SN Computer Science (2023) 4:696
https://doi.org/10.1007/s42979-023-02065-9

ORIGINAL RESEARCH

An Application of the Analytic Hierarchy Process to the Evaluation of Companies' Data Maturity

Simone Malacaria[1] · Andrea De Mauro[1] · Marco Greco[2] · Michele Grimaldi[2]

Received: 1 November 2022 / Accepted: 13 June 2023
© The Author(s) 2023

Abstract
The aim of this study is to evaluate data maturity of a sample of Italian firms of different sectors and sizes, obtained through an online assessment submitted to 261 professionals and entrepreneurs operating in the data/IT domain. The paper's objective is to assess the relative importance of the factors that determine the success of big data initiatives, according to the company structure and managerial perspective. The questionnaire was digitally submitted to IT professionals and decision-makers in Italy through the LinkedIn platform. The assessment was divided into two sections: the first focused on the assessment of 8 critical success factors for big data, whereas the second assigned weights based on an application of the analytic hierarchy process. The result of this process is a weighted-scores system that reflects the relative importance that managers and employees give to different domains. Respondents agreed to the importance of integrated architecture, data-friendly corporate culture, and integrated organization domains. Once the results consider the weights from the AHP, data friendliness becomes the most sought-after characteristic. The findings provide direction for further development of this assessment system.

Keywords Big data · Data analytics · Analytic hierarchy process · Assessment system · Data maturity models

Introduction

Data science is the set of statistical techniques and methods necessary for the extraction, analysis, and interpretation of data. In the era of "Big Data", where a huge amount of information is available to companies, data-driven choices are essential for defining a company's medium and long-term strategies and can turn into a huge competitive advantage [14, 22]. The major internet and manufacturing companies such as Google, Facebook, and Apple hire the best data science talents to work in their vast data science departments. Being a successful company today means making data-driven decisions [12, 41]. Companies that have overlooked the potential of data science have observed their competitors seize market share and enlarge their customer base over the past years. Pioneers such as Facebook, Amazon, and Google instead developed dominant market positions. Nowadays, companies of all sizes are investing heavily in big data and AI initiatives to narrow the gap with the tech giants [8]. Although the value that data analytics brings to companies has been recognized [14, 15, 25], there is still confusion on how to properly integrate big data initiatives within the organization for long-term planning [24, 43]. This is one the main reason for the failure of more than half of big data programs worldwide [38, 43]. Being a data mature organization means being able to spot new data-driven opportunities in advance, while they are still invisible to the competitors, using analytic insights to deliver business outcomes. Survey results from data-driven companies have shown that decisions made with the help of data are more accurate, faster, and more cost-effective. According to a survey conducted by YouGov and commissioned by Tableau [21], business leaders reported that data-driven decisions lead to better decision-making, improved customer satisfaction, increased revenue, and greater cost savings. In addition, the survey found that data-driven decision-making leads to faster decision-making, with decisions taking days or weeks instead

This article is part of the topical collection "Advances on Enterprise Information Systems" guest edited by Michal Smialek, Slimane Hammoudi, Alexander Brodsky and Joaquim Filipe.

✉ Simone Malacaria
simone.malacaria@students.uniroma2.eu

[1] Department of Enterprise Engineering, University of Rome "Tor Vergata", Rome, Italy

[2] Department of Civil and Mechanical Engineering, University of Cassino and Southern Lazio, Cassino, Italy

Published online: 12 September 2023

of months or years. Furthermore, the survey revealed that data-driven decisions are more accurate, with fewer errors and more accurate predictions. Finally, the survey showed that data-driven decision-making leads to greater cost savings, with data-driven companies spending less money on research and development.

In this study, we analyze the data maturity of 261 Italian companies of different sectors and sizes to answer the following business questions:

- RQ1 What are the data maturity estimators considered most important by the managers interviewed for the success of big data initiatives?
- RQ2 Does the relevance of data maturity estimators to industry experts differ depending on the size of the company or industry?

To respond our research questions and score what the ripeness level of the enterprises is, we relied on an eight-dimensional assessment system derived from the literature De Mauro et al. [9]—the CBDAS—consensual with the existing big data maturity models. The CBDAS applies the analytic hierarchy process to assign weights to the critical success factors for big data initiatives. We analyze how respondents (Senior Manager and IT decision-makers) agreed or disagreed with questions that underlined the importance of each success factor proposed. As a result, the paper derives insights on the importance that the managers give on data-driven choices and on the validity of the CBDAS to apply to companies of different sizes and industry sectors.

Big Data Maturity Models

The big data maturity models represent robust frameworks that support the evaluation of old and new big data initiatives among specific aspects or domains to rule whether they can generate new knowledge for a company [14, 28, 32–34].

These models help organizations to create a structure around their big data initiatives and provide guidance on how to structure processes, technologies, and analytics to achieve success [1, 14].

Big data maturity models are comprised of different stages that guide organizations in understanding their current level of big data initiatives and how to move forward with their big data strategy. These models follow a logical progression through defined levels, starting from a foundational stage of weakness to reaching the ultimate goal of full maturity [31]. Each stage is identified by criteria that must be met to advance to the next level. These criteria build upon each other, with each stage requiring the fulfillment of lower-level criteria and additional requirements. This structure enables business managers to determine the necessary steps to achieve further improvement [6, 16, 27].

As organizations mature in their big data journey, they gradually gain the ability to effectively utilize big data to their advantage. The process typically starts with an awareness stage where organizations first become aware of big data and its potential, then progress to a stage where they start building their big data infrastructure and capabilities. With increasing proficiency, organizations start to see success in their big data initiatives and make data-driven decisions. Finally, when organizations reach a mature level in their big data initiatives, they have the ability to harness advanced analytics and machine learning to drive growth and gain a competitive advantage [5, 13].

Overall, big data maturity models provide organizations with the structure to create strategies and achieve success in their big data initiatives. By leveraging these models, organizations can gain a better understanding of their current state and make changes to improve their big data initiatives.

By leveraging a big data maturity model, data maturity can be evaluated at the sub-domain level when it refers to micro-level factors, such as routines and organizational requirements, at the domain level when it refers to the macro-level factors to assess the needed conditions to reach maturity stages. While macrolevels generally assess strategic factors of big data initiatives' success, microlevels make clear the actions to be taken to guide maturity within organizations [6, 16, 27].

The aspects investigated through the maturity models can be many, such as IT management, business intelligence ecosystem, and data warehouse adoption, among others. In general, big data maturity models give the company the maximum value when used to analyze how business processes and strategies integrate with big data initiatives, providing management with the needed information to support strategic and operational decisions [2].

Data models help to outline the optimal choices for a path of improvement of the business management system. The absence of specific procedures regarding the assessment and operation of maturity models may represent a limitation for the use of the model as an organizational and diagnostic-prescriptive management system.

So far, only a few of the big data maturity models present in the literature contain details on the development, validation, and evaluation processes of the model itself, constituting a limit to the validity and usefulness of many proposals [31, 33]. We rely on the Consensual Big Data Assessment System (CBDAS) proposal [9], which starts from a holistic and conceptual integration of existing models. It encompasses the key elements of success that are coherent with big data's essential components and agrees with the most cited models in literature [2, 17]. The CBDAS offers a robust conceptual framework complemented by a practical assessment

and recommendation system to grant usefulness and applicability for industries.

The CBDAS framework was leveraged as judged from the authors the most comprehensive big data maturity model after a comparison of various models was conducted. Previous research, including the systematic reviews by Comuzzi and Patel [6] and Olszak and Mach-Król [28], and dos Santos-Neto and Costa [33], revealed that only a limited number of maturity models effectively illustrate the maturity stages, which is crucial for evaluating an organization's performance. These models are varied in scope, with some focusing on specific aspects and others aiming to provide a holistic assessment. Various maturity models have been proposed for specific aspects of big data, such as IT management [4] and business intelligence [18], Mikalef et al. [26], or data warehouse adoption. However, it is widely acknowledged that the scope of big data encompasses much more than these specific areas and extends to the operational processes of organizations and the support of decision-making processes [6, 12, 30, 40]. The multidisciplinary nature of big data makes aspect-specific maturity models insufficient for a comprehensive and holistic evaluation [6, 36].

