EMPLOYABILITY COMPETENCIES NEEDED BY DATA ANALYTICS GRADUATES: AN ANALYSIS OF ONLINE JOB LISTINGS

W. Coetzee*

https://orcid.org/0000-0002-7418-4581

R. Goede*

https://orcid.org/0000-0001-7255-465X

*Unit for Data Science and Computing
North-West University
Potchefstroom, South Africa

ABSTRACT

With the inception of the fourth industrial revolution, the world has undergone significant changes in terms of data availability and analysis. With the vast amounts of data available and the need for businesses to stay profitable in trying times, there is an ever-increasing requirement for data-driven solutions. Finding suitable data analytics practitioners with the knowledge and skill sets to meet these demands is difficult, posing a challenge to universities to keep up with the demands. South Africa is one of the worst-performing countries regarding mathematics standards at school level, which presents a challenge for its universities to provide employable data scientists for the market. This article identifies the competencies required of data analytics practitioners by companies operating in South Africa. This will assist universities in identifying the gaps between what they offer and what is required by the industry, enabling them to implement much-needed changes to produce graduates who can become world-class data professionals. Data were gathered from online job recruiting advertisements for data analysts, business analysts, data scientists, data engineers, and statisticians. The data were analysed using a three-step pluralistic method by which a qualitative method was used to determine broad themes and identify initial skills. These were combined with skills identified in the literature to perform text mining, which led to the creation of a structured dataset that was analysed. The most sought-after skills cover artificial intelligence, statistics, mathematics, and programming (hard skills), SQL, Python, Excel, and Power BI (data analytics software skills), and interpersonal, intrapersonal, business, analytical, communication and creativity (soft skills). Universities need to incorporate these skills in their academic programmes and inform students about the significance of developing these skills.

Keywords: data scientist, employability, hard skills, pluralistic method, skills gap, soft skills, South Africa, data analytics software skills, text mining.

INTRODUCTION

The advanced technologies of the fourth industrial revolution (4IR) have refashioned conventional technical processes into automated ones and triggered a digital transformation. Technologies such as advanced cloud computing, big data analytics and machine learning empower companies to improve decision-making in a highly competitive world (Aujla, Prodan, and Rawat 2022, 639). These businesses must be able to collect, process, aggregate and analyse large quantities of data, and draw meaningful conclusions from them to make wise business decisions.

Since these technologies constantly improve and new ones emerge, it will not be sufficient to graduate in the field with a particular set of skills without subsequently and purposefully engaging in lifelong learning. The fast pace at which technologies change poses a problem for universities struggling to adapt quickly to the industry's needs. Degrees alone will not be sufficient to guarantee success in the workplace; additional skills will need to be acquired.

Unemployment in South Africa is high: during the third quarter of 2022, it was 32.9 per cent of the labour force and the average from 2000 to 2022 was 26.7 per cent (Statistics South Africa 2022). Many vacancies exist but are mainly for highly qualified people with the appropriate skill sets. Adverts often state that several years of job experience are required, hence even well-qualified graduates find themselves unemployed – particularly in developing nations – due to a lack of job-related skills (Ilori and Ajagunna 2020, 3). During the third quarter of 2022, 2.7 per cent of the 7.7 million unemployed people in South Africa were graduates (Statistics South Africa 2022), leaving over 200 000 with shattered job dreams. Even data analytics graduates sometimes struggle and settle for jobs like teaching or selling funeral policies. Unemployed graduates feel disillusioned and hopeless, which harms their well-being (van Lill and Bakker 2022, 130). Current South African governmental and nongovernmental initiatives that promote work-related skills are not sufficient; more effective interventions are needed (Burnett 2014, 202).

A study on the employability of ICT graduates in South Africa found that a lack of necessary skills affects their ability to find employment; it recommends that universities align their curricula with the skills required by employers (Ohei and Brink 2019, 13500).

Preparing graduates for the changing digital environment is a global challenge (Ohei and Brink 2019, 13504). According to Ilori and Ajagunna (2020, 4), it is essential for universities to provide a focused education and be willing to adjust their curricula and teaching strategies to prepare students for job opportunities in the 4IR.

Since industry is the ultimate client of the product (graduates) of higher education (HE), their needs should be taken into account when tertiary institutions design courses for their

students. Ulrich (2003, 333), the "father" of critical systems heuristics, stresses the importance that all stakeholders who are affected by a system should have a role in the decision-making process. It therefore makes sense to learn from industry what their needs are to inform the improvement of the curricula for data analytics students, thereby hoping to increase the employability of future graduates. The purpose of the study reported here is to determine what skillsets employers are looking for in data analytics practitioners (data and business analysts, data scientists, data engineers, and statisticians), as determined by analysing online job listings.

Such a study provides valuable guidelines to universities on how to re-shape their programmes to better equip students to be employable in these fields and meet the ever-increasing demands of industry and society. Also, to enrich the holistic approach to problem solving promoted by critical systems thinking, the study not only focuses on technologies that students should master, but also on the hard and soft skills required.

Several studies on the employability of data scientists have been conducted internationally, for example, Stanton and Stanton (2020), Mills, Chudoba, and Olsen (2016), Smaldone et al. (2022) and Fettach, Ghogho, and Benatallah (2022). Similar research conducted in South Africa by Mabe and Bwalya (2022) addressed soft skills that are needed for information and knowledge management practitioners; Ohei and Brink (2019) researched the employability of ICT graduates; and Brink, Abiodun, and Ohei (2019) conducted a study to determine strategies for using ICT skills for sustainable youth employability.

However, the authors could find no studies specifically pertaining to the skills required by employers of data analytics practitioners in South Africa. This gap in knowledge indicates the value of the study reported here to determine how our universities should adapt their data analytics programmes to allow our industries to become more competitive in this digital age. Another feature of this study is the use of different research methods (both from the interpretivist and the positivistic paradigms), which, according to Mingers (2001, 240), lead to richer results. The following research questions were posed:

Q1: Which data analytics software skills are required by employers?
Q2: Which hard skills are required by employers?
Q3: Which soft skills are required by employers?
Q4: Which degrees do employers require?
Q5: Do the skills required differ for different job titles?
Q6: How many years of experience do employers require?

In the literature study, the need for data analytics professionals is explored and the concepts of

employability and employability skills are explained. This section also provides some foundational knowledge of critical systems thinking as this study is part of a larger research project guided by critical systems thinking (CST). In the section dealing with the research method, the pluralistic methods (both qualitative and quantitative analysis) used are presented; whereafter the results of the analysis conducted are provided. Finally, the discussion of the findings, conclusion, and limitations of the research are presented.

LITERATURE STUDY

The need for data analytics professionals

The growing digitalisation of the global economy, together with the surge in the use of the internet and social media, has led to data overkill, where an estimated 328.77 million terabytes of data are created daily (Duarte 2023). However, data on their own have no meaning. They have to be transformed into information that can inform decision makers (Stanton and Stanton 2020, 139). This is done by big data analytics – and the people needed to make this possible, are data analytics practitioners: data analysts, business analysts, data scientists, data engineers and statisticians. However, these individuals are expected to leverage data – often heterogeneous in nature, merged from disparate data silos – to improve companies' growth, competitiveness, and sustainability, which is no easy task. It is therefore no wonder that it is difficult to find graduates with the skills and abilities needed for these challenges (Stanton and Stanton 2020, 139).

These data practitioners find themselves in an interdisciplinary field that includes computer science, statistics, and knowledge pertaining to a specific discipline or business (Smaldone et al. 2022, 672). The rapid expansion of technological infrastructure where terabytes of data can be analysed, the advances in data storage and the unprecedented growth of analytical tools have all contributed to the explosion in the demand for data scientists (Mills et al. 2016, 131).

