Decoding DQM for Experimental Insights on Data Quality Metadata's Impact on Decision-Making Process Efficacy

Zeinab Rahimi Rise
Department of Industrial Engineering and Management Systems; Amirkabir University of Technology; Tehran, Iran;
Email: zeinab.rahimi@aut.ac.ir

Mohammad Mahdi Ershadi*
Department of Industrial Engineering and Management Systems; Amirkabir University of Technology; Tehran, Iran;
Email: ershadi.mm1372@aut.ac.ir

Mohammad Javad Ershadi
Information Technology Department; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran;
Email: ershadi@irandoc.ac.ir

Iranian Journal of Information Processing and Management

Received: 08, Jan. 2024 | **Accepted:** 12, Feb. 2024

Abstract: Decision-making processes are significantly influenced by a myriad of factors, with data quality emerging as a crucial determinant. Despite widespread awareness of the detrimental effects of poor-quality data on decisions, organizations struggle with persistent challenges because of the sheer volume of data within their systems. Existing literature advocates for providing Data Quality Metadata (DQM) to help decision-makers communicate data quality levels. However, concerns about potential cognitive overload induced by DQM may hinder decision-makers and negatively impact outcomes. To address this concern, we conducted an experimental exploration into the impact of Data Quality Management (DQM) on decision outcomes. Our study aimed to identify specific groups of decision-makers benefiting from DQM and uncover factors influencing its usage. Statistical analyses revealed that decision-makers with a heightened awareness of data quality demonstrated improved Data Quality Management (DQM) utilization, leading to increased decision accuracy. Nevertheless, a trade-off was observed as the

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Special Issue | Winter 2025 | pp. 41-80
Exploring the Relationship Between Data Quality and Business Process Management
https://doi.org/10.22034/jipm.2025.2019855.1504

* Corresponding Author

efficiency of decision-makers suffered when employing Decision Quality Management (DQM). We propose that the positive impact of incorporating Data Quality Management (DQM) on decision outcomes is contingent on characteristics such as a high level of knowledge about data quality. However, we acknowledge that the inference of this positive impact could be more transparent and thoroughly explained. Our findings caution against a blanket inclusion of Data Quality Management (DQM) in data warehouses, emphasizing the need for tailored investigations into its utility and impact within specific organizational settings.

Keywords: Data Quality Metadata (DQM), Decision Support Systems, Data Quality, Decision Strategy

1. Introduction

The landscape of traditional systems grapples persistently with formidable data quality (DQ) challenges, a struggle amplified within decision support systems. In the era where organizations increasingly rely on these systems for informed decision-making, the integration of data from diverse sources introduces significant hurdles in managing data quality (DQ) issues. Soft data analysis, an integral part of strategic planning, adds another layer of complexity by requiring the integration of subjective assessments and forecasts into decision-making processes.

Despite decision-makers' intuitive understanding of data, the widespread use of data warehouses and the various repurposing of data reduce the reliability of this intuitive approach. Decision-makers who are unfamiliar with the intricacies of the data may hesitate due to legitimate concerns about verifying data quality. These challenges underscore the pivotal role of data quality (DQ) in decision-making, especially within the intricate environment of data warehouses that encounter a variety of DQ issues.

To address the adverse effects of poor data quality on decision outcomes, literature has been vocal in advocating for the integration of Data Quality Metadata (DQM). Recognizing the data quality (DQ) level and understanding its context dependency becomes crucial in this endeavor. Data Quality Management (DQM) emerges as an invaluable tool for decision-makers to assess the appropriateness of DQ levels within the specific context of the task at hand, ultimately streamlining decision-making processes. The vitality of Data Quality Management (DQM) relies on maintaining standards across critical dimensions such as accuracy, completeness, and timeliness. However, the integration of Data Quality Management (DQM) with raw data demands careful scrutiny, as potential financial and operational implications loom large.

Despite the acknowledgment that the positive impact of DQM is not universal, our study takes a laser-focused approach by delving into its influence on decision outcomes, especially in the context of bankruptcy prediction using the Altman-Z model. Our objectives include identifying decision-makers who would benefit the most from the integration of DQM and examining the factors that either facilitate or impede its utilization in decision-making processes. This research aims to provide clear guidelines for system designers by assessing the importance of Data Quality Management (DQM) alongside established dimensions. It also introduces new variables such as Data Quality Assessment (DQA) to enhance the understanding of decision-making processes.

This inclusive approach to our research methodology mitigates potential experimental biases, contributing substantially to the redefinition of decision outcome measures within a homogeneous environment. By providing specific examples or statistics related to data quality challenges and the impact of data quality management (DQM), we aim to further emphasize the critical nature of the research problem and its broader implications in the realm of decision-making. Accordingly, the research questions are as follows:

- ◇ RQ1: To what extent does Data Quality Metadata (DQM) influence decision outcomes in the context of bankruptcy prediction using the Altman-Z model?
- ◇ RQ2: How do decision-makers with varying levels of data quality awareness benefit from the incorporation of DQM in their decision-making processes?
- ◇ RQ3: What factors facilitate or impede the utilization of Data Quality Metadata (DQM) in decision-making, particularly in the realm of bankruptcy prediction?
- ◇ RQ4: Is there a trade-off between decision accuracy and efficiency when decision-makers employ Data Quality Metadata (DQM)?

The paper unfolds as follows: the subsequent section conducts a thorough review of prior DQM research. Following that, the third section delves into the intricacies of our research design, with the fourth section unveiling and elucidating our findings. Ultimately, the paper concludes by offering reflective insights and charting potential avenues for future research endeavors.

2. Background

2-1. Data Quality

In the contemporary landscape, the issue of data quality (DQ) has emerged as a central concern for organizations grappling with vast datasets, garnering significant attention in academic research. DQ studies can be broadly categorized into two streams: intrinsic and contextual. Intrinsic research focuses on the inherent value of data, independent of their context, while contextual studies consider factors such as intended data purposes and characteristics of users. Prior research emphasizes the significant impact of contextual factors on DQ assessments. Recognizing the multidimensional nature of Data Quality (DQ), they advocate for measuring data items based on user perceptions. However, integrating contextual data quality assessment introduces complexity. For instance, a sales sheet without “cost” data may be considered insufficient for making production efficiency decisions, but it may be sufficient for inventory decisions aimed at reconciling quantities. Acknowledging the contextual nature of data quality (DQ) becomes imperative for effective DQ management. Decision-makers need to assess data quality (DQ) levels relevant to specific tasks, emphasizing the rationale behind recent research advocating for the integration of Data Quality Management (DQM) with data in decision support systems. This recognition of context not only poses challenges but also holds the potential to enhance data quality management within databases, aligning data quality with the specific needs and objectives of decision-makers.

2-2. Data quality metadata (DQM)

Data Quality Metadata (DQM) is crucial for evaluating data excellence in organizational databases, encompassing dimensions like precision and timeliness. As a vital facet of data, DQM’s objectivity during production enhances its significance. Its inception involves the intriguing process of DQ tagging, and information systems feature various metadata types, such as varieties of metadata explored in academic and general discourses (see Table 1 and Table 2).

Table 1. Varieties of Metadata Explored in Academic Discourse

Types of metadata	Description
Data quality metadata	It unveils the caliber of precise data housed within databases, like spotlighting that sales figures boast a 90% completeness for the vibrant month of January 2022.
Descriptive Metadata	It unravels the intricacies of data, encompassing facets such as purpose, authorship, title, and more.
Terms and conditions metadata	It delves into the nuanced conditions dictating the permissible (or prohibited) use of data, navigating the realm of intellectual property rights and beyond.
Administrative metadata	It illuminates the temporal and procedural dimensions of data genesis, shedding light on creation timelines and delineating the privileged few who hold access keys.
Data dictionary metadata	It deciphers the tapestry of meanings and interconnections woven within the data, unraveling the narrative that binds each piece.
Structural metadata	It embarks on a journey through the syntactical landscape of data, exploring the structural nuances and foundational types that underpin the data records.

Table 2. Varieties of Metadata Explored in General Discourse

DQM formats	Description
Ordinal	The assessment discerns if data quality surpasses the norm or falls below, neatly classifying it as stellar, commendable, mediocre, and so on.
Interval	Gauging the DQ on a 0-100 spectrum, it cleverly mirrors heightened precision with an elevated DQ level.
Probability	By adopting a 0-1 probability gauge, it unveils the likelihood that the data stands correct, encapsulating the DQ level adeptly.
Range	Pinpointing both the floor and ceiling of a designated data set, it unveils the comprehensive range with precision.
Graphical	Employing a vibrant palette, it paints a vivid picture of the DQ level, ensuring a visual representation that captures the essence of each data set.

