Determining the best maintenance strategy per asset using failure rates at Allinq

University of Twente

B.A. (Britt) Koobs

August, 2022

The following report is intended for Allinq Group B.V. and the examiners from the University of Twente.

University of Twente
Industrial Engineering and Management
Postbus 217
7500 AE Enschede
Tel. (053) 489 9111

Allinq Group B.V.
Fahrenheitstraat 27-29
3846 BM Harderwijk
Tel. (0341) 437 100

Determining the best maintenance strategy per asset using failure rates at Allinq

Student
B.A. (Britt) Koobs
s2331160
Bachelor Industrial Engineering and Management
University of Twente

Supervisors
University of Twente
First supervisor:
dr. D.R.J. Prak
Assistant professor
Faculty of Behavioural Management and Social Sciences (BMS)

Second supervisor:
dr. E. Topan
Assistant professor
Faculty of Behavioural Management and Social Sciences (BMS)

Allinq Group B.V.
Jonna Tromm
Junior Data Scientist
Smart Asset Management

Jasper Habermehl
Project manager
Smart Asset Management

Publication date: 30-08-2022

This report was written as part of the bachelor assignment of the Industrial Engineering and Management educational program at the University of Twente.

Preface

Dear reader,

In front of you lies my bachelor thesis which I executed at Allinq Group B.V. as the graduation assignment for the bachelor program Industrial Engineering and Management at the University of Twente. I worked at Allinq from February 2022 until July 2022. This thesis aims on gaining insights into the lifetime distributions and maintenance strategies of different clients.

First of all, I would like to thank Allinq for giving me the opportunity to execute my bachelor assignment at their company. A special thanks to my direct supervisor Jonna Tromm, who guided me throughout my entire research. You always made time for me whenever I needed something, and you really helped me to think a step further. Furthermore, I would like to thank Jasper Habermehl. Your critical view on my assignment gave me lots of insights. Moreover, I also want to thank Jeroen Hiel, Kalle Raatgever and Sjoerd Leensen for their feedback and ideas during the weekly meetings we had. You all have really made me feel part of the team during my internship period.

I would also like to thank Dennis Prak, my first supervisor from the University of Twente. You always provided me with lots of ideas and helped me further whenever I felt like I got stuck, which I really appreciate. Also, your extensive and clear feedback definitely helped me improving my thesis. Thank you for the nice conversations and discussions we had, I really enjoyed working with you. Furthermore, I want to thank Engin Topan for being my second supervisor, and good to mention, for bringing me into contact with Allinq! During the minor, you made me enthusiastic about this topic and I was happy to do my bachelor assignment in this field.

Lastly, I cannot forget to thank my family and friends for the feedback, support, interest in my project and for some distraction that I needed every now and then. Special thanks to my brother, you were always willing to make some time in your busy schedule. Your ideas and help were very valuable to me.

I hope you enjoy reading this thesis!

Britt Koobs
Enschede, August 2022

Executive summary

Introduction

This thesis describes the research done at Allinq Group B.V. as the final assignment of the program Industrial Engineering and Management at the University of Twente. Allinq is a specialist in telecommunication and has an advisory and executive role for its customers. These customers all use sensor data, alarms and tickets to monitor their asset performance. Thousands of alarms enter the system every day, which makes a good maintenance strategy important. The assets are large and contain lots of components which can fail and have a different impact when failing. When a ticket comes in for such a component a service engineer needs to go to the site and check out the damage, which is corrective maintenance. There is also already some preventive maintenance being done. Allinq wants to increase its knowledge in the field of these different maintenance strategies, in order to get a better picture of the differences between these strategies, how different customers use these strategies and how strategies can be used in asset management. From this request on, we found the following core problem which needs to be solved: "There is insufficient awareness of the different maintenance strategies and how they can be used in asset management among Allinq's customers."

Approach

The first stage of solving the problem was determining the scope. After analyzing the ticket data and consulting some employees, it was decided to only focus on the technical buildings of Company X and Company Y and then to make the distinction between the Large and the Regular locations, to look at all tickets together and to zoom in on solely the tickets for the Fans. These assets all ticked the requirements and had similar data so also, the comparison between Company X and Company Y would be possible.

The data was sorted and analyzed, for example by excluding tickets that were clearly not a failure from the research. All tickets were labelled to make the right distinctions, and all corrective and preventive tickets were identified. When the different assets were sorted to date as well, the number of days until the next ticket arose for the asset could be calculated which resulted in the mean time till next failure for all categories.

Now that the current situation was all mapped out. Estimating the parameters for the lifetime distributions was the next step. The Weibull distribution was chosen for this based on literature. Furthermore, the Kaplan-Meier was also estimated, which is also a method used for survival analysis. With the help of Excel calculations and calculations in RStudio, the parameters of the Weibull distribution and the Kaplan-Meier were estimated and the distributions were plotted.

Findings

The next step was the perform analyses with the solutions found. Using the Weibull distributions, the optimal maintenance timing could be calculated, given that the asset has an increasing failure rate. The cost rate formula was used for this, which also needed costs for corrective and preventive maintenance as input parameters. The optimal time T for maintenance was found to be after 16 weeks for the fans of the Large Company X locations. The other categories did not have an increasing failure rate and therefore, would benefit most from solely corrective maintenance.

The goodness-of-fit test performed showed that the Weibull distribution is not a very good fit for all categories. For four out of eight categories the null hypothesis, which says that the sample data follows a Weibull distribution, was rejected. Further analysis can be done to see whether another distribution is a better fit to the data. Other distributions to consider might be the Gamma distribution or the Exponential distribution. However, another, maybe even more plausible, explanation of why the distribution does not fit is because we have aggregated different types of failures. Differentiating into different failures and components, and looking into that separately will most likely end up with more accurate results.

Contents

List of Figures

List of Tables

Glossary

CM	Corrective maintenance
MPSM	Managerial problem solving method
MTBF	Mean time between failures
PM	Preventive maintenance
P-P plot	Probability - Probability plot

Chapter 1

Introduction

This chapter introduces the research. A description of the company is given as well as why this research is performed. Furthermore, the problems that Allinq is dealing with are identified and explained, resulting in the choice of one core problem. After that, the scope of the research is determined. Moreover, the research design and the corresponding research questions are presented.

1.1 About Allinq

Allinq Group B.V. (Allinq) is an international specialist in telecommunication technology. For sixty years now, they have been responsible for the construction and management of information infrastructures in among others, telecommunication. Their headquarters are in Harderwijk, the Netherlands but they also operate in Germany and Denmark.

Allinq distinguishes three main activities: projects, production, and services. The projects are large development assignments like dismantling an old switchboard. For production, one can think of the construction of a fibre network. Services focus on the customer assets.

Smart Asset Management is a department within Allinq that focuses on that services part. Allinq's mission is to have a constantly connected world. Smart Asset Management aims to do that by extracting intelligence from the data of the assets by using data engineering and data analytics. With that, they can constantly provide the customer with intelligence and value from their assets.

1.2 Research motivation

Customers of Allinq, telecommunication companies, almost all use sensor data, alarms, and tickets to monitor their asset performance. These alarms can indicate different types of conditions; a possible problem, a problem that is already playing now, or just a warning. Thousands of alarms come in every day, making a good maintenance strategy crucial for telecommunication companies. A good maintenance strategy leads to lowering maintenance costs, reducing asset downtime, and increasing asset life and effectiveness.

Allinq wants to increase its knowledge in the field of these different maintenance strategies, in order to get a better picture of the differences between these strategies and how this can be used in asset management. Allinq has large amounts of data from various customers about maintenance, sensors and malfunctions. Allinq wants to gain insights in the differences between the customers, which maintenance strategies they use and how this results in asset performance to arrive at the best practices per asset from there.

1.3 Problem identification

In the telecommunication industry, there are all kinds of problems. First of all, it is a very competitive industry which results in no data or information being shared. This results in little knowledge on maintenance strategies and their effects of their competitors. Furthermore, some companies lack data, or do not manage their data correctly. When there is no data, a deliberate maintenance strategy cannot be derived.

All in all, many telecommunication companies lack incentives, knowledge, or the resources to set up good strategies leading into targets not being met. In order to be progressive and to be able to distinguish themselves from competitors Allinq wants to increase its knowledge in this field as well. Therefore, this results in the following action problem:

Allinq's customers do not reach their desired objectives with their current maintenance practices

1.3.1 Problem cluster and the core problem

Now that the action problem has been identified, the next step is to find the core problem. By conducting some informal interviews with multiple employees, the problems going on in this sector have been listed. In order to gain insights in how the problems influence each other, a problem cluster is created [1]. The problem cluster is shown in Figure 1.1. From this problem cluster, problem 1 is identified as the core problem. The solution to this problem provides Allinq with the knowledge that they requested. The core problem is as follows:

There is insufficient awareness of the different maintenance strategies and how they can be used in asset management among Allinq's customers.

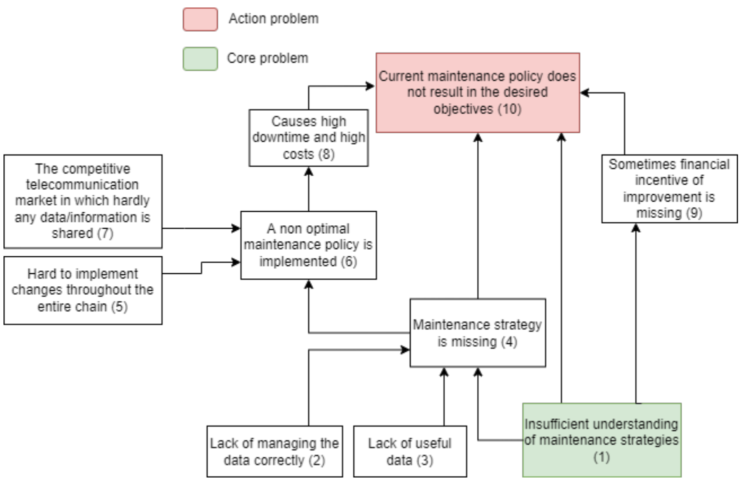


Figure 1.1: Problem cluster

1.3.2 Norm and reality

The current reality is that Allinq wants to enrich their knowledge in maintenance strategies. They hold a valuable position because they have multiple customers and therefore, have the opportunity to compare the different maintenance practices of the different customers. Allinq wishes to gain more insights in maintenance strategies, specifically for certain assets. In this research, per asset category, the best maintenance practice is determined. The norm will then be that they own a clear overview of what strategy works best for what asset and why.

1.4 Research goal

The purpose of this research in to gain insights in the current situation of the customers and how they handle asset management as well as a recommendation on whether it would be more beneficial to change the asset management strategies. The desired level of detail of the study needed to provide this recommendation should be written down as well. Given the research scope in subsection 1.4.1 and the fact that there is not an earlier research that can be elaborated on, this research aims at providing general directions for future research and requirements that should be met to alter the maintenance strategies. Thus, the aim is not to deliver a step-by-step guide to directly implement the possible improvements.

1.4.1 Research scope

The scope of this research is determined in the first phase. In that phase, a selection of certain customers with some of their assets to focus on is determined. Furthermore, given the complexity of the subject, some assumptions are done in the early stage. These limitations are stated in section 6.2.1. Moreover, there is limited time available for this study so it is not possible to analyze all assets of all customers. In the first phase the assets are analyzed to decide which ones are most interesting to look at. Furthermore, implementing the actual recommendations is out of scope of this research. This is not possible to execute within the given time frame.

1.5 Research design and research questions

In order to find the answer to the main research question in a structured way, sub research questions are formulated. Furthermore, the research is split up into different phases, corresponding to the different chapters.

Phase 1: Context analysis
Allinq has multiple customers that all collect data of their assets. However, there are lots of assets so considering all of these assets would be too big of a scope for this research. Therefore, the most interesting assets to look at are selected in this chapter. This is done by analyzing the data of the customers, provided by Allinq. Moreover, an explanation of the asset is given. Furthermore, the current situation on maintenance for the selected assets is plotted.

The input for this chapter is mainly provided by analyzing data and input from employees. To gather input from employees, informal interviews are conducted. The goal of these informal interviews is to see what assets the employees think are interesting to dive into. This results in the first research question:

Research question 1: What assets from different customers can be best focused on for analyzing maintenance strategies based on the current situation?

Phase 2: Literature review
When the scope is determined and the current situation is analyzed, more background information is needed to continue with the research. Theory provides clarification of relevant

terms, statistics and theories. The input for this chapter is mainly provided by scientific articles on time to failure models, maintenance strategies and lifetime distributions.

Research question 2: What maintenance strategies are there for telecommunication assets?
Research question 3: What statistical methods are available for failure modelling?
Research question 4: How is the "time-to-failure" for an asset determined?

Phase 3: Solution design
Once all the fundamental information is collected, the actual analysis starts. In this phase of the study, the lifetime distributions are calculated and fitted to the data set. This is done using the statistical methods found in the literature section and by analyzing the data. The goal is to gain insights in the lifetime distributions of the assets which allows us to continue with other analyses and to draw conclusions. The input for this phase is the data of the customers provided by Allinq and the findings from the literature.

Research question 5: What data within the data set is needed to find the failure distribution?
Research question 6: What does the failure distribution of each asset look like and what does that mean?

Phase 4: Solution analysis
During the solution design phase, the base for the calculations is set. This phase starts with interpreting the results found in the solution design. Also, further analyses are performed such as calculating the optimal time T for performing maintenance based on the lifetime distribution found in the solution design and performing a goodness-to-fit test on the Weibull distributions. These further analyses show the possibilities what is possible to calculate easily once the lifetime distribution is determined.