Furthermore, the lack of alignment between maturity models and assessment models can cause confusion for users and hinder their use as organizational management tools (2019).

To overcome these issues, big data maturity models need to broaden their scope to include industry surveys and gray literature, and an assessment tool is required to complement the maturity model to make it practically applicable in industries [2, 37].

Different authors aimed to make the constitutive elements of their maturity model tangible and measurable to ensure its practical applicability [4, 26, 42].

In general, big data maturity models are mostly used as a descriptive tool, assessing the current situation against a framework, while only a minor part of these models provides normative, prescriptive recommendations for performance improvement [2, 20].

The completeness of the maturity model in terms of scope is a critical factor to consider for the implementation of big data programs in real-world industries as it ensures its practicality and actionability [5, 13]. The CBDAS provides both a descriptive tool for assessing the current state against a framework and a normative, prescriptive component for performance improvement recommendations. It is aligned with the key elements of big data and is consistent with widely recognized big data maturity models. This assessment system has the potential to drive future research by providing clear understanding of the critical success factors for big data programs and the specific requirements for each subfactor to achieve organizational success, fulfilling the call by researchers for consolidation of existing maturity models rather than creating new ones [4, 26, 42].

The Consensual Big Data Assessment System (CBDAS)

The CBDAS is a consensual framework designed on a theoretical framework that outlines the key success factors (CSFs) of data analytics. This framework encompasses the factors and sub-domains, and the criteria of success of big data initiatives. The structure of the CBDAS consists of two sections: the first section assesses the current maturity levels of enterprises at various levels of granularity—from factor to subdomain level—while the second section evaluates the relative importance of the CSFs for industry experts. The assessment system is comprised of 59 questions, with 44 questions focusing on the maturity levels of the CSFs and sub-domains and 15 questions addressing the underlying business needs. The results are obtained through a standardized questionnaire with one closed-ended question per item and analyzed using software, such as Cognito, Zapier, and Google Sheets. The CBDAS is a quantitative method that generates 156 potential outcomes based on the combination of sub-domain maturity levels and their interdependencies. The second section of the questionnaire evaluates the importance of each factor using the analytic hierarchy process (AHP) which evaluates individuals' and groups' judgments and converts them into numerical values for objective comparison. The output of the AHP is used as input to evaluate the relevance of each organizational area from the manager's perspective, recognizing that the relevance of each factor may vary depending on context and may not be applicable to all organizations.

Methodology

Our aim was to explore the critical data maturity indicators that managers consider for the success of big data initiatives and determine if their importance changes based on the company size or industry. To achieve our scope, we gathered input from industry experts through the administration of the CBDAS framework.

We leveraged multilinear regression and Principal Component Analysis (PCA) to understand which domains are strong predictors of the final questionnaire score, we utilized. PCA helps uncover the fundamental structure of the data, while multilinear regression reveals the relationships between variables.

Then, we applied the Analytical Hierarchy Process (AHP) to assess the experts' subjective perception of the relative importance of each domain and answer our RQ1.

AHP assigns weights to factors to determine their relative significance.

We conducted the Analysis of Variance (ANOVA) tests to check for statistical differences among groups and address our RQ2. ANOVA tests the null hypothesis that the averages of different groups come from the same distribution, determining if there is a significant difference in the means of two or more groups.

To summarize, we employed PCA, multilinear regression, AHP, and ANOVA tests to analyze the data from industry experts and address our research questions. These methods were selected for their appropriateness to the study and their widespread use in the field of data analysis.

Data Collection

In this work, we analyzed the responses of 261 employees from Italian companies to a Likert-scale questionnaire (ranging from 1—strongly disagree to 5—strongly agree), where the participants were asked to answer questions that measure a company data maturity. The participants were mostly company managers and IT experts (139 managers, 122 IT experts). We used the LinkedIn platform to draw a sample of professionals worldwide to conduct the online assessment. Although the LinkedIn community is not encompassing the population of industry representatives exhaustively, it might be considered suitable for targeting professionals in scope. The process of sample selection leverage publicly available information about the respondents provided by LinkedIn users, which increases the credibility of the sample and permits control over its composition.

The inclusion criteria were related to:

(a) the seniority of the respondents: Senior Managers and Directors;
(b) the position covered in their organization: IT Director, IT Responsible, IT Specialist, Senior Data Specialist, Senior Data Scientist, Senior Business Analyst, IT Consultant;
(c) their confirmed experiences and skills in the areas of Data Analytics, Big Data, and IT Management.

With the inclusion criteria defined, we identified a potential audience of over 320,000 unique respondents who were targeted through LinkedIn campaigns conducted from October to December 2021. To ensure the validity and reliability of the data collected, we applied exclusion criteria to exclude (a) incomplete assessments and (b) companies operating outside of Italy, as the focus of the study was on the Italian market. Our sample was derived from a starting pool of 297 respondents, and ultimately consisted of 261 complete responses.

Assessment Structure

The CBDAS assessment was structured in two parts: the first part is composed of 40 questions divided into 8 domains and allows the evaluation of data maturity on critical success factors for big data initiatives; the second part is made of 15 questions that focus on the pairwise comparison of data maturity characteristics of the company, which represent a multifactorial combination of the 8 critical success factors.

In the following chapter we present the results of the survey.

Questionnaire Results

The questionnaire results are summarized in Figs. 1 and 2, where the respondent proportion for each question is represented as vertical bars of distinct colors. The full questionnaire answers are available at https://github.com/SimoneMal/AHP_dataset for consultation.

As shown in Figs. 1 and 2 more than half of the respondents agreed, or strongly agreed to the importance of "Integrated architecture", "Data Friendly" and "Integrated organization" domains. This can also be seen in Fig. 2, where the sum of the Likert score is shown. This result is not surprising, since data friendliness and an integrated data architecture in a company are of paramount importance for making data driven decisions. Data integration in a company involves connecting data sources across different systems and applications to create a unified view of data. This process is essential to ensure the accuracy and reliability of data, and to enable data-driven decision-making and helps improve data quality and reduce data silos. To achieve data integration, companies can use a variety of methods, such as data federation, data replication, data virtualization, and data warehouse consolidation. Data friendliness in a company refers to the collective beliefs and behaviors of the people working in the organization regarding the use of data. To create a data friendly culture, organizations should ensure that everyone in the company is aware of the importance and value of data and data-driven decision-making. In addition, organizations should provide data literacy training and workshops, establish data governance policies and processes, and create an environment, where data are easily accessible, trusted, and used to drive decisions. An integrated

