Navigating the Data Science Landscape: Essential Competencies

Mehmet KOKOÇ

Trabzon University

Introduction

Although data science has gained prominence in recent years due to advancements in technology and business, its roots extend further back in time. Data scientists, the experts in this field, are versatile professionals who drive decision-making by extracting meaningful insights from data. To succeed, data scientists must possess a blend of technical, analytical, and soft skills. The increasing demand for data scientists has highlighted the importance of understanding the core competencies required for this profession. These skills can be categorized into technical areas such as programming, statistical and mathematical knowledge, machine learning, data management, and visualization; analytical skills like problem-solving, critical thinking, and domain expertise; and soft skills including communication, teamwork, and adaptability. However, it is important to note that these competencies can vary depending on the industry and specific role within data science. Consequently, data science training programs and curricula play a significant role in shaping these skills. Therefore, it is essential that data science education continuously evolves and adapts to industry developments. This book chapter aims to explore in detail the essential competencies data scientists must possess and contribute to the development and training processes within this field.

Data Science as a Field

Data science can be defined as an interdisciplinary field that processes structured and unstructured large data sets to derive meaningful information from them. As a field, data science integrates statistics, computer science, mathematics and domain-specific expertise to gain insights to solve complex problems as shown in the Venn diagram for data science (Hazzan & Mike, 2023). Data science is not only a process of analysis, but also includes predictions, the development of decision support systems and the optimization of processes, thus offering versatile functionality. The origins of data science can be traced back to the early 20th century and initially arose from statistical analysis. Pioneers in the field of statistics, such as Francis Galton and Karl Pearson, developed fundamental techniques such as correlation and regression (Stanton, 2001). In its modern meaning, however, data science entered a new phase with the advancement of computer technologies in the 1960s. John Tukey, in his seminal work "*The Future of Data Analysis*" published in 1962, argued that data analysis should not be limited to statistical models, but rather should be exploratory in nature (Tukey, 1962). With the advent of database management systems in the 1970s, the storage and processing of data became possible on a large scale and the concept of data mining began to take shape. With the increase in computing power in the 1980s, data analysis became more

sophisticated and enabled not only retrospective analysis but also predictive modeling. In the 1990s, the term "*data science*" officially entered the literature through the work of Peter Naur (Naur, 1992). At that time, however, data science was still considered a sub-discipline of statistics and computer science. In the 2000s, the spread of the internet and the concept of "big data" facilitated the recognition of data science as an independent field (Dhar, 2013).

Figure 1
Data Science as an Interdisciplinary Field (Hazzan & Mike, 2023)
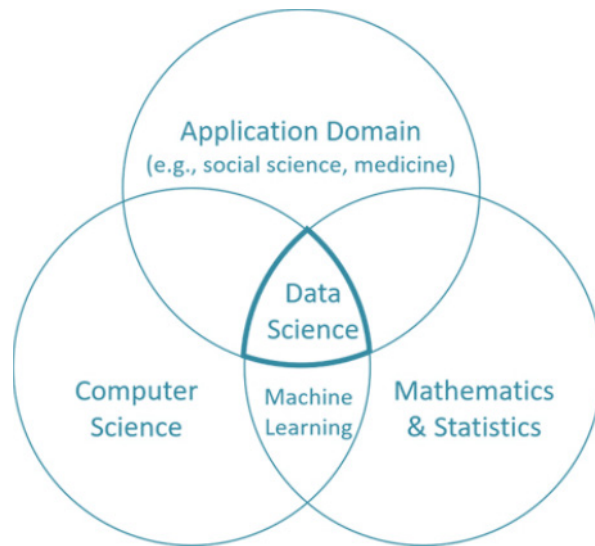
In the 21st century, data science experienced rapid growth driven by digitization and big data technologies. Tools such as Hadoop and Spark enabled data scientists to analyze huge datasets (Zikopoulos et al., 2012). At the same time, artificial intelligence (AI) and machine learning (ML) algorithms expanded the possibilities of data science. Today, data science is not limited to understanding the past but is also a powerful tool for predicting and optimizing the future. Data science now plays a crucial role in various sectors such as healthcare, finance, marketing, education and engineering. In healthcare, for example, AI-powered models are being developed for disease diagnosis and treatment optimization, while in finance, data science is being used for fraud detection and investment strategies (Chen et al., 2021). In the future, data science is expected to focus more on automation and ethical aspects. Dealing with algorithmic bias, ensuring data privacy and promoting a more democratic use of data will be crucial to improve the sustainability of data science. Furthermore, technologies such as quantum computing could herald a new era in data science (Preskill, 2018).

Data science, which emerged at the intersection of statistics and computer science, has evolved into a scientific discipline, characterized by technological advances. During its historical development, data science has integrated various scientific fields, undergone technological transformations and achieved its current crucial importance. In the modern world, data science is transforming decision-making processes and offering innovative solutions that touch all areas of life. Although it is a concept that has gained prominence in recent years along with technological and business changes, the roots of data science lie deep in the past. Experts in the field -data scientists- extract meaningful insights from data to provide effective and timely solutions to real-world problems. So, what are the knowledge and skills that a competent data scientist needs to have? The answers to this question will be insightful for both those who want to become data scientists and educators who want to teach data science.