Research question 7: What is the optimal time T to perform maintenance?

Phase 5: Conclusion
In the last phase of the research, the conclusions are drawn. This is done by looking at all the answers to the different research questions and by looking at the results of the solution designs and tests. Furthermore, recommendations for Allinq are described, the discussion, limitations, scientific and practical contribution and further research are discussed.

1.6 Deliverables

- Determining insights into lifetime distributions and survival probabilities for the selected assets
- Determining the optimal maintenance policies for each of the selected assets
- Report with literature and substantiation of the recommendations and conclusions

Chapter 2

Context analysis

In this chapter, the context analysis for this research is described. This is done by answering the first research question: What assets from different customers can best be focused on for analyzing maintenance strategies based on the current situation? In the first section, the choice for the selected asset is explained. In the next section, more elaboration on the selected asset is provided and this chapter finishes with more context about the data used and the zero measurement based on that data.

2.1 Asset selection

Allinq has multiple clients for which they have an advisory and executive role in terms of handling their assets. In this research, Company X and Company Y will be assessed as these are comparable companies. Both Company X and Company Y collect ticket data of their assets, this is explained later in this chapter. However, there are lots of assets and it will not be possible to consider all of these assets for this research. Several criteria were considered when selecting the most interesting assets to look at:

- *Company X and Company Y must both own such an asset*
  It is not possible to compare the way maintenance and failures are handled if different assets are analyzed that most likely require different maintenance strategies
- *Company X and Company Y must both collect ticket data of the asset*
  If they do not do that, it is not be possible to analyze data and failures.
- *Allinq must preferably own ticket data of the asset of Company X and Company Y in the same period of time*
  Covering approximately the same period of time has two advantages. The same warm and cold months are included. Lots of tickets depend on temperatures so comparing tickets from only cold months with tickets from summer months is not useful. Furthermore, the awareness for using certain maintenance strategies has been growing a lot in the last few years. Most likely, maintenance was different ten years ago than last year. Looking at the same period of time for both Company X and Company Y is most realistic for drawing conclusions about the future.
- *Allinq must own sufficient ticket data of the assets of Company X and Company Y to be able to perform (statistical) analyses*
  In order to, for example, say something about the time to failure of an asset, there must be at least two tickets in the data set of that specific asset and preferably even more tickets.

- *The asset must be interesting and cause impact*
  There are lots of assets and failures of certain assets do not cause too many problems, like having a negative impact on costs for the company or cause major problems for customers. Therefore, for these types of assets it is not even relevant to look at preventive maintenance.

Based on discussions with employees and my supervisors, technical buildings have been chosen. These buildings meet all the requirements and have sufficient data for both Company X and Company Y to perform analyses on. Figures 2.1a and 2.1b show examples of these buildings in the Netherlands of Company X and Company Y respectively.

This picture has been left out due to confidentiality agreements

(a) Company X technical building

This picture has been left out due to confidentiality agreements

(b) Company Y technical building

Figure 2.1: Technical buildings in the Netherlands

2.2 Technical buildings

Technical buildings have been selected as the assets to analyze maintenance data for this research. A technical building is the point at which all of the local telephone circuits converge through a complex system of switching equipment, and where each signal connects to its destination electronically. There are 'large locations' which are national data centers and 'regular locations' which are network nodes or smaller data centers.

According to an internal document provided by Allinq, in January 2020, there were 1395 different Company X technical buildings. Four of these are larger locations. Company Y has about 500 different technical buildings, 13 of these are large locations. These are the largest and most important buildings within this asset category. All signals that are transferred go through one of these larger locations before entering a regular technical building which in its place, transfers it to the consumer at home. These large locations handle so many more signals and connections than the regular technical buildings and therefore, also have a lot more tickets. A major failure at a large location will have a lot more negative impact than for a regular technical buildings. The large locations are designed in such a way that when one would fail, one of the other locations can take over the capacity of the failing location. This is, however, not desired so these locations have teams of a hand full of service men that are always present and continuously perform maintenance. Large locations are categorized under the technical buildings but this difference in size must be noted, and the difference in impact must also be taken into account. Therefore, for further analysis, both the large and regular locations will be reviewed but separately so the differences can be pointed out.

2.3 Tickets and labeling

The data that is used in this study is data about the malfunctioning of assets which are maintained by Allinq. The data about the malfunctioning of the assets is also called ticket data. When such a malfunction of an asset occurs, an alarm is prompted. With the alarm, information is given on why the alarm is triggered. Subsequently, the alarm is converted into a ticket in an ERP system of Allinq. Service employees of Allinq will then take the necessary action on the ticket. The status of the ticket is constantly updated in the system. When the ticket is closed, it gets some labels of which one is whether the maintenance performed has been corrective or preventive. Corrective maintenance is done when problems have occurred already. Until the problem is solved there has been downtime. Preventive maintenance is done to prevent or reduce the probability of problems in the future. More elaboration on these maintenance strategies is provided in Section 3.1.

There are many different reasons why an alarm can be triggered. These reasons vary from loss of power to temperature problems to a door that has been open too long. In order to provide more insights into the grounds for the different tickets, the tickets have been labelled into different categories. These categories and corresponding number of tickets are presented in Table 2.1. These are tickets that have arisen between January 2021 and April 2022. So, for example, Company X had 421 tickets in the category Energy divided over 245 different locations. Thus, some locations had an energy related problem more than once in this given time frame.

Table 2.1: Number of tickets and assets within each label for Company X and Company Y

Labelling of assets				
Labels	Company X - Number of tickets	Company X - Number of assets	Company Y - Number of tickets	Company Y - Number of assets
Energy	421	245	500	256
Emergency power generator	21	21	121	84
Engineering	2	2	356	155
Mechanical engineering	253	185	802	271
Connection problems	1621	16		
Network problems	2938	689		
Hardware problems	1301	160		
SFP, XFP, Patches	1329	330		
Safety			307	74
Security			830	286
Doors and fences			308	165
Other	5429	838	814	185
Non-failure	666	40		

Ticket labels belonging to Company X and Company Y assets are very different so that is why certain categories only have tickets for one of the companies. Overall, from the table it can be concluded that Company X and Company Y have a different ticket system which also results in different kinds of tickets. Company X has many more tickets because it simply has more locations, and thus, more components that can fail over time. On average, Company X and Company Y have almost the same number of tickets per location which makes the comparison doable. Elaboration on the different ticket labels is given below.

Energy: The energy tickets for Company X cover all tickets regarding under or over voltage, power supply issues, battery problems and power failures. For Company Y these tickets mostly include power outages and power supply failures.

Emergency power generator: There are relatively few tickets for this label, however, problems regarding the emergency power generator are significant and therefore, this is mentioned separately. The emergency power generator supports important electrical systems on loss of regular power supply. So, when the energy tickets took place, an appeal was made to the emergency power generator.
Engineering: Company Y has lots of tickets that fall in the engineering category. These tickets go from broken card readers to water detector alarms to leakages. Company X only reported two tickets regarding water logging.
Mechanical engineering: Tickets within the mechanical engineering label are mostly fan or airco related problems.
Connection problems: These tickets include issues with cables, fiber cables, synchronization and (un)link orders.
Network problems: These tickets include issues with the router, switches and glass cords.
Hardware problems: These tickets include problems with blades, servers, DSLAM filters and configurations issues.
SFP, XFP, Patches: SFPs, XFPs and patches problems go often hand in hand. That is why they are grouped.
Safety: Company Y had lots of tickets on the malfunctioning of the fire alarm and other technical alarms.
Security: Security includes the intrusion alarm and alarms of the door not being closed.
Doors and fences: These tickets are about broken locks, broken fences and gate problems.
Other: These are all the tickets that did not fit in one of the existing labels.
Non-failure: These are tickets that require action but are clearly not failures. Most of these tickets say that someone needs access to the site, or to take a package.

For further failure calculations, the non-failure tickets for Company X will be left out. All other tickets will be assumed to be a failure in order to be able to calculate the failure rate. Frequently, this is probably not the case (yet) and thus, we are overestimating the failure rate a bit. Given the data and the time available, it was however, necessary to make this assumption. Like already mentioned in the previous section, the distinction between the larger locations and the regular technical buildings will also be made since they are handled very differently and have significantly more tickets. Furthermore, for further calculations, the focus will be on all tickets as a whole to gain insight in how Company X and Company Y handle tickets and how often they occur. However, it would also be interesting to focus on just one single component. The distributions and analyses of just one component will result in more detailed insights of the behaviour of that certain component than all tickets together. After analyzing Table 2.1, mechanical engineering, the label containing all tickets on the fans, is chosen to be the best component for this. This is because both Company X and Company Y have about the same number of tickets in this category which allows us to compare them later on. Also, fans are important in a technical building because they control the temperature and prevents the building from overheating.

2.4 Zero measurement

As has been noted, the distinction will be made between the larger and the regular buildings, all tickets and only the tickets on one component namely the fans, and lastly, between Company X and Company Y. All in all, this results in eight different categories for which the zero measurement, or the current situation, is presented here.

Table 2.2 shows the current situation per asset category. ‘Number of tickets’ indicates how many tickets have been recorded in total within the category for the time period mentioned before. Only the assets with at least two tickets in the data set have been included, only then the time in between tickets can be observed. So note that all the assets that did not get tickets in this time frame, and the assets with only one ticket are excluded from this research.

Table 2.2: Current situation per asset category

Company	Category	Number of tickets	Percentage of PM done	Average time till next ticket (days)
Company X	All - Large	3917	53.81%	0.4914
	All - Regular	9212	2.92%	33.0506
	Fan - Large	27	54.17%	45.5
	Fan - Regular	77	0.00%	66.3636
Company Y	All - Large	558	25.99%	9.7559
	All - Regular	3480	22.64%	33.4312
	Fan - Large	139	28.06%	17.5038
	Fan - Regular	663	16.89%	51.2179

The fourth column, percentage of PM done, indicates what percentage of the tickets were labelled as preventive maintenance. For now, we still assume that every ticket is a failure, also when it has been labelled as preventive maintenance. This is because although the distinction can be made in the table here, there was still an alarm or warning in the first place which caused the ticket to be prompted. The action that followed from the ticket was later labelled as preventive maintenance. Later in the solution design and the solution tests, methods have been chosen that take these differences in corrective and preventive maintenance actions in consideration. The last column can be read as the average number of days until the next ticket arose on a certain asset. In this study, a ticket is assumed to be a failure, therefore, the average time till the next failure is calculated by taking the average of the days between tickets. This is done to show that it is justified to make the distinction between the larger and the regular locations since here, at first sight, they show large differences. Also, this first observation already gives some interesting insights that we can look into further when the lifetime distribution has been mapped later in this research. In the first category, all tickets of the large Company X locations, the average time to failure is equal to approximately 0.5 which means that on average twice a day a ticket is issued on a certain location. For all tickets on the regular Company X locations, this is on average approximately once a month.

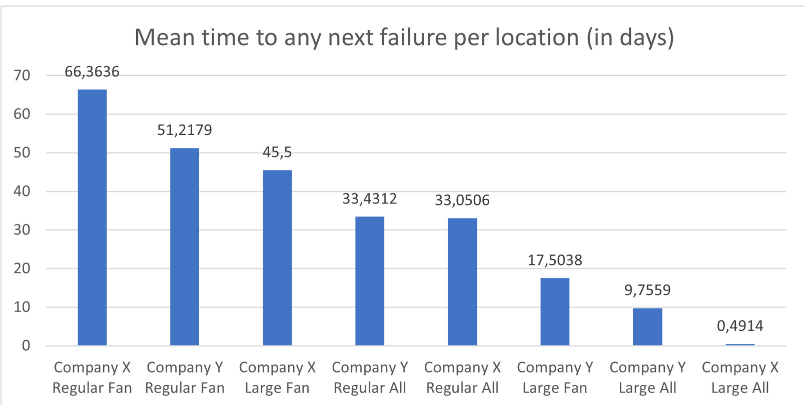


Figure 2.2: Mean time to any next failure per location per category

A couple of the interesting insights in Table 2.2 are that Company Y is a lot more constant in preventive maintenance. Both the larger and the regular locations have some preventive

maintenance. Also, for both Company X and Company Y the larger locations always have a shorter time to failure than the regular locations. This is explained by the fact that larger locations have more components, so also more components that can fail. Furthermore, we record every ticket as a failure of the entire building, while in reality that would not always be the case. More components that can fail within a building would mean that by definition, the building fails more often. Although this may sound very unrealistic, it is still useful to make this assumption since now something can be said about how often 'something is wrong' with an asset.
The mean time to any next failure per location in days are presented in Figure 2.2, sorted from large to small.

Later in this study, in Section 4.1 the lifetime distributions for the different categories are estimated and plotted. These distributions also take into account whether these tickets have been preventive or corrective while these averages do not. In that chapter, these numbers are compared to the results found there.

Chapter 3

Literature review

This chapter focuses on relevant literature to this research. All the concepts and terminology that are used later in the solution design and the solution analysis are reviewed here.

3.1 Maintenance strategies

Corrective maintenance
Corrective maintenance, also called failure-based maintenance, is only carried out after a failure has occurred [2]. Corrective maintenance requires less planning than preventive maintenance. Disadvantages are among others, safety issues that can arise when waiting for something to break down and the unpredictable nature of corrective maintenance.

Time-based preventive maintenance
Time-based preventive maintenance, also called block-based maintenance or constant-interval maintenance, specifies maintenance at a certain interval [3]. The idea is to combine multiple periodic maintenance activities. A replacement cycle is defined by the time between two consecutive preventive maintenance instances. Within a cycle the same picture repeats itself: a preventive replacement at the state of a cycle, and if failure(s) occur, unplanned corrective maintenance(s) until the next cycle [4].

