IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN:2278-487X, p-ISSN: 2319-7668. Volume 26, Issue 12. Ser. 8 (December. 2024), PP 01-13
www.iosrjournals.org

Data-Driven Culture Maturity Assessment: In Which Aspects Does The Use Of Data Technologies Impact Companies In The Current Brazilian Scenario

Douglas Da Silva Mangini[1], Renan Melo De Freitas[1],
Kosisochukwu Annabelle Emeagwali[1], Alexandre Riberti Di Pierro[1],
Romulo Gonçalves Lins[1]
(Federal University Of ABC, 5001, States Avenue, Santo Andre´, São Paulo, Brazil.

Abstract:
Well-structured and available data is one of the most valuable goods for companies. This study aims to understand and evaluate the reality of the Brazilian Market in the Data-Driven Culture scenario in terms of the most evident current data trends, such as Business Intelligence, Big Data, Cloud Computing, and Machine Learning. The proposed methodology is based on a questionnaire regarding data-driven culture adherence to strategic decisions, data tools, and governance and control. The two-stage analysis aims to extract the reality and how the use of data technologies may be able to differentiate them competitively. As a result, it is possible to deduce that the sample of the Brazilian Market shows high similarities of Maturity-Digital with an average to a high score for all the analyzed individual aspects and contains the statistical approach showed that for the sample that the Digital Maturity is strongly associated with the financial results and Internal Approval of Employees. On the other hand, the results were not conclusive for Brand Recognition and Customer Approval and were weakly associated with Operational Efficiency.

Key Word: *Data-Driven Culture; Brazilian Market; Maturity Assessment; Data Technologies; Innovation.*

Date of Submission: 04-12-2024 Date of Acceptance: 14-12-2024

I. Introduction

Lately, data-driven culture (DD) has become one of the most crucial elements for generating significant insights for decision-making [1]. Boosted by this new paradigm, new and traditional companies have been self-adjusting and transforming the current Market over the last two decades, driven by the spread of data technologies' importance. Behind this transformation, considerable effort is being made by researchers from the most advanced companies and universities regarding technologies to support the vital role data has assumed [2].

Led by data processing availability, the term Data-Driven was created to designate business models that rely heavily on data as a critical resource for decision-making [3], directly affecting the business models adopted by companies. An increasing number of researchers consider that using data technologies plays a fundamental role in distinguishing high-performing and low-performing organizations in the current market [4].

Over the past decade, DD has been regularly reported to be a top priority for many top-level managers [5]. Such interest has not been a trend but instead a result of clear proof corroborating the values of data analytics to businesses. Numerous publications in the past few years have been released, with topics related to data-driven technology and technologies that make its application feasible. Among them, business intelligence, big data, cloud computing, machine learning, and artificial intelligence are the tools most commonly used for data analysis.

To sustain these disruptive new technologies, several types of research have been developed to establish friendly frameworks, methods, and standards that sustainably support this exponential growth in market demand. Such modern technologies are utterly impacting several instances in our society, with examples of how elements present in DD can be applied to fields such as E-Commerce and Market Intelligence [6], E-Government and Politics 2.0 [7], Science and Technology [8], Smart Health and Well- being [9], Security and Public Safety [10], among others.

This study proposes positioning the authors' contribution at the intersections of the above discussion and gaps in practice by investigating how the Data-Driven culture is used internally in terms of maturity in data usage for decision-making in Brazilian companies. To answer whether this attribute may represent a significant competitive differentiation in the Brazilian scenario. Thereby, the research problem is declared as follows: From the importance that the Data-Driven culture is globally reaching, this study aims to determine how the internal maturity in data use of current companies for decision-making inserted in the Brazilian Market and to answer whether this characteristic may represent a significant competitive differentiation in the Brazilian scenario.

The study has been divided into three parts to tackle the research question. The first part includes the theoretical background that supports the significance of the Data-Driven Culture transformation, the present technologies involved in this growth, and the current scenario of the Brazilian Market. In the second part, the proposed methodology is described, which includes an explanation of the input analysis, the research criteria for evaluating the Brazilian companies in terms of Key Performance Indicators, and a short questionnaire for gathering perceptions from inside the companies regarding their adherence to the Data-Driven Culture, use of data tools, and terms for governance and control used by them. The third part presents the results of two analyses. The first analysis aims to understand every cluster suggested individually to comprehend its contribution to data-driven culture adherence across different companies and within the same company. The second analysis aims to clarify which aspects of digital differentiation may influence the companies' competitive results regarding financial, operational efficiency, employee satisfaction, and consumer complaint rates. For this study, the Mann-Whitney statistical approach has been applied to a control group within the highlighted companies and another group of companies consulted in the survey.

II. Theoretical Background

The decision-making process in business involves understanding trends and practices supported by data. The capacity to convert gathered data into value-for-economic benefit/profit is essential for business growth [11]. In other words, precise data are acquired not only for the development of valuable insights but also for the execution of strategic business. Therefore, entrepreneurs expect this due to the importance of making proper decisions under unpredictable circumstances to find business prospects [12].

The effective use of data has accompanied the development of information technology architecture, thereby affecting decisions regarding the desirability and viability of entrepreneurial ideas, which makes entrepreneurial decision-making primarily knowledge-based. Thus, information systems are reasonable instruments for entrepreneurs, facilitating decision-making in complex business problems. Moreover, data technologies and techniques for decision-making based on applications are growing in a broad scope of knowledge-based domains. The promotion of such technologies has changed the dynamics of the business environment. The dialogues have centered on its transformational possibility for efficiency in following commercial value creation, which allows entrepreneurs to assess, learn, and exploit options and solutions under business uncertainties. These data technologies include creating computer systems to accomplish tasks that demand human intelligence, such as making decisions. A practical application of such technologies and techniques results in enhanced decision-making and more promising business performance [13]. This paper discusses data technologies and decision-making techniques, which include business intelligence, big data, cloud computing, machine learning, and artificial intelligence technologies.

Business intelligence (BI) was one of the first prominent technologies for data use for companies, and it reached a significant recognized value for assertiveness in decision-making. The start of use for previous data to extract strategic decisions computationally by companies lies in the 90s when a segment of IT professionals specialized in business structures to aggregate more value to their analysis. In the late 2000s, moved by the improvement of Analytics, BI was improved by acquiring new prediction tools to not just analyze past situations [14].

