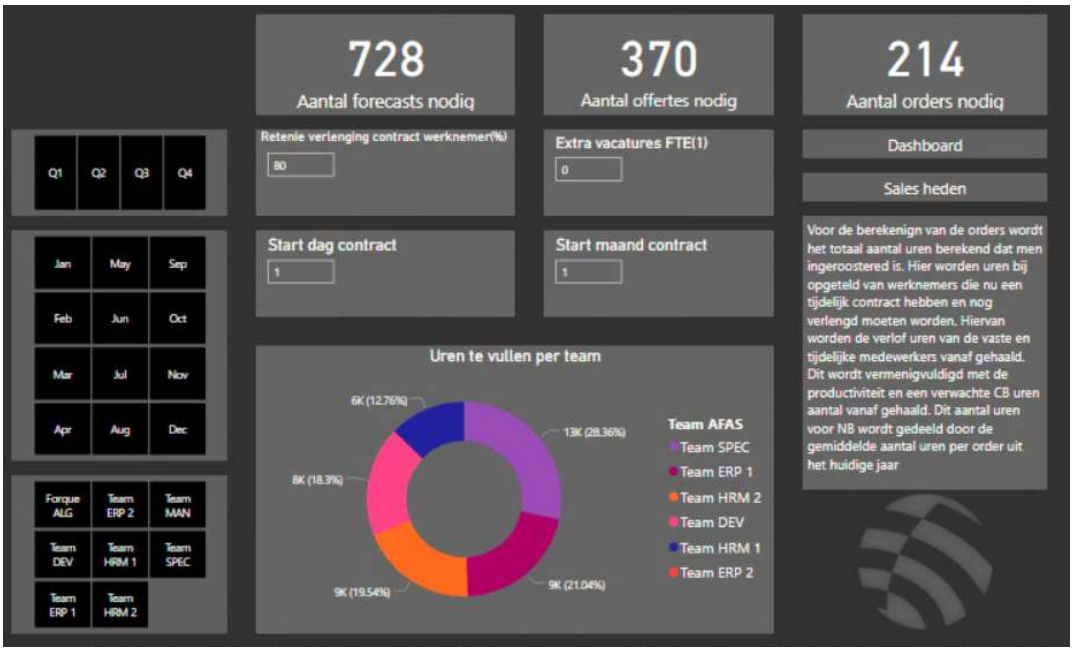

Figure 4: Sales next year panel

In the panel shown in Figure 9, the data shows that the teams special, ERP1 and HRM2 need the most attention to fill their hours. This graph shows that the sales division cannot divide their attention evenly, but that some teams require more than others as they have different amounts of hours. The four input boxes allow the user to create various scenarios with different numbers of employees, different starting dates of new employees, or different number of people who accept an extension of their contract. The outcome of the calculation and the input boxes is shown in the three values at the top of this panel. Here it is shown that forecasts have a low chance of success.

