

**FUNDAÇÃO INSTITUTO DE ADMINISTRAÇÃO
FACULDADE FIA DE ADMINISTRAÇÃO E NEGÓCIOS
PROGRAMA DE MESTRADO PROFISSIONAL EM GESTÃO DE NEGÓCIOS**

FABIO SANTANA REIS

**MENSURANDO A QUALIDADE DE DADOS: UMA PROPOSTA DE
QUESTIONÁRIO COMO INSTRUMENTO**

São Paulo

2025

FABIO SANTANA REIS

**MENSURANDO A QUALIDADE DE DADOS: UMA PROPOSTA DE
QUESTIONÁRIO COMO INSTRUMENTO**

Dissertação apresentada à Banca Examinadora do Programa de Mestrado Profissional em Gestão de Negócios, mantida pela Fundação Instituto de Administração, como requisito para a obtenção do título de Mestre em Gestão de Negócios, sob a orientação do Prof. Dr. Cesar Akira Yokomizo.

São Paulo

2025

Ficha Catalográfica Elaborada pelo Bibliotecário Chefe
Kleber Zornoff Manrubia CRB-8 /7561

Reis, Fabio Santana.

Mensurando a qualidade de dados: uma proposta de
questionário como instrumento. / Fabio Santana Reis. São Paulo, [s.n.]:
2025.

77 f.: il., tab.

Orientador: Prof. Dr. Cesar Akira Yokomizo.

Área de concentração: Administração.

Dissertação (Mestrado Profissional em Gestão de Negócios) –
Faculdade FIA de Administração e Negócios, Programa de
Pós-Graduação em Stricto Sensu, 2025.

1. Qualidade de dados. 2. DAMA-DMBOK. 3. Instrumento de
mensuração. 4. Validação de conteúdo. 5. Focus group. 6.
Gestão da
qualidade de dados. I. Yokomizo, Cesar Akira. II. Mestrado
Profissional. III. Faculdade FIA de Administração e Negócios.

RESUMO

O presente estudo teve como objetivo propor um instrumento de mensuração da Qualidade de Dados (QD) prático, replicável e fundamentado no modelo de seis dimensões do DAMA-DMBOK: completude, unicidade, atualidade, validade, precisão e consistência. Com essa finalidade, a pesquisa adotou uma abordagem mista, realizada em três fases distintas: realização de *focus group* com especialistas para propor sentenças representativas das dimensões; seleção das sentenças mais pertinentes por meio de questionário aplicado a especialistas, complementado por estatística descritiva; e validação de conteúdo com juízes especialistas, utilizando o Índice de Validação de Conteúdo (IVC). Ao término das três fases, desenvolveu-se um instrumento de mensuração da QD composto por 17 sentenças, bem como um produto tecnológico hospedado em meio virtual, com a finalidade de apoiar gestores na avaliação da QD das empresas. Como limitações da pesquisa, destaca-se a amostragem por conveniência e a ausência de técnicas estatísticas mais robustas, como Análise Fatorial Exploratória (AFE). Em pesquisas futuras, recomenda-se a aplicação do instrumento desenvolvido em diferentes setores de atividade econômica, bem como o desenvolvimento de novos instrumentos considerando outros *frameworks*, como o TDQM e a ISO 8000.

Palavras-chave: Qualidade de Dados; DAMA-DMBOK; Instrumento de mensuração; Validação de conteúdo; *focus group*; Gestão da qualidade de dados.

ABSTRACT

The objective of this study was to propose a practical, replicable Data Quality (DQ) measurement tool based on the six dimensions of the DAMA-DMBOK model: completeness, uniqueness, timeliness, validity, accuracy, and consistency. To this end, the research adopted a mixed approach, carried out in three distinct phases: conducting focus groups with experts to propose sentences representative of the dimensions; selecting the most relevant sentences through a questionnaire applied to experts, complemented by descriptive statistics; and validating the content with expert judges, using the Content Validation Index (IVC). At the end of the three phases, a QD measurement instrument consisting of 17 sentences was developed, as well as a technological product hosted in a virtual environment, with the purpose of supporting managers in evaluating the QD of companies. The limitations of the research include convenience sampling and the absence of more robust statistical techniques, such as Exploratory Factor Analysis (EFA). In future research, it is recommended that the developed instrument be applied in different sectors of economic activity, as well as the development of new instruments considering other *frameworks*, such as TDQM and ISO 8000.

Keywords: Data Quality, DAMA-DMBOK, Measurement instrument, Content validation, Focus group, Data quality management.

LISTA DE FIGURAS

Figura 1: Dimensões da qualidade de dados	17
--	----

LISTA DE QUADROS

Quadro 1: Dimensões da qualidade de dados por <i>framework</i>	20
Quadro 2: Definição dos constructos do modelo conceitual.....	25
Quadro 3: Roteiro para Realização de <i>Focus Group</i>	29
Quadro 4: Sentenças geradas nos <i>Focus Groups</i>	31
Quadro 5: Perfil dos participantes da etapa de seleção de sentenças	36
Quadro 6: Perfil dos participantes da etapa de análise de juízes.....	38
Quadro 7: Sentenças seleccionadas pelos juízes.....	49

LISTA DE GRÁFICOS

Gráfico 1: Quantidade de sentenças geradas nos <i>focus groups</i> por dimensão	39
Gráfico 2: Quantidade de sentenças filtradas	45
Gráfico 3: Sentenças avaliadas pelos juízes	47
Gráfico 4: Quantidade de sentenças com ICV maior ou igual a 0,50.....	50

LISTA DE TABELAS

Tabela 1: Ranqueamento das sentenças selecionadas nesta etapa.....	41
Tabela 2: Estatística Descritiva.....	43
Tabela 3: Detalhamento dos dados do Gráfico.....	48
Tabela 4: Exemplo de uso do produto tecnológico.....	55
Tabela 5: Exemplo de apresentação dos resultados da mensuração por dimensão de QD	57

LISTA DE EQUAÇÕES

Equação 1. Índice de Validade de Conteúdo (IVC)	37
--	----

SUMÁRIO

1 INTRODUÇÃO.....	12
1.1 Questão de pesquisa.....	13
1.2 Objetivos geral e específicos.....	13
1.3 Justificativa	14
1.4 Estrutura da pesquisa	15
2 REVISÃO DA LITERATURA.....	16
2.1 Mensuração da Qualidade de Dados: Importância, evolução e aplicações para os segmentos de negócios	16
2.2 Frameworks de mensuração da Qualidade de Dados	19
2.2.1 Data Management Body of Knowledge (DAMA-DMBOK).....	21
2.2.2 Total Data Quality Management (TDQM)	22
2.2.3 ISO 8000	23
2.3 Modelo Conceitual da Pesquisa	24
3 PROCEDIMENTOS METODOLÓGICOS.....	27
3.1 <i>Focus group</i>	28
3.2 Seleção de sentenças.....	30
3.3 Validação de conteúdo.....	37
4 ANÁLISE E DISCUSSÃO DOS RESULTADOS	39
4.1 Etapa 1: <i>focus group</i>	39
4.2 Etapa 2: seleção de sentenças.....	40
4.3 Etapa 3: validação de conteúdo	45
4.4 Discussão dos resultados com a literatura	53
5 PRODUTO TECNOLÓGICO.....	55
6 CONSIDERAÇÕES FINAIS.....	58
REFERÊNCIAS.....	61
APÊNDICE A – Roteiro do <i>focus group</i>	66
APÊNDICE B – Perguntas geradas pelos participantes do <i>focus group</i> 1.....	67
APÊNDICE C – Perguntas geradas pelos participantes do <i>focus group</i> 2.....	68
APÊNDICE D – Perguntas geradas pelos participantes do <i>focus group</i> 3.....	70
APÊNDICE E - Sentenças geradas nos <i>focus groups</i>.....	72
APÊNDICE F - Sentenças submetidas à análise de juízes.....	76
APÊNDICE G – Link para acesso ao produto tecnológico	77

1 INTRODUÇÃO

A mensuração da Qualidade de Dados (QD) tem passado por um processo de evolução constante e notável, impulsionado pelos avanços tecnológicos e pela crescente dependência das organizações por decisões cada vez mais orientadas por dados de boa qualidade. Inicialmente, os processos de mensuração da QD tinham como base métodos tradicionais, especialmente com sistemas construídos por meio de regras e modelos estatísticos, que por um tempo foram eficazes na análise de dados estruturados. Entretanto, diante dos crescentes avanços tecnológicos, o volume, variedade e velocidade com que os dados estão sendo gerados tornaram as abordagens tradicionais insuficientes, fomentando o surgimento de novas metodologias de mensuração da QD, incluindo os dados não estruturados e semiestruturados (Miller *et al.*, 2025; Mishra; Darade, 2024).

A relevância da QD tem sido reforçada nos mais diversos segmentos de atividade econômica, estando presente desde a previsão de demanda no varejo até a medição da qualidade do ar em grandes metrópoles, em todos esses casos a QD é fundamental para análises de maior precisão e tomadas de decisão cada vez mais eficientes. Nessa conjuntura, a pesquisa e inovação contínua dos processos de mensuração da QD evidenciam-se fundamentais para que as organizações possam alavancar seus ativos de dados, obter vantagem competitiva e excelência operacional dado o surgimento de cenários dependentes de dados cada vez mais complexos (Mishra; Darade, 2024; Yang *et al.*, 2025).

No contexto corporativo atual, a QD evidencia-se fundamental para a assecuração de que as decisões sejam baseadas em informações precisas, com vista a tomadas de decisão cada vez mais aprimoradas. Em adição, reitera-se a QD como um pilar da performance dos sistemas de apoio à decisão, que dependem de dados de qualidade para o suporte adequado aos usuários no processo de tomada de decisão baseada em dados (Hägele *et al.*, 2025; Yeung; Clemente; Lambiotte, 2025; Miller; Spiegel, 2025).

A QD pode ser definida como um conjunto de valores e atributos, que são nomeados na literatura acadêmica e no meio profissional, frequentemente, como dimensões. Essas dimensões podem assumir abordagens quantitativas ou qualitativas e tem a finalidade de possibilitar a mensuração da qualidade de estruturas de dados. Nessa conjuntura, a QD demonstra-se fundamental para a garantia da eficiência operacional e melhoria da confiabilidade dos sistemas digitais. A QD abrange uma série de fatores, tais como a adequação, perfeição, atualidade e confiabilidade dos dados. Logo, com vista a garantir a QD, os setores econômicos, da iniciativa pública e privada, que atuam nessas frentes, costumam desenvolver ferramentas de avaliação

considerando padrões, estruturas, diretrizes e, até mesmo, regulamentações (Abhiyasa *et al.*, 2024; Miller *et al.*, 2025; Yulfitri *et al.*, 2024).

A QD está associada à capacidade de alcançar padrões satisfatórios de qualidade para os dados utilizados. Diante da complexidade e da necessidade fundamental de padronização da gestão da qualidade de dados, *frameworks* como o *Data Management Body of Knowledge* (DAMA-DMBOK) ganham notoriedade. O DAMA-DMBOK é uma estrutura abrangente que fornece diretrizes detalhadas para o gerenciamento de dados, com ênfase para a governança de dados enquanto eixo central das atividades de gerenciamento da QD, com vista ao equilíbrio e consistência de todas as áreas, que podem incluir, mas não se resumem, a modelagem, segurança e qualidade dos dados. As dimensões da QD são precisão, completude, consistência, pontualidade, exclusividade e validade representam eixos fundamentais nas estruturas DAMA-DMBOK (Abhiyasa *et al.*, 2024; Miller *et al.*, 2025; Yulfitri *et al.*, 2024).

A importância de empregar o DAMA-DMBOK em avaliações voltadas para a QD é evidenciada pelas contribuições desse *framework* para a melhoria na governança dos dados, uma vez que o DAMA-DMBOK pode contribuir para que as políticas e procedimentos sejam seguidos com uniformidade, o valor dos dados seja maximizado e que os riscos associados à QD sejam gerenciados e abordados com maior eficiência. Nesse contexto, dados de alta qualidade, garantidos por meio da aplicação do DAMA-DMBOK, podem ser essenciais para uma tomada de decisão mais precisa, segura e eficaz, com vista a minimizar as incertezas que permeiam problemas relativos à baixa QD (Miller *et al.*, 2025; Pietersen; Rudman, 2025).

1.1 Questão de pesquisa

Diante do exposto, considerando o potencial de contribuição do DAMA-DMBOK para a QD, o presente estudo parte da seguinte questão norteadora: Como mensurar a qualidade de dados de forma confiável, prática e alinhado à estrutura conceitual fundamentada nas seis dimensões do DAMA-DMBOK?

1.2 Objetivos geral e específicos

Com vista a alcançar a resposta para a questão supracitada, a presente pesquisa tem como objetivo geral propor um instrumento de mensuração da qualidade de dados prático, replicável e fundamentado numa estrutura conceitual sólida, o modelo de seis dimensões propostas pelo DAMA-DMBOK: completude, unicidade, atualidade, validade, precisão e consistência.

Ainda, como objetivos específicos, pretende-se:

1. Propor sentenças representativas das seis dimensões do modelo DAMA-DMBOK por meio de *focus group* com especialistas;
2. Selecionar as sentenças mais pertinentes por meio de questionário sob a perspectiva de especialista; e
3. Validar o conteúdo das sentenças selecionadas por juízes especialistas, utilizando o Índice de Validação de Conteúdo (IVC).

1.3 Justificativa

A QD representa um pilar fundamental para a tomada de decisão eficiente das organizações, abrangendo todos os setores econômicos, especialmente no contexto atual de expansão tecnológica, em que a dependência de sistemas construídos por uma quantidade cada vez maior de dados é acentuada no meio corporativo. Dados com baixa qualidade podem ocasionar informações de baixa precisão, desempenhos insatisfatórios nos negócios e, por conseguinte, levando a perdas financeiras significativas, o que impacta diretamente na sobrevivência das organizações.

Nesse cenário, a mensuração da QD configura-se como uma etapa fundamental para o gerenciamento da QD, facilitando a identificação prévia de problemas, análise das causas e implementação das melhorias necessárias. Apesar da evidenciada importância para o meio corporativo global, lacunas significativas na oferta de instrumentos de mensuração práticos, replicáveis e robustos ainda persistem (Miller *et al.*, 2025; Nugraha *et al.*, 2024).

No ponto de vista teórico, a proposição de um instrumento de mensuração da QD, com base em um modelo conceitual consolidado na literatura, como o das seis dimensões do DAMA-DMBOK, pode contribuir para o estado da arte sobre o tema ao passo que mitiga fragmentações da literatura existente. Conforme evidenciado por Miller *et al.* (2025) a complexidade de terminologias e a falta de padronização nas dimensões da QD entre diferentes *frameworks* representa um desafio substancial para as estruturas de mensuração da QD (Abhiyasa *et al.*, 2024; Miller *et al.*, 2025; Rahmawati *et al.*, 2023). Logo, a proposição de aplicar as seis dimensões do DAMA-DMBOK pode oferecer uma base mais sólida e reconhecida globalmente para a mensuração da QD, ao passo em que fortalece a reprodutibilidade do instrumento desenvolvido em outros contextos, sejam esses acadêmicos ou organizacionais.

Em termos práticos, o instrumento proposto pode constituir uma base inicial para o desenvolvimento de ferramentas voltadas para a avaliação sistemática da QD em organizações que enfrentem dificuldades de avaliar a QD. Ademais, o instrumento proposto, focado nas seis dimensões do DAMA-DMBOK, pode contribuir para identificação e quantificação de problemas relativos à QD, fomentar o desenvolvimento de melhorias na QD mais direcionadas e incentivar culturas organizacionais mais preocupadas com a QD.

Ratifica-se que a escolha do DAMA-DMBOK como estrutura conceitual para o desenvolvimento do instrumento de mensuração da QD proposto justifica-se pela abrangência e reconhecimento global desse *framework*. O DAMA-DMBOK é um grupo de processos e áreas de conhecimento amplamente aceito no fornecimento de práticas de gerenciamento de dados, abrangendo dez áreas de conhecimento fundamentais, que incluem a governança, arquitetura, modelagem, design, armazenamento, operação, segurança, integração e qualidade de dados. Essa amplitude do DAMA-DMBOK pode contribuir para o asseguramento de que o instrumento de mensuração desenvolvido neste estudo não represente um esforço isolado, mas sim parte de uma estratégia holística de gerenciamento de dados (Abhiyasa *et al.*, 2024; Miller *et al.*, 2025; Taryono *et al.*, 2024).

Em adição, ressalta-se que, embora não seja regulatório por si só, o DAMA-DMBOK exerce influência significativa em outros *frameworks*, como o *UK Government Data Quality Framework*, e é amplamente aceito no quesito de fornecimento de melhores práticas em gerenciamento de dados. Dessa forma, representando uma importante ferramenta para o alcance da conformidade dos dados com padrões e regulamentações, bem como a garantia de que os dados empregados em processos de negócios tenham maior qualidade (Miller *et al.*, 2025; Pietersen; Rudman, 2025).

1.4 Estrutura da pesquisa

O presente estudo encontra-se subdividido nos seguintes capítulos: Introdução, abrangendo contextualização do estudo, questão de pesquisa, objetivos geral e específicos, justificativa com contribuições teóricas e práticas, apresentação da proposta de produto tecnológico e estruturação do estudo; Referencial teórico, com os capítulos mensuração da qualidade de dados; DAMA-DMBOK e demais *frameworks* de mensuração da QD; e o uso do *focus group* para desenvolvimento de instrumentos de mensuração; Procedimentos metodológicos; Análise e discussão dos resultados e considerações finais da pesquisa.

2 REVISÃO DA LITERATURA

A presente seção encontra-se subdividida nos seguintes tópicos: inicialmente, apresenta-se discussão teórica acerca da mensuração da QD, considerando a importância do tema, a evolução dos instrumentos de mensuração da QD e as aplicações da mensuração da QD para os segmentos de negócios. Em sequência, tratar-se-á dos principais *frameworks* de mensuração da QD, considerando as convergências e divergências metodológicas e estruturais que esses possuem com o *framework* central deste estudo, o DAMA-DMBOK. Por fim, apresenta-se o modelo conceitual da pesquisa.

2.1 Mensuração da Qualidade de Dados: Importância, evolução e aplicações para os segmentos de negócios

Em um contexto marcado pelo papel da Inteligência Artificial (IA) na gestão empresarial, que tem apresentado influências crescentes na tomada de decisões e na competitividade dos negócios, a QD demonstra-se cada vez mais fundamental. Os modelos baseados em dados constituem uma das principais forças motriz por trás da transformação digital em ascensão, apesar disso, esses têm enfrentado desafios relativos à QD, diante da necessidade de modelos baseados em dados cada vez mais precisos para atender demandas em um contexto corporativo marcado pela complexidade (Liu *et al.*, 2025; Schuessler *et al.*, 2025).

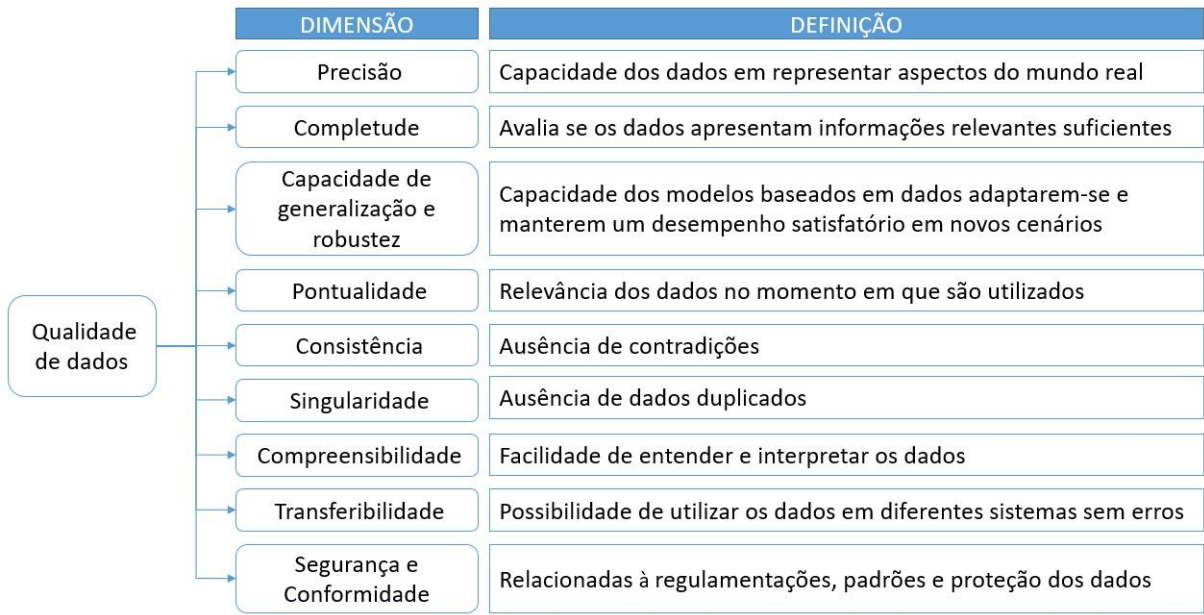
A QD evidencia-se essencial nos mais diversos segmentos de negócios, desde os setores industriais, de saúde, até setores como o da defesa nacional dos países (Polimeno *et al.*, 2025; Schuessler *et al.*, 2025; Wang; Pan; Zhang, 2025). A reprodução e precisão de qualquer pesquisa depende da integridade dos dados que, por sua vez, é influenciada pela manipulação adequada dos dados, desde o planejamento até o processamento e análise (Miller; Spiegel, 2025).

