Degree Project in Building and Real Estate Economics
Second cycle, 30 credits

Defining Underlying Factors Affecting Fault Reports within Residential Real Estate

LI DJURESTÅL

DAVID LEANDER
Abstract

Several studies have been reviewed to get an understanding of where the real estate industry, and more specifically within facility management, stand in regards to digitalization. Implementing digitalization of the Real Estate industry has been researched as a possibility for some time now. Furthermore, the industry needs to make better use of the data in hand created by these digitalized solutions. This thesis uses a quantitative approach through a data analysis, studying underlying factors that affect fault reports, with data on this matter from three Swedish real estate housing companies. Studying this in regards to fault reports in general, non-digitalized fault reports, and digitalized fault reports. The result of the data analysis implies that there are several variables that are statistically proven to affect the amount of fault reports made. This result is then discussed arguing for reasons of this outcome, as well as the literature study related to this subject.
Examensarbete

<table>
<thead>
<tr>
<th>Titel</th>
<th>Kvantitativ studie som definierar underliggande faktorer till felanmälningar inom bostadsfastigheter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Författare</td>
<td>Li Djurestål och David Leander</td>
</tr>
<tr>
<td>Institution</td>
<td>Fastigheter och Byggande, Fastighetsekonomi och Finans</td>
</tr>
<tr>
<td>TRITA nummer</td>
<td>TRITA-ABE-MBT-23430</td>
</tr>
<tr>
<td>Handledare (KTH)</td>
<td>Bertram Steininger</td>
</tr>
<tr>
<td>Handledare (Företag)</td>
<td>Mattias Liljemid</td>
</tr>
<tr>
<td>Nyckelord:</td>
<td>Digital Fastighetsförvaltning, Felanmälan, Data utforskning, Proptech</td>
</tr>
</tbody>
</table>

Sammanfattning

Acknowledgement

We would like to express appreciation to our supervisor Mattias Liljemid at Momentum Software AB assisting us with validated data, genuine interest, willingness to teach, and meaningful insights in progressing this thesis forward.

Additionally, we would like to thank our academic supervisor, Bertram Steininger, who assisted us through this paper with guidance, meaningful insights, and helping us in the progress of finalizing this thesis.
Table of Contents

Abstract ........................................................................................................................... II
Sammanfattning ................................................................................................................. III
Acknowledgement .............................................................................................................. IV
1. Introduction .................................................................................................................. 1
   1.1 Background .............................................................................................................. 1
   1.2 Problem Formulation .............................................................................................. 2
   1.3 Purpose and Questions ........................................................................................... 2
   1.4 Limitations .............................................................................................................. 3
   1.5 Structure of Thesis ................................................................................................. 3
   1.6 Motivation of Thesis .............................................................................................. 4
2. Literature Review ......................................................................................................... 5
   2.1 Digitalizing Real Estate .......................................................................................... 5
      2.1.1 Business Intelligence ....................................................................................... 6
   2.2 Digitalizing Facility Management .......................................................................... 7
   2.3 Fault Reporting within Facility Management ....................................................... 8
   2.4 Creating Value with Data Analytics ....................................................................... 9
   2.5 Visualizing Data ................................................................................................... 10
3. Theoretical Framework .............................................................................................. 12
   3.1 Scientific Management Theory ............................................................................ 12
   3.2 Explanatory Approach .......................................................................................... 12
4. Methodology .............................................................................................................. 13
   4.1 Research Outline .................................................................................................. 13
   4.2 Data Collection ..................................................................................................... 13
   4.3 Digitalized and Non-digitalized Fault Reports ...................................................... 14
      4.3.1 Digitalized Fault Reports ............................................................................... 14
      4.3.2 Non-digitalized Fault Reports ......................................................................... 14
   4.4 Software ................................................................................................................ 14
      4.4.1 Microsoft Excel ............................................................................................... 14
      4.4.2 Python ............................................................................................................. 15
   4.5 Quantitative Approach .......................................................................................... 15
4.5.1 Inductive Approach ........................................................................................................ 15
4.6 Regression Analysis ........................................................................................................... 16
  4.6.1 Linear Regression ......................................................................................................... 17
  4.6.2 Negative Binomial Regression Model ........................................................................... 17
  4.6.3 Poisson Regression Model .......................................................................................... 18
5. Data ........................................................................................................................................ 19
  5.1 Descriptive Statistics ........................................................................................................ 20
6. Results .................................................................................................................................... 26
  6.1 Exploratory Regression Analysis ...................................................................................... 26
  6.2 Ordinary Least Squares (OLS) Regression and Poisson Regression .............................. 27
  6.3 Analysis of the Negative Binomial Coefficients ............................................................. 28
    6.3.1 Underlying Factors for Fault Reports ......................................................................... 28
    6.3.2 Underlying Factors for Digitalized Fault Reports ...................................................... 30
    6.3.3 Underlying Factors for Non-digitalized Fault Reports ............................................. 32
  6.4 Validity ................................................................................................................................ 33
  6.5 Reliability ............................................................................................................................ 34
  6.6 Relevance ............................................................................................................................ 34
7. Discussion ................................................................................................................................. 35
  7.1 Underlying Factors for Fault Reporting .......................................................................... 35
  7.2 Non-digitalized Fault Reports ......................................................................................... 38
  7.3 Digitalized Fault Reports ................................................................................................. 38
  7.4 Acceptance of Digitalization .......................................................................................... 39
  7.5 Data Analytics & Facility Management .......................................................................... 40
8. Summary and Conclusion ....................................................................................................... 42
  8.1 Further Research .............................................................................................................. 43
9. References .............................................................................................................................. 44
10. Appendix ................................................................................................................................. 48
1. Introduction

1.1 Background

Many industries today have in the past decades taken a leap in becoming more digitalized and transformed for the sole purpose of being more effective. Softwares is developed all around us, in all kinds of different industries, and effectively helping professionals in all kinds of positive ways. In light of the current challenges faced for Real Estate owners such as handling rising energy costs and higher interest rates (Sveriges Riksbank, 2022), it is crucial for financial decision makers in all industries to take action. Implementing digitization of the Real Estate industry has been a researched and viable possibility for some time now (Kytömäki, 2020). The real estate industry is not really known for being very positive toward technological change (Baum et al., 2020). The industry has been one of the most resistant industries towards digital transformation within the Constructional & Real Estate industry (Trevik & Nilsson, 2017).

Therefore, real estate companies should move beyond traditional methods and find new alternative ways to generate revenue from their properties. The growing popularity of Proptech and digitalized solutions gives an opportunity to leverage technological advancements to improve the real estate market and drive innovation (Baum et al., 2020).

Coming years within facility management will be important for the growth and expansion of facility management services. According to Araszkiewicz (2017), the next few years will be critical for the facility management segment, as digital technology such as Big Data, IoT, Management systems, and advanced connectivity will impact the efficiency of buildings and affect both facility managers, and also their specific clients. Big data is an extensive amount of data that cannot be processed by conventional methods due to its complexity and size (Araszkiewicz, 2017). It can supply predictive analysis and assistance in creating data-driven decisions (Atkin & Bildsten, 2017). Araszkiewicz (2017) also suggests that the FM market will be worth almost 1,000 billion USD by 2025, and by creating value in the industry also highlighting the potential for digital transformation. Operational data extracted from big data can also be used to arrange preventive maintenance, resulting in more effective and also speedier decision-making (Atkin & Bildsten, 2017).
Trying to find more information specifically regarding facility management and what variables affect the fault reports being made confirms a research gap. However, there seems to be a need for new research to understand how the result of implementing digitization can bring value in specific areas. When taking the leap in digitalizing properties, the effects from this are that loads of data points can be explored in finding how and why certain things have occurred for a property. Therefore, in this Master Thesis, the authors are trying to contribute to what the implementation of digitalized software can contribute to, with the focus towards facility management, and the area of fault reporting for the real estate housing industry.

1.2 Problem Formulation

There has been significant research on the potential benefits of digital software within the real estate industry, particularly also in the area of facility management. These benefits can improve knowledge regarding energy efficiency, better vacancy management, optimization of rental levels, and more efficient fault reporting. By leveraging digital technology, real estate companies can improve their operations, reduce costs, and provide better services to tenants (Momentum, 2023). Prior research has shown that digitalizing certain areas of facility management can improve efficiency and reduce costs, although, there are no studies examining underlying factors that affect the amount of fault reports made and looking for relationships and patterns and explain which factors are related to the cause of them. Based on current knowledge of the authors there is a gap in previous research in understanding why these fault reports are being made, therefore this thesis is being made.

1.3 Purpose and Questions

The purpose of this thesis is to research underlying factors that affect fault reports, also referred to as service orders, within three Swedish residential real estate companies portfolio of properties. The ambition is to understand fault reporting within facility management in a better way by conducting this research. Another purpose is to look if there is any difference in what affects fault reports between the ones made digitalized, and the ones made non-digitalized, through the dataset used containing fault reports from three Swedish residential real estate companies portfolio of properties.
The purpose then leads to the following questions:

(1) Which underlying factors affect fault reports?

(2a) Which underlying factors affect the digitalized fault reports?

(2b) Which underlying factors affect the non-digitalized fault reports?

