LUT University

School of Business and Management

Master’s Degree Program in Strategic Finance and Business Analytics

Antti Nikkinen

Analyzing the Finnish Property Maintenance Business Through Data

Master’s thesis

Examiners : Professor Mikael Collan
Post-Doctoral Researcher Jan Stoklasa

ABSTRACT

Author :	Antti Nikkinen
Title :	Analyzing the Finnish Property Maintenance Business Through Data
Facility :	School of Business and Management
Degree :	Master of Science in Economics and Business Administration
Master's Program :	Strategic Finance and Business Analytics
Year :	2020
Master's Thesis :	LUT University, 89 pages, 28 figures, 12 tables, 4 equations, 2 appendices
Examiners :	Professor Mikael Collan Post-Doctoral Researcher Jan Stoklasa
Key words :	profitability analysis, self-organizing map, principal component analysis, Finnish property maintenance business

The profitability of the Finnish property maintenance business is under-researched by academics. This study focuses on defining the current status of the industry and the most important financial ratios. Study also introduces incentives to improve the business and the profitability of the case Company X. The theoretical part consists of three sections. The first one defines the structure of the market and the second one presents financial ratios used in the field. The last section introduces and discusses the previous literature. The empirical part consists data analytics with the Principal Component Analysis and the Self-Organizing Map. The data consists of financial statements from the years between 2014 and 2018. The main findings are the following. Solvency and performance ratios are the most important ones on the field according to the Principal Component Analysis. The Self-Organizing Map clustered the dataset and revealed characteristics of the companies in terms of profitability. The highest profitability ratios have relationship with strong capital structure and high values on short-term solvency. The field of the property maintenance in Finland is characterized by use of debt.

TIIVISTELMÄ

Tekijä :	Antti Nikkinen
Otsikko :	Suomalaisen kiinteistöhoito-toimialan analysointi datan avulla
Tiedekunta :	School of Business and Management
Tutkinto :	Kauppatieteiden maisteri (KTM)
Maisteriohjelma :	Strategic Finance and Business Analytics
Vuosi :	2020
Pro Gradu -tutkielma :	LUT-yliopisto, 89 sivua, 28 kuvaa, 12 taulukkoa, 4 kaavaa, 2 liitettä
Tarkastajat :	Professori Mikael Collan Tutkijatohtori Jan Stoklasa
Avainsanat :	kannattavuusanalyysi, itseorganisoituva kartta, pääkomponenttianalyysi, suomalainen kiinteistönhoito-toimiala

Suomalaisen kiinteistönhoito-toimialan kannattavuutta on tutkittu vain vähän akateemisessa kirjallisuudessa. Tämä tutkimus keskittyy selvittämään toimialan nykytilaa ja sen tärkeimpiä taloudellisia tunnuslukuja. Tutkimus esittelee Yritys X:lle keinoja liiketoiminnan ja kannattavuuden parantamiseksi. Työn teoreettinen osuus koostuu kolmesta osiosta. Ensimmäinen osa määrittelee markkinan rakenteen, toinen osa esittelee toimialalla käytettävät taloudelliset tunnusluvut ja kolmas osa esittelee aiemman kirjallisuuden. Työn empiirinen osa koostuu data-analyysistä, jossa hyödynnetään pääkomponenttianalyysiä ja itse-ohjautuvaa karttaa. Käytettävä data sisälsi tilinpäätösinformaatiota vuosien 2014 ja 2018 väliseltä ajalta. Tutkimuksen johtopäätökset ovat seuraavanlaiset. Tärkeimpiä taloudellisia tunnuslukuja ovat pääkomponenttianalyysin mukaan maksukyvyn ja suorituskyvyn tunnusluvut. Itse-ohjautuva kartta paljasti toimialan kannattavien yritysten ominaispiirteet. Toimialan korkein kannattavuus on suhteessa vahvaan taserakenteeseen ja korkeaan lyhytaikaiseen maksukykyyn. Velka on toimialalle tyypillistä.

ACKNOWLEDGEMENTS

This master's thesis is the final assignment in completing my master's degree in Strategic Finance and Business Analytics at LUT University. The thesis demanded a lot of time and effort. The process was indeed like a ride in a rollercoaster with ups and downs, but also educational and very intriguing. I'm very happy that I'm graduating in June 2020 and I want to thank LUT University for the opportunity to study and develop myself. My studies here have been the most exciting and challenging time of my life so far.

At first, I would like to thank Company X for this opportunity to write my master's thesis on this project.

I'm very grateful to Professor Mikael Collan for being thesis advisor and giving support on this project.

I want to thank Suomen Asiakastieto Oy and Statistics Finland for providing data for my thesis. I also want to thank Pia Gramén from Kiinteistötyönantajat Ry for providing background information for this study.

Special thanks to Juha for commenting my texts.

Last but not least, I want to thank Jonna for supporting and encouraging me on my studies.

Mikkeli 20th of May 2020

Antti Nikkinen

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1 INTRODUCTION

This master's thesis begins from the introduction chapter and it consists from four parts. Focus of this introduction is to present the topic and incentive for the study. The first part introduces background of the study. The second part defines the focus of the research. The third part presents objectives and research questions. The last part introduces the structure of this master's thesis.

1.1 Background of the study

"Profitability that generates value comes from a firm's business operations."
(Penman 2010)

Rakennusinsinöörien Liitto (2019, 4-5) states that real estate and construction industries count for 15 % of Finland's annual GDP and that buildings comprise 45 % of Finnish national wealth. According to Statistics Finland (2019) building stock consists of 1 530 474 buildings at the end of year 2018. These buildings need continuous maintenance and supervision to serve residents and other users. Nowadays property maintenance business is undergoing a transformation. Internet of Things and digitalization are the future prospects and aging building stock create possibilities for the companies operating in the field of property maintenance.

This thesis focuses on Finnish property maintenance business and especially analyzing it particularly through financial statements. The industry is very interesting due to its role in maintaining the national wealth and the growth prospects it possess. However, the industry is not the most attractive one in the eyes of academia and especially in terms of profitability. The goal of this research is to find the financial ratios most important to the industry and utilize them visually to explain the present state and characteristics of the industry. This research is motivated by a real-life business problem and is conducted as a case study.

The results of this study could be implemented to evaluate the industry and benchmark company performance against the industry. Academia could use this research as basis of further research. Findings are also applicable to be used by managements of companies operating in the industry.

Limited existing research of profitability in the property maintenance business in Finland has caused a remarkable research gap. Regardless of the size of the property maintenance business and the role it assumes in assessing Finnish national wealth, there are only a few research papers available for the subject, most of which are bachelor’s theses. This was a significant incentive to conduct this research.

1.2 Focus of the research

This research focus on an analysis of the property maintenance industry in Finland. The study is quantitative and to addresses the topic in a comprehensive manner. Background information about the field of business and financial ratios is provided as well. Data analysis is in the core of this research and two different analytical methods are used. A key element of this study is to practice with these methods and develop skills further in the field of data-analysis. Figure 1 presents the focus of the thesis.

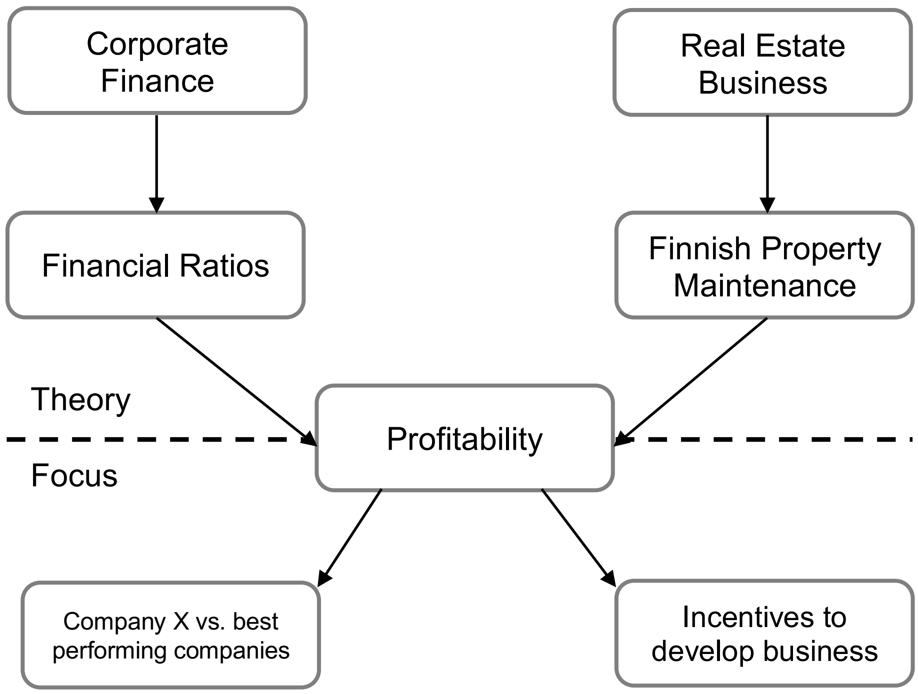

Figure 1 Focus of the research

From a theoretical standpoint this study has two objects. The first object is to examine the Finnish property maintenance business. The second objective is to present financial ratios in the field. These two objects connect to each other from a point of view of profitability assessment. After this study concentrates on the evaluation of Company X against the industry and defining incentives to develop the business to enhance profitability.

1.3 Objectives and research questions

This research has three primary objectives. First, this research studies the importance of financial ratios in the field of property maintenance. Suggested method is the Principal Component Analysis. Second, aim is to create an overall picture of the current status in property maintenance industry in Finland through financial ratios. This part is conducted through Self-Organizing Maps with the most important variables defined with the Principal Component Analysis. Third, case study will be implied on financial ratios between case company and best performing companies and search incentives to develop business and profitability. This last part is conducted with analysis of financial statements and through comparison with the best performing group.

Based on the objectives, the following research questions and sub-questions were formed:

Question 1
Which financial ratios are the most important in the field of property maintenance in Finland?

Question 2
The goal is to study the current status of companies in the field of property maintenance in Finland.

Question 3 and its sub-questions
How does the case company compare to the most profitable ones in the field?
In what kind of cluster does the case company belong to?
What actions can be taken to develop the profitability and the business?

1.4 Structure of the research

The structure of the study can be divided into three main sections. The first section presents basic concepts, theory and methodology to the reader. The second section is the empirical part of the study. The third section presents findings and incentives to develop business to enhance profitability. The more precise structure of the study is as follows and it is also highlighted with Figure 2.

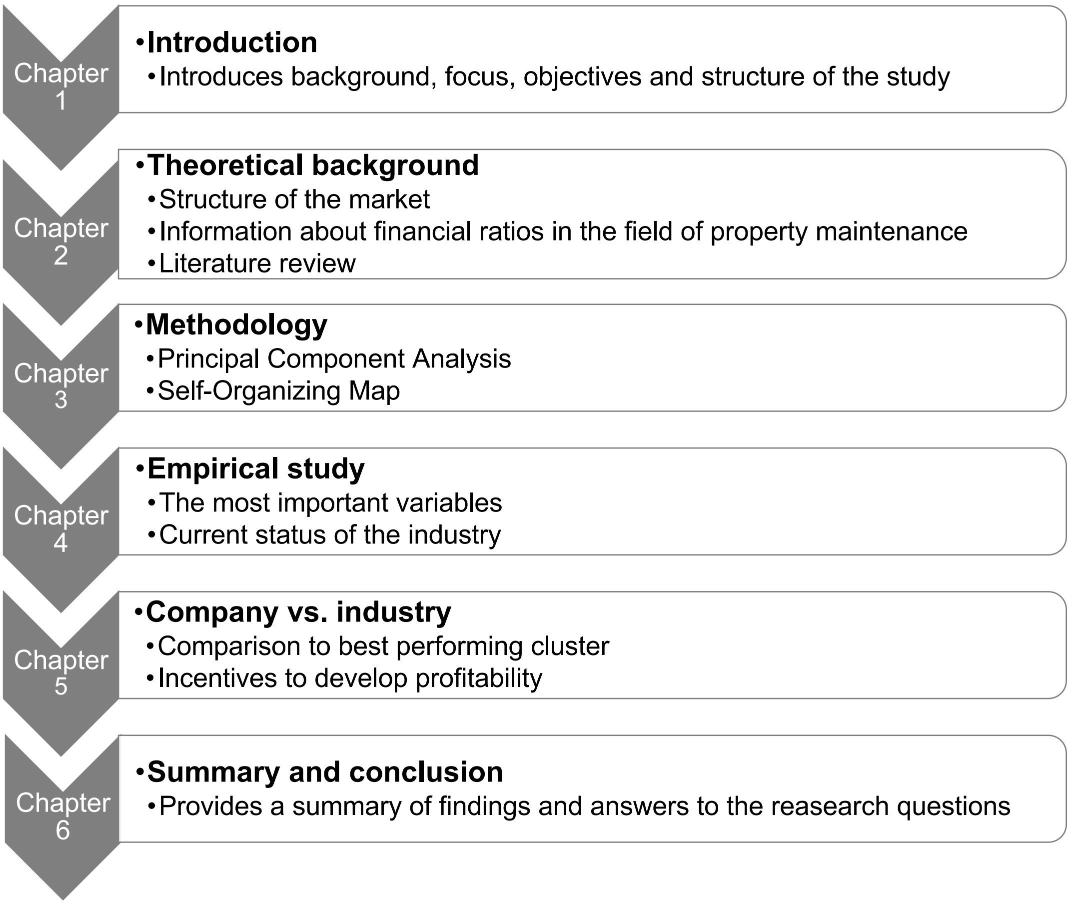

Figure 2 Structure of the research

The first chapter introduces the study and defines its focus, objectives and structure of the study. The second chapter provides theoretical background for the research and it presents the structure of the property maintenance field in Finland and also introduces financial ratios used in the field. This chapter also presents a literature review. The third chapter defines methodology and discusses the methods used in this study. The fourth chapter is the empirical part of the study. It presents the Principal Component Analysis and Self-Organizing Map. Chapter 5 introduces case Company X and discusses how it performs versus the best-performing cluster. The final chapter summarizes and discusses findings of the study.

2 THEORETICAL BACKGROUND

This chapter introduces basic concepts intended to develop knowledge of the research topic. First, an overview of the property maintenance business in Finland is presented. Emphasis is on the structure of the market. Second, financial ratios relevant to property maintenance are introduced focusing on the field's most fundamental ones. Finally, previous research and literature are presented and discussed, paying specific attention to describing the process of selecting and compressing the existing literature.

2.1 Finnish property maintenance business

This chapter aims to draw overall picture of the industry in Finland between years 2014 and 2018. Material for the year 2019 was not yet available at the time of writing of this research. Focus of this chapter is on the structure and size of the market and growth of the industry.

According to Kiinteistötyönantajat Ry (2020a) the field of real estate is a part of wider cluster of real estate and construction in which operations with multiplicative effect are estimated to be over 500 billion € per year, employing nearly one in every five people in Finland. Overall, real estate sector employs 115 000 people and figure 3 presents how workforce is divided between industries (Kiinteistötyönantajat Ry 2020a). Most of the workforce is working in facility services.

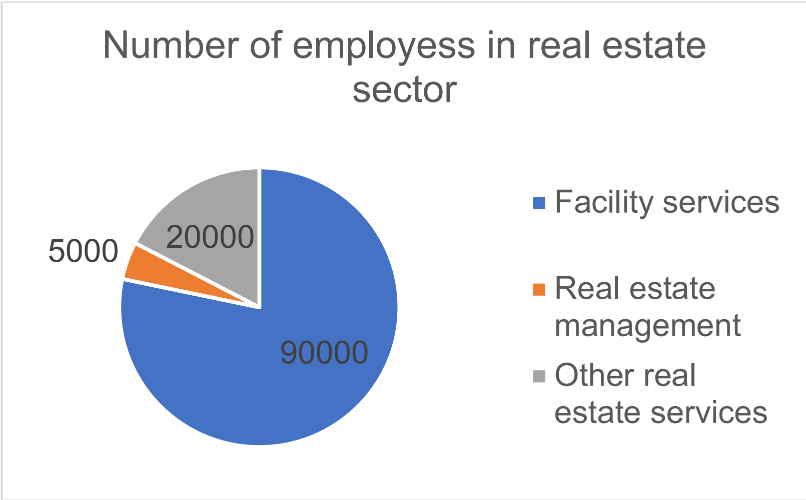

Figure 3 Number of personnel in real estate sector (Kiinteistötyönantajat Ry 2020a)

Field of real estate consists of four different industries: facility services, real estate management, ownership and leasing of real estates, and services for real estate possession and management. Facility services includes property maintenance, technical services, energy management services, cleaning and courtyard upkeep (Kiinteistötyönantajat Ry 2020b). Figure 4 presents the structure of the industry.

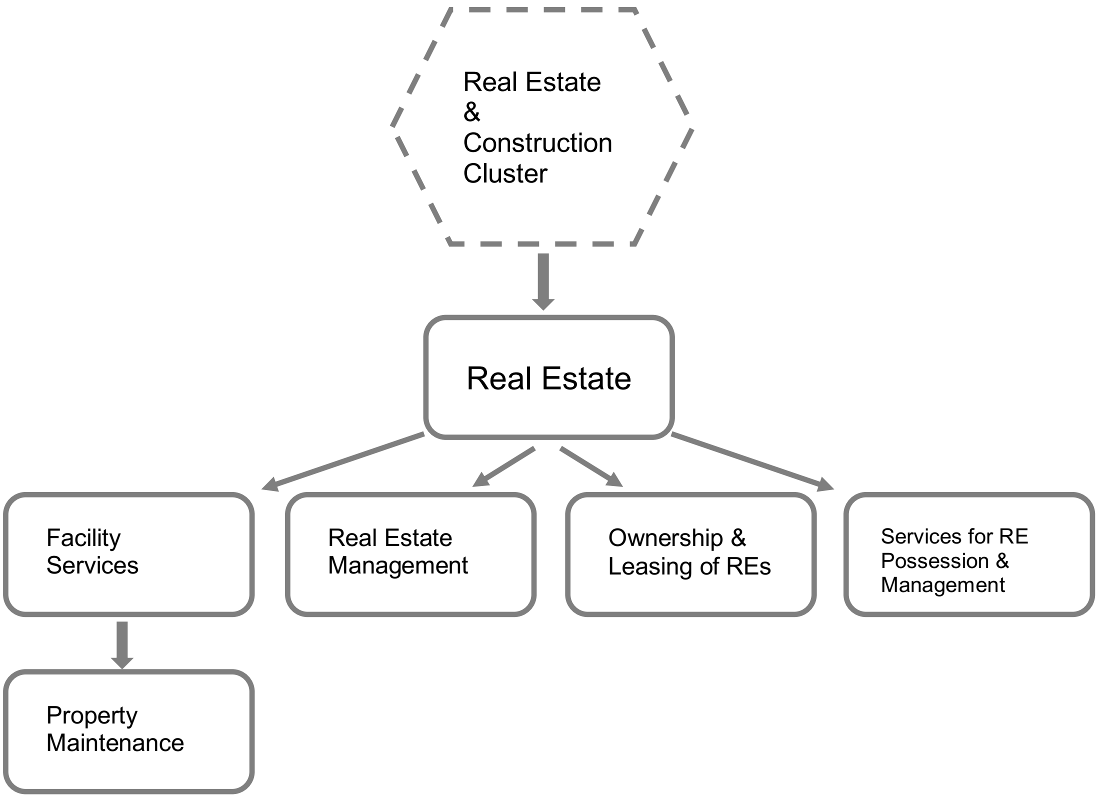

Figure 4 Structure of the market (Kiinteistötyönantajat Ry 2020a & 2020b).