A evolução da qualidade dos dados possui alguns aspectos fundamentais, com destaque para o contexto da evolução e o impacto nos modelos baseados em dados. Por exemplo, no setor de manufatura, a estrutura e a estabilidade dos modelos podem ser melhoradas com o aumento da quantidade de dados disponíveis, ao passo em que melhorar a precisão e capacidade de generalização de dados pode requerer um volume de dados expressivamente ainda maior. O desafio para o setor, no que tange a QD, está na capacidade de equilíbrio entre desempenho e generalização diante de ambientes de negócios industriais cada vez mais dinâmicos (Liu *et al.*, 2025).

Assim como no setor industrial, os segmentos de negócio da área de saúde também são altamente dinâmicos. As práticas médicas, as tecnologias do setor de saúde e as necessidades dos pacientes mudam de forma constante e rápida. Nesse contexto, destaca-se a necessidade de estratégias voltadas para a validação constante com vista a consistência e aplicabilidade desses dados ao longo do tempo. Logo, a relevância dos dados utilizados evidencia-se tão importante quanto o volume de dados disponíveis para a tomada de decisão na área médica. Os modelos de dados podem ser utilizados também para prontuários eletrônicos de saúde. A disponibilidade cada vez maior de dados nesse setor apresenta desafios de qualidade, especialmente relativos à incompletude, inconsistência e imprecisão dos dados. Portanto, evidencia-se, para ambos os setores, a importância da QD (Hunsdieck *et al.*, 2025; Schuessler *et al.*, 2025).

A utilização de dados de qualidade pode contribuir de muitas formas para os resultados empresariais, com melhoria da capacidade de precisão e generalização de sistemas, redução de custos e de tempo necessário para a realização de atividades, detecção e prevenção de eventuais instabilidades, tomadas de decisão informadas, validação e robustez de modelos, otimização do desempenho de modelos e confiabilidade dos dados (Liu *et al.*, 2025; Miller; Spiegel, 2025; Schuessler *et al.*, 2025; Wang; Pan; Zhang, 2025). A Figura 1 apresenta algumas das principais dimensões da qualidade de dados identificadas na literatura discutida.

Figura 1: Dimensões da qualidade de dados



Fonte: Elaborado pelo autor com base em Miller *et al.* (2025).

Conforme evidenciado na Figura 1, uma série de fatores constituem determinantes da QD, tais como a precisão, voltada para a capacidade dos dados em representar aspectos do mundo real; a completude, que avalia se os dados apresentam informações relevantes suficientes; a capacidade de generalização e a robustez, relativas à capacidade dos modelos baseados em dados de adaptar-se e manter um desempenho satisfatório em novos cenários; a pontualidade, que refere-se à relevância dos dados no momento em que são utilizados; a consistência, em outras palavras, ausência de contradições; a singularidade, com vista a evitar dados duplicados; a compreensibilidade, relativa a facilidade de entender e interpretar os dados; a transferibilidade, que refere-se à possibilidade de utilizar os dados em diferentes sistemas sem erros; e, não menos importante, a segurança e conformidade, voltadas para regulamentações, padrões e proteção dos dados (Liu *et al.*, 2025; Miller; Spiegel, 2025; Polimeno *et al.*, 2025; Schuessler *et al.*, 2025; Wang; Pan; Zhang, 2025).

Nessa conjuntura, mensurar a QD é, ao mesmo tempo, um desafio e uma necessidade das organizações modernas. A QD é definida, medida, aprimorada, monitorada e gerenciada por meio da mensuração da qualidade das dimensões (Rahmawati *et al.*, 2023). As empresas medem a QD por meio de combinações de abordagens e ferramentas que têm evoluído de forma significativa, seguindo os avanços tecnológicos acelerados e a crescente complexidade dos dados. A avaliação da QD inicia com a medição de dimensões de qualidade, que são as características inerentes de um grupo de dados, com vista a assegurar uma mensuração orientada para a avaliação da qualidade (Yaman *et al.*, 2024).

Em âmbito internacional, identifica-se uma variedade de dimensões consideradas na mensuração da QD, a *International Organization for Standardization* (ISO) possui um conjunto de normas técnicas nesse sentido, como a ISO 8000, voltada para a qualidade de dados genéricos; ISO 19157, qualidade de dados geoespaciais; e a ISO 25012, voltada para a qualidade de dados em sistemas de computador. Cada uma dessas normas apresenta um conjunto de dimensões e métricas de avaliação da QD, tais como a completude, consistência, acurácia e pontualidade, já citadas anteriormente (Yaman *et al.*, 2024).

As métricas de mensuração tomam como base regras pré-estabelecidas, ferramentas de mensuração da QD comumente aplicam um grupo de regras com vista a avaliar a conformidade dos dados utilizados. Atualmente, os modelos de dados semânticos constituem uma estratégia de integração de critérios de mensuração diversos, baseados em múltiplas ferramentas e padrões. Logo, as empresas possuem uma vasta gama de ferramentas de mensuração da QD, que tomam como base regras pré-estabelecidas para avaliar a QD (Yaman *et al.*, 2024).

A melhoria contínua da QD depende do monitoramento e avaliação de desempenho constantes dos dados utilizados, que constituem uma necessidade crítica, voltada para a identificação e resolução sistemática de problemas. Esse monitoramento e avaliação deve considerar diversos fatores, entre eles as métricas de processo, como eficácia, eficiência e qualidade, objeto desta seção de discussão teórica, que refere-se à conformidade dos dados com padrões pré-definidos; a utilização de modelos de maturidade, com vista a avaliar o nível de sofisticação e eficácia da gestão da QD, identificação de lacunas e planejamento de melhoras de forma continuada; e a visibilidade e rastreabilidade dos dados, que se traduz na capacidade de entender o fluxo de dados de toda a organização, com vista a identificar eventuais causas de problemas de qualidade (Bernardo *et al.*, 2024; Schmuck, 2024; Wings; Harkonen, 2025; Yaman *et al.*, 2024; Zhang; Callaghan-koru; Koru., 2024).

Logo, afirma-se que a mensuração da QD representa um pilar de expressiva importância para as organizações por uma série de motivos, com impactos para a eficiência operacional, tomada de decisões, sustentabilidade, conformidade regulatória, mitigação de riscos, inovação, vantagem competitiva, confiança e colaboração com os *stakeholders* (Bernardo *et al.*, 2024; Meloni *et al.*, 2025; Sharma *et al.*, 2025; Wings; Harkonen, 2025).

A discussão levantada nessa seção, apresenta a mensuração da QD como um processo complexo, contínuo e multifacetado, que incorpora a definição de padrões, implementação e utilização de ferramentas e tecnologia avançada e a integração de práticas de gestão voltadas para a QD. A importância da QD está nos mais diversos segmentos da organização, desde a tomada de decisão, otimização da eficiência operacional, asseguração de maior conformidade, promoção da inovação e relacionamento com os *stakeholders*. A seguir, apresenta-se a seção do referencial teórico que versará sobre *frameworks* de mensuração da QD, com ênfase para o DAMA-DMBOK.

2.2 Frameworks de mensuração da Qualidade de Dados

Essa seção do referencial teórico tem como objetivo apresentar alguns dos principais *frameworks* de mensuração da QD, conforme delimitado na literatura recente. Dessa forma, os *frameworks* serão apresentados com vista à identificação de interseções e diferenças estruturais desses com o *framework* central dessa pesquisa, o DAMA-DMBOK. Os *frameworks* abordados a seguir serão DAMA-DMBOK, Gestão da Qualidade Total de Dados (TDQM) e ISO 8000 (Miller *et al.*, 2025). Além desses *frameworks*, existem outros voltados para a mensuração da QD, tais como a ISO 25012, Princípios de Prática de Informação Justa (FIPPS), Estrutura de

Garantia de Qualidade do Sistema Estatístico Europeu (ESS QAF), Estrutura de Avaliação da Qualidade de Dados do Fundo Monetário Internacional (FMI) (DQAF) e a estrutura de Garantia de Qualidade de Dados da Organização Mundial da Saúde (OMS) (Miller *et al.*, 2025). O Quadro 1 ilustra as dimensões que cada um dos *frameworks* possui em comum ou diferente do DAMA-DMBOK.

Quadro 1: Dimensões da qualidade de dados por *framework*

Dimensão	DAMA DMBOK	ISO 8000	TDQM	ISO 25012	FIPPS	ESS QAF	IMF	BCBS 239	ALCOA+	WHO
Acessibilidade			X	X		X	X			
Acurácia	X	X	X	X	X	X	X	X	X	X
Disponibilidade				X		X			X	
Compleitude	X	X	X	X	X			X	X	X
Conformidade				X						
Confidencialidade			X	X		X				
Consistência	X	X	X	X		X	X		X	X
Atualização							X			
Atualidade	X	X	X	X	X	X		X		X
Eficiência				X		X				
Governança								X	X	X
Portabilidade				X						
Precisão			X	X						
Quantidade			X							
Recuperabilidade				X						
Semântica							X			
Rastreabilidade				X		X	X		X	
Entendimento			X	X		X	X	X	X	
Unicidade	X	X	X		X	X	X			
Validade	X	X	X	X		X	X			

Nota: Dimensões sombreadas correspondem às dimensões do DAMA-DMBOK.

Fonte: Elaborado pelo autor com base em Miller *et al.* (2025).

A seguir, apresenta-se discussão teórica acerca dos principais aspectos do DAMA-DMBOK, *Total Data Quality Management* (TDQM) e ISO 8000, com ênfase para as diferenças e similaridades desses dois últimos com o DAMA-DMBOK.

2.2.1 Data Management Body of Knowledge (DAMA-DMBOK)

O DAMA-DMBOK é um *framework* de importante reconhecimento no campo dos estudos e na área empresarial voltada para a gestão de dados (Miller *et al.*, 2025; Pietersen; Rudman, 2025). Além disso, esse *framework* é reconhecido também por ser voltado para o gerenciamento de dados e para a QD constituído por uma estrutura de conhecimento abrangente (Khairunisak; Kusumasari; Fauzi, 2021), podendo ser utilizado como base para projetos de várias áreas da gestão de dados, incluindo técnicas de avaliação da gestão de segurança de dados, diretrizes voltadas para a gestão de governança de dados e *Business Intelligence* (Febiryani; Kusumasari; Fauzi, 2021; Khairunisak; Kusumasari; Fauzi, 2021).

A literatura apresenta recomendações do DAMA-DMBOK para a preparação de relatórios voltados para a QD, dada a sua capacidade de descrever problemas identificados durante a fase de exploração dos dados, bem como propor abordagens de correção e mitigação desses problemas. Logo, é possível inferir que o DAMA-DMBOK fornece uma base referencial para subsidiar o tratamento da QD, com a identificação de práticas e processos mais eficientes a serem adotados (Ferreira; Silva, 2018).

Em adição, o DAMA-DMBOK também é reconhecido como uma base importante para os processos de revisão sistemática de arquiteturas analíticas de dados, incluindo as arquiteturas de *Big Data* (Priebe; Neumaier; Markus, 2021). Os princípios e objetivos do DAMA-DMBOK podem ser empregados na governança de dados, ao passo em que demonstram capacidade de identificar riscos de forma abrangente, por exemplo, no contexto da aprendizagem de máquina para previsão de demandas de clientes (Pietersen; Rudman, 2025).

No que tange suas características e aplicações específicas, o DAMA-DMBOK pode subsidiar o desenvolvimento de processos de gestão de segurança de dados; desenvolvimento de diretrizes de governança de dados, voltado para a otimização do uso de dados e controles internos das empresas; gestão de metadados; e melhorias gerais voltadas para a qualidade e consistência dos dados (Febiryani; Kusumasari; Fauzi, 2021; Hendrawan; Kusumasari; Fauzi, 2022; Khairunisak; Kusumasari; Fauzi, 2021; Zahabia; Kusumasari; Fauzi, 2022).

Apesar de sua abrangência de aplicações e contribuições para a QD, o *framework* DAMA-DMBOK ainda apresenta limitações, como o caráter genérico de suas diretrizes, que

não especificam de forma precisa o que deve ser medido ou a associação de medidas de QD em contextos específicos (Ferreira; Silva, 2018). Ademais, sua aplicação evidencia-se especialmente útil para empresas de grande porte, não apresentando a mesma facilidade de aplicação em Pequenas e Médias Empresas (PMEs), que demandam modelos melhor adaptados para suas realidades (Hai, 2025).

Em suma, observa-se que o DAMA-DMBOK é um importante *framework*, especialmente em razão de seu caráter abrangente para a avaliação da QD e para a gestão de dados de um modo geral. Esse apresenta contribuições para campos diversos da gestão de dados, como a segurança, a governança, metadados, qualidade e arquitetura dos dados. Apesar disso, sua estrutura voltada para contextos mais gerais pode constituir uma limitação, ao passo em que cenários específicos ou PMEs podem carecer de diretrizes mais específicas para seus escopos, limitando as contribuições do DAMA-DMBOK nesses contextos.

2.2.2 Total Data Quality Management (TDQM)

O *Total Data Quality Management* (TDQM) é um *framework* de abordagem estratégica e holística, voltado para o monitoramento da QD nas organizações. Esse *framework* foi desenvolvido na década de 1980 pelo *MIT Sloan School of Management*, representando um dos *frameworks* pioneiros na área de QD, com influência significativa para o desenvolvimento desse seguimento. O TDQM é um *framework* voltado para o desenvolvimento de metodologias que garantam alta qualidade dos dados, composto por quatro etapas, conhecidas como o ciclo DMAI, em português: Definir, Medir, Analisar e Melhorar (Bowo *et al.*, 2019; Ershadi; Mohseni; Niaki, 2023; Miller *et al.*, 2025).

A etapa de Definição, refere-se à determinação das dimensões da QD que são relevantes para cada organização e para os dados específicos de cada contexto. A etapa de medição visa a avaliação da condição atual da QD, com a identificação de eventuais problemas e compreensão dos efeitos desses problemas para a organização. A etapa de análise visa a investigar as razões por trás dos desafios da QD. Por fim, a etapa de melhoria tem a finalidade de aplicar modificações com vista a melhorar a QD, considerando os desafios identificados nas etapas anteriores (Bowo *et al.*, 2019; Rahmawati *et al.*, 2023; Miller *et al.*, 2025).

Logo, o TDQM pode ser considerado um *framework* de implementação e mudanças na cultura organizacional no que tange a gestão dos dados. Tem a finalidade de contribuir para o estabelecimento de um ambiente corporativo saudável, voltado para a gestão contínua da QD.

Esse *framework*, portanto, apresenta uma estrutura robusta e abrangente de supervisão da QD (Miller *et al.*, 2025).

O TDQM considera quatro áreas fundamentais para a superação dos desafios da QD: sistemas, processos, procedimentos e políticas, sendo intrinsecamente ligado à QD, vez que é um *framework* que fornece uma abordagem voltada para a facilitação e melhoria da QD, por meio da gestão dos dados. Esse, portanto, traz a QD como um produto, com foco em como esses dados são definidos, medidos, analisados e aprimorados em um ciclo contínuo de gestão da QD (Bowo *et al.*, 2019; Ershadi; Mohseni; Niaki, 2023).

O TDQM possui alguns pontos em comum com o DAMA-DMBOK, tais como o foco na qualidade de dados, uma vez que ambos os *frameworks* pautam sobre a gestão da QD nas organizações, com vista a garantir a adequação dos dados; ambos os *frameworks* utilizam dimensões para a descrição e mensuração da QD, com dimensões compartilhadas, tais como completude, validade, precisão, pontualidade e consistência; ambos são caracterizados por apresentar uma abordagem abrangente, com capacidade de contribuições para diversas áreas de negócios e do conhecimento; e ambos são pautados na ideia de melhoria contínua da QD (Bowo *et al.*, 2019; Rahmawati *et al.*, 2023; Miller *et al.*, 2025).

Apesar disso, também apresentam diferenças, com destaque para o escopo e foco de cada *framework*. Enquanto o TDQM é especificamente voltado para a gestão da QD, por meio do ciclo DMAI, o DAMA-DMBOK apresenta um caráter mais amplo, voltado para o gerenciamento amplo de dados, a QD, no caso do DAMA-DMBOK é uma das áreas de conhecimento contempladas, ao lado da governança, arquitetura, modelagem, segurança e design dos dados (Rahmawati *et al.*, 2023; Miller *et al.*, 2025).

2.2.3 ISO 8000

A ISO 8000 é composta por uma série de normas internacionais voltadas para o estabelecimento de padrões globais de QD, com ênfase para esses aspectos no âmbito empresarial. Essa série de normas fornece estruturas voltadas para a melhoria da QD, ao passo em que abrange princípios gerais e específicos de qualidade, com potencial de aplicação em todas as fases do ciclo de vida dos dados (Miller *et al.*, 2025).

A ISO 8000 possui um ciclo semelhante ao DMAI do TDQM, o PDCA (em português, planejar, fazer, verificar e agir), com vista a melhoria da QD. As dimensões da QD que a ISO 8000 abrange são: precisão, completude, consistência, pontualidade, exclusividade e validade. As normas abrangem aspectos diversos, tais como modelos de referência para o gerenciamento

da QD, avaliação da maturidade organizacional no processo de gerenciamento da QD, dentre outros (Lee *et al.*, 2019; Kim *et al.*, 2022; Miller *et al.*, 2025).

No que se refere às diferenças e semelhanças entre o DAMA-DMBOK e a ISO 8000, ressalta-se que ambos são voltados para a gestão da QD e possuem, em suas finalidades, o objetivo de melhorar as práticas de dados. Apesar disso, em termos de escopo e natureza esses são diferentes. Enquanto a ISO 8000 é composta por uma série de normas voltadas para o estabelecimento de padrões internacionais da QD no contexto empresarial, o DAMA-DMBOK é formado por um conjunto detalhado de diretrizes com vista à superação de desafios relacionados à QD. Entende-se que a ISO 8000 é mais focada em padrões e certificações, enquanto o DAMA-DMBOK é um guia abrangente de boas práticas de gestão da QD (Miller *et al.*, 2025).

2.3 Modelo Conceitual da Pesquisa

Considerando a revisão de literatura apresentada, este estudo adota como base teórico-metodológica para o instrumento de mensuração da qualidade de dados o modelo proposto pelo DAMA-DMBOK. Essa escolha se fundamenta no fato de que o DAMA-DMBOK representa um dos *frameworks* mais amplamente utilizados e reconhecidos em diversos setores organizacionais, proporcionando padronização e comparabilidade entre diferentes contextos (DAMA International, 2017). Conforme apresentado nas seções anteriores deste estudo, o DAMA-DMBOK é composto pelas dimensões precisão, completude, consistência, pontualidade, exclusividade e validade (Abhiyasa *et al.*, 2024; Miller *et al.*, 2025; Yulfitri *et al.*, 2024).

Cada dimensão está apresentada considerando seus aspectos constitutivos, voltados para a definição da dimensão, e aspectos operacionais, que se referem aos meios de mensuração que serão adotados durante a realização da pesquisa. O Quadro 2 apresenta cada uma das dimensões do DAMA-DMBOK, seguida da definição e estratégia de mensuração adotada. As dimensões do DAMA-DMBOK exclusividade e pontualidade estão apresentadas no modelo teórico conceitual como, unicidade e atualidade, respectivamente. Essa alteração se deu considerando a nomenclatura utilizada pelo Dama International (2017).

Quadro 2: Definição dos constructos do modelo conceitual

Dimensão	Definição	Como mensurar
QCP – ‘Compleitude’	Representa a proporção de dados armazenados em relação ao seu potencial total, evitando lacunas que possam comprometer análises e decisões.	Será mensurada através de: (1) itens que avaliam a existência de todos os campos obrigatórios; (2) percentual de preenchimento de registros; (3) percepção dos usuários sobre a suficiência das informações disponíveis para suas atividades.
QUN – ‘Unicidade’	Assegura que não haja duplicidade de registros, preservando a integridade e a clareza das informações.	Será avaliada por: (1) existência de controles para identificação de registros duplicados; (2) consistência de identificadores únicos; (3) frequência percebida de problemas relacionados à duplicidade.
QAT – ‘Atualidade’	Diz respeito à disponibilidade dos dados no tempo certo, essencial para a tomada de decisões ágeis e pertinentes.	Será mensurada por: (1) frequência de atualização; (2) defasagem temporal entre a ocorrência dos fatos e seu registro nos sistemas; (3) percepção dos usuários sobre a tempestividade das informações disponíveis.
QVA – ‘Validade’	Indica o grau de conformidade dos dados com regras, formatos e restrições previamente definidos, garantindo que sejam apropriados ao uso.	Será avaliada por: (1) existência de validações automáticas; (2) conformidade com padrões estabelecidos; (3) percepção dos usuários sobre a adequação dos formatos e valores aos requisitos do negócio.
QPR – ‘Precisão’	Reflete o quanto os dados representam de forma fiel a realidade que expressam.	Será mensurada por: (1) correspondência entre os registros nos sistemas e suas contrapartes no mundo real; (2) taxa de erros identificados; (3) percepção dos usuários sobre a confiabilidade das informações utilizadas.
QCS – ‘Consistência’	Trata da ausência de discrepâncias entre diferentes representações do mesmo dado em diferentes sistemas ou bases, mantendo uniformidade dentro do contexto da organização.	Será avaliada por: (1) coerência dos dados entre diferentes plataformas e departamentos; (2) existência de processos de conciliação; (3) percepção dos usuários sobre contradições encontradas nos dados de múltiplas fontes.

Fonte: Elaborado pelo autor com base em DAMA International (2017).