In the domain of Data Quality (DQ) tagging, challenges arise due to the absence of well-defined guidelines for optimal Data Quality Management (DQM) levels within databases. Choices range from individual data items to relational tables, but the literature lacks a comprehensive exploration of the merits and drawbacks associated with these different levels. Past researchers often work at the level of individual data items. Determining the context-dependent data quality dimension(s) for encapsulating quality measures as Data Quality Management (DQM) presents a second challenge. This study aligns with the literature, emphasizing the fundamental nature of the accuracy dimension. The third consideration involves the format of DQM creation, maintenance, and presentation, which is crucial for positively impacting the decision-making process. Previous explorations in Data Quality Management (DQM) have primarily focused on attribute-level tagging, interval representations, task complexities, and decision-making strategies. Innovatively, our research introduces a distinctive approach. While incorporating two tiers of task complexity and attribute DQ tagging for benchmarking, we uniquely scrutinize decision-makers active integration of DQM alongside the provided data. This goes beyond altered decision preferences, providing explicit insights into decision-makers responsiveness to Data Quality Management (DQM) in their strategic choices. Additionally, our paper evaluates how Data Quality Management (DQM) influences the efficacy of decision outcomes, providing a comprehensive analysis of its impact (refer to Table 3).

Table 3. Measurements of decision outcomes discussed in DQM literature

Decision outcome assessment	Description	How to measure
Complacency	Assesses the incorporation of the novel variable, specifically the DQM, into the decision-making process. If decisions yield comparable outcomes with or without DQM, it indicates the decision maker’s indifference to its influence.	Assessing complacency involves gauging whether the favored option varies between cohorts utilizing DQM and those without.
Consensus	Evaluates the consensus among decision-makers regarding outcomes with and without the incorporation of DQM.	Consensus delves into the ratio of optimal choices within both DQM-empowered and non-DQM groups, allowing for potential divergence in the best choices.

Decision outcome assessment	Description	How to measure
Consistency	Addresses the hierarchical arrangement of alternatives based on preference, ranging from the most favored to the least favored.	Evaluating consistency entails correlating the mean rankings of alternatives among decision-makers in groups with and without DQM.
Efficiency	Quantifies the time invested by decision-makers in completing a specific decision task.	Measure by scrutinizing the time discrepancy between groups employing and eschewing DQM to glean insights into the decision-making process.
Confidence	Examines how the presence or absence of DQM influences the confidence levels of decision makers, defining it as their assurance in the accuracy of the choices they make.	Quantify by comparing the confidence levels documented for a specific decision task across diverse test groups, revealing nuanced insights into decision-making dynamics.

2-3. Relevant variables for the use of DQM

2-3-1. Decision-making strategy

In scenarios demanding a systematic approach where decision-makers gather pertinent information, structure decision scenarios with a spectrum of alternatives, and ultimately select the optimum one, the analytical decision strategy takes precedence due to its objectivity. Regarding the literature review, Table 4 is represented for an elaborate description of these decision strategies.

Table 4. Summaries of 3 decision strategies in the literature

Decision strategies	Description
Weighted Additive (WA) - bdolkhani et al., [1]	It opts for the maximum sum value derived from the products of each criterion and their respective values. Let W1a1 & V1a1 and W2a1 & V2a1, and W1a2 & V1a2 and W2a2 & V2a2 represent the weight and value of criterion 1 (C1) and criterion 2 (C2) for alternative 1 (A1) and alternative 2 (A2) respectively. The WA strategy would compare the value of (W1a1 x V1a1) + (W2a1 x V2a1) for A1 and (W1a2 x V1a2) + (W2a2 x V2a2) for A2 and selects the alternative with the highest value as an optimal solution of the decision task.
Elimination By Attribute (EBA) - Valencia Parra [22]	It evaluates various options by weighing them against a key attribute, known as the decisive factor. Subsequently, it dismisses alternatives with inferior values for this crucial attribute. Typically, the deciding attribute is chosen for its substantial weight or influential role in the decision-making process. If for example C1 is a decisive criterion because of its highest weight for a specific decision task, the EBA decision strategy would compare the values of (W1a1 x V1a1) and (W1a2 x V1a2) for A1 and A2 respectively, and selects an alternative with the highest value as an optimal decision outcome.
Conjunctive (CON) - Hamlin [7]	It establishes a threshold for every criterion, handpicking alternatives that surpass this benchmark for each aspect to arrive at an optimal decision. Business experts set the cut-off value, its fluctuation is dependent on the nature of the task. Notably, a cut-off of 50 might be apt for certain tasks, yet prove inadequate for others. Typically, an alternative scoring above the cut-off across the majority of criteria emerges as the optimal solution. Consider a decision task with a specified cut-off of 50 for each criterion. If V1a1 is 50 and V2a1 is 45 for A1 and V1a2 is 65 and V2a2 is 50, the CON strategy would select the alternative with the values for all or most of the criteria above 50. In this case, the optimal decision outcome would be A2.

2-3-2. Experience

It is rational to presume that expertise plays a pivotal role in decision-making, enhancing the process through the integration of a lifetime's worth of knowledge. Seasoned decision-makers excel in error detection and grasp crucial elements of a decision, surpassing the capabilities of their less-experienced counterparts. A critical factor to consider is the cognitive limit of decision-makers, which can be enhanced by a wealth of lifetime experiences. The interplay between working memory (WM) and long-term memory (LM) establishes the cognitive threshold. Working Memory (WM) temporarily stores data, while Long-Term Memory (LM) houses information linked to lifelong encounters. Working memory

(WM) temporarily stores data, integrating it with long-term memory (LM) data to conceptually represent a specific decision task. If this representation proves insufficient, Working Memory (WM) extracts data from Long-Term Memory (LM), applying logical rules to explore alternative conceptualizations until an optimal solution emerges or the capacity of WM is exceeded. Research underscores that skilled decision-makers create well-organized conceptual representations, which lead to optimal solutions. However, assuming that experience guarantees optimal decision outcomes is not universally accurate. Familiarity with data can influence attitudes towards a specific dataset, leading decision-makers to depend more on previous knowledge rather than objectively evaluating all the information at hand. This inclination might prematurely conclude decision-making processes, adversely affecting outcomes. Studies indicate a correlation between higher levels of experience and the utilization of Decision Quality Management (DQM) in decision-making. However, the impact of experience level and type on DQM usage varies.

2-3-3. Time

Decision-making time is a precious commodity, demanding judicious utilization by those in charge. Efficiency is paramount, and the introduction of Decision Quality Management (DQM) should ideally enhance both decision-making time and effectiveness. If the integration of DQM results in a prolonged decision-making process without a commensurate improvement in decision effectiveness, its influence on decision-making may be considered unfavorable. A fascinating realm explored by researchers delves into decision-making under time pressure, an aspect measured diversely across studies. Various researchers have quantified time pressure using different methodologies. For some, it entailed setting a specific time duration for a task. Conversely, others distinguish between time constraints and time pressure, defining the former as a designated time allowance and the latter as a subjective response to the allotted time for decision-making. The perception of time pressure arises when decision-makers perceive the allocated time as insufficient for completing the task at hand. This study, unique in its approach, refrained from imposing explicit time constraints on participants. Instead, participants were asked to record the start and end times of their involvement in the experiment, providing a detailed insight into the temporal dynamics involved.

2-3-4. Data Quality Awareness

One of marketing's primary objectives is cultivating and sustaining brand awareness, a critical endeavor in an age where consumers actively seek information to gauge their brand preferences. The influence of brand awareness on consumer decision-making is notable, with heightened awareness correlating to an increased likelihood of brands being considered and chosen. Customers often express sentiments like, "I opt for the brand I'm familiar with," and, "I've encountered the brand frequently, so it must be of high quality." Similarly, decision-makers lacking Data Quality (DQ) awareness may underutilize available DQ metrics in decision support systems. DQ practitioners emphasize the significance of fostering DQ awareness to address DQ issues effectively. While numerous organizations currently grapple with DQ challenges, there's a tendency to overlook DQ awareness. This oversight may stem from a lack of clear understanding regarding the repercussions of poor DQ on organizational performance. Despite an intuitive sense that DQ awareness enhances the utilization of DQ metrics by keeping decision-makers vigilant, its actual impact on DQM utilization and subsequent effects on decision performance remain unexplored. Consequently, this paper's experiment delves into unraveling the influence of DQ awareness on the utilization and outcomes of DQM in decision performance.

2-3-5. Task complexity

The complexity of a task depends on several factors, such as the amount of relevant information available—measured by the number of decision options and characteristics—and the time given for decision-making. As the data to be processed for a given decision task increases, so does the task complexity. Previous studies have defined task complexity by considering the number of cells in the decision alternatives and criteria matrix. A task with 20 or fewer cells falls into the realm of simplicity, while surpassing this threshold categorizes a decision task as complex. This study uses this benchmark to differentiate tasks as either simple or complex.

2-4. Literature review

The literature review covers a wide range of studies that address critical aspects of data quality (DQ) and its management across various domains. The following is a brief summary of the main findings and contributions of each study.

Wolff [23] focuses on enhancing the metadata quality of scientific publications using literature-based approaches. The study identifies common errors and formulates a measurable process for improvement. The workflow, validated using Open Researcher and Contributor ID (ORCID) data, demonstrates a 56% correction rate, with potential applications to various source systems. Shankaranarayanan & Bin Zhu [18] introduce a prototype decision support system to alleviate cognitive overload caused by data quality metadata (DQM) in decision-making. The study explores the effectiveness of the system in reducing mental demand and improving decision performance, emphasizing the seamless integration of DQM into decision tasks. Serra et al., [16] evaluated 13 state-of-the-art tools for data quality, identifying strengths and potential areas for improvement. In a follow-up study, they emphasize the role of context in data quality management (DQM), investigating recent proposals and proposing a Context-Aware Data Quality Management (CaDQM) methodology. Zhou et al., [26] address the challenge of evaluating data quality in repurposed datasets. Their experiment, utilizing eye-tracking and retrospective think-aloud analysis, offers crucial insights into how users interact with metadata during data repurposing tasks, laying the groundwork for enhanced systems and tools.