Facility services market in Finland has four primary characteristics: it is considered non-technical business, markets are stable, market have few actors (e.g. service provider and customer) and the business is unregulated. Facility tasks are considered labor-intensive and tasks are easy to measure. (Salonen 2004, 56). Salonen (2004, 52) describes that "facility services is a very developed business. Even though the facilities and related systems are becoming more complex, the services are still relatively simple." According to Rakli Ry (2012, 54-57) facility services can be divided into real estate maintenance and facility and user services. The former is defined as services that aim to maintain a desired level in real estate with aspects of value, qualities and state. The latter, facility and user services, refer to services that aim to create necessary conditions for users of real estate. Lith (2019, 35) adds that real estate maintenance consists of proactive procedures which aim to reduce the need for property repairs.

Finland's official business classifications classify property maintenance under the main class of administrative and support service activities, which is located under services to buildings and landscape activities. These services consist property maintenance, cleaning activities and landscape service activities. (Statistics Finland 2020a).

Property maintenance industry produces services to maintain structures of the buildings, to maintain technical systems and to upkeep properties and clean indoors and outdoor grounds aiming to keep properties safe and functioning (Studentum 2018). Lith (2019, 35) adds that maintenance also prevents the occurrance of faults.

Structure of the market in 2018

In 2018, in terms of personnel, approximately 89 % of the companies were classified as micro companies (employing less than 10 people). 9 % of the companies were considered small companies (employing between 10 and 49 people) and remaining 2 % were larger companies employing more than 50 people. Figure 5 illustrates how the property maintenance companies are divided according to number of people employed.

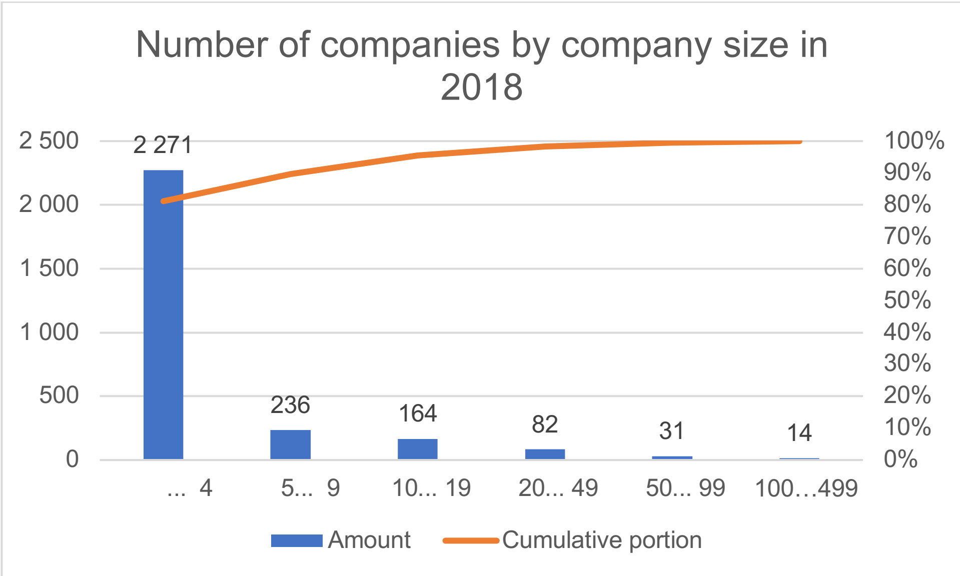

Figure 5 Size of the companies in the property maintenance (Statistics Finland 2020b)

Micro companies and small companies each generate 35 % of total revenues of the market and the remaining 30 % is generated by larger companies. Companies employing less than four employees have the largest share of total revenues totaling 240 million € at the end of 2018. Figure 6 illustrates the market share by company size.

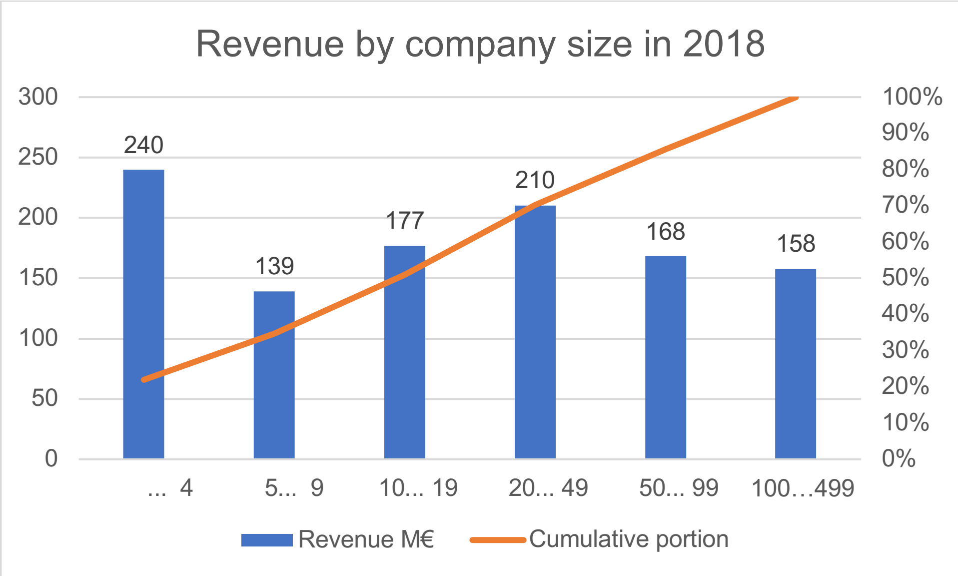

Figure 6 Revenue by company size (Statistics Finland 2020b)

Amount of personnel is distributed relatively evenly across different classes. Micro companies employ 31 % of the total workforce in the industry while the share for small and larger companies is approximately 34% each. Companies employing between 20 to 49 employees have the largest share of total workforce. Figure 7 illustrates the share of workforce in by company size.

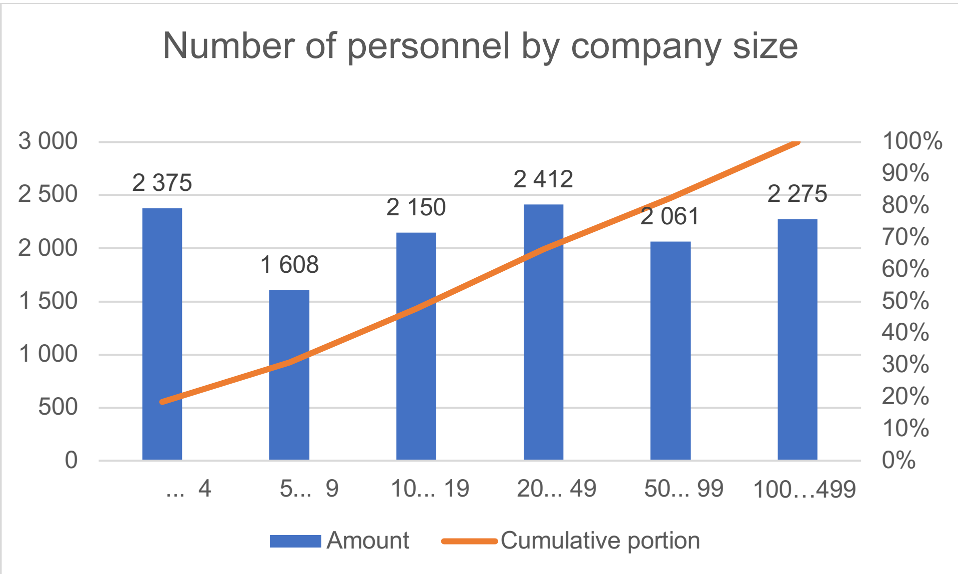

Figure 7 Workforce by company size (Statistics Finland 2020b)

Development in recent years

Size of the property maintenance market in Finland has grown with CAGR of 5,68 % through years 2014 & 2018, reaching 1,09 billion € in 2018 (Statistics Finland 2020b). There is clear positive development in revenue across all categories of company size. The largest relative development (115,5 %) in revenue is in companies employing between 50 and 499 employees. Correspondingly, the smallest relative development (7,6 %) is in companies employing between less than 10 employees. Small companies increased revenue by 19,1 %. Figure 8 illustrates the recent development of the market by company size.

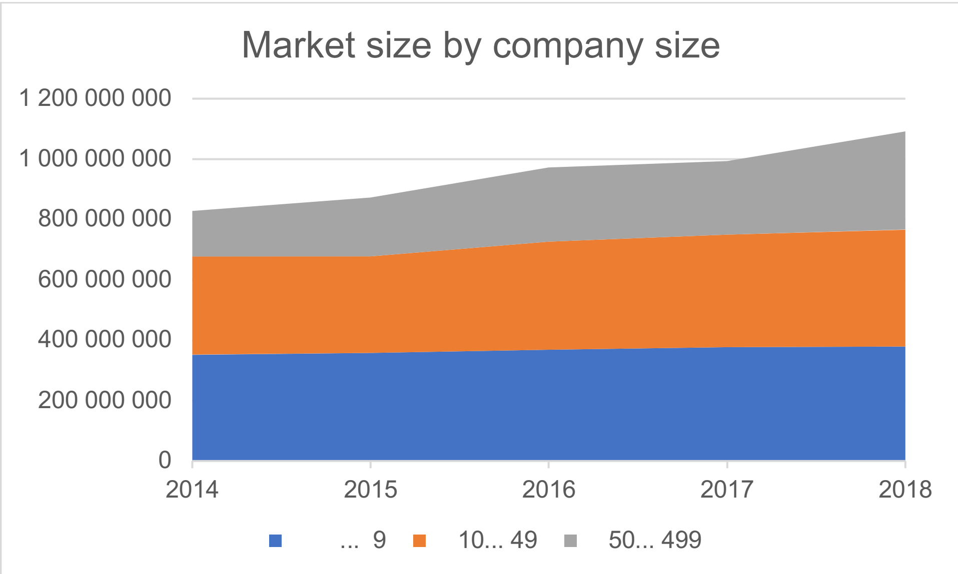

Figure 8 Development of the market size in property maintenance by company size (Statistics Finland 2020b)

Number of the companies in the field of property maintenance has decreased by 1,4 % between years 2014 and 2018, and the total amount of the companies was 2798 at the end of year 2018 (Statistics Finland 2020b). During this time period, the number of micro companies has decreased by 2,3 % while the number of small companies has increased by 0,8 %. The largest relative growth has occurred among companies employing more than 50 people, their amount has grown 66,7 % but as the growth in absolute terms is only 18 companies, the percentage growth seems slightly misleading. Figure 9 shows development of the number of the companies operating in the industry in years 2014-2018.

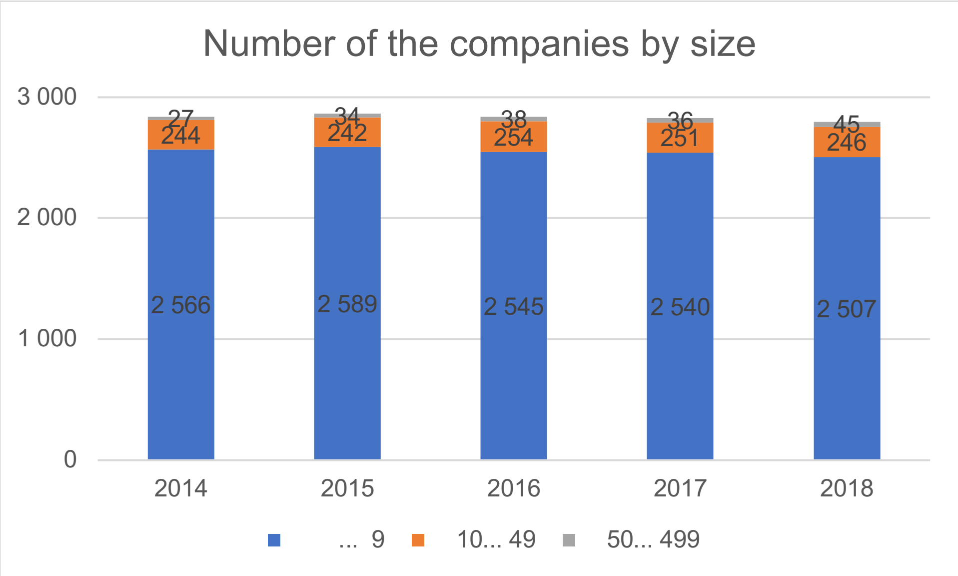

Figure 9 Number of the companies by size (Statistics Finland 2020b)

A growing trend can be seen in the total amount of the personnel. In property maintenance, the total amount of personnel has grown 9,6 % between years 2014 and 2018, by the end of 2018 the number of people employed by property maintenance companies reached 12 881. According to an analysis 11,2 % of the workforce in real estate sector is working in property maintenance. Number of employees in micro companies has decreased by 12,2 %. The same time period has seen the total amount of personnel in small companies decrease 2,4 %. The most significant relative growth in terms of overall workforce has occurred in companies employing more than 50 employees; the workforce has grown by 70,4 % between 2014 and 2018. Figure 10 presents the development of the workforce in property maintenance.

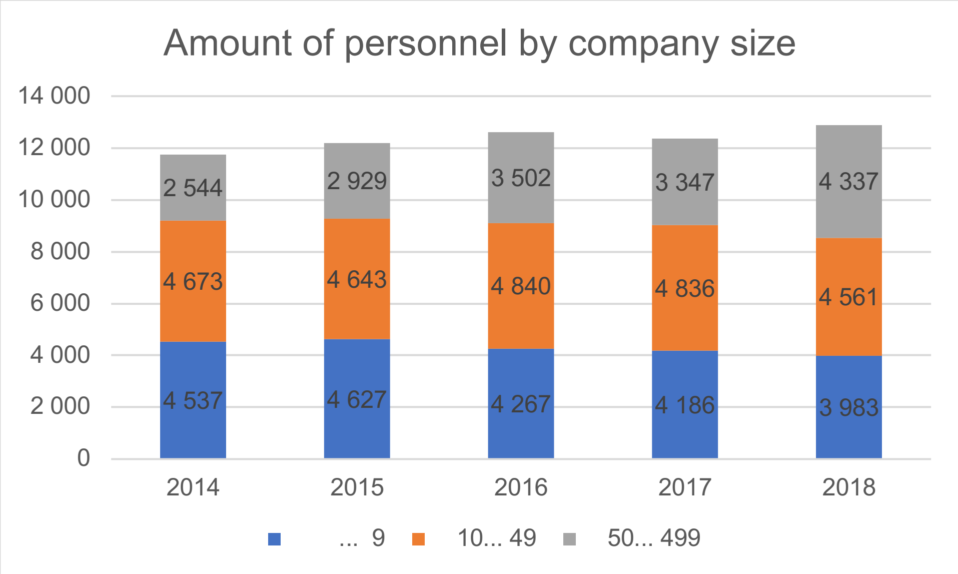

Figure 10 Total amount of workforce (Statistics Finland 2020b)

Growth can also be spotted when examining the development of revenue by personnel, which is illustrated on figure 11. On average the revenue has grown 20,3 % between years 2014 and 2018. Largest revenue growth, over 26 %, occurred in companies employing more than 50 people. Micro companies have been more productive on per-person basis than small companies and larger companies over the same time period.

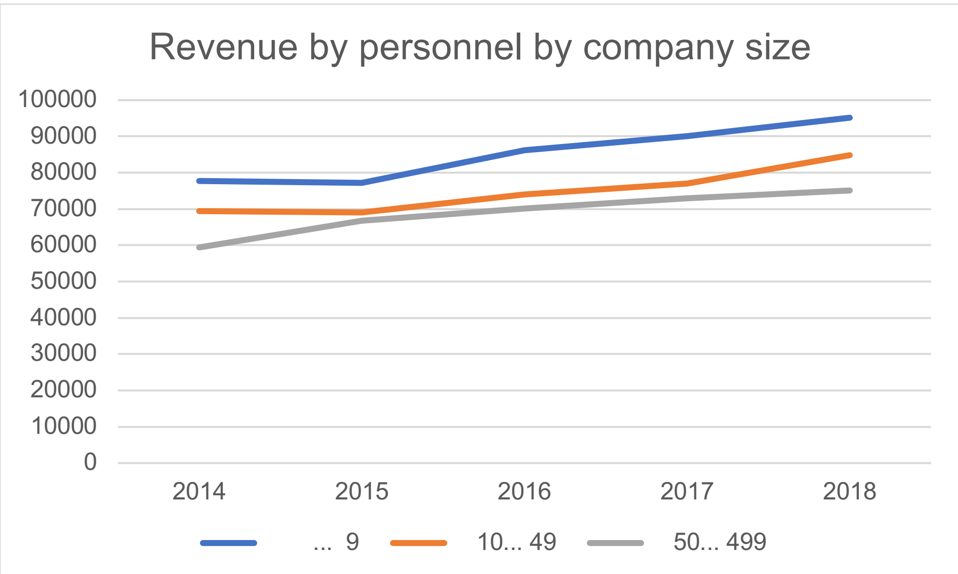

Figure 11 Revenue by personnel by company size (Statistics Finland 2020b)

Rakli ry (2014) states that jobs in real estate sector are in relationship with Finland's building stock and as well as their development and maintenance of them. Therefore, employment in the sector is relatively stable because the need for maintenance and development is constant and not heavily affected by business cycles. The analysis and statistics presented in this section suggest significant growth in workforce development amplifying the labor-intensity of the business.

According to Kauppalehti (2020) the Finnish facility service business has thousands of small companies but that the services are concentrated in the hands of big companies, and this can be verified from the presented statistics. In 2018 the largest 45 companies representing 1,6 % of the industry's total number of companies amassed 30 % of total revenue and 34 % of the total workforce in the field. Furthermore, Lith (2019, 116) states that it is characteristic to the facility service sector that big companies grow through mergers and acquisitions. The analysis presented in this section echoes this statement and underlines the likelihood of future developments being characterized by industry consolidation leading to further growth in the number of large companies similarly to the time between 2014 and 2018.

Elinkeinoelämän Keskusliitto (2020) states in its economic survey that in the last quarter of the 2019 businesses in facility services suffered from four obstacles to achieve growth:

- not enough capable workforce available
- insufficient demand
- problems with financing
- other obstacle

Overall the industry is growing in terms of revenue and number of personnel, but the amount of the companies is slightly decreasing in the field of property maintenance. This analysis indicates that the market is estimated to continue the growth at a similar pace in coming years, largely because the building stock is aging and purposeful maintenance is required. Furthermore, new housing companies and industry buildings require maintenance across all phases of the facility's lifecycle. These two factors drive the growth of the industry.

2.2 Financial ratios in property maintenance

This subchapter presents financial ratios essential to property maintenance and develops understanding of the topic. First, basics of the use of financial ratios are presented and after which ratios are briefly defined and explained. Valuation and market value indicators are excluded from this research. Focus of the section is to define essential financial ratios in property maintenance through characteristics of the business.

Drury & Tayles (2006, 406) state that profitability analysis is "one of the most important management accounting practices". Throughout history financial ratios have been a target of great interest by researchers and academic world and there are several aspects how to approach the topic:

- assessing the financial health of the company or industry
- estimating and predicting future (e.g. bankruptcy)
- assessing creditworthiness and creditrisk
- valuation of the company

Nowadays measurements of company performance are useful tools in managerial decision making. Financial position of the company is discoverable in annual financial statements, and the information in them is central to financial analysis. Figure 12 lists internal and external applications of financial statement analysis.