Com o desenvolvimento do instrumento de mensuração, pretende-se não apenas avaliar objetivamente o estado atual da qualidade dos dados nas organizações, mas também identificar áreas prioritárias para intervenção e melhoria, contribuindo tanto para o avanço teórico no campo da gestão da qualidade de dados quanto para aplicações práticas que visam aprimorar os processos decisórios baseados em dados confiáveis.

Ratifica-se que, conforme evidenciado na seção anterior deste referencial teórico, apesar de existirem outros *frameworks* de grande relevância para a mensuração da QD, como o TDQM e a ISO 8000, a escolha do DAMA-DMBOK como base teórico metodológica desta pesquisa se deu em razão do escopo abrangente, em comparação com os demais *frameworks*. Ademais, observa-se que o TDQM e a ISO 8000 apresentam um escopo mais direcionado, com o TDQM apresentando um ciclo fechado em sua estrutura e a ISO voltada para a padronização da mensuração da QD (Bowo *et al.*, 2019; Lee *et al.*, 2019; Kim *et al.*, 2022).

Apesar de seu caráter abrangente e consolidado, é importante ratificar que o DAMA-DMBOK também possui limitações, como o caráter genérico, muito em função de sua abrangência, bem como sua pouca aplicabilidade para o caso das PMEs (Ferreira; Silva, 2018; Hai, 2025). Diante do exposto, este estudo pretende permitir que o instrumento desenvolvido possa manter o rigor conceitual do DAMA-DMBOK, ao passo em que também reconhece as especificidades de aplicação derivadas dessa escolha pelo *framework* DAMA-DMBOK. A seguir, apresentam-se os procedimentos metodológicos da pesquisa.

3 PROCEDIMENTOS METODOLÓGICOS

De acordo com as recomendações de Hernandez-Nieto (2022) para construção e validação de instrumentos psicométricos, orientada pelo objetivo da pesquisa, adotou-se uma abordagem qualitativa e exploratória, estruturada em três etapas: a primeira etapa foi composta por três sessões de *focus group* com profissionais de diferentes perfis acadêmicos e funções corporativas que influenciam ou até mesmo tomam decisões de negócio com base em análise de dados. Nessas sessões, foram abordadas cada uma das dimensões como tema em rodadas separadas com duração média de dez minutos, em que os participantes deveriam compreender a dimensão abordada e contribuir ao final de cada rodada, de acordo com sua percepção e experiência profissional, com sentenças que representem a dimensão explorada na interação entre os participantes.

A partir das sentenças levantadas na etapa anterior, a segunda etapa consistiu em aplicar um questionário dividido em seis sessões, representando respectivamente as dimensões atualidade, validade, unicidade, precisão, completude e consistência. Em cada sessão, foram agrupadas as sentenças correspondentes à dimensão vinculada. Para cada sentença, o participante devia assinalar o quanto concordava com a pertinência da sentença à dimensão onde está agrupada. Desse modo, devia escolher um número compreendido numa escala entre 1 (discordo totalmente) e 5 (concordo totalmente). Com base nas sentenças elaboradas, dados de estatística descritiva foram gerados por meio do *software* de uso livre R, com vista a auxiliar na avaliação das pontuações que os participantes deram para cada sentença.

Para a última etapa, que consistia na validação de conteúdo por juízes, em que especialistas avaliaram as questões quanto à clareza, relevância e alinhamento a cada uma das dimensões analisadas, foram selecionadas as sentenças mais relevantes de acordo com os apontamentos dos participantes da etapa anterior. Para isso, adotaram-se os seguintes critérios: ao menos dez por cento de sentenças em cada sessão deverá compor a presente etapa, sendo que o primeiro critério de seleção é pela maior quantidade de notas 5 atribuídas à sentença. Caso esse critério não atenda à quantidade estabelecida pela primeira condição, seria necessário adotar como segundo critério a maior quantidade de notas 4. Assim, uma vez que as sentenças tenham sido selecionadas de acordo com as regras enunciadas, seguiu-se para a análise de juízes, onde os participantes leram as sentenças numa ordem aleatória e identificaram, de acordo com sua percepção, a qual dimensão está relacionada cada uma.

A decisão por adotar procedimentos com abordagem mista, por meio do *focus group*, enquanto etapa qualitativa, e do IVC como etapa quantitativa da pesquisa, justifica-se em razão

do objetivo proposto neste estudo e do estágio de construção do instrumento que esta pesquisa propôs alcançar. Diante da finalidade de atender ao objetivo geral do estudo, de propor um instrumento de mensuração da qualidade de dados prático, replicável e fundamentado numa estrutura conceitual sólida, com base no modelo de seis dimensões do DAMA-DMBOK, entende-se que a adoção de uma abordagem mista pode conferir ao instrumento um maior rigor e abrangência metodológica, com a soma da flexibilidade e profundidade da análise qualitativa, voltada para a avaliação conceitual de cada dimensão, ao caráter descritivo da etapa quantitativa, que permite a identificação de padrões de concordância entre especialistas.

Apesar disso, entende-se que uma etapa adicional de validação de conteúdo com base na estrutura interna, por meio de Análise Fatorial Exploratória (AFE), poderia enriquecer a análise, uma vez que a AFE poderia permitir a avaliação sobre em que medida a estrutura empírica do instrumento desenvolvido reflete a estrutura teórica do construto em análise. Apesar disso, ratifica-se que a aplicação da AFE iria requerer um número maior de respostas (Brown, 2006). Nesse sentido, essa recomendação constitui uma indicação para pesquisas futuras que visem a ampliar este estudo.

3.1 *Focus group*

A primeira etapa da pesquisa utilizou a técnica de *focus group* para explorar a percepção de especialistas sobre as seis dimensões da qualidade de dados propostas pelo modelo DAMA-DMBOK: completude, unicidade, atualidade, validade, precisão e consistência. Conforme Morgan (1996), essa técnica é particularmente eficaz para acessar representações sociais e compreender os processos de decisão dos participantes, especialmente quando esses processos envolvem negociação e construção de consenso, favorecendo a construção coletiva de conhecimento por meio da interação entre diferentes pontos de vista. A técnica *focus groups* é particularmente utilizada, na literatura recente, na fase inicial de desenvolvimento de um instrumento de medida, com vista a identificação de dimensões e criação de itens de escala (Hong; Hong; Cho, 2025; Kanthasamy *et al.*, 2025; Lin *et al.*, 2026; Yildiz *et al.*, 2025).

As sessões foram conduzidas por videoconferência, com grupos compostos por quatro a seis participantes. Cada rodada do encontro foi dedicada a uma das seis dimensões, e, ao final, os participantes formularam sentenças que expressassem, em linguagem prática, os conceitos discutidos. A mediação seguiu um roteiro semiestruturado, apresentado no Quadro 3 e no Apêndice A, permitindo equilíbrio entre foco temático e espontaneidade nas contribuições.

Quadro 3: Roteiro para Realização de *Focus Group*

Bloco	Texto lido aos participantes pelo moderador
1. Instruções Iniciais	Você foi convidado(a) a participar de um estudo cujo propósito é identificar e compreender as dimensões que determinam a qualidade dos dados. Dessa forma, busca-se construir um debate estruturado sobre as seis dimensões da qualidade dos dados e elaborar três perguntas para cada uma delas, de modo a sintetizar seus conceitos essenciais para fins de mensuração.
2. Apresentação do Moderador	Meu nome é Fabio Santana Reis e assumo como moderador do debate, que tem duração prevista de 60 minutos.
3. Consentimento dos participantes	Por favor, assegurem-se que estejam confortáveis para que não haja interrupções e sintam-se livres para expressar suas opiniões. Esta sessão está sendo gravada para análise futura, mas se alguém não estiver favorável, por favor solicito que apresente sua objeção. Não havendo objeção, vamos seguir.
4. Formato e duração da sessão (60 minutos)	Teremos 6 rodadas de discussão, cada uma delas focada numa das dimensões de mensuração da qualidade de dados com duração de 10 minutos. Cada rodada inicia-se com a apresentação da dimensão que será objeto de discussão entre vocês com uma breve descrição teórica. E, ao final da rodada, solicito a gentileza de formularem questões relacionadas à dimensão em foco visando obter uma métrica.
5. Em cada rodada (1 a 6), o moderador apresentará brevemente a dimensão em foco para que os participantes colaborem na formulação de sentenças relacionadas	1.Completude: Representa a proporção de dados armazenados em relação ao seu potencial total. Indica a ausência de lacunas que possam comprometer o uso da informação. Dados completos garantem que todas as informações necessárias estejam disponíveis para suportar decisões eficazes 2.Unicidade: Refere-se à ausência de duplicações nos dados. Cada entidade deve ser registrada uma única vez, de acordo com critérios de identificação definidos. Essa dimensão é essencial para evitar inconsistências e redundâncias em análises. 3.Atualidade: Mede o grau em que os dados refletem a realidade no momento em que são necessários. Dados atualizados permitem que decisões sejam tomadas com base em informações recentes e confiáveis. 4.Validade: Avalia se os dados estão em conformidade com regras, padrões e restrições predefinidos. Inclui a adequação ao formato, tipo e contexto de uso pretendido, garantindo que sejam utilizáveis e relevantes. 5.Precisão: Reflete o grau em que os dados representam corretamente a realidade ou um valor específico. Dados precisos devem estar livres de erros e serem fiéis à sua fonte original. 6.Consistência: Trata da ausência de discrepâncias entre diferentes representações do mesmo dado em diferentes sistemas ou bases. Dados consistentes garantem que a informação seja coerente em todos os contextos organizacionais.
6. Agradecimento	Gostaria de expressar meu sincero agradecimento a todos que participaram da sessão de <i>focus group</i> . Sua contribuição foi fundamental para o desenvolvimento desta pesquisa e para o enriquecimento da análise dos dados, trazendo perspectivas valiosas e experiências que certamente fortalecerão a qualidade dos resultados.
7. Encerramento	Informo que, por se tratar de uma pesquisa acadêmica, os resultados serão públicos e poderão ser consultados livremente após a conclusão do trabalho. Aos participantes que desejarem receber uma cópia da versão final da dissertação, peço que, por gentileza, se manifestem para que eu possa encaminhá-la diretamente. Será uma satisfação compartilhar o produto final desta etapa de estudos, que só foi possível graças à colaboração de todos vocês.

Fonte: Elaborado pelo autor.

Os participantes foram selecionados com base em critérios de experiência profissional e formação acadêmica. Todos possuíam, no mínimo, dois anos de atuação em áreas como governança e gestão de dados, estatística, tecnologia da informação ou administração, e exerciam funções diversas, incluindo engenheiros de dados, cientistas de dados, analistas de *business intelligence*, consultores e pesquisadores acadêmicos. A seleção foi realizada majoritariamente por meio da rede de contatos do pesquisador, o que, embora tenha facilitado o recrutamento, representa uma limitação quanto à diversidade de perfis. Ainda assim, buscou-se mitigar possíveis vieses por meio da composição heterogênea dos grupos, abrangendo diferentes níveis hierárquicos, formações e tipos de *expertise*.

A etapa foi fundamental para gerar material empírico representativo das dimensões investigadas, articulando os conceitos teóricos do modelo DAMA-DMBOK com experiências práticas relatadas pelos profissionais. As sentenças coletadas serviram de base para a construção do instrumento de mensuração da qualidade dos dados, sendo refinadas nas etapas posteriores por meio da seleção e validação por especialistas.

3.2 Seleção de sentenças

Concluída a etapa dos *focus groups*, as sentenças geradas foram submetidas à avaliação por meio de um questionário estruturado, com o objetivo de selecionar as afirmações mais representativas de cada uma das seis dimensões da qualidade dos dados. Os participantes desta fase foram convidados a julgar o grau de pertinência de cada sentença em relação à dimensão em que estava agrupada, utilizando uma escala de Likert de 1 (discordo totalmente) a 5 (concordo totalmente). O instrumento foi disponibilizado em formato digital, conforme apresentado no Quadro 4.

Quadro 4: Sentenças geradas nos *Focus Groups*

Dimensão	Sentenças
Atualidade	1. Os dados possuem uma periodicidade de atualização adequada ao seu propósito de uso. 2. A frequência de atualização dos dados (mensal, trimestral, etc.) está alinhada com as necessidades do negócio. 3. Os dados utilizados são classificados corretamente como históricos ou preditivos, conforme o contexto da análise. 4. Os dados são disponibilizados com antecedência suficiente para análise e tomada de decisão. 5. Quando os dados são usados para fins de venda, há um histórico confiável disponível. 6. Quando os dados são utilizados para novas aquisições, há previsões atualizadas disponíveis. 7. O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados. 8. A data de corte dos dados é coerente com o período de análise considerado. 9. O histórico de dados analisado é suficiente para gerar projeções confiáveis para o futuro. 10. Existe percepção clara da sazonalidade nas informações analisadas. 11. As comparações entre dados consideram períodos equivalentes (mesma faixa de tempo). 12. A sazonalidade é considerada adequadamente na análise dos dados. 13. O impacto potencial dos dados está diretamente relacionado à sua atualidade. 14. A recorrência de atualização dos dados é bem definida e mantida conforme o esperado. 15. Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados. 16. Mudanças no tempo ou no contexto são levadas em conta nas análises temporais dos dados. 17. Existe um responsável claramente definido pela atualização dos dados. 18. Potenciais interesses ocultos podem influenciar na manipulação ou atraso dos dados. 19. A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente. 20. A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações. 21. Há capacidade de restringir o volume de dados atualizados para focar em informações mais estratégicas (ex.: produtos mais lucrativos). 22. A omnicanalidade contribui para a atualização simultânea dos dados em diferentes canais. 23. Assim que os dados são carregados, um processo de validação de atualidade (freshness) é realizado. 24. Os dados são entregues no momento adequado para seu uso. 25. A versão mais atualizada dos dados é sempre utilizada nas análises. 26. Os dados foram entregues dentro do prazo acordado (SLA). 27. O provedor dos dados cumpriu os prazos combinados para entrega. 28. Em caso de falhas de sistema, há mecanismos de aviso sobre a indisponibilidade dos dados. 29. Após eventuais instabilidades, os dados gerados off-line são validados e integrados corretamente. 30. A periodicidade acordada entre quem fornece e quem consome os dados é respeitada. 31. Os dados são atualizados em tempo real, ou conforme o modelo estabelecido (ex.: D-1). 32. A sensibilidade à atualização dos dados é considerada nos processos críticos do negócio.
Validade	1. Os dados contábeis seguem normas regulatórias, como o padrão IFRS, quando aplicável. 2. Os dados utilizados estão em conformidade com normas e políticas estabelecidas. 3. Os dados seguem um padrão bem definido que facilita sua interpretação. 4. As informações apresentadas fazem sentido dentro do contexto analisado. 5. Os dados são utilizados de forma coerente com a finalidade para a qual foram gerados.

6. Há um padrão de referência que permite a comparação entre diferentes conjuntos de dados.
7. Os dados analisados são consistentes com os registros históricos anteriores.
8. Os dados estão no formato adequado para análise e processamento.
9. A fonte dos dados é claramente identificada.
10. A fonte dos dados é considerada confiável.
11. O período de referência dos dados está claramente definido.
12. O objetivo da análise está alinhado com a natureza dos dados utilizados.
13. Há clareza sobre o propósito final da análise com base nos dados disponíveis.
14. Existe correspondência entre os dados de entrada e os resultados esperados.
15. As diferentes fontes de dados se cruzam de forma coerente e validável.
16. O dado utilizado está alinhado com a intenção analítica definida.
17. A natureza do dado (ex.: faturamento com ou sem imposto) é explicitamente identificada.
18. As restrições de uso dos dados são conhecidas e respeitadas.
19. O volume de dados disponível é suficiente e apropriado para a análise pretendida.
20. As informações estão apresentadas de forma clara e compreensível.
21. Tenho clareza sobre o significado dos dados que estou utilizando.
22. Os dados passaram por auditoria externa ou revisão formal.
23. Outras áreas da empresa avaliam os dados de forma criteriosa e padronizada.
24. Os dados passaram por validação cruzada entre diferentes fontes ou sistemas.
25. Há esforços para evitar vieses analíticos e garantir que a informação seja válida.
26. Os dados são disponibilizados para consumo conforme esperado.
27. O formato dos dados é compatível com os sistemas de análise e consumo.
28. O processo de integração dos dados é simples e bem estruturado.
29. A simplicidade no tratamento dos dados contribui para maior qualidade e confiança.
30. Conflitos de interesse (ex.: bônus por performance) são considerados ao validar os dados.
31. Conflitos entre áreas da empresa são tratados para não comprometer a validade dos dados.
32. A importância dos dados varia conforme a área da empresa e isso é respeitado.
33. Regras de validação, como *multichecks*, são aplicadas para garantir a conformidade dos dados.
34. Os campos analisados contêm exatamente as informações necessárias.
35. Percentuais estão representados corretamente dentro do intervalo de 0 a 100.
36. O *profiling* dos dados foi realizado para entender padrões e variações.
37. Os limites definidos para os dados estão sendo respeitados.
38. O usuário está corretamente autenticado ao acessar ou consumir os dados.
39. É possível agrupar os dados por partes para facilitar sua validação.
40. Os dados estão limpos, sem erros ou ruídos que comprometam sua interpretação.
41. A criticidade dos dados foi avaliada para priorização e controle de qualidade.
42. O catálogo de dados foi consultado para confirmar definições e regras aplicáveis.
43. As informações estão mapeadas em seus respectivos fluxos e origens.
44. Migrações de sistemas não causaram problemas críticos de validade nos dados.

	45. O nível de certeza necessário sobre a validade dos dados foi previamente definido. 46. Quando problemas são identificados, há correção no sistema legado de origem. 47. Os dados presentes no sistema refletem com precisão as informações da origem. 48. O volume de dados analisado é proporcional à complexidade da demanda. 49. A análise dos dados considera as dores reais do negócio que precisam ser resolvidas. 50. Os dados representam com fidelidade o fenômeno ou processo que se deseja analisar. 51. Quando os dados são fornecidos por terceiros, a coleta é feita de forma sistemática e confiável.
Unicidade	1. A origem dos dados é claramente identificada e contribui para evitar redundância. 2. Os dados são rastreáveis, o que ajuda a garantir que não existam duplicações. 3. A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro. 4. O fato de a fonte ser qualificada reduz a possibilidade de múltiplas versões de um mesmo dado. 5. Dados com caráter objetivo tendem a ter menor risco de duplicação em diferentes sistemas. 6. A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos. 7. A competência da fonte produtora dos dados influencia na garantia de unicidade da informação. 8. A ocorrência dos mesmos dados em mais de um canal indica possível falha na gestão da unicidade. 9. A qualidade do canal de divulgação dos dados contribui para a percepção de que as informações são únicas. 10. Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única. 11. A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados. 12. É possível cruzar diferentes bases de dados sem gerar redundâncias. 13. As áreas da empresa que avaliam os dados consideram a unicidade como um critério essencial. 14. Existe colaboração entre as áreas da organização para evitar a duplicação de dados. 15. O uso de diversos sistemas dificulta a manutenção da unicidade dos dados. 16. A coleta de dados é feita diretamente da fonte apropriada, reduzindo o risco de duplicações. 17. A governança de dados contribui para a garantia da unicidade dos registros. 18. O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação. 19. Existe uma estratégia clara para evitar a duplicação de dados nas diferentes fontes. 20. A captura de dados é realizada a partir de uma única fonte confiável. 21. O objetivo da análise de dados influencia na forma como a unicidade dos dados é tratada. 22. A identificação de desvios em dados duplicados é uma prática comum nas análises da organização. 23. A simplificação da análise de dados pode comprometer a verificação da unicidade. 24. Há clareza sobre quais pontos críticos devem ser analisados para garantir dados não redundantes. 25. A consideração de vieses analíticos (ex.: viés de confirmação) pode impactar a percepção sobre a unicidade dos dados. 26. O viés do analista pode interferir na verificação da unicidade dos dados. 27. Aplicações distintas apresentam os mesmos dados, e é definido qual sistema possui o dado de referência. 28. Cada registro possui um identificador único, como uma chave primária. 29. Existe um processo de reconciliação de dados para eliminar inconsistências e duplicidades. 30. Quando um erro é detectado, o script de geração de dados é ajustado para evitar repetições futuras. 31. Em processos de unificação de bases de dados, existe um procedimento formal definido. 32. A verificação dos processos de negócio é realizada ao unificar bases de dados para garantir unicidade.