Competencies and Skills for Success in Data Science

Success in data science requires a blend of technical expertise (hard skills), analytical thinking and interpersonal skills (soft skills), making it a uniquely interdisciplinary field. As data becomes more complex and voluminous, the ability to gain actionable insights depends on mastering a combination of hard and soft skills. A data scientist must be proficient in programming languages, statistical methods and machine learning techniques, while also excelling in problem solving and communication. These skills not only enable the effective processing and interpretation of data, but also support collaboration across different teams and domains. In addition, ethical considerations and domain-specific knowledge are becoming increasingly important as data science impacts critical areas such as healthcare, finance and environmental studies. This section looks at the key skills that underpin data science expertise and explores their role in driving innovation and delivering impactful solutions.

The KSAO model (Knowledge, Skills, Abilities, and Other Characteristics) provides a foundational framework for understanding the competencies required for effective performance in professional roles, particularly in data science. Introduced by Boyatzis (1982) and expanded in HR and IT literature, the model identifies four key dimensions. Knowledge encompasses the theoretical and domain-specific expertise necessary for data science, such as proficiency in statistics, machine learning, and programming languages like Python and R (Binkley et al., 2012). Skills focus on the practical application of this knowledge, including data cleaning, visualization, and predictive modeling, along with soft skills like communication and teamwork (Hattingh, 2014). Abilities represent innate or developed capacities, such as problem-solving, adaptability, and critical thinking, which enable data scientists to approach complex, interdisciplinary challenges. Finally, Other Characteristics include personal traits and behavioral tendencies, such as creativity, resilience, and ethical awareness, which contribute to success in dynamic environments (Boyatzis & Kolb, 1999). This comprehensive framework allows organizations to structure training programs and performance evaluations while aligning with emerging competency models, such as KSAVE, which adds values and ethics as critical components to meet modern societal and organizational demands (Dhar, 2013; Hattingh, 2014). By integrating these dimensions, the KSAO model ensures that data scientists possess both the technical expertise and interpersonal attributes needed to navigate the complex, evolving landscape of data-driven industries.

The KSAVE (Knowledge, Skills, Attitudes, Values, and Ethics) model, introduced by Binkley et al. (2012) as part of the Assessment and Teaching of 21st Century Skills (ATC21S) initiative, provides a comprehensive framework for developing skills that are critical to the modern workforce, particularly in interdisciplinary fields such as data science. KSAVE model extends traditional models such as KSAO model and incorporates values and ethics as core elements, emphasizing innovation, transparency and continuous learning alongside responsible decision making and accountability. This integration addresses the ethical and societal challenges in areas such as healthcare, finance and government where data science has a significant impact. In addition to technical competence, the model emphasizes attitudes such as curiosity, resilience and growth mindset, recognizing their importance in tackling complex problems. By prioritizing these human-centered dimensions, KSAVE model acknowledges that technical competence alone is not sufficient to meet the demands of ethical and socially responsible practices (Schleicher, 2012). The holistic approach of the KSAVE model is consistent with the interdisciplinary nature of data science and prepares professionals to use their skills responsibly and ethically. It emphasizes the importance of combining technical expertise with a strong moral framework, providing a critical foundation for the development of 21st century skills in a rapidly evolving digital landscape.

The study by Silveira et al. (2020) offers a comprehensive framework for understanding

the ***Core Skills Required in Data Science***. By employing both qualitative and quantitative methods, Silveira et al. systematically analyzed job postings to identify critical competencies for data scientists. Their proposed framework integrates six dimensions of data science, as outlined by Donoho (2017): *Data exploration and preparation, data representation and transformation, computing with data, data visualization and presentation, data modeling, and science about data science*. Each dimension encompasses both technical skills—such as programming, machine learning, and statistical analysis—and soft skills, including communication, creativity, and problem-solving. This dual emphasis underscores the multidisciplinary nature of data science, highlighting the necessity of balancing technical expertise with business acumen. Additionally, the study emphasizes the flexibility in educational requirements within the industry, reflecting a preference for practical skills over formal qualifications. By aligning academic and industry perspectives, the framework aims to guide training and recruitment efforts, thereby addressing the growing demand for skilled professionals in data science.

In the context of defining the competencies required for data scientists, a recent study (2018–2023) conducted an extensive literature review to address the lack of clarity on this topic and created ***Data Scientist Competency Framework*** (Zarefard & Marsden, 2024). The researchers developed a comprehensive framework (see Figure 2) by identifying and categorizing 130 competencies across seven critical domains: Functional, ethical, cognitive, consciousness, social, organizational and behavioral skills.

Figure 2
The Data Scientist Competency Framework (Zarefard & Marsden, 2024)

This framework, as seen in Figure 1, provides valuable insight for aligning data science education and career development with industry needs, emphasizing the interplay between technical expertise and soft skills such as communication, integration, and regulatory compliance. Functional competencies emphasize technical expertise, while ethical competencies ensure adherence to responsible practices throughout the data lifecycle. Cognitive competencies focus on intellectual skills and the use of advanced tools, while awareness competencies emphasize emotional intelligence and understanding of business contexts. Social competencies involve effective collaboration and interpersonal skills, while organizational competencies combine technical expertise with organizational goals and stakeholder relationships. Finally, behavioral competencies emphasize adaptability, entrepreneurial behavior and basic personal attributes. Together, these areas form a holistic

framework that aligns individual skills with organizational needs, encourages innovation and supports responsible and effective performance in a data-driven environment. The study highlights the importance of tailoring skills to specific roles and recognizes that it is unrealistic for one person to master all skills. This perspective is consistent with the broader understanding of data science as an interdisciplinary and collaborative field. Furthermore, the study uses an innovative methodology that combines advanced text mining and statistical analysis, setting a methodological precedent for future interdisciplinary research. By combining theoretical insights and practical applications, this study contributes to the academic discourse on data science skills while providing a robust model for assessing and improving professional performance in both academic and industrial settings.