The exponential increase of BI importance over the years is directly related to digital transformation, which leads to a massive increase in data availability, boosting competitive advantages in strategic decisions [15]. The digital transformation process was gradual and may be related to the high-speed internet network spread for commercial, residential, and personal use and popularization converging with changes in the population's lifestyle and the company's globalization for commercial and organizational purposes.

Fig. 1 illustrates the projection of data available through an internet connection, almost quintupling the number of devices plugged in and able to exchange data in just 15 years. This proves the exponential growth of BI's importance.

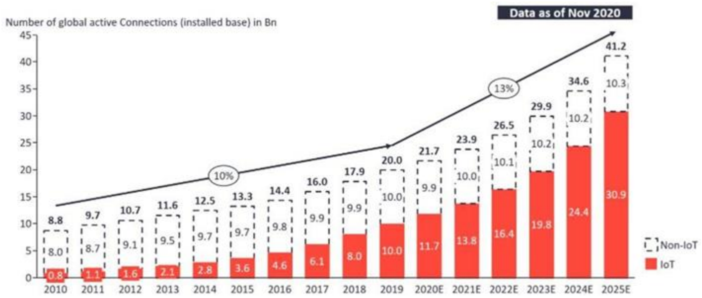


Figure 1: Number of internet-connected devices globally following accompaniments and projections of the IOT-ANALYTICS Consultancy in the 2020 reporting.

This ascendant curve of data availability is fundamental to understanding the internal and external movements of the current companies. Besides the immediate expected results, according to [16], an empirical study of 50 companies shows that the use of BI may be justified more than just for the seeking of low time benefit return as it is expected to bring disruptive returns in terms of anticipating important market movements in medium to long terms.

The boost of assertiveness in decision-making based on a Data-Driven Culture lies directly in the evolution from the "what who makes the decision thinks" to "what the company has data to support the decision" models is just possible and is based on the increase in available data [13].

An important concept presented by [17] is the Digital Transformation strategy, which lies in using proper technologies, changes in value creation, structural changes, and financial aspects. Their study related financial returns directly to the appropriate use of adequate technologies that are applied strategically to change value creation and perform an internal structural change. In this context, ample data use is the strategic approach to reforming the technology for potential businesses, including all the other elements of value creation and structural changes in search of financial aspects. A prominent characteristic of Big Data is the 5 V's definition: 1) Volume, 2) Velocity, 3) Variety, 4) Veracity, and 5) Value.

Furthermore, it is essential to emphasize that using Big Data does not just imply an improvement in the 5 Vs. It represents a disruptive change in data culture, leading to a need for adaptation in techniques and tools to work correctly in this new scenario. This new scenario demands competencies of the work members who extract decisions from exponential data technologies that are also different from traditional BI.

Because of the stunning increase in relevant data available for decision-making, pure BI techniques are becoming obsolete in the current Market. That occurs because standard processing tools present in traditional BI, generally used for punctual processing studies, are not performative enough to deal with and respond at an appropriate time for heavy demands. In these cases, BI may be combined with modern technologies, such as Big Data and Machine Learning, to extract the maximum proposed value of combining business knowledge and modern technical solutions.

Due to the increase in data volume, data combination, manipulation, and storage became much more complex and expensive for companies. Consequently, the Cloud Computing concept has become utterly popular in recent years [18].

The definition of Cloud Computing lies in a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Before cloud services were available, for companies to process their data and transform it into information, the only option was to have their local server. This structure has some disadvantages concerning Cloud or Hybrid (Cloud + Local Server), such as the price for expansion, the need for ultra-specialized professionals, and the server control activity also needs capable professionals working continuously to prevent and act in cases of overload, which could lead the applications to stop running. Therefore, the cloud advantages are 1) on-demand self-service contracts, 2) the contracts appreciate the unlimited elasticity needed, and 3) the architecture and difficulties regarding the size of the services are not an issue for the renting company.

Studies show that the proper use of Cloud Computing may considerably improve business value by directing all the development efforts on the finality of the company actuation in terms of pursuit for aggregation of value and not for money, such as in-infrastructure, where the customers are just concerned when they are impacted for some reason, i.e., unavailability or slowness [19].

Finally, consistent movements for manual process automation occurred in the data-driven culture and digitization of the company's operations. This process started by simulating simple click actions without any embedded intelligence, just repeating actions. The evolution from this model was the RPA (Robotic Process Automation), which allowed the companies to fully simulate an operator routine by analyzing inputs and making simple tabulated actions with limited intelligence.

The concern with this situation is that RPA developments may not learn from past situations. Although they can make decisions based on pre-programmed conditions, they cannot perform appropriately in non-standardized conditions. Data Science adopted the Machine Learning (ML) concept for these situations. The main idea is to develop software that can learn from past situations and keep enrolling in use. The main problem with that technology is the need for considerable effort in training to teach the code enough conditions and results. After that, it turns into an auto-sufficient solution.

It is important to emphasize that when properly used, all quoted automation rep- resents company gains. The choice of the most advisable solution depends on the granularity of difficulty and conditions. The gains can be measured in using less time of operators for repeated actions, releasing their time to making more important tasks related to decisions; in time, as machines may work out of commercial time; and in quality, as well programmed software is not subjected to operational failure.

The types of applications that RPA developments may have in modern companies have had a boost in importance in the recent period thanks to two main reasons: 1) the availability of data in various formats is a foresee aspect of the Big Data key concepts; and 2) the use of Artificial Intelligence (AI) based on ML to create chatbots to attend to customers, by using Natural Language Process or the use of algorithms to understand human language [20].

Therefore, Machine Learning is one possible approach to achieving the Pattern Discovery process. According to Fayyad's model, all these applications represent combinations with complementary technologies in sustainable and connected solutions [21].

Taking as source the most recent survey for Companies Demography and entrepreneurship Statistics released by the Brazilian Institute of Geography and Statistics (in Portuguese IBGE), using as a parameter the following Census from all segments of Brazilian companies in 2017, then a review is conducted to delineate the Market, as follows:

- Annual Survey for Industries and Companies (PIA-E);
- Annual Survey for Construction Industries (PAIC);
- Annual Survey for Commerce (PAC);
- Annual Survey for Services (PAS).