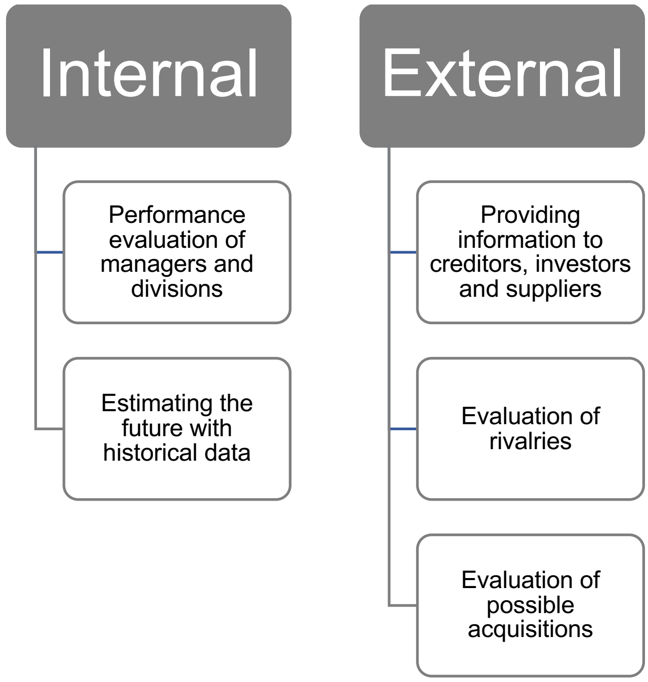

Figure 12 Using financial statement information (Ross et al. 2018, 71)

There are several key ratios to analyze company performance from different aspects. Key ratios are useful in company or industry comparison. Ross et al. (2018, 57) state that financial ratios are typically divided to five groups:

- Liquidity ratios for short-term solvency
- Financial leverage ratios for long-term solvency
- Asset ratios
- Profitability ratios
- Market value ratios

Rist & Pizzica (2015, 1, 3) add performance ratios (also known as activity ratios) to the groups of ratios and continue that some ratios cannot necessarily be allocated to any of groups mentioned above.

As stated in the previous chapter the facility services as an industry suffers from three different problems that curtail the growth opportunities. Based on personal work experience in the banking industry and financing small-and-medium sized companies, problems with financing often arise from an unprofitable business, bad capital structure or difficulties to respond to bank's demand for collaterals.

Therefore, short- and long-term solvency ratios and profitability ratios are selected for further examination. Property maintenance is service business where projects are invoiced primarily post-completion, which is why the collection period of trade receivables is a key metric. Another characteristic feature of property is machinery and their usage efficiency is central in conducting successful business. Thus, asset turnover ratio is also chosen as a metric for this research. Lastly, facility services is a very labor-intensive business, so it is logical to measure personnel performance. Next, the selected ratios are described further. Formulas of the financial ratios are presented in Appendix 1.

Liquidity ratios

Following presents the most common short-term liquidity ratios. Many companies have long-term debt, which maturity is longer than one year and shot-term liabilities, for which maturities are less than one year. Liquidity ratios concern short-term solvency and liquidity. Corporate Analysis reg. assoc. (2013, 81) states that "liquidity position can, at the same time, be both a dynamic and a static concept. A dynamic liquidity measures the amount of internally generated cash in meeting the payment obligations. A static liquidity, on the other hand, measures how the quickly realizable assets of the disposal of the company at any particular time, could be used in servicing the obligations arising from short-term liabilities."

Berk et al. (2015, 70) present current ratio, quick ratio and cash ratio as short-term liquidity ratios. These ratios are useful in assessing liquidity and "whether the firm has sufficient working capital to meets its short-term assets". Current ratio is the least stringent one because it takes the ratio of current assets including inventory to current liabilities. Cash ratio is the strictest measure because it only considers the ratio of cash to current liabilities. Current ratio is excluded from this research because companies in the field rarely have significant inventories.

Financial leverage

Berk et al. (2015, 72) define financial leverage as a financial position which indicates how much the company has debt as a source of financing. Penman (2010, 702) says that regarding long-term solvency ratios, the process “moves to incorporate the noncurrent sections of the balance sheet in ratios.” Brealey et al. (2011, 732) argue that debt creates financial leverage because it increases returns in favourable times and reduces returns in unfavourable times.

Corporate Analysis reg. assoc. (2013, 75-76) defines equity ratio, relative indebtedness and net gearing as indicators of capital structure. Berk et al. (2015, 73), Penman (2010, 702) and Brealey et al. (2011, 733) define debt-to-total assets ratio (also known as total debt ratio and debt ratio).

Efficiency

Brealy et al. (2011, 729) define efficiency as a part of overall company profitability. Ratios present how effectively the business is using its assets.

Brealey et al. (2011, 729) and Berk et al. (2015, 70) define asset turnover ratio as a measure of efficiency. Brealey et al. (2011, 730) state receivables turnover as an indicator of performance.

Profitability ratios

Common to accounting and financial literature is the thought of profitability as a measure of companies’ efficiency and the basis of their actions. Profitability is a key requirement for a company to maintain operations, and it is possible to measure it with absolute or relative values. Financial literature mainly computes ratios in two ways: the first one utilizes businesses’ capital usage, resulting in relative profitability measures, and the second one describes absolute differences between sales and expenses, thus producing absolute profitability measures.

Damodaran (2008, 94-97) defines return on assets (ROA), return on capital (ROC) and return on equity (ROE) as basic indicators of profitability. Corporate Analysis reg. assoc. (2013, 72-73) defines also return on investment (ROI). Corporate Analysis reg. assoc. (2013, 67-69) presents gross margin, operating margin (EBITDA, operating result (EBIT), net result margin as profitability ratios calculated from the income statement. Berk et al. (2015, 69) introduce gross margin, EBITDA-margin and net result margin as profitability ratios. Net result, gross margin, ROI, ROE and ROC will not be used in this research. In my opinion ROA is the toughest measure of profitability, because it also contains information of historical performance of the company.

Performance ratios

Rist & Pizzica (2015, 3) define performance ratios as indicators of business' capacity to generate revenue and create profit from their assets. These ratios are used to assess companies' relative efficiency to harness their assets.

Corporate Analysis reg. assoc. (2013, 85) defines change in net sales (CINS) as ratio which describes development of net sales and invoicing. They add that net sales per employee can be used to assess productivity.

2.3 Previous research and literature review

This chapter describes the literature selection process used in this study. Emphasis is on describing the process using different search terms on different search portals. After the process is described the next subchapter presents previous literature and related discussion.

2.3.1 Literature selection process

Topic of this research is context-dependent and therefore searches were conducted in Finnish. The first assumption was that the field is very under-researched and there are not many international publications available. Initial plan for literature selection was the following:

1. Search for relevant literature using Finnish search terms
2. Search backwards based on citations of articles found in first phase to recognize the most fundamental ones.
3. Get further information about specific subjects based on authors cited in fundamental studies and with more precise search terms that have come up during step 1 and 2.

Process was started by using LUT Finna-service (the Lappeenrannan-Lahden University of Technology's internal article portal) which provides articles from multiple databases and arranges the search results based on their relevancy. The search term primarily used in the first phase was "Kiinteistönhoito" and "Kannattavuus". This term was chosen to get a comprehensive idea of the topic and to map previous studies and articles related to property maintenance and profitability. This criterion yielded zero results. Search terms in next step were "Kiinteistönhoito" and "Tunnuslukuanalyysi". These were chosen because profitability analysis of property maintenance business is a major topic of this research. However, this search did not result in any hits either. Furthermore, search terms "Kiinteistönhoito" and "Talous" and "Tunnuslukuanalyysi" led to zero results. The final search consisted of key words "Kiinteistönhoito" and "Toimialakohtainen tutkimus", but to no avail. It became apparent that not many relevant research papers and articles are available.

In the next phase Google Scholar was implemented using the same search terms. Results from these searches are presented in figure 13. Screening process was conducted in two parts: first, examination of the abstract and second, if topic was relevant, it led to screening the references. Most of examinated results are bachelor's theses from different universities of applied science, with one master's thesis from a university. The first search in Google Scholar returned two bachelor's theses and the second search added two bachelor's theses to previous search totaling four results. Third search resulted one bachelor’s thesis from the first search and another bachelor’s thesis from the second search. Final search resulted in one master’s thesis and one previously discovered bachelor’s thesis.

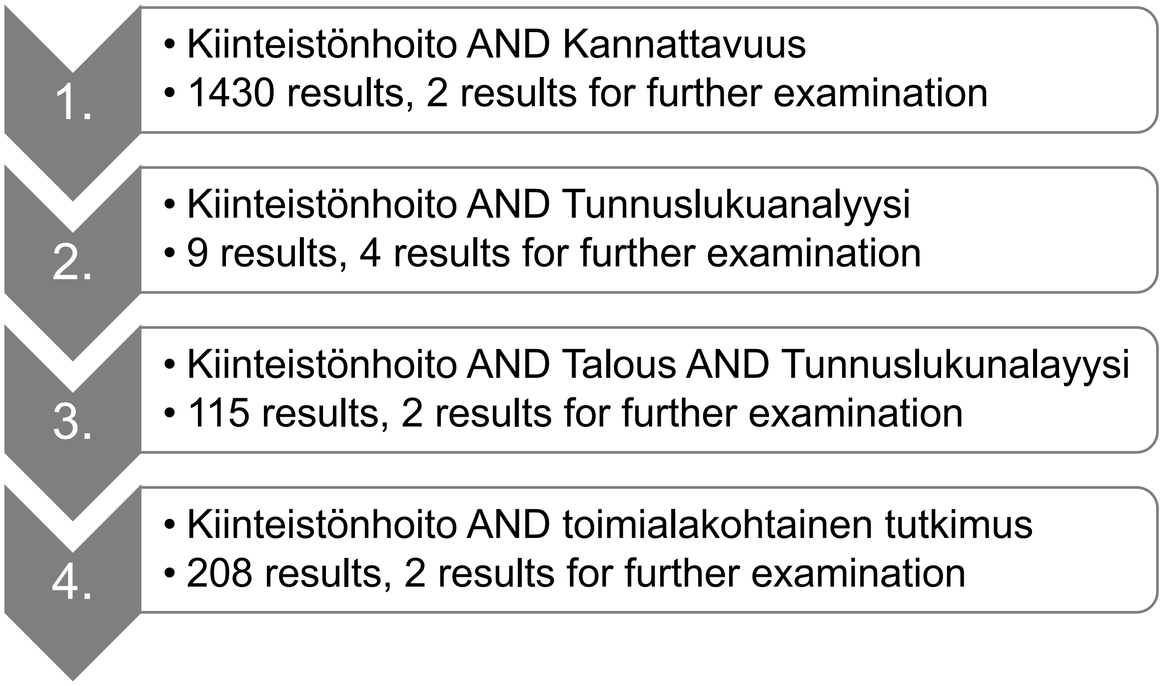

Figure 13 Search process for literature review

The main reason for leaving out articles and research papers at this stage was their irrelevant context. Several of the used search terms yielded results that offered no tangible support for this research. This highlighted the under-researched nature of the topic.

In summary, the search process returned four bachelor’s thesis and one master’s thesis. One key point in screening the references was the discovery of the Pekka Lith’s report Kiinteistöala Suomen kansantaloudessa, which has been cited by several parties such as Kiinteistötyönantajat Ry and companies in the industry, and it can be considered as a foundational research in business policy. It can also be thought of as a fact package of the real estate business. This report is updated annually and the latest version was published in the spring of 2019. Kiinteistötyönantajat Ry’s CEO Pia Gramén was contacted regarding this research and she kindly delivered a copy of the report to be used as a reference.

2.3.2 Previous literature

Leskinen (2009) studied in his bachelor's thesis company profitability in the real estate business. The case company (a limited liability company) operated in three areas: real estate management, property maintenance and apartment leasing. Focus of the thesis is two-fold: financial analysis of the case company and assessment of the profitability of the three operating areas through cost accounting between 2006 and 2008. Return on invested capital (ROI), equity ratio, relative indebtedness and quick ratio were used in financial analysis. These variables were assessed against commonly accepted levels of performance. The case study found the company to be very solvent with an equity ratio of approximately 90 % each year and its liabilities were small. Quick ratio fluctuated greatly during the time period because the total amount of cash was varying in financial statements. All in all, the company was not growing between 2006 and 2008 and its most profitable business was apartment leasing. Real estate management was also profitable segment, but property maintenance was non-profitable in 2006 and 2007. Company was advised to increase business in profitable segments and reduce business in non-profitable operations. Furthermore, pricing of the contracts was untouched between years 2006 and 2008 and there is potential to improve profitability through price checks. Thesis also provided the case company with Microsoft Excel-spreadsheet to monitor the development of financial ratios.

Summanen (2012) also utilized a case study in her bachelor's thesis research on profitability in property maintenance. Case company was a small limited partnership firm and it operated only in field of property maintenance. The principle focus of the thesis was to provide an insight into the financial situation of the company and present incentives to improve profitability. Thesis focused on fiscal years 2010 and 2011 and it used EBITDA-margin, EBIT-margin, return on assets (ROA), return on equity (ROE), equity ratio, relative indebtedness and quick ratio as financial ratios. Profitability was also assessed through cost accounting. Company was assessed to be very solvent with an equity ratio of over 50 % but it had negative EBIT-margins in both years. ROA was negative in fiscal year 2011 and positive in fiscal year 2010.

Quick ratio was assessed to be on excellent levels in both fiscal years. These financial ratios were compared to industry averages in fiscal year 2010 with the industry sample constructed from companies that operate in property maintenance and employ less than 5 people. Analysis showed that the case company performed weakly against industry averages, and that EBITDA-margin was the only metric by which the case company managed to outperform the industry average. The thesis also assessed profitability through cost accounting methods and suggested that improving sales would be the best way to improve profitability. Finally, to attract new customers the study also proposed improvements to the company's marketing practices.

In her bachelor's thesis, Suhonen (2012) studied the financial ratios of SME's with the goal to create a Microsoft Excel-spreadsheet to be used as automated calculation tool of financial ratios on monthly basis. Case company was a limited liability company operating in property maintenance. Analysis was conducted through fiscal years 2010 and 2011. Profitability was assessed through EBIT-margin, ROI, ROE and ROA. Liquidity was assessed using quick ratio, current ratio and working capital to sales ratio. Solvency was measured with equity ratio, relative indebtedness and net gearing. The study found that the case company performed excellently in terms of profitability and in solvency, but liquidity was in lower levels. Conclusion of the thesis emphasized the supportive role of financial ratios in decision-making, and it highlighted the shortness of time between observations. Furthermore, the conclusion discussed historical data of the case company but also data of competitors. Suhonen (2012) stated that comprehensive financial planning calls for the ability to interpret financial ratios and to use this information for the company's benefit.

Kumpula (2015) studied in her bachelor's thesis company's profitability and conducted a trend analysis on financial statements. The case company was a limited liability SME that operated in the fields of property maintenance as well as earthworks. Earthworks is the company's main operating business while its secondary focus is on property maintenance. The study used different solvency-, liquidity- and profitability-based financial ratios. Time period extended from 2011 through 2013. The most essential financial ratios included equity ratio, EBITDA, current ratio and ROI. Case company performance was evaluated against industry averages. Profitability and solvency were determined to be on a good level, but the analysis indicated a possibility to speed up collection of trade receivables. The company's goal was to reach a position in which trade receivables turn around faster than accounts payable. Kumpula discovered that it was difficult to select peer industry for the assessment because companies operate in industries which they consider secondary.

Makkonen (2019) studied in her master's thesis digitalization in facility services. She states that digitalization in the industry centers around enhancing cost efficiencies. A driver for the digitalization is an intra-industry competitive situation between companies. Currently, digital solutions are almost exclusively on improving company's internal effiency, developing the business and also as an answer to customer's expectations. As a conclusion Makkonen states that digitalization in facility services is progressing and it may lead to new data-based businesses and improved internal efficiencies of the companies.

Lith (2019) studied the real estate business in Finland's national economy focusing on facility services. The report pursues to define the state of industry and the economic cycle. This report is updated annually and it relies on official statistics from Statistics Finland. Report consists five parts: defining the industry and the market, market of facility services, business and company structure in facility services, financial situation of the companies in the facility services landscape and facility services in public entities. The most interesting part of the report concerned the business and company structure of the industry as well as the companies' financial situation. Financial situation of the industry was assessed using change in net sales, EBITDA-margin, EBIT-margin, gross margin, sales per employee and equity ratio. Lith (2019, 111) estimated that total costs in 2018 in property maintenance services were 9,51 billion €. Lith (2019, 114-115) also noticed that there was no change in number of real estate maintenance operating branches between years 2013 and 2017. Additionally, the change in net sales in the industry was 16,5 % between years 2013 and 2017.

2.3.3 Discussion of literature and previous studies

This chapter summarizes previous studies. It also connects also connect previous studies to this research. Table 1 lists the most relevant findings from previous literature. The first four findings are bachelor's theses after which there is one master's thesis and one report which is updated annually for the use by authorities and companies.

Connection to thesis	***Author(s) and year***	***Title***	***Main findings***
Financial ratios in the field of property maintenance and incentives to develop profitability.	Leskinen, M. (2009)	Kiinteistöalan pk-yrityksen kannattavuus.	ROI, equity ratio, relative indebtedness and quick ratio were used to assess the profitability. Improvements for the business: increase profitable business, reduce nonprofitable business, check the pricing of the contracts.
	Summanen, K. (2012)	Kiinteistönhoitoyrityksen kannattavuus.	EBITDA-margin, EBIT-margin, ROA, ROE, equity ratio, relative indebtedness and quick ratio were used as financial ratios. Improvements for the business: increase sales.
	Suhonen, L. (2012).	PK-yrityksen tunnuslukuanalyysi ja talouden seuranta.	EBIT-margin, ROI, ROE, ROA, quick ratio, current ratio, working capital to sales ratio, equity ratio, relative indebtedness and net gearing were used as financial ratios. Key takeaway: comprehensive financial planning needs vast understanding of the financial ratios.
	Kumpula, P. (2015).	Tilinpäätösanalyysi investointisuunnitelmien tukena kohdeyrityksessä.	Equity ratio, EBITDA, current ratio and ROI were used. Improvements for the business: speed up the turnover of the trade receivables.
Incentives to develop profitability.	Makkonen, M. (2019)	Kiinteistöpalvelujen digitalisaatio Suomessa.	Digitalization is concentrating on enhancing the cost efficiency of the industry.
Finnish property maintenance.	Lith, P. (2019)	Kiinteistöala Suomen kansantaloudessa – raportti kiinteistöalan yritystoiminnasta, markkinoista ja kehityslinjoista 2018-2019	Overall picture of the market and recent developments in Finland.

Table 1 Previous research papers relevant to the research.

In summary, only six pieces of previous literature were found. Arguably the field is very under-researched and there is a little recent previous research available. Main reason for this hurdle is the very context-dependent topic, which narrows the availability of inter-industry research in Finland. Previous literature is quite recent as the oldest thesis is from 2009 and the newest thesis is from 2019.

Financial ratios used in the bachelor's theses were very similar. Most used one was the equity ratio which was utilized in all the theses. ROI, relative indebtedness and quick ratio were used in three theses. EBITDA, EBIT, ROA and ROE were used in two theses. Least utilized ratios were working capital to sales ratio and net gearing. One might say that short- and long-term solvency as well as profitability have been the most studied and targeted areas of interest in previous literature. However, the theses offered little to no comprehensive background or reasoning as to why certain financial ratios were used, and often the ratios were used in a generalized manner to describe performance. There were also comparison of intra-company ratios over a certain time period and also evaluations of company performance against industry averages. Lith's (2019) report assessed industry performance through change in net sales, EBITDA-margin, EBIT-margin, gross margin, sales per employee and equity ratio. All in all, there are no benchmarks available to decide which financial ratios are the most important ones on the field of property maintenance in Finland.