Another study examines the data science skills, core competencies and career preferences of Nigerian college students with the aim of informing curriculum development for Library and Information Science (LIS) degree programs (Olatokun, Ayanbode, & Oladipo, 2024). According to the findings, *success in data science requires a combination of technical and non-technical (soft) skills*. These skills are not limited to programming or data analysis skills, but also include communication, problem solving and teamwork. Research emphasizes the interdisciplinary nature of data science and points out that knowledge of programming, data management or statistics alone is not enough. To achieve meaningful results in a business context, skills such as analytical thinking, critical evaluation and domain knowledge are equally important. The study also examines the relationship between students' skills and their interest in pursuing a career in data science. It finds that the skills required for data scientists can vary depending on industry needs, highlighting the importance of LIS degree programs equipping students with both technical and non-technical skills. These findings offer valuable insights into the design of curricula for data scientists. Educational programs should include diverse learning opportunities that allow students to develop these competencies and meet industry expectations to ultimately prepare qualified data science professionals.

The previous sections have highlighted various competency frameworks and domains of expertise crucial for data scientists, including the KSAO and KSAVE models, the competency framework by Silveira et al. (2020), and the recent Data Scientist Competency Framework by Zarefard & Marsden (2024), among others. Each of these models presents unique perspectives on the essential skills for success in data science, reflecting the diversity and interdisciplinarity inherent in the field. However, upon careful examination of these competency frameworks, a common thread emerges all of these competencies can fundamentally be classified into two broad categories—technical/professional competencies and soft skills. Technical competencies encompass the core expertise in areas such as programming, machine learning, statistical analysis, data management, and other domain-specific knowledge required to perform data science tasks. Meanwhile, soft skills include interpersonal abilities such as communication, teamwork, adaptability, problem-solving, ethical considerations, and creativity, which are equally crucial for effective practice in real-world, collaborative environments.

Given the overarching categorization that these competencies naturally fall into, this book chapter aims to systematically explore data science skills under two main headings: *Technical/Professional Competencies and Soft Skills*. By structuring the discussion in this way, I provide a comprehensive and accessible examination of the skills data scientists need, highlighting their interplay and importance in the data-driven industry. This approach allows us to critically assess how these competencies work together to support innovation, ensure ethical practice, and deliver impactful solutions in diverse fields such as healthcare, finance, and beyond.

Technical/Professional Competencies

Technical competencies are of critical importance for anyone aspiring to be a successful data scientist. These skills enable data scientists to analyze large datasets, derive

meaningful insights, and support strategic decision-making processes. This section explores the core technical skills that a data scientist must possess to excel in the field, focusing on programming, data analysis, machine learning, big data technologies, data visualization, and more.

Programming Languages

Proficiency in programming is fundamental for data scientists, as it allows them to manipulate data, build models, and create effective solutions. Python is one of the most widely used programming languages in data science, favored for its extensive libraries like Pandas, NumPy, Scikit-learn, and TensorFlow, which provide powerful tools for data manipulation and model building (Costa & Santos, 2017). SQL (Structured Query Language) is equally important, as it is essential for querying databases and retrieving data, making it a critical tool for any data scientist working with large datasets (Da Silveira et al., 2020). Additionally, R is a popular language for statistical analysis and data visualization, providing sophisticated tools for researchers and analysts to effectively explore data.

Data Analysis, Statistics, and Machine Learning

To succeed as a data scientist, it is essential to master a diverse set of competencies spanning data analysis, statistics, and machine learning. These skills encompass both theoretical knowledge and practical abilities, enabling data scientists to handle complex datasets, build models, and derive actionable insights.

Data analysis is a foundational aspect of data science, and it involves understanding, cleaning, transforming, and visualizing data to prepare it for modeling and decision-making. Exploratory Data Analysis (EDA) is one of the most important components of data analysis, as it helps data scientists understand the structure and characteristics of the data. EDA includes processes such as data cleaning, visualization, and the use of summary statistics, which are crucial for identifying patterns, detecting anomalies, and making informed decisions about the next steps in analysis (Das & Mishra, 2021; Zarbin et al., 2021). In addition, data manipulation and transformation skills are critical for converting raw data into a form suitable for analysis. This involves collecting, cleaning, organizing, and transforming data to ensure it is reliable and consistent for further use. Proficiency in data wrangling tools and techniques is essential to make raw data meaningful and actionable, supporting data-driven decision-making (Cady, 2017; Ismail & Abidin, 2016).

Statistics form the backbone of data science, providing the theoretical foundation necessary for making sense of data and developing models. A solid understanding of basic statistical concepts such as variables, sampling, correlation, and outlier handling is crucial for interpreting data and performing accurate analyses (Das & Mishra, 2021; Cady, 2017). This foundational knowledge enables data scientists to correctly apply statistical methods and derive insights that are both valid and reliable. Statistical modeling, which includes techniques such as regression analysis, hypothesis testing, and statistical inference, is another vital competency. These techniques allow data scientists to draw meaningful conclusions from data, develop predictive models, and test hypotheses, thereby enabling a deep understanding of underlying relationships within datasets (Lipovetsky, 2022; Thiruvengadam et al., 2022). For larger and more complex datasets, high-dimensional data analysis is often required. Advanced statistical techniques and data mining methods are employed to effectively analyze large datasets and extract valuable insights. Techniques suitable for high-dimensional data help data scientists handle the increasing volume, velocity, and variety of data in today's big data landscape (Cady, 2017; Donoho, 2017).