Four pillars of analytics were identified by Kang, Holden and Yu (2015) in Mills et al. (2016, 133), who listed the skills required for each. The pillars are a) data storage, retrieval, and pre-processing; b) data exploration; c) analytical models and algorithms; and d) data products. Some of the proposed skills for each of the pillars are a) data warehousing, parallel computing, and NoSQL (i.e., non-tabular databases that don't store data in relational tables); b) statistical analysis and visualisation; c) machine learning, data mining and natural language processing; and d) data and information organisation, knowledge representation and application development.

Another framework was proposed by Anderson et al. (2014, 147) that comprises eight areas, namely, a) handling large data sets; b) databases that include the design, storage and query of the databases; c) AI, which includes genetic algorithms, neural networks, machine learning, and pattern matching; d) software, algorithms and programming; e) information retrieval using data and text mining; f) mathematics that incorporates logic, statistics, modelling, and simulations; g) oral and written communication; and h) social, ethical and legal issues.

From the wide variety of areas addressed by this framework it should be evident that graduates in this field need numerous skills to cope with the challenges they can be presented with in the workplace. By looking at various aspects of the employability concept, one may better understand how HE may improve the programmes they offer to data analytics students. Employability is therefore discussed next.

Employability

According to Yorke (2006, 2), employability can be seen as a set of achievements which represent a necessary but not sufficient condition for obtaining employment. Furthermore, they also reflect a person's capacity to function in a job. Employability is therefore complex, since it goes well beyond the simplistic idea of possessing a few key skills. Yorke (2006, 7) defines employability as

> "a set of achievements – skills, understandings and personal attributes – that make graduates more likely to gain employment and be successful in their chosen occupations, which benefits themselves, the workforce, the community and the economy". (Yorke 2006).

Our study supports the more elaborate definition provided by Smaldone et al. (2022, 673):

> "Employability is herein defined as an individual's portfolio of previously acquired knowledge, skills, attitudes, competencies, experiences, and other qualifications that underpin their ability to be a reliable source of efficiency, innovation, and productivity to an employer."

Employability is a construct that can be used to benchmark an individual's probability of being employed in a certain context.

Employability skills

To be highly employable in the field of data analytics, a person needs to be well equipped in a wide range of qualities. Not all researchers agree on the dimensions that should be used. Smaldone et al. (2022, 674) divides employability skills into hard skills and soft skills, where an employability skill is an asset held by an individual that meets a requirement of employers.

Hard skills can be seen as practical competencies that can be gained through study and experience. It is possible to assess and accredit hard skills (Smaldone et al. 2022, 674). The hard skills required depend on the job. While data practitioners will need some generic hard skills such as extracting data, cleaning data, statistical methods, and modelling tools, there may also be other hard skills required that depend on the industry where the person will be employed – for example, commanding expertise in e-commerce, retail, or ability to model biological data. Some of the hard skills identified include mathematical modelling, forecasting, applied probability, optimisation, statistics, and economics (Schoenherr and Speier-Pero 2015, 124), machine learning, data mining, and programming languages (Kim and Lee 2016, 8166).

Soft skills are defined as "skills, abilities and traits that pertain to personality, attitude and behaviour rather than to formal or technical knowledge" (Moss 1996, 253). They are portrayed in the manner employees are able to manage themselves, their emotions, and others (Hurrell, Scholarios, and Thompson 2013, 8162). Since data scientists are often required to work in teams and present their findings to management or clients, soft skills such as being able to collaborate and to communicate through written reports and verbal presentations, are important (Stanton and Stanton 2020, 141). In 4IR where automation is becoming the norm, soft skills are growing increasingly important, because it is the personal touch that distinguishes a person from a machine (such as in AI and robots) and these personal attributes are increasingly valued in the market (Smaldone et al. 2022, 675).

According to (Heine (2023), there are several reasons why soft skills are considered important in the workplace. For instance, skills such as commitment and motivation can indicate an employee's longevity in a company. Likewise, communication and conflict resolution skills can enhance teamwork, while good interpersonal skills can help build and maintain professional networks. Additionally, organisational skills, attention to detail, time management and the ability to delegate tasks are all crucial in meeting deadlines.

Although there is no consensus among researchers on how soft skills can be sub-divided, similar ideas about what soft skills are, have emerged from the literature. Rawboon et al. (2021, 1) identifies three categories, namely, social, personal, and methodological.

Social skills indicate a person's ability and willingness to cooperate and communicate with others. Personal skills reflect an individual's values, motivation, and attitude. Methodological skills are needed for problem solving and decision making.

According to Caeiro-Rodriguez et al. (2021, 29224), soft skills can be grouped into five categories: technical, metacognitive, intrapersonal, interpersonal, and problem-solving skills. Technical soft skills include basic digital literacy, information and media literacy, health and wellness literacy, economic and financial literacy, ethics, and global awareness. Metacognitive

soft skills comprise autonomous learning, willingness to learn, integrating information, evaluating information through critical thinking, and high-level and innovative thinking. Intrapersonal soft skills consist of open-mindedness, creativity, flexibility and adaptability, openness to criticism, receptive to others' ideas and thoughts, taking initiative, perseverance, self-direction, self-discipline, planning, ability to prioritise, assertiveness, having a positive outlook, and the ability to assess the quality of work done. Interpersonal skills include social interaction and demonstrating empathy, being able to listen, leadership, oral and written communication, ability to apply knowledge in the real world, and presentation skills. Problem-solving soft skills encompass the understanding of a problem, the ability to identify factors contributing to the problem, following systematic thinking, evaluating potential solutions, assessing the effectiveness of a solution, working with limited resources, time management, and project management.

In the data science realm there is an additional skillset that is important: Hale (2018, 4) identified data analytics software skills (programming languages, libraries and tools) as asked for in online job listings. The top ten skills required were mastery of Python, R, SQL, Hadoop, Spark, Java, SAS, Tableau, Hive and Scala (Hale 2018).

For clarity purposes, all the data analytics software that is referred to in this article, together with a brief explanation of each, are presented in Table 1.

Table 1: Explanation of the various data analytics software mentioned in the article.

Data analytics software	Explanation (Similar tools are grouped together)
SQL	Structure Query Language: programming language used to store and process information in a relational database.
NoSQL	Not only SQL: non-tabular databases, storing data in ways other than relational databases.
Access	A database management system that unites the relational Access database engine with a graphical user interface and software development tools.
Excel	Spreadsheets are used to store, sort, manipulate and visualise data.
Power BI	A scalable platform used for business intelligence.
Tableau	A visual analytics platform used for business intelligence.
Google Analytics	Google Analytics is a web analytics service. It provides basic analytical tools for search engine optimization and marketing purposes by tracking website performance and collecting visitor insights.
Python	An object-oriented programming language.
R	A programming language used for statistical computing and graphics.
SAS	A Statistical Analysis System (SAS) is used for information retrieval, data management, data visualisation and modelling.
Java	A programming language for creating web applications.
Scala	A high-level general-purpose programming language.
Hadoop	Manages the storage and parallel processing of big data.
Spark	A distributed processing system used for big data.
Hive	A data warehouse system enabling analytics at a large scale.
ETL	A process that Extracts, Transforms, and Loads data from multiple sources to a data warehouse or data repository.

Data analytics software	Explanation (Similar tools are grouped together)
SSIS	Microsoft SQL Server Integration Services (SSIS)is a platform for building high-performance data integration solutions.
DDS	The Data Distribution Service (DDS) is an application programming interface (API) standard for data-centred connectivity.
Agile	A project management approach that entails the breaking of the project into phases and emphasising continuous cooperation and improvement.
DevOps	A combination of software development (dev) and operations (ops) process.
AWS	Amazon Web Services (AWS) is a cloud-based platform, providing services and tools to support the DevOps process.
Azure	A cloud-based platform, providing services and tools to support the DevOps process.
Kafka	An event store and stream-processing platform that provides high-throughput for handling real-time data feeds.
Git	A version control system used for tracking changes in computer files.