Previous studies presented incentives for business and profitability development. Incentives were logical and can be applied to any industry or in any company. Boosting sales and focusing on profitable business segments while cutting down non-profitable segments are procedures that enables the company to become increasingly profitable. Pricing checks of existing contracts are another common method of enhancing profitability and defining actions towards non-profitable customers. Previous studies also state that comprehensive financial planning requires extensive understanding of financial ratios. Although the facility service business is very labor-intensive the digitalization it undergoes focuses on enhancing cost-efficiency in companies. On a personal note, there is plenty to still gain with digitalization in property maintenance, specifically regarding optimizing processes and automating the manual routines in the office. This adds pressure on further development of enterprise resource planning systems.

Lith's (2019) report provides a comprehensive picture of the real estate business in Finland. The report presents new business aspects and argues that Internet of Things (IoT) and Artificial Intelligence (AI) may have great impact on property maintenance through remote supervision and intelligent buildings. The report also states that uncertainty in the markets is possibly increasing the outsourcing of the facility services to allocate resources on core businesses. Lith's report is a foundational piece of research and acts as central background for further studies in the industry.

The overall situation in the markets is very interesting because there is a lot of untapped potential discoverable through digitalization and process development. Previous studies presented the state of the industry and defined many aspects of the development of company profitability. The extent and limited availability of previous research also indicated that here is need and usage for further profitability and state analysis in the field of property maintenance. Going forward, this research seeks to discover and define the most important financial ratios in the industry.

3 METHODOLOGY

This chapter of the research presents methods deployed in this study. Methodology consists of two parts: first part introduces Principal Component Analysis (PCA) and second part presents Self-Organizing Map (SOM). Abbreviations (PCA and SOM) are used in following chapters. Goal of this chapter is to develop knowledge of the implemented techniques.

Das et al. (2015) state in their paper that PCA and SOM are the most common methods used for reducing data's dimensions. In this research PCA is only used to find out which variables are the most important ones. Extracted Principal

Components will not be used in this study. Most important variables are deployed into SOM for clustering the dataset.

3.1 Principal Component Analysis (PCA)

Leskovec et al. (2020, 437) "view PCA as a data-mining technique." Ivosev et al. (2008) continue that PCA is a comprehensively applied method to reduce high dimensionality into a more practical set of new variables, thus resulting in simplification of data visualization. Brooks (2014, 170) state that PCA is the most common mathematical factor model and it is beneficial in situations where independent variables face multicollinearity, i.e. variables are closely related. Metsämuuronen (2008, 26) adds that factor analysis is one of the oldest multivariate methods, originally introduced by Charles Spearman in the beginning of the 20th century. PCA is a common method in physics, biology, machine learning and statistics.

Brooks (2014, 170) describes that Principal Components (later PC) "are independent linear combinations of the original data where α_{ij} coefficients to be calculated, representing the coefficient on *j*th explanatory variable in the *i*th principal component". Equation 1 presents the formula.

$$p_1 = \alpha_{11}x1 + \alpha_{12}x_2 + \dots \alpha_{1k}x_k$$
$$p_2 = \alpha_{11}x1 + \alpha_{12}x_2 + \dots \alpha_{1k}x_k$$
$$\dots$$
$$p_k = \alpha_{k1}x_1 + \alpha_{k2}x_2 + \dots \alpha_{kk}x_k$$

$$\sum_{j=1}^{k} \alpha_{ij}^2 = 1 \quad \forall \quad i = 1, \dots, k$$

Equation 1 Formula for Principal Component (Brooks 2014, 170)

Coefficients are also called as component loadings and for each component the sum of squares must equal one (see the sigma notation in Equation 1).

Components are built with completely mathematical methods using forced optimization and there are no assumptions regarding distribution, structure or any other properties of the variables. Brooks (2014, 170-171).

PCA tries to find mutual variance from great number of variables and form new interpretable variables leading to a reduction in the number of original variables. PCA in conducted in four steps:

- calculate correlation- or covariancematrix
- estimate PC loadings with matrix created in previous step
- rotate PC loadings
- calculate PC Scores

Last step of calculation of PC Scores is optional and left out of this research because the only interest is on the importance of single variables. There are limitations and assumptions of PCA: variables should have correlation between them, sample size must be adequate (e.g. 300) and outliers must be removed or fixed. Results are more convincing if variables are normalized, but it is not necessary. Multicollinearity is not a problem in PCA. Metsämuuronen (2008, 25-29).

There are two tests that are useful to conduct before moving on to PCA. Bartlett's Test of Sphericity examines are values zero in the correlation- or covariancematrix and whether the matrix is appropriate for conducting the PCA. When dealing with large samples the test tends to easily produce results indicating that correlations differ from zero. Kaiser-Meyer-Olkin's Measure of Sampling Adequacy (abbreviation KMO, also known as MSA) calculates the relationship between correlation and partial correlation. (Metsämuuronen 2008, 32). Sarstedt & Mooi (2019, 265) continue by saying that the KMO-test "indicates whether the other variables in the dataset can explain the correlations between variables".

Goodness of PC can be assessed with eigenvalues. The rule of thumb is that values above 1 are considered acceptable. Goodness of variable can be assessed using communality. It measures the percentage of a variable's variance that can be explained with PC. The stronger the variable's loading on PC, the closer the value of communality gets to value of 1. If a variable's loading is lower than 0,30 on all PCs it can be removed. (Metsämuuronen 2008, 31).

In literature there are two issues regarding PCA. Das et al. (2015) state that PCs consist of linear combinations of variables. Therefore, normal distribution of the variables is not met and this leads to problems with multivariate data analysis techniques. Ivosev et al. (2008) argue that number of variables have an effect on the number of PCs and a large amount of variables can lead to difficulties in interpretation.

3.2 Self-Organizing Map (SOM)

In 1980's Professor D.Sc. Teuvo Kohonen introduced the Self-Organizing Map (also known as Kohonen map). SOM is an unsupervised machine learning technique which aims to identify patterns from the data. Commonly SOM is considered as a type of artificial neural network. It is mostly used in data analysis and pattern recognition problems. It is a widely used tool in different fields of exploratory data analysis and it is also useful in data visualization. Among others, SOM has been implemented in fields of finance, natural sciences and linguistics and it is one of the most popular neural networking techniques in unsupervised machine learning algorithms. Kohonen (2013, 52) states that SOM is applied most in bioinformatics and in the management of textual databases.

Neural networks (later NN) have been inspired by the human brain. NNs mimic the biological structure of the brain, but they are not like brains. Simply put, deep learning refers to NNs. Deep NNs have made a breakthrough in 2010s because there is significant data and computing power available and nowadays mathematical ideas can be implemented practically and efficiently. (Kananen & Puolitaival 2019, 127).

According to Wendler & Gröttrup (2016, 844-845) "NN consists of multiple neurons or units that process and pass information between each other" and "during data processing, a neuron receives weighted signals from some other neurons and transforms the sum of these weighted signals into new information via an activation function". Characteristically the activation function is the sigmoid or the hyberbolic tangent function. Neurons are structured between different layers. There are three different types of layers in NN: input, hidden layer (can be one or multiple) and output layer. Figure 14 presents the basic structure of a neural network. Input layer consists of first neurons that receive unprocessed raw data, transform it using the activation function and send it to the next layer of neurons. Neurons in the output layer finalize the calculation of final score from the received data.

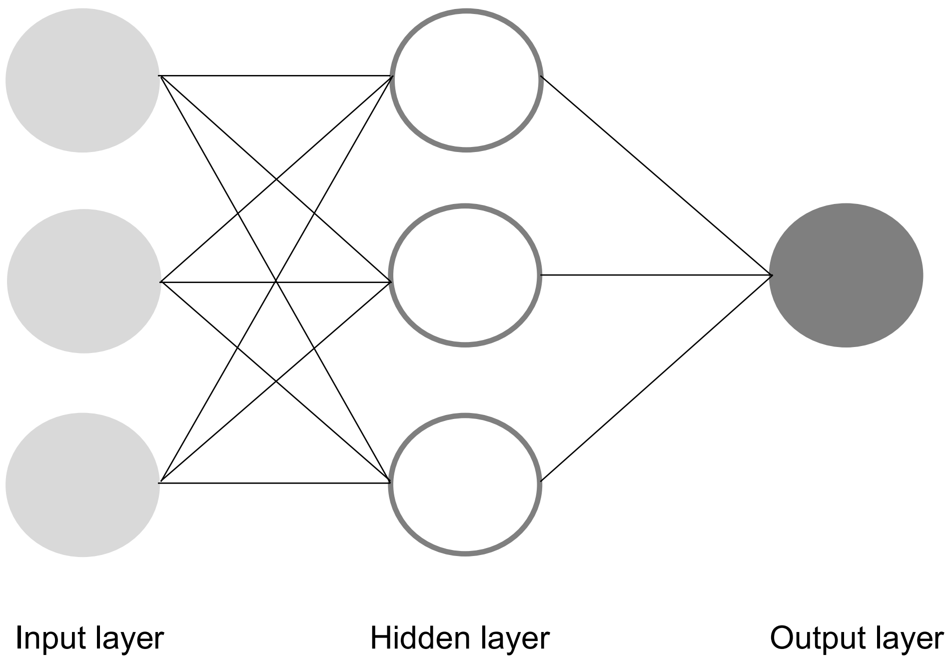

Figure 14 Structure of neural network

"SOM represents distribution of input data items using a finite set of models" and "these models are automatically associated with the nodes of a regular (usually two-dimensional) grid in an orderly fashion such that more similar models become automatically associated with nodes that are adjacent in the grid, whereas less similar models are situated farther away from each other in the grid." These similarities of the models provide insight of topographic relationships of the data. (Kohonen 2013, 52).

SOM models can be divided into two groups depending on computation algorithms. The first type is a recursive and stepwise approximation process in which data is inputted randomly or periodically one by one into the algorithm. This is conducted until a stable state is achieved. The second type of SOM model is a batch-type process in which all the data is simultaneously inputted into the algorithm. After this all computed models are updated in a simultaneous operation. Often this process must be reiterated to create stabilized models. Batch-type process is considered to be faster than the stepwise computation process and it is recommended to be used in practice because it doesn't include any learning parameters. The stepwise process was originally developed only for theoretical reasons and to compare different Self-Organizing models. (Kohonen 2013, 53-54).

First step is to initialize the map. Most commonly the map is a two-dimensional array and have rectangular or hexagonal form. The difference between these two comes from the number of interacting nodes: rectangular has four and hexagonal has six immediate neighbors. Therefore, hexagonal form is considered better for visualization purposes. (Kohonen 1997, 86).

The SOM algorithm has two stages. Kohonen (2014, 21) state that "first the input data item *x(t)* defines or selects the best-matching model (winner)" and "the model at this node as well as its spatial neighbors in the array are modified. The modifications always take place in such a direction that the modified models will match better with the input". First stage uses the formula presented in Equation 2 and the second stage uses the formula presented in Equation 3.

$$c = \text{argmin}_i\{||\mathbf{x}(t) - \mathbf{m}_i(t)||\}$$

Equation 2 Formula for best-matching unit (Kohonen 2014, 21).

In Equation 2 c is the index of best match unit from the equation of $||x(t) - m_i(t)||$ which is the smallest Euclidean distance in this case, x is input data and m_i is the reference vector in node i.

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)]$$

Equation 3 Formula for the modifications in SOM (Kohonen 2014, 21)

In Equation 3 $h_{ci}(t)$ is the so called neighborhood function of the winning best-match unit *c* and *t* is the time coordinate. This research uses the neighborhood function $h_{ci}(t)$ of Gaussian function and can be presented as follows.

$$h_{ci} = \alpha(t) \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right)$$

Equation 4 Formula for Gaussian function (Eklund 2014, 62)

In Equation 4 where $\alpha(t)$ is the learning rate factor and the $\sigma(t)$ is the radius or width of the kernel. Equation 4 and explanation of it is based on Eklund's (2014, 61-62) doctoral dissertation.

Vesanto & Alhoniemi (2000, 599) state that means, medians and ranges of variables are intriguing in cluster analysis but they also say that each cluster can be assessed from different aspects:

- Which variables make the difference between neighboring clusters?
- Which factors make the difference in cluster against the rest of the data?
- What is the effective data dimension of the cluster?
- Does cluster have sub-clusters?
- Is the cluster an outlier or spherical one?
- What are the associations between variables in the clusters?

Data normalization have an essential role with the SOM. Kohonen (2014, 41) state easiest way to normalize the data is to rescale variables so that variance is identical. Normalization with Euclidean distance is applicable to practical studies.

Wendler & Gröttrup (2016, 720, 844) list pros and cons of NNs. They argue that NNs are suitable for large samples and that they can deal with very complex interactions between variables. Furthermore, they are resistant to defective data. Using of NNs is non-parametric which means there are no assumptions regarding distributions. On the other hand, NNs have problems with handling a large number of variables and results that NNs produce can be difficult to understand because of the black box nature of its algorithm. Black box means that an algorithm is hard to interpret because it contains hidden layers. Occasionally NNs are not the optimal solution. Pampalk (2001, 5) discusses that SOM is not applicable in situations where information about existing clusters is available because SOM is an unsupervised technique. Kohonen (2013, 52, 54) presented a new finding that makes it possible to assign inputs between best-matching models more accurately. This is done with a least-squares fitting procedure. Kohonen also states that nowadays there are multiple different versions of SOM.

3.3 Evaluation of methodological choices

This chapter focuses on the evaluation of methodological choices. First the PCA is assessed and after this the SOM is under evaluation. Focus is to define the suitability of used methods in this study.

Defining the most important financial ratios is conducted with the PCA and it is considered to be a good method for data dimensionality reduction. Therefore, it can be seen as a valid method for the purpose. According to Metsämuuronen (2008, 25-26, 28) it is one of the oldest multivariate methods and it is sort of a black box test in which variables are inputs into algorithm and the output is under assessment. Metsämuuronen (2008, 28) continues the PCA is applied in many different materials and types of research when aim is to group a large amount of variables to few groups and reduce phenomen’s scatterness. This verifies the application of the PCA in this master’s thesis and it is suitable for defining the most important variables on the field of property maintenance.

Self-Organizing Map is also a applicable method in data reduction but it is also considered to be very useful in data visualization and unveiling interactions between variables. In this thesis the SOM is used for data visualization and it is also clustered with Ward's method for further analysis. Kohonen (2014, 18) lists main application areas of the SOM where exploratory data analysis and financial applications are among them, which also verify the adequacy of the method in this research. The SOM is used to define the current status of the industry and visualize it and it is considered as a suitable method for this purpose.

4 EMPIRICAL STUDY OF PROPERTY MAINTENANCE INDUSTRY

This chapter presents empirical part of this study. First the used data is presented. This section describes used methods to create a tidy dataset and also presents tools used in this study. Second subchapter presents PCA conducted in this study to define most important variables. These variables will be used and explored in SOM, which is presented in subchapter 3.

4.1 Data

Data used in this thesis consists of two parts. The first part is the dataset of financial statements from Suomen Asiakastieto Oy and the second part was acquired from Statistics Finland. Financial statements were used in PCA and SOM and the data from Statistics Finland was used in the theoretical background. Next, both datasets will be briefly presented.

Financial information used in this thesis was sponsored by Suomen Asiakastieto Oy. Suomen Asiakastieto Oy is a credit report and company information provider in Finland. It serves financial institutions and other companies in their needs regarding risk management, accounting and decision-making. The company is a part of Asiakastieto Group Oyj.

Dataset consists of financial statements from companies operating in the property maintenance industry (sector code 81100) in Finland between the years 2014 and 2018. The total amount of financial statements is 5738 and the dataset consists 173 variables containing rows from the financial statements and also financial ratios. Data was in form of panel data and it was on xlsx-format.

Data from Statistics Finland includes information about the number of companies, total sales, amount of personnel and total amount of payrolls from companies operating in the field of property maintenance. Data was split according to personnel amounts to reflect different company sizes. The observed time period of the dataset was between 2014 and 2018. Data consisted of spreadsheets in xlsx-format.

The following subchapter presents data preparation, transformation and data analysis techniques.

4.1.1 Data preparation, transformation & analysis

The dataset of financial statements consisted of nearly all variables needed for this research. In addition, variables CAR, DR and AT were calculated with Microsoft Excel. The next step was to filter the data to reflect companies in operation and this was done by setting the value of revenue to larger than zero. Focus was on companies that had ongoing operations and it was logical to remove companies with zero revenue to leave out dormant companies. After this, data cleaning was started by removing NA-values from the variables. Moreover, calculations of new variables resulted in some INF- and DIV/0-values which were also removed from the dataset. After the cleaning measures were completed, the dataset consisted of 2077 observations and 11 variables. Data also included the ID of the company, fiscal year, amount of personnel and total revenue. These extra variables were used in filtering the subset of the data and for identification purposes.

The next step was to handle outliers of the data. Size of the companies in the dataset was not homogenous. Therefore, all variables included some extreme outliers. Both analysis methods, PCA and SOM, are outlier sensitive which is why handling outliers is essential in establishing more convincing results. After assessing the boxplots and histograms of the variables a need to manipulate them to specific scale became apparent. In other words this meant forcing the extreme values to a certain maximum or the minimum value in the scale. Scale of -100 to 100 was used for variables CINS, ROA and ER. The scale for variable NSPE was set to between 500 to 250000. Variables QR and CAR were forced to a scale of -3 to 3. Variable DR was scaled between 0 to 100. Scale of -5 to 5 was used with variable NG. Variable RI used a scale of 0 to 200. With variable AT the scale was 0 to 2 and with variable RT 0 to 180. These actions had significant effect in reducing standard deviations of the variables. The last step was to convert the file format from xlsx-spreadsheet to CSV that could be used with R.

Package	***Version***	***Purpose***
radiant	0.9.9.1	Conducting the PCA
kohonen	3.0.10	Conducting the SOM
PerformanceAnalytics	2.0.4	Data analytics

Table 2 R packages

Software version of R used in this thesis was 3.6.3 and utilized packages as well as their versions are listed in the table 2. PCA and SOM required for the data to be normalized using the Z-score method to create zero mean and standard deviation of one for every variable. More specific details of the process are presented in chapters 4.2 and 4.3. The developed code is presented in the Appendix 2 and contains it’s own references.

4.1.2 Descriptive statistics

This subchapter presents descriptive statistics in two separate ways. First, the raw data is introduced as it was before manipulation, after which the manipulated data is represented.

Table 3 presents descriptive statistics of the raw data. In summary, the data was non-normally distributed and contained extreme outliers in both negative and positive values. One can assess that average company in the dataset has annual sales growth of 21,83 % (CINS) annually and has 95 006 € of net sales per employee (NSPE). Average company generates a 12,23 % return on assets (ROA), has quick ratio (QR) of 2,29 and cash ratio (CAR) of 0,51. In terms of capital structure, the average company has an equity ratios 36,24 % and debt ratio of 63,97 %. Furthermore, long-term solvency measures indicate that the average company has a net gearing of 1,56 and relative indebtedness of 72,38 % Finally, average asset turnover (AT) of 2,08 and receivables turnover of 36,81 are used as measures of efficiency. To sum, the field of property maintenance is, on average, characterized by relatively high values of debt but also strong values in short-term solvency and liquidity and good returns on assets.

Variance in the variables are substantial, observable through min- and max-values and standard deviation. This is attributable to the company sample not being homogenous. Largest variance is observed for variable NSPE, with a minimum value of 500 and maximum value of 1 545 000. In turn, variable CAR has the smallest variance with a minimum value of 0,00 and maximum value of 187,00. These observations necessitated variable manipulation. The following presents the descriptive statistics after outlier manipulation.