It shows an amount of 4.458.678 distinct active companies, being 676.444 in less than one year. As employees of these companies, there were 38.4 million registered workers, 31.9 million employees (19% of them in a 1 to 9 employees' company and 81% in a ten or more employees' company), and 6.5 million partners in or owners.

When taking into account the size of the companies, in terms of the number of employees, 46.2% of the companies did not have any employees, just the workforce of partners or owners - a phenomenon which can be heavily related to a Brazilian manner of contract when for the company the employee represents an individual company instead of a hired worker; 43.6% of the companies just had from 1 to 9 employees besides the partners or owners; remaining 10.2% of the companies within ten or more employees.

Regarding gender distribution and qualification, 60.8% of males and 39.2% of females registered to work for those companies. Of them, 14.2% of employees had formal graduation, and 85.8% had no formal graduation. Concerning the primary activity segmentation of the companies, the IBGE grouping criteria for active companies adopt the following results: Commerce and Automotive or Motorcycle Repair - 42.3%; Transformation Industry - 9.0%; Hospitality and Restaurants - 6.9%; Scientific and Technical activities - 6.6%; Administrative activities -6.5%; Construction - 5.3%; Transportation, Storage, and Logistics - 5.0%; Human Health and Social Services - 4.4%; Information and Communication - 3.1%; Education -2.4%; Services in General - 2.2%; Financial, Security and Related Services - 2.0%; Real Estate Activities - 2.0%; Arts, Culture, Sports and Recreation - 1.1%; Agriculture, Cattle, Forestall Production or Fishing - 0.7%; Water or Residuals Management - 0.2%; Extractives Industry - 0.2%; Electricity or Oil and Gas Industry - 0.1%.

Considering the Brazilian geographic division, the active companies are distributed, as shown in Fig 2.

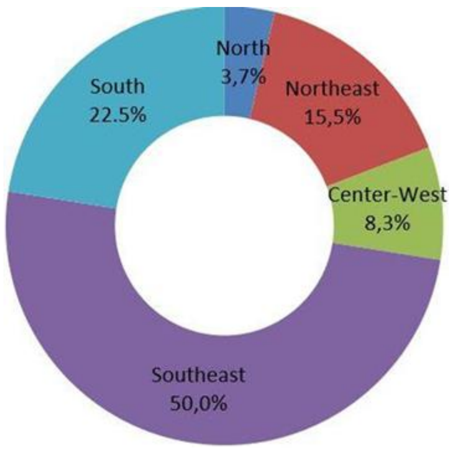


Figure 2: Distribution by Region of Brazilian Companies.

Featuring the top 10 States with more than 100.000 active companies: São Paulo (1.464.544), Minas Gerais (526.543), Paraná (411.669), Rio Grande do Sul (406.470), Rio de Janeiro (350.103), Santa Catarina (281.534), Bahia (225.537), Goiás (167.267), Ceará (128.754) and Pernambuco (122.116). Considering that the States with the highest number of registered employees follow the same top 10 States with a few changes in order positions.

From this perspective, it would be interesting to have respondents from at least 8 of the top 10 highest participatory states regarding the number of active companies and registered employees to have an accurate overview of the Brazilian market situation. Also, it would be an ideal target to have respondents from all the range of employees, from just partners and owners 17% of the workers) to 1 to 9 (16% of the workers) and more than 10 (67% of Brazilian workers).

III. Research Method

The proposed methodology comprises two complementary steps: the Data-Driven Maturity questionnaire elaboration and application and the result analysis.

An important disclaimer is that all the interviews are confidential, and the responses containing names and companies will not be published, nor will the companies' names. The questionnaire starts with a short interview to understand more about the interviewee. The purpose of these questions is merely to avoid duplicate or invalid responses. The first section asks for the social name, residence UF (Federal Unit), academic degree, and whether it is working or has worked in the year of research - being this the first delimiter of this section, as the aim is only on respondents currently acting in the Brazilian Market. The second section asks for the company's information and the interviewee's actuation. The questions are focused on the company name, company activity segmentation, company size in terms of the number of employees, company actuation UF, employee predominant actuation area, and employee role of actuation. Whether the employee works directly on the computer - being this the second filter of this section, as to understand clearly the data politics in the company, the employee must be familiar with them. Besides these two filters, another reason that would invalidate one or more responses is duplicating responses for a single interviewer. This could be achieved by inspecting the answers and searching especially for employee names and company names.

The interviews aim to recognize different data maturity situations for various segments of current Brazilian companies, from digital-born companies such as e-commerce and IT-based companies (Startups, Fintech, or InsureTechs) to classical sectors that are being urged to join the Data-Driven Culture to maintain competitiveness in the modern Market.

Besides, these interviews expect to facilitate the understanding of the difference between:

1. Brazilian demographic location (including the desire to have at least respondents from the top 10 Brazilian States' inactive companies' importance);
2. Companies Segments of activity (including the desire to have respondents from a higher amount of the 18 possible segments);
3. Size of the Company based on the number of employees (following the IBGE clustering criteria defined to classify companies by size);
4. Areas of Acting for the same and from different companies;
5. Employees' Role Positions from the same and different companies;
6. Employees' Academic background.

Also, it is intended to have answers in different levels of detail: from equivalent areas of other companies, from various regions of the same company, from distinct companies from the same market segment, and the equivalent quantity of employees as follows in Fig. 3.

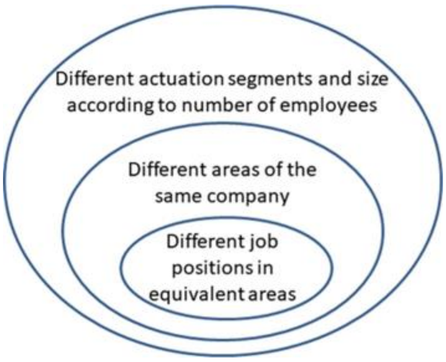


Figure 3: Diagram of expected responses distributed by level of detail.