1.4 Limitations

For limitations the thesis will only use a quantitative approach by using regression analysis for the result. This will be made on a merged dataset containing three different real estate housing companies data regarding fault reports being made on their portfolio of properties, made between May 2021 - February 2023. The fault reports are only based on properties for rental housing. The fault report specifically is based on faults being reported on both apartments for the tenants, and public spaces in the properties (for example hallway, laundry room, store-area). These properties are all being placed in the country of Sweden. Due to the data used is not perfect since human factors are affecting the information being reported, some data points had to be filtered out and are unused in this analysis that does not hold all the values from the variables needed to make the regression model the authors want to achieve for the result. Another limitation is that this study evaluates whether the variables positively or negatively impact the number of service orders; however, it does not go into how much the variables positively or negatively impact the number of service orders.

1.5 Structure of Thesis

This paper is structured with Chapter 2 presenting the previous academic works, which evolves around digitalizing properties, fault reporting for the tenant, facility management challenges, and data analytics within facility management. Chapter 3 resolves the Theoretical Framework that will be in focus for this thesis. Chapter 4 is dedicated to explaining the quantitative method that is used to be able to answer the research questions mentioned earlier. Furthermore, explaining the models and variables used to enable this. Chapter 5 will present the data used by visualizing it. Thereafter, Chapter 6 presents the result achieved, it will show the results from the regression analysis and what has been discovered. Chapter 7 will discuss the findings and the interesting relationship discovered when comparing it to prior studies. The result will
also be analyzed in this part to give thoughts and discussion for the findings that have been conducted. Lastly, Chapter 8 is dedicated to the conclusion where the most important parts will be summarized together with what further studies can research.

1.6 Motivation of Thesis

This thesis comes from the ambition of generating influential knowledge by understanding data that is accumulated from a Business Intelligence platform, called Momentum BI, provided by Momentum Software AB. This platform is made for Real Estate housing companies to get valuable and informative insights and provide their own operation with the right accurate information to steer and measure their business in the right directions (Momentum, 2023). It is made specifically for facility management and to ease many of the economical and technical challenges in taking care of both the property and the operation itself. Momentum BI has multiple areas that get valuable information from their building, with a few being: Vacancy Rates and rental losses, contracts, maintenance planning, Customer Relationship Management, Service orders, and Fault reports (Momentum, 2023). The area of fault reporting is something that is an important and valuable service for the tenant living in an apartment building (Blomé, 2006). This thesis will specifically look into the fault reports that have been made from multiple properties and have received data regarding this coming from three separate Residential Real Estate companies, with the help of Momentum Software AB.
2. Literature Review

2.1 Digitalizing Real Estate

The real estate industry has encountered significant changes and growth due to the emergence of PropTech, which refers to utilizing technology to transform and innovate the industry (Retief et al., 2016). PropTech has evolved into a strong capital market, drawing billions of dollars in investment (Baum, 2017). The real estate industry is presently undergoing change as investors and technology creators have begun shifting their focus toward altering business and product model inventions (Siniak et al., 2020).

Kytömäki’s Licentiate thesis (2020) examines the impact of digitalization on the facility management sectors, and also the facility management sectors in general. This thesis argues that digitalization has the potential to significantly improve resource efficiency and services for tenants and building users. Onwards explaining that by investing in integrated digitalization operations, real estate companies can differentiate themselves in a competitive market, offering innovative solutions that can attract and retain its customers. The thesis talks about the challenges and opportunities presented by digitalization and provides a holistic view on it by shedding light on how the real estate industry can take advantage of this phenomenon to achieve sustainable growth. The result in Kytömäki’s thesis indicates that investing in an integrated digitalization operation can be precisely what the real estate companies need to distinguish in an active competitive market.

Kytömäki (2020) underlines the difficulties of executing changes in the conservative real estate industry, including uncertainties arising from new technologies, and also competition that create obstacles in decision-making. According to Trevik & Nilsson (2017), the Constructional & Real Estate industry is among the most reluctant to adopt digital transformation within its segment. While digitalization can bring significant benefits to the industry, there are challenges in implementing these technologies, particularly in the real estate industry. Another report explains that the real estate firms struggle to grasp and navigate the extensive amount of data they have accumulated and to fully apprehend and understand its value (Atkin & Bildsten, 2017). This being said, the need of implementing data analysis to collect data, such as our own study is doing, seems to be highly relevant in contributing to this area.
Parviainen et al. (2017) however presents that adopting new technologies might also negatively affect companies' operations. The same authors conclude that it is not as easy as putting the current operations into a digitalized process, but instead changing the process of how to operate, from new perspectives to make it work. Parviainen et al. (2017) highlight that all corporations will be affected by digitization; therefore, it would be unwise for companies to assume their current status will remain the same way, in a rapidly developing market. Maintaining market relevance demands proactive acts by companies. This suggests that a master thesis focused on exploring the data provided from a digitalized software can be valuable in contributing to future adoption in a digitalized era.

Blomé (2006) highlights the value of customer relationships within the real estate industry, the study examined how housing companies, that are driven by the municipality, organize their resources to create effective customer relationships. The aim for this was then to increase knowledge of how unlike types of housing accommodation companies work to develop their specific services for all related to the operation. This study found that real estate companies in general recognize the specific weight of fault reporting as one of the key activities in a Facility Manager's work tasks (Blomé, 2006). This finding is also very interesting in understanding more about this subject, such as our own study.

Researchers claim that technology integration in the real estate industry benefits companies and customers significantly. Digitalization delivers even more value-creation for companies and residents, offering more innovative ways to utilize real estate (Andreasson & Mattsson, 2019). According to Andreasson & Mattsson (2019), remote control facilities, communication tools, and services enabled by digitalization make the handling of real estate more accessible and also more convenient for the tenants. It is believed that digital transformation can eventually lead to more content tenants by optimizing the use of resources in the buildings, this resulting in increased income and also reduced expenses over time, as stated by Zalejska-Jonsson (2021).

2.1.1 Business Intelligence

Since the data used in this thesis is conducted through a Business Intelligence platform provided by Momentum Software AB, it can be of relevance for more professionals in the real estate industry to adapt and implement a similar software to collect their data. Business
Intelligence can help the organization in increasing its competitiveness, and also its revenue, by analyzing the operations performances (Gawin & Marcinkowski, 2017). It is a tool that helps inform its user of the environment of the operation by conducting data from multiple sources and putting them together in one place (Tableau, 2023). BI can obtain a large volume of business data, and use that data to drive change. BI is not only applicable within Real Estate and Finance, it is something that has in recent years been recurrently used in other domains (Gawin & Marcinkowski, 2017).

2.2 Digitalizing Facility Management

Researchers have recognized the many benefits of digitalization in the real estate industry, including enhanced efficiency in operations, increased innovation, improved productivity in the workforce, and stronger consumer engagement (Siniak et al., 2020). The integration of technology trends has led to significant changes within the business domain, organizational, and process levels. Thereby also creating opportunities for companies to optimize efficiency and reduce costs (Parviainen et al., 2017).

The significance of facility management in the real estate industry cannot be overstated, particularly when assessing economic forecasts for the upcoming years (Araszkiewicz, 2017). Maintenance and operations are the most costly of a building's life cycle (Wong et al., 2018). When facility management receives a work order, it involves coordinating between real estate, systems, and people, which means to manage and store many details (Korchane & Thorbjörnson, 2022). While some real estate is manually managed and handled, others can operate on automated machines and databases (Cheng et al., 2016). Thus, many researchers indicate that digitalizing the maintenance and operations process could have a significant effect on businesses, which is both costly and time-consuming. Digitalization helps property and facility management to take part in the day-to-day property operations (Baum, 2017). Nonetheless, communication tools have also made it possible for more efficient and useful contact between residents and landlords (Baum, 2017). Digitalization is critical to keep up with the development of facility management services and operations (Araszkiewicz, 2017). The solutions to some facility management challenges can be increased by knowledge of data integration, interoperability, and competence development, which can be enabled via comprehensive training.
Successful maintenance and operation in the real estate industry are highly dependent on the collection and accessibility of relevant information. Specialized employees and different types of data must be integrated and gathered to form a functional facility. During the operation, the property managers must gather the information for maintenance and operations successfully in order to have a sound workflow (Akcamete et al., 2011). The same study continues implying that, with the help of an integrated database that retains facility data, this can help facilitate this exact procedure with operations running smoothly. This, in turn, is anticipated to increase the property's marketability, as it allows easy access to important information for many professionals in the field (Akcamete et al., 2011).

2.3 Fault Reporting within Facility Management

Fault Reporting is an important part as a customer service function within facility management (Åsedal, 2015). The Fault Reporting is used as a tool when the tenant living in the property wants to report something that is broken, or needs to be fixed (Blomé, 2006). Blomé (2006) continues saying that information systems as of today have changed specifically facilities housing management and become an important factor for further development. Nowadays many of the fault reports being made are done on the internet. These technical developments enable many positive factors such as evaluation and also follow-up of what is going on regarding the fault reports (Blomé, 2006).