Bonyadi et al., [3] highlight the pivotal role of data quality in Artificial Intelligence initiatives and propose BIGOWL4DQ, an ontology-driven approach for Big Data quality meta-modeling, selection, and reasoning. They emphasize the importance of using machine-readable and processable vocabularies to streamline data quality analysis in complex data scenarios. Mäkitalo [12] emphasizes the financial implications of poor data quality and advocates for ensuring good data quality to unlock the benefits of data-driven organizations. The paper introduces a customized data quality maturity tool for enhancing data quality within organizations. Pratiktio et al., [14] analyzed data quality methodologies, offering guidelines and processes based on the Data Management Body of Knowledge to improve data quality in applications or organizations. The study contributes to maintaining consistency in high-quality data for companies. Ngueilbaye, Alladoumbaye, et al., [13] propose a data quality model that incorporates a canonical data model and Benford's Law to evaluate the WHO's COVID-19 data reporting in the CEMAC region. The model serves as a reliability indicator for inspecting large datasets.

Khaleghian, Hossein, and Yongwei Shan [9] propose a framework with five metrics for recognizing the importance of high-quality data in decision-making, focusing on sewer inspection data quality. The framework contributes a robust data quality evaluation framework for sewer system data, crucial for effective sewer asset management. Hamlin [7] proposes a consensus-based data quality assessment (DQA) model for PROM information in digital quality measurement programs, addressing the current scarcity of effective guidance. Valencia Parra [22] focuses on Big Data pipelines, Data Preparation, Data Quality assessment, and Data Analysis stages, introducing Domain-Specific Languages (DSLs) for complex data transformation. Implemented in Apache Spark, the methodologies aim to improve the value extraction process from large and heterogeneous datasets. Abdolkhani, Robab, et al., [1] focus on Remote patient monitoring using patient-generated health data (PGHD) from wearables, applying a sociotechnical health informatics approach to develop a data quality management (DQM) guideline.

Zhang et al. [25] delve into the necessity of ensuring high-quality RDF resources in the Semantic Web, introducing an automated approach for assessing the quality of RDF resources. Serra et al., [17] present a comprehensive contextual model specifically designed for data quality (DQ), filling the gap in context modeling within the DQ literature. Hutama et al., [8] investigate Business Intelligence (BI) maturity in an Indonesian automotive company, utilizing the Business Intelligence Maturity Model (BIMM). Siregar et al. [19] assessed the importance of data quality in the COVID-19 vaccination scheduling system in Jakarta, Indonesia, using Total Data Quality Management (TDQM). Kim et al., [10] propose an organizational process maturity model for ensuring data quality in the Internet of Things (IoT), aligning with ISO standards.

Bonyadi et al., [3] emphasize the significance of data in organizational decision-making and processes, presenting a comprehensive data quality management model for data governance derived through meta-synthesis. Sanchez and Reitmeier [15] address the challenges of digital transformation, highlighting the impact of poor data quality on digital operations. Sturm [21] explores the importance of monitoring histograms for data quality assurance in the LHCb experiment, implementing automatic checks during simulation test jobs. Bastie and Sandoz [2] extend the exploration of data quality to Best Source Selector (BSS) performances in flight test scenarios, showcasing results from various modes and highlighting the importance of Data Quality Metrics (DQM). Yaman et al., [24] introduce LinkedDataOps, an approach applying semantic web standards to end-to-end geospatial linked data production governance. Caballero et al., [4] introduce BR4DQ, a novel approach emphasizing the significance of grouping business rules at the outset for enhanced generalization, comparability, and reusability.

Hutama [8] explores the crucial role of data quality in artificial intelligence (AI) development within the AI-TIE project. Ehrlinger et al., [5] emphasize the often-neglected tasks of data integration, management, and quality assurance in data science projects, proposing an approach for successful project execution. AuditDQ, proposed by Souza et al., [20], introduces a framework designed for data quality evaluations of hospital data obtained from clinical coding processes. [11] focuses on the challenges faced by small and medium-sized enterprises (SMEs) in utilizing data for decision-making. The study proposes guidelines and technologies to assist South African SMEs in capturing, storing, and managing data. Fakhrzdaeh et al., [6] contribute to metadata quality with a case study of the registration system of the Iranian Research Institute for Information Science and Technology.

In summary, the literature review provides a comprehensive overview of various studies addressing data quality and its management in different domains, offering valuable insights and methodologies to the field.

3. Research Methodology

The foundation of our research methodology stems from a thoughtful analysis of diverse factors influencing the utilization of Data Quality Management (DQM), as illustrated in Figure 1.

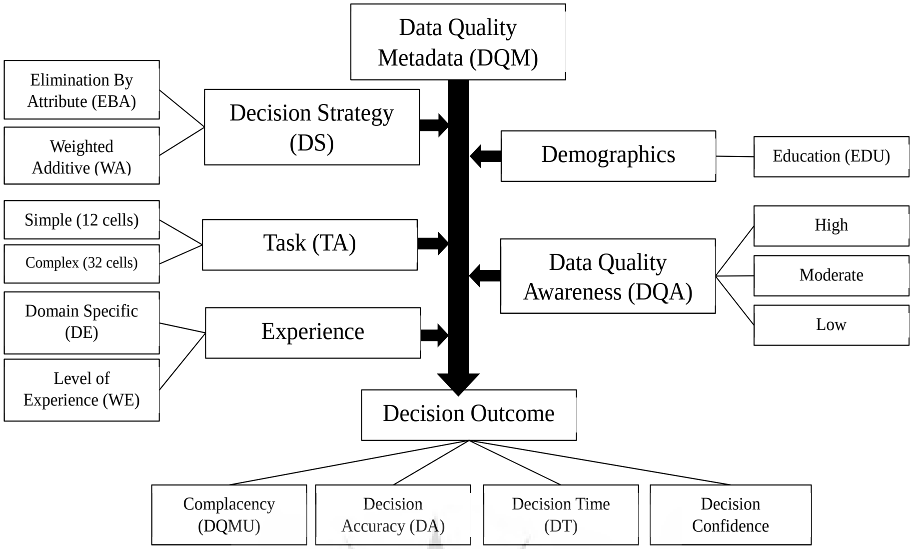

Figure 1. Unveiling the Research Stage: Exploring the Dynamic Influence of DQM on Decision Outcomes and Its Interplay with Varied Variables

3-1. Research objectives

Existing studies underscore the pivotal role of integrating Data Quality Management (DQM) with raw data to enable decision-makers to assess its suitability for the given task. However, before committing to the incorporation of DQM into databases, crucial queries must be addressed due to the associated expenses in its creation and maintenance. The initial query delves into whether DQM significantly enhances the efficacy (accuracy/quality) of decision outcomes. The second query pertains to the receptiveness of decision-makers to novel information, specifically in the context of DQM. While the latter inquiry is well-explored in DQ literature, consensus on results remains elusive. Remarkably, the former query has been inadequately examined, as far as our current understanding goes. Although both questions carry significance, the primary one holds greater weight, given that the influence of DQM on decision outcomes can swing either way—positively or negatively. To delve into the initial research query, we employed three metrics for decision outcomes: decision accuracy (effectiveness), decision confidence, and decision time (efficiency). Essentially, the impact of DQM on decision outcomes is gauged through these three dependent variables. Moreover, this research question delves into how DQM interacts with other pivotal factors, such as the experience level of decision-makers and decision strategy.

- ◇ H1: Decision-makers with lower educational backgrounds integrate Data Quality Management (DQM) into their decisions just as much as their highly educated counterparts.
  - ◆ H1a: Highly educated decision-makers more actively incorporate DQM into their decision-making processes than those with less education. As educational levels rise, the resistance to embracing new information diminishes.
- ◇ H2: Decision-makers with less experience integrate DQM into their decisions just as much as their more experienced counterparts.
  - ◆ H2a: Experienced decision-makers demonstrate a greater incorporation of DQM into their decisions compared to their less experienced counterparts. H2 and H2a address the general work experience levels of decision-makers. Those with experience tend to be more adaptable to new information than novices.
- ◇ H3: Decision-makers without domain-specific experience integrate DQM into their decisions just as much as those with specific domain experience.
  - ◆ H3a: Decision-makers lacking domain-specific experience show a higher integration of DQM into their decisions compared to those with specific domain experience. Specialization can bias decision-making by limiting objective information use.
- ◇ H4: Decision-makers with prior Data Quality (DQ) awareness incorporate DQM into their decisions just as much as those without prior awareness.
  - ◆ H4a: Decision-makers with DQ knowledge actively incorporate DQM into their decisions compared to those without DQ knowledge. Generally, decision-makers tend to rely on known variables, ignoring the unknown ones. Lack of prior DQ awareness makes decision-makers less likely to include DQM in their decisions as it represents an unknown variable. In essence, knowledge about DQ or DQ awareness prompts the use of DQM in decision-making.