Precisão	1. É possível identificar tendências reais nos dados analisados, sem distorções. 2. Os dados não apresentam anomalias quando comparados com a base histórica. 3. A fonte fornecedora dos dados é confiável o suficiente para garantir a precisão. 4. O formato dos dados está adequado para evitar erros de interpretação ou leitura. 5. Os dados não estão corrompidos nem apresentam falhas técnicas. 6. A origem dos dados (única ou múltiplas fontes) não compromete sua precisão. 7. Os dados estão perfeitamente alinhados com o objetivo da análise. 8. Os dados são interpretados dentro do contexto correto de uso. 9. Foi realizada uma análise de similaridade que contribui para validar a precisão dos dados. 10. As fontes que disponibilizam os dados estão identificadas e são conhecidas. 11. Mesmo quando existe um padrão de imprecisão, ele pode ser reconhecido e compensado adequadamente. 12. Os dados analisados representam de forma fiel a realidade do fenômeno observado. 13. Os dados refletem corretamente o desejo ou comportamento real do cliente. 14. Pode haver viés na entrada dos dados que compromete sua precisão. 15. A imprecisão dos dados pode ser causada por vieses humanos. 16. Os dados originados de pesquisas de mercado são confiáveis quanto à precisão. 17. A metodologia de coleta foi adequada para garantir dados precisos. 18. O grau de objetividade dos dados é suficiente para minimizar distorções. 19. Os critérios usados na análise dos dados são claros e favorecem a precisão na tomada de decisão. 20. O viés nem sempre é negativo; o conhecimento e a experiência podem contribuir para maior precisão. 21. Pode haver viés tanto na geração quanto na interpretação dos dados. 22. Os dados são considerados vivos (orgânicos), e sua precisão pode evoluir ao longo do tempo. 23. As pessoas responsáveis pela geração ou análise dos dados possuem qualificação técnica adequada. 24. Outliers foram identificados e tratados corretamente para evitar distorções. 25. Os critérios utilizados para a obtenção dos dados foram rigorosos e bem definidos. 26. Referências de mercado foram utilizadas para validar a precisão dos dados internos. 27. Os dados passaram por uma validação inicial (input) que garante sua fidelidade à realidade. 28. Os dados de destino (target) refletem com precisão os dados de origem (<i>source</i>). 29. Os processos complementares foram validados e não comprometem a precisão dos dados. 30. O flat file utilizado está correto e livre de erros que afetem a precisão. 31. É conhecida a origem dos dados — se primária (interna) ou secundária (auditorias externas). 32. O dado reflete com alto grau de precisão a realidade a partir da fonte original.
Compleitude	1. A base de dados não apresenta registros ausentes. 2. Os campos dos registros estão devidamente preenchidos. 3. O processo de entrada (input) dos dados foi seguido corretamente em todas as etapas. 4. Os campos-chave identificadores (<i>Key Data Elements</i>) estão preenchidos em todos os registros. 5. A análise dos dados pode ser compartilhada com os <i>stakeholders</i> de forma clara, inclusive apontando onde há dados ausentes por grupo (clientes, fornecedores, etc.). 6. Para empresas que atuam em diferentes mercados (ex.: por país ou região), é possível verificar se todos os mercados estão representados na base de dados.

	7. O percentual de presença de elementos por grupo (ex.: países, categorias, regiões) é conhecido e controlado. 8. A base de dados contempla registros de todas as marcas comercializadas pela organização. 9. Todos os dados relevantes estão disponíveis para análise e decisão. 10. Sob a perspectiva da jornada do cliente, existem dados representando todas as etapas dessa jornada. 11. Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor. 12. Os dados disponíveis são suficientes para apoiar decisões com base em um conjunto completo de informações. 13. Há clareza sobre quais dados são necessários para se tomar decisões baseadas em informação completa. 14. Existe preocupação com o excesso de dados que, apesar de completos, podem prejudicar a objetividade da análise.
Consistência	1. Os dados são compatíveis com os padrões do mercado quando comparados entre pares equivalentes. 2. A variação dos dados em relação à base histórica é controlada e justificada. 3. Os dados foram avaliados por mais de um especialista, contribuindo para sua consistência. 4. A opinião técnica dos especialistas reforça a confiança na consistência dos dados analisados. 5. As opiniões dos especialistas são ponderadas com critérios claros e contribuem para validação consistente dos dados. 6. Há diversidade de fontes de dados sem comprometer a uniformidade das informações. 7. A quantidade de temas e categorias que originam os dados é conhecida e monitorada. 8. O sistema utilizado disponibiliza os dados em formato funcional e utilizável, sem distorções. 9. Dados provenientes de sistemas legados não geram divergências nos resultados atuais. 10. Os indicadores de performance atuais estão alinhados com os sistemas que os alimentam. 11. Foi realizada uma comparação entre dados de categorias semelhantes (ex.: comparação entre produtos similares), garantindo coerência. 12. Existe um processo estruturado de reconciliação entre diferentes fontes ou bases de dados. 13. Pontos críticos são considerados no processo de reconciliação de dados. 14. Sempre que os dados são transferidos entre sistemas ou bases, é verificada a integridade das informações. 15. Não houve falhas significativas no processo de migração ou movimentação dos dados entre sistemas. 16. Quando ocorrem anomalias nos dados, há um processo de recuperação eficaz. 17. Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema. 18. Existe um plano de consistência para monitorar a movimentação de dados entre bases, como amostragem periódica. 19. A validação cruzada é parte integrante do processo de verificação da consistência dos dados. 20. As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados. 21. Diferenças nas regras de arredondamento entre sistemas não comprometem somas ou análises agregadas.

Fonte: Elaborado pelo autor.

Participaram dessa etapa profissionais e acadêmicos distintos dos que integraram os grupos focais, garantindo a ampliação das perspectivas avaliativas. Embora não tenham sido caracterizados em profundidade como na etapa anterior, o questionário incluiu campos para coleta de informações como formação, cargo e tempo de experiência profissional. O perfil dos respondentes foi semelhante ao dos especialistas dos *focus groups*, contemplando indivíduos com experiência em gestão, governança e análise de dados. A seleção também ocorreu por meio da rede de contatos do autor, mantendo a limitação já reconhecida quanto ao potencial viés de seleção, porém com mitigação parcial pela diversidade de funções e áreas de atuação dos participantes. Os dados obtidos permitiram ranquear as sentenças com base no número de avaliações máximas (nota 5), complementadas, quando necessário, pelas notas 4, de forma a selecionar os itens mais consensuais para a etapa subsequente de validação por juízes especialistas.

Quadro 5: Perfil dos participantes da etapa de seleção de sentenças

Formação Acadêmica	Cargo	Experiência Profissional (Anos)
Administração	Coordenadora Acadêmica	18
Ciência de Dados	Arquiteto de dados	02
Ciência da computação	Analista de sistemas	18
Engenheiro de Sistemas	Consultor BI	18
Engenharia elétrica	Engenheiro de dados	07
Bacharel em Sistemas de informação	Especialista em <i>Business Intelligence</i>	14
Tecnologia da informação	Especialista de sistemas	15
Informática para Negócios	Analista Planejamento Comercial Sênior	15
Bacharel em Ciências da Computação	<i>Head de Customer Success</i>	25
Ciências da computação	Engenheiro de dados	20
Engenharia de produção	Engenheiro de dados	09
Bacharel em Enfermagem	Analista de dados	02
Ciência da Computação	Engenheiro de dados	20
Matemática	Engenheiro de dados	30
Marketing & Tecnologia	<i>Business Intelligence Specialist</i>	05

Fonte: Elaborador pelo autor.

3.3 Validação de conteúdo

A validação de conteúdo representou uma etapa essencial nesta pesquisa, tendo como objetivo assegurar que os itens do questionário realmente refletiam, de forma clara e precisa, as dimensões teóricas propostas para mensuração. Esse processo foi conduzido com a participação de avaliadores independentes, visando reduzir possíveis vieses do pesquisador e aumentar a confiabilidade do instrumento.

De acordo com Lynn (1986), essa validação seguiu um procedimento sistemático, com o apoio de técnicas quantitativas, sendo o Índice de Validade de Conteúdo (IVC) uma das mais utilizadas. Esse índice permitiu verificar o grau de concordância entre os juízes quanto à correspondência de cada item às dimensões teóricas do modelo, funcionando como um termômetro da clareza e aderência das sentenças ao constructo que pretendiam representar. A aplicação do IVC é amplamente reconhecida na literatura psicométrica por sua simplicidade e efetividade na construção e aprimoramento de escalas e questionários. Neste estudo, o cálculo do IVC considerou a proporção de juízes que associaram corretamente (ou de forma consensual) um item à dimensão teórica correspondente, conforme definido no modelo conceitual adotado. Esse critério contribuiu de forma decisiva para identificar os itens mais consistentes e sinalizar aqueles que precisavam ser revistos em ciclos posteriores de refinamento. O IVC foi calculado utilizando a Equação 1.

Equação 1. Índice de Validade de Conteúdo (IVC)

$$IVC = \frac{\text{Número de classificações da questão dentro do constructo esperado}}{\text{Número total de juízes}}$$

Fonte: retirado de Hernandez-Nieto (2022).

Segundo Hernandez-Nieto (2022), valores de IVC iguais ou superiores a 0,80 indicam forte evidência de validade de conteúdo. Itens com índice inferior a esse valor poderiam ser revisados, reformulados ou, eventualmente, excluídos do instrumento. Esse processo garantiu uma triangulação metodológica entre a percepção coletiva dos especialistas (via *focus group*) e a validação técnica individual (via análise de juízes), conferindo maior robustez ao instrumento proposto. A aplicação da técnica de análise de juízes, combinada com a utilização do IVC, proporcionou uma base sólida para validar o conteúdo do instrumento de pesquisa. Essa abordagem sistemática garantiu que o questionário final fosse claro, relevante e livre de ambiguidades, aumentando a confiabilidade das respostas e robustez dos resultados da pesquisa.

Ao final das duas etapas, foi possível obter uma versão preliminar do instrumento de mensuração da qualidade dos dados, com evidências empíricas de validade de conteúdo e suporte teórico adequado. Esse instrumento poderá ser aplicado futuramente em estudos quantitativos, possibilitando análises estatísticas, como validação fatorial exploratória e análise de confiabilidade. Nesta etapa, 12 novos especialistas da área de dados, que não participaram das etapas anteriores, foram convidados para atuar como juízes independentes. O perfil desses juízes incluía experiência em gestão, governança e qualidade de dados, garantindo que possuíam o conhecimento técnico necessário para avaliar a clareza, relevância e abrangência das questões em relação aos objetivos da pesquisa. A seleção desses juízes, assim como nas etapas anteriores, beneficiou-se da rede de contatos do autor, o que, embora prático, impôs a limitação de potencial viés de seleção e a necessidade de considerar a generalização dos resultados com cautela. O perfil acadêmico, profissional e a experiência de cada participante são detalhados no Quadro 6. Em seguida, apresentam-se a análise e discussão dos resultados da pesquisa.

Quadro 6: Perfil dos participantes da etapa de análise de juízes

Formação Acadêmica	Cargo	Experiência Profissional (Anos)
Bacharel Matemática Aplicada	Analista Programador	27
Mestrado profissional	Diretor de dados	25
Superior Completo	Coordenador de TI	25
Matemática	Diretora de Dados e Análises	20
Graduação em Contabilidade e Matemática, pós-graduação em Estatística e MBA em Big Data	<i>Data & Analytics Manager</i>	24
Engenheiro Mecânico / Mestrando em Gestão de Negócios	Consultor	29
Engenharia Elétrica	Cientista de Dados	20
Bacharel em Física pelo IFUSP	Empresário	30
Doutorado em Letras	Professora	54
Tecnólogo em Processamento de Dados	Especialista de Sistemas	25
Engenharia de Controle e Automação	Gerente Sênior de TI	12
MBA	Gerente de Projetos	28

Fonte: Elaborado pelo autor.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

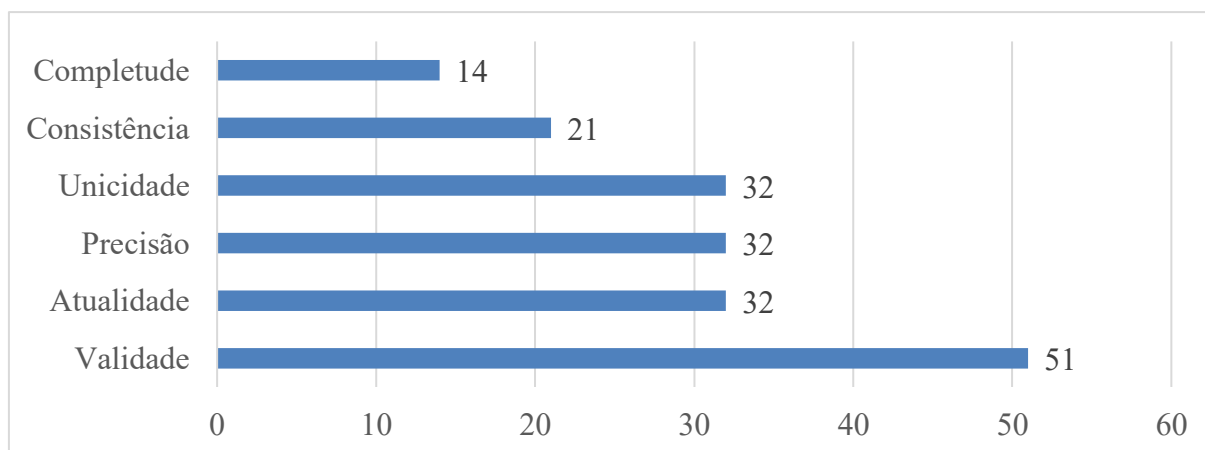
Este capítulo, inicialmente, apresenta as etapas até a construção do instrumento final, objeto de desenvolvimento desta pesquisa. Em sequência, uma análise do instrumento desenvolvido com a literatura sobre o tema de mensuração da qualidade de dados é apresentada. As análises buscam não apenas apresentar os dados, mas também interpretar seu significado à luz da literatura sobre qualidade de dados e do modelo conceitual adotado, que se baseia nas dimensões do DAMA-DMBOK.

4.1 Etapa 1: *focus group*

As sessões *de focus group* foram organizadas em rodadas temáticas, cada uma dedicada a uma das seis dimensões da qualidade de dados: Completude, Unicidade, Atualidade, Validade, Precisão e Consistência. Ao final de cada rodada, os participantes foram convidados a propor sentenças que traduzissem, com base em suas vivências profissionais, o significado prático da dimensão discutida. Essas contribuições tinham como propósito servir de matéria-prima para a construção do instrumento de mensuração proposto nesta pesquisa.

A dinâmica entre os participantes foi caracterizada por trocas de informações e reflexões acerca da forma como cada dimensão se manifesta no cotidiano organizacional. As distintas experiências relatadas contribuíram para que as sentenças elaboradas contemplassem tanto os conceitos teóricos quanto aspectos práticos vivenciados por profissionais que lidam com dados nas organizações. As sentenças elaboradas ao longo dessas sessões estão reunidas nos Apêndices B, C e D desta dissertação. Para ilustrar a diversidade e profundidade do material gerado, o Gráfico 1 apresenta a quantidade de sentenças geradas para cada dimensão.

Gráfico 1: Quantidade de sentenças geradas nos *focus groups* por dimensão



Fonte: Dados da pesquisa.

A ocorrência de "Não houve tempo" para a dimensão "Completeness" no *Focus Group* 1 (Apêndice B) foi uma limitação pontual, indicando que o tempo alocado inicialmente para a discussão e formulação das frases para essa dimensão pode ter sido insuficiente ou que a dinâmica do grupo específico não permitiu a profundidade esperada naquele momento. No entanto, essa lacuna foi compensada pelas contribuições robustas dos *Focus Groups* 2 e 3 (Apêndices C e D), que geraram uma quantidade significativa de sentenças para a dimensão Completeness. Essa variabilidade ressalta a importância de múltiplas sessões de *focus group* para garantir a abrangência das perspectivas e a mitigação de falhas isoladas.

Um achado instigante do último *focus group* foi a sugestão de um participante de que qualquer análise de dados deveria sempre começar pela dimensão Atualidade. A justificativa apresentada foi que, se os dados não estiverem atualizados, não faria sentido analisar as demais dimensões, uma vez que informações defasadas levariam a decisões equivocadas, independentemente de sua completeness, unicidade ou precisão.

Outro achado notável ocorreu no primeiro *focus group*, durante a discussão sobre a dimensão Precisão. Um participante identificou que, mesmo dentro da imprecisão, é possível identificar a precisão desde que se conheça o padrão do dado analisado. Como exemplo, ele citou a prática de um concorrente em seu ramo de atividade hoteleira que consistentemente divulga uma quantidade de reservas 20% acima da disponibilidade real. Desse modo, o dado divulgado pelo concorrente, embora impreciso em sua face, torna-se previsivelmente impreciso, permitindo a dedução do dado preciso real ao desconsiderar o acréscimo usualmente intrínseco. Essa observação levanta uma perspectiva interessante sobre a "precisão relativa" ou "precisão ajustada", sugerindo que o conhecimento do contexto e do "viés" inerente à fonte pode ser uma forma de mitigar a imprecisão e ainda extrair valor do dado.

4.2 Etapa 2: seleção de sentenças

Com base nas notas atribuídas no questionário da Etapa 2 (Apêndice E), as sentenças foram ranqueadas e selecionadas para a próxima fase. Os critérios de seleção foram: primeiro, a maior quantidade de notas 5 atribuídas à sentença; e segundo, caso o critério anterior não atingisse a quantidade desejada, a maior quantidade de notas 4. O objetivo era que ao menos dez por cento das sentenças de cada sessão compusessem a etapa de análise de juízes. Para exemplificar o processo, o Tabela 1 apresenta uma ilustração do ranqueamento de sentenças selecionadas para a etapa seguinte agrupadas por dimensão, considerando as 31 sentenças listadas no Apêndice E.

Tabela 1: Ranqueamento das sentenças selecionadas nesta etapa

	Quantidade de Notas 5	Quantidade de Notas 4	Ranking
Atualidade			
19. A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente.	10	3	1
7. O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados.	8	6	2
11. As comparações entre dados consideram períodos equivalentes (mesma faixa de tempo).	8	4	3
20. A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações.	8	4	4
15. Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados.	8	3	5
Compleitude			
6. Para empresas que atuam em diferentes mercados (ex.: por país ou região), é possível verificar se todos os mercados estão representados na base de dados.	7	7	1
8. A base de dados contempla registros de todas as marcas comercializadas pela organização.	7	5	2
4. Os campos-chave identificadores (Key Data Elements) estão preenchidos em todos os registros.	7	3	3
11. Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor.	7	3	4
Consistência			
14. Sempre que os dados são transferidos entre sistemas ou bases, é verificada a integridade das informações.	7	6	1
17. Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema.	6	8	2
20. As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados.	6	6	3
Precisão			
5. Os dados não estão corrompidos nem apresentam falhas técnicas.	8	6	1
4. O formato dos dados está adequado para evitar erros de interpretação ou leitura.	8	2	2
22. Os dados são considerados vivos (orgânicos), e sua precisão pode evoluir ao longo do tempo.	7	7	3
21. Pode haver viés tanto na geração quanto na interpretação dos dados.	7	6	4
28. Os dados de destino (target) refletem com precisão os dados de origem (source).	7	6	5
7. Os dados estão perfeitamente alinhados com o objetivo da análise.	7	4	6

(continua)

(continuação)

Unicidade			
18. O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação.	10	5	1
11. A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados.	9	4	2
1. A origem dos dados é claramente identificada e contribui para evitar redundância.	9	3	3
7. A competência da fonte produtora dos dados influencia na garantia de unicidade da informação.	8	6	4
10. Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única.	8	6	5
3. A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro.	8	4	6
6. A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos.	8	3	7
Validade-			
35. Percentuais estão representados corretamente dentro do intervalo de 0 a 100.	9	5	1
38. O usuário está corretamente autenticado ao acessar ou consumir os dados.	8	7	2
26. Os dados são disponibilizados para consumo conforme esperado.	8	6	3
18. As restrições de uso dos dados são conhecidas e respeitadas.	8	5	4
10. A fonte dos dados é considerada confiável.	7	8	5
11. O período de referência dos dados está claramente definido.	7	8	6

Fonte: Dados da pesquisa (2025).

Em adição, dados de estatística descritiva foram obtidos para cada uma das 31 sentenças, conforme apresentado na Tabela 2.