There was a report going deeper into fault reporting for real estate companies, doing so by conducting a market study on digital fault reports for housing facility management companies (Korchane & Thorbjörnson, 2022). This study aimed to identify the important attributes that specifically housing facility management companies consider when picking a system within digital fault reporting. The study concluded that the most crucial factors for this was; ease of use, integration, economy, follow-up, compatibility, and structuring. The process of fault reporting, if not being digital, can also be based on a phone call process between the tenant, facility manager, and contractor (Korchane & Thorbjörnson, 2022). This study provides valuable information on the importance of digital fault reporting in the real estate industry. However, the report is not going deeper into underlying factors for the fault reports rather than what software to use in making these reports. Therefore, our own thesis can be seen as an interesting way of taking the beat stick to the next phase in getting knowledge from fault reports.
being made. The fault reporting possibility enables the tenant to have full access to the properties functions, as it is something they pay for each month (Blomé, 2006). It is in a way the extended arm for the tenant to reach the facility manager which enables and streamlines the communication between the tenant and facility manager. Blomé (2006) continues that it is central for the real estate organization to increase quality in the service offer by having a focus on it, that can mean focusing and putting a stake into being reliable, available and also responsive. The fault reporting and the process of making these are seen as something that is important and of great value for tenants according to Åsedal (2015). When comparing the value of facility management outsourcing or having it internally, the important factor and value comes down to the importance of the fault reporting process for the tenants behalf (Korchane & Thorbjörnson, 2022). Digital fault reporting can enable facility managers using a platform to understand, analyze, and forecast future fault reports. This can enable them to look for long-term solutions which can also prevent reasons for future fault reports, if using a digital type of system (Korchane & Thorbjörnson, 2022). Having all data collected on the same platform makes the work for making forecasts easier since everything is saved in the same place (Araszkiewicz, 2017).

2.4 Creating Value with Data Analytics

The amount of data conducted around us as individuals and corporations is always growing and seen as something of great value for the ones having the ownership of it. “Data analytics is defined as the application of computer systems to the analysis of large data sets for the support of decisions” (Runkler, 2020, 2). Big data analytics can also be seen as highly relevant knowledge for this study since it refers to a variety of large volumes of data and technology that is then collected from different sources to use this as an advantage against rivals in the same industry with great performance and overview of the business (Shabbir & Gardezi, 2020). Shabbir & Gardezi (2020) continues that firms are in need of analytical insights to handle and to make better business decisions which can then help to a more successful business operation. One example of this is the company “Amazon” which is generating about 35% of the customers' purchases by effectively using big data analytics from the data coming from personal recommendations, through customers. Therefore, it is easy to say big data analytics can be of great importance in achieving and generating a high business value. Big data has changed the face of information Technology and several industries. It is also something being
used and of great relevance for the real estate industries in making smart business decisions. It is usual in understanding for example real estate prices and seeing what patterns customers follow and which factors drive prices (Singh et al., 2020). One paper (Ahmed et al., 2017) putting focus towards exploring and understanding the use of the Big Data concept by looking at barriers, drivers, and opportunities with an emphasis on facilities management, have made some interesting findings. The report suggests that Big data is recognized by the industry but needs guidance and leadership in applying this. Although, It is presented and discussed that the industry holds great potential in adding value. It is said that the industry needs to make better use of their data in hand (Ahmed et al., 2017). This is also a great confirmation as a value driver of the relevance from our specific study going deeper in the data provided by incoming fault reports to be able to understand the facility management industry better.

2.5 Visualizing Data

The human brain is powerful for analyzing data. To then visualize the data is important in being able to document and communicate what the data analysis results really are (Runkler, 2020). Graphical storytelling is of great importance in understanding and reaching the viewer for visually described data. One person that is of great importance for the historical development of how data is visually presented is Florence Nightingale (Scientific American, 2022). The same article continues and describes that this woman recognized and understood that very few people actually read statistical tables. Together with her team, Mrs Nightingale designed graphics that other sources of media were not possible to provide. By presenting charts that did not only describe it as data, with her approach the presentation made it more about data storytelling, and easier to understand. This made it possible to drive great changes in understanding why certain death mortality in war would occur. For example, she showed that peacetime soldiers that were living in barracks died at higher rates than similar aged civilian men. Her way of presenting the data made it possible for army administration to understand that a sanitary reform with priority to fresh air, clean sewers, and less crowding would help reduce the death mortality of the British Army. Based on this it is clear that there is of great importance in how one presents data in order to reach the audience. Applying this knowledge for this specific thesis is something that the authors will strive for since its relevance is of great importance for this type of study. Furthermore, since our study is solely about data and carrying
through a data analysis, this will be emphasized by presenting the data so the reader can clearly understand the data used as well.
3. Theoretical Framework

3.1 Scientific Management Theory

This theory is a management approach that aims to increase productivity and efficiency within organizations by focusing on the workforce job design and processes. This can also be referred as Taylorism, which was presented by Taylor (1909) as he proposed that by simplifying and optimizing jobs, as an effect productivity would increase. Taylor suggests applying scientific principles for methods of facility management pursuance. This would increase productivity for individuals and as an effect increase productivity within the organization. Taylor considered that particularly optimized techniques and resources was a prerequisite in achieving increased effectiveness (Waring, 2016). Applying this as an economic theory to this study does not give the authors a specific direct theory to work with in supporting the result. However, this can confirm the value of developing digital softwares as making the work of fault reporting more effective, within facility management. The effect of that enables us to hold a high amount of data points to then be able to implement data analysis, such as our own thesis, to predict and forecast future events to be able to implement Taylorism in the daily work.

3.2 Explanatory Approach

Can be made by investigating existing literature or other relevant resources with the ambition to seek and discover something new (Elman et al, 2020). Explanatory research can be defined as the soul of good research. Without looking for new answers, to say something new, research would not move forward. However, doing this type of research is seen as risky by definition, because it is not possible to know in advance if the result will be novel and never concluded before. This type of approach is not so specific, it is rather quite general and abstract and can be enabled with many applications. The two most usual scenarios of practicing this theory is through (1) if research already exists, one can study this through a new idea or hypothesis, and (2) if the topic or subject has never been studied.
4. Methodology

4.1 Research Outline

The goal of this study is to conduct a data analysis. The dataset is holding information regarding fault reports, and as the method, this thesis will be using mathematical regression analysis to interpret the underlying factors of fault reports, within three Swedish real estate housing companies’ portfolio of properties. An initial dialogue with our supervisor from the company Momentum Software AB was held regarding the dataset they provided for us to pin which variables would be interesting to test for this thesis. Three Excel-datasets from three different real estate housing companies were conducted and then merged into one big dataset for a more effective workflow. The data was analyzed with the goal of identifying interesting patterns for variables to study further. This excel-file was further executed with Python. All variables were then standardized using built-in functions in Python. To sort the data, dummies were made to put the fault reports into different categories so that they could be possible to analyze as more clear patterns. Multiple regression models have then been tested and used to see which models fit the data and its variables best in its patterns and relationship. Thereafter we were enabled to draw the conclusions. To see the list of definitions, see Table A1 in the Appendix.

4.2 Data Collection

This thesis will mainly get its data from a quantitative data collection. Three large datasets with multiple columns of data points that are related to one another are conducted from Momentum Software AB. These datasets are provided by three real estate housing companies that are using Momentum software AB’s platform for facility management. This platform has multiple functions for the user to utilize, however, this study will solely focus on the data points logged regarding the fault reports that have been made. These three datasets have then been merged together so that the workflow gets more effective working in one dataset. Using three different real estate housing companies' fault reports data enables us to get an overview of how it looks more generally, and not only for one specific company. In total there are 128,271 observations of fault reports made between May 2021 - February 2023. Of all the observations analyzed in the dataset Company F has 101,739 observations of fault reports, Company A has 17,738 observations of fault reports, and finally, Company L has 8,794 observations of fault reports.
4.3 Digitalized and Non-digitalized Fault Reports

The data collected contains both fault reports that are made through two separate ways. Either digitalized or what we call non-digitalized.

4.3.1 Digitalized Fault Reports

The first way is digitalized fault reporting. This means that the fault reports have been made from a software application platform from the tenant through the function called My Pages. This way of fault reporting is what Momentum Software AB would prefer seeing how their clients' tenants make the fault reports since it enables an efficient workflow in handling the fault report process for both the tenant and the real estate company.

4.3.2 Non-digitalized Fault Reports

The second way which is non-digitalized contains fault reports being made either by a phone call directly to the real estate company, e-mailed fault reports, a direct visit to the real estate companies office to make the fault reports, and also fault reports made from staff working for the real estate company that perceives something needs to be reported.

4.4 Software

Through this thesis mainly two softwares have been used and are described in more detail below. There are two reasons for using these softwares, firstly, the knowledge from our supervisors at both Momentum Software AB and also KTH which have good knowledge in helping using these tools. Secondly, the authors of this thesis have some prior experience of these tools so the procedure will be surmountable. In short three separate datasets have been conducted as excel-files and then been sorted, structured, and combined together as one dataset. Furthermore, this dataset has been analyzed by looking for interesting patterns and what possibilities for an analysis we have with this data. As the dataset was prepared and thoroughly understood, it was then transferred into Python where the process of data analysis and the regression model was completed.

