

Data analytics job skills dynamics: a preliminary longitudinal analysis

Mohamed Almgerbi, Andrea De Mauro, Adham Kahlawi, Valentina Poggioni

Abstract This study investigates the evolving role of Data Analytics between 2019 and 2023, analyzing the shifting job market demands for professionals in this field. Utilizing a corpus of job postings, we employed Latent Dirichlet Allocation (LDA) for topic modeling to discern patterns in required skills and expertise. Our findings indicate a marked increase in data analytics roles specialization and an increasing prominence of Machine Learning.

Keywords: data analytics, machine learning, job market, human resources, job skills

1 Introduction

The rapid evolution in digital technology over recent decades has unlocked vast volumes of data in various sectors, known as Big Data, prompting businesses to seek advanced strategies for its effective utilization. This shift necessitates data professionals to acquire an expanded range of abilities, encompassing expertise in managing extensive databases, statistical analysis, cloud technology, and programming. Such competencies are crucial for enhancing organizational performance and propelling digital innovation. Nonetheless, the pace at which these essential Data Analytics skills are being developed lags behind the growing demands of the corporate world, highlighting the need for an organized, ongoing approach to skill enhancement. The lack of consensus on the precise definitions of roles within the data-centric workforce, such as Business Analytics, Big Data, and Data Science, further complicates the distinction between these areas. Moreover, within corporations, there is a blending of roles that share common ground, leading us to adopt the term “Data Analytics” as a comprehensive reference to the wide spectrum of professional requirements and growth possibilities in the data realm [2].

Mohamed Almgerbi
University of Perugia, e-mail: almgerbi.mohamed@gmail.com

Andrea De Mauro
Tor Vergata University of Rome, e-mail: andrea.de.mauro@uniroma2.it

Adham Kahlawi
University of Florence, e-mail: adham.kahlawi@unifi.it

Valentina Poggioni
University of Perugia, e-mail: valentina.poggioni@unipg.it

In recent years, the role of Data Analytics has undergone considerable evolution, with a marked shift in the skill set necessary for professional success. Between 2019 and 2023, the emergence of new technologies, shifting priorities, and the growing importance of certain skills have reshaped the landscape. This discussion aims to explore the key dimensions of these changes, with a focus on the evolution of technical and analytical skills.

2 Methodology

To achieve the goal of this research, we have formulated a 4-step methodology.

2.1 Data collection

We utilized a dataset obtained by collating 14,495 job advertisements sourced from the internet in 2019 (as described in [2]) and 13,493 additional entries published in 2023.

2.2 Data preprocessing

We have performed the same data preprocessing routine utilized in our previous work (as described in [2]), which included Data cleansing, Deduplication, N-gram removal, and Stemming. We randomly selected 16,060 job advertisements from the original dataset, ensuring an equal distribution between 2019 and 2023, resulting in 8,030 job advertisements each year. Finally, we created the dictionary and the corpus required by Topic Modelling [4].

2.3 Topic modelling and performance evaluation

We run LDA topic modelling algorithm several times with a different number of topics, k , ranging from 5 to 20, to determine the optimal number of topics [1]. Subsequently, we measured the coherence score for each run and, by comparing these scores, determined that the optimal number of topics would be within the range 9, 12, 13. Finally, based on human judgment, we determined $k = 12$ to be the optimal number of topics in terms of interpretability of the topics [3].

3 Findings

We have identified twelve distinct topics that best represent the underlying themes within our corpus. Each topic was named based on an expert judgement review of the top keywords generated by the LDA algorithms and the content of job posts with higher representation within each category. Below, we briefly describe each identified topic and its thematic focus:

1. **Financial Applications:** This topic encompasses the use of data analytics within the financial sector, highlighting roles that combine analytical skills with financial expertise.
2. **Sales and Marketing Applications:** Focuses on integrating data analytics in sales and marketing applications, underscoring the importance of data-driven decision-making in commercial contexts.
3. **Foundational Statistics:** Represents the core statistical techniques and theories underpinning data analysis, emphasising the mathematical foundations of the field.
4. **CyberSec Applications:** Centers on the critical role of cybersecurity in protecting data assets, reflecting the growing importance of security measures in data management.
5. **Project Management:** Highlights the application of project management principles within data analytics projects, stressing the importance of planning.
6. **Business Intelligence:** Focuses on the tools and techniques for transforming data into actionable insights, showcasing the value of BI in corporate decision-making processes.
7. **Databases:** Covers the technical aspects of database management and administration, reflecting the foundational role of databases in storing and managing data.
8. **Scientific Research Applications:** Emphasizes the application of data analytics in scientific research, highlighting the need for precise and methodical data analysis.
9. **Cloud and Big Data Engineering:** Centers on the technologies and frameworks for managing large-scale data sets in cloud environments, underlining the importance of scalability and flexibility in data storage and processing.
10. **Machine Learning:** Deals with the fundamentals and applications of AI and machine learning, particularly noting the prominence of deep learning techniques and the emergence of Large Language Models (LLMs) as significant areas of focus within the field.
11. **Software Engineering:** Highlights the development of software solutions for data analytics, stressing the importance of coding, development practices, and the application of engineering principles.
12. **Senior Management:** Focuses on the strategic roles within organizations that leverage data analytics for business leadership and decision-making, emphasizing the integration of analytical insights with executive strategies.

