SUMMARY

This thesis studies the intersection of small and medium sized enterprises (SMEs) and data operations consultancy services. The thesis begins with recognizing the transformative impact of industry 4.0 technologies on companies, such as enabling them to enhance their efficiency and competitiveness through data-driven decision making. The theoretical background of this research highlights the importance of SMEs to economic development and how they need to be data-driven in order to stay competitive. It highlights that data consultancy firms can bridge the gap between SMEs and data operations expertise. The thesis suggests using open data in the identification of SMEs in need of consultancy support.

Methodologically, the research outlines the process of data extraction from sources like the Estonian Open Business Registry and the Estonian Tax and Customs Board. It talks about the importance of establishing an efficient database infrastructure using PostgreSQL. It also includes considerations for cloud hosting on platforms like Microsoft Azure. Data preprocessing steps are detailed. This includes handling missing values, duplicates, and standardizing formats to ensure data reliability. The data section provides an overview of the datasets used in the analysis. It offers insights into the structure and the scope of the extracted data. Following this, the data cleaning process is discussed. This section focuses on addressing data quality issues and enhancing readability.

Finally, and most importantly, the criteria development and analysis chapter takes place. This chapter establishes the criteria for identifying SMEs which are ready for data operations. Factors such as revenue trends and workforce size are analyzed in this section. Initially, the dataset included over 350,000 companies. Through a systematic filtering process based on criteria such as, ascending revenue over the last four quarters and workforce size, this number was refined to 659 eligible companies. Our goal was to develop a criterion to determine the companies data readiness, and to find a tangible list of companies that would get the most out of data operations consultancy services out of hundred thousand of companies, while doing analysis of their revenues and workforce. We can conclude that these goals have been achieved.

In conclusion, the thesis offers a systematic approach to target SMEs for data operations consultancy, by using open data sources and advanced analytics. It contributes to both academic knowledge and practical applications by addressing the challenges and opportunities in data-driven decision-making for SMEs in the evolving landscape of Industry 4.0.