Machine learning is central to the role of a data scientist, providing the means to create predictive models and automate analytical tasks. A successful data scientist must be adept at applying basic machine learning algorithms such as k-Nearest Neighbors, Naive

Bayes, linear and logistic regression, decision trees, neural networks, and clustering. Mastery of these algorithms is necessary for building models that can classify data, predict future trends, and discover hidden patterns in datasets (Cady, 2017; Zarbin et al., 2021). Furthermore, skills in model evaluation and selection are essential for ensuring that the machine learning models developed are accurate and reliable. Techniques like cross-validation, model selection, and hyperparameter optimization play a critical role in improving the performance and robustness of models (Thiruvengadam et al., 2022). In recent years, deep learning and advanced machine learning techniques have become increasingly important, especially when dealing with large datasets and complex problems. Neural networks and deep learning methods, supported by tools like TensorFlow and PyTorch, have opened new possibilities for solving problems in image recognition, natural language processing, and other domains that involve vast amounts of data (Cady, 2017; Summa et al., 2017).

Database Management and Big Data Technologies

Competencies in database management and big data technologies are crucial for data scientists in effectively managing and analyzing large volumes of data. These skills are essential for data storage, management, accessibility, and analysis, and enable data scientists to derive meaningful insights from both structured and unstructured datasets. Database management competencies include proficiency in SQL, which is fundamental for data retrieval, manipulation, and analysis. Additionally, understanding database design helps ensure data is organized efficiently. Data scientists should also apply FAIR principles (Findable, Accessible, Interoperable, Reusable) to make data accessible and reusable across different contexts (Duke et al., 2022; Mildenberger et al., 2023). Big data technologies competencies involve using tools for big data analytics such as Hadoop and Spark to process large datasets. Proficiency in machine learning helps derive insights from big data, and familiarity with both SQL and NoSQL storage architectures ensures effective data management (Kumar et al., 2022; Nkwanyana et al., 2023). Data scientists need strong competencies in database management, big data technologies to excel in modern data environments. These competencies empower them to manage, analyze, and interpret large datasets effectively, making a meaningful impact across industries.

Data Visualization

Data scientists must possess basic knowledge of statistics and visualization, which is crucial for understanding how to best represent data (Hehman & Xie, 2021). Proficiency in programming languages like R and Python is also essential, as these are widely used for implementing visualization techniques (Reyes, 2022). Additionally, a combination of creativity and technical knowledge is needed to select the most suitable visualization types for presenting data effectively. Advanced skills include creating interactive visualizations to make data more accessible and engaging, which requires familiarity with data management technologies and tools for interactive analysis (Battle & Scheidegger, 2020). Moreover, data scientists should have a comprehensive understanding of visualization tools to create visualizations across different domains such as healthcare, IoT, and business dashboards (Shakeel et al., 2022). Educational programs and resources are essential for developing data visualization skills, preparing students for future professional roles (Kirk et al., 2021). Adopting a step-by-step approach can help data scientists learn visualization techniques effectively, particularly in programming languages like R and Python (Reyes, 2022). Visualization competencies for data scientists encompass basic statistical knowledge, proficiency in programming languages, creativity, interactive visualization skills, and data management techniques. These competencies enable data scientists to present data clearly and effectively, ensuring their analysis results are well communicated. Educational initiatives play a crucial role in the development of these skills, supporting data scientists in their professional growth.

Cloud Computing

Data scientists need a range of skills to effectively utilize cloud computing, which has become essential for processing large data sets and conducting data-intensive experiments. Core competencies include the ability to understand and utilize cloud-based services and resources required to manage analytics without relying on on-premises infrastructure (Jami & Munir, 2021). In addition, data scientists need to master the scalability and flexibility of cloud computing for processing complex data sets. This can be seen in models that use machine learning algorithms to derive insights and support decision-making processes in companies (Archana & Kamalraj, 2024). In addition, mastery of programming languages such as R and their integration with cloud APIs is crucial for performing analytics and effectively managing data on cloud platforms (Ohri, 2020). Cloud computing democratizes access to resources and enables data-driven scientific and engineering applications that were previously constrained by cost. In this context, data scientists need to understand the economic and technical dimensions of cloud services, including the pay-as-you-go model and the challenges of data transfer and storage (Simmhan et al., 2016). Data scientists need to combine technical expertise with a strategic understanding of cloud economics and the ability to integrate cloud technologies into data science workflows to maximize the potential of cloud computing in their field.

The path to becoming a successful data scientist involves mastering a diverse set of technical competencies. These include proficiency in programming languages like Python, SQL, and R, a strong foundation in statistics and machine learning, database management expertise, knowledge of big data technologies, and skills in data visualization and cloud computing. Additionally, competencies in API usage, version control, ethical considerations, and domain-specific knowledge are critical to thriving in this dynamic field (Costa & Santos, 2017; Da Silveira et al., 2020; Hattingh et al., 2019). Together, these skills empower data scientists to analyze complex datasets effectively, derive meaningful insights, and contribute significantly to strategic decision-making across industries. In the ever-evolving domain of data science, possessing these competencies is essential for driving innovation, ensuring ethical practices, and delivering impactful solutions.