Another important validation that needs to be performed is understanding statistical independence by analyzing the separated elements and their relation inside a single company. This analysis is exceptionally fundamental because, in other conditions, it would be imprecise to separate each condition and have its single results disregard the different elements.

To perform this validation, the One Hot Encoded technique is proposed; in this approach, all the individual element labels are converted into individual separate, and their filling is substituted for binary values. In a practical example shown by Fig. 4, if there are two conditions of four possible responses each, these two conditions would be converted into eight possible labels, and the filling would be converted from the two label responses to 1 in the ones which were the response and 0 in the others.

Condition 1	Condition 2
C1 Label 1	C2 Label 1
C1 Label 1	C2 Label 2
C1 Label 2	C2 Label 2
C1 Label 3	C2 Label 2
C1 Label 3	C2 Label 3
C1 Label 4	C2 Label 4

C1 Label 1	C1 Label 2	C1 Label 3	C1 Label 4	C2 Label 1	C2 Label 2	C2 Label 3	C2 Label 4
1	0	0	0	1	0	0	0
1	0	0	0	0	1	0	0
0	1	0	0	0	1	0	0
0	0	1	0	0	1	0	0
0	0	1	0	0	0	1	0
0	0	0	1	0	0	0	1

Figure 4: Example of practical application for One Hot Encoded Technique.

After making this change in the data structure, the next step would be performing a standard correlation operation to understand whether there are relations between the individual labels. This has much more meaning for computational tools to extract the results than in text format, as in the previous situation.

The questionnaire is divided into three pillars to singularize the interviewee's perceptions. The first pillar aims to understand the general Company Maturity in terms of Data-Driven Culture related to the leading technologies, and the questions elaborated for it are as follows:

1. Data with openly discussed calculation methodology support all the critical decisions in my actuation area.
2. In possible situations, not only past data but also statistical projections with future scenario predictions are considered.
3. In my sector, sensible activities are continuously monitored, and in abnormal situations, the support teams are called beforehand the final client is impacted (users/ consumers or other sectors).
4. In my industry, performing activities exploit the creation and management processes at maximum. Repetitive activities without value are discouraged and automated.

The second pillar respects understanding the already established development use of the presented technologies: Business Intelligence, Big Data, Cloud Computing, Machine Learning, and Artificial Intelligence. The questions elaborated for it are as follows:315

5. In my sector, robust data processing tools such as SQL, NoSQL, Hadoop, SAS, and Teradata are used for reading and writing instead of Windows-based solutions, such as Excel, Sharepoint, and Access.
6. The reports that foment my sector's decision-making are developed and made available in indicators publishing tools, such as Power BI or Tableau, instead of being disclosed via email.
7. The company data environments in use can adjust processing for growing amounts of data (a good indicator is to ponder if it is difficult to occur in unavailability or slowness situations).
8. The company bets on modern technologies for importing essentials through Cloud Service platforms, such as Salesforce, Amazon AWS, Google Cloud Platform (GCP), and Microsoft Azure; Big Data solutions, such as Spark; and distributed Solutions, such as Hadoop.
9. The company has projects based on Advanced Analytics, which optimizes sensible decision-making through Machine Learning techniques for self-adaptation in continuous learning and Artificial Intelligence that emulates human behavior.

The last pillar aims to understand the data and automation governance and availability for the different sectors of the company, and the questions elaborated for it are as follows:345

10. The company's relevant data are well cataloged in dictionaries distributed across the areas and follow norms for quality assurance and control of account periodicity updates.
11. The company's relevant data are widely available for projects and studies (possibly subject to access restrictions depending on information sensitivity).
12. In my sector, all the automation solutions are cataloged and hosted at IT-controlled infrastructure (and are under governance and good practices for development and use).

The questionnaire's concept was developed to be aware that some interested participants could desist from answering if it was too long or complicated. The questionnaire comprises 12 questions, with a response range from 1 to 5 (from less to high digital maturity is the company adherence) depending on how strongly the participant agrees with the statement, based on the Likert Scale.

Another disclaimer is that all the questions have individual targets in one aspect of the presented technologies, and no one topic has redundant questions. Because of this prerogative, the approach to calculating the general score will be through the simple sum of each answered question for each maturity pillar and the overall result for the respondent. As the individual score goes from 1 to 5, the highest value obtained in the summarizing, the most digital maturity is the respondent's perception of the company adherence.

Because the Likert Scale and the nonparametric statistical methodology of Mann-Whitney were chosen to analyze the results, no treatment for the outliers will be required for the analysis.

The data input is desired to be achieved by disclosing web forms containing the introduction questions plus a questionnaire through social media to target groups of university students, who are most expected to answer them correctly. Because of the program schedule, the disclosure agenda is about six months long and has a minimum. The target is to achieve at least seven out of the ten mainly Brazilian States regarding company representativeness and at least 75 valid answers.

The research analysis is performed in two complementary analyses: The first is based on studying the individual results and placing them in the already described groupings from companies and interviewee actuation in companies. Another two tests are going to be performed. The foremost aim is to understand whether there is a correlation between the individual elements of the cluster, which could suggest that the basis has unwanted dependent variables. Finally, if there are proven independent variables, the test is used to understand how the answers behave in different roles and acting positions inside the same company with more interviewees.

Furthermore, the second lies in separate companies in two clusters with proper relation and grouping significance. Finally, the Mann-Whitney analysis will focus on comparing the Data Maturity in each cluster with Result Factors according to the highest companies in terms of the following key aspects: 1) Financial results, such as Profit in a fiscal year; 2) Operational results, such as operational efficiency; 3) Employees Approval, such as market indexes of best companies to work on; 4) Customer care, such as the amount of customers complaints and the average time for resoluteness at consumer protection agencies.

The expected result for the second part, which is different from the first, is to understand statistically in terms of significant impact, in which the higher data maturity may represent a significant practical differentiation.

It must be noted that it is impossible to perform any cause-effect analysis of the results. The only expected answer is whether or not they are statistically related. However, the proposed methodology may not affirm causality. For instance, it may be possible to stipulate within secure statistical security that there is a relation between the key aspects and Data Maturity adherence in the Brazilian Market. Still, there will not be evidence to say if they are caused by or result from each other. Over the entire analysis phase, the test result comparisons will be presented over visual dashboards presenting all the result analyses within explanations and legends.