Condition-based preventive maintenance
Condition-based preventive maintenance, also called use-based maintenance or age-based policy, suggests that the total operating time since that previous maintenance action exceeds a preset limit [3]. Both preventive maintenance strategies are based on the renewal theory. The condition-based strategy is the 'extended' time-based strategy. The difference is in the duration of the maintenance interventions. For age-based maintenance, the objective is to find the optimal time T that minimizes the average cost per unit of time. A disadvantage of the age replacement policy is that the replacement times for different items are not synchronized, and one needs to keep track of each item age. This is, therefore, complex to manage if systems have a high number of items [4].

In for example, [5], also different components of telecommunication equipment is analyzed and optimal maintenance strategies have been found as well. The component fan is also looked at in this study. For the fan, an increasing failure rate is found and predictive maintenance would be recommended. In [6], the research concludes that proper maintenance strategies are missing however, are proved to be crucial for the performing of their telecommunication facilities.

3.2 Failure modelling

3.2.1 Failure behavior and failure rate

The cumulative probability of failure is the probability that the product will fail under fixed operational time, which is denoted with $F(t)$ and see Equation 3.1 where T is the lifetime of an asset and t is the time [7]. The probability density function of the first failure is described as $f(t)$. The relationship of $f(t)$ to $F(t)$ is described in Equation 3.2.

$$F(t) = \begin{cases} P(0 < T \leq t) & (t \geq 0) \\ 0 & (t < 0) \end{cases} = \begin{cases} \int\limits_0^t f(t)dt & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{3.1}$$

$$f(t) = F'(t) \tag{3.2}$$

Reliability is the probability that a product will operate properly for a specified period of time under the designed operating conditions. Reliability is a function of time t and can be described as Equation 3.3.

$$R(t) = 1 - F(t) = \begin{cases} P(T > t) & (t \geq 0) \\ 1 & (t < 0) \end{cases} = \begin{cases} \int\limits_t^\infty f(t)dt & (t \geq 0) \\ 1 & (t < 0) \end{cases} \tag{3.3}$$

From Equation 3.3, it can be derived that:

$$R(t) + F(t) = 1 \tag{3.4}$$

By analyzing historical data, it can be shown that an item will break down on average after a specific time. The expected number of times that an item fails in that specified period of time is defined as the failure rate λ [8]. The failure rate is the limit of the probability that a failure occurs per unit time interval t and the conditional probability which can be expressed as Equation 3.5 [7]. For constant failure rates, the failure rate is the inverse of the mean time between failures (MTBF) value which can be expressed as Equation 3.6.

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq +\Delta t | T > t)}{\Delta t} = \frac{\lim\limits_{\Delta t \to 0} [F(t + \Delta t) - F(t)]/\Delta t}{R(t)} = \frac{f(t)}{R(t)} \tag{3.5}$$

$$MTBF = \frac{\sum_{i=1}^{n} t_i}{n} = \frac{1}{\lambda} \tag{3.6}$$

The bathtub curve is an interesting and common concept to use when analyzing failure behavior [9]. The bathtub curve of a non repairable asset can be seen in Figure 3.1.

The bathtub curve can describe the variance of failure rate. The first part, here called 'Burn-in' but also known as 'Early failure' or 'Start-up problems', shows a decreasing failure rate. The second part, 'Useful life', 'Stable period' or 'Random failure', is a constant failure rate. The last and third part is an increasing failure rate, known as 'Wear-out'. The bathtub curve shows that usually, in the beginning a product's failure rate is high due to early failure of components. When the early failures have been discarded, the product is in a stable condition with a low and constant failure rate. During this phase, failures can be caused by random factors. In the last

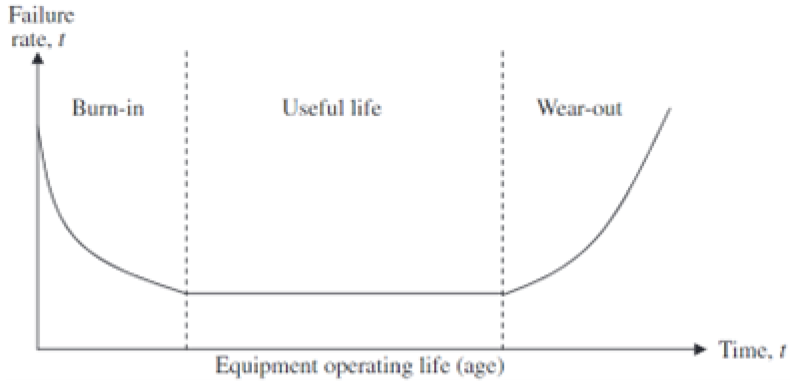


Figure 3.1: The bathtub curve

phase, the failure rate increases because of gradual wear and tear by the products maturation. Later in this research, we will see that the failure rates have influence on what maintenance strategies would be beneficial.

3.2.2 Statistics

The Weibull distribution

The Weibull distribution is the most generally used distribution to model lifetimes and has been considered to provide a good picture of lots of types of lifetime data [10]. A Weibull distribution can be considered with either two or three parameters, a shape parameter k, a scale parameter λ and sometimes an additional location parameter θ. The location parameter is commonly used when there is a failure-free free period, however, in this research the location parameter will be valued zero as there will not be any preventive maintenance during this failure-free period. The distribution function of the Weibull distribution is given in Equation 3.7.

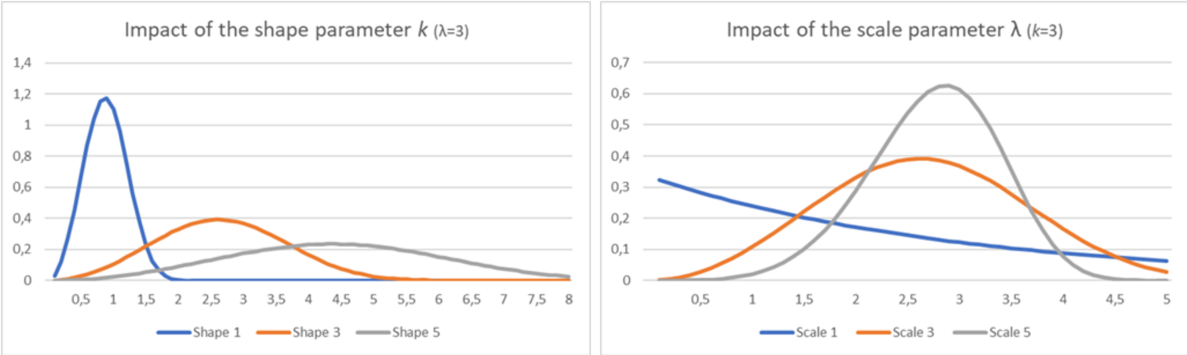
$$F(t) = 1 - e^{-(\frac{t}{\lambda})^k} \tag{3.7}$$

$$z(t) = \frac{f(t)}{R(t)} = -\frac{R'(t)}{R(t)} = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \tag{3.8}$$

Weibull failure rates are given in Equation 3.8. From these Weibull failure rates it can be concluded that a Weibull distribution with a shape parameter smaller than 1 has a decreasing failure rate, indicating that preventive maintenance will not be beneficial. When the shape parameter is equal to 1, the failure rate is constant. When the shape parameter is larger than 1, there is an increasing failure rate.

Figure 3.2a and Figure 3.2b show the what variation in the value of the shape and scale parameter respectively do. It can be seen clearly that with changing one value, the entire graph changes significantly.

In order to be able to perform Weibull analyses on a set of lifetime data, the parameters need to be estimated. This can be done with various methods such as Methods of moments and Regression analysis. In this research, the Maximum Likelihood Estimation will be used to estimate the parameters of the Weibull distribution with both the help of Excel and R. With these parameters, the graphs can be plotted and conclusions can be drawn. The mean of the Weibull distribution can be calculated using Equation 3.9 with t_0 being equal to zero in this case



(a) Variation of the shape parameter

(b) Variation of the scale parameter

Figure 3.2: Impact of the parameters on the Weibull distribution

[4]. The variance can be calculated using Equation 3.10. $\Gamma(r)$ is necessary to solve Equation 3.9 and Equation 3.10, and can be computed using Equation 3.11.

$$E[T] = t_0 + n\Gamma\left(1 + \frac{1}{k}\right) \tag{3.9}$$

$$Var[T] = \lambda^2 \left\{\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right\} \tag{3.10}$$

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx, r > 0 \tag{3.11}$$

The Kaplan-Meier estimator

Next to the Weibull distribution, there is also the Kaplan Meier method which is one of the most popular methods used for survival analysis [11]. The Kaplan-Meier method is a popular method within the medical field but can also well be used for the survival analysis of assets. In [11] it is emphasized that the primary aim of survival analysis is the modeling and analysis of ‘time-to-event’ data, which is, data that has as an end point, the time the event occurs. In this research, the event is the failure occurring. A striking feature of the Kaplan-Meier method is that it takes into account censoring. A censored duration means that the duration does not end with an actual failure. In the context of failure behaviour, it means that preventive maintenance has been performed and thus, no failure is seen which would have been the case for corrective maintenance. Another potential source of censored durations is the end of the observation window without a failure or preventive maintenance. Either way, they are incomplete observations. An uncensored duration is also called an event duration. The event is in this context the failure happening. Thus, the Kaplan-Meier method is a suitable technique for determining appropriate estimates of the MTBF and the reliability function when there is censored data.

The Kaplan-Meier estimator involves computing the probabilities of an event occurring at the certain moment in time. These consecutive probabilities are then multiplied by any earlier computed probabilities to end up with the final estimate. The estimator of the survival function $S(t)$ is given by Equation 3.12 [12].

$$\hat{S}_t = \prod_{i:\, t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{3.12}$$

with t_i being a time when at least one event happened, d_i the number of events that happened at time t_i and n_i the assets that have not yet had an event or been censored up to time t_i.

For this research, a Kaplan-Meier estimator of the reliability function will be made. Remarkable about this estimator is that it exclusively contains horizontal and vertical line segments. This is because the number of 'surviving' assets remains the same in between events. A major difference of the Kaplan-Meier estimator compared to the Weibull distribution is that the Kaplan-Meier estimator is non-parametric while the Weibull distribution is parametric. This means that the Weibull distribution is summarized to two or three parameters, which is not the case for the Kaplan-Meier estimator. The Kaplan-Meier estimator is often used to see whether the parametric models are appropriate and therefore, gives a good impression of what the lifetime distribution looks like. An example of what a Kaplan-Meier estimator can look like can be found in Figure 3.3. In this example graph, the asset has a survival probability of 0.50 left after approximately 10 weeks. The lighter red part around the line is a 95%-confidence interval of the Kaplan-Meier estimator. This interval shows that with 95% certainty we can conclude that the line of the Kaplan-Meier estimator should be within this lighter red area. The function used to estimate the Kaplan-Meier in this case, also directly calculates this confidence interval.

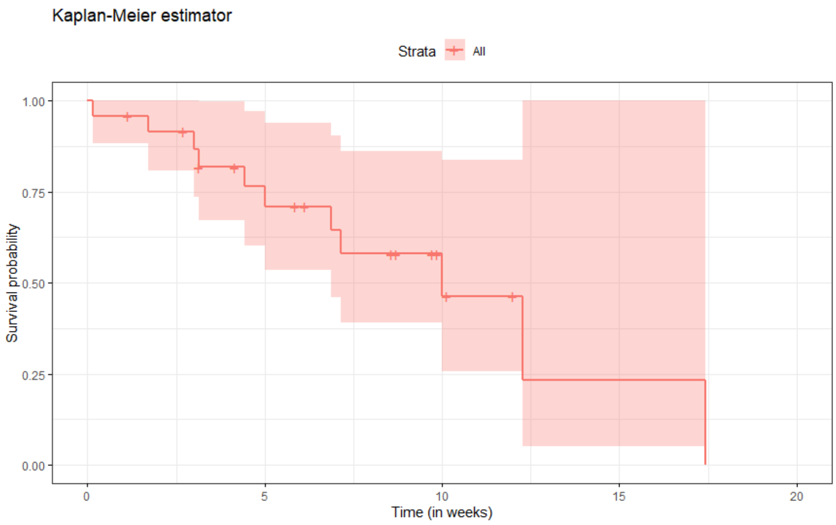


Figure 3.3: Kaplan-Meier estimator

3.3 Finding the optimal time T for maintenance

3.3.1 Finding the optimal time T

In case we find that the assets or some of the assets that we analyze have an increasing failure rate, preventive maintenance is then beneficial. In that case, we want to do a recommendation with what intervals the maintenance should be performed. These intervals can be calculated with the following method.

A maintenance optimization problem involving only one component and an increasing failure rate in time is finding the optimal T for performing maintenance [13]. The goal here is to find the best maintenance schedule for the component over an infinite time horizon using the Weibull distribution and the costs for corrective and preventive maintenance.

$$cost\ rate\ T = \frac{C_{cm}F(T) + C_{pm}(1 - F(T))}{\int_0^T (1 - F(t))dt} \tag{3.13}$$

Equation 3.13 shows how to calculate the cost rate as a function of the cumulative distribution function $F(T)$ for component failure, the corrective maintenance costs C_{cm}, preventive maintenance costs C_{pm} and preventive maintenance interval T [14]. $F(T)$ can be determined using Equation 3.7 from the Weibull distribution. On the computation of the corrective maintenance costs and the preventive maintenance costs will be elaborated in Chapter 4.