Variable	Obs	Min	Max	Mean	Median	St.dev
CINS	2077	-99,50	4420,50	21,83	4,90	155,20
NSPE	2077	500,00	1545000,00	95006,00	72600,00	103033,40
ROA	2077	-132,50	163,00	12,23	10,20	19,69
QR	2077	0,00	187,00	2,29	1,20	7,24
CAR	2077	0,00	187,00	1,40	0,51	6,33
ER	2077	-2266,70	100,00	36,24	39,00	64,19
DR	2077	0,20	2333,33	63,97	61,45	63,51
NG	2077	-533,00	1216,00	1,56	0,10	31,31
RI	2077	0,50	9999,90	72,38	25,00	485,31
AT	2077	0,03	1278,00	2,08	0,50	30,42
RT	2077	-17,00	6205,00	36,81	21,00	161,16

Table 3 Descriptive statistics before outlier manipulation

Table 4 presents descriptive statistics after outlier manipulation. As an overall observation the variance reduced significantly for variables. Largest variance remained in variable NSPE with a minimum value of 500 and maximum value of 250 000. Smallest variance was observed for variable AT with a minimum value of 0,03 and maximum value of 2,00.

Variable	Obs	Min	Max	Mean	Median	St.dev
CINS	2077	-99,50	100,00	9,82	4,90	29,54
NSPE	2077	500,00	250000,00	85829,00	72600,00	53727,68
ROA	2077	-100,00	100,00	12,22	10,20	19,38
QR	2077	0,00	3,00	1,43	1,20	0,9
CAR	2077	0,00	3,00	0,84	0,51	0,89
ER	2077	-100,00	100,00	38,13	39,00	34,81
DR	2077	0,20	100,00	59,21	61,45	28,16
NG	2077	-5,00	5,00	0,81	0,20	2,02
RI	2077	0,50	200,00	38,84	25,00	40,06
AT	2077	0,03	2,00	0,66	0,50	0,5
RT	2077	0,00	180,00	30,46	21,00	32,63

Table 4 Descriptive statistics after outlier manipulation

The following observations were made for an average company of the manipulated dataset: it grows its sales annually by 9,82 % (CINS) and has 85 829

€ of net sales per employee (NSPE). It yields a 12,22 % return on assets (ROA). Short-term solvency ratios show a quick ratio of 1,43 (QR) and a cash ratio of 0,84 (CAR). It has an equity ratio of 38,13 % (ER) and debt ratio of 59,21 % (DR). In terms of long-term solvency, its net gearing is 0,81 (NG) and relative indebtedness is 38,84 % (RI). Lastly, the average company has asset turnover of 0,66 and receivables turnover of (30,46). The manipulated dataset draws similarities to the pre-manipulated dataset in the sense that for both datasets, the average company experiences good profit levels and nearly similar capital structures. On the other hand, short-term solvency, liquidity and efficiency measures resulted in significantly weaker values in the manipulated dataset. All in all, outlier manipulation caused for the values to better describe the sample, and to give a more realistic picture of the industry. The initial extreme values skewed the distribution of the values too much.

4.2 Most important variables in the property maintenance

This chapter presents the PCA conducted in this thesis. The process is first described step-by-step after which the results are interpreted. Pre-component testing and PCA were conducted in R with a package called radiant. Dataset with outlier manipulation was used in the PCA.

The first step was to select variables to be used in the PCA. First, the ROA-variable was removed from the dataset because it is the dependent variable. Second, the causality of the variables was assessed. Variables ER and DR were discovered to have strong causality which is logical: if a company has an equity ratio of 50 % it means that its debt ratio is also 50 % because of the capital structure. Both can be calculated easily if the value of at least one is known. Therefore, variable DR was excluded from the dataset. All other variables are independent, and they are calculated from different figures in the financial statements. However, it must be noted that both QR and CAR describe liquidity and short-term solvency, and NG and RI both measure long-term solvency. Thus, these variables connect with each other by measuring the same thing, but they do it from different angles. QR is a measure of all financial assets against short-term liabilities and CAR is an absolute measure of liquidity because it only considers cash against short-term liabilities. NG measures interest-bearing debt against shareholder’s equity and RI measures how much debt there is against net sales.

Correlation matrix was assessed and there were only two zero correlations: between variables CINS and RT and between variables NSPE and RI. Metsämuuronen (2008, 28) refers to Tabachnik and Fidell who suggest that if all correlations are below 0,30, it is not beneficial to conduct the PCA. This problem was not present in this research, and the next step was to conduct tests to assure that the dataset is suitable for PCA.

Here, the dataset was normalized with the Z-score technique, resulting in zero mean and a standard deviation of one for every variable to reach more convincing results. Bartlett’s Test of Sphericity was conducted to investigate whether there were zero correlations between variables, and it yielded a p-value of less than 0,001, thus rejecting the null-hypothesis of variables not being correlated. In other words, correlations between variables are non-zero. Next, every variable was assessed with the Kaiser-Meyer-Olkin-test (KMO), also known as a Measure of Sampling Adequacy (MSA). Kaiser (1974, 35) states that values below 0,50 are unacceptable, so 0,50 was set as the threshold value for this research. This means that all variables receiving a value of less than 0,50 from the KMO test are excluded. This led to the removal of variables AT (KMO 0,26) and RT (KMO 0,40). The highest value was 0,87 for variable NG and the lowest approved value was 0,50 on variable CINS. The overall KMO value reached 0,75 after removal of the variables, and it can be considered middling according to Kaiser (1974, 35).

The next step was to define how many PCs to extract from the data. Sarstedt & Mooi (2019, 271) state that regarding Kaiser Criterion is “an intuitive way to decide on the number of factors is to extract all the factors with an eigenvalue greater than 1.” Metsämuuronen (2008,31) adds that eigenvalues over 1 are not exact borderline and if PC is easily interpretable the eigenvalue can be under 1. As a rule of thumb, the optimal number of PCs should be small enough to allow compression of the data without significant data loss. Sarsted & Mooi (2019, 272) define that extracted PCs should consist at minimum of 50 % of the total variance explained while the recommended value is 75 % or above. With these rules in mind, the test yielded 3 PCs that together explain 70,9 % of the variance.

The next phase included the extraction and rotation of the PCs. This research utilized the Varimax-rotation to create more interpretable results. Table 5 presents the results from the Varimax-rotated PCA which resulted in three rotated components including seven variables. Rotated Component 1 (RC1) includes variables QR, CAR and ER and it is used to assess solvency in short- and long-term. RC1 explains 36,7 % of the variance and produced an eigenvalue of 2,57. RC2 includes variables NG and RI, and it approximates financial leverage. RC2 explains 17,5 % of the variance and has an eigenvalue of 1,23. RC3 consists of variables CINS and NSPE, and the common factor between them is company performance. RC3 explains 16,7 % of the variance and has an eigenvalue of 1,17. In total, these rotated PCs explain 70,9 % of the variance.

According to Sarstedt & Mooi (2019, 274) when only a few PCs are extracted the loading is recommended to be over the threshold value of 0,50. In this case all loadings were clearly over 0,50 and acceptable. The highest loading was 0,90 on variable QR and the lowest loading was 0,55 for variable NG. Sarstedt & Mooi (2019, 271) continue that values of communality should be higher than the threshold value of 50 %, a condition which was also met for every variable in this study. The highest communality was 83,24 % on variable QR and the lowest value was 50,54 % on variable NG. The highest and lowest loading and communality values were found on the same variables.

Variable	*RC1*	*RC2*	*RC3*		*Communality*	*KMO*
CINS			0,75		71,86 %	0,50
NSPE			0,75		71,69 %	0,54
QR	0,90				83,24 %	0,70
CAR	0,88				80,16 %	0,71
ER	0,82				69,46 %	0,85
NG		0,55			50,54 %	0,87
RI		0,81			69,87 %	0,76
Eigenvalue	2,57	1,23	1,17		Pre-component testing	
Variance	36,7 %	17,5 %	16,7 %		Bartlett's Test of Sphericity	p-value < 0,001
Cumulative variance	36,7 %	54,2 %	70,9 %		KMO	0,75

Table 5 Results from the PCA

As a summary, liquidity as well as short- and long-term solvency are the most important financial ratios in the field of property maintenance, followed by the group of financial leverage and finally performance ratios. Results from PCA are logical because the industry has certain characteristics. First, labor-intensity of the business is clear, therefore variable NSPE is a key metric in measuring employee performance. Second, efficient companies need modern machinery and equipment and often these are financed through debt or finance companies. Thus, short- and long-term solvency ratios have an essential role. Third, the market in property maintenance has grown with a CAGR of 5,68 % between years 2014 and 2018, thus it is logical to monitor the sales growth using variable CINS.

4.3 Clustering / SOM

The following chapter introduces the SOM that was modeled for this research. The stages of SOM are described first, after which the results are interpreted. SOM was modeled using R and its package called kohonen. The outlier-manipulated dataset was used in the model.

4.3.1 Building the SOM

The data was normalized with Z-score to achieve zero mean and standard deviation of one in all variables. This is important because the SOM-algorithm uses Euclidean distance to locate the best matching unit (BMU) and if normalization is not conducted the results become distorted. Additionally, scale of the variables is in key role, because a variable with high variance might end up dominating the SOM and creating biased results. Kohonen (2014, 41) state that Euclidean distance with normalization is the preferred way in practical studies that utilize SOM.

After completion of the PCA, the most important variables were defined and will be used in constructing the SOM. Dependent variable of ROA was also included in the dataset. Total of 8 variables were used in the SOM:

- Short-term solvency: QR and CAR
- Long-term solvency: ER, NG and RI
- Profitability: ROA
- Performance: CINS and NSPE

Dataset was filtered to include only the year 2018, because the goal is to study the most recent financial ratios and the current status of the industry. Moreover, a subset of data was constructed to contain only those companies that generated revenues between 2 000 000 € and 10 000 000 € in year 2018. The purpose of this restriction was to create a picture of small companies in the field.

Once pre-processing the data was completed, the first step in modeling the SOM is defining the grid size. A rectangular grid is considered a good alternative due to its visualization properties, thus it was chosen for this research. Eklund (2004, 77-78) states that a commonly used principle in defining the grid size is that the length of x-axis should be approximately 1,3 times the length of the y-axis. This was used as rule of thumb in constructing the maps. Eklund (2004, 78) says that often the learning rate factor (alpha) starts from a value of 0,5 and end ups at 0,05.

Kohonen (1997, 88) argues that the selection of parameters does not have significant influence on the results when dealing with small maps (less than 200 nodes).

The following concerns only the larger dataset. The dataset included 479 observations. A trial-and-error method was applied by which the grid size was determined to be 12 x 9 resulting in 108 nodes. This met the condition of the x-axis being 1,3 times the length of the y-axis. Learning rate factor (also known as alpha) was determined to be 0,70 in the beginning and 0,01 in the end. These parameters led to well-defined clusters. Other initial parameters resulted in clusters that were not spherical and had unclear borderlines. Algorithm's learning process was also assessed through a plot which presented the development of the mean distance to closest unit. Lynn (2014) states that the "distance should reach a minimum plateau" and that more iterations are needed if the curve in the plot is continually decreasing. In this case the plateau-condition was met with 1000 iterations. The SOM setup used Gaussian neighborhood function and Euclidean distance.

The following section concerns only the subset of the data which included 64 observations. The grid size was determined to be 5 x 4 resulting in 20 nodes. The grid size almost met the length of the x-axis rule. Learning rate factor was set to initial values of 0,7 in the beginning and 0,01 in the end. With 250 iterations the minimum plateau condition was achieved. Subset SOM-setup used also Gaussian neighborhood function and Euclidean distance.

Clustering was conducted using Ward's method. According to Eklund (2004, 79), this method defines the number of natural clusters in the data. He continues by saying that Ward's method "increases the objectivity of the cluster analysis." Commonly Ward's method is considered to find spherical clusters from the dataset. Therefore, Ward's method is applied in this thesis and the results are assessed through dendrograms.

Regarding the larger dataset, 8 clusters seemed to be a good selection, and for the smaller subset the selection included 5 clusters. Figure 15 presents the dendrogram of the dataset and figure 16 presents the dendrogram of the subset.

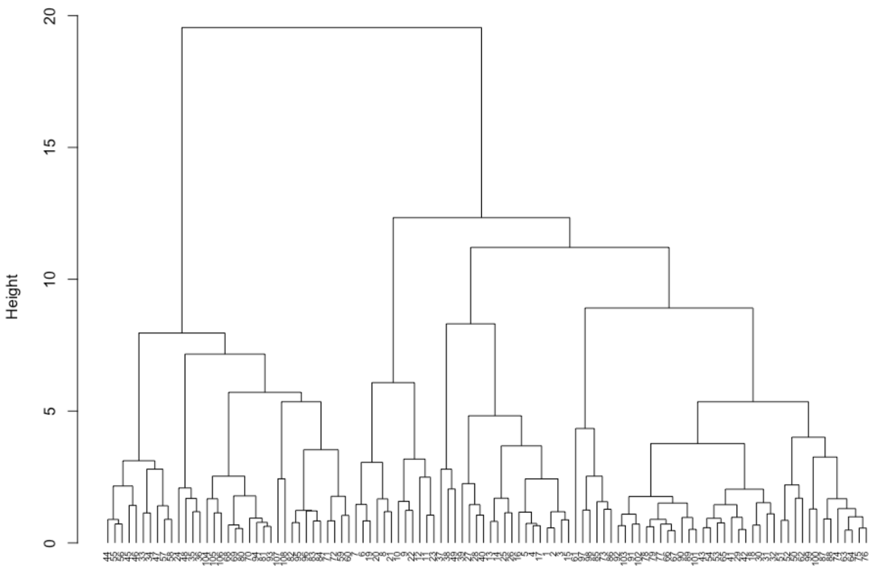

Figure 15 Dendrogram of the dataset

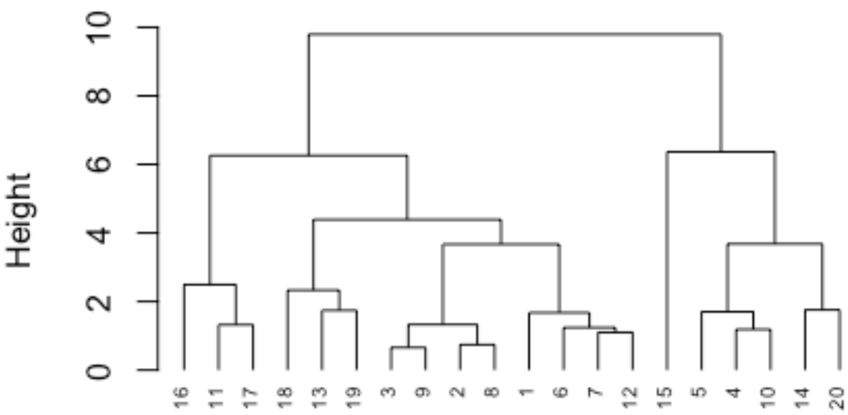

Figure 16 Dendrogram of the dataset (small companies)

With these settings results were consistent with both datasets. Random initialization process caused SOMs to differ in the structure and the best cluster could be found on different sides of the SOM through trial-and-error-phase. Clustering resulted same results no matter where were the best and the worst end of the SOM.

4.3.2 The current status of the industry

This subchapter presents the SOM constructed for the dataset and introduces current status of the industry. First the structure of the SOM is introduced and after this the clustering is presented. Clusters are analyzed based on their characteristics. Finally, heatmaps are presented one variable at a time.

Figure 17 presents mapping of the SOM. There was only one empty node and Lynn (2014) states that the presence of empty nodes means the map is too large for the samples, and that the minimum target for number of samples in a node should be between 5 to 10. It must be remembered that this depends on the dataset used and in general it can be said that this condition was met.

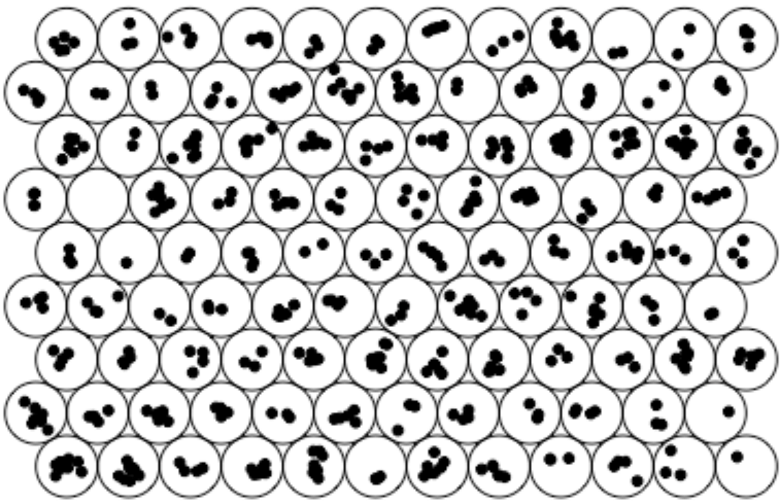

Figure 17 Structure of the SOM nodes

Figure 18 presents the quality plot and Wehrens & Kruisselbrink (2019) states that "quality shows the mean distance of objects mapped to a unit to the codebook vector of that unit. The smaller the distances, the better the objects are represented by the codebook vectors." Based on this description, the results are satisfactory and the highest mean distances (darker grey shadows) are at the edges of the SOM. Overall picture of the quality plot is uniform.

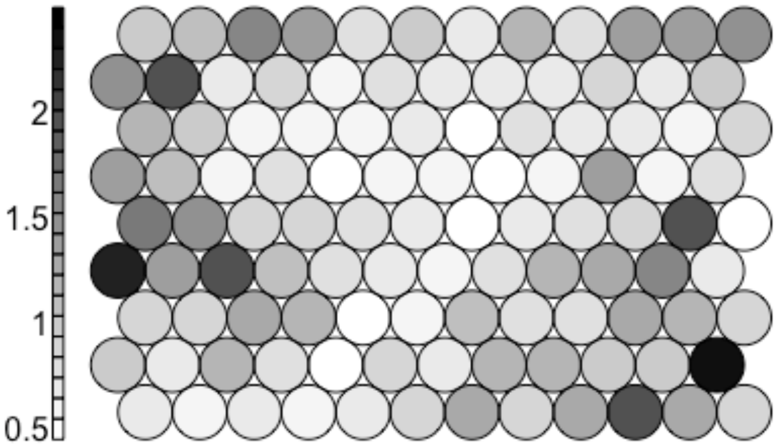

Figure 18 Quality of the SOM

Figure 19 presents the U-matrix which is one of the most popular ways to visualize distances between nodes. Light grey colors indicate nodes are close to each other and darker they get the further the nodes are. It is apparent that there are easily separatable groups known also as clusters. Most of the nodes are quite close to each other and are concentrated around the centre as well as around the left lower corner of the SOM. Highest distances are at the edges of the SOM.

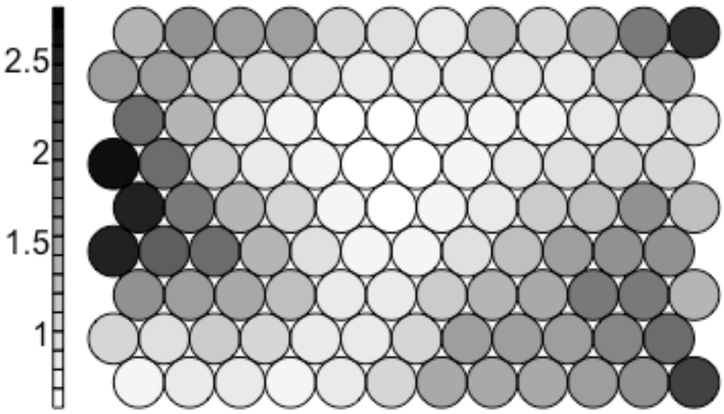

Figure 19 U-matrix

Vesanto & Alhoniemi (2000, 586) state that clustering is an effective way to use information offered by the SOM and that large maps require clusters because the investigation of all nodes leads to far too many summaries. Ward's method was selected to be used for clustering in this thesis because it defines spherical and natural clusters. Figure 20 presents clustering of the SOM. The figure supports the usage of 8 clusters.