The research conclusion is based on exploring the results obtained and the advice of Brazilian companies based on the results raised and inferences.

IV. Method Application And Results Analysis

The questionnaire was applied, and 113 answers from 15 of the 26 + 1 (the Brazilian States + Federal District) were reported. From the answers, 77 were working or have worked in the year of research, and from this public, 74 have activities related to computer use. Furthermore, it computed adherent answers from 57 different companies, including Education, Transports, Telecommunications, Health care, Financial, Research and Development, Legal, Retailers, Third Segment, Chemistry, Building, Textiles, and Agronomy from 13 of the 18 different signaled segments following the IBGE classifying. Fig.5 shows the answers average for each question applied.

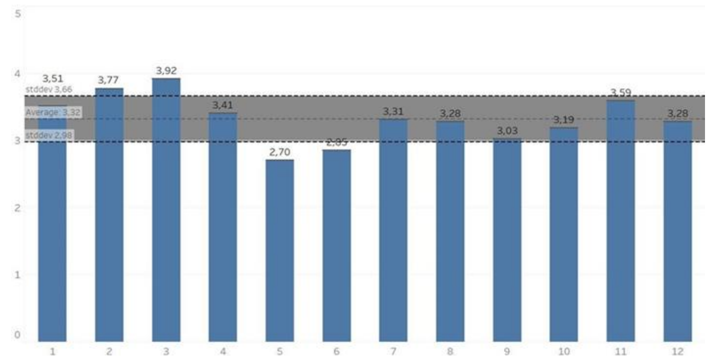


Figure 5: Answers average for each question applied.

The Average respondents' perception of the answers was 3.32, the Global Standard Deviation was around 1.20, and the lower-scored questions were related to the proper use of robust data processing tools instead of Operational System simple-based solutions (question 5 with a score of 2.70, and Standard Deviation of 1.50) and the use of reporting tools based on server solutions instead of shared by email lists (question 6 with a score of 2.85, and Standard Deviation of 1.40).

On the other hand, the highest-scored questions were related to continuously monitoring sensible activities to avoid impact on the final client (question 3 with a score of 3.92) and using statistical projections with future scenario predictions for efficient analysis (question 2 with a score of 3.77). From the results above, it can be stated that companies are increasingly interested in customer centrism (first pillar). However, some did not follow the best practices to reach their goals, mainly regarding best technologies (second pillar) and data governance (third pillar). Summarizing the results, Fig.6 illustrates the analysis of the answers by the pillar.

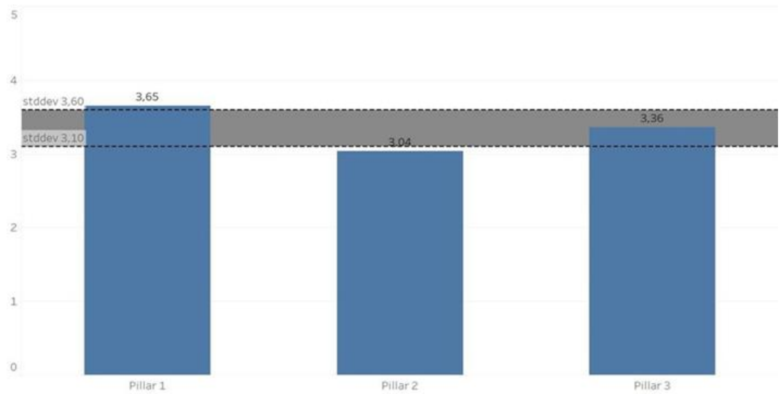


Figure 6: Answers average graph for each pillar.

By analyzing the graph, it is possible to infer that the Data Driven-Culture is not state of the art in the evaluated companies in which respondents participated (expected to be as great as possible from 4 to achieve). However, they have an overall positive perception of their approaches.

Regarding the Data-Maturity perception (pillar 1), it is higher than the perceptions of Technologies in use (pillar 2) and also slightly higher than the governance and data control (pillar 3), which may indicate that the respondents assumed that the companies are firstly transforming their culture before establishing a robust technological framework which could support this transformation. It occurs because the immediate and cheaper transformation begins with the first pillar, which involves cultural questions and not potential investments in technologies, as pillar 2, and governance and availability, as pillar 3.

It must be noted that companies that deeply believe in the cultural aspects without being supported by the other pillars after establishing a data-driven culture could face a risk. Fig.7 depicts the scenario of the Brazilian geographic region.

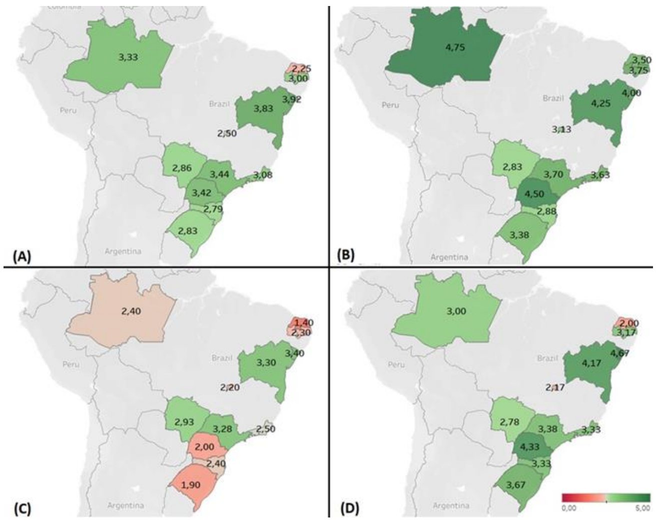


Figure 7: A: Overall; B: Pillar 1; C: Pillar 2; D: Pillar 3; Questionnaire Score over Brazilian Regions.

Following Fig.7, the Geographic Analysis identified a similar behavior to the previous analysis for overall and individual Pillars. However, the higher fluctuation between them was recognized, which can be justified by the most significant number of individuals who analyzed elements. Furthermore, the same positive perception over the First Pillar can be observed, which suggests that, whether there were enough available responses over the other states, it was possible to conclude in a global understanding that the Brazilian Market tends to believe in the Data-Maturity need and benefits.