3.3.2 Maintenance costs

In case of corrective maintenance, an item is broken and there is downtime. Not only the replacement or repair of the broken component is part of the corrective maintenance costs, also the fact that there is downtime is costly. Therefore, the corrective maintenance costs consists of costs of downtime and the costs of sending a service employee who fixes the failure.

In case of preventive maintenance, there has not been a failure yet. Therefore, there has also not been downtime. There are only costs involved for the service employee going to the site and the replacement or repair of the components.

Overall, there are large differences in costs for downtime. If an item has a lot of impact and is therefore of major importance, downtime is very expensive and more preventive maintenance will be likely. In case of a low impact component or a back-up component for example, preventive maintenance will not be likely. Downtime costs can be calculated by assessing the profit that is lost due to the downtime, but other factors can be taken into account as well if relevant. For example, damage to the reputation because of downtime or damage to other components.

Chapter 4

Solution design

In this chapter the design of the solution is presented. All the steps, concepts and calculations that we executed to get to the solution are explained. The lifetime distributions are mapped by estimating both the Weibull distribution and the Kaplan-Meier estimator for all eight categories.

4.1 Failure distributions

In Chapter 2, some things have already been concluded about the average time until the next failure. In order to gain more insights into the failure behaviour of the different assets, the failure distribution needs to be defined. As described before, a Weibull distribution is assumed as the failure distribution. As mentioned in [10], which is a study about the optimal age-based maintenance strategy, the Weibull lifetime distribution is commonly used and appropriate for modeling lifetimes of a wide variety of units and systems. Like mentioned in Chapter 3.3.2, in this research a two-parameter Weibull distribution is used because it is assumed that the location parameter is equal to zero. Later, in Chapter 5 a goodness-of-fit test will be executed in order to see if the Weibull distribution really is a good fit to the data sets. Now, the shape parameter k and the scale parameter λ need to be estimated. The formula for the Weibull distribution is already given by Equation 3.8 but for clarity presented again below. The location parameter t_0 is thus, equal to zero.

$$z(t) = \frac{f(t)}{R(t)} = -\frac{R'(t)}{R(t)} = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \tag{4.1}$$

The parameters of the Weibull distribution are first being estimated in Excel according to the Maximum Likelihood Estimation [15]. In Figure 4.1, a screenshot is shown of the computations done in Excel. This is a screenshot of the data set for the Company X Fan Tickets for the larger locations. The data in the first two columns show all the tickets in the category and they have been sorted on the different locations and the dates from old to new. The third column calculates the difference, or the duration until the next ticket arose, in days between the tickets. These are thus, the 'times to ticket' but given the assumption that a ticket is a failure, 'times to failure'. These durations are the basis of all further calculations done in this research but the options are not even limited to the calculations done here. These durations can also for example, show the differences between Company X and Company Y, differences between locations and testing whether all locations have the same lifetime distribution.

Next, the durations have been put together in column H, note that the unit changed from days to weeks. The column 'Censored' indicates whether the tickets were labelled as either corrective maintenance (Censored = "No") or preventive maintenance (Censored = "Yes"). For Column

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Location ID names have been left out due to confidentiality agreements	Date	Difference in days					k	1,313396				
2		5-1-2021						Lambda	12,51415			Sum(Log(LLH)):	-17,220489
3		6-1-2021	1										
4		18-1-2021	12										
5		9-2-2021	22										
6		3-3-2021	22					**Duration (in weeks)**	**Censored**	**f**	**R**	**LLH**	**Log(LLH)**
7		11-5-2021	69					0,1	No	0,02576336	0,99719380	0,025763355	-1,5889976
8		21-6-2021	41					1,7	No	0,05230285	0,92916224	0,052302846	-1,2814747
9		20-7-2021	29					3,1	No	0,05783566	0,84969513	0,057835656	-1,2378043
10		6-9-2021	48					3,1	Yes	0,05783566	0,84969513	0,849695126	-0,0707369
11		6-1-2022	122					9,9	Yes	0,04688985	0,48146901	0,481469013	-0,3174317
12		25-1-2022	19					5,9	Yes	0,05720515	0,69146863	0,691468633	-0,1602275
13		2-2-2022	8					4,1	Yes	0,05872942	0,79126766	0,791267664	-0,1016766
14		11-4-2022	68					6,9	No	0,05521207	0,63520980	0,055212074	-1,2579659
15		31-3-2021						17,4	No	0,02483522	0,21329896	0,024835224	-1,6049319
16		25-6-2021	86					2,7	Yes	0,05683702	0,87428370	0,874283695	-0,0583476
17		25-8-2021	61					1,1	Yes	0,04748032	0,95778075	0,957780748	-0,0187339
18		4-11-2021	71					9,7	Yes	0,04732824	0,48819890	0,488198901	-0,3114032
19		8-3-2021						12,3	No	0,03931648	0,37678014	0,03931648	-1,4054254

Figure 4.1: Screenshot of the parameter estimations in Excel

J and K both the WEIBULL.DIST function within Excel is used, for column J the cumulative option in the formula was FALSE since it requires the Weibull probability density function. Column K requires R, given that $R = 1 - F$, the formula was 1 minus the Excel function in which the cumulative option was TRUE. Lambda and k were both given the value 1 as initial input parameters. Column L checks with an IF function whether the duration is censored or not. In the case that it is not censored, it is an event duration, the density function in Column J is used. When it is censored, the reliability function in Column K is used. Column L thus presents, for every duration, the likelihood for that duration given the parameters of the Weibull distribution. The product of all these likelihoods per duration is the likelihood of the complete data set. However, this number can get extremely small. To avoid this problem, the logarithm of the likelihoods is maximized. Given the fact that the logarithm of a product is equal to the sum of the logarithms, the sum of all logarithms of the likelihoods is maximized. In Column M the logarithms of the likelihoods are computed, and Cell M2 shows the sum of all values under 'Log(LLH)'. Using the Solver function in Excel, the sum of logarithms of likelihoods is maximized by changing the values for Lambda and k. This results in the optimal values for Lambda and k.

This computation is repeated for all eight categories. The results are summarized in Table 4.1.

Table 4.1: Estimated Weibull parameters of all eight categories

Company	Category	Scale parameter λ	Shape parameter k
Company X	All - Large	0.1530499	0.9553772
	All - Regular	4.08374362	0.7003266
	Fan - Large	12.514572	1.313695
	Fan - Regular	5.6074605	05460265
Company Y	All - Large	1.9935324	0.7770662
	All - Regular	6.5531077	0.7412759
	Fan - Large	3.3683081	0.7889358
	Fan - Regular	8.6115308	0.7953795

In order to check the results from the Excel computations, and to more easily compute the Kaplan-Meier estimator, RStudio is used. With a few lines of code, see Appendix A, the Weibull distribution parameters are estimated again and the Kaplan-Meier is estimated. The result is a plot which shows both the Weibull distribution and the Kaplan-Meier estimator according to the estimated parameters, together with a 95% - confidence interval for the Kaplan-Meier estimator. A confidence interval is used to indicate how confident we are of an estimated value.

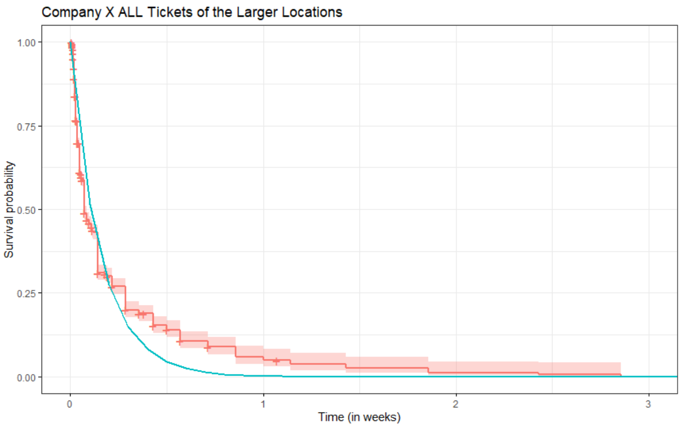
This confidence interval shows us that with 95% confidence we can say that the line should be within the range of the lighter red part. The y-axis shows the survival probability and the x-axis shows the time horizon in weeks.

The blue line is the Weibull distribution and the red line is the Kaplan-Meier estimator, the lighter red part indicates the confidence interval of the Kaplan-Meier estimator.

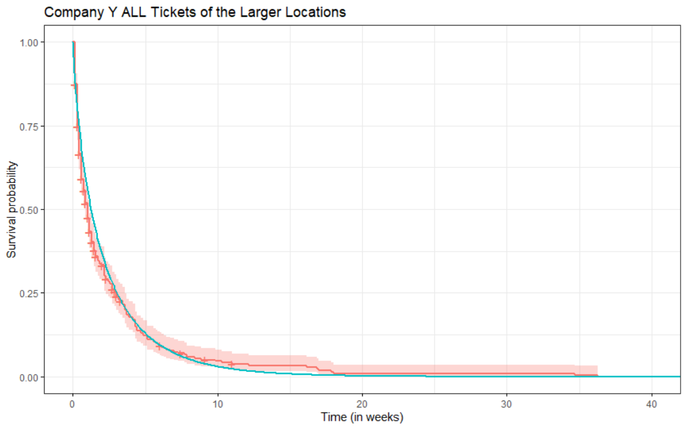
Subsections 4.1.1 and 4.1.2 show these plots for all eight categories. Subsection 4.1.1 presents the plots of all tickets of the different locations and subsection 4.1.2 presents the plots for only the fan tickets of all locations. In Section 5.1, the comparison and interpretation of these plots is discussed.

4.1.1 Distributions of All tickets

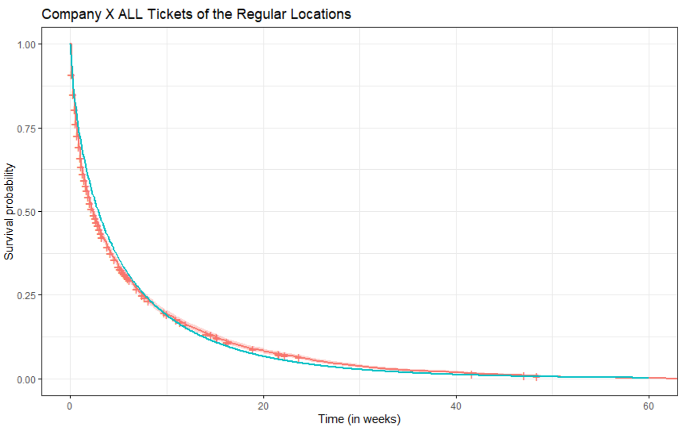
Figure 4.2 shows the Weibull distributions and the Kaplan-Meier estimators for all tickets in total. The y-axis shows the survival probability and the x-axis represents the time in weeks.



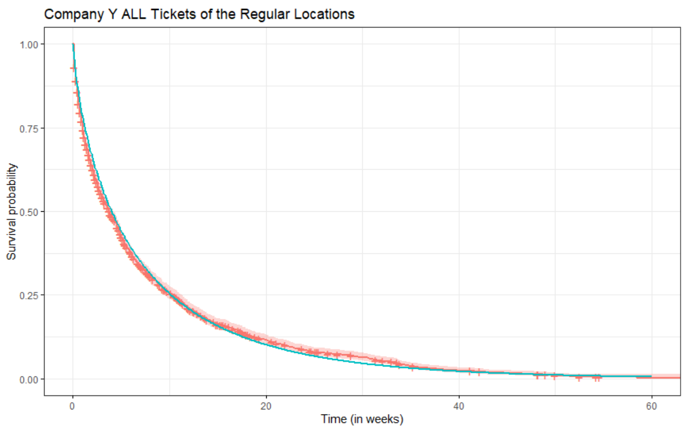
(a) Company X All tickets of the Larger locations



(b) Company Y All tickets of the Larger locations



(c) Company X All tickets of the Regular locations



(d) Company Y All tickets of the Regular locations

Figure 4.2: All 'All tickets' plots

In the first plot, Figure 4.2a, the lifetime distribution of the large locations of Company X is plotted. In Section 2.4, it was already concluded that these locations have a lot of tickets, multiple per day. That is also reflected in the graph. After one week, the survival probability would be zero when no maintenance is done. The 0.50 survival probability is after about 1 day. This can be interpreted as after one day, there is a 50% probability that the location is still performing optimally when no maintenance has been done or after one day, 50% of the locations are still performing optimally when no maintenance has been done.

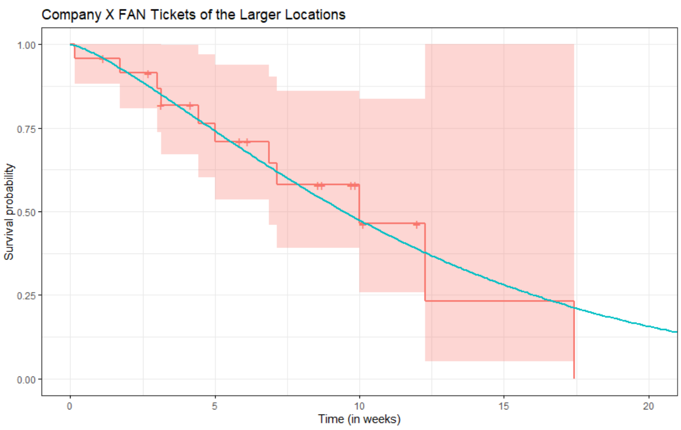
The second plot, Figure 4.2b, shows the life time distribution of the large locations of Company Y. These locations have significantly fewer tickets than the large locations of Company X. Here the 0.50 survival probability is after about 1 week.

The third plot, Figure 4.2c, presents the life time distribution of the regular Company X locations. For these locations, the surival probability is 0.50 after approximately 3 weeks. Thus, within Company X, there is a huge difference between the large and the regular locations.