Analytical Skills

Analytical and problem-solving skills are essential for data scientists, as they enable them to navigate the complexity of modern data landscapes and effectively tackle multi-layered challenges. At the heart of this expertise is ***data-driven thinking***, a mindset in which problems are approached quantitatively, and decisions are based on facts rather than intuition. This approach requires data scientists to interpret raw data, recognize patterns and derive meaningful insights that can inform strategic decisions. In addition, they must be able to break down complex problems into smaller, more manageable components that allow for systematic analysis and identification of critical variables that influence the outcome.

Another essential aspect of ***problem solving*** in data science is the ability to model real-world challenges mathematically or statistically and translate them into forms that can be analyzed computationally. Data-driven thinking further supports this by ensuring that hypotheses to specific questions are rigorously developed and tested against empirical data. This iterative process not only refines their understanding, but also guarantees that the proposed solutions are robust and based on objective analysis. Furthermore, understanding cause-and-effect relationships within datasets is crucial for distinguishing between correlation and causation, allowing data scientists to focus on eliminating causes rather than just treating symptoms.

Effective data scientists are also characterized by ***strategic thinking*** and the ability to assess risks and probabilities, both of which are enhanced by a data-driven approach. By

assessing the possible outcomes of different scenarios, they help companies anticipate challenges and seize opportunities. This includes benchmarking potential solutions, comparing their effectiveness and suggesting alternative methods if necessary. Ultimately, these capabilities are results-oriented and focus on implementing and refining solutions to ensure measurable impact. By integrating data-driven thinking with a structured, evidence-based methodology, data scientists not only solve problems effectively, but also contribute to the long-term success of the organization.

Soft Skills

The success of data scientists in the digital era increasingly depends on a combination of technical expertise and soft skills. While technical skills such as programming, data analysis, and machine learning are foundational, it is the development of soft skills that often distinguishes highly effective data scientists. Essential soft skills for data scientists include *communication, leadership, teamwork, empathy adaptability, and problem-solving* (Afsharian, 2024; Lu, 2024; Oropesa, 2024). These competencies are vital for effective collaboration, efficient project management, and translating complex data insights into actionable recommendations that drive business value (Desidério et al., 2024; Karneli et al., 2024).

Communication skills are particularly important because they allow data scientists to share their technical work with stakeholders who do not have a technical background. This is not just about showing numbers and charts, but rather about writing stories about specific analyses for the audiences for whom the analysis is relevant. In these disciplines, communication can mean several different things, such as a video, an infographic, an oversimplification of statistics, or the level of audience being addressed (Desidério et al, 2024). For example, when data scientists share the results of a predictive model with marketers, they are usually concerned with the application of the results to specific market segments rather than the mathematical models used. The ability to weave disparate information into a cohesive narrative is critical to ensuring that the findings are not only appreciated but also acted upon.

The ability to lead is another important skill that is more relevant today than ever. Data scientists are often required to take the lead on data-related projects, mentor junior staff and even act as advocates for data within an organization. In data science, leadership is not always just about managing people, but also about creating a roadmap of how data can be used to fulfil business needs. The focus here is on creating value through new technologies and increasing the organization's appetite for data. Data scientists who demonstrate leadership qualities describe themselves as "*agents of change*" who are actively involved in transformative processes that prescribe the inclusion of data in decision-making as well as strategic processes to improve the quality of their services (Lu, 2024).

Teamwork is also essential, as most off-the-shelf data science projects require a variety of data engineers, domain experts, stakeholders, customers and software developers. This allows them to see the business context and constraints that lead them to the data analysis relevant to that business (Coners et al., 2024). For example, when developing a predictive maintenance model for an organization, a data scientist needs to interact directly with engineers who know the equipment, operators who know about failures and provide feedback on patterns, and managers who work on the optimal allocation of resources. Such a collaborative approach ensures that the solutions developed are realistic and in line with business principles.

Empathy and flexibility are essential in a field like data science where there is a lot of change. New technologies, new tools, new ways of doing things all lead to new regulations and a data scientist should be willing to learn and unlearn as well as update their paradigm (Lu, 2024). This flexibility is also forward-looking and includes other types of innovation such as reorganization when new information has been added or the

goals of the project have changed. For example, a data scientist might use a particular data set over the course of a project, but then realize halfway through the project that external data is needed to improve the accuracy of the model. The ability to revise the analysis strategy at short notice and efficiently incorporate new sources of information can have a significant impact on the project.

Problem-solving remains at the core of data science landscape. Data scientists are, in essence, problem solvers who apply analytical skills to derive meaningful insights from data. However, problem-solving in this context often extends beyond technical challenges to include defining the right problem, understanding stakeholder needs, and ensuring that the proposed solution is feasible and valuable for the organization (Afsharian, 2024). This involves asking the right questions, breaking down complex problems into smaller, manageable parts, and systematically testing hypotheses to arrive at data-driven conclusions. For instance, when working on customer churn prediction, a data scientist must not only build a predictive model but also identify what interventions the company can realistically implement to reduce churn. Effective problem-solving in data science thus blends technical acumen with practical business considerations.