The following step is to summarize the clusters. Average values for every variable in every cluster were calculated. Clusters are described through a 5-step scale: very low, low, medium, high and very high and these values correspond to calculations. Very high is assigned to the highest values and very low to the lowest values. Medium is assigned to mean values. Low and high values are assigned to values between both sides of the mean. Assessment is done from five perspectives: performance, profitability, short- and long-term solvency and capital structure. The assessment was also conducted visually from the code plot which is presented in figure 21. The code plot can be difficult to assess and therefore, 5-step-scale assessment were conducted.

The first cluster (slateblue) is in the lower left corner. The cluster is characterized by low profitability and low values of capital structure and short-term solvency. The highest values of the cluster are found in variable NG while values for variable RI are on a medium level. Performance ratios are also on a medium level.

The second cluster (springgreen) is in the lower right corner. Cluster is characterized by high levels of performance and profitability ratios. Capital structure and short- and long-term solvency are on medium levels.

The third cluster (steelblue) is in the middle of the SOM and it is the largest cluster of this dataset. It can be thought of as the average cluster of the dataset. It has low values related to performance and profitability ratios. It contains medium values on long-term solvency and on capital structure and short-term solvency is considered to have low to medium values.

The fourth cluster (tan) is in the right side of the SOM and above the second cluster. The cluster is characterized by high to very high performance ratios while short-term solvency and capital structure have very high values. Profitability is assessed to be high. Long-term solvency is assigned with very low to low values.

The fifth cluster (thistle) is located towards the middle of the SOM and is next to the fourth cluster. It can be considered the best cluster from point of view of profitability. Cluster is characterized by highest profitability ratios and low performance and long-term solvency ratios. Short-term solvency is on medium to high levels and capital structure value is high.

The sixth cluster (tomato) is the smallest cluster of the dataset and it is found on the left edge of the SOM. It is characterized by very low values on profitability and very high long-term solvency ratios. There are also low values on short-term solvency and capital structure. Performance ratios are on levels between very low and medium.

The seventh cluster (turquoise) is in the top right corner. The cluster is characterized by low values on performance and on long-term solvency. Capital structure have high values. Profitability is on medium levels. Short-term solvency is between high to very high levels.

The eight and final cluster (violetred) is in the top left top corner of the SOM. It is characterized by the worst performing companies. Profitability, short-term solvency and capital structure have very low values. Performance ratios are between very low and medium. Long-term solvency is assigned values between very low and high.

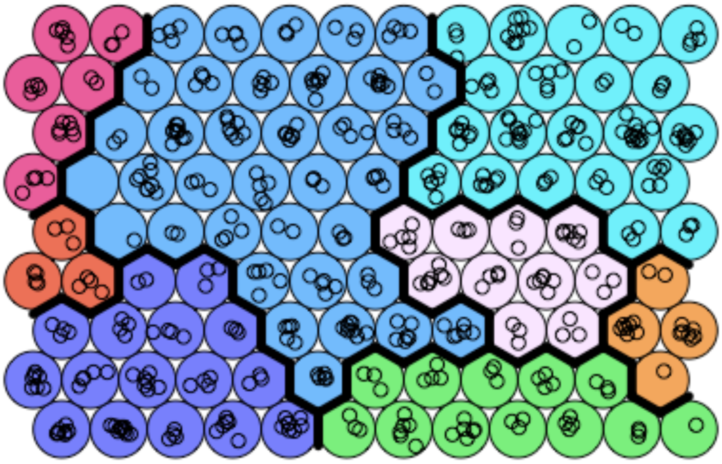

Figure 20 Clusters

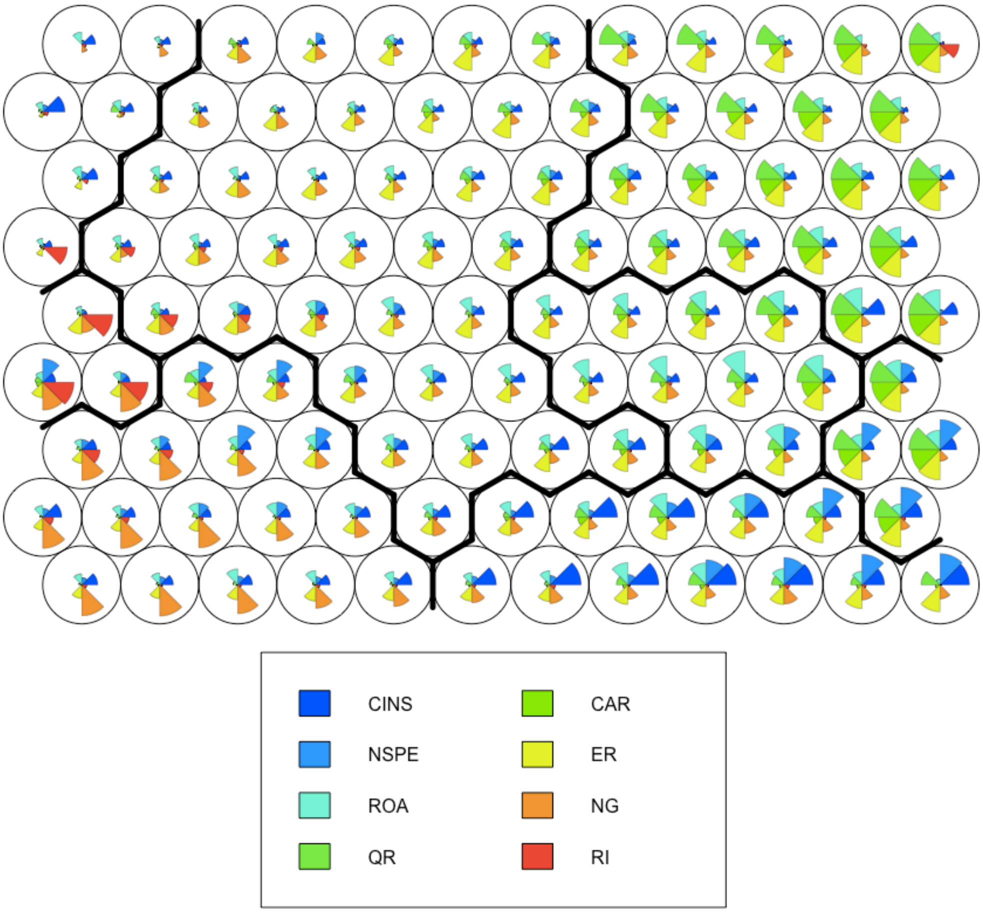

Figure 21 Code plot

Cluster	Performance	Profitability	Short-term solvency	Long-term solvency	Capital strucure
1	M	L	VL-L	M-VH	L
2	H-VH	H	M	M	M
3	L	L	L-M	M	M
4	H-VH	H	VH	VL-L	VH
5	L	VH	M-H	L	H
6	VL-M	VL	VL-L	VH	L
7	L	M	H-VH	L	H
8	VL-M	VL	VL	VL-H	VL

Table 6 Cluster descriptions

Table 6 summarizes findings from the clusters. The third cluster is the average cluster and the fifth cluster is consisted the most profitable companies. The eighth cluster is the worst performing cluster. From the point of view of long-term solvency the first cluster is characterized by highest values of variable NG. The sixth cluster have highest values of variable RI as a common factor.

Figure 22 introduces component planes presenting univariate distribution of the variables. It makes it easier to assess relationships between the variables. Red color indicates high values and blue color low values. These assessments are implicated only from the aspects of the profitability and the growth to gain more knowledge to create incentives to develop the business.

For high profitability it is common that performance ratios are on medium to high levels. Relationship between the profitability and the liquidity is more complex. For high values in the variable ROA the level in variable QR is between medium to very high and with CAR-variable the level is between low to very high. The interaction between high profitability and capital structure is clear too. For high ROA the levels in variable ER are between medium to high. Relationship between profitability and long-term solvency ratios are clear. For high profitability the variable NG is on medium levels and the variable RI have very low values. High values in long-term solvency have relationship with low values in variable ROA.

High growth means low to very low values in liquidity and medium levels in capital structure. Very high values on short-term solvency and capital structure are concentrated on the right side of the SOM and high long-term solvency values can be found from left side of the SOM.

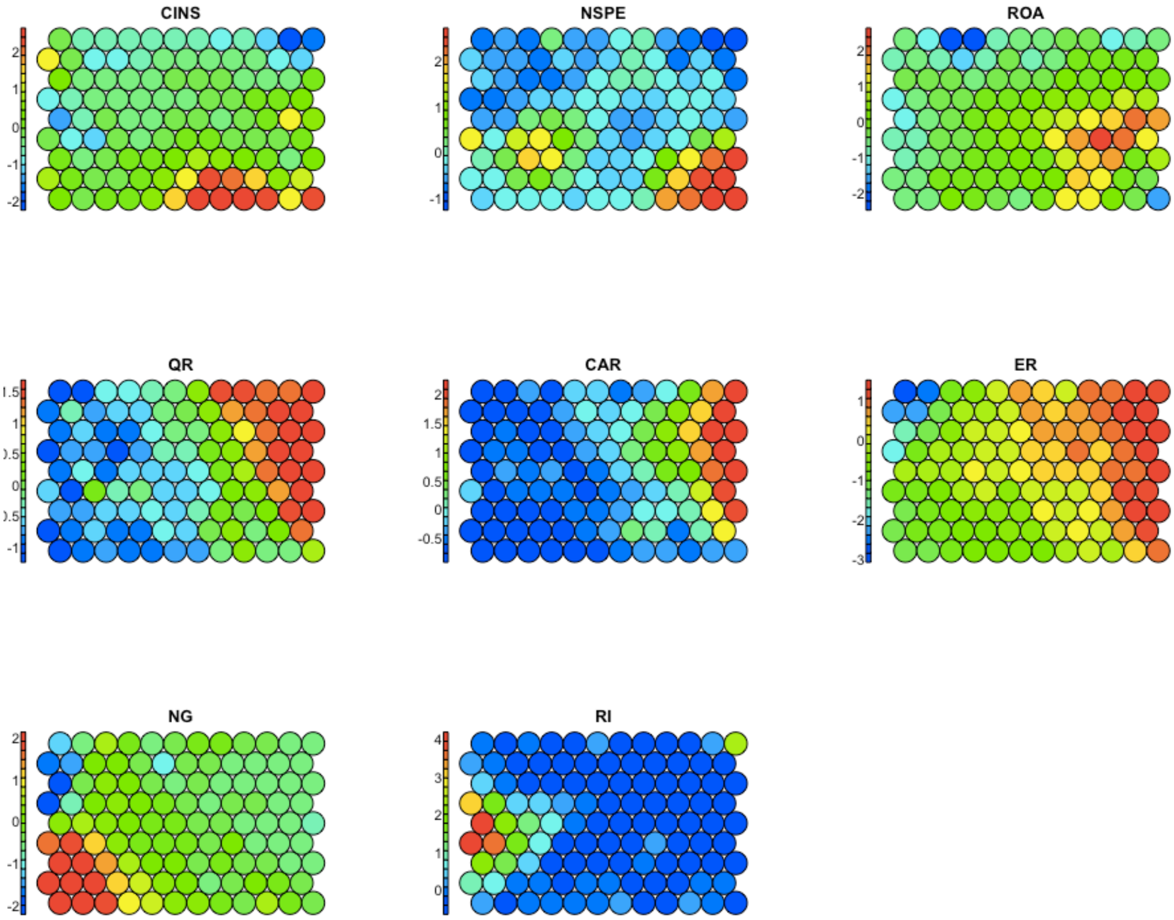

Figure 22 Component planes

As a summary from the assessment the average companies have low profitability and performance ratios. Long-term solvency and capital structure are considered on medium levels and short-term solvency have low to medium values. The best cluster is characterized with low performance and long-term solvency ratios and high capital structure. This group have companies with medium to high values on short-term solvency and long-term solvency have low values. Worst performing clusters are clusters six and eight. Both have almost same characteristics on performance, short-term solvency and on capital structure ratios and the largest difference is on long-term solvency where cluster six have very high values and cluster 8 have very low to high values.

For high profitability company must focus on medium levels on short-term solvency and on high values on capital structure. For high performance companies seems to have very low to medium values on long-term solvency, which means that there is debt which used for growth. At the same time capital structure of the companies is on medium to high levels, which indicates good financial planning and reasonable amount of debt with respect to the balance sheet. Very low to low capital structure is in association with very low to low values on short-term solvency.

4.3.3 The current status of the small companies

This subchapter presents the SOM constructed for the subset consisting of small companies. First, the structure of the SOM is presented and after this the clustering results are presented. Clusters are analyzed based on their characteristics. Lastly, heatmaps are introduced one variable at a time.

Figure 23 presents the mapping of the SOM. There were no empty nodes, but 6 nodes had only 1 sample. The subset consisted of 64 observations and a 5x4-sized map is considered relatively large one for the dataset and therefore, it was expected that the distribution between nodes would be substantial. This map size also made it possible to maintain uniqueness of the samples and highlighted the fact that the subset of the data is not homogenous at all.

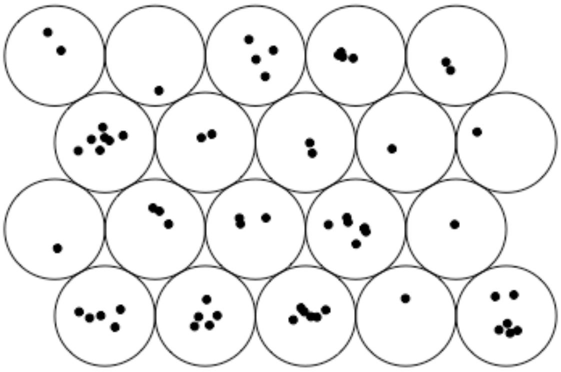

Figure 23 Structure of the SOM nodes (small companies)

Figure 24 introduces the quality plot. The picture is uniform and the highest values are at the edges of the SOM. Based on the plot the results are satisfactory and the highest mean distances (darker grey shadows) are at the edges of the SOM.

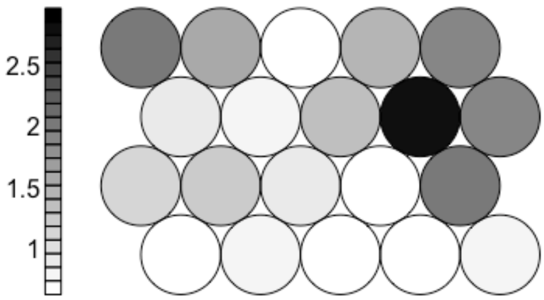

Figure 24 Quality plot (small companies)

Figure 25 introduces the U-matrix and it visualizes the distances between nodes. Light grey colors indicate that the nodes are close to each other and the darker they get the further the nodes are from each other. Most of the nodes are close to each other and are concentrated around the centre and the bottom of the SOM. Highest distances are on the right edge of the SOM.

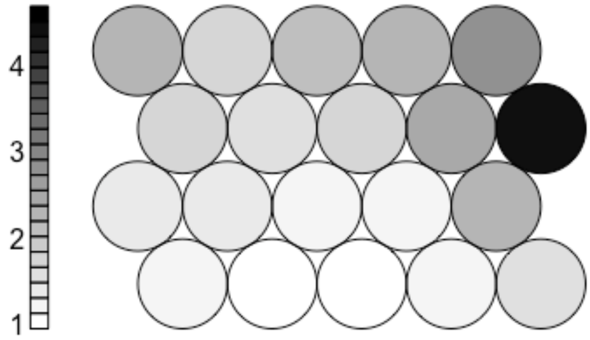

Figure 25 U-matrix (small companies)

The following will present the characteristics of the clusters and figure 26 highlights the SOM with cluster borderlines. With the subset average values were calculated for every variable in every cluster. Values correspond to 5-step scale: very low, low, medium, high and very high. The highest value is assigned to very high and the lowest values is assigned to very low. Mean value is corresponding to medium and low and high values are on the both sides of the mean. Clusters of the subset are also analyzed from five different aspects: performance, profitability, short- and long-term solvency and capital structure. Assessment was also conducted visually using the code plot presented in figure 27.

The first cluster (slateblue) is in the lower left corner. It is the largest cluster of the subset and can be considered as an average cluster. The cluster is characterized by very low performance ratios and low values in solvency ratios. Capital structure ratios have high values and profitability ratios have medium ratios.

The second cluster (springgreen) is at the right edge of the SOM and it contains a subcluster which is cluster number 5. Common for the companies in the cluster are very low levels of short-term solvency and medium levels of performance ratios and long-term solvency. These companies are also characterized by low values regarding profitability and capital structure.

The third cluster (steelblue) is in the top left corner of the SOM. It is considered the best cluster of the dataset. The cluster Is characterized by very high levels of profitability, short-term solvency and capital structure. Long-term solvency is on a medium level and performance is on a low to medium levels.

The fourth cluster (tan) is characterized by high levels of profitability and on capital structure. Performance is on medium to very high levels. Solvency have values between very low to medium.

The fifth cluster (thistle) is the worst performing cluster of the dataset and it is also the smallest. It is a sub-cluster of the second cluster. It consists of only one company that differs heavily from the rest of the companies of the subset. It is worst performing in terms of profitability, long-term solvency and capital structure. However, short-term solvency is on low to medium levels.

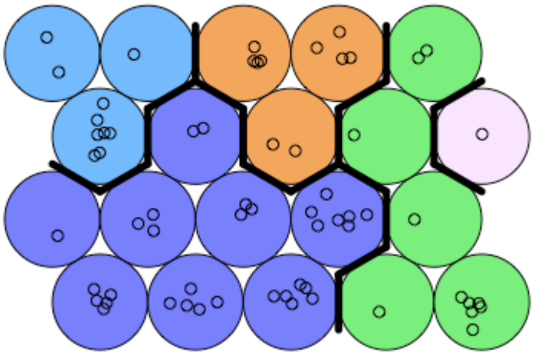

Figure 26 Clusters (small companies)

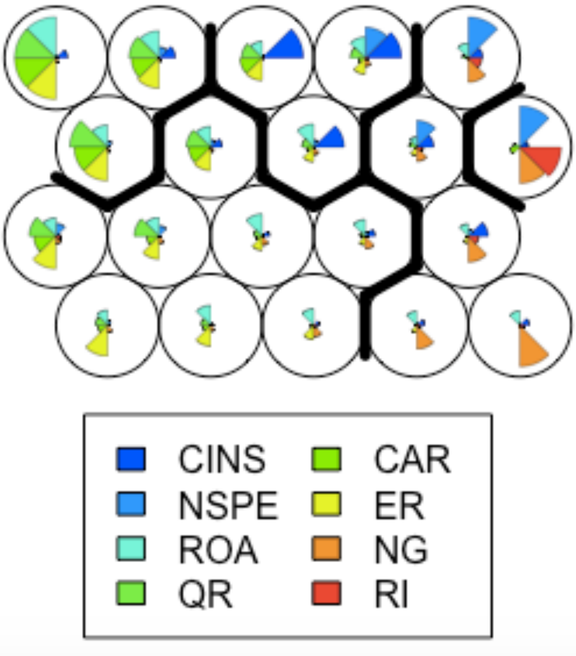

Figure 27 Code plot (small companies)

Table 7 summarizes the findings from clusters of the subset. The first cluster is the average cluster, third cluster is consisted from the most profitable companies and the fifth cluster consist the worst performing companies.