Tabela 2: Estatística Descritiva

Sentença	n	min	max	median	iqr	mean	sd	se	ci
O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados.	16	3	5	4,5	1	4,375	0,719	0,18	0,383
As comparações entre dados consideram períodos equivalentes (mesma faixa de tempo).	16	2	5	4,5	1,25	4,188	0,981	0,245	0,523
Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados.	16	2	5	4,5	2	4,062	1,124	0,281	0,599
A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente.	16	2	5	5	1	4,375	0,957	0,239	0,51
A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações.	16	3	5	4,5	1,25	4,25	0,856	0,214	0,456
A fonte dos dados é considerada confiável.	16	2	5	4	1	4,312	0,793	0,198	0,423
O período de referência dos dados está claramente definido.	16	3	5	4	1	4,375	0,619	0,155	0,33
As restrições de uso dos dados são conhecidas e respeitadas.	16	2	5	4,5	1	4,25	0,931	0,233	0,496
Os dados são disponibilizados para consumo conforme esperado.	16	3	5	4,5	1	4,375	0,719	0,18	0,383
Percentuais estão representados corretamente dentro do intervalo de 0 a 100.	16	2	5	5	1	4,375	0,885	0,221	0,472
O usuário está corretamente autenticado ao acessar ou consumir os dados.	16	3	5	4,5	1	4,438	0,629	0,157	0,335
A origem dos dados é claramente identificada e contribui para evitar redundância.	16	2	5	5	1,25	4,25	1	0,25	0,533
A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro.	16	2	5	4,5	1,25	4,188	0,981	0,245	0,523
A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos.	16	1	5	4,5	2	4,062	1,181	0,295	0,63
A competência da fonte produtora dos dados influencia na garantia de unicidade da informação.	16	2	5	4,5	1	4,312	0,873	0,218	0,465
Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única.	16	3	5	4,5	1	4,375	0,719	0,18	0,383
A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados.	16	2	5	5	1	4,312	0,946	0,237	0,504

O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação.	16	3	5	5	1	4,562	0,629	0,157	0,335
O formato dos dados está adequado para evitar erros de interpretação ou leitura.	16	2	5	4,5	2	3,938	1,237	0,309	0,659
Os dados não estão corrompidos nem apresentam falhas técnicas.	16	3	5	4,5	1	4,375	0,719	0,18	0,383
Os dados estão perfeitamente alinhados com o objetivo da análise.	16	3	5	4	2	4,125	0,885	0,221	0,472
Pode haver viés tanto na geração quanto na interpretação dos dados.	16	3	5	4	1	4,25	0,775	0,194	0,413
Os dados são considerados vivos (orgânicos), e sua precisão pode evoluir ao longo do tempo.	16	3	5	4	1	4,312	0,704	0,176	0,375
Os dados de destino (target) refletem com precisão os dados de origem (source).	16	3	5	4	1	4,25	0,775	0,194	0,413
Os campos-chave identificadores (Key Data Elements) estão preenchidos em todos os registros.	16	2	5	4	2	4	1,033	0,258	0,55
Para empresas que atuam em diferentes mercados (ex.: por país ou região), é possível verificar se todos os mercados estão representados na base de dados.	16	3	5	4	1	4,312	0,704	0,176	0,375
A base de dados contempla registros de todas as marcas comercializadas pela organização.	15	2	5	4	1	4,2	0,941	0,243	0,521
Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor.	16	2	5	4	2,25	3,812	1,276	0,319	0,68
Sempre que os dados são transferidos entre sistemas ou bases, é verificada a integridade das informações.	16	3	5	4	1	4,25	0,775	0,194	0,413
Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema.	16	1	5	4	1	4,125	1,025	0,256	0,546
As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados.	16	2	5	4	1,25	4	1,033	0,258	0,55

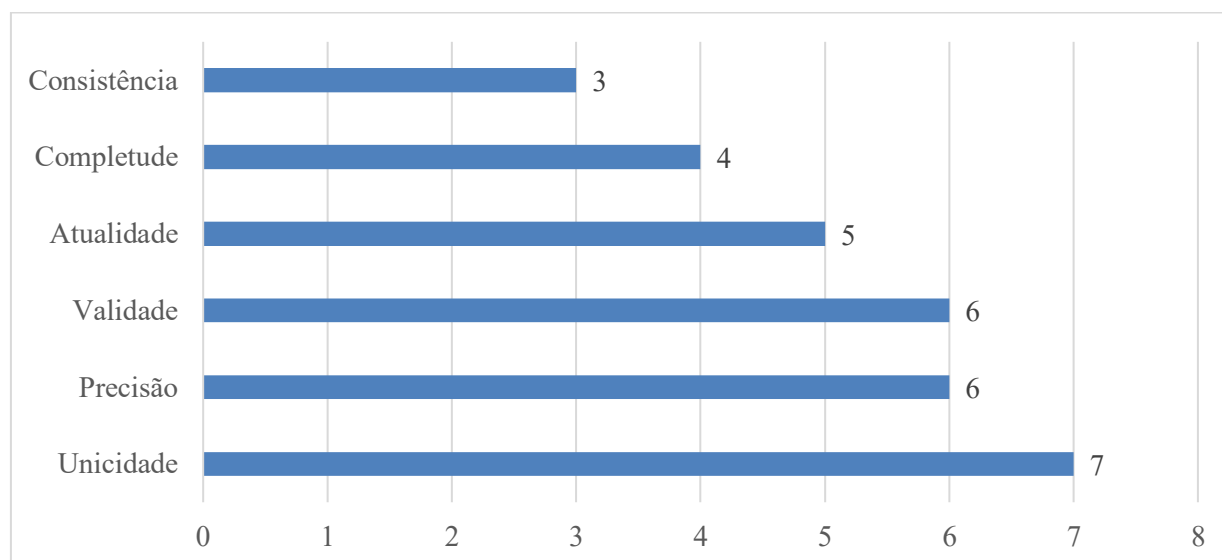
Nota: *n* = número de observações; *min* = valor mínimo; *max* = valor máximo; *median* = mediana; *iqr* = intervalo interquartil; *mean* = média aritmética; *sd* = desvio padrão; *se* = erro padrão da média; *ci* = amplitude do intervalo de confiança (95%).

Fonte: *Software R* com base nos dados da pesquisa.

Observa-se que todas as sentenças, com exceção de uma, apresentam média e mediana acima do valor 4, o que denota uma pontuação satisfatória para as sentenças selecionadas e reforça a qualidade da seleção realizada. As médias de pontuação, para todas as sentenças, variam entre 3,812 e 4,562, o que pode indicar uma tendência geral de concordância ou percepção positiva.

Este processo garantiu que apenas as sentenças com maior grau de concordância e pertinência percebida fossem encaminhadas para a validação de conteúdo por juízes, otimizando o tempo dos especialistas na etapa seguinte e aumentando a probabilidade de um instrumento final robusto. No Gráfico 2 observa-se a quantidade de sentenças selecionadas nesta etapa para cada dimensão de qualidade de dados.

Gráfico 2: Quantidade de sentenças filtradas



Fonte: elaborado pelo autor.

4.3 Etapa 3: validação de conteúdo

As sentenças selecionadas para a etapa de validação de conteúdo foram apresentadas aos juízes de forma aleatória, sem qualquer identificação prévia da dimensão teórica à qual pertenciam, assegurando a imparcialidade da avaliação. Conforme apresentado no Apêndice F, cada juiz foi instruído a associar cada item a uma das seis dimensões da qualidade de dados, conforme sua compreensão conceitual, ou indicar “Não sei” caso não se sentisse apto a realizar a associação. Com base nas respostas, foi calculado o Índice de Validade de Conteúdo (IVC) de cada sentença, conforme metodologia proposta por Hernandez-Nieto (2022), que consiste

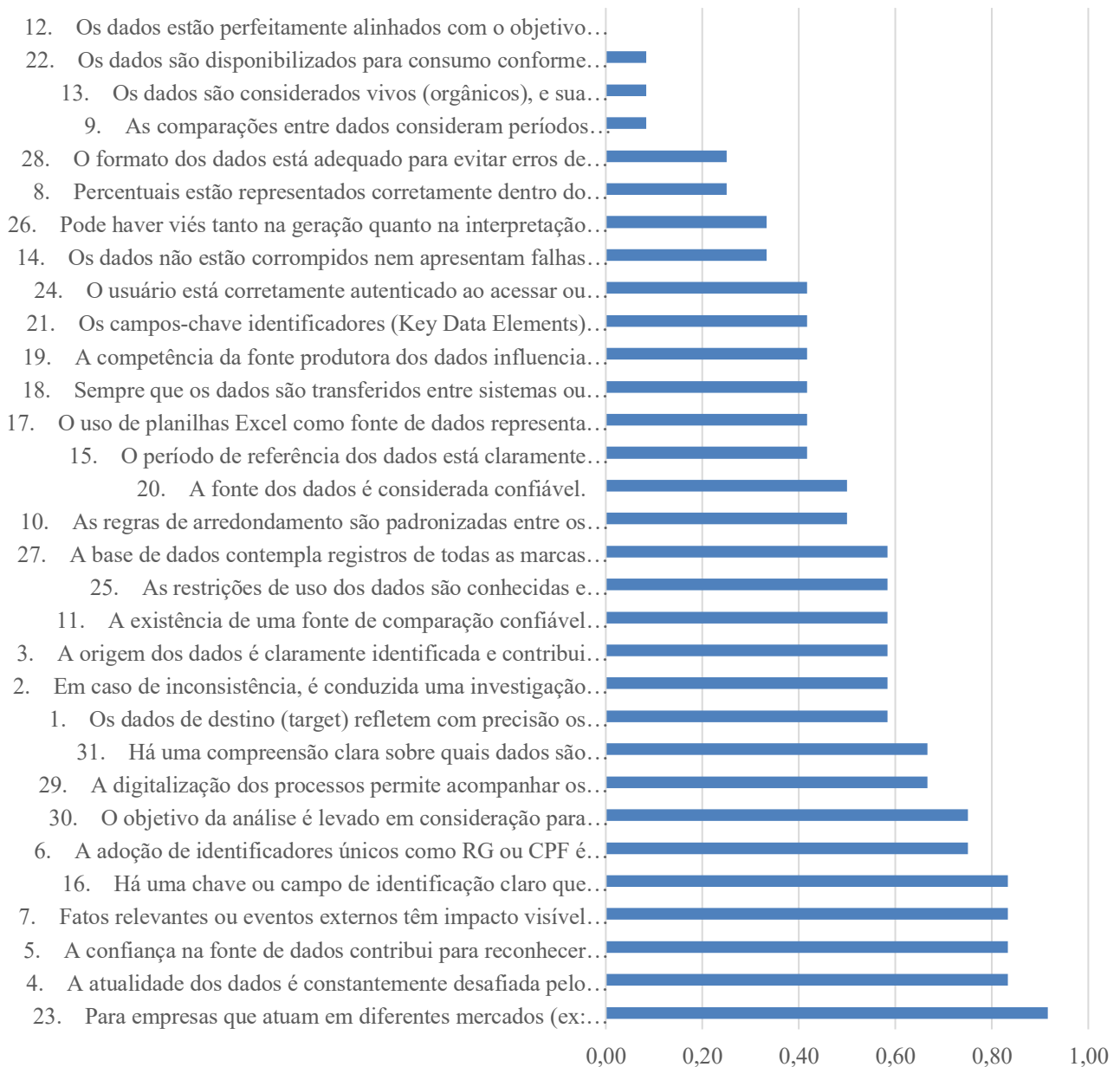
na proporção de juízes que indicaram a mesma dimensão teórica originalmente atribuída à sentença.

Embora a literatura indique o valor de 0,80 como referência para forte evidência de validade de conteúdo, este critério é geralmente aplicado em contextos de validação final de instrumentos com escopo consolidado. No presente estudo, por se tratar de uma proposta em estágio inicial de construção de um instrumento exploratório, optou-se por adotar um ponto de corte de IVC igual ou superior a 0,50. Tal decisão tem respaldo na literatura psicométrica, que reconhece a flexibilidade de critérios em fases exploratórias preliminares de validação, desde que devidamente justificada (Lynn, 1986; Zamanzadeh *et al.*, 2015).

O uso desse limiar mais moderado permite identificar sentenças que, embora ainda demandem ajustes, apresentam um nível razoável de reconhecimento entre especialistas quanto à sua vinculação conceitual. Essa estratégia possibilita o aproveitamento de um número maior de itens potencialmente relevantes, que podem ser refinados em ciclos posteriores, contribuindo para o aprimoramento contínuo do instrumento e garantindo maior representatividade das dimensões do modelo DAMA-DMBOK.

Além disso, evita a exclusão precoce de sentenças que possuem valor interpretativo e aplicabilidade prática, mas que podem ter sido impactadas por variações de linguagem ou ambiguidades pontuais passíveis de revisão. A análise dos resultados da Etapa 3 revelou que, das 31 sentenças avaliadas por 12 juízes, apenas 17 atingiram um Índice de Validade de Conteúdo (IVC) igual ou superior a 0,50, conforme observado no Gráfico 3.

Gráfico 3: Sentenças avaliadas pelos juízes



Fonte: Dados da pesquisa (2025).

Para maior detalhamento, a Tabela 3 contém os dados referentes ao Gráfico 3.

Tabela 3: Detalhamento dos dados do Gráfico

Dimensão Referência	Sentenças	Matches (0 a 12)	%Matches	Produto Tecnológico
Precisão	1.Os dados de destino (target) refletem com precisão os dados de origem (source).	7	58%	Sim
Consistência	2.Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema.	7	58%	Sim
Unicidade	3. A origem dos dados é claramente identificada e contribui para evitar redundância.	7	58%	Sim
Atualidade	4.A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações.	10	83%	Sim
Unicidade	5.A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro.	10	83%	Sim
Unicidade	6.A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados.	9	75%	Sim
Atualidade	7.Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados.	10	83%	Sim
Validade	8.Percentuais estão representados corretamente dentro do intervalo de 0 a 100.	3	25%	Não
Atualidade	9.As comparações entre dados consideram períodos equivalentes (mesma faixa de tempo).	1	8%	Não
Consistência	10.As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados.	6	50%	Sim
Unicidade	11.A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos.	7	58%	Sim
Precisão	12.Os dados estão perfeitamente alinhados com o objetivo da análise.	0	0%	Não
Precisão	13.Os dados são considerados vivos (orgânicos), e sua precisão pode evoluir ao longo do tempo.	1	8%	Não
Precisão	14.Os dados não estão corrompidos nem apresentam falhas técnicas.	4	33%	Não
Validade	15.O período de referência dos dados está claramente definido.	5	42%	Não
Unicidade	16.Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única.	10	83%	Sim
Unicidade	17.O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação.	5	42%	Não
Consistência	18. Sempre que os dados são transferidos entre sistemas ou bases, é verificada a integridade das informações.	5	42%	Não
Unicidade	19.A competência da fonte produtora dos dados influencia na garantia de unicidade da informação.	5	42%	Não
Validade	20.A fonte dos dados é considerada confiável.	6	50%	Sim
Completude	21.Os campos-chave identificadores (Key Data Elements) estão preenchidos em todos os registros.	5	42%	Não
Validade	22.Os dados são disponibilizados para consumo conforme esperado.	1	8%	Não
Completude	23.Para empresas que atuam em diferentes mercados (ex.: por país ou região), é possível verificar se todos os mercados estão representados na base de dados.	11	92%	Sim
Validade	24.O usuário está corretamente autenticado ao acessar ou consumir os dados.	5	42%	Não
Validade	25.As restrições de uso dos dados são conhecidas e respeitadas.	7	58%	Sim
Precisão	26.Pode haver viés tanto na geração quanto na interpretação dos dados.	4	33%	Não
Completude	27.A base de dados contempla registros de todas as marcas comercializadas pela organização.	7	58%	Sim
Precisão	28.O formato dos dados está adequado para evitar erros de interpretação ou leitura.	3	25%	Não
Atualidade	29.A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente.	8	67%	Sim
Atualidade	30.O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados.	9	75%	Sim
Completude	31.Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor.	8	67%	Sim

Fonte: dados da pesquisa.

Para ilustrar o resultado da análise de juízes, o Quadro 7 apresenta as sentenças agrupadas por dimensão de referência com destaque para aquelas cujo IVC foi superior a 0,5, assinaladas na última coluna com “Sim”.

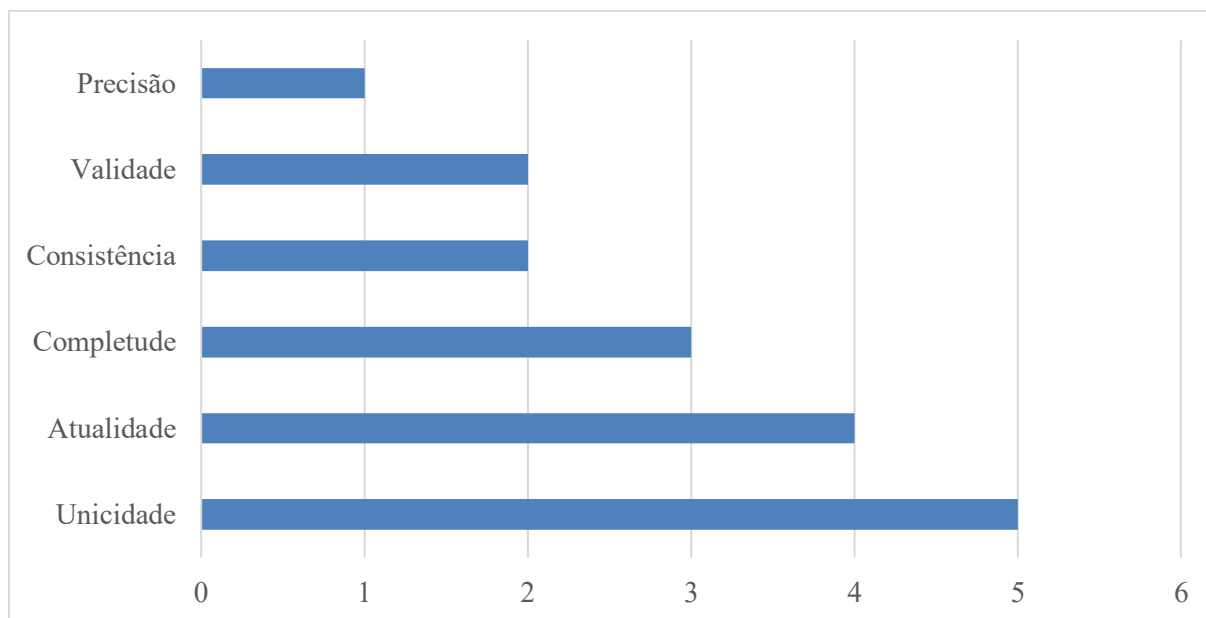
Quadro 7: Sentenças selecionadas pelos juízes

Dimensão	Sentenças	IVC $\geq 0,5$
Atualidade	4. A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações.	Sim
	7. Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados.	Sim
	9. As comparações entre dados consideram períodos equivalentes (mesma faixa de tempo).	Não
	29. A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente.	Sim
	30. O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados.	Sim
Compleitude	21. Os campos-chave identificadores (<i>Key Data Elements</i>) estão preenchidos em todos os registros.	Não
	23. Para empresas que atuam em diferentes mercados (ex.: por país ou região), é possível verificar se todos os mercados estão representados na base de dados.	Sim
	27. A base de dados contempla registros de todas as marcas comercializadas pela organização.	Sim
	31. Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor.	Sim
Consistência	2. Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema.	Sim
	10. As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados.	Sim
	18. Sempre que os dados são transferidos entre sistemas ou bases, é verificada a integridade das informações.	Não
Precisão	1. Os dados de destino (<i>target</i>) refletem com precisão os dados de origem (<i>source</i>).	Sim
	12. Os dados estão perfeitamente alinhados com o objetivo da análise.	Não
	13. Os dados são considerados vivos (orgânicos), e sua precisão pode evoluir ao longo do tempo.	Não
	14. Os dados não estão corrompidos nem apresentam falhas técnicas.	Não
	26. Pode haver viés tanto na geração quanto na interpretação dos dados.	Não
	28. O formato dos dados está adequado para evitar erros de interpretação ou leitura.	Não
Unicidade	3. A origem dos dados é claramente identificada e contribui para evitar redundância.	Sim
	5. A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro.	Sim
	6. A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados.	Sim
	11. A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos.	Sim
	16. Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única.	Sim
	17. O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação.	Não
	19. A competência da fonte produtora dos dados influencia na garantia de unicidade da informação.	Não
Validade	8. Percentuais estão representados corretamente dentro do intervalo de 0 a 100.	Não
	15. O período de referência dos dados está claramente definido.	Não
	20. A fonte dos dados é considerada confiável.	Sim
	22. Os dados são disponibilizados para consumo conforme esperado.	Não
	24. O usuário está corretamente autenticado ao acessar ou consumir os dados.	Não
	25. As restrições de uso dos dados são conhecidas e respeitadas.	Sim

Fonte: Dados da pesquisa (2025).

De modo mais detalhado, o Gráfico 4 mostra as respectivas quantidades de sentenças cujo IVC foi superior ou igual a 0,5 agrupadas por dimensão.

Gráfico 4: Quantidade de sentenças com IVC maior ou igual a 0,50



Fonte: elaborado pelo autor.

A análise da distribuição das sentenças que atingiram um Índice de Validade de Conteúdo (IVC) igual ou superior a 0,50 na Etapa 3 de validação, conforme o Quadro 6, oferece *insights* importantes sobre a operacionalização das dimensões da qualidade de dados do DAMA-DMBOK em um instrumento de mensuração. Das 31 sentenças avaliadas por 12 juízes, 17 alcançaram esse limiar, considerado razoável para fases preliminares de validação, indicando um nível de reconhecimento de sua vinculação conceitual pelos especialistas.

A dimensão Unicidade obteve o maior número de sentenças validadas (5 sentenças). Isso sugere que os conceitos relacionados à ausência de duplicações e à identificação única de registros são mais facilmente compreendidos e operacionalizados pelos especialistas. Sentenças como "A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados" e "Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única" corroboram a importância de mecanismos de controle para evitar redundâncias.

O maior número de sentenças para a dimensão Unicidade apresenta alinhamento com a literatura recente, que enfatiza a importância e caráter fundamental da Unicidade para a QD. A falta de unicidade dos dados pode constituir um problema expressivo para as organizações,

especialmente durante processos de migração de dados, em que essa duplicidade pode impactar negativamente a migração, com aumento dos esforços necessários para o procedimento de transferência dos dados (Pietersen; Rudman, 2025; Priebe; Neumaier; Markus, 2021; Yulfitri *et al.*, 2024).

O achado de que "O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação" complementa a teoria, evidenciando uma vulnerabilidade comum em ambientes corporativos. A utilização de planilhas ou outros meios que constituam sistemas não integrados pode comprometer o risco de roubo de dados, dentre outros problemas, ao passo em que também pode criar silos de dados e dificultar a padronização. Ademais, a utilização de sistemas não integrados, durante processos de migração, pode ocasionar falhas em projetos (Febiryani; Kusumasari; Fauzi, 2021; Hai, 2025; Yulfitri *et al.*, 2024).

Em seguida, a Atualidade teve 4 sentenças validadas. A percepção dos participantes do *focus group* de que a atualidade deveria ser o ponto de partida para qualquer análise, pois dados desatualizados levariam a decisões equivocadas independentemente de outras dimensões, revela a importância da pontualidade dos dados. Sentenças como "A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações" e "Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados" reforçam esse entendimento. A observação de que "A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente" demonstra como a tecnologia moderna impacta diretamente essa dimensão.

As ferramentas tecnológicas modernas são uma aliada para a garantia da atualidade dos dados, ao passo em que se evidenciam capazes de subsidiar análises de dados advindos de múltiplas fontes em um tempo relativamente menor, superando as limitações de sistemas mais antigos. A atualidade dos dados demonstra-se fundamental em um cenário corporativo em que a tomada de decisão em tempo real é um elemento imprescindível para o bom desempenho dos negócios (Pietersen; Rudman, 2025).