The importance of developing these soft skills cannot be understated, as they have a direct impact on project success and career advancement in the field of data science (Oropesa, 2024). Data scientists who excel at communication, leadership, adaptability, and teamwork are better equipped to lead impactful projects, influence decision-making processes, and ensure that data insights translate into tangible actions (Karneli et al., 2024). As the field of data science evolves, the role of these soft skills becomes even more critical, complementing technical capabilities to drive meaningful outcomes in the digital business landscape. Therefore, to remain competitive, data scientists must focus on enhancing both their technical and soft skills, enabling them to navigate complexities, work collaboratively, and contribute to the long-term success of their organizations (Ul Haq et al., 2024).

Ultimately, it is the integration of these soft skills — communication, leadership, teamwork, empathy and flexibility, problem solving, adaptability, and effective decision-making - that characterizes a successful data scientist. These competencies enable them to communicate their analysis persuasively, collaborate effectively and develop impactful solutions that advance their organization's strategic goals. Soft skills combined with technical know-how help turn raw data into meaningful insights that drive innovation and create significant value across industries.

Conclusion and Future Directions

Becoming a great data scientist is not an easy road as it requires a bunch of skills which includes both technical, analytical and soft skills. In this chapter, it is highlighted that data science is a seamless blend of programming, data analysis, machine learning and holistic vision to see and process complex data. In addition, it also refers to soft skills such as communication of analytical results, team cooperation effort and adaptation to the changing environment or roles in data science. The competencies of a data scientist discussed in this study are illustrated in Figure 3 for clarity and emphasis.

Figure 3
Key Skills of Data Scientist
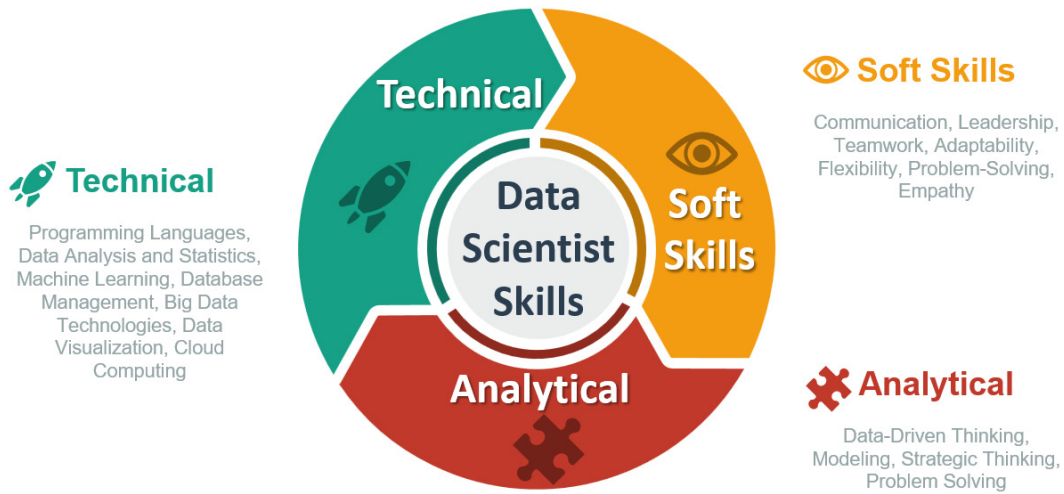
To be considered a data scientist, one should be competent enough to understand how data can contribute to any business decision-making process. The technical skills of a data scientist primarily include programming in languages such as Python and R, machine learning modeling and statistical analysis, and working with large amounts of data. Last but not least, these individuals should also be able to visualize their results not only for themselves but also for others, regardless of technical level. In addition to the technical skills mentioned above, data scientists are also expected to have a certain level of intellect that helps them to tackle various problems effectively, e.g. the ability to break a problem down into sub-problems and solve them one by one. And just as building a team requires skill and flair, so does working in a team with members from other fields. Such successful data scientists can present their findings in writing, respond flexibly to change and make useful business recommendations based on the insights gained from the data. This ensures that regardless of the type of data available to a data scientist, whether qualitative or quantitative, they not only analyze and process data, but also implement changes in the way the organization functions and other social issues.

It is forecasted that the discipline of data science will keep pace with changes owing to advancements in technology, especially within the frameworks of artificial intelligence, cloud computing, and big data technologies. The rise of big data calls for so complex and advanced machine learning algorithms, and as the case is, quantum computing presents significant possibilities in this context. Such changes are forcing data scientists to be dynamic and progressive in their technical abilities to embrace the cutting edge of technology. In addition, it has become apparent that psychosocial issues concerning data science practice cannot be overlooked. With modern societal expectations, issues on data privacy, fairness in algorithms and ethical use of artificial intelligence are becoming prominent and data scientists must be trained in this respect. So, they also must teach future practitioners these things, hence programs will have an emphasis on ethics and values and on social claims making them civil data scientists.

Dismantling unnecessary barriers to the use of data is, and will remain, a central preoccupation of data science in the years to come. With companies increasingly accepting the need to make decisions driven by data, it becomes imperative to find ways of making data science easy for non-technical professionals, who are the majority. This serves to indicate enlarging of the audience in terms of the development of platforms and trainings designed for effective interaction with data for all concerned on a more general scale. Improving the attitude towards dealing with data and encouraging a broader category of workers to embrace data science technology, within, can, in turn, draw a greater value from data science. Helping to establish that relationship would be even more important

given the complex global issues such as climate change, health care system or socio-economic disparities require addressing. In many cases, data scientists are expected to work hand in hand with subject matter experts and leverage their collective expertise to arrive at a broad solution.