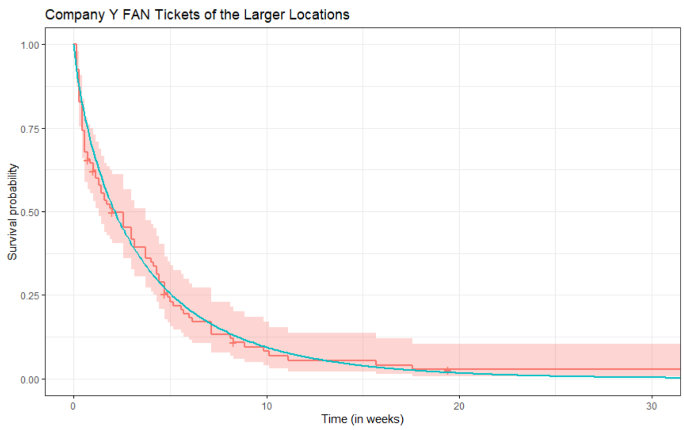
The fourth and last plot in this category, Figure 4.2d, shows the life time distribution of the regular Company Y locations. This one looks quite like the regular Company X locations and has its survival probability of 0.50 after about 4 weeks.

4.1.2 Distributions of Fan tickets

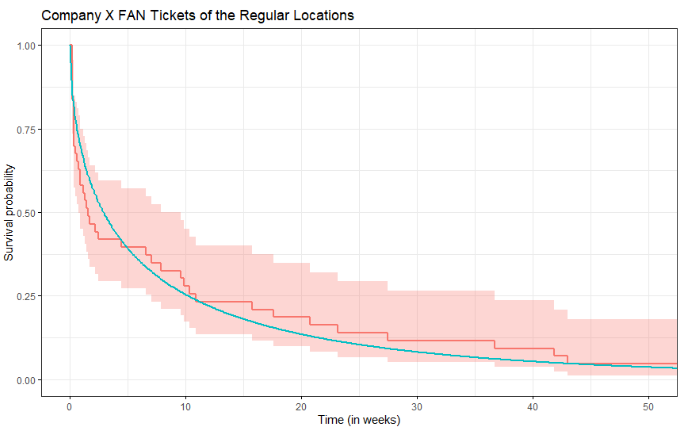
Just like in the previous Subsection the Weibull distributions and the Kaplan-Meier estimators are plotted, but only for the fans this time, see Figure 4.3. Similarly, the y-axis shows the survival probability and the x-axis represents the time in weeks.



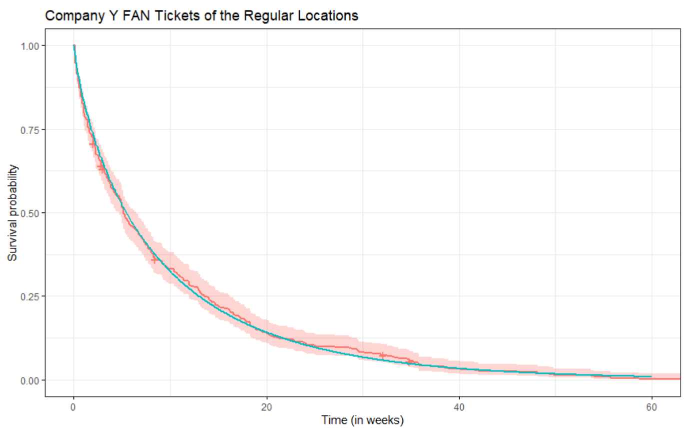
(a) Company X Fan tickets of the Larger locations



(b) Company Y Fan tickets of the Larger locations



(c) Company X Fan tickets of the Regular locations



(d) Company Y Fan tickets of the Regular locations

Figure 4.3: All 'Fan tickets' plots

The first plot presents the fans of the large locations in Figure 4.3a. This category has the least number of tickets and therefore, also the widest confidence interval. The survival probability of 0.50 for this category is after more or less 10 weeks.

The second plot in Figure 4.3b shows the lifetime distribution of the large locations of Company

Y. This survival probability is shorter than the Company X large locations namely a survival probability of 0.50 after roughly 5 weeks.

Figure 4.3c shows the lifetime distribution of the fans of the regular Company X locations. The line is more curved than the line of the large locations. The survival probability is also shorter namely 0.50 after about 3 weeks already.

The last plot, in Figure 4.3d, presents the lifetime distribution of the fans in the regular locations of Company Y. This is the plot with the most narrow confidence interval of the four. The 0.50 survival probability is after more or less 6 weeks for this category.

Chapter 5

Solution analysis

In this chapter, the solution design is taken to the next level and analyses are done. First of all, the results of Chapter 4 are discussed and analyzed. Then, the distribution that is found in Chapter 4 is used to calculate the optimal T to perform maintenance. Furthermore, a goodness-of-fit test is performed to see how well the data actually fits the Weibull distribution.

5.1 Interpretation of the solution design

This section focuses on the results of the survival probabilities found in Chapter 4. The different categories are compared to each other and explanations are found for the differences.

5.1.1 Interpretation of All tickets

First of all, it must be noted that it is very hard to really compare the two different companies, but also the different locations. The number of tickets of the different locations differ so much that it is hard to draw real conclusions. As can be seen in Table 2.2, the large locations of Company X and Company Y are 3917 tickets compared to 558 tickets. For the regular tickets it is 9212 compared to 3480. The size of the data sample is significantly different which makes it hard to compare.

When looking at the large locations of both Company X and Company Y, the survival probability of Company Y is much larger even though Company X does a lot more preventive maintenance. An explanation for this could be that Company X still has about six times as much tickets as Company Y does in the same time period. There can be lots of explanations of why this is the case for example, Company X having a lower benchmark for an alarm resulting in more tickets or Company X having more components that fail regularly. Both confidence intervals seem to be about the same size. Moreover, it is clear that the difference between the Weibull distribution and the Kaplan-Meier estimator is much larger for Company X than for Company Y. This is possible because of the poor fit of the Weibull distribution.

When comparing the large and the regular locations of Company X, it is striking that Company X clearly has its focus on the large locations regarding preventive maintenance. Almost no preventive maintenance is done for the regular locations. Still, the survival probability for the regular locations is much larger. This could be explained given that a regular location is much smaller and thus, has very little components in comparison to the large locations. Fewer components means that on average it takes longer before something in that assets fails. Also, the confidence interval for the large locations is much larger than the confidence interval for the regular locations. This is plausible given the fact that the regular locations have about three times as many tickets as the large locations. The Weibull distribution and the Kaplan-Meier also look a lot more similar for the regular locations than for the large locations.

Comparing both regular locations is again quite complex because of the difference in the number of tickets. However, what we do see is that although, Company Y does a lot more preventive maintenance on the regular locations, the Company X regular locations look much alike regarding the survival probability. So the difference in maintenance strategies does not seem to have effect on the survival probabilities. Lots of factors can the the reason for this but one explanation could be that tickets are specific to one action or one component while in total there are lots of components that are not looked at during the maintenance of one ticket. The actual preventive maintenance therefore only has a small impact. Both regular locations look quite alike regarding the size of the confidence interval and the difference in Weibull distribution versus Kaplan-Meier estimator.

Looking at the percentage of preventive maintenance done by Company Y in Table 2.2, Company Y performs its preventive maintenance pretty equally over the large and the regular locations. The difference in survival probabilities could very well have the same explanation as for Company X. Regular locations are much smaller and have fewer components that can fail, therefore the survival probability is larger. Remarkable for these two graphs is that they both have bump, that the Kaplan-Meier is higher than the Weibull distribution. The fit of the Kaplan-Meier to the Weibull distribution is much better for the regular locations than for the large locations. Furthermore, the difference in the size of the confidence interval is similar to what happened for Company X.

5.1.2 Interpretation of Fans tickets

The ratio preventive maintenance done for the fans of the large locations of Company X and large locations of Company Y is quite similar to that ratio for all tickets. More preventive maintenance is done by Company X on the large locations and Company X also has a larger survival probability. The difference in preventive maintenance here might very well have the positive effect as expected.

When comparing the large locations of Company X by the regular locations of Company X, a large difference is seen. A lot of preventive maintenance is done on the large locations, and no preventive maintenance is done on the regular locations. The survival probability of the large locations is much better. After 10 weeks, the survival probability of the large locations is about 0.50 while for the regular locations, it is only 0.25. The preventive maintenance thus seems to pay off. The confidence interval for the large locations is by far the largest of all graphs, this makes sense as this category only contained 27 tickets. This also explains the fact that the Weibull distribution and the Kaplan-Meier estimator do not fit very neatly together. The fit of the Weibull distribution and the Kaplan-Meier estimator is much better for the regular locations, however, there are still a lot of deviations. The confidence interval is also still quite large.

When comparing the two distributions for the regular locations of Company X and Company Y, the difference in preventive maintenance again tends to make the difference. No preventive maintenance is done for Company X and after 10 weeks the survival probability there is only 0.25 while for Company Y, who do quite some preventive maintenance, after 10 weeks the survival probability is 0.40. For these the graphs, also a major difference is visible in the size of the confidence interval and the fit of the Weibull distribution to the Kaplan-Meier estimator. For Company Y, this all looks much smoother.

The ratio preventive and corrective maintenance is quite comparable for the large and regular locations of Company Y. The survival probability of the large locations is however, a lot shorter. It could be the case that the large locations contain more fans, and more components automatically mean that one fails faster. Another explanation might be that since the fan is an important component, and the large locations have more impact, the benchmark for a failure is lower on the large locations and a ticket comes in faster. The large difference in size of the confidence interval could again be well explained by the difference in number of tickets.

In order to better compare the Weibull distributions, Figure 5.1 shows all the Weibull distributions of the Fans combined. Interesting is the interception point of three lines at about 22 weeks and the differences in survival probability at the duration of 10 weeks, ranging from 0.10 to 0.50. This means that for the large Company X locations after 10 weeks, approximately half of the locations have not failed yet. Simultaneously, after 10 weeks only 10% of the large Company Y locations are still fully functioning. The difference between Company X and Company Y is thus, very large. This difference can (partially) be explained by the difference in preventive maintenance. Company X has 54.17% preventive maintenance on this location and Company Y does 28.06% preventive. However, it must also be noted that Company X only had 27 tickets in this category and Company Y had 139 tickets. The small number of tickets reduces the reliability of these findings. Other explanations mentioned previously can also be the reasons for these differences.

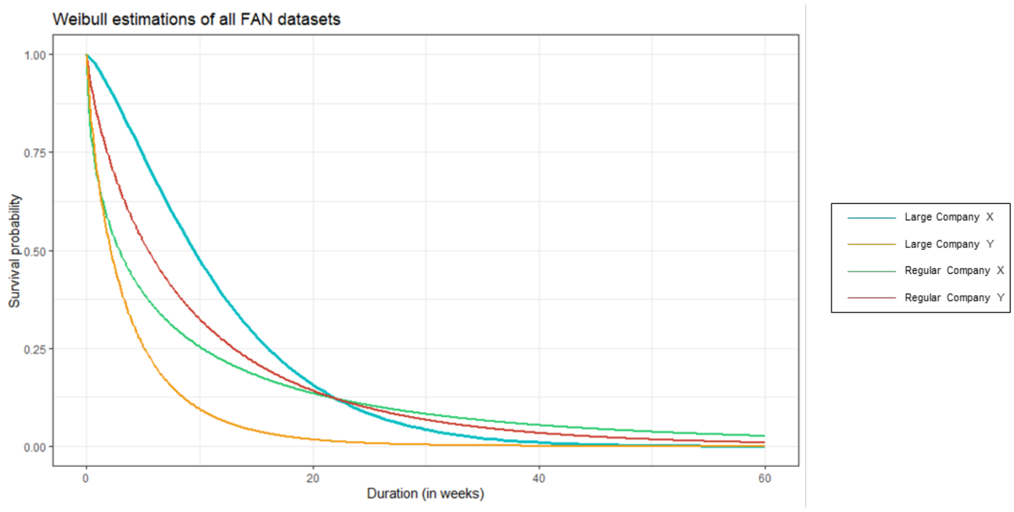


Figure 5.1: All Weibull distributions of the Fan categories

5.2 Optimal time T for performing maintenance

In Section 3.3, the concept of the optimal time T using the lifetime distributions and the costs for preventive and corrective maintenance was already introduced. The formula for calculating this optimal time T is repeated below.

$$cost\ rate\ (T) = \frac{C_{cm}F(T) + C_{pm}(1 - F(T))}{\int_0^T (1 - F(t))dt} \tag{5.1}$$

In this research, the lifetime distribution is assumed to follow a Weibull distribution. Therefore, the following input parameters need to be used. The formula for the $F(t)$ was already shown as Equation 3.7 in Section 3.2.2 and is repeated here for clarity. For the sake of completeness, the formula for $1 - F(t)$ is also shown.

$$F(t) = 1 - e^{-(\frac{t}{\lambda})^k} \tag{5.2}$$

$$1 - F(t) = e^{-(\frac{t}{\lambda})^k} \tag{5.3}$$

The two other parameters are the costs for preventive and corrective maintenance. After a short meeting with the manager of the Company X tickets, it was discovered that for corrective tickets, a fixed price is set out in a contract. This fixed price is a good average of what all the different tickets cost, the value of the downtime and the costs for sending someone to the site. Every corrective maintenance ticket costs €175,- and therefore, C_{cm}=175. Since this is a fixed price, the exact price build-up so the costs for sending a service employee and the downtime are unknown. According to [16], preventive maintenance represents about 10% to 30% of the total maintenance cost, and for the balance between corrective maintenance and preventive maintenance costs the 80/20 rule of thumb can be used. Therefore, we assume the preventive maintenance to be 20% of the total maintenance costs, and the corrective costs to be the other 80%. The value of preventive maintenance would therefore be equal to €43,75.

$$C_{pm} = \left(\frac{175}{80}\right) * 20 = 43.75 \tag{5.4}$$

In Section 3.3, it was stated that calculating the optimal time T is only useful when the failure rate of the asset is increasing. A failure rate is increasing when the value for k larger than 1. When an asset has a decreasing failure rate, corrective maintenance is more beneficial. In this research, for only one category this is the case namely for the Fans of the large Company X locations. In Table 4.1, all parameters are presented and the k for Company X Fan - Large is equal to 1.313695. Therefore, only for this asset category the optimal time T will be calculated.