The critical systems perspective is a useful approach to guide this research project to identify the skills the industry requires from data practitioners. This approach will be discussed in the next section.

Critical systems thinking

Systems thinking involves a holistic approach used to address problems by seeking to understand the relationships and interactions among the different elements of the system (Jackson 1991b, 184). Critical systems thinking (CST) is a strand of systems thinking characterised by its commitments to:

- critical awareness: presupposed assumptions are examined and re-examined
- social awareness: organisational and societal pressures should be considered
- emancipation: take issues of power in account and ensure that research is focused on the maximum development of an individual by giving a voice to parties that are affected by decisions but do not have control over the decisions
- methodological pluralism: make use of different research methodologies and sources of information to extract meaning
- theoretical pluralism: different theoretical points of view must be respected. (Jackson 1991a).

In this project, CST guides us to:
- consider different perspectives from stakeholders;
- do what is best for those at the receiving end of a university education – the students, industry and society;
- use a pluralistic approach in research methodology; and

- Identify the information needs of our audience.

Next, the research method is presented.

RESEARCH METHOD

In CST, it is important to consider the perspectives from different stakeholders. For this study (which forms part of a larger study), we sought the perspective of industry (as communicated in job listings) to determine which skills data professionals need.

Numerous studies worldwide have used text mining of online job advertisements for this purpose. Some of the recently published studies are listed in Table 2.

Table 2: Examples of studies that use text mining of online job advertisements

Reference	Sites scraped	Geographical area	Dates	Search words
Debortoli et al. 2014	Monster.com	US, Canada, Australia, UK	Sept 2013 to Mar 2014	business intelligence, big data
Kim et al. 2016	Indeed.com, Monster.com, and CareerBuider.com	Not specified	Not specified	data scientist
Stanton et al., 2016	LinkedIn	US	Jun 2016	big data, data science, digital analytics
De Mauro et al. (2018)	Dice.com	Not specified	Two months during autumn of 2015	big data
Hale 2018	LinkedIn, Indeed, SimplyHired, Monster, and AngelList	US	10 Oct 2018	data scientist
Stanton et al. 2019	LinkedIn	US	Not specified	15 types of analytics professionals
Verma et al. 2019	Indeed.com	US: Arkansas, Florida, Kansas, and Missouri	Not specified	business analyst, business intelligence analyst, data analyst, data scientist
Stanton and Stanton (2020)	LinkedIn	US	20–30 Apr 2019	55 data science and analytics job titles
Smaldone et al. (2022)	Not specified	US	Not specified	data scientist

Online job listing data are qualitative in nature because they do not follow a particular structure. However, it is important to note that these data do not describe a subjective experience as is usually the case of qualitative data in the interpretive paradigm. Here it is not only possible to analyse qualitative data from the positivistic paradigm because it is objective and quantifiable, but also desirable, because the audience for this study are academics in data science who are accustomed to the positivistic paradigm where quantitative analysis techniques are central.

Schoonenboom (2023, 13) explains that qualitative data can be quantified and then

analysed quantitatively. This cannot (and should not) be done on just any qualitative data, however; the nature and context of the data will determine if it is desirable to transform it into quantitative data. Qualitative data often contain much more detail than the quantitative (Witt 2001, 2). They are usually first coded and then examined to find emerging themes.

Job listing data, however, present an ideal opportunity to use quantitative research methods from a positivistic perspective without much loss of meaning, since these adverts are usually quite concise and succinct. They generally present certain common attributes, such as job title, location, as well as the qualifications, skills and experience required of the applicants sought. We were interested in learning which skills emerge, and how they can be grouped together in meaningful categories (typical qualitative analysis techniques). Following this, what percentage of adverts ask for a particular skill (or category of skill).

The method used therefore followed one of the paths advocated by Schoonenboom (2023, 9), as depicted in Figure 1 and indicated by the darker blocks and arrows.
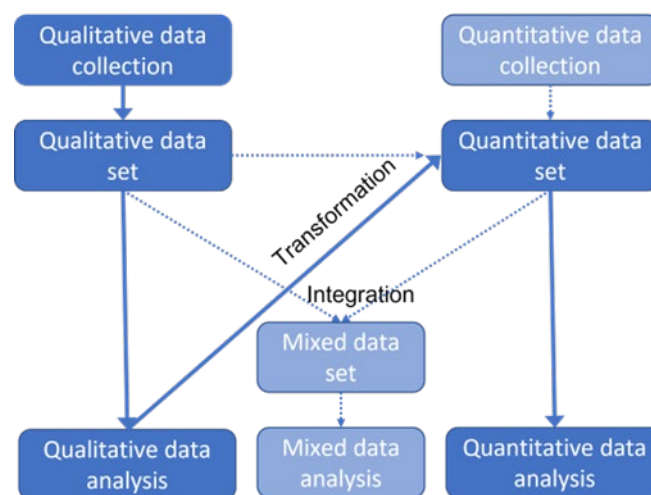

Figure 1: From data collection to data analysis through transformation and/or integration (after Schoonenboom, 2023)

This strategy fits in well with CST that promotes methodological pluralism, a multi-faceted approach that reveals the richness of the data, as explained by Mingers (2001). The following process was followed:

- Perform a qualitative analysis on the qualitative data using Atlas TI.
- Use the results of the qualitative analysis to prepare data for the quantitative assessment.
- Analyse the structured quantitative dataset using quantitative techniques.

This study focused on South Africa as geographical area and made use of job listings found in Glassdoor.com and LinkedIn.com and posted during January and February 2023. Search terms used were "data analyst", "data scientist", "data engineer", "business analyst", and "statistician". These were checked for duplicates and the duplicates were removed. The final dataset contained 156 job listings.

Step 1: Qualitative analysis using Atlas TI

The purpose of the qualitative analysis was to determine broad classes of employability requirements from the online job listings. Specific codes relating to each class also had to be identified, using Atlas TI, for use in the transformation to quantitative data, .

Figure 2 provides an example of a specific listing and demonstrates the coding process. Thus, the requirements for a data scientist were coded. Attributes sought included knowledge of data science, understanding business, willingness to take ownership, and being able to work with structured and unstructured data.

Although a wide range of specific skills were listed, saturation about the broad topics was soon reached. Where adverts stated technical skills required, it usually referred to competency in specific data analytics software. Hard skills, such as involving credit risk and business knowledge, were mentioned, as well as many soft skills like teamwork and communication. Figure 3 shows some of the codes that were classified under the theme "soft skills" as well as some quotations regarding communication skills required. Often, the field of study, level of education, and years of experience were mentioned.
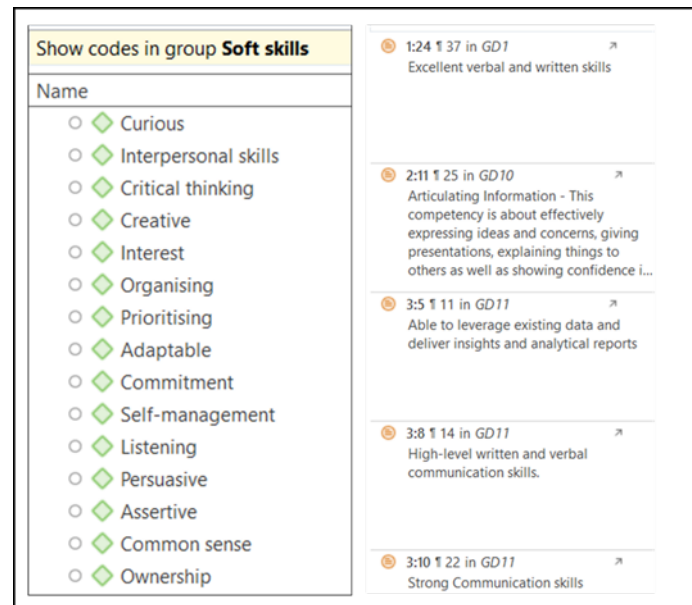