◇ H5: Decision-makers using relatively simple decision-making strategies integrate DQM into their decisions just as much as those using more complex strategies.

- The complexities of decision-making can escalate with an abundance of data and choices, resulting in cognitive overload. To navigate this, decision makers can strategically embrace additional variables, such as Data Quality Management (DQM), sidestepping the perils of cognitive exhaustion. This phenomenon is rooted in two established literary concepts: information overload and cognitive capacity limits. Information overload occurs when decision makers struggle with more data than their cognitive abilities can process. Meanwhile, cognitive capacity limits manifest as a delicate interplay between working and long-term memory. Tackling complex decision scenarios demands a heightened cognitive capacity, potentially resulting in the omission of pertinent information. The experience of information overload and cognitive capacity constraints, however, depends on the idiosyncrasies of decision makers. Past research indicates that inexperienced individuals, when presented with the same amount of data, may experience information overload, while experienced decision-makers may not perceive it.
- H5a: it posits that decision makers employing a relatively simple decision strategy (DS) incorporate DQM more into their decisions compared to those opting for a more complex approach. As revealed earlier, when cognitive capacity is strained, decision makers instinctively simplify their processes by integrating new variables. The simplicity of a decision strategy, as per prior studies, can be contingent on the task type. For instance, the EBA strategy is deemed relatively complex, while WA is considered simple due to its compensatory nature. Consequently, it is anticipated that decision makers utilizing a WA strategy will involve DQM more in their decision processes than those employing an EBA strategy.

From H1 (H1a) to H5(H5a), the hypotheses are uniformly applicable in both simple and complex decision scenarios. Section 4, Results and Discussions, will delve into the outcomes in both environments.

◇ H6: it challenges the assumption that decision makers, irrespective of task complexity, incorporate DQM equally into their decisions.

- H6a: in contrast, conjectures that decision makers handling simpler tasks integrate DQM more extensively than their counterparts grappling with complex tasks.

3-2. Experimental setting

3-2-1. Pilot study

In this investigation, we embarked on a quest to identify the most fitting Data Quality Metric (DQM) representation. Our preliminary exploration featured a dynamic pilot experiment introducing two distinct DQM formats: the interval representation, featuring lower and upper value limits, and the intriguing probability representation. The former showcased the spectrum of potential values for specific data, while the latter conveyed the likelihood that a given data item accurately represented reality. Three cohorts were meticulously assembled for this experimental odyssey—one immersed in interval DQM, another in probability DQM, and a control group devoid of DQM enchantment. With experimental rigor in mind, decision strategies were confined to the realms of additive and EBA, and the participants hailed from the illustrious realm of Ph.D. scholars in Applied Economics. The Ph.D. scholars, a cohort of ten for each group, were whimsically assigned to their DQM destinies. A statistical dance of $\chi 2$ assured us that the use of DQM was comparably indifferent between the interval and probability DQM format factions, maintaining the sanctity of our findings at the 95% confidence level. To delve deeper into the psyche of our scholarly participants, we orchestrated exit interviews with both DQM groups, unraveling the intricacies of their decision-making processes. The revelations danced around (1) the incorporation of DQM in their decisions, (2) their grasp of DQM's essence, and (3) their DQM format preferences. Unsurprisingly, the interval DQM wielders struggled to harmonize its presence in their decisions, a conundrum reflected in their slightly protracted decision-making timelines. On the flip side, the probability DQM wielders exhibited a synchronized understanding of its significance, employing it uniformly within the hallowed halls of statistical confidence. Given these enlightening insights, we steered our research toward the implementation of the probability DQM format for the grand finale of our study.

Drawing inspiration from the navigational wisdom of literature review, we fine-tuned our approach, labeling tags with the term "accuracy" and elaborating on DQM nuances through vivid examples in the experiment's instructional tapestry.

Similarly, a parallel odyssey unfolded in our exploration, involving a pilot experiment featuring a cadre of ten Ph.D. scholars. This preliminary escapade sought to unravel the tapestry of clarity and comprehension woven by our experiment. Moreover, it unveiled the trifecta of decision strategies—Elimination by Attribute (EBA), Weighted Additive (WA), and Conjunctive (CON)—which our subjects wielded to navigate the labyrinth of decision tasks. The compass guiding our final experiment remained true, harmonizing the decision solutions across the trio of strategies, ensuring a fair evaluation of each subject's decision accuracy.

Choosing Ph.D. scholars in Applied Economics for the pilot study was strategic, leveraging their advanced academic training and practical expertise. This cohort's familiarity with intricate decision landscapes and data-driven analyses aligns with our study's objectives. Their nuanced understanding of complex decision tasks allows us to gather insightful feedback on the effectiveness of different Data Quality Metric (DQM) formats. The scholars' experience in decision strategies and statistical methodologies adds depth to the pilot study, contributing to the robustness and relevance of our findings. Importantly, their academic discipline emphasizes the practical implications of DQM in decision tasks, making their insights particularly valuable for our research. In essence, the selection of Ph.D. scholars in Applied Economics ensures that the pilot study aligns closely with real-world applications of DQM in decision-making scenarios, enhancing the quality and applicability of our research outcomes.

3-2-2. Final Task

The inception of Data Quality Management (DQM) experiments has sparked curiosity regarding the impact of the application domain on DQM utilization. The rationale behind this idea is that participants may show less apprehension when decisions are based on inadequate data, assuming that there will be minimal consequences on outcomes within a particular domain. Consequently, existing research advocates exploring the efficacy of Data Quality Management (DQM) in diverse decision-making landscapes, especially in pivotal scenarios. In light of this, we innovated a distinct decision-making arena: a bankruptcy prediction task inspired by the Altman-Z model. Our assessment covers all four criteria for evaluating a firm's financial well-being.

$$\left(\frac{Retained\ earnings}{total\ assets}, \frac{Market\ value\ equity}{Book\ value\ of\ total\ liabilities}, \frac{\text{Earnings before interest \& taxes}}{Total\ assets}\ \text{and}\ \frac{\text{working capital}}{Total\ assets}\right)$$

The Altman-Z model served as our compass in assigning relative importance to each criterion, shaping the decision task into a familiar landscape based on past studies. Picture the decision-making process as a journey, where firms are ranked in terms of financial health, navigating from the summit of excellence to the valleys of fiscal challenge. We enriched this journey with insights from a literature review, a treasure trove of experimental wisdom. In the realm of a pilot study, we designed an experiment with three decision mechanisms (EBA, WA, and CON), aiming to achieve a unified ranking result (refer to Table 4). Surprisingly, the final experiment's voyagers, our subjects, charted their course using only two decision strategies: EBA and WA. Behold Figure 1, providing an exclusive glimpse into their strategic seas. Tasks unfolded in two forms—simple and complex. The participants were asked to rank four banks based on three initial criteria, resulting in a concise grid of 12 cells. The latter, a grander quest, entailed ranking eight banks across four criteria, covering an expansive terrain of 32 cells. Both adventures splintered into two trajectories: DQM upfront or DQM later in the journey. Subjects, oblivious to DQM upfront, encountered it later, prompting a pivotal question: Would their decisions be influenced by this newfound wisdom? A strategic move, not just to uncover complacency but to increase our sample size. Our experimental tetrad included four conditions: simple task with immediate DQM, simple task with delayed DQM, complex task with immediate DQM, and complex task with delayed DQM. These conditions were presented to the subjects randomly, revealing their mysteries. In the realm of statistical scrutiny, a Friedman test at $\alpha = 5\%$ confirmed our findings, revealing no significant differences in DQM usage between those informed upfront and those enlightened later. The experiment, like an intricately woven tapestry, featured a comprehensive guide that demystified each attribute and illuminated subject expectations. Within the cocoon of a controlled environment, our experiment unfolded with an exit survey—a 28-question compass—navigating through subjects' demographics post-expedition. Lastly, our subjects recorded the time coordinates of their journey's start and end—a record of their exploration.

3-2-3. Participants

In this experimental study, a diverse cohort of 106 individuals actively participated, encompassing 80 proficient students specializing in business information systems and an additional 26 individuals with varied backgrounds. The selection process aimed to create a well-rounded participant pool, ensuring representation from both the academic and professional realms.

The intriguing diversity of our participants unfolded as they immersed themselves in a multifaceted decision task. Sixty participants navigated the complexities with Data Quality Metadata (DQM) from the outset, while 46 embraced the challenge with DQM introduced later in the process.

The experiment's intricacies deepened with 42 individuals engaging in the labyrinth of a complex decision task, while 64 navigated the waters of a comparatively simpler task. This strategic variation in task complexity allowed us to capture a spectrum of decision-making scenarios and responses.

The human tapestry of our participants revealed that, among the 106 individuals, 30 brought valuable work experience to the experiment, enriching its dynamic with practical insights. Additionally, 35 participants boasted domain experience, injecting a specialized essence into the proceedings, further enhancing the study's ecological validity.

A notable facet emerged as 77 out of the 106 participants displayed a commendable medium or high level of prior data quality awareness, indicating a depth of understanding in the realm of data quality. In contrast, 29 participants entered the experiment with a clean slate, devoid of any prior data quality awareness, introducing an element of unpredictability and ensuring a diverse perspective in the study.