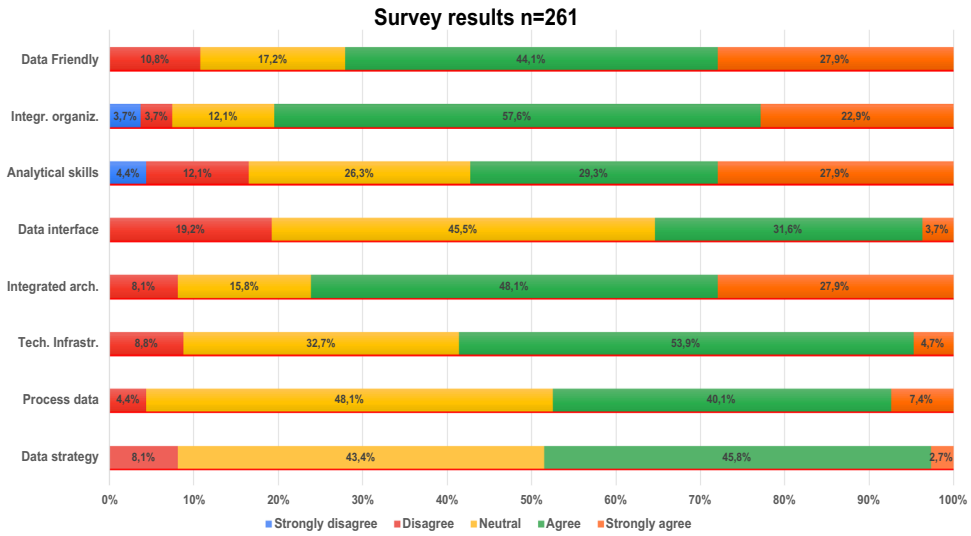


Fig. 1 Relative percentage of answers for each of the eight domains

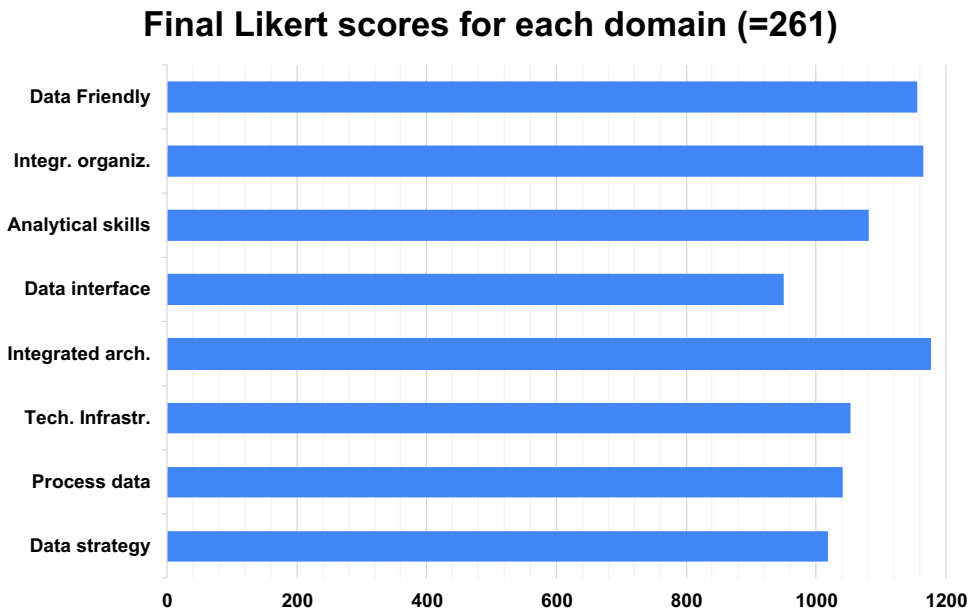


Fig. 2 Final Likert scores for each domain. This graph compares the sum of the Likert scores given by each of the 261 people interviewed

organization in a company refers to a unified approach to the management of data and resources across all departments and business units. This approach ensures that all data are collected, stored, managed, and shared in a consistent and reliable manner.

Notably, most of the respondents have considered the "Data interface" domain the least important among the eight. A data interface is an interface that provides access to data by allowing users to interact with the data in some way. It may consist of web app, e standalone app with (GUI) or a command-line interface (CLI). Data interfaces allow users to view and manipulate data in a structured and organized way. The reason why the respondents have considered this domain less important than the others is maybe due to the omnipresence of such interfaces and the unquestionable high level of user friendliness achieved in the last years.

Correlation Among Parameters

We calculated a correlation coefficient matrix between every domain listed in Table 1. To do that, we used the Spearman correlation coefficient [35], which has the advantage of not being limited to continuous numerical variables but can also be applied to discrete ordinal variables. Moreover,

Table 1 Assessment submitted to the participants

Domain	Questions
DATA STRATEGY	(1) The company has a solid data analysis strategy
	(2) The company uses data analysis to make strategic decisions
	(3) Data analysis is not an important part of the company's transformation strategy
	(4) The Corporate Data Strategy has been documented, approved, and communicated by Top Management to the entire organization
	(5) Leadership promotes the use of data analytics throughout the company
	(6) There is a list of key analytical projects or analysis priorities whose progress is regularly tracked
	(7) The legal procedures on data usage and management are documented and communicated to the entire organization
	(8) There are regular audit processes on data usage and management within the organization
DATA-PROCESS INTEGRATION	(9) Business processes are guided by numerical evidence, which directly impacts the way the company operates
	(10) The Key Performance Indicators related to data processes are stored and could be analyzed in real-time
	(11) The organization uses automated analyses (e.g., systems that suggest in-depth analysis or build models, alert systems based on control levels, reports that automated data processing and output delivery)
	(12) Your company has organizations (internal or external) that focus on data engineering, software development, data quality to ensure proper support to analytical processes
	(13) Managers and process owners know what data are available in the company to support their business decisions
TECH INFRASTRUCTURE	(14) The data infrastructure is adequate to the size of the organization, and the organization is using the following types of data management technology where needed: cloud systems, big data architectures
	(15) The organization is able to monitor more data pipelines. Therefore, the organization is able to manage multiple analytical projects in parallel
	(16) The organization has designed its data architecture to integrate multiple sources and facilitate data access and analysis
	(17) The computational power and the size of the available memory are adapted to the current information injections
	(18) Systems comply with high-security standards and are subject to periodic intrusion tests
	(19) How many of the following technologies use your organization to analyze your data? (Spreadsheets, reports, dashboard, predictive analysis/machine learning, deep learning, and other aspects of the AI)
	(20) Only a few managers in the company have access to the analysis results
INTEGRATED ARCHITECTURE	(21) The organization collects and manages structured (i.e., sales data in tabular format) and unstructured (i.e., Video and Audio) data types for its analysis activities
	(22) Employees can access data as needed, including structured and unstructured data, through a well-defined governance process
	(23) The data formats are standardized and documented
	(24) There is a single data model to which all the business units refer
DATA INTERFACE	(25) Analysis solutions are designed to provide the best user interface to the right person (for example, corporate analysts, business users, data scientists, data engineers, et al.)
	(26) Employees do not receive guides on how to access the data
	(27) It is easy to get data in a format not covered by existing interfaces; the technical support needed is minimal, and the request is standardized
	(28) Corporate data are accessible through a business intelligence interface that allows users to combine different data sources, create graphs and tables with a high degree of customization, and allows users to share the most interesting views with other stakeholders
	(29) Data Scientists and Business Analysts are able to connect any application developed with the latest data available at the needed level of detail
ANALYTICAL SKILLS	(30) The knowledge of data science techniques is widespread in the organization, even outside the business units dedicated to data analysis
	(31) The analysts use tested and documented tools and methodologies to respond to the business questions of the organization, which requires analytical support
	(32) There is a career model for Business Analysts and Data Scientists
	(33) The company has a clear recruitment strategy for data professionals

Table 1 (continued)

Domain	Questions
	(34) There is a broad and modular program for analytical skills development open to all employees and modulated according to career aspirations and personal interests
	(35) The organization invests in the training of data analysts
INTEGRATED ORGANISATION	(36) Business Analysts and Data Scientists operate in an integrated manner with the rest of the organization
	(37) Analysts are in contact with corporate opportunities and challenges; they can directly impact decisions and influence the corporate strategy
	(38) Analysts' priorities are defined according to urgencies and are not linked to the company's opportunities
DATA FRIENDLY APPROACH	(39) The entire organization is pervaded by widespread knowledge and live interest in data, and the role of analytics is recognized in guiding the company to success
	(40) Top Managers are consistently leveraging the recommendations generated by data and algorithms

