

TOOLS AND TECHNOLOGIES UTILIZED IN DATA-RELATED POSITIONS: AN EMPIRICAL STUDY OF JOB ADVERTISEMENTS

NURLAN MUSAZADE

Åbo Akademi University, Faculty of Social Sciences, Business and Economics, and Law
(FSEJ), Turku, Finland
nurlan.musazade@abo.fi

Role, value and amount of data and related tools, technologies, companies and professions in society is rising. Since the required skills for data-related professions are predicted to experience changes, and labor market mismatches create challenges for stakeholders, this research focuses on changes in the required tools and technologies for data-related positions. The presented research defines trends and changes in frequencies of the tools utilized in the data-related professions by applying quantitative content analysis on collected data from job advertisements of Finland, Denmark and Poland. The research findings show that tools used in data-related professions experience significant changes over time. For example, AI and cloud computing-related skills, and SQL started to be required more, whereas Excel, SPSS and similar tools are less expected from the candidates. Furthermore, while R programming language utilization rises in analytics related positions, Python is more common in positions related to data science.

Keywords:

data
analytics,
data
science,
data
professions,
skill
requirements,
labor
market,
job
advertisements



DOI <https://doi.org/10.18690/um.fov.6.2023.10>
ISBN 978-961-286-804-8

1 Introduction

Data volumes have an increasing trend influenced by automation, technological shifts and pandemic (Sheng et al., 2020), which makes it crucial to study big data, related tools and professions. Furthermore, organizations have started to utilize big data, related technologies and Artificial Intelligence (AI) more extensively compelled by pandemic regulations (McKinsey, 2021). As a result, data context and labor market have started to experience significant shifts (McKinsey, 2021). Field experts and related studies indicate that technologies utilized in the workplace for accomplishing tasks, and corresponding required skills, are experiencing changes, and predicted to change in the future on a wider scale (Musazade, 2022, World Economic Forum, 2020). In other words, digitalization drives uncertainty regarding the tools and skills required for professions in the future (Koh, 2020). For example, increasing data volume and unstructured data requires working on new technologies, such as cloud computing (Marr, 2019), and corresponding skills therefore. Some professions, e.g. data entrance clerks, are expected to be completely replaced by emerging technologies (World Economic Forum, 2020). As a result of distribution of some work to machines, the emergence of different professions (e.g. AI specialist) specialized in the new tools and possessing corresponding skills can be observed (World Economic Forum, 2020).

Human resource management in organizations experiences challenges in recruiting to vacant positions because of the existence of skill gaps (Koh, 2020). Furthermore, data-related professions have common definition deficiency, as they are typically specified via personal perspectives of different specialists, which increases the knowledge gap (Miller, 2014, Granville, 2014). This is also a problem sporadically addressed in the literature (Granville, 2014). In other words, there is evident need for the study of data-related professions, their role in organizations, as well as related tools that are utilized by them. Another factor that increases the relevance of this study is the rising trend of data companies (Cattaneo et al., 2020). Furthermore, data volume has a continuous rising trend, which generates the need for usage of advanced tools that are capable of processing big data (Power & Heavin, 2017, Sheng et al., 2020, Kenett & Redman, 2019). While there are some existing studies on this topic, they are focused on specific markets with limited data analysis techniques.

This study concentrates on the Danish, Polish and Finnish labor markets, and aims to answer following research questions:

RQ1: What are the most important skills organizations look for when hiring data professionals?

RQ2: How have the skills required by companies for data-related positions changed in the recent years?

The rest of the paper is structured as follows. Section 2 presents the literature review on the previous similar research in the domain. In Section 3, the research methodology, data and data analysis are described. Section 4 discusses trends and the findings of the study, and Section 5 summarizes the study, and presents contributions and limitations.

2 Literature Review

Data-related positions have been previously analyzed by researchers as listed in Appendix A, typically by means of empirical data collected from job advertisements by web scraping and analyzes have been performed with different techniques of text analytics. For example, the study of De Mauro et al. (2018) shows that big data developer (BDD) and engineer (BDE), business analyst (BA) and data scientist (DS) are related professions in the big data domain. Furthermore, research done by Verma et al. (2019) compares data analyst (DA) and DS, and BA and business intelligence (BI) positions, and concludes that a DS profession requires more technical skills. Yet, in the findings of these two articles there is a mismatch (e.g. in coding skill requirements).