The participant selection aimed for a balanced representation across various dimensions, such as educational background, work experience, and domain expertise, contributing to the external validity of our findings. The criteria for participant selection prioritized diversity to capture a comprehensive understanding of the impact of DQM on decision-making efficacy across different contexts and perspectives.

3-2-4. Variables

Table 5 comprehensively encapsulates both the independent and dependent variables, presenting vivid descriptions and corresponding acronyms for each, providing a succinct and informative overview.

3-3. Statistical Analysis

To gauge the meaningfulness of our findings, we subject them to a battery of statistical examinations, aligning with established literature. Every unique test undergoes scrutiny at a 5% significance level, unless explicitly specified.

3-3-1. Chi-square

The χ2 test serves as a detective of statistical tales, delving into the mysterious realm of hypotheses. It scrutinizes the harmony between the observed events' frequency distribution in a sample and the anticipated frequencies, either derived from a theoretical distribution or gleaned from the control groups within the observed samples. Beyond this analytical prowess, the χ2 analysis ventures into the intrigue of paired observations, unveiling whether two variables are independent. Picture this: the education level of subjects waltzing in tandem with their DQM usage. Our narrative, however, takes a unique twist, utilizing the χ2 test to probe the complacency levels of decision outcomes across diverse groups. It's a quest to unravel the connection between subject characteristics and the utilization of DQM for decision-making purposes.

3-3-2. Regression trees - Leave-one-out-cross validation

Regression trees

Tree-building algorithms craft logical environments to predict or classify cases with varying degrees of accuracy. Unlike linear and parametric methods, regression trees operate non-linearly, predicting continuous dependent variables based on one or more continuous or categorical independent variables. They sidestep assumptions integral to tests like Analysis of Variance (ANOVA) and t-tests, which hinge on data being normally distributed and independent and identically distributed (iid). Interpreting tree results is typically straightforward, and they don't presuppose any specific relationship—linear, non-linear, or monotonic—between predictors and dependent variables.

For instance, decision accuracy might inversely correlate with DQM use, yet the relationship could become positive if subjects possess a high data quality awareness or experience level. Linear regression trees prove valuable when little or no prior knowledge exists about the connection between dependent and independent variables. This makes them ideal for analyzing experimental data where assumptions about three dependent variables (decision accuracy, decision confidence, and decision time) and their predictors are absent.

Table 5. Summaries of the different variables in the experiment

Types of metadata	Description
Decision Accuracy (DA)	It involves a discreet variable gauged through the ranking of firms in tasks both intricate and straightforward. Perfect accuracy stands at 10/10 when the ranking is flawlessly executed. A single ranking misstep results in a precision dip of 1.25 or 2.5 marks for complex and simple tasks, respectively. Real-world scenarios are riddled with uncertainties like low-quality data, making it challenging to predetermine optimal decisions. In our experiment, we aimed for a relatively correct ranking by factoring in Decision Quality (DQ) levels through three widely-used decision-making strategies. Hence, each subject's decision accuracy is appraised using a comparable decision solution. In this experimental context, 100% decision accuracy means achieving the predefined correct ranking without any errors, be it calculation glitches or inconsistencies, using one of the three decision strategies. Essentially, accuracy quantifies the decision maker's errors in navigating the decision task.
Decision Confidence (CONF)	The variable in question is gauged using a 5-point Likert scale, spanning from -2 to 2. Ratings of -2 and -1 signify minimal or moderate assurance regarding the decision's outcome, respectively. A score of 0 indicates a neutral stance toward the decision task. Conclusively, ratings of 1 and 2 reflect substantial and exceptionally high confidence in the decision outcomes.
Decision Time (DT)	It represents a dynamic continuum gauged by calculating the duration between the experiment's commencement and conclusion. The metric is quantified in minutes.
Data Quality Metadata Used (DQMU)	It's a binary variable labeled either "yes" or "no," representing a categorical aspect. The evaluation hinges on responses to two distinct queries. Firstly, participants detail their problem-solving formula, triggering a positive label for the DQMU variable if their method incorporates DQM as a decision variable and aligns with the DQM solution category. Simultaneously, their feedback on crucial decision task variables is scrutinized for coherence with their solutions. Almost universally, respondents exhibit consistency in their answers to both questions. This reinforces the accurate identification of DQM utilization.

Types of metadata	Description
Data Quality Metadata (DQM)	It represents a categorical aspect, marked either as a "yes" or "no." Its determination is contingent upon the nature of the experiment. When DQM is incorporated at the outset of the experiment, the variable boasts a "yes" designation; conversely, in scenarios lacking DQM integration, it bears a "no" classification.
Decision Strategy (DS)	The decision-making landscape is bifurcated into two intriguing realms: Weighted Additive (WA) and Elimination by Attribute (EBA). Participants are tasked with unraveling the mystery of their decision strategy when tackling the decision challenge. Through their narratives, the decision strategy undergoes a fascinating classification, ultimately falling into the distinguished categories of either EBA or WA.
Task Clear (TAclear)	It represents a categorical variable gauged through participants' feedback on the clarity of the experiment. Respondents answer a pivotal question regarding the experiment's lucidity, choosing between a definitive "yes" or a categorical "no."
Task Type (TA)	It represents a categorical variable adorned with the labels "Complex" or "Simple," delineated by the nature of the experiment.
Data Quality Awareness (DQA)	The variable falls into the realm of categoricals, assessed by subjects' responses to six fundamental questions about data quality in our exit survey. Full data quality awareness emerges when a subject accurately answers pivotal questions B, C, D, and E. A subject who successfully answers two or three of these crucial questions enters the realm of moderate data quality awareness. Meanwhile, those who struggle with just one question or fail to answer any find themselves in the abyss of lacking data quality awareness.
Work Experience (WE)	It represents a categorical aspect disclosing the subjects' proficiency in work experience. This classification is derived from their responses to the exit survey question, "How many years of professional experience do you possess?"
Domain Experience (DE)	This categorical variable assesses participants' familiarity with the experiment based on past experiences. It depends on their responses to the question, "How frequently have you dealt with similar tasks in the past?" A response of zero designates a lack of domain experience, while an answer of 1 or more signifies possession of domain expertise.
Education (EDU)	It represents a categorical variable denoting the attained level of education by an individual.

a Refer to Table 4 for details.
b How would you characterize accuracy within the realm of data quality?
c Could you provide an instance of data inaccuracy?
d Kindly elaborate on the concept of data quality dimensions or attributes.
e Can you highlight a few examples of data quality dimensions or attributes?

Leave-one-out cross-validation

Cross-validation serves as the litmus test for a regression or classification model, assessing its performance on new data. Essential in predictive analysis, this method involves forecasting a model's performance in real-world scenarios. Unlike the traditional split of data into training and test sets, cross-validation takes a more dynamic approach. The original dataset splinters into various subsets, each serving as a unique testing ground after training. This strategic maneuver minimizes overfitting, providing a glimpse into how the model will perform against real-world, independent data. Particularly handy when a limited sample size hinders the conventional division into training, validation, and test sets. Enter leave-one-out cross-validation, a notable variant where each observation takes its turn in the validation spotlight, while the rest constitute the training background. In our study, we utilized the predictive power of a regression tree, combined with leave-one-out cross-validation. Our mission is to forecast the values of three dependent variables—decision accuracy (DA), decision confidence (DC), and decision time (DT)—using a variety of independent variables available to us.

3-3-3. Stepwise regression

A progressive linear regression approach was employed to identify influential variables and their interactive impacts on three dependent variables: Decision Accuracy (DA), Confidence (CONF), and Time (DT). Interestingly, the outcomes from leave-one-out cross-validation mirrored those of the stepwise regression, albeit with subtle nuances. Notably, the leave-one-out cross-validation results demonstrated superior performance in comparison to stepwise regressions, especially evident when evaluating mean squared error (MSE). As a consequence, this paper exclusively showcases the findings derived from leave-one-out cross-validation tests.

4. Results and discussions

4-1. The use of DQM in decision making processes

4-1-1. Education, Experience, Data Quality Awareness and Decision Strategy - Simple decision task

As illustrated in Table 3, complacency gauges the extent to which Decision

Quality Metrics (DQM) influences decision-making processes. To delve into the complacency levels of diverse decision makers regarding DQM, a χ2 test is performed on the DQMU variable (refer to Table 5).

Table 6 showcases the χ2 test results, revealing that, for straightforward decision tasks, the complacency levels regarding DQM do not significantly differ between decision makers with varying educational backgrounds at a 95% confidence level. Likewise, the complacency levels remain comparable between decision makers with or without work experience, and those with or without domain experience. Interestingly, no discernible link exists between complacency levels and the implemented decision strategies.

Consequently, H1-H3 and H5 find acceptance, while H1a-H3a and H5a face rejection at the 95% confidence level for the uncomplicated decision task. Notably, a noteworthy correlation emerges between decision makers' complacency levels and their Decision Quality Awareness (DQA) at the 95% confidence level. In simpler terms, individuals with a high DQA incorporate DQM more seamlessly into their decision-making processes compared to those with minimal or no DQA.

Among 24 subjects lacking substantial DQA, 17 refrained from integrating DQM into their decision processes. In contrast, out of 40 subjects boasting high DQA, merely 15 neglected to incorporate DQM in solving the decision task. Consequently, H4 is dismissed in favor of H4a at the 95% confidence level.