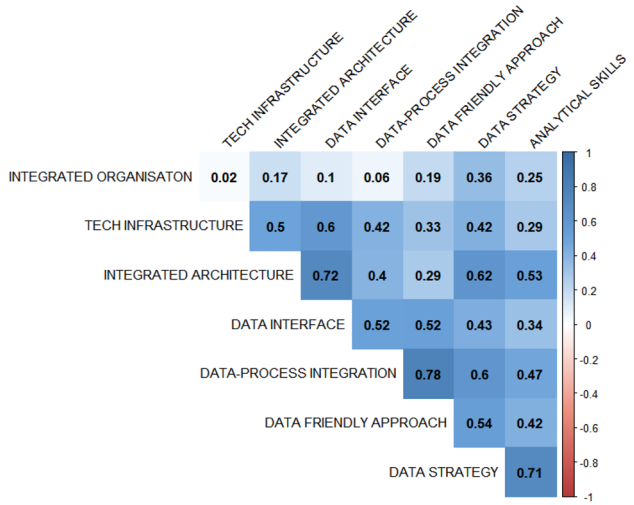


Fig. 3 Spearman correlation matrix among domains scores

this method can spot strictly non-linear correlations and can assess how much two variables are correlated by a monotonic function [44]. For linear relationships, the two methods give similar answers. The value of Spearman's R is between -1 (indicating a perfect negative correlation) and $+1$ (indicating a perfect positive correlation). Weak correlations have R values $0 < |R| < 0.2$, moderate correlations $0.2 < |R| < 0.6$ and strong correlations $0.6 < |R| < 1$ We have created a correlation Matrix using the software R 4.1.2 and the command *rcorr*. The results are shown in Fig. 3.

It was to be expected that only positive correlations had to be found, since all questions in the Likert scale go in the same direction. The strongest correlations happen to be between data process integration and data-friendly approach, where $R = 0.78$ indicates a strong correlation. It also appears a strong correlation between the domain integrated architecture and data interface with a Spearman's $R = 0.72$. A strong correlation among parameters is also referred to as multicollinearity.

Multicollinearity Analysis

To deepen the analysis of the correlation among parameters, we have examined the phenomenon of multicollinearity in our data. Multicollinearity refers to the high correlation between predictors in a multiple regression model, which can hinder statistical analysis and affect the robustness and accuracy of coefficient estimates. To mitigate this challenge, it is necessary to detect and address multicollinearity. This can be achieved through a dimensionality-reduction technique called Principal Component Analysis (or PCA) [19, 29].

The first step in this kind of analysis is understanding how many components are needed to explain most of the variance in the sample. To this purpose a useful tool is the scree plot (see Fig. 4), that depicts the explained sample variance as a function of the number of principal components.

A rule of thumb criteria to understand how many components are really needed to retain most of the information and reduce complexity, is the so-called elbow method. With the elbow method one looks at the start of the bend in the line ("elbow"), that in our case happens at dimension n. 3. We may, therefore, expect that in our case just three domains will be able to explain $> 80\%$ of the sample variance.

The biplots plot shows the domain variables as vectors (arrows). These vectors can be interpreted as follows [10]:

Angle—The angle that each vector makes with an axis is related to how much that variable contributes to that principal component, the smaller the angle the greater the contribution. In our case the process_data variable contributes the most to dimension 2 (and dimension 2 contributes for the 15.1% to the total variance), while the domain process_data contributes negatively to dimension 2.

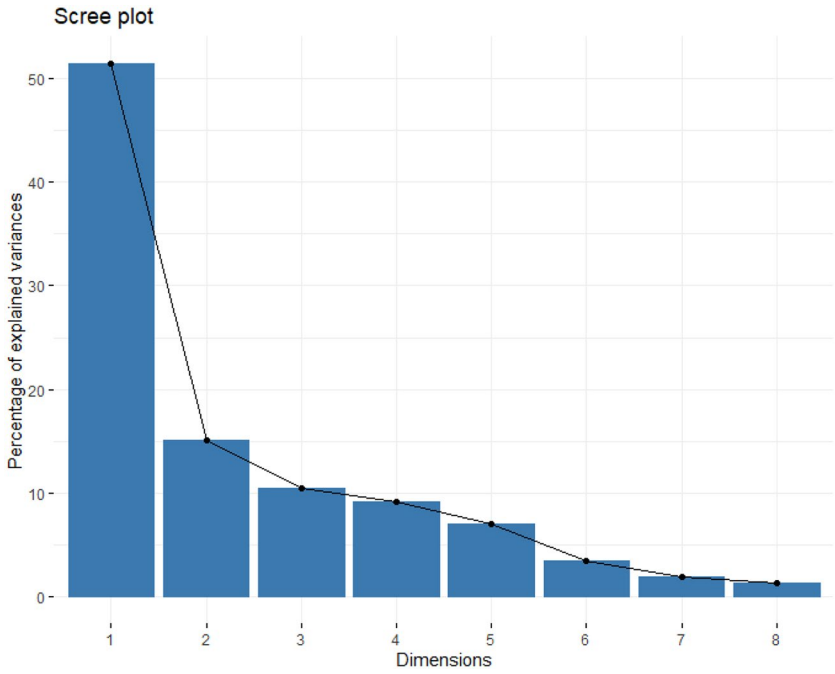


Fig. 4 Scree plot of the principal components. The elbow method suggests that the first 3 PC are enough to explain > 80% of the sample variance

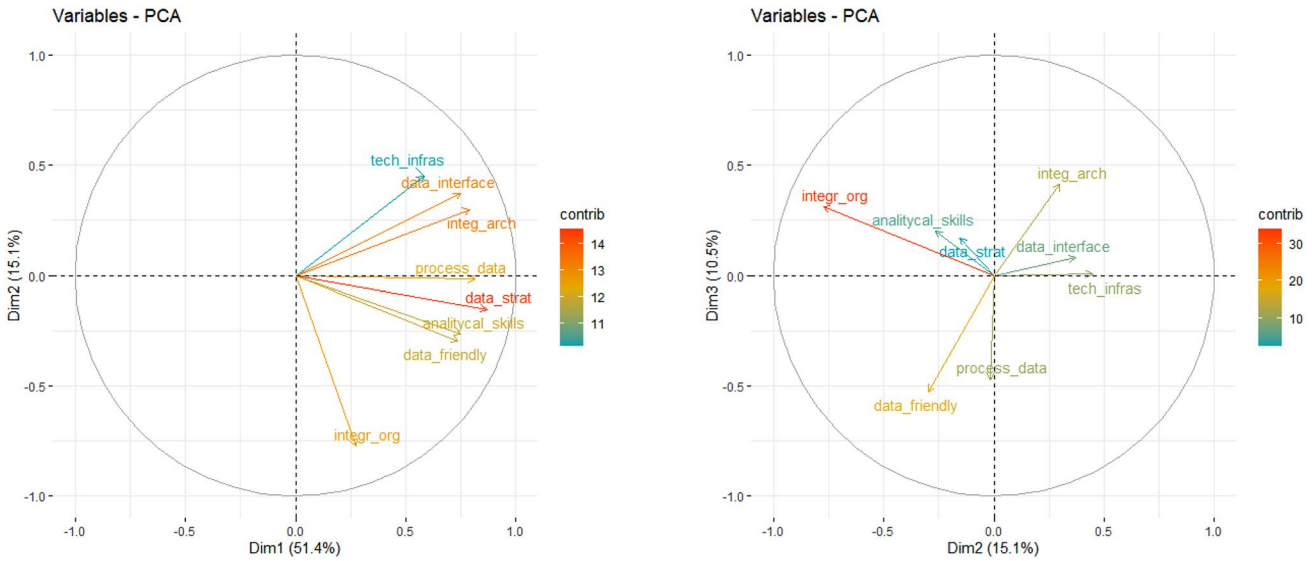


Fig. 5 Biplots of the three principal components. The scalar projection of the single vector on each axis represents the contribution of that variable to the principal component. For ease of visualization a colour scale proportional to the net contribution has been added

Length—The longer the vector the more variability represented by that variable. In our case the variable data_strat contributes the most to the variance of dimension 1.

Directions—The angles/direction between vectors of different variables are proportional to their correlation, the closer the direction the greater the correlation of those parameters. As can be also seen in Fig. 5 the domains Data Interface (data_interface) vs Integrated Architecture (integ_arch) and Analytical Skills (analytical_skills) vs Data Friendly (data_friendly) are highly correlated.