Organization, Excel, SQL, Reporting are the skills that have been found required for the DA profession in multiple studies (Verma et al., 2019, Jiang & Chen, 2021). At the same time, Python and R programming languages, Teamwork and Statistics are the common required skills for DS positions (Verma et al., 2019, Jiang & Chen, 2021). For the BA profession, teamwork, management and project management, testing and analytics are the keywords that studies have identified in job advertisements (De Mauro et al., 2018, Verma et al., 2019, Jiang & Chen, 2021).

Persaud (2020) has analyzed big data analytics-related professions with the aim of defining main in demand competencies and provision of students with them. The study has utilized both quantitative and qualitative research methodology, including text mining of job advertisements and education institutions' courses, and interviewing senior executives. The framework presented consists of cognitive, functional, meta-competence and social competence extensions. Python, SQL, Spark, Hadoop are some of the tools, teamwork, customer relationships, business domain knowledge, communication, presentation skills the research found to be required from candidates. Moreover, for entry-level vacant positions, candidates are required to have a degree from quantitative field, as well as experience in relevant positions. Yet, the research presents general findings without detailed description of the frequencies of the skills.

In a similar research, Smaldone et al. (2022) analyzed job advertisements of the US market. After data preprocessing and text analysis, skills have been categorized in thematic groups. Similar to the study of the Persaud (2020), skills have been classified into different groups, whereas in the research of Persaud (2020) the themes have been compared with the particular theoretical framework. Moreover, Smaldone et al. (2022) queried only the presence of "data science" during data collection. Big data, data processing and AI are the top three frequent terms. Topic modeling has resulted in five topics, including generic skills, hard skills, big data and AI, analytical and team-related skills. Python, R, data analysis, deep learning, communication, teamwork, problem-solving, data management and data mining are some of the skills that research has found existing in job advertisements. Fang and Zhou (2021) have analyzed DA positions in China's market via data collection from recruitment websites and applying text mining. Data have been collected during 2017. Among the programming languages, Python, R and Java have been found as most utilized in DA positions, whereas authors underline R as the most demanded language. Furthermore, Hadoop, Spark, SQL, Hive, Oracle, SPSS, MATLAB, TABLEAU and SAS are other tools that have been found having high frequency in the job postings, whereas the highest is Excel among the tools that serve for "visualization of statistical analysis" or "application" (i.e. latter four tools) as defined by authors. Yet, soft skills and other domain-related terms are completely missing in the study.

Another research, in which data have been collected from Chinese websites, is conducted by Cao and Zhang (2021). The search query has been limited to the “data mining” and “data science” keywords. In the research named entity recognition, as well as Bert-BILSTM-CRF model has been built and trained for extraction of information or named entities. Based on the findings, Computer Science, Statistics and Mathematics are the most required and co-occurred majors by recruiters, based on the data collected in 2020. Yet, majors such as Economic, Finance, Management, Electronics, Information Management, Physics also exist in the requirement list of some advertisements, which mostly co-occur with the mentioned top three majors. Findings indicate that communication, teamwork, logical thinking and responsibility, as well as SQL, Python, R, SPSS, Excel tools are having highest frequencies in the job postings. Furthermore, authors have categorized positions into three groups, which are big data mining, data analysis and data administration, data based on the findings in advertisements, that differ based on the programming languages and softwares (e.g., for data processing and management).

The studies present considerably differing findings. To compare findings, SPSS has only around 4% and SAS 11% frequency in the study of Verma et al. (2019) for the DA profession, whereas they are in the most required skill list in the study of Fang and Zhou (2021). Moreover, different from other studies, Excel keyword lacks in presented findings of Smaldone et al. (2022) and Fang and Zhou (2020). Although two studies, in which data have been collected from China’s job advertisements, have analyzed similar positions, the search query and the data collection years have been different. Yet, all the skills that are described in the study of Fang and Zhou (2021) are mentioned in the study of Cao and Zhang (2021), except Matlab and Tableau. Non-existence of precise frequencies in the studies unable to make accurate observations on the change of or variations in skill requirements.