Table 1 shows the relative presence of each topic in 2019 and 2023 with the relative percentage change.

Table 1: Topic definition and relative presence in 2019 and 2023, with percentage change.

<i>Skill</i>	<i>2019</i>	<i>2023</i>	<i>Delta</i>
Financial applications	7.60%	8.06%	+6%
Sales and Marketing applications	7.55%	8.43%	+10%
Foundational Statistics	9.06%	8.04%	-13%
CyberSec applications	7.70%	8.15%	+6%
Project Management	8.41%	8.06%	-4%
Business Intelligence	9.25%	8.76%	-6%
Databases	9.53%	6.85%	-39%
Scientific Research applications	7.87%	8.10%	+3%
Cloud and Big Data Engineering	8.82%	8.25%	-7%
Machine Learning	8.24%	9.35%	+12%
Software Engineering	8.20%	9.48%	+13%
Senior Management	7.76%	8.48%	+8%

Based on the trends observed between 2019 and 2023 from the job post data, several conclusions can be drawn regarding the shifting landscape of the data analytics job market:

Domain-Specific Applications Rising: We noticed an increased demand for professionals in Financial applications (+6%) and Sales and Marketing applications (+10%). We believe this indicates a trend towards specialization in data analytics. Companies are seeking individuals who possess not only analytical skills but also a deep understanding of specific industry domains. This reflects a move away from generic analytics roles towards more specialized positions where domain knowledge is crucial for leveraging data effectively.

Shift from Theoretical Statistics to Practical Applications: The decline in Foundational Statistics (-13%) alongside the significant rise in Machine Learning (+12%) suggests that the job market favours practical application of statistics over theoretical knowledge. The growth in Machine Learning, particularly with the increasing prominence of Natural Language Processing (NLP) and Large Language Models (LLMs), points towards a demand for skills that can directly contribute to developing and implementing AI-driven solutions.

Commoditization of Database Management and Cloud Technologies: The substantial decrease in demand for expertise in Databases (-39%) and a decrease in Cloud and Big Data Engineering (-7%) may imply that these skills are becoming commoditized. As cloud services and database management become more user-friendly and accessible, the need for specialized skills in these areas may be diminishing, possibly due to increased automation and the availability of off-the-shelf cloud solutions.

Increased Need for Software Engineering: The marked increase in Software Engineering roles (+13%) underscores the need for robust, agile-based software development skills to support the creation of complex applications. This trend may

be driven by the increasing sophistication of data analytics tools and the need for custom solutions that seamlessly integrate with existing systems.

Governance and Leadership Skills in Demand: The growth in CyberSec applications (+6%) and Senior Management roles (+8%) indicates a heightened need for governance and leadership within organisations. As data analytics matures within companies, there is a greater requirement for individuals who can oversee data governance, manage cross-functional teams, and lead strategic initiatives at higher levels of the organisation.

4 Conclusion

This study identified twelve distinct topics that best represented the dynamics of Data Analytics Job Skill between 2019 and 2023. We found that the job market trends point to an evolving field where there is an increasing demand for specialized data analytics skills that are tailored to specific sectors, coupled with a strong emphasis on the practical applications of machine learning and AI.

Nevertheless, this study is not without its limitations. The reliance on job bulletins may introduce a bias towards skills and roles that are more frequently advertised online, potentially overlooking niche or emerging skills that are communicated through other channels or that yet need to emerge. Future research could expand upon this work by incorporating a wider array of data sources, including direct surveys of industry professionals and analysis of educational curricula to anticipate which skills will be in demand.

References

1. Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., Hassan, A.: Topic modeling algorithms and applications: A survey. *Inf. Syst.* (2023) doi: 10.1016/j.is.2022.102131
2. Almgerbi, M., De Mauro, A., Kahlawi, A., Poggioni, V.: A Systematic Review of Data Analytics Job Requirements and Online-Courses. *J. Comput. Inf. Syst.* **00**, 1–13 (2021)
3. Hasan, M., Rahman, A., Karim, M.R., Khan, M.S.I., Islam, M.J.: Normalized approach to find optimal number of topics in Latent Dirichlet Allocation (LDA). In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds.) *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020*, pp. 341-354. Springer, New York (2021)
4. Lind, F., Eberl, J.M., Eisele, O., Heidenreich, T., Galyga, S., Boomgaarden, H.G.: Building the Bridge: Topic Modeling for Comparative Research. *Commun. Methods Meas.* **16**, 96–114 (2022)