

The Data Scientist's Triple Challenge: Quality, Clarity, and Expectations

Muraina, Ismail Olaniyi

Orcid-ID: <https://orcid.org/0000-0002-9633-6080>

Computer Science Department, College of Information and Technology Education, Lagos State
University of Education, Lagos Nigeria

Abstract

Three recurrent strategic issues confront data scientists as organizations increase their reliance on data-driven insights include managing organizational expectations, communicating successfully with non-technical stakeholders, and preserving data quality and availability. These issues have quantifiable effects on business; in fact, the average company loses roughly \$12.9 million annually due to poor data quality. The majority of practitioner time is typically devoted to data preparation alone; cleaning and arranging data accounts for about 60% of analytic effort. Additionally, practitioner surveys consistently identify inability to match analytics with stakeholder needs (including clear communication of results) as a major factor in project success. This study examines the intersections of these three issues in contemporary data science processes and suggests workable solutions that combine expectation-calibration, stakeholder-centered communication, and governance to improve delivery and reduce hidden costs.

Keywords: Data Scientist, Data-Driven, Data Quality, Data Clarity, Data Availability, Data Expectations

1. Introduction

Industrial strategic decisions are now being made by data science, however, without trustworthy inputs and organizational alignment, technical models do not work. The financial risk is significant: besides the average cost of c. 12.9M, companies lose as much as an estimated 15-25% of the revenue due to poor data quality and the business inefficiencies that arise due to this. According to practitioners, data cleaning and preparation also use about 60 percent of their time and leave less bandwidth to model, interpret and engage stakeholders. These facts are the reason why the triad quality, clarity, expectations are repeatedly presented in the surveys and

in case studies as the determining factor in outcomes of the project (See figure 1).

The Synergy of Effective Data Science

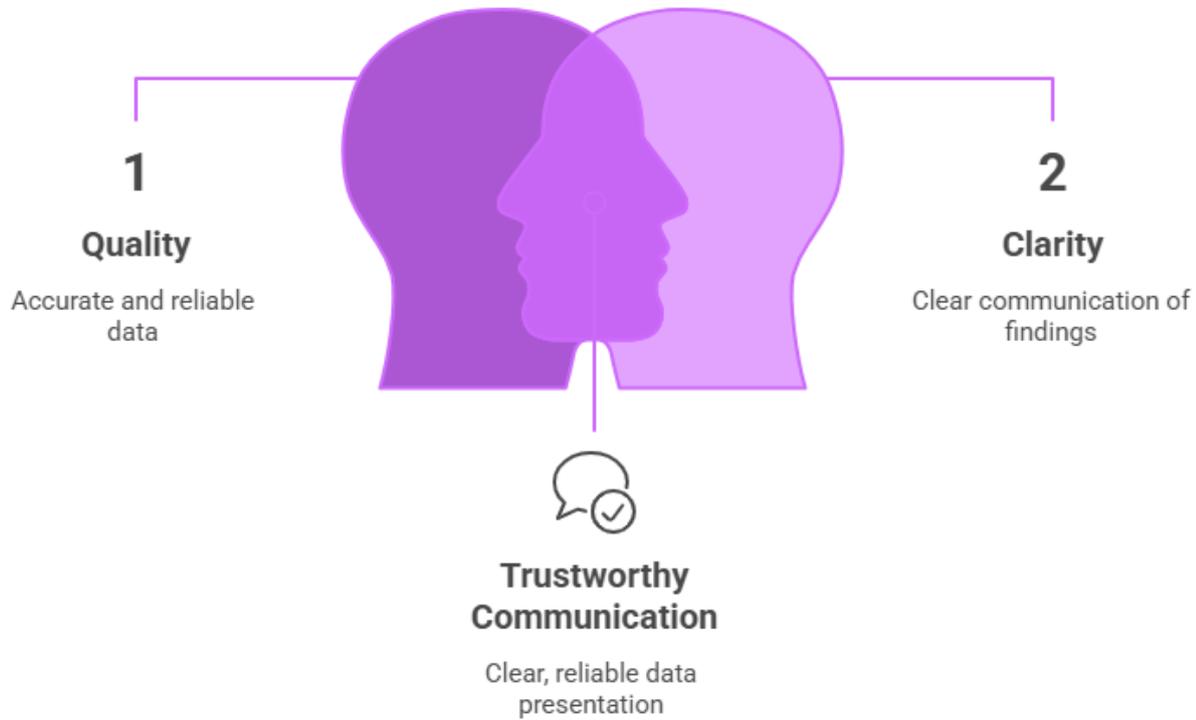


Figure 1: The Synergy of Effective Data Science

The Challenge of Data Quality and Availability

1. Fragmented and Inconsistent Data Sources

The dispersal of data between cloud, on-prem and spreadsheet silos causes missing fields, identifiers that are not consistent as well as the existence of duplicate records - issues directly reflected in operational costs and decision risk. Current industry reports and commentaries still rank data quality at the highest point of enterprise risk records. Gibbs et al, (2005)[12] argued that customer data fragmentation issue in multiple databases which are kept by various functional units is a specialized data quality problem, which impacts on the capability of an organization to meet privacy laws. It was observed that when personal information is distributed and disseminated throughout an organization and the lack of control is noted, such actions may have grave consequences especially in the context of the sensitivity of such information. They added their roles on the issue of fragmentation between databases where the data structure used was not compatible and inconsistent and thus it was hard and time consuming to access and collate personal information. Hoang (2023)[13] investigated how data