Table 6. Exploring Subjects' Contentment Across Varied Groups in Light of Data Quality Metadata (DQM) and Simple Decision Tasks

Simple Task				
Variables		**DQMU**	**Obs.**	**Complacency**
EDU (H1)	Under graduates	Yes	20	$\chi2 = 0.2591$ $p = 0.6107$
		No	18	
	Post graduates	Yes	12	
		No	14	
WE (H2)	No experience	Yes	23	$\chi2 = 1.6967$ $p = 0.1927$
		No	18	
	With experience	Yes	9	
		No	14	

Simple Task				
Variables		**DQMU**	**Obs.**	**Complacency**
DE (H3)	Without DE	Yes	22	$\chi2 = 0.2771$ $p = 0.5986$
		No	20	
	With DE	Yes	10	
		No	12	
DQA (H4)	Without DQA	Yes	7	$\chi2 = 6.6667$ $p = 0.00098$**
		No	17	
	With DQA	Yes	25	
		No	15	
DS (H5)	WA	Yes	18	$\chi2 = 1.0667$ $p = 0.3017$
		No	22	
	EBA	Yes	14	
		No	10	

**= $p < 0.05$

4-1-2. Education, Experience, Data Quality Awareness and Decision strategy- Complex decision task

The findings from the χ2 test, as illustrated in Table 7, divulge intriguing insights into the complacency dynamics surrounding the complex decision task. Astonishingly, the confidence level of decision makers, irrespective of their educational background, exhibits no significant divergence at the 95% threshold. Similarly, the chasm between the complacency levels of decision makers with varying work experiences is not statistically discernible. Remarkably, the choice of decision strategies seems to bear no weight on complacency levels. Thus, we embrace H1, H2, and H5 while discarding H1a, H2a, and H5a for the complex decision task.

Yet, a riveting twist emerges when exploring the complacency dichotomy between subjects endowed with Domain Experience (DE) and those devoid of it. Strikingly, subjects with prior DE exhibit a significantly higher complacency level towards Decision Quality Measures (DQM) in tackling decision tasks. To elaborate, of the 13 subjects with prior DE, only 3 seamlessly integrated DQM, whereas among the 29 novices, a whopping 23 seamlessly incorporated DQM into their decision-making processes.

Equally captivating is the revelation of a robust relationship between decision makers' complacency towards DQM and their Decision Quality Aptitude (DQA) level. The narrative unfolds as subjects boasting higher DQA are more inclined to seamlessly integrate DQM, mirroring the patterns observed in the simple decision task. Among the meager 5 subjects with scant DQA, a staggering 4 refrained from incorporating DQM, while among the 37 subjects flaunting a high DQA, merely 12 abstained from integrating DQM.

In essence, the tenets of H3 and H4 are relinquished in favor of H3a and H4a, steadfast at the 95% confidence level.

Summarizing the revelations encapsulated in Tables 6 and 7, a compelling narrative unfolds for both simple and complex decision tasks. The influence of Decision Quality Aptitude (DQA) on decision makers' complacency towards Decision Quality Measures (DQM) is stark. A higher DQA level seamlessly corresponds to a heightened integration of DQM into decision-making processes.

Table 7. Exploring Subjects' Satisfaction with Decision Outcomes Amidst Complex Tasks and Data Quality Metadata (DQM) Influence.

Complex Task				
Variables		**DQMU**	**Obs.**	**Complacency**
EDU (H1)	Under graduates	Yes	21	$\chi2 = 0.1958$ $p = 0.6581$
		No	12	
	Post graduates	Yes	5	
		No	4	
WE (H2)	No experience	Yes	22	$\chi2 = 0.0808$ $p = 0.7763$
		No	13	
	With experience	Yes	4	
		No	3	
DE (H3)	Without DE	Yes	23	$\chi2 = 12.0361$ $p = 0.0005$**
		No	6	
	With DE	Yes	3	
		No	10	

Complex Task				
Variables		**DQMU**	**Obs.**	**Complacency**
DQA (H4)	Without DQA	Yes	1	$\chi2 = 4.2262$ $p = 0.0398$**
		No	4	
	With DQA	Yes	25	
		No	12	
DS (H5)	WA	Yes	19	$\chi2 = 1.2620$ $p = 0.2613$
		No	9	
	EBA	Yes	7	
		No	7	

**= $p < 0.05$

4-1-3. Task type

Table 8's χ2 test outcomes reveal a surprising revelation: at a 95% confidence level, there's no notable difference in the complacency levels among decision makers tackling simple and complex tasks. Contrary to the widely accepted information overload theory predicting heightened complacency in complex decisions, our results defy expectations. Consequently, H6 stands validated, while H6a faces rejection. This outcome finds its rationale in the shared underlying problem for both decision types—whether simple or complex—where task intricacy solely hinges on the number of alternatives available.

Table 8. Exploring Subject Satisfaction in Decision-Making: The Impact of Data Quality Metadata (DQM) Combined with Decision Task Complexity on Perceived Comfort with Choices.

Complacency for the decision task when DQM is provided				
Variables		**DQMU**	**Obs.**	**Complacency**
TA (H6)	Simple	Yes	32	$\chi2 = 1.4505$ $p = 0.2285$
		No	32	
	Complex	Yes	26	
		No	16	

4-2. Data quality metadata and its impact on decision outcomes

4-2-1. Decision accuracy

The intricate regression tree depicted in Figure 2 unfolds a compelling narrative pertaining to the predictability of the decision accuracy (DA) variable. This analysis indicates that independent variables, namely Decision Quality Management (DQM), Decision Quality Assessment (DQA), Decision Strategy (DS), and Task Dynamics (TA), collectively contribute to forecasting DA with notable precision, encapsulated in a mean squared error of 0.5715. Conceptualizing DA on a scale from 0 to 10, where elevated peaks denote excellence and troughs indicate potential improvements, this study illuminates the orchestration of a symphony of variables in predicting DA outcomes. Exploring decision makers endowed with a robust DQA reveals a notable association with heightened decision accuracy. A steadfast DQA emerges as a key determinant in unlocking optimal decision outcomes. Conversely, for decision architects lacking in DQA, the fate of decision accuracy becomes contingent on the interplay of DQM, DS, and TA variables. In instances where DQA is absent, the integration of DQM into decision-making processes for architects leads to a precipitous decline in decision accuracy. In contrast, decision makers with moderate DQA levels, adeptly incorporating DQM and embracing a weighted additive decision strategy, ascend to the pinnacle of decision accuracy. In summary, the essence of decision accuracy hinges delicately on DQA levels. The narrative extends as the influence of DQA weaves through decision architects, where high DQA correlates with increased utilization of DQM, as evidenced in Tables 6 and 7, culminating in elevated decision accuracy. The findings reveal a story of high DQA decision-makers transcending the intricacies of decision strategy and task complexity. This suggests that high DQA may cultivate a harmonious blend of high decision accuracy, possibly stemming from the enriched foundation of supplementary DQA knowledge acquired through further educational endeavors.

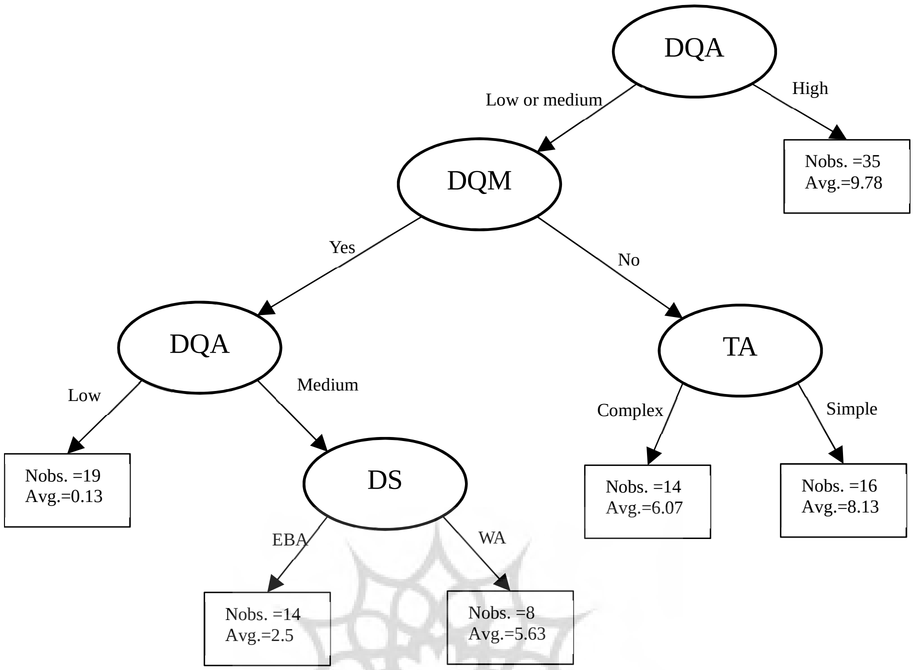

Figure 2. A regression tree for the decision accuracy (DA) with MSE=0.5715. The minimum score is 0 and the maximum is 10

4-2-2. Decision confidence

Figure 3 presents a detailed analysis of decision confidence prediction through a complex regression tree, providing insights into the various influences on this crucial metric. Within this arboreal framework, Decision Quality Management (DQM), Decision Quality Assessment (DQA), Task Clarity (TAclear), Education Level (EDU), and Decision Strategy (DS) emerge as influential factors, collectively contributing to decision confidence with a remarkable Mean Squared Error (MSE) of 0.089.