As a summary, while previous studies have concentrated on current labor market requirements, this study expands literature by focusing on additional different perspectives. Among others, lack of accurate information on frequencies of skills, focus mostly on US and China’s markets, existence of mismatches among the previous studies’ findings have been the rationale of this study.

3 Methodology

3.1 Content analysis and data collection

In total, eight different data-related professions have been studied in the research, as shown in Figure 1.

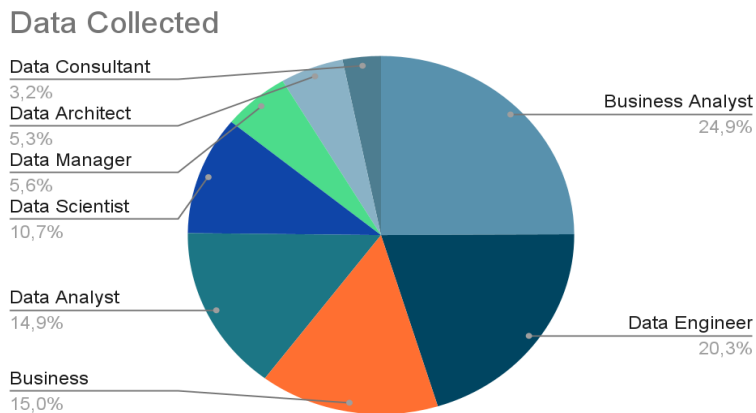


Figure 1: Studied positions

Query and data have been limited to the existence of the term in the title of vacancy advertisements. The titles have been selected by referring to the previous studies, Future of Jobs Report (World Economic Forum, 2020) and LinkedIn's (2020) career explorer portal. Derived from the domain problems that are listed above, including lack of common definition, personal interpretations in the literature and similarity of the professions, frequency technique of quantitative content analysis has been utilized for conducting the research.

Quantitative content analysis has enabled generalization of the findings and elimination of possible biases, caused by subjective interpretations of the positions by field experts, of qualitative analysis (White & Marsh, 2006, Macnamara, 2018). Moreover, the research has been designed to find changes in required skills and utilized tools by differentiating study results with the previous research, and authors

of previous research utilized content analysis and job advertisements as a methodology and research data, correspondingly.

Factors, such as permission requirement for web scraping, usage commonness or number of advertisements, technical feasibility or architecture simplicity of the website, have been decisive for choosing Indeed.com job portal as a data source. Data have been collected between November 2021 and January 2022 by utilizing Python programming language and BeautifulSoup library (i.e. web scraping). Common libraries, such as “requests”, “numpy” and “pandas”, have been deployed during data collection, preparation and processing. In total, 2658 advertisements of Denmark, Finland and Poland have been extracted.

3.2 Data Analysis

As shown in Figure 2, data analysis has been conducted in three interrelated stages based on the research questions and objective. Skills and tools have been collected from previous studies. Among eight studied positions, six of the titles have matched with the titles from the previous research. Subsequently, Spacy library (i.e. Matcher class) has been applied to the collected data for retrieving frequencies of the tools and skills in the job advertisements.

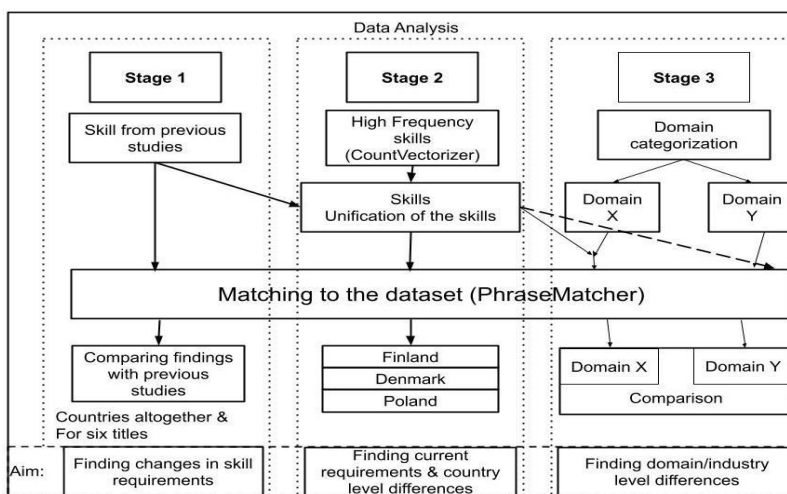


Figure 2: Data Analysis Process

4 Discussion and Result

The study focuses on understanding the required skills for data-related professions, changes in required skills, as well as actual skill requirements in the market. The results show that utilized tools have changed considerably. For instance, AI, cloud computing, Hadoop and SQL skills requirements started to rise, whereas Excel, SPSS, SAS and similar tools less expected from the candidates. Moreover, Python is more common in data science related professions, while R programming language usage increases in analytics related positions.