quality affects the process of data harmonization, especially when companies receive external data of various sources and internal data of different subsidiaries. He cited incompatible data quality as one of the issues which may result in information asymmetry and eventually to the sustainability of the process and the software being operated. Therefore, he proposed that effective harmonization of data is pegged on quality data that lacks inconsistency and errors. The isolated repositories of information that are referred to as data silos are the fundamental issue of incorporating big data analytics in the supply chain management that they do not allow a comprehensive analysis and cross-functional coordination. The data silos cause redundancy and inconsistency of data which results to discrepancy in forecasting and inventory management that is part of bad data quality. It is also due to the use of fragmented data systems whereby the suppliers of a supply chain have been siloed, which leads to inefficiencies hindering decision-making and transparency (Kusumawati, 2025)[17]. In the same vein, Sienkiewicz, (2025)[31] revealed the phenomenon of Data Silos, or separate repositories that are directly linked to individual business segments, as being a clear expression of architectural fragmentation, causing operational inefficiency and problems with governance. Since the presence of Data Silos lowers the informational flow, results in the repetition of resources, decreases the quality and availability of data, and creates inconsistencies and challenges in the organized analysis. This form of fragmentation is a major obstacle to data-driven strategies and digital transformation because the data stored in Data Silos is not easily shared, since it is kept in isolated systems.

2. Data Governance and Quality Assurance

Proper governance must be done on lineage, validation, role-based access and automated monitoring. It can be justified by investing in the capabilities: the business rationale is supported by the quantifiable cost savings with regard to the multi-million-dollar annual loss of bad data. Bassi and Alves-Souza posit that Data Governance (DG) is comprehensive in evaluating and managing data usability, availability, data privacy and security and data quality (DQ) both inside and outside the organization (Bassi and Alves-Souza, (2025)[5]. Authors recognized the need in Data Quality (DQ) assurance as one of the most significant issues in the implementation of Data Governance projects and classifies challenges of Data/Organization concept that comprises Data Quality (DQ) and Data Life Cycle Management as the flaws of the nature of the data or the organization itself. Ates & Garip, (2025)[4] stated that data governance is a collection of roles, processes, policies and tools that are aimed at making sure that quality of data is upheld during its life cycle and is used properly within an organization implying that it is a solution to quality assurance and not necessarily a problem. Their research shows that the key challenges in

business include lack of knowledge on the effect of the data governance process on other business processes, the complexity of the process, and the position of the process manager and lastly stated that the quality of data available and quality are key issues whose data governance initiatives aim to resolve. Lebaea et al, (2024)[19] opined that the issue of complexities that accompany data governance, particularly when dealing with new technological models, is a very challenging one. They found that data integration, scalability, regulatory compliance, and data quality maintenance are among the problems facing organizations in spite of acknowledging the relevance of data governance. They also claimed that the concepts of data governance are to ensure the availability, coherence, and quality of data which are vital in reliable and responsive IT systems hence suggesting that absence of good governance may result in problems of data quality and availability.

3. Ensuring Data Availability

Standardized APIs, documented schemas, and reproducible pipelines are required of data scientists. A shortened prep work tail that wastes time with practitioners is reduced here. Xiao et al, (2017)[35] discovered that the availability of data as an aspect of data quality, which quantifies it as a percentage of voluntary medical male circumcision (VMMC) clients whose client intake form (CIF) is on file at the respective sites. The research found data quality and data availability issues in a VMMC program and indicated that, despite an overall improvement in data quality audits tended to enhance data availability and completeness, certain gaps in documentation persisted, as well as that data availability and completeness tended to improve following data quality audit, indicating the inconsistency of data quality assurance. Cai and Zhu, (2015)[7] defined data availability as user convenience in obtaining data and related information, Identifying the availability as one of five dimensions of a data quality standard, usability, reliability, relevance, and presentation quality, but also breaks down the concept of availability into three subcategories, namely, accessibility, authorization, and timeliness. Data quality management (DQM) is essential in the process of successful data governance because it safeguards the accuracy, reliability and relevance of the data to be used in decision-making, DQM encompasses detecting and fixing data quality problems, the establishment of controls, and ongoing guarantee throughout all the data lifecycle processes. Therefore, the successful data quality management is often threatened with a variety of challenges, and their overcoming is a key factor in ensuring data integrity and utility (Sargiotis, 2024)[30]. Derakhshan et al, (2021)[11] explained that any unavailable or inaccessible data would result in a drop in the percentage of data being completed and more missing data, which is how availability is one of

the aspects of data quality issues. Moreover, the access and availability of data were tested by a skilled panel in the initial stages of data quality testing, as these were identified as possible challenges. This missing information was usually explained by the unconsciousness of patients, the absence of a responsive companion, or the absence of ID-documents which are feasible problems directly related to the availability of the data at the time of collecting them.