4.4.1 Microsoft Excel

Microsoft Excel is a part of an office-package which is used to perform calculation and data management tasks. Excel has been used for this study in order to perform to structure the data
used to make the data analysis in Python Studio which is mentioned below. Excel has also been used to filter categories of the data and explore it in order to understand it better and see which factors would be interesting to look deeper into.

4.4.2 Python
Python is a coding software which enables the user to make powerful data analysis by using different types of functions, called libraries. Python as a software is used to perform the regression analysis by using the three libraries Pandas, NumPy, and Statsmodels. All variables in the dataset used are not numerical, they are representing a categorical variable. Therefore, dummies have been made in Python so that we only have numerical variables to enable the regressions tested, and eventually used for the result.

4.5 Quantitative Approach
Based on the type of research that the authors will investigate, the most reasonable and perhaps obvious choice of method is a quantitative approach. According to Slevitch (2011) the quantitative approach involves using statistics, mathematics, and data to measure and analyze causal relationships in a given phenomenon. Slevitch (2011) continues saying that the approach is based on a realistic perspective that aims to achieve one objective truth and enable an objective and generalizable view.

A quantitative research design can be characterized by the use of variables that have a numerical determined relationship towards each other (Saunders et al., 2019). This is the case for this thesis with the dataset that is applied for the data analysis. Therefore, also a good argument for using this type of approach.

4.5.1 Inductive Approach
Since the dataset being conducted from Momentum Software AB is a dataset the authors of this thesis have not seen and researched prior to this study, the inductive approach came naturally. Even though your research can be clearly defined with an aim and research question, one can be in need of seeing the data to get a feeling of what is actually going on (Saunders et al., 2019). Induction for theory development is more starting with an explorative approach in gaining knowledge and understanding a problem to then later develop a hypothesis, theory, or conceptual framework. The method is based on going from the specific to the general (Saunders
et al., 2019). Using the inductive approach also seemed to be obvious in getting more knowledge of the data to understand what kind of specific factors would have an effect on the fault reporting. And also in finalizing the design of the research question used for the paper that can match a realistic and sound result made.

4.6 Regression Analysis

A regression analysis is a statistical analysis method where one looks for correlations between different variables that are being studied (Blom et al, 2017). It is used to analyze and also measure the relationship between one or more independent variables and a dependent variable. Furthermore, according to Holmes et al (2017), it can also be used to predict and forecast the value of the dependent variable, based on the values from the independent variables.

According to Wegner (2010), regression analysis can be proceeded in different types of ways. There is simple linear regression analysis, which inspects the relationship of two numeric variables solely, this one will be used in this thesis. Correlation analysis, this computes the strength of a certain relationship. And the multiple linear regression analysis, which is a development of the more simple linear regression. For the linear regression there is only one independent variable, x, which is a variable that is assumed to be influencing the fallout of the dependent y variable. However, multiple linear regression analysis is arguably the most widely used statistical modeling approach (Wegner, 2010).

For the use of multiple linear regression analysis, this model looks for the relationship between not only one, but a set of numerical variables and also a dependent variable. If the relationship between the independent and dependent variables are reasonably linear, then this method can be more useful rather than estimating these separately, like the linear regression. (Wegner, 2010).

The following is the form of sample-based multiple linear regression analysis (Blom et al, 2017):

\[ Y_j = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_{ki} + \epsilon_i, \ i = 1, 2, \ldots, n \]  

(1)

Explaining the formula (1), the dependent variable is \( Y_j \), \( \beta_0 \) is the intercept for the regression line and this is the value where the function crosses the \( y \)-axel. \( \beta_1 \) and \( \beta_2 \) is then of purpose in
explaining how the explanatory variables $k$ are affecting the dependent variable. The indicator variable $i$, which can also be called a dummy variable, shows if there is a belonging for any category that can be expected to make a change in the results. The indicator variable is used to sort data in certain categories that will be useful (Blom et al., 2017).

4.6.1 Linear Regression

With the aim to show the relationship between two specific variables, one can use the Linear regression. Doing so by fitting a linear equation to the data in hand. One specific variable is appraised to be an explanatory variable. Furthermore, the other variable will be observed as a dependent variable. It is considered of great importance to interpret and see whether there is a relationship between the intended values by observing the data in a scatterplot. The correlation coefficient is a valuable numerical measure of association between two variables, representing the strength of the observed data association on a scale from -1 to 1. This has also been used to see if our chosen variables would be applicable together in a linear regression (Yale University, 1997).

A precise explanation of the model is described here below:

“A linear regression line has an equation of the form $Y = a + bX$, where $X$ is the explanatory variable and $Y$ is the dependent variable. The slope of the line is $b$, and $a$ is the intercept (the value of $y$ when $x = 0$)” - (Yale University, 1997).

4.6.2 Negative Binomial Regression Model

This model is an unusual statistical model and not very frequently used. Other examples not often used are Poisson regression, logistic regression, or even the probit model (Hilbe, 2011). The negative binomial regression model is very comparable to the multiple regression model (NCSS, 2023). However, the main difference between the regular multiple model and the negative binomial model is that the dependent variable, also known as $Y$, tails the negative binomial distribution as a count variable (NCSS, 2023). Therefore, the representing observed counts are positive values that are zero or higher (NCSS, 2023). Negative binomial regression is applicable when the conditional variance surpasses the conditional mean, for over-dispersed count data (UCLA, 2023). If comparing this model to the Poisson regression model one can identify that the confidence interval tends to be slimmer for the negative binomial regression...
when overdispersion is shown through the outcome variable (UCLA, 2023). This model can be seen as quite similar to the poisson regression because the negative binomial is a generalization of the poisson regression and is popular since it allows the modeling of poisson heterogeneity, by using a gamma distribution (NCSS, 2023).

4.6.3 Poisson Regression Model
A specific case of the generalized linear model is the Poisson model. The Poisson distribution decides the random component. The Poisson regression technique is useful for dependent variables that involve counting events, for instance, the number of customer service calls. In contrast to the binomial distribution, the Poisson distribution doesn't have a maximum limit for the count (Penn State, 2023).
5. Data

The data provided by Momentum Software AB contained three different Excel files. The three separate Excel files were then combined into one single dataset, including different tabs with specific titles to categorize the data. The dataset was initially reviewed using Excel and subjected to exploratory data analysis, which involved visualizing the data to identify any potentially noteworthy patterns or trends. Once interesting findings were identified, specific data from various tabs were selected and moved to a common tab for further analysis. This included columns such as `numbers_of_rooms`, `area_m2`, `rent`, `commencement_date`, `contract_period`, and `tenant_age`. Further, the age of the different buildings age was calculated by subtracting the construction year from 2023. This process permitted us to compile the relevant data for further analysis.

After finalizing the Excel data, we explored it using Python. To be able to conduct data exploration and preprocessing in Python, we started importing the necessary libraries, such as Pandas, NumPy and Statsmodels. We then read/imported the Excel file provided by Momentum Software AB and converted the Swedish titles to English. Next, we filtered the data to only include apartments, as our objective was to analyze fault reporting for this type of dwelling. Additionally, we filtered relevant columns for the analysis; these categories included company, residential area, building name, apartment ID, building age, number of rooms, area in square meters, rent, commencement date, contract period, tenant age, and date of service order.

After filtering, we reviewed the data to identify rows with NaN (Not A Number) values, which accounted for 15,293 of the total 88,327 rows. These NaN-containing rows were dropped, resulting in a total of 73,034 data points. We then grouped the data by apartments, which combined numerous rows for the same apartment into a single line, with the number of service orders summed. Therefore, also the count of used data points is then filtered down and lower since we have not used all for the regression analysis. This is because combining all variables in the dataset into one regression would not enable us to make a good model. Thereafter, a descriptive statistics was created which can be viewed in the Appendix, *Table A2.*
5.1 Descriptive Statistics

We continued the data analysis by visualizing the data to get an even more comprehensive understanding. Specifically, we analyzed the age of the building that we previously calculated in the Excel file. Looking at the graph below, the y-axis describes the count, representing the number of buildings. The x-axis shows the building’s age ranging from 2 to 118. To interpret the graph, the age of 2 indicates that the building was constructed two years prior to the year of 2023.

![Graph](image)

*Figure 1: Number of Buildings by Building Age. Note: This bar graph describes the spread of the buildings ages.*

By studying *Figure 1* above, we observed that the count is the highest when the buildings are 53 years old. By further analyzing the graph one can also detect that many houses were being constructed around the same time. This finding implies that our dataset's highest number of buildings was constructed during the 1970s.

The analysis continued on by looking into the amount of apartments the three different real estate companies of our dataset own. We have named the companies F, L, and A to show this distribution in *Figure 2.*
Figure 2: Number of Service Orders per Company. Note: This pie chart describes the spread of service orders per company in proportion to one another.

Looking at Figure 2 above, Company F stands for the majority of service orders that have been researched with about 76.8% of the data. Thereafter Company L and A stands for the rest of the service order, which is about 8.9% and 14.3% respectively.

Furthermore, we wanted to see the partition of fault reports made. As Figure 3 below describes, on the x-axis the fault reports origin, and on the y-axis the number of fault reports. Here we could see that about 90% of the service order observations came from the category non-digitalized fault reports and roughly 10% came from the category digitalized fault reports.