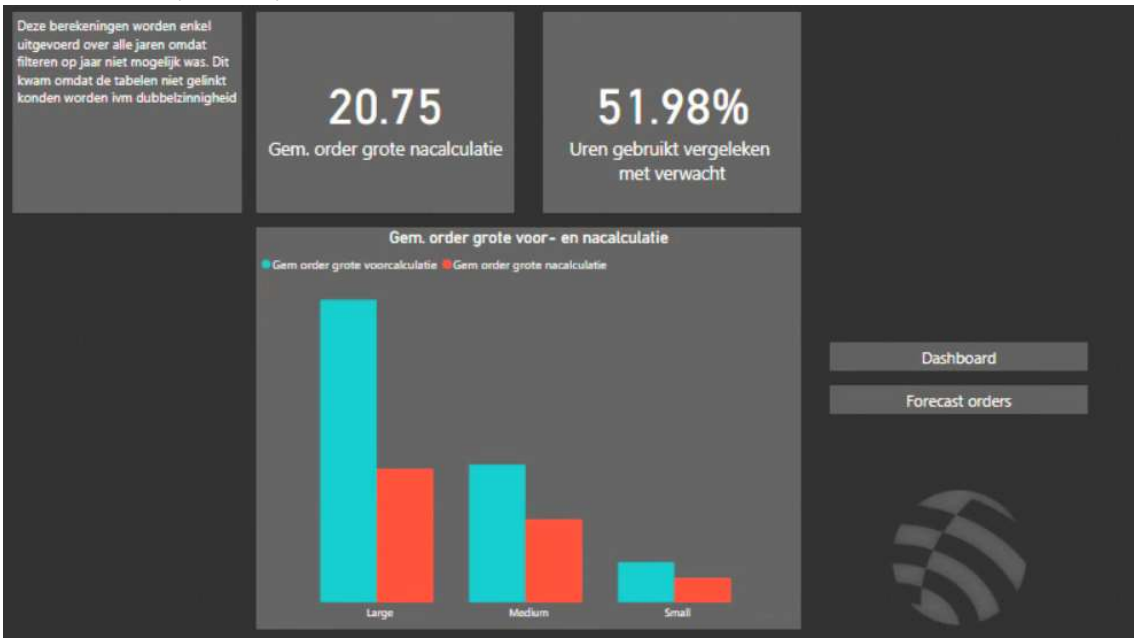

Figure 5: Pre/After-calculation panel

In the panel shown in Figure 10, the data shows the difference between the expected hours spent on projects and the actual hours spent. When the orange bar is higher than the blue one, it means

that more hours are spent than expected. This is also shown in the percentage that is higher than 100%. This is also a good thing as more money is being made than expected. If the orange bar is smaller it means that fewer hours are spent than expected and that more orders are needed to fill the hours. It will be shown when the percentage is lower than 100%. This data can help the sales division make more realistic estimates regarding the size of the projects.

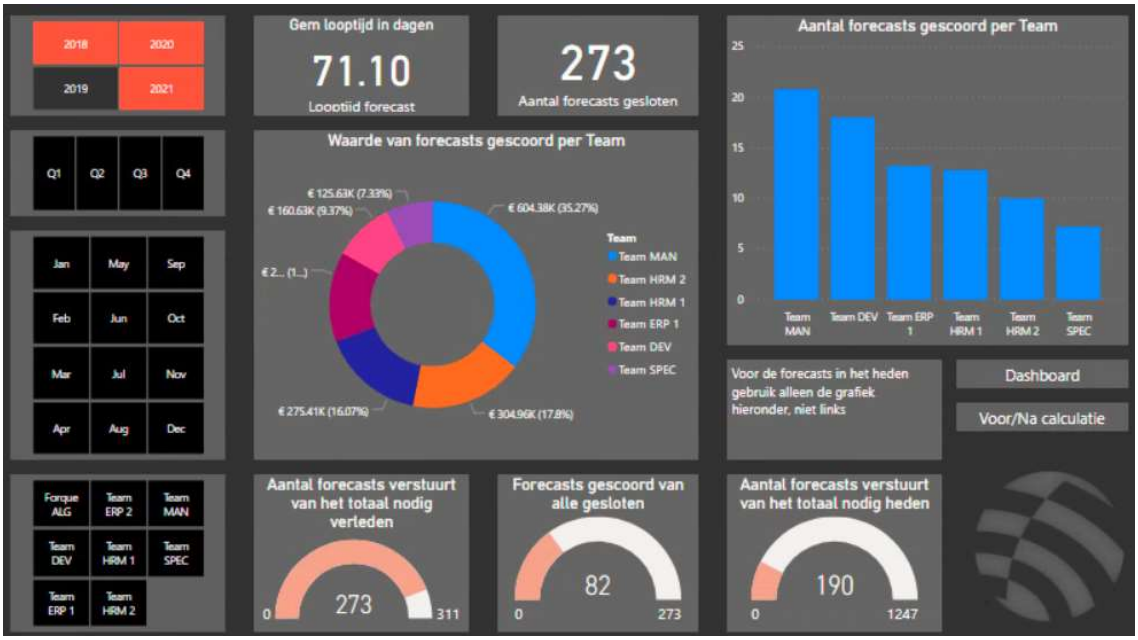

Figure 6: Forecast panel

In the panel of Figure 11, the data shows that an order of team management is on average worth more than an order of team specials. The bigger the part the higher is the average worth in comparison to other teams. In the bottom two half circle diagrams show on the left how many forecasts are sent in comparison to how many should be sent. This should be filled as much as possible to accomplish the order quantity needed, to fill the hours, make enough revenue. The middle half circle diagram shows how many forecasts have ended in success from all forecasts ended. If this bar is filled, then less work is needed as fewer forecasts are needed to achieve the order target.