Communicating with Non-Technical Stakeholders

1. Translation of Technical Concepts

Most successful teams do not analyze model internals by themselves, but rather translate algorithmic output into business measures - dollars saved, lead time reduction, risk reduction, etc. According to surveys, communicating the results to end users is one of the most successful factors in data science projects, which supports the view of technical performance as not being enough. According to Martinez et al (2021)[24] , out of three success factors listed as the most relevant, the authors recognized such as communicating the results to end-users, the precise description of the needs of stakeholders and team collaboration/coordination.

Researcher demonstrated that the emphasis on organizational and socio-technical issues (such as communication) offers a substantial void that was recognized in earlier surveys between technical and organizational operations in data science project management. It was also mentioned that over the past few years, the data science community concentrated on technical issues and did not pay enough attention to the organizational and socio-technical ones. The article by Suh et al (2023)[33] mentioned poor communication between data scientists and subject matter experts as one of the leading impediments to model development and deployment, which substantiates the assumption that technical performance is not a sufficient condition. Their results indicated that the delivery of performance measures is not enough to the Subject Matter Experts (SMEs) that are expected to make decisions and recommendations using the outputs of the model. They also noted that the communication bottleneck may ensure the model fails or even is not used at all implying that poor presentation has negative effect on adoption despite its technical quality. As Weiner (2022)[34] clarified, the successful delivery of a project is the delivery of a model or communication of the analysis results, which implies that communication is one of the success factors that can be discussed on the same level as the delivery of a model. According to him, anytime a project will fail, the inability to explain a model prediction when it matters to the management or customer, he cited examples as to why a project will fail, such as lack of leadership support, access to information, and collaboration across teams, all of which are related to communications and alignment of management as basis

to project success. In a bid to establish a list of critical success factors (CSFs) in data analytics, Demir et al (2024)[10] conducted a systematic literature review to study the diverse data analytics areas such as Data Science. Based on 28 research articles, they determined six major themes of CSFs, namely, People, Strategy, Technology and Data, Organizational Culture, Process Design, and External Factors. Communication as success factor encompasses the success factor of Communication and Collaboration indicating the significance of working relationships and approach to working with team members that are associated with communicating results to end-users.

2. Understanding Stakeholder Priorities

Framings of the same analysis are needed by executives, product managers, and operations teams; aligning the framing to the audience is a repeatable art that has a significant positive effect on adoption. Al-Hasan and Micheli, (2022)[2] examined the role of mental templates that managers employ to provide information with form and meaning in constructing their decision making with respect to the promotion, adaptation or prevention of process improvement approaches (PIAs). They found three cognitive frame types of managers, namely, conflicting, paradoxical, and supporting, which define the extent and scope of PIA application in new product development (NPD), at times opposed to organizational strategy, and proved the existence of how the frames of the organization may be substituted by the frames of a particular manager, revealing the substitution of the organizational frames by the individual managers. The Matrix provided by Mendelow was utilized by Alshawaf, (2025)[3] to position the principal stakeholders, including researchers, RDM professionals, institutional leadership, funding bodies and infrastructure providers, on the basis of their power and concern to establish customized communication strategies. His communication strategy facilitated the various needs of the stakeholders by establishing awareness, creating trust, dealing with obstacles, instilling acceptance, engaging with stakeholders, and leading to action to ensure successful implementation and adoption of Research Data Management (RDM) with a focus on the active involvement of key stakeholders by targeting the communication efforts to each stakeholder based on their respective roles, interests, and influence. Markey, (2024)[23] came up with two salient themes of TA perceptions of writing in statistics and data science that include writing as presentation and writing as contextualization which correlates with the necessity to fit framing to the audience. He also described that writing as contextualization meant that the statistician has to shape data and evidence to the requirements and purposes of a given audience in order to be persuasive, which directly corresponds to the question of the audience-specific framing,

and that when a statistician sees writing as contextualization, then the audience had the capacity to make informed judgments, and the requirements and their objectives shape the way they frame their message, which also confirmed the significance of audience-specific framing. Yanti et al (2024)[36] discussed how interactive capabilities in data visualization (i.e. dynamic filters and drill-downs) enable users to tailor the data display to the particular analysis requirements and underlie such a notion as adapting the framing to the audience. They showed that an interactive data visualization allows users to explore a smaller set of information and learn data in a further and more detailed way and that data visualization in financial reports can definitely enhance better understanding of stakeholders, and thus, showing data in a more digestible and more visual form, as opposed to the classic table, can lead to higher adoption.