Figure 3: Number of Fault Reports by Origin. Note: This bar graph describes the spread of fault reports by either Non-digitalized or Digitalized fault reports.
Thereafter, we analyzed the tenant age category by creating Figure 4 above that shows the number of tenants on the y-axis, and the different age categories on the x-axis. The graph shows that the largest age category of tenants is between 40-54 years old. We could also observe that the age categories are fairly evenly distributed among the different age groups. However, the two categories that differ the most are “missing age” and the 17 - 25 age group, which are significantly smaller compared to the other categories.

![Figure 4: Number of Fault Reports by the Age of Tenant. Note: This bar graph describes the spread of the different tenant age-groups in relationship to one another.](image)

Furthermore, we saw the spread on the sizes of the apartments represented in Figure 5. The largest apartment size category of tenants is between 50-69 m². We could also observe that the majority of apartments were sized between 50-69m² as well as 70-89m².

![Figure 5: Apartment Size Distribution. Note: This pie chart describes the spread of apartment sizes measured in square meters in proportion to one another.](image)
Figure 6 below shows the spread on the number of rooms these apartments have. The number of apartments is being displayed on the y-axis, and the different number of rooms these apartments have on the x-axis. It is interesting to note that the sizes of apartments have a tendency to correlate together with the amount of rooms. Where most of the apartments are 2 rooms, or 3 rooms, and then also some with 4 rooms.

Figure 6: Number of Rooms by Number of Apartments. Note: This bar graph describes the spread of apartments numbers of rooms in relationship to one another.

Figure 7: Monthly Rental Cost by Number of Apartments. Note: This pie chart describes the spread of monthly cost to live in the apartments in proportion to one another.

Figure 7 shows the distribution of the monthly cost tenants pay each month to rent apartments. Values under 1% were excluded from Figure 7; therefore, the values 14 000 - 15 999
SEK/month, which constituted 0.08%, and 16 000+ SEK/month, which constituted 0.01% of the rents of the apartments, were excluded from the graph. The most common monthly cost for the tenants living in the apartments are between 6000-7999 SEK/month. It is also quite usual to see monthly costs around 4000-5999 SEK/month and 8000-9999 SEK/month.

With this information shown in Figure 8 we can observe that the largest categories of tenants have lived in the apartments between 10-30 years. Then followed by a 0-1 year period which is the opposite direction to the one mentioned prior. Many tenants have had their contracts for 1-2 year periods, as well as 2-3 year periods. The patterns regarding this are not very clear when looking at it initially. Many tenants live in the apartments for a long time, and many others live in the apartments for a very short period of time as well according to this observation.
According to Figure 9, in the 4 room category, 57.29% constitutes rents between 8 000 - 9 999 SEK/month, and 63.13% of those with five rooms. Rents between 10 000 - 11 999, 15.55% of 4 room apartments have rents between 10 000 - 11 999 SEK/month, while 28.59% in 5-room apartments fall in the same category. The feature 12 000 - 13 999 SEK/month is observed in apartments with 4, 5, and 6 rooms. However, it has its highest presence of 6 room apartments, where 88.72% have rents in this range. The rent interval 6 000 - 7 999 SEK/month has a notable presence across all categories except for 6 rooms, with 17.94%, 55.04%, 78.12%, and 21.84% in 1, 2, 3, and 4 rooms, respectively. By looking at Figure 9, we are able to see clear associations between rents and the number of rooms with some exceptions. Note that rent categories under 1% have been removed from Figure 9.
6. Results

6.1 Exploratory Regression Analysis

We created a correlation matrix of the features and the dependent variable. We identified features with pairwise correlation and observed that the number of rooms and the area of apartments were correlated. This is displayed by the black circles in Figure 10 below.

Figure 10: Correlation Matrix from the Variables. Note: This correlation matrix shows used variables for the regression analysis with a focus on showing the correlation on the area categories and room categories identified by the two blue ellipses.

To be able to reduce the impact of this correlation, we excluded the area categories 30-49, 50-69, and 70-89 from our data. The updated correlation matrix can be found below in Figure 11:
With the high-correlation values removed, we proceeded to conduct an Ordinary Least Squares (OLS) regression and generated a coefficient plot function based on the OLS results.

6.2 Ordinary Least Squares (OLS) Regression and Poisson Regression

We completed the Ordinary Least Squares (OLS) and generated a coefficient plot function to visualize our results. However, since our dependent variable is count data, we switched to conducting a Poisson regression analysis instead, which is better suited for count data.

The Poisson model does, however, not perform well for overdispersed data; therefore, we needed to check the dispersion. Overdispersion occurs if the variance is greater than the mean (Dean & Lundy, 2016). The first indication of overdispersed data in the Poisson model is if the
deviance is much larger than the degrees of freedom (df). This is true for our data as the deviance of 41863 is much greater than the df of the model with a value of 29.

To confirm the overdispersion indication, we proceeded to calculate the dispersion parameter by using the (2) formula (Payne et al, 2018):

\[
\hat{\phi} = \frac{\sum_i (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i}{n - p}
\] (2)

The resulting dispersion parameter was 1672.17, which is large, confirming that our data is overdispersed. To mitigate this overdispersion, we performed a Negative Binomial regression, as this regression is better suited for overdispersed data.

After conducting the following test on the OLS and Poisson regression models, we concluded that the Negative Binomial regression was the best-suited regression model for us to use in our analysis. Therefore, this data analysis was completed by conducting Negative Binomial regression and generated a coefficient plot function to display the results.

6.3 Analysis of the Negative Binomial Coefficients

6.3.1 Underlying Factors for Fault Reports

The coefficient plot function below displays the coefficient values on the y-axis, representing the features' impact on the number of service orders. To see the summary statistics, see Table A3 in the Appendix.
The blue lines on the markers in NB coefficient graph above represents the coefficient’s confidence interval. It indicates that the deviation of the coefficient. Therefore, the longer the blue line, the higher the uncertainty in the coefficient range. The red markers on the graph represent features with coefficients that negatively influence the number of service orders on the whole coefficient range. Similarly, the green markers represent features with coefficients that positively impact the number of service orders on the whole coefficient range.
Table 1: Statistical Proven Variables for Fault Reports in General. Note: The following coefficient table shows statistically proven variables for fault reports in general.

<table>
<thead>
<tr>
<th>Features</th>
<th>Positive/Negative</th>
<th>coef</th>
<th>[0.025]</th>
<th>[0.975]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 29 m²</td>
<td>Negative</td>
<td>-0.3608</td>
<td>-0.676</td>
<td>-0.045</td>
</tr>
<tr>
<td>30+ year period</td>
<td>Negative</td>
<td>-0.2801</td>
<td>-0.420</td>
<td>-0.141</td>
</tr>
<tr>
<td>4 - 5 year period</td>
<td>Negative</td>
<td>-0.1053</td>
<td>-0.209</td>
<td>-0.002</td>
</tr>
<tr>
<td>building_age</td>
<td>Positive</td>
<td>0.0033</td>
<td>0.002</td>
<td>0.005</td>
</tr>
<tr>
<td>3 - 4 year period</td>
<td>Positive</td>
<td>0.1088</td>
<td>0.015</td>
<td>0.202</td>
</tr>
<tr>
<td>70+ tenant age</td>
<td>Positive</td>
<td>0.1759</td>
<td>0.108</td>
<td>0.243</td>
</tr>
<tr>
<td>3 room</td>
<td>Positive</td>
<td>0.2023</td>
<td>0.049</td>
<td>0.355</td>
</tr>
<tr>
<td>missing age</td>
<td>Positive</td>
<td>0.2127</td>
<td>0.098</td>
<td>0.328</td>
</tr>
<tr>
<td>55 - 69 tenant age</td>
<td>Positive</td>
<td>0.2302</td>
<td>0.164</td>
<td>0.296</td>
</tr>
<tr>
<td>26 - 39 tenant age</td>
<td>Positive</td>
<td>0.2535</td>
<td>0.188</td>
<td>0.319</td>
</tr>
<tr>
<td>5 room</td>
<td>Positive</td>
<td>0.2650</td>
<td>0.032</td>
<td>0.498</td>
</tr>
<tr>
<td>40 - 54 tenant age</td>
<td>Positive</td>
<td>0.2916</td>
<td>0.227</td>
<td>0.356</td>
</tr>
<tr>
<td>10 000 - 11 999 SEK/month</td>
<td>Positive</td>
<td>0.3365</td>
<td>0.058</td>
<td>0.615</td>
</tr>
<tr>
<td>8 000 - 9 999 SEK/month</td>
<td>Positive</td>
<td>0.3700</td>
<td>0.106</td>
<td>0.634</td>
</tr>
<tr>
<td>4 room</td>
<td>Positive</td>
<td>0.3777</td>
<td>0.225</td>
<td>0.530</td>
</tr>
<tr>
<td>2 - 3 year period</td>
<td>Positive</td>
<td>0.3929</td>
<td>0.307</td>
<td>0.479</td>
</tr>
<tr>
<td>6 000 - 7 999 SEK/month</td>
<td>Positive</td>
<td>0.4148</td>
<td>0.151</td>
<td>0.679</td>
</tr>
<tr>
<td>100+ m²</td>
<td>Positive</td>
<td>0.4573</td>
<td>0.276</td>
<td>0.639</td>
</tr>
<tr>
<td>1 - 2 year period</td>
<td>Positive</td>
<td>0.4831</td>
<td>0.410</td>
<td>0.556</td>
</tr>
<tr>
<td>0 - 1 year period</td>
<td>Positive</td>
<td>0.4929</td>
<td>0.424</td>
<td>0.562</td>
</tr>
</tbody>
</table>

To be able to answer research question number (1), we conducted a Negative Binomial regression. From the Figure 12 and Table 1 above, we observed that features building age, 100+ m², 0 - 1 year period, 1 - 2 year period, 2 - 3 year period, 3 - 4 year period, 26 - 39 tenant age, 40 - 54 tenant age, 55 - 69 tenant age, 70+ tenant age, missing age, 3 room, 4 room, 5 room, 6 000 - 7 999 SEK/month, 8 000 - 9 999 SEK/month, and 10 000 - 11 999 SEK/month positively affects the coefficient, meaning these features contribute to more service orders being made. On the other hand, the results showed that the features 0 - 29 m², 4 - 5 year period and 30+ year period negatively influenced the number of service orders, indicating that these features contributed to fewer service orders. It is important to note that the Negative Binomial magnitude of the coefficients needs to be interpreted on the exponential scale.