Figure 2: Example of a job listing and its codes in Atlas TI
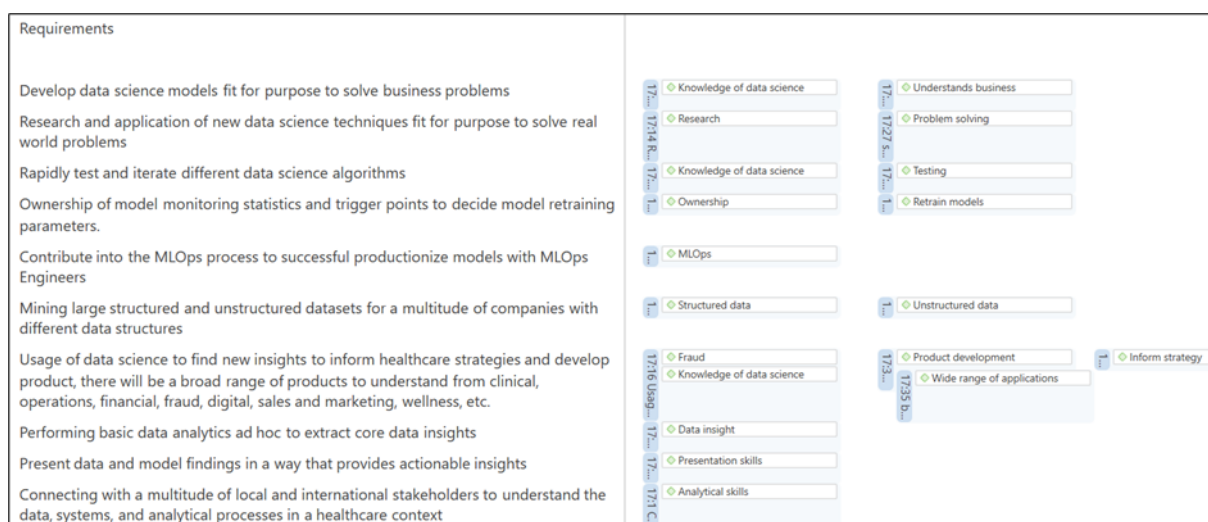
Figure 3: Example of some of the codes that were grouped together under "Soft skills" and some verbatim quotes that are represented by the code "Communication"

Step 2: Preparing data for quantitative analysis using R

The skills identified in the qualitative analysis were used to mine the text of all the adverts sampled using the programming language R. To ensure that important topics were not accidentally missed, skills identified in the comprehensive research conducted by Stanton and Stanton (2020) were also used. A total of 106 keywords were employed as Boolean modifiers. Each advert represented an observation (row) in the structured dataset that was created. If an advert contained a keyword such as "teamwork", a binary variable "teamwork" was set equal to 1, otherwise 0. In addition, key aspects were extracted such as job title, location (Gauteng, Western Cape, or other), minimum number of years of experience required, and qualifications needed. Figure 4 displays an extract of the structured dataset created.

Advert	DataAnalyst	BusinessAnalyst	DataScientist	DataEngineer	Statistician	JobTitle	Area	Python	SAS	SQL	presentation	writing	team	lead	communication	integrity	innovative	creative	learn	statistics	math	programming	AI
1	1	0	0	0	0	Data Analyst	Western Cape	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1
2	1	0	0	0	0	Data Analyst	Other	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1
3	1	0	0	0	0	Data Analyst	Other	0	0	1	0	0	1	1	1	0	1	0	1	0	0	1	1
4	0	1	0	0	0	Business Analyst	Western Cape	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1
5	0	1	0	0	0	Business Analyst	Western Cape	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1
6	1	0	0	0	0	Data Analyst	Western Cape	0	0	0	1	1	1	1	1	0	0	0	0	1	0	0	1
7	0	1	0	0	0	Business Analyst	Western Cape	0	0	0	0	0	1	1	1	1	1	0	1	0	0	0	1
8	1	0	0	0	0	Data Analyst	Western Cape	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0

Figure 4: Excerpt from structured dataset created during transformation step

Step 3: Quantitative analysis using R

The structured dataset created in step 2 was then subjected to quantitative statistical analysis techniques, as illustrated below.

RESULTS

Figure 5 displays the distribution of data analytics jobs by province. Most data analytics job listings are in Gauteng (49.4%) and the Western Cape (27.6%). Data analysts and data scientists are the most sought-after roles (28.8% and 28.2%), followed by business analysts (24.4%). Data engineers accounted for 14.1 per cent of the job openings, while statisticians had a mere 4.5 per cent.
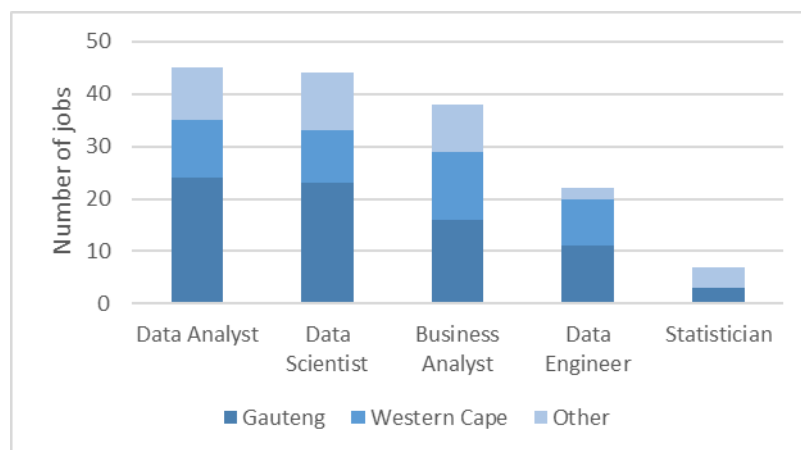
Figure 5: Data analytics job per province

In Figure 6, we see the minimum degree requirements for the sampled job listings. Thirty-three percent did not mention the required degree, whereas 42.3 per cent required a bachelor's degree, 7.7 per cent required an Honours degree, 14.8 per cent required a master's degree, and 2 per cent required a PhD. The majority of the listings required degrees in Statistics (39.5%), Computer Science (36.5%), Mathematics/Applied Mathematics (28.8%), Engineering (21.2%), Information Management/Systems/Technology (17.3%), Business/ Economics (11.5%).