3. Building Stakeholder Trust

Transparency in respect to assumptions, uncertainty and data limits are the most imperative. In the industry reports, there is positive executive concern on the reliability of analytics input e.g. 84% of CEOs interviewed expressed concern about the quality of data on which decisions were made which is a sign of a general question of trust which must be resolved through communication and governance. In a procedure that is depicted in Brous et al (2020)[6] , it is reported that the outcome of a decision made by data science initiatives is not consumed or embraced by decision-makers due to the uncertainty as to the quality of the data input that correlates also with the stated issue of data quality. They assert that data governance is increasingly being implemented by organizations as a means of increasing the levels of adoption and creation of trust in the outcomes of data science decisions and found that having a full-fledged data governance capacity is a pre-condition to being in a position to trust the outcomes of data science decisions to the extent that it is required to make fruitful decisions. According to Mahanti (2019)[21] , the key to data quality success is not so much technical excellence but rather soft, including connecting the top management, education, change resistance, and communication, which significantly contributes to overcoming the issue of trust that many CEOs continue to experience. He discussed how data quality could be a burden to the organizations because there is a lot of operations change involved and it is an intangible factor and why such massive effort is usually difficult even without compliance and regulatory intent and why such an endeavor was necessary in the first place and why data quality is such a concern causing stakeholders to lose hope. Passi and Jackson, (2018)[26] explored organizational actors as a way of generating and renegotiating under uncertain conditions of analysis through taking part in skepticism, assessment, and credibility practices. Investigating four common tensions of

applied data science work that have problems of trust: unambiguous numbers, counterintuitive knowledge, unbelievable data, and un-understandable models and, hence, to the collaborative nature of the work, the author explores. Zhang and Wang, (2025)[36] described how decaying digital trust is fuelled by increasing concerns over data privacy, security breaches, an absence of visibility of which algorithms these are, and inappropriate decision-making that is inaccessible through traditional data governance practices. According to them, to achieve confidence with AI choices and smart contracts, data quality and integrity should be handled carefully, and imperfect data is one of the key obstacles towards creating and sustaining digital trust on a smart market, which implied the creation of a multi-layered Data Governance Framework that was specifically aimed at building and preserving trust.

Managing Organizational Expectations

1. Misunderstandings About Capabilities and Timelines

Unrealistic sponsor expectations are caused by common assumptions, such as instantaneous, flawless projections. Timelines should specifically account for the phases of data engineering and validation since data cleaning usually requires the most work. According to Kordon (2020)[16], the most time-consuming step in the data science pipeline is data preparation, which explains why it takes so much time and effort. The inner, shorter model development loop comprises an iterative sequence of data preparation with corresponding analysis that can lead to model generation and validation, he said. Data preparation is a nontrivial, challenging-to-fully-automate process that includes procedures to collect, integrate, preprocess, and balance available data. According to Martins et al. (2025)[25], data cleaning is still one of the most important and time-consuming processes in contemporary data science since it has a direct impact on the correctness and dependability of downstream analytics. They frequently referenced research demonstrating that duties like identifying and fixing incorrect entries, reconciling duplication, and standardizing conflicting formats account for up to 80% of a data professional's work. Therefore, it has been observed that attempts at automated cleaning are complicated by the varied and "messy" character of real-world data, which ranges from partially organized records to high-frequency transactions, necessitating more advanced techniques. Data pre-processing, which includes cleaning, can take up 50% to 80% of the total classification process, according to Maharana et al. (2022)[22], highlighting its substantial effort and time impact. The concerns with the raw data itself (noise, corruption, missing, and inconsistency) and the necessary processes to follow for the optimal data analysis technique were the two major pre-processing issues that the writers addressed. Additionally, it was stated that data pre-

processing is essential because raw data is susceptible to noise, corruption, missing data, and inconsistency. To improve quality and avoid erroneous predictions, cleaning, integration, transformation, and reduction techniques are needed.

2. Expectation Calibration Framework

Precursors have been workshops we held early on (scoping studies), and 'deliverables', e.g., being available for testing at certain dates, with the data ready, using base models and pilot models, as well as educating stakeholders that could not or did not want to yet accept probability likelihoods etc. This mitigates the risk and temptation that projects may be only evaluated by their finally obtained accuracy, rather than incremental value capture. Lahiri & Saltz, (2024)[18] noted that data science projects fail often because of uncertain inputs and outputs, nebulous objectives, and project complexity suggesting the need for structured processes like early scoping to manage expectations and risk. They identified that data science project failures arise from numerous challenges, including unrealistic expectations, unstructured project execution, poor stakeholder management, and scope creep, which early interventions aim to mitigate and noted that current project management methodologies often overlook the diverse risks of sociotechnical systems and risk articulation inherent in data science lifecycles, justifying the need for a framework that manages risk incrementally rather than focusing solely on final outcomes. At each level of the data science project, Chollangi et al. (2024)[9] set important benchmarks and deliverables to signify the end of important stages, which naturally promotes incremental value capture. In a similar vein, they described a structured project process flow for data science projects that emphasizes defining a clear scope and guaranteeing both technical and business alignment throughout the project lifecycle, which lessens the focus solely on final accuracy. Lastly, they suggested gathering stakeholder feedback to comprehend their viewpoints and refine the model or processes based on their input, supporting iterative refinements and continuous adaptation rather than a single judgment based on final accuracy. According to Afzal et al. (2021)[1], a Data Readiness Report is an artifact that certifies the baseline quality of ingested data by documenting operations and remediations carried out on the data. This naturally enables incremental value evaluation and milestone-based deliverables. A shareable data asset allows stakeholders to comprehend data quality before the final model output by providing a comprehensive overview of data readiness for machine learning tasks, documenting data quality issues, and making it easy to advertise the data in a data-as-a-service model. By completing the AI documentation pipeline with Datasheets,

Dataset Nutrition Label, FactSheets, and Model Cards, which promote communication and transparency beyond final model accuracy, this boosts confidence and dependability.