The ever-changing landscape of data science has its own rewards and challenges. To be relevant and also efficient, a data scientist is expected to possess a multitude of hard and soft skills, be open to change and evolution in technology, as well as behave in a professional and ethical manner in work. If the focus is on the promotion of life-long education, the observance of ethical principles, and the enhancement of outreach and teamwork, data science can be a very powerful force for good in many spheres and within society itself. These outlooks help data scientists build their professional paths and combine them with solving many of the demands existing in the contemporary world.

References

Afsharian, M. (2024). Data science essentials in business administration: A multidisciplinary perspective. *Decision Analytics Journal*, *11*, 100442. https://doi.org/10.1016/j.dajour.2024.100442

Archana, G., & Kamalraj, R. (2024). A Cutting-Edge Data Science Model Leveraging Cloud Computing. *International Journal of Advanced Research in Education & Technology, 11*(3), 863-869. https://doi.org/10.15680/ijarety.2024.1103006

Battle, L., & Scheidegger, C. (2020). A Structured Review of Data Management Technology for Interactive Visualization and Analysis. *IEEE Transactions on Visualization and Computer Graphics, 27*, 1128-1138. https://doi.org/10.1109/TVCG.2020.3028891

Cady, F. (2024). *The data science handbook*. John Wiley & Sons. https://doi.org/10.1002/9781119092919

Chen, J., Asch, S. M., & Asch, D. A. (2021). Machine learning and prediction in medicine — Beyond the peak of inflated expectations. *The New England Journal of Medicine, 385*(2), 190-193. https://doi.org/10.1056/NEJMp2101709

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM, 13*(6), 377-387. https://doi.org/10.1145/362384.362685

Coners, A., Matthies, B., Vollenberg, C., & Koch, J. (2024). Data Skills for Everyone!(?)– An Approach to Assessing the Integration of Data Literacy and Data Science Competencies in Higher Education. *Journal of Statistics and Data Science Education*, 1-29. https://doi.org/10.1080/26939169.2024.2334408

Costa, C., & Santos, M. Y. (2017). The data scientist profile and its representativeness in the European e-Competence framework and the skills framework for the information age. *International Journal of Information Management*, *37*(6), 726-734. https://doi.org/10.1016/j.ijinfomgt.2017.07.010

da Silveira, C. C., Marcolin, C. B., da Silva, M., & Domingos, J. C. (2020). What is a Data Scientist? Analysis of core soft and technical competencies in job postings. *Revista Inovação, Projetos e Tecnologias, 8*(1), 25-39. https://doi.org/10.5585/iptec.v8i1.17263

Desidério, S. B., Lelis, M. R. L., Rodrigues, M. E., & Marques, A. B. (2024). How ready for HCI? A qualitative analysis of the practice of soft skills related to HCI by women involved in the digital girls program partners projects. *Journal on Interactive Systems*, 15(1), 504–516. https://doi.org/10.5753/jis.2024.3852

Dhar, V. (2013). Data science and prediction. *Communications of the ACM, 56*(12), 64-

73. https://doi.org/10.1145/2500499

Donoho, D. (2017). 50 years of Data Science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. https://doi.org/10.1080/10618600.2017.1384734

Donoho, D. (2017). 50 Years of data science. *Journal of Computational and Graphical Statistics, 26*, 745 - 766. https://doi.org/10.1080/10618600.2017.1384734

Duke, R., Bhat, V., & Risko, C. (2022). Data storage architectures to accelerate chemical discovery: data accessibility for individual laboratories and the community. *Chemical Science, 13*, 13646 - 13656. https://doi.org/10.1039/d2sc05142g

Hattingh, M., Marshall, L., Holmner, M., & Naidoo, R. (2019). Data science competency in organisations: A systematic review and unified model. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019 (SAICSIT '19)*, pp. 1–8. https://doi.org/10.1145/3351108.3351110

Hazzan, O., & Mike, K. (2023). What is Data Science? In *Guide to Teaching Data Science: An Interdisciplinary Approach*, pp. 19-34. Cham: Springer International Publishing.

Hehman, E., & Xie, S. (2021). Doing Better Data Visualization. *Advances in Methods and Practices in Psychological Science, 4*(4). https://doi.org/10.1177/25152459211045334

Hersh, W., Hoyt, R., Chamberlin, S., Ancker, J., Gupta, A., & Borlawsky-Payne, T. (2023). Beyond mathematics, statistics, and programming: data science, machine learning, and artificial intelligence competencies and curricula for clinicians, informaticians, science journalists, and researchers. *Health Systems, 12*, 255 - 263. https://doi.org/10.1080/20476965.2023.2237745

Ismail, N., & Abidin, W. (2016). Data scientist skills. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA), 3*(4), 52-61. https://doi.org/10.9790/0050-03045261

Jami, S. I. & Munir, S. (2021). Current Trends in Cloud Computing for Data Science Experiments. *International Journal of Cloud Applications and Computing (IJCAC), 11*(4), 80-99. https://doi.org/10.4018/IJCAC.2021100105

Karneli, O., Handayati, R., & Rijal, S. (2024). Enhancement of soft skills competence in human resources as a key success factor in the digital business era. *Journal of Contemporary Administration and Management (ADMAN)*, *2*(1), 319–324. https://doi.org/10.61100/adman.v2i1.126