The colours in the above plot represent the total contribution of each variable to the total variance. The domains that contribute the most are in order Data Strategy (data_strat) Integrated Architecture (integ_arch) and Data Interface (data_interface). Together these variables explain > 85% of the total sample variance.

To confirm this result, we performed a multilinear regression of these three variables on the total score (Fig. 6).

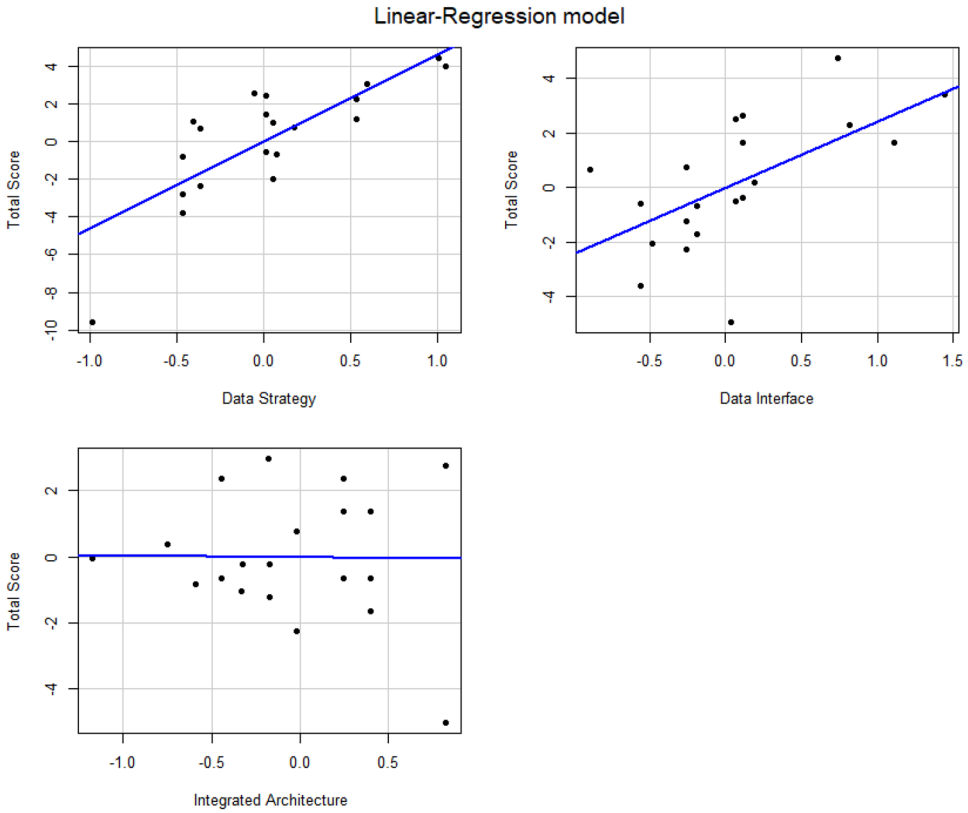


Fig. 6 Graphical representation of the multilinear regression obtained with the R "lm" command. The plotted points are 20 rather than 297, because since the questionnaire has discrete values, the points are overlapped

```
Call:
lm(formula = total_score ~ data_strat + data_interface + integ_arch,
    data = df_scores)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0028 -1.0504 -0.2449  1.3782  2.9536

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.65326    0.56090  10.079   <2e-16 ***
data_strat     4.62304    0.20976  22.039   <2e-16 ***
data_interface 2.42456    0.19200  12.628   <2e-16 ***
integ_arch    -0.04355    0.21169  -0.206    0.837
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.759 on 293 degrees of freedom
Multiple R-squared:  0.861, Adjusted R-squared:  0.8596
F-statistic: 605.1 on 3 and 293 DF,  p-value: < 2.2e-16
```

The coefficient of determination $R^2 = 0.86$ confirms that alone these three variables combined can explain 86% of the sample variability.

From the PCA and multilinear regression emerges that the Data Strategy, Data Interface, and Integrated Architecture domains are good predictors of the questionnaire final score. This means that the broad range of questions posed in these domains are similar in nature to those of other domains and correlate well with other questions. It should be noted that this result does not evaluate the respondent subjective perception of the domains relative importance (evaluated

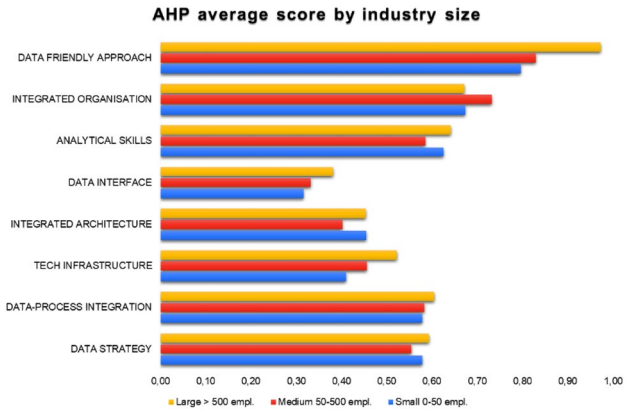


Fig. 7 AHP-weighted average score by industry size. The results look very similar among different company sizes

through AHP) but is only a reflection of the relative correlation among domains.

Stratification by Company Sector and Size

A series of demographic questions were asked to the participants when the assessment was submitted to them. We collected information about the characteristics of the company to which they belonged. The questions were focused on the company size and the sector in which it operates. This allowed us to stratify for such parameters and search for statistically significant differences, to answer our RQ2.

The results of this stratification are shown in Figs. 7 and 8.

The stratification by company size shows that the only domain, where there could verify differences among the different-sized company is the data-friendly approach. However, such differences must be ascertained by means of appropriate statistical tools. We performed an ANOVA (ANalysis Of VAriance) test [11] to search for statistical differences among groups. ANOVA tests the null hypothesis that the averages of the groups belong to the same distribution by testing the variance between and within groups. We first tested for significant differences among domains, the results of which are shown in Table 2.

Since the p value is $p << 0.05$, we can reject the hypothesis that the different domains have equal means (i.e., there are significant differences). Regarding the scores of companies of different sizes, the results were opposite and are summarized in Table 3.

Since the p value $>> 0.05$, we observe no statistically significant differences among different company sizes in this case. This can be interpreted as companies of varied sizes having the same data-maturity aspirations and ambitions. The stratification for the company sector also shows similar results. Figure 8 and Table 4 show no statistically significant differences in data needs and maturity scores among companies operating in different sectors.

The Analytic Hierarchy Process

Introduction to the AHP

The AHP is a structured technique for organizing and analyzing complex decisions. It was developed by Thomas L. Saaty in the 1970s, and is based on the concept of hierarchies, which are structures that can be used to represent a problem in terms of its components and sub-components. AHP uses hierarchical decomposition to break down a complex decision problem into its component parts, and then

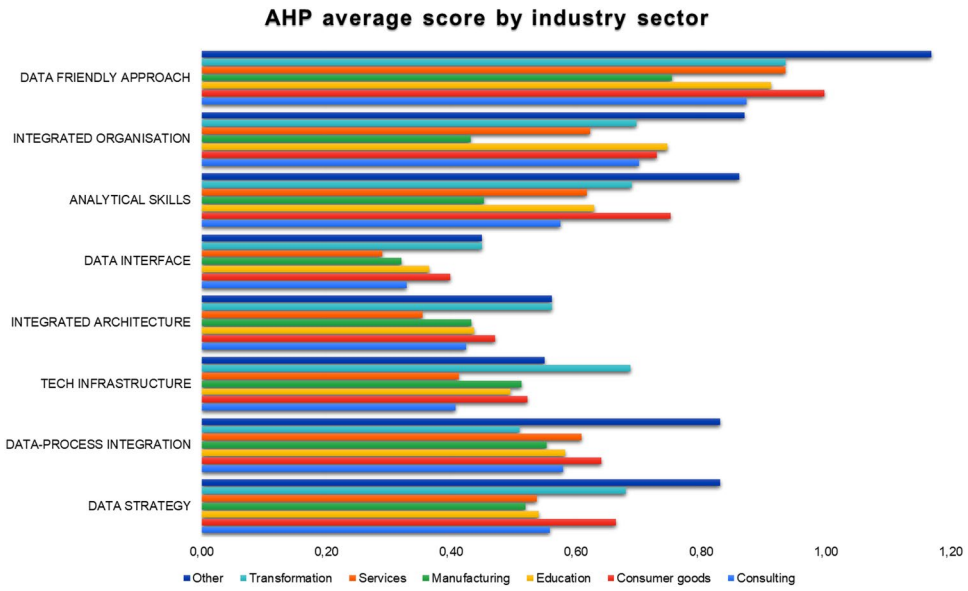


Fig. 8 AHP-weighted average score by industry sector. The results look similar among different company sizes