Figure 5.2 shows the results of the cost rate function, the costs against the time, for the fans of the large Company X locations. The lowest costs are when the age is equal to 16 weeks. Thus, 16 weeks is the optimal interval to perform maintenance given this lifetime distribution and these costs. Appendix B contains the worksheet including the calculations done in Excel to compute this graph.

5.2.1 Sensors

In the Research motivation, in Section 1.2 of the introduction, sensor data is mentioned. A sensor can provide up-to-date information on the status of the asset. Therefore, make the timing for maintenance even more precise and minimizing the probability of performing preventive maintenance too early but also minimizing downtime by performing corrective maintenance too late. In Section 5.2, it was concluded that given the estimated costs for preventive and corrective maintenance, intervals of 16 weeks would be optimal for preventive maintenance. With Equation 5.5 the probability of performing maintenance too early and with Equation 5.6 the probability of performing maintenance too late can be calculated, and thus, this can be prevented by placing a sensor.

$$P(Too\,early) = P(t > 16) = \int_{T=16}^{\infty} f(t)dt \tag{5.5}$$

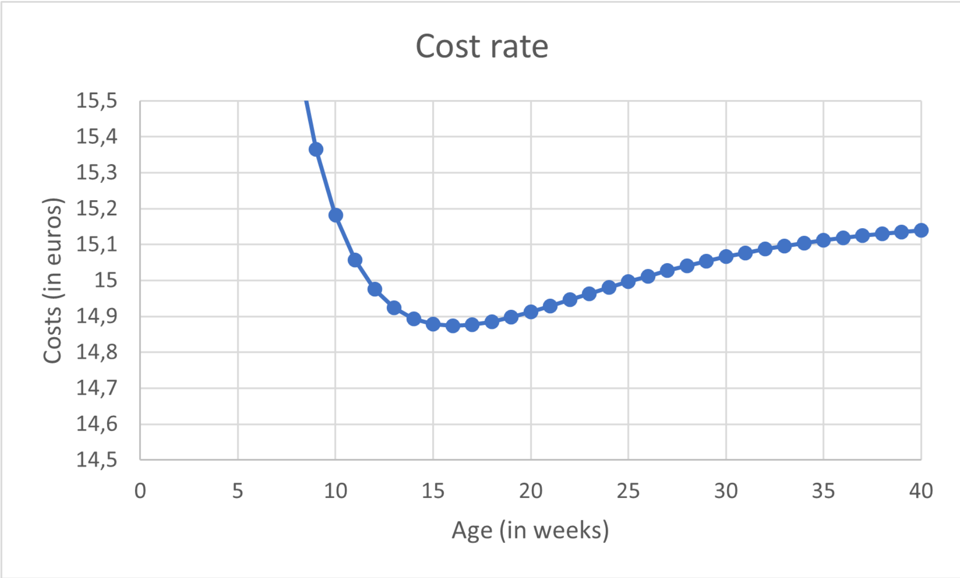


Figure 5.2: Optimal replacement age calculation for the Fans of Company X, Large locations

$$P(Too\,late) = P(t < 16) = 1 - P(t > 16) = \int_{0}^{T=16} f(t)dt \tag{5.6}$$

When filling in the parameters of the lifetime distribution of the fans of the large Company X locations in Equation 5.6, the calculation becomes as follows:

$$1 - P(t > 16) = \int_{16}^{\infty} \frac{1.313}{12.514} * \left(\frac{t}{12.514}\right)^{0.313} * e^{-(\frac{t}{12.514})^k} dt = 0.7503029..... \tag{5.7}$$

So then,

$$P(t > 16) = 1 - 0.7503029... = 0.249697... \tag{5.8}$$

Therefore, in about 75% of the cases, maintenance comes too late when doing maintenance after 16 weeks so there will be some downtime, however, given the costs this is still the optimal solution. In about 25% of the cases, maintenance comes too early which is unnecessary because there was still life potential in the component. Coming much too late or too early could be reduced by placing sensors. With sensors, the timing of when maintenance would be necessary can be optimised even more. Still, it cannot be perfect because other factors must be kept in mind like the time to travel to the site, the availability of service engineers and the urgency of the failure. In this example, in approximately 75% of the times, the service engineer comes too late and there is downtime. If the uptime of a fan is important to the company, placing a sensor might reduce the downtime from a week to only day, for example. Further analyses must then be done to see if that would be beneficial given the costs of buying and monitoring sensors in comparison to the maintenance costs now.

5.3 Goodness of fit test for the Weibull distribution

Based on the literature found in Chapter 3.3.2, it was assumed that the Weibull distribution would be an appropriate distribution to choose for fitting to the data. However, the goodness of fit can be tested statistically as well. The standard statistical tests for this cannot be applied because these do not support having censored data as well, however, there are some modifications of standard tests that do consider censored data. One example of such a modification that attempts to do this is the Modified Kolmogorov-Smirnov test found in [17]. They propose a modification of the classical Kolmogorov one-sample goodness-of-fit and Smirnov two-sample test procedures. These modified procedures are especially appealing as they allow for clear generalization to procedures, rather easy to apply, which are relevant in case the data is subject to random right censorship. With the help of [18], a code in RStudio is written that performs the calculation. The code can be found in Appendix C. The output of the calculation is a p-value. This value tells us whether the null hypothesis can be rejected or not. The null hypothesis and the alternative hypothesis being the following:

$$H_0\text{: The sample data follow the Weibull distribution}$$
$$H_1\text{: The sample data do not follow the Weibull distribution}$$

We choose $p = 0.05$ as our threshold value, this is the most common value to use [19]. When the p-value is below the threshold value, it is statistically significant and indicates strong evidence against the null hypothesis as there is less than 5% probability that the data set follows a Weibull distribution. A p-value larger than 0.05 indicates strong evidence for the null hypothesis. Note that the null hypothesis cannot be accepted, it can only be failed to reject in this case. A statistically significant result is not the same as a 100% certainty. The outcomes of the calculations are presented in Table 5.1.

Table 5.1: P-values of the goodness of fit test for the Weibull distribution per category

Company	Category	$p-value$	H_0 rejected?
Company X	All - Large	0.0000	Yes
	All - Regular	0.0000	Yes
	Fan - Large	0.9797	No
	Fan - Regular	0.4981	No
Company Y	All - Large	0.0000	Yes
	All - Regular	0.0000	Yes
	Fan - Large	0.2837	No
	Fan - Regular	0.6207	No

What immediately stands out when looking at the results is that all 'All' categories reject the null hypothesis. These are the categories with lots of tickets so we can assume that for these categories we are dealing with the common phenomenon that the power of a test increases when the number of observations grow. Given the fact that it is highly unlikely that the actual distribution is the Weibull distribution, the probability is large that with this many observations, or tickets, it is not possible to find a Weibull distribution that will not be rejected. Furthermore, a logical explanation for the fact that the Weibull distribution does not fit for these data sets is that in these categories we are combining a failures of all components and look at it as it behaves the same. Different components do not behave in the same way and therefore, we are most likely dealing with multiple distributions. Alternative more complex distributions could be considered for these data sets, or assuming that we are dealing with multiple distributions, the sum could be a multimodal distribution.

In order to check whether the Weibull distribution is a really poor fit, probability-probability plots, or short P-P plots, are created. The closer the dots are to the straight line, the closer the data sets fit. All the P-P plots are presented in Appendix D. Figure D.1 shows all the plots of rejected tests and Figure D.2 shows the plots of non-rejected tests. When, for example, comparing Figure D.1d which has a rejected null hypothesis but the plot seems to fit quite well and Figure D.2c which failed to reject the null hypothesis but the plot seems quite off. This is therefore, a reason the further research can be done in order to see if the Weibull distribution is a good fit to the data set or not.

Chapter 6

Conclusion

This is the concluding chapter of the research. Section 6.1 states the overall conclusion and also answers all sub research questions. Section 6.2 discusses the findings in the conclusion and it will be put into context of the overall research. Section 6.2.1 reviews all limitations that influence the interpretation of the research, also the assumptions made will be covered in this section. The next section presents the recommendations of the study. Then, the research relevance and the scientific and practical contribution are reviewed. Lastly, opportunities for further research are given.

6.1 Conclusion

In this section, first the overall conclusion will be given by returning to the original core problem. Then, all the research questions will also be answered.

Core problem: There is insufficient awareness of the different maintenance strategies and how they can be used in asset management among Allinq's customers.
We found that only when an asset has an increasing failure rate, it is beneficial to perform preventive maintenance. Given the failure rates and the lifetime distributions that we calculated now, this is only the case for the fans of the large Company X locations. The others all have a decreasing failure rate which, in theory, makes preventive maintenance not useful. Intuitively, this however seems a bit odd. Fans are likely to be components that suffer from wearing out and therefore, are expected to have an increasing failure rate. The decreasing failure rate that is calculated implies that the components get better by age however, we cannot practically justify this. It could be that the limitations and assumptions of this research cause the result to be different than expected. Also, given that the Weibull distribution does not fit well for half of the data sets, the conclusion regarding the decreasing failure rate also does not seem valid anymore.

By calculating the optimal time T for maintenance, the fans of the large Company X locations need maintenance every 16 weeks. For the other categories, just performing corrective maintenance is most beneficial. All in all, the uncertainty of when maintenance is necessary can be reduced by placing sensors. Further analysis is, however, necessary to prove whether this is profitable or not.

Research question 1: What assets from different customers can be best focused on for analyzing maintenance strategies based on the current situation?
We chose technical buildings here since it best fulfilled all the given criteria. Since there was still a lot of distinction to be made within the category, namely the large and the regular locations, that distinction is also made. Then, it is interesting to look at all tickets in total but because a

technical building has so many components and can have numerous different kinds of failures, we chose to also focus on one component namely the fans. Fans are important components because they prevent the buildings from overheating and there were enough and comparable tickets in all categories of both Company X and Company Y on fans.

Research question 2: What maintenance strategies are there for telecommunication assets?
Corrective maintenance, time-based preventive maintenance and condition-based preventive maintenance are covered in this research and these are also the most common maintenance strategies. All can be applied to telecommunication assets, which one is most beneficial depends on costs and failure rate. Company X and Company Y both use a combination of corrective and preventive maintenance. At this point, it is uncertain whether time-based or condition-based preventive maintenance is performed for both. In literature, for example [5], a combination of time-based and condition-based preventive maintenance is recommended, based on different components.

Research question 3: What statistical methods are available for failure modelling?
There seem to be endless statistical methods for failure modelling. The Weibull distribution is very common and is also used in this research. The Kaplan-Meier estimator is used as well so that a parametric and a non-parametric method both the life time distribution of the same data set. Other methods have not been elaborated on in this research. *Research question 4: How is the "time-to-failure" for an asset determined?*
Gaining insights in the time it takes until the next failure occurs can be provided by the many statistical methods for failure modelling. The parameters and the plot of the Weibull distribution provided a lot of information on the survival probability taking into account the difference between the preventive and corrective maintenance tickets. In Section 2.4, the average of the number of days until the next ticket arose was presented. These results however, do not consider the difference of the corrective and preventive maintenance tickets which makes those results less accurate.

Research question 5: What data within the data set is needed to find the failure distribution?
In this research the building number, the date that the ticket came in and the parent line description, this was the label whether a ticket was preventive or corrective, were used for the failure distribution itself. Before this step, the descriptions of all tickets were already analyzed and the tickets were labeled, see Table 2.1. Labeling provided more depth in the categories, but was not necessary for a calculation of a failure distribution in general.

Research question 6: What does the failure distribution of each asset look like and what does that mean?
In order to map the failure distributions, per category, the Weibull distribution parameters are estimated using the historical time to failure data and the Kaplan-Meier is estimated. These distributions are plotted together in a graph representing the failure distribution. Only the category Fans at large location of Company X had an increasing failure rate as outcome, the others all had a decreasing failure rate. It is hard to practically justify the decreasing failure rate. Later in the study, it was found that the Weibull distribution is a poor fit to half of the data sets which could be an explanation for the unexpected results.

Research question 7: What is the optimal time T to perform maintenance?
Since there was only one category with an increasing failure rate, only one optimal time T has been calculated. Given the life time distribution and the costs that are estimated in the research, the optimal time T to perform maintenance for the fans of the large Company X locations is 16 weeks.

6.2 Discussion

Maintenance strategies have a huge impact on the failure behaviour of assets. However, in this research that exact relationship between the maintenance practices done and the resulting failure behaviour is still unknown. In order to actually find that relationship, more differentiation needs to be done to get to the core source of a failure of a certain component, then a good maintenance strategy can be set up. Also then, the impact of certain failures can be taken into account. Some components have a more important task to fulfill than others after all. Nevertheless, he methods described in this study are a good starting point for Allinq to continue their research into this topic.