Kirk, A., Santos, B., & Alford, G. (2021). A recipe of capabilities for pursuing expertise in data visualization: A practitioner's perspective. *IEEE Computer Graphics and Applications, 41*, 58-62. https://doi.org/10.1109/MCG.2020.3034737

Kumar, A., Deutsch, A., Gupta, A., Papakonstantinou, Y., Salimi, B., & Vianu, V. (2022). Database Education at UC San Diego. *ACM SIGMOD Record, 51*, 43 - 46. https://doi.org/10.1145/3572751.3572763

Lipovetsky, S. (2022). Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data. *Technometrics, 64*, 145 - 148. https://doi.org/10.1080/00401706.2021.2020521

Lu, J. (2024). Data Scientist Knowledge and Skills Evaluation Towards a Data-Driven Research Methodology. In *Proceedings of the 23rd European Conference on Research Methodology for Business and Management Studies*, pp.136-144. https://doi.org/10.34190/ecrm.23.1.2321

Mildenberger, T., Braschler, M., Rucksuhl, A., Vorburger, R., & Stockinger, K. (2023). The Role of Data Scientists in Modern Enterprises - Experience from

Data Science Education. *ACM SIGMOD Record, 52*, 48 - 52. https://doi.org/10.1145/3615952.3615966

Naur, P. (1992). *Computing: A human activity*. ACM Press/Addison-Wesley Publishing Co. https://link.springer.com/book/10.1007/978-1-4612-3096-2

Nkwanyana, A., Mathews, V., Zachary, I., & Bhayani, V. (2023). Skills and competencies in health data analytics for health professionals: A scoping review protocol. *BMJ Open, 13*. https://doi.org/10.1136/bmjopen-2022-070596

Ohri, A. (2020). *R for cloud computing: An approach for data scientists*. Springer.

Olatokun, W. M., Ayanbode, O. F., & Oladipo, S. O. (2024). Data science career preference of Nigeria University students. *Education and Information Technologies*, 1-25. https://doi.org/10.1007/s10639-024-12897-4

Oropesa, C. M. (2024). The essential soft skills of project managers and project success. Scientific *Journal of Applied Social and Clinical Science, 4*(20), 1-8. https://orcid.org/0000-0001-9168-998X

Preskill, J. (2018). Quantum computing in the NISQ era and beyond. *Quantum, 2*(1), 79. https://doi.org/10.22331/q-2018-08-06-79

Reyes, J. M. M. (2022). *Data Visualization for Social and Policy Research: A Step-by-step Approach Using R and Python*. Cambridge University Press.

Shakeel, H., et al. (2022). A Comprehensive State-of-the-Art Survey on Data Visualization Tools. IEEE Access, 10, 96581-96601. https://doi.org/10.1109/ACCESS.2022.3205115

Simmhan, Y., Ramakrishnan, L., Antoniu, G., & Goble, C. (2016). Cloud computing for data-driven science and engineering. *Concurrency and Computation: Practice and Experience, 28*(4), 947-949. https://doi.org/10.1002/cpe.3668

Summa, M.G., Bottou, L., Goldfarb, B., Murtagh, F., Pardoux, C., & Touati, M. (Eds.). (2017). *Statistical Learning and Data Science*. Chapman and Hall/CRC. https://doi.org/10.1201/b11429

Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9(3). http://dx.doi.org/10.1080/10691898.2001.11910537

Thiruvengadam, K., Watson, B., Chinnaiyan, P., & Krishnan, R. (2022). A review of statistical modelling and machine learning in analytical problems. *International Journal of Applied Engineering Research, 17*(5), 506-510. https://doi.org/10.37622/ijaer/17.5.2022.506-510

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics, 33*(1), 1-67. https://doi.org/10.1214/aoms/1177704711

Ul Haq, M. U., Frazzetto, P., Sperduti, A., & Da San Martino, G. (2024, April). Improving soft skill extraction via data augmentation and embedding manipulation. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, pp. 987-996.

Zarbin, M., Lee, A., Keane, P., & Chiang, M. (2021). Data science in translational vision science and technology. *Translational Vision Science & Technology, 10*. https://doi.org/10.1167/tvst.10.8.20

Zarefard, M., Marsden, N. (2024). The Essential Competencies of Data Scientists: A Framework for Hiring and Training. In: Mori, H., Asahi, Y. (eds) *Human Interface and the Management of Information. HCII 2024. Lecture Notes in Computer Science*, vol 14691. Springer, Cham. https://doi.org/10.1007/978-3-031-60125-5_27

Zikopoulos, P. C., Eaton, C., DeRoos, D., Deutsch, T., & Lapis, G. (2012). *Understanding Big Data: Analytics for enterprise class Hadoop and streaming data*. McGraw-Hill.

About The Authors

M. KOKOÇ received a Ph.D. degree from the department of Computer Education and Instructional Technologies at Hacettepe University, a highly esteemed institution in Türkiye. He currently holds the position of Associate Professor in the Department of Management Information Systems at the School of Applied Sciences at Trabzon University. His research interests include Management Information Systems, E-Learning, Learning Technologies, Learning Analytics, Cognitive Profiling, Video Lectures, Human-Computer Interaction, and Cognition and Media.
E-mail: kokoc@trabzon.edu.tr **ORCID:** 0000-0002-1347-8033

Similarity Index
The similarity index obtained from the plagiarism software for this book chapter is 4%.