In the top right corner, the forecasts per team are shown which shows where the most forecasts scored are. This graph in combination with the donut chart show where money can be made as it can point out where many profitable orders are. For example, here the team MAN has both a high average worth and a high forecast scored count. In this case here a decent part of the revenue is being made.

Figure 7: Leave usage panel

In the panel shown in Figure 12, the leave usage is shown. In the two bar charts it is shown how the expected leave usage is in reality. In the top bar chart when the orange bar is smaller than the purple one there are fewer sick leave hours than expected. This means that more orders have to be closed to fill the gap. When the orange bar is higher, fewer orders are needed. In the bottom bar chart, when the orange bar is higher than the blue one, fewer hours have to be filled. When it is lower more hours have to be filled than expected so more orders are needed. In the upper right corner is the average value of an hours work per year. An upwards trend means that the work can be sold for a higher price in the future.

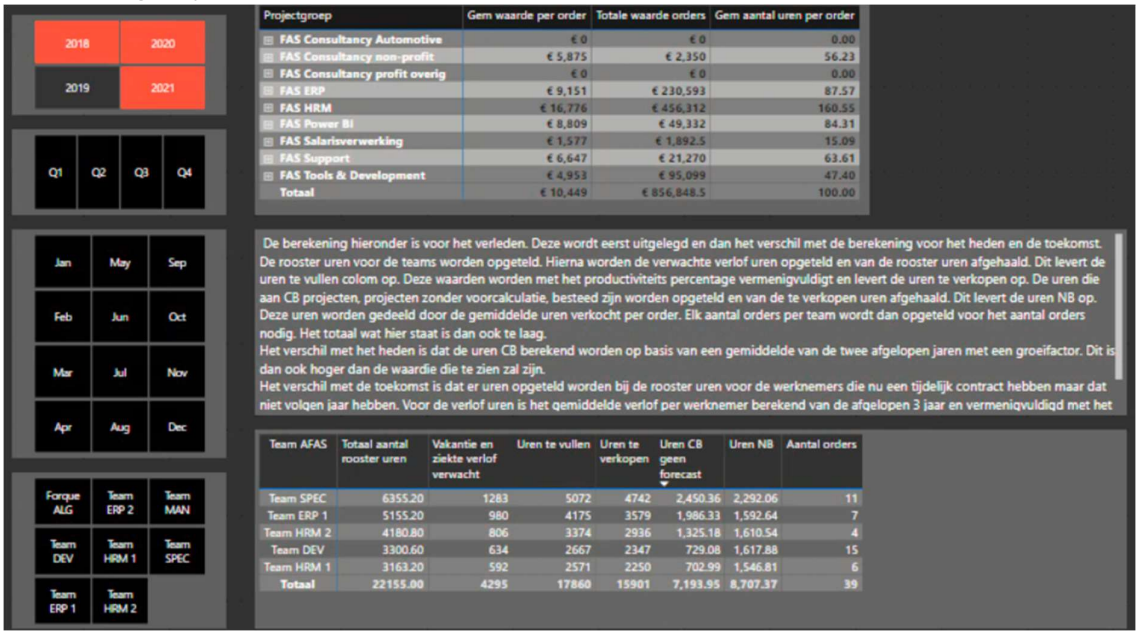

Figure 8: Calculations panel

In the panel shown in Figure 13, the calculations are shown in a table form. In the bottom halve of Figure 13 can a table be seen that functions as visual representation of the calculations per team.

Here in the last column of the table in the bottom halve shows the number of orders needed for that team. The difference between the “Uren te vullen” and the “Uren te verkopen” tells the story of the productivity of the employees. The smaller the difference is between the two columns the more productive the employees are. Productivity here is the percentage of total hours that is being used for invoiceable hours to clients.

Appendix 10 Structure of the model

The model will exist out of different reports. They are a page in PowerBI that shows information of datasets from multiple perspectives. They allow for analysis and visualisation of data. For this project six pages are used that show the story of the Sales and Marketing department. This starts with a sales action continues to forecast and ends with orders. There is one page that functions as a dashboard which has the function of testing possible future cases and show the values that are of importance. The last page is about calculations and is about showing the data behind them. For the designing of these pages, see Appendix 5, where the goals and parameters are conceived. This is a first idea of how the dashboard could look like and is still being developed as the dashboard is designed. This is the more detailed version of the one in Chapter 4.4.