Cluster	*Performance*	*Profitability*	*Short-term solvency*	*Long-term solvency*	*Capital structure*
1	VL-L	M	L	L	H
2	M-H	L	VL	M	L
3	L-M	VH	VH	M	VH
4	M-VH	H	L-M	VL-L	H
5	VL-VH	VL	L-M	VH	VL

Table 7 Cluster descriptions (small companies).

Figure 28 introduces component planes presenting univariate distribution of the variables in the subset. Similarly like with the larger dataset these assessments are implicated only from the aspects of the profitability and the growth to gain more knowledge to create incentives to develop business.

For high profitability it is common that performance ratios are on low to medium levels. Relationship between profitability and liquidity is clearer than with the whole dataset. For high values in the variable ROA the short-term solvency ratios have very high values.

The interaction between the high profitability and the capital structure is clear too. For high ROA the levels in the variable ER are high too. Relationship between the profitability and the long-term solvency ratios are clear. For high profitability the variable NG is on low levels and the variable RI have very low values. High values in the long-term solvency ratios have relationship with low values in the variable ROA.

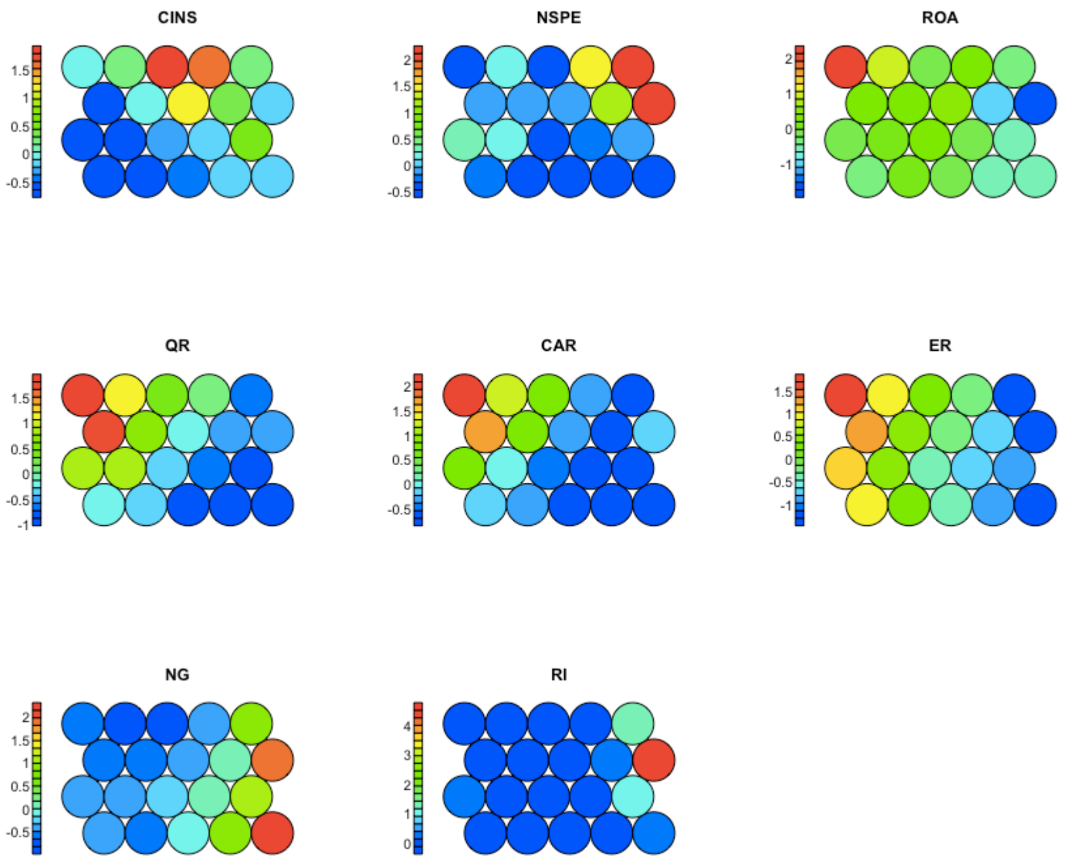

Figure 28 Component planes (small companies)

As a summary average companies in the subset have medium profitability and very low to low performance ratios. Both the short- and the long-term solvency ratios are on low level and the capital structure is considered to be high. Best cluster have very high values on profitability, short-term solvency and on capital structure. The long-term solvency is on medium level and performance ratios on low to medium levels. The worst clusters (cluster two and five) have low to very low capital structure and short-term solvency ratios have very low to medium values.

For high or very high profitability company must focus on medium to high levels on short-term solvency and on high to very high values on capital structure. For high performance companies seems to have very low to medium values on long-term solvency, which means that there is debt which used for growth. At the same time capital structure of the companies is on low to high levels, which indicates financial planning and reasonable amount of debt with respect to balance sheet. Very low or low capital structure is in association with very low or low profitability.

5 CASE FOR COMPANY X

Following chapter introduces the case for Company X. In the first section Company X is introduced. The second section presents comparison of financial ratios between the company and clusters. The third section introduces incentives to develop the business and the profitability further.

5.1 Presentation of the Company X

Goal of this chapter is to present the company and to create an overall picture of its operations. Brief history, management, personnel and operations of the company are introduced.

Company X was established in 1966 in the metropolitan area of Helsinki and it operates in property maintenance industry. In this time the operating region was under construction and there were no providers of real estate management or property management services to serve recently established housing companies in the area. Therefore, Company X was established by housing companies of the area to tackle this problem. The business is divided into four segments, based on cost-accounting: real estate management, property maintenance, parking places and apartments. Company's official industry is property maintenance.

Company is a limited liability company that has 74 owners which all are housing companies. There has been only one transaction between stockholders in the history of the company taking place when deconstructed housing company sold its stocks to the other owner of the company. Most of the owners are also customers of the company and utilizing its property maintenance and real estate management services and rent parking places.

The Board of Directors includes 6 persons who are selected from the stockholder housing companies. The board consists of the chairman of the board and five board members. One of the board members is also acting as a vice-chairman of the board. Management team consists of three people: chief executive officer, chief financial officer and property maintenance manager. At the end of fiscal year 2019 the company employed 56 people on permanent basis, including three senior executives, twelve housing managers, twelve office workers and twenty-nine people in property maintenance.

5.2 Financial ratios of Company X

This subchapter presents brief comparison between Company X and its own cluster and the best cluster of the subset. Mean values of variables are calculated for the both clusters and are compared to Company X's financial ratios. These observations create a foundation for actions to develop profitability. Table 8 introduces financial ratios of Company X, its cluster and the best cluster.

First, the comparison between Company X and its own cluster is conducted. Company X is in cluster 1 which is the average cluster. Company X outperforms its cluster in short-term solvency ratios and capital structure. Company X is underperforming in performance ratios and in profitability. In long-term solvency ratios Company X obtains smaller values than the cluster, which can be also considered as a better performance. In summary cluster 1 has very low to low values in performance ratios and has medium profitability. All solvency ratios are on low level, but capital structure is on high level.

Second, the comparison between Company X and the best cluster is executed. In this comparison the best cluster is outperforming in every variable except in the variable RI. The capital structure is very similar between Company X and the best cluster. The largest differences are in performance measures and in profitability. The smallest differences are in long-term solvency ratios. As a summary the best cluster has very high values on profitability, short-term solvency and capital structure, but medium values regarding on long-term solvency.

Observation	*CINS*	*NSPE*	*ROA*	*QR*	*CAR*	*ER*	*NG*	*RI*
Company X	3,3	72982	7,4	1,3	1,0	73,3	-0,2	16,3
Cluster 1	6,3	83707	14,6	1,2	0,6	39,7	0,1	22,2
Cluster 3	10,1	92907	38,4	2,8	2,4	74,6	-0,5	18,3

Table 8 Performance comparison between company and clusters

As a summary, two of the most essential actions Company X must take is to grow its sales and enhance its profitability to gain access into a better cluster. The next chapter will present concrete actions for this purpose.

5.3 Incentives to develop profitability and business

This section consists of two parts. The first one introduces recommendations assessed through numbers and financial ratios and the second one presents more practical solutions to develop business and profitability.

5.3.1 Through numbers

Turnover of trade receivables have been growing substantially from 9 days to 23 days between 2014 and 2018. During the same period peer companies have kept it on the stable level. Motivation for this is that there is no need to act as a bank for the customers with long payment terms. The efficiency in handling the trade receivables can be achieved with the daily billing routine. It is also essential to close assignments immediately when they are finished speeding up the billing. This recommendation can be considered as a universal one in the service business and it aims to establish an unambiguous process to enhance efficiency.

Value of the CAR-variable stands out from the cluster analysis and Company X has excess cash in comparison to its cluster. Liquidity is generally a good thing but when it is too strong it is not beneficial for the company. Excess cash can be converted to more profits with investments or it can be paid out to owners as dividends. Incentive for this is that cash sitting in a bank account does not yield interest nowadays, it is more likely to create a liability if deposit interest rates are negative. Cash management with low-risk and a liquid solution should be considered. In my experience companies often use fixed income portfolios, corporate bond funds and government bond funds in this purpose. This recommendation is also universal and can be applied to any company that has excess liquidity.

Company X has a very solvent situation and its capital structure is very strong against cluster averages. Use of debt for investments and use of finance company services for equipment purchases should be considered. It can be assessed from the cluster analysis that strong growth is associated with lower equity ratios and medium values in long-term solvency which means that debt is used to create growth. When assessing the profitability a very strong capital structure is unbeneficial for financial ratio calculation and can send a wrong signal about the company’s situation. Enhancing the equity ratio further will not create more benefits after the good level have been reached. Using the debt wisely can be considered as active asset and liability management and this is also a universal suggestion for every company.

5.3.2 Practical procedures in the company

One of the greatest worries a business controller deals with revolves around uncertainty of the right number hours billed. If the hours are not maxed out company will eventually lose money, because it misses out an opportunity for 100 % billing. This is essential for service businesses in which billing is done after the assignment is finished. It is also very important when many employees are in association with different stages of billing. ERP-software and automation are closely related to this matter. Notifications on open bills should be automated to remind of immediate closing when possible. Incentive here is to create an unambiguous process that helps all employees to work through jointly accepted rules. This is mostly done with the building the process and creating supervision for it. This suggestion is also universally applicable for all companies and it is one element of a modern business controller's job.

Pricing should also be considered because net sales per employee is underperforming against the cluster average. It must be mentioned that there are lots of differences between the companies and there is no information available regarding which geographical areas they operate in so the market's price level cannot be defined or assessed. One key element in pricing the agreements is to build an automated spreadsheet-pricing tool. This is beneficial in a case a key person in pricing leaves the company. A pricing tool helps to maintain database of pricing parameters and historical data of pricing. It also can be used to find anomalies in existing pricing agreements and to define future pricing checks.

Company X should define a dividend policy. A company's basic idea is to create wealth for its owners and dividend policy must be defined. This is closely related to active asset and liability management and payout of dividends can be one way to streamlining the balance sheet. Furthermore, growth strategy should be defined. It is essential for managing the company and acts as a basic guideline for its development. The field of business have grown with CAGR of 5,8 % between 2014 and 2018 and it is good guideline for the future objectives.

6 SUMMARY AND CONCLUSIONS

Final chapter of this study summarizes the findings and consists three parts. The first part is focused on answering the hypothesis. The second part presents implications for the industry. The third part introduces limitations of this study and suggest further research subjects.

6.1 Answers to hypotheses

This subchapter is divided into three parts. The first part introduces most important financial ratios. The second part presents the current status of the industry. The last part introduces the comparison between Company X and most profitable companies in the industry. The last part also presents actions to develop the business and the profitability.

6.1.1 Most important financial ratios

This subchapter consists of three parts. The first one states the findings from the previous literature and the second summarizes results from the PCA. The third part compares results from the PCA with the previous literature.

As a reminder the first research question was:

Which financial ratios are the most important in the field of property maintenance in Finland?

According to previous literature the most important financial ratios are presented in the table 9. In summary the most important one is ER followed by RI and QR. The common factor between these variables is solvency in short- and long-term. Importance was assessed based on the amount of times they were used in previous literature. Profitability measures were excluded.

Variable	*Financial ratio group*	*Times used in previous lliterature*
ER	Long-term solvency	4
RI	Long-term solvency	3
QR	Short-term solvency	3
CR	Short-term solvency	2
WCTS	Short-term solvency	1
NG	Long-term solvency	1

Table 9 Most important financial ratios according to previous literature.

Table 10 presents results from the PCA conducted in this study. As a summary, performance ratios, short- and long-term solvency ratios are the most important ones in the field of property maintenance in Finland. Most important single variables were CINS, NSPE, QR, CAR, ER, NG and RI.

Variable	*Financial ratio group*
CINS	Performance
NSPE	Performance
QR	Short-term solvency
CAR	Short-term solvency
ER	Long-term solvency
NG	Long-term solvency
RI	Long-term solvency

Table 10 Most important financial ratios according to PCA.

As a conclusion from the previous literature and from the PCA the solvency in short- and long-term is in a key role for companies operating in the field of property maintenance. The PCA analysis revealed that performance ratios serve an important role. Results from this research verify results from the previous literature and the variables that were used were very similar especially when examining short- and long-term solvency.

6.1.2 Current status of the industry

This subchapter presents the results of analyzing the descriptive statistics of the outlier manipulated dataset as well as results of clustering the dataset. The goal is to define the current status of the industry.

The second research question was:

The goal is to study the current status of companies in the field of property maintenance in Finland.

According to the findings by the SOM average companies have low profitability and performance ratios while long-term solvency and capital structure are on medium levels. Short-term solvency ratios are on levels between low and medium. Cluster averages of the companies are very near the medians of the whole dataset and only variables ROA, QR, NG and RI had significant anomalies against medians from the descriptive statistics. It is logical that most of the companies in the dataset are near these values because the SOM aims to find similarities between the observations.

Best performing companies have low values in terms of performance and long-term solvency ratios while simultaneously having a high capital structure. Short-term solvency is on medium to high levels.

Worst performing companies have very low profitability. Performance ratios are considered to be on very low to medium levels. Short-term solvency and capital structure are on very low levels.

The following discusses the status of the companies using descriptive statistics between the years 2014 and 2018. On average the capital structure of the companies consisted 39 % of equity and 61 % of debt and with net gearing of 0,81 and relative indebtedness of 38,84 %. This means that debt plays a remarkable role in the field of property maintenance. Even though companies have a lot of debt in terms of capital structure, when it is compared to net sales the situation changes and the amount of debt is considered to be on a good level. Additionally, companies in the field are in liquid position to meet their short-term liabilities as illustrated by the variable QR with an average value of 1,43. Values over 1 are universally accepted to be good for this variable. The average value of variable CAR also highlights satisfying short-term solvency. The property maintenance market has grown with CAGR of 5,8 % between 2014 and 2018. On average companies have been able to grow 9,82 % annually, but this is a relative measure. Between 2014 and 2018 the largest relative revenue growth occurred in companies employing more than 50 people, which means that the field is most likely to face consolidation and growth is achieved through mergers and acquisitions.

All in all, the highest profitability in the field is associated with a strong capital structure and high values of short-term solvency ratios. Moreover, the most profitable companies are not growing substantially. Correspondingly low profitability has a relationship with the lowest values for short-term solvency and weak capital structure. The field is characterized by the use of debt and the market has grown with a CAGR of 5,8 % between years 2014 and 2018.

6.1.3 Company X

This subchapter consists of two parts. The first one state the findings from clustering the small companies' dataset and the second one presents the actions to take to develop the profitability and the business.

As a reminder the third research question with its sub-questions were:

How does the case company compare to the most profitable ones in the field?
In what kind of cluster does the case company belong to?
What actions can be taken to develop the profitability and the business?

Observation	***CINS***	***NSPE***	***ROA***	***QR***	***CAR***	***ER***	***NG***	***RI***
Cluster 1	-	-	-	+	+	+	-	+
Cluster 3	-	-	-	-	-	-	-	+

Table 11 Company performance against clusters (small companies)

Table 11 introduces the results for Company X against its own cluster and the best performing cluster. In the table a minus-sign indicates worse performance and a plus-sign indicates better performance than the cluster. Company X is located in cluster one, the average cluster of the small companies. The best performing cluster was cluster three.

Companies in the average cluster have medium profitability, high capital structure and simultaneously short- and long-term solvency are on low levels. Performance of these companies are on very low to low levels. This means that these companies have established their position on the market and are not looking for rapid growth. One can argue that these companies have little debt because their capital structure is on a high level.

Companies in the best performing cluster have very high values of profitability, short-term solvency and capital structure. At the same time long-term solvency is considered to be on medium levels and performance ratios are on low to medium levels. These companies are extremely solvent and liquid and have a very strong balance sheet. It is also common that these companies have established their position on the market and are not looking for substantial growth. This can be seen as a one reason for very high values in profitability, because resources can be used efficiently and there is no need to actively use resources towards growth.

Company X outperformed its own cluster in short-term solvency ratio, capital structure and also with variable RI. Company X underperformed in performance and profitability ratios against the cluster averages. Company X underperformed against the best-performing cluster in every variable except variable RI. All in all, Company X needs to develop performance and profitability ratios to gain a place in the best performing cluster. Next, the actions to enhance profitability are summarized.

Action	Description	Effect
Billing routine	Establishing the daily billing routine and ERP-development.	Performance and efficiency.
Investments	Excess liquidity should be invested with low-risk and a liquid solution.	Short-term solvency and profitability.
Active asset and liability management	Use of debt in investments.	Profitability, solvency and capital structure.
Pricing tool	Creating an automated Excel-spreadsheet pricing tool.	Performance and profitability.
Update of the strategy	Defining the growth and dividend policy.	Performance, efficiency, solvency and capital structure.

Table 12 Summary of actions

Table 12 presents a summary of actions introduced for the Company X. All in all the presented actions have an effect on all financial ratio groups and with the purpose of developing business and profitability. Defined actions have an impact on processes, financial operations and on the strategy of the company. Presented actions are also universal and can be used in any company.

6.2 Implications for the industry

As a take-away message to the property maintenance industry the most important financial ratios are useful indicators of the company performance. Those can be used to build the key performance indicators for the management to monitor company performance. Essential information and knowledge have a significant role in managerial decision-making and the defined variables are useful in this context. The findings from the SOM are useful when benchmarking the companies and when developing and planning the business further. SOM can be considered as a very useful tool to assess the status of the industry. It formed comprehensive visualization with financial ratios and defined characteristic for success. Results work as guidelines in business development. For a company to reach high to very high profitability it must factor in that short-term solvency and capital structure have to be on medium to very high levels. This highlights the importance of good financial planning and it must simultaneously be assessed through the targeted level of growth. If the performance ratios are on high to very high levels the long-term solvency is on very low to medium levels. This indicates use of debt to boost the growth. In terms of short-term solvency it does not necessarily have to be very high to achieve high profitability. This may mean allocating cash to be used for growth, more profitable investments or dividend payout to owners. When assessing the worst performing companies it can also be said that a very low to low capital structure is most likely to lead very low to low profitability and to very low to medium performance ratios. At the same time these companies have very low to low values on short-term solvency ratios.

The PCA and the SOM can be considered as useful tools in intra-industry profitability analysis. Both methods can be easily implemented using R. The code for this research is available in Appendix 2 and can be used for further research.

6.3 Limitations and further research subjects

With scientific approach it is essential to define limitations of this study. It must be understood that formulas of financial ratios vary from source to source and Delen et al. (2013) define that there is no universally accepted list of formulas and grouping of the financial ratios in the previous literature. In Finland Corporate Analysis Reg. Association is the benchmark of defining the financial ratios and their formulas are used in banking sector and in other institutions and companies. This study also relied on their contribution to the matter.