The regression tree reveals a compelling story where the lack of Data Quality Management (DQM) in decision-making, combined with unclear task objectives, leads to a significant decrease in decision confidence. When a clear decision task is presented without Data Quality Management (DQM), the focus shifts to the importance of Data Quality Assurance (DQA). A high level of DQA reduces uncertainty, enabling more confident decision-making.

For decision makers with elevated Data Quality Awareness (DQA), the incorporation of Data Quality Management (DQM) becomes instrumental in achieving high decision confidence. Conversely, individuals with lower Decision Quality Assessment (DQA) scores embark on a unique journey where their confidence in making decisions is closely tied to their level of education and the subtle implementation of decision-making strategies. Notably, a combination of a high level of education and an Evidence-Based Approach (EBA) decision strategy harmoniously culminates in increased decision confidence.

In a climactic resolution, the convergence of DQM integration, a robust DQA, clear task definition, advanced education, and a complex decision strategy emerges as the focal point in the narrative of decision confidence. Deciphering this intricate dance of variables reveals a profound truth: decision makers equipped with prior data knowledge, scholarly acumen, and a clear vision experience elevated decision confidence when intertwining Data Quality Management (DQM), education, and the sophisticated nuances of an Evidence-Based Approach. This study contributes a masterpiece to the exploration of the complex interplay of variables influencing decision confidence, inviting comparisons with other studies in the field.

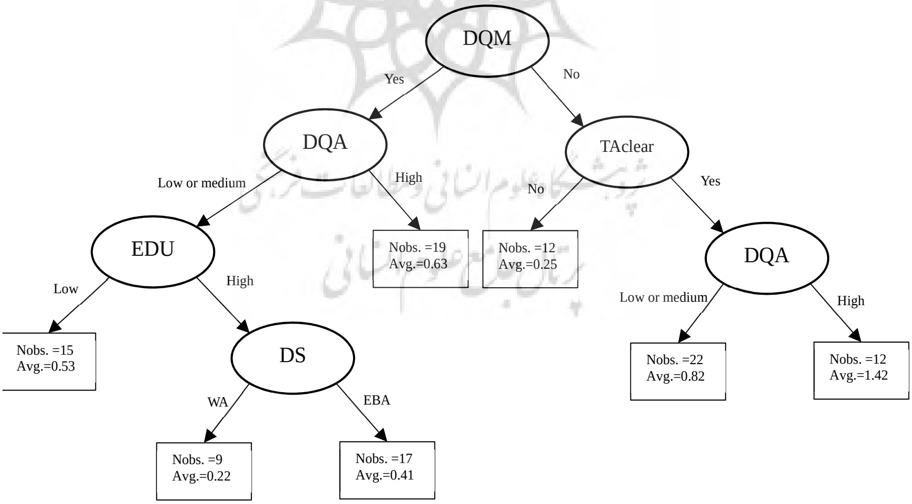

Figure 3. A regression tree for the confidence level of decision makers on their decision outcomes with MSE=0.089

4-2-3. Decision time

The illuminating representation in Figure 4 elucidates the intricacies surrounding the prediction of decision times, weaving a narrative that revolves around various independent variables. Experienced decision-makers, fortified by extensive work experience, tread a deliberate path, investing more time in deciphering tasks compared to their less-experienced counterparts. However, a notable deviation awaits novices without work history; their decision times pivot on the seamless integration of Decision Quality Management (DQM) into their cognitive processes.

For novices eschewing DQM, a swift decision tempo prevails. Yet, for those enlightened individuals who incorporate DQM into their decision-making framework, a nuanced dance unfolds. Decision times intricately correlate with the educational and domain expertise of these neophytes. A robust educational background and domain acumen extend decision times, carving a distinctive rhythm. In essence, a high Decision Quality Assessment (DQA), coupled with a rich educational pedigree and seasoned expertise, orchestrates an extended decision chronicle.

Accordingly, decision experts endowed with a lofty DQA and scholastic grounding tap into reservoirs of prior wisdom, transcending the confines of given information. This expansive approach, compared to the novice's reliance on provided data, elongates the decisional symphony. Similarly, those proficient at intertwining DQM into their decisional fabric revel in a prolonged deliberative ballet, surpassing their counterparts who abstain from this integration. This nuanced understanding contributes to the academic discourse on decision-making processes, underscoring the multifaceted dynamics influenced by experience, education, and the strategic integration of DQM.

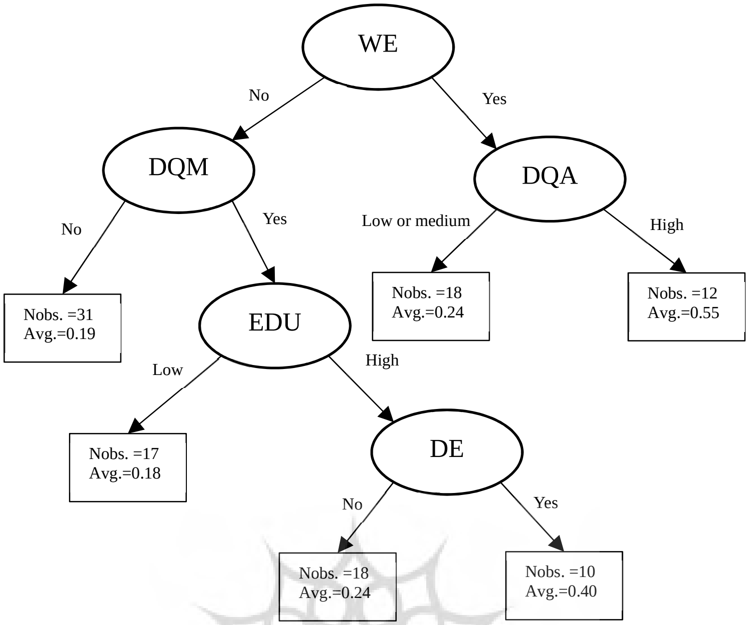

Figure 4. A regression tree for the decision time measured in minutes with MSE=0.0118

5. Conclusion

In conclusion, our thorough investigation of Decision Quality Management (DQM) has revealed significant insights with profound implications for both researchers and industry practitioners. The study not only addresses the conflicting findings from prior research but also presents innovative solutions for optimizing decision-making efficacy through the strategic integration of Data Quality Metadata (DQM).

Our approach departed from conventional methodologies by introducing a multifaceted exploration enriched with novel elements to enhance the comprehensibility and inclusivity of DQM evaluation. Through a unique presentation of Decision Quality Management (DQM) and the introduction of Decision Quality Assessment (DQA) as a crucial variable, we aimed to cultivate a consistent understanding among participants. This inclusive methodology allowed for a nuanced evaluation of DQM's impact on decision outcomes.

A pivotal contribution of our research lies in the redefinition and analysis of decision outcome measures. Instead of restricting complacency assessment to traditional metrics, we evaluated it by analyzing participants' explicit articulation of decision strategies, categorization of decision solutions, and responses to a crucial question: "which variables were pivotal in the decision-making processes." The indicators consistently aligned for each subject, providing a robust foundation for evaluating decision outcomes.

Furthermore, our study utilized a state-of-the-art tree-based algorithm to analyze the complex interaction of DQM and its effects with other independent variables on decision accuracy, time, and confidence. These findings offer fresh perspectives on DQM's profound impact, shedding light on the nuanced dynamics influencing crucial decision metrics. Accordingly, the responses to the research questions are as follows:

- A1: Decision outcomes in the context of bankruptcy prediction using the Altman-Z model are positively influenced by the incorporation of Data Quality Metadata (DQM);
- A2: Decision-makers with an elevated awareness of data quality demonstrate enhanced utilization of DQM, leading to improved decision accuracy;
- A3: Factors such as a high level of data quality knowledge facilitate the utilization of DQM, while concerns about cognitive overload and efficiency issues may impede its adoption;
- A4: Decision-makers experience a trade-off between decision accuracy and efficiency when employing Data Quality Metadata (DQM).

However, it is crucial to acknowledge the complexity of decision-making and the contextual nature of data quality. Our study identified certain limitations, such as the decline in DQM usage with increasing domain experience. This emphasizes the need for targeted interventions to encourage seasoned users to embrace DQM. Additionally, the impact of education level, work experience, and decision strategy on DQM usage remains inconclusive, suggesting potential avenues for future exploration.

For other researchers and industry professionals, our findings advocate for a thoughtful assessment of DQM benefits in unique contexts, recognizing the task-dependent nature of data quality. We recommend targeted training for decision-makers before deploying DQM in data warehouses. The symbiotic relationship between high Data Quality Awareness (DQA), DQM use, and decision outcomes underscores the intricate nature of these dynamics, offering actionable insights for organizational initiatives. This study lays a foundation for future research to delve deeper into the evolving landscape of data quality management, optimizing decision processes across diverse organizational settings.