On the other hand, the same behavior presented by Pillars 2 and 3 was also identified as more motivated by Pillar 2 within technologies in use than Pillar 3 within governance and data control. For these two pillars, concluding the results for the other regions reasonably was impossible. However, predicting behavior from low to moderate inferior for Pillar 3 concerning Pillar 1 and considerably inferior for Pillar 2 relating to the other two Pillars was possible.
To finish the description and analysis of the individual results conducted, Table 1 summarizes all the results.

Table 1: Comparison between individual suggested cluster results.

Suggested Cluster	Average Score	Fluctuations Presence	Outliers Presence
Company Segment	1.67 to 4.17	Considerable	Few
Company Size	2.15 to 4.25	Slightly	No
Employees Actuation Area	2.28 to 4.40	Slightly	One
Acting Role Position	2.58 to 4.00	Slightly	No
Employees Scholar	1.80 to 4.50	Considerable	Yes

These results are similar, corroborating the homogeneity Hypothesis in the Brazilian Data-Driven Maturity.
A fundamental aspect of ensuring the result of all the covered analyses is guaranteeing that all variables are independent. If one or more of the studied variables are dependent, all the dependent results could be invalidated because, statistically, one would affect the others.
The one-hot encoding technique was applied to perform the test since the statistical correlation analysis would not be performative in comparing names as 1 to 9 employees, the UF names, or even the Roles or Segments. To perform it, a new converted base was created following all variables information transformed into labels and the correspondence values placed as 0 in case of not fit and 1 in case of fit. The analysis was performed simultaneously, using Python, for the whole set of variables. However, the results are shown individually to be easiest to analyze. Fig.8 illustrates the study for the Brazilian Geographic UFs.

Figure 8: Study of Independent Variables for the Brazilian Geographic UFs.

Fig.9 shows the study for the Brazilian Segment Activity.

Figure 9: Study of Independent Variables for the Brazilian Segment Activity.

Fig.10 displays the Study of Independent Variables for the Brazilian Size of the Company in the Number of Employees.

Figure 10: Study of Independent Variables for the Brazilian Company Size in Number of Employees.

Fig.11 brings the study for the Interviewed Employees Actuation Area.

Figure 11: Study of Independent Variables for the Interviewed Employees Actuation Area.

Fig.12 depicts the study for the Interviewed Employees Acting Role Position.

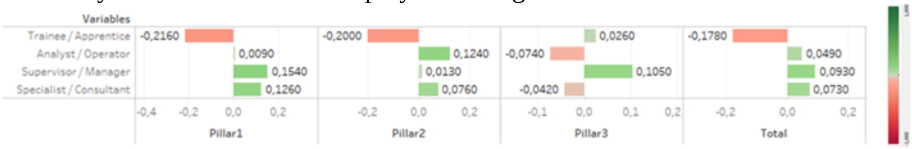


Figure 12: Study of Independent Variables for the Interviewed Employees Acting Role Position.

Fig.13 presents the study for the Interviewed Employees Scholarity.

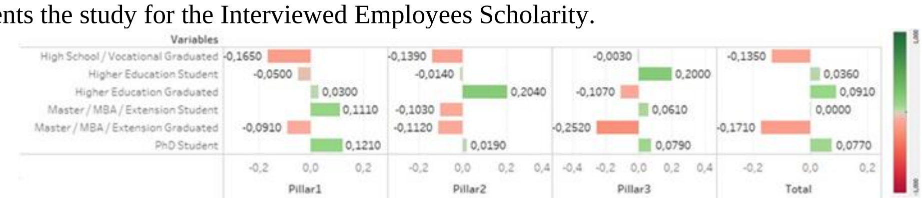
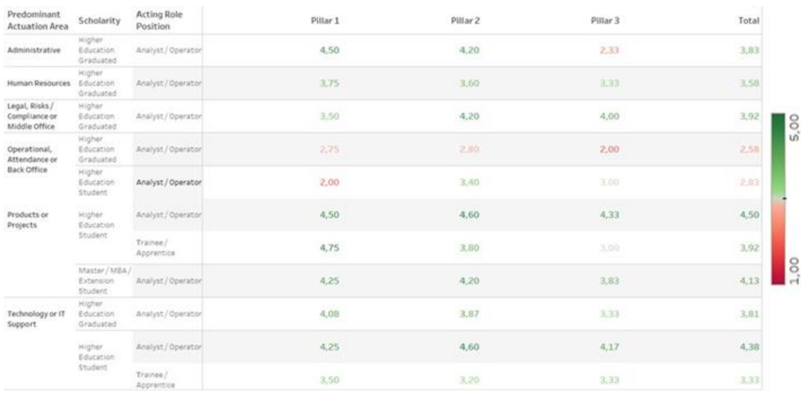


Figure 13: Study of Independent Variables for the Interviewed Employees Scholarity.

As shown in Figs.8 to 13, no one of the individual variables presented a significant correlation (close to 1—as a statistical definition) to the data Maturity Results among the other variables, validating the initial premises of Independence and all the highlighted results. Therefore, this validation corroborated the above results and enabled the situational test below to compare the individual situations inside the same company.

The final individual analysis is performed in the company with more interactions (15) and is related to analyzing the individual personalities of different interviewees in the same company. Fig.14 shows the test results for the 11 different classifications found. This test was important because it aimed to validate the Variable Independence test and to understand if the individual behaviors presented previously were still evident in a more controlled scenario inside the same company.



Predominant Actuation Area	Scholarity	Acting Role Position	Pillar 1	Pillar 2	Pillar 3	Total
Administrative	Higher Education Graduated	Analyst / Operator	4,50	4,20	2,33	3,83
Human Resources	Higher Education Graduated	Analyst / Operator	3,75	3,60	3,33	3,58
Legal, Risks / Compliance or Middle Office	Higher Education Graduated	Analyst / Operator	3,50	4,20	4,00	3,92
Operational, Attendance or Back Office	Higher Education Graduated	Analyst / Operator	2,75	2,80	2,00	2,58
	Higher Education Student	Analyst / Operator	2,00	3,40	3,00	2,83
Products or Projects	Higher Education Student	Analyst / Operator	4,50	4,60	4,33	4,50
		Trainee / Apprentice	4,75	3,80	3,00	3,92
	Master / MBA / Extension Student	Analyst / Operator	4,25	4,20	3,83	4,13
Technology or IT Support	Higher Education Graduated	Analyst / Operator	4,08	3,87	3,33	3,81
	Higher Education Student	Analyst / Operator	4,25	4,60	4,17	4,38
		Trainee / Apprentice	3,50	3,20	3,33	3,33

Figure 14: Individual Perception of 15 Interviewed Over the Same Company.