The first page is the overview dashboard. Here the values of the quotations, forecasts, and the orders that are needed to fill next year's hours are displayed. This is the goal of this page, to give a single overview of these key values. Beneath these values are 5 sliders that can manipulate different aspects of the process. They are used to create different scenarios that might happen in the future to help the Sales division prepare. They are listed and explained why they matter below:

- **Value hour:** The value of an hour that an employee works is not the same. There are three different types of employees: consultant, senior consultant, and project manager. This value can change when more project managers are used, thus increasing the average value of an hour.
- **Value order:** The value of an order has an impact on how many orders are required. When the value goes up there are fewer orders needed. This is because an order of a higher value is allowed more hours. This is now depicted in the three classes of order size and this changeable parameter will help prediction efforts.
- **Success rate:** The different success rates have an impact on how many sales are needed for a certain number of forecasts, which in turn influences how many orders are received. By allowing this parameter to influence this, the company can simulate the state of an economy, as in a recession people are less likely to make a deal.
- **More employees and Fewer employees:** These two sliders allow the company to get an insight into what new amounts of sales, forecasts, and orders are needed for different numbers of employees. This helps making human resource management decisions as feasibility can be tested whether more or fewer orders are expected.

The second page is about the state of Sales in the past years. The goal of this page is to give insight in the past states of the sales targets. This is split in two categories sales actions and hours. Both these values are broken down in the target, currently achieved, and left. The graph beneath the table shows the state per team on how much of the target is achieved in matter of percentages.

The third page is the sales of this year. It is similar as the past years, but now on how the company is currently. The main addition is the insight into how the company should react to extending contracts that end in this year. The number of orders that should be added to the current number is displayed here if all contracts are extended.

The fourth page is sales next year. It is similar to the dashboard in function as it shows the numbers of sales, forecasts, and orders needed for the next year. The main difference is that the numbers are broken down into teams and the three classes of order sizes. The numbers are also spread out over the months to show the busier and calmer periods to help the sales division with their planning and focusing.

The fifth page is the forecast pre/after calculation. This page has the function to help the sales division check if their forecasts are close to reality. This is not of direct importance for the calculation of the number of sales, forecasts, and orders but does have an impact if it turns out that the estimates are always below reality. The page has two tables. The first one is on the difference between the hours in the forecast and reality broken down in teams and the three classes of order sizes. The second one is the value of an order broken down in the same way as the hours table. The graph shows the differences as a visual.

The sixth page is the forecast orders analysis. This page shows the numeric insights into the forecasting process. There are 5 values that are highlighted here that are written down below:

- **Forecasts closed:** This is the total number of forecasts that have been made in a specific year. This does not yet differentiate between whether the deal is made or not. This is used in the calculation.
- **Forecasts not scored:** This is the number of forecasts that have been closed but no deal was made.
- **Forecasts scored:** This is the number of forecasts where a deal has been made. They are also called orders and are used in the success rate of the forecast calculation
- **Order value:** The total order value and average value over the year will be calculated here.
- **Throughput time:** The average throughput time per team per size will be examined here to help the planning of the Sales division.
- **Productivity:** This variable is used to keep track of the expected productivity of an employee that is stated at the beginning of a year.

The seventh page is on leave usage. This page shows how many sick and holiday leave hours are used and expected to be used. There is also a graph that shows how the development of the worth of an hour worked was over the past years.

The last page is on calculations. The content is similar to the calculations shown in Table 1, Table 2, and Chapter 4.1.3. They will be ordered in tables with the addition of the three classes of order sizes. The new addition is the variable on sick leave. As the company allows an employee 24 hours of sick leave, the actual use thereof is not analysed. To this extent a table is made that shows the usage of sick leave on a team and company level. This is because sick leave can be pooled together as employees can help each other with tasks. This is especially true within a team.

Appendix 11 Reviewing the performance of the tool

Redacted due to sensitive data

Figure 1: Bar chart of the accuracy of holiday leave calculation

Redacted due to sensitive data

Figure 2: Bar chart of the accuracy of the order forecast

Redacted due to sensitive data

Table 1: accuracy of holiday leave calculations

Redacted due to sensitive data

Table 2: Accuracy of order forecast