6.3.2 Underlying Factors for Digitalized Fault Reports

To be able to answer the research question (2a), we needed to make a regression separately for this matter. Since we earlier understood that the Negative Binomial Regression model fits our model best, this is the type of regression used. To see the summary statistics, see Table A4 in the Appendix.
Table 2: Statistical Proven Variables for Digitalized Fault Reports. Note: The following coefficient table shows statistically proven variables for digitalized fault reports.

<table>
<thead>
<tr>
<th>Features</th>
<th>Positive/Negative</th>
<th>coef</th>
<th>[0.025]</th>
<th>0.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 1 year period</td>
<td>Positive</td>
<td>0.1767</td>
<td>0.004</td>
<td>0.349</td>
</tr>
<tr>
<td>40 - 54 tenant age</td>
<td>Positive</td>
<td>0.2476</td>
<td>0.074</td>
<td>0.421</td>
</tr>
<tr>
<td>26 - 39 tenant age</td>
<td>Positive</td>
<td>0.2822</td>
<td>0.108</td>
<td>0.456</td>
</tr>
</tbody>
</table>

The Figure 13 and Table 2 above shows that it was statistically proved that the features 0 - 1 year period, 40 - 54 tenant age, and 26 - 39 tenant age affect the number of digitalized service orders made positively. On the contrary, no statistical evidence could show a negative impact on the number of digitalized service orders made.
6.3.3 Underlying Factors for Non-digitalized Fault Reports

To answer the research question (2b), we conducted a Negative Binomial regression and generated a coefficient plot function to display the results. To see the summary statistics, see Table A5 in the Appendix.

![Coefficient Plot for Non-digitalized Fault Reports](image)

Figure 14: Coefficient Plot for Non-digitalized Fault Reports. Note: This table shows the Negative Binomial Coefficient plot for non-digitalized fault reports.
When looking at the coefficient plot for Figure 14 and Table 3 above, we can confirm that there are variables that positively impact the number of fault reports being made since we can see some of the variable dots on the graph above are green. The variables 6 000 - 7 999 SEK/month, 8 000 - 9 999 SEK/month, 10 000 - 11 999 SEK/month, 12 000 - 13 999 SEK/month, 0 - 1 year period, 1 - 2 year period, 2 - 3 year period, 3 - 4 year period, 26 - 39 tenant age, 40 - 54 tenant age, 55 - 69 tenant age, 70+ tenant age, missing age, building age, 3 room, 4 room, 5 room and, 100+ m2 positively affect the number of non-digitalized service orders made. The features that statistically negatively impact the number of non-digitalized service orders made are 14 000 - 15 999 SEK/month, 30+ year period, and 4 - 5 year period.

6.4 Validity

The validity of this research conducted is considered to be good. According to the source from where the data has been received, our supervisor from Momentum Software AB, the data used in this study is collected for the main purpose of its users to find out information and gain input from all the fault reports made on their properties. However, analyzing the data, such as our own thesis has done, has not been prioritized and therefore there is of good relevance in
conducting this research. The result of this study is possible to reproduce as of today, however it might be changed based on the developing use of fault reports being more digitalized or non-digitalized. Furthermore, multiple types of regression models have been used so that the most suitable model is used for our study. Based on the conclusions drawn for the chosen regression model, this was the best suited one.

6.5 Reliability

The reliability of this study is also considered to be good. The objectives of what is going to be researched is clearly stated. The result presents that multiple regressions have been tested to fit the most appropriate one for the importance of this study's reliability. The result is made from the same dataset for all regression models, the one thing that is changed is the origin of the fault reports to be able to answer all the three research questions. Based on the specific dataset it should be possible for future studies within this subject to find similar variables that affect fault reports. Last but not least, the method is clearly thoroughly showing how the result has been conducted.

6.6 Relevance

This research aims to present the positive effects of what digitalizing real estate can bring for the future, and more specifically as an effect what data analysis can bring in economical benefits within facility management. This enables decision makers to see relevant patterns and trends by forecasting and understanding how humans use and think about the use of properties. Thereafter, the decision makers can save money and learn how to prevent cost driving events’ for their own business. .
7. Discussion

The first question for this research paper was broad and the purpose was to understand the underlying factors for the number of fault reports made. In being able to answer this question the explanatory approach was applied so that we could understand the data better. Multiple scientific papers were overviewed to get an understanding of the view of both the possibilities of digitalization for facility management, and also more specifically understand more concerning fault reports for tenants. Starting to review the data makes us quickly understand that there are multiple factors causing fault reports to be made.

Generally, we have come to understand that there are multiple factors causing fault reports and the underlying factors for them being made. This thesis has only discovered this topic from a quantitative approach of data and looked for conjunctions by using regression analysis for this matter. However, there are some interesting things that can be confirmed as of what the data analysis conducted has informed us.

The results in this thesis are exclusively derived from the dataset's analysis, making it difficult to conclude why the results appear the way they do based on the dataset alone. However, the findings in this paper open up for discussions of possible reasons why the results may arise in the way they do.

7.1 Underlying Factors for Fault Reporting

When analyzing the results, we were able to identify that the features 3 room, 4 room, 5 room, 0 - 1 year period, 1 - 2 year period, 2 - 3 year period, 3 - 4 year period, 6 000 - 7 999 SEK/month, 8 000 - 9 999 SEK/month, 10 000 - 11 999 SEK/month, 26 - 39 tenant age, 40 - 54 tenant age, 55 - 69 tenant age, 70+ tenant age, missing age, building age and 100+ m² positively affect the numbers of service orders. In contrast, the results could statistically prove that the features negatively influenced the number of service orders are 0 - 29 m², 4 - 5 year period, and 30+ year period.

Interpreting the results, we can identify that apartments with multiple rooms positively affect the number of service orders. This could be explained by larger apartments typically having more items and a larger area. If the number of items increases, the likelihood that some will
break will also increase. It could be discussed that this consequently leads to more service orders being generated. Kytömäki (2020) argues that digitalization can improve resource efficiency and services for tenants and building users. By conducting the data analysis, property managers can use these results when forecasting and preparing for maintenance needs.

We also discovered that a 0 - 1 year period, 1 - 2 year period, 2 - 3 year period, and 3 - 4 year period positively affect the number of service orders being made. At the same time, the results show that a 4 - 5 year period and a 30+ year period negatively affect service orders. It is noteworthy that there seems to be a shift in service orders made depending on how long the residents have lived in the apartments. Tenants living in the apartment for less or equal to four years make more service orders than tenants staying for more than four years. Even though the features 5 - 10 year period and 10 - 30 year period are not proven to affect service orders negatively, they have more of a tendency to lean towards having a negative effect on service orders. This aligns with the discovery that tenants who have lived up to four years in the apartments will make more service orders than residents who have lived in the apartment longer than four years. While it is difficult to reach a definitive conclusion based on the dataset alone, it is possible that tenants at the beginning of their residency may have higher expectations regarding the apartment's condition. Since the tenants have not stayed in the apartment for as long, they might want to hold the company accountable for any issues that arise since they might believe that they are not the reason for things breaking that early on. Therefore, in the event of a malfunction or defect, the residents might attribute it to the real estate company's negligence or a warranty issue and, consequently, be more inclined to file a service order. It is also possible that tenants that have lived in the apartments longer than four years might already have put in all the necessary service orders, resulting in fewer fault reports the longer they have stayed in the apartment.

The results also show that the rent intervals 6 000 - 7 999 SEK/month, 8 000 - 9 999 SEK/month and, 10 000 - 11 999 SEK/month has a positive outcome on the numbers of service orders. This finding is consistent with Figure 9; which displays the distribution of rent and the number of rooms in the apartments. By looking at Figure 9, we are able to see clear associations between rents and the number of rooms with some exceptions. Therefore, this explains the results fairly well by implying that rents depend on how many rooms an apartment has. Allowing us to
discuss that the more rooms an apartment has, the higher likelihood of service orders due to the greater number of items and the higher possibility of malfunctioning.