From this analysis, some aspects could be detected. The first was the lower score of Operational, Attendance, or Back Office results, which decreased the overall results independent of the other parameters. The second one was the results were in areas with more than one interview with different Scholarly and same Acting Role Positions, in which the results were quite similar with a fluctuation smaller than 0, 6. The third possible analysis that may be performed is comparing the same areas and scholarly but different acting role positions, the score of which is slightly lower for trainee/apprentice positions. All those aspects were according to the results highlighted previously. Also, they corroborated the variable independence validation because even with more than one variable in a controlled test environment, the results followed the same individual tendencies.

In the Mann-Whitney tests, whose goal is to understand the advantages of Data Maturity, all 57 companies were ranked over their overall score in the questionnaire average for each of the five evaluations conducted. For all the tests, the study of each Pillar and Total were validated as proof of concept. However, because of the similar results presented, only the total is demonstrated; otherwise, the analysis would be repetitive and not bring complementary evidence to the study. Table 2 summarizes the tests' results adopting as the criterion for rejecting the null hypothesis the Z values less than −1.96, or greater than 1.96, for a $p < 0.05$.

Table 2: Mann-Whitney test results.

Key Aspect for Evaluation	Made Choice	Hypothesis	Z Value
Known or Not Known Company	by having open capital and being listed on the stock exchange	Are the 20 most Known Companies more Data-Mature than the other 37?	-0,98
Financial Results	Considering revenues higher than 1 billion Dollars in Real to Dollar exchange, considering a 4 to 1 approximation for the 2019 fiscal year	Are the eight most Profitable Companies more Data-Mature than the other 12?	-2,31
Operational Results	because their balance of Revenue and Profit was greater than 15 in the 2019 fiscal year	Are the five most Operationally Efficient companies more Data-Mature than the other 15?	-1,57
Employees Approval	because their score assigned for the company is higher than 4 in a 1 to 5 employee evaluation in the most well-rated website to perform companies' evaluations in Brazil - GlassDoor	Have the five companies most Well Approved by Employees more Data-Mature than the other 15?	-1,87
Customer Care	because their score is higher than 8 in a 0 to 10 customer evaluation for the 2019 year in the most well-rated website to perform companies' evaluations in Brazil - Reclame Aqui	Have the five companies most Well Rated by Customers more Data-Mature than the other 9?	-

As the Normal Distribution Conversion is just indicated for Mann-Whitney Studies within more than 20 samples, this technique was not applied for the Customer Care evaluation. However, because of the not-so-wide difference between the Ub parameter and the value of Mann-Whitney, it is expected that the statistical analysis would not be conclusive in affirming that Companies most Well-Approved by Customers have better Data than the others least Approved.

In a nutshell, Table 3 below shows an overview of the main results obtained through the Mann-Whitney test to compare each individual suggested aspect for evaluation.

Table 3: Comparison between individual suggested key aspects evaluations.

Key Aspect for Evaluation	Statistical Relation
Known or Not Known Company	Not Possible to Affirm
Financial Results	Highly Possible to Affirm
Operational Results	Possible to Infer a Certain Tendency
Employees Approval	Possible to Affirm
Customer Care	Not Possible to Affirm

To conclude the Mann-Whitney analysis, the main findings are as follows:

1. All the comparisons used companies from different activity segments, and only this subject was not expected to affect the results because most of them had a similar score in the individual activity segment evaluation from the previous analysis;
2. The comparison between most and least-known companies showed that most new companies, such as start-ups and digital companies, usually have already initialized their activities as data-maturity adherence, which is one of their strengths. Another important indicator is that the most known and consecutively biggest companies that participated in the research showed to be evolving in the direction of Data-Maturity adherence. However, they need a longer time to self-adapt, which is not expected for the most minor known companies and probably smaller than may perform this transition with lower effort;
3. The comparison between the most and least profitable companies showed the best statistical result obtained for the Mann-Whitney analysis. An aspect of the results is that it is impossible to deduce any relationship between the causes and effects of the results. In other words, whether the companies have a better profit because of Data- Maturity adherence or because they have a better profit, they can invest more to achieve this status. Nevertheless, the test can provide reliable results that the most consolidated companies strive to achieve Data-Maturity knowledge;

4. The poor performance for the comparison between most and least operationally efficient companies was affected mainly by the distinct segments and actuation areas. However, the lack of answers made it impossible to perform an analysis by segment, whose results could be considered committed. Nonetheless, considering only the comparison in the overall Brazilian Market, the results can be signaled as not statistically precise;
5. The comparison between most and most minor companies, which were well-approved by employees who participated in the research, showed a positive statistical result. However, it was impossible to deduce any relationship between cause and effect for the results. In any case, the test could provide an interesting result that suggests that employees may enjoy working in companies with the best data security adherence;
6. The poor performance of the comparison between most and most minor companies well approved by customers could be related again to comparing companies with different actuation areas and customer care and post-sales concerns. However, another factor that could make this analysis difficult was that the internal areas that take action on those cases are the operational, attendance, or back offices, as shown by the previous study, which were areas systemically less adherent to data maturity than the rest of the company. This factor is unjustified because all companies are subject to the same conditions. However, it could complicate the first aspect of comparing different realities scenarios.

V. Conclusion

An important finding of this research was the introduction of modern technologies related to Data-Driven use and their placement in a Data-Maturity adherent scenario. The beginning of the work led to the elaboration of a questionnaire that aims to clarify all the different needs that companies are expected to achieve in the format of questions and pillars. All tests and applications for the questionnaire did not present any reported difficulty or misunderstanding that could guarantee the quality of the answers.