4.1 Falling trends

An analysis of job advertisements shows that the demand for some skills has decreased. For instance, Microsoft Excel has become less required for data professionals, particularly for DA and BI occupations. In contrast, most of the previous studies indicate Excel as one of the most required tools (e.g., Jiang & Chen, 2021; Verma et al., 2019; Cao & Zhang, 2021; Fang & Zhou, 2021). Based on research, some tools and programs, including Microsoft Word or Office, SPSS, and SAS, are being utilized less frequently in data-related jobs, as evidenced by a decrease in their requirement frequencies in job vacancies. For instance, in the study by Verma et al. (2019), SPSS has a frequency of around 4% for DA positions, while in this study it has a frequency of around 1%. Similarly, SAS has a frequency of 11% in Verma et al.'s study, and around 5% in this study. However, Fang and Zhou (2021), who studied China's market, found these tools to be among the most required for DA positions. A similar trend is applicable for Microsoft Office and PPT.

4.2 Rising trends

In contrast, the requirement for some skills and tools has shown an increasing trend. The matching trend to the utilization of cloud computing by data professionals is the rise of usage of keyword “cloud” in job advertisements, in particular for data engineer (DE) and DAR professions. Not only the “cloud” keyword, but also specific cloud-based database management systems, e.g., Azure and AWS, have increased their presence in the job advertisements, in particular for the DAR and DE positions.

The findings show that AI, cloud, deep learning, Mathematica, R, and Python have started to appear in one or more data-related vacancies, which is different from all of the previous studies except for the most recent US market study by Smaldone et al. (2022). In the study by Smaldone et al. (2022), "artificial intelligence" is the third most commonly appearing keyword after "big data" and "data processing," whereas it appears in only around one-third of advertisements in this study. While "deep learning" in Smaldone et al.'s study has a higher frequency than Python, in this study, around 20% of advertisements include "deep learning" compared to 87% for Python. The same trend applies to the term "machine learning." A possible interpretation of this differentiation is that tasks directly related to big data and AI have become the main duties and requirements for data scientists in the US market in recent years.

In addition, Power BI and Tableau have become required and utilized more frequently in DA, BA and BI positions. Fog computing's components and big data tools Spark and Hadoop have started to be utilized more as well, especially in DS, DE and DAR positions. Persaud (2020) lists Spark and Hadoop as one of the most required tools in the data related professions. Increase in the utilization of Hadoop may also be a sign of more extensive application of predictive analytics by organizations, since it is one of the enabling tools of predictive analytics (Pathak et al., 2018). The latter hypothesis is compatible with Power and Heavins' (2017) findings, in which they identify more incorporation of predictive analytics into organizational operations and decision makings.

4.3 Differing trends of programming languages (R, Python, SQL)

There are tools, in particular programming languages, that have become more required for some professions, while less demanded in other data-related positions. To elaborate, Python programming language has become more required in DS and DE positions; R programming language has become more in demand in BI and DA positions; Structured Query Language (SQL) requirement has a rising trend in DS, DE and DA positions.

Jiang and Chen (2021) have found R in one in two job postings, whereas Python in three of four job postings titled as a "data scientist". In China's job advertisements, for DA positions R is more required than Python (Fang and Zhou, 2021), whereas

job postings with “data science” or “data mining” keywords have higher frequency for Python than R (Cao & Zhang, 2021). In other words, inferences match with the findings of Musazade (2022).

Findings of Smaldone et al. (2022) and Verma et al. (2019) shows that R programming language appears more in advertisements than Python. For instance, in the study of Verma et al. (2019) R is higher in the Statistics skill category, whereas Python is higher in the Programming skill category of data scientist positions. Topic modelling of Smaldone et al. (2022) shows that R more frequently appears with the analytics, data management and business intelligence keywords, whereas Python more appears with the keywords such as data mining and processing, or artificial intelligence.