5-1. Limitations

Despite the innovative approach, certain limitations should be acknowledged. DQM usage dwindled with increasing domain experience, suggesting the need for targeted initiatives to encourage seasoned users to embrace DQM. The impact of education level, work experience, and decision strategy on DQM usage was inconclusive, highlighting areas for further exploration. Additionally, the study focused on specific decision tasks, and generalizing findings to diverse contexts requires caution.

5-2. Future Works

Building on these insights, future research avenues are abundant. Further exploration into the nuanced relationship between decision maker characteristics, chosen strategies, and task dynamics is crucial for a comprehensive understanding of DQM utility. Investigating the influence of educational interventions and specialized training on DQM adoption, particularly among experienced decision makers, can provide actionable strategies for organizations. Additionally, expanding the study to different industries and decision contexts will enhance the generalizability of findings. Continuous advancements in DQM tools and techniques necessitate ongoing research to stay abreast of evolving practices and their impact on decision outcomes. Overall, the evolving landscape of data quality management offers a rich terrain for future investigations, aiming to optimize decision processes in diverse organizational settings.

References

Abdolkhani, R., Borda, A., & Gray, K. (2018). Quality Management of Patient Generated Health Data in Remote Patient Monitoring Using Medical Wearables–A Systematic Review. *Connecting the System to Enhance the Practitioner and Consumer Experience in Healthcare*, 1-7; https://doi.org/10.3233/978-1-61499-890-7-1.

Bastie, P., & Sandoz, F. (2022). Understanding Best Source Selector Performances and Results from Two Test Scenarios. *International Foundation for Telemetering*, *ISSN*: 1546-2188, 0884-5123, 0074-9079.

Bonyadi, S., Hariri, N., Taheri, S. M., & Poornaghi, R. (2023). Provide data quality management model for data governance using meta synthesis. *Iranian Journal of Information Processing and Management*, 38 (4), 1533-1564, https://doi.org/10.22034/jipm.2023.698597.

Caballero, I., Gualo, F., Rodríguez, M., & Piattini, M. (2022). BR4DQ: a methodology for grouping business rules for data quality evaluation. *Information Systems*, 109, 102058, https://doi.org/10.1016/j.is.2022.102058.

Ehrlinger, L., Lettner, C., Fragner, W., Gsellmann, G., Nestelberger, S., Rauchenzauner, F., ... & Zeindl, J. (2022, August). Data Integration, Management, and Quality: From Basic Research to Industrial Application. *In International Conference on Database and Expert Systems Applications (pp. 167-178). Cham: Springer International Publishing*, https://doi.org/10.1007/978-3-031-14343-4_16.

Fakhrzdaeh, A., Ershadi, M. J., & Ershadi, M. M. (2023). Data mining methods for quality control of research data; Case study of Iranian Scientific Database (GANJ). *Iranian Journal of Information Processing and Management*, 38 (3), 927-944, https://doi.org/10.22034/jipm.2023.698614.

Hamlin, B. (2022). *A Consensus-based Data Quality Assessment Model for Patient Reported Outcome Information in Digital Quality Measurement Programs* (Doctoral dissertation, Johns Hopkins University), http://jhir.library.jhu.edu/handle/1774.2/67111.

Hutama, A. T., Hidayanto, A. N., & Suryono, R. R. (2023). Business intelligence maturity assessment: a case study on an indonesian automotive company. *Jurnal Teknoinfo*, 17 (2), 379-389, https://doi.org/10.33365/jti.v17i2.2443.

Khaleghian, H., & Shan, Y. (2023). Developing a Data Quality Evaluation Framework for Sewer Inspection Data. *Water*, 15 (11), 2043, https://doi.org/10.3390/w15112043.

Kim, S., Pérez-Castillo, R., Caballero, I., & Lee, D. (2022). Organizational process maturity model for IoT data quality management. *Journal of Industrial Information Integration*, 26, 100256, https://doi.org/10.1016/j.jii.2021.100256.

Mabotja, P. P. (2022). *Improving data quality and management in one South African SME for data analytics* (Doctoral dissertation, North-West University (South Africa)), http://hdl.handle.net/10394/39226.

Mäkitalo, K. (2023). *Using custom data quality maturity tool for improving data quality in an organization*, https://urn.fi/URN:NBN:fi:amk-2023120133508.

Ngueilbaye, A., Huang, J. Z., Khan, M., & Wang, H. (2023). Data quality model for assessing public COVID-19 big datasets. *The Journal of Supercomputing*, 79 (17), 19574-19606, https://doi.org/10.1007/s11227-023-05410-0.

Pratiktio, R. P., Kusumasari, T. F., & Fauzi, R. (2023, February). Design guidelines and process of reference data quality management based on data management body of knowledge*. In AIP Conference Proceedings* (Vol. 2654, No. 1). AIP Publishing, https://doi.org/10.1063/5.0114293.

Sanchez, G., & Reitmeier, J. (2022, April). Multivariate Data Quality Relationships of Geothermal Facilities for Increased Efficiency in Digital Operations. *In Offshore Technology Conference* (p. D041S056R007). OTC, ISBN: 978-1-61399-852-6.

Serra, F., Peralta, V., Marotta, A., & Marcel, P. (2022). Use of context in data quality management: a systematic literature review. *ACM Journal of Data and Information Quality*, https://doi.org/10.1145/3672082.

Serra, F., Peralta, V., Marotta, A., & Marcel, P. (2022, October). Modeling context for data quality management. *In International Conference on Conceptual Modeling (pp. 325-335). Cham: Springer International Publishing*, https://doi.org/10.1007/978-3-031-17995-2_23.

Shankaranarayanan, G., & Zhu, B. (2021). Enhancing decision-making with data quality metadata. Journal of Systems and Information Technology, 23 (2), 199-217, https://doi.org/10.1108/JSIT-08-2020-0153.

Siregar, D. Y., Akbar, H., Pranidhana, I. B. P. A., Hidayanto, A. N., & Ruldeviyani, Y. (2022, January). The importance of data quality to reinforce COVID-19 vaccination scheduling system: A study case of Jakarta, Indonesia. *In 2022 2nd International Conference on Information Technology and Education (ICIT&E)* (pp. 262-268). IEEE, https://doi.org/10.1109/ICITE54466.2022.9759880.

Souza, J., Caballero, I., Lopes, F., Vasco Santos, J., Gualo, F., Merino, J., & Freitas, A. *Auditdq: a Framework Based on International Standards to Enhance the Evaluation of the Quality of Hospital Administrative Data. Available at SSRN* 4598486,

Sturm, M. L. (2023). Simulation Data Quality--Automatic checks on monitoring histograms (No. CERN-STUDENTS-Note-2023-180), CERN-STUDENTS-Note-2023-180.

Valencia Parra, Á. (2022). On the enhancement of big data pipelines through data preparation, data quality, and the distribution of optimization problems, https://hdl.handle.net/11441/141972.

Wolff, S. (2023). Design and implementation of a workflow for quality improvement of the metadata of scientific publications, https://doi.org/10.25366/2023.211.

Yaman, B., Thompson, K., Fahey, F., & Brennan, R. (2024). LinkedDataOps: quality oriented end-to-end geospatial linked data production governance. Semantic Web, (Preprint), 1-27, https://doi.org/10.3233/SW-233293.

Zhang, S., Benis, N., & Cornet, R. (2023). Automated approach for quality assessment of RDF resources. BMC Medical Informatics and Decision Making, 23 (Suppl 1), 90, https://doi.org/10.1186/s12911-023-02182-8.

Zhou, H., Han, L., Dermatini, G., Indulska, M., & Sadiq, S. (2022, October). *A Behavioral Analysis of Metadata Use in Evaluating the Quality of Repurposed Data.* In International Conference on Conceptual Modeling (pp. 310-324). Cham: Springer International Publishing, https://doi.org/10.1007/978-3-031-17995-2_22.

Zeinab Rahimi Rise

Zeinab Rahimi Rise Born in 1992, she holds a master’s degree in Industrial Engineering from Amirkabir University of Technology, earned in 2018. She previously obtained her bachelor’s degree in Industrial Engineering from Ershad-Damavand University in 2014. Her research interests include machine learning, image processing, and healthcare systems.

Mohammad Mahdi Ershadi

Mohammad Mahdi Ershadi Born in 1993, he holds a master’s degree in Industrial Engineering from Amirkabir University of Technology, which he obtained in 2017. He previously earned his bachelor’s degree in Industrial Engineering from Isfahan University of Technology in 2015. Since 2017, he has been an active researcher at Amirkabir University of Technology. His research interests include data engineering, operations research, and healthcare systems.

Mohammad Javad Ershadi

Mohammad Javad Ershadi: Born in 1983, he holds a doctorate in Industrial Engineering from Iran University of Science and Technology, where he graduated in 2015. He received his MSc in Industrial Engineering from Sharif University of Technology in 2009 and his BSc in Industrial Engineering from Isfahan University of Technology in 2005. He began his academic career in 2015 as a faculty member at IranDoc and is currently an associate professor in the IT Management Research Group. His work at IranDoc focuses on applying quality engineering and management principles to information science and services. His research interests include statistical quality control (SQC), total quality management (TQM), business process reengineering (BPR), optimization, meta-heuristic algorithms, systems analysis, and data mining.

Iranian Journal of
Information
Processing and
Management