3. Aligning Business Goals and Analytical Outcomes

Measure downstream impact and link model success indicators to business KPIs. The effort will struggle for funding and adoption if the firm cannot link analytics outputs to quantifiable business outcomes. According to Sudhakar et al. (2025)[32], the idea that initiatives must demonstrate quantifiable financial outcomes in order to succeed is implicitly supported by the argument that as digital marketing grows, transparent and accountable measurement that links marketing investments directly to financial results becomes essential for sustainable growth. The results demonstrated that companies that use integration have more accurate ROI estimations, which helps them manage marketing resources more efficiently and provide clear financial justification for investments. Additionally, it demonstrates that improving Return on Investment (ROI) requires bridging the gap between financial measurements and marketing analytics, and that integrating financial data with marketing insights provides firms with a comprehensive picture of investments and results. According to Liu et al. (2018)[20], a business analytics program can effectively enhance strategic decision-making by obtaining actionable intelligence, suggesting a link between analytics and favorable results. noted that not knowing how to use analytics to improve business performance is the largest obstacle to adoption, implying that tying analytics to performance is essential for adoption. The necessity for funding and resources connected to results is directly addressed by the statement that the analytics leader must successfully convey the advantages of analytics in order to gain the trust and resources required for performance. Rahman (2025)[28] showed that companies with greater analytics maturity had far better performance outcomes, such as profitability, efficiency, decision quality, and return on investment. This suggests a clear connection between success and quantifiable results. He confirmed that the methodical integration of analytics into strategic processes results in quantifiable improvements in efficiency, customer value, and financial performance, supporting the significance of linking analytics to quantifiable outcomes. He also mentioned that financial analytics allows organizations to measure, forecast, and optimize financial performance, demonstrating that the capacity to measure is a fundamental component of successful analytics adoption.

Integrating Solutions: A Unified Approach

Technical investments (data catalogs, lineage, automated tests), communication procedures (model cards, business-facing visualizations, narrative one-pagers), and governance rhythms

(quarterly steering, cross-functional KPIs) are all combined in a single program. The scope of the issue (multi-million annual cost estimates and percent-of-revenue implications mentioned above) supports the ROI for these solutions. By measuring the substantial yearly recurring economic value—calculated at USD 11.3 million—achieved by applying a data governance framework to production data for 1,700 wells, Huff & Lee (2020)[14] illustrated the need for a methodical approach. Describing a nine-step framework for data governance that includes technical elements (e.g., creating an appropriate data catalog, designing enabling technologies), communication processes (e.g., investing in organizational change management, communicating data importance), and governance rhythms (e.g., implementing continuous data quality monitoring, facilitating and stating that the disconnect between departments frequently results in significant "hard-dollar" loss, contributing to unknown potential value loss, which the systematic data governance procedure is designed to address. Redman (2016)[29] calculated that the annual cost of subpar data in the US alone is \$3.1 trillion, an astounding amount that warrants a large investment in data quality enhancement. He found that decision-makers, managers, and data scientists must account for errors in their daily work, which is costly and time-consuming. He also noted that the process of verifying and fixing data errors quickly becomes a routine but ineffective aspect of work life, pointing to a systemic issue that calls for a cohesive program to address.

Case Illustration

After automating point-of-sale data validation rules, holding weekly business-data syncs with straightforward scenario visualizations, and resetting forecast precision targets to realistic bands given data limitations, a retail team was able to reduce stock-outs by 12%. Within three months, quantifiable operational progress was achieved through the mix of communication, governance, and calibrated expectations. Conflicts occur when expectations are not fulfilled, according to Huff & Lee (2020)[14], highlighting the significance of establishing expectations up front on how to address data quality issues in order to reduce misunderstandings. They also found that leadership must inform stakeholders on the state and significance of developing data governance on a regular basis. In a similar vein, they stated that the realized economic impact for a top natural resources business was estimated to be USD 11.3 million in yearly recurring savings for production data alone for 1,700 wells following the use of the nine-step data governance framework. According to Purohit et al. (2024)[27], by securing investment and improving data quality, an automated, integrated, and comprehensive web-based workflow solution helped overcome obstacles and produced savings of almost \$7.5 million. Noting that the

workflow's automation and digitalization, along with the integration of data governance practices and policies, increased the effectiveness of managing the business process, resulting in significant outcomes, and outlining how data governance through automated processes guarantees that significant data assets are managed effectively throughout their life cycle and ensures appropriate time management for the organization, which addresses the problem of geoscientists and engineers wasting valuable time on low-level tasks.

Recommendations

How to enhance data utilization and collaboration?