A percepção dos participantes do *focus group* acerca da dimensão da atualidade constituir o ponto de partida dos processos de mensuração da QD vai ao encontro da literatura recente sobre as dimensões do DAMA-DMBOK, ao passo em que as evidências empíricas apresentadas no referencial teórico deste estudo demonstram que a qualidade dos dados pode ser expressivamente comprometida por atrasos ou falta de atualização dos dados, podendo ocasionar, em última instância, decisões equivocadas (Pietersen; Rudman, 2025).

A Completude apresentou 3 sentenças validadas. Frases como "Os campos-chave identificadores (*Key Data Elements*) estão preenchidos em todos os registros" e "A base de

dados contempla registros de todas as marcas comercializadas pela organização" capturam a essência da completude, abordando tanto a granularidade (campos) quanto a abrangência (registros e entidades) dos dados. A incompletude é um dos principais problemas que podem comprometer a QD. A completude dos dados evidencia-se essencial para uma tomada de decisão baseada em dados de qualidade. A ausência de dados, mesmo quando parcial, pode prejudicar a qualidade e capacidade de suporte dos sistemas de dados, que podem não funcionar diante de dados ausentes (Febiryani; Kusumasari; Fauzi, 2021; Ferreira; Silva, 2018; Pietersen; Rudman, 2025).

As dimensões Consistência e Validade tiveram apenas 2 sentenças validadas cada. Essa baixa taxa de validação sugere uma maior dificuldade em traduzir seus conceitos em sentenças claras e unívocas para a percepção dos juízes. Para a Consistência, que busca a uniformidade entre diferentes sistemas e contextos, foram validadas as sentenças "Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema" e "As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados". Para a Validade, que indica a conformidade com regras e restrições predefinidas, foram validadas as sentenças "Percentuais estão representados corretamente dentro do intervalo de 0 a 100" e "As restrições de uso dos dados são conhecidas e respeitadas". A literatura apresenta a inconsistência dos dados como um problema para o alcance de níveis satisfatórios de QD, especialmente se os dados inconsistentes vierem de muitas fontes e em diversos formatos (Pietersen; Rudman, 2025; Yulfitri *et al.*, 2024).

Por fim, a Precisão foi a dimensão com o menor número de sentenças validadas (apenas 1). Essa baixa validação evidencia uma dificuldade significativa em operacionalizar o conceito de fidelidade à realidade. Por ser um dos pilares da QD, a baixa precisão de um sistema de dados pode impactar diretamente na qualidade dos dados utilizados. A falta de governança de dados pode impactar na qualidade, integridade e segurança dos dados, com reflexos para a precisão (Ferreira; Silva, 2018; Pietersen; Rudman, 2025).

Essa distribuição desigual na validação de conteúdo reforça a conclusão da dissertação de que a tradução de conceitos teóricos em sentenças objetivas de questionário é mais complexa do que o previsto. A discrepância entre as etapas de geração de sentenças nos *focus groups* e a validação por juízes sugere que a etapa de seleção inicial pode não ter sido robusta o suficiente para filtrar sentenças ambíguas. Isso implica a necessidade de um processo iterativo e rigoroso de formulação e validação para garantir que as sentenças sejam compreendidas de forma consistente pelos avaliadores e, futuramente, pelos usuários finais do instrumento.

4.4 Discussão dos resultados com a literatura

Os achados dos *focus groups*, mesmo para as sentenças que não obtiveram validação estatística na etapa de juízes, ainda oferecem *insights* importantes:

- Atualidade como porta de entrada: a sugestão de iniciar qualquer análise pela dimensão Atualidade, uma vez que dados não atualizados invalidariam análises posteriores, representa uma hierarquização da importância das dimensões na prática. Apesar da literatura recente sobre o tema apresentar a atualidade como uma dimensão de grande importância, as evidências deste estudo ratificam e complementam outros estudos, trazendo a compreensão da atualidade como um ponto fundamental para a QD (Pietersen; Rudman, 2025; Rahmawati *et al.*, 2023);
- Precisão dedutível da imprecisão: o achado sobre a capacidade de inferir dados precisos a partir de um padrão de imprecisão conhecido ("dentro da imprecisão, é possível identificar a precisão desde que se conheça o padrão do dado analisado") é particularmente instigante. A literatura conceitua a precisão como a qualidade dos dados em serem exatos, confiáveis e, o máximo possível, livres de erros. Esse achado ratifica a noção, discutida na literatura, sobre a possibilidade de avaliar a QD não somente com base em sua precisão ou exatidão absoluta, mas também por meio da consistência dos desvios identificados (Rahmawati *et al.*, 2023);
- Impacto de vieses humanos: a inclusão de sentenças como "Pode haver viés tanto na geração quanto na interpretação dos dados" e "O viés do analista pode interferir na verificação da unicidade dos dados" (ainda que não validadas pelo IVC) reflete uma preocupação latente entre os profissionais com os aspectos subjetivos da qualidade dos dados, além das métricas puramente técnicas. A consideração de aspectos subjetivos, ao tratar da percepção profissional sobre a QD, evidencia-se importante ao passo em que a literatura apresenta evidências empíricas acerca da influência de abordagens intuitivas, construídas por meio de experiências práticas e na compreensão de atributos essenciais, na avaliação da QD (Pietersen; Rudman, 2025; Rahmawati *et al.*, 2023);
- Ameaças de fontes comuns: a sentença "O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação" (validada pelo IVC) é um achado prático que, embora não seja uma inovação teórica, evidencia uma vulnerabilidade comum em ambientes corporativos. Isso ressalta a importância de fontes de dados estruturadas e governadas para garantir a unicidade;
- Qualidade dos dados como um ativo "vivo": a ideia de que "Os dados são considerados vivos (orgânicos), e sua precisão pode evoluir ao longo do tempo"

(validada pelo IVC) sugere uma perspectiva dinâmica da qualidade que complementa a noção estática das dimensões. A literatura reconhece o gerenciamento de dados como um processo contínuo de desenvolvimento, implementação e supervisão de procedimentos que visem a melhoria e manutenção da QD. Esse processo constante, que incorpora validação, limpeza, integração e controle da qualidade dos dados demonstra-se fundamental para a manutenção de dados confiáveis e, por conseguinte, para uma tomada de decisão eficaz (Taryono *et al.*, 2024; Miller *et al.*, 2025; Nugraha *et al.*, 2024).

Os achados desta pesquisa vão além da construção de um instrumento de mensuração, ao revelar como as dimensões da qualidade de dados se manifestam na prática organizacional. A percepção da Atualidade como porta de entrada evidencia uma hierarquização funcional das dimensões, em que a defasagem de dados compromete a análise subsequente. A seguir, apresenta-se o produto tecnológico construído e as considerações finais da pesquisa.

5 PRODUTO TECNOLÓGICO

Como parte de um mestrado profissional, a relevância deste trabalho não se limita à fundamentação teórica ou à discussão metodológica. Sua relevância se consolida na capacidade de gerar soluções aplicáveis a problemas concretos. É nesse contexto que o presente capítulo destinado à apresentação do produto tecnológico, traduz os resultados da pesquisa em um artefato prático, concebido para uso em ambientes organizacionais.

O produto tecnológico apresentado na Tabela 4 consiste em um instrumento para avaliação da qualidade de dados, elaborado a partir das sentenças que, ao final da etapa de análise de juízes, obtiveram Índice de Validade de Conteúdo (IVC) igual ou superior a 0,50. Esse recurso está disponível em formato *Google Sheet* (Apêndice G), permitindo que os gestores acessem e realizem a avaliação da qualidade dos dados utilizados sempre que seja necessário mensurar a qualidade de uma base de dados.

Tabela 4: Exemplo de uso do produto tecnológico

Dimensão Referência	Sentenças	Pontuação
Precisão	Os dados de destino (<i>target</i>) refletem com precisão os dados de origem (<i>source</i>).	2
Consistência	Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema.	3
Unicidade	A origem dos dados é claramente identificada e contribui para evitar redundância.	5
Atualidade	A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações.	7
Unicidade	A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro.	1
Unicidade	A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados.	3
Atualidade	Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados.	7
Consistência	As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados.	9
Unicidade	A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos.	3
Unicidade	Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única.	9
Validade	A fonte dos dados é considerada confiável.	1
Compleitude	Para empresas que atuam em diferentes mercados (ex.: por país ou região), é possível verificar se todos os mercados estão representados na base de dados.	7
Validade	As restrições de uso dos dados são conhecidas e respeitadas.	1
Compleitude	A base de dados contempla registros de todas as marcas comercializadas pela organização.	3
Atualidade	A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente.	7
Atualidade	O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados.	5
Compleitude	Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor.	7

Fonte: Dados da pesquisa (2025).

Desenvolvido com base nas dimensões estabelecidas pelo *framework* DAMA-DMBOK, o instrumento foi criado no contexto desta dissertação para mensurar, de maneira sistemática e quantitativa, a percepção dos usuários sobre uma base de dados específica. Seu propósito é oferecer um procedimento padronizado capaz de identificar pontos fortes e fragilidades da base avaliada, subsidiando ações de aprimoramento e decisões estratégicas voltadas à governança de dados.

Quanto ao público-alvo, trata-se de gestores e profissionais diretamente envolvidos com a governança e o uso estratégico de dados no ambiente corporativo, que dispõem de condições para avaliar a qualidade dos dados sob sua responsabilidade. Já a unidade de análise refere-se a uma base de dados específica administrado pela área ou departamento do respondente, de acordo com o escopo de sua atuação. Como exemplos de bases de dados poderiam ser avaliados de modo independente o cadastro de clientes, fornecedores, transações comerciais, entre outros.

A primeira interface, representada no quadro acima, corresponde a uma planilha estruturada no *Google Sheets* para a coleta das respostas. A tabela é composta por três colunas principais. A coluna "Dimensão Referência" indica a dimensão de qualidade à qual a sentença está associada, como Precisão, Consistência, Unicidade, Atualidade, Validade e Completude. Em seguida, a coluna "Sentenças" reúne afirmações objetivas que funcionam como critérios de análise, elaboradas para estimular o avaliador a refletir sobre o grau de atendimento da base de dados em relação ao requisito descrito.

Por exemplo, na dimensão Precisão, a sentença “Os dados de destino (*target*) refletem com precisão os dados de origem (*source*)” leva o respondente a ponderar sobre a exatidão das informações armazenadas. Por fim, a coluna "Pontuação" é destinada ao registro da nota atribuída pelo avaliador, variando de 0 (atendimento mínimo) a 10 (atendimento máximo), de acordo com sua percepção. Essa configuração não apenas organiza e padroniza a coleta de respostas, como também viabiliza a análise comparativa entre as diferentes dimensões avaliadas. A segunda interface corresponde ao painel de resultados consolidados conforme Tabela 5.

Tabela 5: Exemplo de apresentação dos resultados da mensuração por dimensão de QD

RESULTADOS		
Dimensão	No. Questões	Pontuação Média
Atualidade	4	6,50
Compleitude	3	5,67
Consistência	2	6,00
Precisão	1	2,00
Unicidade	5	4,20
Validade	2	1,00
Geral	17	4,71

Fonte: Dados da pesquisa (2025).

A Tabela 5 mostra um exemplo de apresentação dos resultados em que cada dimensão de qualidade de dados é apresentada juntamente com a quantidade de sentenças que compuseram sua avaliação (No. Questões) e com a Pontuação Média atribuída pelos participantes. O valor Geral, mostrado na última linha, representa a média de todas as dimensões avaliadas (4,71) e funciona como parâmetro para identificar quais atributos se encontram acima ou abaixo desse índice.

Por um lado, a leitura dos resultados no exemplo fornecido revela que Atualidade (6,50), Consistência (6,00) e Compleitude (5,67) obtiveram pontuações superiores à média geral, sinalizando uma percepção mais favorável em relação a esses aspectos da base de dados. De modo oposto, Unicidade (4,20), Precisão (2,00) e, de forma mais acentuada, Validade (1,00) ficaram abaixo desse patamar, apontando pontos críticos que demandam maior atenção e ações corretivas prioritárias.

Esse resumo numérico não apenas organiza e sintetiza as informações obtidas, mas também oferece subsídios claros para a análise gerencial, auxiliando o responsável pela governança de dados a direcionar esforços e recursos para as dimensões que apresentam maior fragilidade. Ao integrar esse diagnóstico quantitativo ao instrumento de coleta previamente descrito, cria-se um fluxo coerente que vai da avaliação estruturada de critérios específicos até a definição objetiva de prioridades para intervenção, reforçando a utilidade do produto como ferramenta de acompanhamento contínuo e apoio à tomada de decisões fundamentadas em evidências.

6 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo geral propor um instrumento de mensuração da QD prático, replicável e fundamentado numa estrutura conceitual sólida, o modelo de seis dimensões propostas pelo DAMA-DMBOK: completude, unicidade, atualidade, validade, precisão e consistência. Ainda, como objetivos específicos, pretende-se: Propor sentenças representativas das seis dimensões do modelo DAMA-DMBOK por meio de *focus group* com especialistas; selecionar as sentenças mais pertinentes por meio de questionário sob a perspectiva de especialista; e validar o conteúdo das sentenças selecionadas por juízes especialistas, utilizando o Índice de Validação de Conteúdo (IVC).

A pesquisa apresentou uma abordagem mista, integrando técnicas quantitativas e qualitativas durante o desenvolvimento do instrumento de mensuração da QD. Com essa finalidade, a pesquisa dividiu-se em três fases distintas. A primeira fase realizada foi o *focus group*, com a finalidade de explorar a percepção de especialistas sobre as seis dimensões da qualidade de dados propostas pelo modelo DAMA-DMBOK. Em sequência, a segunda fase da pesquisa foi a seleção das sentenças, por fim, com base no IVC as sentenças finais do instrumento foram selecionadas na etapa 3 da pesquisa. Dessa forma, um instrumento composto por 17 sentenças foi o produto desta pesquisa. Com base nas sentenças, foi desenvolvido um produto tecnológico, disponibilizado de forma virtual, que visa a contribuir para que gestores possam avaliar a qualidade dos dados que utilizam em suas empresas, mesmo que de uma forma inicial.

As sessões de *focus group* revelaram que parte dos participantes possuía familiaridade conceitual com o *framework* DAMA-DMBOK, reconhecendo-o como um referencial para a avaliação da qualidade de dados. No entanto, não houve menção espontânea a outros *frameworks*, como ISO 8000, ISO 25012, TDQM ou COBIT. Essa ausência de referência formal sugere que, embora alguns profissionais conheçam a nomenclatura do DAMA-DMBOK, seu contato com modelos alternativos ou complementares é limitado. Por outro lado, a experiência prática demonstrada foi consistente, revelando domínio na aplicação cotidiana de procedimentos como auditorias, conciliações, dados duplicados e rastreabilidade — práticas essenciais para assegurar a integridade e confiabilidade das informações corporativas.

Ainda que não explicitassem a utilização de outros referenciais, foi possível identificar um alinhamento implícito com boas práticas previstas em *frameworks* como a ISO 8000 e o TDQM. Questões relacionadas à rastreabilidade, consistência e completude aproximam-se de diretrizes presentes na ISO 8000, enquanto a menção à definição de métricas e indicadores de desempenho (KPIs) para a qualidade dos dados reflete princípios centrais do TDQM. Essa

convergência indica que, mesmo sem o uso formal das nomenclaturas, parte dos participantes incorpora elementos estruturados de gestão da qualidade em seu trabalho.

A análise das falas também revelou áreas de sobreposição conceitual entre as dimensões do DAMA-DMBOK, indicando que, na prática, as fronteiras entre elas nem sempre são claras. Um exemplo recorrente foi a relação entre validade e completude: alguns participantes mencionaram que, dependendo do *framework* de referência, a verificação de campos nulos (*null check*) poderia ser classificada em qualquer uma das duas dimensões. De modo semelhante, a presença de dados obrigatórios foi, em determinados momentos, discutida como parte da validade, reforçando a percepção de que esses conceitos se interconectam.

Outro ponto de confusão ocorreu entre validade e consistência, quando afirmações como “garantir que o dado reflita a realidade” ou “bater valores entre sistemas” foram associadas à validade, apesar de seu enquadramento mais adequado estar relacionado à consistência. Na dimensão atualidade, observou-se que alguns participantes a associaram diretamente à precisão, tratando dados corretos e atualizados como um único critério de confiabilidade. De modo semelhante, a unicidade foi, em algumas falas, vinculada à consistência, especialmente ao abordar de duplicação e coerência de registros entre diferentes sistemas.

Por fim, as interações sugerem que a base de conhecimento dos participantes está fortemente ancorada na experiência prática, mais do que na formalização teórica. A maior parte dos exemplos relatados decorre de experiência profissional, o que garante um entendimento empírico dos aspectos operacionais da qualidade de dados. Contudo, essa abordagem pragmática, embora eficiente no dia a dia, pode dificultar a associação precisa de cada prática à dimensão conceitual correspondente, sobretudo quando as distinções são sutis e variam entre diferentes modelos de referência.

Enquanto limitações da pesquisa, destaca-se que técnicas mais robustas poderiam ter sido utilizadas para os procedimentos de validação do instrumento, tais como AFE, com vista a avaliar em que medida a estrutura interna do instrumento reflete a estrutura teórica. Ademais, a aplicação deste instrumento em outras pesquisas voltadas para a QD poderia contribuir para a identificação de limitações e potencialidades inerentes de diferentes setores de atividade econômica, o que poderia fortalecer a consistência do instrumento desenvolvido e a compreensão de sua aplicabilidade em diferentes contextos organizacionais. Por fim, é importante destacar que a composição das amostras nos *focus groups*, no questionário de seleção e na análise de juízes foi realizada por conveniência, utilizando-se predominantemente a rede de contatos do pesquisador.

Durante a realização da pesquisa, foi possível compreender que a mensuração da QD não é uma tarefa simples e a compreensão, mesmo em meios seletos como de especialistas, ainda pode ser fragmentada no que se refere às dimensões da QD. Em pesquisas futuras, recomenda-se o desenvolvimento de instrumentos que considerem outros frameworks como base conceitual, com o TDQM ou a ISO 8000. Em adição, a aplicação do instrumento proposto nesta pesquisa em diferentes contextos organizacionais e acadêmicos poderá contribuir para o desenvolvimento e discussões acerca de instrumentos de mensuração da QD, como meios de suporte para a tomada de decisão empresarial.

REFERÊNCIAS

- ABHIYASA, S. D. LUBIS, M. WIDARTHA, V. P. HENDRAWAN, F. R. Assessment Analysis of Data Ticketing Quality of The Multimedia Division of Telecommunication Companies in Indonesia Using Data Quality Assessment (DQA) Based on DAMA-DMBOK. **2024 IEEE International Conference on Computing (ICOCO)**. p. 78-83. <http://doi.org/10.1109/ICOCO62848.2024.10928243>
- BERNARDO, B. M. V. MAMEDE, H. S. BARROSO, J. M. P. SANTOS, V. M. P. D. Data governance & quality management—Innovation and breakthroughs across different fields. **Journal of Innovation & Knowledge**, v. 9, n. 4, p. 100598, 2024. <https://doi.org/10.1016/j.jik.2024.100598>
- BOWO, W. A. SUHANTO, A. NAISUTY, M. MA'MUN, S. HIDAYANTO, A. N. HIDAYANTO, I. C. Data quality assessment: A case study of PT JAS using TDQM Framework. **International Conference on Informatics and Computing (ICIC)**. p. 1-6, 2019. <http://doi.org/10.1109/ICIC47613.2019.8985896>
- BROWN, W. Some experimental results in the correlation of mental abilities. **British Journal of Psychology**, v. 3, p. 296–322, 2006.
- DAMA International, **DAMA-DMBOK: data management body of knowledge**. 2017.
- ERSHADI, M. J.; MOHSENI, A.; NIAKI, S. T. A. Measuring Data Quality of Theses and Dissertations in the Data Preparation Stage of Registration Systems, **International Journal of Information Science and Management**, v. 21, n. 4, p. 189-213., 2023. <https://doi.org/10.22034/ijism.2023.1977986.0>
- FEBIRYANI, W. KUSUMASARI, T. F. FAUZI, R. Analysis and Design of Implementation Guidelines Data Security Management Assessment Techniques Based on DAMA-DMBOKv2. **International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)**, p. 371-375, 2021. <https://doi.org/10.1109/ICITISEE53823.2021.9655782>
- FEREIRA, M. SILVA, L. A. Data quality measurement framework. **Latin American Computer Conference (CLEI)**, p. 455-463, 2018. <http://doi.org/10.1109/CLEI.2018.00061>
- HÄGELE, L. Klier, M. MOESTUE, L. OBERMEIER, A. Aspect-based currency of customer reviews: A novel probability-based metric to pave the way for data quality-aware decision-making. **Electronic Markets**, v. 35, n. 10, p. 1-18, 2025. <https://doi.org/10.1007/s12525-025-00760-4>

HAI, C. T. H. A maturity assessment model for data governance in small and medium enterprises in Vietnam. **Edelweiss Applied Science and Technology**, v. 9, n. 2, p. 1931-1951, 2025. <https://doi.org/10.55214/25768484.v9i2.4982>

HENDRAWAN, F. R. KUSUMASARI, T. F. FAUZI, R. Analysis of design implementation guidelines for data governance management based on DAMA-DMBOKv2. **International Conference on Informatics and Computing (ICIC)**, p. 1-6, 2022. <http://doi.org/10.1109/ICIC56845.2022.10007021>

HERNÁNDEZ-NIETO, R. A. **Contributions to statistical analysis**. Mérida: Universidad de Los Andes, 2022.