All the conclusions drawn are based on the estimated Weibull distributions. However, in the goodness-of-fit test, it was seen that the Weibull distribution is not a very good fit for half of the categories. The ‘All tickets’ categories all do not fit the Weibull distribution which makes sense. In the All tickets categories we are adding all different kinds of failures of different kinds of components. It is not realistic to assume that different kinds of failures behave in the same way, which is shown by the fact that simply one distribution does not fit. We are dealing with different distributions here that should be treated separately for more accurate results.

Furthermore, we only analyze ticket data. Meaning that assets that did not receive a ticket are not included in the data set. This leaves us with an incomplete data set. Furthermore, only locations that had more than one ticket in the time frame, are included in the calculations. This excluded quite some locations from the data set and therefore, created a biased data set. The locations that performed well, meaning having zero or only one ticket, are excluded which gives the results a divergent image since only the ‘poorly’ performing locations are included. Also, not all tickets could be well interpreted and they might have ended up in the wrong category. We have seen that the results still show interesting insights, however, this bias needs to be kept in mind.

6.2.1 Limitations

In this subsection, some limitations due to the method used and the provided data will be discussed. First of all, this research has been performed as the graduation assignment for the bachelor program Industrial Engineering and Management meaning that there was only limited time available. Assumptions were necessary in order to still finish the research, and opportunities for further research are provided. If more time would be available, we would have liked to dive into the data even more to better define the categories which would result in more accurate results. Also, we would have tried to make the section on costs more accurate.

Furthermore, the data used is all subject to human error. Most labels and descriptions have been filled in my hand which causes some bias and noise in the data. People can fill in information wrong or leave blank spaces. Also, we are uncertain of whether the data is complete, not all tickets might have been documented well.

Incomplete data example: we only have ticket data so all the assets that did not receive a ticket in the given time frame are not included in the research. Also the assets that only had one ticket are excluded because the timing of the last or next maintenance action is unknown so nothing can be said about the time till next. Furthermore, in this research we made the assumption that every ticket is a failure, however, looking at it from the other way, we are unsure if there are failures that occur but do not create a ticket. If that is the case, we also have an incomplete image of failures occurring. More investigation in the failure behaviour of the asset needs to be done to rule that out, or not.

Given time restrictions and the difficulty of interpreting some tickets, the tickets are grouped in only a few categories and we did not dive into the sources of the failures, also called the failure mode. This limits the accuracy of the results. Moreover, in the All categories, we surely aggregated different sources of failures, which we say might be the reason for the poor fits of the distributions. Working with aggregate failures is a limitation.

6.2.2 Assumptions

We assume that every ticket is a failure, even though in real life that is not always the case. Sometimes the tickets are simply a warning or the ticket is a failure of such a small component which does not necessarily cause the entire technical building to fail. For this research we assume that every ticket is a failure, or it can be interpreted as 'something is wrong with the asset'. There were a few tickets that we obvious to not be failures of any kind and they have been excluded from the research to make it as accurate as possible. In Table 2.1 these are collected under the label 'non-failure'. By making this assumptions, we have probably overestimated the failure rate. Most tickets were not a real failure yet, or were a failure of only a component that did not have too much impact on the entire building. This overestimation should be kept in mind when interpreting the results.

The costs for calculating the optimal time T are based on talks to employees and an estimation based on literature. This price estimation is made as accurate as possible but with the little information at hand, these assumptions had to be made. The calculation is still presented to show the method of how this can be calculated, also when the costs are different or when they change.

There was no information on whether the maintenance performed did what was expected, or was done correctly. In order to still draw conclusions from tickets and the corresponding maintenance action we needed to make an assumption. We assume that after maintenance is done, preventive or corrective, the condition of the asset is 100% again. This assumption also overestimates the failure rate a bit.

Regarding the tickets, we assume either preventive maintenance or corrective maintenance is performed. The column in the data set that included information on this contained more options than just these two. We were not sure of the definitions of all these other options and therefore, made the assumption that when 'Preventive maintenance' was not mentioned literally, it was corrective maintenance. Clearing up the definitions could change the ratio of corrective and preventive maintenance.

6.3 Recommendations

Based on this research, given all the limitations and assumptions, we would like to make the following recommendations:

First of all, given our model the recommendation would be to only perform corrective maintenance for all categories except for the fans on the large Company X locations. For the fans on the large Company X locations, we would recommend to do maintenance every 16 weeks. At these large locations there are already multiple tickets daily, therefore, in practice this would mean that the service man that comes regularly for these other tickets would get the specific task of replacing the fan every 16 weeks.

Furthermore, we would recommend to do further research in order to make the results more accurate. The goodness-of-fit test showed that for half of the categories, the Weibull distribution was not a good fit. Another distribution might fit better and give more valid results. Also, given that the poor fit could be explained by that we combine different failures and treat them as if they were the same, it would be recommended to make the distinction between the failures and then run the calculations again. This will most likely give more accurate results. Moreover, the estimation of the costs for the corrective and preventive maintenance is simplified a lot. Therefore, finding more precise values for these costs taking into account the value of down time, man hours and repair costs would also improve the results.

Improve data collection of data in such a way that these insights can be gained for all components and assets, keep track of more variables so interpretation is easier. Also to keep monitoring this to have a larger time frame. We only looked at data of 16 months in this study.

6.4 Relevance

Referring back to the research motivation in Section 1.2, this research surely is relevant. Not only within Allinq, but within the entire telecommunication sector, more insights in the lifetime distributions and maintenance strategies are requested. Furthermore, the model created can also be used in other sectors and situations. Creating insight in the lifetime distribution of any asset that needs maintenance, especially when the assets have high costs and high downtimes, would be beneficial because that is the basis of setting up a well thought-out maintenance strategy. This would therefore, be very helpful for companies that do not have a maintenance strategy yet. Moreover, when an asset has an increasing failure rate, considering the optimal replacement age could be very profitable. Performing maintenance according to the optimal replacement age will significantly reduce costs and downtime. This method can easily be applied to other assets. As a recommendation, companies should collect data on the failure behaviour of their components or assets. This data can be as detailed as the company desires but logically, the more detailed, the more detailed the results will be. With the data collected, the lifetime distribution of the asset can be mapped out and further analyses can be done.

6.5 Scientific and practical contribution

The scientific contribution of this research is first of all, applying theory and adjusting methods to data of a real life situation and therefore, performing a case study. Furthermore, there is limited literature available of maintenance strategies in the context of telecommunication. Unique for this research is analyzing telecommunication failure data by combining the Weibull distribution as well as the Kaplan-Meier estimator. Moreover, the way the theory and results are presented can provide insights and contribution to future works.

The practical contribution of this research to Allinq is first of all, inspiring them with this method of predicting failures. This is a method that they have not used before and they are pleased with the results. Furthermore, the model can bring them many benefits. It shows the insights that they requested and it has many possibilities for further analyses. Lastly, the recommendations mentioned earlier can all be of great contribution.

6.6 Further research

Finding out whether the results can be more accurate, like mentioned in the recommendations, are all opportunities for further research. Other possibilities for further research are presented in this section as well.

Also, like already expected, we can draw more realistic conclusions from the fans than from all tickets in total, therefore it would be useful to look into all ticket labels separately to see what the lifetime distribution is. Following the same method as used in this study, it is simply a matter of collecting the data of different components or categories and running the calculations.

Furthermore, more analysis can be done into the fans. Since they control the temperature of a technical building and prevent it from overheating, it could be interesting to see if the seasons have influence on its performance. Maybe they break down faster in the summer time because it is warmer outside and the fans need to work harder, then maintenance strategy can be adjusted to the weather outside as an example.

It has been mentioned a few time in this research already but it might be profitable to place sensors on certain components. Sensors are expensive but for certain components, especially the ones with high impact, it might be profitable to know exactly what the condition is at all times.

Some suggestions for further analyses have been mentioned throughout the research already but there are lots of extra analyses that can be done with the basis of a life time distribution. Just a few possibilities are testing whether the geographic location of an asset influences the failure behaviour, formally testing if Company X and Company Y actually face the same failure behaviour or not and testing if all locations behave similar and if not, why not.

Bibliography

[1] Hans Heerkens and Arnold van Winden. *Solving Managerial Problems Systematically.* Noordhoff Uitgevers, Groningen/Houten, 1st edition, 2017.

[2] Fangfang Ding and Zhigang Tian. Opportunistic maintenance for wind farms considering multi-level imperfect maintenance thresholds. *Renewable Energy*, 45:175–182, 2012.

[3] Geert Waeyenbergh and Liliane Pintelon. A framework for maintenance concept development. *International Journal of Production Economics*, 77(3):299–313, 2002.

[4] Engin Topan. Reliability and maintenance in the minor aerospace management. University Lecture, 2015.

[5] Ali Karevan, Kong Fah Tee, and Mohammadreza Vasili. A reliability-based and sustainability-informed maintenance optimization considering risk attitudes for telecommunications equipment. *International Journal of Quality & Reliability Management*, ahead-of-print, 08 2020.

[6] Olasunkanmi Olapeju. *Planning and the Telecommunication Industry in Nigeria*, pages 29–47. 01 2017.

[7] Shaoping Wang, Mileta Tomovic, and Hong Liu. Chapter 3 - comprehensive reliability design of aircraft hydraulic system. In *Commercial Aircraft Hydraulic Systems*, Aerospace Engineering, pages 115–169. Academic Press, Oxford, 2016.

[8] Gerhard Greeff and Ranjan Ghoshal. Chapter 8 - production capability management. In *Practical E-Manufacturing and Supply Chain Management*, pages 214–242. Newnes, Oxford, 2004.

[9] Rosmaini Ahmad and Shahrul Kamaruddin. An overview of time-based and condition-based maintenance in industrial application. *Computers and Industrial Engineering*, 63(1):135–149, 2012.

[10] Bram de Jonge, Warse Klingenberg, Ruud Teunter, and Tiedo Tinga. Optimum maintenance strategy under uncertainty in the lifetime distribution. *Reliability engineering and system safety*, 133:59–67, 2015.

[11] Kitty J. Jager, Paul C. van Dijk, Carmine Zoccali, and Friedo W. Dekker. The analysis of survival data: the kaplan–meier method. *Kidney International*, 74(5):560–565, 2008.

[12] E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[13] Jonathan S. Tan and Mark A. Kramer. A general framework for preventive maintenance optimization in chemical process operations. *Computers and Chemical Engineering*, 21(12):1451–1469, 1997.

[14] Richard E. Barlow and Larry C. Hunter. Optimum preventive maintenance policies. *Operations Research*, 8:90–100, 1960.

[15] Bram de Jonge. Weibull distribution; Weibull parameter estimation. https://www.youtube.com/watch?v=TCjZ_E1jFbU&t=791s, February 2021.

[16] Christer Stenström, Per Norrbin, Aditya Parida, and Uday Kumar. Preventive and corrective maintenance – cost comparison and cost–benefit analysis. *Structure and Infrastructure Engineering*, 12:603 – 617, 2016.

[17] Thomas R. Fleming, Judith R. O'Fallon, Peter C. O'Brien, and David P. Harrington. Modified kolmogorov-smirnov test procedures with application to arbitrarily right-censored data. *Biometrics*, 36(4):607–625, 1980.

[18] Klaus Langohr, Mireia Besalú, and Guadalupe Gómez. Goodness-of-Fit Methods for Right-Censored Data: Package 'GofCens'. *CRAN*, 2022.

[19] Saul A. McLeod. What a p-value tells you about statistical significance. *Simply Psychology*, 2019.

Appendix A

Code in RStudio for estimating the Weibull distribution and Kaplan-Meier estimator

```
####################################################################
# In this script we calculate the Weibull and Kaplan Meier         #
# estimator of the survival function of a given set of life times  #
####################################################################

### Loading in the data ###
# Package required
library(readxl)

# Load the data
data <- read_excel("Weibull FANS.xlsx", sheet = 4)

### Estimating the Weibull parameters ###

# First implement a function that calculates likelihood of the data while
# assuming a Weibull distribution
# In this function a is the vector with shape and scale parameters created using
# c(shape, scale). This function is meant for non-censored data
loglik <- function(a) {
  shape <- a[1]
  scale <- a[2]
  lik <- dweibull(data$`Duration (in weeks)`, shape, scale) #density
  truelik <- c() #initialise vector to which we will add the likelihood terms
  for (i in 1:nrow(data)) {
    truelik[i] <- lik[i]
  }
  return(-sum(log((truelik))))
}

# This function is meant for censored data
loglikCensored <- function(a) {
  shape <- a[1]
  scale <- a[2]
  lik <- dweibull(data$`Duration (in weeks)`, shape, scale) #density
  truelik <- c() #initialise vector to which we will add the likelihood terms
  r <- pweibull(data$`Duration (in weeks)`, shape, scale, lower.tail = F)
```

```
  for (i in 1:nrow(data)) {
    if (data$Censored[i] == "No") {
      truelik[i] <- lik[i]
    } else {
      truelik[i] <- r[i]
    }
  }
  return(-sum(log((truelik))))
}

# Examples how to run these functions
loglik(c(3, 4)) #for non-censored data
loglikCensored(c(3, 4)) # for censored data

### We now turn to optimising the log-likelihood
# Load package required for numerical optimisation
library(stats)

# Use the function optim to maximise likelihood and find optimal Weibull
# parameters. Play around with the starting values (here 3 and 5)
# For non-censored data:
optim(c(1, 1), loglik)
optimpar <- optim(c(1, 1), loglik)$par

# For censored data:
optim(c(1, 1), loglikCensored)
optimpar <- optim(c(1, 1), loglikCensored)$par


# Generate sequence of numbers from 0 to 30 (adapt this based on the maximum
# lifetime in the data)
x <- seq(0, 60, by = 0.1)
df <- data.frame(x)

# Packages required for plotting
library(ggplot2)
library(ggfortify)


# Create Weibull plot
weibullp <- ggplot(df, aes(x = x)) +
  stat_function(fun =
                  function(x) pweibull(x,
                                       shape = optimpar[1],
                                       scale = optimpar[2],
                                       lower.tail = FALSE),
                color = "#00BFC4", size = 1.3) +
  ggtitle("Fitted Weibull survival function") +
  xlab("Duration (in weeks)") +
  ylab("Survival probability") +
  theme_bw()
```

```r
weibullp


### Kaplan-Meier estimator ###

library(survival)
library(survminer)
# For censored data
Y <- Surv(data$`Duration (in weeks)`, event = data$Censored == "No")
# For non-censored data
Y <- Surv(data$`Duration (in weeks)`)

# Plot Kaplan Meier estimator
ggsurvplot(survfit(Y ~ 1), data = data, ggtheme = theme_bw())$plot +
  ggtitle("Kaplan-Meier estimator")


# Plot Kaplan Meier and Weibull together
# First create dataframe with Weibull values
df <- data.frame(x = seq(0, 60, by = 0.1),
                 y = pweibull(x, optimpar[1], optimpar[2], lower.tail = F))

ggsurvplot(survfit(Y ~ 1), data = data, xlab =( "Time (in weeks)"),
ggtheme = theme_bw())$plot +
  geom_line(data = df, aes(x = x, y = y), size = 1, color = '#00BFC4') +
  theme(legend.position = 'None') +
  ggtitle("Company Y ALL Tickets of the Larger Locations")
```