4.4 Background requirement/specialization

In addition to the particular tools, the study shows that “technical”, “engineering” and “computer science” terms have started to appear less in vacancies of data related positions compared to previous studies (e.g., Verma et al., 2019, Jiang & Chen, 2021), which may imply a diminishing role of technical background in data professionals. For instance, for DE positions “computer science” term has 22% frequency, whereas in the study of Jiang and Chen (2021) the term appears in half of the advertisements. The study of Cao and Zhang (2021) has defined “computer science” and other related terms (e.g., mathematics, statistics) as one of the required background majors from candidates. Keyword “statistics” for the DS profession has started to appear 10-15% less: it has around 60% frequency in study of Verma et al. (2019) and Cao and Zhang (2021). Moreover, the study suggests that data professionals have started to be required to possess a narrower competence and specialization in particular common, unique and extensive tools and technologies. For instance, different from findings of previous research, Python and Python libraries (e.g. Pytorch) have started to occur more frequently in advertisements. Yet, for a business intelligence role, for instance, the same judgment cannot be applied, in which tools utilized are distributed more broadly and shifts are more thorough.

4.5 Grouping of professions

Based on the findings on required tools some professions can be categorized. Firstly, DAR, consultant and partly DE positions involve extensive labor on databases and different cloud platforms. Jiang and Chen's (2021) findings are similar with the findings of this study, except instead of a data consultant position, data scientist profession is a part of the group. Secondly, professions such as DA and BI frequently use BI, data visualization and analytics products, such as Tableau or Power BI. Thirdly, SQL and Python programming languages are frequently used in DE, DS and DA positions. Findings indicate that BI and DA professions are similar to each other in terms of skills they require. Verma et al. (2019) analyze the existence of similarity between DS and DA positions, and between BI and BA professions, whereas the study lacks comparison of positions all together. Finally, both findings of this research and Jiang and Chen (2021) indicate the distinctiveness of the BA profession, which may not be fully matched to a particular category.

5 Conclusions

The research findings show that the skills required and tools utilized in data-related professions have experienced substantial changes in the last two-four years. As the presented results address RQ1, while some tools have common trends in most or all of the data-related professions, differences in frequencies of skills and tools, and of particular tools' trends may enable differentiation between similar data-related professions. Furthermore, regarding RQ2, this study shows that skill requirements for data related positions have changed in the last two-four years, which may imply the change of tools and technologies used in data-related positions, as well as usage objectives of big data by organizations. For example, since Python and machine learning are highly correlated in job ads (Smaldone., 2022), the rising requirement for Python may suggest the rising usage of big data for machine learning. Falling trend of most of the tools belonging to the statistics category defined by Verma et al. (2019), such as Excel, SAS, SPSS may imply less usage of big data for statistical analysis only.

More specifically, cloud computing tools, AI and deep learning skills and knowledge have started to be required more from candidates. SQL, Power BI, R and Tableau have become more common in the job postings for analytics related professions.

Moreover, SQL, Spark, Hadoop and Python have become more in demand for data scientist and engineer professions. However, Microsoft Word or Office, SPSS, SAS, Excel have started to be utilized less in data related positions. Although comparison of finding with the previous studies' results and of results of previous studies outline existence of differences in required tools (e.g. Excel), some common observations can be derived. For example, SQL both in this study and in previous studies is the most required database management tool. While Python is mostly used in data scientist and related positions, data analyst and similar professions utilize R more than Python.

Firstly, the research contributes to the literature by expanding study of data-related professions to the new geographical area and studying the labor market of three countries: Denmark, Finland and Poland. Secondly, the study not only defines current market requirements, but also studies and defines an approach for exploring changes in the market. Thirdly, findings of the research may assist authors in accurate definition of data-related professions based on the skill required for them. Finally, our results can also help in recognizing the importance of the existence of human resources and skills, in addition to the cost, scalability and functionality, security and availability, as the component of multi-criteria decision-making task of tool and system selection (Kachaoui & Belangour, 2019, Grandhi & Wibowo, 2018).