Figure 7 presents a Pareto plot showing which data analytics software skills are in the highest demand. Three clearly stand out, namely, experience of SQL, Excel, and Python, each of which appears in more than 40 per cent of the advertisements. Other examples of software worth mentioning are Power BI, SSIS, ODS, Agile, SAS, and Scala, all of which appear in more than 20 per cent of the listings.
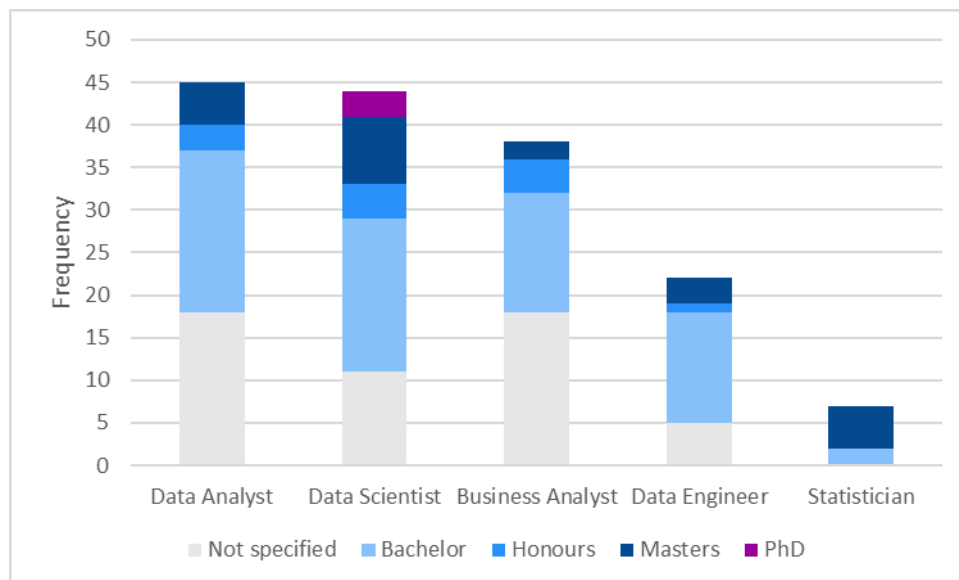
Figure 6: Minimum degree qualification required for a data analytics job
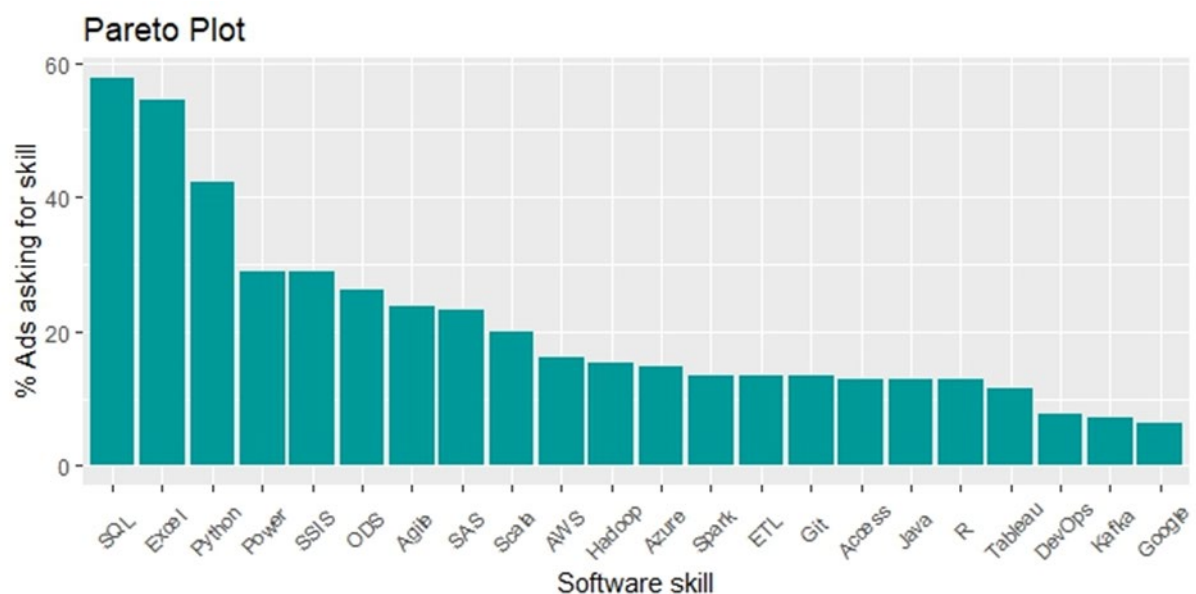

Figure 7: Pareto plot: mastery of data analytics software skills required

The top data analytics software skills can be divided into different categories, the most cited example of which is highlighted: database (***SQL***, Access, SSIS, ODS, and ETL), programming language (***Python***, SAS, R, Java, and Scala), data analysis and visualisation tools (***Excel***, ***Power BI***, and Tableau), cloud (***AWS***, and Azure), big data processing (***Hadoop***, Spark, and Scala), project management tools (***Agile***, DevOps, and GIT) and web analytics (***Google Analytics***).

The most sought hard skill required was in artificial intelligence (82% of adverts asked for it). Other important hard skills featured were statistics, mathematics, programming, database

management, data analysis and machine learning. Figure 8 shows the Pareto plot for the hard skills.
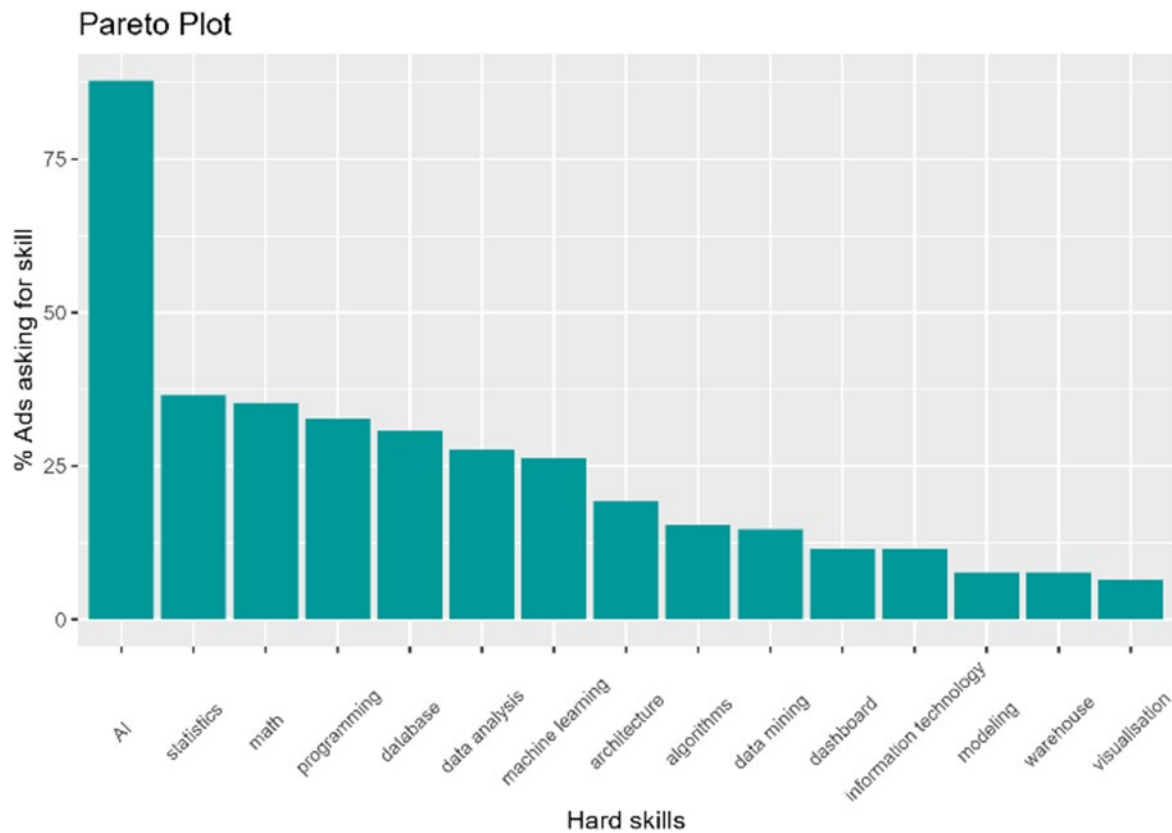

Figure 8: Pareto plot: hard skills required

The top three soft skills required of data analytics workers are the ability to work in a team, having business knowledge, and the ability to manage. Other important skills are the ability to lead, having analytical skills, willingness and ability to learn continuously, and being able to communicate well (Figure 9). The soft skills emerging from the job listings can be grouped into

- interpersonal skills (being able to work well in a team, lead a team, collaborate with others, consult and network);
- business skills (understanding the business, financial mechanisms, retail, marketing and managing a project or people);
- intrapersonal skills (willingness and ability to learn continuously, being self-driven, ability to adapt to change, independent, having a prominent level of integrity, able to work in a fast-paced, deadline-driven environment);
- analytical skills (being an analytical problem-solver);
- creative skills (being creative and innovative); and