The study displays that the age groups 26-39, 40-54, 55-69, missing age, and 70+ positively influence the number of service orders made, while statistical evidence for age features 17-25 is inconclusive. This suggests that the likelihood of service orders being made is positive across all age groups, but the impact is particularly notable for tenants aged 26-70+.

Another feature to be statistically proven to affect the number of service orders positively is $100+ \text{ m}^2$. This indicates that apartments larger than $100+ \text{ m}^2$ positively impact the number of service orders being made. As an outcome, it is possible to discuss the likelihood of a higher number of service orders for larger apartments than for smaller ones. This reflection is strengthened by our results finding that $0 - 29 \text{ m}^2$ tends to affect service orders negatively since it aligns with the discovery that larger apartments affect service orders more than smaller ones. When we conducted the correlation matrix, we excluded the area features $30 - 49 \text{ m}^2$, $50 - 69 \text{ m}^2$, and $70 - 89 \text{ m}^2$ from the dataset. Since these features were excluded, no results could be accomplished on these specific categories. We, therefore, could neither confirm nor deny any positive or negative effects of these features on the number of service orders.

The apartments with more square meters have more service orders but simultaneously take up a larger space in the building. With the results, we are unable to conclude if the service order per square meter differs between large and small apartments.

The feature building age positively affects the number of service orders made, being statistically proven according to the regression results. The result does not tell us whether it is the older buildings or the newer buildings that have affected the number of fault reports. However, initially before the thesis, we would think and argue that older buildings need more maintenance and updates to work well by today's standards; however, this is not something this thesis will go deeper into.
7.2 Non-digitalized Fault Reports

The comparison of the general and non-digitalized regression results shows that the features that positively and negatively affect service orders are the same, except for three instances. The feature 0 - 29 m² negatively affects service orders in the general regression, whereas for non-digitalized regression results, there is no statistical evidence for this claim. In the non-digitalized regression results, the features 12 000 - 13 999 SEK/month and 14 000 - 15 999 SEK/month negatively affects service orders, whereas, in the general regression results, there is no statistical evidence to support this. It is worth noting that although these features lack statistical significance, the corresponding figures suggest that they demonstrate similar movement with the other coefficient graph. The resemblance in the trends of the general regression results and the non-digitalized regression results could potentially be explained by the significant presence of non-digitalized service orders in the dataset. Since roughly 90% of the data points in the dataset came from non-digitalized reports, and less than 10% came from digitalized reports, this skew towards non-digitalized service orders has presumably influenced the general and non-digitalized regression analyses results since most of the dataset consists of shared datapoints.

7.3 Digitalized Fault Reports

The results from the digitalized regression analysis show that tenants aged 26-54 positively influence digital service orders made. It is reasonable to believe that people who grew up with digital devices would be more willing and comfortable using digital solutions. However, while the feature 55-69 tenant age is not statistically significant for digital service orders, it has a tendency of leaning toward a positive effect on service orders. This allows us to discuss that all age categories use the digital platform when reporting service orders. However, since the result was not statistically proven, this is not something we can conclude.

Another finding in the regression results is that tenants that have lived in the apartment for 0-1 year positively affect the number of service orders. This suggests that tenants living in the apartment for a shorter period of time generate more digital service orders compared to tenants who have lived longer than one year. This could be explained by the fact that the companies have been better at informing and guiding new tenants on how to use the digital platform to report service orders and while maybe not have been as good at informing current residents to
use the digital platform. Therefore, if companies want to encourage residents to use the digital platform, launching a campaign targeting residents that have been in the apartment for longer than a year might be beneficial.

The results we got from making separate regressions for non-digitalized and digitalized fault reports were quite different in comparison for them both. Having roughly 90% of all the observations in the dataset used coming from non-digitalized fault reports, and roughly 10% coming from digitalized fault reports, might have affected these results. Generally, the output results seem to be so dominated by observations being non-digitalized fault reports, there are not enough observations for digitalized fault reports to give us answers in underlying factors for those fault reports separately. The results we got from making separate regressions for non-digitalized and digitalized fault reports were quite different in comparison for them both. Having roughly 90% of all the observations in the dataset used coming from non-digitalized fault reports, and roughly 10% coming from digitalized fault reports, might have affected these results. In being able to make this analysis and compare the two separate ways of making the fault reports it might be valuable to make this analysis when there is more data to use for these. Perhaps not that they need to be perfectly balanced in the dataset standing for 50% of observations respectively, but that there are loads of more observations to use for making a data analysis for this cause.

7.4 Acceptance of Digitalization

Parviainen et al. (2017) highlight that all corporations will be impacted by digitization and that it would be unwise for companies to assume their current status will remain the same in a rapidly developing market. By taking the leap into digitalizing the operations of the properties the three real estate companies whose fault reports observations have been analyzed in this thesis are themself breaking ground in adapting to a more digitalized era. Blomé (2006) says that information systems as of today have changed facilities housing management and become an important factor for further development. Since this is the steps that seems to be necessary to take in being able to adapt to the change of how to operate properties, it is interesting to see from the result above that roughly 10% of the observations are “only” made from digitalized fault reports. Even though the other roughly 90% made from non-digitalized matters can come from multiple sources, the numbers are quite unbalanced according to ourselves. The big
difference in how many fault reports are being made via digitalized tools can of course be caused by multiple factors that we might not understand by having solely the information and results gathered above. However, some points that can be the reason for this might be possible to argue for. Parviainen et al. (2017) mentioned that adopting new technologies might negatively affect companies’ operations. In the beginning moving towards change might not be positive directly. Would it be possible to have the amount of digitalized fault reports to be higher as a percentage value from the total amount of fault reports? What would it take for the digitalized fault reports to be 20%, rather than 10%, which is what our data says. It is also important to give the tenants the possibility of utilizing the application and services that the real estate company themself want the tenants to use. The fault report function is an extended arm for the tenant to reach the facility manager which enables and streamlines the communication between the two parties (Blomé, 2006). Åsedal (2015), Blomé (2006), and Korchane & Thorbjörnson (2022) all mention the value of fault reporting as an important function for the tenants behalf. If this process is digitalized and developed as an advantage for all parties, what is the reason for it not to be utilized even better?

7.5 Data Analytics & Facility Management

According to Araszkiewicz (2017) and Atkin & Bildsten (2017) big data can help the development of facility management with efficient maintenance and cutting operation costs. The possibilities should not be underestimated. The industry needs to make better use of their data in hand (Ahmed et al, 2017). Therefore, the value of taking the data that has been collected over the years seems to confirm that it is important to analyze and learn from this data. Many companies active in the space of facility management seem to have numerous data points to be explored regarding many more areas than solely fault reporting, according to our literature study. The data used for our specific thesis has actively been saved for the purpose of analyzing it further and gaining knowledge from it. In being able to do so it seems to be necessary for real estate professionals to embrace and try to use the data to come forward, even though the results and wanted outcomes from the analysis might not come directly. Kytömäki’s Licentiate thesis (2020) strongly suggests focusing on improving resource efficiency and services for tenants and building users. It all seems to come down to making the real effort for the leap of digitalizing the industry of facility management. Getting the most out of real estate comes down to how bad the professionals of the facility management industry are willing to transfer towards
a more digitalized world. The softwares is in use, it seems like the people anyhow involved in one way or another, need guidance from the most knowledgeable professionals in this area, on how to use this data in its best way.
8. Summary and Conclusion

In this thesis we have researched the area of underlying factors for fault reports made non-digitalized as well as digitalized. By applying three research questions we have explored the real estate and facility management industries' stand on digitalization, and also made several regressions to gather necessary information for a result and discussion regarding this topic.

The first question, regarding the underlying factors for fault reports, was answered first by exploring the data of interesting variables to design a regression model. The output from the Negative Binomial Regression model was the one best suited for all our regressions in the dataset. The regression output confirmed multiple variables which were statistically proven which positively affect the number of fault reports being made. The features building age, 100+ m², 0 - 1 year period, 1 - 2 year period, 2 - 3 year period, 3 - 4 year period, 26 - 39 tenant age, 40 - 54 tenant age, 55 - 69 tenant age, 70+ tenant age, missing age, 3 room, 4 room, 5 room, 6 000 - 7 999 SEK/month, 8 000 - 9 999 SEK/month and, 10 000 - 11 999 SEK/month positively affects the coefficient, meaning these features contribute to more service orders. On the other hand, the results showed that the features 0 - 29 m², 4 - 5 year period and 30+ year period negatively influenced the number of service orders.

The second question, regarding which underlying factors affect only the digitalized fault reports had the following findings. The results from the digitalized regression analysis show that tenants aged 26-54 positively influence digital service orders made. Another finding is also that tenants that have lived in the apartment for 0-1 year positively affect the number of service orders.

The third question, regarding which underlying factors affect only the non-digitalized fault reports, also had some findings. The comparison of the general and non-digitalized regression results shows that the features that positively and negatively affect service orders are the same, except for two instances. The feature 0 - 29 m² negatively affects service orders in the general regression, however in the non-digitalized regression results there is no statistical evidence for this claim. In the non-digitalized regression results, the feature 14 000 - 15 999 SEK/month
negatively and the feature 12 000 - 13 999 SEK/month positively affects service orders, whereas, in the general regression results, there is no statistical evidence to support this.