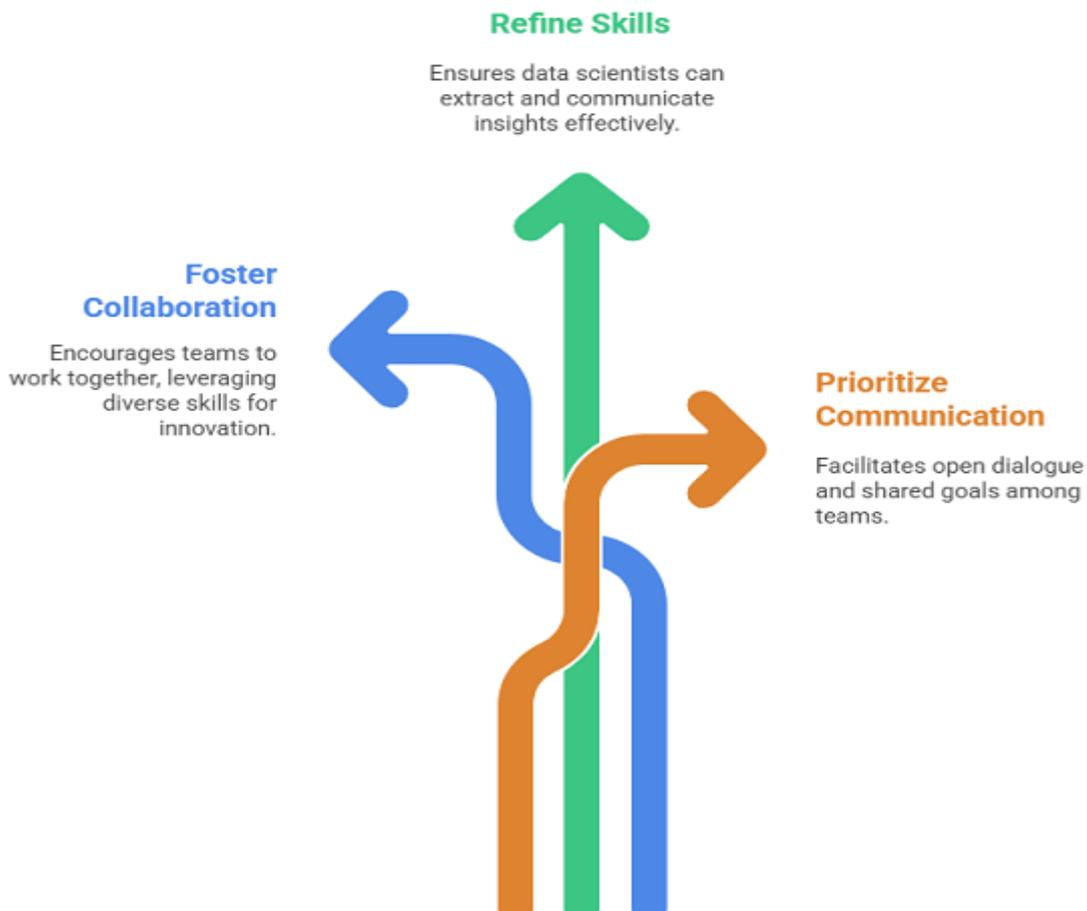


Figure 2: Ways to Enhance Data Utilization and Collaboration

For Organizations

- Invest in governance and data quality tools. Analyst estimates of average enterprise costs of about \$12.9 million annually based on inadequate data serve as the foundation for the business case.
- Calculate the price of inaccurate data. This issue is often underestimated by organizations; monitoring it can free up funds for solutions.

- Arrange the resources needed to prepare the data. Since data preparation typically accounts for about 60% of analysis time, realistic staffing and scheduling should take this into account.

For Data Scientists

- Create model cards that highlight data limitations and uncertainty.
- To make "data quality" a visible, quantifiable focus of management attention, use dashboards with data-health measures (missing rates, freshness, duplicates).

For Cross-Functional Teams

Before committing to timetables, make sure sponsor groups participate in data-readiness milestones and establish agreed KPIs and governance cadence.

Conclusion

Organizations seeking dependable, repeatable analytics must address the triple problem of data quality, stakeholder communication, and expectation management. The evidence is clear: this is a top priority due to the multimillion-dollar annual expenditures and significant time sinks for practitioners. Teams can decrease hidden costs and increase quantifiable economic value by integrating governance, communication, and calibration. This study shows that the success of every data science project depends on the careful coordination of three interrelated pillars: realistic expectation management, clear communication, and high-quality data. Poor data quality continues to be a multi-million-dollar burden, according to industry reports and recent research, and practitioners still spend the majority of their time fixing avoidable data problems. The study also shows that when stakeholders are unable to comprehend results, trust outputs, or incorporate insights into operational decision-making, even technically competent models fall short. Projects are also jeopardized when expectations fail to account for the data-dependent, probabilistic, and iterative character of analytical work. This study provides a single route for businesses looking for reliable, repeatable, and value-generating analytics by combining insights from governance frameworks, socio-technical communication practices, and structured expectation-calibration tactics. The case study demonstrates that when these components work together rather than separately, quantifiable performance benefits result. In the end, addressing the triple challenge is a strategic requirement for businesses hoping to get long-term benefits from data-driven systems in an increasingly complicated digital environment, not a technical luxury.