Findings of the research may contribute to the closing skill gap, re-skilling, correctly defining data-related positions by the job advertisers, specialization for a particular data-related profession, search of relevant vacancies based on the possessed skills by candidates, as well as up-skilling for the professions in the field. Furthermore, education institutions may utilize findings in defining curriculum for the studied specializations based on the market requirements and trends.

Shifts in the skill requirements can be influenced by country-level variations, considering previous studies have been conducted in other markets. Frequency of the words may be influenced by the terms in the language of the advertisements. Moreover, real expectations of recruiters may be represented incompletely or inaccurate in the job postings. Number of collected data for some professions may be considered insufficient. Finally, utilization only quantitative analysis may inhibit possible interpretation of terms, in particular soft skills, and conjunctions among terms have been dismissed.

Future research may study possessed (i.e. supplied) skills for refining the existing skill gap. Secondly, studying each high in demand tool and relationships between them, may present insights on usage objectives of the tools and possible need for a shift between them. Thirdly, categorization of job advertisements based on the position level, company size and different industries may depict a clear picture of the current required skills for each specialization, as well as possible differences depending on listed or other variables. This paper focused on a limited number of professions, whereas the number studied positions can be extended to other professions.

References

- Cao, L., & Zhang, J. (2021, April 1). Skill Requirements Analysis for Data Analysts Based on Named Entities Recognition. *IEEE Xplore*. <https://doi.org/10.1109/ICBDIE52740.2021.00023>
- Cattaneo, G., Micheletti, G., Glennon, M., La Croce, C., & Mitta, C. (2020). The European data market monitoring tool : key facts & figures, first policy conclusions, data landscape and quantified stories : d2.9 final study report. Publications Office of the European Union.
- De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, 54(5), 807–817. <https://doi.org/10.1016/j.ipm.2017.05.004>
- Fang, F., & Zhou, Y. (2021). A Study on Recruitment of Data Analyst Based on Text Mining and Visualization Technology. *Journal of Physics: Conference Series*, 1952(4), 042017. <https://doi.org/10.1088/1742-6596/1952/4/042017>
- Gardiner, A., Aasheim, C., Rutner, P., & Williams, S. (2018). Skill Requirements in Big Data: A Content Analysis of Job Advertisements. *Journal of Computer Information Systems*, 58(4), 374–384. <https://doi.org/10.1080/08874417.2017.1289354>
- Grandhi, S., & Wibowo, S. (2018). A Multi-criteria Group Decision Making Method for Selecting Big Data Visualization Tools. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 10(1-8), 67–72. Retrieved from <https://jtec.utm.edu.my/jtec/article/view/3737>
- Granville, V. (2014). *Developing analytic talent: Becoming a data scientist*. John Wiley & Sons, Inc.
- Jiang, H., & Chen, C. (2021). Data Science Skills and Graduate Certificates: A Quantitative Text Analysis. *Journal of Computer Information Systems*, 1–17. <https://doi.org/10.1080/08874417.2020.1852628>
- Kachaoui, J., & Belangour, A. (2019). A Multi-criteria Group Decision Making Method for Big Data Storage Selection. *Networked Systems*, 381–386. https://doi.org/10.1007/978-3-030-31277-0_25
- Kenett, R., & Redman, T. C. (2019). *The real work of data science : turning data into information, better decisions, and stronger organizations*. Wiley.
- Koh, A. (2020). Anticipating and preparing for the future—one example from higher education: The Singapore Management University (SMU) experience. In B. Panth & R. Maclean (Eds.), *Anticipating and Preparing for Emerging Skills and Jobs: Key Issues, Concerns, and Prospects* (pp. 41–46). Springer. https://doi.org/10.1007/978-981-15-7018-6_6
- LinkedIn. (2020). <https://linkedin.github.io/career-explorer/>. <https://linkedin.github.io/career-explorer/> 88