8.1 Further Research

Regarding the first research question, we suggest that further research should be made regarding the fault report variables. It should be possible to statistically prove more variables that affect the number of fault reports if these variables could be saved and held from other sources. It would also be interesting to get an understanding of how much time and money the real estate companies can save by making the fault report digitalized. This could be enabled through interviews with industry professionals and thereafter get a more specific understanding in the viewpoint of the users from this type of services.

During this study we have only used fault reports ranging from about two years back in time. It would be very interesting to see if making the same type of study with even more data. If the same type of research could be conducted in five years from now, maybe the results will show something new. This could also enable the development of perhaps seeing more fault reports with the origin of being made digitalized. It might be the case that the digitalized fault reports show more variables for a future study that are statistically proven to have an impact or not for fault reports.

Lastly, it would be interesting to make a similar analysis by looking at either one company's fault reports so that the result is specific for only that company, this could be valuable in understanding the use of fault reports for one specific company in a region. Or else, make an analysis on even more fault reports from more companies than our own thesis handles in order to get a more general result and reflect the use of the fault reporting function in a wider context.
9. References


Blom et al. (2017). Sannolikhetsteori och statistikteori med tillämpningar, Studentlitteratur AB.


10. Appendix

Table A1: List of Definitions. Note: This table shows the definitions of the variables from the dataset.

<table>
<thead>
<tr>
<th>Company</th>
<th>The real estate companies represented by a letter (A, F &amp; L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential area</td>
<td>The area in which the property is located</td>
</tr>
<tr>
<td>Building name</td>
<td>The name of the building</td>
</tr>
<tr>
<td>Apartment ID</td>
<td>The apartment ID for the apartment</td>
</tr>
<tr>
<td>Building age</td>
<td>The number of years since the building was constructed</td>
</tr>
<tr>
<td>Number of rooms</td>
<td>The number of rooms the apartments have</td>
</tr>
<tr>
<td>Area in square meters</td>
<td>The apartment size in square meters</td>
</tr>
<tr>
<td>Rent</td>
<td>The rent of the apartment</td>
</tr>
<tr>
<td>Commencement date</td>
<td>Move-in date of the tenant</td>
</tr>
<tr>
<td>Contract period</td>
<td>The period during which the tenant has been a resident in the apartment</td>
</tr>
<tr>
<td>Tenant age</td>
<td>The age of the tenant</td>
</tr>
<tr>
<td>Date of service order</td>
<td>The date the service order was reported</td>
</tr>
</tbody>
</table>

Table A2: Descriptive Statistics. Note: This table displays the descriptive statistics for all variables used in this study.

<table>
<thead>
<tr>
<th>NumberOfServiceOrders</th>
<th>count</th>
<th>mean</th>
<th>std</th>
<th>min</th>
<th>25%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>building age</td>
<td>7303</td>
<td>10.0</td>
<td>9.22</td>
<td>1</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>1 room</td>
<td>7303</td>
<td>41.9</td>
<td>18.11</td>
<td>2</td>
<td>33</td>
<td>46</td>
</tr>
<tr>
<td>2 room</td>
<td>7303</td>
<td>0.4</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3 room</td>
<td>7303</td>
<td>0.38</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4 room</td>
<td>7303</td>
<td>0.11</td>
<td>0.31</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5 room</td>
<td>7303</td>
<td>0.01</td>
<td>0.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 room</td>
<td>7303</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0 - 29 m²</td>
<td>7303</td>
<td>0.01</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100+ m²</td>
<td>7303</td>
<td>0.05</td>
<td>0.23</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>90 - 99 m²</td>
<td>7303</td>
<td>0.05</td>
<td>0.21</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 - 3 999 SEK/month</td>
<td>7303</td>
<td>0.03</td>
<td>0.17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10 000 - 11 999 SEK/month</td>
<td>7303</td>
<td>0.04</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12 000 - 13 999 SEK/month</td>
<td>7303</td>
<td>0.01</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14 000 - 15 999 SEK/month</td>
<td>7303</td>
<td>0.03</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16 000+ SEK/month</td>
<td>7303</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4 000 - 5 999 SEK/month</td>
<td>7303</td>
<td>0.21</td>
<td>0.41</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 000 - 7 999 SEK/month</td>
<td>7303</td>
<td>0.55</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8 000 - 9 999 SEK/month</td>
<td>7303</td>
<td>0.16</td>
<td>0.36</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0 - 1 year period</td>
<td>7303</td>
<td>0.16</td>
<td>0.37</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 - 2 year period</td>
<td>7303</td>
<td>0.13</td>
<td>0.34</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10 - 30 year period</td>
<td>7303</td>
<td>0.28</td>
<td>0.45</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 - 3 year period</td>
<td>7303</td>
<td>0.08</td>
<td>0.27</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3 - 4 year period</td>
<td>7303</td>
<td>0.07</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30+ year period</td>
<td>7303</td>
<td>0.03</td>
<td>0.17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4 - 5 year period</td>
<td>7303</td>
<td>0.05</td>
<td>0.23</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5 - 10 year period</td>
<td>7303</td>
<td>0.18</td>
<td>0.39</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>vacant</td>
<td>7303</td>
<td>0.01</td>
<td>0.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17 - 25 tenant age</td>
<td>7303</td>
<td>0.03</td>
<td>0.17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>26 - 39 tenant age</td>
<td>7303</td>
<td>0.22</td>
<td>0.42</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>40 - 54 tenant age</td>
<td>7303</td>
<td>0.25</td>
<td>0.43</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>55 - 69 tenant age</td>
<td>7303</td>
<td>0.22</td>
<td>0.42</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>70+ tenant age</td>
<td>7303</td>
<td>0.23</td>
<td>0.42</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>missing age</td>
<td>7303</td>
<td>0.05</td>
<td>0.22</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table A3: Summary Statistics for Fault Reports in General. Note: This table describes summary statistics from the negative binomial model for fault reports in general.

<table>
<thead>
<tr>
<th>Dep. Variable</th>
<th>NumberOfServiceOrders</th>
<th>No. Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>GLM</td>
<td>Df Residuals</td>
</tr>
<tr>
<td>Model Family</td>
<td>NegativeBinomial</td>
<td>Df Model</td>
</tr>
<tr>
<td>Link Function</td>
<td>Log</td>
<td>Scale</td>
</tr>
<tr>
<td>Method</td>
<td>IRLS</td>
<td>Log-Likelihood</td>
</tr>
<tr>
<td>Date</td>
<td>Tue, 23 May 2023</td>
<td>Deviance</td>
</tr>
<tr>
<td>Time</td>
<td>22:59:37</td>
<td>Pearson chi2</td>
</tr>
<tr>
<td>No. Iterations</td>
<td>100</td>
<td>Pseudo R-squ. (CS)</td>
</tr>
<tr>
<td>Covariance Type</td>
<td>nonrobust</td>
<td>0.1019</td>
</tr>
</tbody>
</table>

Table A4: Summary Statistics for Digitalized Fault Reports. Note: This table describes summary statistics from the negative binomial model for digitalized fault reports.

<table>
<thead>
<tr>
<th>Dep. Variable</th>
<th>NumberOfServiceOrders</th>
<th>No. Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>GLM</td>
<td>Df Residuals</td>
</tr>
<tr>
<td>Model Family</td>
<td>NegativeBinomial</td>
<td>Df Model</td>
</tr>
<tr>
<td>Link Function</td>
<td>Log</td>
<td>Scale</td>
</tr>
<tr>
<td>Method</td>
<td>IRLS</td>
<td>Log-Likelihood</td>
</tr>
<tr>
<td>Date</td>
<td>Sun, 21 May 2023</td>
<td>Deviance</td>
</tr>
<tr>
<td>Time</td>
<td>11:00:05</td>
<td>Pearson chi2</td>
</tr>
<tr>
<td>No. Iterations</td>
<td>100</td>
<td>Pseudo R-squ. (CS)</td>
</tr>
<tr>
<td>Covariance Type</td>
<td>nonrobust</td>
<td>0.02599</td>
</tr>
</tbody>
</table>

Table A5: Summary Statistics for Non-digitalized Fault Reports. Note: This table describes summary statistics from the negative binomial model for non-digitalized fault reports.

<table>
<thead>
<tr>
<th>Dep. Variable</th>
<th>NumberOfServiceOrders</th>
<th>No. Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>GLM</td>
<td>Df Residuals</td>
</tr>
<tr>
<td>Model Family</td>
<td>NegativeBinomial</td>
<td>Df Model</td>
</tr>
<tr>
<td>Link Function</td>
<td>Log</td>
<td>Scale</td>
</tr>
<tr>
<td>Method</td>
<td>IRLS</td>
<td>Log-Likelihood</td>
</tr>
<tr>
<td>Date</td>
<td>Sun, 21 May 2023</td>
<td>Deviance</td>
</tr>
<tr>
<td>Time</td>
<td>21:06:59</td>
<td>Pearson chi2</td>
</tr>
<tr>
<td>No. Iterations</td>
<td>100</td>
<td>Pseudo R-squ. (CS)</td>
</tr>
<tr>
<td>Covariance Type</td>
<td>nonrobust</td>
<td>0.09570</td>
</tr>
</tbody>
</table>