It is important to emphasize that as much the interview quantity is accurate, it would be the first part of the analysis to understand the Brazilian Market in terms of Geographic Region, Segment Activity, company size in Number of Employees, Pre- dominant Actuation Area, Acting Role Position, and the Scholarly. In the second part of the research, the analysis is based on the Mann-Whitney, and the number of interviews does not represent an issue. Such a statement is related to the number of answers available that would represent each group, and the most accurate would be the responses, but having a higher number of different companies would not necessarily be a source for improving the analysis.

In a nutshell, except for some punctual characteristics, the Brazilian Market Sample has shown relative similarity in terms of Data-Maturity adherence, and the same behavior was found in the companies that participated in the research. Furthermore, some essential outliers are related to the actuation segment, which has Commerce and Automotive or Related Services below the national average, while the Financial, Security and Related Services, Scientific and Technical Activities, and Transportation, Storage, and Logistics are above the national average. The findings for the statistical approach were quite exciting and showed that data maturity in the companies might be strongly related to higher Profit and better employee approval; on the other hand, some hypotheses did not follow the same behavior, showing a low relation to Company Recognition and Customers Complaints and weak relation to Operational Efficiency. Therefore, it is possible to conclude that the Brazilian Data-Maturity adherence for the addressed sample is not fully consolidated yet; however, it is possible to state that respondent companies in the research follow the same good practices and improvement levels.

For future work, the suggestion is to enlarge the number of respondents. It is highly recommended that the continuity of questionnaire application be conducted at constant time intervals in the next years, following the adaptation of the evident technologies, if necessary. This continuity would confirm this evolutionary curve for the evaluated data maturity and the impacts on the National Market.

References

[1]. C. Guan, J. Mou, Z. Jiang, Artificial Intelligence Innovation In Education: A Twenty-Year Data-Driven Historical Analysis, International Journal Of Innovation Studies 4 (4) (2020) 134–147.

[2]. D. Delen, H. M. Zolbanin, The Analytics Paradigm In Business Research, Journal Of Business Research 90 (2018) 186–195.

[3]. P. Hartmann, M. Zaki, N. Feldmann, A. Neely, Big Data For Big Business? A Taxonomy Of Data-Driven Business Models Used By Start-Up Firms.

[4]. A. C. Louro, M. M. Branda˜o, J. Jaklicˇ, A. Sarcinelli, How Can Customer Analytics Capabilities Influence Organizational Performance? A Moderated Mediation Analysis, Bbr. Brazilian Business Review 16 (2019) 369–382.

[5]. P. Mikalef, R. Van De Wetering, J. Krogstie, Building Dynamic Capabilities By Leveraging Big Data Analytics: The Role Of Organizational Inertia, Information & Management 58 (6) (2021) 103412.

[6]. S. Painuly, S. Sharma, P. Matta, Big Data-Driven E-Commerce Application Management System, In 2021 6th International Conference On Communication And Electronics Systems (Icces), Ieee, 2021, Pp. 1–5.

[7]. F. Lemke, K. Taveter, R. Erlenheim, I. Pappel, D. Draheim, M. Janssen, Stage Models For Moving From E-Government To Smart Government, In: Electronic Governance And Open Society: Challenges In Eurasia: 6th International Conference, Egose 2019, St. Petersburg, Russia, November 13–14, 2019, Proceedings 6, Springer, 2020, Pp. 152–164.

[8]. B. Wang, C. Wu, L. Huang, L. Kang, Using Data-Driven Safety Decision-Making To Realize Smart Safety Management In The Era Of Big Data: A Theoretical Perspective On Basic Questions And Their Answers, Journal Of Cleaner Production 210 (2019) 1595–1604.

[9]. L. Carrubbo, A. Megaro, F. Notari, Data-Driven Decisions For A Smarter And Re-Silent Healthcare Service Ecosystem, In Research And Innovation Forum 2021: Managing Continuity, Innovation, And Change In The Post-Covid World: Technology, Politics And Society, Springer, 2021, Pp. 69–77.
[10]. N. Mohamed, J. Al-Jaroodi, I. Jawhar, N. Kesserwan, Data-Driven Security For Smart City Systems: Carving A Trail, Ieee Access 8 (2020) 147211–147230.
[11]. G. George, M. R. Haas, A. Pentland, Big Data And Management (2014).
[12]. F. Stocker, G. Abib, Risk Management In Born Globals: The Case Of Brazilian Craft Breweries, Bbr. Brazilian Business Review 16 (2019) 334–349.
[13]. S. Shamim, J. Zeng, S. M. Shariq, Z. Khan, Role Of Big Data Management In Enhancing Big Data Decision-Making Capability And Quality Among Chinese Firms: A Dynamic Capabilities View, Information & Management 56 (6) (2019) 103135.
[14]. T. H. Davenport, Et Al., Competing On Analytics, Harvard Business Review 84 (1) (2006) 98.
[15]. F. P. S. Surbakti, W. Wang, M. Indulska, S. Sadiq, Factors Influencing The Effective Use Of Big Data: A Research Framework, Information & Management 57 (1) (2020) 103146.
[16]. G. Phillips-Wren, M. Daly, F. Burstein, Reconciling Business Intelligence, Analytics And Decision Support Systems: More Data, Deeper Insight, Decision Support Systems 146 (2021) 113560.
[17]. C. Matt, T. Hess, A. Benlian, Digital Transformation Strategies, Business & Information Systems Engineering 57 (2015) 339–343.
[18]. M. Al-Ruithe, E. Benkhelifa, K. Hameed, Data Governance Taxonomy: Cloud Versus Non-Cloud, Sustainability 10 (1) (2018) 95.
[19]. A. Aljabre, Cloud Computing For Increased Business Value, International Journal Of Business And Social Science 3 (1).
[20]. N. M. Sharef, T. Martin, K. A. Kasmiran, A. Mustapha, M. N. Sulaiman, M. A. Azmi-Murad, A Comparative Study Of Evolving Fuzzy Grammar And Machine Learning Techniques For Text Categorization, Soft Computing 19 (6) (2015) 1701–1714.
[21]. U. Fayyad, R. Uthurusamy, Evolving Data Into Mining Solutions For Insights, Communications Of The Acm 45 (8) (2002) 28–31.