- Macnamara, J. (2018). 11. Content Analysis. *Mediated Communication*, 191–212. <https://doi.org/10.1515/9783110481129-012> Retrieved from: https://www.researchgate.net/publication/327910121_Content_Analysis
- Marr, B. (2019). *Artificial intelligence in practice: How 50 successful companies used artificial intelligence to solve problems (First Edition)*. Wiley.
- McKinsey. (2021). *The future of work after COVID-19*. Retrieved October 23, 2021, from <https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-workafter-covid-19>
- Miller, S. (2014). Collaborative approaches needed to close the big data skills gap. *Journal of Organization Design*, 3(1), 26–30. <https://doi.org/10.7146/jod.9823> 89
- Musazade, N. (2022). Understanding the relevant skills for data analytics-related positions: An empirical study of job advertisements. *www.doria.fi*. <https://urn.fi/URN:NBN:fi-fe2022031323297>
- Pathak, A. R., Pandey, M., & Rautaray, S. (2018). Construing the big data based on taxonomy, analytics and approaches. *Iran Journal of Computer Science*, 1(4), 237–259. <https://doi.org/10.1007/s42044-018-0024-3>
- Persaud, A. (2020). Key competencies for big data analytics professions: a multimethod study. *Information Technology & People*, ahead-of-print(ahead-of-print). <https://doi.org/10.1108/itp-06-2019-0290>
- Power, D. J., & Heavin, C. (2017). *Decision support, analytics, and business intelligence*. Business Expert Press.
- Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2020). Covid-19 pandemic in the new era of big data analytics: Methodological innovations and future research directions. *British Journal of Management*, 32(4), 1164–1183. <https://doi.org/10.1111/1467-8551.12441>
- Smaldone, F., Ippolito, A., Lager, J., & Pellicano, M. (2022). Employability skills: Profiling data scientists in the digital labour market. *European Management Journal*, 40(5). <https://doi.org/10.1016/j.emj.2022.05.005>
- Verma, A., Yurov, K. M., Lane, P. L., & Yurova, Y. V. (2019). An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements. *Journal of Education for Business*, 94(4), 243–250. <https://doi.org/10.1080/08832323.2018.1520685>
- White, M. D., & Marsh, E. E. (2006). Content Analysis: A Flexible Methodology. *Library Trends*, 55(1), 22–45. <https://doi.org/10.1353/lib.2006.0053>
- World Economic Forum. (2020). *The Future of Jobs Report 2020*. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

Appendix A: Literature Review of Data-related Positions

Study - Scope or "Title"	Methodology (tools)	Study focus or findings	Frequent tools
Smaldone et al. (2022) - DS	Web Scraping (Octoparse), Text Mining (R), LDA, Total frequency (TF), Network analysis	Thematic areas for skills, correlated skills, top skills	Python, R
Verma et al. (2019) - DS, DA, BI, BA	Experts for skills' categories, Content analysis, Web scraping (R)	Skills and their categories' comparative importance, comparison of professions	SAS, R, Python, SQL, Excel
Persaud (2020) -Big DA-related professions	Web crawling, text analysis (job postings, university programs), Interviews, semantic relationships (Leximancer)	Domain knowledge, kinds of skills are expected, e.g., social, functional, cognitive, technical.	Hadoop, sql, python, spark
Jiang & Chen (2021) - DA, BA, DAR, DE, DS, DS Researcher	Text analysis with the created list of dictionary, total frequency (Python SpaCy)	Overlapping skills and grouping of professions, match of skills of graduate certificated with the labor market requirements, comparison of top skills of professions	SQL, Excel, R, Python

Fang & Zhou (2021) - DA	Crawling (Bazhuayu), Text mining (dictionary) (Python -Jieba)	Top skills	R, Python, Java, MySQL, Hive, Spark, Oracle, MATLAB, Hadoop, SPSS, Tableau, SAS
Cao & Zhang (2021) - “data science”, “data mining”	Web crawler, BERT - BiLSTM -CRF for NER (Python)	Required majors and their cooccurance, soft and hard skills, kinds and comparisons of positions.	Excel, SQL, SPSS, Python, R, hive, PPT, SPARK, Haddop, Java, SAS;
De Mauro et al., (2018) - “Big Data”	Web Scraping (Portia), Text mining and expert judgment, LDA (R)	Identifying and comparing job families, skill sets and their match with each job family	Hadoop, SQL, Azure, AWS, Python, SQL, Java, Ruby, R, SAS, MATLAB
Gardiner et al. (2018) - “big data”	Text analysis (R), Consensus pile-sort protocol with domain experts	Defining frequent terms and studying term or skills categorization per concept, presenting developed framework	Hadoop, Java;