- communication skills (being able to communicate verbally, give good presentations, and author excellent reports).
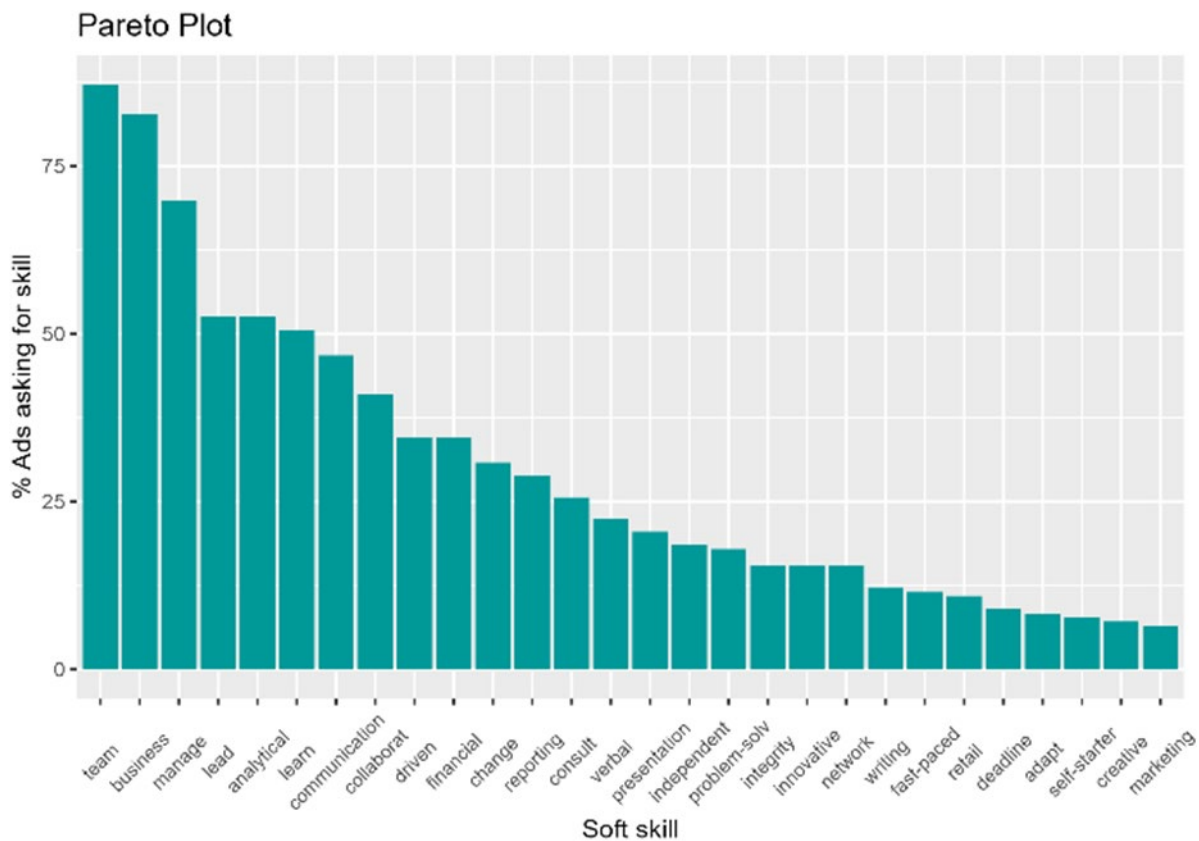

Figure 9: Pareto plot: soft skills required

Figures 10 and 11 present combined violin–boxplots (useful to visualise the distribution of the data and its probability density) that indicate the number of data analytics software skills, hard skills, soft skills and years of experience required per job category. Nonparametric Kruskal–Wallis tests were conducted to determine if there are significant differences between the number of skills required for the different jobs. The number of data analytics software skills required indeed differ significantly $[H(4) = 19.57, p = .0006]$, as well as the hard skills $[H(4) = 35.22, p < .0001]$ and years of experience $[H(4) = 16.89, p = .0020]$. The number of soft skills required, however, do not differ significantly $[H(4) = 3.96, p = .411]$. The median number of data analytics software skills and hard skills required for data engineers and data scientists are significantly higher than for the three other jobs. It should be noted that there is also much variation in the number of skills required per job listing for these two jobs. Although Figure 11 shows some difference between the jobs with regard to soft skills required, it is not statistically significant. As for work experience, posts for statisticians and business analysts require the most experience on average (5.8 and 5.5 years, respectively), and positions for data analyst the least (3.4 years). It is, however, worth pointing out that most of the posts (82.2%) require job
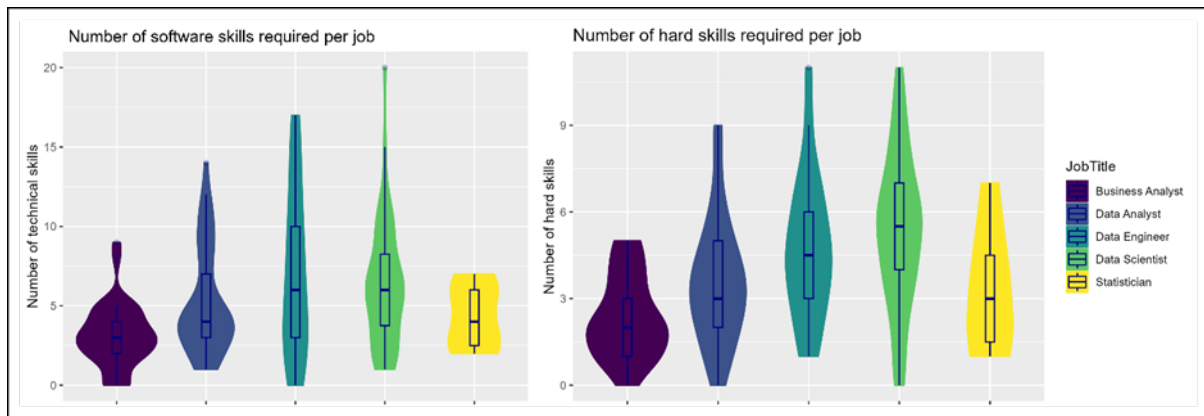
experience of 3 or more years.

Figure 10: Violin–boxplots: Number of technical and hard skills required per job
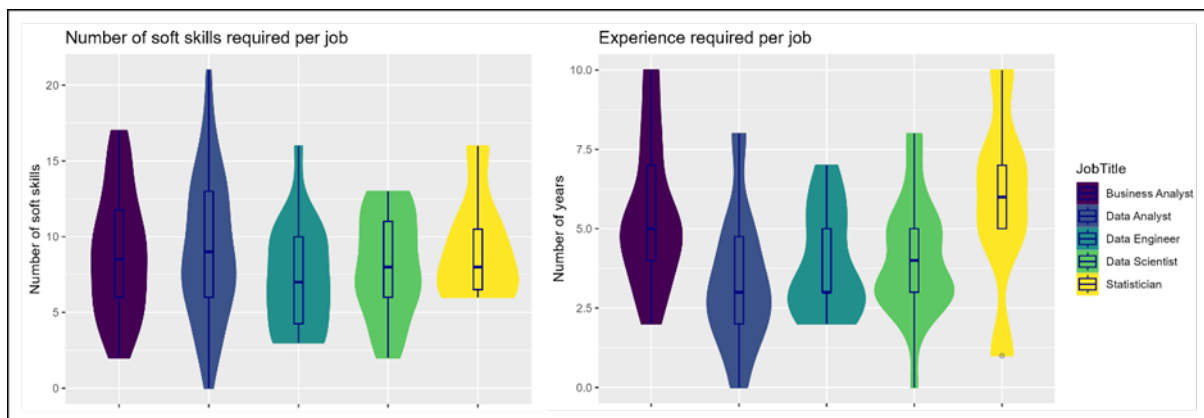

Figure 11: Violin–boxplots: Number of soft skills and years of experience required per job

DISCUSSION

The systems approach involves analysing a problem from multiple perspectives. This study assessed the essential perspective of industry, as expressed through job listings, on what skills are needed for data professionals. Jackson (2003) averred that CSH is an approach that counteracts possible injustice by warranting that parties that are affected by decisions should have a voice to influence these decisions. By analysing the needs of various industries, they can have a say in shaping the design of university degree programs. This will result in graduates who are better equipped for employment and, in turn, benefit both the employed graduates and the industry. Such graduates will be able to make an immediate positive contribution to their employer's bottom line. A university that produces excellent graduates for the industry will eventually gain benefits, as the word will spread that the graduates of that particular institution are highly sought after. This will attract the top students to the university, ultimately leading to its growth and success. This study is significant in determining the requirements of the industry, which will help in designing the academic programme for data analytics students.