References

- [1] Afzal, S., Rajmohan, C., Kesarwani, M., Mehta, S., & Patel, H. (2021). Data Readiness Report. *2021 IEEE International Conference on Smart Data Services (SMDS)*, 42–51. <https://doi.org/10.1109/SMDS53860.2021.00016>
- [2] Al-Hasan R, Micheli P (2022), "How managers' cognitive frames affect the use of process improvement approaches in new product development". *International Journal of Operations & Production Management*, Vol. 42 No. 8 pp. 1229–1271, doi: <https://doi.org/10.1108/IJOPM-12-2021-0758>
- [3] Alshawaf, F. (2025). Strategies for Embedding Research Data Management Through Effective Communication. *Data*, 10(6), 83. <https://doi.org/10.3390/data10060083>
- [4] Ateş, V., & Garip, A. (2025). Data Governance for Businesses: Challenges, Recommendations, and Critical Success Factors. *Acta Infologica*, 9(1), Article 1. <https://doi.org/10.26650/acin.1540759>
- [5] Bassi, C. A., & Alves-Souza, S. N. (2025). Challenges and Solutions in Implementing Data Governance: A Literature Review. *SN Computer Science*, 6(8), 945. <https://doi.org/10.1007/s42979-025-04473-5>
- [6] Brous, P., Janssen, M., Brous, P., & Janssen, M. (2020). Trusted Decision-Making: Data Governance for Creating Trust in Data Science Decision Outcomes. *Administrative Sciences*, 10(4), 81. <https://doi.org/10.3390/admsci10040081>
- [7] Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14(1), 2. <https://doi.org/10.5334/dsj-2015-002>
- [8] Chimedessa, S. G., & Chawla, A. S. (2025). The Status of Servant Leadership, Organizational Commitment, and Organizational Citizenship Behavior: A Case Study of State Owned HEI'S In Southern Ethiopia. *Journal of Philanthropy and Marketing*, 5(1), 199–206.
- [9] Chollangi, H., Vaddhiparthy, S. S., & Vaddhiparthy, S. S. S. (2024). *A Report on Data Science Project: Management, Strategy, and Key Deliverables*. <https://doi.org/10.36227/techrxiv.173092369.97726669/v1>
- [10] Demir, N., Aysolmaz, B., & Özcan-Top, Ö. (2024, September 10). Critical Success Factors in Data Analytics Projects: Insights from a Systematic Literature Review. *Disruptive Innovation in a Digitally Connected Healthy World*. Conference on e-Business, e-Services and e-Society, Switzerland. https://doi.org/10.1007/978-3-031-72234-9_11
- [11] Derakhshan, P., Azadmanjir, Z., Naghdi, K., Habibi Arejan, R., Safdarian, M., Zarei, M. R., Jazayeri, S. B., Sharif-Alhoseini, M., Arab Kheradmand, J., Amirjamshidi, A., Ghodsi, Z.,

- Faghieh Jooybari, M., Mohammadzadeh, M., Khazaeipour, Z., Abdollah Zadegan, S., Abedi, A., Oreilly, G., Noonan, V., Benzel, E. C., ... Rahimi-Movaghar, V. (2021). The impact of data quality assurance and control solutions on the completeness, accuracy, and consistency of data in a national spinal cord injury registry of Iran (NSCIR-IR). *Spinal Cord Series and Cases*, 7(1), 51. <https://doi.org/10.1038/s41394-020-00358-2>
- [12] Gibbs, M. R., Shanks, G., & Lederman, R. (2005). *Data Quality, Database Fragmentation and Information Privacy | Surveillance & Society*. 3(1). <https://doi.org/10.24908/ss.v3i1.3319>
- [13] Hoang, D. (2023). *The emphasis of data quality in the data harmonization process* [fi=AMK-opinnäytetyö|sv=YH-examensarbete|en=Bachelor's thesis]]. <http://www.theseus.fi/handle/10024/799931>
- [14] Huff, E., & Lee, J. (2020). *Data as a Strategic Asset: Improving Results Through a Systematic Data Governance Framework*. <https://doi.org/10.2118/198950-MS>
- [15] *Information Technology Managers' Strategies for Implementing Data Governance—ProQuest*. (n.d.). Retrieved 9 December 2025, from <https://www.proquest.com/openview/3f5b0401e23d6bac423340fd5c4b140f/1?pq-origsite=gscholar&cbl=18750&diss=y>
- [16] Kordon, A. K. (2020). The AI-Based Data Science Workflow. In *Applying Data Science* (pp. 189–202). Springer. https://doi.org/10.1007/978-3-030-36375-8_6
- [17] Kusumawati, R. (2023). Integrating Big Data Analytics into Supply Chain Management: Overcoming Data Silos to Improve Real-Time Decision-Making. *International Journal of Advanced Computational Methodologies and Emerging Technologies*, 15(2), 17–26.
- [18] Lahiri, S., & Saltz, J. (2024). The need for a risk management framework for data science projects: A systematic literature review. *International Journal of Information Systems and Project Management*, 12(4). <https://aisel.aisnet.org/ijispm/vol12/iss4/4>
- [19] Lebaea, R., Roshe, Y., Ntontela, S., & Thango, B. A. (2024). *The Role of Data Governance in Ensuring System Success and Long-Term IT Performance: A Systematic Review*. Business, Economics and Management. <https://doi.org/10.20944/preprints202410.1841.v1>
- [20] Liu, Yi., Han, H., & DeBello, J. (2018). The Challenges of Business Analytics: Successes and Failures. *Hawaii International Conference on System Sciences 2018 (HICSS-51)*. https://aisel.aisnet.org/hicss-51/da/business_intelligence_case_studies/4
- [21] Mahanti, R. (2019). *Data Quality: Dimensions, Measurement, Strategy, Management, and Governance*. Quality Press.