Table 2 ANOVA output

ANOVA						
SUMMARY						
Groups		Count	Sum	Average		Variance
DATA STRATEGY		3	1.726453388	0.575484463		0.000372
DATA-PROCESS INTEGRATION		3	1.767461022	0.589153674		0.000201
TECH INFRASTRUCTURE		3	1.389559865	0.463186622		0.003198
INTEGRATED ARCHITECTURE		3	1.310043357	0.436681119		0.000883
DATA INTERFACE		3	1.030418065	0.343472688		0.001125
ANALYTICAL SKILLS		3	1.852298318	0.617432773		0.000828
INTEGRATED ORGANISATION		3	2.077575264	0.692525088		0.001193
DATA FRIENDLY APPROACH		3	2.597837495	0.865945832		0.008814
ANOVA						
Source of variation	SS	*df*	MS	*F*	*p* value	*F* crit
Between groups	0.557008477	7	0.07957264	38.31854	**7.88E−09**	2.657197
Within groups	0.033225746	16	0.002076609			
Total	0.590234223	23				

The p value $<<0.05$ indicates that we can reject with a high degree of confidence the hypothesis that the average scores of companies by DOMAIN are the same

Table 3 ANOVA output

ANOVA (company size)						
SUMMARY						
Groups		Count	Sum	Average		Variance
Small 0–50 employees		8	4.433638116	0.554204765		0.023713
Medium 50–500 employees		8	4.47730905	0.559663631		0.027267
Large > 500 employees		8	4.840699608	0.605087451		0.031556
ANOVA						
Source of variation	SS	*df*	MS	*F*	*p* value	*F* crit
Between Groups	0.012485788	2	0.006242894	0.226917	**0.798916**	3.4668
Within Groups	0.577748435	21	0.02751183			
Total	0.590234223	23				

The p value $>>0.05$ indicates that we cannot reject the null hypothesis that the average AHP-weighted scores of companies by SIZE are the same

uses a combination of mathematics and psychology to evaluate them and come up with an optimal solution.

The AHP methodology consists of three main steps, namely, pairwise comparisons and local priority vectors, weighted super matrix, and super matrix formation and calculation. In the pairwise comparison step, the user compares each component to every other component and rates them according to their relative importance. In the weighted super matrix step, the relative importance scores from the pairwise comparison step are combined to generate a relative importance score for each component. Finally, the relative importance scores are used in a mathematical optimization process to generate the optimal solution to the decision problem. This solution can then be used to make an informed decision and select the best option.

Here is an example of how the AHP can be used to decide. Let us say we are trying to decide which car to buy. We may have several criteria that we want to consider when making our decision, such as price, safety, performance, and style.

First, we would break down the problem into its component parts. In this case, the component parts would be the criteria that we want to consider when making our decision.

Next, we would use the pairwise comparison method to rate each component relative to every other component. For

Table 4 ANOVA output

ANOVA (company sector)						
SUMMARY						
Groups	Count		Sum	Average		Variance
Consulting	8		4.445729	0.555716		0.030538
Consumer goods	8		5.172092	0.646512		0.035667
Education	8		4.705451	0.588181		0.030794
Manufacturing	8		3.972509	0.496564		0.015973
Services	8		4.37681	0.547101		0.040978
Transformation	8		5.208651	0.651081		0.022189
Other	8		6.123635	0.765454		0.054318
ANOVA						
Source of variation	SS	*df*	MS	*F*	*p* value	*F* crit
Between Groups	0.379039697	6	0.063173	1.918852	0.096414	2.290432
Within Groups	1.613199204	49	0.032922			
Total	1.992238901	55				

The p value > 0.05 indicates that we cannot reject the null hypothesis that the average AHP-weighted scores of company sector are the same

example, we would rate the performance criterion relative to the price criterion, the safety criterion relative to the style criterion, and so on.

Finally, we would combine these ratings and generate a relative importance score for each criterion. The relative importance scores can then be used in a mathematical optimization process to generate the optimal solution to the problem.

In this case, the optimal solution would be the car that best meets all our criteria.

Mathematical Framework and Results

The AHP process is a quantitative method for making decisions based on the relative importance that people arbitrarily assign to certain factors. This process requires answering pairwise comparison questions structured in the following way:

(A) is X times more important than (B)

Or

(A) is as important as (B)

Or

(B) is X times more important than A

Weights for the single criteria are then computed according to the formula:

$$V_1 = \sqrt[3]{x_{11} * x_{12} * x_{13}},$$

which is the geometric mean (criterion 1) of the question's scores. The weights are then calculated by dividing each geometric average for the sum of all the other criteria:

$$W_1 = \frac{V_1}{V_1 + V_2 + V_3}.$$

The 15 questions in the second section of the CBDAS require the respondent to choose between a pairwise comparison of data maturity characteristics and how much it counts on a specific aspect vs one other to improve big data management, according to the organization's reality.

In our specific case, the AHP process was used to evaluate the following company characteristics, derived from the conceptual CBDAS:

1. The proliferation of a data culture across the entire organization.
2. Availability of IT services.
3. Managers' support in data-driven projects.
4. Care of analytical talents within the company.
5. Satisfaction of technological needs.
6. Business sponsorship to facilitate data-driven decision-making.

The respondents were allowed to rank one of the options from equally important to three times more important,

Table 5 AHP matrix results based on the interview of 261 Italian managers and entrepreneurs

	Proliferation of a data culture across the entire organization	Availability of IT services	Managers' support in data-driven projects	Care of analytical talents within the company	Satisfaction of technological needs	Business sponsorship to facilitate data-driven decision making
Proliferation of a data culture across the entire organization	1.0000	1.6061	0.9394	1.0000	1.6902	0.5926
Availability of IT services	0.6226	1.0000	0.5000	0.5000	1.2677	0.5505
Managers' support in data-driven projects	1.0645	2.0000	1.0000	0.8451	0.8923	0.8165
Care of analytical talents within the company	1.0000	2.0000	1.1833	1.0000	0.9646	0.6246
Satisfaction of technological needs	0.5916	0.7888	1.1208	1.0366	1.0000	0.6768
Business sponsorship to facilitate data-driven decision making	1.6875	1.8165	1.2247	1.6011	1.4776	1.0000