The measure of success of the system under consideration is the employability of the graduates that it produces. The success of the system being considered is determined by the employability of the graduates it produces. Universities can take measures to enhance their students' employability by providing them with the highly desirable skills in AI, statistics, mathematics, programming and database management and appropriate data analytics software. Offering these skills to students can help them succeed in the job market and is within the university's control. Universities should prioritise the development of soft skills such as business acumen, interpersonal skills, analytical abilities, creativity, and communication skills wherever possible. However, intrapersonal skills such as motivation, self-drive, independence, and integrity are self-regulation skills that fall outside the university's control. To help students understand the significance of developing these skills, programs that raise awareness about their importance may prove beneficial.

This study confirms that mastery of SQL and Python are among the data analytics software skills in greatest demand and supports the work of Kim and Lee (2016), and Hale (2018). SAS and Scala are also highly sought after skills, similar to the findings of Hale (2018, 4). While South African job listings do mention R, Hadoop, Spark, Java and Hive, it seems that they are not in such high demand as stated by Hale (2018, 4). Power BI has emerged as being relatively popular in South Africa, more than Tableau, which is notably popular abroad.

The required hard skill that stands out by far in this study, is in artificial intelligence (AI); 82 per cent of the adverts ask for this, whereas the next most popular hard skills are in statistics (37%), mathematics (35%), programming (32%) and database management (30%). Although the listings seldom give information on how the applicant will need to use AI, we suppose that it may entail machine learning, deep learning, natural language processing, and computer vision. The high demand for this skill confirms that South Africa is preparing for 4IR. It is therefore essential that universities add AI skills to the courses they offer. If possible, it should already be introduced at undergraduate level to produce graduates that can meet this high demand. Stanton and Stanton (2020, 152) found that 19.6 per cent of entry-level posts require AI skills, which are also ranked as among the top three skills reported for data scientists in a study by Smaldone et al. (2022, 680).

The top three soft skills identified are the ability to work in a team (87%), possessing business acumen (82%), and the ability to manage (69%). Collaborative teamwork was found to be especially important (Caeiro-Rodriguez et al. 2021, 29224; Smaldone et al. 2022, 679; Lundberg et al. 2020, 647). Other studies mention the importance of data practitioners having knowledge of the business sector (De Mauro et al. 2018, 810; Ohei and Brink 2019, 13515) and possessing management skills (De Mauro et al. 2018, 810; Caeiro-Rodriguez et al. 2021,

29224).

Preparing data analytics students is a challenging task for universities. Educators should ensure that their undergraduates learn at least one technology of each of the categories identified: for example, database query language, a programming language, data analysis and visualisation, cloud computing, big data processing, project management, and web analytics. A good combination would be SQL, Python, Excel and Power BI, AWS, Hadoop, Agile and Google Analytics. Since working in teams was found to be important, students should ideally learn how to use Git. The ideal curriculum should ensure that students learn the fundamentals and the application of statistics, mathematics, programming, and business principles. Including various AI applications is crucial to prepare students for 4IR. However, educators tend to prioritize knowledge acquisition over the development of soft skills, according to Ohei and Brink (2019). Collaborative university projects offer students a chance to develop essential soft skills. By working through the entire process of identifying a problem, gathering and cleaning data, building models, writing professional reports, and presenting findings to an audience, students can hone their abilities in communication, teamwork, critical thinking, and problem-solving. Caeiro-Rodriguez et al. (2021, 29225) averred that collaboration-orientated project-based learning may indeed be particularly useful to develop soft skills.

Universities play a vital role in preparing students for employability. However, students themselves also need to take steps to become employable. In other words, anything that universities cannot do to help students become employable is considered "outside the control of the system". Tertiary education institutions play a crucial role in preparing students for the job market. They can do so by making students aware of the expectations of industry and encouraging them to participate in extracurricular activities such as sports, academic societies and cultural events. This would help them develop essential skills like teamwork, leadership and creativity. Additionally, part-time work or volunteering can be highly beneficial for graduates as it helps them gain vocational experience and develop various soft skills. To showcase their achievements and skills, students can create an electronic employability profile that highlights everything they learned and were involved in during their time at university. Mapundu and Musara (2020) claim that an e-profile is valuable as it can improve resourcefulness, flexibility, entrepreneurial skills, and collaboration, and showcase data analytics software skills as well as hard and soft skills to potential employers.

CONCLUSION

Qualitative and quantitative methods were used to gain insight into the skills that industry requires from data professionals. This was achieved through studying South African online job

listings to inform HE on how to improve students' employability. Broad themes were identified that emerged from the job listings. Various skills required were also identified. These skills were combined with those described in the literature and used when mining the advertisements to create a structured dataset, which was analysed quantitatively. Since the context of job listings is relatively narrow, the quantification of qualitative data proved to be useful to gain more insight from the data.

Themes identified were data analytics software skills, hard and soft skills, as well as the type and level of education and years of experience that employers ask for. Understanding of artificial intelligence ranked among the top hard skills, followed by statistics, mathematics, and programming. Competence in working with SQL, Python, Excel, and Power BI were the four most sought-after data analytics software skills. Data engineers and data scientists generally need a wider range of data analytics software skills than other data professionals. The substantial number of soft skills mentioned in the adverts indicates the importance of these skills for a data professional to be successful. Soft skills identified included interpersonal, intrapersonal, business-related, analytical, communication and creative skills. A postgraduate degree was a requirement in a quarter of the listings; and even though 42 per cent indicated that they require a bachelor's degree, many of them stated that it is the minimum and that a higher degree is preferred. Degrees in Statistics and Computer Science were the most sought after, followed by qualifications in Mathematics / Applied Mathematics and Engineering. Graduates applying for posts should be aware that four out of five jobs posted require at least three years of experience. Business analysts and statisticians require on average at least five years of experience to be recruited.

Studying online job postings provided useful information for universities to enhance their data analytics programmes. By working more closely with industries, universities can gain an understanding of their needs and effectively respond to them.

This study identified the needs of the industry. Universities can now use this knowledge to identify gaps in their programmes, decide on how to respond to these gaps and implement the necessary changes to ensure that their programmes produce world-class data professionals.

LIMITATIONS AND FUTURE RESEARCH

This research was based on online job recruitment adverts posted in South Africa. While they give some indication of the needs of industry, they may not cover the full spectrum of skills in demand. Follow-up studies, where data practitioners and management from different companies are asked what their needs are and where they perceive the biggest skills gaps to be, would assist universities in deciding on the content of their curricula for data analytics students.

Permission

Permission was obtained from the Ethics Committee of the North-West University to perform this study, for which no funding was received.

REFERENCES

Anderson, P., J. Bowring, R. McCauley, G. Pothering, and C. Starr. 2014. "An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation Experience." *ACM Special Interest Group on Computer Science Education (SIGCSE), Atlanta, GA.*: 145–150. https://doi.org/https://dl.acm.org/doi/10.1145/2538862.2538936.

Aujla, Gagangeet Singh, Radu Prodan, and Danda B. Rawat. 2022. "Big data analytics in Industry 4.0 ecosystems." *Software: Practice and Experience* 52(3): 639–641. https://doi.org/10.1002/spe.3008.