Sector-coding can lead to wrong assumptions, because it is assigned by the main business of the company. Companies may have secondary businesses, which may have significant impact on overall profitability. This is highlighted by the fact the sample is non-homogenous in this study. Therefore, generalization of results must be done with discretion and these results are only applicable on the field of the property maintenance.

Results from the PCA are acceptable, but the explained variance was only 70,9 % with three Rotated Principal Components. It could have been higher, because it is recommended to be over 75 % according to Sarstedt & Mooi (2019, 272), but also their recommendation for minimum threshold value is 50 %. From this point of view results are in overall acceptable.

Both methods the PCA and the SOM are considered to be black box algorithms and their output can vary from time to time and it highlights the importance of testing the outputs and meeting the assumptions of the methods. From the point of view of software results can slightly vary depending on used software, packages and algorithms. The PCA was tested with Bartlett's Test of Sphericity and with Kaiser-Meyer-Olkin-test. The SOM was assessed only through algorithm's learning process, which is limitation in this study. Although both SOMs were tested with many different amounts of iterations, the mean distance between samples didn't get any better.

This study concentrated on industry and on small companies and it is very context depended. For further research it would be logical to research differences between micro, small, medium and large sized companies in the industry. ROA-variable were used on this study and one aspect can be different measure of profitability (EBITDA, EBIT or net result) in further studies.

The field of Finnish property maintenance is very under-researched. There are only few previous studies available and further research must be waited to verify findings from this study. In overall outcome of the previous studies and this research are similar. All in all this study filled a research gap and is a basis for forthcoming research. Methods presented are applicable and beneficial for further use in any inter-industry analysis.

REFERENCES

Berk, J., DeMarzo, P. & Harford, J. (2015). Fundamentals of Corporate Finance. Third edition, global edition. Harlow: Pearson.

Brealey, R. A., Myers, S. C. & Allen, F. (2011). Principles of Corporate Finance. Tenth edition, global edition. New York: McGraw-Hill/Irwin.

Brooks, C. (2014). Introductory Econometrics for Finance. Third edition. Cambridge: Cambridge University Press.

Corporate Analysis reg. assoc. (2013). The Guide to the Analysis of Financial Statements of Finnish Companies. Second edition. Helsinki, Gaudeamus.

Das, G., Chattopadhyay, M. & Gupta, S. (2016). A Comparison of Self-Organising Maps and Principal Components Analysis. International Journal of Market Research, Vol 58, Issue 6, pp. 815-834. DOI 10.2501/IJMR-2016-039.

Delen, D., Kuzey, C. & Uyar, A. (2013). Measuring Firm Performance Using Financial Ratios: A Decision Tree Approach. Expert Systems With Applications, Vol 40, Issue 10, pp. 3970-3983. DOI 10.1016/j.eswa.2013.01.012.

Drury, C. & Tayles, M. (2006). Profitability Analysis in UK Organizations: An Exploratory Study. The British Accounting Review, Vol 38, Issue 4, pp. 405-425. DOI 10.1016/j.bar.2006.05.003.

Eklund, T. (2004). The Self-Organizing Map in Financial Benchmarking. Doctoral Dissertation. Åbo Akademi University, Turku Centre for Computer Science. Turku. 156 pages.

Elinkeinoelämän Keskusliitto. (2020.) Suhdannebarometri – Tammikuu 2020. [Online document.] Available: https://ek.fi/wp-content/uploads/SB-tammikuu-2020.pdf

Ivosev, G., Burton, L. & Bonner, R. (2008). Dimensionality Reduction and Visualization in Principal Component Analysis. Analytical Chemistry, Vol 80, Issue 13, pp. 4933-4944. DOI 10.1021/ac800110w.

Kaiser, H. (1974). An Index of Factorial Simplicity. Psychometrika, Vol 39, Issue 1, pp. 31-36. DOI 10.1007/BF02291575.

Kananen, H. & Puolitaival, H. (2019). Tekoäly: Bisneksen uudet työkalut. Helsinki: Alma Talent.

Kauppalehti. (2020). Hiekoittaminen, lumenauraus ja nurmikon leikkaaminen takaavat vakauden – pirstaloitunut kiinteistönhoito keskittyy isoille yrityksille. [Online document.] [Accessed 15.03.2020.] Available: https://www.kauppalehti.fi/uutiset/hiekoittaminen-lumenauraus-ja-nurmikon-leikkaaminen-takaavat-vakauden-pirstaloitunut-kiinteistonhoito-keskittyy-isoille-yrityksille/58e4a1d0-654e-4816-a5bc-a888afe8a3b2

Kiinteistötyönantajat Ry. (2020a). Kansantalouden moottori. [Online document.] [Accessed 07.03.2020.] Available: https://www.kiinteistotyonantajat.fi/kiinteistoala/

Kiinteistötyönanatajat Ry. (2020b). Kiinteistöpalvelut. [Online document.] [Accessed 07.03.2020.] Available: https://www.kiinteistotyonantajat.fi/kiinteistoala/kiinteistopalvelut/

Kohonen, T. (2014). MATLAB Implementations and Applications of the Self-Organizing Map. Helsinki: Unigrafia Oy.

Kohonen, T. (2013). Essentials of the Self-Organizing Map. Neural Networks, 37, pp. 52-65. DOI 10.1016/j.neunet.2012.09.018.

Kohonen, T. (1997). Self-Organizing Maps. Second edition. Leipzig: Springer-Verlag.

Kumpula, P. (2015). Tilinpäätösanalyysi investointisuunnitelmien tukena kohdeyrityksessä. Bachelor’s thesis. Seinäjoki University of Applied Sciences, School of Business and Culture. Seinäjoki. 75 pages.

Leskinen, M. (2009). Kiinteistöalan pk- yrityksen kannattavuus. Bachelor’s thesis. Vaasa University of Applied Sciences, Business Management. Vaasa. 60 pages.

Leskovec, J., Rajaraman, A. & Ullman, J. D. (2020). Mining of Massive Datasets. Third edition. Cambridge: Cambridge University Press.

Lith, P. (2019). Kiinteistöala Suomen kansantaloudessa – Raportti kiinteistöalan yritystoiminnasta, markkinoista ja kehityslinjoista 2018-2019.

Lynn, S. (2014). Self-Organising Maps for Customer Segmentation using R. [Online document.] [Accessed 15.03.2020.] Available: https://www.r-bloggers.com/self-organising-maps-for-customer-segmentation-using-r/

Makkonen, M. (2019). Kiinteistöpalvelujen digitalisaatio Suomessa. Master’s thesis. University of Jyväskylä, Information Systems. Jyväskylä. 71 pages.

Metsämuuronen, J. (2008). Monimuuttujamenetelmien perusteet. Second fixed edition. Jyväskylä: Gummerus.

Pampalk, E. (2001). Limitations of the SOM and the GTM. University of Vienna. Department of Medical Cybernetics and Artificial Intelligence. 11 pages.

Penman, S. H. (2010). Financial Statement Analysis and Security Valuation. Fourth edition. New York: McGraw-Hill.

Rakennusinsinöörien Liitto. (2019). Rakennetun omaisuuden tila 2019. [Online document.] [Accessed 13.03.2020.] Available: https://www.ril.fi/media/2019/roti/roti_2019_raportti.pdf

RAKLI ry. (2014). Kiinteistöalan yhteiskunnallinen ja kansantaloudellinen merkitys. [Online document.] [Accessed 07.03.2020]. Available https://kti.fi/wp-content/uploads/Kiinteistöalan-yhteiskunnallinen-ja-kansantaloudellinen-merkitys-2014.pdf

RAKLI ry. (2012). Kiinteistöliiketoiminnan sanasto, 2. laitos. Helsinki: Sanastokeskus TSK Ry.

Rist, M. & Pizzica, A. J. (2015). Financial Ratios for Executives: How to Assess Company Strength, Fix Problems, and Make Better Decisions. Berkeley, CA: Apress.

Ross, S. A., Westerfield, R. W. & Jordan, B. D. (2008). Corporate Finance Fundamentals. Eighth edition. New York: McGraw-Hill/Irwin.

Salonen, A. (2014). Characteristics of Facility Service Industry and Effects on Buyer-Supplier relationships. Nordic Journal of Surveying and Real Estate Research, 2. Available: https://journal.fi/njs/article/view/41486

Sarstedt, M., & Mooi, E. (2019). A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics. Third edition. Berlin, Heidelberg: Springer Berlin Heidelberg.

Statistics Finland. (2020a). Tietoa tilastoista. Kiinteistönhoito. [Online document.] [Accessed 27.01.2020.] Available: https://www.stat.fi/meta/luokitukset/toimiala/001-2008/81100.html

Statistics Finland. (2020b). Rakenne- ja tilinpäätöstilaston aineisto.

Statistics Finland. (2019.) Building Stock 2018. [Online document]. [Accessed 13.05.2020.] Available: http://www.stat.fi/til/rakke/2018/rakke_2018_2019-05-21_kat_002_en.html

Studentum. (2018). Kiinteistöala. [Online document.] [Accessed 27.01.2020.] Available: https://www.studentum.fi/tietoa-tyoelamasta/tyoelamatietoa-ammattialoista/kiinteistoala-6197

Suhonen, L. (2012). PK-yrityksen tunnuslukuanalyysi ja talouden seuranta. Bachelor's thesis. Kymenlaakso University of Applied Sciences, Business Management. Kouvola. 28 + 2 pages.

Summanen, K. (2012). Kiinteistönhoitoyrityksen kannattavuus. Bachelor's thesis. Kymenlaakso University of Applied Sciences, Business Management. Kouvola. 36 + 3 pages.

Wehrens, R. & Kruisselbrink, J. (2019.) Package 'kohonen'. [Online document.] [Accessed 25.03.2020.] Available: https://cran.r-project.org/web/packages/kohonen/kohonen.pdf

Wendler, T. & Gröttrup, S. (2016). Data Mining with SPSS Modeler: Theory, Exercises and Solutions. Cham: Springer International Publishing.

Vesanto, J. & Alhoniemi, E. (2000). Clustering of the Self-Organizing Map. IEEE Transactions on Neural Networks, Volume 11, Issue 3, pp. 586-60

APPENDICES

Appendix 1 Formulas for the financial ratios

Ratio group		Ratio	Abbreviation	Formula
Solvency	Short-term	Quick ratio	QR	(Financial assets – percentage of compeletion receivables) / (Short-term liabilities – short-term advances received)
		Cash ratio	CAR	Cash / Current liabilities
	Long-term	Equity ratio	ER	Adjusted shareholders' equity / (Adjusted balance sheet total – advances received) X 100
		Relative indebtedness	RI	(Adj. balance sheet's debt – advances received) / Net sales X 100
		Net gearing	NG	(Interest-bearing liabilities – cash and financial sec.) / Adjusted shareholders' equity
		Debt ratio	DR	Total liabilities / Total assets X 100
Efficiency		Asset Turnover	AT	Sales / Total assets
		Receivables turnover	RT	Trade receivables x 365 / Sales
Profitability		Return on assets-%	ROA	(Net result + financial expenses + taxes (12 months)) / Total capital (adjusted average for the financial year) X 100
Performance		Change in net sales-%	CINS	Change in net sales / Net sales in previous financial year X 100
		Net sales per employee	NSPE	Net sales / Average number of personnel

Appendix 2 Code for the R

```
# Analyzing the Finnish Property Maintenance Business Through Data
# PCA and SOM

# References:

# KU Leuven Marketing Department. (2019). R for Marketing Students. Available:
https://bookdown.org/content/1340/
# Lynn, S. (2014a). Palette defined by kohonen package. Available:
https://github.com/shanealynn/Kohonen-Self-organising-maps-in-R/blob/master/coolBlueHotRed.R
# Lynn, S. (2014b). Self-Organising Maps for Customer Segmentation using R. Available:
https://www.r-bloggers.com/self-organising-maps-for-customer-segmentation-using-r/
# Rakotomalala, R. (2017). Tanagra Data Mining. Available: http://eric.univ-
lyon2.fr/~ricco/tanagra/fichiers/en_Tanagra_Kohonen_SOM_R.pdf
# Wehrens, R. & Kruisselbrink, J. (2019). Package "kohonen". Available: https://cran.r-
project.org/web/packages/kohonen/kohonen.pdf

# Developed with R 3.6.3
# Author: Antti Nikkinen 27.04.2020

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

# Clear the R environment
rm(list=ls())

# Call libraries (if needed, use command install.packages("name of package"))
library(radiant) #library for PCA
library(kohonen) #library for SOM
library(PerformanceAnalytics) #library for data analytics tools

# Read data from CSV-file
dataraw <- read.csv("Dataset_mod_ready.csv",header=TRUE,sep=";")

# Check the structure of the data
str(dataraw)

# Check if there are any null-values, NA-values or duplicates
any(is.null(dataraw))
any(is.na(dataraw))
anyDuplicated(dataraw)

# Remove rows with missing values and check the structure
data <- na.omit(dataraw)
str(data)

# Remove unnecessary variables, in this case ID, year, personnel, revenue, ROA and DR
data1 <- data[ -c(1:4, 7, 11) ]

# Correlation analysis for the data
chart.Correlation(data1, histogram=TRUE, pch="+")

# Correlation matrix with two-decimal roundings
round(cor(data1),2)

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
```

```
# Principal Component Analysis PCA (KU Leuven Marketing Department (2019))

# Normalise the data with Z-score method (zero mean and st.dev of 1) and create a dataframe
datapca <- data.frame(scale(data1, center = TRUE, scale = TRUE))

# Assign the names of the variables in a vector
variables <- colnames(datapca)

# Pre-PCA for Bartlett's Test of Sphericity and Kaiser-Meyer-Olkin-test.
pre1 <- pre_factor(datapca, vars = variables)
summary(pre1)

# Remove RT-variable
datapca2 <- datapca[ -c(9) ]
variables2 <- colnames(datapca2)
pre2 <- pre_factor(datapca2, vars = variables2)
summary(pre2)

# Remove AT-variable
datapca3 <- datapca2[ -c(8) ]
variables3 <- colnames(datapca3)
pre3 <- pre_factor(datapca3, vars = variables3)
summary(pre3)

# Scree plot from pre-component analysis
plot(pre3, plots = c("scree"))

# Conduct the PCA with Varimax-rotation
# nr_fact = amount of factors (see previous plot and previous tests to decide number)
# dec = amount of decimals
pca1 <- full_factor(datapca3, vars = variables3, method = "PCA", nr_fact = 3, rotation = "varimax")
summary(pca1, dec = 4)

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

# Self-Organizing Map SOM (Lynn (2014b), Rakotomolala (2017) and Wehrens & Kruisselbrink
(2019))

# Color-function for blueshades (Rakotomalala 2017)
blueshades <- function(n){
  return(rgb(0,0.4,1,alpha=seq(0,1,1/n)))
}

# Color-function for greyshades (Rakotomalala 2017)
greyshades <- function(n){
  return(rgb(0,0,0,alpha=seq(0,1,1/n)))
}

# Color-function for the rainbow colours (Lynn 2014a)
coolBlueHotRed <- function(n, alpha = 1) {
  rainbow(n, end=4/6, alpha=alpha)[n:1]
}

# Data set-up for SOM
datasom <- filter(data, YEAR == 2018)

# Subset for 2018 and REV > 2000000 & <= 10000000
datasom210 <- filter(data, YEAR == 2018, REV > 2000000, REV <= 10000000)
```

```
# Remove unnecessary variables from both datasets
dsom1 <- datasom[ -c(1:4,11,14:15) ]
dsom2 <- datasom210[ -c(1:4,11,14:15) ]

# Normalise data with Z-score (zero mean and st.dev of 1) and create a matrix for both datasets
datasom1 <- as.matrix(scale(dsom1, center = TRUE, scale = TRUE))
datasom2 <- as.matrix(scale(dsom2, center = TRUE, scale = TRUE))

# Define the grid setup
som_grid <- somgrid(xdim = 12, ydim=9, #recommendation xdim = 1.3 * ydim
                    topo="hexagonal",
                    neighbourhood.fct = "gaussian")

# Larger dataset: grid size 12x9, alpha 0,7-0,01 and rlen=1000
# Smaller dataset: grid size 5x4, alpha 0,7-0,01 and rlen=250

som_model <- som(datasom1,
                 grid = som_grid,
                 alpha = c(0.7,0.01), #default 0.5 and 0.01
                 rlen = 1000, #default 100
                 dist.fcts = "euclidean",
                 keep.data = TRUE)

# Plot to assess development of algorithm learning after iterations (minimum plateau should be
visible)
plot(som_model, type="changes")

# Count plot presents observations in each node
plot(som_model, type="count")
plot(som_model,type="count", palette.name=greyshades)
plot(som_model,type="count", palette.name=blueshades)

# Statistics and number of observations assigned to each node
obs <- table(som_model$unit.classif)
mean(obs)
min(obs)
max(obs)
sd(obs)
hist(obs)

# Check if there are empty nodes (result should be the amount of nodes in the grid, xdim * ydim)
length(obs)

# Plot the amount of observations in nodes
plot(som_model, type = "mapping", pchs = 20)

# Plot U-matrix
plot(som_model, type="dist.neighbours", palette.name=blueshades)
plot(som_model, type="dist.neighbours", palette.name=greyshades)
plot(som_model, type="dist.neighbours", palette.name=coolBlueHotRed)
plot(som_model, type="dist.neighbours")

# Plot codes (codebook vectors)
plot(som_model, type="codes", palette.name=blueshades)
plot(som_model, type="codes", palette.name=greyshades)
plot(som_model, type="codes", palette.name=coolBlueHotRed)
```

```
# Plot the quality of the map (mean distance of objects)
plot(som_model, type="quality")
plot(som_model, type="quality", palette.name=blueshades)
plot(som_model, type="quality", palette.name=greyshades)
plot(som_model, type="quality", palette.name=coolBlueHotRed)

# Component plane for first variable in the data
plot(som_model, type = "property", property = getCodes(som_model, 1)[,1],
    main = colnames(getCodes(som_model, 1))[1])

# Component planes for all variables in the data (adjust par(mfrow) for suitable size)
par(mfrow=c(3,3))
for (i in 1:ncol(datasom1)){
  plot(som_model, type = "property", property = getCodes(som_model, 1)[,i],
      main = colnames(getCodes(som_model, 1))[i], palette.name=coolBlueHotRed)
}
par(mfrow=c(1,1))

# Hierarchical clustering with Ward's method

# Distances between objects
dist <- object.distances(som_model, type = "codes")

# Create a dendrogram with Ward’s method
hcw <- hclust(dist, method = "ward.D2")

# Plot the obtained dendrogram for assessment
plot(hcw)

# Use Ward's method to cluster the SOM
# Larger dataset with 8 clusters and subset with 5 clusters
som.hcw <- cutree(hclust(object.distances(som_model, "codes"), method = "ward.D2"), 8)

# Plot SOM with clusters (more color options http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf)
plot(som_model, type="mapping", bgcol=c("slateblue1","springgreen1", "steelblue1", "tan1",
"thistle1", "tomato1", "turquoise1", "violetred1", "wheat1", "yellow1")[som.hcw])
add.cluster.boundaries(som_model, clustering=som.hcw)

# Plot codebook vectors with clusters
plot(som_model, type="codes", palette.name=coolBlueHotRed)
add.cluster.boundaries(som_model, clustering=som.hcw)

# Assign each observation to its cluster
clusters <- som.hcw[som_model$unit.classif]
clusters

# Assign each observation to its node-membership in original data
datasom$NMS <- som_model$unit.classif
datasom210$NMS <- som_model$unit.classif

# Assign each observation to its cluster-membership in original data
datasom$CLMS <- clusters
datasom210$CLMS <- clusters

# Write csv-file to get results on a file and for further examination with R
write.csv(datasom,'nodes_and_clusters_subset.csv')
```