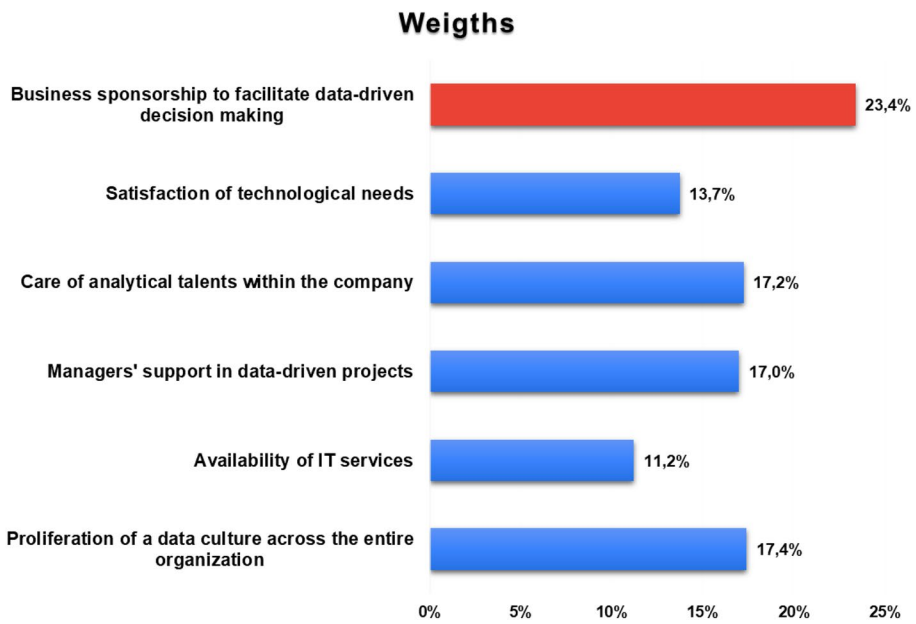


Fig. 9 AHP weights for each of the dimensions chosen in this study. According to more than 261 managers interviewed, the key factors for a company's data maturity are that the business facilitates data-driven decisions and the proliferation of data culture across the entire organization

according to the respondent's perspective on its organization reality. It is crucial to figure out that not all the domains could always be relevant to a particular context [40]. For the same reason, certain factors may be more important than others in specific sectors. To consider these possibilities we allowed the respondents to assign different weights to each organizational need in this section of the assessment. The resulting AHP super matrix is shown in Table 5.

Application to the Survey

The associated weights are depicted in Fig. 9:

The results showed in Fig. 10 answer our RQ1 on the aspects which are judged as most important by the respondents. Looking at the figure, one can see that once one considers the weights from the AHP, data friendliness in the organization becomes the most sought-after characteristic,

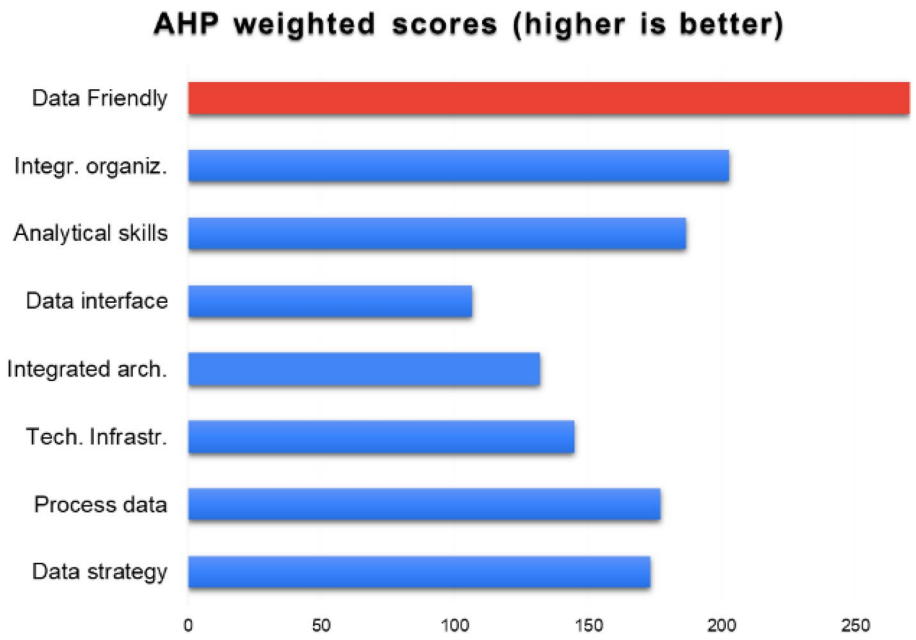


Fig. 10 Final scores weighted using the AHP process. The results now look quite different, since the domain "data-friendly" was considered the most important (higher weights) in the AHP process

followed by "Integrated Organization" and "Analytical Skills."

Discussions and Conclusions

We created an AHP-based evaluation system for estimating companies' data maturity and the importance that their managers and IT professionals assign to data-driven choices. Our results answer RQ1 and suggest that the data-maturity estimator that is considered as the most important by the respondents was "Data friendliness", followed by "Integrated Organization" and "Analytical Skills". Moreover, we found that the relevance of the eight critical success factors included in CBDAS is statistically independent of the size of the company and the sector in which it is operating, making the assessment of general applicability for a broad range of business organizations, responding to our RQ2. Our findings suggest that, when companies look for new opportunities to use analytics, the presence of data-driven culture is of primary importance for making data initiatives able to generate business value [23, 39]. Data culture in a company refers to the collective beliefs and behaviors of the people working in the organization regarding the use of data. Organizations should ensure that everyone in the company is aware of the importance and value of data and data-driven decision-making. To build a successful data culture in a company, it is important to provide data literacy training and workshops, establish data governance policies and processes, and create an environment, where data are easily accessible, trusted, and used to drive decisions. In addition, organizations should incentivize data-driven decision-making and recognize individuals who demonstrate a commitment to data-driven practices. We believe that managers' support rule should be promoting a broad sense of data ownership by all employees and a solid connection between data professionals and business functions [3, 6]. This enables data experts to directly impact business decisions and influence business strategy. By having top managers seeking advice from data analysts, organizations recognize and accept the central role of data in decision-making, business transformation, and innovation. Our research also highlighted how the characteristics identified by managers as relevant (i.e., corporate culture) do not correspond linearly to those with a higher degree of maturity. This mismatch between managers' perceptions and the implementation of concrete actions suggests the usefulness of a system of recommendations for bridging the existing maturity gap in higher priority areas.

The current study is affected by some known limitations that provide opportunities for future research. First, the limited sample size requires the assessment to be tested with a broader audience involving a larger number of enterprises respondents to confirm preliminary insights obtained from the current analysis. Second, the scope of the interviewed audience was limited to Italy, causing its findings to be prone to specific local dynamics. Third, more robust qualitative research is needed to assess the sufficiency of the critical success factors included in the assessment model that was used in this study. A future direction of the study would be to create a specific model for different company contexts capable of thoroughly evaluating how every aspect of data management change according to the complexity of the organizational network [7, 13]. This will allow increasing the practical applicability of the rule-based recommendations,

obtaining specific indications to be implemented in the process of improving business choices.

Funding Open access funding provided by Università degli Studi di Roma Tor Vergata within the CRUI-CARE Agreement.