Brink, Roelien, Alao Abiodun, and Kenneth Nwanua Ohei. 2019. "Information and Communication Technology (ICT) graduates and challenges of employability: A conceptual framework for enhancing employment opportunities in South Africa." *Gender and Behaviour* 17(3): 13500–13521.

Burnett, S. 2014. "Unemployed youth: 'Time bombs' or engines for growth?" *AFRICAN SECURITY REVIEW* 23(2): 196–205.

Caeiro-Rodriguez, Manuel, Mario Manso-Vazquez, Fernando A. Mikic-Fonte, Martin Llamas-Nistal, Manuel J. Fernandez-Iglesias, Hariklia Tsalapatas, Olivier Heidmann, Carlos Vaz De Carvalho, Triinu Jesmin, Jaanus Terasmaa, and Lene Tolstrup Sorensen. 2021. "Teaching Soft Skills in Engineering Education: An European Perspective." *IEEE Access* 9: 29222–29242. https://doi.org/10.1109/ACCESS.2021.3059516.

Debortoli, S., O. Müller, and J. vom Brocke. 2014. "Comparing business intelligence and big data skills: A text mining study using job advertisements." *Business & Information Systems Engineering*, 6(5): 289–300.

De Mauro, Andrea, Marco Greco, Michele Grimaldi, and Paavo Ritala. 2018. "Human resources for Big Data professions: A systematic classification of job roles and required skill sets." *Information Processing and Management* 54(5): 807–817. https://doi.org/10.1016/j.ipm.2017.05.004.

Duarte, Fabio. 2023. "Amount of Data Created Daily (2023)." https://explodingtopics.com/blog/data-generated-per-day. (Accessed 12 June 2023).

Fettach, Yousra, Mounir Ghogho, and Boualem Benatallah. 2022. "Knowledge Graphs in Education and Employability: A Survey on Applications and Techniques." *IEEE Access* 10. https://doi.org/10.1109/ACCESS.2022.3194063.

Hale, Jeff. 2018. "The most in-demand skills for data scientists." https://towardsdatascience.com/the-most-in-demand-skills-for-data-scientists-4a4a8db896db.

Heine, A. 2023. "10 Reasons Why Soft Skills Are Important for Your Career." Indeed.com: Career advice. https://www.indeed.com/career-advice/interviewing/why-are-soft-skills-important. (Accessed 18 January).

Hurrell, Scott A., Dora Scholarios, and Paul Thompson. 2013. "More than a 'humpty dumpty' term: Strengthening the conceptualization of soft skills." *Economic and Industrial Democracy* 34(1): 161–182. https://doi.org/10.1177/0143831X12444934.

Ilori, Matthew Olusoji and Ibrahim Ajagunna. 2020. "Re-imagining the future of education in the era of the fourth industrial revolution." *Worldwide Hospitality and Tourism Themes* 12(1): 3–12. https://doi.org/10.1108/WHATT-10-2019-0066.

Jackson, M. C. 1991a. "Five Commitments of Critical Systems Thinking." In *Systems Thinking in*

Europe, edited by M. C. Jackson, G. J. Mansell, R. L. Flood, R. B. Blackham and S. V. E. Probert, 61–71. Boston, MA: Springer US.

Jackson, M. C. 1991b. *Systems methodology for the management sciences*. Non-fiction. Contemporary systems thinking. Plenum.

Jackson, M. C. 2003. *Systems thinking: Creative holism for managers*. Bibliographies. John Wiley & Sons.

Kim, J. Y. and C. K. Lee. 2016. "An empirical analysis of requirements for data scientists using online job postings." *International Journal of Software Engineering and its Applications* 10(4): 161–172. https://doi.org/10.14257/ijseia.2016.10.4.15.

Lundberg, Gunhild M., Birgit R. Krogstie, John Krogstie, and Portugal April April Ieee Global Engineering Education Conference Porto. 2020. "Becoming Fully Operational: Employability and the Need for Training of Computer Science Graduates." In *2020 IEEE Global Engineering Education Conference (EDUCON)*, 644–651. IEEE.

Mabe, Kagiso and Kelvin J. Bwalya. 2022. "Critical soft skills for information and knowledge management practitioners in the fourth industrial revolution." *South African Journal of Information Management* 24(1). https://doi.org/10.4102/sajim.v24i1.1519.

Mapundu, M. and M. Musara. 2020. "E-Portfolios as a tool to enhance student learning experience and entrepreneurial skills." *South African Journal of Higher Education* 33(6): 191–214. https://doi.org/10.20853/33-6-2990. https://www.journals.ac.za/sajhe/article/view/2990.

Mills, Robert J., Katherine M. Chudoba, and David H. Olsen. 2016. "IS Programs Responding to Industry Demands for Data Scientists: A Comparison between 2011–2016." *Journal of Information Systems Education* 27(2): 131–140.

Mingers, John. 2001. "Combining IS Research Methods: Towards a Pluralist Methodology." *Information Systems Research* 12(3): 240–259.

Moss, P. and C. Tilly. 1996. "Soft skills and race: An investigation of black men's employment problems." *Work and Occupations* 23(3): 252–276. https://doi.org/https://doi.org/10.1177/0730888496023003002.

Ohei, Kenneth Nwanua and Roelien Brink. 2019. "Investigating the prevailing issues surrounding ICT graduate employability in South Africa: A case study of a South African university." *The Independent Journal of Teaching and Learning* 14(2): 29–42. https://doi.org/10.10520/EJC-1d66cb012b.

Rawboon, Khwanruethai, Atsuko K. Yamazaki, Wannaphop Klomklieng, and Wisa Thanomsub. 2021. "Future Competencies for Three Demanding Careers of Industry 4.0: Robotics Engineers, Data Scientists, and Food Designers." *Journal of Competency-Based Education* 6(2). https://nwulib.nwu.ac.za/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1299756&site=eds-livehttp://dx.doi.org/10.1002/cbe2.1253.

Schoenherr, Tobias and Cheri Speier-Pero. 2015. "Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential." *Journal of Business Logistics* 36(1): 120–132. https://doi.org/10.1111/jbl.12082.

Schoonenboom, Judith. 2023. The Fundamental Difference Between Qualitative and Quantitative Data in Mixed Methods Research. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research* 24(1). https://doi.org/10.17169/fqs-24.1.3986.

Smaldone, Francesco, Adelaide Ippolito, Jelena Lagger, and Marco Pellicano. 2022. "Employability skills: Profiling data scientists in the digital labour market." *European Management Journal* 40(5): 671–684. https://doi.org/10.1016/j.emj.2022.05.005.

Stanton, Wilbur W. and Angela D'Auria Stanton. 2020. "Helping Business Students Acquire the Skills Needed for a Career in Analytics: A Comprehensive Industry Assessment of Entry-Level Requirements." *Decision Sciences Journal of Innovative Education* 18(1): 138–165.

Statistics South Africa. 2022. "Quarterly Labour Force Survey Quarter3: 2022." https://www.statssa.gov.za/publications/P0211/P02113rdQuarter2022.pdf.

Ulrich, W. 2003. "Beyond Methodology Choice: Critical Systems Thinking as Critically Systemic Discourse." *The Journal of the Operational Research Society* 54(4): 325–342.

Van Lill, Rinet and Therese Maria Bakker. 2022. "Life at a Stop Sign: Narrative Plots of the Transition to Adulthood During Unemployment Among South African Graduates." *Emerging Adulthood* 10(1): 124–134. https://doi.org/10.1177/2167696820937879.

Verma A, K. M. Yurov, P. L. Lane, and Y. W. Yurova. 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. doi:10.1080/08832323.2018.1520685.

Witt, Harald. 2001. Forschungsstrategien bei quantitativer und qualitativer Sozialforschung. FQS 2(1): Art. 8. DEU.

Yorke, Mantz. 2006. *Employability in Higher Education: What It Is, What It Is Not Learning and Employability Series 1*. York, United Kingdom: The Higher Education Academy.