HONG, R. HONG, P. Y. P. CHO, J. Y. Developing and validating the perceived youth barriers scale (PYBS): A bottom-up approach to assessing barriers faced by underprivileged young adults. **Children and Youth Services Review**, v. 177, p. 1-6, 2025. <https://doi.org/10.1016/j.chilyouth.2025.108465>

HUNSDIECK, B. BENDER, C. ICKSTADT, K. MIELKE, J. Joint models in big data: simulation-based guidelines for required data quality in longitudinal electronic health records. **BioData Mining**, v.18, n. 35, p. 1-27, 2025. <https://doi.org/10.1186/s13040-025-00450-z>

KANTHASAMY, V. SCHILLING, R. J. HORAN, S. BAKER, V. AHLUWALIA, N. HUNTER, R. J. HONARBAKHS, S. FINLAY, M. Development and validation of a new disease-specific quality-of-life measure for atrial fibrillation derived from patient perspectives: The Atrial Fibrillation Patient-Reported Outcome Measures (AF-PROMs) Questionnaire. **Heart Rhythm**. v. 22, n. 8, p. 1922-1934, 2025. <https://doi.org/10.1016/j.hrthm.2025.02.007>

KHAIRUNISAK KUSUMASARI, T. F. FAUZI, R. Design guidelines and process of metadata management based on data management body of knowledge. **International Conference on Information Management (ICIM)**, 2021. <https://doi.org/10.1109/ICIM52229.2021.9417156>

KIM, S. CASTILLO, R. P. CABALLERO, I. LEE, D. Organizational process maturity model for IoT data quality management. **Journal of Industrial Information Integration**, v. 26, p. 1-16, 2022. <https://doi.org/10.1016/j.jii.2021.100256>

KUSUMASARI, T. F. FAUZI, R. Analysis and Design of Data Warehousing and Business Intelligence Guidelines Using DAMA-DMBOKv2. **2022 International Conference on Electrical and Information Technology (IEIT)**, p. 109-114, 2022. <http://doi.org/10.1109/IEIT56384.2022.9967862>

LEE, C. KIM, S. LEE, J. LEE, D. CABALLERO, I. CASTILLO, R. P. Data Quality Assessment for Manufacturing Processes. **ICIC Express Letters**, v.10, n. 12, p. 1107-1113, 2019. <https://doi.org/10.24507/icicelb.10.12.1107>

LIN, B. LEE, W. CHOI, H. C. ANDERECK, K. Developing and validating an ethical risk scale for service robots in hospitality. **International Journal of Hospitality Management**, v. 132, p. 1-13, 2026. <https://doi.org/10.1016/j.ijhm.2025.104383>

LIU, J. JIANG, F. TASHIRO, S. CHEN, S. TANAKA, M. A physics-informed and data-driven framework for robotic welding in manufacturing. **Nature Communications**, v. 16, p. 1-18, 2025. <https://doi.org/10.1038/s41467-025-60164-y>

LYNN, M. R. Determination and quantification of content validity. **Nursing research**, v.35, n. 6, p. 382-386, 1986. <https://doi.org/10.1097/00006199-198611000-00017>

MELONI, C. BLASO, L. BRANCHETTI, S. CALDERA, M. MAESTOSI, P. C. D'AGOSTA, G. FRASCELLA, A. GOZO, N. MASSA, G. MORETTI, F. NOVELLI, C. PIZZUTI, S. ROMANO, S. TOFANI, A. TUNDO, A. ZANGHERI, P. Energy and Digital Transitions for Energy Communities: Tools and Methodologies to Promote Digitalization in Italy. **Electronics**, v. 14, n. 10, p. 1-21, 2025. <https://doi.org/10.3390/electronics14102027>

MILLER, G. SPIEGEL, E. Guidelines for research data integrity (GRDI). **Scientific Data**, v. 12, n. 1, p. 1-8, 2025. <https://doi.org/10.1038/s41597-024-04312-x>

MILLER, R. CHAN, S. H. M. WHELAN, H. GREGÓRIO, J. A Comparison of Data Quality Frameworks: A Review. **Big Data and Cognitive Computing**, v. 9, n. 4, p. 1-21, 2025. <https://doi.org/10.3390/bdcc9040093>

MISHRA, V. DARADE, H. Advancements in Data Quality Management in the Big Data Era: A Comprehensive Review. **Third International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS)**, p. 1-6, 2024. <http://doi.org/10.1109/AICECS63354.2024.10956447>

MORGAN, D. L. Focus groups. **Annual Review of Sociology**, v. 22, p. 129-152, 1996. <https://doi.org/10.1146/annurev.soc.22.1.129>

NUGRAHA, T. F. WIBOWO, W. S. GENIA, V. FADHIL, A. RULDEVIYANI, Y. A Practical Approach to Enhance Data Quality Management in Government: Case Study of Indonesian Customs and Excise Office. **Journal of Information Systems Engineering & Business Intelligence**, v. 10, n. 1, p. 51-69, 2024. <https://doi.org/10.20473/jisebi.10.1.51-69>

PIETERSEN, L. A. RUDMAN, R. J. Data-related risks for the use of machine learning in retail customer demand forecasting. **South African Journal of Business Management**, v. 56, n. 1, p. 1-13, 2025. <https://doi.org/10.4102/sajbm.v56i1.4766>

POLIMENO, A. BRAGHIN, C. ANISETTI, M. ARDAGNA, C. A. Maximizing data quality while ensuring data protection in service-based data pipelines. **Journal of Big Data**, v. 12, n. 62, p. 1-34, 2025. <https://doi.org/10.1186/s40537-025-01118-5>

PRIEBE, T. NEUMAIER, S. MARKUS, S. Finding your way through the jungle of big data architectures. **2021 IEEE International Conference on Big Data (Big Data)**, p. 5994-5996, 2021. <https://doi.org/10.1109/BigData52589.2021.9671862>

RAHMAWATI, R. RULDEVIYANI, Y. ABDULLAH, P. P. HUDOARMA, F. M. R. Strategies to Improve Data Quality Management Using Total Data Quality Management (TDQM) and Data Management Body of Knowledge (DMBOK): A Case Study of M-Passport Application. **CommIT (Communication and Information Technology) Journal**, v. 17, n. 1, p. 27-42, 2023. <https://doi.org/10.21512/commit.v17i1.8330>

SCHMUCK, M. Master Data Management as Part of Data Governance: A Maturity Model to Improve Efficiency and Trust in Master Data and Thus Business Performance. **Business Performance Review**, v. 2, p. 1-15, 2024. <https://doi.org/10.22495/bprv2i2p2>

SCHUESSLER, M. FLEMING, S. MEYER, S. SETO, T. HERNANDEZ-BOUSSARD, T. Diagnostic framework to validate clinical machine learning models locally on temporally stamped data. **Communications Medicine**, v. 5, n. 261, p. 1-12, 2025. <https://doi.org/10.1038/s43856-025-00965-w>

SHARMA, V. JAMWAL, A. AGRAWAL, R. PRATAP, S. A review on digital transformation in healthcare waste management: Applications, research trends and implications. **Waste Management & Research**, v. 43, n. 6, p. 828-849. 2025. <https://doi.org/10.1177/0734242X241285420>

TARYONO, T. KUSUMASARI, T. F. FAUZI, R. FEBRIYANI, W. Analysis and Design of a Data Architecture Management Assessment Guide using DAMA-DMBOKv2 and Process Assessment Model. **2024 Ninth International Conference on Informatics and Computing (ICIC)**, p. 1-6, 2024. <http://doi.org/10.1109/ICIC64337.2024.10957667>

WANG, J. PAN, C. ZHANG, Q. Double weighted combat data quality evaluation method based on CVF optimized FAHP. **Scientific Reports**, v. 15, n. 2516, p. 1-13, 2025. <https://doi.org/10.1038/s41598-025-87266-3>

WINGS, S. HARKONEN, J. Toward dataflow-centric process management of data-rich processes. **Journal of Systems and Information Technology**. V. 27, n. 3, p. 460-488, 2025. <https://doi.org/10.1108/JSIT-07-2024-0243>

YAMAN, B. THOMPSON, K. FAHEY, F. BRENNAN, R. LinkedDataOps: quality oriented end-to-end geospatial linked data production governance. **Semantic Web**, p. 1-27, 2024. <https://doi.org/10.3233/SW-233293>

YANG, Z. LI, Z. CHENG, F. LV, Q. LI, K. ZHANG, T. ZHOU, Y. ZHAO, B. XUE, W. WEI, J. Two-decade surface ozone (O₃) pollution in China: Enhanced fine-scale estimations and

environmental health implications. **Remote Sensing of Environment**, v. 317, p. 1-20, 2025. <https://doi.org/10.1016/j.rse.2024.114459>

YEUNG, W. N. DI CLEMENTE, R., LAMBIOTTE, R. Garbage in garbage out? Impacts of data quality on criminal network intervention. *EPJ Data Science*, v. 14, n. 37, p. 1-18, 2025. <https://doi.org/10.1140/epjds/s13688-025-00553-x>

YILDIZ, C. Ç. CEYLAN, B. K. ERGÜN, Y. METE, M. Development and psychometric testing of the reality shock scale for nursing students. **Nurse Education Today**. v. 153, p. 1-9, 2025. <https://doi.org/10.1016/j.nedt.2025.106795>

YULFITRI, A. SENSUSE, D. I. HIDAYAT, D. S. SURYONO, R. R. KAUTSARINA PURWANINGSIH, E. H. PRABUWONO, A. S. The role of data governance in a high-level approach of data migration to open data. **Business Information Review**, v. 41, n. 3, p. 94-103, 2024. <https://doi.org/10.1177/02663821241264702>

ZAHABIA; KUSUMASARI, T. F.; FAUZI, R. Analysis and Design of Data Warehousing and Business Intelligence Guidelines Using DAMA-DMBOKv2. International conference of electrical and information technology, 2022

ZAMANZADEH, V. GHAHRAMANIAN, A. RASSOULI, M. ABBASZADEH, A. ALAVI-MAJD, H. NIKANFAR, A. R. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. **Journal of caring sciences**, v. 4, n. 2, p. 165-178, 2015. <https://doi.org/10.15171/jcs.2015.017>

ZHANG, Y. CALLAGHAN-KORU, J. A. KORU, G. The challenges and opportunities of continuous data quality improvement for healthcare administration data. **JAMIA open**, v. 7, n. 3, p. 1-8, 2024. <https://doi.org/10.1093/jamiaopen/ooae058>

APÊNDICE A – Roteiro do *focus group*

Etapa 1: Roteiro para Realização de Focus Group

Bloco	Texto lido aos participantes pelo moderador
1.Instruções Iniciais	Você foi convidado(a) a participar de um estudo cujo propósito é identificar e compreender as dimensões que determinam a qualidade dos dados. Dessa forma, busca-se construir um debate estruturado sobre as seis dimensões da qualidade dos dados e elaborar três perguntas para cada uma delas, de modo a sintetizar seus conceitos essenciais para fins de mensuração.
2.Apresentação do Moderador	Meu nome é Fabio Santana Reis e assumo como moderador do debate, que tem duração prevista de 60 minutos.
3.Consentimento dos participantes	Por favor, assegurem-se que estejam confortáveis para que não haja interrupções e sintam-se livres para expressar suas opiniões. Esta sessão está sendo gravada para análise futura, mas se alguém não estiver favorável, por favor solicito que apresente sua objeção. Não havendo objeção, vamos seguir.
4. Formato e duração da sessão (60 minutos)	Teremos 6 rodadas de discussão, cada uma delas focada numa das dimensões de mensuração da qualidade de dados com duração de 10 minutos. Cada rodada inicia-se com a apresentação da dimensão que será objeto de discussão entre vocês com uma breve descrição teórica. E, ao final da rodada, solicito a gentileza de formularem questões relacionadas à dimensão em foco visando obter uma métrica.
5. Em cada rodada (1 a 6), o moderador apresentará brevemente a dimensão em foco para que os participantes colaborem na formulação de sentenças relacionadas	1.Completude: Representa a proporção de dados armazenados em relação ao seu potencial total. Indica a ausência de lacunas que possam comprometer o uso da informação. Dados completos garantem que todas as informações necessárias estejam disponíveis para suportar decisões eficazes 2.Unicidade: Refere-se à ausência de duplicações nos dados. Cada entidade deve ser registrada uma única vez, de acordo com critérios de identificação definidos. Essa dimensão é essencial para evitar inconsistências e redundâncias em análises. 3.Atualidade: Mede o grau em que os dados refletem a realidade no momento em que são necessários. Dados atualizados permitem que decisões sejam tomadas com base em informações recentes e confiáveis. 4.Validade: Avalia se os dados estão em conformidade com regras, padrões e restrições predefinidos. Inclui a adequação ao formato, tipo e contexto de uso pretendido, garantindo que sejam utilizáveis e relevantes. 5.Precisão: Reflete o grau em que os dados representam corretamente a realidade ou um valor específico. Dados precisos devem estar livres de erros e serem fiéis à sua fonte original. 6.Consistência: Trata da ausência de discrepâncias entre diferentes representações do mesmo dado em diferentes sistemas ou bases. Dados consistentes garantem que a informação seja coerente em todos os contextos organizacionais.
6. Agradecimento	Gostaria de expressar meu sincero agradecimento a todos que participaram da sessão de <i>focus group</i> . Sua contribuição foi fundamental para o desenvolvimento desta pesquisa e para o enriquecimento da análise dos dados, trazendo perspectivas valiosas e experiências que certamente fortalecerão a qualidade dos resultados.
7.Encerramento	Informo que, por se tratar de uma pesquisa acadêmica, os resultados serão públicos e poderão ser consultados livremente após a conclusão do trabalho. Aos participantes que desejarem receber uma cópia da versão final da dissertação, peço que, por gentileza, se manifestem para que eu possa encaminhá-la diretamente. Será uma satisfação compartilhar o produto final desta etapa de estudos, que só foi possível graças à colaboração de todos vocês.

APÊNDICE B – Perguntas geradas pelos participantes do *focus group* 1

Etapa 1: Perguntas geradas pelos participantes do Focus Group 1

Dimensão	Perguntas
Unicidade	Qual a origem do dado? Os dados são rastreáveis? Você confia na fonte de dados? A fonte de dados é qualificada? Os dados tem caráter subjetivo ou objetivo? Tem uma fonte de comparação? Como posso garantir que a fonte tem competência para produzir os dados? Os mesmos dados aparecem em mais de um canal? O canal de consulta ao dado tem competência para divulgá-lo?
Atualidade	Qual a periodicidade do dado? Com que frequência deveria ser atualizado? Mensal, Trimestral, etc? Os dados são do passado (histórico) ou do futuro (preditivos)? Com quanto tempo de antecedência consigo o dado? Os dados analisados são para venda e neste caso tenho dados históricos? Os dados analisados são para novas aquisições e neste caso tenho dados preditivos? Qual o objetivo de análise? Qual a data de corte dos dados em relação ao período de análise? Qual o histórico analisado para projetar o futuro?
Validade	No contexto contábil, os dados de uma eventual demonstração financeira seguem o padrão IRFS? Os dados seguem uma norma? Os dados apresentam um padrão? Os dados apresentados fazem sentido? Qual a utilização dos dados? Há um padrão que permite comparações? Os dados analisados tem similaridade com o histórico? Os dados estão no formato esperado? Estão adequados?
Precisão	É possível identificar alguma tendência nos dados analisados? Os dados apresentam alguma anomalia em relação a base histórica? Posso confiar na fonte fornecedora dos dados? O formato do dado está adequado? O dado está corrompido? É um dado singular (uma só fonte) ou plural (múltiplas fontes)? Os dados estão adequados ao objetivo de uso? Os dados estão dentro do contexto de análise? Foi realizada uma análise de similaridade? Quantas e quais as fontes que disponibilizam os dados? É possível identificar um padrão de imprecisão que permite deduzir dados precisos? Exemplo: um concorrente que sempre divulga vendas 20% acima do valor real.
Consistência	Os dados são compatíveis com o mercado quando realizada uma comparação entre pares? Qual a variação em relação a base histórica? Quanto especialistas avaliaram os dados? Qual a opinião do especialista naquele tipo de dado? Quais os especialistas avaliaram os dados e quais são os respectivos pesos de cada opinião?
Completeness	Não houve tempo

APÊNDICE C – Perguntas geradas pelos participantes do *focus group* 2

Etapa 1: Perguntas geradas pelos participantes do Focus Group 2

Dimensão	Perguntas
Unicidade	Qual a chave/campo de identificação dos registros no cadastro? Como posso adotar uma chave, por exemplo, RG, CPF? Como posso cruzar diferentes bases digitados sem redundância? O que é mais importante para quem ou quais áreas da empresa avaliam os dados? Como posso obter colaboração entre as áreas para obtenção de dados não duplicados? Como evitar o uso de diversos sistemas? Estou coletando da fonte apropriada? Os dados têm governança? Os dados são provenientes de planilha Excel? Há uma estratégia clara para os dados? A captura de dados foi realizada de uma única fonte? O que você quer analisar? Você quer identificar desvios? A análise pode ser simplificada? Qual o ponto crítico a ser analisado com base no objetivo? Estão sendo considerados vieses na busca por dados? Exemplo de viés de confirmação: busca por dados para confirmar uma tese pré-estabelecida. Qual o viés de quem analisa o dado?
Atualidade	Pode ser observada uma sazonalidade de vendas? A comparação está sendo feita com a mesma faixa de período? Como considerar a sazonalidade? Qual impacto esses dados poderiam causar? Qual a recorrência de atualização? Um fato relevante pode impactar os dados? O que pode causar uma mudança no tempo em relação aos dados para uma avaliação de evolução temporal? Quem é o responsável pela atualização dos dados? Há interesses ocultos em relação aos dados? A digitalização de processos permite enxergar em tempo real mas eu consigo interferir também em tempo real? Como você considera a atualidade dos dados no atual contexto de mundo volátil com tantas mudanças rápidas? Consigo restringir o volume de dados atualizados, por exemplo produtos mais lucrativos? A omnicanalidade permite disponibilizar os dados em diferentes canais?
Validade	Qual a fonte ? A fonte é confiável? Qual o período dos dados? Qual o objetivo da análise? Onde você quer chegar com esses dados? Qual o output para o input que tenho? As Fontes se cruzam? O que se pretende? Qual é o dado? O faturamento é com imposto ou sem imposto? Quais as restrições? Qual o volume de dados? A informação está clara? Tenho clareza dos dados?

	Os dados possuem auditoria externa? Como outras áreas da empresa avaliam o dado? Foi efetuado o processo de validação cruzada? Como evitar os vieses para conseguir informação válida? Os dados são disponibilizados? Qual o formato dos dados? A integração é um processo simples? Você concorda que o simples bem feito entrega mais? Pode haver conflito de interesses nos dados como bônus de performance? Há conflito de interesses entre as áreas? Qual a importância do dado para cada área?
Precisão	Qual a realidade analisada dos dados? Qual o desejo real do meu cliente? É provável que haja algum viés em relação ao input dos dados causando imprecisão? O viés é causado por pessoas? Os dados são originados da e pesquisa de mercado? Qual a metodologia utilizada na coleta? Qual o grau de objetividade e subjetividade dos dados? Quais os critérios considerados para a análise dos dados na tomada de decisão? O viés é mesmo ruim ou será que conhecimento e experiência podem ajudar? Há viés na geração ou interpretação dos dados? Os dados são vivos/orgânicos? Qual a qualificação de quem gera ou analisa os dados? Os outliers foram desconsiderados? Qual o critério de obtenção do dado? Qual a referência de mercado para cruzar dados internos e externos?
Consistência	Qual é a diversidade de Fontes de dados? Qual a quantidade desses temas que originam os dados? Nosso sistema traz as informações utilizáveis/funcionais? A divergências ocasionadas por sistemas legados? Os atuais indicadores de performance conversam com os sistemas? Foi realizado a comparação com outras frutas?
Completeness	Todos os dados estão disponíveis? Sob a perspectiva da jornada de consumo dos clientes, eu possuo dados correspondentes a todas as etapas? Há total compreensão dessa jornada para tomada de decisão? Quais os dados necessários para compreender essa jornada? Quais são os dados necessários para que é possa se tomar uma decisão acertada com base em dados completos? Será que não há dados em excesso?

APÊNDICE D – Perguntas geradas pelos participantes do *focus group 3*

Etapa 1: Perguntas geradas pelos participantes do *Focus Group 3*

Dimensão	Perguntas
Unicidade	Aplicações distintas apresentam os mesmos dados? Qual delas possuirá o dado de referência? Existe um identificador único por registro como uma chave primária? Foi realizado o processo de reconciliação dos dados? Quando um erro é identificado, o script de geração dos dados é ajustado? Em caso de unificação de bases, existe um processo definido? Nesse caso, os processos de negócio são verificados?
Atualidade	Assim que o dado é carregado, a sua validação de freshness é executada? O dado chegou no momento certo? É o dado na versão mais atualizada? Os dados foram entregues de acordo com o SLA? O provedor dos dados cumpriu o combinado em relação a prazo? Eventualmente, em caso de falha de sistema, existe algum aviso de indisponibilidade de dados? Após alguma instabilidade dos sistemas, as transações realizadas off line são validadas? Qual a periodicidade combinada entre o provedor e consumidor dos dados? São dados real time? D-1? Qual o “atraso” esperado? Qual a sensibilidade de atualização?
Validade	Tem regra de multichек? Os campos dos registros analisados tem o dado que preciso? Os percentuais estão representados com valores de 0 a 100? Qual o profiling do dado? Os limites estão sendo respeitados? O usuário está logado para consumo dos dados? É possível agrupar para validar por partes? Os dados estão limpos? Qual a criticidade dos dados? O catálogo de dados foi verificado? As informações estão mapeadas? Há algum problema crítico decorrente da migração de sistemas? Qual o percentual de certeza que preciso sobre a validade dos dados? Quando identificado algum problema, este é corrigido no sistema legado de origem dos dados? O que tenho no sistema representa a origem? Qual o volume de dados que preciso? Quais as dores preciso resolver com os dados? Os dados representam o fenômeno a ser analisado? Em caso de dados fornecidos por terceiros, a coleta foi feita de forma sistemática?
Precisão	Os dados passaram por uma validação de entrada para garantir a precisão? O target reflete o source? Os processos complementares foram validados? O flat file está correto? Os dados são de origem primária (gerado internamente) ou secundária (fornecido por auditorias contratadas)? Quão preciso o dado é em relação a fonte? Quanto ele reflete a realidade?
Consistência	Há um processo de reconciliação dos dados? O que é crítico a ser considerado na reconciliação? Quando o dado é movimentado entre diferentes bases/sistemas é verificada uma possível mutação? Houve falha em alguma parte do processo durante a mutação de base de dado? Quando ocorre alguma anomalia, há um processo de recuperação dos dados?