<https://books.google.com.ng/books?id=semiEAAAQBAJ&lpg=PT5&ots=pRIJNLFPIh&lr&pg=PT5#v=onepage&q&f=false>

- [22] Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1), 91–99. <https://doi.org/10.1016/j.gltp.2022.04.020>
- [23] Markey, B. (2024). Presenting and Making Relevant: A Semantic Frame Analysis of Teaching Assistant Perceptions of Writing in Statistics. *2024 IEEE International Professional Communication Conference (ProComm)*, 41–47. <https://doi.org/10.1109/ProComm61427.2024.00014>
- [24] Martinez, I., Viles, E., & Olaizola, I. G. (2021). A survey study of success factors in data science projects. *2021 IEEE International Conference on Big Data (Big Data)*, 2313–2318. <https://doi.org/10.1109/BigData52589.2021.9671588>
- [25] Martins, P., Cardoso, F., Váz, P., Silva, J., Abbasi, M., Martins, P., Cardoso, F., Váz, P., Silva, J., & Abbasi, M. (2025). Performance and Scalability of Data Cleaning and Preprocessing Tools: A Benchmark on Large Real-World Datasets. *Data*, 10(5), 68. <https://doi.org/10.3390/data10050068>
- [26] Passi, S., & Jackson, S. J. (2018). Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), 136:1-136:28. <https://doi.org/10.1145/3274405>
- [27] Purohit, P., Al Nuaimi, F., & Nakkolakkal, S. (2024). *Data Governance, Privacy, Data Sharing Challenges*. GOTECH, Dubai, UAE. <https://doi.org/10.2118/219172-MS>
- [28] Rahman, M. M. (2025). Data Analytics for Strategic Business Development: A Systematic Review Analyzing Its Role In Informing Decisions, Optimizing Processes, and Driving Growth. *Journal of Sustainable Development and Policy*, 1(01), 285–314. <https://doi.org/10.63125/he1tfg25>
- [29] Redman, T. C. (2016). *Bad Data Costs the U.S. \$3 Trillion Per Year*. <https://dataladder.com/wp-content/uploads/2019/07/Bad-Data-Costs-the-U.S-3-Trillion-Per-Year.pdf>
- [30] Sargiotis, D. (2024). *Data Quality Management: Ensuring Accuracy and Reliability* (pp. 197–216). https://doi.org/10.1007/978-3-031-67268-2_5
- [31] Sienkiewicz, M. (2025, August 18). From Data Silos to Data Mesh: A Case Study in Financial Data Architecture. *Database and Expert Systems Applications*. International

- Conference on Database and Expert Systems Applications. https://doi.org/10.1007/978-3-032-02049-9_1
- [32] Sudhakar, D., Sarma, J. G., Deepika, A., & Rajesh, K. P. R. (2025). The Status of Servant Leadership, Organizational Commitment, and Organizational Citizenship Behavior: A Case Study of State Owned HEI'S In Southern Ethiopia | Journal of Philanthropy and Marketing. <http://journalofphilanthropyandmarketing.org/index.php/JPM/article/view/111>
- [33] Suh, A., Appleby, G., Anderson, E. W., Finelli, L., Chang, R., & Cashman, D. (2023). *Are Metrics Enough? Guidelines for Communicating and Visualizing Predictive Models to Subject Matter Experts* | IEEE Journals & Magazine | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10077087>
- [34] Weiner, J. (2022). *Why AI/Data Science Projects Fail*. Retrieved 10 December 2025, from <https://books.google.com/books/about/Why AI Data Science Projects Fail.html?id=x4VyEAAAQBAI>
- [35] Xiao, Y., Bochner, A. F., Makunike, B., Holec, M., Xaba, S., Tshimanga, M., Chitimbire, V., Barnhart, S., & Feldacker, C. (2017). *Challenges in data quality: The influence of data quality assessments on data availability and completeness in a voluntary medical male circumcision programme in Zimbabwe*. <https://doi.org/10.1136/bmjopen-2016-013562>
- [36] Yanti, N. L. P. A. T., Antari, G. A. P. D., Rizelputra, P. K. S., & Candra, I. G. S. W. (2024). Implementation of Data Visualization in Presentation of Financial Reports to Improve Stakeholder Understanding. *Journal of Information Systems, Digitization and Business*, 3(3), 32–38. <https://doi.org/10.38142/jisdb.v3i3.1433>
- [37] Zhang, L., & Wang, W. (2025). Data Governance and Digital Trust in Smart Markets. *Journal of Electronic Commerce*. <https://doi.org/%20joecm.3.2.15564.35125656565005>