Appendix B

Optimal T calculation in Excel

	A	B	C	D	E	F	G
1	**k**	1,313695		**C cm**	175		
2	**lambda**	12,51457		**C pm**	43,75		
3							
4			**T**	**F(t)**	**1-F(t)**	**Int(1-F(t))**	**Cost rate**
5			0	0,00000	1,00000	0	#DIV/0!
6			1	0,03552	0,96448	0,982239	49,28769265
7			2	0,08598	0,91402	1,921486	28,64206595
8			3	0,14200	0,85800	2,807494	22,22179615
9			4	0,20028	0,79972	3,636356	19,25996943
10			5	0,25890	0,74110	4,40677	17,63880433
11			6	0,31662	0,68338	5,119014	16,664548
12			7	0,37259	0,62741	5,774409	16,04536453
13			8	0,42624	0,57376	6,374996	15,6382303
14			9	0,47717	0,52283	6,92329	15,36538485
15			10	0,52516	0,47484	7,422124	15,18124751
16			11	0,57006	0,42994	7,874513	15,05753678
17			12	0,61184	0,38816	8,283562	14,97591085
18			13	0,65050	0,34950	8,65239	14,92404949
19			14	0,68612	0,31388	8,984076	14,89344113
20			15	0,71880	0,28120	9,281614	14,87807095
21			**16**	**0,74866**	**0,25134**	**9,547884**	**14,87361098**
22			17	0,77585	0,22415	9,785631	14,87690247
23			18	0,80052	0,19948	9,997448	14,88561513
24			19	0,82284	0,17716	10,18577	14,89801678
25			20	0,84297	0,15703	10,35286	14,91281414

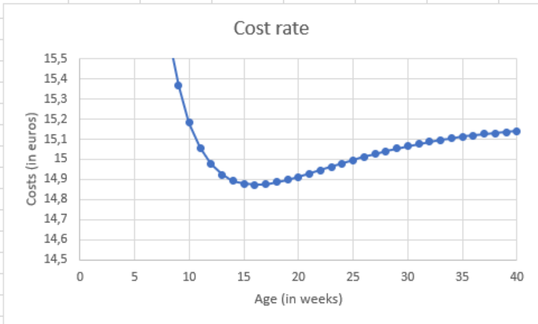


Figure B.1: Optimal T calculation for the Fans of the Large Company X locations in Excel

Appendix C

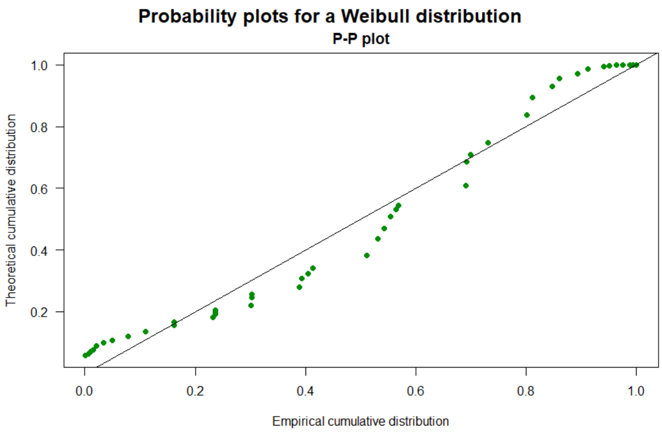
Code in RStudio for the Goodness of fit test

```
###########################
# Weibull Goodness-of-Fit #
###########################

library(readxl)
data <- read_excel("Alle durations, input voor R (1).xlsx", sheet = 8)

# First implement a function that calculates likelihood of the data while
# assuming a Weibull distribution
loglik <- function(a) {
  shape <- a[1]
  scale <- a[2]
  lik <- dweibull(data$`Duration (in weeks)`, shape, scale)
  r <- pweibull(data$`Duration (in weeks)`, shape, scale, lower.tail = F)
  truelik <- c()
  for (i in 1:nrow(data)) {
    if (data$Censored[i] == "No") {
      truelik[i] <- lik[i]
    } else {
      truelik[i] <- r[i]
    }
  }
  return(-sum(log((truelik))))
}

# Load package required for numerical optimization
library(stats)

# Use the function optim to maximize likelihood and find optimal Weibull
# parameters
optim(c(1, 1), loglik)
optimpar <- optim(c(1, 1), loglik)$par

### Implement GOF
# install.packages("Rtools")

install.packages("actuar")

install.packages("GofCens")
library(GofCens)
```

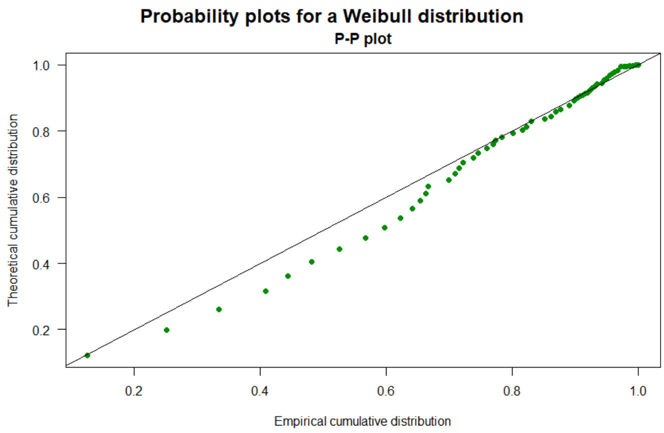
```r
# KScens function uses 1 for exact observation and 0 for censored
# Below, we transform our censored column to the right format
data$Censored <- as.numeric(data$Censored == "No")


KScens(times = data$`Duration (in weeks)`,
        cens = data$Censored,
        distr = "weibull")

# plot P-P plot
probPlot(times =data$`Duration (in weeks)`,
         cens = data$Censored,
         distr = c("weibull"),
         plots = c("PP") ,
)
```

Appendix D

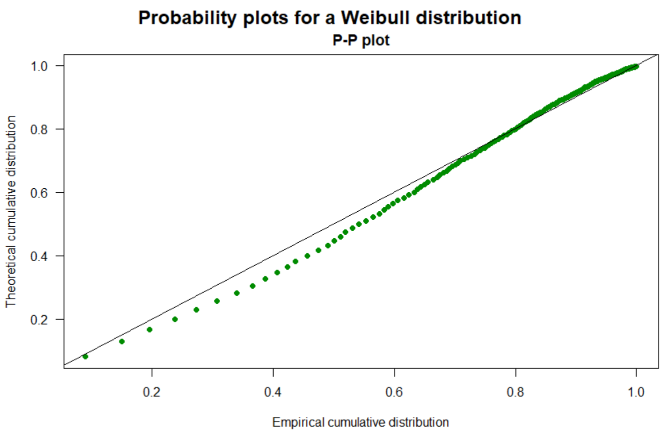
P-P plots for all categories



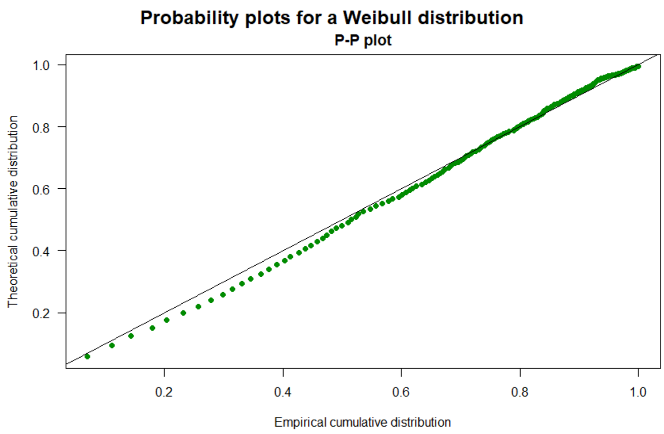
(a) Company X All tickets of the Larger locations



(b) Company Y All tickets of the Larger locations

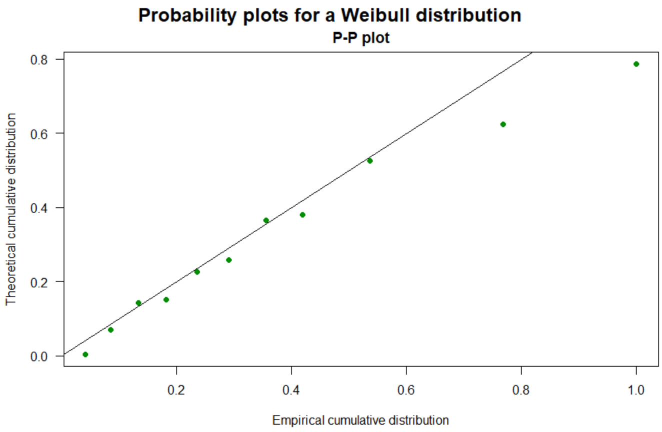


(c) Company X All tickets of the Regular locations

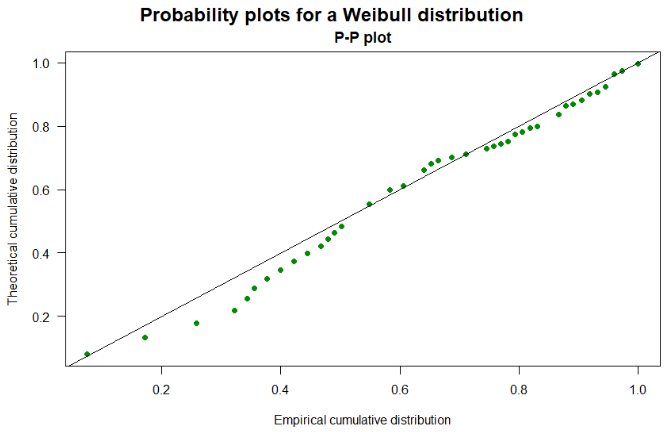


(d) Company Y All tickets of the Regular locations

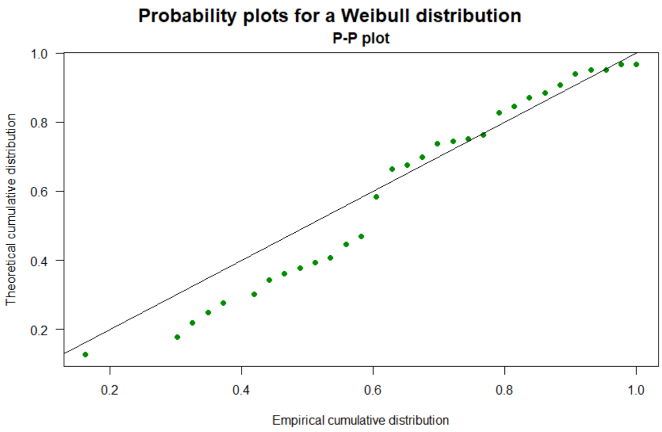
Figure D.1: All 'All tickets' P-P plots



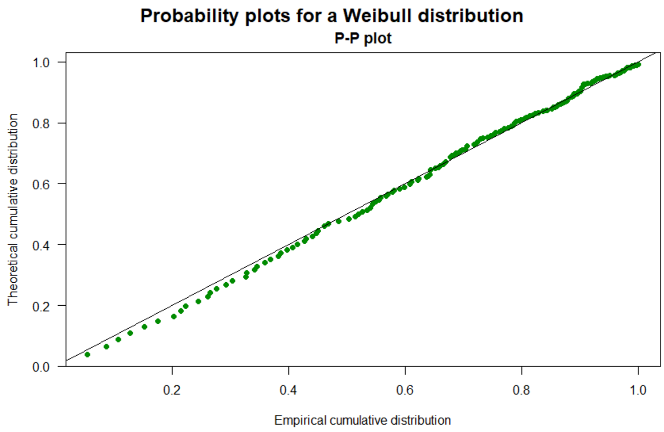
(a) Company X Fan tickets of the Larger locations



(b) Company Y Fan tickets of the Larger locations



(c) Company X Fan tickets of the Regular locations



(d) Company Y Fan tickets of the Regular locations

Figure D.2: All 'Fan tickets' P-P plots