	Em caso de anomalia, há um processo de identificação de causa raiz? Há um plano de consistência dos dados que são movimentados entre bases? Exemplo uma análise por amostragem. Há um processo de validação cruzada? As regras de arredondamento afetam grandes somas? São as mesmas entre os diferentes sistemas que se baseiam nos dados analisados?
Completeness	Há registros ausentes? Os campos estão preenchidos? O processo de input de dados foi seguido corretamente? Os identificadores de registro denominados Key Data Element estão preenchidos? É possível compartilhar a análise de acordo com os <i>stakeholders</i> responsáveis pelos dados? Por exemplo, pode ser verificado quantos clientes ou transações destes falta em cada grupo. De modo similar, pode-se realizar o mesmo procedimento com fornecedores. No caso de uma empresa que atua em diferentes mercados, agrupados por país ou região, é possível verificar se todos os mercados estão na base de dados? Qual o percentual de elementos de cada grupo está presente na base de dados? A base de dados possui registros referentes a todas as marcas comercializadas?

APÊNDICE E - Sentenças geradas nos *focus groups*

Dimensão	Sentenças
Atualidade	1. Os dados possuem uma periodicidade de atualização adequada ao seu propósito de uso. 2. A frequência de atualização dos dados (mensal, trimestral, etc.) está alinhada com as necessidades do negócio. 3. Os dados utilizados são classificados corretamente como históricos ou preditivos, conforme o contexto da análise. 4. Os dados são disponibilizados com antecedência suficiente para análise e tomada de decisão. 5. Quando os dados são usados para fins de venda, há um histórico confiável disponível. 6. Quando os dados são utilizados para novas aquisições, há previsões atualizadas disponíveis. 7. O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados. 8. A data de corte dos dados é coerente com o período de análise considerado. 9. O histórico de dados analisado é suficiente para gerar projeções confiáveis para o futuro. 10. Existe percepção clara da sazonalidade nas informações analisadas. 11. As comparações entre dados consideram períodos equivalentes (mesma faixa de tempo). 12. A sazonalidade é considerada adequadamente na análise dos dados. 13. O impacto potencial dos dados está diretamente relacionado à sua atualidade. 14. A recorrência de atualização dos dados é bem definida e mantida conforme o esperado. 15. Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados. 16. Mudanças no tempo ou no contexto são levadas em conta nas análises temporais dos dados. 17. Existe um responsável claramente definido pela atualização dos dados. 18. Potenciais interesses ocultos podem influenciar na manipulação ou atraso dos dados. 19. A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente. 20. A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações. 21. Há capacidade de restringir o volume de dados atualizados para focar em informações mais estratégicas (ex.: produtos mais lucrativos). 22. A omnicanalidade contribui para a atualização simultânea dos dados em diferentes canais. 23. Assim que os dados são carregados, um processo de validação de atualidade (freshness) é realizado. 24. Os dados são entregues no momento adequado para seu uso. 25. A versão mais atualizada dos dados é sempre utilizada nas análises. 26. Os dados foram entregues dentro do prazo acordado (SLA). 27. O provedor dos dados cumpriu os prazos combinados para entrega. 28. Em caso de falhas de sistema, há mecanismos de aviso sobre a indisponibilidade dos dados. 29. Após eventuais instabilidades, os dados gerados off-line são validados e integrados corretamente. 30. A periodicidade acordada entre quem fornece e quem consome os dados é respeitada. 31. Os dados são atualizados em tempo real, ou conforme o modelo estabelecido (ex.: D-1). 32. A sensibilidade à atualização dos dados é considerada nos processos críticos do negócio.
Validade	1. Os dados contábeis seguem normas regulatórias, como o padrão IFRS, quando aplicável. 2. Os dados utilizados estão em conformidade com normas e políticas estabelecidas. 3. Os dados seguem um padrão bem definido que facilita sua interpretação. 4. As informações apresentadas fazem sentido dentro do contexto analisado. 5. Os dados são utilizados de forma coerente com a finalidade para a qual foram gerados. 6. Há um padrão de referência que permite a comparação entre diferentes conjuntos de dados. 7. Os dados analisados são consistentes com os registros históricos anteriores. 8. Os dados estão no formato adequado para análise e processamento. 9. A fonte dos dados é claramente identificada. 10. A fonte dos dados é considerada confiável. 11. O período de referência dos dados está claramente definido. 12. O objetivo da análise está alinhado com a natureza dos dados utilizados. 13. Há clareza sobre o propósito final da análise com base nos dados disponíveis. 14. Existe correspondência entre os dados de entrada e os resultados esperados. 15. As diferentes fontes de dados se cruzam de forma coerente e válida. 16. O dado utilizado está alinhado com a intenção analítica definida. 17. A natureza do dado (ex.: faturamento com ou sem imposto) é explicitamente identificada.

	18. As restrições de uso dos dados são conhecidas e respeitadas. 19. O volume de dados disponível é suficiente e apropriado para a análise pretendida. 20. As informações estão apresentadas de forma clara e compreensível. 21. Tenho clareza sobre o significado dos dados que estou utilizando. 22. Os dados passaram por auditoria externa ou revisão formal. 23. Outras áreas da empresa avaliam os dados de forma criteriosa e padronizada. 24. Os dados passaram por validação cruzada entre diferentes fontes ou sistemas. 25. Há esforços para evitar vieses analíticos e garantir que a informação seja válida. 26. Os dados são disponibilizados para consumo conforme esperado. 27. O formato dos dados é compatível com os sistemas de análise e consumo. 28. O processo de integração dos dados é simples e bem estruturado. 29. A simplicidade no tratamento dos dados contribui para maior qualidade e confiança. 30. Conflitos de interesse (ex.: bônus por performance) são considerados ao validar os dados. 31. Conflitos entre áreas da empresa são tratados para não comprometer a validade dos dados. 32. A importância dos dados varia conforme a área da empresa e isso é respeitado. 33. Regras de validação, como multichecks, são aplicadas para garantir a conformidade dos dados. 34. Os campos analisados contêm exatamente as informações necessárias. 35. Percentuais estão representados corretamente dentro do intervalo de 0 a 100. 36. O profiling dos dados foi realizado para entender padrões e variações. 37. Os limites definidos para os dados estão sendo respeitados. 38. O usuário está corretamente autenticado ao acessar ou consumir os dados. 39. É possível agrupar os dados por partes para facilitar sua validação. 40. Os dados estão limpos, sem erros ou ruídos que comprometam sua interpretação. 41. A criticidade dos dados foi avaliada para priorização e controle de qualidade. 42. O catálogo de dados foi consultado para confirmar definições e regras aplicáveis. 43. As informações estão mapeadas em seus respectivos fluxos e origens. 44. Migrações de sistemas não causaram problemas críticos de validade nos dados. 45. O nível de certeza necessário sobre a validade dos dados foi previamente definido. 46. Quando problemas são identificados, há correção no sistema legado de origem. 47. Os dados presentes no sistema refletem com precisão as informações da origem. 48. O volume de dados analisado é proporcional à complexidade da demanda. 49. A análise dos dados considera as dores reais do negócio que precisam ser resolvidas. 50. Os dados representam com fidelidade o fenômeno ou processo que se deseja analisar. 51. Quando os dados são fornecidos por terceiros, a coleta é feita de forma sistemática e confiável.
Unicidade	1. A origem dos dados é claramente identificada e contribui para evitar redundância. 2. Os dados são rastreáveis, o que ajuda a garantir que não existam duplicações. 3. A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro. 4. O fato de a fonte ser qualificada reduz a possibilidade de múltiplas versões de um mesmo dado. 5. Dados com caráter objetivo tendem a ter menor risco de duplicação em diferentes sistemas. 6. A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos. 7. A competência da fonte produtora dos dados influencia na garantia de unicidade da informação. 8. A ocorrência dos mesmos dados em mais de um canal indica possível falha na gestão da unicidade. 9. A qualidade do canal de divulgação dos dados contribui para a percepção de que as informações são únicas. 10. Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única. 11. A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados. 12. É possível cruzar diferentes bases de dados sem gerar redundâncias. 13. As áreas da empresa que avaliam os dados consideram a unicidade como um critério essencial. 14. Existe colaboração entre as áreas da organização para evitar a duplicação de dados. 15. O uso de diversos sistemas dificulta a manutenção da unicidade dos dados. 16. A coleta de dados é feita diretamente da fonte apropriada, reduzindo o risco de duplicações.

	17. A governança de dados contribui para a garantia da unicidade dos registros. 18. O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação. 19. Existe uma estratégia clara para evitar a duplicação de dados nas diferentes fontes. 20. A captura de dados é realizada a partir de uma única fonte confiável. 21. O objetivo da análise de dados influencia na forma como a unicidade dos dados é tratada. 22. A identificação de desvios em dados duplicados é uma prática comum nas análises da organização. 23. A simplificação da análise de dados pode comprometer a verificação da unicidade. 24. Há clareza sobre quais pontos críticos devem ser analisados para garantir dados não redundantes. 25. A consideração de vieses analíticos (ex.: viés de confirmação) pode impactar a percepção sobre a unicidade dos dados. 26. O viés do analista pode interferir na verificação da unicidade dos dados. 27. Aplicações distintas apresentam os mesmos dados, e é definido qual sistema possui o dado de referência. 28. Cada registro possui um identificador único, como uma chave primária. 29. Existe um processo de reconciliação de dados para eliminar inconsistências e duplicidades. 30. Quando um erro é detectado, o script de geração de dados é ajustado para evitar repetições futuras. 31. Em processos de unificação de bases de dados, existe um procedimento formal definido. 32. A verificação dos processos de negócio é realizada ao unificar bases de dados para garantir unicidade.
Precisão	1. É possível identificar tendências reais nos dados analisados, sem distorções. 2. Os dados não apresentam anomalias quando comparados com a base histórica. 3. A fonte fornecedora dos dados é confiável o suficiente para garantir a precisão. 4. O formato dos dados está adequado para evitar erros de interpretação ou leitura. 5. Os dados não estão corrompidos nem apresentam falhas técnicas. 6. A origem dos dados (única ou múltiplas fontes) não compromete sua precisão. 7. Os dados estão perfeitamente alinhados com o objetivo da análise. 8. Os dados são interpretados dentro do contexto correto de uso. 9. Foi realizada uma análise de similaridade que contribui para validar a precisão dos dados. 10. As fontes que disponibilizam os dados estão identificadas e são conhecidas. 11. Mesmo quando existe um padrão de imprecisão, ele pode ser reconhecido e compensado adequadamente. 12. Os dados analisados representam de forma fiel a realidade do fenômeno observado. 13. Os dados refletem corretamente o desejo ou comportamento real do cliente. 14. Pode haver viés na entrada dos dados que compromete sua precisão. 15. A imprecisão dos dados pode ser causada por vieses humanos. 16. Os dados originados de pesquisas de mercado são confiáveis quanto à precisão. 17. A metodologia de coleta foi adequada para garantir dados precisos. 18. O grau de objetividade dos dados é suficiente para minimizar distorções. 19. Os critérios usados na análise dos dados são claros e favorecem a precisão na tomada de decisão. 20. O viés nem sempre é negativo; o conhecimento e a experiência podem contribuir para maior precisão. 21. Pode haver viés tanto na geração quanto na interpretação dos dados. 22. Os dados são considerados vivos (orgânicos), e sua precisão pode evoluir ao longo do tempo. 23. As pessoas responsáveis pela geração ou análise dos dados possuem qualificação técnica adequada. 24. Outliers foram identificados e tratados corretamente para evitar distorções. 25. Os critérios utilizados para a obtenção dos dados foram rigorosos e bem definidos. 26. Referências de mercado foram utilizadas para validar a precisão dos dados internos. 27. Os dados passaram por uma validação inicial (input) que garante sua fidelidade à realidade. 28. Os dados de destino (target) refletem com precisão os dados de origem (source). 29. Os processos complementares foram validados e não comprometem a precisão dos dados. 30. O flat file utilizado está correto e livre de erros que afetem a precisão. 31. É conhecida a origem dos dados — se primária (interna) ou secundária (auditorias externas). 32. O dado reflete com alto grau de precisão a realidade a partir da fonte original.
Compleitude	1. A base de dados não apresenta registros ausentes.

	2. Os campos dos registros estão devidamente preenchidos. 3. O processo de entrada (input) dos dados foi seguido corretamente em todas as etapas. 4. Os campos-chave identificadores (Key Data Elements) estão preenchidos em todos os registros. 5. A análise dos dados pode ser compartilhada com os <i>stakeholders</i> de forma clara, inclusive apontando onde há dados ausentes por grupo (clientes, fornecedores, etc.). 6. Para empresas que atuam em diferentes mercados (ex.: por país ou região), é possível verificar se todos os mercados estão representados na base de dados. 7. O percentual de presença de elementos por grupo (ex.: países, categorias, regiões) é conhecido e controlado. 8. A base de dados contempla registros de todas as marcas comercializadas pela organização. 9. Todos os dados relevantes estão disponíveis para análise e decisão. 10. Sob a perspectiva da jornada do cliente, existem dados representando todas as etapas dessa jornada. 11. Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor. 12. Os dados disponíveis são suficientes para apoiar decisões com base em um conjunto completo de informações. 13. Há clareza sobre quais dados são necessários para se tomar decisões baseadas em informação completa. 14. Existe preocupação com o excesso de dados que, apesar de completos, podem prejudicar a objetividade da análise.
Consistência	1. Os dados são compatíveis com os padrões do mercado quando comparados entre pares equivalentes. 2. A variação dos dados em relação à base histórica é controlada e justificada. 3. Os dados foram avaliados por mais de um especialista, contribuindo para sua consistência. 4. A opinião técnica dos especialistas reforça a confiança na consistência dos dados analisados. 5. As opiniões dos especialistas são ponderadas com critérios claros e contribuem para validação consistente dos dados. 6. Há diversidade de fontes de dados sem comprometer a uniformidade das informações. 7. A quantidade de temas e categorias que originam os dados é conhecida e monitorada. 8. O sistema utilizado disponibiliza os dados em formato funcional e utilizável, sem distorções. 9. Dados provenientes de sistemas legados não geram divergências nos resultados atuais. 10. Os indicadores de performance atuais estão alinhados com os sistemas que os alimentam. 11. Foi realizada uma comparação entre dados de categorias semelhantes (ex.: comparação entre produtos similares), garantindo coerência. 12. Existe um processo estruturado de reconciliação entre diferentes fontes ou bases de dados. 13. Pontos críticos são considerados no processo de reconciliação de dados. 14. Sempre que os dados são transferidos entre sistemas ou bases, é verificada a integridade das informações. 15. Não houve falhas significativas no processo de migração ou movimentação dos dados entre sistemas. 16. Quando ocorrem anomalias nos dados, há um processo de recuperação eficaz. 17. Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema. 18. Existe um plano de consistência para monitorar a movimentação de dados entre bases, como amostragem periódica. 19. A validação cruzada é parte integrante do processo de verificação da consistência dos dados. 20. As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados. 21. Diferenças nas regras de arredondamento entre sistemas não comprometem somas ou análises agregadas.

APÊNDICE F - Sentenças submetidas à análise de juízes

Etapa 3: Análise de juízes

Sentenças	Dimensão Esperada
1. Os dados de destino (target) refletem com precisão os dados de origem (source).	Precisão
2. Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema.	Consistência
3. A origem dos dados é claramente identificada e contribui para evitar redundância.	Unicidade
4. A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações.	Atualidade
5. A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro.	Unicidade
6. A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados.	Unicidade
7. Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados.	Atualidade
8. Percentuais estão representados corretamente dentro do intervalo de 0 a 100.	Validade
9. As comparações entre dados consideram períodos equivalentes (mesma faixa de tempo).	Atualidade
10. As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados.	Consistência
11. A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos.	Unicidade
12. Os dados estão perfeitamente alinhados com o objetivo da análise.	Precisão
13. Os dados são considerados vivos (orgânicos), e sua precisão pode evoluir ao longo do tempo.	Precisão
14. Os dados não estão corrompidos nem apresentam falhas técnicas.	Precisão
15. O período de referência dos dados está claramente definido.	Validade
16. Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única.	Unicidade
17. O uso de planilhas Excel como fonte de dados representa um risco para a unicidade da informação.	Unicidade
18. Sempre que os dados são transferidos entre sistemas ou bases, é verificada a integridade das informações.	Consistência
19. A competência da fonte produtora dos dados influencia na garantia de unicidade da informação.	Unicidade
20. A fonte dos dados é considerada confiável.	Validade
21. Os campos-chave identificadores (Key Data Elements) estão preenchidos em todos os registros.	Compleitude
22. Os dados são disponibilizados para consumo conforme esperado.	Validade
23. Para empresas que atuam em diferentes mercados (ex.: por país ou região), é possível verificar se todos os mercados estão representados na base de dados.	Compleitude
24. O usuário está corretamente autenticado ao acessar ou consumir os dados.	Validade
25. As restrições de uso dos dados são conhecidas e respeitadas.	Validade
26. Pode haver viés tanto na geração quanto na interpretação dos dados.	Precisão
27. A base de dados contempla registros de todas as marcas comercializadas pela organização.	Compleitude
28. O formato dos dados está adequado para evitar erros de interpretação ou leitura.	Precisão
29. A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente.	Atualidade
30. O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados.	Atualidade
31. Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor.	Compleitude

APÊNDICE G – Link para acesso ao produto tecnológico

Este apêndice apresenta o produto tecnológico detalhado na seção 5.

A ferramenta desenvolvida encontra-se no link: https://docs.google.com/spreadsheets/d/15tq_-9_L4V10FBR3nZfh8nx70bW-wa6BSY3y0JEFyM/edit?gid=2041917872

A seguir, a tela da ferramenta e um exemplo de preenchimento:

Mensuração QD					
O presente instrumento, elaborado a partir das dimensões do framework DAMA-DMBOK, permite mensurar de forma padronizada a percepção dos gestores sobre a qualidade de bases de dados sob sua responsabilidade, como cadastros de clientes ou fornecedores, por exemplo. O quadro principal, à esquerda, está organizado em três colunas: "Dimensão Referência", que identifica a dimensão de qualidade, "Sentenças", que apresenta os critérios de análise, e a coluna de resposta, onde o avaliador indica o grau de atendimento da base em relação ao requisito descrito. No quadro RESULTADOS, à direita, são calculadas as médias de cada dimensão avaliada, bem como a média geral das notas de classificação que serve como parâmetro para identificar quais dimensões estão acima ou abaixo desse índice.					
Dimensão Referência	Sentenças	Pontos	RESULTADOS		
Precisão	Os dados de destino (target) refletem com precisão os dados de origem (source).	2	Dimensão	No. Questões	Pontuação Média
Consistência	Em caso de inconsistência, é conduzida uma investigação da causa raiz do problema.	3	Atualidade	4	6,50
Unicidade	A origem dos dados é claramente identificada e contribui para evitar redundância.	5	Completeness	3	5,67
Atualidade	A atualidade dos dados é constantemente desafiada pelo cenário atual de rápidas transformações.	7	Consistência	2	6,00
Unicidade	A confiança na fonte de dados contribui para reconhecer um único registro verdadeiro.	1	Precisão	1	2,00
Unicidade	A adoção de identificadores únicos como RG ou CPF é eficaz para garantir a unicidade dos dados.	3	Unicidade	5	4,20
Atualidade	Fatos relevantes ou eventos externos têm impacto visível na atualização dos dados.	7	Validade	2	1,00
Consistência	As regras de arredondamento são padronizadas entre os diferentes sistemas que utilizam os dados.	9			
Unicidade	A existência de uma fonte de comparação confiável ajuda a detectar e eliminar dados repetidos.	3	Geral	17	4,71
Unicidade	Há uma chave ou campo de identificação claro que permite distinguir cada registro de forma única.	9			
Validade	A fonte dos dados é considerada confiável.	1			
Completeness	Para empresas que atuam em diferentes mercados (ex: por país ou região), é possível verificar se todos os mercados estão representados na base de dados.	7			
Validade	As restrições de uso dos dados são conhecidas e respeitadas.	1			
Completeness	A base de dados contempla registros de todas as marcas comercializadas pela organização.	3			
Atualidade	A digitalização dos processos permite acompanhar os dados em tempo real e, quando necessário, intervir rapidamente.	7			
Atualidade	O objetivo da análise é levado em consideração para definir a atualidade necessária dos dados.	5			
Completeness	Há uma compreensão clara sobre quais dados são necessários para representar todas as fases da jornada do consumidor.	7			