Data Availability The full questionnaire answers related to this study are publicly available for consultation at https://github.com/SimoneMal/AHP_dataset.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Adrian C, Abdullah R, Atan R, Jusoh YY. Towards developing strategic assessment model for big data implementation: a systematic literature review. Int J Adv Soft Comput Appl. 2016;8(3):173–92.
2. Al-Sai ZA, Abdullah R, Husin MH. A review on big data maturity models. Institute of Electrical and Electronics Engineers, Jordan international joint conference on electrical engineering and information technology, JEEIT 2019. Amman, Jordan. 2019. p. 156–61. https://doi.org/10.1109/JEEIT.2019.8717398.
3. Bahjat E-D, Koch V, Meer D, Shehadi RTU, Tohme W. Big data maturity: an action plan for policymakers and executives. Weforum. 2014. http://reports.weforum.org/global-information-technology-report-2014/. Accessed 12 June 2021.
4. Becker J, Knackstedt R, Pöppelbuß J. Developing maturity models for IT management. Bus Inf Syst Eng. 2009;1(3):213–22. https://doi.org/10.1007/s12599-009-0044-5.
5. Belghith O, Skhiri S, Zitoun S, Ferjaoui S. A survey of maturity models in data management. Institute of Electrical and Electronics Engineers, 12th International conference on mechanical and intelligent manufacturing technologies, ICMIMT 2021. Cape Town, South Africa. 2021. p. 298–309. https://doi.org/10.1109/icmimt52186.2021.9476197.
6. Comuzzi M, Patel A. How organisations leverage: big data: a maturity model. Ind Manag Data Syst. 2016;116(8):1468–92. https://doi.org/10.1108/IMDS-12-2015-0495.
7. Daryani SM, Amini A. Management and organizational complexity. Procedia Soc Behav Sci. 2016;230(May):359–66. https://doi.org/10.1016/j.sbspro.2016.09.045.
8. Davenport TH, Bean R. How big data and AI are accelerating business transformation. 2019;1–16. www.newvantage.com. Accessed 15 Mar 2021
9. De Mauro A, Greco M, Grimaldi M, Malacaria S, Mignacca B. Toward the implementation of a Consensual Maturity Model for Big Data in Consumer Goods companies. 2022. Proceedings of the International Forum on Knowledge Asset Dynamics, Knowledge Drivers for Resilience and Transformation, IFKAD 2022. Lugano, Switzerland. 2022. p. 2380–402. https://hdl.handle.net/11580/91660.
10. Dunn Iii WJ, Scott DR, Glen WG. Principal components analysis and partial least squares regression. Tetrahedron Comput Methodol. 1989;2(6):349–76.
11. Fisher RA. Statistical methods for research workers. Edinburg: Oliver and Boyd; 1946.
12. Ghasemaghaei M. Are firms ready to use big data analytics to create value? The role of structural and psychological readiness. Enterp Inf Syst. 2019;13(5):650–74. https://doi.org/10.1080/17517575.2019.1576228.
13. Gökalp MO, Gökalp E, Kayabay K, Koçyiğit A, Eren PE. Data-driven manufacturing: an assessment model for data science maturity. J Manuf Syst. 2021;60(March):527–46. https://doi.org/10.1016/j.jmsy.2021.07.011.
14. Grover V, Chiang RHL, Liang T-P, Zhang D. Creating strategic business value from big data analytics: a research framework. J Manag Inf Syst. 2018;35(2):388–423. https://doi.org/10.1080/07421222.2018.1451951.
15. Günther WA, Rezazade Mehrizi MH, Huysman M, Feldberg F. Debating big data: a literature review on realizing value from big data. J Strat Inf Syst. 2017;26(3):191–209. https://doi.org/10.1016/j.jsis.2017.07.003.
16. Halper F, Krishnan K. TDWI big data maturity model guide. 2014. http://www.pentaho.com/sites/default/files/uploads/resources/tdwi_big_data_maturity_model_guide_2013.pdf. Accessed 10 July 2021
17. Helmy M, Mazen S, Helal IM, Youssef W. Analytical study on building a comprehensive big data management maturity framework. Int J Inf Sci Manag. 2022;20(1):225–55.
18. Hostmann B, Hagerty J. ITScore for business intelligence and performance management. Gartner IT Leaders Research Note G00205369. 2010. p. 1–5.
19. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 1933;24:498–520.
20. Hsieh PJ, Lin C, Chang S. The evolution of knowledge navigator model: the construction and application of KNM 20. Expert Syst Appl. 2020;148: 113209. https://doi.org/10.1016/j.eswa.2020.113209.
21. https://www.bloomberg.com. YouGov survey finds 80% of data-driven businesses claim they have a critical advantage as impact of pandemic continues. 2020. https://www.ptonline.com/articles/how-to-get-better-mfi-results. Accessed 17 June 2021.
22. Kubina M, Varmus M, Kubinova I. Use of big data for competitive advantage of company. Procedia Econ Finance. 2015;26(15):561–5. https://doi.org/10.1016/s2212-5671(15)00955-7.
23. McAfee A, Brynjolfsson E. Big data: the management revolution. Harv Bus Rev. 2012;90(10):61–7. https://doi.org/10.1007/s12599-013-0249-5.
24. McShea C, Oakley D, Mazzei C. The reason so many analytics efforts fall. 2016. Retrieved from Harvard business review website: https://hbr.org/2016/08/the-reason-so-many-analytics-efforts-fall-short. Accessed 18 May 2021
25. Mikalef P, Boura M, Lekakos G, Krogstie J. Big data analytics and firm performance: findings from a mixed-method approach. J Bus Res. 2019;98(July 2018):261–76. https://doi.org/10.1016/j.jbusres.2019.01.044.
26. Mikalef P, Pappas IO, Krogstie J, Giannakos M. Big data analytics capabilities: a systematic literature review and research

agenda. Inf Syst E Bus Manag. 2017. https://doi.org/10.1007/s10257-017-0362-y.
27. Nott C, Betteridge N. Big Data & Analytics Maturity Model. 2014. https://www.ibm.com/developerworks/community/blogs/bigdataanalytics/entry/big_data_analytics_maturity_model?lang=en. Accessed 25 June 2021
28. Olszak CM, Mach-Król M. A conceptual framework for assessing an organization's readiness to adopt big data. Sustainability (Switzerland). 2018. https://doi.org/10.3390/su10103734.
29. Pearson K. On lines and planes of closest fit to systems of points in space. Philos Mag Ser. 1901;6:559–72.
30. Popovič A, Hackney R, Tassabehji R, Castelli M. The impact of big data analytics on firms' high value business performance. Inf Syst Front. 2018;20(2):209–22. https://doi.org/10.1007/s10796-016-9720-4.
31. Pöppelbuß J, Röglinger M. What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. 2011. 19th European conference on information systems, ECIS 2011. Helsinky, Finland. https://aisel.aisnet.org/ecis2011/28.
32. Sahoo S. Big data analytics in manufacturing: a bibliometric analysis of research in the field of business management. Int J Prod Res. 2021. https://doi.org/10.1080/00207543.2021.1919333.
33. dos Santos-Neto JBS, Costa APCS. Enterprise maturity models: a systematic literature review. Enterp Inf Syst. 2019;13(5):719–69. https://doi.org/10.1080/17517575.2019.1575986.
34. Sharma R, Mithas S, Kankanhalli A. Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. Eur J Inf Syst. 2014;23(4):433–41. https://doi.org/10.1057/ejis.2014.17.
35. Spearman C. The proof and measurement of association between two things author (s). Am J Psychol. 1904;15(1):72–101.
36. Tarhan A, Turetken O, Reijers HA. Business process maturity models: a systematic literature review. Inf Softw Technol. 2016;75(November 2017):122–34. https://doi.org/10.1016/j.infsof.2016.01.010.
37. Van Looy A, De Backer M, Poels G, Snoeck M. Choosing the right business process maturity model. Inf Manag. 2013;50(7):466–88. https://doi.org/10.1016/j.im.2013.06.002.
38. VentureBeat. Why most big data projects fail. 2019. http://www.lavastorm.com/. Accessed 6 Apr 2021
39. Vidgen R, Shaw S, Grant DB. Management challenges in creating value from business analytics. Eur J Oper Res. 2017;261(2):626–39. https://doi.org/10.1016/j.ejor.2017.02.023.
40. Walls C, Barnard B. Success factors of big data to achieve organisational performance. 2019. https://www.researchgate.net/publication/337991292_Success_factors_of_big_data_to_achieve_organisational_performance. Accessed 26 Nov 2021
41. Wamba SF, Gunasekaran A, Akter S, Ren SJ, fan Dubey R, Childe SJ. Big data analytics and firm performance: effects of dynamic capabilities. J Bus Res. 2017;70:356–65. https://doi.org/10.1016/j.jbusres.2016.08.009.
42. Wendler R. The maturity of maturity model research: a systematic mapping study. Inf Softw Technol. 2012;54(12):1317–39. https://doi.org/10.1016/j.infsof.2012.07.007.
43. White A. Gartner—Our top data and analytics predicts for 2019. 2019. https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/. Accessed 30 Sept 2021
44. Zar JH. Significance testing of the spearman rank correlation coefficient. J Am Stat Assoc. 1972;67(339):578–80. https://doi.org/10.1080/01621459.1972.10481251.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.